FINAL REPORT

COHERENT CONVERSION OF THE SUNLIGHT SPECTRUM

Prof. T. K. Gustafson

NASA Grant NAG 3-88

July 28, 1980 - April 30, 1982

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

TABLE OF CONTENTS

LIST OF FIGURES AND ABBREVIATED FIGURE CAPTIONS

II-1 a) Energy displacement diagram for a metal-barrier-metal tunneling junction structure.
b) Energy diagram of a d.c. biased metal-barrier-metal tunneling junction structure.

II-2 Energy level diagram for a metal-barrier-semiconductor junction utilized to demonstrate the relationship between stimulated transitions and the current-voltage characteristic.

II-3 Energy Level diagram for a junction structure that would prevent back-flow and hot electron formation. Simpler structures might prove possible. This semiconductor metal structure is discussed in more detail in the present proposal.

III-1 Optical properties of Aluminum: (a) Reflectivity and absorption coefficient data over an extended range (Reprinted from Ref. 6) (b) Frequency dependence of ε_1, ε_2, conductivity σ and the energy loss function - Im ε^{-1}. The effect of the 1.6 eV interband peak on ε_1, ε_2 and σ is obvious. (Reprinted from Ref. 7).

III-2 Schematic diagram of an infinite-extent planar interface between a vacuum and a metal.

III-3 Dispersion of surface plasmons on a planar interface for an ideal metal.

III-4 Schematic diagram of a vacuum-metal-vacuum planar structure.

III-5 Dispersion of the symmetric and the anti-symmetric mode of surface plasmons on a planar ideal metal film.

III-6 Dispersion of the symmetric and antisymmetric mode for a thickness given by $(\frac{\omega_p}{c})d = 0.4$.

III-7 Schematic diagram of a metal-vacuum-metal planar structure.

III-8 Dispersion of the anti-symmetric mode of surface plasmons in a small planar gap between ideal metals and a gap width $(\omega_p/c)d = 0.10$.

III-9 Dispersion of the anti-symmetric mode for a gap width given by $(\omega_p/c)d \simeq 0.04$.

III-10 Propagation length of single surface metal-vacuum plasmons and the metal-vacuum-metal anti-symmetric mode throughout the visible portion of the spectrum. ω_m the plasma frequency was chosen to be 11.0 eV and the relaxation time is 2×10^{-15} sec. for the plotted curve. Addition data for Imβ is indicated at 2.23 eV.

III-11 Numerical solution for the dispersion of an anti-symmetric plasmon mode for a 10 Å barrier junction with the metal plasma frequency, ω_m = 11 eV and ω_s the plasma frequency of the second electrode representative of SnTe (.35eV). The relaxation times chosen are representative of a cooled device (≈10°K).

III-12 Numerical solution for the gap-mode dispersion of Ag-Al_2O_3-Al junctions at room temperature.

III-13 Gap mode dispersion characteristics shown for room temperature Al-Al_2O_3-Al junctions.

IV-1 Various techniques for coupling to the surface plasma mode.

IV-2 Idealized metal-barrier-metal coupling structure.

IV-3 Ideal antenna terminations.

IV-4 Radiation patterns at low and high frequencies for the total radiated power.

IV-5 Coupling of a Gaussiam Beam to a long wire antenna lobe. The antenna is oriented along the z-axis at R=0 and coincides with the focal region of the Gaussian beam.

IV-6 Illustration of reciprocity for evaluating the receiving characteristics of a laser illuminated antenna.

IV-7 Theoretically predicted frequency dependence of the optically induced voltage.

IV-8 Schematic diagram of the heterodyne diagnostic experiment.

IV-9 Directions of first two lobes in the 9.3 μm antenna pattern.

IV-10 Equivalent circuit for antenna coupling to the load and the I.F. connection.

IV-11 Planar junction current-voltage characteristics.

IV-12 Photographs of metal-barrier-metal tunneling devices fabricated at Berkeley.

V-1 M-O-P-O-M device with interdigital electrode structure.

Final Report NASA Grant No. N.A.G. 3-88

ABSTRACT

Efforts related to the utilization of tunneling junctions for the conversion of radiative power to electrical power are reported. The theoretical foundations for these particular devices is presented along with a discussion of many of the practical problems associated with the implementation of such devices.

I INTRODUCTION

The enclosed report is sumbitted as an in depth review of work supported by N.A.S.A. grant number NAG 3-88. It in essence is the culmination of a number of years of investigation into the utilization of rectifying contacts for the purpose of converting either a monochromatic or broadband (solar) spectrum into electrical current with efficiencies approaching those achievable or exceeding low frequency classical rectifiers. Although such conversion efficiencies are not achievable in terms of presently available devices, the theoretical understanding and resulting predictions are beginning to indicate junction and coupling requirements which would have to be achieved to demonstrate the possibility. Recently also related areas of development, such as surface enhanced Raman generation, have stimulated interest in the coupling to microstructures that should prove relevant, particularly with regards to the coupling of the incident radiation field to the junction devices. The study of tunneling devices for conversion should also play a relevant role for p-n junction and other devices which have already demonstrated conversion efficiencies in excess of 10%. Such questions as the role of phonons and the influence of phonon collisions on the open circuit voltage problem might be illuminating from a fully quantum mechanical point of view as is necessary when tunneling is considered. Also since tunneling junctions have already been demonstrated to be useful with various types of cells, the question is relevant.

The report is organized in a manner which presents first of all the theoretical foundations of some of the quantum mechanical aspects of the problem. These are particularly important for many of the concepts which will be necessary to pursue. This section is the basis of an understanding of the relation between the quantum mechanical and classical picture of the rectifying process and in particular its relationship to the multiphoton theory of rectification. At optical frequencies this ties together the transition rate across the junctions and its relationship to the elastic current voltage characteristic. This is based upon a stationary state model of tunneling which provides the simplest way of tying the photon assisted tunneling theory of Tien and Gordon to the detailed interactions which occur in microjunction devices. Much of this effort is continuing

since the connections made are recent results.

In the second section of the report we discuss the particular modes which will be of interest in operating devices. This involves the solution of dispersion relations for the modes of the micro-structure junctions and in particular, the derivation of the loss coefficient associated with some of these modes. This is a particularly crucial problem since the modes turn out to be extremely lossy and in addition travel very slowly. An understanding of this dispersion curve for various frequency regions is important.

Open surface (one metal and an overlaying dielectric) modes are also discussed and although they have more reasonable propagation lengths and loss coefficients, they still are highly lossy.

In section IV of the report we discuss the coupling of open surface, free space and confined or gap modes. This involves basically three types of structures, the prism or grating, antenna and aperture, or, taper. Of these the first is very difficult because of the short periodicities required. The second is probably feasible at optical frequencies in spite of the short propagation lengths required. Taper coupling although difficult in terms of present day fabrication techniques, should become feasible and could eventually be incorporated into an antenna coupling scheme.

The next two sections of the report pertain to experiments which have been undertaken to investigate some of the underlying principles and coupling techniques. This begins with Chapter IV, section 5 describing a heterodyne diagnostic experiment. In Chapter V the fabrication of our junction devices which we are attempting to utilize for most of these experiments is described.

We conclude the report with our perception of the prospects of obtaining practical devices which could display a relatively high conversion efficiency. We conclude that the theoretical possibility is present, however the fabrication techniques as well as the ability to obtain junctions exhibiting the desired characteristics are not yet at hand. Many of the ideas developed in the present report have been extended more recently and appear in the renewal proposal. The overall opinion is that whereas a few years ago it was uncertain as to how tunneling junctions were

managing to mix signals together, to couple the radiation field into the junction, and to exhibit the bias dependencies which they have, it is now felt that a rather thorough theoretical understanding of the physical principles involved, the coupling, the propagation of the field along surfaces and into the junction, and the rectification characteristics of the junction devices has been obtained. The main problems requiring solution are those related to the fabrication of the structures required both for the tunneling and for the coupling in order to experimentally demonstrate that the theoretical modelling is correct. These problems are a challenge to microfabrication techniques which are presently evolving, such as molecular beam epitaxy, electron lithography, and chemical vapor deposition.

The conclusion is that three aspects must emerge. First is the quantum mechanical basis of rectification theory and the inclusion of aspects such as phonon collisions and electron-electron collisions. Second, a junction displaying asymmetry (no back current) must be theoretically analyzed and be experimentally fabricated. Lastly, coupling devices for the radiation field must be fabricated to demonstrate the possibility of significant deposition of energy into the region of the junction.

While this report contains results pertaining to the direct conversion of the electromagnetic field to electron current, some of the basic findings have resulted from efforts associated with related investigations of the coupling between the free space radiation field and surface electromagnetic waves and the coupling of the latter to electron tunneling currents. These sources have been NSF Grant No. ECS-7923877-02 pertaining to the investigation of tunneling devices for optical communication, the Joint Services Electronics Program which has provided support for submillmeter and infrared mixing and detection and in particular the comparison of metal-barrier-metal tunneling devices with other types of junction structures. The Jet Propulsion Laboratories has provided support for surface wave amplification devices which utilizes the same basic expressions as for absorption. They have also provided some support for conversion work which has been described in a final report submitted and does not significantly overlap the presently reported effort.

With regard to notation, every attempt has been made to be as consistent and as clear as possible. The equation referencing has been made simple where possible. Where a single number is found reference is to an equation within the same section. If two numbers are given the first is to the section number of the same chapter and the second the equation number. All three numbers are given when reference is made to a different chapter. In section (IV.2), the section number was deleted.

In final printing it was noticed that in some cases although every attempt was made to retain a consistent notation throughout, there is in some cases double notation. The most important of these we list below:

1) β or k is used to represent the propagation vector component in the direction of propagation of the surface electromagnetic mode.
2) $\beta = \beta' - i\beta'' = k = k' - ik''$
3) γ as well as ν is used to represent the propagation vector component in the direction transverse to that of the propagation of the mode.
4) ξ or δ is used to represent the relaxation time of the free electron cloud.
5) Barrier thickness is given by s, d, or b.
6) The elastic junction current is given by $\langle(VO)\rangle$ in Sec. II.2. I_{el} in Sec. II.3 and simply I in Sec. II.5.
7) References are indicated by [] or ().

II THEORETICAL CONSIDERATIONS FOR RECTIFICATION

II-1 Introduction

For the application of tunneling devices for the purposes of converting between modes of the radiation field and low frequency modes of lumped circuit devices, a suitable model describing the tunneling process is required. The model must include the effects of photon assisted tunneling as a classical analysis does not apply when the photon energy is large on the scale of the nonlinearity. The current theory, known as the transfer Hamiltonian model, has the dual disadvantage of being inapplicable in the thin barrier limit, at least for practical considerations, and an inability to treat barrier height modulation produced by the radiation field. A theoretical approach of Tien and Gordon, and used by Tucker relates the inelastic current resulting from an interaction of the radiation field to the low frequency elastic current. This relationship will become central for calculations of the potential efficiencies available. However, it was deduced not by considering the detailed interaction within the barrier region but by discussing the relative modulation of the wave functions on one side of the barrier with respect to those on the other side of the barrier.

We have reviewed this problem and have been able to tie the results of Tien and Gordon to the interaction itself. This has given us an accurate picture of the interaction process and in particular, its influence upon the barrier. We can thus include processes which do not necessarily result in the Tien-Gordon formulas and which will become important in discussing the conversion problem.

This particular approach involves an extension of the stationary state model[1-7] to the case of an a.c. bias excitation, and does not suffer from the above limitations.

II-2 The Stationary State Theory of Tunneling

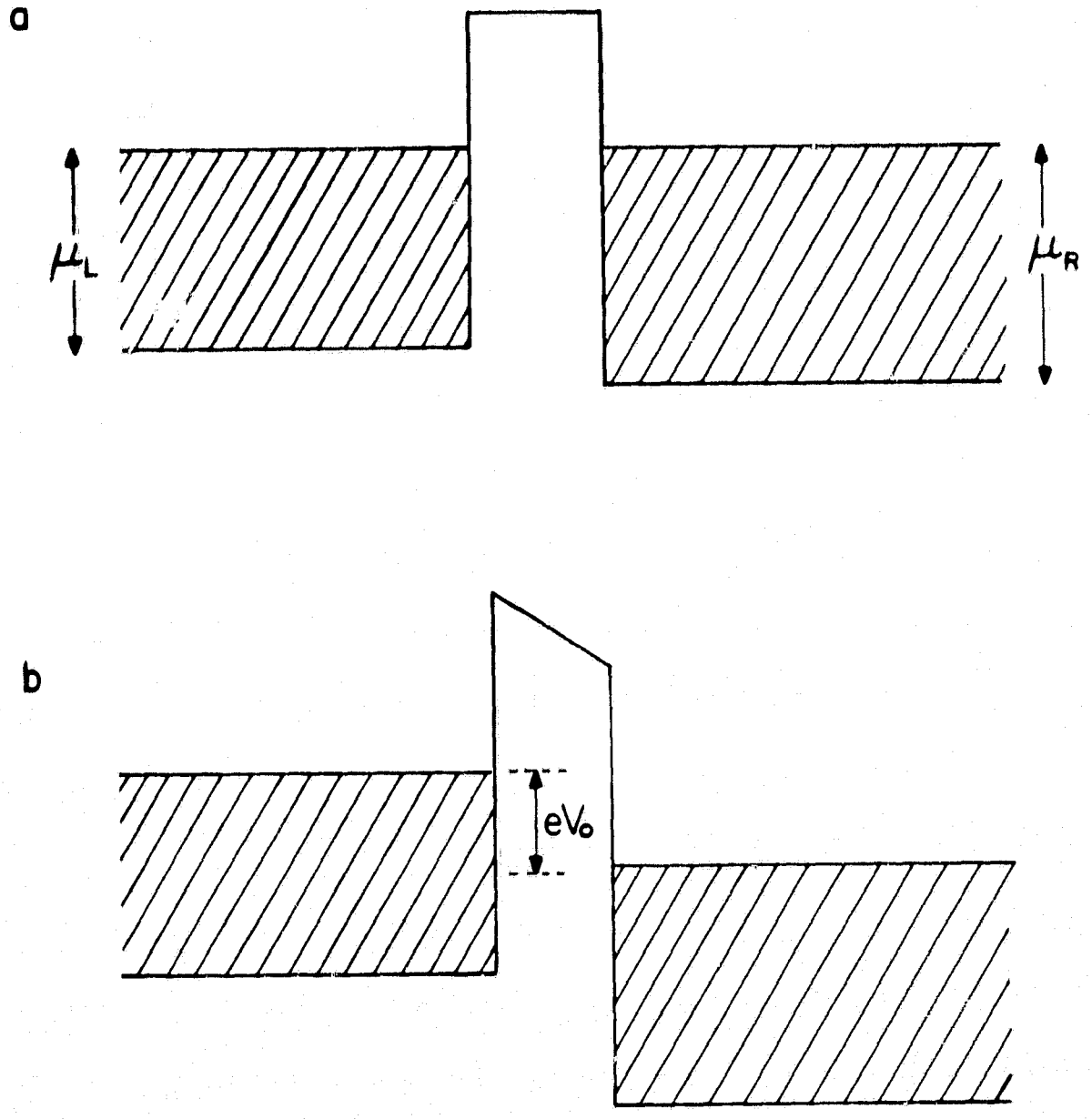
a)Static Bias Case

In the stationary state approach, one initially finds the eigenstates of the Hamiltonian describing the complete structure. The exact form of the barrier potential is a complicated and generally unknown quantity, and

is a strong function of the material properties and interface quality. One generally has to assume a form for this potential, which represents a phenomenological input. It is true, however, that the tunneling probability is relatively independent of the barrier shape, and one can generally assume a phenomonological barrier height and width for an assumed rectangular barrier. A rectangular barrier allows straightforward determination of the eigenstates, although closed form solutions exist for triangular, trapezoidal, parabolic and perhaps other simple geometries. For the treatment of field emission, for instance, a triangular barrier is most appropriate. Furthermore, for arbitrary barrier shapes with gently sloping interfaces, the WKB approximation[8] can be employed.

Initially this chapter will restrict discussion to tunneling between free electron metals. This is not essential and certainly other junction configurations will be important for conversion devices. The metals are characterized by the chemical potential μ, which governs the density of electrons in such a manner as to preserve charge neutrality with the positive lattice in thermal equilibrium. When assembled into the tunneling structure, the Fermi levels must coincide in thermal equilibrium (this condition actually defines thermal equilibrium). The energy difference between the conduction band edges of the two metals is therefore well defined, as is the height of the effective barrier as measured, e.g. from one conduction band edge. The structure is shown schematically as an energy-displacement diagram in Fig. II-1a. An applied bias potential will shift the Fermi levels with respect to each other by an amount eV_0. This assumption is due to the large effective screening in the electrodes, which implies that the electrons will "see" the field due to the applied potential only in the barrier region. This further assumes that there is appreciable reduction of the electron wavefunction amplitude in the barrier, i.e. the charge density and hence the screening are much lower there than in the electrodes. The energy diagram of a d.c. biased (conventionally positive) junction is shown in Fig. II-1b. From this model, one can immediately write down the structure Hamiltonian from which the eigenstates are to be determined.

If interactions between the electrons are neglected, the eigenfunctions can be separated into two disjoint subsets, one associated with each electrode, and treated as independent thermal ensembles characterized



II-1 a) Energy displacement diagram for a metal-barrier-metal tunneling junction structure.

b) Energy diagram of a d.c. bias metal-barrier-metal tunneling junction structure.

by the left and right Fermi energies. Therefore, once one assigns a suitable basis of structure eigenstates, the expectation of observables such as current density can be computed.

In the ideally isolated metal electrodes, each energy level corresponds to a standing wave Bloch state occupied by an opposite spin pair of electrons. The energy density of such states is $\rho_L(\varepsilon)$ and $\rho_R(\varepsilon)$ in the left and right metal, where ε is measured from the respective conduction band edge.

For the real tunneling system, we wish to take into account the fact that for an electronic state which has a large amplitude in the left electrode, there is a finite albeit small amplitude for the state in the right electrode. It is assumed here that the eigenstates have the following form; the states which originally were standing wave states in the left electrode are now matched to a traveling wave in the right electrode carrying flux away from the barrier, and similarly the right electrode states are now matched to a traveling wave in the left electrode carrying flux away from the barrier. The states, of course, are functions of the applied bias voltage V_0 across the barrier. We shall denote the new states corresponding to the original eigenvalues $\{\varepsilon_L\}$ of the left electrode $\{|c_\varepsilon(V_0)\rangle\}$, and those corresponding to the eigenvalues $\{\varepsilon_R\}$ of the right metal $\{|b_\varepsilon(V_0)\rangle\}$. The $\{\varepsilon_L\}$ and $\{\varepsilon_R\}$, measured from the respective conduction band edge, are distributed according to $\rho_L(\varepsilon)$ and $\rho_R(\varepsilon)$.

The Hamiltonian as a function of coordinate x is

$$H = \frac{p^2}{2m} + \phi(x) + eV_0\, f(x), \qquad \text{(II.2.1}$$

with

	$x < 0$	$0 \le x \le b$	$x > b$
$\phi(x)$	δ	$\phi_b(x)$	0
$f(x)$	1	$f_b(x)$	0

(II.2.2

where $\delta = \mu_R - \mu_L$, and $\phi_b(x)$ is a complicated function representing the

effective barrier. The function $f_b(x)$ represents the normalized bias potential in the barrier. We will take the average total barrier height to be $\phi + eV_0/2$, i.e. the Hamiltonian from which the eigenstates will be defined as

	$x < 0$	$0 < x < b$	$x > b$	
$H - \frac{p^2}{2m}$	$eV_0 + \delta$	$\frac{eV_0}{2} + \phi$	0	(II.2.3)

Eigenenergy of H: $\varepsilon + eV_0 + \delta$, for $|C_\varepsilon(V_0)\rangle$ and for $|b_\varepsilon(V_0)\rangle$

It should be reiterated that although a rectangular barrier is assumed other approximations, e.g. WKB, may be used.

For high bias voltage $V_0 \approx \phi$, the barrier potential is more appropriately described as triangular or trapezoidal, hence the assumption that $eV_0 < \phi$ for these eigenstates. Structure energy levels are referenced (arbitrarily) to the conduction band edge of the right metal.

The unnormalized (~) eigenstates have the form

Eigenstate	$x < 0$	$0 \le x \le b$
$\|\tilde{c}_\varepsilon(V_0)\rangle$	$e^{ik_1x} + R_\varepsilon e^{-ik_1x}$	$C_\varepsilon e^{-Kx} + D_\varepsilon e^{K(x-b)}$
$\|\tilde{b}_{\varepsilon+eV_0+\delta}(V_0)\rangle$	$T'_{\varepsilon+eV_0+\delta} e^{-k_1x}$	$C'_{\varepsilon+eV_0+\delta} e^{K(x-b)} + D'_{\varepsilon+eV_0+\delta} e^{-Kx}$

$$x > b$$

$$T_\varepsilon e^{ik_3(x-b)}$$

$$e^{-ik_3(x-b)} + R'_{\varepsilon+eV_0+\delta} e^{ik_3(x-b)} \qquad \text{(II.2.4)}$$

where

$$k_1 = \sqrt{2m\varepsilon/\hbar^2}$$

$$K = \sqrt{2m(\phi - eV_0/2 - \delta - \varepsilon)/\hbar^2}$$

$$k_3 = \sqrt{2m(\varepsilon + eV_0 + \delta)/\hbar^2} \qquad \text{(II.2.5)}$$

and it may be assumed $\varepsilon \in \{\varepsilon_L\}$ for $|c_\varepsilon(V_0)\rangle$ states, and $\varepsilon + eV_0 + \delta \in \{\varepsilon_R\}$ for $|b_{\varepsilon+eV_0+\delta}(V_0)\rangle$ states.

Matching amplitudes and derivatives at the boundaries x = 0 and x = b implies

$$C_\varepsilon = \frac{2ik_1(K-ik_3)}{F_\varepsilon} \quad , \quad D_\varepsilon = C_\varepsilon \frac{K+ik_3}{K-ik_3} e^{-Kb}$$

$$R_\varepsilon = \frac{(K+ik_1)(K-ik_3) - e^{-2Kb}(K-ik_1)(K+ik_3)}{F_\varepsilon} \quad , \quad T_\varepsilon = \frac{4ik_1 K e^{-Kb}}{F_\varepsilon} \qquad \text{(II.2.6)}$$

where

$$F_\varepsilon = -(K-ik_1)(K-ik_3) + e^{-2Kb}(K+ik_1)(K+ik_3). \qquad \text{(II.2.7)}$$

Transforming $k_1 \to k_3$, $k_3 \to k_1$, yields the corresponding quantities $C'_{\varepsilon+eV_0+\delta}$, $D'_{\varepsilon+eV_0+\delta}$, $R'_{\varepsilon+eV_0+\delta}$ and $T'_{\varepsilon+eV_0+\delta}$. The wave functions have the important properties:

$$k_1(1-|R_\varepsilon|^2) = k_3|T_\varepsilon|^2 \qquad \text{(II.2.8)}$$

$$k_3(1-|R'_{\varepsilon+eV_0+\delta}|^2) = k_1|T'_{\varepsilon+eV_0+\delta}|^2 \quad , \qquad \text{(II.2.9)}$$

and we observe also

$$F_\varepsilon = F'_{\varepsilon+eV_0+\delta} \quad , \qquad \text{(II.2.10)}$$

so

$$T_\varepsilon = \left(\frac{k_1}{k_3}\right) T'_{\varepsilon+eV_0+\delta}$$

The states derived have appeared in the literature previously[4,9]. A clear discussion of orthogonality, however, has not been given and is important for the following a.c. interaction considerations. One might assume that the wavefunctions are orthogonal outside of the barrier region, where they assume plane wave character, however this is inaccurate since the barrier region will always contribute an overlap. It might then be argued that the overlap would become insignificant as the normalization length tends toward infinity, as the overlap is independent of the normalization length and hence the contribution should decrease as 1/L. This argument does not provide an adequate degree of orthogonality, as the overlap is always of the same degree (1/L) as the current operator martix elements. It can, in fact, be shown that in the continuum limit the wavefunctions are exactly orthogonal.

Consider for example orthogonality among the |c> states. Assuming an infinite structure and using $\int_0^\infty e^{i\Delta x}dx = \pi\delta(\Delta) + iP(\frac{1}{\Delta})$, one has

$$\langle \tilde{c}_{\varepsilon'}(V_0)|\tilde{c}_\varepsilon(V_0)\rangle = \pi(1+|R_\varepsilon|^2)\delta(k_1-k_1') + \pi|T_\varepsilon|^2\,\delta(k_3-k_3')$$

$$-i(1-R^*_{\varepsilon'}R_\varepsilon)\,P\left(\frac{1}{k_1-k_1'}\right)-i(R^*_{\varepsilon'}-R_\varepsilon)\,P\left(\frac{1}{k_1+k_1'}\right) + iT^*_{\varepsilon'}T_\varepsilon\;P\left(\frac{1}{k_3-k_3'}\right)$$

$$+ \text{barrier contribution} \qquad \text{(II.2.12)}$$

It is arduous but straightforward to verify that the principal part terms cancel the barrier contribution. Using

$$\delta(k_3-k_3') = \delta(k_1-k_1')\frac{\partial k_1}{\partial k_3} = \frac{k_3}{k_1}\delta(k_1-k_1') \qquad \text{one obtains}$$

$$\langle \tilde{c}_{\varepsilon'}(V_0)|\tilde{c}_\varepsilon(V_0)\rangle = 2\pi\delta(k_1-k_1') \qquad \text{(II.2.13)}$$

Similarly one may show

$$\langle \tilde{b}_{\varepsilon'}(V_0) | \tilde{b}_{\varepsilon}(V_0) \rangle = 2\pi\delta(k_3-k_3') \tag{II.2.14}$$

and as $T_{\varepsilon'}'^{*}R_{\varepsilon}\delta(k_1-k_1') + R_{\varepsilon'}'^{*}T_{\varepsilon}\delta(k_3-k_3') = 0$, one has

$$\langle \tilde{b}_{\varepsilon'}(V_0) | \tilde{c}_{\varepsilon}(V_0) \rangle = 0 \tag{II.2.15}$$

Orthogonality holds exactly only in the continuum limit, but the discrete formalism developed in this chapter is applicable with the understanding that evaluation of expressions should be done in the L→∞ limit. Adopting the normalization $\langle c_{\varepsilon}(V_0)| = \frac{1}{\sqrt{2L}}\langle \tilde{c}_{\varepsilon}(V_0)|$, $\langle b_{\varepsilon'}(V_0)| = \frac{1}{\sqrt{2L}}\langle \tilde{b}_{\varepsilon}(V_0)|$, which is usual, one may return to the summation notation with normalization

$$\langle c_{\varepsilon'}(V_0) | c_{\varepsilon}(V_0) \rangle = \delta_{k_1,k_1'} \tag{II.2.16}$$

$$\langle b_{\varepsilon'}(V_0) | c_{\varepsilon}(V_0) \rangle = \delta_{k_3,k_3'} \tag{II.2.17}$$

and we have the associations $\frac{\pi}{L}\delta(k_1-k_1') = \delta_{k_1,k_1'}$, $\frac{\pi}{L}\delta(k_3-k_3') = \delta_{k_3,k_3'}$ as L→∞. In a practical sense, going to the continuum limit involves only changing summations to integrations in the usual manner, i.e. $\sum_{\varepsilon} \rightarrow \int \rho(\varepsilon)d\varepsilon$, and taking the domain of the wavefunctions to be $(-\infty,\infty)$. The change has no effect on the current operator matrix elements.

Charge neutrality in equilibrium is maintained, i.e., the electron density in each metal is the intrinsic density at zero bias. For instance, in the isolated left metal for a given energy the electron density is $n(k) = \frac{2L}{\pi}$. For the left metal in the structure with $V_0 = 0$ we have $n(k) = \frac{1}{2}\left((1+|R_{\varepsilon}|^2 + \frac{k_1}{k_3}|T_{\varepsilon+\delta}'|^2)\frac{2L}{\pi} = \frac{2L}{\pi}\right)$. This is true in the right metal as well.

The current operator is

$$j = \frac{ie\hbar}{2m}(\overleftrightarrow{\nabla}\delta(x-x')-\delta(x-x')\overleftrightarrow{\nabla}). \qquad \text{(II.2.18)}$$

where x' is now taken to be the coordinate variable upon which the wave functions depend and x is the position of observation.

Evaluating the matrix elements, one finds

$$\langle c_\varepsilon(V_0)|j|c_\varepsilon(V_0)\rangle = \frac{e\hbar}{2m}(k_3+k_3^*)\cdot\frac{16K^2k_1^2e^{-2Kb}}{2L|F_\varepsilon|^2} \qquad \text{(II.2.19)}$$

$$\langle b_{\varepsilon+eV_0+\delta}(V_0)|j|b_{\varepsilon+eV_0+\delta}(V_0)\rangle = -\frac{e\hbar}{2m}(k_1+k_1^*)\cdot\frac{16K^2k_3^2e^{-2Kb}}{2L|F_\varepsilon|^2} \qquad \text{(II.2.20)}$$

The fact that the final momentum must be real precludes current flow below the conduction band edge of the electrode into which the electron tunnels. Multiplying the matrix elements by the respective standing wave state densities (including spin degeneracy, and where $\theta(\varepsilon)$ is the Heaviside function)

$$\rho_L(\varepsilon) = \frac{2L}{\hbar^2\pi}\frac{m}{k_1}\,\theta(\varepsilon) \quad , \qquad \text{(II.2.21)}$$

$$\rho_R(\varepsilon+eV_0+\delta) = \frac{2L}{\hbar^2\pi}\frac{m}{k_3}\,\theta(\varepsilon+eV_0+\delta) \qquad \text{(II.2.22)}$$

one has immediately

$$\rho_L(\varepsilon)\langle c_\varepsilon(V_0)|j|c_\varepsilon(V_0)\rangle + \rho_R(\varepsilon+eV_0+\delta)\,\langle b_{\varepsilon+eV_0+\delta}(V_0)|j|b_{\varepsilon+eV_0+\delta}(V_0)\rangle = 0 \quad , \qquad \text{(II.2.23)}$$

the condition of detail balance. This means that for any structure energy level, if both $|c_\varepsilon(V_0)\rangle$ and $|b_{\varepsilon+eV_0+\delta}(V_0)\rangle$ have equal occupation probabilities, then there is no net current across the barrier. This

ensures that there is no net current flow at zero bias.

In the second quantized form,

$$H = \sum_{\varepsilon_L} (\varepsilon + eV_0 + \delta) c_\varepsilon^\dagger c_\varepsilon + \sum_{\varepsilon_R} \varepsilon\, b_\varepsilon^\dagger b_\varepsilon, \tag{II.2.24}$$

where the creation operators $c_\varepsilon^\dagger$, $b_\varepsilon^\dagger$ create $|c_\varepsilon(V_0)\rangle$ and $|b_\varepsilon(V_0)\rangle$ eigenstates with eigenvalues $\varepsilon + eV_0 + \delta$ and ε, respectively, i.e.,

$$H|c_\varepsilon(V_0)\rangle = (\varepsilon + eV_0 + \delta)|c_\varepsilon(V_0)\rangle \tag{II.2.25}$$

$$H|b_\varepsilon(V_0)\rangle = \varepsilon|b_\varepsilon(V_0)\rangle . \tag{II.2.26}$$

One generally utilizes the complete set of eigenstates to form unperturbed Green's functions, as this facilitates the treatment of, e.g., phonon interactions [4]. Here all interactions are neglected, thus there is no need for the introduction of Green's functions at this point.

The time dependence of the field operator is written explicitly

$$\psi = \sum_{\varepsilon_L} c_\varepsilon |c_\varepsilon(V_0)\rangle\, e^{-i(\varepsilon + eV_0 + \delta)t/\hbar} + \sum_{\varepsilon_R} b_\varepsilon |b_\varepsilon(V_0)\rangle e^{-i\varepsilon t/\hbar} \tag{II.2.27}$$

so that ψ satisfies $i\hbar\dot{\psi} = [H,\psi]$ (Heisenberg picture) and the creation operators are time invariant.

It should be pointed out that here Dirac notation is used to represent the Schrodinger wave function.

Closure of the Dirac brackets implies an integration over the coordinate variable (x') upon which the wave functions depend. The density operator is

$$\rho = \frac{1}{Z} e^{-\beta K}, \tag{II.2.28}$$

where

$$K = H - \sum_{\varepsilon_L} (\mu_R + eV_0) c_\varepsilon^\dagger c_\varepsilon - \sum_{\varepsilon_R} \mu_R b_\varepsilon^\dagger b_\varepsilon, \tag{II.2.29}$$

and

$$Z = \mathrm{Tr}\, e^{-\beta K} \quad , \beta = \frac{1}{k_B T} \tag{II.2.30}$$

Thus

$$K = \sum_{\varepsilon_L} (\varepsilon - \mu_L) c_\varepsilon^\dagger c_\varepsilon + \sum_{\varepsilon_R} (\varepsilon - \mu_R) b_\varepsilon^\dagger b_\varepsilon \quad . \tag{II.2.31}$$

This expression for K shows that the bias dependence is in the states, i.e., the density operator is form invariant with changing bias. The second quantized current operator is $J = \psi^\dagger j \psi$, or

$$\begin{aligned} J = & \sum_{\varepsilon_L, \varepsilon_L'} c_\varepsilon^\dagger c_{\varepsilon'} \langle c_\varepsilon(V_0)|j|c_{\varepsilon'}(V_0)\rangle e^{-i(\varepsilon'-\varepsilon)t/\hbar} \\ & + \sum_{\substack{\varepsilon' \\ R'}} \sum_{\substack{\varepsilon \\ R}} b_\varepsilon^\dagger b_{\varepsilon'} \langle b_\varepsilon(V_0)|j|b_{\varepsilon'}(V_0)\rangle e^{-i(\varepsilon'-\varepsilon)t/\hbar} \\ & + \sum_{\varepsilon_L, \varepsilon_R'} c_\varepsilon b_{\varepsilon'+eV_0+\delta} \langle c_\varepsilon(V_0)|j|b_{\varepsilon'+eV_0+\delta}(V_0)\rangle e^{-i(\varepsilon'-\varepsilon)t/\hbar} \\ & + \sum_{\varepsilon_L', \varepsilon_R} b_{\varepsilon+eV_0+\delta} c_{\varepsilon'} \langle b_{\varepsilon+eV_0+\delta}(V_0)|j|c_{\varepsilon'}(V_0)\rangle e^{-i(\varepsilon'-\varepsilon)t/\hbar} \end{aligned} \tag{II.2.32}$$

Spin indices are suppressed for notational simplicity, as the Hamiltonian and density matrix are trivially spin diagonal. The current operator matrix elements are independent of spin, hence the summations over $\varepsilon, \varepsilon'$ imply additional summations over independent spin variables. In this chapter all spin summations can be performed by including spin degeneracy in the density of states, thus one may adopt degenerate state densities and neglect spin throughout.

For the static bias case at hand, the current operator may be evaluated anywhere in the structure, and the total current $J(V_0) = \mathrm{Tr}(\rho J)$ may be

immediately written down.

$$\langle J(V_0)\rangle = \sum_{\varepsilon_L} \frac{e^{-\beta(\varepsilon-\mu_L)}}{1+e^{-\beta(\varepsilon-\mu_L)}} \langle c_\varepsilon(V_0)|j|c_\varepsilon(V_0)\rangle$$

$$+ \sum_{\varepsilon_R} \frac{e^{-\beta(\varepsilon-\mu_R)}}{1+e^{-\beta(\varepsilon-\mu_R)}} \langle b_\varepsilon(V_0)|j|b_\varepsilon(V_0)\rangle , \qquad \text{(II.2.33)}$$

or in the continuum limit

$$\langle J(V_0)\rangle = \int_{-\infty}^{\infty} d\varepsilon \rho_L(\varepsilon) f(\varepsilon-\mu_L) \langle c_\varepsilon(V_0)|j|c_\varepsilon(V_0)\rangle$$

$$+ \int_{-\infty}^{\infty} d\varepsilon \rho_R(\varepsilon) f(\varepsilon-\mu_R) \langle b_\varepsilon(V_0)|j|b_\varepsilon(V_0)\rangle \qquad \text{(II.2.34)}$$

where $f(x) = \frac{e^{-\beta x}}{1+e^{-\beta x}}$ is the Fermi function. Detail balance (Eq. II.2.23) may be used to combine the two integrals. Then,

$$\langle J(V_0)\rangle = \int_{-\infty}^{\infty} d\varepsilon \rho_L(\varepsilon)(f(\varepsilon-\mu_L)-f(\varepsilon+eV_0-\mu_L)) \langle c_\varepsilon(V_0)|j|c_\varepsilon(V_0)\rangle , \qquad \text{(II.2.35)}$$

recalling that $\mu_R - \mu_L = \delta$.

Given $\rho_L(\varepsilon) = \frac{2Lm}{\pi k_1 \hbar^2} \theta(\varepsilon)$,

and evaluating

$$\langle c_\varepsilon(V_0)|j|c_\varepsilon(V_0)\rangle = \frac{4e\hbar k_1^2 K^2 (k_3 + k_3^*) e^{-2Kb}}{m L |F_\varepsilon|^2}$$

one obtains

$$\langle J(V_0)\rangle = \frac{8e}{\hbar\pi} \int_{-\infty}^{\infty} d\varepsilon \frac{k_1 K^2 (k_3 + k_3^*) e^{-2Kb}}{|F_\varepsilon|^2} (f(\varepsilon-\mu_L)-f(\varepsilon+eV_0-\mu_L)). \qquad \text{(II.2.36)}$$

This agrees with the result derived by Harrison [10] to lowest order in the tunneling exponential e^{-Kb}. Fluctuations in the tunneling current[11] can be treated similarly[7] and give the results of Ref. (12) to a good approximation.

b) Nonstatic Bias Excitation

Previous to Ref. (7), the stationary state approach was restricted to static problems, i.e., where the applied bias was time independent. Although scattering from time dependent impurities was considered in a stationary state approach, Ref. (5), the response of tunnel junctions in the presence of a radiation field (effective nonstatic bias) was accomplished only in the transfer Hamiltonian format[13,14,15]. In this section, the stationary state method of Ref. (7) for photon assisted tunneling will be described. Due to the quantum nature of the radiation field, expressions for the tunneling current revert to summations over the number of quanta absorbed or emitted into the field, and nonclassical behavior is exhibited. By "classical behavior," we mean the predicted response of the junction when $V_0 \rightarrow V(t)$ in our previously derived expressions in the static bias case. Throughout the present section, comparison will be made to results obtained from the transfer Hamiltonian format, and certain discrepancies will be noted. The discrepancies are the result of the change in barrier height due to the applied field being neglected in previous analyses, as will be described. The ability of the stationary state model to predict the effects of photon assisted tunneling demonstrates the complete utility of this model for independent tunneling charge transport.

The approach described here is to find time varying eigenstates for a structure with a sinusoidal bias component. The current of the time varying eigenstates is then summed in the usual manner, yielding the current response. The source of the time varying bias is not specified, and therefore could be due to a radiation field or plasmon field, within a certain frequency limit to be described. It is assumed that the frequency is low enough so that the density operator is time invariant, which is equivalent to assuming the maintenance of thermal equilibrium in the population of the $\{|c_\varepsilon\rangle\}$ and $\{|b_\varepsilon\rangle\}$ states. We are neglecting bias field penetration into the electrodes so that ε, the kinetic energy of the incident electron, is still a valid label. Approaching visible frequencies, the assumption of rigid motion of the Fermi levels (rigid occupancy

hypothesis (4)) breaks down, as does the assumption of a piecewise linear potential taken for $V(x,t) = V(t)f(x)$. One should use perturbation methods for excitation at these frequencies, but the expressions observed here are still valid to lowest order in the a.c. field amplitude. The density operator therefore retains its form and time dependence is introduced through explicitly time varying states. Consider the $|c_\varepsilon(V_0)\rangle$ states, which satisfy $H|c_\varepsilon(V_0)\rangle = (\varepsilon+eV_0+\delta)|c_\varepsilon(V_0)\rangle$ for static bias V_0. For an explicitly time varying bias voltage $V = V_0 + V_r \cos \omega t$, $|c_\varepsilon(V)\rangle$ will no longer be associated with a simple exponential time dependence, yet if one now includes the time dependence in the state one must have

$$H|c_\varepsilon(V(t))\rangle = i\hbar \frac{\partial}{\partial t} |c_\varepsilon(V(t))\rangle \ . \qquad \text{(II.2.37)}$$

Thus, states $|c_\varepsilon(V(t))\rangle$ must be found such that

$$\left(\frac{p^2}{2m} + \phi(x)+e(V_0+V_r \cos \omega t)f(x)\right)|c_\varepsilon(V(t))\rangle = i\hbar \frac{\partial}{\partial t} |c_\varepsilon(V(t))\rangle \ . \qquad \text{(II.2.38)}$$

If $|c_\varepsilon(V(t)\rangle$ and the corresponding $|b_\varepsilon(V(t)\rangle$ can be found, the a.c. current response neglecting field penetration may be computed without approximation. To date, exact solutions have not been determined, but an approximate solution may be found as follows. One assumes a solution to Eq. (2.38) of the form

$$|c_\varepsilon(V(t))\rangle = \sum_n A_n \, c_\varepsilon^n\rangle \, e^{-i(\varepsilon+eV_0+\delta+n\hbar\omega)t/\hbar} , \qquad \text{(II.2.39)}$$

where the $|c_\varepsilon^n\rangle$ are normalized, time invariant wave functions not necessarily orthogonal on n. Substitution into Eq. (2.38) above implies, because of this time invariance, that

$$\left(\frac{p^2}{2m} + \phi(x) + eV_0 f(x) - (\varepsilon+eV_0+\delta+n\hbar\omega)\right)A_n|c_\varepsilon^n\rangle + \frac{eV_r}{2} f(x)(A_{n+1}|c_\varepsilon^{n+1}\rangle + A_{n-1}|c_\varepsilon^{n-1}\rangle) = 0. \qquad \text{(II.2.40)}$$

Suppose it is required that

$$\left(\frac{p^2}{2m} + \phi(x)+(eV_0+n\hbar\omega)f(x)\right)|c_\varepsilon^n\rangle = (\varepsilon+eV_0+\delta+n\hbar\omega)|c_\varepsilon^n\rangle$$

then one can see that $|c_\varepsilon^n\rangle$ is simply $|c_\varepsilon(V_0+n\hbar\omega/e)\rangle$, corresponding to the static bias case but with bias $V_0+n\hbar\omega/e$. Then one must have

$$-f(x)n\hbar\omega A_n|c_\varepsilon(V_0+n\hbar\omega/e)\rangle$$
$$+\frac{eV_r}{2} f(x)(A_{n+1}|c_\varepsilon(V_0+(n+1)\hbar\omega/e\rangle + A_{n-1}|c_\varepsilon(V_0+(n-1)\hbar\omega/e)\rangle) = 0. \qquad \text{(II.2.41)}$$

Where $f(x)=1$ (the left electrode, $f(x)$ vanishes in the right electrode)

$|c_\varepsilon(V_0+n\hbar\omega/e)\rangle$ has the form $\frac{1}{\sqrt{2L}}(e^{ik_\varepsilon x}+R_\varepsilon^n e^{ik_\varepsilon x})$, where the reflection coefficient R_ε^n is a slowly varying function of n when $\hbar\omega \ll \phi-M_R$. Neglecting this dependence in (2.41), i.e. assuming the above inequality implies,

$$-nA_n + \tfrac{1}{2}\frac{eV_r}{\hbar\omega}(A_{n+1}+A_{n-1}) = 0 \text{ or}$$
$$A_n = J_n\left(\frac{eV_r}{\hbar\omega}\right), \qquad \text{(II.2.42)}$$

J_n being the ordinary Bessel function. The time varying $|c_\varepsilon(V(t))\rangle$ states are thus taken to be

$$|c_\varepsilon(V(t))\rangle = \sum_n J_n\left(\frac{eV_r}{\hbar\omega}\right)|c_\varepsilon(V_0+n\hbar\omega/e)\rangle\ e^{-i(\varepsilon+eV_0+n\hbar\omega+\delta)t/\hbar} \qquad \text{(II.2.43)}$$

i.e., a weighted sum over the previously derived states for the static bias case, where the bias potential differs by some integer number of

photon energies. Recall that $|c_\varepsilon(V_0+n\hbar\omega/e)\rangle$ satisfies

$$\left(\frac{p^2}{2m}+\phi(x)+(eV_0+n\hbar\omega)f(x)\right)|c_\varepsilon(V_0+n\hbar\omega/e)\rangle = (\varepsilon+eV_0+n\hbar\omega+\delta)|c_\varepsilon(V_0+n\hbar\omega/e)\rangle \quad . \tag{II.2.44}$$

Similarly, one can find time varying $|b(V(t))\rangle$ states. The requirement that the transmitted part of the eigenstates be projected out of the recursion relations, so that one may use the approximate invariance of the reflection coefficient on n as above, implies (with hindsight) that a slightly different form must be adopted for $|b_\varepsilon(V(t))\rangle$. A suitable form is

$$|b_\varepsilon(V(t))\rangle = e^{-i\int^t \frac{eV_r}{\hbar\omega}\cos\omega t'dt'} \sum_n B_n|b_\varepsilon^n\rangle\, e^{-i(\varepsilon+n\hbar\omega)t/\hbar} \tag{II.2.45}$$

and upon substituting this into the appropriate Hamiltonian expression one obtains the recursion relation

$$\left(\frac{p^2}{2m}+\phi(x)+eV_0f(x)-\varepsilon-n\hbar\omega\right)B_n|b_\varepsilon^n\rangle + \frac{eV_r}{2}(f(x)-1)(B_{n+1}|b_\varepsilon^{n+1}\rangle + B_{n-1}|b_\varepsilon^{n-1}\rangle) = 0. \tag{II.2.46}$$

If one requires

$$\left(\frac{p^2}{2m}+\phi(x)+(eV_0-n\hbar\omega f(x))\right)|b_\varepsilon^n\rangle = \varepsilon|b_\varepsilon^n\rangle \quad ,$$

then

$$n\hbar\omega(f(x)-1)B_n|b^n\rangle + \frac{eV_r}{2}(f(x)-1)(B_{n+1}|b^{n+1}\rangle + B_{n-1}|b^{n-1}\rangle) = 0. \tag{II.2.47}$$

Where $f(x)-1 = -1$ (the right electrode, $f(x)-1$ vanishes in the left electrode) the states are approximately invariant on n, as it is evident that $|b_\varepsilon^n\rangle = |b_\varepsilon(V_0-n\hbar\omega/e)\rangle$, thus implying that in this limit one should take

$B_n = J_{-n}\left(\frac{eV_r}{\hbar\omega}\right)$. Thus, we adopt

$$|b_\varepsilon(V(t))\rangle = e^{-i\int^t \frac{eV_r}{\hbar}\cos\omega t'dt'} \sum_n J_n\left(\frac{eV_r}{\hbar\omega}\right)|b_\varepsilon(V_0+n\hbar\omega/e)\rangle e^{-i(\varepsilon-n\hbar\omega)t/\hbar} . \tag{II.2.48}$$

Note that $\langle c_\varepsilon(V(t))|c_\varepsilon(V(t))\rangle = \langle b_\varepsilon(V(t))|b_\varepsilon(V(t))\rangle = 1$, so that as stated previously the density operator retains its time independent form, however now the creation operators create the explicitly time varying states. It should be stressed that our solutions are approximate, and are valid only in the limit $\hbar\omega \ll \phi, \mu_L, \mu_R$. This inequality will be assumed for the remainder of this section.

Our concern here is with the current, so the relevant (diagonal) part of the current operator is

$$\begin{aligned} J = &\sum_{\varepsilon_L} c_\varepsilon^\dagger c_\varepsilon \sum_{n,n'} J_n\left(\frac{eV_r}{\hbar\omega}\right)J_{n'}\left(\frac{eV_r}{\hbar\omega}\right) \\ &\langle c_\varepsilon(V_0+n\hbar\omega/e)|j|c_\varepsilon(V_0+n\hbar\omega/e)\rangle\ e^{-i(n-n')\omega t} \\ + &\sum_{\varepsilon_R} b_\varepsilon^\dagger b_\varepsilon \sum_{n,n'} J_n\left(\frac{eV_r}{\hbar\omega}\right)\ J_{n'}\left(\frac{eV_r}{\hbar\omega}\right) \\ &\langle b_\varepsilon(V_0+n\hbar\omega/e)|j|b_\varepsilon(V_0+n\hbar\omega/e)\rangle\ e^{-i(n-n')\omega t}. \end{aligned} \tag{II.2.49}$$

Evaluation of the trace yields in the same manner as for Eq. (2.34) letting $n' = n+m$

$$\begin{aligned} \langle J\rangle = &\sum_{\varepsilon_L,n,m} f(\varepsilon-\mu_L)J_n\left(\frac{eV_r}{\hbar\omega}\right)J_{n+m}\left(\frac{eV_r}{\hbar\omega}\right) \\ &\langle c_\varepsilon(V_0+(n+m)\hbar\omega/e)|j|c_\varepsilon(V_0+n\hbar\omega/e)\rangle\ e^{-im\omega t} \\ + &\sum_{\varepsilon_R,n,m} f(\varepsilon-\mu_R)\ J_n\left(\frac{eV_r}{\hbar\omega}\right)\ J_{n+m}\left(\frac{eV_r}{\hbar\omega}\right) \\ &\langle b_\varepsilon(V_0+n\hbar\omega/e)|j|b_\varepsilon(V_0+(n+m)\hbar\omega/e)\rangle\ e^{-im\omega t} \end{aligned} \tag{II.2.50}$$

The a.c. current is not a conserved quantity, (i.e. $\frac{\partial}{\partial x}\langle J\rangle \neq 0$),

implying the existence of oscillating charge density. For instance, the $|c_\varepsilon\rangle$ matrix elements have sinusoidal spatial modulation in the right electrode. This is due to modulation of the electron wavefunctions on passage through the time dependent potential. For small ω where classical circuit concepts apply, this spatial variation becomes arbitrarily small.

For the m = 0 (d.c.) terms, one can again use detail balance, i.e.,

$$\rho_L(\varepsilon)\,\langle c_\varepsilon(V_0+n\hbar\omega/e)|j|c_\varepsilon(V_0+n\hbar\omega/e)\rangle$$

$$+\rho_r(\varepsilon+eV_0+n\hbar\omega+\delta)\,\langle b_{\varepsilon+eV_0+n\hbar\omega+\delta}(V_0+n\hbar\omega/e)|j|b_{\varepsilon+eV_0+n\hbar\omega+\delta}(V_0+n\hbar\omega/e)\rangle = 0 \qquad \text{(II.2.51)}$$

to combine the two integrals. One obtains

$$\langle J\rangle_0 = \sum_n \int_{-\infty}^{\infty} \rho_L(\varepsilon)d\varepsilon\,(f(\varepsilon-\mu_L)-f(\varepsilon+eV_0+n\hbar\omega-\mu_L))$$

$$\cdot\, J_n^2\left(\frac{eV_r}{\hbar\omega}\right)\langle c_\varepsilon(V_0+n\hbar\omega/e)|j|c_\varepsilon(V_0+n\hbar\omega/e)\rangle \quad , \qquad \text{(II.2.52)}$$

or

$$\langle J\rangle_0 = \sum_n J_n^2\left(\frac{eV_r}{\hbar\omega}\right)\langle J(V_0+n\hbar\omega/e)\rangle \quad , \qquad \text{(II.2.53)}$$

where $\langle J(V_0+n\hbar\omega/e)\rangle$ is the current for static bias potential $V_0+n\hbar\omega/e$. This is identical to Tucker's expression for the d.c. quasiparticle tunneling current derived from the first order transfer Hamiltonian model[15]. It predicts the famous steps in the I-V characteristic due to photon assisted tunneling[13].

Eq. (II.2.53) shows that the index n can be interpreted as the number of photons absorbed in the structure and that the current is a superposition over the possible multi-photon emission and absorption processes. The total absorption rate γ is thus given by the expression

$$\gamma = \sum n J_n^2\left(\frac{eV_r}{\hbar\omega}\right)\; J(V_0+n\hbar\omega/e)/e \qquad \text{(II.2.54)}$$

One can obtain expressions for the a.c. terms as well. This has been done as well as a consideration of the Kramers-Kronig relationship between the imaginery and real parts. Since these are not immediately relevant to the conversion problem they will not be discussed here.

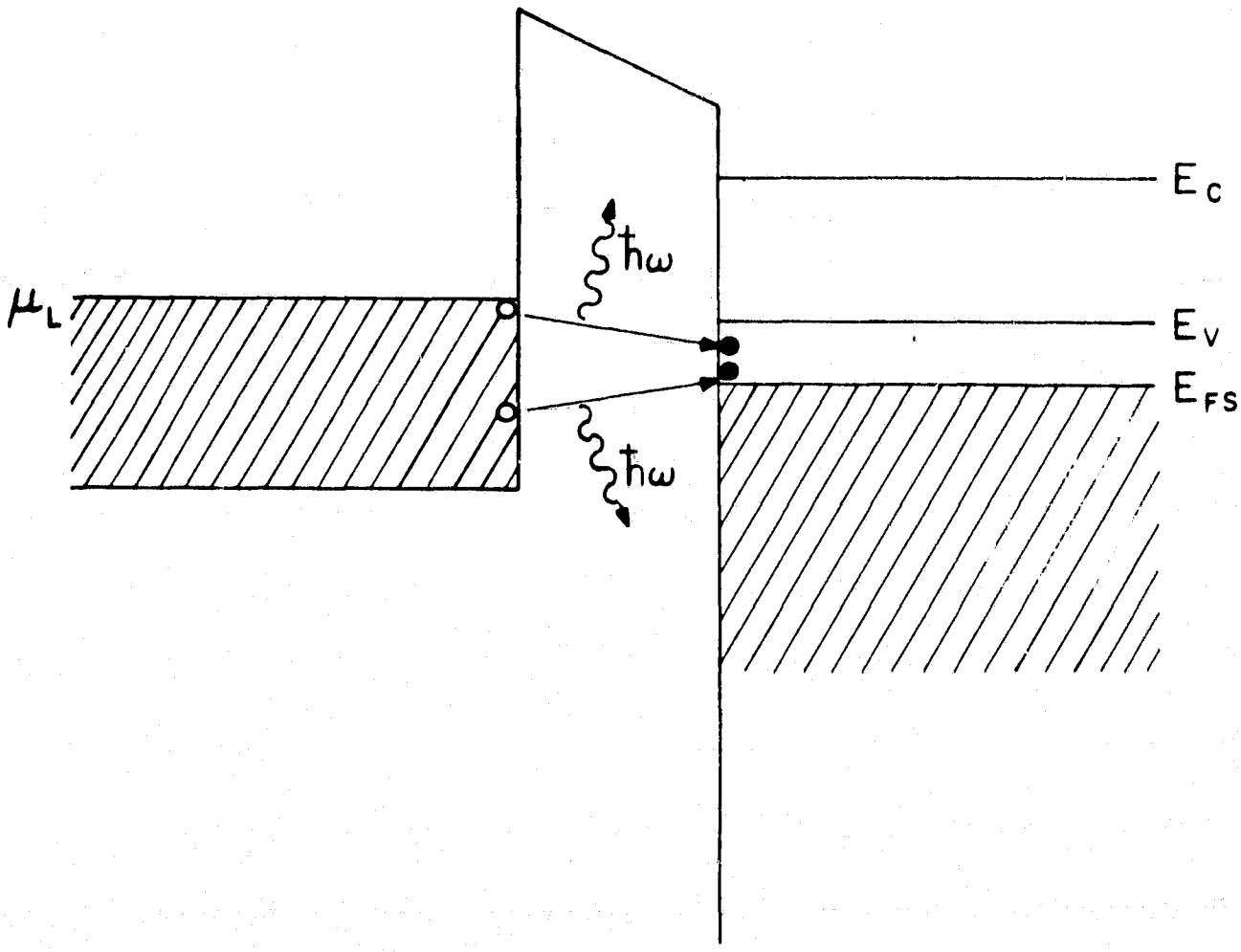
The primary results of the present section are the two above formulae II.2.53 and II.2.54. They tie together the absorption aspects of the problem with the current voltage characteristics. In recent work we have been utilizing these expressions to investigate the conversion efficiency available from a p-n junction as a function of frequency.

The tie in to the transfer Hamiltonian approach is extremely important since it allows one to begin with a microscopic interaction potential and discuss absorption and emission phenomena within the junction region, and within the limits of validity to obtain the Tien-Gordon results, which are obtained by consideration of the interaction outside of the barrier through a phase modulating gauge term.

There are two aspects which allow this to occur. Firstly Eqs. (2.43) and (2.48) show that the n photon absorption and emission processes occur with an effective barrier height equivalent to a shifting of the d.c. potential to $V_0 + (-)\ n\hbar\omega/e$ for the c states and secondly $V_0 - (+) n\hbar\omega/e$ for the b states.

This is what one would expect from an intuitive view as shown in Fig. (II-2) for the bias convention. Considering Eq. (II.2.5), from a transition point of view one can picture the final state as a b-state which for emission has an energy lower than the initial c-state by $n\hbar\omega$ and one might expect the K to differ for the two states. However for the former c-state the value of V_0 must be decreased by $n\hbar\omega/e$. Since the electron is undergoing emission from that state the effective barrier increased by $n\hbar\omega/2$. On the other hand the final b state is first of all lower in energy by $n\hbar\omega$ and absorption occurs from it, thus for K the effective barrier must be increased by $n\hbar\omega/e$. If we use these two facts for the final b-state in Eq. (2.5), we observe that the K value for the b-state is equal to that for the c-state, and is equivalent to a net increase in eV_0 by $n\hbar\omega/2e$ as observed above.

These aspects make the matrix element which one could use in the transfer Hamiltonian approach different from that conventionally used for photon interactions. However it will be seen that the integration



II-2 Energy level diagram for a metal-barrier-semiconductor junction utilized to demonstrate relationship between stimulated transitions and the current-voltage characteristic.

over the densities of states to deduce current-voltage charateristics of various devices for practical consideration will become simpler. It is also this aspect which allows one to relate absorption to the current-voltage characteristic.

II-3 Model of Absorption and Emission in Junctions to First Or in Perturbation

We now establish directly that to first order in perturbation the net rate of stimulated transitions into confined structure modes of tunneling junctions is proportional to a finite difference on the elastic d.c. tunneling current-voltage characteristic. In the low frequency (classical) limit, this reduces to the differential conductance. Since we have specifically done this for a metal-semiconductor junction, we shall considerthis case. The result and approach are however general. The approach demonstrates the relationship of absorption to the current-voltage characteristic (as discussed above) through the interaction within the barrier and thus provides insight into the relationship between the stationary state and transfer Hamiltonian approaches and of these to photon assisted tunneling.

Considering Fig. II-2, electrons in tunneling from left to right can emit a plasmon during the tunneling process and terminate in a lower energy state in the right electrode. Alternatively, the electron can absorb a plasmon from the field and enter an excited state in the right electrode.

To deduce the relationship between the absorption and emission processes and the elastic current we represent the plasmon interaction by a time varying potential $eV_r \cos(\omega t)$ across the barrier region, and neglect the interaction within the electrodes. A one dimensional structure is analyzed, and the plasmon propagation constant is assumed to be negligible. A rectangular barrier is taken, and the structure wave functions are approximated to lowest order in the tunneling exponential.

One can express the net plasmon emission rate to first order of perturbation theory as

$$\gamma = \frac{2\pi}{\hbar}\int_{-\infty}^{\infty} dE\ \rho_L(E)\{\rho_R(E-\hbar\omega)|M_{E,E-\hbar\omega}|^2(f(E-\mu) - f(E+eV_0-\hbar\omega-\mu)) - \rho_R(E+\hbar\omega)|M_{E,E+\hbar\omega}|^2(f(E-\mu) - f(E+eV_0+\hbar\omega-\mu))\} \tag{II.3.1}$$

where μ is the chemical potential of the left electrode, and for parabolic bands the state densities are given by

$$\rho_L(E) = \mathrm{Re}\frac{m}{k_i \hbar^2} \quad , \quad \rho_R(E') = \mathrm{Re}\frac{m}{k_f \hbar^2} \quad . \tag{II.3.2}$$

If the energy E is referenced to the conduction band edge of the left electrode, and we specialize to an M-I-S structure where the final state lies in the valence band, one has

$$k_i(E) = \left(\frac{2mE}{\hbar^2}\right)^{\frac{1}{2}} \quad , \quad k_f(E'+eV_0) = \left(\frac{2m}{\hbar^2}\,(\mu+\Delta-eV_0-E')\right)^{\frac{1}{2}} \quad , \tag{II.3.3}$$

where a single mass parameter has been adopted for simplicity. The parameter Δ represents the energy difference between the semiconductor Fermi level and valence band edge and is positive for the degenerate semiconductor specifically assumed for Fig. 2.

The stationary state model of the last section shows that the decay constants of the initial and final electronic wave functions in the barrier are equal and given by

$$K = \left(\frac{2m}{\hbar^2}\,(\phi-(E+E'+eV_0)/2)\right)^{\frac{1}{2}} \quad . \tag{II.3.4}$$

This relationship will be seen to be important in establishing the desired connection to the elastic current expression.

The matrix elements $M_{E'E}$, where

$$M_{E'E} = \int_0^d dx\ \psi^*_{L,E'}\left(-\frac{e}{m}A\cdot P - \frac{e}{2m}\,(p\cdot A)\right)\psi_{R,E} \quad , \tag{II.3.5}$$

can be expressed for a guage choice of $A = E/i\omega$ as

$$M_{E',E} = \frac{eV_r}{2\omega}\,T_L^* T_R\,\frac{\hbar K}{m}\,e^{-Kd} \tag{II.3.6}$$

where

$$T_L = \frac{-2ik_i}{K-ik_i} \quad , \quad T_R = \frac{-2ik_f}{K-ik_f} \tag{II.3.7}$$

The barrier height ϕ is measured from the left electrode (metal) conduction band edge.

Equation (3.1) is directly related to the d.c. current in the absence of inelastic tunneling, which to lowest order in the tunneling exponential is

$$I_{el}(V_0) = \frac{2\pi e}{\hbar} \int_{-\infty}^{\infty} dE \rho_L(E)\rho_R(E)|M'_{E,E}|^2(f(E-\mu) - f(E+eV_0-\mu) \tag{II.3.8}$$

where

$$M'_{E',E} = \frac{i\hbar}{2m}\int_{-\infty}^{\infty} dx\ \psi^*_{L,E'}(\vec{\nabla}\delta(x-\frac{d}{2}) - \delta(x-\frac{d}{2})\vec{\nabla})\psi_{R,E} \quad . \tag{II.3.9}$$

Upon examination of equations (1) to (7), it is apparent that the energy shift of the final state is equivalent to a change in bias potential, $-\hbar\omega/e$ for the first (emissive) term and $+\hbar\omega/e$ for the second (absorptive) term. Comparison of (1) with equation (8) for the d.c. elastic current yields the relationship

$$\gamma = \left(\frac{eV_r}{2\hbar\omega}\right)^2 \left(I_{el}(V_0-\hbar\omega/e) - I_{el}(V_0+\hbar\omega/e)\right)\frac{1}{e} \tag{II.3.9}$$

which in the classical ($\omega \to 0$) limit is

$$\gamma \cong -\left(\frac{\hbar\omega}{2}\right)\left(\frac{Vr}{\hbar\omega}\right)^2 \frac{\partial}{\partial V_0} I_{el}(V_0) \tag{II.3.10}$$

This is precisely the relation one obtains from the theory of photon assisted tunneling [7,15] in the limit $eV_r << \hbar\omega$. The present derivation depends specifically upon equation (3.4) from the stationary state theory and is obtained by considering the interaction within the barrier. It is a specific example of the general validity of the Tien-Gordon equation when barrier height modulation is included in the theory as in the stationary state approach [7], as opposed to the simple phase modulation imposed on the left electrode wave functions in the transfer Hamiltonian model[15].

Equation (3.4) has a simple physical interpretation which can be obtained by considering in particular a transition from a left state to a right state $\hbar\omega$ greater in energy. Viewed as a stationary process, the electron receives $\hbar\omega$ additional kinetic energy in making the transition

to the right electrode. Thus the barrier height it experiences emerging from the right side of the structure is reduced by $\hbar\omega$ and the average barrier height is reduced by $\hbar\omega/2$. Therefore to be correct, modified wavenumbers in the barrier region corresponding to an effective barrier height $\phi-\hbar\omega/2$ for the left states participating in absorption processes and an effective barrier height $\phi+\hbar\omega/2$ for the corresponding right states should be used. For emissive processes the signs are reversed. These corrections are generally small but are important in establishing the equality implied in equation (9) and its relation to photon assisted tunneling.

II-4 Relationship of the Stationary State Model to the Transfer Hamiltonian Model

In the present section of this chapter, it will be shown that the expression for d.c. elastic tunneling current itself can be written in the form of a transition rate, using a relation between the structure eigenstates which will be derived (17). This correspondence relates the stationary state model to models based on perturbation theory, in particular the transfer Hamiltonian model.

It is assumed that the tunneling particles are independent and have free particle dispersion characteristics throughout the structure. The extension to the effective mass approximation is straightforward, and does not change the nature of the result, so for simplicity a single mass parameter is assumed here. Furthermore a one dimensional structure is assumed, as the extension to three dimensions is also straightforward and again does not alter fundamentally the result of this section. Given that a static bias potential V_0 appears across the structure, an orthogonal set of time independent structure eigenstates are assumed to be known. As in section one, the $|c_\varepsilon(V_0)\rangle$ are eigenstates associated with the left electrode and carry flux to the right, and deep in the right electrode the $|c_\varepsilon(V_0)\rangle$ assume a plane wave form. Similarly the $|b_\varepsilon(V_0)\rangle$ are eigenstates associated with the right electrode and carry flux to the left. The parameter ε is the particle incident kinetic energy. It is assumed that the barrier potential is of finite extent, and outside of this region the states have plane wave behavior with incident, reflected

and transmitted components. The electrodes have equal thickness which will be taken to approach infinity, hence the actual behavior of the states in the barrier region is unimportant with respect to orthogonality and normalization. The normalization used here will be that used previously,

$$\langle c_{\varepsilon'}(V_0)|c_{\varepsilon}(V_0)\rangle = \delta_{k_1',k_1} \quad ,$$

$$\langle b_{\varepsilon'}(V_0)|b_{\varepsilon}(V_0)\rangle = \delta_{k_3',k_3}$$

where one may associate, e.g., $\frac{\pi}{L}\delta(k_1-k_1') \leftrightarrow \delta_{k_1,k_1'}$ as we assume $L \to \infty$. Outside of the barrier region the wavefunctions have the form

	$x \ll 0$	$x \gg 0$
$\sqrt{2L}\lvert c(V_0)\rangle$	$e^{ik_1x}+Re^{-ik_1x}$	Te^{ik_3x}
$\sqrt{2L}\lvert b(V_0)\rangle$	Te^{-ik_1x}	$e^{-ik_3x}+R'e^{ik_3x}$

(II.4.1)

where again the incident kinetic energy subscripts have been suppressed. The states are associated with the standing wave state energy density of the respective electrode, which are, neglecting spin degeneracy,

$$|c_{\varepsilon}(V_0)\rangle \leftrightarrow \rho_L = \frac{L}{\pi\hbar^2}\frac{m}{k_1} \tag{II.4.2}$$

$$|b_{\varepsilon}(V_0)\rangle \leftrightarrow \rho_R = \frac{L}{\pi\hbar^2}\frac{m}{k_3} \tag{II.4.3}$$

For any two degenerate eigenstates of the structure Hamiltonian, it is easily verified that the current operator matrix element between the two wavefunctions is constant throughout the configuration space of the Hamiltonian. Therefore one has $\langle c(V_0)|j|c(V_0)\rangle$, $\langle b(V_0)|j|b(V_0)\rangle$ and $\langle c(V_0)|j|b(V_0)\rangle$ constant throughout the structure for states corresponding to the same structure energy measured, e.g., from the right electrode conduction band edge (the energy subscripts have been suppressed). One may also consider the states $|c^*(V_0)\rangle$, $|b^*(V_0)\rangle$ which are the complex conjugates of the (time independent) $|c(V_0)\rangle$, $|b(V_0)\rangle$. The complex conjugates are orthogonal to each other, but not to $|c(V_0)\rangle$, $|b(V_0)\rangle$ with which they

are linearly dependent. With these states one can prove the following relationships between the exact structure eigenstates:

$$\langle c(V_0)|j|c(V_0)\rangle = -\langle c^*(V_0)|j|c^*(V_0)\rangle \ , \tag{II.4.4}$$

$$\langle c(V_0)|j|c(V_0)\rangle = \frac{2\pi\hbar}{e} |\langle c^*(V_0)|j|b(V_0)\rangle|^2 \rho_R \ , \tag{II.4.5}$$

$$\langle b(V_0)|j|b(V_0)\rangle = -\langle b^*(V_0)|j|b^*(V_0)\rangle \ , \tag{II.4.6}$$

$$\langle b(V_0)|j|b(V_0)\rangle = -\frac{2\pi\hbar}{e}|\langle c^*(V_0)|j|b(V_0)\rangle|^2 \rho_L \ , \tag{II.4.7}$$

where j is the current operator and again all states correspond to the same structure energy. The proof may be obtained through straightforward evaluation of the matrix elements. As all of the matrix elements are independent of position, one may evaluate $\langle c(V_0)|j|c(V_0)\rangle$, $\langle c^*(V_0)|j|c^*(V_0)\rangle$ and $\langle c^*(V_0)|j|b(V_0)\rangle$ deep in the right electrode, yielding

$$\langle c(V_0)|j|c(V_0)\rangle = \frac{1}{2L}\left(\frac{e\hbar k_3}{m}\right)|T|^2, \tag{II.4.8}$$

$$\langle c^*(V_0)|j|c^*(V_0)\rangle = -\frac{1}{2L}\left(\frac{e\hbar k_3}{m}\right)|T|^2 \tag{II.4.9}$$

$$\langle c^*(V_0)\ j\ b(V_0)\rangle = -\frac{1}{2L}\left(\frac{e\hbar k_3}{m}\right)T. \tag{II.4.10}$$

(4,5) then follow immediately. Similarly, evaluating $\langle b(V_0)|j|b(V_0)\rangle$, $\langle b^*(V_0)|j|b^*(V_0)\rangle$ and $\langle c^*(V_0)|j|b(V_0)\rangle$ deep in the left electrode leads directly to (6,7) above. The expressions that are most interesting for the purposes of this section are (5,7), as they relate the current of an eigenstate to an expression in the form of a transition rate, with j/e being the "interaction Hamiltonian."

The connection between the first order transfer Hamiltonian and stationary state models is established mathematically through (5,7). In the transfer Hamiltonian formalism, one generally adopts tunneling matrix elements as derived by Bardeen [6] so that the model reproduces the result of the stationary state approach to lowest order in the tunneling exponential. The prescription applied is to first find eigenstates

of the noninteracting L and R subsystems which are found by taking the barrier to have semi-infinite width, i.e. the L subspace consists of the left electrode and barrier for $x \geq 0$, and the right electrode and barrier for $x \leq b$ constitute the R space. The eigenstates, therefore, are standing waves. The expression for d.c. tunneling current is an occupation weighted sum of the transition rates between L and R states, where $\hbar j/e$, evaluated in the barrier region, is taken as the interaction Hamiltonian.

Although this procedure does reproduce the d.c., current expression of the stationary state model to lowest order in the tunneling exponential, the accuracy of higher order terms in the perturbation expansion is questionable due to the first order eigenstates employed. The eigenstates of the L and R subspaces of the conventional transfer Hamiltonian approach are approximations to lowest order in the tunneling exponential of the $|c^*(V_0)\rangle$, $|b(V_0)\rangle$ states (or equivalently, $|c(V_0)\rangle$ and $|b^*(V_0)\rangle$) above. Thus the fact that this procedure approximates the stationary state result is a consequence of Eq. (5,7). An exact correspondence is obtained if one adopts $|c^*(V_0)\rangle$ states for L and $|b^*(V_0)\rangle$ states for R in the first order transfer Hamiltonian model. This gives:

$$\langle J \rangle = \frac{2\pi e}{\hbar} \int_{-\infty}^{\infty} d\varepsilon \, \rho_L(\varepsilon)\rho_R(\varepsilon+eV_0+\delta)(f(\varepsilon-\mu_L) - f(\varepsilon+eV_0-\mu_L)$$

$$|\langle \; c_\varepsilon^*(V_0)|\frac{\hbar j}{e}|\, b_{\varepsilon+eV_0+\delta}(V_0)\rangle\,|^2 \qquad \text{(II.4.11)}$$

which using (5) is equal to

$$\langle J \rangle = \int_{-\infty}^{\infty} d\varepsilon \, \rho_L(\varepsilon)(f(\varepsilon-\mu_L) - f(\varepsilon+eV_0-\mu_L)) \;\; \langle c_\varepsilon(V_0)|j|c_\varepsilon(V_0)\rangle \qquad \text{(II.4.12)}$$

where δ is the difference in electrode chemical potentials $\mu_L - \mu_R$. This is the exact expression for the current from the stationary state model [7]. Replacement of Bardeen's approximate L and R wavefunctions with the exact $|c^*(V_0)\rangle$, $|b(V_0)\rangle$ thus makes the two models equivalent with respect to the computation of the d.c. current.

One may view the first order independent particle Hamiltonian model as a formal approximation of the stationary state model where the

use of $\hbar j/e$ as the interaction Hamiltonian is justified by Eq. (5,7). With the association of the first order and full structure eigenstates and utilizing (5,7) one can regain the stationary state d.c. current expression from that of the first order transfer Hamiltonian model. This fact may be useful in extending the first order model to, e.g., extremely thin barriers.

In this section we have described how the stationary state and first order transfer Hamiltonian tunneling models are related, in the case of static bias and independent particles. The relationship is a consequence of the properties of the structure eigenstates, in particular that the flux of an eigenstate may be expressed in the form of a transition rate. Using the relationship, one may construct an exact transfer Hamiltonian model for static bias and neglecting collisions simply by using the appropriate structure wavefunctions instead of the commonly used approximate L and R wavefunctions.

II-5 Theoretical Considerations for Practical Converters

The basic relationships developed in the previous sections which are most important as far as mode conversion in tunneling junctions is concerned are first of all that providing the d.c. current, inelastic (induced) and elastic (back-flow). Secondly one has the expression providing the transition rate due to the presence of the mode within the junction itself. These two are:

$$I = \sum_n J_n^2 \left(\frac{eV_r}{\hbar\omega}\right) I(V_0 + n\hbar\omega/e) \quad , \tag{II.5.1}$$

$$\gamma = \sum_n n J_n^2 \left(\frac{eV_r}{\hbar\omega}\right) I(V_0 + n\hbar\omega/e)/e \quad , \tag{II.5.2}$$

One observes that to maximize energy absorption from the field the current-voltage relationship should satisfy $I(V_0+n\hbar\omega/e) >> I(V_0-n\hbar\omega/e)$ for $n > 0$. This is similar to the condition one would expect for classical rectification, although due to the quantum nature of the radiation the change in slope need not be abrupt. Only points on the current-voltage characteristic corresponding to integral numbers of photon potentials contribute, which at optical frequencies are separated by the order of one volt. For

the current-voltage relationship exhibited by an Al-Al_2O_3-Al field emission junction, the rectification efficiency increases with increasing frequency. At visible frequencies one would expect high rectification efficiency as the current change can be several orders of magnitude over a photon potential.

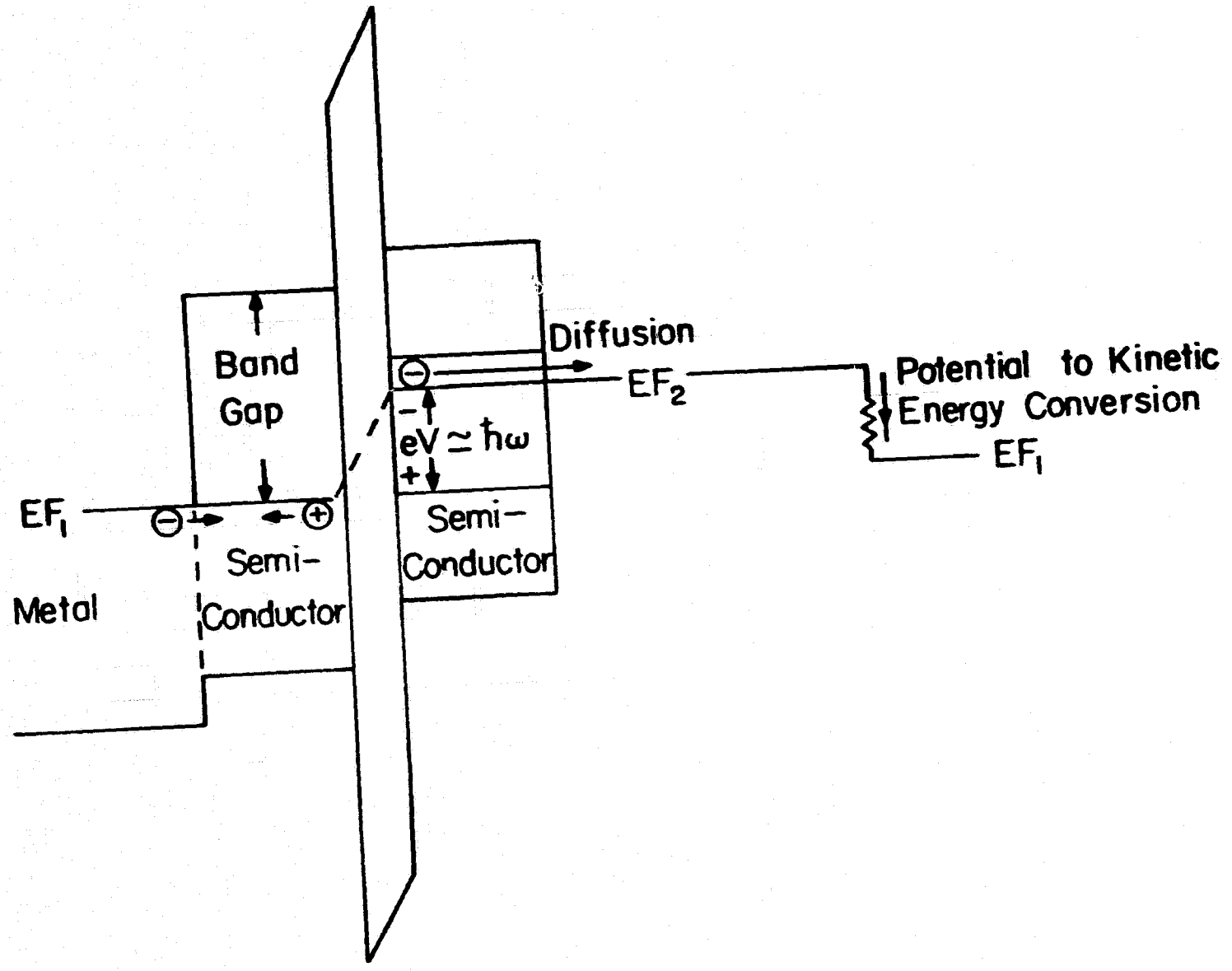
In order for energy to be extracted from the field, the hot electrons produced by photon absorption must be prevented from decaying thermally within the electrode. Electrode band structure can be used to forbid decay through thermalization, for example if the final tunneling state is a narrow impurity band in a semiconductor band gap, the energy loss will be less than the impurity band width. Such a configuration is shown in Fig. II-3 (photon "pump") and is to be considered further.

In order to extract net energy from the field it is necessary to compare the absorption rate given by the above formulae with the losses inherent in the system. The sources of loss can be divided into two types: those caused by the inefficiency inherent within the process itself, and secondly losses associated with other processes accompanying the direct conversion process. An example of the former is the harmonic conversion in the rectifier converter, since it arises as a result of the conversion process itself. The efficiency of such processes are given by the ratio of the useful transition rate over the total transition rate. For the rectifier this is the ratio of the transition rate providing the d.c. component to the total transition rate. This aspect is dealt with in more detail in the recently submitted proposal.

An extremely important example of the second type of loss and one which must be carefully considered for the tunneling converter is the propagational loss associated with the particular electromagnetic mode which is inducing the transitions. This particular loss is given by the imaginary part of the propagation constant associated with the propagation of the mode. This loss is discussed more fully in Chapter III pertaining to the structure modes of interest.

In order to compare the propagational loss with the useful transition rate one must introduce the loss coefficient for the conversion process. This is given by

$$\alpha = \hbar\omega R/P \qquad (II.5.3)$$



II-3 Energy Level diagram for a junction structure that would prevent back-flow and hot electron formation. Simpler structures might prove possible. This semiconductor metal structure is discussed in more detail in the present proposal.

where R is the absorption rate per unit length of the device and P is the power propagating in the mode inducing the transitions. The former is given by γ/L_z, where L_z is the interaction length of the converter. P the power is given in terms of the Poynting vector

$$P = i_z \cdot \int \vec{E} \times \vec{H}/2 \, d\vec{s} \quad .$$

To lowest order in perturbation γ is given by Eq. (II.3.1) which for the three dimensional case can be written as

$$\gamma = \frac{2\pi}{\hbar} \int d^3k_i d^3k_f |M_{i,f}|^2 (f(E_i-\mu) - f(E_f+eV_0-\mu) \times$$

$$(\delta(E_i-E_f+\hbar\omega) - \delta(E_i-E_f-\hbar\omega)) \qquad \text{(II.5.4)}$$

$$= \frac{1}{e}\left(\frac{eV_r}{2\hbar\omega}\right)^2 \left(I\left(V_0+\frac{\hbar\omega}{e}\right) - I\left(V_0-\frac{\hbar\omega}{e}\right)\right) \qquad \text{(II.5.5)}$$

the latter by Eq. (II.3.9)

One observes from Eq. (5.3) that for appreciable absorption to occur, the ratio of R/P must be as large as possible. Since R in the region of interest is proportional to the power propagating within the barrier, this implies that the greater the ratio of power propagating within the tunneling barrier to the total power propagating within the mode, the greater the value of α, and hence the greater the overall conversion efficiency possible. In a region of normal dispersion $P = v_g W$ where W is the energy stored per unit length $\frac{1}{2}V_r^2 \epsilon_0 A/(L_z d)$ where A is the junction area. Thus letting $v_g = S/c$, one obtains

$$\alpha = \left(\frac{d}{\epsilon_0 A}\right) \times \frac{S}{c} \times \frac{(I(V_0+\hbar\omega/e) - I(V_0-\hbar\omega/e))}{(2\hbar\omega/e)} \qquad \text{(II.5.6)}$$

If for parameter values one assumes d = 10Å, S = 20 , $I(V_0+\hbar\omega/e)/A = 5\times10^4$ Amps/cm^2, $I(V_0-\hbar\omega/e) = 0$, $\hbar\omega/e$ = 1 Volt, one obtains $\alpha \simeq 36$ cm^{-1}.

There are two important aspects to this result. Firstly, this α must be able to compete and overwhelm the mode absorption for efficient conversion. At low temperatures this should be no problem, although low temperature loss studies should be performed. At room temperature, the

mode loss is too great for this to occur.

Looking ahead, Fig. (III-8) indicates a mode loss of approximately 10^4 to 10^5 cm^{-1} at room temperature so that significantly reduced loss must be obtained. For low temperatures ($\simeq$10°K) three of four orders of magnitude reduction is easily obtained so that α can readily dominate. The efficiency is then limited by coupling and harmonic conversion.

Because of the second important aspect even if α is large, one can still be left with a low efficiency device. This is clear from Eq. (II.5.1). Since one observes that if $I(V_0+\hbar\omega/e) = -I(V_0-\hbar\omega/e)$, to first order I is equal to zero and no conversion to d.c. occurs. This is the case for a zero biased symmetric junction device.

One should thus define the inherent efficiency as the ratio, $I_{dc}^2 R/\hbar\omega$, of the power delivered to load to the total incident power absorbed. An additional mode loss factor of $\frac{\alpha}{\alpha+2\beta''}$ where β'' is the imaginary part of the mode propagation coefficient should be included.* Thus the overall efficiency for the device, excluding coupling is given by

$$\eta = \frac{\left(\left(\sum J_n^2\left(\frac{eV_r}{\hbar\omega}\right) I(V_0+n\hbar\omega/e)\right)^2 - I^2(V_0)\right)R}{\left(\sum_n n J_n^2\left(\frac{eV_r}{\hbar\omega}\right) I(V_0+n\hbar\omega/e)\right)} \frac{\alpha}{\alpha+2\beta''} \cdot \frac{e}{\hbar\omega} \qquad \text{(II.5.7)}$$

At room temperature $\alpha/(\alpha+2\beta'') \simeq 3.6\times10^{-3}$ indicating a peak efficiency of .36%. In this expression the $- I^2(V_0)$ term accounts for the power required of the bias circuit.

It is interesting that η, theoretically at least, can have a value of 1. This occurs for a situation for which $\alpha \ll \beta''$ and $I(V_0+n\hbar\omega/e) = 0$ for $n \neq +1$ and coupling to the mode is assumed perfect. This in one compact statement expresses the essential features for a 100% converter. That this is possible is interesting, since, according to this, low frequency rectification can be surpassed by preventing multiphoton transitions from occuring. Furthermore it is interesting to note that the above cannot happen for a depletion layer device since phonons inhibit the coherence upon which this result depends.

*While in print it has been observed that recent results obtained by R. E. Drullinger et al, Applied Physics Letter 42, 137 (1983) reports an $\omega^{-\frac{3}{4}}$ dependence of the signal to noise ratio of a point contact diode from 1 GHz to 10 THz. It will be shown that $\beta''\alpha\omega^{\frac{3}{4}}$ and $\beta \gg \alpha$, so that these results tend to verify Eq. (II.5.7).

The low frequency small signal limit of Eq. (II.5.7) is obtained by taking only the n = ±1,0 terms, taking only the first terms in the series expansions for J_0 and J_1 and approximating differences by derivatives. Current terms independent of V_r are subtracted since these are bias terms. Assuming $\alpha >> \beta''$ gives

$$\eta \cong \frac{I''^2-(4I/V_r^2)^2}{8I'} RV_r^2$$

Choosing $I''/I' \cong 7(V^{-1})$, $R^{-1} \cong I' \cong (3000\ \Omega)^{-1}V_r \cong .2V$ and $I \sim 20\mu A$, one obtains $\eta \cong .2^{(11,18)}$ even including the bias power required. These results are highly questionnable however, since first of all for V_r .2V one needs the large signal analysis which will decrease this projected efficiency. Secondly, the results are highly dependent on V_r. Reducing this to .1V causes the conversion to be easily overwhelmed by the bias power required. Finally for visible frequencies, for a correct analysis, the quantum limit is required since $\hbar\omega/e \sim 1V$. We are presently discussing the analysis of this using a more accurate approach for a junction configuration for which backflow has been diminished. Unless this is done it is believed that high conversion is not likely.

In conclusion, we have carried out a stationary state analysis of tunneling and have been able to demonstrate the detailed aspects of the interaction of the tunneling with a confined junction mode. The final equations for the load current, the inherent transition rate from the mode and the expression for the efficiency are simple and provide the essential features required to obtain high conversion efficiencies. We are continuing our studies of these relationships to first of all investigate the fundamental limits of present converters such as p-n junction depletion layer devices, and secondly to investigate junctions which will provide the highly selective constraints imposed by Eq. (II.5.7) = 1.

References for Chapter II

[1] J.G. Simmons, "Generalized formula for the electric tunnel effect between similar electrodes separated by a thin insulating film," J. Appl. Phys., Vol. 34, pp. 1793-1803, June 1963.
J.G. Simmons, "Electric tunnel effect between dissimilar electrodes separated by a thin insulating film," J. Appl. Phys., Vol. 34, pp. 2581-2590, September 1963.

[2] T. Stratton, "Volt-current characteristics for tunneling through insulating films," J. Phys. Chem. Solids, Vol. 23, pp. 1177-1190, 1962.

[3] J.G. Simmons, "Potential barrier and emission-limited current flow between closely spaced parallel metal electrodes," J. Appl. Phys., Vol 35, pp. 2472-2481, August 1964.

[4] C.B. Duke, G.G. Kleinman and T.E. Slaketon, "Microscopic theory of tunneling: general theory and application to the static impurity," Phys. Rev., Vol. B6, pp. 2389-2409, September 1972.

[5] A.M. Andrews, H.W. Korb, N. Holonyak Jr., C.B. Duke and G.G. Kleinman, "Tunnel mechanisms and junction characteristization in III-IV tunnel diodes," Phys. Rev. B, Vol. 5, pp. 2273-2295, March 1972.

[6] J. Bardeen, "Tunneling from a many particle point of view," Phys. Rev. Lett., Vol. 6, pp. 57-59, January 1961.

[7] S.R. Whiteley and T.K. Gustafson, "Stationary state model for normal metal tunnel junction phenomena," Journal of Quantum Elect., QE-18, 1387 (1982).

[8] See, e.g. M. Merzbacher, Quantum Mechanics, John Wiley, New York, 1970.

[9] L.C. Davis, "Theory of surface plasmon excitation in metal-insulator-metal tunnel junctions," Phys. Rev., Vol. B16, pp. 2482-2490, September 1977.

[10] W.A. Harrison, "Tunneling from an independent particle point of view," Phys. Rev., Vol. 123, pp. 85-89, July 1961.

[11] C.W. Slayman, Metal-Barrier-Metal Junctions for Millimeter Wave Mixing and Detection, Ph.D. Thesis, U.C. Berkeley, 1979 (unpublished).

[12] D. Rogovin and D.J. Scalapino, "Fluctuation phenomena in tunnel junctions," Ann. Phys., Vol. 86, pp. 1-90, July 1974.

[13] A.H. Dayem and R.J. Martin, "Quantum interaction of microwave radiation with tunneling between superconductors," Phys. Rev. Lett., Vol. 8, pp. 246-248, March 1962.
P.K. Tien and J.P. Gordon, "Multiphoton processes observed in the interaction of microwave fields with the tunneling between superconductor films," Phys. Rev., Vol. 129, pp. 647-651, January 1963.

[14] N.R. Werthamer, "Nonlinear self coupling of Josephson radiation in superconducting tunnel junctions," Phys. Rev., Vol. 147, pp. 255-263, July 1966.
L.E. Hasselberg, "Theory of photon assisted tunneling in superconconductors," J. Phys. F: Metal Physics, Vol. 3, pp. 1438-1462, July 1973.

[15] J.R. Tucker, "Quantum limited detection in tunnel junction mixers," IEEE J. Quantum Electronics, Vol. QE-15, pp. 1234-1258, November 1979.

[16] L. Esaki and P.J. Stiles, Phys. Rev. Lett. (16), p. 1108 (1966).

[17] S.R. Whiteley and T.K. Gustafson, "Relationship between the stationary state and first order transfer Hamiltonian tunneling models," to be published.

[18] A. Sanchez, C. F. Davies, Jr., K. C. Liu, and A. Javan, J. Appl. Phys. 49, 5271 (1978). By a similar classical analysis the authors estimate an efficiency of 1%.

III PLASMON MODES IN METAL-BARRIER-METAL STRUCTURES

For the various tunneling junctions and coupling devices which have been proposed to directly convert a free space field to an electrical current through the junction, interfaces between dielectrically different media will be important to constrain the spatial extent of the radiation field. Such self-sustained electromagnetic interface modes can exist provided certain dispersion relationships are satisfied.[1-4] Since the dielectric properties of metals and heavily doped semiconductors at optical or infrared frequencies and below are strongly influenced by the collective oscillations of free electrons, these excitations are referred to as surface plasmons. They are T.M. in character and depend upon the presence of a negative dielectric coefficient for at least one of the bounding materials.

Some of these modes couple strongly with tunneling electrons due to their large dipole moment. For suitable conditions they also couple efficiently to an external electromagnetic field. This coupling can be accomplished by phase matching using surface roughness, prism couplers, gratings, or by confining the excitations to finite structures. The various possible coupling schemes are discussed in more detail in Chapter IV. In the present chapter, we review some of the basic idealized plasmon modes and then consider aspects of particular importance to potential energy conversion devices. In particular, we review some recent numerical results which have been obtained.

III-1 The Shape of the Resonance

An important factor in considering the interaction between electrons and surface plasmons is the line-shape of the resonance. This is given by the rate of change in energy, as a function of frequency. Consider the introduction of a small amount of charge ρ_f into a surface plasmon system. This can be experimentally accomplished by transmitting electrons through a thin foil. The total charge ρ_T which consists of this in addition to the polarization charges induced is still zero. For frequencies where the electrostatic treatment is valid, the normal component of the electric field is equal and opposite on opposite sides of the surface.

The latter are related to the charges by:

$$E_0 - E_1 = 0 \tag{III.1.1a}$$

$$\varepsilon_0 E_0 + \varepsilon_1 E_1 = \rho_f$$

$$\therefore \quad E_0 = \rho_f/(\varepsilon_0+\varepsilon_1) \tag{III.1.1b}$$

where E_0 and E_1 are the normal electric fields. The rate of change in energy of surface plasmons is given by:

$$\begin{aligned}\frac{dU}{dt} &= \mathrm{Re}\int dV\ E\,\frac{\partial D^*}{\partial t} \\ &\cong \int dV\ \mathrm{Re}[-i\omega\rho_f^2\ \ (\varepsilon_0+\varepsilon_1)^*] \\ &= \int dV\ \mathrm{Re}[-i\omega\rho_f^2/(\varepsilon_0+\varepsilon_1)] \\ &\propto \mathrm{Im}\left(\frac{\varepsilon_0}{\varepsilon_0+\varepsilon_1}\right)\end{aligned} \tag{III.1.2}$$

Hence the line-shape is given by $\mathrm{Im}\left(\frac{\varepsilon_0}{\varepsilon_0+\varepsilon_1}\right)$. For example, for the Drude model dielectric function $\varepsilon_0(1-\omega_p^2/\omega^2)$, the quantity $(\varepsilon_0+\varepsilon_1)$ has a simple zero at $\omega = \omega_p/\sqrt{2}$. The line shape is thus given by $\delta(\omega-\omega_p/\sqrt{2})$. A more realistic dielectric constant is that for a low-loss plasma. For $\omega >> \frac{1}{\xi}$, where ξ is the mean collision time, this can be written as:

$$\varepsilon_1 = \varepsilon_0\left[\left(1 - \frac{\omega_p^2}{\omega^2}\right) - i\,\frac{\omega_p^2}{\omega^3\xi}\right] \tag{III.1.3}$$

For frequencies near $\omega_p/\sqrt{2}$, where $\mathrm{Im}\left(\frac{\varepsilon_0}{\varepsilon_0+\varepsilon_1}\right)$ is significant, the latter can be written as:

$$\mathrm{Im}\left(\frac{\varepsilon_0}{\varepsilon_0+\varepsilon_1}\right) \propto \frac{\frac{1}{2}\xi}{(\omega-\omega_p/\sqrt{2})^2+(\sqrt{2}/4\xi)^2} \tag{III.1.4}$$

that is, a Lorentzian line-shape with a peak at $\omega = \omega_p/\sqrt{2}$, and a full width at half maximum of $1/\xi$. In real metals, an approximately Lorentzian line-shape is obtained from $\mathrm{Im}\left(\frac{\varepsilon_0}{\varepsilon_0+\varepsilon_1}\right)$, at least on the low frequency side.

On the other side this is usually complicated by interband transitions. For a simple metal-air interface the resonant profile is that of the bulk resonance (given by $\mathrm{Im}(1/\varepsilon_1)$) shifted to a frequency $\omega_p/\sqrt{2}$. This bulk resonance is shown in Fig. (III-1) for illustrative purposes. Additional data is available in Ref. (5). Table 1 shows basic data for several of the important bulk resonances.

III-2 Idealized Lossless Plasmon Modes for Some Important Structures

When one discusses the lineshape as in the previous section, the dispersion is neglected; that is the dependence of the resonance frequency on the wavelength of the radiation. This thus gives the resonances for the portions of the dispersion curves that are independent of β.

To discuss the important aspects of such modes, the full dispersion is extremely important particularly since one generally is away from the resonance frequency. In this section we consider a variety of ideal modes which are basic to tunneling devices and potential coupling devices. To deduce these dispersion relationships we initially neglect the influence of loss. In the following section this (loss) is considered for two of the modes of importance. Recent numerical calculations being carried out to more fully investigate the modes of importance within the junction devices themselves are also considered.

The simplest interface is a metal-dielectric plane, as illustrated in Fig. (III-2). The dielectric constant of the metal is ε_1, while that of the adjacent medium is ε_0. Consider an electromagnetic wave traveling in the z-direction, parallel to the surface. Since a surface mode is desired here, the fields are assumed to be exponentially decaying away from the boundary, $x = 0$. With this assumed form, the tangential components of the fields are written as:

for $x > 0$:

$$E_z = A_0 \exp(i\omega t - ikz - \nu_0 x) \qquad \text{(III.2.1)}$$

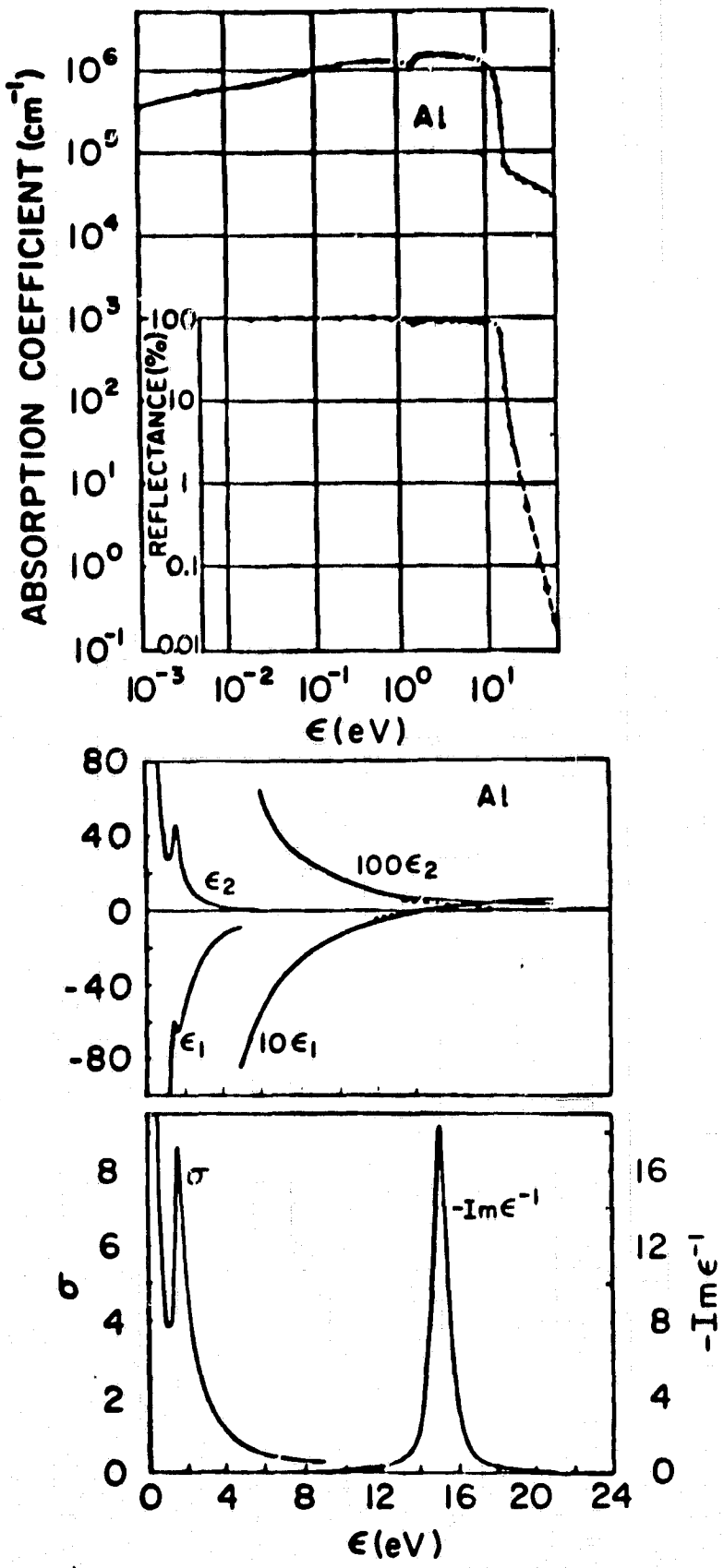


Fig. III-1: Optical Properties of Aluminum: (a) Reflectivity and absorption coefficient data over an extended range (Reprinted from Ref. 6)

(b) Frequency dependence of ϵ_1, ϵ_2 conductivity σ and the energy loss function - Im ϵ^{-1}. The effect of t 1.6 eV interband peak on ϵ_1, ϵ_2 and σ is obvious. (Reprinted from Ref.

$$H_y = -\left(\frac{i\omega\varepsilon_0}{\nu_0}\right) E_z$$

$$E_x = \frac{k}{\omega\varepsilon_0} H_y = +i\frac{k}{\nu_0} E_z$$

for $x < 0$:

$$E_z = A_0 \exp(i\omega t - ikz + \nu_1 x)$$

$$H_y = \left(\frac{i\omega\varepsilon_1}{\nu_1}\right) E_z$$

$$E_x = \frac{k}{\omega\varepsilon_1} H_y = \frac{ik}{\nu_1} E_z \qquad \text{(III.2.2)}$$

where A_0 is the field strength, k is the propagation constant, and $\nu^2 = k^2 - \omega^2\mu\varepsilon$ in the respective media. The dispersion relationship between ω and k, obtained by applying the boundary condition of continuity of these components, is given by:

$$\frac{\varepsilon_0}{\nu_0} + \frac{\varepsilon_1}{\nu_1} = 0 \qquad \text{(III.2.3)}$$

Since ν_0 and ν_1 are assumed to be positive quantities, it follows that ε_0 and ε_1 must be of opposite sign.

The dispersion can also be written as:

$$k^2 = \omega^2\mu\varepsilon_0 \left(\frac{\varepsilon_1}{\varepsilon_0+\varepsilon_1}\right) \qquad \text{(III.2.4)}$$

Of particular interest are those frequencies for which $\varepsilon_1 = -\varepsilon_0$, and thus $k \to \infty$. These are the resonance frequencies of the surface plasmons. For real metals, ε_1 is in general complex; k obtained as the solution to Eq. (III.2.3) or Eq. (III.2.4) is also complex, $k = \beta - i\alpha$. The decay length or range for such surface plasmons, $\ell = 1/\alpha$, which is usually very long below resonance [7,8,9] is discussed in the next section. The resonance frequencies are still given by $\varepsilon_1' = -\varepsilon_0$, where ε_1' is the real part of ε_1.

The dispersion obtained above can be expressed in an explicit form using the Drude model for the metal, where the dielectric constant is given by:

$$\varepsilon_1 = \varepsilon_0\left(1 - \frac{\omega_p^2}{\omega^2}\right) \qquad \text{(III.2.5)}$$

where ω_p is the bulk plasma frequency: $\omega_p^2 = \frac{Ne^2}{\varepsilon_0 m}$. Equation (III.2.3) and Eq. (III.2.4) can then be written as:

$$k^2 = \omega^2\mu\varepsilon_0\left(\frac{\omega_p^2-\omega^2}{\omega_p^2-2\omega^2}\right) \qquad \text{(III.2.6)}$$

or

$$\omega^2\mu\varepsilon_0 = \left(k^2 + \frac{\omega_p^2\mu\varepsilon_0}{2}\right) - \left[(k^2)^2 + \left(\frac{\omega_p^2\mu\varepsilon_0}{2}\right)\right]^{\frac{1}{2}} \qquad \text{(III.2.7)}$$

The resonance frequency is $\omega_p/\sqrt{2}$. Such dispersion is representative of ideal metals, and is illustrated in Fig. (III-3).

For the model above, ν_1 which indicates the decay into the metal and is approximately $\omega_p\sqrt{\mu\varepsilon_0}$ for $\omega \ll \omega_p$, which for typical metals is of the order of $(1000\ \mathring{A})^{-1}$. Such thicknesses are not uncommon for metal film electrodes in metal-barrier-metal junctions. The interaction between the fields of the two interfaces is then significant, and the description of the surface plamon mode as a single interface mode becomes inaccurate. Both surfaces must be considered in order to correctly describe the dispersion.

The penetration into the air above the metal is similarly given by $\nu_0 \cong \frac{\omega}{\omega_p}(\omega\sqrt{\mu\varepsilon_0})$ for $\omega < \omega_p$ and is thus typically 10λ for aluminum throughout the visible region. This however does not take relaxation into account.

A modification of the simple interface of Fig. III-2, the dielectric-metal-dielectric structure, is illustrated in Fig. III-4. The fields in the dielectric regions are again decaying exponentials, whereas those in the metal are a combination of growing and decaying ones. Due to the symmetry of the structure about the mid plane, the fields must either be symmetric or antisymmetric. The dispersion for these two branches is given respectively by:

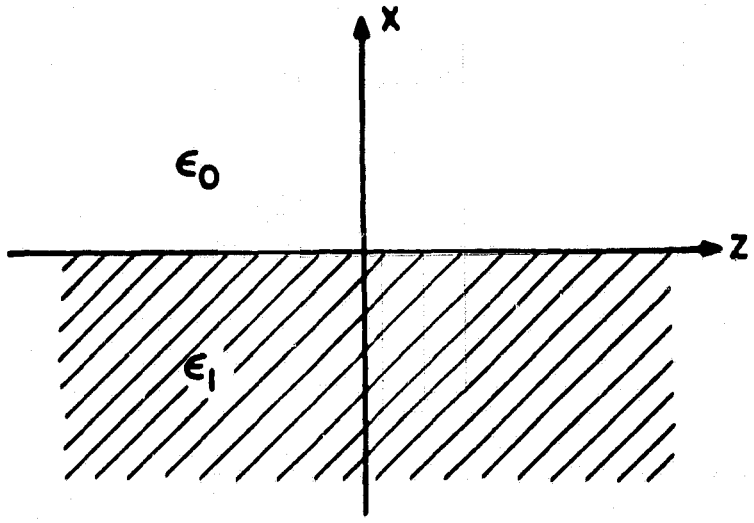


Fig. III-2. Schematic diagram of an infinite-extent planar interface between vacuum and a metal. The region above the y-z plane is occupied by free space of dielectric constant ε_0, and that below is occupied by a metal of dielectric constant ε_1. Surface plasmon excitations are assumed to be propagating along the z-direction.

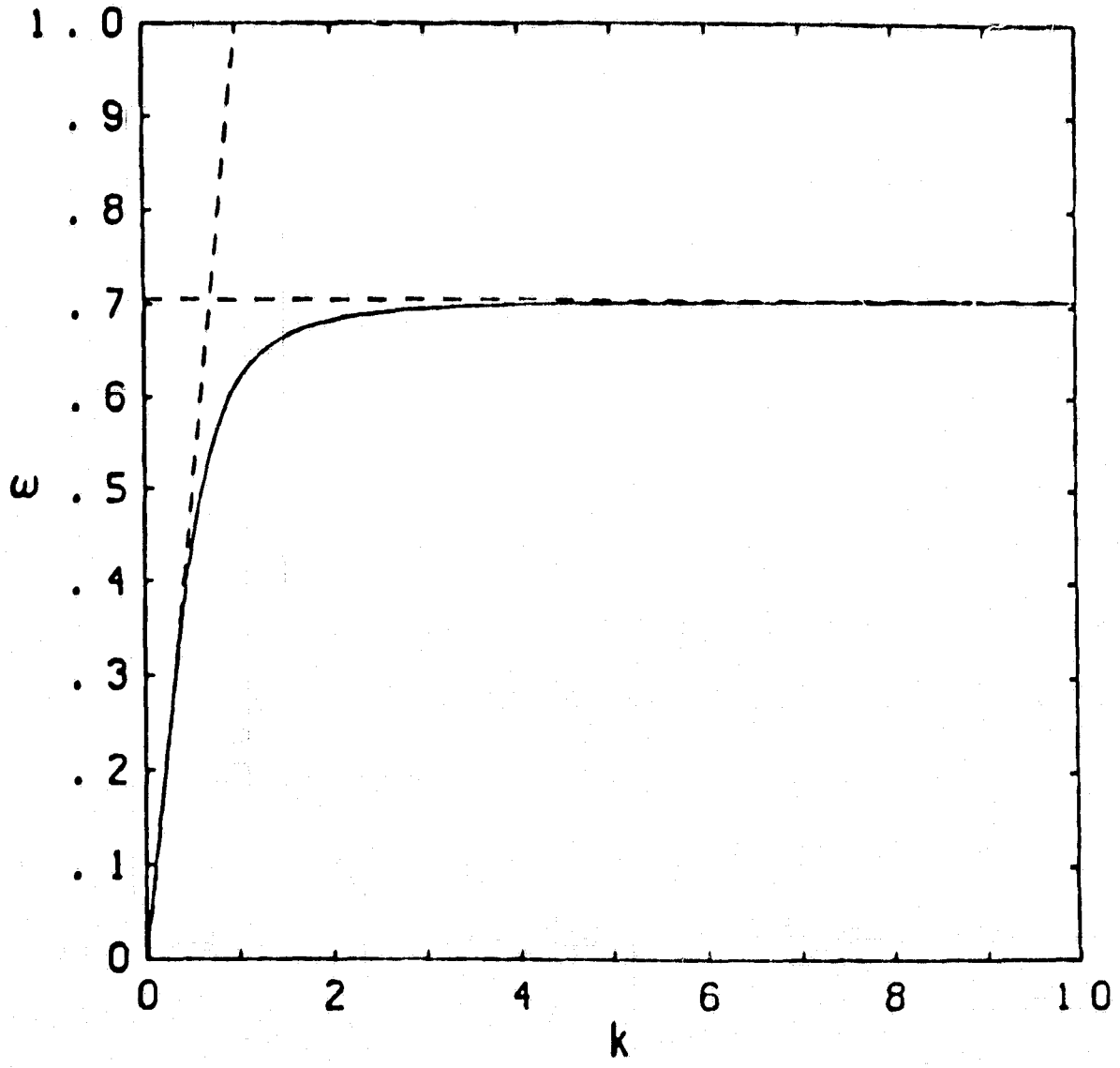


Fig. III-3. Dispersion of surface plasmons on a planar interface for an ideal metal. ω is normalized with respect to ω_p, and k with respect to ω_p/c. The dash lines represent the limits given by $\omega = kc$ and $\omega = \omega_p/\sqrt{2}$.

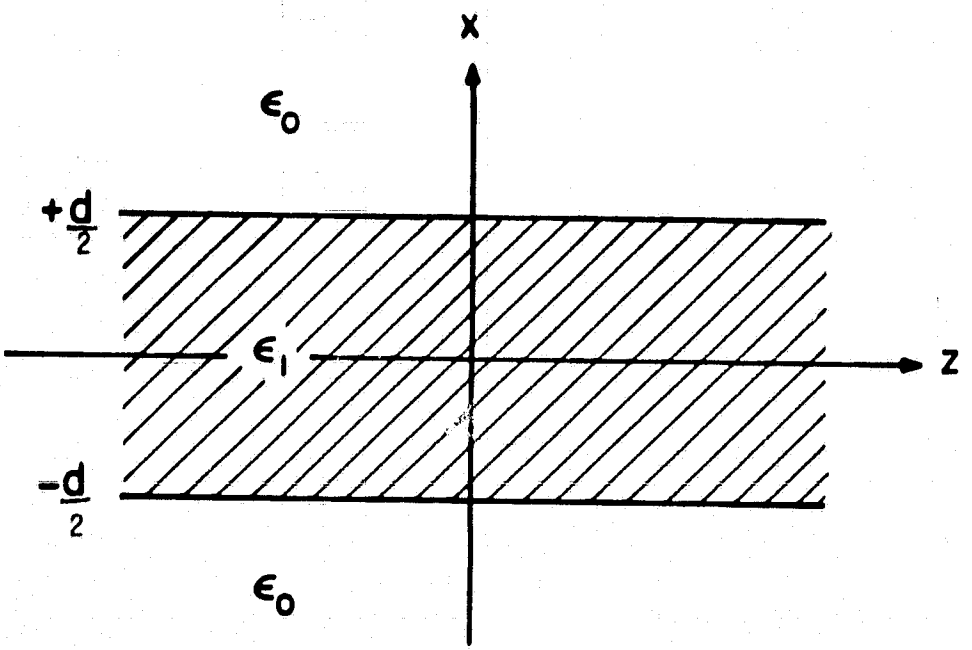


Fig. III-4. Schematic diagram of a vacuum-metal-vacuum planar structure. A metal film of thickness d occupies the space between $x = +\frac{d}{2}$ to $x = -\frac{d}{2}$. Surface plasmon excitations are assumed to be propagating in the z-direction.

$$\left(\frac{\varepsilon_0}{\nu_0}\right) + \tanh(\nu_1 d)\left(\frac{\varepsilon_1}{\nu_1}\right) = 0$$

$$\tanh(\nu_1 d)\left(\frac{\varepsilon_0}{\nu_0}\right) + \left(\frac{\varepsilon_1}{\nu_1}\right) = 0 \qquad \text{(III.2.8)}$$

For small values of d, the branches behave significantly different from the single surface mode. The dispersion for the symmetric branch becomes further relaxed, while that for the anti-symmetric branch becomes closer to the porpagation mode in free space. The resonant behaviour is however very similar in all cases, with the same resonance frequencies. For the simple Drude model, the dispersion of both branches are illustrated in Figs. (III-5) and (III-6).

A closely related but important structure for tunneling junctions is a metal-dielectric-metal gap occurring in the junction, as illustrated in Fig. (III-7). Here only the antisymmetric branch is of interest, since the potential difference between the surfaces along a path perpendicular to them is zero for the symmetric branch, and therefore has little effect on device operation. From Eq. (III.2.8), the dispersion can be written as:

$$\left(\frac{\varepsilon_0}{\nu_0}\right) + \tanh(\nu_0 d)\left(\frac{\varepsilon_1}{\nu_1}\right) = 0 \qquad \text{(III.2.9)}$$

The field configuration and the derivation of this is given in Appendix (II). For very small gaps the phase velocity is much smaller than the speed of light. The dispersion is approximately given by: $\nu_0^2 \simeq (\nu_1\varepsilon_0)/(-\varepsilon_1 d)$. Furthermore, $\nu_1, ((\beta^2-\omega^2\mu\varepsilon_1)^{\frac{1}{2}})$, can be approximated by β in some cases or by $\omega\sqrt{\mu_0}\sqrt{-\varepsilon_1}$ in others resulting in two approximate but useful expression for β. Assuming $\nu_0 \cong \beta$ since $\beta >> \omega\sqrt{\mu\varepsilon_0}$ one obtains respectively

$$\beta \cong \left(\frac{\varepsilon_0}{-\varepsilon_1}\right)/d \quad ; \quad \nu_1 \simeq \beta \quad ; \quad \nu_0 \simeq \beta \qquad \text{(III.2.10a)}$$

and

$$\beta^2 \cong \omega\mu_0^{\frac{1}{2}}\varepsilon_0/((-\varepsilon_1)^{\frac{1}{2}}d) \; ; \; \nu_1 \simeq \omega\sqrt{\mu_0(-\varepsilon_1)} \simeq \omega_p\sqrt{\mu\varepsilon_0} \; ; \; \nu_0 \simeq \beta \qquad \text{(III.2.10b)}$$

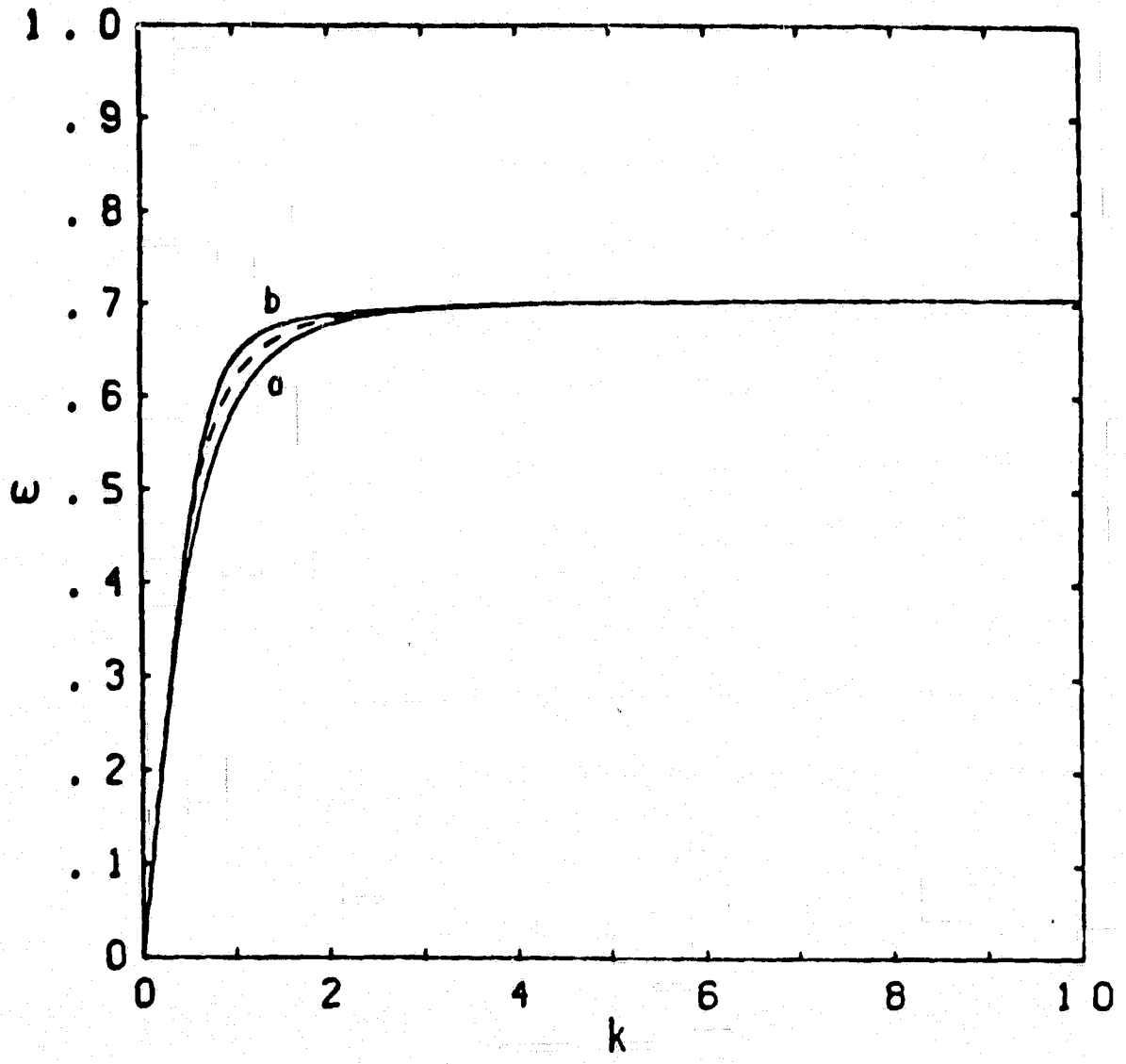


Fig. III-5. Dispersion of the symmetric mode (a) and the anti-symmetric mode (b) of surface plasmons on a planar ideal metal film. The thickness is given by $(\omega_p/c)d = 2$. ω is normalized with respect to ω_p, and k with respect to (ω/c). The dash line represents that of excitations on a single planar surface.

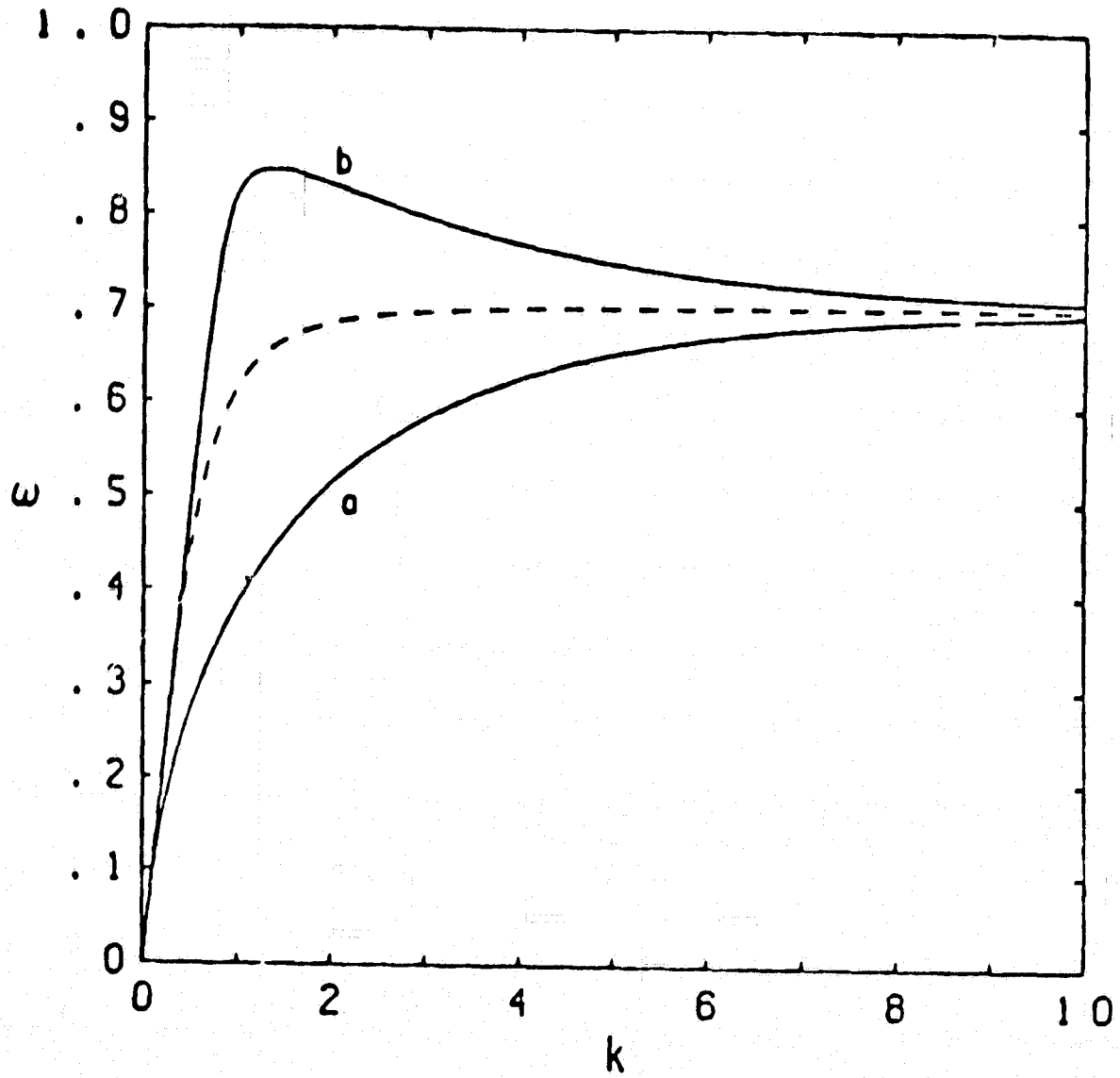


Fig. III-6. Dispersion of the modes for a thickness of $(\omega_p/c)d = 0.4$.

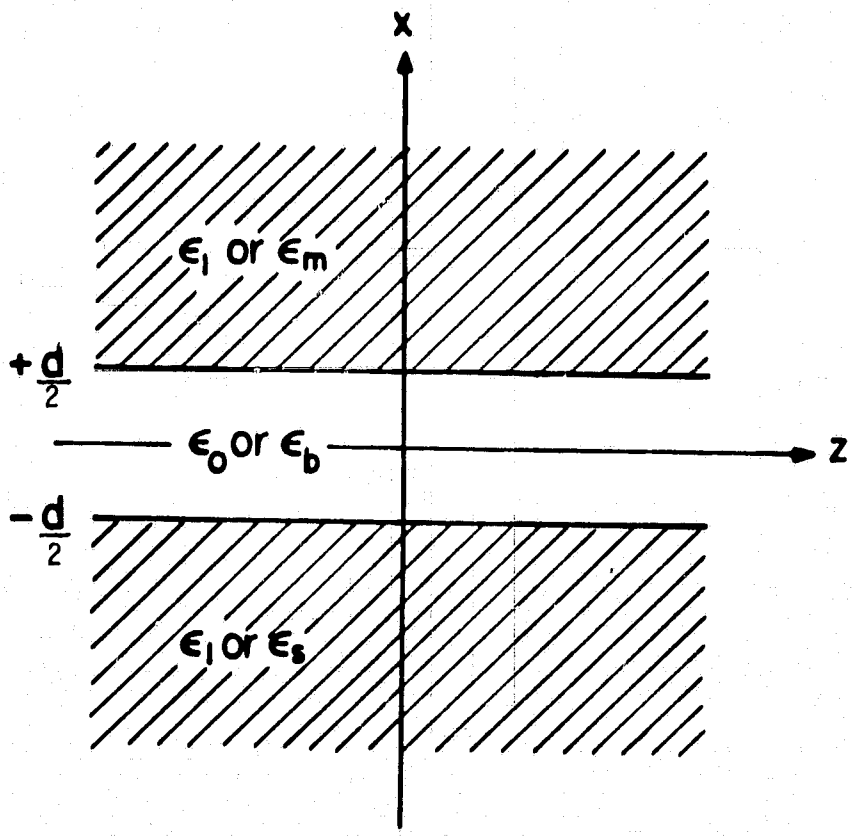


Fig. III-7. Schematic diagram of a metal-vacuum-metal planar structure. A free space gap of width d occupies the space between $x = +\frac{d}{2}$ to $x = -\frac{d}{2}$. Surface plasmon excitations are assumed to be propagating in the z-direction.

In the second case, the penetration into the metal is of the same order as for the single surface mode. It is this latter case which is most prevalent as will be seen. These expressions are valid for $\omega << \omega_p$, where ω_p is the plasma frequency. The resonance frequency for the Drude model is again given by $\omega_p/\sqrt{2}$, and the behaviour near resonance is very similar to the cases previously discussed. The dispersion using the Drude model is illustrated in Figs. (III-8) and (III-9). The circuit properties of this mode are discussed later in terms of its important role in the coupling of radiation to metal-barrier-metal junctions.

Although only planar structures have been considered here, the surface modes have been considered for other geometries such as the right circular cylinder and the ellipsoid of revolution. These could eventually play an important role in the coupling of radiation into tunneling devices.

Recently the plasmon resonance associated with spheroids and ellipsoids have become important because of the coupling through surface periodicities and roughness. This resonance follows straight-forwardly from the problem of a dielectric sphere in a static constant field. If E_0 is the field amplitude outside of the sphere then the potential inside the sphere ϕ_{in} is given by

$$\phi_{in} = -\frac{3}{\varepsilon+2} E_0 r\cos\theta \qquad \text{(III.2.11)}$$

For a dielectric constant which is negative and in particular given by the Drude model one observes that there is indeed a resonance. This has been investigated recently by Liao and others for the enhancement of the Raman effect, as well as the coupling to luminescent junctions in periodic arrays[10].

III-3 Propagation Length Limitations Due to Loss

In all of the idealized structures considered in the previous section the neglect of the loss is the most important approximation which has been made. For a tunneling converter, the stimulated absorption due to the tunneling transition must be much larger than the absorption due to the dielectric losses. The latter of course implies conversion to heat and hence a decrease in the efficiency of the device. As we see in this

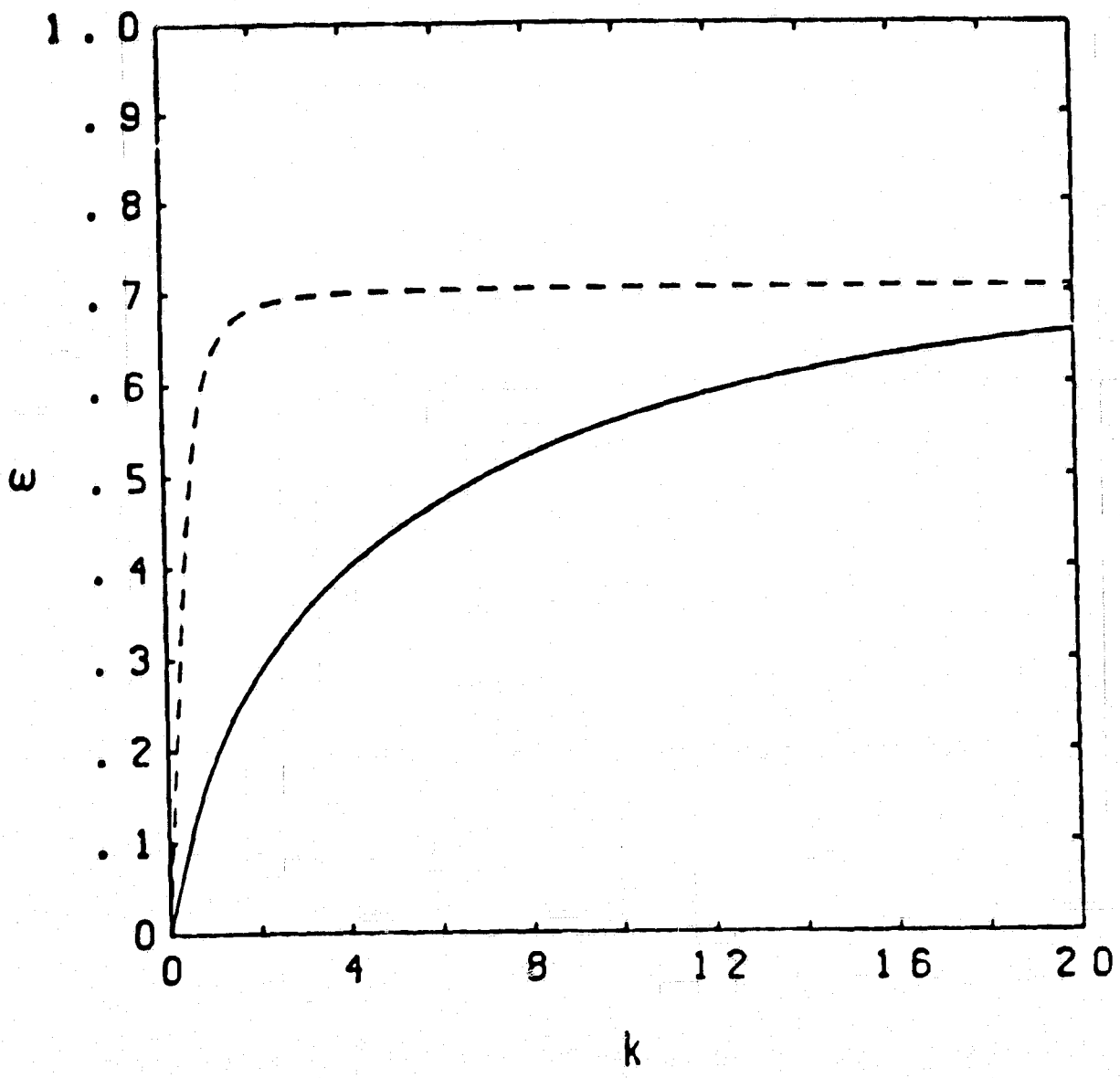


Fig. III-8. Dispersion of the anti-symmetric mode of surface plasmons in a small planar gap between ideal metals. The width is given by $(\omega_p/c)d = 0.10$. ω is normalized with respect to ω_p, and k with respect to (ω_p/c). The dash line represents that of excitations on a single planar surface. The phase velocity for $\omega \to 0$ is approximately 0.022c.

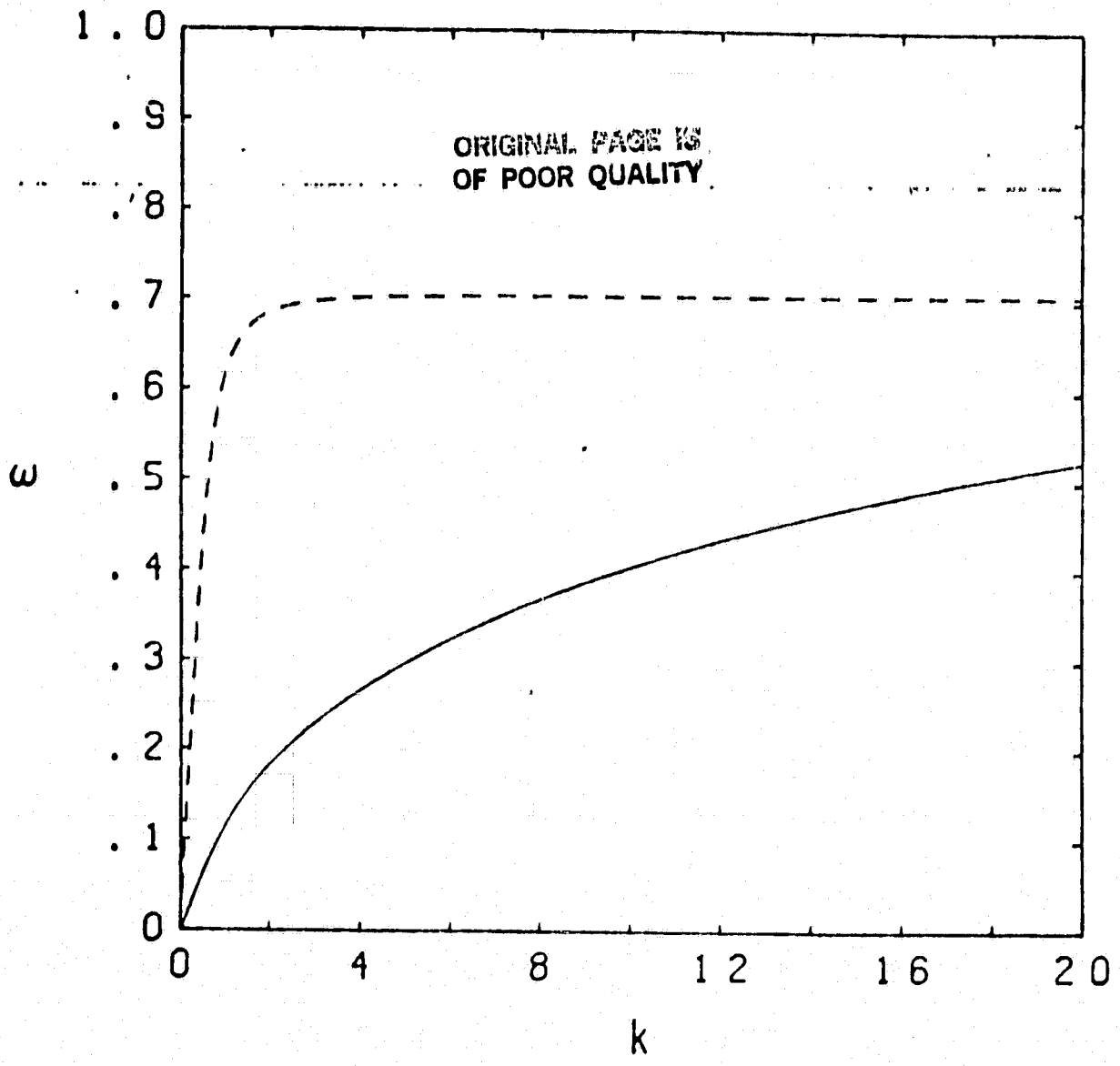


Fig. III-9. Dispersion of the mode for a width of $(\omega_p/c)\ d = 0.04$. The phase velocity for $\omega \to 0$ is approximately 0.014c.

section, this implies severe constraints on the junction devices, since the gap mode in particular propagates a very short distance.

The most straight-forward means by which to examine the loss is to include relaxation effects in the dielectric coefficients of the metals and the semiconductors. The simplest expressions by which to estimate the resultant propagation lengths are Eq. (III.2.4) for the single surface plasmon and Eq. (III.2.10) for the junction mode. The former provides the exact propagation distance, however the latter is an approximation, first of all, assuming that the same dielectric occurs on both sides of the tunneling junction, and secondly that one is not too close to the plasma resonances of the dielectric materials.

One assumes that the dielectric behaviour is described by

$$\varepsilon_1(\omega) = \varepsilon_m(1-\omega_m^2/(\omega^2+i\omega/\xi_m)) \qquad \text{(III.2.12)}$$

where ε_m, ω_m, and ξ_m are the ideal frequency independent dielectric constant as $\omega \to \infty$, the metallic plasma resonance frequency, and the relaxation time, respectively. One can then obtain approximate expressions for the propagation distances for various frequency ranges (a) the low (b) intermediate, (c) relaxation and (d) resonance.

a) Low Frequencies ($\omega\xi_m << 1$)

For low frequencies $\varepsilon_1(\omega) \cong i\varepsilon_m\omega_m^2\xi_m/\omega = i\sigma_m/\omega$ where σ_m is the d.c. conductivity. Then Eq. (III.2.4) gives for the propagation length of the single surface mode k" (k=k'-ik")

$$k'' = \tfrac{1}{2}k^2\left(\sigma_m\sqrt{\frac{\mu_0}{\varepsilon_0}}\right)^{-1} \qquad \text{(III.2.13)}$$

and $k' = k = \omega\sqrt{\mu\varepsilon_0}$. Thus $k''/k' = \tfrac{1}{2}(k)\left(\sigma_m\sqrt{\frac{\mu_0}{\varepsilon_0}}\right)^{-1}$ a long distance ($\sigma_m \cong 10^7$, thus $(k'')^{-1} \cong (10^6)\lambda^2$ cm). The penetration depth into the metal reduces to the skin depth and the penetration into the air is given by $\gamma_1 \cong \omega^{\frac{1}{2}}\mu_0^{\frac{1}{2}}\varepsilon_0/(\sigma_m i)^{\frac{1}{2}} \cong .01\ k_0$ for $\omega \cong 10^{13}$ rad/sec.

The metal-barrier-metal or gap mode propagation distance follows from Eq. (III.2.10b) and in the low frequency conductivity limit gives for k = k'-ik"

$$k \cong \frac{\omega^{\frac{3}{4}}\mu_0^{\frac{1}{4}}\epsilon_0^{\frac{1}{2}}}{d^{\frac{1}{2}}\sigma_m^{\frac{1}{4}}}(.92 - i.38) \qquad \text{(III.2.14)}$$

where the .38 and .92 are respectively $\cos(\frac{\pi}{8})$ and $\sin(\frac{\pi}{8})$. The most striking characteristic is the $\omega^{\frac{3}{4}}$ dependence for both k' and k".

Thus the real part of k in this region is approximately 2.4 times the imaginery part and the ratio of $k'/k_0 \cong (2.42)k''/k_0 \cong 1/(d^2\omega\mu_0\sigma_m)^{\frac{1}{4}} \cong (\delta/d)^{\frac{1}{2}}$. The propagation distance is thus of the order of $\frac{\lambda}{10}(\frac{d}{\delta})^{\frac{1}{2}}$ which is about $10^{-2}\lambda$ at 10^{12} radians per second, an extremely short distance (≅18μm)

b) Intermediate Frequencies

At higher frequencies for which the real part of the dielectric begins to contribute significantly the expression (2.14) for k of the gap mode is replaced by

$$k \cong \frac{\omega^{\frac{3}{4}}\mu_0^{\frac{1}{4}}\epsilon_0^{\frac{1}{2}}}{d^{\frac{1}{2}}\sigma_m^{\frac{1}{4}}}(\cos(\tfrac{\pi}{8} - \omega\xi_m/4) - i\sin(\tfrac{\pi}{8} - \omega\xi_m/4)) \qquad \text{(III.2.15)}$$

where $\omega\xi_m << 1$

This shows that the real part continues to increase as approximately $\omega^{\frac{3}{4}}$ but that the imaginery part increases less rapidly with increasing frequency. If ξ_m is long enough, k" can in fact decrease and thus the propagation distance can remain constant or decrease slightly. For room temperature this is not expected to be observable. For the single surface mode expression (2.13) given above is still valid in this frequency region.

c) Relaxation Region

This corresponds to frequencies for which the real part of the dielectric becomes comparable to the low frequency conductivity contribution or is the dominant contribution. This is expected to be the most important region with regards to convertors since this, aside from interband absorption, is applicable for the optical portion of the spectrum. The dielectric coefficient can be approximately written as

$$\varepsilon_1(\omega) \cong +\varepsilon_m\left(\frac{-\omega_m^2}{\omega^2} + i\frac{\omega_m^2}{\omega^3\zeta_m}\right) \qquad \text{(III.2.16)}$$

where it is assumed that $\omega\zeta_m >> 1$.
Then one obtains from Eq. (III.2.4) for the single surface mode

$$k \cong k_0\left(1+\frac{\varepsilon_0}{2\varepsilon_m}\right) + i\,\frac{\varepsilon_0}{2\varepsilon_m(c\zeta_m)} \qquad \text{(III.2.17)}$$

where c is the speed of light. Consequently the propagation distance is roughly a constant of the order of 3 μm for a relaxation time (ζ_m) of the order of 5×10^{-15} sec. Since for metals $\varepsilon_0 \cong \varepsilon_m$ the propagation distance is roughly given by the relaxation time in wave numbers.

In this region of the spectrum, the penetration of the mode into the dielectric is given by $\gamma_0 = k_0\left(\frac{\varepsilon_0}{\varepsilon_m}\right)^{\frac{1}{2}}\frac{\omega}{\omega_m}$, the high frequency limit, and the penetration into the metal is given by $\gamma_1 \cong \omega\sqrt{\mu\varepsilon_m}\left(\frac{\omega_m}{\omega}\right)$.

For the gap mode Eq. (III.2.10) with this approximate value of $\varepsilon_1(\omega)$ gives

$$k^2 \cong \omega^2\mu_0^{\frac{1}{2}}\varepsilon_0/(\varepsilon_m^{\frac{1}{2}}d\omega_m)\left(1+i\left(\frac{1}{2\omega\zeta_m}\right)\right) \qquad \text{(III.2.18)}$$

One expects for $\omega\zeta_m >> 1$, the real part k' to be proportional to ω and k" to be approximately constant and given by

$$(k') \cong \omega(\mu_0/\varepsilon_m)^{\frac{1}{4}}c_0^{\frac{1}{2}}/(d\omega_m)^{\frac{1}{2}}$$

$$(k'') \cong \frac{1}{(\omega_m\varepsilon_m^{\frac{1}{2}}\mu_0^{\frac{1}{2}})^{\frac{1}{2}}}\frac{1}{d^{\frac{1}{2}}}\frac{1}{2(c\zeta_m)} \cong \frac{1}{(10^{-5})}\ \text{cm}^{-1} \qquad \text{(III.2.19)}$$

assuming $k_m = (\omega_m\sqrt{\varepsilon_m\mu_0}) \cong 10^6\ \text{cm}^{-1}$, $d \cong 10^{-7}$ cm and $\zeta_m \cong 10^{-15}$ sec. To observe the constant k", ζ_m must be long enough to allow this region to be observable without having the wing of the resonance peak dominate the behaviour.

d) Resonance Behaviour

For frequencies for which the influence of the resonance comes into play, one can assume in Eq. (III.2.4) that ε_1 in the numerator is given

predominantly by the real part and that this is approximately given by $-\varepsilon_0$. One then obtains

$$k^2 \cong \frac{\omega_r^3 \mu_0 \varepsilon_0^2}{(\varepsilon_0+\varepsilon_m)^2} \times \frac{1}{(\omega-\omega_r+i\frac{\omega_r}{\omega_m}\times\frac{1}{2\zeta_m})} \qquad \text{(III.2.20)}$$

which displays the resonance profile shown in Eq. (1.4) where $\omega_r = \omega_m(\frac{\varepsilon_m}{\varepsilon_m+\varepsilon_0})^{\frac{1}{2}}$. This expression can be used to obtain the propagation distance. At resonance this is $\omega_r(\frac{\mu_0\varepsilon_0^2\omega_m\zeta_m}{\varepsilon_0+\varepsilon_m})^{\frac{1}{2}}$. Eqs. (2.10a) and (2.10b), although not expected to be strictly valid, show that the gap mode displays a similiar resonance determined approximately by Re $\varepsilon_1(\omega) = 0$, that is, close to ω_m. These two expressions (2.10a and 2.10b) become

$$\beta \cong \frac{1}{d}\frac{\varepsilon_0}{\varepsilon_m} \times \omega^2 \frac{1}{(-\omega^2+\omega_m^2-i(\frac{\omega_m}{\zeta_m}))} \qquad \text{(III.2.21)}$$

and

$$\beta^2 \cong \frac{1}{d}\omega^2 \frac{\mu_0^{\frac{1}{2}}}{\varepsilon_m^{\frac{1}{2}}}\varepsilon_0 \times \frac{1}{(-\omega^2+\omega_m^2-i(\frac{\omega_m}{\zeta_m}))^{\frac{1}{2}}} \qquad \text{(III.2.22)}$$

The first predicts that at resonance $\beta' = 0$ and $\beta'' = \frac{1}{d}\frac{\varepsilon_0}{\varepsilon_m}(\zeta_m\omega_m)$ and that a Lorentzian profile is obtained. The second gives $\beta^2 \cong \frac{\omega_m}{d}(\omega_m\zeta_m)^{\frac{1}{2}}\mu_0^{\frac{1}{2}} \times \frac{\varepsilon_0}{\varepsilon_m^{\frac{1}{2}}}(-i)^{\frac{1}{2}}$ so that $|\beta''| \cong \beta'$ and above the resonance the character of the mode changes. It is this latter behaviour which is found by solving numerically.

In Fig. (III-10), we have shown propagation distances for the open surface mode and the junction mode using numerical evaluation of expressions (2.4) and (2.10b) for the case of Indium ($h\omega_m = 11$eV, $\varepsilon_m \simeq \varepsilon_0$). One observes that the propagation distance of the open surface mode is tolerable throughout the visible portion of the spectrum at room temperature (~ 50. μm).

The gentle rise throughout the visible is due to the resonance which is more pronounced for the single surface mode.

Below resonance the relaxation results are extracted for $\omega << \omega_m$. Eq. (2.22) gives Eq. (2.17) as is seen by expanding the denominator. Eq. (2.21) gives

$$\beta \cong \frac{1}{d} \frac{\varepsilon_0}{\varepsilon_m} \frac{\omega^2}{\omega_m^2}\left(1+i\frac{1}{\varepsilon_m \omega_m}\right) \qquad \text{(III.2.23)}$$

which provides a quadratic dependence on ω as might be expected in the blue portion of the visible spectrum.

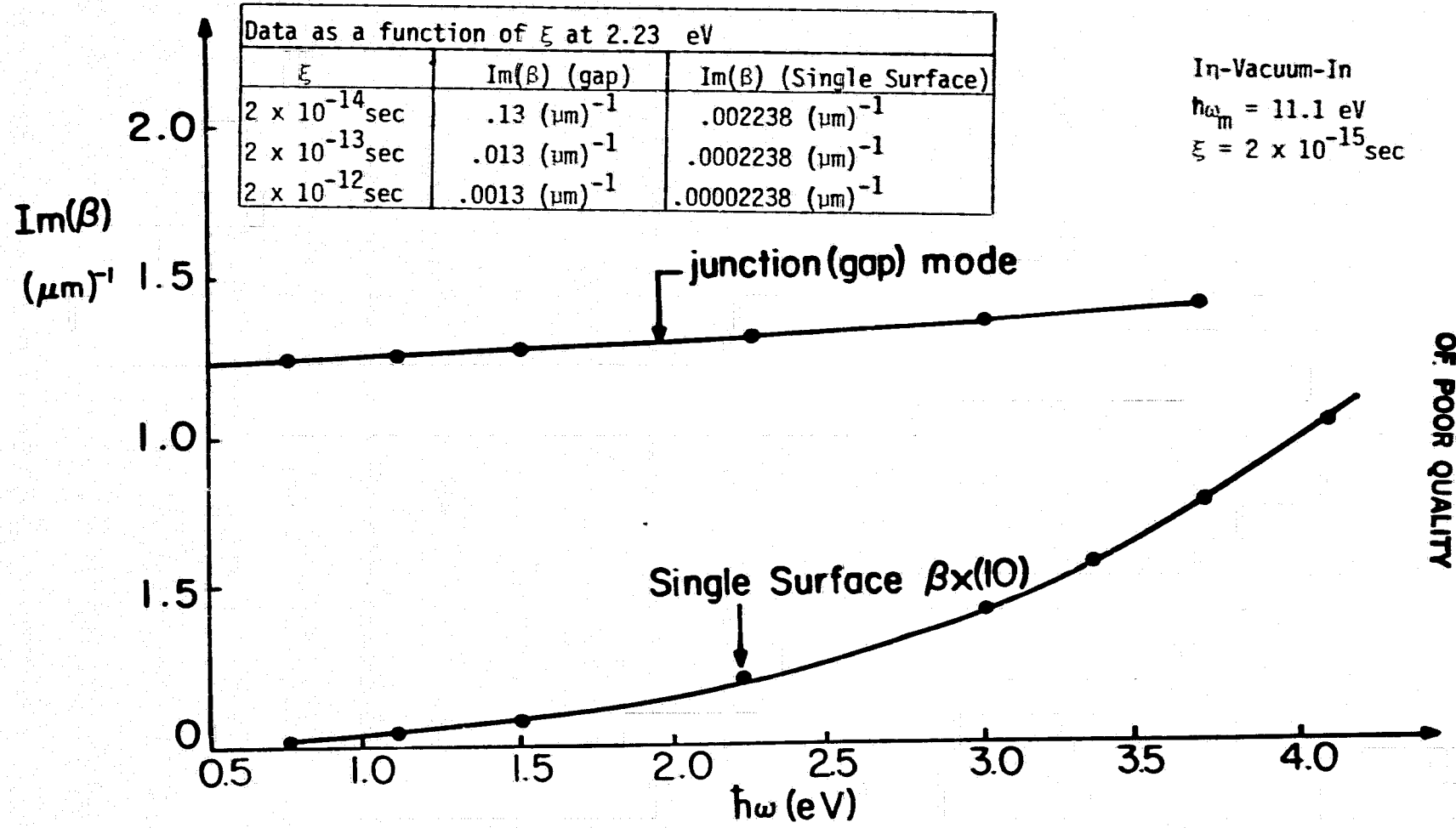
For the case of a metal there is considerable advantage to cooling since the relaxation time increases with a decrease in temperature, and this is proportional to the propagation distance. In Fig. (III-10), we list the propagation distance for the open surface and junction modes as a function of the relaxation times and show that the propagation distance is enhanced considerably. There may be additional inherent advantage to running the junctions cooled to obtain such relaxation times. One would expect the tunneling characteristics to be more ideal as the temperature is lowered and thus it would be expected that the transition probability for the absorption processes should increase slightly with a drop in temperature.

There does not appear to be much data available in the literature relating to the temperature dependence of the relaxation times of metals. However the data which is available does indicate that the relaxation is in fact phonon dominated and can thus be considerably lengthened by operating at reduced temperatures, and in particular the photon pump processes might be able to exceed thermalizing processes reasonably well but that the propagation distance of the junction mode is extremely short ($\cong$ .5μm). More work needs to be done to investigate such relaxation processes at lowered temperatures.

III-4 Numerical Solutions

We have in addition to the above analytical approximations to the junction mode, numerically obtained the dispersion characteristics of the gap mode. This is appropriate when the dielectric material on one

Data as a function of ξ at 2.23 eV		
ξ	Im(β) (gap)	Im(β) (Single Surface)
2×10^{-14} sec	.13 $(\mu m)^{-1}$	.002238 $(\mu m)^{-1}$
2×10^{-13} sec	.013 $(\mu m)^{-1}$	.0002238 $(\mu m)^{-1}$
2×10^{-12} sec	.0013 $(\mu m)^{-1}$	.00002238 $(\mu m)^{-1}$



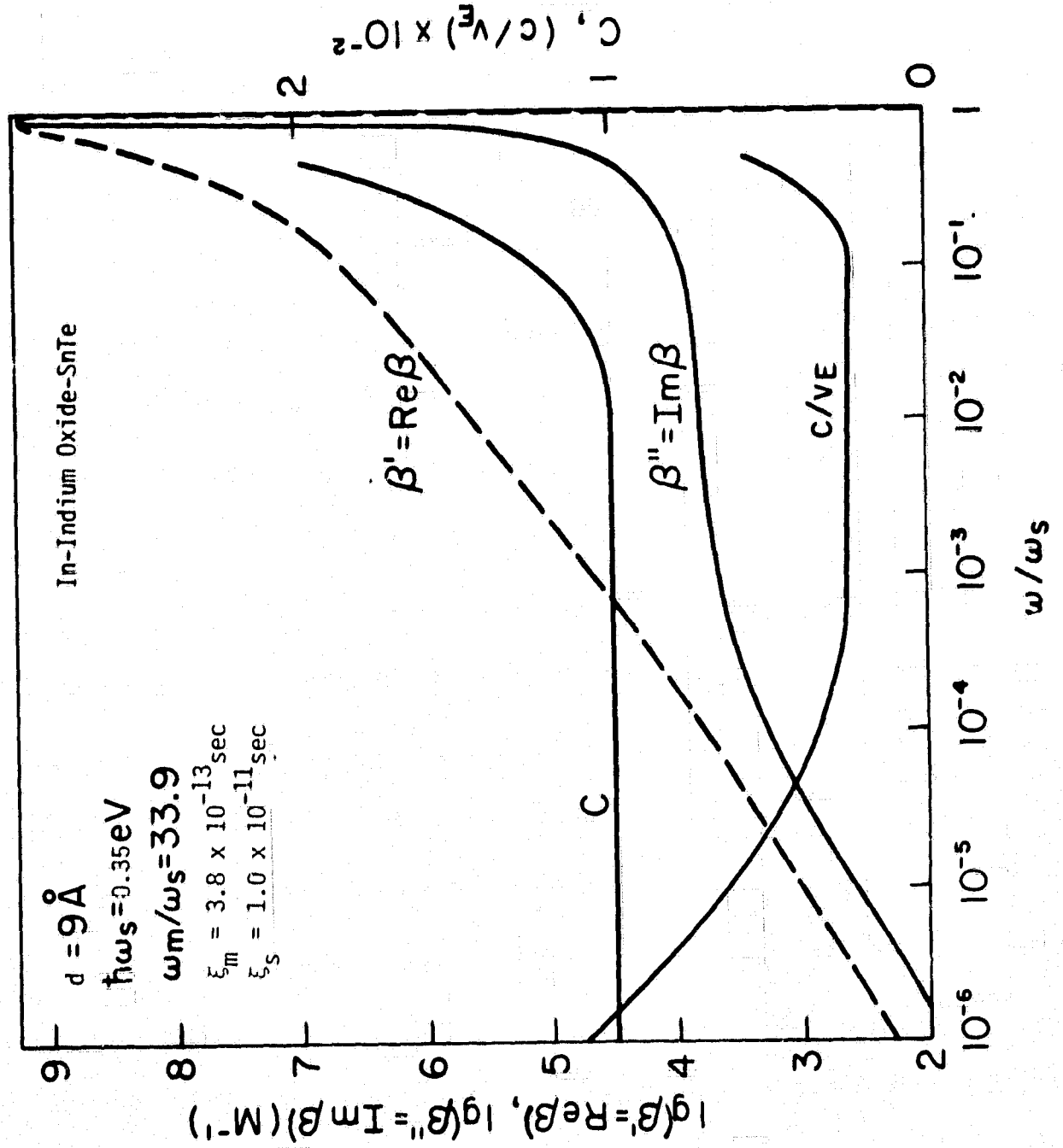
III-10 Propagation length of single surface metal-vacuum plasmons and the metal-vacuum-metal anti-symmetric mode throughout the visible portion of the spectrum. ω_m the plasma frequency was chosen to be 11.0 eV and the relaxation time is 2×10^{-15} sec. for the

side of the tunneling barrier differs from the material on the other side of the barrier or for more complicated barrier structures that will probably need to be considered in the future. For the case which we have explicitly considered, the dispersion expression is given by

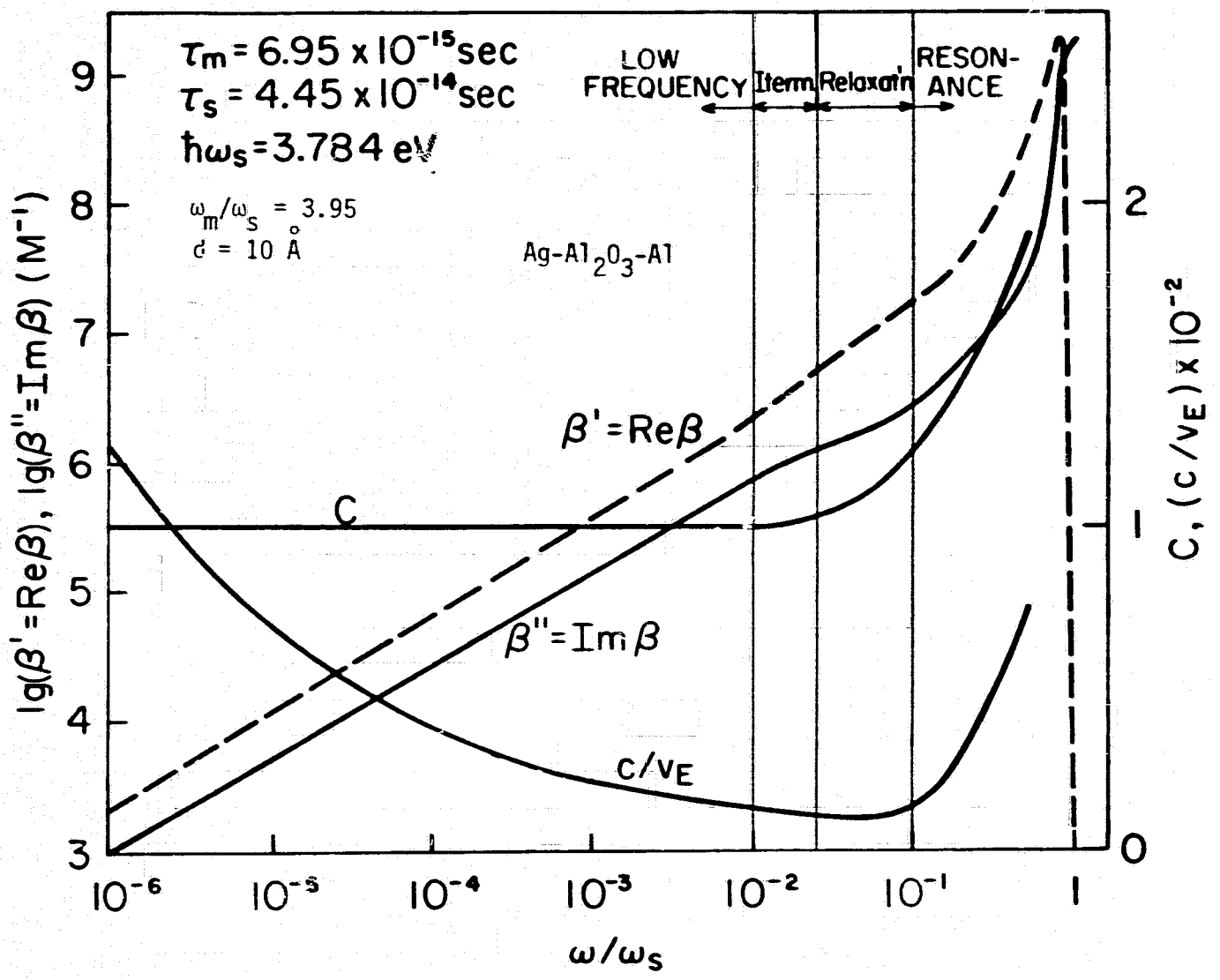
$$\left[\frac{\varepsilon_I}{\gamma_I}+\frac{\varepsilon_s}{\gamma_s}\right]\left[\frac{\varepsilon_I}{\gamma_I}+\frac{\varepsilon_M}{\gamma_M}\right] e^{\gamma_I d} = \left[\frac{\varepsilon_I}{\gamma_I}-\frac{\varepsilon_s}{\gamma_s}\right]\left[\frac{\varepsilon_I}{\gamma_I}-\frac{\varepsilon_M}{\gamma_M}\right] e^{-\gamma_I d} \qquad \text{(III.4.1)}$$

where the transverse propagation constants γ_s, γ_I, γ_m are given by $\gamma = [\beta^2 - \omega^2\mu_0\varepsilon]^{\frac{1}{2}}$ with $Re(\beta) > 0$ and ε_m, ε_b, ε_s the dielectric constants of the respective regions shown in Fig. (III-7). They are taken to be $\varepsilon_m(\omega) = \varepsilon_m(1-\omega_m^2/(\omega^2+i\omega/\zeta_m))$, $\varepsilon_I = \varepsilon_b$ and $\varepsilon_s(\omega) = \varepsilon_s(1-\omega_s^2/(\omega^2+i\omega/\zeta_s))$, where ζ_s, ω_s, ζ_m, ω_m are the relaxation times and plasma frequencies of the semiconductor and metal respectively, and ε_I, ε_s are the ideal frequency independent dielectric constants as $\omega \to \infty$ in this model. Fig. (III-11) illustrates an example calculation for a dispersion curve using this numerical procedure. For this particular case ω_m was arbitrarily chosen for Aluminum ($\hbar\omega_m \cong 11eV$) and the behavior is dominated by ω_s and ζ_s. A relaxation time $\zeta_m \cong 4.0 \times 10^{-13}$ sec. was chosen as well as $\hbar\omega_s$ = .35eV (heavily doped semiconductor ($\cong 5\times10^{20}cm^{-3}$)) and $\zeta_s = 1 \times 10^{-11}$sec. The relaxation times are approximately 10^2 to 10^3 longer than expected but should be achievable by cooling to reduce phonon domination. These results should thus provide an indication of typical behavior expected of cooled metal-barrier-heavily doped semiconductor devices. One observes a much reduced loss. Fig. (III-12) and (III-13) show similiar calculations for Al-Al_2O_3-Ag and Al-Al_2O_3-Al junctions respectively. Fig. (III-12) also indicates the various regions of the dispersion curve.

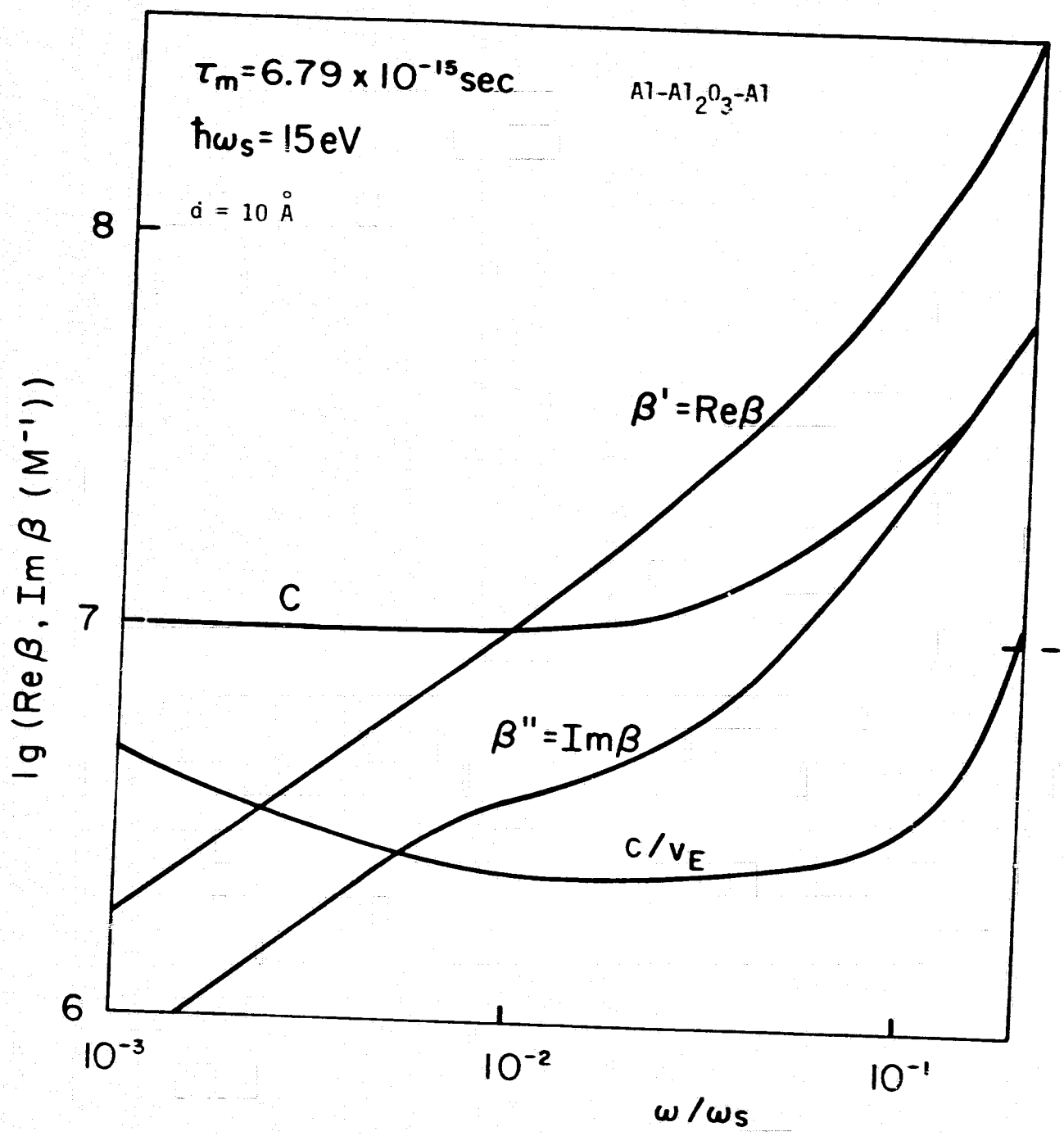
The log-log scale clearly illustrates that at low frequencies one observes the $\omega^{3/4}$ dependence as suggested by Eq. (III.2.14) with the real part being greater than the imaginery part by approximately a factor of 2.5. At higher frequencies for which the real parts of the dielectric coefficients contribute significantly along with the imaginary parts (relaxation region) the dispersion deviates from such a dependence. One observes an ω dependence for β' and an almost frequency independent β'' (as expected in the relaxation region).



III-11 Numerical solution for the dispersion of an anti-symmetric plasmon mode for a 10 Å barrier junction with the metal plasma frequency, ω_m = 11 eV and ω_s the plasma frequency of the second electrode representative of SnTe (.35eV). The relaxation times chosen are representative of a cooled device (≈10°K).



III-12 Numerical solution for the gap-mode dispersion of Ag-Al_2O_3-Al junctions at room temperature.



III-13 Gap mode dispersion characteristics shown for room temperature Al-Al$_2$O$_3$-Al junctions.

One observes from Fig. (III-12) that for $\hbar\omega \simeq$ 1eV, one is in this intermediate range between the relaxation behavior and the resonant behavior. A value of β'' of approximately 2×10^{6} M^{-1} or a propagation distance $\simeq$ .5 μm for the $\zeta_m \simeq 6\times10^{-15}$ sec can be expected. This is a reasonable propagation distance with respect to the absorption which could be anticipated from relatively nonlinear devices (see Eq. II.5.6). This propagation length, also in good agreement with Eq. (2.18), could also be lengthened by cooling. Since the relaxation time decreases linearly down to the Debye Temperature, $|\theta|$, and as T^5 thereafter Eqs. (III.2.18) and (III.2.22) indicate that by cooling to approximately 10° K the propagation distance is increased to at least 200 μm and thus the tunneling transition rate can easily compete with the plasmon propagational loss.

There are two other parameters which are significant when one is concerned with maximizing the transition rate associated with the tunneling phonomena. The first is the confinement factor or the fraction of power that is flowing within the tunneling layer. This can be large and can exceed unity since the negative dielectrics cause reverse power flow. For the region of interest, Fig. (III-11) indicates that this confinement factor C is close to unity. The power confinement is excellent although the energy is principally stored outside of the tunneling barrier. This is of course due to both the large negative dielectric coefficients as well as the longitudinally directed electric field component of the plasmon mode. The latter becomes important near the resonant frequency of the mode. The question of power and energy confinement needs more consideration.

The second parameter of importance is the slowing factor, the ratio of the speed of light to the energy velocity associated with the mode and defined in the first chapter. One observes from Fig. (III-11) that this exhibits divergent behaviour at low frequencies due to the influence of the conductivity but that for the frequency range of interest it is relatively constant with a value of 200 confirming the number used in Chapter II to estimate the absorption coefficient due to the converting transitions. We are presently in the process of calculating the dispersion characteristics utilizing more accurate values for the dielectric parameters as a function of temperature. In addition, in the visible region

of the spectrum, for most metals since interband transitions add at least a small amount of additional absorption, these are being included in the calculations. Finally more complicated multilayer tunneling structures having the form of metal-semiconductor-tunneling barrier-semiconductor-metal might have to be considered and present more difficulty in calculations.

References for Chapter III

[1] K.L. Kliewer and R. Fuchs, Phys. Rev. 153, 498 (1967).

[2] E.N. Economou, Phys. Rev. 182, 539 (1967).

[3] R.H. Ritchie, Surf. Sci. 34, 1 (1973).

[4] C.A. Pfeiffer, E.N. Economou, and K.L. Ngai, Phys. Rev. B10, 3038, (1974).

[5] Ravinder K. Jain, "Photon-Assisted Electron Tunneling in Metal-Barrier-Metal Junctions," ERL Memorandum No. ERL-M522, June 1975.

[6] H.R. Philipp and H. Ehrenreich, J. Appl. Phys. 35, 1416 (1964).

[7] H. Ehrenreich, in The Optical Properties of Solids, edited by J. Tauc (Academic Press, New York, 1966), p. 106.

[8] J. Schoenwald, E. Burnstein, and J.M. Elson, Solid State Commun. 12, 185 (1973).

[9] A.S. Barker Jr., Phys. Rev., B8, 5418 (1973).

[10] See P.F. Liao, Appl. Phy. B28, 229 (1982). This has reference to some of the recent work on surface enhanced Raman scattering. The papers following in this section also provide excellent references to the literature. See also P.F. Liao in "Laser Spectroscopy" edited by A.R.W. McKellar, T. Oka and B.P. Stoicheff, Springer Series in Optical Sciences (Springer-Verlag, Berlin 1981).

[11] J. Daniels, Z. Phys. 203, 235 (1967).

[12] J. Brambring and H. Raether, Phys. Rev. Lett. 15, 882 (1965).

[13] ibid, Z. Phys. 199, 118 (1967).

[14] J. Brambring, Z. Phys. 200, 186 (1967).

[15] C. Kunz, Phys. Lett. 15, 312 (1965).

[16] J. B. Swan, Phys. Rev. 135, A1467 (1964).

[17] C. Kittel, Introduction to Solid State Physics, (Wiley, New York, 4th ed., 1971).

[18] J. Geiger and K. Wittmaack, Z. Phys. 195, 44 (1966).

[19] E.T. Arakawa, R.J. Herickhoff and R.D. Birkhoff, Phys. Rev. Lett. 12, 319 (1964).

[20] B. Feuerbacker and B. Fitton, Phys. Rev. Lett. 24, 499 (1970).

IV COUPLING BETWEEN FREE SPACE RADIATION MODES AND JUNCTION SEW MODES

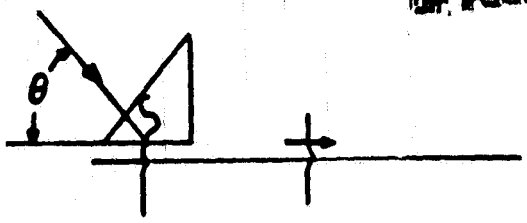
IV-1 Introduction and Overview

In order to extract energy from free space radiation modes, such modes must be coupled with low loss to the well confined surface electromagnetic junction modes. The coupling devices required to accomplish this must basically guide the incident field from dimensions of the order of microns to dimension of the order of Ångstroms, a four order of magnitude change in the confinement distance. It is quite likely that this coupling will involve the intermediary of a single surface plasmon wave. Thus the coupling will involve two steps; the coupling of free space modes to single surface SEW's and subsequently the coupling of these to the highly confined junction modes.

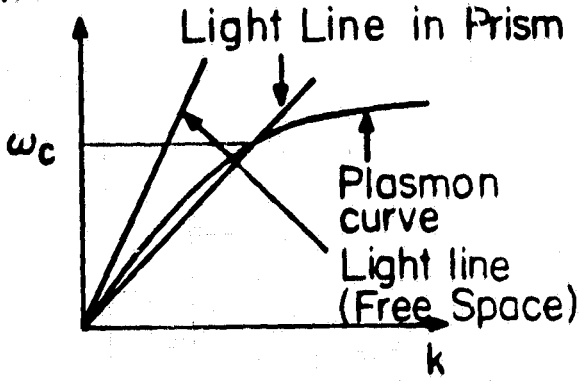
Thus one must basically consider three problems involved with the coupling. 1) The coupling of the incident field to a guided single surface electromagnetic mode; 2) the coupling of the single surface electromagnetic mode to the confined junction mode, and 3) the coupling of the junction mode to the electron tunneling current.

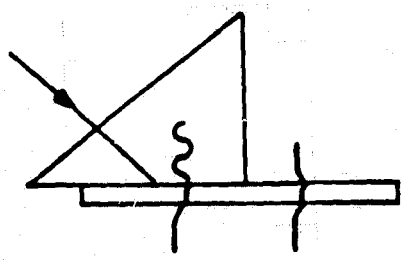
In the present chapter we wish to consider the first two coupling problems and report on the progress which has been made to alleviate them involving both our own efforts and those of other groups.

An overview of various coupling schemes which have been reported is shown in Fig. (IV-1). In the first three the technique is to obtain an overlap of the free space radiation modes propagation dispersion relationship and the dispersion relation for the plasmon modes. In the last Fig. (IV-1(d)) the diffraction pattern of the plasmon mode from an abrupt termination is utilized as a radiative aperture for the coupling. The first is analogous to integrated optical coupling (the Otto configuration[1]) The prism refractive index n is large enough to allow the radiative mode phase velocity $\frac{\omega}{c}n$ to exceed the plasmon mode phase velocity. A finite angle θ is thus obtained for an intersection of the two dispersion curves and hence the possibility of coupling. In the second technique (the Kretschman configuration[2]), the plasmon mode on the bottom surface of the metallic film is evanescently coupled to the radiative mode utilizing the exponential decay in the metallic film. The length

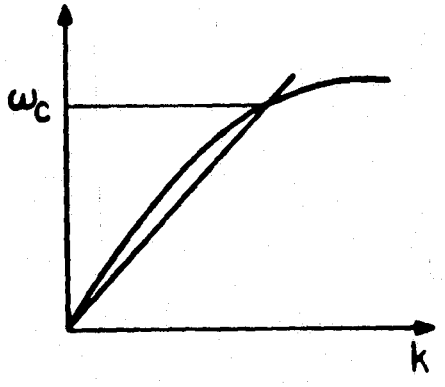


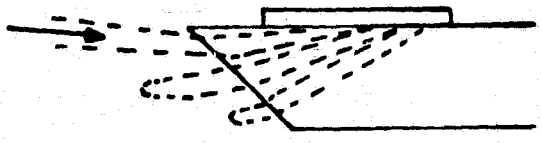
(a)



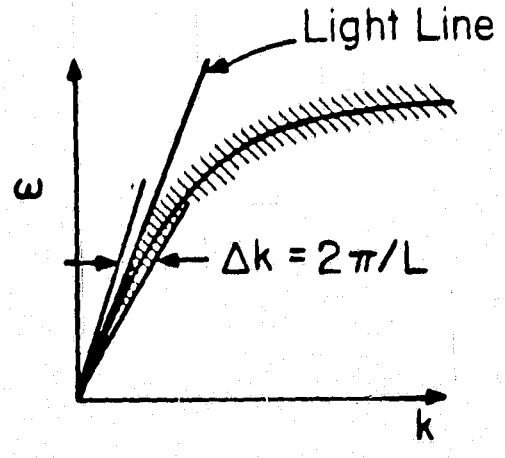


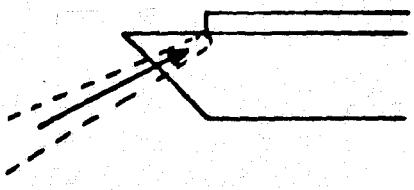
(b)





(c)





(d)

Diffraction from
end illumination
like semi-conductor laser

IV-1 Various techniques for coupling to the surface plasma mode.

confinement coupling of Fig. (IV-1(c)) is easily understood in terms of the uncertainly relationship $\Delta kL \geq 2\pi$. Thus $\Delta k \geq (2\pi/L)$. This smears out the dispersion curve and causes an overlap with the light line and hence coupling. The lobe pattern results from the detailed spatial frequency spectrum of the confined mode, which has a total width of Δk. We shall consider this in detail subsequently. One observes that if the length is infinite the SPW does not radiate in this configuration. Various combinations of the above schemes are of course possible, in particular length confinement and evanescent wave coupling. Length confinement results in side lobes on the evanescent wave coupling characteristics. Similarly, the boundary conditions associated with the end coupling characteristics will influence the length confinement mode pattern. The prism coupling schemes to date have not achieved adequate coupling for efficient energy conversion (on the order of a few percent)[3], although it is felt that optimization could increase the coupling to acceptable levels.

Prism coupling using end radiation has achieved the highest coupling efficiency of any method[4]. Losses are due to parasitic scattering and inefficient mode pattern matching. It is felt that this method is potentially promising for efficient coupling, as optimization reduces to edge preparation and mode matching by a suitable lens. It is also only slightly frequency dependent.

Length confinement as an extension of antenna theory[5,6] takes into account the high frequency dielectric behaviour. It thus should prove possible to have well designed antenna arrays for matching to the radiation field and thus these should be potentially efficient couplers.

Grating coupling[7] has also been used for coupling single frequency radiation to the SPW, but suffers from low efficiency due to reradiation and coupling to higher order modes. Grating fabrication, being a planar process, is however directly compatible with thim film technology which is advantageous for the construction of thin film energy converters and progress has recently been made on increasing the efficiency at submillimeter wavelengths[7]. Gratings are producible through direct photolithography, laser interference, or laser photodeposition[8]. The latter method shows promise as the SPW coupling is optimized, i.e. the grating profile grows in such a manner as to maximize the SPW intensity in the presence of

a radiation field. It has been shown that in the presence of dielectric loss, total absorption of the monochromatic incident wave is possible[9] with a suitable grating profile. As in the evanescent coupling schemes, however most of the energy is absorbed in the materials, hence actual mode conversion efficiency is low.

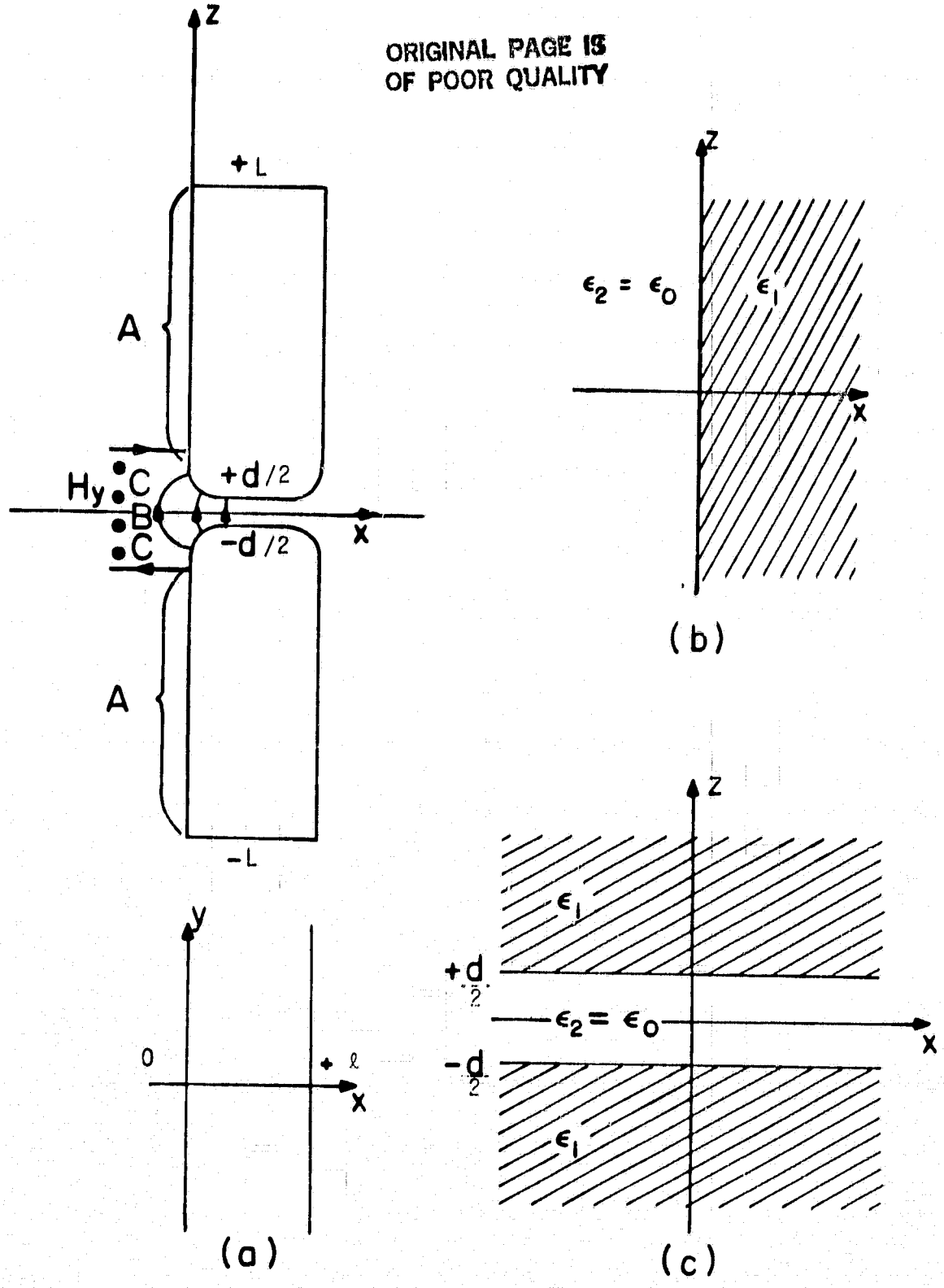
The strong dependence of coupling angle on frequency makes periodic gratings unsuitable for wideband couplers. Related methods, e.g. aperiodic gratings and surface roughness[10], show some promise in this application, but again coupling efficiency is fundamentally low. Nonlinear optical surface studies[11] and investigations of laser induced surface gratings[12] have recently been reported and should result in development in this area. In addition enhanced plasmon mode coupling to an array of junctions at visible wavelength has recently been reported[13].

In the present chapter of this report we consider primarily length confinement coupling since this is the area in which our concentration has been. We briefly consider taper coupling which can provide efficient conversion from the single surface plasmon mode to the well-confined gap mode and which we are continuing to pursue. In the final section we report on recent theoretical investigations involving grating coupling to the single surface mode.

IV-2 Length Confinement Coupling on a Planar Structure

Although a true surface mode on an infinite surface does not interact with the radiation field because of a lack of phase-matching, for a finite structure such a coupling is possible. In this section we show that such surface plasmon excitations on a linear structure properly reduce to conventional antenna characteristics at low frequencies. It is also demonstrated that such plasmon excitation provides efficient coupling between the radiation field and a metal tunneling structure up to a frequency at least as high as the plasma resonance. We have calculated the radiative patterns of such length confinement coupling characteristics.

The ideal structure of interest is that shown in Fig. (IV-2(a)). For simplicity we consider a planar geometry for the plasmon mode on the boundary. Such a structure supports surface modes in regions A which are coupled to the radiation field, and an antisymmetric "gap" mode in region B which can be efficiently excited by conversion from the former mode.



IV-2 Idealized metal-barrier-metal coupling structure.

The basic surface plasma mode is that propagating along an infinite plane surface separating two regions with dielectric constants ε_1 and ε_2 respectively, as shown in Fig. (IV-2(b)). From Chapter III, the dispersion of such a mode at frequency ω is given by:

$$(\varepsilon_1/\nu_1) + (\varepsilon_2/\nu_2) = 0 \qquad \text{(IV-1)}$$

where ν in each respective region is defined by $\nu^2 = k^2 - \omega^2\mu\varepsilon$, and k is the propagation constant for the mode.* For real metals ε is in general complex due to losses, and k is complex: $k = \beta - i\alpha$. The range given by $1/2\alpha$ is usually much larger than a wavelength, except near resonance as we have seen.

In the region of the gap (region B), which is formed by two plane surfaces, of the two non-radiative modes which can exist the anti-symmetric is of interest. It has been shown that the dispersion for infinite extent of the surfaces shown in Fig. (IV-2(c)) is given by:

$$(\varepsilon_1/\nu_1^g)\sinh(\nu_2^g d) + (\varepsilon_2/\nu_2^g)\cosh(\nu_2^g d) = 0 \qquad \text{(IV-2)}$$

as shown in Chapter 3. The corresponding z-polarized component of the electric field is given by $E_z(x.z.t) = E_g(z)e^{i(\omega t - k_g x)}/2 + c.c.$, where k_g is the propagation constant for the gap mode obtained from Eq. (IV-2) (as above, $\nu^2 = k_g^2 - \omega^2\mu\varepsilon$), and the complex amplitude $E_g(z)$ is

$$E_g(z) = E_i \cosh(\nu_2^g z)/\cosh(\nu_2^g d/2), \quad -\frac{d}{2} < z < +\frac{d}{2} \qquad \text{(IV-3)}$$

The magnetic field component associated with this surface electromagnetic mode is y-polarized with amplitude (from $H_y = \frac{i\omega\varepsilon_2 E_z}{ik_g}$) $H_g(z)$ is given by

$$H_g(z) = \frac{\omega\varepsilon_2 E_i}{k_g}\cosh(\nu_2^g z)/\cosh(\nu_2^g d/2), \quad -\frac{d}{2} < z < +\frac{d}{2}$$

$$= H_i \cosh(\nu_2^g z)/\cosh(\nu_2^g d/2) \qquad \text{(IV-4)}$$

Although in general k_g cannot be expressed in closed form, for frequencies such that the penetration depth into the metal is much larger than the gap width, which is generally true for M-B-M structures,

---

* Note: in accordance with Fig. (IV-2(c)) we use $\varepsilon_2 = \varepsilon_0$ and $\gamma_0 = \gamma_2$. A superscript g is used for the gap mode.

$\nu_2^g \approx k_g$, and $\nu_1^g \approx \sqrt{-\mu\omega\varepsilon_1}$ and one thus obtains the approximate results obtained earlier and shown numerically to give at least quantitatively reasonable results:

$$k_g^2 = (k_g' - ik_g'')^2 \approx \omega\sqrt{\mu\varepsilon_2}\,(\varepsilon_2/-\varepsilon_1)^{\frac{1}{2}}(\tfrac{1}{d}) \qquad \text{(IV-5)}$$

This corresponds to the relatively flat region for $\beta''(\omega)$ from Fig. (III-11).

The characteristic impedance of the anti-symmetric mode for a width 2b in the y-direction is given by the ratio of the voltage across the gap to the current induced on the upper electrode; that is

$$Z_g = 2\int_0^d E_g(z)dz/(\omega\varepsilon_2 E_i/k_g)$$

$$= 2(\tfrac{d}{b})\,\frac{k_g}{\omega\varepsilon_2}\;\;\frac{1}{(\nu_2^g d)}\;\;\tanh(\nu_2^g\, d/2) \qquad \text{(IV-6)}$$

For the length confinement coupling of this section one would expect continuity of the direction of the y-component of the magnetic field across the gap as indicated on the upper part of Fig. (IV-2(a)). Since the x-directed component of the electric field in region A becomes the z-component in region B (the gap) this component is in opposite directions at $z = -\frac{d}{2}$ and $z = +\frac{d}{2}$. The z-directed component of the electric field in region A can be calculated from these two components since ($i\omega\varepsilon_2 E_z = \frac{\partial H_y}{\partial x} = -\gamma_2 H_y$) where $\gamma_2 = \sqrt{k^2-\omega^2\mu\varepsilon_2}$ and k is the open surface propagation vector. This component one would expect to be continuous across the gap.

With respect to the relatively long wave-length surface plasmon wave in regions A there is a step discontinuity in the x-component of the electric field due to the anti-symmetry. Thus one can express the field as

$$E_x(z) = E_0\,\mathrm{sgn}(z)(e^{-ik|z|}) \qquad \text{(IV-7)}$$

Similarly

$$E_z(z) = E_0(\tfrac{\gamma_2}{ik})(e^{-ik|z|}) \qquad \text{(IV-8)}$$

where it has been assumed by symmetry that

$$E_x(z=0) \approx 0 \qquad \text{(IV-9)}$$

---

* A more general expression for the characteristic impedance valid for all d is given by

$$Z_g = \frac{k_g}{b\omega\varepsilon_2\nu_2^g}\,(\tanh(\nu_2^g d/2) + (\nu_2^g d/2)\,(\cosh^2(\nu_2^g d/2)))$$

Similarly one can have plasma waves propagating in the opposite direction. The relative amplitudes of the two are determined by the boundary conditions at the ends as well as the power radiated. The boundary condition in general results in an end coupling condition which can provide strong radiation. We have not studied this aspect of the problem, but plan to do so in the future since it appears promising and also is more broadband than length confinement coupling considered here. The latter ideally does not allow end radiation since one generally assumes an ideally terminated open or short circuit. All of the power propagating toward the end is thus reflected back. This is accomplished by terminating the antenna in an abrupt change in metal thickness (Fig. IV-3). The impedance change, even in the visible, is large enough to establish nearly 100% reflection. E_x is forced to become very small far away from the plasma resonance for b) the short circuit termination. For a) the "open circuit termination" $E_z(z=\pm L) \sim 0$, although this can provide strong coupling from the end of the metallic antenna.

Using zero end conditions for $E_z(z)$ gives for the fields outside of the metal

$$E_x(z) = E_0 \, \mathrm{sgn}(z)[e^{-ik|z|}+e^{-ik(2L-|z|)}] \, e^{-\gamma_0 x} \qquad \text{(IV-10a)}$$

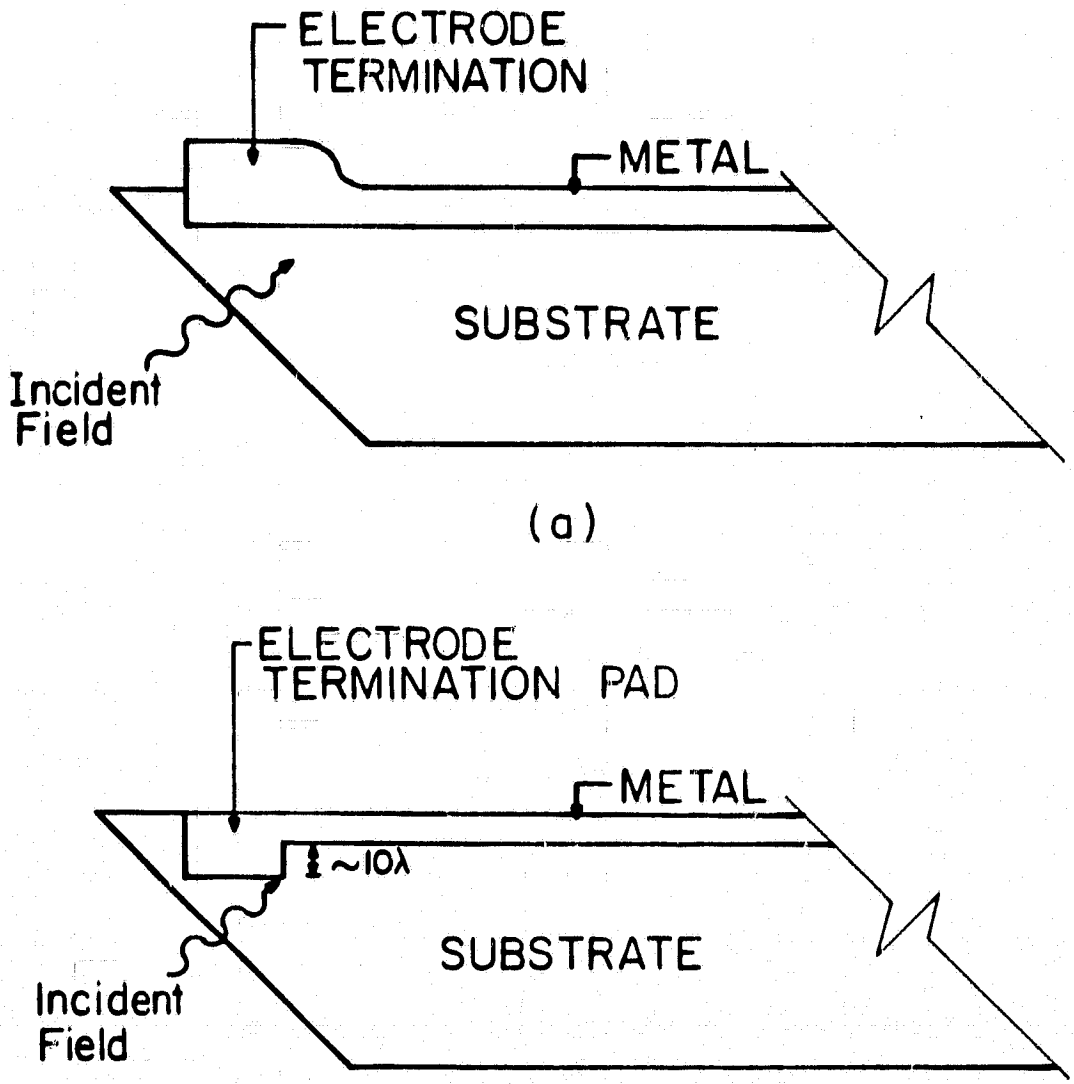
$$H_y(z) = \frac{\omega\varepsilon_0}{k} E_0[e^{-ik|z|}-e^{-ik(2L-|z|)}] \, e^{-\gamma_0 x} \qquad \text{(IV-10b)}$$

$$E_z(z) = \frac{+\gamma_0}{ik} E_0[e^{-ik|z|}-e^{-ik(2L-|z|)}] \, e^{-\gamma_0 x} \qquad \text{(IV-10c)}$$

The fields inside are the same with E_0 replaced by $\frac{\varepsilon_0}{\varepsilon_1}E_0$ in $E_x(z)$ and γ_0 replaced by $-\gamma_1$ to guarantee continuity of H_y and $E_z(z)$ and provide a jump in E_x.

By virtue of the termination one expects the field pattern given by Eq. (10a,b,c) to radiate. To deduce this radiation pattern one can obtain vector potential $\bar{A}$ far from the antenna by replacing the surface mode with a complex curr The current is given by $i\omega(\varepsilon-\varepsilon_0)E$ for the fields shown. Since one expects $E_z(z)$ to be the influential term, we consider only this component. Thus

$$J_z(z) = 0 \qquad x > 0$$



IV-3 Ideal antenna terminations. a) Open circuit b) Short circuit

$$= \frac{+\omega\gamma_0}{k}(\varepsilon_1-\varepsilon_0)E_0[e^{-ik|z|}-e^{-ik(2L-|z|)}]e^{-\gamma_1 x}$$

$$x < 0 \qquad \text{(IV-11)}$$

for the respective current densities in the two media.

The radiation pattern is obtained by solving for the vector potential in the Lorentz guage so that only currents determine A and hence the magnetic field. One observes that the above relationship gives a long wire antenna pattern at low frequencies. Assuming a width L_y and that the decay lengths γ_0 and γ_1 are much less than a wavelength, one can obtain an ideal terminated line current source

$$I_z(z) = \int_{-\infty}^{+\infty} J_z(z)dxL_y = \frac{\omega L_y}{k}\left[\frac{\gamma_0}{\gamma_1}(\varepsilon_1-\varepsilon_0)E_0\right][e^{-ik|z|}-e^{-ik(2L-|z|)}] \text{ Amps.}$$

$$= I_0 e^{-ikL} i \sin k(L-|z|) \qquad \text{(IV-12)}$$

where I_0 has been defined by

$$I_0 = \frac{2\omega L_y}{k}\frac{\gamma_0}{\gamma_1}(\varepsilon_1-\varepsilon_2)E_0 \qquad \text{(IV-13)}$$

In the low frequency limit this reduces properly since $\varepsilon_1 = -j\frac{\sigma}{\omega}$, $\nu_1 \cong \sqrt{j}/\delta$, $k \cong \omega/c$ and $\frac{\gamma_0}{\gamma_1} \cong -j\frac{\varepsilon_0\omega}{\sigma}$ from the single surface dispersion relations. One then finds $I_0 \cong -2L_y\sigma\delta \quad \frac{1}{\sqrt{j}}$ (Amplitude of the z-component of the field $(\frac{\gamma_0}{jk}E_0)$), which one would expect. Here δ is the classical skin depth.

From the vector potential one can calculate both the radiation fields and the total radiated power. The electric field in the θ direction is given by

$$E_\theta(\theta,r,\phi) = \sqrt{\frac{\mu}{\varepsilon}}\frac{I_0}{r}e^{-ikL}\sin^2\theta\frac{(\cos(kL)-\cos(k_0L\cos\theta))(\frac{k_0}{k})}{(1-(\frac{k_0}{k})^2\cos^2\theta)} \qquad \text{(IV-14)}$$

and the radiated power per unit solid angle by

$$\left(\frac{dP_R}{d\Omega}\right) = \frac{n}{8\pi^2} |I_0|^2 \left|\frac{k_0}{k}\right|^2 \exp(-2\alpha L)$$

$$\left|\frac{\sin\theta}{1-(k_0/k)^2\cos^2\theta}[\cos(k_0 L\cos\theta) - \cos(kL)]\right|^2 \qquad \text{(IV-15)}$$

The two extreme values of (k_0/k) provide two limiting cases for the coupling characteristics. For low frequencies $(k_0/k) \cong 1$ and $\alpha\ell \ll 1$. The radiative pattern reduces to:

$$\left(\frac{dP_R}{d\Omega}\right) = \frac{n}{8\pi^2} I_0^2 \left|\frac{\cos(k_0 L\cos\theta) - \cos(k_0 L)}{\sin\theta}\right|^2 \qquad \text{(IV-16)}$$

This is as expected is identical to that of a standing wave long wire antenna, and is shown in Fig. IV-4. The more relevant case for infrared and optical wavelengths at room temperature is the high frequency limit for which $(k_0/k) < 1$ and $\alpha L \gg 1$. Here the radiation pattern as illustrated in Fig. IV-4, is given by:

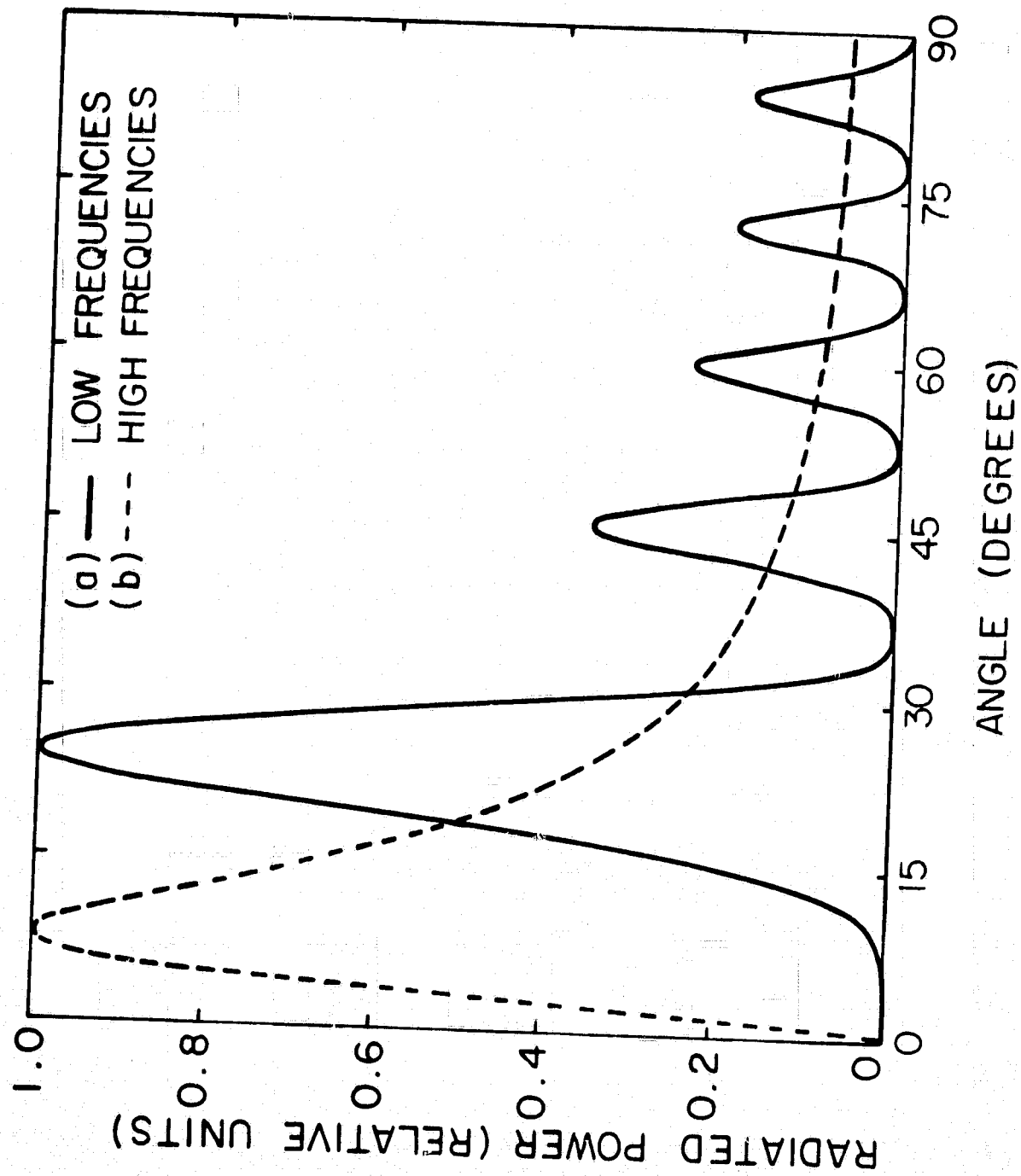
$$\left(\frac{dP_R}{d\Omega}\right) \cong \frac{n}{32\pi^2} I_0^2 \left|\frac{k_0}{k}\right|^2 \left|\frac{\sin\theta}{1-(k_0/k)^2\cos^2\theta}\right|^2 e^{-2\alpha L} \qquad \text{(IV-17)}$$

The most apparent difference from the low frequency pattern is the lack of lobe structure due to the short propagation length of plasmons at high frequencies. This also removes the dependence on L of the radiation pattern, If $1/\beta < L < 1/\alpha$, lobe structure is expected to be observed, but this is in general different from that given in Eq. (IV-16) since $\beta > k_0$.

The coupling of the plasmon surface mode to a laser cavity can be calculated by a modal expansion using the laser cavity modes. To do this consider the electric field of the Gaussian mode in the far field ($|r| \gg z_0 = \frac{\pi\omega_0^2 n}{\lambda}$). The far field of the lowest order mode, E_{00} is given by

$$E_{00} = \frac{1}{\sqrt{2\pi}\omega(z)} \exp\left[\frac{-(x^2+y^2)}{\omega^2(z)} - j\bar{k}\cdot\bar{r} + j\omega t\right] \qquad \text{(IV-18)}$$

where the normalization is such that $\int |E_{00}|^2 dx'dy' = 1$ and $\omega^2(z) = \omega_0^2(1+(\frac{z}{z_0})^2)$



IV-4 Radiation patterns at low and high frequencies for the total radiated power.

and $z_0 = \pi\omega_0^2 n/\lambda$. Similar expression are valid for the higher order modes E_{mn}.

We approach the problem from the point of view of a transmitting antenna exciting the laser cavity as illustrated in Fig. (IV-5). One thus expands the electric field excited in the laser cavity by the antenna in terms of these orthonormal modes of the laser cavity. This is of the form

$$E = \sum a_{mn} E_{mn} \tag{IV-19}$$

where we are particularly interested in a_{00}, the coupling to the lowest order mode. As long as this is done in the far-field of both the laser cavity modes and the antenna radiation field, the coefficients a_{mn} are easily deduced. This is assumed to be the x'-y' plane of Fig. (IV-5). The laser field is assumed polarized in the plane of the antenna. The coupling coefficient of interest a_{00} is then given by

$$a_{00} = \int_{-\infty}^{+\infty} dx'dy' E_{00}^* E_\theta(x',y') \tag{IV-20}$$

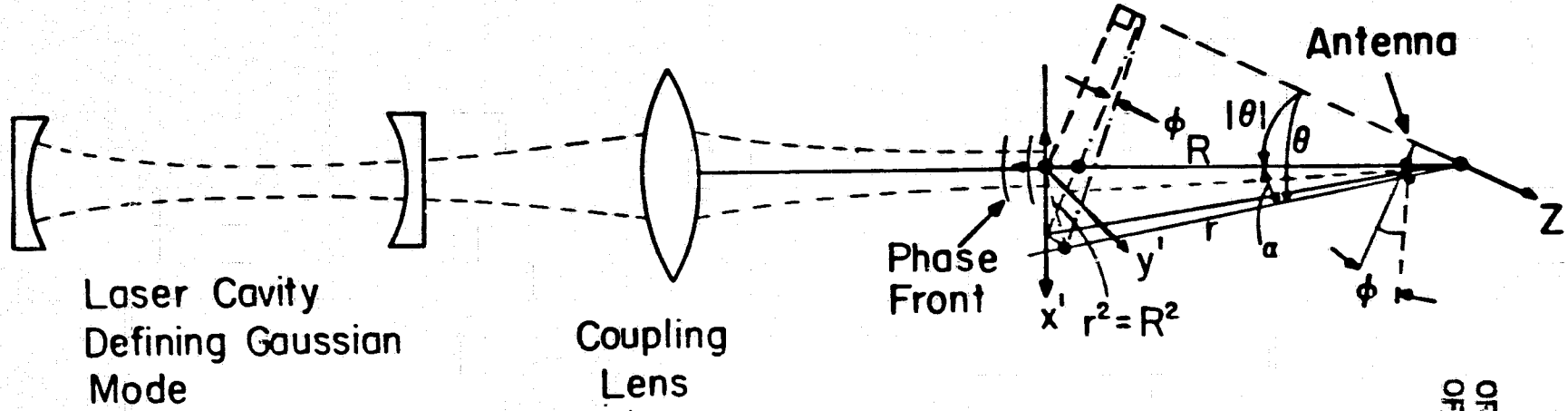
where $E_\theta(x',y')$ is the electric field component of the radiation field of the antenna.

This integral can be easily carried out when the area of the laser is much smaller than the lobe size of the antenna and assuming that the coupling lens is such that the confocal spot is at the antenna as shown. Then $r \cong R$, $\theta \cong |0|$ and $\phi \cong 0$ and one can separate a_{00} into a product of an integral over the laser field and the field of the antenna. Thus

$$a_{00} = \left[\int_{-\infty}^{+\infty} \frac{dx'dy'}{\omega(z)\omega(z)} \quad \frac{1}{(2\pi)^{\frac{1}{2}}}\, \omega(z)\, e^{-((x')^2+(y')^2)/\omega^2(z)}\right] E_\theta(R,|0|,\phi)$$

$$= \frac{1}{(2\pi)^{\frac{1}{2}}} \quad \omega(z)\, 2\sqrt{\pi}\, E_\theta(R,|\theta|,\phi) \tag{IV-21}$$

In the far field $\omega(z) \cong \frac{\omega_0 R}{z_0} = \frac{R\lambda}{\pi\omega_0 n}$ where n is the refractive index and ω_0 is the beam radius at the antenna. Thus



IV-5 Coupling of a Gaussian beam to a long wire antenna lobe. The antenna is oriented along the z-axis at R=0 and coincides with the focal region of the Gaussian beam.

$$a_{00} = \left(\frac{4Rc}{\sqrt{2}\omega_0\omega\eta}\right)E_\theta(R,|\theta|,\Phi) \qquad \text{(IV-22)}$$

The power coupling to the lowest order mode of the laser cavity is thus

$$P_c = \tfrac{1}{2}|a_{00}|^2\sqrt{\frac{\mu}{\varepsilon}}$$

$$P_c = |b_{00}|^2\left(\frac{dP_R}{d\Omega}\right) \qquad \text{(IV-23)}$$

where $\frac{dP_R}{d\Omega}$ is given by Eq. (IV-14) and $|b_{00}|^2 = \frac{8c^2}{\omega_0^2\omega^2\eta^2}$.

As an illustration of this, the low frequency short dipole of length L has $\left(\frac{dP_R}{d\Omega}\right) = \frac{I_0^2}{(4\pi\varepsilon)^2}L^2\frac{\sin^2\theta}{c^2}\eta^2\omega^2\sqrt{\frac{\varepsilon}{\mu}}$ where I_0 is the current amplitude. Thus

$$P_c = \left(\sqrt{\frac{\mu}{\varepsilon}}\ \frac{2}{\pi\omega_0^2}\ \frac{L^2}{4\pi}\sin^2\theta\sqrt{\frac{\mu}{\varepsilon}}\right)\frac{I_0^2}{2} \qquad \text{(IV-24)}$$

Since the radiation resistance of such a short dipole is $R = \sqrt{\frac{\mu}{\varepsilon}}\left(\frac{\omega}{c}\right)^2\frac{L^2}{6\pi}$, for a conjugate match of the antenna the power from the generator is $(I_0^2R/2) = P_A$. Thus from the above

$$P_c = \left(\frac{\left(\frac{\lambda}{2\pi}\right)^2\ 3\sin^2\theta}{S}\right)P_A \qquad \text{(IV-25)}$$

where S is the confocal area of the laser at the antenna $\pi\omega_0^2$. By integrating over all angles $\left(\int_0^\pi \sin^3\theta\, d\theta = \frac{2}{3}\right)$ one obtains the usual result for the effective area for the short dipole antenna.

To complete the calculation of the coupling of the gap mode to the laser field the value of E_0 must be expressed in terms of the gap field E_i by considering the coupling between the open surface mode and the gap mode. Once this relationship is established one can apply the principle of reciprocity which demands that the coupling be equal for receiving and transmitting. Ideally one would desire one hundred percent coupling of the power from the open surface to the gap. One has Power = $\mathrm{Re}\int\tfrac{1}{2}E_xH_y^*dx$ for the gap mode propagation in the x-direction. For the components E_x =

$\frac{k}{\omega\varepsilon_2} H_y$ for the single surface mode and $E_z = \frac{k_g}{\omega\varepsilon_2} H_y$ for the gap mode, thus the one hundred percent coupling constraint implies

$$\frac{k}{\omega\varepsilon_2} \int_{\substack{\text{Single}\\ \text{Surface}}} |H_y|^2 dx = \frac{k_g}{\omega\varepsilon_2} \int_{\text{Gap}} |H_y|^2 dz \qquad \text{(IV-26)}$$

One also has the relations $H_y = \frac{\omega\varepsilon_2}{k_g} E_i$ for the gap (Eq. IV-4) and $\frac{\gamma_2}{ik} E_0 = \frac{\gamma_2}{i\omega\varepsilon_2} H_y$ for the single surface mode.

Since the power flow is primarily outside the negative dielectric regions, these integrals give for the field amplitude constraints

$$\frac{\omega\varepsilon_2}{2k\gamma_2} |E_0|^2 = \frac{k}{2\gamma_2\omega\varepsilon_2} |H_0|^2 = \frac{\omega\varepsilon_2 d}{k_g} |E_i|^2 = \frac{k_g d}{\omega\varepsilon_2} |H_i|^2 \qquad \text{(IV-27)}$$

Thus H_i in the gap can be expressed in terms of E_0.

Such a coupling from the gap to the antenna requires a taper design which can in principle be accomplished for any antenna length. We are continuing to investigate this possibility theoretically and experimentally.

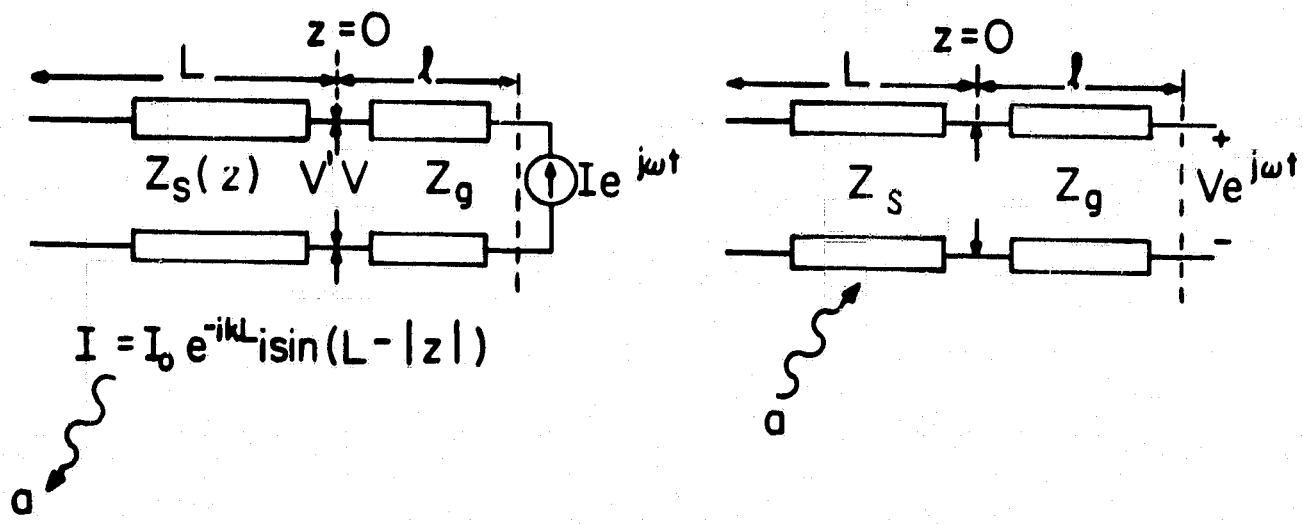
If it is assumed that Kirchoffs Laws are valid for the termination (no significant radiation at the junction antenna boundary), then the junction coupling problem reduces to a transmission line analysis (Fig. (IV-6)). Letting $Z_s(0)$ be the impedance of the single surface mode at $z = 0$, then the relationship between a current source $I_s e^{j\omega t}$ applied at the gap and I_0 is

$$I_s = jI_0 e^{-jkL} \sin kL\left(1+\frac{1}{Z_s(0)} Z_g(0)\right) \qquad \text{(IV-28)}$$

where Z_g is the impedance (assumed lumped) of the gap. Thus I_0 can be expressed in terms of the driving current amplitude I_s associated with the gap through this relationship.

The impedance of the gap $Z_g(0)$ is given by $Z_g/(i\tan(k_g\ell))$ which for a short gap is $Z_g/ik_g\ell$. Thus using Eq. (IV-6) one obtains $\frac{d}{b\ell} \frac{1}{i\omega\varepsilon_2} = \frac{1}{i\omega C_g}$ where C_g is the gap capacitance. $Z_s(0)$ the impedance of the antenna is similarly given by

$$Z_s(0) = \frac{k}{\omega\varepsilon_2} \frac{1}{b\gamma_0} \frac{1}{i\tan(k\ell)} \qquad \text{(IV-29)}$$



IV-6 Illustration of reciprocity for evaluating the receiving characteristics of a laser illuminated antenna.

Using Eqs. (14,22,28) one can thus write

$$a_{00} = ZI_s \tag{IV-30}$$

where Z the coupling constant is given by

$$Z = \sqrt{\frac{\mu}{\varepsilon}} \frac{e^{jkL}}{j\sin(kL)(1+\frac{Z_s(0)}{Z_g(0)})} \frac{4}{\sqrt{2\omega_0}} \frac{1}{k} f(0,L) \tag{IV-31}$$

where $f(\theta,L)$ is $\sin^2\theta(\frac{\cos kL-\cos(k_0 L\cos\theta)}{1-(k_0/k)^2\cos^2\theta})$ for the case of transmission line coupling to the gap mode. As a consequence of reciprocity, one thus has

$$V_i = (Z)\sqrt{\frac{\varepsilon}{\mu}}\, a_{00} \tag{IV-32}$$

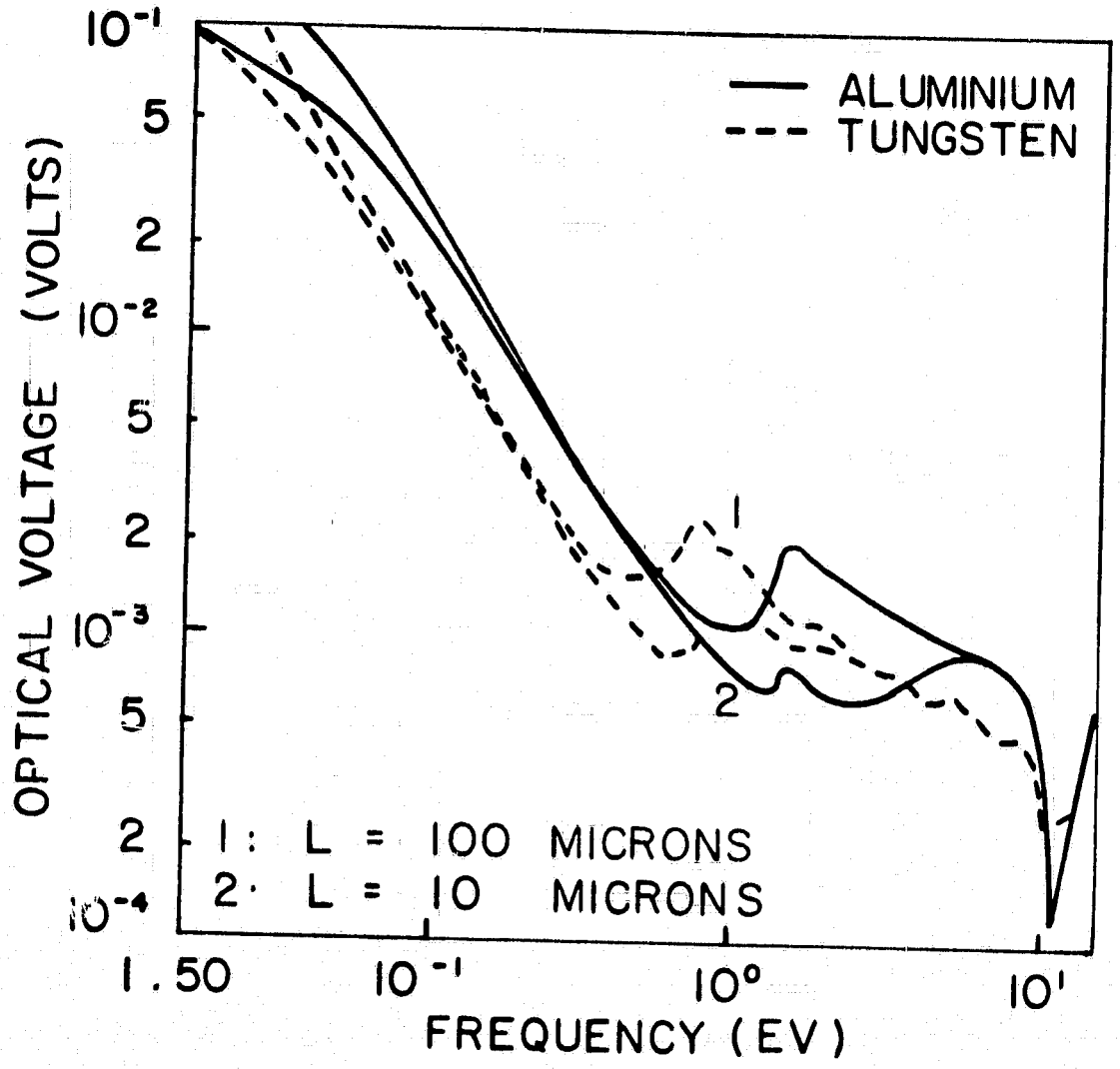
as the voltage induced across the gap by the laser field, since $\sqrt{\frac{\varepsilon}{\mu}}\, a_{00}$ is the magnitude of the magnetic field within the laser cavity and also because of normalization corresponds to current. Knowing the value of the voltage-current characteristic of the junction allows one to thus calculate the rectified signal induced due to nonlinearity.

The above basic approach shows that antenna theory is easily extended to higher frequencies. These results form the basis for our numerical estimates of the signals available from the diagnostic heterodyne experiment considered in a latter section. In the following section we also verify that for the proper taper one can expect nearly one hundred percent coupling between the open surface mode and the gap-mode.

Before leaving this section we briefly consider the gap mode in the presence of electron tunneling and some initial calculations of induced optical voltage using a more conventional antenna approach. For typical junction areas and distributed tunneling conductance $G_J(\Omega^{-1}/m^2)$ the gap mode is only slightly perturbed by the tunneling. The potential amplitude induced across the gap by the incident field can be obtained by considering power flow through the junction for which the relative phases of the waveforms need not be considered. For the ideal structure of Fig. IV-1a assuming for simplicity symmetrical excitation at the edges ($x=\pm a$) a standing wave gap mode is excited. The distributed tunneling conductance can then be replaced by an equivalent lumped

conductance ($G_{eff} = 2b\int_{-a}^{a} G_J e^{-2k_g''(a-|x|)} dx$) in parallel with the impedance of the gap mode as seen from $x = \pm a$. By symmetry the gap mode current must have a null at the center (x=0). The equivalent impedance of the gap observed at the edge is then $Z = Z_g/i \tan(k_g a)$. This reduces to that of a lumped capacitance of the same dimensions at low frequencies since $\tan(k_g a) \approx k_g a$ and $Z \sim 1/i\omega(2\varepsilon_0 ab/d)$ (using Eqs. (VI-4) and (VI-5)), but approaches the characteristic impedance of the anti-symmetric mode at higher frequencies ($\tan(k_g a) \to -i$).

The equivalent circuit consisting of a voltage source with an internal impedance driving the above combined gap mode impedance has been used to calculate the optical voltage excited by an external radiation field. Using conventional antenna theory, the voltage source driving the antenna has an amplitude $2E_i \ell_{eff}$, where E_i is the incident field amplitude $a_{00}/(\sqrt{2\pi}\omega_0)$ and $\ell_{eff} = \sqrt{2\pi} f(\theta,L) \frac{e^{jkL}}{jk \sin kL}$ is the effective antenna length which for a short antenna in the maximum gain direction is given by the shorter of the actual length L or the plasmon propagation length $1/\alpha$. The internal impedance given by the series combination of the radiative, dissipative, and reactive impedances is included in $Z_s(0)$. The results of such a calculation carried out by Siu for the optical voltage across an effective tunneling resistance ($1/G_{eff}$) of 500 ohms in both Aℓ and W structures is illustrated in Fig. IV-7. For these computations the experimental frequency dependence of the dielectric coefficients has been included. Since the dissipative and reactive power in the metal become significant, the optical voltage decreases as the frequency increases in the region where the plasmon propagation length $1/\alpha$ is much longer than the actual length. As the plasmon propagation decreases below the actual length, the dissipative and reactive impedances decrease, and since the effective antenna length also decreases accordingly, the optical voltage is approximately constant. An interesting phenomenon is observed at those frequencies where the interband transitions are strongest. Dus to better impedance matching from the increased reflectivity at the gap, the optical voltage at the surface resonance in Aℓ, since propagation of surface plasmons is inhibited. At frequencies above the plasma resonance these structures behave as dielectric antennas, and the optical voltage developed can be significant.



IV-7 Theoretically predicted frequency dependence of the optically induced voltage.

The decrease in optical voltage as the frequency increases from the infrared to the visible is much less dramatic than that predicted by conventional antenna theory. The ratio of the optical voltages at 10.6 μm and at 6328 Å is approximately 8, both for Aℓ and W, whereas the ratio of the respective wavelengths is approximately 16. It is found that in the relaxation region the reactive power in the metal becomes more important than the dissipative power. This indicates that by properly matching or tuning, the optical voltage can be substantially enhanced.

More extensive investigations of the behaviour of the optical voltage or power coupled into the junction at visible frequencies are presently being pursued on the basis of Eqs. (IV-31) and (IV-32) to calculate V_j directly in terms of the laser field. The important implications of an extension of antenna theory through the use of surface electromagnetic waves are, however, apparent.

IV-3 Modes Along a Corrugated Interface - Grating Coupling Basic Relations

To couple energy to and from a surface plasmon mode at a metal-dielectric interface, one might consider the use of a corrugated interface region. Coupling between the bound mode and free space modes is achieved through Bragg scattering from the periodic structure. In this section, the equations describing the field amplitudes along a sinusoidally corrugated interface will be derived, and will be solved to first order in the corrugation amplitude over the periodicity. Consider first a flat interface, where the dielectric constant is ε_1 for $X > 0$, ε_2 for $X < 0$. If an incident field is to be totally transmitted as $\nabla E = 0$ in the two regions, the fields can be expressed in the form for p-polarized waves

$$E = E_0(\hat{X}\sin\theta_1 - \hat{Z}\cos\theta_1)\, e^{ik_0\sqrt{\varepsilon_1}\sin\theta_1 Z}\, e^{ik_0\sqrt{\varepsilon_1}\cos\theta_1 X} \qquad X > 0$$

$$H = \frac{k_0\sqrt{\varepsilon_1}}{\omega} E_0 \hat{Y}\, e^{ik_0\sqrt{\varepsilon_1}\sin\theta_1 Z}\, e^{ik_0\sqrt{\varepsilon_1}\cos\theta_1 X} \qquad \text{(IV.3.1)}$$

$$E' = E_0'(\hat{X}\sin\theta_2+\hat{Z}\cos\theta_2)\, e^{ik_0\sqrt{\varepsilon_2}\sin\theta_2 Z}\, e^{-ik_0\sqrt{\varepsilon_2}\cos\theta_2 X}$$

$$X < 0$$

$$H' = \frac{k_0\sqrt{\varepsilon_2}}{\omega} E_0'\hat{Y}\, e^{ik_0\sqrt{\varepsilon_2}\sin\theta_2 Z}\, e^{-ik_0\sqrt{\varepsilon_2}\cos\theta_2 X} \qquad \text{(IV.3.2)}$$

Using Snell's law $\sqrt{\varepsilon_1}\sin\theta_1 = \sqrt{\varepsilon_2}\sin\theta_2$ and the boundary conditions (continuity of the fields in the Y-Z plane) one obtains the relations

$$\sqrt{\varepsilon_1}E_0 = \sqrt{\varepsilon_2}E_0' \quad , \quad -\cos\theta_1 E_0 = \cos\theta_2 E_0' \quad ,$$

hence $\sqrt{\frac{\varepsilon_2}{\varepsilon_1}} = \frac{\cos\theta_2}{\cos\theta_1}$. Again using Snell's law, one obtains the relation

$$\frac{\varepsilon_2}{\varepsilon_1} = \frac{1-\frac{\varepsilon_1}{\varepsilon_2}\sin^2\theta_1}{1-\sin^2\theta_1} \quad , \quad \text{or } \sin^2\theta_1 = \frac{\varepsilon_2}{\varepsilon_2+\varepsilon_1} \quad . \qquad \text{(IV.3.3)}$$

This, as expected, gives the Brewster angle constraint for ε_2 and $\varepsilon_1 > 0$. If $\varepsilon_2 < 0$ then $\sin^2\theta_1$ is positive but larger than 1 so $\cos\theta$ is imaginery

If the surface is periodically corrugated, given by $X = F(Z)$ where $F(Z) = F(Z+\frac{ng}{k_0})$ for any integer n, the Bloch-Floquet theorem implies that the fields can be expressed as summations over absorbed grating vectors, i.e., the grating k-vector given by $k_g = k_0 g$

$$E = \frac{1}{\sqrt{\varepsilon_1}} \sum_n E_n(\hat{X}(\sqrt{\varepsilon_1}\sin\theta_1+ng) - \hat{Z}[\varepsilon_1-(\sqrt{\varepsilon_1}\sin\theta_1+ng)^2]^{\frac{1}{2}}$$
$$X\, e^{ik_0(\sqrt{\varepsilon_1}\sin\theta_1+ng)Z}\, e^{ik_0[\varepsilon_1-(\sqrt{\varepsilon_1}\sin\theta_1+ng)^2]^{\frac{1}{2}}X}$$

$$X > 0$$

$$H = \hat{Y}\frac{k_0\sqrt{\varepsilon_1}}{\omega} \sum_n E_n\, e^{ik_0(\sqrt{\varepsilon_1}\sin\theta_1+ng)Z}\, e^{ik_0[\varepsilon_1-(\sqrt{\varepsilon_1}\sin\theta_1+ng)^2]^{\frac{1}{2}}X}$$

$$E' = \frac{1}{\sqrt{\varepsilon_2}} \sum_n E_n'(\hat{X}(\sqrt{\varepsilon_2}\sin\theta_2+ng) + \hat{Z}[\varepsilon_2-(\sqrt{\varepsilon_2}\sin\theta_2+ng)^2]^{\frac{1}{2}})$$
$$e^{ik_0(\sqrt{\varepsilon_2}\sin\theta_2+ng)Z}\, e^{-ik_0[\varepsilon_2-(\sqrt{\varepsilon_2}\sin\theta_2+ng)^2]^{\frac{1}{2}}X}$$

$$H' = \hat{Y}\frac{k_0\sqrt{\varepsilon_2}}{\omega} \sum_n E_n'\, e^{ik_0(\sqrt{\varepsilon_2}\sin\theta_2+ng)Z}\, e^{-ik_0[\varepsilon_2-(\sqrt{\varepsilon_2}\sin\theta_2+ng)^2]^{\frac{1}{2}}X}$$

(IV.3.4)

Here if n = 0 pertains to the surface plasmon mode n = -1 provides a radiated component into dielectric ε_1 and hence E_{-1}/E_0 is the coupling coefficient of interest.

To proceed we apply the boundary conditions and use Snell's Law

Define $(\sqrt{\varepsilon_1}\sin\theta_1+ng) = \sqrt{\varepsilon_1}\sin\theta_{1n} = \sqrt{\varepsilon_2}\sin\theta_{2n}$ (Snells' law),
$[\varepsilon_1-(\sqrt{\varepsilon_1}\sin\theta_1+ng)^2]^{\frac{1}{2}} = \sqrt{\varepsilon_1}\cos\theta_{1n}$, and $[\varepsilon_2-(\sqrt{\varepsilon_2}\sin\theta_2+ng)^2]^{\frac{1}{2}} = \sqrt{\varepsilon_2}\cos\theta_{2n}$.

If $\theta(z) = \frac{\partial}{\partial z} F(z)$, the boundary conditions become

$$E_t: \sum_n E_n\, e^{ink_g z}\, e^{ik_0\sqrt{\varepsilon_1}\cos\theta_{1n}F(z)}\, (\sin\theta_{1n}\sin\theta(z)-\cos\theta_{1n}\cos\theta(z))$$

$$= \sum_n E_n'\, e^{ink_g z}\, e^{-ik_0\sqrt{\varepsilon_2}\cos\theta_{2n}F(z)}\, (\sin\theta_{2n}\sin\theta(z)+\cos\theta_{2n}\cos\theta(z)) ,$$

$$H_t: \sqrt{\varepsilon_1}\sum_n E_n\, e^{ink_g z}\, e^{ik_0\sqrt{\varepsilon_1}\cos\ _{1n}F(z)} = \sqrt{\varepsilon_2}\sum_n E_n'\, e^{ink_g z}\, e^{-ik_0\sqrt{\varepsilon_2}\cos\theta_{2n}F(z} \qquad \text{(IV.3.5)}$$

Specializing now to a sinusoidal corrugation $F(z) = \frac{\varepsilon}{k_g}\sin k_g z$, then

$$\cos\theta(z) = \frac{1}{\sqrt{1+\varepsilon^2\cos^2 k_g z}}, \quad \sin\theta(z) = \frac{\varepsilon\cos k_g z}{\sqrt{1+\varepsilon^2\cos^2 k_g z}} \qquad \text{(IV.3.6)}$$

Substituting into the boundary condition expressions above and using a well known expansion one obtains

$e^{i\alpha\sin\theta} = \sum_m J_m(\alpha)\, e^{im\theta}$, one has

$$E_t: \sum_{n,m} E_n\, e^{i(n+m)k_g z}\, J_m\left(\frac{\varepsilon\sqrt{\varepsilon_1}k_0}{k_g}\cos\theta_{1n}\right)(\varepsilon\sin\theta_{1n}\cos k_g z-\cos\theta_{1n})$$

$$= \sum_{n,m} E_n'\, e^{i(n+m)k_g z}\, J_m\left(\frac{\varepsilon\sqrt{\varepsilon_2}k_0}{k_g}\cos\theta_{2n}\right)(\varepsilon\sin\theta_{2n}\cos k_g z+\cos\theta_{2n}) , \qquad \text{(IV.3.7)}$$

$$H_t:\ \sqrt{\varepsilon_1}\sum_{n,m} E_n\, e^{i(n+m)k_g z}\, J_m\left(\frac{\varepsilon\sqrt{\varepsilon_1}k_0}{k_g}\cos\theta_{1n}\right)$$

$$= \sqrt{\varepsilon_2}\sum_{n,m} E_n'\, e^{i(n+m)k_g z}\, J_m\left(\frac{\varepsilon\sqrt{\varepsilon_2}k_0}{k_g}\cos\theta_{2n}\right)$$

This can be expressed, using $\frac{\alpha}{2}(J_{m+1}(\alpha)+J_{m-1}(\alpha)) = nJ_m(\alpha)$, and associating wave numbers as

$$\sum_n E_n J_{N-n}\left(\frac{\varepsilon\sqrt{\varepsilon_1}k_0}{k_g}\cos\theta_{1n}\right)\left(\frac{(N-n)k_g}{\sqrt{\varepsilon_1}k_0\cos\theta_{1n}} - \cos\theta_{1n}\right)$$

$$= \sum_n E_n' J_{N-n}\left(\frac{\varepsilon\sqrt{\varepsilon_2}k_0}{k_g}\cos\theta_{2n}\right)\left(\frac{(N-n)k_g}{\sqrt{\varepsilon_2}k_0\cos\theta_{2n}} - \cos\theta_{2n}\right), \qquad \text{(IV.3.8)}$$

and

$$\sqrt{\varepsilon_1}\sum_n E_n J_{N-n}\left(\frac{\varepsilon\sqrt{\varepsilon_1}k_0}{k_g}\cos\theta_{1n}\right) = \sqrt{\varepsilon_2}\sum_n E_n' J_{N-n}\left(\frac{\varepsilon\sqrt{\varepsilon_2}k_0}{k_g}\cos\theta_{2n}\right),$$

which must be valid for any integer N.

The system above can be solved in the small ε limit. To first order in ε, E_0, E_0' are the same as in the flat interface case, and E_1, E_1', E_{-1}, E_{-1}', the only other nonvanishing components in this limit, can be determined. The small argument of the Bessel functions is used, i.e., $J_0(\varepsilon X) \to 1$, $J_{\pm1}(\varepsilon x) \to \frac{\pm\varepsilon X}{2}$, all others vanish to first order.

The equations for N=0 yield the flat interface relations between E_0, E_0' and hence yields $(\sin\theta_{10})^2 = \frac{\varepsilon_2}{\varepsilon_1+\varepsilon_2}$ (the terms involving $E_{\pm1}$, $E_{\pm1}'$ may be neglected as they are second order in ε). The equations for $N = \pm1$ are

$$N=1,\ E_t:\ E_1(-\cos\theta_{11}) + E_0\left(\frac{\varepsilon\sqrt{\varepsilon_1}k_0}{2k_g}\cos\theta_{10}\right)\left(\frac{k_g}{\sqrt{\varepsilon_1}k_0\cos\theta_{10}} - \cos\theta_{10}\right)$$

$$= E_1'(\cos\theta_{21}) + E_0'\left(\frac{\varepsilon\sqrt{\varepsilon_2}k_0}{2k_g}\cos\theta_{20}\right)\left(\frac{k_g}{\sqrt{\varepsilon_2}k_0\cos\theta_{20}} + \cos\theta_{20}\right),$$

$$N=-1,\ E_t:\ E_{-1}(-\cos\theta_{1-1}) + E_0\left(\frac{-\varepsilon\sqrt{\varepsilon_1}k_0}{2k_g}\cos\theta_{10}\right)\left(\frac{-k_g}{\sqrt{\varepsilon_1}k_0\cos\theta_{10}} - \cos\theta_{10}\right)$$

(IV.3.9)

C-2

$$= E'_{-1}(\cos\theta_{2-1}) + E'_0\left(\frac{-\varepsilon\sqrt{\varepsilon_2}k_0}{2k_g}\cos\theta_{20}\right)\left(\frac{-k_g}{\sqrt{\varepsilon_2}k_0\cos\theta_{20}} + \cos\theta_{20}\right)$$

$$N=1,\ H_t:\ \sqrt{\varepsilon_1}\left[E_1+E_0\left(\frac{\varepsilon\sqrt{\varepsilon_1}k_0}{2k_g}\cos\theta_{10}\right)\right] = \sqrt{\varepsilon_2}\left[E'_1+E'_0\left(\frac{\varepsilon\sqrt{\varepsilon_2}k_0}{2k_g}\cos\theta_{20}\right)\right],$$

$$N=-1,\ H_t:\ \sqrt{\varepsilon_1}\left[E_{-1}+E_0\left(\frac{-\varepsilon\sqrt{\varepsilon_1}k_0}{2k_g}\cos\theta_{10}\right)\right] = \sqrt{\varepsilon_2}\left[E'_{-1}+E_0\left(\frac{-\varepsilon\sqrt{\varepsilon_2}k_0}{2k_g}\cos\theta_{20}\right)\right].$$

The radiated component of the field is proportional to E_{-1}. The equations of interest can be expressed, with $E'_0 = \sqrt{\frac{\varepsilon_1}{\varepsilon_2}}\ E_0$,

$$-E_{-1}\cos\theta_{1-1} + \frac{\varepsilon}{2}E_0\left[1+\frac{\sqrt{\varepsilon_1}k_0}{k_g}\cos^2\theta_{10}-\sqrt{\frac{\varepsilon_1}{\varepsilon_2}}\left(1-\frac{\sqrt{\varepsilon_2}k_0}{k_g}\cos^2\theta_{20}\right)\right] = E'_{-1}\cos\theta_{2-1},$$

$$\sqrt{\varepsilon_1}E_{-1} - \frac{\varepsilon}{2}E_0\left[\frac{\varepsilon_1 k_0}{k_g}\cos\theta_{10}-\frac{\sqrt{\varepsilon_1\varepsilon_2}k_0}{k_g}\cos\theta_{20}\right] = \sqrt{\varepsilon_2}\,E'_{-1} \qquad \text{(IV.3.10)}$$

These can be combined to eliminate E'_{-1}, yielding

$$\frac{E_{-1}}{E_0} = \frac{\frac{\varepsilon}{2}\left[1+\frac{\sqrt{\varepsilon_1}k_0}{k_g}\cos^2\theta_{10}-\sqrt{\frac{\varepsilon_1}{\varepsilon_2}}\left(1-\frac{\sqrt{\varepsilon_2}k_0}{k_g}\cos^2\theta_{20}\right) + \frac{\cos\theta_{2-1}}{\sqrt{\varepsilon_2}}\frac{k_0}{k_g}\left(\varepsilon_1\cos\theta_{10}-\sqrt{\varepsilon_1\varepsilon_2}\cos\theta_{20}\right)\right]}{\cos\theta_{1-1}+\sqrt{\frac{\varepsilon_1}{\varepsilon_2}}\cos\theta_{2-1}} \qquad \text{(IV.3.11)}$$

Using $\sin\theta_{10} = \sqrt{\frac{\varepsilon_2}{\varepsilon_1+\varepsilon_2}}$, $\cos\theta_{10} = \sqrt{\frac{\varepsilon_2}{\varepsilon_1+\varepsilon_2}}$, $\frac{\cos\theta_{20}}{\cos\theta_{10}} = -\sqrt{\frac{\varepsilon_2}{\varepsilon_1}}$ etc., one finds

$$\frac{E_{-1}}{E_0} = \frac{\frac{\varepsilon}{2}\left[\sqrt{\varepsilon_2}-\sqrt{\varepsilon_1}+\frac{k_0}{k_g}\sqrt{\varepsilon_1\varepsilon_2}+\frac{k_0}{k_g}\left[1-\left(\sqrt{\frac{\varepsilon_1}{\varepsilon_1+\varepsilon_2}}-\frac{k_g}{k_g\sqrt{\varepsilon_2}}\right)^2\right]^{\frac{1}{2}}\left(\frac{\varepsilon_1\sqrt{\varepsilon_1}+\varepsilon_2\sqrt{\varepsilon_1}}{\sqrt{\varepsilon_1}+\varepsilon_2}\right)\right]}{\sqrt{\varepsilon_2}\left[1-\left(\sqrt{\frac{\varepsilon_2}{\varepsilon_1+\varepsilon_2}}-\frac{k_g}{k_0\sqrt{\varepsilon_1}}\right)^2\right]^{\frac{1}{2}}-\sqrt{\varepsilon_1}\left[1-\left(\sqrt{\frac{\varepsilon_1}{\varepsilon_1+\varepsilon_2}}-\frac{k_g}{k_0\sqrt{\varepsilon_2}}\right)^2\right]^{\frac{1}{2}}} \qquad \text{(IV.3.12)}$$

and θ_{1-1} and θ_{2+1} provide the angles for the emergence of the modes. When $|\varepsilon_2| >> |\varepsilon_1|$, this reduces to (for $\varepsilon_1 = 1$)

$$\frac{E_{-1}}{E_0} \cong \frac{\frac{\varepsilon}{2}\left[1+2\frac{k_0}{k_g}\sqrt{\varepsilon_1}\right]}{\frac{k_g}{\sqrt{\varepsilon_1}k_0}\left[2-\frac{k_g}{k_0\sqrt{\varepsilon_1}}\right]^{\frac{1}{2}}}, \text{ and if } \frac{k_g}{k_0} \ll 1, \qquad \text{(IV.3.13)}$$

i.e., $\frac{d}{\lambda_0} \ll 1$, one finds $\frac{E_{-1}}{E_0} \cong \frac{\varepsilon \cdot \varepsilon_1}{i2}\left(\frac{d}{\lambda_0}\right)^2 = \alpha c\ k_0$

The coupling for a small ε can be substantial due to the large magnitude of $\sqrt{\varepsilon_2}$. αk_0 is a simple approximation of the coupling coefficient deduced by Seymour et al[7], who have increased αk_0 by increasing ε_1.

The above results provide detailed coupling parameters. In particular at optical frequencies they indicate that 1% coupling can be achieved for a 300 Å height in the grooves with a spacing of the order of a wavelength.

The basic problem with this particular coupling scheme is the optimal design of the corrugation to prevent scattering into higher order modes. A sinusoidal grating is in particular not optimal. Seymour, Koteles, and Stegemen[7] have experimentally demonstrated a coupling efficiency of 8-9% using overcoated gratings to help confine the radiation field. This effort was at 118 μ and does demonstrate possibilities for visible frequencies.

The overall power efficiency for the coupling for the particular mode can be estimated from the theory of Tien and Uhlrich[14] for integrated optical coupling. The coupling efficiency is given by

$$\eta \cong 2\ell\alpha k_0\left(\frac{1-\exp[-(\beta''+\alpha k_0)\ell]}{(\beta''+\alpha k_0)\ell}\right)^2 \qquad \text{(IV.3.14)}$$

where β'' is the inverse of the plasmon decay length, and ℓ the grating length. At visible frequencies for which β'' is large η can be ≅ 100% for $\ell \gtrsim (\alpha k_0)^{-1}$ however the energy is absorbed. One must guarantee that $\ell \ll \beta''^{-1}$ and $\alpha k_0 \gg \beta''$ for efficient loss free coupling. This is difficult to achieve at room temperature; however, not impossible with careful design.

IV-4 Taper Coupling

An alternative solution to the coupling of radiation to tunnel junctions is the utilization of tapers in analogy to microwave coupling[15]. Due to the slow wave nature of the surface plasmon mode and the abrupt ending of conventional tunnel junctions, a large mismatch in wave impedence exists between the junction gap mode and the surrounding media. This implies that little light can be transmitted or received. By gradually taping the ends of the device a wave impedence transformer is made whereby the gap mode becomes radiative.

Letting the height of the gap, x_0 be sufficiently small such that only the surface plasmon forward and backward modes exist in the junction (a good approximation, as $x_0 \cong 10$ Å) a coupled mode analysis[16] can be performed to obtain the coupling coefficient, κ, between these modes. If κ is high, as in the conventional junction, most of the forward wave will be converted to the backward wave (high reflection) and not much light will enter or leave the device. If κ is low, coupling between the junction and surrounding media is high.

The transverse electric and magnetic fields of the forward and backward waves, E_a, H_a and E_b, H_b respectively, may be represented by,

$$E_a = A(z)\bar{e}_a(x,z)$$
$$H_a = A(z)\bar{h}_a(x,z) \qquad \text{(IV.4.1)}$$

and

$$E_b = B(z)\bar{e}_b(x,z)$$
$$H_b = B(z)\bar{h}_b(x,z) \qquad \text{(IV.4.2)}$$

where A is the forward modal amplitude coefficient, and $\bar{e}_a$, $\bar{h}_a$ are the transverse vector modal functions. B, $\bar{e}_b$, $\bar{h}_b$ have corresponding definitions for the backward wave. Quantities e, h, are normalized using

$$\int \hat{z}\cdot\bar{e}_n \times \bar{h}_m^* ds = \delta_{nm} \qquad \text{(IV.4.3)}$$

m,n = a,b and S is the cross-sectional area of the junction.

Denoting β as the modal propagation constant, the coupled mode differential equations [17] governing A and B are,

$$\frac{dA}{dz} - i\beta A = ik_{ba}B \qquad \text{(IV.4.4)}$$

$$\frac{dB}{dz} + i\beta B = ik_{ab}A \qquad \text{(IV.4.5)}$$

where k_{ab} is the coupling coefficient from A to B, and k_{ba} the same from B to A. By assumption, only two modes exist in the junction, therefore their combined power is conserved. Consequently, $k_{ab} = k_{ba} = k$.

In tapered waveguides the guide medium changes as a function of propagation direction. The varying boundary conditions allow the normally non-interacting (different spatial frequency components) forward and backward modes to couple. As a result, β, $\bar{e}$, $\bar{h}$, all become functions of the propagation direction, z. The coupling coefficient is shown to be given by[16]

$$k_{ba} = \int_S \hat{z} \cdot \bar{e}_n \times \frac{\delta \bar{h}_m^*}{\delta z}\, ds = k_{ab} \qquad \text{(IV.4.6)}$$

The field solutions for the antisymmetric plasmon mode given in Appendix II are good approximations for the vector modal functions. The limitations are imposed by the boundary conditions taking into account the variation in $x_0(z)$ with z. Considering continuity of tangential $\bar{H}$ and $\bar{E}$,

$$[H_z]_{x_0^-}^{x_0^+} = 0 \qquad \text{(IV.4.7)}$$

$$[E_x \frac{dx_0}{dz} + \frac{i}{\omega\varepsilon}\frac{\delta H_y}{\delta x}]_{x_0^-}^{x_0^+} \qquad \text{(IV.4.8)}$$

Eqn. 7 is as before. Eqn. 8 contains an additional term due to the sloping boundary x_0. If this term is small compared with the second, i.e., if $E_x \frac{dx_0}{dz} << \frac{1}{\omega\varepsilon}\frac{\delta H_y}{\delta x}$, then the boundary conditions are identical to those of section (III-2). As a consequence, the aforementioned fields become good approximations to the transverse vector modal functions. Using the

fields given in Appendix II , away from resonance where C ≅ 1, we have

$$\bar{e}_a = \left[\frac{\beta}{\omega\varepsilon_b x_0 L_y}\right]^{\frac{1}{2}} \cosh(\nu_b x - \phi_m)\hat{x} \qquad \text{(IV.4.9)}$$

$$\bar{h}_b = \left[\frac{\omega\varepsilon_b}{\beta x_0 L_y}\right]^{\frac{1}{2}} \cosh(\nu_b x - \phi_m)\hat{y} \qquad \text{(IV.4.10)}$$

and

$$\phi_s = \tanh^{-1}\frac{\varepsilon_b \nu_s}{-\varepsilon_s \nu_b}$$

$$\phi_m = \tanh^{-1}\frac{\varepsilon_b \gamma_m}{-\varepsilon_m \nu_b}$$

Comparing the power in the insulator to the total power for the visible region of the spectrum, it is noted that the fields are concentrated in the insulator (C ≅ 1). Therefore, essentially all coupling will occur there, and only the fields in the insulator are used to calculate the coupling coefficient. Using Eqns. (6,9 and 10), assuming small variations of the fields across the barrier,

$$\kappa \cong \frac{d\nu_b}{dz}\,\frac{(2\nu_b x_0 - 3\phi_s)x_c}{6} \qquad \text{(IV.4.11)}$$

For the visible region of the spectrum, ν_b may be approximated by,

$$\nu_b \cong \frac{1}{x_0(z)}\,(\phi_{s\infty}+\phi_{m\infty}) \qquad \text{(IV.4.12)}$$

$\phi_{s\infty}$, $\phi_{m\infty}$ being equal to ϕ_s, ϕ_m with $\gamma_s/\nu_b \to 1$ and $\gamma_m/\nu_b \to 1$ respectively. Using this expression in Eqn. (11), one obtains

$$\kappa = \frac{(3\phi_s - 2\nu_b x_0)(\phi_{s\infty}+\phi_{m\infty})}{6x_0}\,\frac{dx_0}{dz} \qquad \text{(IV.4.13)}$$

For a symmetrical structure $\phi_{s\infty} = \phi_{m\infty}$ and $\nu_b = \dfrac{\phi_{s\infty}}{x_0}$. Thus

$$\kappa \cong \frac{1}{6} \nu_b^2 x_0 \left(\frac{dx_0}{dz}\right) \cong \frac{1}{6} \frac{\omega_p^2}{c^2} x_0 \frac{dx_0}{dz} \qquad \text{(IV.4.14)}$$

For an exponential taper with profile $x_0 = Ae^{-\alpha z}$ one obtains a peak value of κ of

$$|\kappa_{peak}| = \frac{1}{6} \frac{\omega_p^2}{c^2} A^2 \alpha \cong (4 \times 10^{-2}) \alpha \qquad \text{(IV.4.15)}$$

with α in cm^{-1}. Here we have taken $\omega_p/c \cong 5 \times 10^5$ cm^{-1} and $A \cong 100$ Å being the maximum value of x_0 for which the approximations are valid. Thus $\kappa_{peak} \cong 4\alpha \cong 400$ cm^{-1} for $\alpha = 10^4$ cm^{-1}. This indicates that an exponential taper of the order of 3 μm in length will guarantee low power coupling into the backward mode while going from 100 Å to 10 Å. More detailed numerical calculations of the power converted into the backward mode are presently being carried out.

IV-5 Heterodyne Diagnostic Experiment - Numerical Estimates

To investigate the non-linearity of tunneling junctions at high frequencies as well as the various coupling schemes discussed in the previous sections we have devised a heterodyne mixing experiment which uses conveniently available sources in our lab, and provides a high frequency r.f. output which can be monitored on a spectrum analyzer. This involves the heterodyne mixing of the output of a water vapor laser with frequency (11THz) and the output of a carbon dioxide laser with a frequency of 32 THz. The output frequency $\omega_{IF} = 3\omega_{H_2O} - \omega_{CO_2}$ should be at approximately 20 GHz (Fig. IV-8). The geometry and fabrication of the devices is discussed in section V-5. The preliminary calculations are carried out in three stages. First the input coupling is analyzed, then the conversion efficiency, and finally the output coupling to the spectrum analyzer.

The first step in analyzing the input coupling is to determine the incident power density. We have assumed 75 mwatts of water vapor laser power, and 1.2 watts of CO_2 laser power available. Both of these levels have been achieved in our laboratory. We believe that the water vapor laser power can be increased, and that 75 mwatts is a conservative estimate.

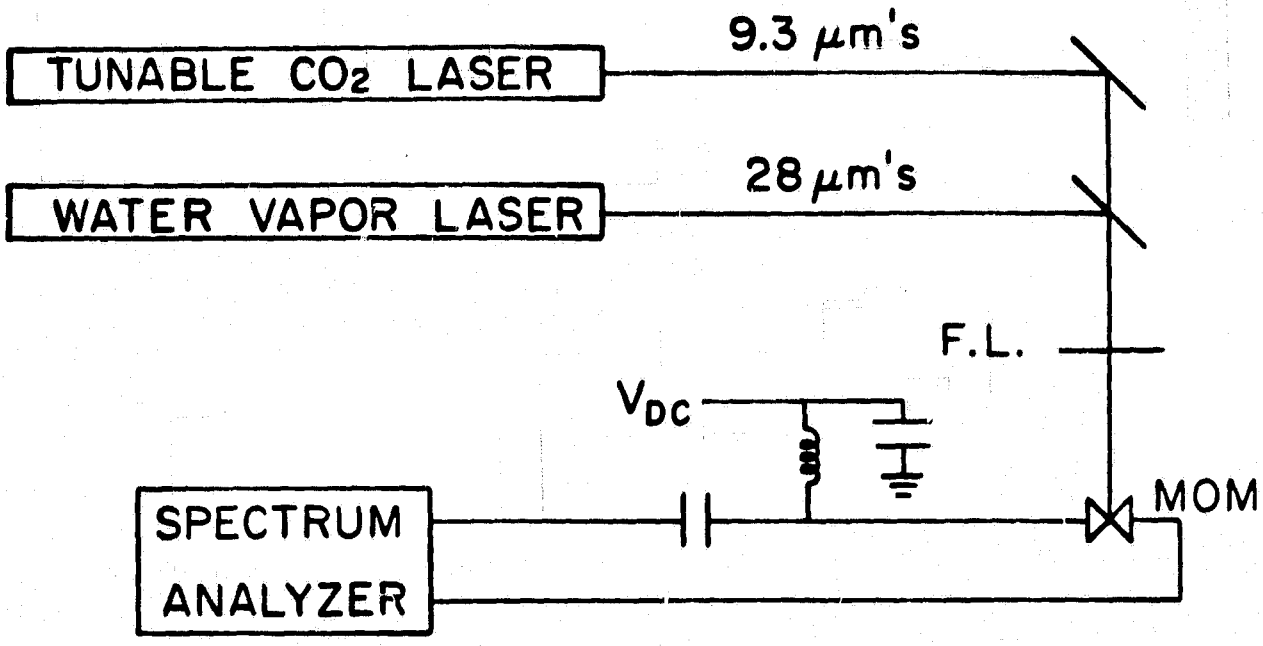
The 28 μm (11THz) beam is to be at almost normal incidence to the silicon substrate and will therefore suffer a reflection loss ρ^2

$$\rho^2 = \left(\frac{n_{si}-1}{n_{si}+1}\right)^2 \qquad \text{(IV.5.1)}$$

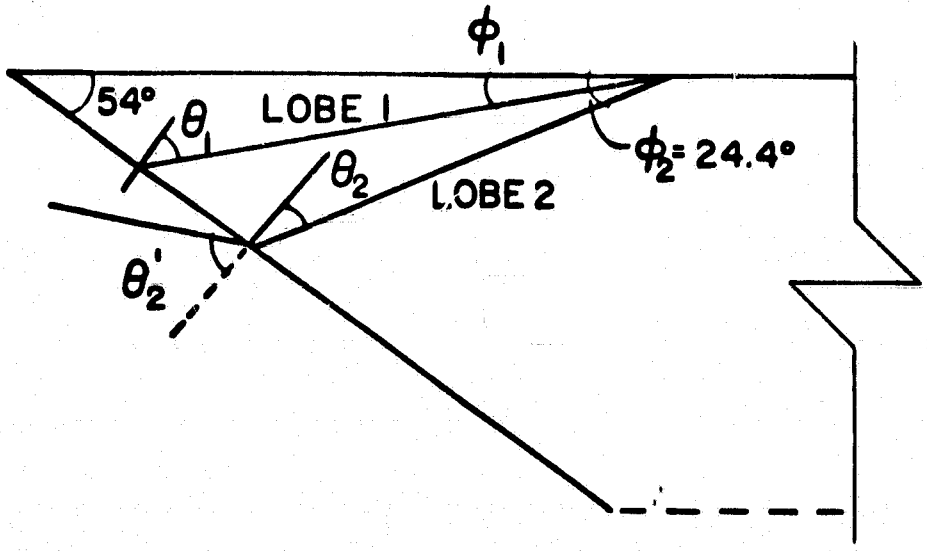
of .274. The 9.3 μm (32THz) beam will be incident at an angle of 40° from the normal, the difference in angle being determined by the relative locations of the gain maximum of the antenna at the two frequencies. At 40° the reflection loss is

$$\rho^2 = \left|\frac{\cos\theta_2 - n_{si}\cos\theta_2'}{\cos\theta_2 + n_{si}\cos\theta_2'}\right|^2 = .210 \qquad \text{(IV.5.2)}$$

where the angles are shown in Fig. (IV-9).



IV-8 Schematic diagram of the heterodyne diagnostics experiment.



IV-9 Directions of first two lobes in the 9.3 μm antenna pattern.

The diffraction limited spot size can be determined from the beam diameter (D), the wavelength, and the focal length of the lens or mirror being used for focussing. The radius of the spot is approximately

$$r_D = 1.22 \frac{\lambda}{D} \sqrt{f^2 + \left(\frac{D}{2}\right)^2} \qquad \text{(IV.5.3)}$$

which is 114 μm for the CO_2 laser and 683 μm for the H_2O laser. The available incident power density at the antenna is then $I_{INC} = P_{laser} |1-\rho^2| \times [\pi r_D^2]^{-1} = 3.7$ watts/cm^2 H_2O power and 2.4×10^3 watts/cm^2 CO_2 power.

In order to estimate the fraction of incident laser intensity coupled into the antenna structure and hence coupled to the junction, we next determine V_A, the voltage induced across the antenna, which requires estimates of the antenna gain at the frequencies of interest and for the particular antenna lobes of interest.

Consideration of the true dielectric constant of the metals yields Eqn. (IV-15) as the analogue of a low frequency antenna calculation. The magnitude of the first order corrections to k_0, is of the order of $(\omega/\omega_p)^2 \times k_0$. For nickel, the metal currently in use, $(\omega/\omega_p)^2 \approx 3 \times 10^{-5}$ at 28 μm and 2.4×10^{-4} at 9.3 μm thus $k \cong k_0$. Also, the decay length (the length for which $\alpha\ell = 1$) is in the centimeter range for both of these wavelengths. Since all of the antenna structures are much smaller than this, $\alpha\ell$ can be assumed to be much less than 1.

With $k_0/k \cong 1$, and $\alpha\ell << 1$ for both frequencies of interest, Eq. IV-15 reduces to the normal low frequency antenna result. This implies that standard antenna theory can be used to calculate the gain of the antenna structures used in the length confinement coupling scheme. We have carried out a computer calculation of the gain at 28 μm [18]. The results indicate that for an antenna of length $L = 5.5\lambda$, the gain should be approximately 34 at an angle which is nearly normal to the preferentially etched edge. At 9.3 μm, formula IV-15 can be integrated over solid angle to give the total power radiated. $\cos(k\ell) = 0$ for antenna of length $k_0 L = 33\pi$ so that

$$P_R = \frac{\eta}{8\pi^2} I_0^2 \int_0^{2\pi} d\phi \int_0^{\pi} \left(\frac{\cos(k_0 L \cos\theta) + 1}{\sin\theta} \right)^2 \sin\theta d\theta \qquad \text{(IV.5.4)}$$

substituting $\Gamma = \cos\theta$ gives

$$P_R = \frac{\eta}{4\pi} I_0^2 \int_{-1}^{+1} \frac{(\cos(33\pi\Gamma)+1)^2}{1-\Gamma^2}\, d\Gamma \qquad \text{(IV.5.4)}$$

Splitting $\frac{1}{1-\Gamma^2}$ into $\frac{1}{2(1-\Gamma)} + \frac{1}{2(1+\Gamma)}$, substituting $\nu = 1 - \Gamma$ and $\omega = 1 + \Gamma$ into the subsequent integrals gives

$$P_R = \frac{\eta}{4\pi} I_0^2 \left[\int_0^2 2\,\frac{1-\cos 33\pi\nu}{\nu}\, d\nu - \int_0^2 \frac{1-\cos 66\pi\nu}{2\nu}\, d\nu \right]$$

$$= \frac{\eta}{4\pi} I_0^2 [2\{\ln 66\pi + .5772 - Ci(66\pi)\} - \tfrac{1}{2}\{\ln(132\pi) + .5772 - Ci(132\pi)\}]$$

$$= 8.52\left(\frac{\eta I_0^2}{4\pi}\right) \qquad \text{(IV.5.5)}$$

The gain at any angle, is then

$$\frac{\left(\frac{\partial P}{\partial \Omega}\right)(\theta,\phi)}{(P_R/4\pi)} = \frac{2}{8.52}\left\{\frac{\cos(k_0 L\cos\theta)+1}{\sin\theta}\right\}^2 \qquad \text{(IV.5.6)}$$

Maxima in the pattern are near angles for which $16.5\pi\cos\theta = m\pi$, $m = 1,2,\ldots,16$. Letting $m = 15$ yields for the gain of the second lobe 5.44. This gain calculation is for an antenna in a uniform surrounding medium. In the actual structure, the antenna current is at the silicon-metal interface, and, according to the results of section (2) radiates very little into the air half space. The total power radiated should thus be half of the above doubling the gain to 10.88 at an angle of 24.4°. The radiation strikes the etched Si edge at an angle of 10.7 degrees (Fig. IV-9) with respect to the normal. Snell's law, shows that the incident angle of light required to couple into this lobe is 40° (θ_2). It is not possible to couple to the primary lobe since the critical angle is exceeded.

At both 9.3 microns and 28 microns, the response of the metal is large enough and the junction small enough that the surface plasmon wave gap mode reduces to the quasi-static limit of a lumped capacitor. The result is that the total diode impedance is

$$Z_D = R_D \,//\, \frac{-j}{\omega C} = [R_D/\omega^2C^2 - jR_D^2/\omega C][R_D^2 + 1/\omega^2C^2]^{-1} \qquad \text{(IV.5.7)}$$

where $C = \frac{\varepsilon A}{d}$ and R_D is the dynamic tunneling resistance of the diode. Assuming A = 1 μmx200Å, d = 20Å and $\varepsilon_{rel} = 2$ yields C = 1.77×10^{-16} farads where ε_{rel} is an estimate of the high frequency dielectric constant of the 20 Å nickel oxide. These formulas lead to the following impedances for the diode at 28 μm and 9.3 μm.

$$Z_D\,(28) = [13.82 - j81.96]\Omega \qquad \text{(IV.5.8)}$$

$$Z_D\,(9.3) = [1.59 - j28.15]\Omega \qquad \text{(IV.5.9)}$$

where $R_D = 500\Omega$.

Integral wavelength antennas have current nodes at the feed point. This leads to a large real antenna impedance (500-2000Ω). Shortening the length of the antenna slightly results in a slightly inductive impedance. For the worst case calculation, we use a 2000Ω impedance and maximize antenna-diode mismatch. For the best case calculation we assume the antenna to be slightly inductive to resonate out the capacitive parts of the terminations.

Standard antenna theory gives[19]

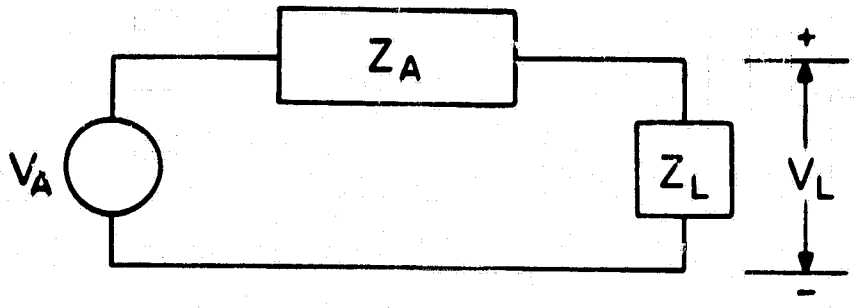
$$P_{ml} = \frac{\lambda^2}{4\pi} GI_{inc} = \frac{1}{8}\frac{V_a^2}{R_a} \qquad \text{(IV.5.10)}$$

for the maximum power coupled to the load. This assumes a load conjugately matched to the antenna. From this formula, one can solve for V_a and calculate the actual coupled power as

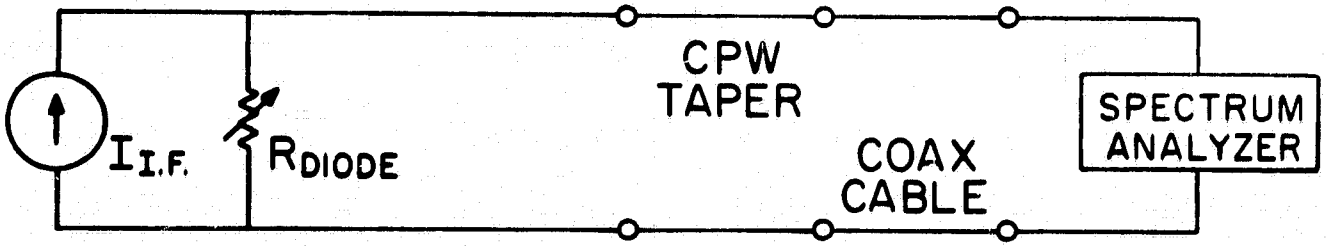
$$V_a = \left(\frac{\lambda^2 GI_{inc} 8R_a}{4\pi}\right)^{\frac{1}{2}} \quad \text{and} \qquad \text{(IV.5.11)}$$

$$|V_D| = \frac{|V_a|\,|Z_D|}{|Z_D + Z_a|} \qquad \text{(IV.5.12)}$$

(see Fig. IV-10)



(a)



(b)

IV-10 Equivalent circuit for antenna coupling to the load and the I.F. connection.

The above numbers yield for the maximum and minimum induced voltages.

Maximum	Minimum
$V_D(CO_2)$ = 148 mv	$V_D(CO_2)$ = 52 mv
$V_D(H_2O)$ = 90 mv	$V_D(H_2O)$ = 46 mv

Knowing the load (diode) voltages allows us to estimate the intermediate frequency (19.56 GHz) signal generated by the diode. Using a Taylor's series expansion about the d.c. bias gives

$$I = \sum_n \frac{1}{n!} I^{(n)} V^n \qquad \text{(IV.5.13)}$$

where V is the total junction voltage and

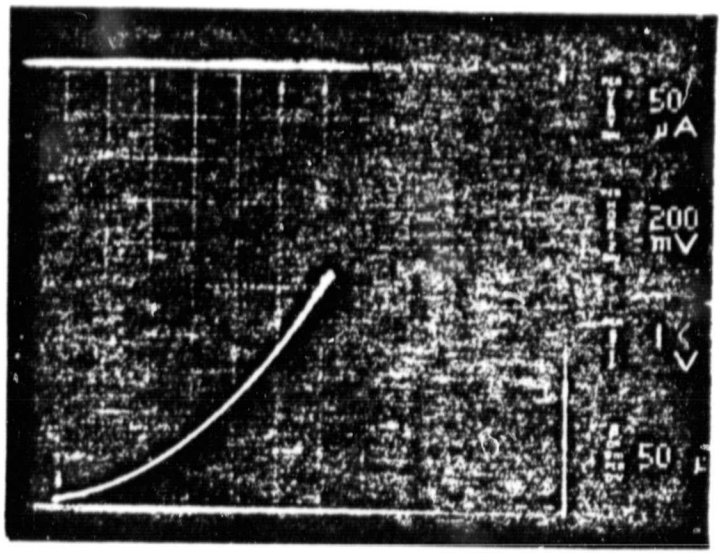
$$I^{(n)} = \frac{\partial^n I}{\partial V^n} \qquad \text{(IV.5.14)}$$

The n=4 term yields $\frac{1}{4!} I^{(4)} \cdot V^4$

V^4 contains the term $4V^3(\omega_1)V(\omega_2)e^{j(3\omega_1-\omega_2)t}$ which contributes the 20 GHz term. As a minimum $I^{(4)}$, we take Javan's[20] number of $10^{-6} \frac{\mu a}{(mv)^4} = 10^{-12} \frac{amps}{(mv)^4}$. Our devices are noticibly more nonlinear (Fig. IV-11), the sensitivity $S = \frac{I''}{2I}$, ranging from two to six per volt. Those of Javan are in the range of 10^{-4} per volt. Numerical calculation gives a minimum of 2×10^{-12} amps $(mv)^{-4}$, for the $I^{(4)}$ of our planar diode, using a low frequency I-V curve. There is some experimental evidence[21] which suggests that the planar diodes are more nonlinear at higher frequencies. In addition, we hope to be able to increase the device nonlinearity, as already discussed.

Using 3×10^{-12} amps/$(mv)^4$ as a reasonable maximum for $I^{(4)}$ results in $I_{max}(19.5\ GHz) \cong 1\times10^{-4}$ amps and $I_{min}(19.5\ GHz) \cong 0.7\times10^{-6}$ amps. The current is split between the diode, and load (Fig. IV-10) as

$$I_{load} = I_{IF} \times \frac{500\Omega}{550\Omega}$$



IV-11 Planar junction current-voltage characteristics.

and so the 20 GHz power launched down the coplanar waveguide is

$$P_{TL} = I^2_{load} \times 50\Omega = I^2_{IF}\left(\frac{500}{550}\right)^2 \times 50$$

Thus for the maximun and minimum values of P_{TL} one obtains

$$P_{TL} = 1.15 \times 10^{-4} \text{ mwatts} = -39 \text{ dbm max}$$
$$= 2.8 \times 10^{-8} \text{ mwatts} = -75 \text{ dmb min}$$

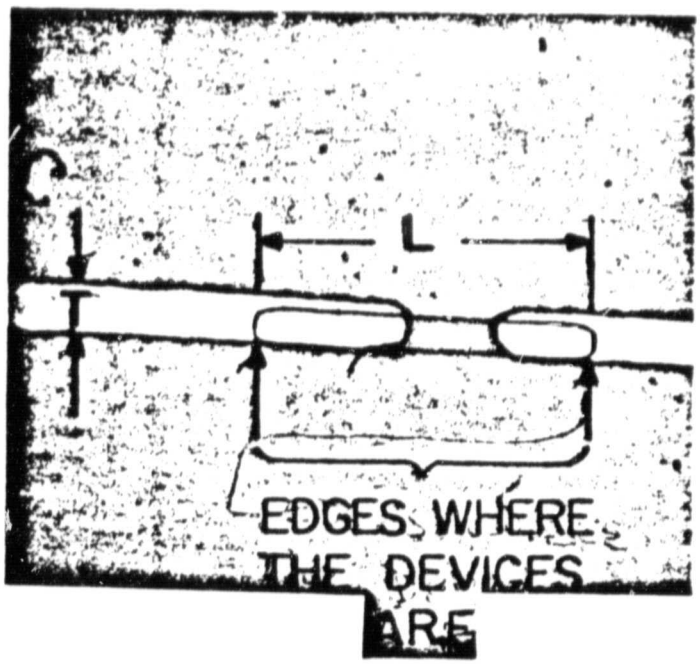
There are several additional sources of loss in the diode to spectrum analyser I.F. connection. These include, propagation loss in the co-planar waveguide (CPW) taper, propagation loss in the coaxial cable which connects the CPW to the spectrum analyzer, and losses at each of the microwave connectors. Using the supplied information on the cables and connectors we have estimated these additional losses to total less than 7.1 db. Subtracting this from the IF power level sent down the transmission guide yields the final minimum and maximum expected IF signals at the spectrum analyzer as P_{SA}(19.5 GHz) = -46 dbm maximum and -83 dbm minimum.

The Tektronix spectrum analyzer which we have will detect signals as low as -95 dbm, if the bandwidth is sufficiently narrow. At its widest bandwidth the sensitivity is -65 dbm. Phase sensitive post detection can also be used to increase the sensitivity further.

These estimates certainly indicate that this particular experiment should be succesful with out continuous wave water vapor and CO_2 laser sources and our present devices shown in Fig. (IV-12).

Since the actual structure consists of an array of two diodes raised from the substrate, the coupling characteristics will differ from those calculated above. The estimates are not expected to change significantly due to the distance between the substrate and the center electrode since this is of the order of a few hundred Å, which is appreciably less than the plasmon wavelength. The thickness of the center electrode is however less than the penetration depth of the single surface plasmon or of the same order and thus the mode, and hence characteristic impedance, will be altered somewhat. In addition, the short stub (center electrode) causes a phase shift so that the junctions are shifted from an end

(a)



(b)

IV-12 Photographs of metal-barrier-metal tunneling devices fabricated at Berkeley. a) Overall view, b) Magnified to show antenna, c) Further magnified to show the junction. Antenna width is 1 μm.

termination. As long as the resonance condition is satisfied, this should not appreciably alter the expected signal strength. We are presently analyzing these two effects.

Finally, shifting the experiment towards optical frequencies is expected to change the results in several ways. The devices and antenna structures will have to be made smaller or traveling wave junctions will have to be exploited. Since the plasmon becomes more lossy as one enters the relaxation region or in the visible is even influenced by the resonance behaviour of the plasmon, the coupling will decrease. A great deal could be learned by considering the coupling as a function of temperature at these short wavelengths. However, it is expected that no fundamental problems exist for coupling at visible frequencies and we anticipate moving in this direction when our basic results have experimental verification.

References for Chapter IV

1. A. Otto, Festhorperprobleme XIV, 1, (1974).
2. E. Kretschmann, Z. Physik 241, 313 (1971).
3. E. Burstein, W.P. Chen, Y.J. Chen and A. Hartstein, J. Vac. Sci. Technol., 11, 1004 (1974).
4. Y. Chabal and A.J. Sievers, Appl. Phys. Lett. 32, 90 (1978).
5. D.P. Siu and T.K. Gustafson, Appl. Phys. Lett. 31, 71 (1977).
6. A. Sanchez, C.F. Davies, Jr., K.C. Liu, and A. Javan, J. Appl. Phys. 49, 5270 (1978).
7. R.H. Bjork, A.S. Karakashian, and Y.Y. Teng, Phys. Rev. B9, 1394 (1974).
   C.E. Wheeler, E.T. Arakawa, and R.H. Ritchie, Phys. Rev. B13, 2372 (1976).
   I. Pockrand and H. Raether, Optics Communications, 18, 395 (1976).
   R.J. Seymour, E.S. Koteles, G.I. Stegeman , Appl. Phys. Lett. 41, 1013 (1982).
8. S.R.J. Brueck and D.J. Ehrlich, Phys. Rev. Lett., 48, 1678 (1982).
9. D. Maystre and R. Petit, Optics Communications, 17, 196 (1976).
10. References to surface roughness include M. Fleischmann, P.J. Hendre and A.J. McQuillan, Chem, Phys. Lett. 26, 163 (1974); D.L. Jeanmaire and R.P. Van Duyne, J. Electroanal. Chem. 84, 1 (1977); P.F. Liao and M.B. Stern, Optics Letters, 7, 483 (1982).
11. Y.R. Shen, C.K. Chen, T.F. Heinz, and D. Ricard, Springer Series in Optical Sciences Laser Spectroscopy V, (editors A.R.W. McKellar, T. Oka and B.P. Stoicheff (1981) page 412; G.I. Stegeman, J.J. Burke, and D.G. Hall, Appl. Phys. Lett. 41, 906 (1982).
12. R.M. Osgood, Jr., and P.J. Ehrlich, Optics Letters, 1, 385 (1982) S.R.J. Brueck and D.J. Ehrlich, Phys. Rev. Letters, 48, 1678 (1982).
13. A. M. Glass, P. F. Liao, D. H. Olsen, and L. M. Humphrey, Optics Letters 1, 575 (1982).
14. P. K. Tien and R. Ulrich, J. Opt. Soc. Am. 60, 1325 (1970); R. Ulrich, J. Opt. Soc. Am. 60, 1337 (1970); R. Ulrich, J. Opt. Soc. Am., 61, 1467 (1971).
15. Collin , Robert E., Foundation for Microwave Engineering, McGraw Hill Book Company, New York; 1966.

16. Allan W. Snyder, "Coupling of Modes on a Tapered Dielectric Cylinder," IEEE Transactions on Microwave Theory and Techniques, Vol. MTT-18, 383 (1970); and see also Ref. (14).
17. T. Tamir (ed.) Integrated Optics, Springer-Verlag, Berlin (1979).
18. B. Hooper - masters report UCB (unpublished).
19. W.L. Stutzman and G.A. Thiele, Antenna Theory and Design, (copyright 1981 - John Wiley & Sons, Inc.).
20. A. Javan and A. Sanchez, preprint providing nonlinear coefficients $\frac{\partial^n I}{\partial V^n}$ for point contact metal-oxide-metal diodes (publication not indicated).
21. C.W. Slayman, Metal-Barrier-Metal Junctions for Millimeter Wave Mixing and Detection, Doctoral dissertation, University of California, Berkeley, 1980 (unpublished).

V FABRICATION OF JUNCTION DEVICES

V-1 Introduction

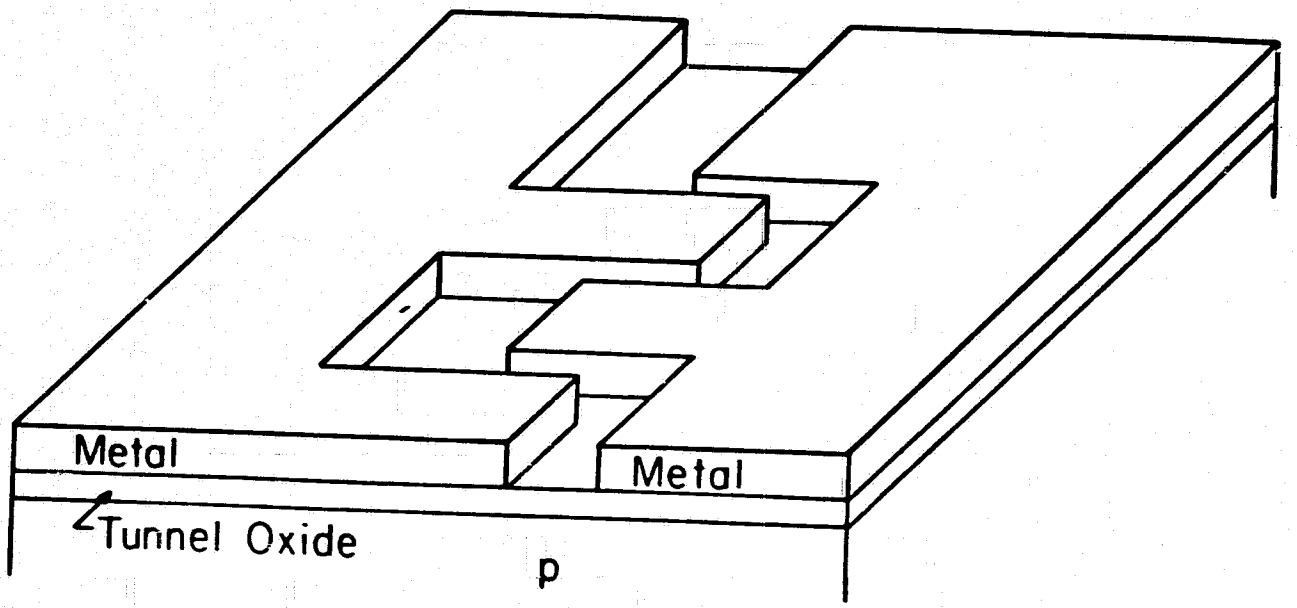
In the present chapter we discuss the fabrication of junction devices presently being used to investigate various coupling and tunneling aspects pertaining to direct conversion through tunneling.

We have generally been using three configurations for our tunneling devices: the metal-oxide-p^{++} semiconductor, the metal-oxide-ptype semiconductor-oxide-metal junction device, and the metal-barrier-metal junction device. The first two have been useful because of their inherently fast response time and thus for the detection of short optical pulses. This requires no coupling to the surface mode and is therefore much easier to demonstrate than a highly efficient direct conversion process. Any utilization of these junctions to exhibit conversion efficiencies theoretically available from tunneling devices will require the development of such coupling schemes.

The primary focus of the present chapter will be the edge metal-oxide-metal device presently being utilized in a fourth order mixing experiment being implemented as a diagnostic tool to measure both the junction nonlinearity and coupling at high frequencies. Unfortunately the metal-oxide-metal (MOM) is not a highly nonlinear junction, but it is expected that it can be improved as fabrication procedures evolve and possibly by combining it with semiconductor technology. Prior to describing its fabrication in detail, we present a short summary of the work on the two other devices.

V-2 The M-O-P-O-M Junction

The M-O-P-O-M (Fig. (V-1)) interdigitated electrode structure presently operates well as a fast conventional photodetector with a response time of approximately 50 picoseconds[1]. The carriers are generated in the bulk semiconductor by flooding the junction with the incident radiation through the interdigitation. This interdigitation can in principle be used as a grating coupling structure for the excitation of surface waves however this is not essential for the detector, hence its simplicity, although we have still referred to it as a surface detector. This is the simplest of



V-1 M-O-P-O-M device with interdigital electrode structure.

the three devices to fabricate and will concievably be utilized as a surface device for possible conversion experiments by attempting this coupling into the oxide layer.

To fabricate the device the silicon wafer is first cleaned and a thin oxide layer is grown thermally in a dry oxygen atmosphere. The thin oxide thickness is monitored using an ellipsometer, typical thicknesses presently ranging from thirty to sixty Å.

Subsequent to oxidation, both aluminum electrodes are evaporated and patterned using a lift-off technique. This particular tunneling device operates by having electrons tunnel from one electrode through the oxide into the semiconductor substrate. In order to complete the electron path from one electrode through the oxides, to the second electrode, free carriers must be supplied to the semiconductor either by excitation or by doping. For the detector, we utilize photo-generation. For a tunneling converter this would be a heavily doped layer or even concievably a metallic layer. We are attempting to modify this structure by first laying the device on a saphire substrate with approximately 3000 Å of silicon. Such a thin Si layer would confine the current path. Requiring no new fabrication steps, the simplicity of the device is retained, while allowing the eventual fabrication of these devices on a metallic substrate for the guiding and coupling.

V-3 The M-O-S (p^{++}) Junction

There are two important factors motivating the fabrication of the second class of (M-O-S(p^{++})) devices. These are expected to have more nonlinear current voltage characteristics than simple M-O-M devices and hence provide better conversion characteristics. Secondly the plasma resonance frequency can be tailored by doping, thus possibly providing a means of enhancing conversion characteristics. Three types of semiconductors (Si, GaAs, SnTe) are being used. For silicon, thermal oxidation is first carried out. Subsequently the oxide is patterned, and the silicon ion implanted to create a p^{++} region. The wafers are then annealed, and a thin oxide layer grown and measured as for the M-O-P-O-M devices. The metal can then either be evaporated or sputtered to form the electrode structure. If p^{++} GaAs is used, it is obtained either by purchasing it

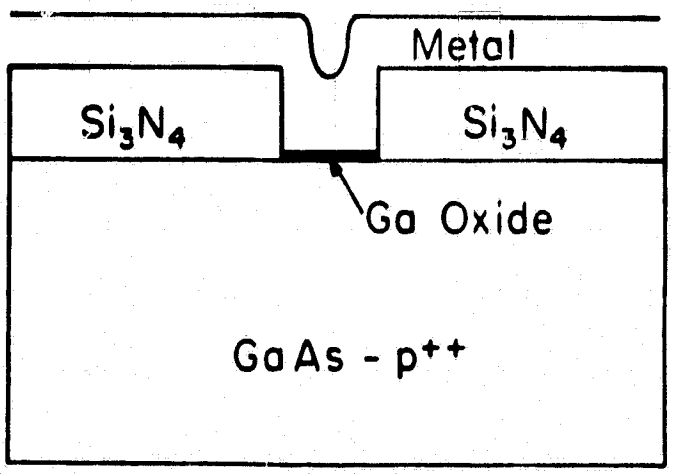
directly from a supplier or by growing epitaxially. SiO_2 pr Si_2N_4 is then deposited and a window opened through the insulator (Fig. (V-2)). Ga_2O_3 is then grown using the sputter oxidation technique described below for the "edge MOM" fabrication procudure. The metal layer can then also be deposited in the sputtering machine without breaking the vacuum.

SnTe has the convenient property that it can be evaporated and still retain semiconducting characteristics, the resultant film being p^{++}(2). This allows the deposition and patterning of the metal layer first and the use of a metal oxide tunneling layer obtained via sputter oxidation, thermal oxidation in air, or by placing it in an oxygen plasma in an evaporator. The last method allows the evaporation of SnTe onto the substrate, or immediately following the oxidation. To finish the device, the SnTe must be patterned, either by lift-off or chemically using HCl.

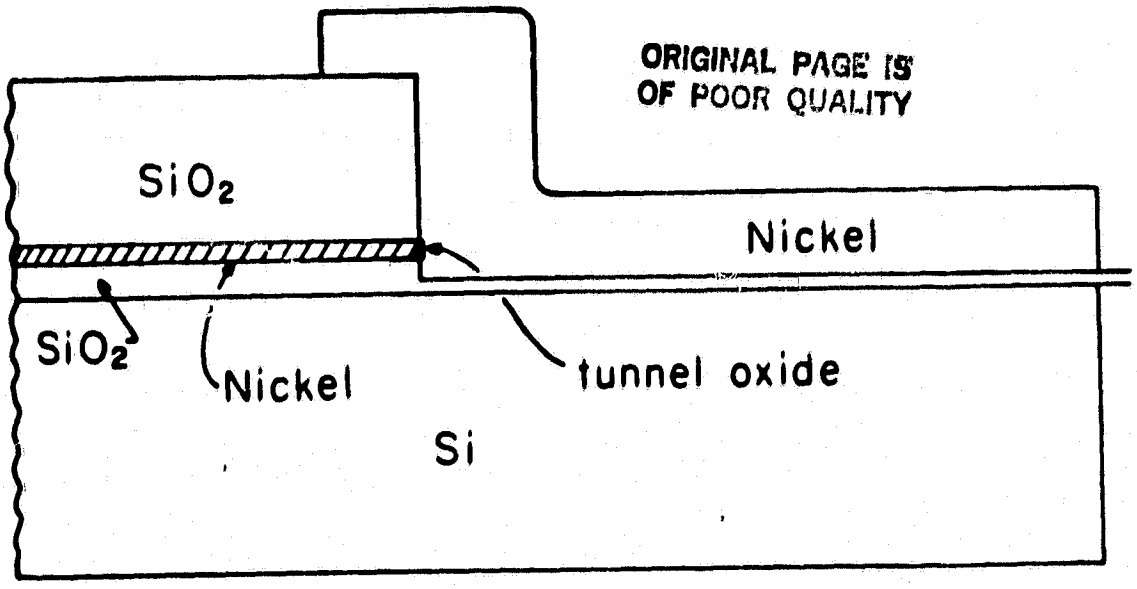
V-4 The M-O-M Tunneling Junction Fabrication

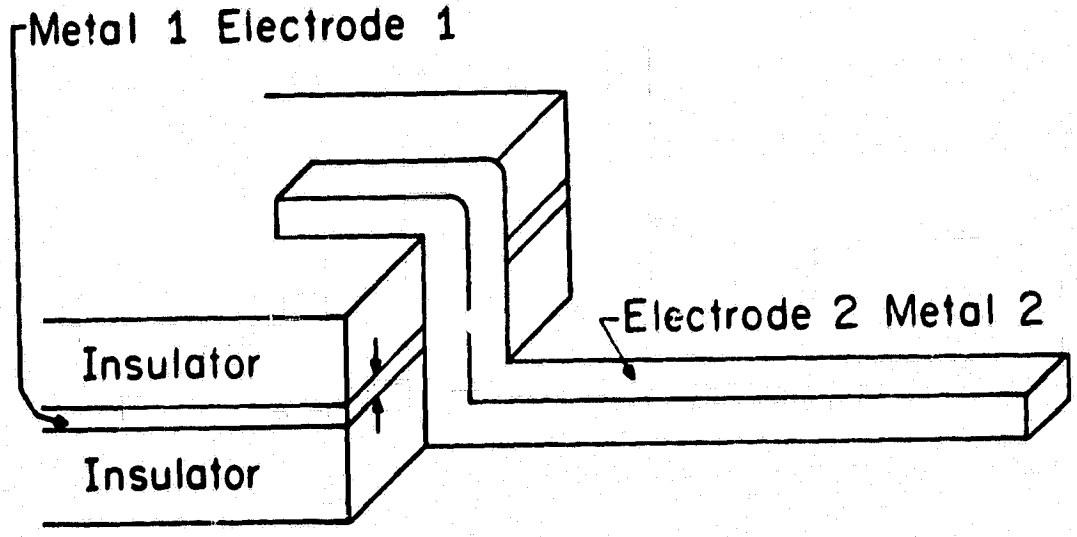
The EDGE MOM is so named because the metal-oxide-metal junction is formed on the edge of one of the two metallic electrodes. In this configuration, one dimension of the device is thus determined by the thickness of the metal on whose edge the junction is fabricated (Fig. (V-3)). This dimension may be as small as 100-200 Å. This device is fabricated using two rf sputtering deposition steps. The first electrode is patterned using an ion milling machine, and the second with a wet chemical etchant. In our fabrication procedure for these junctions rf sputtering has been heavily used. Three different processes are involved: sputter etching, sputter deposition, and sputter-bias deposition. All three of these modes involve a plasma whose ions are accelerated into either the material to be deposited (target), or the substrate, or both. In sputter deposition, the ions are accelerated into the target where they dislodge target molecules which drift down onto the substrate.

When sputtering a metal, a purely nonreactive Ar plasma is used. When depositing SiO_2 (or Si_3N_4), some O_2 (or N_2) must be added to the plasma to keep silicon compounds with fewer oxygen (nitrogen) atoms from forming. In sputter etching, the plasma ions are accelerated into the substrate, with resultant ballistic etching. Ions are accelerated into both the target and substrate in the sputter bias mode. This deposits and etches



V-2 M-O-P^{++} (GaAs) structure used to study interactions between surface electromagnetic waves and tunneling current.





V-3 Edge-MOM structure. a) Front view b) Angular view

simultaneously resulting in smoother, lower resistance metal films.

The sputter deposition mode is used to clean each target before the fabrication process. Material is sputtered off the target onto a blank. Any absorbed impurities are then sputtered away before rotating the substrate under the target.

The first electrode consists of an insulator-metal-insulator sandwich. The substrate is introduced into the sputtering machine, the targets are cleaned, and the insulator is sputter deposited. If SiO_2 is used as the insulator, the metal target to be used for the electrode material must be recleaned to remove any nickel oxide which grew in the argon-oxygen plasma. A thin layer of metal is deposited using the sputter bias mode. It is important to obtain the best possible metal film during this step, since it is so thin, typically 200-400 Å. The upper insulating layer is then deposited. The lower insulating film protects the silicon surface during a subsequent ion milling procedure, and the milling process is easily stopped before etching all the way through this layer and into the wafer. The upper insulating layer keeps the two electrodes from shorting. This insulator must be made thick enough to minimize the electrode to electrode capacitance, typically greater than 1000 Å. Photo resist is next spun on and patterned by standard photolithographic procedures. The first electrode is then ready to be etched.

The sandwich structure is etched in an ion mill, in order to obtain smooth vertical walls. This is essential to building devices on the edge of the metal. The photoresist used to pattern this electrode must be post-baked at 120° for at least 15 minutes, to harden it prior to being placed in the ion mill. The Si wafer is also heat sunk with gallium to a cooled substrate table. If the substrate is not heat sunk, two problems arise in the etching. First, the photoresist becomes too hot, and polymerizes, making it very difficult to remove. Secondly, the etching process is disturbed. The part of the wafer being etched, develops "hot" and "cold" spots which tend to either etch quickly, or very slowly. This results in parts of the wafer effectively not being etched, while others are etched too deeply. The etch process is timed to etch into the lower insulator layer, but not through it. Etching through the lower insulating layer results in a roughening of the surface of the Si, making subsequent

chemical etching of the upper electrode difficult. This problem cannot be eliminated by lowering the incident ion energy, as even low energy ion etching has been observed to cause surface damage[3].

At the end of a usual etching process, the photoresist can be removed by ultrasonically cleaning the wafer in acetone. Occasionally, it is also necessary to do a short plasma ashing to remove the last photoresist residue. The wafer is now ready for the upper electrode deposition.

Subsequent to the lower electrode patterning, the wafer is returned to the rf sputtering system. Without breaking the vacuum, the wafer is sputter cleaned, the oxide barrier grown on the lower electrode edge, and the upper electrode deposited. The wafer is sputter cleaned (etched) for two minutes in an argon atmosphere to remove any oxide which may already have grown on the lower electrode. This also removes a small amount of residual photo resist if the stripping process has not entirely removed it. The lower electrode is then sputter oxidized, and the upper electrode metal deposited.

The rf sputter oxidation procedure developed by Greiner[4,5] involves two competing processes. The substrate is sputter etched using a plasma containing oxygen. The etch process removes material from the substrate at a constant rate R. If X is the thickness of oxide present, $\frac{\partial X}{\partial t} = -R$ for the etch process alone. The presence of oxygen in the plasma causes metal oxide to grow. However, the rate of growth is not constant, since it requires metallic atoms to be available at the growth surface. This exposure decreases with increasing oxide thickness, and the growth rate decreases. The oxide growth rate in a plasma decreases as[3] $\frac{\partial X}{\partial t} = Ke^{-X/X_0}$, giving $\frac{\partial X}{\partial t} = Ke^{-X/X_0} - R$ for the net rate of change in oxide thickness during sputter oxidation. As metal oxide grows, the growth rate slows, until it equals the etch rate, and no further oxidation occurs. Setting $\partial X/\partial t = 0$ gives $X_{final} = X_0 \ln\frac{K}{R}$ as the equilibrium thickness. This thickness is approached exponentially with a time constant of X_0/R. This time constant varies substantially from metal to metal. In some metals, the time constant, or equilibrium thickness are such that for the fabrication of practical tunneling barriers the equilibrium is never reached and the time of growth becomes an

important parameter determining the barrier thickness. In the case of Ni, the constant has been determined to be only a few minutes[6], so that the equilibrium thickness is reproducibly achieved by allowing the process to run for 10-15 minutes. The equilibrium thickness can be varied by adjusting the rf power applied to the substrate, and the amount of oxygen in the plasma. With Ni, the equilibrium thickness can be adjusted to produce edge MOM's with zero bias dynamic resistances from about 750 Ω to about one MΩ. Wafers with lower zero bias resistances have a noticably larger percentage of shorted devices on the wafer.

The oxide chamber is covered as quickly as possible after it is grown. The chamber is evacuated to remove the O_2, and refilled with a pure argon sputtering atmosphere. Ni is first sputtered off of the target, onto a blank table, to remove any nickel oxide formed on the target. A thin layer of Ni (approximately 200-400 Å) is then sputter deposited very quickly onto the substrate (approximately two minutes). After this initial layer, which protects the barrier, the rest of the upper metal layer is deposited using the sputter-bias mode.

The upper electrode is etched using a standard chemical etchant rather than ion milling. It in fact has been experimentally determined that ion milling of the second electrode causes the junctions to short. It has also been observed that if during the ion milling of the lower electrode, the lower insulating layer is etched through, the Ni deposited for the second electrode adheres too strongly to the roughened Si surface. The roughened surface makes it difficult for the chemical etchant to thoroughly remove the Ni. To do so would require the wafer to be left in the etchant so long that all lines less than 5-10 μm's wide are undercut, and etched away. One solution to this problem has been to put the wafer in the ion milling machine for a very short period, to etch this last layer of Ni.

With the completion of this final etching, the edge MOM is ready to be tested on the curve tracer. This basic procedure for the M-O-M fabrication has had several and various modifications, and variations which have been, or are to be, tried to increase reliability.

Because of the inherent geometry chosen, (edge), it is easier to actually fabricate a two junction series array, than to fabricate a single

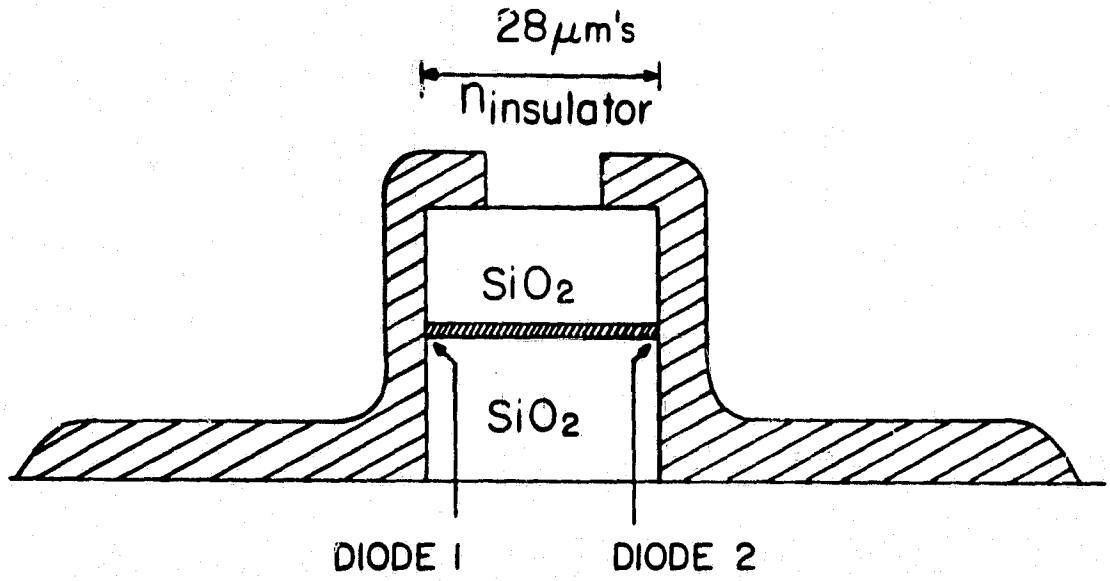
junction. The steps necessary to contact the buried electrode can be eliminated if two junctions are made in series (Fig. V-4). The series array, also forms a natural pillbox structure, which can be designed to be resonant at the frequencies of interest. For the 28 μm mixing experiment two element arrays (Fig. V-4) have been built, with spacings of L = 28μm/n between elements. These boxes should be resonant, at 9.3 μm as well as 28, within the precision of the lithography alignment. Series arrays should improve the overall conversion efficiency of the MOM's.

V-5 Fabrication Procedure for Measuring of the High Frequency Nonlinearity

The procedure described above is for making a single edge MOM diode. The current experiment for measuring the high frequency nonlinearity imposes additional constraints. The 9.3 μm mixing experiment requires efficient input coupling of two laser frequencies, as well as efficient output coupling out of the 20 GHz IF. It is also desirable to be able to construct a variety of combinations of diodes, including arrays, and balanced mixers. The coupling problems and the desireability of multiple junctions has led to the incorporation of a variety of additional processing steps, and to a search to modify somewhat the basic procedure.

The coupling considerations have had three main effects on the construction of the edge MOM devices. The electrode patterns must be designed to maximize the coupling. The devices are built near a preferentially etched edge of the Si wafer and the SiO_2 is being replaced as the insulating layer. These three steps should allow for optimum IF signal generation.

The shaping of the electrodes for good input coupling has been the most difficult design problem. Standard antenna theory can be used to predict receiving patterns at longer wavelengths. At 28 μm and 9.3 μm two problems with this approach begin to manifest themselves. At these shorter wavelengths, the metal does not exhibit ideal loss-less free electron behaviour. Both the displacement current, and the conduction current must be taken into account in calculating any radiation, or receiving patterns, as has already been discussed. The limits of optical lithography are such that it is not possible to make "thin wire" antennas without using e-beam lithography. While theoretical efforts are being made to



V-4 Edge-MOM junction structure showing the buried electrode.

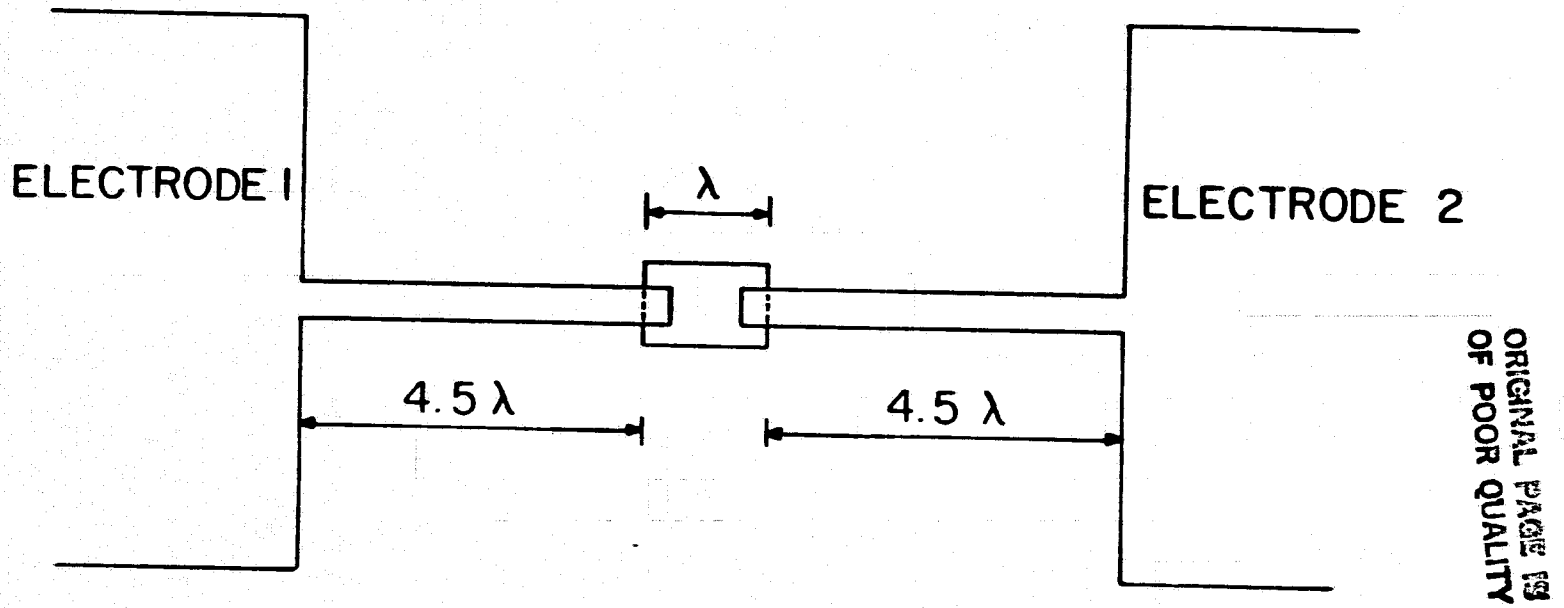
understand these problems, there are, as of this writing, no sure directions for proper design.

This lack of a clear theoretical direction has led us to continue to build long wire antennas (Fig. V-5). At longer wavelengths, where analysis is easier, these show very high gain. There is a need to experiment with other shapes, suggested in another section. These antenna structures must be incorporated inside a 20 GHz IF electrode pattern, which does not radiate.

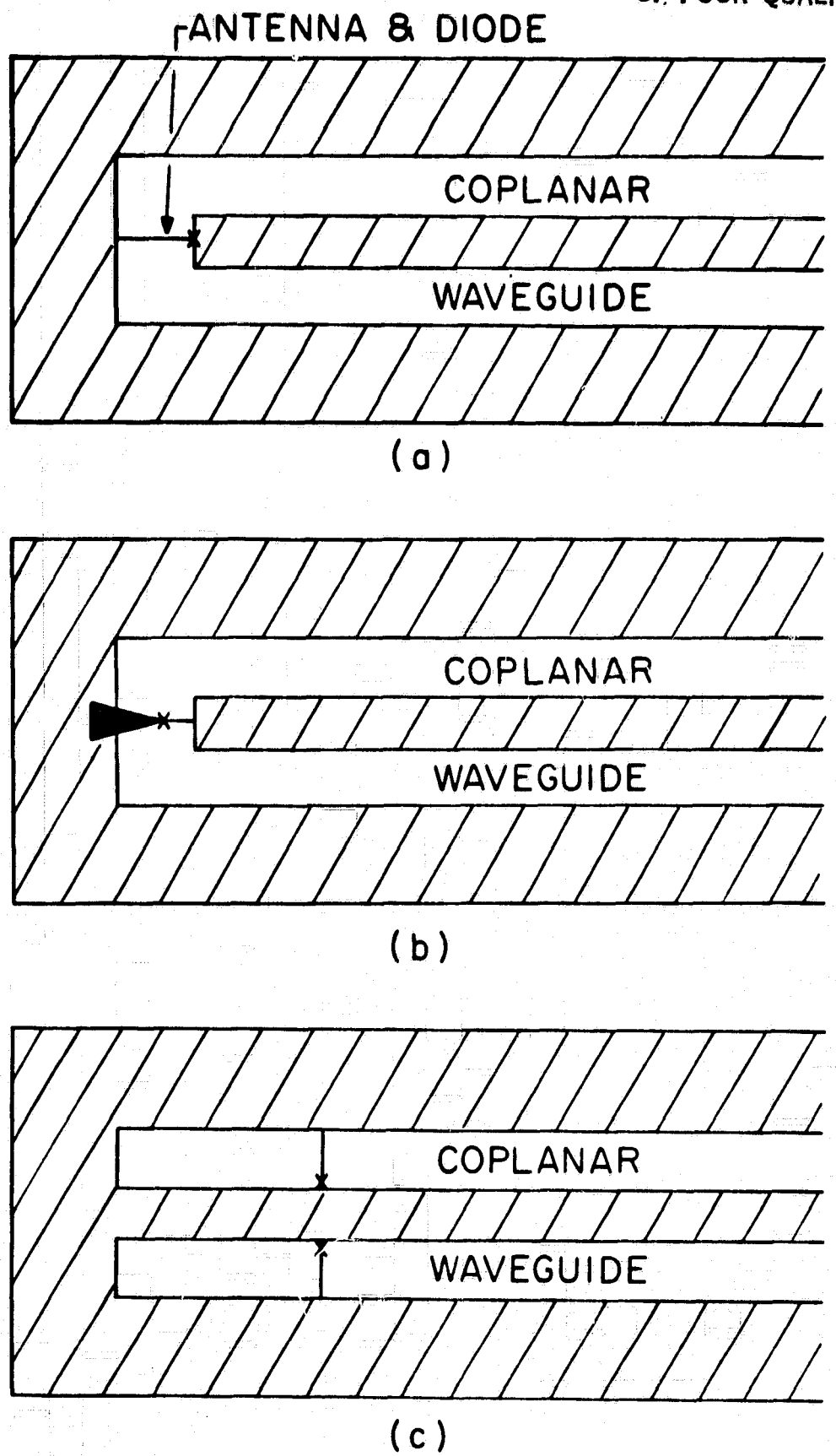
The larger electrode structure is that of a coplanar microwave waveguide (Fig. V-6) [7]. This guide has been designed to have an impedance of 50 Ω's. This waveguide structure offers the possibility of incorporating matching elements directly into the waveguide, to maximize the 20 GHz coupling. This coplanar waveguide will connect to a taper section built on an alumina substrate. The taper changes the dimensions of the waveguide, but not its impedance (Fig. V-7).

This coplanar waveguide structure should couple the 20 GHz IF quite readily into the spectrum analyzer. Input coupling considerations, combined with the desireability of a completely integrated structure have led to the building of these devices above the preferentially etched edge of a Si wafer (Fig. V-8). Calculations show that an integrated antenna receives radiation very poorly from the half plane above the dielectric (air side). In past work a superstrate, Si prism, was placed on top of the antenna. The antenna above a preferentially etched edge allows the substrate to perform the role of the prism and allows one to irradiate the antenna at angles which would be totally neglected from the back of the substrate. The antenna length is chosen to place the beam maximum perpendicular to the preferentially etched edge. The edge is etched prior to the device construction.

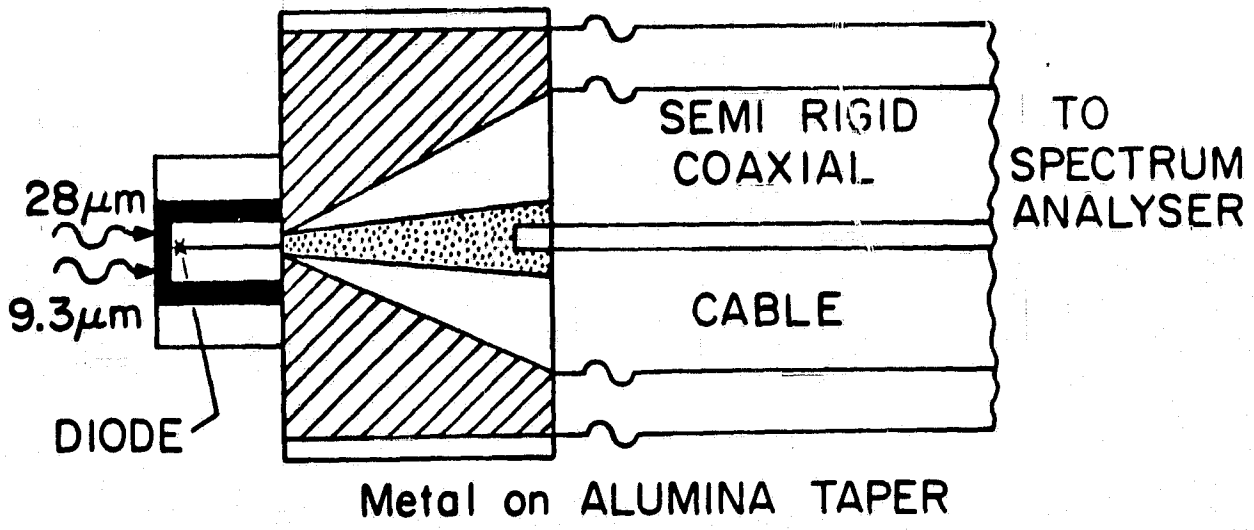
The preferential etching of Si is performed by a standard EDP (Ethelyne Diamine Pyrocatachol) process [8]. The cleaned Si wafer is oxidized in a wet O_2 atmosphere. The atmosphere is changed to N_2 after about 8000 Å of oxide has grown. The oxide is annealed in the N_2 atmosphere for a half an hour, then removed from the oven. The wafer is patterned with photoresist. The SiO_2 is chemically etched in HF, creating the mask for the Si etch step. The photoresist is removed, and the wafer is immersed in a heated solution of EDP. The wafer is left in the EDP until

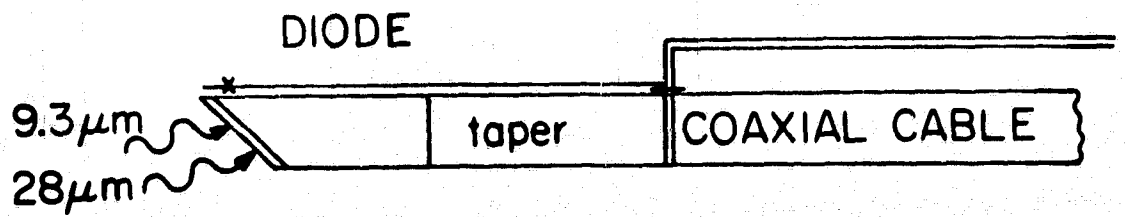


V-5 Top view of series array of two tunneling junctions showing two 4.5 λ long wire antennas on either side of the pill box structure.

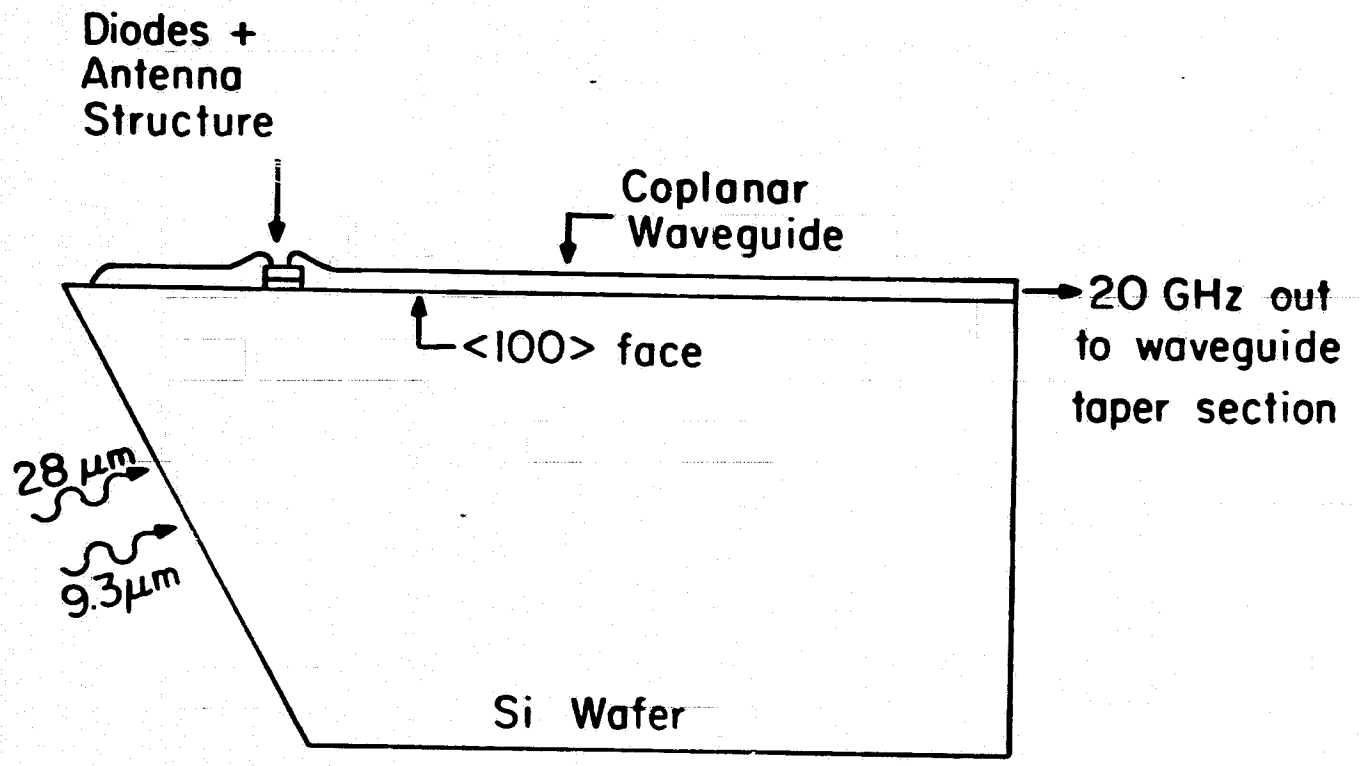


V-6 Three possible configurations for interconnecting the tunneling junctions and the microwave network for the heterodyne diagnostic experiment. The diodes are indicated by X.





V-7 Overall view of coupling scheme between the microwave spectrum analyzer and the tunneling diodes.



V-8 Diodes above a preferentially etched edge of a silicon wafer.

it has been etched through (about four hours). The remaining SiO_2 is now removed with a buffered HF solution. The substrate is then ready for the MOM device to be built on top of it. Having examined the motivation for the electrode shape, and the preferential etching of the substrate, there remains the choice of substrate and insulator best suited for the 28 μm and the 9.3 μm coupling.

There are only two substrate materials which have been found to have the appropriate hardness and solubility requirements for IC processing, while also being transparent to 28 μm and 9.3 μm radiation. These are high resistivity Si, and high resistivity Carbon (diamond). Diamond is both expensive and difficult to preferentially etch or lap. Silicon is, therefore, the substrate of choice to investigate the junction devices. We are using > 700 Ω-cm n-type Si. It is also necessary to choose an insulator which is transparent at 28 μm and 9.3 μm. SiO_2, the insulator we have usually used, absorbs 28 μm light too strongly to be a good choice. Experiments are currently underway to try to eliminate SiO_2 as the insulator.

The most important place to eliminate SiO_2, would be under the antenna structure, since this has a strong adverse effect on both the antenna pattern and the effective area. Eliminating the SiO_2 layer entirely is not feasible. Although the silicon has sufficiently high resistivity to isolate the electrodes, one runs into the previously mentioned surface roughness problem resulting from the ion milling step. A possible way around this is to put down the SiO_2 layer, and then remove it after ion milling. The removal can either be done via chemical (HF) or plasma etching. Plasma etching should be done in a barrel reactor, so that the ions are <u>not</u> accelerated into the substrate. These procedures can be eliminated completely if a more suitable insulator is chosen. A suitable material is Si_3N_4 which is reasonably transparent at both the CO_2 and H_2O laser wavelengths[9]. The sputtering of Si_3N_4 in an r.f. sputtering machine is currently being parametrized.

The resultant junction devices are then designed for a high frequency measurement of the nonlinear mixing characteristics with an I.F. frequency of 20 GHz and an upper frequency corresponding to 10 μm. Various modifications to increase nonlinearity and coupling are expected as a result of these measurements.

References for Chapter V

1. S. Thaniyavarn and T.K. Gustafson, Appl Phys. Lett. 40, 255 (1982).
2. L. Esaki and P.J. Stiles, "New Type of Negative Resistance in Barrier Tunneling," Phys. Rev. Lett. 16, 1108, 1966.
3. D.A. Doane et al, "Semiconductor Technology, A 1982 Symposium," published by Electrochemical Society.
4. J.H. Greiner, "Josephson Tunneling Barriers by RF Sputter-Etching in an Oxygen Plasma," J. Appl. Phys. 42, 5151, 1971.
5. J.H. Greiner, "Oxidation of Lead Films by RF Sputter Etching in an Oxygen Plasma," J. Appl. Phys., 45, 32, 1974.
6. Moredehai Heildum, "Characteristics of Metal-Oxide-Metal Devices." Ph.D. Thesis, University of California.
7. C.P. Wen, "Coplanar Waveguide: A Surface Strip Line," IEEE Trans. on Microwave Theory and Techniques, 17, 1087, 1969, K.C. Gupta, R. Garg, and I.J. Bahl, Microstrip Lines and Slotlines, Artech House Inc., publ., 1979, Chapter 7.
8. K.A. Beam, "Anisotropic Etching of Si," IEEE Trans. on El. Dev., 25, 1185, 1978.
9. Yu. I. Prokhorov, V.A. Sologub, and B.A. Sukhodaev, "Infrared Transmission and Reflectance Spectra of Films Obtained in a High Frequency RF Discharge Plasma," Journ. of Practical Spectroscopy, 19, p. 520.

VI OUTLOOK - ABILITY TO REACH THE THEORETICAL LIMIT

In the present report we have detailed our efforts involved in an investigation of the possibility of utilizing tunneling junction for the direct conversion of the free space radiation field to electron flow. This has involved the theoretical aspects of tunneling using a stationary state theory. This has allowed us to provide a unification of the Tien-Gordon model with other models dealing directly with the interaction within the tunneling region. For practical convertors it has made clear the analogous pictures of multi-photon absorption and Fermi level modulation and has provided us with a quantum mechanical view of the ideal 40% conversion from low frequency rectification. We have seen that the competing processes of multi-photon absorption from the field and dielectric absorption within the bounding materials is a problem which must be dealt with. Experiments on cooled junctions should help clarify some of these aspects.

One of the basic problems with regards to the utilization of tunneling junctions is the fabrication of a diode which does not have appreciable backflow; that is when reverse biased it does not conduct. This seems to run counter to the tunneling phenomena itself and the only way this appears definitively acheivable at the moment is by using the energy gap of a semiconductor as postulated. The current voltage relationship for such a junction should be calculated theoretically to determine whether a sharp enough turn-on of the current can be acheived while retaining a low back flow. There are also some indications that metal-barrier-metal junctions themselves can exhibit a sharp enough turn-on to overcome this.[†]

The surface modes (and the coupling to them) which propagate in metal-barrier-metal structures have been considered in chapters III and IV respectively. We have shown that the coupling between an open surface mode and the junction mode is primarily one of fabrication; that theoretically close to 100% coupling can be expected for ideal tapers. Coupling between the structure modes and the free space radiation field can be accomplished through appropriate antenna design as shown. We have also mentioned an end coupling scheme which becomes possible at high frequencies and which has demonstrated experimentally. Calculations on the latter should be carried out as an extension to the present work.

[†]See C. Slayman, Ph.D. Thesis (U.C. Berkeley, unpublished (1980)), page 30.

The principle problem with surface modes is their loss and particularly short propagation length at visible wavelengths. This was discussed in detail in section (III-3) and has been verified by numerical calculations (III-4). The short propagation length, particularly for the junction gap mode can be alleviated by cooling if the relaxation time is phonon collision dominated. This is true of the metal, which is the most important because of its extremely short relaxation time. Picosecond spectroscopy of thin metallic films at low temperatures would be interesting since it should provide direct evidence that the relaxation time could be extended to at least the picosecond range by cooling. Coupling to surface plasma waves and either propagating the mode or scattering into the free space radiation field could also be used to obtain useful data on losses.

Heavily doped semiconductors should also be investigated since they could possibly be exploited to solve the backflow problem.

In chapter V, we described the processes which we have used to fabricate tunneling devices to be used in our experiments. It is likely that future development of molecular beam epitaxy, chemical vapor deposition, and electron beam deposition will result in highly improved fabrication techniques and will allow much more flexibility in the development of semiconductor barriers.

In concluding this report, one could state that the potential for highly efficient convertors is theoretically present. It is true that in contrast to depletion layer devices which inherently depend upon the internal field to separate the carriers, the upper limit to the efficiency is one hundred percent. The reason that thus far the conversion efficiencies are so low are due primarily to the following:

a) The coupling to the gap mode from a single surface mode has not been definitively demonstrated let alone designed for efficiency, although coupling as well as end coupling and coupling through roughness have been demonstrated.

b) The junction current-voltage characteristic has displayed a highly symmetrical behaviour about the zero-bias point for virtually all tunneling junctions fabricated thus far. Thus the back flow has generally been a problem.

c) The absorption length due to multiphoton absorption through the gap mode is comparable to or greater than the propagation length of the mode in the absence of the multi-photon absorption.

Theoretically, perhaps the most crucial problem at this point is a calculation of the current-voltage characteristics of hypothetical junction configurations which could demonstrate more ideal rectification behaviour than the metal-barrier-metal junction. Eqs. (II.5.6 and II.5.7) could then be used to estimate the absorption length and possible efficiencies which might be realized. We have made a start by considering the metal-barrier-heavily doped semiconductor device; however this needs to be extended to include a further metallic boundary on the semiconductor side.

APPENDIX I PROGRAM FOR CALCULATING THE DISPERSION CHARACTERISTICS OF THE SURFACE ELECTROMAGNETIC WAVE GAP MODE

Input parameters are:

xss - Normalized freqency ω/ω_s where ω_s is the plasma frequency of the heavily doped semiconductor

e(1), f(1) - dielectric constant of the film is $\varepsilon_0(e(1)+jf(1))$

es - dielectric constant of film as $\omega \to \infty$

xs - $1/\omega_s\zeta_s$

ws - $\omega_s/\omega_s = 1$

em - dielectric constant of film at $\omega \to \infty$

xm - $1/\omega_m\zeta_m$

wm - ω_m/ω_s

s - barrier thickness

The most important output is dm, dp, rk, rkk, xss and s.

dm - dm = 0 when a solution is obtained.

dp - n(2π) for a solution (phase condition)

rk - Re $\beta/\omega\sqrt{\mu_0\varepsilon_0}$

rkk - Im $\beta/\omega\sqrt{\mu_0\varepsilon_0}$

Program Disp.f. Notation

$k_0 = \omega_0\sqrt{\mu_0\varepsilon_0}$ = po

= 18878.3 λ_0 = 3.3383 μm ω_0 = 5.6635×10^{14}

β = rk + jrkk

xss = (ω/ω_0) x2 = $(xss)^2$

ts ≤ 0 → test {tr = 0 → 1 / or else → 2
> 0 → 3

$\varepsilon(1) = \varepsilon_b$ = (e(1)+jf(1))

$\varepsilon(2) = \varepsilon_{sc}$ = (e(2)+jf(2))es

$\varepsilon(3) = \varepsilon_M$ = (e(3)+jf(3))em

sk(1), sk(2), sk(3) change

xs = $1/\zeta_s\omega_0$

xm = $1/\zeta_m\omega_0$

ws = ω_{ps}/ω_0

wm = ω_{pm}/ω_0

Normalization on k $\quad \gamma = \sqrt{k^2 - \omega^2\mu\varepsilon}$

$$\frac{\gamma^2}{\omega^2\mu\varepsilon_0} = \sqrt{\frac{k^2}{\omega^2\mu\varepsilon_0} - \frac{\varepsilon}{\varepsilon_0}}$$

Controls

$$\frac{\gamma^2}{\omega_0^2\mu\varepsilon_0} = \sqrt{\frac{k^2}{\omega_0^2\mu\varepsilon_i} - \frac{\varepsilon}{\varepsilon_0}\left(\frac{\omega}{\omega_0}\right)^2}$$

is < 0 write

x, yy, ε(i), f(i)

↑ Reγ^2 ↑ Imγ^2

sk(i) < 0 change b(i) → π-b(i) $\quad (\overset{\uparrow}{rk}+jrkk)xss = \dfrac{k}{\omega_0^2\sqrt{\mu\varepsilon_0}} = \dfrac{k}{po}$

= 0 test for x < 0

> 0 change b(i) → π - b(i) $\quad \therefore (rk+jrkk) = \left(\frac{k}{po}\right)\left(\frac{\omega_0}{\omega}\right)$ → rk+jrkk normalized to k(ω)

a(i) = [Amp($k^2-\omega^2\mu\varepsilon$)]/(po) $\quad$ 1 → barrier 2) s.c. 3) metal

b(i) = Arc($k^2-\omega^2\mu\varepsilon$)

g(i) = Amp ε(i) $\quad$ h(i) = Arc ε(i)

pp' = < ε(i) - < ε(i) - (< γ(i)-< γ(i))

bb = (amp ε(i)/amp ε(1)) x Amp(γ(1)/amp γ(i))

z(j) = 1 + $(bb)^2$ + s(bb) x cos(pp) = squared ampl. of $1+\frac{\varepsilon(i)}{\varepsilon(1)}\frac{\gamma(1)}{\gamma(i)}$

y(j) = squared ampl. of $1-\frac{\varepsilon(i)}{\varepsilon(1)}\frac{\gamma(1)}{\gamma(i)}$

rq(j') = $\sqrt{(1+\frac{\varepsilon(i)}{\varepsilon(1)}\frac{\gamma(1)}{\gamma(i)})}$

r(j) = $\sqrt{(1-\frac{\varepsilon(i)}{\varepsilon(1)}\frac{\gamma(1)}{\gamma(i)})}$

d = 2 a(1) x S ← Normalized to k_0 (dk_0) = S

cb = cos(b(1)) $\quad$ sb = sin(b(1)) $\quad$ dm → ampl. cond. dp → phase

PRINTOUT

Z S(1), Z S(2), y(1), y(2), z(1), z(2)

↑ Amp squared of dnff $\quad$ ↑ sum

ak1, ak2, ak3, bk1, bk2, bk3

Im.(γ_1,γ_2,γ_3) Re(γ_1,γ_2,γ_3)

```
      program Disp
      dimension sk(3),zs(3)
      dimension e(5),f(5),a(5),b(5),g(5),h(5),r(5),rq(5),z(5),y(5)
      double precision xss,x2,dp,es,em,xs,rks,ws,xm,rkm,wm,s
      double precision e(1),e(2),e(3),f(1),f(2),f(3),rk,rkk,sy
      double precision u,v,x,yy,sa,a(1),a(2),a(3),b(1),b(2),b(3)
      double precision g(1),g(2),g(3),h(1),h(2),h(3),p,bb,hh
      double precision z(1),z(2),z(3),y(1),y(2),y(3),d,cb,sb,dm,pp
      double precision cb2,sb2,cb3,sb3,ak1,ak2,ak3,bk1,bk2,bk3
      double precision rq(1),rq(2),r(1),r(2),pi,ss,c,po,e1,e2
      po=18878.3
      pi=3.1415926536
      rk=1.7
      rkk=.225
      xss=.1
      x2=.01
      e1=3.2
      f1=0.0
      es=11.1
      xs=1/sqrt(32.5)
      ws=1.
      xm=sqrt(22.8/32.5/1.54)
      wm=sqrt(2518.*22.8/32.5/1.54)
      em=1.
      s=9.*5.6635/3./10000.
    1 continue
      write (*,'("enter xss,e(1),f(1),es,xs,ws,em,xm,wm,s",$)')
      read(*,*) xss,e1,f1,es,xs,ws,em,xm,wm,s
      write(*,*) xss,e1,f1,es,xs,ws,em,xm,wm,s
    3 continue
      write (*,'("enter sk(1),sk(2),sk(3)",$)')
      read(*,*) sk(1),sk(2),sk(3)
      write(*,*) sk(1),sk(2),sk(3)
    2 continue
      write (*,'("enter rk and rkk,tr,ts,is ",$)')
      read(*,*) rk,rkk,tr,ts,is
      rk=rk*xss
      rkk=rkk*xss
      x2=xss*xss
      rks=xs*xs
      rkm=xm*xm
      e(1)=e1
      f(1)=f1
      e(2)=es*(1.-(ws*ws)/(x2+rks))
      f(2)=es*(xs*ws*ws)/(x2+rks)/xss
      e(3)=em*(1.-(wm*wm)/(x2+rkm))
      f(3)=em*(xm*wm*wm)/(x2+rkm)/xss
      u=rk*rk-rkk*rkk
      v=2.*rk*rkk
      do 10 i=1,3
      x=u-e(i)*x2
      yy=v-f(i)*x2
      sa=dsqrt(x*x+yy*yy)
      a(i)=dsqrt(sa)
      b(i)=dasin(yy/sa)
      if(is)78,79,79
   78 write(*,*) x,yy,e(i),f(i)
   79 continue
      if(sk(i)) 7,9,8
    9 if(-) 5,6,6
    5 if(b(i))7,8,8
    7 b(i)=-pi-b(i)
      go to 6
```

```
      g(i)=dsqrt(e(i)*e(i)+f(i)*f(i))
      h(i)=dasin(f(i)/g(i))
      if(e(i))30,31,31
   30 h(i)=pi-h(i)
   31 continue
   10 continue
      do 20 j=1,2
      k=j+1
      pp=h(k)-h(1)-(b(k)-b(1))
      ss=dsin(pp)
      c=dcos(pp)
      bb=(g(k)/g(1))*a(1)/a(k)
      zs(j)=bb
      z(j)=1.+bb*bb+2.*bb*c
      y(j)=1.+bb*bb-2.*bb*c
      hh=bb*ss
      sz=dsqrt(z(j))
      rq(j)=dasin(hh/sz)
      if(1.+bb*c)49,48,48
   49 if(hh)46,47,47
   47 rq(j)=pi-rq(j)
      go to 48
   46 rq(j)=-pi-rq(j)
   48 continue
      sy=dsqrt(y(j))
      r(j)=dasin(-hh/sy)
      if(1.-bb*c) 53,52,52
   53 if(-hh)54,51,51
   54 r(j)=-pi-r(j)
      go to 52
   51 r(j)=pi-r(j)
   52 continue
   20 continue
      d=2.*a(1)*s
      cb=dcos(b(1))
      sb=dsin(b(1))
      dm=z(1)*z(2)*dexp(2.*d*cb)-y(1)*y(2)
      dp=rq(1)+rq(2)-r(1)-r(2,+d*sb
      rk=rk/xss
      rkk=rkk/xss
      dp=dp/pi
      write (*,*) dm,dp,rk,rkk,xss,s
      if (is)57,58,58
   57 write (*,*) zs(1),zs(2),y(1),y(2),z(1),z(2),d
      write (*,*) a(1),b(1),a(2),b(2),a(3),b(3)
      cb2=dcos(b(2))
      sb2=dsin(b(2))
      cb3=dcos(b(3))
      sb3=dsin(b(3))
      ak1=a(1)*sb
      ak2=a(2)*sb2
      ak3=a(3)*sb3
      write(*,*) ak1,ak2,ak3
      bk1=a(1)*cb
      bk2=a(2)*cb2
      bk3=a(3)*cb3
      c2=e(1)*(e(2)/g(2)/g(2)+rkk*f(2)/g(2)/g(2)/rk)/bk2/s
      c3=e(1)*(e(3)/g(3)/g(3)+rkk*f(3)/g(3)/g(3)/rk)/bk3/s
      c=1./(1.+c2+c3)
      rke=rk/(rk*rk+rkk*rkk)
      rke=rke*e(1)
      write(*,*) bk1,bk2,bk3
      write(*,*) c,c2,c3,rke
```

APPENDIX II THE ANTI-SYMMETRIC GAP MODE

Using Fig. (III-7) and $\varepsilon_{m,s} = \varepsilon_\infty(1 - \frac{(\omega_{m,s})^2}{\omega^2+i\omega/\xi_{m,s}})$ one can represent the fields of the gap mode as

$S(x \geq \frac{d}{2})$	$b(-\frac{d}{2} \leq x \leq \frac{d}{2})$	$m(x \leq -\frac{d}{2})$
H_y: $H_S e^{\nu_S(x-\frac{d}{2})}$	$H_b \cosh(+\nu_b(x+\frac{d}{2})-\phi_s)$	$H_m e^{-\nu_m(x+\frac{d}{2})}$
E_x: $\frac{\beta H_S}{\omega\varepsilon_S} e^{\nu_S(x-\frac{d}{2})}$	$\frac{\beta H_b}{\omega\varepsilon_b} \cosh(\nu_b(x+\frac{d}{2})-\phi_S$	$\frac{\beta H_m}{\omega\varepsilon_m} e^{-\nu_c(x+\frac{d}{2})}$
E_z: $\frac{i\nu_S H_S}{\omega\varepsilon_S} e^{\nu_S(x-\frac{d}{2})}$	$\frac{i\nu_b H_b}{\omega\varepsilon_b} \sinh(\nu_b(x+\frac{d}{2})-\phi_S)$	$\frac{-i\nu_m H_m}{\omega\varepsilon_m} e^{-\nu_c(x+\frac{d}{2})}$
$\nu_S = (\beta^2-\omega^2\mu\varepsilon_S)^{\frac{1}{2}}$	$\nu_b = (\beta^2-\omega^2\mu\varepsilon_b)^{\frac{1}{2}}$	$\nu_m = (\beta^2-\omega^2\mu\varepsilon_m)^{\frac{1}{2}}$

Continuity of Tangential $\bar{E}$ and $\bar{H}$ at x = d and x = -d gives:

$$H_S = H_b\cosh(\nu_b d-\phi_S) \qquad (a)$$

$$H_m = H_b\cosh(+\phi_S) \qquad (b)$$

$$\frac{\nu_s H_s}{\varepsilon_s} = \frac{\nu_b H_b}{\varepsilon_b}\sinh(-\phi_S) \qquad (c)$$

$$\frac{\nu_b H_b}{\varepsilon_b}\sinh(\nu_b d-\phi_S) = -\frac{\nu_m H_m}{\varepsilon_m} \qquad (d)$$

(a) and (d) give

$$\frac{\nu_b}{\varepsilon_b}\tanh(+\nu_b d-\phi_S) = \frac{\nu_m}{\varepsilon_m} \qquad (e)$$

while (b) and (c) give

$$\frac{\nu_S}{-\varepsilon_S}\frac{\varepsilon_b}{\nu_b} = \tanh\phi_S \qquad (f)$$

(e) can be rewritten as

$$2\nu_b d = +\phi_m + \phi_S$$

where we define ϕ_m by

$$\phi_m = \tanh^{-1} \frac{\varepsilon_b}{(-\varepsilon_m)} \frac{\nu_m}{\nu_b}$$

It is straightforward to evaluate the power propagated in the mode P

$$P = \tfrac{1}{2}\mathrm{Re} \iint (\bar{E}\times\bar{H}^*)\cdot\bar{n}dS$$

$$= \tfrac{1}{2}E_b H_b^* L_y x(d) x C$$

where E_b and H_b are the x and y component electric and magnetic field amplitudes respectively. C is the confinement factor given by

$$C = \mathrm{Re}(\beta/(\omega\varepsilon_b)/\mathrm{Re}(\beta/\omega[1/\varepsilon_b + \frac{1}{\varepsilon_m(\nu_m+\nu_m^*)d} + \frac{1}{\varepsilon_S(\nu_S+\nu_S^*)d}])$$

$$\mathrm{Re}(\omega/\beta)/\upsilon_E$$

where υ_E is the energy velocity = c/S, S being the slowing factor.

For a symmetric structure $\phi_m = \phi_S$ and thus $\tanh(\nu_d d) = \frac{\varepsilon_b}{\nu_b}\frac{\nu_m}{(-\varepsilon_m)}$ or Eq. (III.2.8).