A Thesis

entitled

Optical Bistability

with

Surface Plasmons

by

Gary M. Wysin

as partial fulfillment of the requirements of
the Master of Science Degree
in Physics

1980

ACKNOWLEDGEMENTS

I am deeply appreciative and would like to thank Dr. R. T. Deck and Dr. H. J. Simon for the help offered to me in the preparation of this thesis, and to my Mother for her assistance with the typing.

TABLE OF CONTENTS

I. INTRODUCTION

Recently there has been interest in various optical phenomena associated with a nonlinear dielectric medium whose permittivity is dependent on the electric field intensity; especially with the effects it can have on the reflection and refraction of light at a boundary between a linear medium and a nonlinear medium.[1,2] One of the major effects referred to here is the existence of intensity dependent optical bistability at such a boundary; it is possible to have two stable solutions for the electric fields for a given applied light beam, under certain special conditions. The boundary can be made to switch abruptly from one state (a solution for the fields) to the other by a small change in incident light wave intensity; the observed effect of this switching of states is a discontinuous jump in the reflection coefficient of the boundary. Hysteresis effects have also been shown to be related,[1] and in fact are a result of this bistability.

It is intended here to investigate some of these facts, making use of Snell's law and the Fresnel reflection and transmission coefficients for the boundary. A method will be given for finding the reflectivity of the boundary as a function of intensity, and it will be general enough so that it is applicable for any values of the linear and nonlinear permittivities of the two media, and therefore for critical angles which are not necessarily grazing angles, which was the case previously discussed by Kaplan.[1] The advantage of this generality comes in the later half of the paper, which discusses the behavior of these nonlinear

effects when a third medium, a metal, is "sandwiched" between the linear and nonlinear media, in the configuration necessary for the excitation of surface plasmons (Kretschmann configuration). Surface plasmon waves have been widely investigated [3,4], and have been shown to induce decreases in reflectivity related to the enhanced absorption of energy by the metal under the conditions necessary for surface plasmons [3]. Surface plasmons do not require nonlinearities in the system in order to be excited, but it is the intention here to see how nonlinearities affect their production.

The purpose here is to discuss the possibility of bistable effects related to the excitation of these surface plasmon waves. It will be shown that these effects are expected to occur both for the case of positive (nonlinear permittivity increases as intensity increases) nonlinearity and for negative (nonlinear permittivity decreases as intensity increases) nonlinearity. Discontinuous changes in reflectivity can be expected for this case, in a manner similar to that obtained without the metal present, but for incident angles which are close to the plasmon angle, the angle at which the surface plasmon made would be most strongly excited. Because of the variable permittivity in the nonlinear medium, this plasmon angle is not fixed, but varies with intensity. It is usually, however, an angle which is not near grazing incidence, and this may afford the experimentalist some easing of any problems associated with working at incident angles near 90°. The generality of the calculational method used here allows the computations to be performed for any incident angle; this is its advantage over previous methods referred to above.

The paper will be divided into three major sections; the first is a

short discussion of nonlinear polarization, and the related effects this has on the permittivity and index of refraction of a nonlinear medium, the second is a derivation and explanation of the method used to find the reflectivity as a function of the intensity of an incident light beam, for a linear-nonlinear boundary, and the third treats the case of optical bistability near the plasmon angle, with the combined effects of the nonlinearity and the surface plasmon excitation. The major theme of this paper is to discuss the surface plasmon bistability; the linear-nonlinear boundary case is used only as an example to demonstrate how the calculation can be performed, and because the algebra involved is simpler.

II. SIMPLE NON-LINEAR POLARIZATION

Usually the dielectric constant ε in any material is taken to be a constant, independent of the strength of the electromagnetic field within the material. This is derived from the fact that the displacement vector $\vec{D}$ can usually be taken to be linearly related to the electric field $\vec{E}$, and through the definition of ε

$$\vec{D} = \varepsilon \vec{E} \tag{1}$$

it then results that ε must be a constant. The displacement $\vec{D}$, however, is actually defined in terms of the polarization of the material, denoted $\vec{P}$, defined as the net dipole moment per unit volume of the material. $\vec{D}$ then depends on both $\vec{E}$ and $\vec{P}$ through the relation

$$\vec{D} = \vec{E} + 4\pi\vec{P} = \varepsilon \vec{E} \tag{2}$$

Now any microscopic model of the material would necessarily produce a relation between the "averaged" polarization $\vec{P}$ and the "averaged" total electric field $\vec{E}$, and in general the relation obtained could allow $\vec{P}$ to be expanded as a power series in $\vec{E}$. Writing out the components of $\vec{P}$, using the summation convention on repeated indices

$$P_i = \chi^1_{ij} E_j + \chi^2_{ijk} E_j E_k + \chi^3_{ijkl} E_j E_k E_l + \ldots \tag{3}$$

where $i = 1, 2, 3$, and the χ's are coefficients in the series expansion which are independent of $\vec{E}$. The above is the general expression for any anisotropic medium, if we specialize to the case of an isotropic medium, then the polarization can be written in the form

$$\vec{P} = \chi_1 \vec{E} + \chi_3 |E|^2 \vec{E} \tag{4}$$

And using the definition of ε, through $\vec{D}$

$$\vec{D} = (1 + 4\pi\chi_1 + 4\pi\chi_3|E|^2)\vec{E} = \varepsilon\vec{E} \tag{5}$$

It then results that the dielectric "constant" can be written

$$\varepsilon = 1 + 4\pi\chi_1 + 4\pi\chi_3|E|^2 = \varepsilon^0 + \alpha|E|^2 \tag{6}$$

Also, since the index of refraction of the material n is related to ε, it follows that n also depends on the strength of the electric field. From the definition of the index of refraction of a non-magnetic material, put in terms of the wave speed in the medium:

$$v = \frac{c}{\sqrt{\varepsilon\mu}} = \frac{c}{\sqrt{\varepsilon}} = \frac{c}{n} \tag{7}$$

then

$$n = \sqrt{\varepsilon} = \sqrt{\varepsilon^0 + \alpha|E|^2} \tag{8}$$

and since in general any nonlinearity in the medium will be small in some sense, the index n can be written approximately from this expression as

$$n = \sqrt{\varepsilon^0} + \frac{\alpha}{2\sqrt{\varepsilon^0}}|E|^2 \tag{9}$$

Another common way to write n is as the sum of a linear term and a non-linear term in the form

$$n = n^0 + n_{NL}|E|^2 \tag{10}$$

In general both n and ε would have both a real and an imaginary part, and therefore so would the coefficients α and n_{NL}. The real part is related to nonlinear dispersion of the medium, while the imaginary part relates to the damping of the field due to nonlinear absorption.

III. OPTICAL BISTABILITY AT A LINEAR TO NONLINEAR BOUNDARY

First the behavior of a boundary between a linear glass medium and a nonlinear (Kerr effect) medium will be investigated with respect to dependence of the optical reflectivity on the intensity of a wave incident from the linear medium. The geometrical set up is shown in figure 1. A plane boundary is assumed, with the z-axis perpendicular to the boundary and the x-axis along the boundary, consequently the y-axis is perpendicular to the page, pointing out of the paper. The dielectric constant ε in the nonlinear medium has the previously shown dependence on the electric field

$$\varepsilon_2 = \varepsilon_2^0 + \alpha|E|^2 \tag{11}$$

The angle of incidence for the incoming wave is θ_1, with the wave assumed to be polarized in the plane of incidence, the case of P-polarization. The amplitude of the incident wave is E_1, while those of the reflected and transmitted waves are E_{1r} and E_2 respectively. Note that θ_1 is always a real angle, while the transmission angle θ_2 can be imaginary in the case of total internal reflection. We treat here the case of no absorption.

A. Assumed Forms For Waves In The Two Media

A plane wave solution is assumed for both the incident and reflected waves, with $e^{-i\omega t}$ time dependence assumed for all field amplitudes, and the boundary taken to be at $z = 0$. The wave in the nonlinear medium may not be an ordinary plane wave, a solution is assumed of the form

$$\vec{E}_2(\vec{x},t) = \vec{E}_2(z)e^{ik_{2x}x}e^{i\int_o^z k_{2z}(z')dz'} \tag{12}$$

The associated magnetic field can be found using Faraday's Law

$$\vec{\nabla} \times \vec{E}_2 = \frac{i\omega}{c}\vec{H}_2 \tag{13}$$

From which it follows that the y-component of H_2 is given by

$$H_{2y} = -\frac{ic}{\omega}\left(\frac{\partial E}{\partial z} - \frac{\partial E}{\partial x}\right) \tag{14}$$

Finally, all of the E and H fields in the two media can be written out in terms of their respective x, y, z components for P-polarization, as follows:

Incident Wave

$$E_{1x} = E_1\cos\theta_1 e^{i(k_{1x}x + k_{1z}z)} \qquad H_{1x} = 0$$

$$E_{1y} = 0 \qquad H_{1y} = \frac{c}{\omega} k_1 E_1 e^{i(k_{1x}x + k_{1z}z)} \tag{15}$$

$$E_{1z} = -E_1\sin\theta_1 e^{i(k_{1x}x + k_{1z}z)} \qquad H_{1z} = 0$$

Reflected Wave

$$E_{1rx} = -E_{1r}\cos\theta_1 e^{i(k_{1x}x - k_{1z}z)} \qquad H_{1rx} = 0$$

$$E_{1ry} = 0 \qquad H_{1ry} = \frac{c}{\omega} k_1 E_{1r} e^{i(k_{1x} - k_{1z}z)} \tag{16}$$

$$E_{1rz} = -E_{1r}\sin\theta_1 e^{i(k_{1x}x - k_{1z}z)} \qquad H_{1rz} = 0$$

Transmitted Fields

$$E_{2x} = E_2(z)\cos\theta_2 e^{ik_{2x}x} e^{i\int_o^z k_{2z}dz'} \qquad H_{2x} = 0$$

$$E_{2y} = 0 \qquad H_{2y} = \frac{c}{\omega}\left(k_2 E_2(z) - i\cos\theta_2 \frac{dE_2}{dz}\right) e^{ik_{2x}x} e^{i\int_o^z k_{2z}dz'} \tag{17}$$

$$E_{2z} = -E_2(z)\sin\theta_2 e^{ik_{2x}x} e^{i\int_o^z k_{2z}dz'} \qquad H_{2z} = 0$$

In the above field expressions, the wave vectors k_1 and k_2 are determined through use of the wave equation, with the relations

$$k_1 = \frac{\omega}{c}\sqrt{\varepsilon_1} \qquad k_2 = \frac{\omega}{c}\sqrt{\varepsilon_2} \tag{18}$$

In a more exact treatment, the form assumed for the wave in the nonlinear medium may not be correct, and consequently the expression for k_2 may also be invalid, in which case the nonlinear wave equation should be solved,

$$-\nabla^2 \vec{E}_2 + \vec{\nabla}(\vec{\nabla} \cdot \vec{E}_2) - \frac{\omega^2}{c^2}(\varepsilon_2^0 + \alpha|E_2|^2)\vec{E}_2 = 0 \qquad (19)$$

giving a more valid form for the fields in the nonlinear medium. Here, however, the nonlinear wave equation will not be solved, but the approximate expressions for the fields will be used.

The components of the wave-vectors are given by

$$\begin{aligned} K_{1x} &= K_1 \sin\theta_1 & K_{1z} &= K_1 \cos\theta_1 \\ K_{2x} &= K_2 \sin\theta_2 & K_{2z} &= K_2 \cos\theta_2 \end{aligned} \qquad (20)$$

B. Application of Boundary Conditions

For the situation under consideration, application of the continuity of tangential $\vec{E}$ and tangential $\vec{H}$ across the boundary leads to the usual Fresnel reflection and transmission coefficients, but with one additional assumption along the way. Writing the equation for tangential $\vec{E}$ gives

$$E_1 \cos\theta_1 e^{iK_{1x}x} - E_{1r} \cos\theta_1 e^{iK_{1x}x} = E_2(0) \cos\theta_2 e^{iK_{2x}x} \qquad (22)$$

Writing the equation for tangential $\vec{H}$ gives

$$K_1 E_1 e^{iK_{1x}x} + K_1 E_{1r} e^{iK_{1x}x} = [K_2 E_2(0) - i \cos\theta_2 \frac{dE_2}{dz}] e^{iK_{2x}x} \qquad (22)$$

Now it must be assumed that the derivative in this equation is small compared to the other terms in order to obtain the usual Fresnel factors. This is probably not a bad assumption in the case of transmission of the wave into the medium for here it would be expected that the amplitude of the field

changes slowly with respect to position in the material. In the case of total internal reflection (TIR), however, it is not evident that this is a reasonable approximation, and again a more exact treatment would involve first solving the nonlinear wave equation for the fields in the nonlinear medium, and then calculating the required derivative and inserting it into the above equation, and thereby solving the resulting equations for the reflection and transmission coefficients. The simpler (approximate) method will be used here, and the derivative $\frac{dE_2}{dZ}$ will be taken to be zero, evaluated at the boundary $Z = 0$.

With the above assumption, E_1 , E_{1r} , and $E_2(o)$ are related by the Fresnel factors[5] r_{12}, t_{12} according to

$$E_{1r} = r_{12}E_1 = \left(\frac{K_2 \cos\theta_1 - K_1 \cos\theta_2}{K_2 \cos\theta_1 + K_1 \cos\theta_2}\right) E_1$$

$$E_2(0) = t_{12}E_1 = \left(\frac{2K_1 \cos\theta_1}{K_2 \cos\theta_1 + K_1 \cos\theta_2}\right) E_1 \qquad (23)$$

These can also be written in an alternative but equivalent form

$$E_{1r} = \left(\frac{\varepsilon_2 K_{1z} - \varepsilon_1 K_{2z}}{\varepsilon_2 K_{1z} + \varepsilon_1 K_{2z}}\right) E_1$$

$$E_2(0) = \left(\frac{2\sqrt{\varepsilon_1\varepsilon_2}\, K_{1z}}{\varepsilon_2 K_{1z} + \varepsilon_1 K_{2z}}\right) E_1 \qquad (24)$$

C. Calculation of Reflectivity $R = \dfrac{|E_{1r}|^2}{|E_1|^2}$

Now it is useful to write these expressions in terms of only the angle of incidence θ_1 and the dielectric constants in the two media. This can be effected by writing K_{2z} in terms of θ_1 , via the relation

$$K_{2x}^2 + K_{2z}^2 = K_2^2 = \frac{\omega^2}{c^2}\ \varepsilon_2 \qquad (25)$$

Also making use of Snell's Law, which is valid here by the boundary

condition equations, in the form

$$K_{1x} = K_{2x} = \frac{\omega}{c}\sqrt{\varepsilon_1}\sin\theta_1 \tag{26}$$

Then in the case of transmission of the wave, where K_{2z} is real, there results

$$K_{2z} = \frac{\omega}{c}\sqrt{\varepsilon_2 - \varepsilon_1 \sin^2\theta_1} \tag{27}$$

This is real provided $\varepsilon_2 > \varepsilon_1 \sin^2\theta_1$. But since ε_2 depends on the intensity of the wave, it is evident that the critical angle for total internal reflection, which is determined through

$$\varepsilon_2 = \varepsilon_1 \sin^2\theta_{critical} \tag{28}$$

will depend on the intensity of the incident light wave, and consequently it appears possible to optically switch from TIR to transmission, or in reverse, simply by changing the incident power. A further consequence is that the situation exhibits hysteresis effects, in that the power level necessary to switch from TIR to transmission is not the same as that necessary to go from transmission to TIR. This second effect will be made evident in what follows.

For the case of TIR, the z-component of K_2 is pure imaginary, and can be written

$$K_{2z} = i\frac{\omega}{c}\sqrt{\varepsilon_1 \sin^2\theta_1 - \varepsilon_2} \tag{29}$$

In this case $\varepsilon_1 \sin^2\theta_1 > \varepsilon_2$, and the wave amplitude decreases exponentially with distance measured from the boundary.

The ratio of the reflected to the incident wave intensity is found by calculating the quantity R, equal to the reflectivity,

$$R = \frac{|E_{1r}|^2}{|E_1|^2} \tag{30}$$

In the usual linear case, R is given simply as the absolute square of the Fresnel reflection, coefficient r_{12} . This does not work in this nonlinear case, however, since r_{12} depends on ε_2 , which depends on $|E_2|^2$, which is in turn dependent on the incident amplitude E_1 . Thus it is evident that R will be a function of the incident wave amplitude. It is then required to find the reflectivity R as a function of E_1 . This could be accomplished by eliminating E_2 from the equations (24) by first solving the second for E_2 as a function of E_1 , and then substituting the resulting function into the first equation (for E_{1r}). Thus E_{1r} would be obtained as a function of only E_1 ; from this R could then be determined. An equivalent and perhaps more tractable way to determine R will be explained in what follows, it is more useful especially since the equation $E_2(0) = t_{12} E_1$ cannot be very easily solved for $E_2 = f(E_1)$ in closed algebraic form.

The two Fresnel equations under consideration can be solved for E_{1r} and E_1 as functions of E_2 :

$$E_1 = E_2(0)\left(\frac{\varepsilon_2 K_{1z} + \varepsilon_1 K_{2z}}{2\sqrt{\varepsilon_1\varepsilon_2}\, K_{1z}}\right) = E_2(0)\left[\frac{\sqrt{\varepsilon_2}\cos\theta_1 + \sqrt{\varepsilon_1}\sqrt{1 - \frac{\varepsilon_1}{\varepsilon_2}\sin^2\theta_1}}{2\sqrt{\varepsilon_1}\cos\theta_1}\right]$$

$$E_{1r} = E_2(0)\left(\frac{\varepsilon_2 K_{1z} - \varepsilon_1 K_{2z}}{2\sqrt{\varepsilon_1\varepsilon_2}\, K_{1z}}\right) = E_2(0)\left[\frac{\sqrt{\varepsilon_2}\cos\theta_1 - \sqrt{\varepsilon_1}\sqrt{1 - \frac{\varepsilon_1}{\varepsilon_2}\sin^2\theta_1}}{2\sqrt{\varepsilon_1}\cos\theta_1}\right] \qquad (31)$$

Now the explicit form for $\varepsilon_2 = \varepsilon_2^0 + \alpha|E_2|^2$ can be inserted, and writing E_2 for $E_2(0)$, the equations can be absolute squared to change them to intensity equations. Finally a set of dimensionless intensity variables can be defined as follows,

$$\text{Dimensionless incident intensity } U_1 \equiv \frac{\alpha}{\varepsilon_2^0}|E_1|^2$$

$$\text{Dimensionless reflected intensity } U_{1r} \equiv \frac{\alpha}{\varepsilon_2^0}|E_{1r}|^2 \qquad (32)$$

$$\text{Dimensionless transmitted intensity } U_2 \equiv \frac{\alpha}{\varepsilon_2^0}|E_2|^2$$

Also define $\varepsilon \equiv \frac{\varepsilon_2^0}{\varepsilon_1} = \left(\frac{n_2^0}{n_1}\right)^2 = \sin^2\theta_c^0$ (33)

Where θ_c^0 is the linear critical angle, the critical angle if no nonlinearity were present ($\alpha = 0$) .

In terms of these dimensionless variables, there results

$$U_1 = \frac{1}{4} U_2 \left| \sqrt{\varepsilon(1 + U_2)} + \sec\theta_1 \sqrt{1 - \sin^2\theta_1 / \varepsilon(1 + U_2)} \right|^2$$
$$U_{1r} = \frac{1}{4} U_2 \left| \sqrt{\varepsilon(1 + U_2)} - \sec\theta_1 \sqrt{1 - \sin^2\theta_1 / \varepsilon(1 + U_2)} \right|^2 \quad (34)$$

Now R is determined simply from

$$R = \frac{U_{1r}}{U_1} \quad (35)$$

However, using equations (34) by dividing the second by the first will give R as a function of U_2 , but it is much more useful to have R as a function of the incident intensity U_1 , since experimentally U_1 can be varied as is necessary, while U_2 (and also U_{1r}) will only adjust itself to satisfy the boundary conditions, and cannot be thought of as an independent parameter which can be changed to any desired value.

Equations (34) give U_1 and U_{1r} as functions of U_2 , assuming that for any given experimental set up, θ_1 and ε are fixed. As they stand, all quantities are real for the case of transmission of the wave, whereas in the case of TIR, the second radical in each expression is pure imaginary. Here the two cases (TIR and transmission) will be discussed separately. In fact the two cases correspond to the two stable states of the system.

D. Total Internal Reflection Mode

The equations (34) for U_1, U_{1r} can be written in the form

$$U_1 = \frac{1}{4} U_2 \left| \sqrt{\varepsilon(1 + U_2)} + i \sec\theta_1 \sqrt{\frac{\sin^2\theta_1}{\varepsilon(1 + U_2)} - 1} \right|^2$$
$$U_{1r} = \frac{1}{4} U_2 \left| \sqrt{\varepsilon(1 + U_2)} - i \sec\theta_1 \sqrt{\frac{\sin^2\theta_1}{\varepsilon(1 + U_2)} - 1} \right|^2 \quad (36)$$

In this case the quantities to be absolute squared are complex conjugates of each other in the two expressions, and therefore R = 1, since $U_1 = U_{1r}$.

$$U_1 = U_{1r} = \frac{1}{4} U_2 \left[\varepsilon(1 + U_2) + \sec^2\theta_1 \left(\frac{\sin^2\theta_1}{\varepsilon(1+U_2)} - 1 \right) \right]$$

$$R = \frac{U_{1r}}{U_1} = 1 \qquad (37)$$

This is certainly the expected result for TIR, in the case of no absorption. The conditions on U_1 necessary to produce TIR can be found indirectly through restrictions on U_2. The reflectivity will be unity provided θ_1 is greater than the critical angle; stated in terms of U_2, the requirement is

$$\frac{\sin^2\theta_1}{\varepsilon(1+U_2)} - 1 \geq 0 \qquad \text{for TIR Mode}$$

$$\text{or} \quad U_2 \leq \frac{\sin^2\theta_1}{\varepsilon} - 1 \qquad (38)$$

The limiting value of U_2 at which TIR can occur, called U_2^c, is found using the equal sign, and substituting back into the expression for U_1, (36), gives U_1^c, the limiting incident intensity at which TIR can occur:

$$U_1^c = \frac{1}{4} \left(\frac{\sin^2\theta_1}{\varepsilon} - 1 \right) \sin^2\theta_1 \qquad (39)$$

For incident dimensionless intensities less than this value U_1^c, TIR is possible, while any U_1 greater than U_1^c will necessarily produce a transmitted wave. This does not rule out the possibility of transmission for U_1 values less than U_1^c, this can in fact occur in the case of bistability.

E. Transmission Mode

In this case the expressions in the equations (34) for U_1 , U_{1r} are pure real, there is no need for the complex absolute square.

$$U_1 = \frac{1}{4} U_2 \left[\sqrt{\varepsilon(1 + U_2)} + \sec\theta_1 \sqrt{1 - \sin^2\theta_1/\varepsilon(1 + U_2)} \right]^2$$
$$U_{1r} = \frac{1}{4} U_2 \left[\sqrt{\varepsilon(1 + U_2)} - \sec\theta_1 \sqrt{1 - \sin^2\theta_1/\varepsilon(1 + U_2)} \right]^2 \qquad (40)$$

F. Bistability

In the case where the dielectric constant ε_2 is not a function of the electric field intensity E_2, the reflectivity of the boundary is a constant, independent of the power in the light wave (or laser beam). In contrast, when depends on E_2 , the reflectivity will obviously depend on the incident intensity U_1, so that R is not constant, for a given angle of incidence. But a more striking result in this case is that optical bistability is possible, that is, two different reflected wave amplitudes (or two different values of R) can exist for a given incident wave amplitude, under certain conditions. These two values are determined as a result of the fact that for a limited range of U_1 , the boundary can be either transmitting or totally reflecting, giving the two values for R corresponding to a single value of U_1 . The state of the boundary, that is, whether it is transmitting or totally reflecting, for a given U_1 , is determined by the previous values of U_1 , its past history.

Statement of conditions necessary for bistability:

The equations (40) can be used to make graphs of U_1 and U_{1r} versus U_2 , which in the case of no bistability, will be monotically increasing (or decreasing) functions of U_2 . For this case one value for U_1 will determine only one possible value of U_{1r} which satisfies the boundary

conditions. If bistability is to occur, it is necessary that a given value of U_1 correspond to multiple values of U_2, and then each of these U_2 values will determine a different value of U_{1r}, thus producing multiple values for the reflectivity R which relate to a single value of U_1. Another way to state this requirement is to say that the graph of U_1 versus U_2 must have an extremum point, in order that one value of U_1 be determined by more than one value of U_2. Graphs of U_1 versus U_2 make this obvious, as in Figure 2.

In Figure 2, the three values of U_2 which correspond to a single value of U_1 are inserted into the equation (40) for U_{1r}, producing three values for R corresponding to the one value chosen for U_1, thus bistability results. Thus U_1 should be checked to see under what conditions bistability can occur by looking for extremum points. Alternatively it is only necessary to see if U_1 as a function of U_2 first increases and then decreases, or vice versa, this would imply bistability.

By simple inspection, if U_2 is positive, U_1 will always be an increasing function of U_2, and no maximum occurs, therefore no bistability results. For the case of negative values of U_2 both terms within the square are decreasing functions of U_2, and so therefore is the square. When this is multiplied by U_2, the resulting function can evidently start at zero, decrease to some minimum negative value, and then begin to increase again, so that for negative U_2 values bistability is indeed possible. Since U_2 is defined as $U_2 = \alpha|E_2|^2/\varepsilon_2^0$, this means that this form of bistability is possible only for the case of negative nonlinearity ($\alpha < 0$).

G. Generated Curves

Graphs of U_{1r} vs U_1 can be obtained by inserting values for U_2 into the equations (34) to obtain a set of points (U_1, U_{1r}), with each point determined by a different value of U_2. This is in fact just obtaining a parameterization of the curve in the $U_1 - U_{1r}$ plane, with U_2 as the independent parameter. A simple case to consider is that in which $\varepsilon_1 = \varepsilon_2^0$, where the linear dielectric constants on each side of the boundary are matched. In this case it is expected that the boundary is transparent to low intensity waves, but as the intensity increases R should increase from zero towards one.

Figure 3a shows graphs of U_{1r} vs U_1 for $\theta_1 = 89.0^\circ$, an incident angle close to grazing incidence. The solid line represents operating points in the transmission mode, while the dashed line with slope equal to one represents the TIR mode. The dotted line is an unallowed region of the transmission regime. The graph is for negative U_1, U_{1r} values because the case of $\alpha < 0$ is being considered here.

For very small incident intensities, ε_2 will be unaffected, and therefore the critical angle, given by $\sin^2\theta_c = \frac{\varepsilon_2}{\varepsilon_1}$, will be greater than the incident angle θ_1, and the wave will be almost totally transmitted, with very little reflected, R will be close to zero. As the intensity is slowly increased in absolute value, the critical angle, which for small intensities was close to 90°, decreases towards 89°, and eventually will match the incident angle. At this point the reflectivity will suddenly switch to one, as the boundary flips state to the TIR mode.

This occurs at the value of U_1 denoted U_1^s, the point where the transmission mode curve has a vertical tangent. This switching has caused R to change discontinuously from some value around .1 up to unity. Now if

the intensity is further increased, R will remain at one; the operating point will simply move out along the dashed TIR curve. If the intensity is slowly decreased, after TIR has been established, the boundary will remain in the TIR mode even after U_1 has been reduced below U_1^s. It is only when U_1 reaches the value U_1^c previously discussed that the mode will switch to transmission. Thus this simple model predicts that the boundary will exhibit hysteresis effects as a result of the optical bistability present.

It is to be noted that in order for bistability to result, an incident intensity at least as large as U_1^s must be available to initiate the switching from transmission to TIR. Thus it would be convenient to find the approximate value of U_1^s .

H. Approximation for U_1^s

The switching point U_1^s is found from the condition

$$\left.\frac{dU_1}{dU_2}\right|_{U_1 = U_1^s} = 0 \tag{41}$$

Actually this condition determines the value of U_2 at which the slope of U_1 vs U_2 is zero, and then U_1^s is found using this value of U_2.

For very small values of U_2 , U_1 can be approximated by

$$U_1 \approx \frac{1}{4} U_2 \left[\varepsilon(1 + U_2) + \sec^2\theta_1 \left(1 - \frac{\sin^2\theta_1}{\varepsilon}\right)(1 - U_2) + 2\sec\theta_1 \sqrt{\varepsilon(1+U_2) - \sin^2\theta_1}\right] \tag{42}$$

for the transmission mode equation. The value of U_2 , denoted U_2^s , where the slope is zero is found to be approximately

$$U_2^s \approx \frac{-\left(1 + \frac{\sqrt{\varepsilon - \sin^2\theta_1}}{\varepsilon \cos_1}\right)^2}{2\left(1 + \frac{\mathrm{Tan}^2\theta_1}{\varepsilon}\right) + \frac{3\sec\theta_1}{\sqrt{\varepsilon - \sin^2\theta_1}}} \tag{43}$$

For the special case $\varepsilon = 1$, this reduces to

$$U_2^s \approx -\frac{4}{5} \cos^2\theta_1 \quad \text{(for } \varepsilon = 1 \text{ only)} \tag{44}$$

The corresponding switching value U_1^s in this special case is found to be

$$U_1^s \approx -.42 \cos^2\theta_1 \quad \text{(for } \varepsilon = 1 \text{ only)} \tag{45}$$

In the more general case it is not evident that a simple expression for U_1^s will result.

For $\theta_1 = 89°$ with $\varepsilon = 1$, the approximate value of U_1^s obtained by this formula is -1.28×10^{-4} , which agrees almost exactly with the data obtained by a PDP - 11/70 time sharing computer. This would be expected since the approximations made here are most accurate for very small U_1, U_2 values.

I. Conversion of U-Values To Power Per Unit Area

To convert U_1 values to power levels per unit area, it is only necessary to apply the definitions of U_1 and n_{NL} (for which experimental numbers are given for different materials), and calculate the magnitude of the Poynting vector, given as [6]

$$\left|\vec{S}\right| = \frac{c}{8\pi} \left|\vec{E} \times \vec{H}\right| = \frac{c}{8\pi} \sqrt{\varepsilon} \left|E\right|^2 \quad \text{(in CGS units)} \tag{46}$$

(= time averaged power/unit area)

We also have

$$\left|E_1\right|^2 = U_1 \frac{\varepsilon_2^{\circ}}{\alpha} = U_1 \frac{\sqrt{\varepsilon_2^{0}}}{2n_{NL}} \tag{47}$$

Then the incident power is given in terms of U_1 as

$$\left|\vec{S}_1\right| = \frac{c}{16\pi} \frac{\sqrt{\varepsilon_1 \varepsilon_2^{0}}}{n_{NL}} U_1 \tag{48}$$

In figures 3b a graph of R verses U_1 is shown for the case where $\theta_c^\circ = 88°$. The incident angle $\theta_1 = 88.5°$, and the curve is drawn for the case of a positive nonlinearity. Under these conditions the reflectance undergoes no discontinuous changes, and therefore the boundary does not exhibit any bistability.

In figure 8 the incident angle $\theta_1 = 87.5°$, and the curve is drawn for the case of a negative nonlinearity. Bistability is present, and a switching intensity U_1^s equal to about -2.2×10^{-4} is necessary for the boundary to switch from transmission to TIR. Using this number an estimate will be made for the power necessary to initiate this transition, assuming the nonlinear medium has $n_{NL} = 9 \times 10^{-12}$ cm^2/V^2, which is the magnitude of the nonlinear index for CS_2[1]. If we take $\varepsilon_2^\circ = 2.25$, then $\varepsilon_1 = 2.247$, and the switching power is approximately 3.3×10^9 W/cm^2. This numerical value may be compared to the experimental value of 7.5×10^9 W/cm^2 quoted by Smith et al. in their experiment using CS_2, and their theoretical value of 8.1×10^9 W/cm^2 calculated from Kaplan's Theory. Again it is the result of our calculation that a positive Kerr medium shows no bistable effects. Clearly this discrepancy must be resolved!

J. Applications

The nonlinear boundary discussed here has essentially the same input-output characteristics as that of a transistor Schmitt trigger, and therefore could be used in any applications where optical Schmitt trigger operation is necessary. This would include switching applications and digital signal processing applications. Under proper biasing, it may also be useful as an amplifier, and possibly a power stabilizer. But in order for any of these to be more feasible, it would most probably be necessary to make use of a nonlinear medium with a nonlinear index of refraction larger than 10^{-10} cm^2/V^1, in order for the critical switching power to be within a reasonable practical limit.

IV. OPTICAL BISTABILITY NEAR THE PLASMON ANGLE

With the existence of optical bistability near the critical angle at a linear to nonlinear medium boundary, the question is asked: does a similar phenomenon occur when the incident angle is near the plasmon angle for excitation of surface plasmons when a thin metal film is sandwiched between the two media? This question can be theoretically investigated in a manner very similar to the first half of this paper.

The geometrical set up is shown in Figure 4. Medium 1 is a linear glass, medium 2 is a thin metal film of thickness d, with a complex dielectric constant, the imaginary part of which exhibits the absorption of the material. Medium 3 is the nonlinear medium, with a field dependent dielectric constant. A plane wave is assumed incident on the boundary from the linear glass, with electric field amplitude $\vec{E}_1$. The reflected electric field amplitude is denoted $\vec{E}_{1r}$, while that transmitted to the nonlinear medium is denoted $\vec{E}_3$. The fields $\vec{E}_2$ and $\vec{E}_{2r}$ produce standing waves within the metal. Again the case of P-polarization is under consideration, especially since this polarization is necessary to create the conditions responsible for excitation of surface plasmons[3]. Also the set up, referred to as the Kretschmann configuration, allows for the excitation of surface plasmons through the method of frustrated total reflection[7].

Again θ_1 is the incident angle, and is always real, while the angles θ_2 and θ_3 can be complex, especially in the case where evanescent waves exist in media 2 and 3. The goal of this calculation is again to obtain the ratio of the reflected optical power to the incident optical power, i.e. the reflectivity R.

A. Assumed Field Dependences

A coordinate system is chosen with z-axis perpendicular to the boundary, directed into the nonlinear medium, while the x-axis is parallel to the boundary. Monochromatic plane wave solutions to the wave equation are assumed in both the linear glass and the metal; a slightly more complicated form is assumed in the nonlinear medium, of the form

$$\vec{E}_3(\vec{x},t) = \vec{E}_3(z)\, e^{iK_{3x}x}\, e^{i\int_d^z K_{2z}(z')dz'}\, e^{-i\omega t} \tag{49}$$

The magnetic field associated with this E - field is found from Faraday's Law:

$$\vec{\nabla} \times \vec{E}_3 = \frac{i\omega}{c}\, \vec{H}_3 \tag{50}$$

All fields are assumed to have $e^{-i\omega t}$ time dependence, and note that the plane $z = 0$ is at the boundary between the linear glass and the metal, which has a thickness d. Below the fields in the three media are written out, with $e^{-i\omega t}$ suppressed.

Incident Wave $\vec{E}_1$

$$\begin{aligned} E_{1x} &= E_1 \cos\theta_1\, e^{i(K_{1x}x + K_{1z}z)} & H_{1x} &= 0 \\ E_{1y} &= 0 & H_{1y} &= \frac{c}{\omega} K_1 E_1\, e^{i(K_{1x}x + K_{1z}z)} \\ E_{1z} &= -E_1 \sin\theta_1\, e^{i(K_{1x}x + K_{1z}z)} & H_{1z} &= 0 \end{aligned} \tag{51}$$

Reflected Wave $\vec{E}_{1r}$

$$\begin{aligned} E_{1rx} &= -E_{1r} \cos\theta_1\, e^{1(K_{1x}x - K_{1z}z)} & H_{1rx} &= 0 \\ E_{1ry} &= 0 & H_{1ry} &= \frac{c}{\omega} K_1\, E_{1r}\, e^{i(K_{1x}x - K_{1z}z)} \\ E_{1rz} &= -E_{1r} \sin\theta_1\, e^{i(K_{1x}x - K_{1z}z)} & H_{1z} &= 0 \end{aligned} \tag{52}$$

Wave in Metal $\vec{E}_2$

$$E_{2x} = E_2 \cos\theta_2 \, e^{i(K_{2x}x + K_{2z}z)} \qquad H_{2x} = 0$$

$$E_{2ry} = 0 \qquad H_{2ry} = \frac{c}{\omega} K_2 E_{2r} \, e^{i(K_{2x}x - K_{2z}z)}$$

$$E_{2rz} = -E_{2r} \sin\theta_2 \, e^{i(K_{2x}x - K_{2z}z)} \qquad H_{2z} = 0$$

Wave in Metal $\vec{E}_{2r}$

$$E_{2rx} = -E_{2r} \cos\theta_2 \, e^{i(K_{2x}x - K_{2z}z)} \qquad H_{2rx} = 0$$

$$E_{2ry} = 0 \qquad H_{2ry} = \frac{c}{\omega} K_2 E_2 \, e^{i(K_{2x}x - K_{2z}z)} \qquad (54)$$

$$E_{2rz} = -E_{2r} \sin\theta_2 \, e^{i(K_{2x}x - K_{2z}z)} \qquad H_{2rz} = 0$$

Transmitted Wave In Medium 3 $\vec{E}_3$ (Nonlinear)

$$E_{3x} = E_3(z) \cos_3 \, e^{iK_{3x}x} \, e^{i\int_d^z K_{3z}(z')dz'}$$

$$E_{3z} = E_3(z) \sin_3 \, e^{iK_{3x}x} \, e^{i\int_d^z K_{3z}(z')dz'}$$

$$H_{3y} = \frac{c}{\omega}\left[K_3 E_3(z) - i \cos_3 \frac{dE_3(z)}{dz}\right] e^{iK_{3x}x} \, e^{i\int_d^z K_{3z}(z')dz'} \qquad (55)$$

$$E_{3y} = H_{3x} = H_{3z} = 0$$

The comments made previously in Secton III-A about whether this is a valid assumed form of field in the nonlinear medium should be recalled here, the nonlinear wave equation should in fact be solved to obtain the exact form of the fields. For simplicity the above fields will be used, and the wave-vectors K in the three media are related to the dielectric constants through the relations

$$K_1 = \frac{\omega}{c}\sqrt{\varepsilon_1} \qquad K_2 = \frac{\omega}{c}\sqrt{\varepsilon_{2r} + i\varepsilon_{2i}} \qquad K_3 = \frac{\omega}{c}\sqrt{\varepsilon_3^0 + \alpha|E_3(z)|^2} \qquad (56)$$

Application of the boundary conditions here leads to Sneel's law, in the form that the tangential component (x-component) of the wavevector must be continuous across the boundaries, so that

$$K_{1x} = K_{2x} = K_{3x} \qquad (57)$$

From this relation and relations of the form

$K_{1x}^2 + K_{1z}^2 = K_1^2$ similar for K_2, K_3, the z-components

of the wavevectors can be written in terms of the incident angle θ_1 and the dielectric constants $\varepsilon_1, \varepsilon_2, \varepsilon_3$, as follows.

$$K_{1z} = \frac{\omega}{c}\sqrt{\varepsilon_1}\cos\theta_1$$

$$K_{2z} = \frac{\omega}{c}\sqrt{\varepsilon_1}\cos\theta_2 = \frac{\omega}{c}\sqrt{\varepsilon_2 - \varepsilon_1\sin^2\theta_1} = K_{2zr} + i\,K_{2zi} \qquad (58)$$

$$K_{3z} = \frac{\omega}{c}\sqrt{\varepsilon_1}\cos\theta_3 = \frac{\omega}{c}\sqrt{\varepsilon_3 - \varepsilon_1\sin^2\theta_1}$$

The real and imaginary parts of K_{2z} can be written as

$$K_{2zr} = \frac{1}{\sqrt{2}}\left[\frac{\varepsilon_{2r}}{\varepsilon_1} - \sin^2\theta_1 + \sqrt{\left(\frac{\varepsilon_{2r}}{\varepsilon_1} - \sin^2\theta_1\right)^2 + \left(\varepsilon_{2i}/\varepsilon_1\right)^2}\right]^{\frac{1}{2}} \frac{2\pi}{\lambda_1}$$

$$K_{2zi} = \frac{1}{\sqrt{2}}\left[-\frac{\varepsilon_{2r}}{\varepsilon_1} + \sin^2\theta_1 + \sqrt{\left(\frac{\varepsilon_{2r}}{\varepsilon_1} - \sin^2\theta_1\right)^2 + \left(\varepsilon_{2i}/\varepsilon_1\right)^2}\right]^{\frac{1}{2}} \frac{2\pi}{\lambda_1} \qquad (59)$$

where λ_1 is the wavelength of the light in medium 1. In the case where $|\varepsilon_{2i}| \ll |\varepsilon_{2r}|$, these are given to first order in $\varepsilon_{2i/\varepsilon}$, as

$$K_{2zr} \approx \frac{1}{2}\,\frac{\varepsilon_{2i}/\varepsilon_1}{\sqrt{\sin^2\theta_1 - \varepsilon_{2r}/\varepsilon_1}}\,\frac{2\pi}{\lambda_1} \qquad K_{2zi} \approx \sqrt{\sin^2\theta_1 - \varepsilon_{2r}/\varepsilon_1}\,\frac{2\pi}{\lambda_1} \qquad (60)$$

The real part of ε_2 (metal dielectric) is given approximately in terms of the plasma frequency[8] $\omega_p = 4\pi ne^2/m_e$ through the relation

$$\varepsilon_{2r} = 1 - \frac{\omega_p^2}{\omega^2} \qquad (61)$$

For optical frequencies for most metals, the plasma frequency is greater than ω, and ε_{2r} is a negative number. This is then responsible for the heavy damping of the waves in the metal.

B. Boundary Condition Equations

Again the pertinent boundary conditions to be applied at both boundaries, are the continuity of tangential $\vec{E}$ and tangential $\vec{H}$. Also the normal component of $\vec{D}$ must be continuous, but this condition is not necessary.

Tangential $\vec{E}$ (continuity of E_x)

$$\begin{aligned} &\text{at } z = 0 \qquad E_1 \cos\theta_1 - E_{1r} \cos\theta_1 = E_2 \cos\theta_2 - E_{2r} \cos\theta_2 \\ &\text{at } z = 0 \qquad E_2 \cos\theta_2 \, e^{iK_{2z}d} - E_{2r} \cos\theta_2 \, e^{-iK_{2z}d} = E_3(d) \cos\theta_3 \end{aligned} \tag{62}$$

Tangential $\vec{H}$ (continuity of H_y)

$$\begin{aligned} &\text{at } z = 0 \quad K_1E_1 + K_1E_{1r} = K_2E_2 + K_2E_{2r} \\ &\text{at } z = 0 \quad K_2E_2 \, e^{iK_{2z}d} + K_2E_{2r} \, e^{-iK_{2z}d} = K_3(d)E_3(d) - i \, \frac{K_{3z}(d)}{K_3(d)} \left.\frac{dE_3}{dz}\right|_{z=d} \end{aligned} \tag{63}$$

If the assumption is made that the derivative term $\left.\frac{dE_3}{dz}\right|_{z=d}$ is in some sense small compared to the other term on the right hand side of the last equation, then equations (62) and (63) can be put into the matrix form

$$\begin{bmatrix} 1 & \cos\theta_2/\cos\theta_1 & -\cos\theta_2/\cos\theta_1 & 0 \\ -1 & K_2/K_1 & K_2/K_1 & 0 \\ 0 & e^{iK_{2z}d} & -e^{-iK_{2z}d} & -\cos\theta_3/\cos\theta_2 \\ 0 & e^{iK_{2z}d} & e^{-iK_{2z}d} & -K_3/K_2 \end{bmatrix} \begin{bmatrix} E_{1r} \\ E_2 \\ E_{2r} \\ E_3(d) \end{bmatrix} = \begin{bmatrix} E_1 \\ E_1 \\ 0 \\ 0 \end{bmatrix} \tag{64}$$

In the case where $\cos\theta_2$ & $\cos\theta_3$ may be complex, they can be replaced by their equivalents

$$\cos\theta_2 = \frac{K_{2z}}{K_2} \qquad \cos\theta_3 = \frac{K_{3z}}{K_3} \tag{65}$$

The set of four equations can be solved simultaneously, to obtain

$$E_{1r} = E_1 \left[\frac{r_{12} + r_{23}\, e^{i2K_{2z}d}}{1 + r_{12}r_{23}\, e^{i2K_{2z}d}} \right]$$

$$E_2 = E_1 \left[\frac{t_{12}}{1 + r_{12}r_{23}\, e^{i2K_{2z}d}} \right]$$

$$E_{2r} = E_1 \left[\frac{t_{12}r_{23}\, e^{i2K_{2z}d}}{1 + r_{12}r_{23}\, e^{i2K_{2z}d}} \right] \qquad (66)$$

$$E_3(d) = E_1 \left[\frac{t_{12}t_{23}\, e^{iK_{2z}d}}{1 + r_{12}r_{23}\, e^{i2K_{2z}d}} \right]$$

The symbols r_{12}, r_{23}, t_{12} & t_{23} are the Fresnel reflection and transmission coefficients for the two boundaries, defined by

$$r_{12} = \frac{K_2 \cos\theta_1 - K_1 \cos\theta_2}{K_2 \cos\theta_1 + K_1 \cos\theta_2} = \frac{\sqrt{\frac{\varepsilon_2}{\varepsilon_1}}\, K_{1z} - \sqrt{\frac{\varepsilon_1}{\varepsilon_2}}\, K_{2z}}{\sqrt{\frac{\varepsilon_2}{\varepsilon_1}}\, K_{1z} + \sqrt{\frac{\varepsilon_1}{\varepsilon_2}}\, K_{2z}}$$

$$t_{12} = \frac{2K_1 \cos\theta_1}{K_2 \cos\theta_1 + K_1 \cos\theta_2} = \frac{2K_{1z}}{\sqrt{\frac{\varepsilon_2}{\varepsilon_1}}\, K_{1z} + \sqrt{\frac{\varepsilon_1}{\varepsilon_2}}\, K_{2z}} \qquad (67)$$

Expressions for r_{23} and t_{23} are found from the above by changing the subscripts $1 \to 2$ and $2 \to 3$. It is to be noted here that both r_{23} and t_{23} will depend on the field E_3, so that all four fields determined above are not simply directly proportional to E_1.

C. Excitement of Surface Plasmons

1. Linear Case

In the case where medium 3 is considered a linear medium ($\alpha \to 0$), the reflectivity R is given simply by $R = |E_{1r}|^2/|E_1|^2$

$$R = \left| \frac{r_{12} + r_{23}\, e^{i2K_{2z}d}}{1 + r_{12}r_{23}\, e^{i2K_{2z}d}} \right|^2 \tag{68}$$

For a special incident angle termed the plasmon angle, θ_p the denominator of the Fresnel Factor r_{23} vanishes[9] (in the case of $\varepsilon_{2i} = 0$), and this is the condition for excitement of the surface plasmon mode. The plasmon angle is given by [3]

$$\varepsilon_1 \sin^2\theta_p = \frac{\varepsilon_2\varepsilon_3}{\varepsilon_{2+}\varepsilon_3} = K_x^2 \tag{69}$$

This relation is also the dispersion relation for the component of the surface plasmon wavevector which is tangent to the boundary between the metal and medium 3.

A simple interpretation of the behavior of the fields at the plasmon mode is given by realizing that if r_{23} diverges as at the plasmon angle, then this means that the field E_2 is very small compared to the field E_{2r}. Both of these fields normally have exponential z-dependence, the first decreasing as the metal - medium 3 boundary is approached, while the second increases towards this boundary. It is easy to show that the plasmon angle is greater than the critical angle between media 1 and 3, and therefore the field in medium 3 is also evanescent, decaying in amplitude away from the boundary. Thus, under conditions required for the surface plasmon mode, the electric field amplitude decreases exponentially on both sides of the metal - medium 3 boundary, the field E_2 being negligible. Also it is possible that the field at this boundary can be larger than the incident field E_1.

If the metal has no absorption, then $R = 1$ at the plasmon mode. In the case where $\varepsilon_{2i} \neq 0$, however, excitation of the surface plasmon is characterized by a large decrease in R below one, it can in fact approach zero closely. This large drop in R can be attributed to the fact that a collective oscillation of the electron gas in the metal is associated with the plasmon excitation, or resonance, and this allows for absorption of energy by the metal, thereby decreasing the reflected optical power at the plasmon angle.

2. Nonlinear Case

When α is nonzero , the reflectivity R will be a function of the incident field amplitude E_1, since ε_3 depends on E_3 and E_3 will depend on E_1. In this case the relations (66) between E_1, E_{1r}, and E_3 can be rewritten in the form

$$E_{1r} = E_3(d) \left[\frac{r_{12} + r_{23}\, e^{i2K_{2z}d}}{t_{12}\, t_{23}\, e^{iK_{2z}d}} \right]$$

$$E_1 = E_3(d) \left[\frac{1 + r_{12}\, r_{23}\, e^{i2K_{2z}d}}{t_{12}\, t_{23}\, e^{iK_{2z}d}} \right]$$

These two equations express E_{1r} and E_1 as nonlinear functions of E_3 at the metal - nonlinear boundary, since both r_{23} and t_{23} depend on $E_3(d)$. Thus from these a graph of $|E_{1r}|^2$ versus $|E_1|^2$ can be obtained by inserting different values for $E_3(d)$ on the right hand sides of the equations; for each value of $E_3(d)$ used, a different point on the curve will be obtained. By substituting an adequate number of values, a smooth curve will be determined.

Substituting the definitions of the Fresnel Factors into these expressions allows the field dependences of E_{1r} and E_1 on $E_3(d)$ to be

shown more explicitly in the following forms.

$$E_{1r} = \frac{1}{2} E_3(d) \left\{ \sqrt{\frac{\varepsilon_1 \overset{0}{\varepsilon_3}}{\varepsilon_2}} \left[\frac{\varepsilon_2}{\varepsilon_1} \cos(K_{2z}d) + i \sin(K_{2z}d) \sec\theta_1 \sqrt{\frac{\varepsilon_2}{\varepsilon_1} - \sin^2\theta_1} \right] \sqrt{1+U_3} \right.$$

$$\left. - \left[\cos(K_{2z}d) \sec\theta_1 + i \frac{\varepsilon_2}{\varepsilon_1} \sin(K_{2z}d) \left(\frac{\varepsilon_2}{\varepsilon_1} - \sin^2\theta_1\right)^{-1/2} \right] \sqrt{1 - \frac{\varepsilon_1}{\overset{0}{\varepsilon_3}} \left(\frac{\sin^2 1}{1+U_3}\right)} \right\} \quad (71)$$

$$E_1 = \frac{1}{2} E_3(d) \left\{ \sqrt{\frac{\varepsilon_1 \overset{0}{3}}{\varepsilon_2}} \left[\frac{\varepsilon_2}{\varepsilon_1} \cos(K_{2z}d) - i \sin(k_{2z}d) \sec\theta_1 \sqrt{\frac{\varepsilon_2}{\varepsilon_1} - \sin^2\theta_1} \right] \sqrt{1+U_3} \right.$$

$$\left. + \left[\cos(K_{2z}d) \sec\theta_1 - i \frac{\varepsilon_2}{\varepsilon_1} \sin(K_{2z}d) \left(\frac{\varepsilon_2}{\varepsilon_1} - \sin^2\theta_1\right)^{-1/2} \right] \sqrt{1 - \frac{1}{\overset{0}{\varepsilon_3}} \left(\frac{\sin^2 \theta 1}{1+U_3}\right)} \right\} \quad (72)$$

These relations can also be written in terms of the dimensionless intensities U_1, U_{1r}, and U_3.

$$U_1 = \text{function of } U_3 = f(U_3)$$
$$U_{1r} = \text{function of } U_2 = g(U_3) \quad (73)$$

This then gives an alternate view (a graphical description) of how the functional dependence of R on U_1 and bistability can be determined. As in the one boundary problem already discussed, when the graph of U_1 versus U_3 has extremum points, then a single value of U_1 can be determined by more than one value of U_3. Then each of these U_3 values will determine a distinct value of U_{1r}, and all of this results in a single value of U_1 having multiple values of R related to it, each of which characterizes electric field amplitudes which satisfy all of the boundary conditions. (As in Figure 2) Thus the mathematical conditions necessary for bistability here are the same as before, as is the method for producing the graph of R vs. U_1. The algebra is more involved, however, especially due to the fact that since absorption in the metal is allowed for in terms of a complex dielectric constant, most of the factors in equations (70) involve complex numbers. As a result it is essential that the calculations be performed by computer.

D. Discussion of Switching Near the Plasmon Angle

Above is the method for mathematically determining the reflectivity as a function of intensity. A physical interpretation of what happens near θ°_p may be useful here. A case will be discussed in which the angle of incidence θ_1, is slightly greater than the linear plasmon angle θ°_p (plasmon angle at low intensity), defined through

$$\varepsilon_1 \sin^2\theta^\circ_p = \frac{\varepsilon_2 \varepsilon^\circ_3}{\varepsilon_2 + \varepsilon^\circ_3} \tag{74}$$

First, if the incident beam has very low intensity, no nonlinear effects will be evident, and the reflectivity will be some value greater than that which would result if the surface plasmon mode were fully excited. (Changing θ_1, so that it matches θ°_p will fully excite the surface plasmon mode, and result in the lowest possible reflectivity). Note here that since the plasmon angle θ_p depends on the variable dielectric constant ε_3, then the plasmon angle is not fixed, but changes with the intensity of the incoming beam. Now if the intensity of the beam is increased, then it is possible to promote excitation of the plasmon resonance by making the variable plasmon angle change until it matches the incident angle. For the case of a positive nonlinearity, increasing the intensity U_3 will cause θ_p to increase, so that for the example case cited here, the plasmon angle can be forced to approach the angle of incidence; the plasmon mode can be excited even though θ_1 is not equal to the linear plasmon angle θ°_p.

Figure 5 shows a typical plot of R vs. U_1 for values of the parameters ε_1, ε_2, ε_3, etc. as indicated, and for the case of positive nonlinearity. The values are appropriate for a 625 Å silver film on a glass prism of index $n_1 = 1.9$, with an incident wavelength 1.06μ.

An interesting thing happens as the intensity is increased. Instead of the reflectivity simply declining smoothly as the plasmon condition is approached, it is found that a point is reached such that if the intensity is increased only slightly further beyond this point, then the reflectivity undergoes a discontinuous decrease; the operating point of the system will now be on the lower portion of the curve. In fact, it would now be necessary to decrease the intensity to obtain excitation of the surface plasmon, since the operating point will now stay on the lower portion of the curve, and will move along this curve as U_1 changes. If the intensity is decreased enough the plasmon mode is excited (the point where R is a minimum). Decreasing U_1 further only slightly will now cause a discontinuous increase in R up to some value around 0.9; the operating point is now back on the upper portion of the curve. The net effect here is that R exhibits hysteresis behavior with respect to changes in U_1. Also an important and curious detail to notice is that operating points on the portion of the curve which connects the upper and lower sections referred to above are never obtainable, even though the fields they relate to satisfy the boundary conditions. The discontinuous jumps in R will always prevent the operating point from moving onto this intermediate section of the curve.

For this example, and for others, there exists a minimum intensity U_1, called U_1^c, which must be applied to the boundary to initiate the discontinuous downward jump in R. Also this intensity is necessary if it is required to excite the surface plasmon mode, since the surface plasmon point is on the lower section of the curve. It is to be expected that this power level associated with U_1^c will depend on the magnitude of the nonlinearity α, the difference between the linear plasmon angle θ_p° and

the incident angle θ_1, denoted $\Delta\theta = \theta_1 - \theta_p^\circ$, and possibly the strength of the surface plasmon resonance, which is related to the ratio of the real part of ε_2 to the imaginary part of ε_2. Therefore it would be advantageous to produce graphs of U_1^c as a function of one of these factors, with the other held fixed.

The Figures 6 and 7 are graphs of U_1^c as a function of $\Delta\theta$ and as a function of ε_{2i}, with the real part of ε_2 held fixed. As $\Delta\theta$ is increased (moving θ_1 to either side of the plasmon angle θ_p°), the critical switching power also increases, although not in a linear fashion. An important thing to notice about this curve is that there is a gap in it near the origin. No points on the curve are plotted here in this region because U_1^c does not exist for displacement angles $\Delta\theta$ less than some minimum value. If $\Delta\theta$ is too small in magnitude, then the curve of R vs. U_1 shows no bistability and consequently no hysteresis. This is especially evident for the case where $\theta_1 = \theta_p^\circ$ or $\Delta\theta = 0$. Here increasing U_1 will simply cause the boundary to move away from the plasmon condition, and R will smoothly increase from its minimum at the plasmon condition. A similar type of behavior is present for very small angular displacements, until some minimum $\Delta\theta$ is reached at which bistability is just barely possible.

In Figure 7 the critical switching intensity increases with increasing ε_{2i}, again not in a linear manner. It seems that where the plasmon resonance is strong, that is, when the ratio $\varepsilon_{2r}/\varepsilon_{2i}$ is large (when ε_{2i} is small), then a smaller critical intensity is required. In the limit of ε_{2i} going to zero, the critical switching intensity is not zero, but some finite value, however, no plasmon resonance is observable in this case, since R will be unity for all values of U_1.

V. RESULTS AND CONCLUSIONS

A. Bistability Near the Critical Angle:

A New method for generating curves of reflectivity versus incident intensity has been demonstrated here, its advantage over previous methods being that the calculation is possible at any incident angle, not just angles near grazing incidence. This is because the grazing angle approximations do not have to be made here. There are, however, other approximations that are made here, the most obvious being the assumption that the solutions of the wave equation in the nonlinear medium can be expressed as something similar to plane waves, characterized by wavevectors K_x and K_z. Certainly this is not necessarily true, and may be accurate only in the limit of a very small nonlinear effect.

The curves are such that in certain special regions, termed bistable, there exist three possible values of reflectivity for a single given incident field intensity. Two of the values correspond to the transmission mode, while the third value relates to the TIR mode. Of the two transmission mode values, only one can physically be made to occur, since it is impossible to switch to the other (the dotted curve in figure 3a is physically unattainable). Thus there are actually only two reflectance values attainable in these bistable regions. Since both of these are possible for any intensity for which the boundary is bistable, there must be a method for determining which of the two states is present under the given conditions. The method referred to here simply involves having the operating point of the system move smoothly along the R vs. U_1 curve; discontinuous jumps occur only at the special switching points referred to in the discussion. At these switching points, switching occurs where initially there are multiple values of R possible, and either increasing or decreasing the intensity moves the operating point to a region where there is only a single R value possible for the given value of U_1. The

switching can be made from TIR to transmission or from transmission to TIR, for both a negative and a positive nonlinearity α . Typical switching powers would be on the order of 10^8 to 10^9 Watts per cm^2, depending on the difference between the incident angle and the linear critical angle, and also on the linear critical angle itself. This latter fact is indicated in Figure 8, which shows curves of R vs. U_1 for various linear critical angles θ_c° , with associated incident angles each 0.5° less than θ_c° . It is easily seen that the switching intensity U_1^s generally increases as the linear critical angle is set farther from grazing incidence, so that this experiment would be most easily done for the case where ϵ_1 and ϵ_2 are almost identical, so that θ_c° is close to 90°. Also it is noticed that for the cases closer to grazing incidence, the jump in reflectivity in going from transmission to TIR is larger.

B. Bistability Near The Plasmon Angle:

The method used above for the critical angle bistability was also used in the case of a two boundary geometry, with a metal being between a linear and a nonlinear medium. Similar type approximations were made. The difference here is that incident angles near the linear plasmon angle θ_p° are of interest.

Again there can be a bistable region with R being a multivalued function of U_1. Here, however, the two accessible regions of the curve both correspond to the TIR case, since this is necessary for excitation of surface plasmons; the difference between the two regions is that they relate to being on opposite sides of the (variable) plasmon angle θ_p . The discontinuous switching causes the following to happen. First, the electric fields are such that the incident angle θ_1 is either greater than or less than the adjustable plasmon angle θ_p , defined by equation (69).

Now changing the strength of the incident electric field can cause the plasmon angle to either increase or decrease, depending on the particular parameters of the problem, and at some point positive feedback takes over which now forces θ_p to be on the opposite side of θ_1 from which it started. The switching here is in effect "across" the plasmon condition. The surface plasmon resonance can be excited only by first producing an intensity large enough to cause the operating point to switch to the lower portion of the R vs. U_1 curve (as previously noted in the discussion section). Now that the operating point is on this lower portion, the intensity must be decreased somewhat until the fields adjust themselves so that θ_1 matches θ_p (then naturally the surface plasmon must be excited). Decreasing the intensity further will cause θ_p to change enough so that now the operating point will jump back to the upper portion of the curve. Thus the boundary set up has been shown to exhibit optical hysteresis.

The optical power necessary to cause the initial optical switching has been shown to depend on the displacement of the incident angle from θ_p°, on the ratio $\varepsilon_{2r}/\varepsilon_{2i}$, and on the magnitude of the nonlinearity α. Typical powers are on the order of 10^8 to 10^9 Watts per square centimeter.

Finally, it is to be noted that the optical bistability exhibited near the plasmon angle allows the experimentalist to work near angles which are not grazing angles, and possibly would offer some relief from problems associated with working near grazing incidence.

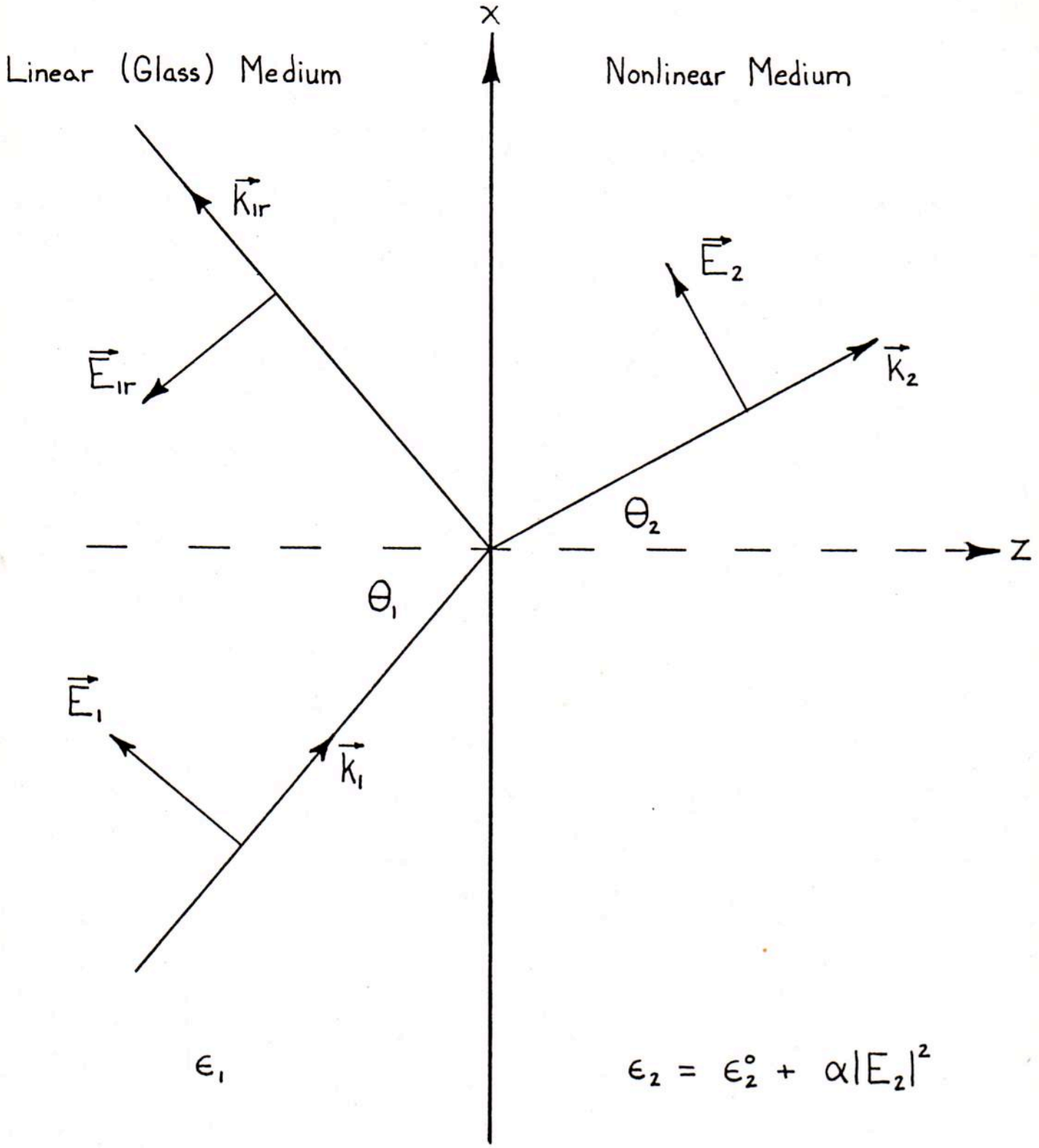

Figure 1. Linear to Nonlinear Boundary Geometry

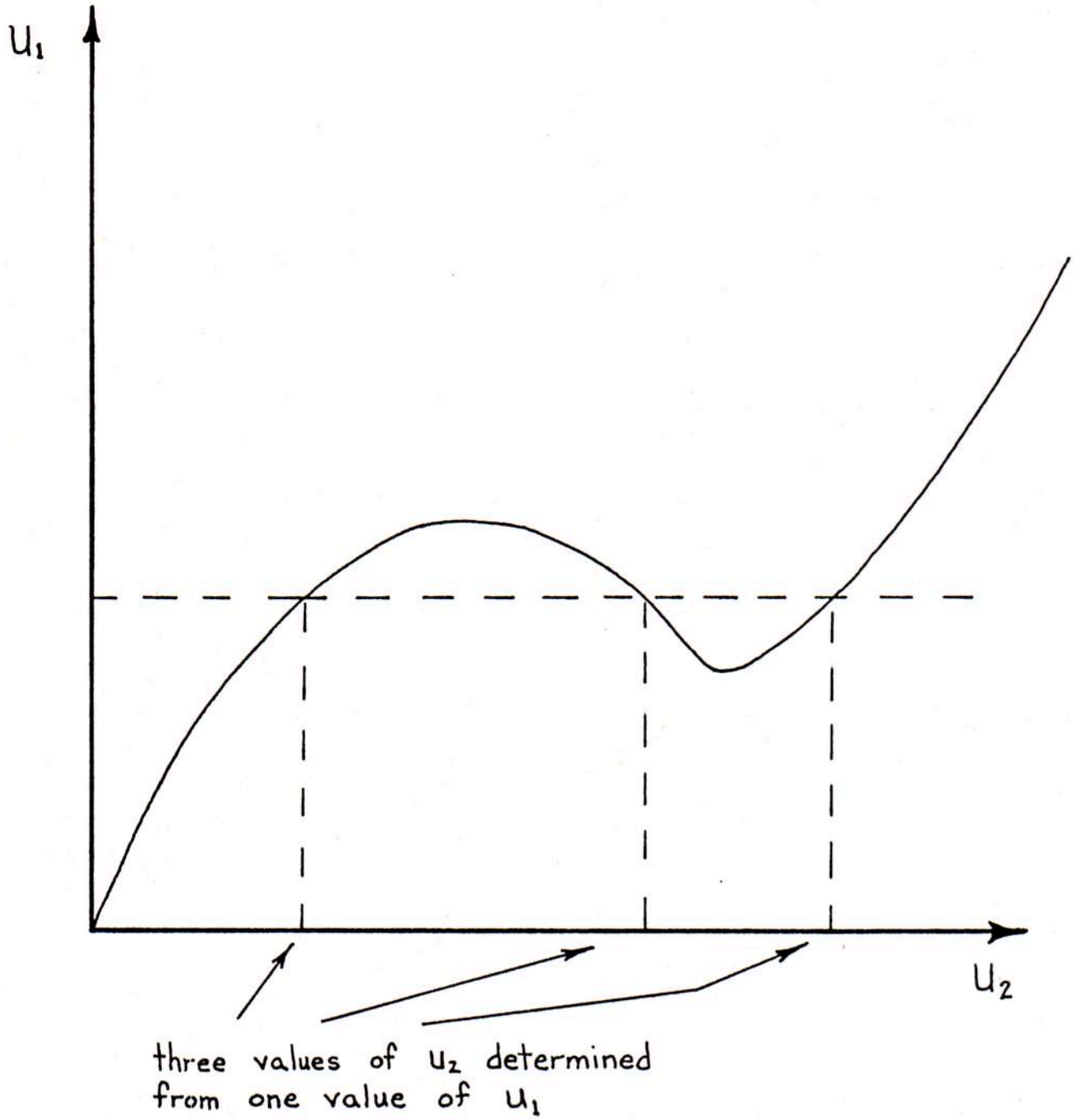

Figure 2. Typical graph of u_1 vs. u_2 with bistability

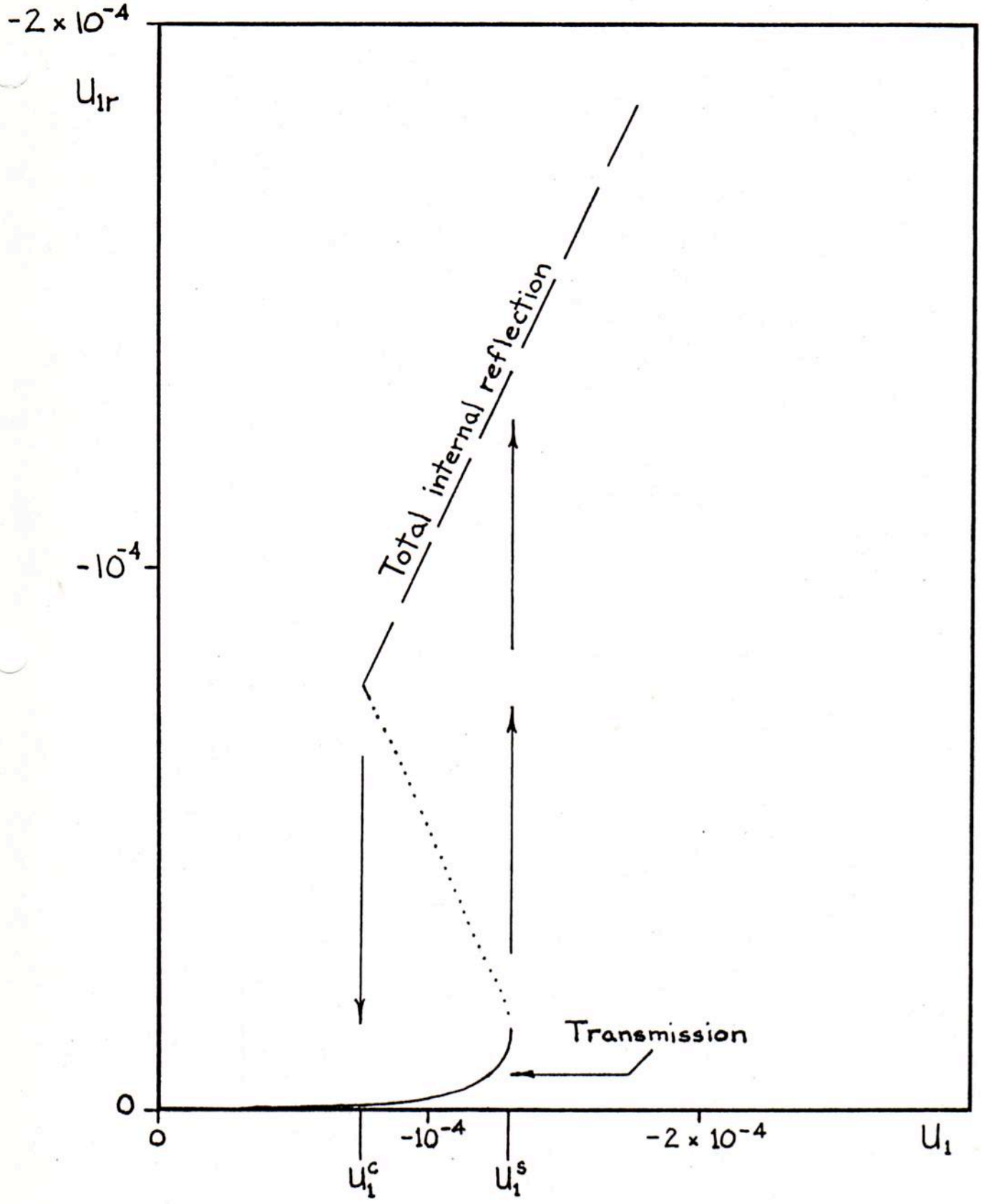

Figure 3a. U_{1r} vs. U_1 for $\theta_1 = 89°$, $\epsilon_1 = \epsilon_2^\circ$

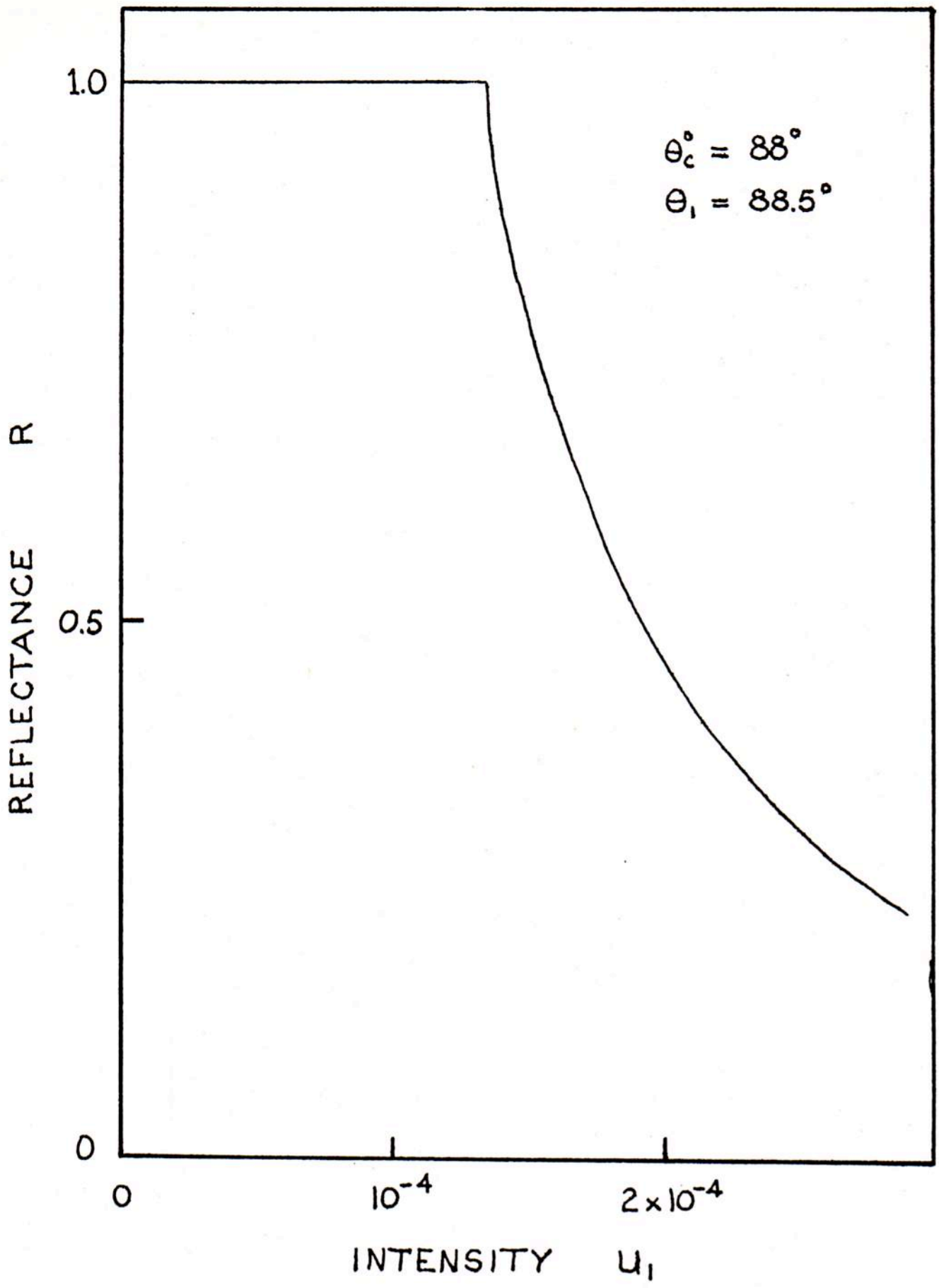
1.0
0.5
0
REFLECTANCE R
$\theta_c^\circ = 88^\circ$
$\theta_1 = 88.5^\circ$
0
10^{-4}
2×10^{-4}
INTENSITY U_1

Figure 3b.

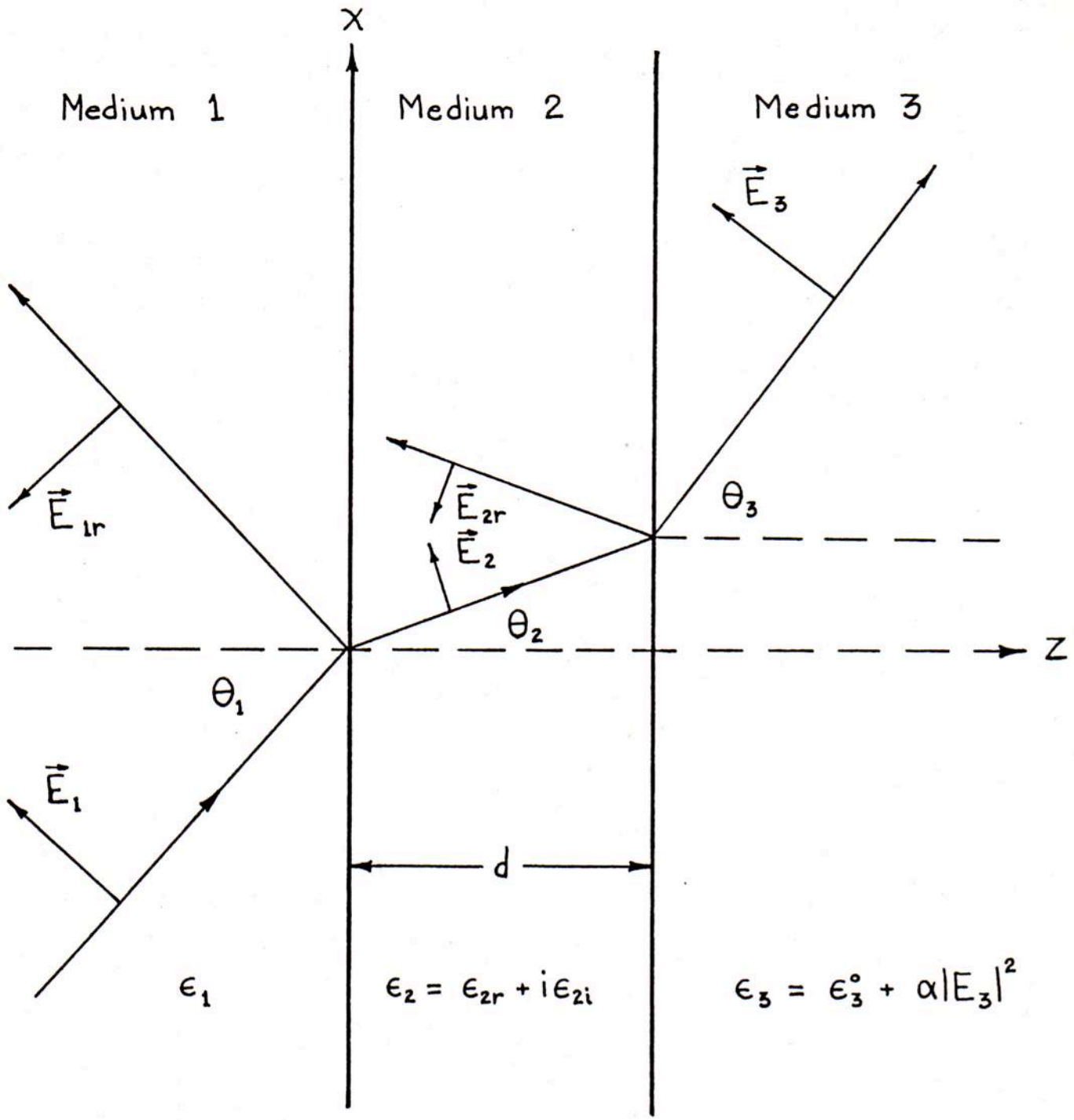

Figure 4. Surface Plasmon Geometry

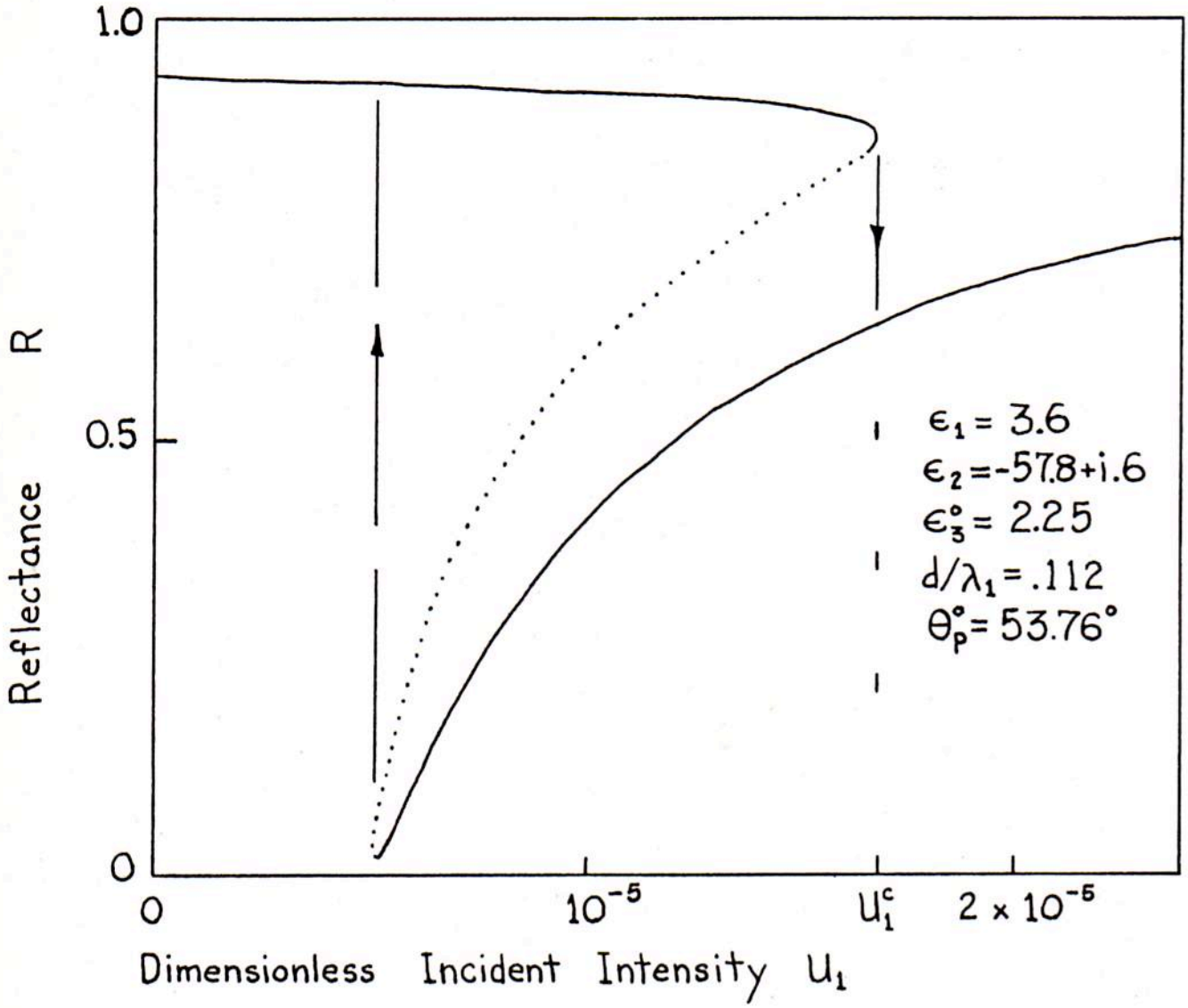

Figure 5. Reflectance vs. Intensity U_1, with $\theta_1 = 53.90^\circ$

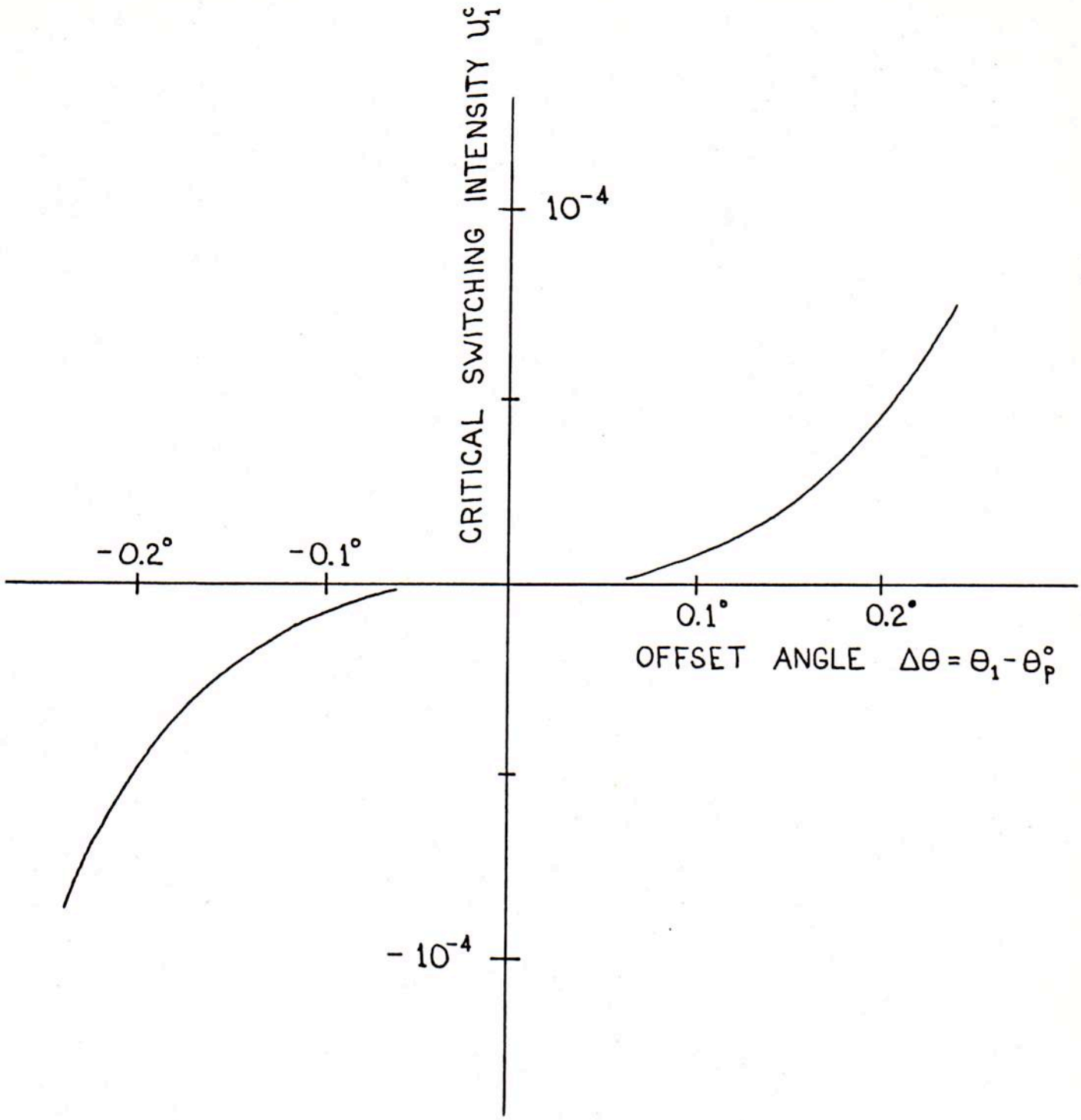

Figure 6. Critical Intensity U_1^c vs. Offset Angle $\Delta\theta$

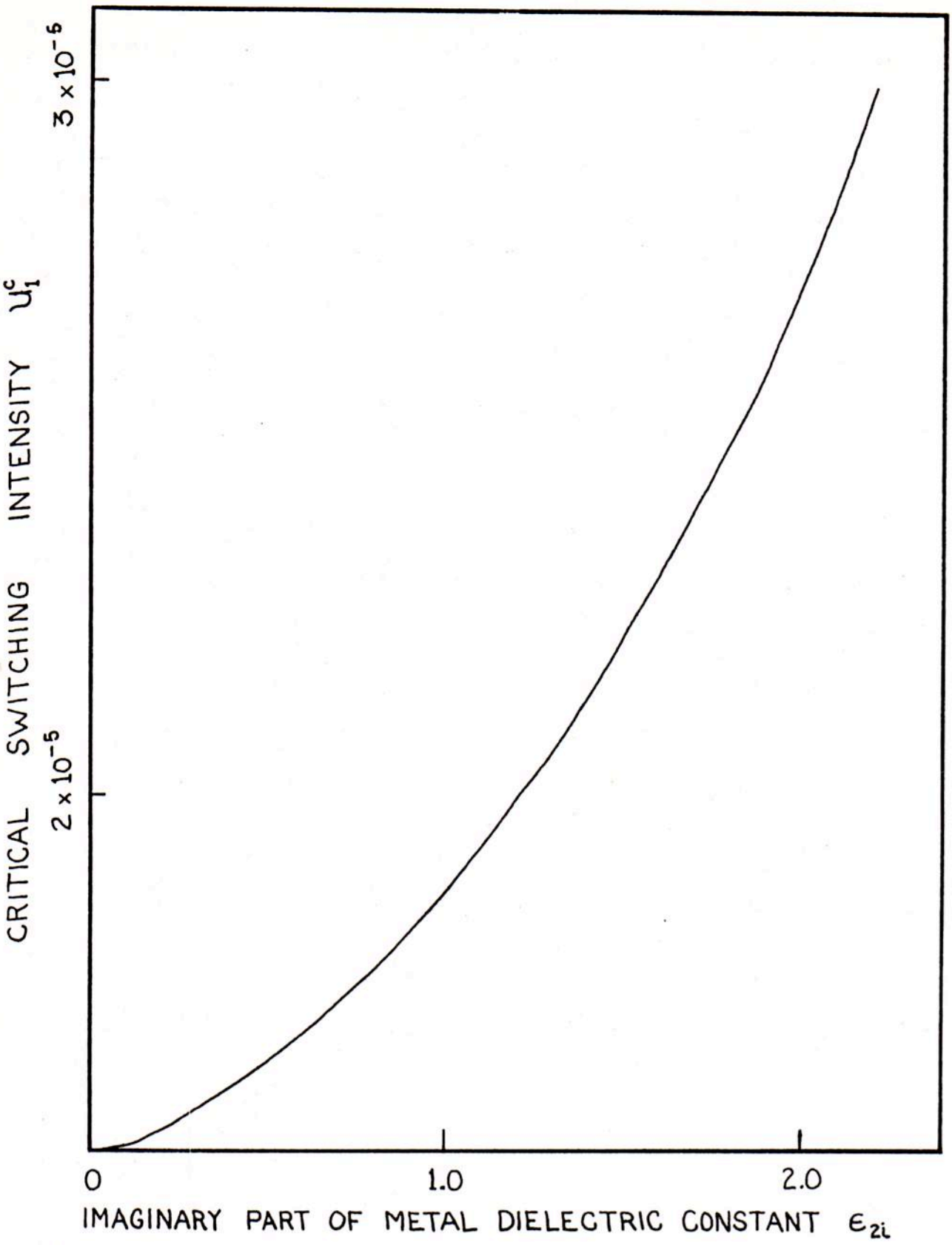

Figure 7. Critical Intensity U_1^c vs. ϵ_{2i}

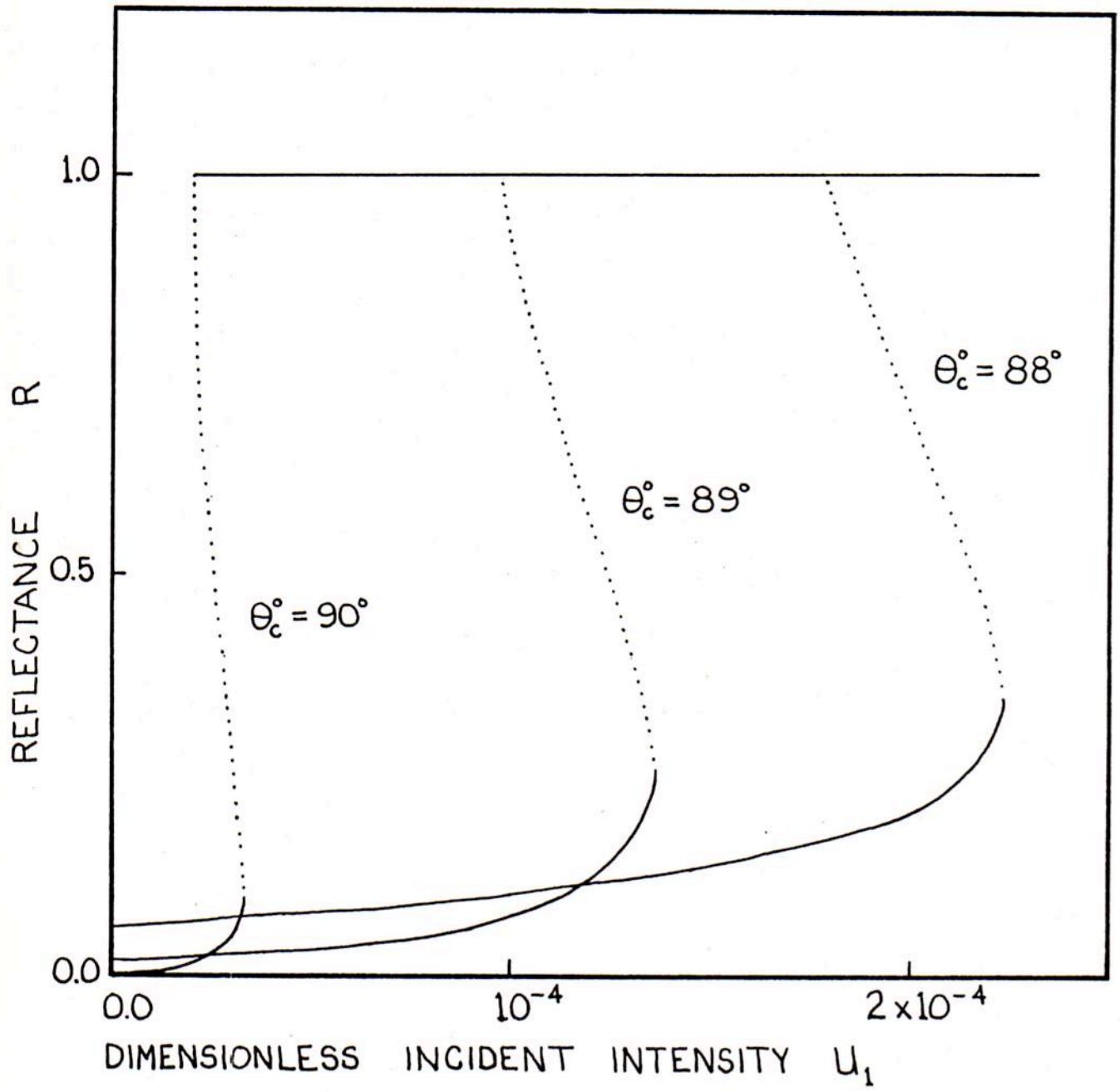

Figure 8. R vs U_1 for various critical angles θ_c°

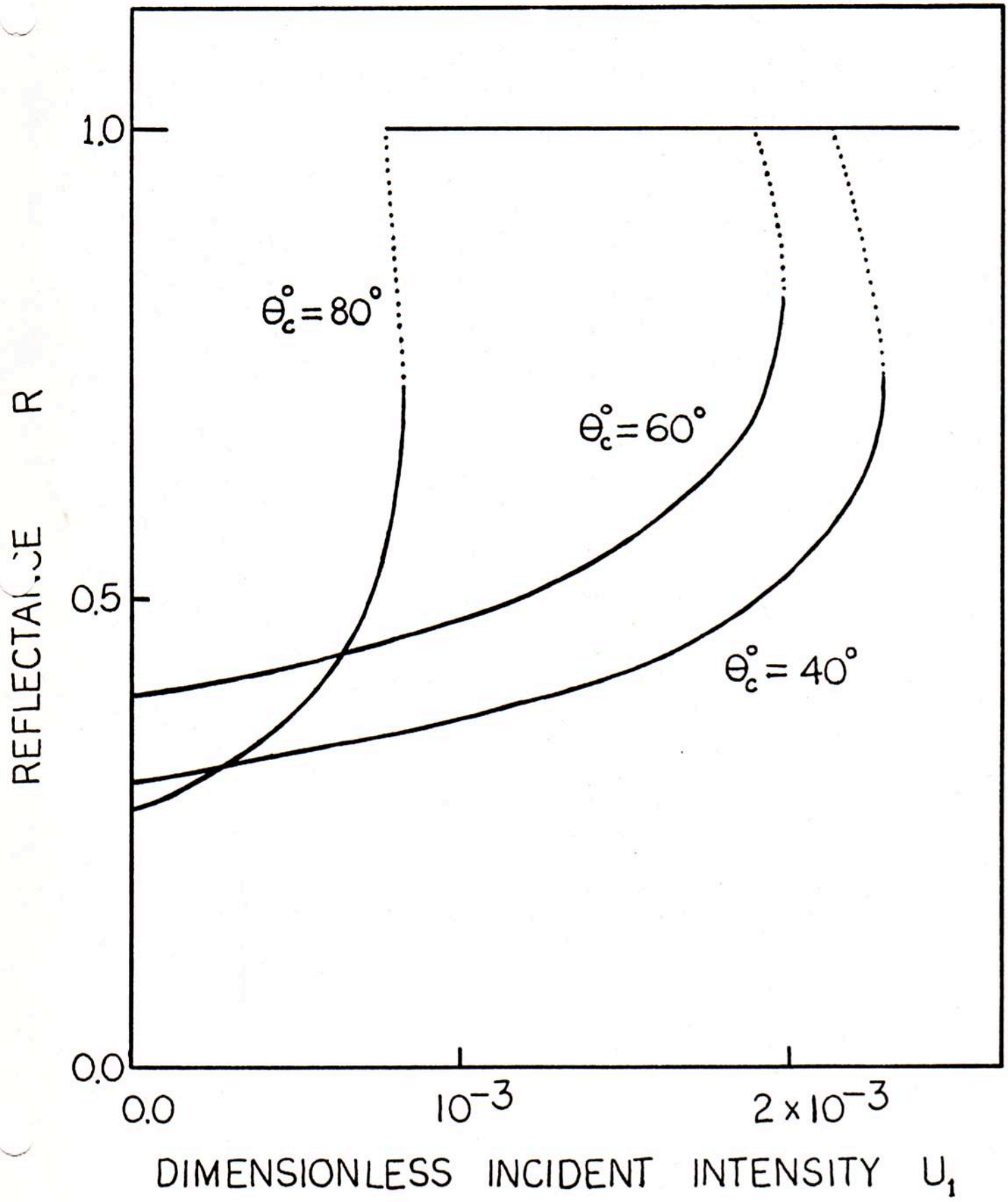
1.0
0.5
0.0
REFLECTANCE R
$\theta_c^\circ = 80^\circ$
$\theta_c^\circ = 60^\circ$
$\theta_c^\circ = 40^\circ$
0.0
10^{-3}
2×10^{-3}
DIMENSIONLESS INCIDENT INTENSITY U_1

Figure 8b.

REFERENCES

1. A.E. Kaplan, Sov. Phys. JETP, 45, 896 (1977).
2. P.W. Smith, J.-P. Hermann, W.J. Tomlinson, P.J. Maloney, Appl. Phys. Lett., 35, 846 (1979).
3. H.J. Simon, D.E. Mitchell, J.G. Watson, Amer. Jour. of Phys., 43, 630 (1975).
4. R.H. Ritchie, Surf. Sci., 34, 1 (1973).
5. M. Born and E. Wolf, Principles of Optics, (MacMillan, New York, 1965), 3rd Edition, p. 62.
6. J.D. Jackson, Classical Electrodynamics, (John Wiley & Sons, Inc., 1975) 2nd Edition, p. 237.
7. A. Otto, Z. Phys., 216, 398, (1968); Phys. Status Solidi, 42, K37, (1970).
8. C. Kittel, Introduction to Solid State Physics, (John Wiley & Sons, N.Y. 1976), 5th Edition, p. 289.
9. M. Cardona, Am. J. Phys., 39, 1277 (1971).