SEMIANNUAL REPORT FOR

Grant No. NAG 1-500

SEMICONDUCTOR SUPERLATTICE PHOTODETECTORS

January 1, 1986 to June 30, 1986

Submitted to

Dr. Ivan Clark

National Aeronautics and Space Administration
Langley Research Center, Mail Stop 473
Hampton, VA 23665

Prepared by

S. L. Chuang
Department of Electrical and Computer Engineering
University of Illinois
Urbana, IL 61801

K. Hess, J. J. Coleman and J. P. Leburton
Department of Electrical and Computer Engineering
and Coordinated Science Laboratory
University of Illinois
Urbana, IL 61801

SEMIANNUAL REPORT FOR

Grant No. NAG 1-500

SEMICONDUCTOR SUPERLATTICE PHOTODETECTORS

January 1, 1986 to June 30, 1986

Submitted to

Dr. Ivan Clark

National Aeronautics and Space Administration
Langley Research Center, Mail Stop 473
Hampton, VA 23665

Prepared by

S. L. Chuang
Department of Electrical and Computer Engineering
University of Illinois
Urbana, IL 61801

K. Hess, J. J. Coleman and J. P. Leburton
Department of Electrical and Computer Engineering
and Coordinated Science Laboratory
University of Illinois
Urbana, IL 61801

TABLE OF CONTENTS

I. INTRODUCTION

The research grant NAG 1-500 entitled "Semiconductor Superlattice Photodetectors" was awarded to the University of Illinois at Urbana-Champaign by National Aeronautics and Space Administration-Langley Research Center on June 27, 1984. The grant was continued on July 1, 1985. Dr. Ivan Clark is the Technical Officer, and Mr. John F. Royall is the Grants Officer.

This report is the fourth semiannual report.

1. Period:

   January 1, 1986 to June 30, 1986

2. Reporting Date:

   July 1, 1986.

3. Technical Personnel:

   S. L. Chuang — Assistant Professor of Electrical and Computer Engineering

   J. J. Coleman — Professor of Electrical and Computer Engineering

   K. Hess — Professor of Electrical and Computer Engineering and Research Professor of Coordinated Science Laboratory

   J. P. Leburton — Assistant Professor of Electrical and Computer Engineering and Research Assistant Professor of Coordinated Science Laboratory

   Two research assistants

II. TECHNICAL PROGRESS

During the past half-year period, our research effort centered on further investigating both the new superlattice photomultiplier with tunneling-assisted impact ionization and the photodetector based on the real space transfer mechanism. The following is a brief outline of the current status of these projects. Detailed analyses are included in Appendices A and B. New results of the tunneling-assisted impact ionization are presented in Appendix B and will be presented in a future publication.

1. Superlattice Photomultiplier

We have, recently, included the effect of tunneling in the impact ionization across the conduction-band-edge discontinuity. The results have been presented at the 1986 Device Research Conference. A copy of the viewgraphs has been included in Appendix B. On pages B2-B4, we review the motivation for our work as well as some previous work. We then discuss our work on a new superlattice photomultiplier on pages B5-B7. The effect of tunneling is discussed on pages B8-B11. It is shown that, to properly take into account the tunneling process, a numerical approach using the Airy function, rather than the WKB method used commonly in the literature, should be followed because the conventional WKB method fails when the electron energy is near the barrier height. It is shown in page B-10 that, especially on the lower energy side, tunneling lowers the threshold energy and enhances the ionization rate. When the energy of the incident electron is very high, the final energies of the two electrons will be high, thus they tend to be above the barrier and can contribute to the current without tunneling. The multiplication factor versus the

applied voltage is shown on page B12. It shows a smoother increase than that of a general avalanche photodiode. Since our device involves only one type of carrier (electrons), it works as an ideal photomultiplier and does not have the problem of the rapid increase of the multiplication factor versus the applied voltage. The detailed theoretical formulation will be shown in a future publication. On pages B13-B14, the experimental results of Capasso et al. are shown to illustrate the plausibility of our predicted phenomenon; i.e., the impact ionization across the conduction-band-edge discontinuity.

2. Superlattice Photodetector Based on the Real Space Transfer Mechanism

We are investigating the free carrier absorption process of a superlattice photodetector which makes use of the real space transfer mechanism. In particular, we have formulated and numerically computed the free carrier absorption coefficient for bulk GaAs in which we considered a second-order process involving both a photon and a phonon. Additionally, we carried out the computations for the free carrier absorption coefficient in a quantum well for the above cited second-order process involving a photon and a phonon. We are, currently, completing the numerical calculations for this case. Finally, we have begun to investigate the free carrier absorption coefficient in a quantum well in which we consider a first-order process involving only a photon.

III. FUTURE WORK

We shall apply our theory to the devices recently presented by Capasso et al. at AT&T Bell Laboratories. They have observed the predicted phenomenon of impact ionization in superlattice structures made of different materials, i.e., AlInAS/GaInAS, AlSb/GaSb, and InP/GaInAs. Though our theory has qualitatively predicted these phenomena, it will be interesting to compare the results quantitatively. The possible experimental approach will also be taken.

In the area of the superlattice photodetector based on the real space transfer mechanism, we will continue to investigate the free carrier absorption process. The numerical computations of the free carrier absorption coefficient in a quantum well involving a second-order process of a photon and a phonon will be completed. The enhancement of the free carrier absorption coefficient for the quantum well over that of bulk GaAs will be examined. Additionally, we will continue our investigation of the free carrier absorption coefficient in a quantum well wherein a first-order process, employing only a photon, is involved. Consequently, we will be able to make a comparison between the free carrier absorption coefficients in a quantum well for the first- and second-order processes and determine which of the two is dominant.

IV. PUBLICATIONS

The following manuscripts submitted for publication were supported either fully or partially by this Grant. The support by this Grant has been acknowledged in these manuscripts.

1. S. L. Chuang and K. Hess, "Impact ionization across the conduction-band-edge discontinuity of quantum-well heterostructures," J. Appl. Phys., vol. 59, pp. 2885-2894, 1986.

2. S. L. Chuang and K. Hess, "Impact ionization across the band-edge discontinuity for a superlattice photomultiplier," presented at the 1986 Device Research Conference, Amherst, Massachusetts, June 21-23, 1986.

3. S. L. Chuang, "Lateral shift of an optical beam due to leaky surface-plasmon excitations," J. Opt. Soc. Am., vol. 3, pp. 593-599, 1986.

4. S. L. Chuang, "A coupled-mode formulation by reciprocity and a variational principle," IEEE J. Lightwave Technology, accepted for publication.

V. TRAVEL

1. S. L. Chuang attended the Conference on Optical Fiber Communication and the topical meeting on Integrated and Guided Wave Optics, Atlanta, Georgia, February 23-28, 1986.

2. S. L. Chuang presented a paper entitled: "Impact ionization across the band-edge discontinuity for a superlattice photomultiplier" at the 1986 Device Research Conference, Amherst, Massachusetts, June 21-23, 1986. He also visited AT&T Bell Laboratories in Murray Hill, N.J., June 20, 1986.

APPENDIX A

Reprint of: "Impact Ionization Across the Conduction-Band-Edge Discontinuity of Quantum-Well Heterostructures" by S. L. Chuang and K. Hess, J. Appl. Phys., vol. 59, pp. 2885-2894, 1986.

Impact ionization across the conduction-band-edge discontinuity of quantum-well heterostructures

S. L. Chuang
Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

K. Hess
Department of Electrical and Computer Engineering and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

(Received 9 September 1985; accepted for publication 7 January 1986)

Impact ionization across the band-edge discontinuity of quantum-well heterostructures is studied theoretically. We consider a heterolayer structure of alternating $Al_x Ga_{1-x} As$ and GaAs layers where the GaAs layers are heavily doped with donors. Thus a large number of electrons is confined to the quantum-well region. Incident electrons are heated up by applied electric fields and collide with the electrons confined in the well regions. Both the ionization rate as a function of the incident energy, and average ionization rates are computed. Device applications of such multiple quantum-well structures and the possibility of a complete analog to the conventional photomultiplier are discussed.

I. INTRODUCTION

Novel types of photodetectors using superlattice structures have been proposed recently.[1-10] These include the enhancement of impact ionization in a superlattice for avalanche photodiodes,[1-6] an infrared detector using free carrier absorption in a heavily doped quantum-well structure,[7] a photodetector using the transient displacement current effect in a sawtooth superlattice,[8] strained-layer superlattice photodetectors,[9,10] and doping (*n-i-p-i*) photodetectors.[11,12] The operational principles of all these new devices are very interesting and additional studies are needed to characterize these photodetectors completely.

In this paper, we consider a heterolayer structure which consists of alternating $Al_x Ga_{1-x} As$ and GaAs layers where the GaAs layers are heavily doped with donors (Fig. 1). The donor are electrons confined mostly to the quantum wells. Incident electrons (which, for example, can be generated optically) are accelerated by the external field and collide with the electrons confined in the wells as shown in Fig. 1. This results in impact ionization of some of the electrons bound in the quantum wells. The wave functions for the incident electrons (three dimensional) and the confined electrons (quasi-two dimensional) are used to calculate the electron–electron interaction due to the Coulomb force with screening effects taken into account. The quantum mechanical transition rate for a system involving an initial state of one free electron above the well and one bound electron in the well and a final state of both electrons free is calculated using the Golden rule. The process is schematically displayed in Fig. 2. The electron–electron interaction in the momentum space is illustrated in Fig. 3 and will be explained in the next section. The average ionization rate over the momentum (energy) distribution of all the incident hot electrons is then calculated. Potential device applications are discussed in a final section.

Our formulation for the impact ionization for the quantum-well structure is similar to the inverse Auger effect in bulk semiconductors[13-17] or in a quantum-well heterostructure.[18-19] The new feature is that we are interested in impact ionization across the conduction-band-edge discontinuity while previous studies are considering transitions across the energy band gap. Thus, the threshold energy for impact ionization is smaller (of the order of conduction-band-edge discontinuity) than that of the interband process. The required bias voltage may be smaller since a smaller electric field than that of the conventional avalanche photodiode is needed.

II. ANALYTICAL CONSIDERATIONS

In this section, we present the detailed formulation of the impact ionization of electrons across the conduction-band-edge discontinuity in a heterolayer structure. First, the transition rate due to the electron–electron interaction is derived, then we discuss the energy distributions of the electrons confined in the quantum-well region and the incident hot electrons. Third, we present the impact ionization rate and its numerical evaluation.

A. Transition rate

Consider an incident free electron interacting with a bound electron in a quantum well; the total average transition per unit volume P_{tr} is given by[19,20]

$$P_{tr} = \frac{1}{V}\sum_{\mathbf{k}_1}\sum_{\mathbf{k}_2}\sum_{\mathbf{k}_1'}\sum_{\mathbf{k}_2'} P_{\mathbf{k}_1,\mathbf{k}_2}^{\mathbf{k}_1',\mathbf{k}_2'} f(\mathbf{k}_1) f(\mathbf{k}_2) \times [1-f(\mathbf{k}_1')][1-f(\mathbf{k}_2')], \quad (1)$$

where f is the Fermi distribution and $P_{\mathbf{k}_1,\mathbf{k}_2}^{\mathbf{k}_1',\mathbf{k}_2'}$ is the quantum mechanical transition rate (per unit time) between the initial state of electron 1 in state $\mathbf{k}_1$, electron 2 in state $\mathbf{k}_2$, and the final state of electron 1 in state $\mathbf{k}_1'$ and electron 2 in state $\mathbf{k}_2'$ (Fig. 2) due to the electron–electron interaction:

$$P_{\mathbf{k}_1,\mathbf{k}_2}^{\mathbf{k}_1',\mathbf{k}_2'} = \frac{2\pi}{\hbar}|\langle 12|H_s|1'2'\rangle|^2 \delta(E_1+E_2-E_1'-E_2'). \quad (2)$$

The square of the matrix element $|\langle 12|H_s|1'2'\rangle|^2$ is given by

 0021-8979/86/082885-10$02.40

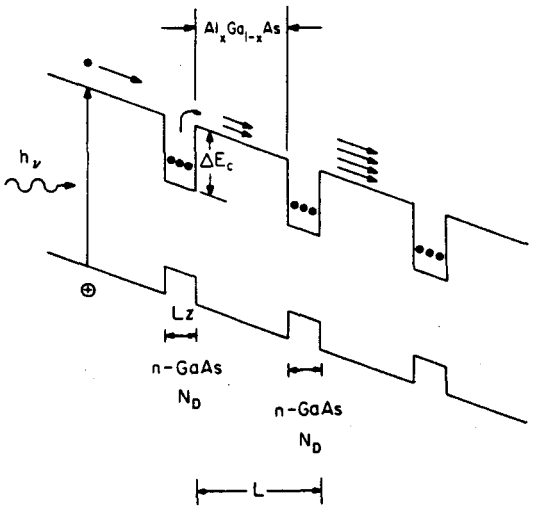


FIG. 1. Geometry of a structure exhibiting impact ionization across the band-edge discontinuity of quantum wells.

$$|\langle 12|H_s|1'2'\rangle|^2 = 4\left|\int \psi^*_{\mathbf{k}_1}(\mathbf{r}_1)\psi^*_{\mathbf{k}_2}(\mathbf{r}_2)\frac{e^2\exp(-q|\mathbf{r}_1-\mathbf{r}_2|)}{4\pi\epsilon|\mathbf{r}_1-\mathbf{r}_2|} \times \psi_{\mathbf{k}'_1}(\mathbf{r}_1)\psi_{\mathbf{k}'_2}(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2\right|^2, \tag{3}$$

for the interaction by a screened Coulomb potential. Here q is the screening parameter which is discussed in Appendix A. We have neglected the difference in the dielectric constants of the two materials which is a good approximation for AlGaAs-GaAs. The factor of 4 in Eq. (3) takes account of effects of the exchange and preservation of the spin,[14,15] considering the integrations over $\mathbf{k}'_1$ and $\mathbf{k}'_2$, and

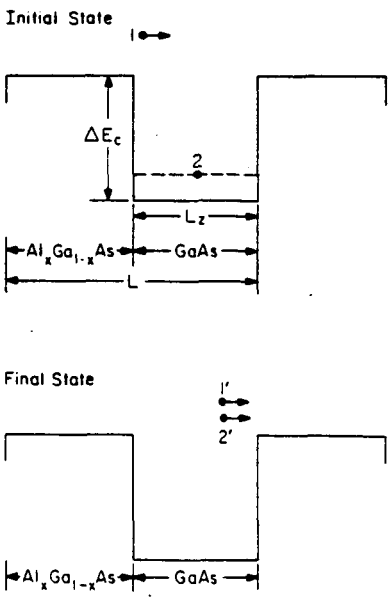


FIG. 2. Initial and final states for the electron–electron interaction in a quantum-well structure.

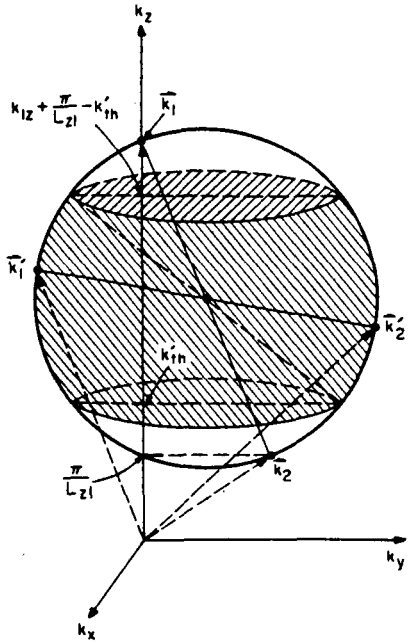


FIG. 3. A geometrical configuration in k space to illustrate the electron–electron interaction integral. With both the energy and the momentum conserved, the four electron wave vectors are on the same spherical surface. The integrations over $\mathbf{k}'_1$ and $\mathbf{k}'_2$ are restricted to the shaded portion with the z component between k'_{th} and $k_{1z}+n\pi/L_{zn}-k'_{th}$, $(n=1)$.

$$\langle \mathbf{r}_1\mathbf{r}_2|12\rangle \equiv \psi_{\mathbf{k}_1}(\mathbf{r}_1)\psi_{\mathbf{k}_2}(\mathbf{r}_2) = \frac{\exp(i\mathbf{k}_1\cdot\mathbf{r}_1)}{\sqrt{V}}\sqrt{\frac{2}{AL_{zn}}} \times \exp(i\mathbf{k}_{2t}\cdot\mathbf{r}_{2t})\sin\frac{n\pi}{L_{zn}}z_2, \tag{4}$$

$$\langle \mathbf{r}_1\mathbf{r}_2|1'2'\rangle \equiv \psi_{\mathbf{k}'_1}(\mathbf{r}_1)\psi_{\mathbf{k}'_2}(\mathbf{r}_2) = \frac{\exp(i\mathbf{k}'_1\cdot\mathbf{r}_1)}{\sqrt{V}}\frac{\exp(i\mathbf{k}'_2\cdot\mathbf{r}_2)}{\sqrt{V}}, \tag{5}$$

where $\mathbf{k}_1, \mathbf{k}'_1, \mathbf{k}'_2, \mathbf{r}_1, \mathbf{r}_2$ are the general three-dimensional vectors and $\mathbf{k}_{2t}, \mathbf{r}_{2t}$ are two-dimensional vectors in the quantum-well interface planes. The periodic parts $u_{\mathbf{k}}(\mathbf{r})$ of the Bloch functions associated with Eqs. (4) and (5) have been ignored since the overlap integral is assumed to be one for states in the conduction band. The effective quantum-well width L_{zn} is defined such that

$$E_{Qn} = \frac{\hbar^2}{2m^*}\left(\frac{n\pi}{L_{zn}}\right)^2, \tag{6}$$

where E_{Qn} is the energy level of the nth subband, which is calculated numerically taking into account the finite barrier height ΔE_c, i.e., the conduction-band-edge discontinuity (see Appendix B). For convenience, we further define

$$\mathbf{k}^{\pm}_{2n} = \mathbf{k}_{2t} \pm \hat{z}(n\pi/L_{zn}) = \hat{x}k_x + \hat{y}k_y \pm \hat{z}(n\pi/L_{zn}), \tag{7}$$

$$\mathbf{r}_2 = \mathbf{r}_{2t} + \hat{z}z_2 = \hat{x}x_2 + \hat{y}y_2 + \hat{z}z_2, \tag{8}$$

and replace the summations $\Sigma_{\mathbf{k}}$ by integrations. One then obtains

$$|\langle 12|H_s|1'2'\rangle|^2 \simeq \left(\frac{e^2}{\epsilon V}\right)^2 \left(\frac{2L_{zn}}{L}\right) \frac{1}{[|\mathbf{k}_1-\mathbf{k}_1'|^2+q^2]^2} (\delta_{\mathbf{k}_1+\mathbf{k}_2^+,\mathbf{k}_1'+\mathbf{k}_2'} + \delta_{\mathbf{k}_1+\mathbf{k}_2^+,\mathbf{k}_1'+\mathbf{k}_2'}) \tag{9}$$

and

$$P_{tr} \simeq \frac{e^4}{\epsilon^2} \frac{4\pi}{\hbar} \left(\frac{L_{zn}}{L}\right)^2 \int \frac{d\mathbf{k}_1}{(2\pi)^3} f(\mathbf{k}_1) \int \frac{d\mathbf{k}_{2t}}{(2\pi)^2 L_{zn}} f(\mathbf{k}_{2t}) [S(\mathbf{k}_1,\mathbf{k}_{2n}^+) + S(\mathbf{k}_1,\mathbf{k}_{2n}^-)]. \tag{10}$$

Here we have used the fact that $f(\mathbf{k}_1')$ and $f(\mathbf{k}_2') \ll 1$, also $f(\mathbf{k}_{2t}) = f(\mathbf{k}_2^+) = f(\mathbf{k}_2^-)$ (which will be discussed later) and

$$S(\mathbf{k}_1,\mathbf{k}_{2n}^{\pm}) \equiv \int_{\substack{k_{1z}'>k_{th}' \\ k_{2z}'>k_{th}'}} \frac{d\mathbf{k}_1'}{(2\pi)^3} \frac{\delta(E_1+E_2-E_1'-E_2')}{[|\mathbf{k}_1-\mathbf{k}_1'|^2+q^2]^2}\Bigg|_{\mathbf{k}_2'=\mathbf{k}_1+\mathbf{k}_{2n}^{\pm}-\mathbf{k}_1'}. \tag{11}$$

The lower limit in the integration for $\mathbf{k}_1'$ is due to the fact that both k_{1z}' and k_{2z}' need to be larger than a certain threshold k_{th}' in order to get both electrons 1′ and 2′ out of the quantum well. The condition $k_{2z}' > k_{th}'$ imposes another condition on k_{1z}' because of the relation

$$k_{2z}' = k_{1z} \pm \frac{n\pi}{L_{zn}} - k_{1z}' > k_{th}'.$$

One finally obtains

$$k_{th}' < k_{1z}' < k_{1z} \pm n\pi/L_{zn} - k_{th}'. \tag{12}$$

One notes that the above condition for k_{1z}' also holds for k_{2z}' obviously.

An upper bound of the integral of Eq. (11) without the constraints on k_{1z}' and k_{2z}' has been obtained analytically for a parabolic band structure.[21]

$$S_u(\mathbf{k}_1,\mathbf{k}_{2n}^{\pm}) \equiv \int_{\text{all } \mathbf{k}_1'} \frac{d\mathbf{k}_1'}{(2\pi)^3} \frac{\delta(E_1+E_2-E_1'-E_2')}{[|\mathbf{k}_1-\mathbf{k}_1'|^2+q^2]^2}\Bigg|_{\mathbf{k}_2'=\mathbf{k}_1+\mathbf{k}_{12n}^{\pm}-\mathbf{k}_1'} = \frac{m^*}{8\pi^2\hbar^2 q^2} \frac{|\mathbf{k}_1-\mathbf{k}_{2n}^{\pm}|}{|\mathbf{k}_1-\mathbf{k}_{2n}^{\pm}|^2+q^2}. \tag{13}$$

The geometry pertinent to the evaluation of $S(\mathbf{k}_1,\mathbf{k}_2)$ and its upper bound $S_u(\mathbf{k}_1,\mathbf{k}_2)$ is shown in Fig. 3 for

$$\mathbf{k}_2 \equiv \mathbf{k}_{2n}^+ = \mathbf{k}_{2t} + \hat{z}(n\pi/L_{zn}) (n=1).$$

Due to energy and momentum conservations, the four vectors $\mathbf{k}_1$, $\mathbf{k}_2$, $\mathbf{k}_1'$, and $\mathbf{k}_2'$ lie on a spherical surface, assuming that the $E-\mathbf{k}$ relations are parabolic for all four vectors.[22] Due to the above mentioned constraints on both $\mathbf{k}_{1z}'$ and $\mathbf{k}_{2z}'$, the $S(\mathbf{k}_1,\mathbf{k}_2)$ integral will be over the unshaded spherical surface in Fig. 3 for which

$$k_{th}' < k_{1z}' < k_{1z} + \pi/L_{zt} - k_{th}',$$

while the integral for $S_u(\mathbf{k}_1,\mathbf{k}_2)$ is over the whole spherical surface. The latter integration can be performed analytically by a coordinate transformation to the polar axis $\mathbf{k}_1 - \mathbf{k}_2$, and results in Eq. (13).

B. Electron distribution function $f(\mathbf{k}_2)$ in the quantum-well region

The electron distribution function in the quantum-well region $f(\mathbf{k}_2^{\pm})$ or $f(\mathbf{k}_{2t})$ is taken as

$$f(\mathbf{k}_{2t}) = \begin{cases} 1 & \text{if } 0 < k_{2t} < K_{2t}^{(n)} \\ 0 & \text{otherwise,} \end{cases} \tag{14}$$

where

$$K_{2t}^{(n)} = \sqrt{(2m^*/\hbar^2)E_F - (n\pi/L_{zn})^2} \tag{15}$$

is the maximum of k_{2t} for electrons in the nth subband.

The electron concentration is given by

$$n = \int_{E_{Q1}}^{\infty} f(E)g(E)dE = \sum_n \frac{m^*}{\pi\hbar^2 L_z} k_B T \ln(1 + e^{(E_F - E_{Qn})/k_B T}), \tag{16}$$

where E_{Qn} is the nth quantized energy level of electron 2 as given in Eq. (6). Defining a characteristic concentration n_c,

$$n_c \equiv m^* k_B T/\pi\hbar^2 L_z. \tag{17}$$

We have for $E_F - E_{Qn} \gg k_B T$,

$$\frac{n}{n_c} = \sum_{\substack{\text{occupied} \\ \text{subbands } n}} (E_F - E_{Qn})/k_B T. \tag{18}$$

From $n \approx N_D$ and the above relation or Eq. (16), we obtain E_F, and thus $K_{2t}^{(n)}$.

C. Electron distribution function $f(\mathbf{k}_1)$ for the incident hot electrons

We assume that the incident hot electrons in the $Al_x Ga_{1-x} As$ region (region a) is Maxwellian in shape, i.e., the spherically symmetrical part is given by the expression

$$f(\mathbf{k}_0) = 4n_0 (\pi\hbar^2/2m_a^* k_B T_e)^{3/2} e^{-E_0(\mathbf{k}_0)/k_B T_e}, \tag{19}$$

with an electron temperature T_e, where m_a^* is the effective mass of the electron in the conduction band in the $Al_x Ga_{1-x} As$ region and the index 0 characterizes the electrons in this region. Since the energy

$$E_0 = (\hbar^2/2m_a^*)(k_{0x}^2 + k_{0y}^2 + k_{0z}^2) \tag{20}$$

is measured from the conduction-band edge in the $Al_xGa_{1-x}As$ region, and

$$\mathbf{k}_0 = k_{0x}\hat{x} + k_{0y}\hat{y} + k_{0z}\hat{z}, \tag{21}$$

the energy E_1 and $\mathbf{k}_1$ vector in the GaAs region (region b) are related to E_0 and $\mathbf{k}_0$ by the equations

$$k_{1x} = k_{0x}, \tag{22}$$

$$k_{1y} = k_{0y}, \tag{23}$$

$$(\hbar^2/2m_a^*)k_0^2 + \Delta E_c = (\hbar^2/2m_b^*)k_1^2, \tag{24}$$

where m_b^* is the average effective mass of electrons in the Γ valley of the GaAs region. (We use the average effective mass m_b^* to include the nonparabolicity effect. The effective mass m^* used in the previous equations is replaced by m_b^* in the computations. (See Appendix C.) Instead of integrating over $d\mathbf{k}_1$, we integrate over $d\mathbf{k}_0$,

$$\int \frac{d\mathbf{k}_1}{(2\pi)^3} f(\mathbf{k}_1) \rightarrow \int \frac{d\mathbf{k}_0}{(2\pi)^3} f_0(\mathbf{k}_0).$$

D. Average impact ionization rate

The average ionization rate $\langle 1/\tau \rangle$ is defined as the total transition rate per incident electron, and therefore

$$\langle 1/\tau \rangle = P_{tr}/n_0 = \int \frac{d\mathbf{k}_0}{(2\pi)^3} 4\left(\frac{\pi\hbar^2}{2m_a^* k_B T_e}\right)^{3/2} \exp\left(-\frac{\hbar^2}{2m_a^* k_B T_e}(k_{0x}^2 + k_{0y}^2 + k_{0z}^2)\right)\frac{1}{\tau(E_0)}$$

$$\simeq \int \frac{dk_{0z}}{2\pi}\left(\frac{\pi\hbar^2}{2m_a^* k_B T_e}\right)^{1/2} \exp\left(-\frac{\hbar^2 k_{0z}^2}{2m_a^* k_B T_e}\right)\frac{1}{\tau(E_{0z})}, \tag{25}$$

where

$$\frac{1}{\tau(E_{0z})} = \frac{e^4}{\epsilon^2}\frac{4\pi}{\hbar}\left(\frac{L_{zn}}{L}\right)^2 \int \frac{d\mathbf{k}_{2t}}{(2\pi)^2 L_{zn}} f(\mathbf{k}_{2t})\left[S(\mathbf{k}_1,\mathbf{k}_{2n}^+) + S(\mathbf{k}_1,\mathbf{k}_{2n}^-)\right], \tag{26}$$

which is the ionization rate for an incident hot electron with the energy

$$E_0 = E_{0z} = (\hbar^2/2m_a^*)k_{0z}^2 \tag{27}$$

and

$$k_{0x} = k_{0y} = 0. \tag{28}$$

Since the integrand in Eq. (25) is sharply peaked near $k_{0x} = k_{0y} \simeq 0$, we have used the approximations

$$\frac{1}{\tau(E_0)} \simeq \frac{1}{\tau(E_0)}\bigg|_{\substack{k_{0x}=0 \\ k_{0y}=0}} \equiv \frac{1}{\tau(E_{0z})}, \tag{29}$$

and carried out the dk_{0x} and dk_{0y} integrations in Eq. (25). By doing so, the incident electrons are essentially z directed.

$$\mathbf{k}_0 = \hat{z}k_{0z}. \tag{30}$$

Thus, for the transmitted electron we have

$$\mathbf{k}_1 = \hat{z}k_{1z} \tag{31}$$

and

$$(\hbar^2/2m_a^*)k_{0z}^2 + \Delta E_c = (\hbar^2/2m_b^*)k_{1z}^2. \tag{32}$$

The average ionization rate $\langle 1/\tau \rangle$ is obtained by integrating Eq. (25) using Eq. (26). An upper bound of $\langle 1/\tau \rangle$, $\langle 1/\tau \rangle_u$, is obtained using the upper bound of $1/\tau(E_0)$. Dropping the subscript z and using E_0 instead of E_{0z}, one obtains for the integrand of Eq. (25)

$$\frac{1}{\tau(E_0)_u} \equiv \frac{e^4}{\epsilon^2}\frac{4\pi}{\hbar}\left(\frac{L_{zn}}{L}\right)^2 \int \frac{d\mathbf{k}_{2t}}{(2\pi)^2 L_{zn}} f(\mathbf{k}_{2t}) \times \left[S_u(\mathbf{k}_1,\mathbf{k}_{2n}^+) + S_u(\mathbf{k}_1,\mathbf{k}_{2n}^-)\right]. \tag{33}$$

Next we first discuss the exact numerical integrations of $1/\tau(E_0)$ and $\langle 1/\tau \rangle$ in Eqs. (25) and (26). Then formulas for upper bounds, $1/\tau(E_0)_u$, which lead to analytical expressions for $\langle 1/\tau \rangle_u$ are presented.

III. THE NUMERICAL INTEGRATIONS

A. Integrations $1/\tau(E_0)$ and $\langle 1/\tau \rangle$ from Eqs. (25) and (26)

To obtain $1/\tau(E_0)$, we need to evaluate the electron–electron interaction integrals $S(\mathbf{k}_1,\mathbf{k}_{2n}^{\pm})$ in Eq. (11). For the following discussion, we use $\mathbf{k}_2 \equiv \mathbf{k}_{2n}^+$. The integral $S(\mathbf{k}_1,\mathbf{k}_2^-)$ is obtained similarly, replacing $\mathbf{k}_{2n}^+$ by $\mathbf{k}_{2n}^-$, or $n\pi/L_{zn}$ by $-n\pi/L_{zn}$. For a parabolic band structure, and using

$$\mathbf{k}_1' = \mathbf{k}_{1t}' + \hat{z}k_{1z}', \tag{34}$$

$$\mathbf{k}_2' = \mathbf{k}_{2t}' + \hat{z}k_{2z}', \tag{35}$$

the $\delta(E_1 + E_2 - E_1' - E_2')$ term in the integrand of Eq. (11) can be rewritten as

$$(m_b^*/\hbar^2)\,\delta[(k_{1z}' - k_{1z}'^{\alpha})(k_{1z}' - k_{1z}'^{\beta})], \tag{36}$$

where the two roots for k_{1z}' are

$$k_{1z}'^{\alpha} \equiv \frac{k_{1z} + n\pi/L_{zn}}{2} + \sqrt{\left(\frac{k_{1z} - n\pi/L_{zn}}{2}\right)^2 + \mathbf{k}_{1t}' \cdot \mathbf{k}_{2t} - k_{1t}'^2}, \tag{37}$$

$$k_{1z}'^{\beta} \equiv \frac{k_{1z} + n\pi/L_{zn}}{2} - \sqrt{\left(\frac{k_{1z} - n\pi/L_{zn}}{2}\right)^2 + \mathbf{k}_{1t}' \cdot \mathbf{k}_{2t} - k_{1t}'^2}. \tag{38}$$

Therefore, the k_{1z}' integration can be performed analytically using the properties of the delta functions, and k_{1z}' is restricted as given in Eq. (12). The other integrations can be simplifed as

$$\int d\mathbf{k}_{2t} \int d\mathbf{k}'_{1t} = \int_0^{K_{2t}^{(n)}} k_{2t}\, dk_{2t} \times \int_0^{K'_{1t}} k'_{1t}\, dk'_{1t}\, 2\pi \int_0^{2\pi} d\phi_{12}, \tag{39}$$

where the two angular variables ϕ_{2t} for $\mathbf{k}'_{2t}$ and ϕ'_{1t} for $\mathbf{k}'_{1t}$ have been transformed to only one angular variable ϕ_{12} which is the angle between $\mathbf{k}'_{1t}$ and $\mathbf{k}_{2t}$, with a 2π factor added as a result of this transformation because the integrand depends on ϕ_{12} only. The upper limit of k'_{1t} variable K'_{1t} is obtained from the condition that k'^{α}_{1z} and k'^{β}_{1z} in Eqs. (37) and (38) should be real, i.e.,

$$\left(\frac{k_{1z} - n\pi/L_{zn}}{2}\right)^2 + k'_{1t} k_{2t} \cos\phi_{12} - k'^2_{1t} > 0. \tag{40}$$

This condition is easily included in the numerical code. If it is violated, k'^{α}_{1z} and k'^{β}_{1z} become imaginary and the integrand vanishes because the variable of integration k'_{1z} is real. Explicit solutions of Eq. (40) can also be used, which give the upper limit for k'_{1t},

$$K'^{+}_{1t} = \frac{k_{2t}}{2} + \sqrt{\left(\frac{k_{1z} - n\pi/L_{zn}}{2}\right)^2 + \left(\frac{k_{2t}}{2}\right)^2}. \tag{41}$$

Then $1/\tau(E_0)$ can be written as

$$\frac{1}{\tau(E_0)} = \frac{e^4 m_b^*}{\epsilon^2 \hbar^3} \left(\frac{L_{zn}}{L}\right)^2 \frac{1}{4\pi^3 L_{zn}} \int_0^{K_{2t}^{(n)}} k_{2t}\, dk_{2t} \int_0^{2\pi} d\phi_{12} \left(\int_0^{K'^{+}_{1t}} k'_{1t}\, dk'_{1t} \int_0^{2\pi} d\phi_{12}\, F^+(k'_{1t}, k_{2t}, \phi_{12}) + \int_0^{K'^{-}_{1t}} k'_{1t}\, dk'_{1t} \int_0^{2\pi} d\phi_{12}\, F^-(k'_{1t}, k_{2t}, \phi_{12}) \right), \tag{42}$$

where

$$F^+(k'_{1t}, k_{2t}, \phi_{12}) \equiv \frac{1}{2\left[\left(\frac{k_{1z} - n\pi/L_{zn}}{2}\right)^2 + k'_{1t} k_{2t} \cos\phi_{12} - k'^2_{1t}\right]^{1/2}} \times \left(\frac{1}{[k'^2_{1t} + (k_{1z} - k'^{\alpha}_{1z})^2 + q^2]^2} + \frac{1}{[k'^2_{1t} + (k_{1z} - k'^{\beta}_{1z})^2 + q^2]^2} \right) \tag{43}$$

for k'^{α}_{1z} and k'^{β}_{1z} between k'_{th} and $k_{1z} + (n\pi/L_{zn}) - k'_{th}$, and zero for the appropriate terms otherwise. $F^-(k'_{1t}, k_{2t}, \phi_{12})$ and K'^{-}_{1t} are obtained by replacing $n\pi/L_{zn}$ by $-n\pi/L_{zn}$ in the F^+ function as well as in K'^{+}_{1t} from Eqs. (41) and (43).

The last integration $\int dk_{0z}$ in $\langle 1/\tau \rangle$ must be performed from a minimum $k^{\pm}_{0z,\min}$ which is obtained from Eq. (12),

$$k_{1z} > 2k'_{th} \mp n\pi/L_{zn}, \tag{44}$$

and the energy conservation relation Eq. (32),

$$k^{\pm}_{0z,\min} = \sqrt{\frac{m_a^*}{m_b^*}\left(2k'_{th} \mp \frac{n\pi}{L_{zn}}\right)^2 - \frac{2m_a^*}{\hbar^2}\Delta E_c}. \tag{45}$$

B. The upper bounds of the integrals $1/\tau(E_0)_u$ and $\langle 1/\tau \rangle_u$ using Eqs. (25) and (33)

The analytical result for the upper bound of $1/\tau(E_0)$ is

$$\frac{1}{\tau(E_0)_u} = \frac{e^4 m_b^*}{\epsilon^2 \hbar^3} \left(\frac{L_{zn}}{L}\right)^2 \frac{1}{4\pi^2 q L_{zn}} (G^+ + G^-) \equiv \frac{1}{\tau(E_0)}\bigg|_u^+ + \frac{1}{\tau(E_0)}\bigg|_u^-, \tag{46}$$

$$G^{\pm} = \sqrt{\left(\frac{K_{2t}}{q}\right)^2 + \left(\frac{k_{1z} \mp n\pi/L_{zn}}{q}\right)^2} - \left(\frac{k_{1z} \mp n\pi/L_{zn}}{q}\right) - \tan^{-1}\sqrt{\left(\frac{K_{2t}}{q}\right)^2 + \left(\frac{k_{1z} \mp n\pi/L_{zn}}{q}\right)^2} + \tan^{-1}\left(\frac{k_{1z} \mp n\pi/L_{zn}}{q}\right). \tag{47}$$

Here the G^+ term is due to the term with $\mathbf{k}^+_{2n}$, and the G_2^- term due to that with $\mathbf{k}^-_{2n}$ in Eq. (33), and

$$\langle 1/\tau \rangle_u = \int_{k^+_{0z,\min}}^{\infty} \frac{dk_{0z}}{2\pi} \left(\frac{\pi\hbar^2}{2m_a^* k_B T_e}\right)^{1/2} \exp\left(-\frac{\hbar^2 k^2_{0z}}{2m_a^* k_B T_e}\right) \frac{1}{\tau(E_0)}\bigg|_u^+ + \int_{k^-_{0z,\min}}^{\infty} \frac{dk_{0z}}{2\pi} \left(\frac{\pi\hbar^2}{2m_a^* k_B T_e}\right)^{1/2} \exp\left(-\frac{\hbar^2 k^2_{0z}}{2m_a^* k_B T_e}\right) \frac{1}{\tau(E_0)}\bigg|_u^-, \tag{48}$$

which can be integrated numerically by Simpson's rule.

IV. NUMERICAL RESULTS AND DISCUSSIONS

The direct numerical integrations of the ionization rate $1/\tau(E_0)$ and the average ionization rate $\langle 1/\tau \rangle$ in Eq. (25) is straightforward. The upper bounds of the ionization rate and the exact numerically integrated results as functions of the energy of the incident hot electrons are shown in Figs. 4(a) and 4(b). One can see that the ionization rate increases rapidly when the energy E_0 is above the threshold value E_i

$$E_i = \frac{\hbar^2}{2m_a^*} k_{0z,\min}^{+2} = \frac{\hbar^2}{2m_b^*}\left(2k'_{\rm th} - \frac{n\pi}{L_{zn}}\right)^2 - \Delta E_c, \quad (49)$$

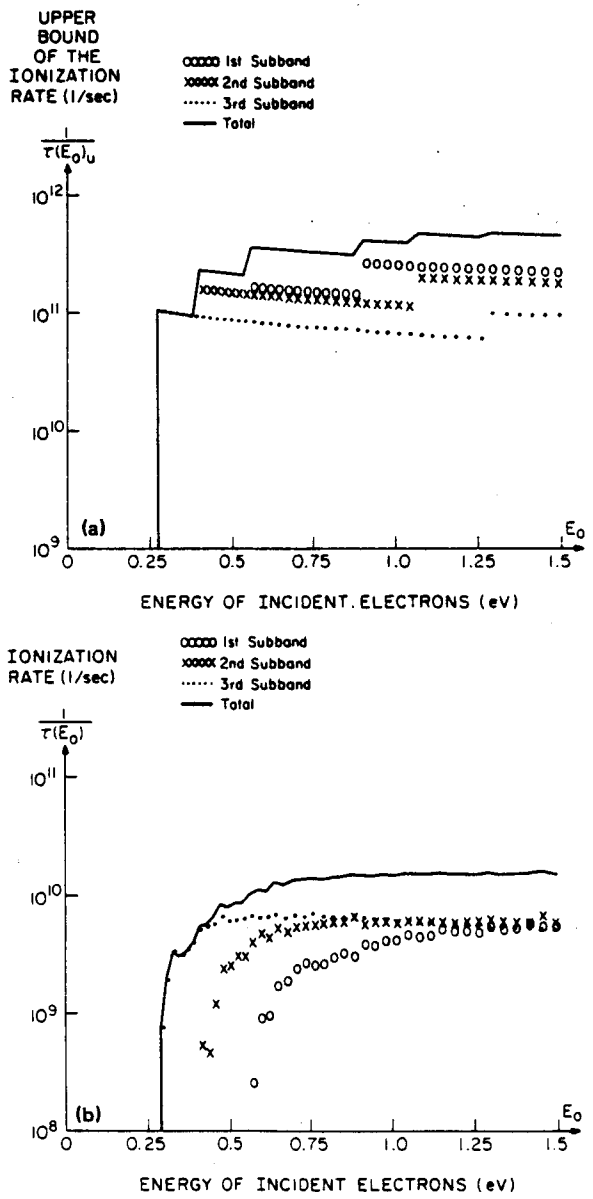


FIG. 4. (a) Upper bound of the impact ionization rates $1/\tau(E_0)_u$ as functions of the incident hot electron energy $E_0(\equiv E_{0z})$ for each subband and the sum of all three subbands. The parameters used are $L_z = 200$ Å, $L = 1200$ Å, $N_D = 5\times10^{18}\,\text{cm}^{-3}$, $x = 0.35$, $T = 77$ K, $T_e = 2300$ K ($m_a^* = 0.09m_0$, $m_b^* = 0.105m_0$, $\Delta E_c = 0.262$ eV). (b) Exact impact ionization rates from numerical integration of Eqs. (25) and (42). The same material parameters as those in (a) are used here. $\langle 1/\tau \rangle_u = 4.4\times10^9$/s using Eq. (48), and $\langle 1/\tau \rangle = 1.4\times10^8$/s using Eqs. (25) and (42).

for each subband $n = 1,2,3$. The overall minimum threshold energy for impact ionization will be due to the highest occupied subband of course. For a quantum-well size $L_z = 200$ Å, Al mole fraction $x = 0.35$ in the $Al_xGa_{1-x}As$ region, and $N_D = 5\times10^{18}\,\text{cm}^{-3}$, there are three occupied subbands among the six quantized levels in the well. The Fermi level is 0.1092 eV measured from the conduction-band edge in the GaAs region. The threshold value for $k'_{\rm th}$ (the minimum value of k'_{1z} or k'_{2z} for electrons 1′ and 2′ to get out of the quantum well) is given by

$$k'_{\rm th} = \sqrt{2m_b^*(\Delta E_c - \delta)}/\hbar, \quad (50)$$

where δ accounts for the tilting of the quantum well due to the large applied field (Fig. 1). Without space-charge effects we have

$$\delta \simeq eFL_z, \quad (51)$$

where e is the magnitude of the electron charge. In all our computations we assume a field strength $F = 10$ kV/cm and a corresponding electron temperature of 2300 K at a lattice temperature of $T = 77$ K. (No accurate data for T_e in $Al_xGa_{1-x}As$ are available to our knowledge. The above value is a result of Monte Carlo simulations in GaAs. In Ref. 23 there are some discussions on F-T_e relations in $Al_xGa_{1-x}As$.) The necessary parameters used here are taken from Refs. 24 and 25.

The average ionization rate $\langle 1/\tau \rangle$ involves the integration over the energy E_0 and the distribution function. The upper bound of the impact ionization rate $1/\tau(E_0)_u$ as a function of the energy E_0 is shown in Fig. 4(a) for each subband and also for all subbands combined. These curves are much easier and faster to generate than the results of the exact numerical integration which are shown in Fig. 4(b) since the results in Eq. (46) are analytical expressions. They also can serve as a means of checking the numerical accuracy of the exact numerical integral in Eq. (42). This is because $1/\tau(E_0)_u$ from Eq. (46) is the upper bound of the numerical integral of $1/\tau(E_0)$ in Eq. (42). $1/\tau(E_0)_u$ is the integration over the whole spherical surface in Fig. 3, for which all $\mathbf{k}'_1$ and $\mathbf{k}'_2$ of the two electrons are allowed everywhere on the spherical surface. $1/\tau(E_0)$ is the integration over only a portion of the spherical surface for which the condition

$$k'_{\rm th} < k'_{1z} < k'_{1z} + n\pi/L_{zn} - k'_{\rm th}$$

is satisfied. If one relaxes this condition in Eq. (43), the results of integration (42) will be over the whole spherical surface, which is the upper bound. Thus a numerical test in Eq. (42) by simply removing the constraints on k'_{1z} can be checked exactly with the analytical expression in Eq. (46). This process determines the number of discretized intervals for each integration variable in Eq. (42). One then adds the constraints on both k'_{1z} and k'_{2z} to obtain the exact numerical integrals.

A. The effect of multiple subbands

As shown in Figs. 4(a) and 4(b), the ionization rates due to higher level subbands are clearly very important because electrons in those subbands are in higher energy levels and thus easier ionized. Figure 4(a) shows the upper bounds

of the ionization rates for each subband and the total contribution is shown as the solid curve. The total rate in Fig. 4(a) has a discontinuity whenever either a new subband with a term corresponding to either $+n\pi/L_{zn}$ or $-n\pi/L_{zn}$ term comes in. The results for the exact numerical integrations for $1/\tau(E_0)$ are shown in Fig. 4(b) and the curves look smoother than those in Fig. 4(a). (The small fluctuations in the curves of Fig. 4(b) are due to numerical errors in the multiple integrations.) The electrons in the highest subband require the lowest threshold energy for impact ionization and there are more incident hot electrons with enough energy to impact ionize these electrons.

B. The effect of the quantum-well width L_z

In Figs. 5(a) and 5(b), we decrease L_z from 200 to 80 Å keeping the doping concentration N_D the same (5.0×10^{18} cm^{-3}). The upper bounds of the ionized rates are shown in Fig. 5(a) and the exact numerical results are shown in Fig. 5(b). There are three subbands for this structure and only the lowest one is occupied at this doping level. The ionization rate is smaller than the previous case in Fig. 4(b). Roughly speaking, it is because the number of electrons per unit area in the quantum-well region $N_D L_z$ is smaller when L_z is decreased and N_D kept the same. Although the ground level of a thinner quantum well is higher than that of a thicker well in Fig. 4(b), the higher subbands in Fig. 4(b) also contribute to the total ionization rate and thus the ionization rate in the case in Fig. 4(b) is larger.

C. The effect of the band-edge discontinuity

In Figs. 6(a) and 6(b), we show the upper bounds of the ionization rates for all subbands and the results of the numerical integrations by decreasing the Al mole fraction x from 0.35 to 0.25 with L_z the same as that in Figs. 4(a) and 4(b). We use the 60%–40% rule[26,27] for the conduction and valence-band-edge discontinuities in our calculation. If one uses a different rule with a number between 60% and 85% for the conduction-band-edge discontinuity, the ionization rate will decrease because the well is deeper, Once ΔE_c is determined, the threshold energy for impact ionization can be obtained. Decreasing x will decrease ΔE_c, thus the threshold energy is lowered and the impact ionization rate is increased. This is clearly shown in Figs. 6(a) and 6(b) as compared to Figs. 4(a) and 4(b).

D. The effect of doping and the total ionization rate

If one increases the doping concentration in the quantum wells, one expects that the ionization rate will increase because there are more electrons available for impact ionization. This is shown in Fig. 7 where the dashed line is for the upper bound of the ionization rate and the solid line is from exact numerical integration. Increasing N_D will increase the number of occupied subbands also since the Fermi level will be raised higher. For each N_D, the computation takes about 1500 cpu seconds as an average in the Cyber 175 Computer using about 1.6×10^6 number of discretized points (40 for variable k_{2t}, 40 for k'_{1t}, 20 for ϕ_{12}, and 50 for k_{0z}). When the Fermi level is raised, there will be more electrons in the higher subbands also and they are easier to be ionized since the minimum threshold energy is decreased. The other effect is that a higher doping concentration may also increase the tunneling probability of the bound electrons. This may also enhance the ionization rate.

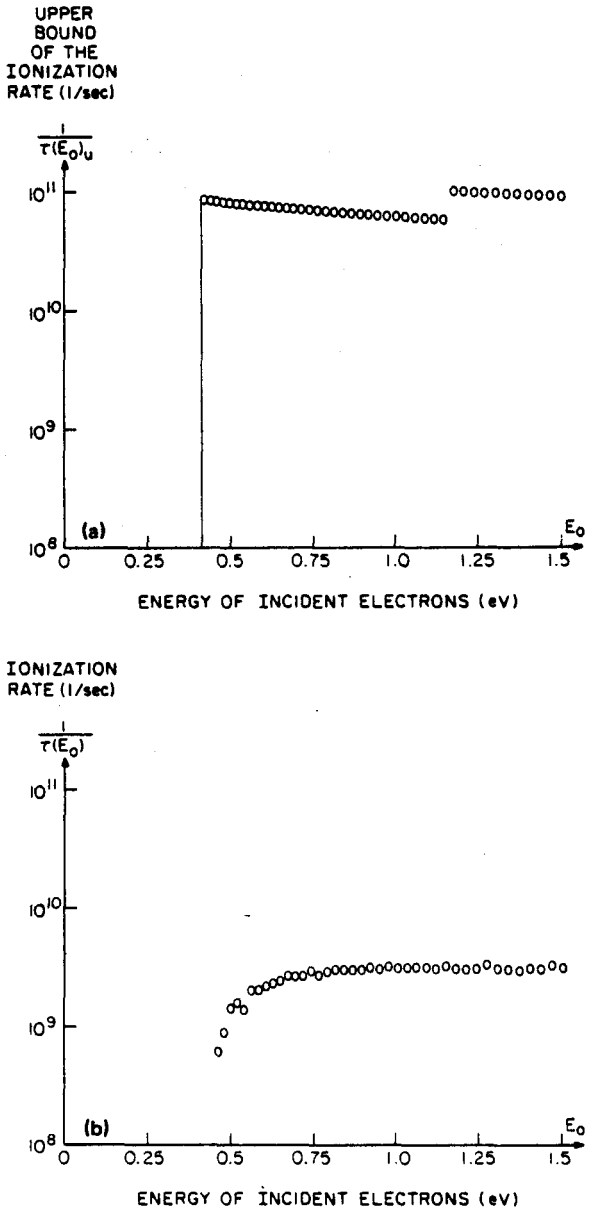


FIG. 5. (a) Upper bound of the impact ionization rate as a function of the incident hot electron energy E_0 for a different quantum-well width $L_z = 80$ Å. ($L = 1200$ Å, $x = 0.35$, $N_D = 5\times10^{18}$ cm^{-3}, $m_a^* = 0.09m_0$, $m_b^* = 0.105m_0$, $\Delta E_c = 0.262$ eV.) There is only one occupied subband at this doping level. (b) Exact impact ionization rate as a function of the incident hot electron energy E_0 obtained from numerical integration of Eq. (25) using Eq. (42). The same parameters as those in (a) are used here. $\langle 1/\tau\rangle_u = 6.7\times10^8$/s, and $\langle 1/\tau\rangle = 1.6\times10^7$/s.

V. POTENTIAL FOR DEVICE APPLICATIONS

Avalanche photodioes which in their operational characteristic come close to the ideal of the conventional photomultiplier have been proposed by Capasso[2–6] and have been termed solid-state photomultipliers. The principle of these

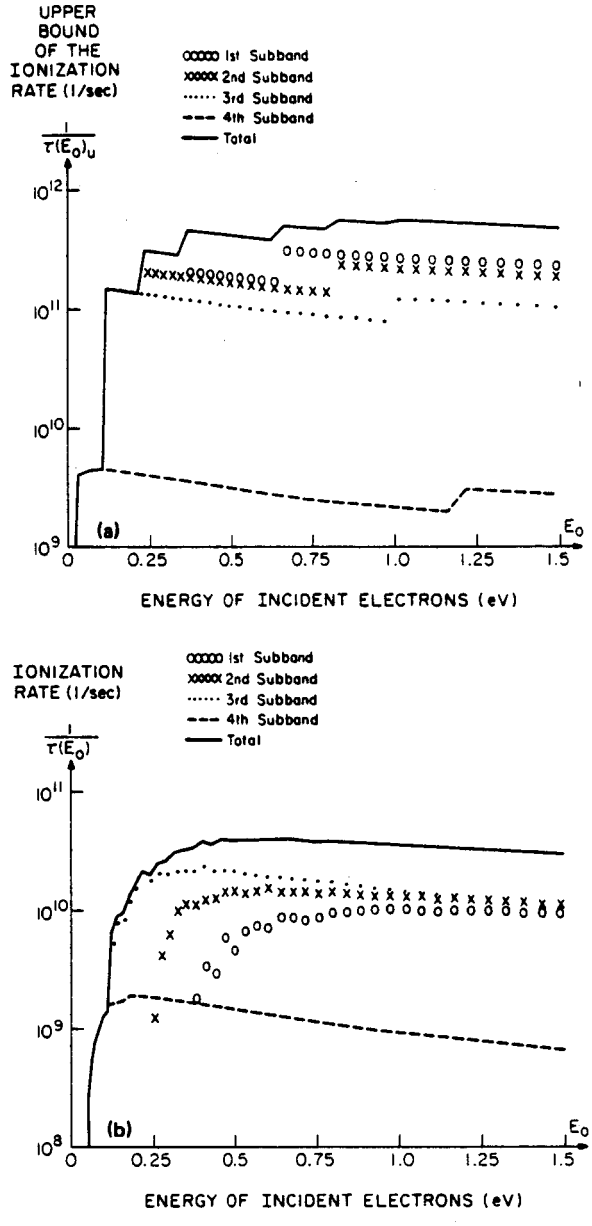


FIG. 6. (a) Upper bound of the impact ionization rates as functions of the incident electron energy E_0 for a smaller Barrier height ($x = 0.25$, $\Delta E_c = 0.187$ eV) than that in Fig. 4. We use $m_a^* = 0.105 m_0$, $m_b^* = 0.083 m_0$ for this value of x and ΔE_c. The other parameters are $L_z = 200$ Å, $L = 1200$ Å, $N_D = 5\times10^{18}$ cm^{-3}. (b) Impact ionization as functions of the incident electron energy E_0. Same parameters are used as in (a). $\langle 1/\tau\rangle_u = 1.8\times10^{10}$/s and $\langle 1/\tau\rangle = 1.5\times10^9$/s.

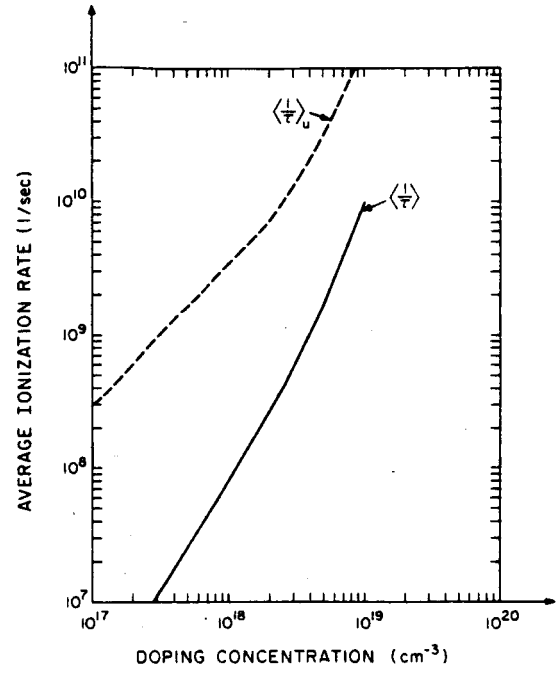


FIG. 7. Average ionization rate $\langle 1/\tau\rangle$ (solid curve) and its upper bound $\langle 1/\tau\rangle_u$ (dashed curve) as functions of the doping concentration N_D in the quantum well. We use $L_z = 200$ Å, $L = 1200$ Å, $x = 0.25$, $T = 77$ K, $T_e = 2300$ K. The parameters m_a^*, m_b^*, and ΔE_c are the same as in Fig. 6(a).

devices is based on a highly asymmetric ionization coefficient for electrons (α) and holes (β). A conventional photomultiplier does not, of course, involve any holes at all. The question arises therefore if the new technological possibilities which are offered by MOCVD and MBE crystal growth techniques can be used to construct the complete solid-state analogy of a photomultiplier. This question was one of the driving forces of our calculation in the previous sections. Instead of photomultiplier electrodes one would use quantum wells and the secondary electrons would be emitted into the neighboring AlGaAs instead of into vacuum. The results for the ionization rate in Figs. 4–7 indeed show that appreciable gain can be achieved if the doping of the quantum well is $\sim 10^{19}$ cm^{-3} and the number of layers is larger than ~ 100. From Fig. 7, the average ionization rate varies from 10^7 to 10^{10} 1/s. Thus the electron ionization coefficients may vary from around 1 to 10^3/cm or higher assuming the average electron velocity is 10^7 cm/s or lower. The ionization rate can also be increased by decreasing the band-edge discontinuity. Primary electrons can be generated by the absorption of infrared radiation either by the electrons in the valence band (interband absorption as shown in Fig. 1) or by the quantum-well electrons which make a transition from bound to propagating states. The latter type of free carrier absorption is of first order (the quantum well provides the momentum and phonons are not necessary) and therefore significantly stronger than the free carrier absorption in bulk semiconductors. This fact has been noted by Chiu *et al.*[28] who also show that multiquantum-well structures (similar to our structure in Fig. 1, except that they also dope the GaAlAs regions) can be used as infrared detectors.[7] Chiu *et al.* did not, however, investigate the possibility of gain by the avalanche process. But, the possibility of varying the band-edge discontinuity (and therefore the range of infrared detection almost at will in the range of energies $0 < \hbar\omega \lesssim 0.4$ eV) and the possibility of gain (analogous to the conventional photomultiplier) are enormously attractive features.

There are, however, also enormous technological problems which need to be overcome to realize such a device. Since holes are not involved, the electrons which are ionized out of the well need to be replenished. This means that separate ohmic contacts have to be applied to a number of sets of quantum wells without changing the high resistivity of the AlGaAs layers in between. It may be possible to accomplish this complicated task with selective Schottky barrier con-

tacts which would be ohmic for the highly doped quantum wells. Another severe problem is the requirement of highly doped quantum wells neighboring to essentially depleted and highly resistive AlGaAs. Both problems are not basic in nature and do not seem out of reach considering the recent progress in microstructure technology. However, more studies (including also the effects of field ionization) are needed to confirm and show the device potential of such structures.

Independently, Capasso has recently performed experiments and has obtained encouraging evidence for impact ionization of the kind discussed in this paper. These experiments may call for an extension of our theory to include effects such as the tunneling-supported impact ionization[29,30] due to the band bending as shown in Fig. 1. The bound electrons in the quantum well may acquire a small amount of energy by collision with an incident hot electron and subsequently tunnel into the conduction band in the AlGaAs region. This process reduces the ionization threshold energy in each subband. Thus the ionization rates, as, for example, shown in Fig. 4(b), may start from a lower threshold value for each subband and may also be enhanced because of the tunnel-impact ionization process. More work needs to be done to predict the precise performance of such a structure.

VI. CONCLUSIONS

The impact ionization across the conduction-band-edge discontinuity of quantum-well heterostructures has been studied. A geometrical configuration to illustrate the contribution to the electron–electron interaction in the momentum space has been shown. The average ionization rate has also been calculated for various doping concentrations and geometries. We have included the contributions from the multisubbands and the nonparabolicity effects in our calculations. The application of this impact ionization mechanism to a solid-state superlattice photomultiplier has also been discussed.

ACKNOWLEDGMENTS

The discussions with J. P. Leburton, J. J. Coleman, and T. Wang are appreciated. This research is supported by the NASA grant NAG 1-500 and the Joint Services Electronics Program.

APPENDIX A: THE SCREENING PARAMETER q

The dielectric screening parameter in the static limit is given by[31]

$$q = \left[\frac{e^2}{\epsilon}\int g(E)\left(-\frac{\partial f^0}{\partial E}\right)dE\right]^{1/2}, \tag{A1}$$

which is the Thomas–Fermi screening parameter, and it reduces to 1/(Debye length) for a Maxwellian–Boltzmann distribution in the case of a nondegenerate electron gas.

When the electron concentration is degenerate, and the temperature is low, we obtain the Thomas–Fermi screening parameter using a three-dimensional density of states:

$$q = (m^* e^2/\epsilon\hbar^2)^{1/2}(3n/\pi^4)^{1/6}, \tag{A2}$$

where n is the electron concentration.

If we consider a two-dimensional density of state $g_{2D}(E)$, we can carry out the integral in Eq. (A1) since $g_{2D}(E)$ is a constant and obtain

$$q = (m^* e^2/\epsilon\hbar^2)^{1/2}(N/\pi L_z)^{1/2}, \tag{A3}$$

where N is the number of occupied subbands. Our calculations for a quantum well of size $L = 200$ Å and $\Delta E_c \cong 0.3$ eV show that Eqs. (A2) and (A3) give almost the same result for a large n.

APPENDIX B: THE QUANTIZED ENERGY LEVELS E_{Qn} IN A QUANTUM WELL INCLUDING THE EFFECTIVE-MASS DIFFERENCE

The energy level E_{Qn} is obtained by using the boundary conditions

$$\psi_a = \psi_b, \tag{B1}$$

$$\frac{1}{m_a^*}\frac{\partial\psi_a}{\partial z} = \frac{1}{m_b^*}\frac{\partial\psi_b}{\partial z}, \tag{B2}$$

at the two interfaces $z = 0$ and $z = L_z$. The result is the well-known eigenequation:

$$\tan kL_z = 2\alpha/(1-\alpha^2), \tag{B3}$$

where

$$k = (2m_b^* E)^{1/2}/\hbar, \tag{B4}$$

$$\alpha = [(m_b^*/m_a^*)(\Delta E_c - E)/E]^{1/2}. \tag{B5}$$

The eigenequation (B3) includes both the even and odd solutions for the wave functions. The number of quantized levels in the quantum well is easily determined from the criterion. If

$$\frac{(N-1)\pi}{2} < \frac{(2m_b^*\Delta E_c)^{1/2}L_z}{2\hbar} < \frac{N\pi}{2}, \tag{B6}$$

the number of quantized levels $= N$.

APPENDIX C: AVERAGE EFFECT MASS m_b^* IN GaAs REGION (REGION b) WITH THE NONPARABOLICITY EFFECT INCLUDED

The E-k relation including the nonparabolic effect is

$$E(1+\alpha E) = \hbar^2 k^2/2m^*, \tag{C1}$$

where we use $\alpha = 0.67\ (\text{eV})^{-1}$, $m^* = 0.063\, m_0$ at 77 K, and m_0 is the electron mass in free space.

Using the definition

$$\frac{1}{m_{\text{eff}}^*(E)} = \frac{1}{\hbar^2}\frac{\partial^2}{\partial k^2}\mathrm{E} \tag{C2}$$

and

$$m_b^* = \int_0^{\Delta E_c} m_{\text{eff}}^*(E)dE/\Delta E_c, \tag{C3}$$

where ΔE_c is the conduction-band-edge discontinuity, we obtain analytically

$$m_b^* = m^*(1+\alpha\Delta E_c)(1+2\alpha\Delta E_c + 2\alpha^2\Delta E_c^2). \tag{C4}$$

The above procedure accounts for the nonparabolic effect up to the energy ΔE_c. We then use this average value m_b^* in the

parabolic band structure for GaAs in our calculation for the ionization rates.

[1]R. Chin, N. Holonyak, Jr., G. E. Stillman, J. Y. Tang, and K. Hess, Electron. Lett. **16**, 467 (1980).
[2]F. Capasso, W. T. Tsang, A. L. Hutchinson, and G. F. Williams, Appl. Phys. Lett. **40**, 38 (1982).
[3]F. Capasso and W. T. Tsang, Int. Electron Devices Meet. Tech. Dig. 334 (1982).
[4]F. Capasso, IEEE Trans. Nucl. Sci. **NS-30**, 424 (1983).
[5]F. Capasso, W. T. Tsang, and G. F. Williams, IEEE Trans. Electron Devices **ED-30**, 381 (1983).
[6]F. Capasso, J. Vac. Sci. Technol. B **1**, 454 (1983).
[7]J. S. Smith, L. C. Chiu, S. Margalit, and A. Yariv, J. Vac. Sci. Technol. **1**, 376 (1983).
[8]F. Capasso, S. Luryi, W. T. Tsang, C. G. Bethea, and B. F. Levine, Phys. Rev. Lett. **51**, 2318 (1983).
[9]G. C. Osbourn, J. Vac. Sci. Technol. B **2**, 176 (1984).
[10]D. R. Myers, J. J. Wiczer, T. E. Zipperian, and R. M. Biefeld, IEEE Electron Device Lett. **EDL-5**, 326 (1984).
[11]Y. Horikoshi, A. Fischer, and K. Ploog, Appl. Phys. Lett. **45**, 919 (1984).
[12]G. H. Dohler, Superlatt. Microstruct. **1**, 279 (1985).
[13]P. T. Landsberg and A. R. Beattie, J. Phys. Chem. Solids **8**, 73 (1959).
[14]A. R. Beattie and P. T. Landsberg, Proc. R. Soc. London. Ser. A **249**, 16 (1959).
[15]P. T. Landsberg, C. Phys-Roberts, and P. Lal, Proc. Phys. Soc. **84**, 915 (1964).
[16]P. T. Landsberg, Phys. Status Solidi **41**, 457 (1970).
[17]M. Takeshima, J. Appl. Phys. **43**, 4114 (1972).
[18]C. Smith, R. A. Abram, M. G. Burt, Superlatt. Microstruct. **1**, 119 (1985).
[19]A. Sugimura, IEEE J. Quantum Electron. **QE-19**, 932 (1983).
[20]J. M. Ziman, *Electrons and Phonons* (Oxford University, London, 1960).
[21]A. Matulionis, J. Pozela, and A. Reklaitis, Solid State Commun. **16**, 1133 (1975).
[22]K. Huang, *Statistical Mechanics* (Wiley, New York, 1963), pp. 59–60.
[23]S. Adachi, J. Appl. Phys. **58**, R1 (1985).
[24]H. C. Casey, Jr., and M. B. Panish, *Heterostructure Lasers, Part A Fundamental Principals* (Academic, New York, 1978).
[25]J. S. Blakemore, J. Appl. Phys. **53**, R123 (1982).
[26]R. C. Miller, A. C. Gossard, D. A. Kleinman, and O. Munteanu, Phys. Rev. B **29**, 3740 (1984).
[27]R. C. Miller, D. A. Kleinman, and A. C. Gossard, Phys. Rev. B **29**, 7085 (1984).
[28]L. C. Chiu, J. S. Smith, S. Margalit, and A. Yariv, Appl. Phys. Lett. **45**, 331 (1983).
[29]A. S. Kyuregyan, Sov. Phys.-Semicond. **10**, 410 (1976).
[30]F. Capasso, *Physics of Avalanche Photodiodes in Semiconductors and Semimetals*, edited by R. K. Willardson and A. C. Beer (Academic, New York, 1985), Vol. 22, Part D, Chap. 1.
[31]J. M. Ziman, *Principles of the Theory of Solids*, second edition (Cambridge University, Cambridge, 1972), Sec. 5.2.

APPENDIX B

Copy of viewgraphs entitled: "Impact Ionization Across the Band-edge Discontinuity For A Superlattice Photomultiplier" by S. L. Chuang and K. Hess, presented at the 1986 Device Research Conference, Amherst, Massachusetts, June 21-23, 1986.

Impact Ionization Across The Band-edge Discontinuity For A Superlattice Photomultiplier

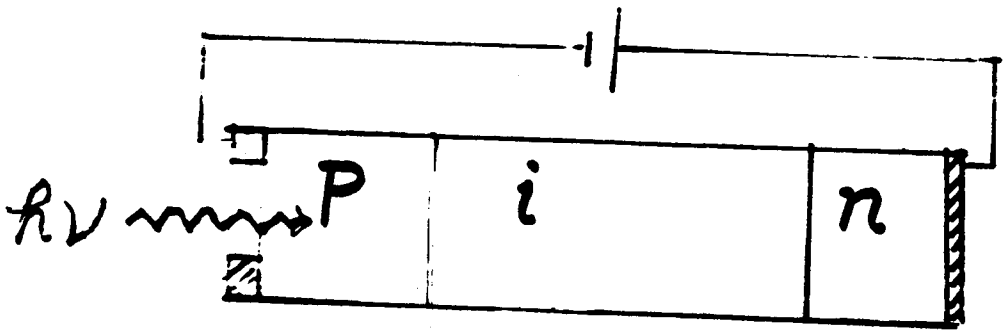
S.L. Chuang and K. Hess

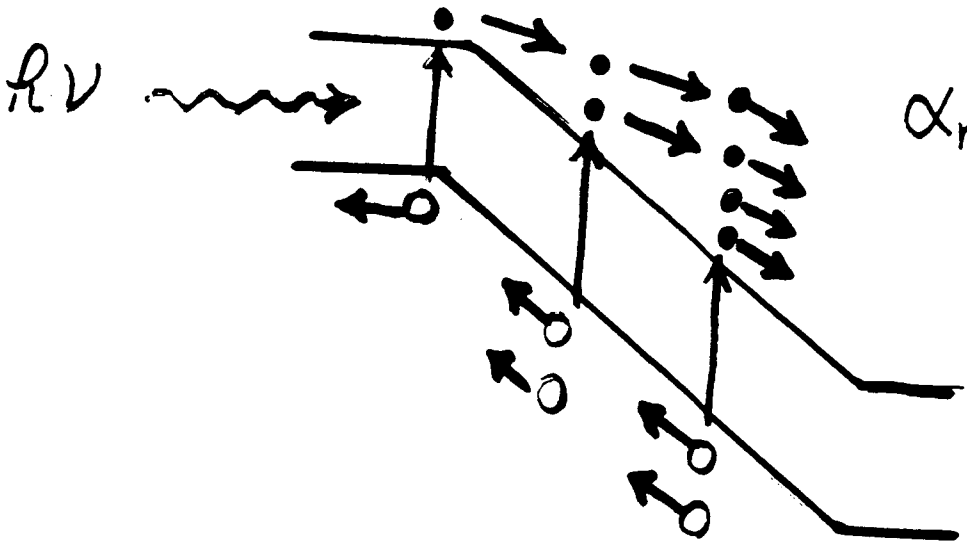
Department of Electrical and Computer Engineering

University of Illinois

B1

Avalanche Photodiodes (APD)

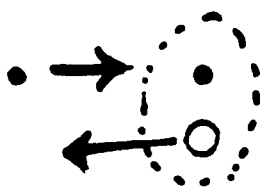




α_n : Average # of secondary elec. created by an electron/unit length

$\beta_p \rightarrow$ hole

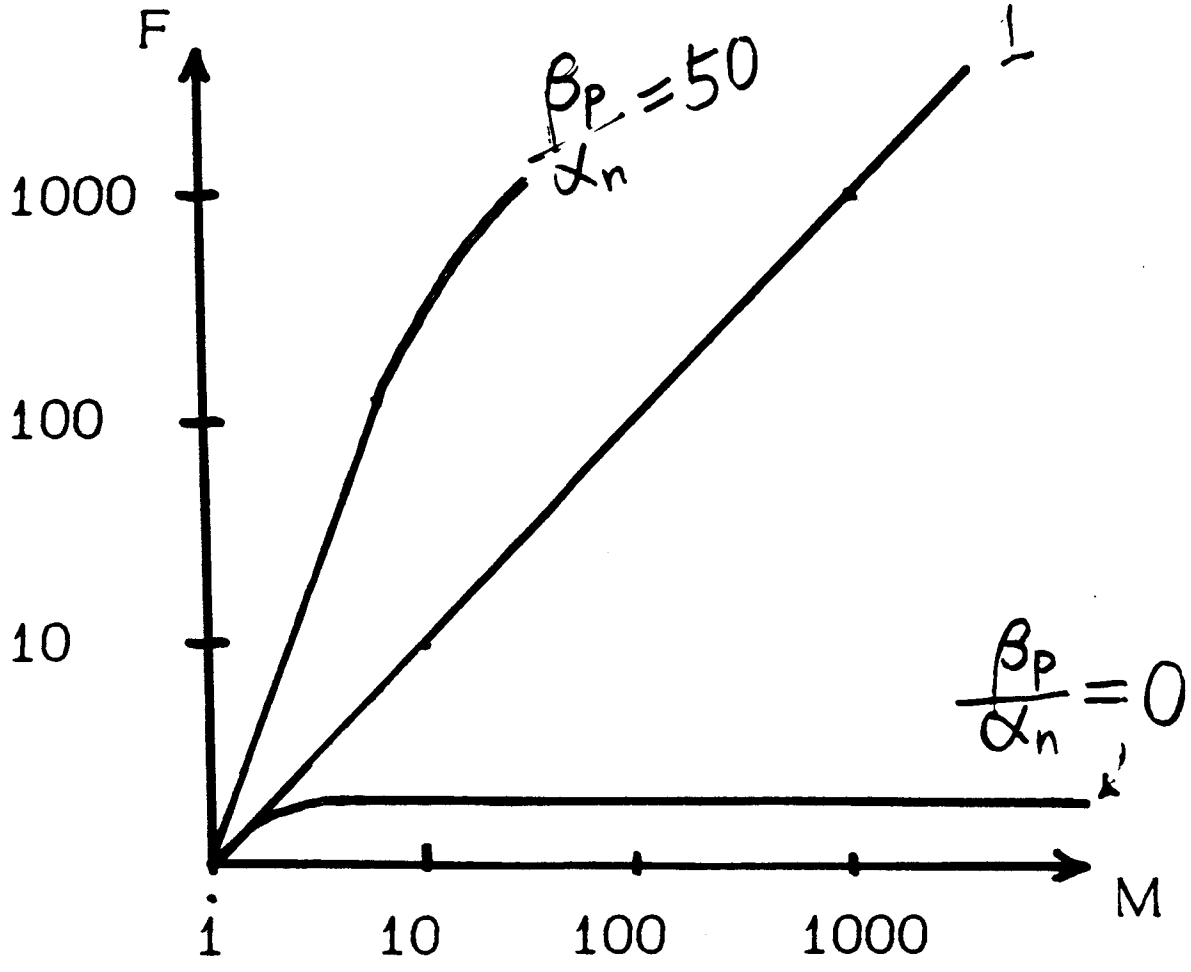
Avalanche multiplication--- a random process



opposite direction

This strong feedback effect "amplifies" the noise fluctuations, (unless $\beta_p \ll \alpha_n$).

The excess noise Factor



If $\frac{\beta_p}{\alpha_n} << 1$, F is small.

For low noise APD's:

need: (1) $\beta_p << \alpha_n$

(2) Electrons initiate the avalanche process.

Previous work:

A superlattice APD:

① R. Chin et al. Electron Lett. 17, 467, 1980.

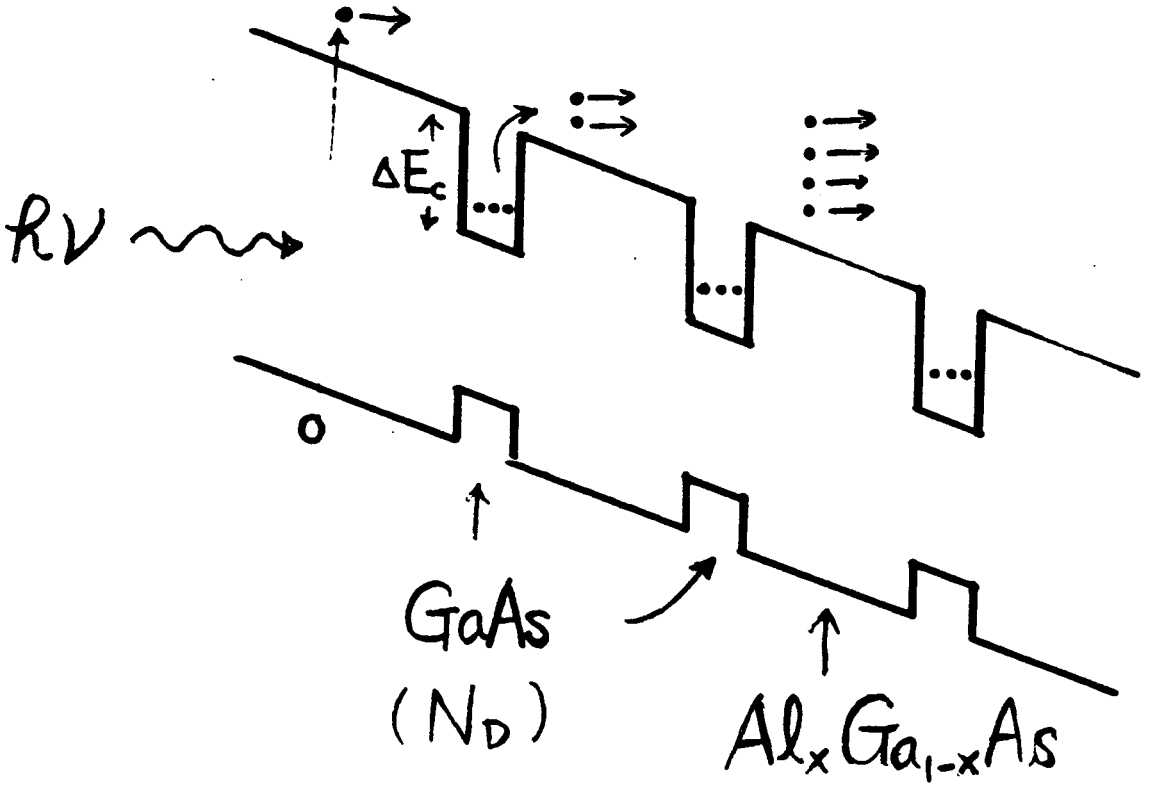
② Capasso et al. Appl. Phys. Lett. 40, 38, 1982.

⋮

Interband impact ionization process is considered.

Our work:

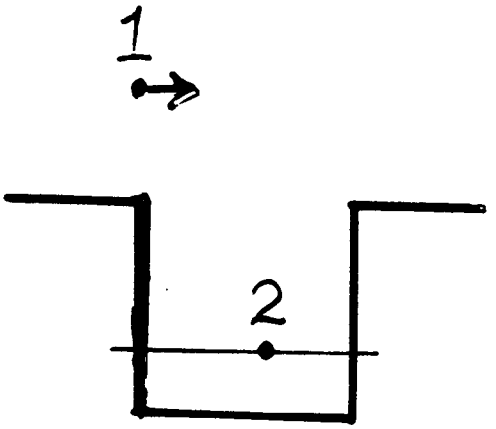
A New Superlattice Photomultiplier ?



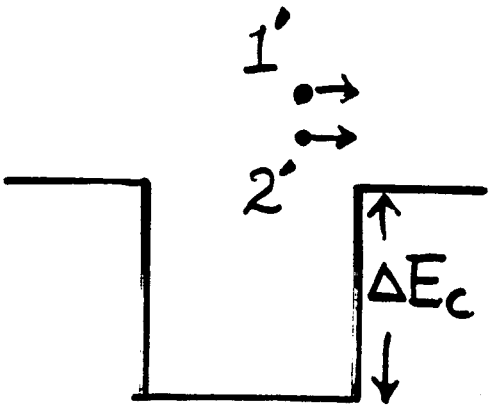
(1) impact ionization across ΔE_c (instead of E_g): <u>intra-band</u> transition

(2) Only electrons are involved ($\alpha_n \gg \beta_p$).

Initial State



Final State



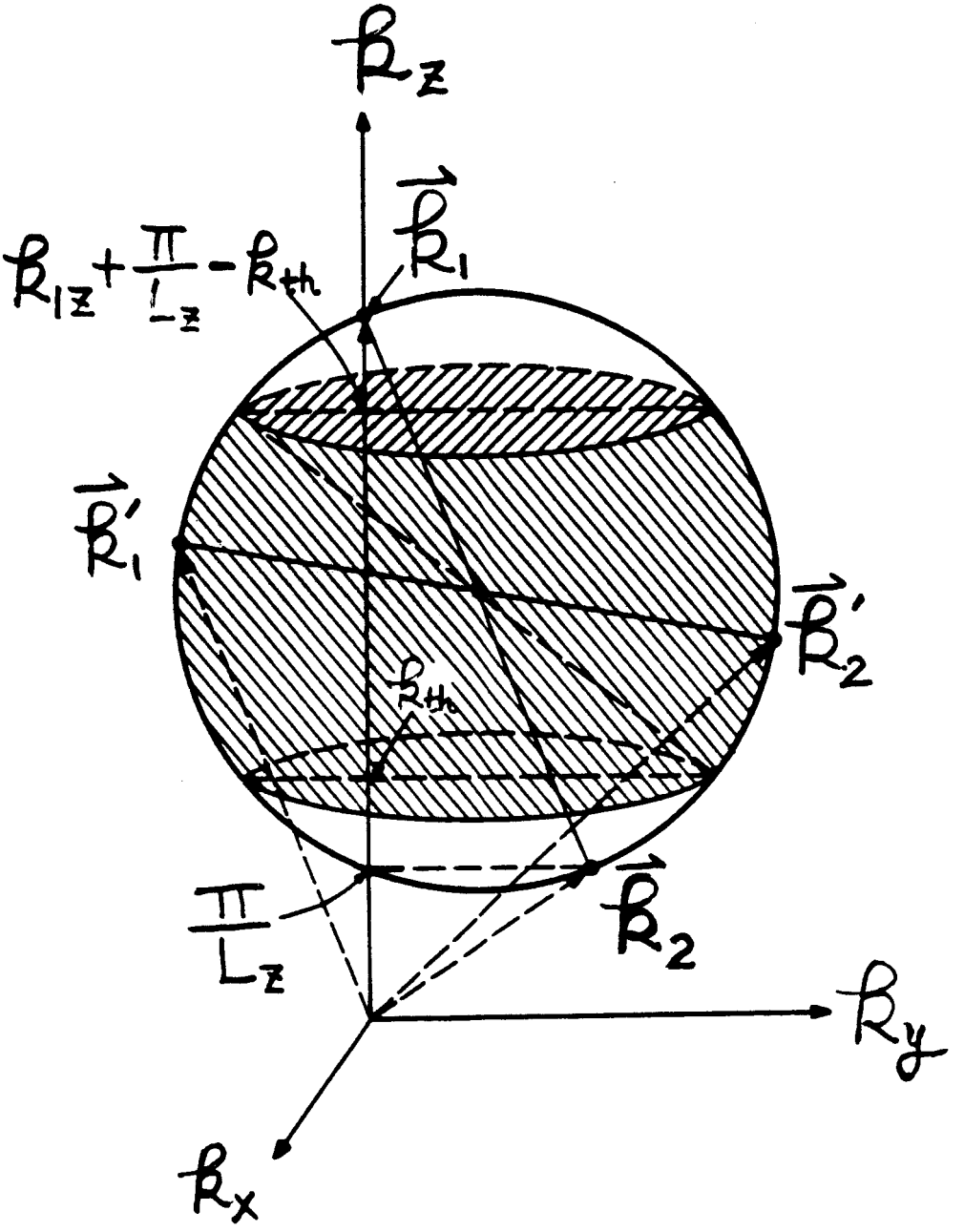
Transition rate/volume

$$\frac{2\pi}{\hbar}\left|\langle 12|H_s|1'2'\rangle\right|^2\delta(E_1+E_2-E_1'-E_2')$$

$$P_{tr}=\frac{1}{V}\sum_{k_1,k_2}^{k_1',k_2'} P\, f(k_1)\, f(k_2)[\,1-f(k'_1)\,][\,1-f(k'_2)\,]$$

(JAP, 59, 2885, 1986)

A momentum space diagram :

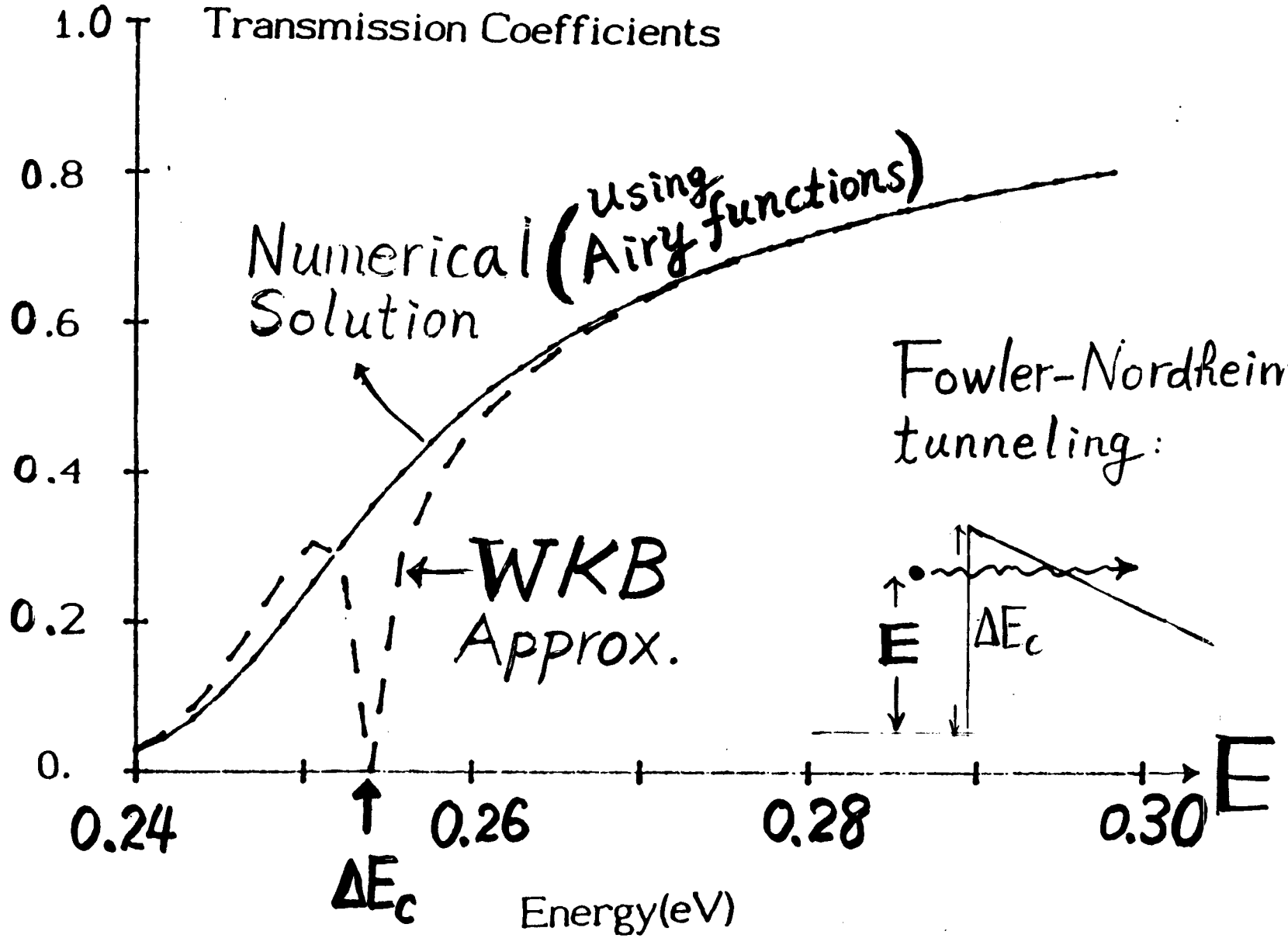


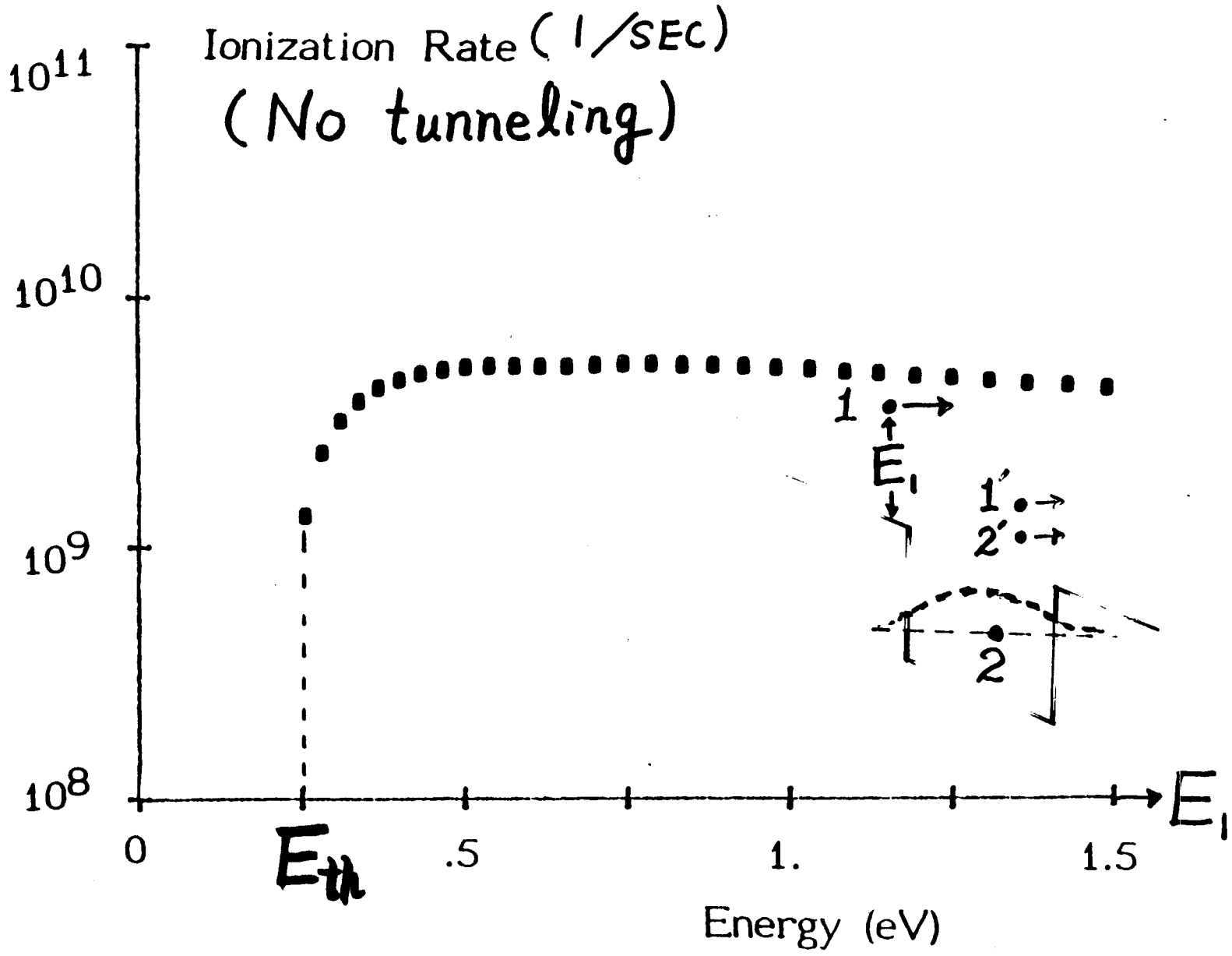
Momentum conservation : $\vec{k}'_1 + \vec{k}'_2 = \vec{k}_1 + \vec{k}_2$

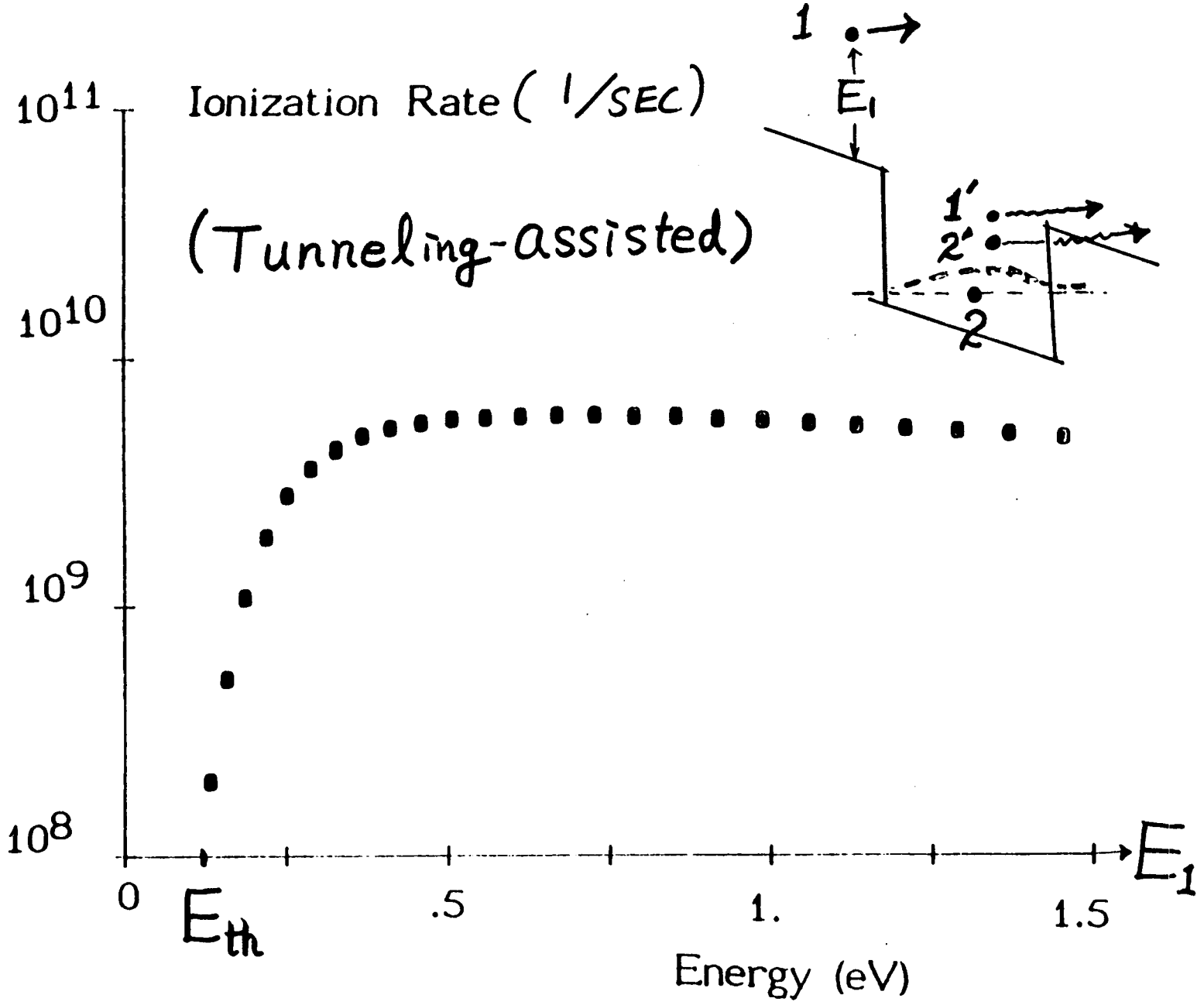
Energy conservation : $E'_1 + E'_2 = E_1 + E_2$

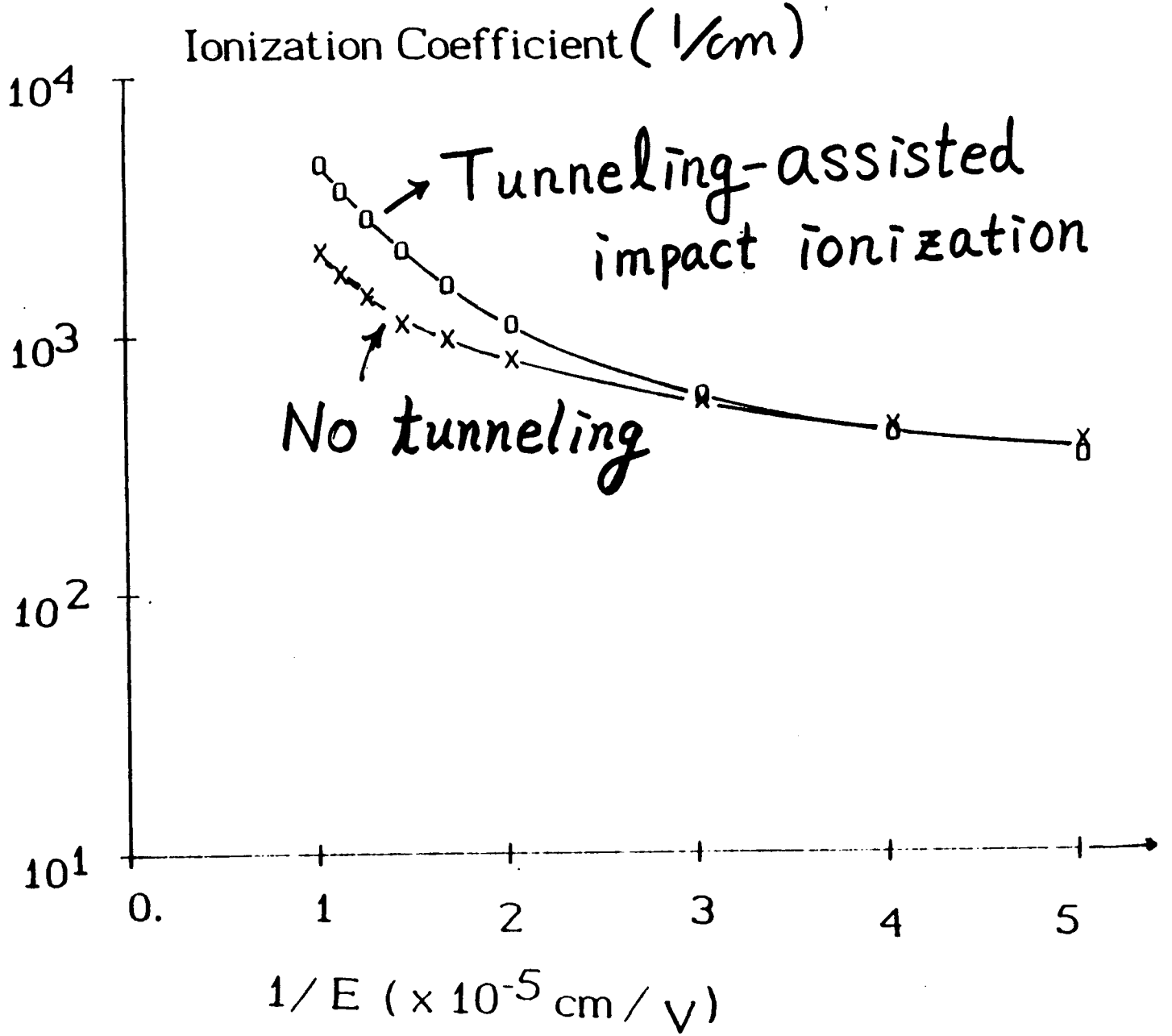
➡ $\vec{k}_1$, $\vec{k}_2$, $\vec{k}'_1$ and $\vec{k}'_2$ lie on a spherical surface

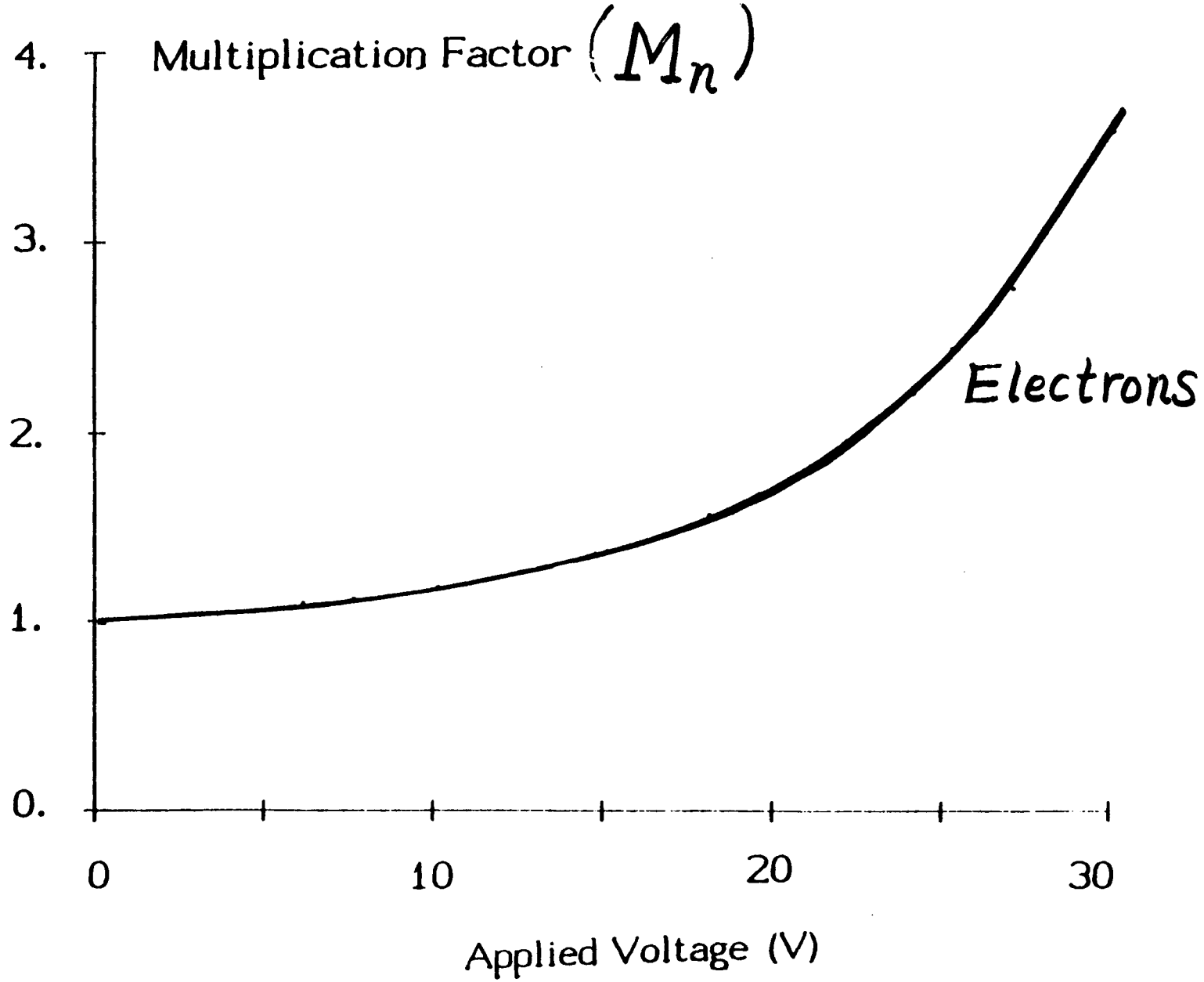
Tunneling effect:



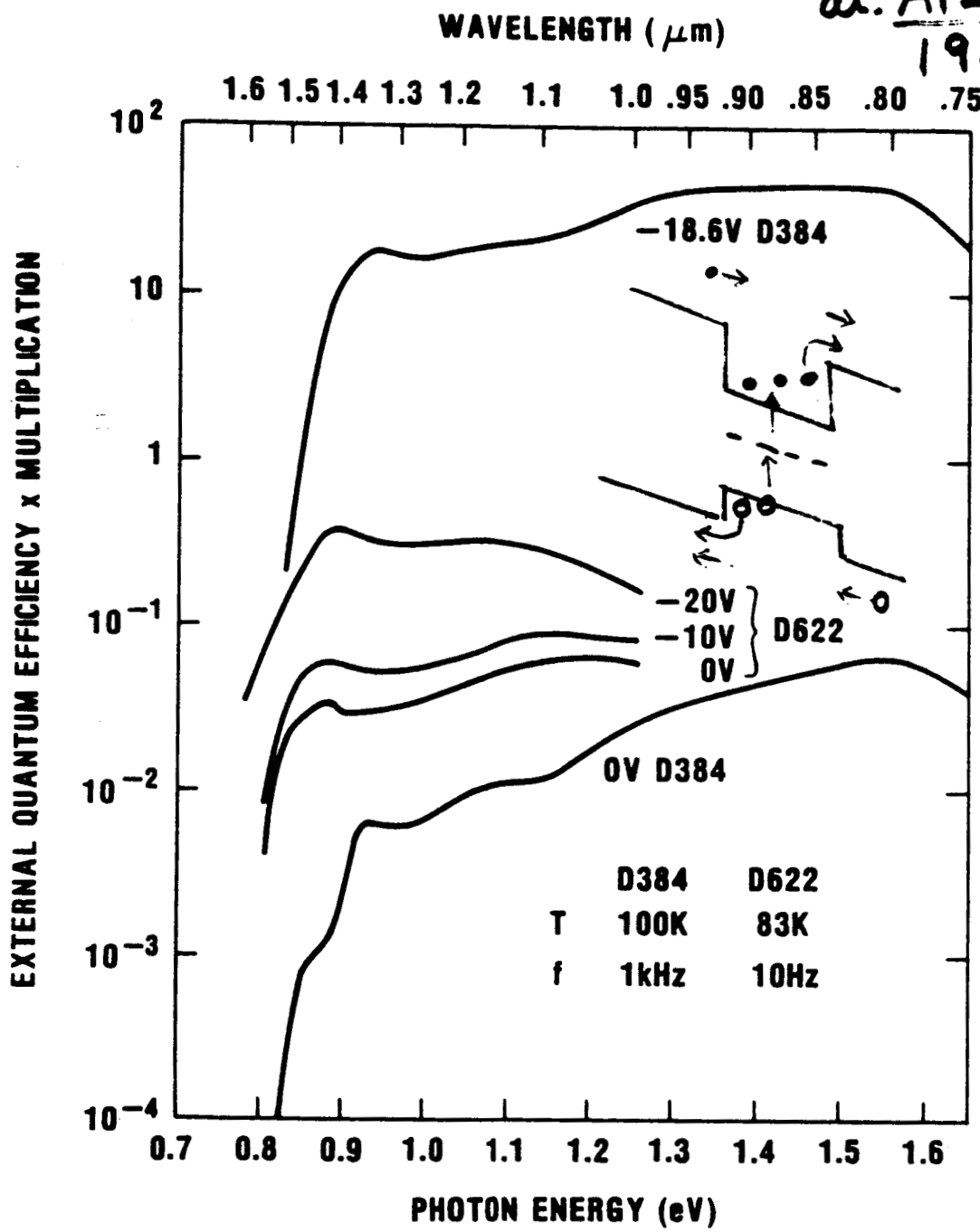

Ionization Rate (1/SEC)
(No tunneling)
10^{11}
10^{10}
10^{9}
10^{8}
1
E_1
1'
2'
2
0
E_{th}
.5
1.
1.5
E_1
Energy (eV)


Ionization Rate (1/SEC)
(Tunneling-assisted)
10^{11}
10^{10}
10^{9}
10^{8}
0
E_{th}
.5
1.
1.5
E_1
Energy (eV)
1
E_1
1'
2'
2

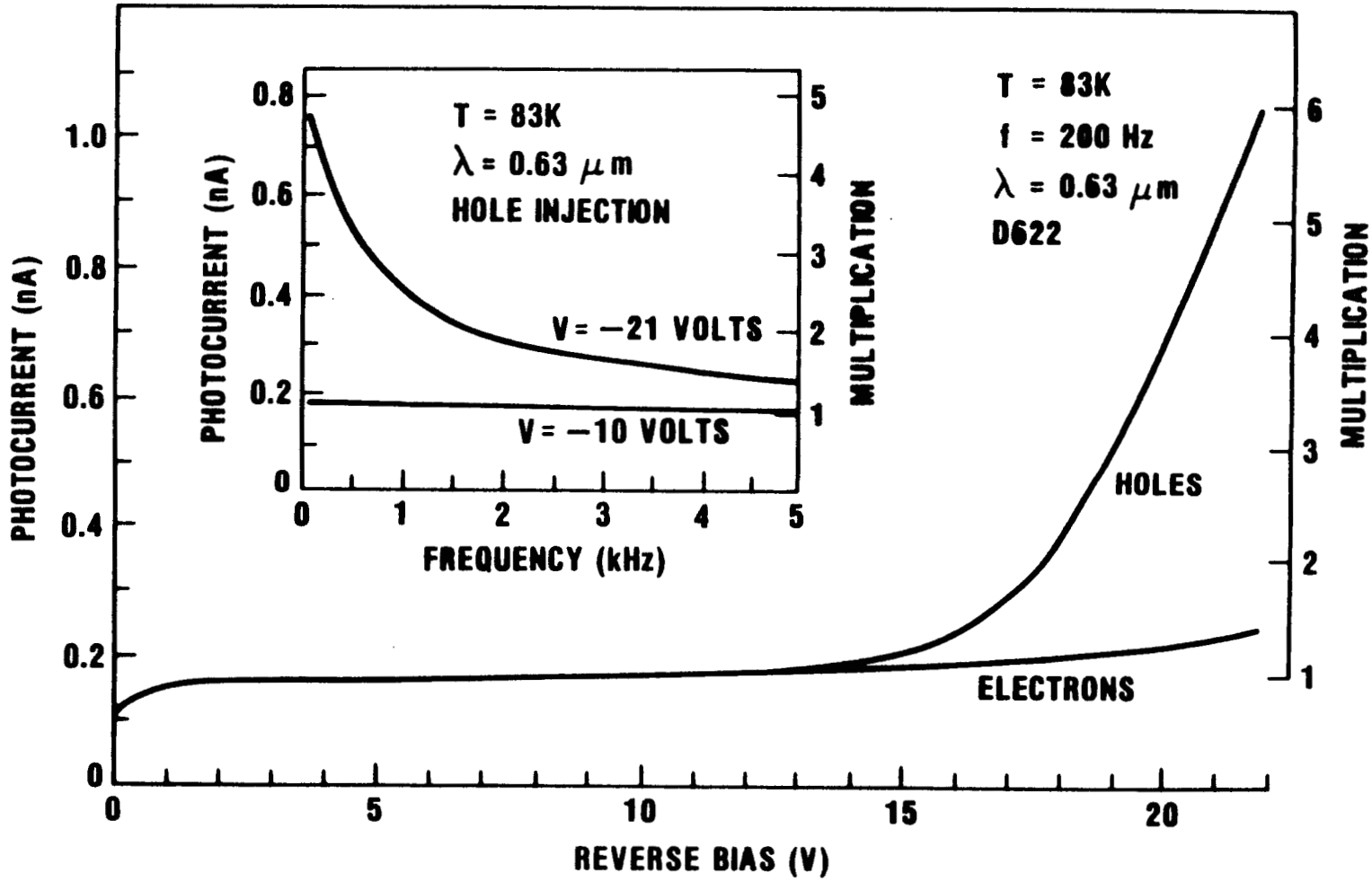

Ionization Coefficient (1/cm)
10^4
10^3
10^2
10^1
Tunneling-assisted impact ionization
No tunneling
0.
1
2
3
4
5
1/E (x 10^{-5} cm / V)


Multiplication Factor (M_n)
4.
3.
2.
1.
0.
Electrons
0
10
20
30
Applied Voltage (V)

Experimental evidence (Capasso et al. APL 48 1986)



$Al_{0.48}In_{0.52}As$ $Ga_{0.47}In_{0.53}As$



Conclusions:

A new superlattice photodetector has been proposed and studied theoretically. A single carrier type (electron) is involved and it has the potential of minimizing the excess noise and achieving an appreciable gain.

APPENDIX C

Reprint of: "Lateral Shift Of An Optical Beam Due To Leaky Surface-Plasmon Excitations" by S. L. Chuang, J. Opt. Soc. Am., vol. 3, pp. 593-599, 1986.

Lateral shift of an optical beam due to leaky surface-plasmon excitations

S. L. Chuang

Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

Received February 12, 1985; accepted December 18, 1985

A leaky-wave theory is applied to the study of the lateral shift of a bounded optical beam reflected from a prism–air–metal (and prism–metal–air) layered configuration with surface-plasmon excitations. The locations of the pole and the zero of the reflection coefficient for the leaky surface-plasmon mode are calculated as one varies the thickness of the center layer. The excited surface-plasmon mode is leaky because of the presence of the prism. Both the real part and the imaginary part of the pole are affected by the energy leakage. The detailed structure of the reflected intensity and the forward and backward shifts of the reflected beam are illustrated.

1. INTRODUCTION

Since the optical properties of a metallic film or a semiconductor sample can be characterized by a permittivity function $\epsilon(\omega)$, the theories and techniques to determine $\epsilon(\omega)$ have been important to our understanding of the optical properties of solids. One useful optical technique is to use the surface-plasmon wave theory in the reflectance measurement.[1] The excitation of surface-plasmon waves is achieved either by using the attenuated-total-reflection method or by using a grating structure with a proper choice of the phase-matching condition. If one measures the reflected power as a function of the incidence angle or the frequency, anomalous absorption will occur when the surface-plasmon mode is excited. Thus the surface-plasmon theory provides a sensitive tool to determine the permittivity function of the solids. For metals, the surface-plasmon mode is excited in the visible frequency region, whereas for degenerate semiconductors it is excited in the infrared region. The study of surface-plasmon excitations may give detailed information about the optical properties of the metals and semiconductors.

However, so far, almost all the studies on the excitation of the surface-plasmon theory assume that the incident wave is an infinite plane wave. In electromagnetics or in acoustics, it is known that lateral beam displacement may occur when the incident wave is a bounded beam and proper phase-matching conditions are satisfied.[2] For example, if the incident angle is at the critical angle between two media, the well-known Goos–Hänchen shift occurs.[3–6] In acoustics, if the incident acoustic beam is from a fluid medium to a solid surface and the angle of incidence is near the Rayleigh critical angle, the excitation of the Rayleigh wave in the solid, which becomes a leaky wave because of the existence of the fluid, causes a beam shift in the reflected wave.[7] The result of the reflected field is different from that predicted by the geometric acoustics. The intensity and phase variations of the shifted beam give information about the acoustic properties of the solids. Thus this idea can be applied in the nondestructive testing of the mechanical properties of solids. A leaky-wave theory has been successful in explaining various observed phenomena in acoustics[7] and for multilayered or grating structures.[8–10]

The properties of surface polaritons in layered structures have been discussed in many references.[1,11–13] The attenuated-total-reflection method[1] is used to determine the optical constants of thin metallic films[14] or a CsBr sample in the infrared region.[15] Grating structures have also been used to study the dispersions of surface plasmons in semiconductors[16,17] and in metals.[18] Both prisms and gratings have been used to couple the electromagnetic wave beams into the waveguide modes.[19–21] Coupling of the finite beam with the surface-plasmon mode has also been studied, including the coupling efficiency[22–24] and the beam-shift phenomenon.[25,26] The study on the lateral beam displacement in Ref. 26 closely follows Ref. 6, which assumes a Taylor-series expansion near the incident angle and is applied to the case of the critical angle of incidence, i.e., the Goos–Hänchen shift. That expansion is not valid if there is a pole, which is due to the leaky surface-plasmon mode, as will be explained later, near the horizontal wave number where the reflection coefficient is expanded into the Taylor series. This has been pointed out in Ref. 7, and a better approximation, which takes into account the singularity of the reflection coefficient, has been proposed. The pole of the reflection coefficient accounts for the leaky surface mode excited by the incident beam. The excitation of the surface electromagnetic waves in an Otto configuration (i.e., prism–air–metal) by using the leaky-wave theory has been studied[25] both theoretically and experimentally. In this paper, we apply the leaky-wave theory to both the Otto configuration and the Kretschmann configurations (i.e., prism–metal–air). The trajectories of the poles and zeros as the middle-layer thickness is varied are studied in detail. Other singularities, e.g., the branch points, are also shown. When the branch-point contribution is important in certain cases of Kretschmann configuration for which the leaky-wave theory for the reflection coefficient is no longer valid, a direct numerical approach is adopted, and the results are also shown.

2. FORMULATION

The theoretical formulation closely follows Ref. 10, except that we deal with the excitation of a leaky surface-plasmon mode that requires the incident wave to be TM polarized and either medium 2 or medium 3 has a negative permittivi-

0740-3232/86/050593-07$02.00

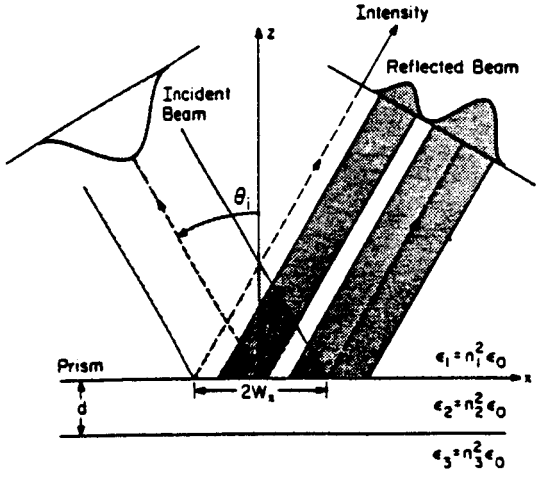


Fig. 1. The reflection of a bounded beam from a prism–air–metal or prism–metal–air configuration.

ty in the real part that is smaller than $-\epsilon_0$. The geometrical configuration of the problem is shown in Fig. 1, where both prism–air–metal and prism–metal–air configurations are considered. A metal such as silver has a permittivity $(-5.19 + i0.28)\,\epsilon_0$ at the wavelength 4358 Å.[14]

Assume that a bounded beam is incident from the prism region with the magnetic field given by

$$\mathbf{H}_i = \hat{y} H_i(x, 0) = \hat{y} \exp\left[-\left(\frac{x}{W_x}\right)^2 + ik_{x_i}x\right] \tag{1}$$

on the $z = 0$ plane, where $W_x = W/\cos\theta_i$, W is the beamwidth, θ_i is the angle of incidence for the center of the beam, $k_{xi} = k_1 \sin\theta_i$, $k_1 = \omega\sqrt{\mu_0\epsilon_1}$, and ϵ_1 is the permittivity of the prism.

Using the plane-wave spectrum of the incident field at $z = 0$

$$H_i(x, 0) = \int_{-\infty}^{\infty} \frac{dk_x}{2\pi} h(k_x)\exp(ik_x x), \tag{2}$$

where

$$h(k_x) = \sqrt{\pi} W_x \exp\left[-(k_x - k_{x_i})^2\left(\frac{W_x}{2}\right)^2\right], \tag{3}$$

the reflected magnetic field is given by

$$\mathbf{H}_R = \hat{y} H_r(x, z) = \hat{y} \int_{-\infty}^{\infty} \frac{dk_x}{2\pi} R(k_x)h(k_x)\exp[ik_x x + ik_{1z}z], \tag{4}$$

where $k_{1z} = (\omega^2\mu_0\epsilon_1 - k_x^2)^{1/2}$, with $\mathrm{Im}\,(k_{1z}) \geq 0$, and the path of integration is on the real axis of the complex k_x plane. $R(k_x)$ is the plane-wave reflection coefficient from the layered medium. From the expression in Eq. (3), one sees that the major contribution to the integral in Eq. (4) is from the visible region with k_x satisfying

$$|k_x - k_{x_i}| < \frac{2}{W_x}. \tag{5}$$

As will be explained in Section 3, if θ_i is near the angle where the surface plasmon is excited, the reflection coefficient can be expressed as the product of a factor containing a zero and a pole in the form

$$R(k_x) \simeq R_0 \frac{k_x - k_n}{k_x - k_p}, \tag{6}$$

where R_0 is independent of k_x as an approximation, provided that there are no other singularities such as the branch points in the visible region. If there are branch points near the region in expression (5), the approximation in expression (6) is no longer valid, and a direct numerical approach using Eq. (4) should be taken. This will be the case for the Kretschmann attenuated-total-reflection configuration, as will be shown later. The zero and the pole of $R(k_x)$ in the complex plane can be obtained numercially, since an exact analytical expression of $R(k_x)$ is known. R_0 can be obtained accurately from $R(k_{xi})(k_{xi} - k_p)/(k_{xi} - k_n)$.[6] Let us denote k_p and k_n as

$$k_p = \beta_p + i\alpha_p, \tag{7}$$

$$k_n = \beta_n + i\alpha_n. \tag{8}$$

The reflected field at $z = 0$ can be evaluated from Eq. (4) and expression (6) to be[10]

$$H_r(x, 0) = R_0\left[1 + i\frac{\sqrt{\pi}}{2}(k_p - k_n)W_x \exp(\gamma^2)\mathrm{erfc}(\gamma)\right] \times \exp\left[-\left(\frac{x}{W_x}\right)^2 + ik_{x_i}x\right], \tag{9}$$

where

$$\gamma = \frac{\alpha_p W_x}{2} - \left(\frac{x}{W_x}\right) + i(k_{x_i} - \beta_p)\frac{W_x}{2} \tag{10}$$

and $\mathrm{erfc}(\gamma)$ is the complementary error function of complex argument γ.

3. THE PLANE-WAVE REFLECTION COEFFICIENTS FROM A PRISM-AIR-SILVER AND A PRISM-SILVER-AIR CONFIGURATION

The plane-wave reflection coefficient $R(k_x)$ from a layered medium shown in Fig. 1 can be written as[27]

$$R(k_x) = \frac{R_{12}^{-}R_{23}^{+} + R_{12}^{+}R_{23}^{-}\exp(i2k_{2z}d)}{R_{12}^{+}R_{23}^{+} + R_{12}^{-}R_{23}^{-}\exp(i2k_{2z}d)}, \tag{11}$$

where

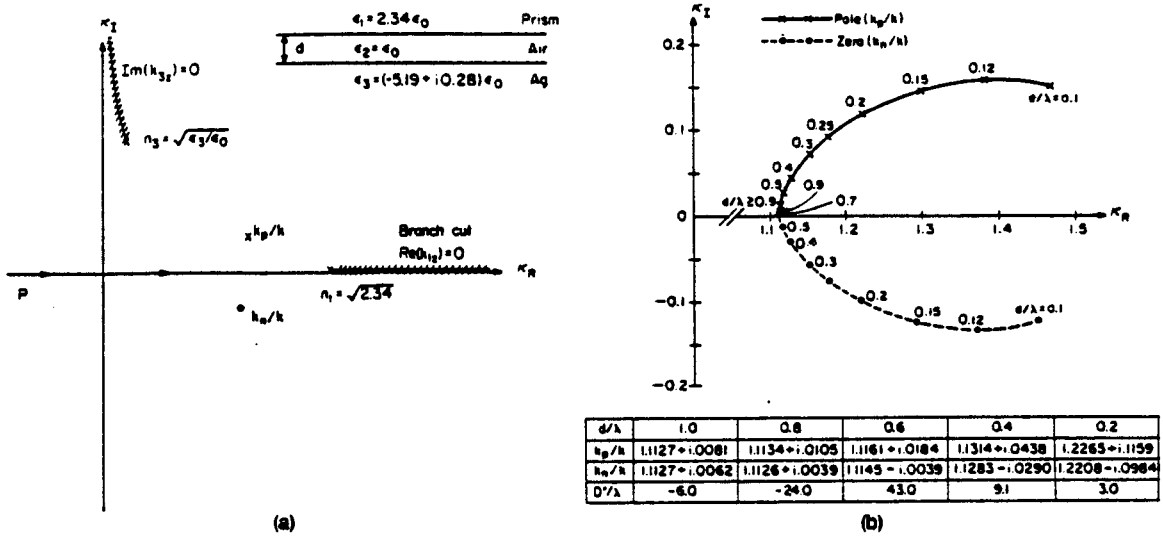
$$R_{ij}^{\pm} \equiv \frac{k_{iz}}{\epsilon_i} \pm \frac{k_{jz}}{\epsilon_j}, \tag{12}$$

$$k_{iz} = (\omega^2\mu_0\epsilon_i - k_x^2)^{1/2}. \tag{13}$$

One notes that there are only two branch points in Eq. (11) occurring at $k_{1z} = 0$ and $k_{3z} = 0$. This has been noted in Ref. 10 and also proved by invoking the uniqueness principle in solving the boundary-value problem for stratified media.[28] Another algebraic method is simply to look at Eq. (11) and change k_{2z} to $-k_{2z}$; one obtains the same expression for the reflection coefficient. Let us define the normalized quantity

$$\kappa = k_x/k, \tag{14}$$

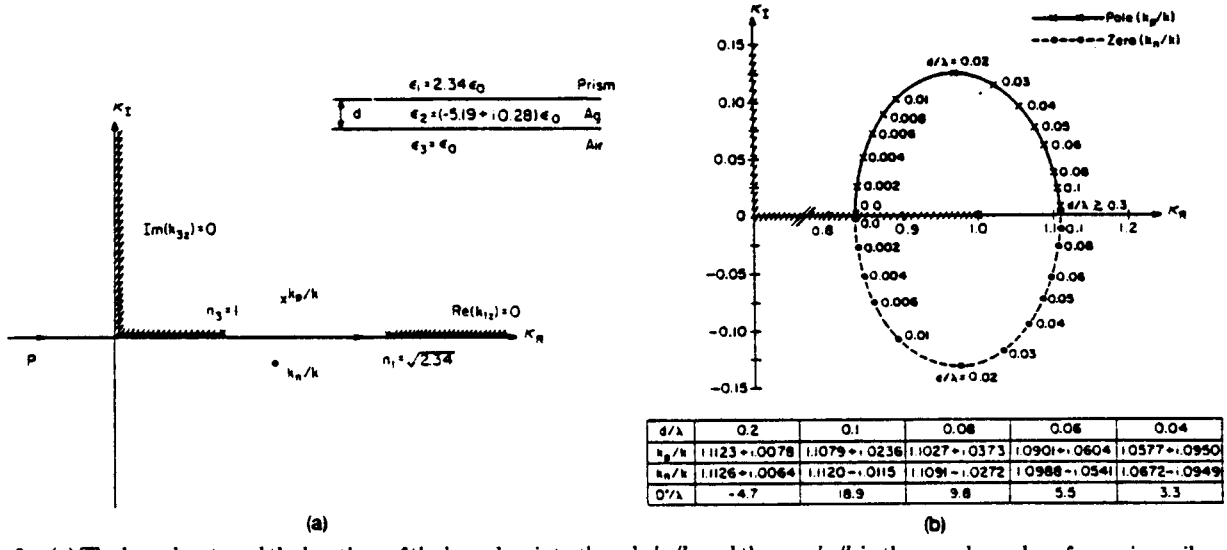
where $k = \omega\sqrt{\mu_0\epsilon_0}$ is the wave number in free space. The two branch points occur at



d/λ	1.0	0.8	0.6	0.4	0.2
k_p/k	1.1127+i.0081	1.1134+i.0105	1.1161+i.0184	1.1314+i.0438	1.2265+i.1159
k_n/k	1.1127+i.0062	1.1126+i.0039	1.1145−i.0039	1.1283−i.0290	1.2208−i.0984
Q′/λ	-6.0	-24.0	43.0	9.1	3.0

(a) (b)

Fig. 2. (a) The branch cuts and the locations of the branch points, the pole k_p/k, and the zero k_n/k in the complex κ plane for a prism–air–silver configuration. (b) The trajectories of the poles and zeros on the κ plane when the air-gap thickness is varied.



d/λ	0.2	0.1	0.08	0.06	0.04
k_p/k	1.1123+i.0078	1.1079+i.0236	1.1027+i.0373	1.0901+i.0604	1.0577+i.0950
k_n/k	1.1126+i.0064	1.1120−i.0115	1.1091−i.0272	1.0988−i.0541	1.0672−i.0949
Q′/λ	-4.7	18.9	9.8	5.5	3.3

(a) (b)

Fig. 3. (a) The branch cuts and the locations of the branch points, the pole k_p/k, and the zero k_n/k in the complex κ plane for a prism–silver–air configuration. (b) The trajectories of the poles and the zeros on the complex κ plane when the silver thickness is varied.

$$n_1 = \sqrt{\epsilon_1/\epsilon_0} \tag{15}$$

and

$$n_3 = \sqrt{\epsilon_3/\epsilon_0} \tag{16}$$

in the complex κ plane. The branch cuts are chosen to be[27]

$$\mathrm{Re}[k_{1z}] = 0 \tag{17}$$

and

$$\mathrm{Im}[k_{3z}] = 0, \tag{18}$$

which are shown in Figs. 2(a) and 3(a).

Case 1. Prism–Air–Metal (Otto Attenuated Total Reflection) Configuration

The branch cuts, as defined by Eqs. (17) and (18), are shown in Fig. 2(a). Once these branch cuts are defined, the pole and the zero of the reflection coefficient for k_z in Eq. (11) can be obtained and written as k_p/k and k_n/k in the κ plane after being normalized to the free-space wave number k.

The pole in the reflection coefficient k_p corresponds to a leaky surface-plasmon mode of this configuration. If the prism is not present, the surface-plasmon mode has a dispersion relation

$$k_z = k\left[\frac{\epsilon_2\epsilon_3}{\epsilon_0(\epsilon_2 + \epsilon_3)}\right]^{1/2}. \tag{19}$$

If both ϵ_2 and ϵ_3 are real and $\epsilon_3 < -\epsilon_2$, the above solution in Eq. (19) is real and corresponds to a nondissipative surface-plasmon mode. Adding the prism will make this pole complex because it corresponds to a leaky mode, i.e., the coupling of the prism and the air–silver surface-plasmon mode would make this mode radiate into the prism region. A complex pole comes from the fact that it is a leaky mode but

not from a power dissipation in the bulk region because ϵ_2 and ϵ_3 are real.

If ϵ_3 is a complex quantity, there is a dissipation in the silver region, and the air–silver surface-plasmon mode without the prism has $(k_x/k)_{\text{air-Ag}} = 1.1125 + i0.0071$ from Eq. (19). With the prism, the leaky surface-plasmon mode that has the wave number k_x/k close to $(k_x/k)_{\text{air-Ag}}$ contains both the leakage and the dissipation effects. As $d/\lambda \to \infty$, we have $\exp(i2k_{2z}d) \to 0$ [one chooses $\text{Im}(k_{2z}) > 0$ for convenience] and

$$R(k_x) \to \frac{R_{12}^- R_{23}^+}{R_{12}^+ R_{23}^+} = R_{12}. \tag{20}$$

Thus both the pole and the zero of $R(k_x)$ approach the root of $R_{23}^+ = 0$ that gives exactly Eq. (19) as it should, since, as $d \gg \lambda$, the reflection coefficient $R(k_x)$ should be independent of the property of the air–silver interface, and the pole and the zero must cancel each other. On the other hand, if $d \to 0$, the reflection coefficient can be rewritten as

$$R(k_x) \to \frac{\frac{1}{\epsilon_1}k_{1z} - \frac{1}{\epsilon_3}k_{3z}}{\frac{1}{\epsilon_1}k_{1z} + \frac{1}{\epsilon_3}k_{3z}} = \frac{R_{13}^-}{R_{13}^+} = R_{13}, \tag{21}$$

which has a pole at $R_{13}^+ = 0$, i.e.,

$$(k_x)_{\text{prism-Ag}} = k\left[\frac{\epsilon_1\epsilon_3}{\epsilon_0(\epsilon_1 + \epsilon_3)}\right]^{1/2}, \tag{22}$$

and we obtain $k_p/k = 2.0604 + i0.0454$. Thus the pole in Fig. 2(b) will approach this value as $d/\lambda \to 0$. The zero of the reflection coefficient does not exist on the complex plane shown in Fig. 2(b). It actually moves to the bottom Riemann sheet defined by $\text{Re}[k_{1z}] < 0$ [note that we always define $\text{Im}(k_{3z}) \geq 0$ in our discussions here] as it crosses the branch cut defined by $\text{Re}(k_{1z}) = 0$ shown in Fig. 2(a). This zero will then approach the value $2.0604 + i0.00454$ in the bottom Riemann sheet with $\text{Re}(k_{1z}) < 0$.

Case 2. Prism–Metal–Air (Kretschmann Attenuated Total Reflection) Configuration

The branch cuts are shown in Fig. 3(a). On the complex κ plane, the branch cut $\text{Im}[k_{3z}] = 0$ extends from the positive κ_I axis to the positive κ_R axis with $0 \leq \kappa_R \leq n_3 = 1$. The other branch cut $\text{Re}(k_{1z}) = 0$ extends from $\kappa_R = n_1$ to $+\infty$ on the real axis since ϵ_1 is real. The locations of the pole and the zero move as one varies the thickness of the metal d. As $d/\lambda \to \infty$ [Fig. 3(b)], both the pole and the zero move to the value defined by Eq. (19), i.e., $(k_x/k)_{\text{Ag-air}} = 1.1125 + i0.0071$, as can be seen from Eq. (20) with the common factor $R_{23}^+ = 0$ giving the pole and the zero in the range k_x of our interest [expression (5)]. As $d/\lambda \to 0$, $R(k_x)$ approaches R_{13}^-/R_{13}^+ as in Eq. (21), and we have

$$\frac{k_p}{k} \to \left(\frac{\epsilon_1\epsilon_3}{\epsilon_0(\epsilon_1 + \epsilon_3)}\right)^{1/2} + i\delta = 0.8370 + i\delta \tag{23}$$

and

$$\frac{k_n}{k} \to 0.8370 - i\delta, \tag{24}$$

where δ is an infinitesimal positive number. In this case, the pole and the zero do not cancel each other because one is slightly above and the other slightly below the branch cut $\text{Im}(k_{3z}) = 0$.

Another difference between the Kretschmann configuration and the Otto configuration is the branch-point contribution. As shown in Fig. 3(b) for d/λ between 0.02 and 0.03, the pole has a real part close to the branch point $n_3 = 1.0$. Thus, if the angle of incidence is such that k_{xi} is close to the branch point, the approximate formula in expression (6) for the reflection coefficient will not be valid since the branch-point singularity will also affect the integral for the reflected field. In this case a direct numerical integration using Eq. (4) is necessary, and the results are also given in Section 4. Both a lateral-wave contribution (due to the branch point) and the surface-plasmon-mode contribution (due to the pole) are important in determining the reflected beam.

4. NUMERICAL RESULTS AND DISCUSSIONS

In Figs. 4–6, we show the reflected-beam intensities $|H_r(x, 0)|^2$ (solid curves) and the incident-beam intensities $|H_i(x, 0)|^2$ (dashed curves) as functions of the position x for the

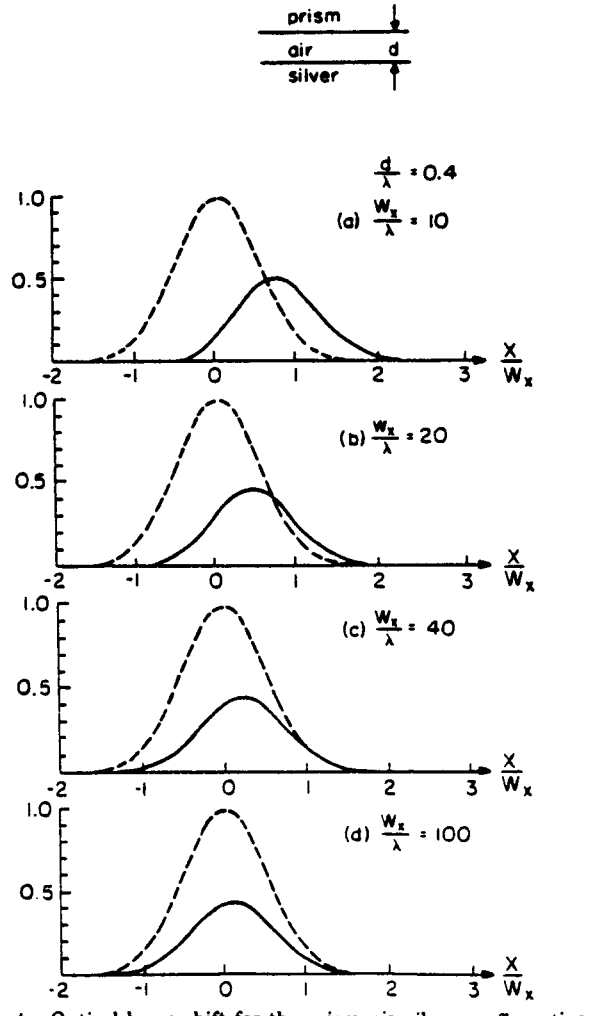


Fig. 4. Optical beam shift for the prism–air–silver configuration with $d/\lambda = 0.4$ for four beamwidths. The dashed curves represent the incident-beam intensities $|H_i(x, 0)|^2$, and the solid curves are for the reflected-beam intensities $|H_r(x, 0)|^2$.

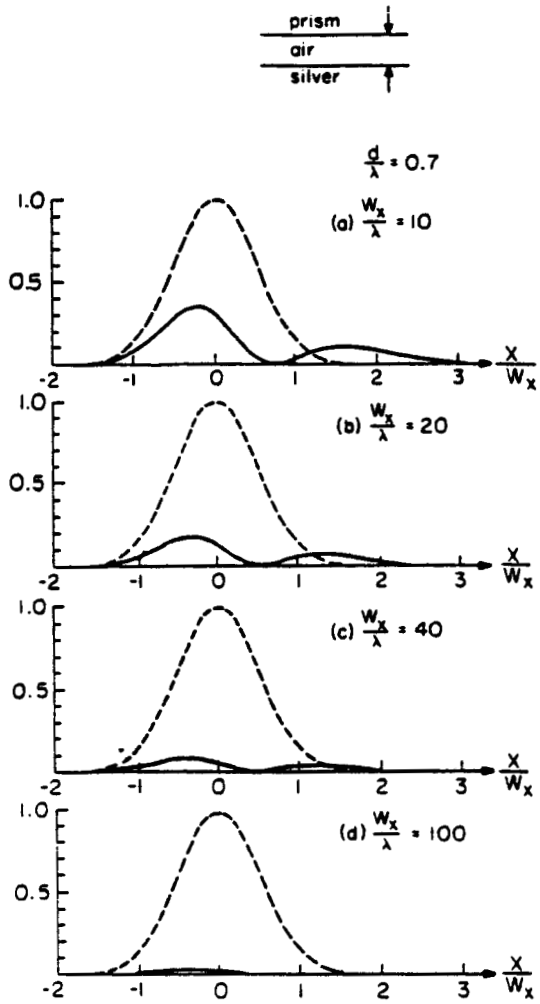


Fig. 5. Optical beam shift for the prism–air–silver configuration with $d/\lambda = 0.7$ for four beamwidths. Backward beam shifts occur here. Dashed curves, incident-beam intensities; solid curves, reflected-beam intensities.

prism–air–silver configuration. The angle of incidence is always assumed to be such that the phase-matching condition is satisfied, i.e., $k_{xi} = \beta_p$. This is the case that is expected to have a maximum displacement. If $k_{xi} \neq \beta_p$, Eq. (9) can still be used, and the beam displacement will be different. Some numerical results for the effect of varying the incidence angle have been discussed in Ref. 8 from a general point of view for a multilayered structure.

A quantity D related to the beam displacement is suggested in Ref. 10:

$$D \equiv i\left[\frac{\mathrm{d}}{\mathrm{d}k_x}\ln R(k_x)\right]_{k_x=k_{x_i}} \simeq \frac{i(k_n - k_p)}{(k_{x_i} - k_n)(k_{x_i} - k_p)}, \quad (25)$$

where the approximate formula for $R(k_x)$ in expression (6) has been used. The real part of D, D', denotes the lateral displacement of the reflected beam. To illustrate the beam-shift phenomenon, D'/λ are calculated for a few values of d/λ and are shown in the table in Fig. 2(b). One sees that the beam shifts can be forward (positive) or backward (negative), depending on the locations of the pole and the zero. For $d/\lambda = 0.4$, as shown in Fig. 4, the beam shift D' is always 9.1 for four beamwidths. If the beamwidth W_x is small, the shift looks large because the horizontal scale is normalized to W_x. A further approximation to D assumes that[10]

$$\beta_n \simeq \beta_p = k_{x_i}, \quad (26)$$

$$D = \frac{\alpha_n - \alpha_p}{\alpha_n \alpha_p}. \quad (27)$$

Thus, if $\alpha_n < 0$ and $\alpha_p > 0$, a positive beam shift occurs as calculated in Fig. 2(b), $d/\lambda = 0.2$, 0.4, and 0.6. When $d/\lambda > 0.7$, $0 < \alpha_n < \alpha_p$, a negative beam shift occurs. These are illustrated in Figs. 4(a)–4(d) and 6(a)–6(b) for positive beam shifts and in Figs. 5(a)–5(d) and 6(c)–6(d) for negative beam shifts with detailed intensity variations. For $W_x/\lambda = 100$, as shown in Figs. 4(d) and 5(d), the incident beam is close to a plane wave; thus the reflected beams are similar to the incident Gaussian beams multiplied by the reflection coefficients at θ_i. In Fig. 5(d), $d/\lambda = 0.7$ and $W_x/\lambda = 100$; the zero k_n is very close to the real axis, as shown in Fig. 2(b). Thus the reflection coefficient is close to zero, and the reflected beam has a small intensity. This is the anomalous absorp-

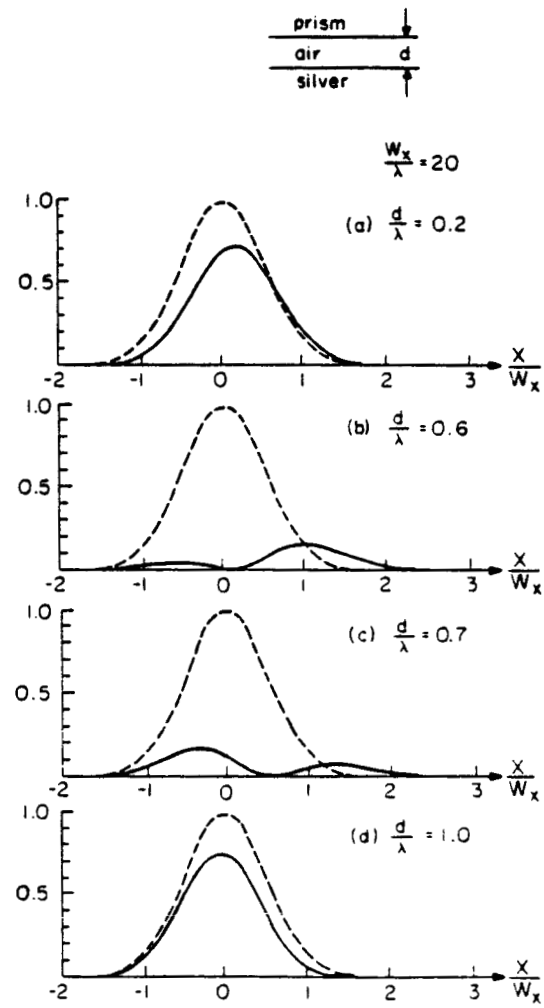


Fig. 6. Optical beam shift for the prism–air–silver configuration with $W_x/\lambda = 20$ and four different air-gap thicknesses, $d/\lambda = 0.2$, 0.6, 0.7, and 1.0. Dashed curves, incident-beam intensities; solid curves, reflected-beam intensities.

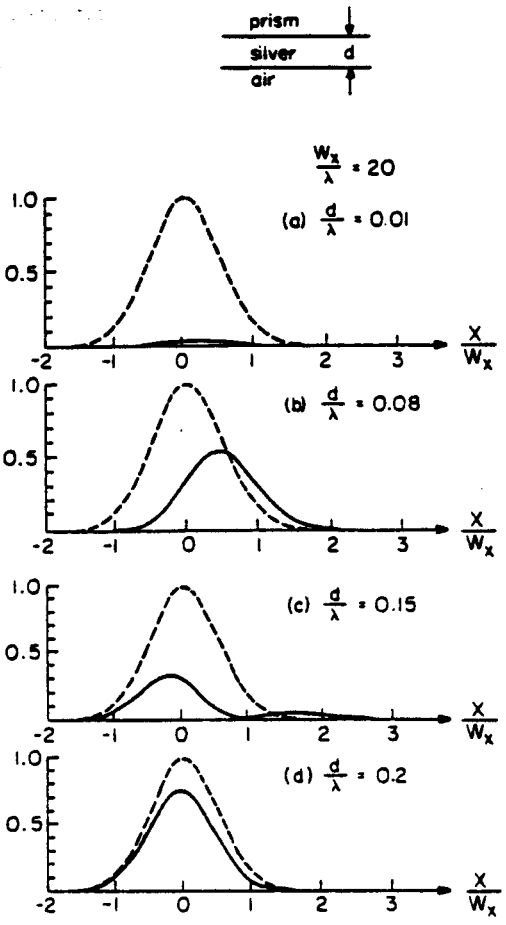


Fig. 7. Optical beam shift for the prism–silver–air configuration with $W_x/\lambda = 20$ and four different silver-film thicknesses, $d/\lambda = 0.01, 0.08, 0.15$, and 0.2.

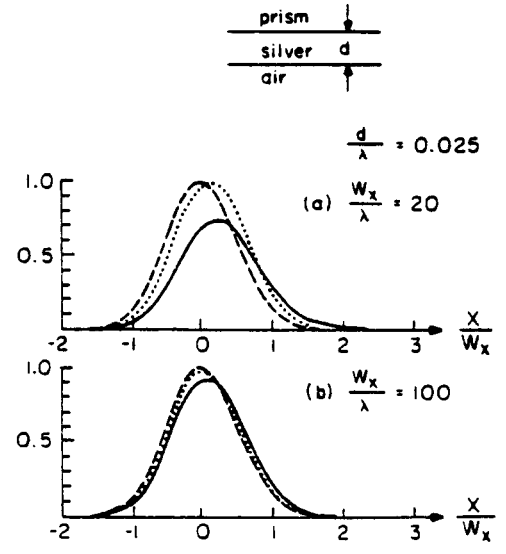


Fig. 8. Optical beam shift for the prism–silver–air configuration with $d/\lambda = 0.025$. As can be seen from Fig. 3(b), the location of the pole is right above the branch point at $\kappa = 1.0$. Thus a direct numerical integration of Eq. (4) is required to obtain the reflected field. Dashed curves, incident-beam intensities; dotted curves, reflected-beam intensities using Eq. (9); solid curves, reflected-beam intensities using the direct numerical integration.

tion, as demonstrated both theoretically and experimentally at microwave frequency in Ref. 9.

In Figs. 7 and 8, the prism–silver–air configuration is considered. D'/λ for both forward and backward beam shifts are calculated in the table of Fig. 3(b). The detailed intensity variations are illustrated in Fig. 7 for different metal thicknesses, $d/\lambda = 0.01, 0.08, 0.15$, and 0.2. When $d/\lambda = 0.025$, the branch-point contribution from $\kappa = n_3 = 1.0$ is important. As mentioned in Section 3, the approximation for the reflection coefficient in expression (6) in the leaky-wave theory[7-10] is not valid. A direct numerical integration using Simpson's rule for the reflected field is performed by using the exact reflection formulas in Eqs. (4) and (11). The beam shift still occurs as expected. This case is different from the Goos–Hänchen shift in the contributions from both the lateral wave (branch-point contribution) and the surface-plasmon mode (pole contribution) in our case. The Goos–Hänchen shift comes only from the branch-point contribution. If one still uses the simple pole approximation, the results (shown as the dotted lines) are quite different from the exact numerical solution (shown as the solid lines) in Fig. 8. For the case in Fig. 8, the simple leaky-wave theory tends to overestimate the reflection coefficient and underestimate the beam shift, if the branch-point contribution is ignored.

At other wavelengths, e.g., $\lambda = 6328$ Å, $\epsilon_{Ag} \simeq (-16.32 + i0.54)\,\epsilon_0$ and $k_{sp}/k \simeq (1.0321 + i0.0011)$. Thus the leaky surface-plasmon pole will be very close to the branch point at $n_3 = 1.0$ for the Kreschmann configuration. Again the leaky-wave approximation in expression (6) will not be appropriate, and a direct numerical integration, as has been done for Fig. 8, is necessary.

5. CONCLUSIONS

The leaky-wave theory has been applied to study both the Otto and the Kreschmann attenuated-total-internal-reflection configurations, when the leaky surface-plasmon modes are excited. It is demonstrated theoretically that the beam shift and intensity variation due to surface-plasmon excitation exist, as predicted from the leaky-wave theory. This beam-shift phenomenon may affect the reflectance measurement if the beamwidth is narrow (for example, much less than 100 wavelengths). A study of the beam-shift phenomenon may also provide information about the dielectric functions of the metal or the semiconductor in the visible or the infrared region, respectively. In addition to the prism configurations, grating structures may also be employed for similar studies.

ACKNOWLEDGMENTS

Stimulating and helpful discussions with G. A. Deschamps are gratefully appreciated. This research was partially supported by grants NAG 1-500 from the National Aeronautics and Space Administration and ECS 83-11345 from the National Science Foundation.

REFERENCES

1. A. Otto, in *Optical Properties of Solids: New Development*, B. O. Seraphin, ed. (North-Holland, Amsterdam, 1975).
2. R. Tamir, *Integrated Optics*, 2nd ed. (Springer-Verlag, New York, 1979), Chap. 3, pp. 83–137.

3. Von F. Goos and H. Hänchen, "Ein neuer und fundamentaler Versuch zur Totalreflexion," Ann. Physik 6, 333–346 (1947).
4. H. K. V. Lotsch, "Reflection and refraction of a beam of light at a plane interface," J. Opt. Soc. Am. 58, 551–561 (1968).
5. T. Tamir and A. A. Oliner, "Role of the lateral wave in total reflection of light," J. Opt. Soc. Am. 59, 942 (1969).
6. B. R. Horowitz and T. Tamir, "Lateral displacement of a light beam at a dielectric interface," J. Opt. Soc. Am. 61, 586–594 (1971).
7. H. L. Bertoni and T. Tamir, "Unified theory of Rayleigh-angle phenomena for acoustic beams at liquid–solid interfaces," Appl. Phys. 2, 157–172 (1973).
8. T. Tamir and H. L. Bertoni, "Lateral displacement of optical beams at multilayered and periodic structures," J. Opt. Soc. Am. 61, 1397–1413 (1971).
9. A. Amittay, P. D. Einziger, and T. Tamir, "Experimental observation of anomalous electromagnetic absorption in thin-layered media," Appl. Phys. Lett. 38, 754–756 (1981).
10. V. Shah and T. Tamir, "Absorption and lateral shift of beams incident upon lossy multilayered media," J. Opt. Soc. Am. 73, 37–44 (1983).
11. P. Halevi, "Polariton modes at the interface between two conducting or dielectric media," Surface Sci. 76, 64–89 (1978).
12. D. L. Mills and A. A. Maradudin, "Properties of surface polaritons in layered structures," Phys. Rev. Lett. 31, 372–375 (1973).
13. A. D. Boardman, ed., *Electromagnetic Surface Modes* (Wiley, New York, 1982).
14. W. P. Chen and J. M. Chen, "Use of surface plasma waves for determination of the thickness and optical constants of thin metallic films," J. Opt. Soc. Am. 71, 189–191 (1981).
15. M. Tacke, "Far infrared laser spectroscopy of surface (Fano) waves for the determination of the optical constants of solid," Sov. J. Quantum Electron. 12, 662–664 (1982).
16. M. Marschall, B. Fischer, and H. J. Queisser, "Dispersion of surface plasmons in InSb," Phys. Rev. Lett. 27, 95–97 (1971).
17. P. M. van den Berg and J. C. M. Borburgh, "Dispersion of surface plasmons in InSb-gratings," Appl. Phys. 3, 55–60 (1974).
18. R. H. Ritchie, E. T. Arakawa, J. J. Cowan, and R. N. Hamm, "Surface-plasmon resonance effect in grating diffraction," Phys. Rev. Lett. 21, 1530–1533 (1968).
19. P. K. Tien and R. Ulrich, "Theory of prism-film coupler and thin-film light guides," J. Opt. Soc. Am. 60, 1325–1337 (1970).
20. R. Ulrich, "Theory of the prism-film coupler by plane-wave analysis," J. Opt. Soc. Am. 60, 1337–1350 (1970).
21. M. L. Dakss, L. Kuhn, P. F. Heidrich, and B. A. Scott, "Grating coupler for efficient excitation of optical guided waves in thin films," Appl. Phys. Lett. 16, 523–525 (1970).
22. D. L. Begley, D. A. Bryan, R. W. Alexander, and R. J. Bell, "Optimization of prism coupling efficiency for surface electromagnetic wave excitation in the infrared," Appl. Opt. 16, 1549–1552 (1977).
23. R. Miller, D. L. Begley, and R. J. Bell, "Surface electromagnetic wave coupling efficiency dependence on prism gap height," Appl. Opt. 16, 3077–3079 (1977).
24. R. T. Deck, D. Sarid, G. A. Olson, and J. M. Elson, "Coupling between finite electromagnetic beam and long-range surface-plasmon mode," Appl. Opt. 22, 3397–3405 (1983).
25. W. P. Chen, G. Ritchie, and E. Burstein, "Excitation of surface electromagnetic waves in attenuated total-reflection prism configurations," Phys. Rev. Lett. 37, 993–997 (1976).
26. P. Mazur and B. Djafari-Rouhani, "Effect of surface polaritons on the lateral displacement of a light beam at a dielectric interface," Phys. Rev. B 30, 6759–6762 (1984).
27. J. A. Kong, *Theory of Electromagnetic Waves* (Wiley, New York, 1975), especially p. 119 for reflection coefficient and p. 218 for branch cuts.
28. W. C. Chew, "The singularities of a Fourier-type integral in a multicylindrical layer problem," IEEE Trans. Antennas Propag. AP-31, 653–655 (1983).