Light Waves in Thin Films and Integrated Optics

P. K. Tien

Integrated optics is a far-reaching attempt to apply thin-film techno ogy to optical circuits and devices, and, by using methods of integrated circuitry, to achieve a better and more economical optical system. The specific topics discussed here are physics of light waves in thin films, materials and losses involved, methods of coupling a light beam into and out of a thin film, and nonlinear interactions in waveguide structures. The purpose of this paper is to review in some detail the important development of this new and fascinating field, and to caution the reader that the technology involved is difficult because of the smallness and perfection demanded by thin-film optical devices.

I. Introduction

Since the invention of the solid and gas lasers a decade ago, there have been immense advances in optoelectronics. For example, laser transitions now cover the spectral region from ultraviolet light to millimeter waves. Optical modulation, frequency mixing, and parametric oscillations have been extensively studied. Almost all the experiments in the past were performed invariably in bulk materials and with a light beam of nearly gaussian intensity distribution. Recently, however, the introduction of the concept of integrated optics[1,2] and the development of the prism–film[3] and grating[4,5] couplers have raised important questions concerning the future needs of optical systems: Can we conform to the idea of the integrated optics and develop optical modulators, frequency converters, and parametric oscillators in a planar thin-film form which are more efficient than their bulk counterparts? Is there any advantage of using a waveguide structure for nonlinear, electrooptical, or light-wave scattering experiments? While many are devoted to the development of ever better bulk crystals for optical devices, could it be that what we really need in the long run are thin crystal films and epitaxial layers?

One cannot answer the above questions until the complex technology involved in the fabrication of thin-film structures is solved. On the other hand, as far as optical systems are considered, certain features of the thin-film devices appear to be definitely advantageous. First, all the elements of a thin-film device are exposed on the surface and are easily accessible for probing, measurement, or modification. Second, compared to microwaves, the optical wavelength is a factor of 10^4 smaller. The thin-film optical devices can be made very small and they can be placed one next to the other on a single substrate, forming an optical system which is naturally more compact, less vulnerable to the environmental changes, and more economical. Third, since the film has a thickness comparable to the optical wavelength and since most of the light energy is confined within the film, the light intensity inside the film can be very large even at a moderate laser power level. For example, 1 W of laser power can easily result in a power density of 22 MW/cm^2 in a ZnS film 0.46 μm thick. This large power density is important in nonlinear interactions. Finally, the phase velocity of a light wave in a thin-film waveguide depends on the thickness of the film and the mode of propagation. This provides new possibilities in the design of experiments and devices.

Although this field of integrated optics or thin-film optoelectronics is still in a very elementary stage, in the past two years there have been a number of very important advances which generated substantial excitement. In what follows we will briefly enumerate some of these advances.

The first advance is the great improvement in the method of coupling a light wave propagating in free space into a well-defined mode of the thin-film guide. This was brought about through the invention of the prism–film coupler as described by Tien *et al.*[3] Other theories and experiments of the prism–film coupler have been given by Tien and Ulrich,[6] Ulrich,[7] Midwinter,[8] Harris *et al.*,[9] and Harris and Shubert.[10] Recently, a coupling efficiency of 88% was reported by Ulrich.[11] Another type of coupler, which uses a grating in place of the prism, was reported by Dakss *et al.*[4] and by Kogelnik and Sosnowski.[5] Recently an extremely

The author is with Bell Telephone Laboratories, Inc., Holmdel, New Jersey 07733.

Received 3 May 1971.

simple technique of coupling which involves a tapered film edge was described by Tien and Martin.[12] A second advance which has occurred involves the materials of the films used for light-wave propagation. Initial experiments of Tien *et al.*[3] used sputtered ZnO and vacuum-evaporated ZnS films, which have large scattering losses. Then, excellent sputtered glass films were developed by Goell and Standley.[13] Other useful materials include polyurethane and polyester epoxy films developed by Harris *et al.*,[9] Ta_2O_5 films by Hensler *et al.*,[14] and polymerized organosilicon films by Tien *et al.*[15] The organosilicon films made from vinyltrimethylsilane and hexamethyldisiloxane monomers have a loss of the order of 0.04 dB/cm as opposed to a number which was several hundred times larger in some of the earlier experiments. Other advances involve nonlinear and electrooptic effects observed in thin films and laser oscillation in iterated film structures. Tien *et al.*[16] have observed optical second harmonic generation using a ZnS film on a single-crystal ZnO substrate. The nonlinear interaction in this experiment was enhanced by the large concentration of light energy in the vicinity of the film. Kuhn *et al.*[17] succeeded in deflection of an optical guided wave in a thin film by its interaction with a surface acoustic wave. Hall *et al.*[18] have shown that it is possible to control the intensity of the light wave which propagates through a thin film such as a GaAs depletion layer by applying an appropriate bias. The latter two experiments indicate the possibility of efficient modulation of light in thin-film waveguides. Finally, laser oscillation was reported by Kogelnik and Shank[19] in a thin-film grating structure which is doped with a dye of Rhodamine 6G.

The basic problem considered here is simply a dielectric or semiconductor film that is deposited on a substrate. When a light beam in space is fed into a film, the light beam adapts itself so that it is confined within the thickness of the film. However, the transverse dimension of the light beam is not restricted. The film can thus be considered as a slab waveguide. Mathematical solutions of the waveguide show many possible modes of light-wave propagation. Within the plane of the film, the light wave in any of these waveguide modes is allowed to propagate in any direction and can be reflected or refracted at any given boundary. The problem can be thought of as two-dimensional optics. Moreover, the light wave can be made to interact with the material of the film or of the substrate, or with externally applied electric or microwave fields so that certain functions such as modulation of light, parametric interaction, etc., can be achieved. Integrated optics is therefore an interdisciplinary science; it involves materials, film fabrication, electronics, and physical optics. In this paper, we select a few topics that are essential to the understanding of this new and complex field and discuss them in sufficient detail, hoping that the paper can serve as an introductory review for newcomers as well as a useful reference for those already in the field. The specific topics chosen are waveguide and radiation modes, the light-wave couplers, materials and losses of thin-film

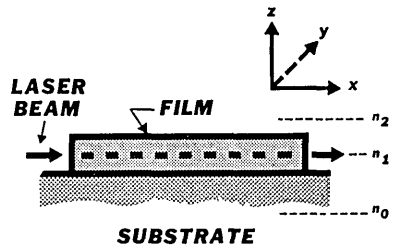

Fig. 1. Coordinate system that will be used throughout the paper. The light wave propagates in the film parallel to the x axis. The surface of the film is in the xy plane and its thickness in the z direction.

waveguides, and nonlinear interactions among optical guided waves. We will not discuss passive optical circuits and their fabrication, which have been reviewed in several excellent papers.[1,2,20]

II. Waveguide and Radiation Modes

The film considered here has a thickness on the order of 1 μ or less; it is so thin that it has to be supported by a substrate. We thus consider three media: a film, an air space above, and a substrate below. As shown in Figure. 1, the thickness of the film is in the X–Y plane. For a thin film to support propagating modes and to act as a dielectric waveguide for the light waves, the refractive index of the film n_1 must be larger than that of the substrate n_0 and naturally that of the air space above n_2. A typical experiment is shown in Fig. 2. Here the entire surface of a 7.6-cm by 2.5-cm microscope glass slide is coated with a layer of an organic film made from a vinyltrimethylsilane monomer by gas discharge. At a wavelength of 6328 Å of the

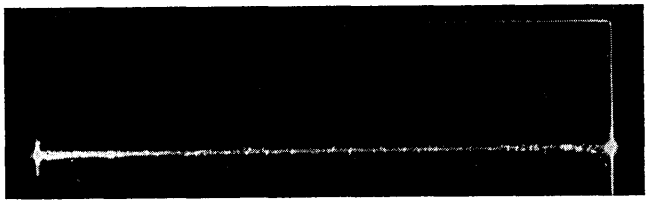

Fig. 2. Light beam in an organic film. The film is coated on a 2.5-cm by 7.6-cm microscope glass slide and it serves as a dielectric waveguide for the light wave.

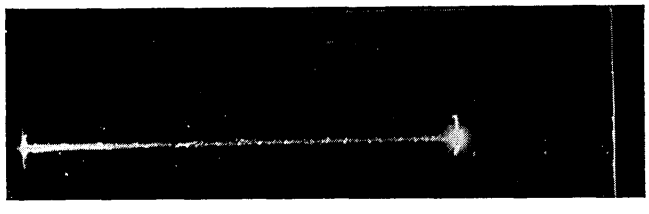

Fig. 3. To show that the light wave was truly propagating in the film, we scratched the film and observed that the light beam stopped immediately at the scratched point, which then radiated brightly.

helium–neon laser, the refractive index of the film is 1.5301, which is larger than that of the glass (1.5125) and also that of the air (1.00). A light beam was fed into the film at the left side of the figure. It propagated through the entire length of the film and then radiated into the free space at the right side of the film. To show that the light wave was truly propagating inside the film, we scratched the film as shown in Fig. 3. The light beam then stopped at the scratched point, which radiated brightly as an antenna. Mathematically, the problem involves a solution of the Maxwell equations that matches the boundary conditions at the film–substrate and film–air interfaces. The solutions indicate three possible modes of propagation. The light wave can be bound and guided by the film as the *waveguide modes*. It can radiate from the film into both of the air and substrate spaces as the *air modes*, or it can radiate into the substrate only as the *substrate modes*. The air and substrate modes are the radiation modes discussed by Marcuse.[21] The modes described above can be explained simply and elegantly by the Snell law of refraction and the related total internal reflection phenomenon in optics.

Let (Fig. 4a) n_0, n_1, and n_2 be the refractive indices and θ_0, θ_1, θ_2 be the angles measured between the light paths and the normals of the interfaces in the substrate, film, and air, respectively. Here $n_1 > n_0 > n_2$. We have then from the Snell law

$$\sin\theta_2/\sin\theta_1 = n_1/n_2, \tag{1}$$

and

$$\sin\theta_0/\sin\theta_1 = n_1/n_0. \tag{2}$$

Let us increase θ_1 gradually from 0. When θ_1 is small, a light wave, for example, starts from the air space above the film, can be refracted into the film, and is then refracted again into the substrate (Fig. 4a). In this case, the waves propagate freely in all the three media—air, film, and substrate—and they are the radiation fields that fill all the three spaces (air modes). Next, as θ_1 is increased to a value larger than the critical angle $\sin^{-1}(n_2/n_1)$ of the film–air interface as shown in Fig. 4(b), the impossible condition incurred in Eq. (1), $\sin\theta_2 > 1$, indicates that the light wave is totally reflected at the film–air boundary. Now the wave can no longer propagate freely in the air space. We thus describe a solution that the light energy in the film radiates into the substrate only (substrate modes). Finally, when θ_1 is larger than the critical angle $\sin^{-1}(n_0/n_1)$ of the film–substrate interface, the light wave as shown in Fig. 4(c) is totally reflected at both the upper and lower surfaces of the film. The energy flow is then confined within the film; that is to be expected in the waveguide modes.

It is interesting to note that in the waveguide modes, the light wave in the film follows a zigzag path (Fig. 4c). The light energy is trapped in the film as the wave is totally reflected back and forth between the two film surfaces. We can represent this zigzag wave motion by two wave vectors A_1 and B_1 in Fig. 5(a). We then divide the wave vectors into the vertical and horizontal components in Fig. 5(b). The horizontal

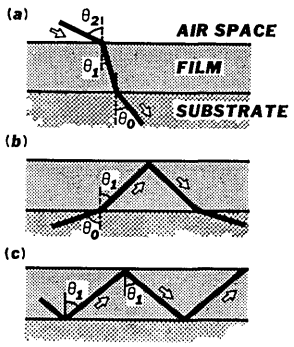

Fig. 4. (a) When $\theta_1 < \sin^{-1}(n_2/n_1)$, the light wave shown represents the air mode. According to ray optics, the light wave originated in the film is refracted into both the substrate and air space. (b) As θ_1 increases so that $\sin^{-1}(n_2/n_1) < \theta_1 < \sin^{-1}(n_0/n_1)$, the light wave shown now represents the substrate mode. It is refracted into the substrate but is totally reflected at the film–air interface. (c) When θ_1 increases further so that $\theta_1 > \sin^{-1}(n_0/n_1)$, the light wave shown is totally reflected at both the film–air and film–substrate interfaces. It is confined in the film as is to be expected in the waveguide mode.

components of wave vectors A_1 and B_1 are equal, indicating that the waves propagate with a constant speed in a direction parallel to the film. The vertical component of the wave vector A_1 represents an upward-traveling wave; that of the wave vector B_1, a downward-traveling wave. When the upward- and downward-traveling waves are superposed, they form a standing wave field pattern across the thickness of the film. By changing θ_1, we change the direction of the wave vectors A_1 and B_1 and thus their horizontal and vertical components. Consequently, we change the wave velocity parallel to the film as well as the standing wave field pattern across the film.

Since we discuss here a planar geometry, the waves described above are plane waves. They are TE waves if they contain the field components E_y, H_z, and H_x; they are TM waves if they contain the field components H_y, E_z, and E_x. Here x is the direction of the wave propagation parallel to the film. The wave vectors A_1 and B_1 discussed above have thus a magnitude kn_1, where $k = \omega/c$ and ω and c are, respectively, the angular frequency of the light wave and the speed of

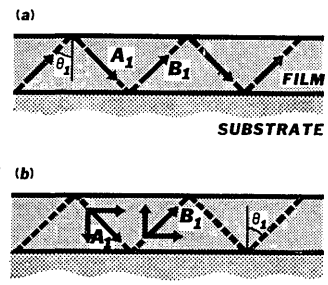

Fig. 5. (a) Light wave in the waveguide mode can be considered as a plane wave which propagates along a zigzag path in the film. The wave can be represented by two wave vectors A_1 and B_1. (b) The wave vectors A_1 and B_1 can be decomposed into vertical and horizontal components. The horizontal components $kn_1 \sin\theta_1$ determine the wave velocity parallel to the film. The vertical components $\pm kn_1 \cos\theta_1$ determine the field distribution across the thickness of the film.

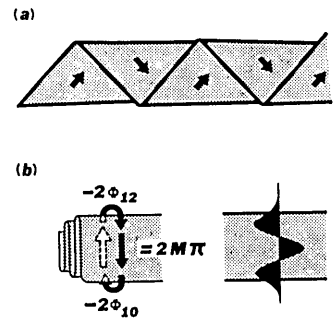

Fig. 6. (a) In wave optics, a light wave in the waveguide mode is an infinitely wide sheet of plane wave which folds back and forth in a zigzag manner between the top and the bottom surface of the film. (b) A light wave propagating inside the film is totally reflected at the two film surfaces. The figure shows that in order for the wave and its reflections to add in phase, the total phase change for the light wave to travel across the thickness of the film, up and down in one round trip, must be equal to $2m\pi$. The figure also shows that the light wave suffers a phase change of $-2\Phi_{12}$ and $-2\Phi_{10}$ at the upper and lower film surfaces, respectively. These phase changes determine the field distribution across the thickness of the film, which is shown at the right of the figure for the $m = 3$ waveguide mode.

light in vacuum. In the picture of the wave optics, the vectors A_1 and B_1 are the normals of the wavefronts, when an infinitely wide sheet of plane wave folds back and forth in a zigzag manner between the two film surfaces (Fig. 6a). Now consider an observer who moves with the wave in the direction parallel to the film. He does not see the horizontal components of the wave vectors. What he observes is a plane wave that folds upward and downward, one directly on top of the other as shown in Fig. 6(b). The condition, then, for all those multiple reflected waves to add in phase, as seen by this observer, is that the total phase change experienced by the plane wave for it to travel one round trip, up and down across the film, should be equal to $2m\pi$, where m is an integer. Otherwise, if after the first reflections from the upper and lower film surfaces, the phase of the reflected wave differs from the original wave by a small phase δ, the phase differences after the second, third,..., reflections would be $2\delta,3\delta,\ldots$, and then the waves of progressively larger phase differences would add finally to zero. As shown in Fig. 5(b), the vertical components of the wave vectors A_1 and B_1 have a magnitude $kn_1\cos\theta_1$. The phase change for the plane wave to cross the thickness W of the film twice (up and down) is then $2kn_1W\cos\theta_1$. In addition, the wave suffers a phase change of $-2\Phi_{12}$ due to the total reflection at the upper film boundary and, similarly, a phase change of $-2\Phi_{10}$ at the lower film boundary. Here, the phases $-2\Phi_{12}$ and $-2\Phi_{10}$ represent, in fact, the Goos-Haenchen shifts.[22] Consequently, in order for the waves in the film to interfere constructively, we have

$$2kn_1W\cos\theta_1 - 2\Phi_{10} - 2\Phi_{12} = 2m\pi, \qquad (3)$$

which is the condition for the waveguide modes. Here $m = 0, 1, 2, 3,\ldots$, is the order of the mode. According to Born and Wolf[23] on the theory of total reflection,

$$\tan\Phi_{12} = (n_1^2\sin^2\theta_1 - n_2^2)^{\frac{1}{2}}/(n_1\cos\theta_1);$$
$$\tan\Phi_{10} = (n_1^2\sin^2\theta_1 - n_0^2)^{\frac{1}{2}}/(n_1\cos\theta_1) \qquad (4)$$

for the TE waves, and

$$\tan\Phi_{12} = n_1^2(n_1^2\sin^2\theta_1 - n_2^2)^{\frac{1}{2}}/(n_2^2n_1\cos\theta_1);$$
$$\tan\Phi_{10} = n_1^2(n_1^2\sin^2\theta_1 - n_0^2)^{\frac{1}{2}}/(n_0^2n_1\cos\theta_1) \qquad (5)$$

for the TM waves.

It is clear that in spite of the zigzag wave motion described above, the wave in a waveguide mode appears to propagate in the horizontal direction only; the vertical part of the wave motion simply forms a standing wave between the two film surfaces. To avoid confusion, it is desirable to use β and v exclusively for the phase constant and the wave velocity parallel to the film. Thus,

$$\beta = kn_1\sin\theta_1, \qquad v = c(k/\beta). \qquad (6)$$

Another quantity which will also be used frequently is the ratio β/k. As shown in Eqs. (6), it is the ratio of the speed of light in vacuum to the speed of wave propagation in the waveguide.

After substituting Eqs. (4) or (5) into Eq. (3), we find that both Eqs. (3) and (6) are transcendental equations. Fortunately, the transcendental functions involve θ_1 only. For a given n_0, n_1, n_2, and m we may easily compute both β/k and W for a common θ_1, and then tabulate β/k and W by assigning different values for θ_1. The curves showing W vs β/k using m as the parameter are the mode characteristics of the waveguide. They will be shown later, for example, in Fig. 23.

To summarize, any radius of the quarter-circle shown in Fig. 7 represents a possible direction for the

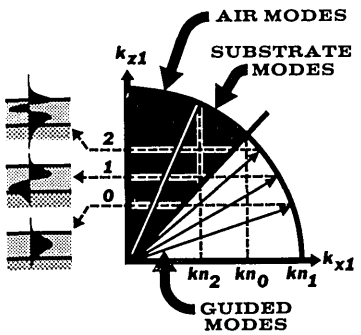

Fig. 7. Any radius of the quarter-circle at the right side of the figure represents a possible direction for the wave vector B_1. In the black region of the circle, the wave vector represents the substrate or air mode. In the white region of the circle, the wave vector represents the waveguide mode, but only a discrete set of the directions in this region satisfies the equation of the waveguide modes. Each direction of this discrete set represents one waveguide mode and each waveguide mode has its own field distribution as shown in the left side of the figure.

wave vector B_1 described above, and θ_1 is the incident angle measured between the wave vector and the vertical axis. The waveguide modes occur in the range $\sin^{-1}(n_0/n_1) < \theta_1 < \pi/2$. Within this range of θ_1, there is a discrete set of the directions which satisfies the equation of the modes (3). Each direction corresponds to one waveguide mode of the film. The horizontal component of the wave vector, $kn_1 \sin\theta_1$, determines the wave motion parallel to the film, while its vertical component, $kn_1 \cos\theta_1$, determines the standing wave field pattern across the film. As shown in the left side of Fig. 7, when $m = 0$, the standing wave pattern has a form similar to a half-sine wave. When $m = 1$, it has a form similar to a full sine wave, and so on. The air and substrate modes occur in the range $0 < \theta_1 < \sin^{-1}(n_0/n_1)$; they occupy the black region of the quarter-circle. As we vary θ_1 continuously from 0 to $\sin^{-1}(n_2/n_1)$ for the air modes and $\sin^{-1}(n_2/n_1)$ to $\sin^{-1}(n_0/n_1)$ for the substrate modes, the corresponding θ_0 and θ_2 sweep through the entire space of the substrate and the air space. It is thus possible to express any radiation field by superposing waves of the air and substrate modes. What we have discussed here is therefore simply an expansion of the solution of the Maxwell equation into plane waves of all possible directions.

III. Wave Equation and the Field Distribution

Having described the modes of light-wave propagation purely on an intuitive basis, we may now derive them mathematically. For simplification, assume the light wave in the film to be infinitely wide in the Y direction so that $\partial/\partial y = 0$ (Fig. 1). Let X be the direction of the wave propagation parallel to the film. The Maxwell equations in E_y for TE waves (or H_y for TM waves) can be reduced to the wave equation below.

$$\partial^2 E/\partial x^2 + \partial^2 E/\partial z^2 = -(kn_j)^2 E, \quad j = 0, 1, \text{or } 2, \qquad (6)$$

where n_j is the refractive index of the medium j. The subscripts $j = 0,1$, and 2 denote the substrate, the film, and the air space, respectively. A time dependence $\exp(-i\omega t)$ is used in Eq. (6), where $i = \sqrt{-1}$. The solution of the wave equation is in the form of $\exp(ik_{xj}x)\exp(\pm ik_{zj}z)$, which may be substituted into Eq. (6) to obtain

$$k_{xj}^2 + k_{zj}^2 = (kn)_j^2. \qquad (7)$$

The boundary conditions at the film–air and film–substrate interfaces demand a same wave motion parallel to the film in all the three media considered; we may thus put

$$k_{x0} = k_{x1} = k_{x2} = \beta. \qquad (8)$$

All the fields thus vary in time and x according to the factor $\exp(-i\omega t + i\beta x)$. This common factor will be omitted in all the later expressions for simplification. Combining Eqs. (7) and (8), we obtain an important relation,

$$k_{zj} = (k^2 n_j^2 - \beta^2)^{\frac{1}{2}}. \qquad (9)$$

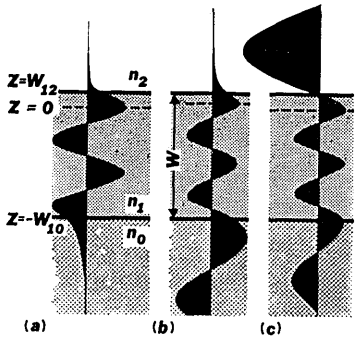

Fig. 8. The electric field distribution of (a) a TE waveguide mode; (b) a TE substrate mode; (c) a TE (even) air mode.

In the film, k_{x1} and k_{z1} are the horizontal and vertical components of the wave vector A_1 or B_1 discussed before. They are, respectively, $k_{x1} = \beta = kn_1 \sin\theta_1$ and $k_{z1} = kn_1 \cos\theta_1$. In the waveguide modes, we find from Eq. (9) and from the condition $\sin^{-1}(n_0/n_1) < \theta_1 < \pi/2$ that $kn_0 < \beta < kn_1$, k_{z1} is real, and k_{z0} and k_{z2} are imaginary. The field distribution in Fig. 8(a) is thus a standing wave in the film and exponential in the substrate and in the air space. Next, for the substrate modes, we find from Eq. (9) and from the condition $\sin^{-1}(n_2/n_1) < \theta_1 < \sin^{-1}(n_0/n_1)$ that k_{z1} and k_{z0} are real, but k_{z2} is imaginary. The fields in this case are standing waves in the film and in the substrate, but exponential in the air space (Fig. 8b). Finally, for the air modes, we find that $0 < \theta_1 < \sin^{-1}(n_2/n_1)$, and k_{z0}, k_{z1}, and k_{z2} are all real. The fields in all the three media are now standing waves (Fig. 8c). It is convenient to denote k_{zj} by b_j when it is real and by ip_j when it is imaginary. For $n_0 \neq n_2$, such as the case that is considered throughout this paper, the waveguide is asymmetric. We choose $z = W_{12}$ and $z = -W_{10}$ as the upper and lower film surfaces. The thickness of the film is then $W = W_{10} + W_{12}$.

The field distributions are derived by choosing $z = 0$ at the position where E_y is maximum for any waveguide, substrate, or even air mode, and $E_y = 0$ for any odd air mode. It is important to note that these positions of $z = 0$ are different for different modes in an asymmetric waveguide. These choices are necessary in order to simplify mathematics so that we can visualize the field distributions of various modes easily. To avoid confusion, we consider below E_y of a TE wave only.

For the waveguide modes, as mentioned earlier, the wave suffers a phase change of $-2\Phi_{12}$ at the upper film surface, and a phase change of $-2\Phi_{10}$ at the lower film surface because of the internal total reflections. The fields at the two film surfaces must therefore be $\pm A\cos\Phi_{12}$ and $\pm A\cos\Phi_{10}$, respectively, where A is a constant. Let the field at $z = 0$ be a maximum value, A. Then, we choose $k_{z1}W_{12}$ (or b_1W_{12}) $= \Phi_{12}$ so that the field at the upper film surface, $z = W_{12}$, can be $A\cos\Phi_{12}$. Similarly we choose $b_1W_{10} = \Phi_{10} + m\pi$ so that the field at the lower film surface, $z = -W_{10}$, can be $A\cos\Phi_{10}$ if m = even and $-A\cos\Phi_{10}$ if m = odd as shown in Fig. 8(a). These choices give $b_1W =$

Table I. Electric Field Distribution in (a) a Waveguide Mode, (b) a Substrate Mode, and (c) the Even and Odd Air Modes[a]

Waveguide Mode

$\sin^{-1}(n_0/n_1) < \theta_1 < \pi/2;\ kn_0 < \beta < kn_1$

Medium	k_{xj}	k_{zj}	E_y (TE wave)
Film	$=\beta$	$=b_1$	$A\cos b_1 z$
Substrate	$=\beta$	$=ip_0$	$A\cos(\Phi_{10}+m\pi)\exp[-p_0(\lvert z\rvert - W_{10})]$
Air-space	$=\beta$	$=ip_2$	$A\cos\Phi_{12}\exp[-p_2(\lvert z\rvert - W_{12})]$

Substrate Mode

$\sin^{-1}(n_2/n_1) < \theta_1 < \sin^{-1}(n_0/n_1);\ \sin^{-1}(n_2/n_0) < \theta_0 < \pi/2;\ kn_2 < \beta < kn_0$

Medium	k_{xj}	k_{zj}	E_y (TE wave)
Film	$=\beta$	$=b_1$	$A\cos b_1 z$
Substrate	$=\beta$	$=b_0$	$\frac{1}{2}A\,[\cos(b_1W_{10}) - i(b_1/b_0)\sin(b_1W_{10})]\cdot\exp[-ib_0(\lvert z\rvert - W_{10})] + \text{c.c.}$
Air space	$=\beta$	$=ip_2$	$A\cos\Phi_{12}\exp[-p_2(\lvert z\rvert - W_{12})]$

Even and Odd Air Modes

$0 < \theta_1 < \sin^{-1}(n_2/n_1);\ 0 < \theta_0 < \sin^{-1}(n_2/n_0);\ 0 < \theta_2 < \pi/2;\ 0 < \beta < kn_2$

Medium	k_{xj}	k_{zj}		
Film	$=\beta$	$=b_1$	Even	$A\cos b_1 z$
			Odd	$A\sin b_1 z$
Substrate	$=\beta$	$=b_0$	Even	$\frac{1}{2}A[\cos(b_1W_{10}) - i(b_1/b_0)\sin(b_1W_{10})]\exp[-ib_0(\lvert z\rvert - W_{10})] + \text{c.c.}$
			Odd	$-\frac{1}{2}A[\sin(b_1W_{10}) + i(b_1/b_0)\cos(b_1W_{10})]\exp[-ib_0(\lvert z\rvert - W_{10})] + \text{c.c.}$
Air space	$=\beta$	$=b_2$	Even	$\frac{1}{2}A[\cos(b_1W_{12}) - i(b_1/b_2)\sin(b_1W_{12})]\exp[-ib_2(\lvert z\rvert - W_{12})] + \text{c.c.}$
			Odd	$\frac{1}{2}A[\sin(b_1W_{12}) + i(b_1/b_2)\cos(b_1W_{12})]\exp[-ib_2(\lvert z\rvert - W_{12})] + \text{c.c.}$

[a] In deriving these expressions, we have chosen $z = 0$ at the position where E_y is either zero or maximum. These positions of $z = 0$ are therefore different for different modes.

$b_1W_{10} + b_1W_{12} = \Phi_{12} + \Phi_{10} + m\pi$, which satisfies Eq. (3). The boundary conditions require E_y and $\partial E_y/\partial_z$ to be continuous at the two interfaces. We have, therefore,

$$E_y = A\cos\Phi_{12}\exp[-p_2(\lvert z\rvert - W_{12})]$$

in the air space and

$$E_y = A\cos(\Phi_{10} + m\pi)\exp[-p_0(\lvert z\rvert - W_{10})]$$

in the substrate.

For the substrate modes, we again assume a maximum field A at $z = 0$ and choose $b_1W_{12} = \Phi_{12}$ (Fig. 8b). The field at $z = W_{12}$ is still $A\cos\Phi_{12}$ and that in the air space is still $A\cos\Phi_{12}\exp[-p_2(z - W_{12})]$. The field at the lower film surface is then $A\cos(b_1W_{10})$ and that in the substrate is

$$\tfrac{1}{2}A[\cos(b_1W_{10}) - i(b_1/b_0)\sin(b_1W_{10})]\exp[-ib_0(\lvert z\rvert - W_{10})] + \text{the complex conjugate.}$$

For the air modes, the even and odd modes must be treated separately. For an asymmetric waveguide, we can choose the $z = 0$ plane anywhere between $z = W_{12}$ and $z = -W_{10}$. However, once it is chosen, the same $z = 0$ plane should be used for all the air modes. For the even modes, the field is a maximum at $z = 0$ and the fields at the two film surfaces are $A\cos b_1W_{12}$ and $A\cos b_1W_{10}$, respectively (Fig. 8c).The boundary conditions require the fields in the substrate and in the air space in the form

$$\tfrac{1}{2}A[\cos(b_1W_{1j}) - i(b_1/b_j)\sin(b_1W_{1j})]\exp[-ib_j(\lvert z\rvert - W_{1j})] + \text{the complex conjugate,}$$

where $j = 0$ and 2. For the odd modes the field is zero at $z = 0$ and is $A\sin(b_1W_{12})$ and $-A\sin(b_1W_{10})$ at the film surfaces. The fields in the substrate and air space are then

$$\pm\tfrac{1}{2}A[\sin(b_1W_{1j}) + i(b_1/b_j)\cos(b_1W_{1j})]\exp[-ib_j(\lvert z\rvert - W_{1j})] + \text{the complex conjugate,}$$

where the plus sign is for $j = 2$ and the minus sign is for $j = 0$. The results discussed above are summarized in Table I.

Mathematically, the field distributions described above are identical to those of the problem of a square potential well in quantum mechanics. Here the air space and the substrate are the potential barriers. We divide the wave energy here into the horizontal and vertical components, keeping the total energy constant. It is the vertical component of the wave energy that negotiates the potential barriers mentioned above. The wave vector represents the momentum and its square, the wave energy. Within the interval $\beta = kn_1$ and $\beta = kn_0$, because of the large horizontal component of the wave vector β, the vertical component of the energy is small enough so that the wave, or the particle, is trapped in the potential well. The mode spectrum or the energy level is thus discrete (waveguide modes). As the horizontal component of the mo-

mentum is reduced to a value $\beta < kn_0$, the vertical component of the wave energy is large enough to overcome the lower potential barrier. The wave function spills over the entire substrate space and we enter into the region of the substrate modes. The mode spectrum or the energy level is now continuous. As we increase further the vertical component of the wave energy by reducing β below kn_2, the wave can spill over the upper and the lower barriers. The mode spectrum remains continuous and it belongs to the air modes.

IV. Light-Wave Couplers

The development of the light-wave couplers in the past two years is an important step forward in thin-film optoelectronics. We can now couple a laser beam efficiently into and out of any thin-film structure and can excite there any single mode of light-wave propagation. In both the prism–film and grating couplers, we feed a light beam into a film through a broad surface of the film and thus avoid the difficult problem of focusing a light beam through a rough film edge. Since the film and the prism (or the grating) are coupled over a length of many optical wavelengths, we can imagine energy transfer taking place continuously between them as waves propagate over the coupled region. It is possible to discuss this type of distributed couplers by a unified theory. We can further show that these couplers have the same optimum coupling efficiency of about 81%, provided that both the coupling strength and the intensity of the incoming laser beam are uniformly distributed over the entire coupling length. An even better efficiency can be achieved by varying the coupling strength along the coupling length in a prescribed manner. By simplifying an earlier theory[6] and by using illustrative figures, we will describe below the principles of the couplers and derive their coupling efficiency in very simple terms.

Figure 9 shows a prism–film coupler. In order to excite all possible waveguide modes in the film, the refractive index of the prism n_3 should be larger than that of the film n_1. An incoming laser beam enters the prism and is totally reflected at the base of the prism. Because of the total reflection, the field in the prism is a standing wave that continues into an exponentially decreasing function below the base of the prism. The part of the field that extends below the prism base is called the evanescent field, since it decreases rapidly away from the prism and does not represent a free radiation. If we represent the incoming wave in the prism by a wave vector A_3 (Fig. 10), it has a magnitude kn_3 and can be decomposed into a horizontal component $kn_3 \sin\theta_3$ and a vertical component $kn_3 \cos\theta_3$. The boundary conditions of the electromagnetic fields at the prism base require that the fields below and above the prism base have the same horizontal wave motion. The evanescent field varies therefore as $\exp(ikn_3x \sin\theta_3)$ in x. Now we place the prism on top of a thin film, maintaining a small but uniform air gap between the base of the prism and the top surface of the film. For effective coupling, the spacing of the air gap is on the order of one-eighth to one-fourth of the vacuum optical wavelength. The evanescent field below the prism then penetrates into the film and excites a light wave into the film. We call this coupling process the optical tunneling. As discussed in Sec. I, the film has many waveguide modes. If the horizontal component of the wave vector A_1 or B_1 of one of the waveguide modes happens to be equal to that of the incoming light wave in the prism $kn_3 \sin\theta_3$, the light wave in the prism is coupled exclusively to this waveguide mode and the laser beam is said to be in a synchronous direction. It is therefore possible to couple the light wave to any waveguide mode by simply choosing a proper direction θ_3 for the incoming laser beam.

When the laser beam is in a synchronous direction, the waves in the prism and in the film have the same horizontal wave motion. The fields at the two opposite sides of the air gap are in phase at every point along x. As shown in Fig. 9, the field in the waveguide mode has an exponential tail extending upward above the film. The evanescent field of the prism is an exponential extending downward below the prism. These two exponential tails overlap in the air gap. The parts of the fields that overlap are common to the prism and the film and constitute the coupling between them.

Let a_3 and b_3 be the field amplitudes of the incoming and reflected waves in the prism and let a_1 and b_1 be the field amplitudes of the zigzag waves in the waveguide mode of the film (Fig. 10). The a_1 and b_1 waves are represented by the wave vectors A_1 and B_1 in the

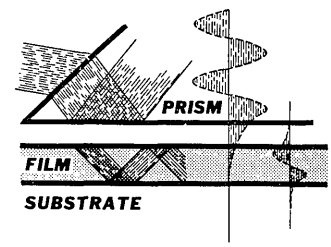

Fig. 9. In a prism–film coupler, the light wave from a laser is totally reflected at the prism base. The field distributions in the prism and in the film show that their evanescent fields overlap each other in the gap region.

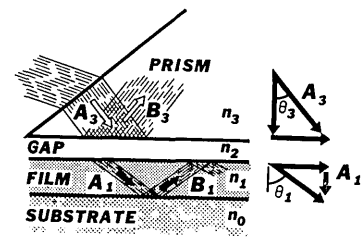

Fig. 10. In order that a waveguide mode be excited in a film by a prism–film coupler, the horizontal component of the wave vector for the wave in the prism must be equal to that of the wave in the film. The prism–film coupler, therefore, excites a single waveguide mode only, and by changing the direction of the incoming light wave, any waveguide mode can be excited.

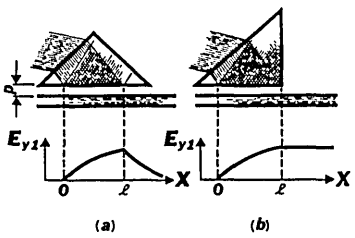

Fig. 11. (a) The light energy transferred from the prism to the film in the region $0 < X < l$ is returned to the prism in the region $X > l$; the net energy retained in the film is therefore zero. (b) By using a right-angle prism, the coupling between the prism and the film discontinues in the region $X > l$. The light wave coupled into the film in the region $0 < X < l$ is therefore retained in the film and continues to propagate in the film.

earlier discussion, and a_1 may be considered as the reflection of the b_1 wave, or vice versa, so that $|a_1| = |b_1|$ at any x. Let all the wave amplitudes be normalized such that $a_j a_j^*$ or $b_j b_j^*$ is the Poynting vector in the direction normal to the film, where $j = 1$ or 3. Because of the coupling described above, the energy is continuously transferred from the prism to the film along the coupling length which starts from $x = 0$. Since the Maxwell equations are linear in field amplitudes, we expect that a_1 (or b_1) increases in x according to a_3, or that da_1/dx should be linearly proportional to a_3. On the other hand, as soon as the wave energy in the film builds up, it continuously leaks into the prism, since the energy transfer is possible in both ways between the prism and the film. We ought to expect, then, that da_1/dx is also proportional $(-a_1)$. We have thus,

$$da_1/dx = Ta_3 - Sa_1, \tag{10}$$

where T and S are the coupling constants that depend on the geometrical configuration and the refractive indices of the media. Near $x = 0$, a_1 is small and so is the term Sa_1 in Eq. (10); a_1 increases linearly from $x = 0$ according to Ta_3x. At a large x, a_1 grows to an amplitude so that Sa_1 approaches a value that nearly cancels the term Ta_3 in Eq. (10); $da_1/dx = 0$ and a_1 reaches a saturation. The wave amplitude in the film cannot therefore increase indefinitely by simply increasing the coupling length.

In Fig. 11(a), we assume that a_3 is uniformly distributed between $x = 0$ and $x = l$, which are the left and right edges of the laser beam incident on the prism base. The amplitude a_1 increases in x until the point $x = l$. Beyond that point, $a_3 = 0$ in Eq. (10) and the equation indicates that a_1 should decrease exponentially to zero according to $\exp[-Sx]$. All the energy fed into the film between $x = 0$ and l is returned to the prism at $x > l$ and therefore the net energy transfer from the prism to the film is zero. If the film is not perfect and scatters the light, a more complex phenomenon occurs. Since the incident laser beam is in the synchronous direction, the light energy is coupled into one of the waveguide modes of the film. However, the energy in the original waveguide mode can be rapidly scattered into other waveguide modes before it is coupled back to the prism. The returned light wave in the prism therefore consists of many waveguide modes; each of them appears in its own synchronous direction. We thus observe a series of bright lines at the right side of the prism. They are called the m lines.[3] For good film, the m lines are thin and weak.

In Fig. 11(b), we use a rectangular prism. Here the rectangular corner of the prism is placed at $x = l$. Contrary to the earlier case, here the coupling between the prism and the film no longer exists beyond $x = l$. The wave energy which is fed into the film between $x = 0$ and l is retained in the film as the wave continues to propagate beyond $x = l$. Therefore, for coupling light energy into or out of a film, we always use a rectangular prism and place the right edge of the laser beam as close as possible to the rectangular corner of the prism.

We notice that because of their directions, the a_3 wave in the prism is coupled only with the a_1 wave in the film and, similarly, the b_1 wave is coupled only to the b_3 wave. Of course, a_1 and b_1 must increase and decrease together in x, since each of them is the reflection of the other. To calculate the coupling efficiency, we consider first the prism-film coupler in Fig. 12, which is used as an output coupler, to couple light energy out of the film. There is a light wave propagating in the film, which is represented by the a_1 and b_1 waves in the film. There is no input laser beam and thus $a_3 = 0$ everywhere. Beyond $x = 0$, the film is coupled to the prism; consequently, the amplitude of the b_1 (or a_1) wave decreases as light energy in the film leaks into the prism. By replacing a_1 by b_1 and putting $a_3 = 0$ in Eq. (10), we have

$$\begin{aligned} b_1(x) &= b_1(0)\exp[-Sx], \quad x > 0, \\ &= b_1(0) \qquad\quad , \quad x < 0. \end{aligned} \tag{11}$$

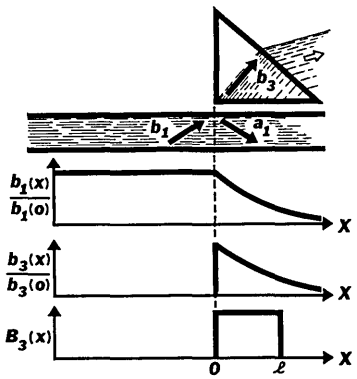

Fig. 12. The figure shows that the prism–film coupler is a perfect output coupler. All the light energy in the film is coupled out and appears as the b_3 wave in the prism. $b_1(x)$ and $b_3(x)$ show respectively how the filed amplitudes in the film and in the prism vary along the coupling length. $B_3(x)$ is the distribution of an incoming laser beam, which is used to demonstrate how to calculate the coupling efficiency.

It is easy to see that the total power flow in the film, which is proportional to $b_1(x)b_1{}^*(x)$, must decrease as $\exp[-2Sx]$. The power lost in the film in a small distance dx is then proportional to $\partial[b_1(x)b_1{}^*(x)]/\partial x$ and it must reappear in the prism as the b_3 wave. We have thus immediately

$$b_3(x) = b_3(0)\exp[-Sx], \quad x > 0$$
$$= 0 \quad , \quad x < 0. \tag{12}$$

Both $b_1(x)$ and $b_3(x)$ are plotted in Fig. 12. We see from Eq. (11) that eventually all the power in the film will be transferred into the prism. An output coupler is always at perfect output coupler, provided that the coupling length is sufficiently long. Now the reciprocity of the linear optics indicates that we can reverse the process. Consequently, if we apply a laser beam in the prism in the direction opposite to that of the b_3 wave discussed above, and if it has an amplitude distribution exactly as $b_3(x)$ in Fig. 12, all the applied laser energy should then enter into the film as to be expected in a perfect input coupler. On the other hand, if we apply a laser beam which is uniform over the cross section as shown in $B_3(x)$ in Fig. 12, we expect that a part of $B_3(x)$ which matches $b_3(x)$ is accepted into the film and the rest is reflected at the prism base. We can therefore define an overlap integral[11]

$$\eta = \frac{\left[\int_{-x}^{+x} B_3(x)b_3{}^*(x)dx\right]^2}{\int_{-x}^{+x} B_3(x)B_3{}^*(x)dx \int_{-x}^{+x} b_3(x)b_3{}^*(x)dx}, \tag{13}$$

which specifies the correlation between the amplitude distribution of the input laser beam and that required for a perfect coupler. In fact, η in Eq. (13) is the coupling efficiency. Since here $B_3(x)$ is constant between $x = 0$ and l, and $b_3(x)$ is exponential beyond $x = 0$, the integral in Eq. (13) can easily be performed. We have

$$\eta = 2/Sl[1 - e^{-Sl}]^2, \tag{14}$$

which is identical to the formula given in the earlier papers.[6,7] By maximizing the expression (14) with respect to Sl, we find the optimum coupling length $Sl = 1.25$, and the optimum coupling efficiency $\eta = 81\%$. We notice that all the properties of the coupler depend on the parameter S which can be computed from the geometrical configuration and the refractive indices of the coupler, or it can be determined experimentally. A detailed calculation[6] shows

$$S = e^{-2p_2 d} \sin 2\Phi_{12} \sin 2\Phi_{32}/(W_{\text{eff}} \tan\theta_1), \tag{15}$$

where $p_2 = [\beta^2 - (kn_2)^2]^{\frac{1}{2}}$; d is the spacing of the air gap between the prism and the film, W_{eff} is the effective thickness of the film, and Φ_{32} may be obtained from Eq. (4) or (5) by replacing the subscript 1 by 3. The effective thickness W_{eff} will be discussed later in Sec. VII. The use of W_{eff} instead of the actual thickness of the film W is due to the Goos-Haenchen shifts.[24] The coupling efficiency does not differ significantly if a gaussian input field distribution is used.

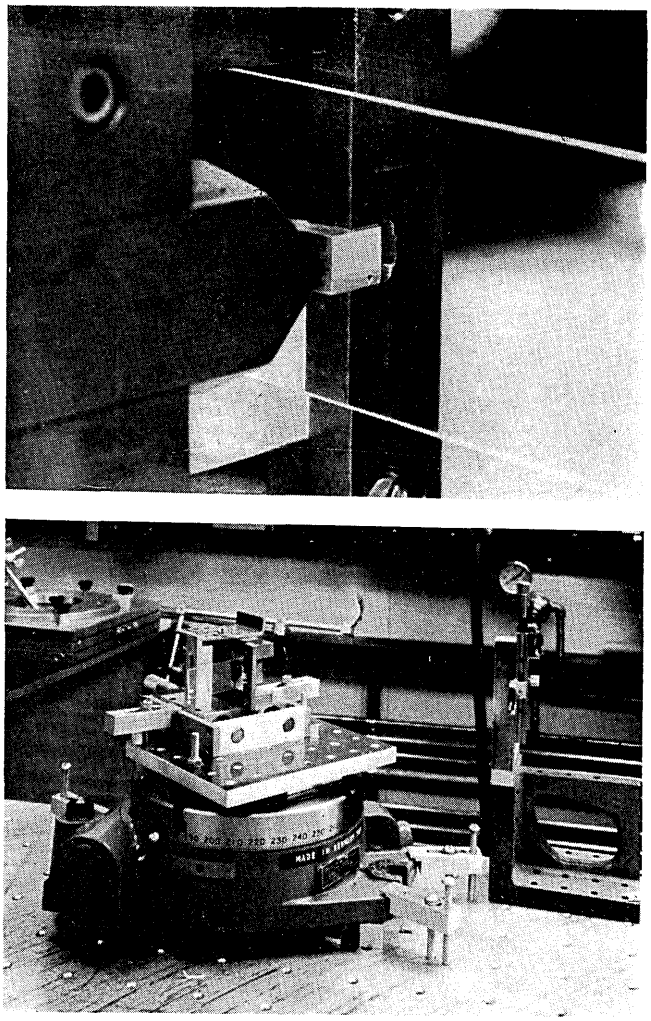

Fig. 13. The top photograph shows how a thin film coated on the glass substrate is pressed against the base of the prism in a prism–film coupler. The lower photograph shows that the entire prism–film assemblage is mounted on a turntable so that the incident laser beam can enter into the prism at any angle.

From the above discussion, we realize that when the prism–film coupler is used as an output coupler, it can easily be made 100% efficient. The light which is not coupled out, say up to the point $x = x_a$, remains in the film. It thus can always be coupled out at $x > x_a$, provided that the coupling length is sufficiently long. In contrast, an input coupler is 100% efficient only when the input light is properly distributed along the coupling gap, since the uncoupled light is immediately lost upon being reflected at the prism base.

The field distribution of a perfect coupler, $b_3(x)$, in Fig. 12 tends to be more uniform and thus matches better the uniform input field distribution, if we can decrease the coupling strength at $x = 0$ and increase it at $x = l$ (refer to Fig. 12) by properly varying the gap spacing between film and prism. We accomplish this in the following way.

Figure 13 shows a prism–film coupler which is mounted on a turntable so that a laser beam can enter

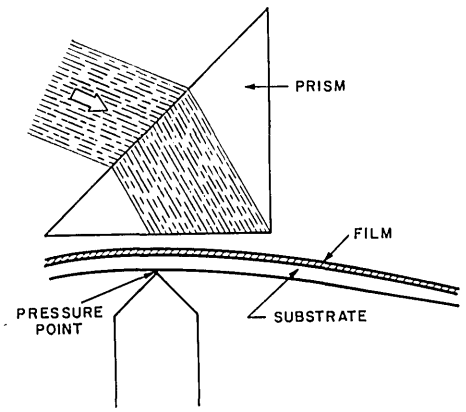

Fig. 14. When the prism–film coupler is used as an input coupler, the pressure is applied at a point about a few tenths of a millimeter from the rectangular corner of the prism. Because of this pressure, the glass substrate bends slightly so that the gap between the prism base and the film is the smallest at the pressure point. To achieve a high coupling efficiency, the incident laser beam should fill the region between the pressure point and the rectangular corner of the prism.

the prism at any angle. The film is coated on a glass slide and pressed against the base of the prism by a knife edge, while the dust particles between the prism and the film act as the spacers. The pressure point is about 1 mm or less away from the rectangular corner of the prism. The pressure applied to the back side of the glass slide actually bends the slide slightly so that the gap between the prism base and the film is smaller at the pressure point and larger at the corner of the prism, as shown in Fig. 14. This provides a stronger coupling at $x = l$ and weaker coupling at $x = 0$ (Fig. 12) and between these two points the input laser beam is located. We thus approach the ideal condition mentioned earlier. In practice, it is easy to obtain a coupling efficiency of about 60%. Beyond that percentage, one needs a prism having a sharp rectangular corner, since any imperfection in the corner would radiate light energy and thus limit the attainable coupling efficiency. Figure 15 shows a laser beam that enters an organic film through the prism–film coupler at the right. The light beam propagated through the film and then was taken out of the film by another prism–film coupler at the left.

The prism–film coupler serves many useful functions in an experiment. By correlating the measured values of the synchronous directions with a theoretical calculation on the waveguide modes, one can independently determine the refractive index and the thickness of the film.[3] The prism–film coupler has been extensively used to determine the refractive index of organosilicon films.[15] The accuracy obtained is within 1 part in 1000 for the refractive index and 1% for the thickness. We often use a rutile prism for semiconductor films and a glass prism for organic and glass films.

Figure 16 shows a grating coupler.[4,5] A phase grating made of photoresist or dichromated gelatin is fabricated by holographic technique on a thin film. A laser beam incident on the phase grating at an angle θ has a phase variation in the x direction according to $\exp[i(2\pi/\lambda_0)(\sin\theta)x]$, where λ_0 is the vacuum laser wavelength. As the beam passes through the grating, it obtains an additional spatial phase modulation $\Delta\Phi \sin(2\pi x/d)$, where $\Delta\Phi$ is the amplitude of the spatial phase modulation caused by the grating and is sometimes called the phase depth of the grating; d is the periodicity of the grating. The light wave reaching to the top surface of the film contains many fourier components, $\exp\{i(2\pi/\lambda_0)[\sin\theta + m(2\pi/d)]x\}$, where m is an integer. If one of these components matches the wave motion of one of the waveguide modes of the film, the light beam is exclusively coupled to this mode and the light energy is fed into the film. We can analyze the grating coupler by the same method used previously for the prism–film coupler. The only difference is the calculation of the parameter S which, in practice, is determined experimentally anyway.

The tapered-film coupler[12] is operated on an entirely different principle. It utilizes the cutoff property of an asymmetric waveguide. We remember in the discussion of the waveguide modes that the thickness of the film W can be divided into two parts W_{12} and W_{10}

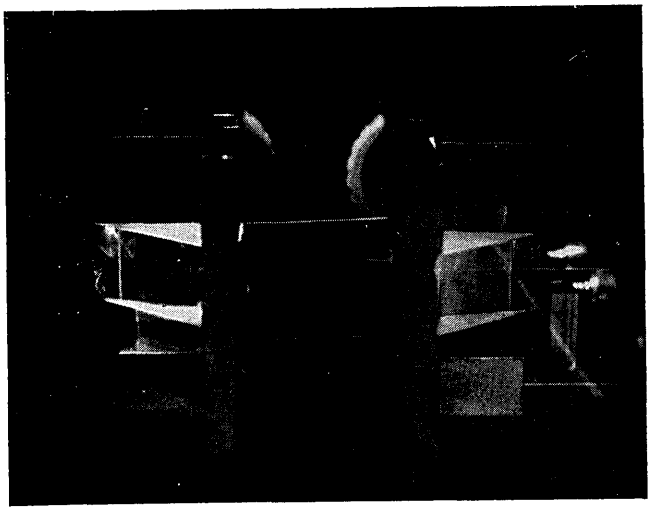

Fig. 15. The photograph shows a light beam which is fed into an organic film by the prism–film coupler at the right. The light wave propagates inside the film and is then coupled out of the film at the left by another prism-film coupler. (This figure is reproduced in color on the cover of the November 1971 *Applied Optics* issue.)

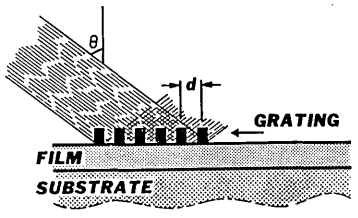

Fig. 16. A grating light-wave coupler.

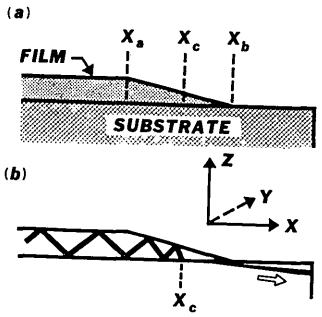

Fig. 17. (a) A tapered film light-wave coupler is simply a tapered film edge deposited on a substrate. The figure shows that the film is tapered to nothing between X_a and X_c. (b) As a light wave originally propagating inside the film enters into the tapered region of the film, the angle between the zigzag light path and the vertical Z axis becomes smaller and smaller. At the cut-off point, $X = X_c$, the angle is smaller than the critical angle of the film–substrate interface and the light wave is refracted into the substrate. The tapered film edge then serves as an output light-wave coupler.

and that they are proportional to Φ_{12} and $(\Phi_{10} + m\pi)$, respectively. As β varies from kn_1 to kn_0, both Φ_{12} and Φ_{10} decrease. At the cutoff point of the waveguide modes, $\beta = kn_0$, $\Phi_{10} = 0$, and W is minimum. Therefore, there is a minimum thickness of the film, less than which a waveguide mode of the order m cannot propagate. Now consider, in Fig. 17(a), a film that is deposited on a substrate and is tapered to nothing in a distance between $x = x_a$ and $x = x_b$, typically in the order of 10 to 100 vacuum wavelengths. The film is in the X–Y plane and the tapered edge is parallel to the Y axis. Consider a light wave in a waveguide mode of the film propagating toward the tapered edge in the direction normal to it. Since the thickness of the film decreases continuously in the tapered region, the waveguide mode is cut off at $x = x_c$. A detailed calculation shows that within a distance of about eight vacuum wavelengths in the vicinity of x_c, the waveguide mode is gradually converted into the substrate modes and the light wave reappears in the substrate as the radiation field. The far-field pattern of the radiation field shows that more than 80% of its energy is concentrated within an angle of 15° below the film–substrate interface. We can understand the problem better by considering again the ray optics. Figure 17(b) shows a light wave that propagates in a zigzag path from the left side of the film toward the tapered film edge. As it enters into the tapered region, the angle between the light path and the Z axis becomes smaller and smaller, and eventually, near $x = x_c$, the angle becomes smaller than the critical angle of the film–substrate interface. The light beam is then refracted into the substrate. In the experiments,[12] it is very easy to couple all the light energy out of a film through the tapered film edge. By reversing the process, we can also feed a light beam into the film by focusing it on the tapered edge through the substrate. The tapered-film coupler is simply the film itself and is particularly useful to the study of the semiconductor epitax layers. Here the refractive index of the film such as the GaAs layer is so large that it becomes difficult to find a prism of a higher refractive index that is also transparent to the radiation used in the experiment.

V. Materials and Losses

Sputtered ZnO films were first used in the light-guiding experiments.[3] They were deposited on a heated glass substrate in an argon–oxygen atmosphere. The method of deposition was developed previously by Foster *et al.*[25,26] for ultrasonic transducers. The films discussed below were grown at a substrate temperature of 400°C. They have very high resistivities. After heat treatment in vacuum, nitrogen, or hydrogen, their resistivities can be reduced from 10^6 to 10^{-2} Ω-cm. The largest mobility measured is 40 cm²/V-sec as compared with that of bulk 200 cm²/V-sec. The films have the hexagonal or Wurzite crystal structure with the c axis normal to the surface of the film and the (0002) planes parallel to it. As determined from x-ray diffraction by using a Debye-Scherrer camera, the c axis of the crystallites is oriented within 5° from the normal of the film. The refractive index of the film is 1.973 ± 0.001 at 6328 Å of the helium–neon laser wavelength as compared with the value 1.988 of bulk ZnO.

It was indeed a surprise when we found that those seemingly perfect films had a loss of more than 60 dB/cm in the light propagation experiment.[3] The excessive loss in the film was not understood until we took electron micrographs of the films. Figure 18(a) shows

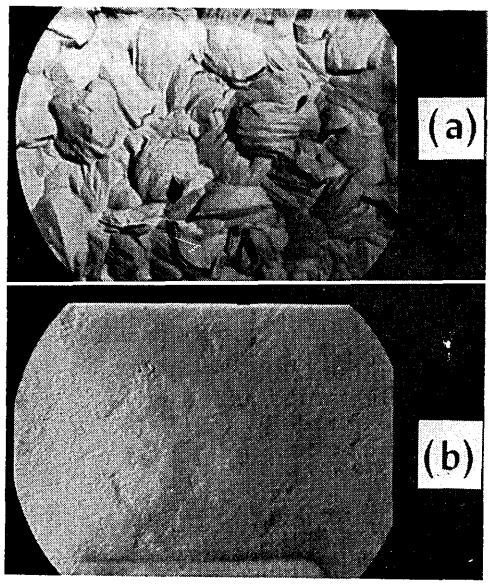

Fig. 18. (a) Electron micrograph of an oriented sputtered ZnO film showing that the sizes of the crystal sites are on the order of 0.5 μm. The entire width of the micrograph is 5 μm. (b) Electron-micrograph showing the surface of the same film after polishing.

the surface profile of a ZnO film 1.5 μ thick. This electron micrograph was taken from a platinum-shadowed carbon replica. It can be seen that the average grain size is on the order of 0.5 μ. This grain size is comparable to the optical wavelength used in the experiment and thus causes excessive scattering to the propagating light wave in the film. The film surface (Fig. 18a) has an irregular profile; the peak-to-peak surface roughness is estimated on the order of 1000 Å. It is interesting to note that it was possible to polish the surface of the film by lapping it with chromium oxide dispersed in water. Figure 18(b) shows the surface of the ZnO film after being lapped. We estimated that the residual roughness was reduced to less than 100 Å. After being polished, the films still have a loss of more than 20 dB/cm.

To avoid scattering caused by large crystals, we started to evaporate ZnS films on glass substrates which were held at room temperature. They are polycrystal films with very small crystal sites. Although we have significantly reduced the scattering loss, unfortunately additional loss was found due to the absorption arising from the long tail of the fundamental band gap. The refractive indices of the film are 2.404, 2.342, and 2.289 at the wavelengths 5322.5 Å, 6328 Å, and 1.0645 Å, respectively. The loss measured at 6328-Å wavelength is on the order of 5 dB/cm.

Now, trying to avoid both the scattering and absorption losses, we chose Ta_2O_5, which has a larger energy gap, 4.6 eV. The Ta_2O_5 films[14] were prepared by first sputtering high purity tantalum in an argon atmosphere. The β-tantalum thus deposited was then heated in pure oxygen at 500°C until completely converted to Ta_2O_5. Debye-Scherrer x-ray diffraction patterns show the films to be amorphous. In spite of many visible defects, the same conclusion can be drawn from the electron micrograph in Fig. 19. The refractive indices of the film are 2.2136, 2.2423, and 2.2767 at the wavelengths 6328 Å, 5145 Å, and 4880 Å, respectively. At the above wavelengths, the losses measured are 0.9 dB/cm for the red, 2.5 dB/cm for the green, and 4.1 dB/cm for the blue.

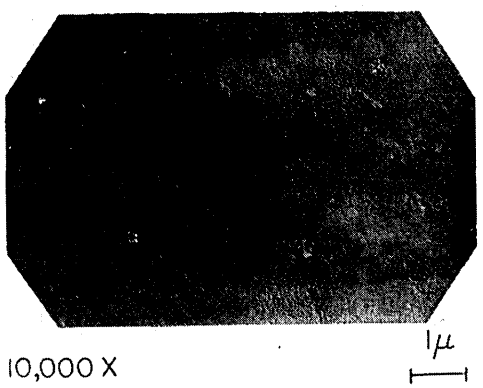

Fig. 19. Electron micrograph of a Ta_2O_5 film showing that the film is generally amorphous, though a number of small crystals visible in the micrograph still exist.

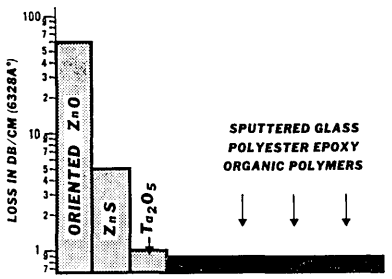

Fig. 20. Losses in decibels per centimeter measured at 6328-Å light wavelength for several semiconductor and organic films.

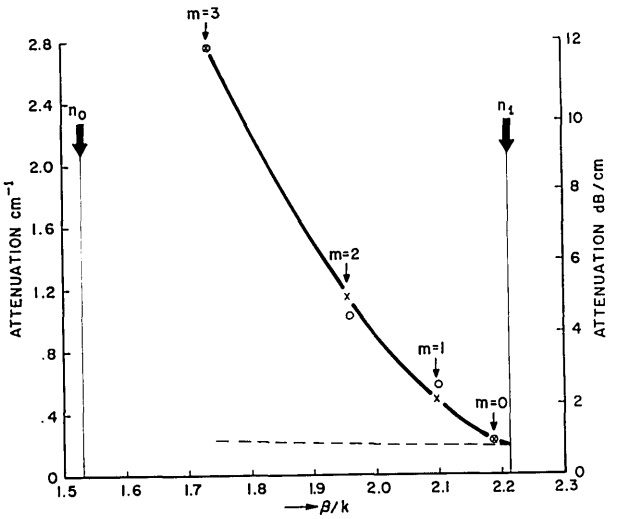

Fig. 21. Measurement (*circles*) and calculation (*crosses*) of the losses in a Ta_2O_5 film at 6328-Å light wavelength. The figure shows that the loss in db/cm or attenuation in cm^{-1} for the $m = 3$ waveguide mode is as much as 14 times that of the $m = 0$ waveguide mode. The dashed line is the volume loss in the film and the vertical distance between the dashed line and the solid curve is the surface scattering.

As shown in Fig. 20, all the low-loss films presently used in the light-guide experiments are amorphous. This includes sputtered glass,[13] polyurethane and polyester epoxy,[9] and organic polymer[15] films. All of these have a loss less than 1 dB/cm. To the author's knowledge, large transparent single-crystal semiconductor films other than epitax layers are not available at the present, although they are essential to the development of thin-film devices.

VI. Simple Theory of the Surface Scattering

The losses of the films quoted above are the losses of the $m = 0$ waveguide mode. In fact, the loss increases rapidly in the higher order modes. The circles in Fig. 21 are the measurement made in a Ta_2O_5 film by Hensler *et al.*[14] for a TE wave at 6328-Å light wavelength. The data show that the loss expressed in decibels per cm (or intensity attenuation in cm^{-1}) for the $m = 3$ mode is as much as fourteen times that of the $m = 0$ mode. In this figure, the losses of different modes are plotted vs β/k, where β is the hori-

zontal component of the wave vector A_1 or B_1 as discussed at the end of Sec. II. It was pointed out earlier that the waveguide modes occur between $\beta/k = n_0$ and n_1, and that the value of β/k associated with each mode depends on the thickness of the film and the refractive indices of the film and substrate. It is evident in Fig. 21 that the loss for the $m = 0$ mode depends on the value of β/k; it represents neither volume loss of the material nor scattering loss caused by surface roughness. Therefore, it seems that there must be a better way to define the loss in a film other than that of a particular waveguide mode.

The losses in Fig. 21 were measured by a system of lenses, filter, slit, and detector which collected the light scattered from a small area of the film into a detector through an adjustable slit. By moving the detector system away from the input coupler but keeping it along a path parallel to the light beam in the film, we measured a decrease of scattered light along the light path, which should represent the intensity attenuation of the light wave propagating inside the film. The losses measured therefore included the volume absorption and scattering as well as the surface scattering. The same method of measurement was used by Goell and Standley[13] for their glass films.

Another method which was used extensively to measure losses in organosilicon films is the transmission measurement.[15] It involved two prism–film couplers in an arrangement similar to that shown in Fig. 15. One prism–film coupler excited a light streak in the film and a second coupler, several centimeters from the first, coupled the light wave out of the film. The efficiency of the input coupler was not measured, but it remained intact during the entire experiment. We accomplished this by monitoring the light scattered from a small section of the streak near the input end so that any change of the input conditions could be detected and corrected. The efficiency of the output coupler was always adjusted to be 100%. We mentioned earlier in Sec. IV that the output coupler can easily be made 100% efficient. In our experiment, the output coupler was applied at different points along the light streak. At each point the coupling was adjusted until the streak disappeared completely beyond the coupling point. The light emerging from the output prism was then detected. The measurements thus obtained at different points along the streak were used to evaluate the loss of the film.

It is possible to estimate the loss of a film based on the sensitivity of the eye. The sensitivity of the eye covers a range of about 27 dB. Thus if the length of the light streak as observed by the naked eye is x cm, the loss should be $27/x$ dB/cm.

The surface scattering of a symmetric slab waveguide has been calculated by Marcuse[21] based on the radiation modes. His results after certain approximations and evaluated to the limit of long correlation length agree with the simple theory that we will develop below. For an asymmetric waveguide ($n_2 \neq n_0$) which is considered in this paper, the two surfaces of the film scatter differently. Since it is practically impossible to measure the scattering losses of the two surfaces separately, we hope to develop a crude theory in which we can lump all the surface properties of the film into a single parameter. We wish further to use this parameter to calculate the losses of different waveguide modes and compare them with the measurement. Of course, the theory that we will discuss is very crude. However, it establishes a guideline by which we can gain insight into this complex problem.

Returning to Fig. 21, first, we must separate the volume loss from surface scattering. As β/k approaches n_1, the fields at two film surfaces vanish. We should then expect surface scattering to vanish and the residue loss at $\beta/k = n_1$ to be volume loss only. At the other values of β/k, the volume loss should be proportional to the length of the zigzag path. We showed earlier that the light wave, in the waveguide mode, is a plane wave which propagates along a zigzag path and a zigzag path is considerably longer than the actual length of the film. We have ignored here, for simplification, the fields extending outside the film. According to the above argument, the volume loss is proportional to $(\sin\theta_1)^{-1}$ or $(\beta/kn_1)^{-1}$. It is plotted as the dashed line in Fig. 21. We find that the volume loss of the $m = 3$ mode is merely 30% larger than that of the $m = 0$ mode as compared with a factor of 14 in the total losses. Consequently, almost all the losses in the higher order modes and the large variation of the losses among different waveguide modes are due to surface scattering only.

To develop a crude theory for surface scattering, we resort to the hundred-year-old Rayleigh criterion.[27] Figure 22 shows a plane wave incident on the upper surface of the film. To cover a unit length of the film in the x direction, the plane wave has a width $\cos\theta_1$ in the direction parallel to the wavefront. Considering a TE wave, the power carried by the incident beam is $(c/8\pi)n_1E_y^2\cos\theta_1$ in gaussian units, where E_y is the field amplitude. Again, we are taking $\partial/\partial y = 0$ and considering a space of unit length in the y direction. According to the Rayleigh criterion, the specularly reflected beam from the upper film surface has a power

$$\frac{c}{8\pi}n_1E_y^2\cos\theta_1\exp\left[-\left(\frac{4\pi\sigma_{12}}{\lambda_1}\cos\theta_1\right)^2\right]. \qquad (16)$$

We use the double subscripts 12 and 10 to denote the film–air and film–substance interfaces, respectively. Note that λ_1 is the wavelength in the film. The surface

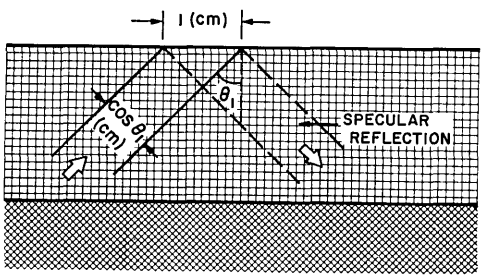

Fig. 22. Plane wave specularly reflected at the top film surface.

scattering is usually characterized by two statistical quantities: the statistical variation of the surface about the mean and the correlation length of the surface variation. Here σ_{12} in Eq. (16) is the variance of the surface roughness. A recent calculation by Marcuse[21] shows that the Rayleigh criterion applies only to the case of long correlation length. Since scattering observed in our loss measurements was always dominated by forward scattering, the assumption of the long correlation length may be correct. The limitation of the expression $\exp[-(4\pi\sigma_{12}\cos\theta_1/\lambda_1)^2]$ is also discussed by Beckmann and Spizzichino.[27] In spite of its shortcomings, the expression is widely used because of its simplicity. The power lost by surface scattering at two surfaces of the film is therefore

$$(c/8\pi)n_1E_y\cos\theta_1\{1-\exp[-K^2(\cos\theta_1)^2]\} \cong (c/8\pi)n_1E_y^2K^2\cos^3\theta_1, \quad (17)$$

where we have assumed that loss per unit length of the film is small, and

$$K = (4\pi/\lambda_1)(\sigma_{12}^2+\sigma_{10}^2)^{\frac{1}{2}}. \quad (18)$$

It can be easily shown from the field distribution discussed in Sec. III that the total power flow in the film for any waveguide mode is

$$(c/4\pi)n_1E_y^2\sin\theta_1[W+(1/p_{10})+(1/p_{12})], \quad (19)$$

where W is the thickness of the film, and p_{13} and p_{12} have been defined earlier. Dividing Eq. (17) by Eq. (19), we obtain the power attenuation per unit length of the film

$$\text{Attenuation} = K^2\left(\tfrac{1}{2}\frac{\cos^3\theta_1}{\sin\theta_1}\right)\left\{\frac{1}{[W+(1/p_{10})+(1/p_{12})]}\right\}. \quad (20)$$

We can also express the attenuation in decibels per unit length after multiplying Eq. (20) by 4.343. It is the loss caused by surface scattering only. We have thus expressed the loss as a product of three independent factors. The first factor, K, depends solely on the surface properties of the film and is a dimensionless quantity. The second factor involves θ_1 only and thus depends on the waveguide mode considered. The third factor is the reciprocal of the effective film thickness and shows explicitly that the loss is inversely proportional to the thickness, as is to be expected. From this theory, it appears that the single parameter K defines all the surface properties of the film. It is a dimensionless parameter and quantitatively it compares surface roughness with the optical wavelength.

The crosses in Fig. 21 are the results calculated from Eq. (20). The parameter $K = 1.27\times10^{-2}$ is evaluated so that the loss of the $m = 3$ mode computed by the theory and that obtained by the measurement coincide. For this Ta_2O_5 film, the agreement between the theory and measurement is excellent. For some other films, the agreement is only moderate. The details will be discussed elsewhere.

The large scattering loss observed in the experiment, particularly that of the higher order waveguide modes, is to be expected. Again, we return to the picture of the zigzag waves. For the Ta_2O_5 film considered in Fig. 21, the light wave is reflected back and forth between the two film surfaces about 2000 times in a length of 1 cm of the film for the $m = 0$ mode and about 10,000 times for the $m = 3$ mode. Thus, for this film to have a scattering loss of 1 dB/cm in the $m = 0$ mode, the loss per reflection should be 1.5×10^{-4}, which should be compared with the loss per reflection as large as 1×10^{-3} for the better dielectric mirrors used in lasers.

VII. Field Concentration and Mode Characteristics

Before discussing nonlinear interactions in thin films, it may be necessary to learn more about mode characteristics. Also included in this section are the discussion on the cutoff of the waveguide mode, field concentration in a film, and finally the concept of the effective film thickness. Throughout this section, we will use a ZnS film on a glass substrate as the example. The light wavelength is 1.06 μm of the YAG:Nd laser. At this wavelength, $n_1 = 2.2899$ and $n_0 = 1.5040$.

We have shown at the end of Sec. II how to calculate W and β/k for a given n_0, n_1 n_2, and m. The result of the calculation for both TE and TM waves is shown in Fig. 23. These W vs β/k curves are the mode characteristics of the waveguide. The ratio β/k is called the effective refractive index, since it measures the ratio of the speed of light in vacuum to that in the waveguide in the same way as the ordinary refractive index measures the ratio of the speed of light in vacuum to that in a dense medium. First, we notice β/k ranging from n_0 to n_1 for the waveguide modes. Let us concentrate on the $m = 0$ TE mode in Fig. 23. When W is large, the effective index (β/k) approaches the refractive index of the film n_1. To this limit, the film acts as a bulk medium and all the light energy is contained within the film. The fields therefore vanish at the two film surfaces. When β/k varies from n_1 to n_0, W decreases continuously as the fields extend more

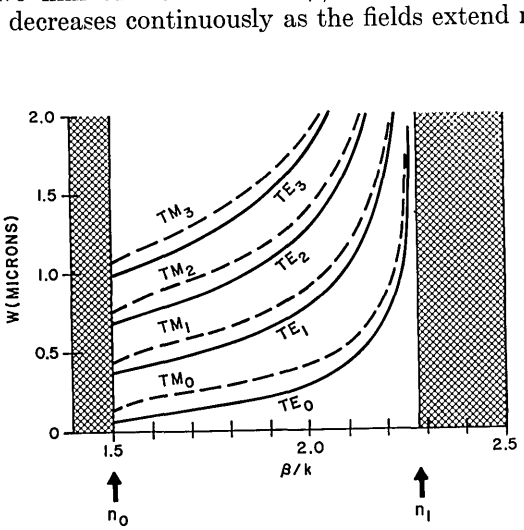

Fig. 23. Thickness W of a ZnS film deposited on a glass substrate is plotted vs the ratio of (β/k) for both TE and TM waveguide modes. Here, the ratio β/k can be considered as the effective refractive index. The left shaded region is for $\beta/k < n_0$, the right shaded region is for $\beta/k > n_1$, and in the space between the two shaded regions the waveguide modes are possible.

and more outside the film. At $\beta/k \to n_0$, the mode becomes cut off and W is the minimum thickness that can support this waveguide mode. At the cutoff, the waveguide mode is turned into a substrate mode as the fields extend infinitely into the substrate.

Next, we notice that for the same β/k, W's of different TE or TM modes are equally spaced. That is, the difference between W(TE, $m=1$) and W(TE, $m=0$) is equal to the difference between W(TE, $m=2$) and W(TE, $m=1$), and so on. We also notice that for a given m and β/k, W(TM) is always larger than W(TE) simply because the TM wave has larger Φ's [see Eqs. (4) and (5)]. The Φ's of the TM wave increase with the ratio n_1/n_0.

In designing an experiment particularly for nonlinear interactions, we often want to calculate the wave velocity accurately. This is difficult in practice. The film may not be homogeneous, and the materials may change with the environment. We do not really know the rafractive indices exactly. Moreover, the film may not be uniform and there are difficulties in measuring the film thickness exactly. It is then important to know the wave velocity, or β/k, varies with small increments in W, n_0, n_1 and n_2. For this purpose, we have calculated $[dW/d(\beta/k)]$, $[d(\beta/k)/dn_1]$, $[d(\beta/k)/dn_0]$, and $[d(\beta/k)/dn_2]$. They are given below for the TE wave only.

$$\frac{dW}{d(\beta/k)} = \frac{\beta k}{b_1^2}\left(W + \frac{1}{p_0} + \frac{1}{p_2}\right); \tag{21}$$

$$\frac{d(\beta/k)}{dn_1} = \frac{kn_1}{\beta}\frac{W + [p_{10}/(p_{10}^2 + b_1^2)] + [p_{12}/(p_{12}^2 + b_1^2)]}{W + \frac{1}{p_{10}} + \frac{1}{p_{12}}}; \tag{22}$$

$$\frac{d(\beta/k)}{dn_j} = \frac{kn_j}{\beta}\cdot\frac{b_1^2/[p_j(p_j^2 + b_1^2)]}{W + \frac{1}{p_0} + \frac{1}{p_2}}, \tag{23}$$

where $j = 0$ or 2. We have calculated the above expressions for the TE $m = 0$ waveguide mode. The results are shown in Fig. 24. In Fig. 24(a), we notice that $dW/d(\beta/k)$ is large for β/k near n_1 and also n_0. A large $dW/d(\beta/k)$ means that the wave velocity is not sensitive to small variations in W. In Fig. 24(b), we find that n_1 does not affect β/k very much for $\beta/k \to n_0$, and similarly n_0 does not affect β/k very much for $\beta/k \to n_1$, as is to be expected. The influence of n_2 on β/k is generally 10 times smaller than that exercised by n_1 or n_0. In spite of its smallness, it is relatively easy to vary n_2 in order to obtain a fine adjustment in the wave velocity, for example, by using a liquid of index of refraction to replace the air space on top of the film.

One can easily calculate the power carried by a TE waveguide mode based on field distribution discussed in Sec. III. We find

$$P = \frac{c}{4\pi} n_1 \sin\theta_1 E_y^2 \left(W + \frac{1}{p_0} + \frac{1}{p_2}\right) d,$$

where E_y is the field amplitude of the A_1 or B_1 wave in the film and d is the width of the light wave in the y

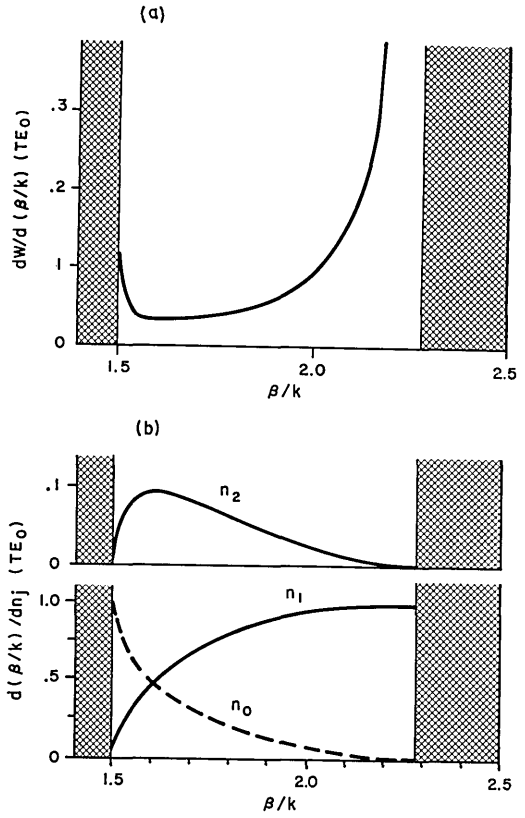

Fig. 24. (a) $dW/d(\beta/k)$ and (b) $d(\beta/k)/dn_j$ vs (β/k) curves for the $m = 0$ TE mode.

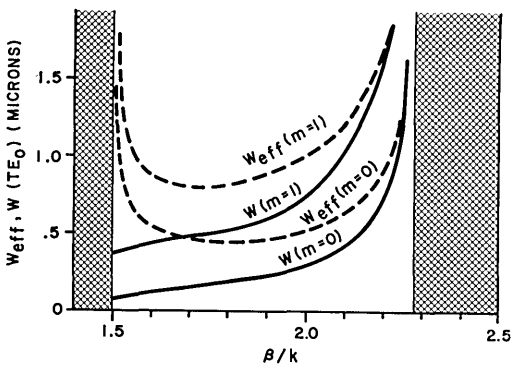

Fig. 25. W_{eff} and W vs (β/k) curves for $m = 0$ and $m = 1$ TE modes.

direction. The equation is written in gaussian units. Because the field extends outside the film, $[W + (1/p_0) + (1/p_2)]$ is the effective thickness of the film, W_{eff}. For a given P, the light intensity E_y^2 is inversely proportional to W_{eff} instead of W. Both W and W_{eff} are plotted in Fig. 25 for the $m = 0$ and m = 1 modes. For large concentration of light intensity inside the film, we prefer the $m = 0$ mode. We see that even for the $m = 0$ mode, we cannot increase light intensity indefinitely by simply reducing thickness of the film. When the film becomes too thin, the fields penetrate deep into the substrate and W_{eff} no longer decreases

with W. It is interesting to note that for the $m = 0$ mode, W_{eff} approaches $\lambda_1/2$ but is never smaller than it, where λ_1 is the optical wavelength in the film medium. The minimum W_{eff} of 0.458 μm occurs at $\beta/k \rightarrow 1.82$, where the average power density and the maximum field amplitude in the film are, respectively, 21.8 MW/cm^2 and 6.78 $\times$ 10^4 V/cm. Here we assume that $d = 10$ μm and that 1 W of the laser power is being fed into the $m = 0$ waveguide mode of the film.

VIII. Phase-Match and Nonlinear Interactions Between Guided Waves

The advantages of performing nonlinear optics in a thin film are many. A thin film can concentrate laser energy for a long distance, whereas a focused gaussian beam diffracts rapidly away from the focused point. The phase velocity of a light wave in a waveguide mode depends on the thickness of the film and the mode of propagation. Thus, for example, by using different waveguide modes for the signal, idler, and pump waves in a parametric oscillator, we can obtain a phase match condition without relying upon the birefringence of the crystal. The crystals such as GaAs, GaP, ZnS, ZnTe, etc., which have large nonlinear coefficients but little birefringences can then be used for nonlinear experiments. The film and the substrate can be immersed in a liquid and the phase-match condition can be varied by varying the refractive index of the liquid. Finally, the nonlinear interaction can take place in the film, in the substrate, or both. All these advantages provide many alternatives to the design of the experiment. Of course, the primary purpose of developing nonlinear devices in thin-film form is that they can be used in integrated optical circuitry. In spite of these advantages, the development of thin-film parametric devices is handicapped by the lack of single-crystal films and by the difficulties in obtaining a long coherence length. We shall illustrate this fully by considering the problem of optical second harmonic generation (SHG).

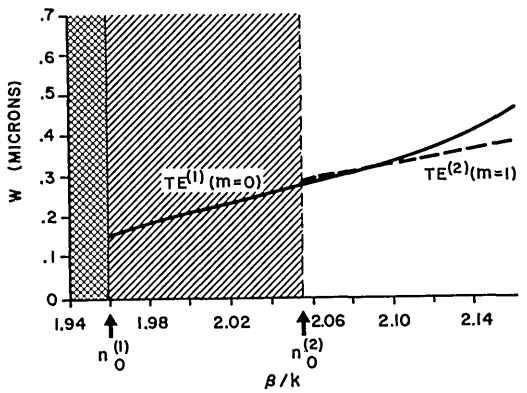

Fig. 26. The phase-match condition is shown as the crossing point of the two W vs (β/k) curves. The solid curve is the $m = 0$ TE mode of the fundamental and the dashed curve is the $m = 1$ TE mode of the harmonic.

In a parametric oscillator, the frequency of the oscillation adjusts itself so that the phase velocities of the signal, idler, and pump are matched. In the SHG, however, we do not have this kind of flexibility, and the film must have exactly the thickness required for the fundamental and the harmonic waves to propagate at the same wave velocity. Any nonhomogeneity in refractive indices and nonuniformity in thickness would reduce the efficiency of the nonlinear interaction. The phase-match condition in nonlinear optics involves only the wave velocities parallel to the film. The phase-match condition of SHG is $(\beta/k)^{(1)} = (\beta/k)^{(2)}$, where the superscripts (1) and (2) denote the fundamental and the harmonic, respectively. Basically, if we plot W vs (β/k) for the fundamental and for the harmonic, the crossing point of the curves is the phase-match condition. The problem becomes simpler if we follow the following simple rules:

1. $dW/d(\beta/k)$ for the fundamental and the harmonic are always positive and the W vs (β/k) curves of the fundamental can cross that of the harmonic only once or not at all.
2. In general, when the W vs (β/k) curves show that $W^{(1)} > W^{(2)}$ near $\beta/k = n_0$, a phase-match condition can be obtained only when $n_1^{(2)} < n_1^{(1)}$. Similarly, when $W^{(1)} < W^{(2)}$ near $\beta/k = n_0$, the phase-match can be obtained only when $n_1^{(2)} > n_1^{(1)}$.

We shall illustrate these rules in the following examples.

As the first example, we consider nonlinear interaction in the substrate. In this case, we use the electric fields that extend into the substrate. The efficiency of the interaction depends on the amount of the fundamental and the harmonic electric fields in the substrate as compared with their distributions over the film, substrate, and air space. In order to obtain a large efficiency, it is obvious that we must operate at β/k near n_0. Since the field distribution in the substrate of any waveguide mode is always exponential, it does not really matter if the fundamental and the harmonic are in the same or different waveguide modes. A good choice is $m = 0$ mode for the fundamental and $m = 1$ mode for the harmonic. In this case, birefringence of the crystal is not required. Both the fundamental and the harmonic can be TE or TM, or one of them TE and the other TM. For example, in Fig. 26, we consider a ZnS film on a single-crystal ZnO substrate. The refractive indices are $n_1^{(1)} = 2.2899$; $n_1^{(2)} = 2.4038$; $n_0^{(1)} = 1.9562$; $n_0^{(2)} = 2.0521$ for the wavelengths 1.064 μm and 0.532 μm, respectively. Here the c axis of ZnO is oriented in the y direction, and the refractive indices quoted are those of the extraordinary ray. To use nonlinear coefficient d_{33} of ZnO, both the fundamental and the harmonic must be TE waves. Phase-match condition is obtained at $(\beta/k) = 2.0877$ where the thickness of the film is 0.314 μm.

A detailed calculation shows that at $(\beta/k) = 2.0877$, $d[(\beta/k)^{(2)} - (\beta/k)^{(1)}]/dW = 0.444$. For a coherence length of 1000 μm, $d[(\beta/k)^{(2)} - (\beta/k)^{(1)}]$ should be equal or less than 2.65 $\times$ 10^{-3}. We find immediately

that the average thickness of the film should be kept within the limits

$$W = 0.314 \pm \frac{2.65 \times 10^{-3}}{0.44} = 0.314 \pm 0.006\ \mu\text{m}.$$

It is not impossible to evaporate a ZnS film within a thickness tolerance of 0.006 μm. Because of the small birefringence, the coherence length in a bulk ZnO is less than 2 μm. Now we have obtained a coherence length of 1000 μm by using the waveguide modes. We have thus improved the coherence length by a factor of 500, which is remarkable. On the other hand, a 1.064-μm light wave can propagate at least a distance of 2.5 cm in ZnS film without suffering appreciable loss. Because of the phase-match problem we have used only a section of 1000 μm of this 2.5-cm-long beam. The advantage of the thin-film waveguide is thus not fully utilized. We have used here the problem of SHG to illustrate the difficulties involved in the integrated optics. We can gain some advantages by using the waveguide principle, but in this and other problems, the full potential of the integrated optics cannot be realized without developing techniques to control the homogeneity and uniformity of the film.

Next, as the second example, we consider nonlinear interaction in film. Although one can use different waveguide modes for the fundamental and the harmonic mismatch between the distribution of nonlinear polarization and that of the fields would reduce the efficiency of interaction sharply. We thus prefer the use of TE and TM waves of the same mode order m for the fundamental and harmonic. A study of the mode characteristics shows that W of the harmonic is always smaller than that of the fundamental at a same value of β/k near n_0. According to our second rule given earlier, we cannot obtain a phase-match condition *unless the refractive index of the film at the harmonic frequency is less than that at the fundamental frequency.* Since most of nonlinear materials have a normal dispersion in the visible spectrum, to satisfy the above condition, we must choose a material of sufficient birefringence for the film. As an example, we consider in Fig. 27 a single-crystal $LiNbO_3$ film on a quartz substrate. The c axis of $LiNbO_3$ is normal to the film. A TE ($m = 0$) wave is used for the fundamental and a TM ($m = 0$) wave for the harmonic. The nonlinear coefficient used is the d_{31} of $LiNbO_3$. The refractive indices involved are than $n_1^{(1)} = 2.365$; $n_0^{(1)} = 1.4614$; $n_1^{(2)} = 2.300$; $n_0^{(2)} = 1.4745$. Again we consider SHG from 1.064 μm to 0.532 μm. We find from Fig. 27 that a phase-match condition is obtained at $(\beta/k) = 2.2274$ where $W = 2.47$ μm. Unfortunately, the single-crystal $LiNbO_3$ film is not available at present. The discussion is therefore academic. We could use the d_{31} nonlinear coefficient of oriented ZnO or CdS films that have been developed for ultrasonic transducers, but they are too lossy, as discussed in Sec. V.

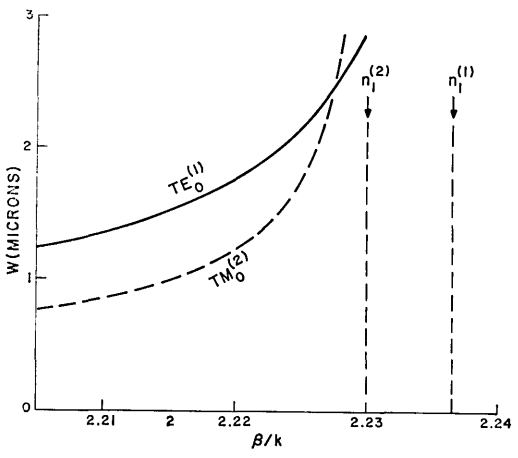

Fig. 27. The phase-match condition for SHG in a $LiNbO_3$ film on a quartz substrate is indicated as the crossing point of the two W vs (β/k) curves. The fundamental at 1.064 μm uses $m = 0$ TE mode and the harmonic at 0.532 μm uses $m = 0$ TM mode.

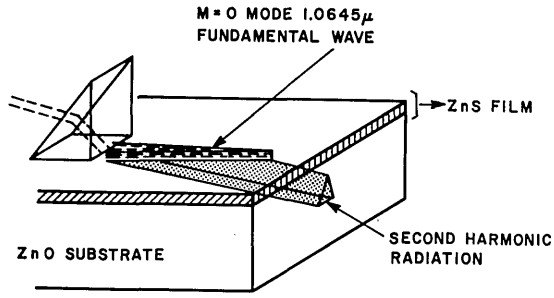

Fig. 28. Experimental arrangement of second harmonic generation in the form of Cerenkov radiation. A fundamental light wave at 1.064 μm is fed into a ZnS polycrystal film which is deposited on a single-crystal ZnO substrate. The evanescent field of the fundamental wave generates a second harmonic Cerenkov radiation in the substrate.

From our earlier discussion on nonlinear interaction in the substrate, we realize that the requirement set by the phase-match condition for the thickness and uniformity of the film is very stringent. To circumvent this problem, a novel method has been used by Tien *et al.*[16] by generating second harmonic in the form of Cerenkov radiation. The nonlinear interaction by their method is not as efficient as that under the phase-match condition, but the interaction extends to the full length of the fundamental wave. As in the case discussed before, a polycrystal ZnS film on a single-crystal ZnO substrated was used and nonlinear interaction took place in the substrate. The c axis of the ZnO was oriented in the y direction, and both the fundamental and the harmonic waves were polarized parallel to the c axis.

The experimental arrangement is shown in Fig. 28. A light beam of a YAG:Nd laser at 1.06 μm was fed into the $m = 0$ waveguide mode of the ZnS film as the fundamental wave. It propagated as $\exp[-i\omega^{(1)} t + i\beta x]$, where $n_0^{(1)} < (\beta/k) < n_1^{(1)}$. The fundamental wave excited a wave of second harmonic nonlinear polarization in the substrate via the d_{33} nonlinear coefficient of ZnO. Here the nonlinear polarization wave was a forced wave and thus it varied as $\exp[-i\omega^{(2)} t + 2i\beta x]$. Because of the normal dispersion of ZnO,

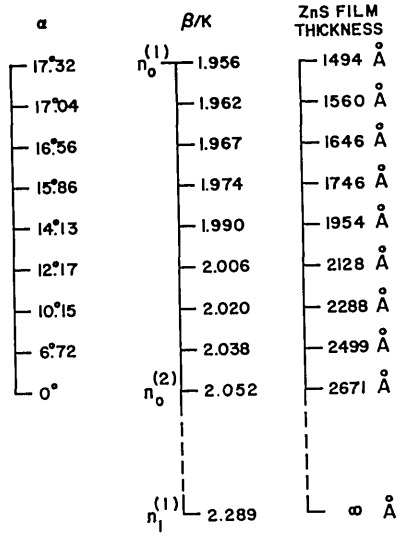

Fig. 29. Any horizontal line on this table indicates the corresponding values for the thickness of the ZnS film, β/k of the fundamental wave, and the Cerenkov angle α.

$n_0^{(2)} > n_0^{(1)}$. Therefore, by using a proper thickness W for the film, it was possible to obtain

$$n_0^{(1)} < (\beta/k) < n_0^{(2)}. \tag{24}$$

Under this condition, the phase velocity of the nonlinear polarization $\omega^{(2)}/2\beta$ exceeds the phase velocity $c/n_0^{(2)}$ of the free second harmonic radiation in the substrate. Consequently, Cerenkov radiation at the second harmonic frequency is emitted and it is emitted at a Cerenkov angle α where.

$$\cos\alpha = \beta/kn_0^{(2)}. \tag{25}$$

To review, the fundamental wave propagating in the ZnS film generates a sheet of nonlinear polarization wave in the substrate immediately below the film–substrate interface. The nonlinear polarization wave then generates a second harmonic radiation in the form of Cerenkov radiation that may be considered simply as a plane wave propagating in the substrate at an angle α below the interface. If the wave vector C represents this plane wave, its horizontal component is equal to the wave vector of the nonlinear polarization wave and is also twice the fundamental wave vector β. The process described above is illustrated in Fig. 29, in which any horizontal line drawn from a value of (β/k) at the middle column gives the corresponding Cerenkov angle α in the left column and the thickness of the film W in the right column. We notice that at the upper limit of the inequality (24), $\beta/k \to n_0^{(2)}$, the Cerenkov angle vanishes, and the phase velocity of the nonlinear polarization wave is equal to that of the free wave in the substrate. At the lower limit of (24), $\beta/k \to n_0^{(1)}$, the waveguide mode of the fundamental wave becomes cut off. Figure 30 is a photograph of the experiment. The bright star in the photograph is the Cerenkov radiation. Figure 31 shows the radiation as it emerged from the side surface of the ZnO crystal. The photograph was taken through a microscope by focusing it on the side surface of the crystal.

IX. Conclusions

Most of the material presented in this paper is drawn from unpublished notes accumulated during the past two years. These notes were prepared for talks and lectures given on various occasions. Much of the time and effort has been spent in developing a method whereby one can visualize easily the waveguide and radiation modes without having to derive the Maxwell equations. A theory that consideres a zigzag plane wave has been developed for that purpose. We have used this theory in Sec. II to derive the mode equation. It was used again in Sec. IV for a unified theory of the prism and grating couplers, and in Sec. VI for a simple theory of surface scattering. This theory turned out to be most useful in the analysis of complex optical devices. For example, the prism–film coupler involves four coupled media: the prism, the air gap, the film, and the substrate. A direct solution of the Maxwell equations for four simultaneously coupled media is not simple. Here, using the theory of the zigzag wave,

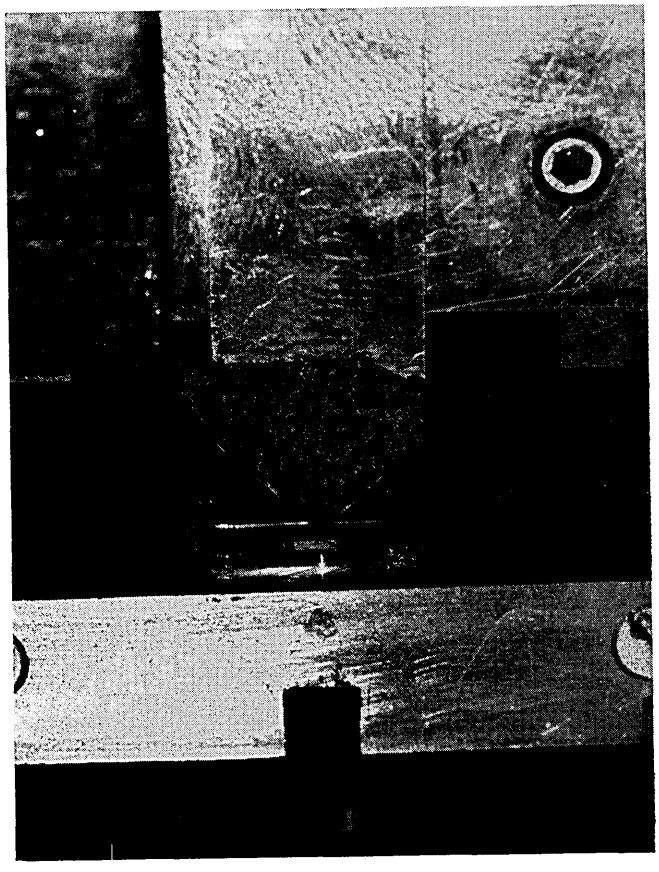

Fig. 30. Experiment of the second harmonic generation. The bright star in the figure is the second harmonic Cerenkov radiation emerging from the side surface of the ZnO substrate. The second harmonic beam thus generated is a coherent light beam of a very small aperture; the fundamental light wave at 1.064 μm is invisible.

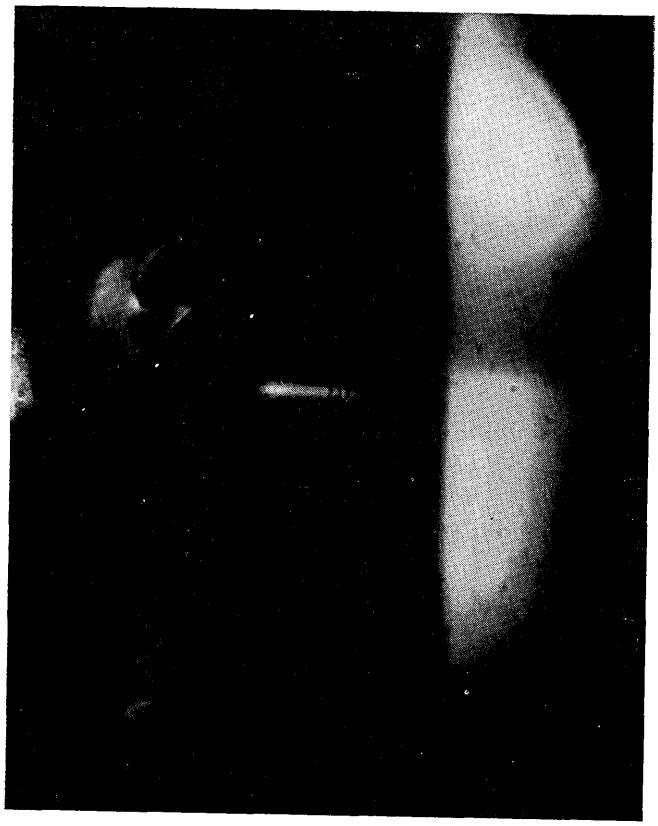

Fig. 31. Photograph of the second harmonic Cerenkov radiation taken through a microscope focused on the side surface of the ZnO substrate where the radiation emerges.

we only have to consider two sets of the waves, and each set consists of only two coupled waves.

We have not discussed in this paper the Goos-Haenchen shift, but it is included in the derivation of the mode equation by introducing the phase changes $-2\Phi_{10}$ and $-2\Phi_{12}$. It is also implied in the expression for the power flow and in the coupling constant of the prism–film coupler by introducing the effective film thickness. A detailed study of the Goos-Haenchen shift should include the discussion of energy flow and is beyond the scope of this paper.

The experiments described in this paper are difficult; one has to learn how to control the small dimensions involved in these experiments. For example, the spacing of the air gap between the prism and the film can be determined by observing Newton's rings near the pressure point. Usually, one should accurately measure the refractive indices of the prism and the substrate, tabulate the incident angles against the values of β/k, and compute the mode characteristics for each film–substrate combination.

The technology involved in integrated optics may be more difficult than that we can visualize today. Large single-crystal films are needed for the development of electrooptical and nonlinear thin-film devices. These single-crystal films, with the exception of epitax layers, simply do not exist today. Methods to control the uniformity and the thickness of the film within the accuracy of one or two atomic layers and techniques for the fabrication of structures 10^4 times smaller than their microwave counterparts are also needed. We are still observing the m lines in our best films; this indicates that considerable energy stored in the film and in the substrate is not in the main mode of propagation, and it would eventually distort the signal carried by the integrated optical circuitry. These problems described above will continue to challange us for some time to come. Looking into the future, we expect the field of integrated optics to grow rapidly, simply because there are needs for optical systems in the electronics and communication industries and there are needs for integrated optics in optical systems.

References

1. S. E. Miller, Bell Syst. Tech. J. **48,** 2059 (1969).
2. R. Shubert and J. H. Harris, IEEE Trans. MMT **16,** 1048 (1968).
3. P. K. Tien, R. Ulrich, and R. J. Martin, Appl. Phys. Lett. **14,** 291 (1969).
4. M. L. Dakss, L. Kuhn, P. F. Heidrich, and B. A. Scott, Appl. Phys. Lett. **16,** 523 (1970).
5. H. Kogelnik and T. Sosnowski, Bell Syst. Tech. J. **49,** 1602 (1970).
6. P. K. Tien and R. Ulrich, J. Opt. Soc. Am. **60,** 1325 (1970).
7. R. Ulrich, J. Opt. Soc. Am. **60,** 1337 (1970).
8. J. E. Midwinter, IEEE J. Quant. Electron. **QE-6,** 583 (1970).
9. J. H. Harris, R. Shubert, and J. N. Polky, J. Opt. Soc. Amer. **60,** 1007 (1970).
10. J. H. Harris and R. Shubert, IEEE Trans. MTT **19,** 269 (1971).
11. R. Ulrich, to be published in J. Opt. Soc. Am.
12. P. K. Tien and R. J. Martin, Appl. Phys. Lett. **18,** 398 (1971).
13. J. E. Goell and R. D. Standley, Bell Syst. Tech. J. **48,** 3445 (1969).
14. D. H. Hensler, J. D. Cuthbert, R. J. Martin, and P. K. Tien, Appl. Opt. **10,** 1037 (1971).
15. P. K. Tien, G. Smolinsky and R. J. Martin, "Thin Organosilicon Films for Integrated Optics," to be published in Appl. Opt.
16. P. K. Tien, R. Ulrich, and R. J. Martin, Appl. Phys. Lett. **17,** 447 (1970).
17. L. Kuhn, M. L. Dakss, P. F. Heindrich, and B. A. Scott, Appl. Phys. Lett. **17,** 265 (1970).
18. D. Hall, A. Yariv and E. Garmine, Appl. Phys. Lett. **17,** 127 (1970).
19. H. Kogelnik and C. V. Shank, Appl. Phys. Lett. **18,** 152 (1971).
20. J. E. Goell and R. D. Standley, Proc. IEEE **58,** 1504 (1970).
21. D. Marcuse, Bell Syst. Tech. J. **48,** 3187 (1969); **48,** 3233 (1969); **49,** 273 (1970); "Dependence of Reflection Loss on the Correlation Function" (private communication).
22. H. K. V. Lotsch, J. Opt. Soc. Am. **58,** 551 (1968).
23. M. Born and E. Wolf, *Principles of Optics* (Pergamon, New York, 1970), p. 49, Eq. (60).
24. J. J. Burke, J. Opt. Soc. Am. **61,** 676A (1971).
25. N. F. Foster, G. A. Coquin, S. A. Rozgonyi, and F. A. Vannatta, IEEE Trans. **S4-15,** 28 (1968).
26. S. A. Rozgonyi and W. J. Polito, Appl. Phys. Lett. **8,** 220 (1966).
27. P. Beckmann and A. Spizzichino, *International Series of Monographs on Electromagnetic Waves* (Oxford, New York, 1963), Chap. 5.