

Measurement of Thin Film Parameters with a Prism Coupler

R. Ulrich and R. Torge

The prism coupler, known from experiments on integrated optics, can be used to determine the refractive index and the thickness of a light-guiding thin film. Both parameters are obtained simultaneously and with good accuracy by measuring the coupling angles at the prism and fitting them by a theoretical dispersion curve. The fundamentals and limitations of this method are discussed, its practical use, and mathematical procedures for the evaluation.

I. Introduction

In various experiments on integrated optics,¹ thin dielectric films are used as planar light guides. The main parameters characterizing such films are the refractive index n and the film thickness W . For the determination of these parameters, numerous methods are possible. One method that is particularly well adapted to this problem is the prism coupling technique.²⁻⁴ It can give accurate results simply and fast, and a prism coupler is often used anyway in such experiments.

In brief, this method works as follows. The coupling of a laser beam by a prism into a planar dielectric light guide is governed²⁻⁵ by the angle θ of incidence of the light onto the prism base (Fig. 1). This angle θ determines the phase velocity in x direction, $v^{(i)} = c/n_p \sin\theta$, of the incident wave in the prism (index n_p) and in the gap. Strong coupling of light into the film occurs only when we choose θ so that $v^{(i)}$ equals the phase velocity v_m of one of the characteristic modes of propagation in the guide ($m = 0, 1, 2 \dots$). Thus, by determining these synchronous angles θ_m of strongest coupling, we can find experimentally the characteristic propagation constants of a given film, relative to the propagation constant $k = \omega/c$ of free space

$$1 \quad \tilde{N}_m = c/v_m = n_p \sin\theta_m. \quad (1)$$

On the other hand, we can calculate theoretical values N_m for the relative propagation constants

from the known dispersion equation of a planar dielectric light guide:

$$N_m = N(m, n, W, k, n_0, n_2, \rho). \quad (2)$$

Details of this calculation are in the Appendix. In Eq. (2), n_0 and n_2 are the refractive indices of the two media adjacent to the film (Fig. 1), and by ρ we indicate the polarization of the laser beam ($\rho = 0$ for TE polarization, and $\rho = 1$ for TM). All these parameters are known or can be measured separately. Therefore, it remains a computational problem to adjust the two unknown parameters n and W until the resulting theoretical values N_m match as closely as possible the experimental values \tilde{N}_m . We regard those parameters \tilde{n} and \tilde{W} , for which the match is best, as the results of this method.

In comparison with other methods, the one described here has two important advantages: (1) it requires only the measurement of angles, which can be done conveniently and with high precision; (2) if the film is thick enough to allow the observation of more than two modes of the same polarization, the method becomes a self-consistent one because the two unknowns n and W are then determined from more than two independent measurements. This improves the accuracy and greatly increases the confidence in the results.

There are, of course, also a number of disadvantages to this method that limit its applicability: (1) The film must be thick enough to permit the propagation of at least two modes. If only one mode can be observed, we may still use the prism coupler to determine one of the parameters n and W if the other is known from an independent measurement. (2) For reasons of convenient observation, a monochromatic laser beam should be available as the light source at the wavelength at which n is to be determined. (3) Because it is usually necessary to press the film mechanically against the base of the prism, the method does not work contactless. Yet, if care

M. Ulrich is with Max-Planck-Institut für Festkörperforschung, 7 Stuttgart, Germany; R. Torge is with Forschungsgruppe, Carl Zeiss, 7082 Oberkochen, Germany.

Received 23 April 1973.

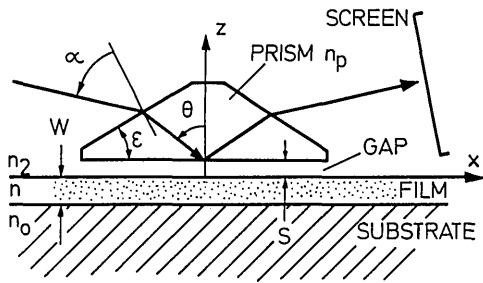


Fig. 1. Schematic cross section through a prism-film coupler.

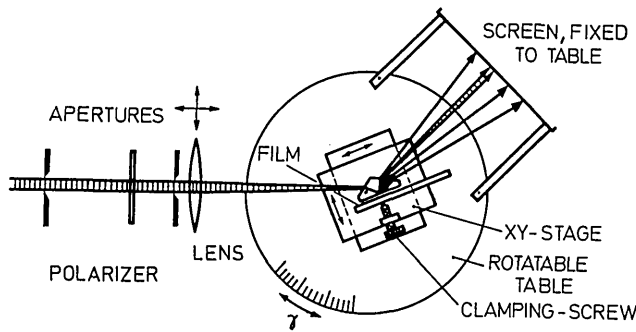


Fig. 2. Schematic arrangement of prism, input optics, and observation screen. The double arrows indicate adjustments.

is taken, the measurement is not destructive, either. In this respect, the method is obviously suited better for hard films than for soft films. (4) Finding the coupling angles requires a certain degree of skill and experience. Difficulties may occur due to excessive losses in the films, causing a broadening of the modes beyond recognition. Typically, an absorption up to 20 cm^{-1} in the direction of propagation (or 80 dB/cm) is tolerable, although it will already reduce somewhat the accuracy of the measurement. (5) The evaluation of the measurements requires lengthy, although simple, calculations. Therefore, the method is practical only where a programmable computer is available.

II. Experimental Technique

A. General Arrangement

The experimental arrangement² for the observation of the coupling and for the measurement of the coupling angles is shown schematically in Fig. 2. By means of a spring-loaded clamp, the film under test is pressed against the base of a coupling prism. The prism can be either symmetric [Fig. 3(a)] or a half-prism [Fig. 3(b)]. It sits on an xy translation stage that is mounted on a precision (<1 min of arc) rotary table or goniometer. Also mounted on the rotary table is an observation screen for use with the symmetric prism. The laser beam is linearly polarized (TE or TM) and must be of TEM_{00} cross section. A lens focuses the beam into the prism so that the beam waist coincides with the prism base. The

point where the beam strikes the prism base is the coupling spot. At this point, the parameters n and W are being measured. In the case of a symmetric prism the coupling spot is preferably near the center of the prism base; in a half prism it must be close to the corner.

The optical system is adjusted so that the coupling spot remains practically stationary on the prism base when the rotary table is rotated through the angular range where coupling is possible. This adjustment requires the direction of the input beam to be pointing slightly off the axis of the rotary table and the prism to be positioned suitably by means of the xy translation stage. Nearly perfect stationarity can be obtained with one fixed adjustment only for the range $\alpha > 0$ of input angles [Fig. 4(a)], while a different fixed adjustment works for the range $\alpha < 0$ [see Fig. 4(b)]. For the first range $\alpha > 0$ the laser beam must be offset from the axis of the table by approximately l_p/n_p toward the top of Fig. 2 and for $\alpha < 0$ by the same amount toward the bottom of that figure. Here l_p is the length of the path of light inside the prism from the entrance face to the coupling spot, measured near $\alpha = 0$. The prism must be adjusted so that it has the positions indicated in Fig. 4 relative to the axis of rotation. For a good stationa-

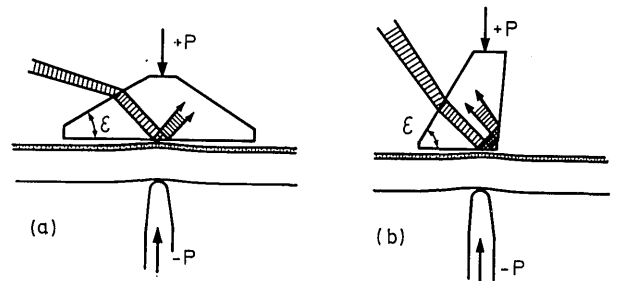


Fig. 3. Two possible prism shapes: (a) symmetric prism, useful with samples that are slightly flexible (e.g., glass substrates ≤ 1 mm) and yield elastically under the clamping pressure; (b) half prism, may also be used with thick samples because the coupling spot is close to the edge.

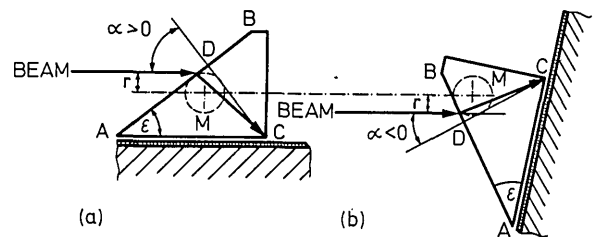


Fig. 4. Arrangement to keep the coupling spot stationary at C when the prism is rotated. The line CD is normal to AB and has length l_p . The axis of the rotary table coincides with the center M of the dashed circles. The radius of these circles is $r = l_p/n_p$, where n_p is the refractive index of the prism. The adjustment is optimum when the lines AB and CD and the incident beam are all tangential to the circles. This requires the beam to be offset by r from the axis of the table, and this offset must be toward B if $\alpha > 0$ as indicated in (a), but toward A if $\alpha < 0$ [see (b)].

rity it is furthermore advantageous for the prism to be small (for example, $l_p < 5$ mm) and have a high refractive index and for the lens to have a long focal length (for example, $f \geq 250$ mm) so that the coupling spot is not too small. A correct adjustment for stationarity is particularly important if the measured film does not have a uniform thickness.

Before performing the actual measurements we scan several times through all modes to watch their appearance and to identify the mode numbers. The mode with the largest angle α of incidence has the lowest mode number. If the measuring range of the prism is wide enough, this is $m = 0$. For such a survey of the modes, it is best to use a fairly high clamping pressure. In the actual measurement, then, the readings Γ_m of the rotary table that give the best coupling are noted. For the reasons given in Sec. III, it is necessary in taking each of these readings to reduce the clamping pressure to the minimum value that still permits an accurate determination of Γ_m . With high-quality films the angles Γ_m can be set reproducibly to within less than 1 min of arc. Because the different modes are measured at different clamping pressures, it is essential that the prism is mounted firmly so that it does not twist or rotate when the clamping pressure is changed. For a calibration of the zero point of the angular scale we determine also the reading Γ_\perp of that angle at which the beam is incident normally on the entrance face of the prism, i.e., where the reflection from the front face of the prism goes back exactly into the incident beam. The actual angles of incidence on the entrance face of the prism are

$$\alpha_m = \pm(\Gamma_m - \Gamma_\perp), \quad (3)$$

where the + or - sign applies depending on whether the scale on the turntable increases clockwise or anticlockwise.

B. Observation of the Modes

The coupling of the laser beam into the film can be observed in various ways, depending on the characteristics of film and substrate and on the type of prism used. Many evaporated or sputtered films with minute roughness or inhomogeneities (e.g., ZnS, ZnO, Al₂O₃, CeO₂) may be measured with the symmetric prism, especially when they are deposited on thin substrates that are slightly flexible [Fig. 3(a)]. In this case the coupling can usually be observed by the appearance of a streak of guided light in the film and simultaneously by the appearance of bright or dark m -lines² on the screen in Fig. 2. In some highly attenuating films, the streak of guided light may be so short that it can hardly be observed at all. Frequently, in such cases, however, the scattering is still strong enough so that the coupling spot itself shows a sudden increase in brightness when the prism is rotated through one of the coupling angles. This is sufficient to determine the positions Γ_m also in such highly scattering films.

The bright m -lines appear on the screen only when

the laser beam is coupled into a mode of the film. The light propagating in that mode is scattered into other directions (in the plane of the film) of the same mode and of the other modes. A fraction of the scattered light is then coupled out again by the prism and produces the bright lines on the screen. In accordance with this explanation it is characteristic for the bright m -lines that they all light up simultaneously when, during rotation of the prism, one of the coupling directions θ_m passes through the direction of the input beam. Moreover, in each of these coupling situations, the reflected beam on the screen coincides with one of the m -lines. Occasionally, one can also observe dark m -lines. On the screen, they appear in the weak background of light scattered at various roughnesses of the prism, particularly at its entrance face. This scattered light has a broad angular distribution, and it would illuminate the screen nearly uniformly. Scattered light, however, which propagates in directions inclined at one of the coupling angles θ_m with respect to the film normal, is coupled into the film. There, it may subsequently be guided away from the coupling spot, or it may be absorbed. Hence, light of these directions is missing from the nearly uniform illumination of the screen, thus causing the dark m -lines. From the above explanation it is clear that the dark m -lines appear at the same positions on the screen as the bright m -lines. Bright vertical lines on the screen may also be caused by scattering of the laser beam at horizontal polishing scratches on the prism faces. Such lines can be distinguished, however, from the m -lines by the fact that only the latter ones have fixed angular positions with respect to the prism.

Films with very smooth surfaces, like the organosilicon films⁶ and many solution-deposited films,⁷ do not show the bright m -lines in an arrangement like Fig. 2 because their scattering is too weak. These films are measured best by observing the intensity of the streak of guided light. Again, with films on slightly flexible substrates either type of prism of Fig. 3 may be used. The clamping pressure may cause a slight deformation of the films, resulting in a tapered coupling gap. As long as this taper is weak (angle $< 10^{-4}$), it is no disadvantage. Rather, it may yield a slightly higher coupling efficiency than a uniform gap.⁸ Smooth films on thick substrates are usually measured best with the half-prism [Fig. 3(b)].

C. Prism

Besides the prism shape (Fig. 3), the most important parameters of a coupling prism are its refractive index n_p and the prism angle ϵ . They determine the range of propagation constants N that can be measured with the prism. The propagation constant $\tilde{N} = n_p \sin\theta$ of the light along the prism base is related to the angle α of incidence on the entrance face of the prism (Fig. 1) by

$$\tilde{N} = \sin\alpha \cos\epsilon + (n_p^2 - \sin^2\alpha)^{1/2} \sin\epsilon. \quad (4)$$

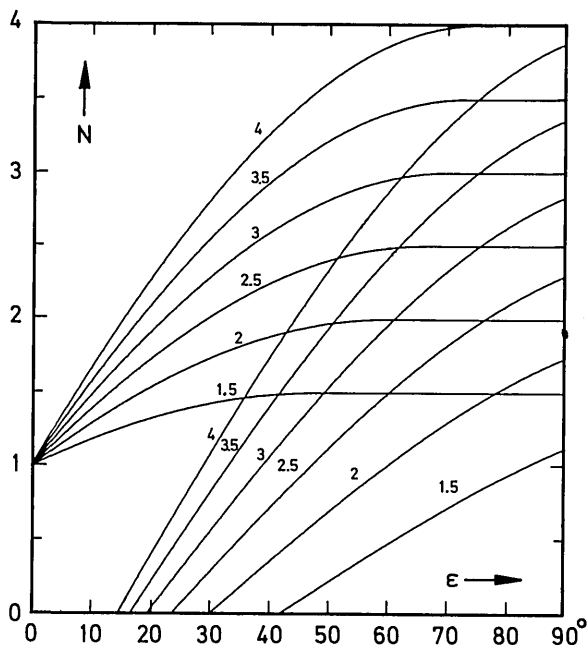


Fig. 5. Range of relative propagation constants N that can be measured with a prism of prism angle ϵ . Parameter at the curves is the prism index n_p . For each value of n_p there are two curves, giving the lower and upper limits of the possible N range.

If we let α vary from $-\pi/2$ to $+\pi/2$, we obtain the range of N -values possible with a given prism. We have shown the limits N_{\min} and N_{\max} of this range in Fig. 5 for a number of prism indices. In a film of index n , deposited on a substrate of index n_0 , all guided modes ($m = 0, 1, 2, \dots$) are in the interval $n_0 < N_m < n$. Therefore, in order to measure all modes of this film, a prism is required in general whose N range covers the interval from n_0 to n . In practice, the N range read from Fig. 5 must even be slightly wider than from n_0 to n , because in the limits of grazing incidence ($\alpha \rightarrow \pm\pi/2$), the usable cross section of the beam vanishes. The choice of N_{\min} may also be governed by the necessity to measure individually the index n_p of the completed coupling prism. This index enters through Eq. (4) into the evaluation, and therefore, it should be known very accurately. It can be determined conveniently by the method of minimum deviation, provided that one of the prism angles is so small that the associated $N_{\min} < 1$. Finally, if possible, one may want to choose the prism angle so that the N range of interest does not contain the case of normal incidence $\alpha = 0$ where the adjustment for a stationary coupling spot must be changed.

These considerations yield in practice prism angles from about 30° to 60° , depending on the index. In order to cover a very wide N range, it may be necessary to use two or even three prisms. Practical experience with various prisms has shown that it is advantageous to have none of the prism angles equal to 30° , 45° , 60° , 90° , or other small-integer fractions of 360° , or multiples thereof. Rather, one should stay

about 1° away from these angles, so that reflexes from the (uncoated) prism faces do not interfere with the main beam. This recommendation includes the largest angle of the half-prism [Fig. 3(b)]. That angle is preferably made slightly larger than 90° to give a clear angular separation of the incident and reflected beams. The edge at that angle should be fairly sharp, because a rounded edge here can cause a wide horizontal fan of scattered light that may illuminate the film in a manner resembling a guided streak.

The prism material should be as hard as possible to avoid damaging the prism base by repeatedly pressing films on it. Some suitable materials are the various high-index glasses, crystals like SrTiO_3 and TiO_2 (rutile), and Si and Ge for the ir. In the case of rutile, a uniaxial birefringent crystal, the optical axis should be oriented parallel to the edge of the angle ϵ . Only this orientation allows us to use in Eq. (4) a constant index n_p , which is the ordinary index for TM polarization, and the extraordinary index for TE. It need hardly be mentioned that the prism faces enclosing the angle ϵ should have a good optical polish. They must be flat to about $\lambda/2$ in order to define the angle ϵ with sufficient accuracy.

The prism may be replaced, in principle, by a hemispherical cylinder. The input beam then would pass the cylindrical surface at nearly normal incidence for all angles θ , thus extending the N range and simplifying the evaluation. An analysis of the hemispherical coupler shows, however, that it is critical with respect to small lateral displacements of the input beam, which must point exactly to the center of the cylinder. The prism-coupler, in comparison, is insensitive to parallel displacements of the input beam or of the prism.

III. Importance of Weak Coupling

It had been mentioned that for taking the final readings of the mode angles Γ_m , the clamping pressure must be reduced with each mode as far as possible without losing the mode. The reason is that a high pressure, i.e., strong coupling, broadens the modes and may shift them. We will show now that these effects become negligible when the coupling pressure is reduced to far below its value for optimum power transfer into the film. Furthermore, we will obtain an estimate of the accuracies possible in the determination of n and W by the prism-coupler method.

We first consider the question of mode broadening due to the prism. The coupling efficiency into the film is given theoretically^{5,8} by the overlap integral of the input field $V_3(x)$ with an output function $\Omega_m(x)$. This function describes the output field of the coupler in a situation where a guided wave in the m th mode is fed into the coupler from the right of Fig. 1. Associated with Ω_m is a certain radiation pattern of the output light. The main lobe of this pattern has an angular width that can be expressed in the N -scale as $\delta N_\Omega \approx \lambda/w_\Omega$. Here w_Ω is the spa-

tial width of the output field $\Omega_m(x)$, measured along the guide. According to the reciprocity theorem for antennas, the radiation pattern also acts as the receiving pattern of the coupler when used for input into the guide. Hence δN_Ω is also the angular input aperture of the coupler, and it may be compared with the angular aperture $\delta N_I = n_p \cos\theta \delta\theta_I$ of the incident laser beam. Here $\delta\theta_I$ is the angular beam spread measured inside the prism. When the coupler is rotated for the measurement of the θ_m , a mode can be observed in the film throughout an angular interval that corresponds to the larger one of δN_Ω and δN_I . While the beam aperture δN_I is practically fixed, we can vary the aperture of δN_Ω of the input pattern by adjusting the clamping pressure. We obtain the largest net power transfer ($\sim 80\%$) into the film when both apertures have about equal widths.⁵ If we reduce the clamping pressure, starting from this condition, until only a small fraction of the laser power enters the film, the input pattern of the coupler is much narrower than the aperture of the laser beam $\delta N_\Omega \ll \delta N_I$. The observed angular widths of the modes are then equal to the angular aperture $\delta\theta_I$ of the laser beam. The broadening due to the prism has become negligible.

The other disturbing influence of too strong coupling is a shift of the modes due to the presence of the prism in the vicinity of the film. According to Ref. 5, this shift can be expressed as

$$N_m^{(p)} - N_m = -K_m \cot 2\phi_{32}. \quad (5)$$

Here the superscript (p) is used to indicate the influence of the prism. K_m is the leak rate of the coupler, and it can be shown that

$$K_m \approx \frac{1}{2} \delta N_\Omega.$$

The angle $2\phi_{32}$ in Eq. (5) is the phase of the total internal reflection at the prism base.⁵ From Eq. (5) it follows that the mode shift vanishes as the coupling pressure is reduced and $\delta N_\Omega \rightarrow 0$, provided the factor $\cot 2\phi_{32}$ remains finite. This factor diverges only when $N \rightarrow n_2$ or $N \rightarrow n_p$. The first case does not occur in practice with films on a substrate because n_2 usually is free space so that $N > n_0 > n_2$. In the other limit $N \rightarrow n_p$, we find

$$\cot 2\phi_{32} < (n_p/n_2)^{2\rho} / 2 \cos\theta.$$

This expression is bounded because we have to stay away from grazing incidence $\theta \rightarrow 90^\circ$ anyway. If we keep $\theta \leq 70^\circ$, for example, we have in a rutile prism $\cot 2\phi_{32} < 1.6$ for TE polarization ($n_p = n_{eo} = 2.86$) and $\cot 2\phi_{32} < 10$ for TM ($n_p = n_o = 2.58$).

IV. Accuracy

The preceding considerations may also help obtain an idea of the accuracy with which the index n of a film can be determined. For films of low attenuation, the accuracy in measuring the mode positions is limited by the aperture δN_I of the input beam. Near normal incidence on the prism $\alpha \approx 0$, we have $\delta N_I \approx \cos\epsilon \delta\alpha$, where $\delta\alpha$ is the angular aperture of

the laser beam outside the prism. By visual observation, the mode positions can be determined to within some fraction of δN_I , typically $\delta N_I/3$. Therefore, we may regard the value $\delta N_I/3$ as the approximate accuracy δn obtainable for the film index n . As an example, we consider a laser beam of 1-mm diam, focused by a lens of $f = 500$ mm into a prism of $\epsilon = 45^\circ$. Here we have $\delta\alpha = 0.002$ and $\delta N_I \approx 0.0014$, so that n should be measurable to an accuracy of $\delta n \approx 0.0005$. If several modes are observed and multiple readings are taken, an accuracy of $\delta n \approx 0.0001$ can be obtained in practice. This is demonstrated in Sec. VI. It is interesting to note that this accuracy is obtained by measuring a film area whose length is diffraction limited by the beam aperture $\delta\alpha$, the width of the coupling spot being $w_I \approx \lambda/\delta N_I \approx \lambda/\cos\epsilon \delta\alpha$. In the above example, we have $w_I \approx 500 \mu\text{m}$. It follows, furthermore, that the attenuation of the guide over the distance w_I must not be too high. If we denote the power attenuation coefficient by A , the accuracy of the measurements begins to be seriously degraded when $A w_I > 1$. In our example, this limit is at $A = 20 \text{ cm}^{-1}$ or 80 dB/cm.

We estimate the accuracy in the determination of the film thickness W from the relation $\delta W \approx \delta n (\partial W / \partial n)_{\bar{m}}$. Here we take the derivative at constant indices and fixed wavelength for a mean mode number \bar{m} . For a not-too-thin film and $n_0 \geq n_2$, the total number of modes of one polarization is of the order

$$M_{\text{tot}} \approx \pi^{-1} k W (n^2 - n_0^2)^{1/2}, \quad (6)$$

and we take $\bar{m} = M_{\text{tot}}/2$. The corresponding mean propagation constant \bar{N} is, in the same approximation, given by $\bar{N}^2 = (n_0^2 + 3n^2)/4$, and we obtain after further simplification

$$\delta W/W \approx 2\delta n/(n - n_0). \quad (7)$$

This formula shows that the possible accuracy of the film thickness is reduced for films of low index differences ($n - n_0$). In the example to be discussed in Sec. VI, we have $(n - n_0) \approx 0.15$ and $\delta n \approx 0.00020$, and Eq. (7) would predict an accuracy of $\delta W/W \approx 0.0025$. The actual error found in Sec. VI is about four times higher.

V. Evaluation of Film Index and Thickness

The results of the measurements are a list containing the observed modes of the film, identified by their mode numbers $m = 0, 1, 2, \dots$, and for each mode the measured coupling angle Γ_m . For simplicity, we assume that all modes have the same polarization. The modes of the other polarization may be treated correspondingly, providing a good check on the results of the first polarization. These pairs (m, Γ_m) , together with the constant parameters of the measurement $(\Gamma_\perp, n_0, n_2, \epsilon, n_p, \lambda, \rho)$, are fed into a computer. It first calculates and prints out the set of observed propagation constants, \bar{N}_m , using Eqs. (3) and (4). The further evaluation depends on

whether the number M of observed modes in the list is $M = 2$ or $M \geq 3$. In the case of $M = 2$, a relatively simple procedure yields unique values n and W . If $M \geq 3$, a more lengthy least square fit of the observations is used to calculate "best" values \bar{n} and \bar{W} of film index and thickness together with their rms deviations.

A. Two Observed Modes

The mode numbers of the two modes are called μ and ν . Usually these will be 0 and 1, but we may allow here μ and ν to be any two modes of equal polarization. The observed propagation constants \tilde{N}_μ and \tilde{N}_ν are related to the unknown n and W by the dispersion equation of the planar dielectric guide. We write this equation as

$$kW(n^2 - N_m^2)^{1/2} = \Psi_m(n, N_m), \quad (8)$$

where

$$\Psi_m(n, N_m) \equiv m\pi + \phi_0(n, N_m) + \phi_2(n, N_m) \quad (9)$$

and

$$\phi_j(n, N_m) \equiv \arctan \left[\left(\frac{n}{n_j} \right)^{2\rho} \left(\frac{N_m^2 - n_j^2}{n^2 - N_m^2} \right) \right]^{1/2}, \quad (10)$$

with $j = 0, 2$. Inserting \tilde{N}_μ and \tilde{N}_ν into Eqs. (8)-(10) yields two equations from which kW can be eliminated. The resulting single equation for n may be written in the form

$$n^2 = F(n^2), \quad (11)$$

with

$$F(n^2) \equiv (\tilde{N}_\mu^2 \Psi_\nu^2 - \tilde{N}_\nu^2 \Psi_\mu^2) / (\Psi_\nu^2 - \Psi_\mu^2). \quad (12)$$

Equation (11) cannot be solved explicitly for n^2 , but a solution is readily found by iterating this equation. We thus calculate the series $n_{[q]}^2$ for $q = 1, 2, 3 \dots$ by the recursion formula

$$n_{[q]}^2 = F(n_{[q-1]}^2). \quad (13)$$

This series converges to the required solution

$$n^2 = \lim_{q \rightarrow \infty} n_{[q]}^2,$$

provided that $|\partial F / \partial(n^2)| < 1$. Because we could not verify this inequality analytically, we have studied the convergence of the iteration Eq. (13) numerically for a very wide range of arguments, including all possible situations of practical interest. We found that Eq. (13) did converge in all cases, provided that the first term $n_{[0]}^2$ of the series satisfied the relations

$$n_{[0]}^2 > \tilde{N}_\mu^2 \quad \text{and} \quad n_{[0]}^2 > \tilde{N}_\nu^2.$$

This is physically reasonable and can always be guaranteed by a suitable choice of the initial term $n_{[0]}^2$. We have used, for example, $n_{[0]}^2 = 1.1\tilde{N}_\mu^2$, where $\tilde{N}_\mu > \tilde{N}_\nu$ was guaranteed by sorting the modes by ascending mode numbers, i.e., descending N_m . The convergence of the series Eq. (13) has an oscillatory character, reaching an error of less than 10^{-6}

after typically ten to twenty steps. The iteration is interrupted after the change of $n_{[q]}^2$ in one step falls below some limit (e.g., 10^{-6}) depending on the numerical accuracy of the computer. With n thus known, we obtain the film thickness W directly from Eq. (8).

B. Three or More Observed Modes

If the number of observed modes is $M \geq 3$, the two unknowns n and W are overdetermined. ($M - 2$) of the M measurements are redundant and may be considered as providing checks on the validity of the other two. If a larger number of modes has been measured, e.g., $M = 5 - 10$, we can therefore obtain experimentally a good idea of the self-consistency of the measurements and of the accuracy of \bar{n} and \bar{W} of a given film.

For the evaluation we define an error sum

$$\sigma(n, w) = \sum_m [\tilde{N}_m - N_m(n, w)]^2. \quad (14)$$

Here the \tilde{N}_m are the observed propagation constants, whereas the $N_m(n, w)$ are the solutions of the dispersion Eq. (8) for a given combination of film index and thickness. Details of computing $N_m(n, w)$ are given in the Appendix. In Eq. (14) we have introduced the dimensionless variable $w = kW$ instead of the physical thickness W . The summation in Eq. (14) runs over all M observed modes. We have defined the error sum σ using the relative propagation constants N_m . More strictly, if all measurements were done with the same angular accuracy, the error sum should be formed with the calculated and observed angles θ_m . However, the use of the N_m instead of the θ_m gives us a simpler formulation at the price of slightly different statistical weights of the measurements.

The best values \bar{n} and \bar{w} are those that minimize the error sum. Necessary conditions for the minimum are

$$\sigma_n(\bar{n}, \bar{w}) = 0 \quad \text{and} \quad \sigma_w(\bar{n}, \bar{w}) = 0, \quad (15)$$

where the subscripts indicate the partial derivatives with respect to n or w , respectively. Equations (15) cannot be solved analytically for \bar{n} and \bar{w} because the dispersion $N_m(n, w)$ cannot be expressed in closed form (see Appendix). Therefore, we determine a solution numerically by a gradient method: We assume that an approximate solution $n_{[q]}$ and $w_{[q]}$ is already known. We expand σ formally into a power series up to second order terms around a point $P_{[q]} = \{n_{[q]}, w_{[q]}\}$ in the $\{n, w\}$ plane, and we insert this series into the Eqs. (15). We obtain an improved solution $P_{[q+1]}$ by progressing from $P_{[q]}$ a suitable distance in the direction of the negative gradient of $\sigma(n, w)$

$$n_{[q+1]} = n_{[q]} + (\sigma_w \sigma_{nw} - \sigma_n \sigma_{ww}) / det, \quad (16)$$

$$w_{[q+1]} = w_{[q]} + (\sigma_n \sigma_{nw} - \sigma_w \sigma_{nn}) / det, \quad (17)$$

where

$$det \equiv (\sigma_{nn} \sigma_{ww} - \sigma_{nw} \sigma_{wn}). \quad (18)$$

The derivatives appearing here are all required at $P_{[q]}$. For the numerical calculation they are approximated by their corresponding central differences

$$\sigma_n = [\sigma(1,0) - \sigma(-1,0)] / 2h_n,$$

$$\sigma_w = [\sigma(0,1) - \sigma(0,-1)] / 2h_w,$$

$$\sigma_{nn} = [\sigma(1,0) - 2\sigma(0,0) + \sigma(-1,0)] / h_n^2,$$

$$\sigma_{ww} = [\sigma(0,1) - 2\sigma(0,0) + \sigma(0,-1)] / h_w^2,$$

$$\sigma_{nw} = \sigma_{wn} = [\sigma(1,1) - \sigma(-1,1) - \sigma(1,-1) + \sigma(-1,-1)] / 4h_n h_w. \quad (19)$$

We have used here an abbreviated notation for the nine argument combinations at which the error sum σ has to be evaluated

$$\sigma(r,s) = \sigma(n_{[q]} + r h_n, w_{[q]} + s h_w).$$

The small increments h_n and h_w in Eqs. (19) should be chosen according to the numerical accuracy of the computer; we have used $h_n = 10^{-6}$ and $h_w = 10^{-5}$. The above procedure is repeated from the new point $P_{[q+1]}$, yielding $P_{[q+2]}$, and so on, until the error sum has become sufficiently stationary (e.g., $|\Delta\sigma/\sigma| < 10^{-3}$) and the changes of n and w in one iteration have dropped below the practically interesting tolerances. According to our experience this happens after five to ten iterations, depending on the accuracy of the initial values $n_{[0]}$ and $w_{[0]}$. A good set of these initial values may be obtained conveniently by dropping $M - 2$ of the measured \tilde{N}_m and applying to the remaining two the method described above for the $M = 2$ case. A closer consideration shows that this method gives the best results if the two modes with the highest and the lowest order m are retained.

When the iteration has converged and the best values \bar{n} and \bar{w} of the film index and thickness are known, we determine their mean square errors. For this purpose we compute first for each observed mode m that value $n_{(m)}$ of the film index which, in combination with \bar{w} , would yield exactly the observed \tilde{N}_m . These film indices are found by the interpolation formula

$$n_{(m)} = \bar{n} + 2h_n [\tilde{N}_m - N_m(\bar{n}, \bar{w})] / [N_m(\bar{n} + h_n, \bar{w})]. \quad (20)$$

The $3M$ values of $N_m(\bar{n}, \bar{w})$ required in these M equations are already existing at this point of the computer program from the last iteration step. Corresponding to the $n_{(m)}$ we also compute for each observed mode that value $w_{(m)}$ of the film thickness, which, in combination with \bar{n} , would yield exactly

the observed \tilde{N}_m . These thicknesses are found directly from Eq. (8). The resulting rms errors of \bar{n} and \bar{w} are then

$$\delta\bar{n} = \left\{ \sum_m [n_{(m)} - \bar{n}]^2 / (M - 1)(M - 2) \right\}^{1/2}, \quad (21)$$

$$\delta\bar{w} = \left\{ \sum_m [w_{(m)} - \bar{w}]^2 / (M - 1)(M - 2) \right\}^{1/2}. \quad (22)$$

In conclusion we note that the existence of a minimum of σ is assured by physical reasons, i.e., by the expectation that the film modes obey the theoretical dispersion relation Eq. (8) and that the measurements have been done with a certain minimum accuracy. The method may fail to converge if the film or the substrate⁹ have an inhomogeneous index distribution, or if the measurements are grossly wrong (e.g., a wrong assignment of the mode numbers m to the observed angles Γ_m). Difficulties of convergence may also occur near the lower limit n_0 of the range in which the N_m exist. If the separation of an observed \tilde{N}_m from this limit is of the order of h_n , the approximations [Eqs. (19)] break down. It is best in such a case to ignore that highest (near-cutoff) mode. The same is necessary if $\tilde{N}_m < n_0$ should be encountered, e.g., due to small errors in reading Γ_m or Γ_\perp . Because of such occasional difficulties we found it more convenient to run the evaluation program on a time-sharing computer terminal than as a batch-processed

Table I. Examples for the Determination of Index n and Thickness W of Two Light-Guiding Films^a

Measurement	1	2	3	4	5
\tilde{N}_m^m	0	0	0	0	1
	1.625259	1.625225	1.625469	1.608802	1.608802
\tilde{N}_m^m	1	1	1	1	2
	1.613519	1.613479	1.613907	1.592410	1.592410
\tilde{N}_m^m	2	2	2	2	3
	1.593590	1.593614	1.593877	1.564619	1.564619
\tilde{N}_m^m	3	3	3	3	4
	1.566527	1.566635	1.567191	1.525426	1.525426
Initial values ^b					
$n_{[0]}$	1.629133	1.629090	1.629313	1.614310	1.624464
(Å) $W_{[0]}$	26006.95	26040.93	26115.59	21491.16	25903.67
Final \bar{n}	1.62901	1.62895	1.62921	1.614421	1.6263
values	$\pm 16 \times 10^{-5}$	$\pm 17 \times 10^{-5}$	$\pm 23 \times 10^{-5}$	$\pm 71 \times 10^{-5}$	$\pm 130 \times 10^{-5}$
\bar{W} [Å]	25999	26033	26086	21485	25844
	± 201	± 221	± 197	± 96	± 680
	1.58×10^{-7}	1.67×10^{-7}	3.15×10^{-7}	0.30×10^{-7}	1.06×10^{-7}

^a Columns 1-3 refer to evaporated Al_2O_3 films on a quartz substrate, column 4 to a film of Kodak Photoresist KPR on quartz. Column 5 is the same as column 4, but using a deliberately wrong assignment of the mode numbers. The error sum and the errors of \bar{n} and \bar{W} , are drastically increased.

^b Calculated from highest and lowest observed mode.

job. In the latter case, the programmer must provide for all these eventualities in advance.

The method described here can easily be extended to the case of birefringent films in which one axis of the index ellipsoid is normal to the film. With such films, one evaluates first the \tilde{N}_m obtained for TE propagation along the two principal axes in the plane of the film. This yields two of the three principal indices of refraction. From these and the TM measurements, the third principal index is found, using a generalized form¹⁰ of Eqs. (8) and (9).

VI. Examples

An early example, although without details, was contained in Ref. 2. A few more examples are in Table I. Measurements 1 to 3 concern an Al_2O_3 film deposited by e-gun evaporation onto a quartz substrate of index $n_0 = 1.45707$. This film was measured at the He-Ne laser wavelength with an SF 13 prism ($\epsilon = 63.0129^\circ$, $n_p = 1.73519$) and showed four TE modes. The film was measured three times at the same spot. Different, but all weak, coupling pressures were used in these measurements. Measurement 3 was taken after removing the film from the prism and readjusting the arrangement from the beginning. Each angle α_m was obtained by averaging four readings. By comparing columns 1 to 3, we see that within the measured tolerances of the resulting index and film thickness we have the same results from all three measurements. This demonstrates that under weak coupling conditions the pressure changes do not influence the results.

In the other example (4 in Table I) we measured a film of Kodak KPR2 Photoresist on quartz.⁷ The resulting mean errors are smaller because this film has much less scattering losses, so that the setting for optimum coupling could be observed with higher precision.

In order to demonstrate the self-consistency of the method, we have repeated in column 5 the coupling angles of column 4, but we have deliberately used a wrong assignment of the mode numbers m to these angles. As a result, the errors of both \bar{n} and \bar{w} are seen to increase drastically.

Except for this last case, the error limits in Table I do not exceed much the errors that were expected for this method by theoretical considerations. The computation time for the evaluation of each measurement (four modes) was approximately 8 sec on a CDC 3200 computer.

Appendix: Inversion of the Dispersion Formula

The dispersion formula [Eq. (8)] cannot be solved analytically to give $N_m(n, w)$. For use in a computer program, however, the inversion may be performed in a subroutine by a simple trial and error procedure. We utilize the fact that the function

$$w_m(N) = (n^2 - N^2)^{-1/2} \Psi_m(N)$$

is strictly monotonic, $\partial w_m / \partial N > 0$, and that we may restrict the discussion to the finite interval $n_0 \leq N \leq$

n of the argument N . We calculate $w_m(N_{1/2})$ at the center point $N_{1/2}$ of this interval and compare the result with that thickness w for which we want to find $N_m(n, w)$. If $w_m(N_{1/2}) = w$, the unknown is $N_m = N_{1/2}$. If $w_m(N_{1/2}) > w$, the unknown N_m must lie in the right half of that interval, $N_{1/2} < N_m \leq n$, otherwise in the left half. That half in which N_m is located is sectioned at its center and so on, until N_m is enclosed into a sufficiently small interval $N_- < N_m < N_+$. We used here an interval of width $(N_+ - N_-) \leq 10^{-4} (n - n_0)$. For a narrower interval, roundoff errors may become significant. In a final step, then, w_m is determined by linear interpolation between $w_m(N_+)$ and $w_m(N_-)$.

We thank R. Kersten for a critical reading of the manuscript and a stimulating discussion about the numerical precision required for the computations.

The work on which this paper is based was partly supported by the Bundesminister für Forschung und Technologie as part of the technical program of this Ministry. The authors are exclusively responsible for the contents of this paper.

Note Added in Proof: It is not clear, however, whether there is only a simple minimum of $\sigma(n, w)$. Rather, indications exist that the $\sigma(n, w)$ surface may have several relative minima. It then depends on the particular initial values $n_{|0}$ and $w_{|0}$ chosen, into which the minima iteration will converge. In this way, slightly different values \bar{n} and \bar{w} may result from the same set of measurements. The only guide to the reliability of a final solution \bar{n} and \bar{w} is the smallness of the error sum σ , for which the values given in Table I are representative.

Another difficulty also is inherent in this method, which uses only the local topology of the $\sigma(n, w)$ surface. In the beginning of the iteration, the steps in the $\{nw\}$ plane are large, and, depending on the global topology of the $\sigma(n, w)$ surface, it may happen that the error sum σ at $P_{|q+1}$ calculated from Eqs. (16) and (17) is larger than at $P_{|q}$. In such cases one can still obtain convergence by using not the full correction terms of Eqs. (16) and (17) but only a fraction of them. The fraction is made so small (e.g., $1/2$, $1/4$, $1/8$, ...) until $\sigma_{|q+1} < \sigma_{|q}$ (damped iteration). From the new point on, one may proceed as before.

References

1. P. K. Tien, Appl. Opt. 10, 2395 (1971).
2. P. K. Tien, R. Ulrich, and R. J. Martin, Appl. Phys. Lett. 14, 291 (1969).
3. J. H. Harris, R. Shubert, and J. N. Polky, J. Opt. Soc. Am. 60, 1007 (1970).
4. P. K. Tien and R. Ulrich, J. Opt. Soc. Am. 60, 1325 (1970).
5. R. Ulrich, J. Opt. Soc. Am. 60, 1337 (1970).
6. P. K. Tien, G. Smolinsky, and R. J. Martin, Appl. Opt. 11, 1313 (1972).
7. R. Ulrich and H. P. Weber, Appl. Opt. 11, 428 (1972).
8. R. Ulrich, J. Opt. Soc. Am. 61, 1467 (1971).
9. F. Zernike, E. L. Sloan, J. C. Webster, R. B. McGraw, and W. L. Knecht, *Topical Meeting on Integrated Optics, Guided Waves, Materials, and Devices*, Paper TuA9, Las Vegas, Nevada (February 1972).
10. M. Tacke and R. Ulrich, Opt. Commun. 6, 234 (1973).