LMSC-D674593

RESEARCH ON SPECTROSCOPIC IMAGING

VOLUME II
REFERENCE LITERATURE

APRIL 1979

Prepared by

ALAN TITLE

WILLIAM ROSENBERG

LOCKHEED PALO ALTO RESEARCH LABORATORIES

3251 Hanover Street

Palo Alto, California 94304

Prepared for

LUNAR AND PLANETARY PROGRAMS

NASA HEADQUARTERS

Washington, D. C. 20546

UNDER CONTRACT NASW-3107

LMSC-D674593

RESEARCH ON SPECTROSCOPIC IMAGING

VOLUME II
REFERENCE LITERATURE

APRIL 1979

Prepared by
ALAN TITLE
WILLIAM ROSENBERG

LOCKHEED PALO ALTO RESEARCH LABORATORIES
3251 Hanover Street
Palo Alto, California 94304

Prepared for
LUNAR AND PLANETARY PROGRAMS
NASA HEADQUARTERS
Washington, D. C. 20546

Permissions

The following article has been reprinted with the permission of the authors and the copyright holders.

"Birefringent Filter for Millimeter Waves", Bernard M. Schiffman and Leo Young

Contents

15. Title, A. M.: Improvement of Birefringent Filters. 2: Achromatic Waveplates
Appl. Opt. 14, 229-237 (Jan. 1975)

16. Title, A. M.: Improvement of Birefringent Filters. 3: Effect of Errors on Wide Field Elements
Appl. Opt. 14, 445-449 (Feb. 1975)

17. Title, A. M.: Improvement of Birefringent Filters. 4: The Alternate Partial Polarizer Filter
Appl. Opt. 15, 2871-2879 (Nov. 1976)

AMERICAN METEOROLOGICAL SOCIETY
45 BEACON STREET
BOSTON, MASSACHUSETTS
02108

TRANSLATION OF

THE BIREFRINGENT FILTER AND ITS APPLICATIONS IN SOLAR PHYSICS

(Le filtre monochromatique polarisant et ses applications en physique solaire)

by

Bernard Lyot

Annales d'Astrophysique, 7 (1/2): 31-79, 1944.

This translation has been made by the American Meteorological Society under Contract AF 19(604)-6113, through the support and sponsorship of the

AIR FORCE CAMBRIDGE RESEARCH LABORATORIES

OFFICE OF AEROSPACE RESEARCH

L. G. HANSCOM FIELD

BEDFORD, MASSACHUSETTS

T-F-59+

1. "The birefringent filter and its applications in solar physics"

2. "Le filtre monochromatique polarisant et ses applications en physique solaire"

3. Lyot, Bernard. Annales d'Astrophysique, 7(1/2): 31-79, 1944.

4. 79 pages

5. Date of translation: December 1963

6. Translator: John Macleay

7. Translated for Air Force Cambridge Research Laboratories, Office of Aerospace Research, United States Air Force, L. G. Hanscom Field, Bedford, Massachusetts, by the American Meteorological Society, Contract number AF 19(604)-6113.

8. Unclassified

9. Complete

THE BIREFRINGENT FILTER AND ITS APPLICATIONS IN SOLAR PHYSICS

By

Bernard Lyot

SUMMARY. Observation of the corona by the light of its emission lines requires the use of a very luminous, highly monochromatic filter with a wide field. None of the instruments in use twenty years ago fulfilled these three conditions simultaneously, and this led the author to conceive and construct a new one, the birefringent filter. The history of this filter is given.

The Theory of the Birefringent Filter.

A description of the principle of the instrument and its operation for rays normal to the entry and exit faces is presented. Variations of the transmitted wavelength as a function of temperature and two methods for varying this wavelength are studied. Variations of the transmitted wavelength as a function of the rays' obliquity and three methods for increasing the field of the instrument are considered.

First Filter. The first variable wavelength filter is described, together with its thick and thin crystalline plates, its general arrangement and its adjustment.

Second Filter. A second filter which isolates six fixed radiations belonging to the chromosphere and to the corona is also described. Its characteristics and mounting are discussed as is the thermostat which adjusts the transmitted wavelengths and insures their constancy. First observations of the corona and prominences are discussed.

Replacement of the filter's polaroids by calcite polarizers, the advantages of this modification and the new results obtained thereby are explained. Details of the corona are studied and the red and green coronas are compared.

Three-Color Cine Photography. A method for separating four radiations transmitted by the filter and for recording three of them on motion picture film is described. A description is also given of the three-color camera for simultaneous photography of the corona with the red line, of prominences with the H_α line and the corona with the green line. Films obtained in this manner are discussed. Changes in the corona and their interpretation through relative variations of intensity are examined.

Study of the Chromosphere. The addition of a plate to the filter to enhance its monochromatic quality is explained as is the addition of a compensator for varying the transmitted wavelengths to observe radial velocities. Observation of the chromosphere at the solar limb and motion pictures of the boiling of the chromospheric surface are discussed. The observation of the chromosphere on the solar disk and motion pictures of flares are also treated.

INTRODUCTION

There are certain heavenly bodies whose observation requires that their image be formed, not by using all their light as is generally done, but by using only one of the radiations they emit to the exclusion of the rest of the spectrum. This is the case, for example, when one wishes to study the chromosphere against the solar disk, prominences through a clear sky or the corona by the light of its bright lines.

The problem consists therefore in observing an extended object while allowing only a more or less narrow region of its spectrum containing a given radiation to be transmitted. Any device which solves this problem is a filter; the narrower the portion of the spectrum transmitted, the more monochromatic such a filter is.

The transparency of such a filter varies rapidly as a function of wavelength and the curve representing its variations passes through a maximum for a given wavelength.

The filter's properties may be summed up by three quantities:

1. Its maximum transparancy.
2. The corresponding wavelength.
3. Its equivalent width, i.e., the width of a rectangle which has the same area and same height as the curve. This third quantity indicates to what degree the filter is monochromatic. It defines the purity of the transmitted light and allows calculation of the proportion of stray radiations that it transmits.

The First Monochromatic Filters. Three types of filters were in use about twenty years ago:

1. Glass or gelatine filters tinted with mineral salts or with organic coloring.

These filters, long known and convenient to use, are only very slightly monochromatic. Their equivalent width rarely drops below 300 Å for a transparency of 10%; it can be reduced slightly by increasing the filter thickness, but at the same time transparency rapidly falls below tolerable values.

For some spectral regions, the salts of rare earths, mainly those of neodymium, improve these filters and reduce their equivalent width to about 100 Å [1]; however, such widths are still at least 10 times too large to allow detection of the chromosphere against the solar disk or to consider observation of the corona with the light of its bright lines.

2. A second type of filter, more monochromatic, was devised by Christiansen.

It is made of powdered glass impregnated with a much more dispersive liquid. For a certain wavelength, the index of the liquid equals that of the glass and the mixture is transparent, like a homogeneous body. When the wavelength varies, the two indices differ more and more, and the mixture diffuses the light at increasingly large angles.

This very luminous filter transmits a band whose equivalent width can drop to a few angstroms; the band can be shifted by temperature variation. Unfortunately, the radiations of other wavelengths are not absorbed, but are diffused and mask the image of the source as soon as the latter's apparent diameter ceases to be very small. This serious drawback makes the Christiansen filter almost useless for solar study.

3. A third type of filter, the most monochromatic of all, is the spectroheliograph invented by Deslandres and by Hale in 1905.

This spectroscope images the spectrum on the jaws of a second slit located where the radiation to be isolated is projected. Both the image of the sun before the first slit and the photographic plate behind the second are displaced simultaneously and continuously by a special mechanism. Thus the instrument records progressively, line by line but without discontinuity, a monochromatic image of the entire sun.

This apparatus may be considered a very monochromatic filter. Its equivalent width depends on the widths of the two slits and on the dispersion and resolving power of the spectrograph. The instrument in the Meudon Observatory for continuous chromospheric observation has an equivalent width of about 2/10 Å in the violet and 4/10 Å in the red, i.e., a thousand times lower than that of the best colored filters. The wavelength of the spectroheliograph is adjustable, but its light yield is very weak because it photographs the sun's different linear regions successively; this reduces the light transmitted by the spectrograph in the relation of the width of the first slit to that of the solar image. Thus the light yield of the spectroheliograph falls to a few ten-thousandths.

On the chromosphere, which is very bright, this instrument has long given excellent results. However, it is not sufficiently luminous for study of the corona, whose brightness is several hundred thousand times weaker. *)

History of the Birefringent Filter. These considerations led me, in 1920, to search for a new type of filter which would be

*) I verified this fact in 1931 at the Pic du Midi, using a coronograph with 8 cm aperture followed by a spectroheliograph having two large 60° prisms with bases of 160 mm and wide slits isolating the coronal green line in an equivalent band width of 2 Å. The image of the sun on the plate was reduced to 16 mm to increase its brightness. A 90-minute exposure gave a weak image of the corona on which a single coronal jet is visible; moreover, the diffusion caused by atmospheric dust varied during the exposure, and the variable illumination of the second slit gave a streaked plate [2].

highly monochromatic and very luminous. After trying to utilize the interference given by a series of half-silvered plates and finding that the images obtained were not sufficiently bright, I studied the interference produced by a series of crystalline plates in polarized light and, in 1927, established the principle of the filter considered in this paper.*) In 1933 I published a brief note [8] stating this principle, describing the properties of the composite crystalline plates and giving the data needed to calculate and construct a filter which, in a very wide field, would isolate a spectral band, variable in wavelength, with an equivalent width of 1 Å in the green. The filter's main plates, 25 mm in diameter, were cut in 1933 and tested in the laboratory. The instrument was to contain 10 large Glazebrook-type polarizers, but I could not obtain the required calcite, and thus my research was halted temporarily.

In 1937, having the advantage of Polaroid polarizing films then newly offered on the market, Öhman (who was unaware of my work on this subject although, shortly afterward, he acknowledged its precedence in point of time [9]) was able to build in a short time a much more elementary filter based on the same principle [10]. His filter isolated a spectral band with an equivalent width of 40 Å and at a fixed wavelength equal to that of chromospheric radiation H_α; this instrument enabled its author to observe and photograph the brightest prominences with a simple telescope.

In 1938, I procured sheets of sufficiently transparent polaroid and built a second monochromatic filter, 36 mm in diameter, which made temporary use of this new type of polarizer

*) Several physicists had used the band spectra given by a single crystalline plate: some, like Mascart [3], Fabry and Perot [4], R. W. Wood [5], used them to eliminate one of the two sodium D-lines; others, like O. Wiener [6], Berek-Rinne, I. G. Priest [7], used them to make complementary colored filters; but they had not studied sets of plates and none of them had produced a monochromatic filter.

and allowed me to isolate at will 4 chromospheric and 2 coronal radiations. It transmits bands with a mean width of 2 Å in the green and 3 Å in the red. Its transparency is 10% in the green and 25% in the red. In July and August 1939 at the Pic du Midi, using this filter, whose wavelength is "tuned" and stabilized by a thermostat, I photographed the chromosphere at the solar limb as well as prominences against a very dark background, using H_α radiation, and, employing radiations 5,303 and 6,374 A, I obtained good monochromatic photographs of the corona, the first of their kind. Publication of these results was delayed by the course of events [11].

Since then, several physicists have constructed filters like that of Öhman; these are increasingly selective but do not achieve the selectivity of the one mentioned above; moreover, they all isolate only H_α radiation.

In 1939 the Zeiss Company of Jena built two filters: the first, 13 mm in diameter, transmits a band of 43 Å mean width with a maximum transparency of 21% [12]; the second, 30 mm in diameter, transmits a band 20 Å wide with a transparency of 30% [12, 13].

In the United States, John Evans constructed a filter 20 mm in diameter which transmits a band with a mean width of 5 Å. This instrument, equipped with a heater, was adapted to a refracting apparatus with an optical system like that of the Lyot coronograph and, since June 1940, has enabled its builder to photograph prominences against a dark background [14].

Edison Pettit then constructed a similar filter provided with a thermostat and has filmed prominences with it since May 1941 [15].

In 1940 and 1941 I obtained two fine samples of calcite and replaced the polaroids temporarily mounted in the 1938 filter by calcite polarizers of the proper type; this increased the instrument's luminosity, doubled its selectivity in the red and reduced

the diffused light, and I could observe the chromosphere against the solar disk and see the corona simultaneously with its green and red radiations. Using this instrument, provided with a radiation separator, I could take three motion picture films simultaneously: the first showing the prominences with the H_a line, the second showing the corona with the 6,374 A line and the third showing the corona with the 5,303 A line.

In 1942, I added another calcite plate to the filter, which reduced the width of the transmitted bands to 1 Å in the green and to 1.5 Å in the red. With this improvement I could observe and film, with H_a radiation, the movements of the chromosphere at the limb, the movements of the filaments on the solar disk, and the evolution of numerous chromospheric flares, and I could bring out their radial velocities with an elliptical polarization analyzer.

My two filters of 1933 and 1938 were described in two very brief notes which seem generally to have escaped the attention of readers; moreover, the other publications cited present the theory of these filters in an incomplete or an inaccurate manner. Therefore, I feel it may be useful to restate in greater detail their principle, their respective properties, construction and the main results which they made possible.

A description of the various solutions tried since 1927 may aid physicists to construct filters suitable for different purposes.

THEORY OF THE BIREFRINGENT FILTER

The Principle. This device makes the light to be filtered pass successively through a series of polarizers P_1, P_2, etc. (fig. 1) whose planes of polarization are parallel. A crystalline plate 1, 2, etc., for example, of quartz, cut parallel to the optical axis of the crystal, is placed between each polarizer and the next one. The faces of these plates are parallel to each other

and perpendicular to the light rays; their optical axes are parallel and form angles of 45° with the planes of polarization of the polarizers. Each plate is twice as thick as the preceding one.

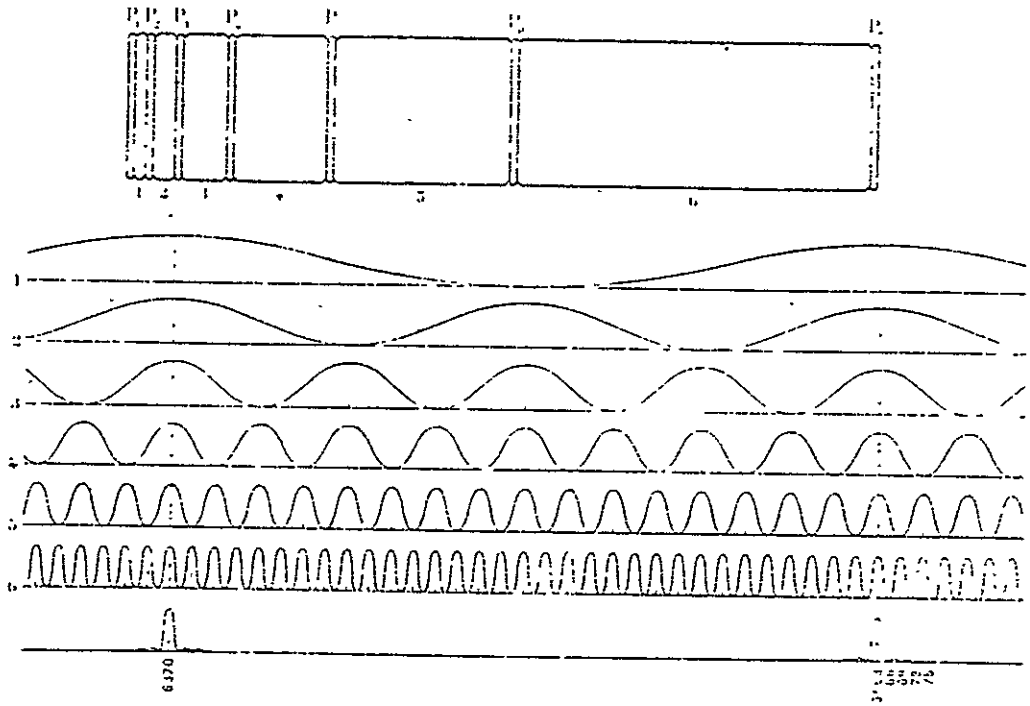


Figure 1

Diagram and principle of the polarizing monochromatic filter.

Light Transmitted in a Normal Direction. Let us calculate the light intensity transmitted by the filter as a function of wavelength, assuming that the incident beam is perpendicular to the plane of the plates and disregarding light losses.

Let A be the amplitude of the light wave which leaves the first polarizer, and let λ be its wavelength, n the number of plates, e the thickness of the thinnest plate, and μ the difference between the ordinary and extraordinary indices of the plates.

Wave A is broken up by the thinnest plate into two vibrations of amplitude $A/\sqrt{2}$, showing a phase difference of $\varphi = \frac{2\pi\mu e}{\lambda}$. The second polarizer makes these two waves parallel and reduces their amplitude to A/2.

We can recombine these two parallel waves at the exit from the second polarizer; the resulting amplitude is $A \cos \frac{\varphi}{2}$ or $A \cos \frac{\pi\mu e}{\lambda}$; the resulting intensity is $A^2 \cos^2 \frac{\pi\mu e}{\lambda}$.

Let us do the same for each plate; the intensity transmitted by the filter equals the product of the intensities transmitted by each plate, or:

$$A'^2 = A^2 \cos^2 \frac{\pi\mu e}{\lambda} \cos^2 \frac{2\pi\mu e}{\lambda} \cos^2 \frac{4\pi\mu e}{\lambda} \ldots\ldots \cos^2 2^{n-1} \frac{\pi\mu e}{\lambda} \qquad (1)$$

Figure 1 shows how a six-plate filter works. The factors of the product, i. e., the transparencies of the six plates for the various values of λ, are represented by the ordinates of curves 1, 2, 3, 4, 5 and 6. These curves may be likened to sine curves whose abscissae scale slowly expands when wavelength increases; each has its maxima twice as close together as the preceding one.

The transparency of the filter is represented by the lower curve, which has, as its ordinate at each point, the product of the corresponding ordinates of the 6 other curves. It passes through some principal maxima equal to unity which are narrow and flanked by some weak secondary maxima whose intensity tends rapidly toward 0. These principal maxima are few in number; indeed the figure shows that the filter transmits one maximum of plate 6 out of 32. Only those maxima common to all plates are transmitted, the others being absorbed by one of the first five plates.

Figure 2 (on Plate I) shows a photo taken between 5,770 A and 6,680 A which gives the band spectra of six quartz plates whose thicknesses are proportional to successive powers of 2. The wavelengths increase from left to right.

Figure 3 (on Plate I) shows the spectra of the light which has passed through plate 6, through plates 6 and 5, through plates 6, 5 and 4, etc. Each plate added to the preceding ones divides the number of bands by two.

The light intensity transmitted by the filter can be calculated numerically by using formula (1), and the calculation can be facilitated by using a second expression, which can be obtained directly by combining the elementary waves at the filter exit only.

After the first plate and the second polarizer, we have two parallel vibrations of amplitude A/2 with a phase difference φ. By the same mechanism, after the second plate and the third polarizer, we have four parallel waves of amplitude A/4 and of phases 0, φ, 2φ, 3φ.

At the filter exit, we have 2^n parallel vibrations of amplitude A/2 and of phases 0, φ, $2\varphi \ldots (2^n - 1)\varphi$.

Their resulting amplitude is given, as in the conventional case of a grating with 2^n grooves, by the formula

$$A' = A \frac{\sin 2^{n-1} \frac{\varphi}{2}}{2^{n-1} \sin \frac{\varphi}{2}} = A \frac{\sin 2^n \frac{\pi \mu e}{\lambda}}{2^{n-1} \sin \frac{\pi \mu e}{\lambda}}.$$

The intensity at the filter exit equals the square of this amplitude, or:

$$A'^2 = A^2 \frac{\sin^2 2^n \frac{\pi \mu e}{\lambda}}{2^{2n} \sin^2 \frac{\pi \mu e}{\lambda}} \qquad (2)$$

According to formulas (1) and (2), the filter transmission passes through a maximum equal to unity for each of the values of λ which make the factor $\frac{\mu e}{\lambda}$ an integer, while it is canceled for the other values of λ which make the factor $\frac{2^n \mu e}{\lambda}$ an integer.

PLATE I

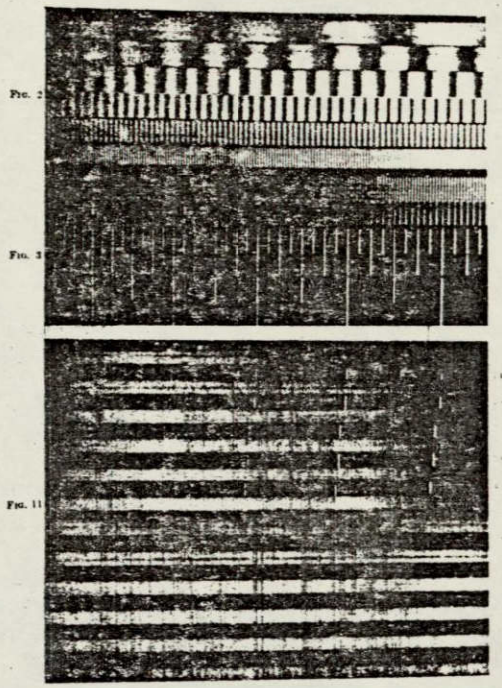


Figure 2. Spectra of the 6 plates in the yellow and red.

Figure 3. Corresponding spectra of plate 6, of plates 6 and 5, plates 6, 5 and 4, etc.

Figure 11. Spectra of the filter at different operating temperatures; agreement of the transmitted bands with solar lines.

PLATE I

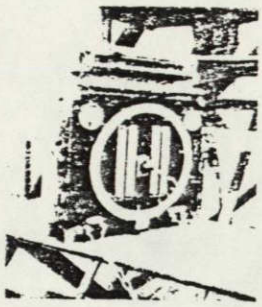


Figure 13. Wheatstone bridge on the observation field.

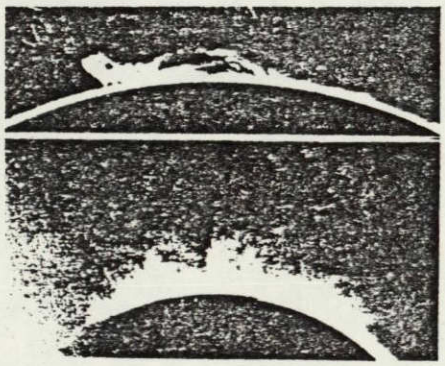


Figure 14. Image taken from one of the first films of prominences obtained with the filter (southeast limb, 12 August 1939, 1126).

Figure 15. The first monochromatic photo of the corona obtained with the filter isolating the green line (west limb, 10 August 1939, 0920).

For the intermediate wavelengths, the transmission passes through a series of secondary maxima, which amount to $(2^n - 2)$ between two principal maxima, but when n is large (e.g., greater than 4), their intensities tend rapidly toward 0 whenever one moves away from a principal maximum. They occur for values of $\frac{\mu e}{\lambda}$ near an integer of the numbers: $1.5/2^n$, $2.5/2^n$, etc., for which the numerator in formula (2) is equal to unity.

In the vicinity of a principal maximum, since $\mu e/\lambda$ is close to an integer, $\sin \frac{\pi \mu e}{\lambda}$ in the denominator of formula (2) can be replaced by the smallest corresponding arc; thus, for the intensities of the successive secondary maxima, we get the approximate values:

$$\frac{1}{1.5\pi^2}, \frac{1}{2.5\pi^2}, \text{ etc. or } \frac{1}{22}, \frac{1}{61}, \frac{1}{121}, \frac{1}{200}, \text{ etc.}$$

With increasing distance from the principal maxima, the ratio of the arc to the sine increases and the intensities diminish a little less rapidly; the figures of this series must be multiplied by a factor which is equal to unity near the principal maxima but which reaches the value $(\frac{\pi}{2})^2$, or 2.47 in the middle of the intervals.

Thus, the spectrum of the light transmitted by the filter is composed chiefly of a small number of bright, narrow bands. The total width of each of these bands is twice the interval between two consecutive zeros, i.e., twice the wavelength variation $\Delta\lambda$, which increases or diminishes the quantity $2^n \frac{\pi e \mu}{\lambda}$ by an integer. The exact calculation can be made with the aid of the dispersion curve of the crystal by deducing from it the curve which gives μ/λ as a function of λ.

Writing that the increase of μ/λ is equal to $\frac{1}{2^n e}$, we get:

$$\Delta\lambda = \frac{\lambda^2}{2^n e \mu} \frac{1}{1 - \frac{\lambda}{\mu} \frac{\partial \mu}{\partial \lambda}} \qquad (3)$$

The first factor contains thickness e of the thinnest plate and the difference μ of its principal indices. The second factor takes into account the dispersion $\frac{\partial\mu}{\partial\lambda}$. In the case of quartz it is 0.923 for line C and 0.90 for line F. For calcite, it is 0.91 for line C and 0.85 for line F.

The spacing of the principal bands is $2^n\Delta\lambda$; it must be calculated by taking the mean values of λ, μ, $\partial\mu/\partial\lambda$ for the spectral region considered.

Formula (3) permits calculation of a filter which isolates a radiation of any given wavelength λ in the center of a band whose mean width differs little from a given value, $\Delta\lambda$. The smallest of the integers for which the closest bands, $2^n\Delta\lambda$ away, can be absorbed with ordinary colored filters will be chosen for n; the multiple of $\frac{\lambda}{\mu}$ nearest the value given by formula (3) will be chosen for the thickness e of the thinnest plate.

An identical result may be obtained by giving the thinnest plate a thickness e which is an uneven multiple of $\lambda/2\mu$, but this plate, which contains an odd number of half-wavelengths, must be placed between crossed polarizers. The other plates, which contain even numbers of half-wavelengths, must remain between parallel polarizers.

The equivalent width of the filter can be derived directly from formula (1). Let us take a band transmitted by the filter for whose center $\mu e/\lambda$ equals an integer K. The interval between K - 1/2 and K + 1/2 corresponds to the separation of the two principal bands; it contains the band K and all the corresponding secondary maxima. Moreover, this interval contains a whole number of periods for each of the n factors in $\cos^2$ of formula (1). It can easily be shown that in this interval the mean value of the product of these n factors is $1/2^n$.

If, therefore, the K band and its secondary maxima are isolated, the equivalent width of the filter, in accordance with the definition given at the beginning of this paper, is equal to

the interval K - 1/2 K + 1/2 divided by 2^n, or to the half-width $\Delta\lambda$ of the isolated K band.

If the polarizers are polaroids which do not polarize the light completely, they transmit stray radiations which increase the filter's equivalent width.

Influence of Temperature. Temperature changes cause both the thickness e and difference μ of its indices to vary for each plate; this modifies the resulting phase difference $\varphi = \frac{2\pi\mu e}{\lambda}$ by the quantity $\Delta\varphi$. For a temperature variation Δt, and assuming λ constant, we have:

$$\frac{\Delta\varphi}{\varphi} = \left(\frac{1}{\mu}\frac{\partial\mu}{\partial t} + \frac{1}{e}\frac{de}{dt}\right)\Delta t, \qquad (4)$$

a quantity having the form A Δt and which is the same for all plates.

This causes a shift in the spectrum of the transmitted bands; their wavelengths vary, causing an additional variation of index μ so that phase difference φ remains constant. Therefore let us write:

$$\frac{\Delta\varphi}{\varphi} = 0 = \left(\frac{1}{\mu}\frac{\partial\mu}{\partial t} + \frac{1}{e}\frac{de}{dt}\right)\Delta t + \frac{1}{\mu}\frac{\partial\mu}{\partial\lambda}\Delta\lambda - \frac{1}{\lambda}\Delta\lambda,$$

whence we get:

$$\frac{\Delta\lambda}{\lambda} = \left(\frac{1}{\mu}\frac{\partial\mu}{\partial t} + \frac{1}{e}\frac{de}{dt}\right)\left(\frac{1}{1 - \frac{\lambda}{\mu}\frac{\partial\mu}{\partial\lambda}}\right)\Delta t, \qquad (5)$$

a quantity having the form A B Δt.

The first factor, A, introduces index μ and its temperature coefficient as well as the coefficient of expansion of the crystal perpendicular to the plane of the plates. The second factor, B, is the dispersion factor of formula (3); we have already given its value.

The calculation made for the D line and for a 1° C rise in temperature gives $\frac{\Delta\varphi}{\varphi} = A = -10^{-4}$ for quartz and $A = -0.62 \times 10^{-4}$ for calcite.

Measurements made on the green line of mercury at 5,461 Å with a quartz plate 60.7 mm thick, between 20° and 71.5°, gave for a 1° temperature increase the more exact value:

$$\frac{\Delta\varphi}{\varphi} = A = -1{,}14.10^{-4},$$

whence one derives

$$\frac{\Delta\lambda}{\lambda} = AB = -1{,}04.10^{-4}.$$

Measurements made with the green line of mercury on a calcite plate 6.625 mm thick between 16.8° and 46.2° gave:

$$\frac{\Delta\lambda}{\lambda} = A = 0{,}646.10^{-4},$$

whence we derive

$$\frac{\Delta\lambda}{\lambda} = AB = 0.56.10^{-4}.$$

All these values are negative, because the relative reduction of index μ exceeds the relative increase of thickness e.

Variable Wavelength Filter. Thus, the wavelength of the filter can be varied merely by changing the temperature of the filter. By this method one of the transmitted bands can be made to coincide exactly with the radiation to be isolated, but the filter cannot be used in a wide spectral region, e.g., between 10° and 60° the bands transmitted by a quartz filter move only 32 Å in the red and 22 Å in the blue.

To explore the whole spectrum, we must change the thicknesses of the plates; we shall see that a slight variation of thickness is enough to give this result.

Let us take a filter, for example, of quartz, which has been built for a wavelength λ_o and which we want to use for another wavelength λ_1. To obtain a transmission maximum of

wavelength λ_1 for each plate, it is sufficient to increase or decrease the plate thickness by an amount less than or equal to $\lambda_1/2\mu$, thus producing a half-wave phase shift; about 40 μ suffice for the whole visible spectrum.

This alteration modifies very slightly the closeness of the bands; in the vicinity of λ_1, the spectrum of the transmitted light is the same as if the filter had been calculated for this wavelength, but as one departs from λ_1, the transmission maxima of the various plates are progressively displaced with respect to each other. The spectrum changes slightly and the intensity of the principal maxima diminishes while new, increasingly intense secondary maxima appear.

Thus, the filter transmits stray radiations whose mean intensity increases more and more rapidly as one departs from λ_1. We shall calculate the approximate values of their intensities in a particularly unfavorable case.

Consider a six-plate filter whose thicknesses are exactly proportional to successive powers of 2; let λ_1 and λ_3 be the wavelengths of two consecutive bands transmitted by this filter. For these bands the thinnest plate produces phase shifts $\varphi'_1 = 2\pi K$ and $\varphi'_3 = 2\pi(K-1)$, K being an integer. Let us suppose that, without changing its temperature, we wish to set it for another wavelength λ_2 so that $\lambda_2 - \lambda_1 = \frac{1}{3}(\lambda_3 - \lambda_1)$. For this wavelength λ_2, filter plates 1, 2, 3, 4, etc..., whose dispersion we assume to be constant in the interval considered, will produce phase shifts:

$$\varphi_1 = 2\pi\left(K - \frac{1}{3}\right), \qquad \varphi_2 = 2\pi\left(2K - \frac{2}{3}\right) = 2\pi\left(2K - 1 + \frac{1}{3}\right),$$

$$\varphi_3 = 2\pi\left(4K - \frac{4}{3}\right) = 2\pi\left(4K - 1 - \frac{1}{3}\right), \quad \varphi_4 = 2\pi\left(8K - \frac{8}{3}\right) = 2\pi\left(8K - 3 + \frac{1}{3}\right) \quad \text{etc.}$$

We want $\cos^2 \varphi/2$ to be unity for all plates; for this, we alternately reduce and increase phase shifts φ_1, φ_2, etc., by $2\pi/3$ and, thus, alternately reduce and increase the thickness of plates 1, 2, etc. by Δe_2 so that $\frac{\mu_2 \Delta e_2}{\lambda_2} = \frac{1}{3}$.

Let λ_4 be the wavelength of one of the two bands close to λ_2 transmitted by the filter thus modified. This band does not have a transparency equal to unity; in fact, this would require the variation in the thickness of the plates to be $\pm\Delta e_4$, so that $\frac{\mu_4 \Delta e_4}{\lambda_4} = \frac{1}{3}$.

The phase shifts introduced by the six plates are not zero but

$$\pm 2\pi \frac{\mu_1}{\lambda_1}(\Delta e_2 - \Delta e_1) = \frac{2}{3}\pi \frac{\mu_1}{\lambda_1}\left(\frac{\lambda_2}{\mu_2} - \frac{\lambda_1}{\mu_1}\right) = \frac{2}{3}\pi \left(\frac{\lambda_2}{\lambda_1}\frac{\mu_1}{\mu_2} - 1\right).$$

Let us assume, for example, that $\lambda_2 = 5,300$ A and $\lambda_4 = 5,900$ A. For quartz, we have $\mu_2 = 0.00920$, $\mu_4 = 0.00911$.

For wavelength λ_4, the phase shifts are $\pm\frac{2\pi}{3} \times 0.108 = 13^\circ$.

The corresponding band has a transparency $(\cos^2 13^\circ/2)^6 = 0.925$. The missing light must be in the secondary maxima, part of which lie between λ_2 and λ_4.

For any wavelength whatever, the phase shifts differ by a quantity $\Delta\varphi$ from those one would have with a perfect filter; $\Delta\varphi$ is canceled out for λ_2, for λ_4 it varies almost linearly and reaches a value of $\pm 13^\circ$.

Let us consider successively plates 1, 2, 3, 4, 5 (fig. 1). Each of their minima must eliminate one of the sixth plate's maxima, which is weakened by the preceding plates and is freely transmitted by the following ones. This absorption is no longer total; in the place of a zero of wavelength λ'_2, the plate considered has a transmission equal to

$$\sin^2 \frac{\pi}{3} 0{,}108 \cdot \frac{\lambda'_2 - \lambda_2}{\lambda_1 - \lambda_2},$$

a quantity of the form A^2P^2 with $A^2 = 0.0128$.

Thus, for plate 1, there is a single secondary maximum located in the middle of the interval for which $P = \frac{1}{2}$; its intensity in relation to that of the principal maximum λ_2 is $I = A^2P^2 = 0.25A^2 = 0.0032$.

For plate 2 we find two maxima partially absorbed by plate 1, located at 1/4 and 3/4 of the interval:

$P = 1/4 \qquad I = A^2P^2 \cos^2 \pi/4 = 0.0313\ A^2 = 0,000401,$
$P = 3/4 \qquad I = A^2P^2 \cos^2 3\pi/4 = 0.281\ A^2 = 0,0036.$

For plate 3, there are four maxima partly absorbed by plates 1 and 2:

$P = 1/8 \qquad I = A^2P^2 \cos^2 \pi/8 \cos^2 \pi/4 = 0,0066\ A^2 = 0,000084,$
$P = 3/8 \qquad I = A^2P^2 \cos^2 3\pi/8 \cos^2 3\pi/4 = 0,0104\ A^2 = 0,000133,$
$P = 5/8 \qquad I = A^2P^2 \cos^2 5\pi/8 \cos^2 5\pi/4 = 0,0286\ A^2 = 0,000366,$
$P = 7/8 \qquad I = A^2P^2 \cos^2 7\pi/8 \cos^2 7\pi/4 = 0,327\ A^2 = 0,0042.$

For plate 4, we find eight maxima partially absorbed by plates 1, 2 and 3, for which P = 1/16, 3/16, ... 15/16. The corresponding coefficients of A^2 are: 0.0016, 0.0018, 0.0071, 0.0008, 0.0052, 0.0132, 0.0317, 0.360.

A plate of n order thus gives rise to 2^{n-1} secondary maxima whose intensities increase, on the average, when one departs from wavelength λ_2 isolated by the filter. Only the one farthest removed, i.e., that closest to λ_4, is intense. For plates 1, 2, 3, 4 and 5, the respective values of its coefficient are: 0.250, 0.281, 0.327, 0.360, 0.381.

The sum of the coefficients of A^2 for each plate is respectively: 0.25, 0.312, 0.372, 0.420, 0.449; when n is large, it tends toward the value 1/2. For all plates, this sum is 1.76; its product times A gives the total light of the secondary maxima, or 0.022. This figure is only an order of magnitude; we have not taken into account the secondary effects, the chief of which is that the bands of the sixth plate shift in a direction opposite those of the fifth; the relative displacement is thus multiplied by 1.5. Allowance can be made for this by multiplying the coefficient of the fifth plate by 1.5^2, which gives the figure of 3% for the global intensity of the secondary maxima. The bands transmitted by plates 5 and 6 together undergo almost no displacement and there is no need to modify the coefficients of the other plates.

Therefore, if we use colored filters to eliminate the two maxima located on either side of the isolated radiation λ_2 as well as the radiations located beyond them, the stray light due to modification of the filter will not exceed 6% of that of the principal maximum. Actually, it is reduced much more, because the colored filters do not have distinct absorption limits; they weaken the secondary maxima farthest from λ_2, which are, at the same time, the most intense.

Light Transmitted in an Oblique Direction. The phase shifts produced by the filter's crystalline plates vary whenever the incident light beam deviates from the normal to their faces. First, let us calculate these variations in the general case of a biaxial crystal, from which we can easily deduce the case of a uniaxial crystal.

Let n_1, n_2, n_3 be the smallest, intermediate and greatest principal indices of a plate of thickness e, assumed to be cut perpendicularly to the principal direction for which the waves have index n_2; since n_2 represents the intermediate index, the faces of the plate are parallel to the crystal's optical axes.

First case. The plane of incidence is parallel to the principal directions of indices n_1 and n_2. Let XX' and YY' (fig. 4) be the faces of the plate; the section of the wave surface in the crystal through the plane of incidence is represented by a circle of radius $OA_3 = 1/n_3$ for vibrations normal to this plane and by an ellipse with semi-axes $OA_2 = 1/n_2$ and $OA_1 = 1/n_1$ for vibrations parallel to it, while the section of the wave surface in air is a circle of radius OB = 1.

Let OD be the normal to the plate at point O and i the angle of incidence; the Huygens construction gives the two rays refracted in the plate, OE_3 and OE_1. Let E_1H be the perpendicular dropped in air from the one ray's point of emergence to the other. The birefringence introduced by the plate is:

β_1 = optical path OE_3 + length $E_3'H$ — optical path OE_1. The vibrations normal to the plane of incidence are refracted along OE_3 with path $OE_3 = n_3 e/\cos r_3$ and $\sin i = n_3 \sin r_3$.

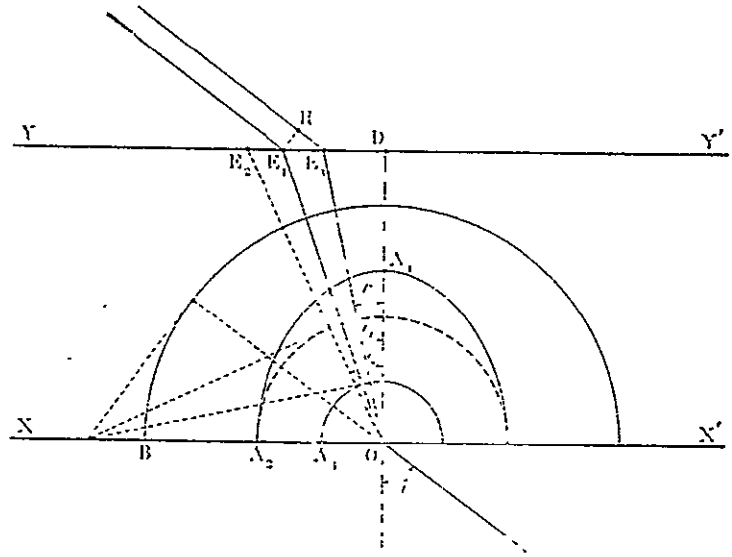


Figure 4

Birefringence of a biaxial plate as a function of angle of incidence.

The vibrations contained in the plane of incidence are refracted along OE_1. If the plate were isotropic with index n_2, the vibrations would be refracted along OE_2, with $OE_2 = n_2 e/\cos r_2$ and $\sin i = n_2 \sin r_2$. The ellipse is a projection of the circle of radius $1/n_2$ expanded perpendicularly to the plane of the plate in the ratio n_2/n_1; hence, in the real case the velocity component perpendicular to the plane of the plate is greater in the ratio n_2/n_1, which gives:

$$\text{optical path } OE_1 = \frac{n_2 e}{\cos r_2}\frac{n_1}{n_2} = \frac{n_1 e}{\cos r_2}.$$

We have, moreover, $E_3H = E_2E_1 \sin i = e(\tan r_1 - \tan r_3) \sin i$. A process like this gives

$$\operatorname{tg} r_1 = \frac{n_2}{n_1}\operatorname{tg} r_2, \text{ whence } E_3H = e\left(\frac{n_1}{n_2}\operatorname{tg} r_2 - \operatorname{tg} r_3\right)\sin i.$$

and the birefringence in plane 1-2 is:

$$\beta_1 = e\left[\frac{n_3}{\cos r_3} + \left(\frac{n_1}{n_2}\operatorname{tg} r_2 - \operatorname{tg} r_3\right)\sin i - \frac{n_1}{\cos r_2}\right].$$

Introduction of the relations $\sin i = n_2 \sin r_2 = n_3 \sin r_3$ simplifies the formula and yields:

$$\beta_1 = e(n_3 \cos r_3 - n_1 \cos r_2). \tag{6}$$

Second case. The plane of incidence is parallel to the principal directions of indices n_3 and n_2.

The calculation can be made as above: the result may be obtained directly by transposing indices 1 and 3 in the preceding formula and by changing the sign of the result, which gives:

$$\beta_3 = e(n_3 \cos r_2 - n_1 \cos r_1). \tag{7}$$

If the incident beam diverges only slightly from the normal to the plate, the cosines can be replaced by the first two terms of their series expansions and we obtain:

$$\beta_1 = e\left[n_3\left(1 - \frac{r_3^2}{2}\right) - n_1\left(1 - \frac{r_2^2}{2}\right)\right].$$

It is also possible to equate the arc with its sine and write $i = nr$, which gives us:

$$\beta_1 = e(n_3 - n_1) + \frac{i^2}{2}\left(\frac{n_1}{n_2^2} - \frac{1}{n_3}\right).$$

or:

$$\beta_1 = e(n_3 - n_1)\left(1 + \frac{i^2}{2}\frac{n_1 n_3 - n_2^2}{(n_3 - n_1)n_2^2 n_3}\right), \tag{8}$$

a quantity having the form

$$e(n_3 - n_1)(1 + K_1 i^2).$$

The first two factors represent the birefringence of the plate for normal incidence; the third factor gives its relative variation. Analogously we have:

$$\beta_2 = -e(n_1 - n_3)\left(1 + \frac{i^2}{2}\frac{n_1 n_3 - n_2^2}{(n_1 - n_3)n_2^2 n_1}\right). \tag{9}$$

a quantity with the form

$$e(n_3 - n_1)(1 + K_3 i^2).$$

Coefficients K_1 and K_3 are opposite in sign, with $K_1/K_3 = -n_1/n_3$. The variations of the birefringence are maxima and have opposite directions in the two rectangular planes containing, respectively, the vibrations of indices n_1 and n_3. Examined in the focal plane of a lens, the isochromatic lines of the second degree are hyperbolas whose asymptotes form an angle α with plane 1 so that

$$\operatorname{tg}^2 \alpha = -\frac{K_1}{K_3} = \frac{n_1}{n_3}.$$

The case of quartz. Let us say that $n_1 = n_2 = n_o$ (ordinary index) and $n_3 = n_e$ (extraordinary index). Formulas (8) and (9) give us:

$$K_1 = +1/2n_o n_e \quad \text{and} \quad K_3 = -1/2n_o^2$$

The birefringence diminishes in the plane containing the optical axis and increases in the perpendicular plane. For the D line, we have

$$n_o = 1.54424 \qquad n_e = 1.55335$$

whence:

$$K_1 = +0.20844 \qquad K_3 = -0.20967$$

The case of calcite. For the D line, $n_1 = n_e = 1.48643$; $n_2 = n_3 = n_o = 1.65837$. In the same way we find $K_1 = -1/2n_o^2 = -0.18180$; $K_3 = +1/2n_o n_e = -0.20283$.

As in the case of quartz, the birefringence decreases in the plane which contains the optical axis and increases in the perpendicular plane.

The relative wavelength variation corresponding to an angle of incidence i is obtained by multiplying the relative

birefringence variation by the dispersion factor of formula (3). We get:

$$\frac{\Delta\lambda}{\lambda} = Ki^2 \left(\frac{1}{1 - \frac{\lambda}{\mu}\frac{\partial\mu}{\partial\lambda}} \right). \qquad (10)$$

Usable field. Let us assume, for example, that the maximum admissible relative wavelength variation is $\pm 1/10,000$; we will have the inequality:

$$\frac{1}{10\,000} > Ki^2 \left(\frac{1}{1 - \frac{\lambda}{\mu}\frac{\partial\mu}{\partial\lambda}} \right).$$

which, for quartz, gives $i < 2.29 \times 10^{-2}$. The field is limited by a nearly equilateral hyperbola in which a $2^\circ 38'$ square can be inscribed.

For calcite, we find $i_1 < 2.50 \times 10^{-2}$ and $i_3 < 2.37 \times 10^{-2}$, giving a field limited by a hyperbola in which a rectangle $2^\circ 52'$ by $2^\circ 43'$ can be inscribed.

A wider field could be obtained by choosing a biaxial crystal in which the quantity $(n_1 n_3 - n_2^2)$ is as low as possible.

Tartaric acid, for example, has indices of $n_1 = 1.495$, $n_2 = 1.535$, $n_3 = 1.604$ and coefficients of $K_1 = +0.055$ and $K_3 = -0.059$; the field would be limited by a hyperbola in which a rectangle of about $5^\circ 10' \times 5^\circ$ could be inscribed.

Composite plates. A very wide field can be obtained by means of the following three arrangements:

1. The first of these arrangements consists in replacing one crystalline plate of thickness e by two identical plates of thicknesses e/2, superposed and oriented in their plane so that their principal directions of the same index are perpendicular. In this relative position, their birefringences would be subtracted if no particular precaution were taken. To cause them to be added, a half-wave plate whose axis is at 45° to the principal directions of the two plates is inserted between them.

This half-wave plate gives a 90° rotation to the planes of the two vibrations transmitted by the first plate; thus, each of them passes through the second plate with the same index as the first. Thus, the birefringences are added, while their variations, which have opposite signs, are largely compensated. In the two principal planes of indices 1 and 3, the relative variation of the birefringence is the same and we have $\beta = \beta_o(1 + Ki^2)$, with

$$K = \frac{K_1 + K_3}{2} = \frac{n_2^2 - n_1 n_3}{4 n_1 n_2^2 n_3} \tag{11}$$

and, in the case of a uniaxial crystal:

$$K = \frac{n_0 - n_e}{4 n_e n_0^2}. \tag{12}$$

The isochromatic surfaces, being second-degree curves, are circles and these circles have much larger diameters than the axes of the hyperbolas given by a simple plate of the same crystal.

For quartz, for example, which has a weak birefringence, $K = -0.0006$; for a relative wavelength variation of $+1/10,000$ in the center and $-1/10,000$ at the edge, this would give a circular field 68° in diameter, if the approximations which we have made were still valid for angles of 34°. The field is 26 times larger than that of a simple plate.

For calcite, which is much more birefringent than quartz, $K = +0.0105$; the circular field is 17° in diameter, or 6 times larger than the field of a simple plate.

This arrangement has the drawback of being applicable only in a restricted spectral region. Indeed, if the wavelength varies by diverging from the value for which the half-wave plate has been calculated, the system of circular fringes produced at infinity by the composite plate is weakened; while its minima cease to be zero, the intensity of its maxima decreases. The transmission of the filter decreases, while it passes an

increasingly large quantity of light outside of the principal maxima; also, the hyperbolic fringes given by a simple plate of thickness e appear at infinity. These fringes, much more closely packed, furrow the field and may make it hard to see details clearly.

Let $\pi \pm a$ be the phase shift introduced by the half-wave plate. If the plate is placed in parallel light, the contrast of these fringes, in the middle of the field, is $\tan^2(a/2)$; a relative wavelength variation of 6% is therefore enough to give them a contrast of 1%.

If the plate is placed, not in parallel light, but sufficiently close to the image of the object to be observed, these parasitic fringes are not in focus; they remain invisible and a greater wavelength variation can be tolerated; on the other hand, plate defects may interfere with visibility of details if they are too near the image.

2. The second arrangement consists in replacing the simple crystalline plate by two plates cut from different crystals whose coefficients K have opposite signs, e.g., a positive uniaxial crystal like quartz and a negative uniaxial one like calcite.

These two plates are simply superposed in such a way that their directions corrsponding to the lowest index are parallel; the optical axes of quartz and calcite, for example, must be perpendicular. The birefringences of the two plates are added without the insertion of the half-wave plate.

Let β'_o be the birefringence of the first plate under normal incidence and K'_1 and K'_3 its coefficients of variation: let β''_o, K''_1 and K''_3 be the corresponding quantities of the second plate.

The birefringence of the double plate, in the incidence plane of the low indices, is:

$$\beta_1 = \beta'_0(1 + K'_1 i^2) + \beta''_0(1 + K''_1 i^2) = (\beta'_0 + \beta''_0)\left(1 + \frac{\beta'_0 K'_1 + \beta''_0 K''_1}{\beta'_0 + \beta''_0} i^2\right),$$

a quantity with the form

$$\beta = \beta_0(1 + K_1 i^2) ;$$

and, in the incidence plane of the high indices:

$$\beta_3 = (\beta'_0 + \beta''_0)\left(1 + \frac{\beta'_0 K'_3 + \beta''_0 K''_3}{\beta'_0 + \beta''_0} i^2\right),$$

a quantity having the form

$$\beta = \beta_0(1 + K_3 i^2).$$

For the isochromatic lines to be circular, it is necessary (and sufficient) that $K_1 = K_3 = K$ (coefficient of the composite plate).

The preceding formulas give us

$$\frac{\beta''_0}{\beta'_0} = -\frac{K'_1 - K'_3}{K''_1 - K''_3} \tag{13}$$

and

$$K = \frac{K''_1(K'_1 - K'_3) - K'_1(K''_1 - K''_3)}{(K'_1 - K'_3) - (K''_1 - K''_3)} = \frac{K'_1 K''_3 - K''_1 K'_3}{K'_1 - K'_3 - K''_1 + K''_3}. \tag{14}$$

The thicknesses e' and e'' of the two component plates are related by the two equations (15) and (16). The first is obtained directly:

$$\frac{e''}{e'} = \frac{\beta''_0}{\beta'_0}\frac{\mu'}{\mu''} = -\frac{K'_1 - K'_3}{K''_1 - K''_3}\frac{\mu'}{\mu''}. \tag{15}$$

Furthermore, let $\Delta\lambda$ be the spacing desired for the bands of the double plate. Spacings $\Delta'\lambda$ and $\Delta''\lambda$ of the bands of the component plates are related by the expressions

$$\Delta'\lambda = \frac{\lambda^2}{e'\mu'\left(1 - \frac{\lambda}{\mu'}\frac{\partial\mu'}{\partial\lambda}\right)} \quad \text{and} \quad \Delta''\lambda = \frac{\lambda^2}{e''\mu''\left(1 - \frac{\lambda}{\mu''}\frac{\partial\mu''}{\partial\lambda}\right)}.$$

We have, moreover: $\frac{1}{\Delta\lambda} = \frac{1}{\Delta'\lambda} + \frac{1}{\Delta''\lambda}$,

whence $$e'\mu'\left(1 - \frac{\lambda}{\mu'}\frac{\partial\mu'}{\partial\lambda}\right) + e''\mu''\left(1 - \frac{\lambda}{\mu''}\frac{\partial\mu''}{\partial\lambda}\right) = \frac{\lambda^2}{\Delta\lambda}. \tag{16}$$

Relations (15) and (16) enable us to calculate thicknesses e' and e''. For a double plate of calcite and quartz and for the D line, we find:

$$K = + 0{,}0052 \qquad \frac{e''}{e'} = 1{,}09. \qquad \frac{e'}{\epsilon'} = 0{,}0577.$$

The spacing of the bands of the composite plate is equal to that of the bands of a quartz plate 2.10 e' thick near 7,000 A, 2.14 e' thick near the D line and 2.21 e' thick near 4,000 A.

Accepting a relative wavelength variation of $\pm 1/10,000$, the field is circular and has a diameter of 25°; it is 9 times as large as that of a simple quartz plate.

This composite plate gives a field that is not so large as that of the double quartz plate, but it is twice as thin and can be used in a very wide spectral region.

3. A third arrangement consists in superposing a plate cut from one crystal and two plates with crossed axes cut from another crystal whose coefficient K is opposite in sign to that of the first. A quartz plate and two calcite plates cut parallel to the optical axis might be superposed, for example; the two calcite plates would be oriented so that the axis of one would be parallel to that of the quartz and the axis of the other perpendicular to it.

By varying the thicknesses of the three plates, the desired birefringence can be given to the composite plate and its two variation coefficients canceled out.

Let us make the following assumptions:

β'_0 is the birefringence of the first plate under normal incidence and K'_1 and K'_3 are its variation coefficients;

β''_0, K''_1 and K''_3 are the corresponding quantities of the second plate;

β'''_0, K'''_1 and K'''_3 are the corresponding quantities of the third plate.

A calculation like the foregoing gives us the two variation coefficients K_1 and K_3 of the composite plate:

$$K_1 = \frac{\beta'_0 K'_1 - \beta''_0 K''_3 + \beta'''_0 K''_1}{\beta'_0 - \beta''_0 + \beta'''_0},$$
$$K_3 = \frac{\beta'_0 K'_3 - \beta''_0 K''_1 + \beta'''_0 K''_3}{\beta'_0 - \beta''_0 + \beta'''_0}.$$

By canceling out the two numerators, we obtain a system of two equations which gives us

$$\frac{\beta''_0}{\beta'_0} = \frac{K'_1 K''_3 - K'_3 K'_1}{K''^2_3 - K''^2_1}$$

and

$$\frac{\beta'''_0}{\beta'_0} = \frac{K'_1 K''_1 - K'_3 K''_3}{K''^2_3 - K''^2_1}.$$

As before, we can calculate the thicknesses of each of the two plates of the second crystal in relation to that of the plate of the first, as well as the spacing of the spectral bands of a given composite plate; for the thickness of the second crystal, we must take the difference of the thicknesses of the two plates cut from it.

In the case of quartz and calcite, we find, for the birefringences of the calcite plates in relation to that of the quartz plate:

$$\frac{\beta''_0}{\beta'_0} = 0{,}514 \qquad \text{and} \qquad \frac{\beta'''_0}{\beta'_0} = 0{,}573,$$

which enables us to calculate their thicknesses:

$$\frac{e''}{e'} = 0{,}0272, \qquad \frac{e'''}{e'} = 0{,}0319.$$

The spacing of the bands of the composite plate is equal to that of a quartz plate 1.089 e' thick near 7,000 A, 1.091 e' thick near the D line and 1.097 e' thick near 4,000 A. This composite plate is as thick as the double quartz plate, but it can be used in a much wider spectral region and it gives even

lower birefringence variations as a function of incidence since such variations are reduced to fourth-order terms near the wavelength for which the plate is calculated.

FIRST FILTER

The preceding calculations were worked out and checked on crystalline plates between 1927 and 1933. They were presented in a sealed letter delivered to the Academy of Science on 29 May 1933. Shortly thereafter, the principal results of those calculations were published in the note already cited [8].

Construction of a first filter based on these results was begun in 1933. This instrument was to have a wide field and was to make possible isolation of a narrow spectral band; its wavelength was to be adjustable from the start of the ultraviolet to the near infrared. In the green, near line 5,303 A, this band was to have a mean width of 1A and was to be located about 500 A from nearby bands which a colored filter would easily isolate.

To produce this result, the filter was to comprise 10 polarizers and 9 crystalline plates with sides 25 mm long; their birefringences could be varied according to the principle given on page 16 of this paper since the three thickest plates were to be formed of two different crystals.

Thick crystalline plates. Certain crystals in combination with calcite will give a high, constant birefringence in a very wide field and can be very thin. Unfortunately I was unable to obtain sufficiently large homogeneous samples of such crystal. Consequently, I adopted for the three most strongly birefringent plates the calcite-and-quartz combination; this is thick but can be made easily and is perfectly transparent in the whole spectral range considered.

Three composite plates were cut in 1934; they are shown in figure 5.

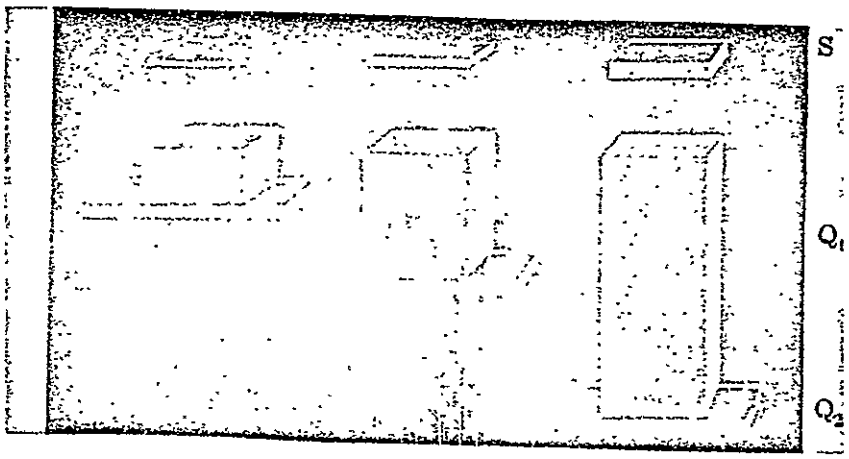


Figure 5.

The three composite plates of the first filter.

The thickest is formed of a calcite plate S, 3.635 mm thick, and two slightly prismatic quartz plates Q_1 and Q_2. Together they form a plate with parallel faces, variable in thickness between 63.46 and 63.76 mm. The next plate, No. 8, less birefringent by a factor of 2, is formed of a calcite plate 1.818 mm thick and a quartz plate for a total variable thickness of 31.66 to 31.96 mm. Plate No. 7 likewise consists of an 0.909-mm calcite plate and a quartz plate; this composite plate is variable from 15.75 to 16.054 mm. The Q_2 prisms are identical for the three plates.

The relative orientation of the S and Q plates requires exact adjustment; this can be effected easily by watching the appearance of the isochromatic lines at infinity given by the ensemble, with the green line of mercury, for example. One rotation of plate S around its optical axis or around an axis normal to the latter and parallel to its plane displaces the fringe system without modifying its general form and allows its center to be brought into the longitudinal axis of plate Q. One rotation of plate S around the longitudinal axis of plate Q, on the other hand, modifies the form of the fringe system without displacing its center of symmetry.

When the angle of the optical axes of the calcite and quartz approach 90°, the fringes, which are hyperbolas at

first, are enlarged and changed into a system of parallel straight lines, then into ellipses whose axes are parallel to the bisectrices of the optical axes. The eccentricity of the ellipses decreases, and when the optical axes are rectangular, it is canceled, providing the quartz and calcite thickness ratio has exactly the value desired, otherwise it passes through a minimum and the long axis of the ellipse is then parallel to the optical axis of the plate that is too thick.

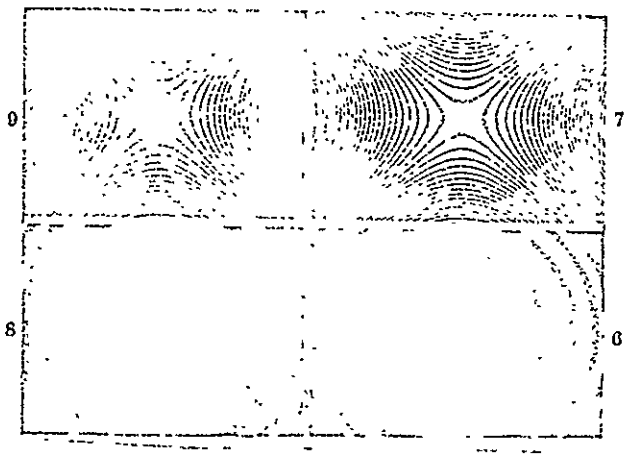
Figure 6 shows a photo, reduced 2x, of the isochromatic lines given by the quartz of the thickest plate (63.6 mm thick). They were photographed with the green line of mercury, $\lambda = 5,460$ A, in the focal plane of a lens with a focal length of 115 mm. The optical axis of the quartz is vertical; the lowest index corresponds to the horizontal direction. The lines are nearly equilateral hyperbolas. The degrees marked at the bottom permit an evaluation of the usable field, about 2.5°.

Figure 7 shows the isochromatic lines of the corresponding calcite plate, which is 3.63 mm thick. They were photographed to the same scale with the green line of mercury. The optical axis is horizontal; as for the quartz, the lowest index corresponds to the horizontal vibrations. The hyperbolas are no longer equilateral; their asymptotes form, with this direction, α angles such as $\tan^2\alpha = n_1/n_3 = n_e/n_o$, whence $\alpha = 43^{\circ}25'$. Also, the differences of coefficients K_1 and K_3 may be seen.

Figure 8 shows, on the same scale, the isochromatic lines obtained with the green line of mercury by superposing the quartz and calcite, oriented as in figures 6 and 7. The birefringence is greater than twice that of the quartz alone, yet the diameter of the isochromatic lines is four times as large. They are slightly elliptic and the long axis of the ellipses is parallel to the optical axis of the quartz.

If the wavelength is changed, the ellipticity varies because the dispersion of the calcite is greater than that of

the quartz. When the wavelength diminishes, the ellipses are shortened along the axis of the quartz, as shown in figure 9, which was taken under the same conditions with the blue line of mercury, $\lambda = 4,358$ A.



Figures 6 and 7.

Isochromatic lines of the quartz plate and the calcite plate (thicknesses 65.8 and 3.63 mm; mercury line 5,460).

Figures 8 and 9

Isochromatic lines of the 2 plates superposed (mercury lines 5,460 and 4,358).

Thin crystalline plates. To give a very wide field, the next plate (No. 6) should consist of a 0.455-mm calcite and a quartz of 7.802 to 8.102 mm. The cutting of very thin calcite plates presents difficulties which complicate the production of plate No. 6 and those that follow. These plates can be made only of two slightly prismatic quartz elements.

The total thickness of plate No. 6 can be calculated by replacing the 0.455-mm calcite by the quartz thickness which gives a band spectrum with the same spacing. This thickness is greater in a ratio which varies with the spectral region considered; this ratio equals 18.9 in the infrared from 14,000 to 10,000 A, 19 to 19.5 in the red, 19.8 in the green and 20 to

20.9 in the ultraviolet. By adopting the extreme figures 18.9 and 20.9, we find minimum and maximum thicknesses of 16.402 and 17.611 mm for plate No. 6; the thicknesses of plates 5, 4, etc... can easily be derived from these thicknesses.

The field of the filter is limited chiefly by plate 6. For 5,303 A, this plate contains 296 wavelengths; a relative birefringence variation of 1/2,960 causes the transmission of this plate to pass from unity to $\cos^2 \pi/10$, or 90.5%. Similarly, plate 5 transmits $\cos^2 \pi/20$, or 97.5%, etc. It finally passes 87%.

This weakening corresponds to an angle of incidence i so that $K_3 i^2 = 1/2,960$, whence i = 0.04 radian, or a field with a diameter of $4^\circ 36'$. Plates 9, 8 and 7 together would give a larger field in the relation $\sqrt{K_3/8K} = 2.2$, or 10° of diameter.

In order to have the maximum field, I studied another solution for plates 6, 5, and 4, viz., to take a calcite thick enough so that it could be cut easily and to combine it with two quartz plates, one of which has its axis parallel to that of the calcite and the other with its axis perpendicular to it. The thicknesses of the two quartz elements were figured in such a way that the fringes would be circular and that these plates together would give a band spectrum whose maxima would have the desired spacing. The sum of their thicknesses would have to be almost the same as if they had parallel axes; the dimensions of the isochromatic lines would remain the same, and only the birefringence would decrease.

Unfortunately, the dispersion of such a combination differs greatly from that of double quartz-and-calcite plates. This increases the stray light as soon as one departs from the wavelength for which the filter is set. To use this arrangement in widely different spectral regions, the thickness of one of the quartz plates would have to be made variable by as much as twice the thickness of the calcite, depending on the spectral region used.

General arrangement of the instrument. The filter was to comprise 10 Glazebrook-type polarizers, measuring 20 mm to a side and 40 mm thick, 3 calcite-and-quartz plates with total thicknesses of 67.3, 33.2 and 16.9 mm, variable by ±0.15 mm, and 6 quartz plates 17, 8.5, 4.25, 2.125, 1.062 and 0.531 mm thick and variable by ±3% ±120 μ (the thickness which corresponds to a half-wave for the 2μ radiation), or ±0.63 mm for the thickest and ±0.15 mm for the thinnest plate.

The total length of the polarizers was therefore to be 400 mm; that of the plates about 153 mm; consequently, the total thickness of all the optical components was to be about 553 mm.

The following arrangement was adopted to keep this great thickness from reducing the field of the instrument: The optical components are divided into 3 groups (fig. 10).

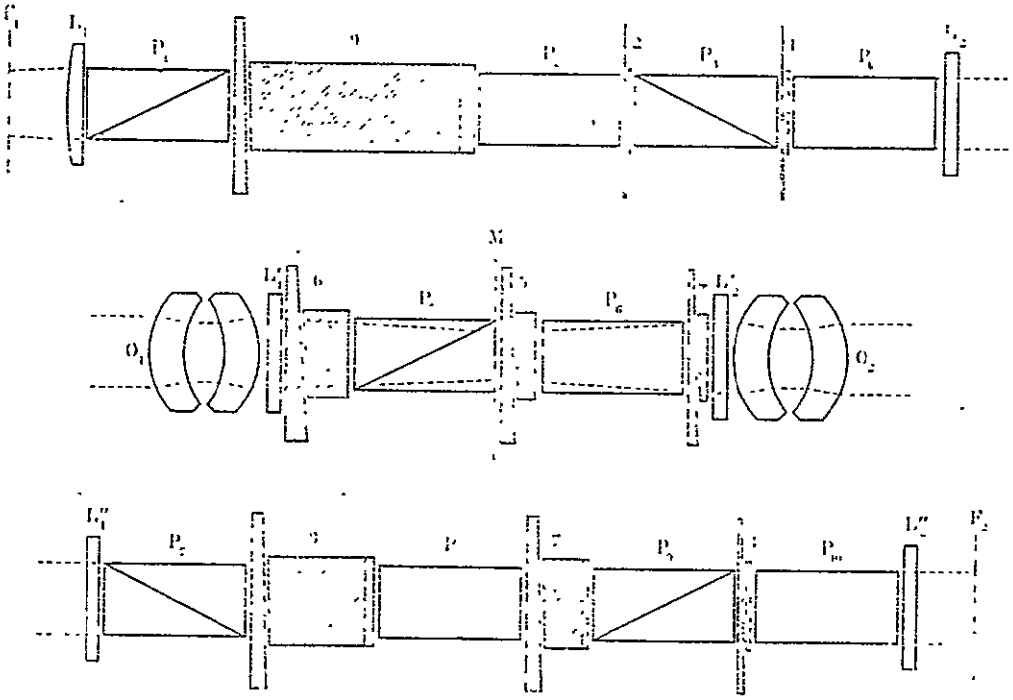


Figure 10

Diagram of the first filter.

The first group includes 4 polarizers P_1, P_2, P_3, P_4 and plates 9, 2, and 1; its overall length is 235 mm;

The second group includes 2 polarizers P_5, P_6 and plates 6, 5, and 4; total length: 111 mm;

The third group includes 4 polarizers P_7, P_8, P_9, P_{10} and plates 8, 7, and 3; its overall length: 217 mm.

Each of these groups is contained in a glass tube closed by two end glasses L_1 and L_2 and filled with a liquid whose index is close to 1.5. This liquid almost entirely eliminates light losses by reflection; moreover, it improves the images by nullifying the wave surface alterations produced by surface defects of the optical components. If a liquid such as carbon tetrachloride, which has no absorption band in the near infrared, is chosen, the filter can be used to 2 microns.

An even number of polarizers is used in each group; they are crossed two by two to avoid the astigmatism which they would otherwise introduce on non-parallel light beams.

The three tubes are placed end to end behind a telescope. The first tube's entry face, located behind focal plane F_1 of the telescope, is a plano-convex lens L_1 with a focal length of about 220 mm in air which images the telescope lens in plane M, near the middle of the second tube.

Between the first and second tube is a lens O_1 with a focal length of about 180 mm in air which forms, at infinity, an image of focal plane F_1 of the telescope. Between tubes 2 and 3 is an identical lens O_2 which forms, outside the instrument in F_2, 17 mm from the exit face, a second image of the telescope's focal plane; this is the final monochromatic image.

The lenses can be worked out so that they correct the curvature of the field introduced by plano-convex lens L_1.

The beam opening is limited by the polarizers, whereas the field is limited chiefly by the birefringence variations of the thickest quartz plates (nos. 6, 5, and 4) as a function of the inclination of the rays.

The rays which are parallel to the filter's axis and which pass through the edges of the polarizers of the first tube are generatrices of a cylinder with a square base, 20 mm x 20 mm. Lens O_1 changes this cylinder into a cone whose vertex is 265 mm from this lens. The section of this cone through plane M, located in the middle of the second tube, 73 mm from O_1, is a square 14.5 mm x 14.5 mm. If we circumscribe this section by a diaphragm with 14.5 mm sides, the field of full light covers the whole entry face of the first polarizer. If we increase this diaphragm up to 20 mm, the field of full light decreases and is finally reduced to the central point.

As we saw on page 34, the angular field of plates 6, 5, and 4 together is a square with sides of 0.08 radian in air, or 0.053 radian in liquid. On the entry face of the first polarizer, it corresponds to a square of 0.053 x 265 = 14 mm per side; hence this face does not limit the field. The same argument is applicable to the exit face of the last polarizer.

Under these conditions, it is advantageous to increase the focal lengths of lenses O_1 and O_2 by a few centimeters. Then the spots can be avoided which would be produced by defects of the optical components located near the entry and exit focal planes; at the same time, the useful aperture of the instrument is increased slightly.

Observed from the entry face of the first polarizer, the 14.5 mm diaphragm which would be placed in the center of the filter would subtend an angle of 14.5/265 = 0.055 radian. In air, this angle becomes 0.055 x 1.5, or 0.082 radian.

The filter, placed behind a telescope with a 12-cm aperture and a 1.50-m focal length, would give a field of 0.0097

radian, or 33' and would permit observation of the whole sun.

Adjustments. In order to isolate a radiation of any given wavelength λ_o, it is necessary to:

1. Give each filter plate a thickness for which one of its transmission maxima has λ_o as its wavelength;

2. Choose for each of the plates 8, 7, etc., of decreasing thickness, that value for which, near λ_o, the quotient of the spacing of its maxima divided by the spacing of the maxima of plate 9 (the thickest) is closest to 2, 2^2, 2^3, etc.

This result could be obtained by trial-and-error, but it is much better to prepare a table which gives, for a given temperature and for each wavelength, the thickness which must be given to each of the 9 plates. This thickness is shown by a scale marked on the edge of plate Q_2; it can be varied by sliding this plate in its plane with the aid of a ratchet. The ratchets can be actuated from outside by metal rods which traverse the end faces through ground joints.

The three tubes are enclosed in a thermostat which keeps temperature constant to within one-tenth of a degree.

SECOND FILTER

Construction of the first filter was interrupted in 1934 because of lack of the calcite needed for cutting the polarizers. Some years later, polaroid sheets appeared and quickly became a commercial product.

With this new, very thin type of polarizer I could reduce the filter length considerably and thus simplify its optical system while preserving an adequate, although weaker, field.

Furthermore, to study the sun, I did not have to isolate every wavelength (as could be done with the first filter), but

only the principal radiations of the chromosphere and corona; however, in certain cases it was useful to isolate them simultaneously.

These conditions induced me to apply the same principle to a second, simplified filter; its characteristics are as follows:

Characteristics of the second filter. The principal part of this instrument, shown in fig. 1, consists of six quartz plates 1, 2, 3, 4, 5, and 6, each with parallel faces 36 mm square. The optical axes of these plates are parallel to the entry and exit faces and to one of their sides. They are superposed, as their optical axes are parallel; each of them is twice as thick as the preceding one. Polaroid sheets P_1, P_2, etc.... are placed between these plates, before the first one and after the last one; their planes of polarization are parallel to each other and are oriented at 45° to the optical axes of the quartz.

As we have seen, the spectrum of the light transmitted by this ensemble consists of a small number of narrow, regularly distributed bands, whose positions and spacing are determined by the thickness of the thinnest plate; the thickness of all the other plates may be derived from this one by doubling.

I tried to give this thinnest plate a thickness such that few radiations would be transmitted, and that six of them should have wavelengths as close as possible to those of the four main chromospheric, and two main coronal, radiations of the visible spectrum, viz.: for hydrogen: H_α and H_β; for helium: D_3; for magnesium: b_1, and for the corona: the green and red lines 5,302.8 and 6,374.5 A.

Fortunately, the distribution of these six radiations and the dispersion law of quartz agree so well that these six conditions are fulfilled satisfactorily for a thickness of 2.22 mm. With this thickness as a design constant, and knowing that wavelengths decrease by 1.04×10^{-4} per degree of temperature

increase, the temperature of the filter need be varied only slightly to make one of its transmitted radiations coincide exactly with the radiation to be isolated.

The first column of Table I shows the six radiations chosen. Column 2 gives their wavelengths λ. The third column gives the retardation δ/λ introduced by the thinnest plate at a temperature of 30° for each of the six radiations. This retardation, expressed in wavelengths, is almost equal to an integer plus 1/2; perfect equality occurs at a temperature T different from 30°, as given in the fourth column. At this temperature the filter isolates exactly the corresponding radiation, if care is taken to place the thinnest plate (with an uneven number of half-wavelengths) between crossed polarizers. The other plates, double, quadruple, etc. the thickness of the thinnest plate, contain whole numbers of waves and must stay between parallel polarizers.

It can be seen from the Table that the coronal red and green lines 6,374.5 and 5,302.8 A are isolated at very similar temperatures: 38.6° and 40.2°; indeed, the retardations they experience in quartz at 30° are in a ratio almost equal to 38.5/31.5, or 11/9. I measured this ratio; it varies very slightly with temperature: 11/9 + 2/10,000 at 49° 11/9 ÷ 1/10,000 at 0°.

TABLE I

Radiation	λ	$\frac{\delta}{\lambda}(30°)$	T	T'
H_α	6 562.8 A	30,557	10,4°	47,5°
K corona	6 374.5	31,531	38,6°	39,4°
D	5 875,6	34,441	14,9°	14,9°
V corona	5 302.8	38,544	40.2°	39,4°
b	5 183.6	39.527	36.1°	35 1°
H_γ	4 861.3	42.457	21.1°	19.6°

This peculiar property of quartz enabled me to isolate simultaneously the two principal coronal radiations by slightly modifying the thickness of the two thickest plates of the filter. I reduced the sixth plate, 71.144 mm thick, by a thickness corresponding to a yellow wavelength, or 64 μ; the fifth plate, 35.572 mm, by a thickness corresponding to a yellow half-wave, or 32μ and placed it between crossed polarizers. This slight modification did not cause any appreciable change of the filter's properties in the visible spectrum.

The last column of the Table gives the T' temperature, at which the filter, thus modified, functions. T' is equal to T in the yellow, is a little higher in the red, a little lower in the blue, so that the two coronal lines are isolated at exactly the same temperature 39.4°. In addition, chromospheric D_3 and H_β radiations are isolated at closer temperatures than before: 14.9° and 19.6°.

Figure 11 (Plate I) shows five spectra of the light transmitted by the filter at the different T' given in Table I, in the red and yellow at the top, in the green and blue at the bottom. Comparison solar spectra taken with the same slit are inserted between these spectra. On them we recognize the principal solar lines: at the top from right to left, H_α and the two D lines of sodium; at the bottom from right to left, the triplet of magnesium and the H_β line.

Between 6,600 and 4,800 A, the filter transmits 13 narrow bands; the sixth principal radiation from the right is not visible on the figure.

At 15°, the 5th band, of 34.5 order, coincides exactly with the D_3 line of helium (invisible in the solar spectrum).

At 19.6°, the 13th band, of 42.5 order, falls exactly on H_β.

At 35.1°, the 10th band, of 39.5 order, coincides with one of the lines of magnesium, b_1.

At 39.4°, it can be seen that the 2nd and 9th bands, of 31.5 and 38.5 orders, coincide simultaneously with the red and green lines of the corona (invisible in the solar spectrum).

Lastly, at 47.5°, the first band, of 30.5 order, coincides with the H_α line.

One of the radiations transmitted by the filter can be separated at will by adding, after the filter, a group of three thin quartz plates, whose retardations for the D_3 line are: 17, 8.5 and 4.5 waves. The first plate must be placed between parallel polarizers, the second between crossed polarizers. If the third is between crossed polarizers, the ensemble transmits only bands D_3 and H_β; if it is between parallel polarizers, the ensemble transmits bands H_α and 5,302.8 A of the corona. Gelatine or colored filters can be used to separate the remaining radiations, if desired.

Another group of three plates, producing retardations of 18, 9 and 4.5 waves for line 6,374 A, transmit the two coronal lines simultaneously, excluding the other radiations.

It is surprising to see the same instrument satisfy such a large number of independent conditions. Indeed, the probability that it would be possible to isolate six radiations in this way in the interval of 15° to 48° without having the total number of bands transmitted between 6,600 and 4,800 A exceed 13, was about 1/400, and the probability that the temperatures relative to the two coronal lines would not differ by more than 2° was about 1/4,000. If one takes into account the filter's other properties, such as the small temperature deviation between D_3 and H_β and the regular distribution of the H_α, D_3, 5,303 corona and H_β bands which makes it possible to separate them at will with a single group of three plates, the probabilities

are of the order of 1/100,000. The simultaneous realization of such a large number of conditions is therefore partly the result of chance.

The filter also has other properties for which it had not been calculated: e.g., with it one can make the bands of 35.5 and 36.5 orders coincide with coronal line 5,694 and line 5,577 of the night sky and polar aurora.

Construction of the filter. The six plates were cut from the most homogeneous parts of some fine quartz samples. The sixth and fifth, 71.080 and 35.540 mm thick, were formed by two plates superposed of half-thickness; hence, if necessary, their field could be increased by inserting a half-wave plate as explained on pages 24-25.

The birefringence of the plates was then compared directly two by two by an interference method by which the difference in index between the samples and between parts of the same sample could be corrected by retouching the faces of the plates. This brought their thicknesses to the desired values at all points with a tolerance of better than one micron and, thus, the phase shifts introduced by the various plates were accurate to a value of less than 2/100 of a wavelength.

The six quartz plates and the seven polaroids were stacked and cemented with Canada balsam between two end glasses to avoid light losses by reflection and to improve the images.

The ensemble formed a square-based parallelepiped, 150 mm long and 36 mm per side. This parallelepiped was placed in a tube of thick aluminum whose cylindrical outer surface was 64 mm in diameter; its conical inner surface measured 48 mm and 52 mm at the ends.

The space between this tube wall and the optical components was filled completely by four aluminum blocks having one

plane side, facing the optical components, and one conical side, facing the tube's inner surface. By pushing slightly on these blocks, one could reduce the play as desired and insure a good thermal contact between the quartz elements and their mounting.

An even layer of silk-insulated copper wire, bakelite-bonded and covered with an insulating sheath of wool 2 mm thick, was wound around the entire cylindrical outer surface of the tube.

This winding, shown schematically as A in fig. 12, had a resistance of 60 ohms and constituted one of the arms of Wheatstone bridge ABCD. The comparison resistance, B, also 60 ohms, made of manganin, was entirely independent of the temperature. Arms C and D, also of 60 ohms, were identical and were made of constantan.

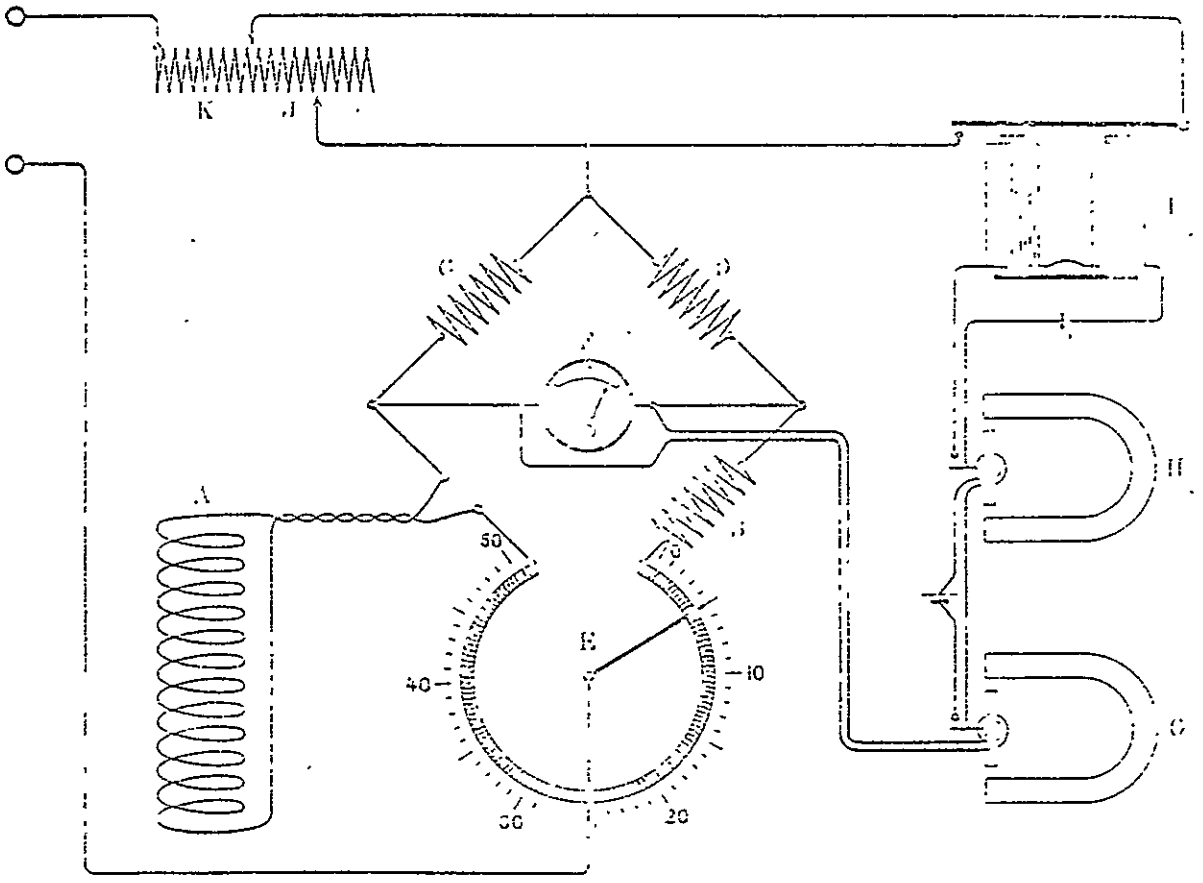


Figure 12

Diagram of Wheatstone bridge and thermostat.

A potentiometer E of 5 ohms and 800 turns of manganin was used to balance the bridge. When this balance was achieved, microammeter F with a range of ±100 μa, connected diagonally, acted as a null detector. A temperature variation of a few hundredths of a degree produced an observable deflection.

The position of the sliding contact which produced the balance depended solely on the resistance of the winding of the filter, hence on its temperature. This sliding contact had a needle which moved over a scale giving the filter temperature to a tenth of a degree, between 0° and 55°.

The mounting had a longitudinal hole in which a thermometer could be placed to calibrate the bridge; it followed the temperature variations of the winding very rapidly. The balancing of the quartz elements, which was checked optically, took longer; for example, if the temperature of the winding were adjusted to 47.5° with the bridge, prominences would appear in 1 minute and shine with full brightness in 3 minutes.

The instrument was provided with a very simple temperature regulator, consisting of three cascade relays. The first relay G had a rotating loop connected to the two terminals of microammeter F located in the diagonal of the bridge; a current of 1 μa sufficed to operate it because it cut off a current of 100 μa only. The current was sent in turn into the rotating loop of a second similar relay H, which sent a current of 0.1 a into the electro-magnet of a third vane-type relay I. The latter short-circuited an adjustable resistor J, cut in on the current feeding the bridge; this current was taken off from a 110-volt d-c line through a resistor K. The auxiliary current of the relays was provided by a battery.

If the temperature was higher than that indicated by the potentiometer needle, the first relay shifted in such a way that contacts were not made; the power current was too weak and temperature decreased. As soon as the latter fell below the

figure shown by the needle, the first relay reversed, all contacts were established and resistor J was short-circuited; this increased the current and raised the temperature.

Thus, the temperature of the winding oscillated from one side to the other of the value indicated by potentiometer E with a period of the order of one second and an amplitude less than one-tenth of a degree. The temperature of the mounting and the quartz elements did not experience these oscillations, but remained remarkably constant during a whole day.

The power consumption of the bridge was always less than 1.5 a.

Figure 13 (Plate II) shows the Wheatstone bridge built in the form of a panel 50 cm square and 16 cm thick. At the bottom is the annular coil of the potentiometer, surrounded by its scale; to the right and left of the needle's axis are the two arms of the bridge; at the top, arranged horizontally, the balance resistor and the two-contact power resistor. To the right is the zero microammeter; to the left, an ammeter which shows power consumption. At the bottom, to the left, the lead-in of the 110-volt supply; at the right, the leads to the filter. Behind the panel are the three relays of the thermostat as well as a fourth relay designed to shunt both the microammeter and the first relay to protect them from surges that might break a conductor in the filter circuit; the rotating loop of this relay is connected to the microammeter terminals.

This thermostatic system is very convenient; the filter, which is not very bulky, can easily be placed behind a coronograph or telescope; a flexwire, 1.2 mm, links it with the Wheatstone bridge; the latter is more bulky but can be placed on the ground or on the observation ladder.

First observations. The filter, completed in June 1939, was adjusted the following month at the Pic du Midi and installed on the coronograph eyepiece tube behind the instrument's lens.

It was preceded by a divergent lens which returned, at infinity, the image of the coronograph's occulting disk and was followed by a lens which formed, in an eyepiece or on a photographic plate, the monochromatic image of this disk surrounded by the corona and the prominences.

The filter was placed in parallel light and each point of the final image was formed by a beam of parallel rays through the filter. Under these conditions, the transmitted wavelengths were the same for all the rays of a beam; on the other hand, the wavelengths varied according to the angle these rays made with that normal to the filter plates, hence from one point of the field to another.

As we have seen, an inclination of $1^{\circ}19'$ produced, in the most unfavorable directions, a relative wavelength variation of 1/10,000, which is admissible; therefore, a field of $2^{\circ}38'$ could be used. The free aperture of the coronograph was usually 120 mm; the corresponding diameter of the beam through the filter was 20 mm; consequently one could observe, without troublesome change of wavelength, a field of $2^{\circ}\ 38' \times 20/120 = 25'$, or 5/6 of the solar diameter.

Observation through the eyepiece disclosed the following phenomena: When all the bands transmitted by the filter were allowed to pass, the sky around the sun showed a dark brown tint, which was due to the polaroids. By increasing the temperature progressively, at 15° the prominences were seen to appear brightly against a dark background in yellow with the helium D_3 radiation; then, at 19.6°, they were illuminated in blue with hydrogen H_{β}; near 17° they showed an intermediate coloration, greenish in cast which turned to blue or yellow on the points which move with positive or negative radial velocities. Near 21° they disappeared.

At 35° the most intense prominences again appeared, in green, with magnesium 5,183 A.

At 39.4° the corona appeared again, simultaneously with its radiations 6,734 and 5,303 A; the latter was dominant and gave the corona a vivid green color; numerous details were visible. Simultaneously, the prominences were already showing in dark red through the polaroids because of the great intensity of H_a.

At 47.5°, finally, the prominences shone bright red with H_a.

When the band corresponding to H_a was isolated, the solar limb could be allowed to exceed the coronograph occulting disk without dazzling the observer; the chromosphere appeared, adding to this edge a multitude of little jets.

The prominences could be photographed even under poor atmospheric conditions, e.g., through cirrus; two good films were obtained thus, on 11 August 1939 from 0836 to 1803 and on 12 August 1939 from 0737 to 1205. Figure 14 (Plate II) shows a picture from the second film: the sun's rim slightly exceeds that of the coronograph's black disk and is fringed by the chromosphere. The watch dial photographed at the left of the film frame gives the time of the view, 1126.

On the other hand, by isolating the bands corresponding to the green or red line, we could photograph the corona with its monochromatic radiations. Figure 15 (Plate II) shows the first monochromatic photograph of the corona thus obtained with the green line on 10 August 1939 at 0920 on the west solar limb. The sun measures 60 mm in diameter on the negative; exposure time was 10 minutes. This image is very different from an ordinary photograph taken in total light; the jets have much greater contrast and appear against a dark background; they show more numerous details, and one of them may be followed up to a height of 8', more than half of the solar radius. The same image, taken shortly after with the red line, shows very different details.

New calcite polarizers. Despite these highly encouraging results, the filter had several drawbacks, attributable to the polaroids.

1. In the green, the polaroids strongly absorbed the light and reduced the filter transparency to 13%; in the blue, they absorbed even more.

2. In the red they were less absorbent and the filter transmitted 21%, but they did not polarize the light completely, and the non-polarized radiations, of various wavelengths, had a higher total intensity than that of the transmitted bands. Therefore, the equivalent width of the filter, which should have been appreciably equal to the half-width of one of these bands, i.e., 3 A, reached 7 A for H_α. The chromosphere was visible on the disk, but could be seen only with difficulty; only the most intense filaments could be observed.

3. Furthermore, the polaroids diffused the light strongly; they did so to such a degree that in order to see the prominences or the chromosphere at the limb well, almost the whole solar disk had to be masked.

4. The thickness of the polaroid sheets was irregular, which usually produced streaks. These defects were lessened considerably by immersing the sheets in Canada balsam of similar index, but they did not disappear entirely and interfered somewhat with the sharpness of the pictures.

In view of all these drawbacks, I decided to return partly to the original project and to replace the polaroids with calcite polarizers as soon as I could obtain the necessary material from which to cut them. Fortunately, in the fall of 1939 the Société l'Optique Scientifique processed a fine rhombohedron of Iceland spar and production of the polarizers was begun at once.

In order to reduce both their thickness and the amount of material needed, these polarizers were composed of birefringent

prisms, one of calcite and one of glass, of the same angle; these were joined by one face with their bases opposed so that their edges and exterior faces were parallel.

The angle of the prisms was $16^{\circ}40'$ and their exterior faces were 36 mm square. Thus the combination formed a parallelepiped 14 mm thick. The optical axis of the calcite was parallel to the exterior faces of the polarizer and to one of their diagonals. The glass prism was cut from crown fluor whose index for the D line, 1.4895, was chosen as close as possible to the extraordinary index of calcite, 1.4864.

We caused a ray of natural light to fall on the polarizer approximately normal to its faces; at the exit we obtained an extraordinary ray which passed through the ensemble almost without deviation, and an ordinary ray, for which the calcite has an index of 1.6583, which was deflected $3^{\circ}7'$ toward the base of the calcite prism.

This arrangement offers two advantages: it is sufficient to substitute the birefringent prism for one of the polaroids of the filter; the undeflected ray vibrates at 45° from the sides of the quartz plates, hence along one of the bisectrices of their optical axes. Also, this plane can be rotated 90° without changing the direction in which the ordinary ray is deflected, simply by turning the polarizer around an axis perpendicular to its edges, thus reversing its faces.

Figure 16 shows the assembly diagram of the filter with the birefringent polarizers: the sign + indicates on which side the optical axis of each of them is ahead of the image plane. The seventh polarizer, placed outside the mounting, is not shown. The filter is operated in the same manner as with the polaroids, except that a diaphragm whose image at infinity circumscribes the field is added in the focal plane of the lens which precedes the filter. Several images of this diaphragm can, in fact, be observed in the focal plane of the lens which follows the filter:

1. A direct, undeflected image, A, which is the one that will be used;

2. Below, in B, 7 superposed images, formed by the rays deflected once by each of the 7 prisms;

3. Below, in C, 21 superposed images, deflected twice by each of the double combinations of the 7 prisms;

4. Also below, in D, 35 superposed images deflected three times by each triple combination of the 7 prisms;

5. In E, 35 superposed images, deflected four times;

6. 21 superposed images, deflected five times;

7. 7 superposed images, deflected 6 times;

8. 1 image, deflected 7 times.

The last images cannot be detected under normal conditions, for their rays do not leave the filter.

The diaphragm which limits the field is given a height which leaves the first image completely separated from the others. This height, which must not exceed $3^{\circ}7'$, is more than sufficient to allow the use of all the field in which the wavelength is practically constant. In the eyepiece one observes only the direct image A; contrary to what one might expect, the 127 unused images do not represent any loss of light; as a matter of fact, the wavelengths are not the same as that of the radiation isolated in the direct image; none of them contains this radiation.

The transparency of the filter is no longer limited except by the need for polarizing the light; it is much greater than with the polaroids, reaching 40% in the entire visible spectrum.

Polarization is complete for all radiations. There is no longer any stray radiation, and the equivalent width of the filter becomes equal to that of a band, i.e., 3 A in the red.

One would expect the 127 unused images to produce a strong diffused light, but this is not the case; the diffused light is much lower than with the polaroids.

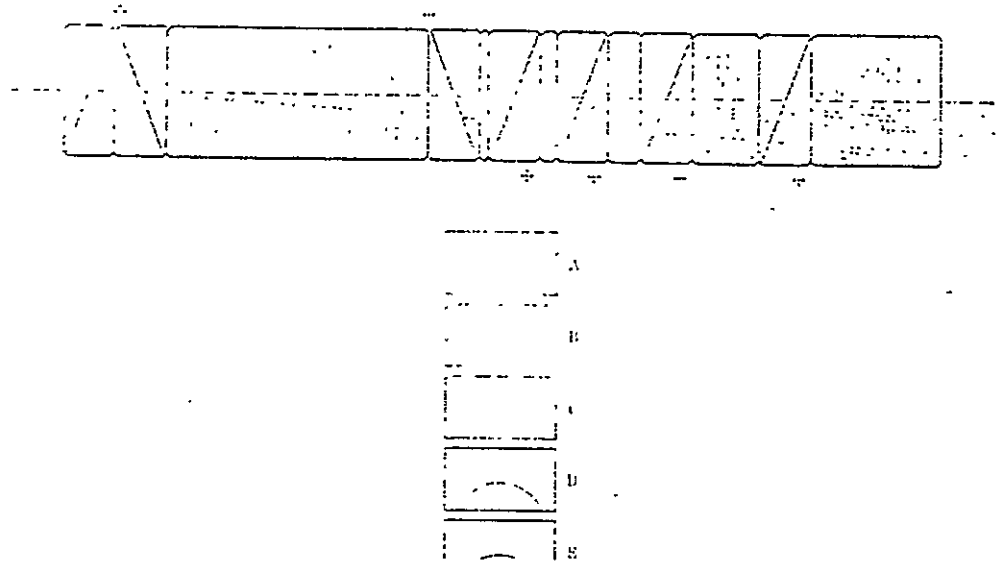


Figure 16

Diagram of the second filter with birefringent polarizers and arrangement of the images given by this instrument.

Lastly, the faces of the calcite prisms are much more regular than those of the polaroids and, cemented with Canada balsam, they make it possible to obtain perfect images, provided the beams passing through them are sufficiently parallel, i. e., that they do not converge at less than 3 meters distance.

The index and especially the dispersion of the crown fluor are not at all the same as those of the calcite for the extraordinary ray; this gives rise to a slight deflection of the direct beam and a dispersion which is troublesome when observations with several radiations are made simultaneously. This dispersion is compensated, for the green and the red, by adding a small-angle direct-vision prism to the left end of the filter.

New coronal observations. The monochromatic filter with its calcite polarizers, mounted behind the coronograph of the Pic du Midi, gave brighter and sharper images in 1941 than in 1939, when it had polaroids.

When H_α was isolated, the chromosphere appeared at the solar limb without any need to mask the rest of the disk; filaments and some flares in sunspot groups were seen simultaneously.

When the red and green coronal radiations were allowed to pass at the same time, the corona appeared with a vivid green hue because the eye is more sensitive to this color, but certain parts were seen to be strongly shaded with red; some of the jets which do not emit green radiation were even distinctly red. In the red, the instrument gave more sharply contrasted photographs than in 1939; exposure times were shorter, especially in the green.

Figure 17 (Plate III) shows the green corona of 3 September 1941 at 0805, recorded at the west limb, with a 6-minute exposure. On the negative the solar image measures 70 mm in diameter; the corona has a very complex structure, seemingly formed of entangled arches, fine jets and small bright clouds.

Below, we see the red corona photographed at the same place 15 minutes later, with a 7-minute exposure. Its structure is much simpler and the arches stand out better. The large jet of the red image, to the left at the bottom, is weak and short on the green image; farther to the left, a parallel jet, more intense in the green, is invisible in the red. The most intense jet of the green image, in the upper right, is very weak on the red image, where only its outline is seen. In the center, the same arches appear on both images, in the green with a knotty structure, and sharp and regular in the red.

Generally speaking, the green and red coronas differ from each other as much as the chromosphere photographed with hydrogen and calcium; these two kinds of images also differ greatly from that obtained without a filter with the continuous spectrum.

Figure 18 (Plate III) shows the green corona of 15 September at 1125, on the east limb. The jets, generally double, seem to fold back toward the interior of the luminous mass, like the petals of a rose.

Figure 19 (Plate III) represents the green corona of 14 September at 0900, to the west. The concentration of light at the equator is much less marked and a very dark interval may even be seen in the center. These is no arch but only incurved jets.

THREE-COLOR CINE PHOTOGRAPHY

Radiation separator. During 1940 and 1941 I studied and adapted for use with the monochromatic filter a radiation separator and a cine camera which allowed three different radiations to be filmed simultaneously: green coronal line 5,302.8 A, red coronal line 6,374.5 A and the H_α line of prominences 6,562.8 A. The first two radiations were isolated at 39.4°, but the third was isolated at 47.5° whereas at 39.4° it was strongly absorbed by the fifth plate of the filter, 36 mm thick, the wavelength of the nearest filter band being 6,562.8 ÷ 5.6 = 6,568.4. I employed the following method to avoid this:

Since the plates of the filter can be placed in any order, I placed the 36 mm plate (A in figure 20) last. I replaced the exit polarizer by a double-image calcite-and-glass prism B, which separated the rays in a plane perpendicular to that of the other polarizers; the sign ÷ indicates on what side the optical axis of the calcite is ahead of the image plane.

PLATE III

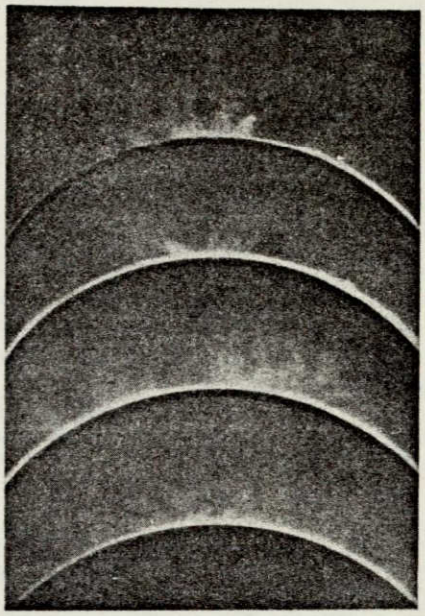


Figure 17. Green corona and red corona at the west limb, 3 September 1941, 0810.

Figure 18. Green corona at the east limb, 15 September 1941, 1125.

Figure 19. Green corona at the west limb, 14 September 1941, 0900.

PLATE IV

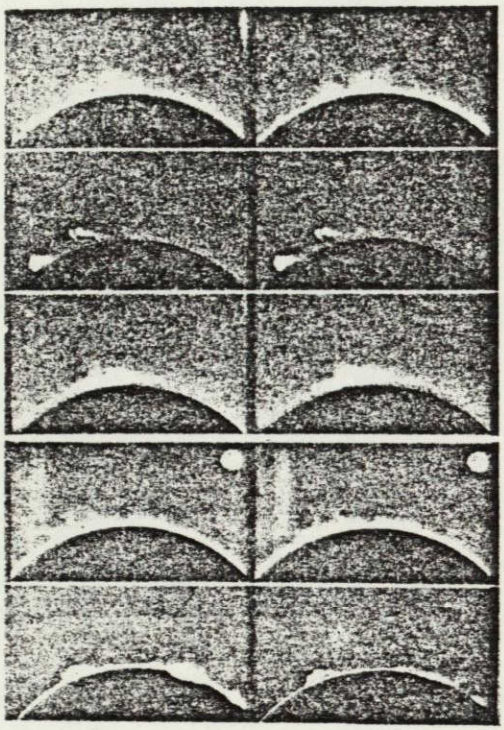


Figure 23. Portion of 3 simultaneous films of 10 August 1941 (0822 and 0824) red line, H_a line, green line.

Figure 24. Portion of 2 simultaneous films of 1 September 1941 (1225 and 1226), red line and H_a line.

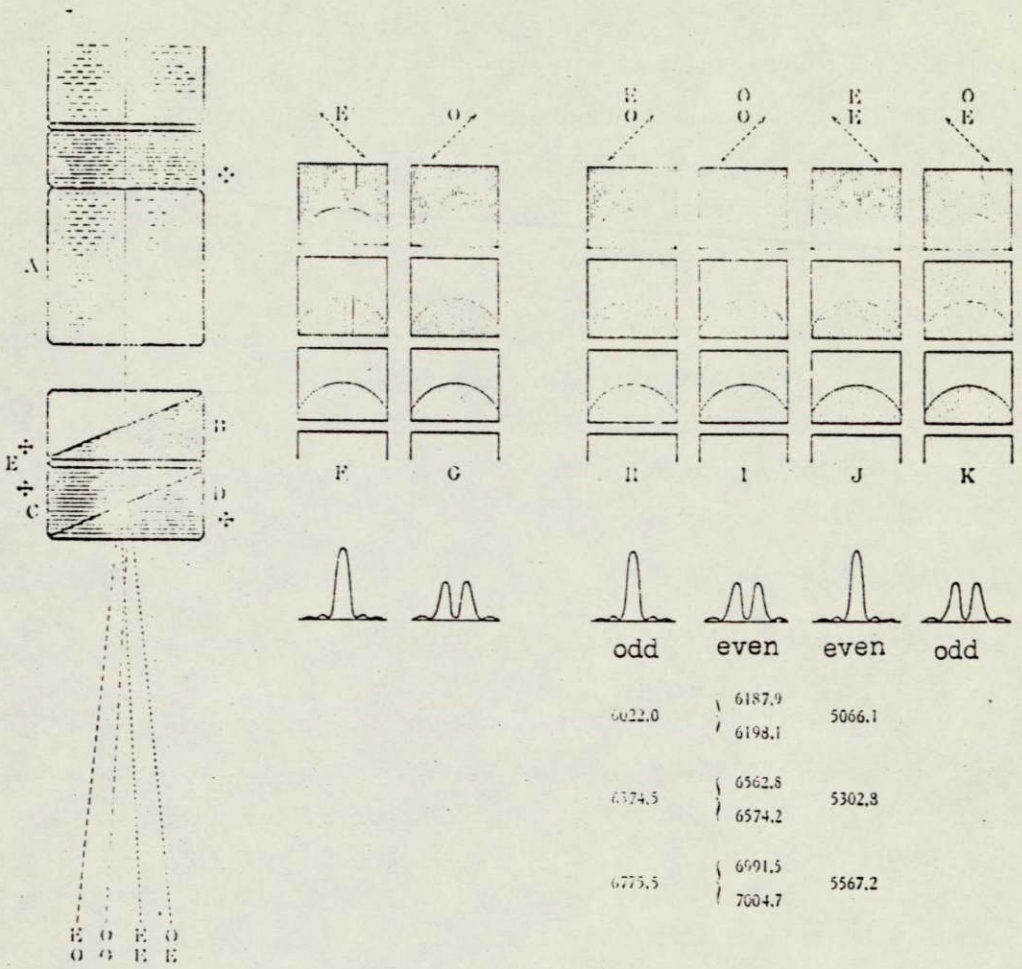


Figure 20

Diagram of the radiation separator placed after the filter. Arrangement of the images.

I obtained two series of pictures, F and G, in this way. Series F was produced by the extraordinary ray of prism B; the upper picture contained the normal radiations of the filter, each consisting, as shown by the curve beneath it, of a principal maximum and weak secondary maxima. At 39.4° one of the transmitted radiations coincided with the green coronal line; another coincided with the red line.

Picture series G was produced by the ordinary ray of prism B; the upper picture of this series contained double

radiations; in fact, for this ray, the 36-mm plate absorbed the center of the radiation transmitted by the rest of the filter; it passed, on each side, two bands with an intensity of 0.44. The band of wavelength 6,568.4 A was split into two components 11.6 A apart,the weaker of which, wavelength component 6,562.6 A, practically coincided with H_α.

In the next position I added a quartz plate E, 1.0966 mm thick, and a Wollaston-type birefringent prism CD whose splitting effect was twice that of the preceding one. This combination separated the whole bands from the split bands and finally obtained the following result:

There were 4 series of pictures, H, I, J, K; only the first picture of each series was of interest.

Image H, which was formed by an extraordinary ray in prism B and an ordinary ray in prism C and will be termed "odd," contains the whole bands, one of which coincides with the red coronal line.

Image I, which was formed by an ordinary ray in prisms B and C and will be termed "even," contains the split bands, one of which, of wavelengths 6,562.6 and 6,574.2 A, has its short wavelength component on H_α.

Image J, which was formed by an extraordinary ray in prisms B and C and will also be called "even," contains the whole bands, one of which coincides with the green coronal line.

The fourth image, K, which was formed by an ordinary ray in prism B and by an extraordinary ray in prism C, contains no interesting radiation.

The first image passed through a suitable red filter to a panchromatic film which recorded only the corona with its red radiation 6,374.5 A.

The second image passed through a darker red filter to a second panchromatic film which recorded only the prominences with the red H_α radiation.

The third image passed through a suitable green filter to an orthochromatic film which recorded only the corona with its green radiation 5, 302. 8 A.

The fourth image was seen in an eyepiece and was used to keep the disk of the coronograph centered on the sun throughout the whole time of filming.

Thus, the principal radiations were separated almost without loss of light.

Camera. Like the mounting for the monochromatic filter, the three-color movie camera was built entirely by A. Martin, who was then the machinist at the Meudon Observatory.

Figure 21 shows a view of the camera attached by clamp A to the end of tube B.

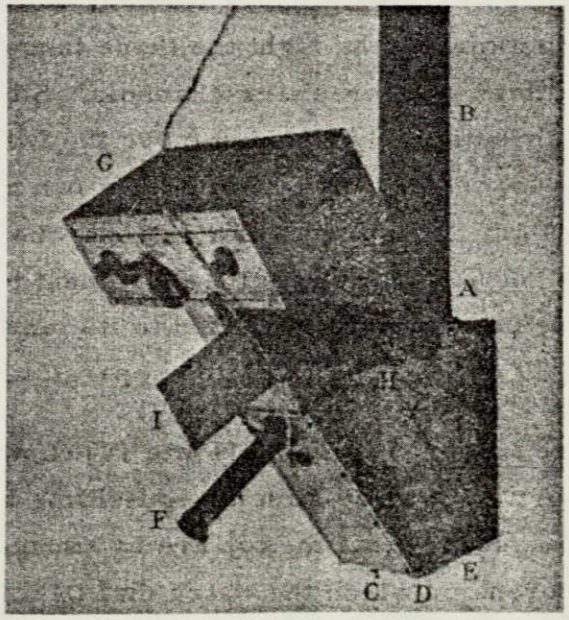


Figure 21

The three-color camera attached to the coronograph.

This tube, carried by the coronograph, contains the monochromatic filter, the radiation separator and a lens with an 800-mm focal length which forms the monochromatic images on the three films and in the eyepiece. In order to install the complete unit in the dome of the Pic du Midi Observatory, we had to bend the light beams back, reflecting them from a mirror at the bottom of the apparatus. The mirror is supported by three set screws, C, D, E, which regulate its orientation.

Each of the four reflected light beams goes through an appropriate filter set in plane AF and reaches the movie films contained in box G, at the top. Eyepiece H permits observation of the fourth image in a fixed mirror and, if need be, the other three in a movable mirror attached to the shutter and operated by a lever; for this purpose it is provided with a slide I. A box containing a watch may be attached to the end of tube F. The watch is then photographed automatically on each frame of one of the films.

Figure 22 shows the inside of the film box: to the left are the three feeder spools, to the right the three take-up spools, and in the middle, the three film channels. Slightly to the right we see the three sprockets which move the film. They are operated by a lever, at the top, by which they can be made to rotate exactly one-half turn, thanks to a ratchet wheel. A chain, located at the bottom, transmits the movement to the take-up reels. At the bottom right, attached to its lens (which has a focal length of 80 cm), is the radiation separator.

Motion picture films. Figure 23 (Plate IV) shows, by way of example, a portion of three simultaneous films taken at the west limb on 10 August 1941; at the top are two successive pictures of the red corona on which the watch dial indicates the hours (0822 at the right and 0824 at the left); in the middle are the corresponding pictures of the prominences with H_α; and at the bottom, those of the green corona.

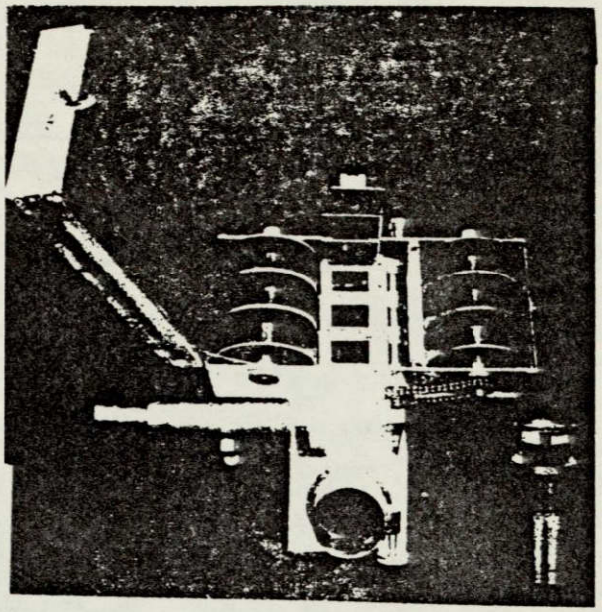


Figure 22

Three-color camera and radiation separator.

These films were taken at the rate of 1 picture every two minutes with an exposure of about 1 minute 50 seconds. With this exposure time, the red corona is overexposed about 3 times; the green corona, on the contrary, is slightly underexposed because of the lower sensitivity of the best emulsions in this region. The prominences of normal intensity are a hundred times overexposed when recorded with the same emulsion as the red corona, but the H_α film of 10 August was taken with an emulsion ten times faster. The overexposure destroys the relief but shows very weak prominences which would otherwise remain invisible.

These simultaneous pictures confirm the dissimilarity of the green and red coronas and they show that the corona and the prominences are almost completely independent.

In the film of 10 August, for example, the arched prominence located at the left does not correspond to any coronal detail; it appears lightly on the picture of the red corona, but only because of its very great brightness. The hook-shaped prominence located more to the right also does not correspond to any definite coronal detail. In contrast with this, the most intense coronal jets coincide only with the smallest and weakest prominences.

On the three-color films, the black disk which masks the sun appears edged with a double fringe, because the main lens of the coronograph is not achromatic; the image of the sun was not in focus on the disk for all radiations at one time; this focusing, carried out in the yellow, was mediocre in the green and red.

To avoid this, in certain cases I neglected the green corona and took only the red corona and the prominences, which allowed me to place the disk at the focal point, reduce the exposure time to 50 seconds, and take one shot per minute.

Figure 24 (Plate IV) shows 2 successive pictures from a film obtained under these conditions at the west limb on 1 September 1941 at the hours indicated by the watch: 1225 at the right and 1226 at the left; at the top is the red corona, and at the bottom, the prominences with H_α. These pictures still show the complete lack of similarity between the prominences and the corona.

On all films, each picture of the red corona is accompanied by a picture of the watch dial. Some of them, like this one, ran for 12 consecutive hours without interruption, thanks to the marvelous atmospheric conditions so frequently to be found at the Pic du Midi.

All pictures, of Leica format 24 x 37 mm, were later reproduced on a positive with a printer which reduced them and

corrected for opacity differences due to the variations of atmospheric diffusion, according to the readings of a galvanometer connected to a photoelectric cell.

Movements of the corona. These cine films allow solution of an important problem: Are there movements in the corona?

Observers of eclipses had tried to show movements in the corona; unfortunately, total solar eclipses are short; they never last more than seven minutes and the coronal details, which do not have definite outlines and which are often projected in front of each other, do not lend themselves to settings exact enough so that their displacements can be recognized with certainty in such short intervals of time.

Indeed, Perrine, comparing photographs of the 1901 eclipse and Hansky, studying those of the 1905 eclipse, found that the speeds of the movements should not exceed about 25 km/sec [16].

Other observers have compared photographs taken at different stations which were crossed successively by the lunar shadow cone.

The photographs of the Lick Observatory, for example, taken during the 1905 eclipse in Spain and in Egypt at an interval of 70 minutes, show well-defined condensations in which structural details seemed to change but with speeds of less than 1.6 km/sec [17].

On the other hand, the photographs of the Lick and Sproul Observatories, taken during the 1918 eclipse in the United States in Washington and Kansas, showed coronal arches which surrounded prominences and which seemed to have moved in 26 minutes with speeds of about 16 km/sec [18].

Furthermore, in 1926, Horn d'Arturo compared the photographs of the Italian expedition in East Africa with those of Sumatra and found movements of coronal domes which confirmed the preceding results [19].

Lastly, during the eclipse of 19 June 1936, which crossed all of Siberia, photographs were taken at six stations with standard instruments at times spaced out over an interval of 2 hours 13 minutes. Four stations had fine weather; the photos show very small displacements of jets, while small coronal clouds reportedly moved more rapidly, at about 4 or 5 km/sec [20].

Apart from eclipses, Waldmeier [21], during the morning of 16 February 1939, with the coronograph of the Arosa Observatory equipped with a wide-slit spectroscope isolating the green line, saw two coronal arches which rose with mean speeds of 7.25 and 11.6 km/sec. This observation would seem to prove the existence of quite rapid movements in the corona.

However, in 1936, 1937, and 1938, years of great solar activity, in my coronal observations I frequently applied the wide-slit method with the Pic du Midi coronograph, which is more powerful than that of Arosa and which has a more dispersive spectroscope [22], without obtaining satisfactory results. The images lacked sharpness because of the great width of the green line. With a fine slit, the forms could not be seen well because the field was too limited; with a wider slit, the continuous spectrum weakened the contrasts and caused the faint details to disappear. Consequently, the method would not seem to give sufficiently clear results to allow definite conclusions about the movements of the corona.

Several hundred photographs of the coronal spectrum were taken at the Pic du Midi. In these photos, the lines of the prominences do show considerable movements at certain points, due to high radial velocities, but the lines of the corona show

only very slight dissymmetries and only very rarely experience observable deviations. Hence it seems improbable that these are rapid movements in the corona. To settle this question, one must obtain numerous good coronal images, spread, as far as possible, over a whole day.

In 1938 I took a series of direct photographs of the corona in total light [23]; they depicted very substantial changes in the space of a few hours, but were too widely-spaced in time to show the evolution of a particular detail and to indicate whether the changes observed were really due to movements.

The photographs of the green line and the red line taken in 1941 with the polarizing filter in monochromatic light show clearer details and more evident changes. Among the most interesting series, we see (fig. 25, Plate V) 8 photos of the green corona, at the east limb, taken on 30 July 1941.

Between 0750 and 0905 a small cloud was illuminated in the main group, to the right; at 1216 it reached maximum brightness, and at 1230 disappeared completely. The arches surrounding this cloud also underwent considerable brightness variations; at 1555 the right side of two of them became bright. At 0905, slightly more to the right, a new arch appeared; at 1116 it was no longer visible, but at 1330 a larger arch, concave toward the left, rose from the same place, by 1555 it had developed in length and at 1720 it surrounded the group of preceding arches.

Figure 26 (Plate V) shows 5 photos of the red corona of 2 September 1941, at the west limb. These images show, to the left of center, a large bright jet which remained practically constant; by the end of the day, its upper half had weakened. To its right is a group of arches, of which the two main ones, elliptical in form, have their long axis vertical and are almost concentric. The outer arch reached a maximum brightness and sharpness at 1150; the inner arch was very distinct at 1050 and at 1150; then its right side disappeared and its broadened left side

coincided with the left side of a larger, paler arch which remained visible all day. To the right of the group of arches is a small cloud which was quite bright at 0915, weakened at 1050 and disappeared at 1150.

These photos show considerable variations of brightness, but they do not show any movement, properly so-called; certain parts grow bright and others dim by turns without one's being able to follow the movement of a given object.

When these views are compared in a stereoscope, photographs taken at intervals up to 4 hours can be examined simultaneously without the difference in appearance of the corona as seen by each eye being too disturbing. The arches show only slight effects of relief, which correspond to relative movements of less than 1 km/sec; these slow movements do not necessarily originate in the corona; the combined effect of perspective and solar rotation are enough to explain them.

Despite these negative results, more complex movements might exist in the corona, e.g., vortices, which would be hard to detect in isolated photographs; only time-lapse photographs would verify them.

In 1939 and 1941, I obtained 8 series of pictures of the corona taken in total light at the rate of 1 picture per minute; each series contains, on the average, 500 pictures and covers a period of 8 to 13 hours. Two of these series have been reproduced on film; when they are projected at the normal rate, i.e., accelerated 1200 times, they disclose no movement but only relative variations of intensity.

A greater number of films was taken in 1941 with the monochromatic filter. They cover a total observation time of 92 hours. Those of 10 and 17 August and 1 September, for example, each last 12 consecutive hours. They show sharper, better contrasted and more abundant details of the corona than those taken in total light; they are also much less affected by variations

of atmospheric diffusion. When they are projected at normal speed, i.e., accelerated 1200 and 2400 times, they fully confirm the preceding conclusions.

While the prominences appear to be animated by very rapid movements, the corona in monochromatic light (either of the green line or of the red line) remains entirely immobile. Arches, jets or clouds brighten by turns like the rays of a polar aurora, but no displacement can be seen.

These series of photos and films have not yet been studied thoroughly; however, the cursory examination given them by no means confirms the rapid movements reported by some observers. Obviously, these series of pictures do not prove that such movements cannot exist, but they do cover a period of 23 hours for the photographs and 92 hours for the moving picture films, which is more than 10 times the total length of eclipse observations; moreover, they show sharper, more contrasted details in the inner corona. If movements of 10 km/sec were as frequent as the observations previously cited would lead us to suppose, the monochromatic images should have revealed them several times.

On the other hand, the films show that the relative variations of intensity of details which partially coincide, the successive appearance of clouds and arches, for example, at different distances from the sun frequently give rise to the semblance of movements. It would seem that one could attribute to phenomena of this kind the movements which have been reported.

At the present, the following conclusions may be drawn from the evidence provided by the birefringent filter.

In general, the corona changes in form and aspect, not like most of the prominences (that is, by relative movements of their various parts), but by the appearance and disappearance and relative intensity variation of the elements which compose it.

The coronal arches and coronal clouds, which often form complex ensembles, are born in one place along invisible trajectories which existed beforehand and whose origin and mode of formation we cannot explain.

This very curious phenomenon seems quite general and theories concerning the corona must take it into consideration.

STUDY OF THE CHROMOSPHERE

As we have seen previously, with H_α radiation the filter showed the chromosphere at the solar limb, but on the solar disk itself the contrasts were insufficient; only the most intense filaments appeared. The chromospheric structure was difficult to discern; hence the filter had to be made more monochromatic.

Arrangement. In August 1932, at the Pic du Midi, I added a plate to the filter; this reduced the equivalent width of the transmitted band on H_α from 3 A to 1.5 A. To keep the instrument within the requisite length, I had to use a calcite plate; it was 8.3 mm thick and was equivalent to a 150 mm quartz plate. I placed it after the filter, in the same mounting as the quartz elements, thus keeping its temperature at a constant value of 47.5°.

This plate had not been cut for the filter: at 47.5° none of its bands coincided exactly with H_α. In order to shift its bands at will in the spectra, I placed an elliptical polarization compensator after it.

I cemented a quarter-wave mica plate for the red between the calcite plate and the end glass, and oriented its axes at 45° to those of the calcite plate. It transformed the elliptical vibration produced by the crystal for a given wavelength into a rectilinear vibration whose plane turned when this wavelength was varied. The interval between the two bands corresponded to a rotation of 180°.

I followed the quarter-wave plate with a final polarizer, a polaroid plate which was placed outside the filter near the eyepiece and which could be turned in its plane. When this was suitably orientated, a bright band could be brought on H_a; the sun then had a minimum glare. Starting from this position, the transmission maximum of the filter could be shifted slightly toward the red or toward the violet; radial velocities could thus be detected without varying the temperature.

About the middle of August 1942, the filter thus modified was installed at the eyepiece end of the Pic du Midi telescope, temporarily equipped with an excellent lens having a 38 cm aperture and a focal length of 6 m, loaned by the Toulouse Observatory. As the unit was too long, the beam had to be bent back.

At the telescope's focal point, a silvered copper plate A (fig. 27), on which the solar image formed, deflected the greater part of the solar rays. An aperture B, 8 mm x 12 mm, circumscribed the field. The rays of this portion of the sun fell on the total-reflection prism C, then on lens D, with a 300-mm focal length, which made them appreciably parallel. The rays passed through the monochromatic filter F, then through lens G (focal length: 300 mm), followed by a divergent lens H. This ensemble constituted a Galilean telescope which formed an enlarged image of the sun with an equivalent focal length variable from 15 to 25 m, depending on the divergent lens used; this image could be focused by displacing lens H with button l.

Next came polaroid J, cemented between two glasses, which could be turned in its plane with knob K to vary the wavelength. Lastly came filter L, which selected the H_a band, and tube M at the end of which an eyepiece or a cine camera was placed.

Observations. During the last part of August and the beginning of September 1942, chiefly in the morning, the solar images were often perfect. The filter did not change appreciably

the optical qualities of the 38 mm telescope, which could be used at full aperture, thus permitting observation of the chromosphere under excellent conditions with magnifications as high as 500 x; very fine details were visible, and the prominences showed, in particular, a remarkable fibrous structure.

In the cine camera the image of the sun on the film measured 15 to 20 cm in diameter. However, the great brightness of the chromosphere permitted the use of high-contrast emulsions and reduction of exposure time to 1/40 sec in order to obtain sharper images.

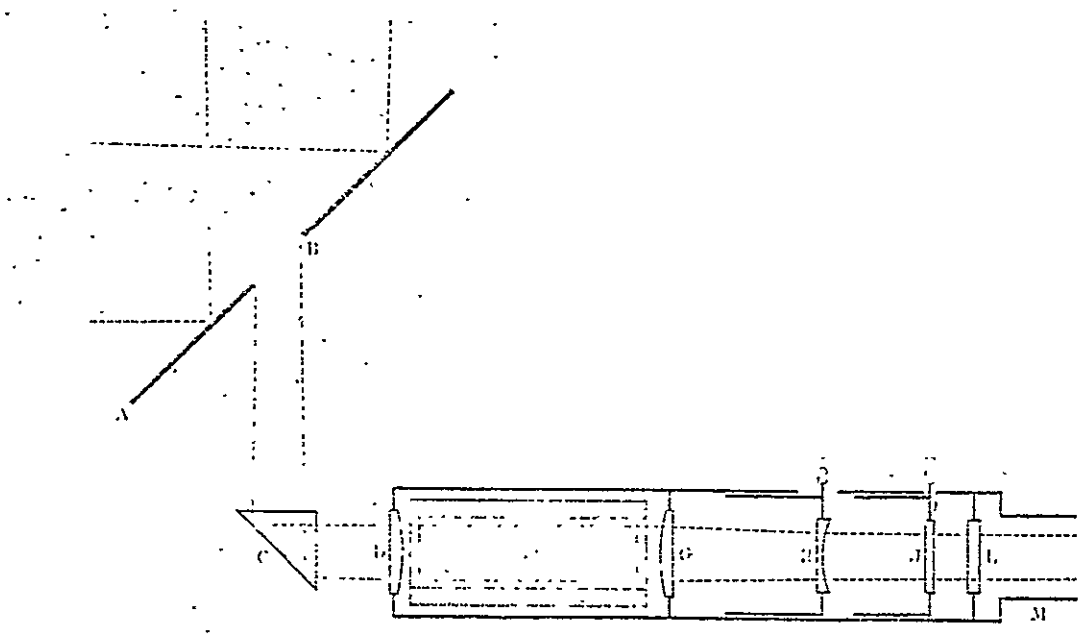


Figure 27
Arrangement of the filter behind the telescope.

Under these conditions, 130 m of film were shot at the rate of 2 pictures per minute; a watch dial photographed on each frame indicates the corresponding time.

Plate VI (figs. 28 through 37) shows portions of some of these pictures, enlarged 3.8 times; on this scale, the diameter of the whole sun would be 57 centimeters.

PLATE V

Variations of the corona

Figure 25. 30 July 1941, the green line, east solar limb.

Figure 26. 2 September 1941, the red line, west solar limb.

PLATE VI

Pictures of the chromosphere in 1942.

Figures 34 through 37 show the solar limb. The latter is double; it includes a very clear, perfectly circular line, the edge of the photosphere, which appeared thanks to the light transmitted by the filter on each side of the H_α line and a less bright fringe whose outer edge is irregular; this fringe is merely the chromosphere seen edgewise with the H_α radiation.

Thus, the chromosphere does not appear as a homogeneous atmosphere, but as a multitude of luminous jets or sheets like small prominences which spurt from the photosphere, sometimes vertically, sometimes in oblique directions. They are projected in front of each other, giving the impression of a continuous fluid, but if the wavelength of the band transmitted by the filter is decreased slightly by turning the last polarizer, the jets which approach the observer appear to be more intense than the others; the chromosphere becomes more transparent and appears striated.

On these photographs, the chromospheric fringe seems to be composed of three parts:

1. A dense layer whose outer edge is quite well-defined; its height, variable from one place to another, is most often from 5" to 7", or a little more than half the total height generally admitted for the chromosphere; it drops at some points as low as 3" (fig. 36);

2. A very faintly luminous layer, visible only in places, in front of prominences whose light it appears to absorb to a height of 1" to 2" (figs. 35 and 36);

3. A series of weak jets, whose height it is hard to state since the highest of them may be considered very small prominences.

When projected at normal speed, the films accelerate the movements from 400 to 600 times; they show that these chromospheric jets are short lived, of the order of a few minutes; they

are extinguished while others spurt out in turn and the dense layer of the chromosphere seems animated by a continuous boiling.

Certain prominences, generally very small, may be brighter than the chromosphere; when they are located behind the solar limb, the chromosphere, which is completely opaque for the radiations which they emit, is projected as a shadow before them (fig. 37).

Over the whole solar disk, the chromosphere appears with a characteristic structure similar to that which one observes in the spectroheliograph with a wide slit; it is covered by small dark plages, variable in size and intensity. Around sunspot groups, these plages stretch out in radiating directions, forming figures reminiscent of magnetic spectra (fig. 31).

The spots are sometimes accompanied by bright, extensive chromospheric flares, emitting a radiation with very short wavelength which profoundly modifies radiowave propagation. The observations made in 1942 with the polarizing filter show that besides these relatively rare and beautiful phenomena, a very great number of flares, whose brightness generally exceeds that of the continuous spectrum outside of H_α but whose diameter often measures less than 1 second of arc, occur around the spots.

In a small group of sunspots which formed at the central meridian on 22 August, as many as 17 flares appeared in one hour; they lasted from 5 to 15 minutes, sometimes longer; often several flares occurred successively in the same place. On 25 August, for example, at 0655 (fig. 28) at the edge of the principal spot, below and to the right, we see a small prominence projected in front of the sun in the form of a dark filament. About 0700 (fig. 29), the filament became very dark, while its left end was illuminated. At 0717 (fig. 30), the latter shone with a bright flash and extended by irradiation. At 0755 (fig. 31), it was completely extinguished.

Beginning at 0700, at the right edge of the penumbra of the main spot, a small point was lighted up for 10 minutes (fig. 29).

At 0710, a small bright line appeared above the third spot from the left (fig. 30); 20 minutes later, it broadened and weakened to form a light plage (fig. 31).

At 0720 two small points formed on the upper edge of the umbra of the spot to the right of the principal one. By 0755 (fig. 31), they had disappeared, but a very small, very bright point can be seen a little higher up: two other points appeared under the second and fourth spots from the left.

At 0828 (fig. 32), the first point was still distinct; the second was diffuse, and a third appeared to the right and a little below the second spot.

These flares were animated at their beginning by strong radial velocities. If the wavelength of the filter was decreased by about 1 A, it passed chiefly the light of the continuous spectrum outside of H_α; the sun became brighter, the chromospheric structure weakened considerably and one could see the much more compact network of the photospheric granulation. Certain flares were disappearing, but others were becoming brighter. The picture in fig. 29 and especially in fig. 33 were obtained under these conditions; a flare measuring 1".5 in diameter can be seen between the second and third spots from the right in fig. 33; it far exceeds the brightness of the continuous spectrum of the photosphere.

The monochromatic filter also showed the filaments. When these dark streamers, formed by the prominences, are projected in front of the brighter background of the chromosphere, generally they show rapid movements on the films; they spurt out at a given point and plunge into the regions near the sunspots.

Only very small filaments appeared during the observations, the most evident ended at the spot to the right of the main spot; it was intense at 0655 (fig. 28) and later weakened.

CONCLUSION

The birefringent monochromatic filter should have a wide variety of applications thanks to its high light yield and to the breadth of field which can be observed with it. The results just presented show that it constitutes a new means of investigation by which we may acquire more complete data on the constitution of the sun.

Manuscript received 5 June 1944

LITERATURE CITED

1. Lyot, Bernard. "Un filtre vert, monochromatique" (A monochromatic green filter), Académie des Sciences, Paris, Comptes Rendus, 200 (9): 738-739, 1935.

1(a) Lyot, Bernard. "L'observation directe des protubérances à Meudon" (Direct observation of protuberances at Meudon), Académie des Sciences, Paris, Comptes Rendus, 195 (21): 943-945, 1932.

2. Lyot, Bernard. "Résultats obtenus avec le coronograph de M. B. Lyot à l'observatoire du Pic du Midi" (Results obtained with B. Lyot's coronograph at the Pic du Midi Observatory), L'Astronomie, 46: 284, 1932.

2(a) Lyot, Bernard. "Étude de la couronne solaire en dehors des eclipses" (Study of the solar corona without eclipses), Zeitschrift für Astrophysik, 5 (2): 73-95, 1932. N. B. p. 91.

3. Mascart, Annales Scientifiques de l'École Normale Superieure, 3: 395, 1874.

4. Fabry, Ch. and A. Perot. "Sur la constitution des raies jaundes du sodium" (Structure of the yellow lines of sodium), Académie des Sciences, Paris, Comptes Rendus, 130 (10): 653-655, 1900.

5. Wood, R. W. "Trennung eng beobachbarer Spektrallinien zum Zwecke monochromatischer Beleuchtung" (Separation of close spectral lines for monochromatic illumination), Physikalische Zeitschrift, 15 (7): 313-317, 1914.

6. Wiener, O. Zeitschrift für Wissenschaftliche Photographie, 11: 13, 1912.

7. Priest, Irwin G. "The colorimetry and photometry of daylight and incandescent illuminants by the method of rotatory dispersion," Optical Society of America, Journal, 7 (12): 1175-1219, 1923.

8. Lyot, Bernard. "Un monochromateur à grand champ utilisant les interférences en lumière polaire" (A wide-field birefringent filter), Académie des Sciences, Paris, Comptes Rendus, 197 (25): 1593-1595, 1933.

9. Öhman, Yngve. "A new monochromator," Nature, 141 (3563): 291, 1938.

10. Öhman, Yngve. "A new monochromator," Nature, 141 (3560): 157-158, 1938.

12. Öhman, Yngve. Populär Astronomisk Tidskrift, No. 1, p. 1., No. 2, p. 27, 1938.

13. [illegible]ot, Bernard. "Un filtre monochromatique spécialement [illegible]apté aux recherches sur le Soleil" (A monochromatic [illegible]ter especially adapted for solar studies), Académie des Sciences, Paris, Comptes Rendus, 212 (24): 1013-1017, 1941.

14. Siedentopf, H. and J. Wempe. Astronomische Nachrichten, 270 (6): 273, 1940.

15. Haase, M. "Interferenzlichtfilter" (Interference light filter), Kristall-Laboratorium der Zeiss-Werke, Jena, Mitteilungen, [Reports of the Crystal Laboratory of the Zeiss plant in Jena]. No date.

15. Evans, J. Astronomical Society of the Pacific. Publications, 52 (309): 305, 1940.

15. Pettit, E. Astronomical Society of the Pacific, Publications, 53 (313): 171, 1941.

16. Perrine, C. D. University of California Publications. Astronomy. Lick Observatory Bulletin. 1: 151, 1902.

16(a) Hansky, Pulkovo Observatory, Mitteilungen, 2 (19): 107, 1907.

17. Campbell, W. W. and C. D. Perrine. University of California Publications. Astronomy. Lick Observatory Bulletin, 4: 121, 1907.

18. Miller, John A. "On changes in the corona of June 8, 1918," Astronomical Society of the Pacific, Publications, 32 (188): 207-214, 1920.

19. Horn d'Arturo, G. "Ecclise solare totale del 14 gennaio 1926 osservata della Missione astronomica italiana nell' Oltregiuba" (The total solar eclipse of 14 January 1926 observed by the Italian Astronomical Mission to Oltre Giuba), Società Astronomica Italiana, Memorie, Nuova Serie, 3: 484-491, 1926.

20. Vsessviatsky, Bugoslavskaya, and Deutsch. "Total eclipse of June 19, 1936," Academy of Sciences of the USSR, Report of Soviet expeditions, 7 February 1939.

21. Waldmeier, M. "Variationen der Koronaform" (Variations of the shape of the corona), Zeitschrift für Astrophysik, 20 (3): 195-213, 1940. N. B. p. 210.

22. Lyot, Bernard. "Quelques observations de la couronne solaire et des protubérances en 1935" (Some observations of the solar corona and protuberances in 1935), L'Astronomie, 51: 203, 1937.

23. Lyot, Bernard. "A study of the solar corona and prominences without eclipses, Royal Astronomical Society, London, Monthly Notices, 99 (8): 580-594, 1939. N.B. p. 586.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 39, NUMBER 3 MARCH, 1949

The Birefringent Filter

JOHN W. EVANS
Harvard Observatory, Harvard University, Cambridge, Massachusetts
(Received November 19, 1948)

Optical birefringent filters, which depend for their action on the interference of polarized light, can be designed to transmit very sharp bands (down to a fraction of an angstrom in width). The elementary theory necessary for their design is given.

Three forms of wide field filters designed by Lyot are described in detail. A more recently developed split element, wide field filter requires only half as many polarizers as the earlier types, which may be an advantage for some applications.

Various methods of adjusting the transmission bands of a birefringent filter, including the use of elements of variable thickness, and phase shifters are discussed. For most purposes the electro-optical phase shifters are probably the most promising. For special purposes, such as spectrophotometry, phase shifters composed of rotating fractional wave plates may be more advantageous. Two such phase shifters and their application in simple and split element filters are described.

A few crystalline materials which have been used or might be used to advantage in birefringent filters are mentioned.

Finally, the possibility of using polarizing interferometers in combination with birefringent elements for filters with extremely sharp transmission bands (in the range of hundredths or thousandths of an angstrom) is very briefly discussed.

I. INTRODUCTION

DURING the past few years the birefringent filter has proved an effective tool in astronomical research. Its utility, however, is not confined to astronomy and the purpose of the present paper is partly to bring it to the attention of investigators in other fields.

Briefly, the birefringent filter serves the purpose of a monochromator over an extended field. It can be designed to transmit a wave-length band of any desired width (down to a fraction of an angstrom) centered at any selected wave-length. It is used very much like an ordinary glass or gelatin filter in either a collimated or a converging beam of light, but with some limitation in field size or focal ratio, according to the type of construction, material, and band width.

The invention of the birefringent filter is one of the many important contributions of the French astronomer, Bernard Lyot,[1] to instrumental astronomy. He first published the basic principles of its operation in 1933. Öhman[2] independently invented the filter and in 1938 constructed the first one to be used for solar observations, with a transmission band about 40-angstroms wide centered on the $H\alpha$ line. With it he succeeded in seeing and photographing the brighter prominences, although it was evident that a much sharper band would be necessary for the best results.

In a later paper Lyot[3] has given a very complete discussion of the history, theory, and construction of birefringent filters. For the benefit of readers to whom his papers are not readily available, the present paper reviews enough of the elementary theory to suffice for the design of filters of any feasible characteristics. The remainder of the paper is a discussion of newer developments which serve to simplify the construction of the filters and extend their field of usefulness.

[1] B. Lyot, Comptes Rendus 197, 1593 (1933).
[2] Y. Öhman, Nature 141, No. 3560, 157 (1938); Nature 141, No. 3565, ... Astronomisk Tidskrift, No. 1–2, pp. 11 and ...
[3] B. Lyot, Ann. Astrop. ... 1–2, p. 31 (1944).

II. THE SIMPLE BIREFRINGENT FILTER

Several forms of the birefringent filter are possible, differing in width of field and complexity of construction. They all depend, however, on the interference of polarized light transmitted through layers of birefringent crystal in the direction perpendicular to the plane of the optic axes if the crystal is biaxial, or any direction perpendicular to the optic axis if the crystal is uniaxial.

Since we can regard the uniaxial crystal as a degenerate biaxial crystal, most of the following discussions will consider only the biaxial case. Let ϵ and ω be the extraordinary and ordinary indices of refraction of any uniaxial crystal, and α, β, γ be the smallest, intermediate, and greatest principal indices of refraction of a biaxial crystal, respectively. Any expression for a biaxial crystal is valid for a uniaxial crystal if one of the following substitutions is made:

$$\alpha=\omega, \quad \beta=\omega, \quad \text{and} \quad \gamma=\epsilon \quad \text{if} \quad \epsilon-\omega>0,$$

or

$$\alpha=\epsilon, \quad \beta=\omega, \quad \text{and} \quad \gamma=\omega \quad \text{if} \quad \epsilon-\omega<0.$$

Unless otherwise specified, the mutually perpendicular directions of vibration of light for which the refractive indices are α, β, and γ will be referred to as the α-axis, β-axis, and γ-axis. These are, of course, the principal axes of the index ellipsoid.

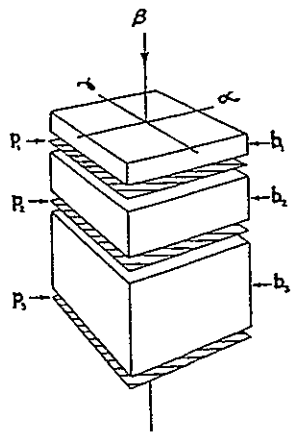


FIG. 1. Birefringent filter of three elements.

The quantity μ is defined by

$$\mu=\gamma-\alpha.$$

The term "retardation" will be used to indicate a path difference in terms of wave-lengths.

For brevity, the direction of vibration of light transmitted by a polarizer (prism or film) will be referred to as the axis of the polarizer.

Consider a block of some birefringent crystal, b_1 in Fig. 1, cut with its surfaces normal to its β-axis. Let light plane polarized at an angle of 45° to the α-axis enter the crystal along the β-axis. In the crystal the light divides into two components polarized with vibrations parallel to the α- and γ-axes, traveling with different velocities, c/α and c/γ. On emerging from the crystal, the two components have therefore a relative retardation of n_1, given by:

$$n_1=(d_1/\lambda)\mu, \tag{II.1}$$

where d is the thickness of the crystal in the β-direction, and λ is the wave-length of the light.

If now the light traverses a polarizer, p_1 (which may be either a Nicol or similar prism, or a film polarizer), with its axis parallel to the vibration plane of the entering light, the two components interfere. The transmission, τ_1, of the b_1, p_1 combination is:

$$\tau_1=\cos^2\pi n_1. \tag{II.2}$$

If white light traverses the combination, the spectrum of the emergent light consists of regularly spaced alternate bright and dark bands at wave-lengths where n_1 is alternately integral and half-integral. The transmission as a function of wave-length is represented by curve a, Fig. 2.

The wave-length interval between successive bright bands is inversely proportional to the thickness of the crystal. To obtain an approximation of the interval, set $\Delta n=1$ in the equation

$$\frac{\Delta\lambda}{\lambda}=\frac{\Delta n}{n}\frac{1}{(\lambda/\mu)(\partial\mu/\partial\lambda)-1}. \tag{II.3}$$

We now add a second crystal, b_2, and a polarizer p_2 oriented parallel to b_1 and p_1. If $d_2=2d_1$, the transmission of the b_2, p_2 combination, represented by curve b, Fig. 2, is:

$$\tau_2=\cos^2\pi n_2=\cos^2\pi 2n_1. \tag{II.4}$$

The transmission of the whole assembly, b_1, p_1, b_2, p_2, shown in curve c, Fig. 2, is therefore:

$$\tau_{12}=\cos^2\pi n_1\cos^2\pi 2n_1. \tag{II.5}$$

A third crystal element, b_3, with $d_3=2d_2$, followed by the polarizer, p_3, has individual transmission shown in curve d. The transmission of the assembly, b_1 to p_3, is then represented by curve e, Fig. 2.

It is evident that further crystal elements and polarizers can be added. The result is the basic type of birefringent filter, which will be termed the simple filter. It is comprised of a series of units, each consisting of a plane-parallel birefringent element (b-element) followed by a polarizer. All b-elements have surfaces normal to their β-axes and are mounted with their α-axes parallel. All polarizers have their axes parallel to the vibration plane of the entering polarized light at 45° to the α-axes. The thickness of the rth b-element is such that

$$n_r=2^{r-1}n_1. \tag{II.6}$$

The spectrum of light transmitted by the filter consists of a series of widely spaced narrow bands. Their separation is equal to the separation of the transmission maxima of the thinnest element alone.

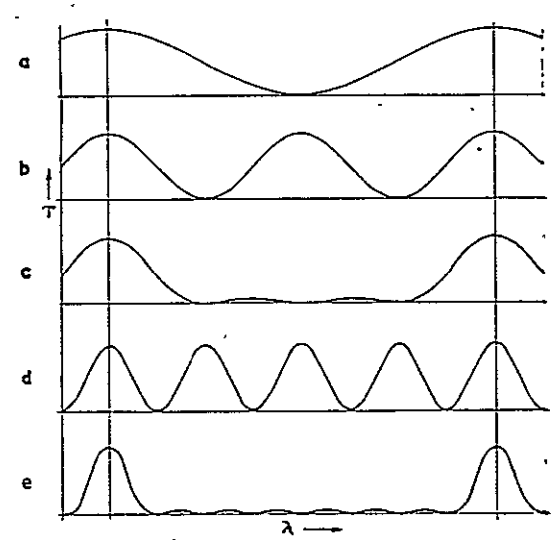


FIG. 2. Transmission curves for elements of Fig. 1. (a) b_1p_1; (b) b_2p_2; (c) $b_1p_1b_2p_2$; (d) b_3p_3; (e) $b_1p_1b_2p_2b_3p_3$.

while their effective width is the half-width of the maxima of the thickest element alone. For polarized entering light, the transmission of a filter of l b-elements (neglecting absorption in the material of the filter) is:

$$\tau = \cos^2\pi n_1 \cos^2\pi 2n_1 \cdots \cos^2\pi 2^{l-1} n_1. \quad \text{(II.7)}$$

The quantity n_1 must, of course, be an integer at the wave-length of the desired transmission band. Its magnitude should be small enough to separate the adjacent bands sufficiently to permit the isolation of the selected band by means of ordinary filters.

It can readily be shown that the total transmission of flux in an equal energy spectrum is 2^{-l}. Regardless of the width and separation of the bands, the total residual flux transmitted between successive principal maxima in a filter with $l>3$ is a substantially constant fraction (about 0.11) of the flux transmitted in a single band.

The filter at the Climax, Colorado station of the High Altitude Observatory of Harvard University and the University of Colorado has been in satisfactory operation in the observation of solar prominences since early 1943. It is a simple filter of six quartz elements with $n_1=23$, $n_6=736$, $d_1=1.677$ mm, and $d_6=53.658$ mm and has a transmission band of effective width 4.1 angstroms centered on the $H\alpha$ line of hydrogen (λ6563) at an operating temperature of 35.5°C. Its purpose is to eliminate the overpowering scattered light (continuous spectrum) near the limb of the sun while still transmitting the $H\alpha$-emission from the prominences, which are otherwise completely invisible. The success of the filter can be judged from the photographs in Plate I.

In practice, a filter should either be cemented or immersed in oil to avoid multiple reflections. Initial polarization is usually obtained by a polarizer mounted in front of the first b-element with its axis parallel to the axes of the other polarizers.

In any birefringent crystal, both the geometrical dimensions and μ are functions of temperature. The result is a small shift in the wave-lengths of the transmission maxima when the temperature changes. In quartz, for instance, $\Delta\lambda/\Delta T = -0.66$ angstrom per degree centigrade in the red. Hence the temperature of the filter must be controlled with sufficient accuracy to keep the maximum excursions of wave-length within tolerable limits. A total range of two-tenths of the effective band width is small enough for most purposes.

III. OFF-AXIS EFFECTS IN SIMPLE FILTERS

It is evident that when light traverses a simple filter at an angle to the instrumental axis, the light path through the birefringent material and the velocity difference of the fast and slow waves are altered. The effect is simply to alter the value of n_1 in Eq. (II.7).

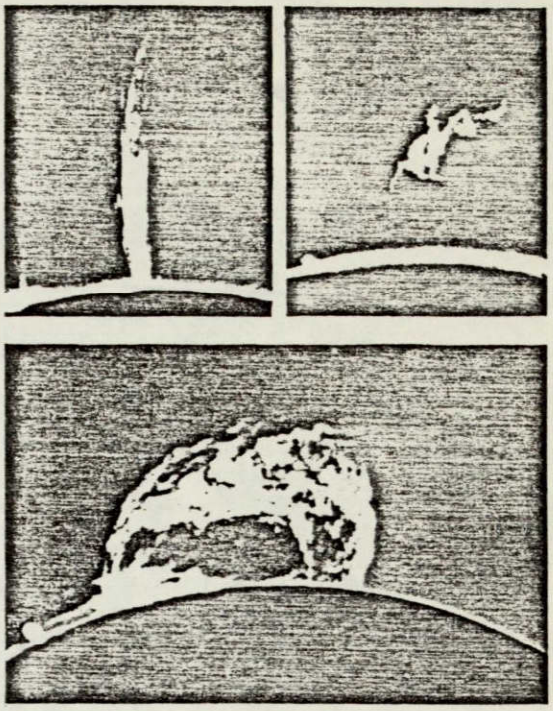


Plate I. Photographs of prominences at the limb of the sun taken through the birefringent filter of the High Altitude Observatory in the light of the $H\alpha$-line of hydrogen.

Lyot[3] has calculated the off-axis effect for light incident in the two principal planes normal to the α- and γ-axes in a biaxial crystal cut with its surfaces normal to the β-axis. Although the equations are not exact, since terms of the fourth and higher degrees in φ (the angle of incidence) are neglected, the approximation is excellent for the moderate angles of incidence encountered in the use of filters.

Lyot's equations can be very simply generalized to give the off axis effects for light incident in any plane normal to the surface of the crystal (and parallel, therefore, to the β-axis). Figure 3 represents a block of biaxial crystal with its α-, β-, and γ-axes in the directions indicated. Let polarized light with vibrations in a plane at 45° to the α-axis enter the crystal in the direction (φ, θ). Here φ is the angle of incidence, and θ is the azimuth of the incident plane measured from the α-axis. The light emerges from the crystal in the direction (φ, θ) in two polarized components with vibrations very closely parallel to the α- and γ-axes. They have a relative retardation, n, which is to be determined as a function of φ, θ, and n_0, where n_0 is the retardation for light entering the crystal from the direction $\varphi=0$.

A consideration of the isochromatic surfaces of

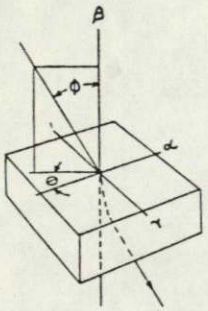


FIG. 3. Off-axis ray in the crystal coordinate system.

biaxial crystals[4] leads to the conclusion that the equations of the curves of constant retardation, n (written in terms of φ and θ), represent hyperbolas if terms in the fourth and higher powers of φ are neglected. Their transverse axes are along the α-axis for $n/n_0 \gtrless 1$ and along the γ-axis for $n/n_0 \lessgtr 1$ for crystals in which $\alpha\gamma-\beta^2 \gtrless 1$. The asymptotes are the lines

$$\tan^2\theta = \alpha/\gamma. \qquad \text{(III.1)}$$

Lyot's equations give the squares of the semi-transverse axes, which are:

$$\varphi_0^2 = \left(\frac{n}{n_0}-1\right)\frac{\gamma}{k} \quad \text{in the plane} \quad \theta = 0,$$
$$\varphi_{\pi/2}^2 = \left(\frac{n}{n_0}-1\right)\frac{\alpha}{k} \quad \text{in the plane} \quad \theta = \pi/2, \qquad \text{(III.2)}$$

where

$$k = \frac{\alpha\gamma-\beta^2}{2(\gamma-\alpha)\beta^2}. \qquad \text{(III.3)}$$

We have, therefore, sufficient information to determine both sets of hyperbolas, which can be represented by a single equation

$$n = n_0\left[1+\varphi^2 k\left(\frac{\cos^2\theta}{\gamma}-\frac{\sin^2\theta}{\alpha}\right)\right]. \qquad \text{(III.4)}$$

The exact expression for n in uniaxial crystals is readily derived by a straightforward application of Huygens' principle and analytic geometry. Consider a plane-parallel uniaxial crystal in a rectangular x, y, z coordinate system with the origin in the first surface. Let it be oriented with its surfaces normal to the z axis. Let the x axis be parallel to the crystal optic axis (i.e., parallel to the α-axis in negative crystals or to the γ-axis in positive crystals). Choose units of time and distance to make the velocity of light in space unity. The equation of an entering plane light wave is then:

$$ax+by+cz-t=0, \qquad \text{(III.5)}$$

where a, b, and c are the direction cosines of the normal to the wave front and t is the time.

As the wave passes the origin in entering the crystal, it initiates a secondary wavelet which expands into an ellipsoid with the equation:

$$\zeta^2x^2+\eta^2y^2+\upsilon^2z^2-t^2=0, \qquad \text{(III.6)}$$

where ζ, η, and υ are reciprocals of the velocities along the x, y, and z directions, respectively.

At a given instant, that portion of the plane wave which is inside the crystal, coincides with a plane tangent to the ellipsoid of Eq. (III.6) and containing the line of intersection of the plane wave of Eq. (III.5) with the first surface of the crystal (i.e., the plane $z=0$). The tangent plane through the point (x_1, y_1, z_1) on the ellipsoid is

$$x_1\zeta^2x+y_1\eta^2y+z_1\upsilon^2z-t^2=0. \qquad \text{(III.7)}$$

The lines of intersection of the planes of Eqs. (III.5) and (III.7) with the first surface of the crystal are, respectively,

$$ax+by-t=0, \quad z=0, \qquad \text{(III.8)}$$

and

$$x_1\zeta^2x+y_1\eta^2y-t^2=0, \quad z=0. \qquad \text{(III.9)}$$

These two lines must coincide. Hence:

$$x_1=(a/\zeta^2)t \quad \text{and} \quad y_1=(b/\eta^2)t. \qquad \text{(III.10)}$$

Since (x_1, y_1, z_1) must be a point on the ellipsoid of Eq. (III.6), we find for z_1:

$$z_1=\frac{t}{\upsilon}\left(1-\frac{a^2}{\zeta^2}-\frac{b^2}{\eta^2}\right)^{\frac{1}{2}}. \qquad \text{(III.11)}$$

Equations (III.10) and (III.11) define the path of a ray through the origin.

Let d be the thickness of the crystal in the z direction. The time, t_1, when a ray through the origin reaches the second surface is, then:

$$t_1=d\upsilon/[1-(a^2/\zeta^2)-(b^2/\eta^2)]^{\frac{1}{2}}. \qquad \text{(III.12)}$$

On emerging from the crystal the plane wave is parallel to the entering wave, with the equation:

$$ax+by+cz-(t-\Delta)=0. \qquad \text{(III.13)}$$

At time, t_1, this plane must contain the point (x_1, y_1, d). Hence,

$$\Delta=t_1-(ax_1+by_1+cd).$$

The distance, p, of the plane wave of Eq. (III.13) from the origin is therefore:

$$p=t-\Delta=t-t_1+ax_1+by_1+cd, \qquad \text{(III.14)}$$

or, from Eqs. (III.10) and (III.12):

$$p=t-d\left[\upsilon\left(1-\frac{a^2}{\zeta^2}-\frac{b^2}{\eta^2}\right)^{\frac{1}{2}}-c\right]. \qquad \text{(III.15)}$$

[4] T. Preston, *Theory of Light*, 3rd Edition (MacMillan Company, Ltd., London, 1901), p. 397.

Now, for the extraordinary wave

$$\zeta=\omega, \quad \eta=\upsilon=\epsilon,$$

and for the ordinary wave

$$\zeta=\eta=\upsilon=\omega.$$

Hence, the distances of the extraordinary and ordinary waves from the origin after their traversal of the crystal can be written, respectively:

$$\begin{aligned} p_\epsilon &= t-d\left[\epsilon\left(1-\frac{a^2}{\omega^2}-\frac{b^2}{\epsilon^2}\right)^{\frac{1}{2}}-c\right], \\ p_\omega &= t-d\left[\omega\left(1-\frac{a^2+b^2}{\omega^2}\right)^{\frac{1}{2}}-c\right]. \end{aligned} \tag{III.16}$$

The retardation, n, is simply $(p_\omega-p_\epsilon)/\lambda$, or:

$$n=\frac{n_0}{\epsilon-\omega}\left[\epsilon\left(1-\frac{a^2}{\omega^2}-\frac{b^2}{\epsilon^2}\right)^{\frac{1}{2}}-\omega\left(1-\frac{a^2+b^2}{\omega^2}\right)^{\frac{1}{2}}\right]. \tag{III.17}$$

Equation (III.17) is the exact expression for the off-axis effect in uniaxial crystals. It is readily reduced to the more convenient approximation of Eq. (III.4). Expanding the radicals, and neglecting fourth and higher powers of a and b, we find:

$$n=\frac{n_0}{\epsilon-\omega}\left[\epsilon-\omega-\frac{\epsilon}{2\omega^2}a^2-\frac{1}{2\epsilon}b^2+\frac{1}{2\omega}(a^2+b^2)\right]. \tag{III.18}$$

The direction cosines can be expressed in terms of φ and θ by the transformation:

$$a=\sin\varphi\sin\theta'; \quad b=\sin\varphi\cos\theta',$$

where

$$\theta'=\theta \quad \text{if} \quad \epsilon-\omega>0,$$

and

$$\theta'=\theta+\pi/2 \quad \text{if} \quad \epsilon-\omega<0.$$

Equation (III.18) becomes, then:

$$n=n_0\left[1+\frac{\varphi^2}{2\omega}\left(\frac{\cos^2\theta'}{\epsilon}-\frac{\sin^2\theta'}{\omega}\right)\right]. \tag{III.19}$$

Equation (III.19) is identical with Eq. (III.4) for uniaxial crystals.

The corresponding exact equation for biaxial crystals can be derived by the same methods, but the resulting expressions become so lengthy and complicated that it has not seemed worth while to push them through. The accuracy of Eq. (III.4) is adequate for all practical purposes whether uniaxial or biaxial crystals are considered.

It should be noted that the use of the equations of isochromatic surfaces in the derivation of off-axis effects does not lead to an exact result, since they are derived on the inexact assumption that the two components of light polarized at right angles traverse the crystal along identical paths.

The usable field of a given filter is determined by the maximum permissible value of $|n-n_0|$ for the thickest b-element.

The maximum permissible angle of incidence in the Climax filter in the $\theta=\pi/2$ plane is

$$\varphi=0.025 \text{ radian},$$

if we require that over the field

$$|n-n_0|\leqslant 0.1$$

for the thickest b-element.

IV. LYOT'S WIDE FIELD FILTERS

The maximum total flux from a given light source that can be squeezed through a filter is roughly proportional to the square of the product of the filter aperture and the maximum usable value of φ. The aperture is limited by the sizes of available birefringent crystals, and it is therefore important to find means for obtaining large fields. The most obvious device is to find a birefringent material for which k is very small. Although the author knows of no such material which is available in useful sizes of optical quality, this is a definite possibility which should be investigated further.

Lyot[3] has described three wide-field filters with compound elements made of available materials. They will be referred to as Lyot's first-type, second-type, and third-type filters.

The first-type filter differs from the simple filter in having each b-element divided into two equal halves by a cut perpendicular to the β-axis. The second half of each element is rotated about the β-axis until the α-axes of the two components are crossed. A half-wave plate is inserted between the components with its α-axis at 45° to the α-axes of the two. It serves to rotate the planes of polarization 90°. Light which enters the first component from the direction (φ, θ) enters the second component from the direction $(\varphi, \theta+\pi/2)$. The retardation introduced by the assembled element is then:

$$\begin{aligned} n &= \tfrac{1}{2}\left[n(\varphi,\theta)+n\left(\varphi,\theta+\frac{\pi}{2}\right)\right] \\ &= \tfrac{1}{2}n_0\left[1+\varphi^2k\left(\frac{\cos^2\theta}{\gamma}-\frac{\sin^2\theta}{\alpha}\right)\right] \\ &\quad +\tfrac{1}{2}n_0\left[1+\varphi^2k\left(\frac{\sin^2\theta}{\gamma}-\frac{\cos^2\theta}{\alpha}\right)\right], \end{aligned}$$

or

$$n=n_0\left[1+\varphi^2\frac{k}{2}\left(\frac{1}{\gamma}-\frac{1}{\alpha}\right)\right]. \tag{IV.1}$$

The loci of constant retardation are circles with radii larger than the axes of the hyperbolas of a simple filter (in the $\theta=\pi/2$ plane) by a factor of

$[2\gamma/(\gamma-\alpha)]^{\frac{1}{2}}$. For a given tolerable value of $|n-n_0|$, the radius of the useful field can be further increased by a factor of $\sqrt{2}$ if we set the retardation at the center of the field at one extreme of the range.

Lyot's first-type filter, unlike the simple filter, can be used only over a small range of wave-lengths. If the wave-length differs greatly from the optimum for which the half-wave plates are made, the residual light between transmission bands increases at the expense of light in the bands. The added residual light appears superposed on the field in the form of faint hyperbolic fringes very similar to the fringes produced by the equivalent simple filter. The fringes are loci of constant retardation, n', given by

$$n'=\frac{n_0}{2}k\varphi^2\left(\frac{1}{\gamma}+\frac{1}{\alpha}\right)\cos 2\theta. \qquad \text{(IV.2)}$$

If, however, the filter is either used for one wave-length only, or supplied with interchangeable half-wave plates for the different spectral regions, its performance is highly satisfactory. This is one of the many instances where the development of an achromatic half-wave plate would be very useful.

Lyot's second-type wide-field filter has compound b-elements of two components of different materials. The quantity k is of opposite sign in the two components, which are mounted with their α-axes parallel. No half-wave plates are required.

Let n_1 and n_2 be the retardations arising from the first and second components for light entering from the direction $\varphi=0$. The retardation for the assembled element is then:

$$n=n_1\left[1+\varphi^2k_1\left(\frac{\cos^2\theta}{\gamma_1}-\frac{\sin^2\theta}{\alpha_1}\right)\right]+n_2\left[1+\varphi^2k_2\left(\frac{\cos^2\theta}{\gamma_2}-\frac{\sin^2\theta}{\alpha_2}\right)\right],$$

$$n=n_0+\varphi^2\left[\cos^2\theta\left(\frac{n_1k_1}{\gamma_1}+\frac{n_2k_2}{\gamma_2}\right)-\sin^2\theta\left(\frac{n_1k_1}{\alpha_1}+\frac{n_2k_2}{\alpha_2}\right)\right], \qquad \text{(IV.3)}$$

where now

$$n_0=n_1+n_2. \qquad \text{(IV.4)}$$

It is evident that while the coefficient of φ^2 cannot be made to vanish by any choice of n_1 and n_2, we can obtain circular fringes by eliminating θ. The condition is

$$\frac{n_1}{n_2}=-\left(\frac{k_2}{k_1}\right)\left[\left(\frac{1}{\gamma_2}+\frac{1}{\alpha_2}\right)\Big/\left(\frac{1}{\gamma_1}+\frac{1}{\alpha_1}\right)\right]. \qquad \text{(IV.5)}$$

Equations (IV.4) and (IV.5) give n_1 and n_2. The retardation of the assembled element can now be written

$$n=n_0+\varphi^2\left(\frac{n_1k_1}{\gamma_1}+\frac{n_2k_2}{\gamma_2}\right). \qquad \text{(IV.6)}$$

The second-type filter can be used over a wide range of wave-lengths, although the fringes do not remain strictly circular throughout the range.

Lyot's third-type filter generally has the largest field. Each b-element consists of three birefringent components. Two of the components are of the same material and are mounted with their α-axes crossed. The third is of a different birefringent material with a k value opposite in sign to the k value for the first two components. It is mounted with its α-axis parallel to that of one of the first two. By a proper choice of thicknesses it is always possible to make n constant over the whole field within the accuracy of Eq. (III.4).

Let α_1, β_1, γ_1 and α_2, β_2, γ_2 be the refractive indices of the crystals composing the single component and the two crossed components, respectively. The crystals must be selected to satisfy the condition

$$\alpha_1\gamma_2>\gamma_1\alpha_2.$$

Let n_a be the retardation of the single component, and n_b and n_c the retardations of the two components of the same material. Let the α-axes of the a and b components be in the $\theta=0$ plane, and the α-axis of the c component in the $\theta=\pi/2$ plane. Then

$$n=n_a\left[1+\varphi^2k_1\left(\frac{\cos^2\theta}{\gamma_1}-\frac{\sin^2\theta}{\alpha_1}\right)\right]+n_b\left[1+\varphi^2k_2\left(\frac{\cos^2\theta}{\gamma_2}-\frac{\sin^2\theta}{\alpha_2}\right)\right]-n_c\left[1+\varphi^2k_2\left(\frac{\sin^2\theta}{\gamma_2}-\frac{\cos^2\theta}{\alpha_2}\right)\right]. \qquad \text{(IV.7)}$$

If we set $n_a+n_b-n_c=n_0$, and require that the coefficient of φ^2 vanish, we find

$$n_a=\frac{n_0'}{A}k_2^2\left(\frac{1}{\alpha_2^2}-\frac{1}{\gamma_2^2}\right),$$

$$n_b=\frac{n_0}{A}k_1k_2\left(\frac{1}{\gamma_1\gamma_2}-\frac{1}{\alpha_1\alpha_2}\right), \qquad \text{(IV.8)}$$

$$n_c=\frac{n_0}{A}k_1k_2\left(\frac{1}{\alpha_1\gamma_2}-\frac{1}{\gamma_1\alpha_2}\right),$$

where

$$A=\begin{vmatrix} \frac{k_1}{\gamma_1} & \frac{k_2}{\gamma_2} & \frac{k_2}{\gamma_2} \\ \frac{k_1}{\alpha_1} & \frac{k_2}{\gamma_2} & \frac{k_2}{\alpha_2} \\ 1 & -1 & 1 \end{vmatrix}. \qquad \text{(IV.9)}$$

The retardation of the assembled element for any direction (φ, θ) is then:

$$n=n_a+n_b-n_c=n_0. \qquad \text{(IV.10)}$$

The third-type filter, like the second-type, can be used over a wide range of wave-lengths. The coefficient of φ^2, however, will generally vanish accurately at only one wave-length.

The design of a wide angle filter does not necessarily require that the thinner elements be compound. Their transmission bands are so broad that the slight shift in wave-length for off-axis rays is negligible in comparison. If the higher order compound elements are made of two materials, however, it may not be possible to use transmission bands in widely separated regions of the spectrum, because the dispersions of different materials are generally not strictly proportional. If the rth element is the thickest simple element, $(n_{r+1})/n_r=2$ at only one wave-length.

The following sections are devoted to the theory of various modifications of birefringent filters which have been recently developed.

V. THE SPLIT ELEMENT FILTER

The split-element filter resembles Lyot's first-type filter, and shares its wide field characteristics. The half-wave plates, however, are replaced by birefringent elements, and successive polarizers are crossed. After the initial polarization, it requires only half as many polarizers as the equivalent simple filter. The result is a considerable reduction in absorption and scattered light if film polarizers are used, or a notable saving in bulk and expense if polarizing prisms are used.

The split-element filter has already been described briefly.[5] A more detailed account of its theory is given here.

A single unit of the split-element filter (which would be mounted between crossed polarizers) is shown schematically in Fig. 4. The x, y, and z axes constitute a rectangular coordinate system. The positive r and s axes in the xy plane bisect the angles between the positive x and y and the positive y and negative x directions, respectively. The unit

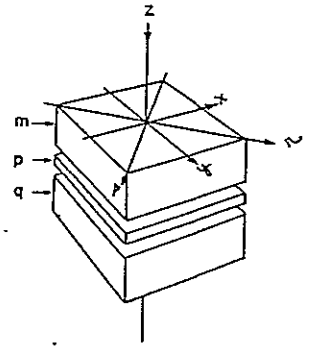


FIG. 4. Birefringent components of a single unit of a split element filter.

consists of a split element with components m and q, and a simple element, p, sandwiched between m and q. They are all mounted with β-axes parallel to the z axis. The γ-axes are aligned parallel to the x, r, and y directions, respectively, in the m, p, and q components. Let the thicknesses of m, p, and q be d_m, d_p, and d_q, and let the unit of time be the vibration period of the light.

Assume that the entering light is polarized in the r plane. The transmissions of the unit for emerging light polarized in the r plane and s plane are to be determined.

The vibration of the entering light is

$$r=a\sin 2\pi t. \qquad \text{(V.1)}$$

This can be resolved along the x and y directions giving

$$x=(a/\sqrt{2})\sin 2\pi t, \quad y=(a/\sqrt{2})\sin 2\pi t. \qquad \text{(V.2)}$$

In traversing m, a phase difference is introduced and the vibration of the emerging light is

$$\begin{aligned} x_m&=(a/\sqrt{2})\sin 2\pi(t-d_m\gamma), \\ y_m&=(a/\sqrt{2})\sin 2\pi(t-d_m\alpha). \end{aligned} \qquad \text{(V.3)}$$

The resultant disturbance along the r and s axes is:

$$\begin{aligned} r_m&=a\cos\pi n_m\sin 2\pi t', \\ s_m&=a\sin\pi n_m\cos 2\pi t', \end{aligned} \qquad \text{(V.4)}$$

where

$$t'=t-(d_m/2\lambda)(\alpha+\gamma):$$

In the traversal of p, an additional phase difference is introduced;

$$\begin{aligned} r_p&=a\cos\pi n_m\sin 2\pi[t'-(d_p/\lambda)\gamma], \\ s_p&=a\sin\pi n_m\cos 2\pi[t'-(d_p/\lambda)\alpha]. \end{aligned} \qquad \text{(V.5)}$$

Resolving this vibration along the x and y axes, and adding the phase difference due to transmission

[5] J. Evans, Ciencia e Investigación (Argentina) Vol. III, No. 9, p. 365 (1947).

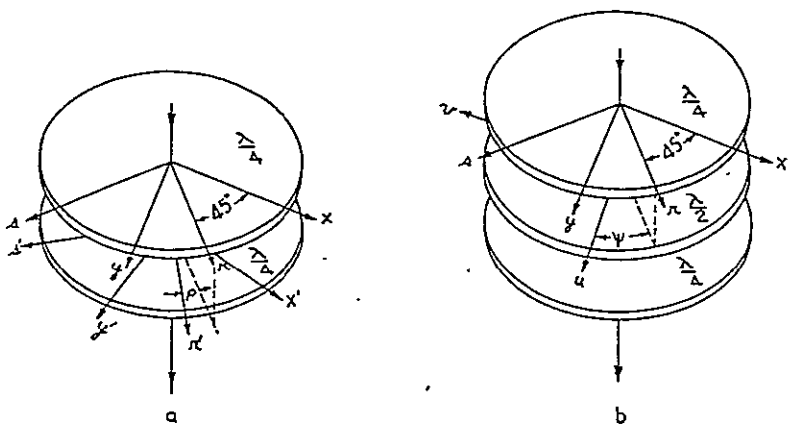


FIG. 5. (a) Phase shifter of two quarter-wave plates. (b) Phase shifter of one half-wave and two quarter-wave plates.

through q, we obtain

$$x_q=\frac{a}{\sqrt{2}}\cos\pi n_m\sin2\pi\left(t'-\frac{d_p}{\lambda}\gamma-\frac{d_q}{\lambda}\alpha\right)$$
$$-\frac{a}{\sqrt{2}}\sin\pi n_m\cos2\pi\left(t'-\frac{d_p}{\lambda}\alpha-\frac{d_q}{\lambda}\alpha\right),$$
$$y_q=\frac{a}{\sqrt{2}}\cos\pi n_m\sin2\pi\left(t'-\frac{d_p}{\lambda}\gamma-\frac{d_q}{\lambda}\gamma\right)$$
$$+\frac{a}{\sqrt{2}}\sin\pi n_m\cos2\pi\left(t'-\frac{d_p}{\lambda}\alpha-\frac{d_q}{\lambda}\gamma\right). \quad (V.6)$$

To determine the final transmission through a polarizer with its axis along either the r or s direction, we must resolve this vibration along the r and s axes:

$$r_q=a\cos\pi n_m\cos\pi n_q\sin2\pi[t''-(d_p/\lambda)\gamma]$$
$$+a\sin\pi n_m\sin\pi n_q\sin2\pi[t''-(d_p/\lambda)\alpha],$$
$$s_q=a\sin\pi n_m\cos\pi n_q\cos2\pi[t''-(d_p/\lambda)\gamma]$$
$$-a\cos\pi n_m\sin\pi n_q\cos2\pi[t''-(d_p/\lambda)\alpha], \quad (V.7)$$

where

$$t''=t-[(d_m+d_q)/2\lambda](\alpha+\gamma).$$

Let the emergent amplitudes be A_r and A_s. The transmissions in the r and s vibration planes are, then:

$$\tau_r=A_r^2/a^2=\cos^2\pi(n_m-n_q)$$
$$-\sin2\pi n_m\sin2\pi n_q\sin^2\pi n_p,$$
$$\tau_s=A_s^2/a^2=\sin^2\pi(n_m-n_q)$$
$$+\sin2\pi n_m\sin2\pi n_q\sin^2\pi n_p. \quad (V.8)$$

In the split-element filter the m and q components are made of equal thicknesses. Hence

$$n_m=n_q.$$

If we let

$$n_j=2n_m=2n_q,$$

Eqs. (V.8) reduce to:

$$\tau_r=1-\sin^2\pi n_j\sin^2\pi n_p,$$
$$\tau_s=\sin^2\pi n_j\sin^2\pi n_p. \quad (V.9)$$

The transmission of an element of Lyot's first-type filter is τ_r in Eqs. (V.9) if we set $n_p=\frac{1}{2}$.

The transmission of a unit of the split-element filter is τ_s. A split-element filter of l elements has exactly the same off-axis characteristics as a filter of Lyot's first type with the first $l/2$ elements simple, and the $l/2$ thicker elements compound. Whether the field is limited by the simple elements or the compound elements depends upon whether or not $(n_l/n_{l/2})[(\gamma-\alpha)/(2\gamma)]$ is greater or less than 1. If the simple elements limit the field, they can, of course, be made compound in any of Lyot's three types.

The transmission of an assembled split-element filter composed of two-element units between crossed polarizers is:

$$\tau=\sin^2\pi n_1\sin^2\pi n_2\cdots\sin^2\pi n_l. \quad (V.10)$$

Since transmission bands occur only at wave-lengths for which all the n's are half-integral, the n's cannot be simply proportional to the powers of 2. If we let $n=n'+\frac{1}{2}$ at the wave-length of a particular band, the best we can do is to make the values of n' proportional to the powers of 2. Thus

$$n_r=2^{r-1}n_1'+\tfrac{1}{2}. \quad (V.11)$$

The transmission can then be written

$$\tau=\cos^2\pi n_1'\cos^2\pi 2n_1'\cdots\cos^2\pi 2^{l-1}n_1'. \quad (V.12)$$

Unfortunately Eq. (V.11) can be strictly valid at only one wave-length, and the usefulness of the filter is restricted to a limited spectral region in the neighborhood of that wave-length. This is a second instance where achromatic half wave plates would be useful. If the r-th element of the filter were made to give a retardation $n_r'=2^{r-1}n_1'$, the addition of an achromatic half-wave plate (two quarter-wave plates for split elements) would satisfy Eq. (V.11) at all wave-lengths.

The thought will doubtless have occurred to the reader that the middle element in each unit of a split-element filter could itself be split, and a third

element inserted between the halves. This plan does not work theoretically, and so far no arrangement has been found that allows more than two elements in a unit between successive polarizers.

VI. FILTERS OF ADJUSTABLE WAVE-LENGTH

It is obvious that the usefulness of the birefringent filter is enormously enhanced if a transmission maximum can be adjusted to center on any desired wave-length. The fine adjustment resulting from the control of temperature is generally quite inadequate as it has a range of only a few angstroms (although Lyot found that with the aid of temperature control it is possible to bring no less than six of the maxima of a quartz filter into coincidence with lines of major importance in the solar spectrum).

The obvious method of controlling the wave-length of the transmission bands is by means of elements of variable thickness, made of pairs of wedges which can be adjusted with respect to each other like the components of a Babinet compensator. It is then possible to set

$$n_1 = \text{an integer},$$

and

$$n_r = 2^{r-1} n_1$$

at any chosen wave-length. Such an arrangement is perfectly feasible and works equally well at all wave-lengths. In the split-element filter, both halves of the split element must, of course, be adjustable since $n_m - n_q = 0$. The range of variation in thickness need be only sufficient to shift the principal transmission maxima of the filter through a range equal to their separation. With a proper choice of wedge angles all the movable wedges can be mounted and adjusted as a single unit.

Although theoretically excellent, the variable-thickness filter requires considerable mechanical refinement, and one wedge in each element must have an aperture much larger than the instrumental aperture (a matter of importance in filters of large aperture). The use of phase shifters for wave-length adjustment is simpler and, for most purposes, equally satisfactory. If achromatic phase shifters can be devised, they will give results as theoretically perfect as variable thickness.

Suppose we equip each b-element of a filter with a phase shifter which permits the addition of a small controllable phase difference, $2\pi\xi$, to the phase difference, $2\pi n$, introduced by the b-element. The transmission of the filter is then

$$\tau = \prod_{r=1}^{r=l} \cos^2 \pi (n_r + \xi_r). \qquad \text{(VI.1)}$$

Again, with the split-element filter, the added phase difference must be divided equally between the two halves of the split elements to keep $n_m - n_q = 0$. A transmission maximum of the filter can then be centered on any given wave-length, λ_1, by adjusting ξ until $n + \xi$ is an integer for each element. This is always possible if ξ can be adjusted over the range $-\frac{1}{2}$ to $+\frac{1}{2}$. If the phase shifter is achromatic, i.e., ξ is independent of wave-length at a given setting, the result is merely a shift of the transmission curve of the filter along the spectrum and its performance is equally good at all wave-length settings. If, on the other hand, ξ is a function of wave-length, the spacings of the transmission maxima of a given element are altered. Hence the relative positions of the transmission maxima and minima of the different elements depart more and more from exact superposition as the wave-length departs from λ_1. The result is an increase in the residual light transmitted in the intervals between principal maxima of the filter as $|\lambda - \lambda_1|$ increases.

Lyot[3] and Billings[6] have both made numerical calculations of the additional residual light resulting from the use of non-achromatic phase shifters. They concluded that over a reasonable wave-length range (which can readily be isolated with glass or gelatine filters) the increase in residual light is negligible. The adjustment of wave-length with phase shifters is therefore a practical possibility whether the phase shifters are achromatic or not.

Several forms of variable phase shifters have been proposed.

Lyot[3] made elements of variable thickness like

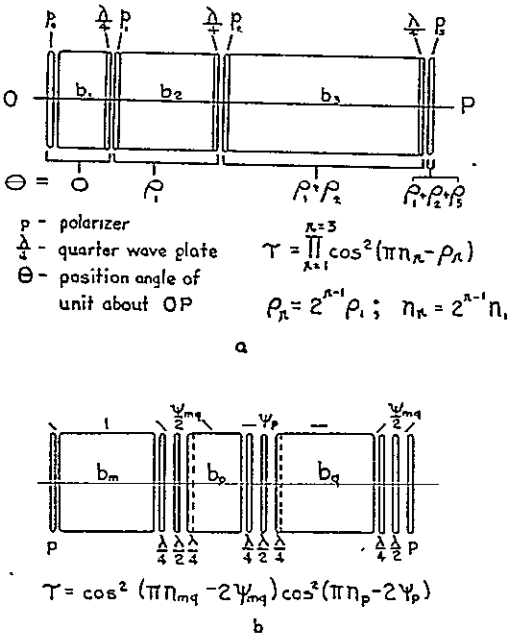


FIG. 6. (a) Simple filter of three elements with quarter-wave plate phase shifters. (b) One unit of a split element filter with fractional wave plate phase shifters.

[6] Bruce H. Billings, J. Opt. Soc. Am. 37, 738 (1947).

those described above for the variable thickness filter, but with the difference that the range of adjustment of retardation was restricted to one wave-length.

Billings[6] made an experimental filter with photo-elastic phase shifters composed of sheets of polyvinyl butyrate under adjustable tension.

While both these arrangements give a satisfactory wave-length adjustment, they are tedious to use. Ordinarily each element must be individually adjusted. The alternative is a complicated mechanical synchronization of the adjustments of all the elements, which would make operation with a single control feasible. Without some such arrangement it would be impossible to vary the wave-length continuously.

A much more promising approach is the use of the electro-optical phase shifters discussed by Billings.[6] A plate of the uniaxial crystal ammonium di-hydrogen phosphate ($NH_4H_2PO_4$), known commercially as PN, cut perpendicular to the optic axis and mounted between transparent electrodes, becomes biaxial and exhibits a retardation when a potential difference is applied to the electrodes. The retardation is proportional to the potential difference and is independent of the thickness of the PN plate. A filter made with a Billings plate added to each element (to each half of the split elements in the split element filter) could be adjusted electrically, and the problem of synchronizing the phase shifts of successive elements would be relatively simple. At the present writing Dr. Billings is actively engaged in the development of such electrically tunable filters.

All three tuning methods have one difficulty in common. It is impracticable to push the phase shift beyond a very limited range. If a range from $-\pi$ to $+\pi$ is adopted, a continuous variation of wave-length involves a discontinuous adjustment of each phase shifter. The phase shift must progress smoothly from $-\pi$ to $+\pi$ (at a rate proportional to the thickness of the associated b-element) and then jump back to $-\pi$. For most purposes there may be no serious disadvantage in this. If, however, the filter is to be used for spectrophotometric work, for example, it may be very difficult to avoid a spurious bump in the filter transmission every time a phase shifter passes a point of discontinuity, even with the electrical tuning. For such special purposes phase shifters composed of rotating fractional wave plates can be used. They have already been described briefly.[5] A fuller account of their theory is given here.

The specific problem is to devise a combination of fractional wave plates which will alter the phase difference between the vibrations along two mutually perpendicular axes, x and y, by any chosen amount, without altering their amplitudes. At a given wave-length, this is equivalent to a variable thickness of birefringent material with its γ-axis along the x or y direction. Such an arrangement is shown at a, Fig. 5. It consists of two quarter-wave plates. The first is fixed with its γ-axis along the r axis (at 45° to the x axis). The second can be rotated around the instrumental axis. At a given setting its γ-axis lies along the r' direction at angle ρ to the r direction.

The vibration of the light entering the system is generally represented by

$$x=b\sin 2\pi t, \quad y=c\sin 2\pi(t+\sigma). \tag{VI.2}$$

Resolving this vibration along the r and s axes and adding a phase difference of $\pi/2$ introduced by the first quarter-wave plate we obtain for the emerging vibration:

$$\begin{aligned} r&=(b/\sqrt{2})\sin 2\pi t+(c/\sqrt{2})\sin 2\pi(t+\sigma),\\ s&=-(b/\sqrt{2})\cos 2\pi t+(c/\sqrt{2})\cos 2\pi(t+\sigma). \end{aligned} \tag{VI.3}$$

Resolving this vibration along the r' and s' axes and adding another phase difference of $\pi/2$ introduced by the second quarter-wave plate, we obtain,

$$\begin{aligned} r'&=(b/\sqrt{2})\sin[2\pi t-\rho]\\ &\quad+(c/\sqrt{2})\sin[2\pi(t+\sigma)+\rho],\\ s'&=(b/\sqrt{2})\sin[2\pi t-\rho]\\ &\quad-(c/\sqrt{2})\sin[2\pi(t+\sigma)+\rho]. \end{aligned} \tag{VI.4}$$

Finally, if we resolve this vibration along the x' and y' axes, at an angle of $\rho+(\pi/2)$ to the x and y axes, we obtain for the emerging vibration:

$$\begin{aligned} x'&=b\sin[2\pi t-\rho],\\ y'&=c\sin[2\pi(t+\sigma)+\rho+\pi]. \end{aligned} \tag{VI.5}$$

A comparison of Eqs. (VI.2) with (VI.5) shows that while the emerging amplitudes along x' and y' are the same as the entering amplitudes along x and y, the phase difference has been increased from $2\pi\sigma$ to $2\pi\sigma+2\rho+\pi$, i.e., the phase shift, $2\pi\xi$, is

$$2\pi\xi=\pi+2\rho. \tag{VI.6}$$

Obviously the phase difference can be set to any desired value by adjusting ρ.

This two-element phase shifter has the disadvantage that the x' and y' axes rotate with the second quarter-wave plate. For some applications this is no inconvenience but in others it renders this phase shifter useless. The x' and y' axes can be restored to parallelism with the x and y axes by the addition of a rotatable half-wave plate, which has the property of reflecting any polarization figure in its γ-axis.

The most convenient system, shown at b, Fig. 5, consists of two fixed quarter-wave plates with the rotatable half-wave plate sandwiched between them. Suppose the γ-axes of both quarter-wave plates are in the r direction, while the γ-axis of the

half-wave plate is along the u direction at an angle ψ to the r direction.

The vibration emerging from the first quarter-wave plate is given by Eq. (VI.3). Resolving this vibration along the u and v axes, and adding a phase difference of π, we obtain for the vibration emerging from the half-wave plate:

$$\begin{aligned} u &= (b/\sqrt{2}) \sin[2\pi t-\psi] \\ &\quad +(c/\sqrt{2}) \sin[2\pi(t+\sigma)+\psi], \\ v &= (b/\sqrt{2}) \cos[2\pi t-\psi] \\ &\quad -(c/\sqrt{2}) \cos[2\pi(t+\sigma)+\psi]. \end{aligned} \quad \text{(VI.7)}$$

Resolving this vibration again along the r and s axes, and adding a phase difference of $\pi/2$, we obtain for the vibration emerging from the second quarter-wave plate:

$$\begin{aligned} r &= (b/\sqrt{2}) \sin[2\pi t-2\psi] \\ &\quad +(c/\sqrt{2}) \sin[2\pi(t+\sigma)+2\psi], \\ s &= -(b/\sqrt{2}) \sin[2\pi t-2\psi] \\ &\quad +(c/\sqrt{2}) \sin[2\pi(t+\sigma)+2\psi]. \end{aligned} \quad \text{(VI.8)}$$

Finally, resolving this vibration along the original x and y axes, we find

$$\begin{aligned} x &= b \sin[2\pi t-2\psi], \\ y &= c \sin[2\pi(t+\sigma)+2\psi]. \end{aligned} \quad \text{(VI.9)}$$

The phase shift introduced by the three-element system is, therefore,

$$2\pi\xi = 4\psi. \quad \text{(VI.10)}$$

The principal advantage in the use of fractional wave plate phase shifters in birefringent filters is in the possibility of a continuous variation of wave-length without discontinuities in the adjustment of the moving elements. Since ρ or ψ can be increased or decreased indefinitely, $2\pi\xi$ is not restricted as it is in the other types of phase shifters discussed above.

It should be noted that the fractional wave plate phase shifter is in a sense achromatic, since ξ is independent of the wave-length for a given value of ρ or ψ—a very desirable property (see the discussion following Eq. (VI.1)). With ordinary quarter- and half-wave plates, however, this advantage is somewhat illusory. Their usefulness is limited to the rather restricted region of the spectrum where their retardations are very close to quarter-wave and half-wave. This is another application where the desirability of achromatic fractional wave plates is evident.

If continuity of adjustment over a large range of the spectrum is a necessity, the fractional wave plates themselves could be made adjustable. The addition of an electro-optical Billings plate to each fractional wave plate would perhaps be the simplest method. A relatively moderate potential applied to the Billings plate would then adjust the retardation accurately to a half-wave or quarter-wave at the wave-length of the transmission band of the filter. This seems a rather desperate measure, however.

The construction of the fractional wave plate phase shifters is considerably simplified when they are used in birefringent filters. Some of the quarter-wave plates simply take the form of an addition to the thickness of the birefringent elements. In instances where the γ-axis of a quarter-wave plate is parallel or perpendicular to the axis of an immediately following polarizer, it is evident that the polarizer utilizes only one component of the vibration emerging from the quarter-wave plate. The $\pi/2$ phase difference therefore serves no real purpose, and the quarter-wave plate can be omitted.

Consider first an element of a simple filter. Suppose the b-element, oriented with its γ-axis along the x direction, is followed by a quarter-wave plate with its γ-axis along the r direction. If we let $b=c=a/\sqrt{2}$, $t=t'-(d/2\lambda)\mu$, and $\sigma=(d/\lambda)\mu$, Eq. (VI.3) for the vibration emerging from the quarter-wave plate reduces to:

$$r = a \cos\pi n \sin 2\pi t', \quad s = a \sin\pi n \sin 2\pi t'. \quad \text{(VI.11)}$$

This is a linear vibration at an angle of πn to the r-axis. We can omit the second quarter-wave plate and let the light enter a polarizer with its plane of polarization at angle ρ to the r axis. The transmission of the assembly is then

$$\tau = \cos^2(\pi n-\rho). \quad \text{(VI.12)}$$

By adjusting ρ (i.e., by rotating the polarizer) until $n_\lambda-\rho/\pi=$ an integer, we can set $\tau=1$ for any chosen wave-length.

Lyot[3] has utilized this device to effect a slight shift in the wave-length of the transmission band of his filter. He used a quarter-wave plate with the last (thickest) element, and provided for the rotation of the final polarizer. The same method can be applied to the whole filter, however.

An adjustable simple birefringent filter would consist, then, of a series of units shown at a, Fig. 6. each composed of a polarizer, a birefringent element with its γ-axis at 45° to the axis of the polarizer, and a quarter-wave plate with its γ-axis parallel to the axis of the polarizer. The three parts of each unit remain fixed with respect to each other, but the unit itself must be rotatable around the instrumental axis. The angle ρ_r is then the angle between the γ-axis of the rth quarter-wave plate and the axis of the immediately following polarizer. The birefringent elements have the same thickness as in the non-adjustable filter. The transmission of the whole is

$$\begin{aligned} \tau = \cos^2(\pi n_1-\rho_1) \cos^2(\pi 2n_1-\rho_2)\cdots \\ \times\cos^2(\pi 2^{t-1}n_1-\rho_t), \end{aligned} \quad \text{(VI.13)}$$

and

$$\rho_r = 2^{r-1}\rho_1. \quad \text{(VI.14)}$$

Since the values of ρ are proportional to the powers of 2, it is a relatively simple matter to devise a gear train by which the wave-length of the transmission band can be adjusted with a single control knob. A continuous variation of wave-length now involves no discontinuity in the adjustment of the various units, since ρ can be made to increase or decrease indefinitely.

Matters are somewhat more complicated in the split-element filter. The wide field characteristics depend upon the m and q components being crossed. Hence the phase shifts must be accomplished without any relative rotation of the two. Various arrangements are possible, some of which involve rotation of the center p-elements, or rotation of the unit as a whole with respect to the polarizers, or both. However, the unit shown at b, Fig. 6, is as simple as any.

The orientation of each element is indicated in the diagram by the short line above it for the fixed elements, or by the symbol $(\psi_{mq}/2)$ or ψ_p for the adjustable half-wave plates. The angle $(\psi_{mq}/2)$ or ψ_p is the angle between the γ-axis of the half-wave plate and the γ-axis of the preceding quarter-wave plate. The second quarter-wave plates following m and p are indicated as an addition to the thicknesses of p and q, while that following q has been omitted, since its γ-axis would be parallel to the axis of the following polarizer. The transmission of a split-element filter composed of such units is

$$\tau=\prod_{r=1}^{r=l}\cos^2\left(\pi n_r-2\psi_r-\frac{\pi}{2}\right). \qquad \text{(VI.15)}$$

It should be noted here that the built-in quarter-wave plates which are added to the thicknesses of the p- and q-elements are not included in the calculation of n for these elements.

The values of ψ_r should be proportional to n_r in Eq. (VI.15). Hence if $n_r=2^{r-1}(n_1-\frac{1}{2})+\frac{1}{2}$ as in the non-adjustable split-element filter, the ψ's are proportional to large odd numbers, and the problem of synchronizing the rotations of the half-wave plates becomes complicated (but not at all impossible). If, on the other hand, the n's are made proportional to the powers of two, the phase changers can compensate for the subtraction of $\frac{1}{2}$ from each value of n in addition to their normal function. Then

$$n_r=2^{r-1}n_1, \qquad \text{(VI.16)}$$

and

$$2\psi_r=\pi/2+2^{r-1}[2\psi_1-(\pi/2)]. \qquad \text{(VI.17)}$$

Since a rotation of the zero point from which angle ψ is measured to $\pi/4$ reduces this equation to

$$2\psi_r'=2^{r-1}(2\psi_1'), \qquad \text{(VI.18)}$$

it is evident that the variable parts of the ψ's are proportional to the powers of two, and the problem of synchronization becomes relatively simple.

The synchronization of the other types of phase shifters (variable thickness, photo elastic, or electro optical) is similarly simplified in a split element filter by constructing it with n's proportional to powers of two. Equations (VI.14) and (VI.17) apply if we substitute $\pi\xi$ for 2ψ.

A final remark about filters of adjustable wave-length seems worth while. The birefringent elements need not be made to any exact thicknesses as in the fixed wave-length filters. It is desirable, but not necessary, to preserve the relation $n_r=2^{r-1}n_1$ as closely as possible, since the synchronization of the various adjustments is then easier. There is no necessity, however, for n_1 to be an integer for any specified wave-length. This simplifies the construction somewhat. If $\mu\lesssim 0.03$, the thicknesses of the elements can be adjusted with sufficient accuracy by mechanical measurements alone. The error tolerance in thickness is inversely proportional to μ and is about ± 0.001 mm for $\mu=0.03$.

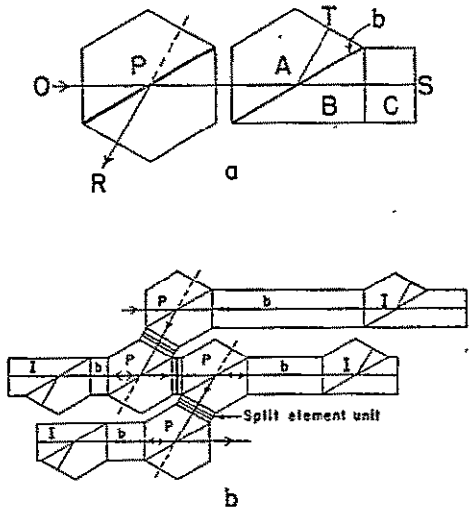


FIG. 7. (a) One form of polarizing interferometer. (b) High resolution filter composed of polarizing interferometers and birefringent elements.

VII. MATERIALS FOR BIREFRINGENT FILTERS

For the benefit of potential builders of birefringent filters, a brief discussion of available materials is given below. It must be emphasized that the list given is certainly far from complete. The author simply lists materials which have come to his attention and either have been successfully used, or look promising. Unfortunately, lack of time has prevented a really thorough search for suitable and available materials, and it would be surprising if some very useful ones had not been overlooked.

Some of the desirable properties of crystals for birefringent filters are a large value of μ with a small temperature coefficient; a high degree of hardness; chemical stability and insolubility in water; high transparency in the region of the spectrum for which the filter is to be used; and availability in large pieces of high optical quality.

For filters with band widths of 3 angstroms or more, quartz is an ideal material. It is excellent on all counts except for its rather small value of $\mu(=0.009)$. The birefringent elements of all the astronomical filters now in operation are made of quartz except for the final element of Lyot's filter, which is calcite.

Calcite would be excellent for elements of large n values if it were readily available in large sizes. Unfortunately it is so difficult to obtain that its general use in filters is probably impossible. While it is not as easily ground and polished as quartz, it presents no real difficulty. $\mu=0.17$.

Gypsum occurs naturally in large crystals and should be readily available. Its birefringence is similar to that of quartz, and it should be useful in the same places. Unfortunately, it is quite soft and might be difficult to polish. $\mu=0.009$.

Ammonium di-hydrogen phosphate has excellent optical characteristics, although it is sensitive to pressure and must be mounted with care. It is available in large sizes. Its optical working has proved rather difficult, though not impossible, and its high solubility in water necessitates careful protection from atmospheric moisture. $\mu=0.045$.

Ethylene diamine tartrate has promising optical characteristics accompanied by the disadvantages of high solubility in water and softness. The author knows of no attempts to polish it, but it would probably be quite difficult. It is available in large sizes. $\mu=0.084$.

Sodium nitrate has a larger μ-value than calcite, and should be useful for elements of large n-values. However, it is very soluble in water and difficult to work. At present it is not available in large sizes with the necessary homogeneity. $\mu=0.25$.

VIII. POLARIZING INTERFEROMETER FILTERS

An account of birefringent filters should not be closed without some mention of the polarizing interferometer, a device which has the effect of an impossibly thick birefringent element. It offers the possibility of filters of very high resolution with band widths in the range of hundredths or thousandths of an angstrom. The advantages of the polarizing over the usual forms of interferometers is in the possibility of an accurate and stable control of the wave-lengths of transmission maxima (by means of phase shifters) and a high light efficiency.

The essential feature of the polarizing interferometer is that the emerging light consists of two coherent sets of waves which differ in phase (because of path difference) and are polarized at right angles to each other. The effect is similar to that of a birefringent element, and a series of polarizing interferometers can be used exactly like a series of birefringent elements to construct a filter. The wave-length of the transmission band can be controlled with adjustable phase shifters, and interferometers can be sandwiched between birefringent elements to form split-element units.

The advantage of the polarizing interferometer over a simple birefringent element is that very large values of n can be obtained in a comparatively compact element. The saving in bulk may not be important, but the difficulty of obtaining birefringent material in very great thicknesses is significant. An element of calcite, for instance, must be about eleven times as thick as a path difference in glass. The principal disadvantage is the expense of construction, common to all interferometers of the split amplitude class. The field is small for large values of n, and while it is theoretically quite simple to make a birefringent field compensator, it is impractical because the thickness of birefringent material required nullifies the advantage of compactness.

Many forms of polarizing interferometers are possible. One type which is well adapted for the construction of filters is shown at a, Fig. 7. It is a modified solid Michelson interferometer with a polarizing beam splitter. It consists of two glass prisms, A and B, with a very thin slip, b, of sodium nitrate (or other highly birefringent material) cemented between them with its optic axis normal to the surface. If the angles are properly chosen, the b-layer totally reflects the light vibrating in the plane of the drawing and transmits the light vibrating at right angles to it. A spacer element, C, introduces a path difference. Surfaces S and T are silvered or aluminized. Light which enters in the direction OS, emerges in the reverse direction, SO, in two components polarized at right angles, with a phase difference given by

$$2\pi n = 4\pi(\mu'/\lambda)dc\cos\varphi \qquad \text{(VIII.1)}$$

where μ' is the refractive index and φ is the angle of incidence on S and T. The prism P (constructed like A, B) has the double function of polarizing entering light and separating out the desired part of the emerging light. It is shown in an incorrect orientation for simplicity in drawing. Actually prism P is rotated about the OS direction, to bring its axis to an angle of 45° to that of prism AB. The transmission of the whole assembly for light emerging in the R direction is then

$$\tau=\sin^2\pi n. \qquad \text{(VIII.2)}$$

The remainder of the light emerges along SO.

The most serious difficulty in the construction of such an interferometer is the optical working and cementing of the b-layer to the required accuracy. The orientation of the S and T surfaces with respect to each other is not so critical, since a slight misalignment can be compensated by a thin wedge of birefringent material between prism P and the interferometer.

One method of using polarizing interferometers combined with birefringent elements in a filter is shown schematically at b, Fig. 7. Between each polarizer, P, and the following interferometer, I, is a b-element, which constitutes the m (for entering light) and q (for emerging light) components of a split element. The interferometer then takes the place of the p component. Between successive polarizers are purely birefringent split element units. The assembly includes 4 interferometers, 4 polarizing prisms, and 10 b-elements. The interferometers and b-elements should be equipped with phase shifters (not shown). As an example, the interferometers might have retardations of 245,760; 122,880; 61,440; 30,720; and the b-elements, retardations from 15360.5 to 30.5 at $\lambda=5000$ angstroms. The system would transmit bands of about 0.01 angstrom effective width, spaced about 150 angstroms apart. Adjustment of the phase shifters will cause a selected band to scan the spectrum.

If the light transmitted by the filter is received on a photoelectric cell, its output gives a high resolution spectrophotometric curve of the entering light. Such a filter would be preferable to a grating spectrograph for spectrophotometric purposes, because, in spite of its small field (maximum usable φ about 0.0012 radian), it can be designed to transmit something like 1000 times as much light—a matter of considerable importance when such sharp bands are used, even in solar studies.

The Birefringent Filter: A Correction

[J. Opt. Soc. Am. 39, 229 (1949)]

JOHN W. EVANS
Harvard Observatory, Harvard University, Cambridge Massachusetts

SHORTLY after this paper went to press, the author was greatly embarrassed to discover an error in his remarks concerning the field of the split element filter. The field, is, in general, approximately the same as that of the equivalent simple filter. Within this limitation, however, the split element filter performs satisfactorily, and for some purposes the reduction in the number of polarizers required is important.

If, using the notation of the original paper, we let

$$\begin{aligned} n_m - n_q &= \delta \\ n_m + n_q &= n_j, \end{aligned} \tag{1}$$

Eqs. (V.8) reduce to

$$\begin{aligned} \tau_r &= 1 - \cos^2\pi n_p \sin^2\pi\delta - \sin^2\pi n_j \sin^2\pi n_p \\ \tau_s &= \cos^2\pi n_p \sin^2\pi\delta + \sin^2\pi n_j \sin^2\pi n_p. \end{aligned} \tag{2}$$

The filter is constructed with n_m and n_q equal for light parallel to the instrumental axis ($\phi=0$). For light inclined to this axis, however, they are no longer equal, and, from Eq. (IV.2),

$$\delta = \frac{n_j}{2} k\varphi^2 \left(\frac{1}{\gamma}+\frac{1}{\alpha}\right)\cos 2\theta. \tag{3}$$

There is, therefore, a system of hyperbolic fringes (composed of light in the wave-length intervals between the transmission maxima of the p component) superposed on a field of broad circular fringes indicated by Eq. (IV.1). If n_p is approximately $\frac{1}{2}$ over the spectral region considered, we have essentially Lyot's first type filter, and the fringes are very weak. The field of the split element filter can be increased by constructing the two components of the split elements individually in any one of Lyot's three wide-field forms. Since the fringes are then circular and identical in each of the two components, $\delta=0$ for all values of ϕ. The off axis effects are then given by the appropriate equations of section IV of the original paper.

Since no gain in field results from crossing the γ-axes of the m and q components, it is pertinent to ask whether this construction is the most advantageous. The transmission of a split element unit with the γ-axes of the m and q components parallel readily follows from Eq. (V.6) of the original paper if we interchange α and γ as multipliers of d_q/λ. We find

$$\begin{aligned} \tau_r &= \cos^2\pi(n_m - n_q) - \sin 2\pi n_m \sin 2\pi n_q \cos^2\pi n_p \\ \tau_s &= \sin^2\pi(n_m - n_q) + \sin 2\pi n_m \sin 2\pi n_q \cos^2\pi n_p. \end{aligned} \tag{4}$$

Expressed in terms of δ and n_j, this becomes:

$$\begin{aligned} \tau_r &= 1 - \sin^2\pi n_p \sin^2\pi\delta - \sin^2\pi n_j \cos^2\pi n_p \\ \tau_s &= \sin^2\pi n_p \sin^2\pi\delta + \sin^2\pi n_j \cos^2\pi n_p. \end{aligned} \tag{5}$$

If δ is zero for light parallel to the instrumental axis, it remains zero at all values of ϕ. The field is the same as that of the equivalent simple filter. The expression for τ_s is then

$$\tau_s = \sin^2\pi n_j \cos^2\pi n_p.$$

This arrangement does offer some advantage over the split element unit with the γ-axes of the m and q components crossed. A filter of l birefringent elements can be constructed with its first $l/2$ elements (i.e., the thin elements) sandwiched between the halves of the $l/2$ thick elements. Since the transmission of the middle element in each unit is $\cos^2\pi n$, the n-values are integral at the desired wave-length, and can be made proportional to the powers of 2. The $l/2$ thin elements therefore function equally well throughout the spectrum. Since the burden of suppressing the light at wave-lengths far removed from those of the transmission bands falls mainly on these thinner elements, this property is a real advantage.

A NEW TYPE OF BIREFRINGENT FILTER

by

Ivan Šolc

Czechoslovak Journal of Physics,

4(1): 53-66, 1954.

A new type of birefringent filter is described in which relatively small birefringent plates, all of equal thickness, and two polaroids are used. A new method of analyzing two complementary models is demonstrated, and an example is given of the actual preparation of the filter.

INTRODUCTION

In 1914, R. W. Wood [1] employed birefringence to isolate the first or second line of the doublet of Na, using a quartz plate cut in the direction of the optic axis and inserted between two Nicol prisms. Since then this experiment has been repeated many times, but only in 1933 did B. Lyot [2] find a method for applying birefringence to obtain monochromatic light. Independently of him, in 1938, Y. Öhman [3] began similar experiments. He employed several quartz plates with interlayers of polaroids and obtained a monochromatic filter. Following this, Lyot [4] again began work in this field and constructed a birefringent filter to be used for observations of solar prominences. Lyot also designed and engineered more complex filters. Recently, J. W. Evans [5], who also suggested constructing an interferometer on the birefringence principle, has made a detailed study of birefringent filters. At a number of astronomical observatories solar investigations are being carried out with birefringent filters; the advantage of a light monochrometer without slits and with a large field of vision affords ever increasing possibilities for application of birefringent filters in science, technology, and industry.

THE PRESENT STATUS OF BIREFRINGENT FILTERS

The simplest kind of birefringent filter has been described independently by Lyot and Öhman. This filter is a system of quartz plates prepared from optically perfect crystal so that the optic axis z lies in the plane of the plate and the thickness of each successive plate is twice that of the preceding. Polaroids are inserted between the individual quartz plates. The shortcomings of this filter (if a narrow transmission band is required) are the negligible permissible convergence of light and the excessive thickness of the last plate. Lyot obtained a large permissible convergence, using new kinds of birefringent filters, which are now referred to as Lyot's first, second, and third type filters. Lyot's first type filter is similar to the above-mentioned basic resolution of the birefringent filter; however, each quartz element is divided in two, and a half-wave plate is inserted between the two elements. In Lyot's second type filter connected birefringent elements are also employed, each of which, however, is prepared from a different material and both materials have to satisfy a definite condition. The most ideal filter is Lyot's third type, in which every birefringent element of the basic filter is replaced by three plates, two of which are prepared from the same material, and the third from a different material.

Evans' work [6] represents a new attainment in the construction of birefringent filters. He developed a filter containing only half the number of polaroids required in previous filters. This filter, which Evans calls the "split element filter," is a modification of Lyot's first type. At

present the Evans filter is considered to have the most ideal construction. Further investigations into the construction of birefringent filters have been directed toward the perfection of a tuning filter. Already, during the initial experiments, first results have been obtained by Lyot and Öhman, who have established that it is possible to shift the position of the transmission-band maxima within known limits by appropriate inclination. Evans, who has indicated the possibility of employing wedge-shaped plates, is concerned with the general solution of the problem of filter tuning. Lyot discovered the possibility of fine filter tuning by temperature change. Finally, B. H. Billings [7] found a new method for tuning birefringent filters using electro-optical materials (e.g., ADP).

Whole series of birefringent plates of different thickness are used in all filters known to date. In developing the birefringent filter with the narrow transmission band, i.e. with great resolving power, the construction of the thickest elements gives rise to considerable difficulties: either very thick plates of rare quartz are required or it is necessary to find very complex combinations. What is more, the thickness of the plates must be kept correct to the order of 10^{-4} mm. Since each filter has an assembly of plates of different thickness, the manufacture of these filters is very difficult and expensive.

In every filter there is a whole series of polaroids, which leads to large absorption (especially when imperfectly colored polaroids are used) and thus there is a large loss of transmission. In this respect, the Evans filter is a step forward, since only half the number of polaroids [normally required] suffice, other conditions being equal.

A NEW TYPE OF BIREFRINGENT FILTER

A new type of birefringent filter has been constructed in the laboratories of the Scientific-Research Institute of Electrotechnical Physics in Prague.[+)] The advantage of this filter over the filters known to date is that comparatively thin birefringent plates of the same thickness are used and are ground from identical material. Another great advantage of this new filter is that two polaroids suffice. Furthermore, the resolution of the filter has two modifications, the second of which gives results complementary to the first.

The general features of the filter are analyzed and described below, and an example is given of its actual construction from quartz.

a) The first modification of the filter

In this filter, m birefringent plates of identical thickness d and 2 polaroids are used. In each plate there are two principal directions in which the linearly polarized light wave is propagated, where the index of refraction in one direction is n_1, and in the other n_2; the directions are perpendicular to each other. The difference of the two indices of refraction determines the birefringence value of the material used.

Using familiar methods we find both the main directions of the prepared plates within an accuracy of $\pm 10'$, and mark them.

[+)] Nauchno-issledovatel'skiĭ Institut Elektrotekhnicheskoĭ Fiziki v Prage.

The examined light passes through the first polaroid. The linearly polarized light wave emerging from the polaroid is defined by the equation

$$r = \sin 2\pi f \cdot t \ , \qquad (1)$$

where r is the instantaneous deflection, f the frequency of the oscillations, and t time. The amplitude of the entering light is assumed to be unity, which does not affect the generality of the result. This linearly polarized wave falls on the first birefringent plate. One of the principal directions of the plate forms a small angle ε, to the right, with the direction of the vibrations of the polarized light. The light, leaving the first plate, falls on the second, which is turned with the same principal direction at the same angle ε to the left of the plane of direction of the vibrations of the polaroid. The third and all the remaining odd-numbered plates are in the same position as the first plate, and all the even-numbered plates are in the same position as the second plate. We will use the conventional symbol for wavelength, λ. The following definitions are introduced for brevity:

$$\sin \varepsilon = A, \ \cos \varepsilon = B, \ \sin 2\varepsilon = C, \ \cos 2\varepsilon = D, \ \sin 4\varepsilon = U,$$

$$\sin 2\pi \cdot \left[f \cdot t - \frac{d}{\lambda} \cdot (kn_1 + jn_2) \right] = S(kn_1 + jn_2),$$

where k and j are whole numbers.

After passing through a certain number of plates, the light always assumes the two principal directions of the last plate. Let us call the instantaneous deflections in these two directions x_g and y_g, when the light

has passed through g-1 plates. After the light has passed through the first plate, the following expressions hold:

$$\begin{aligned} x_1 &= A \cdot S(n_1), \\ y_1 &= B \cdot S(n_2). \end{aligned} \tag{2}$$

The following recurrent formulas may be derived for the passage of light through the succeeding plates: With even g + 1:

$$\begin{aligned} x_{g+1} &= \dot{x}_g \cdot D - \dot{y}_g \cdot C, \\ y_{g+1} &= \bar{x}_g \cdot C + \bar{y}_g \cdot D. \end{aligned} \tag{3a}$$

With odd g + 1:

$$\begin{aligned} x_{g+1} &= \dot{x}_g \cdot D + \dot{y}_g \cdot C, \\ y_{g+1} &= \bar{x}_g \cdot C + \bar{y}_g \cdot D. \end{aligned} \tag{3b}$$

In these equations the dot above a letter means that S is followed by (n_1), a dash means that S is followed by (n_2).

The calculation for the entire series of plates is carried out very simply on the basis of the above formulas; the results obtained, however, are so extensive that they are not included here.

A further step is the reduction of the rays leaving the last plate in two mutually perpendicular directions x_g and y_g to the general direction of the vibrations τ. This is accomplished by means of the second polaroid, whose direction of vibrations is exactly perpendicular to that of the first polaroid. Thus, the polaroids cross. Mathematically this may be expressed by the equation:

$$T = x_g \cdot B - y_g \cdot A.$$

Examining the expression for T, we derive the basic properties of the filter. I will give the results of the long calculation only: the filter shows sharp maxima, whose position in the spectrum is defined by equation

$$d \cdot (n_1 - n_2) = \frac{2v - 1}{2} \lambda \, . \qquad (4)$$

In this equation v is a whole number, $v = 1, 2, 3, \ldots$.

Besides the principal maxima there is a considerable number of secondary maxima, whose intensity, however, may be disregarded. The centers of the minima are defined by the equation:

$$d \cdot (n_1 - n_2) = v \cdot \lambda \, . \qquad (5)$$

Thus, the positions of the maxima and minima of the filter are completely identical with the positions of the maxima and minima of one of its plates placed between the crossed polaroids.

From this we derive the first significant result: The spacing of the maxima and minima in the spectrum is determined only by the birefringence and the thickness of one filter plate. The sharpness of the maxima is determined by the width of the transmission bands: for the principal maxima, the bandwidth $\Delta\lambda$ is determined by the formula:

$$\Delta\lambda \doteq 0.6 \frac{\lambda^2}{g \cdot d \cdot (n_1 - n_2)} \, , \qquad (6)$$

where g is the total number of plates. Formula (6), which is only approximate, indicates that the width of the band is approximately the same as the width of the maxima of one of the plates, the thickness of which equals the

sum of all the filter plates, in a diagonal position. This shows one of the substantial advantages of the new filter. If we place the second polaroid in a position parallel to the first, a complementary phenomenon occurs, in that the maxima and minima interchange in position and width. Thus, a series of sharp dark bands is observed on the spectrum; their position is determined by equation (4) and their width by (6).

Angle ε is relatively small and varies approximately within the limits 0.5° and 10°. It is difficult to compute the optimum value of this angle mathematically. However, it is relatively easy to set it up by experimental means.

If we disregard the absorption in the polaroids and the reflection at the individual polished surfaces (the plates may be cemented with Canada balsam), the transmission of the filter amounts to 50 pct, as indicated by quantitative photometric measurements and an approximate theoretical evaluation. In the case of the cemented filter, we need only consider the total absorption and the reflection from one plane. Therefore, with a given margin /of error/ let us assume the transmission of the filter for the position of the maxima to be 30 pct. Thus far, not one of the selective filters has attained this transmission capacity. A schematic sketch of the filter is given in Fig. 1. In the experimental part we give some results of measurements on a filter made of quartz plates.

b) Second modification of the filter

In the second modification of the filter, m birefringent plates of equal thickness d and two polaroids were used, as in the first modification. The difference between the two types lies in the plate arrangement only. In the second modification, the plates are arranged in such a way that each successive plate is turned at the same angle ε and in the same direction as the preceding plate. Thus, a fan-shaped layout of the plates is obtained. The axis of symmetry of the fan of principal directions runs diagonally to the input /first/ polaroid. The second polaroid is parallel to the first.

Let us once again give a general description of the process taking place in the filter. We will introduce the following notations:

$$\sin\varepsilon = A,\ \cos\varepsilon = B,\ \sin 2\varepsilon = C,\ \cos 2\varepsilon = D,\ \sin\psi = L,\ \cos\psi = M,$$

$$\sin 2\pi\left[f \cdot t - \frac{d}{\lambda} \cdot (kn_1 + jn_2)\right] = S(kn_1 + jn_2).$$

ψ is defined by the expression:

$$\psi = \frac{\pi}{4} - \frac{m-1}{2}\varepsilon .$$

The polarized light from the first polaroid passes through the first plate. After this light has passed through the first plate, the vibration is defined by the expressions:

$$x_1 = L \cdot S(n_1),\quad y_1 = M \cdot S(n_2). \tag{7}$$

The following recurrent formulas are valid for the passage through the remaining plates:

$$x_{g+1} = \dot{x}_g \cdot B + \dot{y}_g \cdot A ,$$
$$y_{g+1} = \bar{y}_g \cdot B - \bar{x}_g \cdot A . \qquad (8)$$

The dots and dashes over the x and y have the same meaning as before.

After passing through a certain number of birefringent quartz plates, the light passes through the second polaroid, after which both waves again reduce to the general direction of vibrations. Mathematically, this process is expressed by the equation:

$$T = x_g \cdot L + y_g \cdot M. \qquad (9)$$

The meaning of the letters is the same as before. Again we give only the results of the analysis of the second filter modification. The filter has sharp maxima, whose position in the spectrum is defined by the equation

$$d \cdot (n_1 - n_2) = v \cdot \lambda . \qquad (10)$$

The centers of the broad minima are defined by a similar equation:

$$d \cdot (n_1 - n_2) = \frac{2v - 1}{2} \lambda . \qquad (11)$$

The width of the transmission band is expressed by equation (6). The secondary minima are quite insignificant. The transmission of this filter is the same as that of the first filter modification. If we place the second polaroid in the crossed position, sharp minima are formed, whose position is determined by (10), and whose broad maxima are determined by (11).

Thus a complementary phenomenon again occurs. In general form, the second modification of the filter is complementary in construction to the

first modification. The filter assembly is represented schematically in Fig. 2.

In the experimental part of this work, several results are given which were obtained from a filter of quartz plates.

OTHER POSSIBILITIES OF EMPLOYING AN ASSEMBLY OF IDENTICAL BIREFRINGENT PLATES

It is possible to carry out a whole series of interesting experiments with an assembly of birefringent plates of identical thickness, ground at the same angle, from the same birefringent material. The filter construction is always identical in that two polaroids are used, between which the plates are placed (Fig. 4b). The various modifications of an assembly of this type differ only in the changes of the position of the directions of vibrations, and in the position of the principal directions of the birefringent plates, i.e. the polaroids and the plates are gradually rotated. In these experiments a very diverse distribution of the transmission and absorption bands in the spectrum is possible. Thus, for example, the whole spectrum consists of sharp doublets, triplets, and quadruplets, light or dark. Instead of a detailed account, in the experimental part of this paper we will give examples of several possibilities, which may be realized by keeping the birefringent plates stationary and simply rotating the polaroids.

EXPERIMENTAL RESULTS

Of the experiments conducted with different birefringent plates, I will cite only the results obtained with one assembly of ten plates. Quartz plates, cut in the direction of the optic axis, were used. The birefringence value of the quartz in this direction, relative to the wavelength of the light is shown in Fig. 3.

In this direction, the quartz also shows a slight capacity for turning the plane of polarization. This capacity of the quartz is zero at a cut of $56^{\circ}10'$. At the specified cut of 90°, a slight capacity for turning remains, which causes an ellipticity of the light with an axis correspondence of about 0.02, so that it may be assumed that the light is propagated practically linearly along the principal directions of the vibrations. The measurements were conducted with plates 1.4417 mm thick. Two slightly colored herapathite polaroids were used. A spectroscope with spectrometer was used for the visual experiment, a spectograph for the photographic experiment. Several of the spectrograms are given here. The spectrograms show that in the red part of the spectrum they are limited by the sensitivity of the photofilms, while in the violet they are limited by the glass of the polaroid.

a) Measurement of the first filter modification

The arc spectrum of iron, which may be used as a scale for wavelength, is given at the top of Spectrogram 1. Below it follow: a) the spectrum obtained from the continuous spectrum after the light has passed through the 10-plate filter, b) the same assembly with only 8 plates, c) with 6 plates,

d) with 4 plates, e) with 3 plates, f) with 2 plates, and g) with 1 plate. For comparison, spectrum h was also photographed using the same plate placed in a diagonal position. In all the other cases, it was assumed that angle $\varepsilon = 5^\circ$. Since a reduction in the number of plates also resulted in a reduction in the number of reflections on the individual surfaces, the exposure time was reduced correspondingly. Obviously, the exposure time was too short in the case of spectrograms e, f, and g. Spectrogram 1 upholds the validity of equations (4) and (5). These equations were also accurately checked by spectrometric measurements. Furthermore, the validity of equation (6) was established through photometric analysis of the spectrogram. The coefficient 0.6 was also determined through this analysis.

The properties of the filter with respect to angle ε were also defined on the basis of Spectrogram 2. The arc spectrum of iron again appears at the top. A gradual change of angle ε took place, namely $1^\circ40'$, $3^\circ30'$, 5°, 7°, 10°, $10^\circ30'$, in the spectra that follow, i.e. a, b, c, d, e, and f, with the same exposure. In spectra d and e not all the ε values were identical, and in some plates the deviations amounted to approximately 20'. The effect of this deviation on the symmetry of the separate passages of light is noticeable on the sketch. On the given spectra, one may see the optimal angle ε, at which additional passages /of light/ can be ignored and where the principal maximum is narrowest. The last two spectra, g and h, were prepared for a final photometric check of equation (6). I will give, briefly, the checking sequence: one of the polaroids was removed and placed in front of the first polaroid. Its direction of vibrations was placed parallel to the direction of the vibrations of the first polaroid. With this

layout, the light first passed through the series of birefringent plates, then through both polaroids, placed parallel to each other. The total result of the transmission was as follows: all reflections in the system remained unchanged, 50 pct of the transmission was lost owing to the absorption in the polaroids; spectrum h was photographed in the same way as spectrum g; however, the polaroids were placed in front of the birefringent plate assembly. Thus there is an identical loss in both cases g and h, owing to absorption and reflection, as in the case of the 10-plate filter. The photometric measurements have shown that the light intensity of the filter maxima is exactly the same as that in the g and h spectra at corresponding wavelengths. Thus, it was established that the filter transmission was exactly 50 pct, if we ignore absorption and reflections. This 50 pct is due to the natural polarization mechanism and therefore cannot be increased. Furthermore, a partial polarization, caused by the plate assembly, appears in spectrum k.

b) Measuring the second filter modification

These measurements were a development of the first filter measurements. The spectrum of iron is again shown at the top of the spectrogram (No. 3). In this spectrum some wavelengths are approximate. Spectrum a was obtained with 6 birefringent plates and two polaroids with the same arrangement as in Spectrogram 2, h. This spectrum was also photographed to evaluate light intensity. Spectrum b is the spectrum of the 6-plate filter of the first modification. Spectrum c represents the spectrum of the 6-plate filter of the second modification. Both spectra taken together show that the filters

complement each other. Spectrum d is also a spectrum of a 6-plate filter of the first modification, but one in which the polaroids were placed parallel to each other. Spectrum e represents the spectrum of the 6-plate filter of the second modification, but in which the polaroids were crossed. Again the complement is evident. Spectrum f represents 3 plates whose directions of vibrations were set uniformly in a diagonal position with respect to the crossed polaroids. Spectrum g is similar, but with parallel polaroids. Both spectra f and g were photographed in order to evaluate the width of the transmission bands. In all previously known complex filters the bandwidth is approximately the same as the bandwidth transmitted by one plate in a diagonal position, the plate being half the total thickness of all the plates. From the above spectra, it follows that the bands of the filter are considerably narrower, if we consider the width of the transmission band of the 6-plate filter and the width of the transmission band with 3 such plates in a diagonal position. The measurements again verify the empirical coefficient 0.6 of equation (6).

c) Examples of several other experiments with measurable plates

The plates were placed as in Fig. 4. In all the spectra of Spectrogram 4 this arrangement was kept constant, only the position of the direction of vibrations of the polaroids was changed. The spectrum of iron is given at the top. Spectrum a represents a 9-plate filter with directions of vibrations of the plates as in Fig. 4.

In the first polaroid as well as in the second, the direction of vibrations is parallel to that of plate No. 5. Spectrum b was obtained with

the same plate assembly; the direction of vibrations of the first polaroid was parallel to plate 5, the second polaroid was crossed. Spectrum c was obtained from the same plate assembly with the directions of vibrations of both polaroids parallel to plates 2 and 9. Spectrum d was photographed with the same assembly; the first polaroid was placed parallel to plates 1 and 9; the second was crossed. Spectrum e was obtained with the direction of vibrations of both polaroids along the axis of symmetry of plates 1, 9 and 5. Spectrum f was photographed with the same assembly, however, the second polaroid was crossed. During the photographing of spectrum g the direction of vibrations of the first polaroid was parallel to plates 1 and 9, the second, to plate 5. Finally, in the case of spectrum h the position of the vibrations of the first polaroid was parallel to plates 1 and 9 and the direction of vibrations of the second polaroid was perpendicular to plate 5.

From these examples it follows that there is the possibility of an enormous number of combinations of positions of the principal directions of the birefringent plates and of the directions of the vibrations of the polaroids. Every new combination gives ever newer possibilities of the distribution of transmission bands in the spectrum. It is also probable that this use of an assembly of identical birefringent plates will find its practical application.

d) Some additional information

In making measurements, I also verified the possibility of shifting the transmission bands of the spectrum by inclining the filter. The results obtained verified the measurement results obtained by Öhman with the simplest filter. If one wishes to include all wavelengths, the first modification of the filter should be changed to the second when tuning the filter to include wider limits.

I also tried quartz plates ground at an angle of 56°10' to the optic axis. A filter can be constructed from such plates, but it is much more sensitive to the convergence of light. This also follows from the equation which expresses the change in birefringence with respect to the change in the angle of cut:

$$(n_1 - n_2)_\varphi = (n_1 - n_2)_{90^\circ} \cdot \sin^2 \varphi .$$

In finding the derivative of this approximate equation, we obtain the relation of the change of birefringence to the change of the angle

$$\Delta (n_1 - n_2)_\varphi = (n_1 - n_2)_\varphi \cdot \sin 2\varphi \cdot \Delta \varphi .$$

From this equation it clearly follows that the smallest change takes place at angle 90°, that is, with the cut parallel to the optic axis. This equation has its maximum at $\varphi = 45^\circ$. Thus, in this region the filters are most sensitive to the convergence of light. When the bundle of light is parallel there are no objections to the use of plates ground at an angle of 56°10', which has zero capacity for turning the plane of polarization. I did not measure the relation of the transmission maximum to the temperature. In the literature cited, a shift of 0.66 A in the direction of shorter waves for a 1°C rise in temperature is given for quartz.

In constructing a filter with great resolving power, the direction of the cut must be maintained very strictly. In cutting quartz parallel to the optic axis it is expedient to keep the tolerance of the angle at ±10'. X-ray verification of the cut, using a Zeeman spectrograph, is well suited

to this purpose. The thickness of the plates must also be strictly controlled. The thickness and planeness can be closely regulated interferometrically.

Prof. Bečvář called my attention to the growing importance of birefringent filters and thus led me to think about new possibilities. In concluding, I wish to express my gratitude to him. The present work is one of a number of dissertations under way at the Scientific-Research Institute of Electrotechnical Physics in Prague.

submitted

20 April 1953

Nauchno-issledovatel'nyi Institut
Elektrotekhnicheskoi Fiziki v Prage.

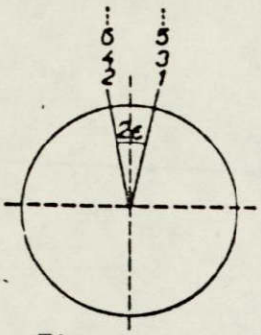


Fig. 1

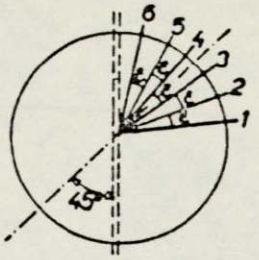


Fig. 2

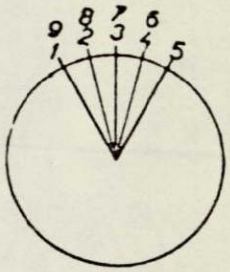


Fig. 4a

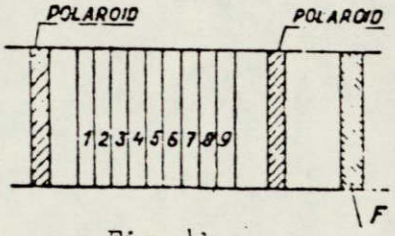


Fig. 4b

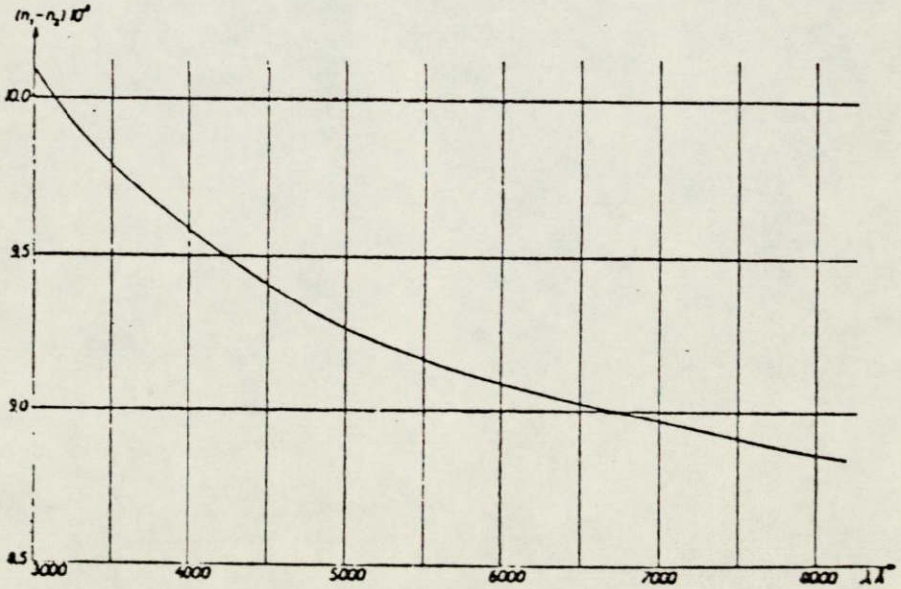
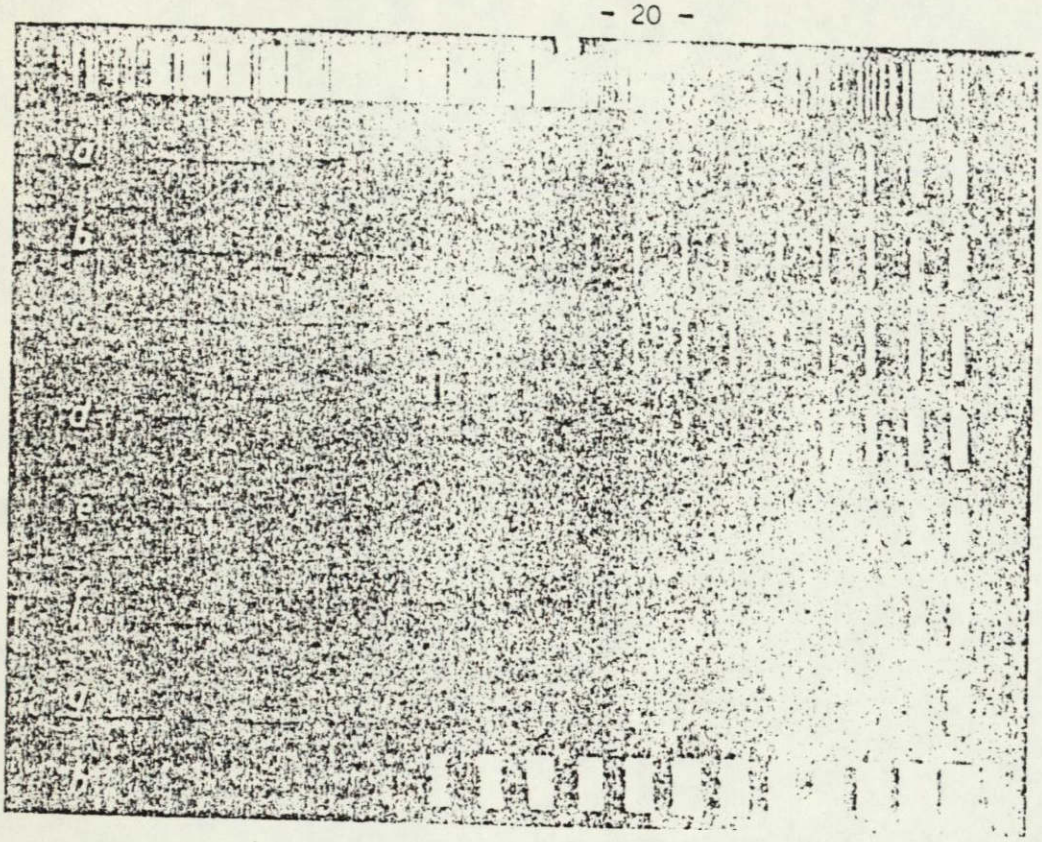
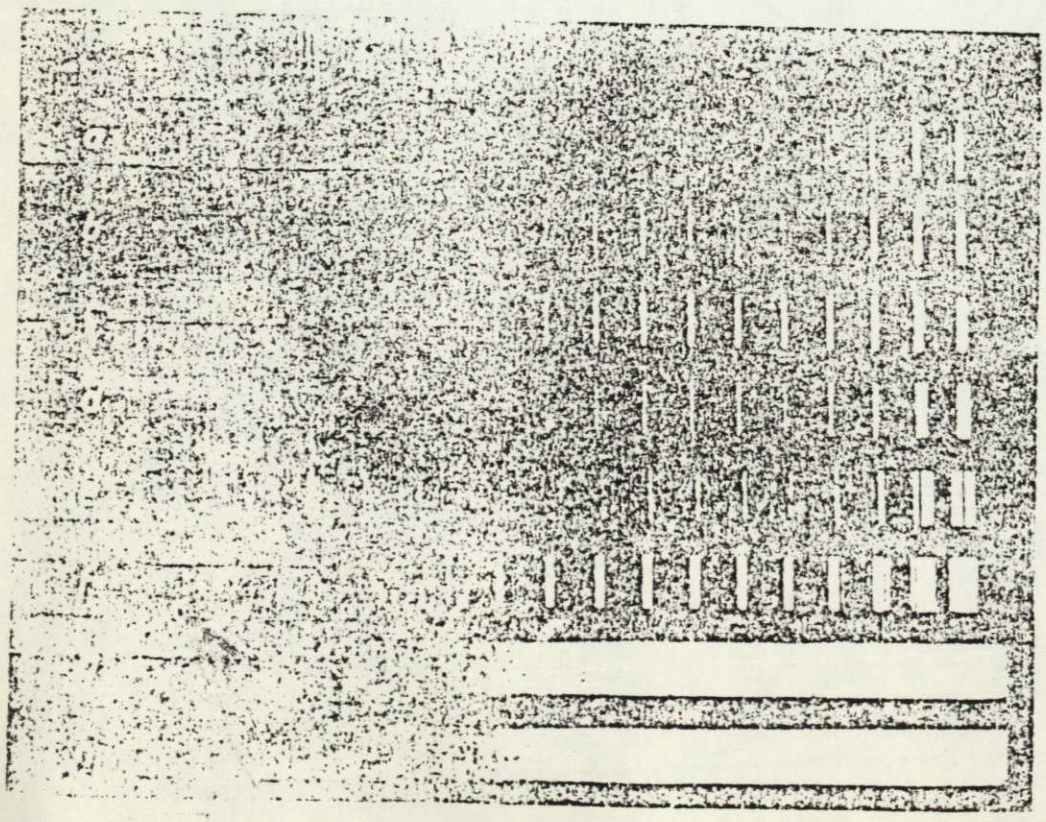


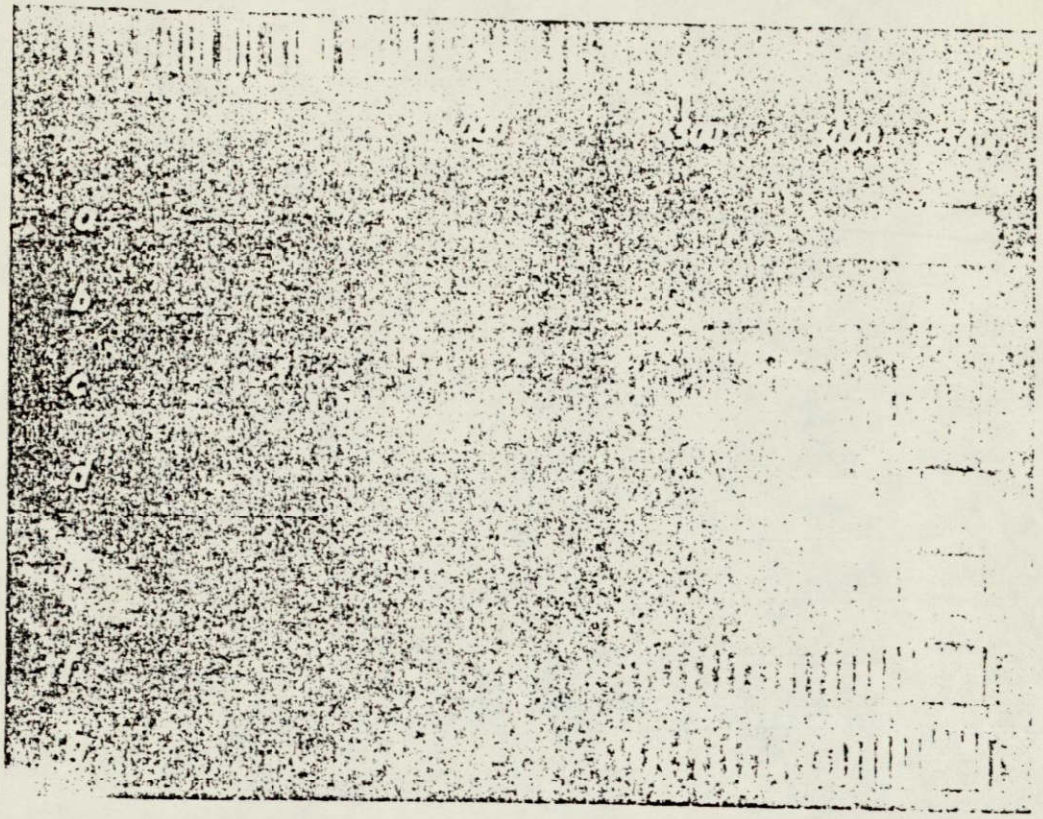
Fig. 3



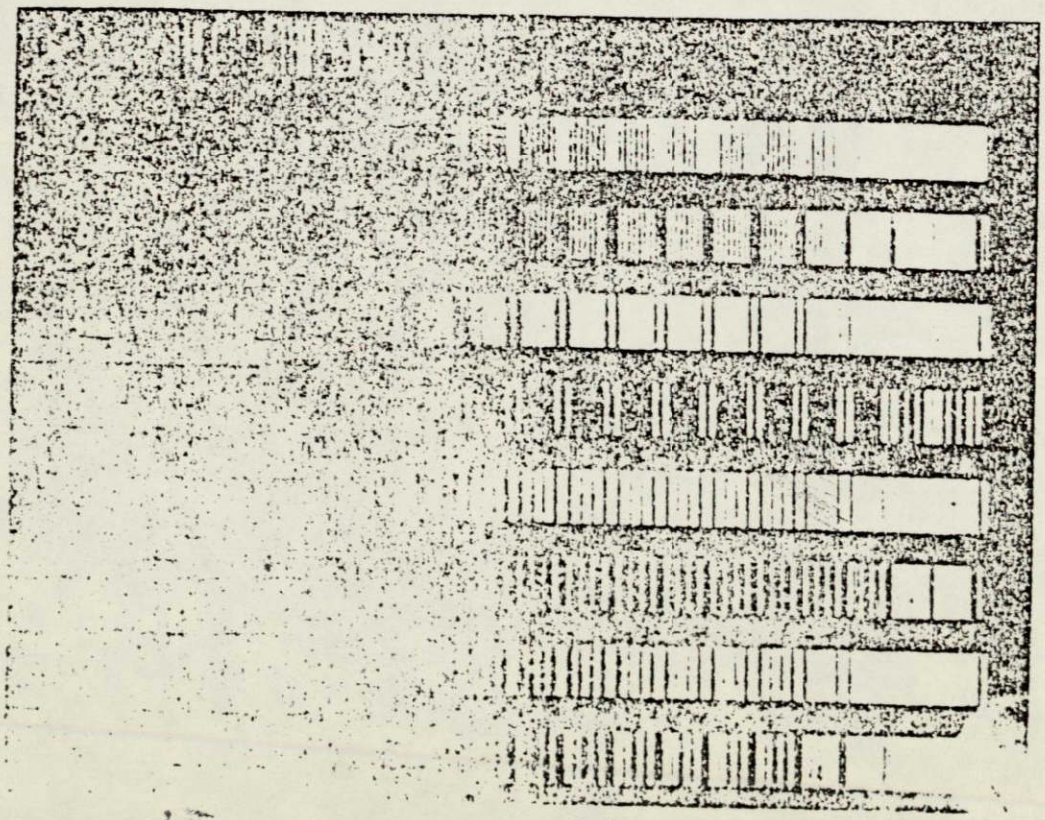
Spectrum 1



Spectrum 2



Spectrum 3



Spectrum 4

LITERATURE

1. Wood, R. W. Phys. Zeitschrift, 15: 313, 1914.

2. Lyot, B. Comptes Rendus, 197: 1953, 1933.

3. Öhman, Y. Nature, 141: 3560, 157, 1938.

4. Lyot, B. Ann. Astrophys., 7(1): 2, 31, 1944.

5. Evans, J. W. J. Opt. Soc. Am., 39: 3, 229, 1949.

6. Evans, J. Ciencia e Investigación (Argentina), 3(9): 365, 1947.

7. Billings, B. H. J. Opt. Soc. Am., 37: 738, 1947.

8. Rozenberg, G. V. Uspekhi fizicheskikh nauk, 17: 1-2, 1952.

FURTHER INVESTIGATION OF THE BIREFRINGENT FILTER

Ivan Šolc

Czechoslovak Journal of Physics,

5(1): 80-86, 1955.

This continues an article which appeared in the Czechoslovak Journal of Physics, 4: 53-66, 1954, where the relationship between the angle and the number of plates was found. A qualitative study is made of the influence of the inclination of the filter and an approximate expression is derived for the maximum permissible convergence of light and for the relation of the wavelength of the transmitted light to the temperature.

In my article "A new type of birefringent filter" which appeared in this journal, the question of the size of the angle between the directions of the vibrations of the individual plates had been left unsolved when this filter was constructed. In this work I wish to show the mechanism by which this angle ε is determined.

Analytical investigations of the angle ε did not lead to clear results. Therefore, I used the experimental method to study the influence of this angle on the birefringent filter. I used the visual spectrometer method, since the weak secondary maxima can be determined more reliably visually than by photographic plate and photoelectric cell. The aim of the measurements was to find the relation between the optimal angle ε and the number of plates, their thickness, and birefringence. Furthermore, I was interested in the influence of the departures of some plates from the optimal angle on the quality of the filter, and the influence of the irregular orientation of the plates, their unequal thicknesses and optic heterogeneity. Concurrently, we carried out investigations of the influence of the inclination of the whole filter with respect to the direction of the rays. The established relations are very interesting and have a practical significance. The

measurements were conducted in air and in immersion. In all experiments quartz was used as the birefringent material.

Below is a short report on the experiments, which often had unexpected results. First of all it was established that the value of the optimal angle $\mathcal{E}$ does not depend on the thickness of the plates, and thus it is also independent of the birefringence value of the material employed. Next it was established that the optimal angle is constant for all wavelengths. The optimal angle depends only on the number of plates, hence its value may be expressed very approximately by the following formula:

$$\mathcal{E}_{opt} = \frac{45^\circ}{g}, \text{ (g being the total number of plates)} \qquad (1)$$

This formula holds for the first and second filter modifications. Since the limits of measurement of angle $\mathcal{E}$ must be defined accurately, I again give sketches of the plan of the first and second filter modifications with final notations (Figs. 1a and 1b). In these sketches the directions of vibrations of the polarizers are depicted by broken lines, the principal optic axes of the plates are represented by solid lines.

In the following I will give some of the details observed during experiments with the first filter modification. The second modification in a number of respects has the same properties.

THE BEHAVIOR OF A FILTER WITH FINE TUNING

Let us examine more closely the effect of the changes of angle $\mathcal{E}$ on the behavior of a filter and let us define more exactly what constitutes the concept of the optimal angle. The interference picture changes with a

change of angle. Thus, the width and intensity of a transmitted line can be changed within given limits and can exert a strong influence on the intensity of the secondary maxima. We tried in so far as possible to suppress the secondary maxima, at the same time keeping the transmission band as narrow and as light as possible. This requirement was satisfied at optimal angle ε_{opt}. By experiment I determined that in general three zones of angle ε may be distinguished, where the following characteristic phenomena take place:

a) Zone 1: ε smaller than ε_{opt}. When the angle becomes smaller, the band of the transmitted line gradually broadens, whereby the characteristic /curve/ of the transmission intensity with respect to the wavelength becomes bell-shaped, while the characteristic of the optimal angle is nearly straight. As the angle decreases, the intensity of the transmission also decreases, at first slowly, then rapidly. However, the intensity of the secondary undesired maxima decreases simultaneously, and considerably more rapidly. The transmission band becomes approximately 60 pct wider than the initial width. As angle ε approaches 0°, the whole process becomes rapidly weaker. When $\varepsilon = 0^\circ$, naturally the whole field of vision is dark.

b) Zone 2: In this zone the angle $\varepsilon = \varepsilon_{opt}$, according to formula (1). In the first filter modification, the optimal angle may be changed within fairly wide limits, without causing any considerable change in the character of the spectrum. It was established by experiment that in this case a change of 10-20 pct in angle ε_0 is permissible. Thus it seems that in the first filter modification it would be expedient to keep the ε values approximately

10 pct below those given by the above formulas. The computed value of the angle is better for the second filter modification.

c) Zone 3: Angle ε larger than ε_{opt}. The width of the transmission band decreases somewhat at first, by several percents, but when the angle increases, the transmission band begins to bifurcate. The intensity of the secondary maxima quickly increases, the phenomenon loses its typical character and the filter ceases to be monochromatic.

Such are the results of the experiments which established the directions of vibrations and in which absolutely identical and homogeneous plates were employed.

THE EFFECTS OF INACCURATE MOUNTING AND HETEROGENEITY

If the directions of vibrations are not established accurately enough, i.e. if the angles of the different plates differ, various changes occur in the initial position of the spectrum of the birefringent filter. If the departure of the mounting angle for the individual plates does not exceed 5 pct of the established ε value, only an insignificant irregularity is observed in the distribution of intensities of the secondary maxima, and the transmission curve of the principal maximum ceases to be symmetrical. If the departure of any plate exceeds the above value, there will be a considerable increase in the intensity of one of the secondary maxima or a general irregularity in the distribution of intensities. With greater departures (20 pct or greater), the secondary maxima cause interference. An incorrect arrangement of the polarizer will also have this effect.

The heterogeneity of the plates, their differences in thickness and twinning also cause undesirable phenomena, particularly in filters with great resolving power. However, the twinnings may be excluded in advance, and the thickness can easily be controlled interferometrically and by polishing within an accuracy of 0.0001 mm, which is sufficient in all cases.

Another question is the possibility of maintaining the direction of the cut. In my first article on filters I wrote of the possibility of using cuts that are not parallel to the optic axis of the crystal. I also introduced a small permissible convergence of the bundle of light. I should amend my statement. Oblique cuts may only be used in filters with a small number of plates. In other cases, oblique cuts should not be used. The reason is very simple: only in cuts parallel to the optic axis does the perpendicularly incident ray pass in a constant direction in the direction of the extraordinary ray, i.e., perpendicular to the plate. With other cuts, as follows from Huygens' construction, the extraordinary ray departs from the initial direction, as a result of which there is an ever larger divergence of the bundle of light in the successive plates, which finally makes monochromatization impossible with a larger number of plates. Thus in constructing birefringent filters it is expedient to use only cuts which are parallel to the optic axis of the crystals. These cuts may be adjusted with an accuracy of 10' (by the optic method or the X-ray method), which is sufficient for filter production.

Before assembling the filter, a simple control can be effected over the optic homogeneity and the uniformity of thickness of the plates. For this purpose two prepared polished plates are placed one on the other so that the

direction of the maximum coefficient of refraction of the first plate will cross that of the second plate. The birefringence of the plates arranged in this manner will cancel. If this pair of plates is examined in parallel polarized light in crossed polarizers, on rotation the pair of plates should appear uniformly dark along the whole surface. This system will be highly sensitive to any heterogeneity, irregularity in thickness, etc. This method, e.g., can also reveal electric twinnings in the quartz.

THE RELATION BETWEEN THE BEHAVIOR OF THE FILTER AND ITS INCLINATION TO THE LIGHT RAYS

Usually a filter is placed in parallel light, whose rays are perpendicularly incident on the filter. However, it is very useful to know the maximum permissible convergence of the bundle of light which still will not disrupt the interference to any appreciable extent. Then the filter can be tilted somewhat so that the rays will fall on the surface of the filter at a different angle; in this way the filter can be tuned within wide limits. Experiments along this line led to the following results:

The first and second filter modifications can be tilted such that the axis of inclination coincides with the bisectrix of the optic axes of the maximum or minimum coefficient of refraction. Thus there are two mutually perpendicular directions around which the filter can be inclined (the chosen direction remains parallel at all times during this process). The following uniform results were established for both filter modifications:

a) If the axis of inclination is the bisectrix of the optical axes of the maximum coefficient of refraction, the transmission bands will shift in the direction of shorter wavelengths, whether or not the filter is tilted in the same or in a different direction from that of the rays.

b) If the axis of inclination is the bisectrix of the optic axes of the minimum coefficient of refraction, the transmission bands will shift in the direction of longer wavelengths, whether or not the filter is tilted in the same or in a different direction to that of the rays.

c) The directions of the wavelength are exactly the same, whether the filter be tilted around the bisectrix in one direction from the direction of the rays or in the other.

d) The absolute value of the change in wavelength during the inclination around the bisectrix of the optic axes of the maximum refraction coefficient is the same as the absolute value of the change with inclination around the bisectrix of the minimum coefficient of refraction.

e) Changes of wavelength in both filter modifications are identical with respect to the angle of inclination.

f) The following general law was found which defines the change in wavelength with respect to the angle of inclination of the filter φ:

$$\overline{\Delta\lambda} = \text{const}\ \lambda \circ F(\varphi) \ . \qquad (2)$$

In the case of not-too-large values of the angle, we can assume $F(\varphi)$ to be approximately φ^2 or $\sin^2\varphi$. The essential relationship of $\overline{\Delta\lambda}$ to the

number of plates or to their thickness was not found.

After the indicated value for $F(\varphi)$ is set up and the numerical constant computed, formula (2) becomes

$$\overline{\Delta\lambda} = 4.5 \cdot \lambda \cdot \varphi^2 \cdot 10^{-3}. \qquad (3)$$

where $\overline{\Delta\lambda}$ is the shift of the known transmission band, λ the wavelength of the transmission band, φ the angle of inclination (in degrees) of the filter from the perpendicular incidence of the light ray.

The measurements were conducted on many filters for various wavelengths, in all cases with the angle ε optimum. By way of example I give the measurements made with the first filter modification consisting of seven quartz plates 1.5852 mm thick. This filter was constructed for observation of the solar corona. The required wavelength of the transmitted light was 5693 A. The relation of wavelength to angle of inclination of the filter was investigated. The measurement results are shown graphically in Fig. 2. The ascending line pertains to the filter inclination around the bisectrix of the principal axes of the minimum index of refraction. The descending line pertains to the inclination of the filter around the bisectrix of the maximum index of refraction. The measured filter was cemented with Canada balsam and was without frame; this made it possible to measure the filter up to an angle of inclination of 85°. The applicability of formula (3) to this case was also studied. The results are shown graphically in Fig. 3. The broken-line hyperbola indicates the computation according to formula (3), the measurements are shown by dots on the graph. The dots with crosses relate to the descending branch of graph 1, the dots in circles to the ascending branch.

g) The width of the transmission band is practically independent of the angle of inclination of the filter, which was established with inclinations up to 85°.

These results show the possibility of tuning filters within wide limits. Since a filter has a whole series of transmission bands, it is easy to build a smoothly tunable birefringent filter.

The inclination of the filter with respect to the direction of the light rays in the ordinary direction causes much more complex spectral changes. However, with a known degree of accuracy, this inclination may be resolved into the sum of the inclinations in the directions of the bisectrix of the optic axes of the maximum coefficient of refraction and the inclination along the bisectrix of the minimum coefficient of refraction, these being mutually perpendicular. As follows from the above, these directions exert opposite effects. Therefore, with greater inclination, the lines should bifurcate, with slight inclination the width of the transmission band should increase, since both parts of the bifurcated line somewhat overlap or come into contact with each other. Such phenomena are actually observed in their general features. This case of inclination of the filter in its general form is of great importance since it gives us data on the maximum permissible convergence of the rays (as long as the required monochromatization is not disrupted). Let us attempt to compute the permissible convergence on the basis of the above relations. If we assume a 10-pct broadening of the transmission band, these formulas make it easy to obtain the expression for the maximum angle of convergence, where $\Delta\lambda$ indicates the original width of the transmission band:

$$\varphi_{conv} = 10 \cdot \sqrt{\frac{\Delta\lambda}{\lambda}} \cdot \qquad (4)$$

The maximum convergence, which in general can be assumed, is that for which the bifurcating parts of the maximum are contiguous. The value of this angle is approximately equal to three times the value computed from (4).

CONCLUSIONS

In the preceding I have shown that a certain optimal angle ε exists, which depends only on the number of birefringent plates from which the filter is composed. This angle may be determined by formula (1); experience, however, has shown that with the first-modification filters it is expedient to employ somewhat smaller angles. The intensity characteristic can even be improved somewhat by small deviations in some of the plates. This adaptation of a filter, however, requires great patience and experience; it may be compared with retouching in sensitive optical systems. Usually it is sufficient to maintain a uniform angle ε in all plates and to be sure the plates are uniform. The tolerances have already been indicated. Further, I proved by experiment that it is possible to tune filters within wide limits in the directions of increasing and decreasing wavelength. The approximate formulas for the computation of these shifts are also given. At the same time, an analysis of the maximum permissible convergence was made on the basis of the foregoing results; this was also expressed by an approximation formula.

The present work is part of a doctoral dissertation done at the Institute of Electrotechnical Physics, Prague.

submitted

20 May 1954

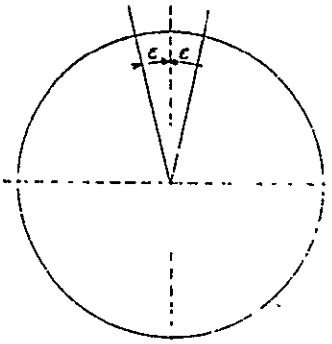


Fig. 1a

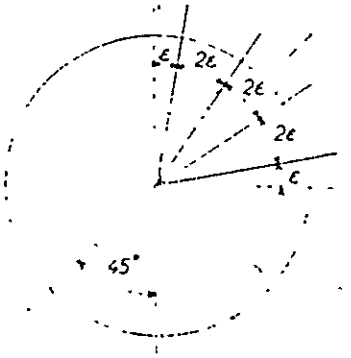


Fig. 1b

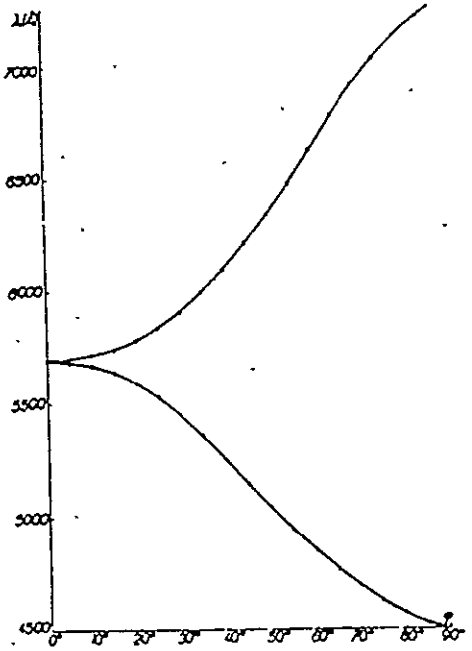


Fig. 2

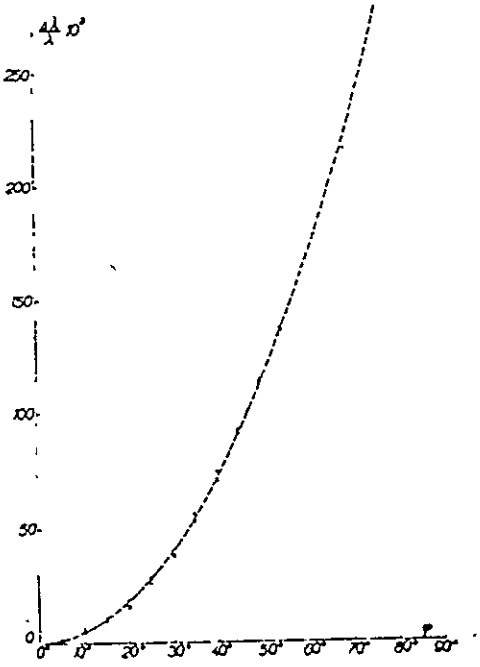


Fig. 3

CHAIN BIREFRINGENT FILTERS

by

Ivan Šolc

INTRODUCTION

We use the term chain birefringent filter to designate the arrangement of birefringent plates in which no plate is separated from another by polaroids and in which the path difference of the individual filter elements is repeated periodically. Hence, the simplest arrangement answering this definition is a system of plates of equal thickness, ground from the same material and with the same orientation of the angle of cut. Filters are either asymmetrical (the filter effect can be changed by changing the direction of the rays) or symmetrical (the filter effect is independent of the direction of the rays), depending on the arrangement of the azimuths of the individual filter elements. For practical reasons, symmetrical filters are the most convenient, so we shall concentrate on them in this paper. Birefringent filters are generally used to make light monochromatic, and thus it is our object to attain the optimum effect along these lines.

BASIC ASSUMPTIONS

A ray may not change its path direction while passing through the filter. Hence it follows that only one cut of a crystal is suitable, namely the cut parallel to the plane of the optic axes (uniaxial crystals are cut parallel with the optic axis). Huygen's construction of the refraction of light in crystals shows clearly that all other cuts are unsuitable. I described the simplest type of chain filter in an article published in 1953 [1]. A more detailed analysis of such filters can be found in [2]. The system of equations derived by Hsien-Yü Hsü, L. Richartz and Yüng-Kang Liang in 1947 [5] from the general formula for computing the intensity of polarized light passing through a series of birefringent plates has proved to be effective for analysis of the filter function

$$J_n = J_0 \cdot \left[\prod_{k=1}^{n} \cos \gamma_k\right]^2 \cdot \left\{\left[\sum_{i=0}^{n/2} (-)^i \sum_{j_0 \ldots j_{2i-1}}^{2i} \prod_{k=1}^{2i} x_{jk} \cos\left(\Phi - 2\sum_{k=1}^{2i} (-)^k \varrho_{jk}\right)\right]^2 + \left[\sum_{i=0}^{n/2} (-)^i \sum_{j_0 \ldots j_{2i-1}}^{2i+1} \prod_{k=1}^{2i+1} x_{jk} \cos\left(\Phi - 2\sum_{k=1}^{2i+1} (-)^{k-1} \varrho_{jk}\right)\right]^2\right\}. \quad (1)$$

In this formula, Φ is the azimuth angle of the analyzer with respect to the azimuth of the polaroid, $x_i = \tan \delta_s$, $\gamma = \delta_s/2$, δ is the phase difference, ρ_1 the azimuth of the i-th plate, J_o the intensity of the light after the first polaroid. The phase difference and the path difference are related by the familiar relationship

$$\delta = \frac{2\pi}{\lambda} \cdot d \cdot D \qquad (2)$$

in which λ is the wavelength of light, d is the plate thickness and D is the index of refraction.

To make satisfactory birefringent filters, one must use suitable, highly homogeneous birefringent crystal material, maintain a strictly accurate angle of cut, and control the grinding such that the plate thickness remains strictly within the assigned tolerances. A condition for the use of birefringent filters is that they must fit into an optical system in which the allowed angle of convergence of the rays is not exceeded.

GENERAL RELATIONSHIPS FOR SIMPLE CHAIN FILTERS

For simple chain filters, with which we are concerned here, the individual elements have the same path difference, which is expressed by the condition

$$\gamma_1 = \gamma_2 = \gamma_3 = \dots = \gamma_s = \gamma,$$

then, also,

$$x_1 = x_2 = x_3 = \dots = x_s = x. \qquad (3)$$

In further computations, intensity J_o will be taken as 1.

We will use X_n for the double brackets in general expression (1). Then, for example,

$$\begin{aligned}
X_1 &= \cos^2 \Phi + [x_1 \cos (\Phi - 2\rho_1)]^2, \\
X_2 &= [\cos \Phi - x_2 x_1 \cos (\Phi - 2\rho_2 + 2\rho_1)]^2 + \\
&\quad + [x_2 \cos (\Phi - 2\rho_2) + x_1 \cos (\Phi - 2\rho_1)]^2, \\
X_3 &= [\cos \Phi - x_3 x_2 \cos (\Phi - 2\rho_3 + 2\rho_2) + x_3 x_1 \cos (\Phi - 2\rho_3 + 2\rho_1) - \\
&\quad - x_2 x_1 \cos (\Phi - 2\rho_2 + 2\rho_1)]^2 + [x_1 \cos (\Phi - 2\rho_1) \\
&\quad + x_2 \cos (\Phi - 2\rho_2) + x_1 \cos (\Phi - 2\rho_1) - x_3 x_2 x_1 \cdot \\
&\quad \cdot \cos (\Phi - 2\rho_3 + 2\rho_2 - 2\rho_1)]^2,
\end{aligned}$$

etc.

Next, let us introduce the expression $I_{u,v}$:

$$L_{u,v} = \sum_{i_u=\ldots i_1=1}^{u} \cos\left(\Phi - 2\sum_{k=1}^{v}(-1)^{k-1}\varrho_{ik}\right). \tag{4}$$

By successive substitution, we get

$$\begin{aligned}
L_{1,1} &= \cos(\Phi - 2\varrho_1),\\
L_{2,1} &= \cos(\Phi - 2\varrho_2) + \cos(\Phi - 2\varrho_1),\\
L_{2,2} &= \cos(\Phi - 2\varrho_2 + 2\varrho_1),\\
L_{3,1} &= \cos(\Phi - 2\varrho_3) + \cos(\Phi - 2\varrho_2) + \cos(\Phi - 2\varrho_1),\\
L_{3,2} &= \cos(\Phi - 2\varrho_3 + 2\varrho_2) + \cos(\Phi - 2\varrho_3 + 2\varrho_1) + \cos(\Phi - 2\varrho_2 + 2\varrho_1),\\
L_{3,3} &= \cos(\Phi - 2\varrho_3 + 2\varrho_2 - 2\varrho_1),
\end{aligned}$$

etc.

For the sake of clarity, we may express these relationships by the following numerical scheme:

$L_{1,1}$ 1
$L_{2,1}$ 2, 1
$L_{2,2}$ 21
$L_{3,1}$ 3, 2, 1
$L_{3,2}$ 32, 31, 21
$L_{3,3}$ 321
$L_{4,1}$ 4, 3, 2, 1
$L_{4,2}$ 43, 42, 41, 32, 31, 21
$L_{4,3}$ 432, 431, 421, 321
$L_{4,4}$ 4321

etc.

By successive substitution of $L_{u,v}$ and relationships (3) for X_n, we arrive at equations

$$\begin{aligned}
X_1 &= \cos^2\Phi + [xL_{1,1}]^2,\\
X_2 &= [\cos\Phi - x^2L_{2,2}]^2 + [xL_{2,1}]^2,\\
X_3 &= [\cos\Phi - x^2L_{2,2}]^2 + [xL_{2,1} - x^3L_{2,2}]^2,\\
X_4 &= [\cos\Phi - x^2L_{4,2} + x^4L_{4,4}]^2 + [xL_{4,1} - x^3L_{4,3}]^2,\\
X_5 &= [\cos\Phi - x^2L_{5,2} + x^4L_{5,4}]^2 + [xL_{5,1} - x^3L_{5,3} + x^5L_{5,5}]^2,\\
X_n &= [\cos\Phi - x^2L_{n,2} + x^4L_{n,4} \cdots \pm x^tL_{n,t}]^2 + [xL_{n,1} - x^3L_{n,3} + \cdots \pm \\
&\quad \pm x^tL_{n,t}]^2.
\end{aligned}$$

In this equation, s is the largest even number that fulfills the condition $s \leqq n$, and t is the largest odd number that fulfills the similar condition $t \leqq n$.

It is convenient to arrange these expressions by decreasing powers of x:

$$\begin{aligned}
X_1 &= x^2L_{1,1}^2 + \cos^2\Phi,\\
X_2 &= x^4L_{2,2}^2 + x^2(L_{2,1}^2 - 2L_{2,2}\cos\Phi) + \cos^2\Phi,\\
X_3 &= x^6L_{3,3}^2 + x^4(L_{3,2}^2 - 2L_{3,1}L_{3,3}) + x^2(L_{3,1}^2 - 2L_{3,2}\cos\Phi) + \cos^2\Phi,\\
X_4 &= x^8L_{4,4}^2 + x^6(L_{4,3}^2 - 2L_{4,2}L_{4,4}) + x^4(L_{4,2}^2 - 2L_{4,1}L_{4,3} + 2L_{4,4}\cos\Phi) + \\
&\quad + x^2(L_{4,1}^2 - 2L_{4,2}\cos\Phi) + \cos^2\Phi,\\
X_5 &= x^{10}L_{5,5}^2 + x^8(L_{5,4}^2 - 2L_{5,3}L_{5,5}) + x^6(L_{5,3}^2 - 2L_{5,2}L_{5,4} + 2L_{5,1}L_{5,5}) + \\
&\quad + x^4(L_{5,2}^2 - 2L_{5,1}L_{5,3} + 2L_{5,4}\cos\Phi) + x^2(L_{5,1}^2 - 2L_{5,2}\cos\Phi) + \\
&\quad \cos^2\Phi,
\end{aligned}$$

etc.

The brackets with functions $L_{u,v}$ with values of x are of the nature of coefficients and express the characteristic properties of a chain filter. Let us call them $A_{k,\ell}$. Then,

$$A_{1,1} = L^2_{1,1},$$
$$A_{2,1} = L^2_{2,2},$$
$$A_{3,1} = L^2_{3,3},$$
$$A_{4,1} = L^2_{4,4},$$

$$A_{k,1} = L^2_{k,k};$$

$$A_{1,2} = \cos\Phi,$$
$$A_{2,2} = L^2_{2,1} - 2L_{2,2}\cos\Phi,$$
$$A_{3,2} = L^2_{3,2} - 2L_{3,1}L_{3,3},$$
$$A_{4,2} = L^2_{4,3} - 2L_{4,2}L_{4,4},$$
$$A_{5,2} = L^2_{5,4} - 2L_{5,3}L_{5,5},$$

$$A_{k,2} = L^2_{k,k-1} - 2L_{k,k-2}L_{k,k};$$

$$A_{1,3} = 0,$$
$$A_{2,3} = \cos\Phi,$$
$$A_{3,3} = L^2_{3,1} - 2L_{3,2}\cos\Phi,$$
$$A_{4,3} = L^2_{4,2} - 2L_{4,1}L_{4,3} - 2L_{4,4}\cos\Phi,$$
$$A_{5,3} = L^2_{5,3} - 2L_{5,2}L_{5,4} + 2L_{5,1}L_{5,5},$$
$$A_{6,3} = L^2_{6,4} - 2L_{6,3}L_{6,5} + 2L_{6,2}L_{6,6},$$

$$A_{k,3} = L^2_{k,k-2} - 2L_{k,k-1}L_{k,k-1} + 2L_{k,k-4}L_{k,k};$$

$$A_{1,4} = 0,$$
$$A_{2,4} = 0,$$
$$A_{3,4} = \cos\Phi,$$
$$A_{4,4} = L^2_{4,1} - 2L_{4,2}\cos\Phi,$$
$$A_{5,4} = L^2_{5,2} - 2L_{5,1}L_{5,3} + 2L_{5,4}\cos\Phi,$$
$$A_{6,4} = L^2_{6,3} - 2L_{6,2}L_{6,4} + 2L_{6,1}L_{6,5} - 2L_{6,6}\cos\Phi,$$
$$A_{7,4} = L^2_{7,4} - 2L_{7,3}L_{7,5} + L2_{7,2}L_{7,6} + 2L_{7,1}L_{7,7},$$

$$A_{k,4} = L^2_{k,k-3} - 2L_{k,k-4}L_{k,k-2} + 2L_{k,k-5}L_{k,k-1} - 2L_{k,k-6}L_{k,k};$$

$$A_{1,5} = 0,$$
$$A_{2,5} = 0,$$
$$A_{3,5} = 0,$$
$$A_{4,5} = \cos\Phi,$$
$$A_{5,5} = L^2_{5,1} - 2L_{5,2}\cos\Phi,$$
$$A_{6,5} = L^2_{6,2} - 2L_{6,1}L_{6,3} + 2L_{6,4}\cos\Phi$$
$$A_{7,5} = L^2_{7,3} - 2L_{7,2}L_{7,4} + 2L_{7,1}L_{7,5} - 2L_{7,6}\cos\Phi,$$

$$A_{8,5} = L^2_{8,4} - 2L_{8,3}L_{8,5} + 2L_{8,2}L_{8,6} - 2L_{8,1}L_{8,7} + 2L_{8,8}\cos\Phi,$$
$$A_{9,5} = L^2_{9,6} - 2L_{9,4}L_{9,6} + 2L_{9,3}L_{9,7} - 2L_{9,2}L_{9,8} + 2L_{9,1}L_{9,9},$$

$$A_{k,5} = L^2_{k,k-4} - 2L_{k,k-5}L_{k,k-3} + 2L_{k,k-6}L_{k,k-2} - 2L_{k,k-7}L_{k,k-1} + 2L_{k,k-8}L_{k,k},$$
etc.

The formulas combine formally if we introduce the symbol $\cos\Phi = L_{k,0}$. Then the general relationship for expression $A_{k,\ell}$ has the form

$$A_{k,l} = L^2_{k,k-l+1} - 2L_{k,k-l}L_{k,k-l+2} + 2L_{k,k-l-1}L_{k,k-l+3} - \\ - 2L_{k,k-l-2}L_{k,k-l+4} \cdots \pm 2L_{k,1}L_{k,k} \qquad (5)$$

If we use coefficients of $A_{k,\ell}$, the formula for X_n assumes the form

$$\begin{aligned} X_1 &= x^2A_{1,1} + \cos^2\varphi, \\ X_2 &= x^4A_{2,1} + x^2A_{2,2} + \cos^2\varphi, \\ X_3 &= x^6A_{3,1} + x^4A_{3,2} + x^2A_{3,3} + \cos^2\varphi, \\ X_4 &= x^8A_{4,1} + x^6A_{4,2} + x^4A_{4,3} + x^2A_{4,4} + \cos^2\varphi, \\ X_5 &= x^{10}A_{5,1} + x^8A_{5,2} + x^6A_{5,3} + x^4A_{5,4} + x^2A_{5,5} + \cos^2\varphi, \end{aligned} \qquad (6)$$

$$X_n = x^{2n}A_{n,1} + x^{2n-2}A_{n,2} + \ldots + x^2A_{n,n} + \cos^2\varphi.$$

With the aid of formula (1) we may convert equation (6) into expressions suitable for direct calculation of intensity:

$$\begin{aligned} J_1 &= \sin^2\gamma \cdot A_{1,1} + \cos^2\gamma\cos^2\Phi, \\ J_2 &= \sin^4\gamma \cdot A_{2,1} + \sin^2\gamma\cos^2\gamma \cdot A_{2,2} + \cos^4\gamma\cos^2\Phi, \\ J_3 &= \sin^6\gamma \cdot A_{3,1} + \sin^4\gamma\cos^2\gamma \cdot A_{3,2} + \sin^2\gamma\cos^4\gamma \cdot A_{3,3} + \cos^6\gamma\cos^2\Phi, \\ J_4 &= \sin^8\gamma \cdot A_{4,1} + \sin^6\gamma\cos^2\gamma \cdot A_{4,2} + \sin^4\gamma\cos^4\gamma \cdot A_{4,3} + \sin^2\gamma\cos^6\gamma \cdot \\ &\quad A_{4,4} + \cos^8\gamma\cos^2\Phi, \end{aligned} \qquad (7)$$

$$J_n = \sin^{2n}\gamma \cdot A_{n,1} + \sin^{2n-2}\gamma\cos^2\gamma \cdot A_{n,2} + \ldots + \sin^2\gamma\cos^{2n-2}\gamma \cdot A_{n,n} + \\ + \cos^{2n}\gamma\cos^2\Phi.$$

Equations (4), (5) and (7) are general relationships describing simple chain filters, both symmetrical and asymmetrical.

In what follows, we shall concentrate on symmetrical filters. We shall use the term "ideal" to define the filter assembly with the simplest arrangement of azimuths of the individual plates. We shall use the term "optimum" to define the filter assembly in which the attempt is made to reduce the secondary maxima as much as possible through arrangement of the azimuths of the individual plates. We shall also distinguish two different cases described by angle Φ. In the first case, angle Φ will be either 0° or 90° (basic positions), in the second case angle Φ will be general (general positions). Of course, all cases will have to be divided into two modifications, the first and the second, which are mutually complementary. We shall take up all the aforementioned cases systematically but separately.

THE FIRST MODIFICATION OF THE CHAIN FILTER IN THE BASIC POSITION, IDEAL ASSEMBLY [OF PLATES]

In this case,

$$\varphi = 90^\circ, \quad \begin{aligned} &\varrho_1 = \varrho_3 = \varrho_5 = \varrho_7 = \ldots = \varrho_{2k+1} = \varrho, \\ &\varrho_2 = \varrho_4 = \varrho_6 = \varrho_8 = \ldots = \varrho_{2k} = -\varrho. \end{aligned} \qquad (8)$$

Through substitution, equation (4) takes on the form

$$
\begin{aligned}
&L_{1,1} = \sin 2\varrho\,,\\
&L_{2,1} = 0\,,\\
&L_{2,2} = -\sin 4\varrho\,,\\
&L_{3,1} = \sin 2\varrho\,,\\
&L_{3,2} = 0\,,\\
&L_{3,3} = \sin 6\varrho\,,\\
&L_{4,1} = 0\,,\\
&L_{4,2} = -2\sin 4\varrho\,,\\
&L_{4,3} = 0\,,\\
&L_{4,4} = -\sin 8\varrho\,,\\
&L_{5,1} = \sin 2\varrho\,,\\
&L_{5,2} = 0\,,\\
&L_{5,3} = 3\sin 6\varrho + \sin 2\varrho\,,\\
&L_{5,4} = 0\,,\\
&L_{5,5} = \sin 10\varrho\,,\\
&\text{etc.}
\end{aligned}
$$

Equation (5) simplifies as follows:

$$
\begin{aligned}
&A_{1,1} = \sin^2 2\varrho\,,\\
&A_{2,1} = \sin^2 4\varrho\,,\\
&A_{3,1} = \sin^2 6\varrho\,,\\
&A_{4,1} = \sin^2 8\varrho\,,\\
&A_{5,1} = \sin^2 10\varrho\,;
\end{aligned}
$$

$$
\begin{aligned}
&A_{1,2} = 0\,,\\
&A_{2,2} = 0\,,\\
&A_{3,2} = -2\sin 6\varrho\sin 2\varrho\,,\\
&A_{4,2} = 4\sin 8\varrho\sin 4\varrho\,,\\
&A_{5,2} = -6\sin 10\varrho\sin 6\varrho - 2\sin 10\varrho\sin 2\varrho\,,
\end{aligned}
$$

$$
\begin{aligned}
&A_{1,3} = 0\,,\\
&A_{2,3} = 0\,,\\
&A_{3,3} = \sin^2 2\varrho\,,\\
&A_{4,3} = 4\sin^4 4\varrho\,,\\
&A_{5,3} = 9\sin^2 6\varrho + \sin^2 2\varrho - 6\sin 6\varrho\sin 2\varrho + 2\sin 10\varrho\sin 2\varrho
\end{aligned}
\qquad (5a)
$$

$$
\begin{aligned}
&A_{1,4} = 0\,,\\
&A_{2,4} = 0\,,\\
&A_{3,4} = 0\,,\\
&A_{4,4} = 0\,,\\
&A_{5,4} = -6\sin 6\varrho\sin 2\varrho - 2\sin^2 2\varrho\,,\\
&\text{etc.}
\end{aligned}
$$

Due to substitution in intensity formulas (7), a condition is created by which the intensity is maximum in the main maxima. This is the case when $A_{1,1} = A_{2,1} = A_{3,1} = \ldots = A_{n,1} = 1$, or

for 1 plate $L^2_{1,1} = \sin^2 2\varrho = 1$,

for 2 plates $L^2_{2,2} = \sin^2 4\varrho = 1$,

for 3 plates $L^2_{3,3} = \sin^2 6\varrho = 1$,

for k plates $L^2_{k,k} = \sin^2 2k\varrho = 1$.

The fundamental relationship for attaining maximum transmission

$$\rho = \frac{45^\circ}{\text{number of plates}} \qquad (9)$$

follows from the above equations. By substitution of condition (9) in equation (4), we can construct the following table of numerical values of $L_{n,v}$.

Table 1

n \ v	1	2	3	4	5	6	7	8	9
1	1								
2	0	−1							
3	0,500	0	1						
4	0	−1,414	0	−1					
5	0,309	0	2,736	0	1				
6	0	−1,500	0	−4,464	0	−1			
7	0,222	0	4,418	0	6,375	0	1		
8	0	−1,630	0	−9,758	0	−9,137	0	−1	
9	0,174	0	6,04	0	16,60	0	12,082	0	1

Values of $L_{n,v}$ with higher indexes can be approximated roughly, e.g., by extrapolation on graph paper.

Using table 1, we can construct table 2 for $A_{k,\ell}$, on the basis of relationship (5).

Table 2

k \ l	1	2	3	4	5	6	7	8	9
1	1								
2	1	0							
3	1	− 1	0,250						
4	1	− 2,818	2,000	0					
5	1	− 5,472	8,104	− 1,691	0.0955				
6		− 8,928	22,927	− 13,392	2,250	0			
7	1	−12,750	49,477	− 56,774	22,350	− 1,962	0.049		
8	1	−18,274	103,001	−181,38	123,179	− 29,859	2,341	0	
9	1	−24,164	178,97	−410,79	411,124	−203,625	42,224	− 2,102	0,026

Substituting numerical values in formula (7), we establish the regularity of the variation of intensity of an ideal assembly of the first-modification chain filter:

$$
\begin{aligned}
J_1 &= \sin^2\gamma\,,\\
J_2 &= \sin^4\gamma\,,\\
J_3 &= \sin^6\gamma - \sin^4\gamma\cos^2\gamma + 0{,}25\sin^2\gamma\cos^4\gamma\,,\\
J_4 &= \sin^8\gamma - 2{,}818\sin^6\gamma\cos^2\gamma + 2\sin^4\gamma\cos^4\gamma\,,\\
J_5 &= \sin^{10}\gamma - 5{,}472\sin^8\gamma\cos^2\gamma + 8{,}104\sin^6\gamma\cos^4\gamma - 1{,}691\,\cdot\\
&\quad \cdot\sin^4\gamma\cos^6\gamma + 0{,}095\sin^2\gamma\cos^8\gamma\,,
\end{aligned}
\qquad (7a)
$$

etc.

By successive substitution of the path differences, we find the curve of the transmission capacity of the filter. The expressions for intensity are polynomial in form. The main maximum is characterized by the first term, the subsequent terms pertain to the secondary maxima and minima. Their number also determines the number of secondary extremes.

By analyzing the first term, we can arrive at the position of the main maximum, which becomes

$$\gamma = \frac{\delta}{2} = \frac{\pi}{2}\,,\ \frac{3}{2}\pi,\ \ldots,\ \frac{2k+1}{2}\pi\,.$$

By substituting this condition in equation (2), we get the familiar relationship for the wavelengths of maxima,

$$d \cdot D = \frac{2k+1}{2}\cdot\lambda\,.$$

The position of the secondary maxima can also be defined roughly as the extremes of the subsequent individual terms of the polynomial, which have the form

$$y = \sin^{2k}\gamma\cdot\cos^{2\ell}\gamma\,.$$

The extremes of this function are given by the equation

$$\sin\gamma = \sqrt{\frac{k}{k+\ell}}\,.$$

In these extremes, the function has the value

$$y_{max} = \frac{k^k\cdot\ell^\ell}{(k+\ell)^{k+\ell}}\,.$$

The true position of the secondary maxima is always somewhat different, since the extremes of the entire polynomial must be considered. Thus, only half the extremes act as maxima, the other half correspond approximately to minima.

The curves of the transmission [of light] for filters with one to nine

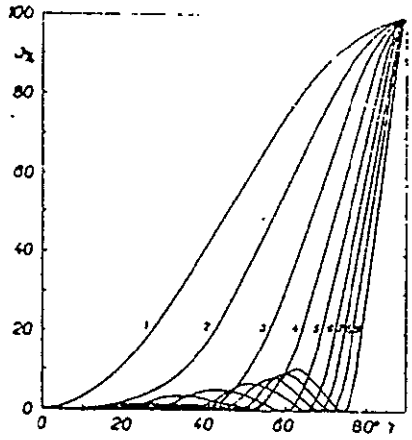


Figure 1. The variation of intensity of first-modification chain filters with one to nine plates, ideal assembly in the basic position.

plates are computed on the basis of equations (7). This is illustrated in figure 1. An interesting fact becomes apparent from equations (7), namely that a filter with two plates has no secondary maximum. This could be used in constructing a Lyot filter. Further, computation shows that the transmission curve is considerably steeper near the main maximum than one would expect from the first term of the polynomial itself. The main maximum is always somewhat broader than the case of a single plate which is as thick as the total number of plates of a chain filter.

THE SECOND MODIFICATION OF A CHAIN FILTER IN THE BASIC POSITION WITH AN IDEAL ASSEMBLY [OF PLATES]

In this case,

$$\Phi = 0^\circ, \quad \varrho_1 = \varrho, \ \varrho_2 = 3\varrho, \ \varrho_3 = 5\varrho, \ \ldots, \ \varrho_n = (2n-1)\,\varrho\,.$$

By substitution in equations (4) we get

$$\begin{aligned}
&L_{1,1} = 0\,,\\
&L_{2,1} = 0\,,\\
&L_{2,2} = -\cos 4\varrho\,,\\
&L_{3,1} = 0\,,\\
&L_{3,2} = 2\sin 2\varrho - \cos 4\varrho\,,\\
&L_{3,3} = 0\,,\\
&L_{4,1} = 0\,,\\
&L_{4,2} = \cos 4\varrho - \cos 12\varrho\,,\\
&L_{4,3} = 0\,,\\
&L_{4,4} = \cos 8\varrho\,,\\
&L_{5,1} = 0\,,\\
&L_{5,2} = 2\sin 6\varrho + 2\sin 2\varrho - 2\cos 8\varrho - \cos 12\varrho + \cos 4\varrho\,,\\
&L_{5,3} = 0\,,\\
&L_{5,4} = \cos 8\varrho\,,\\
&L_{5,5} = 0\,,
\end{aligned} \tag{4b}$$

etc.

Equation (5) simplifies as follows:

$$\begin{aligned}
&A_{1,1} = 0\,,\\
&A_{2,1} = \cos^2 4\varrho\,,\\
&A_{3,1} = 0\,,\\
&A_{4,1} = \cos^2 8\varrho\,,\\
&A_{5,1} = 0\,;
\end{aligned}$$

$$A_{1,2} = 1 ,$$
$$A_{2,2} = 2 \cos 4\varrho ,$$
$$A_{3,2} = 4 \sin^2 2\varrho + \cos^2 4\varrho - 4 \sin 2\varrho \cos 4\varrho ,$$
$$A_{4,2} = \cos 12\varrho \cos 8\varrho - \cos 8\varrho \cos 4\varrho ,$$
$$A_{5,2} = \cos^2 8\varrho ;$$

$$A_{1,3} = 0 ,$$
$$A_{2,3} = 1 ,$$
$$A_{3,3} = 2 \cos 4\varrho - 4 \sin 2\varrho ,$$
$$A_{4,3} = \cos^2 4\varrho + \cos^2 12\varrho - 2 \cos 4\varrho \cos 12\varrho + 2 \cos 8\varrho ,$$
$$A_{5,3} = 4 \cos^2 8\varrho + 2 \cos 12\varrho \cos 8\varrho - 4 \cos 8\varrho \sin 6\varrho - 4 \cos 8\varrho \sin 6\varrho - 2 \cos 8\varrho \cos 4\varrho ;$$

$$A_{1,4} = 0 ,$$
$$A_{2,4} = 0 ,$$
$$A_{3,4} = 1 ,$$
$$A_{4,4} = 2 \cos 12\varrho - 2 \cos 4\varrho ,$$
$$A_{5,4} = 4 \sin^2 6\varrho + 4 \sin^2 2\varrho + 4 \cos^2 8\varrho + \cos^2 12\varrho + \cos^2 4\varrho + 8 \sin 6\varrho \sin 2\varrho - 8 \sin 6\varrho \cos 8\varrho - 4 \sin 6\varrho \cos 12\varrho + 4 \sin 6\varrho \cos 4\varrho + 4 \sin 2\varrho \cos 4\varrho - 4 \sin 2\varrho \cos 12\varrho - 8 \sin 2\varrho \cos 8\varrho + 4 \cos 12\varrho \cos 8\varrho - 4 \cos 8\varrho \cos 4\varrho - 2 \cos 12\varrho \cos 4\varrho + 2 \cos 8\varrho ,$$

etc.

The maximum intensity in the main maxima of the second modification need not be verified separately, since the last term of equations (7) has the coefficient $\cos^2 \Phi = 1$.

However, the second modification can be made complementary to the first by proper choice of angle ρ. This angle is given by relationship (9). Through substitution of its numerical values in equations (4), we can again construct a table for $L_{u,v}$.

Table 3

n \ v	1	2	3	4	5	6	7	8	9
1	0								
2	0	0							
3	0	0,500	0						
4	0	1,414	0	0					
5	0	2,736	0	0,309	0				
6	0	4,464	0	1,500	0	0			
7	0	6,375	0	4,418	0	0,222	0		
8	0	9,137	0	9,758	0	1,530	0	0	
9	0	12,082	0	16,50	0	6,04	0	0,174	

A comparison of this table with table 1 shows the mutual relationship of coefficients $L_{u,v}$ for both filter modifications.

Again we can construct a table for coefficients of $A_{k,\ell}$ on the basis of relationships (5).

Table 4

k \ ℓ	1	2	3	4	5	6	7	8	9
1	0								
2	0	0							
3	0	0,250	−1						
4	0	0	+2	− 2,818					
5	0	0.0955	−1,691	+ 8,104	− 5,472				
6	0	0	+2,250	−13,392	+ 22,927	− 8,928			
7	0	0,049	−1,962	+22,350	− 56,774	+ 49,477	− 12,750		
8	0	0	+2,341	−29,859	+123,179	−181,38	+103,00	− 18,274	
9	0	0,030	−2,102	+42,224	−203,525	+411,124	−410,79	+178,97	−24,164

This time we construct the equation for intensity in the reverse order of terms of the polynomial. This shows the complementary nature of the two modifications more clearly:

$$
\begin{aligned}
J_1 &= \cos^2\gamma\,,\\
J_2 &= \cos^4\gamma\,,\\
J_3 &= \cos^6\gamma - \cos^4\gamma\sin^2\gamma + 0{,}25\cos^2\gamma\sin^4\gamma\,,\\
J_4 &= \cos^8\gamma - 2{,}818\cos^6\gamma\sin^2\gamma + 2\cos^4\sin^4\gamma\,,\\
J_5 &= \cos^{10}\gamma - 5{,}472\cos^8\gamma\sin^2\gamma + 8{,}104\cos^6\gamma\sin^4\gamma - 1{,}691\cos^4\sin^6\gamma +\\
&\quad + 0{,}0955\cos^2\gamma\sin^8\gamma\,,
\end{aligned}
\qquad (7b)
$$

etc.

The system of equations (7b) is formally identical with system (7a), except that the functions of sin γ and cos γ are interchanged. We get complete agreement by introducing a complementary angle. In other words, the transmission curves of the second-modification filter are shifted 90° with respect to those of the first-modification filter, but otherwise are in agreement. The condition for the position of the main maxima of the second-modification filter, $\gamma = \pi, 2\pi, \ldots, k\pi$, follows from equations (7b), which leads to relationship

$$d \cdot D = k \cdot \ell\,.$$

Otherwise, the analysis of the intensity variation is the same as that discussed for the first filter modification.

THE FIRST MODIFICATION OF A SYMMETRICAL CHAIN FILTER IN THE BASIC POSITION, OPTIMUM ASSEMBLY [OF PLATES]

It is quite difficult to define the optimum assembly, for it is difficult to define this case mathematically, since the variation of the main maximum deteriorates somewhat due to the condition that the secondary maxima be suppressed. Before we begin the main analysis of the requirements of an optimum assembly, let us derive the general equation for a symmetrical first-modification filter in the basic position.

In this case,

$$\Phi = \frac{\pi}{2}, \quad |\varrho_1| = |\varrho_n|, \; |\varrho_2| = |\varrho_{n-1}|, \; \ldots, \; |\varrho_i| = |\varrho_{n-i+1}|.$$

Again the signs of the azimuths alternate. The condition for maximum transmission of the filter follows from the general equation

$$A_{1,1} = A_{2,1} = A_{3,1} = \ldots = A_{n,1} = 1,$$

or else

$$L^2_{1,1} = L^2_{2,2} = L^2_{3,3} = \ldots = L^2_{n,n} = 1.$$

According to equation (4), this relationship leads to the result

$$|\varrho_1| + |\varrho_2| + |\varrho_3| + \ldots + |\varrho_n| = 45^\circ, \tag{10}$$

which is the generalized relationship (9). Then, in equations (4), we will substitute, successively,

for 1 plate	ϱ_1,	$\varrho_1 = 45^\circ$,
for 2 plates	$\varrho_2 = -\varrho_1$,	$\varrho_1 - \varrho_2 = 45^\circ$,
for 3 plates	$\varrho_3 = \varrho_1$, ϱ_2,	$2\varrho_1 - \varrho_2 = 45^\circ$,
for 4 plates	$\varrho_4 = -\varrho_1$, $\varrho_3 = -\varrho_2$,	$2\varrho_1 - 2\varrho_2 = 45^\circ$,
for 5 plates	$\varrho_5 = \varrho_1$, $\varrho_4 = \varrho_2$, ϱ_3,	$2\varrho_1 - 2\varrho_2 + \varrho_3 = 45^\circ$,
for 6 plates	$\varrho_6 = -\varrho_1$, $\varrho_5 = -\varrho_2$, $\varrho_4 = -\varrho_3$,	$2\varrho_1 - 2\varrho_2 - 2\varrho_3 = 45^\circ$,

etc.

The general formulas (4) specialize as follows:

$$\begin{aligned} L_{1,1} &= \sin 2\varrho_1 \; (= 1), \\ L_{2,1} &= 0, \\ L_{2,2} &= \sin (2\varrho_2 - 2\varrho_1) \; (= -1), \end{aligned} \tag{4c}$$

$$
\begin{aligned}
&L_{3,1} = 2\sin^2\varrho_1 - \sin 2\varrho_2\,,\\
&L_{3,2} = 0\,,\\
&L_{3,3} = \sin(4\varrho_1 - 2\varrho_2)\ (=1)\,,\\
&L_{4,1} = 0\,,\\
&L_{4,2} = -\sin 4\varrho_1 - \sin 4\varrho_2 - 2\sin(2\varrho_2 + 2\varrho_1) + 2\sin(2\varrho_2 - 2\varrho_1)\,,\\
&L_{4,3} = 0\,,\\
&L_{4,4} = \sin(4\varrho_2 - 4\varrho_1)\ (=-1)\,,\\
&L_{5,1} = 2\sin 2\varrho_1 + 2\sin 2\varrho_2 + \sin 2\varrho_3\,,\\
&L_{5,2} = 0\,,\\
&L_{5,3} = 2\sin(2\varrho_1 - 2\varrho_2 + 2\varrho_3) + 2\sin 2\varrho_1 + 2\sin(4\varrho_1 - 2\varrho_2) +\\
&\qquad + 2\sin(2\varrho_1 + 2\varrho_2 - 2\varrho_3) + \sin(4\varrho_1 - 2\varrho_3) + \sin(4\varrho_2 - 2\varrho_3)\,,\\
&L_{5,4} = 0\,,\\
&L_{5,5} = \sin(4\varrho_1 - 4\varrho_2 + 2\varrho_3)\ (=1)\,,\\
&\text{etc.}
\end{aligned}
\qquad \text{(4c) cont.}
$$

According to equations (7a), one- and two-plate filters in the ideal assembly do not have secondary maxima; therefore, they are, simultaneously, filters in the optimum assembly. The three-plate filter has secondary maxima characterized by the coefficients $A_{3,2}$ and $A_{3,3}$, according to equation (7). Now let us seek such a solution for azimuths ϱ_1 and ϱ_2 in which these coefficients are minimum. According to formulas (5),

$$
\begin{aligned}
A_{3,2} &= L_{3,2}^2 - 2L_{3,1}L_{3,3}\,,\\
A_{3,3} &= L_{3,1}^2 - 2L_{3,2}\cos\Phi
\end{aligned}
$$

According to (4c), the expressions reduce to

$$
\begin{aligned}
A_{3,2} &= -2L_{3,1}\,,\\
A_{3,3} &= L_{3,1}^2\,.
\end{aligned}
$$

Then we need concern ourselves only with the expression $L_{3,1}$. Let us seek the condition for its minimum, which can be used in this case to solve the equations

$$
\begin{aligned}
L_{3,1} &= 2\sin 2\varrho_1 + \sin 2\varrho_2 = 0\,,\\
2\varrho_1 - \varrho_2 &= 45^\circ\,.
\end{aligned}
$$

The quadratic equation

$$2\sin^2 2\varrho_1 + 2\sin 2\varrho_1 - 1 = 0\,,$$

stems from these equations; its solution gives the angles

$$\varrho_1 = 10.7^\circ, \quad \varrho_2 = -23.6^\circ.$$

With these plate azimuths, the three-plate filter has no secondary maximum. The equation for the variation of intensity has the form

$$J = \sin^6\gamma\,.$$

Proceeding in a similar manner, let us seek the solution of the equation $L_{4,2} = 0$ for a four-plate filter, or

$$\sin 4\varrho_1 + \sin 4\varrho_2 + 2\sin(2\varrho_1 + 2\varrho_2) + 2\sin(2\varrho_1 - 2\varrho_2) = 0,$$

$$2\varrho_1 - 2\varrho_2 = 45°.$$

By a simple operation we arrive at the equation

$$\sin 4\varrho_1 \cdot (1 + 2\cos 45°) - \cos 4\varrho_1 (1 + 2\sin 45°) + 2\sin 45° = 0,$$

whose solution gives the angles

$$\rho_1 = 7.6°, \quad \rho_2 = -14.9°.$$

A filter with these azimuths of the plates again has only a main maximum and its intensity variation is given by the equation

$$J = \sin^2 \gamma.$$

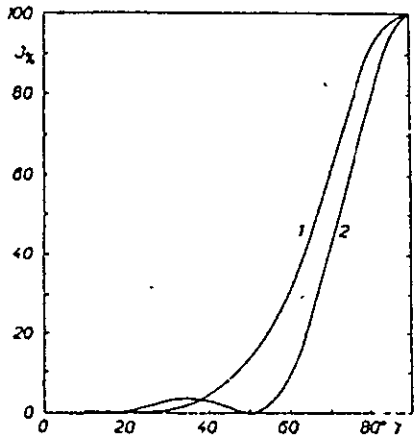


Figure 2. Variation of the intensity of a first-modification chain filter with four plates in an ideal assembly (the steeper curve with the secondary maximum) and in an optimum assembly (the flatter curve without a secondary maximum)

We can continue thus, but with ever-increasing computational difficulties. For instance, the solution of a single trigonometric equation no longer suffices. The distinction between the ideal assembly and the same assembly without secondary maxima is an important one. The two types of filters can best be judged on the basis of the intensity variation, which is shown in figure 2 for a four-plate filter.

It is easy to see that when the secondary maxima are eliminated, the main maximum broadens perceptibly. Therefore we need not concern ourselves with an exact analysis of the possibility of eliminating the secondary maxima in filters with a large number of plates, but we can be satisfied with some compromise solution in the choice of the azimuths of the plates. Furthermore, we are concerned with a considerable reduction of the intensity of the secondary maxima without essentially increasing the width of the main maximum. The following method has proved best: Select the azimuth angle of the first plate; for filters with more than two plates this angle is less than that indicated by formula (9) and, according to circumstances, reaches 50-75% of the value of the angles for an ideal assembly. The absolute values of the azimuth angles of the plates themselves increase progressively toward the center of the filter. The growth functions can be chosen

differently, and the arithmetic progression will also serve the purpose well.

$$\varrho_1, \quad \varrho_2 = -(\varrho_1 + \alpha), \quad \varrho_3 = \varrho_1 + 2\alpha, \quad \varrho_4 = -(\varrho_1 + 3\alpha),$$
$$\varrho_5 = \varrho_1 + 5\alpha \quad \text{etc.}$$

The azimuths of all the plates must then fulfill condition (10) +).

As an example of such an optimum assembly, let us analyze a filter with seven plates. We shall choose the coefficient of reduction of the first azimuth as 70%, the azimuths increase in arithmetic progression. Let us compute the angles

$$\begin{aligned} \varrho_1 &= 4{,}5^\circ, & \varrho_4 &= -9^\circ, \\ \varrho_2 &= -6{,}0^\circ, & \varrho_5 &= +7{,}5^\circ, \\ \varrho_3 &= 7{,}5^\circ, & \varrho_6 &= -6^\circ, \\ & & \varrho_7 &= 4{,}5^\circ. \end{aligned}$$

For seven plates, equations (4) have the form

$$\begin{aligned}
L_{7,1} &= 2\sin 2\varrho_1 + 2\sin 2\varrho_2 + 2\sin 2\varrho_3 + \sin 2\varrho_4, \\
L_{7,2} &= 0, \\
L_{7,3} &= 4\sin 2\varrho_1 + 2\sin 2\varrho_2 + 4\sin(2\varrho_1 - 2\varrho_2 + 2\varrho_3) + \\
&\quad + 2\sin(2\varrho_1 - 2\varrho_2 + 2\varrho_4) + 4\sin(2\varrho_1 + 2\varrho_2 - 2\varrho_3) + \\
&\quad + 2\sin(4\varrho_1 - 2\varrho_3) + 2\sin(4\varrho_1 - 2\varrho_3) + 2\sin(4\varrho_2 - 2\varrho_3) + \\
&\quad + 2\sin(2\varrho_1 - 2\varrho_2 + 2\varrho_4) + 2\sin(2\varrho_2 + 2\varrho_3 - 2\varrho_4) + \\
&\quad + 2\sin(2\varrho_1 + 2\varrho_3 - 2\varrho_4) + 2\sin(2\varrho_1 + 2\varrho_2 - 2\varrho_4) + \\
&\quad + 2\sin(2\varrho_2 - 2\varrho_3 + 2\varrho_4) + \sin(4\varrho_1 - 2\varrho_4) + \\
&\quad + \sin(4\varrho_2 - 2\varrho_4) + \sin(4\varrho_3 - 2\varrho_4), \\
L_{7,4} &= 0, \\
L_{7,5} &= 2\sin 2\varrho_1 + 2\sin(4\varrho_1 - 2\varrho_3) + 2\sin(2\varrho_1 + 2\varrho_3 - 2\varrho_4) + \\
&\quad + 2\sin(2\varrho_1 - 2\varrho_3 + 2\varrho_4) + 2\sin(4\varrho_1 - 2\varrho_2 + 2\varrho_3 - 2\varrho_4) + \\
&\quad + 2\sin(4\varrho_1 - 4\varrho_2 + 2\varrho_3) + 2\sin(2\varrho_1 - 4\varrho_3 + 2\varrho_2 + 2\varrho_4) + \\
&\quad + 2\sin(4\varrho_1 - 2\varrho_2 - 2\varrho_3 + 2\varrho_4) + \sin(4\varrho_1 - 4\varrho_3 + 2\varrho_4) + \\
&\quad + \sin(4\varrho_1 - 4\varrho_2 + 2\varrho_4) + \sin(4\varrho_2 - 4\varrho_3 + 2\varrho_4) + \\
&\quad + 2\sin(2\varrho_1 - 2\varrho_2 + 4\varrho_3 - 2\varrho_4), \\
L_{7,6} &= 0, \\
L_{7,7} &= \sin(4\varrho_1 - 4\varrho_2 + 4\varrho_3 - 2\varrho_4).
\end{aligned}$$

From these formulas we get the following values of $L_{u,v}$

$$\begin{aligned}
L_{7,1} &= +0{,}106, \\
L_{7,2} &= 0, \\
L_{7,3} &= +2{,}851, \\
L_{7,4} &= 0, \\
L_{7,5} &= 4{,}714, \\
L_{7,6} &= 0, \\
L_{7,7} &= 1.
\end{aligned}$$

+) The azimuth of the first plate can even be chosen as 0°; however, in this case the first and last plates will not function and one must reckon with a filter which has two more plates. The first and last plates are omitted in constructing the filter.

The corresponding coefficients of $A_{k,\ell}$ for equations (5) are

$$\begin{aligned} A_{7,1} &= 1, \\ A_{7,2} &= -9{,}428, \\ A_{7,3} &= 27{,}924, \\ A_{7,4} &= -27{,}092, \\ A_{7,5} &= 9{,}127, \\ A_{7,6} &= -0{,}0064, \\ A_{7,7} &= 0{,}0123. \end{aligned}$$

Then the variation of intensity is given by the equation

$$\begin{aligned} J_7 = \sin^{14}\gamma &- 9{,}428\sin^{12}\gamma\cos^2\gamma + \\ &+ 27{,}924\sin^{10}\gamma\cos^4\gamma - \\ &- 27{,}092\sin^8\gamma\cos^6\gamma + \\ &+ 9{,}127\sin^6\gamma\cos^8\gamma - \\ &- 0{,}606\sin^4\gamma\cos^{10}\gamma + \\ &+ 0{,}012\sin^2\gamma\cos^{12}\gamma. \end{aligned}$$

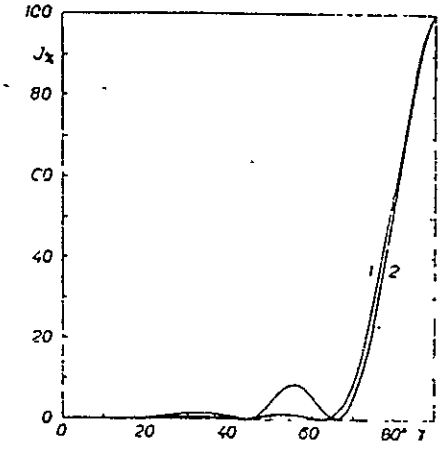


Figure 3. The variation of intensity of a first-modification chain filter with seven plates, in an ideal assembly (the curve with the higher secondary maxima) and in the optimum assembly (the curve with the lower secondary maxima)

A comparison of the intensity variation of the "optimum" assembly of a chain birefringent filter with the "ideal" assembly shows a marked suppression of the secondary maxima (figure 3). It should also be mentioned that when the azimuth of the individual plates is changed, not only the intensity but also the position of the secondary maxima change.

Analysis of the optimum filter assembly shows that the concept of the optimum can be adapted as needed. For example, one may wish to find the sharpest main maximum without considering the secondary maxima. Most often, however, it is important to suppress the secondary maxima. Then the relationship of the azimuths based on arithmetic progression is of use, and this applies not only to filters with a small number of plates but also to filters with tens or hundreds of plates.

THE SECOND MODIFICATION OF A SYMMETRICAL CHAIN FILTER IN THE BASIC POSITION, OPTIMUM ASSEMBLY [OF PLATES]

This example again yields similar results, so we will treat it but briefly. Here

$$\Phi = 0, \quad \varrho_n = 90^\circ - \varrho_1, \ \varrho_{n-1} = 90^\circ - \varrho_2, \ \ldots, \ \varrho_{n-i} = 90^\circ - \varrho_{i+1}.$$

When there is an odd number of plates, due to symmetry the middle plate will have an azimuth of 45°. The condition of maximum transmission need not be determined separately.

Formulas (4) specialize as follows:

$$
\begin{aligned}
&L_{1,1} = 0\,,\\
&L_{2,1} = 0\,,\\
&L_{3,2} = -\cos 4\varrho_1\,,\\
&L_{3,1} = 0\,,\\
&L_{3,2} = 2\sin 2\varrho_1 - \cos 4\varrho_1\,,\\
&L_{3,3} = 0\,,\\
&L_{4,1} = 0\,,\\
&L_{4,2} = -\cos 4\varrho_1 - \cos 4\varrho_2 - 2\cos(2\varrho_2 + 2\varrho_1) + 2\cos(2\varrho_2 - 2\varrho_1)\,,\\
&L_{4,3} = 0\,,\\
&L_{4,4} = \cos(4\varrho_1 - 4\varrho_2)\,,\\
&L_{5,1} = 0\,,\\
&L_{5,2} = 2\cos(2\varrho_1 - 2\varrho_2) - 2\cos(2\varrho_1 + 2\varrho_2) - \cos 4\varrho_1 - \cos 4\varrho_2 +\\
&\qquad + 2\sin 2\varrho_1 + 2\sin 2\varrho_2\,,\\
&L_{5,3} = 0\,,\\
&L_{5,4} = \cos(4\varrho_1 - 4\varrho_2) + 2\sin 2\varrho_1 (4\varrho_1 - 2\varrho_2) - 2\sin 2\varrho_1\,,\\
&L_{5,5} = 0\,,
\end{aligned}
$$

etc.

According to equations (7b), one- and two-plate filters do not have secondary maxima. Again the elimination of a secondary maximum in three-plate filters is formulated by the condition $L_{3,2} = 0$, or

$$2\sin 2\varrho_1 - \cos 4\varrho_1 = 0\,.$$

This equation has the same solution as in the case of the first filter modification; only angle ρ_2 need be computed with respect to symmetry, hence it is different:

$$\varrho_1 = 10{,}7^\circ\,, \quad \varrho_2 = 79{,}3^\circ\,.$$

The curve of the transmission [of light] is the same as for the first filter modification, except that it is shifted by $\pi/2$.

Once again, the solution is similar for a four-plate, second-modification filter without secondary maxima. Conditions $L_{4,2} = 0$ and $L_{4,4} = 0$ must be fulfilled. Then the following equations must be solved:

$$
\begin{gathered}
-\cos 4\varrho_1 - \cos 4\varrho_2 - 2\cos(2\varrho_2 + 2\varrho_1) + 2\cos(2\varrho_2 - 2\varrho_1) = 0\,,\\
4\varrho_1 - 4\varrho_2 = \pm 90^\circ\,.
\end{gathered}
$$

This system leads to equation

$$-\cos 4\varrho_1(1 + 2\cos 45^\circ) + \sin 4\varrho_1(1 + 2\sin 45^\circ) + 2\cos 45^\circ = 0\,,$$

which is equivalent to the equations derived for the first-modification filter. The solution gives the values

$$\varrho_1 = 7{,}6^\circ, \quad \varrho_3 = 59{,}9^\circ,$$
$$\varrho_2 = 30{,}1^\circ, \quad \varrho_4 = 82{,}4^\circ.$$

Again, this filter does not have secondary maxima. Solutions can be reached for multi-plate filters, but greater difficulties are involved; however, what has been said about the first modification also applies to the second modification, so let us be satisfied with the analogy of the arithmetic increase of azimuths applicable to the first modification. In the second modification, the increase of the azimuths from the edge toward the center of the filter can be expressed as follows:

$$\varrho_1, \quad \varrho_2 = 3\varrho_1 + \alpha, \quad \varrho_3 = 5\varrho_1 + (1+3)\alpha, \quad \varrho_4 = 7\varrho_1 + (1+3+5)\alpha,$$
$$\varrho_5 = 9\varrho_1 + (1+3+5+7)\alpha \ \ldots \ \varrho_{n-2} = [2(n-2)-1]\varrho_1 + (1+3+5+7+\ldots+7+5)\alpha,$$
$$\varrho_{n-1} = [2(n-1)-1]\varrho_1 + (1+3+5+7+\ldots+7+5+3)\alpha,$$
$$\varrho_n = (2n-1)\varrho_1 + (1+3+5+7+\ldots+7+5+3+1)\alpha.$$

In selecting the azimuths, ϱ_1 and ϑ can be defined as in the case of the first modification. Thus, the seven-plate first-modification filter of which we spoke in the preceding section has the same azimuths ($\varrho_1 = 4.5^\circ$, $\vartheta = 1.5^\circ$) for the second modification:

$$4.5^\circ, \ 15^\circ, \ 28.5^\circ, \ 45^\circ, \ 61.5^\circ, \ 75^\circ, \ 85.5^\circ.$$

Further analysis of this filter shows that the transmission curve is shifted by $\pi/2$, as in the case of the first filter modification.

CHAIN FILTERS IN THE GENERAL POSITION

Up to this point, all types of birefringent filters have been calculated and constructed for the basic position of the polarizing elements only, whether for parallel or crossed polaroids. However, these two fundamental positions are not a necessary condition for good filter function. Therefore, at this point we will indicate (albeit briefly) a method of solving for chain filters in a general position.

To explain the fundamental relationships, let us begin with one plate. For intensity, according to (7):

$$J_1 = \sin^2\gamma\cos^2(\Phi - 2\varrho_1) + \cos^2\gamma\cos^2\Phi.$$

Angle Φ is arbitrary in the general position. The first step in the analysis

is to find angles ϱ_1 that will correspond to the first and second modification of the filter. The first modification shows secondary maxima for $\gamma = \pi/2$. Then,

$$J_{max} = \cos^2(\Phi - 2\varrho_1),$$

or

$$\varrho_1 = \frac{\Phi}{2} = \Omega_1. \tag{11}$$

Here, too, the maxima reach full values. The minima have the intensity

$$J_{min} = \cos^2 \Phi.$$

The minima are not pure; their intensities increase as the distance of angles Φ from $\pi/2$. The minima vanish when $\Phi = 0^0$.

The second filter modification has its maxima at $\gamma = 0$, its minima at $\gamma = \pi/2$. For one plate,

$$J_{min} = \cos^2(\Phi - 2\varrho_1),$$

or

$$\Phi - 2\varrho_1 = \frac{\pi}{2},$$

$$\varrho_1 = \frac{\pi}{4} - \frac{\Phi}{2} = \Omega_2. \tag{12}$$

The maxima reach values of

$$J_{max} = \cos^2\Phi$$

Therefore, the analogy of the second modification has pure minima; however, the maxima do not reach full values and decrease as angle Φ increases. The maxima vanish when $\Phi = \pi/2$.

This analysis indicates that two distinct azimuths of a birefringent plate, defined by equations (11) and (12), can also be established for a common angle Φ. In the case of equation (11), the plate behaves analogously with the first filter modification; the maxima and minima have the same wavelengths as when $\Phi = \pi/2$. The intensity variation is then given by the relationship

$$J = \sin^2\gamma + \cos^2\gamma\cos^2\Phi.$$

When the azimuth of the plate agrees with equation (12), the result is analogous to the second filter modification and the intensity variation is given by the equation

$$J = \cos^2\gamma\cos^2\Phi.$$

If there are more plates, the general equations can be used in the analysis. For the first filter modification we get the following result: the azi-

muths of the individual plates deviate alternately to the left and right of the mean angle Γ,

$$\Gamma = -\Omega_2 .$$

The following relationship holds for the azimuth angles measured from the mean angle Γ,

$$|\bar{\varrho}_1| + |\bar{\varrho}_2| + |\bar{\varrho}_3| + \ldots - \bar{\varrho}_n| = \Omega_1 . \qquad (13)$$

In filters with a large number of plates, there are also clear minima for angles Φ which differ considerably from $\pi/2$. By selection of angles $\bar{\rho}_i$, a filter with the ideal assembly of plates and one with the optimum assembly can be constructed in a manner completely analogous to that used for the basic position, except that equation (13) will apply instead of equation (8).

In the case of the second filter modification, again we arrange the azimuths of the plates in the shape of a fan. The center of the fan must agree with angle Ω_2, according to equation (12). The azimuth difference of the first and last plate is determined by the equation

$$\varrho_n - \varrho_1 \gtrless \Omega_2 . \qquad (14)$$

One can construct a filter with either the ideal or the optimum assembly, depending on the arrangement of azimuths. When the number of plates is increased, the intensity in the maxima also increases and approaches full value. The special property of filters that operate at a general angle of the polaroids and analyzers is that they offer the possibility of tuning filters in the basic position. Thus, through slight departures from the parallel or crossed positions of the polaroids and analyzer, the filter can be tuned still more. This generalization of the chain filters is also important from the theoretical standpoint.

SUMMARY

In their application, chain birefringent filters fall into the same category as multilayered interference filters prepared by steaming in vacuum. The most selective multilayered interference filters have a transmission band of about 20 Å in the visible region and it can be assumed that, for technical reasons, the selectivity of these filters will not be increased. Birefringent filters are suitable for a range of band widths of about 50 Å to 0.5 Å. Therefore it is desirable to develop birefringent filters with high trans-

mission and, simultaneously, with the purest possible spectral characteristic. Apropos, the numerical analysis performed here shows a method for producing filters with high transmission and nearly pure maxima without secondary transmissions.

Submitted 11 May 1959

LITERATURE CITED

1. Šolc, I. Československý Časopis pro Fysiku, 3: 366, 1953.
2. Šolc, I. Československý Časopis pro Fysiku, 4: 669, 1954.
3. Šolc, I. Československý Časopis pro Fysiku, 4: 607, 1954.
4. Šolc, I. Československý Časopis pro Fysiku, 5: 114, 1955.
5. Hsü, Hsien-Yü, Yüng-Kang Liang, and M. Richartz. Optical Society of America, Journal, 37: 99, 1947.
6. Šolc, I. Czechoslovak Journal of Physics, 9: 237, 1959.
7. Dollfus, A. Revue d'Optique, 35: 539, 1956.
8. Dollfus, A. Revue d'Optique, 35: 625, 1956.
9. Rozenberg, G. V. Uspekhi Fizicheskikh Nauk, 17: 3, 1952.
10. Rozenberg, G. V. Uspekhi Fizicheskikh Nauk, 17: 173, 1952.
11. Mie, K. Optik, 14: 1, 1957.

Journal of the
OPTICAL SOCIETY
of AMERICA

VOLUME 55, NUMBER 6 JUNE 1965

Birefringent Chain Filters

IVAN ŠOLC
Dioptra n.p., Turnov-Námesti, CSSR, Czechoslovakia
(Received 20 August 1964)

Birefringent chain filters were first described 10 years ago. Since then such filters have undergone considerable development, and experience has been gained with their use. This article includes only a brief review of previous work, which is supplemented with new information.

FUNDAMENTAL STATEMENTS

BIREFRINGENT chain filters[1] are composed of a pile of birefringent elements which have the same or periodically repeating phase difference. The azimuths of the individual elements are arranged in prescribed manners, mostly symmetrically with respect to the center of the filter. The filter is placed between two polarizers. The theory of chain birefringent filters was thoroughly described in Ref. 2. The simplest filter is composed of a series of equal birefringent plates, cut parallel to the optic axis of the crystals. The azimuth ρ of the plane of polarization of each plate is measured from the plane of polarization of the first polarizer. The total number of plates is designated N.

There are two modifications of birefringent chain filters. The first modification works between crossed polarizers. The azimuths of the individual plates are:

Element	Azimuth
Entrance polarizer	$P_1=0°$
Plate 1	$\rho_1=\rho$
Plate 2	$\rho_2=-\rho$
Plate 3	$\rho_3=\rho$
Plate 4	$\rho_4=-\rho$
⋮	⋮
Plate i	$\rho_i=-(-1)^i\rho$
Exit polarizer	$P_2=90°$

The angle ρ is:

$$\rho=45°/N. \tag{1}$$

If each plate has thickness d and double refraction $n_e-n_o=D$, its path difference Δ is given by the product:

$$\Delta=d\cdot D. \tag{2}$$

The phase difference ϑ or its half-value γ are connected with the path difference Δ and the wavelength λ by:

$$\vartheta=2\gamma=(2\pi/\lambda)\cdot\Delta. \tag{3}$$

The wavelengths of maximum transmittance of a chain filter of type I are:

$$\Delta=K\cdot\lambda, \tag{4}$$

where $K=0.5,\ 1.5,\ 2.5\cdots$. The detailed course of the transmittance of the chain birefringent filter was calculated with a matrix method by Evans.[3] From his analysis follow all the properties of chain filters, including comparisons of the transmittances of secondary maxima with those of Lyot's filter, in which the transmittances of the secondary maxima are considerably greater. Similar calculations were first made in Ref. 2, based on the theory originally given in Ref. 4. Here we give only the final equations for the transmitted intensity relative to the intensity emerging from the first polarizer, for filters composed of one, two, three, four, or five plates:

$$\begin{aligned}
I_1&=\sin^2\gamma\\
I_2&=\sin^4\gamma\\
I_3&=\sin^6\gamma-\sin^4\gamma\cos^2\gamma+0.25\sin^2\gamma\cos^4\gamma\\
I_4&=\sin^8\gamma-2.818\sin^6\gamma\cos^2\gamma+2\sin^4\gamma\cos^4\gamma\\
I_5&=\sin^{10}\gamma-5.472\sin^8\gamma\cos^2\gamma+8.104\sin^6\gamma\cos^4\gamma\\
&\quad-1.691\sin^4\cos^6\gamma+0.095\sin^2\gamma\cos^8\gamma.
\end{aligned} \tag{5}$$

In Table I the coefficients of Eqs. (5) are given for filters containing as many as nine plates.

[1] I. Šolc, Česk. Časopis Fys. 3, 366 (1953).
[2] I. Šolc, Česk. Časopis Fys. 10, 16 (1960).
[3] J. W. Evans, J. Opt. Soc. Am. 48, 142 (1958).
[4] Hsu Hoien, Liang Yung-Kang, and M. Richartz, J. Opt. Soc. Am. 37, 99 (1947).

TABLE I. Coefficients of Eq. (6) for filters containing as many as 9 plates.

k \ Δ	1	2	3	4	5	6	7	8	9
1	1								
2	1	0							
3	1	−1	0.250						
4	1	−2.818	2.000	0					
5	1	−5.472	8.104	−1.691	0.0955				
6	1	−8.928	22.927	−13.392	2.250	0			
7	1	−12.750	49.477	−56.774	22.350	−1.962	+0.049		
8	1	−17.274	103.001	−181.38	123.179	−29.859	2.341	0	
9	1	−23.164	178.97	−410.79	411.124	−203.525	42.224	−2.102	0.030

The second type of chain filter works between parallel polarizers. The azimuths of the individual plates are:

Element	Azimuth
Entrance polarizer	$P_1=0°$
Plate 1	$\rho_1=\rho$
Plate 2	$\rho_2=3\rho$
Plate 3	$\rho_3=5\rho$
Plate 4	$\rho_4=7\rho$
⋮ ⋮	⋮ ⋮ ⋮
Plate i	$\rho_i=(2i-1)\rho$
Plate N	$\rho_n=(2N-1)\rho=90°-\rho$
Exit polarizer	$P_2=0°$

The value of ρ is again given by Eq. (1). The variation of intensity with phase difference (or wavelength) can be described by equations resembling those given for type I filters:

$$I_1=\cos^2\gamma$$
$$I_2=\cos^4\gamma$$
$$I_3=\cos^6\gamma-\cos^4\gamma\sin^2\gamma+0.25\cos^2\gamma\sin^4\gamma$$
$$I_4=\cos^8\gamma-2.818\cos^6\gamma\sin^2\gamma+2\cos^4\gamma\sin^4\gamma$$
$$I_5=\cos^{10}\gamma-5.472\cos^8\gamma\sin^2\gamma+8.104\cos^6\gamma\sin^4\gamma$$
$$-1.691\cos^4\gamma\sin^6\gamma+0.095\cos^2\gamma\sin^8\gamma. \quad (6)$$

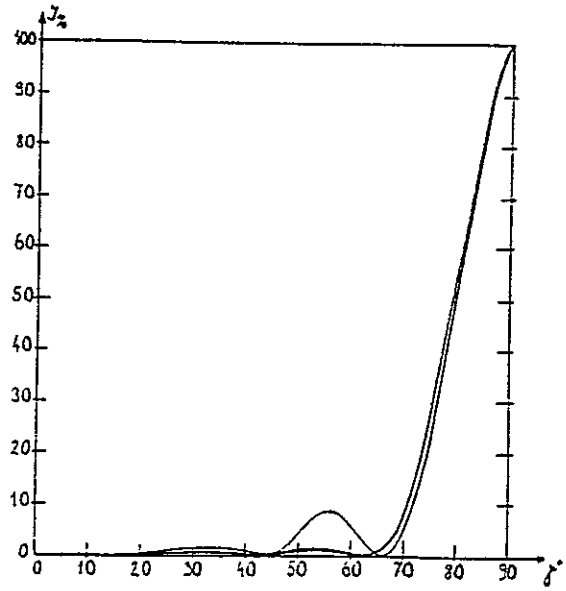


FIG. 1. Variation of intensity of 7-plate chain filter with $\Omega=31.5°$ compared with 7-plate chain filter with $\Omega=45°$ (narrower principal maximum, but greater neighboring maximum at $\gamma=57°$).

The values of the coefficients for filters containing as many as 9 plates are given in Table I.

If we compare Eqs. (5) and (6) we find that only the $\sin\gamma$ and $\cos\gamma$ functions are interchanged, which means that transmittance curves have the same shape; they are only mutually displaced 90°. The coefficients in the expression for the wavelengths of the principal maxima are integers:

$$\Delta=K\cdot\lambda, \quad (4a)$$

where $K=0, 1, 2, 3\cdots$. The disadvantages of chain filters of types I and II are comparatively strong adjacent secondary maxima, as Evans has shown.[3]

IMPROVEMENT OF CHAIN FILTERS

Types I and II of chain filters can be significantly improved by arranging the azimuths of the plates so that the adjacent secondary maxima are suppressed and have negligible values. Modified type-I chain filters with suppressed secondary maxima are constructed by making the azimuths of the end plates smaller than the azimuths of the central plates. In the simplest arrangement, the azimuths increase according to an arithmetic series, from the end towards the center of the filter. Theory (2) imposes the fundamental condition:

$$\sum_{i=1}^{N}|Pi|=45°. \quad (7)$$

If Eq. (7) is not fulfilled, the main maximum is lower than the attainable value. The azimuths of the individual plates are given by:

Element	Azimuth
Entrance polarizer	$P_1=0°$
Plate 1	$\rho_1=\rho$
Plate 2	$\rho_2=-(\rho+\alpha)$
Plate 3	$\rho_3=\rho+2\alpha$
Plate 4	$\rho_4=-(\rho+3\alpha)$
⋮ ⋮	⋮ ⋮ ⋮
Plate $N-1$	$\rho_{N-1}=\mp(\rho+\alpha)$
Plate N	$\rho_N=\mp\rho$
Exit polarizer	$P_2=90°$

Angle ρ is given by $\rho=\Omega/N$, where Ω is an angle smaller

than 45°. The value of Ω depends upon the number of plates and the desired characteristics of the filter. Smaller angles Ω cause better suppression of neighboring maxima, but the main maximum widens somewhat. Larger angles Ω give a sharper principal maximum but the neighboring maxima become more prominent.

Angle α must be such that Eq. (1) is fulfilled. For example, consider a filter with 7 plates. We choose the angle between minimum and central, for example: $\rho=31.5°/7=4.5°$. We use angle $\alpha=1.5°$. The azimuths of the filter plates are then:

$$\rho_1=4.5°, \quad \rho_2=-6.0°, \quad \rho_3=7.5°, \quad \rho_4=-9°,$$
$$\rho_5=7.5°, \quad \rho_6=-6°, \quad \rho_7=4.5°.$$

By using the theory given in Ref. 2 we derive the formula for the intensity variation of this filter:

$$I_7=\sin^{14}\gamma-9.428\sin^{12}\gamma\cos^2\gamma+27.924\sin^{10}\gamma\cos^4\gamma$$
$$-27.092\sin^8\gamma\cos^6\gamma+9.127\sin^6\gamma\cos^8\gamma$$
$$-0.012\sin^2\gamma\cos^{12}\gamma.$$

The main maximum of a filter of this type is given by Eq. (4). The variation of intensity of this filter is compared with that of the 7-plate filter with azimuths according to the simplest arrangement ($\Omega=45°$) in Fig. 1.

The suppression of the neighboring secondary maxima is evident in Fig. 1. Similar modifications of type II filters can be constructed so that neighboring maxima are effectively suppressed. By analogy with the previous case, the azimuths should be arranged as follows:

Element	Azimuth
Entrance polarizer	$P_1=0°$
Plate 1	$\rho_1=\rho$
Plate 2	$\rho_2=3\rho+\alpha$
Plate 3	$\rho_3=5\rho+(1+3)\alpha$
Plate 4	$\rho_4=7\rho+(1+3+5)\alpha$
⋮	⋮
Plate $N-2$	$\rho_{N-2}=[2(N-2)-1]\rho+(1+3+5+7\cdots+7+5+3+1)\alpha$
Plate $N-1$	$\rho_{N-1}=[2(N-1)-1]\rho+(1+3+5+7\cdots+7+5+3+1)\alpha$
Plate N	$\rho_N=(2N-1)\rho+(1+3+5+7\cdots+7+5+3+1)\alpha$
Exit polarizer	$\rho_2=0°$

The angle ρ is again determined by Eq. (8), the angle α is subject to condition (7). As an example of such a filter, consider again a filter with 7 plates:

$$\rho_1=4.5°, \quad \rho_2=15°, \quad \rho_3=28.5°, \quad \rho_4=45°,$$
$$\rho_5=61.5, \quad \rho_6=75°, \quad \rho_7=85.5°.$$

From Ref. 2 we may derive the variation of intensity of this filter:

$$I_7=\cos^{14}\gamma-9.428\cos^{12}\gamma\sin^2\gamma+27.924\cos^{10}\gamma\sin^4\gamma$$
$$-27.092\cos^8\gamma\sin^6\gamma+9.127\cos^6\gamma\sin^8\gamma$$
$$-0.012\cos^2\gamma\sin^{12}\gamma.$$

The transmittance of this filter resembles that shown in Fig. 1, except that it is displaced $\frac{1}{2}\pi$.

Chain filters can also be constructed for azimuths of the two polarizers other than 0° or 90°. Plates also can be used with different path differences. These examples are briefly analyzed in Refs. 1, 2, and 5. The width of the passband $\Delta\lambda$ depends on the thickness of the individual plates and on their number. The high transmission of chain filters in comparison with Lyot filters is a consequence of the use of only two polarizers, which are all that are needed for chain filters.

Numbers of elements Lyot filter	Numbers of elements of chain filters with $\Omega=45°$	Numbers of elements of chain filters with $\Omega=30°$
1 plate+2 polarizers	1 plate+2 polarizers	1 plate+2 polarizers
2 : +3	2 : +2	2 : +2
3 : +4	4 : +2	5 : +2
4 : +5	8 : +2	10 : +2
5 : +6	16 : +2	20 : +2
6 : +7	32 : +2	40 : +2
7 : +8	64 : +2	80 : +2
8 : +9	128 : +2	160 : +2

Theory indicates that the width of the spectrum band transmitted by a chain filter can be approximated by the simple formula:

$$\Delta\lambda\doteq(\lambda/2K\cdot N)\cdot H. \qquad (9)$$

Here K is the same as in Eq. (4). N is the number of plates and H is a correction factor corresponding to the widening of the main maximum, accompanying sup-

[5] I. Šolc, Czech. J. Phys. 9, 237 (1959).

pression of neighboring maxima. The values of H are usually between 1 and 1.5.

CHANGE OF WAVELENGTH BY INCLINATION OF FILTER

For double-refraction filters, plates cut parallel with optical axis are mostly used. (Other cuts are not suitable because the rays inside the plates would deviate from the original direction, which would interfere with the proper function of the filter.) We can derive the dependence of passband wavelength on inclination by consideration of a single plate. The resulting relations are valid for both modifications of chain filters.

(a) The best analysis of wavelength variation was given by Lyot.[6] Approximate theory[5] leads to simple results which are sufficient for the majority of practical calculations. With uniaxial crystals such as quartz, calcite, or ADP, rotation of the plate around the optical axis displaces the wavelength of the maximum toward the red end of spectrum. If the angle of incidence is ω, and if the average index of refraction of the crystal is n, then the wavelength λ_ω transmitted best is:

$$\lambda_{\varphi\parallel}=(d\cdot D/K)\cdot[n/(n^2-\sin^2\omega)^{\frac{1}{2}}]. \qquad (10)$$

If we rotate the plate around a line perpendicular to the optical axis, the maximum is displaced toward the violet end of spectrum, approximately as given by the equation:

$$\lambda_{\varphi\perp}=(d\cdot D/K)\cdot[(n^2-\sin^2\varphi)^{\frac{1}{2}}/n]. \qquad (10a)$$

If we denote by λ_0 the wavelength of maximum transmittance for perpendicular incidence, then approximately:

$$\lambda_{\varphi\parallel}=\lambda_0\cdot[n/(n^2-\sin^2\omega)^{\frac{1}{2}}], \qquad (11)$$

$$\lambda_{\varphi\perp}=\lambda_0\cdot[(n^2-\sin^2\omega)^{\frac{1}{2}}/n]. \qquad (11a)$$

Curves based on Eqs. (11) are shown in Fig. 2. These results apply to type I and to type II chain filters.

(b) The formulas given permit the determination of the transmittances of chain filters for planes of incidence in different azimuths. Therefore, they can be used for analysis of convergent incidence. For small angles we can develop Eq. (10) in series and replace $\sin\varphi$ by the angle φ:

$$\lambda_\varphi\doteq\lambda_0[1-(\varphi^2/2n^2)].$$

We need not consider the second equation, because the curves are symmetrical in the neighborhood of $\omega=0°$. Let us assume further that the change of wavelength of maximum may be at most $1/D$ the width of the passband. We may then write Eq. (9):

$$(\lambda_0/20K\cdot N)\cdot H=\lambda_0\cdot(\omega^2/2n^2).$$

[5] B. Lyot. Ann. Astrophys. 7, 31 (1944).

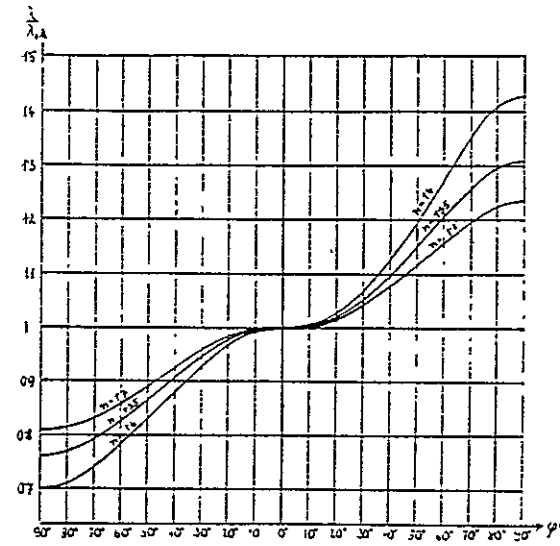


FIG. 2. Tuning of chain filter by inclination. Left: inclination perpendicular to optical axis. Right: inclination parallel to optical axis.

Hence it follows for admissible angles of convergence that:

$$\omega\doteq n\cdot[H/(10K\cdot N)]^{\frac{1}{2}}. \qquad (12)$$

For less severe requirements, when simple chain filters are used, even greater convergence is permissible than indicated by Eq. (12). That is because the interference pattern of the whole chain filter is almost identical with the interference pattern of one plate from which the filter is assembled.

For extraordinary requirements, it is possible to increase the admissible convergence of rays on a chain filter, for example by replacing simple plates with Lyot or Evans split elements or with compound plates consisting of positive and negative crystals.

PRACTICAL USE OF CHAIN FILTERS

(a) Most often it is possible to work with simple plates in type I or II arrangement usual with suppressed secondary maxima. These filters are made with as many as 80 plates, usually quartz, from 0.1 to 15 mm thick, exceptionally with ADP crystals from 1 to 10 mm thick, or with calcite plates 0.3 to 5 mm thick. The desired maximum is often isolated by use of a dielectric interference filter; or two chain filters are combined which fulfill the condition for the desired maximum λ at different orders K. There are two possibilities for changing wavelength by inclination of combined chain filters. For the first the azimuths of both filters are identical. For rays incident at any angle to the axis the wavelength maxima of both filters are shifted by the same amount to the same side of the spectrum. Therefore, both filters can be tuned by common inclination, but convergent rays distort the main maximum.

In the second case, the azimuths of the two filters are

mutually perpendicular. Contrary to the previous case, rays incident at the same angle to the axis shift the wavelengths of the maxima of the two filters in opposite directions in the spectrum. In this case, the main maximum is not distorted by convergent rays, but the transmittance for convergent rays is lower than for the parallel rays. Filters with narrow passbands must be thermostatted.

(b) If large convergence is required with narrow bandpass, there are two possibilities: either a chain filter made from wide-field elements of Lyot or Evans construction, which are technologically very difficult, or a chain filter is combined with one or two Lyot or Evans[7] split elements. The first alternative has the advantage of less weight, but it cannot be tuned by use of a quarter-wave plate (line-shifter); the second alternative is easy to tune. Very much preferred, for instance, is a combination consisting essentially of from one up to three chain filters of quartz, which are placed in a thermostat. At both ends of this filter are calcite split elements with tuning quarter-wave plates. This system is easily tunable over the usual range of wavelengths; at the same time it is possible to suppress quite completely all of neighboring maxima. It is also very easy to isolate the wanted wavelength from all other maxima in a wide spectral range.

(c) With the help of the inclined arrangement, it is possible to construct a filter that can be tuned over a very wide range of wavelengths, but this is usable only for almost parallel rays. Combination of two such filters enables us to choose any arbitrary wavelength without need for dielectric isolating filters.

In special cases, a chain filter can be made of plates cut at an angle to the optical axis of the crystal. This arrangement is convenient only for filters consisting of a few thin plates.

(d) A tuned chain filter made of a pile of wedged plates has a small aperture (e.g., 2 to 10 mm); the convergence of the incident beam can be several degrees. Such a filter is very suitable, for example, for interferometric measurement where by pushing the wedge the chosen wavelength is easily isolated from a line-spectrum source. With a combination of two wedge filters it is again possible to put together a continuously tunable monochromator with resolving power from about 100 up to 300. Such arrangement is usual, for example, for colorimetry or for microscope illumination with a selected wavelength.

Similar arrangements, in which wedged plates are used with supplementary wedged plates, were suggested by Evans[8] for isolation of high orders in grating spectra. Also, it is interesting to note that the "achromatic half-wave plates" described in Ref. 9, are, in fact, identical to the type II chain-birefringent filter with fewer plates.

The described arrangements of chain filters do not exhaust all possibilities and combinations.

[7] J. W. Evans, J. Opt. Soc. Am. 39, 229 (1949).

[8] J. W. Evans, Appl. Opt. 2, 193 (1963).
[9] C. J. Koester, J. Opt. Soc. Am. 49, 405 (1959).

Solc Birefringent Filter

JOHN W. EVANS
Sacramento Peak Observatory, Geophysics Research Directorate, AFCRC, Sunspot, New Mexico
(Received August 26, 1957)

A new birefringent filter composed of a series of retardation plates between a single pair of polarizers has been described by Solc. The filter is analyzed and compared with the Lyot filter. It is found that the Solc filter is inferior to the Lyot filter in the suppression of parasitic light in secondary maxima near the primary transmission bands, although its band width is slightly less, and its transparency is very much better if both filters use Polaroid film for polarizers

BIREFRINGENT filters of the type invented by Lyot, with transmission bands ranging from 0.5 to 5 A in width have become standard tools of solar research. While they are not excessively difficult to construct, and their performance is excellent, any possible simplification or improvement would be welcomed. A new form of birefringent filter recently developed in Czechoslovakia is therefore worth examination, and comparison with the Lyot filter.

In 1953 to 1955 Ivan Solc[1] published three papers on a new form of birefringent filter which he apparently developed and investigated purely by experiment. He does not give a general expression for the transmission of his filter as a function of wavelength because of its mathematical complexity. He did, however, succeed in determining its more important characteristics experimentally with considerable accuracy. The purpose of this paper is to derive an expression for the transmission of the Solc filter, and to compare its performance with that of the Lyot filter.

The Solc filter consists of a pile of identical retardation plates of birefringent material, with only two linear polarizers, one at each end. Figure 1 shows the arrangement. The plates are cut, as in the Lyot filter, with the crystal optic axis parallel to the surfaces. Solc describes two possible arrangements of the orientations of the axes of successive plates, which we shall term the fan and folded filters, respectively.

Let the electric vector of light traversing the first polarizer be the reference direction from which the orientation, ω_j, of the optic axis of the jth plate is measured. In the fan filter the two polarizers are parallel,

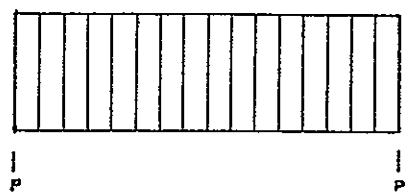


FIG. 1. Section of a Solc filter, consisting of 16 retardation plates between a pair of polarizers, P.

and ω progresses monotonically. The angle ω_j is given by

$$\omega_j=(\alpha/2)+(j-1)\alpha, \quad (1)$$

where α is a small angle which we shall determine.

In the folded filter the polarizers are crossed, and ω alternates between $(\alpha/2)$ and $-(\alpha/2)$. The angle ω_j is given by

$$\omega_j=(-1)^{j+1}(\alpha/2). \quad (2)$$

Solc determined experimentally that both forms of the filter appeared to work best when $\alpha\approx\pi/2n$, where $n=$ the total number of plates. He states further that the curve of transmission as a function of wavelength closely resembles that of a Lyot filter in which the thinnest element (which fixes the spacing of the bands) has the same thickness as one of his retardation plates,

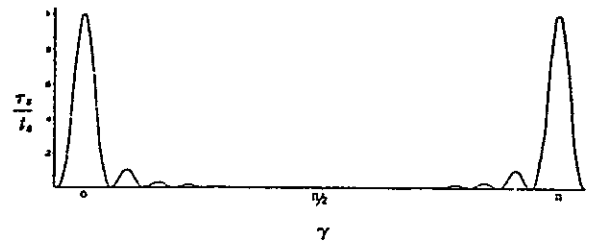


FIG. 2. Transmission curve of a Solc filter of 16 plates as a function of the retardation, γ, of a single plate.

and the thickest element (which determines the band width) has the thickness of the whole pile of Solc plates.

It is difficult to see intuitively how either form of the Solc filter accomplishes its purpose. The problem of deriving an expression for the transmission by the usual methods does not appear to have an easy solution. However, the use of the matrix calculus developed by Jones[2] proves to be effective. The reader is referred to his papers for a very clear explanation of the method. In the following application, we retain his notation for convenience of reference.

Let the light travel along the z axis of a rectangular coordinate system. The faces of the n retardation plates of a Solc filter are normal to the z axis, and the electric vector of light transmitted by the first polarizer is parallel to the x axis. Let the electric vectors of light

[1] I. Solc, Czechoslov. Cosopis pro Fysiku, 3, 366 (1953); 4, 607, 669 (1954); 5, 114 (1955).

[2] R. C. Jones, J. Opt. Soc. Am. 31, 488 (1941); 31, 500 (1941).

entering and emerging from the system be represented by the one column matrices,

$$\epsilon_0=\begin{pmatrix}E_{x0}\\E_{y0}\end{pmatrix} \quad \text{and} \quad \epsilon_n=\begin{pmatrix}E_{xn}\\E_{yn}\end{pmatrix}. \tag{3}$$

Then the initial matrix equations relating ϵ_0 and ϵ_n in the fan and folded filters are,

$$\begin{pmatrix}E_{xn}\\E_{yn}\end{pmatrix}_{\text{Fan}}=\mathrm{P}_x\mathrm{S}(\alpha/2)\mathrm{S}(\pi/2)[\mathrm{S}(-\alpha)\mathrm{G}]^n \times\mathrm{S}(-\alpha)\mathrm{P}_x\begin{pmatrix}E_{x0}\\E_{y0}\end{pmatrix}, \tag{4}$$

$$\begin{pmatrix}E_{xn}\\E_{yn}\end{pmatrix}_{\substack{\text{Folded}\\ n\text{ even}}}=(-1)^{(n/2)}\mathrm{P}_y\mathrm{S}(\alpha/2)[\mathrm{S}(-\alpha)\mathrm{GS}(\alpha)\mathrm{G}]^{(n/2)} \times\mathrm{S}(-\alpha)\mathrm{P}_x\begin{pmatrix}E_{x0}\\E_{y0}\end{pmatrix}, \tag{5}$$

$$\begin{pmatrix}E_{xn}\\E_{yn}\end{pmatrix}_{\substack{\text{Folded}\\ n\text{ odd}}}=i(-1)^{(n-1)/2}\mathrm{P}_y\mathrm{S}(\alpha/2) \times\mathrm{G}[\mathrm{S}(-\alpha)\mathrm{GS}(\alpha)\mathrm{G}]^{(n-1)/2} \times\mathrm{S}(-\alpha)\mathrm{P}_x\begin{pmatrix}E_{x0}\\E_{y0}\end{pmatrix}. \tag{6}$$

Here $\mathrm{S}(\alpha)$ and G are matrices representing rotation through the angle α, and the retardation of a single plate, respectively. They are defined as follows:

$$\mathrm{S}(\alpha)=\begin{pmatrix}\cos\alpha & -\sin\alpha\\ \sin\alpha & \cos\alpha\end{pmatrix}, \tag{7}$$

and

$$\mathrm{G}=\begin{pmatrix}e^{-i\gamma} & 0\\ 0 & e^{-i\gamma}\end{pmatrix}, \tag{8}$$

where $\gamma=\pi(d/\lambda)(\epsilon-\omega)$ in the fan filter, $\gamma=\pi(d/\lambda)(\epsilon-\omega)+(\pi/2)$ in the folded filter, d = thickness of a single retardation plate, λ = wavelength of the light, and ϵ and ω = refractive indexes of the birefringent material.

The matrices P_x and P_y represent polarizers transmitting the electric vector parallel to the x and y axes, respectively, and are defined by

$$\mathrm{P}_x=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} \quad \text{and} \quad \mathrm{P}_y=\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}. \tag{9}$$

The transmission of the filter, τ_S, is the square of the ratio of the amplitudes of the entering and emerging electric vectors, ϵ_0 and ϵ_n. The problem of deriving a useful expression for τ_S is basically the problem of raising the factors in square brackets in Eqs. (4), (5) and (6) to the indicated powers without multiplying them out explicitly, an impossibly laborious task when n is large. These factors are two by two matrices. Jones has devised an ingenious method for raising such matrices to any desired power. His method leads to a single

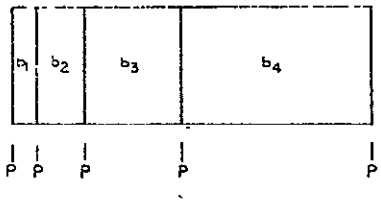


FIG. 3. Section of a Lyot filter of four birefringent elements, b_1–b_4, sandwiched between polarizers, P.

expression for the transmission of initially unpolarized light through the fan and the even and odd folded filters as follows,

$$\tau_S=\frac{t_S}{2}\left[\frac{\sin n\chi}{\sin\chi}\cos\chi\tan\alpha\right]^2. \tag{10}$$

The parameter χ is related to the retardation, γ, by the following expression:

$$\cos\chi=\cos\gamma\cos\alpha. \tag{11}$$

The factor t_S represents absorption and reflection losses in the filter.

Examination of Eq. (10) shows that Solc was correct in his experimental finding that $\alpha=(1/n)(\pi/2)$ is an optimum condition. Adopting this value for α, we obtain the transmission curve of Fig. 2, and derive the following conclusions:

(a) At the principal transmission bands, $\tau_S=(t_S/2)$, when $\chi=\alpha$ or $\chi=\pi-\alpha$. The corresponding values of the retardation are $\gamma=k\pi$, where k is any integer.

(b) $\tau_S=0$ when $\chi=(l/n)\pi=2l\alpha$, where $l=1, 2, 3\cdots(n-1)$.

(c) Secondary maxima occur between the zeros. Their exact positions are not readily determined, although they are very near the midpoints between successive zeros. At these points, $\chi=(2l+1)\alpha$, where $l=1, 2, 3, \cdots(n-2)$. The transmission at a midpoint between zeros is, then,

$$\tau_{Sm}=\frac{t_S}{2}\left[\frac{\sin n(2l+1)\alpha}{\sin(2l+1)\alpha}\cos(2l+1)\alpha\tan\alpha\right]^2. \tag{12}$$

For purposes of comparison, the reader will find the theory of the Lyot filter and its split-element modification in papers by Lyot,[3] Evans,[4] and Dollfus.[5] To review briefly, the Lyot filter consists of a multiple sandwich of birefringent crystal layers and polarizers. The crystal layers, termed b-elements, are cut with the crystal optic axis parallel to the surfaces, perpendicular to the instrumental optical axis. In the simplest form, the thicknesses of the b-elements form a series in powers of 2. Polarizers, usually sheets of Polaroid film, are placed between successive b-elements and at each end, with their axes parallel. The arrangement is shown in Fig. 3. The b-elements are also oriented with their axes parallel, at an angle of 45° to the electric vector trans-

[3] B. Lyot, Ann. astrophys. 7(1), 2 (1944).
[4] J. W. Evans, J. Opt. Soc. Am. 39, 229 (1949).
[5] A. Dollfus, Rev. opt. 35, 625 (1956).

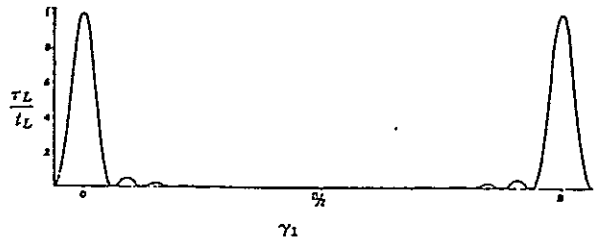


FIG. 4. Transmission curve of a Lyot filter of four elements as a function of the retardation, γ_1, of the thinnest element.

mitted by the polarizers. The transmission, τ_L, of the assembly is given by the equation,

$$\tau_L=(t_L/2)[\cos\gamma_1\cos 2\gamma_1\cos 4\gamma_1\cdots\cos 2^{N-1}\gamma_1]^2, \quad (13)$$

where N=number of b-elements, γ_1=retardation of the thinnest element$=\pi(d_1/\lambda)(\epsilon-\omega)$, d_1=thickness of the thinnest element, and t_L=a factor representing absorption and interface reflection losses in the filter.

Lyot[3] shows that Eq. (13) can be written as follows

$$\tau_L=\frac{t_L}{2}\left[\frac{\sin 2^N\gamma_1}{2^N\sin\gamma_1}\right]^2. \quad (14)$$

He points out that this is identical to the expression for the intensity curve of a diffraction grating of 2^N rulings at fixed angles of incidence and diffraction. The curve, shown in Fig. 4, consists of widely spaced sharp principal maxima or transmission bands. Between successive bands are 2^N-1 zeros interspersed with 2^N-2 secondary maxima. The separation of successive bands is inversely proportional to γ_1, and the band width to $2^{N-1}\gamma_1$. These are the retardations in the thinnest and thickest b-elements, respectively.

Lyot filters, in both the simple and split element forms, made of quartz and calcite, less than 20 cm long, are now widely used for observation of the sun in the light of the Hα line of hydrogen and the 5303 line of Fe XIV in the corona. The band widths range from 0.5 to 5 A.

In assessing the practical value of a birefringent filter, the relevant characteristics are the transmission at the centers of the transmission bands; the band width between the first zeros on each side of the band; the transmission of parasitic light in the secondary maxima, particularly those near the primary bands; and finally, the cost of the filter. It is therefore of interest to compare the Solc and Lyot filters in these particulars. For convenience, the curves of τ_S and τ_L are superposed in Fig. 5.

A comparison of Eqs. (10) and (14) suggests that a Lyot and Solc filter are equivalent if $\gamma_S=\gamma_{1L}$ and $n=2^N$. Then the spacing of the transmission bands in γ (or λ), and the number of zeros and secondary maxima between bands is the same for the two. We therefore compare a Solc filter composed of 2^N identical retardation plates with a Lyot filter of N b-elements, the thinnest of which is identical with one of the Solc retardation plates. The total thickness of birefringent material in the Solc filter exceeds that in the Lyot filter by the thickness of one retardation plate. This is not in accord with Solc's experimental finding that the two are equivalent when the thickness of his filter is approximately half that of a Lyot filter, an item of some economic importance.

In transmission, the Solc filter is definitely superior to the Lyot filter as usually constructed. The polarizers are usually Polaroid film, which has a transmission in the visible spectrum of about 0.75 for light polarized parallel to the transmission direction. Thus in the Solc filter $t_S=0.56$, while in the Lyot filter $t_L=(0.75)^{N+1}$. The split element form of the Lyot filter is considerably better, with $t_L=(0.75)^{(N/2)+1}$. In most practical filters, with transmission band widths between 0.5 and 5 angstroms, N ranges from 6 to 10, and t_L is between 0.14 and 0.05 in the simple Lyot filter, or 0.32 and 0.16 in the split element filter. These losses in the Lyot filter can be avoided, however, by the use of more transparent (and much more expensive) polarizers like Rochon prisms. A Lyot filter of this construction is very nearly as transparent as the equivalent Solc filter, but with the practical disadvantage of some added optical length which vignettes the field unless the aperture is enlarged.

The band widths of the two filters may be most readily compared in terms of $\Delta\gamma$, the increment in retardation of a single Solc plate (or the thinnest Lyot b-element) between the center of a transmission band and the first zero on either side. For the Lyot filter, $\Delta\gamma_L=(\pi/2N)$. For the Solc filter, $\Delta\chi=2\alpha$. If n is large (≥ 16, say), α is small, and $\Delta\gamma_S$ approaches $(\sqrt{3}/2)(\pi/n)$. Since $n=2^N$, we conclude that the Solc filter has the sharper transmission band by a factor of $(\sqrt{3}/2)=0.87$.

The transmitted parasitic light in the secondary maxima is approximately proportional to the transmission at the midpoint between zeros. This is given in Eq. (12) for the Solc filter. For the Lyot filter,

$$\tau_{Lm}=\frac{t_L}{2}\left[\frac{\sin(2l+1)\frac{\pi}{2}}{2^N\sin\frac{\pi}{2}(2l+1)/2^N}\right]^2. \quad (15)$$

We are interested in the ratio $(\tau_{Sm}/t_S)/(\tau_{Lm}/t_L)=\tau_{Sm}'/\tau_{Lm}'$ at the midpoints between corresponding zeros in the two filters (i.e., for the same values of l). This is

$$\frac{\tau_{Sm}'}{\tau_{Lm}'}=\left[2^N\cos\left(\frac{2l+1}{2^N}\right)(\pi/2)\tan\frac{1}{2^N}\frac{\pi}{2}\right]^2. \quad (16)$$

If n is large, the secondaries near the transmission

bands (l small) approach the ratio

$$(\tau_{Sm}'/\tau_{Lm}') = (\pi/2)^2 = 2.47. \quad (17)$$

Thus the parasitic light in the immediate neighborhood of the primary transmission bands is roughly 2.5 times as great in the Solc filter as in the Lyot filter. Since the first secondaries on either side of the primary band of a Lyot filter transmit together about 0.05 as much light as the band itself, it is evident that the parasitic light of the Solc filter may be very serious in certain applications. A textbook example would be the observation of solar flares in the light of the Hα line of hydrogen. Here we must isolate the light at the center of an absorption line roughly one angstrom broad. The light intensity at the center of the line is 0.16. that of the neighboring continuum. Thus a Lyot filter with an interval of 0.5 A between the center of the band and the first zeros on either side has its first secondaries well out in the continuum, and the diluting parasitic light from these secondaries is roughly $0.05/0.16 = 0.3$ times the light in the transmission band. This reduces the contrast of the solar Hα features very appreciably. With the Solc filter the diluting light would be nearly 2.5 times as great, and only the most contrasty of the solar features would de detectable. For the observation of emission line objects against a continuous background, however, like the prominences at the solar limb, the Solc filter should function reasonably well, although it is still definitely inferior to the equivalent Lyot filter in contrast.

One device for reducing the parasitic light in the Solc filter would be the addition of a secondary suppressor plate, in the form of a single thick retardation plate with a transmission maximum coinciding with the passband, and zeros coinciding with the first secondaries. This calls for one additional polarizer. The same device applies equally to the Lyot filter, and in practice it improves the performance quite substantially. The suppressor plate does not change the relative merits of the two filters.

The off-axis performance of the Solc filter has not been investigated analytically. However, we should expect that when n is large, the folded form should have very nearly the same off-axis characteristics as the simple Lyot filter. The field characteristics of the fan filter, however, are not so readily apparent. Solc states that the two forms have the same off-axis characteristics, which is surprising. He further states that the wavelengths of the transmission bands can be shifted over a broad range by tilting the filter. This finding is correct in the same sense that it is true for a Lyot filter. As the angle of inclination increases, however, the angular field over which the wavelength of the transmission band is uniform within a given tolerance (0.1 of the band width, for instance) decreases approximately with the reciprocal of the angle of inclination.

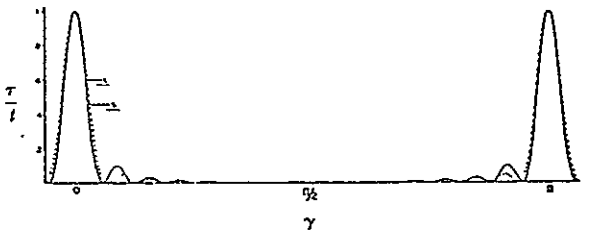


FIG 5. Superposed transmission curves of a 16-plate Solc filter (solid curve) and a 4-element Lyot filter (dotted curve).

Thus this type of tuning would be useful for the observation of a point source at infinity, but is of little use in the observation of an extended field, or in a converging beam of light.

The useful field of the Solc filter can be greatly extended by the same devices that are applicable to the Lyot filter. However, this calls for a compound construction of each retardation plate, and multiplies the complexity of construction several fold.

A Solc filter with only two polarizers is probably more expensive than the equivalent Lyot filter for band widths greater than 3 A, and most certainly so for sharper bands. A filter with a 5-A band width calls for 60 or 70 retardation plates, and the number increases with the reciprocal of band width. Since they are identical, many such plates can be ground and polished in one operation. However, the loss in birefringent material in the sawing process would be very considerable, and the mechanical problem of mounting and cementing so many elements in their proper orientation might be expensive. If the material is quartz, the thicknesses of individual retardation plates presumably would be between 1 and 2 mm. For band widths of less than 2 A it is almost necessary to use calcite to avoid excessively thick filters. The thickness of a single plate is then of the order of 0.06 to 0.12 mm, and the problem of working this rather difficult material to the requisite accuracy in thickness and flatness (about 0.2 micron) in several hundred plates becomes really formidable.

In practice, one would avoid excessively thin plates in the Solc filter by compounding it. For example, a ½-A filter could be made with a first stage of 25 quartz retardation plates about 1.1 mm thick and a second stage of 25 calcite plates 1.4 mm thick (the exact thicknesses depend upon the desired wavelengths of the transmission bands). An additional polarizer must, of course, be inserted between the two stages. I suspect that such a filter would be comparable in cost with an equivalent Lyot filter equipped with Rochon prism polarizers. Of the two the Lyot filter is preferable because of its superior suppression of parasitic light.

In summary, then, the Lyot filter performs better than the Solc filter except for possible applications where the parasitic light of the latter near the transmission bands is of no importance.

A VERSATILE BIREFRINGENT FILTER

by

Kerstin Fredga

Sonnenborgh Observatory, Utrecht, The Netherlands

Stockholm Observatory, Saltsjöbaden, Sweden

and

J. A. Högbom

Kapteyn Astronomical Laboratory, Groningen, The Netherlands

The Netherlands Foundation for Radio Astronomy

ABSTRACT

The theory of the Solc birefringent filter is complicated and the general formulae describing its optical properties do not invite to any simple physical picture. Due to these formal difficulties, many of its inherent possibilities have not been fully appreciated. In this paper we point out some new possibilities for filters based on this general design.

The detailed shape of the transmission profile is a function of the angles of the crystal plate optic axes. It is shown how unwanted transmission sidelobes can be suppressed to any desired level by altering the distribution of plate angles. By the same means, the transmission band can be split into two symmetrically placed replicas, and the distance between the two bands can be varied. In this way the filter can easily be tuned over half the free spectral range.

Some of the error sources that are important to the filter performance are discussed. An expression is given for the amount of parasitic light introduced by errors in the orientation of the plate optic axes and it is found that errors as large as $0\overset{\circ}{.}5$ can be tolerated. A laboratory experiment with 16 birefringent plates has shown that accuracies an order of magnitude better than this figure can easily be achieved. Manufacturing errors in the thickness of the plates can be compensated for by assembling the pile of plates in a particular way. The very strict tolerances usually quoted can therefore be considerably relaxed.

1. INTRODUCTION

Birefringent filters with narrow bandpass have been used for many years in astronomical research, primarily for monochromatic observations of the sun. The main advantage of such filters is their speed: within the selected wavelength band, the light from the whole field of view is available all the time. When observing with a conventional spectroheliograph, only a fraction of the total time is used for any one region as the slit is scanned across the field of view. The instant view of the whole field is also a great advantage when observing rapidly changing phenomena on the solar disc.

The first type of birefringent filter was developed by Lyot (1933, 1944) and by Öhman (1938). A second type was designed by Solc (1953, 1959, 1960, 1965). Both types of filter have been further discussed by Evans (1949, 1958).

Both filters utilise retardation plates (usually of quartz or calcite) cut with the optic axis parallel to the surface of the plate. In the Lyot-Öhman type filter, the plates are sandwiched between linear polarizers and the thickness of each plate is twice that of the preceding plate. All plates are oriented with their optic axes parallel and at 45° to the axes of the polarizers. On entering a plate from the preceding polarizer, the light is divided into vector components which travel through the plate at different phase velocities. The two components are combined in the following polarizer and the transmitted wave will have an amplitude proportional to $\cos\gamma$ where 2γ radians is the phase difference (retardation) introduced by the birefringent plate at the wavelength in question. A complete filter consisting of N such units with plate thicknesses in the ratio $1:2:4:\ldots 2^{N-1}$, will have a transmission proportional to $[\cos\gamma_1 \cdot \cos 2\gamma_1 \cdot \cos 4\gamma_1 \cdots\cdots \cos 2^{N-1}\gamma_1]^2$ which leads to the well-known expression for this filter (Lyot 1944):

$$T_L \sim \left(\frac{\sin 2^N \gamma_1}{2^N \sin \gamma_1} \right)^2 \qquad (1)$$

where $2\gamma_1$ is the retardation of the thinnest plate.

The Solc type of birefringent filter employs a pile of equally thick, but differently oriented, retardation plates and only two polarizers, one at each end of the pile. It is difficult to see directly how the Solc filter accomplishes its purpose. The general formulae describing the optical properties of such a pile of birefringent plates are complicated and do not invite to any simple physical picture of the relation between the orientation of the plates and the resulting transmission profile. Only for a few special cases has it been possible to derive analytical expressions for the filter transmission. However by following, on the Poincaré sphere, the polarization of the light as it passes through the different plates, one can gain some understanding of how the filter works.

Many of the inherent possibilities of this design have not been fully appreciated because of these formal difficulties and consequently Solc filters have not been as widely used as the

Lyot-Öhman type. However, since the Solc filter uses only two polarizers it has obvious advantages, especially in the near ultraviolet region, where available polarizers are either bulky or else give imperfect polarization and significant absorption (Fredga 1969a, b, Fredga and Högbom 1970). We shall here discuss the properties of the Solc filter and point out some new possibilities for filters based upon this general design.

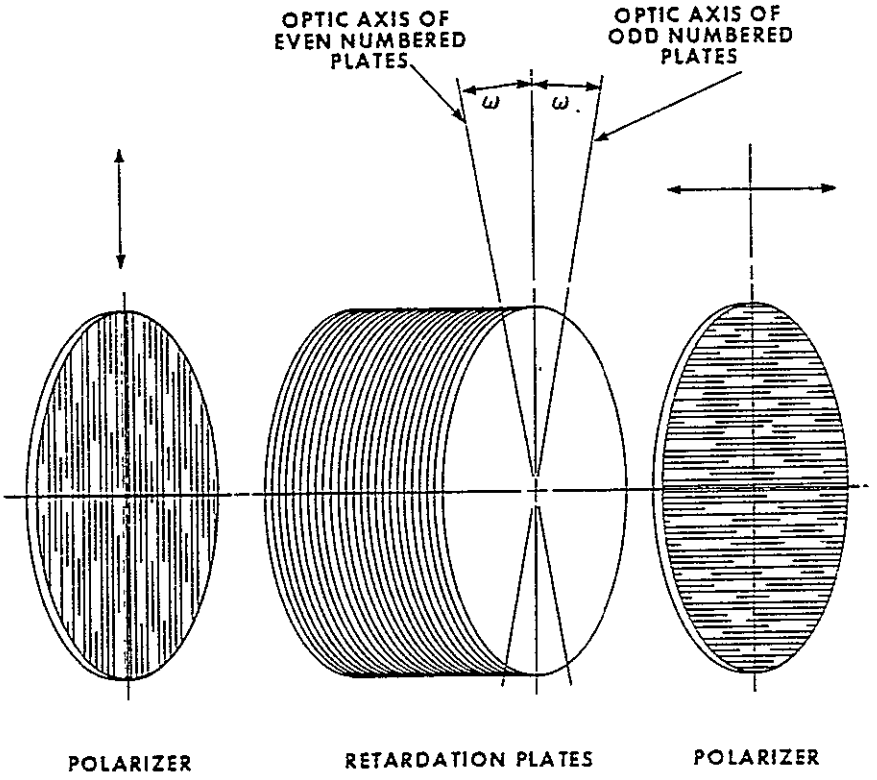


Figure 1. Simple Solc filter of the folded type.

2. MAIN PROPERTIES OF THE SOLC FILTER

Solc describes two main versions of the filter. In his type I filter, called the *folded filter* by Evans, the axes of the plates are all rotated the same angle with respect to the axis of the first polarizer, but alternate between $\omega = +\rho$ and $\omega = -\rho$. The end polarizer is oriented at right angles to the first (Figure 1). In his type II filter, called the *fan filter* by Evans, the angles ω_n of the retardation plates increase monotonically in the series $\rho, 3\rho, 5\rho, \ldots (2N-1)\rho$, and the end polarizer is here oriented parallel to the first polarizer. For both filters the quantity ρ is given by

$$\rho = \pi / 4N \tag{2}$$

where N is the number of plates.

Evans (1958), using the Jones (1941 a, b) matrix calculus, found that the on-axis transmission of the fan filter for unpolarized light can be written:

$$T = \frac{t}{2}\left(\frac{\sin N\chi}{\sin \chi} \cdot \sin 2\rho \cdot \cos \gamma\right)^2 \tag{3}$$

The parameter χ is defined by

$$\cos \chi = \cos \gamma \cos 2\rho \tag{4}$$

2γ (radians of phase) is the retardation of a single plate and related to the wavelength λ by:

$$2\gamma = 2\pi d \mu / \lambda \tag{5}$$

where d is the plate thickness and $\mu = (n_e - n_o)$ the difference between the extraordinary and ordinary refractive indices. The constant t represents the losses due to unwanted reflections and absorption in the filter. The same formulae are valid for the folded filter if $\cos\gamma$ is replaced by $\sin\gamma$. Transmission maxima occur at wavelengths for which $2\gamma = k \cdot 2\pi$ and $2\gamma = (k + \frac{1}{2}) \cdot 2\pi$ for the fan filter and the folded filter respectively (k is an integer). The main transmission bands are accompanied by sidelobes, the closest of which are as high as 12 %. In section 3 we shall see how these can be reduced to any desired level by choosing a modified distribution of angles ω_n for the plates in the pile.

The width of individual transmission bands expressed as an interval in γ can be calculated from Eq. (3). Differentiating Eq. (5) we get the relation between this and the corresponding wavelength interval. The bandwidth between half intensity points becomes

$$\Delta \lambda_{1/2} = 0.80 \cdot \frac{\lambda^2}{N \cdot d \cdot \mu} \cdot q \qquad (6)$$

where $q = 1 / \left(1 - \frac{\lambda}{\mu} \frac{\partial \mu}{\partial \lambda} \right)$

The bandwidth is inversely proportional to the total thickness of the pile of retardation plates, $N \cdot d$.

The distance between successive transmission maxima (the free spectral range) is determined by the properties of the individual plates.

$$\lambda_{m+1} - \lambda_m = \frac{\lambda^2}{d \cdot \mu} \cdot q \qquad (7)$$

and the finesse of the filter is

$$F = (\lambda_{m+1} - \lambda_m) / \Delta\lambda_{1/2} = 1.25 N \qquad (8)$$

For quartz, the factor q is close to unity (≈ 0.9) for wavelengths above 5000 Å, but decreases rapidly as one approaches the ultraviolet part of the spectrum ($q = 0.72$ at 2800 Å and 0.55 at 2000 Å). This makes the ultraviolet transmission bands even sharper and closer spaced than indicated by the λ^2 dependance alone.

The form of the Jones calculus chain of matrices for a pile of birefringent plates shows that the most relevant parameters describing the filter are 1) the distribution of *angle differences* α (between adjacent plates or between plate and polarizer) and 2) the *semi-retardation* γ of the individual plates. The latter is related to the wavelength by Eq. (5). Most equations become unnecessarily complicated when expressed directly in terms of the plate angles ω and the wavelength λ and we shall replace these by the more convenient parameters α and γ.

3. CONTROL OF BANDWIDTH AND SIDELOBES

The 12 % sidelobes accompanying the main transmission bands are too large for many practical applications. Solc (1960, 1965) describes how the sidelobes can be suppressed by tapering the distribution of plate angles. The quantity ρ is made larger near the centre of the pile and smaller towards the ends. Maximum transmission at the center of the band occurs when the distribution of ρ is normalized so that a more general version of Eq. (2) is satisfied:

$$\sum_{n=1}^{N+1} | \alpha_n | = \pi / 2 \tag{9}$$

where α_n is the angle difference between neighbouring elements in the filter:

$$\alpha_n = \omega_n - \omega_{n-1} \tag{10}$$

For this purpose $\omega_o = 0$ represents the orientation of the first polarizer, while the last polarizer is oriented at right angles to ω_{N+1}. Thus for the folded filter, $\omega_{N+1} = 0$ and, for the fan filter, $\omega_{N+1} = \pi / 2$. Eq. (9) can be derived by tracing the polarization of the transmission maximum wavelength on the Poincaré sphere.

From a theoretical point of view it is more satisfactory to consider the tapering as applied to the angle differences α rather than to the quantity ρ. A fan filter that is *untapered in* α will have the plate angles 0 = first polarizer, α, 2α, 3α . . . (the folded filter would have: 0, α, 0, α . . . .) and a transmission:

$$T = \frac{t}{2} \left(\frac{\sin M\chi}{\sin \chi} \cdot \sin \alpha \right)^2 \tag{11}$$

where $\cos \chi = \cos \gamma \cos \alpha$

and $M = N + 1$ is the number of angle differences in the filter. The versions discussed by Solc and Evans (Eq. 3) should then strictly speaking be regarded as tapered in that the two extreme angle differences are only half as large as the others. The transmission profiles are however practically identical when the filters contain a large number of plates. The same is true for the tapered versions of these filters.

It is instructive to investigate the transmission profiles resulting from different distributions of the angle differences α_n. Figure 2 shows computed transmission profiles for differently tapered filters. The second example represents the simple untapered filter described by Eq. (11). The third example shows the "roof" tapering which is similar to that often used by Solc. This gives a sidelobe level which is acceptable for most practical applications.

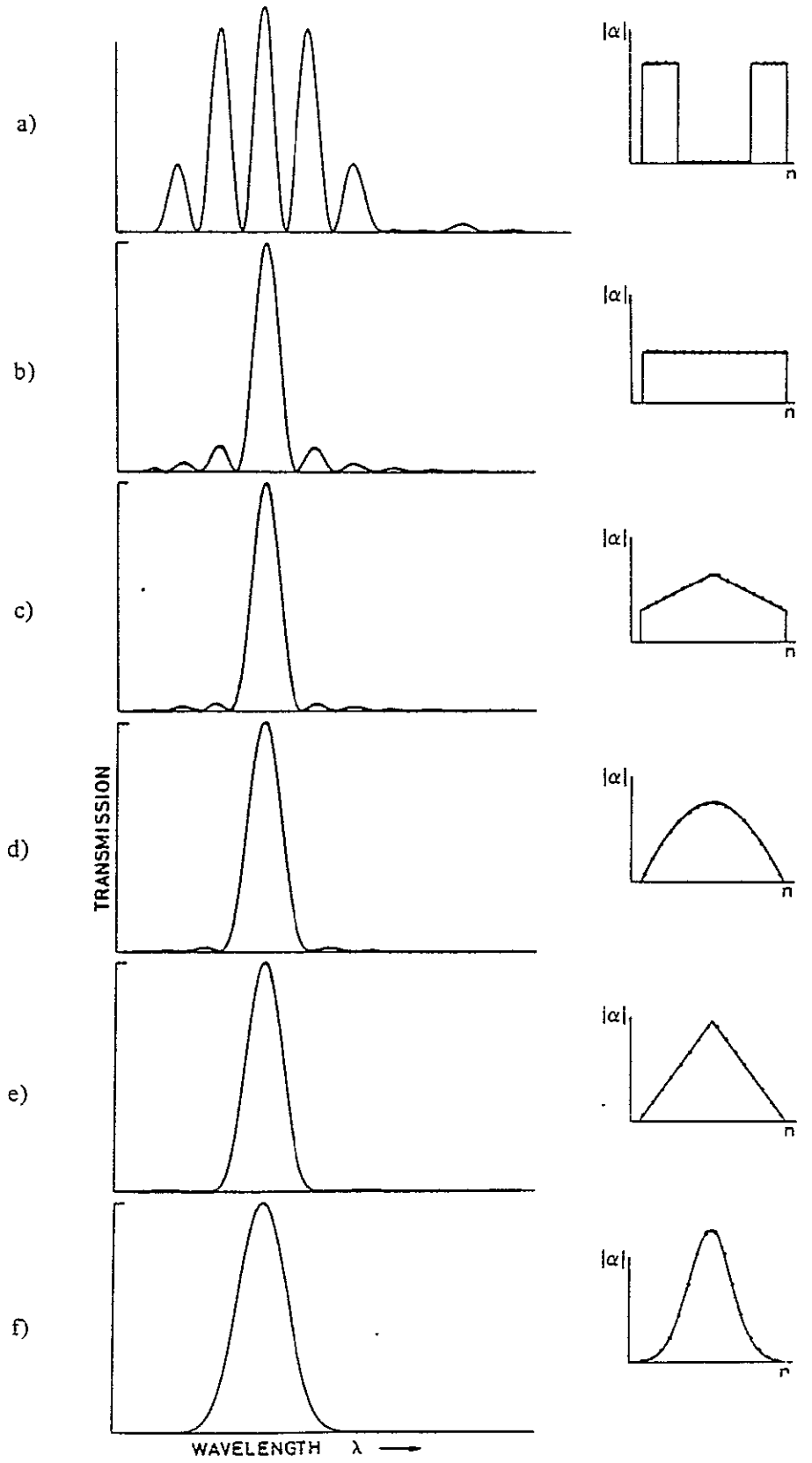


Figure 2. Transmission profiles for differently tapered filters. The absolute value of the plate angle differences α_n are plotted to the right. The tapering functions are *a*) 'interferometer', *b*) no tapering, *c*) roof with a 2:1 amplitude ratio, *d*) cosine, *e*) triangular and *f*) gaussian, truncated at 1/100 of its maximum amplitude.

Table 1. Bandwidth relative to that of an untapered filter and maximum transmission in the first sidelobes for different tapering functions.

Tapering function	Bandwidth	First sidelobe (in % of main peak)	Second sidelobe (in % of main peak)
Interferometer	0.76	(90.7)	(30.8)
No tapering	1.00	11.6	3.9
Roof (2:1)	1.11	3.4	2.1
Cosine	1.36	1.8	.22
Triangular	1.47	.09	.47
Gaussian (100:1)	2.00	.005	.0002

Figure 2 and Table 1 illustrate the general trend towards reduced sidelobes and increased bandwidth as the tapering function is taken from the extreme "interferometer" shape to the smooth gaussian form. Many other shapes of the main peak such as, for instance, a flat-top profile can be produced by suitable tapering functions. In order to obtain wider bands than those shown in Figure 2, one simply applies correspondingly narrower tapering functions of the same shape. A Solc filter in which the plates can be rotated independently can therefore be used as a variable bandwidth device where, in addition, the detailed shape of the profile can be chosen to suit the particular application in question.

The numerical calculations in this paper have been performed using various versions of a computer program developed by Beckers and Dunn (1965) and based on the Jones (1941 a, b) matrix calculus.

4. RELATION TO FOURIER TRANSFORMS

The general shape of the transmission bands and the changes in this shape resulting from the various types of tapering illustrated in Figure 2, suggest that the relation between the distribution of plate angle differences α_n on the one hand and the transmission profile on the other is similar to a Fourier transform.

The general relation to Fourier transforms can be understood with reference to Figure 3. At each boundary between two plates (or between a plate and a polarizer) there is an exchange between the wave components travelling in the ordinary and extraordinary modes. A certain fraction, $\sin \alpha_n$, of the wave component previously carried in the ordinary mode will, after the boundary, add vectorially to the component in the extraordinary mode and vice versa. The wave emerging from the end polarizer can then be described as the vector sum of field components passing from the input to the output along all possible paths. The different paths will produce output field components of different amplitudes, but the phase of every component must have one of the values $k \quad 2\gamma$ radians ($k = 0, 1, 2 \ldots N$) relative to that of the component that has travelled in the ordinary mode all through the pile (dashed in Figure 3).

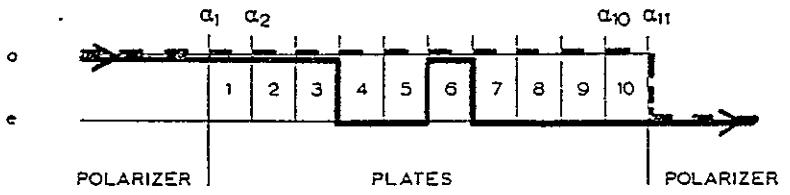


Figure 3. Illustrating the many different paths by which the light can travel from the input to the output of the filter. The heavy line represents an output vector component that has crossed over three times between the two propagation modes. It has passed through a total of six plates in the extraordinary mode and therefore emerges at a phase of $6 \cdot 2\gamma$ relative to the component that has stayed in the ordinary mode all through the filter (dashed).

Let g_n be the sum of the amplitudes of all field components that have passed through the pile in such a way as to produce the phase $n \cdot 2\gamma$. The output wave field can then be written (excluding the frequency factor $e^{j\omega t}$):

$$E(\gamma) = \sum_{n=0}^{N} g_n \exp(j 2\gamma n) \tag{12}$$

This has the form of a Fourier series. The filter transmission is proportional to the square of the field strength

$$T(\gamma) \sim |E(\gamma)|^2 \tag{13}$$

Eq. (12) shows that, independant of the arrangement of the plate angles, the transmission profiles must obey certain well known rules of Fourier series mathematics. The transmission peaks will be repetetive at intervals of π in the coordinate γ, and the profiles will be symmetrical about all points $\gamma = k \cdot \pi$. Furthermore we can deduce that the finesse F for a particular filter cannot be made much greater than the number of plates N. The exact value depends upon how large sidelobes one is prepared to accept. An untapered filter for instance has a finesse of $1.25N$ but also 11.6 % sidelobes, while a cosine tapered filter has $F = 0.92N$ and only 1.8 % sidelobes. No arrangement of the plate angles can give a performance that is significantly better in this respect.

Each coefficient g_n is the sum of the contributions from all possible paths that produce the phase $n \cdot 2\gamma$. However, if all angle differences $\alpha \ll 1$ radian, the by far greatest single contribution will come via the path that makes only one transfer between the two modes. The coefficient g_n will therefore be strongly influenced by the one particular angle difference at which this transfer takes place: this explains why the relation between the tapering function and the transmission behaves in a way that is similar to a Fourier transform.

5. TUNING THE FILTER

Many different shapes of the transmission profile can be produced by altering the distribution of plate angles. An especially interesting possibility, which does not conflict with the symmetry requirement derived in the previous section, is that the normal transmission band can be split into two symmetrically placed replicas of itself. The distance between the two components can be chosen at will and an individual member of such a pair can therefore be tuned in wavelength by rotating the plates.

The modulation theorem in Fourier analysis shows that such a split peak will result if we can change the plate angles in such a way that the new coefficients G_n are related to those of the original filter according to the equation

$$G_n = 2\, g_n \cos 2\pi s n \tag{14}$$

The two component bands will then appear at a distance from the position of the original single band, which is $\pm s$ times the repetition interval. (Eq. 7)

There exist general methods for calculating the plate angle distribution that will produce a set of desired coefficients G_n (Harris, Amman and Chang, 1964), but the mathematics becomes discouragingly complex for large numbers of plates. Fortunately, the exact calculations will not be necessary in this case: the desired cosine modulation of the coefficients will, to a good approximation, result if one simply modulates the angle difference distribution α_n in exactly the same way:

$$A_n = 2\, \alpha_n \cos 2\pi s n \tag{15}$$

In Appendix 1 we discuss the effect on the transmission profile of a modulation of the angle difference distribution according to Eq. (15).

Figure 4 shows the result of numerical calculations for a set of 16 retardation plates when a cosine modulation has been applied to the original angle difference distribution α_n. As the modulation frequency s (expressed in cycles per plate) is carried from 0 to 0.5, the filter changes from the fan type, via split peak versions, into the folded type. For $s = 1$, or any other integer, we have returned to the original angle distribution.

Note that the original angle distribution may be that produced by any desired tapering function. The shifted peaks keep the same shape as the unmodulated peak. In this example the calculations were made for a cosine tapered angle difference distribution characterized by

$$\alpha_n = \frac{\pi^2}{4\,(N+2)} \cdot \sin\left(\pi \frac{n}{N+2}\right) \tag{16}$$

The first factor is the constant required to satisfy Eq. (9) for maximum peak transmission in the unmodulated filter.

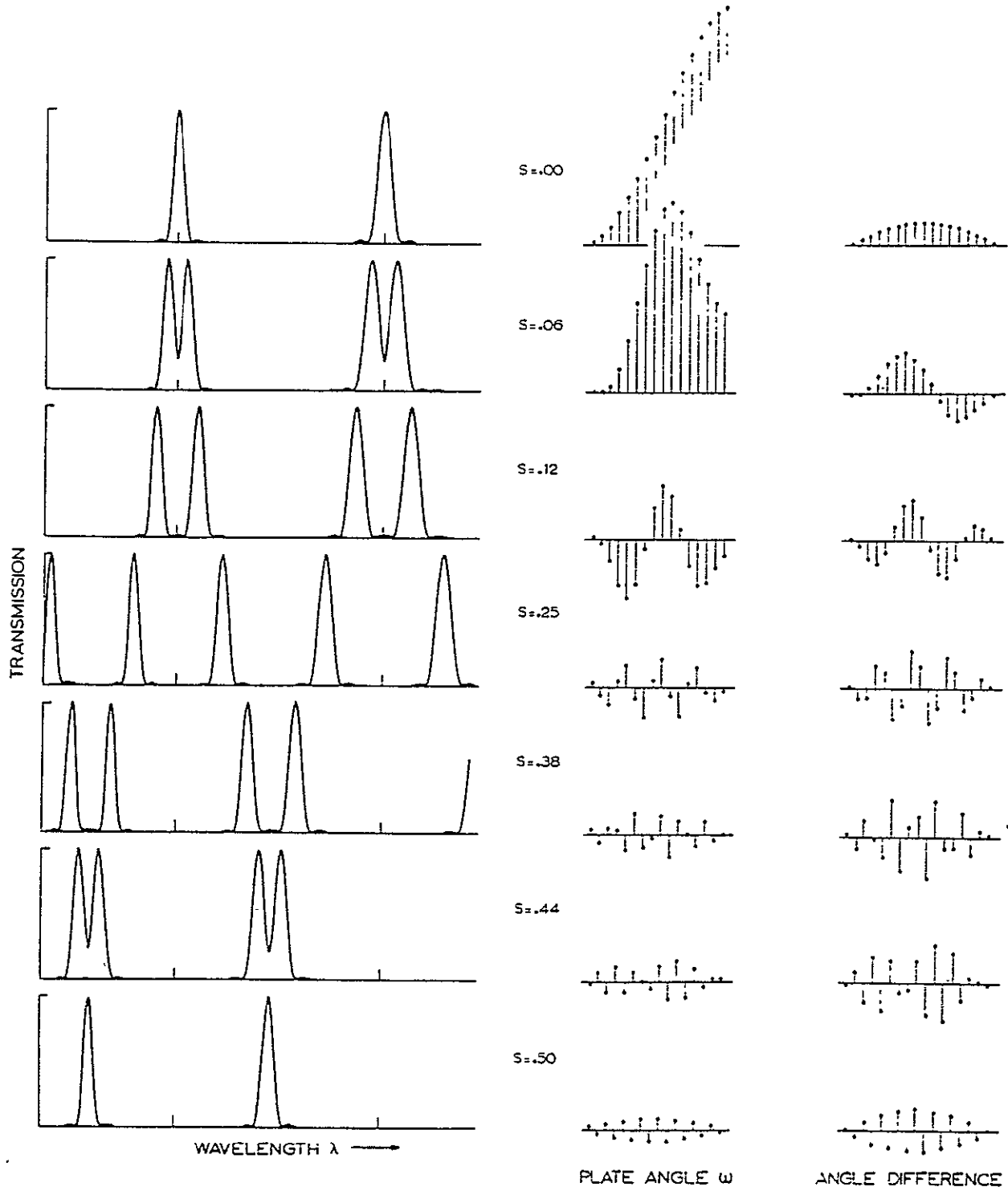


Figure 4. Splitting the transmission bands and tuning the filter. Transmission profiles, plate angles ω and angle differences α are shown for different modulation frequencies s. For $s = 0$, the plate angles go from 0 to 90°. By modulation through half a cycle (s goes from 0 to 0.5) the filter has transformed from a fan filter into a folded filter. Note that the bands become narrower and closer spaced at shorter wavelengths when plotted on a wavelength scale. The computations were made for a 16 plate cosine tapered filter.

The factor 2 in Eqs. (14) and (15) is only valid when the two components of the split peak are well separated. For small values of s, when the bands overlap, the constant deviates from the value 2, approaching unity as $s \to 0$. When calculating the examples in Figure 4 we have used a modified version of Eq. (15):

$$A_n = 2\,\alpha_n \cos\left(2\,\pi\,s\,[n - n_c] + \pi / 3\right) \tag{17}$$

The phase of the modulation at the filter centre, $n_c = (N + 2) / 2$, is then equal to $\pi / 3$. This has little influence on the transmission profiles when the two bands are well separated, but it gives a good approximation to the desired shape also when they are so close as to overlap. For $s = 0$, this equation produces the original fan filter configuration. The actual plate orientations are calculated from Eq. (10):

$$\omega_n = \omega_{n-1} + A_n \tag{18}$$

where, as before, the first polarizer is oriented at $\omega_0 = 0$. Observe that the end polarizer, which is always oriented at right angles to ω_{N+1}, will not necessarily be parallel to or at right angles to the first polarizer (Figure 5).

The region around $s = 0.25$ is best suited for the tuning since neighbouring unwanted bands will then be at their maximum distance and may be rejected by other means. The useful tuning range expressed in bandwidths will clearly be larger than in the example of Figure 4 if the filter contains a larger number of plates.

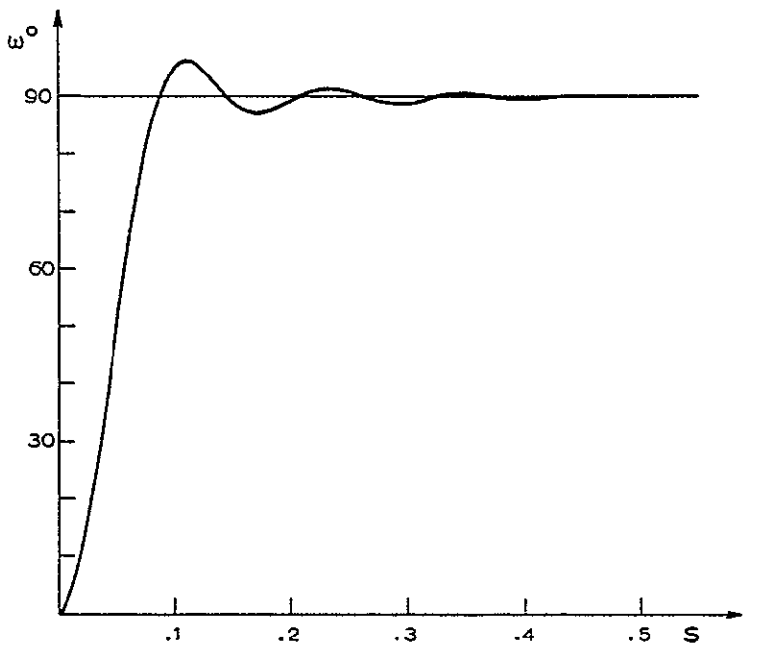


Figure 5. The orientation of the last polarizer as a function of the modulation frequency s for the filter shown in Figure 4.

6. ERROR SOURCES AND TOLERANCES

When designing and building a filter it is important to know the tolerances within which the various design parameters must be kept. Many types of error can be treated as equivalent to errors either in the plate angles ω_n or in the plate thicknesses d_n. These two basic errors affect the transmission profile in different ways. Errors in the angles ω_n will always produce disturbances that are symmetrical with respect to the normal fan or folded filter band positions because Eqs. (12) and (13) remain valid. The first of these equations was derived on the assumption that all plates are identical and errors in the plate thickness can, as we shall see, produce both asymmetrical disturbances and small shifts in the position of the main transmission bands.

6.1 Errors in the plate angles

In Appendix 2 we derive an expression for the average transmission level of spurious sidelobes caused by errors in ω_n (Eq. 42):

$$T_\epsilon = 4N\,\sigma_\omega^2\,\sin^2\gamma \tag{19}$$

σ_ω is the standard deviation of the errors in the plate angles ω_n expressed in radians. If σ_ω is expressed in degrees, the constant 4 should be replaced by $1.22 \cdot 10^{-3}$.

The average parasitic transmission is zero at wavelengths for which a fan filter has its peaks (where γ is a multiple of π) and increases to a maximum at wavelengths for which a folded filter has its peaks (where γ is an odd multiple of $\pi/2$). This is independent of whether the actual filter is of the fan type, the folded type, or has been tuned to some intermediate configuration. Figure 6 shows computed examples of sidelobes resulting from normally distributed random errors with a large standard deviation ($\sigma_\omega = 1°$).

The mean parasitic transmission averaged over all wavelengths (all γ) becomes

$$\epsilon = 2N\,\sigma_\omega^2 \tag{20}$$

We have computed ϵ for several filters with random errors in the plate angles ω_n. The errors added to the correct angles were normally distributed with standard deviations σ_ω equal to 0°.3, 0°.6 and 1°.0 . The calculations were made for cosine tapered fan filters with 20, 40 and 80 plates. The results are shown in Figure 7 and agree well with the values expected from Eq (20).

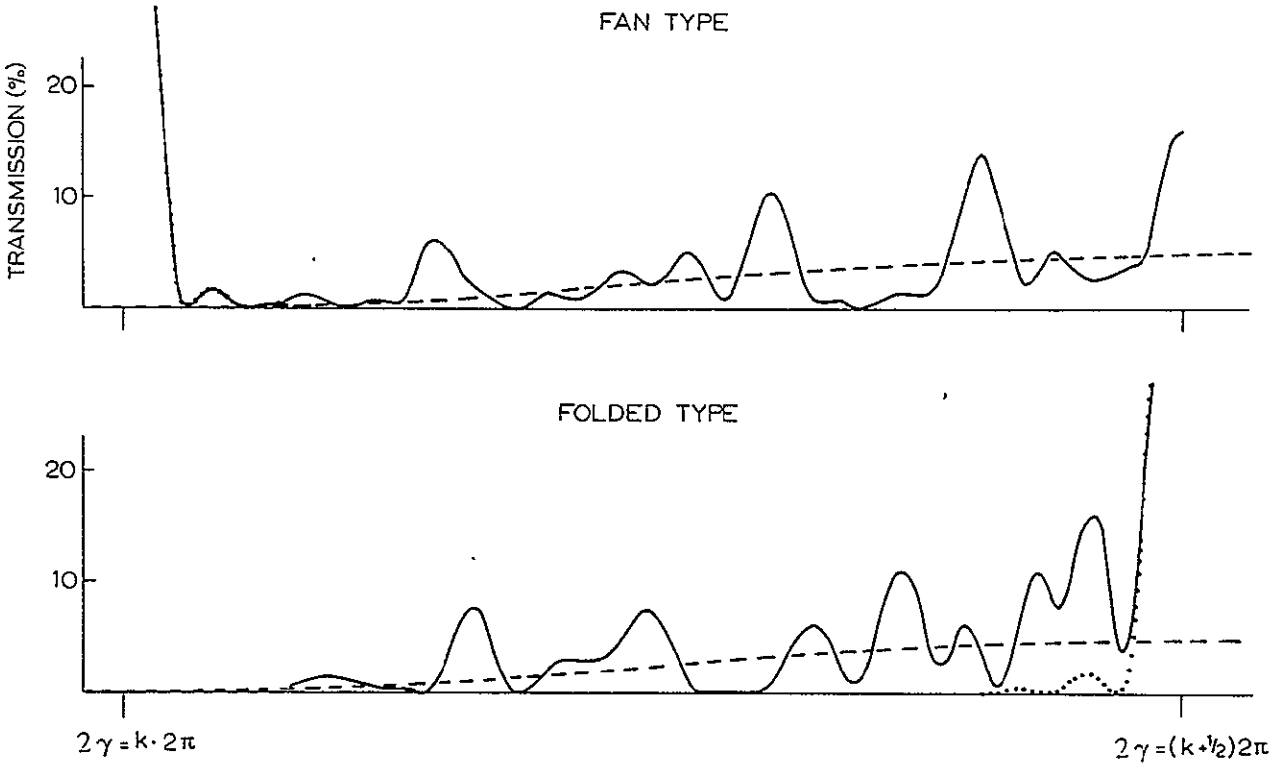


Figure 6. Examples of sidelobe patterns due to errors in the plate angles ω with the large standard deviation $\sigma_\omega = 1°$. The expected mean level according to Eq. (19) is dashed and the errorfree profile is dotted; 40 plate cosine tapered filter.

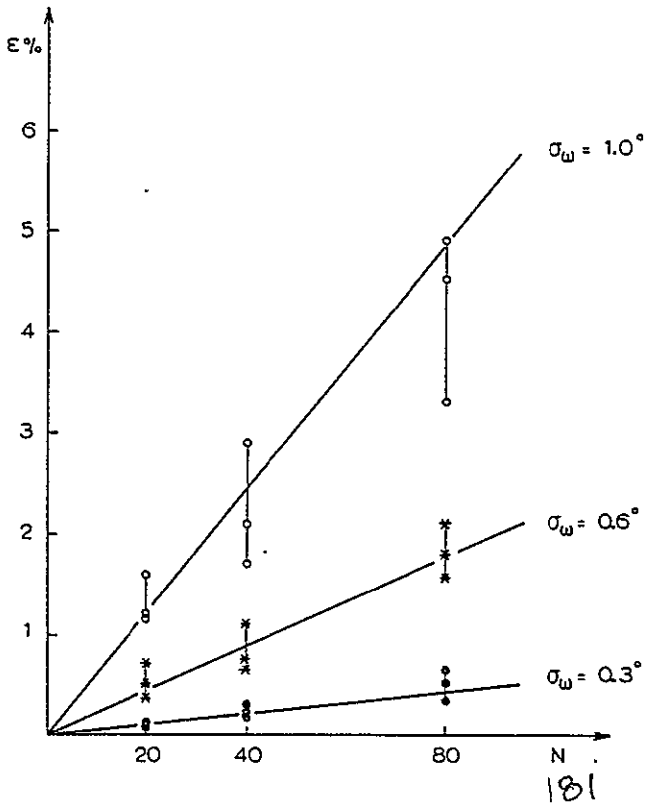


Figure 7.

Mean parasitic transmission averaged over all wavelengths due to errors in ω_n for fan filters with 20, 40 and 80 plates. The standard deviations used in the error distribution were $\sigma_\omega = 0.3°$, $0.6°$ and $1.0°$. Each point represents an independent error distribution. The straight lines represent theoretical values according to Eq. (20).

The ratio of the total energy transmitted in the error sidelobes to that transmitted in the main bands will be approximately equal to ϵ times the finesse F For normally tapered filters $F \approx N$ (sections 2 and 4) and we get

$$\epsilon F \approx 2N^2 \sigma_\omega{}^2 \tag{21}$$

A 40 plate filter, in which the plate angles have been adjusted to a tolerance of $\pm 0^\circ 5$ (corresponding to $\sigma_\omega \approx 0^\circ.3$) will thus have a total parasitic transmission due to the errors in ω_n which is only 8 % of that in the main bands. This agrees with the empirical findings of Beckers and Dunn (1965) that angle errors of this magnitude have little influence on the filter performance.

This result is somewhat surprising since errors of this magnitude are by no means small compared with the angle differences α_n themselves, which for a 40 plate filter are of the order of a few degrees. In practice it is easy to control the plate settings to a considerably higher accuracy (section 7), and the angle errors should therefore have a negligible influence upon the transmission profile of a well designed filter.

Apart from producing unwanted sidelobes, the angle errors can also distort the shape of the main band. However, it appears from computed examples that this effect will usually be less disturbing than the appearence of the sidelobes.

6.2 Errors in the plate thickness

The average sidelobe level away from the main bands of an untapered fan filter is derived in Appendix 2 (Eq. 47). From computed examples it appears that the formula can also be used for tapered filters.

$$T_\epsilon \approx \frac{\pi^4}{4} \frac{\sigma_0{}^2}{N \sin^2 \gamma} \tag{22a}$$

where σ_0 is the standard deviation of the plates from their average thicknesses expressed in orders of birefringence; the latter is related to the standard deviation σ_d in the actual plate thicknesses by:

$$2\pi\sigma_0 = 2\gamma(\sigma_d / d) \tag{22b}$$

The effects of these errors are seen to be more serious for short wavelengths (= larger γ, see Eq. 5). Furthermore, the sidelobes increase very steeply as one approaches the positions of the main bands because $\sin^2\gamma$ becomes very small. The same picture emerges when the calculations are performed for a folded filter or a tuned filter: in contrast to the behaviour of the angle error sidelobes, this error pattern will accompany a main band as it is tuned to a different wavelength.

It should not be concluded from Eq. (22a) that the situation improves for large N. The sidelobe level at any specified wavelength does indeed improve as more plates are added. At the same time, however, the bandwidth decreases and the sidelobe level at a specified number of bandwidths from the main peak gets worse. The result is that the total parasitic trans-

mission becomes larger relative to the transmission in the main bands. Eq. (22a) leads to an expression for the ratio of the total error transmission outside the main bands - here taken to be the region for which $|\sin\gamma| > \pi / N$ - to the transmission within these bands (compare Eq. 21):

$$\epsilon F \approx \frac{\pi^2}{2} N \sigma_0^2 \qquad (22c)$$

According to this formula, a 40 plate filter with $\sigma_0 = 1/40$ order should have a total error transmission equal to 12 % of the main band transmission. Unfortunately, the tolerances must often in practice be set stricter than this because the main bands themselves (and their immediate surroundings) are very sensitive to errors of this kind. The main profile can be seriously distorted and shifted in position by errors which would lead to an acceptable parasitic transmission according to Eq. (22c). Figure 8a shows the computed transmission for a filter with very large errors, $\sigma_0 = 0.12$ orders, in the plate thicknesses. The main peak has collapsed and the profile is dominated by the sidelobes whose general behaviour is described by Eq. (22a).

The very strict tolerances on the plate thicknesses makes it extremely difficult to manufacture filters containing a large number of plates. However, it is relatively easy to measure the errors in the individual plates once they have been made. The whole pile of plates can then be assembled in such a way that the main effects of the known individual plate errors cancel.

An error in the thickness of an individual plate may be looked upon as causing errors in the positions of all the angle differences along the filter axis relative to the positions they would have had in an errorfree filter of the same total length (Appendix 2, Eq. 43). If we now let one neighbouring plate have an error which is equal in magnitude but opposite in sign to that of the first plate, we shall find that all angle differences, except the single one between the two incorrect plates, will fall on their correct positions. The actual pile of plates has therefore become a much better approximation to the design configuration than when there was only one incorrect plate. The main detrimental effects of errors in the plate thicknesses can thus be cancelled if the filter is assembled in such a way that neighbouring plates have errors which are approximately equal in magnitude but opposite in sign. This can be done simply by ranking the plates according to their errors and pairing the plates with the largest errors of opposite sign, those with the second largest errors etc.. A further improvement is obtained by concentrating the plates with large errors towards the two ends of the pile where, as pointed out by Beckers and Dunn (1965), these errors are less critical to the filter performance, especially for tapered filters. The example given in Figure 8 and Table 2 illustrates the general principle and the striking improvement of the filter behaviour that is possible by this method.

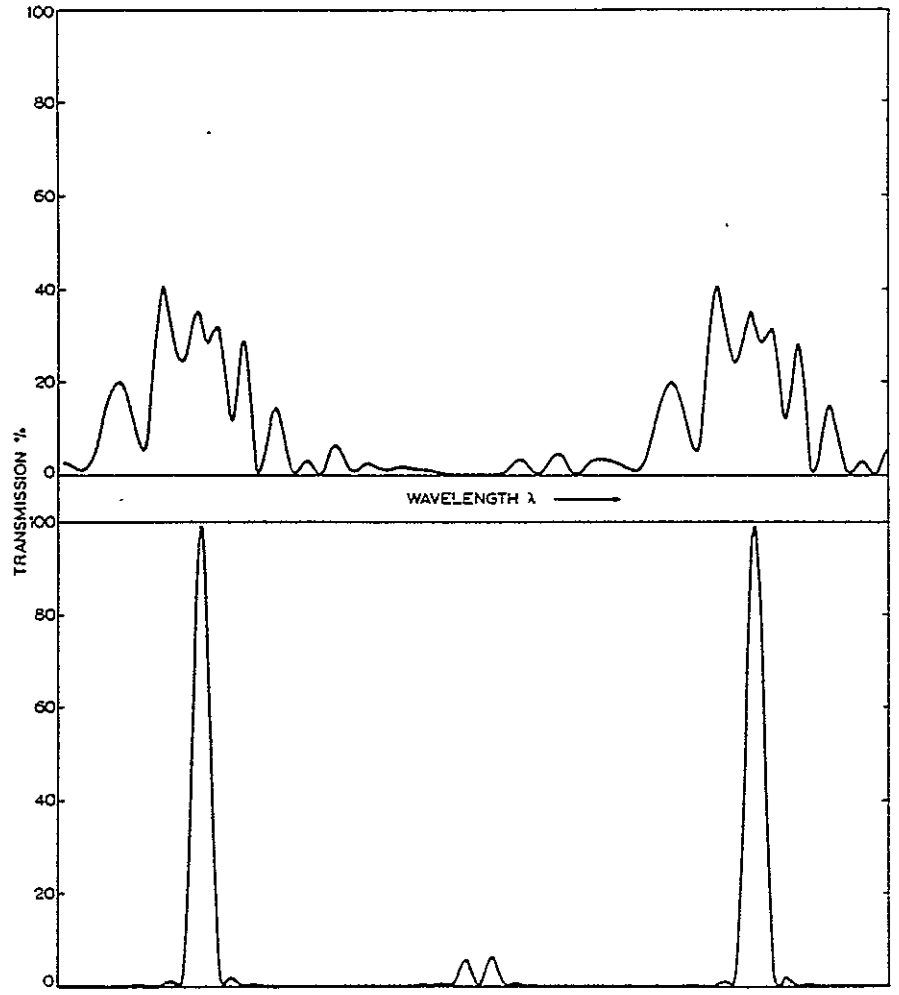


Figure 8. Transmission profiles of a 31 plate cosine tapered filter with large errors in the plate thicknesses d corresponding to a standard deviation σ_0 of 0.12 orders of birefringence. a) the plates assembled without attention to their individual errors, b) the plates assembled so as to cancel the effects of the errors (see Table 2).

Table 2. Characteristics of the cosine tapered filters illustrated in Figure 8. The plate angles are given for both the fan and the folded versions. Column A gives the plate thickness errors expressed in orders of birefringence as produced by a random number generator (normal distribution with $\sigma_0 = 0.12$ orders). Column B gives the rearranged distribution of the same plates.

Plate nr.	Plate angle ω (degrees)		Thickness error (orders)	
	fan	folded	A	B
Pol.	.0	.0		
1	.4	.4	.08	-.29
2	1.2	- .4	-.02	.22
3	2.4	.8	.02	-.19
4	4.0	- .8	.07	.20
5	6.0	1.2	-.08	-.14
6	8.3	-1.1	-.29	.08
7	11.0	1.5	-.09	-.09
8	13.9	-1.5	-.22	.07
9	17.2	1.8	.20	-.05
10	20.6	-1.7	-.09	.05
11	24.4	2.0	.04	-.03
12	28.3	-1.9	.24	.04
13	32.3	2.2	.22	-.02
14	36.5	-2.0	.06	.02
15	40.7	2.2	.12	-.01
16	45.0	-2.0	-.02	.00
17	49.3	2.2	-.02	.01
18	53.5	-2.0	-.03	-.02
19	57.7	2.2	-.04	.02
20	61.7	-1.9	.02	-.02
21	65.6	2.0	.05	.04
22	69.4	-1.7	.20	-.04
23	72.9	1.8	.08	.06
24	76.1	-1.5	-.19	-.08
25	79.1	1.5	-.05	.08
26	81.7	-1.1	-.17	-.09
27	84.0	1.2	.04	.12
28	86.0	- .8	.01	-.17
29	87.6	.8	.00	.20
30	88.8	- .4	-.01	-.22
31	89.6	.4	-.14	.24
Pol.	90.0	.0		

7. EXPERIMENTAL TESTS

The theoretical conclusions arrived at above have been tested in the laboratory. The aim was to investigate the practical difficulties and main sources of error, but not to construct a final filter for some special purpose. 16 birefringent plates of quartz were used. These plates were originally part of a filter for the ultraviolet wavelength region built by Solc's group at the Dioptra Company in Turnov, Czechoslovakia. Each plate is only 0.41164 mm thick which gives a large free spectral range in the ultraviolet.

The plates were mounted in a specially constructed filter holder where each plate could be independently rotated about the filter optic axis by means of micrometer screws (see Figure 9). The plates were immersed in a fluid (Leitz Immersion oil) of refractive index 1.515 (NaD) which reduced the reflection at each surface to 0.01 %. HN 32 linear sheet polaroids were used as polarizers.

The tests were carried out at the Technisch Physische Dienst in Delft, Netherlands, using a 1-m Hilger-Watt scanning monochromator with a Tungsten ribbon lamp as light source and a 1 P21 RCA photomultiplier as detector. The filter was placed in the collimated beam after the exit slit of the monochromator.

For convenience the tests were carried out in the visible region around 5000 Å. The slitwidths of the monochromator were chosen to give a spectral resolution of 1.2 Å which made it possible to study the detailed shape of the profiles, which in the 5000 Å region had a bandwidth of about 30 Å.

The crystal optic axes of the plates were aligned in the following way. The monochromator was set at a wavelength corresponding to maximum transmission for a folded filter, and the polarizers at each end of the filter were oriented at right angles to one another. At this wavelength each plate is a half-wave plate and, if its optic axis is aligned parallel to the axis of one of the polarizers, no light will be transmitted but a small deviation from the parallel position is easily detected. The plates were aligned in this way one after the other, leaving the earlier aligned plates in their parallel position. This procedure turned out to be very accurate and allowed the angles of the optic axes to be controlled to $\pm 0.02°$. This is much better than actually needed (section 6).

The spectral response of the filter was determined relative to the transmission when the end polarizer had been placed in front of, and parallel to, the first polarizer. With this as a reference the transmission of the main peaks (ordinary peaks as well as tapered and shifted peaks) should be 100 %.

The absolute transmission of the filter at 5000 Å was determined to 13 %. The polarizers transmit 34 % of unpolarized light while the quartz plates and the oil (21 cm long path) transmitted 56 % of the incident light. With more transparent polarizers (e.g. Rochon prisms) and a reduced pathlength through the oil, the transmission can be considerably improved. The filter may also be used in any wavelength region where quartz is transparent if a suitable immersion oil and suitable polarizers are chosen.

No absolute wavelength calibration was performed but the experimental maxima as shown on the monochromator wavelength scale fitted the theoretical values to within a few Ångströms. The reduction of the measurements was done assuming that the experimental and theoretical transmission maxima coincided in wavelength for the unshifted bands.

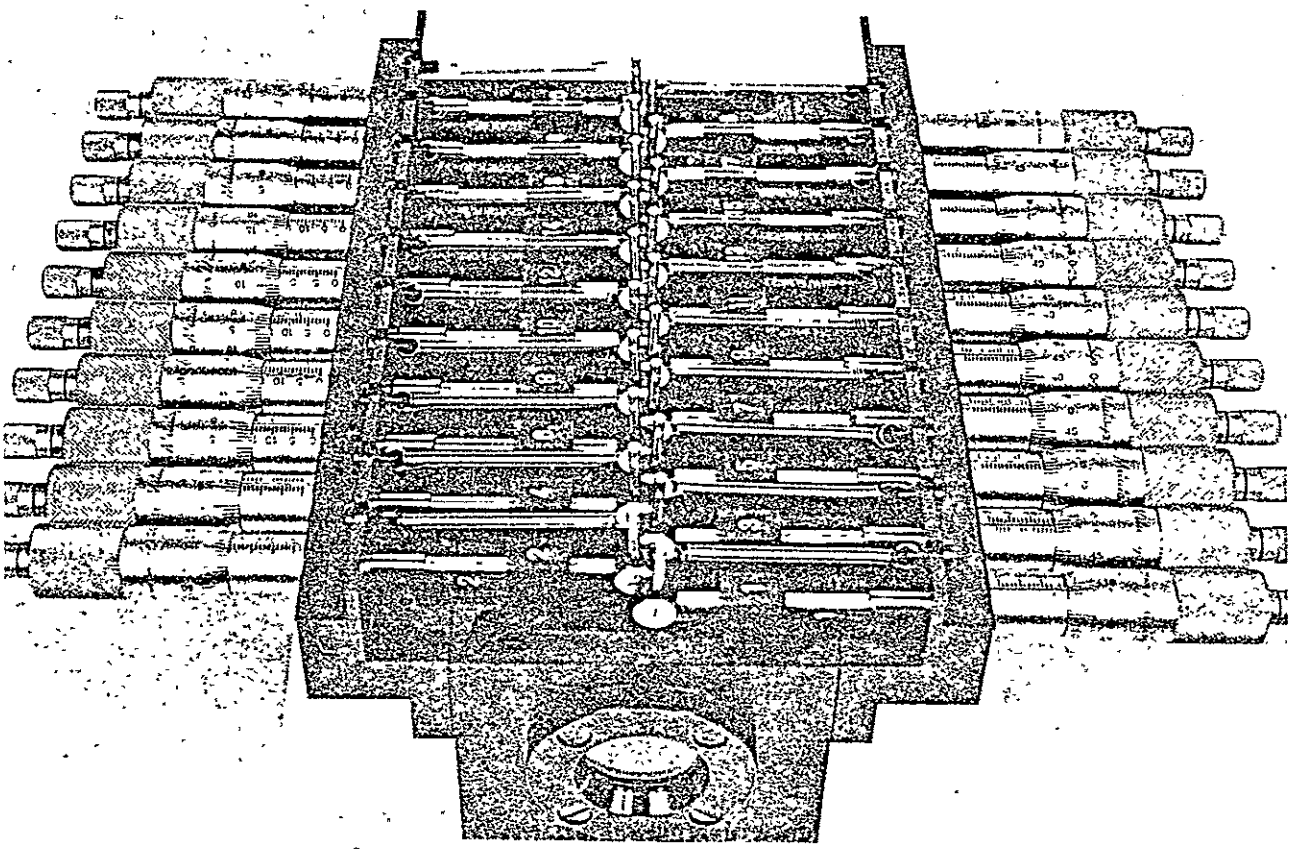


Figure 9. The filter holder used in the experimental tests. The polarizers and plates can be rotated independently about the filter axis by means of the micrometer screws.

The first test run concerned the tapered profiles illustrated in Figure 2. Table 3 summarizes the results giving theoretical and experimental parameters for the main bands and the first two sidelobes. The measured maximum transmission of the main peaks amounted to 100 ± 1 % relative to the reference described above. The error of ± 1 % may be due to slow fluctuations in the output from the light source which occurred between actual registration and calibration runs. When evaluating the rest of the profiles all transmission values are expressed in percent of the main peak. The experimentally determined transmission values fit the theoretical values to within 1.5 %. This means that the difference between the experimental and the theoretical curves would hardly be noticable on the scale of Figure 2.

InFigure 10the sidelobes of the straight, the cosine and the gaussian tapered filters are shown on an expanded scale. The largest deviation from the theoretical curves occur in the second sidelobes of the straight filter and amounts to ± 1 %. The deviations are generally antisymmetrical with respect to the main peak and therefore probably due to small errors in the plate thicknesses.

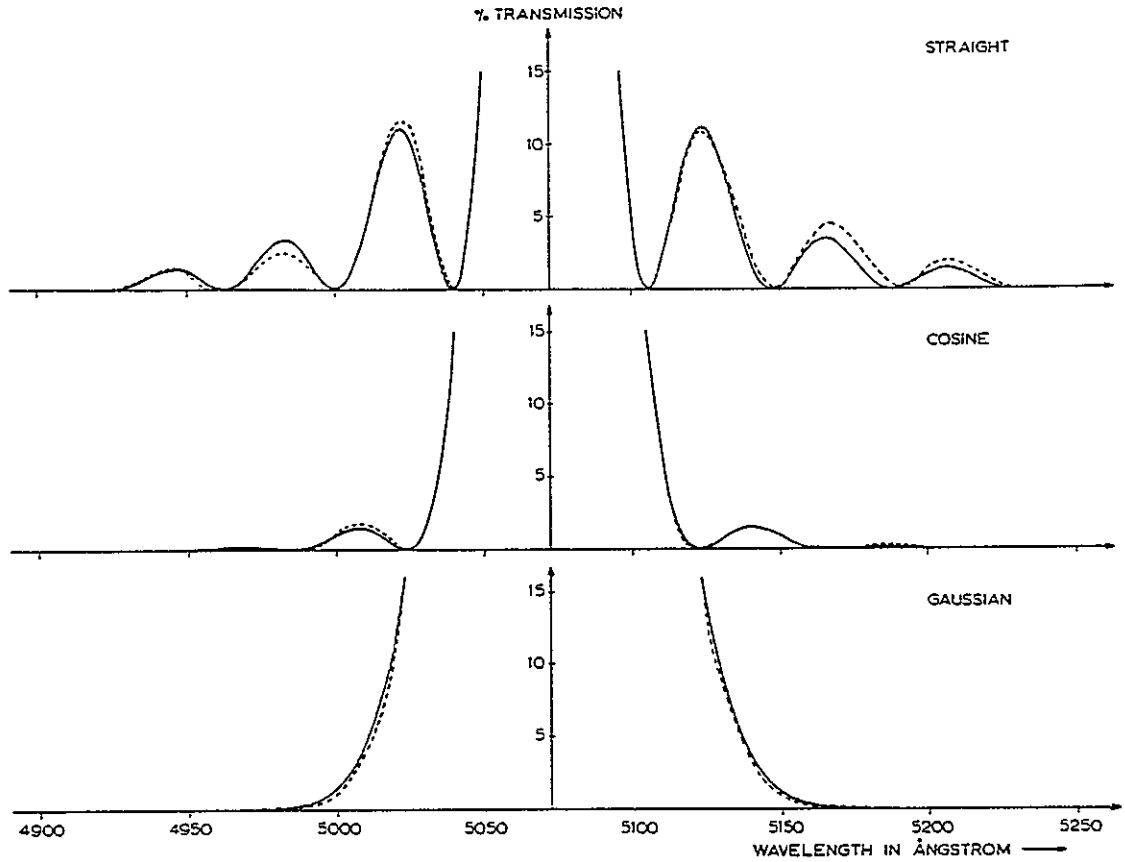


Figure 10. Result of laboratory measurements on a set of 16 retardations plates. Theoretical (full drawn) and experimental (dashed) transmission curves are given for the straight (untapered), the cosine tapered and the gaussian tapered arrangements.

Table 3. Experimental results for a 16 plate filter with different tapering functions. Of the two experimental values given for the sidelobes, the first refers to the long wavelength side and the second to the short wavelength side of the main band. Computed theoretical values are given in brackets.

Tapering function	Main band		First sidelobe	Second sidelobe
	T_{max}	$\Delta\lambda_{1/2}$ Å	T_{max} %	T_{max} %
Interferometer	1.002	24 (23)	91.6 (90.1) 88.9	29.3 (30.5) 32.0
No tapering	1.003	30 (30)	11.6 (11.1) 10.8	2.4 (3.47) 4.4
Roof	0.990	35 (34)	3.4 (2.90) 2.9	1.5 (1.71) 2.4
Cosine	0.997	41 (41)	1.6 (1.52) 1.5	0.1 (0.17) 0.3
Triangular	1.007	46 (44.5)	0.3 (0.08) 0.3	0.3 (0.32) 0.7
Gaussian	1.005	60 (61.5)	- (0.003) -	- - -

Table 4. Experimental results of tuning a cosine tapered filter with 16 plates starting from the folded configuration ($s = 0.5$). Band 1 is tuned towards shorter wavelengths, band 2 towards longer wavelengths. Computed theoretical values are given in brackets.

Modulation frequency s		T_{max}	Distance between the peaks in Å	$\Delta\lambda_{1/2}$ Å
0.50		0.997	0 (0)	41 (41)
0.44	band 1	1.008	63.5 (65)	40 (40)
	2	1.007		41 (41.5)
0.38	1	0.997	139 (139)	40 (39)
	2	1.002		44 (42)
0.25	1	0.995	303 (305)	38 (37.5)
	2	0.998		42 (43)
0.12	1	1.013	561 (564)	37 (36)
	2	1.004		42 (44)

Table 4 summarizes the results of shifting the peaks of a cosine tapered filter for some of the examples illustrated in Figure 4. Also in this case the agreement between the theoretical and experimental profiles is such that the discrepancies would not be noticable on the scale used in that figure. The largest deviation amounts to 1.3 % of the maximum transmission. An adjustable and tunable filter based on these principles is clearly feasible.

Acknowledgements

It is a pleasure for us to thank the people at the Utrecht Astronomical Institute workshop for the excellent design and construction of the filter holder. We are grateful for the possibility to use the optical facilities at the Technisch Physische Dienst in Delft for the experimental tests and for the generous computing facilities provided by the Leiden University computer centre. Parts of this work have been supported by the Swedish Space Research Committee.

APPENDIX 1 The transmission profile of a filter with a modulated angle difference distribution

The state of polarization of light passing through a pile of birefringent plates can be described in terms of the Jones (1941 a, b) matrix calculus (see also Shurcliff 1962). The light, as it emerges from a particular plate, is described by the Jones vector:

$$\begin{pmatrix} E_x + \Delta E_x \\ E_y + \Delta E_y \end{pmatrix} = \begin{pmatrix} e^{+j\gamma} & 0 \\ 0 & e^{-j\gamma} \end{pmatrix} \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} E_x \\ E_y \end{pmatrix} \tag{23}$$

E_x and E_y are here the elements in the Jones vector description of the polarization of the light before entering the plate. α is the angle difference between this plate and the preceding plate, and 2γ radians is the retardation of the plate at the wavelength in question (Eq. 5). We shall apply the Jones equation to Solc filters consisting of a large number of plates and in which all angle differences $\alpha_n \ll 1$ radian. Treating the pile of plates as a continuous birefringent medium in which the optic axis orientation ω is a function of the position l along the pile, we can from Eq. (23) derive the differential equations for the Jones vector elements E_x and E_y:

$$\begin{aligned} E_x' - j\,\kappa/2\; E_x &= +a\,E_y \\ E_y' + j\,\kappa/2\; E_y &= -a\,E_x \end{aligned} \tag{24}$$

E_x' and E_y' are derivatives with respect to l. $a\,(l)$ is the derivative $d\omega / dl$ of the optic axis angle; for a filter consisting of a pile of plates, $a\,(l)$ becomes an array of delta functions α_n. κ is the retardation per unit length of the birefringent material at the wavelength in question. A plate of thickness d has a retardation 2γ given by

$$2\gamma = \kappa \cdot d \qquad \text{radians of phase} \tag{25}$$

Let the light be 100 % polarized and its intensity be normalized to unity as it emerges from the first polarizer at $l = 0$. The Jones vector describing the light at this point is then

$$\begin{pmatrix} E_x\,(0) \\ E_y\,(0) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{26}$$

The end polarizer is oriented at right angles to the optic axis at $l = L$. The intensity of the light at the output of the filter is then given by

$$I = |E_y\,(L)|^2 \tag{27}$$

In order to obtain the transmission of the filter for unpolarized light, this normalized output intensity must be multiplied by the absorption loss factor $t/2$ (Section 2).

Eqs. (24) can be solved for some special cases. $a(l) = \text{const.}$ represents a 'continuous' fan type filter in which the optic axis rotates smoothly throughout the birefringent material. The solution

$$T = \frac{t}{2} \frac{a^2}{\beta^2} \sin^2 \beta L \tag{28}$$

where $\beta = [\,(\kappa/2)^2 + a^2\,]^{1/2}$

can also be derived from Eq. (3) or Eq. (11) as the limiting case for a pile of constant thickness L as the number of plates $N \to \infty$. A single peak with $t/2$ maximum transmission is obtained at $\kappa = 0$ when $a = \pi / 2L$, i.e. the optic axis has turned through one right angle between the input and the output of the filter

There will clearly be other special forms of $a(l)$ which can also be solved explicitly. However we are interested in the effect on the filter transmission of a modulation of the angle difference distribution as specified by Eq. (15). If the original optic axis derivative is $a(l)$, then the modulated version will have the derivative:

$$a_{mod} = 2\, a(l) \cos m l \tag{29}$$

where $l = n \cdot d$ and $m = 2\pi s / d$

Introduce the new variables P and Q by the substitution

$$\begin{aligned} E_x &= P \exp (+ j \kappa/2\ l) \\ E_y &= Q \exp (- j \kappa/2\ l) \end{aligned} \tag{30}$$

Eqs. (24) then take the form

$$\begin{aligned} P' &= + Q\ a \exp (- j \kappa\ l) \\ Q' &= - P\ a \exp (+ j \kappa\ l) \end{aligned} \tag{31}$$

The corresponding eaquations for the modulated version of the same filter become

$$\begin{aligned} P'_m &= + Q_m\ 2\ a \cos m l \exp (- j \kappa\ l) \\ Q'_m &= - P_m\ 2\ a \cos m l \exp (+ j \kappa\ l) \end{aligned} \tag{32}$$

The problem can now be stated as follows. Assuming that we know the solution to Eqs. (31), then what is the solution to Eqs. (32) for the modulated version ?

If $a(l) = 0$, then P_m and Q_m are constants all through the filter. We now make the assumption that $a(l)$ is so small, that the increments ΔP_m and ΔQ_m are always $\ll 1$ when calculated over an interval occupied by one full modulation cycle. We make the further assumption that the tapering function over the same interval can be approximated by its value at the centre of the interval. This essentially means that the angle differences α_n in the unmodulated version of the filter do not change significantly over this interval. These conditions will, in general, be satisfied for normally tapered filters which are modulated in such a way that there are many full cycles along the filter. Eqs. (32) can now be integrated over individual modulation cycles while treating P_m, Q_m and a as constants. The modulation cycle nr. k has its centre at

$$l_c = (k - \tfrac{1}{2})\, 2\pi / m \tag{33}$$

and the increments for the cycle centered at l_c become

$$\begin{aligned} \Delta P_m &= + Q_m \; a \; \frac{2\kappa}{\kappa^2 - m^2} \; \sin(\pi\kappa/m) \; \exp(-j\,\kappa\, l_c) \\ \Delta Q_m &= - P_m \; a \; \frac{2\kappa}{\kappa^2 - m^2} \; \sin(\pi\kappa/m) \; \exp(+j\,\kappa\, l_c) \end{aligned} \tag{34}$$

The same increments over the same intervals (or, for negative κ, their complex conjugates) are produced by the functions P and Q defined by

$$\begin{aligned} P' &= + Q \; a_m \; \exp[-j\,(|\kappa| - m)\, l] \\ Q' &= - P \; a_m \; \exp[+j\,(|\kappa| - m)\, l] \end{aligned} \tag{35}$$

where $a_m = a \cdot 2|\kappa| / (|\kappa| + m)$

Thus, under the assumptions discussed above, these equations can replace the original Eqs. (32). They have the same form as those for the unmodulated filter (31) but with κ replaced by $(|\kappa| - m)$. Consequently we shall now expect maxima at $\kappa = \pm m$ instead of, as before, at $\kappa = 0$. In the neighbourhood of these maxima we have $a_m \approx a$ which shows that the shifted bands will have the same shape as that of the unmodulated version. When the filter consists of a large number of identical plates, the double band pattern must repeat about the normal fan filter positions. In terms of the parameters γ and s we find that $\kappa = \pm m$ corresponds to

$$\gamma = integer \cdot \pi = s\pi \tag{36}$$

= (normal position of fan filter bands) = s (repetition interval)

APPENDIX 2 The influence of errors in the plate angle ω and the plate thickness d

When dealing with filters consisting of a large number of plates between which all angle differences $\alpha_n \ll 1$ radian, we can make some simplifying approximations for wavelengths (values of κ) well removed from the main transmission bands. Under these circumstances, the light will never depart significantly from a state of linear polarization parallel to the local optic axis direction. This becomes obvious if one traces, on the Poincaré sphere, the polarization of such a wave as it passes through the filter. Then $|Q(l)| \ll 1$ and we can use the approximation $P = 1$ since the deviations of P from this value will be of the order of $|Q(l)|^2$. Eq. (31) for Q' can now be integrated directly:

$$Q(L) = -\int_0^L \alpha(l) \exp(+j\kappa l)\, dl \tag{37}$$

If the filter contains N plates of thickness d_n, mounted with the angle difference distribution α_n

$$Q(L) = -\sum_{n=1}^{N+1} \alpha_n \exp(+j\kappa l_n) \tag{38}$$

where l_n is the position along the filter axis at which the corresponding angle difference occurs. The transmission of the filter is (Eqs. 27, 30):

$$\begin{aligned} T &= \frac{t}{2} |E_y(L)|^2 = \frac{t}{2} |Q(L)|^2 \\ &= \frac{t}{2} |Q_0(L) + Q_\epsilon(L)|^2 \end{aligned} \tag{39}$$

Q_0 is here the ideal value according to the design and Q_ϵ the error in this quantity due to errors in the plate angles or the plate thicknesses. We are concerned with wavelengths well removed from the main bands, and the transmission if significant, will be caused mainly by the filter errors. The parasitic transmission T_ϵ, expressed as a fraction of the maximum possible transmission, $t/2$, then becomes:

$$T_\epsilon \approx |Q_\epsilon(L)|^2 \tag{40}$$

Errors in the orientations of the plates

The error $\Delta\alpha_n$ in the angle difference α_n between plates number n and $(n-1)$ is clearly equal to the difference between the errors $\Delta\omega_n$ and $\Delta\omega_{n-1}$ in the actual plate angles. Hence, for a filter with errors only in the plate orientations, Eq. (38) gives

$$Q_\epsilon(L) = -\sum_{n=1}^{N+1} (\Delta\omega_n - \Delta\omega_{n-1}) \exp(+j\,2\gamma\,n)$$

$$= 2j \sin\gamma \sum_{n=1}^{N} \Delta\omega_n \exp(+j\,2\gamma\,[n-\tfrac{1}{2}]) \qquad (41)$$

Assuming that the errors $\Delta\omega_n$ are randomly distributed about zero with a standard deviation σ_ω radians, we can derive an expression for the average sidelobe level $T_\epsilon = |Q_\epsilon(L)|^2$

$$T_\epsilon = 4N\,\sigma_\omega^2 \sin^2\gamma \qquad (42)$$

Errors in the thickness of the plates

Consider a filter in which all plates are identical apart from plate number n_ϵ that has a small error Δd in its thickness. Compared with an ideal filter of the same actual total length L, all angle differences α_n (except the first and the last) will appear slightly displaced. The error Δl_n in the position of the angle difference number n is easily shown to be

$$\Delta l_n = \begin{cases} \Delta d \cdot (n-1)/N & (n \le n_\epsilon) \\ \Delta d \cdot (n-1)/N - \Delta d & (n > n_\epsilon) \end{cases} \qquad (43)$$

The actual position of the angle difference α_n becomes:

$$l_n = (n-1)\,L/N + \Delta l_n \qquad (44)$$

When this is substituted into Eq. (38) we can, after some algebra, derive the error Q_ϵ due to this one plate error. The answer depends in a complicated way on the distribution of the angle differences. For an untapered fan filter, i. e. one in which all angle differences are equal, we get.

$$Q_\epsilon(L) \approx \frac{\alpha \cdot \Delta\gamma}{\sin\gamma} \cdot \exp(j\ 2\gamma\ [n_\epsilon - \tfrac{1}{2}]) \qquad (45)$$

where 2γ is the retardation of a plate whose thickness is exactly L/N, and

$$2\Delta\gamma = \kappa\,\Delta d \ll 1 \qquad (46)$$

is the retardation error in the faulty plate caused by the thickness error Δd. We now turn to

a filter in which all the plates have small errors Δd_n. Adding the various individual error contributions statistically we get the mean sidelobe level. For a normally constructed untapered fan filter, $\alpha = \pi / 2\,(N + 1)$, and we get.

$$T_e \approx \frac{\pi^4}{4} \frac{{\sigma_0}^2}{N \sin^2 \gamma} \tag{47}$$

σ_0 is the standard deviation of the plates from their average thickness expressed in orders of birefringence; this is related to the standard deviation σ_d in the plate thicknesses from their average value $d = L / N$ by·

$$2\pi\sigma_0 = 2\gamma\,(\sigma_d / d) \tag{48}$$

References:

Beckers, J. M. and Dunn, R. B. 1965, Air Force Camb. Res. Lab. Instrumentation papers No. 75

Evans, J. W. 1949, J. O. S. A. *39*, 229

Evans, J. W. 1958, J. O. S. A. *48*, 142

Fredga, K. 1969a, Appl. Optics *8*, 333

Fredga, K. 1969b, Solar Physics *9*, 358

Fredga, K. and Högbom, J. A. 1970, Proc. IAU Symp. No. 41, New Techniques in Space Astronomy, ed. R. Lüst (in press)

Giovanelli, R. G. and Jefferies, J. T. 1954, Aust. J. Phys. *7*, 254

Harris, S. E., Ammann, E. O. and Chang, I. C. 1964, J. O. S. A. *54*, 1267

Jones, R. C. 1941a, J. O. S. A. *31*, 488

Jones, R. C. 1941b, J. O. S. A. *31*, 500

Lyot, B. 1933, Compt. Rend. Acad. Sci. *197*, 1593

Lyot, B. 1944, Ann. Astrophys. *7*, 31

Öhman, Y. 1938, Nature *141*, 157 and *141*, 291

Shurcliff, W. A. 1962, Polarized Light, Harvard University Press.

Solc, I. 1953, Cesk. Casopis Fys. *3*, 366

Solc, I. 1959, Czech. J. Phys. *9*, 237

Solc, I. 1960, Cesk. Casopis Fys. *10*, 16

Solc, I. 1965, J. O. S. A. *55*, 621

Steel, W. H., Smartt, R. N. and Giovanelli, R. G. 1961, Aust. J. Phys. *14*, 201

Reprinted from JOURNAL OF THE OPTICAL SOCIETY OF AMERICA, Vol. 54, No. 10, 1267–1279, October 1964
Printed in U. S. A.

Optical Network Synthesis Using Birefringent Crystals.* I. Synthesis of Lossless Networks of Equal-Length Crystals

S. E. HARRIS,
Department of Electrical Engineering, Stanford University, Stanford, California
AND
E. O. AMMANN,
Sylvania Electronic Systems, Mountain View, California
AND
I. C. CHANG
W. W. Hansen Laboratories, Stanford University, Stanford, California

(Received 7 May 1964)

A procedure for the synthesis of birefringent networks having arbitrarily prescribed transfer functions is presented. The basic network configuration consists of n identical cascaded birefringent crystals between an input and an output polarizer. The crystals are cut with their optic axes perpendicular to their length. The variables determined by the synthesis procedure are the angles of the optic axes of the crystals and the angle of the output polarizer. Any transfer function which is periodic with frequency and whose corresponding impulse response is real and causal can, in theory, be realized. A network of n crystals allows the approximation of a desired function by $(n+1)$ terms of a Fourier exponential series. Bandwidths of less than 1 Å appear possible.

I. INTRODUCTION

THE advent of the laser has made possible various types of optical systems. This has produced a need for optical elements or networks whose transfer functions can be arbitrarily prescribed as a function of frequency. In a manner analogous to that used at radio frequencies, such optical networks could be utilized as discriminators and ratio detectors, equalizers and compensators, frequency selective hybrids, and delay networks, to name just a few. Of particular importance is the possibility of realizing very narrow-band filters having prescribed transmission characteristics.

The purpose of this paper is to present a basic network configuration and synthesis procedure whereby optical networks having arbitrary transfer functions can be constructed using a set of cascaded birefringent crystals. Although synthesis procedures exist for other types of optical devices,[1–4] the very narrow bandwidths and tunability of birefringent devices make them particularly attractive for the above-mentioned applications. The type of network to be considered is shown in Fig. 1. In simplest form, it consists of a number of identical birefringent crystals placed between two polarizers. Although Fig. 1 pictures a network containing four stages (four birefringent crystals), any number can be used. In principle, either uniaxial or biaxial[5] crystals may be employed, but for simplicity we will assume uniaxial crystals are used. Each crystal is cut with its optic axis perpendicular to its length and with end faces which are flat and parallel. The S's and F's in Fig. 1 denote the crystals' "slow" and "fast" axes, respectively. If a negative crystal is used, the fast axis will be the optic axis, while for a positive crystal the slow axis will be the optic axis. The variables to be determined by the synthesis procedure are the angles to which the crystals are rotated, the angle of the output polarizer, and the length L of the crystals used. In the following sections, we will show that by properly choosing these variables, it is possible, in theory, to synthesize any desired transfer function, subject only to the restrictions that it be periodic with frequency and that it satisfy the usual requirements imposed by the necessity for a real and causal impulse response. The basic periodicity of the network response is determined by the type and length of birefringent crystals used. For example, if calcite crystals 1 cm in length are used, the basic period of

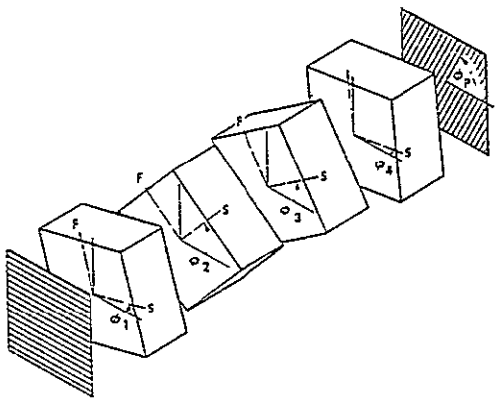


FIG. 1. Basic configuration of optical network (four stages). Polarizers are shown shaded.

* This work was supported at Stanford University by the Space Systems Division of the U. S. Air Force Systems Command under Contract Number AF 04(695)-305 and at Sylvania by the Air Force Avionics Laboratory at Wright-Patterson Air Force Base, Ohio, under Contract AF 33(657)-8995.

[1] H. Pohlack, Jenaer Jahrbuch, 1962, p. 181 (in German).
[2] L. Young, J. Opt. Soc. Am. 51, 967 (1961).
[3] J. S. Seeley, Proc. Phys. Soc. (London) 78, 998 (1961).
[4] R. J. Pegis, J. Opt. Soc. Am. 51, 1255 (1961).
[5] If biaxial crystals are used, crystals in the monoclinic and triclinic systems will probably not be satisfactory since the directions of their principal axes are dependent upon temperature and wavelength.

the response will be about 175 Gc (about 2 Å in the red).

An important modification of the basic configuration of Fig. 1 is the addition of a variable optical compensator[6] before or after each birefringent crystal. The compensators allow one to tune the network transfer function without distortion over its basic period and, in addition, compensate for slightly incorrect crystal lengths.

The optical network described here is a lossless or nondissipative network in that it does not contain any internal polarizers; if the final polarizer is nonabsorbing, e.g., a Rochon prism, then all of the optical energy incident on the first birefringent crystal is, in principle, available at the network output. It is planned to consider the synthesis of dissipative birefringent networks, i.e., networks containing internal polarizers, in a following paper.

A central idea of this paper is the consideration of the impulse response of a system of birefringent crystals. This approach was used by Mertz[7] to analyze the Solc birefringent filter and was independently suggested as an approach to the synthesis problem by Harris.[8] It is first presented and then used to obtain an exact synthesis procedure. The question of tunability is considered and an example given.

II. HISTORY OF BIREFRINGENT DEVICES

Before proceeding further, it is appropriate to note that two birefringent filters having particular transfer functions have been proposed considerably earlier. The first of these was proposed in 1933 by the French astronomer, Lyot,[9] who suggested a birefringent filter consisting of alternating polarizers and birefringent crystals. The length of each crystal is twice that of the preceding crystal. A four-stage Lyot filter is shown in Fig. 2. The transfer function of the Lyot filter has the form $\sin x/x$, repeated at periodic intervals. More recently, Solc[10] proposed two types of birefringent filters, termed fan and folded filters. Figure 3 shows a four-stage Solc fan filter. These filters have the same structural form as our basic network. In the Solc filters, however, the relative rotation angle between each successive crystal is related in a simple manner to the number of birefringent crystals employed. In contrast, the relative angles of the crystals in our network are determined by the choice of optical transfer function—which may be arbitrary. Complete discussions of both the Lyot and Solc filters have been given by Evans.[11,12]

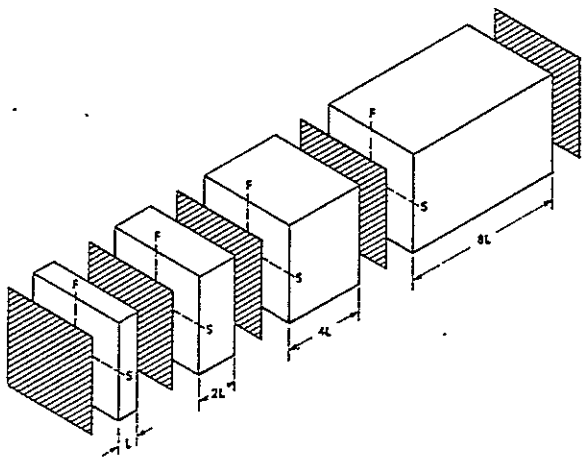


FIG. 2. Four-stage Lyot filter. Polarizers are shown shaded.

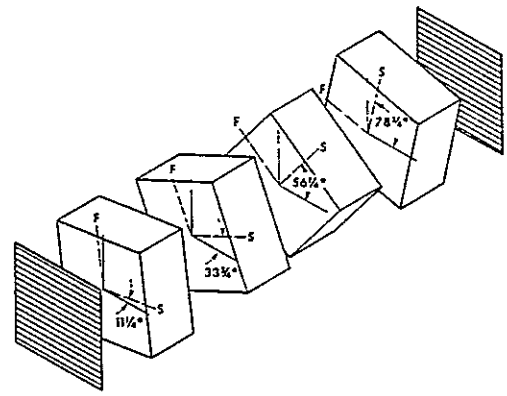


FIG. 3. Four-stage Solc fan filter.

Numerous Lyot and Solc filters have been built and operated.[11,13–18] These filters are used primarily in astronomy where their very narrow bandwidths are utilized to observe solar prominences. Recently, Steel *et al.*[17] have constructed a Lyot filter with a bandwidth of $\frac{1}{8}$ Å in the red. By using the synthesis techniques proposed in this paper, it should be possible to attain similar bandwidths with prescribed transmission characteristics.

III. GENERAL CONSIDERATIONS

A. Impulse Response of a Series of Birefringent Crystals

Analysis by means of impulse response is a concept that is familiar to electrical engineers.[19] If an impulse, i.e., a Dirac delta function in time is applied to a

[6] H. G. Jerrard, J. Opt. Soc. Am. 38, 35 (1948).
[7] L. Mertz, J. Opt. Soc. Am. 50 (June 1960) (advertisement facing p. xii).
[8] S. E. Harris and E. O. Ammann, Proc. IEEE 52, 411 (1964).
[9] B. Lyot, Compt. Rend. 197, 1593 (1933).
[10] I. Solc, Czech. J. Phys. 3, 366 (1953); 4, 607, 669 (1954); 5, 114 (1955).
[11] J. W. Evans, J. Opt. Soc. Am. 39, 229 (1949).
[12] J. W. Evans, J. Opt. Soc. Am. 48, 142 (1958).
[13] Y. Ohman, Nature 141, 157 (1938); Nature 141, 291 (1938); Pop. Astron. Tidskrift, No. 1–2, 11, 27 (1938).
[14] J. W. Evans, Publ. Astron. Soc. Pacific 52, 305 (1940).
[15] J. W. Evans, Ciencia Invest. (Buenos Aires) 3, 365 (1947).
[16] B. H. Billings, J. Opt. Soc. Am. 37, 738 (1947).
[17] W. H. Steel, R. N. Smartt, and R. G. Giovanelli, Australian J. Phys. 14, 201 (1961).
[18] J. W. Evans, Appl. Opt. 2, 193 (1963).
[19] J. A. Aseltine, *Transform Method in Linear System Analysis* (McGraw-Hill Book Company, Inc., New York, 1958).

198

C. 5

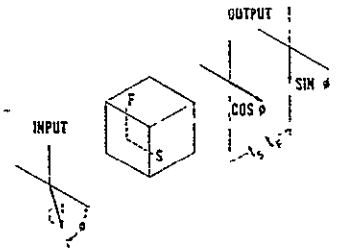


FIG. 4. Impulse response of a single birefringent crystal.

linear network, the Fourier transform of the impulse response of the network is the frequency domain transfer function of the network.

We first consider the impulse response of the single birefringent crystal of Fig. 4. The crystal is cut with its optic axis perpendicular to its length and with end faces flat and parallel. A linearly polarized impulse of optical electric field is assumed to be normally incident on the crystal. Since the incoming signal is normally incident, double refraction will not occur. The impulse will divide into orthogonally polarized ordinary and extraordinary impulses whose amplitudes are dependent on the polarization of the incident impulse with respect to the principal axes of the birefringent crystal. These impulses travel with different velocities, therefore emerging at different times. The difference in the times at which they emerge from the crystal is given by

$$t_S - t_F = L\Delta\eta/c, \tag{1}$$

where $\Delta\eta$ is the difference between the extraordinary and ordinary indices of refraction of the crystal, L is the crystal length, and c is the velocity of light in a vacuum.

We assume here that $\Delta\eta$ is a constant independent of frequency. This is not the actual situation, however, for $\Delta\eta$ will be a function of frequency, at least to some degree. The birefringence of calcite, for example, varies approximately 11% between 4000 and 8000 Å. The effect of the dispersion of $\Delta\eta$ has been ignored in this paper for two reasons. First, to include its effect would greatly complicate the synthesis procedure and obscure the basic ideas. Second, the effects of dispersion upon the resulting transfer function will generally be small, particularly if the synthesized network has a small bandwidth. Existing analyses of the Lyot and Solc filters have also neglected dispersion; yet experimental results have agreed quite well with theory.

Thus, the impulse response of a single birefringent crystal is two orthogonally polarized impulses whose amplitudes depend upon ϕ, the angle between the principal axes of the crystal and the incident optical polarization. If ϕ is equal to zero, all of the light will emerge at time t_S; if ϕ is equal to 45°, the light will emerge as two equal impulses at times t_F and t_S.

We next consider the impulse response of several cascaded birefringent crystals having arbitrary lengths and orientations, as shown schematically in Fig. 5. This figure contains information about the time of emission of the impulses, but none about their polarizations. First consider the case of two crystals. The output of the first crystal is, in general, two orthogonally polarized impulses. Each of these impulses is incident on the second crystal and produces two more impulses. Thus, in general, the impulse response of two cascaded birefringent crystals is four impulses, two of which are polarized along the fast axis and two along the slow axis of the second crystal. With more crystals this process continues, giving us the result that the impulse response of n birefringent crystals having arbitrary lengths and orientations is a set of 2^n impulses. The magnitudes and polarizations of these impulses are determined by the crystal angles, while their relative times of emergence from the crystal are determined by the birefringence and lengths of the crystals used. Thus we reach the important conclusion that the impulse response of a series of birefringent crystals is a train of impulses of finite duration. In contrast, the impulse responses of Fabry–Perot and multilayer dielectric-film filters consist of infinite trains of impulses.

Now suppose that all of the n crystals are chosen to be identical, i.e., the same material and equal lengths. The output will now consist of only $(n+1)$ rather than 2^n impulses. Furthermore, the emerging impulses will be equally spaced in time. The reason that fewer impulses emerge when the crystals are chosen of equal length is seen by examining the two-crystal case. For two crystals of equal length, the impulse which travels along the fast axis of the first and the slow axis of the second will emerge at the same time as the impulse which travels along the slow axis of the first and the fast axis of the second. These two combine, and the output, therefore, consists of three rather than four impulses.

Thus we are led to the network configuration of Fig. 1. The basic idea of the synthesis procedure is to utilize the relative angles of n birefringent crystals and one output polarizer to control the amplitudes of $(n+1)$ equally spaced output impulses. The first step of the procedure is to specify the desired impulse amplitudes at the output of the final polarizer. These amplitudes may be selected arbitrarily as is seen in the following section. We then use a systematic procedure to arrive at angles for the network elements so this final set of impulses is obtained from a single impulse incident on the first crystal of the network. This is equivalent to saying, of course, that the desired set of impulses is the impulse response of the network.

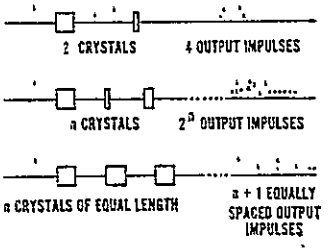


FIG. 5. Impulse response of several birefringent crystals.

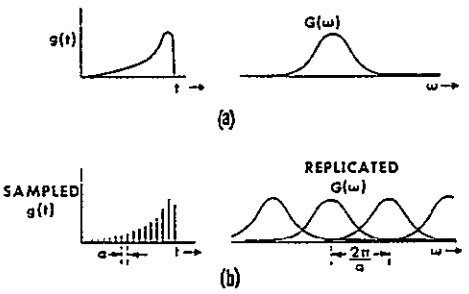


FIG. 6. Impulse responses and corresponding transfer functions for a network whose impulse response is (a) $g(t)$, and (b) $g(t)$ sampled.

B. Specifying the Desired Response

Let us now consider what types of responses we can realize and how we specify them. As in lumped-element circuit theory, a convenient approach is to first choose an ideal response and then approximate this to the necessary degree.

We should note that the frequency transfer function of the optical network must be periodic. This can be seen readily from Fourier theory or sampling theory. Suppose that a network has an impulse response $g(t)$ and a corresponding transfer function $G(\omega)$, where both $g(t)$ and $G(\omega)$ are continuous and aperiodic as shown in Fig. 6(a). Next, suppose another network has an impulse response which is $g(t)$ sampled at a uniform rate of $1/a$ samples/sec. This is the case for a network consisting of a set of birefringent crystals, each of whose length is such that $t_S - t_F$ of Eq. (1) equals a seconds. This network will have a periodic transfer function like that shown in Fig. 6(b), which is the original $G(\omega)$ replicated with a period of $2\pi/a$ rad/sec.[20] Figure 7 shows the transfer function periodicity that can be obtained using readily available lengths of some common crystals.

Assume that a desired periodic transfer function $G(\omega)$ has been chosen. The next step is to find a satisfactory approximation to $G(\omega)$ which can be realized using the optical network of Fig. 1. The approximation is made by an exponential series containing a finite number of terms.

$$C(\omega)=C_0+C_1e^{-ia\omega}+C_2e^{-i2a\omega}+\cdots+C_ne^{-ina\omega}$$
$$=\sum_{k=0}^{n}C_ke^{-ika\omega}. \quad (2)$$

The impulse response corresponding to Eq. (2) is found by taking the inverse Fourier transform, giving

$$C(t)=C_0\delta(t)+C_1\delta(t-a)+C_2\delta(t-2a)+\cdots$$
$$+C_n\delta(t-na) \quad (3)$$
$$=\sum_{k=0}^{n}C_k\delta(t-ka).$$

[20] E. A. Guillemin, *Theory of Linear Physical Systems* (John Wiley & Sons, Inc., New York, 1963), p. 430.

Thus it is clear why an exponential series is used to approximate the desired transfer function. The exponential series has a Fourier transform consisting of uniformly spaced impulses, and this is the form of the impulse response of our optical network. If there are $n+1$ terms in $C(\omega)$ [as there are in Eq. (2)], an n-stage optical network is required.

There are various methods available for finding the C_i of Eq. (3) from a given $G(\omega)$. One obvious possibility is to choose the C_i to be the Fourier coefficients of the series. However, if the desired $G(\omega)$ contains discontinuities, some other approximation such as a Cesàro approximation may well be more desirable. Such topics have been treated in detail elsewhere,[21] so we will not discuss this problem further.

It is likely that $|G(\omega)|^2$ or arg $G(\omega)$ will sometimes be given instead of $G(\omega)$. It will then be necessary to approximate $|G(\omega)|^2$ or arg $G(\omega)$ in a suitable manner and calculate $C(\omega)$ from this.

Two points should be noted concerning the approximating functions $C(\omega)$ and $C(t)$. First, since the impulse response of a physical network must be real, the real and imaginary parts of $C(\omega)$ must be even and odd functions of frequency, respectively. This means that all C_i must be real. Second, it is *not* necessary that $C(\omega)$ and $C(t)$ be causal. While it is true, of course, that the impulse response of a network must be zero for $t<0$, we are free to shift our time scale to a new origin when writing $C(\omega)$ and $C(t)$ if this will be more convenient. Thus, in writing Eqs. (2) and (3), we have neglected most of the uniform time delay associated with the network, i.e., the time delay accumulated by passage of the signal through each crystal, in the space between crystals, and in transit to the point of detection. We have chosen our new time origin to be the time at which the first output impulse occurs. For this choice of origin $C(\omega)$ is causal, but equally well, we could have chosen a time origin which results in a noncausal $C(\omega)$. As far as the synthesis procedure is concerned, the important point is that only the relative positions in time of the various impulses are important. In this paper, we will always

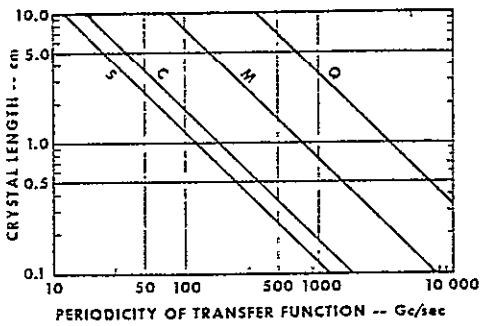


FIG. 7. Periodicity of network response for several types of birefringent crystals. Q: quartz, $\Delta n=0.009$; M: mica, $\Delta n=0.04$; C: calcite, $\Delta n=0.17$; S: sodium nitrate, $\Delta n=0.24$.

[21] Ref. 20, p. 408.

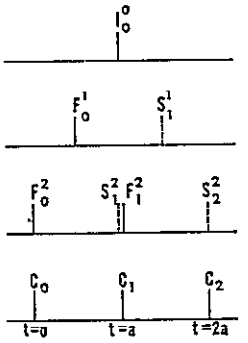


FIG. 8. Summary of impulse notation: Impulse pyramid for a two-stage network. Top: input; next to top: output from first crystal; next to bottom: output from second crystal; bottom: output from polarizer. Solid strokes: polarized along fast axis of crystal. Broken strokes: polarized along slow axis of crystal.

choose our time origin to be synonymous with the occurrence of the first impulse of the train.

The number of birefringent crystals that are necessary to synthesize a desired function will depend on the nature of the function and on the closeness of the approximation desired. Many applications of the synthesis procedure to problems of optical communications will require functions which do not possess discontinuities and whose width is equal to their basic periodicity. (One such function is the triangular waveform of Fig. 11 which might be used to convert a frequency-modulated light signal to an amplitude-modulated light signal.) For functions of this type, the first five or six terms of an exponential series (and, therefore, four or five birefringent crystals) will generally yield a satisfactory approximation.

For narrow-band filter applications, it is necessary to synthesize transfer functions whose basic periodicity is considerably wider than their width. An estimate of the number of crystals necessary for this case may be obtained from sampling considerations and can be written

$$\text{Number of crystals necessary} \approx q\,\frac{\text{periodicity}}{\text{bandwidth}}, \tag{4}$$

where q is an integer which generally will be between 2 and 7. This statement can be understood by noting that the length of the time-domain impulse response is approximately related by the reciprocal width property of Fourier transforms to the bandwidth of the transfer function, and may be written as q/bandwidth, where q is the aforementioned integer. By the length of the impulse response, we mean the time between the first and the last impulses which have significant amplitude. The number of necessary impulses is then the length of the impulse train divided by the spacing between impulses, plus one. Since the spacing between impulses is the reciprocal of the periodicity, and since the number of necessary birefringent crystals is one less than the necessary number of impulses, Eq. (4) follows. The integer q will depend on the function chosen, the degree of approximation desired, and on the definition of bandwidth. As an example, $q=2$ if the desired function is $\sin x/x$ and bandwidth is defined as the number of cycles between its first zeros.

IV. SYNTHESIS PROCEDURE

The object of the synthesis procedure is to find the n birefringent crystal angles and the output polarizer angle which give the desired transfer function $C(\omega)$. The C_i of Eqs. (2) and (3) can have any value, provided that each is real.

A. Notation

The notation and conventions used in the synthesis procedure are discussed here. We refer repeatedly to Fig. 1 which pictures the basic optical network.

Rather than dealing with the φ's of Fig. 1, it is more convenient to solve for the relative angles (additional angles of rotation measured from the preceding component) of the crystals and output polarizer. Therefore, we define

$$\begin{aligned} \theta_1&=\phi_1,\\ \theta_2&=\phi_2-\phi_1,\\ &\vdots\\ \theta_n&=\phi_n-\phi_{n-1},\\ \theta_p&=\phi_p-\phi_n. \end{aligned} \tag{5}$$

The magnitudes of the impulses composing the impulse train emitted from the network are denoted by the C_i of Eqs. (2) and (3). It is also necessary to describe quantitatively the impulse trains which occur between the various stages within the network. In describing them, we must convey information about the polarization of the impulse train, as well as about the magnitudes of the individual impulses. For although we know that $C(t)$ is polarized parallel to the transmission axis of the output polarizer, the impulse train which leaves one of the birefringent crystals on its way toward the output has components polarized parallel to both the S and F axes of that crystal. This points up a fundamental difference between the synthesis procedure described here and conventional synthesis procedures in other fields. Namely, we must be concerned with not only the time variation of the signal, but also with its polarization as it passes through the network.

We illustrate the impulse notation with the aid of the "impulse pyramid" of Fig. 8. Suppose a single impulse (polarized parallel to the transmission axis of the input polarizer) is incident upon a network consisting of two birefringent crystals plus an output polarizer. The resulting output from the second birefringent crystal contains components polarized in both the S and F directions of that crystal.

$$F^2(t)=F_0^2\delta(t)+F_1^2\delta(t-a), \tag{6a}$$

$$S^2(t)=S_1^2\delta(t-a)+S_2^2\delta(t-2a). \tag{6b}$$

S denotes that an impulse is emitted polarized parallel to the slow axis of the crystal, while F denotes polarization parallel to the fast axis. In Fig. 8, slow-axis and fast-axis polarizations are denoted by dotted and solid lines, respectively. The superscript 2 means that we are

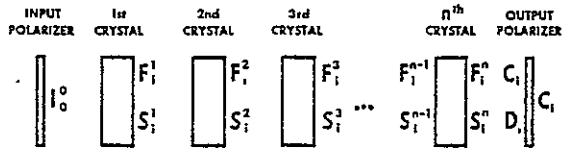


FIG. 9. n-stage network. Compare with two-stage network in Fig. 8.

dealing with the output from the second crystal of the network. The subscripts give the time of occurrence of the impulses. The first impulse, emitted at $t=0$, has the subscript 0; the next two impulses, emitted at $t=a$, have the subscript 1; and so on. Notice, in particular, that S_0^2 and F_2^2 are zero.

Since the impulses are evenly spaced in time, it is not necessary henceforth to write the delta functions when describing an impulse train. All the information of Eqs. (6) is given when F_0^2, F_1^2, S_1^2, and S_2^2 are stated.

As noted earlier, the desired transfer function and corresponding set of impulses are denoted by $C(\omega)$ and C_t, respectively. There is also an orthogonally polarized component which is stopped by the output polarizer. This signal and its corresponding set of impulses is denoted by $D(\omega)$ and D_t. Finally, the area of the impulse incident on the first crystal of the network is denoted by I_0^0. The notation is further summarized in Fig. 9.

B. Procedure

At the outset, two points should be stressed. First, it is assumed that the birefringent crystals of the network are lossless. This means that at all points between the input and output polarizers, energy must be conserved. Energy conservation places certain important restrictions on the F_t and S_t which are derived and listed in Appendix B. Secondly, it should be noted that $F_t^t=S_0^t=0$. This is just a statement of the fact that the first and last impulses out of the ith crystal must have propagated along its fast and slow axes, respectively.

We begin by assuming that $C(\omega)$ and, therefore, the desired C_t of Eqs. (2) and (3) have been chosen. We must next find the orthogonal signal, i.e., the signal $D(\omega)$ that is stopped by the output polarizer. By conservation of energy, we have

$$D(\omega)D^*(\omega)=(I_0^0)^2-C(\omega)C^*(\omega). \tag{7}$$

The left side of this equation must be non-negative for all frequencies and, therefore, for the equation to be valid, $(I_0^0)^2$ must be chosen greater than the maximum value of $C(\omega)C^*(\omega)$. As long as $(I_0^0)^2$ exceeds this value, its choice is arbitrary. However, it will generally be desirable to choose $(I_0^0)^2$ equal to the maximum value of $C(\omega)C^*(\omega)$, since this insures 100% transmission at the frequency at which this maximum occurs. Appendix A shows one method for calculating $D(\omega)$ from $D(\omega)D^*(\omega)$. It is also shown in Appendix A that as long as $(I_0^0)^2$ is chosen sufficiently large, at least one real set of D_t can always be found. Once $(I_0^0)^2$ has been chosen, $D(\omega)$ is calculated and written in the form

$$D(\omega)=D_0+D_1e^{-i\omega a}+D_2e^{-i2a\omega}+\cdots+D_ne^{-ina\omega}. \tag{8}$$

The corresponding orthogonal impulse response is then

$$D(t)=D_0+D_1\delta(t-a)+D_2\delta(t-2a)+\cdots+D_n\delta(t-na). \tag{9}$$

With the C_t and D_t specified, we now have a complete description of the input to the output polarizer. This, of course, is also the output from the last (nth) crystal. It is convenient here to transform this output into the principal axis system of the final crystal. With the help of Fig. 10(a), we have

$$\begin{bmatrix}F_t^n\\ S_t^n\end{bmatrix}=\begin{bmatrix}\sin\theta_p & -\cos\theta_p\\ \cos\theta_p & \sin\theta_p\end{bmatrix}\begin{bmatrix}C_t\\ D_t\end{bmatrix}, \tag{10}$$

where θ_p is the relative angle of the output polarizer.

As mentioned earlier, a requirement is that

$$F_n^n=S_0^n=0. \tag{11}$$

Using Eq. (10), we see that Eq. (11) will be satisfied if

$$\tan\theta_p=D_n/C_n \tag{12a}$$

and

$$\tan\theta_p=-C_0/D_0. \tag{12b}$$

In order for Eqs. (12a) and (12b) to be satisfied simultaneously, it must be true that $C_0C_n+D_0D_n=0$. But we know this is satisfied from conservation of energy, since it is Eq. (B13) of Appendix B.

Thus by using either Eq. (12a) or (12b), the angle of the final polarizer is determined. Then, substituting this calculated value of θ_p into Eq. (10), we obtain F_t^n and S_t^n, the outputs along the fast and slow axes of the last crystal. We now must find the rotation angles of the n crystals.

To accomplish this, we first find expressions relating the input and output of each crystal. This is a matter of taking projections along S and F axes of the crystals. With the help of Figs. 10(b) and 10(c), we find that

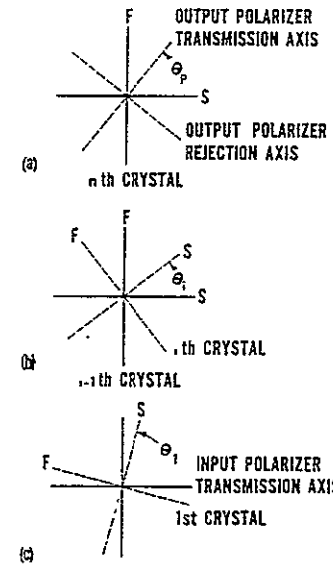


FIG. 10. Angle conventions used in the synthesis procedure: (a) output polarizer; (b) relative crystal angles; (c) input polarizer.

202

First Crystal

$$\begin{bmatrix} F_0^1 \\ S_1^1 \end{bmatrix} = \begin{bmatrix} -\sin\theta_1 \\ \cos\theta_1 \end{bmatrix} [I_0^0], \tag{13a}$$

Second Crystal

$$\begin{bmatrix} F_0^2 \\ F_1^2 \\ S_1^2 \\ S_2^2 \end{bmatrix} = \begin{bmatrix} \cos\theta_2 & 0 \\ 0 & -\sin\theta_2 \\ \sin\theta_2 & 0 \\ 0 & \cos\theta_2 \end{bmatrix} \begin{bmatrix} F_0^1 \\ S_1^1 \end{bmatrix}, \tag{13b}$$

Third Crystal

$$\begin{bmatrix} F_0^3 \\ F_1^3 \\ F_2^3 \\ S_1^3 \\ S_2^3 \\ S_3^3 \end{bmatrix} = \begin{bmatrix} \cos\theta_3 & 0 & 0 & 0 \\ 0 & \cos\theta_3 & -\sin\theta_3 & 0 \\ 0 & 0 & 0 & -\sin\theta_3 \\ \sin\theta_3 & 0 & 0 & 0 \\ 0 & \sin\theta_3 & \cos\theta_3 & 0 \\ 0 & 0 & 0 & \cos\theta_3 \end{bmatrix} \begin{bmatrix} F_0^2 \\ F_1^2 \\ S_1^2 \\ S_2^2 \end{bmatrix}. \tag{13c}$$

From the pattern established, we can write for the ith crystal

ith Crystal

$$\begin{bmatrix} F_0^i \\ F_1^i \\ F_2^i \\ \cdots \\ \cdots \\ \cdots \\ F_{i-3}^i \\ F_{i-2}^i \\ F_{i-1}^i \\ S_1^i \\ S_2^i \\ S_3^i \\ \cdots \\ \cdots \\ \cdots \\ S_{i-2}^i \\ S_{i-1}^i \\ S_i^i \end{bmatrix} = \begin{bmatrix} \cos\theta_i & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \cos\theta_i & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \cos\theta_i & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & -\sin\theta_i & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\sin\theta_i & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & -\sin\theta_i \\ \sin\theta_i & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \sin\theta_i & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \sin\theta_i & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \cos\theta_i & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \cos\theta_i & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cos\theta_i \end{bmatrix} \begin{bmatrix} F_0^{i-1} \\ F_1^{i-1} \\ F_2^{i-1} \\ \cdots \\ \cdots \\ \cdots \\ F_{i-3}^{i-1} \\ F_{i-2}^{i-1} \\ S_1^{i-1} \\ S_2^{i-1} \\ S_3^{i-1} \\ S_4^{i-1} \\ \cdots \\ \cdots \\ \cdots \\ S_{i-2}^{i-1} \\ S_{i-1}^{i-1} \end{bmatrix}. \tag{13d}$$

Our procedure is to start with the output from the last crystal. From these F_i^n and S_i^n, we calculate the crystal angle and the input to the crystal (the F_i^{n-1} and S_i^{n-1}). Since the input to the nth stage is the output from the $(n-1)$th stage, we can repeat the entire process for the $(n-1)$th crystal. Thus we work our way back through the entire network alternately finding crystal angles and crystal inputs.

The calculation of the angles and inputs is accomplished as follows: Consider, for example, Eq. (13c) which relates the input and output of the third crystal. We know the output (the F_i^3 and S_i^3) and wish to find θ_3 and the input (the F_i^2 and S_i^2). In the language of linear equation theory, the problem may be restated as, "Does the system of nonhomogeneous equations (13c) have a solution?"

A set of nonhomogeneous equations has a solution if and only if the rank of the matrix of the coefficients is equal to the rank of the augmented matrix.[22] For Eqs. (13c), this means that a solution exists if the rank of the coefficient matrix

$$\begin{bmatrix} \cos\theta_3 & 0 & 0 & 0 \\ 0 & \cos\theta_3 & -\sin\theta_3 & 0 \\ 0 & 0 & 0 & -\sin\theta_3 \\ \sin\theta_3 & 0 & 0 & 0 \\ 0 & \sin\theta_3 & \cos\theta_3 & 0 \\ 0 & 0 & 0 & \cos\theta_3 \end{bmatrix}$$

equals the rank of the augmented matrix

$$\begin{bmatrix} \cos\theta_3 & 0 & 0 & 0 & F_0^3 \\ 0 & \cos\theta_3 & -\sin\theta_3 & 0 & F_1^3 \\ 0 & 0 & 0 & -\sin\theta_3 & F_2^3 \\ \sin\theta_3 & 0 & 0 & 0 & S_1^3 \\ 0 & \sin\theta_3 & \cos\theta_3 & 0 & S_2^3 \\ 0 & 0 & 0 & \cos\theta_3 & S_3^3 \end{bmatrix}.$$

Since the rank of the coefficient matrix is 4, the rank of the augmented matrix must also be 4 for a solution to exist. Several procedures exist[22] for determining the rank of a matrix. Applying one of these, we find the rank of the augmented matrix to be 4 if

$$\tan\theta_3 = -(F_2^3/S_3^3) \tag{14a}$$

and

$$F_0^3 F_2^3 + S_1^3 S_3^3 = 0. \tag{14b}$$

[22] D. C. Murdoch, *Linear Algebra for Undergraduates* (John Wiley & Sons, Inc., New York, 1947), p. 50–51.

TABLE I. Related sets of D_i and their corresponding θ_i.

	Solutions for D_i					Corresponding crystal and polarizer angles			
	1st set	2nd set	3rd set	4th set		1st set	2nd set	3rd set	4th set
D_0	Δ_0	$-\Delta_0$	Δ_n	$-\Delta_n$	θ_1	Θ_1	$-\Theta_1$	Θ_p	$-\Theta_p$
D_1	Δ_1	$-\Delta_1$	Δ_{n-1}	$-\Delta_{n-1}$	θ_2	Θ_2	$-\Theta_2$	Θ_n	$-\Theta_n$
D_2	Δ_2	$-\Delta_2$	Δ_{n-2}	$-\Delta_{n-2}$	θ_3	Θ_3	$-\Theta_3$	Θ_{n-1}	$-\Theta_{n-1}$
D_3	Δ_3	$-\Delta_3$	Δ_{n-3}	$-\Delta_{n-3}$	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	θ_{n-1}	Θ_{n-1}	$-\Theta_{n-1}$	Θ_3	$-\Theta_3$
D_{n-1}	Δ_{n-1}	$-\Delta_{n-1}$	Δ_1	$-\Delta_1$	θ_n	Θ_n	$-\Theta_n$	Θ_2	$-\Theta_2$
D_n	Δ_n	$-\Delta_n$	Δ_0	$-\Delta_0$	θ_p	Θ_p	$-\Theta_p$	Θ_1	$-\Theta_1$

The first equation gives the angle of the crystal. Using this angle in Eq. (13c), we can now calculate the input (F_0^2, F_1^2, S_1^2, and S_2^2). The calculation is an easy one, involving for any stage no worse than the solution of two simultaneous equations. Appendix C shows a systematic method of performing this calculation. Equation (14b) is seen by comparison with Eq. (B9) to be simply a restatement of the fact that the F_i^3 and S_i^3 must satisfy conservation of energy. This requirement is automatically satisfied by the F_i and S_i of *all* stages since $D(\omega)$ was calculated using conservation of energy.

In a similar manner, the conditions for existence of solutions to Eqs. (13a), (13b), and (13d) result in the equations:

First Crystal

$$\tan\theta_1 = -(F_0^1/S_1^1), \tag{15a}$$

$$(F_0^1)^2+(S_1^1)^2=(I_0^0)^2, \tag{15b}$$

Second Crystal

$$\tan\theta_2 = -(F_1^2/S_2^2), \tag{16a}$$

$$F_0^2F_1^2+S_1^2S_2^2=0, \tag{16b}$$

ith Crystal

$$\tan\theta_i = -(F_{i-1}^i/S_i^i), \tag{17a}$$

$$F_0^iF_{i-1}^i+S_1^iS_i^i=0. \tag{17b}$$

The crystal angles are given by Eqs. (15a), (16a), and (17a), while Eqs. (15b), (16b), and (17b) are statements of conservation of energy. We now have all the information necessary for performing the synthesis. The entire procedure is summarized below.

C. Summary of Synthesis Procedure

(1) Choose the desired output response $C(\omega)$ and write it in the form of Eq. (2). The C_i must be real.

(2) Calculate the crystal length L from $L=ac/\Delta\eta$. The quantity a is obtained by comparing the $C(\omega)$ written in step (1) to $C(\omega)$ as given by Eq. 2. A rough estimate of L can be obtained from Fig. 7.

(3) Choose a value for I_0^0; the choice is arbitrary so long as $(I_0^0)^2$ exceeds the maximum value of $C(\omega)C^*(\omega)$. It will often be advantageous to make $(I_0^0)^2$ equal to the maximum value of $C(\omega)C^*(\omega)$.

(4) Calculate $D(\omega)D^*(\omega)$ from Eq. (7). Solve for $D(\omega)$ from $D(\omega)D^*(\omega)$ using the method of Appendix A (or some equivalent method). The D_i must be real.

(5) Calculate the output polarizer angle θ_p from Eq. (12a).

(6) Calculate the F_i^n and S_i^n from Eq. (10).

(7) Calculate the crystal angle θ_n of the last stage using Eq. (17a). From Eqs. (C1) and (C2) calculate the input to the last stage (which is the output from the preceding stage).

(8) Repeat the procedure of (7) on each succeeding stage until all crystal angles have been found.

D. Number of Possible Networks

It has been stated that at least one real set of D_i can always be found. It would perhaps be more correct to amend this to read that at least four real sets can always be found, for the calculated sets of D_i can always be placed conveniently into groups of four. The relations between the D_i of these four sets and between the corresponding θ_i are shown in Table I. We see that these four sets give four network configurations which are related. For example, the optical network corresponding to the second set is the "mirror image" of the network obtained from the first set. It can be obtained from it simply by rotating each crystal and the output polarizer a negative, instead of positive, angle.

In addition, it is of interest to note that the network of the third set is precisely the same network that is obtained by turning the first set network end for end. This means that the output of a network will be the same, regardless of which end is used as the input end. Finally, the network resulting from the fourth set is the mirror image of that network obtained from the third set. Therefore, these four sets of D_i do not really give four different networks, but rather one network and three variations.

It is stated in Appendix A that a desired transfer function can be realized by $2^{(n-\frac{1}{2}m+1)}$ different networks, where m is the number of complex roots of Eq. (A8). If we consider that the networks of Table I represent

only one, rather than four, networks, the statement should read $2^{(n-\frac{1}{2}m-1)}$ networks.

V. EXAMPLE

A sample calculation will now be performed to illustrate the synthesis procedure of Sec. IV. Suppose that the ideal transfer function $G(\omega)$ which we wish to approximate is the triangular wave of Fig. 11. A network having such a transfer function might be used as a linear discriminator to accomplish the conversion of frequency-modulated light to amplitude-modulated light.[23] As shown in Fig. 11, $G(\omega)$ is real and has a basic period of $2\pi/a$ rad/sec. We must first approximate $G(\omega)$ by a finite exponential series. A series containing six terms will be used. The number of terms is arbitrary, but in the case of the triangular wave six terms give a satisfactory approximation. For this example the Fourier approximation is used to find the series coefficients, although there are other approximations which could have been used.

The exponential Fourier series approximated to the triangular wave is

$$K(\omega)=4/\pi^2[(1/25)e^{+i5a\omega}+(1/9)e^{+i3a\omega}+e^{ia\omega}+e^{-ia\omega}+(1/9)e^{-i3a\omega}+(1/25)e^{-i5a\omega}], \quad (18)$$

which is plotted in Fig. 11. Note that $K(\omega)$ is the transfer function of a noncausal network. It is often more convenient to first calculate the approximating function in a noncausal form such as Eq. (18), and then make the function causal. We can make $K(\omega)$ causal by multiplying it by $e^{-i5a\omega}$, which gives

$$C(\omega)=e^{-i5a\omega}K(\omega)=4/\pi^2[1/25+(1/9)e^{-i2a\omega}+e^{-i4a\omega}+e^{-i6a\omega}+(1/9)e^{-i8a\omega}+(1/25)e^{-i10a\omega}]. \quad (19)$$

Multiplication by $e^{-i5a\omega}$ is equivalent to introducing a pure time delay in the time domain. Thus the network impulse response and transfer function are essentially unchanged by this operation.

Since alternate harmonics in Eqs. (18) and (19) are zero, we may let $2a\omega=b\omega$. Using this in Eq. (19), we obtain the final form for $C(\omega)$

$$C(\omega)=0.01621+0.04503e^{-ib\omega}+0.40528e^{-i2b\omega}+0.40528e^{-i3b\omega}+0.04503e^{-i4b\omega}+0.01621e^{-i5b\omega}. \quad (20)$$

We must now calculate $D(\omega)$. From Eq. (7) we have

$$|D(\omega)|^2=D(\omega)D^*(\omega)=(I_0^0)^2-C(\omega)C^*(\omega),$$
$$=(I_0^0)^2-0.33309-0.40443\cos b\omega-0.09928\cos 2b\omega-0.03034\cos 3b\omega-0.00292\cos 4b\omega-0.000526\cos 5b\omega. \quad (21)$$

The area I_0^0 of the input impulse must now be chosen in order to obtain $|D(\omega)|^2$. It may have any real value as long as $(I_0^0)^2$ is larger than the maximum value of $C(\omega)C^*(\omega)$. From Fig. 11 we see that the maximum value of $C(\omega)C^*(\omega)$ occurs at $\omega=0$ and has a value of 0.87059. Let us choose $I_0^0=1$. Equation (21) then becomes

$$|D(\omega)|^2=0.66691-0.40443\cos b\omega-0.09928\cos 2b\omega-0.03034\cos 3b\omega-0.00292\cos 4b\omega-0.000526\cos 5b\omega. \quad (22)$$

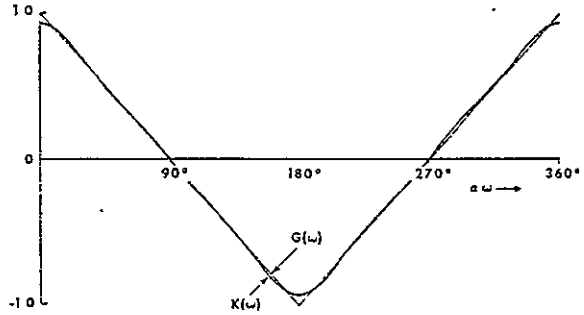


FIG. 11. Ideal and approximating transfer functions of example. Ideal transfer function is shown by dotted line and approximating transfer function by solid line.

We will now use the method outlined in Appendix A to calculate $D(\omega)$. We first form Eq. (A5)

$$-0.000263x^5-0.00146x^4-0.01517x^3-0.04964x^2-0.20222x+0.66691-0.20222x^{-1}-0.04964x^{-2}-0.01517x^{-3}-0.00146x^{-4}-0.000263x^{-5}=0. \quad (23)$$

We next wish to put Eq. (23) in the form of (A6). To determine the B_i, we equate similar coefficients in (A5) and (A6) which gives

$$\begin{aligned} B_5&=A_5 &&=-0.000263,\\ B_4&=A_4 &&=-0.00146,\\ B_3&=A_3-5A_5 &&=-0.01385,\\ B_2&=A_2-4A_4 &&=-0.04380,\\ B_1&=A_1+5A_5-3A_3 &&=-0.15803,\\ B_0&=A_0+2A_4-2A_2 &&=+0.76327. \end{aligned}$$

Substituting these into (A6) and letting $(x+x^{-1})=y$, we have

$$-0.000263y^5-0.00146y^4-0.01385y^3-0.04380y^2-0.15803y+0.76327=0. \quad (24)$$

The roots of (24) are

$$\begin{aligned} &y_1=-4.07379+i3.93269, && y_2=-4.07379-i3.93269,\\ &y_3=0.18957+i6.39532, && y_4=0.18957-i6.39532,\\ &y_5=2.21289. \end{aligned} \quad (25)$$

From Eq. (A7) the corresponding x_i are found to be

$$\begin{aligned} &x_1=-3.95066+i4.05920, && x_2=-3.95066-i4.05920,\\ &x_3=0.18525+i6.54791, && x_4=0.18525-i6.54791,\\ &x_5=1.57997; \end{aligned} \quad (26)$$
$$\begin{aligned} &x_1^{-1}=-0.12313-i0.12652, && x_2^{-1}=-0.12313+i0.12652,\\ &x_3^{-1}=0.00432-i0.15260, && x_4^{-1}=0.00432+i0.15260,\\ &x_5^{-1}=0.63293. \end{aligned}$$

[23] S. E. Harris, Appl. Phys. Letters 2, 47 (1963).

TABLE II. The 16 real sets of D_i.

Set	D_0	D_1	D_2	D_3	D_4	D_5
$x_1x_2x_3x_4x_5$	−0.75607	0.29887	0.07414	0.02091	0.00207	0.00035
$x_1x_2x_3x_4x_5$	0.75607	−0.29887	−0.07414	−0 02901	−0.00207	−0.00035
$x_1^{-1}x_2^{-1}x_3^{-1}x_4^{-1}x_5^{-1}$	−0.00035	−0.00207	−0.02091	−0 07414	−0.29887	0.75607
$x_1^{-1}x_2^{-1}x_3^{-1}x_4^{-1}x_5^{-1}$	0.00035	0.00207	0 02091	0.07414	0.29887	−0.75607
$x_1x_2x_3^{-1}x_4^{-1}x_5$	−0.01762	0.01334	−0.75640	0.29188	0.09415	0.01491
$x_1x_2x_3^{-1}x_4^{-1}x_5$	0.01762	−0.01334	0.75640	−0.29188	−0.09415	−0.01491
$x_1^{-1}x_2^{-1}x_3x_4x_5^{-1}$	−0.01491	−0 09415	−0.29188	0.75640	−0 01334	0.01762
$x_1^{-1}x_2^{-1}x_3x_4x_5^{-1}$	0.01491	0.09415	0.29188	−0.75640	0.01334	−0.01762
$x_1x_2x_3^{-1}x_4^{-1}x_5^{-1}$	−0.01115	0.01901	−0.48006	0.63730	0.17107	0.02356
$x_1x_2x_3^{-1}x_4^{-1}x_5^{-1}$	0.01115	−0.01901	0.48006	−0.63730	−0.17107	−0.02356
$x_1^{-1}x_2^{-1}x_3x_4x_5$	−0.02356	−0.17107	−0.63730	0.48006	−0 01901	0.01115
$x_1^{-1}x_2^{-1}x_3x_4x_5$	0.02356	0.17107	0 63730	−0.48006	0.01901	−0 01115
$x_1x_2x_3x_4x_5^{-1}$	−0.47854	0.64236	0.15462	0 03696	0.00379	0.00055
$x_1x_2x_3x_4x_5^{-1}$	0.47854	−0.64236	−0.15462	−0 03696	−0.00379	−0.00055
$x_1^{-1}x_2^{-1}x_3^{-1}x_4^{-1}x_5$	−0.00055	−0.00379	−0.03696	−0.15462	−0.64236	0.47854
$x_1^{-1}x_2^{-1}x_3^{-1}x_4^{-1}x_5$	0.00055	0.00379	0.03696	0.15462	0.64236	−0.47854

Since there are four complex roots to Eq. (24), there will be $2^{(n-\frac{1}{2}m+1)}=16$ real sets of D_i which can be obtained by multiplying the factors $(x-x_i)$ together in various ways. Eight of these sets are simply the negatives of the other eight. Consider the set that is found by constructing the polynomial

$$(x-x_1)(x-x_2)(x-x_3)(x-x_4)(x-x_5).$$

Performing the indicated multiplication, we obtain

$$x^5+5.95085x^4+60.16845x^3+213.29090x^2 +859.85121x-2175.20862.$$

As stated by Eq. (A11), a set of D_i is proportional to the coefficients of this polynomial.

$$\begin{aligned} D_0&=-2175.20862q, & D_1&=859.85121q,\\ D_2&=213.29090q, & D_3&=60.16845q, \\ D_4&=5.95085q, & D_5&=q. \end{aligned} \tag{27}$$

The value of q is different for each set of D_i. For the above set, q is found from (A12) to be

$$q=\pm 3.47586\times 10^{-4}.$$

Substituting this value back into Eq. (27), we find that one set of D_i is

$$\begin{aligned} D_0&=-0.75607, & D_1&=0.29887,\\ D_2&=0.07414, & D_3&=0.02091,\\ D_4&=0.002068, & D_5&=0.000348. \end{aligned}$$

All 16 real sets of D_i are shown in Table II. We now go through, in detail, the synthesis procedure for the first set.

We first calculate the output polarizer angle from Eq. (12a). Doing so, we obtain

$$\tan\theta_p=D_5/C_5=0.02144,$$

which gives

$$\theta_p=1°14'.$$

Several equations provide checks on the numerical computations and should be used during the synthesis. For example, we should also calculate θ_p from (12b) to verify that (12a) and (12b) do indeed give the same result. These checks are available at various points in the synthesis procedure and will be pointed out when appropriate.

Equation (10) is now used to calculate the F_i^5 and S_i^5, giving

$$\begin{bmatrix} F_0^5\\ F_1^5\\ F_2^5\\ F_3^5\\ F_4^5 \end{bmatrix}=\begin{bmatrix} 0.75625\\ -0.29784\\ -0.06543\\ -0.01222\\ -0.00110 \end{bmatrix}. \qquad \begin{bmatrix} S_1^5\\ S_2^5\\ S_3^5\\ S_4^5\\ S_5^5 \end{bmatrix}=\begin{bmatrix} 0.05143\\ 0.40678\\ 0.40564\\ 0.04507\\ 0.01621 \end{bmatrix}.$$

We are now able to calculate θ_5, the angle of the last birefringent crystal. Using (17a) we find

$$\tan\theta_5=-(F_4^5/S_5^5)=0.06799,$$

which gives

$$\theta_5=3°53'.$$

As a check, we should see that Eq. (17b) is satisfied.

The input impulses to the fifth crystal are calculated next from Eqs. (C1) and (C2).

$$\begin{bmatrix} F_0^4\\ F_1^4\\ F_2^4\\ F_3^4\\ F_4^4 \end{bmatrix}=\frac{1}{\{(F_4^5)^2+(S_5^5)^2\}^{\frac{1}{2}}}\begin{bmatrix} F_0^5 & S_1^5\\ F_1^5 & S_2^5\\ F_2^5 & S_3^5\\ F_3^5 & S_4^5\\ F_4^5 & S_5^5 \end{bmatrix}\begin{bmatrix} S_5^5\\ -F_4^5 \end{bmatrix}=\begin{bmatrix} 0.75799\\ -0.26955\\ -0.03777\\ -0.00913\\ 0 \end{bmatrix},$$

TABLE III. Summary of results of example.

Set	θ_1	θ_2	θ_3	θ_4	θ_5	θ_p
$x_1x_2x_3x_4x_5$	−88°46′	3°53′	29°21′	29°21′	3°53′	1°14′
$x_1x_2v_3v_4x_5$	88°46′	−3°53′	−29°21′	−29°21′	−3°53′	−1°14′
$x_1^{-1}x_2^{-1}x_3^{-1}x_4^{-1}x_5^{-1}$	−1°14′	−3°53′	−29°21′	−29°21′	−3°53′	88°46′
$x_1^{-1}x_2^{-1}x_3^{-1}x_4^{-1}x_5^{-1}$	1°14′	3°53′	29°21′	29°21′	3°53′	−88°46′
$x_1x_2x_3^{-1}x_4^{-1}x_5$	−47°23′	60°25′	−68°34′	−68°34′	60°25′	42°37′
$x_1x_2x_3^{-1}x_4^{-1}x_5$	47°23′	−60°25′	68°34′	68°34′	−60°25′	−42°37′
$x_1^{-1}x_2^{-1}x_3x_4x_5^{-1}$	−42°37′	−60°25′	68°34′	68°34′	−60°25′	47°23′
$x_1^{-1}x_2^{-1}x_3x_4x_5^{-1}$	42°37′	60°25′	−68°34′	−68°34′	60°25′	−47°23′
$x_1x_2x_3^{-1}x_4^{-1}x_5^{-1}$	−34°32′	64°28′	−64°24′	−64°24′	64°28′	55°28′
$x_1x_2x_3^{-1}x_4^{-1}x_5^{-1}$	34°32′	−64°28′	64°24′	64°24′	−64°28′	−55°28′
$x_1^{-1}x_2^{-1}x_3x_4x_5$	−55°28′	−64°28′	64°24′	64°24′	−64°28′	34°32′
$x_1^{-1}x_2^{-1}x_3x_4x_5$	55°28′	64°28′	−64°24′	−64°24′	64°28′	−34°32′
$x_1x_2x_3v_4x_5^{-1}$	−88°04′	7°56′	45°40′	45°40′	7°56′	1°56′
$x_1x_2x_4x_4x_5^{-1}$	88°04′	−7°56′	−45°40′	−45°40′	−7°56′	−1°56′
$x_1^{-1}x_2^{-1}x_3^{-1}x_4^{-1}x_5$	−1°56′	−7°56′	−45°40′	−45°40′	−7°56′	88°04′
$x_1^{-1}x_2^{-1}x_3^{-1}x_4^{-1}x_5$	1°56′	7°56′	45°40′	45°40′	7°56′	−88°04′

$$\begin{bmatrix} S_0^4 \\ S_1^4 \\ S_2^4 \\ S_3^4 \\ S_4^4 \end{bmatrix} = \frac{1}{\{(F_4^5)^2+(S_5^5)^2\}^{\frac{1}{2}}} \begin{bmatrix} F_0^5 & S_1^5 \\ F_1^5 & S_2^5 \\ F_2^5 & S_3^5 \\ F_3^5 & S_4^5 \\ F_4^5 & S_5^5 \end{bmatrix} \begin{bmatrix} F_4^5 \\ S_5^5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.42605 \\ 0.40914 \\ 0.04579 \\ 0.01625 \end{bmatrix}.$$

We can now calculate the angle of the fourth birefringent crystal using Eq. (17a), which gives $\tan\theta_4 = -(F_3^4/S_4^4) = 0.56197$, and $\theta_4 = 29°20'$. Equation (17b) again affords a check.

The synthesis procedure may now be completed by alternately using Appendix C and Eq. (17a) until all crystal angles have been determined. The steps are given below:

$$\begin{bmatrix} F_0^3 \\ F_1^3 \\ F_2^3 \\ F_3^3 \end{bmatrix} = \frac{1}{\{(F_3^4)^2+(S_4^4)^2\}^{\frac{1}{2}}} \begin{bmatrix} F_0^4 & S_1^4 \\ F_1^4 & S_2^4 \\ F_2^4 & S_3^4 \\ F_3^4 & S_4^4 \end{bmatrix} \begin{bmatrix} S_4^4 \\ -F_3^4 \end{bmatrix} = \begin{bmatrix} 0.86952 \\ -0.03454 \\ -0.01049 \\ 0 \end{bmatrix},$$

$$\begin{bmatrix} S_0^3 \\ S_1^3 \\ S_2^3 \\ S_3^3 \end{bmatrix} = \frac{1}{\{(F_3^4)^2+(S_4^4)^2\}^{\frac{1}{2}}} \begin{bmatrix} F_0^4 & S_1^4 \\ F_1^4 & S_2^4 \\ F_2^4 & S_3^4 \\ F_3^4 & S_4^4 \end{bmatrix} \begin{bmatrix} F_3^4 \\ S_4^4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.48873 \\ 0.05842 \\ 0.01864 \end{bmatrix}.$$

$$\tan\theta_3 = -(F_2^3/S_3^3) = 0.56268, \quad \theta_3 = 29°20'.$$

$$\begin{bmatrix} F_0^2 \\ F_1^2 \\ F_2^2 \end{bmatrix} = \frac{1}{\{(F_2^3)^2+(S_3^3)^2\}^{\frac{1}{2}}} \begin{bmatrix} F_0^3 & S_1^3 \\ F_1^3 & S_2^3 \\ F_2^3 & S_3^3 \end{bmatrix} \begin{bmatrix} S_3^3 \\ -F_2^3 \end{bmatrix} = \begin{bmatrix} 0.99746 \\ -0.00145 \\ 0 \end{bmatrix},$$

$$\begin{bmatrix} S_0^2 \\ S_1^2 \\ S_2^2 \end{bmatrix} = \frac{1}{\{(F_2^3)^2+(S_3^3)^2\}^{\frac{1}{2}}} \begin{bmatrix} F_0^3 & S_1^3 \\ F_1^3 & S_2^3 \\ F_2^3 & S_3^3 \end{bmatrix} \begin{bmatrix} F_2^3 \\ S_3^3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.06785 \\ 0.02139 \end{bmatrix}.$$

$$\tan\theta_2 = -(F_1^2/S_2^2) = 0.06802, \quad \theta_2 = 3°53'.$$

$$\begin{bmatrix} F_0^1 \\ F_1^1 \end{bmatrix} = \frac{1}{\{(F_1^2)^2+(S_2^2)^2\}^{\frac{1}{2}}} \begin{bmatrix} F_0^2 & S_1^2 \\ F_1^2 & S_2^2 \end{bmatrix} \begin{bmatrix} S_2^2 \\ -F_1^2 \end{bmatrix} = \begin{bmatrix} 0.99977 \\ 0 \end{bmatrix},$$

$$\begin{bmatrix} S_0^1 \\ S_1^1 \end{bmatrix} = \frac{1}{\{(F_1^2)^2+(S_2^2)^2\}^{\frac{1}{2}}} \begin{bmatrix} F_0^2 & S_1^2 \\ F_1^2 & S_2^2 \end{bmatrix} \begin{bmatrix} F_1^2 \\ S_2^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.02144 \end{bmatrix}.$$

$$\tan\theta_1 = -(F_0^1/S_1^1) = -46.62959, \quad \theta_1 = -88°46',$$

and the synthesis is completed.

The angles calculated from all sixteen real sets of D_i are summarized in Table III. Notice that it is necessary to go through the synthesis procedure for only four of the sets of D_i. The angles for the other 12 sets can be deduced from Table I. The results of the example have been verified by applying the matrix method of Jones[24] to the resulting networks.

VI. DISCUSSION

An important modification of the basic network of Fig. 1 results from associating a variable optical compensator with each crystal of the network. Such compensation can be accomplished either internally to the crystal (for example, by thermal control) or externally[6] (for example, by using a Soleil compensator). To the extent that the transfer function is sufficiently narrow band that the compensation may be considered achromatic, the transfer function may be tuned, without distortion, over its basic period. If we associate a compensation of θ rad with each crystal of the network, the resulting transfer function becomes

$$C_{\text{tuned}}(\omega)=\sum_{k=0}^{n} C_k e^{-ik(a\omega-\theta)} =C_{\text{untuned}}(\omega-\theta/a). \tag{28}$$

The tuned $C(\omega)$ is thus equal to the original $C(\omega)$ shifted by $\theta/2\pi$ of its basic period. Since the required compensation for each crystal is identical, a simple method of tuning such as uniform temperature variation of the entire filter might be attempted. Methods of tuning birefringent filters have been considered by a number of authors.[8-17]

The synthesis procedure is based on the assumption that all crystals have the same length. At first this may seem to be a rather severe restriction, but in reality it is not, for networks containing crystals of different lengths can result from the procedure. It is possible that one or more calculated angles θ_i will be zero, and two crystals with a relative angle of zero degrees are equivalent to a single crystal of twice the length.

An exact procedure for the synthesis of birefringent networks possessing arbitrary transfer functions has been presented. Interesting problems which merit further investigation include: (1) consideration of the effects of crystal misalignment, changes in birefringence, and errors in crystal length; (2) analysis of the angular aperture of these networks and maximization of it; and (3) consideration of the effects of dispersion of $\Delta\eta$.

APPENDIX A

In this Appendix, we give a method for calculating $D(\omega)$ from $|D(\omega)|^2$. In addition, we show that at least one real set of D_i exists, provided $|D(\omega)|^2$ never becomes negative.

[24] R. C. Jones, J. Opt. Soc. Am. 31, 488 (1941).

Suppose we are given the positive semidefinite polynomial

$$|D(\omega)|^2=A_0+2A_1\cos a\omega+\cdots+2A_n\cos na\omega. \tag{A1}$$

Rewriting (A1) we have

$$|D(\omega)|^2=A_ne^{ina\omega}+A_{n-1}e^{i(n-1)a\omega}+\cdots+A_1e^{ia\omega}+A_0 +A_1e^{-ia\omega}+\cdots+A_{n-1}e^{-i(n-1)a\omega}+A_ne^{-ina\omega}. \tag{A2}$$

Notice that the zeros of (A2) appear in reciprocal pairs. Equation (A2) can, therefore, be factored as

$$|D(\omega)|^2=(D_0+D_1e^{ia\omega}+D_2e^{i2a\omega}+\cdots+D_ne^{ina\omega}) \times(D_0+D_1e^{-ia\omega}+\cdots+D_ne^{-ina\omega}). \tag{A3}$$

The D_i are not unique, but rather there are 2^{n+1} possible sets of them. Since $|D(\omega)|^2$ is even and always positive, we may write it as

$$|D(\omega)|^2=D(\omega)D^*(\omega). \tag{A4}$$

Comparing (A3) and (A4), we see that (A4) can be satisfied only if the D_i of (A3) are real. Therefore, at least one *real* set of the coefficients must exist.

The following method of obtaining the D_i is due to Pegis.[4] For more details and explanation of the procedure, the reader should refer to his paper. We begin with $|D(\omega)|^2$ as given by Eq. (A1). Form the equation

$$A_nx^n+A_{n-1}x^{n-1}+\cdots+A_1x+A_0+A_1x^{-1}+\cdots +A_{n-1}x^{-(n-1)}+A_nx^{-n}=0. \tag{A5}$$

Put Eq. (A5) into the form

$$B_n(x+x^{-1})^n+B_{n-1}(x+x^{-1})^{n-1}+\cdots+B_0=0. \tag{A6}$$

and obtain the B_i from the A_i by equating similar coefficients. Make the substitution

$$x+x^{-1}=y \tag{A7}$$

into Eq. (A6), which gives

$$B_ny^n+B_{n-1}y^{n-1}+\cdots+B_0=0. \tag{A8}$$

Solve for the n roots of (A8) and call them $y_1, y_2, \cdots, y_n$. Using Eq. (A7), solve for the reciprocal pairs of roots

$$\begin{array}{ll} x_1, & 1/x_1, \\ x_2, & 1/x_2, \\ \vdots & \vdots \\ x_n. & 1/x_n. \end{array} \tag{A9}$$

Next, construct all possible equations having real coefficients d_i using one root from each row of (A9); e.g., one possibility would be

$$(x-x_1)(x-1/x_2)(x-x_3)\cdots(x-x_n) =x^n+d_{n-1}x^{n-1}+\cdots+d_2x^2+d_1x+d_0. \tag{A10}$$

The D_i are proportional to the d_i

$$\begin{aligned} D_0&=qd_0,\\ D_1&=qd_1,\\ &\vdots\quad\vdots\\ D_n&=qd_n. \end{aligned} \tag{A11}$$

The quantity q is found from

$$q^2(d_0^2+d_1^2+\cdots+d_n^2)=A_0, \tag{A12}$$

and upon substituting this value into (A11), we obtain the D_i.

The number of real sets of D_i will depend upon the number of y_i which are complex. If m of the y_i are complex, there will be $2^{(n-\frac{1}{2}m+1)}$ real sets of D_i. Half of these, however, will just be the negative of the other half, for q can be negative or positive.

APPENDIX B

We derive here the conditions which the F_i and S_i satisfy because of conservation of energy. Assume, for convenience, that we are dealing with a four-crystal network. Since the birefringent crystals are assumed to be lossless, the energy in the fast-axis output plus the energy in the slow-axis output of the fourth crystal must equal the input energy. Mathematically, this is expressed by

$$F^4(\omega)F^{4*}(\omega)+S^4(\omega)S^{4*}(\omega)=(I_0^0)^2. \tag{B1}$$

Writing out (B1) and equating similar coefficients, we obtain

$$(F_0^4)^2+(F_1^4)^2+(F_2^4)^2+(F_3^4)^2+(S_1^4)^2+(S_2^4)^2+(S_3^4)^2+(S_4^4)^2=(I_0^0)^2, \tag{B2}$$

$$F_0^4F_1^4+F_1^4F_2^4+F_2^4F_3^4+S_1^4S_2^4+S_2^4S_3^4+S_3^4S_4^4=0, \tag{B3}$$

$$F_0^4F_2^4+F_1^4F_3^4+S_1^4S_3^4+S_2^4S_4^4=0, \tag{B4}$$

$$F_0^4F_3^4+S_1^4S_4^4=0. \tag{B5}$$

Similarly, we can derive for the ith stage

$$(F_0^i)^2+(F_1^i)^2+\cdots+(F_{i-1}^i)^2+(S_1^i)^2+(S_2^i)^2+\cdots+(S_i^i)^2=(I_0^0)^2, \tag{B6}$$

$$F_0^iF_1^i+F_1^iF_2^i+\cdots+F_{i-2}^iF_{i-1}^i+S_1^iS_2^i+S_2^iS_3^i+\cdots+S_{i-1}^iS_i^i=0, \tag{B7}$$

$$F_0^iF_2^i+\cdots+F_{i-3}^iF_{i-1}^i+S_1^iS_3^i+\cdots+S_{i-2}^iS_i^i=0, \tag{B8}$$

$$\vdots$$

$$F_0^iF_{i-1}^i+S_1^iS_i^i=0. \tag{B9}$$

It should be pointed out that $C(\omega)$ and $D(\omega)$ also satisfy conservation of energy, giving the equations

$$(C_0)^2+(C_1)^2+\cdots+(C_n)^2+(D_0)^2+(D_1)^2+\cdots+(D_n)^2=(I_0^0)^2, \tag{B10}$$

$$C_0C_1+C_1C_2+\cdots+C_{n-1}C_n+D_0D_1+D_1D_2+\cdots+D_{n-1}D_n=0, \tag{B11}$$

$$C_0C_2+C_1C_3+\cdots+C_{n-2}C_n+D_0D_2+D_1D_3+\cdots+D_{n-2}D_n=0, \tag{B12}$$

$$\vdots$$

$$C_0C_n+D_0D_n=0. \tag{B13}$$

APPENDIX C

This Appendix gives a systematic and rapid method of calculating the input to a crystal once the output is known. This is simply a formalized procedure of solving for the F^{i-1} and S^{i-1} of Eq. (13d) once the F^i and S^i are known. In matrix form, the expressions are

$$\begin{bmatrix} F_0^{i-1}\\ F_1^{i-1}\\ F_2^{i-1}\\ \vdots\\ F_{i-1}^{i-1}\end{bmatrix}=\frac{1}{\{(F_{i-1}^i)^2+(S_i^i)^2\}^{\frac{1}{2}}}\begin{bmatrix} F_0^i & S_1^i\\ F_1^i & S_2^i\\ \vdots & \vdots\\ F_{i-1}^i & S_i^i\end{bmatrix}\begin{bmatrix} S_i^i\\ -F_{i-1}^i\end{bmatrix} \tag{C1}$$

$$\begin{bmatrix} S_0^{i-1}\\ S_1^{i-1}\\ \vdots\\ S_{i-1}^{i-1}\end{bmatrix}=\frac{1}{\{(F_{i-1}^i)^2+(S_i^i)^2\}^{\frac{1}{2}}}\begin{bmatrix} F_0^i & S_1^i\\ F_1^i & S_2^i\\ \vdots & \vdots\\ F_{i-1}^i & S_i^i\end{bmatrix}\begin{bmatrix} F_{i-1}^i\\ S_i^i\end{bmatrix}. \tag{C2}$$

One convenient check is that the calculated values F_{i-1}^{i-1} and S_0^{i-1} should always be zero.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the interest, encouragement, and support given throughout the course of this work by Dr. B. J. McMurtry and Professor A. E. Siegman.

Reprinted from JOURNAL OF THE OPTICAL SOCIETY OF AMERICA, Vol. 55, No. 7, 835-841. July 1965
Printed in U. S. A.

Optical Network Synthesis using Birefringent Crystals. II. Synthesis of Networks Containing One Crystal, Optical Compensator, and Polarizer per Stage*

E. O. AMMANN
Electronic Defense Laboratories, Sylvania Electronic Systems, Mountain View, California 94042
AND
I. C. CHANG
W. W. Hansen Laboratories, Stanford University, Stanford, California 94305
(Received 11 January 1965)

A second technique for the synthesis of birefringent networks having arbitrarily prescribed spectral transmittance is presented. The network consists of an input polarizer followed by n cascaded stages. Each stage contains (in general) a birefringent crystal, an optical compensator, and a polarizer. The quantities determined for each stage by the synthesis procedure are the angle to which the crystal is rotated, the amount of delay to be introduced by the optical compensator, and the angle of the polarizer. A desired function consisting of $(n+1)$ terms of an exponential series can be realized by an n-stage network and, in certain cases, by a network containing fewer than n stages. The synthesis procedures of Part I and Part II are compared, and their relative merits discussed. Finally, two examples are given using the procedure of Part II.

I. INTRODUCTION

IN Part I a procedure for the synthesis of birefringent networks having arbitrarily prescribed periodic spectral transmittance was given.[1] The basic network consisted of n identical cascaded birefringent crystals between an input and an output polarizer. Such a network was termed a lossless or nondissipative network, since it does not contain any internal polarizers.

Part II describes a second procedure for the synthesis of birefringent networks, which results in a network containing internal polarizers. As before, this procedure allows realization of an arbitrary spectral transmittance, provided it is periodic.

The type of network obtained from the synthesis procedure of Part II is shown in Fig. 1. It consists of an input polarizer followed by a series of stages, each stage containing a birefringent crystal, optical compensator, and polarizer. This network contains three stages: however, any number of stages may be used. The birefringent crystals in all stages are identical (with a few important exceptions, noted later), each crystal being cut with its optic axis[2] perpendicular to its length and with end faces which are flat and parallel. The crystals' "slow" and "fast" axes are denoted by S and F, respectively, in Fig. 1. The quantities determined for each stage by the synthesis procedure are the angle to which the crystal is rotated, the amount of delay to be introduced by the optical compensator, and the angle of the polarizer. In addition, the length L of the crystals is determined by the periodicity of the desired spectral transmittance.

The optical compensators contained in the network of Fig. 1 are assumed to be achromatic and are therefore impossible to realize in practice. For the range of frequencies over which their behavior is approximately achromatic, the network spectral-amplitude transmittance is the desired $C(\omega)$. Outside this range of frequencies, the actual and desired characteristics diverge. This is not a severe limitation, particularly since birefringent devices are used primarily to obtain extremely narrow bandwidths.

In the synthesis procedure of Part II and the procedure of Part I, we have two different techniques (and two different birefringent networks) for realizing a given spectral transmittance. The relative merits of these two synthesis procedures are discussed, and the expected performances of the two types of networks are compared. Finally, two examples are given, one of which shows how a Lyot filter can be obtained from the synthesis procedure of this paper.

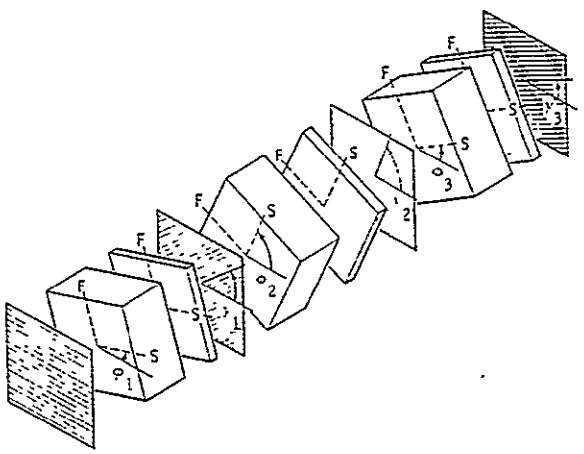


FIG. 1. Basic configuration of optical network (three stages). Polarizers are shown shaded. F and S denote the "fast" and "slow" axes of the birefringent crystals and optical compensators.

* This work was supported at Sylvania by the Air Force Avionics Laboratory at Wright-Patterson Air Force Base, Ohio, under contracts AF 33(657)-8995 and AF 33(615)-1938, and at Stanford University by the U. S. Office of Naval Research under contract NONR 225(24).

[1] S. E. Harris, E. O. Ammann, and I. C. Chang, J. Opt. Soc. Am. 54, 1267 (1964). [Editor's note: the term "transfer function" used in this reference (Part I of this series) has been changed to the more explicit "spectral transmittance" and "amplitude transmittance," as the case may be, in this Part II].

[2] In principal, either uniaxial or biaxial crystals may be employed. For simplicity we assume that uniaxial crystals are used.

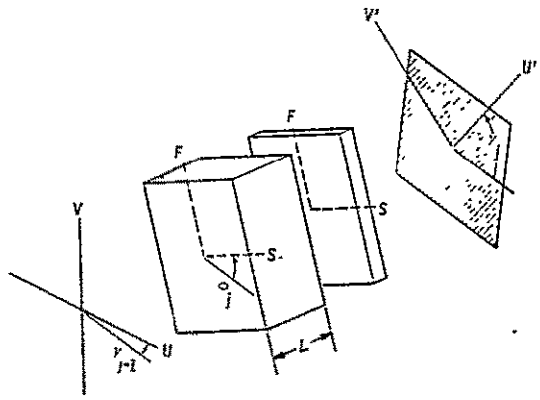


FIG. 2. Components which make up a single stage of the network: birefringent crystal, optical compensator, and polarizer.

II. SYNTHESIS PROCEDURE

In this section we show that the network of Fig. 1 is capable of producing any desired spectral transmittance. Expressions are derived for the crystal angle, the optical compensation, and the polarizer angle for each stage. Several of the topics pertinent to this section are discussed at greater length in Ref. 1.

We begin by assuming that a desired spectral-amplitude-transmittance function, written in the form of Eq. (1), has been chosen,

$$C(\omega)=C_0+C_1e^{-i\alpha\omega}+C_2e^{-i2\alpha\omega}+\cdots+C_ne^{-in\alpha\omega}. \quad (1)$$

The C_j may be real or complex. (Recall that the synthesis procedure of Part I requires all C_j to be real.)

Equation (1) can be considered a polynomial in $e^{-i\alpha\omega}$ and rewritten as

$$C(\omega)=C_n[(C_0/C_n)+(C_1/C_n)e^{-i\alpha\omega}+\cdots+e^{-in\alpha\omega}] \quad (2)$$

$$=C_n(-z_1+e^{-i\alpha\omega})(-z_2+e^{-i\alpha\omega})\cdots(-z_n+e^{-i\alpha\omega}), \quad (3)$$

where the z's are the zeros of the polynomial. These zeros are, in general, complex. If we express each zero in terms of its magnitude $|z|$ and phase angle $e^{i\alpha}$, Eq. (3) becomes

$$C(\omega)=C_n(-|z_1|e^{i\alpha_1}+e^{-i\alpha\omega})\times(-|z_2|e^{i\alpha_2}+e^{-i\alpha\omega})\cdots(-|z_n|e^{i\alpha_n}+e^{-i\alpha\omega}). \quad (4)$$

We now show that each factor in Eq. (4) can be associated with one stage of the network of Fig. 1.

Let us consider one stage (say the jth stage) of the network of Fig. 1. This stage contains a crystal, an optical compensator, and a polarizer as shown in Fig. 2. Using the Jones-calculus formulation,[3] the input and output of the jth stage are found to be related by

$$\begin{pmatrix}E_{u'}\\E_{v'}\end{pmatrix}=\begin{pmatrix}1&0\\0&0\end{pmatrix}\begin{pmatrix}\cos(\gamma_j-\phi_j)&\sin(\gamma_j-\phi_j)\\-\sin(\gamma_j-\phi_j)&\cos(\gamma_j-\phi_j)\end{pmatrix}\times\begin{pmatrix}e^{-ib}&0\\0&1\end{pmatrix}\begin{pmatrix}e^{-i\alpha\omega}&0\\0&1\end{pmatrix}\times\begin{pmatrix}\cos(\phi_j-\gamma_{j-1})&\sin(\phi_j-\gamma_{j-1})\\-\sin(\phi_j-\gamma_{j-1})&\cos(\phi_j-\gamma_{j-1})\end{pmatrix}\begin{pmatrix}E_u\\E_v\end{pmatrix}, \quad (5)$$

where E denotes the complex amplitude of the electric-field vector of the light. The u axis is parallel to the transmission axis of the polarizer of the preceding stage. The birefringent crystal introduces a phase difference of $a\omega$ rad between the S and F components, with $a=L\Delta\eta/c$. (The quantity L is the crystal length, $\Delta\eta$ the birefringence, and c the velocity of light in vacuum.) The quantity b is the phase difference introduced by the optical compensator and is always between 0 and 2π rad. The optical compensator is assumed to be achromatic; this is the only approximation involved in this procedure.

Perhaps a more familiar form than that used in Eq. (5) for the matrix of a birefringent crystal is

$$\begin{pmatrix}e^{-i\alpha\omega/2}&0\\0&e^{i\alpha\omega/2}\end{pmatrix}. \quad (6)$$

If we factor $e^{i\alpha\omega/2}$ out of (6), the matrix used in Eq. (5) results. Furthermore, the factored $e^{i\alpha\omega/2}$ term can be dropped without affecting the results.[1]

From Eq. (5) we see that it is more convenient to deal with relative angles (measured from the preceding component) than absolute angles. Consequently, we define

$$\begin{aligned}\theta_1&=\phi_1,\\ \theta_2&=\phi_2-\gamma_1,\\ \theta_3&=\phi_3-\gamma_2,\\ &\vdots\\ \theta_j&=\phi_j-\gamma_{j-1},\end{aligned} \quad (7)$$

and

$$\begin{aligned}\beta_1&=\gamma_1-\phi_1,\\ \beta_2&=\gamma_2-\phi_2,\\ &\vdots\\ \beta_j&=\gamma_j-\phi_j.\end{aligned} \quad (8)$$

Performing the matrix multiplication indicated in Eq. (5) and noting that E_v is zero, we obtain

$$E_{u'}=e^{-ib}\cos\theta_j\cos\beta_j(-e^{ib}i\tan\theta_j\tan\beta_j+e^{-i\alpha\omega})E_u. \quad (9)$$

The amplitude-transmittance function for the entire network (all stages) is the product of the amplitude-transmittance functions of the individual stages. Thus

[3] R. C. Jones, J. Opt. Soc. Am. 31, 488 (1941).

we have for the entire network,[4]

$$C'(\omega)=e^{-ibT}\cos\theta_1\cos\beta_1\cos\theta_2\cos\beta_2\cdots\cos\theta_n\cos\beta_n \times(-e^{ib_1}\tan\theta_1\tan\beta_1+e^{-ia\omega}) \times(-e^{ib_2}\tan\theta_2\tan\beta_2+e^{-ia\omega})\cdots \times(-e^{ib_n}\tan\theta_n\tan\beta_n+e^{-ia\omega}), \quad (10)$$

where

$$(e^{-ib_1})(e^{-ib_2})(e^{-ib_3})\cdots(e^{-ib_n})=e^{-ibT}. \quad (11)$$

If all C_j are real, then $e^{-ibT}=\pm 1$, since all complex roots will then occur as conjugate pairs.

We see that Eqs. (4) and (10) have the same form. Equating similar terms, we find

$$\tan\theta_1\tan\beta_1=|z_1|, \quad b_1=\alpha_1, \quad (12a)$$

$$\tan\theta_2\tan\beta_2=|z_2|, \quad b_2=\alpha_2, \quad (12b)$$

$$\vdots \qquad \vdots$$

$$\tan\theta_j\tan\beta_j=|z_j|, \quad b_j=\alpha_j. \quad (12c)$$

In Eqs. (4) and (10), the order of the factors is, of course, unimportant. It is only necessary to equate each of the factors in Eq. (10) with one of the factors in Eq. (4). Note that if a root is real and positive, $\alpha=b=0$, and an optical compensator is not required for that stage.

Terms in Eqs. (4) and (10) may also be equated in a slightly different manner, namely,

$$\tan\theta_1\tan\beta_1=-|z_1|, \quad b_1=\alpha_1+\pi, \quad (13a)$$

$$\tan\theta_2\tan\beta_2=-|z_2|, \quad b_2=\alpha_2+\pi, \quad (13b)$$

$$\vdots \qquad \vdots$$

$$\tan\theta_j\tan\beta_j=-|z_j|, \quad b_j=\alpha_j+\pi. \quad (13c)$$

Here if a root is real and negative, an optical compensator is not required for that stage. Either Eqs. (12) or Eqs. (13) may be used for obtaining θ, β, and b for a particular stage.

We have shown that it is possible to choose a crystal angle, polarizer angle, and value of optical compensation to match any factor of Eq. (4). Thus any $C(\omega)$ can be realized *to within a multiplicative constant*, by use of the network of Fig. 1. This multiplicative constant will be real if all C_j are chosen real, and will be complex if one or more of the C_j are chosen complex. That the actual and ideal $C(\omega)$'s differ by a *complex* constant is not of consequence *provided the network spectral transmittance is relatively narrow band.*

Returning to Eqs. (12), we see that θ and β are not uniquely determined for a particular stage. There are an infinite number of combinations of θ_j and β_j which satisfy the equation $\tan\theta_j\tan\beta_j=|z_j|$. Thus an additional criterion may be imposed when choosing the particular θ_j and β_j to be used.

One important possibility is to choose θ_j and β_j so that the magnitude of the output is maximized. When we satisfy Eqs. (12c) [or (13c)], we ensure that the C_j' have the proper *relative* values. However, the magnitude of the entire response depends upon the particular choices for the θ_j and β_j. We therefore wish to maximize the magnitude of one of the C_j' subject to the constraints imposed by Eqs. (12).[5] Let us maximize the magnitude of C_n', which is found by comparing Eqs. (4) and (10), to be

$$|C_n'|=\cos\theta_1\cos\beta_1\cos\theta_2\cos\beta_2\cdots\cos\theta_n\cos\beta_n. \quad (14)$$

This problem can be solved by the method of Lagrange multipliers. We obtain for the result

$$\tan\theta_1=\tan\beta_1=\pm|z_1|^{\frac{1}{2}},$$
$$\tan\theta_2=\tan\beta_2=\pm|z_2|^{\frac{1}{2}}, \quad (15)$$
$$\vdots \qquad \vdots$$
$$\tan\theta_j=\tan\beta_j=\pm|z_j|^{\frac{1}{2}}.$$

If we maximize Eq. (14) using Eqs. (13) instead of Eqs. (12) as the constraint, we obtain

$$\tan\theta_1=-\tan\beta_1=\pm|z_1|^{\frac{1}{2}},$$
$$\tan\theta_2=-\tan\beta_2=\pm|z_2|^{\frac{1}{2}},$$
$$\vdots \qquad \vdots$$
$$\tan\theta_j=-\tan\beta_j=\pm|z_j|^{\frac{1}{2}}. \quad (16)$$

It is interesting to note that if the Eqs. (13) and (16) are used, all of the polarizers of the network will be oriented with their transmission axes parallel to the transmission axis of the input polarizer. This occurs because $\beta_j=-\theta_j$ for each stage of the network. If Eqs. (12) and (15) are employed, the polarizers will be rotated at various angles.

Networks Containing Crystals of Unequal Lengths

In general, the synthesis procedure just described requires an n-stage network to realize a $C(\omega)$ containing $(n+1)$ terms. Occasionally, however, it is possible to realize a $C(\omega)$ using fewer stages with longer crystals in those stages. This comes about in the following manner.

We have shown that each of the factors of Eq. (3) can be associated with one stage of the network. Each stage contains a birefringent crystal of length L, an optical compensator, and a polarizer. Suppose now that two of the factors of Eq. (3), when multiplied together, give a term of the form

$$(-z+e^{-i2a\omega}). \quad (17)$$

It is apparent that (17) can be realized by a *single* stage containing a crystal of length $2L$, an optical compensator, and a polarizer. In general, if r factors of Eq. (3) multiply together to produce a term of the form

$$(-z+e^{-ira\omega}), \quad (18)$$

this term can be realized by a single stage containing a crystal of length rL, an optical compensator, and a

[4] $C'(\omega)$ refers to the actual amplitude-transmittance function of the birefringent network, while $C(\omega)$ refers to the desired amplitude-transmittance function. The two differ only by a constant multiplier.

[5] Since the relative values of the C's are fixed, maximizing any one of them maximizes all of them.

polarizer. As before, the crystal angle, polarizer angle, and the optical compensation are calculated from the magnitude and phase angle of z.

Once the zeros of $C(\omega)$ have been found, it is an easy task to determine whether factors can be combined to produce a term similar to (18). The following rule is useful. If r of the zeros have the same magnitude and have α's which differ by $2\pi/r$, the corresponding factors can be multiplied together to produce a term of the form of (18).

III. DISCUSSION

Suppose we wish to realize some $C(\omega)$ with all C_j real. We now have two methods of accomplishing this—the synthesis procedures of Part I and Part II. It is of interest then to consider the relative merits of the two procedures (and their corresponding networks) in order to determine which should be used in a particular situation. We compare them on the following points.

A. Number of Components Required to Synthesize a Given $C(\omega)$

It is usually desirable to have an optical network composed of as few components as possible. By minimizing the number of components, we reduce the number of surfaces at which reflection can occur, and also minimize the number of crystal surfaces which must be ground and polished. The required number of components is of particular interest when a $C(\omega)$ containing a large number of terms is required. We compare the two synthesis procedures by discussing for each procedure: (1) the number of components required per stage; and (2) the number of stages required to realize a given $C(\omega)$.

Theoretically, each stage of the network of Part I consists of only a birefringent crystal while each stage of the network of Part II contains a birefringent crystal, optical compensator, and polarizer. In practice, however, an optical compensator is probably also needed with each crystal of the network of Part I. Thus the main difference between the two networks is often one component (a polarizer) per stage.

Both types of networks, in general, require n stages to realize a $C(\omega)$ containing $(n+1)$ terms. However, in certain instances, it is possible to reduce the number of stages needed for the network of Part I, while in still other instances, it is possible to reduce the number of stages needed for the network of Part II.

Consider first the network of Part I. When two or more successive crystals are rotated to the same angle, the situation is equivalent to a single crystal of greater length. Unfortunately, efforts to determine what restrictions must be placed on $C(\omega)$ to cause several crystals to have the same angle thus far have been unsuccessful.

On the other hand, it is relatively easy to determine which $C(\omega)$'s result in fewer stages when the synthesis procedure of Part II is employed. Section II notes that if r zeros (written in polar form) have the same magnitude and differ successively by $2\pi/r$ phase angle, then r of the stages can be combined into a single stage having a crystal of length rL. Therefore, once the zeros of $C(\omega)$ are known, it is immediately apparent whether multiple-length crystals can be used. In some cases, it may even be feasible to adjust some of the zeros so that they satisfy the above condition. It is necessary, of course, to make certain that this does not cause an unacceptable change in $C(\omega)$.

B. Magnitude of Output

A second basis of comparison involves the magnitudes of the outputs of the two types of networks. Suppose that we use both the synthesis procedure of Part II and that of Part I to obtain networks having the same desired $C(\omega)$ with all C_j real. Will the amplitudes of the responses of the two networks be the same? If not, what will the relative magnitudes be?

To answer these questions, we first assume that "perfect" polarizers are used in both networks. In addition, we assume that the desired $C(\omega)$ has a maximum magnitude of unity. Consider the network obtained from the procedure of Part I. Its amplitude-transmittance is identical to the desired $C(\omega)$; i.e., both the relative and absolute values of the C_j' are identical to those of the C_j.

Next consider the network obtained by the synthesis procedure of Part II. Its amplitude-transmittance function contains C_j' whose relative values are correct (i.e., C_0'/C_n', C_1'/C_n', $\cdots C_{n-1}'/C_n'$), but whose absolute values probably are not. This discrepancy occurs because we are trying to use n stages to produce the $(n+1)$ different C_j. Thus we have one less degree of freedom than is necessary.

In most instances, the actual C_j' obtained from the network of Part II are smaller than the desired C_j. In a few cases, the two are the same. The C_j' can never be larger than the C_j [unless the maximum magnitude of the desired $C(\omega)$ is less than unity]. We determine below: (1) under what conditions the C_j' and C_j have the same magnitudes; and (2) what the reduction in magnitude is when the C_j' and C_j are not the same.

Let us begin with point (1), keeping in mind that the following discussion assumes that all C_j are chosen real. As stated earlier, the desired $C(\omega)$ has a magnitude of unity at least once per period. In order for the network of Part II to have unit transmittance at some frequency, each stage must have unit transmittance. Consider two stages of the network which correspond to two complex roots (a conjugate pair). From Eq. (9), their amplitude-transmittance functions are

$$(-e^{i\delta}\sin^2\theta+e^{-i\alpha\omega}\cos^2\theta)(-e^{-i\delta}\sin^2\theta+e^{-i\alpha\omega}\cos^2\theta). \quad (19)$$

We have made $\beta=\theta$, as prescribed by Eq. (15), for maximum transmittance.

In order for the first factor of (19) to have unit magnitude, it must be true that

$$-e^{ib}=e^{-ia\omega},$$

which gives

$$-b=a\omega+p\pi, \tag{20}$$

where p is any odd integer. If we similarly require the second factor of (19) to have unit magnitude, we obtain

$$+b=a\omega+q\pi, \tag{21}$$

where q is any odd integer. Solving Eqs. (20) and (21) simultaneously, we find that the only possible values which b can have (between 0 and 2π) are 0 and π. These values result in unit transmittance at

$$a\omega=\pi,\ 3\pi,\ 5\pi,\cdots$$

and

$$a\omega=0,\ 2\pi,\ 4\pi,\cdots,$$

respectively.

In other words, the networks of Part II can have unit transmittance (once per period of their transmittance function) if: (1) all roots of the desired $C(\omega)$ are real and positive; or if (2) all roots of $C(\omega)$ are real and negative. Only under these two conditions do all stages have 100% transmittance at the *same* frequency, thereby resulting in C_j' which are identical in magnitude to the desired C_j.

If the roots of $C(\omega)$ do not fit into one of these two categories, the output from the network of Part II will be smaller than the output from the corresponding network of Part I. We now determine how much smaller it is. To do this, we compare the magnitudes of C_n' of the two networks.

For the network of Part I, C_n' is identical to C_n. For the network of Part II, we find from Eqs. (14) and (15) that

$$C_n'=\cos^2\theta_1\cos^2\theta_2\cdots\cos^2\theta_n. \tag{22}$$

We can solve for $\cos^2\theta_j$ in terms of $|z_j|$ by noting from (15) that

$$\tan^2\theta_j=|z_j|. \tag{23}$$

From (23), it is easily shown that

$$\cos^2\theta_j=1/(1+|z_j|). \tag{24}$$

Equation (22) can now be rewritten as

$$C_n'=[1/(1+|z_1|)][1/(1+|z_2|)]\cdots[1/(1+|z_n|)], \tag{25}$$

which is the desired result. Thus, by calculating C_n' from Eq. (25) and comparing this to C_n, we have our comparison between the amplitudes of the outputs of the two networks.[6]

If nonideal polarizers are used in the birefringent networks, the networks of Part II will fare even more poorly on the basis of output magnitude compared with the networks of Part I. Polarizer losses are especially damaging if the network contains a large number of stages.

The relative merits of the two networks can also be argued on the basis of various other criteria such as angular aperture, effect of crystal misorientation, etc. These and other topics remain to be investigated.

As a final comment, it is of interest to note that the synthesis procedure of Part II is computationally much easier than the procedure of Part I. Once the roots of $C(\omega)$ have been found, the work is essentially completed.

IV. SUMMARY OF SYNTHESIS PROCEDURE

(1) Choose the desired spectral amplitude transmittance $C(\omega)$, and write it in the form of Eq. (1). It is usually desirable to choose the C_j so that $C(\omega)$ has a maximum magnitude of unity. The C_j may be real or complex; however, in most instances it is expedient to use a real set.

(2) Rewrite Eq. (1) in the form of Eq. (2) and solve for the roots (the z_j) of the polynomial. Each root should be written in polar form, i.e., in terms of a magnitude $|z_j|$ and a phase angle $e^{i\alpha_j}$. Each factor of Eq. (3) can be associated with one stage of the network of Fig. 1.

(3) If r of the zeros have the same magnitude and have α's which differ successively by $2\pi/r$, the factors containing these zeros can be multiplied together to produce a term of the form of (18). This makes possible the replacement of r stages by a single stage containing a crystal of length rL, where $rL=rac/\Delta\eta$. (Otherwise, the required crystal length L for a stage will be $L=ac/\Delta\eta$.)

(4) If maximum transmittance through the network is desired, the crystal angle θ_j, polarizer angle β_j, and optical compensation b_j of the jth stage should be calculated from

$$\theta_j=\beta_j=\tan^{-1}(\pm|z_j|^{\frac{1}{2}}),\quad b_j=\alpha_j,$$

or

$$\theta_j=-\beta_j=\tan^{-1}(\pm|z_j|^{\frac{1}{2}}),\quad b_j=\alpha_j+\pi.$$

If the condition of maximum transmittance is not necessary, Eqs. (12c) or (13c) should be used. Some other criterion is then necessary to determine θ_j and β_j uniquely.

(5) If the θ_j and β_j are calculated to give maximum transmittance, the amplitude of the network's actual spectral transmittance can be compared to the amplitude of the ideal spectral transmittance by calculating $|C_n'|$ from Eq. (25) and comparing it to C_n. The actual spectral transmittance usually is smaller than the desired transmittance.

V. EXAMPLES

Two examples are given to illustrate the synthesis procedure. The first example is concerned with obtaining

[6] Equation (25) can also be used to calculate $|C_n'|$ when one or more of the C_j is complex.

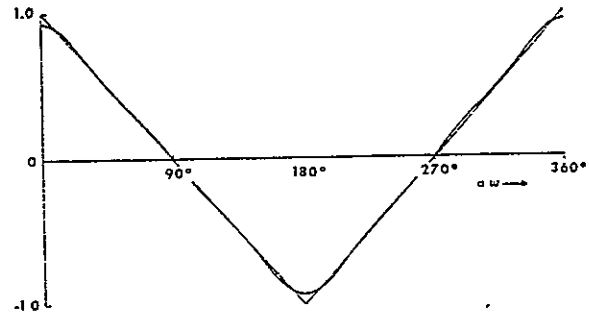


FIG. 3. Ideal and approximating amplitude-transmittance functions of first example. Ideal amplitude-transmittance function is shown by dotted line, and approximating function by solid line.

a network whose spectral amplitude transmittance approximates a triangular wave. This is recognized as the example used also in Part I. As before, we approximate the triangular wave by the first six terms of its exponential Fourier series,

$$K(\omega)=(4/\pi^2)[(1/25)e^{i5a\omega}+(1/9)e^{i3a\omega}+e^{ia\omega}+e^{-ia\omega}+(1/9)e^{-i3a\omega}+(1/25)e^{-i5a\omega}]. \quad (26)$$

The ideal and approximating transmittance functions are plotted in Fig. 3. We next multiply $K(\omega)$ by $e^{-i5a\omega}$ to cast the transmittance function in the form of Eq. (1),

$$C(\omega)=e^{-i5a\omega}K(\omega)=(4/\pi^2)[(1/25)+(1/9)e^{-i2a\omega}+e^{-i4a\omega}+e^{-i6a\omega}+(1/9)e^{-i8a\omega}+(1/25)e^{-i10a\omega}]. \quad (27)$$

Finally, letting $2a\omega=b\omega$, we obtain the final form for $C(\omega)$,

$$C(\omega)=0.01621+0.04503e^{-ib\omega}+0.40528e^{-i2b\omega}+0.40528e^{-i3b\omega}+0.04503e^{-i4b\omega}+0.01621e^{-i5b\omega}. \quad (28)$$

Putting Eq. (28) in the form of Eq. (2), we have

$$\begin{aligned}C(\omega)&=0.01621(1+2.77778e^{-ib\omega}+25.00000e^{-i2b\omega}+25.00000e^{-i3b\omega}+2.77778e^{-i4b\omega}+e^{-i5b\omega}) \quad (29)\\&=0.01621(-|z_1|e^{i\alpha_1}+e^{-ib\omega})(-|z_2|e^{i\alpha_2}+e^{-ib\omega})\times(-|z_3|e^{i\alpha_3}+e^{-ib\omega})(-|z_4|e^{i\alpha_4}+e^{-ib\omega})\times(-|z_5|e^{i\alpha_5}+e^{-ib\omega}).\end{aligned}$$

Solving for the roots of Eq. (29), we obtain

$$\begin{aligned}|z_1|&=0.208275, & \alpha_1&=\ 100°14',\\|z_2|&=0.208275, & \alpha_2&=-100°14',\\|z_3|&=4.801344, & \alpha_3&=\ 100°14',\\|z_4|&=4.801344, & \alpha_4&=-100°14',\\|z_5|&=1.000000, & \alpha_5&=\ 180°00'.\end{aligned}$$

Upon inspecting these roots, we are unable to find r roots whose amplitudes are equal and whose phase angles differ by $2\pi/r$. Five stages are therefore required in the birefringent network and the length of the crystal for each stage is given by $L=bc/\Delta\eta$.

Let us choose the crystal angles θ_j and polarizer angles β_j so the output amplitude is maximized. Using Eq. (13) to calculate the b_j and Eqs. (16) to calculate the θ_j and β_j, we have

$$\begin{aligned}\theta_1&=-\beta_1=\tan^{-1}|z_1|^{\frac{1}{2}}=24°32', & b_1&=\alpha_1+180°=280°14',\\\theta_2&=-\beta_2=\tan^{-1}|z_2|^{\frac{1}{2}}=24°32', & b_2&=\alpha_2+180°=\ 79°46',\\\theta_3&=-\beta_3=\tan^{-1}|z_3|^{\frac{1}{2}}=65°28', & b_3&=\alpha_3+180°=280°14',\\\theta_4&=-\beta_4=\tan^{-1}|z_4|^{\frac{1}{2}}=65°28', & b_4&=\alpha_4+180°=\ 79°46',\\\theta_5&=-\beta_5=\tan^{-1}|z_5|^{\frac{1}{2}}=45°00', & b_5&=\alpha_5+180°=\ 0°00',\end{aligned}$$

and the synthesis is complete. Notice that by using Eqs. (16) and (13) on the fifth stage instead of Eqs. (15) and (12), we have eliminated the need for an optical compensator.

We now compare the amplitudes of the spectral transmittances of the networks of Parts I and II. The network of Part I has $C_5'=C_5=0.01621$. The C_5' for the network of Part II is found from Eq. (25) to be $C_5'=0.01018$. Thus we see that the amplitude-transmittance function of the network of Part I is greater than that of the corresponding network of Part II by a factor of 1.592.

Let us now turn to the second example. Suppose we wish to obtain the amplitude-transmittance function,

$$\begin{aligned}K(\omega)&=\tfrac{1}{8}(\cos a\omega+\cos 3a\omega+\cos 5a\omega+\cos 7a\omega+\cos 9a\omega+\cos 11a\omega+\cos 13a\omega+\cos 15a\omega) \quad (30)\\&=\tfrac{1}{16}(e^{i15a\omega}+e^{-i13a\omega}+e^{+i11a\omega}+e^{+i9a\omega}+e^{-i7a\omega}+e^{+i5a\omega}+e^{+i3a\omega}+e^{+ia\omega}+e^{-ia\omega}+e^{-i3a\omega}+e^{-i5a\omega}+e^{-i7a\omega}+e^{-i9a\omega}+e^{-i11a\omega}+e^{-i13a\omega}+e^{-i15a\omega}).\end{aligned}$$

Equation (30) is plotted in Fig. 4 where it is seen that $K(\omega)$ has the form of a bandpass-filter characteristic. To put the amplitude-transmittance function in the form of Eq. (2), we multiply Eq. (30) by $e^{-i15a\omega}$ and let

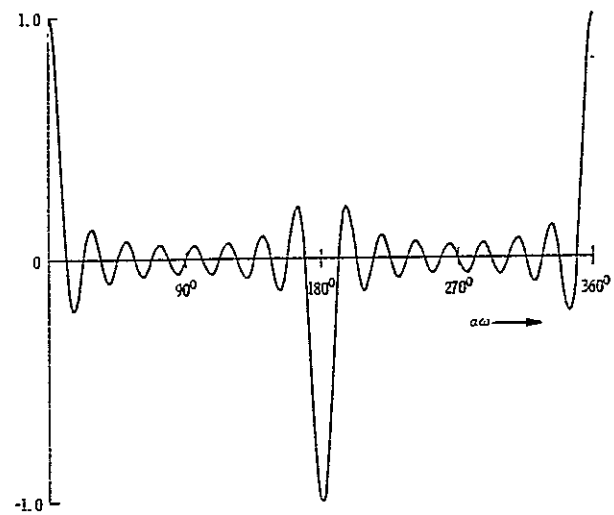


FIG. 4. Amplitude-transmittance function of second example.

$b\omega=2a\omega$, which gives

$$C(\omega)=\tfrac{1}{16}(1+e^{-ib\omega}+e^{-i2b\omega}+e^{-i3b\omega}+e^{-i4b\omega}+e^{-i5b\omega}+e^{-i6b\omega}+e^{-i7b\omega}+e^{-i8b\omega}+e^{-i9b\omega}+e^{-i10b\omega}+e^{-i11b\omega}+e^{-i12b\omega}+e^{-i13b\omega}+e^{-i14b\omega}+e^{-i15b\omega}). \quad (31)$$

Solving for the roots of Eq. (31), we find

$$\begin{aligned}
&|z_1|=1, \quad \alpha_1=22.5^\circ, && |z_8|=1, \quad \alpha_8=180.0^\circ,\\
&|z_2|=1, \quad \alpha_2=45.0^\circ, && |z_9|=1, \quad \alpha_9=202.5^\circ,\\
&|z_3|=1, \quad \alpha_3=67.5^\circ, && |z_{10}|=1, \quad \alpha_{10}=225.0^\circ,\\
&|z_4|=1, \quad \alpha_4=90.0^\circ, && |z_{11}|=1, \quad \alpha_{11}=247.5^\circ,\\
&|z_5|=1, \quad \alpha_5=112.5^\circ, && |z_{12}|=1, \quad \alpha_{12}=270.0^\circ,\\
&|z_6|=1, \quad \alpha_6=135.0^\circ, && |z_{13}|=1, \quad \alpha_{13}=292.5^\circ,\\
&|z_7|=1, \quad \alpha_7=157.5^\circ, && |z_{14}|=1, \quad \alpha_{14}=315.0^\circ,\\
& && |z_{15}|=1, \quad \alpha_{15}=337.5^\circ.
\end{aligned}$$

The fact that all roots have the same magnitude suggests that we should look carefully for possible groupings of roots which will result in fewer stages for the network. We find that the roots can be grouped in the following fashion:

$r=1$ and $r=8$

$$\begin{aligned}
&|z_8|=1, \quad \alpha_8=180^\circ, && |z_1|=1, \quad \alpha_1=22.5^\circ,\\
&r=2 && |z_3|=1, \quad \alpha_3=67.5^\circ,\\
&|z_4|=1, \quad \alpha_4=90^\circ, && |z_5|=1, \quad \alpha_5=112.5^\circ,\\
&|z_{12}|=1, \quad \alpha_{12}=270^\circ, && |z_7|=1, \quad \alpha_7=157.5^\circ,\\
&r=4 && |z_9|=1, \quad \alpha_9=202.5^\circ,\\
&|z_2|=1, \quad \alpha_2=45^\circ, && |z_{11}|=1, \quad \alpha_{11}=247.5^\circ,\\
&|z_6|=1, \quad \alpha_6=135^\circ, && |z_{13}|=1, \quad \alpha_{13}=292.5^\circ,\\
&|z_{10}|=1, \quad \alpha_{10}=225^\circ, && |z_{15}|=1 \quad \alpha_{15}=337.5^\circ.\\
&|z_{14}|=1, \quad \alpha_{14}=315^\circ,
\end{aligned}$$

Each group consists of r roots of the same amplitude whose phase angles differ successively by $2\pi/r$. This means that the 16 factors can be combined to give four factors,

$$C(\omega)=\tfrac{1}{16}(1+e^{-ib\omega})(1+e^{-i2b\omega})(1+e^{-i4b\omega})(1+e^{-i8b\omega}). \quad (32)$$

Thus, the required network consists of only four stages. The length of the crystals in these stages will be

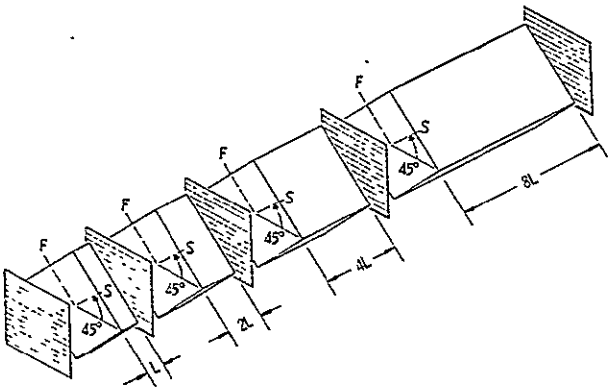


FIG. 5. Birefringent network obtained in second example (4-stage Lyot filter).

L, $2L$, $4L$, and $8L$, where $L=bc/\Delta\eta$. The four zeros of Eq. (32) are

$$\begin{aligned}
&|z_1|=1, && \alpha_1=180^\circ,\\
&|z_2|=1, && \alpha_2=180^\circ,\\
&|z_3|=1, && \alpha_3=180^\circ,\\
&|z_4|=1, && \alpha_4=180^\circ.
\end{aligned} \quad (33)$$

Using Eqs. (13) to calculate the b_i and Eqs. (16) to calculate the θ_i and β_i, we obtain

$$\begin{aligned}
&\theta_1=-\beta_1=\tan^{-1}|z_1|^{\frac{1}{2}}=45^\circ, && b_1=\alpha_1+180^\circ=0^\circ,\\
&\theta_2=-\beta_2=\tan^{-1}|z_2|^{\frac{1}{2}}=45^\circ, && b_2=\alpha_2+180^\circ=0^\circ,\\
&\theta_3=-\beta_3=\tan^{-1}|z_3|^{\frac{1}{2}}=45^\circ, && b_3=\alpha_3+180^\circ=0^\circ,\\
&\theta_4=-\beta_4=\tan^{-1}|z_4|^{\frac{1}{2}}=45^\circ, && b_4=\alpha_4+180^\circ=0^\circ,
\end{aligned}$$

and the synthesis is complete. Optical compensators are not required (in theory) on any of the stages of this network since all b_i are zero. The synthesized network is shown in Fig. 5 and is recognized to be a four-stage Lyot filter.[7] Thus it is interesting to note that the Lyot filter can be obtained by use of the synthesis procedure of Part II, while the Solc filter can be obtained via the synthesis procedure of Part I.

ACKNOWLEDGMENTS

The authors are grateful to S. E. Harris for several contributions to this work, and to W. E. Bicknell for valuable discussions. Also acknowledged are the encouragement and support given by B. J. McMurtry.

[7] J. W. Evans, J. Opt. Soc. Am. 39, 229 (1949).

Reprinted from JOURNAL OF THE OPTICAL SOCIETY OF AMERICA, Vol. 56, No. 12, 1746–1754, December 1966
Printed in U. S. A

Optical Network Synthesis Using Birefringent Crystals. V. Synthesis of Lossless Networks Containing Equal-Length Crystals and Compensators*

E. O. AMMANN AND J. M. YARBOROUGH
Electronic Defense Laboratories, Sylvania Electronic Systems, Mountain View, California 94040
(Received 23 June 1966)

Part I of this series reported a procedure for synthesizing birefringent networks having a prescribed amplitude transmittance. The desired transmittance $C(\omega)$ was written as $C(\omega)=C_0+C_1e^{-i\omega}+C_2e^{-i2\omega}+\cdots+C_ne^{-in\omega}$, where the C_i could be arbitrarily chosen as long as each was real. The synthesis procedure of this paper is a generalization of the procedure of Part I and allows for the realization of $C(\omega)$ having complex C_i. The resulting network consists of n stages between an input and output polarizer, with each stage containing a birefringent crystal and (achromatic) optical compensator. The form of this network is essentially the same as the practical form of the network obtained from Part I, and hence the additional versatility is obtained at no extra cost in network complexity.

INDEX HEADINGS: Polarization; Crystals; Filters; Birefringence.

I. INTRODUCTION

PART I of this series[1] described a procedure for synthesizing birefringent networks whose amplitude transmittance could be specified. The purpose of this paper is to describe a generalization of that procedure which provides still greater flexibility in the synthesis of birefringent networks.

The procedure of Part I allows the realization of a birefringent network whose amplitude transmittance $C(\omega)$ is of the form,

$$C(\omega)=C_0+C_1e^{-i\omega}+C_2e^{-i2\omega}+\cdots+C_ne^{-in\omega}. \quad (1)$$

The number of terms employed in $C(\omega)$ is finite but arbitrary. The choice of the coefficients (the C_i) is also arbitrary as long as each C_i is real. The form of the network obtained from the synthesis procedure of Part I is shown in Fig. 1. The network consists of a series of identical cascaded birefringent crystals between an input and output polarizer. The network may be thought of as composed of several stages, with each stage consisting of one birefringent crystal. A network containing n stages is required for a $C(\omega)$ having $n+1$ terms. Once $C(\omega)$ has been chosen, the rotation angles (the ϕ_i) of the crystals and the output polarizer can be calculated from the synthesis procedure.

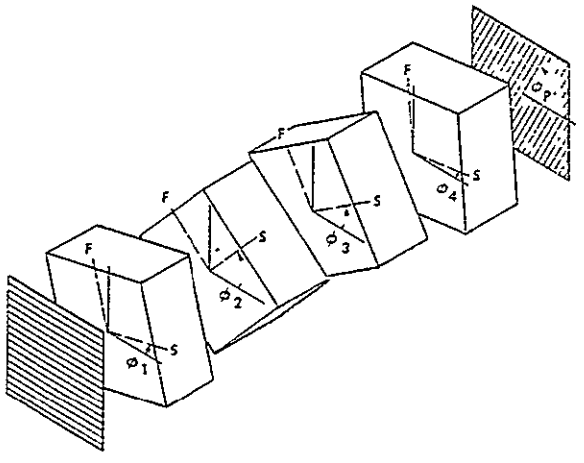


FIG. 1. Basic configuration of birefringent network (4 stages) obtained from the synthesis procedure of Part I. F and S denote the "fast" and "slow" axes of the birefringent crystals.

The synthesis procedure of this paper allows greater freedom in the choice of $C(\omega)$ and results in a network whose basic form is shown in Fig. 2. The desired amplitude transmittance $C(\omega)$ is still written in the form of Eq. (1), but the C_i may now be complex. An n-stage network is again required to realize a $C(\omega)$ having $n+1$ terms, but each stage now consists of an optical compensator[2] and a birefringent crystal. The synthesis procedure determines the rotation angle of each crystal, the retardation introduced by each compensator, and the rotation angle of the output polarizer.

The networks of Part I have been termed lossless birefringent networks since there are no energy-dissipating components between the input and output polarizers. The networks of this paper are lossless in the same sense, since no internal polarizers are required.

The following sections contain a description of the synthesis procedure and give an example of its application.

II. SYNTHESIS PROCEDURE

A. General

The object of the synthesis procedure is to find the n birefringent-crystal angles, the retardations of the $n+1$ optical compensators, and the output-polarizer angle which result in the desired amplitude transmittance $C(\omega)$. For a given $C(\omega)$, $2n+2$ network parameters are to be determined. This matches the number of quantities in $C(\omega)$ which we are free to choose, for we may specify the real and imaginary parts of the $n+1$ coefficients C_i. The length L of the crystals (all crystals have the same length) is determined by the periodicity of the desired amplitude transmittance.

* Work supported by the National Aeronautics and Space Administration under Contract NAS8-20570.

[1] S. E. Harris, E. O. Ammann, and I. C. Chang, J. Opt. Soc. Am. 54, 1267 (1964).

[2] H. G. Jerrard, J. Opt. Soc. Am. 38, 35 (1948).

The notation, conventions, and approaches used here follow closely those used in Part I. Hence for brevity it is assumed that the reader is familiar with that reference, and much of the information contained therein is not repeated here. Because of this, an understanding of Part I is important to the understanding of this paper.

In this paper, optical compensators play an important role. A compensator is used with each crystal of the network and with the output polarizer. Since compensators were not required (in theory) in Part I and hence were not discussed, we briefly describe their operation and analysis. Optical compensators behave essentially the same as very short birefringent crystals. A compensator introduces a phase difference of b radians (where $0<b<2\pi$) between slow-axis (S) and fast-axis (F) components. It is assumed that this phase difference is independent of ω, an assumption which is approximately valid for most cases of interest. If this assumption is valid, light passing through the compensator polarized in the S direction is operated upon by e^{-jb}, while light polarized in the F direction is operated upon by unity.

We assume in this paper (as in Part I) that the birefringent crystals and optical compensators of the network are lossless. This means that energy must be conserved at all points within the network between the input and output polarizers. Energy conservation places certain important restrictions on the F_i and S_i, and on the C_i and D_i. These restrictions are derived and listed in Appendix B.

As in Part I, it is convenient to deal with relative angles (θ_i) of the stages instead of absolute angles (ϕ_i). By relative angle, we mean the additional angle of rotation measured from the preceding stage. The relative angles are given in terms of the ϕ_i of Fig. 2 by

$$\begin{aligned}\theta_1&=\phi_1,\\ \theta_2&=\phi_2-\phi_1,\\ &\vdots\\ \theta_n&=\phi_n-\phi_{n-1},\\ \theta_p&=\phi_p-\phi_n.\end{aligned}$$

B. Procedure

As mentioned in Part I, a useful approach to the synthesis problem is to consider the impulse response of the network. Since the inverse Fourier transform of the amplitude transmittance of a network yields its impulse response, we obtain, by taking the inverse Fourier transform of Eq. (1), the impulse response of the network of Fig. 2:

$$C(t)=C_0\delta(t)+C_1\delta(t-a)+C_2\delta(t-2a)+\cdots+C_n\delta(t-na). \quad (2)$$

Thus the impulse response of our network consists of a series of equally spaced impulses whose areas are given by the C_i. Since the C_i are complex, the impulse response is also complex. The explanation of this apparent paradox and its significance is given in Sec. III.

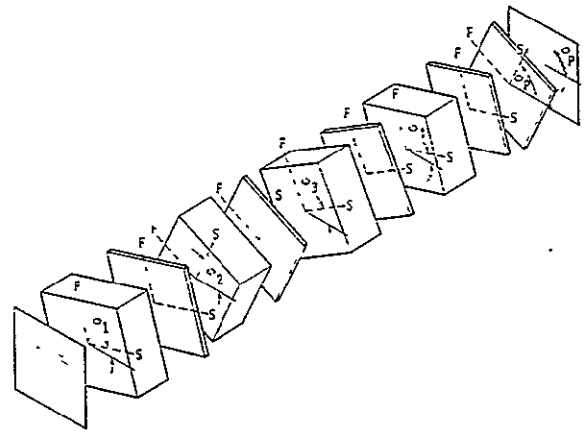


FIG. 2. Basic configuration of birefringent network (4 stages) obtained from the synthesis procedure of this paper.

In the synthesis, we begin with the desired $C(\omega)$ as given by Eq. (1). This is equivalent to prescribing the impulse reponse $C(t)$ of the network. We next proceed from the last component of the network (the output polarizer) back to the first (the input polarizer), calculating the impulse trains which exist at all intermediate points. The areas of the individual impulses of these trains are denoted by the F_i^j and S_i^j of Fig. 3, where the F_i^j impulses are polarized along the fast axis of the preceding (jth) crystal and the S_i^j impulses along the slow axis. In the course of calculating these impulse trains, the crystal angles, compensator delays, and output polarizer angle are determined.

Assume that $C(\omega)$ and therefore the desired C_i of Eqs. (1) and (2) have been chosen. We must next find the signal $D(\omega)$ which is polarized perpendicular to $C(\omega)$ and therefore is stopped by the output polarizer. Since the network is lossless (between the input and output polarizers), the signal energy entering the first crystal must equal the sum of the energies in the $C(\omega)$ and $D(\omega)$ outputs. In equation form, this gives[3]

$$C(\omega)C^*(\omega)+D(\omega)D^*(\omega)=(I_0^0)^2, \quad (3a)$$

where I_0^0 is the area of the impulse which is incident upon the first crystal. Rewriting this, we have

$$D(\omega)D^*(\omega)=(I_0^0)^2-C(\omega)C^*(\omega). \quad (3b)$$

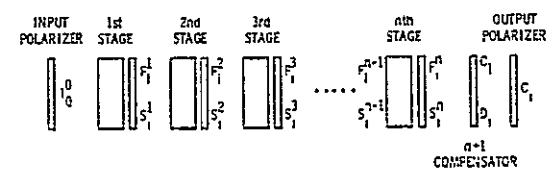


FIG. 3. n-stage network. Each stage contains a birefringent crystal and optical compensator.

[3] Asterisks are used in this paper to denote the complex conjugate of a quantity.

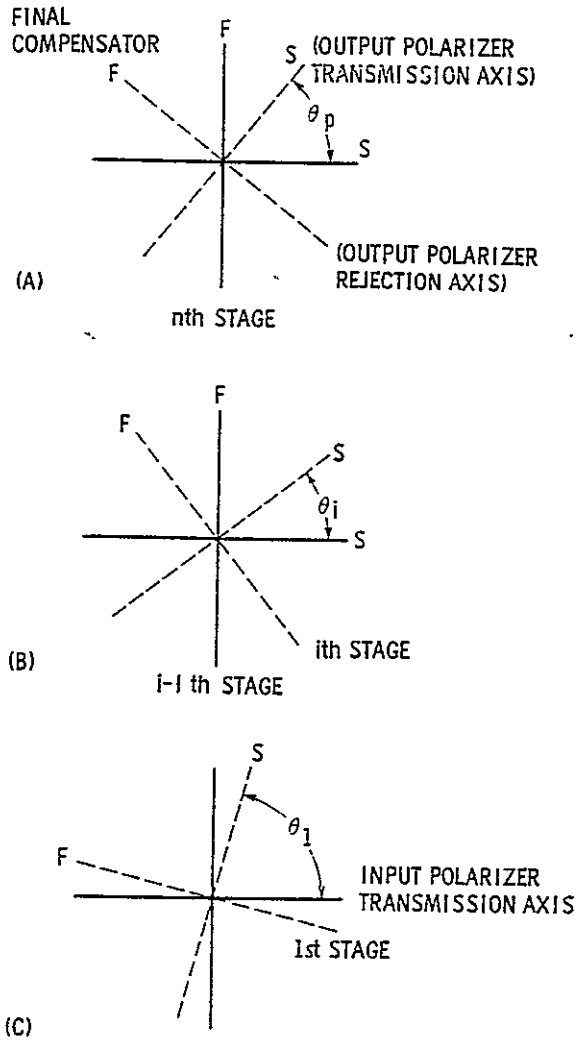


FIG. 4. Angle conventions used in the synthesis procedure: (a) final compensator (and output polarizer); (b) ith stage; (c) input polarizer.

We are now ready to choose a value for I_0^0. The left side of (3b) must be nonnegative for all frequencies; thus $(I_0^0)^2$ must be chosen greater than, or equal to, the maximum value of $C(\omega)C^*(\omega)$. Having chosen I_0^0, we can calculate $D(\omega)$ from $D(\omega)D^*(\omega)$ using the method given in Appendix A.

Doing this, we obtain $D(\omega)$ in the form

$$D(\omega)=D_0'+D_1'e^{-i\omega}+D_2'e^{-i2\omega}+\cdots+D_n'e^{-in\omega},$$

where the D_i' are in general complex. It is important to note, however, that if $D(\omega)$ is a solution of Eq. (3b), then $e^{i\mu}D(\omega)$ is also a solution. Hence a more general solution for $D(\omega)$ is

$$\begin{aligned}D(\omega)&=e^{i\mu}[D_0'+D_1'e^{-i\omega}+D_2'e^{-i2\omega}+\cdots\\&\qquad+D_n'e^{-in\omega}]\\&=D_0+D_1e^{-i\omega}+D_2e^{-i2\omega}+\cdots+D_ne^{-in\omega}.\end{aligned}\tag{4}$$

Although the method of Appendix A gives us the values of the D_i', it does not determine a value for μ. The quantity μ must be determined from other considerations and, as is described shortly, has a value which is fixed by the manner in which the synthesis is formulated.

Let us now relate the inputs (the F_i^n and S_i^n) and outputs (the C_i and D_i) of the final compensator. It should be remembered that the F_i^n and S_i^n are components along the fast and slow axes of the preceding (nth) stage while the C_i and D_i are components along the slow and fast axes of the compensator. With the aid of Fig. 4(a), we find

$$\begin{bmatrix}F_i^n\\S_i^n\end{bmatrix}=\begin{bmatrix}\exp(ib_p)\cdot\sin\theta_p & -\cos\theta_p\\ \exp(ib_p)\cdot\cos\theta_p & \sin\theta_p\end{bmatrix}\begin{bmatrix}C_i\\ e^{i\mu}D_i'\end{bmatrix},\tag{5}$$

where θ_p is the relative angle of the final compensator (and hence also of the output polarizer), and b_p is the compensator delay.

We must next determine the quantities μ, θ_p, and b_p. To do this, we derive and solve three simultaneous equations. The first of these equations is obtained by noting that the first impulse to leave the nth stage must have a real area. This is equivalent to stating that F_0^n must be real. This condition arises from our convention of Sec. IIA which states that light passing through a compensator polarized in the S direction is operated upon by e^{-ib} while light polarized in the F direction is operated upon by unity. Since the first impulse to leave the nth stage must have been polarized along the F axis of each preceding stage, this impulse will have been operated upon by unity in each compensator and will therefore be real. From Eq. (5) we obtain for F_0^n

$$F_0^n=\exp(ib_p)\cdot(\sin\theta_p)\cdot C_0-e^{i\mu}\cdot(\cos\theta_p)\cdot D_0'.$$

Equating the imaginary parts of the left- and right-hand sides of this equation, we obtain the first of our three desired equations,

$$\begin{aligned}0=&\sin\theta_p[\mathrm{Im}(C_0)\cos b_p+\mathrm{Re}(C_0)\sin b_p]\\&-\cos\theta_p[\mathrm{Im}(D_0')\cos\mu+\mathrm{Re}(D_0')\sin\mu],\end{aligned}\tag{6a}$$

where Im and Re denote the imaginary and real parts of the quantity in question. The remaining two equations result because the first and last impulses leaving the nth stage must have been polarized along its fast and slow axes, respectively. This means that

$$F_n^n=S_0^n=0,$$

which, with (5), gives

$$\exp[i(b_p-\mu)]\cdot\tan\theta_p=D_n'/C_n\tag{6b}$$

and

$$\exp[-i(b_p-\mu)]\cdot\tan\theta_p=C_0/D_0'.\tag{6c}$$

Taking the complex conjugate of both sides of Eq. (6c), we obtain

$$\exp[i(b_p-\mu)]\cdot\tan\theta_p=-(C_0^*/D_0'^*).$$

Combining this equation with Eq. (6b), we obtain

$$C_0{}^*C_n+D_0{}'^*D_n'=0, \tag{7a}$$

the relation which must be true if Eqs. (6b) and (6c) are to be satisfied simultaneously. Noting that $D_i'=e^{-i\mu}D_i$, we can rewrite (7a) as

$$C_0{}^*C_n+D_0{}^*D_n=0. \tag{7b}$$

But Eq. (7b) is automatically satisfied from conservation of energy since it is identical to Eq. (B9) of Appendix B.

Since the C_i and D_i' are complex, we can rewrite (6b) in the form

$$\exp[i(b_p-\mu)]\cdot\tan\theta_p=|D_n'/C_n|\exp(i\alpha_p), \tag{8}$$

where in (8) we have expressed (D_n'/C_n) as a magnitude and phase angle. It is apparent from (8) that the rotation angle θ_p of the polarizer and compensator should be chosen to be

$$\tan\theta_p=|D_n'/C_n|. \tag{9}$$

By further manipulations of Eqs. (6a), (6b), and (6c), we obtain

$$\tan b_p=-\mathrm{Im}(C_0)/\mathrm{Re}(C_0) \tag{10}$$

and

$$\mu=b_p-\alpha_p. \tag{11}$$

Having determined α_p, b_p, and μ, we can substitute these values into (5) to obtain $F_i{}^n$ and $S_i{}^n$, the outputs along the fast and slow axes of the nth stage. We must next find the rotation angles and compensator delays of the n stages of the network.

To do this, we write expressions relating the input and output of each stage. With the help of Figs. 4(b) and 4(c), we obtain

First Stage

$$\begin{bmatrix}F_0{}^1\\S_1{}^1\end{bmatrix}=\begin{bmatrix}-\sin\theta_1\\\exp(-ib_1)\cdot\cos\theta_1\end{bmatrix}[I_0{}^0], \tag{12a}$$

Second Stage

$$\begin{bmatrix}F_0{}^2\\F_1{}^2\\S_1{}^2\\S_2{}^2\end{bmatrix}=\begin{bmatrix}\cos\theta_2&0\\0&-\sin\theta_2\\\exp(-ib_2)\cdot\sin\theta_2&0\\0&\exp(-ib_2)\cdot\cos\theta_2\end{bmatrix}\begin{bmatrix}F_0{}^1,\\S_1{}^1\end{bmatrix} \tag{12b}$$

Third Stage

$$\begin{bmatrix}F_0{}^3\\F_1{}^3\\F_2{}^3\\S_1{}^3\\S_2{}^3\\S_3{}^3\end{bmatrix}=\begin{bmatrix}\cos\theta_3&0&0&0\\0&\cos\theta_3&-\sin\theta_3&0\\0&0&0&-\sin\theta_3\\\exp(-ib_3)\cdot\sin\theta_3&0&0&0\\0&\exp(-ib_3)\cdot\sin\theta_3&\exp(-ib_3)\cdot\cos\theta_3&0\\0&0&0&\exp(-ib_3)\cdot\cos\theta_3\end{bmatrix}\begin{bmatrix}F_0{}^2\\F_1{}^2\\S_1{}^2\\S_2{}^2\end{bmatrix}, \tag{12c}$$

and in general,

jth Stage

$$\begin{bmatrix}F_0{}^j\\F_1{}^j\\F_2{}^j\\\vdots\\F_{j-1}{}^j\\F_j{}^j\\S_1{}^j\\S_2{}^j\\S_3{}^j\\\vdots\\S_{j-1}{}^j\\S_j{}^j\end{bmatrix}=\begin{bmatrix}\cos\theta_j&0&0&\cdots&0&0&0\\0&\cos\theta_j&0&\cdots&0&0&0\\0&0&\cos\theta_j&\cdots&0&0&0\\\vdots&\vdots&\vdots&&\vdots&\vdots&\vdots\\0&0&0&\cdots&-\sin\theta_j&0&0\\0&0&0&\cdots&0&-\sin\theta_j&0\\0&0&0&\cdots&0&0&-\sin\theta_j\\\exp(-ib_j)\cdot\sin\theta_j&0&0&\cdots&0&0&0\\0&\exp(-ib_j)\cdot\sin\theta_j&0&\cdots&0&0&0\\0&0&\exp(-ib_j)\cdot\sin\theta_j&\cdots&0&0&0\\\vdots&\vdots&\vdots&&\vdots&\vdots&\vdots\\0&0&0&\cdots&\exp(-ib_j)\cdot\cos\theta_j&0&0\\0&0&0&\cdots&0&\exp(-ib_j)\cdot\cos\theta_j&0\\0&0&0&\cdots&0&0&\exp(-ib_j)\cdot\cos\theta_j\end{bmatrix}\begin{bmatrix}F_0{}^{j-1}\\F_1{}^{j-1}\\F_2{}^{j-1}\\\vdots\\F_{j-2}{}^{j-1}\\F_{j-1}{}^{j-1}\\S_1{}^{j-1}\\S_2{}^{j-1}\\S_3{}^{j-1}\\\vdots\\S_{j-2}{}^{j-1}\\S_{j-1}{}^{j-1}\end{bmatrix}. \tag{12d}$$

Putting $j=n$ in (12d), we have the input and output relations for the nth stage. We know the output (the $F_i{}^n$ and $S_i{}^n$) and wish to find θ_n, b_n, and the input. As discussed in detail in Part I, an input exists which produces our given output provided that

$$\exp(ib_n)\cdot\tan\theta_n=-F_{n-1}{}^n/S_n{}^n=|F_{n-1}{}^n/S_n{}^n|\times\exp(i\alpha_n) \tag{13a}$$

and

$$F_0{}^{n*}F_{n-1}{}^n+S_1{}^{n*}S_n{}^n=0. \tag{13b}$$

Note that α_n includes the effect of the minus sign which precedes $F_{n-1}{}^n/S_n{}^n$.

We can satisfy Eq. (13a) by properly choosing b_n and θ_n, while (13b) is automatically satisfied by conservation of energy. Knowing b_n and θ_n, we can then calculate the input to the nth stage from (12d). This, of course, is also the output from the $n-1$ stage; hence we can repeat the procedure just described to determine b_{n-1} and θ_{n-1}. In this fashion, we can work our way back through the entire network until all rotation angles

and compensator delays have been determined. The general equations for the jth stage are

$$\exp(ib_j)\cdot\tan\theta_j=-F_{j-1}{}^j/S_j{}^j=|F_{j-1}{}^j/S_j{}^j|\times\exp(i\alpha_j), \quad (14a)$$

and

$$F_0{}^{j*}F_{j-1}{}^j+S_1{}^{j*}S_j{}^j=0, \quad (14b)$$

which gives

$$b_j=\alpha_j \quad (15a)$$

and

$$\tan\theta_j=|F_{j-1}{}^j/S_j{}^j|. \quad (15b)$$

As seen from Appendix B, Eq. (14b) is always automatically satisfied by conservation of energy.

Note that if $\alpha_j=0$, a compensator is not required (in theory) for that particular stage. Furthermore it is possible to eliminate the compensator from a stage which has $\alpha_j=\pi$. This is because when $\alpha_j=\pi$, an alternate solution to Eq. (14a) is

$$b_j=0, \quad (15c)$$

and

$$\tan\theta_j=-|F_{j-1}{}^j/S_j{}^j|. \quad (15d)$$

Hence whenever $\alpha_j=\pi$, Eqs. (15c) and (15d), rather than (15a) and (15b), should be used to determine b_j and θ_j.

We now have sufficient information to synthesize a birefringent network. The procedure to be followed is summarized below.

C. Summary of Synthesis Procedure

(1) Choose the desired amplitude transmittance $C(\omega)$ and write it in the form of Eq. (1). The C_i may be complex.

(2) The required length L for all crystals is given by $L=ac/\Delta\eta$, where c is the velocity of light in a vacuum and $\Delta\eta$ is the difference between the extraordinary and ordinary indices of refraction of the crystal. The quantity a is determined by comparing $C(\omega)$ as written in step (1) to $C(\omega)$ as given by Eq. (1).

(3) Choose a (real) value for $I_0{}^0$. The choice is arbitrary as long as $(I_0{}^0)^2$ is greater than or equal to the maximum magnitude of $C(\omega)C^*(\omega)$.

(4) Calculate $D(\omega)D^*(\omega)$ from Eq. (3b). Use the method of Appendix A to solve for $D(\omega)$ from $D(\omega)D^*(\omega)$. This gives the D_i' of Eq. (4), but does not determine μ. Several different $D(\omega)$ result, and each of these, when used with $C(\omega)$ results in an acceptable network. The D_i' of these $D(\omega)$ are, in general, complex. The remaining steps should be carried out for each $D(\omega)$.

(5) Calculate the rotation angle θ_p of the output polarizer and final compensator from Eq. (9), the phase delay b_p of the final compensator from Eq. (10), and μ from Eq. (11).

(6) Calculate the $F_i{}^n$ and $S_i{}^n$ from Eq. (5).

(7) Using Eq. (15b), calculate the rotation angle θ_n of the last stage. The compensator delay b_n for that stage should be computed from (15a). Using Eqs. (C1) and (C2), calculate the input to the last stage (which is the output from the preceding stage).

(8) Repeat the procedure of step (7) on each succeeding stage until the rotation angle and compensator delay of each stage have been determined. If $\alpha_j=\pi$ for a particular stage, Eqs. (15c) and (15d) rather than (15a) and (15b) should be used to calculate b_j and θ_j.

III. DISCUSSION

We now consider the implications of being able to choose C_i which are complex. In Part I, we were limited to amplitude transmittances having all C_i real. This meant that we were limited to $C(\omega)$'s whose real parts were even and whose imaginary parts were odd. These restrictions have now been removed; the real and imaginary portions of $C(\omega)$ can now be asymmetrical.

An objection might be raised that since the C_i are complex, our network has an impulse response, given by Eq. (2), which is complex; but it is well known that the impulse response of a physical network must be real. This dilemma arises because our theory requires the use of achromatic optical compensators in the network. The theory assumes that these compensators introduce a delay which is independent of ω. Such a delay is not realizable in practice. Compensators can approximate this behavior over a limited frequency range however. Hence the response of the synthesized network closely approximates $C(\omega)$ over the frequency range for which the compensators may be considered achromatic. Outside of this frequency range, the transmittance departs from $C(\omega)$. Since birefringent networks are ordinarily designed for use over a limited frequency range, this is an acceptable situation.

Thus we see that $C(\omega)$ accurately describes the network's transmittance over only a limited spectral range. But when we take the inverse Fourier transform of (1) to obtain the impulse response given by (2), we are (incorrectly) assuming that Eq. (1) is valid for all possible values of ω. Hence it is not surprising that the result is a complex impulse response for the network. Even though (2) does not accurately give the network impulse response, the time-domain approach is very useful for visualizing and understanding the synthesis procedure.

Part II of this series[4] described a second synthesis procedure which achieved the same goal as the procedure of Part I, but via a different form of birefringent network. Moreover, the procedure of Part II can be used when complex C_i are present in $C(\omega)$. The network which results, however, contains internal polarizers and hence is not a "lossless" network. For that reason, the network of this paper is preferable to that of Part II for most applications.

The network resulting from the synthesis procedure of this paper contains an optical compensator next to

[4] E. O. Ammann and I. C. Chang, J. Opt. Soc. Am. 55, 835 (1965).

the output polarizer. In practice, it is often possible to remove this optical compensator. Suppose for example that we have synthesized a network which has a desired $C(\omega)$. If we now remove the final compensator from that network, the new transmittance is $\exp(ib_p)\cdot C(\omega)$. Thus the new transmittance differs from the desired transmittance by only a constant phase factor. Often the introduction of this phase factor is of no consequence, and hence the final compensator can be removed. Furthermore, we note from Eq. (10) that if C_0 is chosen to be real, $b_p=0$ and the need for the final compensator is automatically eliminated.

Finally, we note that (as seen in Figs. 1 and 2) the network of this paper contains a greater number of components than the network of Part I. It should be emphasized, however, that Figs. 1 and 2 show the networks predicted by theory. In practice, the network of Part I requires the use of an optical compensator with each crystal of the network to compensate for slightly incorrect crystal lengths. Thus the practical forms of the networks of this paper and of Part I are identical; the additional flexibility is obtained at no expense in actual network complexity. In this paper, each optical compensator serves the dual functions of (a) introducing the delay required by theory, and (b) compensating for incorrect crystal length.

IV. EXAMPLE

A sample calculation is performed to illustrate the synthesis procedure of Sec. II. Suppose we wish to approximate the real transfer function $G(\omega)$ shown in Fig. 5. Since $G(\omega)$ is neither even nor odd, complex coefficients are required in the approximating exponential series. For this example we use a seven-term complex Fourier series.

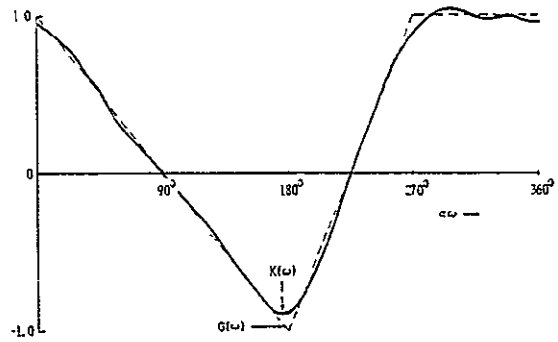


FIG. 5. Ideal and approximating amplitude transmittances of example. Ideal transmittance is shown by dotted line and approximating transmittance by solid line.

The Fourier-series approximation to the ideal transfer $G(\omega)$ is given by

$$K(\omega)=(1/\pi^2)[(\tfrac{1}{9}-i2/9)e^{i3a\omega}-e^{i2a\omega}+(\tfrac{1}{4}+i2)e^{ia\omega}+\pi^2/4+(\tfrac{1}{4}-i2)e^{-ia\omega}-e^{-i2a\omega}+(\tfrac{1}{9}+i2/9)e^{-i3a\omega}], \quad (16)$$

which is plotted in Fig. 5. Following the method of Part I, we convert this noncausal approximating function to a causal function by multiplying by $e^{-i3a\omega}$, which gives

$$C(\omega)=e^{-i3a\omega}K(\omega)=(1/\pi^2)[(\tfrac{1}{9}-i2/9)-e^{-ia\omega}+(\tfrac{1}{4}+i2)e^{-i2a\omega}+(\pi^2/4)e^{-i3a\omega}+(\tfrac{1}{4}-i2)e^{-i4a\omega}-e^{-i5a\omega}+(\tfrac{1}{9}+i2/9)e^{-i6a\omega}]. \quad (17)$$

Multiplication by $e^{-i3a\omega}$ is equivalent to introducing a pure time delay in the time domain, and thus the impulse response and transfer function are essentially unchanged. Since the series contains seven terms, the synthesized network contains six stages.

We now calculate $D(\omega)$. From Eq. (3b) we have

$$|D(\omega)|^2=D(\omega)D^*(\omega)=(I_0^0)^2-C(\omega)C^*(\omega)=(I_0^0)^2-0.44257-(0.11139+i0.14695)e^{ia\omega}-(0.11139-i0.14695)e^{-ia\omega}-(0.09990+i0.12775)e^{i2a\omega}-(0.09990-i0.12775)e^{-i2a\omega}-(-0.05961-i0.05232)e^{i3a\omega}-(-0.05961+i0.05232)e^{-i3a\omega}-0.05589e^{i4a\omega}-0.05589e^{-i4a\omega}-(-0.00913+i0.00456)e^{i5a\omega}-(-0.00913-i0.00456)e^{-i5a\omega}-(0.00152-i0.00203)e^{i6a\omega}-(0.00152+i0.00203)e^{-i6a\omega}. \quad (18)$$

The area I_0^0 of the input impulse must now be chosen in order to obtain $|D(\omega)|^2$. It may have any real value as long as $(I_0^0)^2$ is larger than the maximum value of $C(\omega)C^*(\omega)$. The maximum of $C(\omega)C^*(\omega)$ has been calculated to be 1.035. Thus let us choose $I_0^0=1.050$, which gives $(I_0^0)^2=1.1025$. Equation (18) then becomes, after making the substitution $x=e^{-ia\omega}$,

$$|D(\omega)|^2=-(0.00152+i0.00203)x^6-(-0.00913-i0.00456)x^5-0.05589x^4-(-0.05961+i0.05232)x^3-(0.09990-i0.12775)x^2-(0.11139-i0.14695)x+0.65993-(0.11139+i0.14695)x^{-1}-(0.09990+i0.12775)x^{-2}-(-0.05961-i0.05232)x^{-3}-0.05589x^{-4}-(-0.00913+i0.00456)x^{-5}-(0.00152-i0.00203)x^{-6}, \quad (19)$$

which is in the form of Eq. (A2). Following the procedure of Appendix A, we find the roots of (19) to be

$$\begin{aligned}
x_1&=0.06608-i0.27538, & (1/x_1)^*&=0.82394-i3.43353,\\
x_2&=-0.09690-i0.27436. & (1/x_2)^*&=-1.14455-i3.24064,\\
x_3&=-0.67656-i0.06373, & (1/x_3)^*&=-1.46526-i0.13704,\\
x_4&=0.17633+i0.17387, & (1/x_4)^*&=2.87546+i2.83537,\\
x_5&=0.57518+i0.17898, & (1/x_5)^*&=1.58510+i0.49323,\\
x_6&=0.59387+i1.30936, & (1/x_6)^*&=0.28729+i0.63342.
\end{aligned}$$

There are 128 (2^{n+1}) possible sets of D_i which can be obtained from these roots. However, sixty-four of these sets are simply negatives of the other sixty-four. We consider only the set that is formed by constructing the polynomial

$$(x-x_1)(x-x_2)(x-x_3)(x-x_4)(x-x_5)(x-x_6).$$

Performing the indicated multiplication, we obtain

$$x^6+(-0.63801-i1.04187)x^5+(0.02599-i0.29300)x^4+(0.06553+i0.44610)x^3 \\ +(-0.23903-i0.05436)x^2+(0.04871-i0.00793)x+(-0.00721+i0.00961). \quad (20)$$

As stated in Eqs. (A9), a set of D_i' is proportional to the coefficients of this polynomial. Evaluating $|q|$ using (A10), we find that

$$|q|=0.45943,$$

and so

$$D_0'=-0.00331+i0.00441, \quad D_3'=0.03011+i0.20496. \quad D_5'=-0.29312-i0.48203,$$
$$D_1'=0.02238-i0.00364, \quad D_4'=0.01194-i0.13461, \quad D_6'=0.45943.$$
$$D_2'=-0.10982-i0.02495,$$

From Eqs. (9), (10), and (11) we may now calculate θ_p, b_p, and μ. The results are

$$\theta_p=83^\circ 45', \quad b_p=0.46365 \text{ rad}, \quad \mu=-5.35589 \text{ rad}.$$

Using Eqs. (A9), we find that

$$D_0=e^{i\mu}D_0'=-0.00552+i0, \quad D_3=e^{i\mu}D_3'=-0.14590+i0.14706, \quad D_5=e^{i\mu}D_5'=0.20976-i0.52372,$$
$$D_1=e^{i\mu}D_1'=0.01634+i0.01572, \quad D_4=e^{i\mu}D_4'=0.11485-i0.07122, \quad D_6=e^{i\mu}D_6'=0.27566+i0.36755,$$
$$D_2=e^{i\mu}D_2'=-0.04593-i0.10282,$$

and hence $D(\omega)$ is completely known. Equation (5) is now used to calculate the F_i^6 and S_i^6, giving

$$\begin{bmatrix} F_0^6 \\ F_1^6 \\ F_2^6 \\ F_3^6 \\ F_4^6 \\ F_5^6 \end{bmatrix} = \begin{bmatrix} 0.05065+i0.00000 \\ -0.09187-i0.04675 \\ 0.27526+i0.37154 \\ 0.23817+i0.09512 \\ 0.43791+i0.00776 \\ -0.11293+i0.01201 \end{bmatrix}, \quad \begin{bmatrix} S_1^6 \\ S_2^6 \\ S_3^6 \\ S_4^6 \\ S_5^6 \\ S_6^6 \end{bmatrix} = \begin{bmatrix} 0.00637+i0.01069 \\ -0.01604-i0.06272 \\ -0.12067+i0.15836 \\ 0.16353-i0.07079 \\ 0.19863-i0.52554 \\ 0.27731+i0.36975 \end{bmatrix}.$$

As a check, we should note that F_0^6 must be real and that F_6^6 and S_0^6 must be zero. As a further check, we can verify that Eq. (14b) is satisfied.

We are now able to calculate θ_6 and b_6, the relative angle of the last stage and the optical compensator delay. Using (15b), we find

$$\theta_6=13^\circ 48'$$

and from (15a),

$$b_6=5.24997 \text{ rad}.$$

The input impulses to the sixth stage are now calculated from Eqs. (C3) and (C4). Equations (15b) and (15a) are then applied again, yielding

$$\theta_5=36^\circ 45'$$

and

$$b_5=6.11153 \text{ rad}.$$

By alternately applying Eqs. (C3) and (C4) and Eqs. (15a) and (15b), we obtain the remaining θ_i and b_i. The summarized results of the synthesis are

$$\begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ \theta_6 \\ \theta_p \end{bmatrix} = \begin{bmatrix} 6^\circ 15' \\ 13^\circ 48' \\ 36^\circ 45' \\ 43^\circ 00' \\ 36^\circ 45' \\ 13^\circ 48' \\ 83^\circ 45' \end{bmatrix}, \quad \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_p \end{bmatrix} = \begin{bmatrix} 2.10838 \\ 2.96994 \\ 0.74123 \\ 0.74123 \\ 2.96994 \\ 5.24997 \\ 0.46365 \end{bmatrix} \text{ radians}.$$

The Jones calculus[5] can be used to verify that these angles and compensator delays give the desired transfer function of Eq. (17).

ACKNOWLEDGMENTS

The authors are grateful to S. Barnard for assistance in performing the calculations of Sec. IV, and to Professor S. E. Harris for a careful reading of the manuscript.

[5] R. C. Jones. J. Opt. Soc. Am. 31, 488 (1941).

APPENDIX A

We describe in this appendix a method for calculating $D(\omega)$ from $|D(\omega)|^2$. The method is similar to that given in Appendix A of Part I, but differs in its details. The differences are necessary because (a) we now begin with a $C(\omega)$ containing complex C_i, and (b) complex values of D_i can now be tolerated in $D(\omega)$.

We begin with the positive semidefinite polynomial

$$\begin{aligned}|D(\omega)|^2=D(\omega)D^*(\omega)&=(I_0^0)^2-C(\omega)C^*(\omega),\\ &=A_ne^{in\alpha\omega}+A_{n-1}e^{i(n-1)\alpha\omega}+\cdots+A_1e^{i\alpha\omega}+A_0\\ &\quad+A_1^*e^{-i\alpha\omega}+\cdots+A_{n-1}^*e^{-i(n-1)\alpha\omega}\\ &\quad+A_n^*e^{-in\alpha\omega}. \quad \text{(A1)}\end{aligned}$$

Letting $x=e^{-i\alpha\omega}$ and reversing the order of the terms, Eq. (A1) becomes

$$\begin{aligned}|D(x)|^2&=A_n^*x^n+A_{n-1}^*x^{n-1}+\cdots+A_1^*x+A_0\\ &\quad+A_1x^{-1}+\cdots+A_{n-1}x^{-(n-1)}+A_nx^{-n}. \quad \text{(A2)}\end{aligned}$$

Assume that x_1 is a root of Eq. (A2). Then

$$\begin{aligned}A_n^*x_1^n+A_{n-1}^*x_1^{n-1}+\cdots+A_1^*x_1+A_0+A_1x_1^{-1}+\cdots\\ +A_{n-1}x_1^{-(n-1)}+A_nx_1^{-n}=0. \quad \text{(A3)}\end{aligned}$$

If we now take the complex conjugate of Eq. (A3), we obtain

$$\begin{aligned}A_n(x_1^*)^n+A_{n-1}(x_1^*)^{n-1}+\cdots\\ +A_1x_1^*+A_0^*+A_1^*(x_1^*)^{-1}+\cdots\\ +A_{n-1}^*(x_1^*)^{-(n-1)}+A_n^*(x_1^*)^{-n}=0. \quad \text{(A4)}\end{aligned}$$

Equation (A4) can be rewritten as

$$\begin{aligned}A_n(1/x_1^*)^{-n}+A_{n-1}(1/x_1^*)^{-(n-1)}+\cdots\\ +A_1(1/x_1^*)^{-1}+A_0^*+A_1^*(1/x_1^*)+\cdots\\ +A_{n-1}^*(1/x_1^*)^{n-1}+A_n^*(1/x_1^*)^n=0. \quad \text{(A5)}\end{aligned}$$

But we now see that (A3) and (A5) have identical coefficients, with x_1 being the variable in Eq. (A3) and $(1/x_1^*)$ the variable in (A5). Thus if x_1 is a root of (A2), then $(1/x_1^*)$ is also a root. One of these two roots is associated with $D(x)$ and the other with $D^*(x)$. Hence we associate half of the roots of Eq. (A2) with $D(x)$ and half with $D^*(x)$. $D(x)$ [and hence $D(\omega)$] can then be constructed (to within a multiplicative phase factor) from a knowledge of its roots.

To summarize, begin with $|D(\omega)|^2$ written in the form of Eq. (A1). The A_i are in general complex. Form the equation

$$\begin{aligned}A_n^*x^n+A_{n-1}^*x^{n-1}+\cdots+A_1^*x+A_0+A_1x^{-1}+\cdots\\ +A_{n-1}x^{-(n-1)}+A_nx^{-n}=0. \quad \text{(A6)}\end{aligned}$$

Solve for the $2n$ roots of this equation. These roots always exist in pairs of the form

$$\begin{array}{ll} x_1, & 1/x_1^*,\\ x_2, & 1/x_2^*,\\ x_3, & 1/x_3^*,\\ \vdots & \vdots\\ x_n, & 1/x_n^*. \end{array} \quad \text{(A7)}$$

Construct all possible equations using one root from each row of (A7). One possible grouping, for example, is

$$\begin{aligned}(x-x_1)(x-x_2)(x-1/x_3^*)\cdots(x-1/x_n^*)\\ =x^n+d_{n-1}x^{n-1}+\cdots+d_2x^2+d_1x+d_0. \quad \text{(A8)}\end{aligned}$$

Each different grouping of roots results in a different $D(\omega)$.

The D_i are proportional to the d_i, where q, the constant of proportionality, is in general complex. Writing q in the form $q=|q|e^{i\mu}$, we obtain

$$\begin{aligned}D_0&=|q|e^{i\mu}d_0=e^{i\mu}D_0',\\ D_1&=|q|e^{i\mu}d_1=e^{i\mu}D_1',\\ &\vdots\\ D_n&=|q|e^{i\mu}d_n=e^{i\mu}D_n'=|q|e^{i\mu}, \quad \text{(A9)}\end{aligned}$$

where

$$D_i'=|q|d_i.$$

The necessity of allowing q to be complex can be seen by noting that if $D(\omega)$ is a solution of Eq. (3b), then $e^{i\mu}D(\omega)$ is also a solution.

The quantity $|q|$ is calculated from

$$|q|^2[d_0d_0^*+d_1d_1^*+\cdots+d_{n-1}d_{n-1}^*+1]=A_0. \quad \text{(A10)}$$

In order to calculate the phase angle μ, however, additional information must be provided. The necessary information is obtained from the restriction that F_0^n must be real, a condition which results from our formulation of the synthesis procedure. With this restriction, μ is uniquely determined (see Sec. IIb) and $D(\omega)$ can be obtained.

Thus the method of this appendix allows us to find $D(\omega)$ to within a multiplicative phase factor $e^{i\mu}$. We obtain values for the D_i', where

$$\begin{aligned}D(\omega)&=e^{i\mu}[D_0'+D_1'e^{-i\alpha\omega}+D_2'e^{-i2\alpha\omega}+\cdots\\ &\quad+D_{n-1}'e^{-i(n-1)\alpha\omega}+D_n'e^{-in\alpha\omega}],\\ &=D_0+D_1e^{-i\alpha\omega}+D_2e^{-i2\alpha\omega}+\cdots\\ &\quad+D_{n-1}e^{-i(n-1)\alpha\omega}+D_ne^{-in\alpha\omega}. \quad \text{(A11)}\end{aligned}$$

APPENDIX B

In this appendix, the restrictions placed upon the F_i and S_i (and upon the C_i and D_i) because of conservation of energy are derived. Consider the ith stage of the network of Fig. 2. Since the network is lossless, the energy in the fast-axis output plus the energy in the slow-axis output of the ith stage must equal the energy incident upon the first stage. Stated mathematically, this gives

$$F^i(\omega)F^{i*}(\omega)+S^i(\omega)S^{i*}(\omega)=(I_0^0)^2. \quad \text{(B1)}$$

If we write out Eq. (B1) and equate the coefficients of corresponding terms, we obtain the equations

$$\begin{aligned}F_0^{i*}F_0^i+F_1^{i*}F_1^i+\cdots+F_{i-1}^{i*}F_{i-1}^i+S_1^{i*}S_1^i\\ +S_2^{i*}S_2^i+\cdots+S_i^{i*}S_i^i=(I_0^0)^2, \quad \text{(B2)}\end{aligned}$$

$$F_0^{i*}F_1^i+F_1^{i*}F_2^i+\cdots+F_{i-2}^{i*}F_{i-1}^i+S_1^{i*}S_2^i+S_2^{i*}S_3^i+\cdots+S_{i-1}^{i*}S_i^i=0, \quad \text{(B3)}$$

$$F_0^{i*}F_2^i+F_1^{i*}F_3^i+\cdots+F_{i-3}^{i*}F_{i-1}^i+S_1^{i*}S_3^i+S_2^{i*}S_4^i+\cdots+S_{i-2}^iS_i^i=0, \quad \text{(B4)}$$

$$\vdots$$

$$F_0^{i*}F_{i-1}^i+S_1^{i*}S_i^i=0. \quad \text{(B5)}$$

$C(\omega)$ and $D(\omega)$ must also satisfy conservation of energy, giving the following restrictions on the C_i and D_i.

$$C_0^*C_0+C_1^*C_1+\cdots+C_n^*C_n+D_0^*D_0+D_1^*D_1+\cdots+D_n^*D_n=(I_0^0)^2, \quad \text{(B6)}$$

$$C_0^*C_1+C_1^*C_2+\cdots+C_{n-1}^*C_n+D_0^*D_1+D_1^*D_2+\cdots+D_{n-1}^*D_n=0, \quad \text{(B7)}$$

$$C_0^*C_2+C_1^*C_3+\cdots+C_{n-2}^*C_n+D_0^*D_2+D_1^*D_3+\cdots+D_{n-2}^*D_n=0, \quad \text{(B8)}$$

$$\vdots$$

$$C_0^*C_n+D_0^*D_n=0. \quad \text{(B9)}$$

APPENDIX C

This appendix gives a systematic and rapid method of calculating the input to a stage, once the output is known. This is simply a formalized procedure of solving for the F^{j-1} and S^{j-1} of (12d) once the F^j and S^j are known. The expressions are similar to those of Appendix C of Part I but differ somewhat due to the complex quantities involved.

We begin by defining F_{j-1}^j and S_j^j in polar form:

$$F_{j-1}^j=|F_{j-1}^j|\exp(if_{j-1}^j) \quad \text{(C1)}$$

$$S_j^j=|S_j^j|\exp(is_j^j). \quad \text{(C2)}$$

Using these definitions, we find the expressions for the F^{j-1} and S^{j-1} in matrix form

$$\begin{bmatrix} F_0^{j-1} \\ F_1^{j-1} \\ \vdots \\ F_{j-1}^{j-1} \end{bmatrix}=\frac{\exp(-is_j^j)}{\{|F_{j-1}^j|^2+|S_j^j|^2\}^{\frac{1}{2}}}\begin{bmatrix} F_0^j & S_1^j \\ F_1^j & S_2^j \\ \vdots & \vdots \\ F_{j-1}^j & S_j^j \end{bmatrix}\begin{bmatrix} S_j^j \\ -F_{j-1}^j \end{bmatrix}. \quad \text{(C3)}$$

$$\begin{bmatrix} S_0^{j-1} \\ S_1^{j-1} \\ \vdots \\ S_{j-1}^{j-1} \end{bmatrix}=\frac{\exp(ib_j)\cdot\exp(is_j^j)}{\{|F_{j-1}^j|^2+|S_j^j|^2\}^{\frac{1}{2}}}\begin{bmatrix} F_0^j & S_1^j \\ F_1^j & S_2^j \\ \vdots & \vdots \\ F_{j-1}^j & S_j^j \end{bmatrix}\begin{bmatrix} F_{j-1}^{j*} \\ S_j^{j*} \end{bmatrix}. \quad \text{(C4)}$$

As before, the calculated values F_{j-1}^{j-1} and S_0^{j-1} should always be zero.

Birefringent Filter for Millimeter Waves

BERNARD M. SCHIFFMAN, MEMBER, IEEE, AND LEO YOUNG, FELLOW, IEEE

Abstract—A scale model ($f_0 = 20$ GHz) of a Sole-type birefringent wave filter for millimeter wavelengths is described. The filter consists of five cascaded identical half-wave plates, or crystals, each composed of an artificial anisotropic dielectric medium with its reference axis tilted at some prescribed angle to the plane of the input polarization. The design and analysis of an individual plate, using Collin's second-order theory of the birefringence of artificial anisotropic dielectrics, and the analysis of multielement filters (filters composed of many plates), aided by Evans' matrix method, are discussed. The experimental filter was tested in the range of 18 to 33 GHz, and its measured performance was found to compare well with the theoretical performance over a major portion of the range of frequencies used in the tests.

A synthesis procedure for optimum (equal-ripple stopband) response multielement filters is given, together with tables of plate angles for such filters. In this procedure, the Dolph approximation and the Harris synthesis are combined.

I. Introduction

A. General

THE SCOPE of this research included the design and testing of a scale model of a plate of artificial birefringent medium for the millimeter-wave region, the construction and testing of an experimental filter composed of several plates, and the investigation of methods of synthesizing optimum configurations of the plates of a filter, with a view toward providing convenient tables of filter designs (plate angles).

Optical filter techniques that were originally developed to exploit the birefringence property of certain naturally occurring transparent crystals (such as quartz and calcite) can be adapted for millimeter or shorter wavelengths. This idea was first proposed by one of the authors.[1]

Birefringent filters were first invented in 1933 by Lyot,[2] a French astronomer. These first filters required crystals or plates of unequal length and lossy polarizers between each pair of adjacent plates. Later, Sole,[3]–[6] in Czechoslovakia, invented the form of filter in which all plates are equal length and only two polarizers, one at each terminal, are required. The birefringent filter has recently been further developed at Stanford.[7],[8] The Sole-type filter is the subject of this article, with emphasis on the design of filters with equal ripples in the stopband.[9]

Optical birefringent filters have very narrow passbands,[10] owing to the large number of optical wavelengths in the path through the birefringent material. The presentation of an exact design theory, plus numerical tables, should therefore be of some significance. The experimental confirmation of theories relating to the design of birefringent filters at millimeter wavelengths was carried out at centimeter wavelengths, and it was thus demonstrated that birefringent filters can also be constructed and operated down to microwave frequencies. Here they are equivalent to directional (nonreflecting) filters, and should be suitable for use in higher-mode-free beam waveguides, particularly at millimeter wavelengths.

Other filter design techniques[11]–[13] common to both optical and microwave technology are not discussed herein.

B. The Birefringent Filter

The birefringent filter is a four-port, intrinsically reflectionless passive device. Its mode of operation is similar to that of a directional filter in microwave technology and is different from conventional reflection-type filters. The four "ports" are simply the two orthogonal polarizations of a plane wave at the input and output ends of a birefringent filter. The filter response depends on: 1) the length of a plate, 2) its birefringence (differential phase shift per unit length for the two polarizations, which is a function of frequency), 3) the angles between the plate axes and the input polarization, and 4) the choice of output port (angle of output polarizer).

A transmission line equivalent circuit of the birefringent filter is shown in the Appendix.

C. The Birefringent Plate (*Filter Component Element*)

1) General Properties Required for Filter Work: The term birefringence means double refraction as applied to an unpolarized beam of light striking the surface of an anisotropic plate, as shown, for example, in Fig. 1. However, the condition under which no beamsplitting occurs, as shown in Fig. 2, is precisely that required for birefringent filters. In this report, the reference axis will be a direction normal to the optic axis rather than the optic axis itself. Here the optic axis, also called the fast axis (of polarization), in conformance with the characteristics of the laminated dielectric sandwich type of artificial birefringent medium used in the filters described herein, is normal to a lamination.

2) The Half-Wave Plate; Cascades of Half-Wave Plates: Fig. 2 shows a birefringent plate that we will assume is impedance-matched to free space. We further assume that linearly polarized waves normal and parallel to the plate reference axis suffer a differential phase change of 180 degrees at f_0, the design center frequency.

Note that in Fig. 2 the output wave is at an angle β to the plate axis. Thus, a rotation of the axis of a half-wave plate by an angle β from the plane of polarization of the incoming wave causes the plane of polarization of that wave to be rotated by an angle 2β. The horizontal and verti-

Manuscript received October 26, 1967; revised February 5, 1968. This work was sponsored by the NASA Electronics Research Center, Cambridge, Mass., under Contract NAS 12-126.

The authors are with the Stanford Research Institute, Menlo Park, Calif. 94025

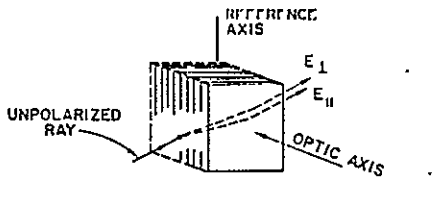


Fig. 1. Optical birefringence An unpolarized ray of light is split into two linearly polarized rays.

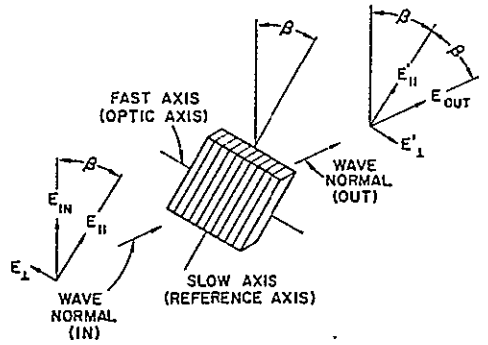


Fig. 2. Rotation of the plane of polarization of a linearly polarized wave by a half-wave plate.

cal components of the output wave are

$$E_V = E_{IN} \cos 2\beta \quad \text{and} \quad E_H = E_{IN} \sin 2\beta. \tag{1}$$

When $\beta = 45$ degrees, it is seen that the half-wave plate can become a one-element filter, because all power at frequency f_0 emerges as a horizontally polarized wave ($E_V = 0$, $E_H = 1$).

Consider now, a cascade of two half-wave plates. Let β_p be the angle of polarization at the output. It can then be shown that

$$\beta_2 - \beta_1 = 1/2\beta_p \tag{2}$$

is the general solution for the necessary plate angles of a two-plate filter.

D. Frequency Response of a One-Plate Filter

Consider now the complete frequency response of a single-plate filter with unit input. Instead of a frequency variable f, we will use the birefringence parameter γ, which is one-half the differential phase shift of the plate. The frequency response of the single-element filter is given by the well-known formulas

$$E_V = \cos \gamma \quad \text{and} \quad E_H = j \sin \gamma. \tag{3}$$

Equation (3) is based on an assumption that at $\gamma = \pi/2$, the fast axis advances the phase by 90 degrees whereas the slow axis retards the phase by the same amount—all referred to some output reference phase which these equations state is zero.[1] The plate angle is $\beta = 45$ degrees.

[1] Physically, the rate of change of phase with frequency is proportional to time delay and must be positive. There is no similar restriction on phase, and the initial value of phase is like a constant of integration that may be assigned an arbitrary value, conveniently zero.

E. Artificial Birefringent Medium

The construction of a birefringent filter for millimeter waves requires an artificial birefringent crystal or medium. Our choice was the air-dielectric sandwich, which has been analyzed by Collin.[1][2] Collin's second-order theory of birefringence in an air-dielectric sandwich material enables us to predict how the birefringence-versus-frequency function deviates from linearity and helps us to choose suitable dimensions for the filter structure.

II. Design of an Artificial Birefringent Plate

A. The Anisotropic Medium

Rough plate dimensions were calculated, based in general on the equipment to be used in testing the filter and in particular on the cross section and extent of the radiant beam that could be generated. On that basis, it was decided to make each plate aperture ten inches square, so that it would encompass the test beam, and to make the length of the filter not much greater than its width to minimize the effects of any beam divergence. An upper limit of about three inches was thus placed on the thickness of a single plate in, say, a four-element filter.

Fig. 3 shows four possible response shapes of a four-plate filter. The lowest passband, at frequency f_0, was chosen as the main passband (the lower right sketch in Fig. 3), in order to keep the overall length of the filter reasonably short. The width of the passband to the half-power points was computed from an approximate formula derived from one given by Solc[3] (also approximate) for filters with equal plate angles

$$\Delta f/f_0 \doteq 1.2/Nk. \tag{4}$$

Here, N is the number of elements or plates and k is an integer representing the order of the passband. In this case, we use $k = 1$. According to (4) for $N = 4$ and $k = 1$, we can expect passbands of the order of 30 percent of the center frequency. Each plate would thus have to be matched over a wide band by tapering or stepping the edges of the dielectric laminations. These were made of Rexolite 1422 (relative dielectric constant $\kappa_d = 2.53$), which has good machinability and dimensional stability, as well as low loss.

B. Design of A Half-Wave Plate

The plan for plate construction, excluding framework, is shown in Fig. 4. Here, t is the maximum thickness of the dielectric material (Rexolite) and S is the spacing of the sheets of this material. In the tapered regions, which occupy two-thirds of the plate length (thickness), the sheet thickness t' varies linearly from 0 to t over more than a wavelength for both polarizations, yielding wide-band impedance matching. The normalized design parameters for the composite material are the ratios t/S for the inner region, t'/S (varying from 0 to t/S) in the two tapered regions, and S/λ where λ is the wavelength in free space. The reference axis is parallel to the thin edge of a sheet of dielectric.

The most efficient use of the dielectric laminations is obtained when the ratio t/S is approximately 0.5. The bire-

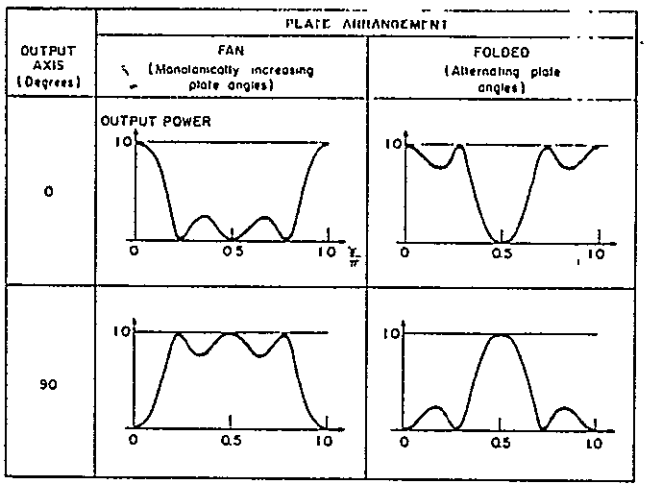


Fig. 3. Fundamental response characteristics of equal-length birefringent filters

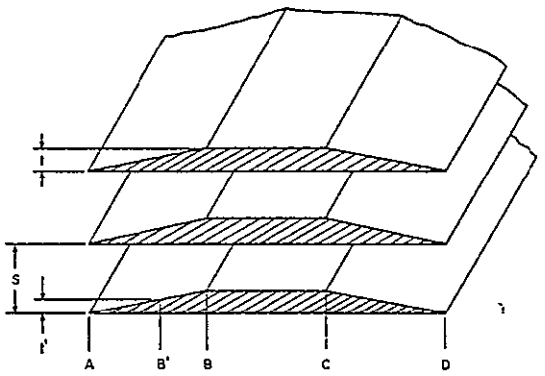


Fig. 4. Sketch of the internal construction of an impedance-matched artificial birefringent plate.

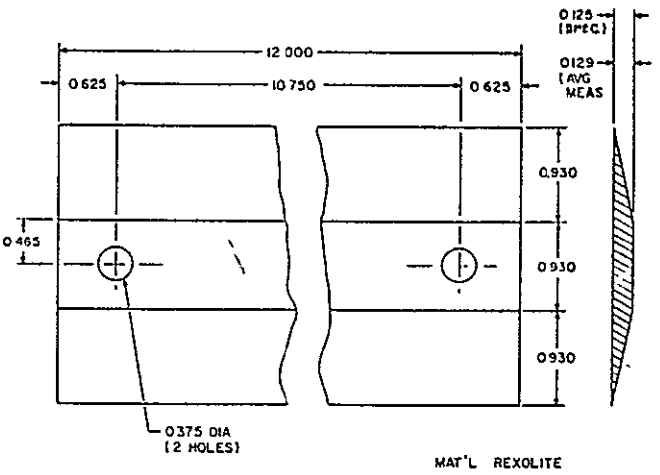


Fig. 5. Sketch of a plate lamination, showing dimensions.

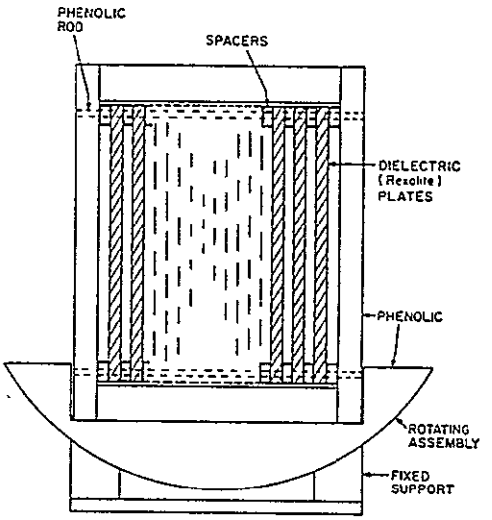


Fig. 6. Front view of a half-wave plate assembly.

fringence for the static case is then near maximum for a given material. When frequency effects are considered, as in Collin's theory,[14] the optimum ratio is a function of frequency and generally is less than 0.5. Excitation of grating lobes will be prevented if S is less than λ_d, where λ_d is the wavelength in the material of the dielectric sheets.[3] With a relative dielectric constant of $\kappa_d = 2.53$ and a free-space wavelength of 0.59 inch (at $f_0 = 20$ GHz), we find that S/λ_0 should be no greater than 0.628. A sheet thickness ($t = 0.125$ inch) was chosen for the plate, and the value t/S was chosen as 0.4, which provided near-optimum efficiency, according to calculations based on Collin's theory. Combined, these values yield $S/\lambda_0 = 0.53$, which is less than the 0.628 calculated above for the allowable maximum with respect to the generation of grating lobes.

The dimensions of each tapered lamination are given in Fig. 5. These dimensions were based on interpolated data with respect to the tapered regions, and the complete plate was designed for $f_0 = 20$ GHz. A direct recalculation of the center frequency, rather than an interpolation from previously computed data, yielded $f_0 = 20.35$ GHz as the half-wave plate center frequency. A sketch of the front view of the plate is shown in Fig. 6.

[3] This limit is for the extreme case, $t/S \to 1$.

C. Test Results on the Half-Wave Plate

Two horn-reflector antennas were lined up in parallel (without the half-wave plate) so that the maximum signal (which was made the reference level) was received. The test frequency used was 19.9 GHz. The half-wave plate was then inserted midway between the transmitting and receiving antennas, and at various positions, the level of the received signal was noted, and compared with the reference level. The results are plotted as attenuation (dB) versus orientation of plate (degrees) in Fig. 7. Two experimental curves of this characteristic, plus a theoretical curve based on (1) are given in Fig. 7.

In addition to testing the plate near the calculated value of f_0 at various angles, tests were made to determine the actual center frequency. These yielded a center frequency $f_0 = 20.65$ GHz, with an estimated measurement accuracy of ± 0.15 GHz. (Later measurements on a filter composed of five identical plates gave $f_0 = 20.4$ GHz.)

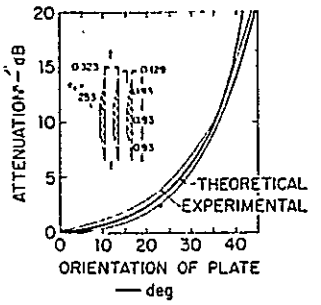


Fig. 7. Theoretical and experimental direct-wave transmission of a linearly polarized beam through a half-wave plate for varying plate angles.

In making the described measurements, it was necessary to fully absorb or divert (that is, reflect out of the system) the output wave orthogonal to the desired wave. Since the undesired component was horizontally polarized, plane gratings made of parallel thin wires spaced one-eighth inch apart were placed on each side of the test plate. The wires were horizontal, and the grating planes were at 45 degrees to the propagation path. The undesired component was thus reflected out of the system, while the desired component was transmitted freely through the two gratings. Also, thin absorbent cards were placed inside each horn so that the desired component was transmitted freely and the undesired component was absorbed. These precautions were necessary to provide proper termination of all four possible ports of the network.

In this first test on a single birefringent element, there appeared to be little or no reflections from the plate; the insertion loss was found to be very low, the measured center frequency was quite close to the design value, and the transmission-versus-plate-angle characteristic was generally close to theoretical.

III. Design and Test of a Five-Element Filter

A. Design

The formulas used here for the plate angles of the folded type of filter are

$$\beta_i = \beta, \quad (i \text{ odd}), \qquad \beta_i = -\beta, \quad (i \text{ even}), \tag{5}$$

and

$$\beta_p = 90 \text{ degrees} \tag{6}$$

where

$$\beta = 45/N \text{ degrees}. \tag{7}$$

Four additional plates identical to the first test plate were constructed so that a filter with up to five plates could be tested. The computed response of a five-plate filter is shown in Fig. 8.

The dashed line in Fig. 8 is the complementary output (absorbed component) of the five-element filter and is labeled "direct wave," since its E field lies in the same plane as the input wave. It would appear that the stopband ripples, other than those closest to the passbands, are missing,

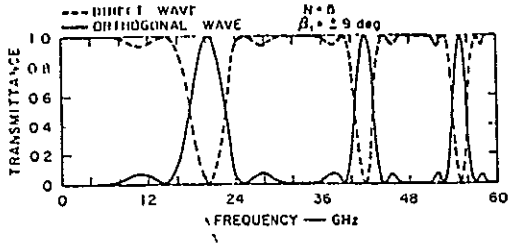


Fig. 8. Computed relative power of the direct and orthogonal outputs of a five-element equal-angle filter with the filter-element structure of Fig. 7.

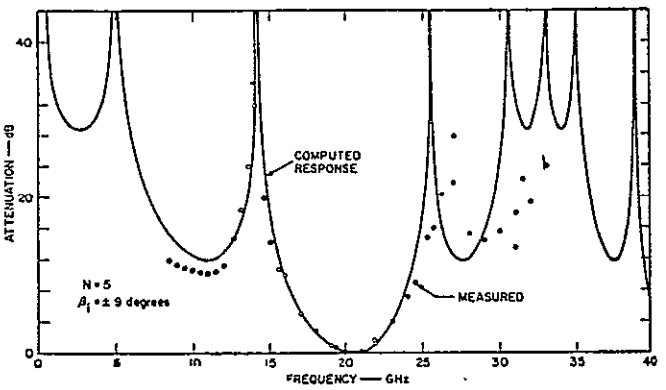


Fig. 9. Measured and computed attenuation response of the orthogonal (bandpass) output of a five-element equal-angle filter.

however, this is only because their amplitude is imperceptible on the scale used in Fig. 8. These minor ripples are shown in some of the following figures. (There are actually two ripples in the lowest stopband—four in all the others of a five-element equal-angle filter.)

B. Test

The five-element folded-type filter was tested as a bandpass filter over the range of 8–35 GHz. The test results together with the calculated response (filter attenuation versus frequency) are shown in Fig. 9. (The method used in calculating the response of birefringent filters is outlined in the next section.) Here, measurement accuracy depended on precision attenuators. Photographs of the assembled filter and the test set-up are shown in Figs. 10 and 11, respectively. The experimental points plotted in Fig. 9 were obtained by two methods. In the frequency range above 22 GHz, the measurements were made at discrete frequencies; at lower frequencies, the frequency was swept electronically. The experimental points in the region below 22 GHz also included some point-by-point measurements as a further check. The reference levels were measured when the test horns were (polarized) parallel and the filter was absent from the path of transmission. The receive horn was then rotated 90 degrees, causing the received signal without the filter in position to be attenuated by about 40 dB or more. The filter was subsequently inserted in the path of transmission and the signal level was again measured.

A plot of the filter response for the bandstop mode is shown in Fig. 12 over the range 16–33 GHz. For reference purposes, the calculated theoretical responses of both modes

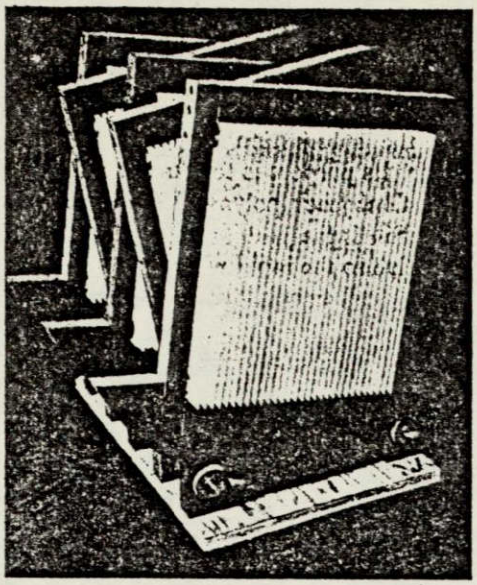


Fig. 10. Photograph of the experimental five-element filter.

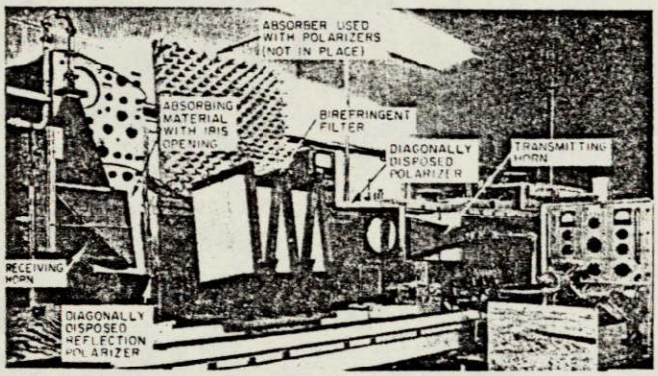


Fig. 11. Photograph of the instrumentation used for measuring the attenuation versus frequency response of the filter of Fig. 10.

are plotted on the same scale in Fig. 13; however, here the abscissa is the birefringence parameter γ, so that there is no crowding of the response shape in the upper part of the spectrum such as occurs when frequency is used as the independent variable (Figs. 8, 9 and 12). The filter response shown in Fig. 12 was made with both antennas aligned parallel, and they remained so throughout the test; no rotation of a horn was required.

The test results (Figs. 9 and 12) tend to confirm the validity of Collin's second-order theory of the birefringence of artificial anisotropic dielectrics,[14] and Solc's theory with respect to equal-length birefringent filters.[6] Solc had, of course, constructed filters at optical frequencies, while the work at Stanford Research Institute was done at frequencies many orders of magnitude lower.

The deviation of the measured points from the computed line in the upper stopband (Fig. 9) may have been caused by trapped resonant modes of the undesired (direct-wave) response, in turn due to imperfect horns and the quasi-optic nature of the experiment.

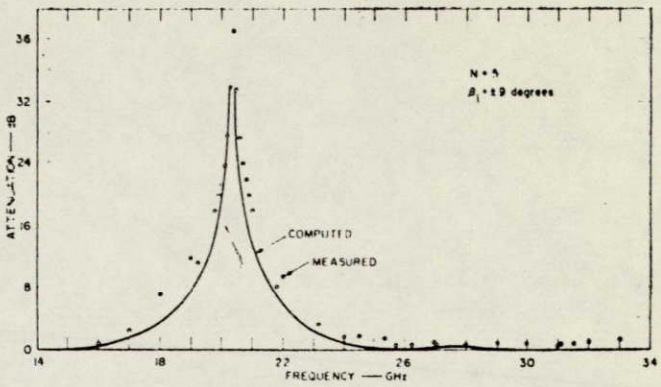


Fig. 12. Measured and computed response of the filter of Fig. 10 for the direct wave (bandstop) mode.

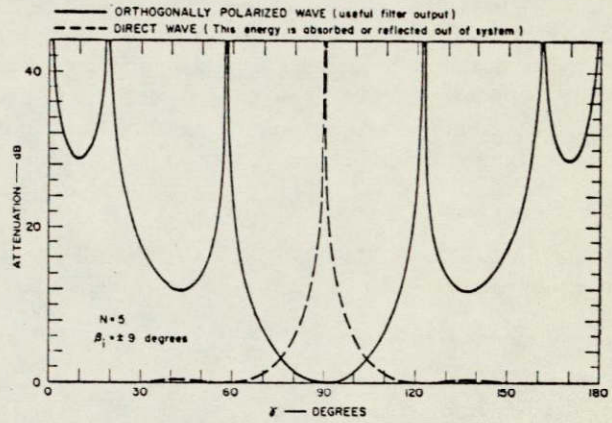


Fig. 13. Computed attenuation response of the five-element equal-angle filter with γ as the independent variable.

IV. Analysis and Synthesis

A. Analysis of Filters with Equal-Length Plates

Evans[15] has described a method of calculating the response of birefringent filters composed of equal-length elements using a matrix multiplication technique attributed to Jones.[16]

A complex output wave vector $\overline{E}_{OUT}$ is found by premultiplying the input wave vector $\overline{E}_{IN}$ by a matrix $[M]$ representing the birefringent filter

$$\overline{E}_{OUT} = [M]\overline{E}_{IN}. \tag{8}$$

The wave vectors are two-element column vectors and M is a 2×2 square matrix. Thus

$$\overline{E}_{OUT} = \begin{pmatrix} E_X \\ E_Y \end{pmatrix} \quad \text{and} \quad \overline{E}_{IN} = \begin{pmatrix} E_V \\ E_H \end{pmatrix}.$$

The X-Y and V-H axes are independent, with the X axis at an angle β_p to the V axis. Expanding (15), we obtain

$$E_X = M_{11}E_V + M_{12}E_H \tag{9}$$

and

$$E_Y = M_{21}E_V + M_{22}E_H. \tag{10}$$

For a vertically polarized input wave of unit amplitude ($E_V = 1$, $E_H = 0$) the output wave components are seen to be

$$E_X = M_{11} \qquad E_Y = M_{21}. \tag{11}$$

The matrix for a plate is given by

$$[P] = \begin{pmatrix} e^{j\gamma} & 0 \\ 0 & e^{-j\gamma} \end{pmatrix}, \tag{12}$$

where, as before, 2γ is the differential phase shift of the two orthogonal waves on the principal axes of the plate. The matrix for a rotation by an angle β is

$$[S(\beta)] = \begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix}. \tag{13}$$

The $[S]$ matrix rotates the wave vector and plate (as a system) so that the plate axis is aligned with the vertical axis, and then rotates them back again to the original angle, so that the vertical and horizontal complex components of the output vector are obtained. This is done for a single plate through a combination of premultiplication and postmultiplication by the $[S]$ matrix as follows.

$$\overline{E}_{OUT} = [S(\beta)][P][S(-\beta)]\overline{E}_{IN}. \tag{14}$$

For several plates we obtain a chain matrix, which—after a simple substitution—can be put in the following form.

$$\overline{E}_{OUT} = [S(-\rho_p)][P][S(-\rho_N)][P][S(-\rho_{N-1})][P] \cdots \cdot[S(-\rho_2)][P][S(-\rho_1)]\overline{E}_{IN}. \tag{15}$$

Here the ρ_i are the plate difference angles. Equation (15) is easily programmed, and it has been used to compute filter responses. Since the independent variable in the $[P]$ matrix is γ, the frequency response of the filter was obtained by substituting frequency values corresponding to values of γ used in the filter response calculations, that had been calculated by Collin's theory[14] for the specific plate structure of the filter.

B. Synthesis of Optimum Response Filters

1) General: Harris *et al.*[7] described a general procedure[3] for the synthesis of lossless birefringent networks.

The first step is to find a Fourier series representation of the desired response. Dolph's[17] method of obtaining a Fourier series with equal-amplitude ripples is employed here as the first step in the synthesis procedure, then the Harris method[7] is used to obtain birefringent filters with equal-amplitude stopband ripples. It has been found convenient to construct a Fourier cosine series with passband at zero, leading to a fan-type filter.

2) Finding the Fourier Series: Dolph's[17] design method is to take the coefficients of a Chebyshev polynomial term by term and make them the coefficients of a polynomial in the variable $(\cos\gamma)$[4] with the same equal-ripple behavior as the Chebyshev polynomial; however, the new function is periodic in γ and of course, unlike a Chebyshev polynomial, it is always finite. The power series is then rewritten in the form of a Fourier series, as required for the Harris synthesis procedure. A scaling procedure must also be used. This is done before the substitution of the variable, described above, is made. Its purpose is to transform the equal-ripple region of the Chebyshev polynomial from peak values of ± 1 to $\pm\epsilon$ where ϵ is the peak amplitude of the wave in the stopband, relative to the input wave, and to simultaneously fix the width of the passband at the ϵ level. The two results are interdependent.

We will not go into the details of the Dolph method since this has been covered very adequately in the literature. The cosine series is next converted to an exponential series as illustrated for the case $N = 5$.

$$M_{11}(\gamma) = \left(\frac{A_5}{2}\right)e^{j5\gamma} + \left(\frac{A_3}{2}\right)e^{j3\gamma} + \left(\frac{A_1}{2}\right)e^{j\gamma} + \left(\frac{A_1}{2}\right)e^{-j\gamma} + \left(\frac{A_3}{2}\right)e^{-j3\gamma} + \left(\frac{A_5}{2}\right)e^{-j5\gamma}. \tag{16}$$

Equation (16) is in a form suitable for use with the remainder of the Harris procedure, which is given in outline form in the following section.

3) Network Synthesis (Finding the Plate Angles): The next step is to find the orthogonal output. (We assume, at this point, that the coefficients of $M_{11}(\gamma)$ have been computed for a particular value of ϵ.) By conservation of energy

$$|M_{11}|^2 + |M_{21}|^2 = 1. \tag{17}$$

This equation states that the total output power normalized to the input power is unity at all frequencies in a lossless, nonreflective, birefringent filter. Equation (17) is then solved for $|M_{21}|^2$

$$|M_{21}|^2 = 1 - |M_{11}|^2. \tag{18}$$

M_{21} is in the form of an exponential series. How to find that series from $|M_{21}|^2$ is explained by Harris,[7] using a procedure described by Pegis that includes complex root-finding methods. (During this process, one-half of a set of $2N$ roots are chosen for use in ensuing operations, and the remaining roots are discarded. The correct roots to choose are those with absolute magnitude less than or equal to one.)

After the exponential series M_{21} has been found, a matrix multiplication method is used to sequentially find the plate difference angles ρ_i. This series of steps starts with finding the output polarizer difference angle $\rho_P = \beta_p - \beta_Y$, and then works backwards from the last plate difference angle, ρ_N, to the first, ρ_1, thereby obtaining the ρ_i of every pair of adjacent elements of a fan-type filter. Then, by changing the signs of alternate ρ_i of the fan-type filter, we obtain the design of a folded-type filter, and finally, the plate angles β_i. As explained earlier, this implies a change of the output plane of

[3] This shall be known throughout the report as the Harris procedure or method.

[4] Part of the Chebyshev function is discarded in this process.

polarization from the direct to the orthogonal wave, but this change is automatically accomplished by the sign reversal of all the alternate ρ_i, including ρ_1.

4) An Equal-Ripple Design ($N=5$): The response shape of an optimum design is shown in Fig. 14 for a folded-type filter with $N=5$. The abscissa is the variable γ. Since the response is symmetrical with respect to $\gamma=90$ degrees, only the lower stopband and half of the first passband are shown. The edge of the stopband is $\gamma_1'=43.4$ degrees, and the ripple level is $\epsilon=0.02$. The exact plate angles obtained from the synthesis by computer are

$$\beta_1' = \beta_6' = 4.90240 \text{ degrees}$$
$$\beta_2' = \beta_4' = -10.1365 \text{ degrees}$$
$$\beta_3' = 14.9221 \text{ degrees}$$
$$\beta_0' = 90.0 \text{ degrees}.$$

The input for the computer synthesis was $\gamma_1=46.6$ degrees for the fan-type filter, corresponding with $\gamma_1'=90-46.6=43.4$ degrees, the edge of the stopband of the folded-type filter response of Fig. 14.

5) Formulas for Equal-Ripple Filters with Any Value of N: In order to compute the plate angles for any value of N and various values of γ_1, the following iterative formulas were developed. These formulas give the Chebyshev coefficients, the coefficients of the two Fourier exponential series M_{11} and $|M_{21}|^2$, and the coefficients of the polynomial (A8) of the Pegis procedure, as described in Harris.[7]

a) Chebyshev Coefficients: The coefficient of y^k of the Nth-order Chebyshev polynomial of the first kind $T_N(y)$ is given the notation t_{Nk} for N both even and odd. An iterative formula for the t_{Nk} is then

$$t_{Nk} = 2t_{N-1,k-1} - t_{N-2,k}, \qquad \text{for } k = 1 \text{ to } N \tag{19}$$

with the following initial conditions:

$$t_{N0} = (-1)^{N/2}, \quad \text{for } N \text{ even},$$
$$t_{N0} = 0, \quad \text{for } N \text{ odd},$$
$$t_{11} = 1,$$
$$t_{N-2,k} = 0, \quad \text{for } k > (N-2).$$

b) Coefficients of the Fourier Exponential Series M_{11}: First compute ϵ from the Chebyshev coefficients[5]

$$\epsilon^{-1} = \sum_{k=1}^{N} t_{Nk} y_1^{-k}, \qquad \text{for } k \text{ odd}. \tag{20}$$

We then have for the coefficients of the cosine series,

$$A_{N-k} = \frac{1}{t_{N-k,N-k}} \left(\epsilon \frac{t_{N,N-k}}{y_1^{N-k}} - \sum_{j=2}^{(k+2)/2} A_{N-2(j-2)} t_{N-2(j-2),N-k} \right). \tag{21}$$

[5] An alternate procedure, replacing (20), would be to choose ϵ instead of γ', and then compute

$$y_1^{-1} = \cosh\left[\frac{1}{N}\cosh^{-1}\frac{1}{\epsilon}\right], \quad \text{and} \quad \gamma_1 = \cos^{-1} y_1.$$

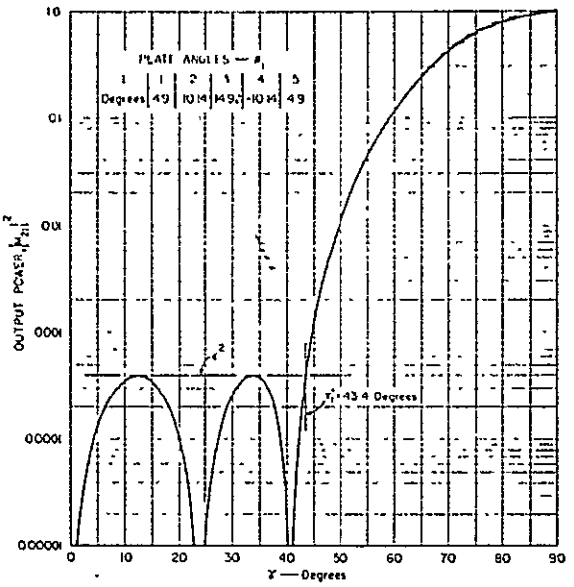


Fig. 14. Computed relative power output of a folded type of equal-ripple five-element filter.

The A_{N-k} values are to be computed for even values of k, over the range $k=0$ to $N-1$. Note that if $k=0$ in this equation, the summation term is understood to be zero. The coefficients of the exponential series M_{11} ((16) in general form) are then obtained from A_{N-k} values as follows

$$C_j = C_{n-j} = \tfrac{1}{2} A_{N-2j}, \qquad j = 0 \text{ to } \frac{(N-1)}{2}. \tag{22}$$

These are analogous to the coefficients of $C(\omega)$ in Harris' eqs. (2) and (20)[7] and are given in the same order. Here, unlike its function in the previous cosine series, the subscript j does not directly indicate the power of the exponential variable.

c) Coefficients of the Exponential Series $|M_{21}|^2$:

$$D_0 = 1 - \sum_{k=0}^{N} C_k^2 \tag{23}$$

and

$$D_k = -\sum_{j=0}^{N-k} C_j C_{j+k}, \qquad \text{for } k = 1 \text{ to } N.$$

Here, again, the D_k coefficients are analogous to those of $D(\omega)$ in Harris' eqs. (8) and (21), and they are given in the same order.[7]

d) Coefficients B_k of eq. (A8) in Harris' Method: First, construct the following set of numbers.

$$R_{jk} = R_{j-1,k-1} + R_{j,k-2}, \quad j = 0 \text{ to } N, \; k = j \text{ to } N, \tag{24}$$

excluding those R_{jk} in which $(j+k)$ is an odd number, with the initial conditions

$$R_{jj} = 1 \qquad \text{for all } j$$
$$R_{0k} = 2 \qquad \text{for } k \neq 0.$$

DESIGN TABLES FOR EQUAL-RIPPLE STOPBAND (FOLDED-TYPE) BIREFRINGENT FILTERS
(*Note:* In these tables BETA [I] = β_I' of the text.)

Table I: N = 5

ATTENUATION (NOMINAL)	10	15	20	30	40
GAMMA/IPRIME (DEGREES)	69.62	63.44	57.55	47.20	38.26
BETA[I] (DEGREES) I=1,2,3,...	15.0843	11.0060	8.3863	5.5708	4.1994
	1.4638	-3.7661	-6.8380	-9.6124	-10.5991
	17.7590	15.4277	14.5513	14.6327	15.4030
ATTENUATION (EXACT)	9.99	14.95	20.03	29.95	39.97

Table II: N = 7

ATTENUATION (NOMINAL)	10	15	20	30	40
GAMMA/IPRIME (DEGREES)	75.27	70.63	66.22	57.94	50.14
BETA[I] (DEGREES) I=1,2,3,...	13.1514	8.9231	6.3180	3.5400	2.2297
	4.1609	-0.4030	-2.8306	-4.7216	-5.2604
	15.0415	12.0154	10.5452	9.6391	9.6754
	3.0840	-2.3182	-5.6122	-9.1972	-11.0692
ATTENUATION (EXACT)	10.01	15.01	20.00	29.96	40.17

Table III: N = 9

ATTENUATION (NOMINAL)	10	15	20	30	40
GAMMA/IPRIME (DEGREES)	78.52	74.81	71.29	64.49	58.01
BETA[I] (DEGREES) I=1,2,3,...	12.2189	7.9079	5.3422	2.6692	1.4881
	5.5991	1.2544	-0.9591	-2.5860	-2.7603
	13.4951	9.9470	8.0865	6.6334	5.9674
	4.6174	-0.4392	-3.3744	-6.3675	-7.7108
	14.0249	10.9210	9.5154	8.8880	9.3104
ATTENUATION (EXACT)	9.97	15.02	20.00	29.99	40.08

Table IV: N = 11

ATTENUATION (NOMINAL)	10	15	20	30	40
GAMMA/IPRIME (DEGREES)	80.60	77.52	74.59	68.88	63.44
BETA[I] (DEGREES) I=1,2,3,...	11.6500	7.3122	4.7790	2.2103	1.1369
	6.4361	2.1904	0.0516	-1.5044	-1.6861
	12.5368	8.7176	6.6216	4.8525	3.8504
	5.6706	0.8958	-1.7551	-4.2131	-5.0760
	13.1038	9.7262	8.0984	7.0802	7.1592
	5.3679	0.3396	-2.5950	-5.6780	-7.1829
ATTENUATION (EXACT)	9.95	15.03	20.02	30.02	39.95

Table V. N = 13

ATTENUATION (NOMINAL)	10	15	20	30	40
GAMMA/IPRIME (DEGREES)	82.01	79.42	76.93	72.06	67.30
BETA[I] (DEGREES) I=1,2,3,...	11.1041	6.9274	4.4348	1.9503	0.9319
	6.9025	2.7831	0.6798	-0.8779	-1.1009
	11.8489	7.9430	5.7442	3.6283	2.7283
	6.3065	1.8026	-0.6588	-2.8045	-3.4221
	12.3454	8.8010	6.9728	5.5685	5.2864
	5.9486	1.1604	-1.6126	-4.4211	-5.7017
	12.5332	9.1461	7.4974	6.4989	6.6558
ATTENUATION (EXACT)	10.00	15.02	19.99	29.93	39.94

Table VI. N = 15

ATTENUATION (NOMINAL)	10	15	20	30	40
GAMMA/IPRIME (DEGREES)	81.05	80.85	78.66	74.35	70.18
BETA[I] (DEGREES) I=1,2,3,...	10.5630	6.6978	4.2016	1.7441	0.7997
	7.7211	3.2224	1.0496	-0.6939	-0.7516
	11.1645	7.6583	5.1765	2.9691	2.0775
	6.7509	2.4701	0.0859	-1.9076	-2.3916
	11.7806	8.1554	6.1554	4.4707	3.9620
	6.4123	1.9815	-0.7770	-3.3279	-4.1251
	12.0202	8.5841	6.8065	5.5944	5.5922
	6.2380	1.6549	-1.1251	-3.9563	-5.2607
ATTENUATION (EXACT)	10.04	14.95	19.95	30.02	39.97

Table VII: N = 17

ATTENUATION (NOMINAL)	10	15	20	30	40
GAMMA/IPRIME (DEGREES)	83.90	81.89	79.91	76.16	72.62
BETA[I] (DEGREES) I=1,2,3,...	10.7168	6.4559	3.9559	1.6278	0.7100
	7.5764	3.4739	1.3341	-0.2301	-0.5267
	11.1272	7.0490	4.7096	2.5424	1.6315
	7.2039	2.8786	0.5400	-1.2964	-1.7256
	11.4683	7.6187	5.5037	3.7034	3.0571
	6.9083	2.3657	-0.2028	-2.6444	-3.2662
	11.7047	8.0419	6.1364	4.7599	4.5339
	6.7432	2.0629	-0.6664	-3.2743	-4.4500
	11.7896	8.2000	6.3821	5.2018	5.1982
ATTENUATION (EXACT)	9.97	15.02	20.13	30.00	40.07

Table VIII: N = 19

ATTENUATION (NOMINAL)	10	15	20	30	40
GAMMA/IPRIME (DEGREES)	84.46	82.69	81.06	77.62	74.25
BETA[I] (DEGREES) I=1,2,3,...	10.3247	6.2072	3.9209	1.5446	0.6519
	7.5363	3.6015	1.6268	-0.0373	-0.3723
	10.6469	6.6856	4.3129	2.2537	1.3524
	7.2246	3.1198	1.0021	-0.8627	-1.2765
	10.9393	7.1600	5.1467	3.1632	2.4177
	6.9609	2.6760	0.3892	-1.8074	-2.4466
	11.1647	7.5504	5.7035	4.0756	3.4782
	6.7826	2.3590	-0.0737	-2.6051	-3.6255
	11.2983	7.7746	6.0163	4.6718	4.5615
	6.7193	2.2430	-0.2481	-2.9259	-4.1108
ATTENUATION (EXACT)	10.19	15.17	19.86	29.92	39.92

Table IX: N = 25

ATTENUATION (NOMINAL)	10	15	20	30	40
GAMMA/IPRIME (DEGREES)	85.79	84.46	83.10	80.50	77.93
BETA[I] (DEGREES) I=1,2,3,...	10.0028	5.9436	3.5236	1.3333	0.5309
	8.0157	4.0553	1.9087	0.2708	-0.1179
	10.1765	6.2152	3.8607	1.7175	0.8482
	7.8453	3.7784	1.5518	-0.1700	-0.5664
	10.3410	6.4925	4.2309	2.2080	1.4263
	7.6895	3.5063	1.1767	-0.6991	-1.1853
	10.4854	6.7529	4.6006	2.7599	2.1086
	7.5596	3.2645	0.8241	-1.2528	-1.9038
	10.5974	6.9691	4.9231	3.2898	2.8257
	7.4689	3.0810	0.5492	-1.7297	-2.5725
	10.6646	7.1132	5.1455	3.6833	3.3933
	7.4257	2.9815	0.3903	-2.0112	-2.9870
	10.6852	7.1640	5.2250	3.8303	3.6125
ATTENUATION (EXACT)	10.15	15.09	20.22	30.15	40.06

The coefficients of the Pegis polynomial (coefficients of y^k in eq. (A8) in Harris *et al.*[7]) are then computed for $k=0$ to n

$$B_k = \sum_{j=0}^{N-k} (-1)^{j/2} R_{k,k+j} D_{k+j}, \qquad (25)$$

with the summation over j to be made only for even values of $j \leq (N-k)$.

The remainder of the synthesis procedure is as given in Harris *et al.*[7] It requires complex root-finding procedures, the reconstruction of the polynomial M_{21} from half of the root factors of $|M_{21}|^2$, and matrix multiplications for determining plate angles. As mentioned in Section IV, B-3) of this paper, the proper roots of $|M_{21}|^2$ to use in the synthesis procedure are those with absolute magnitude less than or equal to one.

e) Discussion: Of further interest are the computer-determined distributions of the one real zero and the $(N-1)/2$ pairs of conjugate zeros of the Pegis eqs. (A8) and (A10) in Harris,[7] of the fan-type filter. Here, the variables are Y and X, where $Y = X + X^{-1}$. The zeros of eq. (A8) are found to lie on an ellipse, the major axis of which lies on the real axis of the Y-plane. The roots are equispaced on the ellipse in the sense that they are projections (toward the real axis) of N equispaced points on the circumscribing circle,

with conjugate roots in both the right and left half-planes, and the one real root at the vertex of the ellipse $Y=2$. The semimajor diameter of the ellipse is $\alpha=2+h$, where $0<h<2$. The point $Y=-2$ appears to be the opposite focus of the ellipse. Furthermore, the center of the ellipse is at $Y=-h$. The locus of the half-cycle in γ, $0<\gamma<\pi/2$ (see the lower left sketch of Fig. 3, which shows a full cycle of the orthogonal response of the fan-type filter) seemingly is transformed in the variable Y to $-2<Y<2$ on the real axis. The equation for such a displaced ellipse in the (Y_1, Y_2) plane is

$$\frac{(Y_1+h)^2}{(2+h)^2}+\frac{Y_2^2}{8h}=1 \tag{26}$$

where $Y=Y_1+jY_2$. This distribution of zeros is to be compared with the zeros of $|M_{21}|^2$ in the variable y/y_1

$$|M_{21}|^2=1-\epsilon^2 T_N^2\left(\frac{y}{y_1}\right). \tag{27}$$

Here, the zeros can be shown to lie on an ellipse centered on the origin, with $2N$ zeros (2 of which are real) that are equispaced on the ellipse in the same sense as stated above. The locus $0<\gamma<\pi/2$ is transformed to $0<y/y_1<1/y_1$ on the real axis of y/y_1. The exact location of the zeros, derived by a method illustrated by Weinberg,[18] are

$$\left(\frac{y}{y_1}\right)_k=\cos\left(\frac{k\pi}{N}\right)\cosh\gamma_1-j\sin\left(\frac{k\pi}{N}\right)\sinh\gamma_1 \tag{28}$$
$$\text{for } k=0, 1, 2, \cdots, 2N-1.$$

It should be possible to find the coefficients of M_{21} by formula, rather than by general (polynomial) foot-finding procedures, as was done here for the tabulated optimum-response filter designs. The two zero distributions are linked by eq. (A8)[7] whose zeroes were found to lie on a pear-shaped curve. The vertex of the curve is at $X=1$ and there is one real zero at $X=1$, and $(N-1)/2$ pairs of conjugate complex zeros (absolute magnitude less than one) arrayed around the origin, on the rounded portion of the curve, which is almost circular in shape. The locus of the half-cycle in γ, $0<\gamma<\pi/2$, is here the unit circle starting at $X=1(\gamma=0)$ and ending on $X=-1$. As explained in Harris,[7] these zero locations are obtained from those of eq. (A8) by the solution of a quadratic equation.

V. Design Tables for Optimum Response Filters

The design tables in this section give the plate angles β_i' of the folded-type filter for all odd values of N from $N=5$ to 19, and for $N=25$ (see Figs. 3 and 14). In each case, five separate designs are given for various values of the stopband ripple parameter ϵ. Tables I–IX give the value of ϵ in terms of attenuation L_s, defined by

$$L_s=20\log_{10}\epsilon^{-1}\text{ dB}. \tag{29}$$

The output wave is orthogonal to the input wave for these designs. Heading each column are nominal values of L_s from 10 to 40 dB. The exact attenuation values, from which the tables were computed, are within 0.1 dB of the nominal values for most of the cases listed, the greatest deviation from the nominal value being 0.22 dB, and these exact values are given directly below the last given value of β' for each case. Since the folded-type filter is symmetrical about a central plane, only the first half of the β' values, including $i=(N+1)/2$, is given in the tables. The angle γ_1' [from which L_s (exact) was computed] is also given for each design.

The approximate values of attenuation (nominal values) were found by interpolation in published tables of the squared Chebyshev function[13] listed against values of N and y_1^{-1}. A more direct approach would have been to solve the equation in footnote 5 for exact values of y_1 for given values of ϵ. Several of the designs, including $N=25$, were analyzed by the matrix multiplication technique of (15), and the responses at the ripple peaks were found to be within a few hundredths of a dB of the exact value of attenuation. The β_i' values ($i=1$ to N) were computed in sequential order from β_N' to β_1' and were found to be symmetrical to better than 0.001 degree for the $N=25$ designs, and to better than 0.0001 degree for the designs with lower values of N, with increasing symmetry as N decreases. Since the tables give β' values that are rounded off to the fourth decimal place (in degrees), all listed values except for $N=25$ retain the full accuracy of the computations; for $N=25$ there is some (but not complete) loss of accuracy in the last decimal place. In any case, the precision of the tables generally exceeds the state-of-the-art of setting devices to precise angles.

A previously unsuspected feature of equal-ripple designs that is brought out by the tables is the fact that many of the designs (those with the smaller L_s values) have β' values that do not alternate in sign. They do, however, alternate in position about some average (nonzero) value of β' and are therefore of the folded-type design. The corresponding fan-type filters have monotonically increasing β values, and both types, of course, are symmetrical, or antisymmetrical, about a central plane.

A formula for computing the plate angles of the fan-type filter from those of the folded-type design is given below.

$$\left.\begin{aligned}
\beta_1&=\beta_1'\\
\beta_2&=\beta_1+(\beta_1'-\beta_2')\\
\beta_3&=\beta_2-(\beta_2'-\beta_3')\\
\beta_4&=\beta_3+(\beta_3'-\beta_4')\\
&\vdots\\
\beta_i&=\beta_{i-1}\pm(\beta_{i-1}'-\beta_i')\\
&\vdots\\
\beta_p&=\beta_N+(\beta_N'-\beta_p')
\end{aligned}\right\} \tag{30}$$

where the prime indicates the folded-type filter. The output polarizer angle β_p will be found to equal zero, and the filter response for the fan-type filter will be identical to that of the folded-type filter but shifted by $\pi/2$ in γ.

The value of γ' for a 3-dB loss in the passband may be computed from the following formula.

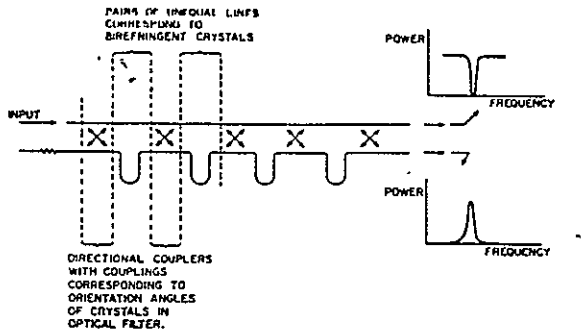


Fig. 15. Equivalent circuit for a birefringent filter.

$$\gamma_{3dB}' = \sin^{-1}\left[(\sin\gamma_1')\cosh\left\{\frac{0.115}{N}(L_s+3)\right\}\right]. \quad (31)$$

The primed values of γ denote the folded-type filter, and the values of γ_1' and L_s (attenuation) are given in the design tables.

Equation (31) is accurate for ripple levels of 10 dB or greater and may therefore be used with all the design tables in this report.

VI. Conclusion

Extrapolation of the results of research on the microwave scale model (reported herein) to millimeter wavelengths should not be difficult. The necessarily smaller size and greater fragility of the resulting structure would then demand different methods of manufacture from the standard machining techniques used for the scale model.

The optimum-response birefringent-filter design tables, their derivation, and the numerical tables, represent an important advance in the state-of-the-art. The analytical techniques and the numerical results should prove useful at optical as well as millimeter wavelengths. Topics relating to artificial anisotropic dielectrics that are worthy of investigation include: 1) improved methods of construction aimed at making them light, compact, and inexpensive, and 2) study of the effects of reflections from interfaces between plates and between a plate and free space, and means of dealing with such effects.

Appendix

Transmission Line Equivalent Circuit

An equivalent circuit for a birefringent filter is shown in Fig. 15.[11] It could be constructed in waveguide, for instance, using cascaded forward directional couplers. To make the correspondence exact, the coupling of the directional couplers should ideally be independent of frequency (corresponding to the difference angles between adjacent optical crystals, which are independent of frequency).

The difference in line length between each upper and lower connecting line is equivalent to the difference in path length between the two polarized components in each optical crystal. The amplitude coupling coefficient is equivalent to either the sine or to the cosine of the difference angle between adjacent optical crystals.

Acknowledgment

The authors wish to gratefully acknowledge stimulating discussions with C. A. Hacking on the subject of artificial anisotropic dielectrics.

The assistance of J. Herndon and W. Wiebenson, Jr., of the Mathematical Sciences Department, and P. Reznek of the Electromagnetic Techniques Laboratory, in programming the filter analysis and synthesis, and of M. Di Domenico, Sr., in performing the laboratory measurements on the birefringent plate and filter, are gratefully acknowledged.

References

[1] L. Young, B. M. Schiffman, and C. A. Hacking, "Novel millimeter and submillimeter wave filters," Stanford Research Institute, Menlo Park, Calif., February 10, 1965 (unpublished note).

[2] B. Lyot, "Optical apparatus with wide field using interference of polarized light," *Compt. Rend.*, vol 197, pp. 1593–1595, 1933.

[3] I. Solc, "A new type of birefringent filter," *Czech. J. Phys*, vol. 4, pp. 53–66, 1954, AD 140-058.

[4] I. Solc, "Further investigation of the birefringent filter," *Czech. J. Phys.*, vol. 5, pp. 80–86, 1955, AD 140-058.

[5] I. Solc, "Chain birefringent filters," *Czech. J. Phys.*, vol. 9, pp. 237–249, 1959.

[6] I. Solc, "Birefringent chain filters," *J. Opt. Soc. Am.*, vol. 55, pp. 621–625, June 1965.

[7] S. E. Harris, E. O. Ammann, and I. C. Chang, "Optical network synthesis using birefringent crystals: I. Synthesis of lossless networks of equal-length crystals," *J. Opt. Soc. Am*, vol. 54, pp. 1267–1279, October 1964.

[8] S. E. Harris and C. M. McIntyre, "Achromatic waveplates for the visible spectrum," presented at the Conf. on Electron Device Research, McGill University, Montreal, Canada, June 21–23, 1967.

[9] B. M. Schiffman, "Design of optimum-response birefringent filters," *J. Opt. Soc. Am*, vol. 57, p. 1390, November 1967.

[10] W. H. Steel, R. N. Smartt, and R. G. Giovanelli, "A 1/8 angstrom birefringent filter for solar research," *Australian J. Phys.*, vol. 14, pp. 201–211, June 1961.

[11] J. J. Taub and J. Cohen, "Quasi-optical filters for millimeter and submillimeter wavelengths," *Proc. IEEE*, vol. 54, pp. 647–656, April 1966.

[12] W. Culshaw, "Reflectors for a microwave Fabry-Perot interferometer," *IRE Trans. Microwave Theory and Techniques*, vol. MTT-7, pp. 221–228, April 1959.

[13] L. Young, "Multilayer interference filters with narrow stop bands," *Appl. Opt.*, vol. 6, pp. 297–315, February 1967.

[14] R. E. Collin, "A simple artificial anisotropic dielectric medium," *IRE Trans. Microwave Theory and Techniques*, vol. MTT-6, pp. 206–209, April 1958.

[15] J. W. Evans, "Solc birefringent filter," *J. Opt. Soc. Am.*, vol. 48, pp. 142–145, March 1958.

[16] R. C. Jones, "New calculus for the treatment of optical systems. I. Description and discussion of the calculus," *J. Opt. Soc. Am.*, vol. 31, p. 488, 1941.

[17] C. L. Dolph, "A current distribution for broadside arrays which optimizes the relationship between beamwidth and side-lobe level," *Proc. IRE*, vol. 34, pp. 335–348, June 1946.

[18] L. Weinberg, *Network Analysis and Synthesis*. New York: McGraw-Hill, 1962, pp. 507–532.

Reprinted from JOURNAL OF THE OPTICAL SOCIETY OF AMERICA, Vol. 58, No. 12, 1575–1580, December 1968
Printed in U. S. A.

Achromatic Wave Plates for the Visible Spectrum*

C. M. MCINTYRE AND S. E. HARRIS
Stanford University, Stanford, California 94305
(Received 19 June 1968)

We report the experimental demonstration of an optical quarter-wave plate having a retardation of 90°±1° in the region from 4000 to 8000 Å. The device is based on the synthesis procedure of Harris, Ammann, and Chang and consists of six sapphire wave plates with appropriately oriented principal axes. The device does not suffer from thermal or angular problems as do longer, narrow-band birefringent networks. Results comparing a 10-plate unit with the above 6-plate unit are given.

INDEX HEADINGS: Wave plate; Birefringence; Sapphire; Crystals; Filters

OPTICAL networks having prescribed transmittance-vs-frequency characteristics can be synthesized by using recently reported procedures.[1–6] The procedure for the synthesis of networks consisting of cascaded, equal-length, birefringent crystals between an input and output polarizer,[1] is also applicable to the design of certain networks which do not contain polarizers. In particular, wave plates which are nearly achromatic over a large portion of the spectrum may be synthesized.

In this paper we present theoretical considerations and experimental verification of the design of such achromatic wave plates. A quarter-wave plate consisting of six sapphire wave plates which has a retardation of 90°±1° in the region from 4000 to 8000 Å is described. We include a systematic method for the synthesis of wave plates of any desired retardation as a function of optical frequency. The degree of approximation of the synthesized wave plate to the desired transfer function is determined by the number of plates employed in its construction.

Achromatic combinations of birefringent plates have been discussed previously by several authors. Combinations of plates having different dispersions of birefringence are considered by West and Makas,[7] who also cite references to earlier related work. Destriau and Prouteau,[8] and Pancharatnam[9,10] describe achromatic combinations of two and three birefringent plates, respectively, where the plates are of the same material but different thickness. These procedures are however far less accurate than those reported here. As an example, Destriau and Prouteau give results for an achromatic quarter-wave plate with a retardation which varies from 83° at λ=6100 Å to 95° at λ=5890 Å to 84° at λ=4360 Å. Achromatic retardation may also be

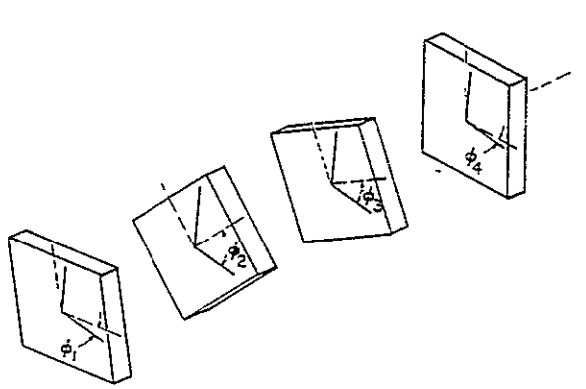


FIG. 1. Basic configuration of achromatic wave plate.

* Work was supported wholly by the U. S. Army Research Office, the U. S. Air Force Office of Scientific Research, and the Office of Naval Research under the Joint Services Electronics Program under Contract Nonr-225(83).

[1] S. E. Harris, E. O. Ammann, and I. C. Chang, J. Opt. Soc. Am. 54, 1267 (1964).
[2] E. O. Ammann and I. C. Chang, J. Opt. Soc. Am. 55, 835 (1965).
[3] E. O. Ammann, J. Opt. Soc. Am. 56, 943 (1966).
[4] E. O. Ammann, J. Opt. Soc. Am. 56, 952 (1966).
[5] E. O. Ammann, J. Opt. Soc. Am. 56, 1081 (1966).
[6] E. O. Ammann and J. M. Yarborough, J. Opt. Soc. Am. 57, 349 (1967).
[7] C. D. West and A. S. Makas, J. Opt. Soc. Am. 39, 791 (1949).
[8] M. G. Destriau and J. Prouteau, J. Phys. Radium 10, 53 (1949).
[9] S. Pancharatnam, Proc. Indian Acad. Sci. A41, 130 (1955).
[10] S. Pancharatnam, Proc. Indian Acad. Sci. A41, 137 (1955).

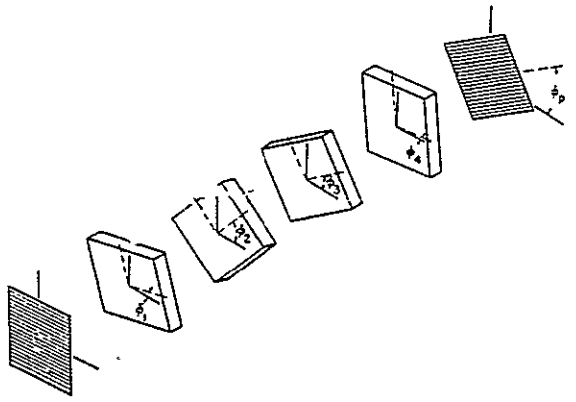


FIG. 2. Basic configuration of optical network including polarizers.

obtained by means of a Fresnel rhomb[11] which makes use of the phase difference between orthogonal polarizations upon total internal reflection. A disadvantage of the Fresnel rhomb is that the optical beam is displaced laterally during passage through the device.

The physical configuration of our achromatic wave plate is shown in Fig. 1. Four crystals are shown here, but any number may be used. In the figure, a set of reference axes, the x, y axes, are shown as solid lines for each crystal. The broken lines indicate the principal axes of the crystals. The angles to which the principal axes of the crystals are rotated are the variables of the synthesis procedure.

ANALYTICAL CONSIDERATIONS

A. Review of Synthesis Procedure

The examination of the impulse response of optical networks composed of birefringent crystals is the basis for the development of a synthesis procedure for these networks.[1] The crystals are assumed to be dispersion free and to have their end faces flat, parallel, and normal to the incident light. Under these conditions, the network shown in Fig. 2, consisting of N equal-length crystals between an input and output polarizer, has an impulse response given by

$$\begin{aligned} h'(t)&=\sum_{m=0}^{N} C_m\delta(t-m\tau),\\ \tau&=L(\Delta n)/c,\\ C_m&=f(\varphi_1,\varphi_2,\cdots\varphi_n,\varphi_p), \end{aligned} \tag{1}$$

where L is the crystal length, Δn is the difference of index of refraction between the orthogonal principal axes of each crystal, φ_i is the angle to which the slow axis of the ith crystal is rotated, φ_p is the angle to which the transmission axis of the final polarizer is rotated, and c is the velocity of light.

In Eq. (1), the time origin has been chosen such that the first impulse appears at the output at $t=0$. This choice is arbitrary; a more convenient choice is

$$h(t)=\sum_{m=-N/2}^{N/2} C_m\delta(t-m\tau). \tag{2}$$

where the impulse response is centered around $t=0$.

The frequency-domain transfer function for this network is the Fourier transform of Eq. (2)

$$H(f)=\sum_{m=-N/2}^{N/2} C_m e^{-im\omega\tau}. \tag{3}$$

Here $H(f)$ is a periodic function of frequency with a period τ determined by the crystal length and birefringence. The desired arbitrary transfer function must first be made periodic by extending it outside of the frequency range of interest. The transfer function is then expanded in the form of Eq. (3) and the synthesis procedure[1] is used to determine the crystal angles to yield the coefficients C_m.[12] If n is the number of crystals used, then the exponential series approximation to the desired function may have $n+1$ terms.

B. Achromatic Wave Plates

In applying the synthesis procedure of Harris, Ammann, and Chang,[1] the transfer function $H(f)$ is obtained between an input and an output polarizer. In general, the transmission axes of these polarizers are not parallel, and the angle between them is φ_p. This angle is determined by the synthesis procedure and is dependent on the value of $H(f)$ at $f=0$ (see Appendix B). Though achromatic wave plates do not have either an input or an output polarizer, they may be synthesized by using the procedure of Ref. 1, by correctly choosing the desired transfer function, and then simply leaving the polarizers out of the finished network. To understand this, it is convenient to look at the two-dimensional descriptions of an optical network without polarizers, using the conventional Jones calculus.[13]

In order to utilize earlier results,[1] we define an x–y coordinate system where the x axis is parallel to the transmission axis of the input polarizer and a u–v coordinate system where the u axis is parallel to the transmission axis of the output polarizer. We define $H(f)$ as the transfer function relating incident light which is linearly polarized along the x axis to output light which is linearly polarized along the u axis. We also define $G(f)$ as the transfer function relating incident light which is linearly polarized along the x axis to output light which is linearly polarized along the v axis. By conservation of energy, these transfer functions must satisfy the relation

$$H(f)H^*(f)+G(f)G^*(f)=1. \tag{4}$$

Appendix A shows that the Jones matrix for the network of crystals without polarizers can be written in

[11] See, for example, M. Born and E. Wolf, *Principles of Optics* (Pergamon Press, Inc., New York, 1964), p. 51.

[12] Experimental results for networks of this type are reported by J. M. Yarborough and E. O. Ammann, J. Opt. Soc. Am. 58, 776 (1968).

[13] R. C. Jones, J. Opt. Soc. Am. 31, 488 (1941).

the factored form

$$M=\begin{bmatrix}\cos\varphi_p & -\sin\varphi_p\\ \sin\varphi_p & \cos\varphi_p\end{bmatrix}\begin{bmatrix}H(f) & G^*(f)\\ -G(f) & H^*(f)\end{bmatrix} \quad (5)$$

when expressed in the x–y coordinate system. The first matrix represents a rotation through the angle φ_p and is independent of frequency. The second factor depends only on the desired transfer function $H(f)$.

To synthesize an achromatic wave plate of retardation θ, we take

$$H_d(f)=e^{j\theta/2}, \quad f_1\leq f\leq f_2, \quad (6)$$

where θ is independent of frequency. From Eq. (4) we have $G(f)=0$, and thus

$$M=\begin{bmatrix}\cos\varphi_p & -\sin\varphi_p\\ \sin\varphi_p & \cos\varphi_p\end{bmatrix}\begin{bmatrix}e^{j\theta/2} & 0\\ 0 & e^{-j\theta/2}\end{bmatrix} f_1\leq f\leq f_2. \quad (7)$$

The second factor is identical to the Jones matrix for a conventional wave plate except that here θ is independent of frequency over the interval f_1 to f_2. The first factor in Eq. (7) represents an achromatic rotation of the principal axes of the wave plate.

Note that, unlike the conventional wave plate, the achromatic wave plate includes a rotation through the angle φ_p. For many applications of wave plates, for instance the conversion of linearly to circularly polarized light or vice versa, this rotation is of no consequence. However, Appendix B shows that by placing a restriction on $H(f)$, namely $H(0)=1$, then φ_p will equal 0 and there will be no rotation. In general, this restriction will require more crystals in the network in order to achieve a desired degree of approximation. In the special case of an achromatic half-wave plate, the network may have principal axes at no cost, in the desired degree of approximation.

C. Approximation Problem

The coefficients, C_m, in the network transfer function of Eq. (3) are chosen to give the best approximation to the desired transfer function of Eq. (6). These coefficients must be real since they are the magnitudes of the delta functions in the impulse response of the network. Because of this, the real part of $H(f)$ must be even and can therefore be expressed as a cosine series

$$\mathrm{Re}[H(f)]=\sum_{m=0}^{N/2} B_m\cos m\omega\tau. \quad (8)$$

Comparing this with $\mathrm{Re}[H_d(f)]$, we see that the choice $B_0=\cos\theta/2$, $B_m=0$ for $m\geq 1$ is exact; therefore,

$$C_0=B_0=\cos\theta/2. \quad (9)$$

The $\mathrm{Im}[H(f)]$ must be an odd function and can be represented as a sine series

$$\mathrm{Im}[H(f)]=\sum_{m=1}^{N/2} K_m\sin m\omega\tau. \quad (10)$$

We now choose the K_m's to give the best rms approximation to $\mathrm{Im}[H_d(f)]$. To do this, we could incorporate $\mathrm{Im}[H_d(f)]$ into a periodic function and compute the usual Fourier-series coefficients. However, specifying a particular function outside the interval $f_1\leq f\leq f_2$ which would give the best approximation within the region of interest is difficult and unnecessary. The optimum approximation is obtained by minimizing the rms error only over the region of interest. With $\mathrm{Im}[H_d(f)]=\sin\theta/2$, we have

$$\frac{\partial}{\partial K_l}\left\{\int_{f_1}^{f_2}\left[\sin\frac{\theta}{2}-\sum_{m=1}^{N/2} K_m\sin m\omega\tau\right]^2 df\right\}_{1\leq l\leq N/2}=0. \quad (11)$$

This leads to N equations in N unknowns for the K_m's

$$-A_l+\sum_{m=1}^{N/2} K_m B_{ml}=0. \quad (12)$$

where

$$A_l=\int_{f_1}^{f_2}\sin\frac{\theta}{2}\sin l\omega\tau\, df$$

and

$$B_{ml}=\int_{f_1}^{f_2}\sin m\omega\tau\sin l\omega\tau\, df.$$

The solutions to Eq. (12) give the coefficients K_m which minimize the rms error over the range $f_1\leq f\leq f_2$. For a wave plate of retardation θ, the coefficients in the exponential series (3) are

$$\begin{aligned} C_0&=\cos\theta/2\\ C_m&=-\tfrac{1}{2}K_m \quad m>0\\ C_m&=\tfrac{1}{2}K_m \quad m<0. \end{aligned} \quad (13)$$

The system of Eq. (12) could include another equation

$$\frac{\partial}{\partial\tau}\left\{\int_{f_1}^{f_2}\left[\sin\frac{\theta}{2}-\sum_{m=1}^{N/2} K_m\sin m\omega\tau\right]^2 df\right\}=0 \quad (14)$$

to optimize the period of $H(f)$. This is a transcendental equation and does not lend itself to normal solution techniques. Numerical results show that the optimum period lies in a very broad minimum. Intuitively, we expect that the optimum period would correspond to a crystal length such that each crystal would have a retardation θ at the center of the band. Because of the rms criterion, this is not precisely the case but is very nearly so, and this choice of crystal length is very close to optimum.

EXPERIMENT

A. Example

An achromatic quarter-wave plate covering the region from 4000 to 8000 Å was chosen as an important example of the procedure. To synthesize this particular network, we chose τ so that the individual crystals have a retardation of $\pi/2$ (not $\pi/2\pm 2n\pi$) at the center of the band. This interval of interest extends

TABLE I. Coefficients in the expansion of $H(f)$ for a 6- and 10-element achromatic quarter-wave plate.

	Quarter-wave plate	
	6 Crystal	10 Crystal
C_0	−0.052707	−0.011329
C_1	0.000000	0.000000
C_2	−0.405376	−0.077093
C_3	0.707107	0.000000
C_4	0.405376	−0.419377
C_5	0.000000	0.707107
C_6	0.052707	0.419377
C_7		0.000000
C_8		0.077093
C_9		0.000000
C_{10}		0.011329

from $f_1=3.75\times10^{14}$ hertz to $f_2=7.5\times10^{14}$ hertz; therefore $2\pi f_0\tau=\pi/2$, where $f_0=5.625\times10^{14}$ hertz corresponds approximately to the center of the band; thus $\tau=0.444\times10^{-15}$ sec. Equations (12) and (13) were used to calculate the coefficients of the exponential series that approximates the desired transfer function over the region f_1 to f_2. These coefficients are given in Table I for two networks, one containing six elements and the other containing ten elements.

The procedure of Ref. 1, was then used to calculate the crystal angles. These angles are tabulated in Table II for both the six- and ten-element networks.

Finally, the resulting networks were analyzed using a computer program which includes dispersion. The results of this analysis show that for the six-crystal network, $\theta=90°\pm1°$ from 4000 to 8000 Å; and for the ten-crystal network $\theta=90°\pm15'$ from 4000 to 8000 Å. (These results are indicated by the solid curve in Fig. 5.)

B. Preparation of Crystals

The material used for both the six- and ten-element wave plates was sapphire. This was chosen from a consideration of the periodicity requirements. From $\tau=L(\Delta n)/c$, we obtain $L(\Delta n)=1.333\times10^{-5}$ cm. Crystals 1 mm thick were chosen for experimental convenience, yielding $\Delta n=1.33\times10^{-4}$. One method of obtaining this small Δn is to cut the crystals so that light propagates in a direction which makes an angle β with the optic axis. For sapphire, a β of approximately 7° gives $(\Delta n)=1.333\times10^{-4}$.

Crystals 1 mm thick and 8 mm square were ground and polished. The retardation was measured with a Sénarmont compensator, using a He–Ne laser (6328 Å) and a conventional quarter-wave plate for that line. The grinding and measuring process was continued until all crystals were within 1/100th wave of the desired retardation.

The individual crystals were then held as shown in Fig. 3. This gimbal-type mounting provides a rotation adjustment around three orthogonal axes for aligning the crystals. These were mounted in an index-matching oil bath. The overall length of the device is 457 mm. The size was for experimental convenience only and has been reduced to 51 mm in recent work.

C. Measurements

The measurements on the achromatic quarter-wave plate employed a xenon arc lamp in conjunction with a Leiss prism monochromator having 50-Å resolution to scan the visible spectrum. If linearly polarized light is incident at 45° to the principal axes of a quarter-wave plate, then this light is circularly polarized after passage through the plate. For the achromatic quarter-wave plate, this property should hold approximately true from 4000 to 8000 Å. To confirm this, the wave plate was placed between two polarizers, with the input polarizer at 45° to the principal axes of the plate. At each wavelength, the output polarizer was rotated through 360° and the ratio of maximum to minimum transmittance was recorded. This ratio may vary from one for circularly polarized light to infinity for linearly polarized light. A plot of this ratio vs wavelength for our 10-element achromatic quarter-wave plate is shown in Fig. 4. The computed ratio is indicated by a solid line. The broken line indicates the equivalent ratio for a single crystal which is a quarter-wave plate at 6000 Å. In this figure, the measured transmittance ratio is 1.13 or less, throughout the range of interest.

TABLE II. Crystal angles for a 6- and 10-element achromatic quarter-wave plate.

	Quarter-wave plate	
	6 Crystal	10 Crystal
φ_1	74°0′	42°6′
φ_2	73°36′	93°10′
φ_3	15°36′	98°56′
φ_4	−29°18′	25°43′
φ_5	28°42′	19°57′
φ_6	29°6′	−25°02′
φ_7		−19°16′
φ_8		53°57′
φ_9		48°11′
φ_{10}		−2°53′

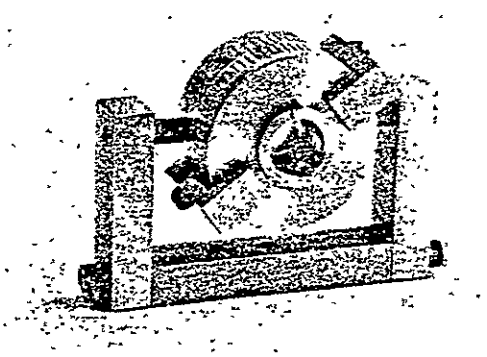


FIG. 3. Individual crystal holder.

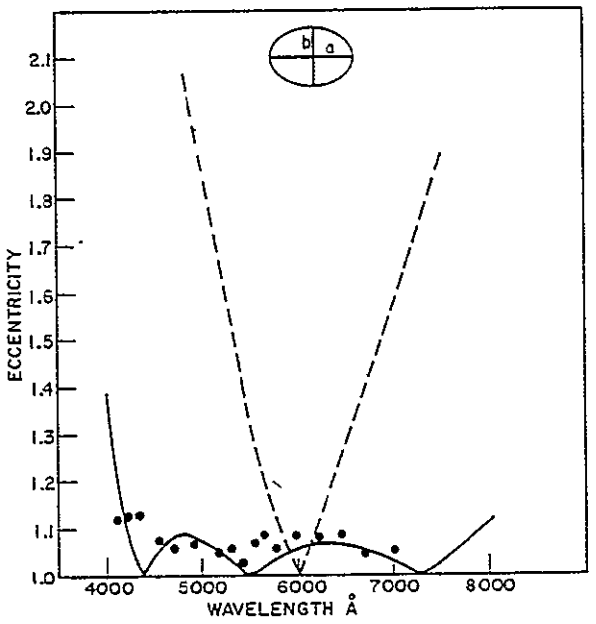
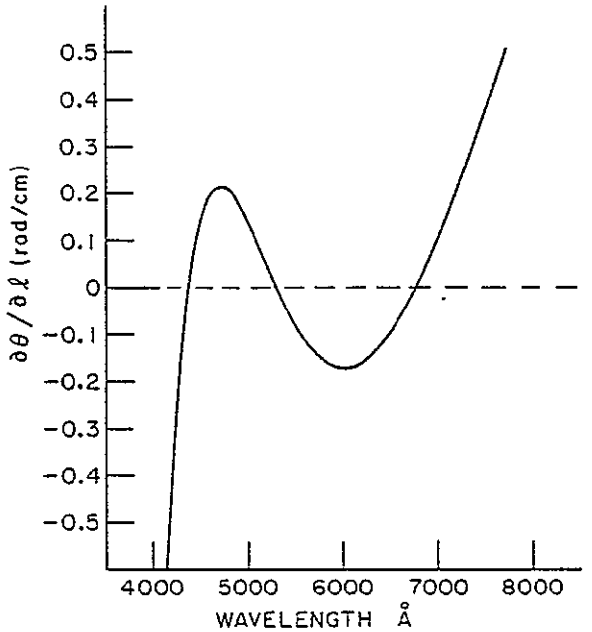


FIG. 4. Eccentricity (a/b) vs wavelength for 10-element quarter-wave plate. – – – single-crystal λ/4 plate at 6000 Å. —— theory. ·· experiment, 10-crystal λ/4 plate.



[FIG. 6. Rate of change of retardation with crystal length vs wavelength for 6-element quarter-wave plate.

From the transmittance data, the retardation of the networks can be inferred. The results of this calculation are shown in Fig. 5. Again, the broken line corresponds to a single crystal which is a quarter-wave plate at 6000 Å. The retardation of this single-crystal wave plate varies from 60° at 8000 Å to 130° at 4000 Å, or 70° over the portion of the spectrum where the retardation of the achromatic quarter-wave plate is approximately constant. The solid line indicates the computed retardation, which is 90°±15′ from 4000 to 8000 Å. The experimental points shown agree with the theory within the expected experimental error. The computation leading to the solid curves takes dispersion into account,

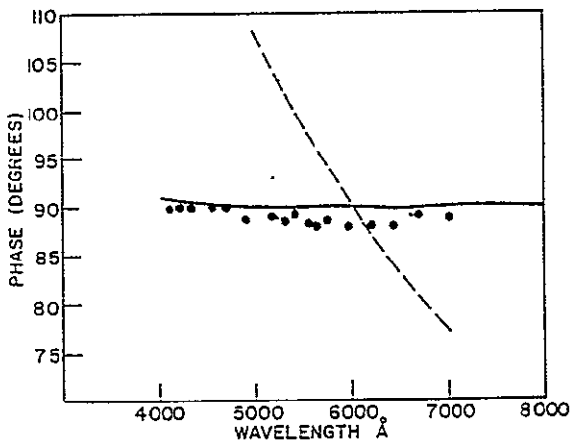


FIG. 5. Retardation vs wavelength for 10-element quarter-wave plate. – – – single-crystal λ/4 plate at 6000 Å. —— theory. ··· experiment, 10-crystal λ/4 plate.

and we believe that the small discrepancies between theory and experiment result from experimental errors.

DISCUSSION

Because of the flat transfer function of the achromatic wave plate, changes of the variables that affect the period of the transfer function [explicitly, l and (Δn): and implicitly, temperature] will not significantly affect the retardation except near the ends of the interval $f_1 \leq f \leq f_2$. Small changes of these parameters simply result in small shifts of the interval over which the retardation is constant: regions away from the edges of the interval are unaffected.

For example, consider the variation of retardation with respect to crystal length. For a single crystal, we have

$$\theta = 2\pi f l(\Delta n)/c$$

$$d\theta/dl = 2\pi f(\Delta n)/c = 15.7\,(\text{rad/cm})$$

at

$$f_0 = 5.625\times 10^{14}\ \text{Hz}.$$

For the achromatic wave plate, we have

$$\frac{\partial\theta}{\partial l} = \frac{\partial}{\partial l}\tan^{-1}\frac{\text{Im}[H(f)]}{\text{Re}[H(f)]}.$$

This expression has been evaluated numerically for the 6-crystal network (see Fig. 6) for all wavelengths throughout the region from 4000 to 8000 Å. With the exception of 200 Å at each edge of the band, we find

$$\frac{d\theta}{dl} \leq 0.5\frac{\text{rad}}{\text{cm}} \qquad 4200\ \text{Å} \leq \lambda \leq 7800\ \text{Å}.$$

The sensitivity of the retardation to crystal length is therefore better by a factor of 30:1 as compared to the sensitivity of a single-crystal wave plate.

The methods presented here are applicable to the synthesis of achromatic wave plates of any desired retardation over any portion of the spectrum. The number of crystals necessary is determined by the required degree of approximation and the interval over which the network is desired to be achromatic.

ACKNOWLEDGMENTS

The authors gratefully acknowledge many helpful discussions with Dr. Ammann. We thank Cora Barry for carrying out the numerical calculations and Bob Griffin for his excellent polishing of the sapphire crystals.

APPENDIX A

If the x, y coordinate system is chosen with the x axis parallel to the transmission axis of the input polarizer, then the Jones matrix for the network of Fig. 2 is

$$M_{x,y}=S(\varphi_p)PS^{-1}(\varphi_p)S(\varphi_n)NS^{-1}(\varphi_n) S(\varphi_{n-1})NS^{-1}(\varphi_{n-1})\cdots S(\varphi_1)NS^{-1}(\varphi_1)P =S(\varphi_p)PAP, \quad \text{(A1)}$$

where N is the Jones matrix for a birefringent plate with its fast axis parallel to the x axis, P is the Jones matrix for a polarizer with its transmission axis parallel to the x axis, $S(\varphi)$ is the unitary matrix for rotation through an angle φ, and the matrix A with elements $a_{i,j}$ is defined by the above relation. Then the frequency-domain relationship between the input and output light vectors is

$$\mathbf{D}_0=M_{x,y}\mathbf{D}_i. \quad \text{(A2)}$$

In order to express the matrix representing an identical network, but without polarizers, in terms of $H(f)$ and $G(f)$ as previously defined, it is necessary to determine the elements $a_{i,j}$ in terms of these transfer functions. By definition of the transfer function $H(f)$ we have

$$\begin{bmatrix}D_{0x}\\D_{0y}\end{bmatrix}=H(f)D_{ix}\begin{bmatrix}\cos\varphi_p\\\sin\varphi_p\end{bmatrix},$$

where $\cos\varphi_p=H(0)$. From Eqs. (A1) and (A2), we obtain

$$\begin{bmatrix}D_{0x}\\D_{0y}\end{bmatrix}=a_{11}D_{ix}\begin{bmatrix}\cos\varphi_p\\\sin\varphi_p\end{bmatrix}$$

and therefore $a_{11}=H(f)$.

By definition of the transfer function $G(f)$, if the output polarizer in the network of Fig. 2 is rotated 90° clockwise, we have

$$\begin{bmatrix}D_{0x}\\D_{0y}\end{bmatrix}=G(f)D_{ix}\begin{bmatrix}\sin\varphi_p\\-\cos\varphi_p\end{bmatrix}.$$

If this 90° rotation of the output polarizer is included, then from Eqs. (A1) and (A2) we obtain

$$\begin{bmatrix}D_{0x}\\D_{0y}\end{bmatrix}=a_{21}D_{ix}\begin{bmatrix}-\sin\varphi_p\\\cos\varphi_p\end{bmatrix}$$

and $a_{21}=-G(f)$. With the polarizers in their original positions and if all of the crystals are rotated 90°, the Jones matrix of the network becomes

$$M_{x,y}'=S(\varphi_p)PS^{-1}(\varphi_p)S(\varphi_n)N^*S^{-1}(\varphi_n)S(\varphi_{n-1}) N^*S^{-1}(\varphi_{n-1})\cdots S(\varphi_1)N^*S^{-1}(\varphi_1)P,$$

where N^* is the complex conjugate of N. Then we have

$$\begin{bmatrix}D_{0x}\\D_{0y}\end{bmatrix}=H^*(f)D_{ix}\begin{bmatrix}\cos\varphi_p\\\sin\varphi_p\end{bmatrix}.$$

In terms of the elements of the matrix A, we find

$$\begin{bmatrix}D_{0x}\\D_{0y}\end{bmatrix}=a_{22}D_{ix}\begin{bmatrix}\cos\varphi_p\\\sin\varphi_p\end{bmatrix}$$

and $a_{22}=H^*(f)$. In a similar manner, it can be shown that $a_{12}=G^*(f)$. Then the Jones matrix for the identical network, but without polarizers, is

$$M=S(\varphi_p)A=\begin{bmatrix}\cos\varphi_p & -\sin\varphi_p\\\sin\varphi_p & \cos\varphi_p\end{bmatrix}\begin{bmatrix}H(f) & G^*(f)\\-G(f) & H^*(f)\end{bmatrix}.$$

APPENDIX B

The requirement that an achromatic wave plate have principal axes (i.e., axes such that light which is polarized along them at the crystal input, remains linearly polarized in the same direction at the crystal output) is equivalent to the requirement that the Jones matrix for the network have linear eigenvectors. For an achromatic wave plate, $G(f)\cong 0$ over the frequency interval of interest and the Jones matrix may be written in the form

$$M=\begin{bmatrix}H(f)\cos\varphi_p & -H^*(f)\sin\varphi_p\\H(f)\sin\varphi_p & H^*(f)\cos\varphi_p\end{bmatrix},$$

where $H(f)$ and φ_p are defined in the text. The requirement that this matrix have linear eigenvectors leads to the condition

$$\mathrm{Re}[H(f)]\sin\varphi_p=0.$$

For the special case of an achromatic half-wave plate, $H(f)\cong e^{j\pi f/2}$ is purely imaginary and thus this case will automatically have principal axes. Achromatic wave plates with any other retardation will have principal axes only if $\varphi_p=0$.

Referring back to the network of Fig. 2 with a transfer function $H(f)$, we notice in the long-wavelength limit, when $f\to 0$, that the phase difference between the fast and slow axes goes to 0. In this limit, the network transfer function is determined by only the input polarizer, i.e., $H(0)=\cos\varphi_p$.[14] Therefore, if we require that $H(0)=1$, the input and output polarizers must be parallel and the wave plate will have two principal axes.

[14] E. O. Ammann, private communication.

Reprinted from APPLIED OPTICS, Vol. 14, page 229, January 1975

Improvement of Birefringent Filters. 2:Achromatic Waveplates

Alan M. Title

By use of Jones' matrix techniques, nine-element achromatic waveplates are developed. These combination plates are achromatic to within 1° throughout the visible (3,500–10,000 Å). In addition, a two-section general retarder rotator is demonstrated. The retardation of the combination is twice the complement of the angle between the halves, while the rotation is equal to the angle between the halves. For a 90° retardation, the dual five-element combination is also achromatic to within 1° throughout the visible.

Introduction

The ability to readily produce good scatter-free waveplates from polyvinyl alcohol makes it useful to investigate properties of multielement achromatic combination waveplates.[1] A number of authors have previously investigated properties of achromatic combinations of simple waveplates. In particular, Pancharatnam[2] developed a procedure for three-element combination waveplates. Further, Harris and McIntyre[3] have developed a scheme for achromatic combinations of n waveplates. The Harris scheme, however, does not always produce a pure waveplate, but rather a retarder plus a rotator.

In this paper, combinations of three and nine elements that can yield achromatic waveplates and four- and ten-element combinations that can yield achromatic rotator waveplate combinations will be discussed. Jones' matrices will be used to develop the conditions for achromaticity.

Pancharatnam assumes that for three-element combinations, the first and third waveplates must be equal and parallel, and the center plate must be a half wave at the center frequency. With matrix techniques it will be demonstrated that these assumptions are consequences of achromatic retardation and stability of the axis of retardation.

More importantly, the matrix form shows that the three-element half waveplate is extendable to superachromatic nine-element waveplates. Also, the matrix form makes clear that it is possible to construct four- and ten-element achromatic rotator waveplates. The four- and ten-element combinations are especially interesting because of their use as variable retardation plates. It will be shown that if the combinations are made in two halves, that is, a pair of two-element plates or a pair of five-element plates, all retardations can then be achieved by rotating the halves with respect to each other. The retardation is just twice the complement of the angle between the plates.

Mathematical Preliminaries

In order to simplify mathematical operations, it is useful to state some mathematical relations. Following the notation of Jones,[4] a retardation plate of retardation 2γ at angle θ with respect to the x axis is represented by the matrix product

$$M(\gamma, \theta) = R(-\theta)G(\gamma)R(\theta). \tag{1}$$

The product of Eq. (1) can be rewritten in the form

$$M(\gamma, \theta) = \cos\gamma I + i \sin\gamma ER(2\theta), \tag{2}$$

where

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{3}$$

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{4}$$

$$E = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{5}$$

$$G(\gamma) = \begin{pmatrix} \exp(+i\gamma) & 0 \\ 0 & \exp(-i\gamma) \end{pmatrix}. \tag{6}$$

The matrix E behaves like a mirror, thus:

$$ER(\theta) = R(-\theta)E. \tag{7}$$

and

$$E^2 = I. \tag{8}$$

It has been shown by Jones that any combination of retardation plates is equal to a retardation and a rotation. That is

The author is with Lockheed Solar Observatory, Lockheed Palo Alto Research Laboratory, Department 52-14, Palo Alto, California 94304.

Received 25 February 1974.

$$\prod_{i=1}^{n} M(\gamma_i,\theta_i) = R(\omega)M(\bar{\gamma},\Omega). \quad (9)$$

or

$$\prod_{i=1}^{n} M(\gamma_i,\theta_i) = \cos\bar{\gamma}R(\omega) + i\sin\bar{\gamma}ER(2\Omega-\omega). \quad (10)$$

where γ_i are an arbitrary set of retardations and θ_i are an arbitrary set of angular positions, ω is the amount of rotation, Ω is the position angle of the retardation axis, and $\bar{\gamma}$ is the retardation of the combination. Since the γ_i are functions of frequency, $\bar{\gamma}$, ω, and Ω will in general vary with frequency. It will be our aim in this paper to demonstrate waveplate combinations for which $\bar{\gamma}$ varies more slowly with frequency than it does for a single waveplate, while Ω, the tilt of the retardation axis, remains nearly fixed.

For the purpose of this paper, we will assume that the birefringence of an individual retarder is constant, so that the variation in retardance of a single plate can be written

$$\gamma(\epsilon) = \gamma^0(1+\epsilon). \quad (11)$$

where γ^0 is the half retardation at the design (center) frequency ν^0 and

$$\Delta\nu = \nu - \nu^0. \quad (12)$$

$$\epsilon = \Delta\nu/\nu^0. \quad (13)$$

Epsilon will be referred to as the relative frequency difference. The superscript zero will denote retardation at the design frequency.

Three-Element Combinations

It is easy to show that there are no achromatic waveplates with only two elements. The general three-element combination $M3$ can be written as

$$\begin{aligned} M3 &= M(\gamma_3,\theta_3)M(\gamma_2,\theta_2)M(\gamma_1,\theta_1) \\ &= \cos\gamma_3\cos\gamma_2\cos\gamma_1 I \\ &\quad -\{\cos\gamma_3\sin\gamma_2\sin\gamma_1 R[2(\theta_1-\theta_2)] \\ &\quad +\cos\gamma_2\sin\gamma_3\sin\gamma_1 R[2(\theta_1-\theta_3)] \\ &\quad -\cos\gamma_1\sin\gamma_3\sin\gamma_2 R[2(\theta_2-\theta_3)]\} \\ &\quad +iE\{-\sin\gamma_3\sin\gamma_2\sin\gamma_1 R[2(\theta_3-\theta_2+\theta_1)] \\ &\quad +\sin\gamma_3\cos\gamma_2\cos\gamma_1 R(2\theta_3) \\ &\quad +\sin\gamma_2\cos\gamma_3\cos\gamma_1 R(2\theta_2) \\ &\quad +\sin\gamma_1\cos\gamma_3\cos\gamma_2 R(2\theta_1)\}. \end{aligned} \quad (14)$$

Although Eq. (14) looks formidable, it really has a quite tractable form. The requirement that the combination be a pure waveplate means from Eq. (10) that

$$\mathrm{Real}(M3_{12}) = \mathrm{Real}(M3_{21}) = 0. \quad (15)$$

With no loss of generality, θ_1 can be set equal to zero, since choosing θ_1 only locates the combination in space. Then, from Eqs. (14) and (15)

$$\sin2(\theta_2-\theta_3) = (\tan\gamma_1/\tan\gamma_3)\sin2\theta_2 + (\tan\gamma_1/\tan\gamma_2)\sin2\theta_3. \quad (16)$$

The result of Eq. (16) has the form of the sine addition relation, which implies

$$\cos(2\theta_3) = \tan\gamma_1/\tan\gamma_3, \quad (17)$$

and

$$\cos(2\theta_2) = -(\tan\gamma_1/\tan\gamma_2). \quad (18)$$

Both θ_2 and θ_3 are fixed in space and are thus independent of frequency. In order that the relation of Eq. (16) hold, either the ratio of the tangents given by Eq. (17) and (18) must be constant and equal to plus or minus one or the relation of Eq. (16) must be independent of Eq. (17) or (18). If

$$\gamma_1 = \pm\gamma_3, \quad (19)$$

$\cos(2\theta_3)$ has magnitude unity and the sine of $2\theta_3$ is zero. Hence, the relation of Eq. (16) holds. If

$$\gamma_1 = \pm\gamma_2. \quad (20)$$

similarly the sine of $2\theta_2$ will be zero and the relation of Eq. (16) will also hold. But Eq. (20) represents a redundant solution in that the first and second plates are parallel and could be replaced by a single plate.

For the purposes of this paper, it will be assumed that all plates are of the same material. Only the positive solution of Eq. (19) will be used. The relation of Eq. (17) then implies

$$\theta_1 = \theta_3. \quad (21)$$

Thus, it has been shown that the parallelism and equality of the input and output plates is a consequence of the requirement that the combination be a waveplate and be made of materials with the same sign index difference.

Then, with Eqs. (20) and (21), the matrix Eq. (14) becomes

$$\begin{aligned} M3 &= (\cos2\gamma_1\cos\gamma_2 - \sin2\gamma_1\sin\gamma_2\cos2\Delta)R(0) \\ &\quad + iE[-\sin^2\gamma_1\sin\gamma_2 R(-2\Delta) + \cos^2\gamma_1\sin\gamma_2 R(2\Delta) \\ &\quad + \sin2\gamma_1\cos\gamma_2 R(0)] \end{aligned} \quad (22)$$

where $\Delta = \theta_2$.

From Eq. (22) it is possible to calculate the conditions under which the combination is achromatic. For a simple waveplate, the first derivative of gamma with respect to epsilon is non zero. At the least, for a waveplate combination to be achromatic, it is required that

$$\left.\frac{\partial\bar{\gamma}}{\partial\epsilon}\right|_{\epsilon=0} = 0.$$

From Eqs. (22) and (10),

$$\cos\bar{\gamma} = \cos2\gamma_1\cos\gamma_2 - \sin2\gamma_1\sin\gamma_2\cos2\Delta. \quad (23)$$

so that

$$\begin{aligned} \left.\frac{\partial\bar{\gamma}}{\partial\epsilon}\right|_{\epsilon=0} &= \frac{1}{\sin\bar{\gamma}}[\sin2\gamma_1\cos\gamma_2(\gamma_2\cos2\Delta - 2\gamma_1) \\ &\quad - \cos2\gamma_1\sin\gamma_2(2\gamma_1\cos2\Delta + \gamma_2)] \\ &= 0. \end{aligned} \quad (24)$$

Equation (24) has three groups of solutions:

(1): (1) $\cos2\Delta = 2\gamma_1^0/\gamma_2^0$ (25)

(2) $\cos2\gamma_1^0\sin\gamma_2^0 = 0$: (26)

(2): (1) $\cos2\Delta = -(\gamma_2^0/2\gamma_1^0)$. (27)

(2) $\sin2\gamma_1^0\cos\gamma_2^0 = 0$: (28)

(3): (1) $2\gamma_1^0 = \pi/2$, (29)

(2) $\gamma_2^0 = \pi/2$. (30)

All of the solutions, (1), (2), and (3) yield waveplates that are independent, to first order, of the relative difference. However, the location of the retardation axis Ω may vary to first order. The condition that the retardation axis be stationary to first order is

$$\frac{\partial \Omega}{\partial \epsilon}\Big|_{\epsilon=0} = 0. \quad (31)$$

From the imaginary part of Eq. (10),

$$\tan 2\Omega = \text{Imag } (M_{21})/\text{Imag } (M_{11}). \quad (32)$$

so

$$\frac{\partial \Omega}{\partial \epsilon}\Big|_{\epsilon=0} = \frac{\cos^2 2\Omega}{2M_{11}}\Big[\gamma_2 \cos\gamma_2 \sin 2\Delta - \frac{M_{21}}{M_{11}} \times [-\sin 2\gamma_1 \sin\gamma_2 (2\gamma_1 \cos 2\Delta + \gamma_2) + \cos 2\gamma_1 \cos\gamma_2 (\gamma_2 \cos 2\Delta + 2\gamma_1)]\Big] = 0. \quad (33)$$

Only solution (2) and the condition

$$\gamma_2^0 = \pi/2 \quad (34)$$

cause the first derivative of Ω to be zero for all values of achromatic retardation. There are, however, isolated solutions for specific values of γ^0 that also yield a zero value of Eq. (33). Thus, it is seen that the requirement that the middle element be a half waveplate is necessary so that the retardation axis be stationary and the solution be general.

With solution (2) and Eq. (34), the second derivatives of γ and Ω can be calculated by Taylor's formula to second order.

$$\bar{\gamma} = \bar{\gamma}^0 - \frac{1}{2}\left(\frac{1}{\tan\bar{\gamma}^0}\right)[(2\gamma_1^0)^2 - (\pi/2)^2]\epsilon^2, \quad (35)$$

and

$$\Omega = \Omega^0 + \left(\frac{\sin 4\Delta}{8}\right)\left(\frac{\cos 2\gamma^0}{\sin^2\bar{\gamma}^0}\right)[(2\gamma_1^0)^2 - 2(\pi/2)^2]\epsilon^2 \quad (36)$$

From Eqs. (35) and (36), it can be seen that for all retardations other than a half wave, a three-element plate can be made with the desired retardation at two frequencies. For a half waveplate, the second-order term in the retardation is identically zero.

Achromatic Range of Three-Element Waveplates

For waveplates other than half wave, it is clear from Eq. (35) that for frequencies near the design frequency, the retardation decreases quadratically with the relative frequency difference. Hence, if the retardation at the design frequency is chosen to be greater than the desired retardation, the desired retardation will be attained at two frequencies; one above and one below the design frequency. With Eq. (35) it is simple to calculate the range in the relative frequency ϵ, so that the retardation is within a tolerance of $\pm\beta$. By recalling that γ is half the retardation of the combination, at the design frequency the retardation should be $\bar{\gamma}^0 + \beta/2$, while at ϵ_β the retardation should be $\bar{\gamma}^0 - \beta/2$. So,

$$\epsilon_\beta = \{\beta \tan(\bar{\gamma}^0 \pm \beta/2)/[(2\gamma_1)^2 - (\pi/2)^2]\}^{1/2}, \quad (37)$$

where γ_1^0 and Δ are obtained from

$$\cos(\bar{\gamma}^0 \pm \beta/2) = (\pi/2)(\sin 2\bar{\gamma}_1^0/2\bar{\gamma}_1^0), \quad (38)$$

and

$$\cos 2\Delta = -(\pi/2)[1/(2\gamma_1^0)]. \quad (39)$$

Equation (38) is derived by evaluating Eq. (23) at the design frequency and by substituting for cosine 2Δ with Eq. (39).

Plotted in Figs. 1 and 2 are the differences in retardation and tilt, respectively, from their values at the design frequency for several values of $2\bar{\gamma}^0$ as a function of epsilon, the relative frequency difference. The values for the figures were calculated with the program WPLATE, which can calculate properties of up to one hundred waveplate combinations. Program WPLATE uses the Jones matrix equations to exactly calculate the properties of combination waveplates. However, for retardation tolerances of 2° or less, the Taylor series solutions are adequate.

For retardation of other than a half wave, from Eq. (38) the outer waveplates will have a retardation less than a half wave. For example, for a quarter waveplate $2\bar{\lambda}_1^0 = 115.5°$. Because of the availability of

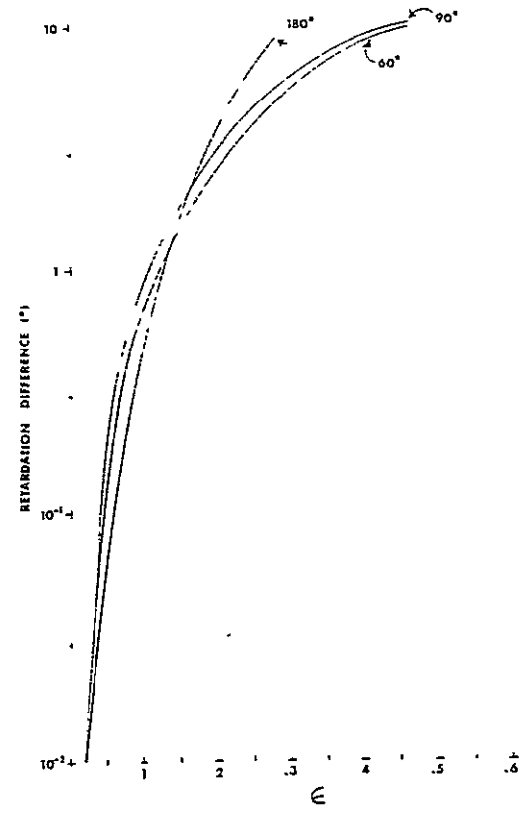


Fig. 1. Retardation difference vs relative frequency difference ϵ for 180°, 90°, and 60° three-element combinations. For 180° retardation $2\gamma_1 = 180°$, $2\gamma_2 = 180°$, $\Delta = 60°$; for 90° retardation $2\gamma_1 = 115.5°$, $2\gamma_2 = 180°$, $\Delta = 70.6°$; for 60° retardation $2\gamma_1 = 101.75°$, $2\gamma_2 = 180°$, $\Delta = 76.096°$.

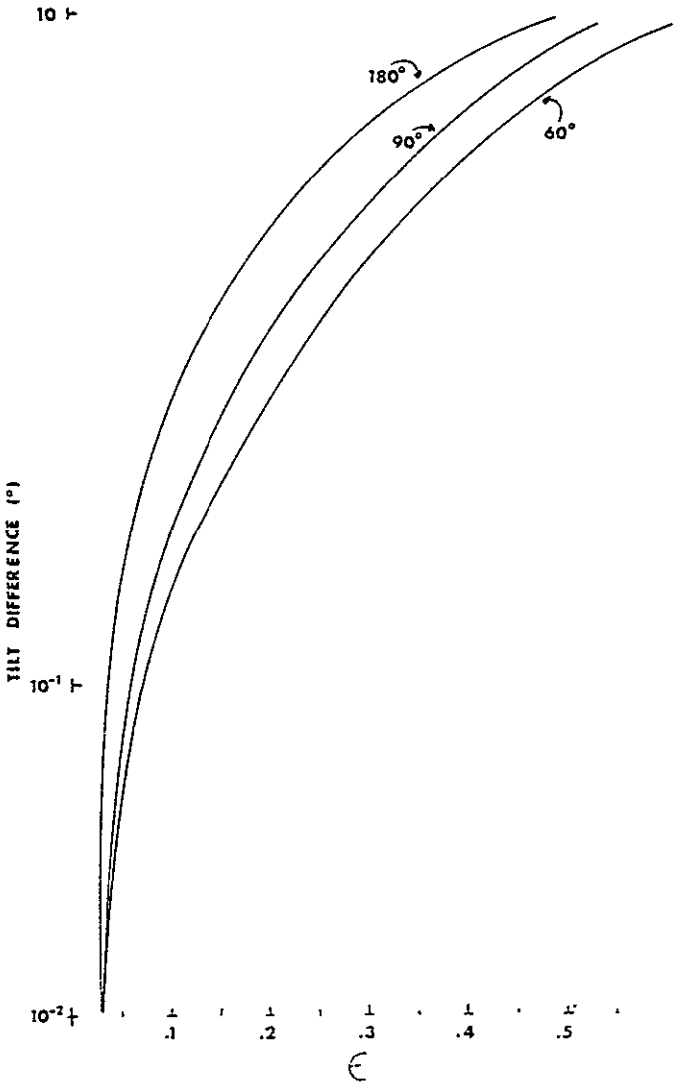


Fig. 2. Tilt difference vs relative frequence difference ϵ for the above three-element combinations.

quarter and half wave material, the use of four-element designs discussed below should be considered, since any retardation can be achieved with only quarter and a half wave material.

For half waveplates, the coefficient of the quadratic term in Eq. (35) vanishes, because $2\bar{\gamma}_1{}^0$ is a half wave at the design frequency. Evaluation of the third derivative combination shows that the variation in retardation is cubic. For nonhalf wave combinations, range adjustment was accomplished by adjusting the center frequency retardation. For the half wave case there is no advantage to this. However, another possibility exists: Δ can be varied to allow a first-order term that can cancel the third-order term in the vicinity of the tolerance range frequency.

It is quite straightforward to calculate the effect of adjusting Δ. By expanding Eq. (23) in terms of epsilon,

$$\cos\gamma = \sin(\pi/2)\epsilon[A\cos^2(\pi/2)\epsilon - \sin^2(\pi/2)\epsilon], \quad (40)$$

where

$$A = (1 + 2\cos 2\Delta). \quad (41)$$

By expanding the sines and cosines to third order,

$$\cos\bar{\gamma} = (\pi/2)\epsilon\{A - [1 + (7/6)A][(\pi/2)\epsilon]^2\}. \quad (42)$$

In general, A will be small compared to 1. This can be seen by evaluating A in terms of small changes in plate angles. For a half waveplate the unadjusted delta is $\pi/3$. So if

$$\Delta = (\pi/3) - \delta, \quad (43)$$

where δ is the small variation,

$$A = 2\sqrt{3}\delta. \quad (43)$$

For δ equal to 1°,

$$A = 0.061. \quad (44)$$

With the assumption that δ is small, Eq. (41) becomes

$$\cos\bar{\gamma} = (\pi/2)\epsilon\{A - [(\pi/2)\epsilon]^2\}. \quad (45)$$

The retardation is $\pi/2$ when

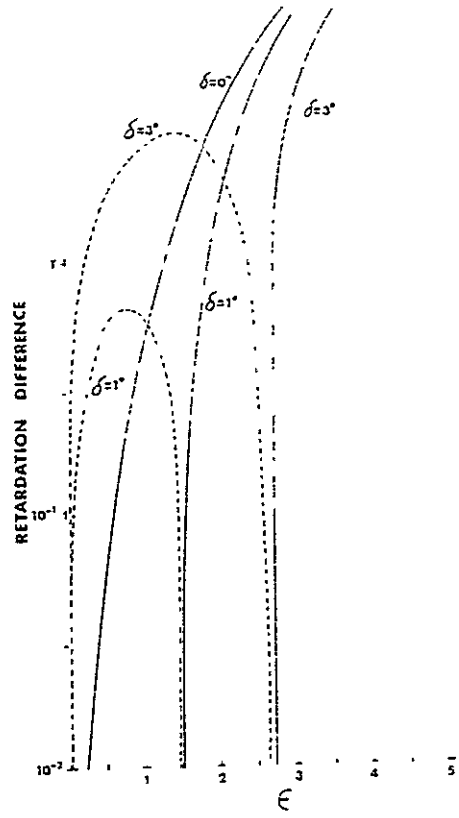


Fig. 3. Retardation difference vs relative frequency difference ϵ for three-element combination half waveplates for several adjustments of the central plate. The dashed curves indicate negative values of the relative retardation differences. All three waveplates have 180° retardation. The angle from the central plate is 60° minus δ.

246

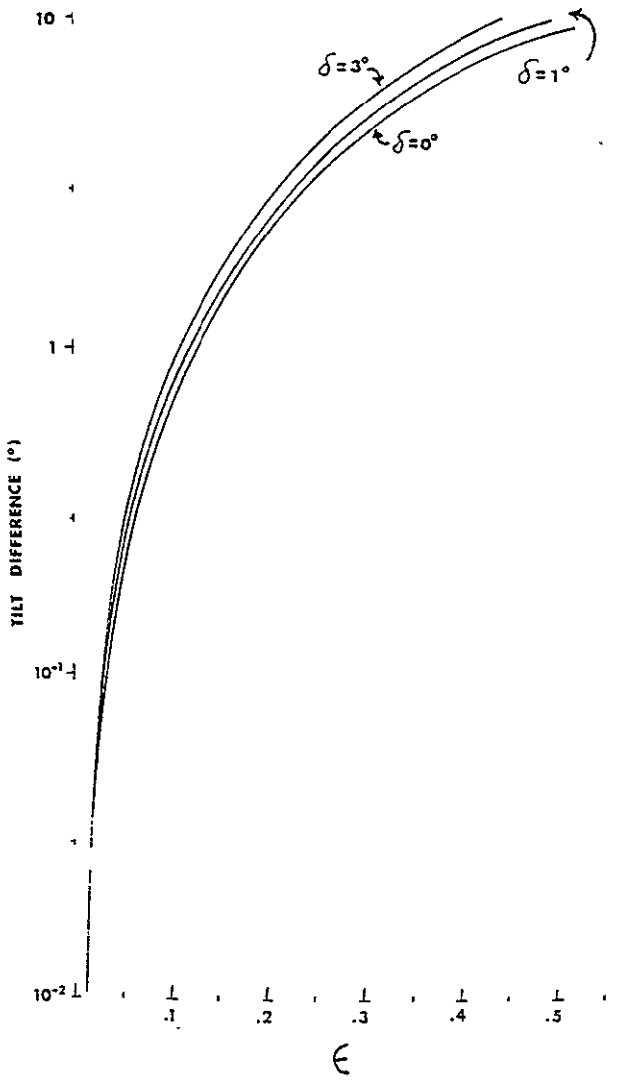


Fig. 4 Tilt difference vs relative frequency difference ϵ for three-element combination half waves for several adjustments of the central plate.

$$\epsilon|_{\tau/2} = \pm A^{1/2}(2/\pi) = \pm (2/\pi)(2\sqrt{3}\delta)^{1/2} = \pm 1.185\delta^{1/2}. \quad (46)$$

The local extremal value of the retardation occurs when

$$\epsilon_{ex} = \pm (2/\pi)[(1/\sqrt{3})A^{1/2}] = \pm 0.684\delta^{1/2} \quad (47)$$

and the extremal values are

$$\cos\overline{\gamma}_{ex} = \pm 2^{5/2}3^{-3/4}\delta^{3/2} = \pm 2.482\,\delta^{3/2}. \quad (48)$$

Now, if the retardation tolerance is β,

$$\beta/2 = \cos^{-1}[(2/3\sqrt{3})(2\sqrt{3}\delta)^{3/2}]. \quad (49)$$

or

$$\delta = 0.55(\beta/2)^{2/3}. \quad (50)$$

The tolerance β will be exceeded when

$$\epsilon > \epsilon_{ex} \pm \beta/4A. \quad (51)$$

Hence, the range in which the retardation is within some tolerance β can be rewritten as

$$\epsilon_\beta \simeq 1.04(\beta/2)^{1/3}. \quad (52)$$

Shown in Figs. 3 and 4 are the retardation and tilt differences from the values at the design frequence vs epsilon for several values of δ. In Fig. 5 is shown the tolerance β vs the change in angle of the central plate δ. In Fig. 6 is the tolerance vs range in frequency difference ϵ_R.

In order to get an idea of the improvement in range obtained by adjusting δ, note that if A is zero, the first derivative equals zero. Then a glance at Eq. (40) shows that

$$\epsilon_\beta \approx (2/\pi)(\beta/2)^{1/3}$$
$$\epsilon_\beta = 0.637(\beta/2)^{1/3}. \quad (53)$$

Thus, by optimizing δ for a given tolerance, the range for that tolerance increased by 57%.

The price paid for nonzero values of δ, as seen in Fig. 3, is that for small values of ϵ the retardation difference is greater than for zero δ.

Nine-Element Plate

Since the half waveplate of three elements has a cubic retardation variation and is made of three half waveplates that have a linear retardation variation,

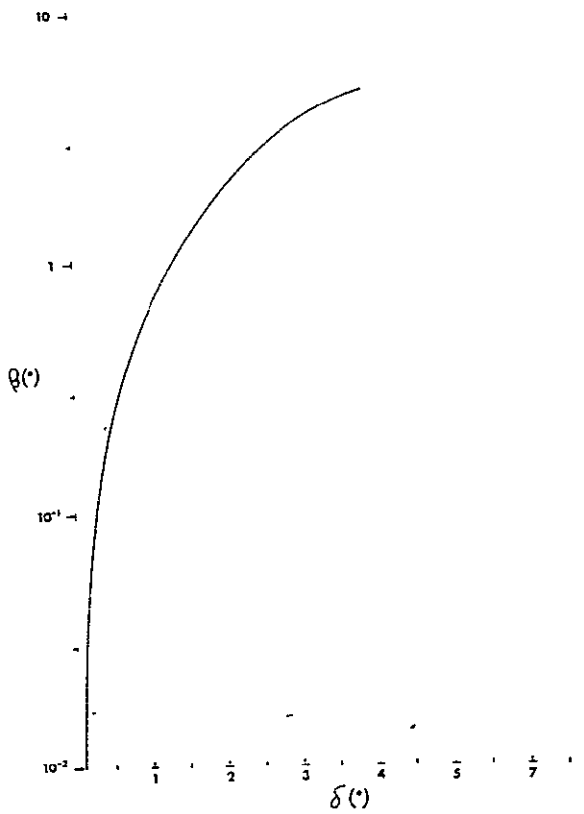


Fig. 5. Tolerance β vs adjustment angle δ for three-element combination half waveplates.

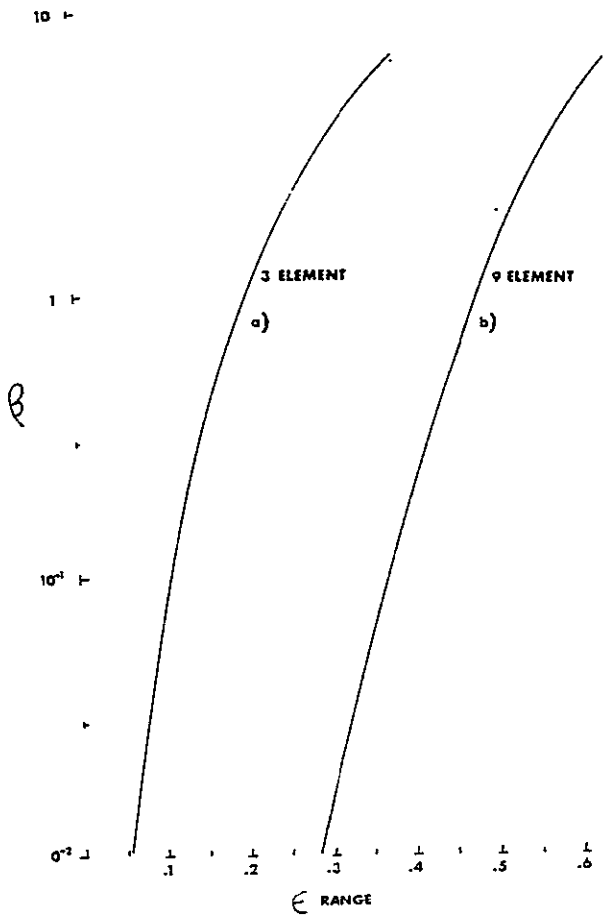


Fig. 6. Tolerance β vs the range of the relative frequency difference ϵ_β for (a) three-element and (b) nine-element combination half waveplates.

there is reason to suspect that a combination of three three-element plates may be superachromatic. The matrix for the three-element half waveplate is

$$M3 = \cos^3\gamma_2 1 + i \sin\gamma_2 E[-R(-2\Delta + \cos^2\gamma_2 R(0)], \quad (54)$$

where $\gamma_2{}^0 = \pi/2$ and $\Delta = \pi/3$, which has an appearance similar to that of a simple waveplate, except for the third-order dependence of the real (retardation) term and an additional second-order term in the imaginary (tilt) term.

There are two possibilities for the nine-element plate. The center three-element group may be rotated in the same direction as the central element of the three-element group, or in the opposite direction. The two nine-element matrices are

$$M9S = M3R[(-\pi/3)]M3R(\pi/3)]M3, \quad (55)$$

and

$$M90 = M3R(\pi/3)M3R[-(\pi/3)]M3. \quad (56)$$

With appropriate rotations the matrices can be put in the form

$$M9S = \cos^9\gamma_2 1 + i \sin\gamma_2 E[R(0) - \cos^2\gamma_2 R(120)(1 - \cos^6\gamma_2) + \cos^8\gamma_2 R(-120)], \quad (57)$$

and

$$M90 = \cos^9\gamma_2 1 + i \sin\gamma_2 E[R(0) - \cos^2\gamma_2 R(120) + \cos^6\gamma_2 R(0) + \cos^8\gamma_2 R(0)]. \quad (58)$$

For both systems the retardation variation will be ninth-order in epsilon:

$$\cos\bar{\gamma} = \sin^9(\pi/2)\epsilon. \quad (59)$$

So, for a tolerance β, the range in epsilon is

$$\epsilon_\beta = (2/\pi)\sin^{-1}[\sin(\beta/2)]^{1/9}. \quad (60)$$

As with the three-element plate, the range for a given tolerance can be increased by varying δ. A computer evaluation shows that either all the angles between plates can be changed by δ or the angle of the central group alone can be adjusted. There is a slight advantage in range to adjusting all the plates. However, if the nine-element system is made of three three-element subsystems, there is a significant practical advantage to adjusting just the central group.

The value of δ for a given tolerance β is the same for nine-element systems as for three-element half waveplates (see Fig. 5). However, the value ϵ_β for a given β is greatly increased. To show the increase more clearly, β vs ϵ_β for a nine-element system is shown in Fig. 6(b), while the curve for a three-element system is in Fig. 6(a). In the calculation for Fig. 6(b), the angle of the center three-element group was varied.

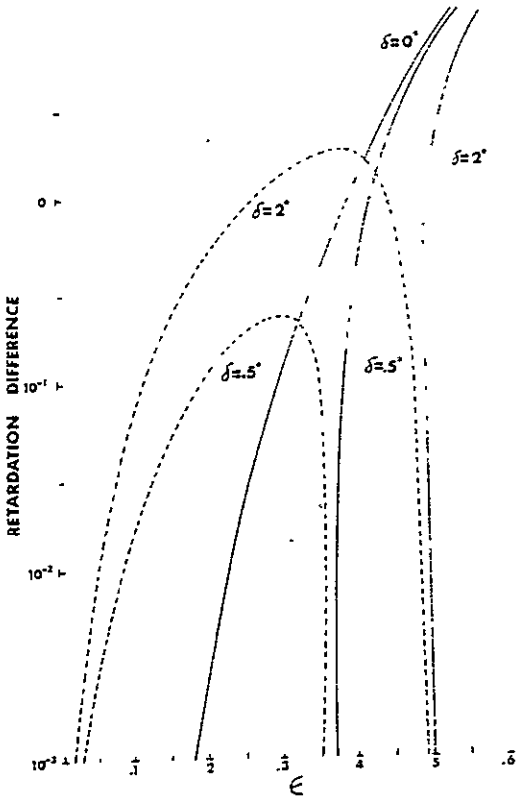


Fig. 7. Retardation difference vs relative frequency difference ϵ for nine-element combination half waveplates for several adjustments of the central group of three elements.

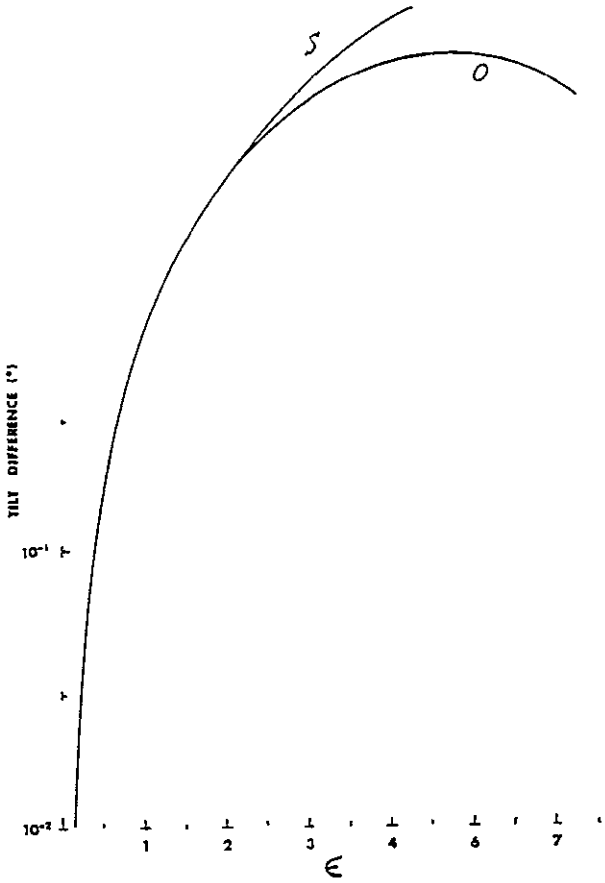


Fig. 8. Tilt difference vs relative frequency difference for the S and O configurations of the central group of three-elements.

From Fig. 6(b), $\epsilon = 0.48$ for a tolerance of 1°. If the design wavelength is λ 5200 Å, the range in wavelength for 1° tolerance band extends from 3514 Å to 10,000 Å. For a 2° tolerance the band is from 3510 Å to 10,947 Å.

The difference in retardation vs ϵ for several values of δ is shown in Fig. 7.

Although the retardation variation for $M9S$ and $M9O$ are the same, the variation in tilt is not. Shown in Fig. 8 are plots of difference in tilt vs ϵ for the S and O cases, for $\delta = 0$. Both the S and the O cases have less tilt variation than the three-element half waveplate, as can be seen by comparison with Fig. 4.

Four-Element Waveplate Rotators

The three-element half waveplate has a very symmetric structure.

$$M3 = M(\pi/2, 0)M(\pi/2, \pi/3)M(\pi/2, 0). \quad (61)$$

If this structure is split in half and two halves rotated with respect to each other, the resulting matrix has the form

$$M4 = R(-\alpha)M(\pi/2, 0)M(\pi/4, \pi/3)R(\alpha) \times M(\pi/4, \pi/3)M(\pi/2, 0). \quad (62)$$

This equation can be put in the form

$$M4 = \sin\alpha\, R(90 - \alpha) - \sin^3(\pi/2)\epsilon\, \cos\alpha\, R(-\alpha) \pm (i/2)\sin\pi\epsilon\, \cos\alpha\, E\{-R[\alpha - 2(\pi/3)] \pm \sin^2(\pi/2)\epsilon\, R(\alpha)\}. \quad (63)$$

From the real part of Eq. (63), the retardation and rotation can be obtained, since by Eq. (10),

$$\sin\omega\, \cos\bar{\gamma} = \text{Real}\,(M4_{21}).$$
$$\cos\omega\, \cos\bar{\gamma} = \text{Real}\,(M4_{11}). \quad (64)$$

After some algebra,

$$\bar{\gamma} = \cos^{-1}[\sin^2\alpha \pm \cos^2\alpha\, \sin^6(\pi/2)\epsilon]. \quad (65)$$

So that when

$$\sin^6(\pi/2)\epsilon \ll \sin^2\alpha, \quad (66)$$

$$\cos\bar{\gamma} \simeq \cos(90 - \alpha)\left\{1 \pm \frac{1}{2}[\sin^6(\pi/2)\epsilon/\tan^2\alpha]\right\}, \quad (67)$$

$$\tan\omega \simeq \tan(90 - \alpha)\{1 \pm [\sin^3(\pi/2)\epsilon/\sin^2\alpha]\}. \quad (68)$$

Equation (67) indicates the particularly pleasant result that all retardations can be achieved by merely rotating a pair of combination plates with respect to each other. The resulting retardation is simply twice the complement of the angle between the plates. Accompanied by the retardation, as shown by Eq. (68), is a rotation by the complement of the separation

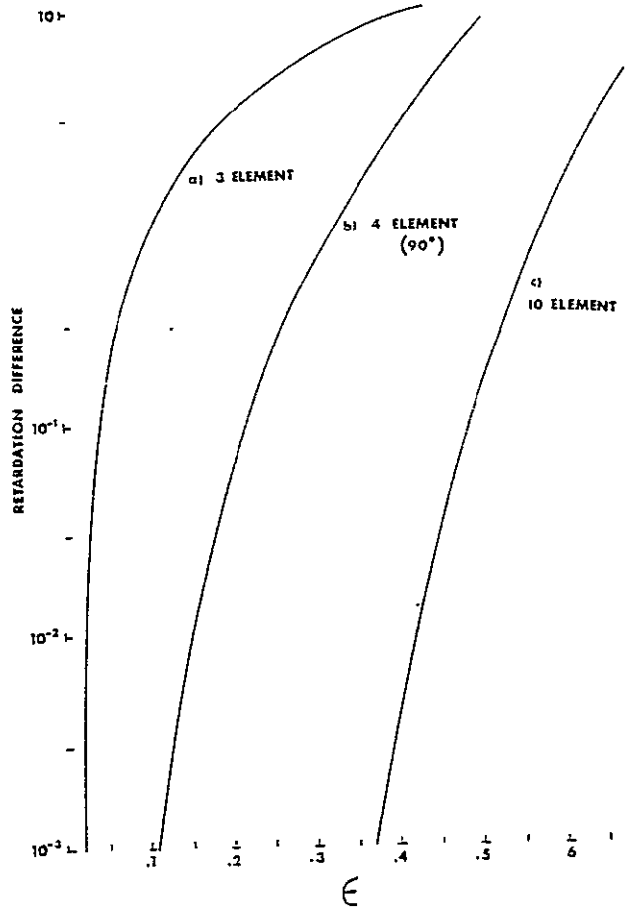


Figure 9. Retardation difference versus relative frequency difference for 3(a), 4(b), and 10(c). element combination quarter waveplates.

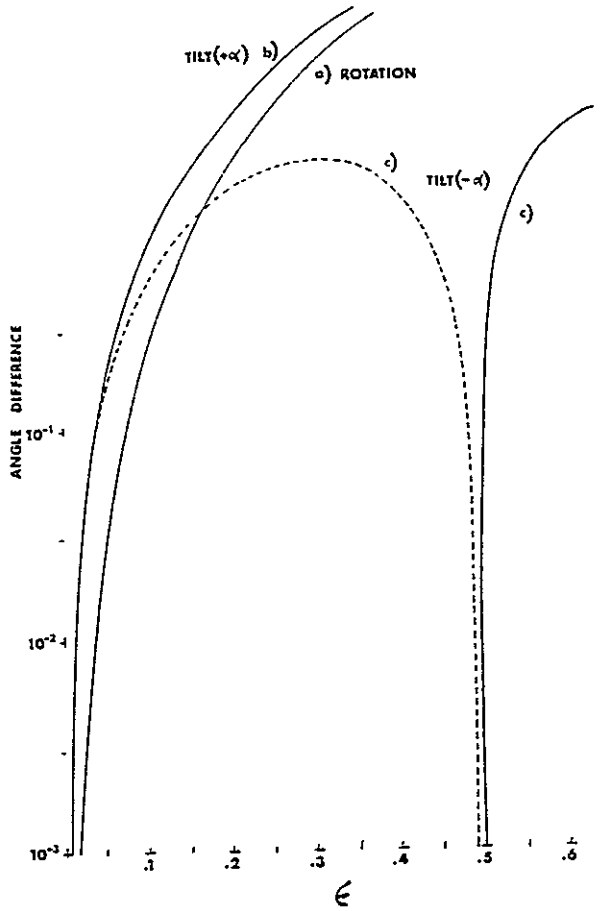


Figure 10. Rotation difference (a), and tilt difference for positive (b), and negative (c), values of α versus relative frequency difference, ϵ.

angle. The behavior of the retardation axis can be obtained from the imaginary part of Eq. (63). Thus

$$\tan[2\Omega - (90 - \alpha)] \approx \tan[(2\pi/3) + \alpha] \times \left[1 - \sqrt{3}\sin^2(\pi/2)\epsilon/\sin 2\left(\frac{2\pi}{3} + \alpha\right)\right] \quad (69)$$

Note that in Eq. (69) to second order the location of the retardation axis is independent of α, while from Eqs. (67) and (68), it is clear that the magnitude of the retardation and the rotation variations are unaffected by the sign of α. The retardation differences vs ϵ for three- and four-element quarter waveplates are shown in Fig. 9(a) and (b). Shown in Fig. 10(a), (b), and (c) are the differences in rotation and tilt with positive and negative values of α, respectively, vs ϵ for the four-element quarter wave retarder.

Equation (62) was written down with only a short plausibility argument. However, it is possible to show, with arguments similar to those for three-element combinations, that there are no other combinations of four half and quarter waveplates that are achromatic in both retardation and tilt.

Ten-Element Systems

Since the three-element half waveplate was directly extended to a superachromatic nine-element plate, it is reasonable to expect that by splitting the nine-element half waveplate in a manner similar to the three-element combination, a superachromatic wave plate rotator may be constructed. In fact, such a ten-element combination is quite successful. A matrix representation of a ten-element plate is

$$M10 = R(-\alpha)M3M(\pi/2, \pi/3)M(\pi/4, 2\pi/3)R(\alpha) \times M(\pi/4, 2\pi/3)M(\pi/2, (\pi/3)M3. \quad (70$$

The mathematical verification of properties of the ten-element plate will not be carried out here, but the results will be stated. The retardation is

$$\cos\bar{\gamma} = \cos(90 - \alpha)\{1 + [\sin^{18}(\pi/2)\epsilon/2\tan^2\alpha]\}, \quad (71$$

and the rotation is

$$\tan\omega \simeq \tan(90 - \alpha)\{1 + [\sin^9(\pi/2)\epsilon/\sin^2\alpha]\}. \quad (72$$

Shown in Fig. 9(c) is the retardation variation of a ten-element quarter waveplate vs. ϵ. As can be seen from Fig. 9(c), the ten-element quarter waveplate gives $\epsilon_\beta = 0.59$ for $\beta = 1°$. If the center wavelength is 5500 Å, the waveplate is within a degree from 3400 Å to 13,500 Å.

Equation (70) represents a split of $M9S$, the central three-element group rotated in the same direction as the central plate; a split of $M90$, the central group rotated opposite to central plate, also gives rise to an achromatic rotator retarder. However, the positive and negative rotation of the two halves has slightly different effects on tilt.

Experimental

A number of sets of three-element half waveplates have been constructed of Polaroid Retarder plastic waveplate with the procedure described in Ref. 1. They are quite achromatic and appear to agree generally with theoretical predictions. Nine-element combinations have been constructed from three groups of three-element plates. While it is clear that they are more achromatic than the three-element combinations, precise measurements have not yet been made. There is some difficulty because such plates, made of the standard plastic plates, are within a degree in retardation from 3510 Å to 10,000 Å. Arrangements have been made to use an accurate polarimeter in order to measure the nine-element plates. This work will be reported on as soon as completed.

The transmission of three group combinations is typically 99% throughout the visible.

Combinations of four- and ten-element groups have also been constructed. These also generally behave as predicted. But again, they are impossible to measure accurately without a more accurate polarimeter.

Discussion

The prime goal in this series of papers is to improve Lyot filters. A key aim is to achieve tunability over a wide spectral range. Tunability requires (1) achromatic half waveplates; (2) achromatic quarter waveplates; and (3) broad-band polarizers.

With nine-element half waves and ten-element quarter waves, it is possible to achieve achromaticity to within a degree from 3500 Å to 11,000 Å. The only remaining problem is polarizing materials that can operate over the entire visible spectrum. The problem of polarizers will be discussed in some of the following papers in this series.

The matrix method described above is very useful for searching out achromatic plate combinations. The form of the achromatic three-element matrix was the suggestive clue for the superachromatic nine-element combination. The three-element matrix split in half formed the basis for the four- and ten-element achromatic waveplate rotator combinations.

Evaluation of three-element systems also shows that achromatic rotators cannot be formed. However, systems that are achromatic rotators with respect to a single fixed axis can be formed. Koester[5] has described such single-axis 90° achromatic rotators. The Koester-type devices do not work as rotators for Lyot wide-field elements, since they only rotate one axis of polarization achromatically.

I would like to thank H. E. Ramsey for the construction of the three-, four-, nine-, and ten-element plates and for tests on Lyot filter elements.

This work was supported by Lockheed independent research funds.

References

1. A. Title, Solar Phys. in press.
2. S. Pancharatnam, Proc. Indian Acad. Sci. A41, 137 (1955).
3. S. E. Harris, and C. M. McIntyre, J. Opt. Soc. Am. 58, 1575 (1968).
4. R. C. Jones, J. Opt. Soc. Am. 31 500 (1941).
5. Koester, C. J. J. Opt. Soc. Am. 49, 560 (1959).

Reprinted from APPLIED OPTICS, Vol. 14, page 445, February 1975

Improvement of Birefringent Filters. 3: Effect of Errors on Wide Field Elements

Alan M. Title

The properties of nontunable and tunable Lyot wide field elements are examined when the components of the elements deviate from their proper values. Special emphasis is put on determining what variations cause light to be transmitted at the transmission minima. The analysis shows that the nine- and ten-element plastic waveplates described in Paper 2 of this series can be used to make a Lyot filter that is tunable from 3500 Å to 10,000 Å.

Introduction

This is the third in a series of papers dealing with the improvement of birefringent filters. The principal purpose of this paper is to critically examine the effects of deviations from perfection of tunable and nontunable wide field elements. The method of the analysis is to expand the error effect terms in powers of the sine while retaining the analytic form of the wavelength-dependent terms. The value of such an analysis is that it clearly distinguishes between those classes of imperfections that cause asymmetric transmission functions and those that cause additional transmission at the transmission minima. A pure power series analysis such as carried out by Jefferies and Giovanelli[1] does not distinguish so clearly between the asymmetry and transmission minima effects. In general, asymmetry errors introduce less light outside the primary maximum than do transmission-minima errors.

The prime goal of this paper is to provide the design tradeoffs for widefield Lyot elements that use achromatic waveplates such as those discussed in Paper 2[2] of this series. For multielement achromatic waveplates the properties that vary are the retardation, the rotation, and the location of the fast axis. The analysis that follows will allow deviations in retardation and in the fast axis location. Rotation is neglected because half waveplates have none; for quarter waveplates rotation is unimportant. In this and all successive papers in the series the word *tilt* will be used to describe deviation of the fast axis from its nominal position.

In the first section below, a standard wide field element whose halves may not have exactly the same retardation and whose central half-waveplate may not have exactly 180° retardation or be exactly aligned at equal angles to the two halves of the element will be discussed. This configuration is examined for three basic reasons. First, the two halves of a wide field element are seldom exactly the same in real elements. Second, the central half-waveplate will in general deviate from 180° as a function of wavelength. Third, when multiple plate achromatic half-waveplates are used, the retardation axis of the combination plate can vary.

The following section concerns tunable elements. It is current practice to tune Lyot elements by rotating a polarizer with respect to a quarter-waveplate at the end of the element. The quarter-waveplate's retardation axis is at 45° to the element's axis. The tuning properties will be examined as a function of deviation from quarter-wave retardation and exact angular location with respect to the element. Again the rationale for the investigation is the understanding of the error effects so that achromatic plates can be used most effectively.

There are other methods for tuning elements that involve additional waveplates. In one the entrance linear polarizer is replaced by a circular polarizer. Then each element can be tuned individually. In the single quarter-waveplate method rotating the exit polarizer forces all succeeding elements to be rotated. A second method of tuning is accomplished by rotation of a half-waveplate between the exit polarizer and the quarter-waveplate. With this technique only the half-waveplate need be rotated.

Wide Field Element

With the notation of Jones,[3] the matrix for a wide field element with the variances noted above is

The author is with Lockheed Solar Observatory, Lockheed Palo Alto Research Laboratory, Department 52-14, Palo Alto, California 94304.

Received 11 March 1974.

$$M = R(-45)G(\alpha)R(45)R(-\delta)G(\gamma)R(\delta)R(45)G(\beta)R(-45), \quad (1)$$

where

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad (2)$$

$$G(\gamma) = \begin{pmatrix} \exp(+i\gamma) & 0 \\ 0 & \exp(-i\gamma) \end{pmatrix}, \quad (3)$$

$$\alpha = 2\pi\Delta n d_\alpha/\lambda, \quad (4)$$

$$\beta = 2\pi\Delta n d_\beta/\lambda, \quad (5)$$

$$\gamma = (\pi/2) + \epsilon, \quad (6)$$

d_α and d_β are the thickness of the two halves of the element, λ is the wavelength, Δn the index difference of the crystal, ϵ the half deviation from half-wave retardation, and δ the deviation from correct location of the axis. In this analysis it will be assumed that the birefringence is wavelength-independent. The matrix mathematics simplifies somewhat if the expression for a retardation plate at angle θ is written in the form

$$R(-\theta)G(\gamma)R(\theta) = \cos\gamma\, 1 + i \sin\gamma\, ER(2\theta), \quad (7)$$

where

$$E = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (8)$$

Note that

$$E^2 = 1, \quad (9)$$

and

$$ER(\theta) = R(-\theta)E. \quad (10)$$

Rewriting Eq. (1) using the form of Eq. (7), the matrix for the element is

$$M = [\cos\alpha\, 1 + i \sin\alpha\, ER(90)][\cos\gamma\, 1 + i \sin ER(2\delta)] \times [\cos\beta 1 + i \sin\beta ER(-90)]. \quad (11)$$

Let $\beta = \alpha + \xi$, and $\eta = 2\alpha + \xi$, then the total retardation of the element is η and the difference in retardation of the two halves is ξ.

When used between parallel perfect polarizers, the transmission amplitude of the element is just M_{11}. Direct evaluation of Eq. (11) shows

$$M_{11} = \cos\gamma \cos\xi + \sin\gamma \sin2\delta \sin\xi + i \sin\gamma \cos2\delta \cos\eta. \quad (12)$$

It is also reasonable to operate a wide field element with crossed polarizers, in which case the transmission amplitude is

$$M_{12} = -\sin\eta \cos2\delta \sin\gamma + i[\cos\xi \sin2\delta \sin\gamma + \cos\gamma \sin\xi]. \quad (13)$$

Since the main region of concern is when the retardation of the central element is nearly a half-wave, it is reasonable to expand in terms of ϵ. Then

$$T_\parallel = M_{11}M_{11}^{\#} = \cos^2\eta \cos^2\epsilon \cos^2 2\delta + \sin^2\epsilon \cos^2\xi + \cos^2\epsilon \sin^2\xi \sin^2 2\delta - \frac{1}{2} \sin2\epsilon \sin2\xi \sin2\delta \quad (14)$$

and

$$T_\perp = M_{12}M_{12}^{\#} = \sin^2\eta \cos^2\epsilon \cos^2 2\delta + \sin^2\epsilon \sin^2\xi + \cos^2\epsilon \cos^2\xi \sin^2 2\delta - \frac{1}{2} \sin2\epsilon \sin2\delta \sin2\xi. \quad (15)$$

Before a further expansion is made, we must determine which terms are small. If waveplates whose birefringent axes and tilts are not wavelength-dependent are used, the terms in δ can be made small. For the polarizers used in normal Lyot filters, the crossed transmission is between 10^{-3} and 10^{-4}. For the remainder of this paper, tolerable errors will be on the order of the cross polarizer leakage. In order to make the terms in δ that small, 0.344 min of arc $< \delta <$ 3.44 min of arc. This limit is placed by the linear terms. In fact, at least ϵ and probably ξ will be small, and δ can be at least ten times the minimum value. The effect of δ will be negligible if δ is less than 2 min of arc.

If we neglect the effect of δ,

$$T_\parallel = T^1 \cos^2\eta + \sin^2\epsilon (1 - \sin^2\xi), \quad (16)$$

$$T_\perp = T^1 \sin^2\eta + \sin^2\epsilon \sin^2\xi, \quad (17)$$

where

$$T^1 = (1 - \sin^2\epsilon). \quad (18)$$

If ξ is close to a multiple of $\pi/2$, there will be a clear advantage to Eq. (16) if ξ is nearly an odd multiple or Eq. (17) if ξ is an even multiple. For quartz elements it is quite straight forward to make ξ nearly zero, since the ratio of the index difference to the index is 5.8×10^{-3} at 6000 Å. Hence if the two quartz elements are equal to within 1.72 waves (1 μ) the birefringence difference will be $\frac{1}{100}$ wave. That is, $\xi = 1.8°$. Hence, for well-matched (<1 μ difference) quartz elements it is much wiser to use crossed polarizers on the wide field element.

For calcite the ratio of the index difference to the mean index is 0.109 at 6000 Å. To get $\frac{1}{100}$ wave birefringence difference requires that the individual thicknesses be the same to within $\frac{1}{10}$ wave. This is not impossible, since the two element halves can be crossed and polished for a black center fringe in white light. But even if the calcite element halves are only the same to a wave, the errors introduced by the half-waveplate are down by a factor of 9 with crossed polarizers. Another possibility is to make ξ nearly an odd or even multiple of $\pi/2$ at the wavelength of the filter and then to choose the appropriate polarizer configuration.

To demonstrate the symmetry between δ and ϵ of the error situation, suppose that ϵ and not δ can be neglected. Then from Eqs. (14) and (15)

$$T_\parallel = T^1 \cos^2\eta - \sin^2\xi \sin^2 2\delta. \quad (19)$$

and

$$T_\perp = T^1 \sin^2\eta + \sin^2 2\delta (1 - \sin^2\xi), \quad (20)$$

where

$$T^1 = (1 - \cos^2 2\delta). \quad (21)$$

So by the same arguments used above, the crossed polarizer mode is again the most insensitive to tilt errors when the thickness error is minimal.

When the element is used in the crossed polarizer mode [Eq. (15)], the dominant error term will be $\sin^2\epsilon \sin^2\xi$ if

$$\delta << \epsilon < \xi. \quad (22)$$

It is desired that

$$\sin^2\epsilon \sin^2\xi \leq 10^{-4}.$$

If ξ is a $1/10$ wave or less,

$$\sin^2\epsilon \leq 10^{-3},$$
$$\epsilon \leq 1.8°. \quad (23)$$

This implies that an element could be tuned thermally (δn is a function of temperature) over a range of ± 34 Å near 6000 Å.

In Paper 2 it was shown that a nine-element half-waveplate could be made achromatic to within a degree from 3500 Å to 10,000 Å. However, the tilt axis varied by as much as several degrees at the ends of the range. It is therefore necessary to have mechanical adjustment for tilt when achromatic multielement half-waveplates are used. If mechanical adjusters are not used, there is no advantage to greater than three-element waveplates.

Tunable Elements

In the standard tuning configuration, a quarter-waveplate is placed behind the wide field element and a polarizer rotated for tuning. If the matrix for the ordinary wide field element is denoted by M, the tuned element is

$$Q = PR(\theta)R(-l)G(\rho)R(l)M, \quad (24)$$

where

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad (25)$$

and

$$\rho = (\pi/4) + q, \quad (26)$$

where θ is the angle of the polarizer, q is half the retardation error and l is the tilt error of the quarter-waveplate. In Eq. (24) $R(-\theta)$ is omitted and when this is done, Q_{11} is the proper transmission amplitude.

In order to understand the basic operation of the quarter-waveplate, suppose for the moment q and l are zero and the tuned element is fed by linearly polarized light. Then it can be shown that the transmission amplitude is

$$Q_{\parallel} = (1 + i)/\sqrt{2}[\cos\epsilon \cos 2\delta \sin(\xi - \theta) - \sin\epsilon \cos(\xi - \theta) + i \cos\epsilon \cos 2\delta \cos(\eta + \theta)]. \quad (27)$$

The transmission is

$$T = Q_{\parallel} Q_{\parallel}^{\#},$$

or

$$T = \cos^2(\eta - \theta) \cos^2\epsilon \cos^2 2\delta + \cos^2\epsilon \sin^2 2\delta \sin^2(\xi + \theta) + \sin^2\epsilon \cos^2(\xi + \theta) - \frac{1}{2}\sin 2\epsilon \sin 2\delta \sin 2(\xi + \theta). \quad (28)$$

Comparison of Eq. (28) with Eqs. (14) and (15) shows that varying θ tunes the element and that as θ goes from zero to π the dominant error producer changes from δ to ϵ. The error term in ϵ dominates when $(\xi + \theta)$ is an even multiple of $\pi/2$. By making $\delta \leq 2'$ the dominant error will come from ϵ, so that if $\epsilon \leq 1.8°$ the error will always be below 10^{-4}. If achromatic waveplates are used, the angular variation of the tilt must be adjusted out if the tilt errors are not to dominate.

If δ is made small, as the exit polarizer is rotated the transmission at a minima caused by the inaccuracy in the half-waveplate will modulate between zero and $\sin^2\epsilon$. The minima will occur at

$$(\xi + \theta) = n(\pi/2), \quad n \text{ odd}, \quad (29)$$

or

$$\theta = n(\pi/2) - \xi. \quad (30)$$

By measuring the angle modulo $(\pi/2)$ at which the minima and/or maxima occur, the thickness difference error ξ can be easily measured to better than $1/100$ wave. This measurement can be used in a production technique for precisely matching element halves.

The ability to make accurate element halves is useful if the elements are temperature-tuned or the filter works at a fixed wavelength. For tunable wide-field elements, exactly matched halves is of marginal utility, since the thickness error only changes the angle at which the maximum half-waveplate error occurs.

In order to understand the errors introduced when q and l are nonzero, it is convenient to write

$$Q = \frac{1}{\sqrt{2}}\begin{pmatrix} \cos\theta & \sin\theta \\ 0 & 0 \end{pmatrix}\left[\cos q \begin{pmatrix} 1 + i & 0 \\ 0 & 1 - i \end{pmatrix} + \sin q \begin{pmatrix} 1 - i & 0 \\ 0 & -1 - i \end{pmatrix} + 2i \sin l(\sin q + \cos q)\begin{pmatrix} -\sin l & \cos l \\ \cos l & \sin l \end{pmatrix}\right]\begin{pmatrix} \cos\eta & -i\sin\eta \\ +i\sin\eta & -\cos\eta \end{pmatrix}. \quad (31)$$

Here the error terms of the wide field element have been neglected because they are already second-order.

The transmission amplitude is

$$\sqrt{2}Q_{11} = (1 + i)[\cos q \cos(\eta - \theta) - i \sin q \cos(\eta + \theta)] + 2i \sin l(\sin q + \cos q)[\cos\eta \sin(\theta - l) - i \sin\eta \cos(\theta - l)]. \quad (32)$$

The transmission is the complex square of Q_{11}. Unfortunately, the complex algebra for the transmission is rather messy. To clarify the algebra, note that

$$\sqrt{2}Q_{11} = (1 + i)(a - ie) + ib(c + id). \quad (33)$$

So

$$2Q_{\parallel}Q_{\parallel}^{\#} = 2(a^2 + e^2) + b^2(c^2 + d^2) + 2ab(c - d) + 2be(c + d), \quad (34)$$

where

$$a = \cos q \cos(\eta - \theta), \quad (35)$$

$$e = -\sin q \cos(\eta + \theta), \quad (36)$$

$$b = 2 \sin l(\sin q + \cos q), \quad (37)$$

$$c = \cos\eta \sin(\theta - l), \quad (38)$$

and

$$d = \sin\eta \cos(\theta - l). \quad (39)$$

The transmission can be put in the form

$$T = a^2 + e^2 + b(c - d)[a + (b/2)(c + d)] + be(c + d). \quad (40)$$

From Eqs. (35) through (40) the effect of a pure error in retardation (no tilt) is

$$T = a^2 + e^2 = \cos^2 q \cos^2(\eta - \theta) + \sin^2 q \cos^2(\eta + \theta). \quad (41)$$

Writing the cosine of q in terms of the sine of q and expressing

$$(\eta + \theta) = (\eta - \theta + 2\theta), \quad (42)$$

expression (41) becomes

$$T = \cos^2(\eta - \theta) + \sin^2 q[-\cos^2 2(\eta - \theta) \sin^2 2\theta + \sin^2(\eta - \theta) \sin^2 2\theta - \tfrac{1}{2}\sin 2(\eta - \theta) \sin 4\theta], \quad (43)$$

or

$$T = T^1 \cos^2(\eta - \theta) + \sin^2 q[\sin^2 2\theta - \tfrac{1}{2}\sin 2(\eta - \theta) \sin 4\theta], \quad (44)$$

where

$$T^1 = (1 - 2 \sin^2 q \sin^2 2\theta). \quad (45)$$

Result (44) demonstrates that transmission asymmetry can be introduced by an imperfect quarter-waveplate. The term in sine $2(\eta$-$\theta)$ is zero when cosine $(\eta$-$\theta)$ is zero. Therefore, the term does not add transmission at the minima, but rather transmission asymmetry. The term in sine$^2 2\theta$ does introduce additional transmission at the minima.

If the retardation error is zero but a tilt of the quarter-waveplate exists

$$T = a^2 + b(c - d)[a + (b/2)(c + d)]. \quad (46)$$

After some algebra and expansion in terms of $(\eta - \theta)$ to second order,

$$T = T^1 \cos^2(\eta - \theta) - \sin l \sin 2(\eta - \theta) - \sin^2 l[2 \cos 2\theta + \sin 2\theta \sin 2(\eta - \theta)], \quad (47)$$

where

$$T^1 = (1 - 4 \sin^2 l \sin^2 \theta). \quad (48)$$

Note that tilt of the quarter-waveplate causes a first-order error. Fortunately, the error causes an asymmetry of the transmission profile and does not create additional transmission at the minima. The second-order terms in the tilt cause both asymmetry and transmission at the minima. The term in sine$2(\eta - \theta)$ only introduces asymmetry, while the term in cosine2θ causes transmission at minima. The quarter-waveplate should therefore be adjusted at least as well as the half-waveplate, or about 2 min of arc.

When both nonzero tilt and retardation are considered, additional second-order terms occur from the expressions $ba(c - d)$ and $be(c + d)$. The additional tilt and retardation terms are to second order

$$-2 \cos^2(\eta - \theta) \sin l \sin q - \sin l \sin q[2 \cos^2 2\theta \sin \times 2(\eta - \theta) - \sin 4\theta] \quad (49)$$

Here again both asymmetry and transmission at minima occur. The error term in sine2 $(\eta - \theta)$ causes pure asymmetry, and the cosine2 $(\eta - \theta)$ term allows additional transmission at minima. Finally then, the transmission effects of the quarter-wave errors are

$$T = T^1 \cos(\eta - \theta) - \tfrac{1}{2}\sin 2l \sin 2(\eta - \theta) \times \sin^2 l[2 \cos 2\theta + \sin 2\theta \sin 2(\eta - \theta)] + \sin^2 q[\sin^2 2\theta - \tfrac{1}{2}\sin 4\theta \sin 2(\eta - \theta)] - \sin l \sin q[2 \cos^2 2\theta \sin 2(\eta - \theta) - \sin 4\theta], \quad (50)$$

where

$$T' = (1 - 2 \sin^2 q \sin^2 2\theta - 4 \sin^2 l \sin^2 \theta - 2 \sin l \sin q \sin 4\theta). \quad (51)$$

Using a single quarter-waveplate to tune requires that the exit polarizer rotate with respect to the element. But since the exit polarizer for one element is the entrance polarizer for the following element, the entire following element must rotate in order to maintain an angle of 45° between the entrance polarizer and the fast axis of that element. In order to obviate rotating entire elements, a quarter-waveplate is often attached at 45° to the entrance polarizer on the side toward the first section of an element. In this case circularly polarized light enters the element, so the properties of the element are independent of orientation of the entrance polarizer. The exit polarizer still allows tuning because the linear polarizer faces the last portion of the second quarter-waveplate.

An analysis of the errors that can be introduced by a quarter-waveplate on the entrance polarizer proceeds in the same manner as for the quarter plate on the element. The same order of errors is introduced by the second quarter-waveplate. The analysis will not be carried out here. Instead an alternate method to tune elements independently is suggested. Instead of rotating the final polarizer, it is equivalent (except for a plate factor) to rotate a half-waveplate between the quarter-waveplate and the exit polarizer. The half-waveplate need only rotate by half the angle at which the polarizer must rotate. In matrix notation,

$$\mathrm{PR}(\theta) = -i\mathrm{PR}[-(\theta/2)]G(\pi/2)R(\theta/2). \quad (52)$$

The left-hand side of Eq. (52) represents a rotated polarizer and the right-hand side represents a half-waveplate at an angle $\theta/2$ in front of a fixed polarizer.

he $-i$ in Eq. (52) removes the arbitrary plate factor. Use of the half-waveplate does not introduce any ldditional first-order errors. Another advantage of ıe half-waveplate occurs if prism polarizers are sed, because rotating the half-waveplate eliminates ıe necessity of rotating the mass of the prism and mplifies the problem of eliminating the rejected ght from the prism.

scussion

The analysis of a single wide field element shows ıat the nine- and ten-element achromatic half- and ıarter-waveplates described in paper 2, it is possible construct a tunable filter over the range 3500–),000 Å if appropriate correction of tilt is accomished. When tuning is accomplished by rotation of alf-waveplates, all the normally nonrotating eleents could be coupled to a single shaft, which could e used to adjust the angles of all the fixed wavepıtes. The rotation variation of the quarter-waves nd the axis location of the tuning half-wave eleents can be corrected for by the programming sysm used for tuning the filter.

Besides mechanical tuning of filter elements, electrical tuning by electrooptical crystals is also possible. If a wide field element that has a nine-element achromatic half-waveplate is tuned by a pair of electrooptical crystals, complete tuning throughout the visible would be possible without any mechanical adjustments. This would occur because the entrance and exit polarizers could be placed in a configuration that minimizes the effect of rotation of the retardation axis of the half-waveplate. Good electrooptical crystals are now available. Their use in Lyot filters will be discussed in a subsequent article in this series.

This work was supported by Lockheed Missiles and Space Corporation independent research funds.

The author is deeply indebted to H. E. Ramsey for construction of a large family of wide field elements with known errors. These provided experimental verification of the calculations presented here.

References

1. J. Jefferies and R. Giovanelli, Aust. J. Phys. 1, 254 (1954).
2. A. Title, Appl. Opt. 14, 00 (Jan. 1975).
3. R. C. Jones, J. Opt. Soc. Am. 31, 488 (1941).

Reprinted from APPLIED OPTICS, Vol. 15, page 2871, November 1976

Improvement in birefringent filters. 4: The alternate partial polarizer filter

Alan M. Title

A design for a birefringent filter is proposed in which alternate polarizers are partial polarizers. Calculated performance characteristics of alternate partial polarizer filters (APP) are compared with those of Lyot and contrast element Lyot filters. These calculations show that the APP design has significant advantages in both transmission and profile shape. Using pulse techniques, partial polarizer systems are shown to be a natural evolution from the standard Lyot and contrast element Lyot systems. The APP filter using achromatic waveplates discussed in earlier papers of this series has been used to construct a universal alternate partial polarizer filter. This filter has a measured full width at half-maximum (FWHM) of 0.09 Å at 5500 Å and a transmission in polarized light of 38%. It is tunable from 4500 Å to 8500 Å. The measured characteristics of the filter agree well with theoretical predictions.

Introduction

In previous papers of this series,[1-3] components of birefringent filters have been discussed. Here a new design is suggested that utilizes partial polarizers. The characteristics of birefringent filters with intermediate partial polarizers have not had much attention, although the cases of filters with perfect or no intermediate polarizers have been studied extensively.[4-10] The standard birefringent filter, the Lyot-Ohman, has perfect intermediate polarizers. The Sölc filter, which has seen significant but much less use, has no intermediate polarizers. In this paper the properties of several filter configurations with imperfect intermediate polarizers shall be investigated. It will be shown that such filters have significant advantages in transmission and profile shape to both the standard Lyot and the contrast element Lyot filters. These theoretical advantages are demonstrated by the measured properties of an actual filter that uses partial polarizers.

It has been analytically shown by Giovanelli and Jefferies,[11] and verified using a computer technique by Beckers and Dunn,[12] that interior imperfect polarizers do not have deleterious effects on Lyot filter performance if the ratio of polarized to unpolarized light is greater than $(10N^2 - 1):1$, where N is the number of intermediate polarizers. The physical reason for this effect has been explained by the author.[13] For the Lyot case, it was shown that the partial polarizers between the longest and next longest element are helpful while all other partial polarizers have a deleterious effect.

The Lyot and contrast element Lyot filters are discussed from the point of view of their pulse response. This is done to provide a basis of comparison with the partial polarizer systems. Partial polarizer filters are then shown to be a natural evolution from the previous designs. Finally, the measured performance of an alternate partial polarizer filter is shown to agree with theoretical prediction.

Lyot Systems

A Lyot filter is formed by two or more modules, each of which consists of an entrance polarizer, a birefringent crystal, and an exit polarizer. When a pulse of light is incident on the first polarizer, pulses are propagated down the fast and slow axis of the crystal with amplitudes

$$A_f = A_i \cos\theta,$$

$$A_s = A_i \sin\theta,$$

where A_i, A_f, and A_s are the amplitudes of the incident, fast axis, and slow axis pulses, and theta is the angle the entrance polarizer makes with the fast axis of the crystal. Exiting from the final polarizer are two pulses with amplitudes

$$A_{of} = A_i \cos\theta \cos\theta',$$

$$A_{os} = A_i \sin\theta \sin\theta',$$

The author is with Lockheed Palo Alto Research Laboratory, Electro-Optics Laboratory, Palo Alto, California 94304.
Received 27 March 1976.

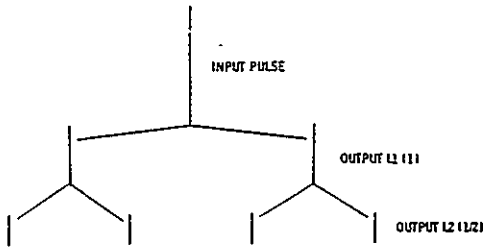


Fig. 1. Pulse diagram of two Lyot modules with length ratios of 1: 1/2.

where theta prime is the angle the exit polarizer makes with the fast axis of the crystal. Both outgoing pulses have the same polarization state. In order that the two pulses have equal amplitudes, theta and theta prime must be equal in magnitude, to 45°. The time difference between the fast and slow pulses is

$$t = \mu d/c, \tag{1}$$

where μ is the difference of the indices of refraction of the crystal, d is the crystal length, and c is the speed of light.

Lyot modules have equal output pulses, and hence they have parallel or perpendicular polarizers, and the crystal fast axis is at ±45 deg to polarizer axis.

For convenience, the origin of time is usually redefined so that one pulse is delayed by half of the time difference while the other is advanced by the same amount. The pulse response and frequency response are Fourier transform pairs, so the transmission amplitude of a Lyot module that gives rise to a pair of pulses separated by the time given in Eq. (1) is

$$A(\nu) = \cos(2\pi\mu d\nu). \tag{2}$$

The simplest Lyot filter consists of two modules. If the crystals are not of equal length, for every input pulse to the first module four pulses emerge from the second. The situation is shown in Fig. 1 for the case in which the crystal lengths are 1:1/2. The 1:1/2 length ratio has the effect of making the output pulses equally spaced in time.

The number of modules in a Lyot filter can be increased indefinitely and equal pulse spacing maintained, if the length ratios of adjoining elements are 1:1/2, that is, equal pulse spacing occurs when the crystal lengths are in the ratio $1:1/2:1/4:\ldots:1/2^{N-1}$.

The output pulse structure for a single input pulse into an N element filter is

$$P(t) = R_L(t) \text{III}\left(\frac{t}{t_N}\right), \tag{3}$$

where

$$R_L(t) = 1 |t| \leq \frac{L}{2}$$

$$= 0 |t| > \frac{L}{2} \tag{4}$$

and

$$L = \sum_{k=1}^{N} \frac{d_1}{2^{k-1}} \approx 2d_1, \tag{5}$$

$$t_N = (\mu d_1)/(c2^{N-1}), \tag{6}$$

where d_1 is the length of the longest crystal, that is, the pulse response can be considered as the product of an infinite series of pulses with a time separation equal to the time delay of the shortest crystal times a function that is unity for times less than twice the time delay of the longest crystal and zero at all other times. The pulse response is written in the form given by Eq. (3) because the Fourier transform of Eq. (3) is just a convolution of the sinc function and a dirac comb. The amplitude frequency response, the Fourier transform of the pulse response, is

$$A(\nu) = \text{sinc}(L\nu) \text{III}\left(\frac{\nu}{\nu_N}\right). \tag{7}$$

From Eq. (7) the Lyot filter has a series of transmission maximum, each of which is separated in frequency by ν_N. [The separation of successive transmission maxima is called the free spectral range (FSR).] Near each transmission maximum, the transmission intensity has the form of sinc squared.

Because of the substantial secondary maxima of sinc squared, it is desirable to modify the standard Lyot profile. In terms of the pulse response, this requires introduction of a tapering or apodizing function. The adjustment of the pulse amplitudes can be done in two ways: by adjusting the angles of the entrance and exit polarizers with respect to the crystals which adjusts the relative amplitude of the fast and slow pulses and/or by adjusting the crystal lengths such that pulse overlap occurs.

Normally, the technique used to reduce the transmission side lobes of Lyot filters is to add another module with a crystal whose length is equal to the second longest crystal in the filter.[14,15] Figure 2 shows the pulse response tree of a series of modules with length ratios 1:1/2:1/2:1/4:1/8.

From the figure there are only five pulses emerging from the contrast element instead of the eight that would normally occur from the third module of a normal

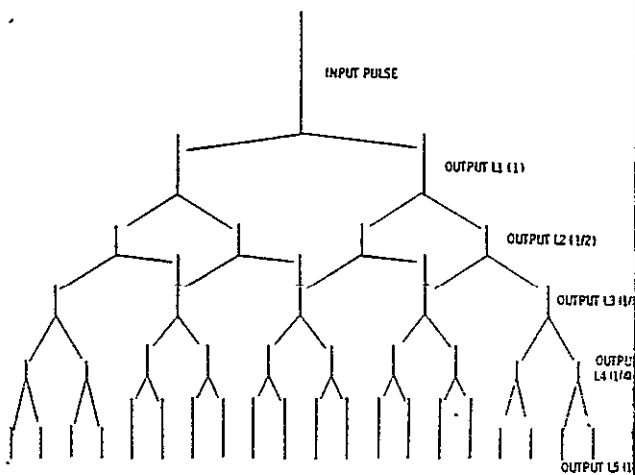


Fig. 2. Pulse diagram for a contrast element Lyot module. Note that the second and third Lyot modules have equal lengths, and that pulse overlap occurs at the output of $L3$.

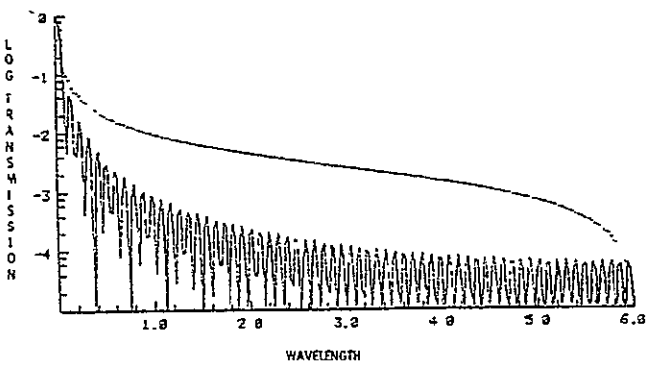


Fig. 3. Transmission vs wavelength (solid) and contribution function (dotted) from a transmission peak for a pure Lyot filter. The graph is plotted to half of the FSR.

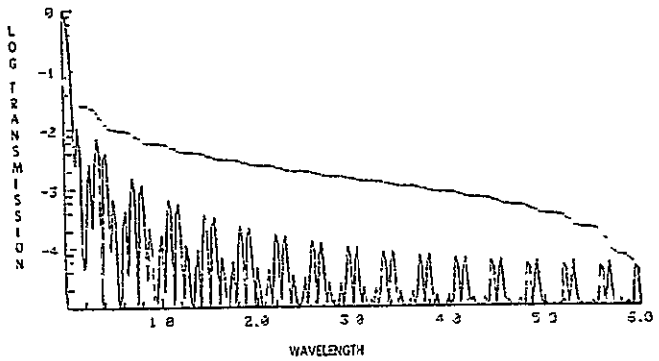


Fig. 4. Transmission (solid) and contribution function (dotted) vs wavelength for a contrast element Lyot.

Lyot filter. Because pulses that arrive at the same time are indistinguishable, the effect of more than one pulse arriving at a given time is to create pulses whose amplitudes are the sum of pulses that arrive at that time. Here the time overlapping pulses have the same amplitude: thus center three pulses from contrast element have twice the amplitude of the outer two.

After the contrast element each successive module doubles the number of output pulses and preserves the amplitude distribution. As can be seen from Fig. 2, the contrast element tapers the pulse response.

The analytic form of the contrast element pulse response is

$$P(t) = [R_{L_1}(t) + R_{L_2}(t)]\amalg\left(\frac{t}{t_N}\right), \tag{8}$$

where

$$L_2 = \frac{3}{5}L_1 \tag{9}$$

and

$$L_1 = \sum_{k=1}^{N} \frac{d_1}{2^{k-1}} + \frac{d_1}{2} = \frac{5}{2}d_1. \tag{1}$$

The amplitude frequency response is then

$$A(\nu) = [\mathrm{sinc}(L_1\nu) + \mathrm{sinc}L_2(\nu)]\amalg\left(\frac{\nu}{\nu_N}\right). \tag{11}$$

The transmission vs wavelength for Lyot and a contrast element Lyot are shown in Figs. 3 and 4. The dotted curve in the figures is the contribution function, the ratio of the light outside the wavelength to the total light in half of the free spectral range (FSR), that is,

$$c(\lambda) = \int_{\lambda}^{\mathrm{FSR}/2} T(\lambda)d\lambda \Big/ \int_{0}^{\mathrm{FSR}/2} T(\lambda)d\lambda. \tag{12}$$

The main design points of a Lyot system are (1) the ±45° angle the fast axis of the crystals make with respect to the parallel entrance and exit polarizers, which is to obtain equal amplitude pulses from a single module, and (2) the 1:1/2 length ratios of adjoining crystals, which is to obtain equally spaced equal amplitude pulses.

Alternate Partial Polarizer Systems

The basic Lyot module produces two output pulses for each input pulse. Hence, the first logical extension is to examine properties of filters built from modules that have three or four output pulses for each input pulse. Such modules must have at least two crystals with an intermediate element and/or a relative orientation of the two crystals. Here only the case of a partial polarizer intermediate element shall be discussed. The designs that have either an intermediate perfect polarizer or no polarizer have already been handled by the general techniques in the literature.

A single partial polarizer module consists of an entrance polarizer, a crystal, a partial polarizer, a crystal, and an exit polarizer. All the polarizers are parallel, while the crystals have their fast axis at plus and minus 45° from the polarizer axis, that is, the crystals' fast axis are orthogonal. The reason for the orthogonality of the crystals will be made clear below. The configuration is shown in Fig. 5. Because a sequence of modules would form a filter with alternate partial polarizers, the basic partial polarizer module shall be called an alternate partial polarizer (APP) module.

Unless the two crystals of the APP have equal lengths, there will be four output pulses for every input pulse. The relative amplitude of the pulses is easy to estimate because the pulse response of the partial polarizer is just the pulse response of a perfect polarizer plus a neutral density filter. The perfect polarizer gives rise to four equal intensity pulses, since it makes the configuration just a pair of Lyot modules. The relative orientation of the crystals has no effect for Lyot modules. The neutral density filter, however, allows the orthogonal crystals to subtract birefringence which has

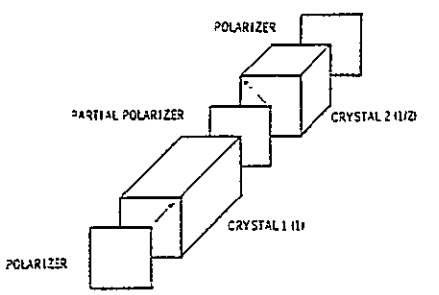


Fig. 5. Optical schematic of an alternate partial polarizer module.

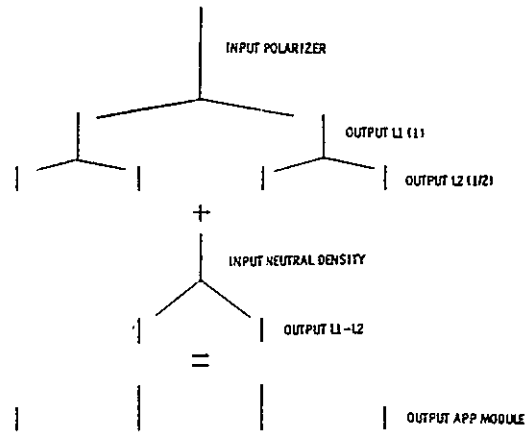


Fig. 6. Pulse diagram of an APP module illustrating the perfect polarizer and neutral density concept.

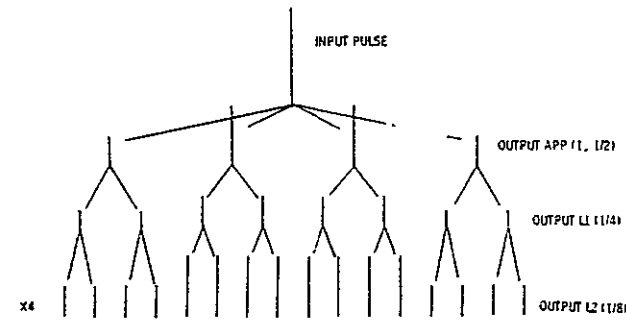


Fig. 7 Pulse diagram for an APP module followed by two Lyot modules with length ratios of 1:1/2:1/4:1/8.

the result that the system behaves like a single Lyot module whose length is equal to the difference of the two crystal lengths. Relative to the perfect polarizer pulses, the neutral density pulses are decreased in amplitude proportional to the neutral density factor. In the case where the two crystals have 1:1/2 length ratios, the neutral density pulses coincide in time with the inner Lyot pulses. The situation is illustrated in Fig. 6.

The argument above can be carried out quantitatively. The Jones matrix for a partial polarizer is

$$\mathrm{PP} = \begin{pmatrix} \rho_x & 0 \\ 0 & \rho_y \end{pmatrix}, \tag{13}$$

where

$$T_x = \rho_x^2,$$
$$T_y = \rho_y^2, \tag{14}$$

T_x and T_y are the transmissions of the partial polarizer in the high and low transmission directions. The matrix can be rewritten as

$$\mathrm{PP} = (\rho_x - \rho_y)\left[\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \rho_y/(\rho_x - \rho_y)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right]. \tag{15}$$

The terms in brackets are the matrices for perfect polarizer and a neutral density filter. The transmission amplitude then can be written as

$$A(\nu) = [\mathrm{PM}(\alpha_1,45)\mathrm{PPM}(\alpha_2,-45)\mathrm{P}]_{11}, \tag{16}$$

or using Eq. (15),

$$A(\nu) = (\rho_x - \rho_y)[\mathrm{PM}(\alpha_1,45)\mathrm{PM}(\alpha_2,-45)\mathrm{P} + \epsilon\mathrm{PM}(\alpha_1,45)\mathrm{M}(\alpha_2,-45)\mathrm{P}]_{11}, \tag{17}$$

where

$$\epsilon = \rho_y/(\rho_x - \rho_y), \tag{18}$$
$$\mathrm{M}(\alpha,\pm 45) = \cos\alpha\,\mathrm{I} + \imath\,\sin\alpha\,\mathrm{ER}(\pm 90),$$
$$\alpha_i = 2\pi\mu d_i\nu,$$
$$\mathrm{E} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

R is the rotation matrix, and I the unit matrix, so that

$$A(\nu) = (\rho_x - \rho_y)[\cos\alpha_1\cos\alpha_2 + \epsilon\cos(\alpha_1 - \alpha_2)], \tag{19}$$

which is the expected result. Expanding the cosine product term, Eq. (19) becomes

$$A(\nu) = (\rho_x - \rho_y)[\cos(\alpha_1 + \alpha_2) + (1 + 2\epsilon)\cos(\alpha_1 - \alpha_2)]. \tag{20}$$

Equation (20) demonstrates that the response of the single partial polarizer module is also identical to the sum of the responses of a Lyot module of length d_1 plus d_2 and one of d_1 minus d_2 and that the amplitude of the inner pulses are greater by a factor of $(1 + 2\epsilon)$ than the outer pulses. From Eq. (20) equal pulse spacing occurs when the crystal length ratios are 1:1/2. The relative pulse amplitude time sequence is $1, 1 + 2\epsilon, 1 + 2\epsilon, 1$.

Since the pulse output of the APP is tapered, a tapered pulse sequence quite similar to that of a contrast element Lyot can be achieved by using Lyot modules in conjunction with an APP module. The pulse response of a filter consisting of an APP module and two Lyot modules in which the crystal length ratios are 1:1/2:1/4:1/8 is shown in Fig. 7. For the figure, the partial polarizer factor epsilon is ½. A comparison of Figs. 7 and 2 shows that the pulse response, and hence the amplitude response, of a Lyot filter plus an APP module are very similar to a contrast element Lyot filter. The transmission vs length near a transmission peak of an APP plus Lyot is shown in Fig. 8, which should be compared to the transmission of the contrast Lyot shown in Fig. 4.

Note that an APP plus Lyot can be made from a pure

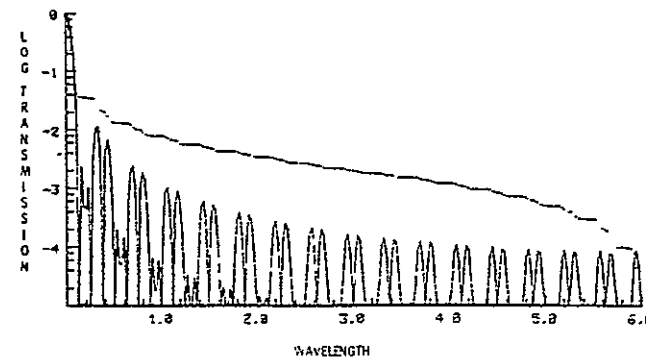


Fig. 8 Transmission (solid) and contribution function (dotted) vs wavelength for APP-Lyot system.

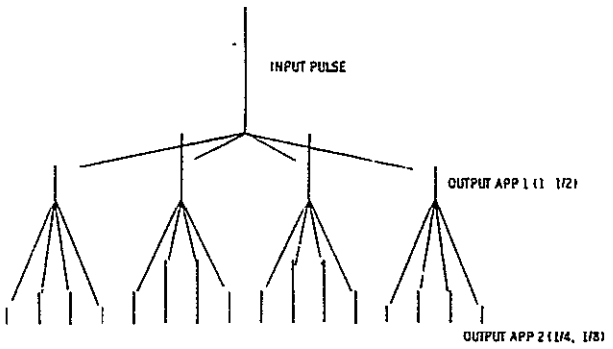


Fig. 9. Pulse diagram for a pair of APP modules with Lyot filter length ratios.

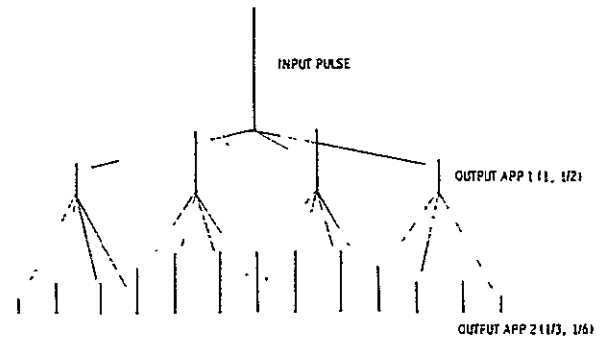


Fig. 10. Pulse diagram for a pair of APP modules with length ratios 1:1/2:1/3:1/6.

Lyot by removing the perfect polarizer from between the longest and next longest crystal and replacing it by a partial polarizer and rotating the first two crystals so that their fast axis are orthogonal. It is tempting to imagine that the pulse response of the Lyot could be tapered even more advantageously by replacing the polarizers between the other crystals with a partial polarizer. The pulse sequence for the case of a pair of APP modules with length ratios of 1:1/2:1/4:1/8 is shown in Fig. 9.

The relative amplitudes of the time sequence of sixteen pulses are 1, $(1 + 2\epsilon_2)$, $(1 + 2\epsilon_2)$, 1, $(1 + 2\epsilon_1)$, $(1 + 2\epsilon_1)$ $(1 + 2\epsilon_2)$, $(1 + 2\epsilon_1)$ $(1 + 2\epsilon_2)$, $(1 + 2\epsilon_1)$, . . ., where ϵ_1 is the partial polarizer factor of the first module, and ϵ_2 is the factor of the second. Unless ϵ_2 is zero, the pulse amplitude sequence amplitude oscillates in magnitude, which gives rise to objectionable secondary maxima in the filter's transmission.

A monotonic taper can be easily restored, however, by increasing the lengths of the crystals in the second module relative to the first. Figure 10 illustrates the case where the crystal lengths have been adjusted so that output pulses from the second APP modules due to successive input pulses just overlap.

In the situation illustrated by Fig. 10, there are three net output pulses from the second APP module of the filter for every input pulse instead of four. Thus, the interpulse space is 4/3 the interpulse space of the Lyot case. Since in each module the crystal lengths are 1:1/2, the ratio of the lengths of crystals that adjoin a perfect polarizer must be 1:2/3. Therefore, the ratios of the crystal lengths are 1:1/2:1/3:1/6. The relative pulse amplitude sequence is 1, $(1 + 2\epsilon_2)$, $(1 + 2\epsilon_2)$, $(2 + 2\epsilon_2)$, $(1 + 2\epsilon_1)$ $(1 + 2\epsilon_2)$, $(1 + 2\epsilon_1)$ $(1 + 2\epsilon_2)$, 2, $(1 + 2\epsilon_1)$, ...

In order that the pulse amplitude sequence does not oscillate, it is required that $\epsilon_2 = \frac{1}{2}$. Epsilon one-half causes the overlapped pulse to have an amplitude that is the mean of the pulses immediately preceding and following the overlapped pulse. An APP filter can be made of an arbitrary number of APP modules when crystals on either side of a perfect polarizer are in the ratio of 1:2/3, and all partial polarizers after the first have epsilon equal to one-half. As more modules are added, the taper of the pulse sequence becomes smoother. The transmission amplitude of the APP is

$$A(\nu) = \left(\frac{2}{3}\right)^{N/2} \prod_{k=1}^{N/2} \left[\cos\left(\frac{\pi\mu d_1\nu}{3^{k-2}}\right) + 2\cos\left(\frac{\pi\mu d_1\nu}{3^{k-1}}\right)\right]^2, \quad (21)$$

and the crystal lengths are

$$d_{\text{odd}k} = \left(\frac{d_1}{2}\right) \Big/ 3^{\frac{k-1}{2}}$$

and

$$d_{\text{even}k} = \left(\frac{3}{2}d_1\right) \Big/ 3^{k/2}, \quad (22)$$

where d_{odd} and d_{even} are the lengths of the odd and even numbered crystals. The transmission near a peak of a four-module (eight-crystal) APP filter is shown in Fig. 11.

The desire for equally spaced pulses has caused all

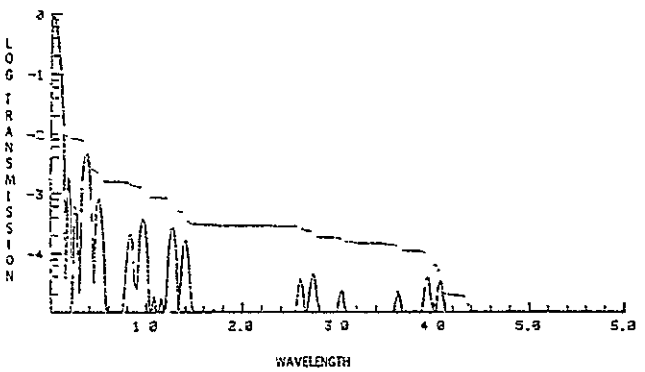


Fig. 11. Transmission (solid) and contribution function (dotted) for a four module APP filter. Half of the FSR occurs at 6 wavelength units.

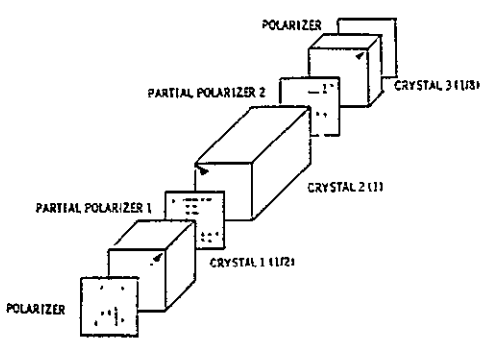


Fig. 12. Optical schematic of double partial polarizer module.

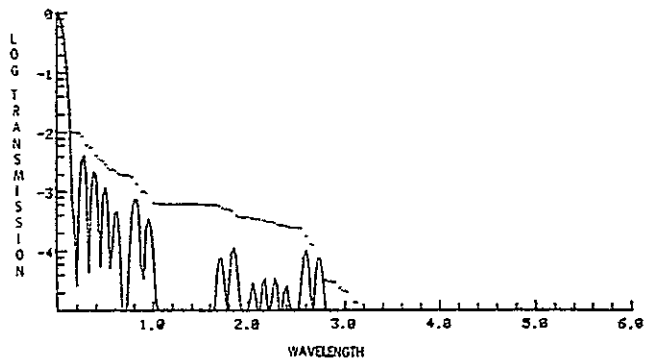


Fig. 13. Transmission (solid) and contribution function (dotted) for a DPP plus two APP module filters. The half FSR occurs at 6 wavelength units.

previous analysis to be limited to birefringent networks in which all crystals are multiples of a constant length; however, as seen above, equally spaced pulses can be achieved without common multiple length crystals.

Three Crystal Modules

Having obtained a useful configuration with a single partial polarizer, it is worthwhile to investigate how many partial polarizers can be in a single module. Using pulse sequence diagrams, it appears that only the system of three crystals and two partial polarizers provide useful pulse tapers. The three crystal system consists of entrance polarizer, crystal, partial polarizer, crystal, partial polarizer, crystal, and exit polarizer. All polarizers are parallel while the fast axis of the outer crystals is at 45° and the inner at minus 45°. The configuration is shown in Fig. 12. The module shall be called a double partial polarizer (DPP).

The transmission amplitude for the module is from the Jones matrix calculus:

$$A(\nu) = [(\rho_x - \rho_y)(\rho_x' - \rho_y') \\ [PM(\alpha_1,45)\mathrm{PPM}(\rho_2,-45)\mathrm{PPM}(\alpha_3,45)\mathrm{P}]_{11}, \quad (23)$$

where

$$\alpha_i = 2\pi\mu d_i\nu \quad (24)$$

Writing the partial polarizer matrices as the sum of perfect polarizer and neutral density filter and expanding, Eq. (23) becomes

$$A(\nu) = c[\mathrm{PM}(\alpha_1,45)\mathrm{PM}(\alpha_2,-45)\mathrm{P}M(\alpha_3,45)\mathrm{P} \\ + \epsilon\mathrm{PM}(\alpha_1,45)\mathrm{M}(\alpha_2,-45)\mathrm{P}M(\alpha_3,45)\mathrm{P} \\ + \epsilon'PM(\alpha_1,45)\mathrm{PM}(\alpha_2,-45)\mathrm{M}(\alpha_3,45)\mathrm{P} \\ + \epsilon\epsilon'\mathrm{PM}(\alpha_1,45)\mathrm{M}(\alpha_2,-45)\mathrm{M})\alpha_3,45)\mathrm{P}]_{11}, \quad (25)$$

where

$$c = (\rho_x - \rho_y)(\rho_x' - \rho_y'),$$
$$\epsilon = \rho_y/\rho_y - \rho_x,$$
$$\epsilon' = \rho_y'/\rho_y' - \rho_x'.$$

The expression in brackets represents just a sum of different numbers of Lyot modules. Hence, by inspection, Eq. (25) is

$$A(\nu) = c[\cos\alpha_1\cos\alpha_2\cos\alpha_3 + \epsilon\cos(\alpha_1 - \alpha_2)\cos\alpha_3 \\ + \epsilon'\cos\alpha_1\cos(\alpha_2 - \alpha_3) + \epsilon\epsilon'\cos(\alpha_1 - \alpha_2 + \alpha_3)]. \quad (26)$$

Expanding Eq. (26) into a series of sums of cosines, sines,

$$A(\nu) = \frac{c}{4}[\cos(\alpha_1 + \alpha_2 + \alpha_3) + (1 + 2\epsilon')\cos(\alpha_1 + \alpha_2 - \alpha_3) \\ + (1 + 2\epsilon)\cos(\alpha_1 - \alpha_2 - \alpha_3) \\ + [1 + 2(\epsilon + \epsilon') + 4\epsilon\epsilon']\cos(\alpha_1 - \alpha_2 + \alpha_3)]. \quad (27)$$

In order to create eight equally spaced pulses, the ratios of $d_1{:}d_2{:}d_3$ must be some permutation of (1,2,4). If $\epsilon > \epsilon'$ in order for the pulse response to taper, the ratios of $d_1{:}d_2{:}d_3$ must be 1/2:1:1/4, the relative pulse amplitude sequence is $1.\ (1 + 2\epsilon'), (1 + 2\epsilon), 1 + 2(\epsilon + \epsilon') + 4\epsilon\epsilon', \ldots$. Using a similar analysis one can show that a network of four crystals and three partial polarizers does not result in a tapered pulse response.

Regardless of the amount of pulse overlap allowed, a monotonic taper cannot be achieved with more than one DPP module. However, the taper can be carried on by indefinitely adding APP modules with epsilon one-half. The transmission near a peak vs wavelength for DPP module followed by several APP modules is shown in Fig. 13.

Pulse Interleaving

In the sections above, pulse amplitude shaping and pulse overlapping have been investigated. Another useful operation on the pulse structure is pulse interleaving. Using pulse interleaving, it is possible to double the number of pulses and half the pulse spacing and hence double the FSR without using shorter crystals. This is a significant advantage because very thin crystals are both difficult to manufacture and extremely fragile. Consider a string of pulses separated by $t_1/2^N$,

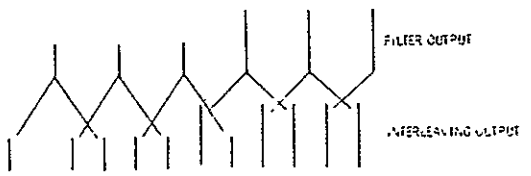


Fig. 14. Segment of a pulse diagram illustrating end and taper effects introduced by pulse interleaving.

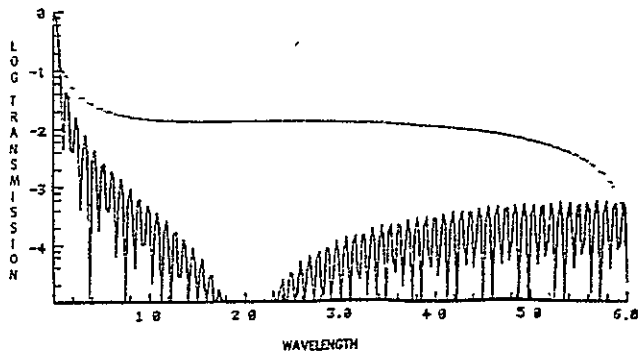


Fig. 15. Transmission (solid) and contribution function (dotted) vs wavelength for a seven element Lyot filter in which the last element is three times normal length. The length ratios are 1:1/2:1/4:1/8:1/16:1/32:3/64

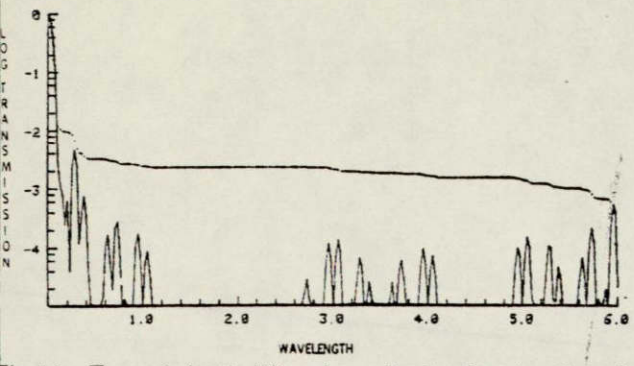


Fig. 16. Transmission (solid) and contribution function (dotted) vs wavelength for three module APP plus two Lyot modules, the last of which is three times normal length. The length ratios are 1:1/2: 1/3:1/6:1/9:1/18:1/32:3/64.

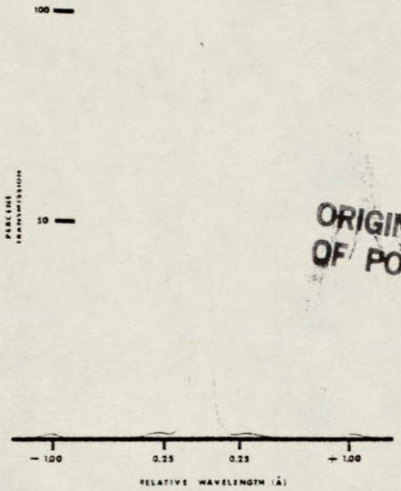


Fig. 17. Measured filter transmission in the neighborhood of λ5324. The FWHM is 0.11 Å. The spectrograph instrumental profile is 0.02 Å wide, so that the measurement agrees with the theoretical FWHM.

if the next crystal has a time difference of $t_1/2^N(1+2p)$, interior pulses will be separated by $t_1/2^{N+1}$, and only the pulses at the ends of the string will be missing. The effect of interleaving is shown in Fig. 14. Unfortunately, pulse interleaving also affects the taper, so that for apodized systems there is some disadvantage. A pure Lyot filter transmission profile in which the seventh module introduces interleaving is shown in Fig. 15. An APP filter with two Lyot modules, the second of which introduces interleaving is shown in Fig. 16. Figure 15 should be compared with Fig. 3 which is the corresponding filter without interleaving. Similarly, Fig. 16 should be compared with Fig. 4.

Experimental

A four module, eight crystal, APP filter has been built using Polaroid HN-38 for the perfect polarizers and two laminated sheets of HN-55 for the partial polarizers. The largest crystal of the filter is 79.312 mm long. At λ5324, the filter has a full width at half-maximum of 0.09 Å. The transmission vs wavelength at λ5324 is shown in Fig. 17.

The measured maximum transmission is 0.38 of the incident polarized light at λ6328 (He–Ne laser). The theoretical maximum transmission of the filter is 0.43 of the incident polarized light when the losses due to the transmission factors of the perfect and partial polarizers are taken into account. When the Fresnel reflection loss of the index matching fluid–calcite interface is added, the theoretical transmission becomes 0.41. Thus, the filter has 93% of its theoretical transmission. The use of four partial polarizers raises the maximum transmission by a factor of $(0.94/0.86)^4 = 1.43$, compared to a Lyot.

The APP filter constructed is designed as an operational filter for use on a solar telescope. In addition to the partial polarizers, it contains achromatic-half wave and quarterwave plates made of polyvinyl alcohol. The filter can be tuned to any wavelength from 4500 Å to 8500 Å on command from a teletype console. The filter can be stabilized to 0.01 Å by referencing the λ6328 Helium–Neon laser line. The filter is tuned by stepping motors that are commanded by a PDP11/10 minicom-

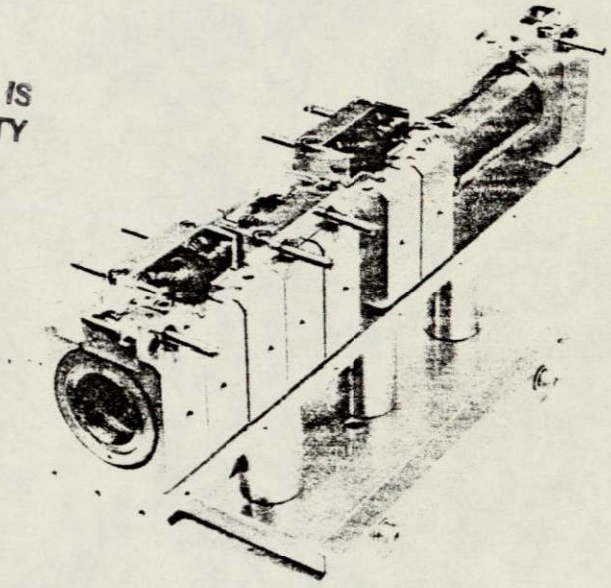


Fig. 18. Photograph of LAPPU filter before installation of outer corner.

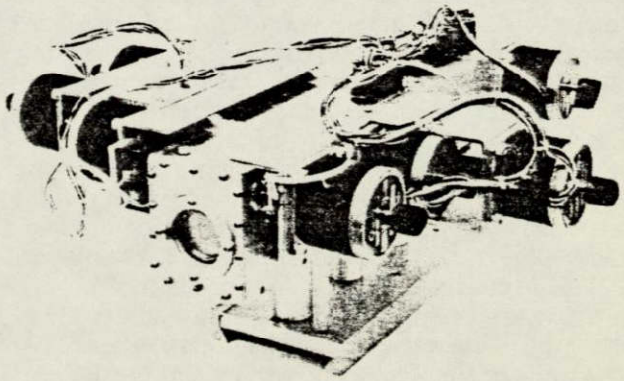


Fig. 19. Photograph of complete LAPPU filter.

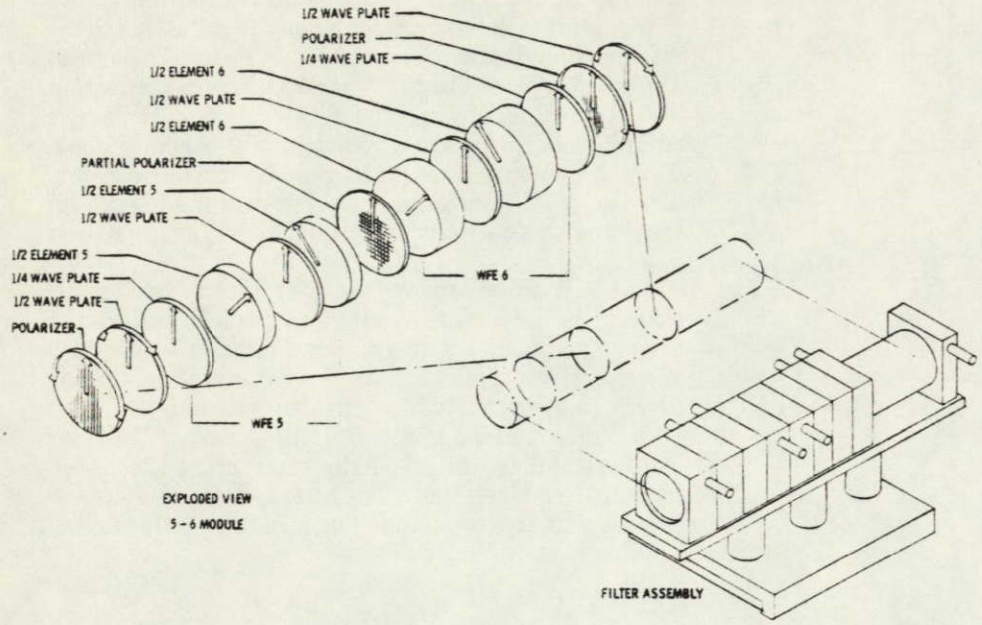


Fig. 20. Explode view of an APP module that is both wide field and tunable.

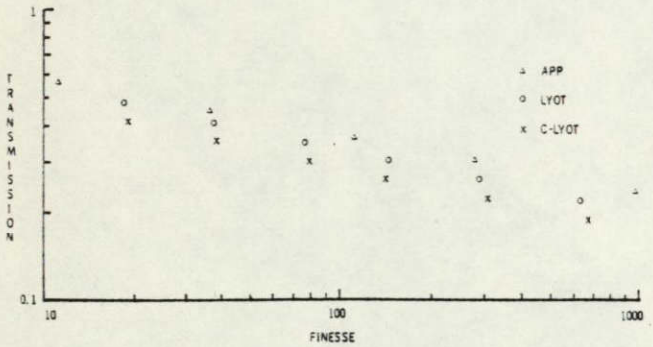


Fig. 21. Transmission vs finesse for APP, Lyot, and contrast element Lyot filters.

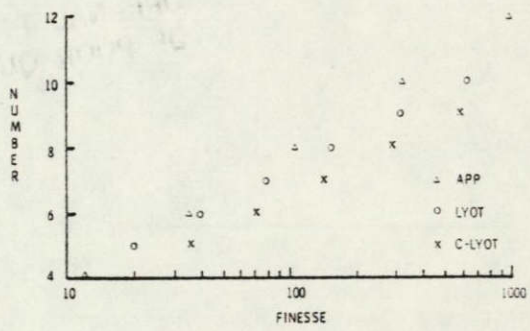


Fig. 22. Number of crystals vs finesse for APP, Lyot, and contrast element Lyot filters.

puter. The filter is shown without its case in Fig. 18. The completed filter with stepping motor drivers is shown in Fig. 19. Figure 20 shows an exploded view of a typical partial polarizer module.

Discussion

From measurements above, a filter using the alternate partial polarizer design operates substantially as predicted. The partial polarizers yield an increase in maximum transmission, and the nonstandard crystal length ratios result in an apodized transmission profile. For a filter with the same number of crystals, the APP is clearly superior in transmission to the Lyot. However, the pulse overlap of the APP diminishes the finesse, the ratio of FSR to FWHM, of the APP compared to the Lyot. This diminition, however, is not sufficient to remove the transmission advantage. The maximum theoretical transmission vs finesse for the APP, Lyot, and contrast element Lyot is shown in Fig. 21. The data for the figure were obtained by using 0.86 and 0.94 as the transmission factors of perfect and partial polarizers. The plot shows there is a transmission advantage to the APP design that increases with finesse. The reason for the behavior illustrated by Fig. 21 is that the contrast element filter adds a polarizer to achieve apodization, whereas the APP uses higher transmission polarizers to achieve apodization.

If one measures apodization quality by the ratio of the light outside the first zero of the transmission profile to the total light transmitted to the FSR, again the APP is superior to the contrast element Lyot. The ratio for the APP is 0.0095, while for the contrast element Lyot it is 0.035. A pure Lyot has, for comparison, a ratio of 0.085.

From the standpoint of both transmission and profile shape, the APP is clearly a superior design to a Lyot system. However, the design does have some practical disadvantages. More crystals are required in the APP for a required finesse. The number of crystals vs finesse for the APP, Lyot, and contrast Lyot is shown in Fig. 22. Also, slightly more calcite is required for the APP than the contrast Lyot. But since extra calcite is required only in elements after this third, there is not a significant disadvantage. By far the most serious drawback of the APP is the problem of compensating any residual birefringence of the partial polarizers. A detailed

analysis of the effect of birefringence on partial polarzers modules will be discussed in a later paper. However, anyone attempting to build an APP should be aware that Polaroid polarizing material is birefringent, and birefringence at the location of the partial polarizer surprisingly quickly destroys the apodization taper.

The concept of pulse response and its great utility was introduced to me by Larry Mertz. I gratefully acknowledge his patience and understanding. All the optical elements of the LAPPU were made by Harry Ramsey. The mechanical structure was designed and constructed by Ralph Reeves, the electronics and logic by Russell Lindgren. The computer design and programming was done by Stephen Schoolman. I gratefully acknowledge the skill, patience, and tenacity used in solving the construction problems of the filter.

The theoretical analysis of birefringent systems has been supported by Lockheed Independent Research funds. The filter was constructed under NASA contract NAS5-20783. The aid and cooperation of Goetz Oertel and John Mangus of NASA are gratefully acknowledged.

References

1. A. M. Title, Solar Phys. 33, 521 (1973).
2. A. M. Title, Appl. Opt. 14, 229 (1975).
3. A. M. Title, Appl. Opt. 14, 445 (1975).
4. S. E. Harris, E. O. Ammann, and I. C. Chang, J. Opt. Soc. Am. 54, 1267 (1964).
5. E. O. Ammann and I. C. Chang, J. Opt. Soc. Am 55, 835 (1965).
6. E. O. Ammann, J. Opt. Soc. Am. 56, 943 (1966).
7. J. W Evans, J. Opt. Soc. Am. 39, 229 (1949).
8. J. W Evans, J. Opt. Soc. Am. 48, 142 (1958).
9. L. Mertz, J. Opt. Soc. Am. Adv. 49 (Dec. 1959).
10. L. Mertz, J. Opt. Soc. Am. Adv. 50 (June 1960).
11. R. G. Giovanelli and J. T. Jefferies, Astron. J. Phys. 7, 254 (1954).
12. J. M. Beckers and R. B. Dunn, AFCRL-65-605 (Aug. 1965).
13. A M. Title, Solar Phys. 38, 523 (1974)
14. B. Lyot, Ann. Astron. 7, 31 (1944).
15. S. A. Schoolman, Solar Phys. 30, 255 (1973).