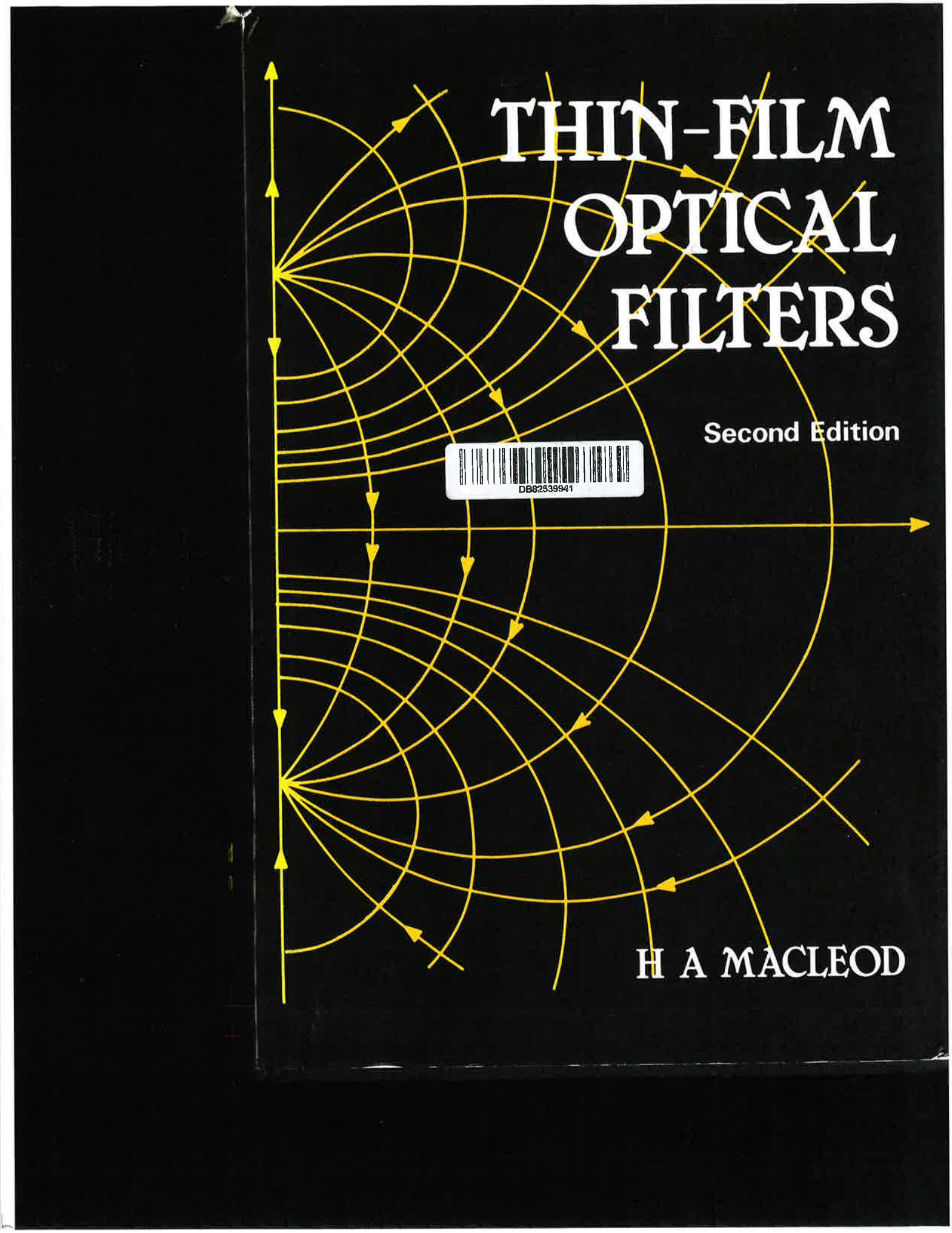
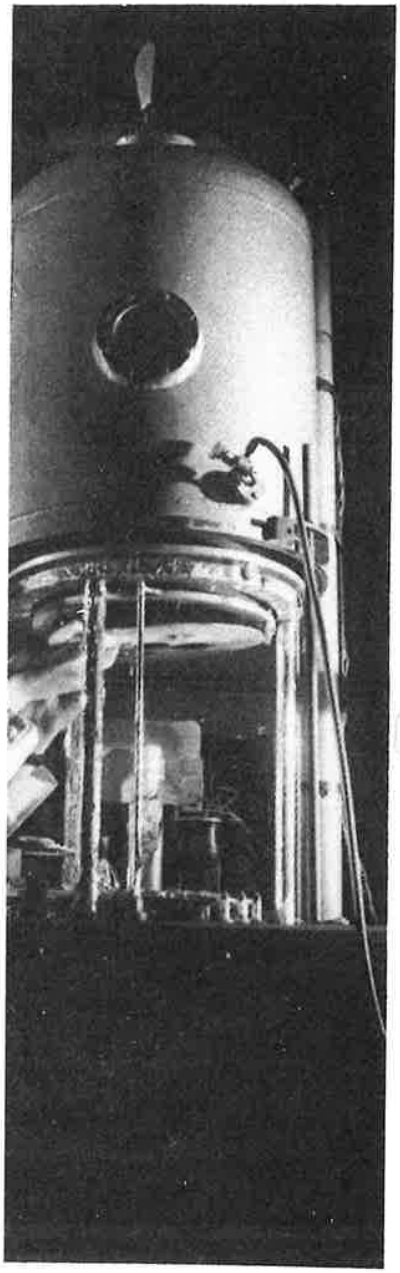

THIN-FILM
OPTICAL
FILTERS
Second Edition
DB62539941
H A MACLEOD



rowband filters. (Courtesy of Walter
ng, and Sir Howard Grubb, Parsons &

Thin-Film Optical Filters

Second Edition

H A Macleod

Professor of Optical Sciences
University of Arizona

McGraw-Hill Publishing Company,
New York, St. Louis, San Francisco,
Colorado Springs, Oklahoma City, Montreal,
Toronto
Adam Hilger, Bristol and Philadelphia

Published in the USA and Canada by McGraw-Hill Publishing Company, 1221 Avenue of the Americas, New York, NY 10020. For ordering information in the USA call 1-800-2-MCGRAW

Published in the rest of the world by IOP Publishing Ltd under the Adam Hilger imprint
Techno House, Redcliffe Way, Bristol BS1 6NX
242 Cherry Street, Philadelphia, PA 19106, USA

Phototypeset by Macmillan India Ltd, Bangalore
Printed in Great Britain by J W Arrowsmith Ltd, Bristol

Library of Congress Catalog Card Number 85-043569

ISBN 0-07-044694-6 (McGraw-Hill Publishing Company)
ISBN 0-85274-784-5 (IOP Publishing Ltd)

Consultant Editor: **Professor W T Welford**
Imperial College, London

First published 1969
Second edition 1986
Reprinted (US edition) 1989

my M
Agnes
J

Contents

7 Band-pass filters

A filter which possesses a region of transmission bounded on either side by regions of rejection is known as a band-pass filter. For the broadest band-pass filters, the most suitable construction is a combination of longwave-pass and shortwave-pass filters, which we discussed in chapter 6. For narrower filters, however, this method is not very successful because of difficulties associated with obtaining both the required precision in positioning and the steepness of edges. Other methods are therefore used, involving a single assembly of thin films to produce simultaneously the pass and rejection bands. The simplest of these is the thin-film Fabry–Perot filter, a development of the interferometer already described in chapter 5. The thin-film Fabry–Perot filter has a pass band shape which is triangular and it has been found possible to improve this by coupling simple filters in series in much the same way as tuned circuits. These coupled arrangements are known as multiple cavity filters or multiple half-wave filters. If two simple Fabry–Perot filters are combined, the resultant becomes a double cavity or double half-wave filter, abbreviated to DHW filter, while, if three Fabry–Perot filters are involved, we have a triple cavity filter, abbreviated normally to THW for triple half-wave. In the earlier part of this chapter, we consider single cavity filters. First of all, we examine combinations of edge filters.

BROADBAND-PASS FILTERS

Band-pass filters can be very roughly divided into broadband-pass filters and narrowband-pass filters. There is no definite boundary between the two types and the description of one particular filter usually depends on the application and the filters with which it is being compared. For the purposes of the present work, by broadband filters we mean filters with bandwidths of perhaps 20% or more which are made by combining longwave-pass and shortwave-pass filters. The best arrangement is probably to deposit the two components on opposite

sides of a single substrate. To give maximum possible transmission, each edge filter should be designed to match the substrate into the surrounding medium, a procedure already examined in chapter 6. Such a filter is shown in figure 7.1.

It is also possible, however, to deposit both components on the same side of the substrate. This was a problem which Epstein[1] examined in his early paper on symmetrical periods. The main difficulty is the combining of the two stacks so that the transmission in the pass band is a maximum and also so that one stack does not produce transmission peaks in the rejection zone of the other. The transmission in the pass band will depend on the matching of the first stack to the substrate, the matching of the second stack to the first, and the matching of the second stack to the surrounding medium. Depending on the equivalent admittances of the various stacks it may be necessary to insert quarter-wave matching layers or to adopt any of the more involved matching techniques.

In the visible region, with materials such as zinc sulphide and cryolite, the combination $[(H/2)L(H/2)]^s$ acts as a good longwave-pass filter with an equivalent admittance at normal incidence and at wavelengths in the pass region not too far removed from the edge of near unity. This can therefore be used next to the air without mismatch. The combination $[(L/2)H(L/2)]^s$ acts as a shortwave-pass filter, with equivalent admittance only a little lower than the first section, and can be placed next to it, between it and the substrate, without any matching layers. The mismatch between this second section and the substrate, which in the visible region will be glass of index 1.52, is sufficiently large to require a matching layer. Happily, the $[(H/2)L(H/2)]$ combination with a total phase thickness of 270°, i.e. effectively three quarter-waves, has an admittance exactly correct for this. The transmission of the final design is shown in figure 7.2(*b*) with the appropriate admittances of the two sections in figure 7.2(*a*). Curve A refers to a $[(L/2)H(L/2)]^4$ shortwave-pass section and B to a $[(H/2)L(H/2)]^4$ longwave-pass. The complete design is shown in table 7.1. The edges of the two sections have been chosen quite arbitrarily and could be moved as required.

To avoid the appearance of transmission peaks in the rejection zones of either component, it is safest to deposit them so that high-reflectance zones do not overlap. The complete rejection band of the shortwave-pass section will always lie over a pass region of the longwave-pass filter, but the higher-order bands should be positioned, if at all possible, clear of the rejection zone of the longwave-pass section. The combination of edge filters of the same type has already been investigated in chapter 6 and the principles discussed there apply to this present situation. It should also be remembered that, although in the normal shortwave-pass filter the second-order reflection peak is missing, a small peak can appear if any thickness errors are present. This can, if superimposed on a rejection zone of the other section, cause the appearance of a transmission peak if the errors are sufficiently pronounced. The expression

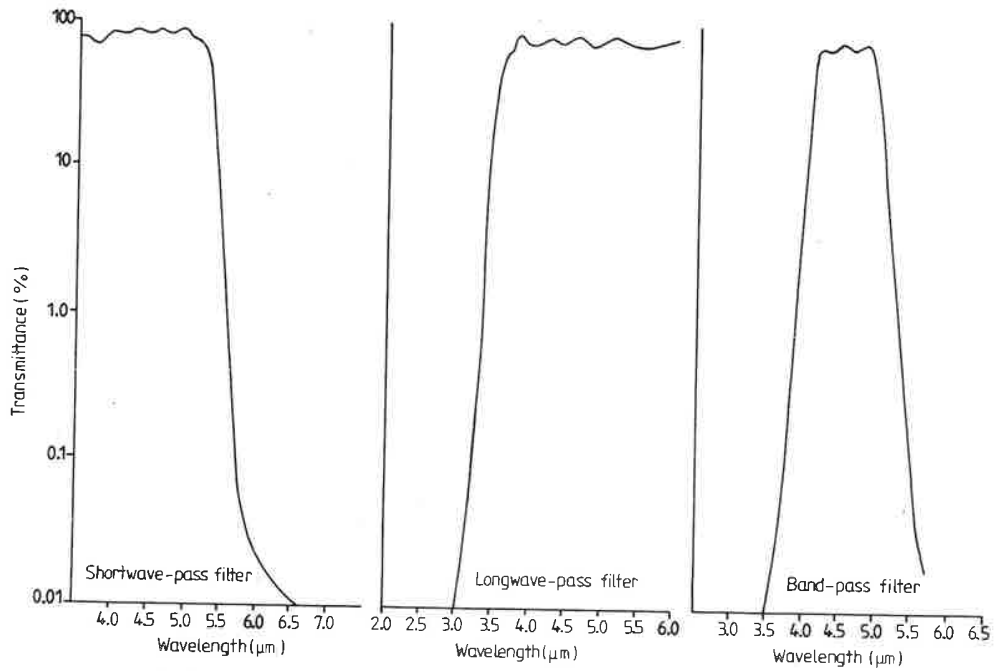


Figure 7.1 The construction of a band-pass filter by placing two separate edge filters in series. (Courtesy of Standard Telephones and Cables Ltd.)

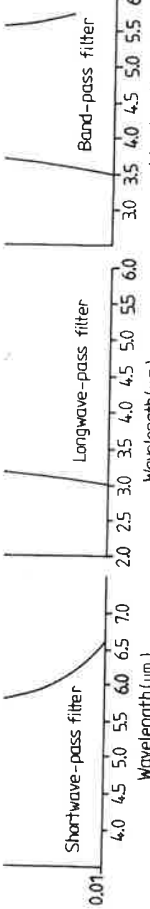


Figure 7.1 The construction of a band-pass filter by placing two separate edge filters in series. (Courtesy of Standard Telephones and Cables Ltd.)

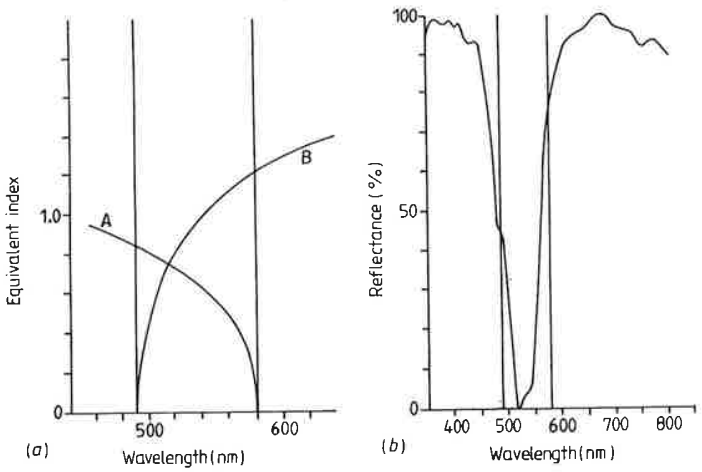


Figure 7.2 (*a*) Equivalent admittances of two stacks made up of symmetrical periods used to form a band-pass filter. A: $(0.5L\,H\,0.5L)$; B: $(0.5H\,L\,0.5H)$, where $n_L = 1.38$; $n_H = 2.30$. (*b*) Calculated reflectance curve for a band-pass filter. For the complete design of this filter, made up of two superimposed stacks, one of type A and one of type B, refer to table 7.1. (After Epstein[1].)

Table 7.1†

Index	Phase thickness of each layer measured at 546 nm (degrees)	Index	Phase thickness of each layer measured at 546 nm (degrees)
1.52	Massive	1.38	55.4
1.38	67.3	2.30	33.9
2.30	134.5	1.38	67.9
1.38	122.7	2.30	67.9
2.30	110.8	1.38	67.9
1.38	110.8	2.30	67.9
2.30	110.8	1.38	67.9
1.38	110.8	2.30	67.9
2.38	110.8	1.38	67.9
1.38	110.8	2.30	33.9
2.30	110.8	1.00	Massive

† From Epstein[1].

for maximum transmission is

$$T_{\max} = \frac{T_a T_b}{[1-(R_a R_b)^{1/2}]^2}$$

but this only holds if the phase conditions are met.

NARROWBAND FILTERS

When we speak of narrowband filters we are referring to filters with bandwidths of perhaps 15 % or less. The main difference between these and the broadband filters described above is that here we are relying on a single assembly to give the pass band and both the adjoining stop bands.

The metal–dielectric Fabry–Perot filter

The simplest type of narrowband thin-film filter is based on the Fabry–Perot interferometer discussed in chapter 5. In its original form, the Fabry–Perot interferometer consists of two identical parallel reflecting surfaces spaced apart a distance d. In collimated light, the transmission is low for all wavelengths except for a series of very narrow transmission bands spaced at intervals that are constant in terms of wavenumber. This device can be replaced by a complete thin-film assembly consisting of a dielectric layer bounded by two metallic reflecting layers (figure 7.3). The dielectric layer takes the place of the spacer and is known as the spacer layer. Except that the spacer layer now has an index greater than unity, the analysis of the performance of this thin-film filter is exactly the same as for the conventional etalon, but in other respects there are a few significant differences.

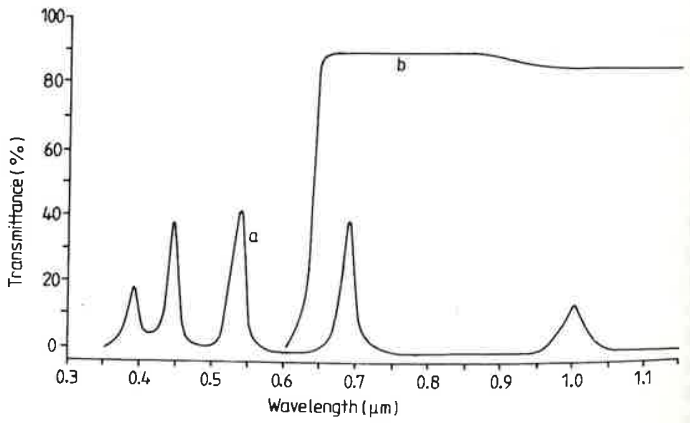


Figure 7.3 Characteristics of a metal–dielectric filter for the visible region (curve a). Curve b is the transmittance of an absorption glass filter that can be used for the suppression of the short wavelength sidebands. (Courtesy of Barr & Stroud Ltd.)

While the surfaces of the substrates should have a high degree of polish, they need not be worked to the exacting tolerances necessary for etalon plates. Provided the vapour stream in the plant is uniform, the films will follow the contours of the substrate without exhibiting thickness variations. This implies

that it is possible for the t
order than the conventic
orders must be used, be
than the fourth order or
pass band and reduces th
the higher order is com
metal–dielectric Fabry–P
described later.

It is worthwhile briefly
again, this time including
point for this analysis is

$$T_F = \frac{}{[1 -}$$

$$F = \frac{4(}{[1 -}$$

where the notation is gi
slightly by removing the
follows is similar to that
including the effects of

$$\frac{2\pi nd\cos\theta}{\lambda} - \phi$$

where we have chosen
definition. The analysis
wavenumber instead of
by

$$\frac{1}{\lambda} = \nu = \frac{m\pi -}{2}$$

Depending on the pa
substrate, and the index
either in the first or seco
between 0 and 1 and rou
filter will therefore be shi
expected simply from th

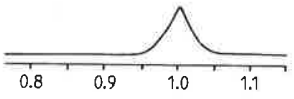
The resolving power
exactly the same way a
convenient definition is

where the halfwidth is

FILTERS

we are referring to filters with
in difference between these and the
here we are relying on a single
he adjoining stop bands.

filter is based on the Fabry–Perot
s original form, the Fabry–Perot
arallel reflecting surfaces spaced
the transmission is low for all
ow transmission bands spaced at
umber. This device can be replaced
g of a dielectric layer bounded by
e dielectric layer takes the place of
. Except that the spacer layer now
s of the performance of this thin-
onventional etalon, but in other
ces.



filter for the visible region (curve a).
glass filter that can be used for the
(Courtesy of Barr & Stroud Ltd.)

have a high degree of polish, they
nces necessary for etalon plates.
uniform, the films will follow the
thickness variations. This implies

that it is possible for the thin-film Fabry–Perot filter to be used in a much lower order than the conventional etalon. Indeed, it turns out in practice that lower orders must be used, because the thin-film spacer layers begin, when thicker than the fourth order or so, to exhibit roughness. This roughness broadens the pass band and reduces the peak transmittance so much that any advantage of the higher order is completely lost. This simple type of filter is known as a metal–dielectric Fabry–Perot to distinguish it from the all-dielectric one to be described later.

It is worthwhile briefly analysing the performance of the Fabry–Perot once again, this time including the effects of phase shift at the reflectors. The starting point for this analysis is equation (2.81):

$$\left.\begin{aligned} T_F &= \frac{T_a T_b}{[1-(R_a R_b)^{1/2}]^2}\,\frac{1}{1+F\sin^2[\tfrac{1}{2}(\phi_a+\phi_b)-\delta]} \\ F &= \frac{4(R_a R_b)^{1/2}}{[1-(R_a R_b)^{1/2}]^2} \qquad \delta = \frac{2\pi nd\cos\theta}{\lambda} \end{aligned}\right\} \tag{7.1}$$

where the notation is given in figure 2.12. We have adapted equation (2.81) slightly by removing the + and − signs on the reflectances. The analysis which follows is similar to that already performed in chapter 5 except that here we are including the effects of ϕ_a and ϕ_b. The maxima of transmission are given by

$$\frac{2\pi nd\cos\theta}{\lambda} - \frac{\phi_a+\phi_b}{2} = m\pi \qquad m = 0, \pm 1, \pm 2, \pm 3, \ldots \tag{7.2}$$

where we have chosen $-m$ rather than $+m$ because $(\phi_a+\phi_b)/2 < \pi$ by definition. The analysis is marginally simpler if we work in terms of wavenumber instead of wavelength. The positions of the peaks are then given by

$$\frac{1}{\lambda} = \nu = \frac{m\pi + (\phi_a+\phi_b)/2}{2\pi nd\cos\theta} = \frac{1}{2nd\cos\theta}\left(m + \frac{\phi_a+\phi_b}{2\pi}\right). \tag{7.3}$$

Depending on the particular metal, the thickness, and the index of the substrate, and the index of the spacer, the phase shift on reflection ϕ will be either in the first or second quadrant. $(\phi_a+\phi_b)/(2\pi)$ will therefore be positive, between 0 and 1 and roughly in the region of 0.5. The peak wavelength of the filter will therefore be shifted to the shortwave side of the peak which would be expected simply from the optical thickness of the spacer layer.

The resolving power of the thin-film Fabry–Perot filter may be defined in exactly the same way as for the interferometer. As we saw in chapter 5, a convenient definition is

$$\frac{\text{Peak wavelength}}{\text{Halfwidth of pass band}}$$

where the halfwidth is the width of the band measured at half the peak

transmission. Now let the pass bands be sufficiently narrow, which is the same as F being sufficiently large, so that near a peak we can replace

$$\frac{\phi_a + \phi_b}{2} - \delta \quad \text{by} \quad -m\pi - \Delta\delta$$

and

$$\sin^2\left(\frac{\phi_a + \phi_b}{2} - \delta\right) \quad \text{by} \quad (\Delta\delta)^2.$$

We are assuming here that ϕ_a and ϕ_b are constant or vary very much more slowly than δ over the pass band.

The half-peak bandwidth, or halfwidth, can be found by noting that at the half-peak transmission points

$$F\sin^2\left(\frac{\phi_a + \phi_b}{2} - \delta\right) = 1.$$

Using the approximation given above, this becomes

$$(\Delta\delta_h)^2 = \frac{1}{F}$$

i.e. the halfwidth of the pass band

$$2\Delta\delta_h = 2/F^{1/2}.$$

The finesse is defined as the ratio of the interval between fringes to the fringe halfwidth, and is written $\mathscr{F}$. The change in δ in moving from one fringe to the next is just π, and the finesse, therefore, is

$$\mathscr{F} = \frac{\pi F^{1/2}}{2}. \tag{7.4}$$

Now $\nu_0/\Delta\nu_h = \delta_0/2\Delta\delta_h$ because $\nu \propto \delta$, where ν_0 and δ_0 are respectively the values of the wavenumber and spacer layer phase thickness associated with the transmission peak, and $\Delta\nu_h$ and $2\Delta\delta_h$ are the corresponding values of halfwidth. The ratio of the peak wavenumber to the halfwidth is then given by

$$\frac{\nu_0}{\Delta\nu_h} = \mathscr{F}\left(m + \frac{\phi_a + \phi_b}{2\pi}\right) \tag{7.5}$$

for a peak of order m, since

$$\delta_0 = m\pi + \frac{\phi_a + \phi_b}{2}.$$

The ratio of peak position to halfwidth expressed in terms of wavenumber is exactly the same in terms of wavelength,

$$\frac{\nu_0}{\Delta\nu_h} = \frac{\lambda_0}{\Delta\lambda_h}$$

where λ_0 is given by

λ_0

and this was discussed in ch
parameter with which to speci
be converted very quickly into
rather than resolving power f
whether or not they are Fabr
expressed as a percentage, is
facturers and users alike. Oth
along with the halfwidth are t
at 0.1 × peak transmission, a
Perot filter, provided the p
layers are effectively constan
respectively by one-third of th
times the halfwidth. The othe
indication of the extent to whic
band, compared with those of t

The manufacture of the met
point to watch is that the met
possible on to a cold substrate.
results are probably achieved
the best combination is alumi
Wherever possible the layers
over them as soon as possible
assembly by equalising the re
layers.

Turner[2] quoted some result
visible region which may be tak
The filters were constructed
spacers. For a first-order s
transmission of 30 % was obta
filter with a second-order s
transmission of 26 % at 535 nm
usually the highest used. Bec
becomes increasingly apparen
which would otherwise arise

A typical curve for a metal–
figure 7.3. The particular peak
third order. The shortwave si
suppressed quite easily by the
be cemented over the metal–d
filter is also shown in the figure

ciently narrow, which is the same
peak we can replace

$-m\pi - \Delta\delta$

$(\Delta\delta)^2$.

onstant or vary very much more

an be found by noting that at the

$= 1.$

becomes

rval between fringes to the fringe
in moving from one fringe to the

(7.4)

e ν_0 and δ_0 are respectively the
ase thickness associated with the
the corresponding values of
to the halfwidth is then given by

$$\left(\frac{+\phi_b}{2\pi}\right) \qquad (7.5)$$

ϕ_b.

ressed in terms of wavenumber is

where λ_0 is given by

$$\lambda_0 = \frac{2nd\cos\theta}{m + (\phi_a + \phi_b)/2\pi} \qquad (7.6)$$

and this was discussed in chapter 5. The halfwidth is thus a most useful parameter with which to specify a narrowband Fabry–Perot filter since it can be converted very quickly into a measure of resolution. It has come to be used rather than resolving power for all types of narrowband filter, regardless of whether or not they are Fabry–Perot type. Usually, therefore, $\Delta\lambda_h/\lambda_0$, often expressed as a percentage, is the parameter which is quoted by the manufacturers and users alike. Other measures of bandwidth sometimes quoted along with the halfwidth are the widths measured at 0.9 × peak transmission, at 0.1 × peak transmission, and at 0.01 × peak transmission. For a Fabry–Perot filter, provided the phase shifts on reflection from the reflecting layers are effectively constant over the pass band, these widths are given respectively by one-third of the halfwidth, three times the halfwidth, and ten times the halfwidth. The other measures of bandwidth are used to give some indication of the extent to which, in any given type of filter, the sides of the pass band, compared with those of the Fabry–Perot, can be considered rectangular.

The manufacture of the metal–dielectric filter is straightforward. The main point to watch is that the metallic layers should be evaporated as quickly as possible on to a cold substrate. In the visible and near infrared regions the best results are probably achieved with silver and cryolite, while in the ultraviolet the best combination is aluminium and either magnesium fluoride or cryolite. Wherever possible the layers should be protected by cementing a cover slip over them as soon as possible after deposition. This also serves to balance the assembly by equalising the refractive indices of the media outside the metal layers.

Turner[2] quoted some results for metal–dielectric filters constructed for the visible region which may be taken as typical of the performance to be expected. The filters were constructed from silver reflectors and magnesium fluoride spacers. For a first-order spacer a bandwidth of 13 nm with a peak transmission of 30 % was obtained at a peak wavelength of 531 nm. A similar filter with a second-order spacer gave a bandwidth of 7 nm with peak transmission of 26 % at 535 nm. With metal–dielectric filters the third order is usually the highest used. Because of scattering in the spacer layer, which becomes increasingly apparent in the fourth and higher orders, any benefit which would otherwise arise from using these orders is largely lost.

A typical curve for a metal–dielectric filter for the visible region is shown in figure 7.3. The particular peak to be used is that at 0.69 μm, which is of the third order. The shortwave sidebands due to the higher-order peaks can be suppressed quite easily by the addition of an absorption glass filter, which can be cemented over the metal–dielectric element to act as a cover glass. Such a filter is also shown in the figure and is one of a wide range of absorption glasses

which are available for the visible and near infrared and which have longwave-pass characteristics. There are, unfortunately, few absorption filters suitable for the suppression of the longwave sidebands. If the detector which is to be used is not sensitive to these longer wavelengths, then no problem exists and commercial metal–dielectric filters for the visible and near infrared usually possess long-wavelength sidebands beyond the limit of the photocathodes or photographic emulsions, which are the usual detectors for this region. If the longwave-sideband suppression must be included as part of the filter assembly, then there is an advantage in using metal–dielectric filters in the first order, even though the peak transmission for a given bandwidth is much lower, since they do not usually possess long-wavelength sidebands. Theoretically, there will always be a peak corresponding to the zero order at very long wavelengths, but this will not usually appear, partly because the substrate will cut off long before the zero order is reached, and also because the properties of the thin-film materials themselves will change radically. We shall discuss later a special type of metal–dielectric filter, the induced transmission filter, which can be made to have a much higher peak transmission, though with a rather broader halfwidth, without introducing long-wavelength sidebands, and which is often used as a long-wavelength suppression filter.

Silver does not have an acceptable performance for ultraviolet filters and aluminium has been found to be the most suitable metal with magnesium fluoride as the preferred dielectric. In the ultraviolet beyond 300 nm there are few suitable cements (none at all beyond 200 nm) and it is not possible to use cover slips which are cemented over the layers in the way in which filters for the visible region are protected. The normal technique, therefore, is to attempt to protect the filter by the addition of an extra dielectric layer between the final metal layer and the atmosphere. These layers are effective in that they slow down the oxidation of the aluminium which otherwise takes place rapidly and causes a reduction in performance even at quite low pressures. This oxidation has already been referred to in chapter 4. They cannot completely stabilise the filters, however, and slight longwave drifts can occur, as reported by Bates and Bradley[3]. A second function of the final dielectric layer is to act as a reflection-reducing layer at the outermost metal surface and hence to increase the transmittance of the filter. This is not a major effect—the problem of improving metal–dielectric filter performance is dealt with later in this chapter—but any technique which helps to improve performance, even marginally, in the ultraviolet, is very welcome. Some performance curves of first-order metal–dielectric Fabry–Perot filters are shown in figure 7.4.

The formula for transmission of the Fabry–Perot filter can also be used to determine both the peak transmission in the presence of absorption in the reflectors and the tolerance which can be allowed in matching the two reflectors. First of all, let the reflectances be equal and let the absorption be denoted by A, so that

$$R+T+A=1. \tag{7.8}$$

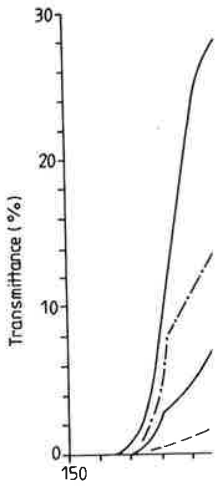


Figure 7.4 Experimental t
for the far ultraviolet deposi

The peak transmission w

and, using equation (7.8

exactly as for the Fab
absorption is present the
A/T.

To estimate the accura
we assume that the abso
expression

where the subscripts a a

where Δ_a is the error in

$(T_F)_{pe}$

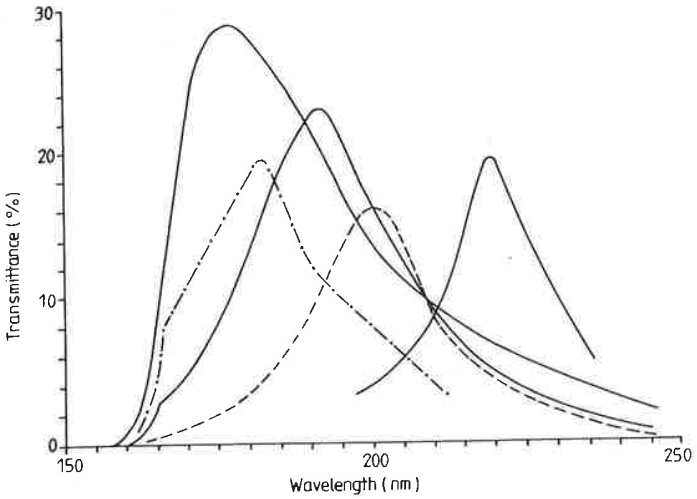


Figure 7.4 Experimental transmittance curves of first-order metal–dielectric filters for the far ultraviolet deposited on Spectrosil B substrates. (After Bates and Bradley[3].)

The peak transmission will then be given by

$$(T_F)_{\text{peak}} = \frac{T^2}{(1-R)^2}$$

and, using equation (7.8),

$$(T_F)_{\text{peak}} = \frac{1}{(1+A/T)^2} \tag{7.9}$$

exactly as for the Fabry–Perot interferometer, which shows that when absorption is present the value of peak transmission is determined by the ratio A/T.

To estimate the accuracy of matching which is required for the two reflectors we assume that the absorption is zero. The peak transmission is given by the expression

$$(T_F)_{\text{peak}} = \frac{T_a T_b}{[1-(R_a R_b)^{1/2}]^2} \tag{7.10}$$

where the subscripts a and b refer to the two reflectors. Let

$$R_b = R_a - \Delta_a \tag{7.11}$$

where Δ_a is the error in matching, so that $T_b = T_a + \Delta_a$. Then we can write

$$(T_F)_{\text{peak}} = \frac{T_a(T_a+\Delta_a)}{\{1-[R_a(R_a-\Delta_a)]^{1/2}\}^2}$$

$$= \frac{T_a(T_a+\Delta_a)}{\{1-R_a[1-\frac{1}{2}(\Delta_a/R_a)+\ldots]\}^2}. \tag{7.12}$$

Now assume that Δ_a is sufficiently small compared with R_a so that we can take only the first two terms of the expansion in equation (7.12). With some rearrangement the equation becomes

$$(T_F)_{\text{peak}} = \frac{T_a^2}{(1-R_a)^2} \frac{1+(\Delta_a/T_a)}{[1+\frac{1}{2}(\Delta_a/T_a)]^2}. \tag{7.13}$$

The first part of the equation is the expression for peak transmission in the absence of any error in the reflectors, while the second part shows how the peak transmission is affected by errors. The second part of the expression is plotted in figure 7.5 where the abscissa is $T_b/T_a = 1 + \Delta_a/T_a$. Clearly, the Fabry–Perot filter is surprisingly insensitive to errors. Even with reflector transmittance unbalanced by a factor of 3, it is still possible to achieve 75% peak transmission.

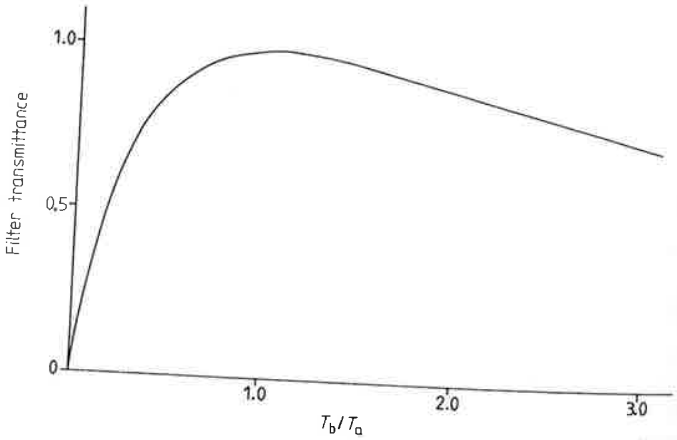


Figure 7.5 Theoretical peak transmittance of a Fabry–Perot filter with unbalanced reflectors.

The all-dielectric Fabry–Perot filter

In the same way as we found for the conventional Fabry–Perot etalon, if improved performance is to be obtained, then the metallic reflecting layers should be replaced by all-dielectric multilayers.

An all-dielectric filter is shown in diagrammatic form in figure 7.6. Basically, this is the same as the conventional etalon with dielectric coatings and with a solid thin-film spacer, and the observations made for the metal–dielectric filter are also valid. Again, the substrate need not be worked to a high degree of flatness although the polish must be good, because, provided the plant geometry is adequate, the films will follow any contours without showing changes in thickness.

pared with R_a so that we can take
in equation (7.12). With some

$$\frac{+(\Delta_a/T_a)}{+\frac{1}{2}(\Delta_a/T_a)]^2}. \qquad (7.13)$$

ion for peak transmission in the
the second part shows how the
second part of the expression is
$T_b/T_a = 1 + \Delta_a/T_a$. Clearly, the
e to errors. Even with reflector
t is still possible to achieve 75%

Fabry–Perot filter with unbalanced

entional Fabry–Perot etalon, if
ien the metallic reflecting layers
ers.
natic form in figure 7.6. Basically,
ith dielectric coatings and with a
ade for the metal–dielectric filter
be worked to a high degree of
d, because, provided the plant
any contours without showing

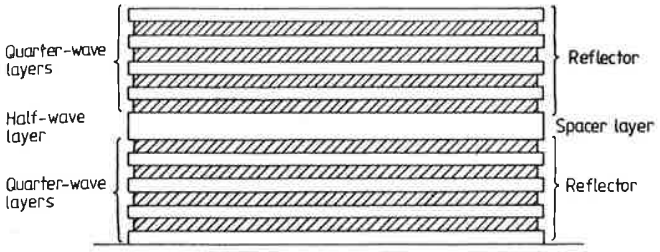


Figure 7.6 The structure of an all-dielectric Fabry–Perot filter.

The bandwidth of the all-dielectric filter can be calculated as follows. If the reflectance of each of the multilayers is sufficiently high, then

$$F = \frac{4R}{(1-R)^2} \simeq \frac{4}{T^2}$$

and

$$\frac{\lambda_0}{\Delta\lambda_h} = m\mathscr{F} = \frac{m\pi F^{1/2}}{2} \simeq \frac{m\pi}{T}. \qquad (7.14)$$

Since the maximum reflectance for a given number of layers will be obtained with a high-index layer outermost, there are really only two cases which need be considered and these are shown in figure 7.7. If x is the number of high-index layers in each stack, not counting the spacer layer, then in the case of the high-index spacer, the transmission of the stack will be given by

$$T = \frac{4n_L^{2x} \cdot n_s}{n_H^{2x+1}}$$

and in the case of the low-index spacer by

$$T = \frac{4n_L^{2x-1} n_s}{n_H^{2x}}.$$

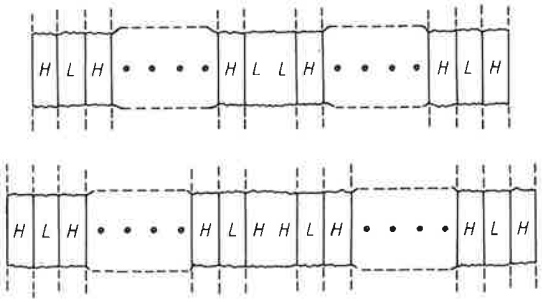


Figure 7.7 The structure of the two basic types of all-dielectric Fabry–Perot filter.

Substituting these results into the expression for bandwidth we find, for the high-index spacer,

$$\frac{\Delta\lambda_h}{\lambda_0} = \frac{4n_L^{2x}n_s}{m\pi n_H^{2x+1}} \tag{7.15}$$

and, for the low-index spacer,

$$\frac{\Delta\lambda_h}{\lambda_0} = \frac{4n_L^{2x-1}n_s}{m\pi n_H^{2x}} \tag{7.16}$$

where we are adopting the fractional halfwidth $\Delta\lambda_h/\lambda_0$ rather than the resolving power $\lambda_0/\Delta\lambda_h$ as the important parameter. This is customary practice.

In these formulae we have completely neglected any effect due to the dispersion of phase change on reflection from a multilayer. As we have already noted in chapter 5, the phase change is not constant. The sense of the variation is such that it increases the rate of variation of $[(\phi_a+\phi_b)/2]-\delta$ with wavelength in the formula for transmission of the Fabry–Perot filter and, hence, reduces the bandwidth and increases the resolving power in equations (7.15) and (7.16). Seeley[4] has studied the all-dielectric filter in detail and by making some approximations in the basic expressions for the filter transmittance, has arrived at formulae for the first-order halfwidths, which, with a little adjustment, become equal to the expressions in (7.15) and (7.16) multiplied by a factor $(n_H-n_L)/n_H$. We can readily extend Seeley's analysis to all-dielectric filters of order m.

We recall that the half-peak points are given by

$$F\sin^2[(2\pi D/\lambda)-\phi] = 1 \tag{7.17}$$

where, since the filter is quite symmetrical, we have replaced $(\phi_1+\phi_2)/2$ by ϕ. It is simpler to carry out the analysis in terms of $g = \lambda_0/\lambda = \nu/\nu_0$. At the peak of the filter we have $g = 1.0$. We can assume for small changes Δg in g that

$$2\pi D/\lambda = m\pi(1+\Delta g)$$

and

$$\phi = \phi_0 + \frac{d\phi}{dg}\Delta g$$

so that equation (7.17) becomes

$$F\sin^2\left(m\pi(1+\Delta g)-\phi_0-\frac{d\phi}{dg}\Delta g\right) = 1.$$

ϕ_0, we know, is 0 or π, and so, using the same approximation as before,

$$F\left(m\pi\Delta g-\frac{d\phi}{dg}\Delta g\right)^2 = 1$$

or

$$\Delta g = F^{-1/2}\left(m\pi-\frac{d\phi}{dg}\right)^{-1}.$$

The halfwidth is $2\Delta g$ so that

$$2\Delta g = \frac{\Delta\nu_h}{\nu_0} =$$

$$= \frac{2}{m\pi F^1}$$

We now need the quantity $d\phi$ follow him exactly, we choos shall require the results later.

[

where, as usual, we are writin space units. Now, for layers v

where ε is small. Then

$\cos\delta$

so that the matrix can be wri

We limit our analysis to quarte to the substrate. There are tw

Case I: even number (2x) of lay

The resultant multilayer matr

$$\begin{bmatrix} B \\ C \end{bmatrix} = [L]$$

where

[

[H

Then

$$\begin{bmatrix} B \\ C \end{bmatrix} = \{[$$

$$= \left[$$

sion for bandwidth we find, for the

$$\frac{{}^{2x}n_s}{\frac{2x+1}{H}} \tag{7.15}$$

$$\frac{{}^{-1}n_s}{n_H^{2x}} \tag{7.16}$$

halfwidth $\Delta\lambda_h/\lambda_0$ rather than the rameter. This is customary practice. y neglected any effect due to the om a multilayer. As we have already constant. The sense of the variation riation of $[(\phi_a+\phi_b)/2]-\delta$ with on of the Fabry–Perot filter and, s the resolving power in equations all-dielectric filter in detail and by expressions for the filter transmit- rder halfwidths, which, with a little ns in (7.15) and (7.16) multiplied by d Seeley's analysis to all-dielectric

given by

$$-\phi] = 1 \tag{7.17}$$

we have replaced $(\phi_1+\phi_2)/2$ by ϕ. ns of $g = \lambda_0/\lambda = \nu/\nu_0$. At the peak ne for small changes Δg in g that

$+\Delta g)$

Δg

$$-\frac{d\phi}{dg}\Delta g\Big) = 1.$$

ame approximation as before,

$$\Big)^2 = 1$$

$$\frac{d\phi}{dg}\Big)^{-1}.$$

The halfwidth is $2\Delta g$ so that

$$2\Delta g = \frac{\Delta\nu_h}{\nu_0} = \frac{\Delta\lambda_h}{\lambda_0} = 2F^{-1/2}\left(m\pi - \frac{d\phi}{dg}\right)^{-1}$$
$$= \frac{2}{m\pi F^{1/2}}\left(1 - \frac{1}{m\pi}\frac{d\phi}{dg}\right)^{-1}. \tag{7.18}$$

We now need the quantity $d\phi/dg$. We use Seeley's technique, but, rather than follow him exactly, we choose a slightly more general approach because we shall require the results later. The matrix for a dielectric quarter-wave layer is

$$\begin{bmatrix} \cos\delta & (i\sin\delta)/n \\ in\sin\delta & \cos\delta \end{bmatrix}$$

where, as usual, we are writing n for the optical admittance, which is in free space units. Now, for layers which are almost a quarter-wave we can write

$$\delta = \pi/2 + \varepsilon$$

where ε is small. Then

$$\cos\delta \simeq -\varepsilon \qquad \sin\delta \simeq 1$$

so that the matrix can be written

$$\begin{bmatrix} -\varepsilon & i/n \\ in & -\varepsilon \end{bmatrix}.$$

We limit our analysis to quarter-wave multilayer stacks having high index next to the substrate. There are two cases, even and odd number of layers.

Case I: even number (2x) of layers

The resultant multilayer matrix is given by

$$\begin{bmatrix} B \\ C \end{bmatrix} = [L][H][L]\ldots[L][H]\begin{bmatrix} 1 \\ n_m \end{bmatrix}$$

where

$$[L] = \begin{bmatrix} -\varepsilon_L & i/n_L \\ in_L & -\varepsilon_L \end{bmatrix}$$

$$[H] = \begin{bmatrix} -\varepsilon_H & i/n_H \\ in_H & -\varepsilon_H \end{bmatrix}.$$

Then

$$\begin{bmatrix} B \\ C \end{bmatrix} = \{[L][H]\}^x \begin{bmatrix} 1 \\ n_m \end{bmatrix}$$
$$= \begin{bmatrix} -\left(\frac{n_4}{n_L}\right) & -i\left(\frac{\varepsilon_L}{n_H}+\frac{\varepsilon_H}{n_L}\right) \\ -i(n_L\varepsilon_H + n_H\varepsilon_L) & -\left(\frac{n_L}{n_H}\right) \end{bmatrix}^x \begin{bmatrix} 1 \\ n_m \end{bmatrix}$$

$$= \begin{bmatrix} M_{11} & iM_{12} \\ iM_{21} & M_{22} \end{bmatrix} \begin{bmatrix} 1 \\ n_m \end{bmatrix}.$$

Our problem is to find expressions for M_{11}, M_{12}, M_{21} and M_{22}. In the evaluation we neglect all terms of second and higher order in ε. Terms in ε appearing in M_{11} and M_{22} are of second and higher order and therefore

$$M_{11} = (-1)^x \left(\frac{n_H}{n_L}\right)^x$$

$$M_{22} = (-1)^x \left(\frac{n_L}{n_H}\right)^x.$$

M_{12} and M_{21} contain terms of first, third and higher orders in ε. The first-order terms are

$$\begin{aligned} M_{12} = &-\left(\frac{\varepsilon_L}{n_H}+\frac{\varepsilon_H}{n_L}\right)\left(-\frac{n_L}{n_H}\right)^{x-1} + \left(-\frac{n_H}{n_L}\right)\left[-\left(\frac{\varepsilon_L}{n_H}+\frac{\varepsilon_H}{n_L}\right)\right]\left(-\frac{n_L}{n_H}\right)^{x-2} + \dots \\ &+ \left(-\frac{n_H}{n_L}\right)^p \left[-\left(\frac{\varepsilon_L}{n_H}+\frac{\varepsilon_H}{n_L}\right)\right]\left(-\frac{n_L}{n_H}\right)^{x-p-1} + \dots \\ &+ \left(-\frac{n_H}{n_L}\right)^{x-1}\left[-\left(\frac{\varepsilon_L}{n_H}+\frac{\varepsilon_H}{n_L}\right)\right] \\ = &(-1)^x \left(\frac{\varepsilon_L}{n_H}+\frac{\varepsilon_H}{n_L}\right)\left[\left(\frac{n_L}{n_H}\right)^{x-1} + \left(\frac{n_L}{n_H}\right)^{x-3} + \dots + \left(\frac{n_H}{n_L}\right)^{x-1}\right] \\ = &(-1)^x \left(\frac{\varepsilon_L}{n_H}+\frac{\varepsilon_H}{n_L}\right)\left(\frac{n_H}{n_L}\right)^{x-1}\left[\left(\frac{n_L}{n_H}\right)^{2x-2} + \left(\frac{n_L}{n_H}\right)^{2x-4} + \dots + \left(\frac{n_L}{n_H}\right)^2 + 1\right] \\ = &(-1)^x \left(\frac{\varepsilon_L}{n_H}+\frac{\varepsilon_H}{n_L}\right)\left(\frac{n_H}{n_L}\right)^{x-1}\left[1-\left(\frac{n_L}{n_H}\right)^{2x}\right]\left[1-\left(\frac{n_L}{n_H}\right)^2\right]^{-1} \end{aligned}$$

since $(n_L/n_H) < 1$.

Now, provided x is large enough and (n_L/n_H) small enough, we can neglect $(n_L/n_H)^{2x}$ in comparison with 1, and after some adjustment, the expression becomes

$$M_{12} = \frac{(-1)^x n_H n_L (n_H/n_L)^x (\varepsilon_L/n_H + \varepsilon_H/n_L)}{(n_H^2 - n_L^2)}.$$

A similar procedure yields

$$M_{21} = \frac{(-1)^x n_H n_L (n_H/n_L)^x (n_L\varepsilon_H + n_H\varepsilon_L)}{(n_H^2 - n_L^2)}.$$

Case II: odd number $(2x+1)$ *of l*

The resultant matrix is given

$$\begin{bmatrix} B \\ C \end{bmatrix} = [H]$$

$$= [H]$$

which we can denote by

$$\begin{bmatrix} \cdot \\ i \end{bmatrix}$$

and which is simply the previ

Then

$$N_{11} = -\varepsilon_H M_{11} - M_{21}/$$

$$N_{12} = -\varepsilon_H M_{12} + M_{22}/$$

$$N_{21} = n_H M_{11} - \varepsilon_H M_{21}$$

$$N_{22} = -\varepsilon_H M_{22} - n_H M_1$$

where terms in $(n_L/n_H)^x$ are r

Phase shift: case I

We are now able to compute th index of the incident medium

$$\begin{bmatrix} B \\ C \end{bmatrix} =$$

$$=$$

$$\rho = \frac{n_0 B - C}{n_0 B + C} = \frac{n_0}{n_0}$$

$$= \frac{(n_0 M_{11} - n_m}{(n_0 M_{11} + n_m}$$

Case II: odd number $(2x+1)$ *of layers*

The resultant matrix is given by

$$\begin{bmatrix} B \\ C \end{bmatrix} = [H][L][H] \ldots [L][H]\begin{bmatrix} 1 \\ n_m \end{bmatrix}$$

$$= [H]\{[L][H]\}^x \begin{bmatrix} 1 \\ n_m \end{bmatrix}$$

which we can denote by

$$\begin{bmatrix} N_{11} & iN_{12} \\ iN_{21} & N_{22} \end{bmatrix}\begin{bmatrix} 1 \\ n_m \end{bmatrix}$$

and which is simply the previous result multiplied by

$$\begin{bmatrix} -\varepsilon_H & i/n_H \\ in_H & -\varepsilon_H \end{bmatrix}.$$

Then

$$N_{11} = -\varepsilon_H M_{11} - M_{21}/n_H = (-1)^{x+1}\left(\frac{n_H}{n_L}\right)^x \frac{(\varepsilon_L n_H n_L + \varepsilon_H n_H^2)}{(n_H^2 - n_L^2)}$$

$$N_{12} = -\varepsilon_H M_{12} + M_{22}/n_H = (-1)^x \left(\frac{n_L}{n_H}\right)^x \frac{1}{n_H}$$

$$N_{21} = n_H M_{11} - \varepsilon_H M_{21} = (-1)^x \left(\frac{n_H}{n_L}\right)^x n_H$$

$$N_{22} = -\varepsilon_H M_{22} - n_H M_{12} = (-1)^{x+1}\left(\frac{n_H}{n_L}\right)^x \frac{n_H^2 n_L(\varepsilon_L/n_H + \varepsilon_H/n_L)}{(n_H^2 - n_L^2)}$$

where terms in $(n_L/n_H)^x$ are neglected in comparison with $(n_H/n_L)^x$.

Phase shift: case I

We are now able to compute the phase shift on reflection. We take, initially, the index of the incident medium to be n_0. Then

$$\begin{bmatrix} B \\ C \end{bmatrix} = \begin{bmatrix} M_{11} & iM_{12} \\ iM_{21} & M_{22} \end{bmatrix}\begin{bmatrix} 1 \\ n_m \end{bmatrix}$$

$$= \begin{bmatrix} M_{11} + in_m M_{12} \\ n_m M_{22} + iM_{21} \end{bmatrix}$$

$$\rho = \frac{n_0 B - C}{n_0 B + C} = \frac{n_0(M_{11} + in_m M_{12}) - n_m M_{22} - iM_{21}}{n_0(M_{11} + in_m M_{12}) + n_m M_{22} + iM_{21}}$$

$$= \frac{(n_0 M_{11} - n_m M_{22}) + i(n_0 n_m M_{12} - M_{21})}{(n_0 M_{11} + n_m M_{22}) + i(n_0 n_m M_{12} + M_{21})} \qquad (7.19)$$

$$\tan\phi = \frac{2n_0 n_m^2 M_{12} M_{22} - 2n_0 M_{11} M_{21}}{n_0^2 M_{11}^2 - n_m^2 M_{22}^2 + n_0^2 n_m^2 M_{12}^2 - M_{21}^2}.$$

Inserting the appropriate expressions and once again neglecting terms of second and higher order in ε and terms in $(n_L/n_H)^x$, we obtain for ϕ

$$\tan\phi = \frac{-2n_H n_L(n_L\varepsilon_H + n_H\varepsilon_L)}{n_0(n_H^2 - n_L^2)} \tag{7.20}$$

(for $LH \ldots LHLH\,|\,n_m$).

Phase shift: case II

ρ is given by an expression similar to (7.19), in which M is replaced by N. Then, following the same procedure as for case I we arrive at

$$\tan\phi = \frac{-2n_0(\varepsilon_L n_L + \varepsilon_H n_H)}{(n_H^2 - n_L^2)} \tag{7.21}$$

(for $HLH \ldots LHLH\,|\,n_m$).

Equations (7.20) and (7.21) are in a general form which we will make use of later. For our present purposes we can introduce some slight simplification.

$$\delta = \frac{2\pi nd}{\lambda} = 2\pi ndv = 2\pi nd\nu_0(\nu/\nu_0) = (\pi/2)g$$

so that

$$\varepsilon_H = \varepsilon_L = (\pi/2)g - \pi/2 = (\pi/2)(g-1).$$

Also, when we consider the construction of the Fabry–Perot filters we see that the incident medium in case I will be a high-index spacer layer and in case II a low index spacer. Thus, for Fabry–Perot filters,

$$\tan\phi = \frac{-\pi n_L}{(n_H - n_L)}(g-1)$$

for both case I and case II.

Now, ϕ is nearly π or 0. Then

$$\frac{d\phi}{dg} = \frac{-\pi n_L}{(n_H - n_L)}$$

which is the result obtained by Seeley. This can then be inserted in equation (7.18) to give

$$\frac{\Delta\nu_h}{\nu_0} = \frac{\Delta\lambda_h}{\lambda_0} = \frac{2}{m\pi F^{1/2}}\left(\frac{n_H - n_L}{n_H - n_L + n_L/m}\right).$$

Then the expressions for the halfwidth of all-dielectric Fabry–Perot filters of mth order become

High-index spacer:

$$\left(\frac{\Delta\lambda_h}{\lambda_0}\right)_H =$$

Low-index spacer:

$$\left(\frac{\Delta\lambda_h}{\lambda_0}\right)_L =$$

which are simply the earlier re
$+ n_L/m$). It should be noted
stacks and mth-order spacer.
the closer the two indices are i
common visible and near inf
factor for first-order spacers i
as zinc sulphide and lead tellu
show the characteristics of
filters.

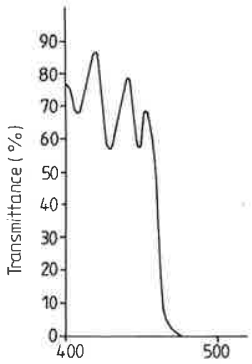


Figure 7.8 Measured transmi
unsuppressed sidebands. Zinc sul
(Courtesy of Sir Howard Grubb

Since the all-dielectric mult
only, sidebands of transmissic
applications must be suppres
very easily by adding to the
available in the form of p
manufacturers. Unfortunately
absorption filters and the rath
considerably to reduce the pe

$\frac{{}_1 M_{21}}{{}_2 - M_{21}^2}$.

once again neglecting terms of $n_L/n_H)^x$, we obtain for ϕ

$\frac{{}_H + n_H \varepsilon_L)}{n_L^2)}$ (7.20)

which M is replaced by N. Then, ve arrive at

$\frac{+ \varepsilon_H n_H)}{{}^2_L)}$ (7.21)

form which we will make use of duce some slight simplification.

$(\nu/\nu_0) = (\pi/2)g$

$)\,(g-1)$.

: Fabry–Perot filters we see that dex spacer layer and in case II a :rs,

$-1)$

ın then be inserted in equation

$\frac{{}_H - n_L}{n_L + n_L/m}\Big)$.

ielectric Fabry–Perot filters of

High-index spacer:

$$\left(\frac{\Delta\lambda_h}{\lambda_0}\right)_H = \frac{4n_m n_L^{2x}}{m\pi n_H^{2x+1}} \frac{(n_H - n_L)}{(n_H - n_L + n_L/m)} \tag{7.22}$$

Low-index spacer:

$$\left(\frac{\Delta\lambda_h}{\lambda_0}\right)_L = \frac{4n_m n_L^{2x-1}}{m\pi n_H^{2x}} \frac{(n_H - n_L)}{(n_H - n_L + n_L/m)} \tag{7.23}$$

which are simply the earlier results multiplied by the factor $(n_H - n_L)/(n_H - n_L + n_L/m)$. It should be noted that these results are for first-order reflecting stacks and mth-order spacer. Clearly the effect of the phase is much greater the closer the two indices are in value and the lower the spacer order m. For the common visible and near infrared materials, zinc sulphide and cryolite, the factor for first-order spacers is equal to 0.43, while for infrared materials such as zinc sulphide and lead telluride it is greater, around 0.57. Figures 7.8 and 7.9 show the characteristics of typical all-dielectric narrowband Fabry–Perot filters.

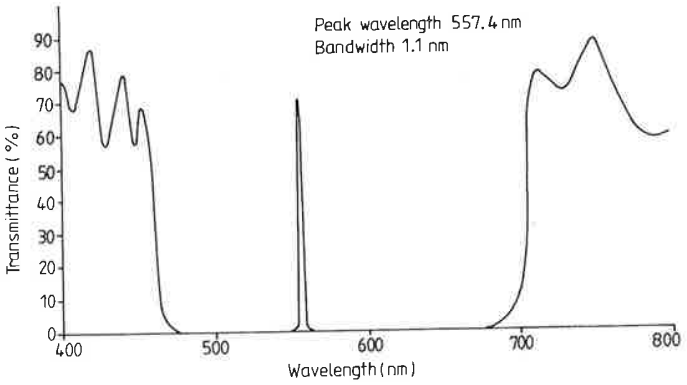


Figure 7.8 Measured transmittance of a narrowband all-dielectric filter with unsuppressed sidebands. Zinc sulphide and cryolite were the thin-film materials used. (Courtesy of Sir Howard Grubb, Parsons & Co Ltd.)

Since the all-dielectric multilayer reflector is effective over a limited range only, sidebands of transmission appear on either side of the peak and in most applications must be suppressed. The shortwave sidebands can be removed very easily by adding to the filter a longwave-pass absorption filter, readily available in the form of polished glass disks from a large number of manufacturers. Unfortunately, it is not nearly as easy to obtain shortwave-pass absorption filters and the rather shallow edges of those which are available tend considerably to reduce the peak transmission of the filter if the sidebands are

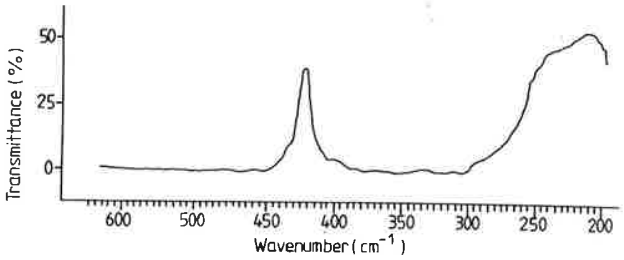


Figure 7.9 Measured transmittance of a Fabry–Perot filter for the far infrared. Design: Air|$LHLHHLH$|Ge with H indicating a quarter-wave of germanium and L of caesium iodide. The rear surface of the substrate is unbloomed so that the effective transmission of the filter is 50%. (Courtesy of Sir Howard Grubb, Parsons & Co Ltd.)

effectively suppressed. The best solution to this problem is not to use an absorption type of filter at all, but to employ as a blocking filter a metal–dielectric filter of the type already discussed or of the multiple cavity type to be considered shortly. Because metal–dielectric filters used in the first order do not have longwave sidebands, they are very successful in this application. The metal–dielectric blocking filter can, in fact, be deposited over the all-dielectric filter in the same evaporation run provided that the layers are monitored using the narrowband filter itself as the test glass—this is known as direct monitoring—but more frequently a completely separate metal–dielectric filter is used. The various components which go to make up the final filter are cemented together in one assembly.

Before we leave the Fabry–Perot filters we can examine the effects of absorption losses in the layers in a manner similar to that already employed in chapter 5, where we were concerned with quarter-wave stacks. The problem has been investigated by many workers. The account which follows relies heavily on the work of Hemingway and Lissberger[5], but with slight differences.

We apply the method of chapter 5 directly. There, we recall, we showed that the loss in a weakly absorbing multilayer was given by

$$A = (1-R)\sum \mathscr{A}$$

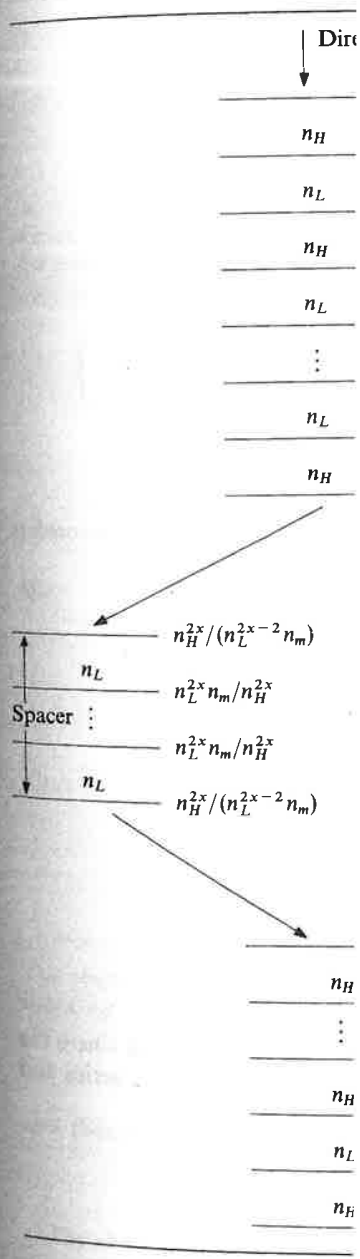
where, for *quarter-waves*,

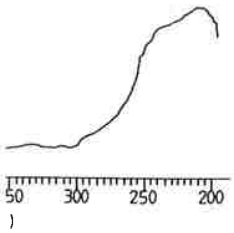
$$\mathscr{A} = \beta\left(\frac{n}{y_e} + \frac{y_e}{n}\right)$$

$$\beta = \frac{2\pi kd}{\lambda} = \frac{2\pi nd}{\lambda}\frac{k}{n} = \frac{\pi}{2}\frac{k}{n}.$$

y_e is the admittance of the structure on the emergent side of the layer, in free space units, $n - ik$ is the refractive index of the layer and d is the geometrical thickness. For quarter-waves, $nd = \pi/4$.

The scheme is shown in table 7.2 where the admittance y_e is given at each





-Perot filter for the far infrared. arter-wave of germanium and L of s unbloomed so that the effective ɔward Grubb, Parsons & Co Ltd.)

his problem is not to use an ıploy as a blocking filter a ıssed or of the multiple cavity lielectric filters used in the first y are very successful in this · can, in fact, be deposited over ı run provided that the layers tself as the test glass—this is uently a completely separate ɔnents which go to make up the bly.

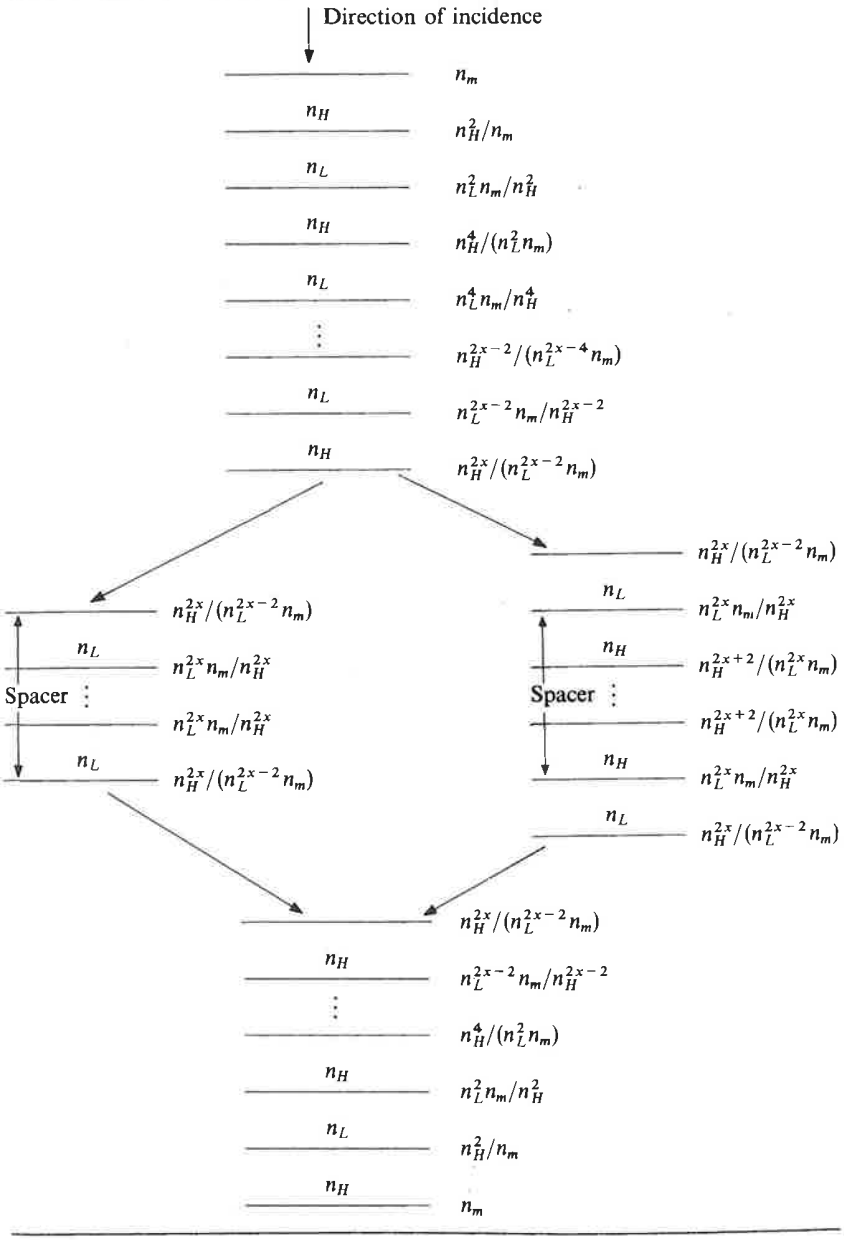
e can examine the effects of lar to that already employed in r-wave stacks. The problem has nt which follows relies heavily ɔut with slight differences. here, we recall, we showed that given by

$$= \frac{\pi}{2}\frac{k}{n}.$$

ergent side of the layer, in free layer and d is the geometrical

admittance y_e is given at each

Table 7.2



interface and where alternative schemes for either high- or low-index spacers are included. The reflecting stacks are assumed to begin with high-index layers of which there are x per reflector, not counting the spacer.

We consider the case of low-index spacers first.

$$\begin{aligned}\sum\mathscr{A} = {} & \beta_H\left(\frac{n_m}{n_H}+\frac{n_H}{n_m}\right)+\beta_L\left(\frac{n_H^2}{n_L n_m}+\frac{n_L n_m}{n_H^2}\right)\\ & +\beta_H\left(\frac{n_L^2 n_m}{n_H^3}+\frac{n_H^3}{n_L^2 n_m}\right)+\beta_L\left(\frac{n_H^4}{n_L^3 n_m}+\frac{n_L^3 n_m}{n_H^4}\right)+\ \ldots\\ & +\beta_L\left(\frac{n_H^{2x-2}}{n_L^{2x-3}n_m}+\frac{n_L^{2x-3}n_m}{n_H^{2x-2}}\right)+\beta_H\left(\frac{n_H^{2x-1}}{n_L^{2x-2}n_m}+\frac{n_L^{2x-2}n_m}{n_H^{2x-1}}\right)\\ & +m\left[\beta_L\left(\frac{n_H^{2x}}{n_L^{2x-1}n_m}+\frac{n_L^{2x-1}n_m}{n_H^{2x}}\right)+\beta_L\left(\frac{n_L^{2x-1}n_m}{n_H^{2x}}+\frac{n_H^{2x}}{n_L^{2x-1}n_m}\right)\right]\\ & +\beta_H\left(\frac{n_H^{2x-1}}{n_L^{2x-2}n_m}+\frac{n_L^{2x-2}n_m}{n_H^{2x-1}}\right)+\ \ldots+\beta_H\left(\frac{n_H}{n_m}+\frac{n_m}{n_H}\right)\end{aligned}$$

where the final set of terms is a repeat of the first and where the spacer consists of $2m$ quarter-waves. Rearranging, we find

$$\begin{aligned}\sum\mathscr{A} = {} & 2\beta_H\left(\frac{n_m}{n_H}+\frac{n_L^2 n_m}{n_H^3}+\frac{n_L^4 n_m}{n_H^5}+\ \ldots\ +\frac{n_L^{2x-2}n_m}{n_H^{2x-1}}\right)\\ & +2\beta_H\left(\frac{n_H}{n_m}+\frac{n_H^3}{n_L^2 n_m}+\frac{n_H^5}{n_L^4 n_m}+\ \ldots\ +\frac{n_H^{2x-1}}{n_L^{2x-2}n_m}\right)\\ & +2\beta_L\left(\frac{n_L n_m}{n_H^2}+\frac{n_L^3 n_m}{n_H^4}+\ \ldots\ +\frac{n_L^{2x-3}n_m}{n_H^{2x-2}}\right)\\ & +2\beta_L\left(\frac{n_H^2}{n_L n_m}+\frac{n_H^4}{n_L^3 n_m}+\ \ldots\ +\frac{n_H^{2x-2}}{n_L^{2x-3}n_m}\right)\\ & +2m\beta_L\left(\frac{n_H^{2x}}{n_L^{2x-1}n_m}\,\frac{n_L^{2x-1}n_m}{n_H^{2x}}\right)\end{aligned}$$

where we have combined similar terms due to the two mirrors and where the final term is due to the spacer. The first four terms are geometric series and therefore, since $(n_L/n_H) < 1$,

$$\begin{aligned}\sum\mathscr{A} = {} & 2\beta_H\frac{n_m}{n_H}\frac{[1-(n_L/n_H)^{2x}]}{[1-(n_L/n_H)^2]}\\ & +2\beta_H\frac{n_H^{2x-1}}{n_L^{2x-2}n_m}\frac{[1-(n_L/n_H)^{2x-2}]}{[1-(n_L/n_H)^2]}\end{aligned}$$

$$+ 2\beta_L \frac{n_L n_m}{n_H^2} \frac{[1-(n_L/n_H)^{2x-2}]}{[1-(n_L/n_H)^2]}$$

$$+ 2\beta_L \frac{n_H^{2x-2}}{n_L^{2x-3} n_m} \frac{[1-(n_L/n_H)^{2x-2}]}{[1-(n_L/n_H)^2]}$$

$$+ 2m\beta_L \left[\frac{n_H^{2x}}{n_L^{2x-1} n_m} + \frac{n_L^{2x-1} n_m}{n_H^{2x}} \right].$$

(n_L/n_H) will usually be rather less than unity and x will normally be large and so we can make the usual approximation and neglect terms such as $(n_L/n_H)^{2x}$ in the numerators and also those terms which have (n_m/n_H) as a factor compared with $(n_L/n_m)\,(n_H/n_L)^{2x-1}$ etc. Then the expression simplifies to

$$\sum \mathscr{A} = 2\beta_H \frac{n_H^{2x-1}}{n_L^{2x-2} n_m} \frac{1}{[1+(n_L/n_H)^2]}$$

$$+ 2\beta_L \frac{n_H^{2x-2}}{n_L^{2x-3} n_m} \frac{1}{[1+(n_L/n_H)^2]}$$

$$+ 2m\beta_L \frac{n_H^{2x}}{n_L^{2x-1} n_m}.$$

But

$$\beta_H = \frac{2\pi n_H d}{\lambda} \frac{k_H}{n_H} = \frac{\pi}{2} \frac{k_H}{n_H}$$

$$\beta_L = \frac{\pi}{2} \frac{k_L}{n_L}.$$

Thus

$$\sum \mathscr{A} = \frac{\pi k_H (n_H^{2x}/n_m n_L^{2x-2}) + \pi k_L (n_H^{2x}/n_m n_L^{2x-2})}{(n_H^2 - n_L^2)} + \frac{\pi m k_L n_H^{2x}}{n_L^{2x} n_m}$$

$$= \frac{\pi n_H^{2x}}{n_m n_L^{2x}} \left(\frac{n_L^2 k_H + n_L^2 k_L}{(n_H^2 - n_L^2)} + m k_L \right).$$

The absorption is then given by $A = (1-R)\sum \mathscr{A}$. If the incident medium has index n_0, then, since the terminating admittance in table 7.2 is n_m,

$$R = \left(\frac{n_0 - n_m}{n_0 + n_m} \right)^2$$

and therefore

$$(1-R) = \frac{4 n_0 n_m}{(n_0 + n_m)^2}.$$

The above expression for $\sum \mathscr{A}$ should, therefore, be multiplied by the factor $4n_0 n_m/(n_0+n_m)^2$ to yield the absorption. However, the filters should be

designed so that they are reasonably well matched into the incident medium and therefore this factor will be unity, or sufficiently near unity. The absorption is then given by $\sum \mathscr{A}$. That is:

$$A = \frac{\pi n_H^{2x}}{n_m n_L^{2x}} \left(\frac{n_L^2 k_H + n_L^2 k_L}{(n_H^2 - n_L^2)} + m k_L \right) \tag{7.24}$$

for low-index spacers.

For high-index spacers we work through a similar scheme and, with the same approximations, we arrive at

$$A = \frac{\pi n_H^{2x}}{n_m n_L^{2x}} \left(\frac{n_L^2 k_H + n_H^2 k_L}{(n_H^2 - n_L^2)} + m k_H \right) \tag{7.25}$$

for high-index spacers.

It should be noted that, since x is the number of high-index layers, the filter represented by equation (7.25) will be narrower than that represented by equation (7.24) for equal x.

A useful set of alternative expressions can be obtained if we substitute equations (7.22) and (7.23) into equations (7.24) and (7.25) to give:

High-index spacer

$$A = 4 \frac{\lambda_0}{\Delta\lambda_h} \frac{\{k_L + k_H [m + (1-m)(n_L/n_H)^2]\}}{(n_H + n_L)[m + (1-m)(n_L/n_H)]} \tag{7.26}$$

Low-index spacer

$$A = 4 \frac{\lambda_0}{\Delta\lambda_h} \frac{\{k_L (n_H/n_L)[m + (1-m)(n_L/n_H)^2] + (n_L/n_H) k_H\}}{(n_H + n_L)[m + (1-m)(n_L/n_H)]}. \tag{7.27}$$

Figure 7.10 shows the value of A plotted for Fabry–Perot filters with $n_H = 2.35$ and $n_L = 1.35$, typical of zinc sulphide and cryolite. $(\lambda_0/\Delta\lambda_h)$ is taken as 100 and k_H and k_L as either zero or 0.0001. The effect of other values of $(\lambda_0/\Delta\lambda_h)$ or k can be estimated by multiplying by an appropriate factor. The approximations are reasonable for $k(\lambda_0/\Delta\lambda_h)$ less than around 0.1.

It is difficult to draw any general conclusions from figure 7.10 because the results depend on the relative magnitudes of k_H and k_L. However, except in the case of very low k_L, the high-index spacer is to be preferred. There are very good reasons connected with performance when tilted, with energy grasp and with the manufacture of filters, for choosing high- rather than low-index spacers.

In the visible and near infrared regions of the spectrum, materials such as zinc sulphide and cryolite are capable of halfwidths of less than 0.1 nm with useful peak transmittance. Uniformity is, however, a major difficulty for filters of such narrow bandwidths. At the 90 %-of-peak points, the Fabry–Perot filter has a width which is one third of the halfwidth. It is a good guide that the uniformity of the filter should be such that the peak wavelength does not vary

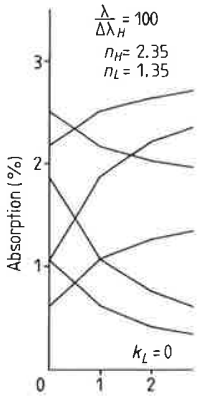


Figure 7.10 The value (e
the order number m, of Fa
coefficients k_H and k_L o
multiplying by an approp
derived from equations (

by more than one thir
This means that the effe
is kept within some
transmittance to less
expressions derived la
collimated incident ligl
implies a variation of
thickness, a very severe
0.3–0.5 nm are less der
considerable care is tak
etalon filters now to b

The solid etalon filter

A solid etalon filter, or,
high-order Fabry–Perc
worked plate or a cleav
side of the spacer in
substrate. The probler
rowband filters are av
increased scattering los
thin-film spacers. The s
than the conventional
ing difficulties are com

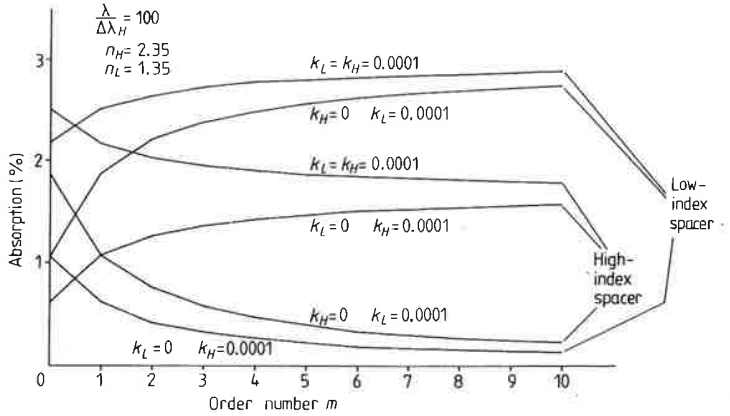


Figure 7.10 The value (expressed as a percentage) of the absorptance, as a function of the order number m, of Fabry–Perot filters with $\lambda_0/\Delta\lambda_h$ of 100 and values of extinction coefficients k_H and k_L of 0.0001 or zero. Other values can be accommodated by multiplying by an appropriate factor. n_H is taken as 2.35 and n_L as 1.35. The results are derived from equations (7.26) and (7.27).

by more than one third of the halfwidth over the entire surface of the filter. This means that the effective increase in halfwidth due to the lack of uniformity is kept within some $4\frac{1}{2}\%$ of the halfwidth and the reduction in peak transmittance to less than 3% (these figures can be calculated using the expressions derived later for assessing the performance of filters in uncollimated incident light). For filters of less than 0.1 nm halfwidth this rule implies a variation of not more than 0.03 nm or 0.006% in terms of layer thickness, a very severe requirement even for quite small filters. Halfwidths of 0.3–0.5 nm are less demanding and can be produced more readily provided considerable care is taken. For narrower filters use is often made of the solid etalon filters now to be described.

The solid etalon filter

A solid etalon filter, or, as it is sometimes called, a solid spacer filter, is a very high-order Fabry–Perot filter in which the spacer consists of an optically worked plate or a cleaved crystal. Thin-film reflectors are deposited on either side of the spacer in the normal way, so that the spacer also acts as the substrate. The problems of uniformity which exist with all-thin-film narrowband filters are avoided and the thick spacer does not suffer from the increased scattering losses which always seem to accompany the higher-order thin-film spacers. The solid etalon filter is very much more robust and stable than the conventional air-spaced Fabry–Perot etalon, while the manufacturing difficulties are comparable. The high order of the spacer implies a small

interval between orders and a conventional thin-film narrowband filter must be used in series with it to eliminate the unwanted orders.

An early account of the use of mica for the construction of filters of this type is that of Dobrowolski[6] who credits Billings with being the first to use mica in this way, achieving halfwidths of 0.3 nm. Dobrowolski obtained rather narrower pass bands and his is the first complete account of the technique. Mica can be cleaved readily to form thin sheets with flat parallel surfaces, but there is a complication due to the natural birefringence of mica which means that the position of the pass band depends on the plane of polarisation. This splitting of the pass band can be avoided by arranging the thickness of the mica such that it is a half-wave plate, or multiple half-wave, at the required wavelength. If the two refractive indices are n_0 and n_e, this implies

$$\frac{2\pi(n_0 - n_e)d}{\lambda} = p\pi \qquad p = 0, \pm 1, \pm 2, \ldots$$

The order of the spacer will then be given by

$$m = \frac{n_0 p}{(n_0 - n_e)} \qquad \text{or} \qquad \frac{n_e p}{(n_0 - n_e)}$$

depending on the plane of polarisation. The difference between these two values is p, but, since p is small, the bandwidth will be virtually identical. The separation of orders for large m is given approximately by λ/m. Dobrowolski found that the maximum order separation, corresponding to $p = 1$, was given by 1.64 nm at 546.1 nm. With such spacers, around 60 μm thick, filters with halfwidths around 0.1 nm, the narrowest 0.085 nm, were constructed. Peak transmission ranged up to 50% for the narrower filters and up to 80% for slightly broader ones with around 0.3 nm halfwidth.

More recent work on solid etalon filters has concentrated on the use of optically worked materials as spacers. These must be ground and polished so that the faces have the necessary flatness and parallelism. The most complete account so far of the production of such filters is by Austin[7]. Fused silica spacers as thin as 50 μm have been produced with the necessary parallelism for halfwidths as narrow as 0.1 nm in the visible region, while thicker discs can give bandwidths as narrow as 0.005 nm. A 50 μm fused silica spacer gives an interval between orders of around 1.4 nm in the visible region which allows the suppression of unwanted orders to be fairly readily achievable by conventional thin-film narrowband filters.

The process of optical working tends to produce an error in parallelism over the surface of the spacer which is ultimately independent of the thickness of the spacer. Let us denote the total range of spacer thickness due to this lack of parallelism and to any deviation from flatness by Δd. This variation in spacer thickness causes the peak wavelength of the filter to vary. We can take an absolute limit for these variations as half the bandwidth of the filter. Then the resultant halfwidth will be increased by just over 10% and the peak

transmittance reduced by just over 7% (once again using the expressions which we will shortly establish for filter performance in uncollimated light). We can write

$$\Delta\lambda_0/\lambda_0 = \Delta D/D = \Delta d/d \leqslant 0.5\Delta\lambda_h/\lambda_0$$

where D is the optical thickness nd of the spacer, $\Delta\lambda_0$ is the error in peak wavelength and $\Delta\lambda_h$ is the halfwidth. But

$$\text{Resolving power} = \lambda_0/\Delta\lambda_h = m\mathscr{F}$$

and hence, since

$$D = m\lambda_0/2$$

$$\mathscr{F} \leqslant \frac{0.25\lambda_0}{\Delta D}.$$

Now the attainable ΔD in the visible region is of the order of $\lambda/100$ and this means that the limiting finesse is around 25, independent of the spacer thickness. High resolving power then has to be achieved by the order number m which determines both the spacer thickness $D = m\lambda_0/2$ and the interval between orders λ_0/m. For a halfwidth of 0.01 nm at, say, 500 nm the resolving power is 50 000. The finesse of 25 implies an order number of 2000, a spacer optical thickness of 500 μm and an interval between orders of 0.25 nm. This very restricted range between orders means that it is very difficult to carry out sideband blocking by a thin-film filter directly. Instead, a broader solid etalon filter can be used with its correspondingly greater interval between orders. It, in its turn, can be suppressed by a thin-film filter. For a halfwidth of 0.1 nm, a spacer optical thickness of 50 μm is required which gives an interval between orders of 2.5 nm.

The temperature coefficient of peak wavelength change of solid etalon filters with fused silica spacers is 0.005 nm °C^{-1} and the filters may be finely tuned by altering this temperature.

Candille and Saurel[8] have used Mylar foil as the spacer. Their filters were strictly of the multiple cavity type described later in this chapter. The Mylar acted as substrate and a high-order spacer. One of the reflectors included a low-order Fabry–Perot filter which served both as blocking filter to eliminate the additional unwanted orders of the Mylar section and as an additional cavity to steepen the sides of the pass band. The position of the pass band could be altered by varying the tension in the Mylar. The filters were not as narrow as the other solid etalon filters which have been mentioned, halfwidths of 0.8–1.0 nm being obtained.

Solid etalon filters have also been constructed for the infrared. Smith and Pidgeon[9] used a polished slab of germanium some 780 μm thick working at around 700 cm^{-1} in the 400th order. Both faces were coated with a quarter-wave of zinc sulphide followed by a quarter-wave of lead telluride to give a reflectance of 62%, a fringe halfwidth of 0.1 cm^{-1} and an interval between orders of 1.6 cm^{-1}. This particular arrangement was designed so that the lines

in the R-branch of the CO_2 spectrum, which are spaced 1.6 cm^{-1} apart at around 14.5 μm, should be exactly matched by a number of adjacent orders. Order sorting was not, therefore, a problem.

Roche and Title[10] have reported a range of solid etalon filters for the infrared. These filters are some 13 mm in diameter, have resolving powers in the region of 3×10^4 and the techniques used for their construction are as reported by Austin[7]. For wavelengths equal to or shorter than 3.5 μm, fused silica spacers are quite satisfactory. For longer wavelengths Yttralox, a combination of yttrium and thorium oxides, was found most satisfactory. With this material, solid etalon filters were produced which at 3.334 μm had halfwidths as low as 0.2 nm and at 4.62 μm, 0.8 nm. At these wavelengths, the attainable finesse was 30–40 and the current limit to the halfwidth which can be achieved is the permissible interval between orders which determines the arrangement of subsidiary blocking filters.

The effect of varying the angle of incidence

As we have seen with other types of thin-film assembly the performance of the all-dielectric Fabry–Perot varies with angle of incidence, and this effect is particularly important when considering, for instance, the allowable focal ratio of the pencil being passed by the filter or the maximum tilt angle in any application. The variation with angle of incidence is not altogether a bad thing because the effect can be used to tune filters which would otherwise be off the desired wavelength—very important from the manufacturer's point of view because it enables him to ease a little the otherwise almost impossibly tight production tolerances.

The effect of tilting has been studied by a number of workers, particularly by Dufour and Herpin[11], Lissberger[12], Lissberger and Wilcock[13] and Pidgeon and Smith[14]. For our present purposes we follow Pidgeon and Smith since their results are in a slightly more suitable form.

Simple tilts in collimated light

The phase thickness of a thin film at oblique incidence is

$$\delta = 2\pi nd \cos\theta/\lambda$$

which can be interpreted as an apparent optical thickness of $nd\cos\theta$ which varies with angle of incidence so that layers seem thinner when tilted. Although the optical admittance changes with tilts, in narrowband filters the predominant effect is the apparent change in thickness which moves the filter pass band to shorter wavelengths.

For an ideal Fabry–Perot filter with spacer layer index n^*, where the reflectors have constant phase shift of zero or π regardless of the angle of incidence or wavelength, we can write for the position of peak wavelength in the mth order

$$2\pi n^* d \cos\theta/\lambda = m\pi$$

i.e.

$$(2\pi n^* d/\lambda_0)\ \ g\cos\theta = m\pi$$

i.e.

$$g\cos\theta = 1$$

$$\Delta g = \left(\frac{1}{\cos\theta} - 1\right).$$

If the angle of incidence is θ_i in air then

$$\theta = \sin^{-1}(\sin\theta_i/n^*)$$

and Δg is given in terms of θ_i and n^*. The effect of tilting, then, in this ideal filter can be estimated simply from a knowledge of the index of the spacer and the angle of incidence. For small angles of incidence, the shift is given by

$$\Delta g = \Delta\nu/\nu_0 = \Delta\lambda/\lambda_0 = \theta_i^2/2n^{*2}. \tag{7.28}$$

The index of the spacer n^* determines its sensitivity to tilt: the higher the index, the less the filter is affected.

In the case of a real filter, the reflectors are also affected by the tilting and so the calculation of the shift in peak wavelength is more involved. It has, however, been shown by Pidgeon and Smith that the shift is similar to that which would have been obtained from an ideal filter with spacer index n^*, intermediate between the high and low indices of the layers of the filter. n^* is known as the effective index. This concept of the effective index holds good for quite high angles of incidence, up to 20° or 30° or even higher, depending on the indices of the layers making up the filter.

We can estimate the effective index for the filter by a technique similar to that already used for metal–dielectrics (equation (7.3)). We retain our assumption of small angle of incidence and small changes in g around the value which corresponds to the peak at normal incidence.

The peak position is given, as before, by

$$\sin^2[(2\pi nd\cos\theta/\lambda) - \phi] = 0 \tag{7.29}$$

with, at normal incidence,

$$\sin^2[(2\pi nd/\lambda_0) - \phi_0] = 0. \tag{7.30}$$

Now ϕ_0 is 0 or π and so equation (7.30) is satisfied by

$$2\pi nd/\lambda_0 = m\pi \qquad m = 0, 1, 2 \ldots$$

The analysis is once again easier in terms of g ($= \lambda_0/\lambda = \nu/\nu_0$). Equation (7.29) becomes

$$\sin^2[(2\pi nd/\lambda_0)g\cos\theta - \phi_0 - \Delta\phi] = 0.$$

We write

$$g = 1 + \Delta g \qquad \text{and} \qquad \cos\theta \simeq 1 - \theta^2/2.$$

However, we should work in terms of θ_i, the external angle of incidence, which

we assume is referred to free space (if not, then we make the appropriate correction). Then

$$n \sin\theta = n_i \sin\theta_i = \sin\theta_i$$

and, using equation (7.31),

$$\sin^2 [(2\pi nd/\lambda_0) - \phi_0 + m\pi\Delta g - (m\pi\theta_i^2/2n^2) - \Delta\phi] = 0$$

is the condition for the new peak position. This requires

$$m\pi\Delta g - (m\pi\theta_i^2/2n^2) - \Delta\phi = 0. \tag{7.32}$$

Now $\Delta\phi$ is a function of θ and Δg and to evaluate it we return to equations (7.20) and (7.21). The layers in the reflectors are all quarter-waves and so ε is given by

$$\pi/2 + \varepsilon = (2\pi nd/\lambda_0) g \cos\theta = (\pi/2)(1+\Delta g)(1-\theta^2/2)$$

but

$$\theta = \theta_i/n$$

so that

$$\varepsilon = (\pi/2)\Delta g - \pi\theta_i^2/4n^2$$

with n being either n_L or n_H for ε_L or ε_H respectively.

At this stage we are forced to consider high-index and low-index spacers separately.

Case I: high-index spacers

From equation (7.20) we have, inserting n_H for n_0,

$$\begin{aligned}\Delta\phi &= -\frac{2n_L^2}{(n_H^2 - n_L^2)}\varepsilon_H - \frac{2n_H n_L}{(n_H^2 - n_L^2)}\varepsilon_L \\ &= \frac{2n_L^2}{(n_H^2 - n_L^2)}\left(\frac{\pi}{2}\Delta g - \frac{\pi\theta_i^2}{4n_H^2}\right) - \frac{2n_H n_L}{(n_H^2 - n_L^2)}\left(\frac{\pi}{2}\Delta g - \frac{\pi\theta_i^2}{4n_L^2}\right) \\ &= -\frac{\pi n_L}{(n_H - n_L)}\Delta g + \frac{\pi}{2}\frac{(n_L^2 - n_L n_H + n_H^2)}{n_H^2 n_L (n_H - n_L)}\theta_i^2\end{aligned}$$

and equation (7.32) becomes

$$m\pi\Delta g - \frac{m\pi\theta_i^2}{2n_H^2} + \frac{\pi n_L \Delta g}{(n_H - n_L)} - \frac{\pi}{2}\frac{(n_L^2 - n_L n_H + n_H^2)}{n_H^2 n_L (n_H - n_L)}\theta_i^2 = 0$$

giving, after some manipulation and simplification

$$\Delta g = \frac{1}{n_H^2}\frac{[(m-1) - (m-1)(n_L/n_H) + (n_H/n_L)]}{[m - (m-1)(n_L/n_H)]}\left(\frac{\theta_i^2}{2}\right).$$

But, comparing the expression with equation (7.28) we find

$$n^{*2} = \frac{n_H^2 [m - (m-1)(n_L/n_H)]}{[(m-1) - (m-1)(n_L/n_H) + (n_H/n_L)]}$$

or

$$n^* = n_H \left(\frac{m-(m-1)(n_L/n_H)}{(m-1)-(m-1)(n_L/n_H)+(n_H/n_L)} \right)^{1/2}. \tag{7.33}$$

For first-order filters

$$n^* = (n_H n_L)^{1/2} \tag{7.34}$$

which is the result obtained by Pidgeon and Smith. As $m \to \infty$ then $n^* \to n_H$, as we would expect.

Case II: low-index spacer

The analysis is exactly as for case I except that equation (7.21) is used and the n in equation (7.32) becomes n_L.

$$n^* = n_L \left(\frac{m-(m-1)(n_L/n_H)}{m-m(n_L/n_H)+(n_L/n_H)^2} \right)^{1/2}. \tag{7.35}$$

For first-order filters

$$n^* = \frac{n_L}{[1-(n_L/n_H)+(n_L/n_H)^2]^{1/2}} \tag{7.36}$$

which is, again, the expression given by Pidgeon and Smith and we note again that as $m \to \infty$ then $n^* \to n_L$.

Typical curves showing how the effective index n^* varies with order number for both low- and high-index spacers are given in figure 7.11.

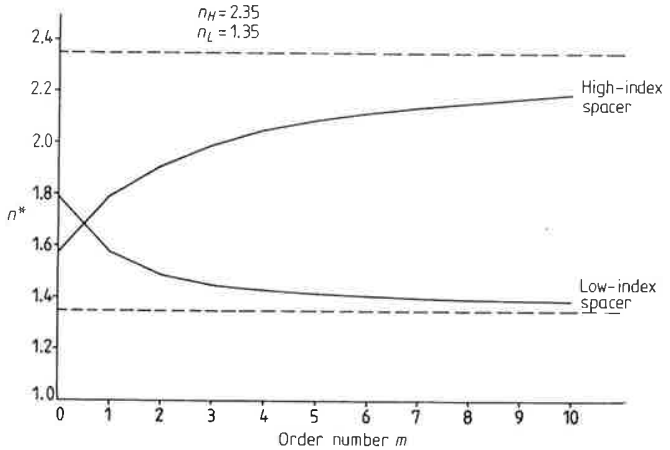


Figure 7.11 The effective index n^* plotted against order number m for Fabry–Perot filters constructed of materials such as zinc sulphide, $n = 2.35$, and cryolite, $n = 1.35$. The results were calculated from expressions (7.35) and (7.36).

Pidgeon and Smith made experimental measurements on narrowband filters for the infrared. The designs in question were

$$(a) \qquad L\,|\,\mathrm{Ge}\,|\,LHLH\ LL\ HLH\,|\,\mathrm{Air}$$

$$(b) \qquad L\,|\,\mathrm{Ge}\,|\,LHLHL\ HH\ LHLH\,|\,\mathrm{Air}$$

where H represents a quarter-wave thickness of lead telluride and L of zinc sulphide, and where the peak wavelength was in the vicinity of 15 μm. Calculations of shift were carried out by the approximate method using n^* and by the full matrix method without approximations. The results using n^* matched the accurate calculations up to angles of incidence of 40° to an accuracy representing $\pm 2\%$ change in n^*. The experimental points showed good agreement with the theoretical estimates. Some of the results are shown in figures 7.12 and 7.13.

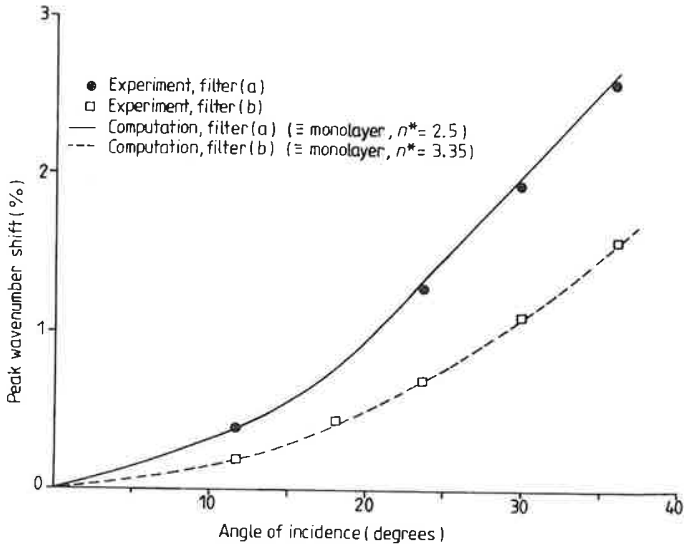


Figure 7.12 The shift of peak wavenumber with scanning angle for two Fabry–Perot filters in collimated light. In both cases the monolayer curves fit the computed curves to $\pm 2\%$ in n. (After Pidgeon and Smith[14].)

The angle of incidence may be in a medium other than free space, in which case equation (7.28) becomes

$$\Delta g = \Delta\lambda_0/\lambda = \Delta\nu_0/\nu = \tfrac{1}{2}(n_i\theta_i/n^*)^2 \tag{7.37}$$

where θ_i is measured in radians.

If θ_i is measured in degrees, then

$$\Delta g = \Delta\lambda_0/\lambda_0 = \Delta\nu_0/\nu_0 = 1.5 \times 10^{-4}\,(n_i/n^*)^2\,\theta_i^2. \tag{7.38}$$

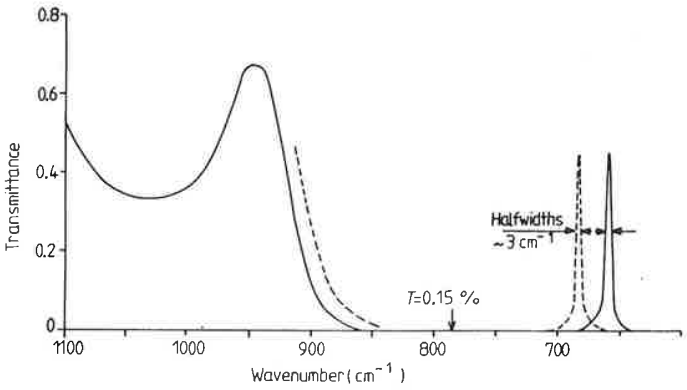


Figure 7.13 Measured transmittance of two filters of type (*b*) (see p 264). Design: Air|*HLHL HH LHLHL*|Ge substrate |*L*|Air (*H* = PbTe, *L* = ZnS). (After Pidgeon and Smith[14].)

Effect of an incident cone of light

The analysis can be taken a stage further to arrive at expressions for the degradations of peak transmission and bandwidth which become apparent when the incident illumination is not perfectly collimated. Essentially the same results have been obtained by Lissberger and Wilcock[13] and by Pidgeon and Smith[14].

It is assumed first of all that, in collimated light, the sole effect of tilting a filter is a shift of the characteristic towards shorter wavelengths or greater wavenumbers, leaving the peak transmittance and bandwidth virtually unchanged. The performance in convergent or divergent light is then given by integrating the transmission curve over a range of angles of incidence. The analysis is simpler in terms of wavenumber or of g, rather than wavelength. If ν_0 is the wavenumber corresponding to the peak at normal incidence and ν_Θ to the peak at angle of incidence Θ, then it is plausible that the resultant peak, when all angles of incidence in the cone from 0 to Θ are included, should appear at a wavenumber given by the mean of the above extremes. We shall show, shortly, that this is indeed the case. The new peak is given by

$$\nu_m = \nu_0 + \tfrac{1}{2}\Delta\nu' \tag{7.39}$$

where

$$\Delta\nu' = \nu_\Theta - \nu_0 = \nu_0\,\Theta^2/2n^{*2}.$$

The effective bandwidth of the filter will, of course, appear broader and, since the process is, in effect, a convolution of a function with bandwidth W_0, which is the width of the filter at normal incidence, and another function with bandwidth $\Delta\nu'$, the change in peak position produced by altering the angle of incidence from 0 to Θ, it seems likely that the resultant bandwidth might be

given by the square root of the sum of their squares. This too is indeed the case, as we shall also show.

$$W_\Theta^2 = W_0^2 + (\Delta\nu')^2. \tag{7.40}$$

The peak transmission falls and is given by

$$\hat{T}_\Theta = \left(\frac{W_0}{\Delta\nu'}\right)\tan^{-1}\left(\frac{\Delta\nu'}{W_0}\right). \tag{7.41}$$

The analysis is as follows.

We consider incident light in the form of a cone with semiangle Θ, that is a cone of focal ratio $1/(2\tan\Theta)$. We assume that in collimated light the effect of tilting the filter is simply to move the characteristic towards shorter wavelengths, leaving the bandwidth and peak transmittance unchanged.

For small values of θ, the flux incident on the filter is proportional to $\theta\,\mathrm{d}\theta$. The resultant transmittance of the filter is then given by the total flux transmitted divided by the total flux incident.

The total flux incident is proportional to

$$\int_0^\Theta \theta\,\mathrm{d}\theta = \tfrac{1}{2}\Theta^2.$$

The total flux transmitted is proportional to

$$\int_0^\Theta \theta T\,\mathrm{d}\theta.$$

We can, for small values of θ and Δg, set

$$T = \frac{1}{1-\{(2/\Delta g_h)[\Delta g - (\theta_i^2/2n^{*2})]\}^2}$$

where Δg_h is the halfwidth at normal incidence of the filter in units of g. This expression follows directly from the concept of n^*. The transmittance of the filter is then given by

$$\begin{aligned} T &= \frac{2}{\Theta^2}\int_0^\Theta \frac{\theta\,\mathrm{d}\theta}{1+\{(2/\Delta g_h)(\Delta g - (\theta_i/2n^{*2})]\}^2} \\ &= -\frac{2}{\Theta^2}\frac{n^{*2}\Delta g_h}{2}\left[\tan^{-1}\left\{\frac{2}{\Delta g_h}\left(\Delta g - \frac{\theta_i}{2n^{*2}}\right)\right\}\right]_0^\Theta \\ &= \frac{1}{2}\frac{\Delta g_h}{(\Theta^2/2n^{*2})}\left\{\tan^{-1}\left(2\frac{\Delta g}{\Delta g_h}\right) - \tan^{-1}\left[2\left(\frac{\Delta g}{\Delta g_h} - \frac{\Theta^2}{2n^{*2}}\frac{1}{\Delta g_h}\right)\right]\right\} \qquad (7.42) \\ &= \frac{1}{2}\frac{\Delta g_h}{(\Theta^2/2n^{*2})}\left[\tan^{-1}\left(\frac{(2/\Delta g_h)(\Theta^2/2n^{*2})}{1+(2/\Delta g_h)^2\{\Delta g[\Delta g - (\Theta^2/2n^{*2})]\}}\right)\right]. \qquad (7.43) \end{aligned}$$

This is a maximum when

$$\Delta g = \frac{1}{2}\frac{\Theta^2}{2n^{*2}}.$$

But $\Theta^2/(2n^{*2})$ is the shift in the position of the peak at angle of incidence Θ. Thus in a cone of light of semiangle Θ, the peak wavelength of the filter is given by the mean of the value at normal incidence and that at the angle Θ corresponding to equation (7.39). The value of the peak transmittance is then, from equation (7.42),

$$\frac{\Delta g_h}{(\Theta^2/2n^{*2})}\tan^{-1}\left(\frac{\Theta^2/2n^{*2}}{\Delta g_h}\right)$$

which corresponds to equation (7.41).

The half-peak points are given by

$$(7.43) = \tfrac{1}{2}\,(\text{peak } T)$$

i.e.

$$\frac{1}{2}\cdot\frac{\Delta g_h}{(\Theta^2/2n^{*2})}\tan^{-1}\left(\frac{(2/\Delta g_h)(\Theta^2/2n^{*2})}{1+(2/\Delta g_h)^2\{\Delta g[\Delta g-(\Theta^2/2n^{*2})]\}}\right)$$

$$=\frac{1}{2}\frac{\Delta g_h}{(\Theta^2/2n^{*2})}\tan^{-1}\left(\frac{\Theta^2/2n^{*2}}{\Delta g_h}\right)$$

which is satisfied by

$$1+\left(\frac{2}{\Delta g_h}\right)\left[\Delta g\left(\Delta g-\frac{\Theta^2}{2n^{*2}}\right)\right]=2$$

i.e.

$$\Delta g\left(\Delta g-\frac{\Theta^2}{2n^{*2}}\right)-\left(\frac{\Delta g_h}{2}\right)^2=0.$$

We are interested in the difference between the roots of the equation which is the width of the characteristic

$$(\Delta g_1-\Delta g_2)=\left[\left(\frac{\Theta^2}{2n^{*2}}\right)^2+(\Delta g_h)^2\right]^{1/2}$$

which corresponds exactly to equation (7.40).

Since

$$\tan^{-1}x = x-\frac{x^3}{3}+\frac{x^5}{5}-\frac{x^7}{7}+\ldots \qquad \text{for } |x|\leqslant 1$$

for small values of $(\Delta\nu'/W_0)$ we can write

$$\hat{T}_\Theta = 1-\frac{1}{3}\left(\frac{\Delta\nu'}{W_0}\right)^2. \tag{7.44}$$

If FR denotes the focal ratio of the incident light, then, for values of around 2 to infinity, it is a reasonably good approximation that

$$\Theta = 1/[2(FR)].$$

Using this, we find another expression for $\Delta\nu'$ which can be useful:

$$\Delta\nu' = \frac{\nu_0}{8n^{*2}(FR)^2}.$$

We can extend this analysis still further to the case of a cone of semiangle Θ incident at an angle other than normal, provided we make some simplifying assumptions. If the angle of incidence of the cone is χ then the range of angles of incidence will be $\chi \pm \Theta$.

If $\chi < \Theta$ then we can assume that the result is simply that for a normally incident cone of semiangle $\chi + \Theta$.

If $\chi > \Theta$ then we have three frequencies, ν_0 corresponding to normal incidence, ν_1 to angle of incidence $\chi - \Theta$, and ν_2 to angle of incidence $\chi + \Theta$. The new filter peak can be assumed to be

$$\frac{1}{2}(\nu_1 + \nu_2) = \frac{\chi^2 + \Theta^2}{2n^{*2}}\nu_0 \qquad (\chi \text{ and } \Theta \text{ in radians})$$

$$= \frac{1.52 \times 10^{-4}(\chi^2 + \Theta^2)}{n^{*2}}\nu_0 \qquad (\chi \text{ and } \Theta \text{ in degrees}). \tag{7.45}$$

The halfwidth is

$$[W_0^2 + (\nu_2 - \nu_1)^2]^{1/2}$$

where

$$(\nu_2 - \nu_1) = \frac{2\chi\Theta}{n^{*2}}\nu_0 \qquad (\chi \text{ and } \Theta \text{ in radians})$$

$$= \frac{6.09 \times 10^{-4}\chi\Theta}{n^{*2}}\nu_0 \qquad (\chi \text{ and } \Theta \text{ in degrees}) \tag{7.46}$$

and the peak transmittance is

$$\frac{W_0}{(\nu_2 - \nu_1)}\tan^{-1}\left(\frac{(\nu_2 - \nu_1)}{W_0}\right) \simeq 1 - \frac{1}{3}\left(\frac{(\nu_2 - \nu_1)}{W_0}\right)^2. \tag{7.47}$$

$(\nu_2 - \nu_1)$ is proportional to $\Theta\chi$ and Hernandez[15] has found excellent agreement between measurements made on real filters and calculations from these expressions for values of $\Theta\chi$ up to 100 degrees^2.

We can illustrate the use of these expressions in calculating the performance of a zinc sulphide and cryolite filter for the visible region. We assume that this is a low-index first-order filter with a bandwidth of 1%.

For this filter we calculate that $n^* = 1.55$. We take 10% reduction in peak transmittance as the limit of what is acceptable. Then, from equation (7.47)

$$(\nu_2 - \nu_1)/W_0 = 0.55$$

and the increased halfwidth which corresponds to this reduction in peak transmittance is

$$(1+0.55^2)^{1/2}\,W_0 = 1.14\,W_0$$

or an increase of 14% over the basic width.

At normal incidence, the cone semiangle which can be tolerated is given by

$$1.5\times10^4(\Theta^2/n^{*2}) = \Delta\nu = 0.55\,W_0 = 0.55\times0.01 \qquad (\Theta \text{ in degrees})$$

i.e.

$$\Theta = [1.55^2\times0.55\times0.01/(1.5\times10^{-4})]^{1/2} = 9.4^\circ.$$

Such a cone at normal incidence will cause a shift in the position of the peak towards shorter wavelengths or higher frequencies of

$$\tfrac{1}{2}(\Delta\nu'/\nu_0) = (\tfrac{1}{2}\times0.55\times0.01) = 0.275\%.$$

Used at oblique incidence in a cone of illumination we have

$$(6.09\times10^{-4}\chi\Theta/n^{*2})\nu_0 = \nu_2-\nu_1 = 0.55\times0.01$$

i.e.

$$\chi\Theta = \frac{1.55^2\times0.55\times0.01}{6.09\times10^{-4}} = 21.7 \text{ degrees}^2$$

which means that the filter can be used in a cone of semiangle 2° up to an angle of incidence of 21.7/2 = 10.9° or of semiangle 3° up to an angle of incidence of 7° and so on.

One very important result is the shift in peak wavelength in a cone at normal incidence which indicates that if a filter is to be used at maximum efficiency in such an arrangement, its peak wavelength at normal incidence in collimated light should be slightly longer to compensate for this shift.

Sideband blocking

There is a disadvantage in the all-dielectric filter that the high-reflectance zone of the reflecting coating is limited in extent and hence the rejection zone of the filter is also limited. In the near ultraviolet, visible and near infrared regions, the transmission sidebands on the shortwave side of the peak can usually be suppressed, or blocked, by an absorption filter with a longwave-pass characteristic in the same way as for metal–dielectric filters. The longwave sidebands are more of a problem. These may be outside the range of sensitivity of the detector and therefore may not require elimination, but if they are troublesome then the usual technique for removing them is the addition of a metal–dielectric first-order filter with no longwave sidebands. It is usually very much broader than the narrowband component in order that the peak transmittance may be high. The metal–dielectric component is usually added as a separate component, but it can be deposited over the basic Fabry–Perot. Rather than a simple Fabry–Perot filter, a double cavity metal–dielectric is commonly used. Multiple cavity filters are the next topic of discussion.

MULTIPLE CAVITY FILTERS

The transmission curve of the basic all-dielectric Fabry–Perot filter is not of ideal shape. It can be shown that one half of the energy transmitted in any order lies outside the halfwidth (assuming an even distribution of energy with frequency in the incident beam). A more nearly rectangular curve would be a great improvement. Further, the maximum rejection of the Fabry–Perot is completely determined by the halfwidth and the order. The broader filters, therefore, tend to have poor rejection as well as a somewhat unsatisfactory shape.

When tuned electric circuits are coupled together, the resultant response curve is rather more rectangular and the rejection outside the pass band rather greater than a single tuned circuit, and a similar result is found for the Fabry–Perot filter. If two or more of these filters are placed in series, much the same sort of double peaked curve is obtained; it has, however, a much more promising shape than the single filter. The filters may be either metal–dielectric or all-dielectric and the basic form is

| reflector | half-wave | reflector | half-wave | reflector |

known as a double half-wave or DHW filter or as a double cavity or *two-cavity* filter. Some typical examples of all-dielectric DHW or two-cavity filters are shown in figure 7.14.

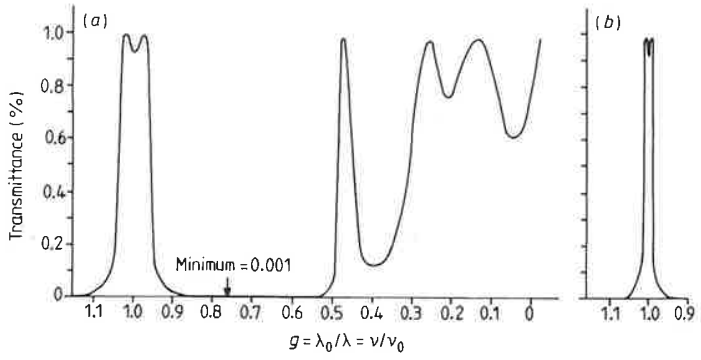


Figure 7.14 (*a*) Computed transmittance of *HLLHLHLLH*. (*b*) Computed transmittance of *HLHHLHLHHLH*. In both cases $n_H = 4.0$ and $n_L = 1.35$. (After Smith[18].)

Such filters were certainly constructed by A F Turner and his co-workers at Bausch and Lomb in the early 1950s but the results were published only as quarterly reports in the Fort Belvoir Contract Series over the period 1950–68[16]. The earliest filters were of the triple half-wave type, known at Bausch and Lomb as WADIS (Wide Band All Dielectric Interference filters)[17]. Double half-wave, or two-cavity, filters came later but were in routine use at

Bausch and Lomb certainly by 1957. They were initially known as TADIS. The Fort Belvoir Contract† Reports make fascinating reading and show just how advanced the work at Bausch and Lomb was at that time. Use was being made of the concept of equivalent admittance for the design both of WADI filters and of the edge filters for blocking the sidebands. Multilayer antireflection coatings were also well understood.

The first complete account of a theory applicable to multiple half-wave filters was published by Smith[18] and it is his method that we follow first here.

The reflecting stacks in the classical Fabry–Perot filter have more or less constant reflectance over the pass band of the filter. A dispersion of phase change on reflection does, as we have seen, help to reduce the bandwidth, but this does so without altering the basic shape of the pass band shape. Smith suggested the idea of using reflectors with much more rapidly varying reflectance to achieve a better shape. The essential expression for the transmission of the complete filter has already been derived on p 52 where we have assumed $\beta = 0$, that is, no absorption in the spacer layer. From Smith's formula, equation (2.80),

$$T = \frac{|\tau_a^+|^2|\tau_b^+|^2}{(1-|\rho_a^-||\rho_b^+|)^2}\left[1+\frac{4|\rho_a^-||\rho_b^+|}{(1-|\rho_a^-||\rho_b^+|)^2}\sin^2\frac{\phi_a+\phi_b-2\delta}{2}\right]^{-1} \quad (7.48)$$

it can be seen that high transmission can be achieved at any wavelength if, and only if, the reflectances on either side of a chosen spacer layer are equal. Of course the phase condition must be met too, but this can be arranged by choosing the correct spacer thickness to make

$$\left|\frac{\phi_a+\phi_b}{2}-\delta\right| = m\pi.$$

In these expressions, the symbols have the same meanings as given in figure 2.12.

Smith pointed out the advantage of having reasonably low reflectance in the region around the peak wavelength, which means that absorption is less effective in limiting the peak transmittance. In the Fabry–Perot filter, low reflectance means wide bandwidth, but Smith limited the bandwidth by arranging for the reflectances to begin to differ appreciably at wavelengths only a little removed from the peak. This is illustrated in figure 7.15. The figure shows what is the simplest type of DHW filter, which has construction *HHLHH*. The *HH* layers are the two half-wave spacers and the *L* layer is a coupling layer. In the discussion which follows, for simplicity we shall ignore any substrate. The behaviour of the filter is described in terms of the reflectances on either side of one of the two spacers. R_1 is the reflectance of the interface between the

† These reports were obtainable from the Engineer Research and Development Laboratories, Fort Belvoir, Virginia 22060, USA, but are now out of print.

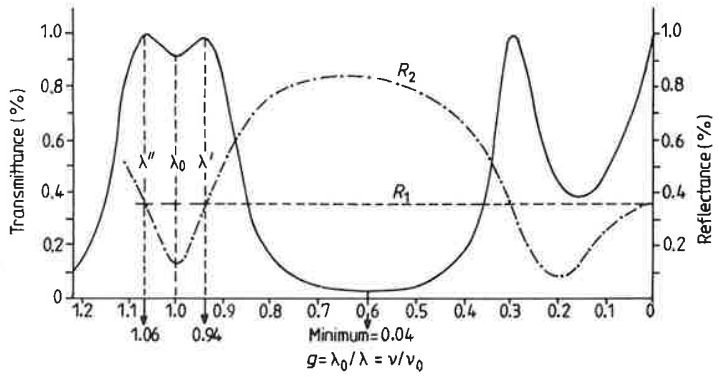


Figure 7.15 Computed transmittance of $HHLHH$ and explanatory reflectance curves R_1 and R_2 ($n_H = 4.0$, $n_L = 1.35$). (After Smith[18].)

high index and the surrounding medium, which we take as air with index unity, and is a constant. R_2 is the reflectance of the assembly on the other side of the spacer and is low at the wavelength at which the spacer is a half-wave and rises on either side. At wavelengths λ' and λ'', the reflectances R_1 and R_2 are equal and we would expect to see high transmission if the phase condition is met, which in fact it is. The transmission of the assembly is also shown in the figure and the shape can be seen to consist of a steep-sided pass band with two peaks close together and only a slight dip in transmission between the peaks, much more like the ideal rectangle than the shape of the Fabry–Perot filter.

Smith's formula for the transmittance of a filter can be written:

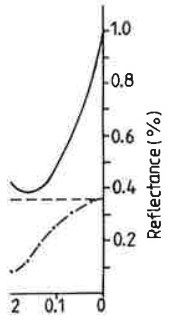
$$T(\lambda) = T_0(\lambda)\frac{1}{1+F(\lambda)\sin^2[(\phi_1+\phi_2)/2-\delta]} \tag{7.49}$$

where

$$T_0(\lambda) = \frac{(1-R_1)(1-R_2)}{[1-(R_1R_2)^{1/2}]^2} \tag{7.50}$$

$$F(\lambda) = \frac{4(R_1R_2)^{1/2}}{[1-(R_1R_2)^{1/2}]^2}. \tag{7.51}$$

Both these quantities are now variable since they involve R_2, which is a variable. The form of the functions is also shown in figure 7.16. At wavelengths removed from the peak, $T_0(\lambda)$ is low and $F(\lambda)$ is high, the combined effect being to increase the rejection. In the region of the peak, T_0 is high, and, just as important, F is low, producing high transmittance which is not sensitive to the effects of absorption. As we have shown before, the peak transmittance is dependent on the quantity A/T, where A is the absorptance and T the transmittance of the reflecting stacks. Clearly, the greater T is, the higher A can be for the same overall filter transmittance.



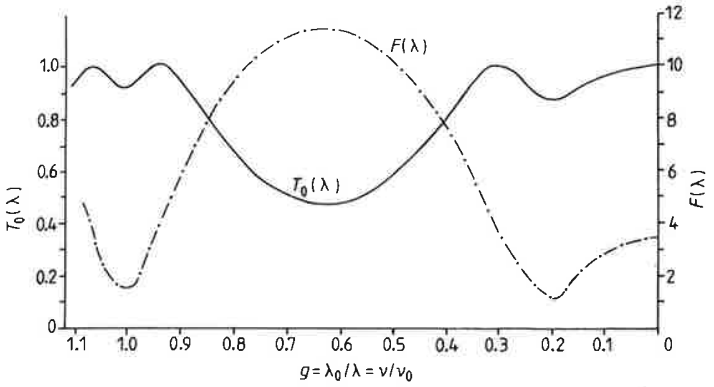


Figure 7.16 $T_0(\lambda)$ and $F(\lambda)$ for *HHLHH*. (After Smith[18].)

The typical double-peaked shape of the double half-wave filter results from the intersection of the R_1 and R_2 curves at two separate points. Two other cases can arise. The curves can intersect at one point only, in which case the system has a single peak whose transmittance is theoretically unity, or the curves may never intersect at all, in which case the system will show a single peak of transmittance rather less than unity, the exact magnitude depending on the relative magnitudes of R_1 and R_2 at their closest approach. This third case is to be avoided in design. For the twin-peaked filter, a requirement is that the trough in the centre between the two peaks should be shallow, which means that R_1 and R_2 should not be very different at λ_0.

Having examined the simplest type of DHW filter, we are in a position to study more complicated ones. What we have to look for is a system of two reflectors, where one of the reflectors remains reasonably constant over the range of interest and where the other should be equal, or nearly equal, to the first over the pass band region, but should increase sharply outside the pass band. The straightforward Fabry–Perot filter has effectively zero reflectance at the peak wavelength, but the reflectance rapidly rises on either side of the peak. If, then, a simple quarter-wave stack is added to the Fabry–Perot, the resultant combination should have the desired property, that is, the reflectance equal to that of the simple stack at the centre wavelength and increasing sharply on either side. We can therefore use a simple stack as one reflector, with more or less constant reflectance, on one side of the spacer, and, on the other side, an exactly similar stack combined with a Fabry–Perot filter. This will result in a single-peaked filter since the reflectances in this way will be exactly matched at λ_0. The double-peaked transmission curve will be obtained if the reflectance of the stack plus the Fabry–Perot filter is arranged to be just a little less than the reflectance of the stack by itself. This is the arrangement that is more often used and it involves the insertion of an extra quarter-wave layer between the stack

and the Fabry–Perot. This layer appears as a sort of coupling layer in the filter. Figure 7.17 should make the situation clear.

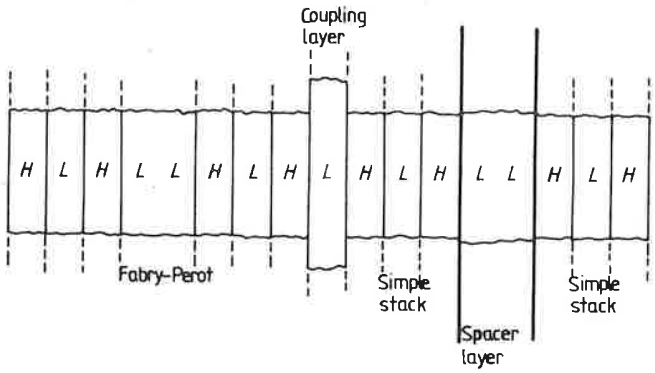


Figure 7.17 The construction of a DHW filter.

So far we have not given any consideration to the substrate of the filter. The substrate will be on one side of the spacer and will alter the reflectance on that side. This change in reflectance can easily be calculated, particularly if the substrate is considered to be on the same side of the spacer as the simple stack. The constant reflectance R_1 of the simple stack will generally be large, and if the substrate index is given by n_s, then the transmittance of the stack on its own, $(1 - R_1)$, will become either $(1 - R_1)/n_s$ if the index of the layer next to the substrate is low, or $n_s(1 - R_1)$ if it is high.

Since this change in reflectance could be considerable, especially if n_s is large, the substrate must be taken into account in the design and this should be done right from the beginning. The substrate can be considered part of the simple stack and R_1 can be adjusted to include it. Provided the reflectances of the two assemblies on either side of the spacer layer are arranged always to be equal at the appropriate wavelengths, the transmittance of the complete filter will be unity.

For example, let us consider the case of a filter deposited on a germanium substrate using zinc sulphide for the low-index layers and germanium for the high ones. Let the spacer be of low index and let the reflecting stack on the germanium substrate be represented by Ge | $LHLL$, where the LL layer is the spacer. The transmittance of the stack into the spacer layer will be approximately $T_1 = 4n_L^3/n_H^2 n_{Ge}$, which, since the substrate is the same material as the high-index layer, becomes $4n_L^3/n_H^3$. On the other side of the spacer layer we make a start with the combination $LLHLH$|air, representing the basic reflecting stack, where LL once again is the spacer layer. This has transmission $T_2 = 4n_L^3/n_H^4$, which is $1/n_H$ times T_1. Clearly this is too

unbalanced and an adjustment to this second stack must be made. If a low-index layer is added next to the air, then the transmission becomes $T_2 = 4n_L^5/n_H^4$. Since n_L^2 is approximately equal to the index of germanium, the transmittances T_1 and T_2 are now equal and the Fabry–Perot filter can be added to the second stack to give the desired shape to the reflectance curve. The Fabry–Perot can take any form, but it is convenient here to use a combination almost exactly the same as the combination of two stacks and a spacer layer which has already been arrived at. The complete design of filter is then:

$$\text{Ge}|LH\ LL\ HLH\ L\ HLH\ LL\ HLH|\text{air}$$

and the performance of the filter is shown in figure 7.18.

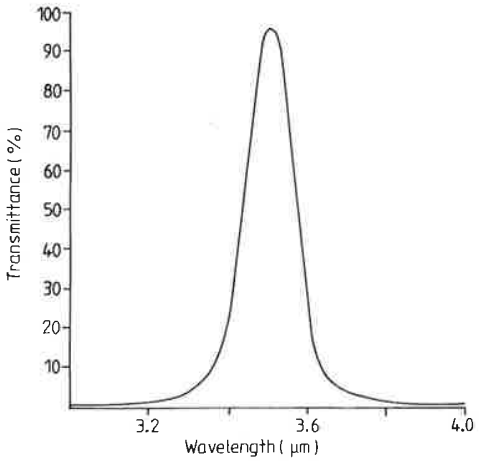


Figure 7.18 Computed transmittance of the double half-wave filter. Design: Air|$LHLLHLHLHLLHLH$|Ge. The substrate is germanium ($n = 4.0$); H = germanium ($n = 4.0$), L = zinc sulphide ($n = 2.35$) and the incident medium is air ($n = 1.0$).

An alternative way of checking whether or not the filter is going to have high transmission uses the concept of absentee half-wave layers. The layers in DHW filters are usually either of quarter- or half-wave thickness at the centre of the pass band, as in the above filter, and we can take it as an example to illustrate the method. First we note that the two spacers are both half-wave layers and that they can be eliminated without affecting the transmission. The filter, at the centre wavelength, will have the same transmittance as

$$\text{Ge}|LH\ HLH\ L\ HLH\ HLH|\text{air}.$$

In this there are two sets of HH layers which can be eliminated in the same way, leaving two sets of LL layers which can be removed in their turn. Almost all the

layers in the filter can be eliminated in this way leaving ultimately

$$\text{Ge}\,|\,L\,|\,\text{air}.$$

As we already know, a single quarter-wave of zinc sulphide is a good antireflection coating for germanium, and so the transmittance of the filter will be high in the centre of the pass band. Any type of DHW filter can be dealt with in this way.

Knittl[19,20] has used an alternative multiple beam approach to study the design of DHW filters. Basically he has applied a multiple beam summation to the first cavity, the results of which are then used in a multiple beam summation for the second cavity. This yields an expression which is not unlike Smith's, although slightly more complicated, but which has the advantage that it is only the phase which varies across the pass band. The magnitude of the reflection and transmission coefficients can be safely assumed constant and this means that the parameters which involve these quantities are also constant. The form of the expression for overall transmittance is then very much easier to manipulate so that the positions and values of maxima and minima in the pass band can be readily determined. We shall not deal further with the method here, because it is already well covered by Knittl[20].

Of course, the possible range of designs does not end with the DHW filter. Other types of filter exist involving even more half-waves. An early type of filter, which we have already mentioned, was the WADI which was devised by Turner and which consisted of a straightforward Fabry–Perot filter, to either side of which was added a half-wave layer together with several quarter-wave layers. The function of these extra layers was to alter the phase characteristics of the reflectors on either side of the primary spacer layer, so that the pass band was broadened and at the same time the sides became steeper. Similarly, it is possible to repeat the basic Fabry–Perot element used in the DHW filter once more to give a triple half-wave or THW filter, which has a similar bandwidth but steeper sides. WADI and THW filters are much the same thing, although the original design philosophy was a little different, and usually the term THW is taken as referring to all types having three half-wave spacers. Even more spacer layers may be used giving multiple half-wave filters. The method which we have been using for the analysis of the filters becomes rather cumbersome when many half waves are involved—even the simple method for checking that the transmittance is high in the pass band breaks down, for reasons which will be made clear in the next section, where we shall consider a very powerful design method which has been devised by Thelen.

Thelen's method of analysis

We have not yet arrived at any ready way of calculating the bandwidth of DHW and THW filters. The design method has merely ensured that the transmittance of the filter is high in the pass band and that the shape of the transmission curve

is steep-sided. The bandwidth can be calculated, but to arrive at a prescribed bandwidth in the design has to be achieved by trial and error. It can indeed be calculated using the formula for transmittance

$$T = T_0 \frac{1}{1 + F_0 \sin^2 \delta}$$

but this can be very laborious as the phases of the reflectances have to be included in δ. This expression has been very useful in achieving an insight into the basic properties of the multiple half-wave filter, but, for systematic design, a method based on the concept of equivalent admittance will be found much more useful.

As was shown in chapter 6, any symmetrical assembly of thin films can be replaced by a single layer of equivalent admittance and optical thickness which both vary with wavelength, but which can be calculated. This concept has been used by Thelen[21] in the development of a very powerful systematic design method which predicts all the performance features of the filters including the bandwidth. The basis of the method is the splitting of the multiple half-wave filter into a series of symmetrical periods, the properties of which can be predicted by finding the equivalent admittance. Take for example the design we have already examined

$$\text{Ge}\,|\,LHLLHLHLHLHLLHLH\,|\,\text{air}.$$

This can be split up into the arrangement

$$\text{Ge}|LHL \quad LHLHLHLHL \quad LHLH|\text{air}.$$

The part of the filter which determines its properties is the central section $LHLHLHLHL$ which is a symmetrical assembly. It can therefore be replaced by a single layer having the usual series of high-reflectance zones where the admittance is imaginary, and pass zones where the admittance is real. We are interested in the latter because they represent the pass bands of the final filter. The symmetrical section must then be matched to the substrate and the surrounding air, and matching layers are added for that purpose on either side. This is the function of the remaining layers of the filter. The condition for perfect matching is easily established because the layers are all of quarter-wave optical thicknesses.

A most useful feature of this design approach is that the central section of the filter can be repeated many times, steepening the edges of the pass band and improving the rejection without affecting the bandwidth to any great extent.

In order to make predictions of performance straightforward, Thelen has computed formulae for the bandwidth of the basic sections. We use Thelen's technique here, with some slight modifications, in order to fit in with the pattern of analysis already carried out for the Fabry–Perot. In order to include filters of order higher than the first, we write the basic period as

$$H^m LHLHLH \ldots LH^m \qquad \text{or} \qquad L^m HLHLHL \ldots HL^m$$

where there are $2x+1$ layers, $x+1$ of the outermost index and x of the other, and m is the order number. We have already mentioned how Seeley[4], in the course of developing expressions for the Fabry–Perot filter, arrived at an approximate formula for the product of the characteristic matrices of quarter-wave layers of alternating high and low indices. Using an approach similar to Seeley's, we can put the characteristic matrix of a quarter-wave layer in the form:

$$\begin{bmatrix} -\varepsilon & \mathrm{i}/n \\ \mathrm{i}n & -\varepsilon \end{bmatrix} \tag{7.52}$$

where $\varepsilon = (\pi/2)(g-1)$ and $g = \lambda_0/\lambda$. This expression is valid for wavelengths close to that for which the layer is a quarter-wave. First let us consider m odd, and write m as $2q+1$. Then, to the same degree of approximation, the matrix for H^m or L^m is

$$(-1)^q \begin{bmatrix} -m\varepsilon & \mathrm{i}/n \\ \mathrm{i}n & -m\varepsilon \end{bmatrix}.$$

Neglecting terms of second and higher order in ε, then the product of the $2x-1$ layers making up the symmetrical period is

$$\begin{bmatrix} M_{11} & \mathrm{i}M_{12} \\ \mathrm{i}M_{21} & M_{22} \end{bmatrix} \tag{7.53}$$

where

$$M_{11} = M_{22} = (-1)^{x+2q}(-\varepsilon)\left[m\left(\frac{n_1}{n_2}\right)^x + \left(\frac{n_1}{n_2}\right)^{x-1} + \left(\frac{n_1}{n_2}\right)^{x-2} + \ldots + \left(\frac{n_2}{n_1}\right)^{x-1} + m\left(\frac{n_2}{n_1}\right)^x \right]$$

$$\mathrm{i}M_{12} = \mathrm{i}(-1)^x/[(n_1/n_2)^x n_1]$$

and

$$\mathrm{i}M_{21} = \mathrm{i}(-1)^x[(n_1/n_2)^x n_1].$$

Now it is not easy from this expression to derive the halfwidth of the final filter analytically. Instead of deriving the halfwidth, therefore, Thelen chose to define the edges of the pass band as those wavelengths for which

$$\frac{1}{2}\left|M_{11}+M_{22}\right| = 1$$

or, since $M_{11} = M_{22}$,

$$|M_{11}| = 1.$$

These points will not be too far removed from the half peak transmission points, especially if the sides of the pass band are steep. Applying this to equation (7.53), we obtain

$$|M_{11}| = \varepsilon\left[m\left(\frac{n_1}{n_2}\right)^x + \left(\frac{n_1}{n_2}\right)^{x-1} + \ldots + \left(\frac{n_2}{n_1}\right)^{x-1} + m\left(\frac{n_2}{n_1}\right)^x \right]. \tag{7.54}$$

Now, this expression is quite symmetrical in terms of n_1 and n_2. Then if we replace n_1 and n_2 by n_H and n_L, regardless of which is which, we will obtain the same expression

$$\varepsilon\left[m\left(\frac{n_H}{n_L}\right)^x+\left(\frac{n_H}{n_L}\right)^{x-1}+\left(\frac{n_H}{n_L}\right)^{x-2}+\ldots+\left(\frac{n_L}{n_H}\right)^{x-1}+m\left(\frac{n_L}{n_H}\right)^x\right]=1$$

i.e.

$$\varepsilon\left[(m-1)\left(\frac{n_H}{n_L}\right)^x+(m-1)\left(\frac{n_L}{n_H}\right)^x+\left(\frac{n_H}{n_L}\right)^x\left(\frac{1-(n_L/n_H)^{x+1}}{1-(n_L/n_H)}\right)\right]=1$$

where we have used the formula for the sum of a geometric series just as in the case of the Fabry–Perot. We now neglect terms of power x or higher in (n_L/n_H) to give

$$\varepsilon\left(\frac{n_H}{n_L}\right)^x\left((m-1)+\frac{1}{1-(n_L/n_H)}\right)=1$$

i.e.

$$\varepsilon=\left(\frac{n_L}{n_H}\right)^x\frac{[1-(n_L/n_H)]}{[m-(m-1)(n_L/n_H)]}. \tag{7.55}$$

The bandwidth will be given by

$$\left|\frac{\Delta\lambda_B}{\lambda_0}\right|=\left|\frac{\Delta\nu_B}{\nu_0}\right|=2(g-1)=\frac{4\varepsilon}{\pi}$$

so that, manipulating equation (7.55) slightly,

$$\left|\frac{\Delta\lambda_B}{\lambda_0}\right|=\frac{4}{m\pi}\left(\frac{n_L}{n_H}\right)^x\frac{(n_H-n_L)}{(n_H-n_L+n_L/m)}. \tag{7.56}$$

The equivalent admittance is given by

$$\eta_E=\left(\frac{M_{21}}{M_{12}}\right)^{1/2}=\left(\frac{n_1}{n_2}\right)^x n_1. \tag{7.57}$$

The case of m even, i.e. $m=2q$, is arrived at similarly. Here the matrix of H^m or L^m is

$$(-1)^q\begin{bmatrix}1 & im\varepsilon/n\\ im\varepsilon n & 1\end{bmatrix}$$

and a similar multiplication, neglecting terms higher than first in ε gives

$$\frac{\Delta\lambda_B}{\lambda_0}=\frac{4}{m\pi}\left(\frac{n_L}{n_H}\right)^x\frac{(n_H-n_L)}{(n_H-n_L+n_L/m)}$$

that is, exactly as equation (7.56), but

$$\eta_E=\left(\frac{n_2}{n_1}\right)^{x-1}n_2 \tag{7.58}$$

for equivalent admittance. This is to be expected since the layers L^m or H^m act as absentees because of the even value of m.

Expression (7.56) should be compared with the Fabry–Perot expressions (7.22) and (7.23). If we consider multiple cavity filters to be a series of Fabry–Perot cavities then the number of layers in each reflector is half that in the basic symmetrical period. Equations (7.22), (7.23) and (7.58) are, therefore, consistent.

In order to complete the design we need to match the basic period to the substrate and the surrounding medium. We first consider the case of first-order filters and the modifications which have to be made in the case of higher order will become obvious. For a first-order filter, then, matching will best be achieved by adding a number of quarter-wave layers to the period. The first layer should have index n_1, the next n_2 and so on, alternating the indices in the usual manner. The equivalent admittance of the combination of symmetrical period and matching layers will then be

$$\frac{n_1^{2y}}{n_2^{2(y-1)}}\left(\frac{n_2}{n_1}\right)^x \frac{1}{n_2} \qquad \text{or} \qquad \left(\frac{n_2}{n_1}\right)^{2y} n_2 \left(\frac{n_1}{n_2}\right)^x \tag{7.59}$$

where there are y layers of index n_1 and either $(y-1)$ or y layers of index n_2 respectively. We have also used the fact that the addition of a quarter-wave of index n to an assembly of equivalent admittance E alters the admittance of the structure to n^2/E.

This equivalent admittance should be made equal to the index of the substrate on the appropriate side, and to the index of the surrounding medium on the other. The following discussion should make the method clear.

When we try to apply this formula to the design of multiple half-wave filters, we find to our surprise that quite a number of designs which we have looked at previously, and which seemed satisfactory, do not satisfy the conditions. For example, let us consider the design arrived at in the earlier part of this section:

$$\text{Ge}\,|\,LH\ LL\ HLH\ L\ HLH\ LL\ HLH\,|\,\text{air}$$

where L indicates zinc sulphide of index 2.35 and H germanium of index 4.0. The central period is $LHLHLHLHL$, which has equivalent admittance n_L^5/n_H^4. The LHL combination alters this equivalent admittance to

$$\frac{n_L^4}{n_H^2}\frac{n_H^4}{n_L^5} = \frac{n_H^2}{n_L}$$

which is a gross mismatch to the germanium substrate. The $LHLH$ combination on the other side alters the admittance to

$$\frac{n_H^4}{n_L^4}\frac{n_L^5}{n_H^4} = n_L$$

which in turn is not a particularly good match to air.

The explanation of this apparent paradox is that in this particular case the total filter, taking the phase thickness of the central symmetrical period into

account, has unity transmittance because it satisfies Smith's conditions given in the previous section, but, over a wide range of wavelengths, pronounced transmission fringes would be seen if the bandwidth of the filter were not much narrower than a single fringe. Adding extra periods to the central symmetrical one has the effect of decreasing this fringe width, bringing them closer together. Eventually, given enough symmetrical periods, the width of the fringes becomes less than the filter bandwidth and they appear as a pronounced ripple superimposed on the pass band. This is illustrated clearly in figure 7.19. The triple half-wave version is still acceptable when an extra L layer is added, but this quintuple half-wave version is quite unusable. The presence or absence of an outermost L layer has no effect on the performance, other than inverting the fringes. The simple method of cancelling out half-waves for predicting the pass band transmission therefore breaks down, because it merely ensures that λ_0 will coincide with a fringe peak.

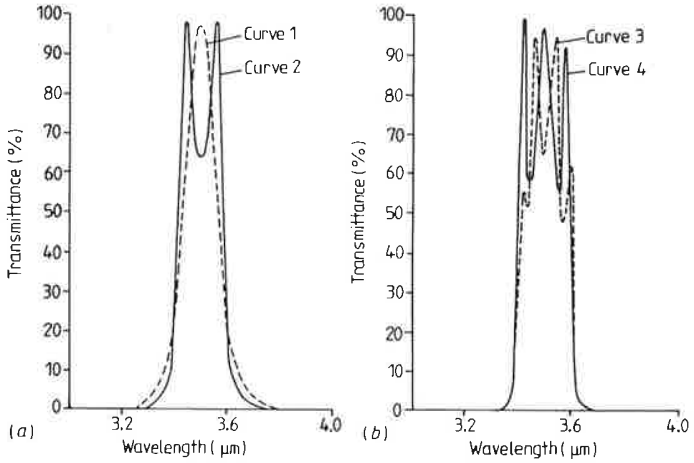


Figure 7.19 (*a*) Curve 1: Computed transmittance of the triple half-wave filter: Air|$LHLHL(LHLHLHLHL)^2$ LHL|Ge. Curve 2: shows the effect of omitting the L layer next to the air in the design of curve 1: Air|$HLHL(LHLHLHLHL)^2$ LHL|Ge. (*b*) Computed transmittance of quintuple half-wave filters. Curve 3: Air|$HLHL(LHLHLHLHL)^4$ LHL|Ge. Curve 4: As curve 3 but with an extra L layer: Air|$LHLHL(LHLHLHLHL)^4$ LHL|Ge. The presence or absence of the L layer has little effect on the ripple in the pass band. For all curves, H = germanium ($n_H = 4.0$) and L = zinc sulphide ($n_L = 2.35$).

It is profitable to look at the possible combinations of the two materials which can be made into a filter on germanium and where the centre section can be repeated as many times as required. The combinations for up to eleven layers in the centre section are given in table 7.3.

Table 7.3

Matching combination for germanium	Symmetrical period	Matching combination for air
Ge|L	LHL	|air (already matched)
Ge|LH	$HLHLH$	H|air
Ge|LHL	$LHLHLHL$	LH|air
Ge|$LHLH$	$HLHLHLHLH$	HLH|air
Ge|$LHLHL$	$LHLHLHLHLHL$	$LHLH$|air

L: ZnS, $n_L = 2.35$ H: Ge, $n_H = 4.0$

The validity of any of these combinations can easily be tested. Take for example the fourth one, with the nine-layer period in the centre. Here the equivalent admittance of the symmetrical period is $E = n_H^5/n_L^4$. The $LHLH$ section between the germanium substrate and the centre section transforms the admittance into

$$\frac{n_L^4}{n_H^4}\frac{n_H^5}{n_L^4} = n_H$$

which is a perfect match for germanium. The matching section at the other end is HLH and this transforms the admittance into

$$\frac{n_H^4}{n_L^2}\frac{n_L^4}{n_H^5} = \frac{n_L^2}{n_H}$$

which, because zinc sulphide is a good antireflection material for germanium, gives a good match for air.

For higher-order filters, the method of designing the matching layers is similar. However, we can choose, if we wish, to add half-wave layers to that part of the matching assembly next to the symmetrical period in order to make the resulting cavity of the same order as the others. For example, the period $HHHLHLHLHLHHH$, based on the fourth example of table 7.3, can be matched either by Ge|$LHLH$ and HLH|air, as shown, or by Ge|$LHLHHH$ and $HHHLH$|air, making all cavities of identical order regardless of the number of periods.

This method, then, gives the information necessary for the design of multiple half-wave filters. The edge steepness and rejection in the stop bands will determine the number of basic symmetrical periods in any particular case. Usually, because of the approximations which have been used in establishing the various formulae, and also because the definition used for bandwidth is not necessarily the halfwidth although it would not be too far removed from it, it is advisable to check the design by accurate computation before actually

manufacturing the filter. It may also be advisable to make an estimate of the permissible errors which can be tolerated in the manufacture because it is pointless attempting to achieve a performance beyond the capabilities of the process. The result will just be worse than if a less demanding specification had been attempted. The estimation of manufacturing errors is a subject which has not received much attention in the literature on thin-film filters. A brief discussion of permissible errors is given in chapter 10, pp 434–43, with some examples of calculations applied to multiple half-wave filters. Typical multiple half-wave filters are shown in figure 7.20.

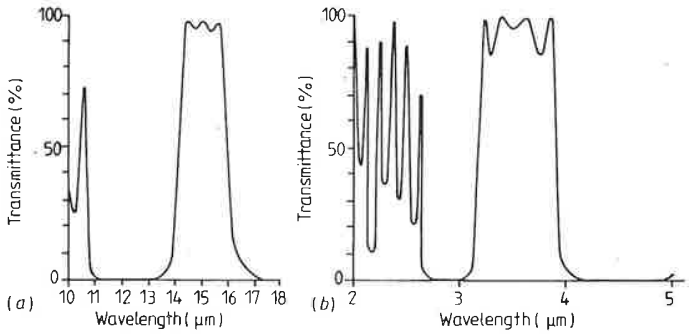


Figure 7.20 (*a*) Transmittance of a multiple half-wave filter. Design: Air|*HHL H LHHL LHHL H*|Ge with H = PbTe (n = 5.0); L = ZnS (n = 2.35), λ_0 = 15 μm. (*b*) Transmittance of a multiple half-wave filter. Design: Air|*HHL H LHHL H LHHL H LHHL H LHH*|silica H = Ge (n = 4.0); L = ZnS (n = 2.35); silica substrate (n = 1.45) λ_0 = 3.5 μm (courtesy of Sir Howard Grubb, Parsons & Co Ltd).

Effect of tilting

A feature of the design not so far mentioned is the sensitivity to changes in angle of incidence. Thelen[21] has examined this aspect and for those types which involve symmetrical periods consisting of quarter-waves of alternating high and low index and where the spacers are of the first order, he arrived at exactly the same expressions as those of Pidgeon and Smith for the Fabry–Perot. As far as angular dependence is concerned, the filter behaves as if it were a single layer with an effective index of

$$n^* = (n_1 n_2)^{1/2}$$

where $n_1 > n_2$ or

$$n^* = \frac{n_1}{[1-(n_1/n_2)+(n_1/n_2)^2]^{1/2}}$$

where $n_2 > n_1$.

For higher-order filters, therefore, we should be safe in making use of expressions (7.33) and (7.35).

Losses in multiple cavity filters

Losses in multiple cavity filters can be estimated in the same way as for the Fabry–Perot filter. There are so many possible designs that a completely general approach would be very involved. However, we can begin by assuming that the basic symmetrical unit is perfectly matched at either end. The scheme of admittances through the basic unit will then be as shown in table 7.4.

Table 7.4

	n_1^x/n_2^{x-1}
n_1	n_2^{x-1}/n_1^{x-2}
n_2	n_1^{x-2}/n_2^{x-3}
n_1	n_2^{x-3}/n_1^{x-4}
⋮	
	n_1^{x-2}/n_2^{x-3}
n_2	n_2^{x-1}/n_1^{x-2}
n_1	n_1^x/n_2^{x-1}

x layers of n_1
$(x-1)$ layers of n_2

Then, in the same way as for the Fabry–Perot, we can write

$$\begin{aligned}\sum \mathscr{A} = {} & \beta_1\left[\left(\frac{n_1}{n_2}\right)^{x-1}+\left(\frac{n_2}{n_1}\right)^{x-1}\right]+\beta_2\left[\left(\frac{n_2}{n_1}\right)^{x-2}+\left(\frac{n_1}{n_2}\right)^{x-2}\right] \\ & +\beta_1\left[\left(\frac{n_1}{n_2}\right)^{x-3}+\left(\frac{n_2}{n_1}\right)^{x-3}\right]+\ldots+\beta_2\left[\left(\frac{n_2}{n_1}\right)^{x-2}+\left(\frac{n_1}{n_2}\right)^{x-2}\right] \\ & +\beta_1\left[\left(\frac{n_1}{n_2}\right)^{x-1}+\left(\frac{n_2}{n_1}\right)^{x-1}\right] \\ = {} & \beta_1\left\{\left[\left(\frac{n_1}{n_2}\right)^{x-1}+\left(\frac{n_1}{n_2}\right)^{x-3}+\ldots+\left(\frac{n_2}{n_1}\right)^{x-1}\right]\right. \\ & \left.+\left[\left(\frac{n_2}{n_1}\right)^{x-1}+\left(\frac{n_2}{n_1}\right)^{x-2}+\ldots+\left(\frac{n_1}{n_2}\right)^{x-1}\right]\right\}\end{aligned}$$

$$+\beta_2\left\{\left[\left(\frac{n_2}{n_1}\right)^{x-2}+\left(\frac{n_2}{n_1}\right)^{x-4}+\ldots+\left(\frac{n_1}{n_2}\right)^{x-2}\right]\right.$$
$$\left.+\left[\left(\frac{n_1}{n_2}\right)^{x-2}+\left(\frac{n_1}{n_2}\right)^{x-4}+\ldots+\left(\frac{n_2}{n_1}\right)^{x-2}\right]\right\}.$$

We note that the second expression of each pair is the same as the first with inverse order.

The layers are quarter waves and so we can write, as before,

$$\beta_1=\frac{\pi}{2}\frac{k_1}{n_1} \qquad \text{and} \qquad \beta_2=\frac{\pi}{2}\frac{k_2}{n_2}.$$

Once again we divide the cases into high- and low-index spacers.

Case I: high-index spacers

We replace n_1 by n_H, k_1 by k_H, n_2 by n_L and k_2 by k_L. Then, neglecting, as before, terms in $(n_L/n_H)^x$ compared with unity,

$$\sum \mathscr{A}=\frac{\pi(k_H/n_H)(n_H/n_L)^{x-1}}{1-(n_L/n_H)^2}+\frac{\pi(k_L/n_L)(n_H/n_L)^{x-2}}{1-(n_L/n_H)^2}$$
$$=\pi\left(\frac{n_H}{n_L}\right)^x\frac{n_L(k_H+k_L)}{(n_H^2-n_L^2)}$$

or, using

$$\Delta\lambda_B/\lambda_0=\frac{2}{\pi}\sin^{-1}\left(\frac{2(n_1/n_2-1)}{(n_1/n_2)^x}\right)$$

and replacing $\sin^{-1}\theta$ by θ

$$\frac{\Delta\lambda_B}{\lambda_0}=\frac{4}{\pi}\frac{(n_H/n_L-1)}{(n_H/n_L)^x}$$

i.e.

$$\sum \mathscr{A}=4\left(\frac{\lambda_0}{\Delta\lambda_B}\right)\frac{(k_H+k_L)}{(n_H+n_L)}.$$

Now, this is the loss of one basic symmetrical unit. If further basic units are added each will have the same loss. In addition, there are the matching stacks at either end of the filter. We will not be far in error if we assume that they add a further loss equal to one of the basic symmetrical units. We can also assume that $R=0$ so that the absorption loss becomes

$$A=(p+1)\pi\left(\frac{n_H}{n_L}\right)^x\frac{n_L(k_H+k_L)}{(n_H^2-n_L^2)} \tag{7.61}$$

or

$$A=4(p+1)\left(\frac{\lambda_0}{\Delta\lambda_B}\right)\frac{(k_H+k_L)}{(n_H+n_L)} \tag{7.62}$$

where p is the number of basic units, i.e. $(p+1) = 2$ for a DHW and $(p+1) = 3$ for a THW and so on.

Case II: low-index spacer

In the same way

$$A = (p+1)\pi\left(\frac{n_H}{n_L}\right)^x \frac{(n_H^2 k_L + n_L^2 k_H)}{n_H(n_H^2 - n_L^2)} \tag{7.63}$$

or

$$A = 4(p+1)\left(\frac{\lambda_0}{\Delta\lambda_B}\right)\frac{[k_L(n_H/n_L) + k_H(n_L/n_H)]}{(n_H + n_L)}. \tag{7.64}$$

Expressions (7.62) and (7.64) are $(p+1)$ times the absorption of Fabry–Perot filters with the same halfwidth, a not unexpected result.

Further information

The examples of multiple half-wave filters so far described have been for the infrared, but of course they can be designed for any region of the spectrum where suitable thin-film materials exist. An account of filters for the visible and ultraviolet is given by Barr[22]. All-dielectric filters, both of the Fabry–Perot and multiple cavity types for the near ultraviolet are described by Nielson and Ring[23]. They used combinations of cryolite and lead fluoride and of cryolite and antimony trioxide, the former for the region 250–320 nm and the latter for 320–400 nm. Apart from the techniques required for the deposition of these materials, the main difference between such filters and those for the infrared is that the values of the high and low refractive indices are much closer together, requiring more layers for the same rejection. Nielson and Ring's filters contained basic units of seventeen or nineteen layers, in most cases, so that complete DHW filters consisted of 31 or 39 layers respectively. Malherbe[24] has described a lanthanum fluoride and magnesium fluoride filter for 202.5 nm in which the basic unit had 51 layers (high-index first-order spacer), the full design being $(HL)^{12}H\ H(LH)^{25}H(LH)^{12}$ with a total number of 99 layers, giving a measured bandwidth of 2.5 nm.

PHASE DISPERSION FILTER

The phase dispersion filter represents an attempt to find an approach to the design of narrowband filters which would avoid some of the manufacturing difficulties inherent in Fabry–Perot filters. The Fabry–Perot becomes increasingly difficult to manufacture as halfwidths are reduced below 0.3% of peak wavelength. Attempts to improve the position by using higher-order spacers are not effective when the spacer becomes thicker than the fourth order because of what has been described as increased roughness of the spacer. Much more is now known about the Fabry–Perot filter and the causes of manufacturing

difficulties, and those will be dealt with in some detail in a subsequent chapter. Although the phase dispersion filter was not, as it turned out, the solution to the narrowband filter problem, nevertheless it does have very interesting properties and the philosophy behind the design is worth discussing.

The reflecting stack with extended bandwidth which was originally intended for classical Fabry–Perot plates and was described in chapter 5 shows a large dispersion of the phase change on reflection and this suggested to Baumeister and Jenkins[25] that it might form the basis for a new type of filter in which the narrow bandwidth would depend almost entirely on this phase dispersion rather than on the very high reflectances of the reflecting stacks. They called this type of filter a 'phase dispersion filter'. It consists quite simply of a Fabry–Perot all-dielectric filter which has, instead of the conventional dielectric quarter-wave stacks on either side of the spacer layer, reflectors consisting of the staggered multilayers. The rapid change in phase causes the bandwidth of the filter and the position of the peak to be much less sensitive to the errors in thickness of the spacer layer than would otherwise be the case.

The results which they themselves[25] and also with Jeppesen[26] eventually achieved were good, although they never quite succeeded in attaining the performance possible in theory. This prompted a study[27] of the influence of errors in any of the layers of a filter on the position of the peak. The idea behind this study was that random errors in both thickness and uniformity in layers other than the spacer might be responsible for the discrepancy between theory and practice. If, in a practical filter, the errors were causing the peak to vary in position over the surface of the filter, then the integrated response would exhibit a rather wider bandwidth and lower transmittance than those of any very small portion of the filter, which might well be attaining the theoretical performance. It seemed possible that there might be a design of filter which could yield the minimum sensitivity to errors and therefore give the minimum possible bandwidth with a given layer 'roughness'.

Giacomo *et al*'s findings[27] can be summarised as follows (the notation in their paper has been slightly altered to agree with that used throughout this book): the peak of an all-dielectric multilayer filter is given by

$$\frac{\phi_a + \phi_b}{2} - \delta = m\pi \tag{7.65}$$

where

$$\delta = \frac{2\pi n d_s}{\lambda} = 2\pi n d_s \nu$$

the symbols having their usual meanings.

For a change Δd_i in the ith layer, Δd_j in the jth layer and Δd_s in the spacer, the corresponding change in the wavenumber of the peak $\Delta\nu$ is given by

$$\sum_i \frac{\partial \phi_a}{\partial d_i}\Delta d_i + \sum_j \frac{\partial \phi_b}{\partial d_j}\Delta d_j - 2\frac{\partial \delta}{\partial d_s}\Delta d_s + \left(\frac{\partial \phi_a}{\partial \nu} + \frac{\partial \phi_b}{\partial \nu} - 2\frac{\partial \delta}{\partial \nu}\right)\Delta\nu = 0. \tag{7.66}$$

Now

$$\frac{\partial\delta}{\partial d_s} = 2\pi n\nu = \frac{\delta}{d_s} \tag{7.67}$$

and

$$\frac{\partial\delta}{\partial\nu} = 2\pi n d_s = \frac{\delta}{\nu} \tag{7.68}$$

and also, since d_i and ν appear in the individual thin-film matrices only in the value of $\delta_i = 2\pi n_i d_i \nu$, then

$$\sum_i \frac{\partial\phi_a}{\partial d_i}\Delta_0 d_i = \frac{\partial\phi_a}{\partial\nu}\Delta_0\nu$$

and similarly for ϕ_b, where Δ_0 indicates that the changes in d_i are related by

$$\frac{\Delta_0 d_i}{d_i} = \frac{\Delta_0\nu}{\nu}.$$

This gives

$$\frac{\partial\phi_a}{\partial\nu} = \sum_i\left(\frac{\partial\phi_a}{\partial d_i}\frac{d_i}{\nu}\right) \tag{7.69}$$

which is independent of the particular choice of Δ_0 used to arrive at it. A similar expression holds for ϕ_b. Using equations (7.67), (7.68) and (7.69) in equation (7.66):

$$\sum_i \frac{\partial\phi_a}{\partial d_i}\Delta d_i + \sum_j \frac{\partial\phi_b}{\partial d_j}\Delta d_j - 2\delta\frac{\Delta d_s}{d_s} + \left(\sum_i \frac{\partial\phi_a}{\partial d_i}d_i + \sum_j \frac{\partial\phi_b}{\partial d_j}d_j - 2\delta\right)\frac{\Delta\nu}{\nu} = 0$$

i.e.

$$\frac{\Delta\nu}{\nu} = -\left[-2\delta\alpha_s + \sum_i\left(\frac{\partial\phi_a}{\partial d_i}d_i\alpha_i\right) + \sum_j\left(\frac{\partial\phi_b}{\partial d_j}d_j\alpha_j\right)\right] \times\left[-2\delta + \sum_i\left(\frac{\partial\phi_a}{\partial d_i}d_i\right) + \sum_j\left(\frac{\partial\phi_b}{\partial d_j}d_j\right)\right]^{-1} \tag{7.70}$$

where

$$\alpha_i = \frac{\Delta d_i}{d_i} \qquad \text{etc.}$$

Now, in a real filter, the fluctuations in thickness, or 'roughness', will be completely random in character, and in order to deal with the performance of any appreciable area of the filter, we must work in terms of the mean square deviations. Each layer in the assembly can be thought of as being a combination of a large number of thin elementary layers of similar mean thicknesses but which fluctuate in a completely random manner quite independently of each other. The RMS variation in thickness of any layer in the filter can then be considered to be proportional to the square root of its thickness. This can be written:

$$\varepsilon_i = k a_i^{1/2}$$

where k can be assumed to be the same for all layers regardless of thickness. If a_i is the RMS fractional variation of the ith layer, then

$$a_i = \frac{\varepsilon_i}{d_i} = \frac{k}{d_i^{1/2}} \tag{7.67}$$

where

$$a_i^2 = \overline{\alpha_i^2}. \tag{7.68}$$

We now define β as being

$$\beta^2 = \overline{\left(\frac{\Delta\nu}{\nu}\right)^2}.$$

Then

$$\beta^2 = \left\{4\delta^2 a_s^2 + \sum_i \left[\left(\frac{\partial\phi_a}{\partial d_i} d_i\right)^2 a_i^2\right] + \sum_j \left[\left(\frac{\partial\phi_b}{\partial d_j} d_j\right)^2 a_j^2\right]\right\} \times \left[-2\delta + \sum_i \left(\frac{\partial\phi_a}{\partial d_i} d_i\right) + \sum_j \left(\frac{\partial\phi_b}{\partial d_j} d_j\right)\right]^{-2}$$

which gives

$$\beta^2 = \left(k^2 \sum_{k=1}^{q} \frac{1}{d_k} A_k^2\right)\left(\sum_{k=1}^{q} A_k\right)^{-2} \tag{7.69}$$

where

$$A_k = \frac{\partial\phi_a}{\partial d_k} d_k \quad \text{or} \quad \frac{\partial\phi_b}{\partial d_k} d_k \quad \text{or} \quad -2\delta$$

whichever is appropriate. q is the number of layers in the filter. This expression will be a minimum when

$$A_k/d_k = A_l/d_l = \ldots \tag{7.70}$$

Then

$$\beta^2 = k^2/T \tag{7.71}$$

where T is the total thickness of the filter.

In the general case,

$$\beta \geqslant k/T^{1/2}$$

and one might hope to attain a limiting resolution of

$$R = T^{1/2}/k. \tag{7.72}$$

The condition written in equation (7.62) can be developed with the aid of equation (7.59) into

$$\frac{\partial\phi_a}{\partial d_k} = \frac{\partial\phi_b}{\partial d_l} = -4\pi n\nu \tag{7.73}$$

so that

$$\nu\left(\frac{\partial\phi_a}{\partial\nu}\right) = \sum_i \frac{\partial\phi_a}{\partial d_i} d_i = -4\pi n\nu d_m \tag{7.74}$$

and likewise for reflector b, where d_m = total thickness of the appropriate

reflector and a is the index of the spacer. This gives

$$\frac{\partial \phi_a}{\partial \nu} = -4\pi n d_m. \tag{7.75}$$

This condition is necessary but not sufficient for the resolution to be a maximum and it can be used as a preliminary test of the suitability of any particular multilayer reflector which may be employed.

The classical quarter-wave stack is very far from satisfying it but the staggered multilayer is much more promising. In their paper, Giacomo *et al* compare a staggered multilayer reflector with a conventional quarter-wave stack. Both reflectors have fifteen layers, and the results are quoted for the broadband reflector at 17 000 cm^{-1} and for the conventional reflector at 20 000 cm^{-1}.

Equation (7.75) can be written

$$\sum_i \frac{\partial \phi_a}{\partial d_i} d_i = \sum_i \frac{\partial \phi_a}{\partial \alpha_i} = -4\pi n \nu d_m.$$

Now, from table 7.5,

$$-\sum_i \frac{\partial \phi_a}{\partial \alpha_i} = 30.662$$

and

$$4\pi n \nu d_m = 34.5$$

so that on the preliminary basis of equation (7.75) the prospects look extremely good. However this is not a sufficient condition. We must calculate the actual relationship between β and k and compare it with the theoretical condition given by equation (7.73). Now

$$A_i = d_i \frac{\partial \phi}{\partial d_i} = \frac{\partial \phi}{\partial \alpha_i}$$

which is the last column given for each reflector. This can be used in equation (7.71) giving for a filter using the broadband reflector

$$\beta = 1.023k$$

which can be compared with the value obtained in the same way for the conventional quarter-wave stack of table 7.5:

$$\beta = 1.289k.$$

For a total filter thickness of 2.35 μm the theoretical minimum value of β is given by (7.73) as

$$\beta = 0.652k$$

(k having units of $\mu m^{1/2}$).

Thus, although the phase dispersion filter using the reflectors shown in table 7.5 appears to be promising on the basis of the criterion (7.75), in the event its

Band-pass

Layer number	Thickn (μr
Substrate	—
1	0.07
2	0.12
3	0.07
4	0.12
5	0.06
6	0.12
7	0.06
8	0.09
9	0.05
10	0.08
11	0.05
12	0.08
13	0.04
14	0.07
15	0.04
Medium of incidence	—
$\sum$	1.19

After Giacomo *et al*

performance is son
the straightforwar
condition of equat
Some otherwise
figure 7.21. This filt

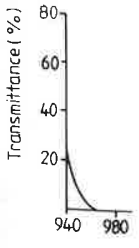


Figure 7.21 The n
design is given in ta

Table 7.5

Layer number	Broadband film				Classical film		
	Thickness d_i (μm)	Index n	$\partial\phi/\partial d_i$ (μm^{-1})	$\partial\phi/\partial\alpha_i$	Thickness d_i (μm)	$\partial\phi/\partial d_i$ (μm^{-1})	$\partial\phi/\partial\alpha_i$
Substrate	—	1.52	—	—	—	—	—
1	0.0751	2.30	0.32	0.024	0.0543	0.01	0.001
2	0.1279	1.35	0.60	0.076	0.0926	0.02	0.002
3	0.0751	2.30	1.97	0.148	0.0543	0.05	0.003
4	0.1235	1.35	1.85	0.229	0.0926	0.06	0.005
5	0.0626	2.30	4.75	0.298	0.0543	0.16	0.009
6	0.1299	1.35	4.60	0.597	0.0926	0.16	0.015
7	0.0681	2.30	11.68	0.795	0.0543	0.48	0.026
8	0.0957	1.35	10.63	1.018	0.0926	0.48	0.044
9	0.0566	2.30	30.85	1.746	0.0543	1.39	0.075
10	0.0859	1.35	30.37	2.608	0.0926	1.39	0.128
11	0.0504	2.30	78.33	3.948	0.0543	4.03	0.219
12	0.0805	1.35	62.33	5.019	0.0926	4.03	0.373
13	0.0450	2.30	121.58	5.471	0.0543	11.69	0.635
14	0.0767	1.35	65.41	5.015	0.0926	11.69	1.082
15	0.0450	2.30	81.59	3.672	0.0543	33.92	1.843
Medium of incidence	—	1.35	—	—	—	—	—
$\sum$	1.1978	—	506.8	30.662	1.0829	69.53	4.460

After Giacomo *et al*[27].

performance is somewhat disappointing. It is, however, certainly better than the straightforward classical filter. So far no design which better meets the condition of equation (7.72) has been proposed.

Some otherwise unpublished results obtained by Ritchie[28] are shown in figure 7.21. This filter used zinc sulphide and cryolite as the materials on glass as

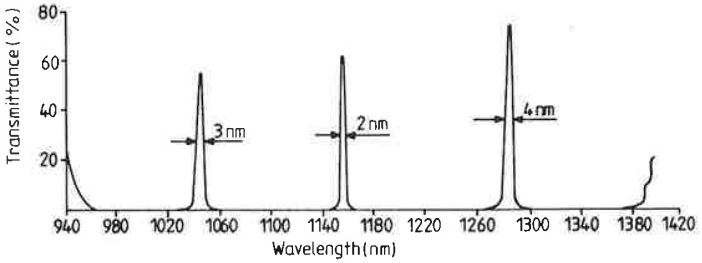


Figure 7.21 The measured transmittance of a 35-layer phase-dispersion filter. The design is given in table 7.5. (After Ritchie[28].)

Table 7.6

Layer number	Material	Optical thickness as fraction of monitoring wavelength
1	ZnS	0.2375
2	Na_3AlF_6	0.2257
3	ZnS	0.2143
4	Na_3AlF_6	0.2036
5	ZnS	0.1934
6	Na_3AlF_6	0.1838
7	ZnS	0.1746
8	Na_3AlF_6	0.1649
9	ZnS	0.1576
10	Na_3AlF_6	0.1498
11	ZnS	0.1423
12	Na_3AlF_6	0.1352
13	ZnS	0.1285
14	Na_3AlF_6	0.1220
15	ZnS	0.1159
16	Na_3AlF_6	0.1101
17	ZnS	0.1046
Spacer	Na_3AlF_6	0.5000

These seventeen layers are followed by another seventeen layers which are a mirror image of the first seventeen.

substrate. Its design is given in table 7.6. An experimental filter monitored at 1.348 μm gave peaks with corresponding bandwidths of

1.047 μm, bandwidth 3.0 nm
1.159 μm, bandwidth 2.5 nm
1.282 μm, bandwidth 4.0 nm.

Theoretically, the bandwidths should have been 0.8 nm, 1.7 nm and 4.6 nm respectively.

MULTIPLE CAVITY METAL–DIELECTRIC FILTERS

Metal–dielectric filters are indispensable in suppressing the longwave sidebands of narrowband all-dielectric filters, and as filters in their own right, especially in the extreme shortwave region of the spectrum. Unlike all-dielectric filters, however, they possess the disadvantage of high intrinsic absorption. In single Fabry–Perot filters this means that the pass bands must be wide in order

to achieve reasonable peak transmission and the shape is far from ideal. It is possible to combine metal–dielectric elements into multiple cavity filters which, because of their more rectangular shape, are more satisfactory but, again, losses can be high.

The accurate design procedure for such metal–dielectric filters can be lengthy and tedious and frequently they are simply designed by trial and error as they are manufactured. We have already mentioned the metal–dielectric Fabry–Perot filter. These filters may be coupled together simply by depositing them one on top of the other with no coupling layer in between.

We can illustrate this by choosing silver as our metal, which we can give an index of $0.055 - i3.32$ at 550 nm[29]. The thickness of the spacer layer in the Fabry–Perot filter, as we have already noted, should be rather thinner than a half-wave at the peak wavelength to allow for the phase changes in reflection at the silver/dielectric interfaces. This phase change varies only slowly with silver thickness when it is thick enough to be useful as a reflector and we can assume, as a reasonable approximation, that it is equal to the limiting value for infinitely thick material. We can then use equation (4.5) to calculate the thickness of the spacer layer. Equation (4.5) calculates for us exactly one half of the filter because it gives the thickness of the dielectric material to yield real admittance with zero phase change at the outer surface of the metal–dielectric combination. Adding a second exactly similar structure with the two dielectric layers facing each other, so that they join to form a single spacer, yields a Fabry–Perot filter in which the phase condition, equation (7.2), is satisfied.

Let us choose a spacer of index 1.35, similar to that of cryolite. Then half the spacer thickness is given by

$$D_f = \frac{1}{4\pi} \tan^{-1}\left(\frac{2\beta n_f}{n_f^2 - \alpha^2 - \beta^2}\right) \tag{7.76}$$

where $\alpha - i\beta$ is the index of the metal and n_f that of the cryolite and the angle is in the first or second quadrant.

With $\alpha - i\beta = 0.055 - i3.32$ and $n_f = 1.35$ we find

$$D_f = 0.188\,55$$

so that the spacer thickness should be 0.3771 full waves.

We can choose a metal layer thickness of 35 nm, quite arbitrarily, simply for the sake of illustration. Our Fabry–Perot filter is then

$$\text{Glass} \left| \begin{matrix} \text{Ag} \\ 35\text{ nm} \end{matrix} \right| \begin{matrix} \text{Cryolite} \\ 0.3771 \text{ full waves} \end{matrix} \left| \begin{matrix} \text{Ag} \\ 35\text{ nm} \end{matrix} \right| \text{Glass}$$

(the geometrical thickness being quoted for the silver and the optical thickness for the cryolite) and the DHW filter is exactly double this structure:

$$\text{Glass} \left| \begin{matrix} \text{Ag} \\ 35\text{ nm} \end{matrix} \right| \begin{matrix} \text{Cryolite} \\ D = 0.3771 \end{matrix} \left| \begin{matrix} \text{Ag} \\ 70\text{ nm} \end{matrix} \right| \begin{matrix} \text{Cryolite} \\ D = 0.3771 \end{matrix} \left| \begin{matrix} \text{Ag} \\ 35\text{ nm} \end{matrix} \right| \text{Glass.}$$

Curves of these filters are shown in figure 7.22. The peaks are slightly displaced from 550 nm because of the approximations inherent in the design procedure.

The Fabry–Perot has reasonably good peak transmission but its typical triangular shape means that its rejection is quite poor even at wavelengths far from the peak. The DHW filter has better shape but rather poorer peak transmittance. The rejection can be improved by increasing the metal thickness, but at the expense of peak transmission.

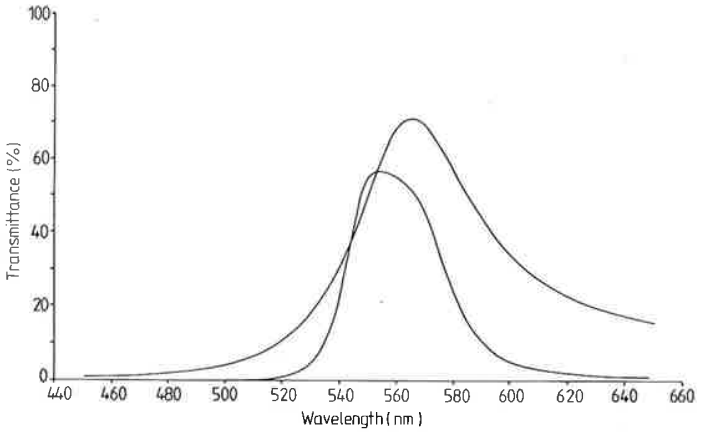


Figure 7.22 The transmittance as a function of wavelength of filters of design:

$$\text{Glass}\left|\begin{matrix}\text{Silver}\\ 35\text{ nm}\end{matrix}\right|\left|\begin{matrix}\text{Cryolite}\\ 0.3771\lambda_0\end{matrix}\right|\left|\begin{matrix}\text{Silver}\\ 35\text{ nm}\end{matrix}\right|\text{Glass}$$

and

$$\text{Glass}\left|\begin{matrix}\text{Silver}\\ 35\text{ nm}\end{matrix}\right|\left|\begin{matrix}\text{Cryolite}\\ 0.3771\lambda_0\end{matrix}\right|\left|\begin{matrix}\text{Silver}\\ 70\text{ nm}\end{matrix}\right|\left|\begin{matrix}\text{Cryolite}\\ 0.3771\lambda_0\end{matrix}\right|\left|\begin{matrix}\text{Silver}\\ 35\text{ nm}\end{matrix}\right|\text{Glass}$$

where $\lambda_0 = 550$ nm, $n - ik = 0.055 - i3.32$ and $n_{\text{cryolite}} = 1.35$. Dispersion in the materials has been neglected.

The design approach we have described is quite crude and simply concentrates on ensuring that the peak of the filter is centred near the desired wavelength. Peak transmittance and bandwidth are either accepted as they are or a new metal thickness is tried. Performance is in no way optimised.

The unsatisfactory nature of this design procedure led Berning and Turner[30] to develop a new technique for the design of metal–dielectric filters in which the emphasis is on ensuring that maximum transmittance is achieved in the filter pass band. For this purpose they devised the concept of potential transmittance and created a new type of metal–dielectric filter known as the induced-transmission filter.

The induced-transmission filter

Given a certain thickness of metal in a filter, what is the maximum possible peak transmission, and how can the filter be designed to realise this transmission? This is the basic problem tackled and solved by Berning and Turner[30]. The development of the technique as given here is based on their approach, but it has been adjusted and adapted to conform more nearly to the general pattern of this book.

The concept of potential transmittance has already been touched on in chapter 2 and used in the analysis of losses in dielectric multilayers. We recall that the potential transmittance ψ of a layer or assembly of layers is defined as the ratio of the intensity leaving the rear surface to that actually entering at the front surface, and it represents the transmittance which the layer or assembly of layers would have if the reflectance of the front surface were reduced to zero. Thus, once the parameters of the metal layer are fixed, the potential transmittance is determined entirely by the admittance of the structure at the exit face of the layer. Furthermore, it is possible to determine the particular admittance which gives maximum potential transmittance. To achieve this transmittance it is sufficient to add a coating to the front surface to reduce the reflectance to zero. The maximum potential transmittance is a function of the thickness of the metal layer.

The design procedure is then as follows: the optical constants of the metal layer at the peak wavelength are given. Then the metal layer thickness is chosen and the maximum potential transmittance together with the matching admittance at the exit face of the layer which is required to produce that level of potential transmittance is found. Often a minimum acceptable figure for the maximum potential transmittance will exist and that will put an upper limit on the metal layer thickness. A dielectric assembly which will give the correct matching admittance when deposited on the substrate must then be designed. The filter is then completed by the addition of a dielectric system to match the front surface of the resulting metal–dielectric assembly to the incident medium. Techniques for each of these steps will be developed. The matching admittances for the metal layer are such that the dielectric stacks are efficient in matching over a limited region only, outside which their performance falls off rapidly. It is this rapid fall in performance that defines the limits of the pass band of the filter.

Before we can proceed further, we require some analytical expressions for the potential transmittance and for the matching admittance. This leads to some lengthy and involved analysis, which is not difficult but rather time-consuming.

(*a*) *Potential transmittance*

We limit the analysis to an assembly in which there is only one absorbing layer, the metal. The potential transmittance is then related to the matrix for the

assembly, as shown in chapter 2

$$\begin{bmatrix} B_i' \\ C_i' \end{bmatrix} = [M] \begin{bmatrix} 1 \\ Y_e \end{bmatrix}$$

where $[M]$ is the characteristic matrix of the metal layer and Y_e is the admittance of the terminating structure. Then the potential transmittance ψ is given by

$$\psi = \frac{T}{(1-R)} = \frac{\mathrm{Re}(Y_e)}{\mathrm{Re}(B_i' C_i'^*)}. \tag{7.77}$$

Let

$$Y_e = X + \mathrm{i}Z.$$

Then

$$\begin{bmatrix} B_i' \\ C_i' \end{bmatrix} = \begin{bmatrix} \cos\delta & (\mathrm{i}\sin\delta)/y \\ \mathrm{i}y\sin\delta & \cos\delta \end{bmatrix} \begin{bmatrix} 1 \\ X + \mathrm{i}Z \end{bmatrix}$$

where

$$\begin{aligned} \delta &= 2\pi(n - \mathrm{i}k)d/\lambda = 2\pi nd/\lambda - \mathrm{i}2\pi kd/\lambda \\ &= \alpha - \mathrm{i}\beta \\ \alpha &= 2\pi nd/\lambda \\ \beta &= 2\pi kd/\lambda. \end{aligned}$$

If free space units are used, then

$$y = n - \mathrm{i}k.$$

Now,

$$\begin{aligned} (B_i' C_i'^*) &= [\cos\delta + \mathrm{i}(\sin\delta/y)(X + \mathrm{i}Z)][\mathrm{i}y\sin\delta + \cos\delta(X + \mathrm{i}Z)]^* \\ &= [\cos\delta + \mathrm{i}(\sin\delta/y)(X + \mathrm{i}Z)][-\mathrm{i}y^*\sin\delta^* + \cos\delta^*(X - \mathrm{i}Z)] \\ &= -\mathrm{i}y^*\cos\delta\sin\delta^* + \frac{\sin\delta\sin\delta^* y^{*2}(X + \mathrm{i}Z)}{yy^*} \\ &\quad + \cos\delta\cos\delta^*(X - \mathrm{i}Z) + \frac{\mathrm{i}\sin\delta\cos\delta^* y^*(X - \mathrm{i}Z)(X + \mathrm{i}Z)}{yy^*}. \end{aligned}$$

We require the real part of this and we take each term in turn.

$$-\mathrm{i}y^*\cos\delta\sin\delta^* = -\mathrm{i}(n + \mathrm{i}k)(\cos\alpha\cosh\beta + \mathrm{i}\sin\alpha\sinh\beta)(\sin\alpha\cosh\beta + \mathrm{i}\cos\alpha\sinh\beta)$$

and the real part of this, after a little manipulation, is

$$\mathrm{Re}(-\mathrm{i}y^*\cos\delta\sin\delta^*) = n\sinh\beta\cosh\beta + k\cos\alpha\sin\alpha.$$

Similarly

$$\mathrm{Re}\left(\frac{\sin\delta\sin\delta^* y^{*2}(X + \mathrm{i}Z)}{yy^*}\right) = \frac{X(n^2 - k^2) - 2nkZ}{(n^2 + k^2)}(\sin^2\alpha\cosh^2\beta + \cos^2\alpha\sinh^2\beta)$$

$$\mathrm{Re}[\cos\delta\cos\delta^*(X - \mathrm{i}Z)] = X(\cos^2\alpha\cosh^2\beta + \sin^2\alpha\sinh^2\beta)$$

$$\mathrm{Re}\left(\frac{\mathrm{i}\sin\delta\cos\delta^* y^*(X-\mathrm{i}Z)(X+\mathrm{i}Z)}{yy^*}\right)$$

$$=\frac{X^2+Z^2}{(n^2+k^2)}(n\sinh\beta\cosh\beta-k\sin\alpha\cos\alpha).$$

The potential transmittance is then

$$\psi=\left(\frac{(n^2-k^2)-2nk(Z/X)}{(n^2+k^2)}(\sin^2\alpha\cosh^2\beta+\cos^2\alpha\sinh^2\beta)\right.$$
$$+(\cos^2\alpha\cosh^2\beta+\sin^2\alpha\sinh^2\beta)$$
$$+\frac{1}{X}(n\sinh\beta\cosh\beta+k\cos\alpha\sin\alpha)$$
$$\left.+\frac{X^2+Z^2}{X(n^2+k^2)}(n\sinh\beta\cosh\beta-k\cos\alpha\sin\alpha)\right)^{-1}. \tag{7.78}$$

(b) *Optimum exit admittance*

Next we find the optimum values of X and Z. From equation (7.78)

$$\frac{1}{\psi}=\left(\frac{q[n^2-k^2-2nk(Z/X)]}{[n^2+k^2]}+r+\frac{p}{X}+\frac{s(X^2+Z^2)}{X(n^2+k^2)}\right) \tag{7.79}$$

where p, q, r and s are shorthand for the corresponding expressions in equation (7.78). For an extremum in ψ, we have an extremum in $1/\psi$ and hence

$$\frac{\partial}{\partial X}\left(\frac{1}{\psi}\right)=0 \qquad \text{and} \qquad \frac{\partial}{\partial Z}\left(\frac{1}{\psi}\right)=0$$

i.e.

$$\frac{q2nkZ}{X^2(n^2+k^2)}-\frac{p}{X^2}+\frac{s}{(n^2+k^2)}-\frac{sZ^2}{X^2(n^2+k^2)}=0 \tag{7.80}$$

and

$$\frac{q(-2nk)}{X(n^2+k^2)}+\frac{2sZ}{X(n^2+k^2)}=0. \tag{7.81}$$

From equation (7.81):

$$Z=nkq/s$$

and, substituting in equation (7.80),

$$X^2=p(n^2+k^2)/s-n^2k^2q^2/s^2.$$

Then, inserting the appropriate expressions for p, q and s, from equation (7.79)

$$X=\left(\frac{(n^2+k^2)(n\sinh\beta\cosh\beta+k\sin\alpha\cos\alpha)}{(n\sinh\beta\cosh\beta-k\sin\alpha\cos\alpha)}\right.$$
$$\left.-\frac{n^2k^2(\sin^2\alpha\cosh^2\beta+\cos^2\alpha\sinh^2\beta)^2}{(n\sinh\beta\cosh\beta-k\sin\alpha\cos\alpha)^2}\right)^{1/2} \tag{7.82}$$

$$Z = \frac{nk(\sin^2\alpha\cosh^2\beta + \cos^2\alpha\sinh^2\beta)}{(n\sinh\beta\cosh\beta - k\sin\alpha\cos\alpha)}. \tag{7.83}$$

We note that for β large $X \to n$ and $Z \to k$, that is:

$$Y_e \to (n + ik) = (n - ik)^*.$$

(c) *Maximum potential transmittance*

The maximum potential transmittance can then be found by substituting the values of X and Z, calculated by equations (7.82) and (7.83), into equation (7.78). All these calculations are best performed by computer or calculator and so there is little advantage in developing a separate analytical solution for maximum potential transmittance.

(d) *Matching stack*

We have to devise an assembly of dielectric layers which, when deposited on the substrate, will have an equivalent admittance of

$$Y = X + iZ.$$

This is illustrated diagramatically in figure 7.23 where a substrate of admittance $(n_s - ik_s)$ has an assembly of dielectric layers terminating such that the final equivalent admittance is $(X + iZ)$. Now, the dielectric layer circles are executed in a clockwise direction always. If we therefore reflect the diagram in the x-axis and then reverse the direction of the arrows, we get exactly the same set of circles—that is, the layer thicknesses are exactly the same—but the order is reversed (it was ABC and is now CBA) and they match a starting admittance of $X - iZ$, i.e. the complex conjugate of $(X + iZ)$, into a terminal admittance of $(n_s + ik_s)$, i.e. the complex conjugate of the substrate index. In our filters the substrate will have real admittance, i.e. $k_s = 0$, and it is a more straightforward problem to match $(X - iZ)$ into n_s than n_s into $(X + iZ)$.

There is an infinite number of possible solutions, but the simplest involves adding a dielectric layer to change the admittance $(X - iZ)$ into a real value and then to add a series of quarter-waves to match the resultant real admittance into the substrate. We will illustrate the technique shortly with several examples. At the moment we recall that the necessary analysis was carried out in chapter 4. There we showed that a film of optical thickness D given by

$$D = \frac{1}{4\pi}\tan^{-1}\left(\frac{2Zn_f}{(n_f^2 - X^2 - Z^2)}\right) \tag{7.84}$$

(where the tangent is taken in the first or second quadrant) will convert an admittance $(X - iZ)$ into a real admittance of value

$$\mu = \frac{2Xn_f^2}{(X^2 + Z^2 + n_f^2) + [(X^2 + Z^2 + n_f^2)^2 - 4X^2n_f^2]^{1/2}}. \tag{7.85}$$

(7.83)

d by substituting the
(7.83), into equation
uter or calculator and
nalytical solution for

h, when deposited on

here a substrate of
terminating such that
lectric layer circles are
reflect the diagram in
e get exactly the same
same—but the order
a starting admittance
erminal admittance of
dex. In our filters the
more straightforward
Z).
the simplest involves
$-iZ$) into a real value
ch the resultant real
echnique shortly with
ecessary analysis was
of optical thickness D

(7.84)

drant) will convert an

$\overline{+X^2n_f^2]^{1/2}}$. (7.85)

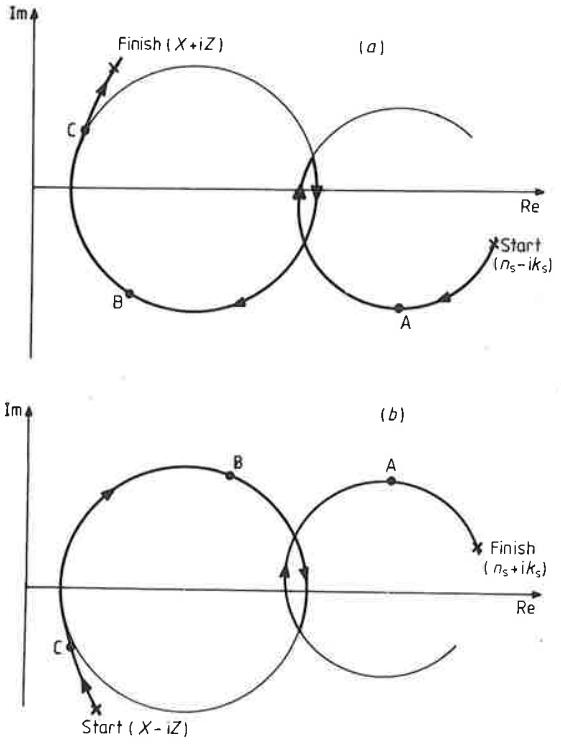


Figure 7.23 (*a*) A sketch of the admittance diagram of an arbitrary dielectric assembly of layers matching a starting admittance of $(n_s - ik_s)$ to the final admittance of $(X + iZ)$. (*b*) The curves of figure 7.23(*a*) reflected in the real axis and with the directions of the arrows reversed. This is now a multilayer identical to (*a*) but in the opposite order and connecting an admittance of $(X - iZ)$ (i.e. $(X + iZ)^*$) to one of $(n_s + ik_s)$ (i.e. $(n_s - ik_s)^*$).

n_f can be of high or low index, but μ will always be lower than the index of the substrate (except in very unlikely cases) because it is the first intersection of the locus of n_f with the real axis which is given by equations (7.84) and (7.85). Since the substrate will always have an index greater than unity, then the quarter-wave stack to match μ to n_s must start with a quarter-wave of low index. Alternate high- and low-index layers follow, the precise number being found by trial and error.

In order to complete the design, we need to know the equivalent admittance at the front surface of the metal layer and then we construct a matching stack to match it to the incident medium.

(*e*) *Front surface equivalent admittance*

If the admittance of the structure at the exit surface of the metal layer is the

optimum value $(X + iZ)$ given by equations (7.82) and (7.83), then it can be shown that the equivalent admittance which is presented by the front surface of the metal layer is simply the complex conjugate $(X - iZ)$. The analytical proof of this requires a great deal of patience, although it is not particularly difficult. Instead, let us use a logical justification.

Consider a filter consisting of a single metal layer matched on either side to the surrounding media by dielectric stacks. Let the transmittance of the assembly be equal to the maximum potential transmittance and let the admittance of the structure at the rear of the metal layer be the optimum admittance $(X + iZ)$ given by equations (7.82) and (7.83). Let the equivalent admittance at the front surface be $(\xi + i\eta)$ and let this be matched perfectly to the incident medium. Now we know that the transmittance is the same regardless of the direction of incidence. Let us turn the filter around, therefore, so that the transmitted light proceeds in the opposite direction. The transmittance of the assembly must be the maximum potential transmittance once again. The admittance of the structure at what was earlier the input, but is now the new exit face of the metal layer, must therefore be $(X + iZ)$. But, since the layers are dielectric and the medium is of real admittance, this must also be the complex conjugate of $(\xi + i\eta)$, that is, $(\xi - i\eta)$. $(\xi + i\eta)$ must therefore be $(X - iZ)$, which is what we set out to prove.

The procedure for matching the front surface to the incident medium is therefore exactly the same as that for the rear surface and, indeed, if the incident medium is identical to the rear exit medium, as in a cemented filter assembly, then the front dielectric section can be an exact repetition of the rear.

Examples of filter designs

We can now attempt some filter designs. We choose the same material, silver, as we did for the Fabry–Perot and the DHW filters earlier. Once again, arbitrarily, we select a thickness of 70 nm. The wavelength we retain as 550 nm, at which the optical constants of silver are $0.055 - i3.32$.

The filter is to use dielectric materials of indices 1.35 and 2.35 corresponding to cryolite and zinc sulphide respectively. The substrate is glass, $n = 1.52$, and the filter will be protected by a cemented cover slip so that we can also use $n = 1.52$ for the incident medium.

$$\alpha = 2\pi nd/\lambda = 0.043\,98$$

$$\beta = 2\pi kd/\lambda = 2.6549$$

and from equations (7.82) and (7.83) we find the optical admittance

$$X + iZ = 0.4572 + i\,3.4693.$$

Substituting this in equation (7.78) gives

$$\psi = 80.50\,\%.$$

We can choose to have either a high- or a low-index spacer. Let us choose first a low index and from equation (7.84) we obtain an optical thickness for the 1.35 index layer of 0.19174 full waves. Equation (7.85) yields a value of 0.05934 for μ which must be matched to the substrate index of 1.52. We start with a low-index quarter-wave and simply work through the sequence of possible admittances:

$$\frac{n_L^2}{\mu}, \quad \frac{n_H^2\mu}{n_L^2}, \quad \frac{n_L^4}{n_H^2\mu}, \quad \frac{n_H^4\mu}{n_L^4} \quad \text{etc}$$

until we find one sufficiently close to 1.52. The best arrangement in this case involves three layers of each type.

$$\frac{n_H^6\mu}{n_L^6} = 1.6511$$

equivalent to a loss of 0.2% at the interface with the substrate.

The structure so far is then

$$|\text{Ag}|\, L''\, LHLHLH\,|\text{Glass} \tag{7.86}$$

with $L'' = 0.19174$ full waves. This can be combined with the following L layer into a single layer $L' = 0.25 + 0.19174 = 0.44174$ full waves, i.e.

$$|\text{Ag}|\, L'\, HLHLH\,|\text{Glass}.$$

Since the medium is identical to the substrate then the matching assembly at the front will be exactly the same as that at the rear so that the complete design is

$$\text{Glass}|\, HLHLH\; L'\,\text{Ag}\; L'\; HLHLH\,|\,\text{Glass}$$

with

Ag	70 nm (geometrical thickness)
L'	0.44174 full waves (optical thickness)
H, L	0.25 full waves
λ_0	550 nm.

The performance of this design is shown in figure 7.24(*a*). Dispersion of the silver has not been taken into account to give a clearer idea of the intrinsic characteristics. The peak is indeed centred at 550 nm with transmittance virtually that predicted.

A high-index matching layer can be handled in exactly the same way. For an index of 2.35, equation (7.84) yields an optical thickness of 0.1561 and equation (7.85) gives a value of 0.1426 for μ. Again, the matching quarter-wave stack should start with a low-index layer. There are two possible arrangements, H' representing 0.1561 full waves.

$$\text{Ag}\; H'\; LHLH\,|\,\text{Glass} \tag{a}$$

with $n_H^4\mu/n_L^4 = 1.310$, i.e. a loss of 0.6% at the glass interface, or

$$\text{Ag } H' \, LHLH \,|\, \text{Glass} \qquad (b)$$

with $n_L^6/n_H^4\mu = 1.392$ representing a loss of 0.2% at the glass interface.

We choose alternative (*b*) and the full design can then be written

$$\text{Glass}\,|\,HLHL\,H'\,\text{Ag}\,H'\,LHLH\,|\,\text{Glass}$$

with

Ag	70 nm (geometrical thickness)
H'	0.1561 full waves (optical thickness)
H, L	0.25 full waves.

The performance of this design is shown in figure 7.24(*b*), where, again, the dispersion of silver has not been taken into account. Peak transmission is virtually as predicted.

When, however, we plot the performance of any of these designs, including

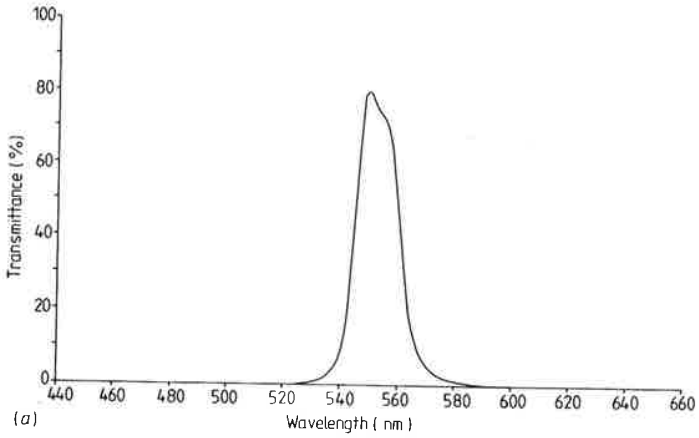


Figure 7.24 (*a*) Calculated performance of the design:

$$\text{Glass}\,|\,HLHLHL'\,\text{Ag}\,L'\,HLHLH\,|\,\text{Glass}$$

where

$n_{\text{Glass}} = 1.52$

Ag = 70 nm	(geometrical thickness) of index	0.055-i3.32
$H = 0.25\,\lambda_0$	(optical thickness) of index	2.35
$L = 0.25\,\lambda_0$	(optical thickness) of index	1.35
$L' = 0.4417\,\lambda_0$	(optical thickness) of index	1.35

$\lambda_0 = 550$ nm.

Dispersion has been neglected.

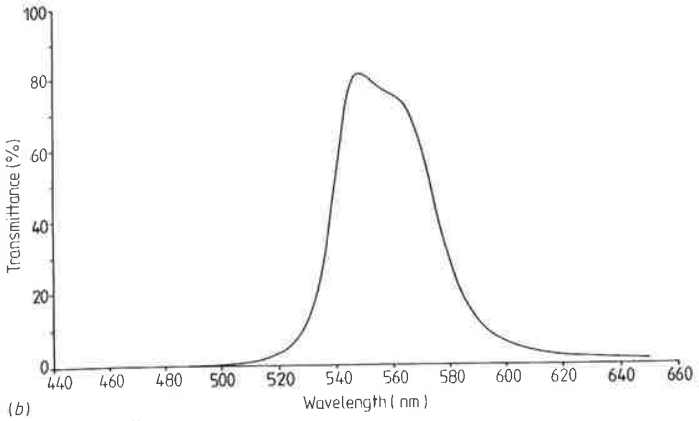


Figure 7.24 (*b*) Calculated performance of the design:

Glass|$HLHL\ H'$ Ag $H'\ LHLH$|Glass

where

$n_{\text{Glass}} = 1.52$		
Ag = 70 nm	(geometrical thickness) of index	0.055-i3.32
$H = 0.25\,\lambda_0$	(optical thickness) of index	2.35
$L = 0.25\lambda_0$	(optical thickness) of index	1.35
$H' = 0.1561\lambda_0$	(optical thickness) of index	2.35
$\lambda_0 = 550$ nm.		

Dispersion has been neglected.

the metal–dielectric Fabry–Perot and DHW filters over an extended wavelength region, we find that the performance at longer wavelengths appears very disappointing. One example, the low-index matched induced-transmission filter, is shown in figure 7.25(*a*). In the case of the Fabry–Perot and the DHW, the rise is smoother, but is of a similar order of magnitude. The reason for the rise is, in fact, our assumption of zero dispersion. This means that β is reduced as λ increases. α is always quite small and the performance of the metal layers is determined principally by β. Silver, however, over the visible and near infrared, shows an increase in k which corresponds roughly to the increase in λ so that k/λ is roughly constant (to within around $\pm 20\%$) over the region 400 nm–2.0 μm. This completely alters the picture and is the reason why the first-order metal–dielectric filters do not show longwave sidebands.

Taking dispersion into account, the performance of the induced transmission filter improves considerably and is shown in figure 7.25(*b*). The rejection is, however, not particularly high, being between 0.01 and 0.1 % transmittance over most of the range with an increase to 0.15 % in the vicinity of 860 nm. This level of rejection can be acceptable in some applications and the induced-transmission filter represents a very useful, inexpensive general purpose filter. The dispersion which improves the performance on the longwave side of the

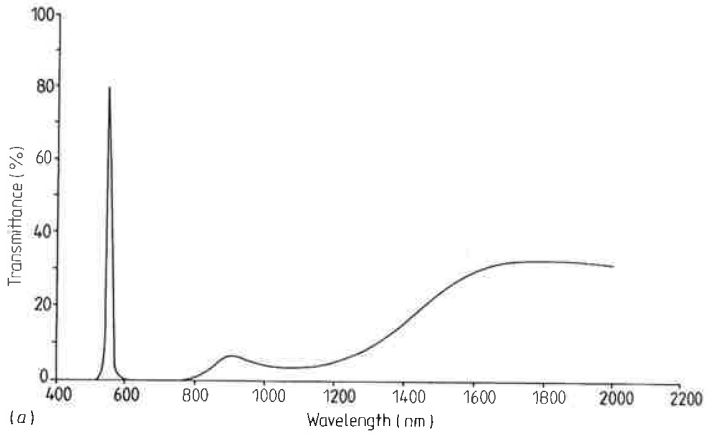


Figure 7.25 (*a*) The design of figure 7.24(*a*) computed over a wider spectral region neglecting dispersion.

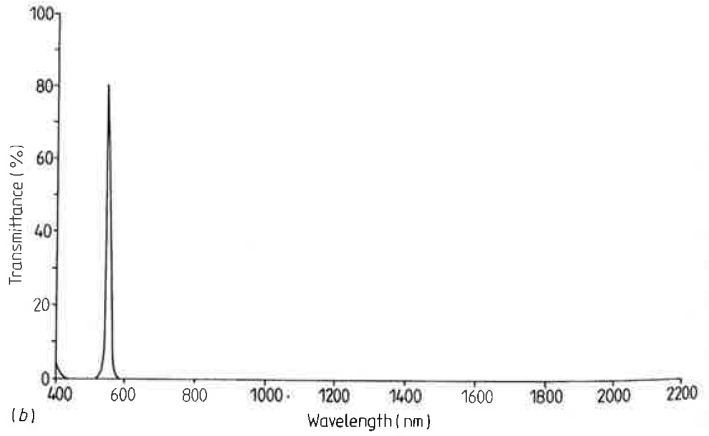


Figure 7.25 (*b*) The design of figure 7.24(*a*) computed this time including dispersion. The rise in transmittance at longer wavelengths has vanished but there is now obvious transmittance at 400 nm.

peak degrades it on the shortwave side, and to complete the filter it is normal to add a longwave-pass absorption glass filter which is cemented to the induced transmission component.

To improve the rejection of the basic filter it is necessary to add further metal layers. The simplest arrangement is to have these extra metal layers of exactly the same thickness as the first. The potential transmittance of the complete filter will then be the product of the potential transmittances of the individual

layers. The terminal admittances for all the metal layers can be arranged to be optimum quite simply, giving optimum performance for the filter. All that is required is a dielectric layer in between the metal layers which is twice the thickness given by equation (7.84) for the first matching layer. We can see why this is by imagining a matching stack on the substrate overcoated with the first metal layer. Since its terminal admittance will be optimum, the input admittance will be the complex conjugate, as we have discussed already. Addition of the thickness given by equation (7.84) renders the admittance real, that is, the admittance locus has reached the real axis. Addition of a further identical thickness must give an equivalent input admittance which is the complex conjugate of the metal input admittance and hence is equal to the optimum admittance. This can be repeated as often as desired.

Returning to our example, a two-metal layer induced-transmission filter will have peak transmittance, if perfectly matched, of $\psi = (0.80501)^2$, that is, 64.8 %, a three-metal layer should have $\psi = (0.805\,01)^3$ that is, 52.17 %, and so on.

The designs, based on the low-index matching layer version, are then, from equation (7.86)

$$\begin{aligned}&\text{Glass}|\,HLHLHL\,L''\,\text{Ag}\,L''L''\,\text{Ag}\,L''\,LHLHLH\,|\text{Glass}\\&=\text{Glass}|\,HLHLHL'\,\text{Ag}\,L'''\,\text{Ag}\,L'\,HLHLH\,|\text{Glass}\end{aligned}\tag{7.87}$$

where

$$\begin{aligned}L' &= 0.25+0.191\,74 = 0.441\,74 \text{ full waves}\\L'' &= 0.191\,74 \text{ full waves}\\L''' &= 2\times 0.191\,74 = 0.383\,48 \text{ full waves}\\\text{Ag} &= 70 \text{ nm}\end{aligned}$$

and

$$\text{Glass}|\,HLHLH\,L'\,\text{Ag}\,L'''\,\text{Ag}\,L'''\,\text{Ag}\,L'\,HLHLH\,|\text{Glass}.\tag{7.88}$$

Unfortunately, these designs, although they do have the peak transmittance predicted, possess a poor pass band shape, in that it has a hump on the longwave side. To eliminate this hump, it is necessary to add an extra half-wave layer to each of the layers marked L''', i.e.

$$\text{Glass}|\,HLHLHL'\,\text{Ag}\,L''''\,\text{Ag}\,L'\,HLHLH\,|\text{Glass}\tag{7.89}$$

and

$$\text{Glass}|\,HLHLH\,L'\,\text{Ag}\,L''''\,\text{Ag}\,L''''\,\text{Ag}\,L'\,HLHLH\,|\text{Glass}\tag{7.90}$$

where

$$L'''' = 0.5+0.383\,48 = 0.883\,48 \text{ full waves.}$$

Figure 7.26 shows the form of designs (7.87) and (7.88) and the hump can clearly be seen together with the improved shape of designs (7.89) and (7.90).

Dispersion was not included in the computation of figure 7.26. To examine

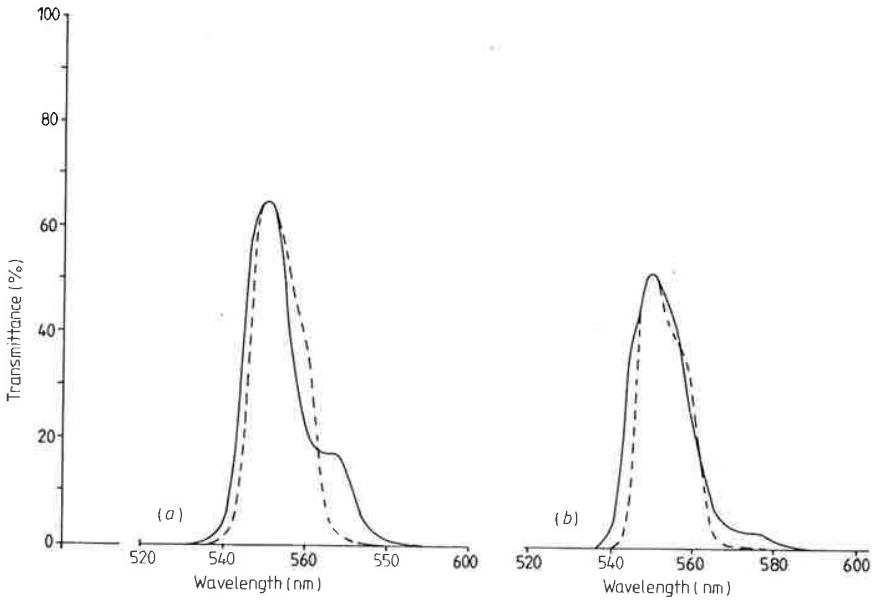


Figure 7.26 Performance, neglecting dispersion, of (*a*) two-metal-layer designs and (*b*) three-metal-layer designs of induced-transmission filter. The full curves denote (7.87) and (7.88) and there is a spurious shoulder on the longwave side of the peak in each case. This can be eliminated by the addition of half-wave decoupling layers as the broken lines show. They are derived from (7.89) and (7.90) respectively.

the rejection over an extended region, we must include the effects of dispersion. Unfortunately, the modified designs (7.89) and (7.90) act as metal–dielectric–metal (M–D–M is a frequently used shorthand notation for such a filter) and metal–dielectric–metal–dielectric–metal (M–D–M–D–M) filters at approximately 1100 nm which gives a very narrow leak, rising to around 0.15% in the former and 0.05% in the latter. Elsewhere, the rejection is excellent, of the order of 0.0001% at 900 nm and 0.000 015% at 1.05 μm for the former and 0.000 0001% at 900 nm and 3×10^{-9}% at 1.05 μm for the latter.

If the leak is unimportant, then the filter can be used as it is with the addition of a longwave-pass filter of the absorption type as before. For the suppression of all-dielectric filter sidebands, it is better to use filters of type (7.87) and (7.88) since the shape of the sides of the pass band is relatively unimportant. The rejection of these filters is slightly better than that of (7.89) and (7.90) and, of course, the leak is missing (figure 7.27).

The bandwidth of the filters is not an easy quantity to predict analytically and the most straightforward approach is simply to compute the filter profile.

Berning and Turner[30] show that a figure of merit indicating the potential

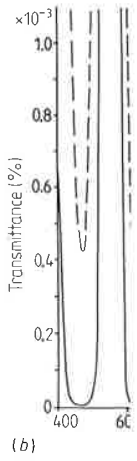
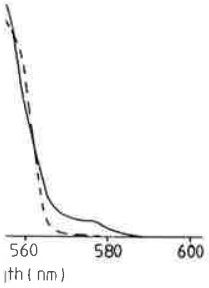


Figure 7.27 (*a*) Ca
of (7.89) (broken c
include the half-wa
shape is the narro
dispersion, of the o
spike is no longer t

usefulness of a m
performance of t
Induced-transm
metal layer are rel
metal layer can b



l-layer designs and
urves denote (7.87)
e peak in each case.
yers as the broken

cts of dispersion.
act as metal–
ation for such a
–D–M) filters at
ising to around
the rejection is
t 1.05 μm for the
ım for the latter.
vith the addition
the suppression
(7.87) and (7.88)
ıimportant. The
nd (7.90) and, of

dict analytically
the filter profile.
ng the potential

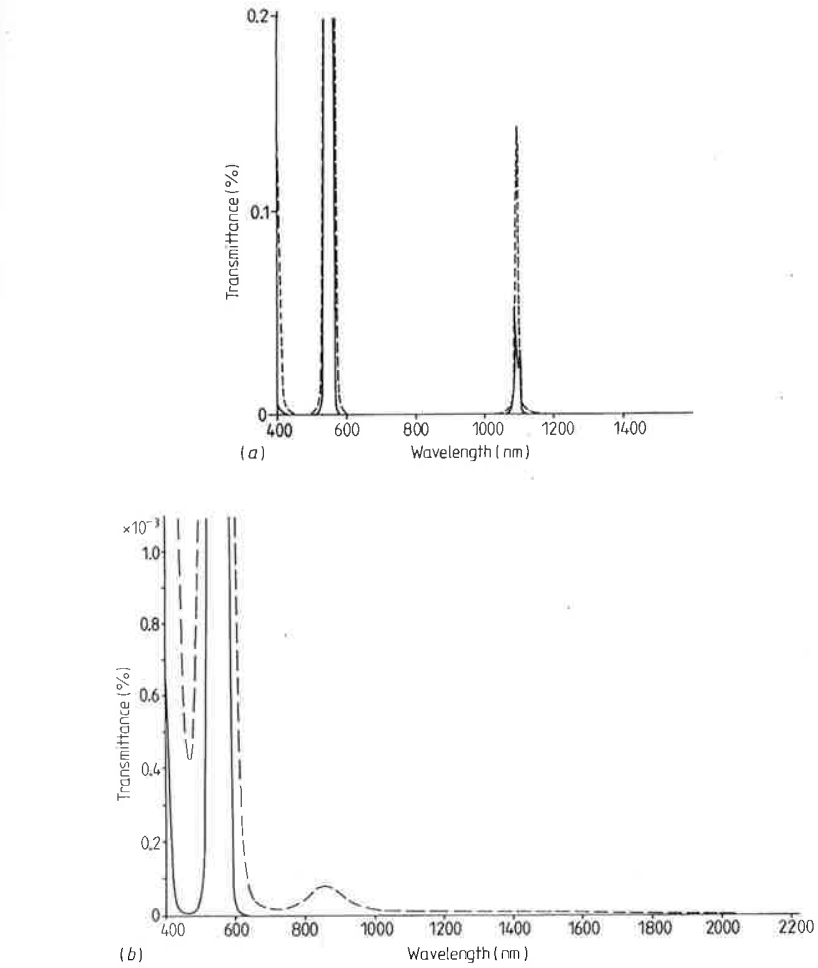


Figure 7.27 (*a*) Calculation, including dispersion, of the performance of the designs of (7.89) (broken curve) and (7.90) over an extended spectral range. These designs include the half-wave decoupling layers and the penalty for the improved pass band shape is the narrow transmission spike near 1.05 μm. (*b*) Calculation, including dispersion, of the original designs (7.87) (broken curve) and (7.88). The transmission spike is no longer there but the pass band shape includes the shoulder (off scale).

usefulness of a metal is the ratio k/n. The higher this ratio, the better is the performance of the completed filter.

Induced-transmission filters for the visible region having only one single metal layer are relatively straightforward to manufacture. The thickness of the metal layer can be arrived at by trial and error. If the metal layer is less than

optimum in thickness, the effect will be a broadening of the pass band and a rise in peak transmission at the expense of an increase in background transmission remote from the peak. A splitting of the pass band will also become noticeable with the appearance eventually, if the thickness is further reduced, of two separate peaks. If, on the other hand, the silver layer is made too thick, the effect will be a narrowing of the peak with a reduction of peak transmission. The best results are usually obtained with a compromise thickness where the peak is still single in shape but where any further reduction in silver thickness would cause the splitting to appear. A good approximation in practice, which can be used as a first attempt at a filter, is to deposit the first dielectric stack and to measure the transmission. The silver layer can then be deposited using a fresh monitor glass so that the optical density is twice that of the dielectric stack. The second spacer and stack can then be added on yet another fresh monitor. A measurement of the transmission of the complete filter will quickly indicate which way the thickness of the silver layer should be altered in order to optimise the design. Usually, one or two tests are sufficient to establish the best parameters. If, after this optimising, the background rejection remote from the peak is found to be unsatisfactory, then not enough silver is being used. As the thickness was chosen to be optimum for the two dielectric sections, a pair of quarter-wave layers should be added to each in the design and the trial and error optimisation repeated. This will also narrow the bandwidth, but this is usually preferable to high background transmission.

In the ultraviolet the available metals do not have as high a performance as, for instance, silver in the visible, and it is very important, therefore, to ensure that the design of a filter is optimised as far as possible; otherwise a very inferior performance will result. An important paper in this field is that by Baumeister *et al*[31]. Aluminium is the metal commonly used for this region and measured and computed results obtained by these workers for filters with aluminium layers are shown in figure 7.28. The performance which has been achieved is most satisfactory and the agreement between practical and theoretical curves is good.

Induced-transmission filters have been the subject of considerable study by many workers. Metal–dielectric multilayers are reviewed by MacDonald[32]. A useful, recent account of induced-transmission filters is given by Lissberger[33]. Multiple cavity induced-transmission filters have been described by Maier[34]. An alternative design technique for metal–dielectric filters involving symmetrical periods has been published by Macleod[35]. Symmetrical periods for metal–dielectric filter design have also been used by McKenney[36] and by Landau and Lissberger[37].

MEASURED FILTER PERFORMANCE

Not a great deal has been published on the measured performance of actual filters and the main source of information for a prospective user is always the

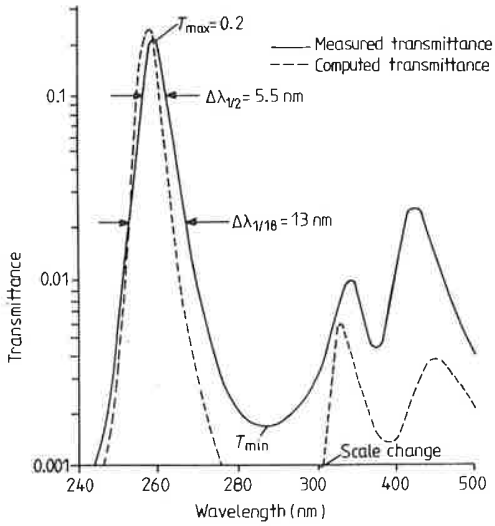


Figure 7.28 Computed and measured transmittance of an induced transmission filter for the ultraviolet. Design:

$$\text{Air}|HLHLHLH\ 1.76L\ \text{Al}\ 1.76L\ HLHLHLH|\text{Quartz}$$

where $H = PbF_2$ ($n_H = 2.0$) and $L = Na_3AlF_6$ ($n_L = 1.36$). The physical thickness of the aluminium layer is 40 nm and $\lambda_0 = 253.6$ nm. (After Baumeister, Costich and Pieper[31].)

literature issued by manufacturers. Performance of current production filters tends to improve all the time so that inevitably such information does not remain up to date for long. Two papers[38,39] quote the results of a number of tests on commercial filters, and, although they were written some time ago, they will still be found useful sources of information.

Blifford examined the performance of the products of four different manufacturers, covering the region 300–1000 nm. The variation of peak wavelength with angle of incidence was found to be similar to the relationship already established (see p 260). Unfortunately, information on the design and materials is lacking, so that the expression for the effective index cannot be checked. The sensitivities to tilt varied from $P = 0.22$ to $P = 0.51$, where P corresponds to the quantity $1/n^{*2}$ in equation (7.39). Blifford suggests that an average value of 0.35 for P would probably be the best value to assume in any case where no other data were available. Changes in peak transmittance with angle of incidence were found, but were not constant from one filter to another and apparently must always be measured for each individual filter. Possibly, the effect is due to the absorption filters which are used for sideband suppression and which, because they do not show any shift in edge wavelength with angle of incidence, may cut into the pass band of the interference section

at large angles of incidence. In most cases examined, the change in peak transmission was less than 10% for angles of 5°–10°.

The variation in peak transmittance over the surface of the filter was also measured in a few cases. For a typical filter with a peak wavelength of 500 nm and a bandwidth not explicitly mentioned, but probably 2.1 nm (from information given elsewhere in the paper), the extremes of peak transmission were 54% and 60%. This is, in fact, one aspect of a variation of peak wavelength, bandwidth and peak transmittance which frequently occurs, although the magnitude can range from very small to very large. The cause is principally the adsorption of water vapour from the atmosphere before a cover slip can be cemented over the layers and it is dealt with in greater detail in chapter 9. Infrared filters appear to suffer less from this defect than visible and near infrared filters.

Another parameter measured by Blifford was the variation of peak wavelength with temperature. Variation of the temperature from -60°C to $+60$°C resulted in changes of peak wavelength of from $+0.01$ nm °C^{-1} to $+0.03$ nm °C^{-1}. The relationship was found to be linear over the whole of this temperature range with little, if any, change in the pass band shape and peak transmittance. In most cases, the temperature coefficients of bandwidth and peak transmittance were found to be less than 0.001 nm °C^{-1}. Filters for the visible region have also been the subject of a detailed study by Pelletier and his colleagues[40]. The shift with temperature for any filter is a function of the coefficients of optical thickness change with temperature, depending on the design of the filter and especially on the material used for the spacers. Measurements made on different filter designs yielded the following coefficients of optical thickness for the individual layer materials:

zinc sulphide	$(4.8 \pm 1.0) \times 10^{-5}$ °C^{-1}
cryolite	$(3.1 \pm 0.7) \times 10^{-5}$ °C^{-1}.

Hysteresis is frequently found when temperature cycling narrowband filters over an extended temperature range. The hysteresis is particularly pronounced when the filters are uncemented and when they are heated towards 100°C. It is usually confined to the first cycle of temperature, takes the form of a shift of peak wavelength towards shorter wavelengths and is caused by the desorption of water which is discussed again in chapter 9.

An effect of a different kind, although related, is the subject of a contribution by Title and his colleagues[41, 42]. A permanent shift of a filter characteristic towards shorter wavelengths amounting to a few tenths of nanometres accompanied by a distortion of pass band shape was produced by a high level of illumination. The filters were for the H_α wavelength, 656.3 nm, and the changes were interpreted as due to a shift in the properties of the zinc sulphide material, the fundamental nature of the shift being unknown. Zinc sulphide

can be transformed int
in the presence of mois
have been caused by s

The possibility of va
filter and as a functio
clearly be borne in mi

A useful survey whic
types of narrowband f

A study was carried
studied were those of
both theoretical and e

Accurate calculatior
yielded a variation of p
variation of bandwidth
that the peak transmit
unchanged for angles u
peak transmittance gr
by measurements on r

The effects of varyin
and practically. As in
Blifford, they measure
temperature. For temp
of peak transmittance
serious losses in transm
in the paper, this is pr
or one of the layer
transmittance at elevat
point that filters desig
incidence are usually
temperature coefficien
$+0.0035$% °C^{-1} to $+$
used in the filters nor th
figures will apply to m

Similar measuremen
at Grubb Parsons. Th
and the filters which
coefficients of peak wa
type used in the select
negative temperature c
of the layer materials.
useful as it tends to con
Seeley *et al*[44] have suc
telluride which have z

can be transformed into zinc oxide by the action of ultraviolet light, especially in the presence of moisture, and the shifts that were observed could probably have been caused by such a mechanism.

The possibility of variations in filter properties both over the surface of the filter and as a function of time, temperature and illumination level should clearly be borne in mind in the designing of apparatus incorporating filters.

A useful survey which compares the performance achievable from different types of narrowband filters was the subject of a report by Baumeister[43].

A study was carried out by Baker and Yen on infrared filters. The effects studied were those of variation in angle of incidence and temperature, and both theoretical and experimental results were quoted.

Accurate calculation of the effects of changes in the angle of incidence yielded a variation of peak wavelength of the expected form, but no significant variation of bandwidth for angles of incidence up to 50°. They also calculated that the peak transmittance and the shape of the pass band should remain unchanged for angles up to 45°. For angles above 50°, both the shape and the peak transmittance gradually deteriorated. The calculations were confirmed by measurements on real filters.

The effects of varying temperatures were also investigated both theoretically and practically. As in the case of the shorter wavelength filters examined by Blifford, they measured a shift towards longer wavelengths with increasing temperature. For temperatures down to liquid helium the filters show little loss of peak transmittance or variation of characteristic pass band shape. However, serious losses in transmittance occurred above 50°C. Although not mentioned in the paper, this is probably due to the use of germanium, either as substrate or one of the layer materials, which always exhibits a marked fall in transmittance at elevated temperatures above 50°C. Baker and Yen make the point that filters designed to be least sensitive to variations in the angle of incidence are usually most sensitive to temperature and vice versa. The temperature coefficients of peak wavelength which they quote vary from $+0.0035\,\%\,°C^{-1}$ to $+0.0125\,\%\,°C^{-1}$. Unfortunately, neither the materials used in the filters nor the designs are quoted in the paper, but it is likely that the figures will apply to most interference filters for the infrared.

Similar measurements of the temperature shift of infrared filters were made at Grubb Parsons. The materials used were zinc sulphide and lead telluride, and the filters which had first-order high-index spacers gave temperature coefficients of peak wavelength of $-0.0135\,\%\,°C^{-1}$. These filters were of the type used in the selective chopper radiometer described in chapter 12. The negative temperature coefficient is usual with filters having lead telluride as one of the layer materials. This negative coefficient in lead telluride is especially useful as it tends to compensate for the positive coefficient in zinc sulphide, and Seeley *et al*[44] have succeeded in designing and constructing filters using lead telluride which have zero temperature coefficient.

REFERENCES

1 Epstein L I 1952 The design of optical filters *J. Opt. Soc. Am.* **42** 806–10
2 Turner A F 1950 Some current developments in multilayer optical films *J. Phys. Radium* **11** 443–60
3 Bates B and Bradley D J 1966 Interference filters for the far ultraviolet (1700 to 2400 Å) *Appl. Opt.* **5** 971–5
4 Seeley J S 1964 Resolving power of multilayer filters *J. Opt. Soc. Am.* **54** 342–6
5 Hemingway D J and Lissberger P H 1973 Properties of weakly absorbing multilayer systems in terms of the concept of potential transmittance *Opt. Acta* **20** 85–96
6 Dobrowolski J A 1959 Mica interference filters with transmission bands of very narrow half-widths *J. Opt. Soc. Am.* **49** 794–806 and 1963 Further developments in mica interference filters *J. Opt. Soc. Am.* **53** 1332 (summary only)
7 Austin R R 1972 The use of solid etalon devices as narrowband interference filters *Opt. Eng.* **11** 65–9
8 Candille M and Saurel J M 1974 Réalisation de filtres "double onde" a bandes passantes très étroites sur supports en matière plastique (mylar) *Opt. Acta.* **21** 947–62
9 Smith S D and Pidgeon C R 1963 Application of multiple beam interferometric methods to the study of CO_2 emission at 15 μm *Mém. Soc. R. Sci. Liège 5ième série* **9** 336–49
10 Roche A E and Title A M 1974 Tilt tunable ultra narrow-band filters for high resolution photometry *Appl. Opt.* **14** 765–70
11 Dufour C and Herpin A 1954 Applications des méthodes matricielles au calcul d'ensembles complexes de couches minces alternées *Opt. Acta* **1** 1–8
12 Lissberger P H 1959 Properties of all-dielectric filters. I—A new method of calculation *J. Opt. Soc. Am.* **49** 121–5
13 Lissberger P H and Wilcock W L 1959 Properties of all-dielectric filters. II—Filters in parallel beams of light incident obliquely and in convergent beams *J. Opt. Soc. Am.* **49** 126–38
14 Pidgeon C R and Smith S D 1964 Resolving power of multilayer filters in non-parallel light *J. Opt. Soc. Am.* **54** 1459–66
15 Hernandez G 1974 Analytical description of a Fabry–Perot spectrometer, 3. Off-axis behaviour and interference filters *Appl. Opt.* **13** 2654–61
16 For example, Reports 4, 5 and 6 of Contract DA-44-009-eng-1113 covering the period January–October 1953
17 Turner A F 1952 Wide pass band multilayer filters *J. Opt. Soc. Am.* **42** 878(*a*)
18 Smith S D 1958 Design of multilayer filters by considering two effective interfaces *J. Opt. Soc. Am.* **48** 43–50
19 Knittl Z 1965 Dielektrische Interferenzfilter mit recheckigen Maximum *Proc. Coll. Thin Films, Budapest* 153–61 (The method is described in detail in reference 20 also)
20 Knittl Z 1976 *Optics of Thin Films* (London: Wiley)
21 Thelen A 1966 Equivalent layers in multilayer filters *J. Opt. Soc. Am.* **56** 1533–8
22 Barr E E 1974 Visible and ultraviolet bandpass filters, in *Optical Coatings, Applications and Utilization* ed G W DeBell and D H Harrison (*Proc. SPIE* **50** 87–118)

23 Neilson R G T and Ring J 1967 Interference filters for the near ultra-violet *J. Physique* **28** C2-270–C2-275 (supplement to no 3–4 March–April)

24 Malherbe A 1974 Interference filters for the far ultraviolet *Appl. Opt.* **13** 1275–6

25 Baumeister P W and Jenkins F A 1957 Dispersion of the phase change for dielectric multilayers. Application to the interference filter *J. Opt. Soc. Am.* **47** 57–61

26 Baumeister P W, Jenkins F A and Jeppesen M A 1959 Characteristics of the phase-dispersion interference filter *J. Opt. Soc. Am.* **49** 1188–90

27 Giacomo P, Baumeister P W and Jenkins F A 1959 On the limiting band width of interference filters *Proc. Phys. Soc.* **73** 480–9

28 Ritchie F S Unpublished work on Ministry of Technology Contract KX/LSO/C.B.70(a)

29 Hass G and Hadley L 1972 Optical constants of metals *American Institute of Physics Hand Book* ed D E Gray (New York: McGraw-Hill) pp 6-124–6-156

30 Berning P H and Turner A F 1957 Induced transmission in absorbing films applied to band pass filter design *J. Opt. Soc. Am.* **47** 230–9

31 Baumeister P W, Costich V R and Pieper S C 1965 Bandpass filters for the ultraviolet *Appl. Opt.* **4** 911–13

32 MacDonald J 1971 Metal–Dielectric Multilayers (London: Adam Hilger)

33 Lissberger P H 1981 Coatings with induced transmission *Appl. Opt.* **20** 95–104

34 Maier R L 1967 2M interference filters for the ultraviolet *Thin Solid Films* **1** 31–7

35 Macleod H A 1978 A new approach to the design of metal–dielectric thin-film optical coatings *Opt. Acta.* **25** 93–106

36 McKenney D B 1969 Ultraviolet interference filters with metal–dielectric stacks *PhD Dissertation* Optical Sciences Center, University of Arizona

37 Landau B V and Lissberger P H 1972 Theory of induced transmission filters in terms of the concept of equivalent layers *J. Opt. Soc. Am.* **62** 1258–64

38 Blifford I H Jr 1966 Factors affecting the performance of commercial interference filters *Appl. Opt.* **5** 105–11

39 Baker M L and Yen V L 1967 The effect of the variation of angle of incidence and temperature on infrared filter characteristics *Appl. Opt.* **6** 1343–51

40 Pelletier F, Roche P and Bertrand L 1974 On the limiting bandwidth of interference filters: influence of temperature during production *Opt. Acta* **21** 927–46

41 Title A M, Pope T P and Andelin J P 1974 Drift in interference filters. 1 *Appl. Opt.* **13** 2675–9

42 Title A M 1974 Drift in interference filters. 2: radiation effects *Appl. Opt.* **13** 2680–4

43 Baumeister P W 1973 Thin films and interferometry *Appl. Opt.* **12** 1993–4

44 Seeley J S, Evans C S, Hunneman R and Whatley A 1976 Filters for the v2 band of CO_2: monitoring and control of layer deposition *Appl. Opt.* **15** 2736–45