17 OPTICAL MATERIAL

17.1 INTRODUCTION

17.1.1 General. The only properties of an optical material which are used in the actual design of a system are its indices of refraction, and the quantities derived therefrom (such as the ν-value). However, frequently the designer must select materials which meet requirements that are not concerned with the quality of imagery, but which are important with respect to satisfactory fabrication and performance of the instrument. Often he must make a compromise between desirable optical properties, and such characteristics as weight, availability, durability and cost.

17.1.2 Coverage. The following discussion of optical materials will involve the properties of refracting materials, and those of reflecting materials.

17.2 REFRACTING MATERIAL CHARACTERISTICS

17.2.1 Transmission. A refracting material, to be useful, obviously must transmit radiation in the wavelength region in which it is to be used, or it may be used to absorb the undesirable radiation of other wavelengths. In some instances, the refracting material transmits imperfectly in the region of use, and the designer must determine what thicknesses (if any) he can use without greatly impairing the performance of the instrument. The amount of light transmitted through a lens or plate is limited by surface reflection, by absorption within the medium, and by diffusion.

17.2.2 Surface reflection. When light is incident on the boundary surface between two refracting media, part of the light is transmitted into the second medium and part is reflected back into the first. The ratio of the reflected light to the incident light, sometimes called the reflection coefficient or reflectance or

$$R = \frac{1}{2}\left[\frac{\sin^2 (I - I')}{\sin^2 (I + I')} + \frac{\tan^2 (I - I')}{\tan^2 (I + I')}\right] , \qquad (1)$$

reflectivity, is where I is the angle of incidence and I' is the angle of refraction. If the light is incident normally, so that $I = 0$, and if one medium is air, or a vacuum with index of refraction = 1, then the

$$R = \left[\frac{n - 1}{n + 1}\right]^2 , \qquad (2)$$

expression reduces to wherein n is the index of refraction of the other medium. This formula is the one most frequently used in estimating reflection losses. Since the reflection coefficient depends on the indices of refraction, and these in general depend on the wavelength, the reflection coefficient itself varies with wavelength. Thus for an extra dense flint, having a refractive index of 1.720 at the sodium D line and 1.673 at 2.5 microns, equation (2) yields reflection coefficients of 7.0% and 6.3% respectively at the two wavelengths. Whether the variation with wavelength is important depends on the application. Many times it is satisfactory to use a single value of the index throughout the range of wavelengths under consideration.

17.2.3 Absorption. When radiant energy passes through a medium other than a vacuum, a portion of it is usually converted into another form of energy. This phenomenon is called absorption, and the energy so absorbed is no longer available for image formation.

NOTE

In this discussion a terminology will be used which is becoming standard. However, the reader is warned that in other publications on this subject the use of terms and symbols may vary considerably with those presented herein. In consulting other published data on optical materials be sure of the exact meaning given to the terms and symbols used, regardless of their apparent similarity to the terms and symbols in this handbook. The basic physical concepts are the same in all cases.

17.2.3.1 According to Bouguer's law, the amount of light not absorbed in the passage through a homogenous medium, i.e., the transmitted light, is a decreasing exponential function of the thickness. That is,

$$W = W_o \; e^{-\alpha t} \qquad (3)$$

In this equation W_o is the original flux. W is the amount remaining unabsorbed after passage through a thickness t of the medium; α is a quantity called the absorption coefficient of the medium. (Note that this equation refers to what happens within the medium - it is not complicated by the reflection losses which occur when the light passes through a boundary surface into or out of the medium.) In general, α is a function of the wavelength.

17.2.3.2 The extinction coefficient κ and the absorption constant k are also commonly used constants. These are related to the absorption coefficient by the equation

$$\kappa = \frac{k}{n} = \frac{\alpha\lambda}{4\pi n} \quad , \tag{4}$$

λ being the wavelength of light and n the index of refraction. For computation purposes it is sometimes convenient to replace the base e in equation (3) by base 10, and write

$$W = W_o \ 10^{-\beta t} \tag{5}$$

This permits using tables of common logarithms. We shall call β the "absorption coefficient to base 10", when necessary for clarity, and simply the "absorption coefficient" when the distinction is clear from the context. Therefore, $\alpha = 2.303\beta$. The quantity,

$$T = \frac{W}{W_o} = e^{-\alpha t} = 10^{-\beta t} \quad . \tag{6}$$

is called the internal transmittance of the thickness t.

17.2.3.3 The absorption characteristics of a material are usually measured by placing a sample in the form of a polished, plane-parallel plate in the beam of a spectrophotometer, and determining what portion of the radiation incident on the first surface of the sample emerges from the second. This determination is made for as many wavelengths as desired. Most modern spectrophotometers automatically draw a curve of transmittance as a function of wavelength. It is evident that the results from the spectrophotometer include the effects of reflection loss and absorption loss. (They also include the effect of loss due to diffusion, but this is usually negligible in comparison with the others.) It is frequently necessary to be able to separate the two effects in order to predict the benefit to be gained by low-reflection coatings, or the absorption to be expected from other thicknesses. Of the light incident on the first surface of the sample, the fraction R is reflected and lost, and the fraction (1-R) passes into the sample. Of this, the fraction T passes through the sample without being absorbed and reaches the second surface. Here a fraction R is reflected and only (1-R) passes through the surface. That is, of the original radiation, the fraction

$$T_1 = T\ (1-R)^2 \tag{7}$$

passes through the second surface. However, the radiation reflected at the second surface passes back through the medium toward the first surface, where a part of it is reflected back into the medium, and so on. Summing over all such passages one obtains for the total fraction T_∞ passing through the second surface of the sample,

$$T_\infty = \frac{(1-R)^2\ T}{1 - T^2\ R^2} \quad . \tag{8}$$

It is the quantity T_∞ (usually expressed as a percent) which is measured by the spectrophotometer.

17.2.3.4 Since R can be determined from the refractive index, the value of T can be computed from equation (8) which can conveniently be rearranged as the quadratic equation:

$$T^2 + \left(\frac{1}{R} - 1\right)^2 \cdot \left(\frac{T}{T_\infty} - \frac{1}{R^2}\right) = 0 \tag{9}$$

If R is not large, the amount of light contributed by the re-passages through the medium may be negligible, and it will then be simpler and satisfactory to use equation (7) for T, using the spectrophotometer reading

for the value of T_1 . Having determined the value T for the sample and knowing its thickness, one can determine the absorption coefficient from equation (7) or equation (9). Then one may compute the fraction which will be transmitted through other thicknesses, with the same or different surface reflectivities.

17.2.3.5 Consider the following example. Spectrophotometer curves were run on a sample of ordinary plate glass, 6.18mm thick, and on a sample of "waterwhite" plate, 6.6mm thick. At the wavelength of 600mμ the former transmitted 88.3% and the latter, 91.4% (the measured transmittances are significant to a few tenths of a percent). Both glasses have an index of refraction of about 1.52 at this wavelength. How would their transmittances compare in thicknesses of 25.4mm, if the surfaces in use are to have low-reflection coatings with reflectivities of 1% ? For the ordinary plate glass

$$R = \left(\frac{0.52}{2.52}\right)^2 = 0.043 .$$

Substituting in equation (7) yields

$$T = 0.964.$$

Use of the more precise equation (9) would give $T = 0.960$ but the difference is hardly significant in view of the limited precision of the initial data. Then,

$$\beta = \frac{-\log 0.964}{6.18\text{mm}} = 0.00257/\text{mm} .$$

The internal transmittance of a 25.4mm thick piece would then be

$$\text{antilog } (-0.00257 \times 25.4) = 0.860 .$$

Since the coated surfaces are to have reflectivities of 1%, substitution in equation (7) shows that 84% of the incident 600mμ radiation would pass through the plate. For the waterwhite plate the surface reflectivity is again 0.043. Substitution in equation (7) yields

$$T = 0.998 .$$

This value is so near unity that the difference is not reliable in view of the limited accuracy of the data from which it was computed. However, it is safe to assume that T_1 is no worse than 0.995 and that hence

$$\beta \leq \frac{-\log 0.995}{6.61\text{mm}} = 0.00033/\text{mm}.$$

Consequently the internal transmittance of the 25.4mm thickness will be at least 0.981 and with the surfaces coated at least 96% of the radiation will pass through the piece.

17.2.4 Diffusion. Some light, in passing through a medium is deviated from its path due to the presence of fine inhomogeneities in the medium. This effect is called diffusion. In extreme cases it causes the medium to be translucent rather than transparent. Some of the diffused light is lost from the optical system. That which remains is not image-forming but, being spread over the image area, reduces contrast. For this discussion it is not necessary to consider the physics of the phenomenon beyond remarking that the amount of scattering is a function of the ratio of the size of the inhomogeneity to the wavelength of light, the amount decreasing as the ratio decreases. (The effect can be noticed in a long line of automobile headlights at night, unless the air is very clear. Since more light is scattered from the blue end of the spectrum, the distant lights look more yellow or orange than the near ones.) As a result, inhomogeneities which would be bothersome in the visible region may be of negligible importance in infrared work.

17.3 REFRACTIVITY AND DISPERSION

17.3.1 Selection of materials.

17.3.1.1 From the available media which transmit satisfactorily in the wavelength region with which he is concerned, the designer must select those with index and dispersion characteristics best suited for his needs. In the visible region, it is usually sufficient for the designer to know the indices of refraction for a few conventionally specified wavelengths, and to do much of his calculations with the quantities derived from them, (such as the ν value and the partial dispersion ratios). Glass-makers' catalogs customarily

describe the refractive properties of the glasses in terms of these standard quantities.

17.3.1.2 In the ultraviolet and infrared regions, procedures and requirements are not so well standardized. Use of standard wavelengths for index measurement, and of dispersion constants based on such measurements, has not become common. It is often necessary to work from such tables of values of refractive index as are available, and to interpolate for the wavelengths one wishes to consider in the design. To aid in selecting either the derivative $dn/d\lambda$ or the related quality

$$\frac{(n-1)}{\frac{dn}{d\lambda}} .$$

The latter is analogous to the ν-value; for achromatism of a thin doublet, the ratio of the powers of the two elements should be the negative of the ratio of the values of $(n-1)/(dn/d\lambda)$ of the two media. Use of the former is similar to the use in the visible region of $n_F - n_C$; the ratio of the total curvatures of the two elements of a thin achromatized doublet should equal the negative reciprocal of the ratio of the two derivatives. Some publications tabulate values of the derivative $dn/d\lambda$.

17.3.2 Optical homogeneity.

17.3.2.1 It is important that the refractive index of a lens or prism be constant throughout the piece. Usually the requirement for uniformity within the piece is more rigid than the requirement for uniformity from piece to piece. There are two principal causes of optical inhomogeneity: chemical inhomogeneity and improper annealing.

17.3.2.2 Since most glassy substances are complex mixtures, rather than precise chemical compounds, it is difficult to make them chemically homogeneous throughout. The presence of streaks of slightly varying composition results in striae. The harm done by striae depends on their loaction in a system. If they are so placed and oriented that all parts of each image-forming ray bundle pass through about the same optical path in the striae, the effect may be negligible. Thus a moderate amount of flat striae, approximately parallel to the plane of a weak lens, may be tolerable. On the other hand, in a reflecting prism in which the beams pass through the same volume in at least two different directions, it is impossible to meet this condition, and material for such prisms should be free of striae.

17.3.3 Mechanical strain. The thermal history of the piece of material may also cause optical inhomogeneity. If it has been such as to result in the presence of mechanical strains, the material becomes locally polarizing. The presence of this defect can be observed by examining the piece between crossed polarizers. However, even if there is no detectable strain, the material still may be optically inhomogeneous, and a more careful fine annealing of the glass, to accomplish the effect known as compacting, is necessary. (The fine annealing is also necessary to bring the index of the glass to its maximum value. Melt sheets supplied by manufacturers generally give index and dispersion measurements made on fine-annealed samples). Heating the glass to softening for making molded blanks, slumpings, etc. cancels out its previous thermal history and necessitates re-annealing to ensure the quality desired.

17.3.4 Optical isotropy and anisotropy.

17.3.4.1 In most cases the designer requires that his materials be optically isotropic - that is, the index of refraction at any point must be constant, regardless of the direction in which the radiation is passing the point. Only occasionally does he want anisotropic or polarizing media. Optical glasses are isotropic, except as made locally anisotropic by improper annealing. Many crystalline materials with otherwise attractive characteristics are anisotropic and hence unsuitable for making lenses and non-polarizing prisms. However, crystals belonging to the cubic system, when free from mechanical strain, are optically isotropic, and some such are noted below.

17.3.4.2 Occasionally a weakly anisotropic material such as sapphire can be used satisfactorily, by orienting its optic axis parallel to the optic axis of the system.

17.3.4.3 A crystal which is isotropic optically is not necessarily so in all mechanical properties, and it may be desirable to have blanks for lens making cut with preferred orientation with respect to the cleavage planes. This will ensure uniform grinding or minimize losses from fracture along the cleavage planes.

17.4 INCLUSIONS

17.4.1 Imperfections in optical materials. Optical materials may have imperfections in the form of tiny opaque or refracting inclusions. In ordinary optical glass the most common of these are called, according to their nature, bubbles, seeds or stones. Other refracting media may have similar imperfections. The

harm done by such defects depends on their position in the optical system. An inclusion near an image plane, as in a field lens or in the plane-parallel plate on which a reticle is formed, will appear as a bothersome out-of-focus object in a visual system, or may give a false signal in an infrared system. The same inclusion in an objective lens may have negligible effect on the performance. Tolerances on such imperfections should be set with the specific use of the part in mind.

17.5 ENVIRONMENTAL CHARACTERISTICS

17.5.1 Optical system requirements. Withing recent years, requirements for the performance of optical systems under extreme environmental conditions, such as in airborned equipment or under exposure to desert, jungle and artic conditions, and also the necessity of using available refracting materials other than ordinary optical glass outside the visible spectrum, have made it necessary for the designer to be conscious of the thermal, mechanical and chemical characteristics of the materials he proposes to use. The most important of these are the following:

(1) Softening characteristics - cold flow. A lens or window obviously should hold its shape, including the figure of its refracting surfaces, under storage and service conditions. The softening temperature of the material should be high enough to ensure that this will be the case. A few materials exhibit the phenomenon of cold flow, a tendency to deform even at ordinary temperatures.

(2) Resistance to thermal shock. Materials vary in their ability to undergo rapid changes in temperature without fracture. Some media, unless precautions are taken, are likely to crack during ordinary grinding and polishing. Others can withstand the changes involved in exposure on the exterior of supersonic airframes. The larger the piece of material, the more subject it is to damage from this cause.

(3) Coefficient of thermal expansion. This characteristic, while related to the ability of a material to withstand thermal shock, is also important if the instrument must withstand a wide range of temperatures. The coefficient material should be matched with that of the cell in which it is to be held or to which it is to be cemented.

(4) Specific gravity. Especially in airborne equipment, weight is an important factor. Hence knowledge of the specific gravity of the material is useful. However, the significance of the specific gravity depends on the effect of the optical characteristics of the design of the system. If, for example, a material of higher density permits using a single lens instead of two, there still may be a weight advantage with the dense material.

(5) Hardness. The hardness of the material is important both during fabrication and in service. A very hard material, such as fuzed quartz, is difficult and time-consuming to grind. On the other hand, a very soft material is likely to develop scratches or sleeks during polishing. Soft materials should be avoided in locations where they will be exposed to surface abrasion, as in exposed domes in airborne equipment.

(6) Surface deterioration. Polished optical surfaces may deteriorate from a number of causes, the susceptibility depending on the material. Staining due to the action of atmospheric moisture and carbon diozide is known as weathering. Closely related to weathering is susceptibility to tarnish or etching by weak acids, which may frequently occur in the atmospheres encountered. In hot, humid climates mold may grow on the surface, leaving marks which cannot be removed except by re-polishing.

(7) Devitrification. Some glassy materials have a tendency to devitrify, and the tiny crystals formed make the glass diffusing. Glasses for instruments which must be stored for long periods should be chosen with this in mind.

17.6 REFRACTIVE MATERIALS FOR SPECIFIC WAVELENGTH RANGES

17.6.1 Classifications. It is convenient to discuss materials under the following classifications:

(a) for the visible region including the near ultraviolet and the near infrared, from about 0.36μ to about 2.20μ.

(b) for the ultraviolet region at wavelengths shorter than 0.36μ .

(c) for the infrared at wavelengths longer than 2.2μ .

17.6.2 Applicable materials for visible spectrum. Most optical design work is done in the visible spectrum and more types of media have been developed for it than for the others. These may be classified as ordinary optical glasses (including rare earth glasses), crystals (natural or synthetic), and plastics. Some of the media used mainly for the other wavelength ranges are transparent in this region and could be used here, but their disadvantages make them unattractive as compared with optical glass.

17.6.2.1 Optical glass. The properties of the ordinary optical glasses are well catalogued, and the designer should obtain the catalogs of several manufacturers for reference. These lists vary widely in the amount of information provided. The most elaborate lists gives indices of refraction for a number of wavelengths distributed through the visible region, the Abbe or ν-value, and several partial dispersions and dispersion ratios, along with information on specific gravity, weathering characteristics, amount of internal imperfections, and thermal characteristics. There is much similarity between the glasses of various manufacturers. However, if a catalog does not give the desired information on a characteristic which is critical for a special application, it is well to inquire of the manufacturer rather than to take data from another manufacturer's catalog.

17.6.2.1.1 Figure 17.1 shows the range of index values and ν-values within which most of the commercially available glasses fall. However, manufacturers' lists vary widely in the variety offered. The available varieties are adequate for most purposes. If the importance of the project warrants the expense, melts of glasses with properties intermediate between those listed can sometimes be arranged for.

17.6.2.1.2 As pointed out elsewhere in this handbook, complete chromatic correction of a simple system puts a condition on the partial dispersion ratios of the glasses involved, as well as on the ν-values. Thus, in a doublet, to bring light of three wavelengths to a single focus, the ν values of the two glasses should differ, but the partial dispersion ratios should be equal. Unfortunately the partial dispersion ratio for most glasses (and other substances as well) is practically a function of the ν-value, so this requirement cannot be met. However, a few manufacturers offer a small number of glasses which depart from the rule sufficiently to be useful.

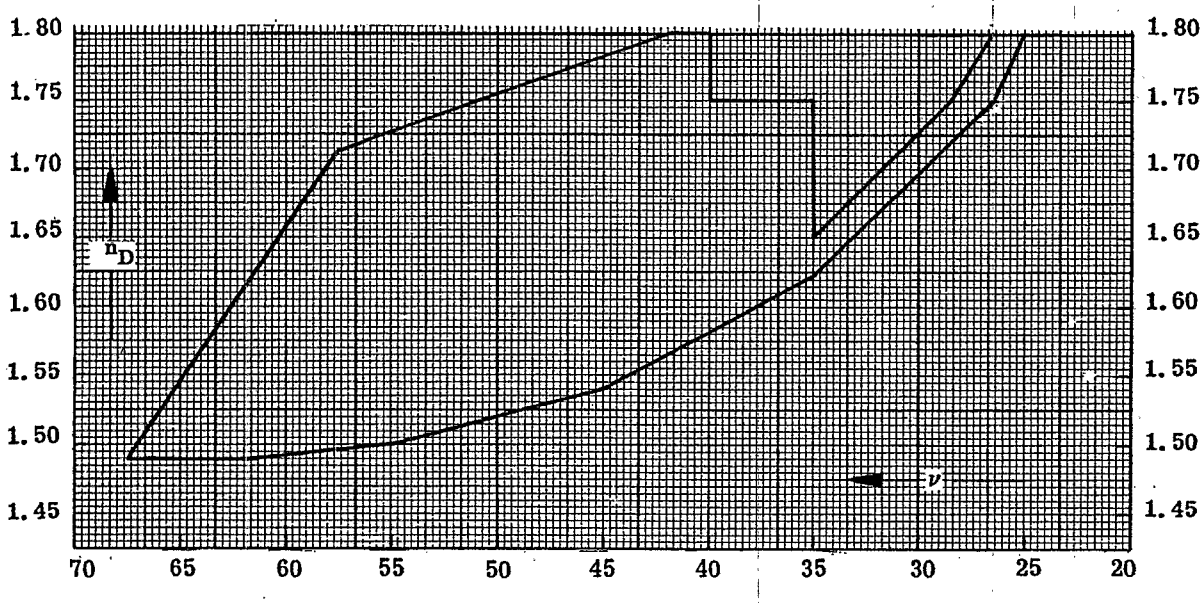


Figure 17.1- The range of commercially available glass, with respect to index and ν values.

17.6.2.1.3 Although one foreign manufacturer's catalog has recently begun listing index values for the wavelengths 0.365μ in the ultraviolet and 1.014μ in the near infrared, most manufacturers give only values in the visible region and to 0.405μ in the ultraviolet. In selecting glass types for the ultraviolet or the infrared without resorting to actual measurement it is necessary to resort to some sort of extrapolation. Kingslake and Conrady* measured the indices of 17 glasses at various wavelengths from 0.365μ to 2.5μ. An approximate value of the index of another glass can usually be derived by interpolation between the values for glasses with similar n_D and ν in Kingslake and Conrady's list. A number of formulae for demonstrating the refractive index as a function of wavelength have been proposed, the constants of the formula being determined from the indices of the glass in the visible region. A fairly recent formula has been proposed by Herzberger**

17.6.2.2 Crystals. Crystals, natural or synthetic, are rarely used for this wavelength region. The principal application is in the reduction, or elimination, of secondary spectrum in apochromatic systems. Fluorite (CaF) has been used for many years for this purpose since it has an index $n_D = 1.434$, and a $\nu = 95.4$. The partial dispersion ratio

$$\tilde{P} = \frac{n_D - n_C}{n_F - n_C} = 0.293$$

which is equal to that of an ordinary glass with $\nu = 57$ to 59. Synthetic crystals, with diameters up to 7.5 inches are available. Synthetic alum crystals ($KAl(SO_4)_2 \cdot 12H_2O$) has also been used. For this material $n = 1.456$, $\nu = 58.2$, $\tilde{P} = 0.315$. Since alum is highly water soluble, the material cannot be used in exposed locations. It has usually been made the central member of a cemented triplet.

17.6.2.3 Plastics. In recent years there has been considerable interest in some of the synthetic resins which can be made transparent and colorless. These materials have a number of attractive features, foremost of which are its lightweight, and the possibility of fabricating elaborate forms quite inexpensively by casting.

17.6.2.3.1 Specific gravities of typical optical plastics are 1.05 to 1.19, compared with 2.48 for crown glass and 3.41 for dense flint. The weight savings are obvious.

17.6.2.3.2 Since the optical elements can be fabricated by casting or molding (processes which are not practical with glass for any, except crude optical elements) it is possible to aspherize mold inserts in order to aspherize lenses, and to cast optical elements complete with mounting flanges, or other mechanically convenient additions.

17.6.2.3.3 Counterbalancing these advantages, available plastics have a number of disadvantages which so far have worked against their use in any but very simple and non-critical systems. They are quite soft, and scratch easily. Many attempts have been made to improve this defect, such as coating the lenses with harder materials, but low scratch resistance remains an important defect. They are quite subject to change of index with humidity, and to significant change of index and surface figure with temperature. They also deform easily under mechanical force, and some tend to turn yellow with age.

17.6.2.3.4 The principal injection-molding materials now used are acrylics, with $n_D = 1.49$, $\nu = 58$, and styrenes with $n_D = 1.59$, and $\nu = 31$. Allyl-diglycol carbonate is used for casting.

17.6.2.3.5 Cast spectacle lenses have some popularity because they are light weight, but the principle defect preventing their wider acceptance is that they scratch easily.

17.6.2.3.6 Large cast blocks are used in unit-power, tank periscopes to shorten the optical path through the periscope, but the end prisms, which ideally should be cast in a single piece with the block, are still made of glass.

17.6.2.3.7 Acrylic lenses have been used in large numbers in simple one and two lens slide viewers, and in inexpensive camera viewfinders. Their quality is adequate for these applications, and their lightness and cheapness are attractive. An additional advantage in the view finders is the aspherization of one lens surface to obtain the elimination of barrel distortion. Injection-molded acrylics are also widely used as singlet meniscus objective lenses in box cameras. One camera manufacturer has introduced a plastic f/8 triplet. It is possible that more applications of this sort will be made as fabrication processes are developed and improved.

---

* R. Kingslake and H. G. Conrady, J. Opt. Soc. Am. 27; 257 (1937).
** M. Herzberger, J. Opt. Am. 32; 70 (1942).

17.6.3 Materials suitable for wavelengths longer than 2.2μ .

17.6.3.1 For application reasons, optical systems in the infrared have been designed mainly for three wavelength regions: the region near 1μ , for use with image converter tubes and infrared photography; the region from 2 to 3μ , for use with lead sulphide cells; and the region near 4.2μ . (Some applications have called for coverage of longer ranges.) Current activity shows an interest in longer wavelengths. As noted above, ordinary optical glasses are satisfactory for the region near 1μ . However, they begin to absorb strongly at about 2.5μ , and their usefulness in the 2 to 3 micron region depends on the requirements of the application, and on the thicknesses needed. For longer wavelengths, other materials must be used.

17.6.3.2 The search for satisfactory infrared transmitting materials has been active and is continuing vigorously. The situation is complicated by the fact that many of the applications, for airborne equipment and especially for windows for such equipment, require excellent mechanical characteristics, and large pieces of material.

17.6.3.3 The properties of approximately fifty materials which are of potential usefulness in the infrared, and which constitute nearly all such materials which had been investigated up to the end of 1958, have been gathered and tabulated by Ballard et al* and the designer should provide himself with a copy of this reference. It lists for each material, to the extent to which information was available at the time of publication, the composition, molecular or atomic weight, specific gravity, crystal class, transmission, reflection loss, refractive index, dispersion, dielectric constant, melting temperature, thermal conductivity, thermal expansion, specific heat, hardness, solubility and elastic moduli. The transmission is usually presented as a curve showing external transmittance as a function of wavelength. Refractive index as a function of wavelength is given in tabular form. The dispersion (which is, of course, implicitly contained in the refractive index table) is for many substances plotted as a curve showing the derivative of index with respect to wavelength as a function of wavelength. One chapter is devoted to tables each listing the substances arranged in order with respect to a single characteristic such as thermal conductivity or coefficient of linear expansion, thus permitting easy comparison. The last chapter is devoted to a brief discussion of glasses and plastics.

17.6.4 Materials suitable for wavelengths shorter than 0.36μ . Work in ultraviolet optics is much less active than in infrared, and the existing applications for the most part impose much less stringent requirements on non-optical characteristics of the materials, and in size of pieces required. A modest number of suitable materials is available, some of them synthetic crystals. Important ones are listed in Table 17.1 together with some of their properties. Index and dispersion in the visible region are given to show the general optical position of the material. Literature references to ultraviolet index and transmission information are included.

Material	N_D	ν	Cutoff	Max. Piece Diameter	Remarks
Sodium Chloride	1.544	42.8	0.25μ	7.5 in.	Highly water soluble
Potassium Bromide	1.560	33.4	0.21μ	7.5 in.	"
Potassium Iodide	1.667	23.2	0.25μ	7.5 in.	"
Lithium Fluoride	1.392	99.3	0.11μ	6.0 in.	
Calcium Fluoride	1.434	95.1	0.125μ	6.0 in.	
Fused Quartz	1.458	67.8	0.22μ	several in.	
Barium Fluoride	1.474	81.8	0.145μ		

Table 17.1 - Materials suitable for ultraviolet beyond 0.36 μ.

17.7 REFLECTING MATERIALS

17.7.1 Thin films. Nearly all reflecting surfaces in optical instrumentation are made by forming thin films, usually by evaporation but sometimes by chemical means, on glass or some other appropriate substrate. Most frequently simple metal films are used. For special purposes, multilayer films are sometimes provided, which give enhanced reflectance in a particular wavelength region, or may serve as

* Stanley S. Ballard, Kathryn A. McCarthy, William L. Wolfe: Optical Materials for Infrared Instrumentation; IRIA Report #2389-11-S, U. of Michigan, Ann Arbor, 1959.

filters, reflecting in one region and transmitting in others.

17.7.2 First surface versus second surface coatings. Depending on the application, the radiation may be incident on the surface either on the side in contact with the glass, or in the side exposed to the air on a vacuum. The latter use is commonly called a "first-surface" reflection; the former, "second surface." When used as a second surface reflector the film can be protected by such means as plating and painting. For use as a first surface reflector, any protecting coating must be transparent. A film of silicon monoxide is frequently employed. Such a film is so thin as to have very little effect on the reflectance in the visible and in the infrared out to 9 or 10μ . However, the thickness required for protection of the surface is sufficiently large to produce interference effects which may decrease the reflectance at various ultraviolet wavelengths.

17.7.3 Simple metal coatings. The situation with respect to simple metal coatings has been summarized by Haas*. Aluminum, silver, gold, copper, and rhodium are considered to be the most important mirror metals. The only material that has a high reflectance in all useful regions, the ultraviolet, visible and infrared, is aluminum. The reflectance of all other metals drops rapidly in the visible or ultraviolet. In the near infrared between 1 and 2μ the average reflectance of silver, gold, copper, and aluminum is higher than 95% but the reflectance of aluminum is about 2% to 3% lower than that of the other three materials. In the far infrared at 10μ all four metals have a reflectance of 98% to 99% and even rhodium reflects about 96%. Today, the most frequently used high reflecting coating for first surface mirrors is vacuum deposited aluminum. It adheres better to glass and other substrates than the other high reflecting materials, it does not tarnish in normal air, and it is very easy to evaporate. Obviously, aluminum coatings are especially important for astronomical mirrors and reflection gratings where high reflectance in the ultraviolet is required. Figure 17.2 of Haas (loc cit) shows the reflectance of freshly deposited films of Ag, Al, Au, Cu and Rh as functions of the wavelength from 0.22 to 10μ . For second-surface reflectors in the visible and near infrared silver is frequently used. As a second-surface reflector it can be adequately protected by copper-plating and painting.

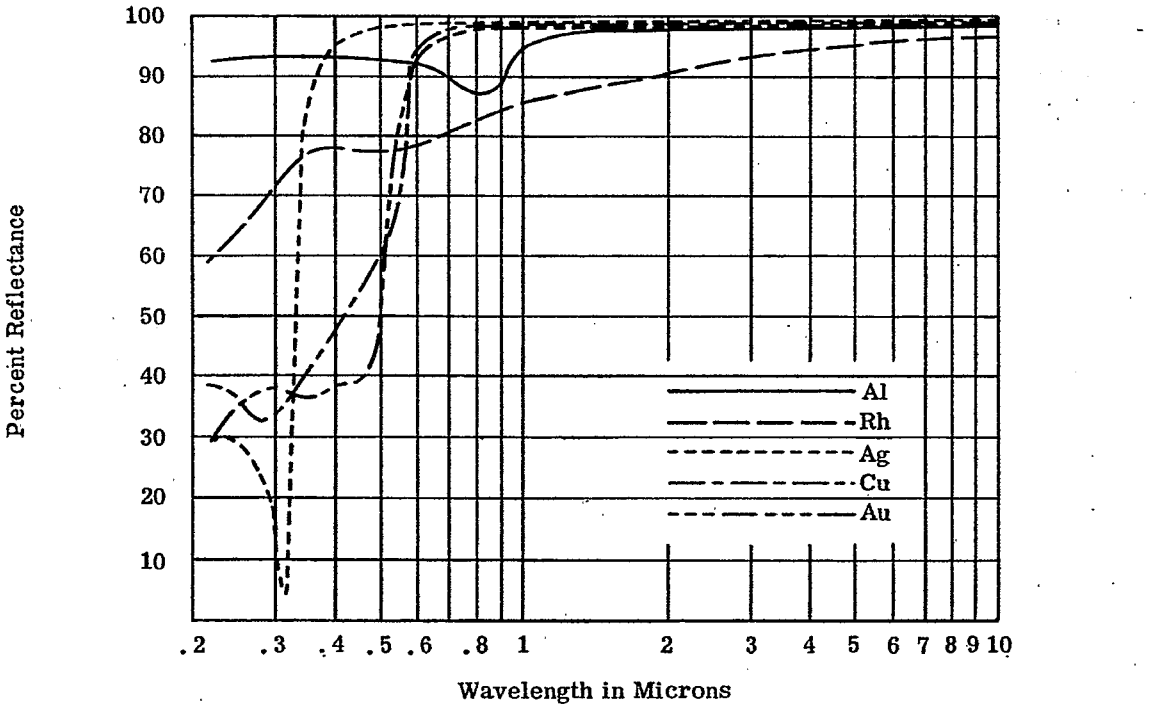


Figure 17.2- Reflectance of freshly deposited films of Ag, Al, Au, Cu, and Rh as function of wavelength from 0.22 to 10μ. (From Jour. Optical Soc. America, G. Haas 945; 945, 1955)

* Haas, George: Filmed Surfaces for Reflecting Optics; JOS 945-945, November, 1955.

17.7.4 Reference. The subject of reflective and anti-reflective coatings is treated extensively in Sections 20 and 21 of this handbook.

17.8 AVAILABILITY; COST; EASE OF WORKING

17.8.1 General. Materials with attractive optical properties are sometimes unsuited to particular applications because of unavailability, either in quantity or in size of pieces required, because of high cost, or because of the difficulty of working the material satisfactorily. Even in cases in which the cost itself is not an objection, the designer of military instruments should as far as possible avoid the use of materials which would be in critical supply, or cause excessive demands on manpower in times of national emergency.

18 ATMOSPHERIC OPTICS

18.1 INTRODUCTION

The use of optical instruments, and one may include the eye in this definition for the present purpose, is limited by the transmission of the atmosphere. As will be developed in the following section, the degree to which the usefulness of an instrument is limited by the atmosphere is roughly related to the aperture. Under specialized conditions however, aperture is of no consequence since image information content is reduced to zero by scattering.

Two generalized types of optical systems may be considered with regard to atmospheric effects. First is the information gathering type of system which depends upon image formation for fulfillment of its purpose. Second are photometric devices which have as their purpose only the detection of amount of radiant energy. Except for highly specialized instrumentation, most optical systems in which atmospheric contributions are significant are of the first class. This discussion will deal only with instruments intended for the former purpose.

18.2 EXTINCTION

18.2.1 Types of transparency losses contributing to extinction.

18.2.1.1 The light of the stars, planets and sun is observed to be weakened as it penetrates the earth's atmosphere. The investigation of this effect, termed the Astronomical Extinction, is one method which has been used in investigating the scattering properties of the atmosphere. Further information has been obtained by measuring the distribution of light in the daylight sky. Both of these types of measurement are of course repeated under laboratory conditions with artificial light sources, and are then useful in the study of extinction produced by media such as rain and fog. The extinction in the clear sky has three contributing components:

(1) Absorption by air molecules.

(2) Rayleigh scattering by air molecules.

(3) Scattering by dust and haze.

The scattered component observed from the clear sky is due to factors 2 and 3 only while the first factor includes the continuous Chappuis bands of Ozone that cover a substantial section of the visual spectrum. Also included in the first factor are the molecular absorption bands of Oxygen, Water Vapor, Carbon Dioxide etc. These later bands are of particular interest to those dealing with Infra-Red optical systems. The so-called atmospheric "windows" are merely areas without such absorption. The design of Infra-Red systems is considered a highly specialized application, and further discussion of this type of atmospheric absorption will therefore not be included here. Atmospheric transmission as a function of wavelength is presented in the International Critical Tables.

18.2.1.2 Extinction can be measured by observing the intensity I of a light source as seen through a volume with scattering particles. If I_o is the intensity of the same light source seen through the same volume without scattering particles, the ratio is:

$$I/I_o = e^{-l\kappa} \qquad (1)$$

where l is the path length through the particulate medium and κ is the extinction coefficient. Even with a very dilute smoke, for example, a considerable intensity of scattered light I_s may easily be detected at right angles to the direction of transmission. A certain amount of true absorption does of course occur, and this represents the actual disappearance of light -- the energy of which is lost as heat. (The kinetic energy is transformed into heat motion of the molecules of the absorbing material.) The term κ may therefore be considered as consisting of two components, a (absorption) and s (scattering). The lateral scattering of a beam of light as it passes through a cloud or aerosol may be easily demonstrated. This phenomenon is closely associated with both diffraction and reflection. The light scattered at right angles to a primary beam is found to be partially plane polarized. The reason that the scattered component normal to the primary beam is not completely polarized is because of multiple scattering where light is scattered several times and therefore yields a non-constant plane of polarization. The integrated effect is therefore one of partial polarization. The polarization produced in the scattered component has been used as the basis of a compass designed by Pfund. The fact that the maximum polarization is observed perpendicular to the incident sunlight allows the approximate direction of the sun to be measured even though the sun itself be invisible. This information coupled with

time data allows direction to be computed in areas where magnetic data are unreliable.

18.2.1.3 The transmissivity of the atmosphere for information-gathering optical systems is an inverse function of the extinction and is limited by particulate scattering such as that caused by haze, clouds and fog, and dusts. The subsequent discussion covers the extinction and, more particularly, scattering by these media.

18.2.2 Particulate scattering.

18.2.2.1 Two of the three factors which contribute to the extinction in the clear sky are concerned only with scattering:

(1) Rayleigh scattering by air molecules.

(2) Scattering by dust and haze.

In the unclear or turbid sky, additional scattering is produced by rain, clouds or fog. If one disregards absorption by air molecules, the extinction may be regarded as a function of the above two listed causes.

18.2.2.2 The first quantitative evaluation of the laws governing the scattering of light by small particles was made by Lord Rayleigh in 1871. His mathematical investigation yielded a general law for the intensity of scattered light. The law derived is generally applicable to particles of any index of refraction different from that of the dispersing medium. The one restriction on the application of the law is that the particle size must be greatly smaller than the wavelength of light.

18.2.2.3 As might be almost intuitively arrived at, the intensity of the scattering is found to be directly proportional to the incident intensity and also directly proportional to the square of the volume of the scattering particle that is typical of the scattering medium so long as the maximum particle dimension is small with respect to a wavelength.

18.2.2.4 A most interesting result of the work of Rayleigh is that the degree of scattering is dependent upon wavelength. Thus, for a given size of particle, long waves are scattered less than short ones, because the particles present obstructions to the light waves which are smaller compared to the longer wavelengths than to the shorter ones. The general expression is given as:

$$I_S = k\lambda^{-4} \qquad (2)$$

where k is a constant of proportionality and I_S the intensity of the scattered component. For example, red light at a wavelength of 0.72 microns has a wavelength 1.8 times as great as that of violet light at 0.40 microns. The law predicts:

$$I_{S_v} = (1.8)^4 \quad I_{S_r} = 10 \quad I_{S_r}$$

assuming that the particles doing the scattering are small compared to either wavelength. As was pointed out previously, if the particulate scattering is due to large particles, the wavelength dependence does not follow the law expressed in equation (2). The relative intensity of the scattered component (I_S) as a function of wavelength is shown in Figure 18.1.

18.2.2.5 It is because of this wavelength dependence that so-called "haze" filters are used. The scattered blue light in the sky may be removed by use of a minus-blue filter so that the sky will appear darker in photographs. Indeed, a dark red filter -- corresponding to the wavelength of least scattering -- will show the clear sky as nearly black. Much here of course depends upon the definition of "clear" sky. When the particulate size is such that it is no longer small with respect to the wavelengths involved, white light scattering will occur. This is the result of ordinary diffuse reflection from the surface of the particles. When transparent large particles are considered, more complex results are obtained. In general however, the final result is that white light is scattered as white light and not to a greater extent at the shorter wavelengths.

18.2.2.6 Haze, together with dusts, forms the atmospheric contaminant referred to as the "aerosol". The aerosol consists of airborne particles of microscopic and sub-microscopic size. This aerosol contributes to scattering and therefore to the extinction.

18.2.2.7 The scattering pattern for atmospheric haze cannot be arrived at analytically with the same accuracy as the extinction law given previously. This is due to the fact that secondary scattering occurs. Therefore, the skylight cannot be simply separated into a factor due to molecular scattering and one due to haze.

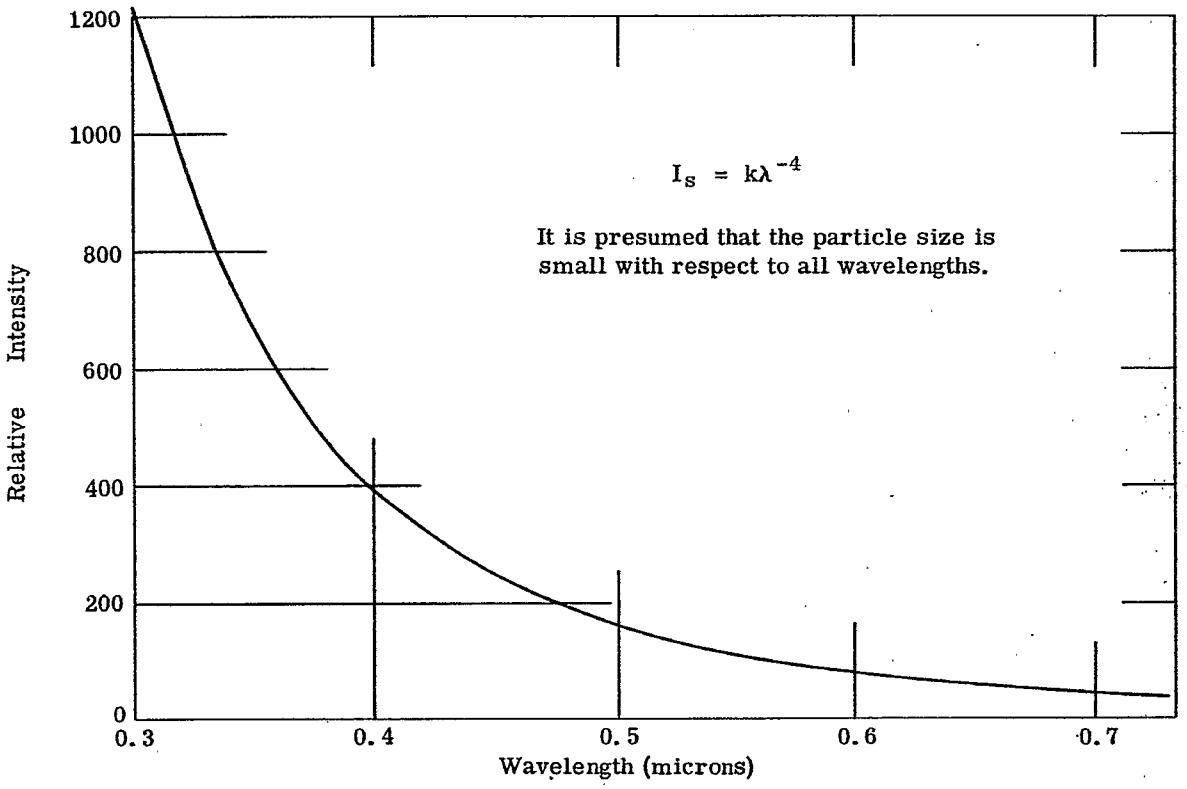


Figure 18.1- Relative intensity vs wavelength.

18.2.2.8 There are two special sources which contribute to the world-wide aerosol. These are volcanic eruptions and forest fires. The eruption of Krakatoa in 1883 was the cause of beautiful sunsets in most parts of the world. A large eruption at Iceland in 1783 caused a particularly red sun to be observed. A more recent eruption of Katmai in 1912 was of special interest since it allowed measurements of particle size to be made. The particles of haze were from 1.0 to 1.2 microns in diameter. Dust particles of the size 0.34 to 0.44 microns diameter were also found and measured.

18.2.2.9 Summer heat haze is a special case. It covers large portions of the earth's surface which are free from dust, fires, or even human habitation. For example, a more or less dense haze hangs over jungle areas in Colombia and the Amazon basin the year around. In the summer months, the heat haze covers most of the continental United States except the desert areas of the Southwest. This haze is of such a density that the open country is almost as hazy as the cities. This heat haze has nothing to do with so-called smog, fog, smoke or even moisture. It is just as dense near the ground, where the relative humidity may be 40 per cent, as near the inversion layer 10,000 feet up, where the humidity may be near 100 per cent. This blue summer haze looks so much like the man-made "Los Angeles" smog that it may almost be considered a "natural" smog. The most usual explanation given for this phenomenon is one based on the production of the "natural" smog by organic emanations from plants. These plants release vast quantities of material into the air. Most of the aromatic substances emitted by plants are hydrocarbons, many of them essential oils. Also, many plants contain terpenes, which after evaporation produce the pungent odors of the deserts and pine forests. Since such summer heat haze is not found over water areas, this explanation is logical.

18.2.2.10 Consider the blue light of a clear sky as seen at sunset. Observing directly overhead, the light scattered will be partially plane-polarized with a maximum perpendicular to the incident sunlight. The reason that completely polarized light is not observed is due to multiple scattering -- scattered light that is scattered a second, third time etc. before reaching the observer. The observation of a red sunset is attributed to the scattering of light by fine dust and smoke particles in the earth's atmosphere near the surface. Since the amount of atmosphere traversed by the sunlight is much greater at sunset due to the low angle, the dust path to the observer is greatly increased with the result that blue and violet have been scattered and the sun appears in the remaining colors. A by-product of the sunset effect is that, due to the low total illumination reaching an observer from the sun and the relatively large amount still present from sky light, the spectral shift in illumination is toward that quality of light yielded by the sky. Under these circumstances, the sky overhead is still blue. The shift is therefore toward the blue end of the spectrum. Due to the so-called "blue myopia" various vision difficulties may be encountered at this time of day even though the total illumination level is quite adequate. The optical instrument designer must consider the use of a spectral balancing filter in

visual instruments to be used at dusk. The obvious problem of total light level vs. detectability requires compromise which must be decided upon the merits of each particular application.

18.2.2.11 The drops of rain, clouds, and fog are much larger than the particles that go to make up dusts and haze. A cloud or fog may contain a proportion of very small droplets, but the diameters of the drops that form the largest portion of the cloud or fog are those that dominate the extinction characteristics of the medium. These particle diameters range from 10 to 40 microns.

18.2.2.12 One extremely important fact to be drawn from this, as may be concluded from the mathematical discussion, is that the extinction contribution of clouds, fog and rain where particle sizes are from 10 to 40 microns or larger is virtually constant throughout the ultraviolet, visual, and near infrared regions. Exactly where one may establish the maximum of the extinction curve will fall is difficult to predict. Measurements throughout the 0.4 to 5.5 micron range have been made by many observers. The extinction remains constant within ± 10 per cent. The efficacy of special filters, for example, under these conditions is not real if penetration is considered. A more complex case exists where cloud cover is not complete, and it is desired only to obtain the best contrast condition against this broken cover. In "typical" rain, the droplet size is of course again much larger than in the case of fogs and clouds. Again, the extinction is essentially constant throughout the spectrum. It is for this reason that special color filters are of little or no value in the penetration of clouds, rain or fogs wherein the particle size is of the order of from 10 to 40 microns. A slight gain is obtained by the use of filters in visual instruments which peak the response to fit the response curve of the eye. Likewise, the response of a given photographic material or photo-cathode substance may be matched in order to give some small penetration advantage. Filters are of a definite advantage whenever the particle size is small with respect to the wavelength.

18.3 EXTINCTION AND VISUAL INSTRUMENTS

18.3.1. Imposed limitations on visual instruments.

18.3.1.1 The main limitation produced by extinction exists in two degrees. First, if the transparency is very low (extinction coefficient high) no art exists whereby penetration in the visual spectrum may be achieved. This is the case in dense fogs, clouds and rain. Since it is impossible to design for this case, the best compromise must be made for the second condition, namely; when transparency is reduced by smaller concentrations of aerosol or water droplet dispersions. We will consider instruments of limited aperture and power. Such instruments will not be seriously hampered by poor "seeing" conditions (such as will be discussed subsequently), but will be limited by extinction.

18.3.1.2 Visual optical systems having apertures of up to two inches and magnifications up to 10 diameters are not greatly affected by differences in air density, inhomogeneity, turbulence etc. when used at elevation angles in excess of 15 degrees. Larger systems are more seriously affected, nearly proportional to their aperture.

18.3.2 Filters.

18.3.2.1 The main thing which can be done in visual optical instrument design to optimize performance under less than ideal atmospheric conditions is the addition of a series of suitable filters. The purpose of the instrument will of course govern the choice of filter combinations made available.

18.3.2.2 The color filter which will prove most advantageous depends upon the object being viewed and upon the background. For ground objects, where the background and subject are both subjected to the same scattering effects, filter choice will depend upon the color of the object. For light or dark objects against the sky, the filter choice will depend upon the brightness of the target and the color of the sky.

(1) When the target is white and bright against the clear blue sky, a filter (red or yellow) which will make the sky appear relatively darker with respect to the object is desired.

(2) Observation of dark objects against a blue sky presents a more complex problem. Some observers maintain that if the blue myopia of the observer is compensated, best results are obtained with a blue filter. Others, and the writer is one of them, fail to observe any improvement in visibility of dark objects against the blue sky with any filter, and subsequently suggest that a "clear" condition is best under these circumstances. The same is true of the compromise required in the observation of mixed white and dark objects against the clear blue sky. For visual purposes the "clear" condition is best or nearly so.

(3) While a white target is usually lighter than a clear blue sky background, it may easily be darker than a white sky background. In this instance it is found that color filters are of little or no help. If the target is colored, the choice of filter to obtain the best result is that which darkens the target more than the background.

18.3.2.3 As we have noted, a great proportion of scattered light is partially polarized perpendicular to the incident radiation causing the scattering. The use of polarizing filters may allow an increase in contrast under special conditions where the scattered component may thusly be at least partially eliminated.

18.3.2.4 At the risk of redundancy, it will be pointed out once again that when large particle media are producing scattering, no filter will appreciably increase penetration. In the case of true haze, where the particle size is small, removal of the scattered light is practical by filtration. A minus-blue filter, or even a red filter may be used under these circumstances to improve clarity.

18.3.3 Light gathering power.

18.3.3.1 One chief result of high background illumination is reduced contrast when bright objects are to be viewed. Also, intervening particles cause scattering which further reduces contrast. The later cause of contrast reduction applies to both bright and dark objects viewed against a comparatively bright background. If the visual consequences of this loss of contrast are to be minimized, it is necessary that further contrast loss not be produced by having an exit pupil smaller than the pupil of the eye. It is found to be advantageous, therefore, to accomplish effective aperture reduction by means of neutral density filters rather than by reduction in aperture by means of adjustable or fixed aperture stops. The use of such neutral density filters has a further advantage, not directly connected with atmospheric optics, in that the resolving power of the instrument is not reduced as it might well be if the aperture were reduced by physical limitation of diameter.

18.4 EXTINCTION AND PHOTOGRAPHIC INSTRUMENTS

18.4.1 Effect of instrument orientation. Photographic instruments, for the present purpose, will be placed in two groups -- those used looking up, such as missile tracking cameras, ballistic cameras etc, and those looking down, such as aerial cameras, satellite borne cameras etc. The atmospheric effects due to extinction coefficient variations are quite different in the two cases. Some generalizations can be made which apply to both groups, but the two are better treated separately.

18.4.2 Photographic instruments "looking up".

18.4.2.1 Missile and satellite tracking cameras and telescopes always have the sky as a background. Depending upon the conditions of a particular firing or satellite pass, the sky may be either clear or cloudy, bright or dark. Due to a gain in performance at greater scale, missile tracking and satellite surveillance cameras are usually of great focal length. Satellite acquisition cameras, on the other hand, are of high optical speed and short focal length.

18.4.2.2 As would be expected, the contrast between the target and the sky background is a function of the distance from the camera to the target. Horizontal and vertical components of this distance however, are not of equal importance in the reduction of contrast. For a dark target on a clear day (blue sky background) subject contrast is decreased more by an increase in altitude than by an increase in horizontal range. For a bright target on a clear day the opposite is true.

18.4.2.3 As is the case for visual instruments, the greatest gain in performance is obtained through the use of suitable filters. The type of filter to be used with an optical system "looking up" depends upon the relative contrast of the object and the background.

(1) Filters for photography of white targets against a blue clear sky require effective darkening of the sky background and heightening of the relative brightness of the target. For a spectrally non-selective white target, the use of a red filter (Wratten 25 or 29) ordinarily gives excellent contrast results. A blue filter will give poor results in this situation since this will effectively reduce contrast. A yellow filter will yield satisfactory results under these conditions, but the red filter will be superior provided that the filter factor or effective density is not so great as to reduce the available illumination below operable limits.

(2) Filters for photography of dark targets against a blue clear sky require effective darkening of the target and heightening of the relative brightness of the background. This can be achieved by the use of a blue filter (Wratten 47). The photographic material does not suffer from the blue myopia of the visual observer. Also, photographic contrast in blue is essentially the same as for other colors -- in fact for certain color materials more tone scales are available from the blue sensitive layer than from any other -- so that the contrast rendition can be excellent under these circumstances if a blue filter is used. If the target is in an orientation 180 degrees opposed to the sun, and if the sun is very low on the horizon, a yellow or red filter may actually be better than the blue, due to the blue deficiency of the sunlight which must pass through a greatly extended atmospheric path. In the case of a satellite that is effectively outside the earth's atmosphere however the choice of filters is more complex, since the scattering path length for light reaching the optical system from the target is nearly the same as the scattering path length of the light from the sun (here the atmospheric path length only is considered).

(3) When photography is done on either white or gray objects against a sky background which is whitened by scattering, filters are found to be of little if any assistance. If the target is colored, the choice of filters should be such as to insure the darkening of the target more than the background. For example, a blue filter would be used with yellow or red targets.

18.4.3 Photographic instruments "looking down".

18.4.3.1 It is well known that photographs taken with large and small lenses of similar quality have nearly the same microscopic contrast in the presence of haze, but have ground resolution modules approximately proportional to focal length. There are many other factors to be considered, such as the quality of the lens, the lens mounting, the type and speed of the shutter, the filter used if any, etc. In general however, if all these factors are equal or nearly so, and if the laboratory performance resolution wise (in lines per millimeter) remains independent of focal length as is usually the case with present day aerial lenses, then the focal length is the most important factor in obtaining a given desired ground resolution.

18.4.3.2 The only extinction or scattering phenomenon which can be partially eliminated by means of filters in the case of photographic instruments "looking down" is haze. Minus-blue filters are commonly employed for this purpose. Since most ground objects have extremely low contrast ratios (of the order of log 0.1 and thereabouts) the elimination of haze effects is extremely important. Red filters have a better haze elimination characteristic in the case of true haze, but these have a tendency to cause natural greenery to appear too black for satisfactory interpretation. The usual compromise has been to select a yellow or orange filter for the purpose of haze minification.

18.5 "SEEING"

18.5.1 Atmospheric factors affecting "seeing".

18.5.1.1 "Seeing," as differentiated from transparency, (reciprocal extinction) is concerned with those factors which limit the information content of images by causing a lack of bulk homogeneity in the optical medium preceding the optical system.

18.5.1.2 At least two basic causes for differences in refractive index of air may be demonstrated. First, there is the "suryp" analogy, in which the density -- and thus the refractive index -- is changed by local differences in temperature throughout the air mass. A second method of producing differences in refractive index is by the condensation and rarification of air by sonic means. When a sound wave moves through air, the air mass instantaneously consists of a series of sections of air having in turn increased and decreased densities -- and thus increased and decreased refractive index. These two causes of inhomogeneity in air are the basic causes of "seeing" difficulties.

18.5.1.3 Any observer who has made observations with a medium size or larger telescope is aware that the performance of his instrument is seriously limited by the astronomical "seeing". Images of stars are much larger than is the case if the image diameter were to be limited by the diffraction of the telescope objective alone. Lunar and planetary detail is badly smeared when the seeing is poor. For example, the average seeing disc of the Hale telescope of 200 inch diameter is about 2.5 seconds of arc while theory indicates that the resolution should be on the order of 0.04 seconds of arc. The optical quality of the telescope is not at fault. The difference is due to the quality of the "seeing." And, it must be remembered that the location of the 200 inch telescope was chosen for the good "seeing" conditions existing on Mt. Palomar.

18.5.1.4 The total amount of light received from a bright star by a telescope of moderate size fluctuates in an irregular fashion, and in a manner which can be shown to be due not to any intrinsic fluctuations in brightness of the star, but to the fact that the starlight must pass through the atmosphere of the earth wherein there are regions of density irregularity. A 12 inch aperture telescope, for example, will exhibit variations in intensity of ± 10 per cent of the average value. The frequency of variation may change from several seconds per cycle up to thousands of cycles per second.

18.5.1.5 Two types of effect are attributable to differences in the air mass preceeding the objective. The first consists of oscillation or image motion. This is due to the movement of relatively large air masses at comparatively low velocities. The second is scintillation. Scintillation is the fluctuation in the light of stars known to be caused by motion across the telescope objective of a complex pattern of light and dark bands known as shadow bands or the shadow pattern.

18.5.2 Oscillation.

18.5.2.1 The change in position of an image in the focal plane of an objective system due to atmospheric disturbance is known as oscillation. This image defect does not of necessity destroy visual resolution. There is a good likelyhood that photographic resolution will be seriously curtailed if substantial amounts of oscillation are present however, since the photographic plate is inherently an integrating device and does not by itself compensate for shifts in position of a high quality image.

18.5.2.2 On a simplified basis, oscillation may be thought of as being caused by the passage of various lens shaped or prism shaped "chunks" or modules of atmosphere in front of the objective. If each air module is of a size greater than the diameter of the telescope objective, a comparatively good image will be seen whenever the entire objective diameter is covered by a single homogeneous module. Since such air modules occur in various layers at different altitudes and move with different velocities, it is obvious that seeing becomes a very complex thing that is difficult of analysis.

18.5.2.3 In a real situation, several cross-currents of air may contain air modules of various sizes. When, in accordance with the laws of chance, the optical path through a diameter covering the objective is equal to within the tolerance of one quarter wavelength of light, an excellent image will be found in the focal plane of the telescope.

18.5.2.4 In the case of relatively slow moving air masses which give rise to oscillation, some sort of compensation is practical. Photometric guided tracking instruments have been developed to compensate for oscillation. These instruments detect the angular movement of the image of an astronomical object, and move the photographic plate or the image so that the effective position of the image on the plate is constant, and a good photographic image is produced.

18.5.2.5 The larger the telescope, the less the probability that the air mass over it will be homogeneous within a quarter wavelength path difference at any given time. Thus follows the hard fact that smaller telescopes may frequently give better resolution than larger ones even though their theoretical resolving power is not nearly so great.

18.5.2.6 Even when the air modules passing over the objective are of sufficient size so that theoretical or nearly theoretical resolution may be obtained, the image motion caused by the shifting air masses is such that long period photography may fail to yield anything approaching what may be seen visually. Very short exposure photography has been tried in efforts to circumvent this difficulty. Also, guidance of the photographic plate and/or the image forming beam have been tried to yield better photographic results. Both of these methods have yielded some success. Excellent planetary photographs have been taken recently with the 60 inch aperture reflector on Mount Wilson after the adaptation of a seeing compensator at the photographic focus. This device moves the final image in accordance with the image shift so that a non-moving image on the plate is the result. Short exposures using telescopes of large aperture have also given promising results in planetary photographs.

18.5.3 Scintillation.

18.5.3.1 Those fluctuations in the light of a star that are not due to any inherent change in brightness of the object but instead to the motion across the telescope objective of a complex pattern of lights and darks, known as the shadow bands or shadow pattern are called scintillation. This shadow pattern is caused by the passage of starlight through atmospheric irregularities which must occur at a considerable height. These irregularities diffract the light and cause rarification and reinforcement of the wavefront at various points along the ground. A fairly complete theory of the relationships between the atmospheric irregularities and the pattern of lights and darks produced in the shadow pattern has been developed.

18.5.3.2 It is common knowledge that the theoretical diffraction pattern can only be observed with telescopes of small aperture and under good conditions of "seeing." With instruments of moderate size -- 36 inches and upwards -- such theoretical resolution is rarely if ever achieved. Using the 36 inch aperture example, the theoretical resolution limit would be less than 0.15 seconds of arc. In practice, however, the starlight is usually spread over a disk of from 2.0 to 5.0 seconds of arc in diameter. This is called the "seeing disk." This "seeing disk" is simply the circle of confusion for the rays reaching the focus (physical optics disregarded for the moment), and its diameter is a measure of the lack of parallelism of the rays when they arrive at the objective from the star. We may thus choose to consider the "seeing disk" as the summation of the diffraction patterns formed by each element of the objective and the air column over it. These elementary diffraction patterns are in rapid oscillatory movement both along and at right angles to the optical axis of the telescope. If the aperture of the large telescope is stopped down to the aperture which would yield the theoretical limit of resolution equal to the actual resolution of the large system, the usual diffraction rings will become clearly visible and sharply defined. The amplitude of brightness scintillation will be noted to have increased roughly inversely to the ratio of new to old diameters. This leads to the conclusion that the "seeing disk" formation is a phenomenon largely independent of brightness scintillation. It seems probable that the "seeing disk" arises from refraction of the rays in their passage through our atmosphere, while brightness scintillation is mainly a diffraction effect arising from the presence of much smaller reinforcement and rarification irregularities within the beam.

18.6 THERMAL EFFECTS

18.6.1 Types.

18.6.1.1 Differences in refractive index due to fluctuations in density which are in turn due to thermal effects play a considerable role in the limitation of vision and photography through optical instruments used for ground level observations. This so-called "ground seeing" frequently limits what can be seen even with low power instruments. In some instances these effects seriously reduce the information which can be gathered by such a small aperture optical system as the human eye.

18.6.1.2 Also, thermal effects in and around larger optical systems are frequently the limiting factor in the performance of these systems. Thus, tube currents and dome currents in astronomical and missile tracking telescopes may reduce the performance of the instrument by a factor of two or more if measures are not taken to circumvent the degradation.

18.6.1.3 The "mirage" or atmospheric striae noticed over the desert floor is an example of what may occur when an air mass is heated by radiation and conduction. In desert areas, even low power telescopic systems of small aperture are very limited in use at low elevation angles. A pair of 7 x 50 binoculars will show much image degradation due to this heat shimmer.

18.6.2 Tube currents.

18.6.2.1 Insofar as the final image is concerned, it matters little whether the density discontinuity occurs without or within the tube of the optical instrument itself. Any telescope tube exposed to thermal radiation or temperature differences of any kind will have variations in density of the air within itself. When these differences become large enough, air flow will occur between the dense and rare areas setting up tube currents. Typically, the air just inside the outside covering of the tube is heated or cooled most rapidly, and a laminar convection current forms wherein air circulates from the perifery to the center. If the tube is an open one, the warmed air -- considering that the tube is being heated by the surroundings -- passes out the end of the tube. Provided the tube is sufficiently larger than the free aperture of the optical system, the air which flows out the end will disturb that remaining in the tube but little. This argument has been presented in favor of open tubes. Unfortunately the issue is not nearly so clear cut. The (assumed) warm air flowing out from the opening of a tube will, since it must obey the gas laws, rise after exiting from the tube. Unfortunately, the path in which it must rise is directly in front of the direction of pointing of the instrument, and while the warm air from the top of the tube causes no difficulty in this regard, that from the bottom flows directly past that area which it is desired to protect from any density differences.

18.6.2.2 A solution which answers the objections to both the closed and open tube arrangements is found in the evacuation of the light path volume. The degree of vacuum need not be high in order to accomplish a substantial increase in performance. A striking experiment may be readily performed to illustrate this fact. If a relatively small telescope of three or four inch aperture is set up for knife-edge test by autocollimation, nearly complete degradation of the image forming qualities of the instrument may be produced by heating the tube of the instrument with a small flame. If the tube is then evacuated to a pressure of 1 PSIA the image quality will be restored to near that present before the heat was applied. It is assumed that the objective is sufficiently thick as to withstand the pressure differential without optical deformation or that a thick optical window has been placed near the objective.

18.6.2.3 The minimization of thermal heating of any tube is desirable from a tube deformation and image quality standpoint. It is good design practice to require some air space beyond the actual optical clearance lines in most optical instruments. A layer of insulation inside the tube also tends to even out the thermal effects so that, while heating will still occur, the unevenness of the heating will not augment thermal disturbances.

18.6.2.4 Most of the energy reaching the earth from the sun is contained within the spectral region between 0.36 and 2.0 microns. The use of highly reflective paints is desirable when thermal effects from radiation are to be reduced. The once prevalent idea that solar heating came mostly from the infrared and that therefore only good infrared reflectivity was required of an instrument paint is quite wrong. The total energy (insolation) from the sun may be as high as 0.028 calories per square centimeter per second. Over 90 per cent of this energy is in the wavelength region below two microns, and nearly 80 per cent is in the region below one micron. So, while good infrared reflectivity is desirable in an instrument white, good visual reflectivity is certainly equally required.

18.6.3 Dome currents.

18.6.3.1 As is the case with telescope tubes, observatory and other housing domes have currents associated with their construction and situation. Particularly in the case of domes opened in the daytime for use, currents may be of such magnitude as to render the housed instrument nearly useless.

18.6.3.2 Painting and insulation are commonly resorted to minimize the effects of thermal heating of the dome. Again, a highly reflective white paint with a reflectivity which matches the solar spectrum as well as possible is desired. Insulation produces a thermal lag which, unless coupled with temperature control as noted later, may actually impair performance rather than improve it.

18.6.3.3 Dome currents may be nearly eliminated if the air inside of the dome is maintained at the same temperature as the air outside. This is an isothermal situation not unlike that in the isothermal jacket of an accurate calorimeter. Since, under conditions of solar radiation, the interior of the dome will usually be considerably warmer than the outside air, it is necessary to provide refrigeration if the desired state of isothermal conditions is to be obtained. At times however -- such as sunrise -- the interior temperature will be much below that of the outside, so that heating capability is also required. It is not sufficient to grossly heat or cool the air within the dome. It is essential that the mixing of air be complete and that inhomogeneities do not exist in the air mass within the dome itself.

18.6.3.4 Next to air temperature control, the best practice is to allow the dome air to arrive at some approximation of equilibrium with the outside air before the housed instrument is to be used. Before observations begin, astronomers open the dome for some period of time to allow the dome, telescope and associated equipment to come to a steady thermal state.

18.7 ATMOSPHERIC CONTAMINANTS

18.7.1 Sources and effects.

18.7.1.1 In addition to the natural products which go up to make the aerosol -- water, dust, smoke etc. -- there are man-made atmospheric contaminants which may produce very undesirable effects on seeing and transparency. For example, industrial smog may limit aerial photography. Certainly the sky light from large cities has greatly reduced the effectiveness of the observatories located near them. Here the man-made effect is two fold. Industrial smog scatters the light which is a by-product of the city so that both transparency and contrast are reduced in that area.

18.7.1.2 Atmospheric contamination by radioactive particles and dusts dispersed by explosions can be viewed as a potential source of seeing degradation if the quantity ever reaches sufficient levels to produce light scattering and even perhaps a very low level of direct radiation. As things now stand, astronomers are counting single photons in the course of measurements on some stars. In these circumstances it is obvious that background light must be kept at an absolute minimum.

18.7.1.3 The increasing amount of carbon dioxide in the atmosphere has undoubtedly increased the "greenhouse" effect present in the atmosphere. While this can have no noticeable effect on visible observations, the increase in infrared absorption may cause detectable variations in the performance of infrared systems. Also, if the increase is sufficient, local heating in areas of high carbon dioxide concentration may cause seeing deterioration due to thermal gradients.

18.8 EFFECT OF ATMOSPHERIC OPTICS ON INSTRUMENT DESIGN

(1) A knowledge of atmospheric optics is important to the optical instrument designer so that he may take advantage of suitable filters, paints and housings in the overall instrument design.

(2) The designer who knows that seeing conditions limit the performance of a telescope more than does the theoretical resolving power will not specify aperture on the basis of theoretical resolution without first considering the actual conditions of use of the instrument.

(3) The design of an optical instrument must go beyond the physical boundaries of the device, it must include at least an approximation of the air column which forms just as much a part of the instrument as does the objective lens. When the limitations placed on performance of high quality optical systems by atmospheric optics are considered in the design phase, a more satisfactory instrument cannot help but result.

19. OPTICS FOR MISSILE TRACKING

19.1 INTRODUCTION

19.1.1 Functions. The main function of optical missile tracking instrumentation is to determine, with precision, the location and trajectory of missiles and satellites. A second important function is to observe the physical appearance and orientation of the space object, and its alterations over short periods of time. In addition, the instrumentation must provide a permanent record of the object's flight for study and later analysis of the trajectory.

19.1.2 Problems.

19.1.2.1 As can be seen from the preceeding paragraph, the problems of tracking and recording the object's flight are closely related to those in astronomy, particularly to those encountered in the observation of planets and planetary detail. However, in missile tracking, additional problems are encountered, since the objects to be observed are not precisely known with respect to location and trajectory. In the proper solution to the problems involved, many contradictory requirements exist, and the correct choice must be made among these requirements for a design necessary to fulfill any particular specification. These requirements are listed as follows:

(1) The field of view must be wide enough so that the missile image is picked up.
(2) The relative aperture must be high enough to see and record the image under the prevailing lighting and atmospheric conditions.
(3) The physical aperture must be large enough so that the Rayleigh limit does not apply to the detail desired.
(4) The focal length must be long enough to have sufficient detail appear in the final image.
(5) The recorded images must properly follow one another fast enough to determine the trajectory, and/or to disclose the physical condition of the missile.

19.1.2.2 A few remarks are in order concerning these points. The requirement for wide field (1) is so contradictory to all the rest that separate instrumentation in the form of tracking telescopes are needed for picking up the object. In almost all cases, two tracking telescopes are used, as shown in Figure 19.1, one for azimuth and one for elevation and each with its own operator. The function of these operators is to keep the

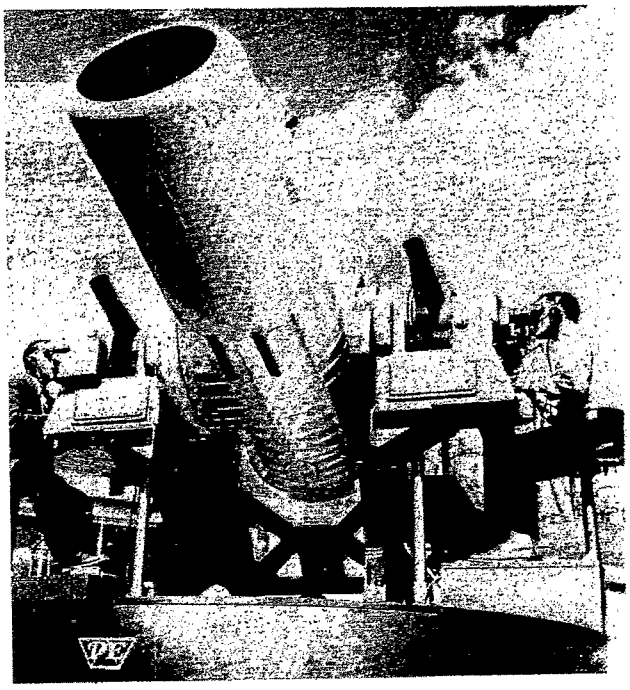


Figure 19.1- Tracking telescopes used with ROT1.
(Courtesy of Perkin Elmer Corp.)

space object close enough to the cross lines so that it will be within the much narrower field angle of the main optical system. This main optical system is completely photographic in nature, so that the final information is recorded on film. Also recorded on the same film, are certain necessary data, particularly an accurate time signal, so that the data from at least two separated stations may be correlated to determine the trajectory of the object by means of triangulation. If errors are to be minimized, a third station may be employed, and suitable data reduction and evaluation means, by high speed computers, used to obtain the trajectory of the object to a high degree of accuracy.

19.1.3 Scope. In this section, the main optical photographic system, which may be either refractive or reflective, will be the basis of the discussion. In addition, emphasis will also be given to those particular optical devices which have been used to obtain desirable properties. The relative merits of refractive and reflective systems will not be gone into here, except to say that for large aperture systems, the arguments are all in favor of the reflective system. In the region of smaller physical apertures, i.e., up to 12 inches diameter, the advantages of greater stability, greater depth of focus, and superior diffraction pattern of the refractive system may overcome the much lower size, weight and freedom from color aberrations of the reflective systems. At larger apertures, however, the balance is so decidedly reversed that the reflective systems have enjoyed an almost complete monopoly in the first half of the twentieth century in the field of astronomy. More advanced designs are now being proposed and constructed in which combined reflective and refractive systems will be of increasing importance, as will be seen from some of the systems described herein.

19.2 REFRACTIVE SYSTEMS

19.2.1 Correction of secondary aberrations. The problems of designing purely refractive systems, of high performance over a relatively small field angle, has been treated extensively in Section 1 of this volume, as has the necessity of correcting for primary spherical aberration, coma, and axial color. Off-axis aberrations, such as astigmatism, distortion, and lateral color, are usually of less importance for a narrow field angle, although they must not be entirely neglected. However, when the focal lengths become large, i.e., in excess of 12 inches, and the relative apertures are of the value of f/8 or lower, great consideration must be given to the correction of the so-called secondary aberrations, the most important of which are secondary spectrum, zonal spherical aberration, and sphero-chromatism.

19.2.2 Secondary spectrum. The methods of handling this troublesome aberration have been discussed in Section 11. For the relatively large apertures we are concerned with in this discussion, the three-glass method leads to fairly high curvatures and impractically thick crown elements. This is also true in combining ordinary crown glass with the crown-flints, where the difference in the Abbe constants is relatively small. It is hoped that glass will eventually be manufactured which will help to alleviate this troublesome aberration. Some years ago, the writer of this section was able to achieve a fairly decent paper design, for secondary color correction, by combining dense barium crown glass with one of the Eastman glasses (EK-320) which gave a higher Abbe spread. However, success was frustrated by the inability to obtain the latter glass, at that time the only available of its type, in better than low-grade C quality. Presently, however, manufacturers are claiming better quality for their lanthanum equivalents, and the situation may change. In addition, other designers are now finding that the later KZF glasses of Schott show considerable promise in this regard and it is imperative therefore, that the designer keep abreast of the newest glass types.

19.2.3 Use of the air space. We turn now to a discussion of the two remaining secondary aberrations namely, zonal spherical aberration and sphero-chromatism, together with a qualitative explanation as to how these aberrations may be minimized or eliminated. With the very long focal lengths involved in the missile tracking problem, either of these can be large enough to ruin the image on the axis. Therefore, the solution to the control of these aberrations is through the judicious use of the air spaces in the interior of the optical design. While these aberrations are of particularly high importance in the design of missile optics, the same considerations apply for other applications throughout the entire field of optics, and a thorough discussion of the corrective properties of the air space is very much in order.

19.2.4 Zonal spherical aberration.

19.2.4.1 In order to understand how the air space may be used to correct for zonal aberration, a short discussion on the nature and origin of the zonal bulge is in order. The spherical aberration of a single surface is given by the expression

$$\text{Sph} = a_1 y^2 + a_2 y^4 + a_3 y^6 + \ldots\ldots\ldots, \qquad (1)$$

where y is the height of the ray above the axis at the surface, and the a's are constant and all of the same sign. The first term is the primary or "Seidel" term and the remaining ones are of so-called higher order.

19.2.4.2 If we now plot the spherical aberration of a single refracting surface, it will look like the solid line in Figure 19.2. The dotted line is the Seidel contribution and since the latter is restricted to the first term in equation (1) it will be a pure parabola. The interval between the solid and dotted lines is due to the presence of higher order terms.

19.2.4.3 Now consider how the zonal bulge originates in so simple an optical design as the corrected doublet. Such a doublet (crown leading) is shown in Figure 19.3. Surfaces 1 and 3 are undercorrecting and relatively weak in curvature as compared to surface 2 which is overcorrecting. It is to be expected that 1 and 3 surfaces will have largely Seidel contributions, while the contribution of the second surface will be relatively rich in higher order terms. If we bend the doublet so that the marginal aberration is zero, we show the sum of the spherical aberration of the undercorrecting surfaces 1 and 3 in Figure 19.4 (a), and that of the overcorrecting surface 2 in (b). The requirement for corrected marginal aberration dictates that $OM_1 = OM_2$. However, the preponderance of Seidel aberration in Figure 19.4 (a) results in a near parabola for the curve, while the presence of higher orders in Figure 19.4 (b) gives a more extreme type of curve as shown, with a milder parabola (dotted extension) plus higher orders. Adding abscissas such as along the line XX to give the final result shown in Figure 19.4 (c) reveals a spherical aberration curve with characteristic zonal bulge, with a maximum of undercorrection at approximately 0.7 of full aperture.

19.2.4.4 Figure 19.5 shows a single positive lens with a marginal ray M and a zonal ray Z. This lens has a large amount of spherical undercorrection and thus the M ray crosses the axis at F_m and the Z ray at F_z. The two rays intersect at the point P. If we set the entrance height of the Z ray at 70 percent of that of the M ray, this ratio is very nearly maintained when the rays leave the lens. However, this percentage ratio gradually increases towards the right and finally reaches 100 percent at the point P, actually reversing beyond this point. If we consider this lens as the first of the doublet shown in Figure 19.3 the negative element of the doublet is located so that the zonal ray Z traverses it at very nearly the original 70 percent height. However, if we allow an air space to exist between the elements as shown in Figure 19.6, the negative overcorrecting element is located at a position where the Z ray is at a height in excess of the original 70 percent value as related to the marginal ray. It thus is acted upon by the negative lens at a point higher than its original assigned value in the aperture, and receives a trifle more overcorrection than it would have received were it a height of 70 percent of the marginal ray. This overcorrection of the zonal ray can be adjusted to reduce or completely eliminate the zonal bulge, or even reverse it, if desired. This control of the zonal bulge obviously depends upon the amount of undercorrected spherical aberration produced by the positive lens and of

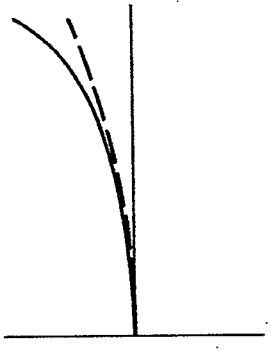


Figure 19.2- A plot of spherical aberration of a single refracting surface.

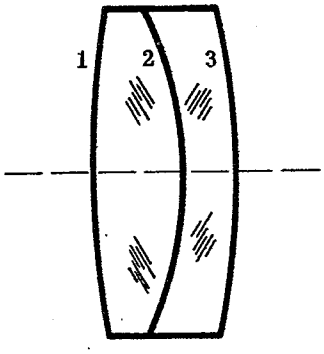


Figure 19.3- Corrected doublet-surfaces 1 and 3 undercorrecting, surface 2 overcorrecting.

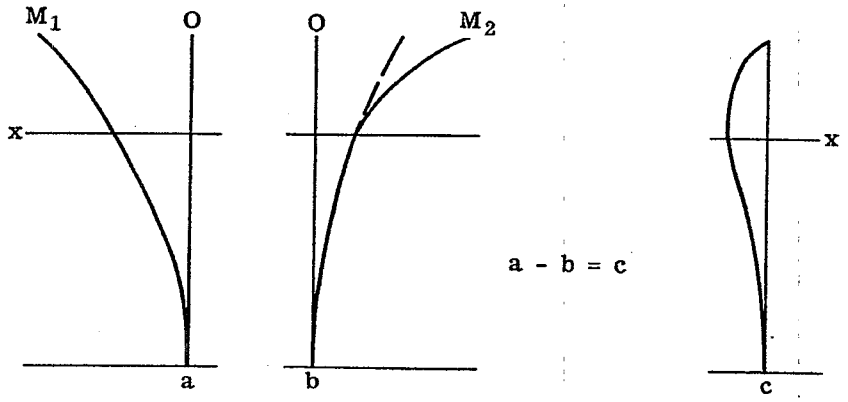


Figure 19.4- Plots of spherical aberration for the lens in Figure 19.3, with marginal aberration equal to zero.

the size of the air space.

19.2.5 Sphero-chromatism.

19.2.5.1 If the cemented doublet in Figure 19.3 is designed to eliminate spherical aberration for a wavelength in the middle of the spectrum, e.g., sodium D, we will discover that the C (red) rays will show considerable undercorrected spherical aberration, while the F (blue) rays will exhibit overcorrected spherical aberration. If the doublet is color-corrected to bring the paraxial C and F rays together, the spherical aberration curves will appear as in Figure 19.7 (a). Also, a better overall color correction will result if the paraxial rays are allowed to be slightly undercorrected, so that the two curves will intersect at approximately the 0.7 zone, or even a little higher as shown in Figure 19.7(b). The cause of this change in spherical aberration with wavelength will be made clear by the following discussion.

19.2.5.2 The refracting power of the doublet is very nearly the same for red (C), yellow (D), and blue (F). However, this refracting power is the sum of those of the positive and negative elements. If the lens is spherically corrected for yellow (D), the positive and negative elements will have certain refracting powers. For red (C) light, the individual refracting powers will be decreased (the sum remaining the same) because of the lower index of refraction. Conversely, for blue (F) light the elements' refracting powers will be increased. The nature of these changes are such, as to have the doublet spherically undercorrected for red (C) light and overcorrected for blue (F) light.

19.2.5.3 The air space can be used to correct this aberration. In Figure 19.8, white light entering a positive single element is refracted by the lens so that the F ray is bent more than the C ray. If the negative overcorrecting element is placed in contact with the positive element, no benefit is achieved. However, if the air space is present as illustrated in Figure 19.9, the F ray strikes the negative element at a lower height than the C ray and thus is subject to less overcorrection by this element. In this way, the naturally spherical undercorrection for red (C) and overcorrection for blue (F) can be neutralized or even reversed by increasing the air space. The amount of the correcting effect depends on the size of the air space and the angular interval between the (C) and the (F) ray upon emerging from the positive element. This angle depends upon the refracting power of the element and the dispersion of the glass.

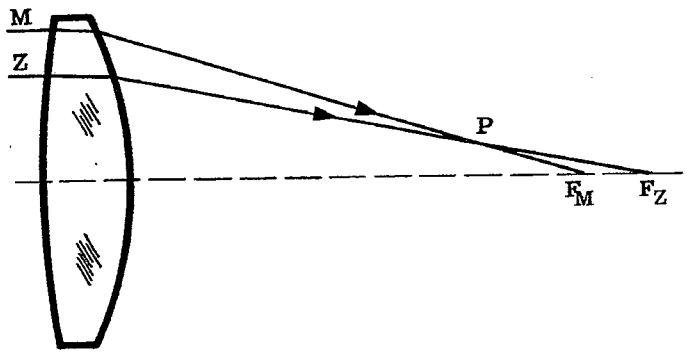


Figure 19.5- Single positive lens used in doublet of Figure 19.3.

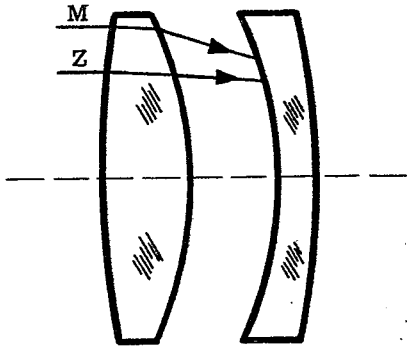


Figure 19.6 - Clark lens, illustrating air spacing to increase over-correction of zonal ray.

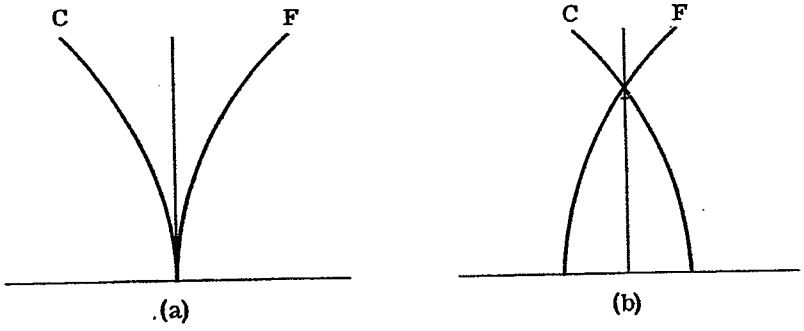


Figure 19.7- The doublet used in Figure 19.2 corrected for soherical aberration and color.

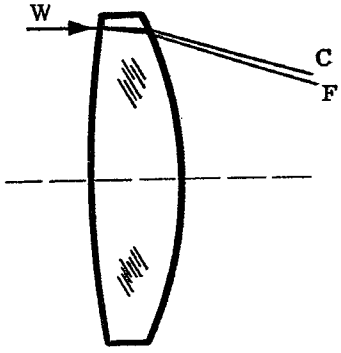


Figure 19.8- Refraction of monochromatic light by the positive element of the doublet shown in Figure 19.3.

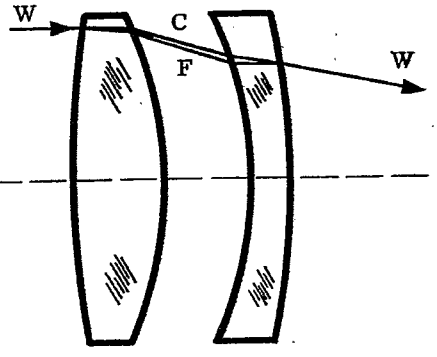


Figure 19.9- Refraction and correction of monochromatic light by the Clark lens shown in Figure 19.6.

19.2.5.4 The air spaced doublet shown in Figures 19.6 and 19.9 is generally known as the Clark lens, produced by Alvan Clark and his successors in the last half of the nineteenth century in the form of large aperture astronomical refractors of exceptional performance. However, there are several objections to the Clark lens. It is extremely difficult to correct both for zonal spherical and for sphero-chromatism with one and the same air space. Further, while in a cemented or contacted doublet the highly curved contact face is of low power, the separation into two elements means that the original contact face has become two strong opposing surfaces of high power. The result is that the axial adjustment and centration become quite critical and the lens become difficult to mount, adjust and maintain. These objections can be overcome by splitting the positive element into two parts and attaching one of these to the negative element as shown in Figure 19.10. Excellent performance with this combination can be achieved up to much higher apertures than is possible with either cemented or contracted doublets. See Section 11. The splitting up of the positive power into two elements means that all of the curves can be made quite mild, and additional degrees of freedom are available to the designer. The designer may choose the relative powers of the two positive components and, also, can investigate the possibility of making these components of different glass types, particularly, with respect to the dispersion factor. *

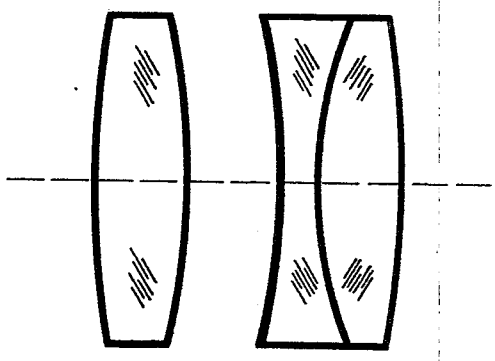


Figure 19.10-Modification of the Clark lens.

19.2.5.5 The lens form shown in Figure 19.11 has incorporated these principles into a simple and effective design used on one of the smaller theodolites. The actual construction has been used for a 40 inch f/6 lens and for a 24 inch f/5.6 lens. The front lens grouping incorporates the general principles discussed in the preceed-

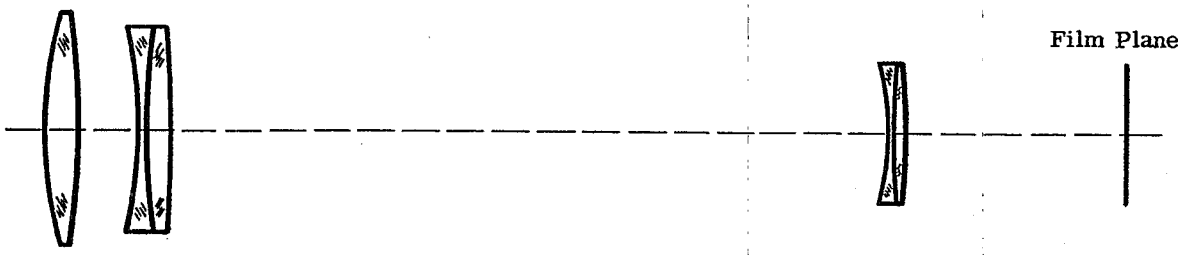


Figure 19.11- Optical layout, 24" high resolution lens.

* JOSA Vol. 42,7 pg. 451, S. Rosin.

ing paragraph. The rear negative doublet neutralizes the Petzval curvature and astigmatism of the front group so that high resolution is maintained over a 70 mm. film format. The spherical aberration curves for all visible colors are straight lines perpendicular to the axis, without any zonal bulge whatsoever (see Figure 19.12). This may be contrasted with the more usual type of curve such as shown in Figure 19.7b.

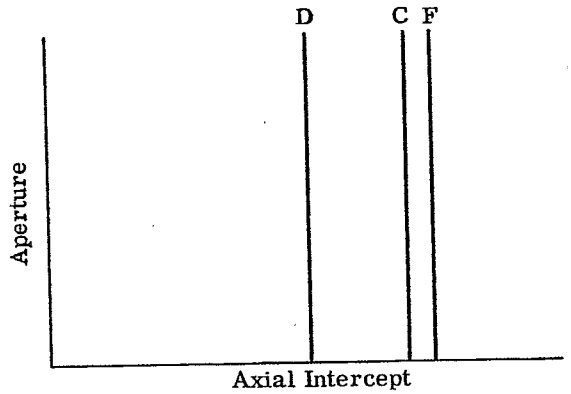


Figure 19.12- Spherical aberration curve for design shown in Figure 19.9.

19.3 REFLECTIVE SYSTEMS

19.3.1 Evolution of the reflective system.

19.3.1.1 The science of astronomy, based on the telescope as it is, discovered early in the twentieth century that the refractive objective had reached the limit of its development. To carry the physical apertures to ever higher values required the use of the reflective system, of which the principal converging element is the reflecting mirror. In this section, a few examples of advanced reflective designs will be given, but in order to properly orient the reader, and to enable him to effect his own designs for particular requirements, a short history of the evolution of reflecting systems and a discussion of some of their fundamental properties will be given.

19.3.1.2 The simplest of all of reflective systems is the concave mirror as shown in Figure 19.13. Rays from a distant object are converged by the mirror to a focus (F). If a film were placed at (F), an image at this point, which is located on the axis half-way between the instantaneous center of curvature (C) at the axis of the

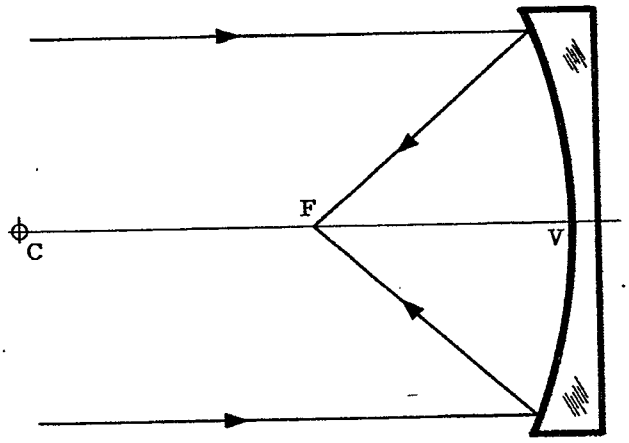


Figure 19.13- A concave mirror.

mirror and the vertex (V) of the mirror itself would be obtained. The simplest of all such mirrors is a portion of a sphere. However, a spherical mirror gives a very poor image at (F), since it is afflicted with a large amount of spherical aberration. This may be verified by ray trace, or by reference to the Seidel expression for spherical aberration. In order to sharpen the image at (F) in Figure 19.13, an aspheric surface for the reflector must be used, and this surface will be a paraboloid with its focus at (F).

19.3.1.3 The great astronomical instruments of the first half of the twentieth century fall into two main classes; the simple paraboloids, the greatest of which is the giant 200 inch diameter mirror at Mt. Palomar, and the Schmidt telescope.

19.3.2 The Newton system.

19.3.2.1 There is one serious mechanical disadvantage to the basic arrangement of Figure 19.13. If an eyepiece for visual observation, or a photographic plate is positioned at F, the center of the beam would be seriously interrupted. To overcome this difficulty, Newton, as early as the seventeenth century, proposed a plane mirror (M), to be positioned as shown in Figure 19.14, to bring the focus outside the beam where observation could be made. If a photographic motion picture film is placed at (F) in Figure 19.14, the recording apparatus can be made as large as necessary. A number of successful missile tracking devices have been constructed in accordance with this arrangement. It will be noted that the beam is partially obscured by the mirror (M). All reflective systems, except for the unimportant off-axis parabola, are characterized by this hole in the pupil, whether it is caused by a photographic plate or a mirror.

19.3.2.2 While it is true that a paraboloidal mirror forms a perfect image on the axis of the system, there remain important limitations. As has been discussed, physical optics and the finite wavelength of light impose a limitation on the resolving power of all optical systems. A good rule for this limitation is the simple equation

$$R = \frac{4.5}{a}, \qquad (2)$$

where R is the resolving power in seconds of arc, and a is the mirror aperture in inches.

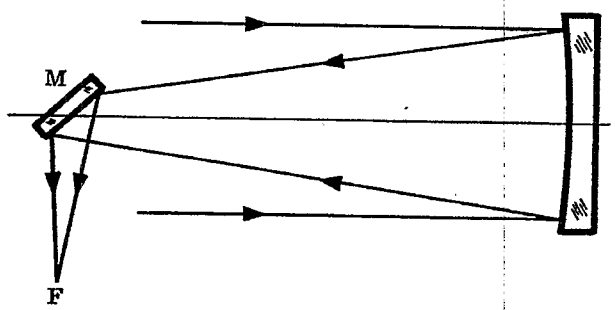


Figure 19.14- Newtonian mirror arrangement.

19.3.2.3 Another serious limitation on the properties of the paraboloid, is the presence of field aberrations, particularly coma. Indeed, the size of the useful field of view of a paraboloid, where the resolution over the field is in accordance with equation (2), can be given by

$$v = \left(\frac{f}{h}\right)^2 \Big/ 25 \qquad (3)$$

where v is the size of the field in inches, f is the focal length of the mirror, and h is the diameter of the mirror. The ratio $\frac{f}{h}$ is the f number of the mirror. Thus an f/5 mirror has a useful field of view of one inch, and an f/10 mirror has a useful field of four inches. The Mt. Palomar 200 inch diameter mirror works at f/3.3, so that its useful field of view is only 0.4 inch. However, over this 0.4 inch field, the Mt. Palomar paraboloid can theoretically resolve better than 0.025 seconds, which for a focal length of 660 inches, would amount to 0.00008 inch.

19.3.3 The Cassegrain system.

19.3.3.1 Another class of reflecting system is the two mirror Cassegrain arrangement.* This system is extremely popular in missile tracking instruments and is shown in Figure 19.15. Rays from a distant object strike a concave mirror (M_1) and are reflected towards a focus at (F_1). Before the rays are converged, a second mirror (M_2), which is convex, interrupts the beam and reflects it to a second focus at (F_2). The position of (F_2) outside the system puts it in an extremely convenient and favorable position for image recording. The hole in the mirror (M_1), is in the region which is blocked out of the original bundle by the physical presence of the convex mirror (M_2). The convex mirror is supported by a mechanical spider, or is cemented to a flat glass plate. In all cases, the mirror (M_2) magnifies the image considerably, since the distance from (F_1) to (M_2) is considerably less than from (M_2) to (F_2). A favorite value for this ratio is 4x, although this figure may vary considerably. This factor lengthens the focal length over that of the mirror (M_1) by the same amount, and similarly increases the focal ratio or f/#. Thus, the Cassegrain arrangement is very suitable for systems of long focal length and relatively low illumination.

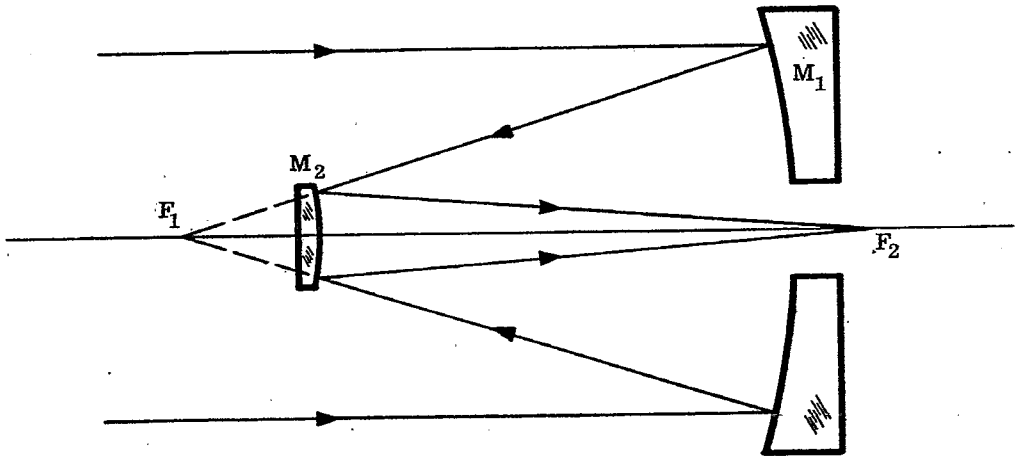


Figure 19.15- Cassegrainian Mirror arrangement.

19.3.3.2 However, there is always the requirement that the image at (F_2) be sharp, that is to say, free from spherical aberration. There are an infinite number of combinations of the two surfaces which will achieve this. For example, the mirror (M_1) can be spherical, in which case (M_2) is a complex, higher order curve. Conversely, the mirror (M_2) can be spherical, in which case the mirror (M_1) will be of a complex nature. Or, in its favorite form, (M_1) is a paraboloid with focus at (F_1), and (M_2) is a hyperboloid with foci at (F_1) and (F_2). Which of these combinations is the best? In most cases, the arrangement which gives freedom from coma will be the most desirable.

* While the system originally proposed by Cassegrain consisted of a paraboloidal primary with an hyperboloidal secondary, accepted usage today has broadened the term "Cassegrain" to apply to any system consisting of a concave primary and a convex secondary.

19.3.3.3 A power equation may be set up for (M_1) with

$$-x = a_1 y^2 + b_1 y^4 + c_1 y^6 + \ldots\ldots\ldots, \quad (4)$$

and the equation cut off after the first two terms. For all surfaces

$$a_1 = 1/2\ r_1\ , \quad (5)$$

where r_1 is the instantaneous radius at the vertex of (M_1). Then let b_1 assume a continuous set of values in the equation

$$-x = y^2/2r_1 + b_1 y^4 \quad (6)$$

and for $b_1 = 0$, the surface is a paraboloid. For $b_1 = 1/8\ r_1^{\,3}$, the surface is spherical, provided the semi-diameter of the mirror is not too large a fraction of the radius r_1.

19.3.3.4 A similar equation can be set up for the mirror (M_2) with

$$-x = y^2/2r_2 + b_2 y^4 \quad (7)$$

where r_2 is determined by the positions of the paraxial arrangement (F_1), (F_2), and (M_2), and b_2 is determined by requiring the marginal ray to pass through (F_2). The form of the mirror (M_2) for any given value of b_1 can then be found. The calculation is difficult and can best be effected on an electronic calculator if available. Then one paraxial and one marginal ray can be traced for each combination (the paraxial ray can be the same for all the combinations) and the departure from the sine condition can be determined. In this procedure, it will be found that the combination most nearly free from coma is not too far from the paraboloid-hyperboloid arrangement, in cases where the magnification of (M_2) is not too different from 4x. If complete freedom from coma is desired by the designer, some departure from this combination may be indicated. However, in all designs known to the author, the paraboloid - hyperboloid form is used, except for extremely low aperture systems, where the spherical aberration is unimportant, and two spheres may be employed.

19.3.3.5 Up to this point, the simple mirror arrangements have been discussed chiefly in the form of the paraboloid and the Cassagrain two mirror system. No color aberrations are involved in purely reflective systems. After spherical aberration is corrected, the paraboloid affords no more degrees of freedom to correct coma. In the Cassagrainian arrangement, the proper choice of form allows both spherical aberration and coma to be controlled. However, no mention has been made of the remaining aberrations, namely astigmatism, field curvature and distortion. These aberrations are handled in precisely the same way as in refractive systems. There is one important difference, relating to the Petzval field curvature.

19.3.3.6 It will be recalled that in refractive systems with a large excess of positive power, the Petzval curvature is concave towards the incident light. The reverse is true in reflective optics. A converging element (concave mirror) has associated with it a heavy Petzval curvature convex towards the incident light. This affords the possibility of combining converging reflective and refractive systems to achieve a flat field, as shall be discussed later. If r is the radius of the mirror, and c ($= 1/r$) its curvature, the contribution to Petzval curvature of any reflecting surface is given by

$$P = 2Nc \quad (8)$$

where N is the index of refraction of the medium in contact with the mirror. For a single concave mirror, the Petzval surface is concentric with the mirror surface as shown in Figure 19.16.

19.4 CATADIOPTRIC SYSTEMS

19.4.1 Introduction. Now consider the second class of reflective systems, which include the Schmidt arrangement and some of its variations and developments. There were two fundamental principles discussed in previous sections of this handbook relating to the change in aberration with a shift in the stop position. In brief, these principles are,

(1) The change in the Seidel coma coefficient, due to a change in the position of the stop, is proportional to the spherical coefficient, multiplied by the shift in the stop position.

(2) The change in the Seidel astigmatic coefficient is proportional to the coma coefficient, multiplied by the shift in the stop position.

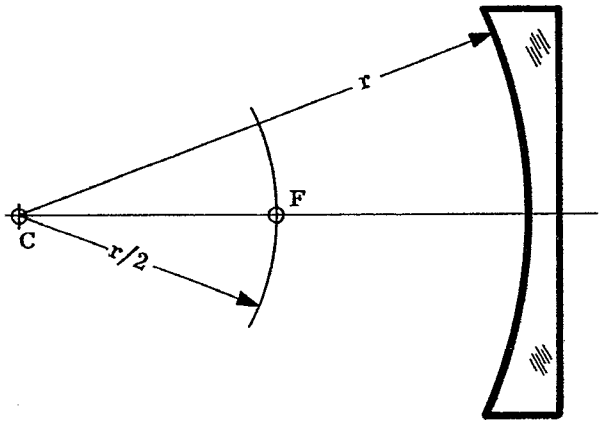


Figure 19.16- Curved focal surface, concave mirror.

A very important principle, relating to aspheric surfaces may then be added,

(3) Any change in form, without a change in the vertex radius, has no effect on the off-axis Seidel coefficients for an optical surface located at the stop position.

The third principle has far reaching consequences in the design of optical systems, since it allows the optical designer to correct spherical aberration by aspherizing the surface at the stop position. At this point, it will only have a mild effect on axial color (no effect at all in the case of reflecting systems), and no effect on coma, astigmatism, field curvature, distortion, and lateral color. It is not to be implied that the optical designer need limit the aspherizing of elements to the stop position, since the effects of such a procedure are subject to calculation, but the use of such surfaces in positions other than the stop brings them into the general juggling procedure characteristic of optical design.

19.4.2 The Schmidt system.

19.4.2.1 With respect to the contributions of Schmidt to optical science, it will be recalled that the single paraboloidal reflector had no spherical aberration but an extremely large amount of coma. In accordance with principle 1, the stop can be anywhere, without changing this coma, since the spherical aberration of the paraboloid is zero. If the stop is considered to be at the mirror, which would be the case if there are no artificial diaphragms in front of it, and the form of the mirror is allowed to change to something else, principle 3 states that the off-axis aberrations, including coma, will be unaffected. Indeed all concave mirrors of equal vertex radius, be they paraboloid, sphere, hyperboloid or off-beat curve will have equal amounts of coma, astigmatism and field curvature. The Seidel coefficient for astigmatism is equal to the refracting power as it is in the case of all thin systems at the stop. For the simple mirror, this is equal to 2c.

19.4.2.2 Now consider the apparently strange consequence of principle 3, and also consider principles 1 and 2. The optical designer might say that perhaps the large amounts of off-axis aberrations can be reduced by shifting the stop position. The designer will be frustrated in the case of the paraboloid, since there is no change in coma due to the absence of spherical aberration. He will probably choose the sphere, since as an accomplished technician he knows there will be less difficulty in making the sphere, than is the case with the hyperboloid or off-beat curve. If the designer allows the stop to recede from the sphere, he will notice a decrease in both coma and astigmatism, until the center of curvature of the mirror is reached. At the center of

curvature, these aberrations will vanish and the configuration will be as shown in Figure 19.17.

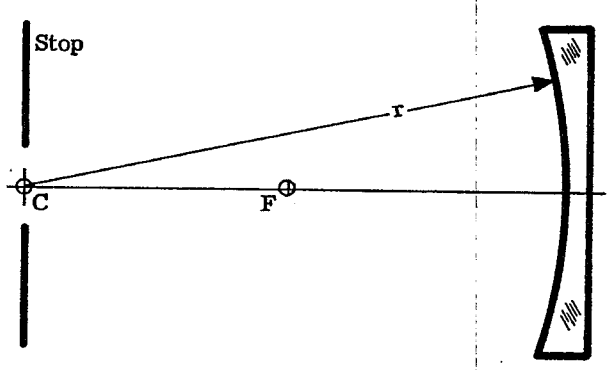


Figure 19.17- Schmidt arrangement, correction plate not shown.

19.4.2.3 At this stage, the design is free of coma and astigmatism, but still contains a large amount of field curvature, which must be tolerated, and a considerable amount of spherical aberration. Principle 3 will allow for the correction of the latter, by placing a plane parallel refracting plate at the stop at c in Figure 19.17, and deforming one surface to correct the spherical aberration. Since the spherical mirror undercorrects the marginal ray, the edge of the plate must have a slight negative power to neutralize it, as compared to the center. In practice, it is common to impart a tiny central positive power to the plate, resulting in a parallel zonal section as shown in Figure 19.18, and a reduced negative marginal region, thus shortening the overall focal length by a very small amount. With the parallel zonal section approximately 0.8 of the distance from the center to the margin, the plate imparts the minimum axial color to the beam.

19.4.2.4 It is fairly certain that Schmidt did not follow this rather involved reasoning to conceive his system. It is quite obvious that for a stop placed at the center of curvature of a spherical mirror, the coma and astigmatism must be zero, since the chief rays strike the mirror normally and can define a new axis just as valid as the central axis. There can be no coma or astigmatic difference on the axis of an optical system. However, the more involved reasoning first given is capable of further extension and application as shown below in the case of the oblate spheroid.

19.4.2.5 The Schmidt telescope has the very great advantage over the paraboloid of enormously extending the field of view over which the image remains sharp. The largest made to date is located on Mt. Palomar, and has a 48 inch diameter aperture with an ability to take sharp pictures over an area 14 inches square. The plates must assume the shape of the field curvature. The focal ratio is F/2.5 with a focal length of 120.9 inches. However, the Schmidt suffers from other defects. Since it has its corrector at the center of the curvature of the mirror, the system must be twice as long as its focal length. Also, to prevent vignetting, the primary mirror must be considerably larger than the aperture. In the Mt. Palomar instrument, the spherical mirror is 72 inches in diameter.

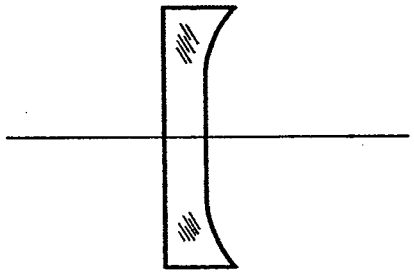
19.4.2.6 If only moderate extension of the field of view is desired, the Schmidt type arrangement, with an oblate spherical primary, can achieve this in a much shorter structure than the Schmidt with a spherical primary. In the following equations,

$$-x = \frac{1}{2r} y^2 , \qquad (9)$$

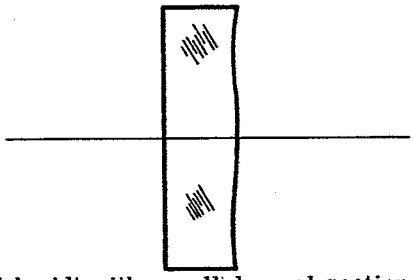
$$-x = \frac{1}{2r} y^2 + \frac{1}{8r^3} y^4 \quad , \tag{10}$$

$$-x = \frac{1}{2r} y^2 + \frac{2}{8r^3} y^4 \quad . \tag{11}$$

Equation (9) is the equation of the paraboloid, equation (10) represents a sphere (expansion of its equation to two terms), and equation (11) represents a surface twice as heavily curved away from the paraboloid as the sphere. Any surface in which the coefficient of the y^4 term is in excess of $1/8r^3$ is called an oblate spheroid and equation (11) defines one such surface.



Schmidt with parallel central section.



Schmidt with parallel zonal section.

Figure 19.18- Schmidt plate with parallel central section, (a); and parallel zonal section, (b).

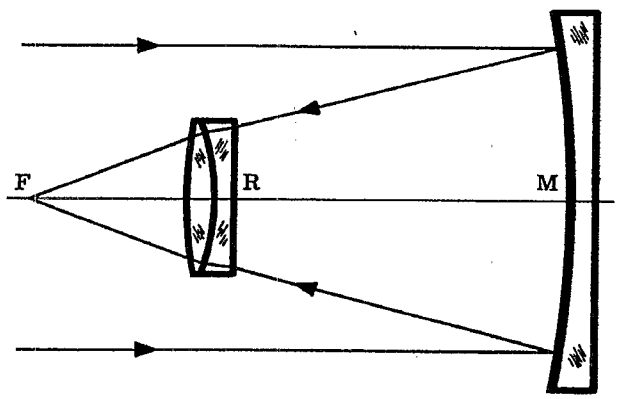
19.4.2.7 The Seidel theory of aberrations as applied to aspheric surfaces states that if equation (9) defines, as it does, a surface with no spherical aberration, and equation (10) defines a surface with a certain amount, then equation (11) is a surface with twice the spherical aberration of that defined in equation (10). Referring back to principle 1, it will be recalled that spherical aberration is needed in order to correct coma by shifting the stop, and that it is possible to correct the coma of the sphere by shifting the stop from the mirror to the center. If now, the designer starts with a surface in accordance with equation (11) having twice the spherical aberration of the sphere, we need to shift the stop only back to the focus to correct the coma, shortening the instrument to half of the Schmidt. The correction plate will now have to be made to correct the doubled spherical aberration.

19.4.2.8 The situation with respect to astigmatism is not so fortunate. Since the stop-at-mirror coma is identical for equation (10) and (11), principle 2 states that shifting the stop back to the focus will correct only half of the astigmatism. For maximum field the stop may be shifted a little further, thus reducing the astigmatism, but allowing the coma of opposite sign to creep back in, until a desirable compromise is reached. The Schmidt principle is capable of a number of variations which will not be discussed further.

19.4.3 The Ross-Baker system.

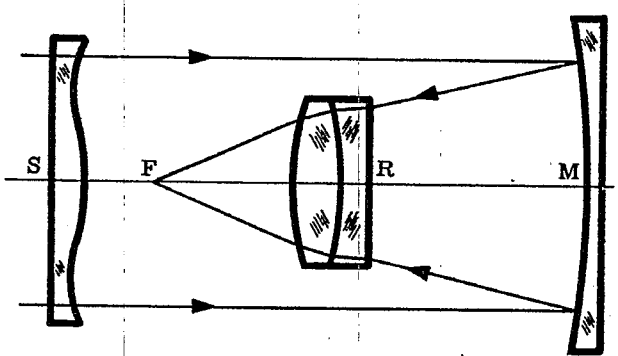
19.4.3.1 Efforts to extend the field of view of the paraboloid have met with some degree of success. Ross constructed some lenses spaced close to the focal plane in accordance with the arrangement in Figure 19.19. A color corrected doublet, placed at R as shown in Figure 19.19, is given sufficient power so that its positive Petzval field curvature contribution will just neutralize that of the mirror M. Its bending and spacing from F offer two degrees of freedom, which are used for the correction of coma and astigmatism. Unfortunately, it proved impossible to prevent R from reintroducing undercorrected spherical aberration into the system so that the images tended to be soft. However the field of view of the paraboloid was greatly extended by this maneuver.

19.4.3.2 James G. Baker has proposed the addition of a Schmidt corrector plate to the Ross system as shown in Figure 19.20. The combination of M and R in Figure 19.20 is designed for correction of coma and astigmatism for a stop position at S. Then insertion of the Schmidt plate takes out the residual spherical aberration. An analysis of this system shows that before the insertion of the Schmidt plate, the residual spherical aberration amounted to a substantial fraction of that of a spherical mirror. However, the amount of depth needed to be hollowed out of the Schmidt was only a few wavelengths of light. This illustrates the tremendous leverage exerted on the rays by systems of this type, and the demanding exactitude required for their construction. Baker proposed this system, which gives definition over a field of view an order of magnitude larger than that of the paraboloid, as a means of correcting simple paraboloids already in existence by the addition of R and S. A similar requirement makes this system desirable for missile recording. The complete system can be used for photographic tracking, while the removal of the refracting elements S and R allows conversion of the system for the detection of targets in the medium infra-red and in the ultraviolet, where the glass would be opaque.



NOT TO SCALE
The size ratio of element R to element M is approximately 5 to 1

Figure 19.19- Ross arrangement.



NOT TO SCALE
The size ratio of element R to element M is approximately 5 to 1

Figure 19.20- Baker arrangement.

19.4.4 Modification of the Ross-Baker arrangement.

19.4.4.1 The author of this section has proposed, in an exchange of letters with Dr. Baker, the elimination of the correcting plate S and transferring its function to the mirror M, thus deforming the paraboloid. The combination of deformed paraboloid, and Ross lens R would have the disadvantage of making the mirror unusable by itself. However, it would have the advantage of eliminating the only large refracting element and enabling the system to be carried to higher physical dimensions.

19.5 APPLIED SYSTEMS

19.5.1 Satellite tracking camera.

19.5.1.1 The principles on which the design of the more complex optical systems used in missile and satellite tracking are based, are illustrated in the Figure 19.21. The design is a classical Schmidt system with just a few variations. For the purpose intended, this camera was designed for high light gathering power and large field, particularly in one direction, preferably the direction of satellite path. The physical aperture of the system is 20 inches, and with a focal length of approximately the same value, the system operates at f/1. To prevent vignetting the primary spherical mirror is 31 inches in diameter.

19.5.1.2 It will be noted that the aperture of the system is very close to the center of curvature of the primary mirror, but the single correcting plate, which normally is located there, is split up into a color corrected triplet for the purpose of eliminating the small amount of residual axial color in the single Schmidt plate. The four inner surfaces of this system are aspheric.

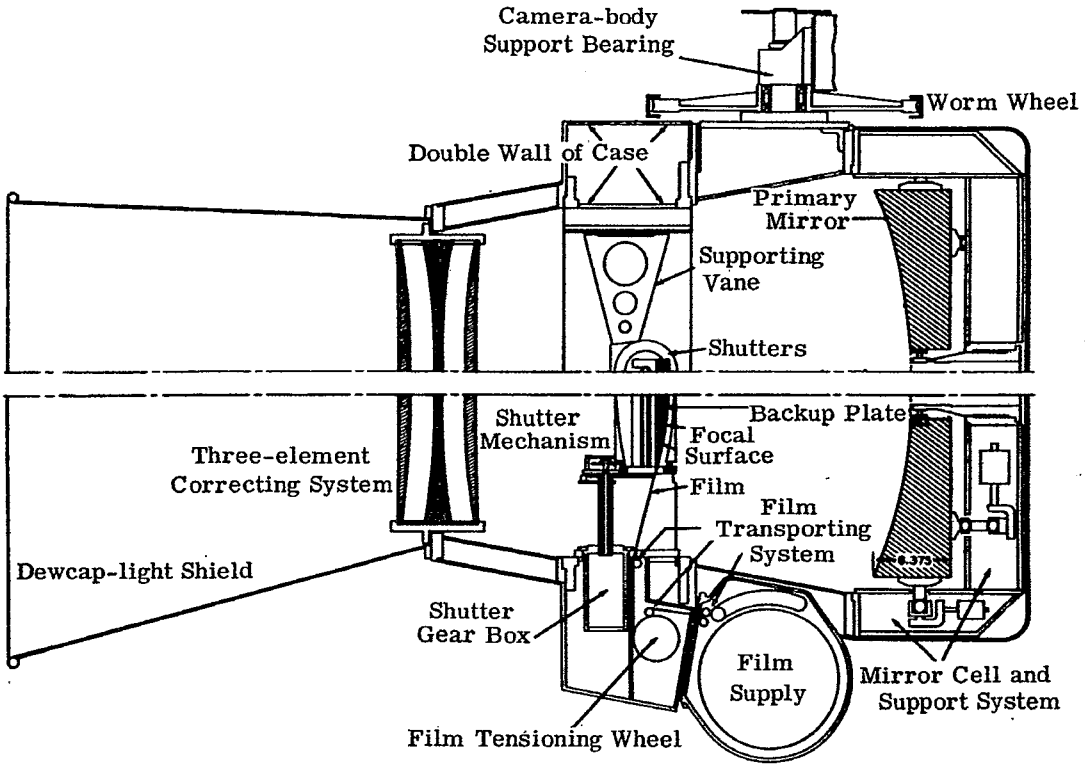


Figure 19.21-Section view of a satellite-tracking camera. (Courtesy of Perkin Elmer Corp.)

19.5.1.3 It is presumed that, because of the high relative aperture of this system (f/1) the curvature of the Schmidt plate required would be more extreme than usual, leading to more axial color than the designer could tolerate. The splitting up of the single plate into three, with the central glass different from the outside, and the distribution of the Schmidt curvature among four surfaces would tend to alleviate this situation.

19.5.1.4 It will be noted from Figure 19.21, that the film is transported over a spherically curved gate, which matches the curved focal plane of the image. The curvature in the plane at right angles must necessarily be zero, because of the mechanical impossibility of bending the moving film into a compound curve. Consequently the field coverage in this direction is limited to only 5 degrees, while in the direction of film travel it reaches the amazing value of 31 degrees. It was found, that at the edges of this extreme field the focal surface departs slightly from a spherical shape so that the film runners are not quite circular. The combination of careful design and excellent execution resulted in a system wherein 80 percent of the point energy anywhere in the field is within a circle 0.001 inch in diameter. This instrument was conceived for the purpose of tracking the U. S. Vanguard satellite, and the first instrument arrived just in time to be used for the original Sputniks.

19.5.2 ROTI Mark II (Recording optical tracking instrument).

19.5.2.1 This instrument and the Igor were made to similar specifications but each has features worth discussing. The original requirements envisioned a versatile instrument capable of a series of fixed focal lengths ranging up to 500 inches in value. Another requirement was that the instrument could be adapted for infrared which necessitates the choice of a paraboloid for the primary mirror.

19.5.2.2 It would be thought that some form of the reflector-corrector system of Baker (Figure 19.20) would be indicated and such is indeed the case. However, the corrector feature is introduced in a rather unique fashion.

19.5.2.3 Referring to Figure 19.22, light enters from the left, passes through the window and strikes the primary mirror, a paraboloid of 100 inch focal length and 24 inch aperture. After reflection by two Newtonian mirrors as shown, the rays are brought to a focus at the reticle. Just before this point a pair of sliding wedges introduce a variable amount of glass into the path. By adjustment of these wedges, the instrument focus can vary between 3000 yards and infinity, the focus being automatically controlled by range data determined from the associated radar equipment.

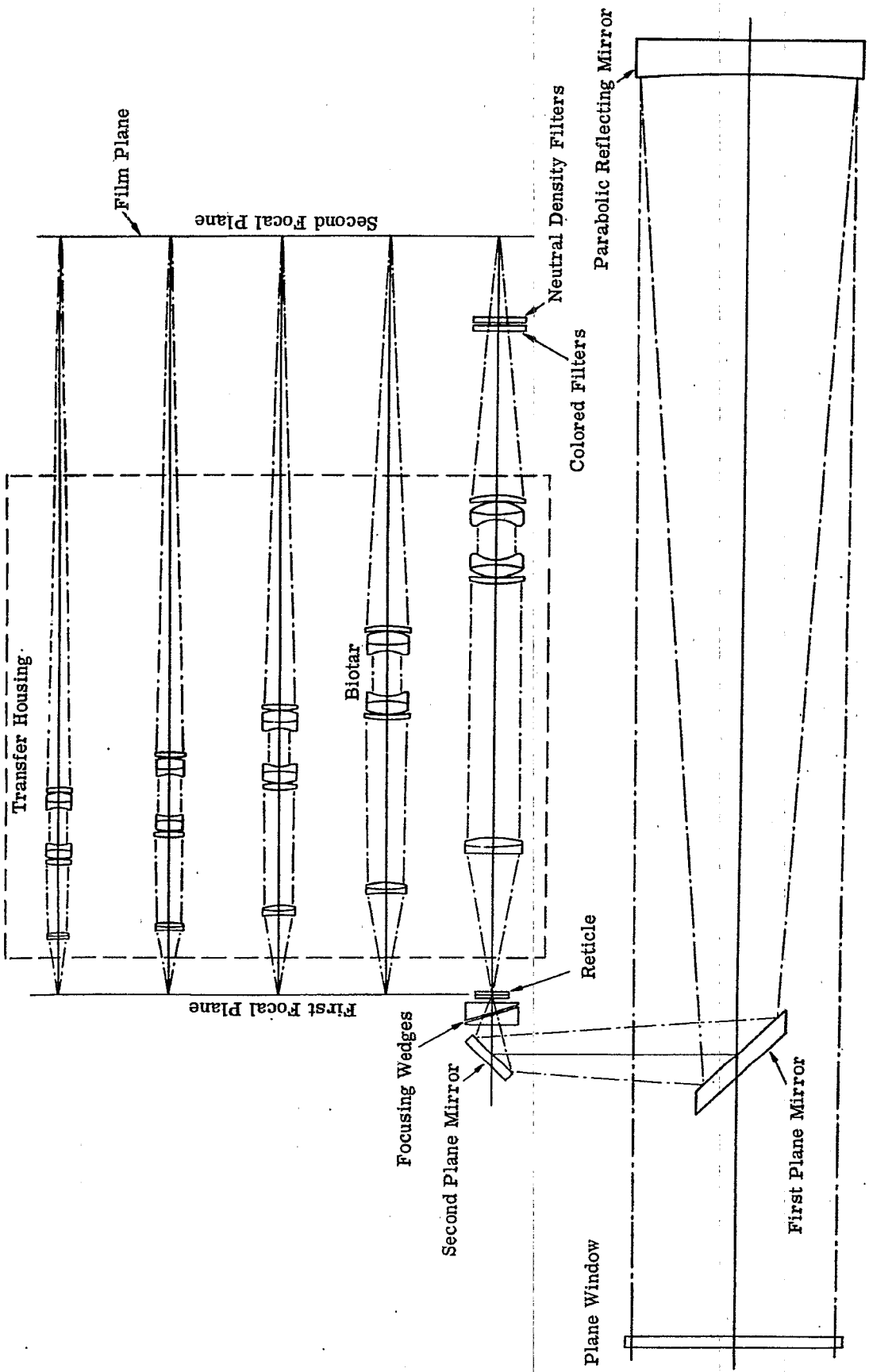


Figure 19.22- R. O. T. I. Mark II, Optical arrangement. (Courtesy of Perkin Elmer Corp.)

19.5.2.4 As is known from the previous discussion, the primary image is heavily afflicted with coma, and the astigmatism is only partially corrected by placement of the stop at the window. The doublet lens, placed behind the focal plane, serves three functions. First it acts as a Ross lens to neutralize the coma resulting from the paraboloid; second, it is a collector lens to turn the rays into the reimaging Biotar system; finally it imparts an initial magnification (2x) to the image at the prime focus.

19.5.2.5 The five imaging systems can be inserted into the system at will. These have magnifications ranging from 1x to 5x in unit power steps. With the doublet lenses working at 2:1, the Biotar type imaging lenses work at magnifications of 1/2, 1 . . . . 2-1/2x.

19.5.2.6 Where is the Schmidt type correcting surface characteristic of the reflector corrector design? Normally, it would be at approximately the location of the window. However a conjugate stop position exists in this system, namely the midpoint of the Biotar imaging lens. The aspheric correction required is put on the two innermost surfaces of this lens, those adjacent to the aforementioned stop position. This aspheric correction is for the purpose of eliminating the zonal spherical aberration only, since the primary spherical is taken care of in the design of the refractive elements. KZF Schott glass is employed in the system to reduce considerably the secondary spectrum.

19.5.2.7 After passing through the imaging lens, the rays traverse filters, colored or neutral as desired. The latter are automatically controlled by photoelectric means. The final focal plane is at the gate of the 70 mm camera, with a 2-1/4 inch square field of view.

19.5.2.8 Recapitulating, the five systems give a range of focal lengths from 100 inches F/4 to 500 inches F/20 (approximate figures, not allowing for the occlusion by the first Newtonian mirror). When enough light is available, the 500 inch focal length reaches into extremely long distances for the target missile, and the writer believes this has proved the most used focal length. Tracking is effected by means of two operators, one for azimuth and one for elevation, with appropriate telescopes. An elaborate electrical control system is very effective in the accuracy of tracking. A photograph of ROTI is shown in Figure 19.1 .

19.5.3 Igor.

19.5.3.1 This instrument is somewhat smaller than ROTI but serves essentially the same purpose, i.e., the observation of missiles in flight to extreme distances. The original specifications envisaged some situations where a 70 mm camera (2-1/4" x 2-1/4") was to take pictures under some circumstances at the prime focus (P) as shown in Figure 19.23. Consequently the image at this point had to have complete correction for all aberrations over its 2-1/4 inch by 2-1/4 inch format. The combination of Schmidt Plate (S), primary mirror (M), Newtonian mirrors (N1) and (N2), and Ross lens (R) as shown in Figure 19.23 accomplished this purpose very well. Removal of (S) and (R) allows for infrared and ultraviolet measurements if desired. Focussing of the system from infinity to 3000 yards is effected by motion of the Ross lens (R) towards the mirror (N1). The total travel to cover this range is 3/16 inch.

19.5.3.2 The system works at F/5 with a clear aperture of 18 inches, so that the focal length of the system at prime focus is 90 inches. The focal length of the paraboloid itself is 118 inches, reduced to the required 90 inches by the Ross lens. A collector lens at (P) and a re-imaging system at (L) enables final imagery on a 70 mm camera at (F).

19.5.3.3 For a change to longer focal lengths and magnified images a system of Barlow lenses is employed. A Barlow lens is illustrated in Figure 19.24. Suppose rays are converging to a prime focus at P. A negative lens B is inserted into the path and diverges the rays to a more distant focus at F. The image is magnified by the ratio BF/BP.

19.5.3.4 Referring to Figure 19.23, the 1x system of collector and re-imaging lenses are removed and one of the Barlows, B_2 (2x), B_4 (4x) or B_5 (5.7x) is inserted, transferring the final image to the camera at (F) under the particular desired magnification. The highest available magnification corresponds to a focal length in excess of 500 inches.

19.5.3.5 The variable density filter arrangement at (D) is comprised of two oppositely rotating continuously varying density filters, slightly inclined to each other and to the axis, to eliminate multiple reflections. Two are needed to keep the field uniform. As in ROTI, the image is quite good over the field of view in all powers.

19.5.4 S.M.T. (Small Missile Telecamera).

19.5.4.1 The requirements for this instrument are similar to those for the satellite tracking camera except that a relatively small field of view is required, namely that of a 70 mm camera. High light gathering power (low F/#) is indicated for the tracking of small, high velocity missiles. The focal length of the system is 100 inches with a 30 inch aperture, so that the system is working at F/3.3, and the field is 1° 18'. This system

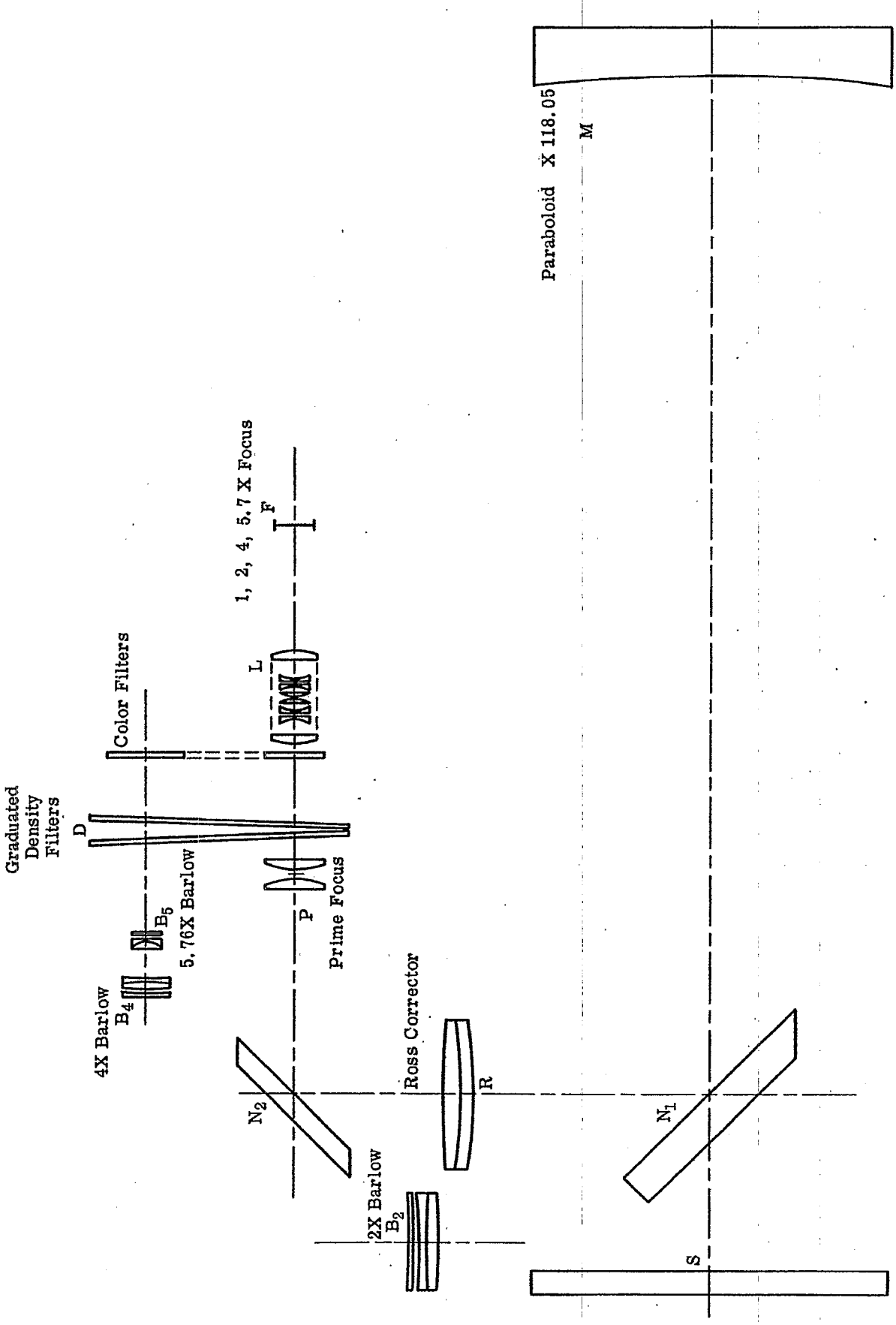


Figure 19.23-I.G.O.R. Optical arrangement. (Courtesy of American Optical Co: Drawing No. 013-0032)

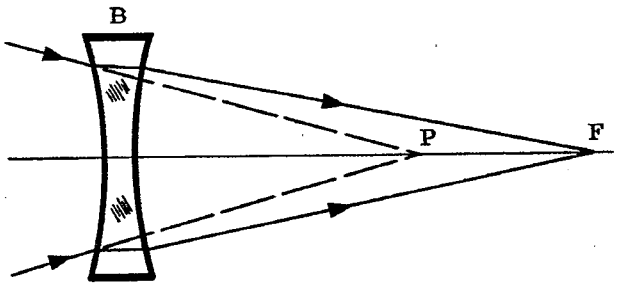


Figure 19.24- A Barlow lens.

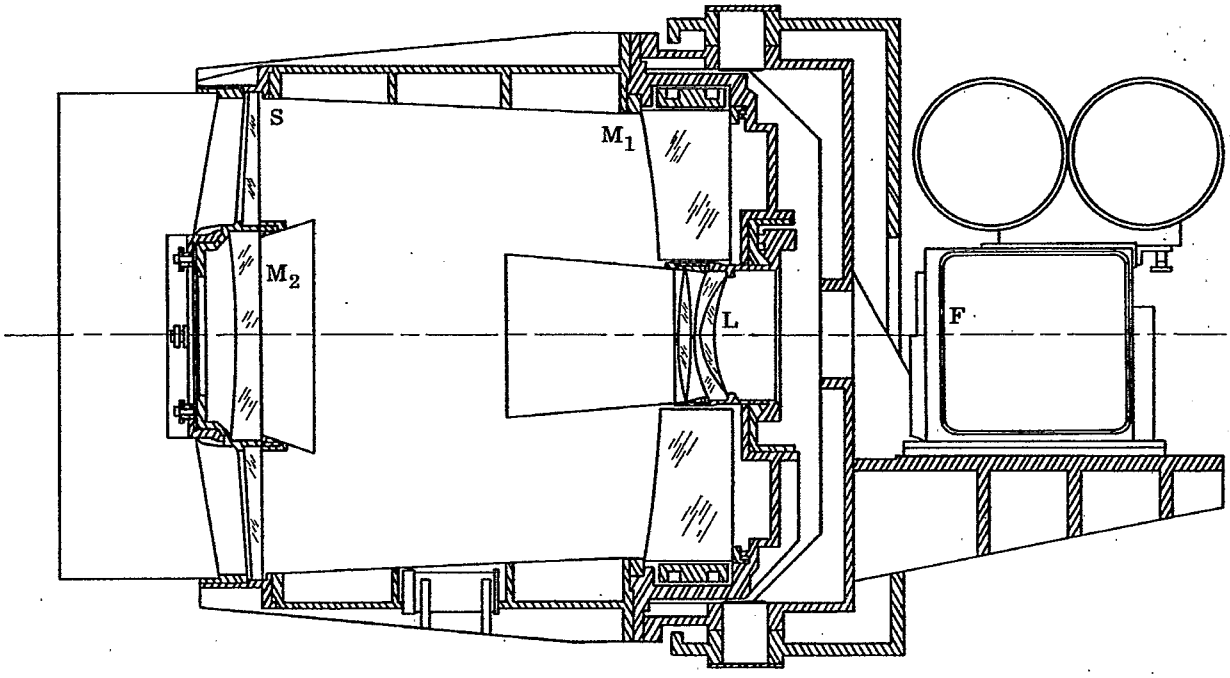


Figure 19.25- S. M. T. Optical arrangement.
(Courtesy of Perkin Elmer Corp.)

is similar to the Cassegrain arrangement in that two mirrors are involved, with the final image behind the hole in the primary as shown in Figure 19.25.

19.5.4.2 Light enters from the left and strikes the spherical mirror (M), after passing through the Schmidt plate (S). The convergent beam is interrupted by the mirror (M_2), which is actually a lens with its rear surface reflecting (often known as a Mangin mirror). The beam thus reflected a second time traverses the lenses placed in the central hole in the primary before reaching the film plane at (F).

19.5.4.3 This arrangement delivers an excellent flat field over its 2-1/4 inch x 2-1/4 inch field. The Mangin mirror and hole lenses afford enough degrees of freedom to correct for the coma and astigmatism of the shortened Schmidt arrangement, as well as to balance out the color aberrations they themselves bring in. The missile trajectory is predicted beforehand and the instrument is aimed from stellar observations made with a calibration camera.

19.5.4.4 It will be noted that heavy emphasis has been given to the optical arrangement of these complex instruments and their illustration of the optical principles discussed earlier. No space is available here for the description of their complex mechanical, electrical and electronic construction which tax all the resources of modern technology.

20 APPLICATIONS OF THIN FILM COATINGS

20.1 INTRODUCTION

20.1.1 General uses of thin films. Quite frequently optical components are coated with thin layers of various solid materials for the purpose of altering either their physical or optical properties. As an example of the former purpose, aluminum mirrors are often coated with a thin layer of silicon monoxide in order to increase their resistance to abrasion and chemical attack. The addition of this layer alters the spectral reflectivity of the mirror, although this is not the primary purpose of such a coating. More frequently, however, thin film coatings are used for the primary purpose of altering the spectral reflection and transmission of optical components. Sometimes a thin film coating consists of only one layer deposited upon a suitable substrate. In other cases, many layers, often as many as forty or fifty, are used to produce a given optical filter. Hence this type of filter is called a multilayer filter, or simply a multilayer. In Section 20, the term multilayer is used as a generic term for such thin film coatings, even though the "multilayer" coating may consist of a single-layer antireflection coating on a glass surface.

20.1.2 Typical applications.

20.1.2.1 Antireflection coatings. Whenever light traverses an interface between two media of different refractive index, such as an air-glass interface of a lens, some of the light is reflected. Often the spacing of optical elements is such that these reflections are manifested in the image plane as "flare images". Before the advent of antireflection coatings, many otherwise acceptable lens configurations were rejected because they produced these flare images. The coating of optical surfaces with antireflection coatings has practically eliminated this problem. It is important that antireflection coatings be applied to infrared optical components, such as lenses or domes, which contain germanium, silicon, or other materials with a high refractive index. The loss of light at uncoated surfaces would be prohibitive otherwise. Antireflection coatings are discussed in more detail in Section 20.3.

20.1.2.2 Achromatic beam splitters. Many optical devices, such as interferometers, range finders, optical gunsights, utilize beam splitters which divide a light beam and divert it into two directions. Thin metal films have been used as beam splitters for many years, but they are inefficient because the metal absorbs part of the light. More recently, multilayer beam splitters have been developed which are much more efficient, because they contain only non-absorbing materials. Less than one percent of the light is absorbed in a typical multilayer beam splitter; the remaining 99% of the light is either reflected or transmitted. The properties of multilayer beam splitters are expounded in Section 20.7.

20.1.2.3 Color filters and band-pass filters. Multilayer filters are used to transmit (or "pass") a broad band of wavelengths in one spectral region, but attenuate in other regions. For example, a multilayer filter is available which transmits more than 90% in the blue but has a transmission of less than 0.5% in the green and red. This multilayer filter is superior to the conventional glass or dyed-gelatin absorption filters, which have a much lower transmission in the blue. Similar types of band-pass filters have been developed for the ultraviolet and infrared spectral regions. The spectral transmission of some typical multilayer filters is shown in Figures 20.86, 20.91 and 20.92; a general discussion is given in Sections 20.5.2 and 20.6.2.

20.1.2.4 Color-selective beam splitters. Figure 20.1 shows a multilayer which is used as a color-selective beam splitter. In this example, the beam splitter transmits blue light, but reflects the green and red. Such beam splitters are useful as color separation devices in color photography and color television. This type of beam splitter is often called a dichroic mirror; its properties are explained in Section 20.7.2.

20.1.2.5 Narrow pass-band (interference) filters. Multilayer filters are used to transmit a narrow band of wavelengths. For reasons which are described in 20.10.2.2, these narrow-band filters are called interference filters, although in a technical sense all multilayer filters are interference filters, because they depend upon the interference of light reflected from the various films. One type of interference filter which is manufactured commercially in large quantities has a pass band which is from ten to twenty millimicrons wide in the visible spectral region. Custom-made filters have been produced which have a pass band as narrow as 0.1 mμ. A filter of this type has been used to isolate one of the sodium D lines at 589.0 mμ from its neighbor at 589.6 mμ. Such filters have many potential uses in the field of spectro-chemical analysis and are discussed in more detail in Section 20.10.

20.1.2.6 Semi-transparent mirrors. Multilayer mirrors have been produced which not only have a high reflectivity, but also transmit almost all of the light which is not reflected with a small absorption loss. A typical multilayer mirror might reflect 95% of the incident light and transmit 4.5%, the remaining 0.5% being absorbed or scattered. These multilayer mirrors have a much lower (absorption) loss than the conventional semi-transparent films of silver or aluminum and are useful for coating the plates of a Fabry-Perot interferometer or the ends of an optical maser. The spectral reflectivity of a semi-transparent metal mirror

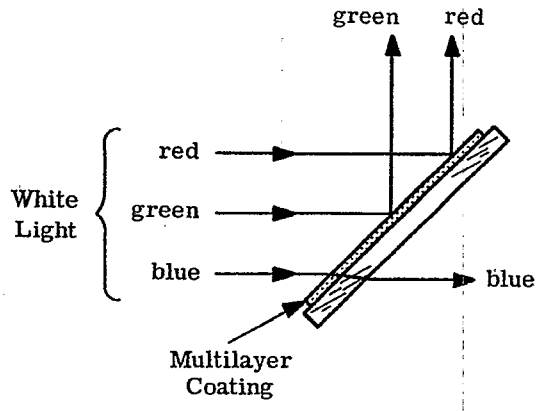


Figure 20.1-A color selective beam splitter which reflects the red and green, but transmits the blue.

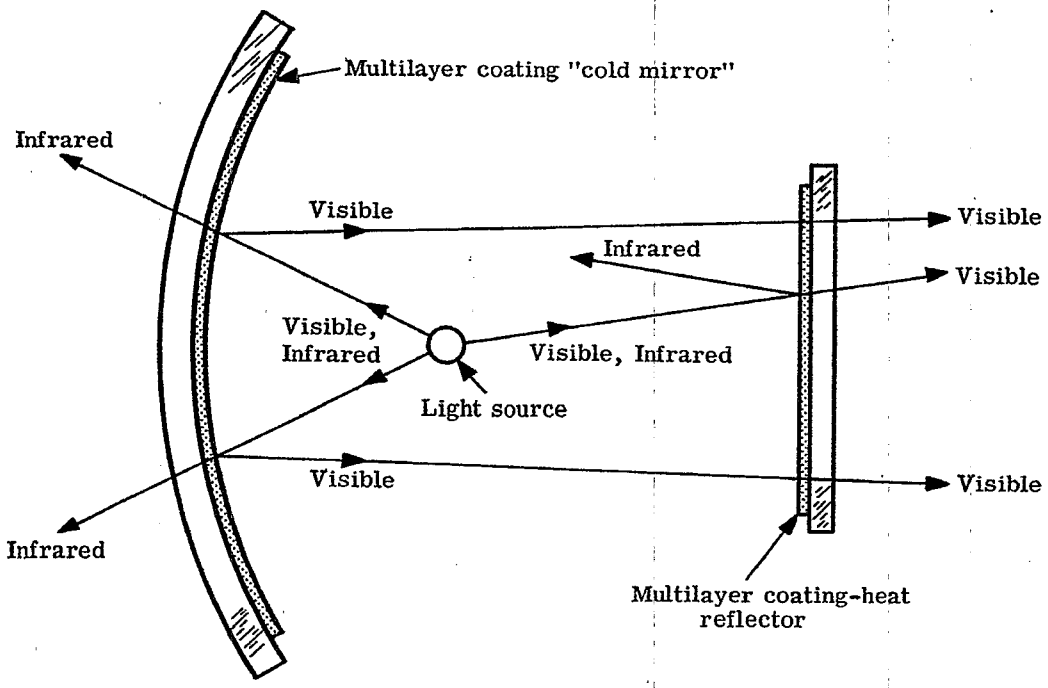


Figure 20.2-A cold mirror and heat reflector reduce the heat directed towards the film in a projection system.

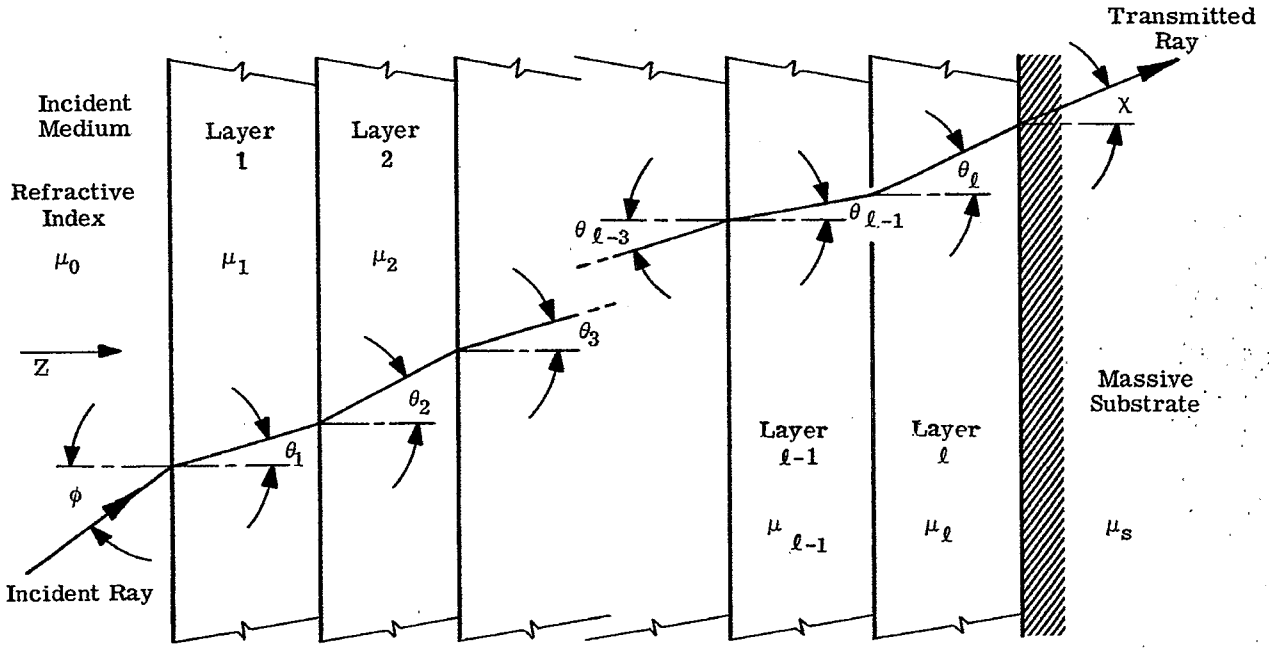
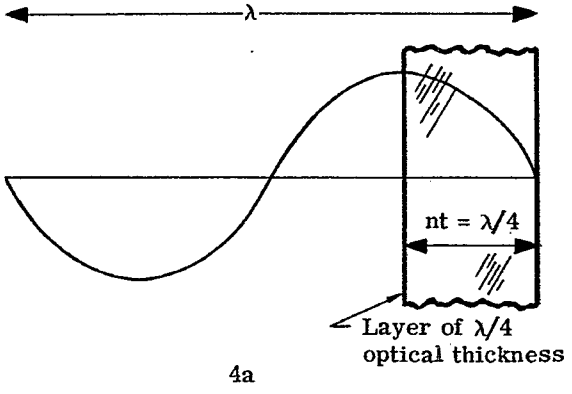
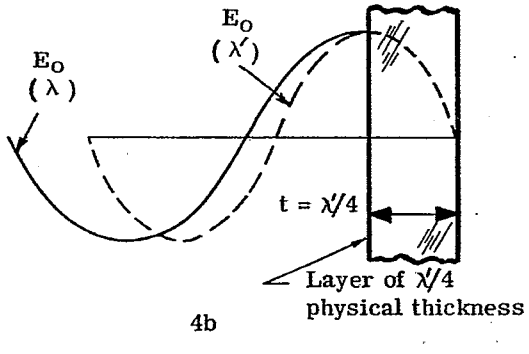


Figure 20.3 - Nomenclature used in designating the thickness, refractive index, and angle of refraction θ in each of the layers. For sake of clarity, the reflections which take place at each interface are not shown.



4a

Figure 20.4 (a) - A comparison of the wavelength of light (in vacuo) with the optical thickness of a quarter-wave layer.



4b

Figure 20.4 (b) - An instantaneous picture of the electric vector of a wave entering a denser medium. (see text 20.1.3.4)

is usually quite "flat," whereas the reflectivity of a typical multilayer mirror changes quite rapidly with wavelength. Further information is presented in Section 20.8.

20.1.2.7 Heat control filters. One type of a multilayer mirror, called a cold mirror, is used to reflect the visible light and transmit the infrared. Figure 20.2 shows a typical use of a cold mirror as a reflector behind a light source in a film projector. The cold mirror reflects the visible light towards the film, but permits the heat in the infrared to pass out the back. This system is even more efficient if a heat reflector is inserted between the arc and the film. This reflector, which has a high transmission in the visible spectral region but a high reflectivity in the infrared, reflects the heat away from the film but permits the visible light to pass through to the film with little attenuation. Another type of heat control filter is the cover glass which is placed over solar cells which are used to power a space vehicle. These cover glasses are designed to reflect the radiant energy at wavelengths longer than 1.2 μ. The radiant energy from the sun in the wavelength range from 1.2 μ to 2.5 μ does not generate power, but only increases the temperature of the solar cell, thereby decreasing its efficiency. Heat control filters are discussed briefly in Sections 20.5.3 and 20.6.2.

20.1.2.8 High-reflectivity mirrors. By overcoating aluminum and other metals with dielectric films, it is possible to obtain a reflectivity as high as 99.5%. The spectral reflectivity of a typical overcoated mirror is shown in Section 20.8.

20.1.2.9 Polarizers. Multilayers can be used to produce linearly polarized light. They are particularly useful in the infrared, where conventional polarizers which utilize birefringence cannot be used because most optical materials are optically isotropic. This application of multilayers is not discussed in Section 20; the reader can refer to Heavens[1] for further details.

20.1.2.10 Reflection filters. A multilayer has been developed which has a high reflectivity in certain spectral regions, but absorbs strongly in other regions. Such a mirror has been used to absorb the visible light and reflect the infrared. It should be noted that the type of reflection filter we are discussing here is different from the heat reflector described in 20.1.2.7. Both types of multilayer filter reflect the infrared, but the former absorbs the visible light, whereas the latter transmits the visible. This type of filter is not discussed in Section 20; further information is given in Heavens[2].

20.1.3 Nomenclature

20.1.3.1 General considerations. In Section 20, we consider only the case where the thin films are optically homogeneous; that is, the optical constants of a given layer do not vary along the direction of the propagation of the light, which is shown as the Z-direction in Figure 20.3. At the present time, almost all commercially manufactured multilayer filters are composed of films which are homogeneous. However, multilayer coatings which contain optically inhomogeneous films have some very interesting properties and it is possible that they will come into more extensive use in the future. A simple antireflection coating which contains an optically inhomogeneous film, sometimes called a "graded film", is described by Strong[3].

20.1.3.2 The multilayer stack. An idealized multilayer stack is shown in Figure 20.3. It consists of a total of ℓ layers deposited upon a substrate which has an optical constant $\mu_s = n_s - j k_s$. Each of the layers has a physical thickness t_i and optical constant $\mu_i = n_i - j k_i$, where n_i is the refractive index and k_i is the absorption coefficient. It should be noted that the absorption coefficient (represented by a lower case "k") is related to the parameter "K" in Section 21.2 by the relationship $k_i = -n_\nu K_\nu$. The light is incident at an angle ϕ from a non-absorbing incident medium of refractive index n_o. Since the layer boundaries are parallel, the angle of refraction in the i^{th} layer, θ_i, is determined simply from Snell's law, if the layers are non-absorbing. Thus each layer in the stack is specified by three parameters, namely t_i, n_i, and k_i. These quantities, along with n_o, n_s, and k_s, completely specify the optical properties of the multilayer. Given these quantities and ϕ it is a straightforward, although tedious, task to calculate the reflectivity R and transmission T of a multilayer as a function of the wavelength λ of the incident light. Methods of computing R and T are presented in Sections 20.1.5, 21.2.8, and 21.2.12.

20.1.3.3 The wave number. The retardation of phase of the i^{th} layer is defined* as

$$\delta_i = 2\pi \sigma n_i t_i \qquad (1)$$

where σ is the wave number of the incident light, $\sigma = 1/\lambda$. The wave number is proportional to the

* S_i is related to the parameter B_ν defined in Equation 21 - (47) by the relationship $2\delta_i = \beta_\nu$.

(1) All references are listed separately at the end of this section.

frequency of the light. In computing the spectral transmission of a multilayer filter, there are many advantages to using some parameter, such as σ, which is proportional to frequency rather than wavelength. One advantage is that many spectral transmission curves have even symmetry about some point, when plotted on a frequency scale, whereas the curves are quite asymmetrical when plotted versus wavelength. This is illustrated by Figures 20.32 and 20.34, which show the reflectivity of the same coatings. It is evident that the curves plotted versus frequency, as in Figure 20.34, have even symmetry about the center, whereas the wavelength plot of Figure 20.32 is quite asymmetrical. Another advantage of using frequency as a variable is that very often maxima and minima in the reflectivity curve are spaced at equal intervals on a frequency scale, whereas on a wavelength scale they are spread out in the long wavelength region and compressed together in the short wavelength region. This is illustrated in reflectivity curve shown in Figure 20.59. The maxima and minima near 500 mμ are so close together that they cannot be plotted accurately, whereas they would be spaced at equal intervals if plotted versus σ. In many cases in Section 20 we use as a variable a dimensionless quantity, $g = \lambda_o/\lambda$, which is proportional to frequency. The disadvantage of using σ as a variable rather than λ lies in our educational system; most persons are unfamiliar with wave number. Most of us have been educated to think in terms of wavelength and thus we associate "green light" with a wavelength of 540 mμ rather than with a wave number of 18519 cm^{-1}. A useful factor to remember in converting wavelength into wave number is that

$$1\mu = 1000\ m\mu = 10{,}000\ A \quad \text{is equivalent to } 10{,}000\ cm^{-1}.$$

Thus using the fact that "ten thousand angstroms 'equals' ten thousand wave numbers", and the fact that the wave number is inversely proportional to wavelength, we see that

5000A is equivalent to $20{,}000\ cm^{-1}$

2500A is equivalent to $40{,}000\ cm^{-1}$

and so on.

20.1.3.4 The QWOT. Not infrequently multilayers are composed entirely of dielectric (non-absorbing) materials. In this case it is convenient to refer to the thickness of the layers in terms of their optical thickness, which is defined as the product of the geometrical thickness, t_i, and the refractive index, n_i. Reference is frequently made to the quarter-wave optical thickness, QWOT, which is defined as

$$QWOT = 4\, n_i\, t_i \quad . \tag{2}$$

Since n_i is a dimensionless quantity and t_i has the dimensions of length, the QWOT also has the dimensions of length and is usually expressed in units of microns or millimicrons. Thus, if a layer has a QWOT of 550 mμ, this means that one-quarter wavelength of light at 550 mμ has the same length as the optical thickness of the layer, as illustrated in Figure 20.4(a). Sometimes one refers to a film of quarter-wave optical thickness simply as a "quarter wave". A note is interpolated to delineate clearly exactly what is drawn in diagrams such as in Figures 20.4, 20.55, 20.107, and others. This represents a comparison of the optical thickness of the film with the wavelength in vacuo of the incident light. To elucidate this point, at a given instant of time, the electric vector E of a light wave propagating in a homogeneous film of refractive index n can be represented by

$$E = E_o \cos\left(\frac{2\pi}{\lambda'} z\right)$$

where z is the coordinate shown in Figure 20.3. The wavelength λ' in the foregoing equation has a "primed" superscript as a reminder that λ' is the wavelength in that medium, of index n. λ' is related to the wavelength λ in vacuum by

$$\lambda' = \frac{\lambda}{n}$$

and consequently λ' changes when the wave enters a medium of different refractive index, as is shown in Figure 20.4b. Thus Figure 20.4b represents the actual E field in the film and vacuum at a given instant of time and the thickness of the film is represented by its actual physical thickness. Another approach, which is extensively used in Section 20, is to compare the optical thickness of the film with the wavelength in vacuo. This is accomplished by writing equation as

$$E = E_o \cos\left(\frac{2\pi}{\lambda} n z\right).$$

The quantity n z is the optical thickness. The film shown in Figure 20.4a is drawn thicker in proportional to its optical thickness, but this is compensated for by the fact that the wavelength which is shown is not the wavelength in the film, but the vacuum wavelength.

20.1.3.5 "H" and "L" layers. Quite often multilayers are composed of films which are all of quarter-wave optical thickness, or some multiple thereof. In this case it is convenient to use a shorthand notation to specify the design. The letters "H" and "L" are used to specify films of high and low refractive index, respectively, which have the same QWOT. For example, the design of the double-layer coating shown in Figure 20.5 is designated as

glass HH L air,

where $n_H = 1.70$ and $n_L = 1.38$. The optical thickness of the film next to the glass is two quarter waves, or a single half wave. Similarly, the design of the double-layer coating shown in Figure 20.37 is

silicon H L air,

where $n_H = 2.40$ and $n_L = 1.38$.

20.1.3.6 The reflectivity and transmission. In Section 20 reference is made to computed values of the reflectivity, R, and the transmission, T. R and T are synonymous with the "reflection coefficient", and the "transmission coefficient" which are defined in terms of a time average of the Poynting vector, as is specified in Equations 45 and 45a in Section 21.2.6. In Section 20 the terms "reflectance" and "transmittance" are reserved for measured, rather than computed, values.

20.1.3.7 Non-normal incidence. When light is incident upon a multilayer at oblique incidence, both R and T must be computed separately in each plane of polarization. Thus one refers to R_p and T_p in the "p" plane of polarization when the electric vector is parallel to the plane of incidence and to R_s and T_s in the "s" plane of polarization when the electric vector is perpendicular to the plane of incidence. In general, if unpolarized light is incident upon a multilayer at non-normal incidence, both the reflected and transmitted light is partially plane-polarized. If the incident light is elliptically polarized, then the degree of elliptical polarization of both the reflected and transmitted light is altered. This is because not only the reflection and transmission are different in the two planes of polarization, but also because the phase shift upon reflection is different for the two planes of polarization. If the light which is obliquely incident upon a multilayer is initially unpolarized and if the light detector, such as the human eye or a photographic plate, is not sensitive to the polarization of the light, then the polarizing effect of the multilayer can be neglected. In this case we simply refer to the average reflectivity, R_{av}, and the average transmission, T_{av}.

$$R_{av} = \frac{1}{2}(R_p + R_s), \qquad T_{av} = \frac{1}{2}(T_p + T_s). \tag{3}$$

20.1.4 Analogies.

20.1.4.1 Electrical transmission lines. It is useful to note the interesting analogy between the propagation of light through a thin film and the propagation of radio waves in an electrical transmission line. The propagation equation for the thin film is identical with the transmission line equation if one identifies the electric vector and magnetic vector in the thin film with the electrical voltage and current in the transmission line. The optical thickness* of the thin film is analogous to the "electrical length" of the transmission line, while the refractive index of the thin film is analogous to the "characteristic admittance" of a section of transmission line. A dielectric thin film corresponds to a "lossless" transmission line. The refractive index of the substrate and incident medium are analogous to the load admittance and the "characteristic admittance of the generator". The equations for the "optical admittance" of a thin film are identical to the equations for admittance of a transmission line. Graphical devices which have been invented for computing the voltage standing ratio of an electrical transmission line, such as the Smith Chart and the Admittance Chart, are used for computing the reflectivity of a stack of thin films[4,5,6]. This analogy cannot be extended too far, however. Shunt transmissions lines and lumped constant elements such as resistors, capacitors, and inductors can be added to an electrical transmission line. No exact analogous devices exist in thin film optics, although at certain wavelengths a thin gold or silver film can be represented as an inductance and an inconel film as a pure resistance.

20.1.4.2 Similarities with geometrical optics. There are some similarities between problems in multilayer filters and lens design. In both cases, it is a straightforward task to compute the performance of a given system, once the design has been specified. The design of a multilayer filters is given by specifying the thickness and optical constants of each layer, and the substrate and incident medium. The design of a lens is specified by the physical dimensions of each optical component and its refractive index. In both cases, it is quite difficult to synthesize a design. In order to synthesize a multilayer filter design, it is

* Note that the electrical length of a transmission line does not change with admittance as the optical thickness of a film changes with index.

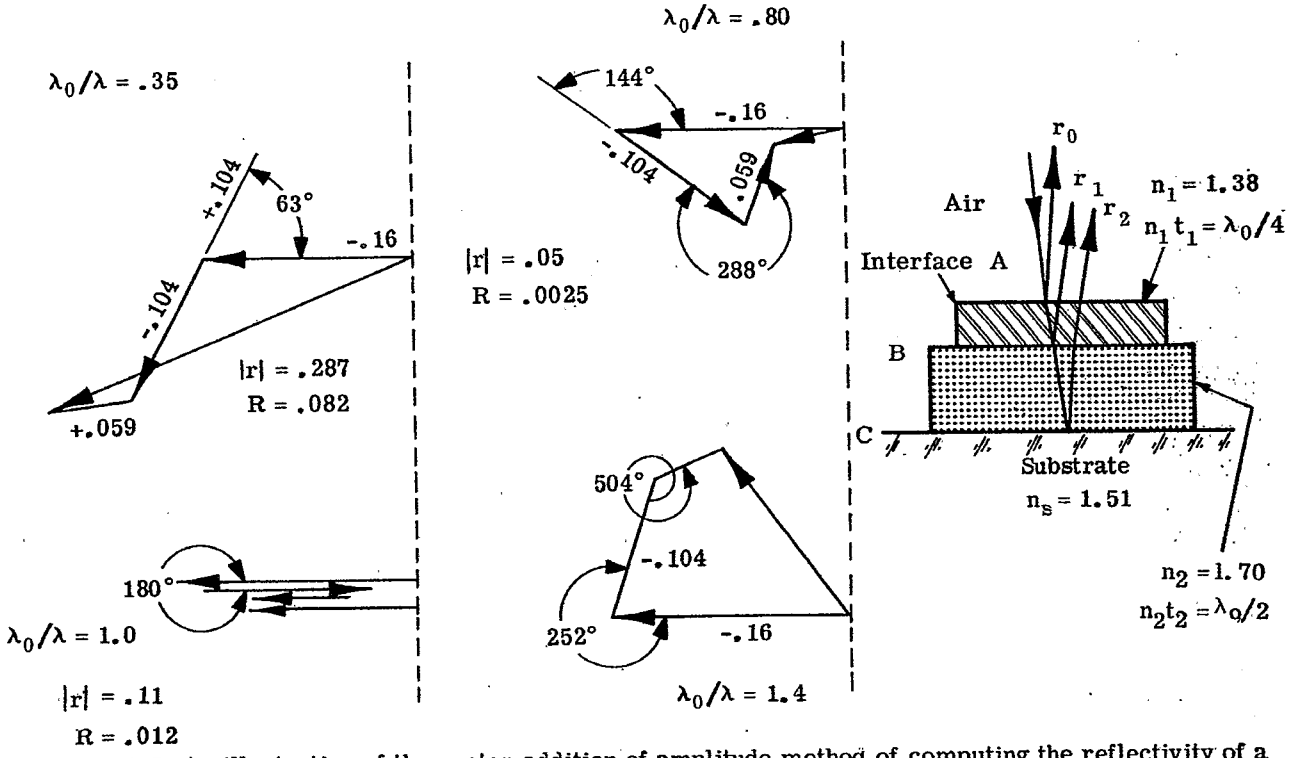


Figure 20.5 - An illustration of the vector addition of amplitude method of computing the reflectivity of a antireflection coating. (Vectors are not drawn exactly to scale.)

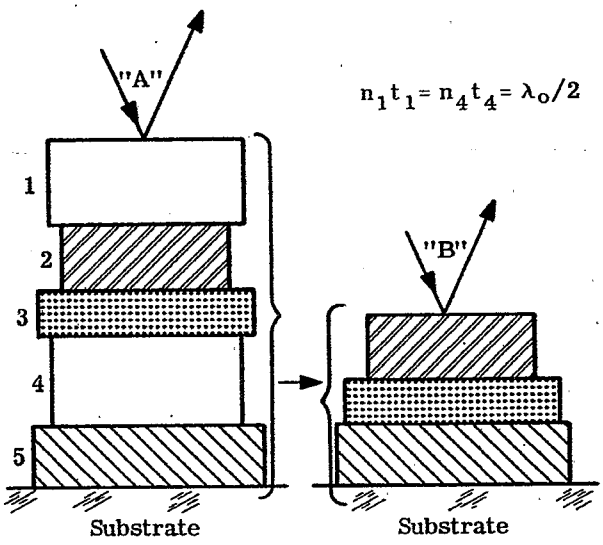


Figure 20.6 - Stack "B" is equivalent to stack "A" when layers 1 and 4 in stack "A" are absentee.

necessary to choose the thickness and optical constants of each layer so that the filter has a prescribed spectral transmission or reflectivity. In synthesizing a lens, one must choose the curvature of the surfaces, spacing and refractive index of the optical elements so that the lens has a prescribed amount of aberration at various points in the image plane. In both lens design and multilayer filters, only approximate methods have been developed to accomplish this synthesis.

20.1.4.2.2 There are several differences between multilayer filters and lenses. The first difference is that the parameters which specify the design of a lens are given to a high degree of precision. The physical dimensions and the refractive indices of the optical components are usually specified to an accuracy of a few thousandths of a percent. In thin film optics, however, it is usually difficult to control the thickness of individual layers to a precision of better than one percent. The refractive index of thin films can vary widely, because the optical constants of thin film materials depend not only upon the thickness of the layer, but also quite markedly upon the conditions in the vacuum when the film is evaporated. For example, the refractive index of the material "silicon monoxide" can vary from 1.40 to 1.90 depending upon the partial pressure of oxygen during the evaporation.[7,8]

20.1.4.2.3 Another difference is that the design of a lens is usually patent, whereas the design of a multilayer can be kept secret quite easily. Even if the design of a lens were not specified, one could disassemble a lens and measure the refractive index and dimensions of its components. In order to ascertain the design of a multilayer filter, however, it would be necessary to "peel off" each of the layers, which might be as thin as a few hundred angstroms. This is practically impossible to accomplish. The amount of material in a layer which has a QWOT of 500 mμ is about thirty micrograms per square centimeter. Hence, it is quite difficult to perform a chemical analysis to determine the composition of the layers. Because of these facts, one finds quite frequently that the designs of multilayer filters and the materials which are used in their manufacture are kept secret for proprietary reasons. This secrecy has been a detriment to the progress in this field.

20.1.5 Methods of computing R and T*

20.1.5.1 Vector addition of amplitudes. This is an approximate method which is most useful for stacks which contain a small number of layers. It essentially neglects the multiple reflections which take place between various interfaces and hence is most accurate when the difference between the refractive indices of adjacent layers is small. As an example of the application of this method, consider the problem of computing reflectivity of the double-layer antireflection coating shown in Figure 20.5 . The incident light reflects from each of the three interfaces. The amplitude of the wave which is reflected at each interface is proportional to the Fresnel amplitude coefficient

$$r_i = (n_i - n_{i+1})(n_i + n_{i+1})^{-1} \qquad (4)$$

where n_{i+1} and n_i refer to the refractive index on each side of the interface. However, these waves are not in phase and hence their amplitudes must be added vectorially. The difference in phase between the wave reflected from interfaces A and B is $2\delta_1$, and similarly for the other interface. In the example in Figure 20.5 , the r_i are -0.16, -0.104, and 0.059, respectively for interfaces A, B, and C. From Equation 20-(1) we readily determine that 2_1 is 63° and 2_2 is 126° when λ_o/λ = 0.35. Extending this procedure to other values of λ_o/λ, we can construct the vector diagrams which are shown in Figure 20.5 . When λ_o/λ = 1.0, the vectors are colinear and hence the r_i can be added algebretically. In Figure 21.8 the reflectivity computed by the vector method is compared with the results of the more exact matrix method described in the next section.

20.1.5.2 The characteristic matrix.[11,12]

20.1.5.2.1 The electric field E and the magnetic field H at one boundary of a film are related to the fields E' and H' at the other boundary by two linear simultaneous algebretic equations. These equations can be written in matrix form:

$$\begin{bmatrix} E \\ H \end{bmatrix} = M_i \begin{bmatrix} E' \\ H' \end{bmatrix} \qquad (5)$$

where the matrix M for a non-absorbing film at normal incidence is

$$M_i = \begin{bmatrix} \cos\delta_i & j\, n^{-1}_i \sin\delta_i \\ j\, n_i \sin\delta_i & \cos\delta_i \end{bmatrix} \qquad (6)$$

* The derivations of the equations which are cited in this section for computing R and T are given in section 21.4 or in references 9 and 10.

and where $j = \sqrt{-1}$ and $\begin{bmatrix} E \\ H \end{bmatrix}$ and $\begin{bmatrix} E' \\ H' \end{bmatrix}$ represent column vectors. As is shown in Section 21.4, Equation 20-(5) can readily be extended to more than one layer and thus this can be regarded as a recursion relationship. The reflectivity of a multilayer is computed by first writing down the matrix M_i for each layer according to Equation 20-(6). Then the matrix product is computed, as for example in Equation 21-(129) in Section 21.4. For example, if a stack consists of three layers, the matrix product M is

$$M = M_1 * M_2 * M_3 \tag{7}$$

where the symbol * denotes a matrix multiplication. The matrix product M has in general four elements:

$$M = \begin{bmatrix} A & j\,B \\ j\,C & D \end{bmatrix} . \tag{8}$$

The four variables, A, B, C, and D, are all real variables if all the films are non-absorbing. However, only three of the variables are independent, since the determinate of the matrix M is unity and hence $A \cdot D + B \cdot C = 1$. The reflectivity R is computed from

$$R = \frac{(X - U)^2 + (Y - V)^2}{(X + U)^2 + (Y + V)^2} \tag{9}$$

where

$$X = n_o A + n_o k_s B \qquad U = n_s D$$
$$Y = n_o n_s B \qquad V = C - k_s D \tag{10}$$

If the substrate is non-absorbing, i.e. $k_s = 0$, conservation of energy requires that $R + T = 1$ and in this case T can be written:

$$T = \frac{4}{2 + A^2 \frac{n_o}{n_s} + D^2 \frac{n_s}{n_o} + \frac{C^2}{n_o n_s} + B^2 n_o n_s} . \tag{11}$$

20.1.5.2.2 Whenever the optical thickness of any layer is $\lambda/2$, λ, $3\lambda/2$, 2λ, etc. the δ of that layer is 180°, 360°, 540°, etc. and the characteristic matrix for that layer reduces to the unit matrix:

$$M = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} . \tag{12}$$

However, the unit matrix has no effect upon the matrix product and hence this "half-wave" film does not contribute to the reflectivity of the multilayer stack at that wavelength. In other words, any "half-wave" layer acts as though it is absent from stack - that is, it is an absentee layer (see Section 21.2.14). As an example, consider the five-layer stack which is shown in Figure 20.6 . It is specified that at some wavelength, λ_o, optical thickness of layers #1 and #4 is $\lambda_o/2$. Then at λ_o layers #1 and #4 are absentee, and the reflectivity of the five layer stack is equivalent to the reflectivity of the three-layer stack shown in Figure 20.6 , which is the five-layer stack with layers #1 and #4 removed.

20.1.6 The computation of R and T at non-normal incidence.

20.1.6.1 Extension of the normal-incidence equations. Equations 20-(4), 20-(6), 20-(10) and 20-(11) strictly apply to normal incidence. However, they can be extended to include a non-absorbing multilayer stack at non-normal incidence. This is accomplished as follows: [13]

Step 1. Given the angle ϕ in the incident medium, the angle of refraction, θ_i, is computed from Snell's law.

Step 2. The "s" plane of polarization is considered first. The refractive index of the layer, n_i, as it appears explicitly in Equations 20-(4) and 20-(6) is replaced by an effective index,

$$n_{eff} = n_i \cos \theta_i . \tag{13}$$

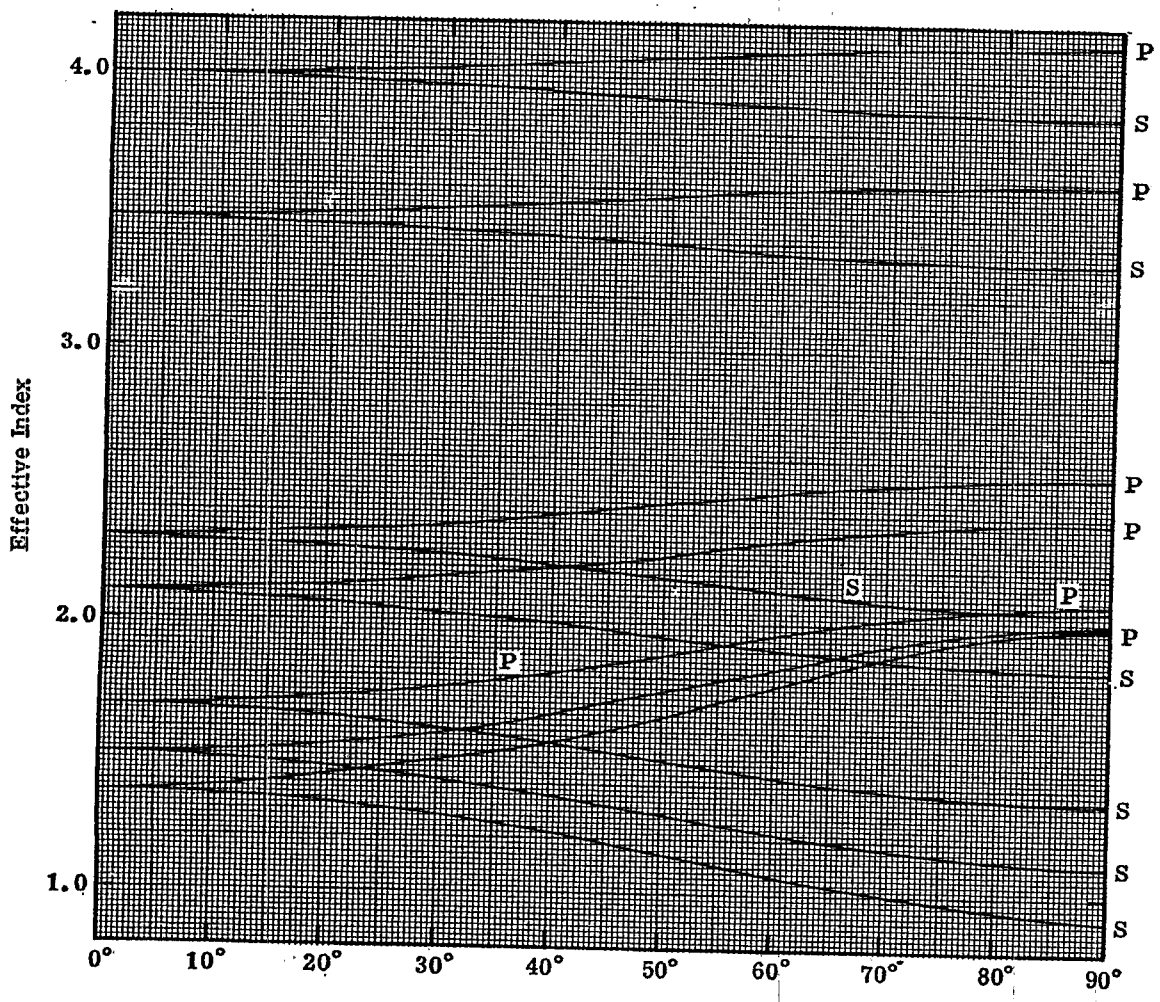


Figure 20.7 - The effective index (given by Eqs. 20.13 or 20.15) for the " p "and "s" planes of polarization, as a function of the angle of incidence in air.

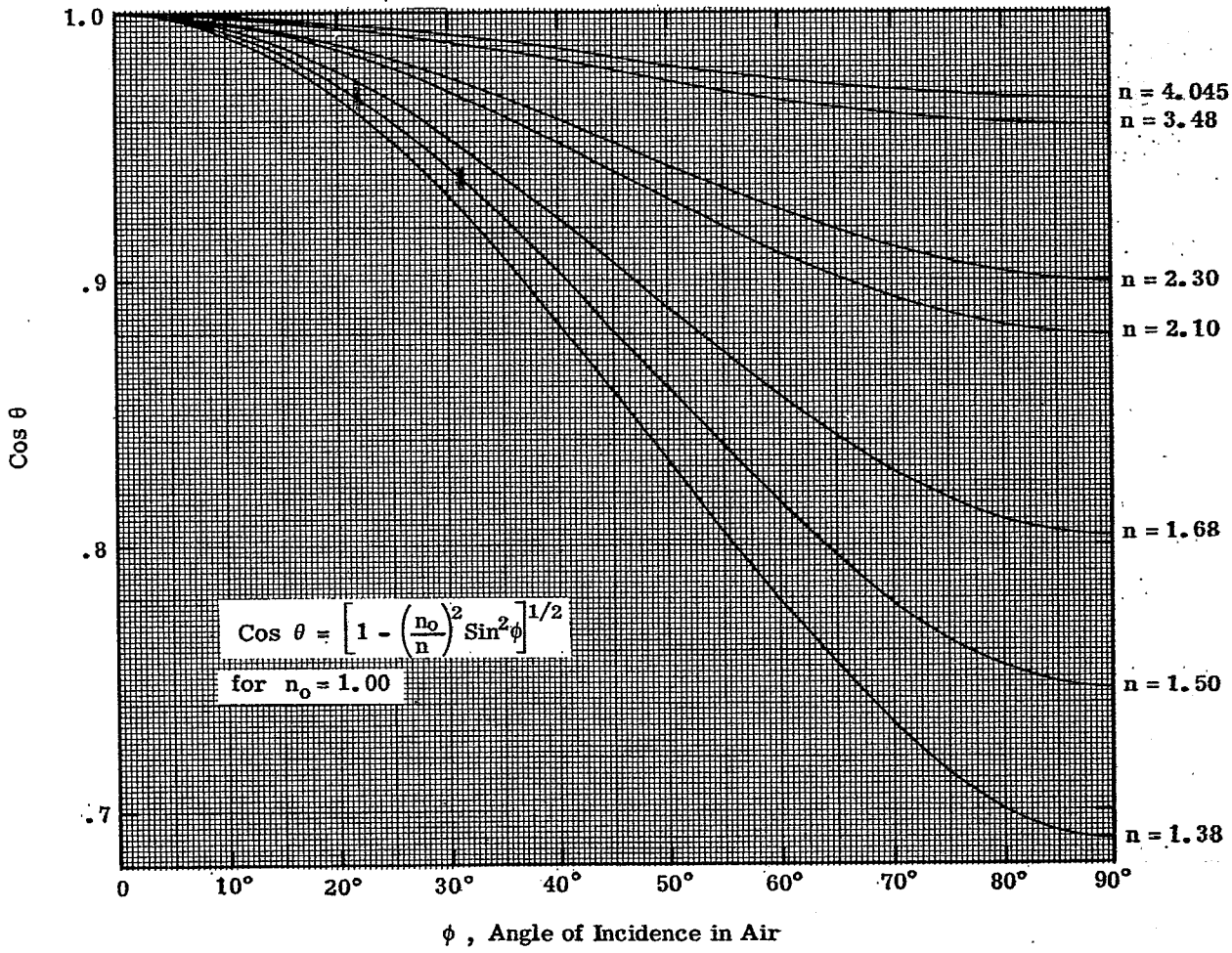


Figure 20.8 The effective thickness (normalized to 1.00 at normal incidence) as a function of the angle of incidence in air.

This substitution does not apply to the n_i which appears in Equation 20-(1). The indices n_o and n_s are replaced by the effective indices $n_o \cos \phi$ and $n_s \cos \chi$, respectively.

Step 3. The optical thickness of each layer is replaced by an effective thickness,

$$(n_i \; t_i)_{eff} = n_i \; t_i \; \cos \theta_i \; . \qquad (14)$$

Step 4. Having executed steps 1, 2, and 3, R_s and T_s are computed, using either the vector addition of amplitude method or the matrix method.

Step 5. The "p" plane of polarization is considered next. Rather than Equation 20-(13), we use for the effective index:

$$n_{eff} = n_i / \cos \theta_i \; . \qquad (15)$$

Again, Equation 20-(15) does not apply to the n_i in Equation 20-(1). The indices n_o and n_s are replaced by the effective indices $n_o / \cos \phi$ and $n_s / \cos \chi$, respectively.

Step 6. Having made the substitutions indicated in steps 3 and 5, R_p and T_p are computed, using either the vector method or the matrix method.

20.1.6.2 Use of the effective index.

20.1.6.2.1 Figure 20.7 shows the plot of the effective index as a function of ϕ for an incident medium of refractive index $n_o = 1.00$. As ϕ approaches 90°, the effective index approaches the limiting value of

$$n_{eff} = \sqrt{n_i^2 - n_o^2} \qquad (16)$$

and

$$n_{eff} = n_i^2 / \sqrt{n_i^2 - n_o^2} \qquad (17)$$

for the "s" and "p" planes of polarization, respectively. From Figure 20.7 we see that the effective index of a material with a large refractive index, such as germanium, changes by less than three percent between $\phi = 0$ and $\phi = 90°$. The materials with a lower index show a much larger change.

20.1.6.2.2 Figure 20.8 shows the fractional change in the effective thickness, $\cos \theta_i$, as a function of ϕ for various values of the index n_i. As one might expect, the change in the effective thickness between $\phi = 0$ and 90° is much greater for low-index materials than for high-index materials.

20.1.6.2.3 Since the effective thickness at oblique incidence is always less than at normal incidence, this means that the reflectivity and transmission peaks of multilayer filters shift to shorter wavelengths as ϕ increases. Although the author does not know of a rigorous proof of the statement made in the foregoing sentence. he has never found any exception to it in his work with multilayer filters.

20.1.6.3 Matched layers.

20.1.6.3.1 By definition, two or more layers in a multilayer stack are matched when there is a prescribed ratio between the optical thickness, $n_i \; t_i$, and hence between the δ_i of those layers.

20.1.6.3.2 As an example, consider the three-layer stack which is shown in Figure 20.9. For the purpose of this illustration, we shall arbitrarily define a matched condition to be:

$$n_1 \; t_1 \; : \; n_2 \; t_2 \; : \; n_3 \; t_3 \; = \; 1 \; : \; 2 \; : \; 2.5 \; . \qquad (18)$$

Equation 20-(18) states that when the optical thickness of the second layer is twice that of the first layer and the $n_3 \; t_3$ of layer three is 2.5 times $n_1 \; t_1$. If this film combination is tipped at an angle of 60°, then the optical thickness of each layer must be replaced by its effective thickness. Since the refractive index for each layer is different, the percentage change in the effective thickness is different. Referring to Figure 20.8 we see that the optical thickness of layer one is multiplied 0.98, layer two by 0.78, and layer three by 0.86.

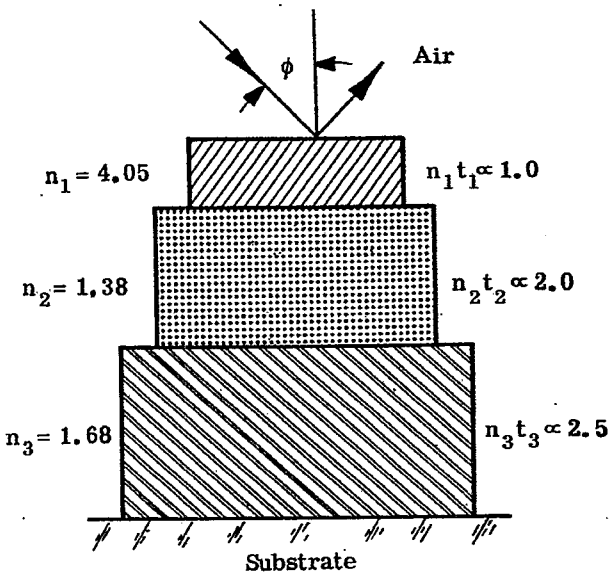


Figure 20.9 - Diagram of a three-layer stack. The layers are drawn proportional to their optical thickness.

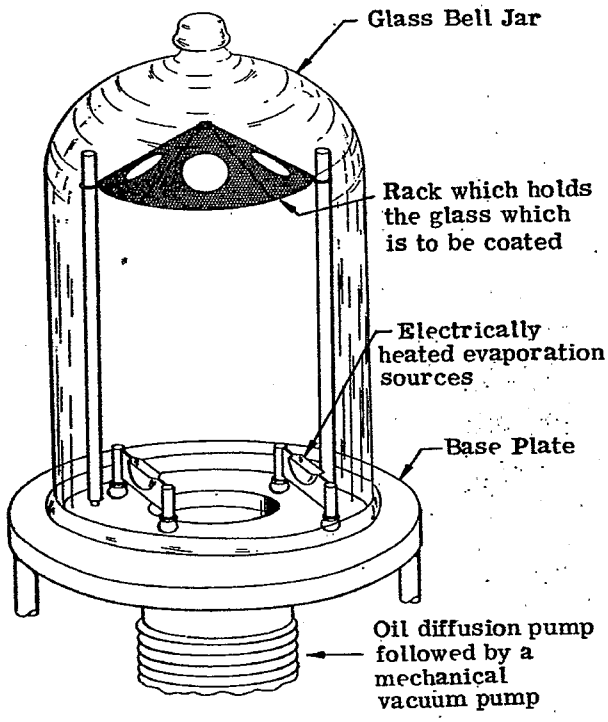
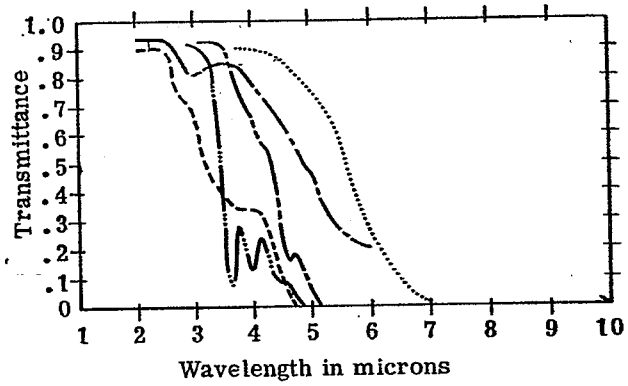


Figure 20.10 - The essential parts of an evaporator which is used to deposit thin films.



..... Sapphire, 2.6 mm

- - - - - Microglass, .005"

——— Quartz, 2 mm

- - - - - - - Corning No 0160, 2 mm

——— Vycor, 2 mm

Figure 20.11 - The optical transmission in the infrared of some substrate materials.

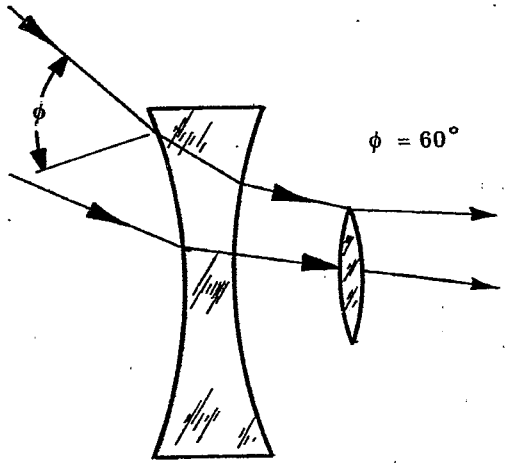


Figure 20.12 - The light is incident at a large angle upon the front surface of a wide-angle lens.

This means that at 60° incidence, the ratio of the thickness of the layers is no longer 1 : 2 : 2.5, but is

$$\left[n_1 \; t_1 \right] : \left[n_2 \; t_2 \right] : \left[n_3 \; t_3 \right] = 0.977 : 1.56 : 2.14 = 1 : 1.60 : 2.19 \qquad (19)$$

where the brackets $\left[\; \right]$ indicate "effective thickness". This means that the layers are no longer matched at 60° incidence, or for any other non-zero angle, for that matter. If all of the layers were composed of high refractive index materials, the deviation of the optical thickness ratios from the matched condition prescribed in Equation 20-(18) would be much smaller.

20.1.6.4 Layers matched at oblique incidence.

20.1.6.4.1 From the discussion in the foregoing subsection, we see that if the layers were deliberately made thicker in proportion to $\sec \theta_i$ then the layers are matched at non-normal incidence for a particular value of ϕ. In the foregoing example, if the $n_i \; t_i$ of the layers were in the ratio of

$$n_1 \; t_1 : n_2 \; t_2 : n_3 \; t_3 = 1.024 : 2.57 : 2.92 \qquad (20)$$

then at an incidence angle of 60° the layers will be matched with the prescribed ratio stated in Equation 20-(18).

20.1.6.4.2 It only makes sense to speak of matched layers provided there is more than one layer. In the case of a single layer if the $\delta_1 = 2\pi/\lambda \; (n_1 \; t_1)$ of that layer has some value at normal incidence, say, for example, 1.25 radians at some wavelength, λ_1. Then as ϕ increases, the effective thickness decreases. It is always possible to find some new wavelength λ_2, at which δ_1, is still 1.25 radians. If $n_o < n_1$, then λ_2 is less than λ_1. This shift to shorter wavelengths is clearly illustrated in reflectivity curves of single-layer reflecting coatings which are shown in Figure 20.20.

20.2 THE MANUFACTURE OF MULTILAYER FILTERS

20.2.1 Practical considerations. Using the computational methods which were described in 20.1.5, the spectral transmission and reflectivity of a given multilayer filter can be computed. However, if this filter is not to be a mere theoretical abstraction, but is to be actually manufactured, then we must keep in mind that there are certain practical problems which are encountered. Just as the lens designer is limited by the fact that the optical glasses which are presently manufactured have refractive indices within in a certain range, the films of a multilayer filter must be composed of materials which have certain specific values of refractive index.

20.2.2 Methods of deposition.

20.2.2.1 Progress in optical filming. Multilayer filters are manufactured by depositing solid films of various materials on an appropriate substrate. Although there are many methods which are used to deposit these films, such as chemical reaction in a vapor or liquid, sputtering, or anodization, the most important and widely-used methods is evaporation in a vacuum.* Consequently, progress in the production and manufacture of multilayer filters has for many years been related to advances in vacuum technology. Although the theory of multilayer filters has been known for more than a century, the production of such useful devices as antireflection coatings for lenses and silver-dielectric-silver interference filters did not start until the decade of the 1930's, when high-capacity oil diffusion pumps were developed which could evacuate a large volume to a pressure of less than 10^{-7} of atmospheric pressure. Antireflection coatings for lenses were used extensively during the Second World War. After the war, all-dielectric interference filters and multilayer beam splitters for use in the visible spectral region became commercially available. In the decade from 1950 to 1960, multilayer coatings have been produced for wavelengths as short as 110 mμ in the ultraviolet[15] and for wavelengths as long as 20μ in the infrared.[16]

20.2.2.2 The evaporator. Figure 20.10 shows the essential parts of an evaporator which is used to deposit thin films by evaporation in a vacuum. The circular pieces which are to be coated (i.e. the substrates) are placed in the holes in holder at the top of the chamber. The glass bell jar is sealed to the base plate and the entire chamber is evacuated by means of the oil diffusion pump and the mechanical pump to a pressure of less than 10^{-4} mm of mercury. The boats which contain the material which is to be evaporated are then electrically heated causing the material in the boat to evaporate and deposit in a thin solid film on the substrates. No mention has been made here of how the substrates are cleaned so that the films adhere well, or how the films are evaporated to a predetermined thickness. These topics are covered in detail in references 17 and 18.

* Methods of depositing thin films are summarized in reference 17.

MATERIAL	REFRACTIVE INDEX	TRANSMISSION RANGE	COMMENTS
Hot pressed Mg F_2 (Irtran 1)	1.37	200 mμ to 7.5 μ	
Calcium fluoride	1.42	180 mμ to 6 μ	
Barium fluoride	1.42	150 mμ to 13 μ	
Irtran 3	1.42	1000 mμ to 9 μ	
Fused quartz	1.46	180 mμ to 3.5 μ	1
Vycor (high-silica glass)	1.46	250 mμ to 3.5 μ	
Glass	1.51 to 1.70	320 mμ to 2.5 μ	2
Sapphire	1.70	<20 mμ to 7.5 μ	
Hot pressed ZnS (Irtran 2)	2.26	2000 mμ to 14 μ	
Arsenic trisulfide glass	2.40	800 mμ to 12 μ	
Irtran 4	2.40	1000 mμ to 20 μ	
Silicon	3.50	1100 mμ to 8 μ	
Germanium	4.05	1900 mμ to >22 μ	

NOTES:

1. The wavelength of the absorption edge in the ultraviolet depends upon the purity of the quartz.
2. In general, glasses with a lower refractive index are transparent to shorter wavelengths than glasses with a higher index. However, there are exceptions to this rule. Most glasses have a strong absorption band in the vicinity of 2.6 μ. Very thin glass plates (i.e. "cover slips") are transparent between 2.7 μ and 4.5 μ.

Table 20.1 - Common substrate materials.

MATERIAL	REFRACTIVE INDEX	RANGE OF TRANSPARENCY (See Note 10)		COMMENTS
		From	To	
Cryolite	1.35	<200 mμ	10 μ	1
Chiolite	1.35	<200 mμ	10 μ	1
Magnesium fluoride	1.38	230 mμ	5 μ	2, 3
Thorium fluoride	1.45	<200 mμ	10 μ	
Cerium fluoride	1.62	300 mμ	> 5 μ	4
Silicon monoxide	1.45 to 1.90	350 mμ	8 μ	5
Sodium chloride	1.54	180 mμ	> 15 μ	6
Zirconium dioxide	2.10	300 mμ	> 7 μ	2
Zinc sulfide	2.30	400 mμ	14 μ	7
Titanium dioxide	2.40 to 2.90	400 mμ	> 7 μ	8
Cerium dioxide	2.30	400 mμ	5 μ	2, 3
Silicon	3.50	900 mμ	8 μ	
Germanium	3.80 to 4.20	1400 mμ	> 20 μ	9
Lead telluride	5.10	3900 mμ	> 20 μ	

NOTES:

1. Both materials are sodium-aluminum fluroide compounds, but differ in the ratio of Na to Al and have different crystal structure. Chiolite is preferable in the infrared, because it has less stress than cryolite (see section 20.2.3.2.4).
2. These materials are hard and durable, especially when evaporated onto a hot substrate.
3. The long wavelength is limited by the fact that when the optical thickness of the film is a quarter-wave at 5 μ, the film cracks due to the mechanical stress (see 20.2.3.2.4).
4. There are other fluorides and oxides of rare earths which have refractive indices in this range from 1.60 to 2.0. See reference 22a.
5. The refractive index of SiO_x (called silicon monoxide) can vary from 1.45 to 1.90 depending upon the partial pressure of oxygen during the evaporation. Films with a refractive index of 1.75 and higher absorb at wavelengths below 500 mμ.
6. Sodium chloride is used in interference filters out to a wavelength of 20 μ. It has very little stress.
7. The refractive index of zinc sulfide is dispersive. [23]
8. The refractive index of TiO_2 rises sharply in the blue spectral region. [24]
9. The higher refractive index of 4.2 is given in reference 23. The lower index is quoted by Dr. A. F. Turner of Rochester, N.Y. (private communication).
10. The range of transparency is for a film of quarter-wave optical thickness at this wavelength. These values are approximate and also depend quite markedly upon the conditions in the vacuum during the evaporation of the film.

Table 20.2 - Commonly used thin film materials.

20.2.3 Substrates for multilayer filters.

20.2.3.1 Optical characteristics. If transmission-type filters are used, the substrate must not absorb in the spectral transmission range of the filter. Very often the fact that a substrate material absorbs in certain wavelength regions can be used to advantage. For example, suppose a filter is required for the infrared which has a high transmission at wavelengths longer than 7.0 μ, and a high attenuation at shorter wavelengths. If a germanium substrate were used, then the germanium would absorb at wavelengths shorter than 1.8 μ. Thus the multilayer which is deposited on the germanium would be required to attenuate between 1.8 μ and 7.0 μ. On the other hand, if a substrate of silicon were used, which absorbs at wavelengths below 1 μ, then the multilayer coatings would have to attenuate over a wider range of wavelengths and hence the multilayer coatings would be more complex and expensive to fabricate. The infrared transmission of some optical materials which can be used as multilayer substrates is shown in Figure 20.11. Ballard, McCarthy,and Wolfe [19] have published transmission data on other materials. The refractive index of the substrate is also important. A high-index substrate has high reflection losses at its surfaces if it is not properly coated with antireflection coatings.

20.2.3.2 Chemical and physical properties.

20.2.3.2.1 Certain substrate materials, such as sapphire and fused quartz are hard, durable, and relatively inert to chemical attack. Other materials, notably the rare-earth glasses and some of the alkali halide crystals, are rather soft and delicate. Multilayer coatings are deposited on a substrate and then sometimes rejected because their optical transmission does not meet specifications. If the substrate is expensive, it is desirable to remove the coating from the substrate and recoat it. It is advantageous to use a substrate which is chemically inert, because in this case the coating can be removed with acid or alkali solutions. Otherwise, it is necessary to remove the coating mechanically by the more expensive method of optical polishing.

20.2.3.2.2 The adhesion of the coatings to the substrate is also important. For example, sapphire is a hard and durable substrate, but it has the disadvantage that some thin film materials do not adhere well to its surface.

20.2.3.2.3 The fragility of the substrate is also a consideration. Some thin film materials, such as magnesium fluoride and cerium dioxide, are much harder and more durable when they are evaporated onto a substrate which is heated to a relatively high temperature, often as high as 300° C. It is not an easy task to heat a large piece of optical glass to this temperature without fracturing it, whereas a fused quartz substrate could easily withstand this heating.

20.2.3.2.4 The number of available substrate materials is legion and an exhaustive list would be quite long. Some of the commonly used materials are listed in Table 20.1.

20.2.4 Thin film materials.

20.2.4.1 Optical properties. In most cases, the thickness and refractive index of the films in a multilayer filter are chosen from theoretical considerations. In order to translate this design into a practical filter, it is necessary to select for each layer a thin film material which can be evaporated to the desired thickness and which has a refractive index which is close to the theoretical value. We see from Table 20.2 that in the visible spectral region, non-absorbing films are available with a refractive index in the range from 1.35 to 2.70. In the infrared, materials such as silicon, germanium, and lead telluride are available which have a considerably higher refractive index. In most cases the film does not have the same refractive index as the bulk material. A comprehensive list of thin film coating materials is compiled by Heavens. [22]

20.2.4.2 Physical properties.

20.2.4.2.1 In some applications it is possible to protect a multilayer by cementing on top of it another transparent plate. Thus most silver-dielectric-silver FP type filters (described in 20.10.3.1) are protected in this manner by "sandwiching" the multilayer between two glass plates. If a multilayer is protected, then it is possible to use materials in the stack which are "soft", such as antimony trioxide. In other cases, the multilayer must be "hard" to resist scratching and abrasion because it is exposed and subjected to extreme environmental conditions. Such a coating is called a "hard coating" or simply a "hard coat". A commonly used means of testing the durability of a multilayer is to see if it can withstand rubbing with a soft rubber eraser. Certain coating materials, notably magnesium fluoride, silicon monoxide, cerium dioxide, titanium dioxide, germanium, and silicon, are extremely hard and durable. This does not imply that these materials are the only "hard" coatings. In fact, we must resist the temptation to classify every coating material as either "hard" or "soft" and remember that there is a continuous variation of the durability between the extremely hard materials mentioned in the foregoing sentence and a fragile material such as antimony trioxide. The durability of the coating also depends markedly on the conditions in the vacuum during the evaporation. For

example, techniques have been developed to evaporate zinc sulfide so that it forms a quite durable layer. The durability of a multilayer filter depends not only on the materials, but also on the technical competence of its manufacturer.

20.2.4.2.2 Resistance to moisture. Some coating materials, such as magnesium fluoride and silicon monoxide, can be immersed in water and even in hot saline solutions for some length of time, with no deleterious effects. Other materials, such as cryolite or chiolite, have quite desirable optical properties, but are not widely used as antireflection coatings because they are slightly water soluble. Even though some coating materials are relatively insoluble, moisture can still destroy a multilayer by destroying the bond between the films and the substrate.

20.2.4.2.3 Adhesion to the substrate. It is important that the multilayer coating adhere strongly to the substrate. This is especially true of coatings which have a mechanical stress. Over the years a vast amount of lore has been acquired by the manufacturers of thin film coatings as to what film materials adhere well to various substrates. The adhesion of a film to a substrate depends quite markedly such parameters as the cleanliness of the substrate, its temperature during the evaporation, and the partial pressure of residual gases in the vacuum chamber during the evaporation. There is no substitute for experience in acquiring the "know-how" of producing durable and adherent coatings, although Holland[18] is a source of much useful information.

20.2.4.2.4 Mechanical stress.[20] Early workers found that after a multilayer filter had been evaporated and was moved from the vacuum into the humid air of the laboratory, the entire coating would separate from substrate or craze with many fine cracks. Further investigation showed that this was due to a mechanical stress in the film which is proportional to the thickness of the film.[20] This stress can be demonstrated quite easily by observing that a thin substrate bends as a film is deposited upon it. The substrate bends concave towards the evaporation source if the stress is compressive. It was found that almost all thin film materials have a tensive stress; one exception is zinc sulfide, which has a compressive stress of 0.02 (in arbitrary units). Magnesium fluoride, on the other hand, has a tensive stress of 0.11 units.

20.2.4.2.5 Limit on the thickness[20] It is clear that this stress is one of the principal factors which limits the thickness of films which are deposited by evaporation. When the adhesion of the film to the substrate can no longer balance the stress which builds up as the film grows thicker during an evaporation, the film either parts from the substrate or crazes. Films of either magnesium fluoride or cerium dioxide tend to craze when their physical thickness exceeds one micron. Thus film materials which have a high stress are limited in their application in the infrared, where the films must be quite thick. Another factor which limits the use of some films in the infrared is called the clouding effect. This effect is not observed in a film which is relative thin, say 100 mμ in physical thickness, and hence it is transparent. However, as a much thicker film is evaporated, the film becomes cloudy and scatters light like a ground slab of glass. This is presumably due to the growth of large crystallites in the film which scatter light.

20.2.4.2.6 An effective way of avoiding stress is to use thin film materials which have little or no stress, such as silicon monoxide or sodium chloride. The latter material could only be used in the laboratory, due to its water solubility. Another approach is to deposit a film with compressive stress next to a film with a tensive stress, the thicknesses of each layer being chosen so that the total stress of the pair of films is zero.[20,21]

20.3 ANTIREFLECTION COATINGS

In this section we shall consider the problem of reducing the reflectivity of a dielectric substrate by the addition of one or more non-absorbing thin films. In order to determine how much the transmission of the substrate has been improved by the addition of the multilayer coating, we must first consider the reflectivity of the uncoated substrate.

20.3.1 Reflectivity of an uncoated surface. The reflectivity R for a bare dielectric surface at normal incidence is given by the Fresnel coefficient,

$$R = 1 - T = (n_o - n_s)^2 / (n_o + n_s)^2 \qquad (21)$$

where n_o is the refractive index of the incident medium (which is usually air) and n_s is the refractive index of the substrate. If the light is obliquely incident, then Equation 20-(21) still applies if the appropriate effective index is used in place of n_o and n_s .

20.3.2 Choice of type of coating.

20.3.2.1 Number of layers. In Section 20.3, the spectral reflectivity of antireflection coatings consisting

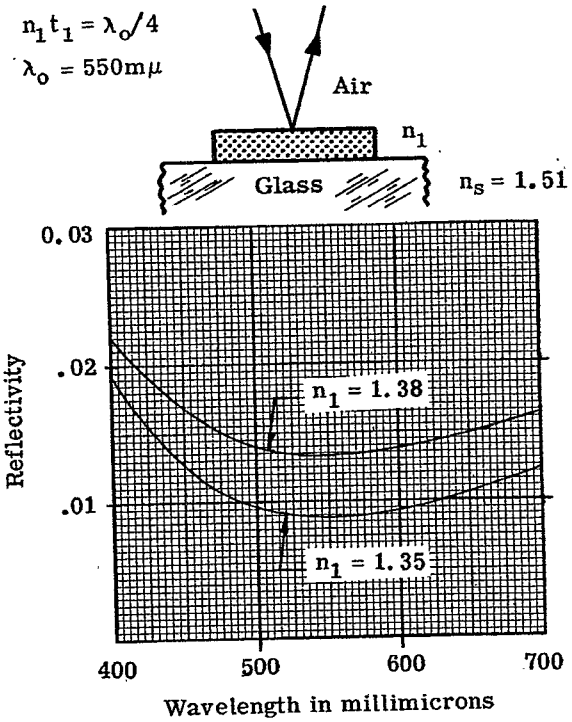


Figure 20.13 - Computed spectral reflectivity of single layers at normal incidence.

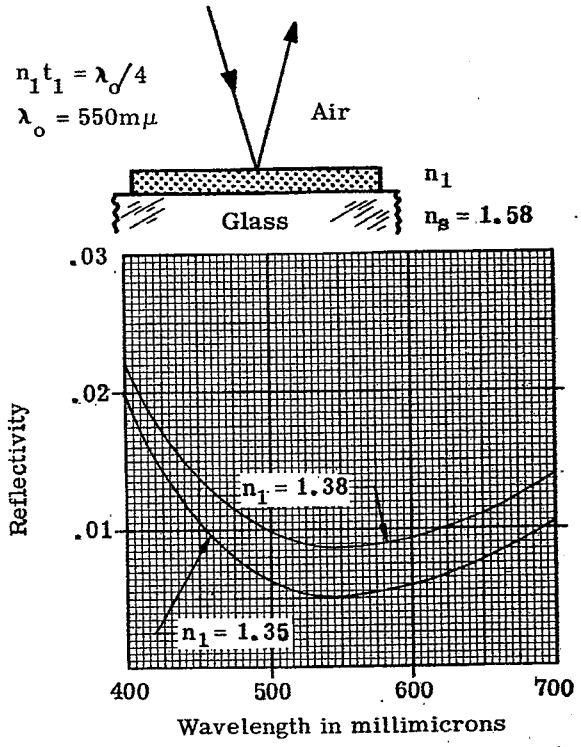


Figure 20.14 - Computed spectral reflectivity of single layers at normal incidence.

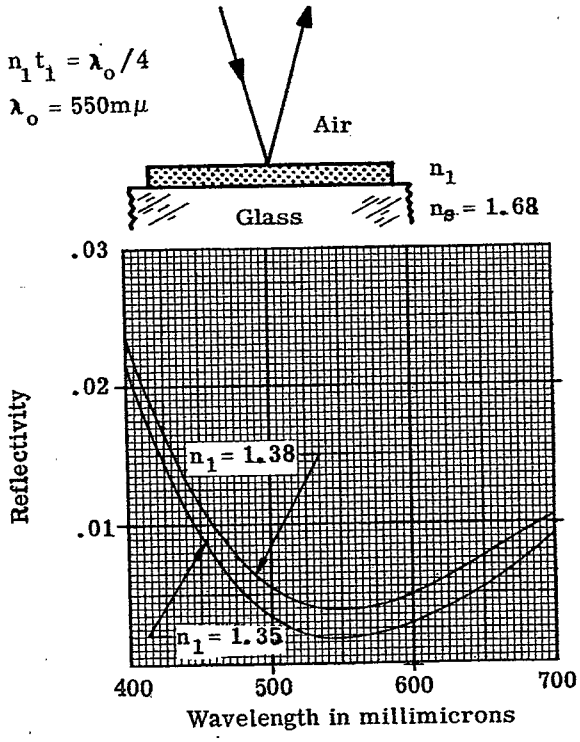


Figure 20.15 - Computed spectral reflectivity of single layers at normal incidence.

of one, two, or three layers is considered. Coatings consisting of more than three layers can be devised, but coatings with less than three layers have proved adequate for most applications; hence coatings with more layers are rarely used. In deciding what type of coating to use, the following points are usually considered:

20.3.2.2 Spectral reflectivity. The type of coating which is selected is often determined by breadth of the spectral region, over which the surface is to have a low reflectivity. For example, suppose that a lens with a large number of elements is used to image a source which emits a considerable amount of radiant energy in the narrow wavelength range from 500 mμ to 550 mμ and a negligible amount of radiant energy at other wavelengths. The best type of antireflection coating for the glass surfaces of this lens would be the two-layer coating which is described in Section 20.3.4.3.2. This coating has a very low reflectivity over a narrow wavelength region and a reflectivity which rises sharply outside of that region. On the other hand, if the source were to emit radiation over a much broader spectral region, say from 400 to 800 mμ, then other types of coatings should be considered, such as a single-layer or three-layer coating, which have a low reflectivity over a much wider wavelength region.

20.3.2.3 Angle of incidence. The reflectivity of all thin film coatings changes with the angle of incidence of the light. The type of antireflection coating which is selected might depend upon the angle of incidence of the light and the amount of convergence in the beam. For example, suppose that a coating is selected for the surfaces of the large negative lens which is the first element in a wide-angle camera lens, as shown in Figure 20.12. The angle of incidence ϕ on the first surface of this lens can easily be as high as sixty or seventy degrees. Thus if excessive vignetting is to be avoided, the antireflection coating on that surface should have a low reflectance at high angles of incidence as well as at normal incidence.

20.3.2.4 Cost. The cost of an antireflection coating is related to complexity of the equipment necessary to evaporate the layers to a prescribed thickness and also to the number of layers in the coating. For example, single-layer coatings of magnesium fluoride are quite easy to produce. The thickness of the magnesium fluoride layer can be easily determined visually during the coating process by examining the color of the reflected light. There are many facilities which have the capability of depositing these coatings because they are quite easy to produce. The production of a three-layer coating requires more elaborate equipment, such as photoelectric monitoring equipment to measure the thickness of the layers and hence the coatings are more expensive.

20.3.3 Single-layer coatings.

20.3.3.1 Basic equations.

20.3.3.1.1 The reflectivity of a dielectric surface coated with a single layer of refractive index n_1 is,

$$R = 1 - T = \frac{a_1 \cos^2 \delta_1 + a_2 \sin^2 \delta_1}{a_3 \cos^2 \delta_1 + a_4 \sin^2 \delta_1} \qquad (23)$$

where

$$a_1 = (n_o - n_s)^2 \qquad a_2 = (n_1 - n_o\, n_s/n_1)^2$$

$$a_3 = (n_o + n_s)^2 \qquad a_4 = (n_1 + n_o\, n_s/n_1)^2$$

where n_o, n_s, and δ_1 have been defined previously in Section 20.1.3.2. The foregoing equation can be derived from Equations (6), (9), and (10) in Section 20.1.5.2. From Equation (23) one can see that when $n_1 t_1 = \lambda/2, 3\lambda/2, 5\lambda/2$, etc., δ_1 is 180°, 360°, 540°, etc. and hence the layer is an absentee layer (defined in 20.1.5.2.2). In this case, the reflectivity is the same as an uncoated surface and Equation (23) reduces to Equation (21).

20.3.3.1.2 When the $n_1 t_1 = \lambda/4, 3\lambda/4, 5\lambda/4$, etc., the reflectivity is either a maximum or a minimum and is given by

$$R_m = \left[\frac{n_1^2 - n_o n_s}{n_1^2 + n_o n_s} \right]^2 . \qquad (24)$$

Some curves of the spectral reflectivity of a single layer on a dielectric substrate are shown in Figures 21.11, 21.12, and 21.13. In Section 20.3, however, we will only consider the case where the reflectivity

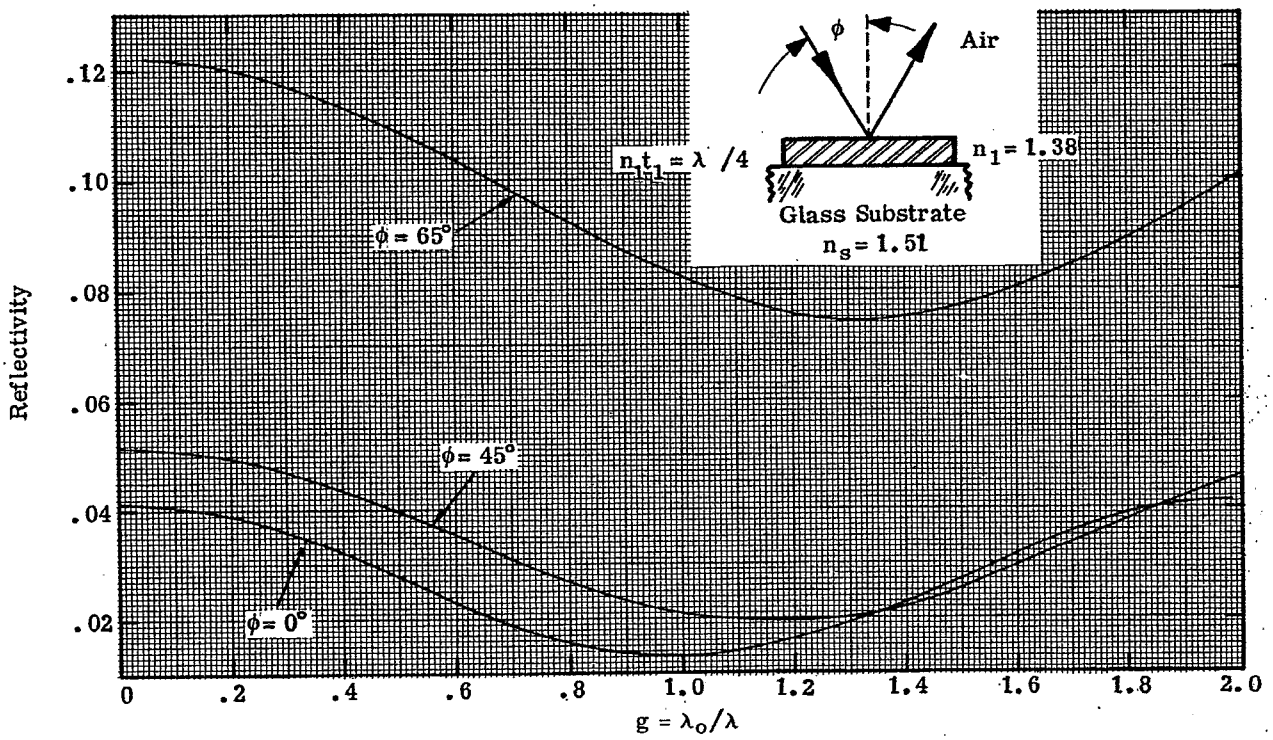


Figure 20.16- Computed spectral reflectivity at various angles of incidence of a single-layer antireflection coating.

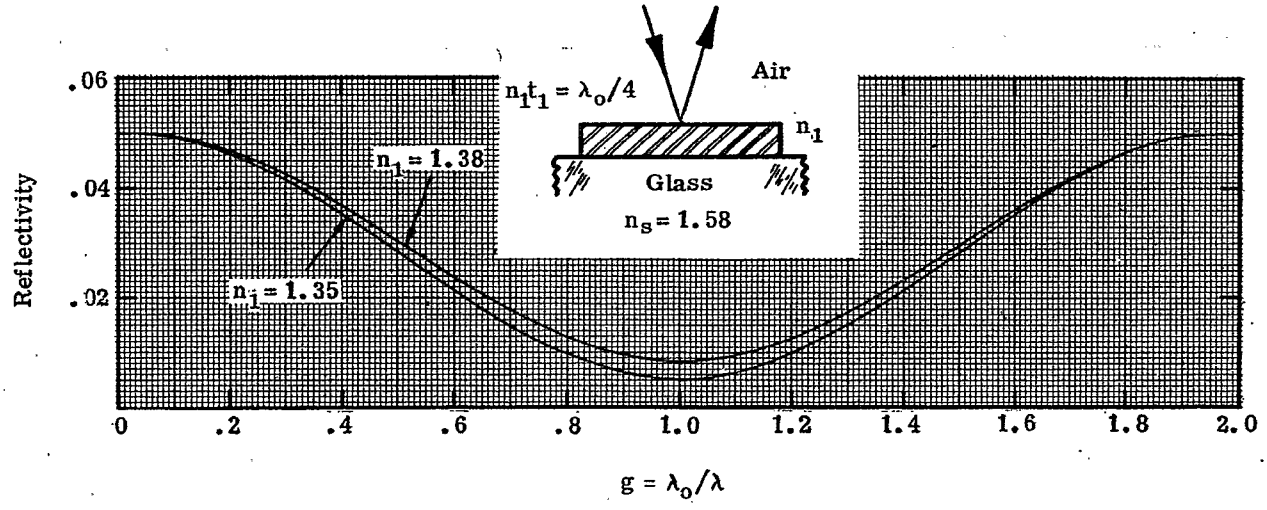


Figure 20.17- Computed spectral reflectivity at normal incidence of a single-layer antireflection coating.

is a minimum, which occurs when

$$n_o > n_1 > n_s \quad \text{or} \quad n_o < n_1 < n_s \ . \tag{25}$$

It is evident from Equation (24) that the reflectivity is zero when

$$n_1 = \sqrt{n_o \, n_s} \ . \tag{26}$$

20.3.3.2 Coatings for substrates with low refractive index.

20.3.3.2.1 We will consider single-layer antireflection coatings for substrates which have a refractive index in the range from 1.5 to 1.7, which includes most of the optical materials which are commonly used in the visible, near-infrared and near-ultraviolet. Crown glass and fused quartz have a refractive index which is close to 1.50, while dense flint glass and sapphire have a refractive index of 1.70.

20.3.3.2.2 In order to obtain a coating which would produce zero reflectivity at a surface between air and glass of refractive index 1.51 at some particular wavelength, say 550 mμ, it is necessary to deposit a film which has an optical thickness of a quarter-wave, three-quarter wave, etc. at 550 mμ and a refractive index given in Equation (26), namely 1.23. No durable coating material has been found which has a refractive index of this low a value. Thus a single-layer coating on such a substrate is a compromise between a coating which is hard and durable and a coating which has an extremely low, if not zero, reflectivity. Two commonly used low-index films are magnesium fluoride (index 1.38) and cryolite (index 1.35). The cryolite film produces a lower reflectivity than the film of magnesium fluoride, because of the lower refractive index of the former. This advantage, however, is offset by the superior physical properties of the magnesium fluoride film. Cryolite films are soft and slightly water soluble, whereas magnesium fluoride films are quite hard and durable, especially when evaporated onto a hot substrate.

20.3.3.2.3 Figures 20.13, 20.14, and 20.15 show the computed spectral reflectivity of some antireflection coatings of both magnesium fluoride and cryolite on substrates of refractive index 1.51, 1.58, and 1.68. The reflectivity of a single uncoated surface at normal incidence in each of these cases is 4.12%, 5.05%, and 6.44%, respectively. The optical thickness of the film is $\lambda_o/4$ at λ_o = 550 mμ so that the minimum reflectivity is in the green region of the spectrum where the eye is most sensitive. Also, 550 mμ is in the center of the visible spectrum (i.e. 400 mμ to 700 mμ) on a wavelength scale. However, 550 mμ is not in the center of the visible spectrum on a wave number (frequency) scale and hence the reflectivity in Figures 20.13, 20.14, and 20.15 is higher at 400 mμ than at 700 mμ. The curves are not symmetrical, about λ_o because they are plotted on a wavelength rather than on a wave number (frequency) scale on the abscissa. The reflectivity of the cryolite coating is lower than the magnesium fluoride throughout the visible spectral region. The minimum reflectivity, R_m, decreases as the index of the substrate progresses from 1.51 to 1.68, because the condition in Equation (26) becomes closer to being satisfied.

20.3.3.2.4 The spectral reflectivity curves shown in Figures 20.13, 20.14, and 20.15 are useful because magnesium fluoride films are used so extensively to coat lenses in the visible spectral region. However, the reflectivity is shown only for a limited spectral region and for a film which has a quarter-wave optical thickness at a particular wavelength, namely 550 mμ. Suppose that a lens is designed to transmit radiant energy not only in the visible, but also in the near infrared to 950 mμ. If a designer wanted to know the reflectivity of a coated surface at 950 mμ, the data in Figures 20.13, 20.14, and 20.15 are of little use to him. Of course, he could compute the reflectivity at this wavelength from Equation (23), but it is much easier to obtain information from a graphical presentation. Suppose that a lens images radiant energy upon a detector which has a maximum response at a wavelength of 700 mμ. This means that the antireflection coatings should have a minimum reflectivity at 700 mμ and hence reflectivity versus wavelength plot depicted in Figures 20.13, 20.14, and 20.15 could not be directly used.

20.3.3.2.5 In order to circumvent the difficulties mentioned in the foregoing paragraph, it is useful to plot the reflectivity as a function of dimensionless parameter, $g = \lambda_o/\lambda$, which is proportional to frequency. Here λ_o is the wavelength at which optical thickness of the film is a quarter-wave. Figures 20.16, 20.17, and 20.18 depict such a plot, for a single layer of magnesium fluoride deposited on substrates of refractive index 1.51, 1.58, and 1.68, respectively. The R_{av} at non-normal incidence is shown for substrates of index 1.51 and 1.68. More extensive data for cryolite is not shown because it is not widely used as an antireflection coating because of the reasons mentioned in 20.3.3.2.2. When plotted versus "g", the reflectivity has even symmetry about $g = 1.0$ (at normal incidence).

20.3.3.2.6 Several examples are given on how these curves are used to compute the reflectivity at a wavelength λ of a single-layer coating which has $n_1 t_1 = \lambda_o/4$.

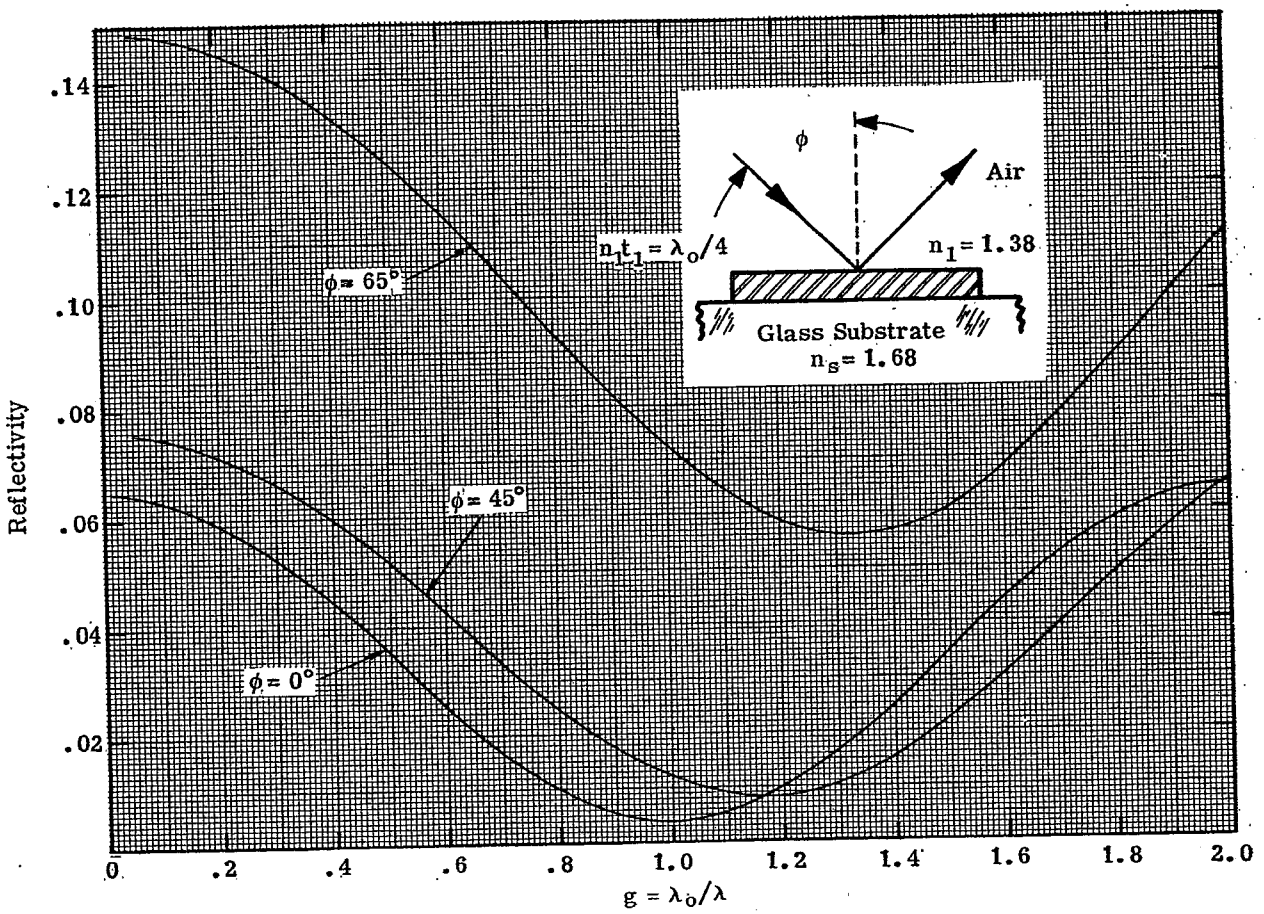


Figure 20.18 - Computed spectral reflectivity at various angles of incidence of a single-layer antireflection coating.

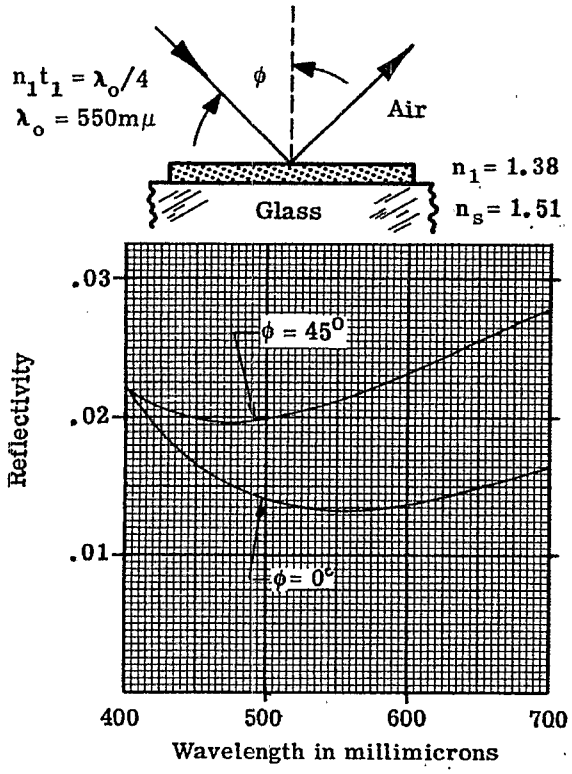


Figure 20.19 - Computed spectral reflectivity of a single layer at $\phi = 0$ and $\phi = 45^\circ$.

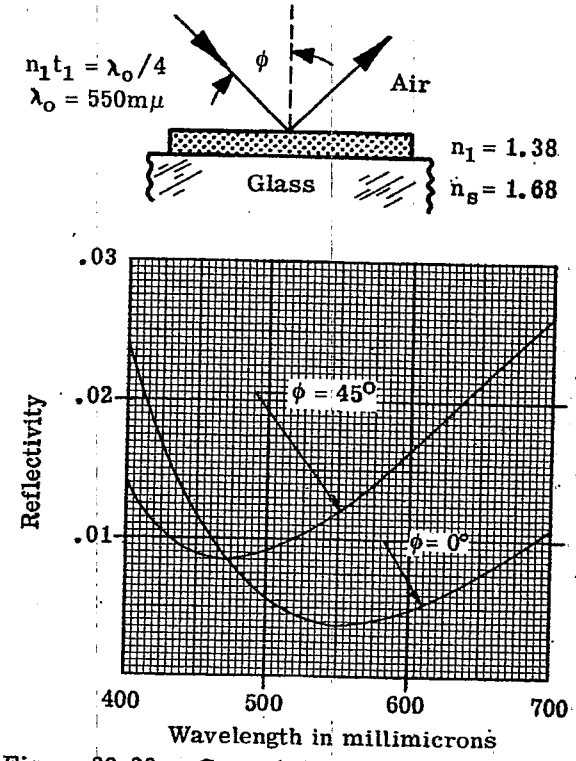


Figure 20.20 - Computed spectral reflectivity of a single layer at $\phi = 0$ and $\phi = 45^\circ$.

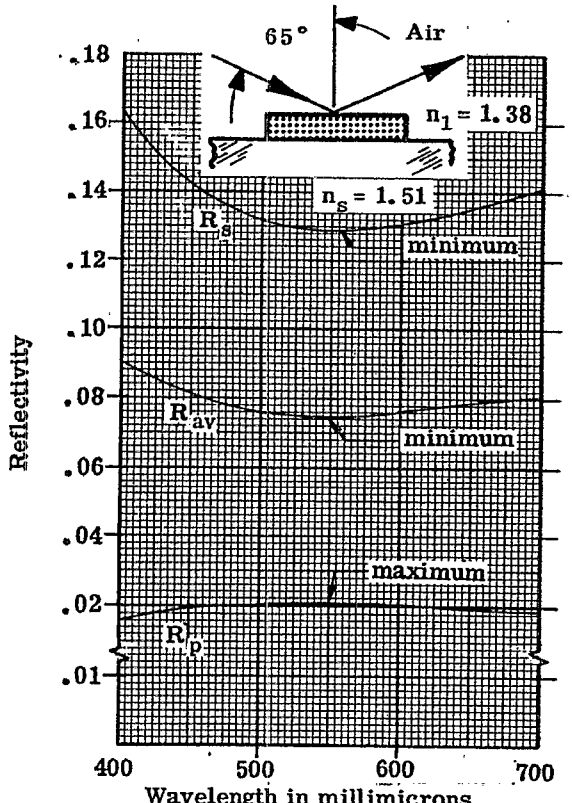


Figure 20.21 - Computed R_s, R_p, and $R_{av} = 1/2\ (R_p + R_s)$ at $\phi = 65^\circ$. The scale of the ordinate changes at 0.02.

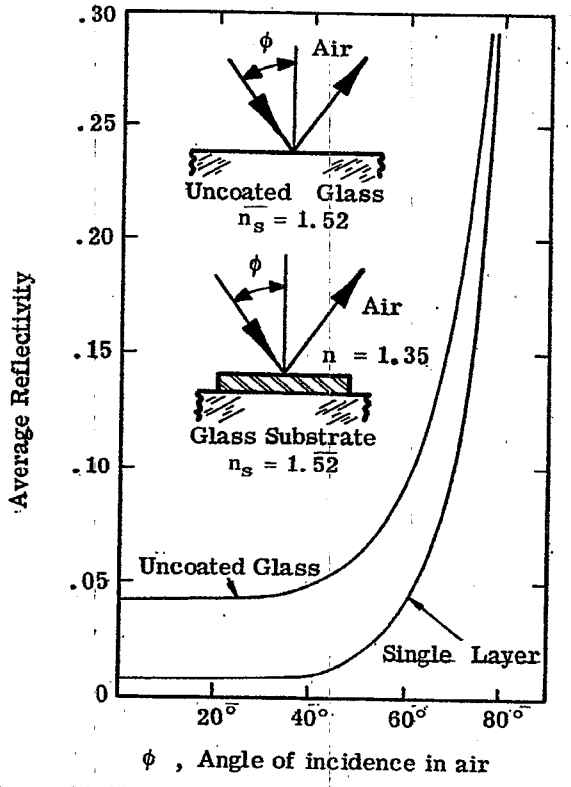


Figure 20.22 - Angle of incidence ϕ versus R_{av} of uncoated glass and the minimum R_{av} of the same glass coated with a single layer.

Example (a) A glass substrate of index 1.51 is coated with a single-layer magnesium fluoride antireflection coating with a minimum reflectivity at 700 mμ. What is the reflectivity at 950 mμ?

Solution: A quarter-wave film with λ_o = 700 mμ, has a minimum R at λ_o. Then at λ = 950 mμ, $g = \lambda_o/\lambda$ = 700/950 = 0.737. From Figure 20.16, at g = .737 R = .017.

Example (b) For the same coating and substrate, what is the reflectivity at 400 mμ?

Solution: Here $g = \lambda_o/\lambda$ = 700/400 = 1.75, and from Figure 20.16, R = .038.

20.3.3.2.7 The reflectivity versus "g" curves are periodic and at normal incidence the curve repeats at g = 2.0, 4.0, 6.0, etc. This fact can be used to find the reflectivity at values of "g" outside the range of the graphs. Example: For the thin film coating described in Paragraph 20.3.3.2.6, what is the reflectivity at 300 mμ? Here $g = \lambda_o/\lambda$ = 700/300 = 2.33. Due to the periodicity, the reflectivity at g = 2.33 is the same as at g = 0.33 and from Figure 20.16 the reflectivity is 0.035.

20.3.3.3 Coatings for substrates with low refractive index, at non-normal incidence.

20.3.3.3.1 As is discussed in Section 20.1.6, Equations (23), (24), and (26) can be used at non-normal incidence, provided that an effective thickness is substituted for the optical thickness and the effective index appropriate to each plane of polarization is used. Regardless the angle of incidence, the reflectivity curve still has either a maxium or a minimum when the phase of retardation of the layer δ_1 = 90°, 270°, etc. Because of the reasons cited in 20.1.6.4.2, the minimum in reflectivity curve shifts towards shorter wavelengths (i.e. the blue). This shift of the minimum reflectivity to shorter wavelengths is illustrated in Figures 20.19 and 20.20, which depict the spectral reflectivity of a layer of magnesium fluoride, ($n_1 t_1 = \lambda_o/4$, λ_o = 550 mμ) on glass substrates of index 1.51 and 1.68 respectively. At non-normal incidence, R_{av} is shown. The minimum reflectivity shifts from 550 mμ at ϕ = 0 to λ_2 = 465 mμ at ϕ = 45°. This shift can also be determined from the graph of the effective thickness versus ϕ in Figure 20.8. From this graph we find that $(n\ t)_{effective} = n_1 t_1 \cos\theta_1 = 0.859\ n_1 t_1$ for a film index of 1.38. Given λ_o of 550 mμ, we determine that $\lambda_2 = (.859)(550) = .472$ mμ.

20.3.3.3.2 The effective indices at ϕ = 65° of the substrate, film, and incident medium of a single-layer MgF_2 coating on glass, are shown in Table 20.3. Its spectral reflectivity in each plane of polarization and the R_{av} is shown in Figure 20.21. The optical thickness of the film is made thicker than $\lambda_o/4$ at normal incidence so that at ϕ = 65° the minimum reflectivity occurs at λ_o = 550 mμ. It is interesting to note that R_p attains a maximum rather than a minimum at λ_o. The reason for this is seen in Table 20.3. The effective indices in the "p" plane of polarization do not satisfy the condition for a minimum stated in Equation (25). However, the R_s curve drops to a sharp minimum at λ_o and hence the R_{av} has a minimum, rather than a maximum, at λ_o. Figures 20.16 and 20.18 show the average reflectivity at ϕ = 45° and ϕ = 65° of a film of refractive index 1.38 deposited on glass of index 1.51 and 1.68, respectively. The optical thickness is $\lambda_o/4$ at ϕ = 0. These curves have even symmetry about their minimum. For example, the curve in Figure 20.16 at ϕ = 65° has a minimum at g = 1.32 and hence the reflectivity is the same at g = 2.12 and g = 0.52.

20.3.3.3.3 Graphs of the average reflectivity, such as are shown in Figures 20.16, 20.18, 20.19 and 20.20, are useful in the case where the incident light is unpolarized and where the detector is not sensitive to polarization, as for example a photographic plate. If the incident light is polarized, or if polarization has been introduced by other elements of an optical system, such as prisms and beam splitters, then it is necessary to compute the reflectivity in each plane of polarization separately, as in Figure 20.27.

20.3.3.3.4 Single-layer antireflection coatings (which satisfy Equation (25)) on glass always decrease the average reflectivity to lower values than the uncoated surface. This is illustrated in Figure 20.22 which shows a plot of the angle of incidence versus R_{av} of an uncoated glass surface, and the minimum reflectivity of the same substrate covered with a single layer of refractive index 1.35. At any angle of incidence, whether it be 20° or 80°, the coated surface has a lower R_{av} than the bare substrate. This subject of antireflection coatings at non-normal incidence is treated in more detail in references 25 and 26.

20.3.3.4 Coatings for substrates with a higher refractive index.

20.3.3.4.1 Single-layer antireflection coatings for substrates with a higher refractive index are considered, which includes materials which are used principally in the infrared, such as arsenic trisulfide glass (index 2.40), silicon (index 3.48), and germanium (index 4.045). The refractive index of all of these materials changes with wavelength. The foregoing values are representative of some mean values in the infrared and even though the calculations do not account for the dispersion in the refractive index, they give some idea of what can be accomplished in the way of antireflection coatings for these materials.

LAYER	INDEX AT NORMAL INCIDENCE	EFFECTIVE INDEX AT $\phi = 65^\circ$	
		POLARIZATION P	POLARIZATION S
Incident medium (Air)	1.00	2.366	0.423
Magnesium fluoride	1.38	1.83	1.041
Glass substrate	1.51	1.89	1.208

Table 20.3 - The effective indices at $\phi = 65^\circ$ incidence of a single-layer antireflection coating.

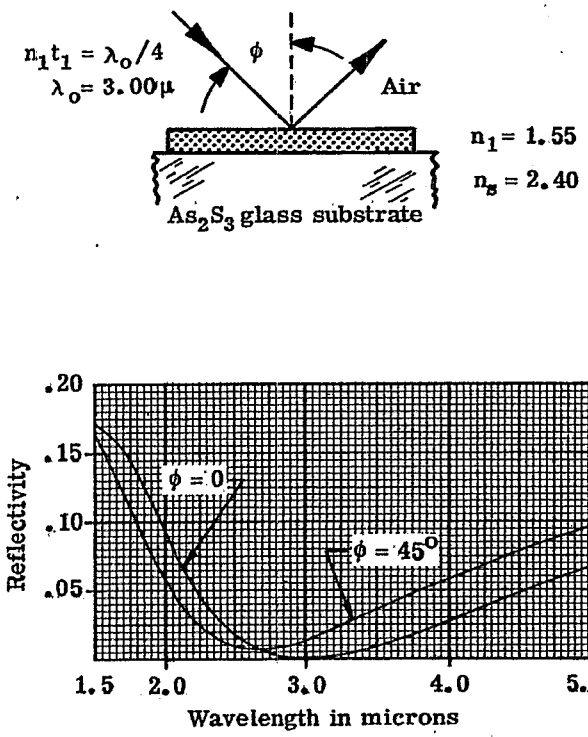


Figure 20.23 - Computed spectral reflectivity of a single-layer antireflection coating at $\phi = 0$ and $\phi = 45^\circ$.

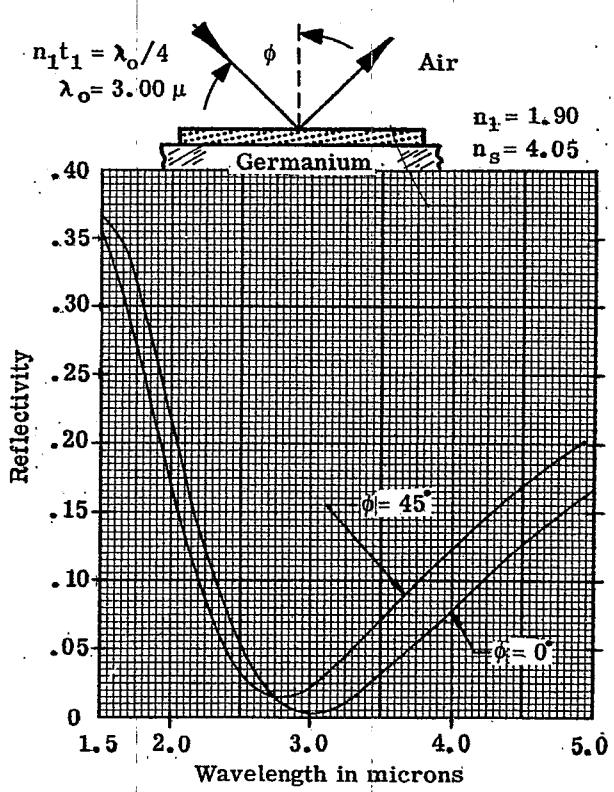


Figure 20.24 - Computed spectral reflectivity of a single-layer antireflection coating at $\phi = 0$ and $\phi = 45^\circ$.

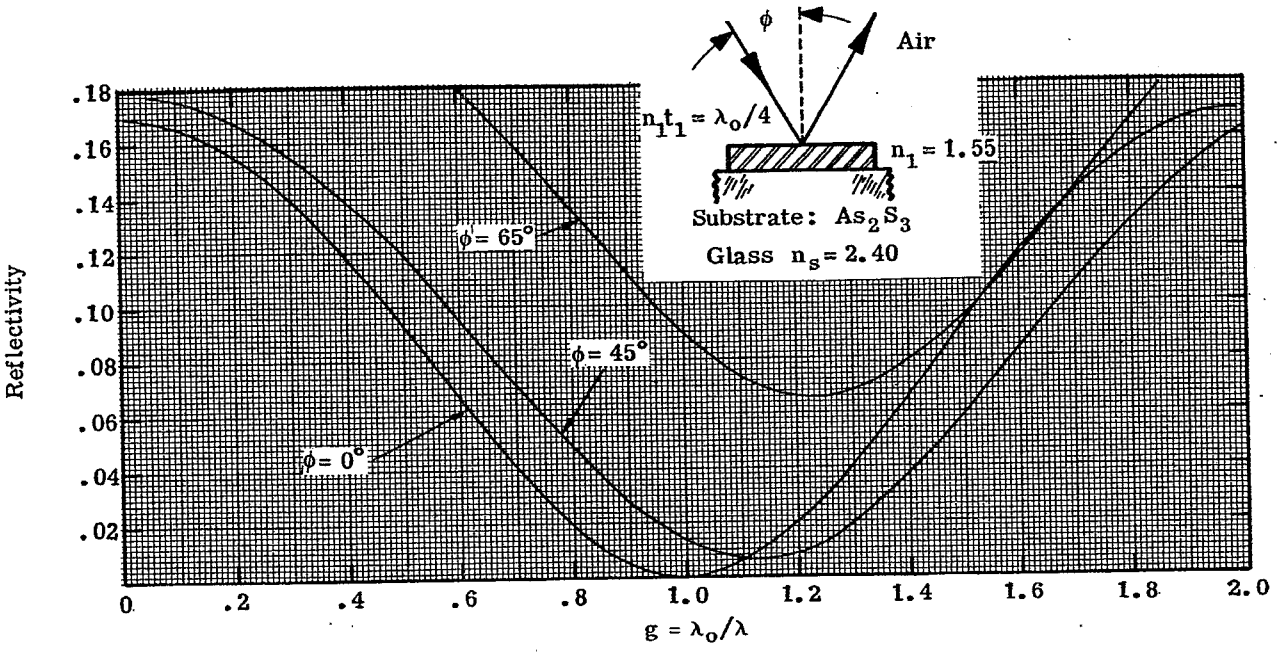


Figure 20.25 – Computed spectral reflectivity of a single-layer antireflection coating at ϕ = 0, 45° and 65°.

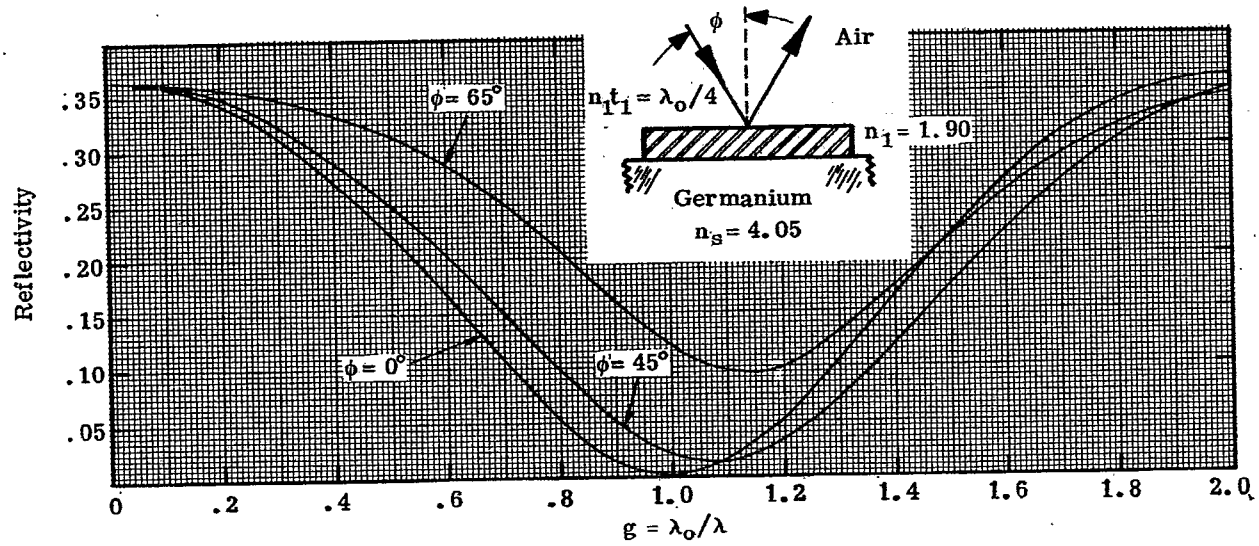


Figure 20.26 – Computed spectral reflectivity of a single-layer antireflection coating at ϕ = 0, 45° and 65°.

20.3.3.4.2 The difference between the coatings for low-index substrates and these higher index materials is that in the latter case thin film materials are available which satisfy Equation (26). Thus the antireflection layer on arsenic trisulfide glass should have a refractive index of $(2.40)^{1/2} = 1.55$, and so on for the other substrates. Figures 20.23 and 20.24 show single layer antireflecting coatings with a quarter-wave optical thickness of 3.0 μ, on substrates of arsenic trisulfide glass and germanium. The reflectivity of the uncoated surface is 0.17 and 0.364 respectively. It should be noted that in practice the reflectivity curve in Figure 20.24 is higher than shown in the region from 1.5 to 1.8 μ, because the absorption constant of the germanium is increasing. Figures 20.25 and 20.26 show a plot of the reflectivity versus "g". In each case the optical thickness of the films is $\lambda_o/4$. These curves can be used to find the reflectivity at various wavelengths of a film which has a quarter-wave optical thickness, say at 4.0 μ, in the manner described in Section 20.3.3.2.6.

20.3.3.5 Coatings for substrates with a higher refractive index, at non-normal incidence.

20.3.3.5.1 The behavior at non-normal incidence of coatings for substrates with a low refractive index was discussed in 20.3.3.3. That discussion applies equally well to these coatings for high-index substrates. The main difference is that the angle shift of the reflection minima to shorter wavelengths is considerably less for the high-index coatings. This is merely a manifestation of the fact that the change in the effective thickness is much less for high-index films than for the low index films, as one can see from Figure 20.8.

20.3.3.5.2 Figure 20.27 shows the spectral reflectivity at 65° in the two planes of polarization and also the R_{av} of a single layer coating on germanium. As in the case of the film shown in Figure 20.21, the effective indices of the incident medium, film, and substrate in the "p" plane of polarization do not satisfy Equation (25) and hence the reflectivity attains a maximum rather than a minimum. However, in the case of the germanium film, R_s is less than R_p. This means that at $\phi = 65°$ the polarizing effect of this coated plate is exactly opposite to that produced by an uncoated dielectric surface, which satisfies the condition that $R_s > R_p$, for all values of $\phi > 0$. This coating on germanium could be used to compensate for the polarization introduced by other uncoated surfaces in an optical system. Figures 20.25 and 20.26 show the average reflectivity of these coatings at various angles of incidence. Hass[27,28] and his co-workers have measured the transmittance in the infrared of some antireflected high-index substrates.

20.3.3.6 Coatings with a higher order of interference. As was pointed out in 20.3.3.1.2, a minimum in the reflectivity of a single-layer coating occurs when the optical thickness of the coating is $\lambda/4$, $3\lambda/4$, $5\lambda/4$, $m\lambda/4$, where the order number "m" is an odd integer. Hitherto we have only shown coatings which show a first-order interference minimum, that is, for $m = 1$. A minimum reflectivity will also occur for higher order interference coatings, such as films which have an optical thickness of three-quarter or five-quarter waves. Figure 20.28 shows the spectral reflectivity of these thicker films deposited on a glass substrate. As one might expect, the reflectivity rises sharply on either side of the minimum at 550 mμ. Hence, there is little advantage to using such higher order interference coatings, with the following exception. In an infrared optical system which is designed to transmit a narrow band of wavelength, some additional attenuation of wavelengths outside of the desired range could be obtained by using higher order interference antireflection coatings.

20.3.3.7 Analogy with electrical transmission lines. In this section we have considered the problem of light impinging upon a substrate of index n_s from an incident medium of index n_o. The problem has been to select the film of proper optical thickness and refractive index so that the reflectivity is reduced to zero. The analogous problem in transmission line theory to match a load of admittance n_s to a transmission line of characteristic admittance n_o so that there will be no standing waves. It is shown in many texts on transmission line and microwave theory that a "quarter-wave transformer" or "quarter-wave matching line" is required to do this. The electrical length of the line should be a quarter-wave and the admittance, n_1, of the line should satisfy Equation (26).

20.3.4 Double-layer coatings.

20.3.4.1 Types of coatings. It is often desirable to use double-layer antireflection coatings because in certain cases these coatings have a lower reflectivity over a wider spectral region than do single-layer coatings. The following types of coatings will be considered:

(1) Double-quarter, single minimum.

(2) Double-quarter, double minimum.

(3) Quarter-half.

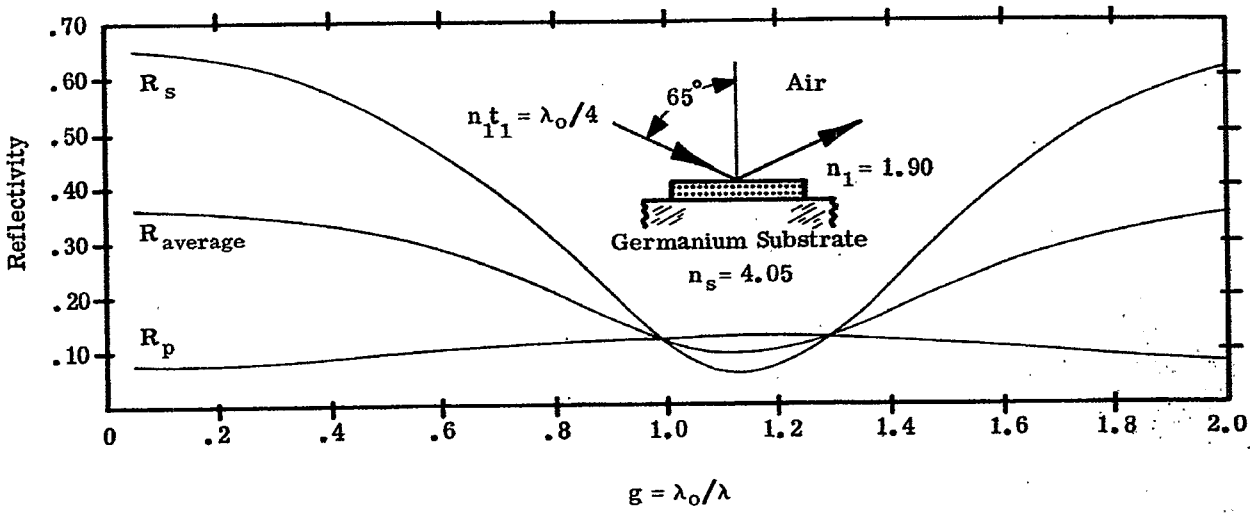


Figure 20.27 Computed R_s, R_p and $R_{av} = 1/2\ (R_s + R_p)$ at $\phi = 65°$.

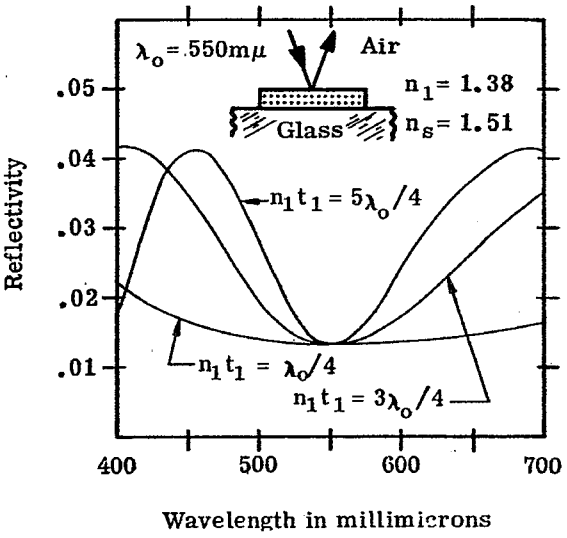


Figure 20.28 Computed spectral reflectivity of single-layer antireflection coatings of various thickness.

Type of Coating	n_1	n_1t_1	n_2	n_2t_2
Single Quarter	1.38	$\lambda_o/4$	—	—
Double Quarter	1.38	$\lambda_o/4$	1.70	$\lambda_o/4$
Quarter-Half	1.38	$\lambda_o/4$	1.80	$\lambda_o/2$

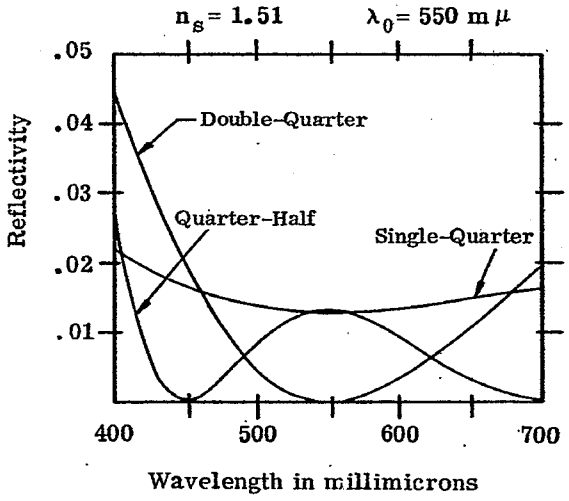


Figure 20.29 Computed spectral reflectivity of single and double-layer antireflection coatings on a glass substrate.

20.3.4.2 Basic equations.

20.3.4.2.1 It is possible to write an equation for the reflectivity of a double-layer coating, but there is little advantage because the equation is lengthy and cumbersome.[29] For our purposes, it is preferable to use the vector addition of amplitude or the characteristic matrix method to compute the reflectivity.

20.3.4.2.2 In the special case where the optical thickness of each film is a quarter-wavelength, that is $n_1 t_1 = n_2 t_2 = \lambda/4$, then $\delta_1 = \delta_2 = 90°$ and the characteristic matrices (Equation (6)) have non-zero elements only along the antidiagonal. The matrix product (Equations (7), (8)) becomes:

$$\begin{bmatrix} A & jB \\ jC & D \end{bmatrix} = \begin{bmatrix} 0 & jn_1^{-1} \\ jn_1 & 0 \end{bmatrix} \begin{bmatrix} 0 & jn_2^{-1} \\ jn_2 & 0 \end{bmatrix} . \qquad (29)$$

After taking the matrix product and substituting into Equations (9) and (10), we obtain for the minimum (or maximum) reflectivity

$$R = 1 - T = \left[\frac{n_o n_2^2 - n_s n_1^2}{n_o n_2^2 + n_s n_1^2} \right]^2 . \qquad (30)$$

Thus for R to be zero, the condition must be satisfied that:

$$\left(\frac{n_2}{n_1}\right) = \left(\frac{n_s}{n_o}\right)^{1/2} . \qquad (31)$$

The condition for zero reflectivity of a double quarter, single minimum type of coating only involves the ratio of the indices of the two films. For example, suppose glass of index 1.51 is coated with a double-quarter-wave coating with the following indices:

$$n_1 = 1.38 \qquad n_2 = 1.69 \qquad (32a)$$

$$n_1 = 1.65 \qquad n_2 = 2.03 \qquad (32b)$$

In both cases Equation (31) is satisfied and the reflectivity is zero at $\lambda = \lambda_o$. However, it can be shown that the reflectivity rises quite sharply on either side of the minimum in the case of Equation (32b), whereas the spectral range over which the reflectivity has a low value is much larger in the case of Equation (32a), and thus Equation (32a) is the better coating to use. Hence it is preferable to use as low index materials as possible.

If the indices satisfy Equation (31), then this is called a double-quarter, single minimum coating. The single minimum means that it has only one reflectivity minimum for a given order of interference. For example, the plot of the reflectivity versus λ_o/λ of the double-quarter coating (curve I in Figure 20.30) has one minimum for the first order interference for $0 \leq g \leq 2.0$, another minimum in the second order for $2.0 \leq g \leq 4.0$, and so on.

20.3.4.2.3 Another type of coating which contains quarter-wave layers has a maximum R at λ_o and two minimum symmetrically spaced about λ_o on a frequency scale. Such a coating is called a double-quarter double minimum coating. A reflectivity curve for a typical coating is shown in Figure 20.35. There are several methods of determining the relationship between the indices of the films of such a coating, such as the solution of complicated algebretic equations [27], [29], [30] or alternatively the use of vector diagrams, as done in Figure 20.5. One type of double-quarter double minimum coating is obtained when the indices satisfy the equations

$$n_1^3 = n_o^2 n_s , \qquad n_2^3 = n_o n_s^2 . \qquad (33)$$

The foregoing equations can be reduced to

$$n_1 n_2 = n_o n_s . \qquad (34)$$

Equations (33) and (34) are derived by Berning [30].

20.3.4.2.4 The spectral reflectivity curve of the quarter-half coating is similar to the curve of the double-quarter, double minimum coating mentioned in the foregoing paragraph, to the extent that it has a maximum at λ_o and two minimum spaced equally about λ_o on a frequency scale. The spectral reflectivity curve of such a coating is depicted in Figure 20.30 (curve III). It is required that the half-wave layer must have a

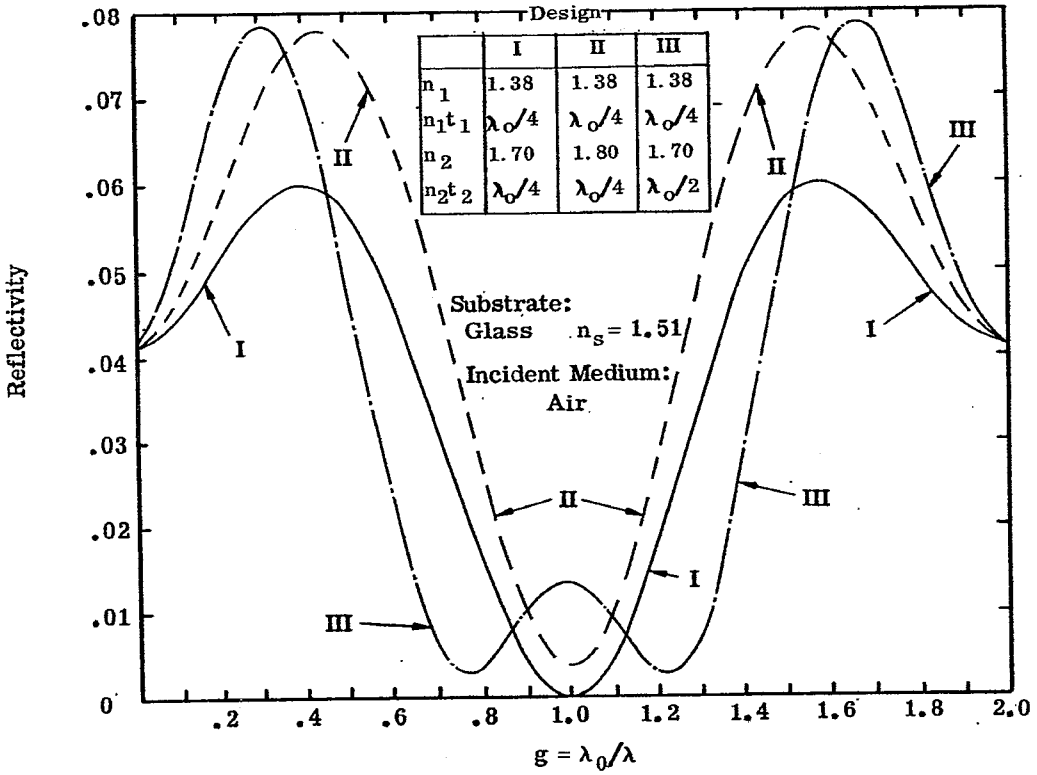


Figure 20.30 – Computed spectral reflectivity of single and double-layer antireflection coatings on a glass substrate.

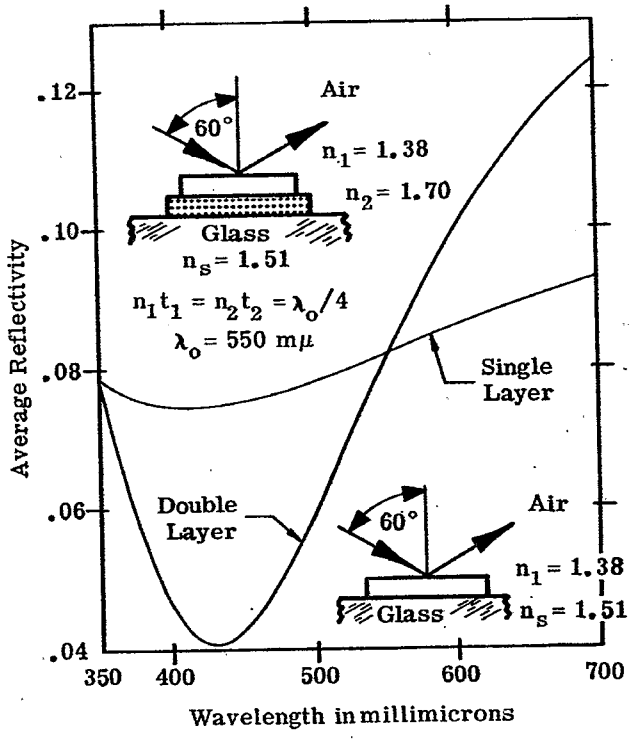


Figure 20.31- Computed average spectral reflectivity of a single-layer and a double-layer coating at $\phi = 60°$.

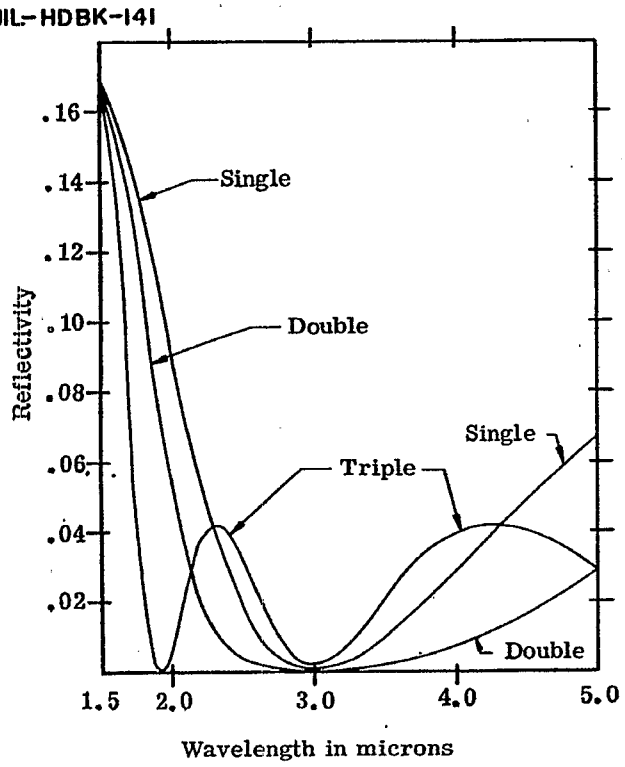


Figure 20.32 — Computed spectral reflectivity of the antireflection coatings whose designs are depicted in Fig. 20.33.

All layers have a quarter-wave optical thickness at $\lambda_0 = 3.0\mu$

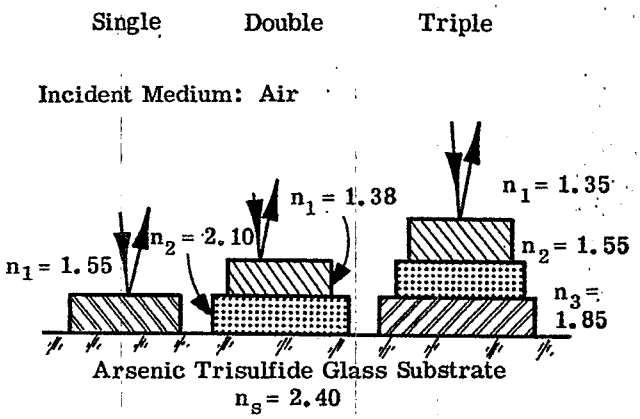


Figure 20.33 — The designs of single, double, and triple-layer antireflection coatings whose reflectivity curves are shown in Figure 20.32.

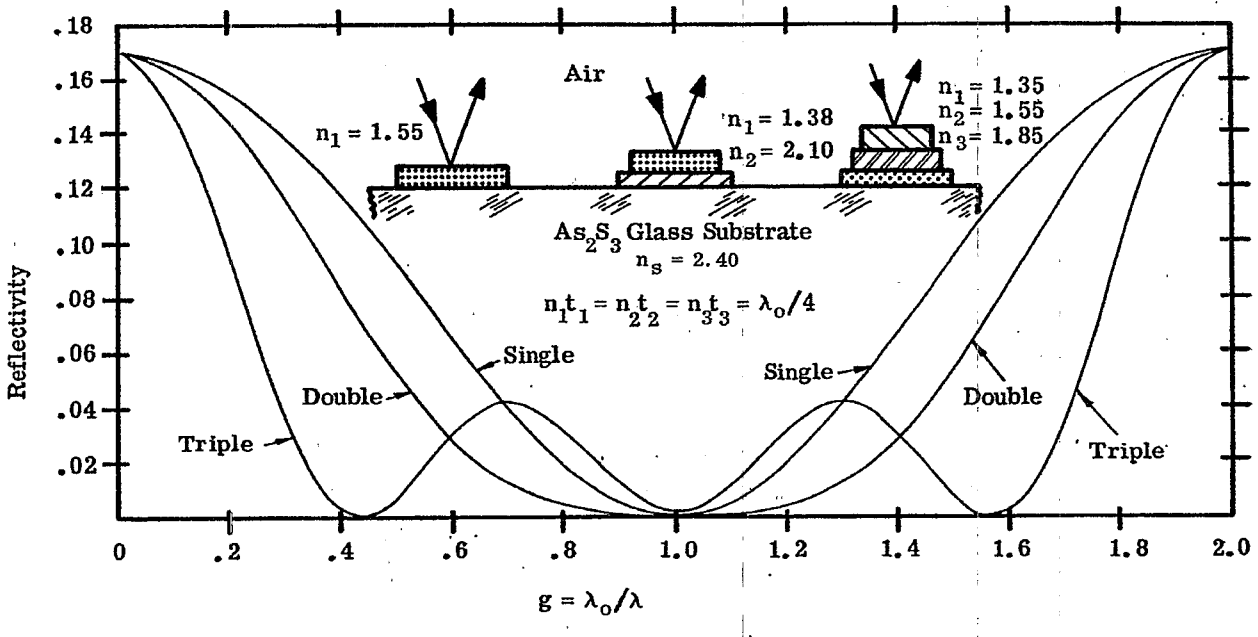


Figure 20.34 — Computed spectral reflectivity of antireflection coatings at normal incidence.

higher refractive index than the substrate. The refractive indices of the two layers can be determined either from a vector diagram (i.e. Figure 20.5) by adjusting the lengths of the vectors so that the vector polygon should close at some value of δ_1 and δ_2. The length of the vectors is proportional to the Fresnel amplitude coefficients and hence one can solve for the indices of the layers. The refractive indices can also be found by solving transendental equation.[29] Whatever indices are chosen, the reflectivity when $\lambda = \lambda_o$ is determined easily from Equation (24) because the half-wave layer is absentee.

20.3.4.3 Double-layer antireflection coatings for low index substrates.

20.3.4.3.1 In this subsection, the term low index substrate is used in the same sense as in 20.3.3.2. The fact that the lowest index film material which is available has an index of 1.35 means that double quarter, double minimum coatings cannot be produced for substrates in this index range.

20.3.4.3.2 The double quarter curve in Figure 20.29 is the spectral reflectivity of a double-quarter, single minimum coating on glass of refractive index 1.51. The optical thickness of the films is a quarter-wave at 550 mμ and because the indices satisfy Equation (31), the reflectivity at this wavelength is zero. Such a coating could be manufactured using magnesium fluoride as the low index film and silicon monoxide as the high index film. Unfortunately, when the index of the silicon monoxide is this large (see Note 4 in Table 20.2), it has a slight absorption in the blue and hence the films are yellowish in appearance. On the same graph is shown for comparison, the reflectivity of a single quarter-wave coating of magnesium fluoride. The reflectivity of the double-quarter coating is below the single layer in the green, but rises considerably above it at 400 mμ. Figure 20.30 shows the reflectivity of the double-quarter coating on a frequency scale. At certain wavelengths, the reflectivity is greater than that of uncoated substrates. Shown also in Figure 20.30 is the reflectivity of a double-quarter coating composed of films of indices 1.38 and 1.80. This shows what happens to the reflectivity when the relationship in Equation (31) are not precisely satisfied. Hass[29] shows many computed reflectivity curves in which variations have been made in both the thickness and refractive indices from the optimum condition specified in Equation (31).

20.3.4.3.3 Figure 20.29 shows the spectral reflectivity of a quarter-half coating. As was stated in 20.3.4.2.4, at λ_o (550 mμ) the half-wave layer is absentee and the reflectivity is the same as the single quarter-wave layer of index 1.38. Since the two minimum are equally spaced about the maximum on a frequency scale, they are unequally spaced on the wavelength scale. Figure 20.30 shows the spectral reflectivity of the same configuration, but on a λ_o/λ scale.

20.3.4.3.4 The limitation on the film index mentioned 20.3.4.3.1 means that the reflectivity of both the double-quarter and the quarter-half antireflection coatings exceeds the reflectivity of the uncoated substrate at some wavelengths, as is shown in Figure 20.30. This is not true of the single-layer low index coating, whose reflectivity never exceeds that of the uncoated substrate. The question as to which type of coating to use depends upon the range of wavelengths over which the reflectivity is to have a low value. If the range is narrow, a double-quarter coating might be preferable. However, if the range is quite large, then even though the single-layer coating does not achieve as low a reflectivity as the double-quarter coating in certain spectral regions, the better overall performance of the single layer over a wide range of wavelengths would make it preferable.

20.3.4.4 Double-layer antireflection coatings for low index substrates at non-normal incidence. At non-normal incidence the reflectivity of coatings which have two or more layers is influenced by the fact that if the optical thicknesses of the layers are matched at normal incidence, they are no longer matched at any other angle. This point is discussed in more detail in 20.1.6.3 and 20.1.6.4. If the layers are matched at some angle of incidence, ϕ, then one can compute the reflectivity at $\lambda = \lambda_o$ by substituting the effective index appropriate to each plane of polarization into Equation (30). In the case of double layer coatings, there are many possible combinations of incident angles and matched or mismatched layers which can be considered and hence a complete analysis of the behavior of the double-layer coatings of non-normal incidence would be quite lengthy. As an illustration of a typical case, Figure 20.31 shows the spectral reflectivity of a double-quarter coating on glass at 60° incidence. Both layers have a quarter-wave optical thickness at normal incidence at 550 mμ. The reflectivity minimum has shifted to the blue and notwithstanding the fact that the optical thickness of the layers is mismatched, the minimum reflectivity is still considerably lower than the minimum of the single-layer coating.

20.3.4.5 Double-layer antireflection coatings for high index substrates.

20.3.4.5.1 In this subsection, the term high index substrate is used in the same sense as in 20.3.3.4.1. For high-index substrates, coating materials are available which not only satisfy Equations (31), (33), or (34), but also the relationship:

$$n_o < n_1 < n_2 < n_s . \qquad (35)$$

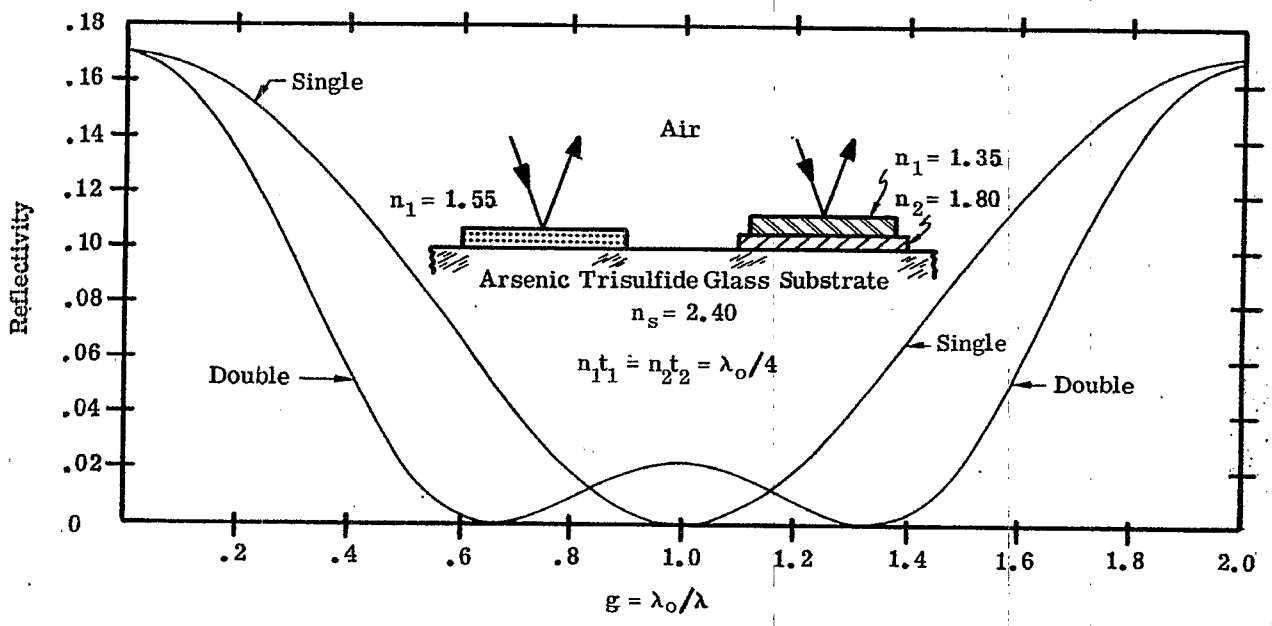


Figure 20.35 – Computed spectral reflectivity of antireflection coatings at normal incidence.

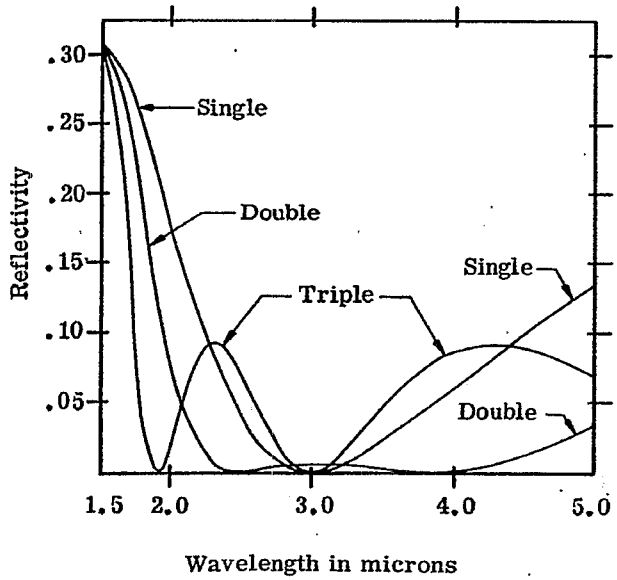


Figure 20.36 – Computed spectral reflectivity of the antireflection coatings whose designs are depicted in Fig. 20.37.

All layers have a quarter-wave optical thickness at $\lambda_0 = 3.0\mu$

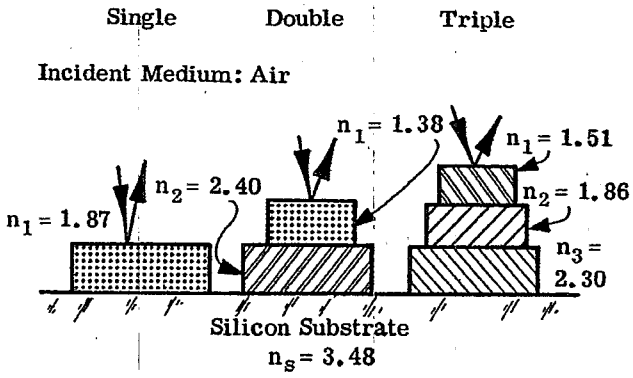


Figure 20.37 – The designs of single, double, and triple-layer antireflection coatings whose reflectivity curves are shown in Figure 20.36.

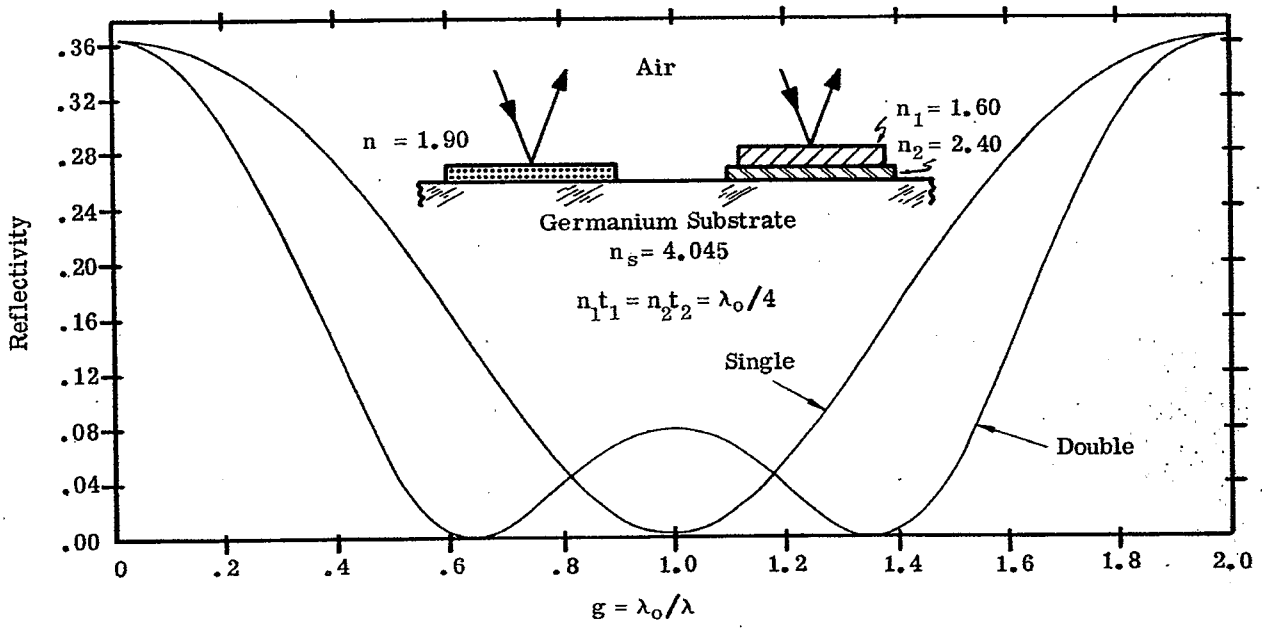


Figure 20.38 – Computed spectral reflectivity of antireflection coatings at normal incidence.

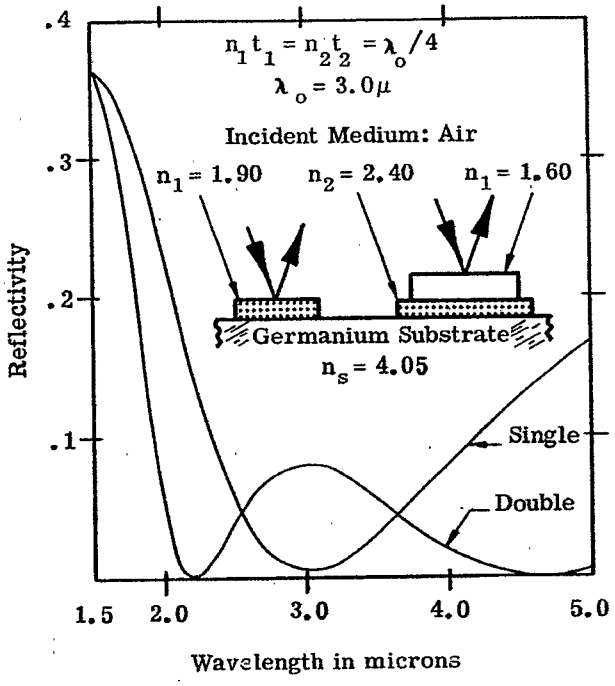


Figure 20.39 – Computed spectral reflectivity of antireflection coatings at normal incidence.

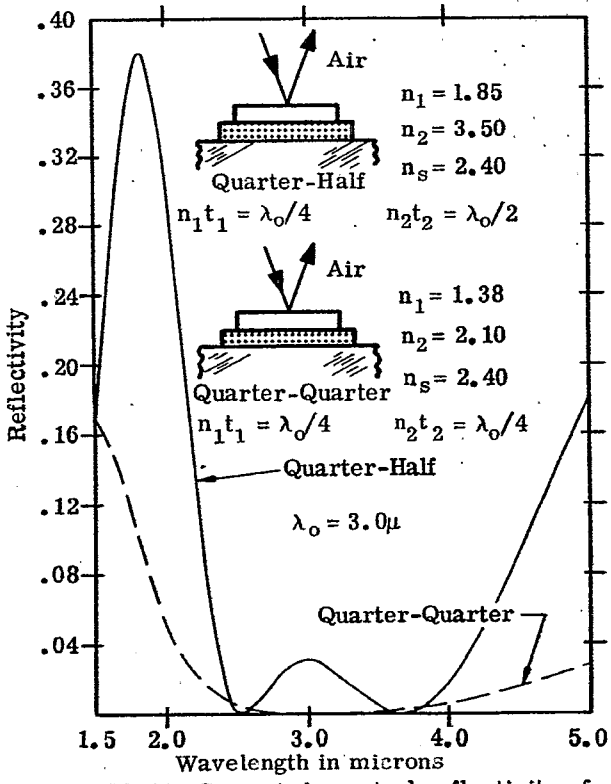


Figure 20.40 – Computed spectral reflectivity of antireflection coatings at normal incidence.

In this case a wide region of low reflectivity is achieved and the reflectivity does not exceed the reflectivity of the uncoated substrate. This fact is illustrated in Figure 20.34, which depicts a double-quarter, single minimum coating on a substrate of arsenic trisulfide glass.

20.3.4.5.2 With higher index substrates it is possible to use film materials which satisfy Equation (33) and hence it is possible to produce double-quarter, double minimum coatings. Figure 20.35 shows the spectral reflectivity of such a coating on arsenic trisulfide glass; the indices of the layers satisfy Equations (33) and (34). Figure 20.36 shows a double-quarter, double minimum coating on a silicon substrate in which the indices of the films satisfy Equation (34) to a fair approximation. Figures 20.38 and 20.39 show the spectral reflectivity of a double-quarter, double minimum coating on a germanium substrate. The indices of the films in this coating satisfy Equations (33) and (35).

20.3.4.5.3 Figure 20.40 shows the spectral reflectivity of a half-quarter coating on a substrate of arsenic trisulfide glass. The layer with the half-wave optical thickness is silicon, which has a higher refractive index than the substrate. The reflectivity rises considerably above that of the uncoated substrate and hence there seems to be little advantage of this type of coating over the double-quarter coatings.

20.3.4.5.4 Figure 20.41 shows the spectral transmittance of a germanium plate with both sides antireflected by a two-layer coating.[28] With the double-quarter single minimum coating, the ratio of the index of the silicon film to the index of the didymium fluoride film closely approximates the ratio specified in Equation (31). Figure 20.42 shows the transmittance of a silicon plate which has been coated in a similar manner.[28] The index of the cerium dioxide film and the magnesium fluoride film of this double-quarter double minimum coating satisfy Equation (34). The double reflectivity minimum of this type of coating are manifested in the transmission maxima at 1.8 and 2.7 μ. The bump in the curve near 3 μ is due to the water adsorbed on the coating. In both cases, the spectral region in which the transmission exceeds 0.9 is wider for these double-layer coatings than for a single layer coating.

20.3.5 Triple-layer antireflection coatings.

20.3.5.1 Types of coatings. The number of types of three-layer antireflection coatings which can be produced is legion, because there are six parameters which can be varied - three thicknesses and the three refractive indices. The discussion is confined to two types of coatings:

(1) The quarter-half-quarter.

(2) The triple-quarter, triple minimum.

20.3.5.2 Basic equations. Triple-layer coatings are usually designed by the use of vector diagrams (Section 20.1.5) or with the more sophisticated methods of Section 21.7.4. A few of the many conditions for zero reflectivity of three-layer coatings are derived by Berning.[30] One of the conditions for a stack of three quarter waves to attain a triple reflectivity minimum is:

$$n_1 n_3 = n_o n_s \qquad (36a)$$

$$n_2^2 = n_o n_s \ . \qquad (36b)$$

20.3.5.3 Triple-layer coatings for substrates with low refractive index. In this subsection, the term low index substrate is used in the same sense as in 20.3.3.2.1. For such substrates, a quarter-half-quarter coating is quite effective in reducing the reflectivity to below .005 throughout the entire visible spectral region. The computed spectral reflectivity of such a coating on a glass substrate of index 1.51, is shown in Figure 20.43. The reflectivity of single-layer, double-quarter, and quarter-half coatings are also shown for purposes of comparison. The reflectivity of the triple-layer coating is well below the reflectivity of the single-layer coating over an octave (i.e. from $g = 0.6$ to $g = 1.2$). However, outside of this region, the reflectivity rises to quite large values. Hass[31] presents a detailed study of this type of coating and shows how the spectral reflectivity is altered as the refractive index and the thickness of each layer is varied. Figures 20.44, 20.45, 20.46, and 20.47 show the measured spectral reflectance of glass covered with this type of triple-layer coating.

20.3.5.4 Triple-layer coatings for substrates with high refractive index. The term high refractive index is used in the same sense as in 20.3.3.4.1. For such substrates, it is possible to use coating materials whose refractive indices satisfy

$$n_o < n_1 < n_2 < n_3 < n_s \ . \qquad (37)$$

Figures 20.32 and 20.34 show the spectral reflectivity of a triple layer coating on a substrate of arsenic trisulfide glass whose refractive indices satisfy Equations (36) and (37). The reflectivity does not exceed

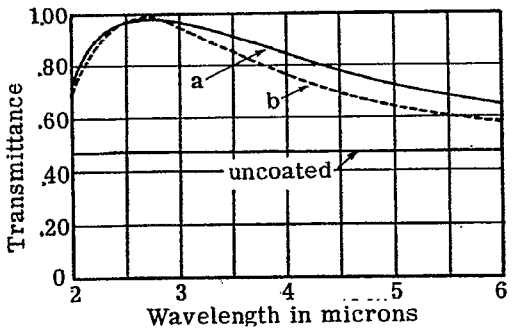


Figure 20.41 – Measured transmittance of Ge plate with antireflection coatings of (a) Si + didymium fluoride, and (b) SiO; (nt = $\lambda/4$ at 2.7 μ). From ref. 28 .

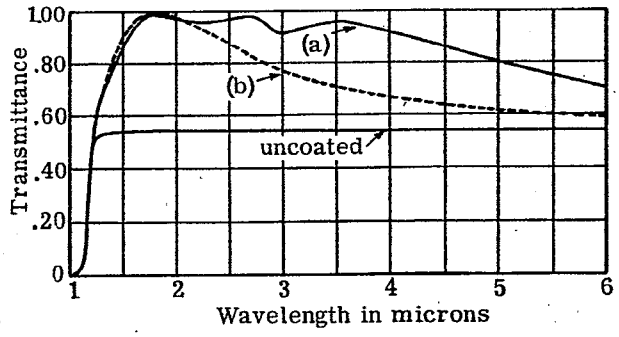


Figure 20.42 – Measured transmittance of Si plate with antireflection coatings of (a) CeO_2 + MgF_2 (nt = $\lambda/4$ at 2.2 μ), and (b) SiO (nt = $\lambda/4$ at 1.8 μ). From ref. 28.

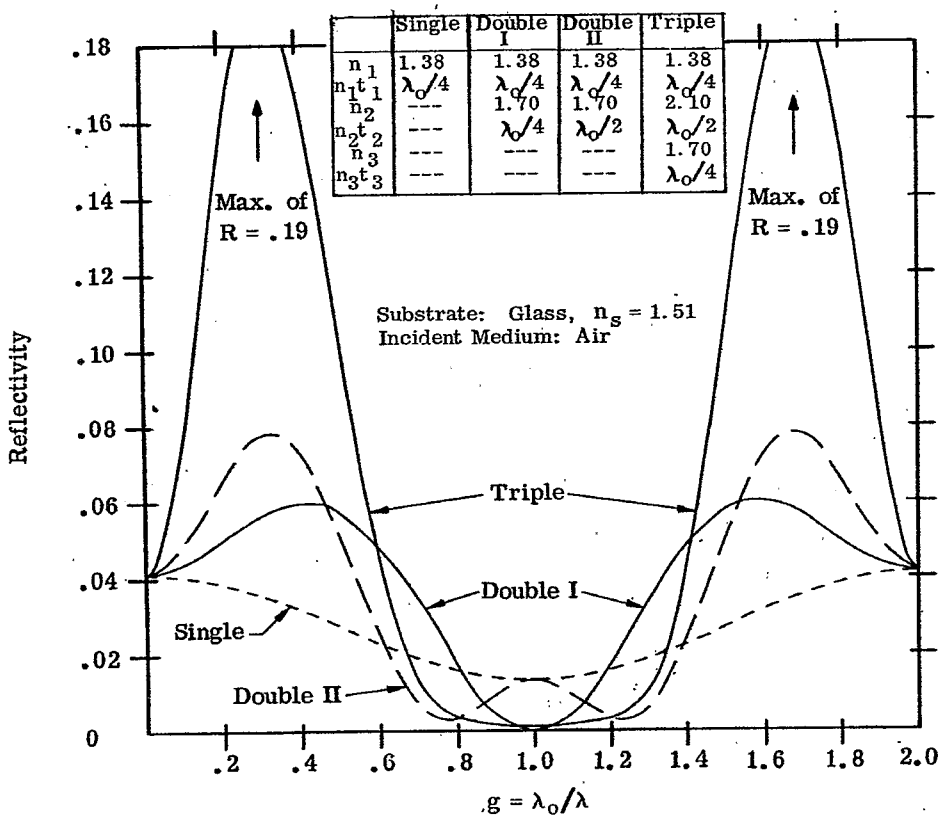


Figure 20.43 – Computed spectral reflectivity of antireflection coatings at normal incidence.

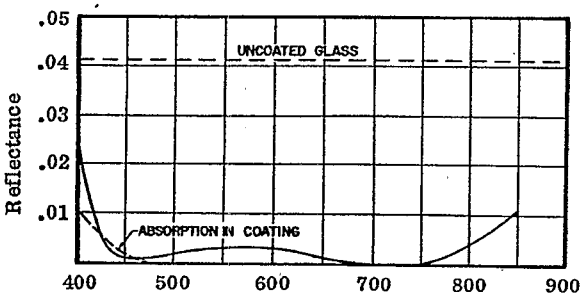


Figure 20.44 - Measured reflectance of a quarter-half-quarter antireflection coating consisting of MgF_2 + ZrO_2 + CeF_3 on glass. From ref. 31.

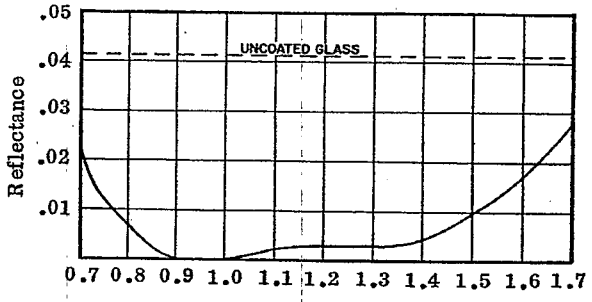


Figure 20.45 - Measured reflectance of a quarter-half-quarter antireflection coating consisting of MgF_2 + Nd_2O_3 + CeF_3 on glass. From ref. 31.

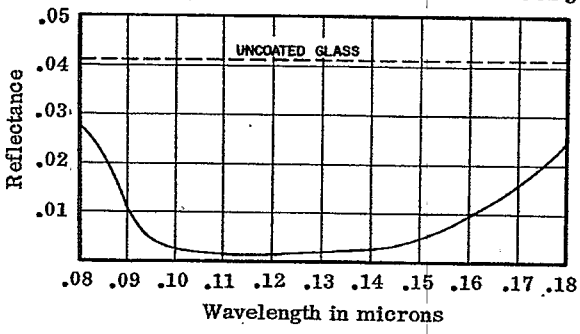


Figure 20.46 - Measured reflectance of a quarter-half-quarter antireflection coating consisting of MgF_2 + SiO + CeF_3 on glass. From ref. 31.

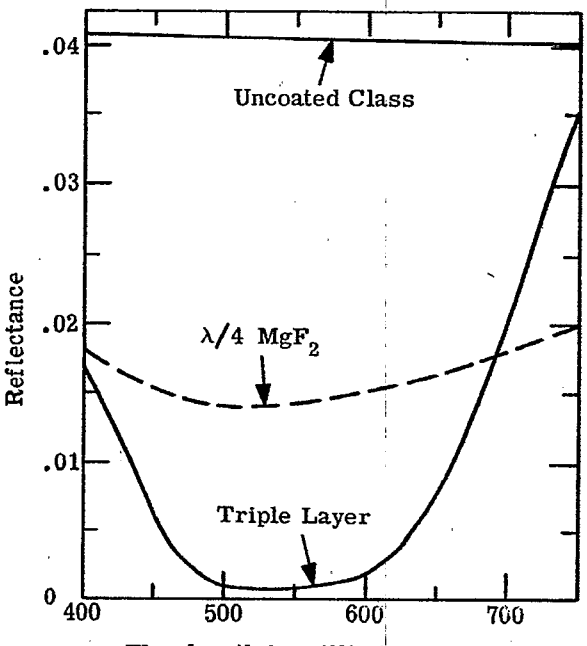


Figure 20.47 - Measured spectral reflectivity of single and triple-layer antireflection coatings. Courtesy of Fish-Schurman, Corporation.

.005 over a frequency range of more than two octaves. Silicon monoxide and chiolite could be used as films in such a coating. Figure 20.36 shows a coating of similar design on a silicon substrate.

20.4 THE REFLECTIVITY OF MULTILAYERS WITH PERIODIC STRUCTURE

20.4.1 The quarter-wave stack.

20.4.1.1 The basic period.

20.4.1.1.1 Before delving into the subject of multilayer mirrors, color filters, beam splitters, etc., it is helpful to understand some of the basic concepts relating to the propagation of light in a multilayer stack with a periodic structure.

20.4.1.1.2 The simplest type of stack with a periodic structure is a quarter-wave stack, which consists of layers with the same optical thickness, but alternating between two refractive indices, n_a and n_b. A diagram of a quarter-wave stack consisting of eight layers is shown in Figure 20.48. At some wavelength λ_o, the optical thickness of each layer is $\lambda_o/4$, that is

$$n_i \; t_i = \lambda_o/4 \qquad (38)$$

for all values of i. The basic period of the quarter wave stack consists of two layers: HL, using the notation of Section 20.1.3.5. The design of the stack depicted in Figure 20.48 can be written:

Glass LHLHLHLH Air .

The foregoing design can also be written:

Glass $(L\,H)^4$ Air

or as

Glass $(L\,H)^m$ Air, where $m = 4$,

to emphasize the fact the basic period, L H, is repeated "m" times.

20.4.1.2 Reflectivity of a typical quarter-wave stack. Figures 20.49 and 20.50 show the computed reflectivity versus $g = \lambda_o/\lambda$ of a quarter-wave stack with the same n_a and n_b but with different numbers of periods. The following observations are made:

(1) The reflectivity in the range of g within the crosshatched area monotonically increases, as the number of basic periods, m, increases from 2 to 5. This region is called the high-reflectance zone.

(2) The reflectivity outside of the high reflectance zone is an oscillating function. By this we mean that for any arbitrary value of g (outside of the high-reflectance zone) the reflectivity can either increase or decrease if two addition layers (i.e. L H) are added to the stack so that m increases to $m + 1$.

(3) In the region between $g = 0$ and the edge of the high-reflectance zone, there are $m - 1$ maximum and $m - 1$ minimum in the reflectivity.

20.4.1.3 Properties of an infinite stack. It is of interest to consider the limiting case where the number of periods, m, becomes infinite. This is called the infinite stack. The statements in 20.4.1.2 can be generalized to include any quarter-wave stack:

(1) The reflectivity in the high-reflectance zone approaches 1.0 as m becomes infinite. As the number of periods becomes large, the reflectivity curve in the high-reflectance zone becomes very flat. Regardless of the number of periods the reflectivity attains a maximum value R_{max} at the center of the zone at $g = 1.0$. The width of the high-reflectance zone depends only on the index ratio, n_a/n_b.

(2) The reflectivity curve in the region outside of the high-reflectance zone oscillates between maximum and minimum values. The number of oscillations between $g = 0$ and the edge of the high-reflectance zone depends upon the relationship between n_o, n_s, n_a, and n_b, but is proportional to m for $m >> 1$.

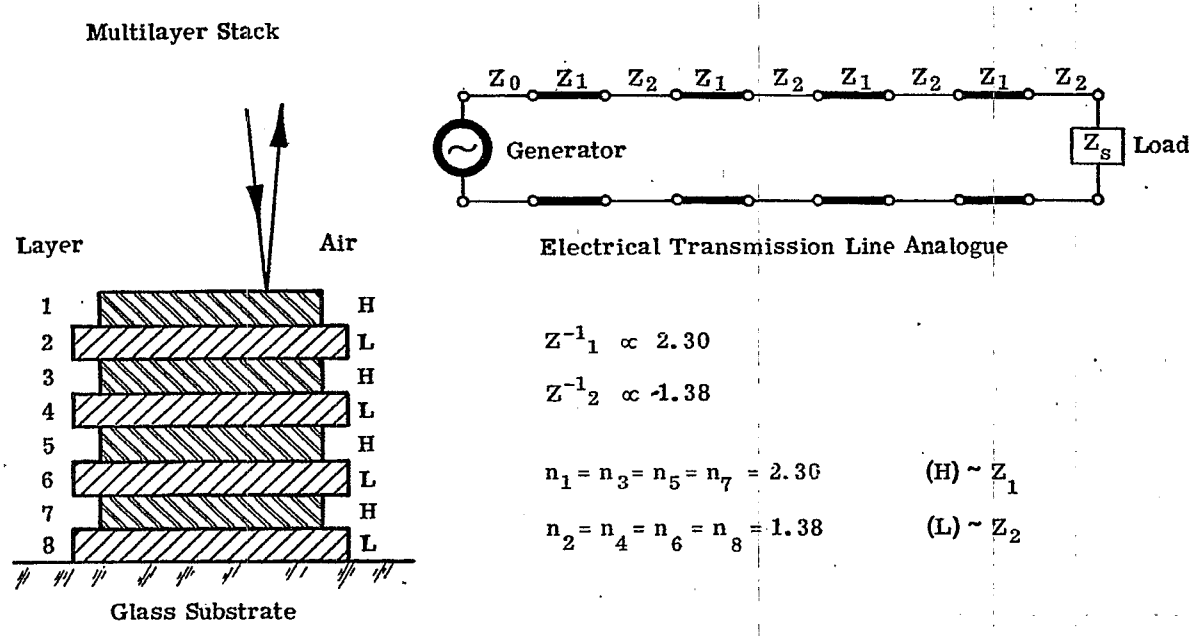


Figure 20.48 - Diagram of a quarter-wave stack and its transmission line analogue.

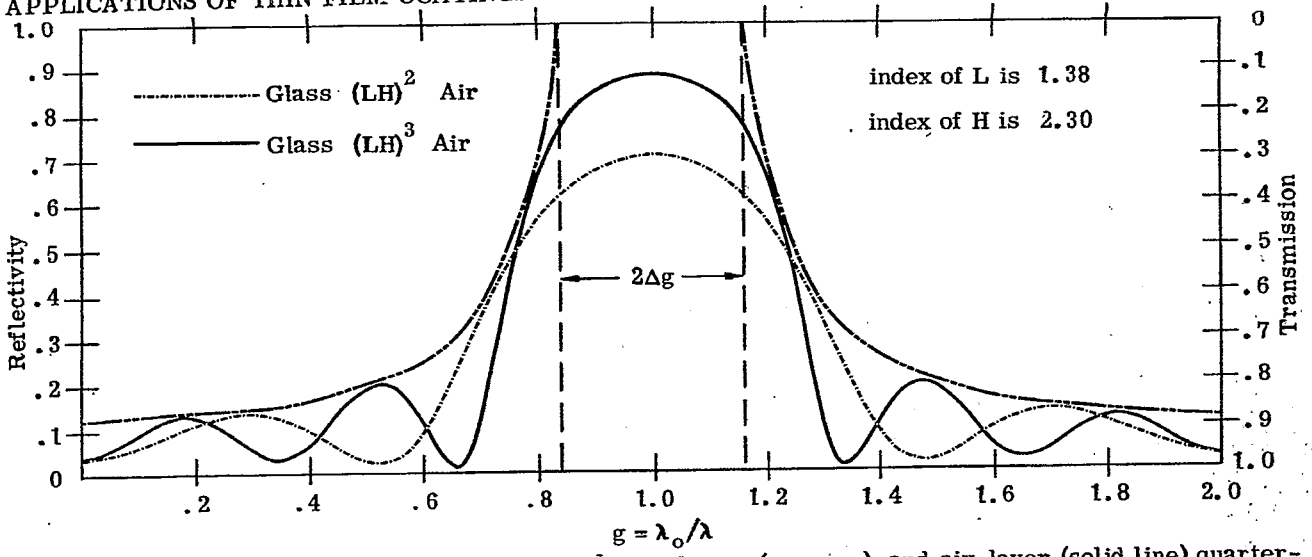


Figure 20.49 - Computed spectral reflectivity of a four-layer (----------) and six-layer (solid line) quarter-wave stack and the envelope of maximum reflectivity (----------).

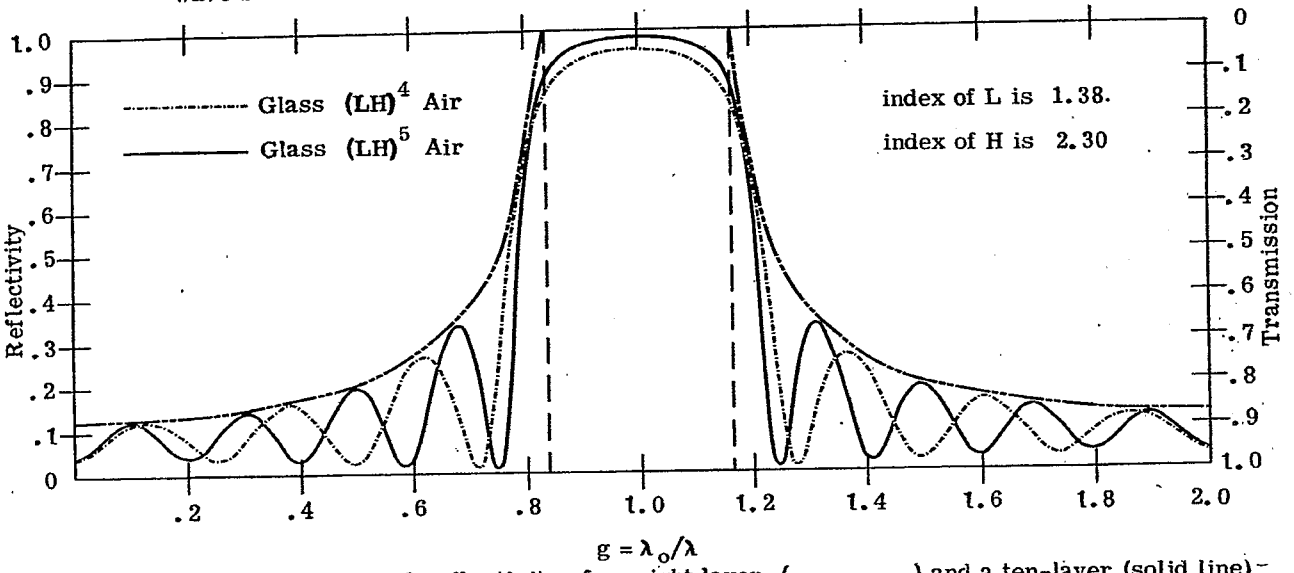


Figure 20.50 - Computed spectral reflectivity of an eight layer (----------) and a ten-layer (solid line) quarter-wave stack, and the envelope of maximum reflectivity (----------).

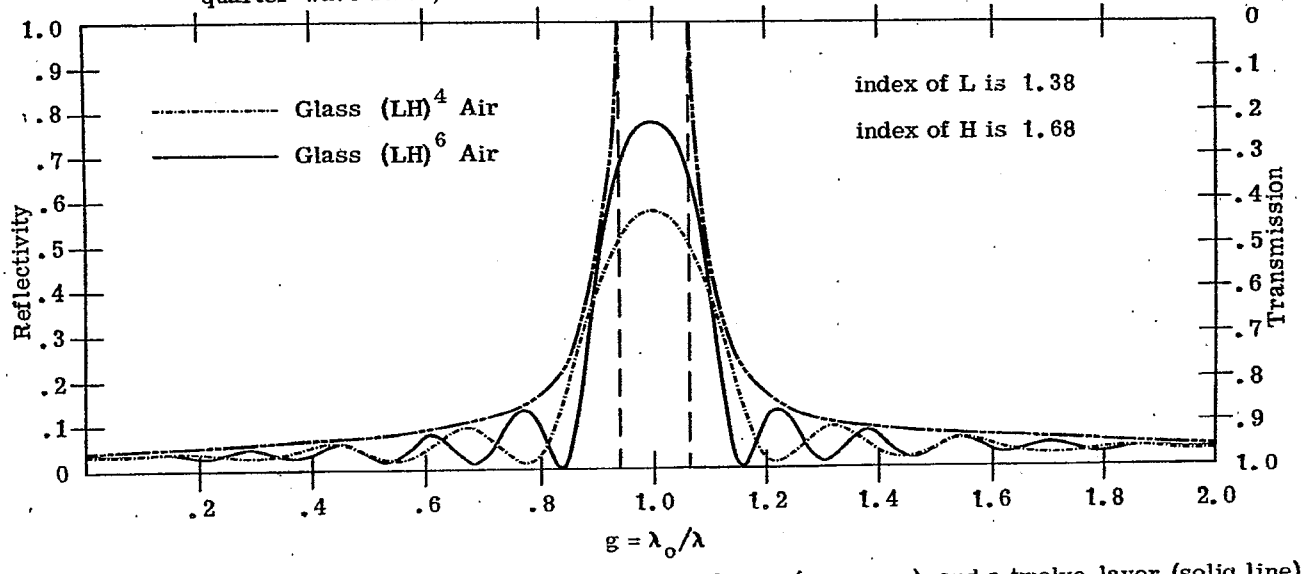


Figure 20.51 - Computed spectral reflectivity of an eight-layer (----------) and a twelve-layer (solid line) quarter-wave stack, and the envelope of maximum reflectivity (----------).

The reflectivity oscillates between two envelope curves. The reflectivity maximum envelope curve is shown as dashed line ———— in Figures 20.49 to 20.52. Both the maximum and minimum envelope curves are shown in Figure 20.53.

(3) The reflectivity of a multilayer with a periodic structure is adequately described by plotting the width of the high-reflectance zone and the maximum and minimum reflectivity envelopes, as is shown in Figure 20.54.

20.4.1.4 Width of the high-reflectance zone. The high-reflectance zone is symmetrical about $g = 1.0$. In the case of a quarter-wave stack, there is a simple expression for the distance Δg from the center of high-reflectance zone at $g = 1.0$ to its edge:

$$\Delta g = \frac{2}{\pi} \arcsin \left| \frac{1 - n_a/n_b}{1 + n_a/n_b} \right| \tag{39}$$

where the $|\ |$ denotes an absolute value; the principal value of the arcsin is to be used. The total width of the high-reflectance zone is $2\Delta g$. For example, the quarter stacks shown in Figures 20.49 and 20.50 have $n_a = 1.38$ and $n_b = 2.30$; from Equation (39) we find that $\Delta g = 0.161$. As another example, consider the quarter-wave stacks shown in Figures 20.51 and 20.52. The index ratio, n_a/n_b of these stacks is approximately the same, although the stack in Figure 20.51 is composed of low-index layers, while the latter stack has high-index layers. As one might expect, the width of the high-reflectance zone of the two stacks is essentially equal. The index ratio n_a/n_b is large for the stack shown in Figure 20.53 and consequently the high-reflectance zone is quite wide. This large index ratio can only be attained with materials such as germanium and chiolite in the infrared.

20.4.1.5 Maximum reflectivity of a quarter-wave stack. Consider a more general type of quarter-wave stack in which the optical thickness of each of the layers is $\lambda_o/4$, but the refractive index, n_i, of each of the layers can be different. From Equations (5), (7), (9) and (10), it follows that the transmission T_{min} and reflectivity R_{max} at $g = 1.0, 3.0, 5.0$, etc. is

$$R_{max} = 1 - T_{min} = (P + P^{-1} - 2)(P + P^{-1} + 2)^{-1} \tag{40}$$

$$T_{min} = 4/(P + P^{-1} + 2) \tag{41}$$

where the variable P is defined as

$$P = \left[\frac{n_\ell}{n_{\ell-1}} \frac{n_{\ell-2}}{n_{\ell-3}} \cdots \frac{n_4}{n_3} \frac{n_2}{n_1} \right]^2 \frac{n_o}{n_s} \tag{42}$$

when the total number of layers ℓ is even and

$$P = \left[\frac{n_\ell}{n_{\ell-1}} \cdot \frac{n_{\ell-2}}{n_{\ell-3}} \cdots \frac{n_3}{n_2} \frac{n_1}{} \right]^2 \frac{1}{n_o\ n_s} \tag{43}$$

when ℓ is odd. From Equation (41) we see that as the number of layers becomes large, P also becomes large and $p >> p^{-1}$, $p >> 2$ and thus

$$T_{min} \sim 4P^{-1} . \tag{44}$$

Equations 40 to 43 are for a quite general type of quarter-wave stack, and can be easily applied to specific cases. For example, P for the stack in Figure 20.48 is:

$$P = \left[\frac{n_b}{n_a} \right]^8 \frac{n_o}{n_s} . \tag{45}$$

In the case where the stack has many layers and hence the index ratio n_a/n_b is raised to a large power, T_{min} is a very sensitive function of this ratio.

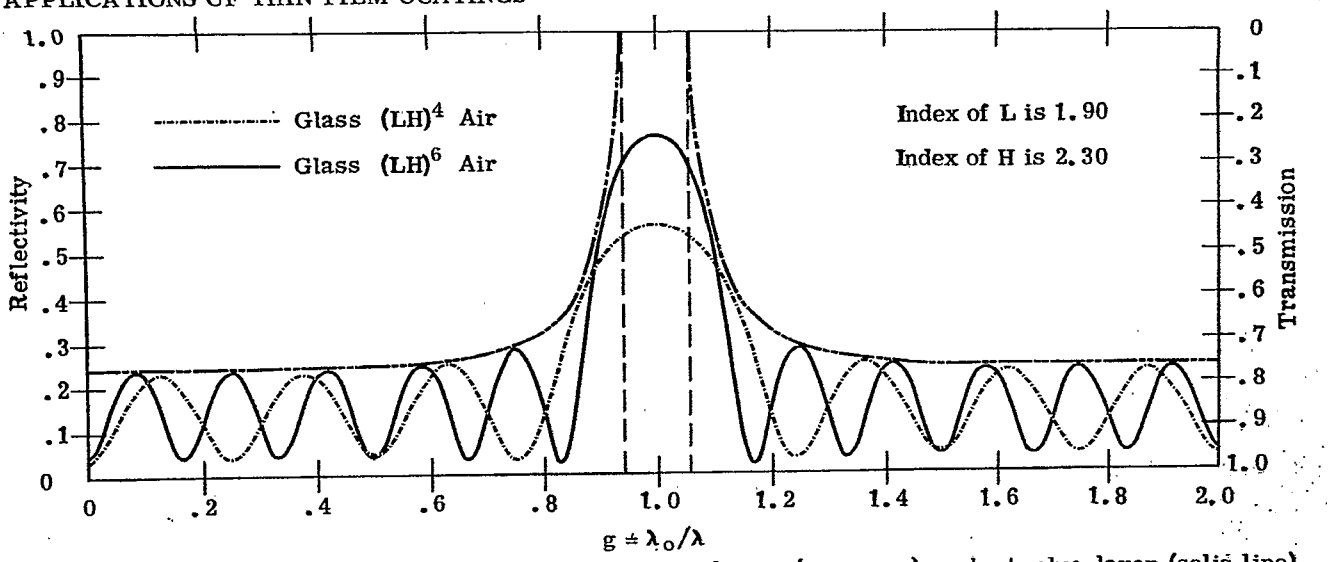


Figure 20.52 - Computed spectral reflectivity of an eight-layer (----------) and a twelve-layer (solid line) quarter-wave stack, and the envelope of maximum reflectivity (----------).

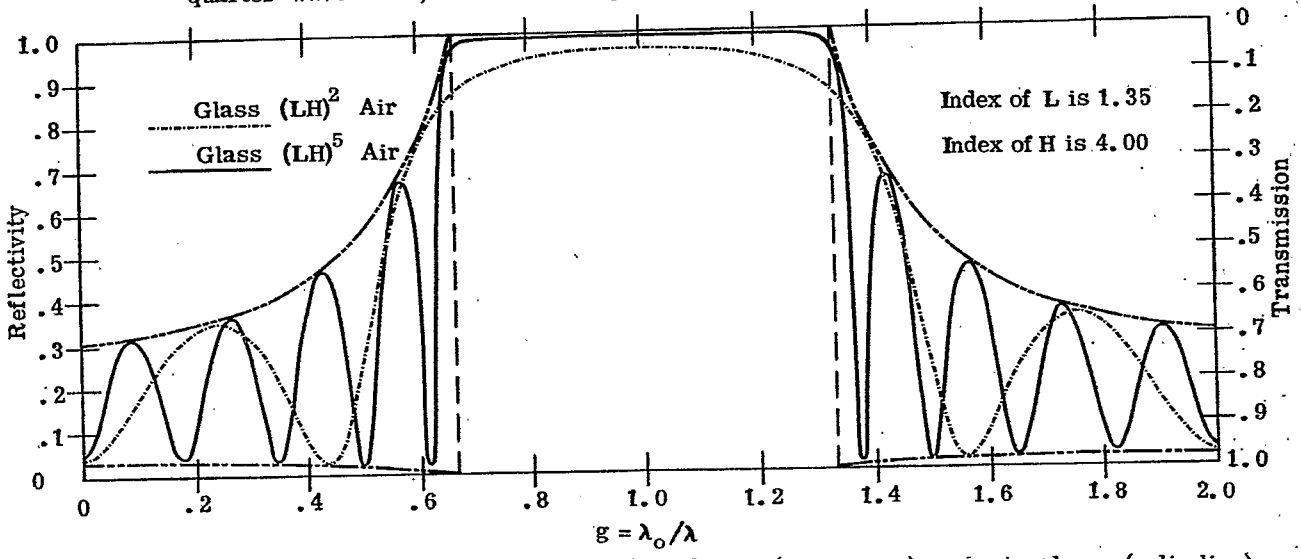


Figure 20.53 - Computed spectral reflectivity of a four-layer (----------) and a ten-layer (solid line) quarter-wave stack, and envelope of maximum reflectivity (----------).

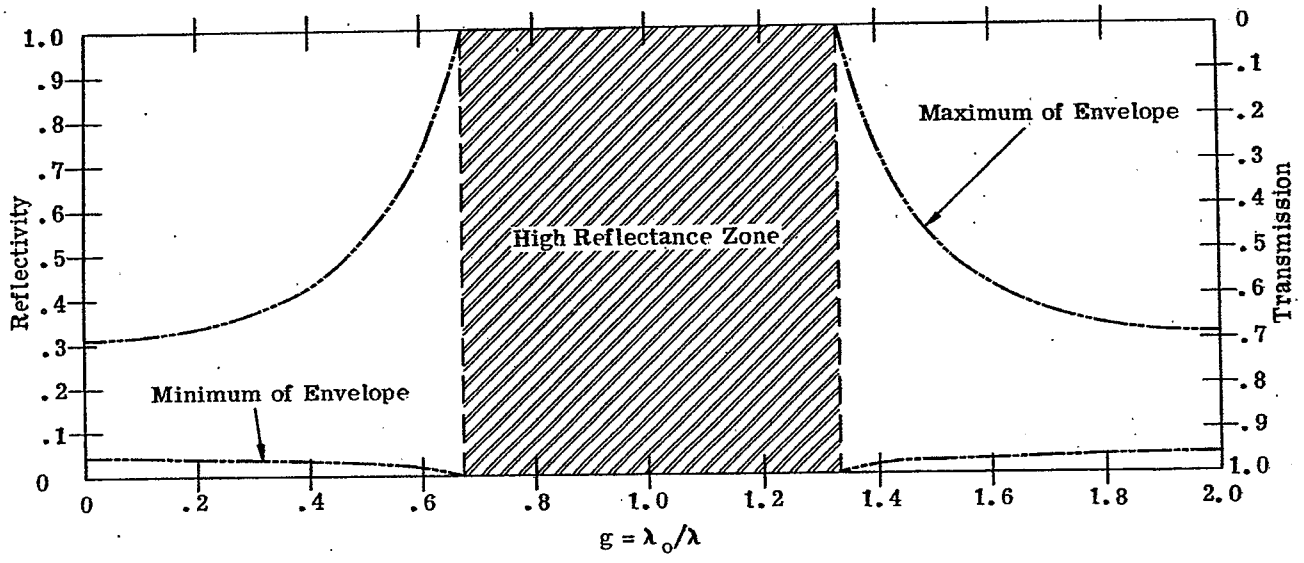


Figure 20.54 - Showing the high-reflectance zone, and the minimum and maximum of the reflectivity envelope of the quarter-wave stacks shown in Fig. 20.53

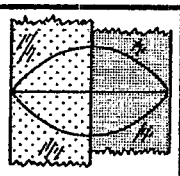
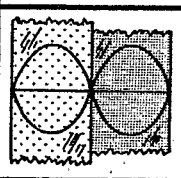
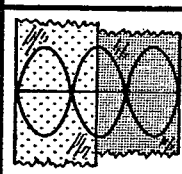
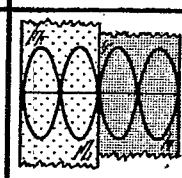
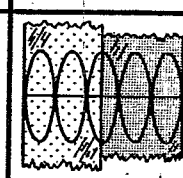
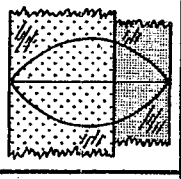
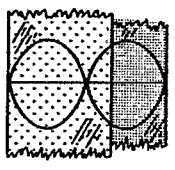
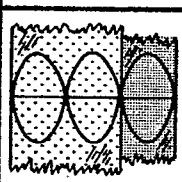
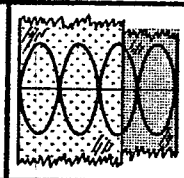
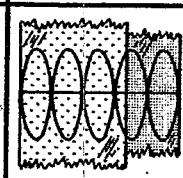
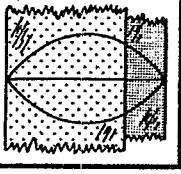
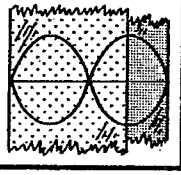
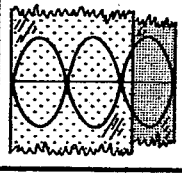
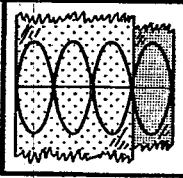
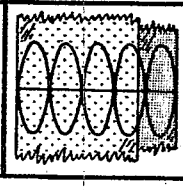
Type of Stack	g = 1.0	g = 2.0	g = 3.0	g = 4.0	g = 5.0	R_{max}, First Order, m=4
Quarterwave Basic Period L H						.957
2 : 1 Basic Period L' L' H'						.925
3 : 1 Basic Period L'' L'' L'' H''						.858

Figure 20.55 A comparison of the wavelength (in vacuo) with the optical thickness of a basic period of a multilayer with a periodic structure, for various orders of interference.

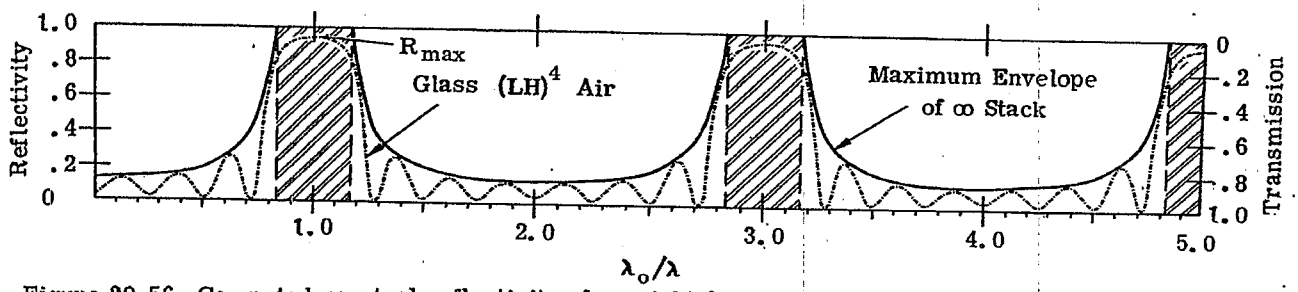


Figure 20.56 - Computed spectral reflectivity of an eight-layer quarter-wave stack (- - - - - - - -) and its envelope of maximum reflectivity (solid line). $n_H t_H + n_L t_L = \lambda_o/2$

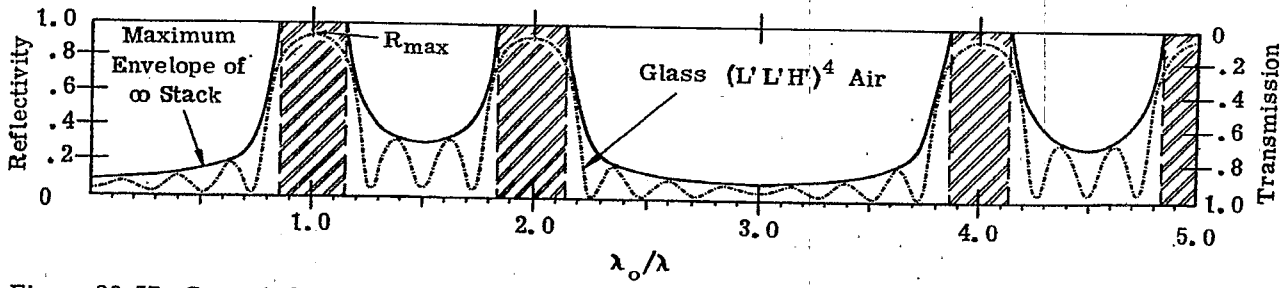


Figure 20.57 - Computed spectral reflectivity of an eight-layer 2:1 stack (- - - - - - - -) and its envelope of maximum reflectivity (solid line). $2\ n_L t_L + n_H t_H = \lambda_o/2$

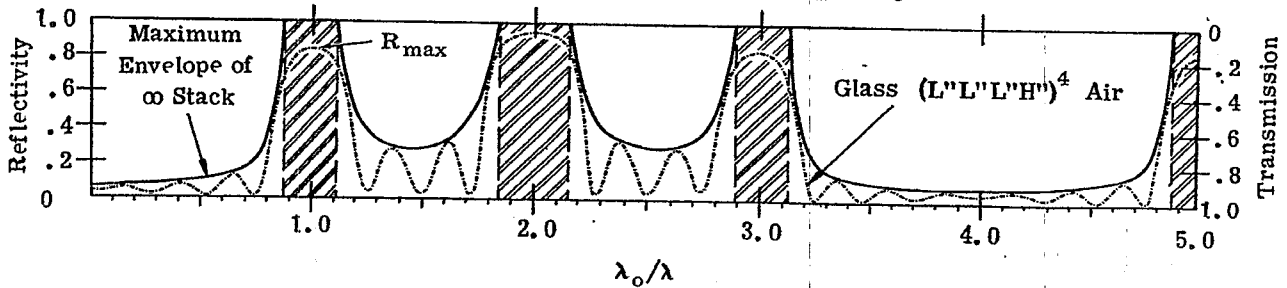


Figure 20.58 - Computed spectral reflectivity of an eight-layer 3:1 stack (- - - - - - - -) and its envelope of maximum reflectivity (solid line). $3\ n_L t_L + n_H t_H = \lambda_o/2$

20.4.2 Stacks with unequal thickness ratios.

20.4.2.1 General analysis.

20.4.2.1.1 Thus far we have considered the reflectivity properties of only a very specialized type of multilayer with a periodic structure, namely the quarter-wave stack, in which the two layers in a basic period have equal optical thickness. The high-reflectance zone occurs when the optical thickness of each layer is $\lambda_o/4$. Another way of stating this is to say that the high-reflectance zone occurs when the optical thickness of an entire period, LH equals $\lambda_o/2$. As is shown schematically in Figure 20.55, the high-reflectance zone occurs when a half-wave length fits into a basic period of stack. Consider the more general case where the two layers which compose the basic period of the stack do not have equal optical thickness, as for example, the stack

$$\text{glass } (HL')^6 \text{ air}$$

where the optical thickness of the L' layer is arbitrarily chosen to be 23% greater than the H layer. Does such a stack possess a high-reflectance zone ? The answer is yes. The position of the high-reflectance zones in any multilayer with a periodic structure can be found with the aid of the rule stated in the following paragraph.

20.4.2.1.2 A necessary but not sufficient condition for a high-reflectance zone to occur at a wavelength λ_o in a stack with a periodic structure is

$$\sum_i n_i t_i = q \frac{\lambda_o}{2} \qquad (46)$$

where q is an integer and the summation is over the basic period of the stack. Stated in other terms, the sum of the optical thicknesses of the layers comprising a basic period should equal an integral number of half wavelengths. This fact and the concept of an absentee layer (see Section 20.1.5.2.2) enables one to determine where the pass bands and high-reflectance zones of multilayer with a periodic structure occur.

20.4.3 Quarter-wave stack. Figure 20.56 shows the envelope function of the same quarter-wave stack shown in Figure 20.50, but over a larger range of g. Equation (46) is satisfied at $g = 1.0$ and hence this is in a high-reflectance zone. When $g = 2.0$, the optical thickness of the basic period LH is two half-waves and Equation (46) is satisfied for $q = 2$. Hence a high-reflectance zone could exist, but does not, because each of the layers in the stack is an absentee layer. Thus the reflectivity at $g = 2.0$ is the same as that of uncoated glass, namely .041. When $g = 3.0$, the total optical thickness of the basic period LH is three half-waves, and another high-reflectance zone occurs. When $g = 4.0$, the optical thickness of each layer is two half-waves and consequently all of the layers are absentee. Another high-reflectance zone occurs at $g = 5.0$, when the total optical thickness of the basic period LH is five-half waves. In the quarter-wave stack, high-refeitance zones occur at $g = 1, 3, 5, 7, 9, \ldots$ that is at odd integers. The high-reflectance zone which occurs at $g = 3$ is the third harmonic of the high-reflectance zone which occurs at $g = 1$, or to use the terminology of physical optics, this is a third-order interference peak. This is analogous to an open-ended organ pipe, which can sustain only odd harmonics.

20.4.3.1 The 2:1 stack. Let us study the reflectivity properties of the 2:1 stack:

$$\text{glass } L'L'H' \quad L'L'H' \quad L'L'H' \quad L'L'H' \quad L'L'H' \text{ etc.}$$

which can also be written as

$$\text{glass } (L'L'H')^m \text{ air}$$

where m is an integer. Here we have used the primed superscript, i.e. L' and H', to show that the optical thickness of these layers is different from the L and H which were used in the quarter-wave stack described in Section 20.4.1. In both cases, the optical thickness of the layers has been chosen so that a first-order interference high-reflectance zone occurs at λ_o. One must remember that the combination L'L' represents a single layer because each L' layer has the same refractive index. Thus the optical thickness of the L'L' layer is twice that of the H' layer and hence this is called a 2:1 stack. The spectral reflectivity curve of a 2:1 stack and its maximum reflectivity envelope are shown in Figure 20.57. Applying the rule which was stated in Section 20.4.2.1, the first-order high-reflectance zone occurs at $g = 1.0$ when the total optical thickness of the basic period L'L'H' is $\lambda_o/2$, as is shown in Figure 20.55. The second-order high-reflectance zone occurs at $g = 2.0$ when the total optical thickness of the basic period is two half-waves of λ_o. One might expect another high-reflectance zone at $g = 3.0$, but the optical thickness of the H' layer is a half-wave and the optical thickness of the L'L' layer is two-half waves. Hence, at $g = 3.0$, all of the layers are absentee layers and the reflectivity is that of

uncoated glass. At $g = 4.0$ and $g = 5.0$, the total optical thickness of the basic period equals four half-waves and five half-waves, respectively and high-reflectance zones occur. In this example, we have chosen L'L'H' as a basic period in which the optical thickness of the low-index layer is twice that of the high-index layer. The width of the high-reflectance zone is the same as the 2:1 stack, H'H'L'. As is shown in 20.4.6 however, the reflectivity properties at non-normal incidence of the two types of 2:1 stacks are quite different.

20.4.3.2 The 3:1 stack.

20.4.3.2.1 As a final illustration, consider the properties of stack with a periodic structure in which the optical thickness ratio is 3:1 :

$$\text{glass } (\text{L''L''L''H''})^m \text{ air}.$$

The double prime superscripts, H'' and L'' are used to show that the optical thickness of the L'' and H'' layers are different from both the H and L layers in the quarter-wave stack and the H' and L' layers of the 2:1 stack. The optical thickness of the H'' and the L'' layers is chosen so that the total optical thickness of the basic period of the stack, L''L''L''H'' equals $\lambda_o/2$, as is shown in Figure 20.55. Figure 20.58 shows a spectral reflectivity curve and the reflectivity envelope of a 3:1 stack. A first-order high-reflectance zone occurs at $g = 1.0$ and high-reflectance zones also occur at $g = 2.0$ and $g = 3.0$ when the total optical thickness of the basic period is two and three half-waves of λ_o, respectively. However, when $g = 4.0$ the optical thickness of the L''L''L'' layer is three half-waves and the optical thickness of the H' layer is a single half-wave. Hence all of the layers are absentee layers and the reflectivity at $g = 4.0$ is the same as uncoated glass. At $g = 5.0$ the total optical thickness of the basic period is five half-waves and hence another high-reflectance zone occurs. Thus every fourth high-reflectance zone is missing. Another way of stating this is to say that high-reflectance zones occur at the 1, 2, 3, 5, 6, 7, 9, 10, 11th etc. harmonics of the frequency of the fundamental.

20.4.3.2.2 The reflectivity curves of the multilayers with a periodic structure have been plotted on a frequency scale. It is possible that many readers are more accustomed to thinking in terms of wavelength, and so in Figure 20.59 is depicted the reflectivity versus wavelength of the 3:1 shown in Figure 20.58 with λ_o chosen to be 2.0 μ. A first-order reflectivity peak occurs at the fundamental wavelength, 2.0 μ. A second-order peak occurs at 1.0μ, which is one-half the fundamental and a third-order peak occurs at 0.667 μ, which is one-third of the fundamental. A fourth-order peak does not occur at $2.0/4 = 0.5\ \mu$ because the layers are absentee. A fifth order peak occurs at $\lambda_o/5 = 0.400\ \mu$, and so on. In comparing Figures 20.58 and 20.59, we note that on a frequency scale, the width of the first and third order reflectivity peaks is the same, whereas this is not true on a wavelength scale.

20.4.4 The general p:q stack.

20.4.4.1 General properties. It is patent that the analysis which we have made on the properties of the 1:1 (i.e. quarter-wave stack), 2:1 and 3:1 stack could be easily extended to any stack in which the ratio of the optical thickness of the layers is p:q, where q and p are integers. The following comparisons can be made between stacks which have a periodic structure, but different thickness ratios:

(1) The high-reflectance zone of the quarter-wave stack has even symmetry about $g = 1.0, 3.0$ etc. This is not necessarily true for other types of stacks.

(2) For a stack composed of layers of alternating refractive index n_a and n_b the width of the high-reflectance zone for any given ratio of $n_a : n_b$ is the largest when the optical thickness is equal. In other words, the high-reflectance zone of the quarter-wave stack is wider than high-reflectance zone of 2:1, 3:1, 3:2, etc. stacks.

(3) The width of the high-reflectance zone of a quarter-wave stack is given by a rather simple expression (i.e. Equation (39)). No such simple equations exist for other types of stacks. For example, the width of the high-reflectance zone of a 2:1 stack is given by the roots of a cubic equation.

(4) In the spectral region outside of the high-reflectance zones, the number of oscillations of the reflectivity increases as the number of periods (and hence the number of layers) of the stack increases. In any case, the oscillations will lie between the maximum and minimum reflectivity envelope of the infinite stack. The shape of this envelope function depends upon:

(a) The ratio of the optical thickness of the layers in a basic period.

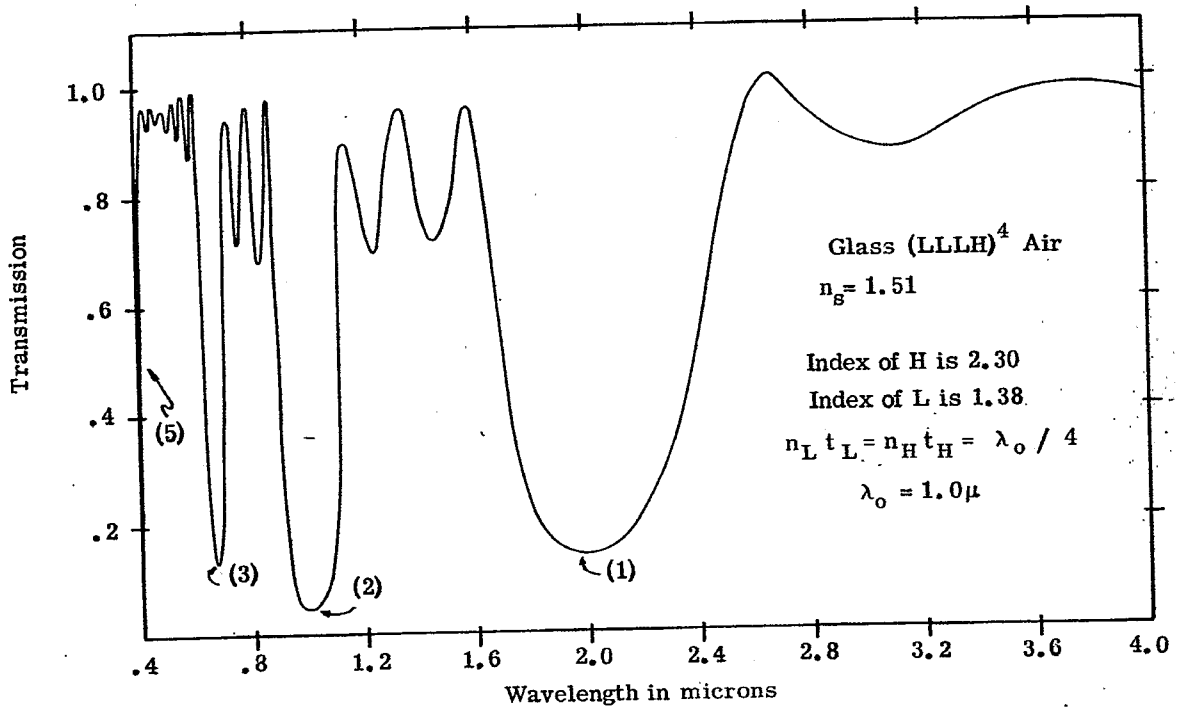


Figure 20.59 – Computed spectral transmission of an eight-layer 3:1 stack.

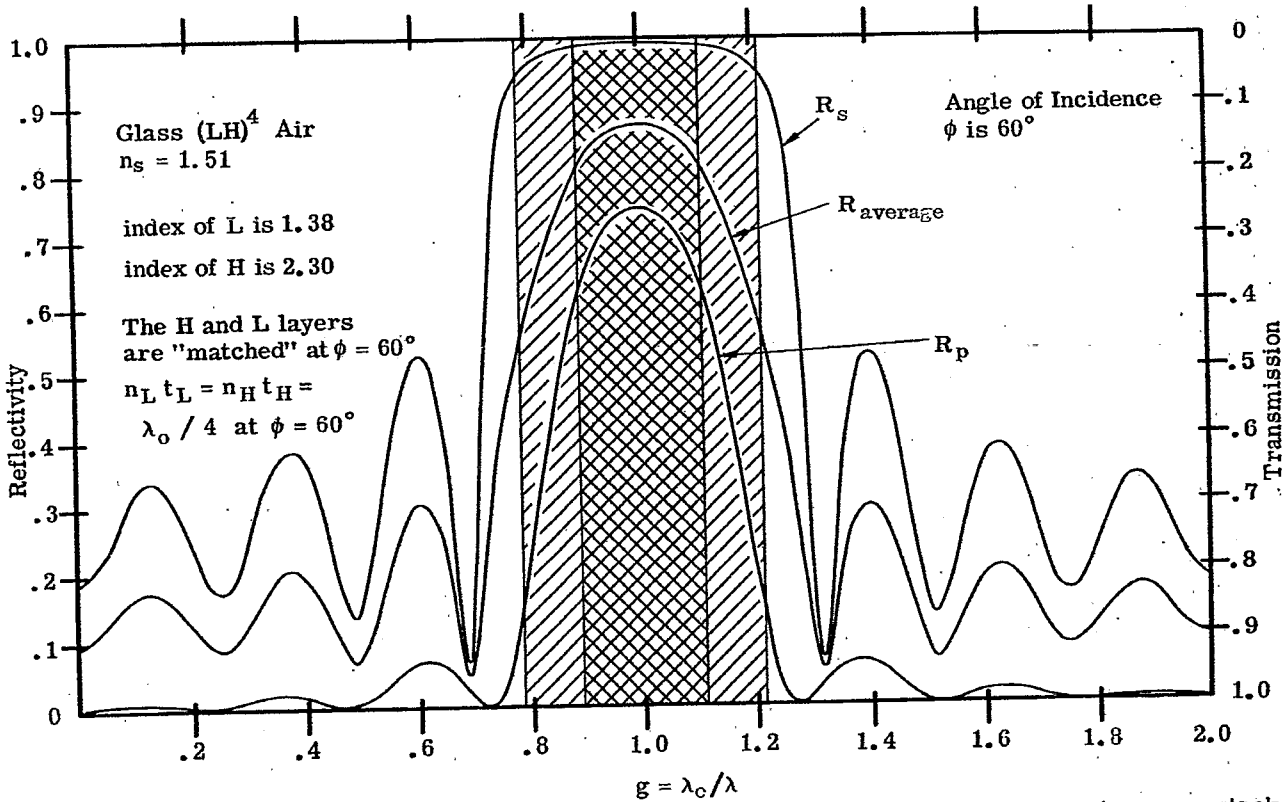


Figure 20.60 – Computed R_p, R_s, and $R_{av} = 1/2\ (R_p + R_s)$ of an eight-layer quarter-wave stack at $\phi = 60°$. The optical thickness of the H and L layers is matched 1:1 at $\phi = 60°$.

(b) The index of the two layers, n_a and n_b .

(c) The refractive index of the substrate, incident medium, and the refractive index and optical thickness of any layers which are added to the either end of the basic stack. This topic is discussed in Section 20.4.8.

(5) Given the indices n_o , n_s , n_a , n_b for any given order of interference the highest reflectivity for a given number of layers is obtained with a quarter-wave stack. This is illustrated in Figure 20.55, where the maximum reflectivity in the first order is listed for the quarter-wave, 2:1 and 3:1 stacks with an equal number of layers, with the same n_a, n_b, n_o, and n_s.

20.4.4.2 As an example of how the concept of absentee layers can be used to find the reflectivity at specific wavelengths, consider the ten-layer stack:

glass H'H'L' H'H'L' H'H'L' H'H'L' H'H'L' air

in which the optical thickness of the H'H' and L' layers has been chosen so that a first-order high-reflectance zone occurs at $g = 1.0$ as is shown in Figure 20.55. The problem is to find the reflectivity of this stack at $g = 1.5$. The optical thickness of each of the H'H' layers at $g = 1.5$ is a half-wave and hence each of these layers is absentee. Thus the stack reduces to:

glass L' L' L' L' L' air .

However, at $g = 1.5$ each of the L' layers is a quarter-wave in optical thickness and hence each pair of L' layers represents a half-wave. Thus four of the L' layers can be removed from the stack leaving:

glass L' air .

Thus the reflectivity of this ten-layer stack at $g = 1.5$ is the same as the reflectivity of a single quarter-wave layer on glass and can be readily computed from Equation (30).

20.4.4.3 The treatment given here of the propagation of waves in a medium with a periodic structure has been descriptive and qualitative. A rigorous mathematical treatment is given by Epstein[32], Brillouin[33], and Seitz[34].

20.4.5 Analogies.

20.4.5.1 Electrical transmission line. It is helpful to give some analogies between the propagation of light in a multilayer with a periodic structure and other fields, such as electrical engineering, x-ray crystallography, and solid state physics. Using the concepts which were presented in Section 20.1.4.1, the electrical transmission line shown in Figure 20.48 is the analogue of the quarter-wave stack shown in the same Figure. The transmission line consists of eight sections, each of which has an electrical length of a quarter-wave and which alternate between a high and low impedance. We recall that the load admittance of the transmission line is analogous to the refractive index of the substrate. As one can readily show by using a Smith chart, the load impedance is reflected back to the terminals of the preceeding section as a maximum value when the electrical length of the line is a quarter-wave. In this case the substitutional impedance at the input terminals is very large. Consequently there is large impedance mismatch between the characteristic impedance of the generator (which is 1.0) and the substitutional impedance of the line. This means that the voltage standing wave ratio is large and hence the voltage reflection coefficient to close to one.

20.4.5.2 Bragg reflection of x-rays. A beam of x-rays travels through a vacuum with the velocity of light. This velocity is pertubed slightly when the beam travels through a cloud of electrons, the amount of the perturbation being nearly proportional to the density of the electrons. In a crystal, there is a high density of electrons near each atomic lattice site. Since atoms are regularly spaced in a crystal, the x-rays travel through a periodically stratified medium. Although the change in the velocity is not abrupt, as it is for light propagating through a multilayer, nevertheless Bragg reflection of the x-rays is observed when the path difference between adjacent reflecting crystal planes is an integral number of wavelengths. This point is further amplified by Brillouin.[33]

20.4.5.3 The propagation of electrons in a crystal. The propagation of a single electron in a crystal is treated quantum mechanically by solving the Schrödinger equation, in which the electron is represented by a traveling wave with a deBroglie wavelength, λ_b . The velocity of the electron is perturbed by the electrostatic repulsion of electron cloud at each lattice site and consequently the electron moves in a potential which varies periodically. Regardless whether one chooses a simple one-dimensional periodical potential of Kronig and Penney, or a more sophisticated potential computed from atomic wave functions, the electron

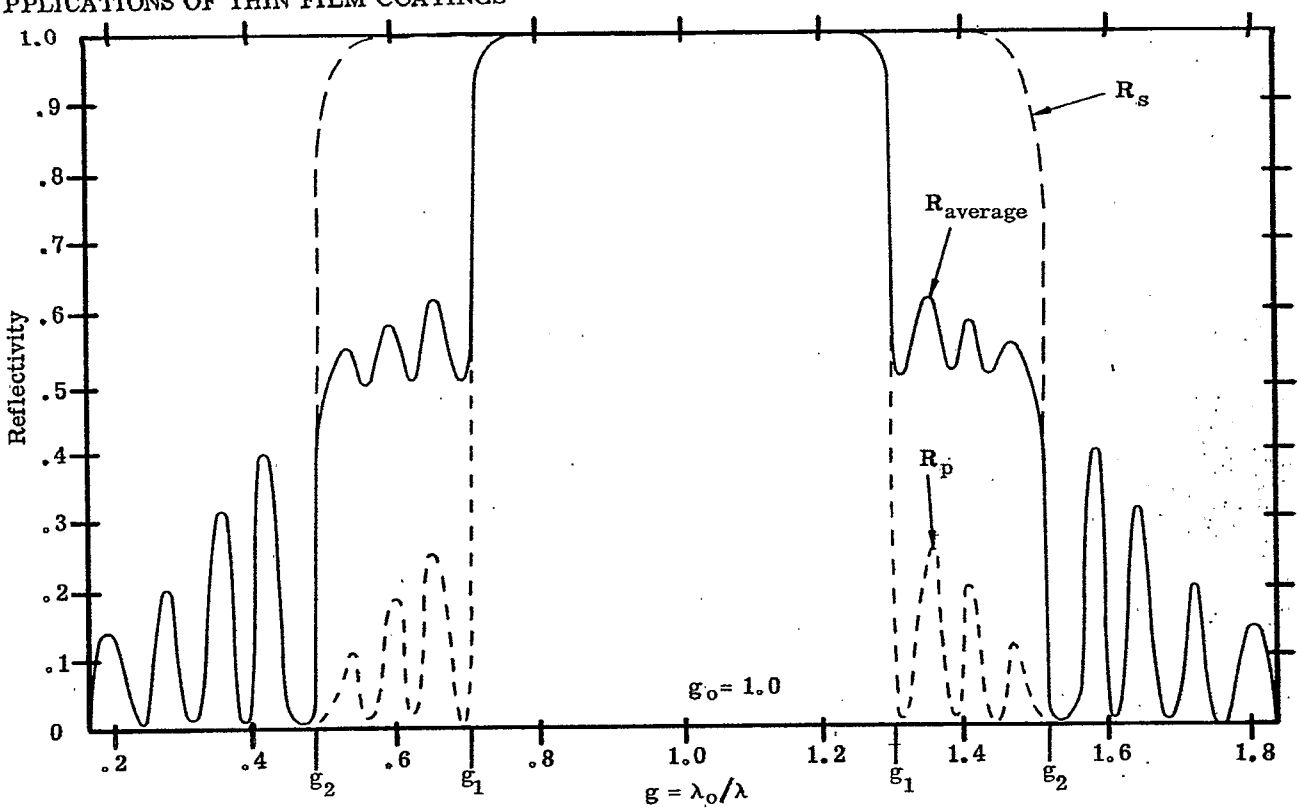


Figure 20.61 - R_p, R_s, and $R_{av} = 1/2 (R_p + R_s)$ at $\phi = \phi_o$ of a ficticious quarter-wave stack with the optical thicknesses of the layers matched 1:1 at ϕ_o.

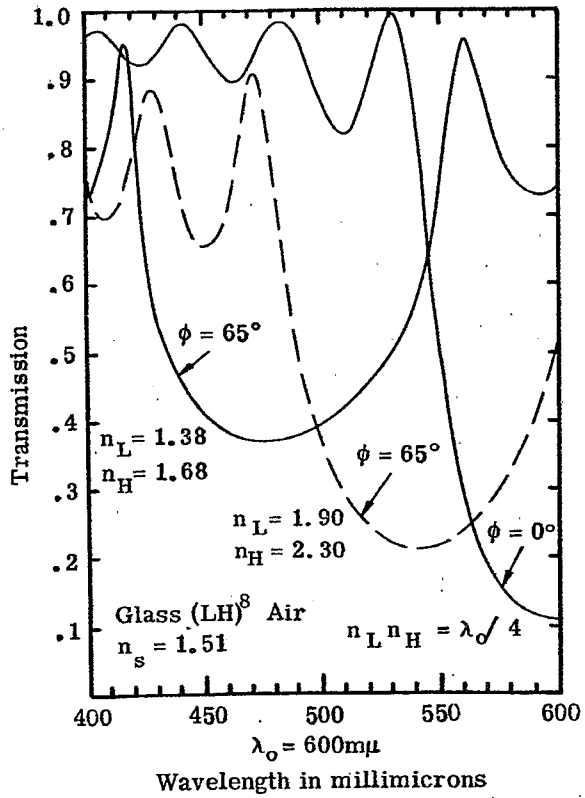


Figure 20.62 - Computed T_{av} at $\phi = 0$ and $\phi = 65°$ of sixteen-layer quarter-wave stacks with low index (solid line) and high index (dashed line).

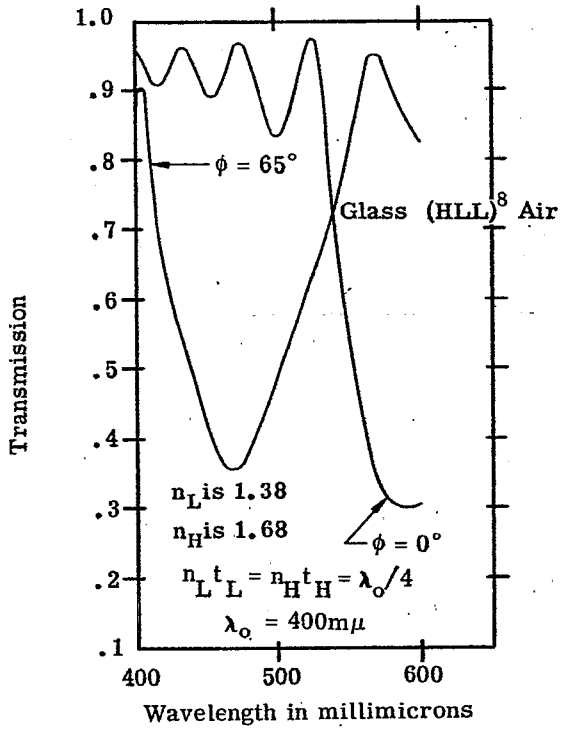


Figure 20.63 - Computed T_{av} at $\phi = 0$ and $\phi = 65°$ of a sixteen-layer 2:1 stack.

is reflected whenever $\lambda_b/2$ equals a multiple of the period of the lattice. An equivalent way of saying this is that the edge of the Brillouin zone occurs when the wave-vector $\underset{\sim}{K} = \pi/a$, where a is a lattice space in a particular direction and $\underset{\sim}{K}$ is wave-vector (Seitz)[34]. This is treated in detail by Brillouin [33], Seitz [34], and others.

20.4.6 The reflectivity of quarter-wave stacks at non-normal incidence.

20.4.6.1 Layers matched at angle. In considering the reflectivity of quarter-wave stacks at non-normal incidence, we must consider separately the case where thickness of the layers is matched (see Section 20.1.6.3, for a definition of this term) at normal incidence and the case where the layers are matched at a particular angle ϕ_o. In the latter case, the layers are deliberately mismatched so that at normal incidence the ratio of the optical thickness of the H and L layer is not 1:1, but the ratio is 1:1 at ϕ_o.

20.4.6.1.1 As an example of a match at angle, consider the eight-layer quarter-wave stack whose reflectivity curve is shown in Figure 20.60. The optical thickness of the low-index layer L has been made thicker than the high-index layer H in a ratio $n_H t_H : n_L t_L = 1.0 : 1.19$. In this case the layers are matched at $\phi_o = 60°$. Equation (39) can be used to compute the width of the high-reflection zone, using the effective index appropriate to each plane of polarization. For example, at $\phi = 60°$, in the "p" plane of polarization the effective index of the L layer is 1.773 and the effective index of the H layer is 2.483. From Equation (39) we find that the half-width, Δg, of the high-reflectance zone is 0.107. Similarly Δg in the "s" plane of polarization is 0.214, which is larger than the Δg at normal incidence, which is 0.161. The high-reflectance zones are shown in Figure 20.60 as cross hatched areas. The generalizations stated in 20.4.6.1.1 and 20.4.6.1.2 apply to quarter-wave stacks which matched at a particular angle ϕ_o, and are not confined to the eight-layer stack which has been used as an example.

20.4.6.1.2 The width of the high-reflectance zone increases in the "s" plane of polarization and decreases in the "p" plane of polarization. If many periods, say thirty or forty, are used in a stack, so that both R_p and R_s are close to 1.0 within their high-reflectance zones, the average reflectivity, $R_{av} = 1/2 (R_p + R_s)$ has the shape which is depicted in Figure 20.61. In the region between $g = 1.0$ at the center of the stack and g_1, a high-reflectance zone exists for both planes of polarization and hence R_{av} is close to 1.0. At g_1, the high-reflectance zone for the "p" plane of polarization ends and R_p fluctuates at low values outside of this zone. The high-reflectance zone for the "s" polarization extends to g_2 and in this intermediate region between g_1 and g_2 R_{av} attains a minimum value of 0.50. We observe that this shoulder on the R_{av} curve is due to the dissimilar width of the high-reflectance zones in the two planes of polarization. This shoulder has been somewhat exaggerated for purposes of illustration in Figure 20.61. The reflectivity curve of the eight-layer quarter-wave stack shown in Figure 20.60 does not show this shoulder, because the reflectance R_p does not drop to zero rapidly enough outside of the high-reflectance zone. If more periods were added to the stack, then eventually the R_{av} curve would show such a shoulder.

20.4.6.1.3 Maximum reflectivity. Since the thickness of the layers is matched at ϕ_o, the maximum reflectivity at $g = 1.0, 3.0, 5.0$, etc. can be computed in each plane of polarization from Equations (40) to (43), using the effective index appropriate to each plane of polarization.

20.4.6.2 Layers matched at normal incidence.

20.4.6.2.1 If a stack which is matched 1:1 at normal incidence is viewed at non-normal incidence, the position of the high-reflectance zone can be found by substituting effective thicknesses in Equation (46). The high-reflectance zone is no longer centered at a wavelength computed from Equation (46) and also the width of the zone (in each plane of polarization) is different than at normal incidence.

20.4.6.2.2 As an example of the application of Equation (46) to a stack at non-normal incidence, consider the 16 layer stack:

$$\text{glass} \quad (\text{LH})^8 \quad \text{air}$$

in which both the L and H have a QWOT of 600 mμ at normal incidence, that is,

$$n_i t_i = 150 \text{ m}\mu = 600 \text{ m}\mu/4 .$$

The transmission curve, shown in Figure 20.62, has a minimum at 600 mμ which is the center of the high-reflectance zone. Two cases are considered separately at $\phi = 65°$.

Case I $n_L = 1.90$, $n_H = 2.30$. From Figure 20.8 we find that the change in effective thickness is 0.92 for the H layer and 0.891 for the L layer, the latter value being found by linear interpolation.

Substituting these values for the effective thickness into Equation (46), we obtain

$$(.92)\ (150\ m\mu)\ +\ (.879)\ (150\ m\mu)\ =\ \frac{\lambda_1}{2}\ , \qquad (47)$$

whence $\lambda_1 = 540\ m\mu$. From Figure 20.62 we see that the minimum T_{av} at $\phi_o = 65°$ is very close to $540\ m\mu$.

Case II $n_L = 1.38$ and $n_H = 1.68$. The change in the effective thickness is found from Figure 20.9 to be 0.754 and 0.842. Solving an equation similar to Equation (47) gives the result that $\lambda_1 = 480\ m\mu$. Actually, Figure 20.62 shows that the wavelength at which T_{av} is a minimum is shifted to shorter wavelengths by about $10\ m\mu$.

20.4.7 Minimization of the angle shift.

20.4.7.1 The basic problem. The transmission curves of all multilayer filters change with the angle of incidence and usually exhibit a shift to shorter-wavelengths. This angle shift is of little importance if the multilayer filter is illuminated with collimated light at one angle of incidence. However, serious problems often arise when the light is highly convergent. For example, either of the multilayers shown in Figure 20.62 would attenuate the sodium yellow lines (at $589\ m\mu$) if used at normal incidence, but they would be quite ineffective at $\phi = 65°$. The problem of the angle shift of a multilayer filter is similar to the problem of the chromatic aberration of a lens. In either case, the effect is ubiquitous and at best the designer can only minimize it. Two methods of minimizing the angle shift are: (1) The use of high-index materials. (2) The use of more of the higher index material in the basic period.

20.4.7.2 The use of high-index materials. From the examples in 20.4.6.2.2, it is patent that the change in the effective thickness of each of the layers is much less for high-index layers than for low-index layers. Thus the angle shift of the quarter-wave stack with $n_L = 1.90$, $n_H = 2.30$ is much less than the stack with $n_L = 1.38$, $n_H = 1.68$, even though the width of the high-reflectance zone at normal incidence is the same because the index ratio n_L/n_H is the same. From this it follows that if a multilayer with a periodic structure is used as a "cutoff" filter - that is to pass, either in the long-wave or short-wavelengths region - it should contain high index materials if the angle shift is to be minimized. For example, an infrared filter which contains cryolite and germanium has a larger angle shift than a filter which contains zinc sulfide and germanium, because the refractive index of zinc sulfide is nearly twice as large as that of cryolite.

20.4.7.3 The use of more high-index material in a basic period. If only quarter-wave stacks were considered, then the discussion would terminate with 20.4.7.2. If, however, the optical thickness of the high-index material is different from the thickness of the low index material, such as occurs in a 2:1 or a 3:1 stack, then it is possible to reduce the angle shift below that of a quarter-wave stack (using the same materials) by using the high-index material in the thicker layer. That is to say, a $(LHH)^m$ stack has a lower angle shift than the stack $(HLL)^m$, even though the width of the high-reflectance zone at normal incidence of these two stacks is identical. Similarly, the stack $(LHHH)^m$ has a smaller angle shift than the stack $(HLLL)^m$. These facts are illustrated in Figure 20.63 which shows the spectral transmission of the 2:1 stack

$$\text{glass}\ (H\,LL)^8\ \text{air}$$

at normal incidence and at 65°. In this stack, the optical thickness of the low-index layer is greater than that of the high-index layer. The 2:1 stack whose transmission curve is shown in Figure 20.64 has more high-index material in the basic period:

$$\text{glass}\ (L\,HH)^8\ \text{air}.$$

In Figure 20.64 the T_{av} at 65° of the $(L\,HH)$ stack (shown as solid line) is compared with the H LL stack (dashed line). Although both stacks show a considerable angle shift to the blue, the shift of the LHH stack is definitely less. The angle shift of the HLL stack is considerably greater than the comparable quarter-wave stack shown in Figure 20.62 which uses the same refractive indices. Figures 20.79, 20.86, and 20.91 show the angle shift of transmission curve of various multilayers.

20.4.8 Variations on the basic periodic structure.

20.4.8.1 General considerations. Most of the quarter-wave stacks and other multilayers with a periodic structure shown in Figures 20.48 to 20.64 have been selected primarily as illustrations and should be modified slightly if they are used as practical filters. Additional layers can be added to these stacks for either

of two purposes:

(1) To increase the reflectivity in the high reflectance zone.

(2) To increase the transmission in the spectral region outside of the high-reflectance zone.

If a quarter-wave stack were used as a semi-transparent mirror coating for a Fabry-Perot interferometer, then primary objective would be to obtain a high reflectivity over a specific spectral region and little interest is paid to the transmission outside of this region. On the other hand, if a quarter-wave stack were used as a long-wave pass filter, then it is important to optimize the transmission in the long-wave region.

20.4.8.2 Increasing the reflectivity. It is evident from Equations (42) and (43) that a quarter-wave stack has a higher R_{max} when a high-index layer is next to both the substrate and the incident medium. The R_{max} of this odd-layered stack is greater than the R_{max} of an even-layered stack. This is illustrated in Figure 20.65, shows the computed reflectivity of a six-layer and a seven-layer stack which use the same n_a and n_b. Even though an additional H layer has been added to the basic stack so that its design can no longer be represented as $(HL)^m$, the seven-layer stack is still called a quarter-wave stack. It is evident that a considerable increase in the reflectivity has been achieved by the addition of the extra H layer. If an additional L layer were added so that the multilayer design is,

glass $(HL)^4$ air ,

the R_{max} would be less that of the seven-layer stack. An increase in the maximum reflectivity is also achieved if an odd number of layers are used rather than an even number in the 2:1 and 3:1 stacks shown in Figures 20.57 and 20.58 .

20.4.8.2.2 The effect on R of a mismatch in layer thickness. Closely allied to the problem of attaining the maximum reflectivity is the problem of a mismatch in the thickness of the layers. If the optical thickness of any one of the layers in a quarter-wave stack, or other type of multilayer with a periodic structure, deviates by a small amount from $\lambda_o/4$, then the reflectivity throughout the entire high-reflectance zone falls below the value which would be attained if all the layers were perfectly matched. For example, suppose that a quarter-wave stack is manufactured by evaporating each of the layers in a vacuum and that in this process errors of a random nature are made in controlling the thickness of each of the layers. Consequently, the optical thickness of each of the layers differs from $\lambda_o/4$ by a random amount. As long as these errors are not excessive, say greater than ten percent, the region of high-reflectivity about λ_o which is characteristic of a quarter-wave stack, is still observed. However, the R_{max} of such a multilayer is not as high as it would be for a perfect stack and also the cutoff at the edge of the high-reflectance zone is not as steep. The fact that rather large errors can be made in controlling the thickness of the layers without serious detrimental effects upon the reflectivity is the factor which permits certain types of band pass multilayer filters to be manufactured with relatively crude monitoring equipment to control the thickness of the layers.

20.4.8.3 Enhancing the transmission in the spectral region outside of the high-reflectance zone. Several methods are used to enhance the transmission in the spectral region outside of the high-reflectance zone:

(1) Additional layers of non-quarter-wave optical thickness can be added to the stack. This is discussed in 20.4.8.3.1.

(2) It is possible to vary the thickness of each of the layers by a small amount so that the transmission increases outside of high-reflectance zone. Since this method improves upon, or refines, an existing multilayer design, it is called the refining method. This is covered in 20.4.8.4.

20.4.8.3.1 An effective method of enhancing the transmission in the spectral region outside of the high-reflectance zone is to add additional layers to the basic stack (which has a periodic structure). For example, one could start with the twelve-layer quarter-wave stack

glass $(HL)^6$ air

and add several layers to either end of the stack:

glass $L_1\ H_2\ (HL)^6\ H_3\ L_4$ air

where the additional layers L_1 , H_2 , H_3 L_4 are tagged with subscripts to emphasize the fact that they do not necessarily have the same thickness or refractive index as the H and L layers in the

basic stack. As an example of this procedure, suppose that a quarter-wave stack is used as a short-wave pass filter which is intended to pass the blue and green, but attenuate in the yellow and red. A seven-layer stack of zinc sulfide ($n = 2.30$) and magnesium fluoride ($n = 1.38$) is considered:

glass H L H L H L H air .

The reflectivity versus frequency curve is shown in Figure 20.65, and has even symmetry about λ_o . The region of high reflectivity extending from $g = 0.8$ to 1.2 would attenuate the red and yellow, but this multilayer would be much more effective as a short-wave pass filter if the reflectivity peaks at $g = 1.39$ and 1.63 could be decreased. This accomplished if a low-index layer of eighth-wave optical thickness is added to each end of the stack:

glass $\frac{L}{2}$ H L H L H L H $\frac{L}{2}$ air .

The spectral transmission of such a stack with $\lambda_o = 700$ mμ is shown in Figure 20.66. The effect of adding the eighth-wave layers is to increase the transmission in the short-wave region. The reflectivity peak at $g = 1.39$ (at 504 mμ in Figure 20.66) has been decreased from 0.25 to 0.11 and the peak at $g = 1.63$ (430 mμ in Figure 20.66) is barely perceptible. The reflectivity in the long wavelength region has actually been increased slightly from .25 to .27 . The addition of the eighth-wave layers to the seven-layer stack has decreased R_{max} from .95 to .94 . Additional layers can be added to this nine-layer stack to increase the transmission even further in the short-wavelength region. Epstein[35] elaborates on methods of accomplishing this. As an example of how the transmission can be increased in the long-wave region, consider the multilayer

glass $\frac{H}{2}$ L H L H L H L $\frac{H}{2}$ air .

The spectral transmission of this multilayer is shown in Figure 20.67. The transmission in the red and infrared is quite high but the transmission in the near ultraviolet is considerably lower than that of a quarter-wave stack LHLHLHL . This is usually the case, that when the transmission is increased on the long-wave side of the high-reflectance zone, it is decreased on the short-wave side, and vice versa. In practice, the ultraviolet transmission of the stack shown in Figure 20.67 would be much lower than the computed values if zinc sulfide were used as the high-index layer material, since this material absorbs strongly below 400 mμ . The design of the multilayer shown in Figure 20.67 can also be written as:

glass $\left(\frac{H}{2} \; L \; \frac{H}{2} \right)^4$ air .

This can be regarded as a multilayer with a periodic structure with $\left(\frac{H}{2} \; L \; \frac{H}{2} \right)$ as a basic period. Equation (46) still applies and consequently a high-reflectance zone can occur when an integral number of half-waves fit into the basic period. When the design of the stack can be written in this manner, it can be shown that at any wavelength outside of the high-reflectance zone, it is possible to replace the entire stack of nine layers by a single layer with a ficticious index n_h and ficticious optical thickness τ . That is, for the purposes of computing the reflectivity (outside of the high-reflectance zone), the nine-layer stack is equivalent to a single-layer of index n_h and optical thickness τ . The index n_h is called the Herpin equivalent index. Space does not permit us to describe how the concept of the Herpin equivalent index is used to design multilayer combinations which have a high transmission outside of the high-reflectance zone. For further details, the reader can refer to Weinstein[5] or Epstein[35] . To cite some additional examples, eighth-wave layers of high-index material have been added to the quarter-wave stacks shown in Figures 20.73, 20.78, 20.80, 20.81,and 20.82 to increase the transmission in the long-wave region. The multilayers shown in Figures 20.82 and 20.83 are quarter-wave stacks modified so that the transmission is optimized in the short-wave region. An eighth-wave layer of low index material is added to one end of the quarter-wave stack (next to the incident medium), while a layer of optical thickness $1.28 \; \lambda_o/4$ is inserted between the quarter-wave stack and the substrate.

20.4.8.4 The refining method.[36] The refining method consists of varying by a small amount the thickness of each of the layers of a multilayer of specified design, so that the transmission is increased (or decreased, as the case may be) at certain specified wavelengths. This method can be used to increase the transmission of a quarter-wave stack on either the short-wave or long-wave side of the high-reflectance zone. As is mentioned in 20.4.8.1.1, the thickness of each of the layers can be changed by as much as 10% without seriously affecting the reflectivity in the high-reflectance zone. The refining method is similar to the relaxation method which is used to solve complex engineering problems. The computations are sufficiently lengthy and tedious that it is necessary to employ an electronic digital computer. As an example of the application of the refining method, suppose it is desired to increase the transmission of a quarter-wave stack which transmits in the blue but attenuates longer wavelengths. The computed reflectivity of a modified quarter-wave stack is shown in Figure 20.68. The effect of using a large number of layers is to achieve a high attenuation in the green and red ($T_{min} = .001$) and a sharp cutoff at

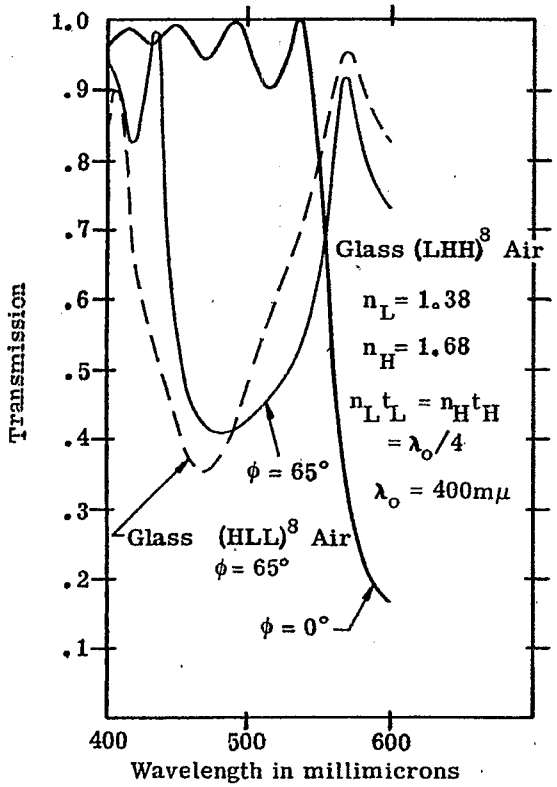


Figure 20.64- T_{av} of a sixteen-layer 2:1 stack at ϕ = 0 and 65°. The ϕ = 65° curve from Fig. 64 is shown as a dashed line.

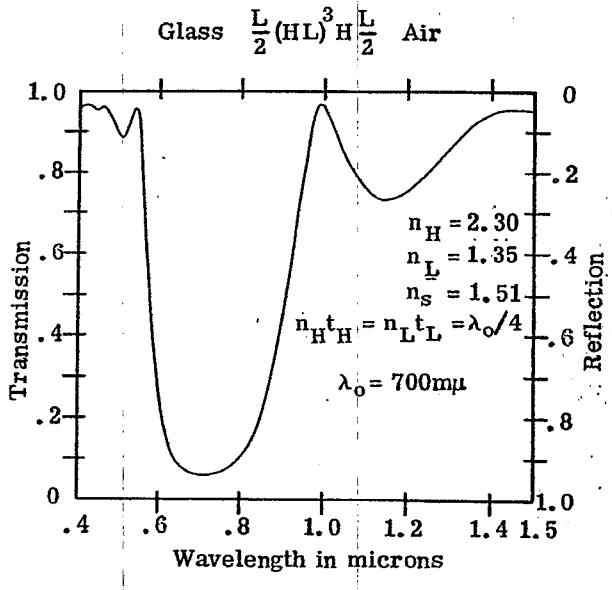


Figure 20.66- Computed spectral transmission of a short-wave pass filter consisting of a modified quarter-wave stack.

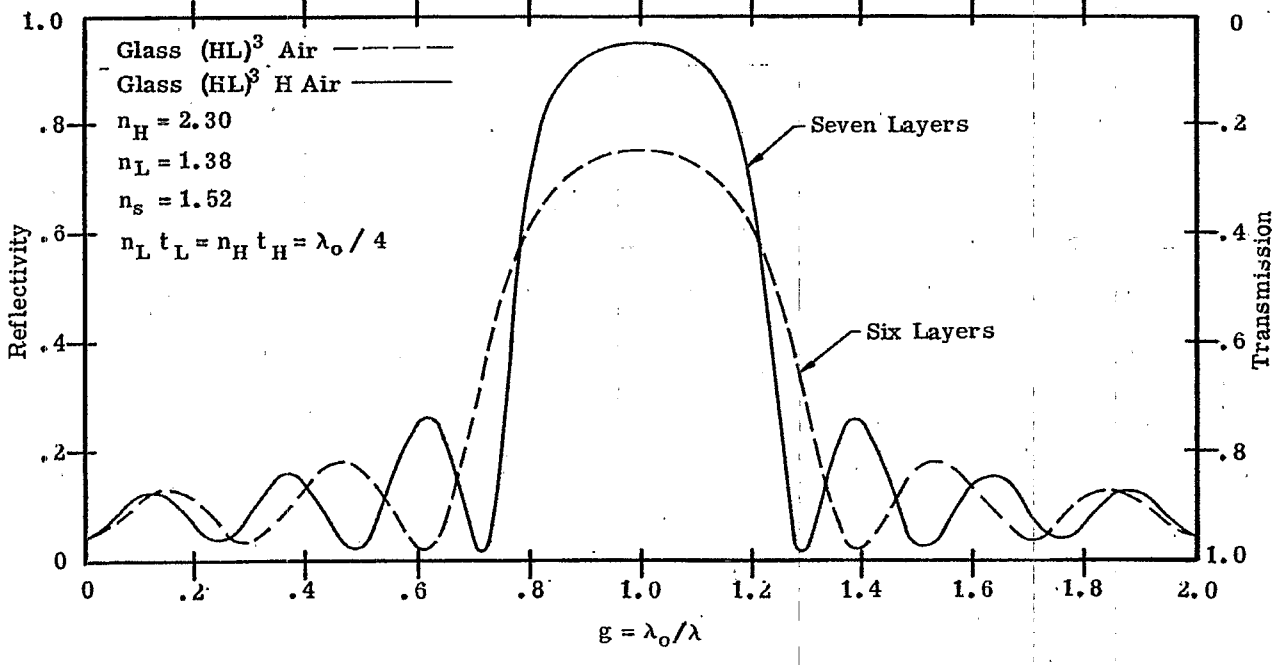


Figure 20.65- Computed spectral reflectivity of six-layer (dashed line) and a seven-layer (solid line) quarter-wave stack.

500 mμ . The reflectivity of the modified quarter-wave stack (DESIGN I) attains a maximum of .33 in this region. This multilayer would be much more effective as a blue pass filter if the undesirable peaks in the reflectivity at 484 mμ and 446 mμ could be eliminated. This is accomplished by varying the thickness of each of the layers by a relaxation process. Table 20.4 shows the quarter-wave optical thickness of each of the layers of the initial stack (Design I) and the refined design (Design II). The maximum change in the thickness of any of the layers is only 12% and the thickness of most of the layers has been changed by only a few percent. This mismatch in the optical thickness of the layers has caused the peak reflectivity to decrease from .9990 (in Design I) to .9988 (in Design IX). Also, the steepness of the transition from the high-reflectance zone to the pass region has been slightly decreased. As another example, the curve designated as Design III in Figure 20.69 depicts the spectral transmission of a quarter-wave stack composed of germanium and silicon monoxide which is used as a long-wave pass filter for the infrared. Table 20.4 shows the thickness of each of the layers (as a fraction $\lambda_o/4$) of the refined multilayer (Design IV) whose transmission curve is also shown in Figure 20.69 .

20.5 LONG-WAVE PASS FILTERS

20.5.1 General properties. From the discussion in 20.4 about the properties of a multilayer with a periodic structure, the method of designing either a long-wave or a short-wave pass filter is fairly obvious. One simply chooses a quarter-wave stack or other type of stack with periodic structure, so that the high-reflectance zone covers the region to be attenuated. It is necessary to choose the materials which are used in the stack and also the number of layers. This can be accomplished after the properties of the multilayer stack have been specified. In establishing the specifications of a long-pass, short-pass, or a band-pass multilayer filter, some of the following properties are considered:

(1) The optical density in the attenuation region. This is discussed in 20.5.1.1.

(2) The transmission in the pass region. In 20.4.8.3 two methods of enhancing transmission in the pass region are presented.

(3) The steepness of the cutoff.* This is discussed in 20.5.1.2.

(4) The change of the transmission with angle, i.e. the angle shift. If a minimum amount of angle shift is required, then high-index materials should be used in the stack, as is mentioned in 20.4.7.

(5) The change of transmission with temperature. In certain applications multilayer filters are used in environments which are either warmer or cooler than room temperature, and the shift of the spectral transmission must be taken into account. Multilayer filters are sometimes placed in thermal contact with an infrared detector which is cooled with liquid nitrogen. Figure 20.79 shows the shift with temperature of the transmission of an infrared filter.

20.5.1.1 The optical density in the attenuation region increases as the index mismatch between the layers of a stack increases and also as the total number of periods increases. In the special case of a quarter-wave stack, the maximum attenuation can be computed from Equations (41) through (43).

20.5.1.2 The sharpness or the steepness of the cutoff * is also an important parameter. The cutoff is the region in which the transmission drops from "high" values in the pass region to "low" values in the attenuation region (in the high reflectance zone). The words low and high in the preceeding sentence are enclosed in quotation marks because the criterion of what constitutes a low or high transmission is rather arbitrary. For example, in the multilayer shown in Figure 20.76 the high value is chosen as 0.70 and the low value as .05. Similarly, the wavelength at which the transmission has decreased to some arbitrary value is called the cutoff wavelength. For example, in Figure 20.79 the wavelength of the $T = .05$ point is chosen as the cutoff wavelength. Regardless of what criterion is chosen, the steepness of the cutoff increases as the number of periods increases. Let us apply as a criterion of the sharpness of the cutoff the wavelength difference $\Delta\lambda$ between the $T = 0.9$ point and $T = 0.1$. Each of the multilayers, whose transmission curves are shown in Figures 20.66 and 20.68, contain the same materials, but have a different number of periods. The $\Delta\lambda$ for the nine-layer stack (Figure 20.66) is 90 mμ, whereas the $\Delta\lambda$ for the seventeen-layer stack (Figure 20.68) is 20 mμ. Although the comparison does a slight unjustice to the nine-layer stack because its cutoff is at longer wavelengths, nevertheless this serves to illustrate the point that the steepness of the cutoff increases at the number of basic periods increases. The order of interfer-

* Some writers use the word "cuton" to denote the onset of the attenuation region and "cutoff" to denote the "cutoff" (or stopping) of the attenuation region. In section 20, this distinction is not made.

LAYER	INDEX OF LAYER	QUARTER-WAVE OPTICAL THICKNESS IN mμ		LAYER	INDEX OF LAYER	OPTICAL THICKNESS OF LAYER IN UNITS OF $\lambda_o/4$	
		DESIGN I	DESIGN II			DESIGN III	DESIGN IV
AIR	1.00	MASSIVE		AIR	1.00	MASSIVE	
1	1.38	300	306	1	4.00	0.5	0.419
2	2.30	599	639	2	1.80	1.0	1.181
3	1.38	599	630	3	4.00	1.0	1.272
4	2.30	599	610	4	1.80	1.0	0.967
5	1.38	599	606	5	4.00	1.0	0.879
6	2.30	599	597	6	1.80	1.0	1.060
7	1.38	599	582	7	4.00	1.0	1.194
8	2.30	599	573	8	1.80	1.0	1.023
9	1.38	599	577	9	4.00	1.0	0.874
10	2.30	599	589	10	1.80	1.0	0.961
11	1.38	599	596	11	4.00	1.0	1.143
12	2.30	599	590	12	1.80	1.0	0.979
13	1.38	599	585	13	4.00	0.5	0.405
14	2.30	599	601	GLASS	1.52	MASSIVE	
15	1.38	599	646				
16	2.30	599	672				
17	1.38	599	616				
GLASS	1.52	MASSIVE					

Table 20.4- The design of multilayer filters, whose transmission and reflectivity curves, are shown in Figures 20.68 and 20.69 .

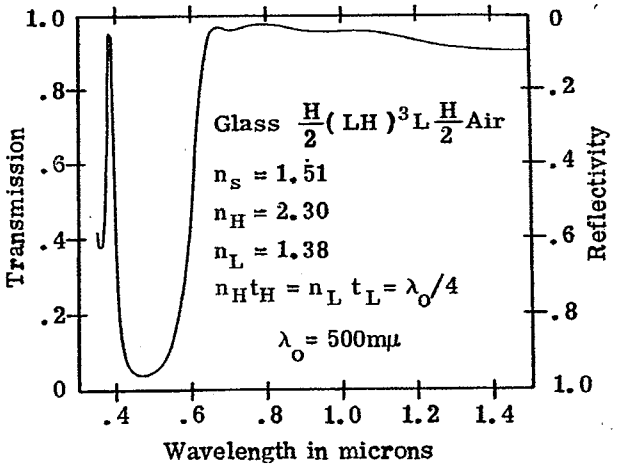


Figure 20.67- Computed spectral transmission of a long-wave pass filter consisting of a modified quarter-wave stack.

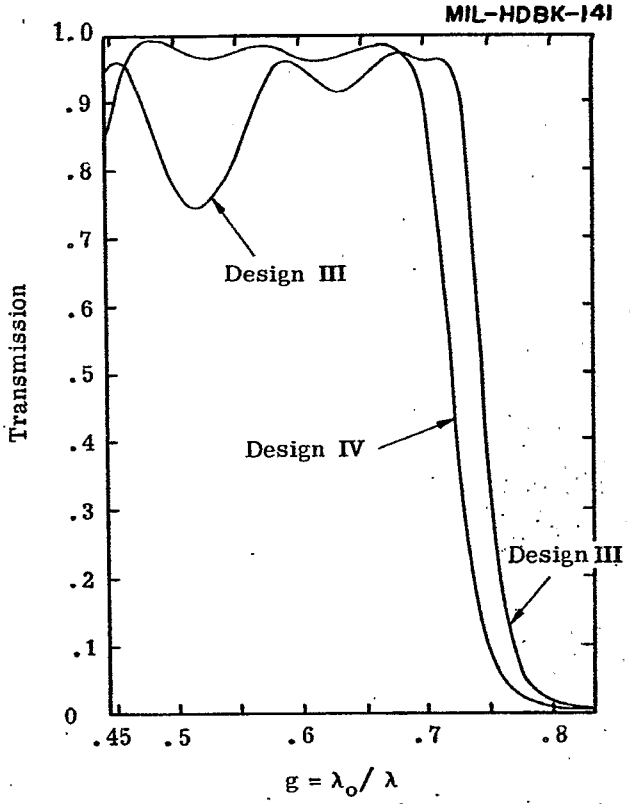


Figure 20.69- Computed spectral transmission of multilayers designated as Design III and Design IV in Table 20.4.

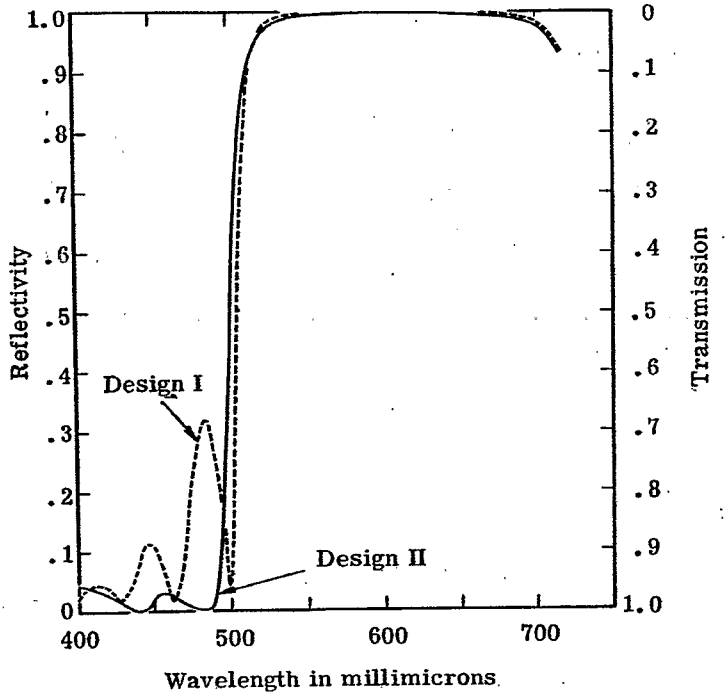


Figure 20.68- Computed spectral reflectivity of the multilayers designated as Design I (dashed line) and Design II (solid line) in Table 20.4.

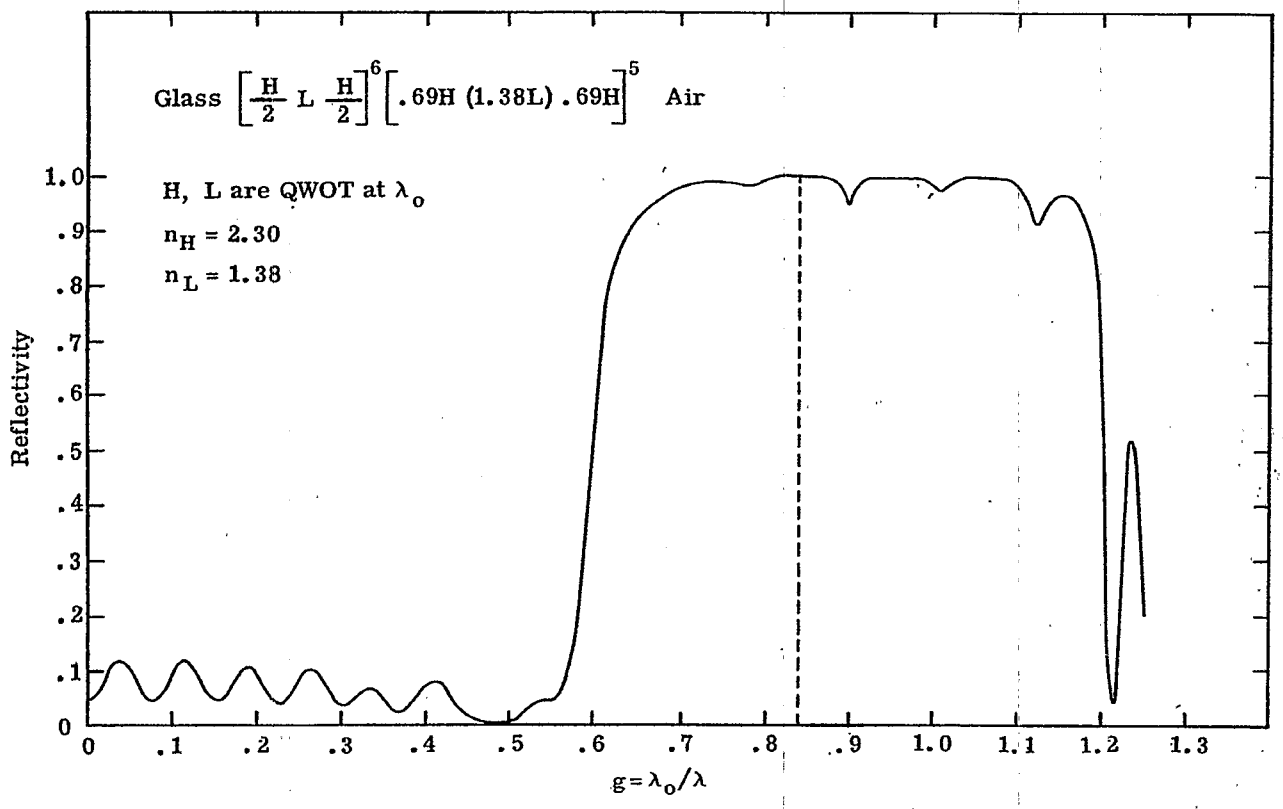


Figure 20.70- Computed spectral reflectivity of a multilayer consisting of an ensemble to two stacks.

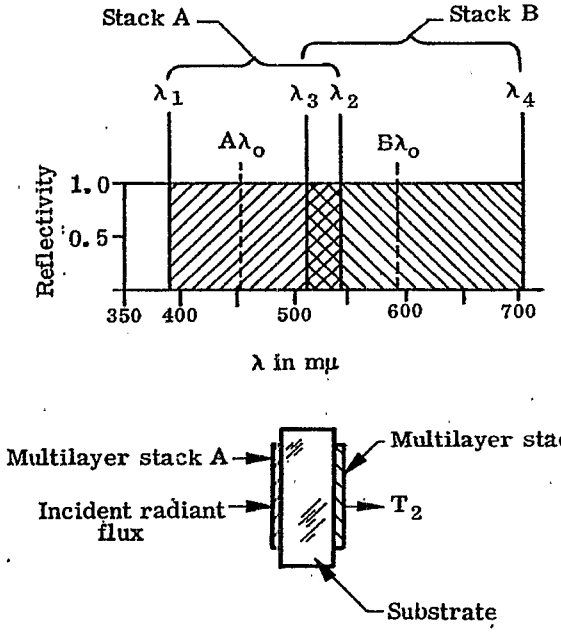
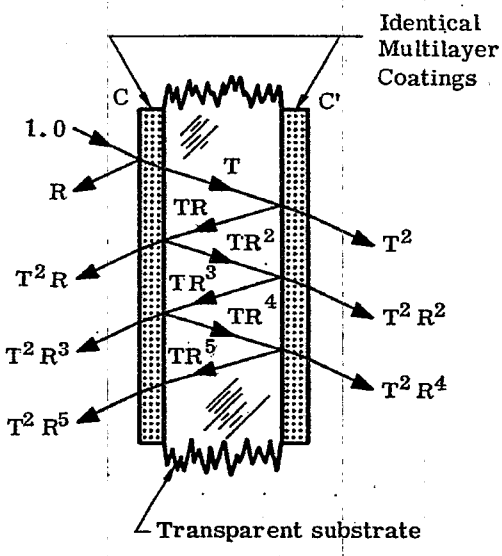


Figure 20.71- The positions of the high-reflectance zones of the two multilayers which constitute a composite filter.



$$T_2 = T^2\left[1 + R^2 + R^4 + \ldots\right]$$

$$= \frac{T^2}{1 - R^2} = \frac{(1 - R)^2}{(1-R)(1+R)}$$

$$= \frac{1 - R}{1 + R} = \frac{T}{1 + R}$$

Figure 20.72- An illustration of the method of computing the total transmission, T_2 of a composite filter consisting of two identical non-absorbing coatings.

ence of the high-reflectance peak influences the steepness of the cutoff. At any given wavelength, a higher-order reflectance peak has a sharper cutoff than a first order peak. The disadvantage of using a higher-order reflectance peak is that the width of the region of high transmission at longer wavelengths is decreased. However, if a first-order high-reflectance peak is used to provide the attenuation region, then there are no regions of high-reflectivity between the long-wave cutoff and infinite wavelength. For example, reflectance peak at 500 mμ of the multilayer shown in Figure 20.67 is first order. In this case the computed transmission of this multilayer is greater than 0.85 mμ in the spectral region from 650 mμ to wavelengths as long as 50 μ in the infrared, or for wavelengths in the microwave region for that matter. Of course in practice the glass substrate and the materials in the multilayer would absorb strongly at wavelengths much shorter than 50 μ.

20.5.1.3 Attenuation over a broad spectral region. It was shown in 20.4.1.4 that the width of the high-reflectance zone is determined by the index ratio of the two materials which constitute the stack and hence the width of the high-reflectance zone is limited by the index of available thin film materials. The optimum width is achieved with a quarter-wave stack which has a large index mismatch. For example, in the region from 300 to 400 mμ the non-absorbing thin film materials have an index which does not exceed 2.10. The optimum width of the high-reflectance zone is achieved if this high-index material is used in conjunction with a low index material such as cryolite (index 1.35). Sometimes it is desired to attenuate a spectral region which is greater than this optimum width. One obvious solution is to combine two or more such stacks. The thickness of the layers in each of the stacks is chosen so that the high-reflectance zone of each one covers a different part of the spectral region to be attenuated. Here the term stack has been used to denote a group of films which have periodic structure. There are two ways of arranging the stacks so that a broad attenuation region is achieved.

(1) The stacks are deposited upon the same substrate, that is, they are piled on top of each other. Since this type of multilayer is an ensemble of individual stacks, it will be referred to as an ensemble multilayer.

(2) Each stack is deposited on a different surface. This configuration will be referred to as a composite filter.

20.5.1.3.1 Ensemble of stacks. As an example of a multilayer which is an ensemble of two individual (modified) quarter-wave stacks, consider

$$\text{glass} \quad \left(\frac{L}{2} \; H \; \frac{L}{2}\right)^6 \quad \left(\frac{L'}{2} \; H' \; \frac{L'}{2}\right)^5 \quad \text{air}$$

where the index of the L land H layers is 1.38 and 2.30, respectively. Using our terminology, the group of layers $(L/2 \; H \; L/2)^6$ constitutes one stack and the group $(L'/2 \; H' \; L'/2)^5$, another stack. If the optical thickness of the L, H and L', H' layers is chosen in the ratio of 1.38: 1, then the high-reflectance zones to the two stacks are contiguous. The reflectivity versus $g = \lambda_o/\lambda$ of this combination is depicted in Figure 20.70. The dashed line shows the boundary of the high-reflectance zones of each stack. The attenuation region covers nearly an octave. In general, two difficulties are encountered in constructing multilayers of this type. First, interference effects often occur between the various stacks and the ensemble filter acts like a Fabry-Perot type filter (described in 20.10), to the extent that narrow transmission bands appear in the attenuation region. Second, such filters are often difficult to manufacture. Each of the films in the filter has a small amount of mechanical stress (refer to 20.2.4.2.4) and if the number of layers is large, the total stress can build up to the point where the multilayer will no longer adhere to the substrate. Multilayers of this type which have a high-reflectivity over a wide spectral region are often called broad-band reflectors. The spectral reflectivity of such a reflector is shown as curve B in Figure 20.95.

20.5.1.3.2 The composite filter. The composite filter consists of two or more individual multilayers which are deposited on separate substrates and arranged so that the incident radiant flux passes through each one. Because the multilayers are on separate substrates, it is possible to separate them physically by a distance of many thousands of wavelengths, thus avoiding interference effects between the various multilayers. If only two multilayers are used in a composite filter, it is often convenient to deposit each multilayer on a side of the substrate, as is shown in Figure 20.71. As an example of this procedure, suppose a filter is required which attenuates the entire visible spectral region from 400 mμ to 700 mμ. Since a single quarter-wave stack of zinc sulfide and magnesium fluoride has a high-reflectance zone which covers approximately one-half of this region, a composite filter composed to two quarter-wave stacks is requisite. Each stack is deposited on a side of the glass substrate, as is shown in Figure 20.71. The high-reflectance zone of multilayer stack "A" extends from λ_1 to λ_2 and the zone of stack "B" from λ_3 to λ_4. Figure 20.71 shows how the high-reflectance zones are arranged so they not only cover the spectral region to be attenuated but also overlap in the center so that there is no possibility of a transmission leak. Stack "A" has a QWOT = 453 mμ and stack "B" a QWOT = 594 mμ. These values are chosen so that region

of attenuation extends from 390 mμ to 706 mμ. Having determined the QWOT and materials of the two stacks, the only remaining problem is to choose the number of periods. This number is determined by the amount of attenuation required, the steepness of the cutoff required at 700 mμ, or by economic considerations. The latter point should be considered if the filter is to produce in large numbers; the cost of manufacture depends upon the number of layers in the stack. It is noted that the transmission beyond 700 mμ would be enhanced if eighth-wave layers were added to each stack, so that stacks "A" and "B" are similar to the multilayer shown in Figure 20.67. The addition of these eighth-wave layers shifts the wavelength of maximum reflectivity of each stack, but does not affect the position of the high-reflectance zones.

20.5.1.4 Since the multilayers which are composed of dielectric materials have a negligible amount of absorption, the transmission properties of these filters, when placed in tandem, are considerably different from absorption-type filters. As an example, suppose that an absorption-type filter consisting of a slab of colored glass has a transmission T of 0.01 at some wavelength, say 610 mμ. If an identical filter is placed in tandem, then the transmission T_2 of the entire system is reduced to the low value of

$$T_2 = T^2 = 0.0001 .$$

Next, consider the case of a multilayer filter which is composed of dielectric (non-absorbing) layers, so that $T = 1 - R$. For purposes of comparison, suppose that this multilayer is designed so that it also has a $T = .01$ at 610 mμ. If an identical multilayer is deposited on the opposite side of the same transparent substrate, as is shown in Figure 20.72, then the total transmission T_2 of the tandem arrangement is close to 0.005, or more precisely:

$$T = (1 - R)/(1 + R)$$

$$\simeq T/2 \text{ as } R \rightarrow 1.0 \qquad (48)$$

The reason for this behavior can be seen in Figure 20.72, which shows identical multilayer coatings C and C' deposited on each side of the transparent substrate. Since the substrate is quite thick, interference effects can be neglected and the intensity of the beams which emerge can be added. As is shown in Figure 20.72, a radiant flux proportional to $T = 1 - R$ penetrates into the medium (i.e. the substrate) between the two multilayers. If the reflectivity is close to 1.0, then this flux is trapped and bounces back and forth between the two multilayers many times, decreasing in intensity only a small amount at each reflection. Eventually all of the flux escapes through interfaces C and C'. If R is close to 1.0, then approximately one half of it escapes through C and one-half through C'. Since T_2 is proportional to the radiant flux which penetrates C', T_2 is decreased by only a factor of two. The total transmission T_2 is decreased below the value given in Equation (48) if there is a small amount of scattering in the multilayers or if there is absorption in either the multilayer or the substrate. T_2 can also be decreased by using a wedge-shaped substrates so that the rays "walk off" the ends of the filter.

20.5.2 Long-wave pass color filters. From the discussion in 20.5.1 and 20.4.8.2.1, it is evident that it is a straightforward task to produce filters for the visible spectral region which pass in the long-wave region. However, it should be remembered that in the visible region there are many absorption filters of colored glass or organic dyes which rival multilayers in the sharpness of their cutoff, high attenuation in the short-wave region and high transmission at long wavelengths. Not only are these absorption filters often less expensive than a multilayer filter, but they have the additional advantage that they are virtually free from the variation of the transmission with the angle of incidence (i.e. the angle shift) which is inherent in multilayer filters. However, multilayer filters have the advantage that they can be manufactured to any specification so that the cutoff can be positioned at any wavelength, whereas the cutoff of glass and dye absorption filters is only at wavelengths which nature has provided. As an example of a long-wave pass color filter, Figure 20.73 shows the measured spectral transmittance of a modified quarter-wave stack composed of fifteen layers of zinc sulfide and magnesium fluoride. Eighth-wave layers of zinc sulfide on either end of the stack increase the transmission in the long-wave region. The transmittance of this filter is similar to the computed curve shown in Figure 20.67.

20.5.3 The cold mirror. The cold mirror is a multilayer composed of dielectric materials which has a high-reflectivity in the visible spectral region. The reflectivity drops abruptly at 700 mμ so that a high transmission is achieved in the near infrared. The discussion in 20.1.2.7 describes briefly how this cold mirror can be placed behind a light source in a projection system so that the light in the visible spectral region is deflected towards the lens, but the heat, (the radiant energy in the infrared), passes out of the system. There are many designs for cold mirrors; they usually consist of an ensemble of stacks, as described in 20.5.1.3.1. An effective cold mirror has a smooth reflectivity curve in the visible region so that the color of the reflected light is not altered. The sharpness of the cutoff at 700 mμ and the amount of transmission in the infrared are also important. A cold mirror which is used in conjunction with an arc lamp should have a sharp cutoff, since a typical arc lamp has a larger portion of its radiant energy concentrated in the region between 700 mμ and 800 mμ than does a tungsten filament. Since 85% of the radiant energy of

a tungsten incandescent lamp operated at 3350°K is outside of the visible region, a combination of a cold mirror and a heat reflector could theoretically reduce the heat six-fold. In practice, a two-fold to three-fold reduction is achieved by using a cold mirror and heat reflector in place of an aluminum mirror. Figures 20.74 and 20.75 depict the spectral transmittance and reflectance of some cold mirrors which are manufactured commercially. Both Dimmick[37] and Turner[38,39a] describe the use of cold mirrors in projection and illumination systems. A basic U.S. patent on the cold mirror has been issued to Koch[39b]

20.5.4 Long-wave pass filters for the infrared. Multilayer filters are used extensively in the infrared as components of missiles which have a heat-seeking guidance mechanism, infrared surveillance systems, and in the spectrochemical analysis of organic vapors. In the infrared, a limited number of absorption-type filters are available and hence multilayer filters are widely used. Infrared multilayers can be composed of the same materials which are used in the visible, such as zinc sulfide and chiolite. However, the use of certain materials is restricted by the stress in the films (see 20.2.3.2.4). Films of germanium can be used at wavelengths longer than 1.3 μ and lead telluride beyond 3.9 μ. Both of these materials have high refractive index. As is illustrated in Figure 20.53, a stack which has a quite broad high-reflectance zone is obtained when either of these materials is used in conjunction with a material of low refractive index, such as chiolite or silicon monoxide. The spectral width of the attenuation region can be increased by using two or more stacks, combined either as a ensemble (20.5.1.3.1) or as a composite filter (see 20.5.1.3.2). Figures 20.76 to 20.81 show the spectral transmittance of some long-wave pass filters which are manufactured by several commercial firms. It is interesting that the cutoff region of the stack shown in Figure 20.79 is provided by a 3:1 stack, which has the advantage of having a lower angle shift than the quarter-wave stack (see Section 20.4.7). It might be conjectured that the spike in the transmission at 3.8μ of the multilayer shown in Figure 20.80 is due to the fact that the edge high-reflectance zone occurs at this wavelength but that the transmission remains low because the lead telluride films are absorbing in this region. Using the refractive indices in Table II, from Equation (39) it is found that the ratio $\lambda_2/\lambda_1 = 1.66$, where λ_2 and λ_1 are the wavelengths at the long and short wave edge of the high-reflectance zone. From Figure 20.82 we see that $\lambda_2 = 6.5\ \mu$, and thus $\lambda_1 = 3.9\ \mu$; this conjecture is probably correct. The transmission peaks at shorter wavelengths in Figures 20.80 and 20.81 could be eliminated by depositing another long-wave pass multilayer on the opposite side of the substrate, thus forming a composite filter. The transmission in the long-wave region in both of these filters could be improved if an antireflection coating were deposited on the opposite side of the arsenic trisulfide glass substrate.

20.6 SHORT-WAVE PASS FILTERS

In Section 20.5.1, we enumerated some of the properties of long-wave pass filters, such as the attenuation, sharpness of the cutoff, angle shift, and so on. These general considerations apply equally well to short-wave pass filters, with the following important exception: The spectral width of the high-transmission region of a quarter-wave stack or other stack with a periodic structure is limited by the fact that at shorter wavelengths higher order high-reflectance zones always occur. For example, suppose that the 3:1 stack shown in Figure 20.59 is used to attenuate in the 1.8 μ to 2.2 μ region. Additional periods can be added to the stack if a higher attenuation in this region is required. The pass region on the short-wave side extends from 1.15 to 1.7 μ. At 1.15 μ the second-order high-reflectance peak occurs and the multilayer attenuates from 0.9 to 1.15 μ. It is evident from Figures 20.56, 20.57, and 20.58 that these higher order attenuation regions are not confined to the 3:1 stack, but occur in all stacks which have a periodic structure, regardless of the composition of the basic period. A comparison of Figures 20.56, 20.57, and 20.58 shows that the quarter-wave stack has the widest region of high transmission on the short-wave (high-frequency) side of the first order high-reflectance zone. The third-order high-reflectance zone occurs at one-third of the wavelength of the first order zone. The spectral width of high transmission of the 2:1 stack is somewhat smaller; the second-order high-reflectance zone occurs at one-half the wavelength of the first order zone. However, we have considered only the case where two layers are used in the basic period of the stack. If more than two layers are used in the basic period of a stack with a periodic structure, then a short-wave pass spectral region which is even wider than that of a quarter-wave stack can be achieved.[40] Regardless of whether a quarter-wave, 2:1 or 3:1 stack is used, it is desirable to add some additional layers to the stack to enhance the transmission in the short-wave region, as is shown in Figure 20.66.

20.6.1 Short-wave pass color filters. Multilayer filters are particularly useful as short-wave pass filters in the visible spectral region, because most of the colored glass and dyed gelatin filters have a low transmission in the pass region and a cutoff which is not sharp. For example, a typical blue glass absorption filter which transmits below 480 mμ has a peak transmission of 0.38 at 410 mμ and a bell-shaped transmission curve. Also, most absorption filters which transmit the blue and near ultraviolet have a leak- a transmission band in the red. Figures 20.66, 20.68, and 20.82 show the spectral transmission of some short-wave pass filters. The multilayer shown in Figure 20.82 is a quarter-wave stack with films added to enhance the transmission in the short-wave region, as described in 20.4.8.3.1.

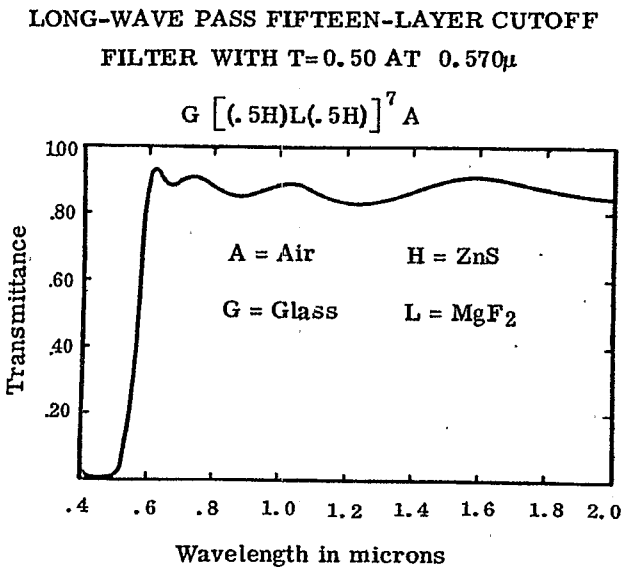


Figure 20.73- Measured spectral transmittance of a long-wave pass filter. Courtesy of Bausch and Lomb, Inc.

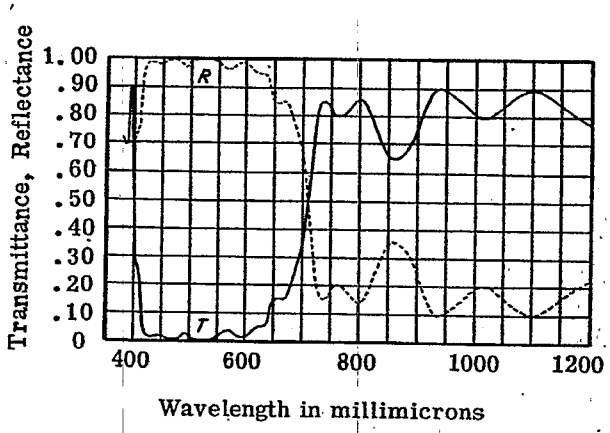


Figure 20.74- Measured reflectance and transmittance of a cold mirror multilayer coating. Courtesy of Balzers Aktiengesellschaft, Liechtenstein

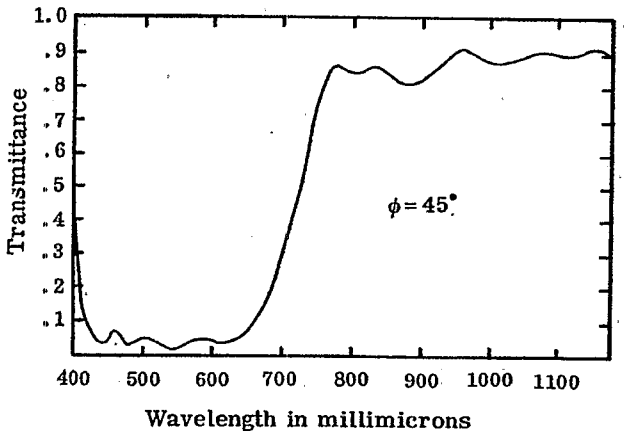


Figure 20.75- Measured spectral transmittance at $\phi = 45°$ of a cold mirror of unspecified design. Courtesy of Fish-Schurman, Corporation.

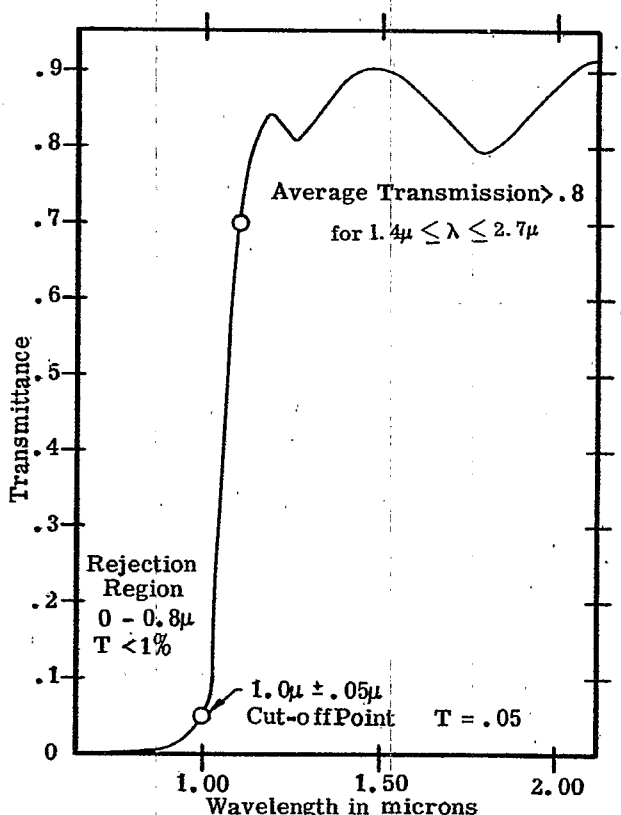


Figure 20.76- Measured spectral transmittance of a long-wave pass filter of unspecified design. Courtesy of Eastman Kodak Company.

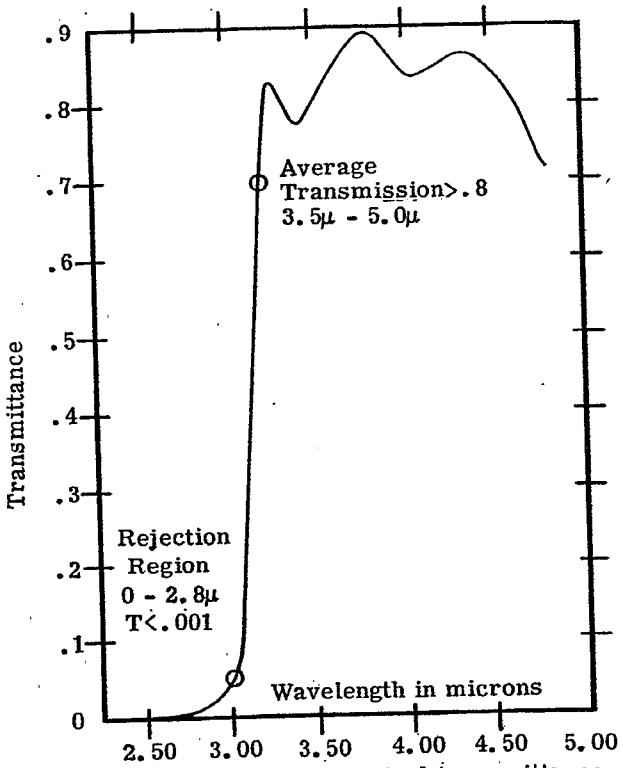


Figure 20.77- Measured spectral transmittance of a long-wave pass filter of unspecified design. Courtesy of Eastman Kodak Company.

MEASURED SPECTROPHOTOMETRIC TRANSMITTANCE OF A LONG-WAVE PASS ELEVEN-LAYER CUTOFF FILTER

G $[(.5H)L(.5H)]^5$ A

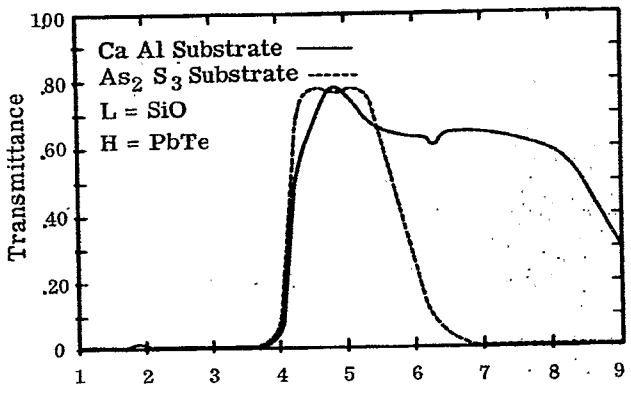


Figure 20.78- Measured spectral transmittance of a modified quarter-wave stack on two different types of substrate, namely As_2S_3 glass (solid curve) and calcium aluminate (dotted curve). Courtesy of Bausch and Lomb, Inc.

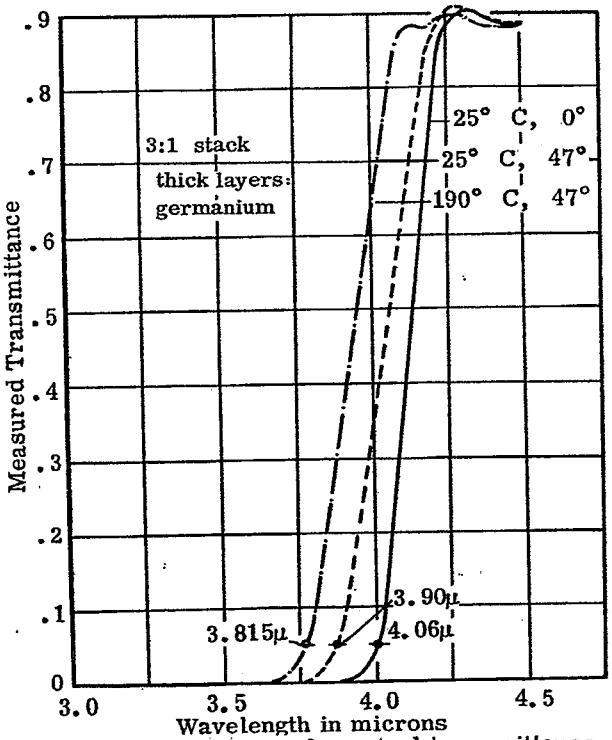
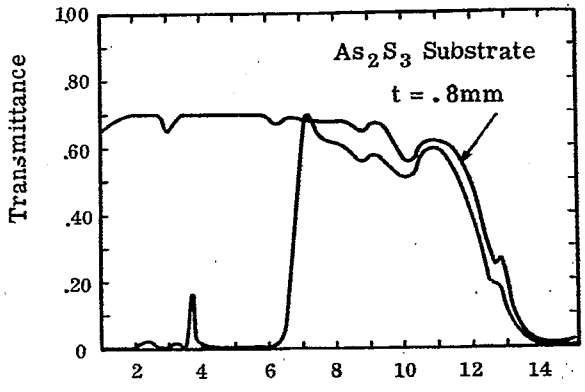


Figure 20.79- Measured spectral transmittance of a long-wave pass filter at ϕ = O and ϕ = 45° and at low temperature. The design is a modified 3:1 stack. Courtesy of Optical Coating Laboratory, Inc.

LONG-WAVE PASS NINE-LAYER CUTOFF FILTER AT 6.5 MICRONS

G $[(.5H)L(.5H)]^4$ A



A = Air L = ZnS
G = As_2S_3 Glass H = PbTe

Figure 20.80- Measured spectral transmittance of a long-wave pass filter on an As_2S_3 glass substrate. The transmission of both the filter and the bare substrate would be improved if the substrate were antireflection coated. Courtesy of Bausch and Lomb, Inc.

20.6.2 Heat reflectors. Section 20.1.2.7 describes briefly how a multilayer heat reflector deflects the heat from a projection lamp away from the film gate. For many years heat absorbing glass has been used to absorb the radiant energy in the infrared, but this has the disadvantage that the heat must be removed from the glass by air cooling or other means. If this is not done, the glass becomes quite hot and often fractures. Multilayer heat reflectors have the advantage that they reflect, rather than absorb, the near infrared and thus remain comparatively cool. Figures 20.83, 20.84, 20.85, 20.86, and 20.87 show the spectral transmission of some infrared reflectors. The multilayer in Figure 20.83 is identical to that of Figure 20.82, with the exception that the QWOT of the former has been moved to longer wavelengths. The multilayer shown in Figure 20.84 attenuates far into the infrared, but unfortunately it also attenuates in part of the red spectral region. The width of the attenuation region in the infrared of the multilayer shown in Figure 20.85 is typical of what can be attained with a simple quarter-wave stack. The multilayer shown in Figure 20.86 not only reflects the infrared, but also has a higher order high-reflectance zone in the blue. Although the design of the multilayer is unspecified, the fact that the maximum of the red reflection band occurs at twice the wavelength of the blue reflection band leads one to suspect that this is a 2:1 stack. It should be pointed out that the heat reflector does not have to be used in conjunction with a cold mirror, as shown in Figure 20.82 Some advantage is gained if only a heat reflector is inserted into the beam at an angle.

20.6.2.1 Another type of heat reflector is used in conjunction with the solar energy cells which are used in satellites and space vehicles. The problem is that only the radiant energy in the range from 0.4 to 1.2 μ produces appreciable electrical energy. The radiant energy outside of this region merely heats the cell, thereby decreasing its efficiency. Therefore, solar cells used in space vehicles are usually protected with a short-wave pass filter which has a high-reflectivity from 1.2 μ out to longer wavelengths. The advantages of using these filters are discussed by Thelen.[41] It is remarked that heat reflecting mirrors are often called hot mirrors. It should be remembered that both the cold mirror (described in Section 20.5.3) and the hot mirror are made of non-absorbing materials. Thus neither of them absorbs an appreciable amount of radiant energy and thus they both remain comparatively cool.

20.6.3 Short-wave pass filters for the infrared. One difficulty which is encountered in using a multilayer as a short-wave pass filter for the infrared is that the spectral width of the pass region at short wavelengths is narrow, especially if high-index materials such as silicon, germanium, or lead telluride are used in the filter. In the latter case, the attenuation region is quite wide, and this width is at the expense of the width of the pass region. One way to avoid this difficulty is to use materials in the stack which have smaller ratio of refractive index, n_a/n_b. Another approach is to use a stack which has a basic period which contains more than two refractive indices.[40]

20.7 BEAM SPLITTERS

20.7.1 A beam splitter is used to divide a wavefront into two portions and direct each portion in a different direction. Beam splitters can be arbitrarily divided into two classifications:

(1) Achromatic, or neutral, beam splitters (see Section 20.7.2).

(2) Color selective beam splitters. These are sometimes called dichroic mirrors and are discussed in Section 20.7.3.

20.7.1.1 The properties of a beam splitter are usually specified by the average transmission, T_{av}, and average reflectivity, R_{av}, (defined in 20.1.3.7). However, as is shown in 20.7.1.2, it is often important to know T_s, T_p and R_s, R_p in the two planes of polarization so that the amount of polarization produced by the beam splitter can be determined. If the beam splitter is used as a component of an interferometer, such as a Michelson or Twyman-Green, then the variation with wavelength of the phase shift upon reflection should be considered, since the position of the fringes in the interferometer depends upon this phase shift. The phase shift of a silver film is shown in Figure 21.17.

20.7.1.2 Polarization effects. If T_s and T_p are unequal, or if $R_s \neq R_p$, then the beam splitter has a polarizing effect, which can be quite important. For example, a beam splitter which is used in a Michelson interferometer should divide the beam equally in order to produce fringes with maximum intensity. Thus a beam splitter composed of dielectric films which has an R_{av} of .50 might look attractive as a beam splitter for this purpose. However, it is important to know the R and T in each plane of polarization, since the beams interfere separately in each plane of polarization. For example, if a beam splitter for a Michelson interferometer has $R_{av} = 0.50$, but $R_s = 0.10$ and $R_p = 0.90$, then the fringe intensity, which is proportional to the product $R \cdot T$, is quite low. The dissimilarity of the reflectivity in the two planes of polarization can also affect other devices. For example, suppose a beam splitter in a glass cube is mounted behind the objective lens of a camera, as is shown in Figure 20.88. Such a beam splitter is manufactured by evaporating the multilayer coating onto the hypotenuse face of a 45-90-45 glass prism, and then cementing an identical prism onto it. The purpose of the beam splitter is to reflect 20% of the light into

MEASURED TRANSMITTANCE OF AN ELEVEN LAYER LONG-WAVE PASS FILTER ON As_2S_3 GLASS SUBSTRATE

G $\left[(.5H)L(.5H)\right]^5$ A

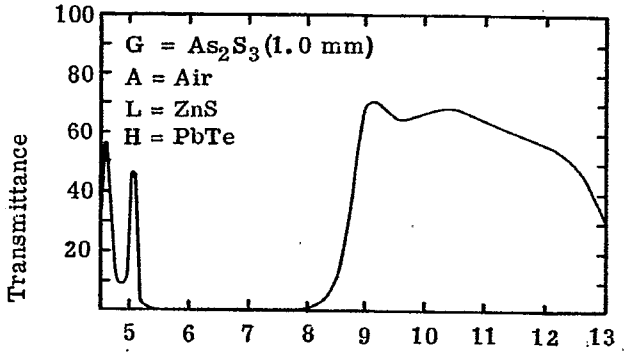


Figure 20.81- Measured spectral transmittance of a long-wave pass filter deposited on an As_2S_3 glass substrate. The transmission beyond 8 μ would improve if the "back" side of the substrate were antireflection coated. Courtesy of Bausch and Lomb, Inc.

SHORT-WAVE PASS THIRTEEN-LAYER CUTOFF FILTER WITH T=.5 AT .572μ

G (.78L) $\left[(.5L)H(.5L)\right]^6$ A

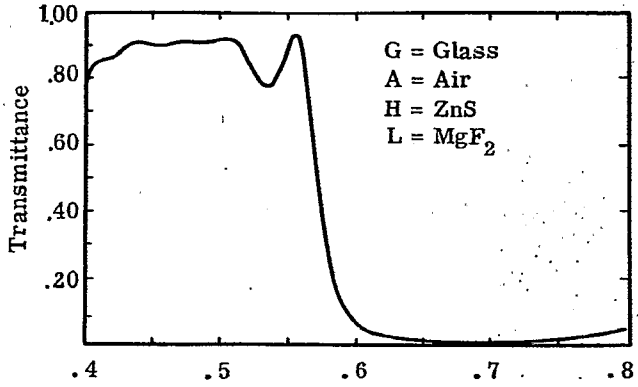


Figure 20.82- Measured spectral transmittance of a short-wave pass filter. Courtesy of Bausch and Lomb, Inc.

SHORT-WAVE PASS THIRTEEN-LAYER CUTOFF FILTER WITH T=.5 AT .80μ

G (.78L) $\left[(.5L)H(.5L)\right]^6$ A

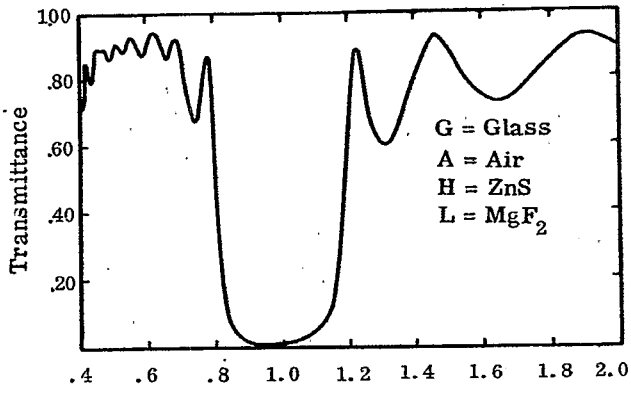


Figure 20.83- Measured spectral transmittance of a short-wave pass heat reflecting filter. Courtesy of Bausch and Lomb, Inc.

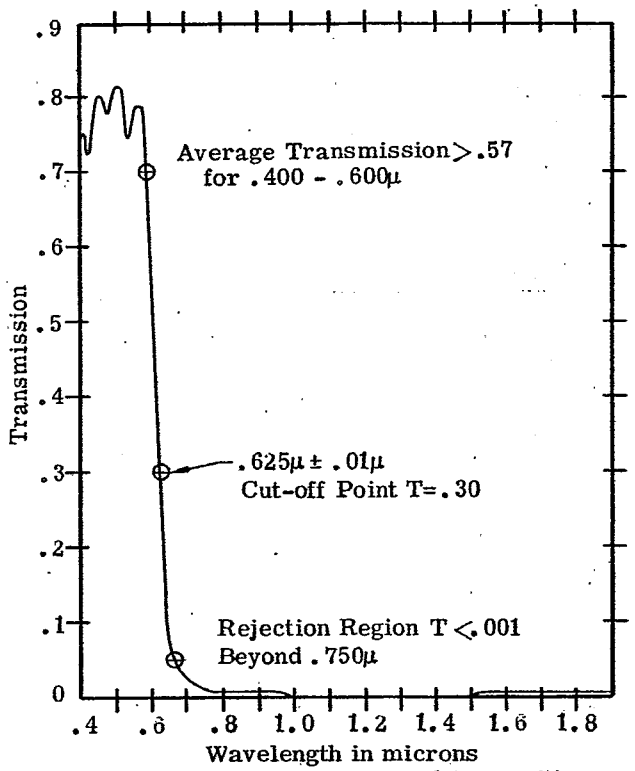


Figure 20.84- Measured spectral transmittance of a short-wave pass filter. Courtesy of Eastman Kodak Co.

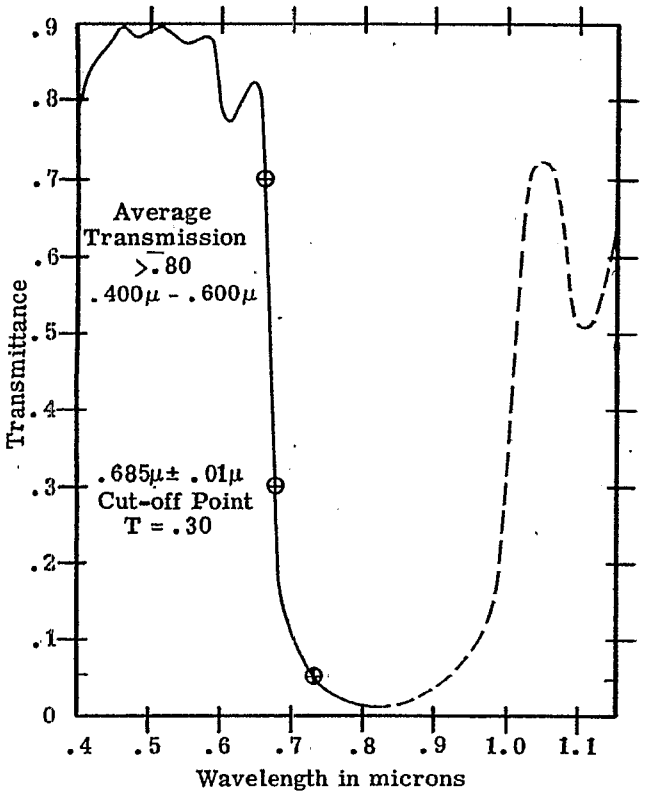


Figure 20.85- Measured spectral transmittance of a short-wave pass heat reflecting filter. Courtesy of Eastman Kodak Company.

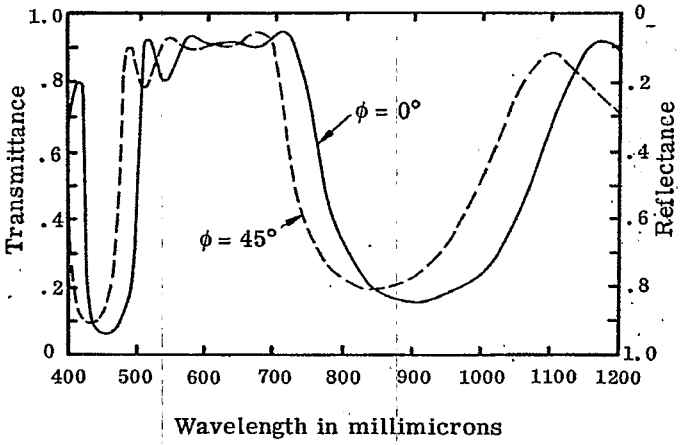


Figure 20.86- Measured spectral transmittance multilayer which reflects the blue and near infrared. Courtesy of Fish-Schurman, Corporation.

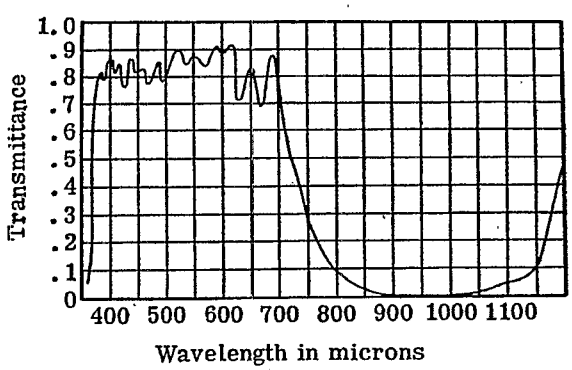


Figure 20.87- Measured spectral transmittance of a multilayer heat reflecting filter. Courtesy of Balzers Aktiengesellschaft, Liechtenstein.

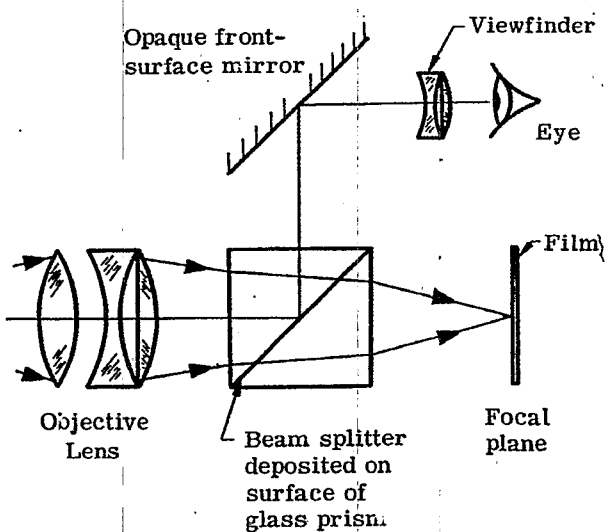


Figure 20.88- The utilization of a beam splitter in a camera to deflect an image into the view finder.

the view finder. As long as the light from the scene which is being photographed is unpolarized, then any beam splitter which has the specified $T_{av} = 0.80$ and $R_{av} = 0.20$ is satisfactory. However, suppose that $R_s = 0.08$ and $R_p = 0.32$ so that the beam splitter is polarizing. If a scene is photographed which produces polarized light, such as certain portions of the sky or the light which is reflected from a wet surface, then the brightness of the scene which is seen in the view finder and the scene which is recorded on the film are quite different. Another effect is that the brightness in the view finder changes as the camera is rotated about the axis of symmetry of the objective lens.

20.7.2 Achromatic beam splitters. An achromatic, or neutral, beam splitter should reflect and transmit equally at all wavelengths. In practice, the transmission of most beam splitters changes slightly with wavelength.

20.7.2.1 Silver films. Thin films of silver are widely used as beam splitters. For many years only chemical methods were used for depositing the films, but more recently superior quality silver films have been produced by evaporation in a vacuum. The optical constants, n and k, vary considerably with wavelength and consequently the R and T of a film of a given thickness also change. The R and T of silver films of varying thickness at a wavelength of 550 mμ is shown in Figure 21.16. Due to the dispersion of the optical constants, silver films which have a given value of t/λ have a much lower absorption in the red than in the blue. Silver films have the advantage that they are easy to prepare. The disadvantages are that they are less efficient than a dielectric beam splitter because they absorb part of the light and that they deteriorate after they are removed from the evaporator. A typical rule of thumb is that silver beam splitter which is designed to divide the light equally reflects 1/3, transmits 1/3 and absorbs 1/3. Sennett and Scott[42] have measured the transmission and reflectivity of some silver films. In the visible spectral region metals other than silver have a high absorption and are generally not used as beam splitters.

20.7.2.2 Dielectric films.

The simplest type of dielectric beam splitter is a single film of zinc sulfide or titanium dioxide deposited on a glass substrate. These beam splitters are neutral in color and divide the light in a ratio of $R : T = 0.4 : 0.6$. Figure 20.89 shows the spectral reflectivity and transmission at $\phi = 45°$ of a single film of TiO_2 on a glass substrate. The dispersion of the refractive index of the TiO_2 has been taken into account in this calculation. The curves in Figure 20.89 give some idea of the flatness of the reflectivity curve and also the amount of polarization which is produced by the beam splitter. Holland[43] describes in detail how beam splitters of this type are prepared. Figure 20.90 shows the spectral reflectivity and transmission at $\phi = 45°$ of two types of beam splitters which are produced commercially. It should be remembered that the uncoated side of the substrate of a beam splitter introduces a transmission loss and also some polarization into the beam. This effect is not large and is reduced even more if the opposite side of the substrate is covered with an antireflection coating.

20.7.3 Color selective beam splitters (dichroic mirrors). Color selective beam splitters are sometimes called dichroic mirrors. One of the definitions of the word dichroism refers to the selective absorption and transmission of light as a function of wavelength regardless of the plane of vibration.

20.7.3.1 It could be argued that section on dichroic mirrors logically belongs in Sections 20.5 and 20.6, which covered long-wave and short-wave pass filters. Actually, this is true, because any of the multilayers described in those chapters could be used a dichroic mirror simply by tilting them at an angle. The difference is that when these multilayers were used as pass filters, we were concerned only with the transmitted light and gave little attention to what happened to the reflected light. If the multilayer is used as a beam splitter, then the reflected light is utilized. Color selective beam splitters are used extensively in optical systems where different kinds of information are identified by a different color and combined so that they are projected simultaneously. Many thousands of such beam splitters are used in radar sets to superimpose the image of a cathode ray oscilloscope screen (which might be green) with a map of another color (magenta, for example). It is easy to see an additional color selective beam splitter could be added to the configuration shown in Figure 20.1 to form a three-color separation system which could be used in color photography or color television.[44] Figure 20.91 shows a measured transmission curve of a blue reflector, which is similar to the long-wave pass filter shown in Figure 20.73, but viewed at non-normal incidence. The transmission curves at various angles of incidence also illustrate the angle shift to shorter wavelengths as ϕ increases. Figure 20.92 shows the reflectivity of green reflector at 45° incidence. One method of achieving this narrow reflection band is to use a $3\lambda/4$ stack. This is illustrated in Figure 20.93, which shows the computed transmission of nine layers of zinc sulfide and magnesium fluoride which have an optical thickness of $3\lambda_o/4$ where $\lambda_o = 546$ mμ, matched at 60°. A first order reflectivity peak occurs at 3×546 mμ $= 1.64$ μ in the infrared. The third order high-reflectance zone is centered at 546 mμ. The use of high-order reflectivity peaks is an effective method of achieving a narrow reflection band.

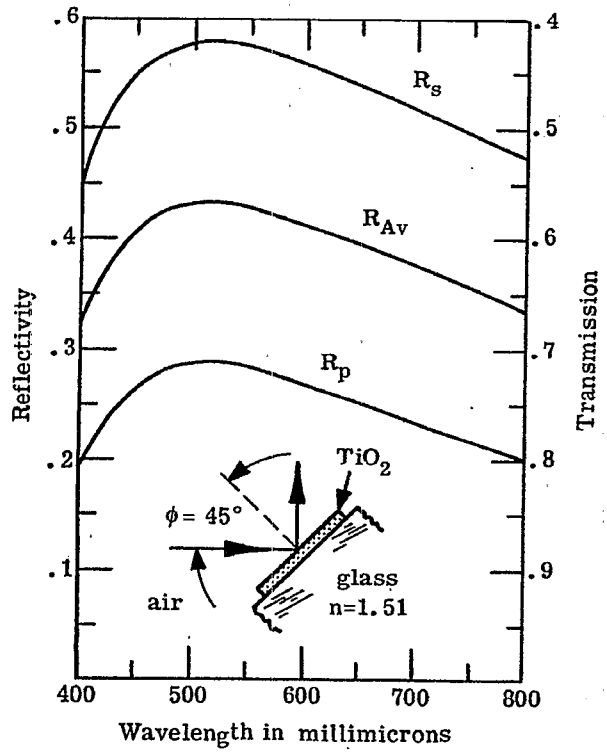


Figure 20.89- Computed R_p, R_s, and $R_{av} = 1/2 (R_p + R_s)$ of a single quarter-wave layer $Ti O_2$ at $\phi = 45$. The dispersion of the index of the $Ti O_2$ is taken into account.

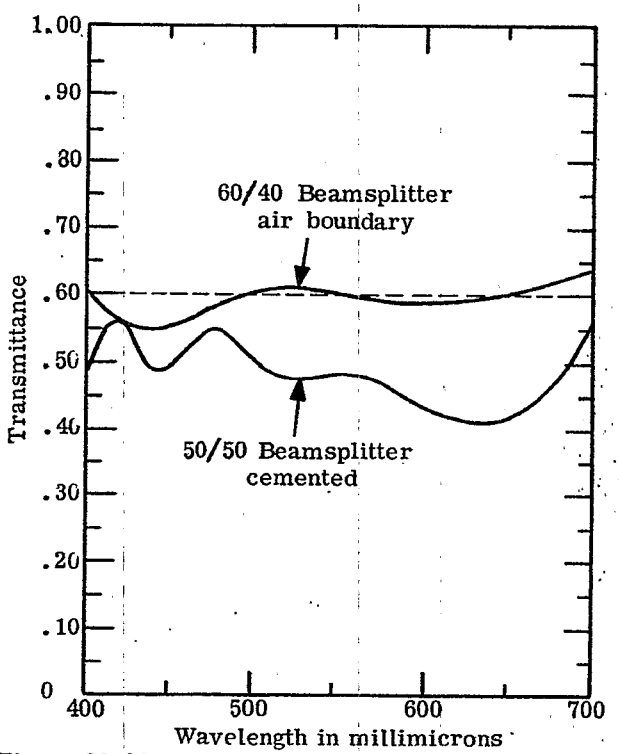


Figure 20.90- Measured transmittance of a 60/40 and a 50/50 multilayer beam splitter at $\phi = 45°$. R is nearly 1 - T. Courtesy of Fish-Schurman Corporation.

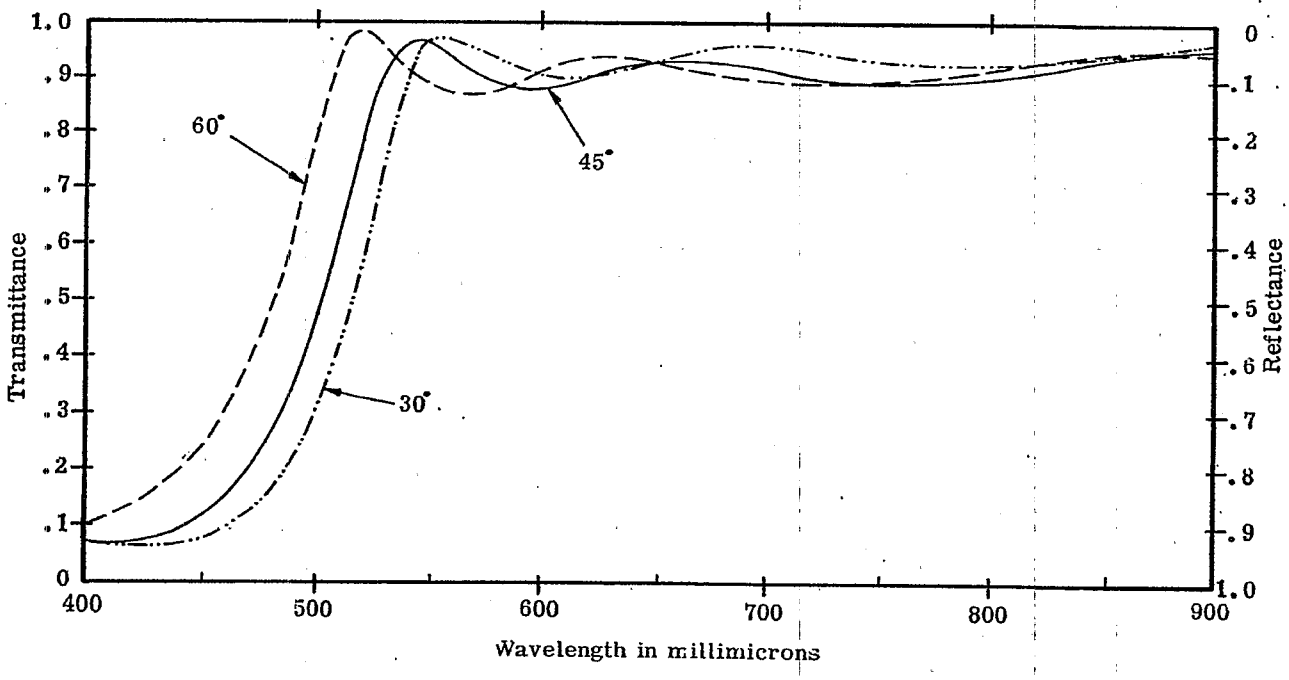


Figure 20.91- Measured spectral transmittance at various ϕ of a color-selective beam splitter. Courtesy of Fish-Schurman Corporation.

20.8 MIRRORS

The distinction between a mirror and a beam splitter is indeed tenuous. Some of the semi-transparent mirrors described in Section 20.8.2 could be used as color selective beam splitters if they were illuminated at non-normal incidence. Also, some of the long-wave and short-wave pass filters described in Sections 20.5 and 20.6 could be equally well used as semi-transparent mirrors. Thus the classification of a multilayer as a long-pass filter, color selective beam splitter, or a semi-transparent mirror depends upon its ultimate use and is not an inherent property of the device itself. Mirrors can be classified as either opaque or semi-transparent. Opaque mirrors are used if one is only concerned with the reflected light, whereas a portion of the light which is not reflected from a semi-transparent mirror, is transmitted.

20.8.1 Opaque mirrors. Silver, aluminum, and rhodium are commonly used as reflecting films in the visible spectral region. All three of these metals can be evaporated in a vacuum. An opaque coating is produced by a film 0.2 μ (eight millionths of an inch) thick. A freshly deposited film of silver has the highest reflectivity, particularly in the red, but the reflectivity deteriorates rapidly in air. An extremely thin, tough film of aluminum oxide forms naturally on the surface of an aluminum coating which protects it against further oxidation. Thus aluminum coatings last a long time under normal environmental conditions. The reflectivity of a rhodium coating is about 10% below that of aluminum. Rhodium is quite inert to attack from salt water and notwithstanding its lower reflectivity it is used in applications where the environmental conditions are severe. Gold films have a high reflectivity in the infrared. Hass [45] has published reflectivity data on the various metal coatings.

20.8.1.2 Protective coatings. A single layer of silicon monoxide(SiO) is often deposited on an aluminum film to protect it from abrasion and chemical attack. The addition of this layer does not alter the reflectivity of the aluminum if the optical thickness of the layer is either much smaller than the wavelength of the incident light or if its optical thickness is a half-wave, in which case it is absentee (see Section 20.1.5.2.2). The former condition can easily be obtained at long wavelengths in the infrared. Figure 21.14 shows the computed spectral reflectivity of aluminum overcoated with a single layer of silicon monoxide. At certain wavelengths the reflectivity of the aluminum is decreased from 0.90 to 0.78 by the addition of the protective overcoat.

20.8.1.3 Reflection enhancing overcoatings. The purpose of adding the single-layer overcoat described in the foregoing paragraph is to increase the resistance of the mirror to abrasion and chemical attack. It is also possible to overcoat a metal mirror with a multilayer coating in order to increase the reflectivity to values as high as .995. Such mirrors are useful in optical systems in which the light is reflected many times. Figure 20.94 shows the computed reflectivity of a bare aluminum mirror. It should be noted that the scale of the ordinate changes at 0.90. The dispersion of the optical constants is taken into account in the calculation; the reflectivity compares well with the published data of Hass [45] . On the same graph is shown the reflectivity of an aluminum mirror overcoated with a six-layer dielectric stack of magnesium fluoride and zinc sulfide. The dispersion of the refractive index of the latter material has also been included in the calculation. The computed reflectivity attains a maximum value of 0.996. Jenkins [46] measured accurately the reflectivity of such overcoated mirrors and found a maximum reflectivity of 0.994. Other types of overcoatings can be used to obtain a broader region of high reflectivity [46] , so that the reflectivity does not decrease in the blue, as it does in Figure 20.94.

20.8.2 Semi-transparent mirrors. The mirrors which are discussed in Section 20.8.1 are opaque - that is, the light which is not reflected is absorbed in the metal coating. There are some applications where mirrors are required which not only have a high reflectivity, but also transmit with a high efficiency the light which is not reflected. Such multilayers are useful as coatings for the plates of the Fabry-Perot interferometer and also as coatings for the ends of an optical maser (sometimes called a Laser).

20.8.2.1 Silver films. Silver films have been used for more than six decades as coatings for the Fabry-Perot interferometer. They have the advantage that the single film of silver can be deposited much more quickly and easily than the many films of a multilayer mirror. Kuhn and Wilson [47] measured carefully the R and T of layers of silver on a glass substrate and found that a freshly evaporated film has small absorption loss, particularly in the red spectral region. At a wavelength of λ = 680 mμ, typical values were R = 0.89, T = 0.08, and the remaining 3% is absorbed. However, in the blue (420 mμ) film with same transmission would absorb more than twice as much. The disadvantage of using silver films is that the reflectivity deteriorates in time. Multilayer coatings have the advantage that they have a lower amount of absorption and that a higher reflectivity can be attained.

20.8.2.2 Dielectric multilayers. Quarter-wave stacks and other types of multilayer coatings have been used quite successfully as semi-transparent mirrors, [46, 48, 49] principally for coating Fabry-Perot interferometers. Figure 20.95 shows the first-order high-reflectance peak of quarter-wave stacks consisting of five, seven, and nine layers of zinc sulfide and cryolite. For a given plate separation, the maximum resolution of a Fabry-Perot is limited by the flatness of the interferometer plates and increasing the reflec-

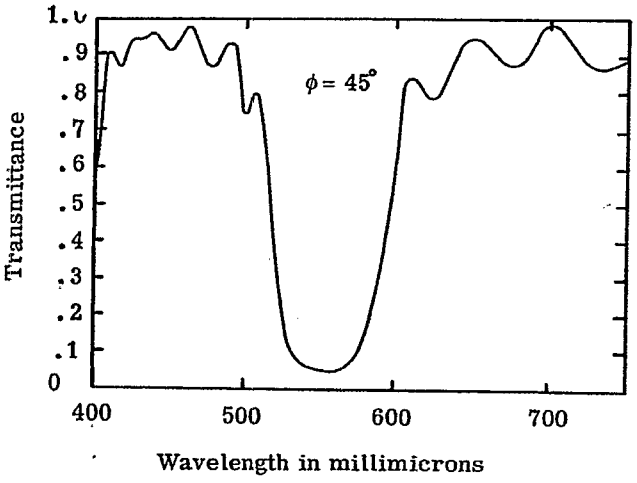


Figure 20.92- Measured spectral transmittance at $\phi = 45°$ of a color selective beam splitter which reflects the green. Courtesy of Fish-Schurman, Inc.

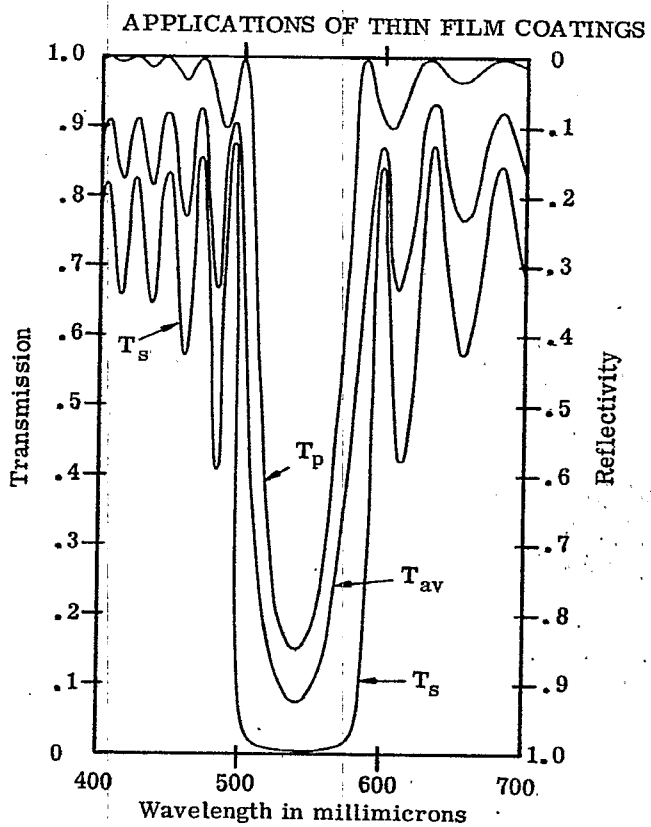


Figure 20.93- Computed T_p, T_s, and $T_{av} = 1/2 (T_p + T_s)$ of a color-selective beam splitter at $\phi = 65°$. The design is glass $H (LH)^4$ air, $n_L = 1.35$, $n_H = 2.30$, $n_L t_L = n_H t_H = \frac{3\lambda_o}{4}$ at $\phi = 65°$, for $\lambda_o = 550\ m\mu$.

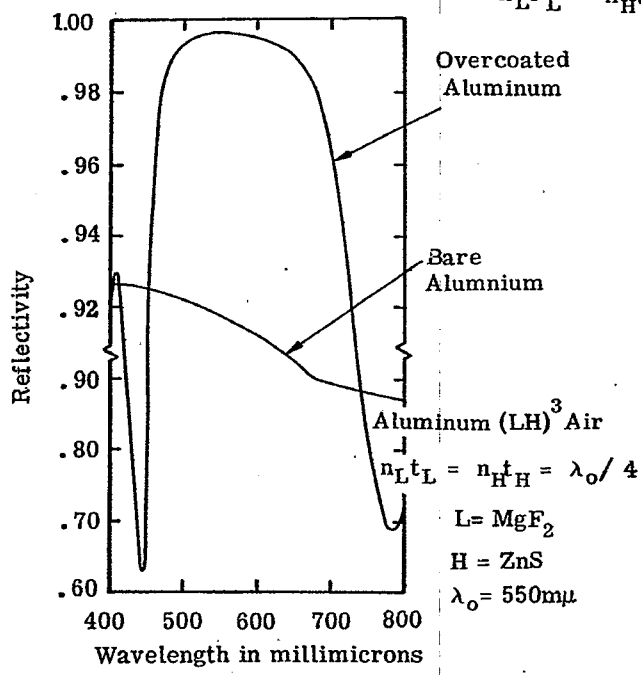


Figure 20.94- Computed spectral reflectivity of bare aluminum and aluminum overcoated with six dielectric layers. The scale of the ordinate changes at 0.90. The disperson of the optical constants is accounted for.

tivity over a certain value merely results in a loss of light without any gain in resolving power. Hence the seven-layer films are adequate for most plates. The advantage of using dielectric coatings over silver coatings is that the absorption loss of the former is quite low. A typical multilayer dielectric stack has an absorption of 0.005.[50] The disadvantage of using quarter-wave stack is readily apparent from Figures 20.71 or 20.95; the region of high reflectivity covers only about one-half of the visible spectrum. This difficulty is avoided if a broad-band reflector [49] is used, such as the one whose spectral reflectivity is shown as curve "B" in Figure 20.95. The design of this reflector is given in reference 49. The reflectance of this multilayer is close to 0.95 in the visible spectrum, which is the optimum value for most interferometer plates. The reflectivity of the cold mirror shown in Figure 20.74 is too high and would not be suitable for coating most interferometer plates.

20.8.2.3 From a theoretical point of view, (i.e. Equations (40) - (43)) the reflectivity of a quarter-wave stack can be made as close to 1.0 as we choose by simply using a sufficient number of layers in the stack. At the present state of thin film technology, it is found that the reflectivity of a $ZnS - MgF_2$ stack does not increase appreciably beyond fifteen layers. The fact which limits the ultimate reflectivity is that the layers are composed of small crystallites which are randomly oriented. This means that the layers have random fluctuations in their optical thickness and also that there are irregularities at the surfaces of the layers which scatter light. Giacomo [51, 52] has studied this problem in detail. As is discussed in 20.10.6.1, it is this scattering which limits the band width of dielectric interference filters. It is possible that in the future techniques will be developed to reduce this scattering, thereby making it possible to increase the reflectivity to values very close to 1.0 and consequently to produce multilayer interference filters with extremely narrow band widths.

20.9 BAND PASS FILTERS

A band pass filter is a multilayer which has a high transmission in a specific spectral region and attenuates in both the short and long wave regions on either side of the pass band. In a certain sense, all of the short-wave pass filters described in 20.6 are band pass filters, because the higher order high-reflectivity peaks limit the width of the short-pass region, although this is not the intent of such a filter. Included in the class of band pass filters are Fabry-Perot type filters or interference filters. This filter, which is discussed in 20.10, was the first type of filter whose transmission characteristics depended upon the interference of light, rather than the absorption of light. The name interference filter is retained for primarily historical reasons. Strictly speaking, every multilayer filter, from the single-layer coating to a sixty-layer stack, is an interference filter, in the sense that its transmission characteristics depends upon the interference of the light reflected from various layers within the filter. For reasons described in 20.10, we prefer to call this type of filter a Fabry-Perot type filter.

20.9.1 Filters with a wide pass band. From the discussion of long-wave and short-wave pass filters in 20.5 and 20.6, it is evident that a band pass filter can be constructed simply by combining a long-pass and short-pass filter. They can be either deposited on the same substrate or they can be deposited on separate substrates as a composite filter. The advantage of constructing a band pass filter in this manner is that width and position of the pass band can be changed at will by altering the cutoff of either the short or long-wave stack. Also, the amount of attenuation outside of the pass region can be easily controlled by changing the number of layers in either of the stacks. The width of the pass band is essentially independent of the attenuation outside the pass region. This is not true of the Fabry-Perot type filter described in 20.10. There are also other possibilities, such as combining a short-wave pass multilayer filter with a glass or dye absorption type filter which passes in the long-wave region. Figure 20.95 shows the spectral transmission of some assorted filters with a broad pass band in the visible and near ultraviolet spectral regions, while Figure 20.96 shows some similar filters for the infrared.

20.10 FABRY-PEROT TYPE FILTERS (INTERFERENCE FILTERS)

20.10.1 Basic concepts - the Fabry-Perot interferometer. In order to explain how an interference filter functions, it is first useful to understand how a Fabry-Perot interferometer works and to become familiar with such concepts as the "Q", free spectral range, order number, and so on. The basic idea of the interference filter has been known since the turn of the century, when two French physicists, C. Fabry and A. Perot, invented the interferometer which bears their name. Although this device is used mostly to measure fine details of emission line spectra, it can also be used as a filter with a very narrow pass band. The Fabry-Perot interferometer, which is shown in Figure 20.98, consists of two plates of glass or fused quartz which have been polished to a high degree of flatness. The plates are held apart by three separator pins so that the inner faces of the plates are parallel. These faces are coated with a semi-transparent mirror, such as the type described in 20.8.2. The transmission T_f of the interferometer is derived in most books on physical optics [53, 54] and in Section 5.17.

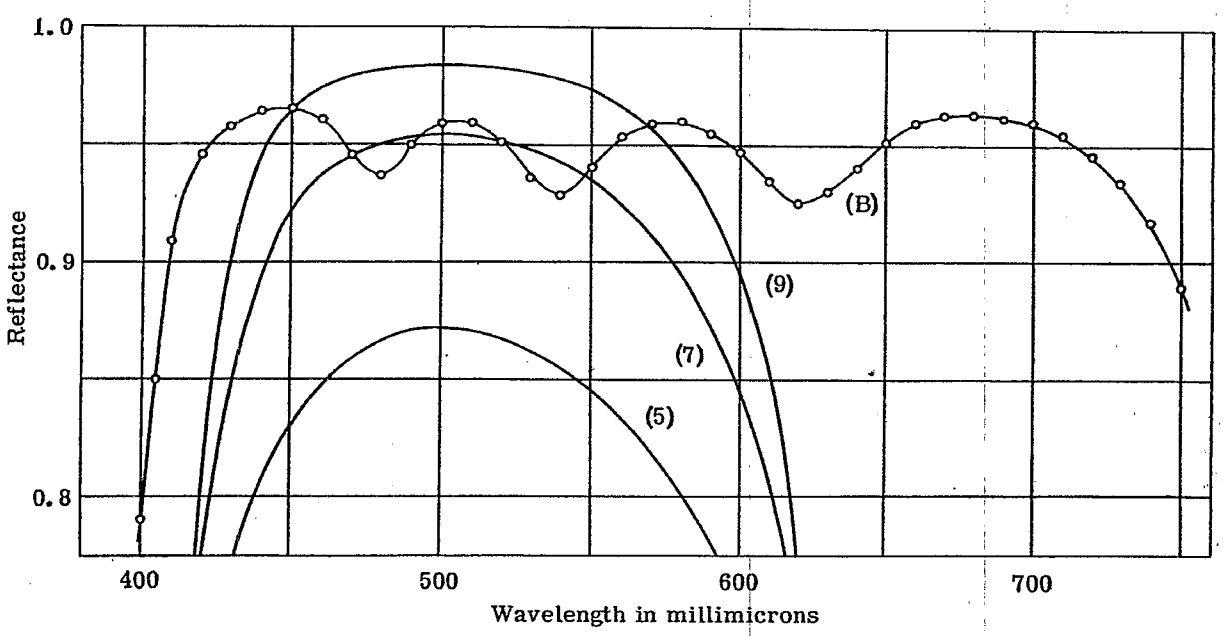


Figure 20.95- Measured spectral reflectivity of quarter-wave stacks with 5, 7, and 9 layers and also a broad-band reflector consisting of 15 layers of ZnS and cryolite. From ref. 49.

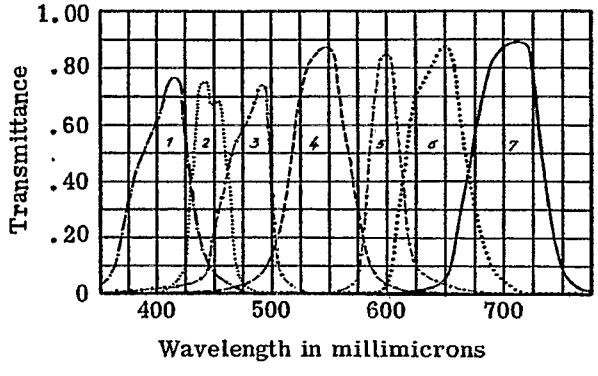


Figure 20.96- Measured transmittance of band pass filters which utilize both multilayer and glass absorption filters. Courtesy of Balzers Aktiengesellschaft, Liechtenstein.

MEASURED TRANSMITTANCE OF A SERIES OF BAND PASS FILTERS COMPLETE WITH AUXILIARIES

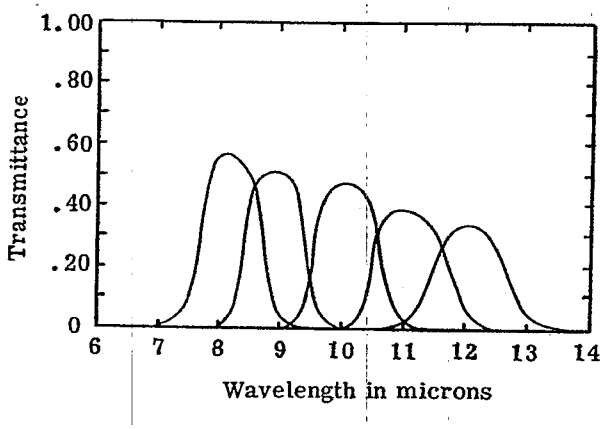


Figure 20.97- Measured spectral transmittance of some infrared band pass filters, complete with auxilliaries. Courtesy of Bausch and Lomb, Inc.

$$T_f = T_{max} \ (1 + F \sin^2 \eta)^{-1} \tag{49}$$

where

$$T_{max} = \frac{T_1 \ T_2}{(1 - R)^2} \ , \tag{50}$$

$$F = \frac{4R}{(1 - R)^2} \ , \tag{51}$$

$$R = \sqrt{R_1 R_2} \ , \tag{52}$$

and

$$\eta = 2\pi\sigma \ n_g \ t_g - \frac{\epsilon_1 + \epsilon_2}{2} \ . \tag{53}$$

T_1 , T_2 and R_1 R_2 are the reflectivity (measured on the side of the mirror coating facing the gap) and transmission of the semitransparent mirror coatings on each of the plates. The coatings need not have the same R and T, and can also be absorbing, so that $R + T \neq 1$. ϵ_1 and ϵ_2 are the phase shift upon reflection from each of the coatings, which is explained in Section 20.10.1.4. t_g is the physical separation of the plates, measured from the surface of the coatings and n_g is the refractive index of the medium in the gap, which is usually air or a vacuum. In the most general case, R_1, T_1, ϵ_1, R_2, T_2, and ϵ_2 all vary with wavelength.

Since the $\sin^2$ function in the denominator of Equation (49) is always a positive quantity, T_f attains a maximum value when $\eta = m\pi$; the integer m is called the order number. If we neglect for the moment the phase shifts, ϵ_1 and ϵ_2, in Equation (53), then T_f is a maximum when

$$m\pi = 2\pi\sigma \ n_g \ t_g \ , \tag{54}$$

or

$$m \ \frac{\lambda_o}{2} = n_g \ t_g \ . \tag{55}$$

Equation (55) states that the resonant, or maximum transmission condition, occurs when the separation of the plates is an integral number of half-waves. The Fabry-Perot interferometer can be considered as a resonant cavity for light waves. Similar resonance conditions occur when an integral number of half-waves fit into a microwave cavity or an organ pipe. One difference is that the resonance in microwave cavities is usually in the first or second mode, that is, $m = 1$ or 2. A very high mode - or to use the language of physical optics, a high order of interference - is obtainable in the Fabry-Perot interferometer. For example, a plate separation of $t_g = 1.0$ cm. is commonly used, in which case $m = 40{,}000$ at $\lambda = 500$ mμ; that is, 40,000 half-waves fit in between the plates.

20.10.1.2 Spectral transmission of F.P. interferometer. If the Fabry-Perot interferometer is illuminated with highly collimated light, as shown in Figure 20.99, then its transmission, T_f, as a function of wave number (frequency) consists of a series of transmission peaks which are equally spaced, as is shown in Figure 20.100. The spacing σ_f between adjacent transmission peaks is called the free spectral range of the interferometer and is inversely proportional to the separation of the plates:

$$\sigma_f = 1/2t_g \tag{56}$$

For most plate separations which are attainable, this free spectral range is rather small and hence many transmission peaks are packed close together. For example, suppose that two plates are separated by three thin pieces of metal foil which are a little less than one thousandth of an inch thick, so that $t_g = 0.0025$ cm. From Equation (56), we find that $\sigma_f = 200$ cm^{-1}. This means that in the green region of the spectrum, transmission maxima would occur at wave numbers of 20,000 cm^{-1}, 19,800 cm^{-1}, 19,600 cm^{-1}, and so on. Converting these wave numbers into wavelength (see 20.1.3.3), we find that transmission peaks occur at 500.00 mμ, 505.05 mμ, 510.10 mμ, and so on. Thus the separation of the peaks is about 5 mμ in this spectral region. If this interferometer were used to pass a spectrum line at 505.05 mμ, then it would be necessary to have an auxiliary blocking filter (a composite filter, to use the language of 20.5.1.3.2) to attenuate the unwanted transmission peaks at 500mμ, 510mμ, and at other wavelengths.

Figure 20.101 shows the shape of the transmission band of an F.P. interferometer whose plates are coated with highly-reflecting films so that $R_1 = R_2 = 0.95$. This is merely an enlarged view of one of the transmission bands shown in Figure 20.100, translated to a wavelength scale. The maximum transmission of the band is T_{max} and occurs at a wavelength λ_o. By definition, the total width of the band at $1/2\ T_{max}$ is $\Delta\lambda_{1/2}$ *. Thus $\Delta\lambda_{1/2}$ gives an indication of the narrowness, or the sharpness, of the resonance line of the interferometer, and consequently its ability to transmit radiant energy at its resonant wavelength, λ_o, and to attenuate radiant energy of nearby wavelengths. This is allied to the chromatic resolving power of the interferometer.

20.10.1.3 Resolving power of the F.P. interferometer. The narrowness of the resonance line in the frequency spectrum of any resonant device is proportional to the ability of the device to store energy with low loss of energy per cycle of oscillation. The dimensionless quantity "Q" of a resonant system is defined as:[55a]

$$Q = 2\pi \frac{\text{Energy stored in system}}{\text{Energy lost in one cycle}}, \qquad (57)$$

and is inversely proportional to the line width, $\Delta\sigma_{1/2}$. For example, the simple "LC" resonant circuit shown in Figure 20.101 stores energy in the electric field in the capacitor and in the magnetic field of the inductor. Each cycle a fraction of this energy is dissipated in the form of heat due to the finite resistance of the wire in the inductor and to losses in the dielectric of the capacitor. If this loss can be decreased, then the Q of the circuit will improve and its sharpness as a tuning element, say in a radio receiver, will improve. Similarly, the energy of a microwave cavity is usually extracted through a small hole in the wall of the cavity. During each cycle of oscillation, a small amount of the energy which is stored in the electromagnetic fields in the cavity leaks out of this hole and is also absorbed as Joule heat in the walls of the cavity. The Q of the cavity is improved by silver plating the cavity walls and thus reducing this loss. The case for the F.P. interferometer is completely analogous. Standing light waves are established in the gap between the two reflecting plates, and electromagnetic energy is stored in this gap. During each cycle of oscillation, which in the case of visible light is more than 10^{14} cycles per second, a small amount of this energy in the cavity is depleted by either being absorbed or transmitted through the coatings.

From the foregoing considerations, we would expect that the Q of a F.P. interferometer would increase as the reflectivity of the coatings on the plates is enhanced. Another method of improving the Q is to store more energy in the cavity, which is accomplished by simply making cavity larger - that is, by using a larger separation of the plates. A more quantitative analysis,[55b] such as the one in Section 5.16, leads to the relation

$$Q = \left(\frac{\lambda_o}{\Delta\lambda_{1/2}}\right)\left(\frac{\sigma_o}{\Delta\sigma_{1/2}}\right) = \left(\frac{\sqrt{R}}{1 - R}\right) m\pi . \qquad (58)$$

Thus an interferometer with $R = 0.95$ and a plate separation of 1 cm. has a Q over two million, which is larger by several orders of magnitude than any microwave device. The Q is a quite sensitive function of R, as R gets close to unity. The Q computed from Equation (58) can either be increased or decreased by the variation with wavelength of the phase shift on reflection of the two semi-transparent mirrors. A more exact formula is given in reference 56. It should be remarked that Q is used as a criterion of the narrowness of the pass band in preference to the chromatic resolving power, which is used in most books on physical optics. The reason is that there are many different criteria for defining the chromatic resolving power and it is not desirable that Equation (58) be confused with them.**

20.10.1.4 The phase shift upon reflection. A standing wave is created when a light wave is incident upon a reflecting surface. The positions of the nodes of this standing wave remain fixed in space. Suppose an incident light wave, whose electric vector is represented by

$$E = E_o\, e^{i(\omega t - 2\pi z/\lambda)} . \qquad (59)$$

* If the width $\Delta\lambda_{1/2}$ is translated into the corresponding width, $\Delta\sigma_{1/2}$, on a frequency (wave number) scale, they are related by: $\frac{\Delta\lambda_{1/2}}{\lambda_o} = \frac{\Delta\sigma_{1/2}}{\sigma_o}$, were $\lambda_o = \sigma_o^{-1}$

** For example, compare Eq. 58 in this text with Eq. 43 on page 334 of Born and Wolf[11].

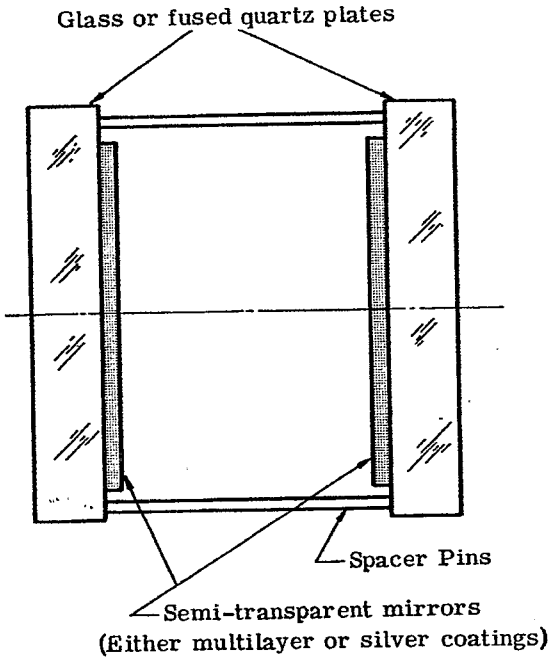


Figure 20.98- The essential parts of a Fabry-Perot interferometer.

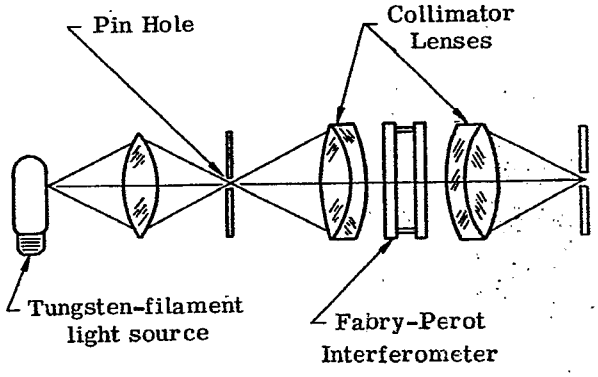


Figure 20.99- A method of illuminating the Fabry-Perot interferometer with collimated light so that it can be used as an optical filter.

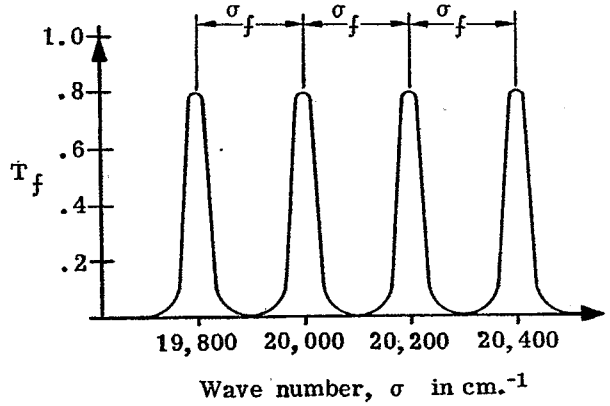
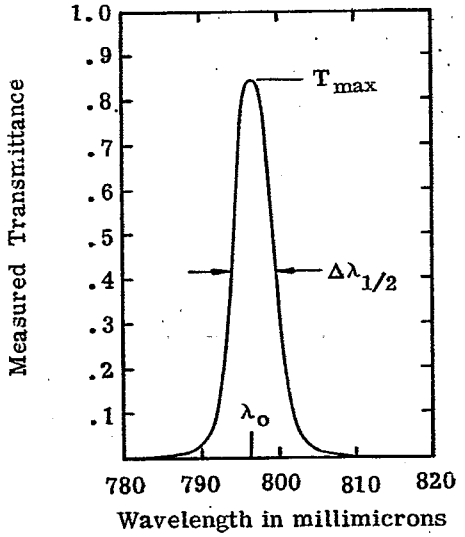


Figure 20.100 – The transmission as a function of wave number of a Fabry-Perot interferometer with t_g = 0.0025 cm .



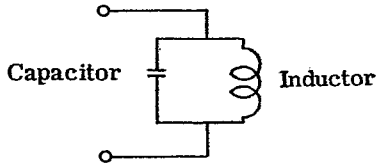


Figure 20.101- Measured spectral transmittance of Fabry-Perot type multilayer filter. Courtesy of Dr. C. Alley.
A parallel resonant circuit.

is reflected from a surface. If the reflected wave represented by:

$$E_r = R^{1/2} E_o e^{i(\omega t + \epsilon_1 + 2\pi z/\lambda)} , \qquad (60)$$

this defines the phase shift upon reflection, ϵ_1 . If the reflecting surface is a metal with infinite conductivity - that is, a perfect reflector, - then the standing wave has a node at the surface and the phase shift, ϵ_1, is 180°. At some instant of time, the amplitude of the light wave varies sinusoidally in space, as shown by the solid line in Figure 20.102; a half-cycle later it is represented by the dashed line. If ϵ is less than 180°, as would be the actual case for a silver mirror, then the node of the standing wave is to the right of the reflecting surface. If ϵ is greater than 180°, it is to the left, as shown in Figure 20.102. The phase shift upon reflection is exactly 180° for an air-glass interface, and does not vary with wavelength as long as the glass is optically non-absorbing. The ϵ for silver films changes slowly with wavelength, but this change is so small that it is usually neglected. The phase shift upon reflection for multilayer is much larger and, as is shown in 20.10.5, modifies the shape of the transmission band of a multilayer filter. It can also alter the wavelength at which the pass band occurs.[56]

20.10.1.5 Effect of ϵ on a resonant cavity. The effect of the phase shift upon reflection is to shrink or expand the walls of a resonant cavity. If the variable η in Equation (53) were defined as $\eta = 2\pi\sigma n_g t_g$, thus omitting the second term, this would tacitly assume that there is a node of the standing wave at each of the two reflecting surfaces. The addition term in Equation (53), $1/2 (\epsilon_1 + \epsilon_2)$ can be regarded as a correction term which accounts for the fact that the ϵ of each of the mirrors shifts the node of the standing wave. This shift in the node of the standing wave changes the resonant frequency of the cavity. If ϵ_1 and ϵ_2 change with wavelength, this alters the shape of the transmission band from the Lorentzian shape shown in Figure 20.101.

20.10.1.6 Shape of the transmission band. If the condition is fulfilled that ϵ_1 and ϵ_2 are constant with wavelength and that R_1 and R_2 are large enough so that F is much greater than one, then over the narrow wave number range near a pass band at σ_o , the sine function in Equation (49) can be replaced by its argument and T_f written as

$$T_f = (T_{max}) (1 + G \Delta^2)^{-1} \quad \text{where} \quad \Delta + \sigma_o = \sigma \qquad (59)$$

and constants such as F , 2π , t_g , etc. have been lumped together into G , which is essentially constant over the range of Δ, in which the approximation is valid . The point we wish to make is that this is a Lorentzian line shape (shown in Figure 20.101), and is similar to the shape power absorption curve of a series resonant "LC" circuit with a high Q[57] . A characteristic of the Lorentzian line shape is that a long tail which decreases slowly in amplitude extends to both the short and long-wave side. There are applications of F.P. type filters which require that the transmission on either the short-wave or long-wave side, decrease much more rapidly than is provided by a filter with a Lorentzian-shape transmission band. A few multilayers with a non-Lorentzian transmission band are shown in 20.10.7.

20.10.2 Fabry-Perot type multilayer filters. The Fabry-Perot interferometer described in the foregoing section is actually used as a band-pass filter in the laboratory. Notwithstanding the high Q which can be attained, it is not widely used because the optically polished plates are quite expensive and small mechanical vibrations or temperature changes cause the plates to warp out of parallel, thus degrading the Q of the instrument. These difficulties are partially avoided if the material in between the plates (i.e. the spacer) is a solid material. This is accomplished by depositing a mirror on a substrate by evaporation in a vacuum, then evaporating a spacer layer whose optical thickness satisfies Equation (55), and then evaporating another mirror. Another method is to use a thin piece of mica as the spacer material and to evaporate a mirror on both sides. This technique is discussed in 20.10.5.2.

20.10.2.1 Method of analysis. In the case of the Fabry-Perot interferometer, it is quite natural to analyze the performance in terms of Equation (49). This has the advantage that one does not need to know any of the details about the construction of the semitransparent mirrors on the two faces - it is only necessary to know the R, T, and ϵ of each mirror, and n_g , t_g . However, in the case where the spacer is evaporated as an integral part of the filter, such as the filters depicted in Figures 20.103 or 20.108, then there are many methods of analyzing its spectral transmission, as for example the admittance method[58] , the matrix method[59] or by considering it as a Fabry-Perot interferometer[60] . There is a large class of multilayer filters whose spectral transmission is most easily analyzed by treating them like a Fabry-Perot interferometer. This is accomplished by selecting one of the films in the stack - usually the center film if the stack has an odd number of layers and is symmetrical - and considering this film as a spacer layer with an optical thickness $n_g t_g$. The remaining films in the stack are divided into two groups - those between the spacer layer and the substrate form an effective interface "A" and the films between the spacer layer and the incident medium form the other effective interface.[60] The spacer layer is usually, but not

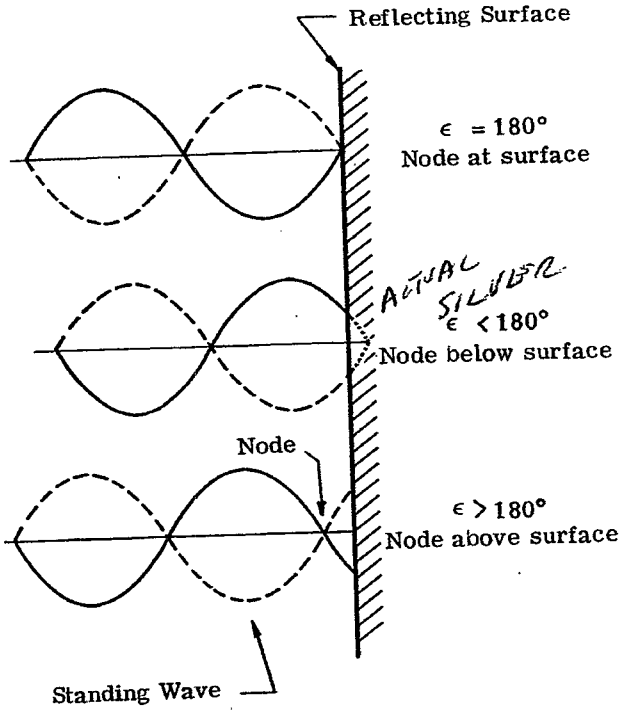


Figure 20.102- Showing the node of the standing wave as a function of the phase shift upon reflection, ϵ .

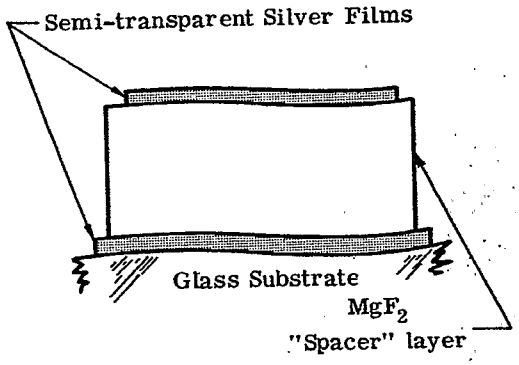
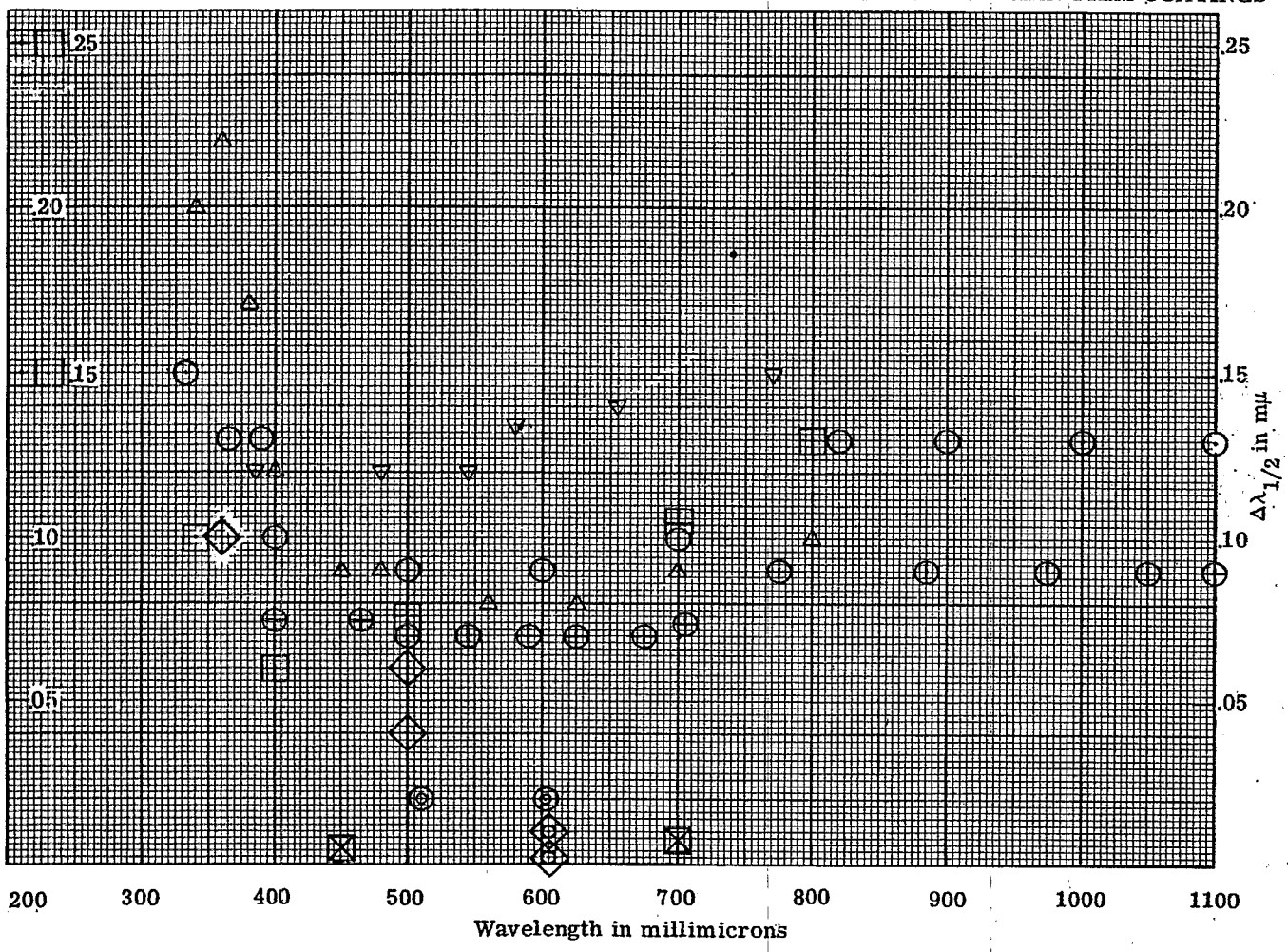


Figure 20.103- An enlarged view of a silver-dielectric-silver Fabry-Perot type filter, the lack of flatness of the substrate being exaggerated for purposes of illustration.



LEGEND: Filters contain metal unless specified otherwise.

- Baird-Atomic, Inc. Multilayer dielectric with Q = 1000 . $T_{max.}$ = 0.45 to 0.60 , blocking filters included.
- Baird-Atomic, Inc. Multilayer dieletric filter with rectangularly shaped pass band. $T_{max.} > 0.60$, blocking filters included.
- Baird-Atomic, Inc. Type "A"

$T_{max.}$			$\Delta\lambda 1/2$
0.15	→	0.20	25 mμ
0.10	→	0.15	15 mμ
0.05	→	0.10	10 mμ

- Bausch and Lomb, Inc. "Standard series" , second order interference, $T_{max.}$ = 0.30 to 0.35 .
- Carl Zeiss, Jena^{er}, "Single filters" , $T_{max.}$ = 0.25 in the u.v. and near i.r. . $T_{max.}$ = 0.35 in the visible.
- Carl Zeiss, Jena^{er}, "Double filter" (a composite filter) Type DSIF , $T_{max.}$ = 0.08 to 0.15 .
- Multilayer dielectric, $T_{max.}$ = 0.93 , without blocking filters. See reference 64 .
- Multilayer dielectric, consisting of from 21 layers ($T_{max.}$ = 0.55) to 29 layers ($T_{max.}$ = 0.15) . See reference 65 .
- Schott (Jena^{er} Glaswerke Schott & Gen., Mainz) These are "Line filters" of unspecified order of interference.
- Specially prepared F-P type filter consisting of nine dielectric films and two silver films . $T_{max.} \sim 0.40$. See reference 64 .
- Spectrolab, Inc. Published curves show $T_{max.} \sim 0.90$.

Figure 20.104- The total width at .5 $T_{max.}$, $\Delta\lambda 1/2$, and wavelength of the passband, λ_o , of some representative Fabry-Perot type filters composed of the metal-dielectric-metal type and all-dielectric type. See 20.10.2.2 for details.

always, an integral number of half-waves in optical thickness. References 60 and 61 show some exceptions. Thus by dividing a multilayer into a spacer layer and two effective interfaces, it can be analyzed as a Fabry-Perot interferometer and Equation (49) can be used to compute its transmission, provided R_1, T_1, ϵ_1, R_2, T_2, ϵ_2 for the interfaces are known.

The simplest type of multilayer whose performance can be analyzed as a F.P. interferometer is the silver-dielectric-silver interference filter (shown in Figure 20.103). When its discovery was announced, it was called an interference filter, since it was the first type of filter which operated on the basis of the interference of light, rather than absorption. When other types of multilayers came into use a decade later, it was recognized that all multilayers are in a sense "interference" filters, since their transmission characteristics depend upon the interference of light reflected from various layers within the multilayer. Thus in this section, the term "Fabry-Perot type filter" (F-P type) is used in preference to"interference filter".

20.10.2.2 Criteria for evaluating F-P type filters. Most F-P type filters are band pass filters and have a narrow transmission spike in the pass region and a high attenuation outside of that region. In order to compare the performance of different types of F-P filters, the following attributes are sometimes considered:

(1) The wavelength λ_o of the maximum transmission of the pass band.

(2) The maximum transmission of the pass band, T_{max}.

(3) The total width of the band at half intensity (i.e. at 1/2 T_{max}) $\Delta\lambda_{1/2}$. This is related to Q: $\Delta\lambda_{1/2} = \lambda_o/Q$.

(4) The Q of the filter, or alternatively, Q^{-1}.

(5) The extent of the attenuation region, and whether blocking filters need be added to extend this region.

(6) The shift of the transmission band as a function of the angle of incidence.

In comparing the literature of various manufacturers of multilayer F-P type filters, it is evident that even if all of these data are given, there is still no substitute for a spectral transmission curve. $\Delta\lambda_{1/2}$ is not necessarily a good criterion for comparing filters, because different types of filters have transmission bands of different shapes. For example, the pass band of the silver-dielectric-silver F-P type filter shown in Figure 20.105 has a Lorentzian line shape and shows the characteristic long tail of high transmission towards the blue, whereas the filter in Figure 20.118 has a pass band of a different shape and although its $\Delta\lambda_{1/2}$ is twice as large, the transmission in the blue decreases much faster. Sometimes the T_{max} given in the specifications of a manufacturer may or may not include the blocking filters which should be placed in series with the F-P type filter in order to eliminate unwanted transmission bands, as appear in the filters shown in Figures 20.105 and 20.115. Figure 20.114 shows how such a blocking filter is used to eliminate an unwanted transmission band at short wavelengths. The $\Delta\lambda_{1/2}$ at various λ_o of F-P type band-pass filters is shown in Figure 20.104. These data are compiled from the scientific literature and from the catalogues of manufacturers. In the latter case, they represent some of the filters produced by some manufacturers. This is definitely not a comprehensive list, but rather it is intended to give some idea of the range of the $\Delta\lambda_{1/2}$ and T_{max} which can be achieved at various wavelengths. Neither is this list intended to be encyclopedic, including all manufacturers. As was mentioned previously, the parameters T_{max} and $\Delta\lambda_{1/2}$ often do not adequately describe the performance of a filter, and hence it would be incorrect to conclude from Figure 20.104 that manufacturer "A's" filters are superior to "B's" filters. Evidently there is some variation in both the T_{max} and $\Delta\lambda_{1/2}$ between individual filters of a given type, because most manufacturers gave a range of values. Average values are used for the data shown in Figure 20.104

20.10.3 Band pass filters containing metal films.

20.10.3.1 The metal-dielectric-metal (M-D-M) Fabry-Perot type filter. The simplest filter of the F-P type is a three-layer filter consisting of a dielectric film, such as magnesium fluoride, sandwiched between two semitransparent metal films. In 1939 Geffcken [62,63] applied for a patent on the device in which both the metal films and the dielectric spacer layer film are evaporated. The beauty and simplicity of this method is that such a filter can be deposited on a substrate of common window glass, rather than an extremely precise optical flat, as is requisite for the F-P interferometer. The point is, that the surface of a piece of window glass deviates many wavelengths from being optically flat, but that the three or more layers in the multilayer coatings follow the contour of the substrate. The drawing in Figure 20.103 shows this effect; the lack of planeness in the substrate is greatly exaggerated. Actually, the lack of planeness in the substrate does broaden the transmission bands slightly, but this effect is certainly not noticeable with broad 10 mμ band widths typical of this type of filter.

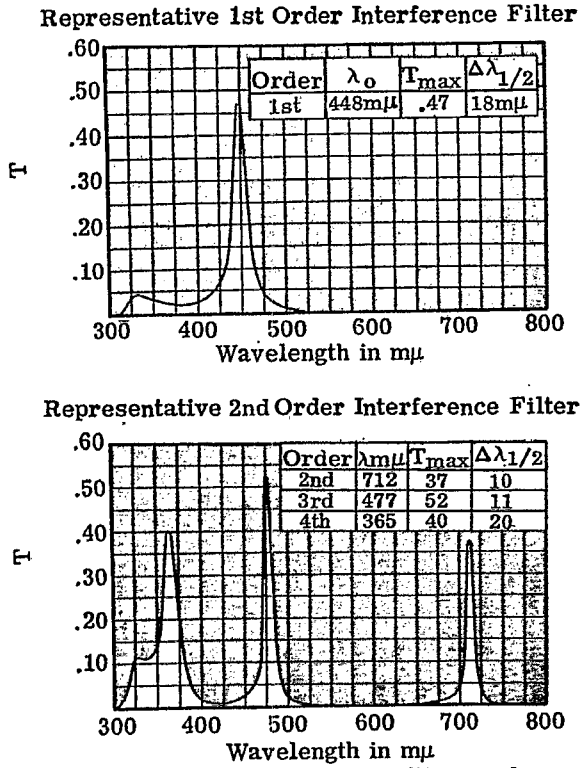


Figure 20.105- Measured transmittance of some metal-dielectric-metal Fabry-Perot type filters. Courtesy of Bausch and Lomb, Inc.

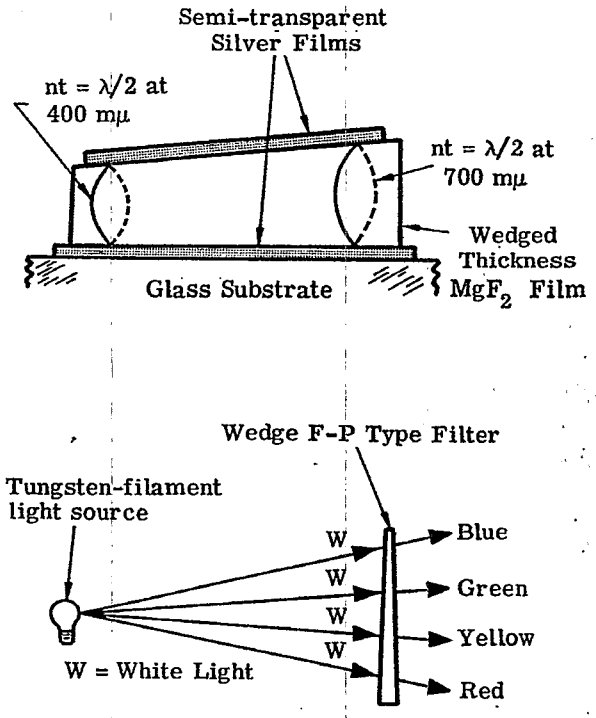


Figure 20.107- (Upper) A cross section of a wedge Fabry-Perot type filter. (Lower) Showing the use of this type of filter as a monochromator.

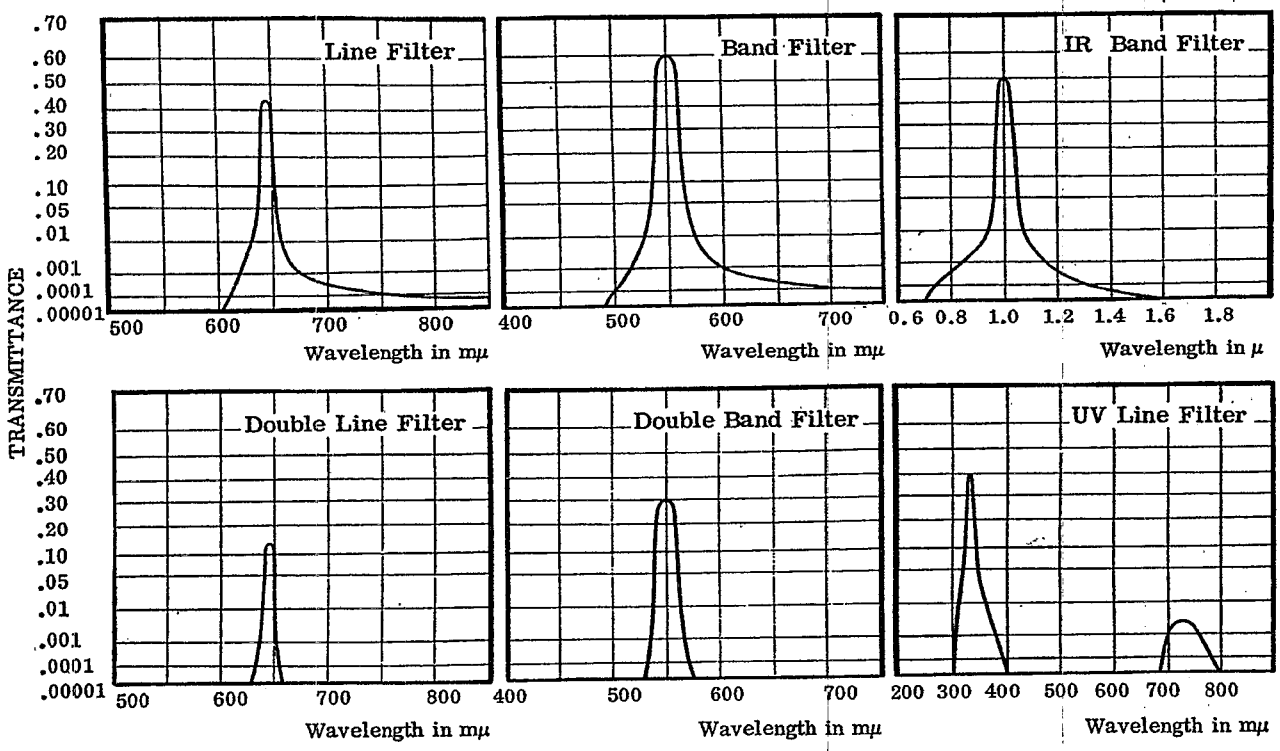


Figure 20.106- Measured transmittance (on a logarithmic scale) of silver-dielectric-silver Fabry-Perot type filters. Courtesy of Schott & Gen., Mainz, West Germany.

For wavelengths greater than 340 mμ , silver is the best metal film to use, having the highest reflectivity and lowest amount of absorption. Below 340 mμ the optical constants of silver change rapidly and the reflectivity drops to very low values. This is why silver-dielectric-silver F-P type filters have a transmission "leak" in this region, as is shown in Figure 20.105. At these shorter wavelengths aluminum is generally used as the metal film, although it absorbs a large fraction of the radiant energy and consequently these filters usually have a T_{max} of .10 to 20. Figure 20.105 shows the spectral transmittance of a silver-dielectric-silver F-P type filter with a first order transmission maximum at 450 mμ . The transmission "leak" in the ultraviolet is due to the loss in the reflectivity of the silver in this region. Also shown in Figure 20.105 is the transmittance of a filter which has a second order peak at 712 mμ , and third and fourth order peaks at shorter wavelengths. In Figure 20.106 is depicted the transmittance (on a logarithmic scale) of various types of M-D-M filters. The "double filter" consists of two identical filters cemented together to form a composite filter. Although the peak transmittance is low, (T_{max} is from 0.10 to 0.30), an extremely high attenuation in the reject region is attained - the optical density is greater than five. This high attenuation is achieved because the filters are absorbing and hence the considerations which apply to dielectric films (see 20.5.1.4) do not hold here.

20.10.3.2 The wedge M-D-M filter. Another interesting form of a M-D-M filter is the wedge Fabry-Perot type filter,[66] which is depicted in Figure 20.107. Two silver films are deposited on either side of a layer of magnesium fluoride, which is wedged shaped so that its thickness varies in a linear fashion along the length of the filter. At one end the optical thickness is a half-wave of violet light (400 mμ), so this portion passes the violet. At other positions along the filter the dielectric film is thicker and these portions pass the blue, green, yellow, and finally the red.* In actual practice, the filter is usually manufactured a second order filter, rather than first order, as shown, thus achieving a narrower band width. In this case, at the thick portion of the wedge the second order red and the third order blue overlap, and it is necessary to remove the blue by appropriate dye or glass filters. The $\Delta\lambda_{1/2}$ of this second order filter is 10 mμ , independent of wavelength. The slope of the wedge and the length of the filter are chosen so that a one millimeter slit gives this pass band. Thus by inserting a slit 1 mm wide in front of this wedge filter and illuminating it with white light, a rather inexpensive source of quasi-monochromatic light is obtained. This type of wedge can be deposited in an annular ring on a disk.[67] The wavelength scanning is accomplished by rotating the disk past a slit.

20.10.3.3 Other types of narrow pass-band filters which contain metal films. A higher peak transmission and narrower band width is attained with F-P type filters if several dielectric films are added to the stack in addition to the spacer layer.[64] Turner and Berning have devised some band-pass filters which contain a single silver film and many dielectric films.[68] M-D-M filters are also useful as reflection filters, particularly in the infrared spectral region.[64, 69]

20.10.4 All-dielectric Fabry-Perot type filters. The simplest form of this type of filter is shown in Figure 20.108. The silver films are replaced by semitransparent mirrors composed of dielectric materials, such as a quarter-wave stack. Thus the design of a filter consisting of seven layers would be:

glass H L H LL H L H air .

Here the H L H combination is a three-layer quarter-wave stack (H and L have a QWOT at λ_o) and LL represents a spacer layer of half-wave optical thickness at λ_o . The spectral transmission versus frequency of such a multilayer is shown in Figure 20.109. Although this and other multilayers which are used as illustrations can be used only in the infrared because they contain germanium as a high index material, the principles involved here apply to any spectral region.

20.10.4.1 T_{max} . The concept of an absentee layer (20.1.5.2.2) is useful in determining the transmission of this filter at the wavelength λ_o (i.e. g = 1.0) where the maximum of the pass band is located. At this wavelength the LL layer is absentee and hence it can be removed from the stack, leaving:

glass H L H H L H air.

This leaves two of the H layers next to each other, resulting in the layer HH which has an optical thickness of a half-wave. After removing this HH combination from the stack we are left with four layers:

glass H L L H air.

* In actual practice, the optical thickness of the dielectric spacer layer is slightly thinner than a half-wave, due to the phase shift upon reflection of the silver films. Also, the thickness of silver films varies along the wedge, due to the dispersion of the optical constants of the silver.

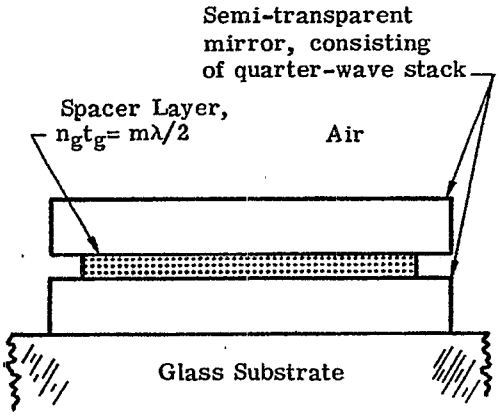


Figure 20.108- Design of a conventional all-dielectric Fabry-Perot type filter.

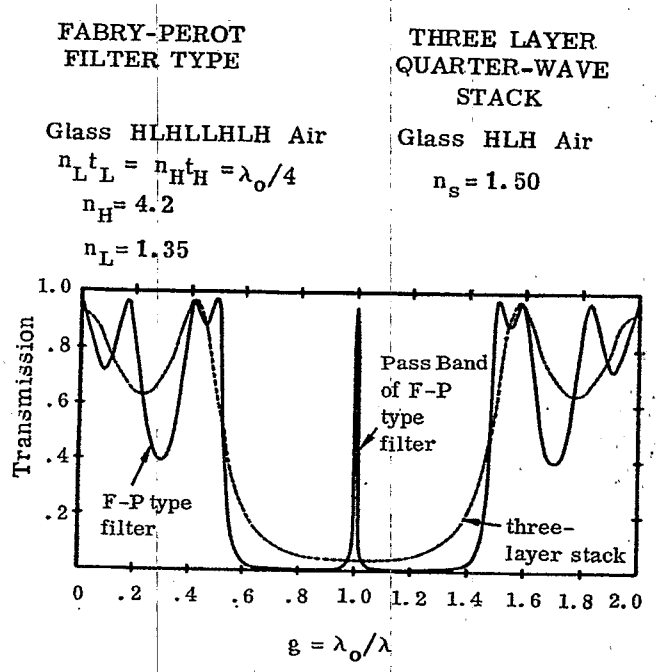


Figure 20.109- Computed spectral transmission of an all-dielectric Fabry-Perot type filter (solid curve) and a three-layer quarter-wave stack (dashed line).

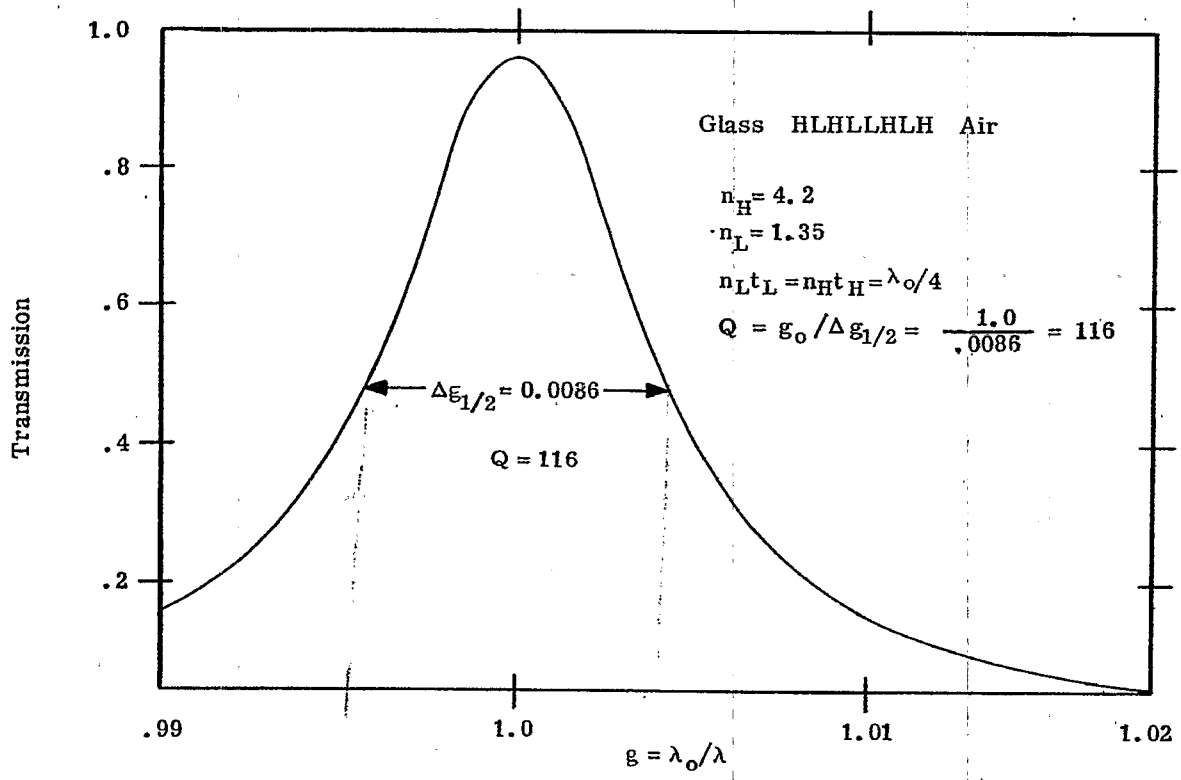


Figure 20.110- Computed transmission of the pass band of the Fabry-Perot type filter shown in Fig. 109.

The half-wave combination LL is absentee and can also be removed from the stack. Repeating this process, we see that at λ_o the transmission is the same as that of a single surface of uncoated glass of index 1.50, and thus from Equation (21), $T = 0.96$. This does not include the reflection loss at the second surface of the substrate.

20.10.4.2 It is instructive to compute the Q of this filter. First, it is necessary to compute the reflectivity, R_1 and R_2, of each of the effective interfaces. The first effective interface is:

glass H L H cryolite

and thus R_1 is computed from an incident medium of cryolite. R_2 is computed for a stack:

air H L H cryolite

where the substrate is air and cryolite ($n_o = 1.35$) is the incident medium. Utilizing Equations (40) and (43), we find that $R_1 = 0.953$ and $R_2 = 0.968$, whence $R = (R_1 R_2)^{1/2} = 0.961$. From Equation (58) a Q of 80 is computed. Figure 20.110 shows the computed spectral line shape of this filter, on a frequency scale. The Q, measured from this graph, is 116. The additional narrowing of the transmission band is attributed to the phase shift upon reflection of the three layers which constitute the effective interfaces of this F-P type filter. References 56 and 61 show how to include the effect of the phase shift in computing the Q of the system.

20.10.4.3 Effect of phase shift upon reflection. As an example of how the phase shift upon reflection can influence the shape of the transmission band of a F-P type filter, consider the following multilayers, which are designated as Design I and Design II :

I glass L H L HH L H L air

II glass HH L H L HH L H L HH air .

Design II is essentially Design I with an extra half-wave layer added to each end of the stack. At $g = 1.0$, the half-wave layers are absentee and the reflectivities of the effective interfaces of each of the two stacks is exactly the same. However, if we examine the transmission bands for these two stacks, shown in Figure 20.111, it is evident that the width $\Delta \lambda_{1/2}$ at $1/2\, T_{max}$ is somewhat less for Design I than for Design II. This can be attributed to the variation with wavelength phase shift upon reflection of the effective interfaces, which is shown in Figure 20.112. In each case the phase shift upon reflection is measured from inside of the germanium spacer layer. At $g = 1.0$ the phase shift is zero, which means that the node lies at the surface of the multilayer. At lower frequencies than $g = 1.0$ the node lies to the right of the surface. The shift of this node as the wavelength changes alters the shape of the transmission band.

20.10.4.4 The use of blocking filters. All-dielectric F-P type filters are usually used in conjunction with blocking filters. These blocking filters must in general have a much higher attenuation than the blocking filters which are used in conjunction with the M-D-M type filters. The reason is that in the former case, the unwanted transmission bands cover a wide spectral region, whereas the unwanted transmission bands of a M-D-M type are usually quite narrow. For example, suppose it is desired to use the transmission band at 477 mμ in the M-D-M type second order filter shown in Figure 20.105. In this case it is necessary to use an auxiliary blocking filter to eliminate the unwanted transmission bands below 425 mμ and the band at 712 mμ . There are many absorption type filters which could be used to attenuate below 425 mμ. The band at 712 mμ is comparatively narrow and hence the total amount of radiant flux which "leaks" through this band is not large. Consequently, the amount of attenuation required in the blocking filter is not as great as it would be if the pass band were wide. In the case of all-dielectric F-P type filters, the quarter-wave stacks which are used for the semitransparent mirrors do not reflect over a wide range of wavelengths. Thus a quite appreciable amount of radiant energy is liable to "leak" through the filter in the region outside of the high-reflectance zone of the mirrors. The narrower the pass band of the F-P type filter, the more effective should be the blocking filters to insure that the total amount of radiant flux transmitted in the spectral region of the pass band of the filter should be much greater than the flux which leaks through at other wavelengths. For example, the all-dielectric F-P type filters shown in Figures 20.115 and 20.116 have a substantial transmittance below 8.0 μ. The total transmitted radiant flux below 8.0 μ to 3.9 μ (where the PbTe films start to absorb) is considerably larger than the radiant flux transmitted through the pass band near 10 μ.

20.10.5 All-dielectric F-P type filters for the visible. All-dielectric F-P type filters are produced commercially for most of the visible spectral region with a wide range of Q . Filters with a Q from 10 to 1000 are available. Values of T_{max} from 0.45 to 0.60 are commonly attained, which includes the appropriate blocking filters. The spectral transmission of an all-dielectric F-P type filter which has its pass

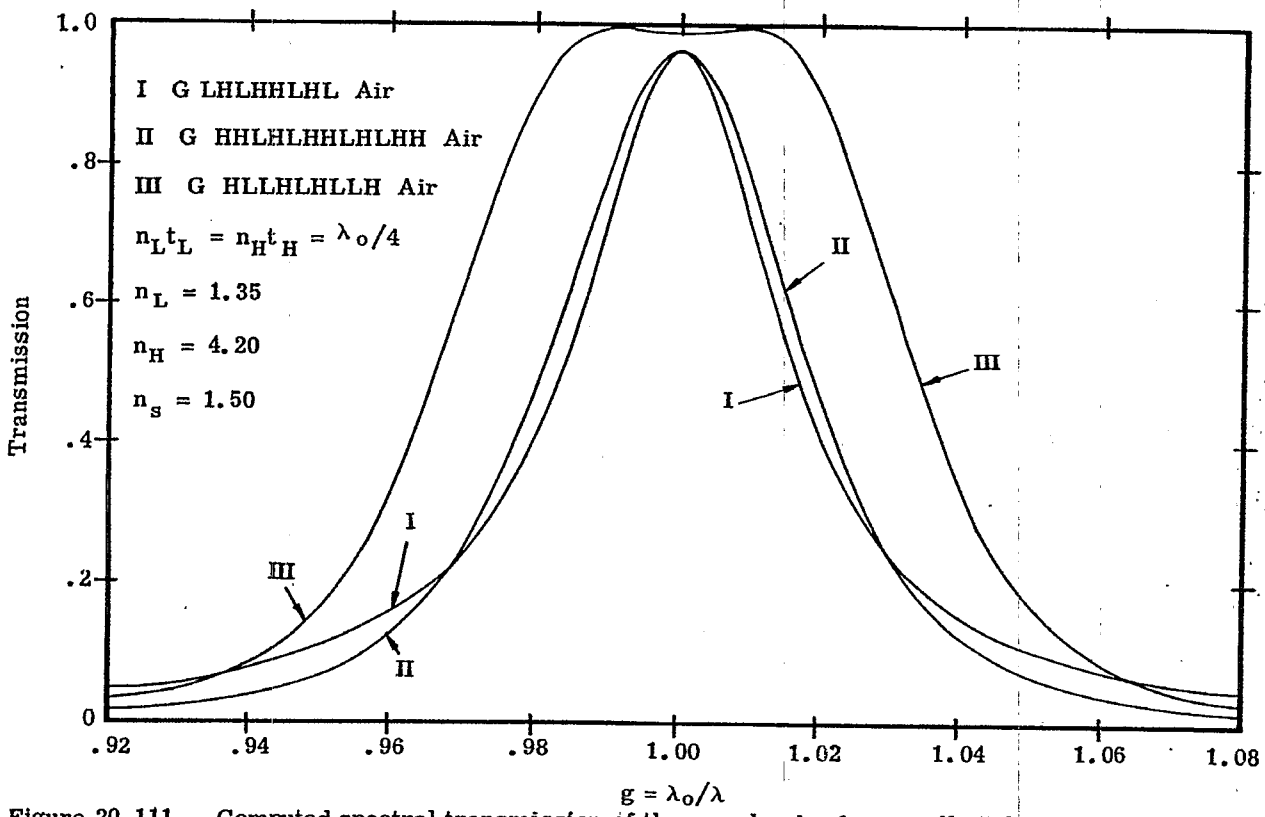


Figure 20.111- Computed spectral transmission of the pass bands of some all-dielectric Fabry-Perot type filters. Filter III has non-Lorentzian shaped pass band.

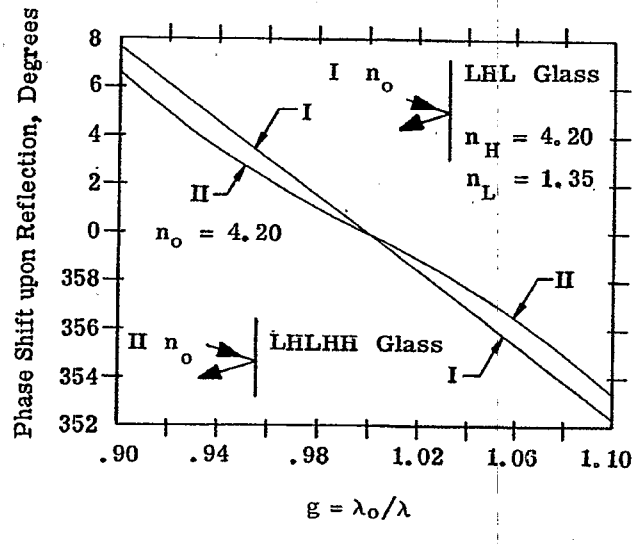


Figure 20.112- The phase shift upon reflection, ϵ , of the "effective interface" of the filters I and II in Fig. 111.

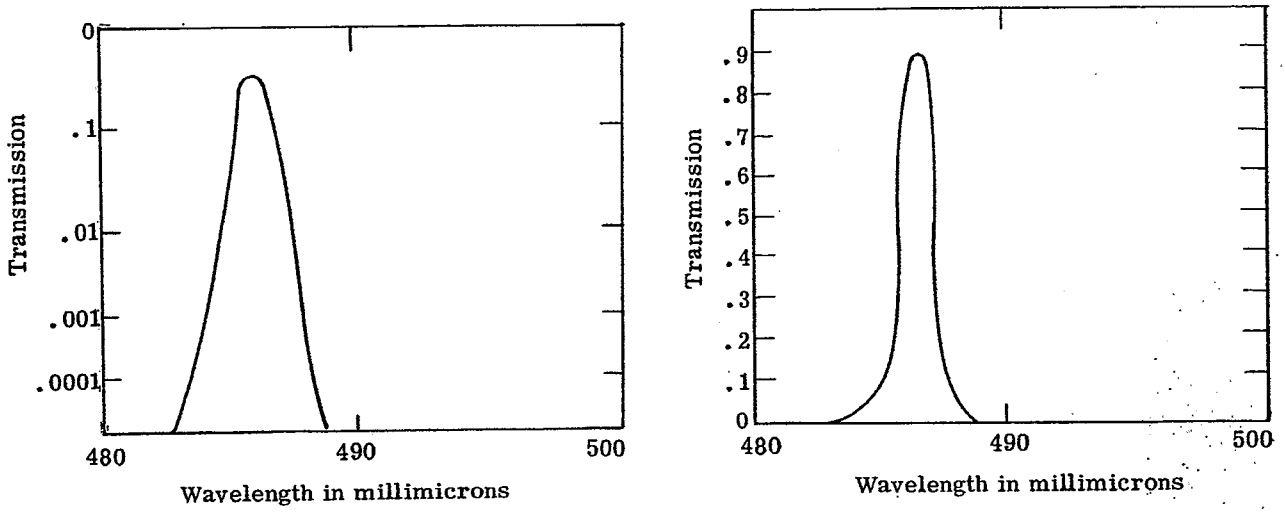


Figure 20.113- Measured transmittance of an all-dielectric Fabry-Perot type filter on a linear and logarithmic scale. Courtesy of Spectrolab, a Division of Textron Electronics, Inc.

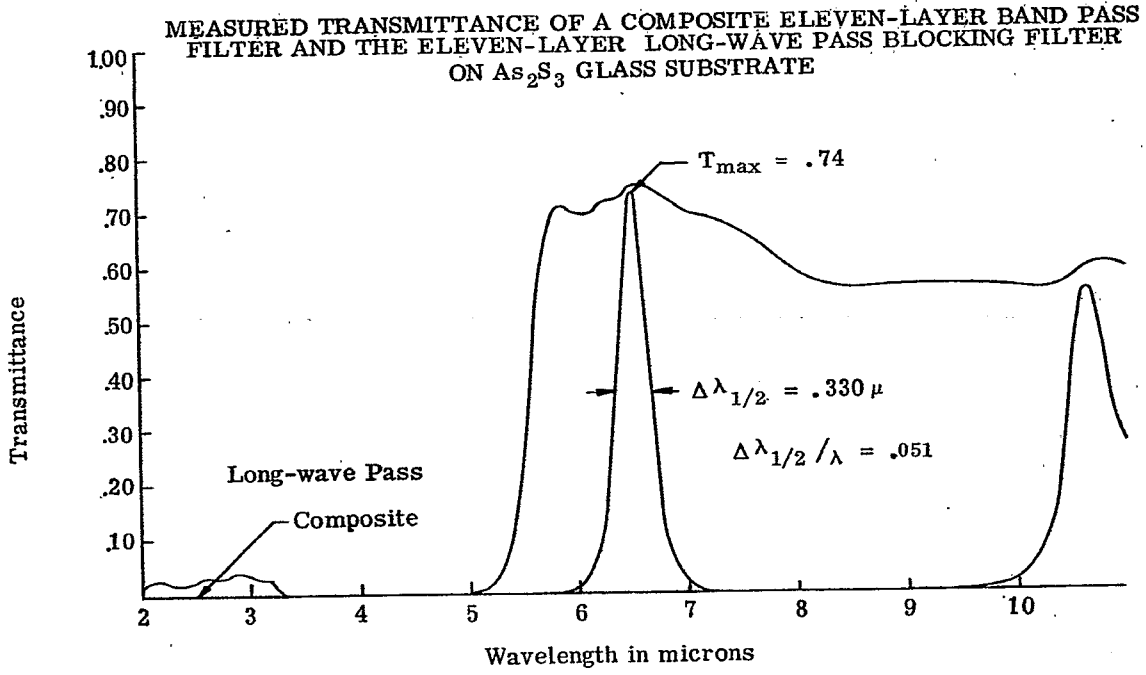


Figure 20.114- Measured spectral transmittance of a Fabry-Perot type filter and auxillary blocking filter composed of ZnS and PbTe .

band in the blue, is shown in Figure 11-. Although the peak transmission would be lowered slightly if a blocking filter were added in tandem, the T_{max} is still quite high. Also, the attenuation is quite high; an optical density of 4.0 is achieved at $3\,\Delta\lambda_{1/2}$ from the wavelength of maximum transmission.

20.10.5.1 Extremely narrow bandwidth filters. From theoretical considerations alone, one might conclude from Equation (58) that it is possible to construct all-dielectric multilayer F-P filters which have Q's of 10,000 and hence a band width of 0.05 mμ at 500 mμ in the visible. Such filters would be quite useful to astronomers, who have been using the expensive Lyot type polarization filters to isolate the H^{α} line. These filters would also find many applications in spectrochemical analysis; in some instances they could replace costly spectrometers which contain diffraction gratings. One method of attaining such large values of Q would be to increase the number of layers in the quarter-wave stacks which constitute the effective interfaces, and thus obtain higher reflectivities. The same practical difficulties which were described in 20.8.2.3 are encountered. The small amount of absorption and scattering in each film degrades the reflectivity and consequently sets a lower limit on the bandwidth of F-P type filters. Thus it is not too difficult to manufacture all-dielectric F-P type filters for the visible spectral region with a $\Delta\lambda_{1/2}$ of from 1.0 mμ to 1.5 mμ and a T_{max} of greater than 0.50 (this does not include blocking filters). A Russian publication [65] reports that filters with a $\Delta\lambda_{1/2}$ of 0.13 mμ and a T_{max} of 0.15 have been produced.

20.10.5.2 Filters which use a mica spacer. Another method of increasing the Q of a filter is to increase the order of interference m. The mechanical stress in the films (Section 20.2.3.2.4) makes it impractical to use thick layers of evaporated material to manufacture a spacer layer with a high order of interference. By using a thin sheet of mica as a spacer in a F-P type filter,[70,71] it is possible to attain values of m in the range from 70 to 700. Both sides of the mica spacer, which is usually from 0.005 to 0.0005 inches in thickness are coated with a semitransparent multilayer mirror, such as a quarter-wave stack. Using this technique, a filter which isolates one of the lines of the yellow sodium doublet has been produced.[71] A filter with a pass band at 570mμ, a $\Delta\lambda_{1/2}$ of 0.1 mμ and a T_{max} of 0.25 is reported.[71] The difficulty of using filters with a high order of interference is mentioned in 20.10.1.2, namely the problem of blocking out adjacent transmission bands which, in the case of the filter cited in the foregoing sentence, occur at intervals of 1.1 mμ on either side of the main pass band. This can be accomplished by inserting additional mica F-P type filters in tandem, but this reduces the T_{max}.

20.10.5.3 F-P filters at non-normal incidence. If a F-P type filter of the type shown in Figure 20.108 or 20.103 is inserted in a collimated beam of light at non-normal incidence, the following effects are observed as ϕ increases:

(1) The transmission pass band broadens and shifts to shorter wavelengths (i.e. a blue shift). This is because $n_g t_g$ in Equation (55) is replaced by an effective thickness (see 20.1.6.2) which is less than its original value. Hence a smaller λ_o satisfies this equation.

(2) The transmission band is partially linearly polarized. If the incidence angle ϕ is increased to large enough angles, two distinct bands are seen, each at a different wavelength. The light in one band is linearly polarized in the "s" plane, and the other in the "p" plane.

This angle shift of the maximum of the pass band can often be used to good advantage. For example, suppose it is desired to isolate the mercury green line at 546μ, and a F-P type filter which is available has a pass band at 550mμ. The spectral position of the pass band can be easily shifted so that it passes the Hg green line by tilting the filter less than ten degrees. Of course, the performance of the filter has been degraded by this tipping because the pass band has been broadened, but the loss is not serious if the pass band is wide to begin with. It is also evident that if a F-P type filter is placed in a beam of convergent light, then the angle shift broadens the transmission band asymmetically towards shorter wave-lengths. Thus a filter which is placed in a convergent beam should have its λ_o at normal incidence at a slightly longer wave-length. For example, Lissberger and Wilcock,[72a],[72b] calculate that a filter which is to have its optimum performance at 5000 mμ is placed in an f 2.0 beam, then at normal incidence its T_{max} should be located at $\lambda_o = 502$ mμ. It is also evident that filters which have extremely narrow band widths should be used at normal incidence, in a well collimated beam to prevent the loss of the narrow bandwidth by the angle shift broadening. The spectral transmittance curve in Figure 20.115 shows the angle shift of an all-dielectric F-P type filter for the infrared.

20.10.6 All-dielectric F-P type filters for the infrared. All-dielectric F-P filters are available for the infrared spectral region with a Q as large as 200. Figures 20.114, 20.115, and 20.116 show the spectral transmittance of some all-dielectric F-P type filters which are used in the infrared. These curves are intended to present a sample of what can be accomplished. The filter shown in Figure 20.114 is intended to have a

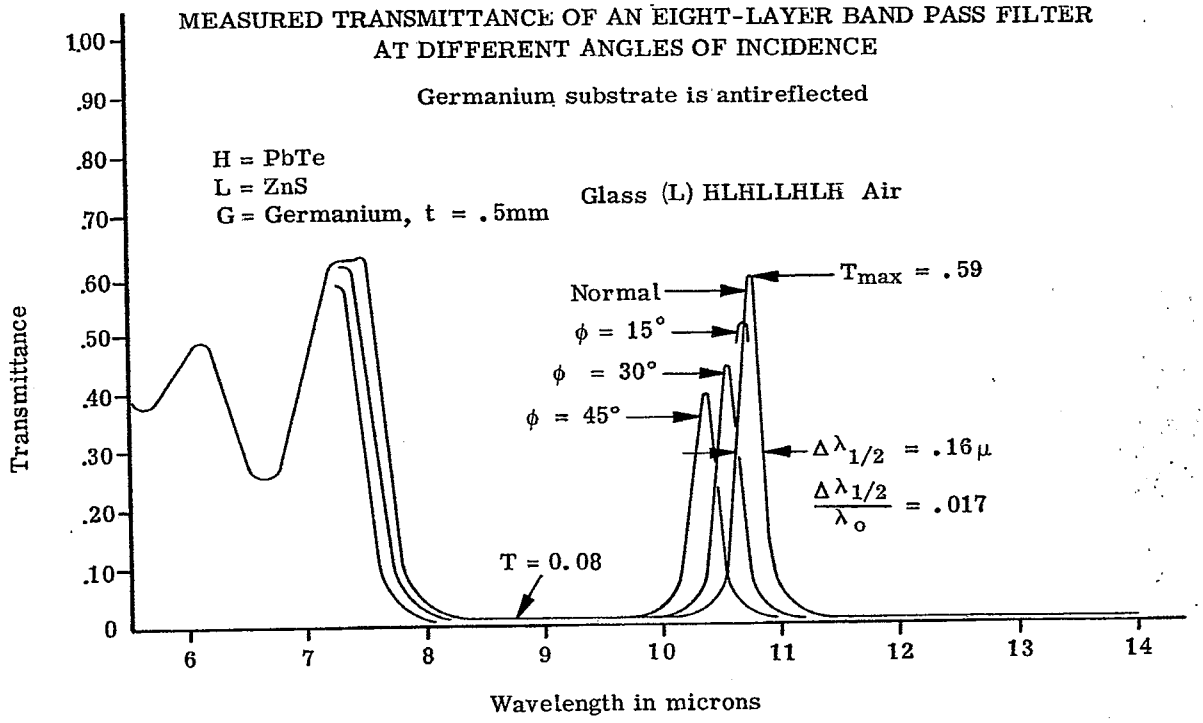


Figure 20.115- Measured spectral transmittance of a Fabry-Perot type filter at various values of ϕ. The decrease in T at $\phi = 45°$ is due to vignetting in the spectrophotometer. Courtesy of Bausch and Lomb, Inc.

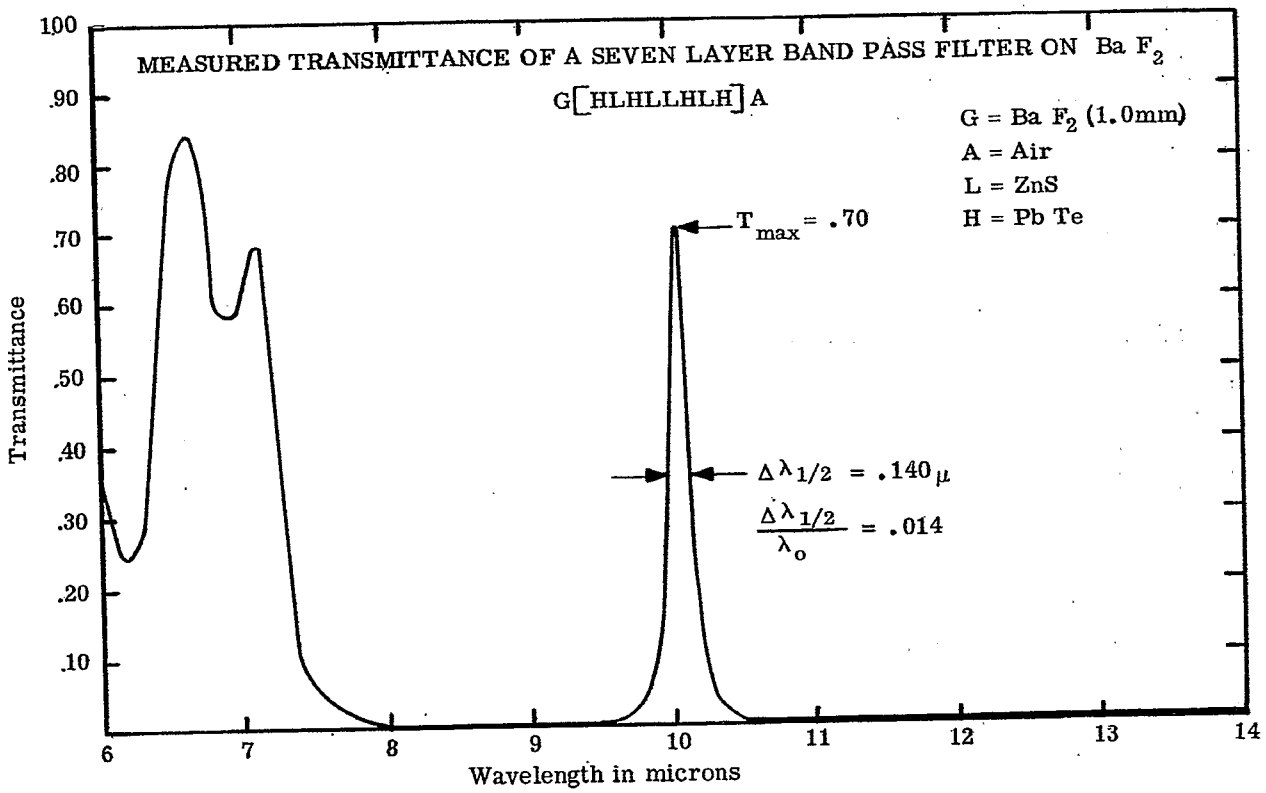


Figure 20.116- Measured spectral transmittance of a Fabry-Perot type filter on a BaF_2 substrate. Courtesy of Bausch and Lomb, Inc.

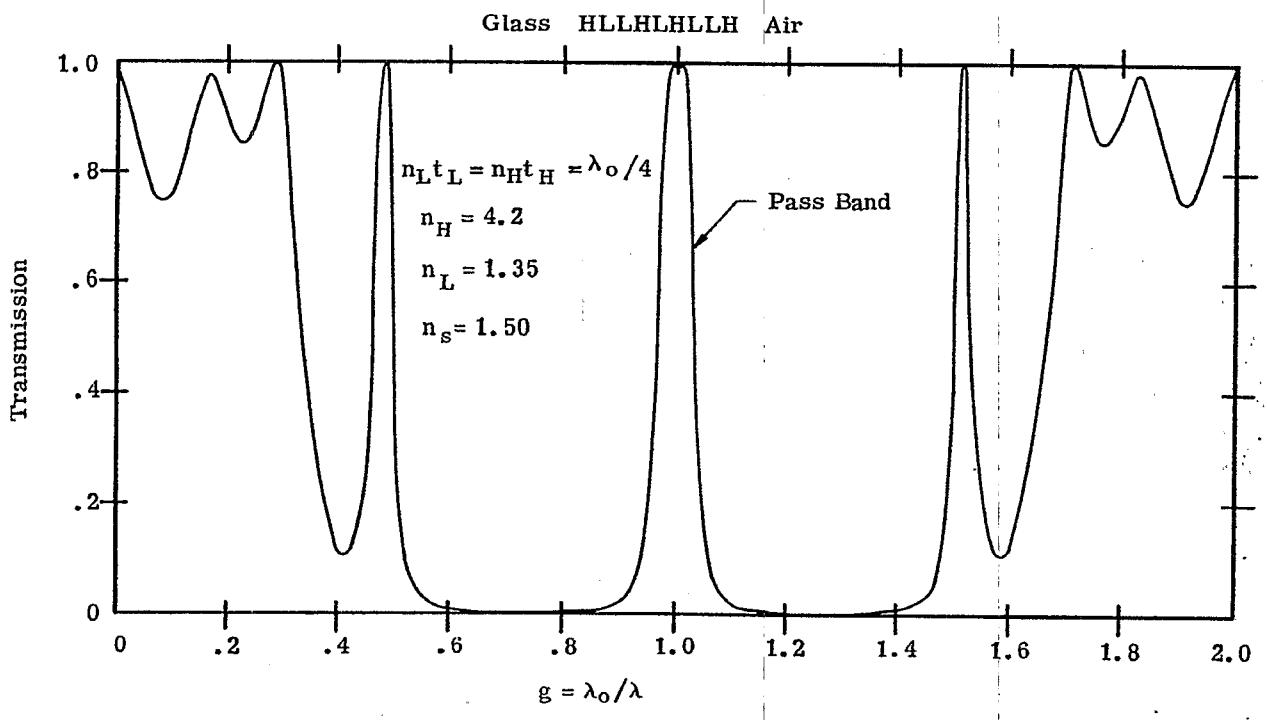


Figure 20.117- Computed spectral transmission of a Fabry-Perot type filter which has a non-Lorentian shaped pass band.

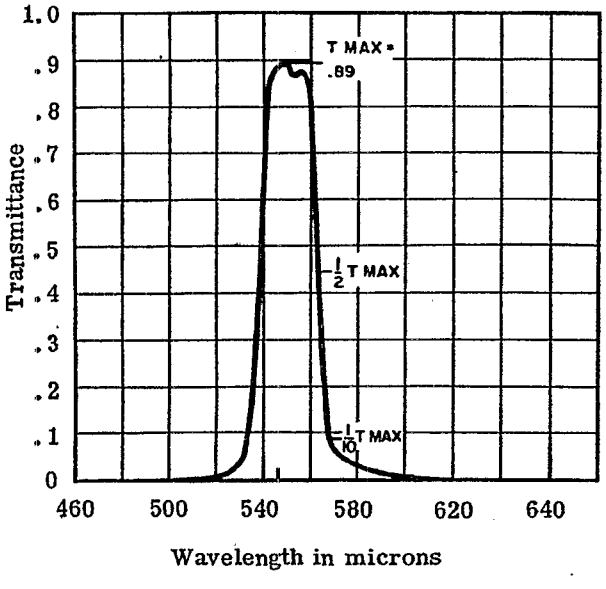


Figure 20.113 – Measured spectral transmittance of a narrow band-pass filter which has a nearly rectangular shaped pass band. Courtesy of Baird-Atomic, Inc.

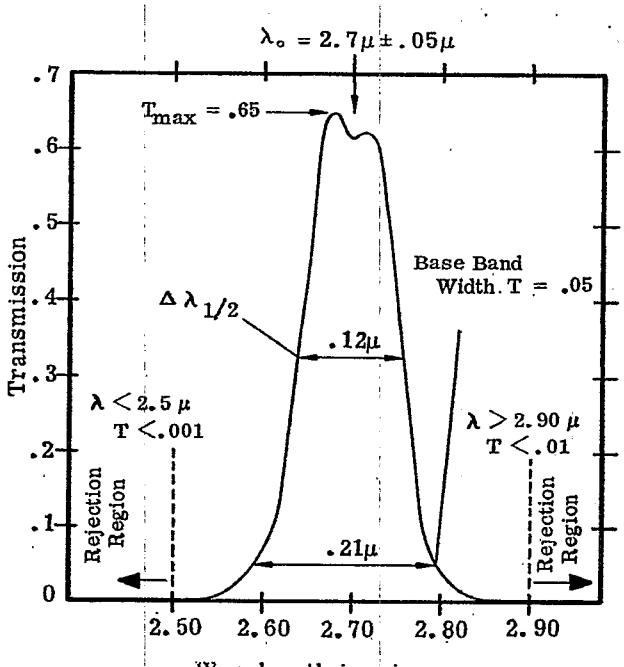


Figure 20.119- Measured spectral transmittance of a narrow band-pass filter which has a nearly rectangular shaped pass band. Courtesy of Eastman Kodak Company.

broad pass band; the Q is about 20. A long-wave pass blocking filter is inserted in tandem to attenuate the unwanted transmission "leak" below 3.3 μ. The region of high transmission above 10 μ is outside of the high-reflectance zone of the dielectric mirrors in the effective interfaces. The F-P type multilayer shown in Figure 20.115 is deposited on a substrate of germanium. The "L" layers which are in parenthesis, sic. (L) , are antireflection coatings for the germanium substrate. At λ_o , which is close to 10.8 μ at normal incidence, the same analysis which was applied in 20.10.5 is used here to remove absentee layers from the stack, thus leaving (L) germanium (L) , a germanium slab with an "L" antireflection coating on either side. Although the refractive index of this L layer (see 20.3.3) is not the optimum value, it still improves considerably transmission at λ_o . If these (L) coatings were not added, then T_{max} would be less than 0.47, instead of the value of 0.59 which is shown. The angle shift of the pass band to shorter wavelengths is also shown in Figure 20.115. The decrease in the T_{max} at non-normal incidence can be attributed to vignetting in the spectrophotometer which measured the transmittance. Figure 20.116 shows essentially the same type of coating, but on a substrate of barium fluoride. The substrate is a low-index material, and hence it is unnecessary to add the (L) coatings to antireflect the substrate, as is necessary for the multilayer shown in Figure 20.115. Comparing Figures 20.115 and 20.116, the effects of using a different substrate are manifested in the greater T_{max} for the multilayer on the CaF_2 substrate, a slightly narrower width of the pass band, and the higher transmission in the short-wave region below eight microns. Of course the latter effect is undesirable in many applications; the "leak" in the transmission in the short-wave region of the multilayers shown in Figures 20.115 and 20.116 could be removed by the addition of suitable blocking filters, as is done in Figure 20.114. Lead telluride is used as a high-index film material in this spectral region because it is transparent and has a large refractive index - and hence a filter with quite respectable Q is obtained with a small number of layers. Zinc sulfide is used as the low index material because it does not have a large mechanical stress (see 20.2.4.2.4) and also because it is transparent in this long-wave region. Silicon monoxide is not used in this spectral region because it has a strong absorption band starting at 8 μ .[72c] Greenler[16] has fabricated Fabry-Perot type filters with pass bands in the 10 μ region by using tellurium as a high-index material and sodium chloride as a low-index material. Using the same materials it is possible to manufacture filters with a pass band at wavelengths as long as 20μ .

20.10.7 F-P type filters with a pass band of non-Lorentzian shape. The F-P type filters which are described in 20.10.4 and 20.10.5 have a transmission pass band which is essentially Lorentzian in shape. The main drawback of this type of line shape is mentioned in 20.10.1.6, namely that unless the pass band is narrow, the filter has a long transmission tail which decreases in amplitude quite slowly. For example, suppose it is desired to use a multilayer filter to isolate the emission line at 491.6 mμ of a mercury discharge lamp. The lamp does not emit lines of any strength for at least 40 mμ on both the short-wave and long-wave side of 491.6 mμ. Therefore, it is not important that the filter have a narrow transmission band, in fact, the $\Delta\lambda_{1/2}$ could easily be as large as 20 mμ , provided the discharge tube does emit an appreciable amount of continuum radiation. However, it is quite important that the emission lines at 435.8 mμ in the blue and at 546.1 mμ in the green be attenuated very effectively, because these lines are at least a thousand times more intense than the 491.6 line. Thus, if the 491.6 mμ which passes through the filter is to be merely ten times more intense than the light from the blue and green lines, then the transmission of the filter is at 496 mμ and 546 mμ must be at least 10^{-4} of T_{max} . A M-D-M type filter would not furnish this much attenuation. A double filter of this type would furnish this much attenuation, but at the expense of a very low T_{max} . An all-dielectric F-P multilayer would furnish this degree of attenuation, provided that a large number of layers were used in the filter. However, this would mean that the transmission band would be quite narrow, whereas a narrow transmission band is not requisite. Such a filter with a narrow band would be expensive for two reasons: First, it contains a large number of layers, and is expensive to manufacture. Second, because the transmission band is narrow, the thickness of the layers must be controlled to quite close tolerances, so that the peak transmission of the filter occurs at exactly the desired wavelength of 491.6 mμ. The latter difficulty is avoided if a multilayer has a pass band which is essentially rectangular in shape, as is shown in Figures 20.117 to 20.121. Space does not permit us to elaborate some of the methods which are used to achieve transmission bands of this shape. One filter of this type is called a double half-wave system; the theory of such filters is discussed by Smith.[60] The spectral transmission of such a filter is shown in Figure 20.117. When analyzed as a F-P type filter, each of the effective interfaces is the film combination H LL H and contains a half-wave film; hence the name, double half-wave. The spacer layer is a quarter-wave optical thickness, rather than a half-wave layer. Figure 20.111 shows the transmission in the pass region so that it can be compared with the Lorentzian-shaped transmission bands of a conventional all-dielectric filter shown in the same Figure. Figure 20.118 shows the transmittance of a multilayer of this type with its pass band in the visible spectral region, while Figures 20.119, 20.120, and 20.121 depict infrared filters which pass at 2.70 μ , 4.50 μ , and 10.8 μ , respectively. Additional blocking filters have been added to the filters whose transmission curves are shown in Figures 20.119 and 20.120.

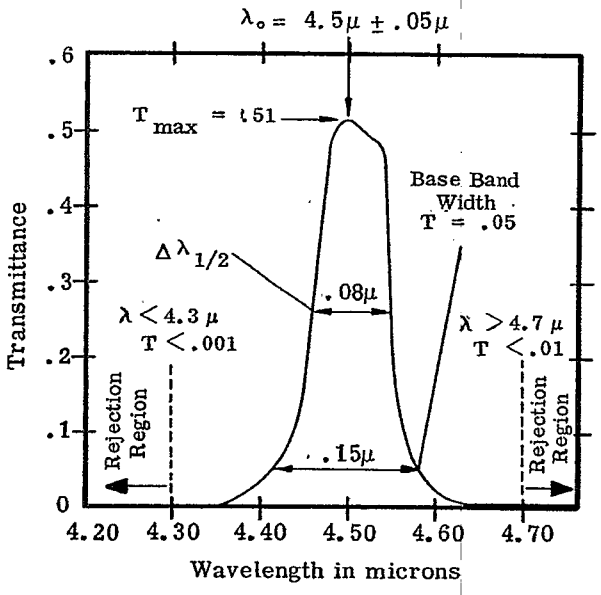


Figure 20.120- Measured spectral transmittance of a narrow band-pass filter which has a nearly rectangular shaped pass band. Courtesy of Eastman Kodak Company.

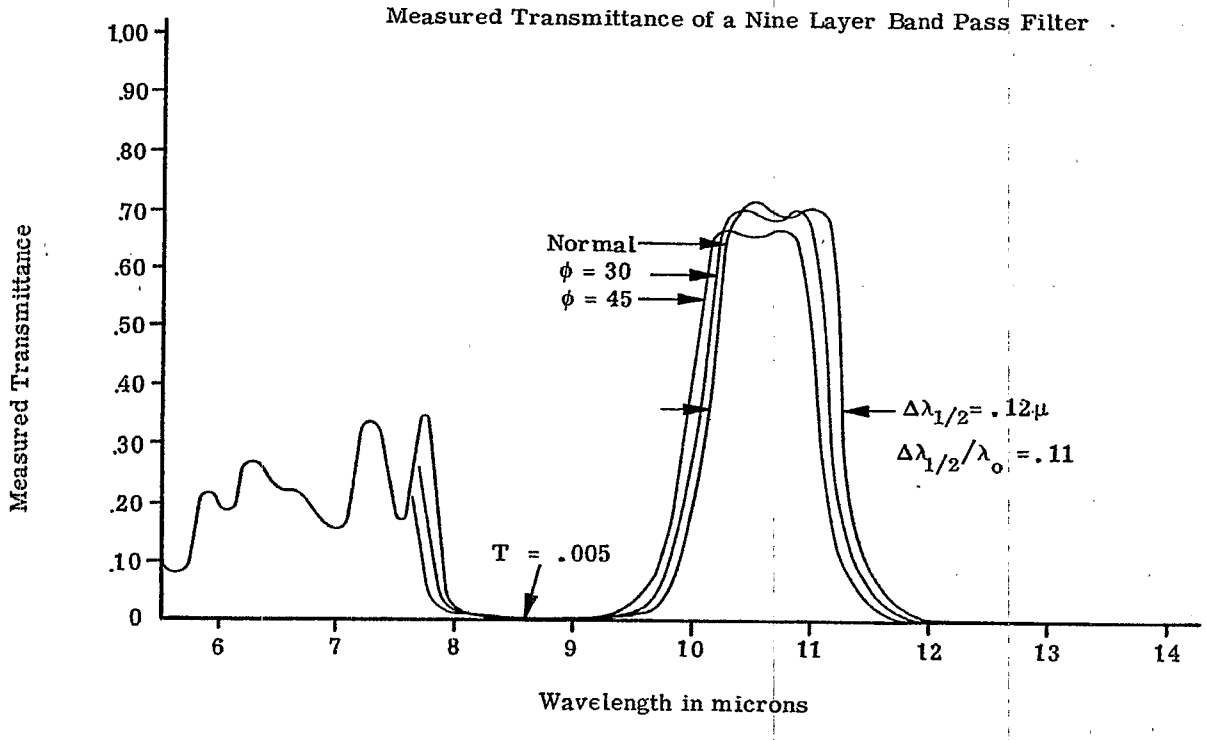


Figure 20.121- Measured spectral transmittance of a narrow band-pass filter which has a nearly rectangular shaped pass band. Courtesy of Bausch and Lomb, Inc.

20.11 REFERENCES FOR FURTHER STUDY

Publications on thin film optics can usually be placed in either of two categories:

(1) Thin film science: This is a study of the fundamental properties of thin films, with an emphasis on knowledge for the sake of knowledge, rather than on knowledge for the sake of ultimately producing some gadget. This includes a study of the physical structure and optical properties of thin films, the thermodynamics and chemistry of their formation, the properties of the films as related to the physics of solids, and so on.

(2) Thin film technology. Section 20 has been devoted to one aspect of this topic, namely how our knowledge of thin films can be utilized to provide useful optical components, such as filters, beam splitters, and other devices. Included in this broad classification are methods and techniques of preparing films, controlling their thickness, methods of multilayer filter design, and so on. The optical applications of thin film technology began to expand rapidly after 1946. A conference on thin film optics was held in 1950 and the papers which were presented at this conference [73] are a good summary of the state of the art up to that time. In 1955 Heavens [1] published a book which not only covers many aspects of thin film science, but he also devotes the later part of this book to some applications of thin films. Therein is presented some of the topics which have not been covered in Section 20, such as multilayer polarizers, frustrated total reflection filters, and so on. Another publication of Heavens [21] surveys more recent developments. Both Weinstein (Welford) [5] and Abeles [12] both present terst and correct mathematical treatments of the theory of the reflectivity of multilayers. Vasicek's book [74] contains some useful information, but very often it is hidden in endless pages of redundant and repetitive derivations of recursion formulae. The practical aspects of depositing thin film coatings are lucidly presented in a Navy Department pamphlet. [75] Holland's book [18] is more recent and presents a vast amount of valuable lore on how to evaporate thin films in a vacuum.

REFERENCES

1. Heavens, O. Optical Properties of Thin Solid Films (Butterworth Scientific Publications, 1955) p. 238
2. ibid, p. 221
3. Strong, John. Concepts of Optics (Freeman, 1958) p. 251
4. Heavens, O. op cit p. 66
5. Weinstein, W. (W. Welford) Vacuum 4, 3 (1954)
6. Terman, F. E. Radio Engineering (McGraw Hill, 1947) 3rd. ed. p. 75
7. Drumheller, Carl "Silicon Monoxide Evaporation Techniques" - A monograph available from the Kemet Company, Cleveland, Ohio. (1960)
8. Drumheller, Carl "Properties and Application of Silicon Monoxide" - A monograph available from the Kemet Company, Cleveland, Ohio. (1960)
9. Weinstein, W. op cit p. 5
10. Heavens, O. op cit Chapter 4
11. Born, M and Wolf, E. - Principles of Optics (Pergamon, 1959) p. 54
12. Abelés, F. Ann. de physique, 5, 103 (1950)
13. Welford, op cit p. 5
14. Pohlack, Hubert "Zum Problem der Reflexionsminderung optisher Gläser bei nichtsenkrechtem Lichteinfall" Jenaer Jahrbuch (VEB Optik Carl Zeiss Jena, 1952)
15. Hass, G. and Tousey, R. J. Opt. Soc. Am. 49, 593 (1959)
16. Greenler, R. G. J. Opt. Soc. Am. 47, 130 (1957)
17. Hass, George and Turner, A. F. "Preparation of Thin Films" - in Volume 6 of Methods of Experimental Physics, L. Marton, Editor in Chief. (Academic Press, 1959)
18. Holland, L. - Vacuum Deposition of Thin Films (Chapman and Hall, 1956)
19. Ballard, S., McCarthy, K. A. and Wolfe, W. L. State-of-the-Art Report: Optical Materials for Infrared Instrumentation. (Report No. 2389-11-S: I.R.I.A, Univ. of Michigan, 1959)
20. Turner, A. F. et al - Thick Thin Films - Quarterly Technical report #4 under contract with U.S. Army Engineer Research and Development Laboratories, Fort Belvoir, Va. (1951)
21. Turner, A. F. and Truby, F. K. U. S. Patent 2,858, 240 (Issued October 1958)

22a Heavens, O. Reports on Progress in Physics, 23,60 (1960)

22b Hass, G., Ramsey, J. B. and Thun, R. J. Opt. Soc. Am. 49, 116 (1959)

23. Huldt, Lennart and Staflin, T. Optica Acta 6, 27 (1959)
24. Hass, G. Vacuum 2, 331 (1952)
25. Pohlack, Hubert. loc cit p. 106
26. Baumeister, P. Optica Acta 8, 105 (1961)

REFERENCES (continued)

27. Hass, George and Turner, A. F. "Coatings for Infrared Optics" in Ergebnisse der Hochvakuumtechnik und der Physik dünner Schichten (Wissenschaftliche Verlagsgesellschaft Stuttgart, 1957)

28. Cox, J. T., Hass, G. and Jacobus, G. F., J. Opt. Soc. Am. 51, 714 (1961)

29. Cox, J. T., Hass, G., and Rowntree, R. F. Vacuum 4, 445 (1954)

30. Turner, A. F., Berning, P. et. al. Infrared Transmission Filters, Quarterly Technical Report #6, under contract DA44-009-eng-1113 with the U. S. Army Engineer Research and Development Laboratories, Fort Belvoir, Va. (1953)

31. Hass, G. and Cox, J. T. Published in Vol. 52 of J. Opt. Soc. Am. (1962)

32. Turner, A. F., Epstein, I., et. al. Optical Properties of Multilayer Films and Interference Filters in the 10 Micron Region. Quarterly Technical report #5 under contract with the U. S. Army Engineer Research and Development Laboratories, Fort Belvoir, Va. (1951)

33. Brillouin, Leon. Wave Propagation in Periodic Structures. (McGraw-Hill, 1946 or Dover, 1953)

34. Seitz, Frederick, The Modern Theory of Solids Chapter 9 (McGraw-Hill, 1940)

35. Epstein, I. J. Opt. Soc. Am. 42, 806 (1952)

36. Baumeister, P. J. Opt. Soc. Am. 48, 955 (1958)

37. Dimmick, G. L. and Widdop, W. E. J. Soc. Motion Picture Engrs. 58, 36 (1952)

38. Carlson, F. E., Howard, G. T., Turner, A. F. and Schroeder, H. H. J. Soc. Motion Picture Engrs. 65, 136 (1956)

39a Turner, A. F. & Schroeder, H. H. J. Soc. Motion Picture Engrs. 69, 351 (1960)

39b Koch, George U. S. Patents 2,552, 184 and 2,552, 185

40. Epstein, L. I. J. Opt. Soc. Am. 45, 360 (1955)

41. Thelen, A. "The Use of Vacuum Deposited Coatings to Improve the Conversion Efficiency of Silicon Solar Cells in Space." Progress in Astronautics and Rocketry Academic Press, (1961) p. 373

42. Sennett, R. S. and Scott, G. D. J. Opt. Soc. Am. 40, 203 (1950)

43. Holland, L. Vacuum 3, 159 (1953)

44. Turner, A. F. and Schroeder, H. H. J. Soc. Motion Picture Engrs. 61, 628 (1953)

45. American Institute of Physics Handbook (McGraw-Hill, 1957) p. 6-108

46. Jenkins, F. A. J. phys. radium 19, 301 (1958)

47. Kuhn, H. and Wilson, B. A. Proc. Phys. Soc. (London) B 63, 754 (1950)

48. Stone, J. M. J. Opt. Soc. Am. 43, 927 (1953)

49. Baumeister, P. W. and Stone, J. M. J. Opt. Soc. Am. 46, 228 (1956)

50. Ring, J. and Wilcock, W. L. Nature 171, 648 (1953)

51. Giacomo, P. J. phys. radium 19, 307 (1958)

52. Giacomo, P. Rev. Opt. 35, 317 (1956)
35, 442 (1956)

REFERENCES (continued)

53. Jenkins, F. A. and White, H. E. Fundamentals of Optics (McGraw-Hill, 1957) Third Ed. p. 273

54. Born, M. and Wolf, E. op cit p. 322

55a. Terman, F. E. op cit p. 39

55b. Fox, A.G. and Li, T. Bell System Tech. J. 40, 453 (1961)

56. Baumeister, P. W. and Jenkins, F. A. J. Opt. Soc. Am. 47, 57 (1957)

57. Terman, F. E. op cit p 42

58. Stone, J. M. Ph. D. Thesis, University of California, Berkeley, 1953 (unpublished)

59. Mielenz, K. D. J. Opt. Soc. Am. 50, 1014 (1960)

60. Smith, S. D. J. Opt. Soc. Am. 48, 43 (1958)

61. Baumeister, P. W., Jenkins, F. A. and Jeppesen, M. A. J. Opt. Soc. Am. 49, 1188 (1959)

62. Geffcken, W. agnew. Chem., A, 60 1 (1948)

63. Geffcken, W. Deutsches Reich Pat. #716,153 (Dec. 1939)

64. Turner, A. F. J. phys. radium 11, 444 (1950)

65. Korolev, F. A., Klement'eva, A. Yu., and Meshcheryakova, T. F. Optics and Spectroscopy (translation of Optika i Spectroskopia) 6, 341 (1960)

66. Turner, A. F. and Ullrich, O. A. J. Opt. Soc. Am. 38, 662 (A) (1948)

67. Mann, A. E. and Rock, F. C. J. Opt. Soc. Am. 38, 280 (A) (1958)

68. Berning, P. H. and Turner, A. F. J. Opt. Soc. Am. 47, 230 (1957)

69. Hadley, L. N. and Dennison, D. M. J. Opt. Soc. Am. 38, 483 (1948)

70. Ring, J., Beer, R., and Hewison, V., J. phys. radium 19, 321 (1958)

71. Dobrowolski, J. J. Opt. Soc. Am. 49, 794 (1959)

72a Lissberger, P. H. J. Opt. Soc. Am. 49, 121 (1959)

72b Lissberger, P. H. and Wilcock, W. L. J. Opt. Soc. Am. 49, 126 (1959)

72c Hass, G. and Salzburg, C. D. J. Opt. Soc. Am. 44, 181 (1954)

73. J. phys. radium 11, 305-480 (July 1950)

74. Vasicek, A. Optics of Thin Films (North-Holland, 1960)

75. Naval Ordnance Pamphlet (OP) 1952, Optics Filming, U.S. Gov. Printing Office, (1945)