Chapter 1

Introduction

The process of designing a thin-film coating can be approached from an analytical or a scientific basis, a numerical or an actual experience, or—with the capability of present thin-film design software—perhaps with no basis (knowledge) at all. Over many decades, several design, analysis, and other related books and technical articles have been published on this subject. As more optical systems have become multispectral with performance requirements at several wavelengths or wavelength regions, the required coatings have become more complex. A particular subset or type of coating design is typically used to highly reflect at several wavelengths and may also have high transmittance at others. Classical designs, such as bandpass, cavity, notch, minus filters, rugate, etc., are routinely used for these applications. This chapter is a brief review of well-established thin-film theory as it relates specifically to the above subset of coating designs.

1.1 Review of Mathematics for Thin-Film Design

The basic physics of optical thin films involves the interaction of one or more wavelengths of light and real media boundaries. Without these boundaries, Maxwell's equations describe properties of light waves that are typically modeled as planar or Gaussian. With boundary conditions, where different dielectric or conductive media are present, Maxwell's equations are solved accordingly to describe the reflected and transmitted waves. Specifically, the complex reflection coefficient is the amplitude ratio of the reflected to incident electric fields. From this coefficient, the reflectance (which is the ratio of the reflected wave's power to the incident wave's power) and the phase shift between the reflected and incident waves are determined. The rigorous mathematical development of the electromagnetic theory for propagating waves and interactions with boundaries for various materials is covered in other texts on electromagnetics[1] and optics.[2]

Beyond the fundamental electromagnetic concepts of propagating waves or beams and media and boundary conditions, single or multiple layers of so-called optical "thin" films have two or more boundaries with relatively small spacing or distance between the boundaries (i.e., this spacing is approximately on the order of the wavelength of the light source). These boundaries or interfaces cause a plurality of reflected waves that can interfere with each other in amplitude and phase. These reflected waves interfere, or add coherently, because the path differences within thin films are less than the coherence length, given by Eq. (1.1):[3]

$$l \approx \frac{\lambda^2}{\Delta\lambda}. \tag{1.1}$$

Here, l is the coherence length, λ is the wavelength of light, and $\Delta\lambda$ is the bandwidth. Over the short distances of thin films, most light sources are coherent and

interference is observed. For larger distances (e.g., a millimeter) and for quasi-monochromatic light sources, interference can also be observed (the path length is still less than the coherence length). However, for path differences larger than the coherence length, multiple waves interfere with each other in intensity only (in an incoherent manner).[3] Still, the observation of interference effects depends on the detection conditions. This text investigates thin-film designs based on the optical interference of thin films, where an incident wave of light is multiplied into groups of reflected and transmitted waves.

The spectral performance of optical thin-film designs can be computer-modeled using any one of four methods or models (electric field matrix, characteristic matrix, reflection recursion, and admittance recursion).[4] For the purpose of this text, the characteristic matrix is used to model thin-film designs for the particular case of nonabsorbing media. The derivation of the characteristic matrix is not included because it is thoroughly covered in other thin-film texts. However, for completeness, definitions of characteristic matrix, admittance, and the calculation of reflectance are provided.[5]

From Ref. [5], the characteristic matrix for any isotropic and homogeneous thin-film layer is given by

$$\begin{bmatrix} \cos\delta_L & (i\sin\delta_L)/\eta_L \\ i\eta_L \sin\delta_L & \cos\delta_L \end{bmatrix}, \tag{1.2}$$

where the phase term δ_L is given by

$$\delta_L = \frac{2\pi N_L t_L}{\lambda}\cos(\theta_L). \tag{1.3}$$

N_L is the complex refractive index of the layer given by

$$N_L = n_L - ik_L, \tag{1.4}$$

where n_L and k_L are the refractive index and extinction coefficient, respectively, and t_L is the layer's physical thickness. Using Snell's law, the internal angle θ_L is readily found from the incident angle of the beam θ_o, from the complex refractive index of the incident medium N_o, and from N_L, where

$$N_o \sin(\theta_o) = N_L \sin(\theta_L). \tag{1.5}$$

The optical admittance of the layer, η_L, is dependent on the polarization of the incident beam and is given by

$$\text{s-polarization:} \quad \eta_L = N_L \cos(\theta_L) \tag{1.6}$$

and

$$\text{p-polarization:} \quad \eta_L = N_L/\cos(\theta_L). \tag{1.7}$$

It should be noted that the term for optical admittance of free space was omitted from Eqs. (1.6) and (1.7) because it cancels when solving for the admittance of the film-substrate optical system at the interface of the ambient medium and film. For reference, the optical admittance of free space Y_0 is the reciprocal of its optical impedance of free space Z_0 given by

$$Y_0 = \frac{1}{Z_0} = \sqrt{\frac{\varepsilon_0}{\mu_0}}, \tag{1.8}$$

where ε_0 and μ_0 are the permittivity and permeability of free space, respectively.

The general form of the optical system described has an ambient medium and thin-film layer(s) deposited onto a substrate as shown in Fig. 1.1. A monochromatic light wave or beam in the ambient medium impinges on the thin-film layer(s) and substrate and is subsequently reflected and transmitted. By solving for the electric and magnetic field vector amplitudes at the ambient-film (outer layer #1, see Fig. 1.1) interface, the optical admittance for the film-substrate system is determined from its characteristic matrix. From this film-substrate's admittance and that of the ambient medium, the reflectance and transmittance are determined.

As one could infer from Fig. 1.1, a small lateral shift of a reflected beam could occur. Departing from plane wave theory, for the special case of total internal reflection (TIR), a shift does occur. This shift is called the Goos-Hänchen shift.[6] However, this shift is typically ignored in thin-film modeling.

Without going through the derivation, the general form of the characteristic matrix for a multilayer film-substrate system is given by the matrix

$$\begin{bmatrix} E_A \\ H_A \end{bmatrix}, \tag{1.9}$$

where

$$\begin{bmatrix} E_A \\ H_A \end{bmatrix} = \left\{ \prod_{L=1}^{n} \begin{bmatrix} \cos\delta_L & (i\sin\delta_L)/\eta_L \\ i\eta_L \sin\delta_L & \cos\delta_L \end{bmatrix} \right\} \begin{bmatrix} 1 \\ \eta_S \end{bmatrix} \tag{1.10}$$

and η_S is the admittance of the substrate medium given by

$$\eta_S = \frac{H_S}{E_S}. \tag{1.11}$$

As shown in Eq. (1.10), the corresponding multilayer thin film-system matrix is the product of the characteristic 2×2 matrices for each layer in the system. The subscript L denotes the Lth layer, where $L = 1$ is the adjacent layer to the ambient medium. A multiplicative factor is not shown in Eq. (1.10) because it is not required to calculate the optical admittance of the film-substrate system as defined below.

The optical admittance of the film-substrate system is

$$Y = \frac{H_A}{E_A}. \tag{1.12}$$

The resulting product matrix of the 2×2 characteristic matrices from Eq. (1.10) takes the general form

$$\begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}. \tag{1.13}$$

By substituting Eq. (1.13) into Eq. (1.10) and using Eq. (1.12), the optical admittance Y for the film-substrate system can be determined, where

$$Y = \frac{p_{21} + p_{22}\eta_S}{p_{11} + p_{12}\eta_S}. \tag{1.14}$$

From the admittance Y of the ambient-coating-substrate system, the complex amplitude reflection coefficient r is determined from Eq. (1.14), where

$$r = \frac{\eta_A - Y}{\eta_A + Y}. \tag{1.15}$$

The optical admittance of the ambient medium, η_A, is determined from Eq. (1.6) or (1.7). Again, note that the optical admittance for each thin-film layer and ambient and substrate media are determined from Eqs. (1.6) and (1.7), and they depend on the incidence angle and polarization state of the incident beam.

The ratio of the reflected beam's power to that of the incident beam, or the reflectance, is given by

$$R = r \times r^*. \tag{1.16}$$

In the case of a nonabsorbing, scatter-free coating-substrate system, the transmittance is determined by

$$T = 1 - R. \tag{1.17}$$

The reflectance can be determined from Eq. (1.15) for the p- and s-polarization states, namely r_p and r_s, respectively. At oblique incidence angles, the differential phase shift Δ is determined from

$$\Delta = \delta_p - \delta_s, \tag{1.18}$$

where

$$\delta_{p,s} = \arg(r_{p,s}). \tag{1.19}$$

At normal incidence, the differential phase shift is either zero or π radians depending on the choice of phase convention: either Abelès (thin film) or Nebraska-Muller[7](ellipsometry), respectively.

1.2 Analytical, Discrete-layer, Thin-Film Designs

Several types of classical thin-film designs achieve prespecified spectral performances based on analytical equations or known layer sequences. Even without the use of a computer and an algorithm to find or optimize an initial coating design, these thin-film coatings are readily designed.[8] Some very basic types of analytical coating designs are shown in Table 1.1.

The basic design types shown in Table 1.1 are written in common symbolic notation, where L, M, and H represent low, medium, and high refractive index layers, respectively. These layers are specified as quarter-wave optical thickness (QWOT), where

$$T_L = 4N_L t_L \cos\theta_L. \tag{1.20}$$

T_L is the thickness (QWOT) of a layer with refractive index N_L, physical thickness t_L, and incidence angle θ_L. Refractive-index profile plots show the refractive index of discrete film layers as a function of optical thickness, as shown in Figs. 1.2(a)–(e).

Table 1.2 lists some specific design types that are similar to designs presented in this text.

Figure 1.2 shows refractive-index profile plots for several of these designs. Also shown are optical-thickness (QWOT) profile plots as a function of the layer

Table 1.1 Some basic thin-film designs.

Coating	Common Purpose or Use	Some Basic Design Types
Antireflection (AR)	Windows	Ambient/L/Sub Ambient/LML/Sub Ambient/LMHL/Sub
Partial reflector (PR)	Laser output couplers Beamsplitters	Ambient/xL $(HL)^n$/Sub
High reflector (HR)	Laser cavity reflectors Corner mirrors	Ambient/$(HL)^n$/Sub
Polarizers	Separate polarization states	Ambient /$(HL)^n$/Sub
Dichroic filters	Separate wavelength bands Color filters	Ambient/$(0.5L\ H\ 0.5L)^n$/Sub Ambient/$(0.5H\ L\ 0.5H)^n$/Sub
Bandpass filters	Wavelength separation	Ambient/$(LH)^n\ 2L\ (HL)^n$/Sub

Table 1.2 Selected thin-film designs.

Coating	Some Basic Design Types	Typical Variations
High reflector (HR)	Ambient /$(HL)^n$/Sub	High-order quarter-wave stacks Multiple stacks Arithmetic or geometric stacks (Ref. [9])
Dichroic filters [short-wave pass (SWP), long-wave pass (LWP)]	Ambient /$(L/2\ H\ L/2)^n$/Sub Ambient /$(H/2\ L\ H/2)^n$/Sub	Combined SWP and LWP Nonpolarizing
Bandpass filters	Ambient /$(LH)^n\ 2L(HL)^n$/Sub	Single cavity Multiple cavity (Ref. [4])
Minus filters	Ambient /$(aL/2\ bH\ aL/2)^n$/Sub where $a + b = 2$	Ref. [10]

number, which correspond to the refractive-index profiles. Refractive-index profile plots are useful to simultaneously visualize optical thickness and refractive index for individual discrete layers in a given coating design. In addition, electric-field amplitude curves are frequently overlaid on refractive-index profile plots, as shown in Fig. 1.2(a). However, for refined or complex analytical designs, optical-thickness profile plots can provide better visualization of layer thickness and layer thickness patterns than refractive-index profiles. The former is introduced and both are utilized in this text for the discrete layer designs presented.

The analytical designs shown in Table 1.2 and Fig. 1.2 are of interest in this text, but many other types of classical, analytical, and numerical coating designs are not covered. Instead, the concept of some specific analytical designs is included to show the classical designs that are related to the designs developed here. Next, we might ask the question "Is this design ready to be manufactured?" In some cases, analytical designs are not ready because of additional issues such as availability of film materials with the desired physical properties and dispersion of the selected materials. Typically the next step toward readying the design for manufacture is computer refinement.

1.3 Thin-Film Design Methods

This section first discusses the typical design process for thin-film coatings. A flowchart for the design of a thin-film coating is shown in Fig. 1.3. It is of particular interest to look at how the preliminary design is determined. Here, the analytical path allows the designer to select from classical designs or to derive designs from *a priori* analytical methods. Alternatively, a design may already exist that meets or performs very close to the required specifications (empirical basis). Last, using computer software, the designer can refine or optimize an initial empirical or analytical starting design.[11] With newer algorithms and faster computers, the computer

can then "create" a design; that is, given the layer, ambient and substrate materials, and spectral performance specifications, the algorithm can determine the number, thickness, and sequence of layers. However, it is debatable whether any, some, or all of the coating designs created by the latter process are "good" designs that are ready to be manufactured.

This text deals primarily with new analytical methods for discrete thin-film coatings along with some refinement of these initial designs. To provide insight and background information for some of the particular analytical methods covered, the last section in Chapter 1 covers rugate and synthesized rugate coatings. This discussion departs from the premise of this text on discrete thin-film designs, but rugate coatings and theory are closely related to the discrete-layer design methods presented.

1.4 Inhomogeneous Coating Designs and Synthesis

Early work on inhomogeneous thin films was done by Jacobsson et al.[12] The characteristic matrix for an inhomogeneous layer[13] is similar in form to Eq. (1.10) but rather complicated. Typically, inhomogeneous layers are computer-modeled by subdividing the given inhomogeneous layer into discrete, homogeneous ones.[14] Next, the quasi-exact spectral performance is readily found from Eq. (1.10). Furthermore, Ref. [14] describes how an inhomogeneous layer can be synthesized by optimizing the refractive index of the subdivided homogeneous layers to achieve a desired spectral performance (Snedaker). A brief review of rugate and synthesized rugate coatings is covered in Sec. 1.4.1. In addition to rugate designs, an inhomogeneous refractive-index profile can be determined for a given spectral curve or performance. The well-known Fourier transform method is presented in Sec. 1.4.2.

1.4.1 Rugate and synthesized rugate designs

Rugate coatings typically consist of a single layer deposited onto a substrate where the refractive index of the layer continuously changes as a function of depth. The nonabsorbing rugate film layer's refractive index is a function of the layer's thickness. A typical index profile of a rugate design is given in Eq. (1.21):[15]

$$n(x) = n_a + \frac{n_{pv}}{2} \sin\left(\frac{2\pi x}{n_a T} + \theta\right). \tag{1.21}$$

The basic sinusoidal refractive-index profile of a rugate filter (see Fig. 1.4) is determined from Eq. (1.21), where n_a and n_{pv} are the average and range of minimum-to-maximum refractive index, respectively; $n_a T$ is the period; θ is a phase term; and x is the optical distance from the substrate. The period of the sine wave is of interest since the wavelength of the single high-reflectance stopband produced with a rugate filter is determined. Specifically, the center wavelength of this stopband is twice the period, or

$$\lambda = 2n_a T, \tag{1.22}$$

where the period is defined in terms of optical thickness. By superimposing several sine waves for the refractive-index profile, a rugate coating produces several stopbands that correspond to each of the individual sine waves.

One particular enhancement of rugate coatings greatly reduces or suppresses reflection ripples or sidelobes in all passband wavelength regions, as shown in Fig. 1.5. This enhancement was defined by Southwell[16] as apodization, which utilizes envelope functions on sinusoidal thickness profiles to suppress sidebands. Apodization envelope functions can be Gaussian, sine, linear, or other functions.[15] An example of an envelope function from Ref. [11] is shown in Fig. 1.6 and applied to a sine wave-modulated index profile as determined by Eq. (1.21), where

$$n(x) = n_a + \left\{ \exp\left[-\beta \left(x - \frac{N n_a T}{2} \right)^2 \right] \frac{n_{pv}}{2} \sin\left(\frac{2\pi x}{n_a T} + \theta \right) \right\}. \tag{1.23}$$

In Eq. (1.23), β is a phase term and N is the number of periods in the sine wave.

The use of wavelets was proposed by Southwell[17] to design several types of apodized rugate coatings. One application, discussed in more detail in Chapter 4, increases the bandwidth of reflector designs.

1.4.2 Fourier transform method of inhomogeneous coating design

Liddell[9] summarized the different types of exact and approximate Fourier transform synthesis methods. Dr. Pierre Verly developed a Fourier transform method and software for the design of the inhomogeneous coatings first proposed by Delano and Sossi.[18] As stated in Ref. [18], the analysis of a given inhomogeneous coating with the refractive-index profile $n(x)$ is exact: The analytical solution to the corresponding Maxwell's equations uniquely determines the spectral performance. A brief description of this Fourier transform method is given below.

To synthesize or determine the required inhomogeneous $n(x)$ profile that produces a known spectral performance, loss and dispersion are excluded. Based on these conditions, the $n(x)$ profile that produces a desired transmittance is given by[18]

$$n(x) = n_0 \exp\left[\frac{i}{\pi} \int_{-\infty}^{\infty} \frac{Q(T,k)}{k} e^{-ikx} dk \right], \tag{1.24}$$

where $Q(T,k)$ is a complex function (commonly referred to as the Q-function), $k = 2\pi/\lambda$ is the wave number, λ is the wavelength, and x is twice the optical thickness. $T(k)$ is typically defined over a finite wavelength range and set to the value of 1 (unity) outside this range ($Q = 0$ when $T = 1$). Here, several forms of the Q-function have been used to overcome limitations, such as when the desired transmittance is near zero (high reflectance). It should be noted that it is necessary to neglect multiple reflections (multiple integrals) to obtain Eq. (1.24).[18] The National Research Council (NRC) of Canada, Bovard,[19] and others have reported

several methods that solve for $n(x)$ for almost any given spectral performance, and even with high or low reflectance. For example, one Q-function useful for high reflectance was reported by Verly et al.,[20] where

$$|Q| = w\sqrt{1-T} + (1-w)\sqrt{\frac{1}{T}-1} \tag{1.25}$$

and

$$0 \leq w \leq 1. \tag{1.26}$$

In Eq. (1.25), T is the desired transmittance and w is an empirically determined weighting factor that is adjusted to achieve a best fit between the desired and calculated transmittance. It is well known that Q-functions are only approximate, and additional methods must be used to correct for the resulting errors. The phase of the Q-function is a function of both reflection and transmission phases, which are usually unknown. Depending on different transmission parameters, k, or the phase of the Q-function, the resulting refractive-index profile will be different.[21] Still, accurate solutions are readily obtained for most transmission parameters. Once the inhomogeneous $n(x)$ profile is determined, an approximate, discrete-layer design can be determined (i.e., discretization). A complete discussion of this method is provided in Refs. [18] and [20].

1.5 Summary

In this chapter, the mathematics for one method to model thin-film performance—namely, the characteristic matrix—was presented first, along with the optical admittance of a multilayer (general case) film-substrate system in order to determine reflectance and transmittance. This swift treatment of electromagnetic and thin-film physics was by no means an attempt to trivialize it; quite simply, the extensive details, development, special cases, etc., of thin-film theory are readily available in several texts. The above presentation's format quickly develops Eqs. (1.10) and (1.14)–(1.19) in a format that is readily usable for the computation of reflectance and transmittance of multilayer thin-film designs (see Appendix A for useful forms of these equations).

Next, discrete-layer thin-film coating design types were presented, with emphasis on dichroic, bandpass, notch, and band-reject filters. The general design method was included to complete the connection between analytical designs and computer refinement methods.

Lastly, in order to provide the reader with the essential background information necessary for the design methods presented, rugate and inhomogeneous coating design methods were reviewed. Although the design methods presented later in this text are specifically for discrete-layer designs, a close connection exists between these designs and inhomogeneous coating designs.

1.6 Exercises

1. Two polished, transparent glass monitor chips are stuck together or optically contacted. What is the maximum spacing between these chips (air gap) where interference patterns can be observed at $\lambda = 550$ nm? Assume $\Delta\lambda = 1$ nm.
2. An uncoated 1-mm thick ZnSe, plano/plano window ($n = 2.403$ at $\lambda = 10.6\,\mu$m) has a measured transmittance of 68.9% in air at normal incidence. Derive the equation for total transmittance (assuming optical interference between the two surfaces) and determine the total transmittance.
3. Starting with Eq. (1.10), where the electric and magnetic field vectors are determined at the surface boundary of a single layer stack, replace the optical admittance of the layer η with NY_0, where Y_0 and $\boldsymbol{N}$ are the optical admittance of free space and the complex refractive index of the layer, respectively. Solve for the optical admittance of the system, Y, for normal incidence.
4. Derive the optical admittance for a single, nonabsorbing layer for normal incidence and for arbitrary p- and s-polarizations.
5. Write a thin-film program that calculates reflectance and transmittance for any given incidence angle, polarization, ambient medium, multilayer thin-film design, and substrate.
6. Optimize any starting design (or no design) with two thin-film materials to produce a dual-wavelength high reflector, multilayer dielectric coating with stopbands at wavelengths 1064 nm and 780 nm, with reflectance greater than 99.8% at normal incidence. Assume nondispersive, nonabsorbing refractive indices of 1.0, 1.46, 2.25, and 1.52 for the ambient, the two layer materials, and the substrate. How many layers were required? What are the maximum and minimum layer thicknesses (if any) in the final design compared with the original design?
7. Synthesize a rugate design (homogeneous layers). Using thin-film software, set up 20 nondispersive coating material files with equally spaced refractive- index values between 1.46 and 2.25. Next, using Eqs. (1.21) and (1.22), determine the required layer thickness and period needed to produce a stopband at $\lambda = 1\ \mu$m with reflectance greater than 99%.
8. Plot the reflectance of the design from Exercise 5 within the wavelength range of $\lambda = 0.5 - 2\ \mu$m. Using Gaussian apodization [Eq. (1.23)] on the same design, plot the reflectance again and compare reflectance ripples in the passbands.
9. Explain why a Q-function would not produce unique solutions for a reflectance profile with high reflectance.

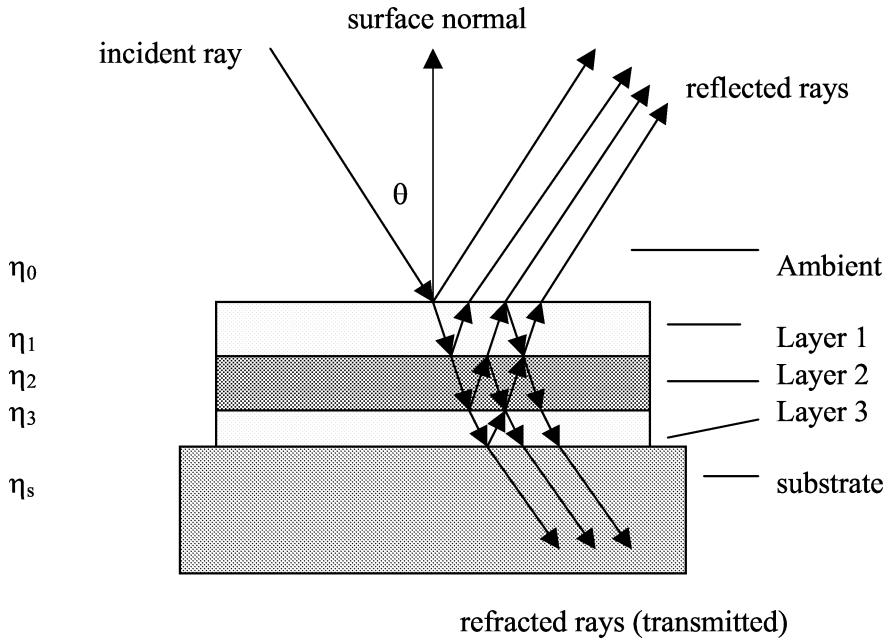

Figure 1.1 Typical cross-section ray trace diagram for an arbitrary, multilayer thin film-substrate system.

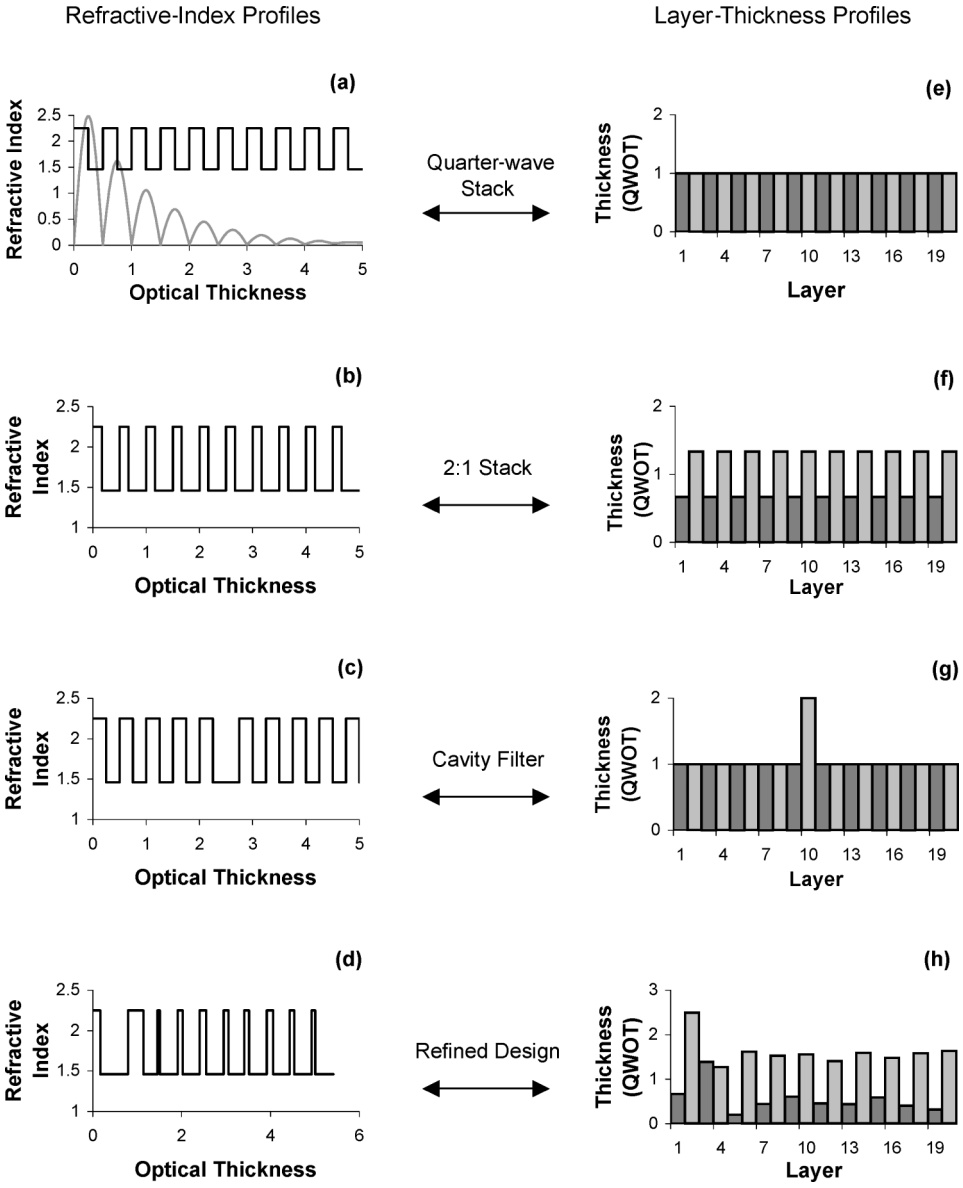

Figure 1.2 Refractive-index and layer-thickness profiles for four common designs. (a)–(d): refractive-index profiles as a function of optical thickness (waves); (e)–(h): corresponding layer-thickness profiles with dark and light shaded bars representing high and low refractive index. In (a), the parallel electric field is overlayed (arbitrary units).

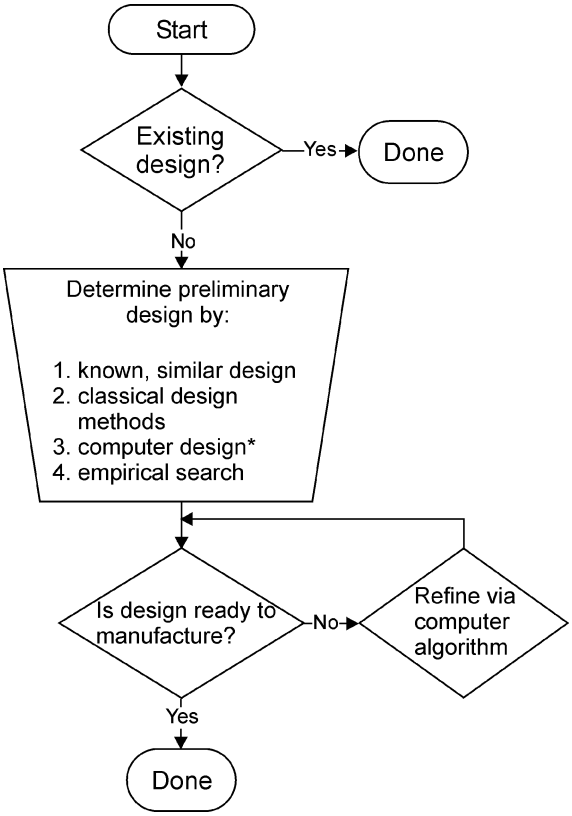

Figure 1.3 Generic flowchart for thin-film design options.
*A computer determines the design from preselected materials.

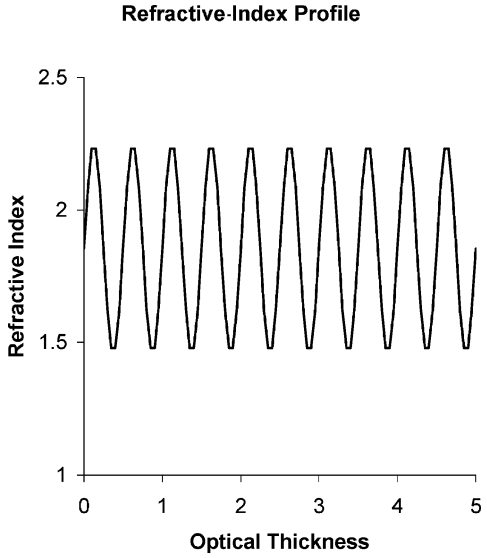

Figure 1.4 Refractive-index profile of a rugate coating.

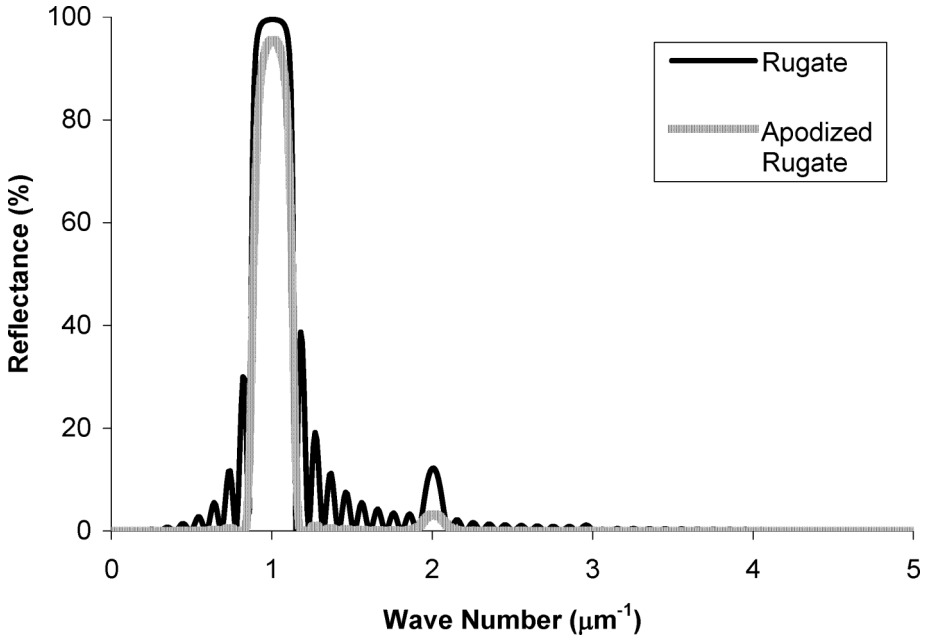

Figure 1.5 Spectral reflectance of rugate and apodized rugate coatings (Figs. 1.4 and 1.6, respectively). Note the supression of high-order stopbands.

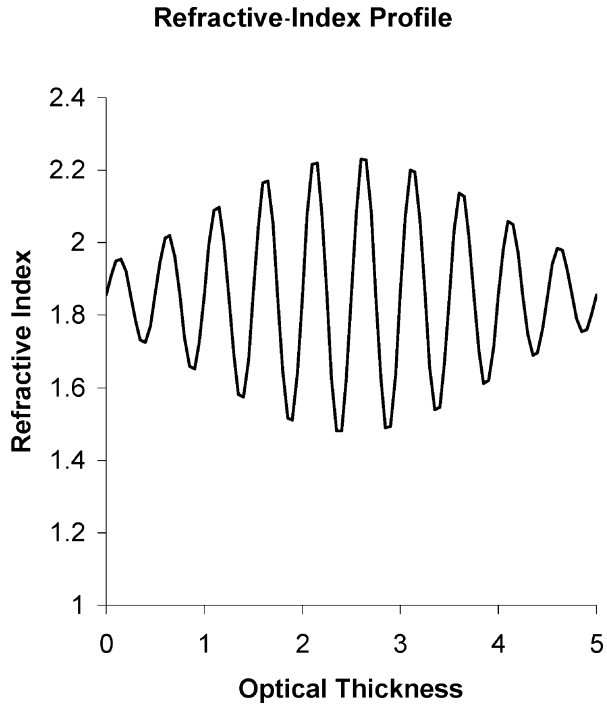

Figure 1.6 Refractive index of a Gaussian-apodized rugate coating.