Springer Series on Atomic, Optical, and Plasma Physics 87

John Lekner

Theory of Reflection

Reflection and Transmission of Electromagnetic, Particle and Acoustic Waves

Second Edition

Springer

Springer Series on Atomic, Optical, and Plasma Physics

Volume 87

The Springer Series on Atomic, Optical, and Plasma Physics covers in a comprehensive manner theory and experiment in the entire field of atoms and molecules and their interaction with electromagnetic radiation. Books in the series provide a rich source of new ideas and techniques with wide applications in fields such as chemistry, materials science, astrophysics, surface science, plasma technology, advanced optics, aeronomy, and engineering. Laser physics is a particular connecting theme that has provided much of the continuing impetus for new developments in the field, such as quantum computation and Bose-Einstein condensation. The purpose of the series is to cover the gap between standard undergraduate textbooks and the research literature with emphasis on the fundamental ideas, methods, techniques, and results in the field.

More information about this series at http://www.springer.com/series/411

John Lekner

Theory of Reflection

Reflection and Transmission of Electromagnetic, Particle and Acoustic Waves

Second Edition

Springer

John Lekner
School of Chemical and Physical Sciences
Victoria University of Wellington
Wellington
New Zealand

Previously published as Vol. 3 in "Developments in Electromagnetic Theory and Applications" (1987), ISBN 978-90-247-3418-4

ISSN 1615-5653 ISSN 2197-6791 (electronic)
Springer Series on Atomic, Optical, and Plasma Physics
ISBN 978-3-319-23626-1 ISBN 978-3-319-23627-8 (eBook)
DOI 10.1007/978-3-319-23627-8

Library of Congress Control Number: 2015950446

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

for Isla, Romy and Tomas

Preface

The second edition is an enlarged and updated version of the book I completed in Canberra in June 1986. There are six new chapters, *Uniaxial anisotropy*, *Ellipsometry*, *Periodically stratified media*, *Neutron and X-ray reflection*, *Acoustic waves* and *Chiral isotropic media*. A first edition chapter has been split into two, dealing with *Pulses* and *Finite beam*s separately. The chapters on matrix methods and on numerical methods have been combined into one. The former appendix is now the chapter *Particle waves*, preceding that on neutron and X-ray reflection. The second edition contains 20 chapters, some with their own appendices, compared with 13 chapters and one appendix in the first edition.

The aim remains the same: to present the theory of reflection and transmission of waves from and through (mainly) planar stratifications in a simple and physical way, from first principles. By that I mean from the Maxwell or Schrödinger equations, for instance. As a theorist, I have naturally favoured exact results and have emphasized universal conservation and invariance properties. However, many particular cases are made explicit in graphs and formulae. That's where the theory connects with reality (as revealed by experiment), and where one gets a physical feel for the meaning of the formulae. Applied topics do appear: two examples are the important phenomenon of attenuated total reflection in Chap. 10, and the reflectivity of multilayer dielectric mirrors in Chaps. 12 and 13.

I have tried to maintain a logical progression throughout, rather than a historical one. Nevertheless, due credit is given to the pioneers of the subject of wave reflection. Rayleigh (John William Strutt, 3rd Baron Rayleigh, 1842–1919) features prominently, as may be expected given the influence of his work, especially of his *Theory of sound*. Even so, some of his reflection papers seem to have been forgotten and his results keep being rediscovered, often in inferior form. The Rayleigh (or weak reflection) approximation is an example, and appears frequently throughout the book.

Rayleigh was of privileged birth and made the most of the consequent opportunities. Not so privileged was George Green (1793–1841), the baker's and later miller's son. He was almost entirely self-taught, having just one year of formal

schooling as a child, between the ages of 8 and 9, and becoming a Cambridge undergraduate when nearly 40. Green's functions form the basis of the perturbation theories for long waves in Chap. 3 and for short waves in Chap. 6. No surprise there. But who talks of the Liouville–Green wavefunctions, or who has heard of Green's angle? The former are the high-frequency waveforms dating back to 1837. Green's angle, as I have called it in Sect. 1.4, is the acoustic analogue of the Brewster angle, at which one polarization has zero reflectance from a sharp interface.

Rayleigh's use of k for wavenumber has become the standard, and I have built on that to maintain a consistent notation throughout the book, as far as possible. The normal and tangential components of the wavevector $\boldsymbol{k}$ are always labelled q and K; the latter is special in being an invariant for waves in plane-stratified media, with the laws of reflection and transmission consequent from that invariance. Greek letters are used (not exclusively, but in preference) for dimensionless quantities.

The book is written for scientists and engineers whose work involves wave reflection or transmission. Most of the chapters are in the language of electromagnetic theory, but many of the electromagnetic results can be applied to particle waves, specifically to those satisfying the Schrödinger equation. The mathematical connection between electromagnetic s (or TE) waves and quantum particle waves is established in Chap. 1. The main results for s waves are translated into quantum mechanical language in the Chap. 15. There is also a close analogy between acoustic waves and electromagnetic p (or TM) waves, as shown in Sect. 1.4, and in detail in Chap. 17. Thus the book, though primarily intended for researchers working in optics, microwaves or in neutron or X-ray optics, will be of use to physicists, chemists and electrical engineers studying reflection and transmission of particles at potential barriers, and also to those working in acoustics, oceanography and seismology.

Chapter 1 is recommended for all readers: it introduces reflection phenomena, defines the notation and previews (in Sect. 1.6) the contents of the rest of the book. The reader can then go to any other chapter in the book, all of which are intended to be sufficiently self-contained so that only occasional reference to other parts of the book is needed.

The first edition was written at the Department of Applied Mathematics of the Australian National University, Canberra. In the Preface I had the pleasure of thanking two Australians, Barry Ninham and Colin Pask. The second edition was written in New Zealand, but I again have pleasure of thanking two Australians, this time Tony Klein and Andrew Wildes, for their suggestions and comments on the new chapter on X-ray and neutron reflection.

Wellington John Lekner

Contents

Chapter 1
Introducing Reflection

Electromagnetic, acoustic and particle waves all scatter, diffract and interfere. Reflection is the result of the constructive interference of many scattered or diffracted waves originating from scatterers in a stratified medium. This fundamental many-body approach is hard to apply (two illustrations are given in Sect. 1.5). Usually one replaces the collection of scatterers by an effective medium whose properties are represented, as far as wave propagation is concerned, by a function of position and frequency (or energy), such as the dielectric function ε in the electromagnetic case, or the effective potential V in the quantum particle case. Electromagnetic and particle waves then satisfy the same kind of linear partial differential equation, with ε and V playing similar roles.

In a medium with planar stratification the functions ε and V depend on only one spatial variable, and the partial differential equations then separate. Snell's Law is a direct consequence of this separability of the spatial dependence, or equivalently, of the invariance of the reflecting material with respect to translations along the surface. The differential equations, and the elementary reflection properties which follow from them, are derived for electromagnetic, particle, and acoustic waves in the first four sections. The many-body, constructive interference, aspect of reflection is outlined in Sect. 1.5. Finally, Sect. 1.6 previews some of the main results in Chaps. 2–20.

1.1 The Electromagnetic *s* Wave

The reflection of a plane electromagnetic wave at a planar interface between two media is completely characterized when solutions for two mutually perpendicular polarizations are known. The polarizations conventionally chosen are: one with its electric vector perpendicular to the plane of incidence (labelled *s*, from the German *senkrecht*, perpendicular), and the other with its electric vector parallel to the plane of incidence (labelled *p*).

We consider monochromatic waves, of angular frequency ω. The reflection of a general electromagnetic wave (a pulse, for example) can be analysed as that of a superposition of monochromatic waves. For a given ω the time dependence of all

J. Lekner, *Theory of Reflection*, Springer Series on Atomic, Optical, and Plasma Physics 87, DOI 10.1007/978-3-319-23627-8_1

fields is carried in the factor $e^{-i\omega t}$. (This is the convention in quantum and solid state physics, and much of optics. In radio and electrical engineering the factor $e^{i\omega t}$ is often used. With the convention used here the dielectric function has positive imaginary part in the case of absorption.) We will mostly consider *non-magnetic* media in this book. The electrodynamic properties of a medium are then contained in the dielectric function $\varepsilon(\mathbf{r}, \omega)$ which is the ratio of the permitivity of the medium at position $\mathbf{r}$ and angular frequency ω to that of the vacuum. The wave equations follow from Maxwell's two curl equations relating the electric field $\mathbf{E}$ and the magnetic field $\mathbf{B}$:

$$\nabla \times \mathbf{E} = i\omega \mathbf{B} \qquad \text{or} \qquad \nabla \times \mathbf{E} = i\frac{\omega}{c}\mathbf{B}, \tag{1.1}$$

$$\nabla \times \mathbf{B} = -i\varepsilon \frac{\omega}{c^2}\mathbf{E} \qquad \text{or} \qquad \nabla \times \mathbf{B} = -i\varepsilon \frac{\omega}{c}\mathbf{E}. \tag{1.2}$$

(The equations on the left are in SI units, those on the right in Gaussian units; the difference lies in the positioning of the speed of light c. In reflection studies, theory and experiment deal in dimensionless ratios, which are independent of the choice of units. Even the formal distinction disappears from (1.5) onward.)

For a planar interface lying in the xy plane, and an electromagnetic wave propagating in the x and z directions, the s wave has $\mathbf{E} = (0, E_y, 0)$ and (1.1) gives

$$-\frac{\partial E_y}{\partial z} = i\frac{\omega}{c}B_x, \qquad \frac{\partial E_y}{\partial x} = i\frac{\omega}{c}B_z, \tag{1.3}$$

and $B_y = 0$. The other curl equation gives

$$\frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x} = -i\varepsilon \frac{\omega}{c}E_y. \tag{1.4}$$

On eliminating B_x and B_z from (1.3) and (1.4), we obtain a second order partial differential equation for E_y,

$$\frac{\partial^2 E_y}{\partial x^2} + \frac{\partial^2 E_y}{\partial z^2} + \varepsilon \frac{\omega^2}{c^2}E_y = 0. \tag{1.5}$$

For planar stratifications the dielectric function depends on one spatial variable, z. The partial differential equation is then separable, with

$$E_y(x, z, t) = e^{i(Kx - \omega t)}E(z), \tag{1.6}$$

where $E(z)$ satisfies the ordinary differential equation

$$\frac{d^2 E}{dz^2} + q^2 E = 0, \qquad q^2 = \varepsilon \frac{\omega^2}{c^2} - K^2 = k^2 - K^2. \tag{1.7}$$

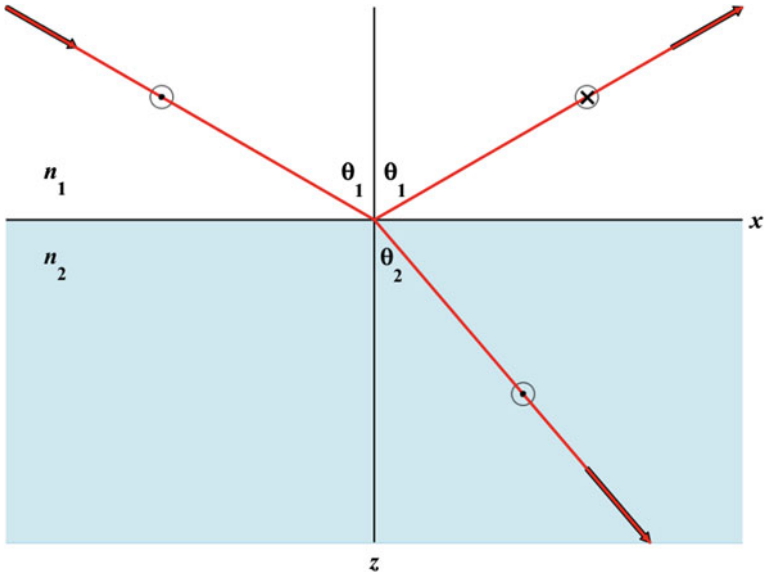


Fig. 1.1 Reflection of the electromagnetic s wave at a planar interface between media characterized by dielectric constants $\varepsilon_1 = n_1^2$ and $\varepsilon_2 = n_2^2$. The figure is drawn the air|water interface at optical frequencies, with $\varepsilon_1 \approx 1, \varepsilon_2 \approx (4/3)^2$

The meanings of k, K and q are evident from (1.5), (1.6) and (1.7): $k = \varepsilon^{1/2}\omega/c$ is the local magnitude of the wavevector, $K = k_x$ is the component of the wavevector along the interface, and $q = k_z$ is the component of the wavevector normal to the interface. For a plane wave incident from medium 1 as shown in Fig. 1.1, the existence of the separation-of-variables constant $K(= k_{1x} = k'_{1x} = k_{2x})$ implies

$$\varepsilon_1^{1/2} \sin\theta_1 \ = \varepsilon_1^{1/2} \sin\theta'_1 \ = \varepsilon_2^{1/2} \sin\theta_2, \tag{1.8}$$

where θ_1, θ'_1, and θ_2 are the angles of incidence, reflection, and transmission (or refraction).

Thus the fact that ε is a function of one spatial coordinate only, and the consequent separation of variables, implies the laws of reflection and refraction: the angle of reflection is equal to the angle of incidence, and the angles of incidence and refraction are related by Snell's Law. The refractive indices of the two media, defined as coefficients in Snell's Law $n_1 \sin\theta_1 = n_2 \sin\theta_2$, are $n_1 = \sqrt{\varepsilon_1}$ and $n_2 = \sqrt{\varepsilon_2}$. Note that the laws of reflection-refraction do not depend on the transition between the two media being sharp: they are valid for an arbitrary variation of $\varepsilon(z)$ between the asymptotic values ε_1 and ε_2.

As ε attains its limiting values $\varepsilon_1 = n_1^2$ and $\varepsilon_2 = n_2^2$, $q = (\varepsilon\omega^2/c^2 - K^2)^{1/2}$ takes the limiting values

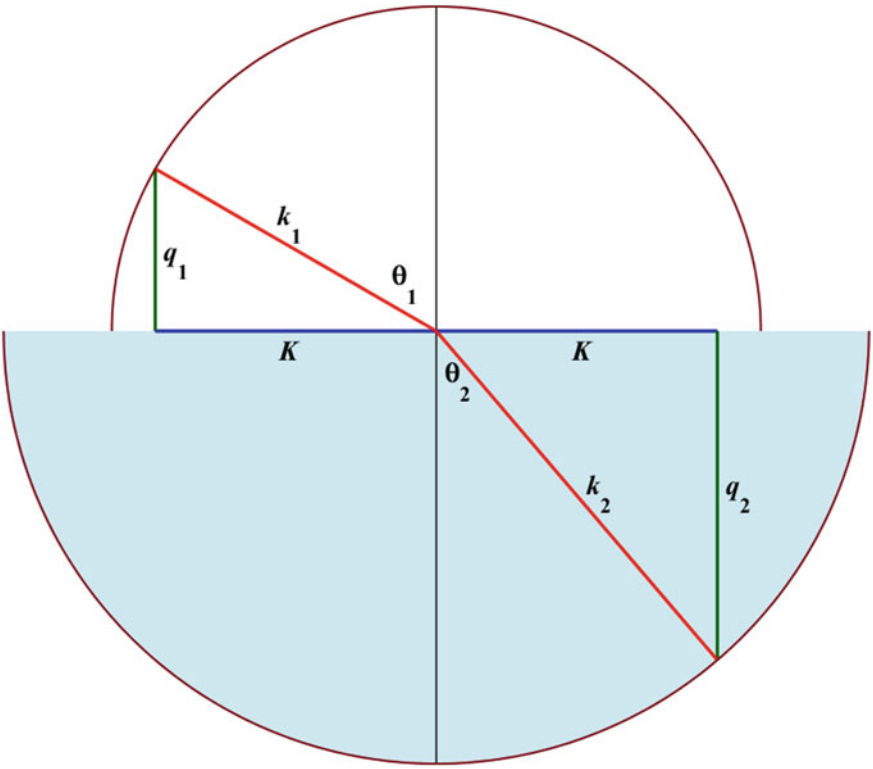


Fig. 1.2 Graphical representation of $k^2 = q^2 + K^2$ and $K = k_1 \sin\theta_1 = k_2 \sin\theta_2$. The figure is drawn for the air|water interface, as in Fig. 1.1. For incidence from the optically denser lower medium, as the angle of incidence θ_2 increases the magnitude of the tangential component K of the wavevector will increase beyond the magnitude k_1 of the wavevector in the upper medium. No transmitted wave is then possible, and there will be total internal reflection

$$q_1 = n_1 \frac{\omega}{c} \cos\theta_1, \qquad q_2 = n_2 \frac{\omega}{c} \cos\theta_2. \tag{1.9}$$

(For $\theta_1 > \theta_c = \arcsin(n_2/n_1)$ there is total reflection, q_2 is imaginary, and θ_2 is complex. This is discussed along with the particle case in Sect. 1.3.) Snell's Law and the relationships between the wavevector components are incorporated together in Fig. 1.2.

We now define the reflection and transmission amplitudes r_s and t_s in terms of the limiting forms of the solution of (1.7):

$$\mathrm{e}^{iq_1 z} + r_s\, \mathrm{e}^{-iq_1 z} \leftarrow E(z) \rightarrow t_s\, \mathrm{e}^{iq_2 z}. \tag{1.10}$$

The reflection amplitude is thus defined as the ratio of the coefficient of $\mathrm{e}^{-iq_1 z}$ to that of $\mathrm{e}^{iq_1 z}$, the transmission amplitude as the coefficient of $\mathrm{e}^{iq_2 z}$ when the incident wave

$e^{iq_1 z}$ has unit amplitude. Theory aims to obtain general properties of the reflection and transmission amplitudes, and to develop methods for calculating these for a given dielectric function profile. The calculation is simple for the important *step profile*

$$\varepsilon_0(z) = \begin{cases} \varepsilon_1 & (z<0) \\ \varepsilon_2 & (z>0) \end{cases} \tag{1.11}$$

For this profile we obtain r_s and t_s from the continuity of E and $\mathrm{d}E/\mathrm{d}z$ at $z=0$. (If, for example, $\mathrm{d}E/\mathrm{d}z$ were discontinuous, $\mathrm{d}^2E/\mathrm{d}z^2$ would have a delta function part, and (1.7) would not be satisfied.) For the step profile, E is given by the left and right sides of (1.10) for $z<0$ and $z>0$, respectively. The continuity of E and $\mathrm{d}E/\mathrm{d}z$ at the origin gives

$$1 + r_{s0} = t_{s0}, \qquad iq_1(1 - r_{s0}) = iq_2 t_{s0}. \tag{1.12}$$

Thus

$$r_{s0} = \frac{q_1 - q_2}{q_1 + q_2}, \qquad t_{s0} = \frac{2q_1}{q_1 + q_2}. \tag{1.13}$$

On using (1.8) and (1.9), the expressions (1.13) may be put into the Fresnel forms (Fresnel 1823)

$$r_{s0} = \frac{\sin(\theta_2 - \theta_1)}{\sin(\theta_2 + \theta_1)}, \qquad t_{s0} = \frac{2\sin\theta_2 \cos\theta_1}{\sin(\theta_2 + \theta_1)} \tag{1.14}$$

The phases of the reflected and transmitted waves are specified only when the phase of the incident wave *and* the location of the interface are specified. The above equations are for the discontinuity in $\varepsilon(z)$ located at $z=0$. In general, for the step located at z_1,

$$r_{s0} = e^{2iq_1 z_1} \frac{q_1 - q_2}{q_1 + q_2}, \qquad t_{s0} = e^{i(q_1 - q_2)z_1} \frac{2q_1}{q_1 + q_2}. \tag{1.15}$$

A special situation arises at grazing incidence ($\theta_1 \to \pi/2, q_1 \to 0$), when the incident and reflected waves are propagating in the same direction. Then the phase of the reflected wave is well-defined without specification of the interface location, and $r_{s0} \to -1$ (even in the case of the total internal reflection, when q_2 is imaginary). The fact that $r_s \to -1$ at grazing incidence is a general property of reflection from all interfaces, as will be shown in Sect. 2.3.

The classical electromagnetic fields **E** and **B** are real quantities, and the complex notation is used for mathematical convenience. (Complex fields are intrinsic in the quantum theory of particles, however.) The physical reflected s wave is, for unit amplitude of the incident wave,

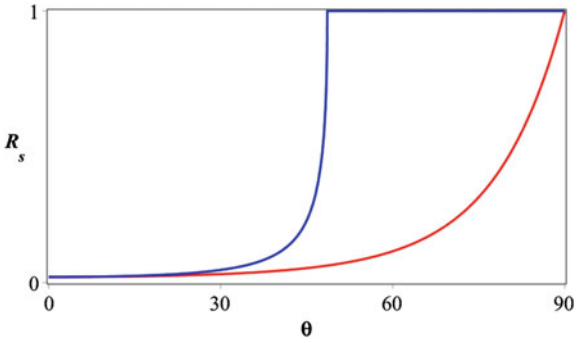


Fig. 1.3 Step profile reflectivity for the s wave. The parameters are for the air|water interface at optical frequencies, as in Figs. 1.1 and 1.2. The *lower curve* is for light incident from air; the *upper curve* for light incident from water shows total internal reflection for angle of incidence greater than $\theta_c = \arcsin\left(\frac{3}{4}\right) \approx 48.6^\circ$

$$\mathrm{Re}\{r_s\, e^{i(Kx-q_1 z-\omega t)}\} = \mathrm{Re}(r_s)\cos(Kx - q_1 z - \omega t) - \mathrm{Im}(r_s)\sin(Kx - q_1 z - \omega t)$$

The reflected intensity is proportional to the time average of the square of this, namely

$$\frac{1}{2}[\mathrm{Re}(r_s)]^2 + \frac{1}{2}[\mathrm{Im}(r_s)]^2 = \frac{1}{2}|r_s|^2$$

The incident intensity is proportional to the time average of $\cos^2(Kx + q_1 z - \omega t)$, which is $1/2$. Thus, $R_s = |r_s|^2$ is the ratio of the reflected intensity to the incident intensity. This quantity is called the reflectivity, or reflectance. Figure 1.3 shows R_s for a sharp transition between air and water, with light incident from air, and from water.

1.2 The Electromagnetic *p* Wave

We again take the incident and reflected waves propagating in the zx plane, and the stratifications lying in xy planes. For the p wave, $\mathbf{B} = (0, B_y, 0)$; the Maxwell equation (1.1) gives

$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = i\frac{\omega}{c}B_y, \tag{1.16}$$

while (1.2) implies $E_y = 0$ and

$$\frac{\partial B_y}{\partial z} = i\varepsilon\frac{\omega}{c}E_x, \qquad \frac{\partial B_y}{\partial x} = -i\varepsilon\frac{\omega}{c}E_z. \tag{1.17}$$

Elimination of E_x and E_z gives

$$\frac{\partial}{\partial x}\left(\frac{1}{\varepsilon}\frac{\partial B_y}{\partial x}\right) + \frac{\partial}{\partial z}\left(\frac{1}{\varepsilon}\frac{\partial B_y}{\partial z}\right) + \frac{\omega^2}{c^2}B_y = 0. \tag{1.18}$$

When ε is a function of one spatial coordinate z, the laws of reflection and refraction again follow from the separability of (1.18). We set

$$B_y(x, z, t) = \mathrm{e}^{i(Kx-\omega t)}B(z) \tag{1.19}$$

where K has the same meaning as for the s wave; then $B(z)$ satisfies the ordinary differential equation

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{\varepsilon}\frac{\mathrm{d}B}{\mathrm{d}z}\right) + \left(\frac{\omega^2}{c^2} - \frac{K^2}{\varepsilon}\right)B = 0 \tag{1.20}$$

When ε is constant (outside the interfacial region), the p wave equation has the same form as the s wave equation, with the same wavevector component q perpendicular to the interface. But within the interface there is an additional term proportional to the product of $\mathrm{d}\varepsilon/\mathrm{d}z$ and $\mathrm{d}B/\mathrm{d}z$. This term may be removed (and (1.20) converted to the form of the s wave (1.7)) in two ways. The first involves defining a new dependent variable

$$b = \left(\frac{\varepsilon_1}{\varepsilon}\right)^{12}B \tag{1.21}$$

(The factor $\varepsilon_1^{1/2}$ makes identical the limiting forms of b and B in medium 1.) The equation satisfied by b is

$$\frac{\mathrm{d}^2 b}{\mathrm{d}z^2} + q_b^2 b = 0, \qquad q_b^2 = q^2 - \varepsilon^{1/2}\frac{\mathrm{d}^2\varepsilon^{-1/2}}{\mathrm{d}z^2} = q^2 + \frac{1}{2\varepsilon}\frac{\mathrm{d}^2\varepsilon}{\mathrm{d}z^2} - \frac{3}{4}\left(\frac{1}{\varepsilon}\frac{\mathrm{d}\varepsilon}{\mathrm{d}z}\right)^2 \tag{1.22}$$

This form of the p polarization equation is useful for special profiles, in particular the exponential profile, which has $\ln\varepsilon$ linear in z, and the Rayleigh profile, which has $\varepsilon^{-1/2}$ linear in z. These are discussed in Chap. 2. It is also useful at short wavelengths, in the derivation of a perturbation theory for the p wave (Chap. 6).

The second transformation which removes the $(\mathrm{d}\varepsilon/\mathrm{d}z)(\mathrm{d}B/\mathrm{d}z)$ term is a dilation of the z variable in proportion to the local value of $\varepsilon(z)$: we define a new independent variable Z by

$$\mathrm{d}Z = \varepsilon\,\mathrm{d}z \tag{1.23}$$

Then, as may be seen on division of (1.20) by ε, the p wave equation reads

$$\frac{\mathrm{d}^2 B}{\mathrm{d}Z^2} + Q^2 B = 0, \qquad Q^2 = \frac{1}{\varepsilon}\frac{\omega^2}{c^2} - \frac{K^2}{\varepsilon^2} = \left(\frac{q}{\varepsilon}\right)^2. \tag{1.24}$$

This equation, in terms of the dilated z variable, and a reduced normal component of the wavevector, $Q = q/\varepsilon$, will be useful in many applications throughout this book.

The p wave reflection and transmission amplitudes are defined in terms of the limiting forms of $B(z)$:

$$\mathrm{e}^{iq_1 z} - r_p\,\mathrm{e}^{-iq_1 z} \leftarrow B(z) \rightarrow \frac{n_2}{n_1} t_p\,\mathrm{e}^{iq_2 z} \tag{1.25}$$

The reason for the factors -1 and $n_2/n_1 = (\varepsilon_2/\varepsilon_1)^{1/2}$ multiplying r_p and t_p is that we wish r_s and r_p and t_p and t_s to refer to the same quantity, here chosen to be the electric field. (This is not the only convention in use: some authors have the opposite sign on r_p.) The electric field components for the p wave are found from (1.2), (1.19) and (1.25) to have the limiting forms

$$\varepsilon_1^{-1/2}\cos\theta_1\ \mathrm{e}^{i(Kx-\omega t)}\left(\mathrm{e}^{iq_1 z} + r_p\,\mathrm{e}^{-iq_1 z}\right) \leftarrow E_x \rightarrow \varepsilon_1^{-1/2}\cos\theta_2\,t_p\,\mathrm{e}^{i(Kx+q_2 z-\omega t)}, \tag{1.26}$$

$$-\varepsilon_1^{-1/2}\sin\theta_1\ \mathrm{e}^{i(Kx-\omega t)}\left(\mathrm{e}^{iq_1 z} - r_p\,\mathrm{e}^{-iq_1 z}\right) \leftarrow E_z \rightarrow -\varepsilon_1^{-1/2}\sin\theta_2\,t_p\,\mathrm{e}^{i(Kx+q_2 z-\omega t)}. \tag{1.27}$$

The x-component of the electric field (tangential to the interface) thus has the reflection amplitude r_p, while the z-component (normal to the interface) has reflection amplitude $-r_p$.

At normal incidence there is no physical difference between the s and p polarizations: both have electric and magnetic fields tangential to the interface. For our geometry, E_z is zero at normal incidence, and (1.1) implies $\partial E_x/\partial z = i(\omega/c)B_y$. Thus B, the solution of (1.20) and (1.25), must be proportional to $\mathrm{d}E/\mathrm{d}z$, where E is the solution of (1.7) and (1.10). On substituting $\mathrm{d}E/\mathrm{d}z$ for B in (1.20) (with K set equal to zero) the left side becomes

$$\frac{\mathrm{d}}{\mathrm{d}z}\left\{\frac{1}{\varepsilon}\left(\frac{\mathrm{d}^2 E}{\mathrm{d}z^2} + \varepsilon\frac{\omega^2}{c^2}E\right)\right\}$$

and this is zero, by (1.7). Thus (1.20) is satisfied by $\mathrm{d}E/\mathrm{d}z$ at normal incidence. The proportionality of B and $\mathrm{d}E/\mathrm{d}z$ at normal incidence, when applied to the limiting forms (1.10) and (1.25), gives the equality of r_p with r_s and of t_p with t_s.

(Proportionality of B and $\mathrm{d}E/\mathrm{d}z$ could be replaced by equality of B and $(c/i\omega)\mathrm{d}E/\mathrm{d}z$, but then (1.25) would have to be modified by the factor n_1.)

At a discontinuity in the dielectric function, B and $\varepsilon^{-1}\mathrm{d}B/\mathrm{d}z = \mathrm{d}B/\mathrm{d}Z$ are continuous (from (1.20) or (1.24)). For the step profile $\varepsilon_0(z)$ defined by (1.11), B is equal to

$$B_0(z) = \begin{cases} \mathrm{e}^{iq_1 z} - r_{p0}\,\mathrm{e}^{-iq_1 z} & (z < 0) \\ \frac{n_2}{n_1} t_{p0}\,\mathrm{e}^{iq_2 z} & (z > 0) \end{cases} \tag{1.28}$$

The continuity of B and $\varepsilon^{-1}\mathrm{d}B/\mathrm{d}z$ at the origin gives

$$1 - r_{p0} = \frac{n_2}{n_1} t_{p0}, \tag{1.29}$$

$$iQ_1\left(1 + r_{p0}\right) = iQ_2 \frac{n_2}{n_1} t_{p0}, \tag{1.30}$$

where $Q_1 = q_1/\varepsilon_1$ and $Q_2 = q_2/\varepsilon_2$. Thus (compare (1.13))

$$-r_{p0} = \frac{Q_1 - Q_2}{Q_1 + Q_2}, \qquad \frac{n_2}{n_1} t_{p0} = \frac{2Q_1}{Q_1 + Q_2}. \tag{1.31}$$

On using (1.8) and (1.9) we obtain the Fresnel forms

$$r_{p0} = \frac{\tan(\theta_2 - \theta_1)}{\tan(\theta_2 + \theta_1)}, \qquad t_{p0} = \frac{2 \sin\theta_2 \cos\theta_1}{\sin(\theta_2 + \theta_1)\cos(\theta_2 - \theta_1)} \tag{1.32}$$

The reflectivity of the p polarization off a discontinuity in the dielectric function is shown in Fig. 1.4.

From (1.31) we see that the p wave shows zero reflection when $Q_1 = Q_2$, that is at the Brewster angle

$$\theta_B = \arctan\frac{n_2}{n_1}. \tag{1.33}$$

It is apparent from (1.24) that this angle has special significance not only for a sharp transition between two media, but for diffuse profiles as well. This is because the wave equation in the dilated variable Z links two media with effective wavevector components Q_1, and Q_2 which are equal at this angle. The s and p effective wavevector components q and Q are shown in Fig. 1.5, which also illustrates the reason for small p reflectivity at the Brewster angle. The Figure shows q^2 versus z and Q^2 versus Z for the hyperbolic tangent profile

$$\varepsilon(z) = \frac{1}{2}(\varepsilon_1 + \varepsilon_2) - \frac{1}{2}(\varepsilon_1 - \varepsilon_2)\tanh\frac{z}{2a}, \tag{1.34}$$

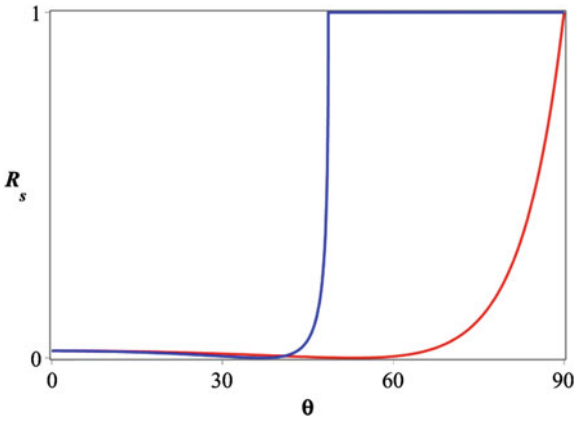


Fig. 1.4 Step profile reflectivity for the p wave, for the air-water interface. The curve for light incident from air is zero at the Brewster angle $\arctan(4/3) \approx 53.1^\circ$. For incidence from water the reflectivity is zero at the Brewster angle $\arctan(3/4) \approx 36.9^\circ$, and unity beyond the critical angle $\theta_c = \arcsin(3/4) \approx 48.6^\circ$

for which the dilated z coordinate is

$$Z = \frac{1}{2}(\varepsilon_1 + \varepsilon_2)z - (\varepsilon_1 - \varepsilon_2)a \ln\cosh\left(\frac{z}{2a}\right). \tag{1.35}$$

At the Brewster angle θ_B,

$$Q_1^2 = Q_2^2 = \frac{(\omega/c)^2}{\varepsilon_1 + \varepsilon_2} = Q_B^2, \tag{1.36}$$

$$K^2 = \varepsilon_1\varepsilon_2 Q_B^2 = K_B^2 = \frac{\varepsilon_1\varepsilon_2}{\varepsilon_1 + \varepsilon_2}\left(\frac{\omega}{c}\right)^2. \tag{1.37}$$

From (1.24), a general profile $\varepsilon(z)$ has Q^2 at the Brewster angle given by

$$Q^2(\theta_B, z) = \frac{\frac{\omega^2}{c^2}\left\{\varepsilon(z) - \frac{\varepsilon_1\varepsilon_2}{\varepsilon_1 + \varepsilon_2}\right\}}{\varepsilon^2(z)}. \tag{1.38}$$

Thus the bump in Q^2 at the Brewster angle (see Fig. 1.5) has the analytic form

$$Q^2(\theta_B, z) - Q_B^2 = \frac{\omega^2}{c^2}\frac{(\varepsilon_1 - \varepsilon)(\varepsilon - \varepsilon_2)}{\varepsilon^2(\varepsilon_1 + \varepsilon_2)}. \tag{1.39}$$

The p wave equation in the Z, Q notation has reflection at θ_B due to the small variation in the effective wavevector component Q as given by (1.39). For the step profile, ε equals either ε_1 or ε_2, and there is no variation in Q and thus no reflection.

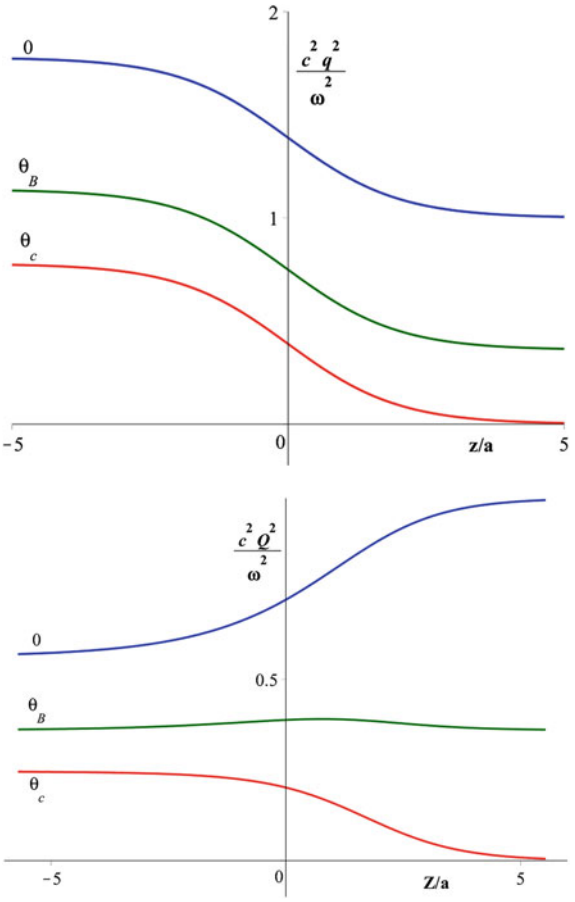


Fig. 1.5 Squares of the normal wavevector component q and of the effective normal component Q for the s and p waves. The figure shows $q^2(z)$ and $Q^2(Z)$ for the hyperbolic tangent dielectric function profile, at three angles of incidence. The *upper curve* (in each case) is for normal incidence, the *middle curve* is at the Brewster angle $\theta_B = \arctan(n_2/n_1)$ and the *lower curve* is at the critical angle for total internal reflection, $\theta_c = \arcsin(n_2/n_1)$. The refractive indices $n_1 = 4/3$ and $n_2 = 1$ approximate the water|air interface. Water is on the *left* in both diagrams

A common explanation for the small reflection of the p polarization at θ_B is in terms of the angular dependence of the dipole radiation from each atom or molecule which produces the transmitted and reflected waves. The far-field radiation pattern of a dipole has zero amplitude along the line of oscillation of the dipole (see Sect. 1.5, (1.78)). We see from (1.32) that r_{p0} is zero when $\theta_1 + \theta_2 = \pi/2$, that is when the refracted and reflected waves are at a right angle (see Fig. 1.6). The argument goes that at this angle of incidence there is no radiation from the accelerated electrons in the material to produce a p-polarized signal in the direction of specular reflection (upper part of Fig. 1.6). But zero reflection also exists in the reverse case of material to vacuum (lower figure). In this case the explanation in terms of electrons radiating along the transmitted beam to produce (or fail to produce) the reflected beam does not apply. Further, a similar case of zero reflection at the interface between two unlike media occurs with acoustic waves (as will be discussed in Sect. 1.4), and in that case the radiation from each scatterer does not have a dipole character.

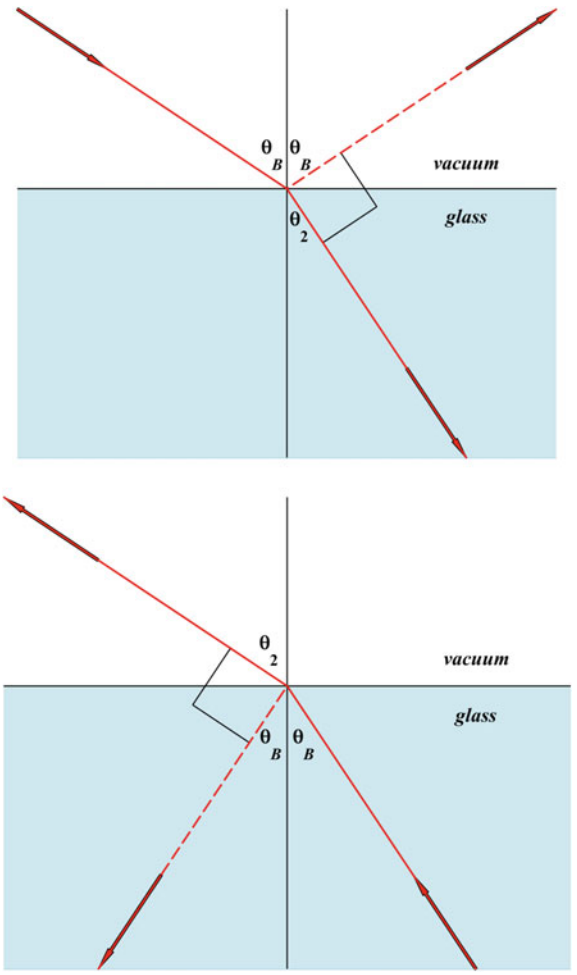


Fig. 1.6 Illustrating complete transmission of the p wave at the Brewster angle. In each case $\theta_2 = \pi/2 - \theta_B$, so the transmitted and non-reflected rays are at right angles

1.3 Particle Waves

In non-relativistic quantum mechanics, the motion of a particle of mass m and energy $\mathcal{E}$ in a potential V is determined by Schrödinger's equation for the probability amplitude Ψ,

$$-\frac{\hbar^2}{2m}\nabla^2\Psi + V\Psi = \mathcal{E}\Psi. \tag{1.40}$$

($\hbar$ is Planck's constant divided by 2π.) We shall consider reflection at a planar stratified boundary region between two uniform media characterized by potentials V_1 and V_2. Examples of the particles and interfaces to which this description applies

are: electrons at a junction between two metals (with possibly an oxide layer in between); neutrons reflecting off a solid or liquid surface; and helium atoms reflecting at a liquid helium surface. In each of these examples the potential V in the single-particle equation (1.40) is an effective potential, representing the net effect of all the interactions between the particle and the scatterers in the medium through which it moves. An example of how this effective potential is determined is given in Sect. 1.5.

We again consider plane waves propagating in the zx plane, incident on a planar interface, with stratification in the z direction. For this geometry, V depends on one spatial variable z, and Ψ is independent of y. The z, x variable dependence in (1.40) is then separable, with

$$\Psi(z,x) = \mathrm{e}^{iKx}\psi(z) \tag{1.41}$$

(it is usual to suppress the time dependence $\mathrm{e}^{-iEt/\hbar}$). Substitution of (1.41) into (1.40) gives an ordinary differential equation for ψ:

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}z^2} + q^2\psi = 0, \qquad q^2(z) = \frac{2m}{\hbar^2}[\mathcal{E} - V(z)] - K^2. \tag{1.42}$$

From (1.41), K is the x-component of the wavevector in either medium, and is an invariant of the motion, because of the absence of transverse components of the force, $\partial V/\partial x = 0 = \partial V/\partial y$. If the angles of incidence, reflection and refraction are θ_1, θ_1', and θ_2, the laws of reflection and refraction follow from the invariance of $K = k_{1x} = k_{1x}' = k_{2x}$:

$$k_1 \sin\theta_1 \ = k_1 \sin\theta_1' \ = k_2 \sin\theta_2, \tag{1.43}$$

where

$$k_i^2 = K^2 + q_i^2 = \frac{2m}{\hbar^2}[\mathcal{E} - V_i]. \tag{1.44}$$

As before, q is the component of the wavevector normal to the interface, with limiting values

$$k_1 \cos\theta_1 \ = q_1 \leftarrow q(z) \rightarrow q_2 = k_2 \cos\theta_2. \tag{1.45}$$

These relations are summarized in Fig. 1.7.

On comparison of (1.7) and (1.42) we see that there is a one-to-one correspondence between the reflection problems for the electromagnetic s wave and particle waves obeying Schrödinger's equation, with the replacement

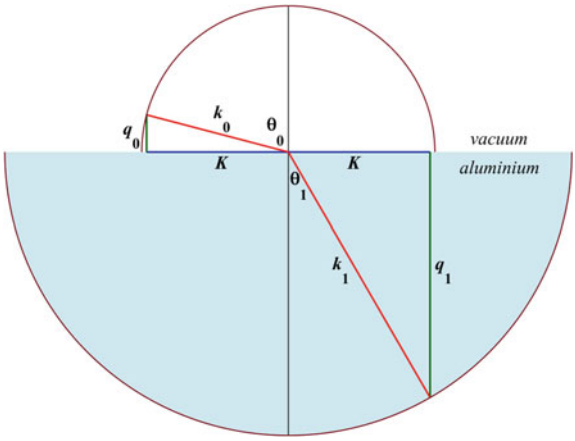


Fig. 1.7 Graphical representation of $k^2 = q^2 + K^2$ and of $K = k_1 \sin\theta_1 = k_0 \sin\theta_0$. (We use zero as subscript since in this example the upper medium is the vacuum; V_0 is the vacuum potential, usually taken as zero.) The figure is drawn for electrons at 10 eV above the Fermi level in bulk aluminium, at the aluminium-vacuum interface. $\mathcal{E}_F - V_1 \approx 11.7\,\text{eV}$, so $\mathcal{E} - V_1 \approx 21.7\,\text{eV}$; $V_0 - \mathcal{E}_F \approx 4.2\,\text{eV}$, so $\mathcal{E} - V_0 \approx 5.8\,\text{eV}$; the ratio of the refractive indices is $\{(\mathcal{E} - V_1)/(\mathcal{E} - V_0)\}^{\frac{1}{2}} \approx 1.934$

$$\varepsilon(z)\frac{\omega^2}{c^2} \leftrightarrow \frac{2m}{\hbar^2}[\mathcal{E} - V(z)]. \tag{1.46}$$

The reflection amplitude r and the transmission amplitude t are defined in terms of the limiting forms of the solution of (1.42):

$$\mathrm{e}^{iq_1 z} + r\,\mathrm{e}^{-iq_1 z} \leftarrow \psi(z) \rightarrow t\,\mathrm{e}^{iq_2 z}. \tag{1.47}$$

For example, for the potential step

$$V_0(z) = \begin{cases} V_1 & (z<0) \\ V_2 & (z>0) \end{cases} \tag{1.48}$$

continuity of ψ and $\mathrm{d}\psi/\mathrm{d}z$ at $z = 0$ gives the Fresnel-type equations

$$r_0 = \frac{q_1 - q_2}{q_1 + q_2}, \qquad t_0 = \frac{2q_1}{q_1 + q_2}. \tag{1.49}$$

Note that, as in the case of electromagnetic waves, the boundary conditions follow from the differential equations; they are not an additional assumption of the theory.

A refractive index can be defined for particles. From (1.43) and (1.44) we see that the refractive index is proportional to $(\mathcal{E} - V)^{1/2}$, that is to the square root of the kinetic energy, or to the local value of the wavevector k. The proportionality to

$(\mathcal{E} - V)^{1/2}$ is also a classical result: the equations for the conservation of energy and transverse momentum for a particle incident at angle θ_1 onto a planar stratification between media 1 and 2 read

$$\frac{1}{2}mv_1^2 + V_1 = \mathcal{E} = \frac{1}{2}mv_2^2 + V_2, \tag{1.50}$$

$$mv_1 \sin\theta_1 \ = mv_2 \sin\theta_2 \,. \tag{1.51}$$

Equation (1.51) shows that the refractive indices are proportional to v_i, which from (1.50) are equal to $[2(\mathcal{E} - V_i)/m]^{1/2}$. However, partial reflection does not exist for classical particles: there is either total reflection (when $V > \mathcal{E} - \frac{1}{2}m(v_1 \sin\theta_1)^2$ anywhere), or no reflection (when $V < \mathcal{E} - \frac{1}{2}m(v_1 \sin\theta_1)^2$ everywhere).

In contrast, total reflection occurs in the wave theory only if $V_2 > \mathcal{E} - \hbar^2K^2/2m$; q_2 is then imaginary, leading to exponential decay of the probability amplitude in medium 2. Regions of imaginary q (negative $q^2 = (2m/\hbar^2)(\mathcal{E} - V) - K^2$) where $V > \mathcal{E} - \hbar^2K^2/2m$), do not lead to *total* reflection when q_2 is real, because of tunneling. Electromagnetic waves are likewise totally reflected when q_2 is imaginary, that is when $\varepsilon_2\omega^2/c^2 < K^2$, or $\sin^2\theta_1 > \varepsilon_2/\varepsilon_1$. Thus the critical angle for total reflection is given by

$$\theta_c = \arcsin\left(\frac{\varepsilon_2}{\varepsilon_1}\right)^{1/2}, \qquad \theta_c = \arcsin\left(\frac{\mathcal{E} - V_2}{\mathcal{E} - V_1}\right)^{1/2} \tag{1.52}$$

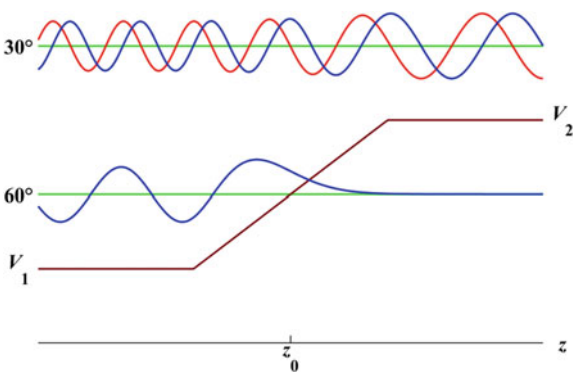


Fig. 1.8 Probability amplitudes, at two angles of incidence, for particle waves incident from the *left* onto a linear ramp potential. The energy and potential values are such that $\theta_c = 45°$ (V_1:V_2:$\mathcal{E}$ = 1:3:5). The *upper two waves* are the real and imaginary parts of the probability amplitude ψ for incidence at 30°. The *lower curve* is the imaginary part of the probability amplitude for a totally reflected wave, incident at 60°. The real part is not shown, since the real and imaginary parts of ψ are proportional to each other in total reflection: $\mathrm{Im}\,\psi/\mathrm{Re}\,\psi = \tan\delta/2$ when $r = \mathrm{e}^{i\delta}$ (Sect. 2.2). The classical turning point z_0 (where $q^2 = 0$) is halfway up the ramp

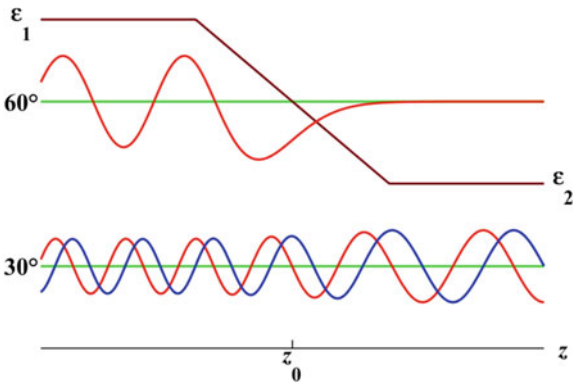


Fig. 1.9 The electromagnetic s wave at two angles of incidence onto a linear dielectric function profile. The radiation is incident from the *left*. The dielectric constants are $\varepsilon_1 = 2, \varepsilon_2 = 1$, so that $\theta_c = 45°$. The *lower two waves* are the real and imaginary parts of the electric field $E(z)$, at 30° angle of incidence. The *upper curve* is the real part of $E(z)$ for a totally reflected wave incident at 60°. The curves are drawn at the level of $(cK/\omega)^2$ for each angle of incidence. The wavefunctions for the electromagnetic s wave and for the particle waves of Fig. 1.8 are the same

in the electromagnetic and particle wave cases. Partial and total reflection of particle and electromagnetic s waves is compared in Figs. 1.8 and 1.9. Note that at 30° incidence the net flux at the right of the barrier, $q_2|t|^2$, is the same as the net flux on the left, $q_1(1 - |r|^2)$, despite the visible increase in the real and imaginary parts of the probability amplitude to the right. At 60° incidence the wave is totally reflected. The probability amplitudes are drawn about the levels $\mathcal{E} - \hbar^2 K^2/2m$, the energy available for motion in the z direction.

1.4 Acoustic Waves

There is an interesting close correspondence between the reflection of sound and the reflection of the electromagnetic p wave. This will be demonstrated in the simplest case of fluid, non-viscous media. Dissipation via viscosity and scattering can be accommodated by the use of a complex sound speed.

Sound waves propagate changes in density and pressure which are usually very small compared to the mean values. The equations of motion, continuity, and state can then be linearized by setting

$$\text{density} = \varrho + \varrho_a, \qquad \text{pressure} = p + p_a, \tag{1.53}$$

where ϱ and p are the mean local values of the density and pressure, and ϱ_a and p_a are the small excess time-dependent values due to the presence of acoustic waves. On dropping second order terms in ϱ_a, p_a and in the velocity of a fluid element, and

neglecting the force due to gravity apart from its effect on stratification according to density, one obtains the equation (Bergmann 1946)

$$\nabla^2 p_a - \frac{1}{v^2}\frac{\partial^2 p_a}{\partial t^2} - \frac{1}{\varrho}\nabla\varrho \cdot \nabla p_a = 0. \tag{1.54}$$

Here $v^2 = (\partial p/\partial\varrho)_s$ is the adiabatic derivative of the pressure with respect to density, and gives the square of the local phase velocity in the medium.

Consider now the reflection of sound at an interface characterized by a density profile $\varrho(z)$ and an adiabatic pressure derivative $(\partial p/\partial\varrho)_s = v^2(z)$. For a plane monochromatic wave propagating in the zx plane, we have

$$p_a(z, x, t) = P(z)\, \mathrm{e}^{i(Kx-\omega t)}. \tag{1.55}$$

K is again the component of the wavevector along the interface, and is a constant of the motion:

$$K = \frac{\omega}{v_1}\sin\theta_1 \; = \frac{\omega}{v_2}\sin\theta_2\,, \tag{1.56}$$

where v_1, v_2 are the limiting values of $\left\{(\partial p/\partial\varrho)_s\right\}^{1/2}$ in the two media, and θ_1, θ_2 are the angles of incidence and refraction. The differential equation for P is obtained by substitution of (1.55) into (1.54):

$$\varrho\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{\varrho}\frac{\mathrm{d}P}{\mathrm{d}z}\right) + q^2 P = 0, \tag{1.57}$$

with

$$q^2(z) = \frac{\omega^2}{v^2(z)} - K^2; \tag{1.58}$$

q is again the normal component of the wavevector, with limiting values $q_1 = (\omega/v_1)\cos\theta_1, q_2 = (\omega/v_2)\cos\theta_2$.

The term $(\mathrm{d}\varrho/\mathrm{d}z)(\mathrm{d}P/\mathrm{d}z)$ in (1.57) may be removed by introducing a new dependent variable $P/\sqrt{\varrho}$, as Bergmann notes. This is analogous to the transformation to $B/\sqrt{\varepsilon}$ discussed in Sect. 1.2. A more fruitful approach is analogous to the transformation to a dilated z variable in the p wave case: (1.57) has the same form as the electromagnetic p wave equation

$$\varepsilon\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{\varepsilon}\frac{\mathrm{d}B}{\mathrm{d}z}\right) + q^2 B = 0$$

In terms of a new independent variable Z, defined by $\mathrm{d}Z = \varrho \mathrm{d}z$, (1.57) becomes

$$\frac{\mathrm{d}^2 P}{\mathrm{d}Z^2} + Q^2 P = 0, \qquad Q = \frac{q}{\varrho}. \tag{1.59}$$

(As defined here, Z and Q no longer have the dimensions of length and of $(\text{length})^{-1}$; this can be remedied by respectively dividing Z and multiplying Q by some density, for example $(\varrho_1 + \varrho_2)/2$.)

It is clear from the form of (1.59), and the discussion of reflection at the Brewster angle given in Sect. 1.2, that weak reflection of acoustic waves (zero reflection, in the case of a sharp transition between the two media) is expected whenever $Q_1 = Q_2$. This holds when

$$\frac{\cos\theta_1}{\varrho_1 v_1} = \frac{\cos\theta_2}{\varrho_2 v_2}. \tag{1.60}$$

This result was first given (for a sharp interface) by George Green (1838). On eliminating θ_2 from (1.60) and Snell's Law (1.56), one finds that weak reflection occurs at an angle of incidence $\theta_1 = \theta_G$ (which we will call Green's angle) given by

$$\tan^2\theta_G = \frac{(\varrho_2 v_2)^2 - (\varrho_1 v_1)^2}{\varrho_1^2 (v_1^2 - v_2^2)}. \tag{1.61}$$

In contrast to the electromagnetic p wave case, weak reflection of acoustic waves does not happen at a certain angle at a boundary between *any* two media: the quantities $\varrho_1 v_1 - \varrho_2 v_2$ and $v_1 - v_2$ must have opposite signs.

At Green's angle θ_G (where $Q_1 = Q_2$), K^2 is equal to

$$K_G^2 = \frac{\omega^2}{\varrho_1^2 - \varrho_2^2}\left\{\left(\frac{\varrho_1}{v_2}\right)^2 - \left(\frac{\varrho_2}{v_1}\right)^2\right\}, \tag{1.62}$$

and the common value of Q_1 and Q_2 is given by

$$Q_G^2 = \frac{\omega^2}{\varrho_1^2 - \varrho_2^2}\left\{\frac{1}{v_1^2} - \frac{1}{v_2^2}\right\}. \tag{1.63}$$

According to (1.59), the acoustic wave in the Z variable then reflects from the bump in Q^2, given by

$$Q^2 - Q_G^2 = \frac{\omega^2}{\varrho^2\left(\varrho_1^2 - \varrho_2^2\right)}\left\{\frac{\varrho_1^2 - \varrho_2^2}{v^2} - \left(\frac{\varrho_1}{v_2}\right)^2 + \left(\frac{\varrho_2}{v_1}\right)^2 - \varrho^2\left(\frac{1}{v_1^2} - \frac{1}{v_2^2}\right)\right\}. \tag{1.64}$$

One can define an acoustic reflection amplitude r and a transmission amplitude t in terms of the pressure:

$$e^{iq_1 z} + r\, e^{iq_1 z} \leftarrow P(z) \rightarrow t\, e^{iq_2 z}. \tag{1.65}$$

For a sharp transition between media 1 and 2, P and $dP/\varrho dz = dP/dZ$ are continuous at the boundary. (This is evident from (1.57); note that, as in the electromagnetic and particle wave cases, the differential equation dictates the boundary conditions, which are not an additional input to the theory.) Thus, for a sharp boundary located at the origin,

$$r = \frac{Q_1 - Q_2}{Q_1 + Q_2}, \qquad t = \frac{2Q_1}{Q_1 + Q_2}. \tag{1.66}$$

These may be rewritten as

$$r = \frac{\varrho_2 \tan\theta_2 - \varrho_1 \tan\theta_1}{\varrho_2 \tan\theta_2 + \varrho_1 \tan\theta_1}, \qquad t = \frac{2\varrho_2 \tan\theta_2}{\varrho_2 \tan\theta_2 + \varrho_1 \tan\theta_1}. \tag{1.67}$$

Total reflection occurs for angles of incidence greater than

$$\theta_c = \arcsin\left(\frac{v_1}{v_2}\right). \tag{1.68}$$

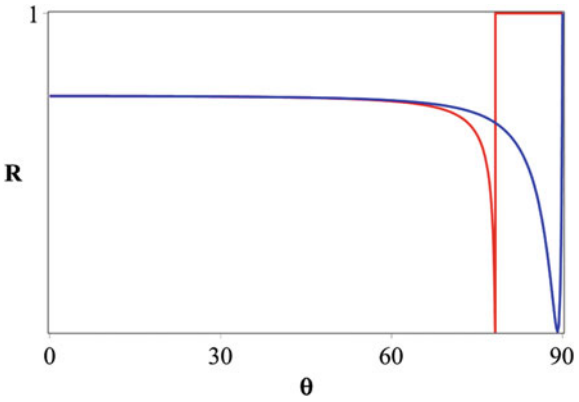


Fig. 1.10 Reflectivity of acoustic waves at a mercury-water interface, according to (1.66). For sound incident from the slower medium (mercury) total reflection occurs beyond the critical angle $\theta_c \approx 78.21°$. Very close is the Green's angle $\theta_G \approx 78.18°$, so the reflectivity changes from zero to unity in 0.03°. For sound incident from water the Green's angle is very close to glancing incidence, $\theta_G \approx 89.12°$. Thus again the reflectivity changes from zero to unity very rapidly. The curves are drawn for $\rho_{Hg}/\rho_{H_2O} = 13.57, v_{Hg}/v_{H_2O} = 0.9789$

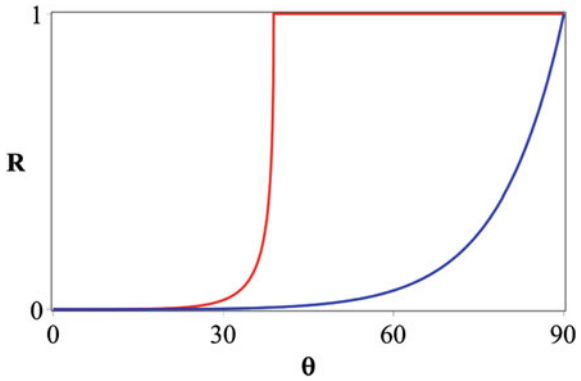


Fig. 1.11 Reflectivity of acoustic waves at a water-carbon tetrachloride interface, obtained from (1.66). For sound incident from the slower medium (CCl_4) the Green's angle is $\theta_G \approx 1.73°$, and total reflection occurs beyond $\theta_c \approx 38.86°$. For sound incident from water the Green's angle is $\theta_G \approx 2.76°$. The curves are drawn for $\rho_{CCl_4}/\rho_{H_2O} = 1.595, v_{CCl_4}/v_{H_2O} = 0.6274$

This result follows on setting $\theta_2 = \pi/2$ in (1.56); it holds for any interface, no matter how diffuse, provided absorption can be neglected. The critical angle θ_c will be close to Green's angle θ_G, if the latter exists, when $v_1 \simeq v_2$ The reflectivity of a step profile then rapidly goes from zero at θ_G to unity at θ_c and beyond, as illustrated for the mercury-water interface in Fig. 1.10.

When $\varrho_1 v_1 \simeq \varrho_2 v_2$ the reflectivity at normal incidence is small, and θ_G (if it exists) will also be small. This is the case for carbon tetrachloride and water, illustrated in Fig. 1.11.

1.5 Scattering and Reflection

Most of the results in this book come from analysis of the differential equations for waves in material media, the media being characterized by a dielectric function, or an effective potential, or the density and speed of sound, in the case of electromagnetic, particle or acoustic waves. This approach hides the many-body complexity of the real physics: specular reflection, for example, is the result of the constructive interference of many scattered or re-radiated waves. A discussion of reflection from this point of view will be given here; it leads to values for the functions characterizing the media, such as ε and V, in terms of the properties of the particles comprising the system. Such approaches go back to Lorentz (1909), Darwin (1924) and Hartree (1928) in the electromagnetic case. We will begin with an adaptation of Fermi's (1950) argument for the effective potential of a collection of neutron scatterers, since this is simpler.

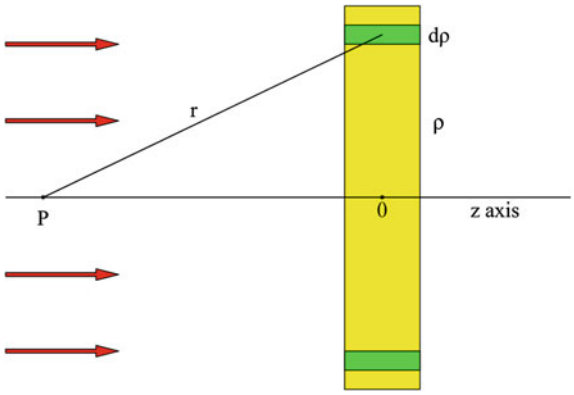


Fig. 1.12 Reflection of neutrons by a slab of scatterers. The thickness Δz of the slab is such that $k\Delta z$ is small, so that the phase of the plane wave e^{ikz}, (incident from the *left*) is nearly constant over the slab

Consider the reflection of a *beam of neutrons* by a thin slab of material. The neutrons interact with the nuclei in the slab. For slow neutrons this interaction is characterized by a length b, the scattering length for neutrons off a bound scatterer. An incident plane wave e^{ikz} causes each scatterer to radiate a spherical wave $-be^{ikr}/r$. The reflected wave is found by summing up the scattered waves from all parts of the slab. The geometry is illustrated in Fig. 1.12.

If n is the number density of the scatterers, $(2\pi\varrho d\varrho\Delta z)n$ is the number of scatterers within an annulus between ρ and $\varrho + d\varrho$, where $\rho = \sqrt{x^2 + y^2}$ is the distance from the z axis. The reflected wave at P is thus

$$\psi_r = \int_0^\infty d\varrho 2\pi\varrho\Delta z\, n \left(-\frac{b\, e^{ikr}}{r}\right). \tag{1.69}$$

For fixed z we have $\varrho d\varrho = rdr$, so that

$$\psi_r = -2\pi nb\Delta z \int_{-z}^\infty dr\, e^{ikr}. \tag{1.70}$$

The integral over r is not defined as it stands, because we have used e^{ikz} as the incident wave, namely a plane wave extending to infinity in the x and y directions. In practice the incident wave would be a finite beam, with an amplitude decreasing with $\varrho = (x^2 + y^2)^{1/2}$. The resulting integral for ψ_r then is well-defined. When such

decrease is slow on the scale of k^{-1} (the beam is many wavelengths wide), the integral is equal to $-\mathrm{e}^{-ikz}/ik$, and

$$\psi_r = \frac{2\pi nb\Delta z}{ik}\,\mathrm{e}^{-ikz} \tag{1.71}$$

This is a reflected wave, with reflection amplitude equal to the coefficient of e^{-ikz}.

We will now show that the reflection amplitude due to a thin slab of thickness Δz and effective potential V is (to lowest order in V)

$$r_1 = \left(\frac{\Delta z}{2ik}\right)\frac{2mV}{\hbar^2}. \tag{1.72}$$

(What follows is heuristic; a rigorous proof is given in Chap. 3; see in particular (3.14).) Consider the effect of a potential hump, or well, which is small in extent in comparison with the wavelength. Seen on the scale of the wavelength, the hump appears as a spike, and its main effect is to create a change of slope in the wavefunction: on integrating (1.42) (at normal incidence) across the hump, we have

$$\psi'(z_2) - \psi'(z_1) = -\int_{z_1}^{z_2} \mathrm{d}z\,\frac{2m}{\hbar^2}\,[\mathcal{E} - V(z)]\,\psi(z). \tag{1.73}$$

For wavelengths long compared to the extent of the hump, ψ is nearly constant over its effect, so when $z_2 - z_1$ is small compared to $k^{-1} = \hbar(2m\mathcal{E})^{-1/2}$, and the hump is centred on the origin,

$$\psi'(z_2) - \psi'(z_1) \simeq \frac{2m}{\hbar^2}\psi(0)\int_{z_1}^{z_2} \mathrm{d}z\,V(z) \tag{1.74}$$

(The zero of energy has been chosen so that V goes to zero on either side of the hump.) From (1.47) the left side is equal to $ik(t-1+r)+O(k^2)$. The assumption that ψ remains nearly constant from z_1 to z_2 also implies $1+r \simeq \psi(0) \simeq t$. Thus (1.74) gives

$$r \simeq \left(\frac{1}{2ik}\right)\frac{2m}{\hbar^2}\int_{z_1}^{z_2} \mathrm{d}z\,V(z). \tag{1.75}$$

For V constant inside the hump (of extent Δz), and zero outside, this reduces to (1.72).

We can now give an expression for the effective potential of a collection of scatterers: (1.71) and (1.72) together imply that this is

$$V = 4\pi\left(\frac{\hbar^2}{2m}\right)nb. \tag{1.76}$$

The scattered waves interfere constructively to give a reflected and a transmitted wave, as if the medium were completely homogeneous and acted on the particles with a potential given by (1.76). We have considered normal incidence; at oblique incidence the constructive interference of the spherically diverging waves from the scatterers within the slab is in the specular and straight-through directions.

We now turn to the more complex question of *electromagnetic radiation* interacting with the atoms or molecules in a thin slab of material. We will calculate the field at P in front of a slab, as in Fig. 1.12. The incident electric field propagates in the z direction, and is taken to be polarized along the x direction,

$$\mathbf{E} = \mathrm{e}^{i(kz-\omega t)}(E_0, 0, 0) \tag{1.77}$$

When the wavelength is large compared to atomic size, each atom radiates predominantly as a dipole. For a given atom with dipole $\mathbf{p}$, oscillating at the impressed angular frequency ω, the electric field at $\mathbf{r} = r\hat{\mathbf{r}}$ from the atom is (see for example Jackson (1975), Sect. 9.2)

$$\mathbf{E} = \frac{\mathrm{e}^{ikr}}{r}\left\{k^2(\hat{\mathbf{r}} \times \mathbf{p}) \times \hat{\mathbf{r}} + [3(\hat{\mathbf{r}} \cdot \mathbf{p})\hat{\mathbf{r}} - \mathbf{p}]\left(\frac{1}{r^2} - \frac{ik}{r}\right)\right\}, \tag{1.78}$$

where $k = \omega/c$. The far field (given by the first term) is a spherically diverging wave, with $\mathbf{E}$ transverse to $\mathbf{r}$. We do not omit the near field, since we do not wish to assume that $kr \gg 1$. All dipoles are taken to lie along the direction of the incident electric field, and to have the same strength αE_0, where α is the polarizability of an atom:

$$\mathbf{p} = \mathrm{e}^{-i\omega t}(\alpha E_0, 0, 0) \tag{1.79}$$

The point P is at $(0, 0, z)$, with $z < 0$. The contribution to the electric field at P from a dipole at $(x, y, 0)$ is then

$$E_x = \frac{\alpha E_0\, \mathrm{e}^{i(kr-\omega t)}}{r^3}\left\{k^2(y^2 + z^2) + (1 - ikr)\left(\frac{3x^2}{r^2} - 1\right)\right\} \tag{1.80}$$

with E_y and E_z odd in x and thus integrating to zero when we sum over the dipolar fields. Thus the net field at P due to all the dipoles (of number density n) in the thin slab is, on changing to the cylindrical coordinates ϱ and ϕ and integrating over ϕ,

$$E_x^{\text{dipoles}} = \mathrm{e}^{-i\omega t}\alpha E_0 n\pi\Delta \mathrm{z}\int_0^\infty \mathrm{d}\varrho\,\varrho\,\frac{e^{ikr}}{r^3}\left\{k^2(\varrho^2+2z^2)+(1-ikr)\left(\frac{3\varrho^2}{r^2}-2\right)\right\}$$
$$= \mathrm{e}^{-i\omega t}\alpha E_0 n\pi\Delta \mathrm{z}\int_{-z}^\infty \mathrm{d}r\,\frac{e^{ikr}}{r^2}\left\{k^2(r^2+z^2)+(1-ikr)\left(1-\frac{3z^2}{r^2}\right)\right\}. \tag{1.81}$$

The first term is an integral of the form (1.70); the others may be obtained from it by integration by parts. The result is

$$E_x^{\text{dipoles}} = \mathrm{e}^{-i(kz+\omega t)}2\pi ik\alpha n\Delta z E_0. \tag{1.82}$$

The reflection amplitude for the slab is the coefficient of $E_0\mathrm{e}^{-i(kz+\omega t)}$ in (1.82). We compare this with the result analogous to (1.72) for the reflection amplitude due to a thin slab of dielectric constant ε,

$$r_1 = \frac{i}{2}k\Delta z(\varepsilon-1). \tag{1.83}$$

Thus the effective dielectric constant of a slab of atoms of polarizability α and number density n is

$$\varepsilon \simeq 1+4\pi\alpha n. \tag{1.84}$$

We have neglected the effects of the dipolar fields on each other. When these are taken into account, the resulting dielectric constant for a uniform medium becomes.

$$\varepsilon = \frac{1+\frac{8}{3}\pi\alpha n}{1-\frac{4}{3}\pi\alpha n}. \tag{1.85}$$

This expression is known as the Clausius-Mossotti or Lorentz-Lorenz formula (Lorentz 1909). The result (1.84) is the first-order term in the αn expansion of (1.85). The form of (1.85), with $n = n(z)$, remains valid with a high degree of accuracy in a stratified medium of polarizable atoms (Castle and Lekner 1980; Lekner 1983).

1.6 A Look Ahead

In the preceding sections we have introduced the definitions and basic equations for the reflection of electromagnetic, particle and acoustic compressional waves by planar stratified media. The remainder of the book is written predominantly in

electromagnetic notation; a translation of the main results into particle-wave language is made in Chap. 15, and Chap. 16 deals with neutron and X-ray reflection. The final Chaps. 17–20 are on acoustic waves, chiral media, pulses and wavepackets, and finite beams. Here we preview the chapters, stating and discussing some of their results and techniques in order to give the reader a feel for the structure and content of the book.

Chapter 2 contains both general results, true for reflection and transmission at any transition between two homogeneous media, and some specific results for exactly solvable profiles. Among the general results are the conservation law

$$q_1(1-|r_{12}|^2)=q_2|t_{12}|^2, \tag{1.86}$$

and reciprocity relations such as

$$r_{21}=-\frac{t_{12}}{t_{12}^*}r_{12}^* \tag{1.87}$$

and

$$q_2t_{12}=q_1t_{21}. \tag{1.88}$$

The conservation law (1.86), which holds for real q_1 and q_2 and in the absence of absorption within the interface, represents conservation of energy in the electromagnetic case, and conservation of probability density current in the particle case. The relation (1.87) holds under the same conditions, and implies that the reflectance $R=|r|^2$ is the same from either side of incidence on a nonabsorbing interface. The relation (1.88) is more general, being valid also in the presence of absorption within the interface. It implies the equality of the transmittances $T_{12}=(q_2/q_1)|t_{12}|^2, T_{21}=(q_1/q_2)|t_{21}|^2$, representing the energy or particle flux through the inhomogeneity, for incidence from medium 1 or from medium 2. (When the polarization subscripts s and p are omitted, the relation quoted is understood to be valid for either wave.)

For inhomogeneous interfaces extending from z_1 to z_2, with $\varepsilon=\varepsilon_1$ for $z\le z_1$ and $\varepsilon=\varepsilon_2$, the s wave reflection and transmission amplitudes may be expressed in terms of the values and derivatives of two linearly independent solutions F and G of (1.7) within $z_1\le z\le z_2$, evaluated at z_1 and z_2:

$$r_s=\mathrm{e}^{2iq_1z_1}\frac{q_1q_2(F_1G_2-G_1F_2)+iq_1(F_1G_2'-G_1F_2')+iq_2(F_1'G_2-G_1'F_2)-(F_1'G_2'-G_1'F_2')}{q_1q_2(F_1G_2-G_1F_2)+iq_1(F_1G_2'-G_1F_2')-iq_2(F_1'G_2-G_1'F_2)+(F_1'G_2'-G_1'F_2')}, \tag{1.89}$$

$$t_s = \frac{e^{i(q_1 z_1 - q_2 z_2)} 2iq_1(F_2 G_2' - G_2 F_2')}{q_1 q_2(F_1 G_2 - G_1 F_2) + iq_1(F_1 G_2' - G_1 F_2') - iq_2(F_1' G_2 - G_1' F_2) + (F_1' G_2' - G_1' F_2')}. \tag{1.90}$$

Similar expressions can be written down for the p polarization. These results are useful for specific profiles for which the solutions are known functions, such as the Airy functions for the linear profile, and the Bessel functions for the exponential profile. General results may also be deduced from (1.89) and (1.90), for example that $r_s \to -1$ and $t_s \to 0$ at grazing incidence, and that r_s and t_s tend to the Fresnel values (1.15) as $\Delta z = z_2 - z_1$ tends to zero. From the p polarization expressions one finds that $r_p \to 1$ and $t_p \to 0$ at grazing incidence. Thus r_p/r_s always moves in the complex plane from $+1$ at normal incidence to -1 at grazing incidence, and the number of principal angles (or ellipsometric Brewster angles), defined by $\mathrm{Re}(r_p/r_s) = 0$, is therefore always odd.

Chapter 2 also lists the exact solutions for three dielectric function profiles which are solvable for both the s and p polarizations, and another (the important hyperbolic tangent profile) which is solvable for the s wave only. Two other cases which are solvable for the s wave case, the $\mathrm{sech}^2(z/a)$ and the linear profile, are discussed in Sects. 4.3 and 5.2 respectively, where their solution is relevant to the problem at hand.

Chapter 3 treats the reflection of long waves, that is those whose wavelength is large compared to the thickness of the reflecting inhomogeneity. The long-wave results are obtained from perturbation theories, which in turn derive from exact integral and integro-differential equations obeyed by the s and p waves. For example, from the perturbation theory for the s wave one finds that the reflection amplitude, to second order in the interface thickness, is given by

$$r_s = r_{s0} + \frac{2q_1\omega^2/c^2}{(q_1+q_2)^2}\left\{i\lambda_1 - 2q_2\lambda_2 - \frac{\omega^2/c^2}{q_1+q_2}\lambda_1^2\right\} + \cdots, \tag{1.91}$$

where the λ_n are integrals of dimension $(\text{length})^n$,

$$\lambda_n = \int_{-\infty}^{\infty} \mathrm{d}z[\varepsilon(z) - \varepsilon_0(z)]z^{n-1}. \tag{1.92}$$

In (1.92), $\varepsilon(z)$ is the dielectric function profile under consideration, and $\varepsilon_0(z)$ is the step dielectric function defined in (1.11), which has the reflection amplitude r_{s0} given in (1.13). The integrals λ_n depend on the relative positioning of the actual profile ε and the step profile ε_0. A theory which calculates reflection amplitudes as a perturbation series about a reference profile (here ε_0) must obtain results for observables, such as $|r_s|^2$, which are invariant to the relative positioning of the actual and reference profile. If $r = r_0 + r_1 + r_2 + \cdots$,

$$R = |r|^2 = |r_0|^2 + 2\,\mathrm{Re}(r_0^* r_1) + \{|r_1|^2 + 2\,\mathrm{Re}(r_0^* r_2)\} + \cdots, \tag{1.93}$$

and we see from (1.91) that the first order term r_1 is imaginary in the absence of absorption or total internal reflection for the s wave. (The same is true for the p wave, as is shown in Sect. 3.4.) Then there is no term in R of first order in the interface thickness, for either polarization. The second order term is given by the expression within the braces in (1.93); from (1.91) we have

$$R_1 = \left(\frac{q_1 - q_2}{q_1 + q_2}\right)^2 - \frac{4q_1 q_2 \omega^4/c^4}{(q_1 + q_2)^4} i_2 + \cdots, \tag{1.94}$$

where the *integral invariant* i_2 is given by

$$i_2 = 2(\varepsilon_1 - \varepsilon_2)\lambda_2 - \lambda_1^2. \tag{1.95}$$

(The subscript 2 denotes dimensionality $(\text{length})^2$.) The integrals λ_1 and λ_2 which enter into r_s and R_s, depend on the relative positioning of the actual and reference profiles, but the combination of integrals which comprise i_2 is invariant with respect to the choice of positioning. Similar results are obtained for the observables r_p/r_s, and $R_p = |r_p|^2$:

$$r_{s0}\left(\frac{r_p}{r_s}\right) = r_{p0} - \frac{2iQ_1}{(Q_1 + Q_2)^2}\frac{K^2}{\varepsilon_1 \varepsilon_2} I_1 + \cdots, \tag{1.96}$$

$$R_p = \left(\frac{Q_1 - Q_2}{Q_1 + Q_2}\right)^2$$

$$- \frac{4Q_1 Q_2}{\varepsilon_1 \varepsilon_2 (Q_1 + Q_2)^4}\left\{\frac{\omega^4}{c^4} i_2 - \frac{\omega^2}{c^2} K^2 \left[j_2 + \left(\frac{1}{\varepsilon_1} + \frac{1}{\varepsilon_2}\right) i_2\right] + \frac{K^4}{\varepsilon_1 \varepsilon_2}\left[(\varepsilon_1 + \varepsilon_2) j_2 - I_1^2\right]\right\} + \cdots, \tag{1.97}$$

In (1.97) j_2 is another second order invariant, and the first order invariant I_1 is defined by

$$I_1 = \int_{-\infty}^{\infty} \mathrm{d}z \frac{(\varepsilon_1 - \varepsilon)(\varepsilon - \varepsilon_2)}{\varepsilon} = \int_{-\infty}^{\infty} \mathrm{d}z \left\{\varepsilon_1 + \varepsilon_2 - \frac{\varepsilon_1 \varepsilon_2}{\varepsilon} - \varepsilon\right\} \tag{1.98}$$

These results show that, in the long wave limit, the observables R_s, R_p, and r_p/r_s take universal form. The integral invariants I_1, i_1, and i_2 depend on the profile shape and extent only. All frequency and angular dependence is contained in the coefficients of I_1, i_1, and i_2, and is the same for all non-singular profiles. (The degenerate

case $\varepsilon_1 = \varepsilon_2$ requires special consideration for the ellipsometric ratio r_p/r_s however.)

Equations (1.94), (1.96) and (1.97) illustrate how theory answers the question "what information can be obtained from a given experiment?" From (1.94) we see that measurement of the s reflectivity in the long wave case can determine only one characteristic of the interface, the invariant i_2. Experimental data at different angles of incidence give no new information (we are assuming that the interface has no roughness, and the absence of absorption in the interface or substrate), merely the opportunity to reduce the uncertainty in i_2. The same is true for ellipsometry to lowest order in the interface thickness: one parameter (the invariant I_1) can be determined, at any angle of incidence. The p wave reflectance (1.97) carries more information, because the direction of the electric field relative to the interface changes with the angle of incidence. In principle, the values of I_1^2, i_2 and i_2 may be determined by intensity measurements at a mimimum of three angles of incidence.

The long wave results described above were obtained from perturbation theory, the perturbation being the deviation of the actual profile $\varepsilon(z)$ from the step profile $\varepsilon_0(z)$. The simplest example of a perturbation theory expression for the reflection amplitude is that for reflection by a film between like media:

$$r_s^{\text{pert}} = -\frac{\omega^2/c^2}{2iq_0} \int_{-\infty}^{\infty} \mathrm{d}z(\varepsilon - \varepsilon_0)\, \mathrm{e}^{2iq_0 z}. \tag{1.99}$$

Here q_0 is the common value of q_1 and q_2; ε_0 is likewise the common value of ε_1 and ε_2. The normal incidence, thin film version of this result has been used in Sect. 1.5 ((1.72) and (1.83)). Note that r_s^{pert} diverges at grazing incidence (as $q_0 \to 0$). This is unphysical: for passive media the reflection amplitude must stay within the unit circle, and in fact we saw that the exact r_s tends to -1 at grazing incidence.

This troublesome divergence at grazing incidence remains in higher order perturbation expressions, but is removed by the variational theory developed in Chap. 4. The simplest trial function, $\psi_0 = \mathrm{e}^{iq_0 z}$, leads to the variational expression

$$r_s^{\text{var}} = \frac{-\frac{\omega^2/c^2}{2iq_0}\lambda(2q_0)}{1 + \frac{\omega^2/c^2}{2iq_0}\frac{\sigma(2q_0)}{\lambda(2q_0)}} \tag{1.100}$$

In this expression $\lambda(2q_0)$ is the Fourier integral in (1.99), and $\sigma(2q_0)$ is a double integral defined in Chap. 4. The variational result (1.100) is not divergent at grazing incidence; in fact it tends to the correct value of -1 as $q_0 \to 0$, since the integrals λ and σ have the property that $\sigma(0) = \lambda^2(0)$. Further, r_s^{var} is correct to second order in the film thickness, whereas r_s^{pert} is not. These properties are shared by the variational expressions, derived in Chap. 4, for s and p wave reflection amplitudes between unlike media.

Non-linear, first order differential equations (of the Riccati type) for the reflection amplitudes are derived in Chap. 5. Two kinds of equations are used: those for a quantity $\varrho(z)$ which tends to $re^{-2iq_1 z}$ as $z \to -\infty$, and those for $r(z)$, tending to r in the same limit. For the s wave, the respective equations are

$$\varrho' + 2iq\varrho = \frac{q'}{2q}\left(1 - \varrho^2\right), \tag{1.101}$$

$$r' = \frac{q'}{2q}\left(e^{2i\phi} - r^2 e^{-2i\phi}\right), \tag{1.102}$$

where primes denote differentiation with respect to z, and the phase integral ϕ is defined by

$$\phi(z) = \int^{z} d\zeta\, q(\zeta). \tag{1.103}$$

The corresponding equations for the p wave reflection amplitudes have Q'/Q instead of q'/q on the right-hand side. From (1.101) it is shown in Sect. 5.4 that R_s has the Fresnel reflectivity as an upper bound for all monotonic profiles:

$$R_s \le R_{s0} = \left(\frac{q_1 - q_2}{q_1 + q_2}\right)^2 \tag{1.104}$$

A similar bound holds for R_p when $Q(z) = q(z)/\varepsilon(z)$ is monotonic.

Integration of (1.102) from $z = -\infty$ to $+\infty$ gives

$$r_s = -\int_{-\infty}^{\infty} dz \frac{q'}{2q}\left(e^{2i\phi} - r^2(z)\, e^{-2i\phi}\right). \tag{1.105}$$

The $r(z)$ on the right-hand side is the reflection amplitude of a profile truncated at z. If the reflection is weak, one can get an approximate expression for r_s by omitting the term proportional to r^2 on the right. This is the weak reflection or Rayleigh approximation,

$$r_s^R = -\int_{-\infty}^{\infty} dz \frac{q'}{2q} e^{2i\phi}. \tag{1.106}$$

The corresponding approximation in the p wave case is

$$r_p^R = \int_{-\infty}^{\infty} \mathrm{d}z \frac{Q'}{2Q} \mathrm{e}^{2i\phi}. \tag{1.107}$$

At normal incidence both (1.106) and (1.107) reduce to

$$r_n^R = -\frac{1}{2} \int_{-\infty}^{\infty} \mathrm{d}z \frac{n'}{n} \mathrm{e}^{2i\phi_n}, \qquad \phi_n(z) = \frac{\omega}{c} \int^z \mathrm{d}\zeta\, n(\zeta), \tag{1.108}$$

where $n = \varepsilon^{1/2}$ is the refractive index. If one makes the further (and drastic) approximations of replacing $2n$ by $n_1 + n_2$ and ϕ_n by $(\omega/c)(n_1 + n_2)z = (k_1 + k_2)z$, (1.108) simplifies to

$$r_n \simeq -\frac{1}{n_1 + n_2} \int_{-\infty}^{\infty} \mathrm{d}z \frac{\mathrm{d}n}{\mathrm{d}z} \mathrm{e}^{i(k_1 + k_2)z}. \tag{1.109}$$

Expressions closely related to (1.109) have been used by Buff et al. (1965) and by Huang and Webb (1969) in the analysis of reflection from the diffuse interface of a binary mixture.

The Rayleigh approximation works very well when the reflection is weak, but fails near grazing incidence. The Rayleigh approximation (1.106) and the long wave limiting form (1.94) are compared in Fig. 1.13 with the exact reflectivity for the hyperbolic tangent profile

$$\varepsilon(z) = \frac{1}{2}(\varepsilon_1 + \varepsilon_2) - \frac{1}{2}(\varepsilon_1 - \varepsilon_2) \tanh \frac{z}{2a}\,. \tag{1.110}$$

For this profile the phase integral can be evaluated analytically (see Sect. 6.4), $i_2 = (\pi^2/3)(\varepsilon_1 - \varepsilon_2)^2 a^2$ from Table 3.1, and the exact reflectivity is (from Sect. 2.5)

$$R_s = \left\{ \frac{\sinh \pi(q_1 - q_2)a}{\sinh \pi(q_1 + q_2)a} \right\}^2. \tag{1.111}$$

The figure illustrates the strengths and limitations of the long wave and weak reflection approximations: the long wave expression is good at glancing incidence, where the effective wavelength $2\pi/q$ is large, while the Rayleigh approximation is good near normal incidence, but fails near glancing incidence, since the reflection is then strong (as always).

The reflection of short waves, that is those whose wavelength is small compared to the thickness of the interface, is discussed in Chap. 6. In the short wave limit the reflection properties usually approach the behaviour of classical particles, which are

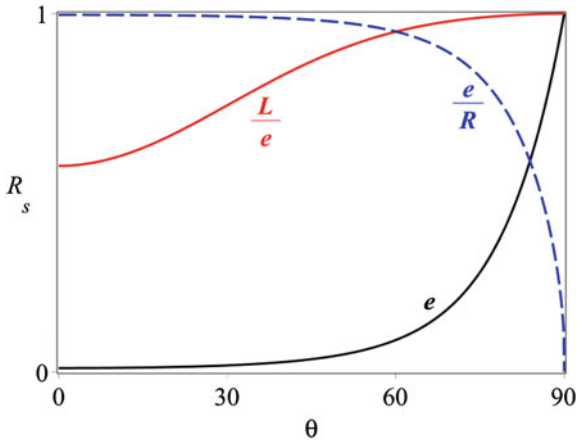


Fig. 1.13 Reflectivity of the s wave by the tanh profile (1.110), for $\varepsilon_1 = 1$ and $\varepsilon_2 = (4/3)^2$ and $\omega a/c = 0.2$. For this value of $\omega a/c$ the distance in which the dielectric function changes over 80 % of its range (from $(9\varepsilon_1 + \varepsilon_2)/10$ to $(\varepsilon_1 + 9\varepsilon_2)/10$) is about one seventh of the wavelength of the incident radiation. The curve e is the exact reflectivity (1.111), the *dashed curve* e/R gives the ratio of the exact to the Rayleigh reflectivities, and the curve L/e gives the ratio of the long-wave limiting form (1.94) to the exact value

either totally reflected or not reflected. Away from classical turning points, which located at the zeros of $q^2(z)$, approximate solutions of (1.7) are the Liouville-Green functions

$$\psi_+ = \left(\frac{q_1}{q}\right)^{\frac{1}{2}} \mathrm{e}^{i\phi}, \qquad \psi_- = \left(\frac{q_2}{q}\right)^{\frac{1}{2}} \mathrm{e}^{-i\phi} \tag{1.112}$$

(ϕ is the phase integral defined in (1.103)). A perturbation theory based on a Green's function constructed from $\psi_\pm$ gives the first order reflection amplitude

$$r_s^{(1)} = -\frac{1}{2}\int_{-\infty}^{\infty} \mathrm{d}\phi\, \mathrm{e}^{2i\phi}\left\{\gamma - \frac{\gamma^2}{4i}\right\}, \tag{1.113}$$

where

$$\gamma = \frac{1}{q^2}\frac{\mathrm{d}q}{\mathrm{d}z} = \frac{1}{q}\frac{\mathrm{d}q}{\mathrm{d}\phi} \tag{1.114}$$

is a dimensionless function which must be small for the short wave approximation to hold. The perturbation theory result is closely related to the Rayleigh approximation (which is the first term of another perturbation approach, the Bremmer series discussed in Sect. 6.5), as can be seen by writing (1.106) in the form

$$r_s^R = -\frac{1}{2}\int_{-\infty}^{\infty} \mathrm{d}\phi\, \gamma\, \mathrm{e}^{2i\phi}. \tag{1.115}$$

Unlike the long wave case, the reflection properties of short waves depend on the detail of the reflecting inhomogeneity, and do not take a universal form. For example: in the case of the profiles of finite range, which have a discontinuity in slope at the endpoints z_1 and z_2, both $r_s^{(1)}$ and r_s^R give

$$r_s = \frac{1}{4i}\,\mathrm{e}^{i(\phi_1+\phi_2)}\left\{\gamma_1\,\mathrm{e}^{-i\Delta\phi} - \gamma_2\,\mathrm{e}^{i\Delta\phi}\right\} + \cdots, \tag{1.116}$$

where ϕ_1 and ϕ_2 are the values of the phase integral at z_1 and z_2, $\Delta\phi = \phi_2 - \phi_1$, and ... denotes that exponentially small terms have been omitted. (The function γ changes in value from 0 to γ_1 at z_1, and from γ_2 to 0 at z_2). A similar result holds for the p wave:

$$r_p = \frac{1}{4i}\,\mathrm{e}^{i(\phi_1+\phi_2)}\left\{\gamma_1 \cos 2\theta_1\ \mathrm{e}^{-i\Delta\phi} - \gamma_2 \cos 2\theta_2\ \mathrm{e}^{i\Delta\phi}\right\} + \cdots. \tag{1.117}$$

Both the s and p reflectivities are thus oscillatory functions of $\Delta\phi$, and decay as the inverse square of the vacuum wavenumber ω/c. The dominant part of the s reflectivity is

$$R_s = \frac{1}{16}\left\{\gamma_1^2 + \gamma_2^2 - 2\gamma_1\gamma_2\,\cos 2\Delta\phi\right\} + \cdots. \tag{1.118}$$

(The p reflectivity has the same form, with $\gamma\cos 2\theta$ replacing γ.) This oscillatory behaviour, with amplitude decreasing with frequency, is characteristic of profiles with discontinuities in slope or higher order derivatives. Profiles with no such discontinuities, such as the hyperbolic tangent, show exponential decrease with $\omega a/c$ in the short wave limit, a being characteristic of the profile thickness.

Approximations such as (1.116) and (1.117), and the Rayleigh approximation, fail at grazing incidence, and in the presence of turning points. When there is a single turning point ($q^2 \le 0$ for $z \ge z_0$, say) there is total reflection. For the s wave

$$r_s = \mathrm{e}^{i\delta_s}, \qquad \delta_s \simeq 2\int_0^{z_0} \mathrm{d}z\, q(z) - \frac{\pi}{2}, \tag{1.119}$$

the phase decrement $\pi/2$ being universal for smooth profiles. In the case of two turning points ($q^2 < 0$ for $z_1 \le z \le z_2$), the classically forbidden region $q^2 < 0$ is tunneled through by a portion of the wave. The transmission amplitude then varies approximately as $\exp(-2\Delta\Phi)$, where $\Delta\Phi$ is the increment in the imaginary part of the phase integral between the turning points:

$$\Delta\Phi = \int_{z_1}^{z_2} \mathrm{d}z|q(z)|. \tag{1.120}$$

Reflection from anisotropic media is considered in Chaps. 7 and 8. Uniaxial systems are characterized by two dielectric functions $\varepsilon_o(z,\omega)$ and $\varepsilon_e(z,\omega)$. The most general uniaxial reflection problem, with arbitrary orientation of the optic axis relative to the reflecting surface and the plane of incidence, is discussed in Chap. 8. In the simplest case where the system retains azimuthal symmetry about the normal to the interface (Chap. 7), the optic axis is also normal to the interface, and $\varepsilon_o(z,\omega)$ and $\varepsilon_e(z,\omega)$ give the response of the system to electric field components respectively parallel and perpendicular to the interface. The resolution of electromagnetic waves into s and p components remains valid in this case, with the equations to be satisfied modified from (1.7) and (1.20) to

$$\frac{\mathrm{d}^2E}{\mathrm{d}z^2} + \left(\varepsilon_o\frac{\omega^2}{c^2} - K^2\right)E = 0, \tag{1.121}$$

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{\varepsilon_o}\frac{\mathrm{d}B}{\mathrm{d}z}\right) + \left(\frac{\omega^2}{c^2} - \frac{K^2}{\varepsilon_e}\right) = 0. \tag{1.122}$$

Equation (1.121) has the same form as (1.7), with ε_o replacing ε, but (1.122) differs from the isotropic case, since it contains both ε_o and ε_e. There are corresponding changes in the p wave reflection amplitude, and in r_p/r_s. The ellipsometric ratio, for example, still takes the form (1.96) in the long wave case, but the invariant I_1 is now given by

$$I_1 = \int_{-\infty}^{\infty} \mathrm{d}z\left\{\varepsilon_1 + \varepsilon_2 - \frac{\varepsilon_1\varepsilon_2}{\varepsilon_e} - \varepsilon_o\right\}. \tag{1.123}$$

(this applies to the case of an anisotropic thin film between isotropic media 1 and 2.) For reflection at a sharp boundary between an isotropic medium 1 and an anisotropic medium characterized by ε_o and ε_e, with its optic axis normal to the reflecting surface, there is zero reflection for the p polarization at

$$\theta_B = \arctan\left\{\frac{\varepsilon_o(\varepsilon_o - \varepsilon_1)}{\varepsilon_1(\varepsilon_e - \varepsilon_1)}\right\}^{1/2}. \tag{1.124}$$

In the case of an anisotropic film, characterized by $\varepsilon_o(z)$ and $\varepsilon_e(z)$, on a homogeneous anisotropic substrate characterized by ε_{2o} and ε_{2e}, the form (1.96) is still valid, with ε_{2e} replacing ε_2 in the factor multiplying I_1 and

$$I_1 = \int_{-\infty}^{\infty} \mathrm{d}z \left\{ \frac{\varepsilon_1^2 - \varepsilon_{2o}\varepsilon_{2e}}{\varepsilon_1 - \varepsilon_{2o}} - \frac{\varepsilon_1 - \varepsilon_{2e}}{\varepsilon_1 - \varepsilon_{2o}} \varepsilon_o - \frac{\varepsilon_1 \varepsilon_{2e}}{\varepsilon_e} \right\}. \tag{1.125}$$

Ellipsometry is discussed in Chap. 9. The emphasis is on the analysis of what various ellipsometric configurations measure. We discuss transmission as well as reflection ellipsometry.

The effect of absorption (the dissipation of electromagnetic energy within the medium) is discussed in Chap. 10. Absorption is included phenomenologically in the Maxwell equations by means of a complex dielectric function, $\varepsilon = \varepsilon_r + i\varepsilon_i$. This simple change has far-reaching consequences for reflection properties. In the case of reflection at the sharp surface of an absorbing medium (a metal, for example), the Fresnel equations (1.13) and (1.31) retain their form, but now $q_2 = q_r + iq_i$ and $Q_2 = Q_r + iQ_i = (q_r + iq_i)/(\varepsilon_r + i\varepsilon_i)$, where

$$\left(\frac{cq_r}{\omega}\right)^2 = \frac{1}{2}\left\{\varepsilon_r - \varepsilon_1 \sin^2\theta_1 + [(\varepsilon_r - \varepsilon_1 \sin^2\theta_1)^2 + \varepsilon_i^2]^{\frac{1}{2}}\right\}, \tag{1.126}$$

$$\frac{cq_i}{\omega} = \frac{\varepsilon_i/2}{cq_i/\omega}. \tag{1.127}$$

The ellipsometric ratio r_p/r_s no longer has the real axis as its trajectory, but lies within the upper half of the unit circle:

$$\frac{r_p}{r_s} = \frac{q_1^2(q_r^2 + q_i^2) - K^4 + 2iq_1q_iK^2}{(q_1q_r + K^2)^2 + q_1^2q_i^2}. \tag{1.128}$$

Some of the general results derived in Chap. 2 still hold, notably the fact that $r_s \to -1$ and $r_p \to 1$ at grazing incidence, and the implication that there is an odd number of principal angles of incidence at which $\mathrm{Re}(r_p/r_s) = 0$. The reciprocity relation (1.88) also holds, and thus the transmittance of an absorbing system is independent of the direction of propagation of the radiation.

Zero reflection is not possible off an absorbing medium with a sharp boundary, for either polarization. If however a dielectric layer is deposited on the absorber, zero reflectance is possible for both polarizations (at different angles of incidence); this interference-absorption effect thus produces reflection polarizers (Sect. 10.3).

A thin absorbing film on a transparent substrate always decreases the transmittance, but reflectance can be either increased or decreased, depending on the polarization and whether $\varepsilon_1 < \varepsilon_2$ or $\varepsilon_1 > \varepsilon_2$. For example, the s reflectivity to first order in the film thickness is given by

$$R_s = \left(\frac{q_1 - q_2}{q_1 + q_2}\right)^2 - \frac{4q_1(q_1 - q_2)}{(q_1 + q_2)^3}\frac{\omega^2}{c^2}\int_{-\infty}^{\infty} \mathrm{d}z\, \varepsilon_i(z) + \cdots. \tag{1.129}$$

The form (1.96) for the ellipsometric ratio is still valid, with I_1 complex:

$$I_1 = \int_{-\infty}^{\infty} \mathrm{d}z \left(\varepsilon_1 + \varepsilon_2 - \frac{\varepsilon_1 \varepsilon_2 \varepsilon_r}{\varepsilon_r^2 + \varepsilon_i^2} - \varepsilon_r \right) + i \int_{-\infty}^{\infty} \mathrm{d}z \left[\frac{\varepsilon_1 \varepsilon_2}{\varepsilon_r^2 + \varepsilon_i^2} \right] \varepsilon_i. \tag{1.130}$$

An important and dramatic effect due to absorption is that of *attenuated total reflection*, discussed in Sect. 10.6. An absorbing layer (typically a metal film) deposited between two dielectrics can turn a total reflection configuration into one whose p reflectance is small at resonance, and can be zero for proper choice of thickness of metal film and angle of incidence. This phenomenon is an interference-attenuation effect, associated with the resonant excitation of electromagnetic surface waves at a metal-dielectric interface.

Chapter 11 deals with the inversion of reflectance and ellipsometric data to obtain the parameters of the reflector. For example, if the real and imaginary parts ϱ_r and ϱ_i of r_p/r_s are measured at angle of incidence θ_1 and the interface is known to be sharply defined on the scale of the wavelength, the real and imaginary parts of ε may be found from

$$\frac{\varepsilon_r + i\varepsilon_i}{\varepsilon_1} = \sin^2\theta_1 + \sin^2\theta_1 \tan^2\theta_1 \frac{(1-\varrho_r^2)^2 - 4\varrho_i^2 + 4i(1-\varrho_r^2)\varrho_i}{[(1-\varrho_r)^2 + \varrho_i^2]^2}. \tag{1.131}$$

If a model reflection amplitude is constructed as a function of wave vector component in medium 1, and analytically continued to negative q_1 via $r(-q_1) = r^*(q_1)$, an explicit inversion is possible (Sect. 11.3) to obtain the dielectric function profile which would give this reflection amplitude in the Rayleigh approximation. In the s wave case the result is

$$\frac{\varepsilon(x)}{\varepsilon_1} \simeq \sin^2\theta_1 + \cos^2\theta_1 \exp\left(-4 \int_{-\infty}^{2x} \mathrm{d}y\, F_s(y) \right), \tag{1.132}$$

where $x = q_1^{-1} \int^z \mathrm{d}\zeta q(\zeta)$ and F_s is the Fourier transform of r_s:

$$F_s(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}q_1 \mathrm{e}^{-iq_1 y} r_s(q_1). \tag{1.133}$$

Matrix and numerical methods are developed in Chap. 12. Any stratified medium may be approximated by a set of homogeneous layers. The matrix methods connect, via a two-by-two matrix, the coefficients of either the two independent solutions, or the field amplitude and its derivative, at the entry and exit points of a layer. In the latter case these matrix relations for a homogeneous layer between z_n and z_{n+1} are as follows: for the s wave, with $D = \mathrm{d}E/\mathrm{d}z$,

$$\begin{pmatrix} E_{n+1} \\ D_{n+1} \end{pmatrix} = \begin{pmatrix} \cos\delta_n & q_n^{-1}\sin\delta_n \\ -q_n\sin\delta_n & \cos\delta_n \end{pmatrix} \begin{pmatrix} E_n \\ D_n \end{pmatrix}. \tag{1.134}$$

For the p wave, with $C = \varepsilon^{-1}\mathrm{d}B/\mathrm{d}z$, $Q_n = q_n/\varepsilon_n$, the matrix relation is

$$\begin{pmatrix} B_{n+1} \\ C_{n+1} \end{pmatrix} = \begin{pmatrix} \cos\delta_n & Q_n^{-1}\sin\delta_n \\ -Q_n\sin\delta_n & \cos\delta_n \end{pmatrix} \begin{pmatrix} B_n \\ C_n \end{pmatrix}. \tag{1.135}$$

For a profile approximated by N homogeneous layers, the reflection and transmission properties are determined by the profile matrix, which is a product of N layer matrices such as those in (1.134) or (1.135). If the elements of the profile matrix for the s polarization are s_{ij}, for example, the reflection and transmission amplitudes for an interface between media a and b are

$$r_s = \mathrm{e}^{2i\alpha}\frac{q_aq_bs_{12} + s_{21} + iq_as_{22} - iq_bs_{11}}{q_aq_bs_{12} - s_{21} + iq_as_{22} + iq_bs_{11}}, \tag{1.136}$$

$$t_s = \mathrm{e}^{i(\alpha-\beta)}\frac{2iq_a}{q_aq_bs_{12} - s_{21} + iq_as_{22} + iq_bs_{11}}. \tag{1.137}$$

(Here $\alpha = q_az_1$ and $\beta = q_bz_{N+1}$, z_1 and z_{N+1} being the boundaries of the inhomogeneity.) In the absence of absorption the matrix elements are real. The matrix formulation, and the results (1.136) and (1.137), remain valid in the presence of absorption also, but the matrix elements are now complex.

The matrices in (1.134) and (1.135) are unimodular (have unit determinant); this fact simplifies the treatment of *periodically stratified media* (Sect. 12.3 and Chap. 13), which in turn has important application to the multilayer dielectric mirrors. Numerical methods based on the matrix formulation are also discussed in Chap. 12. Reflection of s waves by an arbitrary layer extending from a to b can be represented, to second order in the layer thickness, by the s matrix

$$\begin{pmatrix} 1 - \int_a^b \mathrm{d}z\, q^2(z)(b-z) & b-a \\ -\int_a^b \mathrm{d}z\, q^2(z) & 1 - \int_a^b \mathrm{d}z\, q^2(z)(z-a) \end{pmatrix}. \tag{1.138}$$

(This result, and a similar one for the p matrix, are derived in Sect. 12.4.) In Sect. 12.8, a given interface is approximated by a set of layers within which the dielectric function $\varepsilon(z)$, and thus also $q^2(z)$, vary linearly with z. The matrix methods can be applied without modification to total reflection and tunneling; reflection and transmission through absorbing layers requires computation with complex matrix elements, the formalism being otherwise unaltered. Wavefunctions within the stratification may be obtained as a by-product of the profile matrix calculation.

Chapter 14 deals with reflection from rough surfaces. A planar stratified surface, no matter how diffuse, gives specular reflection of an incident plane wave, but rough surfaces scatter as well as reflect. The Rayleigh criterion for negligible roughness is (Sect. 14.1)

$$qh \ll 1 \tag{1.139}$$

where q (as always) is the normal component of the wavevector, and h is a measure of the variation in the height of the surface. When (1.139) is satisfied the surface will reflect specularly. According to the Rayleigh criterion, for given roughness and angle of incidence long waves may be reflected specularly and short waves diffusely, or for given roughness and wavelength there may be diffuse scattering near normal incidence and specular reflection near grazing incidence. Chapter 14 treats the reflection from corrugated surfaces (diffraction gratings), from liquid metal and liquid dielectric surfaces (scattering by thermally excited surface waves), in both cases using the methods of Rayleigh, and gives an outline of the application of the Helmholtz theorem to the scattering by rough surfaces (the Kirchhoff or surface integral method).

Chapter 15 adapts the content of the previous chapters to the language of particle waves obeying the Schrödinger equation. There follow three new chapters on special topics: *16 Neutron and X-ray reflection*, *17 Acoustic waves*, and *18 Chiral isotropic media*.

In the last two chapters we finally move away from the assumption that the incident field consists of unbounded plane waves: the reflection of electromagnetic pulses and particle wavepackets is considered in Chap. 19, and that of finite beams in Chap. 20. We find that nearly monochromatic pulses reflect (in the first approximation) without change of shape, with a time delay Δt determined by the frequency variation of the phase of the reflection amplitude:

$$\Delta t = \frac{\mathrm{d}\delta}{\mathrm{d}\omega}. \tag{1.140}$$

(The derivative is to be evaluated at the dominant angular frequency of the pulse.) For example: total reflection at normal incidence has $r_n = \mathrm{e}^{i\delta_n}$, and the short wave limiting form is found from (1.119) to be

$$\delta_n = 2\frac{\omega}{c}\int_0^{z_0} \mathrm{d}z\, n(z,\omega) - \frac{\pi}{2}, \tag{1.141}$$

where $n(z,\omega)$ is the refractive index, and z_0 is the turning point determined by $n(z_0,\omega) = 0$. (This formula applies to reflection from the ionosphere, for example, in which case $n^2 = \varepsilon \simeq 1 - \omega_p^2/\omega^2$, where ω_p is the plasma angular frequency, proportional to the square root of ionospheric electron density.) From (1.140) and

(1.141) we find that the time delay is the same as for pulse travel to the turning point and back at the group velocity $u(z,\omega) = \mathrm{d}\omega/\mathrm{d}k$, where $k = n\omega/c$:

$$\Delta t = 2\int_0^z \frac{dz}{u} = \frac{2}{c}\int_0^{z_0} \mathrm{d}z\left[n + \omega\frac{\partial n}{\partial \omega}\right]. \tag{1.142}$$

The Appendix of Chap. 19 summarises the universal properties of electromagnetic pulses. The other parts of Chap. 19 deal with the reflection of particle wavepackets, illustrated by exact solutions.

Pulses are built up from waves of differing frequencies. Bounded beams (Chap. 20) can be regarded as superpositions of plane waves of differing directions of propagation. Just as the reflection of pulses is determined by the frequency dependence of the reflection amplitude, the reflection of beams depends on the angular dependence of the reflection amplitude. There is a lateral shift on reflection of a beam of radiation,

$$\Delta x = -\frac{\mathrm{d}\delta}{\mathrm{d}K}, \tag{1.143}$$

where K is the lateral component of the wavevector ($K = k_x$), and the derivative is to be evaluated at the dominant value of K for the incident beam. A particularly interesting case is total reflection at a sharp boundary. For $\theta_1 > \theta_c$ the phase of the s wave reflection amplitude is

$$\delta_s = -2\ \arctan\frac{|q_2|}{q_1}, \tag{1.144}$$

and (1.143) leads to the beam shift

$$\Delta x_s = \frac{2K}{q_1|q_2|} = \frac{\lambda_1}{\pi}\frac{\tan\theta_1}{(\sin^2\theta_1 - \sin^2\theta_c)^{1/2}}, \tag{1.145}$$

where λ_1 is the wavelength $2\pi/k_1 = 2\pi c/n_1\omega$ in the first medium. This beam shift is divergent at the critical angle (where $q_2 = 0$), and in fact the formula (1.139) fails there, since (1.143) is derived on the assumption of a slow variation of the phase shift with angle. In practice (1.145) works well to close proximity of the critical angle, as discussed in Sect. 20.2. Appendix 1 in Chap. 20 we show that the $|q_2|$ singularity in the phase shift at $\theta_1 = \theta_c^+$ is universal for nonabsorbing profiles. Finally, Appendix 2 in Chap. 20 outlines the somewhat surprising polarization properties of finite electromagnetic beams.

References

Bergmann PG (1946) The wave equation in a medium with a variable index of refraction. J Acoust Soc Amer 17:329–333
Buff FP, Lovett RA, Stillinger FH (1965) Interfacial density profile for fluids in the critical region. Phys Rev Lett 15:621–623
Castle PJ, Lekner J (1980) Variation of the local field through the liquid-vapour interface. Physica 101A:99–111
Darwin CG (1924) The optical constants of matter. Trans Camb Phil Soc 23:137–167
Fermi E (1950) Nuclear physics (Notes compiled by Orear J, Rosenfeld AH, Schluter RA). University of Chicago Press, pp 201–202
Fresnel A (1823) Mémoire sur la loi des modifications que la réflexion imprime a la lumière polarisée. Mémoires de l'Académie 11:393–433
Green G (1838) On the reflexion and refraction of sound. Trans Camb Phil Soc 6:403–412
Hartree DR (1928) The propagation of electromagnetic waves in a stratified medium. Proc Camb Phil Soc 25:97–120
Huang JS, Webb WW (1969) Diffuse interface in a critical fluid mixture. J Chem Phys 50:3677–3693
Jackson JD (1975) Classical electrodynamics. John Wiley and Sons, New York
Lekner J (1983) Anisotropy of the dielectric function within a liquid-vapour interface. Mol Phys 49:1385–1400
Lorentz HA (1909/1952) The theory of electrons. Dover Publications, p 145

Further Readings

Books and review articles dealing with aspects of reflection in stratified media
Abeles F (1950) Recherches sur la propagation des ondes électromagnetiques sinusoidales dans les milieux stratifiés. Application aux couches minces. Annales de Physique 5:596–640
Born M, Wolf E (1965) Principles of optics, 3rd edn. Pergamon, New York
Brekhovskikh LM (1980) Waves in layered media, 2nd edn. Academic Press, New York
Budden KG (1985) The propagation of radio waves. Cambridge University Press, Cambridge
Drude P (1902/1959) Theory of optics. Dover publications
Froman N, Froman PO (1965) JWKB approximation. North-Holland, Amsterdam
Ginzburg VL (1964) The propagation of electromagnetic waves in plasmas. Pergamon, New York
Heading J (1962) An introduction to phase-integral methods. Methuen, London
Heading J (1975) Ordinary differential equations, theory and practice. Elek Science, London
Jacobsson R (1966) Light reflection from films of continuously varying refractive index. Prog Opt 5:247–286
Kennett BLN (1983) Seismic wave propagation in stratified media. Cambridge University Press, Cambridge
Knittl Z (1976) Optics of thin films. Wiley, New York
Landau LD, Lifshitz EM (1965) Quantum mechanics. Pergamon, New York
Levine H (1978) Unidirectional wave motions. North-Holland, Amsterdam
Rayleigh JWS (1912) On the propagation of waves through a stratified medium, with special reference to the question of reflection. Proc Roy Soc 86:207–266
Tolstoy I (1973) Wave propagation. McGraw-Hill, New York
Vašíček A (1960) Optics of thin films. North-Holland, Amsterdam
Wait JR (1962) Electromagnetic waves in stratified media. Pergamon, New York

A detailed discussion of all aspects of ellipsometry can be found in Chapter 9 and in

Azzam RMA, Bashara NM (1977) Ellipsometry and polarized light. See also Muller RH (1972) "Principles of ellipsometry", Advances in Electrochemistry and Electrochemical Engineering, vol. 9, Interscience

Sections 1.1 and 1.3 are based in part on

Lekner J (1982a) Reflection of long waves by interfaces. Physica 112A:544–556

The polarization of light by reflection from glass at 56° incidence, and from water at 53°, was discovered by Malus in 1808. In 1814 Brewster found, as the result of extensive experiments, that "the Index of refraction is the tangent of the angle of polarization". See

Brewster D (1853) *A treatise on optics*, Longman, Brown and Green (Chapter 24)

Lekner J (1990) Non-reflecting stratifications. Can J Phys 68:738–741

The explanation of near transparency of a general interface to the p wave given in Section 1.2, and of the similar effect in the reflection of acoustic waves (Section 1.4), is based on

Lekner J (1982b) Second-order ellipsometric coefficients. Physica 113A:506–520

The result (Section 1.5), for the reflection of electromagnetic waves in terms of the superposition of fields radiated by oscillating dipoles within a material slab, is an adaptation of a similar calculation for the transmitted wave, given by

Sargent M, Scully MO, Lamb WE (1974) Laser physics. Addison-Wesley, Appendix A

For a treatment of the molecular basis of reflection see

Hynne F, Bullough RK (1984) The scattering of light. I. The optical response of a finite molecular fluid. Phil. Trans. Roy. Soc. Lond. A312:251–293

Chapter 2
Exact Results

We shall first derive general results, namely those valid for an arbitrary interfacial profile (Sects. 2.1–2.3); some of these results will be restricted to non-absorbing interfaces and substrates. The remainder of the chapter is devoted to important special profiles for which the reflection amplitude may be obtained exactly. Both the general and the specific exact results are useful in testing approximate theories and numerical calculations.

The title of this chapter is not intended to imply that all other results in the book are approximate: for example in Chap. 3 we shall be deriving exact integral equations, and results which are exact (and general) to second order in the interface thickness.

2.1 Comparison Identities, and Conservation and Reciprocity Laws

In Sects. 1.1 and 1.2 we saw that, for planar stratified media, the s and p wave equations may be put in the form

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}z^2} + q^2\psi = 0, \tag{2.1}$$

with the reflection and transmission amplitudes defined in terms of the asymptotic forms of ψ at large negative and positive z:

$$\mathrm{e}^{iq_1 z} + r\mathrm{e}^{-iq_1 z} \leftarrow \psi \rightarrow t\mathrm{e}^{iq_2 z} \tag{2.2}$$

For the s wave $\psi = E$ and $q_E^2 = \varepsilon\omega^2/c^2 - K^2$. For the p wave, $\psi = b = (\varepsilon_1/\varepsilon)^{1/2}B$, and q^2 is given by (1.22):

$$q_b^2 = q_E^2 + \frac{1}{2\varepsilon}\frac{\mathrm{d}^2\varepsilon}{\mathrm{d}z^2} - \frac{3}{4}\left(\frac{1}{\varepsilon}\frac{\mathrm{d}\varepsilon}{\mathrm{d}z}\right)^2 \tag{2.3}$$

J. Lekner, *Theory of Reflection*, Springer Series on Atomic, Optical, and Plasma Physics 87, DOI 10.1007/978-3-319-23627-8_2

For the s wave, the r and t of (2.2) are r_s and t_s as defined in (1.10); for the p wave, from (1.25), $r = -r_p$ and $t = t_p$.

Let $\tilde{\psi}$ be the solution for another dielectric function profile $\tilde{\varepsilon}(z)$, which has the same limiting values ε_1 and ε_2 as $\varepsilon(z)$:

$$\frac{\mathrm{d}^2\tilde{\psi}}{\mathrm{d}z^2} + \tilde{q}^2\tilde{\psi} = 0, \quad \mathrm{e}^{iq_1 z} + \tilde{r}\mathrm{e}^{-iq_1 z} \leftarrow \tilde{\psi} \rightarrow \tilde{t}\mathrm{e}^{iq_2 z}. \tag{2.4}$$

We can obtain *comparison identities* relating r and $\tilde{r}$, and t and $\tilde{t}$ from the differential equations and their boundary conditions. The relation between r and $\tilde{r}$ is obtained as follows. We multiply the wave equation for $\tilde{\psi}$ by ψ, the wave equation for ψ by $\tilde{\psi}$, and subtract, to get

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\psi\frac{\mathrm{d}\tilde{\psi}}{\mathrm{d}z} - \tilde{\psi}\frac{\mathrm{d}\psi}{\mathrm{d}z}\right) = \left(q^2 - \tilde{q}^2\right)\psi\tilde{\psi}. \tag{2.5}$$

We now integrate from a point z_1 deep inside medium 1, to a point z_2 deep inside medium 2 (that is, z_1 and z_2 are such that ψ and $\tilde{\psi}$ have attained their asymptotic forms (2.2) and (2.4)). The integral of the left-hand side of (2.5) gives $2iq_1(\tilde{r} - r)$, all dependence on z_1 and z_2 cancelling out. Thus

$$r = \tilde{r} - \frac{1}{2iq_1}\int_{-\infty}^{\infty} \mathrm{d}z\left(q^2 - \tilde{q}^2\right)\psi\tilde{\psi}, \tag{2.6}$$

where we have replaced z_1 by $-\infty$ and z_2 by $+\infty$.

Another comparison identity may be obtained, in the same way, for ψ and $\tilde{\psi}^*$ (the complex conjugate of $\tilde{\psi}$):

$$2iq_1(1 - r\tilde{r}^*) - 2iq_2 t\tilde{t}^* = \int_{-\infty}^{\infty} \mathrm{d}z\left(q^2 - \tilde{q}^2\right)\psi\tilde{\psi}^*. \tag{2.7}$$

On setting $\tilde{\varepsilon} = \varepsilon$ (and thus $\tilde{q} = q$), we obtain

$$q_1\left(1 - |r|^2\right) = q_2|t|^2. \tag{2.8}$$

The energy density of a plane electromagnetic wave is proportional to $\varepsilon|E|^2$, and the speed to $1/\sqrt{\varepsilon}$; thus the intensity is proportional to $\sqrt{\varepsilon}|E|^2$. The amount of energy in the primary wave which is incident on a unit area of the interface in unit time is proportional to $\sqrt{\varepsilon_1}\cos\theta_1$. The energy carried away by the reflected wave is proportional to $\sqrt{\varepsilon_1}\cos\theta_1|r|^2$, and that carried by the transmitted wave is proportional to $\sqrt{\varepsilon_2}\cos\theta_2|t|^2$. Since $cq_1/\omega = \sqrt{\varepsilon_1}\cos\theta_1$ and $cq_2/\omega = \sqrt{\varepsilon_2}\cos\theta_2$, (2.8) expresses

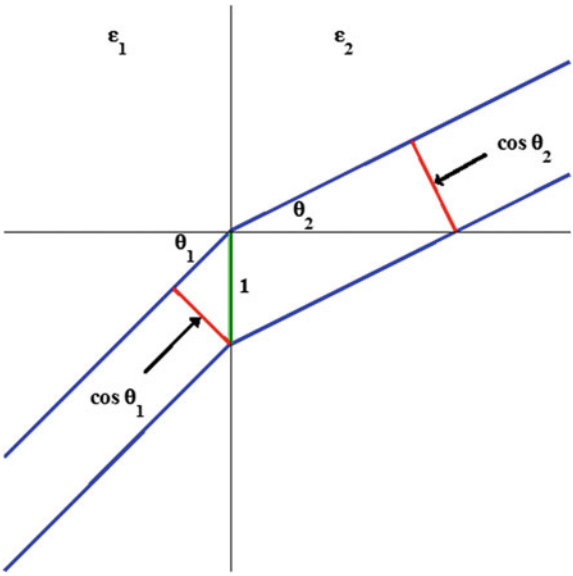


Fig. 2.1 A strip of unit width on the boundary is illuminated by a beam of width $\cos\theta_1$; the transmitted beam has width $\cos\theta_2$ (the reflected beam is not shown)

the law of conservation of energy. The geometry leading to the factors $\cos\theta_1$ and $\cos\theta_2$ is shown in Fig. 2.1.

In quantum mechanics (2.8) expresses the conservation of the probability density current. The conservation of energy is guaranteed, since (1.40), which leads to (2.1) on the substitution $\Psi = e^{iKx}\psi(z)$, is the Schrödinger equation for an energy eigenstate. The probability density current is

$$\mathbf{J} = \frac{\hbar}{2im}(\Psi^*\nabla\Psi - \Psi\nabla\Psi^*) = \frac{\hbar}{m}\mathrm{Im}(\Psi^*\nabla\Psi), \tag{2.9}$$

and has (for our geometry) zero y component, and a constant x component

$$J_x = \frac{\hbar K}{m}. \tag{2.10}$$

As ψ takes the asymptotic forms specified in (2.2), the z component of $\mathbf{J}$ takes the limiting values

$$\frac{\hbar q_1}{m}\left(1 - |r|^2\right) \leftarrow J_z \rightarrow \frac{\hbar q_2}{m}|t|^2. \tag{2.11}$$

Thus the form of the wavefunction guarantees conservation of probability density current tangential to the (smooth) interface, while (2.8) and (2.11) demonstrate the conservation of the normal component.

The relation (2.8) is valid only for real q: the reality of $\tilde{q}$ was assumed in obtaining (2.7). The extension of (2.8) to include absorption will be discussed in

Chap. 10; here we comment only on the case of total internal reflection, when $q_2 = i|q_2|$. Then $|r|^2 = 1$ (as we shall see in the next section), so that the left-hand side of (2.8) is zero. But $|t|^2$ is not zero in general: for example $|t_s|^2$ takes the value $4q_1^2/(q_1^2 + |q_2|^2)$ for the step profile. When total internal reflection occurs, there is no energy (or particle flux) propagated into the second medium, and t becomes the coefficient of an exponentially decaying wave: it is no longer a transmission amplitude of a propagating wave.

Further comparison identities may be obtained by comparing waves incident from opposite directions. We define ψ_{21} as a solution of (2.1) with asymptotic forms

$$t_{21}\mathrm{e}^{-iq_1 z} \leftarrow \psi_{21} \rightarrow \mathrm{e}^{-iq_2 z} + r_{21}\mathrm{e}^{iq_2 z}. \tag{2.12}$$

The previously used ψ and $\tilde{\psi}$, with asymptotics given by (2.2) and (2.4), will now be referred to as ψ_{12} and $\tilde{\psi}_{12}$ to emphasize that these represent motions originating in medium 1, and partially transmitted into medium 2. On comparing $\tilde{\psi}_{12}$ with ψ_{21}, we find

$$2i(q_2\tilde{t}_{12} - q_1 t_{21}) = \int_{-\infty}^{\infty} \mathrm{d}z\,(q^2 - \tilde{q}^2)\psi_{21}\tilde{\psi}_{12}. \tag{2.13}$$

When $\tilde{q} = q$, this gives (for real q_1, q_2, but possibly complex q within the interface) the reciprocity relation

$$q_2 t_{12} = q_1 t_{21}, \tag{2.14}$$

so that (2.13) can be written in a form similar to (2.6):

$$t_{12} = \tilde{t}_{12} - \frac{1}{2iq_2}\int_{-\infty}^{\infty} \mathrm{d}z\,(q^2 - \tilde{q}^2)\psi_{21}\tilde{\psi}_{12}. \tag{2.15}$$

Comparing ψ_{12} with $\tilde{\psi}_{21}$ gives a relation like (2.15), with $\psi_{12}\tilde{\psi}_{21}$ in the integrand. Finally, comparing ψ_{21} with ψ_{12}^* gives

$$-2i\left(q_2 r_{21}\tilde{t}_{12}^* + q_1 t_{21}\tilde{r}_{12}^*\right) = \int_{-\infty}^{\infty} \mathrm{d}z\,(q^2 - \tilde{q}^2)\psi_{21}\tilde{\psi}_{12}^*. \tag{2.16}$$

On setting $\tilde{q} = q$, we find

$$q_2 r_{21} t_{12}^* + q_1 r_{12}^* t_{21} = 0, \tag{2.17}$$

which together with (2.14) gives another reciprocity relation,

$$r_{21} = -\frac{t_{12}}{t_{12}^*} r_{12}^*. \tag{2.18}$$

The relations (2.17) and (2.18) are valid only in the absence of absorption or total internal reflection. In that case the $1 \rightarrow 2$ and $1 \leftarrow 2$ reflection amplitudes have equal absolute value.

The result $|r_{12}|^2 = |r_{21}|^2$, implied by (2.18), is remarkable in juxtaposition with our intuitive picture of particles going up or down a potential gradient (as in Fig. 1.8). In the semiclassical limit, one might expect stronger reflection for particle waves moving uphill than for those going downhill. In fact the reflectivity is exactly the same in the two cases, unless there is total internal reflection.

Some of the above relations were given (in a restricted form) by Stokes (1849), using the idea of reversing the wavemotions, as illustrated in Fig. 2.2. In the blue lines of Fig. 2.2, an incident wave of unit amplitude is split into a reflected and a transmitted wave, with amplitudes r_{12} and t_{12}. Stokes argued that on reversing the wavemotions, the amplitudes given in the caption result. On comparing the resultant amplitudes, one gets $r_{12}^2 + t_{12}t_{21} = 1$ and $r_{21} = -r_{12}$. The first of these, together with (2.14), gives $q_1\left(1 - r_{12}^2\right) = q_2 t_{12}^2$, which agrees with (2.8) when r_{12} and t_{12} are real. The second agrees with (2.18), under the same condition. Thus Stokes' results are valid when the phases associated with r and t are restricted to 0 and π.

A modification of Stokes' argument produces the general results: we replace his "reversion" by *time reversal plus complex conjugation*. The effect, shown below, is

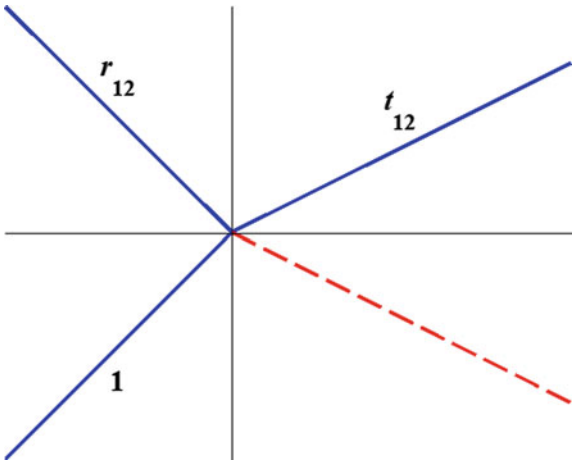


Fig. 2.2 Stokes' *principle of reversion*. The *solid lines* represent an incoming wave of unit amplitude and reflected and transmitted waves of amplitudes r_{12}, t_{12}. Reversing the wave motions leads to two incident waves, of amplitudes r_{12} from the *left*, and t_{12} from the *right*, and a new wave at *lower right* (*dashed*). According to Stokes, the wave at *lower left* now has amplitude $r_{12}^2 + t_{12}t_{21}$, while the wave at *lower right* has amplitude $t_{12}r_{21} + r_{12}t_{12}$

to give the correct analytical forms of the reversed beams. The first line gives the original waveforms, the second the time-reversed and complex-conjugated waveforms.

$$e^{iKx+iq_1z-i\omega t}+r_{12}e^{iKx-iq_1z-i\omega t},\quad t_{12}e^{iKx+iq_2z-i\omega t}$$

$$e^{-iKx-iq_1z-i\omega t}+r_{12}^{*}e^{-iKx+iq_1z-i\omega t},\quad t_{12}^{*}e^{-iKx-iq_2z-i\omega t}$$

Note that we have used the full space-time dependence in the waveforms. The arguments used above now give

$$r_{12}^{*}r_{12}+t_{12}^{*}t_{21}=1, \tag{2.19}$$

and

$$r_{12}^{*}t_{12}+t_{12}^{*}r_{21}=0. \tag{2.20}$$

We now have agreement with the general results derived earlier in this section: (2.19) with (2.14) reproduces (2.8), and (2.20) is the same as (2.18).

Other comparison identities can be obtained: for the p wave one analogous to (2.6) will be used in Sect. 3.4, and another analogous to (2.15) will be derived and used in Sect. 10.5.

2.2 General Expressions for r_s and r_p

Consider an interface for which $\varepsilon=\varepsilon_1$ for $z\le z_1$, and $\varepsilon=\varepsilon_2$ for $z\ge z_2$; the thickness $\Delta z=z_2-z_1$ of the inhomogeneous region can be large. By a limiting process, the results (to be derived below) for such finite-ranged interfaces can be extended to continuously varying interfaces. For example: for the hyperbolic tangent dielectric function

$$\varepsilon(z)=\frac{1}{2}(\varepsilon_1+\varepsilon_2)-\frac{1}{2}(\varepsilon_1-\varepsilon_2)\tanh[(z-z_0)/2a], \tag{2.21}$$

one can take $z_1=z_0-\Delta z/2$, $z_2=z_0+\Delta z/2$, and by making $\Delta z/a$ large enough, any desired accuracy can be achieved.

For the s wave, $\mathbf{E}=(0,E(z)e^{i(Kx-\omega t)},0)$ where $E(z)$ satisfies

$$\frac{d^2E}{dz^2}+q^2E=0. \tag{2.22}$$

This second-order linear differential equation has, for arbitrary form of $\varepsilon(z)$, two linearly independent solutions. We call these $F(z)$ and $G(z)$ in the region $z_1 \le z \le z_2$. Then

$$E(z) = \begin{cases} e^{iq_1 z} + r_s e^{-iq_1 z} & (z < z_1) \\ \alpha F(z) + \beta G(z) & (z_1 \le z \le z_2) \\ t_s e^{iq_2 z} & (z > z_2). \end{cases} \tag{2.23}$$

When $\varepsilon(z)$ has no delta-function singularities (or worse), as we assume here, E and dE/dz are continuous. Continuity of E and dE/dz at z_1 and z_2 gives us four linear equations in the four unknowns r_s, t_s, α, β; namely

$$\begin{aligned} e^{iq_1 z_1} + r_s e^{-iq_1 z_1} &= \alpha F_1 + \beta G_1, \\ iq_1 \left(e^{iq_1 z_1} - r_s e^{-iq_1 z_1}\right) &= \alpha F_1' + \beta G_1' \\ \alpha F_2 + \beta G_2 &= t_s e^{iq_2 z_2}, \\ \alpha F_2' + \beta G_2' &= iq_2 t_s e^{iq_2 z_2}. \end{aligned} \tag{2.24}$$

We have written F_1 for $F(z_1), F_1'$ for dF/dz evaluated at z_1, etc. Solving for r_s we find

$$r_s = e^{2iq_1 z_1} \frac{q_1 q_2 (F_1 G_2 - G_1 F_2) + iq_1 \left(F_1 G_2' - G_1 F_2'\right) + iq_2 \left(F_1' G_2 - G_1' F_2\right) - \left(F_1' G_2' - G_1' F_2'\right)}{q_1 q_2 (F_1 G_2 - G_1 F_2) + iq_1 \left(F_1 G_2' - G_1 F_2'\right) - iq_2 \left(F_1' G_2 - G_1' F_2\right) + \left(F_1' G_2' - G_1' F_2'\right)}. \tag{2.25}$$

The corresponding expression for the transmission amplitude is

$$t_s = \frac{e^{i(q_1 z_1 - q_2 z_2)} 2iq_1 (F_2 G_2' - G_2 F_2')}{q_1 q_2 (F_1 G_2 - G_1 F_2) + iq_1 \left(F_1 G_2' - G_1 F_2'\right) - iq_2 \left(F_1' G_2 - G_1' F_2\right) + \left(F_1' G_2' - G_1' F_2'\right)}. \tag{2.26}$$

The bilinear form in the numerator of t_s is the Wronskian, W, of the solutions F and G of (2.22), and is a constant:

$$W = FG' - GF', \; W' = FG'' - GF'', \tag{2.27}$$

and $F'' = -q^2 F, \; G'' = -q^2 G$ from (2.22); thus W' is zero and W is independent of z.

Some general properties of r_s and t_s follow directly from the forms of (2.25) and (2.26), for arbitrary nonsingular profiles.

We first note that in the absence of absorption (ε real), F and G may be chosen to be real. This is because they are solutions of a linear differential equation with real coefficients: if F is a complex solution of (2.22), so is F^* and therefore so is $F^* + F$, which is real. Henceforth, F and G are taken to be real functions, unless otherwise indicated.

Total internal reflection occurs when $q_2^2 \leq 0$ (that is, when $\sin^2\theta_1 \geq \varepsilon_2/\varepsilon_1$, as discussed in Sect. 1.3); when $q_2 = i|q_2|$, the wave in medium 2 decays exponentially as $e^{-|q_2|z}$. Thus no *propagation* of the wave into medium 2 takes place, and we expect the reflection amplitude to lie on the unit circle. This is true, as may be verified directly from (2.25), which takes the form $e^{2iq_1z_1}(-f+ig)/(f+ig)$, with f and g real, when $q_2 = i|q_2|$. From the last two equations of (2.24) we find that $\beta/\alpha = -(|q_2|F_2 + F_2')/(|q_2|G_2 + G_2')$, which is real if F and G are chosen to be real. Thus $\psi = \alpha F + \beta G$ has $\text{Im}\,\psi/\text{Re}\,\psi = \text{Im}\,\alpha/\text{Re}\,\alpha$, and it follows from (2.24) that this ratio is equal to $\text{Im}\,r_s/(1+\text{Re}\,r_s) = \tan(\delta_s/2)$ when $r_s = \exp\, i\delta_s$. The real and imaginary parts of the wavefunction are proportional to each other in total reflection. The wave motion normal to the interface is then represented by a standing wave.

The first equality in (2.27) may be regarded as a first order linear differential equation for G:

$$G' - \frac{F'}{F}G = \frac{W}{F}. \tag{2.28}$$

This has the solution

$$G(z) = WF(z)\int^{z} \mathrm{d}\zeta/F^2(\zeta). \tag{2.29}$$

For non-zero W, this is a second solution, linearly independent of F, of the differential equation (2.22). Let $\int_1^2$ denote $\int_{z_1}^{z_2}\mathrm{d}z/F^2(z)$, and set $W = 1$. Then

$$\begin{aligned}
F_1G_2 - G_1F_2 &= F_1F_2\int_1^2,\\
F_1G_2' - G_1F_2' &= F_1/F_2 + F_1F_2'\int_1^2,\\
F_1'G_2 - G_1'F_2 &= -F_2/F_1 + F_1'F_2\int_1^2,\\
F_1'G_2' - G_1'F_2' &= \frac{F_1'}{F_2} - \frac{F_2'}{F_1} + F_1'F_2'\int_1^2.
\end{aligned} \tag{2.30}$$

As Δz tends to zero, the first and last bilinear forms tend to zero, and the second and third tend to $+1$ and -1, respectively. Thus the reflection and transmission amplitudes for an arbitrary non-singular profile of extent Δz approach the step profile values (1.15) as $\Delta z \to 0$.

For passive media we must have $\left|r^2\right| \leq 1$. After some algebra, the physical requirement that $|r_s|^2 \leq 1$ reduces to $W(z_1)W(z_2) \geq 0$, where W is the Wronskian $FG' - GF'$. We have seen that $W(z_1) = W(z_2)$, so this condition is satisfied identically.

The conservation law $q_1\left(1 - |r|^2\right) = q_2|t|^2$ proved in the last section may be verified directly from (2.25) to (2.26). As in the result $|r_s|^2 \leq 1$, the proof involves the identity

$$(F_1G_2 - G_1F_2)\left(F_1'G_2' - G_1'F_2'\right) - \left(F_1G_2' - G_1F_2'\right)\left(F_1'G_2 - G_1'F_2\right) = W^2. \quad (2.31)$$

We now turn to the reciprocity laws (2.14) and (2.18), which relate t_{21} to t_{12} and r_{21} to t_{12} and r_{12} (the suffix s will be dropped for the moment). The results (2.25) and (2.26) may be rewritten as

$$r_{12} = \mathrm{e}^{2iq_1z_1}\frac{N}{D}, \quad (2.32)$$

$$t_{12} = \mathrm{e}^{i(q_1z_1 - q_2z_2)}\frac{2iq_1W}{D}, \quad (2.33)$$

where N is the numerator of (2.25), and D is the denominator common to (2.25) and (2.26). The corresponding results for r_{21} and t_{21} follow from the continuity of E and $\mathrm{d}E/\mathrm{d}z$ at z_1 and z_2, where E now has the forms outside the interval $[z_1, z_2]$ as given by (2.12). We find

$$r_{21} = \mathrm{e}^{-2iq_2z_2}\frac{N^*}{D}, \quad (2.34)$$

$$t_{21} = \mathrm{e}^{i(q_1z_1 - q_2z_2)}\frac{2iq_2W}{D}. \quad (2.35)$$

The reciprocity law (2.14) follows on comparison of (2.33) and (2.35); note that the reality of F and G (and thus the lack of absorption within the interface) is not needed. Both media 1 and 2 must be nonabsorbing, however, for the asymptotic forms (2.2) and (2.12) to be valid.

The reciprocity law (2.18) follows from (2.32) to (2.34); in this case the reality of F and G has been assumed in cancelling W with W^* and in writing the numerator of r_{21} as N^*, the general expression being

$$\begin{aligned} N_{21} = {} & q_1q_2(F_1G_2 - G_1F_2) - iq_1\left(F_1G_2' - G_1F_2'\right) - iq_2\left(F_1'G_2 - G_2'F_2\right) \\ & - \left(F_1'G_2' - G_1'F_2'\right) \end{aligned} \quad (2.36)$$

We now turn to the p wave reflection and transmision amplitudes. For the standard geometry defined in Chap. 1, $\mathbf{B} = (0, e^{i(Kx-\omega t)}B(z), 0)$ where $B(z)$ satisfies

$$\frac{d}{dz}\left(\frac{1}{\varepsilon}\frac{dB}{dz}\right) + \left(\frac{\omega^2}{c^2} - \frac{K^2}{\varepsilon}\right)B = 0 \tag{2.37}$$

We will derive general expressions for r_p and t_p, analogous to the results for r_s and t_s. Let $C(z)$ and $D(z)$ be two linearly independent solutions of (2.37) within the interval $[z_1, z_2]$. Then

$$B(z) = \begin{cases} e^{iq_1 z} - r_p e^{-iq_1 z} & z < z_1 \\ \gamma C(z) + \delta D(z) & z_1 \le z \le z_2 \\ \left(\frac{\varepsilon_2}{\varepsilon_1}\right)^{1/2} t_p e^{iq_2 z} & z > z_1. \end{cases} \tag{2.38}$$

The form of (2.37) shows that $\varepsilon^{-1}dB/dz$ must be continuous (discontinuity would give rise to a delta-function term, which we assume to be absent from $\varepsilon(z)$). From the continuity of B and $\varepsilon^{-1}dB/dz$ at z_1 and z_2 we obtain four equations in the four unknowns r_p, t_p, γ and δ. These are, for ε continuous at z_1 and z_2,

$$\begin{aligned} e^{iq_1 z_1} - r_p e^{-iq_1 z_1} &= \gamma C_1 + \delta D_1, \\ iq_1\left(e^{iq_1 z_1} + r_p e^{-iq_1 z_1}\right) &= \gamma C_1' + \delta D_1', \\ \gamma C_2 + \delta D_2 &= \left(\frac{\varepsilon_2}{\varepsilon_1}\right)^{1/2} t_p e^{iq_2 z_2}, \\ \gamma C_2' + \delta D_2' &= \left(\frac{\varepsilon_2}{\varepsilon_1}\right)^{1/2} t_p iq_2 e^{iq_2 z_2}. \end{aligned} \tag{2.39}$$

Solving for r_p and t_p, we find

$$-r_p = e^{2iq_1 z_1}\frac{q_1 q_2(C_1 D_2 - D_1 C_2) + iq_1(C_1 D_2' - D_1 C_2') + iq_2(C_1' D_2 - D_1' C_2) - (C_1' D_2' - D_1' C_2')}{q_1 q_2(C_1 D_2 - D_1 C_2) + iq_1(C_1 D_2' - D_1 C_2') - iq_2(C_2' D_2 - D_1' C_2) + (C_1' D_2' - D_1' C_2')} \tag{2.40}$$

$$\left(\frac{\varepsilon_2}{\varepsilon_1}\right)^{1/2} t_p = \frac{e^{i(q_1 z_1 - q_2 z_2)} 2iq_1(C_2 D_2' - D_2 C_2')}{q_1 q_2(C_1 D_2 - D_1 C_2) + iq_1(C_1 D_2' - D_1 C_2') - iq_2(C_1' D_2 - D_1' D_2) + (C_1' D_2' - D_1' C_2')}. \tag{2.41}$$

All of the general properties derived in Sect. 2.1 may be verified for the p wave. The proofs are as for the s wave, with a slight difference in the case of the reciprocity relations (2.14) and (2.18), which we will make explicit. We write (2.40) and (2.41) as

$$-r_{12} = \mathrm{e}^{-2iq_1z_1}\frac{N_{12}}{D}, \tag{2.42}$$

$$\left(\frac{\varepsilon_2}{\varepsilon_1}\right)^{1/2} t_{12} = \mathrm{e}^{i(q_1z_1-q_2z_2)}\frac{2iq_1W_2}{D}, \tag{2.43}$$

where N_{12} is the denominator of (2.40), D is the denominator common to (2.40) and (2.41), and $W_2 = C_2D_2' - D_2C_2'$ is the Wronskian at z_2 of the pair of solutions of (2.37). (The Wronskian for (2.37) is not independent of z, as we shall see shortly.) The corresponding expressions for a wave incident from medium 2 are

$$-r_{21} = \mathrm{e}^{-2iq_2z_2}\frac{N_{21}}{D}, \tag{2.44}$$

$$\left(\frac{\varepsilon_1}{\varepsilon_2}\right)^{1/2} t_{21} = \mathrm{e}^{i(q_1z_1-q_2z_2)}\frac{2iq_2W_1}{D}, \tag{2.45}$$

where

$$\begin{aligned} N_{21} = {} & q_1q_2(C_1D_2 - D_1C_2) - iq_1\left(C_1D_2' - D_1C_2'\right) - iq_2\left(C_1'D_2 - D_1'C_2\right) \\ & - \left(C_1'D_2' - D_1'C_2'\right) \end{aligned} \tag{2.46}$$

From (2.37), the Wronskian $W = CD' - DC'$ has the derivative

$$W' = CD'' - DC'' = \frac{\varepsilon'}{\varepsilon}W, \tag{2.47}$$

and so W is proportional ε, or W/ε is constant. The relation $q_2t_{12} = q_1t_{21}$ follows from (2.43) and (2.45) on using

$$\frac{W_1}{\varepsilon_1} = \frac{W_2}{\varepsilon_2}, \tag{2.48}$$

which we have just proved. The relation (2.18) or (2.20) follows from (2.42) to (2.44) provided $N_{21} = N_{12}^*$. It thus holds in the absence of absorption within the interface.

On using the identity

$$(C_1D_2 - D_1C_2)\left(C_1'D_2' - D_1'C_2'\right) - \left(C_1D_2' - D_1C_2'\right)\left(C_1'D_2 - D_1'C_2\right) = W_1W_2 \tag{2.49}$$

we find that $|r_p|^2 \leq 1$ provided $(W_1/\varepsilon_1)(W_2/\varepsilon_2) \geq 0$, which follows from (2.48). Finally, the conservation law $q_1\left(1-|r_p|^2\right) = q_2|t_p|^2$ follows from (2.40) and (2.41) on using (2.48) and (2.49).

Generalization of these results to profiles with discontinuities in the dielectric function is straightforward: see Lekner (1990b) for the electrodynamic case, and also Sect. 17.1 for acoustic waves.

The case of *symmetric stratifications* is interesting: suppose that the inhomogeneity extends from $-z_1$ to z_1, and that $z = 0$ is a plane of symmetry. Then the two linearly independent solutions $[F(z), G(z)\ or\ C(z), D(z)]$ can be taken as even and odd, for example $F(-z) = F(z),\ G(-z) = -G(z)$. The reflection amplitudes for a non-absorbing profile then have a real numerator, in contrast to the non-symmetric case when both numerator and denominator are complex. Thus zero reflection is much easier to accomplish for symmetric profiles (Lekner 1990a).

2.3 Reflection at Grazing Incidence, and the Existence of a Principal Angle

We will show that $r_s \to -1$ and $r_p \to +1$ at grazing incidence, exactly and without ambiguity of phase. A direct consequence is that a principal angle (the ellipsometric Brewster angle, defined by location of the zero of the real part of r_p/r_s) always exists. These results hold for interfaces with arbitrary dielectric function profiles, for internal as well as external reflections, and in the presence of absorption within the reflecting layer or its substrate. In Chap. 7 we shall see that the results also hold for those anisotropic media for which the s and p wave characterization is adequate. For the more general anisotropic or chiral cases, see Sects. 9.6 and 18.4.

At grazing incidence, $\theta_1 \to \pi/2$ and $q_1 = \sqrt{\varepsilon_1}(\omega/c)\cos\theta_1 \to 0$. The functions F and G in the general expression (2.25) also depend on angle of incidence, through $q^2(z) = \varepsilon(z)\omega^2/c^2 - K^2$, where $K = \sqrt{\varepsilon_1}(\omega/c)\sin\theta_1 = \sqrt{\varepsilon_2}(\omega/c)\sin\theta_2$. Thus at grazing incidence $q^2(z_1) = \varepsilon_1\omega^2/c^2 - K^2 \to 0$, but this does not imply singular behaviour of F or G through (2.22). On letting $q_1 \to 0$ in (2.25), we find that $r_s \to -1$.

Note that there is usually an arbitrariness of the phase of a reflection amplitude, associated with the arbitrariness of choice of the origin of coordinates. For example, the reflection amplitude of a step profile located at z_1 is given by (1.15):

$$r_{s0} = \mathrm{e}^{2iq_1z_1}\frac{q_1 - q_2}{q_1 + q_2}, \tag{2.50}$$

and carries the origin-dependent phase $2q_1z_1$. But as $q_1 \to 0$, this phase arbitrariness disappears. We have just shown this to be true for all profiles: at grazing incidence the reflection amplitude is known in magnitude and in phase. The incident and reflected waves are then both moving parallel to the interface, and there is no motion perpendicular to the interface to give rise to a phase shift associated with

the path difference $2z_1$ between the incident and reflected waves. The reality of F and G, or of q_2, has not been assumed. Thus there is total reflection at grazing incidence, with reversal of the electric field, even in the presence of absorption and irrespective of the sign of $\varepsilon_1 - \varepsilon_2$.

A similar result holds for the p wave. On letting $q_1 \to 0$ in (2.40), we find $r_p \to +1$. This result, together with (1.27) shows that again the electric field is reversed on reflection at grazing incidence.

Thus the reflected electric fields of both the s and p waves are exactly out of phase with the incident electric fields, whether the reflecting surface is metallic or dielectric, sharp or diffuse, and for internal as well as external reflection. It follows that Lloyd's mirror experiment will produce diffraction fringes, with destructive interference at the mirror's edge, under these very general conditions. This is in accord with experiment (Jenkins and White 1950, Sections 13.8 and 28.10). Section 16.5 deals with neutron Lloyd's mirror experiments.

The convention in use throughout this book, and established in Chap. 1, has $r_p = r_s$ at normal incidence, where the s and p waves are physically indistinguishable. Thus the ratio $r_p/r_s = +1$ at normal incidence, and tends to -1 at grazing incidence. At general incidence r_p/r_s is a complex number, with no ambiguity of phase, since in taking the ratio one cancels out the arbitrary phase factors associated with the choice of origin. The ratio r_p/r_s is measured by ellipsometry. In polarization modulation ellipsometry (Jasperson and Schnatterly 1969; Beaglehole 1980), it is experimentally most convenient to measure $\mathrm{Im}(r_p/r_s)$ at the angle where $\mathrm{Re}(r_p/r_s) = 0$ (the principal angle). The vanishing of the real part of r_p/r_s is one of several possible operational definitions of generalized Brewster angles; other possibilities are locations of minima of $|r_p|^2$ or of $|r_p/r_s|^2$. For the step profile, with r_s given by (2.50) and

$$r_{p0} = -\mathrm{e}^{2iq_1z_1}\frac{Q_1 - Q_2}{Q_1 + Q_2}, \tag{2.51}$$

all these definitions reduce to the Brewster angle (1.33), determined by $Q_1 = Q_2$, that is, purely in terms of the dielectric functions of media 1 and 2.

The question arises as to whether the principal angle or ellipsometric Brewster angle, defined by the location of $\mathrm{Re}(r_p/r_s) = 0$, always exists (Lekner 1985). The answer is *yes*: we have seen that r_p/r_s moves in the complex plane from the point $+1$ at normal incidence to the point -1 at grazing incidence, and it follows that it must cross the line $\mathrm{Re}(r_p/r_s) = 0$ at least once (and in general an odd number of times). This is a consequence of the continuity of solutions of linear differential equations as a function of the parameters of the equations. Here the parameter is the angle of incidence, appearing in the differential equations through K^2. See for example Birkhoff and Rota (1969), Sections 4 and 10 of Chapter 6. Some particular paths of r_p/r_s in the complex plane as θ_1 varies are shown in Fig. 2.3; a further case, illustrating triple principal angles, appears in Fig. 2.4. Both are based on the homogeneous layer, discussed in the next section.

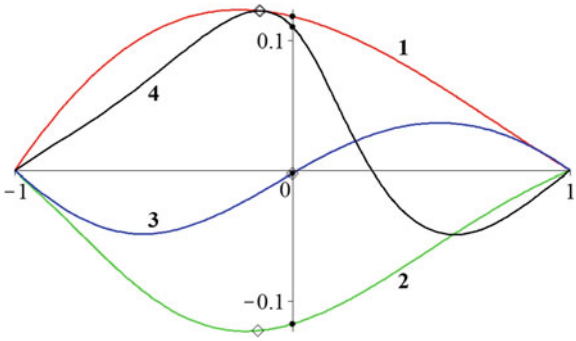


Fig. 2.3 Four paths of r_p/r_s in the complex plane, for a homogeneous layer of thickness Δz. The real and imaginary axes are *horizontal* and *vertical*. The curves for four values $\left(\frac{\omega}{c}\right)\Delta z = 1, 2, 3, 4$ (indicated on the paths) are shown. At normal incidence all paths coincide at $+1$, at glancing incidence they coincide at -1; they cut the vertical axis at the principal angle, where $\mathrm{Re}(r_p/r_s) = 0$ (*solid circles*). The *diamonds* are located at $\theta_B = \arctan(\varepsilon_2/\varepsilon_1)^{1/2}$, the Brewster angle for a film of vanishing thickness. In the case $\left(\frac{\omega}{c}\right)\Delta z = 3$ the Brewster and principal angles are nearly coincident. The curves are drawn for $\varepsilon_1 = 1, \varepsilon = (4/3)^2$ and $\varepsilon_2 = (3/2)^2$, approximating a layer of water on glass

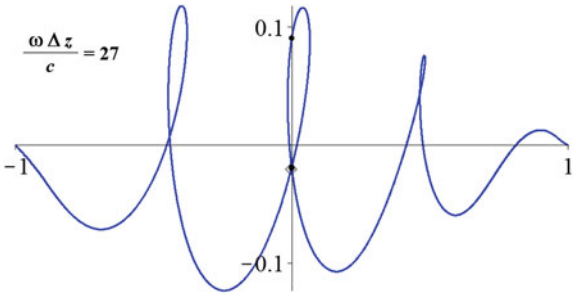


Fig. 2.4 Illustration of triple principal angles. The curve is the locus of r_p/r_s in the complex plane, as a function of the angle of incidence. The thickness of the homogeneous layer is about four wavelengths, $(\omega/c)\Delta z = 27$. The values of ε_1, ε and ε_2 are as in Fig. 2.3, representing a layer of water on glass. The diamond is located at the zero-thickness Brewster angle, $\theta_B = \arctan(\varepsilon_2/\varepsilon_1)^{1/2}$, the *solid circles* at the three principal angles (the first and third crossings of $\mathrm{Re}(r_p/r_s) = 0$ are nearly coincident). The tangents of the three principal angles are $1.110, 1.167, 1.497$. The last is nearly equal to $\tan\theta_B = 1.5$

2.4 Reflection by a Homogeneous Layer

After the step dielectric function profile, the simplest and most commonly occurring is the two-step profile, representing a homogeneous layer between media 1 and 2 (Fig. 2.5).

In the interval $z_1 \leq z \leq z_2$ the s wave equation (2.22) has $q^2 = \varepsilon\omega^2/c^2 - K^2$, with ε constant. The solutions are thus $e^{\pm iqz}$ or $\cos qz$, $\sin qz$. On matching E and dE/dz at z_1 and $z_2 = z_1 + \Delta z$ to $e^{iq_1 z} + r_s e^{-iq_1 z}$ and $t_s e^{iq_2 z}$ respectively, we find

$$r_s = e^{2iq_1 z_1} \frac{q(q_1 - q_2)c + i(q^2 - q_1 q_2)s}{q(q_1 + q_2)c - i(q^2 + q_1 q_2)s}, \tag{2.52}$$

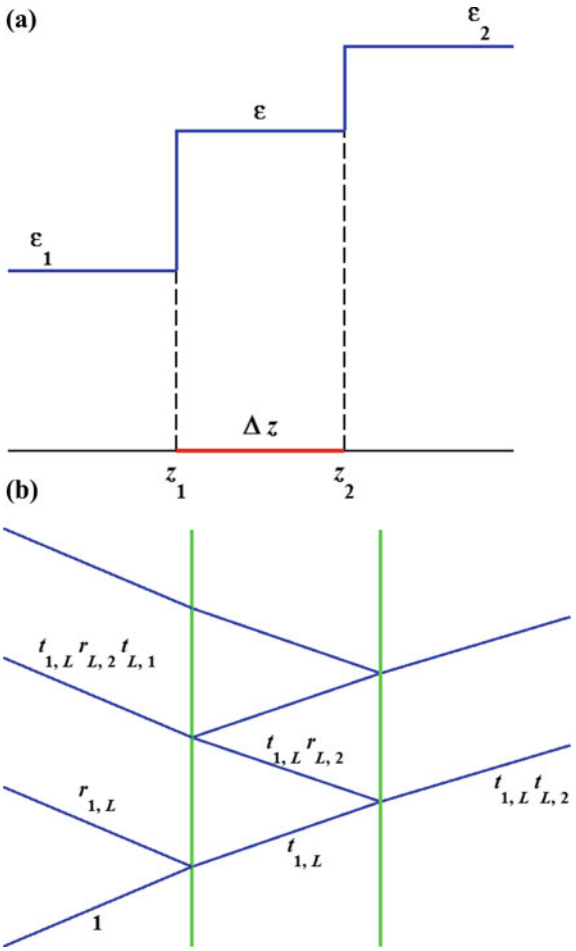


Fig. 2.5 **a** The dielectric function profile of a homogeneous layer. **b** Schematics of the multiple reflection method for calculation of the reflection and transmission amplitudes

$$t_s = \mathrm{e}^{i(q_1 z_1 - q_2 z_2)} \frac{2q_1 q}{q(q_1+q_2)c - i(q^2+q_1 q_2)s}, \tag{2.53}$$

where $c = \cos q\Delta z$ and $s = \sin q\Delta z$. Equivalently, we may substitute $F = \cos qz$, $G = \sin qz$ in (2.25) and (2.26); for these solutions of (2.22) we have $W = q,\ F_1 G_2 - G_1 F_2 = s,\ F_1 G_2' - G_1 F_2' = qc,\ F_1' G_2 - G_1' F_2 = -qc,\ F_1' G_2' - G_1' F_2' = q^2 s$.

It is instructive to consider another derivation of these results, using the multiple reflection method shown schematically in Fig. 2.5b. An incident wave of unit amplitude will produce a reflected wave of amplitude r_{1L} (this being the reflection amplitude at the step at z_1) and a transmitted wave of amplitude t_{1L} within the layer, this being the transmission amplitude at the step from medium 1 to the layer. This wave is in turn partly transmitted at z_2 (with amplitude $t_{1L}t_{L2}$), and partly reflected (amplitude $t_{1L}r_{L2}$). The reflected wave is then partly transmitted at z_1, giving a reflected wave amplitude $t_{1L}r_{L2}t_{L1}$, and partly reflected. The continuation of this process gives

$$r_s \;=\; r_{1L} + t_{1L}t_{L1}(r_{L2} + r_{L2}r_{L1}r_{L2} + \ldots) \;=\; r_{1L} + \frac{t_{1L}t_{L1}r_{L2}}{1 - r_{L1}r_{L2}}, \tag{2.54}$$

and

$$t_s = t_{1L}t_{L2}\Big(1 + r_{L2}r_{L1} + (r_{L2}r_{L1})^2 + \ldots\Big) = \frac{t_{1L}t_{L2}}{1 - r_{L1}r_{L2}}. \tag{2.55}$$

The various reflection and transmission amplitudes are for reflection at a single step, and can be found from (1.15), and the reciprocity relations (2.14) and (2.18):

$$r_{1L} = \mathrm{e}^{2iq_1 z_1}\frac{q_1 - q}{q_1 + q}, \quad r_{L1} = \mathrm{e}^{-2iqz_1}\frac{q - q_1}{q + q_1}, \quad r_{L2} = \mathrm{e}^{2iqz_2}\frac{q - q_2}{q + q_2}, \tag{2.56}$$

$$t_{1L} = \mathrm{e}^{i(q_1 - q)z_1}\frac{2q_1}{q_1 + q}, \quad t_{L1} = \mathrm{e}^{i(q_1 - q)z_1}\frac{2q}{q_1 + q}, \quad t_{L2} = \mathrm{e}^{i(q - q_2)z_2}\frac{2q}{q + q_2}. \tag{2.57}$$

If we write $r = (q_1 - q)/(q_1 + q)$ and $r' = (q - q_2)/(q + q_2)$, r_s and t_s reduce to

$$r_s = \mathrm{e}^{2iq_1 z_1}\frac{r + r'\mathrm{e}^{2iq\Delta z}}{1 + rr'\mathrm{e}^{2iq\Delta z}}, \tag{2.58}$$

$$t_s = \mathrm{e}^{i(q_1 z_1 - q_2 z_2)}\frac{(1+r)(1+r')\mathrm{e}^{iq\Delta z}}{1 + rr'\mathrm{e}^{2iq\Delta z}}, \tag{2.59}$$

and these are readily shown to be equivalent to (2.52) and (2.53).

We see from the above equations that $|r_s|^2$ and $|t_s|^2$ are periodic functions of the thickness Δz of the film, at given $\varepsilon_1, \varepsilon_2, \varepsilon$ and angle of incidence. The period in Δz is π/q, which increases from $\pi c/\omega\sqrt{\varepsilon}$ at normal incidence to $(\pi c/\omega)(\varepsilon - \varepsilon_1)^{-1/2}$ at grazing incidence. The phase factor $Z = e^{2iq\Delta z}$ moves on the unit circle in the complex plane as the thickness of the layer or the angle of incidence change, and r_s is a fractional or bilinear transform of Z, so it too moves on a circle in the complex plane (in the absence of absorption).

Zero reflection is possible if $r' = r$ and $e^{2iq\Delta z} = -1$ and also if $r' = -r$ and $e^{2iq\Delta z} = 1$. The first of these pairs of conditions holds if $q^2 = q_1 q_2$ and $2q\Delta z$ is an odd multiple of π. At normal incidence, these give the familiar characteristics of an antireflection coating:

$$\varepsilon^2 = \varepsilon_1 \varepsilon_2 \quad \text{and} \quad \Delta z = \lambda/4, 3\lambda/4, \ldots \tag{2.60}$$

The refractive index of the layer has to be the geometric mean of the refractive indices of the two outer media, and the thickness has to be equal to an odd multiple of a quarter wavelength (λ is the wavelength within the layer). At oblique incidence the condition $q^2 = q_1 q_2$ can be satisfied only if $\varepsilon^2 < \varepsilon_1 \varepsilon_2$; it then holds at

$$\theta_1 = \arcsin\left\{\left(\varepsilon_1 \varepsilon_2 - \varepsilon^2\right)/\varepsilon_1(\varepsilon_1 + \varepsilon_2 - 2\varepsilon)\right\}^{1/2}.$$

The second pair of conditions holds if $q_1 = q_2$ and $2q\Delta z$ is an even multiple of π. These are equivalent to

$$\varepsilon_1 = \varepsilon_2 \quad \textit{and} \quad \Delta z = \lambda/2, \lambda, \ldots, \tag{2.61}$$

where $\lambda = 2\pi/q$ is the wavelength corresponding to the motion normal to the interface. At normal incidence this happens when $\varepsilon^{1/2}(\omega/c)\Delta z = n\pi$ (n an integer). At oblique incidence a homogeneous film between like media does not reflect the s polarization at the angles of incidence

$$\theta_n = \arcsin\left\{\left(\varepsilon - \left(\frac{n\pi}{(\omega/c)\Delta z}\right)^2\right)/\varepsilon_1\right\}^{1/2}$$

An example of reflectivity at normal incidence as a function of layer thickness is shown in Fig. 2.6. R_s as a function of angle of incidence will be shown together with R_p in Fig. 2.7.

The reflectivity in Fig. 2.6 displays the periodicity in Δz of π/q, mentioned above. For the case shown, the maximum of R_n is the Fresnel (zero thickness) value

$$R_{n0} = \left(\frac{\sqrt{\varepsilon_1} - \sqrt{\varepsilon_2}}{\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2}}\right)^2 = \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2. \tag{2.62}$$

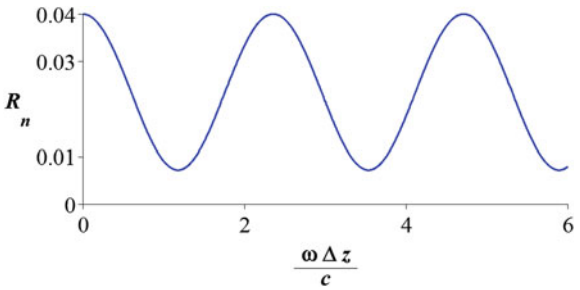


Fig. 2.6 Normal incidence reflectivity of a homogeneous layer as a function of layer thickness. R_n stands for the common value of R_s and R_p at normal incidence. The refractive index values used are $\sqrt{\varepsilon_1} = 1, \sqrt{\varepsilon} = 4/3$ and $\sqrt{\varepsilon_2} = 3/2$, representing a layer of water on glass

(As usual, $n = \sqrt{\varepsilon}$ denotes the refractive index.) In fact the homogeneous layer reflectivity is never greater than the Fresnel reflectivity R_{s0}, at any angle of incidence, provided ε lies between ε_1 and ε_2. This intuitively plausible result follows from the equivalence of

$$R_s \leq \left(\frac{q_1 - q_2}{q_1 + q_2}\right)^2 \tag{2.63}$$

to

$$q^4 + q_1^2 q_2^2 \leq q^2\left(q_1^2 + q_2^2\right), \tag{2.64}$$

which in turn is equivalent to

$$(\varepsilon - \varepsilon_1)(\varepsilon - \varepsilon_2) \leq 0. \tag{2.65}$$

Note that the corresponding result for the p wave reflectivity is not true at all angles: the single-step reflectivity is zero at the Brewster angle $\theta_1 = \theta_B = \arctan(\varepsilon_2/\varepsilon_1)^{1/2}$, at which angle the homogeneous layer or two-step reflectivity is not zero (in general). The s wave reflectivity may be written in the form

$$R_s = |r_s|^2 = \frac{r^2 + 2rr'\cos 2q\Delta z + (r')^2}{1 + 2rr'\cos 2q\Delta z + (rr')^2} \tag{2.66}$$

(provided r and r' are real; this requires the absence of absorption within the film, and $\theta_1 < \arcsin(\varepsilon/\varepsilon_1)^{1/2}, \theta_1 < \arcsin(\varepsilon_2/\varepsilon_1)^{1/2}$). It thus has extrema when $\sin 2q\Delta z = 0$ (when $\cos 2q\Delta z = \pm 1$). These extrema take the values

$$R_s^+ = \left(\frac{q_1 - q_2}{q_1 + q_2}\right)^2 = R_{s0}, \quad R_s^- = \left(\frac{q_1 q_2 - q^2}{q_1 q_2 + q^2}\right)^2. \tag{2.67}$$

Note that R_s^- is zero when $q^2 = q_1 q_2$, the antireflection coating condition. At normal incidence this reads $n^2 = n_1 n_2$: the refractive index of the layer equal to the geometric mean of the bounding indices.

R_s^- is less than R_s^+ provided (2.64) holds, that is when ε lies between ε_1 and ε_2. When ε is outside this range, R_s^+ becomes the minimum value, and R_s^- the maximum. They are equal when ε is equal to either ε_1 or ε_2, in which case the layer is non-existent as far as reflection is concerned.

The p wave reflection and transmission amplitudes are obtained by matching B and $\mathrm{d}B/\varepsilon \mathrm{d}z$ at z_1 and z_2; we find

$$-r_p = \mathrm{e}^{2iq_1 z_1} \frac{Q(Q_1 - Q_2)c + i(Q^2 - Q_1 Q_2)s}{Q(Q_1 + Q_2)c - i(Q^2 + Q_1 Q_2)s}, \tag{2.68}$$

$$\left(\frac{\varepsilon_2}{\varepsilon_1}\right)^{1/2} t_p = \mathrm{e}^{i(q_1 z_1 - q_2 z_2)} \frac{2Q_1 Q}{Q(Q_1 + Q_2)c - i(Q^2 + Q_1 Q_2)s}, \tag{2.69}$$

where $Q_i = q_i/\varepsilon_i$ and $Q = q/\varepsilon$, and $c = \cos q\Delta z, s = \sin q\Delta z$ as before. The multiple reflection method gives the alternative forms

$$-r_p = \mathrm{e}^{2iq_1 z_1} \frac{r + r'\mathrm{e}^{2iq\Delta z}}{1 + rr'\mathrm{e}^{2iq\Delta z}}, \tag{2.70}$$

$$\left(\frac{\varepsilon_2}{\varepsilon_1}\right)^{1/2} t_p = \mathrm{e}^{i(q_1 z_1 - q_2 z_2)} \frac{(1 + r)(1 + r')\mathrm{e}^{iq\Delta z}}{1 + rr'\mathrm{e}^{2iq\Delta z}}, \tag{2.71}$$

where now $r = (Q_1 - Q)/(Q_1 + Q), r' = (Q - Q_2)/(Q + Q_2)$. At normal incidence, $r_p = r_s$ and $t_p = t_s$.

Zero reflection occurs when $r' = r$ and $\mathrm{e}^{2iq\Delta z} = -1$, and also when $r' = -r$ and $\mathrm{e}^{2iq\Delta z} = 1$. The equality of r and r' holds if $Q^2 = Q_1 Q_2$, which at normal incidence is equivalent to $\varepsilon^2 = \varepsilon_1 \varepsilon_2$, as for s wave. The other possibility, $r' = -r$ and $\mathrm{e}^{2iq\Delta z} = 1$, holds if $Q_1 = Q_2$ and $q\Delta z$ is an integer multiple of π (the same condition can be read off from (2.68)). The equality of Q_1 and Q_2 is satisfied at all angles if $\varepsilon_1 = \varepsilon_2$, or at the Brewster angle $\theta_B = \arctan(\varepsilon_2/\varepsilon_1)^{1/2}$ for general values of the dielectric constants $\varepsilon_1, \varepsilon_2$. Zero reflection by a homogeneous layer between like media has the same condition (2.61) as for the s wave. Zero reflection by a homogeneous film between unlike media at the Brewster angle for vanishing thickness, $\theta_B = \arctan(\varepsilon_2/\varepsilon_1)^{1/2}$, will occur for thicknesses such that $q\Delta z$ is an integer times π. This gives, on using (1.37)

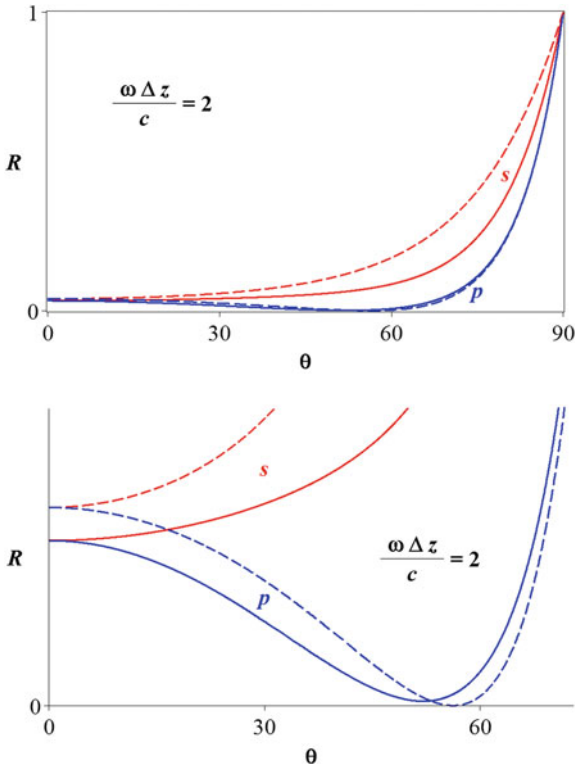


Fig. 2.7 Angular variation of the s and p reflectivities, for $\varepsilon_1 = 1, \varepsilon_2 = (3/2)^2, \varepsilon = (4/3)^2$ and $(\omega/c)\Delta z = 2$. These parameters approximate a layer of water on glass, about one third of a wavelength thick. The corresponding r_p/r_s curve is one of those displayed in Fig. 2.3. The dashed curves are for zero thickness of water (air|glass only)

$$\frac{\omega}{c}\Delta z = \frac{\text{integer} \times \pi}{(\varepsilon - \varepsilon_r)^{1/2}}, \quad \varepsilon_r = \frac{\varepsilon_1 \varepsilon_2}{\varepsilon_1 + \varepsilon_2}. \tag{2.72}$$

We see that zero reflection at θ_B is possible for non-zero thickness only if $\varepsilon > \varepsilon_r$ (which we can call the reduced dielectric constant of the bounding media, in analogy to the two-body reduced mass $m_r = m_1 m_2/(m_1 + m_2)$). Since ε_r is always smaller than either of $\varepsilon_1, \varepsilon_2$, this is not a strong constraint. The variation of R_p with angle is compared with that of R_s in Fig. 2.7.

$R_p = |r_p|^2$ has extrema when $2q\Delta z = \pm 1$; these take the same form as the s wave values (2.67):

$$R_p^+ = \left(\frac{Q_1 - Q_2}{Q_1 + Q_2}\right)^2 = R_{p0}, \quad R_p^- = \left(\frac{Q_1 Q_2 - Q^2}{Q_1 Q_2 + Q^2}\right)^2. \tag{2.73}$$

The extrema are zero when $Q_1 = Q_2$ [$\theta_1 = \arctan(\varepsilon_2/\varepsilon_1)^{1/2} = \theta_B$], and $Q^2 = Q_1 Q_2$, respectively. The extrema are equal (to R_{p0}) when $Q = Q_1$ or Q_2, that is when $\theta_1 = \arctan(\varepsilon/\varepsilon_1)^{1/2}$ or when $\theta_1 = \arcsin(\varepsilon\varepsilon_2/\varepsilon_1(\varepsilon + \varepsilon_2))^{1/2}$. At these two angles,

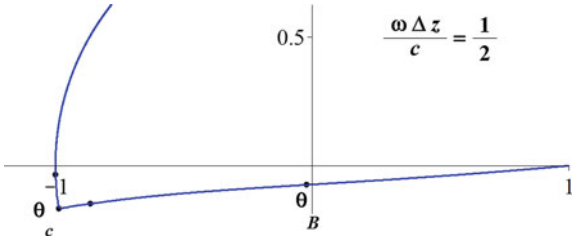


Fig. 2.8 Locus of r_p/r_s in the complex plane, for light incident from glass onto a layer of water bounded by air ($\sqrt{\varepsilon_1} = 3/2, \sqrt{\varepsilon} = 4/3, \varepsilon_2 = 1$). The thickness of the homogeneous water layer is such that $(\omega/c)\Delta z = 1/2$. Note the rapid variation with angle near $\theta_c = \arcsin 2/3 \approx 41.81°$, where 0.1 degree intervals are indicated (the points neighbouring θ_c are at 41.71°, 41.91°). The value of r_p/r_s at angle of incidence equal to the zero thickness Brewster angle $\theta_B = \arctan(\varepsilon_2/\varepsilon_1)^{1/2} = \arctan 2/3 \approx 33.69°$) is also shown

the reflectivity of the p wave is independent of the thickness of the layer. The separation of variables constant K^2 at these angles takes the values

$$K_1^2 = \left(\frac{\omega}{c}\right)^2 \frac{\varepsilon_1 \varepsilon}{\varepsilon_1 + \varepsilon}, \quad K_2^2 = \left(\frac{\omega}{c}\right)^2 \frac{\varepsilon \varepsilon_2}{\varepsilon + \varepsilon_2}, \tag{2.74}$$

appropriate to the Brewster angle values at the first or second interface, respectively.

When incidence is from the medium with higher dielectric constant, *total internal reflection* occurs for $\theta_1 > \theta_c = \arcsin(\varepsilon_2/\varepsilon_1)^{1/2}$. For $\theta_1 > \theta_c$ both r_p and r_s lie on the unit circle, and so does r_p/r_s. The path of r_p/r_s starts at $+1$ at normal incidence, as always. At the critical angle it makes a right-angle turn in the complex plane, moves out to an extremum on the unit circle, and then retraces its path back to its limiting value of r_p/r_s at grazing incidence, always -1. An example is shown in Fig. 2.8.

The reflection and transmission ellipsometry of a homogenous layer is discussed in Sect. 9.8. The phases of the reflection amplitudes are considered in detail in Appendix 20.1 of Chap. 20.

2.5 The Tanh, Exp and Rayleigh Profiles

It is possible to construct an infinite number of dielectric function profiles for which the reflection amplitude at normal incidence is known analytically. For a given function F, define $\varepsilon(\omega^2/c^2)$ as $-F''/F$ in the interval $z_1 \leq z \leq z_2$. A dielectric function so defined has the normal incidence reflection amplitude given by (2.25) and (2.30). Continuity of ε at z_1 or z_2 is not demanded. For example, $F = z^p$ gives

the profile $\varepsilon(\omega^2/c^2) = p(1-p)/z^2$; $F = \mathrm{e}^{ikz}$ gives $\varepsilon(\omega^2/c^2) = k^2$, the homogeneous layer dielectric function discussed in the last section.

The latter example can be applied to oblique incidence as well, by setting $-F''/F = \varepsilon(\omega^2/c^2) - K^2$. How can one construct other solutions which are valid at oblique incidence? This was answered by Heading (1965) in the electromagnetic case. The same question has been examined in quantum mechanics, as the problem of constructing solvable potentials for the Schrödinger equation, and in acoustics (construction of solvable velocity profiles) (Bose 1964; Deavenport 1966; Vasudevan et al. 1967). The method developed consists in transforming an equation whose solutions are known into the wave equation, and then stating the solutions of the wave equation in terms of the original equation. This systematic development has been extended to the electromagnetic p wave by Westcott (1969) (see also Heading 1970).

We are most interested in profiles which are solvable for both the s and p waves, to which we will turn shortly. But first we give one example of a profile solvable for the s wave, which is included in the systematic development, but predates it by more than 30 years. This is the useful *hyperbolic tangent profile* (Eckart 1930; Epstein 1930; Landau and Lifshitz 1965, Sect. 25),

$$\begin{aligned}\varepsilon(z) &= \frac{1}{2}(\varepsilon_1+\varepsilon_2) - \frac{1}{2}(\varepsilon_1-\varepsilon_2)\tanh\frac{z}{2a} \\ &= \frac{\varepsilon_1+\varepsilon_2\mathrm{e}^{z/a}}{1+\mathrm{e}^{z/a}} = \frac{\varepsilon_1}{1+\mathrm{e}^{z/a}} + \frac{\varepsilon_2}{1+\mathrm{e}^{-z/a}}.\end{aligned} \tag{2.75}$$

Figure 2.9 shows this profile, together with the 10–90 thickness $\Delta z = (2\ln 9)a$, which extends from the point z_1 where $\varepsilon = \varepsilon_1 + \frac{\Delta\varepsilon}{10}$ to the point z_2 where

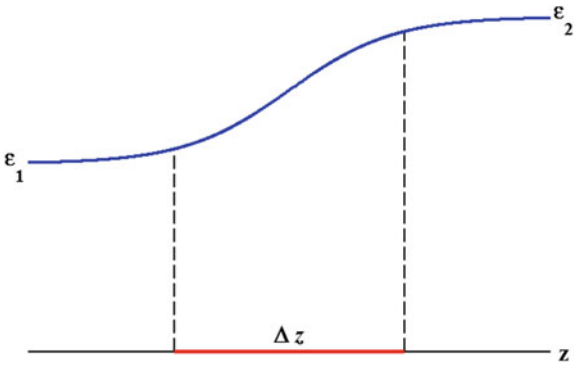


Fig. 2.9 The hyperbolic tangent dielectric function profile (2.75), also known as the Fermi profile. Here $\varepsilon_1 = 1$ and $\varepsilon_2 = (4/3)^2$, representing the air-water interface at optical frequencies. The vertical lines and the thickness Δz refer to the 10–90 thickness described in the text

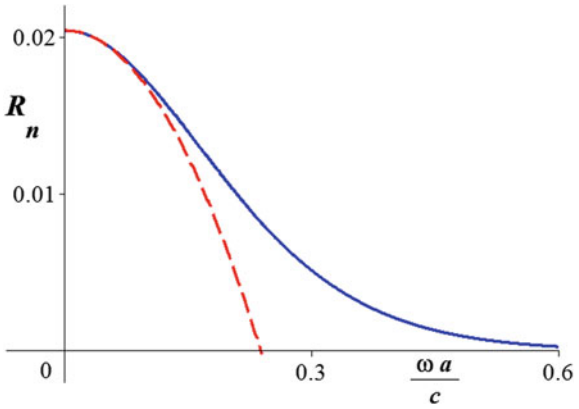


Fig. 2.10 The reflectivity at normal incidence, $R_n = [\sinh \pi a(k_1 - k_2)/\sinh \pi a(k_1 + k_2)]^2$, for the hyperbolic tangent profile, as a function of interface thickness. The dielectric function values are as in Fig. 2.9. The dashed curve is the long-wave expression given in the text

$\varepsilon = \varepsilon_1 + \frac{9\Delta\varepsilon}{10}$. (A comparison of the various measures of surface thickness may be found in Lekner and Henderson 1978.)

Figure 2.10 shows the corresponding reflectivity at normal incidence, to be derived shortly. Also shown are the leading terms in the long-wave expansion: $R_n = R_{n0}[1 - (4\pi^2/3)a^2k_1k_2 + \ldots]$. This is an example of a general result to be derived in Chap. 3. The equation (3.51) reduces, at normal incidence, to $R_n = \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2 - \frac{4n_1n_2}{(n_1+n_2)^4}\left(\frac{\omega}{c}\right)^2 i_2 + \ldots$, where i_2 is a profile integral invariant, and $n_i = \sqrt{\varepsilon_i}$ as usual. The integral invariant $i_2 = \frac{\pi^2}{3}(\varepsilon_1 - \varepsilon_2)^2a^2$ for the *tanh* profile, from Table 3.1.

The s wave equation for the hyperbolic tangent profile (2.75),

$$\frac{\mathrm{d}^2E}{\mathrm{d}z^2} + \left(\varepsilon\frac{\omega^2}{c^2} - K^2\right)E = 0, \tag{2.76}$$

can be transformed to the hypergeometric differential equation by the substitutions

$$\zeta = -\mathrm{e}^{-z/a},\ E = \zeta^{-iq_2a}w(\zeta), \tag{2.77}$$

where $w(\zeta)$ tends to a constant as $\zeta \to 0$ $(z \to \infty)$. The function w satisfies

$$\zeta(1-\zeta)\frac{\mathrm{d}^2w}{\mathrm{d}\zeta^2} + (1 - 2iy_2)(1-\zeta)\frac{\mathrm{d}w}{\mathrm{d}\zeta} + \left(y_2^2 - y_1^2\right)w = 0, \tag{2.78}$$

where $y_1 = q_1 a$ and $y_2 = q_2 a$. (The references quoted above give solutions for normal incidence; the generalization to oblique incidence is given here.) The hypergeometric function

$$F(\alpha, \beta; \gamma; \zeta) = 1 + \frac{\alpha\beta}{\gamma}\frac{\zeta}{1!} + \frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1)}\frac{\zeta^2}{2!} + \ldots \tag{2.79}$$

satisfies the equation

$$\zeta(1-\zeta)\frac{\mathrm{d}^2 F}{\mathrm{d}\zeta^2} + [\gamma - (\alpha+\beta+1)\zeta]\frac{\mathrm{d}F}{\mathrm{d}\zeta} - \alpha\beta F = 0, \tag{2.80}$$

so that $w(\zeta)$ is equal to $F(i(y_1 - y_2), -i(y_1 + y_2); 1 - 2iy_2; \zeta)$. To extract the reflection amplitude we need the limiting form as $z \to -\infty$, i.e., as $\zeta \to -\infty$. This is obtained from the formula (Oberhettinger 1964, 15.3.7)

$$\begin{aligned} F(\alpha, \beta; \gamma; \zeta) = &\frac{\Gamma(\gamma)\Gamma(\beta-\alpha)}{\Gamma(\beta)\Gamma(\gamma-\alpha)}(-\zeta)^{-\alpha} F\left(\alpha, 1+\alpha-\gamma; 1+\alpha-\beta; \frac{1}{\zeta}\right) \\ &+ \frac{\Gamma(\gamma)\Gamma(\alpha-\beta)}{\Gamma(\alpha)\Gamma(\gamma-\beta)}(-\zeta)^{-\beta} F\left(\beta, 1+\beta-\gamma; 1+\beta-\alpha; \frac{1}{\zeta}\right), \end{aligned} \tag{2.81}$$

valid for $|\arg(-\zeta)| < \pi$. As $\zeta \to -\infty$, the leading terms in (2.81), on using the expansion (2.79), give the limiting form

$$(-)^{-iq_2 a}\Gamma(1-2iy_2)\left\{\frac{\Gamma(-2iy_1)\mathrm{e}^{iq_1 z}}{\Gamma(-i(y_1+y_2))\Gamma(1-i(y_1+y_2))} + \frac{\Gamma(2iy_1)\mathrm{e}^{-iq_1 z}}{\Gamma(i(y_1-y_2))\Gamma(1+i(y_1-y_2))}\right\} \leftarrow E. \tag{2.82}$$

The reflection amplitude is defined as the ratio of the coefficient of $\mathrm{e}^{-iq_1 z}$ to that of $\mathrm{e}^{iq_1 z}$. On using the formula

$$\Gamma(z)\Gamma(1-z) = \pi/\sin \pi z, \tag{2.83}$$

we find

$$r_s = -\frac{\Gamma(2iy_1)\Gamma(-i(y_1+y_2))\Gamma(-i(y_1-y_2))\sinh \pi(y_1-y_2)}{\Gamma(-2iy_1)\Gamma(i(y_1+y_2))\Gamma(i(y_1-y_2))\sinh \pi(y_1+y_2)}. \tag{2.84}$$

Ratios of the form $\Gamma(-iy)/\Gamma(iy)$ can be evaluated by using the infinite product representation of the gamma function (Whittaker and Watson (1927), Sect. 12.1)

$$\frac{1}{\Gamma(z)} = z\,\mathrm{e}^{\gamma z}\prod_{n=1}^{\infty}\left(1+\frac{z}{n}\right)\mathrm{e}^{-z/n}. \tag{2.85}$$

Here γ is Euler's constant, $\gamma \approx 0.5772$. We find

$$\frac{\Gamma(-iy)}{\Gamma(iy)} = -\exp 2i\{\gamma y - \phi(y)\}, \tag{2.86}$$

where

$$\phi(y) = \sum_{n=1}^{\infty} \left(\frac{y}{n} - \arctan\frac{y}{n}\right). \tag{2.87}$$

Thus

$$r_s = \exp 2i\{\phi(2y_1) - \phi(y_1+y_2) - \phi(y_1-y_2)\}\frac{\sinh \pi(y_1-y_2)}{\sinh \pi(y_1+y_2)}. \tag{2.88}$$

The combination of ϕ functions within the braces simplifies to

$$\sum_{n=1}^{\infty} \arctan\left\{\frac{2y_1}{n}\cdot\frac{y_1^2-y_2^2}{n^2+3y_1^2+y_2^2}\right\}. \tag{2.89}$$

In this form it is clear that the phase of r_s is third order in the interface thickness when the profile is centred on the origin, and also that $|r_s| = 1$ when $q_2 = i|q_2|$ (total internal reflection).

We have given some detail for this model profile, since it is frequently used and has the virtue that the reflection amplitude, complete with phase, is expressible in terms of elementary functions. Another interesting feature is the relationship between *reflection at oblique incidence to that at normal incidence*. The solution at oblique incidence is obtained from that at normal incidence by replacing $k_i = \sqrt{\varepsilon_i}(\omega/c)$ by $q_i = k_i \cos\theta_i$ in the formulae above. This is a general property of dielectric function (or potential energy) profiles of the form

$$\varepsilon(z) = \frac{1}{2}(\varepsilon_1+\varepsilon_2) - \frac{1}{2}(\varepsilon_1-\varepsilon_2)f(z,a), \tag{2.90}$$

where the function f depends on parameters (such as the length a characterizing the interface thickness) which are independent of ε_1 and ε_2. When this holds, $q^2 = \varepsilon\omega^2/c^2 - K^2$ may be written as

$$q^2(z) = \frac{1}{2}(q_1^2+q_2^2) - \frac{1}{2}(q_1^2-q_2^2)f(z,a), \tag{2.91}$$

and information concerning the dielectric constants ε_1 and ε_2, the vacuum wavenumber ω/c, and the angle of incidence is contained within the normal components of the wavevector, q_1 and q_2. From (2.90), the function f is given by

$$f = \frac{\varepsilon_1 + \varepsilon_2 - 2\varepsilon}{\varepsilon_1 - \varepsilon_2}. \tag{2.92}$$

Any ε will give a function f, but only profiles which can be put in the form (2.90) will have f independent of ε_1 and ε_2. We shall shortly see examples of profiles which do not have this scaling property, and for which the reflection amplitude at oblique incidence cannot be obtained from the formula for normal incidence.

We now turn to profiles for which both s and p wave solutions may be obtained analytically, and concentrate on two continuous dielectric function profiles of finite range: an exponential variation with z of the refractive index or dielectric function, and a linear variation with z of the reciprocal of the refractive index. The *exponential profile* was considered by Galejs (1961), Burman and Gould (1963), and Abelès (1964). The dielectric function is given by

$$\varepsilon(z) = \begin{cases} \varepsilon_1 & z < z_1 \\ (\varepsilon_1 \varepsilon_2)^{1/2} \exp\left\{ \frac{(z-\bar{z})}{\Delta z} \ln \frac{\varepsilon_2}{\varepsilon_1} \right\} & z_1 \le z \le z_2 \\ \varepsilon_2 & z > z_2 \end{cases} \tag{2.93}$$

where $\bar{z} = (z_1 + z_2)/2$ and $\Delta z = z_2 - z_1$. A simpler but less symmetric form for ε is $\varepsilon_1 \exp[(z - z_1)/a]$, where $a = \Delta z / \ln(\varepsilon_2/\varepsilon_1)$. Figure 2.12 shows an exponential profile representing the air-water interface at optical frequencies.

A change from z to a dimensionless independent variable proportional to the refractive index,

$$u = 2a\frac{\omega}{c}\sqrt{\varepsilon} \equiv 2ka, \quad a = \Delta z / \ln(\varepsilon_2/\varepsilon_1) \tag{2.94}$$

transforms the s wave equation into Bessel's equation

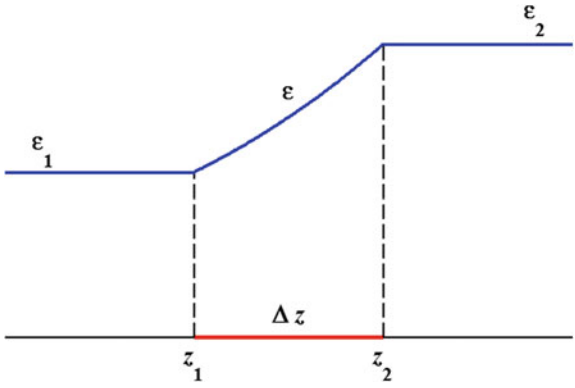


Fig. 2.11 The exponential function profile, (2.93), with $\sqrt{\varepsilon_1} = 1, \sqrt{\varepsilon_2} = 4/3$

$$\frac{\mathrm{d}^2E}{\mathrm{d}u^2}+\frac{1}{u}\frac{\mathrm{d}E}{\mathrm{d}u}+\left(1-\frac{(2Ka)^2}{u^2}\right)E=0. \tag{2.95}$$

The general solution within $[z_1,z_2]$ is $\alpha J_s(u)+\beta Y_s(u)$, with $s=2Ka$. The order s of the Bessel functions depends on the angle of incidence (it is proportional to $\sin\theta_1$); both s and u are proportional to the interface thickness. The s wave reflection and transmission amplitudes may be obtained from (2.25) and (2.26), with $F(z)=J_s(u)$ and $G(z)=Y_s(u)$.

The p wave equation reads, in the u variable,

$$\frac{\mathrm{d}^2B}{\mathrm{d}u^2}-\frac{1}{u}\frac{\mathrm{d}B}{\mathrm{d}u}+\left(1-\frac{(2Ka)^2}{u^2}\right)B=0, \tag{2.96}$$

and is satisfied by $\alpha uJ_p+\beta uY_p(u)$, where $p^2=(2Ka)^2+1$. The reflection amplitude may be found from (2.40), with $C(z)=uJ_p(u)$ and $D(z)=uY_p(u)$. It is useful to work in terms of the cross products

$$\begin{aligned}
A_v&=J_v(u_1)Y_v(u_2)-Y_v(u_1)J_v(u_2),\\
B_v&=J_v(u_1)Y_v'(u_2)-Y_v(u_1)J_v'(u_2),\\
C_v&=J_v'(u_1)Y_v(u_2)-Y_v'(u_1)J_v(u_2),\\
D_v&=J_v'(u_1)Y_v'(u_2)-Y_v'(u_1)J_v'(u_2),
\end{aligned} \tag{2.97}$$

where the primes denote differentiation with respect to u. The p wave reflection amplitude then reads

$$-r_p=\mathrm{e}^{2iq_1z_1}\frac{q_1q_2A_p+iq_1k_2\left(B_p+\frac{A_p}{u_2}\right)+iq_2k_1\left(C_p+\frac{A_p}{u_1}\right)-k_1k_2\left(D_p+\frac{B_p}{u_1}+\frac{C_p}{u_2}+\frac{A_p}{u_1u_2}\right)}{q_1q_2A_p+iq_1k_2\left(B_p+\frac{A_p}{u_2}\right)-iq_2k_1\left(C_p+\frac{A_p}{u_1}\right)+k_1k_2\left(D_p+\frac{B_p}{u_1}+\frac{C_p}{u_2}+\frac{A_p}{u_1u_2}\right)}. \tag{2.98}$$

The s wave result is

$$r_s=\mathrm{e}^{2iq_1z_1}\frac{q_1q_2A_s+iq_1k_2B_s+iq_2k_1C_s-k_1k_2D_s}{q_1q_2A_s+iq_1k_2B_s-iq_2k_1C_s+k_1k_2D_s}. \tag{2.99}$$

At normal incidence $s=0$ and $p=1$. On using the identities (compare Olver 1964, 9.1.32, 33)

$$A_1=D_0,\quad B_1+\frac{A_1}{u_2}=-C_0,\quad C_1+\frac{A_1}{u_1}=-B_0,$$

$$D_1 + \frac{B_1}{u_1} + \frac{C_1}{u_2} + \frac{A_1}{u_1 u_2} = A_0, \tag{2.100}$$

we find that both the reflection amplitudes at normal incidence reduce to

$$r_n = \mathrm{e}^{2ik_1 z_1} \frac{A_0 + iB_0 + iC_0 - D_0}{A_0 + iB_0 - iC_0 + D_0}. \tag{2.101}$$

The corresponding reflectivity is

$$R_n = |r_n|^2 = \frac{A_0^2 + B_0^2 + C_0^2 + D_0^2 - \frac{8}{\pi^2 u_1 u_2}}{A_0^2 + B_0^2 + C_0^2 + D_0^2 + \frac{8}{\pi^2 u_1 u_2}}. \tag{2.102}$$

In obtaining (2.102) we have used the identity (2.49), and the fact that the Wronskian $J_\nu(u)Y'_\nu(u) - J'_\nu(u)Y_\nu(u)$ is equal to $2/\pi u$. The reflectivity at normal incidence as a function of interfacial thickness is shown in Fig. 2.12, together with the long-wave expression as discussed in relation to Fig. 2.10.

The second dielectric function profile for which a solution is known for-both the *s* and *p* waves was first considered by Rayleigh (1880) (for normal incidence only), and a solution for general incidence of both polarizations was given by Burman and Gould (1963). For the *Rayleigh profile* the reciprocal of the refractive index varies linearly with distance between the interfacial boundaries z_1 and z_2; as usual we have $\varepsilon = \varepsilon_1$ for $z < z_2$ and $\varepsilon = \varepsilon_2$ for $z > z_2$. This profile is shown in Fig. 2.13. (The Rayleigh profile with discontinuities at its boundaries is considered in Lekner 1990c.)

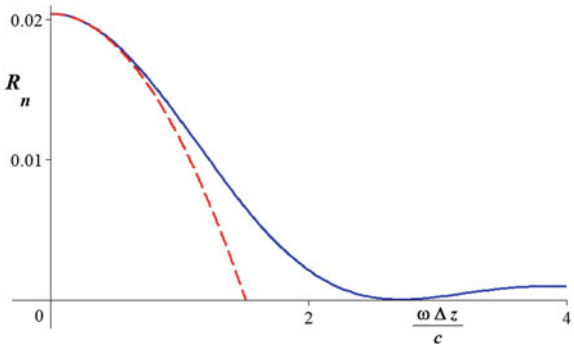


Fig. 2.12 Normal incidence reflectivity for the exponential profile, as a function of the interface thickness. The values of ε_1 and ε_2 are as in Fig. 2.11. The first minimum is at $(\omega/c)\Delta z \approx 2.71$. For the same ε_1 and ε_2 the similar Rayleigh profile (Figs. 2.13 and 2.14) has its first zero at $(\omega/c)\Delta z \approx 2.73$. The *dashed curve* gives the long-wave expression to second order in the interface thickness

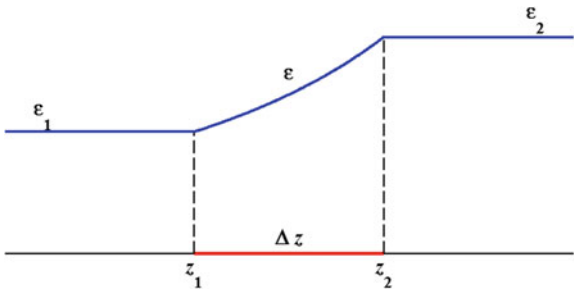


Fig. 2.13 Dielectric function $\varepsilon(z)$ for the Rayleigh profile. The values $\varepsilon_1 = 1$ and $\varepsilon_2 = (4/3)^2$ approximate air on the *left* and water on the *right* (at optical frequencies)

Since $\varepsilon^{-1/2}$ varies linearly with z, it will be useful to work in terms of this function, which we will call η:

$$\varepsilon^{-\frac{1}{2}}(z) = \eta(z) = \bar{\eta} + (z - \bar{z})\frac{\Delta\eta}{\Delta z}, \tag{2.103}$$

where $\Delta\eta = \eta_2 - \eta_1 = \varepsilon_2^{-\frac{1}{2}} - \varepsilon_1^{-\frac{1}{2}}$, $\Delta z = z_2 - z_1$, $\bar{\eta} = (\eta_1 + \eta_2)/2$, and $\bar{z} = (z_1 + z_2)/2$.

At normal incidence the s wave equation becomes, on changing the independent variable from z to η,

$$\frac{\mathrm{d}^2 E}{\mathrm{d}\eta^2} + \left(\frac{1}{4} - \nu^2\right)\frac{E}{\eta^2} = 0, \tag{2.104}$$

where

$$\nu^2 = \frac{1}{4} - \left(\frac{\omega}{c}\frac{\Delta z}{\Delta\eta}\right)^2. \tag{2.105}$$

The equation (2.104) has the power-law solutions $E_\pm = \eta^{\frac{1}{2}\pm\nu}$, and the reflection amplitude can be found from (2.25):

$$r_n = \mathrm{e}^{2ik_1 z_1}\frac{\left(\frac{\varepsilon_1}{\varepsilon_2}\right)^\nu - 1}{2\nu\left[\left(\frac{\varepsilon_1}{\varepsilon_2}\right)^\nu + 1\right] - 2i\frac{\omega\Delta z}{c\Delta\eta}\left[\left(\frac{\varepsilon_1}{\varepsilon_2}\right)^\nu - 1\right]}. \tag{2.106}$$

The reflectivity at normal incidence takes different forms according as $\left(\frac{\omega}{c}\frac{\Delta z}{\Delta\eta}\right)^2$ is smaller or greater than $\frac{1}{4}$. When $\left(\frac{\omega}{c}\frac{\Delta z}{\Delta\eta}\right)^2 < \frac{1}{4}$, ν is real and

$$R_n = \frac{\left[\left(\frac{\varepsilon_1}{\varepsilon_2}\right)^{\nu} - 1\right]^2}{4\nu^2\left[\left(\frac{\varepsilon_1}{\varepsilon_2}\right)^{\nu} + 1\right]^2 + 4\left(\frac{\omega\Delta z}{c\Delta\eta}\right)^2\left[\left(\frac{\varepsilon_1}{\varepsilon_2}\right)^{\nu} - 1\right]^2} = \frac{\left[\left(\frac{\varepsilon_1}{\varepsilon_2}\right)^{\nu} - 1\right]^2}{\left[\left(\frac{\varepsilon_1}{\varepsilon_2}\right)^{\nu} + 1\right]^2 - 16\left(\frac{\omega\Delta z}{c\Delta\eta}\right)^2\left(\frac{\varepsilon_1}{\varepsilon_2}\right)^{\nu}}. \tag{2.107}$$

For $\left(\frac{\omega}{c}\frac{\Delta z}{\Delta\eta}\right)^2 > \frac{1}{4}$, $\nu = i|\nu|$, and

$$R_n = \frac{\sin^2\left(\frac{1}{2}|\nu|\ln\frac{\varepsilon_1}{\varepsilon_2}\right)}{4|\nu|^2 + \sin^2\left(\frac{1}{2}|\nu|\ln\frac{\varepsilon_1}{\varepsilon_2}\right)}. \tag{2.108}$$

At $\nu = 0$ these two forms take the common value

$$R_n(\nu = 0) = \frac{\left(\ln\frac{\varepsilon_1}{\varepsilon_2}\right)^2}{16 + \left(\ln\frac{\varepsilon_1}{\varepsilon_2}\right)^2}. \tag{2.109}$$

We note from (2.108) that the reflectivity is *zero* whenever $|\nu||\ln(\varepsilon_1/\varepsilon_2)| = 2n\pi$ $(n = 1, 2, \ldots)$, that is when

$$\frac{\omega}{c}\frac{\Delta z}{\Delta\eta} = \pm\left\{\frac{1}{4} + \left(\frac{2n\pi}{\ln\left(\frac{\varepsilon_1}{\varepsilon_2}\right)}\right)^2\right\}^{1/2}. \tag{2.110}$$

The reflectivity at normal incidence as a function of interface thickness is shown in Fig. 2.14.

The exponential profile, which also has discontinuities in slope at its boundaries and has a similar shape, has minima at points approximated by (2.110) (see the caption to Fig. 2.12). The reflectivities for both profiles show oscillatory decay with increasing thickness. The homogeneous layer discussed in the last section, which has discontinuities in value at the boundaries, has its reflectivity strictly periodic in the thickness. In contrast, the hyperbolic tangent profile, which is continuous in value and in all its derivatives, shows a monotonic decrease in reflectivity with interface thickness.

For oblique incidence, the s wave equation (2.22) for the Rayleigh profile becomes, in the η variable,

$$\frac{\mathrm{d}^2E}{\mathrm{d}\eta^2} + \left[\frac{\frac{1}{4} - \nu^2}{\eta^2} - \left(K\frac{\Delta z}{\Delta\eta}\right)^2\right]E = 0, \tag{2.111}$$

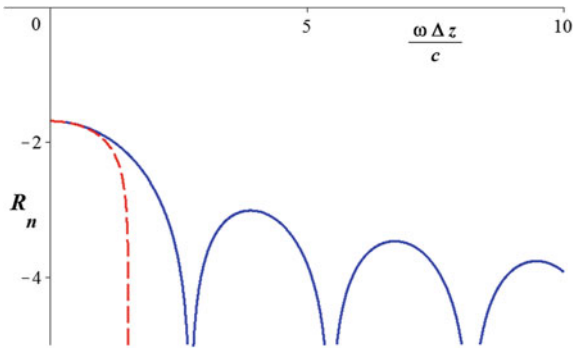


Fig. 2.14 Logarithmic plot of normal incidence reflectivity for the Rayleigh profile, as a function of the interface thickness, together with the long-wave expression (*dashed curve*). The values $\varepsilon_1 = 1, \varepsilon_2 = (4/3)^2$ have been used, representing an air-water interface, in common with Fig. (2.41). The scale is logarithmic to base 10; for example, -2 corresponds to a reflectivity of 10^{-2}. The logarithmic scale emphasizes the zeros in reflectivity, given by (2.110)

and has solutions proportional to $\eta^{1/2}$ times a Bessel function of order ν and imaginary argument $\pm iK(\Delta z/\Delta\eta)\eta$. Thus r_s may be obtained from the general formula given in Sect. 2.2.

The p wave is most conveniently discussed in terms of the variable $b = (\varepsilon_1/\varepsilon)^{1/2}B$, which satisfies (1.22):

$$\frac{\mathrm{d}^2 b}{\mathrm{d}z^2} + \left[q^2 - \varepsilon^{\frac{1}{2}}\frac{\mathrm{d}^2\varepsilon^{-1/2}}{\mathrm{d}z^2}\right]b = 0. \tag{2.112}$$

Since $\varepsilon^{-1/2}$ is linear for z for the Rayleigh profile, the E and b equations are the same, except at the end-points z_1 and z_2. There, because of the discontinuity in the slope of $\varepsilon^{-1/2}$, the equation for b contains additional delta-function terms:

$$\varepsilon^{1/2}\frac{\mathrm{d}^2\varepsilon^{-1/2}}{\mathrm{d}z^2} = \frac{\Delta\eta}{\Delta z}\left\{\frac{1}{\eta_1}\delta(z - z_1) - \frac{1}{\eta_2}\delta(z - z_2)\right\}. \tag{2.113}$$

As a consequence, $\mathrm{d}b/\mathrm{d}z$ is discontinuous at z_1 and z_2. Within the interface, b has the same Bessel function solutions as the s wave. Expressions for r_s and r_p, graphs of r_s, r_p and of r_p/r_s, and a comparison with theory for the reflection of long waves (to be discussed in the next chapter), are given in Lekner (1982). A generalization of the theory of Sect. 2.2 to discontinuous profiles is given in Lekner (1990c).

We have concentrated on general results, and on the reflection by four special profiles, three of which are solvable for both the s and p waves. Discussion of other special profiles (which are solvable for the s wave only) may be found in Sects. 4.3 and 5.2, in Heading (1965), and in the texts listed in the references for this section.

Other exact and general results appear throughout this book. As regards reflection by special profiles, we mention in particular the sech^2 potential (Sect. 4.3), and the linear profile (Sect. 5.2).

References

Abelès F (1964) Optical properties of inhomogeneous films. NBS Misc Publ 265:41–58
Beaglehole D (1980) Ellipsometric study of the surface of simple liquids. Physica 100B:163–174
Birkhoff G, Rota G-C (1969) Ordinary differential equations, Blaisdell. Waltham, Massachusetts
Bose AK (1964) A class of solvable potentials. Nuovo Cimento 32:679–688
Burman R, Gould RN (1963) On the propagation of vertically polarized electromagnetic waves in a horizontally stratified medium. J Atm Terr Phys 25:543–549
Deavenport RL (1966) A normal mode theory of an underwater acoustic duct by means of Green's function. Radio Sc 1:709–724
Eckart C (1930) The penetration of a potential barrier by electrons. Phys Rev 35:1303–1309
Epstein PS (1930) Reflection of waves in an inhomogeneous absorbing medium. Proc Nat Acad Sci 16:627–637
Galejs JG (1961) ELF waves in the presence of exponential ionospheric conductivity profiles. IRE Trans Ant Prop 9:554–562
Heading J (1965) Refractive index profiles based on the hypergeometric equation and the confluent hypergeometric equation. Proc Camb Phil Soc 61:897–913
Heading J (1970) The equality of the moduli of certain ratios occurring in the connexion formulae of solutions of some transcendental differential equations. Proc Camb Phil Soc 67:347–361
Jasperson SN, Schnatterly SE (1969) An improved method for high reflectivity ellipsometry based on a new polarization modulation technique. Rev Sci Instr 40:761–767
Jenkins FA, White HE (1950) Fundamentals of optics. McGraw-HilI, New York. Sections 13.8 and 28.10
Landau LD, Lifshitz EM (1965) Quantum mechanics. Pergamon, Oxford
Lekner J, Henderson JR (1978) Theoretical determination of the thickness of the liquid-vapour interface. Physica 94A:545–558
Lekner J (1982a) Exact reflection amplitudes for the Rayleigh profile. Physica 116A:235–247
Lekner J (1985) Reflection at oblique incidence and the existence of a Brewster angle. J Opt Soc Am A: 2:186–188
Lekner J (1990a) Non-reflecting stratifications. Can J Phys 68:738–742
Lekner J (1990b) Reflection of waves by a profile with discontinuities. J Phys A: Math Gen 23:2897–2904
Oberhettinger F (1964) Hypergeometric functions. In: Abramowitz M, Stegun IA (eds) Chapter 15 of Handbook of mathematical functions. NBS Applied Mathematics Series No. 55
Olver FWJ (1964) Bessel functions of integer order. In: Abramowitz M, Stegun IA (eds) Chapter 9 of Handbook of mathematical functions. NBS Applied Mathematics Series No. 55
Rayleigh JWS (1880) On the reflection of vibrations at the confines of two media between which the transition is gradual. Proc Lond Math Soc 11:51–56
Stokes GG (1849) On the perfect blackness of the central spot in Newton's rings, and on the verification of Fresnel's formulae for the intensities of the reflected and refracted rays. Camb Dublin Math J 4:1–15
Vasudevan R, Venkatesan K, Jagannathan G (1967) Construction of solvable potentials and some aspects of regularization of singular potentials. Nuovo Cimento 5(suppl.):621–643

Westcott BS (1969) Exact solutions for vertically polarized electromagnetic waves in horizontally stratified media. Proc Camb Phil Soc 66:675–684
Whittaker ET, Watson GN (1927) A course of modem analysis. Cambridge University Press, Cambridge

Further Readings

The derivation of conservation laws and reciprocity relations from comparison identities (Section 2.1) is based on

Lekner J (1982b) Reflection of long waves by interfaces. Physica 112A:544–556

For alternative derivations, see Landau and Lifshitz (1965), Section 25 and

Heading J (1975) Ordinary differential equations, theory and practice. Elek Science, Londin. Chapter 4

Extension of Stokes' idea of reversing the wavemotions is discussed by Šantavý and Knittl

Knittl Z (1962) The principle of reversibility and thin film optics. Optica Acta 9:33–45
Knittl Z (1976) Optics of thin films. Wiley, London. Chapter 6
Šantavý U (1961) On the reversibility of light beams in conducting media. Optica Acta 8:301–307

Sections 2.2 and 2.3 are based on the author's paper on the Rayleigh profile (quoted above) and on Lekner (1985).The ellipsometry of a homogenous layer between two bulk media (Section 2.4) is considered more detail in

Dorf MC, Lekner J (1987) Reflection and transmission ellipsometry of a uniform layer. J Opt Soc Am 4:2096–2100
Lekner J (1994) Inversion of reflection ellipsometric data. Appl Opt 33:5159–5165

Exactly solvable profiles (Section 2.5) are listed in the texts by heading (listed above) and

Brekhovskikh LM (1980) Waves in layered media, 2nd edn. Academic Press, New York
Budden KG (1961) Radio waves in the ionosphere. Cambridge University Press, Cambridge
Ginzburg VL (1964) The propagation of electromagnetic waves in plasmas. Pergamon, Oxford
Wait JR (1970) Electromagnetic waves in stratified media, 2nd edn. Pergamon, New York

The phenomenon of zero reflection is discussed in

Lekner J (1990) The phase relation between reflected and transmitted waves, and some consequences. Am J Phys 58:317–320, and Lekner (1990a)

Chapter 3
Reflection of Long Waves

We have seen in Sect. 2.2 that the reflection amplitudes of an arbitrary profile tend to the Fresnel values as the thickness Δz of the profile tends to zero. An equivalent limit to consider is that of reflection by a profile of fixed extent, as the wavelength increases. We might expect the reflection amplitudes to be well represented, in the long wave limit, by the first few terms of a series in the ratio of the interface thickness to the wavelength. This expectation turns out to be essentially correct, with the coefficient of a given power of $(\omega/c)\Delta z$ depending on the angle of incidence, as well as on the profile characteristics.

In the long wave limit, a given dielectric function profile reflects as a step profile, plus a small correction which depends on the deviation of the profile from a single step. The long wave theory treats the deviation as a *perturbation;* the perturbation theory of reflection is developed in Sect. 3.1 (for any type of perturbation), and then used to obtain the long wave expansion in the following sections.

3.1 Integral Equation and Perturbation Theory for the *s* Wave

The results of this section hold for the electromagnetic *s* wave, and for Schrödinger particle waves. We wish to express ψ, the solution of

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}z^2} + q^2\psi = 0, \quad \mathrm{e}^{iq_1 z} + r\,\mathrm{e}^{-iq_1 z} \leftarrow \psi \rightarrow t\,\mathrm{e}^{iq_2 z} \tag{3.1}$$

in terms of a known function ψ_0, the solution of

$$\frac{\mathrm{d}^2\psi_0}{\mathrm{d}z^2} + q_0^2\psi_0 = 0, \quad \mathrm{e}^{iq_1 z} + r_0\,\mathrm{e}^{-iq_1 z} \leftarrow \psi_0 \rightarrow t_0\,\mathrm{e}^{iq_2 z} \tag{3.2}$$

J. Lekner, *Theory of Reflection*, Springer Series on Atomic, Optical, and Plasma Physics 87, DOI 10.1007/978-3-319-23627-8_3

Note that q and q_0 share the same asymptotic values, q_1 and q_2: the reference dielectric function $\varepsilon_0(z)$, or potential $V_0(z)$, must be chosen to tend to the same limits ε_1 and ε_2 as $\varepsilon(z)$ (or the same V_1 and V_2 as $V(z)$). Write $q^2 = q_0^2 + \Delta q^2$, and $\psi = \psi_0 + \psi_1 + \psi_2 + \cdots$, a series in powers of Δq^2. From (3.1) and (3.2), the correction to ψ_0 of nth order in Δq^2 satisfies the equation

$$\frac{\mathrm{d}^2\psi_n}{\mathrm{d}z^2} + q_0^2\psi_n = -\Delta q^2\psi_{n-1} \tag{3.3}$$

for $n = 1, 2, \ldots$ To solve (3.3) we need to construct a Green's function $G(z, z')$ which satisfies

$$\frac{\partial^2 G}{\partial z^2} + q_0^2(z)G = \delta(z - z') \tag{3.4}$$

and has the appropriate asymptotic behaviour to make each ψ_n take the limiting forms given by (3.1), with $r = r_n$ and $t = t_n$. When such a Green's function has been constructed, the corrections $\psi_1, \psi_2, \ldots$ are given by

$$\psi_n(z) = -\int_{-\infty}^{\infty} \mathrm{d}z' \Delta q^2(z') G(z, z') \psi_{n-1}(z') \tag{3.5}$$

so that one can solve for ψ_1 in terms of the known ψ_0, for ψ_2 in terms of ψ_1, and so on. An equivalent formulation is to write (3.1) as an integral equation, using (3.4):

$$\psi(z) = \psi_0(z) - \int_{-\infty}^{\infty} \mathrm{d}z' \Delta q^2(z') G(z, z') \psi(z') \tag{3.6}$$

The sequence (3.5) is then obtained by iteration of (3.6). (The reader not familiar with Green's functions may verify that (3.6) solves (3.1) by operating on both sides with $\frac{\partial^2}{\partial z^2} + q_0^2(z)$, and using (3.4).)

The above is for *any* reference q_0 and ψ_0. For long waves, the natural choice for q_0 and ψ_0 are the functions corresponding to the step dielectric function profile

$$\begin{aligned} \varepsilon_0(z) &= \begin{cases} \varepsilon_1 & (z < 0) \\ \varepsilon_2 & (z > 0) \end{cases} \\ &= \frac{1}{2}(\varepsilon_1 + \varepsilon_2) - \frac{1}{2}(\varepsilon_1 - \varepsilon_2)\ \mathrm{sgn}\ (z), \end{aligned} \tag{3.7}$$

for which q_0 takes the values q_1 and q_2 for $z \lessgtr 0$, and

$$\psi_0(z) = \begin{cases} e^{iq_1 z} + r_0 e^{-iq_1 z} & (z < 0) \\ t_0 e^{iq_2 z} & (z > 0) \end{cases}. \tag{3.8}$$

From the continuity of ψ_0 and $d\psi_0/dz$ at $z = 0$,

$$r_0 = \frac{q_1 - q_2}{q_1 + q_2}, \quad t_0 = \frac{2q_1}{q_1 + q_2} = 1 + r_0. \tag{3.9}$$

The appropriate Green's function $G(z, z')$ must satisfy (3.4), and be such that ψ, as given by (3.6), have the asymptotic forms of (3.1). We see from (3.4) that G is built up from the functions $e^{\pm iq_0 z}$, and that $\partial G/\partial z$ must have a unit discontinuity along the line $z = z'$. For each of $z' > 0$ and $z' < 0$, G has three different analytic forms, the dividing lines being at $z = 0$, and $z = z'$. The continuity of G at both boundaries, and the respective continuity and discontinuity of $\partial G/\partial z$ at $z = 0$ and z' impose conditions on the coefficients. When the coefficients are evaluated, we find the six analytic forms

$$G(z, z') = \begin{cases} \dfrac{e^{iq_2 z'}}{2iq_2}\left(e^{-iq_2 z} - r_0 e^{iq_2 z}\right) & z' \\ \dfrac{e^{i(q_2 z' - q_1 z)}}{i(q_1 + q_2)} & z' = z \\ \dfrac{e^{iq_2 z}}{2iq_2}\left(e^{-iq_2 z'} - r_0 e^{iq_2 z'}\right) & \\ \dfrac{e^{-iq_1 z}}{2iq_1}\left(e^{iq_1 z'} + r_0 e^{-iq_1 z'}\right) & z \\ \dfrac{e^{i(q_2 z - q_1 z')}}{i(q_1 + q_2)} & \\ \dfrac{e^{-iq_1 z'}}{2iq_1}\left(e^{iq_1 z} + r_0 e^{-iq_1 z}\right) & \end{cases} \tag{3.10}$$

From (3.10) and (3.6) we find the analytic form of ψ as $z \to -\infty$:

$$\psi(z) \to e^{iq_1 z} + r_0 e^{-iq_1 z} - e^{-iq_1 z}\left\{\frac{1}{2iq_1}\int_{-\infty}^{0} dz' \Delta q^2(z')\left(e^{iq_1 z'} + r_0 e^{-iq_1 z'}\right)\psi(z') + \frac{1}{i(q_1 + q_2)}\int_{0}^{\infty} dz' \Delta q^2(z') e^{iq_2 z'}\psi(z')\right\}. \tag{3.11}$$

On comparing (3.11) with the asymptotic form in (3.1), we identify the reflection amplitude as r_0 minus the expression within the braces. The nth order term r_n is therefore

$$\begin{aligned} r_n = \frac{i}{2q_1} \int_{-\infty}^{0} \mathrm{d}z' \Delta q^2(z') \left(\mathrm{e}^{iq_1 z'} + r_0 \mathrm{e}^{-iq_1 z'} \right) \psi_{n-1}(z') \\ + \frac{i}{q_1 + q_2} \int_{0}^{\infty} \mathrm{d}z' \Delta q^2(z') \mathrm{e}^{iq_2 z'} \psi_{n-1}(z'). \end{aligned} \tag{3.12}$$

In particular, the first order term r_1 is given by

$$\begin{aligned} r_1 = \frac{i}{2q_1} \int_{-\infty}^{0} \mathrm{d}z' \Delta q^2(z') \left(\mathrm{e}^{iq_1 z'} + r_0 \mathrm{e}^{-iq_1 z'} \right)^2 \\ + \frac{i}{q_1 + q_2} (1 + r_0) \int_{0}^{\infty} \mathrm{d}z' \Delta q^2(z') \mathrm{e}^{2iq_2 z'} \end{aligned} \tag{3.13}$$

When $\varepsilon_1 = \varepsilon_2$ (and q_1, q_2 take the common value q_0) r_1 takes the simple form

$$r_1 = \frac{i}{2q_0} \int_{-\infty}^{\infty} \mathrm{d}z' \Delta q^2(z') \mathrm{e}^{2iq_0 z'}, \tag{3.14}$$

from which the reflection formulae used in Sect. 1.5 may be obtained.

The general expression for r implicit in (3.11) may be put into a simpler form by using (3.8) and (3.9):

$$r = r_0 - \frac{1}{2iq_1} \int_{-\infty}^{\infty} \mathrm{d}z' \Delta q^2(z') \psi(z') \psi_0(z'). \tag{3.15}$$

This formula can be obtained directly from the comparison identity (2.6), but without the perturbation theory derived above is capable of giving only r_1.

We have developed the perturbation series in terms of Δq^2. Since $q^2(z) = \varepsilon(z)\omega^2/c^2 - K^2$,

$$\Delta q^2 = \frac{\omega^2}{c^2} (\varepsilon - \varepsilon_0), \tag{3.16}$$

and the *perturbation is independent of the angle of incidence*. This simple result holds only for the s and particle waves; the more complex p wave perturbation theory has terms in K^2, as we shall see in Sect. 3.4.

3.2 The *s* Wave to Second Order in the Interface Thickness

We see from (3.16) that the perturbation Δq^2 is small, for arbitrary $\Delta\varepsilon$, in the long wavelength limit. In this section we will obtain corrections to r_0 in terms of integrals over the difference $\varepsilon - \varepsilon_0$ between the actual and the step profile. It will be convenient to define the integrals λ_n, of dimension (length)^n, as

$$\lambda_n = \int_{-\infty}^{\infty} \mathrm{d}z[\varepsilon(z) - \varepsilon_0(z)]z^{n-1}. \tag{3.17}$$

Consider the first order (in Δq^2) expression for r_1, as given by (3.13). The integrals in (3.13) contain a factor $\Delta q^2 = (\omega^2/c^2)(\varepsilon - \varepsilon_0)$, assumed to be of short range, by which we mean that the λ_n converge for all n. (This includes profiles such as the hyperbolic tangent defined in (2.75), for which $\varepsilon - \varepsilon_0$ tends to zero exponentially in z as $z \to \pm\infty$.) Let the range of $\varepsilon - \varepsilon_0$ be characterized by a length Δz, an interface thickness. In the long wave limit, $(\omega/c)\Delta z$ is a small dimensionless parameter, as are $q_1\Delta z$ and $q_2\Delta z$. We expand the factors $\mathrm{e}^{iq_1z} + r_0\mathrm{e}^{-iq_1z}$ and e^{iq_2z} in (3.12) and (3.13) in powers of q_1z, to obtain an expansion in the above small parameters. For r_1 we have, to second order in the interface thickness,

$$r_1 = \frac{2iq_1\omega^2/c^2}{(q_1+q_2)^2}\{\lambda_1 + 2iq_2\lambda_2 + \cdots\}. \tag{3.18}$$

The corresponding expression for r_2 is

$$r_2 = \frac{-2q_1\omega^4/c^4}{(q_1+q_2)^3}\lambda_1^2 + \cdots, \tag{3.19}$$

obtained from (3.12) and the result

$$\psi_1(0) = \frac{2iq_1\omega^2/c^2}{(q_1+q_2)^2}\lambda_1 + \cdots \tag{3.20}$$

The leading term in r_3 is of third order in Δz:

$$r_3 = -\frac{2iq_1\omega^6/c^6}{(q_1+q_2)^4}\lambda_1^3 + \cdots, \tag{3.21}$$

and can thus be omitted from the second order expression, which reads

$$r = r_0 + \frac{2iq_1\omega^2/c^2}{(q_1+q_2)^2}\left\{\lambda_1 + 2iq_2\lambda_2 + \frac{i(\omega^2/c^2)\lambda_1^2}{q_1+q_2}\right\} + \cdots. \tag{3.22}$$

We now need a notation to distinguish between terms which are nth order in Δq^2, and those which are nth order in Δz. We write the latter as r_{sn} (for s wave reflection amplitude component which is nth order in Δz); r_{s0} and r_0 are equal, but from $n = 1$ onwards r_n and r_{sn} are different. In the Δz expansion we have

$$r_{s1} = \frac{2iq_1\omega^2/c^2}{(q_1+q_2)^2}\lambda_1, \tag{3.23}$$

$$r_{s2} = \frac{-2iq_1\omega^2/c^2}{(q_1+q_2)^2}\left\{2q_2\lambda_2 + \frac{\omega^2/c^2}{q_1+q_2}\lambda_1^2\right\}. \tag{3.24}$$

The s wave reflection amplitude is thus determined to second order in the interface thickness in terms of the two integrals λ_1 and λ_2 over the difference between the given profile and the step profile. These integrals depend on the choice of relative positioning of $\varepsilon(z)$ and $\varepsilon_0(z)$, as shown in Fig. 3.1.

The result obtained by theory for any observable, such as $|r_s|^2$, must be independent of an arbitrary choice made in calculating that observable. It follows that

$$\begin{aligned}|r_s|^2 = |r_{s0} + r_{s1} + r_{s2} + \cdots|^2 &= |r_{s0}|^2 + 2\,\mathrm{Re}\left(r_{s0}^* r_{s1}\right) \\ &+ \left\{|r_{s1}|^2 + 2\,\mathrm{Re}\left(r_{s0}^* r_{s2}\right)\right\} + \cdots\end{aligned} \tag{3.25}$$

must be invariant with respect to the choice of relative positioning of $\varepsilon(z)$ and $\varepsilon_0(z)$. Consider for example the case where ε is real, and also q_2 is real (thus excluding total internal reflection). Then r_{s0} and r_{s2} are both real, while r_{s1} is purely imaginary. The lowest order correction to r_{s0}^2 is then the second order term within the braces in (3.25). From (3.23) and (3.24), this is equal to

$$\left\{|r_{s1}|^2 + 2r_{s0}r_{s2}\right\} = -\frac{4q_1q_2\omega^4/c^4}{(q_1+q_2)^4}\left[2(\varepsilon_1-\varepsilon_2)\lambda_2 - \lambda_1^2\right]. \tag{3.26}$$

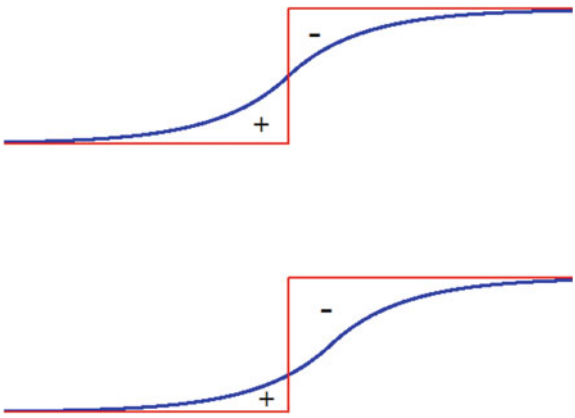


Fig. 3.1 Dependence of the integral $\lambda_1 = \int_{-\infty}^{\infty} \mathrm{d}z(\varepsilon - \varepsilon_0)$ on relative positioning of $\varepsilon(z)$ and the step function $\varepsilon_0(z)$. The sign of the contributions is indicated; the magnitude of the positive and negative contributions is equal to the area enclosed between the actual dielectric function profile and the reference step profile. In the *upper diagram* the areas are equal, and $\lambda_1 = 0$; such positioning is always possible when $\varepsilon(z)$ is real and $\varepsilon_1 \neq \varepsilon_2$. The profile drawn is the double exponential, defined in Sect. 3.6, equation (3.69)

We thus expect the combination $2(\varepsilon_1 - \varepsilon_2)\lambda_2 - \lambda_1^2$ to be invariant with respect to the relative positioning of ε and ε_0. This turns out to be true, as we will show in the next section.

3.3 Integral Invariants

Let $\varepsilon(z)$ be a function with asymptotic values $\varepsilon(-\infty) = \varepsilon_1$ and $\varepsilon(+\infty) = \varepsilon_2$, and $\varepsilon_0(z)$ the step function taking values ε_1 for $z < 0$ and ε_2 for $z > 0$. Consider the dependence of the integral

$$\lambda_{n+1}(s) = \int_{-\infty}^{\infty} \mathrm{d}z[\varepsilon(z - s) - \varepsilon_0(z)]z^n \tag{3.27}$$

on the shift parameter s. (In changing s we change the relative positioning of the two profiles.) We have

$$\begin{aligned}\lambda_{n+1}(s) &= \int_{-\infty}^{\infty} \mathrm{d}z[\varepsilon(z)-\varepsilon_0(z)+\varepsilon_0(z)-\varepsilon_0(z+s)](z+s)^n \\ &= \int_{-\infty}^{\infty} \mathrm{d}z[\varepsilon(z)-\varepsilon_0(z)](z+s)^n + (\varepsilon_1-\varepsilon_2)\int_{-s}^{0} \mathrm{d}z(z+s)^n \\ &= \int_{-\infty}^{\infty} \mathrm{d}z[\varepsilon(z)-\varepsilon_0(z)]\left[z^n + \binom{n}{1} z^{n-1}s + \cdots + s^n\right] + (\varepsilon_1-\varepsilon_2)\frac{s^{n+1}}{n+1} \\ &= \lambda_{n+1}(0) + \binom{n}{1}\lambda_n(0)s + \cdots + \lambda_1(0)s^n + (\varepsilon_1-\varepsilon_2)\frac{s^{n+1}}{n+1}\end{aligned} \tag{3.28}$$

The shift-dependence of the first three integrals is thus

$$\begin{aligned}\lambda_1(s) &= \lambda_1(0) + (\varepsilon_1-\varepsilon_2)s, \\ \lambda_2(s) &= \lambda_2(0) + \lambda_1(0)s + (\varepsilon_1-\varepsilon_2)\frac{s^2}{2}, \\ \lambda_3(s) &= \lambda_3(0) + 2\lambda_2(0)s + \lambda_1^2(0)s^2 + (\varepsilon_1-\varepsilon_2)\frac{s^3}{3}.\end{aligned} \tag{3.29}$$

Note that

$$2(\varepsilon_1-\varepsilon_2)\lambda_2(s) - \lambda_1^2(s) = 2(\varepsilon_1-\varepsilon_2)\lambda_2(0) - \lambda_1^2(0), \tag{3.30}$$

so that

$$i_2 = 2(\varepsilon_1-\varepsilon_2)\lambda_2 - \lambda_1^2 \tag{3.31}$$

is *invariant* with respect to the relative positioning of ε and ε_0. Similarly

$$i_3 = 3(\varepsilon_1-\varepsilon_2)^2\lambda_3 - 6(\varepsilon_1-\varepsilon_2)\lambda_2\lambda_1 + 2\lambda_1^3 \tag{3.32}$$

is an integral invariant. There is an infinite hierarchy of such invariants (assuming the existence of λ_n for all n). The general formula has been given by Lekner (1984).

The invariance of i_2 was suggested on physical grounds, and follows in a straightforward way from the analysis given above. Nevertheless, the form (3.31) is not obviously invariant, and it is useful to express i_2 in a form which *is* manifestly invariant:

$$i_2 = -\int_{-\infty}^{\infty} \mathrm{d}z_1 \int_{-\infty}^{\infty} \mathrm{d}z_2[\varepsilon(z_1) - \varepsilon_0(z_1 - z_2)][\varepsilon(z_2) - \varepsilon_0(z_2 - z_1)]. \tag{3.33}$$

This form immediately suggests how other integral invariants may be generated: multiplication of the integrand in (3.33) by $f(z_1 - z_2)$ will produce integral invariants, for any f (subject only to the convergence of the integral). An example is $f = \mathrm{e}^{ikz}$, which produces another set of invariants via its expansion in powers of k (Lekner 1984, Appendix A).

Any single-valued function of a step function is also a step function. The invariants developed above thus have endless generalizations. For the p wave, we will need invariants arising out of integrals over the difference between the reciprocals of ε and ε_0. We define, in parallel with (3.27),

$$\Lambda_{n+1}(s) = \varepsilon_1\varepsilon_2 \int_{-\infty}^{\infty} \mathrm{d}z \left[\frac{1}{\varepsilon_0(z)} - \frac{1}{\varepsilon(z-s)}\right] z^n.$$

Their transformation properties are the same as those of the λ's: for example

$$\begin{aligned} \Lambda_1(s) &= \Lambda_1(0) + (\varepsilon_1 - \varepsilon_2)s, \\ \Lambda_2(s) &= \Lambda_2(0) + \Lambda_1(0)s + (\varepsilon_1 - \varepsilon_2)\frac{s^2}{2}. \end{aligned} \tag{3.35}$$

Thus the analogue of i_2,

$$J_2 = 2(\varepsilon_1 - \varepsilon_2)\Lambda_2 - \Lambda_1^2 \tag{3.36}$$

is an integral invariant, etc. We note also that, from (3.29) and (3.35), the difference $I_1 = \Lambda_1 - \lambda_1$ is an invariant. This invariance is made explicit by means of the identity

$$I_1 = \Lambda_1 - \lambda_1 = \varepsilon_1\varepsilon_2 \int_{-\infty}^{\infty} \mathrm{d}z \left(\frac{1}{\varepsilon_0} - \frac{1}{\varepsilon}\right) - \int_{-\infty}^{\infty} \mathrm{d}z(\varepsilon - \varepsilon_0) = \int_{-\infty}^{\infty} \mathrm{d}z \frac{(\varepsilon_1 - \varepsilon)(\varepsilon - \varepsilon_2)}{\varepsilon}. \tag{3.37}$$

The next mixed invariant is

$$I_2 = \varepsilon_1\varepsilon_2 \int_{-\infty}^{\infty} \mathrm{d}z_1 \int_{-\infty}^{\infty} \mathrm{d}z_2[\varepsilon(z_1) - \varepsilon_0(z_1 - z_2)]\left[\frac{1}{\varepsilon(z_2)} - \frac{1}{\varepsilon_0(z_2 - z_1)}\right] \\ = (\varepsilon_1 - \varepsilon_2)(\lambda_2 + \Lambda_2) - \lambda_1\Lambda_1 = \frac{1}{2}(i_2 + J_2 + I_1^2). \tag{3.38}$$

Finally, we shall need another invariant, related to I_2, which enters in the expressions for $|r_p|^2$ and r_p/r_s to second order. This is

$$j_2 = \int_{-\infty}^{\infty} \mathrm{d}z_1 \int_{-\infty}^{\infty} \mathrm{d}z_2\varepsilon_0(z_2 - z_1)[\varepsilon(z_1) - \varepsilon_0(z_1 - z_2)]\left[\frac{1}{\varepsilon(z_2)} - \frac{1}{\varepsilon_0(z_2 - z_1)}\right]. \tag{3.39}$$

3.4 $|r_p|^2$ and r_p/r_s to Second Order

The last section developed the tools which enable us to complete the characterization of the reflection of electromagnetic waves to second order in the interface thickness. For the p wave, $B = (0, \mathrm{e}^{i(Kx-\omega t)}B(z), 0)$, with

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{\varepsilon}\frac{\mathrm{d}B}{\mathrm{d}z}\right) + \left(\frac{\omega^2}{c^2} - \frac{K^2}{\varepsilon}\right)B = 0, \quad \mathrm{e}^{iq_1z} - r_p\mathrm{e}^{-iq_1z} \leftarrow B \rightarrow \left(\frac{\varepsilon_2}{\varepsilon_1}\right)^{1/2} t_p\mathrm{e}^{iq_2z}. \tag{3.40}$$

A perturbation theory for the p wave will be sketched below. For the first order results, however, the full perturbation theory with its Green's function is not needed. We can obtain the first order results directly from a comparison identity based on (3.40) and its companion for B_0:

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{\varepsilon_0}\frac{\mathrm{d}B_0}{\mathrm{d}z}\right) + \left(\frac{\omega^2}{c^2} - \frac{K^2}{\varepsilon_0}\right)B_0 = 0, \quad \mathrm{e}^{iq_1z} - r_{p0}\mathrm{e}^{-iq_1z} \leftarrow B_0 \rightarrow \left(\frac{\varepsilon_2}{\varepsilon_1}\right)^{1/2} t_{p0}\mathrm{e}^{iq_2z}. \tag{3.41}$$

This is (Lekner 1982a, equation (46)),

$$r_p = r_{p0} + \frac{1}{2iQ_1}\int_{-\infty}^{\infty} \mathrm{d}z\left\{\left(\frac{1}{\varepsilon_0} - \frac{1}{\varepsilon}\right)K^2BB_0 + (\varepsilon - \varepsilon_0)CC_0\right\}, \tag{3.42}$$

where $Q_1 = q_1/\varepsilon_1, C = \mathrm{d}B/\varepsilon \mathrm{d}z$ and $C_0 = \varepsilon_0^{-1}\mathrm{d}B_0/\mathrm{d}z$. (Note that C_0 is continuous at the discontinuity of ε_0, while $\mathrm{d}B_0/\mathrm{d}z$ is not.) To first order in the interface thickness, it is suffices to replace BB_0 by $B_0^2(0)$ and CC_0 by $C_0^2(0)$, where, from (1.28) and (1.31),

$$B_0(0) = \frac{2Q_1}{Q_1 + Q_2}, \qquad C_0(0) = \frac{2iQ_1Q_2}{Q_1 + Q_2}. \tag{3.43}$$

We obtain

$$r_{p1} = \frac{-2iQ_1}{(Q_1 + Q_2)^2}\left\{\frac{K^2\Lambda_1}{\varepsilon_1\varepsilon_2} - Q_2^2\lambda_1\right\}. \tag{3.44}$$

This result and (3.23) are together sufficient to determine r_p/r_s to first order (in contrast to the expressions for $|r_s|^2$ and $|r_p|^2$, which have no first order terms when ε and q_2 are real). We have

$$\begin{aligned}\frac{r_p}{r_s} &= \frac{r_{p0} + r_{p1} + r_{p2} + \cdots}{r_{s0} + r_{s1} + r_{s2} + \cdots} \\ &= \frac{r_{p0}}{r_{s0}}\left\{1 + \left(\frac{r_{p1}}{r_{p0}} - \frac{r_{s1}}{r_{s0}}\right) + \left(\frac{r_{p2}}{r_{p0}} - \frac{r_{p1}r_{s1}}{r_{p0}r_{s0}} + \frac{r_{s1}^2}{r_{s0}^2} - \frac{r_{s2}}{r_{s0}}\right) + \cdots\right\}.\end{aligned} \tag{3.45}$$

The first order term is given by (for $\varepsilon_1 \neq \varepsilon_2$)

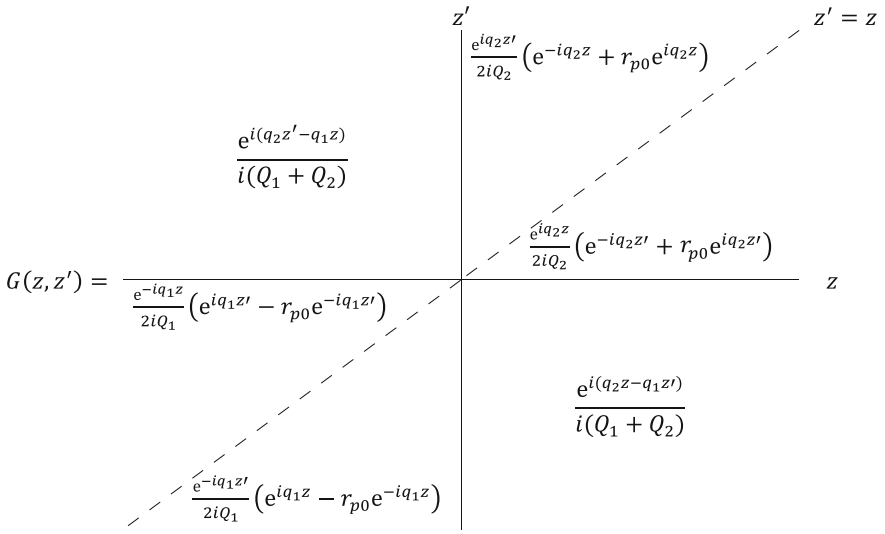
$$r_{s0}\left(\frac{r_p}{r_s}\right)_1 = r_{p0}\left(\frac{r_{p1}}{r_{p0}} - \frac{r_{s1}}{r_{s0}}\right) = -\frac{2iQ_1K^2}{(Q_1 + Q_2)^2}\frac{I_1}{\varepsilon_1\varepsilon_2} \tag{3.46}$$

(on using $r_{s0} = (q_1 - q_2)/(q_1 + q_2)$, $-r_{p0} = (Q_1 - Q_2)/(Q_1 + Q_2)$). The special case of reflection by a thin film between like media (when $\varepsilon_1 = \varepsilon_2$) will be discussed in the next section.

For the second order term in r_p/r_s and $|r_p|^2$ we need r_{p2}. This requires a perturbation theory for (3.40) based on (3.41), which is more complex than the perturbation theory for the s wave. We will give an outline here; further details may be found in Lekner (1982a, b, 1984). One constructs a Green's function $G(z, z')$ satisfying

$$\frac{\partial}{\partial z}\left(\frac{1}{\varepsilon_0}\frac{\partial G}{\partial z}\right) + \left(\frac{\omega^2}{c^2} - \frac{K^2}{\varepsilon_0}\right)G = \delta(z - z'), \tag{3.47}$$

and incorporating the required boundary conditions. This is (compare (3.10))



(3.48)

$B(z)$ satisfies an integro-differential equation

$$B(z) = B_0(z) - \int_{-\infty}^{\infty} \mathrm{d}z' \left\{ \left(\frac{1}{\varepsilon_0(z')} - \frac{1}{\varepsilon(z')} \right) K^2 B(z') G(z,z') + (\varepsilon(z') - \varepsilon_0(z')) C(z') \frac{1}{\varepsilon_0(z')} \frac{\partial G}{\partial z'} \right\}, \tag{3.49}$$

from which (3.42) may be obtained by taking the limit $z \to -\infty$. Considerable work is required to extract the second order term r_{p2}. The result for $|r_p|^2$ to second order in the interface thickness is, for real ε and q_2,

$$|r_p|^2 = r_{p0}^2 - \frac{4Q_1Q_2}{\varepsilon_1\varepsilon_2(Q_1+Q_2)^4} \left\{ \frac{\omega^4}{c^4} i_2 - \frac{\omega^2}{c^2} K^2 \left[j_2 + \left(\frac{1}{\varepsilon_1} + \frac{1}{\varepsilon_2} \right) i_2 \right] + \frac{K^4}{\varepsilon_1\varepsilon_2} \left[(\varepsilon_1 + \varepsilon_2) j_2 - I_1^2 \right] \right\} + \cdots. \tag{3.50}$$

This is to be compared with the simpler expression for the s wave:

$$|r_s|^2 = r_{s0}^2 - \frac{4q_1q_2\omega^4/c^4}{(q_1+q_2)^4} i_2 + \cdots. \tag{3.51}$$

The p wave is more complicated, because the orientation of the electric field relative to the interface changes with the angle of incidence, while for the s polarization the electric field is always parallel to the interface. Measurement of

$R_s = |r_s|^2$ for an interface in the long wave limit can give the value of one invariant, i_2, irrespective of the angle of incidence. Measurement of R_p, at a minimum of three angles, can give the values of i_2, j_2 and I_1^2.

As we saw in (3.46), the invariant I_1 (complete with sign) can be found from the ellipsometric determination of r_p/r_s to first order in the interface thickness. The expression for r_p/r_s to second order is, for $\varepsilon_1 \neq \varepsilon_2$,

$$r_{s0}\left(\frac{r_p}{r_s}\right) = r_{p0} - \frac{2iQ_1K^2/\varepsilon_1\varepsilon_2}{(Q_1+Q_2)^2} I_1 + \frac{2Q_1(K^2/\varepsilon_1\varepsilon_2)^2}{(Q_1+Q_2)^3} I_1^2 + \frac{2Q_1Q_2K^2}{(Q_1+Q_2)^2} \frac{j_2 - \left(\frac{1}{\varepsilon_1}+\frac{1}{\varepsilon_2}\right) i_2}{\varepsilon_1 - \varepsilon_2} + \cdots. \tag{3.52}$$

Note that the first and second order terms vanish at normal incidence ($K = 0$), as they must since there $r_p/r_s = 1$, identically.

We saw in Sect. 1.2 that an interface is expected to be nearly non-reflecting to the p wave at the angle of incidence defined by $Q_1 = Q_2$, namely at the Brewster angle $\theta_B = \arctan(\varepsilon_2/\varepsilon_1)^{1/2}$. In Sect. 2.3 we proved the existence of at least one ellipsometric Brewster angle, the *principal angle* θ_P, defined as the angle at which the real part of r_p/r_s is zero. We can obtain an expression for the difference $\Delta\theta = \theta_P - \theta_B$ in the long wave limit. Assume absorption to be absent (real ε), in which case I_1, i_2 and j_2 are real; and take $\theta_1 < \theta_c$, so q_2 and Q_2 are real. Then from (3.52) we find that, to second order in the interface thickness, $\mathrm{Re}(r_p/r_s)$ is zero when the expression

$$r_{p0} + \frac{2Q_1(K^2/\varepsilon_1\varepsilon_2)}{(Q_1+Q_2)^3}\left\{\frac{K^2}{\varepsilon_1\varepsilon_2} I_1 + \frac{\varepsilon_1\varepsilon_2 Q_2(Q_1+Q_2)}{\varepsilon_1 - \varepsilon_2}\left[j_2 - \left(\frac{1}{\varepsilon_1}+\frac{1}{\varepsilon_2}\right) i_2\right]\right\}$$

is zero. Thus $\Delta\theta$ is second order in the interface thickness. On using (1.36) and (1.37) we find

$$\Delta\theta = \frac{(\varepsilon_1\varepsilon_2)^{5/2}\omega^2/c^2}{\varepsilon_1\left(\varepsilon_1^2 - \varepsilon_2^2\right)^2}\left[j_2 - \left(\frac{1}{\varepsilon_1}+\frac{1}{\varepsilon_2}\right) i_2 - \frac{1}{2}\left(\frac{1}{\varepsilon_1}-\frac{1}{\varepsilon_2}\right) I_1^2\right] + \cdots \tag{3.53}$$

The reader may have noted divergences in the above expressions when $\varepsilon_1 \to \varepsilon_2$. The reflection from interfaces between like media require special treatment, as discussed in the next section, and also in the context of variational theory in Chap. 4.

3.5 Reflection by a Thin Film Between Like Media

When media 1 and 2 are the same (as for a soap film in air), the formulae given above for $|r_s|^2$ and $|r_p|^2$ take a simpler form, and the formula for r_p/r_s takes a different form, since both r_{s0} and r_{p0} are now zero. Let ε_0 now denote the common value of ε_1 and ε_2, q_0 the common value of q_1 and q_2, et cetera. The integrals λ_1 and Λ_1 are now separately invariant (translating the constant functions ε_0 or ε_0^{-1} makes no difference), and the second order invariants i_2 and j_2 take the values

$$i_2 = -\lambda_1^2, \quad j_2 = -\frac{2\lambda_1\Lambda_1}{\varepsilon_0}. \tag{3.54}$$

The reflectivities to second order in the interface thickness are now

$$|r_s|^2 = \frac{\omega^4/c^4}{4q_0^2}\lambda_1^2 + \cdots \tag{3.55}$$

$$|r_p|^2 = \frac{\omega^4/c^4}{4q_0^2}\lambda_1^2\left\{1 - \frac{2(cK/\omega)^2}{\varepsilon_0}\left[\frac{\Lambda_1}{\lambda_1}+1\right] + \frac{(cK/\omega)^4}{\varepsilon_0^2}\left[4\frac{\Lambda_1}{\lambda_1} + \left(\frac{I_1}{\lambda_1}\right)^2\right]\right\} + \cdots. \tag{3.56}$$

The ellipsometric ratio r_p/r_s cannot be found from (3.45), since now r_{s0} and r_{p0} are zero. We have

$$\frac{r_p}{r_s} = \frac{r_{p1}+r_{p2}+\cdots}{r_{s1}+r_{s2}+\cdots} = \frac{r_{p1}}{r_{s1}} + \frac{r_{p2}r_{s1} - r_{s2}r_{p1}}{r_{s1}^2} + \cdots, \tag{3.57}$$

the leading term now being zero order in the interface thickness:

$$\frac{r_{p1}}{r_{s1}} = \cos^2\theta_0 - \frac{\Lambda_1}{\lambda_1}\sin^2\theta_0. \tag{3.58}$$

This ratio correctly tends to unity as θ_0, the angle of incidence and refraction, tends to zero. When λ_1 and Λ_1 have the same sign, as would normally be the case, there is an angle at which the film is non-reflecting to the p wave (to lowest order in the film thickness). This angle is an analogue of the Brewster angle $\arctan(\varepsilon_2/\varepsilon_1)^{1/2}$ for unlike media. From (3.58), the thin film p wave zero reflection occurs at

$$\theta_0 = \arctan\left(\frac{\lambda_1}{\Lambda_1}\right)^{1/2} = \arctan\left\{\frac{\int_{-\infty}^{\infty} \mathrm{d}z(\varepsilon-\varepsilon_0)/\varepsilon_0}{\int_{-\infty}^{\infty} \mathrm{d}z(\varepsilon-\varepsilon_0)/\varepsilon}\right\}^{1/2}. \tag{3.59}$$

The above results may be compared to the exact formulae for a homogeneous layer of dielectric constant ε, with waves incident from (and transmitted into) a medium with dielectric constant ε_0. From (2.52) to (2.68),

$$r_s = \mathrm{e}^{2iq_0z_1}\frac{i\left(q^2-q_0^2\right)\tau}{2qq_0 - i\left(q^2+q_0^2\right)\tau}, \tag{3.60}$$

and

$$-r_p = \mathrm{e}^{2iq_0z_1}\frac{i\left(Q^2-Q_0^2\right)\tau}{2QQ_0 - i\left(Q^2+Q_0^2\right)\tau}, \tag{3.61}$$

where $q^2 = \varepsilon(\omega^2/c^2) - K^2$, $Q = q/\varepsilon$, and $\tau = \tan q\Delta z$. For the homogeneous film of thickness Δz, $\lambda_1 = (\varepsilon - \varepsilon_0)\Delta z$ and $\Lambda_1 = (\varepsilon_0/\varepsilon)\lambda_1$, and (3.55) and (3.56) agree with the reflectivities obtained from (3.60) and (3.61) to second order in Δz. The ratio r_p/r_s may be written as

$$\frac{r_p}{r_s} = \left[1 - \left(\frac{cK}{\omega}\right)^2\left(\frac{1}{\varepsilon}+\frac{1}{\varepsilon_0}\right)\right]\left\{\frac{1-\frac{i}{2}\left(\frac{q}{q_0}+\frac{q_0}{q}\right)\tau}{1-\frac{i}{2}\left(\frac{Q}{Q_0}+\frac{Q_0}{Q}\right)\tau}\right\}. \tag{3.62}$$

The square bracket in (3.62) is equal to $\cos^2\theta_0 - (\varepsilon_0/\varepsilon)\sin^2\theta_0$, and is zero at $\theta_0 = \arctan(\varepsilon/\varepsilon_0)^{1/2}$, which is the same as the Brewster angle for a sharp boundary between ε_0 and an *infinite* medium ε. It follows that a homogeneous film between like media is always non-reflecting to the p wave at the same angle, irrespective of its thickness. An inhomogeneous film, on the other hand, is non-reflecting at θ_0 given by (3.59) only to the lowest order in the film thickness.

There is a difficulty associated with the reflectivity formulae (3.55) and (3.56) which is shared by the first order perturbation expressions

$$r_s^{\mathrm{pert}} = \frac{i\omega^2/c^2}{2q_0}\int_{-\infty}^{\infty}\mathrm{d}z[\varepsilon(z)-\varepsilon_0]\mathrm{e}^{2iq_0z}, \tag{3.63}$$

$$r_p^{\mathrm{pert}} = \frac{1}{2iQ_0}\int_{-\infty}^{\infty}\mathrm{d}z\left\{\left(\frac{1}{\varepsilon_0}-\frac{1}{\varepsilon}\right)K^2 - (\varepsilon-\varepsilon_0)Q_0^2\right\}\mathrm{e}^{2iq_0z}. \tag{3.64}$$

((3.63) is the same as (3.14), and (3.64) may be obtained from (3.42), as will be discussed in Chap. 4). The difficulty occurs at grazing incidence, when $\theta_0 \to \pi/2$ and $q_0 = \sqrt{\varepsilon_0}(\omega/c)\cos\theta_0 \to 0$, and all these expressions are *divergent.* This divergence is unphysical (r_s and r_p must stay within the unit circle), and persists in higher order perturbation theory. One of the virtues of the variational expressions for r_s and r_p to be derived in Chap. 4 is that these are no longer divergent at grazing

incidence, and in fact take the universal limiting values -1 and $+1$ as θ tends to $\pi/2$ (as shown in Sect. 2.3).

The reader may be puzzled as to how an expression like (3.55), which we have stated to be exact to second order in the interface thickness, could be divergent (and wrong) at grazing incidence. This paradox is resolved by considering the order of the two limiting processes (the long wave limit, and the limit of $\theta_0 \to \pi/2$). An illustration of the application of these two limits is provided by the profile

$$\varepsilon(z) = \varepsilon_0 + \Delta\varepsilon \operatorname{sech}^2(z/a), \tag{3.65}$$

which will be discussed in detail in Sect. 4.3. The s wave reflectivity depends on two dimensionless parameters,

$$\alpha = \Delta\varepsilon(\omega/c)^2 a^2, \quad \beta = q_0\, a. \tag{3.66}$$

For $\alpha > -1/4$ it is given by

$$R_s = \frac{\cos^2\left[\frac{\pi}{2}(1+4\alpha)^{1/2}\right]}{\cos^2\left[\frac{\pi}{2}(1+4\alpha)^{1/2}\right] + \sinh^2(\pi\beta)}. \tag{3.67}$$

When α and β are both small compared to $1/\pi$, this takes the form $\alpha^2/\left(\alpha^2+\beta^2\right)$. For fixed small α, and $\beta \to 0$ (fixed interface thickness and angle of incidence tending to $\pi/2$), $R_s \to \alpha^2/\alpha^2 = 1$. For $\alpha/\beta \to 0$, which is the long wave limit $(\omega/c)\alpha \to 0$ with θ_0 fixed, $R_s \to \alpha^2/\beta^2 = [\Delta\varepsilon(\omega^2/c^2)a/q_0]^2$, and is correctly given by (3.55), since $\lambda_1 = 2a\Delta\varepsilon$ for this profile.

3.6 Six Profiles and Their Integral Invariants

We have seen that, to second order in the interface thickness, the s wave reflectivity is characterized by one integral invariant i_2, while the p wave reflectivity and r_p/r_s, are characterized by the invariants I_1, i_2 and j_2. Knowledge of these three invariants is thus sufficient to determine the reflection properties to this order. Conversely, reflection and ellipsometric studies with wavelengths that are large compared to the profile thickness can determine no more than these three integrals over the profile.

We shall list six dielectric function profiles and their invariants. Four of these have been studied already; the others (the linear and double exponential profiles) are useful simple models in the statistical mechanics and electrodynamics of interfaces.

In the following definitions z_1 is arbitrary, and $z_2 = z_1 + \Delta z$. For the first four profiles Δz is the total extent of the interface, while for the double exponential and hyperbolic tangent profiles Δz is a measure of the interface thickness. In the interval $z_1 \le z \le z_2$, the first four profiles take the form

$$\begin{gathered}\text{homogeneous layer: } \varepsilon(z) = \varepsilon \,(\text{a constant for } z_1 < z < z_2) \\ \text{linear: } \varepsilon(z) = \varepsilon_1 + (\varepsilon_2 - \varepsilon_1)(z - z_1)/\Delta z \\ \text{Rayleigh: } \varepsilon(z) = \left[\varepsilon_1^{-1/2} + \left(\varepsilon_2^{-1/2} - \varepsilon_1^{-1/2}\right)(z - z_1)/\Delta z\right]^{-2} \\ \text{single exponential: } \varepsilon(z) = \varepsilon_1 \exp[\ln(\varepsilon_2/\varepsilon_1)(z - z_1)/\Delta z]\end{gathered} \tag{3.68}$$

The double exponential profile (illustrated in Fig. 3.1) has two analytic forms:

$$\varepsilon(z) = \begin{cases} \varepsilon_1 + \frac{1}{2}(\varepsilon_2 - \varepsilon_1)\exp(z - z_1)/\Delta z & (z < z_1) \\ \varepsilon_2 + \frac{1}{2}(\varepsilon_1 - \varepsilon_2)\exp(z_1 - z)/\Delta z & (z > z_1) \end{cases} \tag{3.69}$$

The hyperbolic tangent profile is given by

$$\varepsilon(z) = \frac{1}{2}(\varepsilon_1 + \varepsilon_2) - \frac{1}{2}(\varepsilon_1 - \varepsilon_2)\tanh[(z - z_1)/2\Delta z]. \tag{3.70}$$

The evaluation of invariants is usually made simpler by choosing z_1 so as to make $\lambda_1 = 0$. For the symmetric profiles (linear, double-exponential, and hyperbolic tangent) the required positioning is at the origin ($z_1 = 0$). The invariants i_2 and j_2 then reduce to

$$i_2 = 2(\varepsilon_1 - \varepsilon_2) \int_{-\infty}^{\infty} \mathrm{d}z[\varepsilon(z) - \varepsilon_0(z)]z \quad (\lambda_1 = 0) \tag{3.71}$$

$$j_2 = 2(\varepsilon_1 - \varepsilon_2) \int_{-\infty}^{\infty} \mathrm{d}z[\varepsilon(z) - \varepsilon_0(z)]\left\{\frac{z}{\varepsilon(z)} + \int_0^z \frac{d\zeta}{\varepsilon(\zeta)}\right\} \quad (\lambda_1 = 0) \tag{3.72}$$

The integrations leading to the invariants are mostly elementary. Further details may be found in the references to Table 1 of Lekner (1984), and in Appendix C of that reference. The results are summarized in Table 3.1. The integrals in the j_2 expressions for the double exponential and tanh profiles can be evaluated in terms of the dilog function,

$$\mathrm{dilog}(x) = \int_1^x \mathrm{d}y \frac{\ln y}{1 - y}.$$

However, such evaluation does not lead to simplification of the results.

Table 3.1 First and second order integral invariants for six profiles

Profile	$I_1/\Delta z$	$i_2/(\Delta z)^2$	$j_2/(\Delta z)^2$
Homogeneous layer (two-step profile)	$(\varepsilon_1 - \varepsilon)(\varepsilon - \varepsilon_2)/\varepsilon$	$(\varepsilon_1 - \varepsilon)(\varepsilon - \varepsilon_2)$	$2(\varepsilon_1 - \varepsilon)(\varepsilon - \varepsilon_2)/\varepsilon$
Linear	$\frac{1}{2}(\varepsilon_1 + \varepsilon_2) - \frac{\varepsilon_1\varepsilon_2}{\varepsilon_1 - \varepsilon_2}\ln\frac{\varepsilon_1}{\varepsilon_2}$	$\frac{1}{12}(\varepsilon_1 - \varepsilon_2)^2$	$\frac{1}{2}(\varepsilon_1 + \varepsilon_2) - \frac{\varepsilon_1\varepsilon_2}{\varepsilon_1 - \varepsilon_2}\ln\frac{\varepsilon_1}{\varepsilon_2}$
Rayleigh	$\frac{2}{3}\left(\sqrt{\varepsilon_1} - \sqrt{\varepsilon_2}\right)^2$	$\varepsilon_1\varepsilon_2\left\{\frac{\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2}}{\sqrt{\varepsilon_1} - \sqrt{\varepsilon_2}}\ln\frac{\varepsilon_1}{\varepsilon_2} - 4\right\}$	$\frac{2}{3}\left(\sqrt{\varepsilon_1} - \sqrt{\varepsilon_2}\right)^2$
Single exponential	$\varepsilon_1 + \varepsilon_2 - \frac{2(\varepsilon_1 - \varepsilon_2)}{\ln\varepsilon_1/\varepsilon_2}$	$\left(\frac{\varepsilon_1 - \varepsilon_2}{\ln\varepsilon_1/\varepsilon_2}\right)^2 - \varepsilon_1\varepsilon_2$	$\varepsilon_1 + \varepsilon_2 - \frac{2(\varepsilon_1 - \varepsilon_2)}{\ln\varepsilon_1/\varepsilon_2}$
Double exponential	$\varepsilon_1\ln\frac{\varepsilon_1 + \varepsilon_2}{2\varepsilon_2} + \varepsilon_2\ln\frac{\varepsilon_1 + \varepsilon_2}{2\varepsilon_1}$	$2(\varepsilon_1 - \varepsilon_2)^2$	$2(\varepsilon_1 - \varepsilon_2)\left\{\ln\frac{\varepsilon_1}{\varepsilon_2} + \int_{(\varepsilon_2 - \varepsilon_1)/2\varepsilon_1}^{(\varepsilon_1 - \varepsilon_2)/2\varepsilon_2}\frac{\mathrm{d}x}{x}\ln(1+x)\right\}$
tanh	$(\varepsilon_1 - \varepsilon_2)\ln\frac{\varepsilon_1}{\varepsilon_2}$	$\frac{\pi^2}{3}(\varepsilon_1 - \varepsilon_2)^2$	$2(\varepsilon_1 - \varepsilon_2)^2\left\{\frac{\pi^2}{6}\left(\frac{1}{\varepsilon_1} + \frac{1}{\varepsilon_2}\right) + \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1\varepsilon_2}\int_0^1\frac{\mathrm{d}x}{1+x}\left[\ln\left(\frac{\varepsilon_1 x + \varepsilon_2}{\varepsilon_1 + \varepsilon_2 x}\right) + \frac{\left(\varepsilon_1^2 - \varepsilon_2^2\right)x\ln x}{(\varepsilon_1 x + \varepsilon_2)(\varepsilon_1 + \varepsilon_2 x)}\right]\right\}$

The first four interfaces have extent Δz, while for the double exponential and hyperbolic tangent profiles Δz is a measure of the interface thickness

Note that all the invariants are symmetric with respect to the interchange of ε_1 and ε_2, and that they are all non-negative when ε is positive and lies between ε_1 and ε_2 for all z. These properties are evident for I_1 in the form

$$I_1 = \int_{-\infty}^{\infty} \mathrm{d}z \frac{(\varepsilon_1 - \varepsilon)(\varepsilon - \varepsilon_2)}{\varepsilon}. \tag{3.73}$$

For i_2 and j_2 they may be deduced from the definitions (3.33) and (3.39). For example, the integrands in (3.33) and (3.39) always keep the same sign when ε is positive and $\min(\varepsilon_1, \varepsilon_2) \leq \varepsilon \leq \max(\varepsilon_1, \varepsilon_2)$.

References

Lekner J (1984) Invariant formulation of the reflection of long waves by interfaces. Physica 128A:229–252

Lekner J (1982a) Second-order ellipsometric coefficients. Physica 113A:506–520

Further Readings

The perturbation theory of Sect. 3.1 is based on

Lekner J (1982b) Reflection of long waves by interfaces. Physica 112A:544–556

The s wave Green's function for reflection at an inhomogeneity between like media (leading to (3.14)) was given by

Morse PM, Feshbach H (1953) Methods of theoretical physics. McGraw-Hill, New York, p 1071

Perturbation theory for reflection of the s wave at an interface between unlike media has previously been developed by

Triezenberg TG (1973) Capillary waves in a diffuse liquid-gas interface. Ph.D thesis, University of Maryland (unpublished)

There have been several (mainly formal) derivations of expansions for reflection properties in powers of the interface thickness (Sects. 3.2 and 3.4)

Maclaurin RC (1905) Theory of the reflection of light near the polarising angle. Proc Roy Soc A76:49–65

Rayleigh JWS (1912) On the propagation of waves through a stratified medium, with special reference to the question of reflection. Proc Roy Soc A86:207–266

Abelès F (1950) Recherches sur la propagation des ondes électromagnétiques sinusoïdales dans les milieux stratifiés. Application aux couches minces. Ann de Phys 5:596–640

Drazin PG (1963) On one-dimensional propagation of long waves. Proc Roy Soc A273:400–411

The formula (3.46) for the first order contribution to r_p/r_s is more than a century old. It is often attributed to Drude, but was, in essence, first obtained by L. Lorenz in 1860. Rayleigh (1912, quoted above) derives this result, and gives references to earlier work of Lorenz, Van Ryn, Drude, Schott and Maclaurin

Chapter 4
Variational Theory

In the previous chapter we have developed perturbation theories for dealing with reflection of electromagnetic or particle waves whose wavelength is long compared to the thickness of the interface. The variational theory of this chapter builds on these perturbation theories to provide formulae for the reflection amplitudes which have a greater range of validity, and which do not have the difficulties at grazing incidence such as the divergence of r^{pert}, as discussed in Sect. 3.5 in relation to (3.63) and (3.64). The application of variational theory to the short wave case is also discussed.

4.1 A Variational Expression for the Reflection Amplitude

We shall first give a general formulation of variational theory for reflection of waves described by the equation

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}z^2} + q^2\psi = 0, \quad \mathrm{e}^{iq_1 z} + r\mathrm{e}^{-iq_1 z} \leftarrow \psi \rightarrow t\mathrm{e}^{iq_2 z}, \tag{4.1}$$

and later specialize to the long wave and short wave s and p cases. Suppose that a perturbation theory has been constructed, based on the solution ψ_0 of

$$\frac{\mathrm{d}^2\psi_0}{\mathrm{d}z^2} + q_0^2\psi_0 = 0, \quad \mathrm{e}^{iq_1 z} + r_0\mathrm{e}^{-iq_1 z} \leftarrow \psi_0 \rightarrow t_0\mathrm{e}^{iq_2 z}. \tag{4.2}$$

In particular, a Green's function $G(z, \zeta)$ is assumed known, satisfying

$$\frac{\partial^2 G}{\partial z^2} + q_0^2 G = \delta(z - \zeta), \tag{4.3}$$

and giving the correct asymptotic forms in the solution of the integral equation

J. Lekner, *Theory of Reflection*, Springer Series on Atomic, Optical, and Plasma Physics 87, DOI 10.1007/978-3-319-23627-8_4

$$\psi(z) = \psi_0(z) - \int_{-\infty}^{\infty} \mathrm{d}\zeta\, \Delta q^2(\zeta) G(z,\zeta)\psi(\zeta) \tag{4.4}$$

($\Delta q^2 = q^2 - q_0^2$). Such Green's functions are given in Sects. 3.1 and 6.6 for the long and short wave cases. The results of perturbation theory follow by iteration of (4.4). Here we adapt Schwinger's variational method for the tangent of the phase shift produced by scattering off a central potential (Schwinger 1947; Blatt and Jackson 1949) to the reflection problem. We rewrite (4.4) as

$$\psi(z) + \int_{-\infty}^{\infty} \mathrm{d}\zeta\, \Delta q^2(\zeta) G(z,\zeta)\psi(\zeta) = \psi_0(z), \tag{4.5}$$

multiply by $\Delta q^2(z)\psi(z)$, and integrate over the whole range of z. The result is of the form

$$S = F \tag{4.6}$$

S, of second degree in the unknown ψ, is given by

$$S = \int_{-\infty}^{\infty} \mathrm{d}z\, \Delta q^2(z)\psi^2(z) + \int_{-\infty}^{\infty} \mathrm{d}z\, \Delta q^2(z)\psi(z) \int_{-\infty}^{\infty} \mathrm{d}\zeta\, \Delta q^2(\zeta) G(z,\zeta)\psi(\zeta). \tag{4.7}$$

F, of the first degree in ψ, is given by

$$F = \int_{-\infty}^{\infty} \mathrm{d}z\, \Delta q^2(z)\psi(z)\psi_0(z). \tag{4.8}$$

From the comparison identity (2.6), which we can write as

$$r = r_0 - \frac{1}{2iq_1} \int_{-\infty}^{\infty} \mathrm{d}z\, \Delta q^2(z)\psi(z)\psi_0(z), \tag{4.9}$$

we see that, for the exact ψ,

$$F = 2iq_1(r_0 - r). \tag{4.10}$$

For the exact ψ we also have $S = F$. Consider now a shift (in the variational sense) to a neighbouring function $\psi + \delta\psi$, where ψ is the exact solution. The integrals F and S shift by

$$\delta F = \int_{-\infty}^{\infty} \mathrm{d}z\, \delta\psi(z) \Delta q^2(z) \psi_0(z), \tag{4.11}$$

$$\delta S = 2 \int_{-\infty}^{\infty} \mathrm{d}z\, \delta\psi(z)\, \Delta q^2(z) \left\{ \psi(z) + \int_{-\infty}^{\infty} \mathrm{d}\zeta\, \Delta q^2(\zeta) G(z,\zeta) \psi(\zeta) \right\}. \tag{4.12}$$

From (4.5) the quantity in braces is equal to $\psi_0(z)$, and so $\delta S = 2\delta F$. But $S = F$, so $\delta S/S = 2\delta F/F$, or

$$\delta(F^2/S) = 0. \tag{4.13}$$

This is the variational principle: the correct ψ will extremize F^2/S. For an approximate ψ the extremal value of F^2/S approximates $F = 2iq_1(r_0 - r)$ and thus we have a variational estimate for the reflection amplitude:

$$r^{\mathrm{var}} = r_0 - F^2/2iq_1 S. \tag{4.14}$$

In general one has a parametrized trial function $\psi^{\mathrm{var}}(z)$, which when substituted for ψ gives the values F^{var} and S^{var}. The parameters which extremize $(F^{\mathrm{var}})^2/S^{\mathrm{var}}$ then give the best value (in the space spanned by the trial function) of the reflection amplitude. However, a useful variational estimate can be obtained without any variational parameters in ψ, provided ψ is well chosen.

The simplest choice is $\psi^{\mathrm{var}} = \psi_0$. Denote the corresponding values of F and S by F_0 and S_0, and the resulting variational estimate for r by $r_0 + r_1^{\mathrm{var}}$. Then

$$r_1^{\mathrm{var}} = -F_0^2/2iq_1 S_0. \tag{4.15}$$

From (4.7) and (4.8)

$$S_0 = F_0 + \int_{-\infty}^{\infty} \mathrm{d}z\, \Delta q^2(z) \psi_0(z) \int_{-\infty}^{\infty} \mathrm{d}\zeta\, \Delta q^2(\zeta) G(z,\zeta) \psi_0(\zeta), \tag{4.16}$$

while from (4.9) we see that

$$F_0 = -2iq_1 r_1, \tag{4.17}$$

where r_1 is the first order term in the perturbation series

$$r = r_0 + r_1 + r_2 + \cdots. \tag{4.18}$$

Thus

$$r_1^{\text{var}} = \frac{r_1 F_0}{S_0} = \frac{r_1}{1 + \frac{i}{2q_1 r_1}\int_{-\infty}^{\infty} \mathrm{d}z\, \Delta q^2(z)\psi_0(z) \int_{-\infty}^{\infty} \mathrm{d}\zeta\, \Delta q^2(\zeta) G(z,\zeta)\psi_0(\zeta)}. \quad (4.19)$$

We will apply this general expression to reflection of the electromagnetic s wave in the next two sections.

4.2 Variational Estimate for r_s in the Long Wave Case

The formulae of the previous section apply directly to the electromagnetic s wave, with $\psi = E$, $q^2 = \varepsilon\omega^2/c^2 - K^2$. In the long wave case the appropriate starting point for perturbation theory is the step profile

$$q_0^2(z) = \frac{1}{2}\left(q_1^2 + q_2^2\right) - \frac{1}{2}\left(q_1^2 - q_2^2\right)\mathrm{sgn}(z), \quad (4.20)$$

and the corresponding solution

$$E_0(z) = \begin{cases} \mathrm{e}^{iq_1 z} + r_0 \mathrm{e}^{-iq_1 z} & (z<0) \\ t_0 \mathrm{e}^{iq_2 z} & (z>0) \end{cases} \quad (4.21)$$

where $r_0 = (q_1 - q_2)/(q_1 + q_2)$ and $t_0 = 2q_1/(q_1 + q2) = 1 + r_0$. The Green's function for this case was given in Sect. 3.1 and (3.10). The first order term of perturbation theory is given by (3.13):

$$r_1 = \frac{i}{2q_1}\int_{-\infty}^{0} \mathrm{d}z\, \Delta q^2(z)\left(\mathrm{e}^{iq_1 z} + r_0 \mathrm{e}^{-iq_1 z}\right)^2 + \frac{i}{q_1 + q_2}(1 + r_0)\int_{0}^{\infty} \mathrm{d}z\, \Delta q^2(z)\mathrm{e}^{2iq_2 z}. \quad (4.22)$$

The variational estimate of r_s obtained from (4.19) on substituting (4.21), (4.22) and the Green's function of Sect. 3.1, has built into it two important general properties: it is correct to second order in the interfacial thickness/wavelength expansion, and it is correct at grazing incidence.

To verify the first statement, we express the expansion of r_s^{var} in terms the integrals

$$\lambda_n = \int_{-\infty}^{\infty} \mathrm{d}z[\varepsilon(z) - \varepsilon_0(z)]z^{n-1}, \quad (4.23)$$

defined in Sect. 3.2. We find

$$r_s^{\text{var}} = r_0 + \frac{2iq_1\omega^2/c^2}{(q_1+q_2)^2}\left\{\lambda_1 + 2iq_2\lambda_2 + \frac{i\omega^2/c^2}{q_1+q_2}\lambda_1^2 + \cdots\right\}. \tag{4.24}$$

This is correct to second order in the interface thickness, for all positioning of the reference profile $\varepsilon_0(z)$ (see (3.22)). In contrast, the expansion of r_1 is correct to this order only if the relative positioning of ε and ε_0 is such as to make λ_1 equal to zero. Such positioning is in general possible only when $\varepsilon_1 \neq \varepsilon_2$ and ε is real everywhere.

To verify that r_s^{var} correctly tends to -1 at grazing incidence (an exact general result, as shown in Sect. 2.3) we need to consider the $\varepsilon_1 \neq \varepsilon_2$ and $\varepsilon_1 = \varepsilon_2$ cases separately. When $\varepsilon_1 \neq \varepsilon_2$, r_1 and r_1^{var} become proportional to q_1 as $q_1 \to 0$, while $r_0 \to -1$ and so $r_s^{\text{var}} \to -1$. When $\varepsilon_1 = \varepsilon_2$, let q_0 be the common value of q_1 and q_2. As $q_0 \to 0, r_1 \to i(\omega/c)^2\lambda_1/2q_0$ and $S_0 - F_0 \to (\omega/c)^4\lambda_1^2/2iq_0$. From (4.19) we see that r_1^{var} correctly tends to -1 as q_0 tends to zero. This is in contrast to the *divergence* of the perturbation expression r_1 at grazing incidence, as can.be seen by setting $q_1 = q_2 = q_0$ in (4.22):

$$r_{1s} = \frac{i}{2q_0}\int_{-\infty}^{\infty} \mathrm{d}z\, \Delta q^2(z)\mathrm{e}^{2iq_0 z}. \tag{4.25}$$

4.3 Exact, Perturbation and Variational Results for the sech^2 Profile

The comparative accuracy of the perturbation and variational expressions for the reflectivity is conveniently demonstrated by the sech^2 profile

$$\varepsilon(z) = \varepsilon_0 + \Delta\varepsilon\,\mathrm{sech}^2\, z/a, \tag{4.26}$$

for which the exact, perturbation and variational results can all be found analytically for the s wave. The s wave equation reads

$$\frac{\mathrm{d}^2E}{\mathrm{d}z^2} + \left[q_0^2 + \Delta\varepsilon\frac{\omega^2}{c^2}\mathrm{sech}^2\, z/a\right]E = 0, \tag{4.27}$$

where

$$q_0^2 = \varepsilon_0\frac{\omega^2}{c^2} - K^2 = \varepsilon_0\frac{\omega^2}{c^2}\cos^2\theta \tag{4.28}$$

gives the common value at $z = \pm\infty$ of the wavenumber component q_0 perpendicular to the interface, and θ is the common value of the angles of incidence and transmission.

A solution of (4.27) in terms of the hypergeometric function (defined and discussed in Sect. 2.5) can be found by changing to the variable $\tau = \tanh z/a$ (Epstein 1930; Budden 1961, 1985; Landau and Lifshitz 1965). It is an imaginary power of $\operatorname{sech} z/a$ times a hypergeometric function,

$$E = \left(1-\tau^2\right)^{-\frac{i\beta}{2}} F(-s-i\beta, 1+s-i\beta; 1-i\beta; \frac{1}{2}(1-\tau)). \tag{4.29}$$

The dimensionless parameters in (4.29) are

$$\alpha = \Delta\varepsilon\left(\frac{\omega a}{c}\right)^2, \quad \beta = q_0 a, \quad s = \frac{1}{2}\left[(1+4\alpha)^{\frac{1}{2}}-1\right]. \tag{4.30}$$

As noted in Sect. 3.5, the problem is characterized by the two parameters α (or s) and β. As $z \to -\infty, 1-\tau \to 2\mathrm{e}^{2z/a}$, and E tends to $\mathrm{e}^{iq_0 z}$ times a constant $(2^{-i\beta})$, as required for the reflection problem. The limiting form of E as $z \to -\infty$ (and $\tau \to -1, 1+\tau \to 2\mathrm{e}^{2z/a}$) is found by using the relation between hypergeometric functions of ζ and $1-\zeta$,

$$\begin{aligned} F(a,b;c;\zeta) = {} & \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} F(a,b;1+a+b-c;1-\zeta) \\ & + \frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)} (1-\zeta)^{c-a-b} F(c-a,c-b;1+c-a-b;1-\zeta), \end{aligned} \tag{4.31}$$

which holds for $|\arg(1-\zeta)| < \pi$ (Oberhettinger 1964). With $\zeta = (1-\tau)/2$ and $1-\zeta = (1+\tau)/2 \to \mathrm{e}^{2z/a}$, this gives

$$2^{-i\beta}\Gamma(1-i\beta)\left\{\frac{\Gamma(-i\beta)}{\Gamma(-s-i\beta)\Gamma(1+s-i\beta)}\mathrm{e}^{iq_0 z} + \frac{\Gamma(i\beta)}{\Gamma(1+s)\Gamma(-s)}\mathrm{e}^{-iq_0 z}\right\} \leftarrow E. \tag{4.32}$$

The reflection amplitude r_s is the ratio of the coefficient of $\mathrm{e}^{-iq_0 z}$ to that of $\mathrm{e}^{iq_0 z}$, and is therefore given by

$$r_s = \frac{\Gamma(i\beta)\Gamma(-s-i\beta)\Gamma(1+s-i\beta)}{\Gamma(-i\beta)\Gamma(1+s)\Gamma(-s)} = \frac{\Gamma(i\beta)\Gamma(1+s-i\beta)\sin\pi s}{\Gamma(-i\beta)\Gamma(1+s+i\beta)\sin\pi(s+i\beta)}. \tag{4.33}$$

(The second form is obtained by using $\Gamma(z)\Gamma(1-z) = \pi/\sin\pi z$.) We see that $r_s \to -1$ at grazing incidence (Sect. 2.3), since as $\beta = q_0 a$ tends to zero, $\Gamma(i\beta)/\Gamma(-i\beta) \to -1$ (see (2.86)).

We note that s is an integer n when $\alpha = n(n+1)$. The reflection amplitude (4.33) is then zero, for any β (for all angles of incidence). This remarkable

reflection-less property is explored in Lekner (2007), and also in Sect. 19.2 in the quantum particle wavepacket context.

The transmission amplitude is the ratio of the coefficients of $e^{iq_0 z}$ in the limiting forms as $z \to \pm\infty$. We find

$$t_s = \frac{\Gamma(-s - i\beta)\Gamma(1 + s - i\beta)}{\Gamma(-i\beta)\Gamma(1 - i\beta)} = -\frac{\Gamma(\mathrm{i}\beta)\Gamma(1 + s - i\beta)\sin \pi i\beta}{\Gamma(-\mathrm{i}\beta)\Gamma(1 + \mathrm{s} + \mathrm{i}\beta)\sin \pi(s + i\beta)}. \tag{4.34}$$

When s is real ($\alpha \geq -1/4$),

$$|r_s|^2 = \frac{\cos^2\left[\pi/2(1 + 4\alpha)^{1/2}\right]}{\cos^2\left[\pi/2(1 + 4\alpha)^{1/2}\right] + \sinh^2 \pi\beta}, \tag{4.35}$$

and $|t_s|^2 = 1 - |r_s|^2$, which is a special case of the conservation law $q_1\left(1 - |r|^2\right) = q_2|t|^2$ of Sect. 2.1.

When $\alpha = \Delta\varepsilon(\omega a/c)^2 < -1/4$, s is complex:

$$s = -\frac{1}{2} + i\sigma, \quad \sigma = \frac{1}{2}(4|\alpha| - 1)^{\frac{1}{2}}. \tag{4.36}$$

On setting $z = \frac{1}{2} + iy$ in $\Gamma(z)\Gamma(1 - z) = \pi/\sin \pi z$ we obtain

$$\left|\Gamma\left(\frac{1}{2} + iy\right)\right|^2 = \frac{\pi}{\cosh \pi y}. \tag{4.37}$$

Using this in (4.33) and (4.34) gives

$$|r_s|^2 = \frac{\cosh^2 \pi\sigma}{\cosh^2 \pi\sigma + \sinh^2 \pi\beta}, \quad |t_s|^2 = \frac{\sinh^2 \pi\sigma}{\cosh^2 \pi\sigma + \sinh^2 \pi\beta}. \tag{4.38}$$

For the comparison with perturbation and variational theories we restrict ourselves to $\Delta\varepsilon > 0$ (and thus $\alpha > 0, s$ real). The reflectivity is then given by (4.35), with the corresponding perturbation and variational expressions found from (4.25) and (4.19):

$$|r_{1s}|^2 = (\pi\alpha/\sinh \pi\beta)^2, \tag{4.39}$$

$$\left|r_{1s}^{\mathrm{var}}\right|^2 = \frac{|r_{1s}|^2}{(1 + \alpha)^2 + (\alpha/\beta)^2}. \tag{4.40}$$

These expressions are compared in Figs. 4.1 and 4.2. We note that both the first order perturbation theory and the variational theory based on the free-space

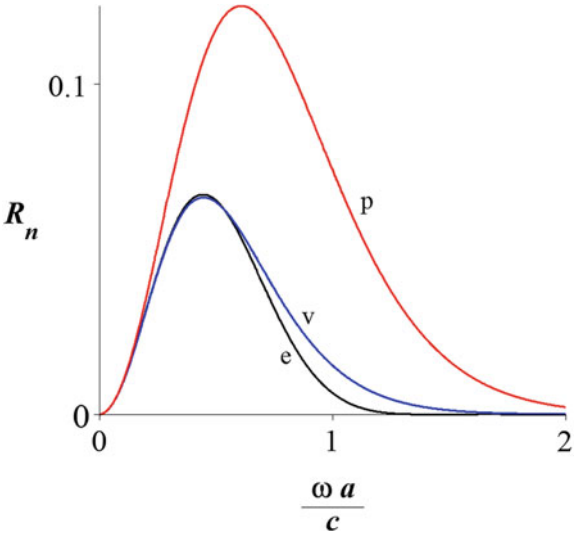


Fig. 4.1 Normal incidence reflectivity for the $\text{sech}^2 z/a$ profile, as a function of the profile thickness. The reflectivities are plotted for $\varepsilon_0 = 1, \Delta\varepsilon = 1$. For these parameters, the exact reflectivity is zero at $(\omega a/c)^2 = 2, 6, 12, \ldots n(n+1)$. The *curve* labelled *e* gives the exact reflectivity, (4.35); the *curve v* the variational estimate, (4.40); and the *curve p* the perturbation expression, (4.39)

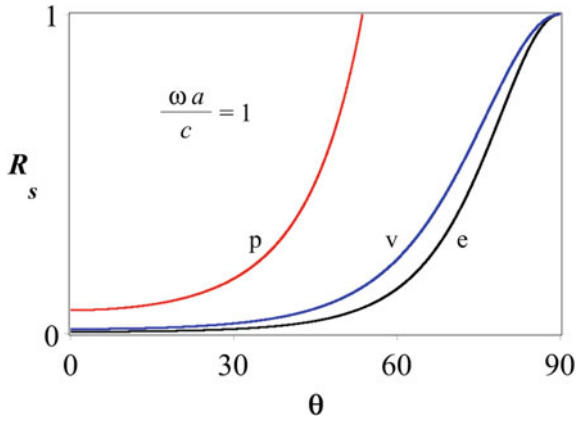


Fig. 4.2 Reflectivity as a function of the angle of incidence for the $\text{sech}^2 z/a$ profile. The exact, variational and perturbation results are denoted by *e*, *v* and *p*. The *curves* are drawn for $\varepsilon_0 = 1, \Delta\varepsilon = 1, \omega a/c = 1$

wavefunction are good at small values of $\omega a/c$. This is to be expected, since the reflection is then small. Neither theory gives the reflectivity zeros at $\alpha = \Delta\varepsilon(\omega a/c)^2 = n(n+1)$ (integer n), but the variational expression is closer to the exact values at all values of $\omega a/c$, and gives the correct unit reflectivity at glancing incidence.

4.4 Variational Theory for the p Wave

As in the s wave case, a variational estimate for the reflection amplitude builds on the corresponding perturbation theory. The long wave perturbation theory for the p wave was outlined in Sect. 3.4, and will be summarized here. The p wave has $\mathbf{B} = (0, \mathrm{e}^{iKx}B(z), 0)$, where K has the same meaning as for the s wave, and B satisfies

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{\varepsilon}\frac{\mathrm{d}B}{\mathrm{d}z}\right) + \left(\frac{\omega^2}{c^2} - \frac{K^2}{\varepsilon}\right)B = 0, \tag{4.41}$$

$$\mathrm{e}^{iq_1 z} - r_p \mathrm{e}^{-iq_1 z} \leftarrow B(z) \rightarrow \left(\frac{\varepsilon_2}{\varepsilon_1}\right)^{\frac{1}{2}} t_p \mathrm{e}^{iq_2 z}. \tag{4.42}$$

The required Green's function satisfies

$$\frac{\partial}{\partial z}\left(\frac{1}{\varepsilon_0}\frac{\partial G}{\partial z}\right) + \left(\frac{\omega^2}{c^2} - \frac{K^2}{\varepsilon_0}\right)G = \delta(z-\zeta), \tag{4.43}$$

where $\varepsilon_0(z)$ is the step function profile, and $G(z,\zeta)$ was given in (3.48). B satisfies the integro-differential equation

$$B_0(z) = B(z) - \int_{-\infty}^{\infty} \mathrm{d}\zeta\, \Delta v(\zeta)\left\{K^2 B(\zeta)G(z,\zeta) + \frac{\mathrm{d}B}{\mathrm{d}\zeta}\frac{\partial G}{\partial \zeta}\right\}, \tag{4.44}$$

In (4.44) $\Delta v = 1/\varepsilon - 1/\varepsilon_0$ and

$$B_0(z) = \begin{cases} \mathrm{e}^{iq_1 z} - r_{p0}\mathrm{e}^{-iq_1 z} & (z<0) \\ \left(\frac{\varepsilon_2}{\varepsilon_1}\right)^{\frac{1}{2}} t_{p0}\mathrm{e}^{iq_2 z} & (z>0) \end{cases} \tag{4.45}$$

In (4.45) the reflection and transmission amplitudes of the step profile $\varepsilon_0(z)$ are

$$-r_{p0} = \frac{Q_1 - Q_2}{Q_1 + Q_2}, \quad \left(\frac{\varepsilon_2}{\varepsilon_1}\right)^{\frac{1}{2}} t_{p0} = \frac{2Q_1}{Q_1 + Q_2}. \tag{4.46}$$

As usual, $Q_1 = q_1/\varepsilon_1$ and $Q_2 = q_2/\varepsilon_2$.

An exact expression for r_p is obtained from (4.44) by extracting the coefficient of $\mathrm{e}^{-iq_1 z}$ as $z \to -\infty$. This is

$$r_p = r_{p0} - \frac{1}{2iQ_1}\int_{-\infty}^{\infty} \mathrm{d}\zeta\, \Delta v(\zeta)\left\{K^2 B(\zeta)B_0(\zeta) + \frac{\mathrm{d}B}{\mathrm{d}\zeta}\frac{\mathrm{d}B_0}{\mathrm{d}\zeta}\right\}$$
$$= r_{p0} - \frac{1}{2iQ_1}\int_{-\infty}^{\infty} \mathrm{d}\zeta\left\{K^2 \Delta v B B_0 - \Delta\varepsilon C C_0\right\} \tag{4.47}$$

where $\Delta\varepsilon = \varepsilon - \varepsilon_0, C = \varepsilon^{-1}\mathrm{d}B/\mathrm{d}\zeta$ and $C_0 = \varepsilon_0^{-1}\mathrm{d}B_0/\mathrm{d}\zeta$. The first order perturbation theory expression for r_p is obtained by replacing B by B_0 and C by C_0 in (4.47). (This is equivalent to lowest order in Δv to replacing $\mathrm{d}B/\mathrm{d}\zeta$ by $\mathrm{d}B_0/\mathrm{d}\zeta$ but is preferable since C is continuous at a discontinuity in the dielectric function, as can be seen from (4.41), while $\mathrm{d}B/\mathrm{d}\zeta$ is not, and since the resulting reflection amplitude gives the correct first order term in the interface thickness/wavelength expansion, to all orders in Δv). Thus in the expansion $r_p = r_{p0} + r_1 + r_2 + \cdots$,

$$r_1 = -\frac{1}{2iQ_1}\int_{-\infty}^{\infty} \mathrm{d}\zeta\left\{K^2 \Delta v B_0^2 - \Delta\varepsilon C_0^2\right\}. \tag{4.48}$$

The variational expression for r_p is obtained by operating on (4.44) with

$$\int_{-\infty}^{\infty} \mathrm{d}z\left\{\Delta v K^2 B - \frac{\mathrm{d}}{\mathrm{d}z}\left(\Delta v \frac{\mathrm{d}B}{\mathrm{d}z}\right)\right\}. \tag{4.49}$$

As in the s wave case, the resulting equation can again be put in the form $F = S$, where the term of the first degree in B is

$$F = \int_{-\infty}^{\infty} \mathrm{d}z\, \Delta v\left\{K^2 B(z)B_0(z) + \frac{\mathrm{d}B}{\mathrm{d}z}\frac{\mathrm{d}B_0}{\mathrm{d}z}\right\} = 2iQ_1(r_{p0} - r_p) \tag{4.50}$$

(the second equality follows from (4.47)). The term of second degree in the unknown B is

$$S = \int_{-\infty}^{\infty} \mathrm{d}z\, \Delta v K^2 B\left\{B - \int_{-\infty}^{\infty} \mathrm{d}\zeta\, \Delta v\left[K^2 B G + \frac{\mathrm{d}B}{\mathrm{d}\zeta}\frac{\partial G}{\partial \zeta}\right]\right\}$$
$$+ \int_{-\infty}^{\infty} \mathrm{d}z\, \Delta v \frac{\mathrm{d}B}{\mathrm{d}z}\left\{\frac{\mathrm{d}B}{\mathrm{d}z} - \int_{-\infty}^{\infty} \mathrm{d}\zeta\, \Delta v\left[K^2 B\frac{\partial G}{\partial z} + \frac{\mathrm{d}B}{\mathrm{d}\zeta}\frac{\partial^2 G}{\partial z \partial \zeta}\right]\right\}. \tag{4.51}$$

We again find $\delta S = 2\delta F$ and hence the variational principle $\delta(F^2/S) = 0$. Thus the device of operating on the integral equation with (4.49) produces a variational estimate

$$r_p^{\text{var}} = r_{p0} - \frac{F^2/S}{2iQ_1}, \tag{4.52}$$

in parallel with the corresponding expression for the s wave, (4.14).

The simplest trial function for $B(z)$ is $B_0(z)$. This gives the values F_0 and S_0 for F and S, where

$$F_0 = \int_{-\infty}^{\infty} \mathrm{d}z\{\Delta\nu K^2 B_0^2 - \Delta\varepsilon C_0^2\} = -2iQ_1 r_1. \tag{4.53}$$

In the evaluation of S we must take care to include the $-\varepsilon_0(z)\delta(z-\zeta)$ singularity in $\partial^2 G/\partial z\partial\zeta$. We find, for general B,

$$\begin{aligned} S = &\int_{-\infty}^{\infty} dz\{\Delta\nu K^2 B^2 - \Delta\varepsilon C^2\} - K^4 \int_{-\infty}^{\infty} \mathrm{d}z\Delta\nu B \int_{-\infty}^{\infty} \mathrm{d}\zeta\Delta\nu BG \\ &+ 2K^2 \int_{-\infty}^{\infty} \mathrm{d}z\Delta\nu B \int_{-\infty}^{\infty} \mathrm{d}\zeta\Delta\varepsilon C\varepsilon_0^{-1}\partial G/\partial\zeta \\ &- \int_{-\infty}^{\infty} \mathrm{d}z\frac{\Delta\varepsilon}{\varepsilon_0}C \int_{-\infty}^{\infty} \mathrm{d}\zeta\frac{\Delta\varepsilon}{\varepsilon_0}C\left(\frac{\partial^2 G}{\partial z\partial\zeta}\right)_r, \end{aligned} \tag{4.54}$$

where

$$\left(\frac{\partial^2 G}{\partial z\partial\zeta}\right)_r = \frac{\partial^2 G}{\partial z\partial\zeta} + \varepsilon_0(z)\delta(z-\zeta) \tag{4.55}$$

is the regular part of $\partial^2 G/\partial z\partial\zeta$. The value of S_0 is found by replacing B with B_0 in (4.54), and the resulting variational estimate for the reflection amplitude has the form

$$r_p^{\text{var}} = r_{p0} + \frac{F_0}{S_0} r_1. \tag{4.56}$$

This expression gives a reflectivity which is correct to second order in the interface thickness/wavelength expansion, as may be shown by comparing the expansion of (4.56) with (3.50) (some reduction is required). Also built-in to the variational estimate (4.56) is the correct limiting value $r_p \to 1$ at grazing incidence

(Sect. 2.3). The cases of equal and unequal ε_1 and ε_2 must be considered separately. When $\varepsilon_1 = \varepsilon_2, r_{p0} = 0$ and $(F_0/S_0)r_1 \to 1$ as $Q_0 \to 0$, Q_0 being the common value of Q_1 and Q_2. When $\varepsilon_1 \neq \varepsilon_2, (F_0/S_0)r_1 \to 0$ as $Q_1 \to 0$, and $r_p^{\text{var}} \to 1$ because $r_{p0} \to 1$.

4.5 Reflection by a Layer Between Like Media

The variational formulae of Sects. 4.2 and 4.4 simplify considerably in the important special case of a reflecting layer between like media. As usual we set ε_0 equal to the common value of ε_1 and ε_2, and likewise for q_0; $\Delta\varepsilon$ stands for $\varepsilon - \varepsilon_0$, the deviation of the dielectric function from the ambient value. The variational theory based on long wave perturbation theory then gives r_s^{var} in terms of two integrals,

$$\lambda(k) = \int_{-\infty}^{\infty} \mathrm{d}z\, \mathrm{e}^{ikz} \Delta\varepsilon, \tag{4.57}$$

$$\sigma(k) = \int_{-\infty}^{\infty} \mathrm{d}z\, \Delta\varepsilon \left\{ \mathrm{e}^{ikz} \int_{-\infty}^{z} \mathrm{d}\zeta \Delta\varepsilon + \int_{z}^{\infty} \mathrm{d}\zeta \mathrm{e}^{ik\zeta} \Delta\varepsilon \right\}. \tag{4.58}$$

In terms of these integrals,

$$r_s^{\text{var}} = -\frac{\frac{\omega^2/c^2}{2iq_0}\lambda(2q_0)}{1 + \frac{\omega^2/c^2}{2iq_0}\frac{\sigma(2q_0)}{\lambda(2q_0)}}. \tag{4.59}$$

The analogous result for the p wave is

$$r_p^{\text{var}} = -\frac{\left[q_0^2\lambda(2q_0) - K^2\Lambda(2q_0)\right]/2iq_0}{1 + \frac{q_0^4\sigma(2q_0) - K^4\Sigma(2q_0) - 2q_0^2K^2\Gamma(2q_0)}{2iq_0\left[q_0^2\lambda(2q_0) - K^2\Lambda(2q_0)\right]}}, \tag{4.60}$$

where

$$\Lambda(k) = \varepsilon_0 \int_{-\infty}^{\infty} \mathrm{d}z\, \mathrm{e}^{ikz} \Delta\varepsilon/\varepsilon, \tag{4.61}$$

$$\Sigma(k) = \varepsilon_0^2 \int_{-\infty}^{\infty} \mathrm{d}z\, \Delta\varepsilon/\varepsilon \left\{ \mathrm{e}^{ikz} \int_{-\infty}^{z} \mathrm{d}\zeta\, \Delta\varepsilon/\varepsilon + \int_{z}^{\infty} \mathrm{d}\zeta\, \mathrm{e}^{ik\zeta} \Delta\varepsilon/\varepsilon \right\}, \tag{4.62}$$

$$\Gamma(k) = \varepsilon_0 \int_{-\infty}^{\infty} \mathrm{d}z\, \Delta\varepsilon/\varepsilon \left\{ \mathrm{e}^{ikz} \int_{-\infty}^{z} \mathrm{d}\zeta\, \Delta\varepsilon - \int_{z}^{\infty} \mathrm{d}\zeta\, \mathrm{e}^{ik\zeta} \Delta\varepsilon \right\}. \tag{4.63}$$

(In both the σ and Σ expressions, the first and second terms are equal because of the z, ζ symmetry of the integrands.) Of the five integrals, λ and Λ are Fourier transforms of $\Delta\varepsilon$ and $\varepsilon_0 \Delta\varepsilon/\varepsilon$, respectively, and have the dimension of length. The integrals σ, Σ and Γ have dimensions of length squared. At grazing incidence, when $q_0 \to 0$, the results $r_s^{\mathrm{var}} \to -1$ and $r_p^{\mathrm{var}} \to 1$ follow from

$$\sigma(0) = \lambda^2(0), \quad \Sigma(0) = \Lambda^2(0). \tag{4.64}$$

At normal incidence, when $K \to 0$ and $q_0 \to \sqrt{\varepsilon_0}\, \omega/c = k_0$, both reflection amplitudes tend to

$$r_n^{\mathrm{var}} = -\frac{\frac{k_0}{2i\varepsilon_0} \lambda(2k_0)}{1 + \frac{k_0}{2i\varepsilon_0} \frac{\sigma(2k_0)}{\lambda(2k_0)}}. \tag{4.65}$$

From (4.60) we see that a layer between like media does not reflect the p wave (according to both the first order perturbation and variational theories) at an angle

$$\theta = \arctan\{\lambda(2k_0)/\Lambda(2k_0)\}^{\frac{1}{2}}. \tag{4.66}$$

This is an approximate extension of the rigorous result obtained in Sect. 3.5, that, to lowest order in the interface thickness, there is zero reflection of the p wave at $\theta_0 = \arctan\{\lambda_1/\Lambda_1\}^{\frac{1}{2}}$. Note that the ratio of $\lambda(2k_0)$ to $\Lambda(2k_0)$ is not real in general. Zero reflection at a certain angle is thus characteristic of thin films; as we saw in Sect. 3.5, it also characterizes homogeneous films of any thickness.

We shall compare the variational and perturbation theories with exact results for the *homogeneous* layer with dielectric constant ε for $z_1 \le z \le z_2 = z_1 + \Delta z$; exact expressions for the reflection amplitudes are given in Sect. 2.4. In this case $\Delta\varepsilon$ and $\Delta\varepsilon/\varepsilon$ are both constant within the layer, and only the two integrals λ and Λ are required for the perturbation and variational reflection amplitudes, since

$$\Lambda(k) = \frac{\varepsilon_0}{\varepsilon} \lambda(k), \quad \Sigma(k) = \left(\frac{\varepsilon_0}{\varepsilon}\right)^2 \sigma(k), \quad \Gamma(k) = 0. \tag{4.67}$$

The expression (4.60) then reduces to

$$r_p^{\mathrm{var}} = -\frac{\frac{\omega/c}{2i\sqrt{\varepsilon_0}\cos\theta} \left\{\cos^2\theta - \frac{\varepsilon_0}{\varepsilon}\sin^2\theta\right\} \lambda(2q_0)}{1 + \frac{\omega/c}{2i\sqrt{\varepsilon_0}\cos\theta} \left\{\cos^2\theta + \frac{\varepsilon_0}{\varepsilon}\sin^2\theta\right\} \frac{\sigma(2q_0)}{\lambda(2q_0)}}, \tag{4.68}$$

correctly giving zero reflection at $\theta = \arctan(\varepsilon/\varepsilon_0)^{\frac{1}{2}}$. For the integrals λ and σ we find

$$\lambda(2q_0) = \Delta\varepsilon\Delta z e^{iq_0(z_1+z_2)} j_0(q_0\Delta z), \tag{4.69}$$

$$\sigma(2q_0) = (\Delta\varepsilon\Delta z)^2 e^{iq_0(z_1+z_2)}\{j_0(q_0\Delta z) + ij_1(q_0\Delta z)\}, \tag{4.70}$$

where $j_0(x) = x^{-1}\sin x$ and $j_1(x) = x^{-2}\sin x - x^{-1}\cos x$ are spherical Bessel functions.

In the figures below we compare the exact (e), perturbation (p) and variational (v) expressions for the reflectivity R_n at normal incidence as a function of the layer thickness (Fig. 4.3), R_s and R_p as a function of the angle of incidence (Figs. 4.4 and 4.5), and r_p/r_s in the complex plane as a function of the angle of incidence (Fig. 4.6). The comparison is made for the values $\varepsilon_0 = 1, \varepsilon = 2(\Delta\varepsilon = 1)$.

In the final figure we show the ellipsometric ratio, r_p/r_s. In the perturbation theory based on the free space wavefunction, $\frac{r_p}{r_s} \approx 1 - \left(\frac{cK}{\omega}\right)^2 \frac{\varepsilon+\varepsilon_0}{\varepsilon\varepsilon_0}$, a real quantity which starts at unity at normal incidence, and tends to $-\varepsilon_0/\varepsilon$ at glancing incidence. (The perturbation expressions for r_s and r_p are given in (4.13) and (4.37) of Lekner 1986, for a general inhomogeneous layer between like media. Equation (4.25) of this chapter gives r_{1s}.)

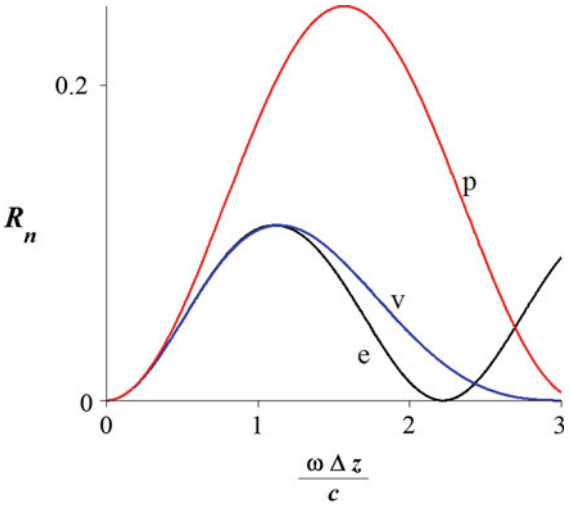


Fig. 4.3 Normal incidence reflectivity for a homogenous layer, as a function of the layer thickness Δz. The exact reflectivity is the labelled e, the variational result v, and the perturbation result p. In this and the following figures, $\varepsilon_0 = 1$ and $\varepsilon = 2$

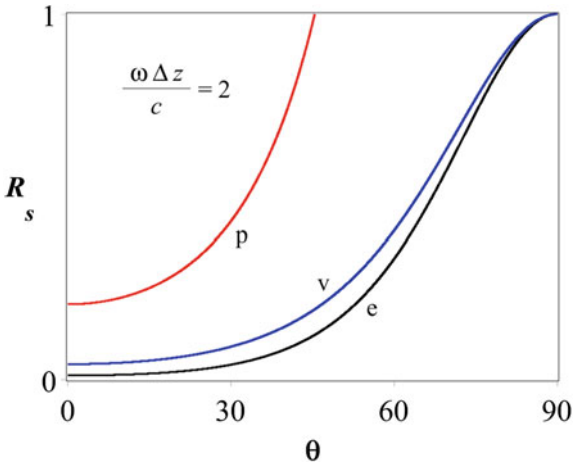


Fig. 4.4 Reflectivity for the *s* wave from a homogeneous layer, as a function of the angle of incidence, at $(\omega/c)\Delta z = 2$. The exact, perturbation, and variational results are denoted by *e*, *v* and *p* as before. Note that the perturbation theory, based (as is the variational theory) on the free space wavefunction, is in poor agreement with the exact reflectivity, and fails completely at large angles of incidence, where the approximate reflectivity exceeds unity

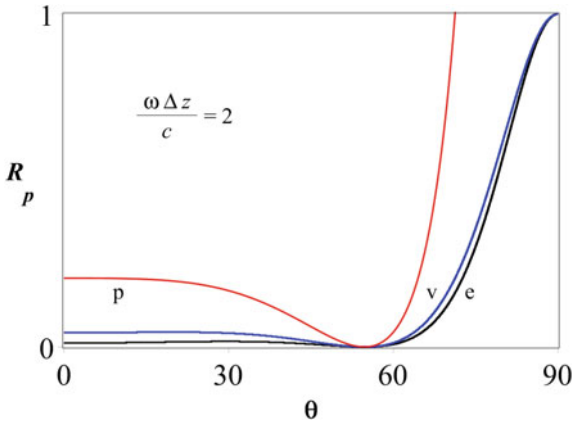


Fig. 4.5 Reflectivity for the *p* wave as a function of the angle of incidence, at $(\omega/c)\Delta z = 2$. The exact, perturbation and variational reflectivities are all zero at $\theta = \arctan\sqrt{2} \approx 54.7^\circ$. Again the perturbation theory fails completely near glancing incidence

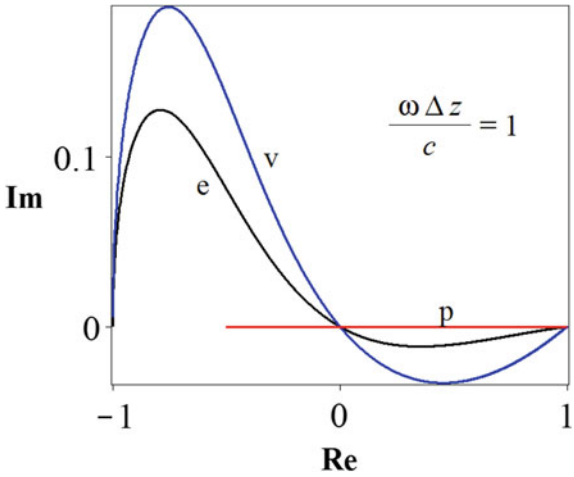


Fig. 4.6 The real and imaginary parts of the ratio r_p/r_s, plotted in in the complex plane, for $(\omega/c)\Delta z = 1$, as the incidence varies from normal to glancing. The exact and variational trajectories are the *curves* labelled *e* and *v*. The perturbation trajectory (*p*) lies along the real axis between $+1$ and $-\varepsilon_0/\varepsilon = -1/2$. All three trajectories start at $+1$ at normal incidence and pass through the origin at $\theta = \arctan\sqrt{\varepsilon/\varepsilon_0} = \arctan\sqrt{2} \approx 54.7°$. Only the perturbation trajectory does not end at -1 at glancing incidence

4.6 The Hulthén-Kohn Variational Method Applied to Reflection

We have seen that the adaptation of Schwinger's variational technique in scattering theory to reflection has led, with the simplest trial function, to *s* and *p* reflection amplitudes which are correct to second order to the interface thickness, and are correct at grazing incidence. These desirable features have been obtained at the cost of some complexity, and we shall now show how the simpler method developed by Hulthén (1948) and Kohn (1948) for scattering problems may be applied to reflection.

We begin with the *s* wave, for which the exact field amplitude E satisfies

$$\frac{\mathrm{d}^2E}{\mathrm{d}z^2} + q^2E = 0, \quad \mathrm{e}^{iq_1z} + r\mathrm{e}^{-iq_1z} \leftarrow E \rightarrow t\mathrm{e}^{iq_2z} \tag{4.71}$$

(we drop the subscript *s* except where needed to distinguish the *s* and *p* results). Consider the functional

$$\Phi[E_t] = \int_{-\infty}^{\infty} \mathrm{d}z\, E_t\left(\frac{\mathrm{d}^2E_t}{\mathrm{d}z^2} + q^2E_t\right) \tag{4.72}$$

of the trial function E_t, which we take to have the limiting forms

$$e^{iq_1 z} + r_t e^{-iq_1 z} \leftarrow E_t \rightarrow t_t e^{iq_2 z}. \tag{4.73}$$

The trial function differs from the exact E by $\delta E = E_t - E$, with

$$e^{iq_1 z} + \delta r e^{-iq_1 z} \leftarrow \delta E \rightarrow \delta t e^{iq_2 z}, \tag{4.74}$$

where $\delta r = r_t - r$ and $\delta t = t_t - t$. We find

$$\delta\Phi = \Phi[E + \delta E] - \Phi[E] = \left[E\frac{d\delta E}{dz} - \delta E\frac{dE}{dz}\right]_{-\infty}^{\infty} + O(\delta E)^2. \tag{4.75}$$

(This result follows on two integrations-by-parts, and the use of the fact that, from (4.71), $\Phi[E] = 0$.) From (4.71) to (4.75) we obtain the result $\delta\Phi = 2iq_1\delta r$, which can be written in the form of a variational principle:

$$\delta(\Phi - 2iq_1 r) = 0. \tag{4.76}$$

In the application of this principle, we use a trial E_t and the corresponding r_t and t_t, to evaluate $\Phi[E_t]$; then from (4.76)

$$\delta\Phi = \Phi[E_t] = 2iq_1(r_t - r) + O(\delta E)^2 \tag{4.77}$$

The variational estimate for the reflection amplitude is thus

$$r^{\text{var}} = r_t - \Phi[E_t]/2iq_1. \tag{4.78}$$

As an example, consider the simplest long-wave trial function $E_t = E_0$, the step profile solution given in (4.21). With $q^2 = q_0^2 + \Delta q^2$ as before,

$$\Phi[E_0] = \int_{-\infty}^{\infty} dz\, \Delta q^2 E_0^2, \tag{4.79}$$

which we recognise as the F_0 of the previous variational treatment. Thus the trial function E_0 leads to the first order perturbation result (compare (4.17))

$$r_s^{\text{var}} = r_{s0} - \frac{F_0}{2iq_1} = r_{s0} + r_1. \tag{4.80}$$

Similarly, in the short wave case the trial function $E_t = (q_1/q)^{1/2} e^{i\phi}$ produces the perturbation result (6.55):

$$r_s^{\text{var}} = \frac{1}{4i}\int_{-\infty}^{\infty} \mathrm{d}z \frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{\mathrm{d}q/\mathrm{d}z}{q^{\frac{3}{2}}}\right) q^{-\frac{1}{2}} \mathrm{e}^{2i\phi}. \tag{4.81}$$

The corresponding results for the p wave are not as satisfactory: if one defines the functional

$$\Phi[B_t] = \int_{-\infty}^{\infty} \mathrm{d}z\, B_t \left\{ \frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{\varepsilon}\frac{\mathrm{d}B_t}{\mathrm{d}z}\right) + \left(\frac{\omega^2}{c^2} - \frac{K^2}{\varepsilon}\right) B_t \right\} \tag{4.82}$$

of the trial function B_t, the variational principle takes the form

$$\delta\left(\Phi + 2iQ_1 r_p\right) = 0. \tag{4.83}$$

For the zeroth-order trial function B_0 defined in (4.45),

$$r_p^{\text{var}} = r_{p0} + \frac{\Phi[B_0]}{2iQ_1} = r_{p0} - \frac{1}{2iQ_1}\int_{-\infty}^{\infty} \mathrm{d}z\, \Delta v \left\{ K^2 B_0 + (\mathrm{d}B_0/dz)^2 \right\}. \tag{4.84}$$

This agrees with the perturbation result (4.48) only to lowest order in $\Delta v = 1/\varepsilon - 1/\varepsilon_0$. In consequence, (4.84) does not give the correct result to first order in the interface thickness (given by (3.44)), and does not agree with r_s^{var} at normal incidence.

The adaptation of the Hulthén-Kohn variational method to reflection problems is thus seen to give results which are inferior, for the simplest trial functions, to that obtained from adapting the Schwinger method. However, the greater simplicity of the Hulthén-Kohn method makes possible the use of more sophisticated trial functions (Joachain 1975, Chap. 10).

4.7 Variational Estimates in the Short Wave Case

We consider the s wave first. The variational theory is built on the perturbation theory of Sect. 6.5. The appropriate Green's function is given by (6.69):

$$2i[q(z)q(\zeta)]^{\frac{1}{2}} G(z,\zeta) = \begin{cases} \exp(i[\phi(\zeta) - \phi(z)]) & z < \zeta \\ \exp(i[\phi(z) - \phi(\zeta)]) & z > \zeta \end{cases} \tag{4.85}$$

where ϕ is the phase integral

$$\phi(z) = \int^{z} \mathrm{d}\zeta q(\zeta) \tag{4.86}$$

which is discussed in Sect. 6.2. The simplest variational expression uses

$$\psi_1^+ = [q_1/q]^{\frac{1}{2}} \mathrm{e}^{i\phi} \tag{4.87}$$

as the trial function. Setting $\psi_0 = \psi_1^+$ in (4.19) we obtain the variational estimate

$$r_s^{\mathrm{var}} = r_s^1 \Big/ \left\{ 1 + \frac{i}{2r_s^1} \int_{-\infty}^{\infty} \mathrm{d}z\, q^{-\frac{1}{2}} \Delta q^2 \mathrm{e}^{i\phi} \int_{-\infty}^{\infty} \mathrm{d}\zeta\, q^{-\frac{1}{2}} \Delta q^2 \mathrm{e}^{i\phi} G(z,\zeta) \right\}, \tag{4.88}$$

where the numerator is the short wave fist-order perturbation result,

$$r_s^1 = \frac{1}{4i} \int_{-\infty}^{\infty} \mathrm{d}z\, q^{-\frac{1}{2}} \frac{\mathrm{d}}{\mathrm{d}z} \left(\frac{\mathrm{d}q/\mathrm{d}z}{q^{\frac{3}{2}}} \right) \mathrm{e}^{2i\phi}. \tag{4.89}$$

On using (4.85) and the expression (6.73), namely

$$\Delta q^2 = -\frac{1}{2} q^{\frac{1}{2}} \frac{\mathrm{d}}{\mathrm{d}z} \left(\frac{\mathrm{d}q/\mathrm{d}z}{q^{\frac{3}{2}}} \right),$$

the double integral in (4.88) reduces to

$$\frac{1}{4i} \int_{-\infty}^{\infty} \mathrm{d}z \left(\frac{\mathrm{d}\gamma}{\mathrm{d}z} + \frac{1}{2} q\gamma^2 \right) \mathrm{e}^{2i\phi} \int_{-\infty}^{\infty} \mathrm{d}\zeta \left(\frac{\mathrm{d}\gamma}{\mathrm{d}\zeta} + \frac{1}{2} q\gamma^2 \right). \tag{4.90}$$

Here, as in Chap. 6, γ is the dimensionless function $q^{-2}\mathrm{d}q/\mathrm{d}z$, and we have used the fact that

$$q^{-\frac{1}{2}} \frac{\mathrm{d}}{\mathrm{d}z} \left(\frac{\mathrm{d}q/\mathrm{d}z}{q^{\frac{3}{2}}} \right) = \frac{\mathrm{d}\gamma}{\mathrm{d}z} + \frac{1}{2} q\gamma^2. \tag{4.91}$$

The evaluation of (4.90) involves a triple integration (unless the phase integral ϕ can be evaluated analytically), making it more difficult to apply than the long wave variational expression.

A variational theory for r_p in the short wave case may be derived along the same lines, since $b = (\varepsilon_1/\varepsilon)^{\frac{1}{2}}B$ satisfies an equation of the same form as E. The results are however so complex that they are unlikely to have practical value.

References

Blatt JM, Jackson JD (1949) On the interpretation of neutron-proton scattering data by the Schwinger variational method. Phys Rev 26:18–28

Budden KG (1961) Radio waves in the ionosphere. Cambridge University Press, Cambridge Section 17.6

Budden KG (1985) The propagation of radio waves. Cambridge University Press, Cambridge Section 15.17

Epstein PS (1930) Reflection of waves in an inhomogeneous absorbing medium. Proc Natl Acad Sci 16:627–637

Huthen L (1948) On the Sturm-Liouville problem connected with a continuous spectrum. Arkiv Mat Astr Fys 35A:14 paper 25

Joachain CJ (1975) Quantum collision theory. North Holland, Amsterdam

Kohn W (1948) Variational methods in nuclear collision problems. Phys Rev 74:1763–1772

Landau LD, Lifshitz EM (1965) Quantum mechanics. Pergamon Press, Oxford, p 79

Lekner J (1986) Reflection of light by a non-uniform film between like media. J Opt Soc Am A 3:9–15

Lekner J (2007) Reflectionless eigenstates of the $sech^2$ potential. Am J Phys 75:1151–1157

Oberhettinger F (1964) Hypergeometric functions. In: Abramowitz M, Stegun IA (eds) Handbook of mathematical functions, NBS Applied Mathematics Series No. 55 (Chap. 15). Dover, New York

Schwinger J (1947) A variational principle for scattering problems. Phys Rev 72:742

Further Readings

A general discussion of variational methods may be found in

Morse PM, Feshbach H (1953) Methods of theoretical physics. McGraw-Hill, New York Section 9.4

This chapter is based on three papers by the author

Lekner J (1985) Variational theory of reflection. Aust J Phys 38:113–123

Lekner J (1986a) Reflection of light by a non-uniform film between like media. J Opt Soc Am A 3:9–15

Lekner J (1986b) Variational theory of the reflection of light by interfaces. J Opt Soc Amer A 3:16–21

Chapter 5
Equations for the Reflection Amplitudes

For some purposes, both analytical and numerical, it is useful to transform the linear second order differential equation for the wave amplitude into a non-linear first order Riccati type differential equation for a quantity related to the reflection amplitude. The advantage lies in dealing directly with the quantity one wants to calculate. A disadvantage is the non-linearity of the resulting equations.

Early applications of this approach to the calculation of reflection amplitudes were by Walker and Wax (1946), Kofink (1947), Brekhovskikh (1949, 1980) and Schelkunoff (1951). This was preceded by development of related techniques in scattering theory, beginning with Morse and Allis (1933), and fully covered in Calogero's book (1967). Analogous methods were used by Courant and Hilbert (1953) to obtain the asymptotic forms of Bessel functions.

5.1 A First Order Non-linear Equation for an *s* Wave Reflection Coefficient

We first rewrite the second order differential equation for the electromagnetic s wave (and, equivalently, for particle waves) as a pair of coupled first order equations: the equation

$$E'' + q^2 E = 0 \tag{5.1}$$

is equivalent to the pair

$$E' = D, \quad D' = -q^2 E \tag{5.2}$$

(primes denote differentiation with respect to z). In turn, new functions F and G are introduced, defined by (5.2) and the equations

$$E = F + G, \quad D = iq(F - G). \tag{5.3}$$

J. Lekner, *Theory of Reflection*, Springer Series on Atomic, Optical, and Plasma Physics 87, DOI 10.1007/978-3-319-23627-8_5

We shall see shortly that F and G have the character of incident and reflected waves, tending to a constant times $\exp(\pm iq_1 z)$ as $z \to -\infty$. The ratio of these functions thus tends to r_s times $\exp(-2iq_1 z)$ as $z \to -\infty$, since r_s is defined as the ratio of the coefficient of $\exp(-iq_1 z)$ to that of $\exp(+iq_1 z)$ as $z \to -\infty$.

On substituting (5.3) into (5.2) we obtain a pair of coupled first order equations for F and G, which may be solved for F' and G' to give

$$F' = iqF - \frac{q'}{2q}(F - G), \tag{5.4}$$

$$G' = -iqG + \frac{q'}{2q}(F - G). \tag{5.5}$$

(When q is constant we see that F and G are proportional to $\exp(\pm iq_1 z)$). We now multiply (5.4) by G, (5.5) by F, subtract, and divide the result by F^2, obtaining an equation for $\rho = G/F$:

$$\rho' + 2iq\rho - \frac{q'}{2q}\left(1 - \rho^2\right) = 0. \tag{5.6}$$

The limiting forms of ρ as $z \to \pm\infty$ are

$$r_s \exp(-2iq_1 z) \leftarrow \rho \to 0. \tag{5.7}$$

The absolute square of ρ at $z \to -\infty$ gives the reflectivity:

$$|\rho(-\infty)|^2 = |r_s|^2 = R_s. \tag{5.8}$$

When q is real everywhere, the equation for the complex conjugate of ρ is

$$\rho^{*\prime} - 2iq\rho^* - \frac{q'}{2q}\left(1 - \rho^{*2}\right) = 0. \tag{5.9}$$

From (5.6) and (5.9) we may obtain an equation for the reflectivity function $R = \rho\rho^*$:

$$R' = \frac{q'}{q}\mathrm{Re}(\rho)(1 - R). \tag{5.10}$$

The boundary conditions on $R(z)$ are

$$R(\infty) = 0, \quad R(-\infty) = R_s. \tag{5.11}$$

On dividing both sides of (5.10) by $1 - R$, and integrating from $-\infty$ to $+\infty$, we obtain

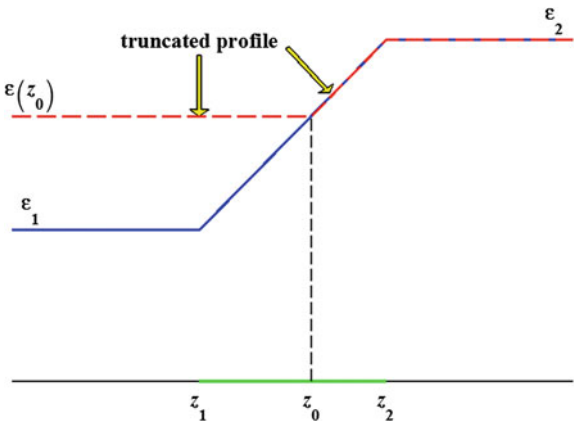


Fig. 5.1 The functions $\rho(z_0)$ and $R(z_0)$ correspond to a profile truncated at z_0, that is having $\varepsilon = \varepsilon(z)$ for $z \geq z_0$, and $\varepsilon = \varepsilon(z_0)$ for $z \leq z_0$. As z_0 increases from $-\infty$ the reflectivity $R(z_0)$ changes from the value R_s for the full profile to zero (the profile variation need not be monotonic; a linear profile is illustrated.)

$$\ln(1 - R_s) = \int_{-\infty}^{\infty} \mathrm{d}z \frac{q'}{q} \mathrm{Re}(\rho). \tag{5.12}$$

The right-hand side is real, and thus $R_s \leq 1$, as we proved first in Sect. 2.2. More generally, (5.10) divided by $1 - R$ can be integrated from z_0 to infinity, giving

$$\ln(1 - R(z_0)) = \int_{z_0}^{\infty} \mathrm{d}z \frac{q'}{q} \mathrm{Re}(\rho), \tag{5.13}$$

so that $R(z)$ is less than unity everywhere. This is in accord with the physical interpretation of $|\rho(z_0)|^2 = R(z_0)$, namely that of the reflectivity of a profile truncated at z_0, as shown in Fig. 5.1.

5.2 An Example: Reflection by the Linear Profile

We will illustrate the variation of R with z_0 for the linear profile of Fig. 5.1

$$\varepsilon(z) = \begin{cases} \varepsilon_1 & z \leq z_1 \\ \varepsilon_1 + (\Delta\varepsilon/\Delta z)(z - z_1) & z_1 \leq z \leq z_2 \\ \varepsilon_2 & z \geq z_2 \end{cases} \tag{5.14}$$

As usual, $\Delta\varepsilon = \varepsilon_2 - \varepsilon_1, \Delta z = z_2 - z_1$. At the same time we will give some of the properties of the solutions of (5.1) when ε is linear in z (the Airy functions), which will be useful later as solutions in the neighbourhood of classical turning points (Sects. 6.7 and 6.8). Within the interval (z_1, z_2) ε is linear in z. As an intermediate step we make ε the independent variable. Equation (5.1) transforms to

$$\frac{\mathrm{d}^2 E}{\mathrm{d}\varepsilon^2} + \left(\frac{\Delta z}{\Delta\varepsilon}\frac{\omega}{c}\right)^2 \left[\varepsilon - \left(\frac{cK}{\omega}\right)^2\right] E = 0. \tag{5.15}$$

We now transform to the variable $x = s\varepsilon$, where

$$s = \left(\left|\frac{\Delta z}{\Delta\varepsilon}\right| \frac{\omega}{c}\right)^{\frac{2}{3}}. \tag{5.16}$$

Equation (5.15) becomes

$$\frac{\mathrm{d}^2 E}{\mathrm{d}x^2} + \left[x - s\left(\frac{cK}{\omega}\right)^2\right] E = 0. \tag{5.17}$$

This equation has the general solution

$$E = \alpha Ai\left[s(cK/\omega)^2 - x\right] + \beta Bi\left[s(cK/\omega)^2 - x\right], \tag{5.18}$$

where Ai and Bi are Airy functions, the solutions of

$$\frac{\mathrm{d}^2 E}{\mathrm{d}\zeta^2} - \zeta E = 0. \tag{5.19}$$

This is known as the Airy differential equation (Heading 1962, Appendix A.3; Olver 2010). It has two power series solutions which are convergent for all ζ:

$$f(\zeta) = 1 + \frac{\zeta^3}{3!} + \frac{1.4\zeta^6}{6!} + \frac{1.4.7\zeta^9}{9!} + \cdots \tag{5.20}$$

$$g(\zeta) = \zeta + \frac{2\zeta^4}{4!} + \frac{2.5\zeta^7}{7!} + \frac{2.5.8\zeta^{10}}{10!} + \cdots \tag{5.21}$$

The standard pair of independent solutions are

$$Ai(\zeta) = c_1 f(\zeta) - c_2 g(\zeta), \quad Bi(\zeta) = \sqrt{3}[c_1 f(\zeta) + c_2 g(\zeta)], \tag{5.22}$$

where

$$\begin{aligned} c_1 &= Ai(0) = Bi(0)/\sqrt{3} = \left[3^{\frac{2}{3}}\Gamma\left(\frac{2}{3}\right)\right]^{-1} = 0.355028\ldots \\ c_2 &= -Ai'(0) = Bi'(0)/\sqrt{3} = \left[3^{\frac{1}{3}}\Gamma\left(\frac{1}{3}\right)\right]^{-1} = 0.258819\ldots \end{aligned} \tag{5.23}$$

From the solution (5.18) and the general formula (2.25) we can immediately obtain the s wave reflection amplitude, care being taken to convert between derivatives with respect to z and x via

$$\frac{\mathrm{d}E}{\mathrm{d}z} = \frac{\Delta\varepsilon}{\Delta z}\frac{\mathrm{d}E}{\mathrm{d}\varepsilon} = \frac{\Delta\varepsilon}{\Delta z}s\frac{\mathrm{d}E}{\mathrm{d}x}. \tag{5.24}$$

The result is

$$r_s = \mathrm{e}^{2iq_1z_1}\frac{q_1q_2(A_1B_2 - B_1A_2) + iq_1\left(A_1B_2' - B_1A_2'\right) + iq_2\left(A_1'B_2 - B_1'A_2\right) - \left(A_1'B_2' - B_1'A_2'\right)}{q_1q_2(A_1B_2 - B_1A_2) + iq_1\left(A_1B_2' - B_1A_2'\right) - iq_2\left(A_1'B_2 - B_1'A_2\right) + \left(A_1'B_2' - B_1'A_2'\right)} \tag{5.25}$$

where

$$\begin{aligned} A_1 &= Ai\left[s\left(\frac{cK}{\omega}\right)^2 - s\varepsilon_1\right] = Ai\left[-s\left(\frac{cq_1}{\omega}\right)^2\right], \\ A_2 &= Ai\left[s\left(\frac{cK}{\omega}\right)^2 - s\varepsilon_2\right] = Ai\left[-s\left(\frac{cq_2}{\omega}\right)^2\right], \end{aligned} \tag{5.26}$$

with similar definition of B_1 and B_2 in terms of Bi, and

$$A_1' = -\frac{\Delta\varepsilon}{\Delta z}sAi'\left[-s\left(\frac{cq_1}{\omega}\right)^2\right], \text{etc.} \tag{5.27}$$

The reflectivity at normal incidence is shown in Fig. 5.2. Here we are interested in the variation of $|\rho(z)|^2$ as z varies between z_1 and z_2. This may be obtained from the above by treating ε_1 as a variable. More instructive in the present context is a calculation of $\rho = G/F$, where the functions F and G are found in terms of the known E: from (5.3)

$$F = \frac{1}{2}\left[E + \frac{1}{iq}\frac{\mathrm{d}E}{\mathrm{d}z}\right], \quad G = \frac{1}{2}\left[E - \frac{1}{iq}\frac{\mathrm{d}E}{\mathrm{d}z}\right]. \tag{5.28}$$

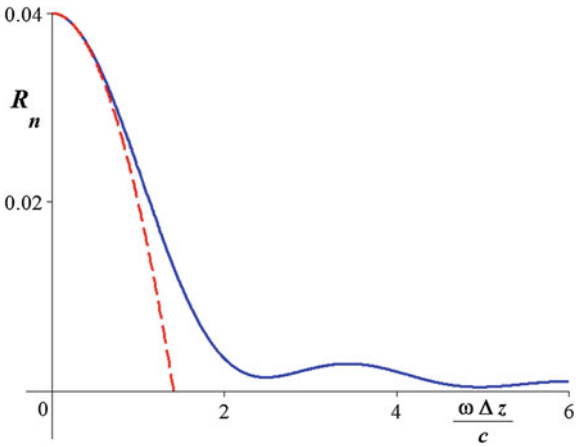


Fig. 5.2 Normal incidence reflectivity of the linear profile, as a function of the layer thickness. As in Fig. 5.1, $\varepsilon_1 = 1, \varepsilon_2 = \left(\frac{3}{2}\right)^2$, representing an inhomogeneous dielectric layer on glass. The *dashed curve* is the long-wave expression to second order in the interface thickness

The condition $G = 0$ at $z = z_2$ (or $\varepsilon = \varepsilon_2$) determines the ratio of the coefficients α and β in (5.18):

$$\frac{\alpha}{\beta} = -\frac{B_2 - B_2'/iq_2}{A_2 - A_2'/iq_2}. \tag{5.29}$$

The function $\rho(z)$ and the reflectivity function $R = \rho\rho^*$ are then obtained from

$$\rho(z) = \frac{G}{F} = \frac{iqE - \mathrm{d}E/\mathrm{d}z}{iqE + \mathrm{d}E/\mathrm{d}z}. \tag{5.30}$$

The results are equivalent to (5.25), with z_1, ε_1 being replaced by z, ε. Some reflectivity function curves are shown in Fig. 5.3.

5.3 Differential Equation for a *p* Wave Reflection Coefficient

The second order equation for the p wave,

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{\varepsilon}\frac{\mathrm{d}B}{\mathrm{d}z}\right) + \left(\frac{\omega^2}{c^2} - \frac{K^2}{\varepsilon}\right)B = 0, \tag{5.31}$$

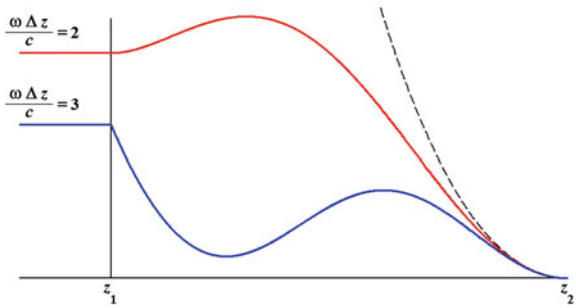


Fig. 5.3 Reflectivity function $R = \rho\rho^*$ for the linear profile of extent $\Delta z = z_2 - z_1$, with $\varepsilon_1 = 1$ and $\varepsilon_2 = (3/2)^2$. The *curves* are drawn for $(\omega/c)\Delta z = 2, 3$ at normal incidence. Also shown as the *dashed curve* is $\left[\left(\sqrt{\varepsilon} - \sqrt{\varepsilon_2}\right)/\left(\sqrt{\varepsilon} + \sqrt{\varepsilon_2}\right)\right]^2$, the reflectivity of a step from $\varepsilon(z)$ to ε_2, which all such curves approach as $z \to z_2$. For the linear profile, $\varepsilon(z) = \varepsilon_1 + (z - z_1)\Delta\varepsilon/\Delta z$

may be written as a pair of coupled first order equations

$$\frac{1}{\varepsilon}\frac{\mathrm{d}B}{\mathrm{d}z} = C, \quad \frac{\mathrm{d}C}{\mathrm{d}z} = -\frac{q^2}{\varepsilon}B. \tag{5.32}$$

We now write

$$B = F + G, \quad C = \frac{iq}{\varepsilon}(F - G). \tag{5.33}$$

On substitution into (5.32), with the usual definition of $Q = q/\varepsilon$, we find the companion relations to (5.4) and (5.5):

$$F' = iqF - \frac{Q'}{2Q}(F - G), \tag{5.34}$$

$$G' = -iqG + \frac{Q'}{2Q}(F - G). \tag{5.35}$$

Thus F and G again have the character of incident and reflected waves as $z \to -\infty$, being proportional to $\exp(\pm iq_1 z)$. The ratio $\rho = G/F$ now has the limiting forms

$$-r_p \exp(-2iq_1 z) \leftarrow \rho \to 0, \tag{5.36}$$

and satisfies the nonlinear first order equation

$$\rho' + 2iq\rho - \frac{Q'}{2Q}\left(1 - \rho^2\right) = 0. \tag{5.37}$$

When q is real everywhere, the p reflectivity function $R = \rho\rho^*$ satisfies

$$R' = \frac{Q'}{Q}\mathrm{Re}(\rho)(1-R), \tag{5.38}$$

with the boundary conditions

$$R(\infty) = 0, \quad R(-\infty) = R_p. \tag{5.39}$$

The fact that $R_p \leq 1$ follows from integrating (5.38) over the whole range of z:

$$\ln(1-R_p) = \int_{-\infty}^{\infty} \mathrm{d}z \frac{Q'}{Q}\mathrm{Re}(\rho). \tag{5.40}$$

This physically necessary upper bound of unity can be much improved, as we shall see in the next section.

5.4 Upper Bounds on R_s and on R_p

It is intuitively plausible that the s wave will reflect less from a monotonically increasing or decreasing profile than from a step profile with the same values of ε_1 and ε_2. This is in accord with the long wave result of Chap. 3,

$$R_s = R_{s0} - \frac{4q_1q_2(\omega/c)^4}{(q_1+q_2)^4} i_2 + \cdots, \tag{5.41}$$

where the invariant i_2 was shown to be non-negative if $\varepsilon(z)$ lies between ε_1 and ε_2. We will show now that a monotonic profile cannot reflect more of the s wave than the corresponding step profile, at any angle of incidence:

$$R_s \leq R_{s0} = \left(\frac{q_1-q_2}{q_1+q_2}\right)^2. \tag{5.42}$$

To prove this result we write $\rho = |\rho|\mathrm{e}^{i\theta}$ in (5.6), and obtain a pair of coupled equations for the modulus $|\rho|$ and the phase θ by separating the real and imaginary parts:

$$|\rho|' - \frac{q'}{2q}\left(1-|\rho|^2\right)\cos\theta = 0, \tag{5.43}$$

$$\theta' + 2q + \frac{q'}{2q}\left(|\rho|^{-1} + |\rho|\right)\sin\theta = 0. \tag{5.44}$$

We can rewrite (5.43) as

$$|\rho|'\left\{\frac{1}{1-|\rho|} + \frac{1}{1+|\rho|}\right\} = \frac{q'}{q}\cos\theta. \tag{5.45}$$

Integrating (5.45) from $-\infty$ to ∞ and using (5.8), we obtain

$$\ln\frac{1+|r_s|}{1-|r_s|} = -\int_{-\infty}^{\infty} \mathrm{d}z \frac{\mathrm{d}q}{q\mathrm{d}z}\cos\theta = -\int_{q_1}^{q_2} \mathrm{d}q\, q^{-1}\cos\theta. \tag{5.46}$$

This holds for any profile. Suppose now that $\varepsilon(z)$ increases monotonically from ε_1 to ε_2; the normal component of the wavevector then increases monotonically from q_1 to q_2, and the right-hand side of (5.46) has the upper bound $\ln(q_2/q_1)$. Thus

$$\frac{1+|r_s|}{1+|r_s|} \le \frac{q_2}{q_1}, \quad |r_s| \le \frac{q_2 - q_1}{q_2 + q_1}, \tag{5.47}$$

and (5.42) follows. The same bound on R_s holds for monotonic decrease from ε_1 to ε_2.

The corresponding result for the p wave reflectivity cannot be true without restriction, since we know that the reflectivity due to a sharp interface is zero at the Brewster angle, whereas an arbitrary interface has a principal angle (or angles) where $\mathrm{Re}(r_p) = 0$, with $\mathrm{Im}(r_p)$ nonzero in general (Sect. 2.3). Nevertheless a useful result can be obtained from the p wave equation corresponding to (5.43), namely

$$|\rho|' = \frac{Q'}{2Q}\left(1 - |\rho|^2\right). \tag{5.48}$$

On integrating this as before we find

$$\ln\frac{1+|r_p|}{1-|r_p|} = -\int_{-\infty}^{\infty} \mathrm{d}z \frac{\mathrm{d}Q}{Q\mathrm{d}z}\cos\theta = -\int_{Q_1}^{Q_2} \mathrm{d}Q\; Q^{-1}\cos\theta. \tag{5.49}$$

An upper bound of $\ln[\max(Q_1, Q_2)/\min(Q_1, Q_2)]$ again follows, *provided* $Q(z)$ is monotonic. Thus

$$R_p \le \left(\frac{Q_1 - Q_2}{Q_1 + Q_2}\right)^2, \quad Q \text{ monotonic} \tag{5.50}$$

Suppose $\varepsilon(z)$ is monotonic. Under what circumstances is then Q monotonic also? We have

$$Q^2 = \frac{q^2}{\varepsilon^2} = \frac{1}{\varepsilon}\left(\frac{\omega}{c}\right)^2 - \frac{K^2}{\varepsilon^2}, \tag{5.51}$$

and so

$$\frac{\mathrm{d}Q^2}{\mathrm{d}z} = -\frac{\mathrm{d}\varepsilon/\mathrm{d}z}{\varepsilon^3}\left[\varepsilon\left(\frac{\omega}{c}\right)^2 - 2K^2\right]. \tag{5.52}$$

Thus if $\mathrm{d}\varepsilon/\mathrm{d}z$ does not change sign, Q^2 will increase or decrease monotonically provided $\varepsilon - 2\varepsilon_1 \sin^2\theta_1$ does not change sign. This will be true if $2\varepsilon_1 \sin^2\theta_1 \le \varepsilon_1$ ($\theta_1 \le 45°$) and also if $2\varepsilon_1 \sin^2\theta_1 \ge \varepsilon_2$ (we have assumed $\varepsilon_1 \le \varepsilon \le \varepsilon_2$). Thus $R_p \le R_{p0}$ is guaranteed for monotonic ε at angles of incidence in the ranges

$$\sin^2\theta_1 \le \frac{1}{2}, \qquad \sin^2\theta_1 \ge \frac{\varepsilon_2}{2\varepsilon_1}. \tag{5.53}$$

Note that the Brewster angle given by $\sin^2\theta_B = \varepsilon_2/(\varepsilon_1 + \varepsilon_2)$ lies between these two limits. In the opposite case, when $\varepsilon_1 \ge \varepsilon \ge \varepsilon_2$, $R_p \le R_{p0}$ is guaranteed in the ranges

$$\sin^2\theta_1 \ge \frac{1}{2}, \quad \sin^2\theta_1 \le \frac{\varepsilon_2}{2\varepsilon_1}. \tag{5.54}$$

The Brewster angle again lies between these two limits. Figure 5.4 illustrates the reflectivity ratios R_s/R_{s0} and R_p/R_{p0} as a function of the angle of incidence, for the homogeneous layer for which r_p/r_s was displayed in Fig. 2.4.

Incidentally, the minimum of R_p/R_{p0} in Fig. 5.4 close to 48° is not a true zero, but results from the near coincidence of $Q^2 = Q_1 Q_2$ (at 47.93°) with the condition $q\Delta z = \left(m + \frac{1}{2}\right)\pi$, satisfied for $m = 9$ at 48.21°. Likewise, the dip close to 56° results from the near coincidence of $q\Delta z = m\pi$ (at 55.62°) with the Brewster angle condition for the substrate $Q_1 = Q_2$(at $\theta_B = \arctan\frac{3}{2} \approx 56.31°$). More details about extrema of the reflectances may be found in Sect. 2.4.

5.5 Long Wave Expansions

Systematic approximations based on the non-linear equation for ρ have been developed by Brekhovskikh (1949, 1980), and will be outlined here. We have seen that the physical meaning of $\rho(z)\exp(2iq(z)z)$ is that of the reflection amplitude of a profile truncated at z (Fig. 5.1). In the long wave limit this would be

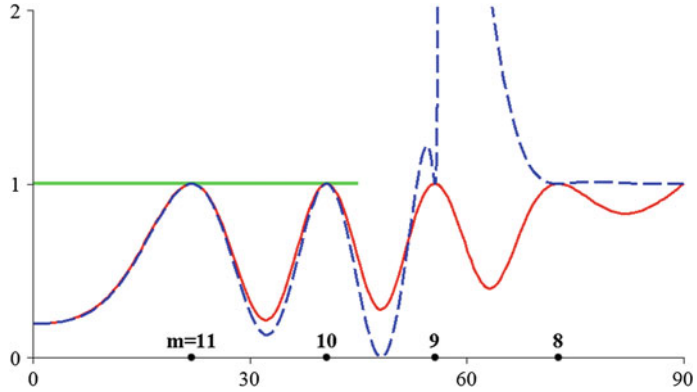


Fig. 5.4 The ratios R_s/R_{s0} and R_p/R_{p0} (*continuous* and *dashed curves*, respectively) for a homogeneous layer, with $\varepsilon_1 = 1, \varepsilon = (4/3)^2, \varepsilon_2 = (3/2)^2, (\omega/c)\Delta z = 27$, representing a layer of water (about four wavelengths thick) on glass. These parameters are the same as in Fig. 2.4. Unity is the upper bound for R_s/R_{s0}, at all angles of incidence, and for R_p/R_{p0} for $\theta_1 \leq 45°$. Unity is attained by both the s and p reflectivity ratios when $q\Delta z = m\pi$, (m integer) that is at angles of incidence given by $\sin^2\theta_1 = \{\varepsilon - [m\pi c/\omega\Delta z]^2\}/\varepsilon_1$. In this case the values $m = 11, 10, 9$ and 8 give the angles indicated

$$r(z) \approx \mathrm{e}^{2iq(z)z}\frac{q(z)-q_2}{q(z)+q_2}, \quad \rho(z) \approx \frac{q(z)-q_2}{q(z)+q_2}. \tag{5.55}$$

Brekhovskikh writes the exact $\rho(z)$ in terms of two functions $u(z), v(z)$, in analogy with (5.55):

$$\rho = \frac{qu - q_2 v}{qu + q_2 v}. \tag{5.56}$$

Then from (5.6) and (5.56) it follows that

$$\frac{u'}{u} - \frac{v'}{v} = iq\left\{\frac{q_2 v}{qu} - \frac{qu}{q_2 v}\right\}. \tag{5.57}$$

This equation is satisfied by

$$u' = iq_2 v, \quad v' = iq^2 u/q_2. \tag{5.58}$$

As $z \to z_2, \rho(z) \to 0$, it being assumed that $\varepsilon = \varepsilon_2$ for $z \geq z_2$. Thus the boundary conditions on u and v may be taken as

$$u(z_2) = 1 = v(z_2) \tag{5.59}$$

(any nonzero constant other than unity would do as well). The equations (5.58) and (5.59) are equivalent to

$$u(z) = 1 - iq_2 \int_z^{z_2} \mathrm{d}\zeta v(\zeta),\ v(z) = 1 - \frac{i}{q_2} \int_z^{z_2} \mathrm{d}\zeta q^2(\zeta) u(\zeta). \tag{5.60}$$

These coupled integral equations may be iterated to give $u = \sum u_n$ and $= \sum v_n$, starting with $u_0 = 1 = v_0$. The first iterates are

$$u_1(z) = -iq_2 \int_z^{z_2} \mathrm{d}\zeta = -iq_2(z_2 - z), \tag{5.61}$$

$$v_1(z) = -\frac{i}{q_2} \int_z^{z_2} \mathrm{d}\zeta q^2(\zeta). \tag{5.62}$$

The nth order is, for $n \geq 1$,

$$u_n(z) = -iq_2 \int_z^{z_2} \mathrm{d}\zeta v_{n-1}(\zeta),\ v_n(z) = -\frac{i}{q_2} \int_z^{z_2} \mathrm{d}\zeta q^2(\zeta) u_{n-1}(\zeta). \tag{5.63}$$

This iteration gives a series in powers of interface thickness/wavelength, and should thus duplicate the long wave results of Chap. 3. We will verify this to second order, for a profile of extent $\Delta z = z_2 - z_1$ (it is assumed that $\varepsilon = \varepsilon_1$ for $z \leq z_1$). For such a profile, the exact ρ is given by

$$\rho(z_1) = \frac{q_1 u(z_1) - q_2 v(z_1)}{q_1 u(z_1) + q_2 v(z_1)}. \tag{5.64}$$

To second order we have

$$u(z_1) = 1 - iq_2 \Delta z - \int_{z_1}^{z_2} \mathrm{d}z (z - z_1) q^2(z) + \cdots, \tag{5.65}$$

$$v(z_1) = 1 - \frac{i}{q_2} \int_{z_1}^{z_2} \mathrm{d}z\, q^2(z) - \int_{z_1}^{z_2} \mathrm{d}z (z_2 - z) q^2(z) + \cdots. \tag{5.66}$$

After some simplification, the reflectivity $R_s = |\rho(z_1)|^2$ reduces to

$$R_s = R_{s0} + \frac{4q_1q_2\omega^4/c^4}{(q_1+q_2)^4}\left[(\varepsilon_2-\varepsilon_1)\int_{z_1}^{z_2} \mathrm{d}z(2z-z_1-z_2)\varepsilon + \int_{z_1}^{z_2}\mathrm{d}z(\varepsilon-\varepsilon_1)\int_{z_1}^{z_2}\mathrm{d}\zeta(\varepsilon-\varepsilon_2)\right] + \cdots. \tag{5.67}$$

The quantity within the square bracket in (5.67) is independent of the angle of incidence, as it should be, since we know from Chap. 3 that the universal form for the s wave reflectivity is

$$R_s = R_{s0} - \frac{\frac{4q_1q_2\omega^4}{c^4}}{(q_1+q_2)^4} i_2 + \cdots. \tag{5.68}$$

Here i_2 is the second order invariant of Sect. 3.3:

$$i_2 = 2(\varepsilon_1-\varepsilon_2)\int_{-\infty}^{\infty}\mathrm{d}z(\varepsilon-\varepsilon_0)z - \left\{\int_{-\infty}^{\infty}\mathrm{d}z(\varepsilon-\varepsilon_0)\right\}^2, \tag{5.69}$$

where $\varepsilon_0(z)$ is the step function: $\varepsilon_0 = \varepsilon_1$ for $z<0$, $\varepsilon_0 = \varepsilon_2$ for $z>0$. The integrands in (5.67) do not go to zero at the end-points z_1 and z_2, and thus (5.67) appears to have no meaning as it stands for profiles which attain the limiting values ε_1 and ε_2 at infinity. To convert (5.67) to a universally applicable form we replace ε by $\varepsilon - \varepsilon_0 + \varepsilon_0$ in the integrands, and use the identities

$$(\varepsilon_2-\varepsilon_1)\int_{z_1}^{z_2}\mathrm{d}z(2z-z_1-z_2)\varepsilon_0 + \int_{z_1}^{z_2}\mathrm{d}z(\varepsilon_0-\varepsilon_1)\int_{z_1}^{z_2}\mathrm{d}\zeta(\varepsilon_0-\varepsilon_2) = 0, \tag{5.70}$$

$$\int_{z_1}^{z_2}\mathrm{d}z(2\varepsilon_0-\varepsilon_1-\varepsilon_2) = (\varepsilon_2-\varepsilon_1)(z_1+z_2). \tag{5.71}$$

The negative of the quantity within the square brackets of (5.67) then reduces to i_2 as given by (5.69), and we have regained the second order s wave result of Chap. 3 in its invariant form.

The p wave results are obtained similarly, but are much more complex. Both s and p thickness/wavelength expansions will be discussed again briefly in the chapter on matrix methods.

5.6 Differential Equations for the Reflection Amplitudes

In the preceding sections we have derived and used generalized Riccati equations for the quantity $\rho = G/F$, which takes the values $\pm r[\varepsilon(z), \varepsilon_2]\mathrm{e}^{-2iqz}$ at z for the s and p waves. Thus as $z \to -\infty$, ρ tends to $\pm r(\varepsilon_1, \varepsilon_2)$ times an oscillatory function of unit modulus. As we shall see here, it is sometimes advantageous to work directly with the reflection amplitude itself. This is particularly so in the short wave limit, which is discussed briefly here and forms the subject of the next chapter.

A non-linear first order equation for the s wave reflection amplitude may be obtained as follows: we set

$$E = f\mathrm{e}^{i\phi} + g\mathrm{e}^{-i\phi}, \quad D = iq(f\mathrm{e}^{i\phi} - g\mathrm{e}^{-i\phi}), \tag{5.72}$$

where f and g are functions determined from (5.2) and (5.72), and ϕ is the phase integral (discussed in detail in Chap. 6),

$$\phi(z) = \int^{z} \mathrm{d}\zeta q(\zeta). \tag{5.73}$$

In this instance it is convenient to choose the normally unspecified lower limit of integration in (5.73) so as to make $\phi(z) \to q_1 z$ as $z \to -\infty$. For example, if $\varepsilon(z) = \varepsilon_1$ for $z \leq z_1$, one can locate the origin at z_1 and set $\phi(z) = \int_0^z \mathrm{d}\zeta q(\zeta)$. We shall see shortly that f and g tend to constants as z tends to minus infinity; since r_s is defined as the ratio of the coefficient of $\mathrm{e}^{-iq_1 z}$ to that of $\mathrm{e}^{iq_1 z}$, for this choice for ϕ the ratio $r = g/f$ tends to r_s as $z \to -\infty$. On eliminating E and D from (5.2) and (5.72), using $\phi' = q$, we find

$$f' + \frac{q'}{2q}\left(f - g\mathrm{e}^{-2i\phi}\right) = 0, \tag{5.74}$$

$$g' + \frac{q'}{2q}\left(g - f\mathrm{e}^{2i\phi}\right) = 0. \tag{5.75}$$

Thus f and g are changing only where ε (and thus q) are changing, verifying that f and g tend to constants at $\pm\infty$. An equation for $r = g/f$ is obtained by multiplying (5.74) by g, (5.75) by f, subtracting, and dividing the result by f^2. It is

$$r' = \frac{q'}{2q}(\mathrm{e}^{2i\phi} - r^2\mathrm{e}^{-2i\phi}), \tag{5.76}$$

with limiting values of $r(z)$ as $z \to \pm\infty$ given by

$$r_s \leftarrow r \rightarrow 0. \tag{5.77}$$

We note that, in contrast to the equation (5.6) for ρ, variation of r occurs only where the dielectric function is varying.

The analogous equation for the p wave reflection amplitude is obtained by setting

$$B = f\mathrm{e}^{i\phi} + g\mathrm{e}^{-i\phi}, \quad C = iQ\left(f\mathrm{e}^{i\phi} - g\mathrm{e}^{-i\phi}\right). \tag{5.78}$$

The result of eliminating B and C from (5.32) and (5.78) is

$$f' + \frac{Q'}{2Q}\left(f - g\mathrm{e}^{-2i\phi}\right) = 0, \tag{5.79}$$

$$g' + \frac{Q'}{2Q}\left(g - f\mathrm{e}^{2i\phi}\right) = 0. \tag{5.80}$$

Thus the p wave reflection amplitude $r = -g/f$ satisfies

$$r' = -\frac{Q'}{2Q}(\mathrm{e}^{2i\phi} - r^2\mathrm{e}^{-2i\phi}), \tag{5.81}$$

with limiting values

$$r_p \leftarrow r \rightarrow 0. \tag{5.82}$$

At normal incidence $q'/q = \varepsilon'/2\varepsilon = -Q'/Q$, and so the equations for the s and p wave amplitudes are the same, as they must be. On integrating (5.76) and (5.81) from $-\infty$ to $+\infty$ we find

$$r_s = -\int_{-\infty}^{\infty} \mathrm{d}z \frac{q'}{2q}(\mathrm{e}^{2i\phi} - r^2\mathrm{e}^{-2i\phi}), \tag{5.83}$$

$$r_p = \int_{-\infty}^{\infty} \mathrm{d}z \frac{Q'}{2Q}(\mathrm{e}^{2i\phi} - r^2\mathrm{e}^{-2i\phi}). \tag{5.84}$$

These exact relationships lead naturally to the approximations of the next section.

5.7 Weak Reflection: The Rayleigh Approximation

We have seen in Sect. 5.1 that the meaning of $|\rho(z_0)|^2$ is that of the reflectivity of a profile truncated at z_0 (i.e., one which has $\varepsilon = \varepsilon(z_0)$ for $z \leq z_0$). The quantity $|r(z_0)|^2$ has the same meaning, and does not exceed unity (compare (5.13)). When the reflection is weak, as for a monotonic profile in the long wave limit with $|q_1 - q_2| \ll q_1 + q_2$, or for a smooth profile in the short wave limit (the absence of total internal reflection or regions of negative q^2 is assumed in both cases), it is reasonable to assume that $|r(z)|^2 \ll 1$ everywhere. Good approximations for r_s and r_p are then obtained by neglecting the terms in r^2 in (5.83) and (5.84):

$$r_s \approx r_s^R = -\int_{-\infty}^{\infty} \mathrm{d}z \frac{q'}{2q} \mathrm{e}^{2i\phi}, \tag{5.85}$$

$$r_p \approx r_p^R = \int_{-\infty}^{\infty} \mathrm{d}z \frac{Q'}{2Q} \mathrm{e}^{2i\phi}. \tag{5.86}$$

We have called these Rayleigh approximations, since they were first derived by Rayleigh (1912). They could also be called the weak reflection approximations, or associated with the names of Brekhovskikh (1949) or Bremmer (1951), who independently derived closely related approximations.

The physical basis of (for example) (5.85) can be seen by considering the profile as a series of small steps. As z changes by δz, the dielectric function changes by $\delta\varepsilon$ and the normal component of the wavenumber by δq. The contribution to the total reflection amplitude from this change is $\delta r = -(\delta q/2q)\mathrm{e}^{2i\phi}$, assuming that the reflection at all preceding steps is weak enough to be ignored. The contribution written down above follows from the single-step formula (1.15), namely

$$\delta r = \frac{q - (q + \delta q)}{q + (q + \delta q)} \mathrm{e}^{2i\phi},$$

with ϕ (the accumulated phase at z) replacing qz. Adding up the contributions δr gives, in the limit of a large number of small steps, the result (5.85).

The weak reflection approximations lead to, in the long wave limit, for a profile located near the origin,

$$r_s \approx \frac{1}{2} \ln \frac{q_1}{q_2}, \quad r_p \approx \frac{1}{2} \ln \frac{Q_2}{Q_1}. \tag{5.87}$$

These expressions are good representations of the exact limiting values (see Sect. 2.2)

$$r_{s0} = \frac{q_1 - q_2}{q_1 + q_2}, \quad r_{p0} = \frac{Q_2 - Q_1}{Q_2 + Q_1}, \tag{5.88}$$

provided these are small compared to unity; that is, provided $\delta = q_1/q_2 - 1$ and $\Delta = Q_2/Q_1 - 1$ are small. More precisely, the s wave reflection amplitudes (5.87) and (5.88) agree to second order in δ, both having the leading terms $\delta - \delta^2/4 + \cdots$. The Rayleigh expressions fail completely in the long wave limit when δ or Δ are not small: for sufficiently large or small values of the ratios q_1/q_2 and Q_2/Q_1 the expressions (5.87) will give reflectivities greater than 1. Since $|\ln x|/2$ is no smaller than $|(x-1)/(x+1)|$ for $x > 0$, the expressions (5.87) give reflectivities which are never less than the Fresnel values.

On comparing the exact expression (5.83) with the Rayleigh approximation (5.85), we see that

$$r_s = r_s^R + \int_{-\infty}^{\infty} \mathrm{d}z \frac{q'}{2q} r^2 \mathrm{e}^{-2i\phi} \equiv r_s^R + \Delta r_s. \tag{5.89}$$

Since $|r| \leq 1$ everywhere,

$$|\Delta r_s| \leq \int_{-\infty}^{\infty} \mathrm{d}z \left| \frac{q'}{2q} \right|. \tag{5.90}$$

When $\varepsilon(z)$ is monotonic,

$$|\Delta r_s| \leq \frac{1}{2} \left| \ln \frac{q_2}{q_1} \right|. \tag{5.91}$$

Similar results follow for the p wave, with Q replacing q in (5.90) and (5.91), the latter holding only if Q is monotonic. Thus simple bounds may be put on the error in the Rayleigh approximation. An example of the accuracy of r^R is given in the next section.

5.8 Iteration of the Integral Equation for r

The differential equation (5.76), together with the condition that $r \to 0$ as $z \to \infty$, may be integrated from z to ∞ to give

$$r(z) = -\int_z^\infty \frac{q'}{2q}(e^{2i\phi} - r^2 e^{-2i\phi}). \tag{5.92}$$

Iteration of this non-linear integral equation gives successive approximations for $r(z)$ and thus for $r_s = r(-\infty)$. If we label these functions $r^{(n)}(z)$, then

$$r^{(n+1)}(z) = -\int_z^\infty d\zeta \frac{q'}{2q}(e^{2i\phi} - \left[r^{(n)}(\zeta)\right]^2 e^{-2i\phi}). \tag{5.93}$$

The natural starting point for this iteration is $r^{(0)}(\zeta) = 0$, giving

$$r^{(1)}(z) = -\int_z^\infty d\zeta \frac{q'}{2q} e^{2i\phi}. \tag{5.94}$$

Thus $r_s^{(1)} = r_s^R$, the Rayleigh approximation.

As an example, we apply this method to the Rayleigh profile of Sect. 2.5, for which the reciprocal of the refractive index is linear in z in the interval (z_1, z_2):

$$\varepsilon^{-\frac{1}{2}}(z) \equiv \eta(z) = \eta_1 + (z - z_1)\Delta\eta/\Delta z, \tag{5.95}$$

where $\eta_1 = \varepsilon_1^{-\frac{1}{2}}, \eta_2 = \varepsilon_2^{-\frac{1}{2}}, \Delta\eta = \eta_2 - \eta_1, \Delta z = z_2 - z_1$. At normal incidence the phase integral is given by

$$\phi(z) = \frac{\omega}{c}\int_{z_1}^z dz\, \eta^{-1}(z) = \frac{\omega\Delta z}{c\Delta\eta}\ln\frac{\eta}{\eta_1}, \tag{5.96}$$

and $q'/q = \varepsilon'/2\varepsilon = -\eta'/\eta = -\Delta\eta/\eta\Delta z$. Thus

$$r^{(1)}(z) = \frac{1}{2}\int_\eta^{\eta_2} d\eta\, \eta^{-1}\left(\frac{\eta}{\eta_1}\right)^{2i\alpha}, \tag{5.97}$$

where α stands for the dimensionless parameter $(\omega/c)(\Delta z/\Delta\eta)$. The Rayleigh approximation to the normal incidence reflection amplitude is

$$r_n^R = r^{(1)}(z_1) = \frac{1}{4i\alpha}\left[(\eta_2/\eta_1)^{2i\alpha} - 1\right] = \left(\frac{\eta_2}{\eta_1}\right)^{i\alpha}\frac{\sin[\alpha\ln(\eta_2/\eta_1)]}{2\alpha}, \tag{5.98}$$

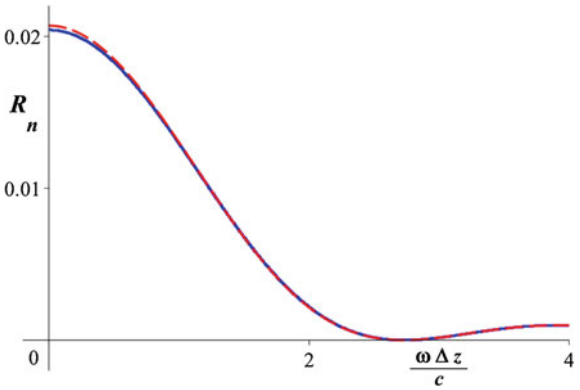


Fig. 5.5 Exact and weak reflection approximation reflectivities for the Rayleigh profile (*continuous* and *dashed curves*, respectively). The normal incidence reflectivities are shown, for $\varepsilon_1 = 1, \varepsilon_2 = (4/3)^2$. The exact and approximate curves have their first zero at $(\omega/c)\Delta z \approx 2.733$ and 2.730; successive exact and approximate zeros come progressively closer

and gives the reflectivity

$$R_n^R = \left(\frac{\sin[\alpha \ln(\eta_2/\eta_1)]}{2\alpha}\right)^2 = \left(\frac{\sin[\alpha/2 \ln(\varepsilon_2/\varepsilon_1)]}{2\alpha}\right)^2. \tag{5.99}$$

The exact reflectivity at normal incidence is given by the two formulae (2.107) and (2.108) according as α^2 is smaller or greater than 1/4. The exact and Rayleigh approximation normal incidence reflectivities are compared in Fig. 5.5 (Fig. 2.14 showed the reflectivity over a larger range of $\omega\Delta z/c$, on a logarithmic scale). The exact and Rayleigh approximation curves are not distinguishable on this scale for $(\omega/c)\Delta z$ greater than unity.

References

Brekhovskikh LM (1949) Reflection of plane waves from layered inhomogeneous media. Zhurnal Technicheskoi Fiziki 19:1126–1135

Brekhovskikh LM (1980) Waves in layered media, 2nd edn. Academic Press, New York

Bremmer H (1951) The W.K.B. approximation as the first term of a geometric-optical series. Commun Pure Appl Math 4:105–115

Calogero F (1967) Variable phase approach to potential scattering. Academic Press, New York

Courant R, Hilbert D (1953) Methods of mathematical physics, vol 1. Interscience Publishers, New York, p 332

Heading J (1962) An introduction to phase integral methods. Methuen, London

Kofink W (1947) Reflexion elektromagnetischer Wellen an einer inhomogenen Schicht. Ann Phys 1:119–132

Morse PM, Allis WP (1933) The effect of exchange on the scattering of electrons. Phys Rev 44:269–276

Olver FWJ (2010) Airy and related functions. In: Olver FWJ et al (eds) NIST Handbook of mathematical functions, Chap. 9, Cambridge University Press, Cambridge
Rayleigh JWS (1912) On the propagation of waves through a stratified medium, with special reference to the question of reflection. Proc R Soc 86:207–266
Schelkunoff SA (1951) Remarks concerning wave propagation in stratified media. Commun Pure Appl Math 4:117–128
Walker LR, Wax N (1946) Non-uniform transmission lines and reflection coefficients. J Appl Phys 17:1043–1045

Chapter 6
Reflection of Short Waves

Consider electromagnetic waves of angular frequency ω, incident on a planar inhomogeneity of thickness (or characterizing length) Δz. When $(\omega/c)\Delta z \gg 1$, and at normal incidence, there are many wavelengths within the inhomogeneity and (for smooth profiles) the change in the dielectric function within a wavelength is small. This is known as the short wave limit. We shall see that at general angle of incidence the short wave limiting forms are attained when $q\Delta z$ is large. Since $q_1 \to 0$ as $\theta_1 \to \pi/2$, the short wave approximations fail at grazing incidence. Special techniques are also needed when $q^2(z)$ passes through zero, and when $\varepsilon(z)$ has discontinuities in gradient.

6.1 Short Wave Limiting Forms for Some Solvable Profiles

It will be useful to look at some profiles for which the reflection amplitude is known analytically, both to orient ourselves and to have examples for comparison with the approximate expressions to be derived. We shall see that there is no *universal* form for the short wave expressions, in contrast to the long wave case of Chap. 3, where we showed that (for instance) the s wave reflectivity always takes the form

$$R_s = R_{s0} - \frac{4q_1q_2\omega^4/c^4}{(q_1+q_2)^4} i_2 + \cdots \tag{6.1}$$

There is greater variety and complexity in the short wave case, because short waves are sensitive to details in the dielectric function profile, while long waves are not.

Hyperbolic tangent profile: from (2.88) and (2.89) we have

$$r_s = \exp\left[2i\sum_{n=1}^{\infty} \arctan\left(\frac{2y_1}{n}\frac{y_1^2-y_2^2}{n^2+3y_1^2+y_2^2}\right)\right]\frac{\sinh\pi(y_1-y_2)}{\sinh\pi(y_1+y_2)}, \tag{6.2}$$

J. Lekner, *Theory of Reflection*, Springer Series on Atomic, Optical, and Plasma Physics 87, DOI 10.1007/978-3-319-23627-8_6

where $y_1 = q_1\Delta z$ and $y_2 = q_2\Delta z$. Suppose $\varepsilon_1 < \varepsilon_2$. Then $q_1^2 < q_2^2$, there is no total internal reflection, and (for large $y_2 - y_1$)

$$R_s = \mathrm{e}^{-4\pi q_1 \Delta z} + \cdots \tag{6.3}$$

This short wave limiting form, the long wave limiting form

$$R_s = R_{s0}\left(1 - \frac{4\pi^2}{3} q_1 q_2 (\Delta z)^2 + \cdots\right), \tag{6.4}$$

(obtained from (6.1) and the value $i_2 = (\pi^2/3)(\varepsilon_1 - \varepsilon_2)^2(\Delta z)^2$ from Table 3.1), and the exact reflectivity are compared in Fig. 6.1. The much more accurate Rayleigh approximation is compared, for the same profile, with the exact result in Fig. 6.3.

When $\varepsilon_1 > \varepsilon_2$, $q_1^2 > q_2^2$ and for $\theta_1 > \theta_c = \arcsin\,(\varepsilon_2/\varepsilon_1)^{1/2}$, $q_2 = i|q_2|$, with total internal reflection. The leading terms of (6.2) for large $y_1 = q_1\Delta z$ are thus unity for $\theta_1 > \theta_c$ and

$$R_s = \mathrm{e}^{-4\pi q_2 \Delta z} + \cdots \quad (\theta_1 < \theta_c) \tag{6.5}$$

Rayleigh profile: from (2.108) the reflectivity at normal incidence is

$$R_n = \frac{\sin^2\left[\frac{1}{2}|v|\ln(\varepsilon_1/\varepsilon_2)\right]}{4|v|^2 + \sin^2\left[\frac{1}{2}|v|\ln(\varepsilon_1/\varepsilon_2)\right]}, \tag{6.6}$$

where

$$|v| = \left[\left(\frac{\omega}{c}\frac{\Delta z}{\Delta\eta}\right)^{1/2} - \frac{1}{4}\right]^{1/2}, \quad \Delta\eta = \varepsilon_2^{-1/2} - \varepsilon_1^{-1/2}. \tag{6.7}$$

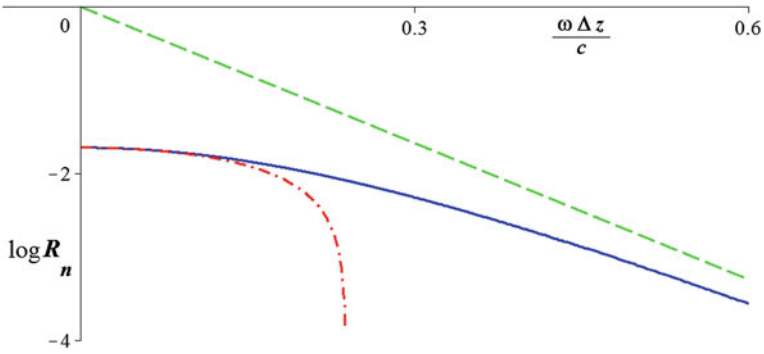


Fig. 6.1 Normal incidence reflectivity for the hyperbolic tangent profile. The logarithm to base 10 of the reflectivity is plotted. The *solid*, *dashed* and *dot-dashed curves* are respectively the exact, short wave and long wave forms. The *curves* are drawn for $\varepsilon_1 = 1$, $\varepsilon_2 = (4/3)^2$

(these forms apply for $((\omega/c)\Delta z)^2 > (\Delta\eta/2)^2$, which is the appropriate case in the short wave limit.) For this profile the $(\omega/c)\Delta z \gg 1$ limiting form shows oscillatory decay:

$$R_n \to \left\{ \frac{\sin\left(\frac{1}{2}\frac{\omega}{c}\frac{\Delta z}{\Delta\eta}\ln\frac{\varepsilon_1}{\varepsilon_2}\right)}{2\frac{\omega}{c}\frac{\Delta z}{\Delta\eta}} \right\}^2. \tag{6.8}$$

Note the power law decrease with increasing $(\omega/c)\Delta z$, in contrast to the exponential decay for the smoother hyperbolic tangent profile.

Exponential profile: the reflectivity at normal incidence is given by (2.102):

$$R_n = \frac{A_0^2 + B_0^2 + C_0^2 + D_0^2 - \frac{8}{\pi^2 u_1 u_2}}{A_0^2 + B_0^2 + C_0^2 + D_0^2 + \frac{8}{\pi^2 u_1 u_2}}. \tag{6.9}$$

When $(\omega/c)\Delta z \gg 1$, the variable $u = 2\sqrt{\varepsilon}(\omega/c)\Delta z/\ln(\varepsilon_2/\varepsilon_1)$ is large, and we can use Hankel's asymptotic expansions (Olver and Maximon 2010, Sect. 10.17)

$$\begin{aligned} J_0(u) &= \left(\frac{2}{\pi u}\right)^{1/2}(\alpha\cos w - \beta\sin w), \quad J_0'(u) = \left(\frac{2}{\pi u}\right)^{1/2}(-\gamma\sin w - \delta\cos w) \\ Y_0(u) &= \left(\frac{2}{\pi u}\right)^{1/2}(\alpha\sin w + \beta\cos w), \quad Y_0'(u) = \left(\frac{2}{\pi u}\right)^{1/2}(\gamma\cos w - \delta\sin w), \end{aligned} \tag{6.10}$$

where $w = u - \pi/4$. The functions α, β, γ and δ have asymptotic expansions which are conveniently expressed in terms of $v = (8u)^{-1}$:

$$\alpha \sim 1 - \frac{9}{2}v^2 + \cdots, \quad \beta \sim -v + \cdots, \quad \gamma \sim 1 + \frac{15}{2}v^2 + \cdots, \quad \delta \sim 3v + \cdots. \tag{6.11}$$

The cross products A_0 to D_0, defined in (2.97), are given by

$$\begin{aligned} A_0 &= \frac{2}{\pi(u_1u_2)^{1/2}}\{(\alpha_1\alpha_2 + \beta_1\beta_2)\sin(u_2 - u_1) + (\alpha_1\beta_2 - \beta_1\alpha_2)\cos(u_2 - u_1)\}, \\ B_0 &= \frac{2}{\pi(u_1u_2)^{1/2}}\{(\alpha_1\gamma_2 + \beta_1\delta_2)\cos(u_2 - u_1) + (\beta_1\gamma_2 - \alpha_1\delta_2)\sin(u_2 - u_1)\}, \\ C_0 &= \frac{2}{\pi(u_1u_2)^{1/2}}\{-(\gamma_1\alpha_2 + \delta_1\beta_2)\cos(u_2 - u_1) - (\delta_1\alpha_2 - \gamma_1\beta_2)\sin(u_2 - u_1)\}, \\ D_0 &= \frac{2}{\pi(u_1u_2)^{1/2}}\{\gamma_1\gamma_2 + \delta_1\delta_2)\sin(u_2 - u_1) + (\gamma_1\delta_2 - \delta_1\gamma_2)\cos(u_2 - u_1)\}. \end{aligned} \tag{6.12}$$

The leading terms of the asymptotic expansions give

$$A_0^2 + B_0^2 + C_0^2 + D_0^2 \sim \frac{8}{\pi^2 u_1 u_2}\left\{1 + 16\left[v_1^2 + v_2^2 - 2v_1 v_2 \cos 2(u_2 - u_1)\right]\right\}. \quad (6.13)$$

The short wave limiting form of R_n is thus

$$\begin{aligned} R_n &\sim 4\left[v_1^2 + v_2^2 - 2v_1 v_2 \cos 2(u_2 - u_1)\right] \\ &= \frac{1}{16}\left[\frac{1}{u_1^2} + \frac{1}{u_2^2} - \frac{2}{u_1 u_2}\cos 2(u_2 - u_1)\right]. \end{aligned} \quad (6.14)$$

Like the Rayleigh profile expression (6.8), this decreases as the inverse square of $(\omega/c)\Delta z$. The period of the oscillatory term (that is, the change in $(\omega/c)\Delta z$ during which the oscillatory term goes through one cycle) is, for large $(\omega/c)\Delta z$,

$$P_e = \frac{\frac{\pi}{2}\ln\frac{\varepsilon_2}{\varepsilon_1}}{\sqrt{\varepsilon_2} - \sqrt{\varepsilon_1}}, \quad (6.15)$$

while the Rayleigh profile reflectivity has the period

$$P_R = \frac{2\pi\left(\sqrt{\varepsilon_2} - \sqrt{\varepsilon_1}\right)}{(\varepsilon_1\varepsilon_2)^{1/2}\ln(\varepsilon_2/\varepsilon_1)}. \quad (6.16)$$

These expressions look dissimilar, but give similar values provided ε_1 and ε_2 are not too different. For example, when $\varepsilon_1 = 1$ and $\varepsilon_2 = (4/3)^2$, $P_e \simeq 2.71$ and $P_R \simeq 2.73$. For small $|\varepsilon_1 - \varepsilon_2|$ the two profiles are both approximately linear in z, and $P_e \simeq P_R \simeq \pi/(\varepsilon_1\varepsilon_2)^{1/4}$.

The short and long wave limiting forms for the exponential profile are compared with the exact reflectivity in Fig. 6.2.

The three examples discussed in this section are sufficient to show the greater variety and complexity in the short wave limiting forms compared to the long wave case, where the reflectivity is always the Fresnel value plus a term proportional to the square of the interfacial thickness. The long wave limit is simpler because the difference between the actual profile and a step is sensed only in an averaged or cumulative way, through integrals over the difference between the two profiles. When the wavelength is short compared to the profile extent, finer details are sensed. For example, discontinuities in slope in the Rayleigh and exponential profiles give a sinusoidal times inverse square dependence of the reflectivity on the parameter $(\omega/c)\Delta z$, in contrast to the exponential decrease with $(\omega/c)\Delta z$ for the smooth tanh profile.

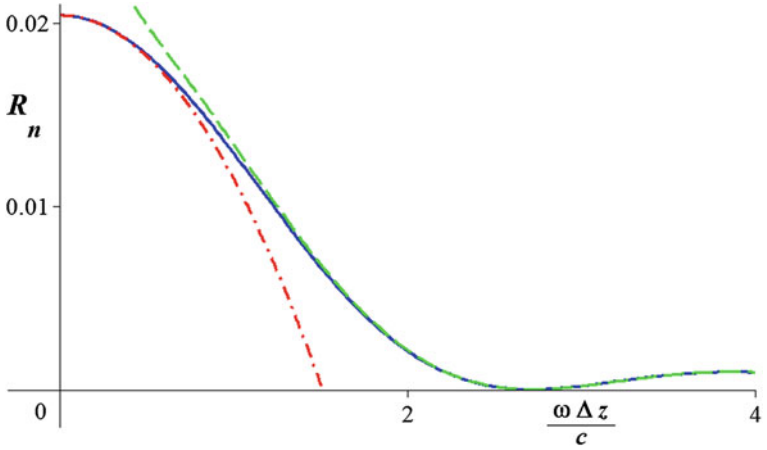


Fig. 6.2 Normal incidence reflectivity for the exponential profile. The *solid curve* is the exact reflectivity derived in Sect. 2.5, the *dashed curve* is the short wave approximation (6.3), and the *dot-dash curve* is the long wave approximation (6.4). The dielectric function values are $\varepsilon_1 = 1$ and $\varepsilon_2 = (4/3)^2$

6.2 Approximate High-Frequency Waveforms

We can consider together the electromagnetic *s* and *p* waves and particle waves, as in Sect. 2.1. The z variation in all three cases is given by solutions of equations of the form

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}z^2} + q^2\psi = 0. \tag{6.17}$$

We assume here that $q^2(z)$ is everywhere positive; solutions in the neighbourhood of zeros of q^2 will be discussed in Sects. 6.7 and 6.8.

When q is constant the propagating solutions are $e^{\pm iqz}$. When q varies we may expect solutions of the form $A(z)e^{\pm i\phi(z)}$, where the *phase integral* ϕ is given by

$$\phi(z) = \int^z \mathrm{d}\zeta q(\zeta). \tag{6.18}$$

The phase integral gives the accumulated phase at z, being the sum of phase differences equal to $(2\pi/\lambda) \times$ (path difference), where for motion in the z direction described by (6.17) the effective local wavelength λ is $2\pi/q$. Thus $q(\zeta)\mathrm{d}\zeta$ is the increase in the phase on going from ζ to $\zeta + \mathrm{d}\zeta$, and (6.18) gives the accumulated phase at z.

The x dependence for planar stratified media is contained in the factor e^{iKx}. Thus the total spatial phase is

$$\Phi(z,x) = \phi(z) + Kx. \tag{6.19}$$

Wave functions whose dominant z and x dependence is contained in the phase factor $\exp i\Phi$ correspond to geometrical optics rays, or to semiclassical particle trajectories. For, consider the wavefronts of $\exp i\Phi$. These are surfaces of equal phase, $\Phi(z, x)$ = constant. By definition, the rays are normal to the wavefronts, and will thus have the direction of

$$\nabla\Phi = \left(K, 0, \frac{\mathrm{d}\phi}{\mathrm{d}z}\right) = (K, 0, q(z)). \tag{6.20}$$

Thus the function $\exp i\Phi$ can be interpreted as a system of rays propagating in the z, x plane, with x and z components of the wavevector being K and $q(z)$. This is precisely the geometrical optics or semiclassical particle picture of propagation in a medium which is stratified with z.

A sequence of approximations for the solution of (6.17) may be obtained by setting $\psi = \exp\int^z \mathrm{d}\zeta\chi(\zeta)$. Then χ satisfies an equation of the generalized Riccati type:

$$\frac{\mathrm{d}\chi}{\mathrm{d}z} + \chi^2 + q^2 = 0 \tag{6.21}$$

By the short wave limit we mean that for an inhomogeneous region of extent Δz, the normal component of the wavevector q is large compared to $(\Delta z)^{-1}$. The physical meaning of $q\Delta z \gg 1$ is that the inhomogeneity extends over many wavelengths. In the zeroth approximation one neglects $\mathrm{d}\chi/\mathrm{d}z$ in (6.21), since (if χ varies smoothly) this is of order $(q\Delta z)^{-1}$ smaller in magnitude than the other two terms. Thus $\chi_0^{\pm} = \pm iq$, and

$$\psi_0^{\pm}(z) = \mathrm{e}^{\pm i\phi(z)}. \tag{6.22}$$

In the next approximation we set

$$\chi_1^{\pm} = \left(-q^2 - \frac{\mathrm{d}\chi_0^{\pm}}{\mathrm{d}z}\right)^{1/2} = \pm iq\left(1 \pm \frac{i\mathrm{d}z/\mathrm{d}z}{q^2}\right)^{1/2} \simeq \pm iq\left(1 \pm \frac{i\mathrm{d}q/\mathrm{d}z}{2q^2}\right), \tag{6.23}$$

on the assumption that the dimensionless quantity $\gamma = q^{-2}\mathrm{d}q/\mathrm{d}z$ is small. The corresponding wavefunctions $\psi_1 = \exp\int^z \mathrm{d}\zeta\chi_1$ are thus

$$\psi_1^+(z) = \left(\frac{q_1}{q(z)}\right)^{1/2} e^{i\phi(z)}, \quad \psi_1^-(z) = \left(\frac{q_2}{q(z)}\right)^{1/2} e^{-i\phi(z)} \tag{6.24}$$

(The square roots of q_1 and q_2 are inserted for later convenience.) It is instructive to compare the differential equations satisfied by the approximate waveforms $\psi_0^\pm$ and $\psi_1^\pm$ with the original wave (6.17). These are

$$\frac{d^2\psi_0^\pm}{dz^2} + \left(q^2 \mp \frac{i dq}{dz}\right)\psi_0^\pm = 0, \tag{6.25}$$

and

$$\frac{d^2\psi_1^\pm}{dz^2} + \left[q^2 + \frac{d^2q/dz^2}{2q} - \frac{3}{4}\left(\frac{dq/dz}{q}\right)^2\right]\psi_1^\pm = 0. \tag{6.26}$$

Note that ψ_0^+ and ψ_0^- satisfy different equations, while $\psi_1^\pm$ are two solutions of the same equation.

The approximate solutions $\psi_1^\pm$ go back to Liouville (1837) and Green (1837), with later contributions by Rayleigh (1912) and Gans (1915). A historical survey can be found in Heading (1962). The initials WKB or JWKB are often associated with $\psi_1^\pm$, the initials standing for Jeffreys (1924), Wentzel (1926), Kramers (1926), and Brillouin (1926). However, as Olver (1974) remarks, the contribution of these authors was not the construction of the approximation, which was already known, but the determination of connection formulae for linking exponential and oscillatory approximations across a zero of q^2 on the real axis. The latter problem is discussed in Sects. 6.7 and 6.8.

Exact solutions of (6.17) satisfy the flux conservation condition, the physical basis of which (conservation of energy in the electromagnetic case, and of probability density current in the particle case) was discussed in Sect. 2.1. The mathematical statement follows directly from (6.17) and its complex conjugate, which for real q^2 is

$$\frac{d^2\psi^*}{dz^2} + q^2\psi^* = 0. \tag{6.27}$$

We multiply (6.20) by ψ^* and (6.28) by ψ, and subtract. The result is

$$\frac{d}{dz}\left(\psi^*\frac{d\psi}{dz} - \psi\frac{d\psi^*}{dz}\right) = 0, \tag{6.28}$$

so that $\psi^*\frac{d\psi}{dz} - \psi\frac{d\psi^*}{dz} = 2i\,\mathrm{Im}\left(\psi^*\frac{d\psi}{dz}\right)$ is a constant. For reflection problems the asymptotic forms of ψ are

$$e^{iq_1 z} + re^{-iq_1 z} \leftarrow \psi \rightarrow te^{iq_2 z}, \tag{6.29}$$

and the fact that Im $(\psi^* d\psi/dz)$ is independent of z implies the flux conservation law

$$q_1\left(1 - |r|^2\right) = q_2|t|^2. \tag{6.30}$$

Any approximations to ψ which satisfy an equation of the form $\left(d^2\psi/dz^2\right) + \tilde{q}^2\psi = 0$ with real $\tilde{q}^2$ (as do $\psi_1^{\pm}$) will have $\mathrm{Im}(\psi^* d\psi/dz) = \text{constant}$, and thus conserve flux. The zeroth approximations $\psi_0^{\pm}$ on the other hand have

$$\mathrm{Im}\left(\psi_0^{\pm *} d\psi_0^{\pm}/dz\right) = \pm q(z), \tag{6.31}$$

which varies through the interface. This variation is associated with the absorption-type term $i dq/dz$ in (6.25). Nevertheless, we will find it instructive to examine the zeroth approximations obtained with $\psi_0^{\pm}$ for the reflectivity in the short wave limit.

6.3 Profiles of Finite Extent with Discontinuities in Slope at the Endpoints

Consider profiles for which q varies continuously in the interval $[z_1, z_2]$, and takes the values q_1 for $z \leq z_1$ and q_2 for $z \geq z_2$; and let $F(z)$ and $G(z)$ be two linearly independent solutions of (6.17) within $[z_1, z_2]$. We showed in Sect. 2.2 that the exact reflection amplitude is given by

$$r = e^{2iq_1 z_1} \frac{q_1 q_2 (F_1 G_2 - G_1 F_2) + iq_1 \left(F_1 G_2' - G_1 F_2'\right) + iq_2 \left(F_1' G_2 - G_1' F_2\right) - \left(F_1' G_2' - G_1' F_2'\right)}{q_1 q_2 (F_1 G_2 - G_1 F_2) + iq_1 \left(F_1 G_2' - G_1 F_2'\right) - iq_2 \left(F_1' G_2 - G_1' F_2\right) + \left(F_1' G_2' - G_1' F_2'\right)}. \tag{6.32}$$

In this section we consider the approximations $r^{(0)}$ and $r^{(1)}$ to r, obtained by setting F, G equal to ψ_0^+, ψ_0^- or ψ_1^+, ψ_1^- respectively. We assume $q^2(z)$ to be continuous at z_1 and z_2; this excludes (for example) the homogeneous layer (or two-step) profile.

Consider $r^{(0)}$ first, obtained by approximating F by ψ_0^+ and G by ψ_0^-. Then $F = e^{i\phi}$, $F' = iq\, e^{i\phi}$, $G = e^{-i\phi}$, $G' = -iq\, e^{-i\phi}$. On comparing (6.25) with (6.17), we see that such an approximation can be good in general only if $\gamma = q^{-2} dq/dz$ is small within the interface. If $q^2(z)$ varies smoothly within $[z_1, z_1 + \Delta z]$, γ is of order $(q\Delta z)^{-1}$. Thus when $q\Delta z$ is large, and the profile is smooth, $|r - r^{(0)}|$ is expected to be small. Let $\Delta\phi = \phi(z_2) - \phi(z_1) = \int_{z_1}^{z_2} dz\, q(z)$ be the phase difference between z_2 and z_1. Then

$$\begin{aligned} F_1G_2 - G_1F_2 &= -2i\sin\Delta\phi, \quad F_1G_2' - G_1F_2' = -2iq_2\cos\Delta\phi, \\ F_1'G_2 - G_1'F_2 &= 2iq_1\cos\Delta\phi, \quad F_1'G_2' - G_1'F_2' = -2iq_1q_2\sin\Delta\phi. \end{aligned} \tag{6.33}$$

On substitution of (6.33) into (6.32) to find $r^{(0)}$ we see that the zeroth order reflectivity is *zero*. This is the correct short wave limiting value in the absence of turning points (values of z where $q^2(z) = 0$, at which a classical particle or a ray described by geometrical optics would turn back).

The next approximation is obtained by substituting $F = \psi_1^+$ and $G = \psi_1^-$ into (6.32). Using $F' = q(i - \gamma/2)F$ and $G' = -q(i + \gamma/2)G$, where γ stands for the dimensionless quantity $q^{-2}\mathrm{d}q/\mathrm{d}z$, we find

$$\begin{aligned} F_1G_2 - G_1F_2 &= -2i\sin\Delta\phi \\ F_1G_2' - G_1F_2' &= iq_2(-2\cos\Delta\phi + \gamma_2\sin\Delta\phi) \\ F_1'G_2 - G_1'F_2 &= iq_1(2\cos\Delta\phi + \gamma_1\sin\Delta\phi) \\ F_1'G_2' - G_1'F_2' &= iq_1q_2[-2\sin\Delta\phi + (\gamma_1 - \gamma_2)\cos\Delta\phi - \frac{1}{2}\gamma_1\gamma_2\sin\Delta\phi]. \end{aligned} \tag{6.34}$$

On keeping first order terms in the small quantities γ_1, γ_2, we find

$$r^{(1)} = \frac{\mathrm{e}^{2iq_1z_1 + i\Delta\phi}}{4i}\{(\gamma_1 - \gamma_2)\cos\Delta\phi - i(\gamma_1 + \gamma_2)\sin\Delta\phi\}. \tag{6.35}$$

The corresponding reflectivity is

$$R^{(1)} = \left|r^{(1)}\right|^2 = \frac{1}{16}\left\{\gamma_1^2 + \gamma_2^2 - 2\gamma_1\gamma_2\cos 2\Delta\phi\right\}. \tag{6.36}$$

We can compare this formula with the short wave limiting forms obtained for the Rayleigh and exponential profiles in Sect. 6.1. These results were given for normal incidence, for which $q = \sqrt{\varepsilon}\omega/c$, and

$$\gamma(z) = q^{-2}\mathrm{d}q/\mathrm{d}z = (c/\omega)\varepsilon^{-1}\mathrm{d}\sqrt{\varepsilon}/\mathrm{d}z. \tag{6.37}$$

For the Rayleigh profile defined by (2.103), γ is independent of z at normal incidence:

$$\gamma = \frac{-\Delta\eta}{(\omega/c)\Delta z}, \qquad \Delta\eta = \varepsilon_2^{-1/2} - \varepsilon_1^{-1/2}, \tag{6.38}$$

and

$$\Delta\phi = \frac{1}{2}\frac{\omega}{c}\frac{\Delta z}{\Delta\eta}\ln\frac{\varepsilon_1}{\varepsilon_2}. \tag{6.39}$$

Thus (6.36) is in accord with (6.8).

For the exponential profile given by (2.93),

$$\gamma = \frac{\ln(\varepsilon_2/\varepsilon_1)}{2(\omega/c)\Delta z\sqrt{\varepsilon}} \equiv \frac{1}{u}, \tag{6.40}$$

and

$$\Delta\phi = \frac{2(\omega/c)\Delta z}{\ln(\varepsilon_2/\varepsilon_1)}(\sqrt{\varepsilon_2} - \sqrt{\varepsilon_1}) \equiv u_2 - u_1. \tag{6.41}$$

Again the expression (6.36), derived by approximating the solutions by $\psi_1^{\pm}$, is in agreement with the exact limiting form (6.14).

It is interesting that the factor $q^{-1/2}$, which transforms the approximate waveforms $\psi_0^{\pm}$ into the better approximations $\psi_1^{\pm}$, takes us from zero reflectivity to the useful expression (6.36). This can be better understood by looking at the total wavefunctions $\alpha F + \beta G$ within the inhomogeneous region, with F, G respectively given by ψ_0^+, ψ_0^- and ψ_1^+, ψ_1^-. With ψ given by

$$\psi(z) = \begin{cases} e^{iq_1 z} + re^{-iq_1 z} & (z < z_1) \\ \alpha F(z) + \beta G(z) & (z_1 \le z \le z_2) \\ te^{iq_2 z} & (z > z_2) \end{cases} \tag{6.42}$$

the continuity of ψ and $d\psi/dz$ at z_1 and z_2 gives us four linear equations in the four unknowns α, β, r, t. In Sect. 2.2 we found r and t (Equations 2.25 and 2.26). The corresponding formulae for α and β are

$$\alpha = 2iq_1 e^{iq_1 z_1}(G_2' - iq_2 G_2)/D, \qquad \beta = -2iq_1 e^{iq_1 z_1}(F_2' - iqF_2)/D, \tag{6.43}$$

where D is the denominator of (6.32). We see at once that $\beta = 0$ for $F = \psi_0^+ = e^{i\phi}$, since $F' = iqF$. Thus there is no backward propagating wave, consistent with zero reflection. The values of α, β when $F = \psi_0^+$ and $G = \psi_0^-$ are

$$\alpha_0 = e^{i(q_1 z_1 - \phi_1)}, \quad \beta_0 = 0. \tag{6.44}$$

The corresponding expressions obtained with $F = \psi_1^+$, and $G = \psi_1^-$, are, to first order in the small quantity γ,

$$\alpha_1 = \left(1 - \frac{i\gamma_1}{4}\right)e^{i(q_1 z_1 - \phi_1)}, \qquad \beta_1 = \left(\frac{q_1}{q_2}\right)^{1/2}\frac{i\gamma_2}{4}e^{i(q_1 z_1 - \phi_1 + 2\phi_2)}. \tag{6.45}$$

The coefficient of the backward propagating wave is now nonzero, and the coefficient of the forward propagating wave has a first order correction term in it.

The transmission amplitude for a profile of finite extent may be found from (2.26) and (6.34). The result is

$$t^{(1)} = \frac{4(q_1/q_2)^{1/2} e^{i(q_1 z_1 - q_2 z_2)}}{4\cos\Delta\phi + (\gamma_1 - \gamma_2)\sin\Delta\phi - i(4\sin\Delta\phi - (\gamma_1 - \gamma_2)\cos\Delta\phi + \frac{1}{2}\gamma_1\gamma_2\sin\Delta\phi)}. \tag{6.46}$$

Thus, to second order in γ_1 and γ_2,

$$\left|t^{(1)}\right|^2 = \frac{q_1}{q_2}\left[1 - \frac{1}{16}(\gamma_1^2 + \gamma_2^2 - 2\gamma_1\gamma_2\cos 2\Delta\phi)\right]. \tag{6.47}$$

On comparing with (6.36), we see that the conservation law $q_1(1 + |r|^2) = q_2|t|^2$ is obeyed to this order in γ.

Profiles with discontinuities at its boundaries in value as well as in slope can also be treated by this method (Lekner 1990).

6.4 Reflection Amplitude Estimates from a Comparison Identity

Let ψ be the solution of

$$\frac{d^2\psi}{dz^2} + q^2\psi = 0, \quad e^{iq_1 z} + r\,e^{-iq_1 z} \leftarrow \psi \rightarrow t\,e^{iq_2 z}, \tag{6.48}$$

and $\tilde{\psi}$ the solution of

$$\frac{d^2\tilde{\psi}}{dz^2} + \tilde{q}^2\tilde{\psi} = 0, \quad e^{iq_1 z} + \tilde{r}\,e^{-iq_1 z} \leftarrow \tilde{\psi} \rightarrow \tilde{t}\,e^{iq_2 z}. \tag{6.49}$$

The functions $q(z)$ and $\tilde{q}(z)$ have the same asymptotic values q_1 and q_2 at $+\infty$. We showed in Sect. 2.1 that r and $\tilde{r}$ are related by

$$r = \tilde{r} - \frac{1}{2iq_1}\int_{-\infty}^{\infty} dz\,(q^2 - \tilde{q}^2)\psi\tilde{\psi}. \tag{6.50}$$

We will use this identity to obtain approximations for r in the short wave limit. Consider first the result which comes from substituting $\tilde{\psi} = \psi_0^+$ into (6.50), where $\psi_0^+ = e^{i\phi}$ is the zeroth approximation to ψ in the short wave limit. Since ψ_0^+ satisfies (6.25), $\tilde{q}^2 = q^2 - i\,dq/dz$; also $\tilde{r} = 0$ since there is no backward propagating component in ψ_0^+, as can be verified by examining its limiting form as $z \rightarrow -\infty$. Thus we have the identity

$$r = -\frac{1}{2q_1}\int_{-\infty}^{\infty} \mathrm{d}z \frac{\mathrm{d}q}{\mathrm{d}z}\psi\psi_0^+ . \tag{6.51}$$

This is an exact relation, for all $q(z)$ and the corresponding ψ and ψ_0^+. Contributions to r come from regions where q is changing; since $q^2(z) = \varepsilon(z)\omega^2/c^2 - K^2$, these are the regions where the dielectric function is changing.

We saw in Sect. 6.2 that ψ_0^+ is a fair approximation to ψ provided the dimensionless quantity $\gamma = q^{-2}\mathrm{d}q/\mathrm{d}z$ is small. Assuming this is so, the expression

$$r^{(0)} = -\frac{1}{2q_1}\int_{-\infty}^{\infty} \mathrm{d}z \frac{\mathrm{d}q}{\mathrm{d}z}\left(\psi_0^+\right)^2 = -\frac{1}{2q_1}\int_{q_1}^{q_2} \mathrm{d}q\, \mathrm{e}^{2i\phi} \tag{6.52}$$

should be a fair approximation to r. We will not discuss the properties of $r^{(0)}$, since there is better theoretical basis for the similar approximation $r^{(1)}$, which is obtained from substituting $\tilde{\psi} = \psi_1^+ = (q_1/q)^{1/2}\mathrm{e}^{i\phi}$ into the comparison identity (6.50). (The factor $\sqrt{q_1}$ in the expression for ψ_0^+ gives the correct coefficient of $\mathrm{e}^{iq_1 z}$ in $\tilde{\psi}$ as $z \to -\infty$.) Since ψ_1^+ satisfies (6.26),

$$\tilde{q}^2 = q^2 + \frac{1}{2q}\frac{\mathrm{d}^2 q}{\mathrm{d}z^2} - \frac{3}{4q^2}\left(\frac{\mathrm{d}q}{\mathrm{d}z}\right)^2 = q^2 + \frac{1}{2}q^{1/2}\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{\mathrm{d}q/\mathrm{d}z}{q^{3/2}}\right), \tag{6.53}$$

and we obtain the identity

$$r = \left(4i\sqrt{q_1}\right)^{-1}\int_{-\infty}^{\infty} \mathrm{d}z \frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{\mathrm{d}q/\mathrm{d}z}{q^{3/2}}\right)\psi\mathrm{e}^{i\phi}. \tag{6.54}$$

On approximating ψ by ψ_1^+ we have another estimate for r, which can be put into several equivalent forms by integration by parts and a change of variable from z to ϕ:

$$\begin{aligned} r^{(1)} &= \frac{1}{4i}\int_{-\infty}^{\infty} \mathrm{d}z \frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{\mathrm{d}q/\mathrm{d}z}{q^{3/2}}\right)q^{-\frac{1}{2}}\mathrm{e}^{2i\phi} = -\frac{1}{2}\int_{-\infty}^{\infty} \mathrm{d}z \frac{\mathrm{d}q}{q\mathrm{d}z}\mathrm{e}^{2i\phi} + \frac{1}{8i}\int_{-\infty}^{\infty} \mathrm{d}z\left(\frac{\mathrm{d}q/\mathrm{d}z}{q^{3/2}}\right)^2\mathrm{e}^{2i\phi} \\ &= -\frac{1}{2}\int_{-\infty}^{\infty} \mathrm{d}\phi\left(\frac{\mathrm{d}q}{q\mathrm{d}\phi}\right)\mathrm{e}^{2i\phi} + \frac{1}{8i}\int_{-\infty}^{\infty} \mathrm{d}\phi\left(\frac{\mathrm{d}q}{q\mathrm{d}\phi}\right)^2\mathrm{e}^{2i\phi} \end{aligned} \tag{6.55}$$

We note that the part equal to $-\frac{1}{2}\int \mathrm{d}q\, q^{-1}\mathrm{e}^{2i\phi}$ is the Rayleigh (or weak reflection) approximation of Sect. 5.7, r_R. This is the dominant part of $r^{(1)}$: in the last line of (6.55) the dimensionless quantity $q^{-1}\mathrm{d}q/\mathrm{d}\phi$ is precisely the function $\gamma = q^{-2}\mathrm{d}q/\mathrm{d}z$ which has

to be small for ψ_0 and ψ_1 to approximate ψ well. The last term in (6.55) could thus be omitted in the calculation of the reflection amplitude to lowest order in this quantity. However, the expression (6.55) came directly from the identity (6.54), and will be seen to follow from first order perturbation theory in the next section, so there are reasons both for retaining and for dropping the last term in (6.55).

In the short wave limit the factor $e^{2i\phi}$ oscillates rapidly with z, and a smooth variation of q with z will lead to exponentially small values of the integral. If however q has discontinuities in any of its derivatives, the main contributions to the integral come from the neighbourhood of the discontinuities. We will give examples of these possibilities.

Consider the case of a medium which is inhomogeneous only within the interval $[z_1, z_2]$, and which has discontinuities in the slope or higher derivatives of ε only at the end points z_1 and z_2. The exponential and Rayleigh profiles are in this category. In the short wave limit the factor $e^{2i\phi}$ oscillates rapidly with z, averaging to an exponentially small value any slowly varying contribution in the integrals defining $r^{(1)}$. There will be a contribution in (6.55) arising from discontinuities in slope: we write (6.55) in the form

$$r^{(1)} = \frac{1}{4i}\int_{-\infty}^{\infty} \mathrm{d}z \left\{ \frac{\mathrm{d}^2 q/\mathrm{d}z^2}{q^2} - \frac{3}{2}\frac{(\mathrm{d}q/\mathrm{d}z)^2}{q^3} \right\} e^{2i\phi}, \tag{6.56}$$

and note that for a profile in which the dimensionless function $\gamma = q^{-2}\mathrm{d}q/\mathrm{d}z$ changes from 0 to γ_1 at z_1, and from γ_2 to 0 at z_2, the second derivative term in (6.56) has the delta function singularities $\gamma_1\delta(z - z_1) - \gamma_2\delta(z - z_2)$. Thus (6.56) takes the value

$$r^{(1)} = \frac{1}{4i}\left\{\gamma_1 e^{2i\phi_1} - \gamma_2 e^{2i\phi_2}\right\} = \frac{e^{i(\phi_1+\phi_2)}}{4i}\left\{\gamma_1 e^{-i\Delta\phi} - \gamma_2 e^{i\Delta\phi}\right\}, \tag{6.57}$$

plus exponentially small terms from the smooth part of the profile. This expression gives the same reflectivity as the theory of Sect. 6.3, (Equation 6.36). It is in agreement with (6.35) in phase as well as in absolute magnitude if $\phi_1 = \phi(z_1) = q_1 z_1$. This condition is in fact a requirement arising in the setting up of the comparison identity (6.54), in which the asymptotic form $e^{iq_1 z}$ was assumed for ψ_1^+ in the limit as $z \to -\infty$. This requirement fixes the phase function to be

$$\phi(z) = q_1 z_1 + \int_{z_1}^{z} \mathrm{d}\zeta q(\zeta). \tag{6.58}$$

In the case of a smooth profile, with all of its derivatives continuous everywhere, the approximation $r^{(1)}$ of (6.55), and its dominant part, the Rayleigh approximation

$$r_R = -\int_{-\infty}^{\infty} \mathrm{d}z \frac{\mathrm{d}q/\mathrm{d}z}{2q} \mathrm{e}^{2i\phi}, \tag{6.59}$$

work well down to surprisingly small values of $(\omega/c)\Delta z$. However, these approximations give, in general, only the correct exponent in the short wave limiting form, and not the correct prefactor. We shall illustrate with the hyperbolic tangent profile, for which the phase integral may be evaluated analytically. We have

$$q^2(z) = \frac{1}{2}\left(q_1^2 + q_2^2\right) + \frac{1}{2}\left(q_2^2 - q_1^2\right) \tanh z/2\Delta z. \tag{6.60}$$

In the phase integral

$$\phi(z) = \int_0^z \mathrm{d}\zeta q(\zeta) \tag{6.61}$$

we transform first to the variable $\tau = \tanh(\zeta/2\Delta z)$, and then to the variable $y = q\Delta z$. We find

$$\begin{aligned}\phi(z) &= \int_0^{\tanh z/2\Delta z} \mathrm{d}\tau \left(\frac{1}{1-\tau} + \frac{1}{1+\tau}\right) y(\tau) = 2\int_{\bar{y}}^{q\Delta z} \mathrm{d}y\, y^2 \left\{\frac{1}{y_2^2 - y^2} + \frac{1}{y^2 - y_1^2}\right\} \\ &= f(y) - f(\bar{y}),\end{aligned} \tag{6.62}$$

where

$$\bar{y}^2 = \frac{1}{2}\left(y_1^2 + y_2^2\right), \quad f(y) = y_2 \ln\frac{y_2 + y}{y_2 - y} - y_1 \ln\frac{y + y_1}{y - y_1} \tag{6.63}$$

(this form applies for $y_1 < y_2$, which is the case when $\varepsilon_1 < \varepsilon_2$). The Rayleigh approximation for the reflection amplitude thus becomes

$$r_R = -\frac{1}{2}\mathrm{e}^{-2if(\bar{y})} \int_{y_1}^{y_2} \frac{\mathrm{d}y}{y} \left(\frac{y_2 + y}{y_2 - y}\right)^{2iy_2} \left(\frac{y - y_1}{y + y_1}\right)^{2iy_1} = -\frac{1}{2}\mathrm{e}^{-2if(\bar{y})} \int_{y_1}^{y_2} \frac{\mathrm{d}y}{y} \mathrm{e}^{2if(y)}. \tag{6.64}$$

The long-wave limiting form is (in accord with the results of Sect. 5.7)

$$r_R \to -\frac{1}{2}\ln\frac{q_2}{q_1}. \tag{6.65}$$

The Rayleigh approximation is compared with the exact result in Fig. 6.3.

We see that the Rayleigh approximation gives good agreement with the exact reflectivity over the entire range of wavelengths. One might expect that at short wavelengths the agreement becomes perfect. This is not so: the asymptotic value has been given by Berry and Mount (1972), and they find that the reflectivity resulting from this approximation is not expression (6.3), but $\pi^2/9 \simeq 1.0966$ times that value. The exponential is given correctly, but not the prefactor. The theoretical reason for such discrepancies was shown by Pokrovskii, Savvinykh and Ulinich (1958) to lie in the nature of the perturbation series of which r_R is the first term. This is called the Bremmer series; a related approach is the Brekhovskikh iteration method of Sect. 5.8. These will be discussed further in the next section. Here we note only that the discrepancy is usually of little practical importance. For example, for the hyperbolic tangent profile the correct asymptotic form exp $(-4\pi q_1 \Delta z)$ is better than the Rayleigh approximation only beyond $(\omega/c)\Delta z \simeq 2$, where the reflectivity is so small (about 10^{-11}) as to be extremely difficult to measure.

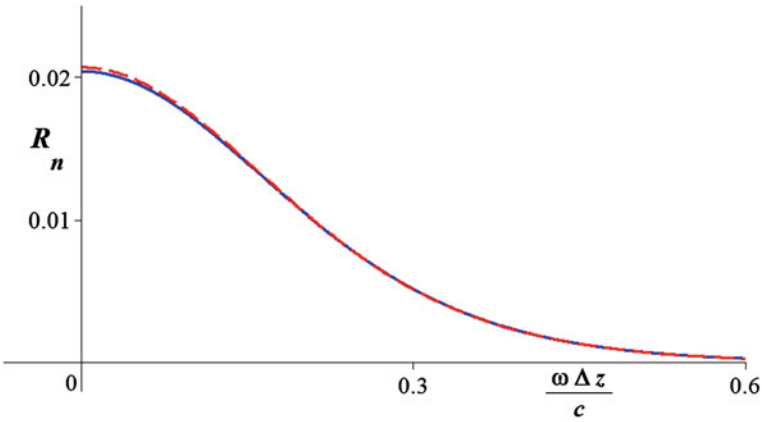


Fig. 6.3 Reflectivity at normal incidence for the *tanh* profile. The *solid curve* is the exact reflectivity from (6.2), the *dashed curve* the Rayleigh approximation reflectivity from (6.64). The values $\varepsilon_1 = 1$, $\varepsilon_2 = (4/3)^2$ are used, as in Fig. 6.1 (where the long and short wave limiting forms were shown)

6.5 Perturbation Theory for Short Waves

As in Sect. 3.1 we wish to express ψ, the solution of

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}z^2} + q^2\psi = 0, \quad \mathrm{e}^{iq_1 z} + r\,\mathrm{e}^{iq_1 z} \leftarrow \psi \rightarrow t\,\mathrm{e}^{iq_2 z}, \tag{6.66}$$

in terms of a known function $\tilde{\psi}$, the solution of

$$\frac{\mathrm{d}^2\tilde{\psi}}{\mathrm{d}z^2} + \tilde{q}^2 = 0, \quad \mathrm{e}^{iq_1 z} + \tilde{r}\,\mathrm{e}^{iq_1 z} \leftarrow \tilde{\psi} \rightarrow \tilde{t}\,\mathrm{e}^{iq_2 z}. \tag{6.67}$$

To do this we need a Green's function $G(z, \zeta)$ satisfying the equation

$$\frac{\partial^2 G}{\partial z^2} + \tilde{q}^2 G = \delta(z - \zeta), \tag{6.68}$$

and giving the appropriate boundary conditions for ψ.

In Sect. 3.1 we used $\mathrm{e}^{\pm iq_0 z}$ to construct G, with $q_0(z) = q_1$, for $z < 0$ and q_2 for $z > 0$. These functions are appropriate for the long wave case. Here we use $\psi_1^+ = (q_1/q)^{1/2}\mathrm{e}^{i\phi}$, $\psi_1^- = (q_2/q)^{1/2}\mathrm{e}^{-i\phi}$ to construct G, these being wavefunctions which approximate ψ in the short wave case. The required Green's function is

$$G(z, \zeta) = \begin{cases} \frac{1}{2i(q_1 q_2)^{1/2}}\psi_1^-(z)\psi_1^+(\zeta) & z < \zeta \\ \frac{1}{2i(q_1 q_2)^{1/2}}\psi_1^+(z)\psi_1^-(\zeta) & z > \zeta \end{cases} \tag{6.69}$$

This G satisfies (6.68) when $z \neq \zeta$, with

$$\tilde{q}^2 = q^2 + \frac{\mathrm{d}^2 q/\mathrm{d}z^2}{2q} - \frac{3}{4}\left(\frac{\mathrm{d}q}{q\mathrm{d}z}\right)^2 = q^2 + \frac{1}{2}q^{1/2}\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{\mathrm{d}q/\mathrm{d}z}{q^{3/2}}\right), \tag{6.70}$$

because of (6.26). The derivative of G is given by (again using $\gamma = q^{-2}\mathrm{d}q/\mathrm{d}z$)

$$\frac{\partial G}{\partial z} = \begin{cases} -q(z)(i + \gamma(z)/2)G(z, \zeta) & z < \zeta \\ q(z)(i - \gamma(z)/2)G(z, \zeta) & z > \zeta \end{cases} \tag{6.71}$$

When $z = \zeta$, G takes the value $1/2iq(z)$. Thus $\partial G/\partial z$ has the required unit discontinuity at $z = \zeta$, leading to the delta function on the right-hand side of (6.68). The integral equation satisfied by ψ appropriate to the reflection problem is

$$\psi(z) = \psi_1^+(z) - \int_{-\infty}^{\infty} \mathrm{d}\zeta G(z, \zeta)\Delta q^2(\zeta)\psi(\zeta), \tag{6.72}$$

where

$$\Delta q^2 = q^2 - \tilde{q}^2 = -\frac{1}{2} q^{1/2} \frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{\mathrm{d}q/\mathrm{d}z}{q^{3/2}}\right). \tag{6.73}$$

As $z \to -\infty$ this gives the asymptotic form (on choosing the lower limit of integration in ϕ to make $\phi \to q_1 z$ as $z \to -\infty$)

$$\begin{aligned}\psi(z) &\to \mathrm{e}^{iq_1 z} - \frac{1}{2i(q_1 q_2)^{\frac{1}{2}}} \int_{-\infty}^{\infty} \mathrm{d}\zeta\ \psi_1^-(z)\psi_1^+(\zeta)\Delta q^2(\zeta)\psi(\zeta) \\ &= \mathrm{e}^{iq_1 z} + \mathrm{e}^{-iq_1 z} \frac{1}{4iq_1^{\frac{1}{2}}} \int_{-\infty}^{\infty} \mathrm{d}\zeta \frac{\mathrm{d}}{\mathrm{d}\zeta}\left(\frac{\mathrm{d}q/\mathrm{d}\zeta}{q^{3/2}}\right)\mathrm{e}^{i\phi(\zeta)}\psi(\zeta).\end{aligned} \tag{6.74}$$

The coefficient of $\mathrm{e}^{-iq_1 z}$ is the exact reflection amplitude r; the expression above is equivalent to (6.54), obtained from the comparison identity (6.50). The first order perturbation result is obtained by setting $\psi = \psi_1^+$ on the right hand sides of (6.72) and (6.74), and reproduces $r^{(1)}$ as given by (6.55). Higher order approximations are obtained by iteration of the integral equation, as in the long wave case of Chap. 3.

The integral equation (6.72) and the resulting perturbation series are closely related to the coupled equations derived by Bremmer (1951), and the resulting Bremmer series. This in turn has an intimate connection with the Brekhovskikh series of Sect. 5.8, as we shall show by deriving the Bremmer equations from the results of Sect. 5.6. In (5.74) and (5.75) we put $f = q^{-\frac{1}{2}}F$, $g = q^{-\frac{1}{2}}G$, to obtain the coupled, first order, linear differential equations

$$F' = \frac{q'}{2q} G\mathrm{e}^{-2i\phi}, \quad G' = \frac{q'}{2q} F\mathrm{e}^{2i\phi}. \tag{6.75}$$

These have the same form as the Bremmer equations (Bremmer (1951), (6.15)); the relation of F and G to the Bremmer functions $u_\uparrow$ and $u_\downarrow$ is

$$F = q^{\frac{1}{2}}\mathrm{e}^{-i\phi}u_\uparrow, \quad G = q^{\frac{1}{2}}\mathrm{e}^{i\phi}u_\downarrow. \tag{6.76}$$

For waves incident from $z = -\infty$ we have $G(+\infty) = 0$. If further we set $F(-\infty) = 1$, then $G(-\infty) = r_s$, and integrating the equation for G' from $-\infty$ to $+\infty$ gives

$$r_s = -\int_{-\infty}^{\infty} \mathrm{d}z \frac{q'}{2q} F\mathrm{e}^{2i\phi}. \tag{6.77}$$

When we approximate F by unity (weak reflection) we regain the Rayleigh expression for r_s.

The Bremmer series, obtained by iteration of the pair (6.75), has been investigated by Bellman and Kalaba (1959) and Atkinson (1960), the latter showing that the series converges if

$$\int_{-\infty}^{\infty} \mathrm{d}z \left| \frac{\mathrm{d}q/\mathrm{d}z}{q} \right| < \pi. \tag{6.78}$$

When $\varepsilon(z)$ is monotonically increasing, so is $q(z)$, and (6.78) may be written as

$$\int_{q_1}^{q_2} \frac{\mathrm{d}q}{q} = \ln \frac{q_2}{q_1} < \pi. \tag{6.79}$$

At normal incidence this condition reads $\varepsilon_2/\varepsilon_1 < e^{2\pi} \approx 535$, a rather weak constraint. The condition for convergence progressively tightens as the angle of incidence is increased: taking the example of the air-water interface at optical frequencies, $\varepsilon_1 = 1,\ \varepsilon_2 \approx \left(\frac{4}{3}\right)^2$, we find that the inequality (6.79) is satisfied by a factor of more than three hundred at normal incidence, but is violated at $\theta_1 \approx 87.8°$, that is at about 2.2° from grazing incidence. The failure of the short wave approximation near grazing incidence was implied in Sect. 6.2, where we noted that the function $\gamma = q^{-2}\mathrm{d}q/\mathrm{d}z$ had to remain small. Near grazing incidence, or near the critical angle (where respectively q_1 and q_2 tend to zero) γ becomes large within the interface, and the short wave perturbation theories fail.

6.6 Short Wave Results for r_p and r_p/r_s

We shall first summarize the results obtained for the s wave, rewriting the results in electromagnetic notation. The electric field is $[0, E(z)\exp(Kx - \omega t), 0]$, with

$$\frac{\mathrm{d}^2 E}{\mathrm{d}z^2} + q^2 E = 0, \quad \mathrm{e}^{iq_1 z} + r_s \mathrm{e}^{-iq_1 z} \leftarrow E \rightarrow t\mathrm{e}^{iq_2 z}. \tag{6.80}$$

Approximate solutions of (6.80) are

$$\psi_1^+ = \left(\frac{q_1}{q}\right)^{\frac{1}{2}} \mathrm{e}^{i\phi},\ \psi_1^- = \left(\frac{q_2}{q}\right)^{\frac{1}{2}} \mathrm{e}^{-i\phi},\ \phi(z) = \int^z \mathrm{d}\zeta q(\zeta), \tag{6.81}$$

which in fact satisfy

$$\frac{\mathrm{d}^2\psi_1^{\pm}}{\mathrm{d}z^2}+\left[q^2+\frac{1}{2q}\frac{\mathrm{d}^2q}{\mathrm{d}z^2}-\frac{3}{4}\left(\frac{\mathrm{d}q}{q\mathrm{d}z}\right)^2\right]\psi_1^{\pm}=0. \tag{6.82}$$

The comparison identity (6.54) applied to r_s

$$r_s=\left(4i\sqrt{q_1}\right)^{-1}\int_{-\infty}^{\infty}\mathrm{d}z\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{\mathrm{d}q/\mathrm{d}z}{q^{3/2}}\right)E\,\mathrm{e}^{i\phi}, \tag{6.83}$$

or the perturbation theory of the previous section, lead to the approximation

$$r_s^{(1)}=\frac{1}{4i}\int_{-\infty}^{\infty}\mathrm{d}z\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{\mathrm{d}q/\mathrm{d}z}{q^{3/2}}\right)q^{-\frac{1}{2}}\mathrm{e}^{2i\phi}. \tag{6.84}$$

For the interfaces which extend from z_1 to z_2, and have discontinuities in the slope of ε at the end points but are otherwise smooth, (6.84) gives

$$r_s^{(1)}=\frac{\mathrm{e}^{i(\phi_1+\phi_2)}}{4i}\left\{\gamma_1\mathrm{e}^{-i\Delta\phi}-\gamma_2\mathrm{e}^{i\Delta\phi}\right\}+\cdots \tag{6.85}$$

(exponentially small terms from the smooth part of the interface being omitted).

We now wish to derive corresponding results for the p wave, for which

$$\mathbf{B}=\left[0,B\mathrm{e}^{i(Kx-\omega t)},0\right]=\left[0,\left(\frac{\varepsilon}{\varepsilon_1}\right)^{\frac{1}{2}}b\mathrm{e}^{i(Kx-\omega t)},0\right], \tag{6.86}$$

The function $b(z)=(\varepsilon_1/\varepsilon)^{\frac{1}{2}}B(z)$, introduced in Sect. 1.2, satisfies

$$\frac{\mathrm{d}^2b}{\mathrm{d}z^2}+q_b^2b=0,\quad \mathrm{e}^{iq_1z}-r_p\mathrm{e}^{-iq_1z}\leftarrow b\rightarrow t_p\mathrm{e}^{iq_2z}, \tag{6.87}$$

$$q_b^2=q^2+\frac{\varepsilon^{\frac{1}{2}}}{2}\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{\mathrm{d}\varepsilon/\mathrm{d}z}{\varepsilon^{\frac{3}{2}}}\right)=q^2-\varepsilon^{\frac{1}{2}}\frac{\mathrm{d}^2\varepsilon^{-\frac{1}{2}}}{\mathrm{d}z^2}=q^2+\frac{1}{2\varepsilon}\frac{\mathrm{d}^2\varepsilon}{\mathrm{d}z^2}-\frac{3}{4}\left(\frac{\mathrm{d}\varepsilon}{\varepsilon\mathrm{d}z}\right)^2. \tag{6.88}$$

(compare (1.22) or (2.3)). For smooth profiles the difference $q_b^2-q^2$ is of order $(\Delta z)^{-2}$, and thus smaller by the factor $[(\omega/c)\Delta z]^{-2}$ in comparison with q^2. It is therefore negligible in the short wave case, except at grazing incidence or near

turning points. For profiles which have discontinuities in their first derivative (typically at the end points z_1 and z_2), the second derivative has delta functions contributions at such points.

We first derive a comparison identity linking b and ψ_1^+, which satisfies (6.82), with the lower limit in the integral defining ϕ chosen so that ϕ tends to $q_1 z$ as $z \to -\infty$:

$$\mathrm{e}^{iq_1 z} \leftarrow \psi_1^+ \rightarrow \left(\frac{q_1}{q_2}\right)^{\frac{1}{2}} \mathrm{e}^{i(q_2 z + \Delta\phi)} \tag{6.89}$$

The expression in brackets in (6.82) will again be shortened to $\tilde{q}^2$. The identity resulting from multiplying (6.87) by ψ_1^+, (6.82) by b, subtracting, and integrating the result from $-\infty$ to ∞ is

$$r_p = \frac{1}{2iq_1} \int_{-\infty}^{\infty} \mathrm{d}z (q_b^2 - \tilde{q}^2) \psi_1^+ b. \tag{6.90}$$

At normal incidence $q \to k = \sqrt{\varepsilon}\omega/c$. Denoting derivatives with respect to z by primes,

$$\tilde{q}^2 - q^2 \to \frac{1}{4}\frac{\varepsilon''}{\varepsilon} - \frac{5}{16}\left(\frac{\varepsilon'}{\varepsilon}\right)^2, \tag{6.91}$$

$$q_b^2 - \tilde{q}^2 \to \frac{1}{4}\frac{\varepsilon''}{\varepsilon} - \frac{7}{16}\left(\frac{\varepsilon'}{\varepsilon}\right)^2. \tag{6.92}$$

Thus the expression (6.90) for r_p gives the normal incidence reflection amplitude

$$r_n = \frac{1}{2ik_1} \int_{-\infty}^{\infty} \mathrm{d}z \left[\frac{1}{4}\frac{\varepsilon''}{\varepsilon} - \frac{7}{16}\left(\frac{\varepsilon'}{\varepsilon}\right)^2\right] \psi_1^+ b, \tag{6.93}$$

which is to be compared with the corresponding expression obtained from (6.84):

$$r_n = \frac{1}{2ik_1} \int_{-\infty}^{\infty} \mathrm{d}z \left[\frac{1}{4}\frac{\varepsilon''}{\varepsilon} - \frac{5}{16}\left(\frac{\varepsilon'}{\varepsilon}\right)^2\right] \psi_1^+ E. \tag{6.94}$$

These are both identities, but give slightly different values for r_n when b and E are both approximated by ψ_1^+.

At general angle of incidence, (6.90) reads

$$r_p = \frac{1}{4iq_1}\int_{-\infty}^{\infty} \mathrm{d}z\left[\frac{\varepsilon''}{\varepsilon} - \frac{3}{2}\left(\frac{\varepsilon'}{\varepsilon}\right)^2 - \frac{q''}{q} + \frac{3}{2}\left(\frac{q'}{q}\right)^2\right]\psi_1^+ b, \tag{6.95}$$

and when b is approximated by ψ_1^+, it leads to

$$r_p^{(1)} = \frac{1}{4i}\int_{-\infty}^{\infty} \mathrm{d}z\left[\frac{\varepsilon''}{\varepsilon} - \frac{3}{2}\left(\frac{\varepsilon'}{\varepsilon}\right)^2 - \frac{q''}{q} + \frac{3}{2}\left(\frac{q'}{q}\right)^2\right]q^{-1}\mathrm{e}^{2i\phi}. \tag{6.96}$$

For profiles of finite extent which have discontinuities in the slope of $\varepsilon(z)$ at the boundaries z_1 and z_2, the dominant short wave contribution comes from these discontinuities. Then, as in the s wave case, the quantity $\gamma = q^{-2}\mathrm{d}q/\mathrm{d}z$ changes from 0 to γ_1 at z_1, and from γ_2 to 0 at z_2. The resulting contribution to the integrand of (6.96) is

$$\gamma_1 \cos 2\theta_1 \delta(z - z_1) - \gamma_2 \cos 2\theta_2 \delta(z - z_2), \tag{6.97}$$

(at z_1 the q''/q^2 term has delta function strength γ_1 and the $\varepsilon''/\varepsilon q$ term has delta function strength $2\gamma_1 \cos^2 \theta_1$.) If the profile is smooth everywhere except at z_1 and z_2, the leading term in the short wave limit is

$$r_p^{(1)} = \frac{1}{4i}\left\{\gamma_1 \cos 2\theta_1 \mathrm{e}^{2i\phi_1} - \gamma_2 \cos 2\theta_2 \mathrm{e}^{2i\phi_2}\right\} + \cdots \tag{6.98}$$

(…denotes exponentially small term terms). This may be rewritten, with $\Delta\phi = \phi_2 - \phi_1$, as

$$r_p^{(1)} = \frac{\mathrm{e}^{i(\phi_1+\phi_2)}}{4i}\left\{\gamma_1 \cos 2\theta_1 \mathrm{e}^{-i\Delta\phi} - \gamma_2 \cos 2\theta_2 \mathrm{e}^{i\Delta\phi}\right\} + \cdots. \tag{6.99}$$

At normal incidence (6.99) is in agreement with $r_s^{(1)}$ as given by (6.85), and the ratio $r_p^{(1)}/r_s^{(1)}$ correctly takes the value +1. At grazing incidence the approximations $r_s^{(1)}$ and $r_p^{(1)}$ both fail, since the assumption that $\gamma = q^{-2}\mathrm{d}q/\mathrm{d}z$ is small compared to unity cannot hold as $q_1 = \sqrt{\varepsilon}(\omega/c)\cos\theta_1$ tends to zero. Neither does $r_p^{(1)}$ tend to 1 nor does $r_s^{(1)}$ tend to -1 (the correct limiting values at grazing incidence, as shown in Sect. 2.3), but $r_p^{(1)}/r_s^{(1)}$ does tend to the correct limiting value of -1, since the γ_1 term dominates and $\cos 2\theta_1$ tends to -1.

The s and p reflectivities in the short wave limit are

$$R_s^{(1)} = \left|r_s^{(1)}\right|^2 = \frac{1}{16}\left\{\gamma_1^2 + \gamma_2^2 - 2\gamma_1\gamma_2 \cos 2\Delta\phi\right\}, \tag{6.100}$$

$$R_p^{(1)} = \left|r_p^{(1)}\right|^2 = \frac{1}{16}\{\gamma_1^2 \cos^2 2\theta_1 + \gamma_2^2 \cos^2 2\theta_2 - 2\gamma_1\gamma_2 \cos^2 2\theta_1 \cos^2 2\theta_2 \cos 2\Delta\phi\}. \tag{6.101}$$

We will compare these formulae with the exact results for the exponential profile defined in (2.93), in which the dielectric function ε changes exponentially with z (compare Sects. 2.5 and 6.1) and

$$\gamma = \frac{1}{2} q^{-3} \frac{\mathrm{d}q^2}{\mathrm{d}z} = \frac{\varepsilon\omega^2/c^2}{2aq^3}, \quad a = \frac{\Delta z}{\ln\,(\varepsilon_2/\varepsilon_1)}, \tag{6.102}$$

$$\Delta\phi = 2a\{q_2 - q_1 - K[\arctan(q_2/K) - \arctan(q_1/K)]\}. \tag{6.103}$$

The reflectivities as a function of angle of incidence are shown in Figs. 6.4 and 6.5.

We see from these figures that the short wave approximations work well at normal incidence with the rather small value $(\omega/c)\Delta z = 2$, but at this value their accuracy is poor near the Brewster angle and beyond. Thus care must be taken in the ellipsometric application of the formulae (6.85) and (6.99) in the intermediate region when the interfacial thickness is of the same order of magnitude as the

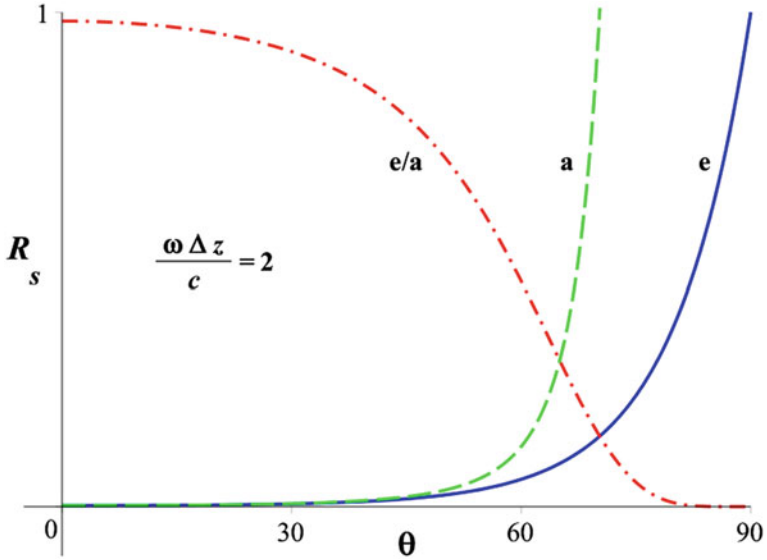


Fig. 6.4 Angular dependence of the s wave reflectivity for the exponential profile, at $(\omega/c)\Delta z = 2$ and with $\varepsilon_1 = 1$, $\varepsilon_2 = (4/3)^2$. The *solid curve* (marked *e*) is the exact reflectivity obtained from (2.99); the *dashed curve* (*a*) is the short wave approximation (6.100). The ratio $R_s/R_s^{(1)}$ is also shown as the *dash-dot curve*; $R_s^{(1)}$ is about 2 % too large at normal incidence, and about a factor of 2 too large at 60°

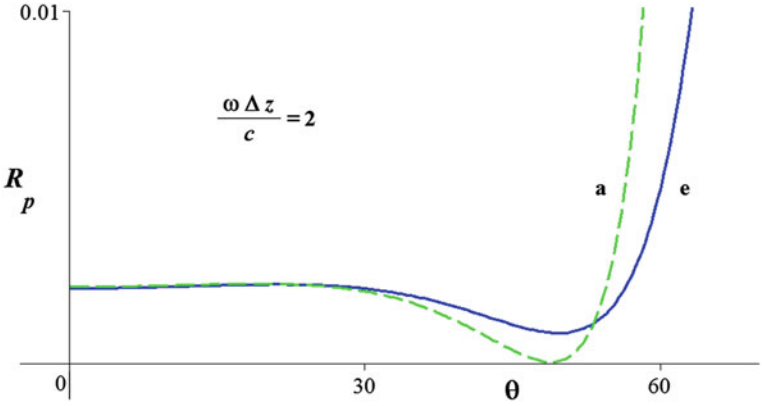


Fig. 6.5 Angular dependence of the p wave reflectivity for the exponential profile, at $(\omega/c)\Delta z = 2$ and with $\varepsilon_1 = 1,\ \varepsilon_2 = (4/3)^2$. The *solid curve* is the exact reflectivity obtained from (2.98); the *dashed curve* is the short wave approximation (6.101). Note the vertical scale is enlarged one hundred times relative to Fig. 6.4. The minima of R_p and $R_p^{(1)}$ are near 49.7° and 48.9°; the zero-thickness Brewster angle is $\arctan(4/3) \approx 53.1°$

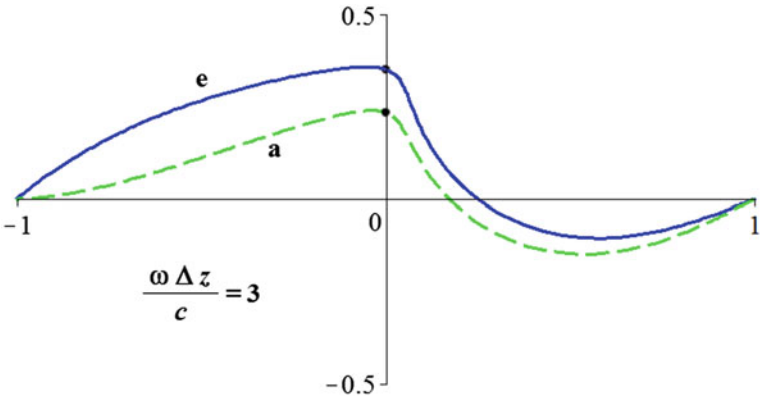


Fig. 6.6 Exact (*solid curve*, *e*) and approximate (*dashed curve*, *a*) trajectories of r_p/r_s in the complex plane, as a function of the angle of incidence. The *curves* are drawn from the exponential profile, with $\varepsilon_1 = 1,\ \varepsilon_2 = (4/3)^2,\ (\omega/c)\Delta z = 3$. The curves start at 1 at normal incidence, pass through their principal angles of 41.9° and 39.7° (where the respective real parts of r_p/r_s are zero, points shown on the graph), and end at −1 at glancing incidence

wavelength ($(\omega/c)\Delta z = 2$ corresponds to a wavelength about three times the interfacial thickness, giving the normal incidence values $\gamma_1 \approx 1/7$ and $\gamma_2 \approx 1/10$ for the profile used in Figs. 6.4 and 6.5).

From (6.85) and (6.99) we find that the real and imaginary parts of the ellipsometric ratio have the short wave limiting forms

$$\mathrm{Re}\left(\frac{r_p}{r_s}\right)^{(1)} = \frac{\gamma_1^2\cos 2\theta_1 + \gamma_2^2\cos 2\theta_2 - \gamma_1\gamma_2(\cos 2\theta_1 + \cos 2\theta_2)\cos 2\Delta\phi}{\gamma_1^2 + \gamma_2^2 - 2\gamma_1\gamma_2\cos 2\Delta\phi}, \quad (6.104)$$

$$\mathrm{Im}\left(\frac{r_p}{r_s}\right)^{(1)} = \frac{\gamma_1\gamma_2(\cos 2\theta_1 - \cos 2\theta_2)\sin 2\Delta\phi}{\gamma_1^2 + \gamma_2^2 - 2\gamma_1\gamma_2\cos 2\Delta\phi}. \quad (6.105)$$

The trajectory of r_p/r_s in the complex plane, as a function of the angle of incidence, is shown in Fig. 6.6 for the exponential profile. We see that while the short wave approximations for the reflectivities work well down to $(\omega/c)\Delta z = 2$ at normal incidence, the agreement is poor at intermediate angles even at $(\omega/c)\Delta z = 3$.

So far in this section we have given results based on the approximations $r_s^{(1)}$ and $r_p^{(1)}$, which may be written in the form

$$r_s^{(1)} = \frac{1}{4i}\int_{-\infty}^{\infty} \mathrm{d}z\left[\frac{q''}{q} - \frac{3}{2}\left(\frac{q'}{q}\right)^2\right]q^{-1}\mathrm{e}^{2i\phi}, \quad (6.106)$$

$$r_p^{(1)} = \frac{1}{4i}\int_{-\infty}^{\infty} \mathrm{d}z\left[\frac{\varepsilon''}{\varepsilon} - \frac{3}{2}\left(\frac{\varepsilon'}{\varepsilon}\right)^2 - \frac{q''}{q} + \frac{3}{2}\left(\frac{q'}{q}\right)^2\right]q^{-1}\mathrm{e}^{2i\phi}. \quad (6.107)$$

We will compare these with the Rayleigh approximations of Sect. 5.7:

$$r_s^R = -\int_{-\infty}^{\infty} \mathrm{d}z\frac{q'}{2q}\mathrm{e}^{2i\phi}, \quad (6.108)$$

$$r_p^R = \int_{-\infty}^{\infty} \mathrm{d}z\frac{Q'}{2Q}\mathrm{e}^{2i\phi}. \quad (6.109)$$

By changing the variable of integration temporarily to ϕ, integrating by parts, and changing back, these may be written in the form

$$r_s^R = \frac{1}{4i}\int_{-\infty}^{\infty} \mathrm{d}z\left[\frac{q''}{q} - 2\left(\frac{q'}{q}\right)^2\right]q^{-1}\mathrm{e}^{2i\phi}, \quad (6.110)$$

$$r_p^R = \frac{1}{4i}\int_{-\infty}^{\infty} \mathrm{d}z\left[\frac{\varepsilon''}{\varepsilon} - \left(\frac{\varepsilon'}{\varepsilon}\right)^2 - \frac{\varepsilon' q'}{\varepsilon q} - \frac{q''}{q} + 2\left(\frac{q'}{q}\right)^2\right]q^{-1}\mathrm{e}^{2i\phi}. \quad (6.111)$$

When written in this form the difference between the two approximate sets is seen to lie in the coefficients of the square of the first derivatives, but not in the second derivative. Thus for profiles with a discontinuity in the first derivative of ε, and consequently a delta function contribution to ε'' or q'', there is no difference between the short wave limiting forms of the Rayleigh approximation and those based on the Liouville-Green waveforms.

At normal incidence the two Rayleigh forms agree, giving

$$r_n^R = \frac{1}{8i}\int_{-\infty}^{\infty} \mathrm{d}z \left[\frac{\varepsilon''}{\varepsilon} - \frac{3}{2}\left(\frac{\varepsilon'}{\varepsilon}\right)^2\right] k^{-1} \mathrm{e}^{2i\phi}, \tag{6.112}$$

while there is a difference between the expressions (6.106) and (6.107) when $q \to \sqrt{\varepsilon}\omega/c = k$, the two averaging to (6.112):

$$r_s^{(1)} \to \frac{1}{8i}\int_{-\infty}^{\infty} \mathrm{d}z \left[\frac{\varepsilon''}{\varepsilon} - \frac{5}{4}\left(\frac{\varepsilon'}{\varepsilon}\right)^2\right] k^{-1} \mathrm{e}^{2i\phi}, \tag{6.113}$$

$$r_p^{(1)} \to \frac{1}{8i}\int_{-\infty}^{\infty} \mathrm{d}z \left[\frac{\varepsilon''}{\varepsilon} - \frac{7}{4}\left(\frac{\varepsilon'}{\varepsilon}\right)^2\right] k^{-1} \mathrm{e}^{2i\phi}. \tag{6.114}$$

6.7 A Single Turning Point: Total Reflection

The preceding sections have dealt with the case $\varepsilon_1 < \varepsilon_2$ (or $V_1 > V_2$ in the quantum particle case), where $q^2(z) = \varepsilon(z)\omega^2/c^2 - K^2$ is positive everywhere. We now examine the opposite case where $\varepsilon_1 > \varepsilon_2$ (or $V_1 < V_2$), of which examples are: light incident on an interface from the optically denser medium, radio waves incident on an ionospheric layer, or particles moving up a potential gradient. This was illustrated in Figs. 1.8 and 1.9. For $\theta_1 > \theta_c = \arcsin(\varepsilon_2/\varepsilon_1)^{\frac{1}{2}}$ there will be total internal reflection, since then $q_2^2 < 0$, and the wave deep inside medium 2 decays exponentially as exp $(-|q_2|z)$. In geometrical optics and classical particle physics the reflection occurs at the point z_0 defined by $q^2(z_0) = 0$. This is called the *turning point*: a classical particle turns back at z_0, being unable to penetrate into a region where the kinetic energy of the motion in the z direction would become negative. Waves do penetrate beyond this point, but decay exponentially for $z > z_0$, and there is no propagating wave at infinity.

For a given profile, the location of the turning point is a function of the angle of incidence. The location is given by $q^2(z_0) = \varepsilon(z_0)\omega^2/c^2 - K^2 = 0$, and since $K^2 = \varepsilon_1(\omega^2/c^2)$ sin $^2\theta_1$, z_0 is determined by

$$\varepsilon(z_0) = \varepsilon_1 \sin^2 \theta_1. \tag{6.115}$$

For example, for the hyperbolic tangent profile, for which

$$\varepsilon(z) = \frac{\varepsilon_1 + \varepsilon_2 e^{z/\Delta z}}{1 + e^{z/\Delta z}}, \tag{6.116}$$

$$z_0(\theta_1) = \Delta z \ln \frac{\cos^2 \theta_1}{\sin^2 \theta_1 - \varepsilon_2/\varepsilon_1} = \Delta z \ln \frac{\cos^2 \theta_1}{\sin^2 \theta_1 - \sin^2 \theta_c}. \tag{6.117}$$

This varies from $+\infty$ at θ_c to $-\infty$ at grazing incidence. Three curves of q^2 versus z for this profile are shown in Fig. 6.7.

A turning point, and the consequent total reflection, may be present even at normal incidence if the dielectric function passes through zero. An example is provided by the dielectric function of an electron plasma, approximating the electron gas in metals or electrons in the ionosphere. If electron collisions and the consequent damping are neglected, this takes the form (see for example Budden 1985 or Kittel 1976)

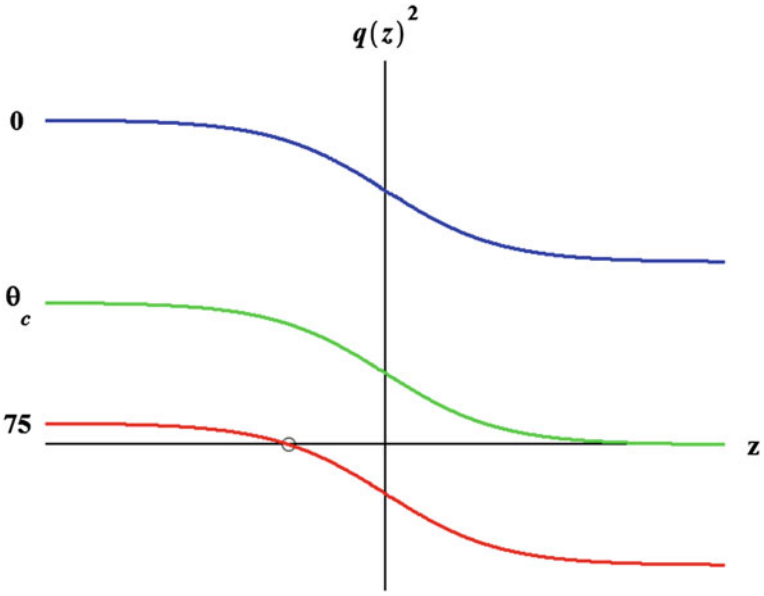


Fig. 6.7 Variation of $q^2(z)$ with angle of incidence. The *curves* are drawn for the hyperbolic tangent profile, with $\varepsilon_1 = (4/3)^2$, $\varepsilon_2 = 1$, $(\theta_c = \arcsin 3/4 \approx 48.59^\circ)$, representing the water-air interface at optical frequencies. The turning point for $\theta_1 = 75^\circ$ is circled

$$\varepsilon(z,\omega) = 1 - \frac{\omega_p^2(z)}{\omega^2}, \tag{6.118}$$

where ω_p is the plasma angular frequency, and is a function of z through its proportionality to the square root of the electron density. For this simplified dielectric function, there is a turning point at normal incidence at z_0 given by $\omega_p(z_0) = \omega$, and at a general angle of incidence at $z_0(\omega, \theta_1)$ given by

$$\omega_p(z_0) = \omega \cos\theta_1 \tag{6.119}$$

(the limiting value $\varepsilon_1 = 1$ is assumed in (6.118)). This model will be considered again, with dissipation included, in Chap. 10. An analogous case of total reflection at normal incidence occurs for particles when their energy is less than V_2, the potential energy in the second medium.

When total reflection occurs we know that $|r|^2 = 1$ (in the absence of dissipation) so there is little point in calculating the magnitude of r. But there is information in the phase of r: it gives for example the location of the ellipsometric ratio r_p/r_s on the unit circle, and determines the time of arrival and shape of reflected pulses (as we shall see in Chap. 19). A simple argument shows that the phase of r_s is always a bit less than $2(\phi_0 - \phi_-)$, where $\phi(z) = \int^z \mathrm{d}\zeta q(\zeta)$ is the phase integral, and takes the limiting form $q_1 z + \phi_-$ as $z \to -\infty$, and ϕ_0 is the value of ϕ at z_0. The argument is based on the behaviour of the wave function near a turning point, illustrated in Fig. 6.8 for the linear profile considered in Sect. 5.2.

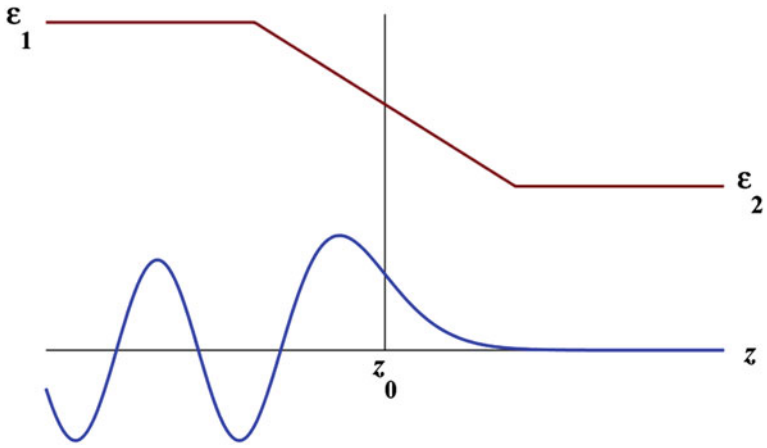


Fig. 6.8 Wavefunction $E(z)$ for total internal reflection by the linear dielectric function profile. The parameters used are $\varepsilon_1 = 2$, $\varepsilon_2 = 1$, $\theta_1 = 60°$. The real and imaginary parts of $E(z)$ are proportional to each other when there is total reflection: $\mathrm{Im}(E)/\mathrm{Re}(E) = \tan(\delta_s/2)$ when $r_s = \mathrm{e}^{i\delta_s}$ (see Sect. 2.2). Only the imaginary part is shown. The turning point z_0 is half-way down the ramp at this angle of incidence

The fact that the phase of r_s is always a bit less than $2(\phi_0 - \phi_-)$ follows from the shape of the wavefunction near the turning point, where it changes from oscillatory behaviour to monotonic decay, with a consequent extremum at $z_0 - \delta z$. The zeroth approximation for the wave is, for $z < z_0$,

$$E \approx \psi_0 = \mathrm{e}^{i(\phi-\phi_-)} + r_s \mathrm{e}^{-i(\phi-\phi_-)}. \tag{6.120}$$

This expression fails near z_0, but in the short-wave limit its region of validity approaches it. Since E has an extremum at $z_0 - \delta z$, we obtain an estimate of r_s by setting $\mathrm{d}\psi_0/\mathrm{d}z = 0$ at this point, where ϕ takes the value $\phi_0 - \delta\phi$. This gives $r_s \approx \mathrm{e}^{2i(\phi_0-\delta\phi-\phi_-)}$. For profiles which can be approximated by a linear variation near the turning point, it turns out that $\delta\phi$ takes the *universal value* $\pi/4$ in the short wave limit:

$$r_s \approx \mathrm{e}^{2i(\phi_0-\pi/4-\phi_-)}, \quad \delta_s \approx 2(\phi_0 - \phi_-) - \frac{\pi}{2}. \tag{6.121}$$

Equation (6.121) is derived by constructing an accurate solution in the neighbourhood of the turning point, on the assumption that $q^2(z) = \varepsilon(z)\omega^2/c^2 - K^2$ is approximately linear near its zero z_0,

$$q^2(z) \approx (z - z_0)\frac{\omega^2}{c^2}\left(\frac{\mathrm{d}\varepsilon}{\mathrm{d}z}\right)_0. \tag{6.122}$$

For waves incident from the left and totally reflected, q^2 and ε are decreasing functions of z, and the derivative of ε at z_0 is negative. We set

$$\zeta = (z - z_0)\left[\frac{\omega^2}{c^2}\left(-\frac{\mathrm{d}\varepsilon}{\mathrm{d}z}\right)_0\right]^{\frac{1}{3}}, \tag{6.123}$$

and the wave equation $\mathrm{d}^2E/\mathrm{d}z^2 + q^2E = 0$ transforms, in the neighbourhood of the turning point, to Airy's equation

$$\frac{\mathrm{d}^2E}{\mathrm{d}\zeta^2} - \zeta E = 0. \tag{6.124}$$

The solutions $Ai(\zeta)$ and $Bi(\zeta)$ of (6.124) were discussed in Sect. 5.2. Here we need only the asymptotic forms for large $|\zeta|$:

$$\begin{aligned} Ai(\zeta) &\sim \tfrac{1}{2}\pi^{-\frac{1}{2}}\zeta^{-\frac{1}{4}}\exp\left[-\tfrac{2}{3}\zeta^{\frac{3}{2}}\right] && (|\arg\zeta|<\pi) \\ Ai(-\zeta) &\sim \pi^{-\frac{1}{2}}\zeta^{-\frac{1}{4}}\sin\left[\tfrac{2}{3}\zeta^{\frac{3}{2}}+\tfrac{\pi}{4}\right] && (|\arg\zeta|<2\pi/3) \\ Bi(\zeta) &\sim \pi^{-\frac{1}{2}}\zeta^{-\frac{1}{4}}\exp\left[\tfrac{2}{3}\zeta^{\frac{3}{2}}\right] && (|\arg\zeta|<\pi/3) \\ Bi(-\zeta) &\sim \pi^{-\frac{1}{2}}\zeta^{-\frac{1}{4}}\cos\left[\tfrac{2}{3}\zeta^{\frac{3}{2}}+\tfrac{\pi}{4}\right] && (|\arg\zeta|<2\pi/3) \end{aligned} \tag{6.125}$$

The main features of the asymptotics (for real ζ) can be obtained from the approximate waveforms $\psi_1^{\pm}$ of Sect. 6.2. For example when $\zeta > 0$ these are proportional to $\zeta^{-\frac{1}{4}}\exp\left[\mp\frac{2}{3}\zeta^{\frac{3}{2}}\right]$. For more details see Heading (1962, Appendix A3), Budden (1985, Chap. 8), and Olver (2010).

We are now able to complete the derivation of (6.121) by matching the approximate solution $q^{-\frac{1}{2}}\psi_0$ (ψ_0 given by (6.120)), which breaks down in the neighbourhood of z_0, to the asymptotic form of the solution $Ai(\zeta)$ which is accurate near z_0. (The coefficient of $Bi(\zeta)$ goes to zero in the short wave limit, since $Bi(\zeta)$ diverges for large positive ζ.) For large negative ζ, $Ai(\zeta)$ is, from (6.125), proportional to $(-\zeta)^{-\frac{1}{4}}\sin\left[\frac{2}{3}(-\zeta)^{\frac{3}{2}}+\frac{\pi}{4}\right]$. The quantity $\frac{2}{3}(-\zeta)^{\frac{3}{2}}$ is, from (6.122), (6.123) and the definition of the phase integral, equal to $\phi_0 - \phi$. Also $(-\zeta)^{-\frac{1}{4}}$ is proportional to $q^{-\frac{1}{2}}$. Thus $\sin\left(\phi_0 - \phi + \frac{\pi}{4}\right)$ must be proportional to ψ_0 as given by (6.120), which proportionality holds when r_s is given by (6.121). The result (6.121) is, in essence, due to Hartree (1931).

The derivation has assumed an overlap of regions of validity of the Airy function solution, and of the ψ_0 form. The Airy function is the solution of the equation (6.124), which is itself an approximation, valid when $(z - z_0)(\mathrm{d}\varepsilon/\mathrm{d}z)_0$ is small. On the other hand, the variable ζ must be large to ensure the accuracy of the asymptotic forms assumed in the derivation. As for the short wave form $q^{-\frac{1}{2}}\psi_0$, we know this to fail when $\gamma = q^{-2}\mathrm{d}q/\mathrm{d}z = \frac{1}{2}q^{-3}(\omega^2/c^2)\mathrm{d}\varepsilon/\mathrm{d}z$ is not small. We note from (6.122) that the quantity $\left|\gamma\zeta^{\frac{3}{2}}\right|$ is of order unity, and thus overlap of the regions of validity of the two forms will exist only if $\left[(\omega/c)/|\mathrm{d}\varepsilon/\mathrm{d}z|_0\right]^{\frac{2}{3}}$ is substantially larger than the value of $|\zeta|$ for which the asymptotic forms become useful. If this is (say) 4, we need $(\omega/c)/|\mathrm{d}\varepsilon/\mathrm{d}z|_0$ substantially larger than 8 as a necessary condition for the validity of (6.121).

In the case of total reflection, with $q_2 = i|q_2|$, it is convenient to set the lower limit in the phase integral equal to the turning point z_0:

$$\phi(z) = \int_{z_0}^{z} \mathrm{d}\zeta\, q(\zeta), \tag{6.126}$$

making $\phi_0 = 0$. When the inhomogeneity in dielectric function or potential is finite, and extends from $z = 0$ to $z = \Delta z$, $\phi_- = -\int_0^{z_0} \mathrm{d}z\ q(z)$, we can write the phase of r_s in the form

$$\delta_s = 2\int_0^{z_0} \mathrm{d}z\ q(z) - \frac{\pi}{2}. \tag{6.127}$$

For the hyperbolic tangent profile

$$\varepsilon(z) = \frac{1}{2}(\varepsilon_1 + \varepsilon_2) - \frac{1}{2}(\varepsilon_1 - \varepsilon_2)\tanh z/2\Delta z \tag{6.128}$$

the phase integral may be evaluated exactly. As in Sect. 6.4 we use the variables $\tau = \tanh z/2\Delta z$ and then $y = q\Delta z$. We need ϕ for $z \le z_0$; this is given by

$$\phi(y) = 2|y_2| \arctan\frac{y}{|y_2|} - y_1 \ln\frac{y_1 + y}{y_1 - y}. \tag{6.129}$$

To evaluate ϕ_-, defined by $\phi \to q_1 z + \phi_-$ as $z \to -\infty$, we set $y = y_1 - \delta y$ in (6.129), using $\delta y \to (y_1^2 + |y_2|^2)/2y_1 \mathrm{e}^{z/\Delta z}$. This gives ϕ_-:

$$\phi_- = 2|y_2| \arctan\frac{y_1}{|y_2|} - y_1 \ln\frac{4y_1^2}{y_1^2 + |y_2|^2}. \tag{6.130}$$

The approximate phase of the reflection amplitude is $2(\phi_0 - \frac{\pi}{4} - \phi_-)$, and we have chosen to make ϕ_0 zero. At grazing incidence $q_1 = \sqrt{\varepsilon}(\omega/c)\cos\theta_1$ tends to zero, so ϕ_- tends to zero, and (6.121) gives the incorrect limiting value $r_s \to -i$ (we saw in Sect. 2.3 that the exact r_s always tends to -1 at grazing incidence). The incorrect glancing value is due to the breakdown of the short wave approximations at grazing incidence, where even if $(\omega/c)\Delta z$ is large $q_1\Delta z$ eventually tends to zero. The approximate phase is compared in Fig. 6.9 with the exact phase, calculated from (A26) of Chap. 20, or directly from (2.82).

For profiles of finite extent the short-wave expression (6.127) applies, and is always incorrectly equal to $-\frac{\pi}{2}$ at glancing incidence. Lekner (1996) derives an exact expression from the general result of Sect. 2.2, rewritten in this chapter as (6.32). (The continuity of the dielectric function at the profile boundaries is assumed.) In (6.32) we set $z_1 = 0$ and $q_2 = i|q_2|$. This gives

$$r_s = \frac{q_1\left[\left(F, G^{'}\right) + |q_2|(F, G)\right] + i\left[\left(F^{'}, G^{'}\right) + |q_2|(F^{'}, G)\right]}{q_1\left[\left(F, G^{'}\right) + |q_2|(F, G)\right] - i\left[\left(F^{'}, G^{'}\right) + |q_2|(F^{'}, G)\right]}. \tag{6.131}$$

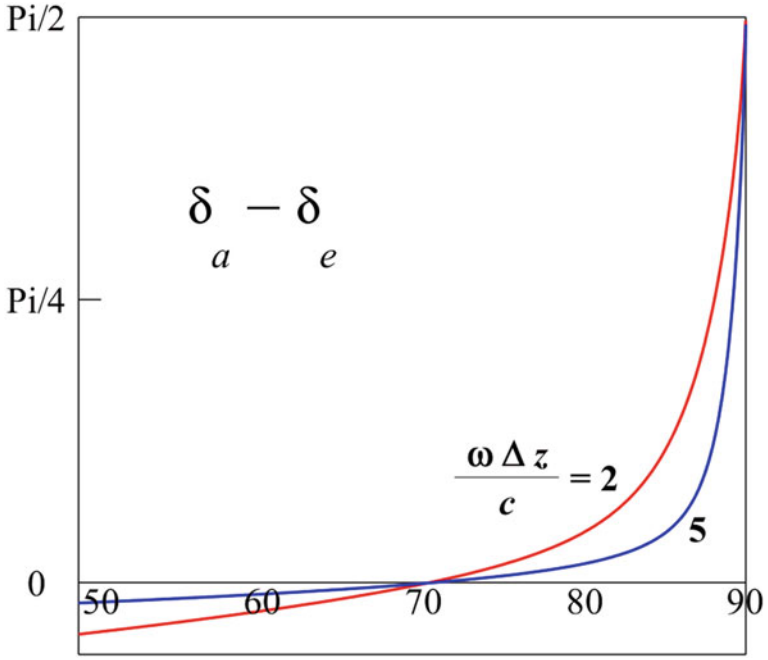


Fig. 6.9 The difference between the approximate and exact phases of the reflection amplitude for the hyperbolic tangent profile, with $\varepsilon_1 = (4/3)^2$, $\varepsilon_2 = 1$, $(\omega/c)\Delta z = 2$ and 5. The phase difference is plotted as a function of the angle of incidence for $\theta_1 > \theta_c \approx 48.6°$, in the total reflection region

We have used the shorthand notations $(F, G) = F_1 G_2 - G_1 F_2$, $\left(F, G^{'}\right) = F_1^{'} G_2 - G_1^{'} F_2$, et cetera. Since the wave equation is linear, with real coefficients, F and G may be taken to be real. The modulus of r_s is clearly unity, and the phase is

$$\delta_s = 2\arctan\left\{\frac{\left(F^{'}, G^{'}\right) + |q_2|(F^{'}, G)}{q_1[(F, G^{'}) + |q_2|(F, G)]}\right\}. \tag{6.132}$$

As $q_1 \to 0, \delta_s \to \pm\pi$ and r_s correctly tends to -1. Lekner (1996) compares the approximate formula (6.127) with the exact solutions for the exponential profile, for both s and p polarizations, in the context of Lloyd's mirror fringes. More detail may be found in Sect. 16.5, in the chapter on neutron reflection.

The p wave reflection amplitude in the case of total reflection, again assuming $z_1 = 0$ and continuity of the dielectric function at the profile boundaries, has similar form to the s wave amplitude of (6.131):

$$r_p = -\frac{q_1\left[\left(F, G^{'}\right) + |q_2|(F, G)\right] + i\left[\left(F^{'}, G^{'}\right) + |q_2|\left(F^{'}, G\right)\right]}{q_1[(F, G^{'}) + |q_2|(F, G)] - i[(F^{'}, G^{'}) + |q_2|(F^{'}, G)]}. \tag{6.133}$$

The functions F and G in (6.133) are now solutions of the equation (1.20) or (2.37) for the magnetic field. As $q_1 \to 0$, $\delta_p \to 0(\mathrm{mod}\, 2\pi)$ and r_p correctly tends to 1.

6.8 Two Turning Points, and Tunneling

The penetration of wave motion into a region where $q^2 < 0$ leads to the possibility of *tunneling*, where the classically forbidden region is traversed by a small portion of the wave, which exits into a region where $q^2 > 0$ and propagates on. Such transmission via tunneling involves an even number of turning points. We will restrict our consideration to two classical turning points z_1 and z_2, defined by $q^2(z_1) = 0 = q^2(z_2)$. Our aim is to derive short wave approximations for the reflection and transmission amplitudes in this case. But first we will show an analytically solvable example of tunneling, provided by the sech^2 profile

$$\varepsilon(z) = \varepsilon_0 + \Delta\varepsilon\, \mathrm{sech}^2 z/a, \tag{6.134}$$

which was discussed in Sect. 4.3, and will again appear as a solvable example in the reflection and transmission of quantum particle wavepackets (Sect. 19.2). The s wave equation $\mathrm{d}^2E/\mathrm{d}z^2 + q^2E = 0$ has

$$q^2(z) = q_0^2 + \Delta\varepsilon \frac{\omega^2}{c^2} \mathrm{sech}^2 z/a, \tag{6.135}$$

where q_0 is the common value at $\pm\infty$ of the wavevector component perpendicular to the interface,

$$q_0^2 = \varepsilon_0 \frac{\omega^2}{c^2} - K^2 = \varepsilon_0 \frac{\omega^2}{c^2} \cos^2\theta, \tag{6.136}$$

and θ is the common value of the angles of incidence and refraction. Thus q^2 is given by

$$q^2(z) = \varepsilon_0 \frac{\omega^2}{c^2} \left[\cos^2\theta + \frac{\Delta\varepsilon}{\varepsilon_0} \mathrm{sech}^2 z/a \right]. \tag{6.137}$$

When $\Delta\varepsilon$ is positive there are no turning points (it is assumed $\varepsilon_0 > 0$), but when $\Delta\varepsilon$ is negative a pair of symmetrically placed turning points come into existence for

$$\theta > \theta_t = \arccos(-\Delta\varepsilon/\varepsilon_0)^{\frac{1}{2}}. \tag{6.138}$$

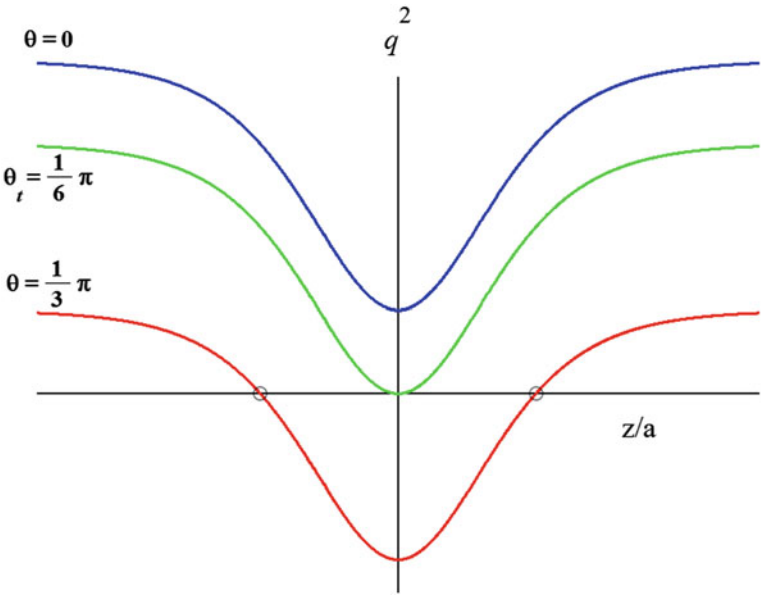


Fig. 6.10 The variation of $q^2(z)$, for the sech^2 profile, with the angle of incidence. The *curves* are drawn for $\varepsilon_0 = 1$, $\Delta\varepsilon = -3/4$, for which $\theta_t = 30°$. The two turning points for $\theta = 60°$ are circled. Negative $\Delta\varepsilon$ corresponds to a positive potential barrier in the quantum particle case

The zeros of $q^2(z)$ are at $\pm z_0$, with

$$z_0 = a \ln \frac{\cos\theta_t + (\cos^2\theta_t - \cos^2\theta)^{\frac{1}{2}}}{\cos\theta}. \tag{6.139}$$

(When $\Delta\varepsilon < -\varepsilon_0$ the dielectric function becomes negative near the origin, and there are two turning points at any angle of incidence. Their location is still given by (6.138) and (6.139), with θ_t now imaginary). The variation of q^2 with z and angle of incidence is shown in Fig. 6.10.

The reflection properties of the sech^2 profile are characterized by two dimensionless parameters α and β, or s and β, where

$$\alpha = \Delta\varepsilon(\omega a/c)^2, \quad \beta = q_0 a, \quad s = \frac{1}{2}\left[-1 + (1+4\alpha)^{\frac{1}{2}}\right]. \tag{6.140}$$

For negative α there will be tunneling at a large enough angle of incidence. The transition from no tunneling to tunneling (in quantum particle language, from *over* the potential barrier to *through* the barrier) takes place when $\beta = (-\alpha)^{\frac{1}{2}}$ if $-\varepsilon_0 < \Delta\varepsilon < 0$. We saw in Sect. 4.3 that the reflectivity takes different analytic

forms according as $\alpha > -\frac{1}{4}$ (s real), in which case $R_s = |r_s|^2$ is given by (4.35), or $\alpha < -\frac{1}{4}$ in which case s is complex and the reflectivity is given by (4.38):

$$s = -\frac{1}{2} + i\sigma, \quad \sigma = \frac{1}{2}(4|\alpha| - 1)^{\frac{1}{2}}, \quad R_s = \frac{\cosh^2 \pi\sigma}{\cosh^2 \pi\sigma + \sinh^2 \pi\beta}, \tag{6.141}$$

When $\alpha = -\frac{1}{4}$ the reflectivity is $R_s = [1 + \sinh^2 \pi\beta]^{-1} = \mathrm{sech}^2 \pi\beta$.

Reflectivity contours for the sech^2 profile in the α, β plane are shown in Fig. 6.11. We note the rapid rise in the reflectivity on passage deeper into the tunneling region (below the $\beta = (-\alpha)^{\frac{1}{2}}$ dashed curve on the left), and the rapid fall on passage out of it. The transition in or out of tunneling is interesting. In particle terms, there is just enough available kinetic energy for the particle to reach the top of the potential barrier. Classically, it is the transition between total reflection and zero reflection. Quantum mechanically, the exact reflectivity when $\beta = (-\alpha)^{\frac{1}{2}}$ is always less than its large-thickness asymptotic value 1/2. On the right-hand side of the figure, corresponding to positive $\Delta\varepsilon$ or a potential well in the particle case, the contour pattern is caused by the interplay of resonance reflectivity zeros with strong reflectivity at grazing incidence ($\beta \to 0$).

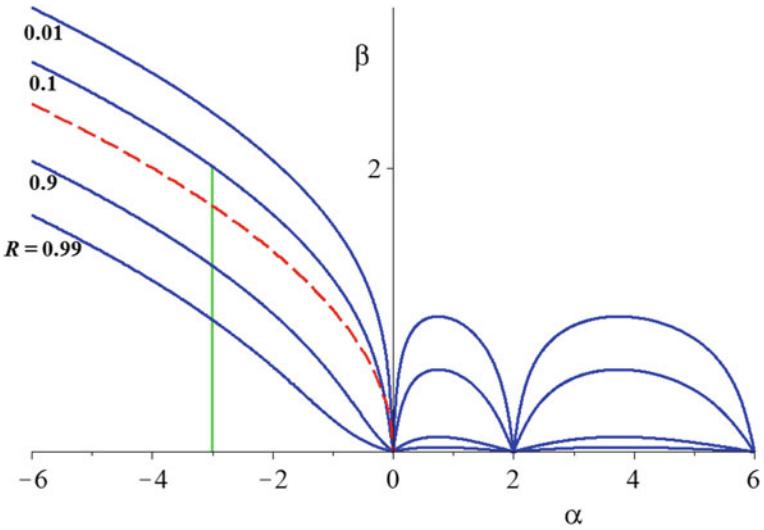


Fig. 6.11 Contours of constant reflectivity for the sech^2 profile. Grazing incidence, corresponding to $\beta = q_0 a \to 0$, generally has high reflectivity. However, for positive α (a positive $\Delta\varepsilon$ or negative ΔV in the particle case) there are resonance zeros of reflectivity when $\alpha = n(n+1)$, $n = 0, 1, 2, \ldots$, as shown in (4.35). The tunneling region is below the *dashed curve* on the *left*, given by $\beta = (-\alpha)^{\frac{1}{2}}$, which is asymptotic to the $R_s = \frac{1}{2}$ contour for large $|\alpha|, \beta$. The q^2 curves of Fig. 6.10 at $\theta = 0°, 30°$ *and* $60°$ correspond to $\beta/(-\alpha)^{\frac{1}{2}} = 2/\sqrt{3}$, 1, and $1/\sqrt{3}$. The *vertical line* at $\alpha = -3$ relates to Fig. 6.12

We shall now derive general but approximate expressions for r_s and t_s when there are two turning points, and then apply these to the sech^2 profile, comparing the reflectivity with (6.141). Away from the turning points, now denoted by z_1 and z_2, the waveforms will be approximated by $q^{-\frac{1}{2}}e^{\pm i\phi}$ for real q, and by $|q|^{-\frac{1}{2}}e^{\pm\Phi}$ for imaginary q (in between the turning points). The real and imaginary parts of the phase integral, ϕ and Φ, are defined by

$$\begin{aligned}
\phi(z) &= \int_{z_1}^{z} \mathrm{d}\zeta\, q(\zeta) + \phi_1 \qquad (z \le z_1),\\
\Phi(z) &= \int_{z_1}^{z} \mathrm{d}\zeta\, |q(\zeta)| + \Phi_1 \qquad (z_1 \le z \le z_2),\\
\phi(z) &= \int_{z_2}^{z} \mathrm{d}\zeta\, q(\zeta) + \phi_1 \qquad (z \ge z_2).
\end{aligned} \tag{6.142}$$

Note that the real part of the phase has the same value ϕ_1 at z_1 and z_2, being continuous across the tunneling interval (only the imaginary part changes). The limiting forms of the real part of ϕ are given by

$$q_0 z + \phi_- \leftarrow \phi(z) \rightarrow q_0 z + \phi_+ \tag{6.143}$$

as $z \to \pm\infty$ (for simplicity we consider the $\varepsilon_1 = \varepsilon_2$ case).

The method used to obtain the reflection and transmission amplitudes is similar to that used in Sect. 6.7, namely matching the approximate wave functions across the turning points. Since the approximate wave functions fail at z_1 and z_2, the matching is via the locally accurate Airy function solutions across z_1 and z_2. Near z_1 the dielectric function is decreasing with z; we approximate q^2 by its leading term linear in $z - z_1$, and define a local variable ζ_1:

$$q^2(z) \approx (z - z_1)\frac{\omega^2}{c^2}\left(\frac{\mathrm{d}\varepsilon}{\mathrm{d}z}\right)_1, \quad \zeta_1 = (z - z_1)\left[\frac{\omega^2}{c^2}\left(-\frac{\mathrm{d}\varepsilon}{\mathrm{d}z}\right)_1\right]^{\frac{1}{3}}. \tag{6.144}$$

The accurate solutions in the neighbourhood of z_1 are then $Ai(\zeta_1)$ and $Bi(\zeta_1)$. Near z_2 the dielectric function is increasing with z and we set

$$q^2(z) \approx (z - z_2)\frac{\omega^2}{c^2}\left(\frac{\mathrm{d}\varepsilon}{\mathrm{d}z}\right)_2, \quad \zeta_2 = (z - z_2)\left[\frac{\omega^2}{c^2}\left(\frac{\mathrm{d}\varepsilon}{\mathrm{d}z}\right)_2\right]^{\frac{1}{3}}. \tag{6.145}$$

The accurate solutions in the neighbourhood of z_2 are $Ai(\zeta_2)$ and $Bi(\zeta_2)$. We again assume the existence of regions of overlap, where both the approximate solutions and the asymptotic forms of the Airy functions hold simultaneously. The condition for the existence of such regions was discussed in the last section. We use the asymptotic forms (6.125), the relations

$$\begin{aligned} &\frac{2}{3}(-\zeta_1)^{\frac{3}{2}} \approx \phi_1 - \phi \quad (z<z_1), \quad \frac{2}{3}(\zeta_1)^{\frac{3}{2}} \approx \Phi - \Phi_1 \quad (z > z_1), \\ &\frac{2}{3}(-\zeta_2)^{\frac{3}{2}} \approx \Phi_2 - \Phi \quad (z<z_2), \quad \frac{2}{3}(\zeta_2)^{\frac{3}{2}} \approx \phi - \phi_1 \quad (z > z_2), \end{aligned} \tag{6.146}$$

and match at four places (to the left and to the right of both z_1 and z_2). After removal of all common factors, the four matchings give

$$\begin{aligned} &\mathrm{e}^{i(\phi-\phi_-)} + r_s \mathrm{e}^{-i(\phi-\phi_-)} = A_1 \sin(\phi_1 - \phi + \pi/4) + A_1 \cos(\phi_1 - \phi + \pi/4), \\ &\frac{1}{2}A_1 \mathrm{e}^{\Phi_1 - \Phi} + B_1 \mathrm{e}^{\Phi - \Phi_1} = A\mathrm{e}^{-\Phi} + B\mathrm{e}^{\Phi}, \\ &A\mathrm{e}^{-\Phi} + B\mathrm{e}^{\Phi} = \frac{1}{2}A_2 \mathrm{e}^{\Phi - \Phi_2} + B_2 \mathrm{e}^{\Phi_2 - \Phi}, \\ &A_2 \sin(\phi - \phi_1 + \pi/4) + A_2 \cos(\phi - \phi_1 + \pi/4) = t_s \mathrm{e}^{i\left(\phi - \phi_+\right)}. \end{aligned} \tag{6.147}$$

At each point we equate the coefficients of $e^{\pm i\phi}$ or $e^{\pm\Phi}$. Thus we have eight conditions to determine the eight coefficients $r_s, A_1, B_1, A, B, A_2, B_2, t_s$. The result of solving for r_s and t_s is

$$r_s = \mathrm{e}^{2i(\phi_1 - \phi_- - \pi/4)} \tanh(\Delta\Phi + \ln 2) \tag{6.148}$$

$$t_s = \mathrm{e}^{i(\phi_+ - \phi_-)} \operatorname{sech}(\Delta\Phi + \ln 2), \tag{6.149}$$

where

$$\Delta\Phi = \Phi_2 - \Phi_1 = \int_{z_1}^{z_2} \mathrm{d}z |q(z)|. \tag{6.150}$$

(The ln 2 comes from the factor $\frac{1}{2}$ multiplying A_1 and A_2 in the tunneling region, which in turn comes from the first equation in (6.125)). Note that $1 - \tanh^2 x = \operatorname{sech}^2 x$, so that the conservation law $1 - |r_s|^2 = |t_s|^2$ of Sect. 2.1 is satisfied by the short-wave approximation (6.148), (6.149) for the reflection and transmission amplitudes. When $\Delta\Phi$ is large,

$$|r_s|^2 \to 1 - \mathrm{e}^{-2\Delta\Phi}, \quad |t_s|^2 \to \mathrm{e}^{-2\Delta\Phi}. \tag{6.151}$$

$\Delta\Phi$ may be evaluated analytically for the sech^2 profile. We have, from (6.135) and (6.140),

$$\Delta\Phi = \int_{z_1}^{z_2} \mathrm{d}z|q(z)| = 2\int_0^{x_0} \mathrm{d}x\left[|\alpha|\mathrm{sech}^2 x - \beta^2\right]^{\frac{1}{2}}, \tag{6.152}$$

with x_0 being the value of $x = z/a$ for which the integrand is zero; this corresponds to z_0 as given by (6.139). The substitution $|\alpha|\mathrm{sech}^2 x = \beta^2 y^2$ reduces (6.152) to elementary form, leading to

$$\Delta\Phi = \pi(|\alpha|^{\frac{1}{2}} - \beta). \tag{6.153}$$

This expression holds for negative α, with, $-\alpha > \beta^2$. When $-\alpha = \beta^2 (\theta = \theta_t)$ (see (6.138)) $\Delta\Phi$ is zero, the turning points having merged at the origin. This is the transition between tunneling and no tunneling discussed in relation to Fig. 6.11. The approximate treatment given above then fails, since it was based on the assumption of well-separated turning points. Nevertheless, we note that when $\Delta\Phi = 0$ the reflection amplitude according to (6.148) has modulus equal to 3/5.

When $|\alpha|$ and β are both large, the exact reflectivity formula (6.141) gives

$$R_s \to \frac{1}{1 + \mathrm{e}^{-2\Delta\Phi}}, \tag{6.154}$$

which is in agreement with (6.151) provided $\Delta\Phi$ is large. When $|\alpha|$ is large but β small (grazing incidence), (6.141) leads to

$$R_s \to \frac{1}{1 + (2\pi\beta)^2 \mathrm{e}^{-2\pi|\alpha|^{\frac{1}{2}}}}. \tag{6.155}$$

This is not in agreement with (6.151), the latter having been based on short wave approximate waveforms which fail at grazing incidence. Thus a large $\Delta\Phi$ is not a guarantee of the accuracy of (6.151). The approximate reflectance and transmittance

$$R_a \approx \tanh^2(\Delta\Phi + \ln 2), \quad T_a \approx \mathrm{sech}^2(\Delta\Phi + \ln 2), \tag{6.156}$$

are compared with the exact R_e of (6.141) and $T_e = 1 - R_e$ in Fig. 6.12.

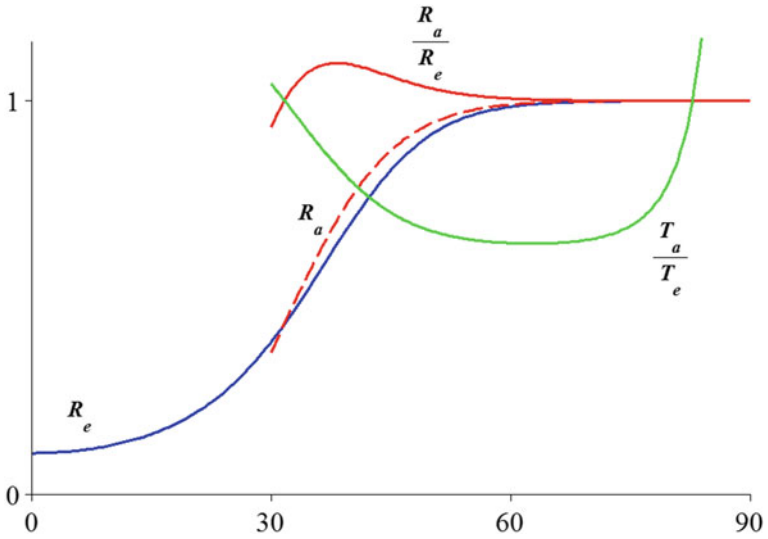


Fig. 6.12 Reflectivity of the sech^2 profile as a function of the angle of incidence, shown for $\varepsilon_0 = 1,\ \Delta\varepsilon = -\frac{3}{4},\ \omega a/c = 2$. These dielectric function values are as for Fig. 6.10, and together with $\omega a/c = 2$ correspond to the *vertical line* shown in Fig. 6.11, with angle of incidence increasing downwards. The exact reflectivity R_e ((6.141), *solid curve*) is shown for all angles, while the approximate reflectivity R_a, given by (6.153) and (6.156) is shown by a *dashed curve* in its range of applicability, $\theta \geq \theta_t = 30°$. The ratios R_a/R_e, and $T_a/T_e = (1 - R_a)/(1 - R_e)$ are also shown. The latter ratio demonstrates the poor accuracy obtained for the tunneling probability, especially near grazing incidence

References

Atkinson FV (1960) Wave propagation and the Bremmer series. J Math Anal Appl 1:255–276
Bellman R, Kalaba R (1959) Functional equations, wave propagation and invariant imbedding. J Math Mech. 8:683–704
Berry MV, Mount KE (1972) Semiclassical approximations in wave mechanics. Rep Prog Phys 35:315–397
Bremmer H (1951) The W. K. B. approximation as the first term of a geometric-optical series. Commun Pure Appl Math 4:105–115
Brillouin L (1926) Remarques sur la méchanique ondulatoire. J Phys Radium 7:353–368
Budden KG (1985) The propagation of radio waves. Cambridge University Press, New York
Green G (1837) On the motion of waves in a variable canal of small depth and width. Trans Camb Phil Soc 6:457–462
Gans R (1915) Fortpflanzung des Lichts durch ein inhomogenes Medium. Ann Physik 47:709–736
Hartree DR (1931) Optical and equivalent paths in a stratified medium treated from a wave standpoint. Proc Roy Soc A131:428–450
Heading J (1962) An introduction to phase-integral methods. Methuen, London
Jeffreys H (1924) On certain approximate solutions of linear differential equations of the second order. Proc London Math Soc 23:428–436
Kittel C (1976) Introduction to solid state physics, 5th edn. Wiley, New York
Kramers HA (1926) Wellenmechanik und halbzahlige Quantisierung. Z Physik 39:828–840

Lekner J (1990) Reflection of waves by a profile with discontinuities. J Phys A: Math Gen 23:2897–2904
Lekner J (1995) Neutron reflection interferometry: extraction of the phase in total reflection from stratified media. Phys B 215:329–336
Lekner J (1996) Analysis of Lloyd's mirror fringes for graded-index reflectors. J Opt Soc Am 13:1809–1815
Liouville J (1837) Sur le développement des fonctions ou parties de fonctions en séries. J Math Pures Appl 2:16–35
Olver FWJ (1974) Asymptotics and special functions. Academic Press, New York, p 228
Olver FWJ (2010) Airy and related functions. In: Olver FWJ et al (eds) NIST handbook of mathematical functions. Cambridge University Press, Cambridge (Chapter 9)
Pokrovskii VL, Savvinykh KS, Ulinich FR (1958) Reflection from a barrier in the quasiclassical approximation. Sov Phys JETP 34:879–882, 1119–1120
Rayleigh JWS (1912) On the propagation of waves through a stratified medium with special reference to the question of reflection. Proc Roy Soc A86:207–266
Wentzel G (1926) Eine Verallgemeinerung der Quantenbedingungen für die Zwecke der Wellenmechanik. Z Physik 38:518–529

Further Readings

Short wave perturbation theory (Sect. 6.5) is also developed in

Levine H (1978) Unidirectional wave motions. North-Holland, Amsterdam

Additional references for Sect. 6.7

Ginzburg VL (1964) The propagation of electromagnetic waves in plasmas. Pergamon, Oxford (Section 30)
Heading J (1975) Ordinary differential equations, theory and practice. Elek Science, London (Section 7.12)
Landau LD, Lifshitz EM (1965) Quantum mechanics, 2nd edn. Pergamon, Oxford (Section 52)

The results on tunneling date from the early days of quantum mechanics. See

Kemble EC (1937) The fundamental principles of quantum mechanics. McGraw-Hill, New York (Section 21j)

Alternative derivations are given by Ginzburg (1964), Section 33, Landau and Lifshitz (1965), Section 50, and

Froman N, Froman PO (1965) JWKB approximation, contributions to the theory. North-Holland, Amsterdam (Section 9.1)
Merzbacher E (1998) Quantum mechanics, 3rd edn. Wiley, New York (Section 7.4)

Chapter 7
Simple Anisotropy

Up till now we have assumed the electrodynamics of a non-magnetic stratified system to be characterized by a single dielectric function $\varepsilon(z)$. This is often a very good approximation: for example in the case of monatomic fluids, where a liquid-vapour interface needs two dielectric functions $\varepsilon_o(z)$ and $\varepsilon_e(z)$ for specification of the electrodynamics, the difference between these is small (Lekner 1983). On the other hand molecular liquids can have strong anisotropy due to orientation of the molecules, extreme examples being liquid crystals. Historically the first anisotropy noted was that of Iceland spar (calcite), and Huygens in his *Treatise on Light* (1690) gives a remarkably prescient discussion of the possible molecular (spheroidal corpuscles, as Huygens termed them) arrangements in anisotropic crystals. Simple examples of reflection in the presence of anisotropy will be discussed here, with emphasis on the interplay of anisotropy and stratification in their effect on reflectivities and ellipsometric measurements. The first five sections deal with a special case of uniaxial anisotropy, where the optic axis is normal to the reflecting surface (the system has azimuthal symmetry on rotation about the surface normal). The last Section discusses anisotropy in ionospheric propagation of radio waves due to the earth's magnetic field. A full treatment of uniaxial anisotropy will be given in Chap. 8.

7.1 Anisotropy with Azimuthal Symmetry

When the reflecting system is symmetric with respect to rotation about the normal to the interface (azimuthal symmetry), the electrodynamics is characterized by $\varepsilon_o(z,\omega)$ and $\varepsilon_e(z,\omega)$, corresponding to the electric field vector aligned respectively along and perpendicular to the interface. The convention used throughout this book is that the interface lies parallel to the xy plane, and propagation is in the zx plane; isotropic systems have $\varepsilon_o = \varepsilon_e$. Examples of the systems discussed in Sects. 7.1–7.5 are: uniaxial crystals with the optic axis along the surface normal, and the surfaces of atomic fluids or of molecular fluids in which surface molecular orientation is symmetric about the surface normal.

J. Lekner, *Theory of Reflection*, Springer Series on Atomic, Optical, and Plasma Physics 87, DOI 10.1007/978-3-319-23627-8_7

Symmetry with respect to rotation about the surface normal conserves the s and p wave characterizations: these two polarizations, with $\mathbf{E} = (0, E_y, 0)$ and $\mathbf{B} = (0, B_y, 0)$ respectively, are together sufficient to represent any plane wave incident onto such an anisotropic planar stratified medium. To derive equations for the s and p waves we repeat the analysis of Sects. 1.1 and 1.2, with (1.1) unchanged, and the dielectric function ε in (1.2) now to be interpreted as the diagonal tensor

$$\varepsilon = \begin{pmatrix} \varepsilon_o & 0 & 0 \\ 0 & \varepsilon_o & 0 \\ 0 & 0 & \varepsilon_e \end{pmatrix}. \tag{7.1}$$

For the s wave (1.1) and (1.2) give

$$i\frac{\omega}{c}B_x = -\frac{\partial E_y}{\partial z}, \quad B_y = 0, \quad i\frac{\omega}{c}B_z = \frac{\partial E_y}{\partial x}, \tag{7.2}$$

$$-i\frac{\omega}{c}\varepsilon_o E_y = \frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x}. \tag{7.3}$$

On eliminating B_x and B_z from (7.2) and (7.3) we find

$$\frac{\partial^2 E_y}{\partial z^2} + \frac{\partial^2 E_y}{\partial x^2} + \varepsilon_o \frac{\omega^2}{c^2} E_y = 0. \tag{7.4}$$

Since the system retains invariance with respect to translation in the x or y directions, the x dependence of E_y is contained in the factor e^{iKx} as before,

$$E_y(z, x, t) = \mathrm{e}^{i(Kx-\omega t)} E(z). \tag{7.5}$$

Substitution of (7.5) into (7.4) gives the usual form for the s wave equation,

$$\frac{\mathrm{d}^2 E}{\mathrm{d}z^2} + q_s^2 E = 0, \quad q_s^2 = \varepsilon_o \frac{\omega^2}{c^2} - K^2. \tag{7.6}$$

Thus all the results we have derived in the last six chapters for the s wave also apply to the s wave in the presence of azimuthally symmetric anisotropy, with the replacement of $\varepsilon(z)$ by $\varepsilon_o(z)$.

The p wave is more complicated, since it samples (at a general angle of incidence) both ε_o and ε_e. The Maxwell equation (1.1) gives

$$i\frac{\omega}{c}B_y = \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x}, \tag{7.7}$$

as before, but (1.2) now implies

$$i\frac{\omega}{c}\varepsilon_o E_x = \frac{\partial B_y}{\partial z}, \quad E_y = 0, \quad -i\frac{\omega}{c}\varepsilon_e E_z = \frac{\partial B_y}{\partial x}. \tag{7.8}$$

Elimination of E_x and E_z from (7.7) and (7.8) gives

$$\frac{\partial}{\partial z}\left(\frac{1}{\varepsilon_o}\frac{\partial B_y}{\partial z}\right) + \frac{\partial}{\partial x}\left(\frac{1}{\varepsilon_e}\frac{\partial B_y}{\partial x}\right) + \frac{\omega^2}{c^2}B_y = 0. \tag{7.9}$$

The substitution

$$B_y(z, x, t) = \mathrm{e}^{i(Kx-\omega t)}B(z) \tag{7.10}$$

gives us a modified p wave equation,

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{\varepsilon_o}\frac{\mathrm{d}B}{\mathrm{d}z}\right) + \left(\frac{\omega^2}{c^2} - \frac{K^2}{\varepsilon_e}\right)B = 0. \tag{7.11}$$

This differs fundamentally from the isotropic case: (7.11) contains both dielectric functions, and thus results previously obtained for the p wave cannot be used directly to obtain results even for this restricted anisotropy.

Anisotropy implies birefringence (double refraction): consider the example of an electromagnetic wave incident onto a homogeneous azimuthally symmetric anisotropic material. The incident and reflected waves have the z dependence $\mathrm{e}^{\pm iq_1 z}$ as usual, but the s and p transmitted waves have the z dependence $\mathrm{e}^{\mathrm{i}q_s z}$ and $\mathrm{e}^{\mathrm{i}q_p z}$, where q_s is given in (7.6) and, from (7.11),

$$q_p^2 = \varepsilon_o\frac{\omega^2}{c^2} - \frac{\varepsilon_o}{\varepsilon_e}K^2. \tag{7.12}$$

Since q_p differs from q_s, the angles θ_p and θ_s giving the *wavevector directions* of the s and p waves are different: $K = q_1 \tan\theta_1 = q_s \tan\theta_s = q_p \tan\theta_p$, and from (7.6) and (7.12),

$$\varepsilon_1 \sin^2\theta_1 = \varepsilon_o \sin^2\theta_s = \frac{\varepsilon_o\varepsilon_e \sin^2\theta_p}{\varepsilon_o \sin^2\theta_p + \varepsilon_e \cos^2\theta_p} \tag{7.13}$$

We shall see in Chap. 8 that for the s (or ordinary) wave the wavevector direction is always the same as the *ray direction*, given by the energy flux or Poynting vector $\mathbf{E} \times \mathbf{B}$. For the p (or extraordinary) wave, the wavevector and $\mathbf{E} \times \mathbf{B}$ directions are different. The latter, when the optic axis is normal to the reflecting plane, is given by

$$\tan\theta'_p = \frac{\varepsilon_o K}{\varepsilon_e q_p} = \frac{\varepsilon_o}{\varepsilon_e}\tan\theta_p \tag{7.14}$$

We shall see in Sect. 8.2 that the extraordinary ray and wavevector are coplanar with the optic axis, which here coincides with the normal to the interface. Figure 7.1 shows the wavevector directions; the extraordinary wave ray vector is also shown. The differences between the three directions is enhanced by index matching: the Figure is drawn to scale for calcite ($n_o = 1.658, n_e = 1.486$, at the sodium yellow line), immersed in an oil of index $n_1 = 1.48$. The angle of incidence is 60°. The magnitudes of the normal components of the wavevectors are shown on the horizontal axis.

The s and p reflection amplitudes for a sharp interface between an isotropic medium and an anisotropic medium, characterized by ε_o and ε_e as above, are (for an interface at $z = 0$)

$$r_{s0} = \frac{q_1 - q_s}{q_1 + q_s}, \quad -r_{p0} = \frac{Q_1 - Q}{Q_1 + Q}, \tag{7.15}$$

where q_s is given by (7.6), $Q_1 = q_1/\varepsilon_1$, and $Q = q_p/\varepsilon_o$. The p wave reflection amplitude follows from the continuity of B and $(1/\varepsilon_o)(\mathrm{d}B/\mathrm{d}z)$ at the boundary, a

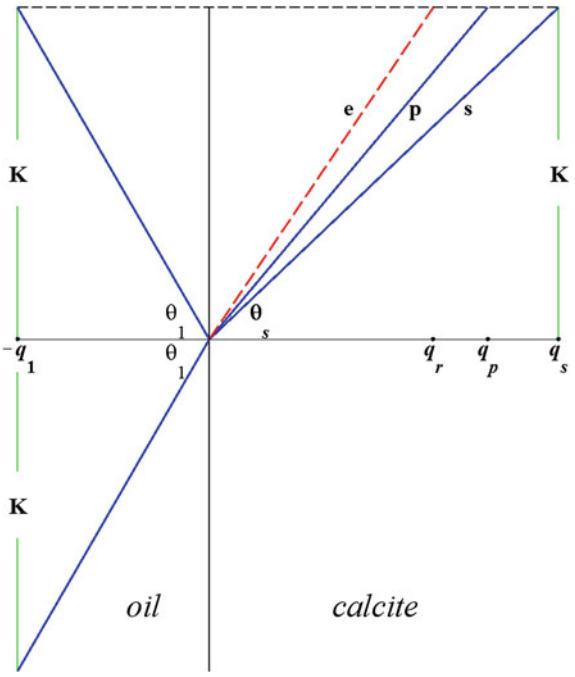


Fig. 7.1 Wavevector and ray directions in an azimuthally symmetric anisotropic system, such as a uniaxial crystal with its optic axis coincident with the z-axis. The wavevectors are *solid lines*, the extraordinary ray vector is the *dashed line*

consequence of the differential equation (7.11). The Brewster angle θ_B, at which $r_p = 0$, is given by $Q_1 = Q$, which leads to

$$\theta_B = \arctan\left\{\frac{\varepsilon_e(\varepsilon_o - \varepsilon_1)}{\varepsilon_1(\varepsilon_e - \varepsilon_1)}\right\}^{1/2}. \tag{7.16}$$

7.2 Ellipsometry of a Thin Film on an Isotropic Substrate

Ellipsometry determines the complex number r_p/r_s. (Chapter 9 gives details of various ellipsometric confugurations, and what they measure with general anisotropy; for reflection from isotropic media, and for the simple anisotropy considered in this chapter, these all give r_p/r_s). In Sect. 3.5, equation (3.52), we saw that, in the isotropic case, this ratio is given to lowest order in the interface thickness/wavelength expansion by

$$r_{s0}\left(\frac{r_p}{r_s}\right) = r_{p0} - \frac{\frac{2iQ_1K^2}{\varepsilon_1\varepsilon_2}}{(Q_1+Q_2)^2} I_1 + \cdots, \tag{7.17}$$

The invariant I_1 is given by

$$I_1 = \int_{-\infty}^{\infty} \mathrm{d}z \frac{(\varepsilon_1 - \varepsilon)(\varepsilon - \varepsilon_2)}{\varepsilon} = \int_{-\infty}^{\infty} \mathrm{d}z \left\{\varepsilon_1 + \varepsilon_2 - \frac{\varepsilon_1\varepsilon_2}{\varepsilon} - \varepsilon\right\}. \tag{7.18}$$

For the case of an anisotropic film characterized by ε_o and ε_e (with cylindrical symmetry about the surface normal) and resting on an isotropic substrate with dielectric constant ε_2, we shall show that the formula (7.17) remains valid, with

$$I_1 = \int_{-\infty}^{\infty} \mathrm{d}z \left\{\varepsilon_1 + \varepsilon_2 - \frac{\varepsilon_1\varepsilon_2}{\varepsilon_e} - \varepsilon_o\right\}. \tag{7.19}$$

This generalization was given (without proof) by Buff (1966) and independently by Beaglehole (1980), the latter crediting Abelès (1976). The proof given here follows Lekner (1983), Appendix A. We begin by deriving an anisotropic generalization of the comparison identity (3.42). Let $\varepsilon_0(z)$ be the step function $\varepsilon_0 = \varepsilon_1$ for $z < 0$, $\varepsilon_0 = \varepsilon_2$ for $z > 0$, and $B_0(z)$ the solution of

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{\varepsilon_0}\frac{\mathrm{d}B_0}{\mathrm{d}z}\right) + \left(\frac{\omega^2}{c^2} - \frac{K^2}{\varepsilon_0}\right)B_0 = 0, \tag{7.20}$$

namely

$$B_0(z) = \begin{cases} e^{iq_1 z} - r_{p0} e^{-iq_1 z} (z < 0) \\ \left(\frac{\varepsilon_2}{\varepsilon_1}\right)^{1/2} t_{p0} e^{iq_2 z} (z > 0). \end{cases} \tag{7.21}$$

$B(z)$ is the solution of (7.11), with limiting forms

$$e^{iq_1 z} - r_p e^{-iq_1 z} \leftarrow B(z) \rightarrow \left(\frac{\varepsilon_2}{\varepsilon_1}\right)^{\frac{1}{2}} t_p e^{iq_2 z}. \tag{7.22}$$

We multiply (7.11) by $B_0(z)$, (7.20) by $B(z)$, and subtract. The result is

$$\frac{\mathrm{d}}{\mathrm{d}z}(B_0 C - B C_0) = K^2 \left(\frac{1}{\varepsilon_e} - \frac{1}{\varepsilon_0}\right) B B_0 - (\varepsilon_o - \varepsilon_0) C C_0, \tag{7.23}$$

where (please note the different subscripts 0 and o)

$$C_0 = \frac{1}{\varepsilon_0} \frac{\mathrm{d}B_0}{dz}, \quad C = \frac{1}{\varepsilon_o} \frac{\mathrm{d}B}{\mathrm{d}z}. \tag{7.24}$$

Integration of (7.23) from z_1 (deep in medium 1) to z_2 (deep in medium 2) and use of (7.21) and (7.22) then gives the identity

$$r_p = r_{p0} + \frac{1}{2iQ_1} \int_{-\infty}^{\infty} \mathrm{d}z \left\{ \left(\frac{1}{\varepsilon_0} - \frac{1}{\varepsilon_e}\right) K^2 B B_0 + (\varepsilon_o - \varepsilon_0) C C_0 \right\}. \tag{7.25}$$

To lowest order in the interface thickness we may replace B by $B_0(0)$ and C by $C_0(0)$, as given by (3.43). The result is

$$r_p = r_{p0} - \frac{2iQ_1}{(Q_1 + Q_2)^2} \left\{ K^2 \int_{-\infty}^{\infty} \mathrm{d}z \left(\frac{1}{\varepsilon_0} - \frac{1}{\varepsilon_e}\right) - Q_2^2 \int_{-\infty}^{\infty} \mathrm{d}z (\varepsilon_o - \varepsilon_0) \right\} + \cdots \tag{7.26}$$

The s wave reflection amplitude is given by (3.18) on replacing ε by ε_o:

$$r_s = r_{s0} + \frac{2iq_1 \omega^2/c^2}{(q_1 + q_2)^2} \int_{-\infty}^{\infty} \mathrm{d}z (\varepsilon_o - \varepsilon_0) + \cdots \tag{7.27}$$

The ellipsometric ratio r_p/r_s may now be found from (7.26), (7.27), and (3.45):

$$r_{s0}\left(\frac{r_p}{r_s}\right) = r_{p0} - \frac{2iQ_1K^2}{(Q_1+Q_2)^2}\left\{\int_{-\infty}^{\infty} \mathrm{d}z\left(\frac{1}{\varepsilon_0}-\frac{1}{\varepsilon_e}\right) - \frac{1}{\varepsilon_1\varepsilon_2}\int_{-\infty}^{\infty} \mathrm{d}z(\varepsilon_o-\varepsilon_0)\right\} + \cdots \tag{7.28}$$

The result (7.17, 7.19) follows from generalization of the identity (3.37):

$$\varepsilon_1\varepsilon_2\int_{-\infty}^{\infty} \mathrm{d}z\left(\frac{1}{\varepsilon_0}-\frac{1}{\varepsilon_e}\right) - \int_{-\infty}^{\infty} \mathrm{d}z(\varepsilon_o-\varepsilon_0) = \int_{-\infty}^{\infty} \mathrm{d}z\left\{\varepsilon_1+\varepsilon_2-\frac{\varepsilon_1\varepsilon_2}{\varepsilon_e}-\varepsilon_o\right\}. \tag{7.29}$$

Another two equivalent ways of writing I_1 (given by (7.29)) are

$$\begin{aligned} I_1 &= \int_{-\infty}^{\infty} \mathrm{d}z\frac{(\varepsilon_1-\varepsilon_e)(\varepsilon_e-\varepsilon_2)}{\varepsilon_e} + \int_{-\infty}^{\infty} \mathrm{d}z(\varepsilon_e-\varepsilon_o) \\ &= \int_{-\infty}^{\infty} \mathrm{d}z\frac{(\varepsilon_1-\varepsilon_o)(\varepsilon_o-\varepsilon_2)}{\varepsilon_o} + \varepsilon_1\varepsilon_2\int_{-\infty}^{\infty} \mathrm{d}z\left(\frac{1}{\varepsilon_o}-\frac{1}{\varepsilon_e}\right). \end{aligned} \tag{7.30}$$

The first form of the isotropic I_1 in (7.18) shows that I_1 is positive when ε lies between ε_1 and ε_2. This is not necessarily so in the presence of anisotropy, even if both ε_o and ε_e lie between ε_1 and ε_2. For example, if on average ε_e is smaller than ε_o, the first form of (7.30) shows that I_1 may be negative for sufficiently large anisotropy. In general, if $\varepsilon_e < \varepsilon_o$ on average, the anisotropy will give the appearance of a thinner film (a smaller I_1), or can even make I_1 negative. Conversely, if $\varepsilon_e > \varepsilon_o$ on average, the anisotropy will increase I_1, giving the same signal as a thicker isotropic film. For a homogeneous anisotropic film we can be more definite: if Δz is the film thickness, $I_1/\Delta z = \varepsilon_1+\varepsilon_2-\varepsilon_o-\varepsilon_1\varepsilon_2/\varepsilon_e$, and is positive provided the sum of ε_o and $\varepsilon_1\varepsilon_2/\varepsilon_e$ is less than the sum of ε_1 and ε_2. The contours of constant $I_1/\Delta z$ are lines of slope -1 in the $(\varepsilon_o, \varepsilon_1\varepsilon_2/\varepsilon_e)$ plane. In the same plane, the contours of fixed anisotropy $\varepsilon_e-\varepsilon_o$ are also shown in Fig. 7.2.

We see that $I_1/\Delta z$ increases with anisotropy. The $\varepsilon_e=\varepsilon_o$ contour has a maximum value of $I_1/\Delta z$ equal to $\left(\sqrt{\varepsilon_2}-\sqrt{\varepsilon_1}\right)^2$. (This is reached when the common value of ε_o and ε_e is $\sqrt{\varepsilon_1\varepsilon_2}$.) Thus if I_1 is measured, Δz for the homogeneous film is known independently, and $I_1/\Delta z$ is found to be bigger than $\left(\sqrt{\varepsilon_2}-\sqrt{\varepsilon_1}\right)^2$, the anisotropy must be positive ($\varepsilon_e > \varepsilon_o$).

For a homogeneous anisotropic film, with its optic axis along the surface normal, the exact reflection amplitudes may be found by the methods of Sect. 2.4. The s wave takes the same form as in the isotropic film, with q replaced by q_s, defined in

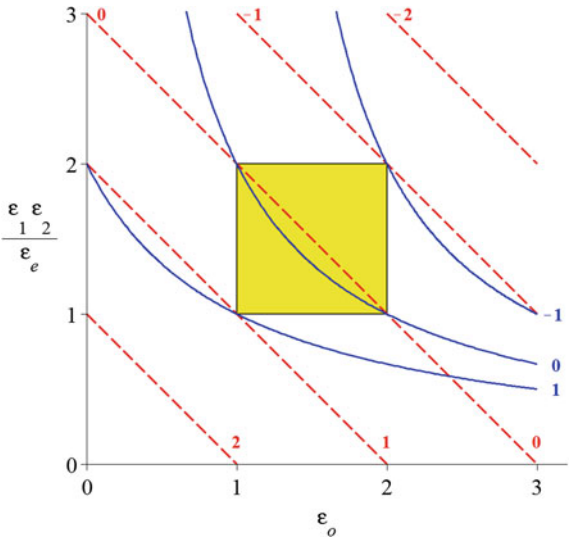


Fig. 7.2 Lines of fixed $I_1/\Delta z$, ranging from -2 to 2, and contours of fixed anisotropy $\varepsilon_e - \varepsilon_o$, ranging from -1 to 1, drawn for a homogeneous anisotropic layer between media with $\varepsilon_1 = 1$, $\varepsilon_2 = 2$, in the $(\varepsilon_o, \varepsilon_1\varepsilon_2/\varepsilon_e)$ plane. The ε_o and ε_e values lie between ε_1 and ε_2 in the *shaded box*

(7.6). For a homogeneous layer located between z_1 and $z_1 + \Delta z$, the s wave reflection amplitude is

$$r_s = \mathrm{e}^{2iq_1z_1}\frac{q_s(q_1 - q_2)c + i\left(q_s^2 - q_1q_2\right)s}{q_s(q_1 + q_2)c - i\left(q_s^2 + q_1q_2\right)s}, \tag{7.31}$$

where $c = \cos q_s\Delta z$ and $s = \sin q_s\Delta z$. For the p wave, the solutions within the film are $\mathrm{e}^{\pm iq_pz}$, with q_p given by (7.12), and the boundary conditions are the continuity of B and of $C = (1/\varepsilon_o)(\mathrm{d}B/\mathrm{d}z)$ at z_1 and z_2. We find that r_p has the same form as (2.68):

$$-r_p = \mathrm{e}^{2iq_1z_1}\frac{Q(Q_1 - Q_2)c + i(Q^2 - Q_1Q_2)s}{Q(Q_1 + q_2)c - i(Q^2 + Q_1Q_2)s}, \tag{7.32}$$

where now $c = \cos q_p\Delta z$, $s = \sin q_p\Delta z$, and $Q = q_p/\varepsilon_o$. From (7.31) and (7.32) we can verify that r_p/r_s takes the form (7.17) with I_1 given by (7.19).

7.3 Thin Film on an Anisotropic Substrate

We now include the possibility of substrate anisotropy, still keeping azimuthal symmetry in both stratified surface and in the homogeneous substrate (and thus retaining the s and p characterization of electromagnetic waves). The electromagnetic response of the system is determined by the three dielectric constants

$\varepsilon_1, \varepsilon_{2o}, \varepsilon_{2e}$ (the latter two for the substrate) and the two dielectric functions $\varepsilon_o(z), \varepsilon_e(z)$ with

$$\varepsilon_1 \leftarrow \varepsilon_o(z) \rightarrow \varepsilon_{2o}, \quad \varepsilon_1 \leftarrow \varepsilon_e(z) \rightarrow \varepsilon_{2e}. \tag{7.33}$$

We now need to define two step functions

$$\varepsilon_{0o}(z), \varepsilon_{0e}(z) = \begin{cases} \varepsilon_1 & (z < 0) \\ \varepsilon_{2o}, \varepsilon_{2e} & (z > 0). \end{cases} \tag{7.34}$$

The results of the previous section for the s wave need only the modification $\varepsilon_0 \rightarrow \varepsilon_{0o}$ (for example in (7.27)). For the p wave we use $B_0(z)$, the solution of

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{\varepsilon_{0o}}\frac{\mathrm{d}B_0}{\mathrm{d}z}\right) + \left(\frac{\omega^2}{c^2} - \frac{K^2}{\varepsilon_{0e}}\right)B_0 = 0. \tag{7.35}$$

The modified version of (7.23) is

$$\frac{\mathrm{d}}{\mathrm{d}z}(B_0 C - B C_0) = K^2\left(\frac{1}{\varepsilon_e} - \frac{1}{\varepsilon_{0e}}\right) - (\varepsilon_o - \varepsilon_{0o})CC_0\,, \tag{7.36}$$

with

$$C_0 = \frac{1}{\varepsilon_{0o}}\frac{\mathrm{d}B_0}{\mathrm{d}z}, \quad C = \frac{1}{\varepsilon_o}\frac{\mathrm{d}B}{\mathrm{d}z}, \tag{7.37}$$

and leads to the comparison identity

$$r_p = r_{p0} + \frac{1}{2iQ_1}\int_{-\infty}^{\infty} \mathrm{d}z\left\{\left(\frac{1}{\varepsilon_{0e}} - \frac{1}{\varepsilon_e}\right)K^2 B B_0 + (\varepsilon_o - \varepsilon_{0o})CC_0\right\}. \tag{7.38}$$

The appropriate values of $B_0(0)$ and $C_0(0)$ are now

$$B_0(0) = \frac{2Q_1}{Q_1 + Q_2}, \quad C_0(0) = \frac{2iQ_1Q_2}{Q_1 + Q_2}, \tag{7.39}$$

where $Q_1 = q_1/\varepsilon_1$, and $Q_2 = q_{2p}/\varepsilon_{2o}$, with

$$q_{2p}^2 = \varepsilon_{2o}\frac{\omega^2}{c^2} - \frac{\varepsilon_{2o}}{\varepsilon_{2e}}K^2. \tag{7.40}$$

To lowest order in the interface thickness,

$$r_p = \frac{Q_2 - Q_1}{Q_2 + Q_1} - \frac{2iQ_1}{(Q_1 + Q_2)^2}\left\{K^2 \int_{-\infty}^{\infty} \mathrm{d}z\left(\frac{1}{\varepsilon_{0e}} - \frac{1}{\varepsilon_e}\right) - Q_2^2 \int_{-\infty}^{\infty} \mathrm{d}z(\varepsilon_o - \varepsilon_{0o})\right\}. \tag{7.41}$$

On combining this with the (7.27) modified by replacing ε_0 by ε_{0o}, we find the generalization of (7.28):

$$r_{s0}\left(\frac{r_p}{r_s}\right) = r_{p0}$$

$$-\frac{2iQ_1K^2}{(Q_1 + Q_2)^2}\left\{\int_{-\infty}^{\infty} \mathrm{d}z\left(\frac{1}{\varepsilon_{0e}} - \frac{1}{\varepsilon_e}\right) - \left(\frac{1}{\varepsilon_1} - \frac{1}{\varepsilon_{2e}}\right)(\varepsilon_{2o} - \varepsilon_1)^{-1} \int_{-\infty}^{\infty} \mathrm{d}z(\varepsilon_o - \varepsilon_{0o})\right\} + \cdots \tag{7.42}$$

The expression in braces may be written as

$$\frac{1}{\varepsilon_1 \varepsilon_{2e}} \int_{-\infty}^{\infty} \mathrm{d}z\left[\frac{\varepsilon_{2o}\varepsilon_{2e} - \varepsilon_1^2}{\varepsilon_{2o} - \varepsilon_1} - \frac{\varepsilon_{2e} - \varepsilon_1}{\varepsilon_{2o} - \varepsilon_1}\varepsilon_o - \frac{\varepsilon_{2e}\varepsilon_1}{\varepsilon_e}\right], \tag{7.43}$$

and thus the form of (7.17) is retained:

$$r_{s0}\left(\frac{r_p}{r_s}\right) = r_{p0} - \frac{\frac{2iQ_1K^2}{\varepsilon_1\varepsilon_{2e}}}{(Q_1 + Q_2)^2} I_1 + \cdots, \tag{7.44}$$

with I_1 now being given by the integral in (7.43).

For a homogeneous anisotropic film on an anisotropic substrate, the formulae (7.31) and (7.32) remain valid, with q_2 being interpreted as q_{2s}, and Q_2 as q_{2p}/ε_{2o}. For a homogeneous film of thickness Δz, $I_1/\Delta z$ is equal to the content of the square bracket in (7.43).

7.4 General Results for Anisotropic Stratifications with Azimuthal Symmetry

In the previous two sections we have derived results for the ellipsometric properties of *thin* anisotropic films on isotropic and anisotropic substrates, respectively. Here we examine general properties of reflection by anisotropic stratified media, still

keeping the constraint of cylindrical symmetry about the normal to the stratification (the z axis). As in Sect. 2.2 we take the inhomogeneous or interfacial region to be within the interval $[z_1, z_2]$; the results for such finite-ranged interfaces can be extended to continuously varying unbounded interfaces by a limiting process.

For the s wave the results of Sect. 2.2 follow directly, since anisotropy modifies the s wave equations only by the replacement of $\varepsilon(z)$ by $\varepsilon_o(z)$. Thus the general expressions for the reflection and transmission amplitudes, (2.25) and (2.26), remain valid, with q_1 and q_2 being the values of $q_s = (\varepsilon_o \omega^2/c^2 - K^2)^{\frac{1}{2}}$ for $z \leq z_1$ and $z \geq z_2$. In consequence, the result $|r_s| \leq 1$, the conservation law $q_1\left(1 - |r_s|^2\right) = q_2|t_s|^2$, and the reciprocity laws all follow. The result that $r_s \to -1$ at grazing incidence (Sect. 2.3) also holds, as does the inequality $|r_s| \leq |r_{s0}|$ for monotonic profiles (Sect. 5.4).

The p wave case involves both dielectric functions $\varepsilon_o(z)$ and $\varepsilon_e(z)$, which take the values ε_1 for $z \leq z_1$ and $\varepsilon_{2o}, \varepsilon_{2e}$ for $z \geq z_2$. $B(z)$, the solution of (7.11), now has the limiting forms

$$B(z) = \mathrm{e}^{iq_1 z} - r_p \mathrm{e}^{-iq_1 z} (z \leq z_1), \quad B(z) = \left(\frac{\varepsilon_{2o}}{\varepsilon_1}\right)^{1/2} t_p \mathrm{e}^{iq_{2p} z} (z \geq z_2) \tag{7.45}$$

The sign of r_p and the factor $\sqrt{\varepsilon_{2o}/\varepsilon_1}$ multiplying t_p are chosen to make r_p and r_s, and t_p and t_s all apply to electric field components, and to agree at normal incidence. The electric field components for the p wave are found from (7.8); the effect of anisotropy of the substrate is the replacement of ε_2 by ε_{2e} in the square root of the ratio multiplying t_p (compare with Sect. 1.2).

For profiles which have ε_o continuous at z_1 and z_2, the equations (2.40) and (2.41) remain valid, with q_2 replaced by q_{2p} and $\sqrt{\varepsilon_2/\varepsilon_1}$ by $\sqrt{\varepsilon_{2o}/\varepsilon_1}$. The Wronskian of two solutions of (7.11) is now proportional to ε_o (compare with (2.47) and (2.48)). Again $|r_p|^2 \leq 1$ and the reciprocity relations remain valid, with q_2 replaced by q_{2p}. The conservation law for the p wave now reads

$$q_1\left(1 - |r_p|^2\right) = q_{2p}|t_p|^2. \tag{7.46}$$

The range of validity of the inequality $|r_p|^2 \leq |r_{p0}|^2$ will be examined in the next section.

7.5 Differential Equations for the Reflection Amplitudes

We shall briefly examine some of the consequences of the non-linear first order differential equations of Chap. 5. The s wave need not be considered in detail, since all results remain valid on the replacement of $\varepsilon(z)$ by $\varepsilon_o(z)$. In the p wave case we set

$$\frac{1}{\varepsilon_o}\frac{\mathrm{d}B}{\mathrm{d}z} = C, \quad \frac{\mathrm{d}C}{\mathrm{d}z} = -\frac{q_p^2}{\varepsilon_o}B\,, \tag{7.47}$$

in analogy with (5.32). This pair of coupled first order equations is equivalent to (7.11), with

$$q_p^2 = \varepsilon_o\frac{\omega^2}{c^2} - \frac{\varepsilon_o}{\varepsilon_e}K^2. \tag{7.48}$$

The anisotropic version of (5.33) is

$$B = F + G, \quad C = i\frac{q_p}{\varepsilon_o}(F - G). \tag{7.49}$$

Elimination of B and C gives equations of the same form as (5.34) and (5.35),

$$F' = iq_pF - \frac{Q'}{2Q}(F - G), \tag{7.50}$$

$$G' = -iq_pG + \frac{Q'}{2Q}(F - G), \tag{7.51}$$

where now $Q = q_p/\varepsilon_o$. The reflection coefficient $\rho = G/F$ (as distinct from the reflection amplitude to be discussed shortly) satisfies the equation

$$\rho' + 2iq_p\rho - \frac{Q'}{2Q}\left(1 - \rho^2\right) = 0. \tag{7.52}$$

We write $\rho = |\rho|\mathrm{e}^{i\theta}$; the absolute magnitude $|\rho|$ satisfies

$$|\rho|' = \frac{Q'}{2Q}\left(1 - |\rho|^2\right)\cos\theta. \tag{7.53}$$

Integration of (7.53) gives the exact result

$$\ln\frac{1 + |r_p|}{1 - |r_p|} = -\int_{-\infty}^{\infty}\mathrm{d}z\frac{Q'}{Q}\cos\theta. \tag{7.54}$$

For *monotonic* Q the inequality

$$R_p \le \left(\frac{Q_1 - Q_2}{Q_1 + Q_2}\right)^2 \tag{7.55}$$

follows, with $Q_1 = q_1/\varepsilon_1$ and $Q_2 = q_{2p}/\varepsilon_{2o}$. We have

$$Q^2 = \frac{q_p^2}{\varepsilon_o^2} = \frac{1}{\varepsilon_o}\frac{\omega^2}{c^2} - \frac{K^2}{\varepsilon_o \varepsilon_e}, \tag{7.56}$$

and whether Q is monotonic or not depends on the variation of both ε_o and ε_e with z, as well as on the angle of incidence.

In Chap. 5 the useful Rayleigh approximations were obtained from differential equations for the reflection amplitude. We will give their anisotropic generalization here. There are now two phase integrals:

$$\phi_s = \int^z \mathrm{d}\zeta\, q_s(\zeta), \quad \phi_p = \int^z \mathrm{d}\zeta\, q_p(\zeta). \tag{7.57}$$

The differential equation for the s wave reflection amplitude is (5.76) with ϕ replaced by ϕ_s and q by q_s. The equation for the p wave reflection amplitude is (5.81) with ϕ_p replacing ϕ and $Q = q_p/\varepsilon_o$. These equations lead to the Rayleigh or weak reflection approximations

$$r_s \approx r_s^R = -\int_{-\infty}^{\infty} \mathrm{d}z \frac{q_s'}{2q_s} \mathrm{e}^{2i\phi_s}, \tag{7.58}$$

$$r_p \approx r_p^R = \int_{-\infty}^{\infty} \mathrm{d}z \frac{Q'}{2Q} \mathrm{e}^{2i\phi_p}, \tag{7.59}$$

in parallel with (5.85) and (5.86).

This concludes our preliminary discussion of the optical aspects of reflection by stratified anisotropic media. Only the very simplest form of anisotropy has so far been treated: for more general cases (but restricted to systems with sharp boundaries) the reader is referred to Landau and Lifshitz (1960, Chap. 11), Born and Wolf (1970, Chap. 14), and Azzam and Bashara (1977). A full treatment of reflection and transmission by *uniaxial* media is given in the next chapter. We shall end this chapter by considering anisotropy in ionospheric radio propagation.

7.6 Reflection from the Ionosphere

In the days before satellite communication systems, radio propagation round the earth, using the ionosphere as a reflecting layer, was the only form of long distance "wireless" communication. The simplest model of the ionosphere, that of a plasma

of free electrons in a neutralizing background of ions, leads to the dielectric function (Budden 1985)

$$\varepsilon(z,\omega) = 1 - \frac{\omega_p^2(z)}{\omega^2}. \tag{7.60}$$

The height dependence of ω_p^2, the square of the plasma angular frequency, arises through its proportionality to the electron density. We noted in Sect. 6.7 that for waves radiated at θ_1 to the vertical, this model gives a turning point at height z_0 given by $\omega_p(z_0) = \omega\cos\theta_1$. For fixed θ_1 and ionospheric electron density profile, with maximum $\omega_p(z)$ equal to $\omega_p^{\max}$, frequencies below $\omega_p^{\max}/\cos\theta_1$ will be strongly reflected, while those above this value will be weakly reflected. When (for example) the electron density can be approximated by $\mathrm{sech}^2(z-h)/a$, the resulting reflectivity is that given in Sect. 6.7, with reflectivity contours shown in Fig. 6.11.

The above assumes absence of electron collisions, and neglect of the earth's magnetic field. The effect of dissipation resulting from electron collisions will be discussed in Chap. 10, while the anisotropy resulting from propagation in the earth's magnetic field will be briefly treated here. As in the case of anisotropic dielectrics, there is double refraction. The magneto-ionic case is more complex, since for neither of the two polarizations which propagate unchanged do the wave normal and ray diffractions coincide (Ratcliffe 1959; Budden 1964).

The simplest example of anisotropy arises for wave propagation along the direction of the earth's magnetic field, $\mathbf{B}_0$. This is referred to as the longitudinal case. Jackson (1962, Sect. 7.9) gives a simple argument which shows that for transverse electromagnetic waves propagating along $\mathbf{B}_0$ the two waves which propagate unchanged are left or right circularly polarized, with effective dielectric constants

$$\varepsilon_\pm = 1 - \frac{\omega_p^2}{\omega(\omega \pm \omega_B)}. \tag{7.61}$$

Here ω_B is the frequency of electron gyration round the magnetic field (the gyro-frequency), and is proportional to B_0. The gyro and plasma frequencies can have comparable magnitudes; Fig. 7.3 shows ε_+ and ε_- for the case where $\omega_p = \omega_B$.

We see from the Figure that there is a dramatic difference between the two circular polarizations. The $+$ wave (with positive helicity) is strongly reflected for

$$\omega < \frac{1}{2}\left\{\left(\omega_B^2 + 4\omega_p^2\sec^2\theta_1\right)^{1/2} - \omega_B\right\}, \tag{7.62}$$

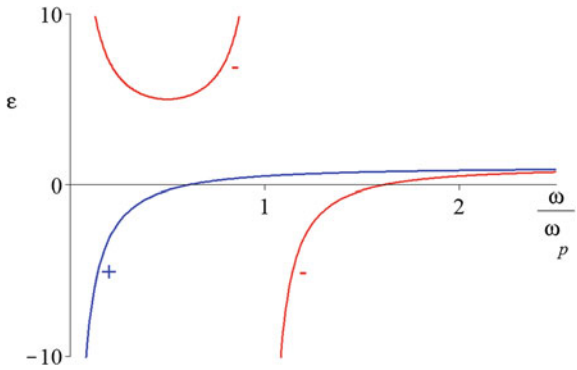


Fig. 7.3 Dielectric functions ε_+ and ε_- for the two circular polarizations, shown as a function of frequency when $\omega_B = \omega_p$

while the − wave is strongly reflected in the interval

$$\omega_B < \omega < \frac{1}{2}\left\{\left(\omega_B^2 + 4\omega_p^2 \sec^2\theta_1\right)^{1/2} + \omega_B\right\}. \tag{7.63}$$

(In these formulae ω_p stands for the maximum value of the plasma frequency, and ω_B for the value attained within the region of maximum electron density.) The physical reason for the difference in the propagation of the two polarizations is that one reinforces and the other opposes the precessional motion of the electrons in the earth's magnetic field. Heading (1975, Sects. 2.5 and 9.8) gives analytic results for reflection by an exponential ionosphere, in the longitudinal case.

There are also interesting effects in the transverse case, where propagation is perpendicular to the earth's magnetic field. Barber and Crombie (1959) have shown that the reflection from the ionosphere is greater for very low frequency waves travelling west to east around the magnetic equator, than for those travelling from east to west. Exact solutions for this case have been given by Westcott (1970).

The general case, with arbitrary angle between the direction of propagation and the earth's magnetic field, is discussed by Heading and Whipple (1952), Heading (1955, 1963), Ratcliffe (1959), Ginzburg (1964), Booker (1984), Budden (1985).

References

Abelès F (1976) Optical properties of very thin films. Thin Solid Films 34:291–302
Azzam RMA, Bashara NM (1977) Ellipsometry and polarized light. North Holland, Amsterdam
Barber NF, Crombie DD (1959) V.L.F. reflections from the ionosphere in the presence of a transverse magnetic field. J Atmos Terr Phys 16:37–45
Beaglehole D (1980) Ellipsometric study of the surface of simple liquids. Physics 100B:163–174
Booker HG (1984) Cold plasma waves. Martinus Nijhoff, Dordrecht
Born M, Wolf E (1970) Principles of optics, 4th edn. Pergamon, New York
Budden KG (1964) Lectures on the magnetoionic theory. Gordon and Breach
Budden KG (1985) The propagation of radio waves. Cambridge

Buff FP (1966) Saline water conversion report. U.S. Govt. Printing Office, 26–27
Ginzburg VL (1964) The propagation of electromagnetic waves in plasmas. Pergamon, New York
Heading J, Whipple RTP (1952) The oblique reflexion of long wireless waves from the ionosphere at places where the earth's magnetic field is regarded as vertical. Philos Trans R Soc 244A:469–503
Heading J (1955) The reflexion of vertically-incident long radio waves from the ionosphere when the earth's magnetic field is oblique. Proc Roy Soc A231:414–435
Heading J (1963) Composition of reflection and transmission formulae. J Res Nat Bur Stand 67D:65–77
Heading J (1975) Ordinary differential equations, theory and practice. Elek Science
Huygens C (1690/1912) Treatise on light. In which are explained the causes of that which occurs in reflexion, and in refraction. And particularly in the strange refraction of Iceland crystal. (trans: Thompson SP)
Jackson JD (1962) Classical electrodynamics. Wiley, New York
Landau LD, Lifshitz EM (1960) Electrodynamics of continuous media. Pergamon, Oxford
Lekner J (1983) Anisotropy of the dielectric function within a liquid-vapour interface. Mol Phys 49:1385–1400
Ratcliffe JA (1959) The magneto-ionic theory and its applications to the ionosphere. Cambridge University Press, Cambridge
Westcott BS (1970) Exact solutions for vertically polarized electromagnetic waves in a horizontally stratified magneto-plasma. Proc Camb Phil Soc 67:491–501

Further Readings

Notes: Sections 7.1 and 7.2 are based on Lekner (1983) and

Lekner J (1986) Reflection of light by a non-uniform film between like media. J Opt Soc Am A3:9–15. (This paper also gives a variational theory for the anisotropic case.)

Chapter 8
Uniaxial Anisotropy

In Chap. 7 we have treated only the simplest case: reflection by a uniaxial system which has its optic axis coinciding with the normal to the reflecting surface. It is time to consider the general case, with arbitrary orientation of the optic axis or axes (Lekner 1991). As usual we shall take the reflecting surface to be the xy plane $(z = 0)$, and the plane of incidence to be the zx plane, with the z-axis normal to the crystal face, and directed into it. After the first section the discussion is restricted to uniaxial media.

8.1 Propagation Within Homogeneous Anisotropic Media

We first formulate the general problem (uniaxial or biaxial). When a plane wave is incident from an isotropic medium onto the anisotropic medium, there will be a reflected wave, and (in general) two transmitted plane waves. All components of the electric and magnetic field vectors **E** and **B** have the x and t dependence in the factor $\exp i(Kx - \omega t)$. Because of the translational symmetry in the y direction, there is no y dependence. Within the anisotropic medium (assumed non-magnetic) the two curl Maxwell curl equations read, after time differentiation,

$$\nabla \times \mathbf{E} = i\frac{\omega}{c}\mathbf{B}, \qquad \nabla \times \mathbf{B} = -i\frac{\omega}{c}\mathbf{D}. \tag{8.1}$$

The electric displacement **D** is found from **E** via to the dielectric tensor:

$$\begin{pmatrix} D_x \\ D_y \\ D_z \end{pmatrix} = \begin{pmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix}. \tag{8.2}$$

The dielectric tensor is symmetric: Born and Wolf (1970, Sect. 14.1) show that the symmetry of the dielectric tensor is related to the form taken by conservation of energy in the electromagnetic field, and that this symmetry implies the existence of a principal coordinate system in which the dielectric tensor is diagonal.

J. Lekner, *Theory of Reflection*, Springer Series on Atomic, Optical, and Plasma Physics 87, DOI 10.1007/978-3-319-23627-8_8

The six equations (8.1) give, after differentiation with respect to x is performed,

$$\begin{array}{cc} -\partial E_y/\partial z = i\frac{\omega}{c}B_x & -\partial B_y/\partial z = -i\frac{\omega}{c}D_x \\ \partial E_x/\partial z - iKE_z = i\frac{\omega}{c}B_y & \partial B_x/\partial z - iKB_z = -i\frac{\omega}{c}D_y \\ iKE_y = i\frac{\omega}{c}B_z & iKB_y = -i\frac{\omega}{c}D_z \end{array} \tag{8.3}$$

We can eliminate **B**, to obtain three coupled equations for the components of **E**:

$$\partial^2 E_x/\partial z^2 - iK\partial E_z/\partial z + \left(\frac{\omega}{c}\right)^2 D_x = 0 \tag{8.4}$$

$$\partial^2 E_y/\partial z^2 - K^2 E_y + \left(\frac{\omega}{c}\right)^2 D_y = 0 \tag{8.5}$$

$$-iK\partial E_x/\partial z - K^2 E_z + \left(\frac{\omega}{c}\right)^2 D_z = 0. \tag{8.6}$$

In the isotropic case (ε_{ij} diagonal) E_y is decoupled from E_x and E_z.

The differential equations (8.4–8.6) imply boundary conditions to be satisfied at a discontinuity. The derivative of a discontinuous function would give a delta function, which cannot be cancelled by any other term in the equation, nor can the derivative of a delta function be cancelled. Thus from (8.4) it follows that $\partial E_x/\partial z - iKE_z$ and E_x are continuous across a boundary (the continuity of E_x is also implied by (8.6)), and from (8.5) that $\partial E_y/\partial z$ and E_y are continuous. Comparison with (8.3) shows these conditions to be the continuity of the tangential components of **E** and **B**, as expected. As regards **D**, we note that iK times (8.4) plus the z-derivative of (8.6) implies

$$\partial D_z/\partial z + iKD_x = 0, \tag{8.7}$$

from which we deduce that the normal component of **D** is continuous, also a familiar result.

The above equations are for an arbitrary z-stratified anisotropic material. We now specialize to homogeneous anisotropic media. We wish to find the *normal modes*, namely those fields which propagate as plane waves in the medium. Such modes have all field components with the z-dependence $\exp iqz$, q being the component of the wavevector normal to the surface. Substitution of this functional form into (8.4–8.6) gives

$$-q^2 E_x + qKE_z + \left(\frac{\omega}{c}\right)^2 D_x = 0 \tag{8.8}$$

$$-(q^2 + K^2)E_y + \left(\frac{\omega}{c}\right)^2 D_y = 0 \tag{8.9}$$

$$qKE_x - K^2 E_z + \left(\frac{\omega}{c}\right)^2 D_z = 0. \tag{8.10}$$

The components of **D** are given as linear combinations of the components of **E**, by (8.2). Thus we have three linear homogeneous equations in E_x, E_y and E_z; a solution is possible only if the determinant of the coefficients, as given in (8.11), is zero:

$$\begin{vmatrix} \varepsilon_{xx}\left(\frac{\omega}{c}\right)^2 - q^2 & \varepsilon_{xy}\left(\frac{\omega}{c}\right)^2 & \varepsilon_{xz}\left(\frac{\omega}{c}\right)^2 + qK \\ \varepsilon_{yx}\left(\frac{\omega}{c}\right)^2 & \varepsilon_{yy}\left(\frac{\omega}{c}\right)^2 - K^2 - q^2 & \varepsilon_{yz}\left(\frac{\omega}{c}\right)^2 \\ \varepsilon_{zx}\left(\frac{\omega}{c}\right)^2 + qK & \varepsilon_{zy}\left(\frac{\omega}{c}\right)^2 & \varepsilon_{zz}\left(\frac{\omega}{c}\right)^2 - K^2 \end{vmatrix} = 0. \tag{8.11}$$

Equation (8.11) is a quartic in q. Its four solutions, the normal mode eigenvalues, correspond to two inward-propagating modes and two backward-propagating modes. At normal incidence it simplifies into a quadratic for q^2. The biaxial case at normal incidence is discussed in the Appendix of Lekner (1994a). The general uniaxial case is the subject of the following sections.

8.2 Dielectric Tensor and Normal Modes in Uniaxial Crystals

In the principal coordinate system the dielectric tensor is diagonal, with diagonal elements ε_a, ε_b and ε_c (giving the dielectric response along the principal axes **a**, **b** and **c**). For uniaxial crystals two of these values are equal. Conventionally $\varepsilon_a = \varepsilon_b$ and their common value is ε_o, while ε_c is called ε_e; the subscripts o and e stand for *ordinary* and *extraordinary*. The direction **c** is the *optic axis*. Let α, β and γ be direction cosines of the optic axis **c** relative to the cartesian (xyz) laboratory frame:

$$\mathbf{c} = \alpha\hat{\mathbf{x}} + \beta\hat{\mathbf{y}} + \gamma\hat{\mathbf{z}}\,, \tag{8.12}$$

where $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ are unit vectors along the x, y and z positive axes. Since **c** is also a unit vector,

$$\alpha^2 + \beta^2 + \gamma^2 = 1. \tag{8.13}$$

With $\Delta\varepsilon = \varepsilon_e - \varepsilon_o$, the dielectric tensor in the laboratory frame, shown in (8.2), reduces to (Lekner 1991, Sect. 3)

$$\begin{pmatrix} \varepsilon_o + \alpha^2\Delta\varepsilon & \alpha\beta\Delta\varepsilon & \alpha\gamma\Delta\varepsilon \\ \alpha\beta\Delta\varepsilon & \varepsilon_o + \beta^2\Delta\varepsilon & \beta\gamma\Delta\varepsilon \\ \alpha\gamma\Delta\varepsilon & \beta\gamma\Delta\varepsilon & \varepsilon_o + \gamma^2\Delta\varepsilon \end{pmatrix}. \tag{8.14}$$

The determinant (8.11) now factors into two quadratics, one of which is $q^2 - q_o^2$, where

$$q_o^2 = \varepsilon_o \left(\frac{\omega}{c}\right)^2 - K^2. \tag{8.15}$$

The normal components $\pm q_o$ of the ordinary wavevector $\mathbf{k}_o$ have the same simple form as in the isotropic case, and do not depend on how the uniaxial crystal is oriented: the expression (8.15) is independent of the direction cosines α, β and γ of the optic axis. The normal components of the extraordinary wavevector $\mathbf{k}_e$ do depend on the direction cosines. They are the solutions of the remaining quadratic,

$$\varepsilon_\gamma q^2 + 2\alpha\gamma\Delta\varepsilon K q + \varepsilon_\alpha K^2 - \varepsilon_o \varepsilon_e \left(\frac{\omega}{c}\right)^2 = 0. \tag{8.16}$$

Here and in the following we use the shorthand (Lekner 1992b, 1993)

$$\begin{aligned} \varepsilon_\gamma &= \varepsilon_o + \gamma^2\Delta\varepsilon, \qquad \varepsilon_\alpha = \varepsilon_o + \alpha^2\Delta\varepsilon, \\ \varepsilon_{\alpha\gamma} &= \varepsilon_o + (\alpha^2 + \gamma^2)\Delta\varepsilon = \varepsilon_o + \left(1 - \beta^2\right)\Delta\varepsilon = \varepsilon_e - \beta^2\Delta\varepsilon. \end{aligned} \tag{8.17}$$

The solutions of (8.16) are

$$q_\pm = \pm\bar{q} - \alpha\gamma K\Delta\varepsilon/\varepsilon_\gamma, \tag{8.18}$$

where

$$\bar{q}^2 = \frac{\varepsilon_o}{\varepsilon_\gamma^2}\left[\varepsilon_e\varepsilon_\gamma\left(\frac{\omega}{c}\right)^2 - \varepsilon_{\alpha\gamma}K^2\right]. \tag{8.19}$$

We note here an important difference between the ordinary and extraordinary waves: if $\varepsilon_1 > \varepsilon_o$ then for $\theta_1 \geq \theta_c^o$, where the ordinary critical angle θ_c^o is given by

$$\sin^2\theta_c^o = \frac{\varepsilon_o}{\varepsilon_1}, \tag{8.20}$$

the wavevector component q_o becomes purely imaginary, and there will be no propagating ordinary wave. But for $\theta_1 \geq \theta_c^e$ where

$$\sin^2\theta_c^e = \frac{\varepsilon_e\varepsilon_\gamma}{\varepsilon_1\varepsilon_{\alpha\gamma}} \tag{8.21}$$

only $\bar{q}$ becomes imaginary. The other part of q_e, namely $-\alpha\gamma K\Delta\varepsilon/\varepsilon_\gamma$, remains real in the absence of absorption. Thus, beyond angle of incidence θ_c^e, the extraordinary wave will have a real and an imaginary part of its normal wavevector component, as if inside an absorber, and an exponentially damped extraordinary wave can propagate into the medium. This holds unless the product $\alpha\gamma\Delta\varepsilon$ is zero, namely for

vanishing anisotropy ($\Delta\varepsilon = 0$) or when one of α or γ is zero. The latter occurs when the optic axis lies in the reflecting plane.

For a given q_o or q_e, the electric field vectors $\mathbf{E}^o$ or $\mathbf{E}^e$ satisfy the homogeneous equations obtained by putting $q = q_o$ or $q = q_e$ in the matrix of coefficients shown in (8.11). If this matrix is M, we have $M\mathbf{E} = 0$, from which the field components can be obtained, up to a common factor. For example, we can set $E_x/E_z = X$, $E_y/E_z = Y$. Then $M\mathbf{E} = 0$ becomes a (redundant) trio of inhomogeneous equations for X and Y. There are three sets of solutions, given in equation (26) of Lekner (1991), all equivalent since the determinant of the coefficients is zero. Let N_o and N_e be normalization factors, determined by giving the square of the electric field vector some assigned value. Then the ordinary and extraordinary electric field vectors are found to be

$$\mathbf{E}^o = N_o(-\beta q_o,\ \alpha q_o - \gamma K,\ \beta K), \tag{8.22}$$

$$\mathbf{E}^e = N_e\left(\alpha q_o^2 - \gamma q_e K,\ \beta\varepsilon_o\left(\frac{\omega}{c}\right)^2,\ \gamma\left[\varepsilon_o\left(\frac{\omega}{c}\right)^2 - q_e^2\right] - \alpha q_e K\right). \tag{8.23}$$

Note that $\mathbf{E}^o$ is always perpendicular to the optic axis $\mathbf{c} = (\alpha, \beta, \gamma)$. The scalar product of these normal mode field vectors is

$$\mathbf{E}^o \cdot \mathbf{E}^e = N_o N_e \beta K(\alpha K + \gamma q_e)(q_o - q_e). \tag{8.24}$$

Thus the ordinary and extraordinary electric fields are orthogonal when the optic axis lies in the plane of incidence ($\beta = 0$), at normal incidence ($K = 0$), in the isotropic limit ($q_o = q_e$), and also when $\alpha K + \gamma q_e = 0$. When the last condition is satisfied, the extraordinary wavevector $(K, 0, q_e)$ and ray direction (given below) are both that of $(\gamma, 0, -\alpha)$, and thus perpendicular to the optic axis (α, β, γ). Also $\mathbf{E}^e$ is then parallel to the optic axis.

The wavevector gives the direction of the normal to the surfaces of constant phase, and is $(K, 0, q)$ with $q = q_o$ or $q = q_e$. The *ray vector* gives the energy flow direction, which is that of $\mathbf{E} \times \mathbf{B}$. From (8.3) we have

$$\frac{\omega}{c}B_x = -qE_y, \qquad \frac{\omega}{c}B_y = qE_x - KE_z, \qquad \frac{\omega}{c}B_z = KE_y, \tag{8.25}$$

so that $\mathbf{E} \times \mathbf{B}$ is proportional to the vector

$$\left(K[E_y^2 + E_z^2] - qE_xE_z, \quad -KE_xE_y - qE_yE_z, \quad q[E_x^2 + E_y^2] - KE_xE_z\right). \tag{8.26}$$

For the ordinary mode the ray direction is that of $(K, 0, q_o)$, the same as the wavevector direction. For the extraordinary wave the ray is parallel to the vector with components

$$\begin{aligned}&(\alpha q_e - \gamma K)\left\{\alpha q_e K - \gamma\left(\varepsilon_o\left(\tfrac{\omega}{c}\right)^2 - q_e^2\right)\right\} + \beta^2 K\varepsilon_o,\\ &\beta(\alpha K + \gamma q_e)(q_e^2 - q_o^2),\\ &(\alpha q_e - \gamma K)\left(\alpha q_o^2 - \gamma q_e K\right) + \beta^2 q_e \varepsilon_o\left(\tfrac{\omega}{c}\right)^2.\end{aligned} \tag{8.27}$$

The plane containing the optic axis and the extraordinary wavevector has its normal along the cross product of (α, β, γ) and $(K, 0, q_e)$, namely $(\beta q_e, \gamma K - \alpha q_e, -\beta K)$. This normal is perpendicular to the ray vector (8.27), so the extraordinary ray and wavevector are coplanar with the optic axis.

8.3 Uniaxial Crystal Reflection and Transmission Amplitudes

As usual we decompose the incoming field into its s and p components, with $\mathbf{E}_s$ perpendicular to the plane of incidence (the zx plane) and $\mathbf{E}_p$ lying in it. For the s polarization the z-dependence of the electric field components is

$$\begin{array}{rc}\text{incoming:} & (0,\ 1,\ 0)\mathrm{e}^{iq_1 z}\\ \text{reflected:} & (r_{sp}\cos\theta_1,\ r_{ss},\ r_{sp}\sin\theta_1)\mathrm{e}^{-iq_1 z}\\ \text{transmitted:} & t_{so}(E_x^o, E_y^o, E_z^o)\mathrm{e}^{iq_o z} + t_{se}(E_x^e, E_y^e, E_z^e)\mathrm{e}^{iq_e z}\end{array} \tag{8.28}$$

The reflection and transmission amplitudes for the s polarization are r_{ss}, r_{sp}, t_{so}, t_{se}; θ_1 is the angle of incidence, and q_1 is the z-component of the incoming wavevector, $q_1 = n_1\ \omega/c\ \cos\theta_1$. In Sect. 8.1 we deduced the boundary conditions to be applied at the crystal face $z = 0$, namely the continuity of E_x, E_y, $\partial E_x/\partial z - iKE_z$ and of $\partial E_y/\partial z$. These four conditions, applied to the waveforms given in (8.28), lead to four equations linear in the unknowns r_{ss}, r_{sp}, t_{so}, t_{se}. The solutions can be put in the form (Lekner 1992b, 1993)

$$r_{ss} = \frac{A(q_1 - q_o) + B(q_1 - q_e)}{A(q_1 + q_o) + B(q_1 + q_e)} \tag{8.29}$$

$$r_{sp} = \frac{2\beta(\alpha q_o + \gamma K)(q_o - q_e)k_1 k_o^2}{A(q_1 + q_o) + B(q_1 + q_e)} \tag{8.30}$$

$$N_o t_{so} = \frac{2q_1\left[\alpha(k_o^2 q_e + q_t q_o^2) - \gamma K(k_o^2 + q_t q_e)\right]}{A(q_1 + q_o) + B(q_1 + q_e)} \tag{8.31}$$

$$N_e t_{se} = \frac{2\beta q_1(k_o^2 + q_t q_o)}{A(q_1 + q_o) + B(q_1 + q_e)} \tag{8.32}$$

In order to obtain compact forms for these amplitudes, the following shorthand is used (with $\varepsilon_1 = n_1^2$, $\varepsilon_o = n_o^2$):

$$k_1 = n_1 \frac{\omega}{c}, \qquad k_o = n_o \frac{\omega}{c}, \qquad q_t = q_1 + K \tan\theta_1 = k_1^2/q_1. \tag{8.33}$$

The coefficients A and B are defined by

$$A = (\alpha q_o - \gamma K)\big\{\alpha\big(k_o^2 q_e + q_t q_o^2\big) - \gamma K\big(k_o^2 + q_t q_e\big)\big\}, \tag{8.34}$$

$$B = \beta^2 k_o^2\big(k_o^2 + q_t q_o\big). \tag{8.35}$$

We note that the reflection amplitudes are independent of the normalization factors N_o and N_e of the ordinary and extraordinary electric fields, whereas it is the product of the transmission amplitudes and the respective field magnitudes N_o and N_e that features in (8.31) and (8.32). When the incident field has unit modulus (which is the choice made here), the amplitudes of the fields transmitted into the crystal are $N_o t_{so}$ and $N_e t_{se}$.

Turning now to the p polarization, the waveforms are

$$\text{incoming: } (\cos\theta_1,\ 0,\ -\sin\theta_1)\mathrm{e}^{iq_1 z} \tag{8.36}$$

$$\text{reflected: } (r_{pp}\cos\theta_1,\ r_{ps},\ r_{pp}\sin\theta_1)\mathrm{e}^{-iq_1 z} \tag{8.37}$$

$$\text{transmitted: } t_{po}(E_x^o, E_y^o, E_z^o)\mathrm{e}^{iq_o z} + t_{pe}(E_x^e, E_y^e, E_z^e)\mathrm{e}^{iq_e z} \tag{8.38}$$

The continuity conditions applied at $z = 0$ give four linear equations in the four unknowns r_{pp}, r_{ps}, t_{po}, t_{pe}. The solution is

$$r_{pp} = -\frac{A'(q_1 + q_o) + B'(q_1 + q_e)}{A(q_1 + q_o) + B(q_1 + q_e)} \tag{8.39}$$

$$r_{ps} = \frac{2\beta(\alpha q_o - \gamma K)(q_o - q_e)k_1 k_o^2}{A(q_1 + q_o) + B(q_1 + q_e)} \tag{8.40}$$

$$N_o t_{po} = -\frac{2\beta(q_1 + q_e)k_1 k_o^2}{A(q_1 + q_o) + B(q_1 + q_e)} \tag{8.41}$$

$$N_e t_{pe} = \frac{2(\alpha q_o - \gamma K)(q_1 + q_o)k_1}{A(q_1 + q_o) + B(q_1 + q_e)} \tag{8.42}$$

The coefficients A' and B' are defined by

$$A' = (\alpha q_o - \gamma K)\left[\alpha(k_o^2 q_e - q_t q_o^2) - \gamma K(k_o^2 - q_t q_e)\right], \tag{8.43}$$

$$B' = \beta^2 k_o^2 (k_o^2 - q_t q_o). \tag{8.44}$$

The formulae given above give explicit closed-form expressions for the reflection and transmission amplitudes for an arbitrary face of a uniaxial crystal, at any angle of incidence.

Special geometries are discussed in Lekner (1991, 1992b); here we give only the simplest, *reflection from a basal plane* (one perpendicular to the optic axis). An elementary treatment of this special case was given in Sect. 7.1. The optic axis coincides with the normal to the reflecting surface, so $\gamma^2 = 1$ and α and β are zero. The normal component of the extraordinary wavevector is given by

$$q_e^2 = \varepsilon_o \left(\frac{\omega}{c}\right)^2 - \frac{\varepsilon_o}{\varepsilon_e} K^2 = \frac{\varepsilon_o}{\varepsilon_e}\left[\varepsilon_e \left(\frac{\omega}{c}\right)^2 - K^2\right]. \tag{8.45}$$

If we take $\gamma = -1$, corresponding to the optic coinciding with the outward normal, the normalized field eigenstates are

$$\mathbf{E}^o = (0, 1, 0), \qquad \mathbf{E}^e = \left(q_e,\ 0,\ -\frac{\varepsilon_o}{\varepsilon_e} K\right)\left[q_e^2 + \left(\frac{\varepsilon_o}{\varepsilon_e} K\right)^2\right]^{-\frac{1}{2}}. \tag{8.46}$$

The cross-reflection amplitudes r_{ps} and r_{sp} are zero, and

$$r_{ss} = \frac{q_1 - q_o}{q_1 + q_o}, \qquad r_{pp} = \frac{Q - Q_1}{Q + Q_1}. \tag{8.47}$$

In (8.47) $Q = q_e/\varepsilon_o$ evaluated at $\gamma^2 = 1$ and $Q_1 = q_1/\varepsilon_1$. In the absence of absorption r_{pp} is zero at the Brewster angle θ_{pp} given by

$$\tan^2 \theta_{pp} = \frac{\varepsilon_e(\varepsilon_o - \varepsilon_1)}{\varepsilon_1(\varepsilon_e - \varepsilon_1)} \qquad (\gamma^2 = 1). \tag{8.48}$$

A real θ_{pp} will not exist when ε_1 lies between ε_o and ε_e (for $\gamma^2 = 1$). As noted in Sect. 7.1, there is no coupling between the s and p (equivalent in this case to the o and e), so $t_{se} = 0 = t_{po}$. The direct transmission amplitudes are

$$t_{so} = \frac{2q_1}{q_1 + q_o}, \qquad t_{pe} = \frac{2n_1 Q_1}{n_o(Q_1 + Q)}\left[1 - K^2(\varepsilon_e - \varepsilon_o)/\varepsilon_e^2\right]. \tag{8.49}$$

At normal incidence the common values of the nonzero reflection and transmission amplitudes are

$$\frac{n_1 - n_o}{n_1 + n_o} \quad \text{and} \quad \frac{2n_1}{n_1 + n_o}. \tag{8.50}$$

8.4 Bounds and Zeros of the Reflection Amplitudes, the Polarizing Angle

We return to the general case, and summarize first some of the reflection properties derived in Lekner (1992b). The square of the extraordinary wavevector is bounded by

$$q_o^2 = \varepsilon_o \left(\frac{\omega}{c}\right)^2 - K^2 \quad \text{and} \quad q_m^2 = \varepsilon_e \left(\frac{\omega}{c}\right)^2 - K^2. \tag{8.51}$$

The reflection amplitudes r_{ss} and r_{pp} are bounded by

$$r_{ss}(\beta = 0) = \frac{q_1 - q_o}{q_1 + q_o} \quad \text{(optic axis in plane of incidence)} \tag{8.52}$$

$$r_{ss}(\beta^2 = 1) = \frac{q_1 - q_m}{q_1 + q_m} \quad \text{(optic axis perpendicular to plane of incidence)} \tag{8.53}$$

$$r_{pp}(\gamma^2 = 1) = \frac{Q_m - Q_1}{Q_m + Q_1} \quad \text{(optic axis normal to reflecting plane)} \tag{8.54}$$

$$r_{pp}(\alpha^2 = 1) = \frac{Q - Q_1}{Q + Q_1} \quad \text{(optic axis along intersection of plane of incidence and the reflecting plane)} \tag{8.55}$$

In (8.54) $Q_m = q_e(\gamma^2 = 1)/\varepsilon_o = q_m/n_o n_e$; in (8.55) $Q = q_o/n_o n_e$. The bounds apply only when all the wavenumbers are real $[\theta_1 < \theta_c^o, \theta_c^e]$. Figure 8.1 shows these bounds for reflection from calcite in air. The bounds are much wider for calcite in oil (Fig. 3 of Lekner 1992b), for which the wavefront and ray directions were shown in Fig. 7.1.

The other reflection and transmission amplitudes are discussed in Lekner (1992b). Bounds on the Brewster angle at which $r_{pp} = 0$ are found from (8.54) and (8.55). These are $\theta_{pp}(\gamma^2 = 1)$ given in (8.45) and

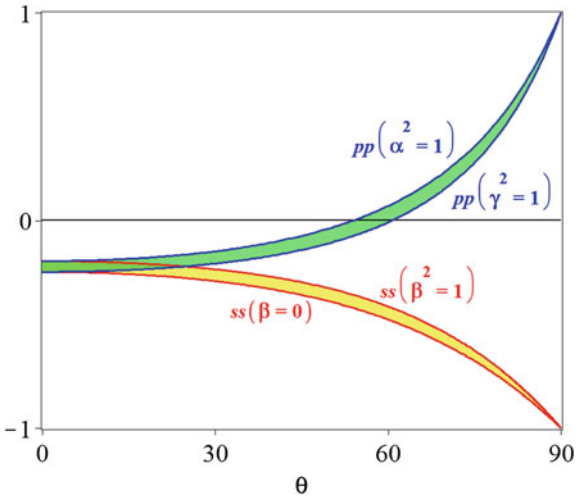


Fig. 8.1 Bounds on the reflection amplitudes, for sodium yellow light incident from vacuum onto calcite ($n_1 = 1, n_o = 1.658, n_e = 1.486$). The possible values of the reflection amplitudes lie within the *shaded bands*

$$\tan^2\theta_{pp}(\alpha^2 = 1) = \frac{\varepsilon_o(\varepsilon_e - \varepsilon_1)}{\varepsilon_1(\varepsilon_o - \varepsilon_1)}. \tag{8.56}$$

These bounds are special cases of the more general formula

$$\tan^2\theta_{pp}(\beta = 0) = \frac{\varepsilon_o\varepsilon_e - \varepsilon_1\varepsilon_\gamma}{\varepsilon_1(\varepsilon_\gamma - \varepsilon_1)}, \tag{8.57}$$

or its equivalent

$$\sin^2\theta_{pp}(\beta = 0) = \frac{\varepsilon_o\varepsilon_e - \varepsilon_1\varepsilon_\gamma}{\varepsilon_o\varepsilon_e - \varepsilon_1^2}. \tag{8.58}$$

The expressions (8.57, 8.58) in turn follow from the simple expression

$$r_{pp}(\beta = 0) = \frac{Q_\gamma - Q_1}{Q_\gamma + Q_1} \tag{8.59}$$

where $Q_1 = q_1/\varepsilon_1$ as always, and

$$Q_\gamma = \frac{q_\gamma}{n_o n_e}, \qquad q_\gamma^2 = \varepsilon_\gamma\left(\frac{\omega}{c}\right)^2 - K^2. \tag{8.60}$$

The possible zeros of the reflection amplitudes are discussed in more detail in Lekner (1993). It is shown there that when the index of the medium of incidence

lies between the ordinary and extraordinary indices of the crystal, it is possible for r_{ss} to be zero at an angle θ_{ss}, and that there exist four equivalent orientations of the crystal optic axis for which r_{pp}, r_{ss} and either r_{ps} or r_{sp} are simultaneously zero, at angle of incidence equal to $\arctan(n_o/n_1)$.

There is another angle of interest: the polarizing angle θ_{pol}, defined as the angle of incidence at which an incident wave of arbitrary polarization becomes linearly polarized on reflection. In terms of the reflection amplitudes, θ_{pol} is given by (Sect. 18.4)

$$r_{pp}r_{ss} - r_{ps}r_{sp} = 0 \tag{8.61}$$

This leads to a quartic equation (Lekner 1999), closely related to the quartics associated with the angles θ_{pp} and θ_{ss} discussed in the previous paragraph.

8.5 External Reflection from an Immersed Crystal

Suppose the crystal is immersed in a liquid of refractive index n_1 greater than n_o. Then q_o becomes imaginary for angle of incidence greater than the ordinary critical angle $\theta_c^o = \arcsin(n_o/n_1)$, as we saw in Sect. 8.2. There will then be strong external reflection of both polarizations for $\theta_1 > \theta_c^o$. Likewise when n_1 is greater than $n_e n_\gamma / n_{\alpha\gamma}$ (where $n_\gamma^2 = \varepsilon_\gamma = \varepsilon_o + \gamma^2\Delta\varepsilon$, $n_{\alpha\gamma}^2 = \varepsilon_{\alpha\gamma} = \varepsilon_o + (\alpha^2 + \gamma^2)\Delta\varepsilon$) then $\bar{q}$ will be imaginary for $\theta_1 > \theta_c^e = \arcsin(n_e n_\gamma / n_1 n_{\alpha\gamma})$, and again there will be strong external reflection. The general case is complicated, but when the optic axis lies in the plane of incidence ($\beta = 0$), the reflection amplitudes simplify greatly. When $\beta = 0$ the coefficients B and B' are both zero, so we have from Sect. 8.3

$$r_{ss} = \frac{q_1 - q_o}{q_1 + q_o}, \qquad r_{pp} = -\left(\frac{A'}{A}\right)_{\beta=0} \tag{8.62}$$

Also both r_{sp} and r_{ps} are zero when $\beta = 0$. Thus when $n_1 > n_o$ and $\theta_1 > \arcsin(n_o/n_1)$, q_o is imaginary, which implies $|r_{ss}|^2 = 1$. Also $\bar{q}$ is imaginary for $\theta_1 > \arcsin(n_\gamma/n_1)$. When this is so, we find after much reduction that $|r_{pp}|^2 = 1$ also.

To sum up: when the optic axis lies in the plane of incidence, the s-wave will be totally reflected for $\theta_1 > \arcsin(n_o/n_1)$, and the p-wave will be totally reflected for $\theta_1 > \arcsin(n_\gamma/n_1)$. If both inequalities hold, the crystal will reflect all light incident upon it. Figure 8.2 shows the s and p reflection amplitudes r_{ss} and r_{pp} for calcite with its optic axis in the plane of incidence and normal to the reflecting surface ($\gamma^2 = 1$), immersed in liquid carbon disulphide. The refractive indices for sodium yellow wavelength (589 nm) are $n_1 = 1.628$, $n_o = 1.658$, $n_e = 1.486$. The extraordinary critical angle is $\theta_c^e = \arcsin(n_e/n_1) \approx 65.89°$ (from (8.21), with

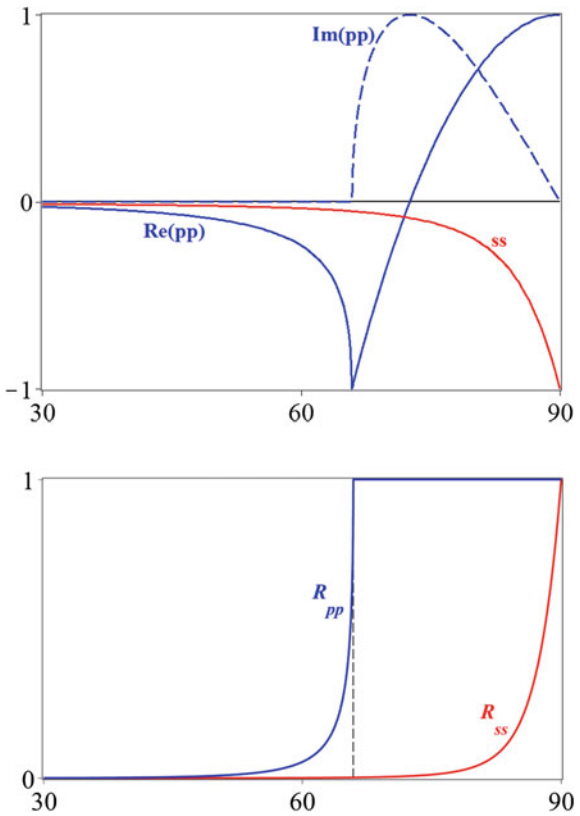


Fig. 8.2 Reflection amplitudes r_{ss} and r_{pp} for calcite immersed in liquid carbon disulphide, in the *upper diagram*. The *lower diagram* shows the reflectivities $R_{ss} = r_{ss}^2$ and $R_{pp} = |r_{pp}|^2$. The latter is unity for $\theta_1 \geq \theta_c^e$, indicated by the *dashed vertical line*, for the particular crystal orientation chosen (given in the text)

$\gamma^2 = 1$, $\alpha = 0 = \beta$). Beyond this angle of incidence, r_{pp} has an imaginary part. The real part of r_{pp} passes through zero at the angle at which the imaginary part is unity. For $\theta_1 > \theta_c^e$ the absolute value of r_{pp} is unity, as shown on the lower figure.

8.6 Normal-Incidence Reflection and Transmission

At normal incidence there is no physical difference between the s and p polarizations: for any incident polarization the electric field vector is parallel to the reflecting surface. Instead of four reflection amplitudes r_{ss}, r_{sp} r_{pp} and r_{ps}, the reflection is characterized by two, which (following Lekner 1992a) we shall call r and r'. Likewise two transmission amplitudes t_o and t_e suffice to characterize the transmission properties. As before the optic axis is defined in terms of its direction cosines α, β and γ with respect to the x, y and z axes. From our previous results of Sect. 8.2, the normal modes which can propagate as plane waves inside the crystal are

$$\mathbf{E}^o = N_o(-\beta, \alpha, 0), \qquad \mathbf{E}^e = N_e\big(\alpha, \beta, -\gamma(1-\gamma^2)\Delta\varepsilon/\varepsilon_\gamma\big). \tag{8.63}$$

Note that $\mathbf{E}^o \cdot \mathbf{E}^e = 0$ (compare the general result (8.24)). The ordinary and extraordinary wavevectors both lie along the inward normal $\mathbf{n} = (0, 0, 1)$, and have magnitudes

$$k_o = n_o \frac{\omega}{c}, \qquad k_e = \frac{n_o n_e}{n_\gamma}\frac{\omega}{c} \qquad (n_\gamma^2 = \varepsilon_\gamma = \varepsilon_o + \gamma^2\Delta\varepsilon). \tag{8.64}$$

The vector $\mathbf{n} \times \mathbf{c} = (0, 0, 1) \times (\alpha, \beta, \gamma) = (-\beta, \alpha, 0)$ lies in the reflecting plane $z = 0$, and is parallel to $\mathbf{E}^o$. We shall take the direction of $\mathbf{E}_1$ to define the x-axis, so $\mathbf{E}_1 = (E_1, 0, 0)$, and let ϕ be the angle between $\mathbf{E}_1$ and the $\mathbf{n} \times \mathbf{c}$ or $\mathbf{E}^o$ direction. Thus, with $\cos\phi = -\beta/(\alpha^2+\beta^2)^{\frac{1}{2}}$, $\sin\phi = \alpha/(\alpha^2+\beta^2)^{\frac{1}{2}}$,

$$\mathbf{E}^o = (\cos\phi, \sin\phi, 0), \quad \mathbf{E}^e = (\sin\phi\cos\delta, -\cos\phi\cos\delta, \sin\delta), \tag{8.65}$$

where δ is the angle between the ray and wavevector directions for the extraordinary wave:

$$\sin\delta = \mathbf{E}^e \cdot \mathbf{n}, \qquad \tan\delta = -\gamma(1-\gamma^2)^{\frac{1}{2}}\Delta\varepsilon/\varepsilon_\gamma. \tag{8.66}$$

(The extraordinary ray direction for normal incidence is that of $(\alpha\gamma\Delta\varepsilon,\ \beta\gamma\Delta\varepsilon,\ \varepsilon_\gamma)$, from (8.27) or Lekner (1992a).)

The incident, reflected and transmitted waves, for light incident normally onto an arbitrary crystal face, are

$$\begin{array}{rc} \text{incident:} & (1,\ 0,\ 0)\mathrm{e}^{ik_1 z} \qquad \big(k_1 = n_1\frac{\omega}{c}\big) \\ \text{reflected:} & (r,\ r',\ 0)\mathrm{e}^{-ik_1 z} \\ \text{transmitted:} & t_o\mathbf{E}^o\mathrm{e}^{ik_o z} + t_e\mathbf{E}^e\mathrm{e}^{ik_e z} \end{array} \tag{8.67}$$

We apply the continuity of E_x, E_y, $\partial E_x/\partial z$ and $\partial E_y/\partial z$ at the interface $z = 0$. This gives four equations, which can be put in matrix form: define the column vectors

$$\mathbf{r} = \begin{pmatrix} r \\ r' \end{pmatrix}, \qquad \mathbf{t} = \begin{pmatrix} t_o \\ t_e \end{pmatrix}, \qquad \mathbf{u} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \tag{8.68}$$

and the matrices

$$M = \begin{pmatrix} E_x^o & E_x^e \\ E_y^o & E_y^e \end{pmatrix}, \qquad K_1^{-1} = \begin{pmatrix} k_o/k_1 & 0 \\ 0 & k_e/k_1 \end{pmatrix}. \tag{8.69}$$

M is the *mode* matrix, formed from the transverse components of the eigenstates (modes) $\mathbf{E}^o$ and $\mathbf{E}^e$. The four equations arising from the boundary conditions can now be written as a pair of coupled vector equations in $\mathbf{r}$ and $\mathbf{t}$:

$$\mathbf{u}+\mathbf{r}=M\mathbf{t}, \qquad \mathbf{u}-\mathbf{r}=MK_1^{-1}\mathbf{t}. \tag{8.70}$$

The solution, expressed in terms of a reflection matrix R, is

$$\mathbf{r}=R\mathbf{u}, \qquad \mathbf{t}=M^{-1}(I+R)\mathbf{u}, \tag{8.71}$$

where I is the identity (unit 2×2 matrix) and R a reflection matrix (more detail is given in Lekner 1992a)

$$R=M(K_1+I)^{-1}(K_1-I)M^{-1}=\begin{pmatrix} r_o\cos^2\phi+r_e\sin^2\phi & (r_o-r_e)\cos\phi\sin\phi \\ (r_o-r_e)\cos\phi\sin\phi & r_o\sin^2\phi+r_e\cos^2\phi \end{pmatrix}. \tag{8.72}$$

Note that R does not contain δ, even though M does, through $\mathbf{E}^e$:

$$M=\begin{pmatrix} \cos\phi & \sin\phi\cos\delta \\ \sin\phi & -\cos\phi\cos\delta \end{pmatrix}. \tag{8.73}$$

The reflection coefficients r_o and r_e are the same as the normal incidence reflection amplitudes for isotropic media with refractive indices n_o and $n_o n_e/n_\gamma$, respectively:

$$r_o=\frac{n_1-n_o}{n_1+n_o}, \qquad r_e=\frac{n_1-n_o n_e/n_\gamma}{n_1+n_o n_e/n_\gamma}. \tag{8.74}$$

When $\gamma^2=1$ the reflection amplitudes r_o and r_e are equal. Also $r_e=(n_1-n_e)/(n_1+n_e)$ when $\gamma=0$.

From (8.71) and (8.72) we find

$$r=r_o\cos^2\phi+r_e\sin^2\phi, \qquad r'=(r_o-r_e)\cos\phi\sin\phi. \tag{8.75}$$

When there is no absorption r_o and r_e are real, and thus when the incident light is linearly polarized (as we have assumed), the reflected light is also linearly polarized, with electric field direction rotated by $\arctan(r'/r)$. Note that r', which gives the amplitude of the electric field reflected into the y-polarization when the incident electric field $\mathbf{E}_1$ is x-polarized, is zero when $\mathbf{E}_1$ is either parallel or perpendicular to $\mathbf{E}^o$ ($\sin\phi=0$ or $\cos\phi=0$). It has maximum amplitude at ϕ equal to odd multiples of 45°; the maximum value of $|r'|$ at $\gamma=0$ (optic axis in the reflecting plane) is $n_1|n_e-n_o|/[(n_1+n_o)(n_1+n_e)]$.

The transmission amplitudes are found from (8.71) and (8.72):

$$\begin{aligned} t_o &= (1 + r_o)\cos\phi = \tfrac{2n_1}{n_1 + n_o}\cos\phi, \\ t_e &= (1 + r_e)\sin\phi/\cos\delta = \tfrac{2n_1 n_\gamma}{n_1 n_\gamma + n_o n_e}\sin\phi/\cos\delta. \end{aligned} \tag{8.76}$$

If the incident wave is polarized with $\mathbf{E}_1$ along the $\mathbf{E}^o$ (or $\mathbf{n} \times \mathbf{c}$) direction, $\sin\phi = 0$ and only the ordinary wave will propagate into the crystal. Only the extraordinary wave will be excited when $\phi = \pm 90°$.

We have introduced the 2×2 matrix method for dealing with normal incidence partly because it generalizes to the treatment of normal incidence on crystal plates, as discussed in the next section.

8.7 Normal Incidence on a Uniaxial Plate

We wish to find the reflection and transmission amplitudes for normal incidence on a crystal plate of thickness Δz. We now have backward as well as forward propagation within the plate, and interference between them. We suppose the medium of incidence to have refractive index n_1 as before, and the substrate to have index n_2. For incident light linearly polarized along the x direction, the electric fields are

$$\begin{aligned} &\textit{incident} && (1,\ 0,\ 0)\mathrm{e}^{ik_1 z} \\ &\textit{reflected} && (r,\ r',\ 0)\mathrm{e}^{-ik_1 z} \\ &\textit{within crystal} && (a_o\mathrm{e}^{ik_o z} + b_o\mathrm{e}^{-ik_o z})\mathbf{E}^o + (a_e\mathrm{e}^{ik_e z} + b_e\mathrm{e}^{-ik_e z})\mathbf{E}^e \\ &\textit{transmitted} && (t,\ t',\ 0)\mathrm{e}^{ik_2(z-\Delta z)} \end{aligned} \tag{8.77}$$

(We have simplified the results by absorbing a phase $k_2\Delta z$ into the transmission coefficients t and t'.) The continuity of E_x, E_y, $\partial E_x/\partial z$ and $\partial E_y/\partial z$ at $z = 0$ and $z = \Delta z$ leads to eight equations in the eight unknowns r, r', a_o, b_o, a_e, b_e, t, t'. In Lekner (1992a) this 8×8 problem is reduced to four coupled 2×2 problems, which are then solved in terms of 2×2 matrices (these include the mode matrix M of the previous section, and also K_1 and its analogue K_2). The solutions are

$$r = r_o\cos^2\phi + r_e\sin^2\phi, \qquad r' = (r_o - r_e)\cos\phi\sin\phi, \tag{8.78}$$

(the same forms as (8.75), but with different r_o and r_e, defined below), and

$$t = t_o\cos^2\phi + t_e\sin^2\phi, \qquad t' = (t_o - t_e)\cos\phi\sin\phi. \tag{8.79}$$

As in the previous section, ϕ is the angle between the incident electric field $\mathbf{E}_1$ and the $\mathbf{E}^o$ or $\mathbf{n} \times \mathbf{c}$ direction. The component reflection amplitude r_o is found to be

$$r_o = \frac{k_o(k_1 - k_2)\cos(k_o\Delta z) + i(k_o^2 - k_1k_2)\sin(k_o\Delta z)}{k_o(k_1 + k_2)\cos(k_o\Delta z) - i(k_o^2 - k_1k_2)\sin(k_o\Delta z)}. \tag{8.80}$$

This is the reflection amplitude for a homogeneous isotropic layer of index n_o and thickness Δz (Equation 2.52, with $z_1 = 0$). The amplitude r_e, obtained by replacing k_o by k_e in (8.80), is likewise the reflection amplitude for an isotropic slab of index $n_o n_e/n_\gamma$. An alternative form of (8.80) is

$$r_o = \frac{r_1^o + r_2^o e^{2ik_o\Delta z}}{1 + r_1^o r_2^o e^{2ik_o\Delta z}}, \qquad r_1^o = \frac{k_1 - k_o}{k_1 + k_o}, \qquad r_2^o = \frac{k_o - k_2}{k_o + k_2}, \tag{8.81}$$

(expressed in terms of the Fresnel amplitudes r_1^o and r_2^o for reflection at the entry and exit faces of a slab of index n_o).

The component amplitudes t_o and t_e are likewise the transmission amplitudes for isotropic slabs of thickness Δz and refractive indices n_o and $n_o n_e/n_\gamma$ (Equations 2.53 and 2.59)

$$t_o = \frac{2k_1k_o}{k_o(k_1 + k_2)\cos(k_o\Delta z) - i(k_o^2 + k_1k_2)\sin(k_o\Delta z)} = \frac{(1 + r_1^o)(1 + r_2^o)e^{ik_o\Delta z}}{1 + r_1^o r_2^o e^{2ik_o\Delta z}}. \tag{8.82}$$

The transmission amplitude t_e is obtained by replacing k_o by k_e, r_1^o and r_2^o by r_1^e and r_2^e, as for the reflection amplitude r_e.

The reflection properties of a uniaxial layer are discussed in detail by Lekner (1992a). Both r_o and r_e have the bilinear (or fractional) complex transformation form

$$\frac{r_1 + r_2 Z}{1 + r_1 r_2 Z}, \qquad Z = \exp 2ik\Delta z). \tag{8.83}$$

Thus as $k\Delta z$ increases Z moves on the unit circle in the complex plane (assuming non-absorbing crystal and substrate) r_o and r_e will also move on circles. The periods of Δz of r_o and r_e are π/k_o and π/k_e. The $Z = 1$ ($\Delta z = 0$ or multiples of π/k_o or π/k_e) common value of r_o and r_e is, as expected, the Fresnel amplitude for a sudden transition from index n_1 to index n_2:

$$r^+ = \frac{r_1 + r_2}{1 + r_1 r_2} = \frac{k_1 - k_2}{k_1 + k_2}. \tag{8.84}$$

When $Z = 1$ there will be no reflection if $n_1 = n_2$. At $Z = -1$ ($k\Delta z$ an odd multiple of $\pi/2$), r_o and r_e take the different values obtained by setting $k = k_o$ or k_e in

$$r^- = \frac{r_1 - r_2}{1 - r_1 r_2} = \frac{k_1 k_2 - k^2}{k_1 k_2 + k^2}. \tag{8.85}$$

Thus r_o^- is zero when $n_o^2 = n_1 n_2$, r_e^- is zero when $(n_o n_e / n_\gamma)^2 = n_1 n_2$.

The centre of the circles, which in the absence of absorption lie on the real axis, is at the point $(r^+ + r^-)/2$. Their radius is $(r^+ - r^-)/2$. Figure 8.3 shows the r_o and r_e circles for calcite in air $(n_1 = 1 = n_2)$. The r_o circle is fixed, while the r_e circle depends on γ, the direction cosine of the optic axis to the surface normal. All the circles have the origin $(Z = 1)$ as common point, since this is shared by the zero-thickness value of the reflection amplitudes (when $n_1 = n_2$). The other intersections with the real axis are at $Z = -1$.

The situation is different for $n_1 \neq n_2$, and for absorbing substrates. The reflection amplitudes r_o and r_e still move on circles, however; examples of a calcite slab on Al or Si are shown in Fig. 2 of Lekner (1992a).

We now briefly discuss the *transmission amplitudes* t_o and t_e. These move on a quartic in the complex plane; their reciprocals move on ellipses (details may be found in Lekner 1992a). For incident electric field $\mathbf{E}_1$ either parallel or perpendicular to $\mathbf{n} \times \mathbf{c}$ (the direction of $\mathbf{E}^o$), the exit polarization is the same as that on entry. These orientations thus give zero transmission between crossed polarizers. In the general case of a crystal plate between polarizer and analyser, with angle χ between their easy axes, the electric field transmitted through the analyser is, for incident field of unit amplitude,

$$(t \cos\chi + t' \sin\chi) e^{i(k_2 z - \omega t)}. \tag{8.86}$$

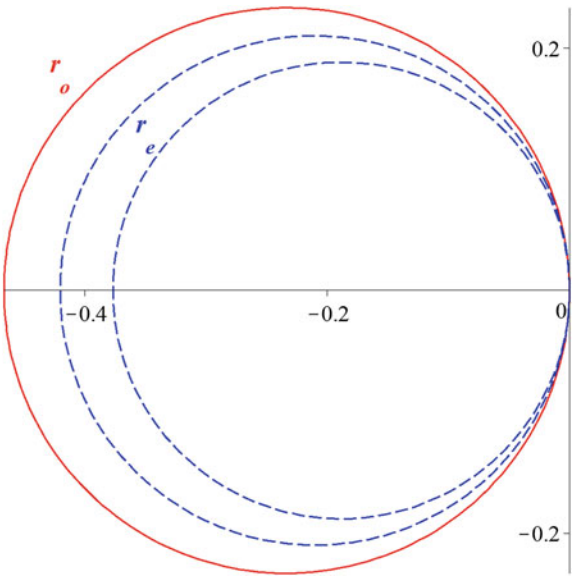


Fig. 8.3 The loci of r_o and r_e in the complex plane, for calcite ($n_o = 1.658$, $n_e = 1.486$) in air. The r_e circles (*dashed*) are drawn for $\gamma = 0$ (*inner circle*) and for $\gamma^2 = \frac{1}{2}$

Since $\mathbf{E}_1$ defines the x-axis, this is also the polarizer easy axis, and the analyser is at angle χ to the x-axis. The transmitted intensity of a broad beam, measured well within the beam, is the absolute square of (8.86),

$$|t|^2\cos^2\chi + \left|t^{'}\right|^2\sin^2\chi + 2\left|tt^{'}\right|\cos\chi\sin\chi\cos\left[ph(t^{'}/t)\right], \tag{8.87}$$

where we use the modulus-phase notation $\xi = |\xi|\exp[i\,ph(\xi)]$. When the polarizer and analyser are crossed ($\chi = 90°$), the intensity is

$$\left|t^{'}\right|^2 = |t_o - t_e|^2\cos^2\phi\sin^2\phi, \tag{8.88}$$

which is zero when ϕ is zero or a multiple of 90°. Both t_o and t_e depend on the thickness of the plate, and t_e also depends on γ through n_γ. The absolute square of their difference which features in (8.88) is

$$|t_o - t_e|^2 = |t_o|^2 + |t_e|^2 - 2|t_o t_e|\cos[ph(t_e/t_o)]. \tag{8.89}$$

The phase difference between t_e and t_o is often given as $(k_e - k_o)\Delta z$ (see for example Born and Wolf (1970), 14.4 (8.15)). From (8.82) we see that, if the entry and exit planes of the crystal are accurately parallel, it is actually

$$\begin{aligned} ph\left(\frac{t_e}{t_o}\right) &= ph(t_e) - ph(t_o) \\ &= \mathrm{atn}\left[\frac{k_o^2 + k_1k_2}{k_o(k_1 + k_2)}\tan(k_o\Delta z)\right] - \mathrm{atn}\left[\frac{k_o^2 + k_1k_2}{k_o(k_1 + k_2)}\tan(k_o\Delta z)\right]. \end{aligned} \tag{8.90}$$

The difference between (8.90) and the approximate value $(k_e - k_o)\Delta z$ is due to multiple reflections within the crystal (which could be eliminated at a given wavelength by antireflection coatings). The thin film limit of (8.90) is

$$ph(t_e/t_o) = \frac{k_e^2 - k_o^2}{k_1 + k_2}\Delta z + O(\Delta z)^3, \tag{8.91}$$

which differs from the approximate value $(k_e - k_o)\Delta z$ by the factor $(k_e + k_o)/(k_1 + k_2)$.

We note finally that the characterization of the transmitted light in terms of the transmission amplitudes t and $t^{'}$ applies to beams which are wide enough to be accurately represented by the last line of (8.77). For narrow beams passing through thick crystals there will be complete separation of the o and e rays within the crystal, and two parallel beams will exit the crystal, perpendicularly polarized in the $\mathbf{E}^o$ and $\mathbf{E}^e$ directions. The transmission amplitudes for these exit beams are $t_o\cos\phi$ and $t_e\sin\phi$. The identity

$$|t_o|^2\cos^2\phi + |t_e|^2\sin^2\phi = \left|t_o\cos^2\phi + t_e\sin^2\phi\right|^2 + |t_o - t_e|^2\cos^2\phi\sin^2\phi \tag{8.92}$$

shows that for given incident power, the transmitted total power in the two beams is the same as it would be for a single very broad beam. (The right-hand side of (8.92) equals $|t|^2 + |t'|^2$.)

8.8 Isotropic Layer on a Uniaxial Substrate

The optical properties of a homogeneous isotropic layer on an isotropic substrate are discussed in Sect. 2.4. They are contained in two reflection amplitudes r_s and r_p, and two transmission amplitudes t_s and t_p. When the isotropic layer rests on an anisotropic substrate, four reflection amplitudes r_{ss}, r_{sp}, r_{pp} and r_{ps}, and four transmission amplitudes t_{so}, t_{se}, t_{po}, t_{pe} are required.

An example, of considerable geophysical importance, is that of a thin layer of water on the surface of ice below 0 °C. The compaction of snow, frost heave, rock fracture, water transport at subzero temperatures, and charge transfer in the electrification of thunder clouds are some of the aspects of premelting of ice discussed by Dash (1989). Premelting of crystals is widespread, if not universal (Dash 1999). Reflection anisotropy spectroscopy is covered in the review by McGilp (1995).

The isotropic layer has dielectric constant $\varepsilon = n^2$ and is bounded by the medium of incidence $(\varepsilon_1 = n_1^2)$ at $z = 0$, and by a uniaxial substrate $(\varepsilon_o = n_o^2,\ \varepsilon_o = n_e^2)$ at $z = \Delta z$. The plane of incidence is the zx plane; the direction cosines of the optic axis of the substrate are the components of the unit vector $\mathbf{c} = (\alpha, \beta, \gamma)$. Thus the ordinary and extraordinary modes have the wave vectors and electric fields derived in Sect. 8.2. With the common factor $\exp i(Kx - \omega t)$ suppressed, and an s-polarized wave of unit amplitude incident at angle θ, the electric fields are

$$\begin{aligned}
&\textit{incident}\colon (0,\ 1,\ 0)\mathrm{e}^{iq_1 z}\\
&\textit{reflected}\colon \left(r_{sp}\cos\theta_1,\ r_{ss},\ r_{sp}\sin\theta_1\right)\mathrm{e}^{-iq_1 z}\\
&\textit{within layer}\colon \left(\cos\theta\left[a\mathrm{e}^{iqz} + b\mathrm{e}^{-iqz}\right],\ A\mathrm{e}^{iqz} + B\mathrm{e}^{-iqz},\ -\sin\theta\left[a\mathrm{e}^{iqz} - b\mathrm{e}^{-iqz}\right]\right)\\
&\textit{within crystal}\colon t_{so}\mathbf{E}^o\mathrm{e}^{iq_o(z-\Delta z)} + t_{se}\mathbf{E}^o\mathrm{e}^{iq_e(z-\Delta z)}
\end{aligned} \tag{8.93}$$

Within the layer the inward propagating part has Poynting vector (proportional to $\mathbf{E}\times\mathbf{B}$) along $(K, 0, q)$, and the outward propagating part has $\mathbf{E}\times\mathbf{B}$ along $(K, 0, -q)$, with proportionality constants $A^2 + a^2$ and $B^2 + b^2$, respectively. These follow from $(K = k_1\sin\theta_1 = k\sin\theta,\ q = k\cos\theta)$

$$q\cos\theta + K\sin\theta = n\,\omega/c = k. \tag{8.94}$$

There are eight unknowns in (8.93) (r_{ss}, r_{sp}, a, b, A, B, t_{so} and t_{se}) and eight equations follow from the continuity of E_x, E_y, $\partial_z E_x - iKE_z$ and $\partial_z E_y$ at $z = 0$ and at $z = \Delta z$. The resulting s to s reflection amplitude may be written in the form of that for an isotropic layer on an isotropic substrate:

$$r_{ss} = \frac{f_1 + fZ}{1 + f_1 fZ}, \qquad f_1 = \frac{q_1 - q}{q_1 + q}, \qquad Z = \mathrm{e}^{2iq\Delta z}. \tag{8.95}$$

The reflection amplitude f_1 is the Fresnel amplitude at the $z = 0$ interface; f is more complicated (Lekner 1992c, Sect. 4). However, we can see from (8.95) that as Δz changes, Z moves on the unit circle, and thus r_{ss} will also move around on a circle in the complex plane (refer to the discussion in Sect. 8.7; absence of absorption is assumed). The period Δz of this notion is π/q.

For incident p-polarization, the electric fields are

$$\begin{gathered}
\textit{incident}\colon (\cos\theta_1,\ 0,\ -\sin\theta_1)\mathrm{e}^{iq_1 z} \\
\textit{reflected}\colon (r_{pp}\cos\theta_1,\ r_{ps},\ r_{pp}\sin\theta_1)\mathrm{e}^{-iq_1 z} \\
\textit{within layer}\colon \left(\cos\theta[a\mathrm{e}^{iqz} + b\mathrm{e}^{-iqz}],\ A\mathrm{e}^{iqz} + B\mathrm{e}^{-iqz},\ -\sin\theta[a\mathrm{e}^{iqz} - b\mathrm{e}^{-iqz}]\right) \\
\textit{within crystal}\colon t_{po}\mathbf{E}^o\mathrm{e}^{iq_o(z-\Delta z)} + t_{pe}\mathbf{E}^e\mathrm{e}^{iq_e(z-\Delta z)}
\end{gathered} \tag{8.96}$$

Again r_{pp} may be written in the form of r_p for an isotropic layer on an isotropic substrate (Sect. 2.4)

$$r_{pp} = \frac{F_1 + FZ}{1 + F_1 FZ}, \qquad F_1 = \frac{Q - Q_1}{Q + Q_1} \qquad \left(Q_1 = \frac{q_1}{\varepsilon_1},\ Q = \frac{q}{\varepsilon}\right). \tag{8.97}$$

(F is made explicit in Sect. 4 of Lekner 1992c). $Z = \mathrm{e}^{2iq\Delta z}$ as before; as $q\Delta z$ changes Z moves on the unit circle, and thus r_{pp} will move on a circle in the complex plane (in the absence of absorption within the layer). The off-diagonal reflection amplitudes r_{sp} and r_{ps} are given in (86) of Lekner (1992c); they also move on circles in the complex plane as the thickness of the isotropic layer varies, as is seen in the example of air|water|calcite at 30° and 60° angle of incidence, shown in Fig. 1 of that paper.

8.9 Optical Properties of a Uniaxial Layer

In Sect. 8.7 we have discussed the normal incidence case, which was amenable to 2×2 matrix treatment. The general case requires 4×4 matrix representation, but an algebraically explicit solution is again possible (Lekner 1994b). The medium and the substrate have dielectric constants $\varepsilon_1 = n_1^2$ and $\varepsilon_2 = n_2^2$, the crystal has ordinary and extraordinary indices ε_o and ε_e. For s-polarization incident, and with the common factor $\exp i(Kx - \omega t)$ suppressed, the electric fields are

$$\begin{gathered}
\textit{incident}: (0,\ 1,\ 0)\mathrm{e}^{iq_1 z} \\
\textit{reflected}: (r_{sp}\cos\theta_1,\ r_{ss},\ r_{sp}\sin\theta_1)\mathrm{e}^{-iq_1 z} \\
\textit{within layer}: a_o\mathbf{E}_o^+\mathrm{e}^{iq_o z} + b_o\mathbf{E}_o^-\mathrm{e}^{-iq_o z} + a_e\mathbf{E}_e^+\mathrm{e}^{iq_e^+ z} + b_e\mathbf{E}_e^-\mathrm{e}^{iq_e^- z} \\
\textit{transmitted}: (t_{sp}\cos\theta_2,\ t_{ss},\ -t_{sp}\sin\theta_2)\mathrm{e}^{iq_2(z-\Delta z)}
\end{gathered} \tag{8.98}$$

As usual, $q_o^2 = k_o^2 - K^2$, $k_o = n_o\omega/c$. The four plane waves that can propagate within the crystal have the z-dependence

$$\begin{gathered}
\mathrm{e}^{iq_o z},\ \mathrm{e}^{-iq_o z},\ \mathrm{e}^{iq_e^+ z} \text{ and } \mathrm{e}^{iq_e^- z}, \\
q_e^{\pm} = \pm\bar{q} - \tfrac{\gamma K\Delta\varepsilon}{\varepsilon_\gamma},\ \bar{q}^2 = \tfrac{\varepsilon_o}{\varepsilon_\gamma^2}\left\{\varepsilon_e\varepsilon_\gamma\left(\tfrac{\omega}{c}\right)^2 - K^2\varepsilon_{\alpha\gamma}\right\}.
\end{gathered} \tag{8.99}$$

There are eight unknowns, r_{ss}, r_{sp}, a_o, b_o, a_e, b_e, t_{ss} and t_{sp}, and eight equations arising from the continuity of E_x, E_y, $\partial_z E_x - iKE_z$ and $\partial_z E_y$ at $z = 0$ and at $z = \Delta z$. The solution is facilitated by a diagonal phase matrix,

$$P = \mathrm{diag}\left(\mathrm{e}^{iq_o\Delta z},\ \mathrm{e}^{-iq_o\Delta z},\ \mathrm{e}^{iq_e^+\Delta z},\ \mathrm{e}^{iq_e^-\Delta z}\right), \tag{8.100}$$

and by 4×4 mode matrices M and layer matrix

$$L = MPM^{-1}. \tag{8.101}$$

These same matrices also give the solution (in terms of the matrix elements L_{ij}) for the p-polarization, for which the fields inside the crystal have the form given in (8.98), the others being

$$\begin{array}{rc}
\textit{incoming}: & (\cos\theta_1,\ 0,\ -\sin\theta_1)\mathrm{e}^{iq_1 z} \\
\textit{reflected}: & (r_{pp}\cos\theta_1,\ r_{ps},\ r_{pp}\sin\theta_1)\mathrm{e}^{-iq_1 z} \\
\textit{transmitted}: & (t_{pp}\cos\theta_2,\ t_{ps},\ -t_{pp}\sin\theta_2)\mathrm{e}^{iq_2 z}
\end{array} \tag{8.102}$$

The reflection and transmission amplitudes at oblique incidence and at arbitrary orientation of the optic axis with respect to the reflecting surface normal are given

in Sect. 3 of Lekner (1994b). For non-absorbing media they satisfy the energy conservation conditions

$$\begin{aligned} q_1\left(1-|r_{ss}|^2-|r_{sp}|^2\right) &= q_2\left(|t_{ss}|^2+|t_{sp}|^2\right) \\ q_1\left(1-|r_{pp}|^2-|r_{ps}|^2\right) &= q_2\left(|t_{pp}|^2+|t_{ps}|^2\right) \end{aligned} \tag{8.103}$$

There is much simplification when the *optic axis lies in the plane of incidence* ($\beta = 0$). Then the ordinary electric field vector is perpendicular to the plane of incidence (the **s** direction), and from (8.24), $\mathbf{E}^e$ is perpendicular to $\mathbf{E}^o$. Thus the *s*-polarization converts fully to the ordinary mode, and r_{sp} and t_{sp} are zero. The reflection amplitude r_{ss} is the same as that of an isotropic layer of index n_o:

$$r_{ss} = \frac{s_1 + s_2 e^{2iq_o\Delta z}}{1 + s_1 s_2 e^{2iq_o\Delta z}}, \qquad s_1 = \frac{q_1 - q_o}{q_1 + q_o}, \qquad s_2 = \frac{q_o - q_2}{q_o + q_2}. \tag{8.104}$$

The amplitude r_{pp} also takes the isotropic layer form,

$$r_{pp} = \frac{p_1 + p_2 e^{2i\bar{q}\Delta z}}{1 + p_1 p_2 e^{2i\bar{q}\Delta z}}, \qquad p_1 = \frac{Q - Q_1}{Q + Q_1}, \qquad p_2 = \frac{Q_2 - Q}{Q_2 + Q}. \tag{8.105}$$

with the definitions

$$\begin{aligned} &Q_1 = \tfrac{q_1}{\varepsilon_1}, \quad Q_2 = \tfrac{q_2}{\varepsilon_2}, \quad Q = \tfrac{q_\gamma}{n_o n_e}, \\ &q_\gamma = \varepsilon_\gamma(\omega/c)^2 - K^2, \quad \bar{q} = (n_o n_e/\varepsilon_\gamma) q_\gamma. \end{aligned} \tag{8.106}$$

(The value for $\bar{q}$ given is that taken by the general $\bar{q}$ defined in (8.19) when $\beta = 0$; q_γ is given in (8.60).) The transmission amplitudes, when the optic axis lies in the plane of incidence, are

$$\begin{aligned} t_{ss} &= \tfrac{(1+s_1)(1+s_2)e^{iq_o\Delta z}}{1 + s_1 s_2 e^{2iq_o\Delta z}}, \\ t_{pp} &= e^{-i\alpha\gamma K\Delta z\Delta\varepsilon/\varepsilon_\gamma}\, \tfrac{n_1}{n_2}\, \tfrac{(1-p_1)(1-p_2)e^{i\bar{q}\Delta z}}{1 + p_1 p_2 e^{2i\bar{q}\Delta z}}. \end{aligned} \tag{8.107}$$

References

Born M, Wolf E (1970) Principles of optics, 3 edn. Pergamon

Dash JG (1989) Surface melting. Contemporary Phys 30:89–100

Dash JG (1999) History of the search for continuous melting. Rev Mod Phys 71:1737–1743

Lekner J (1991) Reflection and refraction by uniaxial crystals. J Phys: Condens Matter 3:6121–6133

Lekner J (1992a) Normal incidence reflection and transmission by uniaxial crystals and crystal plates. J Phys: Condens Matter 4:1387–1398

Lekner J (1992b) Bounds and zeros in reflection and refraction by uniaxial crystals. J Phys: Condens Matter 4:9459–9468
Lekner J (1992c) Optical properties of an isotropic layer on a uniaxial crystal substrate. J Phys: Condens Matter 4:6569–6586
Lekner J (1993) Brewster angles in reflection by uniaxial crystals. J Opt Soc Am 10:2059–2064
Lekner J (1994a) Normal incidence transmission ellipsometry of anisotropic layers. Pure Appl Opt 3:307–321
Lekner J (1994b) Optical properties of a uniaxial layer. Pure Appl Opt 3:821–837
Lekner J (1999) Reflection by uniaxial crystals: polarizing angle and Brewster angle. J Opt Soc Am 11:2763–2766
McGilp JF (1995) Optical characterization of semiconductor surfaces and interfaces. Prog Surf Sci 49:1–106

Chapter 9
Ellipsometry

Ellipsometry is a sensitive technique based on bringing s and p reflected or transmitted polarizations into interference. Its main use is in the characterization of solid surfaces and liquid-vapour or liquid-liquid interfaces. Azzam (1991) has compiled a selection of papers on ellipsometry. A survey of the various optical techniques used to study surfaces was given by McGilp (1995). We shall first consider five reflection configurations, in the general anisotropic case where all of r_{pp}, r_{ss}, r_{ps} and r_{sp} may be active (Lekner 1993); reflection ellipsometry of uniaxial crystals is applied in Sect. 9.9 to the extraction of the ordinary and extraordinary indices and of the orientation of the optic axis. Reflection from isotropic media, with only r_p and r_s active, will follow as special cases. Transmission ellipsometry will be considered in Sects. 9.7 and 9.8.

The essence of all reflection ellipsometric methods is as follows. A polarizer produces a known proportion of in-phase p and s incident waves. The amplitude and phase of these waves are altered by reflection, as specified by the complex reflection amplitudes r_{pp}, r_{ps} and r_{ss}, r_{sp}. The reflected light is passed through an analyser, which combines the components of the orthogonal p and s polarizations along the analyser easy direction. The intensity then measured by a detector is the result of interference of the p and s components and thus contains information about the relative phases of the reflection amplitudes, as well as about their magnitudes.

9.1 Polarizer–Sample–Analyser

This is the simplest ellipsometer configuration. Let P be the angle between the polarizer easy axis and the **p** direction (which is in the plane of incidence, and perpendicular to the incident ray). The angles P and A are measured from the **p** and $\mathbf{p}'$ directions towards the **s** directions, with the vectors **p**, **s** and **k** (the wave vector) forming a right-handed triplet. Likewise the $\mathbf{p}', \mathbf{s}$ and $\mathbf{k}'$ directions form a right-handed triplet. Note that the reflected $\mathbf{p}'$ and $\mathbf{k}'$ directions are different from the incident ones, whereas the **s** vector is fixed. Figure 9.1 gives the schematics of the polarizer–sample–analyser configuration.

J. Lekner, *Theory of Reflection*, Springer Series on Atomic, Optical, and Plasma Physics 87, DOI 10.1007/978-3-319-23627-8_9

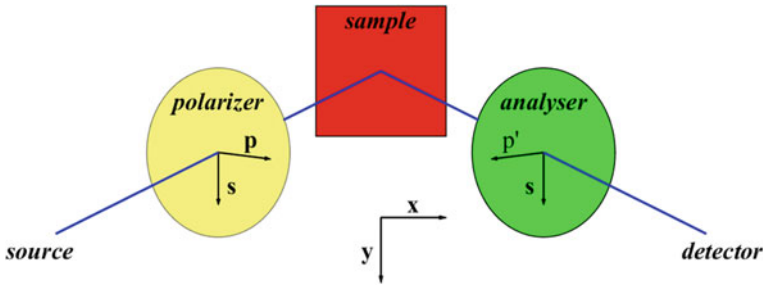


Fig. 9.1 The simplest reflection ellipsometry configuration. The polarizer easy axis is at azimuthal angle P, measured from the **p** toward the **s** directions. The analyser easy axis is at azimuthal angle A, measured from the $\mathbf{p}'$ toward the **s** directions. The **p** and $\mathbf{p}'$ directions are antiparallel at normal incidence, parallel at glancing incidence

Then, on removal of common factors which cancel from ratios of detected intensities, the electric field has p and s components $\cos P$ and $\sin P$ after passing through the polarizer. After reflection, the p and s components are

$$r_{pp}\cos P + r_{sp}\sin P, \quad r_{ps}\cos P + r_{ss}\sin P. \tag{9.1}$$

These components are combined by the analyser. If this is set at angle A to the (reflected) $\mathbf{p}'$ direction, the field transmitted by the analyser is

$$\begin{aligned}&(r_{pp}\cos P + r_{sp}\sin P)\cos A + (r_{ps}\cos P + r_{ss}\sin P)\sin A\\ &= \cos P\cos A\left[r_{pp} + r_{sp}\tan P + (r_{ps} + r_{ss}\tan P)\tan A\right]\\ &= \cos P\cos A\left[r_{pp} + r_{ps}\tan A + (r_{sp} + r_{ss}\tan A)\tan P\right].\end{aligned} \tag{9.2}$$

We define two ellipsometric ratios (which will serve in all of the configurations we discuss)

$$\rho_P = \frac{r_{pp} + r_{sp}\tan P}{r_{ps} + r_{ss}\tan P}, \qquad \rho_A = \frac{r_{pp} + r_{ps}\tan A}{r_{sp} + r_{ss}\tan A}. \tag{9.3}$$

Consider two intensity measurements at analyser angles A_1 and A_2, and fixed polarizer angle P. Their ratio is

$$\frac{I(A_1)}{I(A_2)} = \left(\frac{\cos A_1}{\cos A_2}\right)^2 \left|\frac{\rho_P + \tan A_1}{\rho_P + \tan A_2}\right|^2. \tag{9.4}$$

Similarly, if two intensity measurements are made at fixed analyser angle A, at two polarizer angles P_1 and P_2, their ratio will be

$$\frac{I(P_1)}{I(P_2)} = \left(\frac{\cos P_1}{\cos P_2}\right)^2 \left|\frac{\rho_A + \tan P_1}{\rho_A + \tan P_2}\right|^2. \tag{9.5}$$

Measurements of this kind can thus give the absolute squares and the real parts of ρ_P or of ρ_A, respectively. The signs of the imaginary parts of ρ_P and ρ_A are not determined. In the isotropic case ρ_P becomes $(r_p/r_s)\cot P$, and ρ_A becomes $(r_p/r_s)\cot A$.

9.2 Polarizer–Compensator–Sample–Analyser

A compensator (or waveplate, or retarder) is a crystal plate, or arrangement of plates, that produces a known phase difference between two orthogonal components. For example, a wave normally incident onto a uniaxial crystal will split into two orthogonal components that travel in the crystal with phase speeds c/n_o and $cn_\gamma/n_o n_e$, where n_o and n_e are the ordinary and extraordinary refractive indices of the crystal, and $n_\gamma^2 = \varepsilon_\gamma = \varepsilon_o + \gamma^2\Delta\varepsilon$, γ being the direction cosine of the optic axis with the inward normal. Section 8.7 gave the transmission amplitudes of the ordinary and extraordinary modes, and their relative phase allowing for all internal reflections in (8.90). The phase difference when reflections at the plate surfaces are removed by antireflection coatings is

$$\delta_e - \delta_o = \frac{\omega}{c}\left(\frac{n_o n_e}{n_\gamma} - n_o\right)\Delta z. \tag{9.6}$$

Consider the polarizer–compensator–sample–analyser configuration, with the polarizer easy axis at angle P to the **p** direction, and the compensator with its **o** direction at angle C to the **p** direction. The field components along the **o** and **e** directions are

	along **o**	along **e**
after polarizer	$\cos(C-P)$	$-\sin(C-P)$
after compensator	$t_o\cos(C-P)$	$-t_e\sin(C-P)$

(9.7)

We now resolve along the **s** and **p** directions. The respective field components are

$$\begin{aligned} E_p &= t_o\cos C\cos(C-P) + t_e\sin C\sin(C-P), \\ E_s &= t_o\sin C\cos(C-P) - t_e\cos C\sin(C-P). \end{aligned} \tag{9.8}$$

After reflection these become

$$E_p^r = r_{pp}E_p + r_{sp}E_s, \qquad E_s^r = r_{ss}E_s + r_{ps}E_p. \tag{9.9}$$

These reflected components are combined by the analyser, set at angle A to the $\mathbf{p}'$ direction. The field transmitted by the analyser is

$$\left(r_{pp}E_p + r_{sp}E_s\right)\cos A + \left(r_{ss}E_s + r_{ps}E_p\right)\sin A. \tag{9.10}$$

The intensity measured is proportional to the absolute square of the quantity (9.10). In null ellipsometry this is made zero (in practice minimized) by adjustments of the angles P, C and A. It is useful to define the complex ratio τ and the complex angle D by

$$\tau = t_e/t_o, \qquad \tan D = \tau \tan(C - P). \tag{9.11}$$

Then from (9.8) the ratio of E_s to E_p is given by

$$E_s/E_p = \tan(C - D). \tag{9.12}$$

It follows that the zero of expression (9.10) occurs when

$$\rho_A = \tan(D - C). \tag{9.13}$$

Thus a null setting of the polarizer–compensator–sample–analyser reflection ellipsometer determines the real and imaginary parts of $\rho_A = (r_{pp} + r_{ps}\tan A)/(r_{sp} + r_{ss}\tan A)$. In the isotropic case (9.13) reduces to

$$r_p/r_s = \tan A \tan(D - C). \tag{9.14}$$

(Compare (3.33) of Azzam and Bashara (1987).) In null ellipsometry, such as that just described, one obtains information purely by angle measurement: one does not measure light intensity, but rather locates its minimum.

9.3 Polarizer–Sample–Compensator–Analyser

In this configuration the field components along **s** and along **p** are $\cos P$ and $\sin P$ after the polarizer, and after reflection these become

$$E_p' = r_{pp}\cos P + r_{sp}\sin P, \qquad E_s' = r_{ss}\sin P + r_{ps}\cos P. \tag{9.15}$$

With the compensator **o** direction at angle C to $\mathbf{p}'$, the components along the **o** and **e** directions after passing through the compensator are

$$t_o\left(E_p^{'}\cos C + E_s^{'}\sin C\right), \qquad t_e\left(E_s^{'}\cos C - E_p^{'}\sin C\right). \tag{9.16}$$

Thus the electric field transmitted by the analyser set at angle A to the $\mathbf{p}^{'}$ direction is

$$t_o(E_p^{'}\cos C + E_s^{'}\sin C)\cos(C-A) - t_e(E_s^{'}\cos C - E_p^{'}\sin C)\sin(C-A). \tag{9.17}$$

We again define $\tau = t_e/t_o$, with the understanding that the complex number τ can represent any compensator, and introduce the complex angle $D^{'}$, defined by

$$\tan D^{'} = \tau\tan(C-A). \tag{9.18}$$

The field (9.17) is zero when

$$\rho_P = \tan(D^{'} - C). \tag{9.19}$$

A null setting of the ellipsometer thus determines the real and imaginary parts of $\rho_P = (r_{pp} + r_{sp}\tan P)/(r_{ps} + r_{ss}\tan P)$. In this isotropic case, (9.19) reduces to

$$r_p/r_s = \tan P\tan(D^{'} - C). \tag{9.20}$$

(Compare (3.55) of Azzam and Bashara (1987).)

9.4 Polarizer–Modulator–Sample–Analyser

We now turn to polarization–modulation ellipsometry (Jasperson and Schnatterly 1969; Beaglehole 1980), in which the polarization state of the light is varied sinusoidally, with synchronous detection of the intensity by lock-in amplifiers.

When the polarizer pass direction is at angle P to the **p** direction, the p and s components are $\cos P$ and $\sin P$. The birefringent modulator is oriented so that its o and e directions lie along **p** and **s** respectively. After passing through the modulator the p and s components are $t_o\cos P$ and $t_e\sin P$. The modulator is (for example) a piezo-electric transducer, consisting of a block of fused quartz through which the light passes, joined onto a block of crystal quartz which oscillates at its fundamental frequency of say 50 kHz. The resulting sinusoidal uniaxial strain in the fused quartz modulates the phase difference between the transmitted o and e components,

$$\delta(t) = M\sin(\Omega t). \tag{9.21}$$

M is the maximum phase shift induced by the modulator, $\Omega/2\pi$ the modulation frequency. In practice the transmission amplitude magnitudes $|t_o|$ and $|t_e|$ are almost identical, so the p and s components after passing through the modulator are, up to a common factor,

$$\cos P \quad \text{and} \quad e^{i\delta} \sin P. \tag{9.22}$$

After reflection, the $\mathbf{p}'$ and $\mathbf{s}$ components are

$$\begin{aligned} E_p &= r_{pp} \cos P + r_{sp} e^{i\delta} \sin P \\ E_s &= r_{ps} \cos P + r_{ss} e^{i\delta} \sin P \end{aligned} \tag{9.23}$$

The analyser pass direction is at angle A to the $\mathbf{p}'$ direction, so the final field amplitude is

$$E_p \cos A + E_s \sin A = \cos P \cos A (r_{sp} + r_{ss} \tan A)(\rho_A + e^{i\delta} \tan P). \tag{9.24}$$

The last factor contains the modulation. Let $\rho_A = \rho_r + i\rho_i$; the detected intensity is proportional to

$$\rho_r^2 + \rho_i^2 + 2(\rho_r \cos\delta + \rho_i \sin\delta) \tan P + \tan^2 P. \tag{9.25}$$

The terms $\cos\delta$ and $\sin\delta$ are sinusoidal functions of sinusoidal argument. Being periodic they may be expanded in Fourier series, the coefficients of which are Bessel functions (Olver and Maximon 2010, formulae 10.12.1, 2):

$$\begin{aligned} \cos[M \sin \Omega t] &= J_0(M) + 2 \sum_{n=1}^{\infty} J_{2n}(M) \cos(2n\Omega t) \\ \sin[M \sin \Omega t] &= 2 \sum_{n=0}^{\infty} J_{2n+1}(M) \sin[(2n+1)\Omega t] \end{aligned} \tag{9.26}$$

Therefore the DC, Ω and 2Ω parts of the intensity are respectively proportional to

$$\begin{aligned} &\rho_r^2 + \rho_i^2 + 2\rho_r J_0(M) \tan P + \tan^2 P \\ &4\rho_i J_1(M) \tan P \sin \Omega t \\ &4\rho_r J_2(M) \tan P \cos 2\Omega t \end{aligned} \tag{9.27}$$

Thus polarization modulation ellipsometry, with the modulator placed between the polarizer and the sample, measures the real and imaginary parts of $\rho_A = (r_{pp} + r_{ps} \tan A)/(r_{sp} + r_{ss} \tan A)$, which reduces to $(r_p/r_s) \cot A$ in the isotropic case.

9.5 Polarizer–Sample–Modulator–Analyser

This is the final reflection ellipsometer configuration we shall examine. We can abbreviate the discussion, since it is similar to the preceding configuration. The $\mathbf{p}'$ and $\mathbf{s}$ components after reflection are

$$E_p' = r_{pp}\cos P + r_{sp}\sin P, \qquad E_s' = r_{ps}\cos P + r_{ss}\sin P. \tag{9.28}$$

After the modulator, aligned with its o and e directions along $\mathbf{p}'$ and $\mathbf{s}$, the $\mathbf{p}'$ and $\mathbf{s}$ components become E_p' and $E_s' e^{i\delta}$, with $\delta(t)$ given by (9.21) as before (we again neglect the very small difference in the t_o and t_e magnitudes). The field amplitude passing through the analyser is thus

$$E_p'\cos A + E_s' e^{i\delta}\sin A = \cos P\cos A(r_{ps} + r_{ss}\tan P)(\rho_P + e^{i\delta}\tan A). \tag{9.29}$$

With $\rho_P = \rho_r + i\rho_i$; the detected intensity is proportional to

$$\rho_r^2 + \rho_i^2 + 2(\rho_r\cos\delta + \rho_i\sin\delta)\tan A + \tan^2 A. \tag{9.30}$$

Hence the DC, Ω and 2Ω signals are respectively proportional to

$$\begin{aligned} &\rho_r^2 + \rho_i^2 + 2\rho_r J_0(M)\tan A + \tan^2 A \\ &4\rho_i J_1(M)\tan A\sin\Omega t \\ &4\rho_r J_2(M)\tan A\cos 2\Omega t \end{aligned} \tag{9.31}$$

Thus when the modulator is placed between the sample and the analyser, the polarization modulation ellipsometer measures the real and imaginary parts of $\rho_P = (r_{pp} + r_{sp}\tan P)/(r_{ps} + r_{ss}\tan P)$, which reduces to $(r_p/r_s)\cot P$ in the isotropic case.

9.6 Ellipsometric Measurements: The Principal Angle

We have seen that reflection ellipsometry with either compensator or modulator can determine the real and imaginary parts of ρ_P or of ρ_A. Thus measurements at least three polarizer or three analyser settings give three ratios of the reflection amplitudes, for example r_{pp}/r_{ss}, r_{ps}/r_{ss} and r_{sp}/r_{ss}. The individual amplitudes all carry the same arbitrary phase, which depends on conventions such as the choice of origin. The common phase factor cancels out in the ratios, which depend on the angle of incidence, the frequency of the light, the variation of the dielectric functions through the interface and (in the anisotropic case) on the orientation of the optic axis or axes of the reflector.

In the *isotropic* case ellipsometry can determine the real and imaginary parts of r_p/r_s. Often it is more useful to find the principal angle θ_P at which the real part of r_p/r_s is zero, and then measure the imaginary part of r_p/r_s. For thin films we saw in Chap. 3 that $\mathrm{Im}(r_p/r_s)$ is proportional (with known proportionality constant, see (3.46)) to the profile integral invariant

$$I_1 = \int_{-\infty}^{\infty} \mathrm{d}z \frac{(\varepsilon_1 - \varepsilon)(\varepsilon - \varepsilon_2)}{\varepsilon} \tag{9.32}$$

(see Sects. 7.2 and 7.3 for anisotropic generalizations). The integration in (9.32) is through a dielectric profile $\varepsilon(z)$, for example that of a vapour-liquid interface between vapour (ε_1) and liquid (ε_2). For interfaces thin compared to the wavelength of light being used, this integral invariant is *all* that can be determined. The principal angle at which $\mathrm{Re}(r_p/s) = 0$ is, for thin films, just the Brewster angle $\theta_B = \arctan(n_2/n_1)$.

We showed in Sect. 2.3 that, for an arbitrary dielectric profile, at least one principal angle exists. In general there are an odd number of principal angles. For a homogeneous nonabsorbing isotropic layer of dielectric constant ε and of thickness Δz, $\mathrm{Re}(r_p/r_s)$ is a ratio of two quadratics in $\cos 2q\Delta z$ with $q^2 = \varepsilon(\omega/c)^2 - K^2$, and equating the numerator to zero gives the principal angles (Lekner 2000). In general the difference between the principal angle θ_P and the Brewster angle θ_B is second order in the interface thickness, as shown in Sect. 3.5; (3.53) is of the form

$$\theta_P - \theta_B = \alpha\left(\frac{\omega}{c}\Delta z\right)^2 + O\left(\frac{\omega}{c}\Delta z\right)^4. \tag{9.33}$$

The proportionality coefficient α is given in terms of integral invariants in (3.53), and explicitly for the homogeneous layer in Sect. 3 of Lekner (2000).

9.7 Transmission Ellipsometry

We shall consider the normal-incidence transmission ellipsometry of a uniaxial layer resting on an isotropic substrate. The reflection and transmission amplitudes for an unsupported uniaxial layer were given in Sect. 8.7. With an isotropic substrate of index n_2, the reflection and transmission amplitudes at normal amplitude after the analyser, assuming unit amplitude field is incident, incidence may be found in Lekner (1994b). We shall give the electric field for five ellipsometer configurations. The transmitted amplitude depends on the polarizer and analyser angles P and A and on the compensator or modulator angle C, all measured in the same sense from the uniaxial layer $\mathbf{E}^o$ direction, which is perpendicular to the inward normal $\mathbf{n}$ to the crystal, being along the $\mathbf{n} \times \mathbf{c}$ direction ($\mathbf{c}$ is the optic axis, as in Chap. 8). Figure 9.2 shows the configuration and the angles P and A.

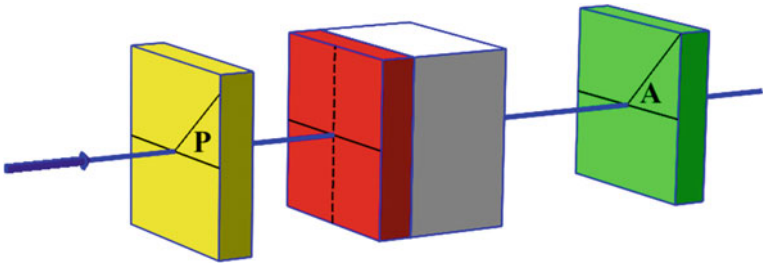


Fig. 9.2 The polarizer–sample (on substrate)–analyser transmission ellipsometry schematics. The angles P and A are measured from the $\mathbf{E}^o$ direction, shown as the *horizontal line* on the sample

9.7.1 Polarizer–Sample–Analyser

After passing through the analyser, the field amplitude is

$$t_o \cos P \cos A + t_e \sin P \sin A. \tag{9.34}$$

The intensity, given by the absolute square of (9.34), contains the polarizer and analyser angles, and $|t_o|$, $|t_e|$ and $\mathrm{Re}(t_o t_e^*)$. The unknowns are the absolute magnitudes of t_o and t_e, and their phase difference. Three intensity measurements at different polarizer and analyser settings, plus an intensity measurement with the sample absent, are the minimum required to determine the unknowns.

9.7.2 Polarizer–Compensator–Sample–Analyser

Figure 9.3 shows the configuration; the compensator ordinary and extraordinary field directions (in the latter case, actually the projection of $\mathbf{E}^e$ onto the crystal surface, see Sect. 8.6) are at right angles, with the ordinary field at angle C to the $\mathbf{E}^o$ direction of the crystal.

The field amplitude after the analyser is

$$t_o E_o \cos A + t_e E_e \sin A, \tag{9.35}$$

where, with t_o' and t_e' the complex compensator transmission amplitudes,

$$\begin{aligned} E_o &= t_o' \cos(P - C) \cos C - t_e' \sin(P - C) \sin C \\ E_e &= t_o' \cos(P - C) \sin C + t_e' \sin(P - C) \cos C \end{aligned} \tag{9.36}$$

In the null setting of the ellipsometer, the intensity is made zero (in practice minimized). The intensity will be zero when the real and imaginary parts of (9.35) are zero. From (9.36) we see that

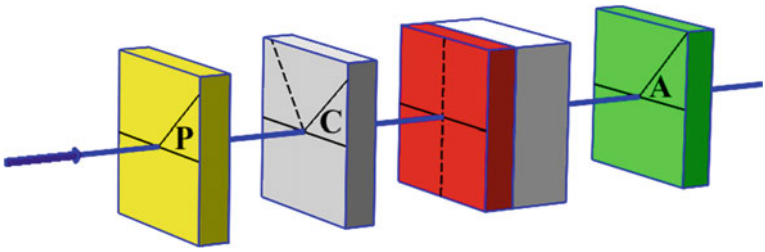


Fig. 9.3 The polarizer–compensator–sample–analyser ellipsometer arrangement. The angles P, C and A are measured from the sample $\mathbf{E}^o$ direction, shown as the *horizontal line* on the sample. The compensator ordinary and extraordinary field directions $\mathbf{o}'$ and $\mathbf{e}'$ are shown by *solid* and *dashed lines* on the compensator

$$\frac{E_e}{E_o} = \frac{t_o' \tan C + t_e' \tan(P - C)}{t_o' - t_e' \tan C \tan(P - C)} = \tan(C + D), \tag{9.37}$$

where D is a complex angle defined by

$$\tan D = \frac{t_e'}{t_o'} \tan(P - C). \tag{9.38}$$

Thus a null setting determines the complex ratio t_o/t_e in terms of the compensator transmission amplitude ratio and the angles P, C and A:

$$t_o/t_e = -\frac{E_e}{E_o} \tan A = -\tan(C + D) \tan A. \tag{9.39}$$

9.7.3 Polarizer–Sample–Compensator–Analyser

We resolve first along the sample **o** and **e** directions. The **o** and **e** components of the field passing through the sample are $t_o \cos P$ and $t_e \sin P$. We now resolve along the $\mathbf{o}'$ and $\mathbf{e}'$ directions of the compensator. The field components are

$$\begin{aligned} E_o' &= t_o \cos P \cos C + t_e \sin P \sin C \\ E_e' &= -t_o \cos P \sin C + t_e \sin P \cos C \end{aligned} \tag{9.40}$$

After passing through the compensator the $\mathbf{o}'$ and $\mathbf{e}'$ components are $t_o' E_o'$ and $t_e' E_e'$. After the analyser the field (along the analyser easy direction) is

$$t_o' E_o' \cos(A - C) + t_e' E_e' \sin(A - C). \tag{9.41}$$

We again define a complex angle related to the ratio t_e'/t_o':

$$\tan D' = t_e'/t_o' \tan(A - C). \tag{9.42}$$

A null setting is obtained when the real and imaginary parts of (9.41) are zero, that is when $E_o'/E_e' = -\tan D'$. Then the complex ratio t_o/t_e is given by

$$t_o/t_e = -\tan P \tan(C + D'). \tag{9.43}$$

9.7.4 *Transmission Ellipsometry with a Polarization Modulator*

The action of a modulator (a compensator with a sinusoidally varying phase) has been discussed in Sect. 9.4. The ratio of the e and o amplitudes passed through the modulator is

$$\frac{t_e'}{t_o'} \approx e^{i\delta}, \qquad \delta(t) \approx M \sin \Omega t. \tag{9.44}$$

M is the maximum phase shift, $\Omega/2\pi$ is the modulation frequency. The equalities in (9.44) are not exact, since the modulas of the ratio t_e'/t_o' is not unity but very close to it. Also the phase shift is not exactly sinusoidal (Archer et al. 1989). Lekner (1994a) assumes (9.44) to be true as equalities, for ease of analysis. Then, as in Sects. 9.4 and 9.5, measurement of the DC, Ω and 2Ω signals with lock-in amplifiers gives the real and imaginary parts of the relevant ellipsometric ratio, in this case the transmission amplitude ratio t_o/t_e of the sample. Details may be found in Sects. 6 and 7 of Lekner (1994a). Section 8 of that paper discusses the properties of the ratio t_o/t_e, and in particular its orbit in the complex plane as the thickness of the crystal plate increases.

9.8 Reflection and Transmission Ellipsometry of a Homogeneous Layer

We consider an isotropic homogeneous layer of thickness Δz and dielectric constant ε, bounded by homogeneous media of dielectric constants $\varepsilon_1 = n_1^2$ and $\varepsilon_2 = n_2^2$, and focus on the ellipsometric quantities $\rho = r_p/r_s$ and $\tau = t_p/t_s$. Isotropy in all media is assumed here; the optical properties of a uniaxial layer were outlined in Sect. 8.9,

but only the results for the optic axis lying in the plane of incidence were given explicitly there.

From Sect. 2.4, the ellipsometric ratios are given by

$$\rho = \frac{r_p}{r_s} = \frac{p_1 + p_2 Z}{1 + p_1 p_2 Z} \cdot \frac{1 + s_1 s_2 Z}{s_1 + s_2 Z}, \tag{9.45}$$

$$\tau = \frac{t_p}{t_s} = \frac{n_1}{n_2} \frac{(1 - p_1)(1 - p_2)(1 + s_1 s_2 Z)}{(1 + s_1)(1 + s_2)(1 + p_1 p_2 Z)}, \tag{9.46}$$

where s_1, p_1 and s_2, p_2 are the reflection amplitudes at the layer boundaries with media 1 and 2:

$$s_1 = \frac{q_1 - q}{q_1 + q}, \quad s_2 = \frac{q - q_2}{q + q_2}, \quad p_1 = \frac{Q - Q_1}{Q + Q_1}, \quad p_2 = \frac{Q_2 - Q}{Q_2 + Q}. \tag{9.47}$$

As usual, q_1, q and q_2 are the normal components of the wavevector medium 1, the layer, and medium 2; $Q_1 = q_1/\varepsilon_1$, $Q = q/\varepsilon$ and $Q_2 = q_2/\varepsilon_2$. In (9.45) and (9.46), $Z = e^{2iq\Delta z}$ and moves periodically on the unit circle in the complex plane when q is real (no absorption, and $\sin^2\theta_1 < \varepsilon/\varepsilon_1$). The period in Δz is π/q. The transmission ratio is a bilinear conformal transformation of the variable Z, and thus τ also moves on circles in the complex plane as the layer thickness increases (Fig. 9.4). The largest circle is at glancing incidence. More detail may be found in Dorf and Lekner (1987).

When $n_1 = n_2$ the reflection ellipsometric ratio simplifies to

$$\rho = \frac{r_p}{r_s} = \frac{p_1}{s_1} \frac{1 - s_1^2 Z}{1 - p_1^2 Z} \qquad (n_1 = n_2), \tag{9.48}$$

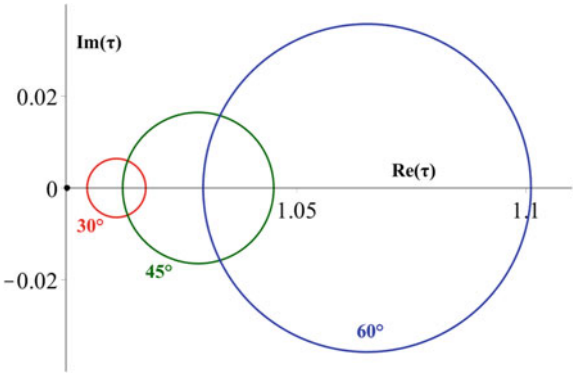


Fig. 9.4 Circular paths of $\tau = t_p/t_s$ in the complex plane at angles of incidence $30°, 45°$ and $60°$, drawn for $n_1 = 1$, $n = 4/3$ and $n = 3/2$, approximating air|water|glass. At normal incidence $t_p = t_s$ and the circle collapses to the point $(1, 0)$ shown in the figure

and thus also moves on circles in the complex plane as the layer thickness Δz increases (at fixed angle of incidence).

In the general case $n_1 \neq n_2$, the path of ρ in the complex plane as Δz increases becomes a closed quartic curve, obtained by eliminating Z from (9.45) by repeated use of $ZZ^* = 1$ (Lekner (1994c), Sect. 4). Lack of absorption is assumed, otherwise Z spirals in towards the origin as Δz increases. The quartic, in $x = \mathrm{Re}(\rho)$ and $y = \mathrm{Im}(\rho)$, is even in y. It passes through $y = 0$ at points $(\rho_\pm, 0)$ on the x-axis, where

$$\rho_+ = \rho(Z = +1) = \frac{q_1 q_2 - K^2}{q_1 q_2 + K^2} = \frac{p_0}{s_0} \tag{9.49}$$

(p_0 and s_0 are the reflection amplitudes of the bare substrate). The point $(\rho_+, 0)$ does not depend on the layer properties. The other intersection with $y = \mathrm{Im}(\rho) = 0$ is at

$$\rho_- = \rho(Z = -1) = \frac{(q^2 + q_1 q_2)(Q_1 Q_2 - Q^2)}{(q^2 - q_1 q_2)(Q_1 Q_2 + Q^2)}. \tag{9.50}$$

This goes to infinity when $q^2 = q_1 q_2$; together with $2q\Delta z$ being an odd multiple of π, $q^2 = q_1 q_2$ leads to $r_s = 0$, as seen in Sect. 2.4. The condition $q^2 = q_1 q_2$ is satisfied at angle of incidence θ_0 given by

$$\tan^2 \theta_0 = \frac{\varepsilon_1 \varepsilon_2 - \varepsilon^2}{(\varepsilon_1 - \varepsilon_2)^2} \tag{9.51}$$

(a real θ_0 is possible only if $\varepsilon^2 < \varepsilon_1 \varepsilon_2$).

The quartic curve in the real and imaginary parts of ρ is simultaneously a quintic equation in ε (Lekner 1994c, Sect. 2), reducing to a cubic in ε when $n_1 = n_2$.

A much simpler inversion of *combined reflection and transmission data* is possible (Azzam 1983; Lekner 1994a). Returning to (9.45) and (9.46), we define

$$P = p_1 p_2, \quad S = s_1 s_2, \quad t = \frac{n_2 \tau}{n_1}, \quad f = \frac{(1 - p_1)(1 - p_2)}{(1 + s_1)(1 + s_2)}. \tag{9.52}$$

Then $\rho = r_p / r_s$ and $t = n_2 t_p / n_1 t_s$ may be written as

$$\rho = \frac{p_1 + p_2 Z}{s_1 + s_2 Z} \frac{1 + SZ}{1 + PZ}, \qquad t = f \frac{1 + SZ}{1 + PZ}. \tag{9.53}$$

The inversion problem consists in the extraction of the unknowns ε and Δz from experimental values of ρ and τ (or t). The solution is as follows: the equality for t is a linear equation for $Z = e^{2iq\Delta z}$ with solution

$$Z = \frac{f - t}{tP - fS}. \tag{9.54}$$

This value for Z is now substituted into the expression for ρ, thus eliminating the unknown thickness Δz. The remaining equation contains the experimental complex quantities ρ and t (or τ), and the unknown dielectric constant ε, which also appears in q, since $q^2 = \varepsilon(\omega/c)^2 - K^2$. Substitution for $\varepsilon = (c/\omega)^2(q^2 + K^2)$ gives an equation with the unknown q, which after removal of common factors leaves an equation *linear* in q^2 (or equivalently, linear in ε). Thus a complete unambiguous solution for the dielectric constant of the layer is obtained. The details are given in Lekner (1994a), where extraction of the film thickness from (9.54), and stability of the solutions with respect to experimental error, are also discussed. In the simplest case of unsupported films ($n_1 = n_2$) the solution is

$$\varepsilon = \frac{\varepsilon_1 \sin^2\theta_1}{\cos^2\theta_1 - \rho/\tau}. \tag{9.55}$$

9.9 Reflection Ellipsometry of Uniaxial Crystals

Reflection ellipsometry is capable of determining the ordinary and extraordinary dielectric constants ε_o and ε_e of a uniaxial crystal, and the orientation of the optic axis, if at least one clean face of the crystal is exposed. The quantities measured are the ratios of the four reflection amplitudes r_{ss}, r_{sp}, r_{ps} and r_{pp}, and angles. Figure 9.5 defines the geometry.

If the crystal is nonabsorbing the reflection amplitudes are all real and the four real numbers ε_o, ε_e, χ and ϕ (refer to the caption of Fig. 9.5 for the definition of χ and ϕ) are found by measuring a minimum of four real quantities (three reflection amplitude ratios, and one angle). The same method applies to absorbing uniaxial crystals, but then ε_o, ε_e and the reflection amplitudes are all complex.

In Chap. 8 we characterized the optic axis **c** by its direction cosines α, β and γ:

$$\mathbf{c} = (\alpha, \beta, \gamma), \qquad \alpha^2 + \beta^2 + \gamma^2 = 1. \tag{9.56}$$

The equivalent notation of this section is

$$\mathbf{c} = (\sin\chi\cos\phi,\ \sin\chi\sin\phi,\ \cos\chi). \tag{9.57}$$

One first tests whether the crystal is indeed uniaxial, or whether it is isotropic or biaxial. For uniaxial crystals the r_{sp} and r_{ps} reflection amplitudes are known (Equations (8.30) and (8.40) of Chap. 8): they have the forms

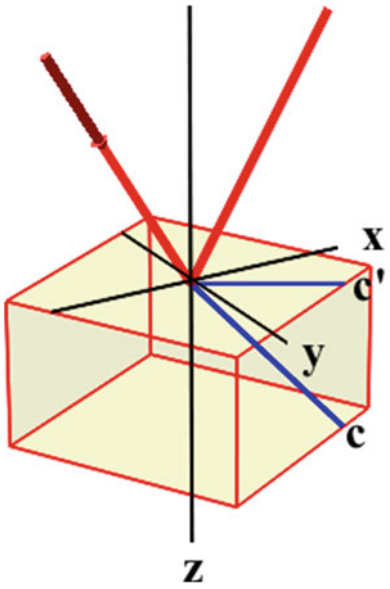


Fig. 9.5 Reflection by a uniaxial crystal: xy is the reflecting face, zx is the plane of incidence. The positive z axis coincides with the inward normal of the reflecting face. The optic axis $\mathbf{c}$ is at angle χ to the inward normal, and the plane containing $\mathbf{c}$ and the z axis cuts the xy plane on the *line* $\mathbf{c}'$, at angle ϕ to the x axis

$$r_{sp} = \beta(\alpha q_o + \gamma K)F, \qquad r_{ps} = \beta(\alpha q_o - \gamma K)F. \tag{9.58}$$

(F is the same for both amplitudes). To be specific, we suppose the ellipsometer is in one of the polarizer–sample–compensator–analyser or polarizer–sample–modulator–analyser configurations. Then the quantity measured is (see Sects. 9.3 and 9.5)

$$\rho_p = \frac{r_{pp} + r_{sp}\tan P}{r_{ps} + r_{ss}\tan P}, \tag{9.59}$$

where P is the angle between the polarizer easy and the incident $\mathbf{p}$ direction. Measurement of ρ_p at N different values of the polarizer angle P gives N linear homogeneous linear equations for the unknowns r_{ss}, r_{sp}, r_{ps} and r_{pp}. Only the ratios of the reflection amplitudes can be found from these experimental values. Three measurements are thus sufficient to determine the three independent ratios. For example, if we set $P = 0$, $\pi/4$ and $\pi/2$ (so $\tan P$ takes the values 0, 1 and ∞) and measure the corresponding complex numbers ρ_0, ρ_1 and ρ_∞,

$$\frac{r_{sp}}{r_{ss}} = \rho_\infty, \quad \frac{r_{ps}}{r_{ss}} = \frac{\rho_1 - \rho_\infty}{\rho_0 - \rho_1}, \quad \frac{r_{pp}}{r_{ss}} = \rho_0\frac{\rho_1 - \rho_\infty}{\rho_0 - \rho_1}. \tag{9.60}$$

Extra measurements provide a check on the accuracy of the data. For example, at $P = -\frac{\pi}{4}(\tan P = -1)$ the resulting measurement ρ_{-1} will be consistent with the previous measurements ρ_0, ρ_1 and ρ_∞ if

$$\rho_{-1} = \frac{\rho_0\rho_1 - 2\rho_0\rho_\infty + \rho_1\rho_\infty}{2\rho_1 - \rho_0 - \rho_\infty}. \tag{9.61}$$

In practice this consistency relation will not be satisfied exactly, because of experimental error. For $N > 3$ measurements there will be $N!/(N-3)!3!$ solution sets, which can be averaged to give the reflection amplitude ratios. Alternatively, one can measure repeatedly at three fixed values of the polarizer angle and average the ρ_P values obtained at each P, with occasional measurement at a fourth angle to provide a consistency check.

We return to the determination of the crystal parameters. The crystal is mounted on a support stage which can be rotated about the normal to the reflecting face (the z axis in Fig. 9.5). It is important that the stage can be adjusted so that the angle of incidence does not change on rotation. The ratios r_{sp}/r_{ss} and r_{ps}/r_{ss} are found first.

1. If these are zero at all values of the azimuthal angle ϕ, the crystal is isotropic, in which case the inversion of r_p/r_s to obtain the dielectric function ε is relatively straightforward (see Sect. 11.1 of Chap. 11), or the reflection is from a basal plane (one perpendicular to the optic axis, $\chi = 0$, so $\beta = 0$ in (9.58)). Lekner (1997) shows how the remaining unknowns ε_o and ε_e may then be found.
2. If the amplitudes r_{sp} and r_{ps} are not identically zero, then from (9.58) they will have two common zeros as the uniaxial crystal is rotated through 360° about the surface normal. These zeros occur at $\beta = 0$ ($\phi = 0$ or π). (If r_{sp} and r_{ps} are not zero together twice in a full rotation, the crystal is not uniaxial.) When the crystal is aligned so that r_{ps} and r_{sp} are both zero, the optic axis lies in the plane of incidence. The known reflection amplitudes for $\beta = 0$ then allow extraction of the remaining unknowns ε_o, ε_e and χ (or γ) (see Lekner 1997), in terms of reflection amplitude ratios at $\beta = 0$ and at $\alpha = 0 (\cos\phi = 0)$, and of

$$\frac{r_{sp} + r_{ps}}{r_{sp} - r_{ps}} = \frac{\alpha q_0}{\gamma K} = \tan\chi\cos\phi\frac{q_0}{K}. \tag{9.62}$$

Sensitivity to errors in measurement is explored in Lekner (1997) for reflection by calcite and selenium, the latter a strongly absorbing crystal. An alternative method for thin uniaxial samples has been proposed by Yang et al. (1995).

References

Archer O, Bigan E, Drévillon B (1989) Improvements of phase-modulated ellipsometry. Rev Sci Instr 60:65–77

Azzam RMA (1983) Ellipsometry of unsupported and embedded thin films. J de Physique C 10:67–70

Azzam RMA (1991) (ed) Selected papers on ellipsometry. SPIE Press

Azzam RMA, Bashara NM (1987) Ellipsometry and polarized light. North-Holland

Beaglehole D (1980) Ellipsometric study of the surface of simple liquids. Phys B 100:163–174
Dorf MC, Lekner J (1987) Reflection and transmission ellipsometry of a uniform layer. J Opt Soc Am A 4:2096–2100
Jasperson SN, Schnatterly SE (1969) An improved method for high-reflectivity ellipsometry based on a new polarization modulation technique. Rev Sci Instr 40:761–767
Lekner J (1993) Ellipsometry of anisotropic media. J Opt Soc Am A 10:1579–1581
Lekner J (1994a) Determination of complex refractive index and thickness of a homogeneous layer by combined reflection and transmission ellipsometry. J Opt Soc Am A 11:2156–2158
Lekner J (1994b) Normal incidence transmission ellipsometry of anisotropic layers. Pure Appl Opt 3:307–321
Lekner J (1994c) Inversion of reflection ellipsometric data. Appl Opt 33:5159–5165
Lekner J (1997) Reflection ellipsometry of uniaxial crystals. J Opt Soc Am A 14:1359–1362
Lekner J (2000) Multiple principal angles for a homogeneous layer. J Opt A: Pure Appl Opt 2:239–245
McGilp JF (1995) Optical characterization of semiconductor surfaces and interfaces. Prog Surf Sci 49:1–106
Olver FWJ, Maximon LC (2010) Bessel functions. In: Olver FWJ et al (eds) Ch10 of NIST handbook of mathematical functions. Cambridge University Press
Yang F, Bradberry GW, Sambles JR (1995) A method for the optical characterization of thin uniaxial samples. J Mod Opt 42:763–774, 1241–1252, 1447–1458

Chapter 10
Absorption

This chapter deals with the effect of absorption on reflection properties. The absorption, or dissipation of electromagnetic energy within the medium, can be due to conductivity (as in metals, and in the ionosphere). However, good insulators can also be absorbers at high frequencies, where the electromagnetic field energy is converted to heat via molecular or electronic excitations. The absorption is included in the Maxwell equation (1.2) by allowing the dielectric function ε to take complex values. In general, the curl of $\mathbf{B}$ is the sum of terms proportional to $\partial\mathbf{E}/\partial t$ and to the total current density. For non-magnetic media, and fields with the time variation $e^{-i\omega t}$, the form of (1.2) is retained, with the imaginary part of ε now proportional to the conductivity divided by the frequency (Born and Wolf 1970, Sect. 13.1). The simplest model for conducting media is that of an electron gas, with mean free time between collisions τ. This leads to the dielectric function (see for example Kittel 1966; Booker 1984; Budden 1985)

$$\varepsilon(\omega, z) = 1 - \frac{\omega_p^2}{\omega^2 + i\omega/\tau}, \tag{10.1}$$

where ω_p is the plasma frequency. In the ionosphere, for example, ε is a function of height z through the proportionality of ω_p^2 to the electron density, as well as through the dependence of τ on the electron, ion, and neutral species densities.

We will represent the real and imaginary parts of physical variables such as ε by the subscripts r and i:

$$\varepsilon = \varepsilon_r + i\varepsilon_i. \tag{10.2}$$

The real and imaginary parts of ε are directly related to the electronic properties of the material under study. Either $\varepsilon_r, \varepsilon_i$, or the real and imaginary parts of the square root of ε (the complex refractive index) can be used in writing the reflectivity formulae. We shall use both, with the refractive index notation being particularly convenient at normal incidence. The relationship between the two is found from

J. Lekner, *Theory of Reflection*, Springer Series on Atomic, Optical, and Plasma Physics 87, DOI 10.1007/978-3-319-23627-8_10

$$\varepsilon_r + i\varepsilon_i = (n_r + in_i)^2 \tag{10.3}$$

giving

$$\varepsilon_r = n_r^2 - n_i^2, \quad \varepsilon_i = 2n_r n_i. \tag{10.4}$$

The real and imaginary parts of ε are related in their frequency dependence by the Kramers-Krönig relations: $\varepsilon_r(\omega) - 1$ and $\varepsilon_i(\omega)$ are Hilbert transforms of each other, because the response of any system to an arbitrary signal must be causal (see, for example, Landau and Lifshitz 1960, Sect. 62).

10.1 Fresnel Reflection Formulae for an Absorbing Medium

For the s wave, with $\mathbf{E} = (0, E_y, 0)$ for propagation in the zx plane,

$$E_y(z, x, t) = \mathrm{e}^{i(Kx-\omega t)} E(z), \tag{10.5}$$

with

$$\frac{\mathrm{d}^2 E}{\mathrm{d}z^2} + q^2 E = 0, \quad q^2(z) = \varepsilon(z)\frac{\omega^2}{c^2} - K^2. \tag{10.6}$$

The separation of variables constant K is the component of the wavevector along the interface, and its invariance leads to Snell's Law:

$$\sqrt{\varepsilon_1}\frac{\omega}{c}\sin\theta_1 = K = \sqrt{\varepsilon_2}\frac{\omega}{c}\sin\theta_2. \tag{10.7}$$

Here we consider radiation incident from a non-absorbing medium (real ε_1) onto an absorbing medium (complex ε_2). Thus the angle of refraction is complex, and has a formal meaning only. The behaviour of the refracted wave is found from its waveform

$$E_y(z, x, t) = \mathrm{e}^{i(Kx + q_2 z - \omega t)}. \tag{10.8}$$

We write $\varepsilon_2 = \varepsilon_r + i\varepsilon_i$; the real and imaginary parts of q_2 are found from

$$q_2^2 = \varepsilon_2\frac{\omega^2}{c^2} - K^2 = \frac{\omega^2}{c^2}(\varepsilon_r + i\varepsilon_i - \varepsilon_1 \sin^2\theta_1). \tag{10.9}$$

Setting $q_2 = q_r + iq_i$, so that $q_2^2 = q_r^2 - q_i^2 + 2iq_r q_i$, we have

$$q_r^2 - q_i^2 = \frac{\omega^2}{c^2}\left(\varepsilon_r - \varepsilon_1 \sin^2\theta_1\right), \quad 2q_r q_i = \frac{\omega^2}{c^2}\varepsilon_i. \tag{10.10}$$

Thus, for $\varepsilon_i \neq 0$ and $\varepsilon_r > 0$,

$$\left(\frac{cq_r}{\omega}\right)^2 = \frac{1}{2}\left\{\varepsilon_r - \varepsilon_1 \sin^2\theta_1 + \left[\left(\varepsilon_r - \varepsilon_1 \sin^2\theta_1\right)^2 + \varepsilon_i^2\right]^{\frac{1}{2}}\right\}, \tag{10.11}$$

$$\frac{cq_i}{\omega} = \frac{\varepsilon_i/2}{cq_r/\omega}. \tag{10.12}$$

(When $\varepsilon_i = 0$ we have either q_i or $q_r = 0$, depending on whether $\theta_1 < \theta_c$ or $\theta_1 > \theta_c$.) The waveform in the absorbing medium is

$$E_y(z, x, t) = e^{-q_i z} e^{i(Kx + q_r z - \omega t)}. \tag{10.13}$$

Thus q_i must be non-negative, which implies that ε_i and n_i must be non-negative. Surfaces of constant amplitude are planes parallel to the interface ($z =$ constant), while surfaces of constant real phase are the planes $Kx + q_r z =$ constant. The normal to the surfaces of constant phase is at an angle θ_2' to the normal to the interface, where

$$\tan^2\theta_2' = \frac{K^2}{q_r^2} = \frac{2\varepsilon_1 \sin^2\theta_1}{\varepsilon_r - \varepsilon_1 \sin^2\theta_1 + \left[\left(\varepsilon_r - \varepsilon_1 \sin^2\theta_1\right)^2 + \varepsilon_i^2\right]^{\frac{1}{2}}}. \tag{10.14}$$

The real angle θ_2' and the angle θ_2 (in general complex) coincide only for real ε_2, or at normal incidence.

For a sharp boundary between media 1 and 2, represented by a step dielectric function at $z = 0$, the continuity of E and $\mathrm{d}E/\mathrm{d}z$ at the boundary (implied by the differential equation (10.6)) give the s wave reflection amplitude

$$r_s = \frac{q_1 - q_2}{q_1 + q_2} = \frac{q_1 - q_r - iq_i}{q_1 + q_r + iq_i}. \tag{10.15}$$

The s reflectivity is thus

$$R_s = \frac{(q_1 - q_r)^2 + q_i^2}{(q_1 + q_r)^2 + q_i^2}. \tag{10.16}$$

At normal incidence this reduces to

$$R_n = \frac{(n_1 - n_r)^2 + n_i^2}{(n_1 + n_r)^2 + n_i^2}, \tag{10.17}$$

since then $q_r = n_r\omega/c$ and $q_i = n_i\omega/c$.

The p wave, which has $\mathbf{B} = (0, B_y, 0)$, has the form

$$B_y(z, x, t) = \mathrm{e}^{i(Kx-\omega t)} B(z), \tag{10.18}$$

with the same separation of variables constant K as the s wave. The equation satisfied by B is

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{\varepsilon}\frac{\mathrm{d}B}{\mathrm{d}z}\right) + \left(\frac{\omega^2}{c^2} - \frac{K^2}{\varepsilon}\right)B = 0. \tag{10.19}$$

At a sharp boundary between two media B and $\varepsilon^{-1}\mathrm{d}B/\mathrm{d}z$ are continuous; the reflection amplitude is thus

$$-r_p = \frac{Q_1 - Q_2}{Q_1 + Q_2} = \frac{Q_1 - Q_r - iQ_i}{Q_1 + Q_r + iQ_i}, \tag{10.20}$$

where $Q_1 = q_1/\varepsilon_1$ and $Q_2 = q_2/\varepsilon_2$. The latter is the ratio of the two complex quantities $q_r + iq_i$ and $\varepsilon_r + i\varepsilon_i$, and thus has the real and imaginary parts

$$Q_r = \frac{\varepsilon_r q_r + \varepsilon_i q_i}{\varepsilon_r^2 + \varepsilon_i^2}, \quad Q_i = \frac{\varepsilon_r q_i - \varepsilon_i q_r}{\varepsilon_r^2 + \varepsilon_i^2}. \tag{10.21}$$

We note that $r_s \to -1$ and $r_p \to +1$ at grazing incidence, as in the case of non-absorbing media. The p reflectivity is

$$R_p = \frac{(Q_1 - Q_r)^2 + Q_i^2}{(Q_1 + Q_r)^2 + Q_i^2}, \tag{10.22}$$

and reduces to (10.17) at normal incidence, where the real and imaginary parts of Q take the values

$$Q_r = \frac{n_r\omega/c}{n_r^2 + n_i^2}, \quad Q_i = \frac{-n_i\omega/c}{n_r^2 + n_i^2}. \tag{10.23}$$

We shall see later in this section that R_p is never greater than R_s for a step profile. Both R_s and R_p differ from their zero-absorption Fresnel values by terms second order in ε_i.

In the absense of absorption, the p wave reflectivity is zero when $Q_1 = Q_2$ (at the Brewster angle $\theta_B = \arctan(n_2/n_1)$). The condition $Q_1 = Q_2$ cannot be satisfied for absorbing reflectors, since this would imply both $Q_1 = Q_r$ and $Q_i = 0$. The latter condition is satisfied at angle of incidence θ_1 such that

$$\sin^2\theta_1 = \frac{\varepsilon_r^2 + \varepsilon_i^2}{2\varepsilon_1\varepsilon_r}, \tag{10.24}$$

and is thus possible if $\varepsilon_r > 0$ and $\varepsilon_r^2 + \varepsilon_i^2 < 2\varepsilon_1\varepsilon_r$. But when (10.24) holds, the condition $Q_1 = Q_r$ could be satisfied only if $(\varepsilon_1 - \varepsilon_r)^2 + \varepsilon_i^2 = 0$. Thus, when there is absorption, zero reflection at a single sharp boundary is not possible. When a dielectric layer is placed over an absorbing medium, zero reflectivity is possible, for both polarizations (at different angles), as we shall see in Sect. 10.3.

The s and p step profile reflectivities are shown in Fig. 10.1, for a metal (Al) and a semiconductor (Si) at the visible He–Ne laser wavelength, $\lambda_0 = 0.633\,\mu$m. The minima of R_p are determined by a cubic equation in $\sin^2\theta_1$, given in (11.69) at the end of the next chapter. For the Al and Si reflectances shown in Fig. 10.1, they occur at about 83° and 76°.

We note the high metallic reflectivities, which are due to wavenumber mismatch: q_1 is real and $q_2 = q_r + iq_i$ has a large imaginary part. An example of this real|imaginary type of mismatch was seen in total internal reflection in dielectric materials, where for $\theta_1 > \theta_c = \arcsin(n_2/n_1)$ the wavenumber normal component q_2 is pure imaginary, giving total reflection for both polarizations. We see from (10.11), (10.12) and (10.16) that when $\varepsilon_i \neq 0$ total reflection is not possible for the s wave (except at grazing incidence, when $q_1 \to 0$). The same result follows for the p wave from (10.11), (10.12), (10.21) and (10.22).

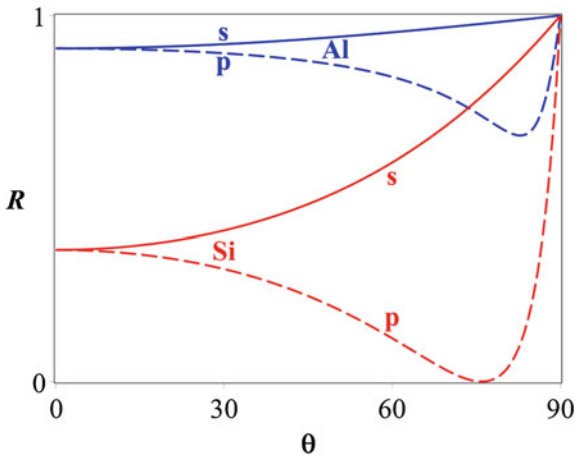


Fig. 10.1 Reflectivity as a function of the angle of incidence, for the s and p polarizations at 633 nm. The refractive indices are Al: $1.566 + 7.938i$, Si: $4.0 + 0.12i$. The corresponding dielectric functions are Al: $-60.56 + 24.86i$, Si: $16.0 + 0.96i$ (The Al values are for bulk metal; vapour-deposited values are different: see Allen 1976)

The ellipsometric quantity r_p/r_s is obtained from (10.15) and (10.20). It has the real and imaginary parts

$$\mathrm{Re}\left(\frac{r_p}{r_s}\right) = -\frac{\left\{\left(Q_1^2 - Q_r^2 - Q_i^2\right)\left(q_1^2 - q_r^2 - q_i^2\right) + 4Q_1Q_iq_1q_i\right\}}{\left[\left(Q_1 + Q_r\right)^2 + Q_i^2\right]\left[\left(q_1 - q_r\right)^2 + q_i^2\right]}, \tag{10.25}$$

$$\mathrm{Im}\left(\frac{r_p}{r_s}\right) = 2\frac{Q_1Q_i\left(q_1^2 - q_r^2 - q_i^2\right) - q_1q_i\left(Q_1^2 - Q_r^2 - Q_i^2\right)}{\left[\left(Q_1 + Q_r\right)^2 + Q_i^2\right]\left[\left(q_1 - q_r\right)^2 + q_i^2\right]}. \tag{10.26}$$

The computation of these quantities is simplified by the identity

$$Q_r^2 + Q_i^2 = \frac{q_r^2 + q_i^2}{\varepsilon_r^2 + \varepsilon_i^2}. \tag{10.27}$$

Equivalent and somewhat simpler formulae (in terms of q_1, q_r, q_i and K) are given in Sect. 11.1.

The trajectories of r_p/r_s in the complex plane for variable angle of incidence are shown in Fig. 10.2 for Al and Si at 633 nm. There is rapid variation in the real and imaginary parts of r_p/r_s at large angles of incidence, for Al particularly: the paths cross the real axis at the principal angles of about 83°(Al) and 76°(Si).

The trajectory of r_p/r_s always lies within the upper half of the unit circle for an arbitrary absorbing medium with a sharply defined surface. To see this, it is convenient to define a complex angle of refraction, $\theta_2 = \theta_r + i\theta_i$, via

$$q_2 = \varepsilon_2^{1/2}\left(\frac{\omega}{c}\right)\cos\theta_2 \tag{10.28}$$

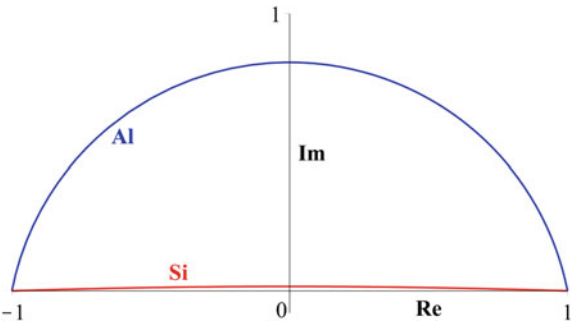


Fig. 10.2 The ellipsometric ratio r_p/r_s, in the complex plane; trajectories for Al and Si are shown. The refractive indices are for 633 nm, as in Fig. 10.1. For a perfect dielectric (no absorption) the trajectory is the real axis from 1 (at 0°) to −1 (at 90°)

(this definition is consistent with (10.6) and (10.7)). Then (10.15) and (10.20) may be written in the Fresnel forms (1.14) and (1.32),

$$r_s = \frac{\sin(\theta_2 - \theta_1)}{\sin(\theta_2 + \theta_1)}, \quad r_p = \frac{\tan(\theta_2 - \theta_1)}{\tan(\theta_2 + \theta_1)}, \tag{10.29}$$

and the ellipsometric ratio r_p/r_s as $\cos(\theta_2 + \theta_1)/\cos(\theta_2 - \theta_1)$, or

$$\frac{r_p}{r_s} = \frac{\cos(\theta_1 + \theta_r)\cosh\theta_i - i\sin(\theta_1 + \theta_r)\sinh\theta_i}{\cos(\theta_1 - \theta_r)\cosh\theta_i + i\sin(\theta_1 - \theta_r)\sinh\theta_i}. \tag{10.30}$$

The fact that $R_p \leq R_s$ follows from $\sinh^2\theta_i < \cosh^2\theta_i$. The sign of $\mathrm{Im}(r_p/r_s)$ is opposite to that of θ_i. From (10.28) we have

$$\left(\frac{c}{\omega}\right)(q_r + iq_i) = (n_r + in_i)(\cos\theta_r \cosh\theta_i - i\sin\theta_r \sinh\theta_i). \tag{10.31}$$

The real and imaginary parts of (10.31) give

$$\cos\theta_r \cosh\theta_i = \frac{c}{\omega}\frac{(n_r q_r + n_i q_i)}{n_r^2 + n_i^2}, \tag{10.32}$$

$$\sin\theta_r \sinh\theta_i = \frac{c}{\omega}\frac{(n_i q_r - n_r q_i)}{n_r^2 + n_i^2}. \tag{10.33}$$

These relations may in turn be used to find θ_r and θ_i as a function of the angle of incidence, θ_1. Here we are interested mainly in the sign of θ_i, which is that of $n_i q_r - n_r q_i$. We noted below (10.13) that q_r and q_i are non-negative. Thus, from (10.10), ε_i is also non-negative, and so is n_i (both n_r and n_i are ≥ 0, since $q_r = n_r\omega/c$, $q_i = n_i\omega/c$ at normal incidence). It thus follows from (10.11), (10.12) and (10.33) that θ_i is never positive, so that r_p/r_s always stays in the upper half of the unit circle.

For non-absorbing dielectrics R_p is zero at the Brewster angle. In the presence of absorption the reflectivity ratio R_p/R_s has a minimum at what is known as the pseudo or second Brewster angle. The extraction of the optical constants n_r and n_i (or ε_r and ε_i) from measurements of this angle and of the minimum reflectivity ratio is discussed by Potter (1969). The equations determining the principal angle of an absorber, where $\mathrm{Re}\left(\frac{r_p}{r_s}\right) = 0$, and also the angle where R_p is minimum, are given in Sect. 11.7.

10.2 General Results for Reflection by Absorbing Media

In Sect. 2.1 we derived the conservation law

$$q_1\left(1-|r|^2\right)=q_2|t|^2, \tag{10.34}$$

valid for s and p electromagnetic waves and quantum particle waves in the absence of absorption. This relation represents conservation of energy in the electromagnetic case, and conservation of the probability density current in the particle case. In the presence of absorption the conservation law is no longer valid, since energy or particles are removed by the absorbing medium. This was noted in Sect. 2.1; it is mathematically more explicit in the approach of Sect. 6.2, where it is clear that the derivation of (10.34) depends on the reality of $q^2(z)$.

The quantity $T_{12}=(q_2/q_1)|t_{12}|^2$ is called the transmittance (see the discussion following (2.8); equivalently one may take the ratio of $\mathrm{Im}(\psi^*\mathrm{d}\psi/\mathrm{d}z)$ for $\psi_2=t_{12}\mathrm{e}^{iq_2z}$ and $\psi_1=\mathrm{e}^{iq_1z}$, as in (2.9)). For an arbitrary inhomogeneous and absorbing layer between the nonabsorbing media 1 and 2, we showed in Sect. 2.1 that the reciprocity relation $q_2t_{12}=q_1t_{21}$ holds. Thus the transmittances for propagation in either direction through an absorbing layer are equal:

$$T_{12}=\frac{q_2}{q_1}|t_{12}|^2=\frac{q_1}{q_2}|t_{21}|=T_{21}. \tag{10.35}$$

The corresponding result for reflectivities, $R_{12}=R_{21}$, holds only in the absence of absorption (Sect. 2.1, (10.18)).

The result (10.34) may be written as $1-R=T$. For an absorbing layer between nonabsorbing media, the ratio $(1-R)/T$ is greater than unity, since the conservation law $1=R+T$ is replaced by $1=R+T+A$, where A represents absorption. Abelès (1950) has shown that if an arbitrary nonabsorbing layer is inserted in the front of the absorbing layer, causing the reflectance to change to R' and the transmittance to T', the ratio of $1-R$ to T is unaltered:

$$\frac{1-R}{T}=\frac{1-R'}{T'}. \tag{10.36}$$

This result is proved by matrix methods in Sect. 12.5.

The general formulae for r_s, t_s, r_p and t_p given in Sect. 2.2 remain true in the presence of absorption, with the solutions F, G and C, D now complex. Thus it remains true that $r_s\to-1$ and $r_p\to1$ at grazing incidence. Since $r_p=r_s$ at normal incidence, it follows that the trajectory of r_p/r_s still starts at $+1$ and ends at -1 in the complex plane, and consequently there always exists a *principal angle* (ellipsometric Brewster angle) where $\mathrm{Re}(r_p/r_s)=0$. In general there can be an odd number of such angles, as noted in Sect. 2.3.

10.3 Dielectric Layer on an Absorbing Substrate

The reflection from a homogeneous dielectric layer on a transparent substrate was discussed in Sect. 2.4. When the substrate is absorbing (typically it is metallic), the formulae for the reflection amplitudes derived in Sect. 2.4 remain valid, with $q_2 = q_r + iq_i$ and $Q_2 = Q_r + iQ_i$. The fact that q_2 and Q_2 are complex changes the reflectivity properties markedly. Of particular interest is the design of reflection polarizers, in which the reflectance of one of the components of polarization is extinguished by interference effects, while that of the other is not. See, for example, Ruiz-Urbieta and Sparrow (1972), Bennett and Bennett (1978), and Azzam (1985).

Consider a non-absorbing homogeneous film, of thickness Δz, located between z_1 and $z_1 + \Delta z$. The light is incident at angle θ_1 from medium l, of dielectric constant ε_1. The film has dielectric constant ε, and the refracted ray within it makes an angle θ to the normal. Snell's Law (the invariance of K^2) gives $\varepsilon_1 \sin^2\theta_1 = \varepsilon \sin^2\theta$. The substrate has dielectric constant $\varepsilon_2 = \varepsilon_r + i\varepsilon_i$. The normal components of the wavevector in the three media are q_1, q, and $q_2 = q_r + iq_i$, where the real and imaginary parts of q_2 are given by (10.11) and (10.12). The s wave reflection amplitude for this system is given by (2.58):

$$r_s = \mathrm{e}^{2iq_1z_1}\,\frac{r + r'\mathrm{e}^{2iq\Delta z}}{1 + rr'\mathrm{e}^{2iq\Delta z}}, \tag{10.37}$$

where

$$r = \frac{q_1 - q}{q_1 + q}, \quad r' = \frac{q - q_2}{q + q_2}, \tag{10.38}$$

are the reflection amplitudes (without phase factors associated with location) for the ambient-film and film-substrate interfaces. From (10.37), r_s will be zero when $r' = -r\mathrm{e}^{-2iq\Delta z}$; on equating the real and imaginary part we find

$$\frac{q_r^2 + q_i^2 - q^2}{(q + q_r)^2 + q_i^2} = r\cos 2q\Delta z, \quad \frac{-2qq_i}{(q + q_r)^2 + q_i^2} = r\sin 2q\Delta z \tag{10.39}$$

The angle of incidence at which zero reflection occurs (and the corresponding wavevector components to be inserted into (10.39) to determine the appropriate values of Δz) is found from $|r'|^2 = r^2$, which leads to

$$\left(q^2 - q_1q_r\right)(q_r - q_1) = q_1q_i^2. \tag{10.40}$$

This equation reduces to a quadratic in $\cos^2\theta_1$ (the coefficients are given in Sect. 10.9), or may solved numerically. For metallic substrates the solution lies near grazing incidence; for example, for a layer of Al_2O_3 of refractive index 1.6 on aluminium (with the optical parameters used in Figs. 10.1 and 10.2), zero reflection

for the s wave occurs at 87.95°. Since the reflectivity is always unity at grazing incidence, its variation with angle of incidence is necessarily rapid between the polarizing angle and 90°.

Kitajima et al. (1984) give numerical and experimental examples of the reflectivity near the extinction point as a function of film thickness. The dependence is strong, so observation of the oblique incidence reflectance during film deposition is a sensitive thickness monitor.

For the p wave the reflection amplitude is given by (2.70):

$$-r_p = \mathrm{e}^{2iq_1z_1}\,\frac{r + r'\mathrm{e}^{2iq\Delta z}}{1 + rr'\mathrm{e}^{2iq\Delta z}}, \tag{10.41}$$

where now

$$r = \frac{Q_1 - Q}{Q_1 + Q},\ r' = \frac{Q - Q_2}{Q + Q_2}\ \left(Q_1 = \frac{q_1}{\varepsilon_1},\ Q = \frac{q}{\varepsilon},\ Q_2 = \frac{q_2}{\varepsilon_2}\right). \tag{10.42}$$

The real and imaginary parts of Q_2 are given by (10.21). The condition for zero p reflectivity is $r' = -r\mathrm{e}^{-2iq\Delta z}$, which is equivalent to the equations derived for the s wave with Q_1, Q, Q_2 replacing q_1, q, q_2 except in the oscillatory functions of $2q\Delta z$. The equation analogous to (10.40),

$$\left(Q^2 - Q_1Q_r\right)(Q_r - Q_1) = Q_1Q_i^2, \tag{10.43}$$

reduces to a sextic in $C = \cos^2\theta_1$, and again has a solution close to grazing incidence: for the Al_2O_3 on Al case, extinction of the p wave occurs at $\theta_1 \approx 88.68°$. Azzam (1985) has used the fact that zero reflection occurs near 90° to obtain approximate but explicit solutions of (10.40) and (10.43). If one keeps just the constant and linear terms in the quadratic and sextic for $\cos^2\theta_1$, the accuracy in the examples just quoted is about 2 % in the glancing angle.

10.4 Absorbing Film on a Non-absorbing Substrate

The derivation of the reflection and transmission amplitudes for a homogeneous layer between two homogeneous media given in Sect. 2.4 remains valid when the layer, the substrate, or both, are absorbing. Here we examine the case when the layer is absorbing, and the substrate is not. An example is a metallic film on glass. The s wave results may be obtained from (2.52) and (2.53) or from (2.58) and (2.59). The latter are more convenient when ε (the dielectric constant of the layer) is complex. We set $\varepsilon = \varepsilon_r + i\varepsilon_i,\ q^2 = (q_r + iq_i)^2 = \varepsilon\omega^2/c^2 - K^2$, to obtain

$$\left(\frac{cq_r}{\omega}\right)^2 = \frac{1}{2}\left\{\varepsilon_r - \varepsilon_1 \sin^2\theta_1 + \left[(\varepsilon_r - \varepsilon_1 \sin^2\theta_1)^2 + \varepsilon_i^2\right]^{\frac{1}{2}}\right\}, \tag{10.44}$$

$$\frac{cq_i}{\omega} = \frac{\varepsilon_i/2}{cq_r/\omega}. \tag{10.45}$$

(These formulae are the same as (10.11) and (10.12), but here $\varepsilon_r, \varepsilon_i$ and q_r, q_i refer to a film rather than to a bulk medium.) The reflection amplitude is given formally by (10.37) and (10.38), with $q = q_r + iq_i$ now complex. To simplify the analysis we write

$$r = \rho e^{i\delta}, \quad r' e^{2iq\Delta z} = \rho' e^{i\delta'}, \tag{10.46}$$

with ρ' including the exponential decay factor $e^{-2q_i\Delta z}$, and δ' the $2q_r\Delta z$ phase increment. Then r_s takes the simple form

$$r_s = e^{2iq_1 z_1}\frac{\rho e^{i\delta} + \rho' e^{i\delta'}}{1 + \rho\rho' e^{i(\delta+\delta')}}, \tag{10.47}$$

and gives reflectivity

$$|r_s|^2 = \frac{\rho^2 + 2\rho\rho'\cos(\delta - \delta') + (\rho')^2}{1 + 2\rho\rho'\cos(\delta + \delta') + (\rho\rho')^2}. \tag{10.48}$$

The transmission amplitude is found from (2.59):

$$t_s = e^{i(q_1 z_1 - q_2 z_2)}\frac{(1 + \rho e^{i\delta})(e^{iq\Delta z} + e^{-iq\Delta z}\rho' e^{i\delta'})}{1 + \rho\rho' e^{i(\delta+\delta')}} \tag{10.49}$$

We set $f = e^{-q_i\Delta z}$, $\phi = q_r\Delta z$ so

$$e^{iq\Delta z} = e^{-q_i\Delta z}e^{iq_r\Delta z} = f e^{i\phi}, \tag{10.50}$$

and obtain

$$|t_s|^2 = \frac{(1 + 2\rho\cos\delta + \rho^2)(f^2 + 2\rho'\cos(2\phi - \delta) + (\rho'/f)^2)}{1 + 2\rho\rho'\cos(\delta + \delta') + (\rho\rho')^2}. \tag{10.51}$$

When the film is "thick", by which we mean here that the absorption within it is large, with $f = e^{-q_i\Delta z} \ll 1$, the reflection properties reduce to those of Sect. 10.1 for waves incident on a semi-infinite absorbing medium. The transmitted flux is then proportional to $e^{-2q_i z}$. In such a film the effect of interference of multiply reflected light (see Fig. 2.5) is negligible because of the decrease in the amplitude due to absorption.

In the absence of absorption the above formulae reduce to those of Sect. 2.4, with $|r_s|^2$ and $|t_s|^2$ periodic functions of Δz of period $\pi/q = \lambda/2$, λ being the wavelength within the film associated with motion in the z direction (perpendicular to the surface). An example was shown in Fig. 2.6. Absorption damps the periodicity, and strong absorption removes the oscillations altogether.

The p wave formulae may be obtained from (2.70) and (2.71), with q and Q now complex and given by (10.44, 10.45) and (10.21). The results are completely analogous to the s wave case, with the exception of the $\sqrt{(\varepsilon_2/\varepsilon_1)}$ factor multiplying t_p in (2.71).

Figure 10.3 shows the normal incidence reflectances and transmittances for films of Al and Si on glass. For both materials the transmittance tends to zero as the film thickness increases, and the reflectance tends to $\rho^2 = [(q_1 - q_r)^2 + q_i^2]/[(q_1 + q_r)^2 + q_i^2]$, with respective values of about 0.91(Al) and 0.36(Si). The approach to the thick film limit is rapid and monotonic for Al, which has a large imaginary part of the refractive index. With its small imaginary part of refractive index, the Si reflectance and transmittance oscillate (because of interference of forward and backward propagating waves within the film) until the exponential decay factor $e^{-2q_i\Delta z}$ dominates. For the examples shown, the limiting absorbance values $A = 1 - R - T \to 1 - R \to 1 - \rho^2$ are about 0.09(Al) and 0.64(Si). For thick films the more weakly absorbing material absorbs more energy, because more light penetrates into the film, to be absorbed at depth.

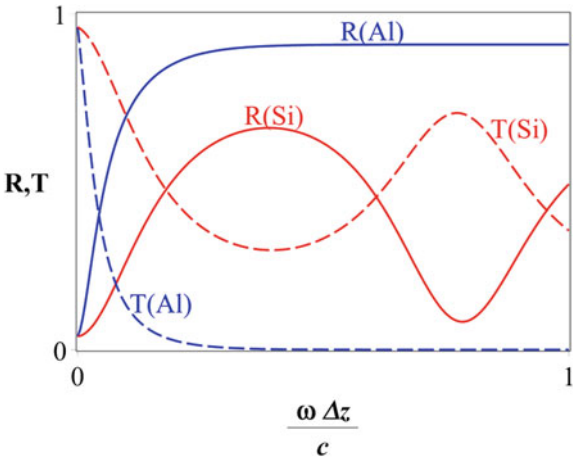


Fig. 10.3 Reflectance $R = |r|^2$ and transmittance $T = (q_2/q_1)|t|^2$ for light of wavelength 633 nm incident normally on films of Si($n = 4.0 + 0.12i$) and vapour-deposited Al($n = 1.212 + 6.924i$) on glass ($n = 1.5$). At normal incidence $q_i = n_i\omega/c$; for large $q_i\Delta z$ the transmitted intensity varies as $e^{-2q_i\Delta z}$. When $\Delta z = c/\omega = \lambda_0/2\pi$, this factor is approximately 10^{-6} for Al; the thickness is then about 0.1 μm

10.5 Thin Inhomogeneous Absorbing Films

So far we have considered homogeneous media, and homogeneous layers of arbitrary thickness. We now specialize to thin films or interfaces ("thin" now meaning that the film thickness times ω/c is small), which however may have arbitrary depth dependence in the real and imaginary parts of the dielectric function, $\varepsilon_r(z)$ and $\varepsilon_i(z)$.

The presence of absorption within the interface and/or substrate has major effects on the reflection and ellipsometric properties. We saw in Chap. 3 that for nonabsorbing media the reflectivities are unchanged to first order in the interface thickness/wavelength expansion: the interfacial profile characteristics appear only in the second order. For absorbing media, the wavenumber components q and $Q = q/\varepsilon$ are complex. The s and p reflection amplitudes to first order in the interface thickness are given by (3.23) and (3.44):

$$r_s = r_{s0} + \frac{2iq_1\omega^2/c^2}{(q_1 + q_2)^2}\lambda_1 + \cdots, \tag{10.52}$$

$$r_p = r_{p0} - \frac{2iQ_1}{(Q_1 + Q_2)^2}\left\{\frac{K^2\Lambda_1}{\varepsilon_1\varepsilon_2} - Q_2^2\lambda_1\right\} + \cdots. \tag{10.53}$$

The integrals λ_1 and Λ_1 are given by

$$\lambda_1 = \int_{-\infty}^{\infty} \mathrm{d}z(\varepsilon - \varepsilon_0), \quad \Lambda_1 = \varepsilon_1\varepsilon_2 \int_{-\infty}^{\infty} \mathrm{d}z\left(\frac{1}{\varepsilon_0} - \frac{1}{\varepsilon}\right), \tag{10.54}$$

where the step function $\varepsilon_0(z)$, which takes the values ε_1 for $z<0$ and ε_2 for $z > 0$, may now be complex, since ε_2 may be complex.

We will consider in detail the non-absorbing substrate case (ε_2 real). The integrals λ_1 and Λ_1 are still complex, since the interface is absorbing, with complex $\varepsilon(z)$. The s and p reflectivities thus contain first order correction terms proportional to the imaginary parts of λ_1 and Λ_1:

$$R_s = R_{s0} - \frac{4q_1(q_1 - q_2)\omega^2/c^2}{(q_1 + q_2)^3}\mathrm{Im}\,\lambda_1 + \cdots, \tag{10.55}$$

$$R_p = R_{p0} - \frac{4Q_1(Q_1 - Q_2)}{(Q_1 + Q_2)^3}\left\{\frac{K^2}{\varepsilon_1\varepsilon_2}\mathrm{Im}\,\Lambda_1 - Q_2^2\mathrm{Im}\,\lambda_1\right\} + \cdots \tag{10.56}$$

(these formulae and the following discussion apply to the nonabsorbing substrate case only). The step function ε_0 is real when ε_2 is real, so

$$\mathrm{Im}\,\lambda_1 = \int_{-\infty}^{\infty} \mathrm{d}z\; \varepsilon_i(z), \quad \frac{1}{\varepsilon_1 \varepsilon_2}\mathrm{Im}\,\Lambda_1 = \int_{-\infty}^{\infty} \mathrm{d}z \frac{\varepsilon_i}{\varepsilon_r^2 + \varepsilon_i^2}. \tag{10.57}$$

The reflectivity corrections are proportional to integrals over the imaginary part of ε through the absorbing region, as may be expected. At normal incidence both (10.55) and (10.56) reduce to

$$R_n = R_{n0} - \frac{4n_1(n_1 - n_2)}{(n_1 + n_2)^3}\frac{\omega}{c}\mathrm{Im}\,\lambda_1 + \cdots, \tag{10.58}$$

where $n_1 = \varepsilon_1^{1/2}$ and $n_2 = \varepsilon_2^{1/2}$ are the real refractive indices of the media bounding the inhomogeneous region.

For passive media the absorption term $\mathrm{Im}\,\lambda_1$ is non-negative, and so (to first order in the film thickness) absorption in the film increases the system reflectivity at normal incidence if $n_1 < n_2$, and decreases it if $n_1 > n_2$. This statement remains true at all angles of incidence for the s wave, but not for the p wave, for which the correction term in (10.56) changes sign at the Brewster angle $\theta_B = \arctan(n_2/n_1)$ (at which $Q_1 = Q_2$), and also when

$$\sin^2\theta_1 = \left[\frac{\varepsilon_1}{\varepsilon_2} + \frac{\mathrm{Im}\,\Lambda_1}{\mathrm{Im}\,\lambda_1}\right]^{-1}.$$

At these two angles for the p wave there is no contribution to the reflectivity in first order in the film thickness. The first order term is also absent for absorbing films between like media ($\varepsilon_1 = \varepsilon_2$), for both polarizations and at all angles of incidence.

In all cases there is however a first order effect in the transmission: for the s wave we have from (2.15) that

$$t_s = t_{s0} + \frac{2iq_1\omega^2/c^2\lambda_1}{(q_1 + q_2)^2} + \cdots, \tag{10.59}$$

which gives the transmittance

$$T_s = \frac{q_2}{q_1}|t_s|^2 = \frac{4q_1q_2}{(q_1 + q_2)^2}\left\{1 - \frac{2\omega^2/c^2\mathrm{Im}\,\lambda_1}{q_1 + q_2} + \cdots\right\}. \tag{10.60}$$

The p wave result is a little more complicated. To obtain an identity similar to (2.15) we start with two p wave equations, with dielectric functions ε and $\tilde{\varepsilon}$, and incident from media 1 and 2, respectively:

$$\frac{\mathrm{d}C_{12}}{\mathrm{d}z} + \left(\frac{\omega^2}{c^2} - \frac{K^2}{\varepsilon}\right)B_{12} = 0, \quad e^{iq_1z} - r_{12}e^{iq_1z} \leftarrow B_{12} \rightarrow \tau_{12}e^{iq_2z}, \tag{10.61}$$

$$\frac{\mathrm{d}\tilde{C}_{21}}{\mathrm{d}z} + \left(\frac{\omega^2}{c^2} - \frac{K^2}{\tilde{\varepsilon}}\right)\tilde{B}_{21} = 0, \quad \tilde{\tau}_{21}e^{-iq_1 z} \leftarrow \tilde{B}_{21} \rightarrow e^{-iq_2 z} - \tilde{r}_{21}e^{iq_2 z}. \tag{10.62}$$

Here C stands for $\varepsilon^{-1}\mathrm{d}B/\mathrm{d}z$, τ_{12} for $(\varepsilon_2/\varepsilon_1)^{1/2}t_{12}$. On multiplying (10.61) by $\tilde{B}_{21}$ and (10.62) by B_{12}, subtracting, and integrating from $-\infty$ to $+\infty$, we get the comparison identity

$$2i(Q_1\tilde{\tau}_{21} - Q_2\tau_{12}) = \int_{-\infty}^{\infty} \mathrm{d}z\left\{K^2\left(\frac{1}{\tilde{\varepsilon}} - \frac{1}{\varepsilon}\right)B_{12}\tilde{B}_{21} + (\varepsilon - \tilde{\varepsilon})C_{12}\tilde{C}_{21}\right\}. \tag{10.63}$$

This holds for any pair of profiles ε and $\tilde{\varepsilon}$ (with the same limiting values ε_1 and ε_2), and thus also for $\tilde{\varepsilon} = \varepsilon$, when the right-hand side is zero. Thus $Q_1\tau_{21} = Q_2\tau_{12}$ (equivalent to $q_1 t_{21} = q_2 t_{12}$, (2.14)), and (10.63) may be rewritten as

$$\tau_{12} = \tilde{\tau}_{12} - \frac{1}{2iQ_2}\int_{-\infty}^{\infty} \mathrm{d}z\left\{K^2\left(\frac{1}{\tilde{\varepsilon}} - \frac{1}{\varepsilon}\right)B_{12}\tilde{B}_{21} + (\varepsilon - \tilde{\varepsilon})C_{12}\tilde{C}_{21}\right\}. \tag{10.64}$$

We now set $\tilde{\varepsilon} = \varepsilon_0$, the step function profile. To lowest order in the film thickness, it suffices to replace B and C by the values taken by B_0 and C_0 at the origin:

$$B_{12} \rightarrow \frac{2Q_1}{Q_1 + Q_2},\ \tilde{B}_{21} \rightarrow \frac{2Q_2}{Q_1 + Q_2},\ C_{12} \rightarrow \frac{2iQ_1Q_2}{Q_1 + Q_2},\ \tilde{C}_{21} \rightarrow -\frac{2iQ_1Q_2}{Q_1 + Q_2}. \tag{10.65}$$

Thus

$$\tau = \tau_0 + \frac{2iQ_1}{(Q_1 + Q_2)^2}\left[\frac{K^2}{\varepsilon_1\varepsilon_2}\Lambda_1 + Q_1Q_2\lambda_1\right] + \cdots. \tag{10.66}$$

The corresponding p wave transmittance is, on using $\tau_0 = 2Q_1/(Q_1 + Q_2)$,

$$\begin{aligned} T_p &= \frac{q_2}{q_1}|t_p|^2 = \frac{Q_2}{Q_1}|\tau|^2 \\ &= \frac{4Q_1Q_2}{(Q_1 + Q_2)^2}\left\{1 - \frac{2}{Q_1 + Q_2}\left[\frac{K^2}{\varepsilon_1\varepsilon_2}\mathrm{Im}\,\Lambda_1 + Q_1Q_2\mathrm{Im}\,\lambda_1\right] + \cdots\right\}. \end{aligned} \tag{10.67}$$

The imaginary parts of λ_1 and Λ_1 are both positive, so (10.60) and (10.67) show that, to first order in the film thickness, the transmission through a film is always decreased by absorption within the film (in contrast to the reflection, which we saw could be either decreased or increased by absorption). At normal incidence both (10.60) and (10.67) reduce to

$$T_n = \frac{4n_1 n_2}{(n_1 + n_2)^2}\left\{1 - \frac{2}{n_1 + n_2}\frac{\omega}{c}\mathrm{Im}\,\lambda_1 + \cdots\right\}. \tag{10.68}$$

The conservation law $R + T = 1$ for non-dissipative media can be generalized to $R + T + A = 1$, where A represents absorption within the system, and is nonnegative for passive media. From (10.55), (10.56), (10.60) and (10.67) we find the absorptance for the two polarizations:

$$A_s = \frac{4q_1}{(q_1 + q_2)^2}\frac{\omega^2}{c^2}\mathrm{Im}\,\lambda_1 + \cdots, \tag{10.69}$$

$$A_p = \frac{4Q_1}{(Q_1 + Q_2)^2}\left\{\frac{K^2}{\varepsilon_1\varepsilon_2}\mathrm{Im}\,\Lambda_1 + Q_2^2\mathrm{Im}\,\lambda_1\right\} + \cdots. \tag{10.70}$$

We now turn to the *ellipsometric characterization* of thin absorbing films. The derivation given in Sect. 3.4 remains valid for complex ε. To first order in the film thickness we have

$$r_{s0}\left(\frac{r_p}{r_s}\right) = r_{p0} - \frac{2iQ_1K^2/\varepsilon_1\varepsilon_2}{(Q_1 + Q_2)^2}I_1 + \cdots, \tag{10.71}$$

where the integral invariant $I_1 = \Lambda_1 - \lambda_1$ is given by

$$I_1 = \int_{-\infty}^{\infty} dz\frac{(\varepsilon - \varepsilon_1)(\varepsilon_2 - \varepsilon)}{\varepsilon} = \int_{-\infty}^{\infty} \mathrm{d}z\left(\varepsilon_1 + \varepsilon_2 - \frac{\varepsilon_1\varepsilon_2}{\varepsilon} - \varepsilon\right). \tag{10.72}$$

We again consider the simplest case where only the film is absorbing, with ε_1 and ε_2 real. Then, with $\varepsilon(z) = \varepsilon_r(z) + i\varepsilon_i(z)$,

$$I_1 = \int_{-\infty}^{\infty} \mathrm{d}z\left(\varepsilon_1 + \varepsilon_2 - \frac{\varepsilon_1\varepsilon_2\varepsilon_r}{\varepsilon_r^2 + \varepsilon_i^2} - \varepsilon_r\right) + i\int_{-\infty}^{\infty} \mathrm{d}z\left(\frac{\varepsilon_1\varepsilon_2}{\varepsilon_r^2 + \varepsilon_i^2} - 1\right)\varepsilon_i. \tag{10.73}$$

Since ε_r has the limiting values ε_1 and ε_2, and ε_i is zero outside the absorbing region, both integrands go to zero at the end-points. For non-absorbing films the principal angle θ_P (the ellipsometric Brewster angle) at which $\mathrm{Re}(r_p/r_s) = 0$, differs in second order in the film thickness from $\theta_B = \arctan\sqrt{(\varepsilon_2/\varepsilon_1)}$ (determined by $Q_1 = Q_2$), as we saw in Sects. 3.4 and 9.6. When the film is absorbing there is a first order correction: from (10.71) we find

$$\Delta\theta = \theta_P - \theta_B = \frac{(\omega/c)\mathrm{Im}\,I_1}{(\varepsilon_1 + \varepsilon_2)^{1/2}\left(\frac{\varepsilon_1}{\varepsilon_2}\right)^{1/2}\left(\frac{\varepsilon_1}{\varepsilon_2} - \frac{\varepsilon_2}{\varepsilon_1}\right)} + \cdots. \tag{10.74}$$

This difference between the principal and the Brewster angles is proportional to $\mathrm{Im}\,\Lambda_1 - \mathrm{Im}\,\lambda_1$, and may be large even for thin films if $\varepsilon_1 \approx \varepsilon_2$.

The case $\varepsilon_1 = \varepsilon_2$ (absorbing film between optically identical media) requires special consideration, since then both r_{s0} and r_{p0} are zero. The leading term in the ellipsometric ratio now depends on the ratio of Λ_1 to λ_1: from (10.52) and (10.53),

$$\frac{r_p}{r_s} = \frac{r_{p1}}{r_{s1}} + \cdots = \cos^2\theta_0 - \frac{\Lambda_1}{\lambda_1}\sin^2\theta_0 + \cdots . \tag{10.75}$$

where θ_0 is the common angle of incidence and transmission. The zero-thickness Brewster angle is $\theta_B = \pi/4$, while the principal angle at which $\mathrm{Re}(r_p/r_s) = 0$ is given by

$$\cot^2\theta_P = \mathrm{Re}\left(\frac{\Lambda_1}{\lambda_1}\right) + \cdots = \frac{\Lambda_r\lambda_r + \Lambda_i\lambda_i}{\lambda_r^2 + \lambda_i^2} + \cdots . \tag{10.76}$$

Here $\Lambda_1 = \Lambda_r + i\Lambda_i$ and $\lambda_1 = \lambda_r + i\lambda_i$, and the real and imaginary parts may be extracted from (10.54), with $\varepsilon_1 = \varepsilon_2 = \varepsilon_0$. This differs in second order in ε_i from the angle at which a non-absorbing film has zero reflection of the p wave, given by (3.59).

10.6 Attenuated Total Reflection, Surface Waves

When light is incident from a dielectric of refractive index n_1 onto another dielectric of refractive index $n_2 < n_1$ (for example from glass to air) there will be total reflection when $\theta_1 > \theta_c = \arcsin(n_2/n_1)$. This holds for both polarizations, and irrespective of whether the transition between the dielectrics is sharp or gradual, *provided there is no absorption within the interface*. When an absorbing layer (typically a metal film) is deposited between the two dielectrics, the transmission is still zero (since $q_2 = (\varepsilon_2\omega^2/c^2 - K^2)^{1/2}$ is imaginary for $\theta_1 > \theta_c$) but the p wave reflectance can be very much less than total. A sharp resonance in the absorption can appear, and this is the basis of an experimental technique for the determination of the optical constants of metal and semiconductor films. The technique is due to Otto (1968) and Kretschmann and Raether (1968). Otto originally referred to "frustrated total reflection"; it is now called *attenuated total reflection*. Two basic configurations are illustrated in Fig. 10.4.

We will consider the two configurations of Fig. 10.4 in this chapter. The symmetric high|low|complex|low|high configuration will be treated in Sect. 12.6, since it is more easily treated by matrix methods.

For the *dielectric|absorbing layer|dielectric* case (upper diagram in Fig. 10.4), we can use the results of Sect. 10.4, taking account of the fact that q_2 and Q_2 become imaginary for $\theta_1 > \theta_c = \arcsin(n_2/n_1)$. The s and p reflectivities have the form (10.48),

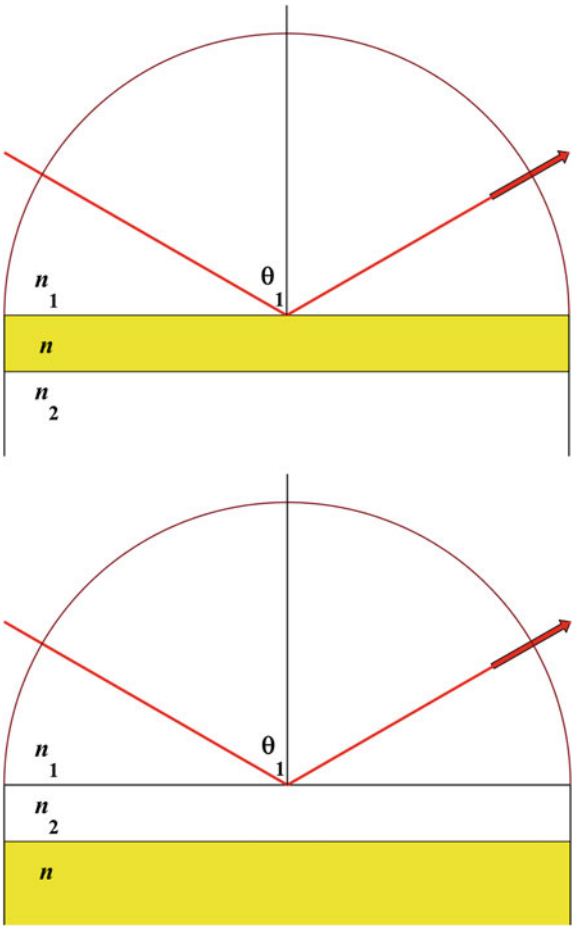


Fig. 10.4 Two configurations for attenuated total reflection. Prisms are often used instead of the half-cylinders illustrated. The shaded material is a conductor, with $n = n_r + in_i$. The medium of incidence has the higher index: $n_1 > n_2$

$$R_s, R_p = \frac{\rho^2 + 2\rho\rho' \cos(\delta - \delta') + (\rho')^2}{1 + 2\rho\rho' \cos(\delta + \delta') + (\rho\rho')^2}, \tag{10.77}$$

where ρ, δ and ρ', δ' are defined for the two polarizations by

$$\rho_s e^{i\delta_s} = \frac{q_1 - q}{q_1 + q}, \quad \rho'_s e^{i\delta'_s} = \frac{q - q_2}{q + q_2} e^{2iq\Delta z}, \tag{10.78}$$

$$\rho_p e^{i\delta_p} = \frac{Q_1 - Q}{Q_1 + Q},\ \rho'_p e^{i\delta'_p} = \frac{Q - Q_2}{Q + Q_2} e^{2iq\Delta z}. \tag{10.79}$$

The wavenumber normal components q and Q are complex, with real and imaginary parts given as before by (10.44, 10.45) and (10.21). We have

$$\rho_s^2 = \frac{(q_1 - q_r)^2 + q_i^2}{(q_1 + q_r)^2 + q_i^2},\ \rho_p^2 = \frac{(Q_1 - Q_r)^2 + Q_i^2}{(Q_1 + Q_r)^2 + Q_i^2}. \tag{10.80}$$

We interpret ρ_s and ρ_p as the positive square roots of these expressions, and define $\text{atn}(y, x)$ as the arctangent of y/x placed in the correct quadrant according to the signs of x and y. The corresponding phases are then

$$\begin{aligned} \delta_s &= \text{atn}\left(-2q_1 q_i,\ q_1^2 - q_r^2 - q_i^2\right), \\ \delta_p &= \text{atn}\left(-2Q_1 Q_i,\ Q_1^2 - Q_r^2 - Q_i^2\right). \end{aligned} \tag{10.81}$$

The primed variables take different forms depending on whether θ_1 is less or greater than θ_c. For $\theta_1 < \theta_c$, we find

$$\rho'_s = e^{-2q_i\Delta z}\left\{\frac{(q_r - q_2)^2 + q_i^2}{(q_r + q_2)^2 + q_i^2}\right\}^{1/2},\ \rho'_p = e^{-2q_i\Delta z}\left\{\frac{(Q_r - Q_2)^2 + Q_i^2}{(Q_r + Q_2)^2 + Q_i^2}\right\}^{1/2}. \tag{10.82}$$

The positive square roots are again understood. The corresponding phases are

$$\begin{aligned} \delta'_s &= 2q_r\Delta z + \text{atn}\left(2q_i q_2,\ q_r^2 + q_i^2 - q_2^2\right), \\ \delta'_p &= 2q_r\Delta z + \text{atn}\left(2Q_i Q_2,\ Q_r^2 + Q_i^2 - Q_2^2\right). \end{aligned} \tag{10.83}$$

For $\theta_1 > \theta_c$ we set $q_2 = i|q_2|$ and $Q_2 = i|Q_2|$. Then

$$\rho'_s = e^{-2q_i\Delta z}\left\{\frac{q_r^2 + (q_i - |q_2|)^2}{q_r^2 + (q_i + |q_2|)^2}\right\}^{1/2},\ \rho'_p = e^{-2q_i\Delta z}\left\{\frac{Q_r^2 + (Q_i - |Q_2|)^2}{Q_r^2 + (Q_i + |Q_2|)^2}\right\}^{1/2}, \tag{10.84}$$

$$\begin{aligned} \delta'_s &= 2q_r\Delta z + \text{atn}\left(-2q_r|q_2|,\ q_r^2 + q_i^2 - |q_2|^2\right), \\ \delta'_p &= 2q_r\Delta z + \text{atn}\left(-2Q_r|Q_2|,\ Q_r^2 + Q_i^2 - |Q_2|^2\right). \end{aligned} \tag{10.85}$$

Figure 10.5 shows the s and p reflectivities for a high refractive index glass|silver film|lithium fluoride system at $\lambda_0 = 546\,\text{nm}$. The refractive indices are $n_1 =$

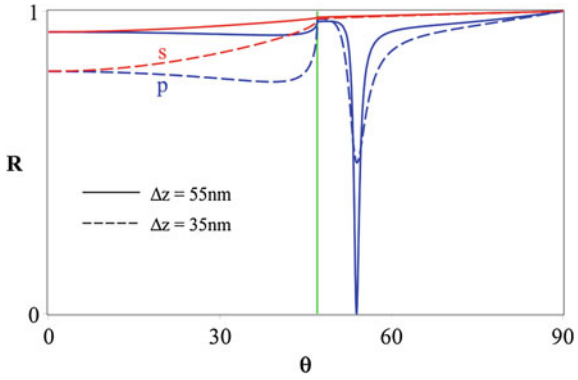


Fig. 10.5 Attenuated total reflection for the upper configuration in Fig. 10.4, for incidence from a high refractive index glass onto a silver film, with a lithium fluoride substrate. The p reflectivity shows a sharp minimum in the region where total reflection would occur in the absence of the metal film of thickness Δz (the critical angle is $\theta_c \approx 47°$, indicated by the vertical line in the Figure)

$1.9018, n = 0.055 + 3.28i, n_2 = 1.392$ (these correspond to those used in Fig. 13b of Otto 1976). The values of $(\omega/c)\Delta z$ for the silver film thicknesses of 35 and 55 nm are 0.403 and 0.633.

An estimate of the location of the minimum in the p reflectance can be obtained from the thin film formula (10.53). For $\theta_1 > \theta_c$ we have $Q_2 = i|Q_2|$, and (10.53) gives

$$R_p = 1 - \frac{4Q_1}{Q_1^2 + |Q_2|^2}\left[\frac{K^2}{\varepsilon_1\varepsilon_2}\mathrm{Im}\,\Lambda_1 + |Q_2|^2\mathrm{Im}\,\lambda_1\right] + \cdots . \tag{10.86}$$

For a homogeneous metallic film of thickness Δz this becomes (on using (10.57))

$$R_p = 1 - \frac{4Q_1}{Q_1^2 + |Q_2|^2}\left(\frac{K^2}{\varepsilon_r^2 + \varepsilon_i^2} + |Q_2|^2\right)\varepsilon_i \Delta z + \cdots . \tag{10.87}$$

This has a minimum when

$$\left(\frac{cK}{\omega}\right)^2 = \varepsilon_1 \sin^2\theta_1 = \frac{1}{2}\left\{3v - u + \left[(3v-u)^2 - 4uv + 8\varepsilon_1(u-v)\right]^{1/2}\right\}, \tag{10.88}$$

where

$$u = \frac{(\varepsilon_r^2 + \varepsilon_i^2)\varepsilon_2}{\varepsilon_r^2 + \varepsilon_i^2 + \varepsilon_2^2}, \quad v = \frac{\varepsilon_1\varepsilon_2}{\varepsilon_1 + \varepsilon_2}. \tag{10.89}$$

For the parameters of Fig. 10.5 these expressions locate the minimum at about 59°. The actual minima are at smaller angles: the 55 nm metal film (for which $(\omega/c)\Delta z \approx 0.633$) has a reflectance minimum at about 54°. One reason for the lack of precision in this estimate is that the thin film formulae contain no direct information about the complex wavenumber component $q = q_r + iq_i$ within the absorber. For metals, ε_r can be large and negative, and q_i is then much larger than q_r, making the effect of the $\exp(-2q_i\Delta z)$ factor in the reflection formulae very strong. The same comments apply to (10.74), which is accurate for metallic films only when these are unrealistically thin.

The ellipsometric quantity r_p/r_s shows remarkable behaviour in the vicinity of strong attenuated total reflection. It is given by

$$\frac{r_p}{r_s} = -\frac{\rho_p e^{i\delta_p} + \rho'_p e^{i\delta'_p}}{1 + \rho_p \rho'_p e^{i\left(\delta_p + \delta'_p\right)}} \cdot \frac{1 + \rho_s \rho'_s e^{i(\delta_s + \delta'_s)}}{\rho_s e^{i\delta_s} + \rho'_s e^{i\delta'_s}}, \tag{10.90}$$

where the magnitudes and phases of the component amplitudes are given by (10.78)–(10.85). For thick metal layers the trajectories r_p/r_s tend to those of Fig. 10.2; for very thin metal films the trajectories approach those of Fig. 2.9. The layers of intermediate thickness which show strongly attenuated total reflection have a variety of trajectories between these two limiting cases. Two examples, corresponding to those in Fig. 10.5, are shown in Fig. 10.6: note the very rapid variation with angle near θ_c and in the vicinity of the reflection minimum. The 55 nm film gives an r_p/r_s trajectory which passes close to the origin, corresponding to the very small R_p value shown in Fig. 10.5.

We mentioned in the discussion following (10.89) the reason why the thin film formulae (10.85–10.89) for the location of the p reflectivity minimum might give accurate results only for unrealistically thin metal films. Likewise (10.74) predicts a small negative shift $\Delta\theta = \theta_P - \theta_B$ for the silver film case illustrated in Figs. 10.5 and 10.6. For vanishing thickness of silver $\theta_P = \theta_B \approx 36°$, while for thick silver layers $\theta_P \approx 65°$, an increase of nearly 30°. This large increase is almost complete when $(\omega/c)\Delta z \approx 1$, and swamps the small predicted decrease even for monolayers of silver.

The phenomenon of attenuated total reflection is due to the generation of *electromagnetic surface waves* in situations where total reflection would occur in the absence of the metal layer. Consider the simplest possible case of an idealized conductor with negative dielectric function, for example, $\varepsilon = 1 - \omega_p^2/\omega^2$ with $\omega < \omega_p$, bounded by a dielectric with $\varepsilon_2 > 0$. A surface wave solution for the p polarization is possible: for an interface in the xy plane $z = 0$, let

$$B_y(z, x, t) = \begin{cases} e^{i(Kx-\omega t)} e^{|q|z} & (z < 0) \\ e^{i(Kx-\omega t)} e^{-|q_2|z} & (z > 0). \end{cases} \tag{10.91}$$

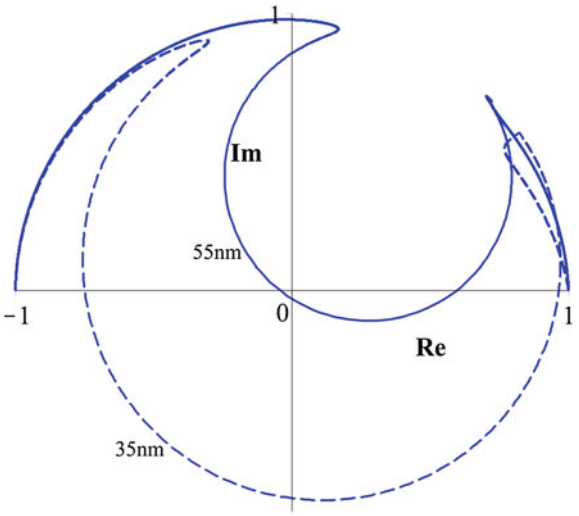


Fig. 10.6 The ellipsometric ratio r_p/r_s in the complex plane, for the prism|metal layer|dielectric configuration illustrated in the upper diagram of Fig. 10.4. The parameters used are those of Fig. 10.5. The 35 nm silver film has a single principal angle $\theta_P \approx 53.4°$, while the 55 nm thick film shows triple principal angles ($\theta_P \approx 53.9°$, 55.1°, 63.4°). (The possibility of multiple principal angles was discussed in Sect. 2.3). As the thickness of silver is increased the indentation near 54° diminishes, and eventually the trajectory becomes a simple arc in the upper half plane, and crosses the $Re(r_p/r_s) = 0$ axis at $\theta_P \approx 65°$

This function satisfies (1.18) and its consequent boundary conditions, namely the continuity of B_y and $\varepsilon^{-1}\partial B_y/\partial z$ at $z = 0$, provided

$$K^2 - |q|^2 = \varepsilon\frac{\omega^2}{c^2}, \quad K^2 - |q_2|^2 = \varepsilon_2\frac{\omega^2}{c^2}, \quad \frac{|q|}{\varepsilon} = -\frac{|q_2|}{\varepsilon_2} \tag{10.92}$$

(note that a surface wave solution for s polarization is not possible, since this would require $|q| = -|q_2|$). The first two equations may be regarded as the usual relation between the tangential and normal components of the wave-number, $K^2 + q^2 = \varepsilon\omega^2/c^2$, except that q and q_2 are imaginary and give exponential decay away from the surface rather than propagation in the z direction. The last relation in (10.92) can be satisfied if $|\varepsilon| > \varepsilon_2$, since $|q| = (K^2 + |\varepsilon|\omega^2/c^2)^{1/2} > (K^2 - \varepsilon_2\omega^2/c^2)^{1/2} = |q_2|$. On eliminating $|q|$ and $|q_2|$ from (10.92) we find for the wavevector component K (for propagation along the interface) the dispersion relation

$$K^2 = \frac{|\varepsilon|\varepsilon_2}{|\varepsilon| - \varepsilon_2}\frac{\omega^2}{c^2}. \tag{10.93}$$

The hypothetical electromagnetic surface wave described above has no real normal component of its wavevector and thus cannot be coupled into by an incident plane wave. When incidence is from an optically denser medium, so as to produce total reflection in the absence of the metal film, strong coupling is possible with p-polarization for a special combination of angle of incidence and thickness of metal. The modulus of the magnetic field is shown within the silver layer and the lithium fluoride substrate in Fig. 10.7. Inside the Ag layer the magnetic field equals $Ce^{iqz} + De^{-iqz}$, with $C = \frac{2Q_1}{Q+Q_1}\left[1 + e^{2iq\Delta z}p_1p_2\right]^{-1}, D = e^{2iq\Delta z}p_2C$, where $p_1 = \frac{Q_1-Q}{Q_1+Q}, p_2 = \frac{Q-Q_2}{Q+Q_2}$. The waves decay exponentially inside LiF, since the angles of incidence exceed the critical angle.

We see that near the angle for minimum reflectivity (here about 53.9°) the fields peak at the interface between the metal and the second dielectric, as in the idealized metal|dielectric case discussed above. As usual the fields decay exponentially into the second dielectric, but at resonance increase (approximately exponentially) into the metal, as opposed to the usual decay away from the illuminated surface.

The conditions for minimum reflection, and thus maximum absorption within the metal film, can be seen from (10.77). For R_p to be minimum we need ρ_p and ρ'_p approximately equal, and $\cos(\delta_p - \delta'_p)$ near -1. Since δ'_p contains the factor $\exp(-2q_i\Delta z)$ it would normally be much smaller than ρ_p, especially as q_i is large for metals. But by varying the angle it is possible to make $Q_i + Q_2$ zero or very small, Q_i being usually negative for metals (note that $\theta_1 > \theta_c$ is needed here). Since Q_r is small, this makes the factor multiplying $\exp(-2q_i\Delta z)$ in ρ'_p large, and $\rho'_p \approx \rho_p$. Both angle and metal layer thickness adjustment are involved in attaining approximate equality of ρ_p and ρ'_p. The other condition, $\cos\left(\delta_p - \delta'_p\right) \approx -1$,

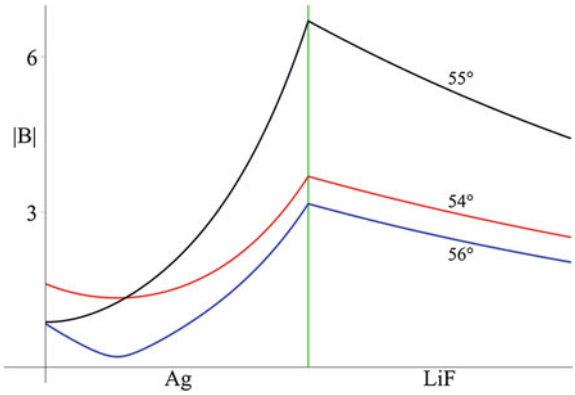


Fig. 10.7 The modulus of the magnetic field **B** as a function of z, for the prism|silver|lithium fluoride configuration, with the Ag layer being 55 nm thick. Note the sensitivity to angle of incidence: in this case even a 1° shift is sufficient to decrease the peak amplitude by a factor of 2. The parameters are as in Figs. 10.5 and 10.6, and the incident field has unit amplitude. The boundary between silver and lithium fluoride is marked by the vertical line

depends mainly on angle, since here the metal thickness enters as $q_r\Delta z$, and q_r is small.

These rough arguments can be made more precise. For example, we can ask whether a certain combination of prism|metal|dielectric can give zero reflection (total absorption) at some angle and thickness combination. This amounts to satisfying $\rho_p = \rho'_p$ and $\cos\left(\delta_p - \delta'_p\right) = -1$ simultaneously. On eliminating Δz between these two relations, we find the condition

$$\mathrm{atn}(-2Q_1Q_i, Q_1^2 - Q_r^2 - Q_i^2) - \mathrm{atn}(-2Q_r|Q_2|, Q_r^2 + Q_i^2 - |Q_2^2|)$$

$$-\frac{1}{2}\frac{q_r}{q_i}\ln\left\{\frac{Q_r^2 + (Q_i - |Q_2|)^2}{Q_r^2 + (Q_i + |Q_2|)^2}\cdot\frac{(Q_1 - |Q_r|)^2 + Q_i^2}{(Q_1 - |Q_r|)^2 - Q_i^2}\right\} = (2m+1)\pi \qquad (10.94)$$

where m is a positive or negative integer, or zero. For the case illustrated in Figs. 10.5, 10.6 and 10.7 we find that (10.94) is satisfied (with $m = 0$) at $\theta_0 \approx 53.9°$, and gives the optimum thickness $\Delta z = (4q_i)^{-1}\ln\{\}_0 \approx 55.6\,\mathrm{nm}$, where $\{\}_0$ denotes the contents of the braces in (10.94) evaluated at θ_0. For perfectly attenuated total reflection the trajectory of r_p/r_s passes through the origin at θ_0. The trajectory for the 55 nm thickness of silver shown in Fig. 10.6 passed close to the origin, but this thickness does not quite give a perfect absorber at θ_0.

For vapour-deposited Al at 633 nm (refractive index $1.212 + 6.924i$, as in Fig. 10.3) between the same two dielectrics, (10.94) gives $\theta_0 \approx 49.5°$ (again with $m = 0$), and an optimum thickness of about 13 nm.

10.7 Attenuated Total Reflection: Second Example

We now turn to the lower configuration in Fig. 10.4, in which the prism is followed by a low-index material (or an air gap) and then by a metallic substrate. In the absence of the metal we would have exponential decay of the fields into the second dielectric for $\theta_1 > \arcsin(n_2/n_1)$, since $q_2 = i|q_2|$ is then imaginary. When the metal is present, both exponential increase and decrease are possible, these going as $\exp(\pm|q_2|z)$. Attenuated total reflection occurs when the increase dominates, producing large fields at the second dielectric|metal boundary, and thus large absorption. The reflection amplitudes may be obtained as before. We have, for the $n_1|n_2|n_r + in_i$ configuration with boundaries at $z = z_1$ and at $z_1 + \Delta z$,

$$r_s = \mathrm{e}^{2iq_1z_1}\frac{r + r'\mathrm{e}^{2iq_2\Delta z}}{1 + rr'\mathrm{e}^{2iq_2\Delta z}},\quad r = \frac{q_1 - q_2}{q_1 + q_2},\quad r' = \frac{q_2 - q}{q_2 + q}, \qquad (10.95)$$

where

$$q_1^2 = \varepsilon_1\omega^2/c^2 - K^2 = \varepsilon_1\left(\omega^2/c^2\right)\cos^2\theta_1,\ q_2^2 = \varepsilon_2\omega^2/c^2 - K^2,\ q^2 = \varepsilon\omega^2/c^2 - K^2 \tag{10.96}$$

The real and imaginary parts of q are given by (10.11) and (10.12) or (10.44) and (10.45). We again set $r = \rho e^{i\delta}$ and $r' e^{2iq_2\Delta z} = \rho' e^{i\delta'}$; when q_2 is imaginary ρ' includes the exponential factor $e^{-2|q_2|\Delta z}$. The p wave reflection amplitude is

$$-r_p = e^{2iq_1 z_1}\frac{r + r' e^{2iq_2\Delta z}}{1 + rr' e^{2iq_2\Delta z}}, \tag{10.97}$$

where now

$$r = \frac{Q_1 - Q_2}{Q_1 + Q_2},\quad r' = \frac{Q_2 - Q}{Q_2 + Q}, \tag{10.98}$$

with $Q_1 = q_1/\varepsilon_1$, $Q_2 = q_2/\varepsilon_2$ and $Q = q/\varepsilon$ has real and imaginary parts given by (10.21). The amplitudes ρ, ρ' and the phases δ and δ' are defined as for the s wave. The reflectivities then take the form (10.77). For $\theta_1 < \theta_c = \arcsin(n_2/n_1)$ we have

$$\rho_s = \frac{q_1 - q_2}{q_1 + q_2},\ \rho_p = \frac{Q_1 - Q_2}{Q_1 + Q_2}$$

$$\rho_s' = \left\{\frac{(q_2 - q_r)^2 + q_i^2}{(q_2 + q_r)^2 + q_i^2}\right\}^{1/2},\ \delta_s' = \operatorname{atn}\left(-2q_2 q_i, q_2^2 - q_r^2 - q_i^2\right) + 2q_2\Delta z.$$

$$\rho_p' = \left\{\frac{(Q_2 - Q_r)^2 + Q_i^2}{(Q_2 + Q_r)^2 + Q_i^2}\right\}^{1/2},\ \delta_p' = \operatorname{atn}\left(-2Q_2 Q_i, Q_2^2 - Q_r^2 - Q_i^2\right) + 2q_2\Delta z. \tag{10.99}$$

(In the $\theta_1 < \theta_c$ case it is convenient to set $\delta_p = 0$ and allow ρ_p to carry the change of sign at the Brewster angle $\theta_B = \arctan(n_2/n_1)$, where $Q_1 = Q_2$.) For angle of incidence greater than the critical angle, $q_2 = i|q_2|, Q_2 = i|Q_2|$, and

$$\rho_s = 1,\quad \delta_s = \operatorname{atn}\left(-2q_1|q_2|, q_1^2 - |q_2|^2\right),$$

$$\rho_p = 1,\quad \delta_p = \operatorname{atn}\left(-2Q_1|Q_2|, Q_1^2 - |Q_2|^2\right), \tag{10.100}$$

$$\rho_s' = \left\{\frac{q_r^2 + (|q_2| - q_i)^2}{q_r^2 + (|q_2| + q_i)^2}\right\}^{1/2} e^{-2|q_2|\Delta z},\quad \delta_s' = \operatorname{atn}\left(2|q_2|q_r, |q_2|^2 - q_r^2 - q_i^2\right),$$

$$\rho'_p = \left\{\frac{Q_r^2 + (|Q_2| - Q_i)^2}{Q_r^2 + (|Q_2| + Q_i)^2}\right\}^{1/2} \mathrm{e}^{-2|q_2|\Delta z}, \quad \delta'_p = \mathrm{atn}\left(2|Q_2|Q_r, |Q_2|^2 - Q_i^2 - Q_r^2\right).$$

The p wave reflectivity can be zero if $\rho_p = \rho'_p$ and $\cos\left(\delta_p - \delta'_p\right) = -1$ are satisfied simultaneously. The angle θ_0 at which this can happen is (for $\theta_1 > \theta_c$) found from

$$\mathrm{atn}\left(-2Q_1|Q_2|, Q_1^2 - |Q_2|^2\right) - \mathrm{atn}\left(2|Q_2|Q_r, |Q_2|^2 - Q_i^2 - Q_r^2\right) = (2m+1)\pi. \tag{10.101}$$

The thickness of the second dielectric which gives perfectly attenuated total reflection (that is, total absorption) is given by the following expression evaluated at θ_0:

$$\Delta z = \frac{1}{4|q_2|} \ln\left\{\frac{Q_r^2 + (|Q_2| - Q_i)^2}{Q_r^2 + (|Q_2| + Q_i)^2}\right\}. \tag{10.102}$$

For the high refractive index prism|lithium fluoride|silver system at a wavelength of 546 nm, with the refractive indices $1.9018, 1.392, 0.055 + 3.28i$, (10.101) gives $\theta_0 \approx 54°$ (with $m = -1$) and (10.102) gives $\Delta z \approx 307$ nm. There is also zero reflection of the p wave near grazing incidence, with $\theta_0 \approx 89°$ (again $m = -1$) and $\Delta z \approx 39.3$ nm: compare the discussion of reflection polarizers consisting of a dielectric layer on a metal substrate in Sect. 10.3. The reflectivities are given by (10.77) and r_p/r_s by (10.90). Figure 10.8 shows R_p for the above combination,

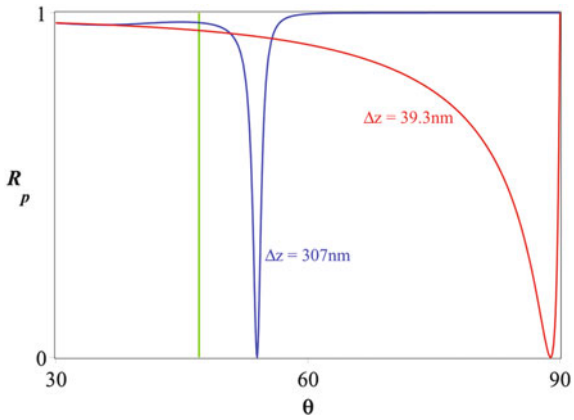


Fig. 10.8 The p reflectivity as a function of angle of incidence, for the lower configuration in Fig. 10.4, shown for the two thicknesses of the second dielectric (LiF) which give total absorption of the p wave at one angle. The thicknesses and angles are 307 nm, 54° and 39.3 nm, 89° ($at\ \lambda_0 = 546$ nm). The substrate is silver. The critical angle $\theta_c \approx 47°$ is marked by the vertical line

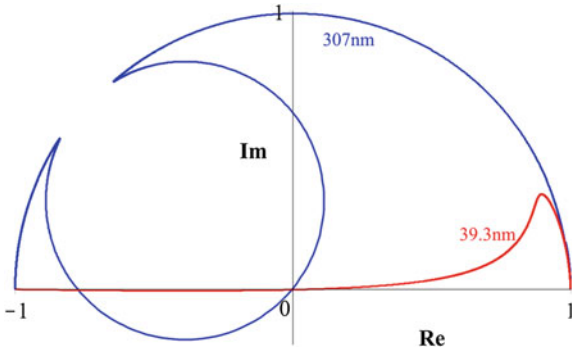


Fig. 10.9 The ratio r_p/r_s is the complex plane, for the two dielectric thicknesses which give perfectly attenuated reflection at one angle. The thicknesses are 39.3 and 307 nm; the corresponding curves pass through the origin at about 89° and 54° respectively. The 307 nm diagram shows the triple principal angle phenomenon, as in Fig. 10.6

at the two optimum thicknesses of the second dielectric, which correspond to $(\omega/c)\Delta z = 3.53$ and 0.452. Figure 10.9 shows the corresponding r_p/r_s curves.

There is wide variety in both the reflectance and the ellipsometric curves as the thickness of the second dielectric varies. For thin layers of the dielectric the curves tend to those of Sect. 10.1.

The phenomenon of attenuated total reflection has been treated here purely by classical electrodynamics. It was seen to be an interference-attenuation effect, linked to the excitation of electromagnetic surface waves. These are coupled into by means of the exponential decay or growth in the metallic layer or the second dielectric which is possible for $\theta_1 > \theta_c$. In radio physics the electromagnetic surface waves sometimes go by the names of Zenneck or Sommerfeld-Zenneck (Barlow and Brown 1962), or ground waves (Budden 1985). In solid state physics the terms surface polariton, surface plasmon, or sometimes surface polariton-plasmon or phonon-polariton are used. There the phenomenon of attenuated total reflection has many applications: for example the determination of the optical constants of metals and semiconductors (Otto 1976), and the study of adsorbates (McIntyre 1976). The literature on the solid state aspects of surface wave phenomena is very large; see for example the collections of papers edited by Burstein and DeMartini (1974), Seraphin (1976), Boardman (1982), and Agranovich and Mills (1982).

10.8 Reflection by a Diffuse Absorbing Interface: The Tanh Profile

In Sect. 2.5 we considered the reflection properties of the hyperbolic tangent profile

$$\varepsilon(z) = \frac{1}{2}(\varepsilon_1 + \varepsilon_2) - \frac{1}{2}(\varepsilon_1 - \varepsilon_2)\tanh\frac{z}{2a} = \frac{\varepsilon_1 + \varepsilon_2 e^{z/a}}{1 + e^{z/a}}. \tag{10.103}$$

The s wave reflectivity was shown to be

$$R_s = \left\{\frac{\sinh \pi a(q_1 - q_2)}{\sinh \pi a(q_1 + q_2)}\right\}^2 \tag{10.104}$$

for real q_2, and unity for imaginary q_2 [$q_2 = i|q_2|$ for $\theta_1 > \theta_c = \arcsin(\varepsilon_2/\varepsilon_1)^{1/2}$]. Here we shall discuss reflection by the profile (10.103) when ε_2 is complex, $\varepsilon_2 = \varepsilon_r + i\varepsilon_i$. The particular example we have in mind is reflection by an ionospheric layer in which the electron density at the lower ionospheric boundary approximately takes the functional form (10.103). The electron gas dielectric function (10.1) has the real and imaginary parts

$$\varepsilon_r = 1 - \frac{\omega_p^2}{\omega^2 + 1/\tau^2}, \quad \varepsilon_i = \frac{\omega_p^2/\omega\tau}{\omega^2 + 1/\tau^2} \tag{10.105}$$

Since ω_p^2 is proportional to the electron density, both ε_r and ε_i take the form (10.103), and so does $\varepsilon(z)$ with $\varepsilon_2 = \varepsilon_r + i\varepsilon_i$, if the variation of τ through the inhomogeneity can be neglected.

The theory leading to r_s as given by (2.84) remains valid when ε_2 is complex, with q_2 being replaced by $q_r + iq_i$. The reflection amplitude in the absorbing case is thus given by (with $y_1 = q_1 a$ as before and $q_2 a = y_r + iy_i$)

$$r_s = -\frac{\Gamma(2iy_1)\Gamma(y_i - i(y_1 + y_r))\Gamma(-y_i - i(y_1 - y_r))\sinh \pi(y_1 - y_r - iy_i)}{\Gamma(-2iy_1)\Gamma(-y_i + i(y_1 + y_r))\Gamma(y_i + i(y_1 - y_r))\sinh \pi(y_1 + y_r + iy_i)} \tag{10.106}$$

From (2.86) the ratio $\Gamma(2iy_1)/\Gamma(-2iy_1)$ has modulus unity, and so

$$R_s = |r_s|^2 = G\frac{\sinh^2 \pi(y_1 - y_r) + \sin^2 \pi y_i}{\sinh^2 \pi(y_1 + y_r) + \sin^2 \pi y_i} \tag{10.107}$$

where G is the modulus squared of the gamma function ratios in (10.106):

$$G = \left|\frac{\Gamma(y_i - i(y_1 + y_r))\Gamma(-y_i - i(y_1 - y_r))}{\Gamma(-y_i + i(y_1 + y_r))\Gamma(y_i + i(y_1 - y_r))}\right|^2. \tag{10.108}$$

To evaluate G we consider the ratio $\Gamma(-z)/\Gamma(z)$. From the infinite product representation (2.85) we have

$$\frac{\Gamma(-z)}{\Gamma(z)} = -e^{2\gamma z}\prod_{n=1}^{\infty}\left(\frac{n+z}{n-z}\right)\mathrm{e}^{-2\pi/n} \tag{10.109}$$

Thus, with $z = x + iy$,

$$\frac{\Gamma(-z)}{\Gamma(z)} = e^{4\gamma x}\prod_{n=1}^{\infty}\left(\frac{(n+x)^2+y^2}{(n-x)^2+y^2}\right)^2\mathrm{e}^{-4x/n} \tag{10.110}$$

Both $\Gamma(-z)/\Gamma(z)$ ratios in G have the same real part of z, so the exponential factors cancel and

$$G(y_1, y_r, y_i) = \prod_{n=1}^{\infty}\frac{(n-y_i)^2+(y_1+y_r)^2}{(n+y_i)^2+(y_1+y_r)^2}\cdot\frac{(n+y_i)^2+(y_1-y_r)^2}{(n-y_i)^2+(y_1-y_r)^2} \tag{10.111}$$

Since q_1, q_r and q_i are all non negative, G is always greater than unity in the presence of absorption. G tends to 1 as $\varepsilon_i \to 0$, and also at grazing incidence where q_1 and thus y_1 tend to zero.

Figure 10.10 shows the s wave reflectivity for a *tanh* profile with $\varepsilon_r = 0.25$ and $\varepsilon_i = 0.001$, corresponding roughly to a frequency a bit above ($2/\sqrt{3}$ larger than) the maximum plasma frequency, with absorption typical of the ionospheric E layer. In the absence of absorption there would be total reflection for angle of incidence greater than $\arcsin(1/2) = 30°$, for any layer thickness. Electron collisions decrease the reflectivity, the decrease being greater for greater thickness of the transition, there being more penetration into the absorbing region.

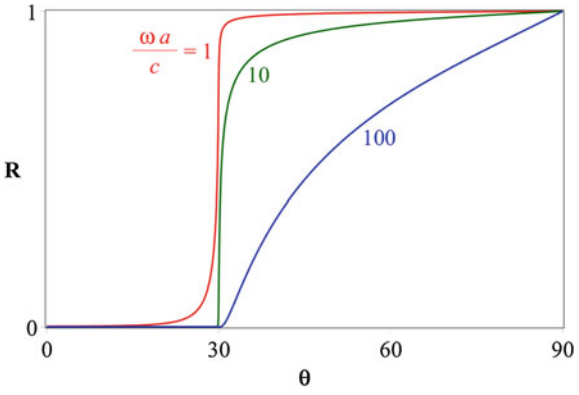


Fig. 10.10 The *s* wave reflectivity as a function of the angle of incidence for the *tanh* profile with fixed absorption and varying thickness a ($\varepsilon_1 = 1, \varepsilon_r = 0.25, \varepsilon_i = 0.001, (\omega/c)a = 1, 10$ *and* 100)

An interesting phenomenon appears in the reflection at a gradual transition to negative ε_r. In the absence of absorption both s and p waves would be totally reflected. The turning point where $q^2(z) = 0$ is given by $\varepsilon(z) = \varepsilon_1 \sin^2\theta_1$, and the s wave shows exponential decay beyond this point. The p wave however has a singularity at $\varepsilon(z) = 0$ arising from the $\varepsilon^{-1}\mathrm{d}B/\mathrm{d}z$ term in the equation satisfied by B, and this leads to a logarithmic singularity in B, and infinities in E_x and E_z. Absorption removes the infinities, but the fields at the point where $\varepsilon_r(z) = 0$ can still be large. This problem is discussed by Landau and Lifshitz (1960, Sect. 68) for real ε; the effect of absorption is considered by Ginzburg (1964, Sect. 20) and Budden (1985, Sect. 15.6) who also give references to earlier work.

10.9 Zero Reflection from Dielectric Layer on Absorbing Substrate

In Sect. 10.3 we mentioned that (10.40) can be reduced to a quadratic in the variable $C = \cos^2\theta_1$, which we write as $c_0 + c_1 C + c_2 C^2 = 0$. The coefficients of the quadratic are

$$\begin{aligned} c_0 &= -\varepsilon_i^2(\varepsilon - \varepsilon_1)^4 \\ c_1 &= 4\varepsilon_1(\varepsilon - \varepsilon_1)^2(\varepsilon_1 - \varepsilon_r)[(\varepsilon - \varepsilon_r)^2 + \varepsilon_i^2] \\ c_2 &= 4\varepsilon_1^2[2\varepsilon(\varepsilon_1 - \varepsilon_r) - \varepsilon_1^2 + \varepsilon_r^2 + \varepsilon_i^2][(\varepsilon - \varepsilon_r)^2 + \varepsilon_i^2] \end{aligned} \tag{10.112}$$

When C is small, as it is at glancing incidence, the solution is approximately $C = -c_0/c_1$. For the example quoted in the text, this gives the glancing angle at which the s wave reflection is zero as 2.09°, instead of the exact 2.05°. When we keep only the constant and linear terms in the p-wave sextic, the glancing angles for zero reflection are 1.32° (approximate) and 1.34° (exact).

References

Abelès F (1950) Recherches sur la propagation des ondes électromagnetiques sinusoidales dans les milieux stratifiés. Application aux couches minces. Annales de Physique 5:596–640

Agranovich VM, Mills DL (eds) (1982) Surface polaritons: electromagnetic waves at surfaces and interfaces. North-Holland, Amsterdam; In: Abelès F, Lopez-Rios T (eds) Surface polaritons at metal surfaces and interfaces

Allen TH (1976) Study of Al with a combined Auger electron spectrometer-ellipsometric system. J Vac Sci Technol 13:112–115

Azzam RMA (1985) Explicit equations for the polarizing angles of a high-reflectance substrate coated by a transparent thin film. J Opt Soc Amer A 2:480–482

Barlow HM, Brown J (1962) Radio surface waves. Oxford University Press, Oxford

Bennett JM, Bennett HE (1978) Polarization, Chap. 10. In: Driscoll WG and Vaughan W (eds) Handbook of optics. McGraw Hill, New York
Boardman AD (ed) (1982) Electromagnetic surface modes. Wiley, New York
Born M and Wolf E (1970) Principles of optics, 4th edn. Pergamon, New York
Booker H G (1984) Cold plasma waves. Martinus Nijhoff, Dordrecht
Budden K G (1985) The propagation of radio waves. Cambridge University Press, Cambridge
Burnstein E and DeMartini F (eds.) (1974) Polaritons. Pergamon, New York
Ginzburg V L (1964) The propagation of electromagnetic waves in plasmas. Pergamon, Oxford
Kitajima H, Fujita K, Cizmic H (1984) Zero reflection from a dielectric film on a metal substrate at oblique angles of incidence. Appl Opt 23:1937–1939
Kittel C (1966) *Introduction to solid state physics*, Wiley
Kretschmann E, Raether H (1968) Radiative decay of non radiative surface plasmons by light, Zeit. für Naturforschung, **23a**, 2135–2136
Landau LD, Lifshitz EM (1960) Electrodynamics of continuous media, Pergamon, Oxford
McIntyre JDE (1976) Optical reflection spectroscopy of chemisorbed monolayers, Chap. 11, In: Seraphin (ed)
Otto A (1968) Excitation of nonradiative surface plasma waves in silver by the method of frustrated total reflection. Zeit. für Physik 216:398–410
Otto A (1976), Spectroscopy of surface polaritons by attenuated total reflection, Chap. 13, In: Seraphin (ed)
Potter RF (1969) Pseudo-Brewster angle technique for determining the optical constants, Ch. 16. In: Nudelman S et al. (eds) Optical properties of solids. Springer, Berlin
Ruiz-Urbieta M, Sparrow EM (1972) Reflection polarization by a transparent-film-absorbing substrate system. J Opt Soc Am 62:1188–1194
Seraphin BO (ed) (1976) Optical properties of solids: new developments, North Holland, Amsterdam

Chapter 11
Inverse Problems

The *direct* problem in reflection is the calculation of the reflection amplitudes (and thence the reflectivities and the ellipsometric ratio), given the characteristics of the reflecting profile. *Inverse* or *inversion* problems consist in the estimation of the profile characteristics, given some experimental reflection or transmission data (or both). In general, the wider the range of the experimental data (in polarization, angle, and frequency), the more can be said about the reflecting profile. But the information is never complete, and if sparse, can be ambiguous. For example: suppose we measure the s or p reflectivity from a homogeneous layer between two other homogeneous media, all three refractive indices being known. What can be said about the thickness of this layer? Only that it has one of an infinity of possible values, since the reflectivity is periodic in the thickness (see (2.66) and Fig. 2.6). One measurement does not guarantee the evaluation of one parameter, even if it is the only unknown in the model.

There can also be measurements which give no profile information whatever, serving only to verify experimental accuracy or to calibrate the apparatus. An example is the reflectivity at grazing incidence, this being unity for either polarization, for arbitrary profiles with or without absorption (Sect. 2.3). More surprising is the fact that null reflectivity at any given angle of incidence, from an interface between media of given dielectric constants ε_1 and ε_2, can be produced in a non-denumerable infinity of ways. The prescription is to pick any function $\varepsilon(z)$ which takes the values ε_1 at $z = -\infty$ and ε_2 at $z = +\infty$, and form $q^2(z) = \varepsilon(z)\omega^2/c^2 - K^2$ (K has the usual meaning, being given by $\sqrt{\varepsilon_1}(\omega/c)\sin\theta_1$]. Then the profile

$$\varepsilon_s(z) = \varepsilon(z) + \left(\frac{c}{\omega}\right)^2 \left\{ \frac{q''}{2q} - \frac{3}{4}\left(\frac{q'}{q}\right)^2 \right\} \tag{11.1}$$

will give zero reflection for the s wave. This was noted by Kofink (1947); the result follows from the fact that the Liouville-Green approximations to the s wave-functions both satisfy (6.26), and that ψ_1^+ tends to $e^{iq_1 z}$ as $z \to -\infty$, thus having zero component of the reflected wave (see Sect. 6.2). The analogous result

J. Lekner, *Theory of Reflection*, Springer Series on Atomic, Optical, and Plasma Physics 87, DOI 10.1007/978-3-319-23627-8_11

for the p polarization is obtained from the equation satisfied by $q_b^{-1/2}\mathrm{e}^{i\phi_b}$, $\phi_b' = q_b$, where q_b is given by (6.88).

After these cautionary notes we will examine some inverse problems relating to reflection, beginning with restricted and thus simple examples, and progressing to more general results. References to related inverse problems in other fields are given at the end of the chapter.

11.1 Reflection at a Sharp Boundary

The case of a sharp boundary between two homogeneous non-absorbing media is straightforward: we have

$$R_s = \left(\frac{q_1 - q_2}{q_1 + q_2}\right)^2, \quad R_p = \left(\frac{Q_1 - Q_2}{Q_1 + Q_2}\right)^2, \tag{11.2}$$

and thus

$$\frac{q_2}{q_1} = \frac{1 \pm R_s^{1/2}}{1 \mp R_s^{1/2}}, \quad \frac{Q_2}{Q_1} = \frac{\varepsilon_1}{\varepsilon_2}\frac{q_2}{q_1} = \frac{1 \pm R_p^{1/2}}{1 \mp R_p^{1/2}}, \tag{11.3}$$

with the upper signs to be taken for $q_2 > q_1$ and $Q_2 > Q_1$, respectively. From the wavenumber ratios one can extract the dielectric constant ratio $\varepsilon_2/\varepsilon_1$ via $q_1^2 = \varepsilon_1(\omega^2/c^2) - K^2 = (\omega^2/c^2)\varepsilon_1 \cos^2\theta_1$, $q_2^2 = \varepsilon_2(\omega^2/c^2) - K^2 = (\omega^2/c^2)(\varepsilon_2 - \varepsilon_1 \sin^2\theta_1)$. This gives, for example,

$$\frac{\varepsilon_2}{\varepsilon_1} = \sin^2\theta_1 + \cos^2\theta_1 \left(\frac{1 \pm R_s^{1/2}}{1 \mp R_s^{1/2}}\right)^2. \tag{11.4}$$

The ellipsometric ratio r_p/r_s moves on the real axis from $+1$ at normal incidence to -1 at grazing incidence, passing through the origin at $\theta_B = \arctan(\varepsilon_2/\varepsilon_1)^{1/2}$, from which the ratio of the refractive indices can be obtained.

When the second medium is absorbing, $\varepsilon_2 = \varepsilon_r + i\varepsilon_i$, the wavevector normal components q_2 and Q_2 are also complex, and

$$R_s = \frac{(q_1 - q_r)^2 + q_i^2}{(q_1 + q_r)^2 + q_i^2}, \qquad R_p = \frac{(Q_1 - Q_r)^2 + Q_i^2}{(Q_1 + Q_r)^2 + Q_i^2}, \tag{11.5}$$

with q_r, q_i, Q_r, Q_i being given by (10.11), (10.12) and (10.21). Explicit inversion of (11.5), with the real and imaginary parts ε_r and ε_i of the dielectric function expressed in terms of R_s, R_p and the angle of incidence, is given in Lekner (1997). As may be expected, no information is gained at normal incidence where $R_s = R_p$,

and at glancing incidence when both tend to unity. Mathematically, this manifests itself as a zero/zero instability. For a sharp interface (with both media isotropic) it is also true for all ε_r and ε_i that $R_s^2 = R_p$ at 45° angle of incidence, as may be verified from (11.2). This again leads to a zero/zero instability. Detailed error analysis shows that measurement at quite large angles of incidence (between 60° and 80°) minimises the error multipliers, but that even then the error multipliers are large for both ε_r and ε_i when the medium is strongly absorbing, for example for metals. Potter (1969) has developed an inversion procedure based on the values of R_p/R_s and θ_1 at the minimum of R_p/R_s, this angle of incidence being known as the pseudo-Brewster angle.

A simple explicit inversion is possible for the real and imaginary parts of ε_2 in terms of the *ellipsometric ratio*, variously written as (Vašiček 1960; Aspnes 1976)

$$\frac{r_p}{r_s} = \rho = \tan\psi\, \mathrm{e}^{i\Delta}, \tag{11.6}$$

(The phase difference $\Delta = \delta_p - \delta_s$ is shown in Fig. 20.5 for a glass-air interface.) The ellipsometric ratio is given by

$$\rho = -\frac{Q_1 - Q_2}{Q_1 + Q_2} \cdot \frac{q_1 + q_2}{q_1 - q_2} = \frac{q_1 q_2 - K^2}{q_1 q_2 + K^2}, \tag{11.7}$$

where we have used $Q_1 = q_1/\varepsilon_1$, $Q_2 = q_2/\varepsilon_2$ and

$$\varepsilon_2 q_1^2 - \varepsilon_1 q_2^2 = (\varepsilon_1 - \varepsilon_2)K^2. \tag{11.8}$$

Thus

$$\frac{1+\rho}{1-\rho} = \frac{q_1 q_2}{K^2}, \quad \left(\frac{1+\rho}{1-\rho}\right)^2 = \frac{\varepsilon_2/\varepsilon_1 - \sin^2\theta_1}{\sin^2\theta_1 \tan^2\theta_1}, \tag{11.9}$$

and therefore

$$\frac{\varepsilon_2}{\varepsilon_1} = \sin^2\theta_1 + \sin^2\theta_1 \tan^2\theta_1 \left(\frac{1+\rho}{1-\rho}\right)^2. \tag{11.10}$$

This equation gives the real and imaginary parts of ε_2 in terms of the real and imaginary parts of $(1+\rho)^2/(1-\rho)^2$. If we write $\rho = \rho_r + i\rho_i$, then

$$\frac{\varepsilon_r}{\varepsilon_1} = \sin^2\theta_1 + \sin^2\theta_1 \tan^2\theta_1 \frac{(1-\rho_r^2-\rho_i^2)^2 - 4\rho_i^2}{\left[(1-\rho_r)^2 + \rho_i^2\right]^2}, \tag{11.11}$$

$$\frac{\varepsilon_i}{\varepsilon_1} = \sin^2\theta_1 + \tan^2\theta_1 \frac{4(1-\rho_r^2-\rho_i^2)\rho_i}{\left[(1-\rho_r)^2+\rho_i^2\right]^2}. \tag{11.12}$$

In terms of the ellipsometric angles ψ and Δ, these formulae read

$$\frac{\varepsilon_r}{\varepsilon_1} = \sin^2\theta_1 + \sin^2\theta_1 \tan^2\theta_1 \frac{\cos^2 2\psi - \sin^2 2\psi \sin^2\Delta}{(1-\sin 2\psi\cos\Delta)^2}, \tag{11.13}$$

$$\frac{\varepsilon_i}{\varepsilon_1} = \sin^2\theta_1 + \tan^2\theta_1 \frac{\sin 4\psi \sin\Delta}{(1-\sin 2\psi\cos\Delta)^2}. \tag{11.14}$$

The ψ, Δ angle representation is ambiguous without the specification of the range of one of them; we take $0 \le \psi \le \pi/2$, in which case $\tan\psi = |\rho|$. From (11.6) and (11.7) we have

$$\rho_r = \tan\psi\cos\Delta = \frac{q_1^2(q_r^2+q_i^2) - K^4}{(q_1 q_r + K^2)^2 + q_1^2 q_i^2}, \tag{11.15}$$

$$\rho_i = \tan\psi\sin\Delta = \frac{2q_1 q_i K^2}{(q_1 q_r + K^2)^2 + q_1^2 q_i^2}. \tag{11.16}$$

We showed in Sect. 10.1 that r_p/r_s always lies within the upper semicircle of unit radius (for a sharp boundary between a dielectric and an absorbing medium). Thus $\psi \le \pi/4$; the value $\pi/4$ is attained at normal incidence and at grazing incidence. The angle Δ increases from 0 at normal incidence to π at grazing incidence. According to Aspnes (1976), the attainable precision in ψ and Δ is about $\delta\psi \approx \delta\Delta/2 \approx 1$ millidegree. Figure 11.1 shows an example of the fractional errors $\Delta\varepsilon_r/\varepsilon_r$ and $\Delta\varepsilon_i/\varepsilon_i$ in ε_r and ε_i as a function of the angle at which measurement is carried out, assuming the larger random scatter of up to 0.01° in ψ and 0.02° in Δ, at all angles of incidence. The errors diverge at normal incidence, and also at glancing incidence, since there the ratio $\rho = r_p/r_s$ always takes the values 1 and -1, respectively. The accuracy in ε_r and ε_i is best near the the principal angle θ_p, at which $\rho_r = 0$. The principal angle for an absorber is determined by a cubic equation, given in the Note at the end of this chapter. Only the region near θ_p is shown in Fig. 11.1.

The above ellipsometric extraction of ε_r and ε_i, carried out over a range of frequencies, gives $\varepsilon_r(\omega)$, $\varepsilon_i(\omega)$, or $n_r(\omega)$ and $n_i(\omega)$. Another method is to measure the reflectivity at normal incidence (given in terms of n_r and n_i by (10.17)). This determines the modulus $|r(\omega)|$ of the reflection amplitude $r = |r|e^{i\delta}$. The phase $\delta(\omega)$ is found from a Kramers-Krönig relation between $\ln|r|$ and δ (extrapolation of measured reflectivity data is required), and finally $n_r(\omega)$ and $n_i(\omega)$ are found from $r = (n_1 - n_r - in_i)/(n_1 + n_r + in_i)$. Details of this procedure are given by Wooten (1972, Chap. 6 and Appendix G).

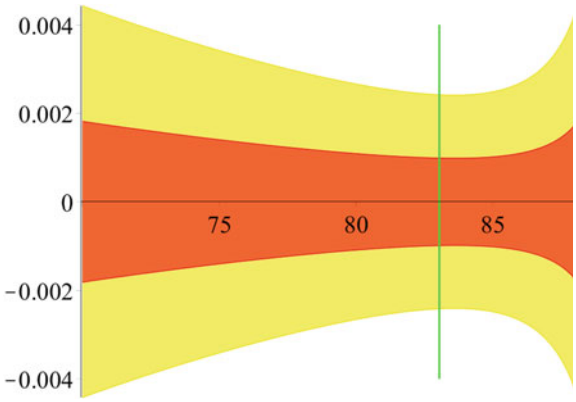


Fig. 11.1 Fractional errors $\Delta\varepsilon_r/\varepsilon_r$ (*dark shading*) and $\Delta\varepsilon_i/\varepsilon_i$ (*light shading*) in ε_r and ε_i values deduced from (11.13) and (11.14), on the assumption of uniformly distributed random errors of up to 0.01° in ψ and 0.02° in Δ. The "true" values of ψ and Δ are calculated from (11.15) and (11.16), using the bulk Al parameters at 633 nm, $\varepsilon_r = -60.56$, $\varepsilon_i = 24.86$ (as in Figs. 10.1 and 10.2), for which $\theta_p \approx 83°$, indicated by the *vertical line*

11.2 Homogeneous Film Between Like Media

An explicit inversion of *reflection* ellipsometric data for a homogeneous nonabsorbing film has been discussed in Sect. 9.8, and also the inversion of combined *reflection and transmission* ellipsometric data. Where the media bounding the film have the same dielectric constant the solution was given by Azzam (1983); the general case is solved in Lekner (1994a). Both reflection and transmission ellipsometric coefficients are used. Here we shall consider a homogeneous film of thickness Δz and dielectric constant $\varepsilon_r + i\varepsilon_i$, embedded in a medium of dielectric constant ε_0. From (2.58), (2.59) and (2.70), (2.71) we have

$$\rho = \frac{r_p}{r_s} = \frac{p}{s} \cdot \frac{1 - s^2 \mathrm{e}^{2iq\Delta z}}{1 - p^2 \mathrm{e}^{2iq\Delta z}}, \tag{11.17}$$

$$\tau = \frac{t_p}{t_s} = \frac{1 - p^2}{1 - s^2} \cdot \frac{1 - s^2 \mathrm{e}^{2iq\Delta z}}{1 - p^2 \mathrm{e}^{2iq\Delta z}}, \tag{11.18}$$

where

$$s = \frac{q_0 - q}{q_0 + q}, \qquad -p = \frac{Q_0 - Q}{Q_0 + Q}, \tag{11.19}$$

are the s and p reflection amplitudes at a step between media with dielectric constants ε_0 and ε, q_0 and q being the corresponding real and complex normal components of the wavevector. From (11.17) and (11.18) we see that the ratio of the

transmission and reflection ellipsometric ratios is independent of the thickness of the layer:

$$\frac{\tau}{\rho} = \frac{s}{p}\frac{1-p^2}{1-s^2}. \tag{11.20}$$

This equation is to be solved for $\varepsilon = \varepsilon_r + i\varepsilon_i$. The identity (Azzam 1979)

$$p = \frac{s(\cos 2\theta - s)}{1 - s\cos 2\theta}, \tag{11.21}$$

(which may be verified by solving (11.21) for $\cos 2\theta$ and using (11.19) and $q^2 = (\omega/c)^2(\varepsilon - \varepsilon_0 \sin^2\theta)$) serves to eliminate p from (11.20), which reduces to a quadratic for s in terms of the measured θ and τ/ρ:

$$s^2 - 2\sigma s + 1 = 0, \qquad \sigma = \frac{\cos 2\theta - \left(\frac{\tau}{2\rho}\right)(1+\cos^2 2\theta)}{1 - \left(\frac{\tau}{\rho}\right)\cos 2\theta}. \tag{11.22}$$

This has the solutions

$$s_\pm = \sigma \pm (\sigma^2 - 1)^{1/2}, \tag{11.23}$$

of which one needs to select the root with $|s| < 1$. That one and only one such root exists can be seen by writing $\sigma = \cosh\zeta$: then $s_\pm = \mathrm{e}^{\pm\zeta}$. In general, ζ is complex, and one takes s_+ or s_- according as $\mathrm{Re}\,\zeta < 0$ or $\mathrm{Re}\,\zeta > 0$. (In the special case where $\mathrm{Re}\,\zeta = 0$ and σ is equal to $\cos(\mathrm{Im}\,\zeta)$, both roots would have unit modulus. But for the absorbing media $|s|^2$ cannot be unity except at grazing incidence, where $q_0 \to 0$.) Having obtained the complex value of s, the dielectric function may be found from

$$\frac{\varepsilon}{\varepsilon_0} = \sin^2\theta + \cos^2\theta\left(\frac{1-s}{1+s}\right)^2. \tag{11.24}$$

This relation is obtained by squaring $(1-s)/(1+s) = q/q_0$; compare (11.4). ((11.24), (11.20) and (11.21) are together equivalent to (9.55) in Sect. 9.8, in which the inversion of ρ and τ when $n_1 \neq n_2$ is discussed.) It then remains to evaluate the thickness Δz. Since s and p (from the identity (11.21)) are now known, $\mathrm{e}^{2iq\Delta z}$ may be found from ρ or τ as given by (11.17) or (11.18):

$$\mathrm{e}^{2iq\Delta z} = \frac{1}{sp}\cdot\frac{p-\rho s}{s-\rho p}, \tag{11.25}$$

$$e^{2iq\Delta z} = \frac{1 - p^2 - \tau(1 - s^2)}{s^2(1 - p^2) - \tau p^2(1 - s^2)}. \tag{11.26}$$

If the right-hand side of either of these is written as $e^{2i(\alpha + i\beta)}$, then

$$q_r \Delta z = \alpha + m\pi, \qquad q_i \Delta z = \beta, \tag{11.27}$$

where m is zero or a positive or negative integer, and q_r and q_i have been found from (10.11) and (10.12). Only one of (11.25) and (11.26) and only the second relation in (11.27) need be used to obtain Δz; the others provide a check on the accuracy. Azzam (1983) gives an example of the application of this technique to the determination of the thickness and optical parameters of a thin gold foil.

11.3 Inversion of Transmission Ellipsometric Data for a Homogeneous Nonabsorbing Layer

We wish to find the (real) dielectric constant ε and the thickness Δz of a layer, from the real and imaginary parts of the transmission ellipsometry ratio t_p/t_s. The full solution, complete with analysis of the effect of measurement errors, is given in Lekner (1994a). We shall just give the essence of the inversion method. From Sect. 2.4 or directly from (9.46) of Sect. 9.8 we have

$$\frac{t_p}{t_s} = \frac{n_1(1 - p_1)(1 - p_2)(1 + s_1 s_2 Z)}{n_2(1 + s_1)(1 + s_2)(1 + p_1 p_2 Z)}, \tag{11.28}$$

where p_1, p_2, s_1 and s_2 are the Fresnel reflection amplitudes at the $n_1|n$ and $n|n_2$ interfaces, given in (9.47), and $Z = e^{2iq\Delta z}$. The experimentally determined ratio $t = (n_2/n_1)(t_p/t_s) = x + iy$ can be written as

$$t = f\frac{1 + SZ}{1 + PZ}, \quad P = p_1 p_2, \quad S = s_1 s_2, \quad f = \frac{(1 - p_1)(1 - p_2)}{(1 + s_1)(1 + s_2)}. \tag{11.29}$$

For nonabsorbing layers $Z = e^{2iq\Delta z}$ lies on the unit circle (we are excluding the possibility of total external reflection, with imaginary q). Since $t = f(1 + SZ)/(1 + PZ)$ is linear in Z, we can solve for Z and eliminate it by using $ZZ^* = 1$. After some algebraic manipulation this reduces to a quadratic in q^2, or equivalently to a quadratic in ε, because $\varepsilon(\omega/c)^2 = q^2 + K^2$. When this is solved there are two roots, each of which in turn will give two values of Z and therefore of Δz:

$$\exp(2iq\Delta z) = Z = \frac{f - t}{tP - fS}. \tag{11.30}$$

To distinguish the physical root from the nonphysical root would require some knowledge about the likely values of ε and Δz. This extra knowledge is not essential, however: for measurements at two or more angles of incidence, the physical roots will agree (to within experimental error), the nonphysical roots will not. Figure 1 of Lekner (1994a) illustrates the behaviour of the physical and nonphysical ε and Δz, and of the effects of experimental errors. In the limit of thin films, with $q\Delta z$ small compared to unity ($q\Delta z = (2\pi\Delta z/\lambda)(n^2 - n_1^2 \sin^2\theta_1)^{\frac{1}{2}}$, where λ is the vacuum wavelength), transmission ellipsometry does not give ε and Δz separately, but only the integral invariant for the thin film,

$$I_1 = \frac{(\varepsilon_1 - \varepsilon)(\varepsilon - \varepsilon_2)}{\varepsilon}\Delta z. \tag{11.31}$$

This invariant is the same as enters into the first-order expression for the ellipsometric ratio r_p/r_s.

11.4 Inversion of Reflection Ellipsometric Data for a Homogeneous Nonabsorbing Layer

Again we wish to find the real dielectric constant ε and the layer thickness Δz, this time from the real part and the absolute square of the ratio of the reflection amplitudes,

$$\rho = \frac{r_p}{r_s} = \frac{p_1 + p_2 Z}{1 + p_1 p_2 Z}\frac{1 + s_1 s_2 Z}{s_1 + s_2 Z}. \tag{11.32}$$

It is shown in Lekner (1994b) that both $|\rho|^2$ and $\mathrm{Re}(\rho)$ have the form (quadratic in $\cos 2q\Delta z$)/(quadratic in $\cos 2q\Delta z$). Thus when $|\rho|^2$ and $\mathrm{Re}(\rho)$ are determined experimentally, $\cos 2q\Delta z$ satisfies two quadratic equations. The condition that the two quadratics share a common root implies a relation between the coefficients of the two quadratics. After removal of common factors (which reduce the degree in the unknown q by 21), one is left with a quintic in q^2, or equivalently a quintic in ε. This factors to a quadratic and a cubic when $\varepsilon_1 = \varepsilon_2$, previously discussed in Lekner (1990). The quintic in ε, which is at the same time a quartic plane curve in $x = \mathrm{Re}(\rho)$ and $y = \mathrm{Im}(\rho)$, may be conveniently expressed in terms of dimensionless variables u, v and w, defined by

$$q^2 = q_1 q_2 u, \quad K^2 = q_1 q_2 v, \quad q_2 = q_1 w. \tag{11.33}$$

The variable u contains the unknown ε; v and w depend on the angle of incidence and on the ε_1 and ε_2 values. The equation, with $r^2 = x^2 + y^2 = |\rho|^2$, reads

$$a\left[r^2-(\rho_+ +\rho_-)x+\rho_+\rho_-\right]\left[r^2-2x_2x+x_2^2\right] +4v^2(1+vw)(u-w)^4(uw-v^2)y^2=0 \tag{11.34}$$

All parameters in (11.34) are dimensionless; $\rho_\pm=\rho(Z=\pm1)$, a and x_2 (which gives the location of an isolated point of the quartic) may be found in Lekner (1994b). Once the dielectric constant ε has been determined, the thickness of the layer may also be extracted, *except* when $Z=\exp(2iq\Delta z)$ is close to unity. Just as in the previous section, only the invariant I_1 of (11.31) may then be extracted.

11.5 Synthesis of a Profile from *r* as a Function of Wavenumber

A general solution to the inverse reflection problem (or the inverse scattering problem in quantum mechanics) has been found by Gelfand and Levitan (1951/1955) and others. More references are given at the end of this chapter; here we shall give only a brief description of the theory, discuss some results which follow from it, and give an approximate but explicit solution which is simple enough to have practical application.

The inversion procedure assumes the knowledge of the reflection amplitude as a function of wavenumber (the latter ranging from zero to infinity) and in addition, coefficients relating to any bound states that may exist. Since experiments generally give $|r|^2$, not r, and that over a finite range of wavenumbers, we prefer the term *synthesis* to *inversion* in this case: a model reflection amplitude, complete with phase, can be constructed to have some desired properties, such as high reflectance in one wavenumber region and low reflectance in another region. The theory then gives a procedure for synthesising the refractive index profile which will give the desired reflectance. In the general theory, one constructs an integral equation from the Fourier transform of the reflection amplitude (analytically continued to negative wavenumbers). The solution of the integral equation then gives the refractive index profile. In special cases an explicit solution can be found, for example when the reflection amplitude is a rational function of the wavenumber (Kay 1960; Jordan 1980). Another special case is the construction of an infinite set of profiles which do not reflect the s wave, at fixed frequency but *for any angle of incidence* (Kay and Moses 1956). The simplest of these is the sech^2 profile discussed in Sect. 4.3, for certain special values of its parameters. We saw there that for profile

$$\varepsilon(z)=\varepsilon_0+\Delta\varepsilon\,\mathrm{sech}^2(z/a), \tag{11.35}$$

the s reflection amplitude was a function of two dimensionless parameters $\alpha=\Delta\varepsilon(\omega a/c)^2$ and $\beta=q_0a=\varepsilon_0^{1/2}(\omega a/c)\cos\theta$, θ being the angle of incidence. When $\alpha\geq-1/4$,

$$R_s = \frac{\cos^2\left[\frac{\pi}{2}(1+4\alpha)^{1/2}\right]}{\cos^2\left[\frac{\pi}{2}(1+4\alpha)^{1/2}\right] + \sinh^2 \pi\beta}. \tag{11.36}$$

This is zero when $\alpha = m(m+1)$, m an integer, for any θ, the angle of incidence appearing only through β.

We now give an approximate solution of a synthesis problem, due to Hirsch (1979). Only the essence of the method will be given, since we then show how a more general result can be obtained in a simpler way. The problem is that of constructing a refractive index profile $n(z) = \sqrt{\varepsilon(z)}$ so as to give a desired reflection amplitude $r(k_1)$ at normal incidence. Here the wave is incident from a medium of unit refractive index, and transmitted into a medium of refractive index n_2. The respective wavenumbers are $k_1 = \omega/c$ and $k_2 = n_2(\omega/c)$. The E field satisfies

$$\frac{\mathrm{d}^2E}{\mathrm{d}z^2} + n^2k_1^2E = 0, \quad \mathrm{e}^{ik_1z} + r\,\mathrm{e}^{-ik_1z} \leftarrow E \rightarrow t\,\mathrm{e}^{ik_2z}. \tag{11.37}$$

The geometric path increment $\mathrm{d}z$ is replaced by the optical path increment $\mathrm{d}x = n\mathrm{d}z$. Also E is replaced by the function $w = n^{1/2}E$. The resulting equation is

$$w'' + \left[k_1^2 - U(x)\right]w = 0, \quad U(x) = \frac{1}{2}\frac{n''}{n} - \frac{1}{4}\left(\frac{n'}{n}\right)^2. \tag{11.38}$$

(Throughout this section primes will denote differentiation with respect to x.) Now E and w are proportional as z and x tend to $-\infty$; if further the refractive index is taken to be equal to 1 for $z \leq 0$ and x is defined by

$$x = \int_0^z \mathrm{d}\zeta\, n(\zeta), \tag{11.39}$$

then x and z are equal for $z \leq 0$ and $r(k_1)$ is the reflection amplitude for $w(x)$ as well as for $E(z)$. (What we have just done is ensure that the phase of r is the same for both.) From $r(k_1)$ and its analytic continuation to negative k_1 via $r(-k_1) = r^*(k_1)$, we form the Fourier transform

$$F(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}k_1 r(k_1)\mathrm{e}^{-ik_1x}. \tag{11.40}$$

The first two terms of $U(x)$ in a series expansion formally equivalent to the Gelfand-Levitan equation are (Moses 1956)

$$U_a(x) = -2\frac{\mathrm{d}}{\mathrm{d}x}F(2x) + 4F^2(2x). \tag{11.41}$$

The second relation in (11.38) may be written as

$$(n^{1/2})'' = n^{1/2}U. \tag{11.42}$$

When U is approximated by U_a, this differential equation for $n^{1/2}(x)$ may be integrated; the solution incorporating the boundary condition $n \to 1$ at $-\infty$ and $n \to$ finite constant at $+\infty$ is then

$$n_a(x) = \exp\left\{-2\int_{-\infty}^{2x} \mathrm{d}y\ F(y)\right\}. \tag{11.43}$$

We note that $F(x)$ is real, and also that since r is the inverse Fourier transform of F,

$$r(k_1) = \int_{-\infty}^{\infty} \mathrm{d}x\, F(x)\, \mathrm{e}^{ik_1 x}, \tag{11.44}$$

the final value of the refractive index is approximately

$$n_a(\infty) = \exp\{-2r(0)\}. \tag{11.45}$$

As an example of these relations, consider the application of the inverse of (11.43),

$$F(2x) \approx -\frac{1}{4}\frac{n'(x)}{n(x)}. \tag{11.46}$$

We will use (11.46) to obtain the approximate reflection amplitude for the Rayleigh profile studied in Sect. 2.5, for which n^{-1} is linear in z:

$$n^{-1}(x) \equiv \eta(z) = \eta_1 + (\Delta\eta/\Delta z)z, \qquad 0 \le z \le \Delta z. \tag{11.47}$$

For this profile we have, in the interval $0 \le z \le \Delta z$,

$$x = \frac{\Delta z}{\Delta \eta}\ln\eta, \quad \frac{n'(x)}{n(x)} = -\frac{\Delta\eta}{\Delta z}, \quad F(2x) = \frac{1}{4}\frac{\Delta\eta}{\Delta z}, \tag{11.48}$$

and so from (11.44) the approximate reflection amplitude is

$$r_a(k_1) = \frac{1}{2}\ln\frac{\eta_2}{\eta_1} e^{ik_1\Delta x}\frac{\sin k_1\Delta x}{k_1\Delta x}, \quad \Delta x = \frac{\Delta z}{\Delta\eta}\ln\left(\frac{\eta_2}{\eta_1}\right). \tag{11.49}$$

This is precisely the Rayleigh or weak reflection approximation result obtained in Sect. 5.8 for this profile (see (5.98) and Fig. 5.4).

11.6 Inversion of the Rayleigh Approximation

We will now show that this surprising accord is not accidental: the approximate solution of the synthesis problem given above is identical to the result obtained by inverting the Rayleigh approximation for the reflection amplitude. The latter was given in Sect. 5.7; at normal incidence (5.85) and (5.86) reduce to

$$r \approx -\int_{-\infty}^{\infty} d\phi \frac{dn/d\phi}{2n} e^{2i\phi}, \quad \phi = \int^{z} d\zeta\, k(\zeta). \tag{11.50}$$

Thus the Rayleigh approximation gives the reflection amplitude at normal incidence as the Fourier transform in the ϕ variable of the logarithmic derivative of $n^{1/2}$. To keep common notation with the Hirsch inversion we set $d\phi = kdz = k_1 dx$. Then (11.50) reads

$$r(k_1) \approx -\int_{-\infty}^{\infty} dx \frac{n'}{2n} e^{2ik_1 x}, \tag{11.51}$$

and has the inverse

$$-\frac{n'}{2n} \approx \frac{1}{2\pi}\int_{-\infty}^{\infty} d\,k_1\, r(k_1)\, e^{-2ik_1 x} \equiv F(2x), \tag{11.52}$$

$$n(x) \approx n_1 \exp\left\{-2\int_{-\infty}^{2x} dy\, F(y)\right\}. \tag{11.53}$$

This equation is slightly more general than (11.50), in that $n_1 = 1$ has not been assumed. Discussion of the validity of the Rayleigh approximation, and error

bounds on the resulting reflection amplitudes, may be found in Sect. 5.7. In general it is expected to break down when the reflection is strong.

Inversion of the Rayleigh approximation reflection amplitudes is possible for all angles of incidence for both polarizations. In the s wave case we have, from (5.85) $\mathrm{d}\phi = q_1\,\mathrm{d}x$ or $x = q_1^{-1}\int^z \mathrm{d}\zeta\, k(\zeta)$,

$$r_s(q_1) \approx -\int_{-\infty}^{\infty} \mathrm{d}x \frac{q'}{2q} \mathrm{e}^{2iq_1x}, \tag{11.54}$$

which has the Fourier inverse

$$-\frac{q'}{2q} \approx \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}q_1 \mathrm{e}^{-2iq_1x} r_s(q_1) \equiv F_s(2x). \tag{11.55}$$

Thus

$$q(x) \approx q_1 \exp\left(-2\int_{-\infty}^{2x} \mathrm{d}y\, F_s(y)\right), \tag{11.56}$$

or

$$\frac{\varepsilon(x)}{\varepsilon_1} \approx \sin^2\theta_1 + \cos^2\theta_1 \exp\left(-4\int_{-\infty}^{2x} \mathrm{d}y\, F_s(y)\right) \tag{11.57}$$

(note the formal similarity with (11.4) and (11.24)). The p wave (5.86) is inverted using the same variable x:

$$r_p(q_1) \approx \int_{-\infty}^{\infty} \mathrm{d}x \frac{Q'}{2Q} \mathrm{e}^{2iq_1x}, \tag{11.58}$$

$$\frac{Q'}{2Q} \approx \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}q_1 \mathrm{e}^{2iq_1x} r_p(q_1) \equiv F_p(2x), \tag{11.59}$$

$$Q(x) \approx Q_1 \exp\left(2\int_{-\infty}^{2x} \mathrm{d}y\, F_p(y)\right). \tag{11.60}$$

Since $Q = q/\varepsilon$ and $(cq/\omega)^2 = \varepsilon - \varepsilon_1 \sin^2\theta_1$, (11.60) gives a quadratic for $\varepsilon(x)$. We take the root which agrees with (11.57) at normal incidence:

$$\frac{\varepsilon(x)}{\varepsilon_1} \approx \frac{\exp\left(-4\int_{-\infty}^{2x} \mathrm{d}y\ F_p(y)\right)}{2\cos^2\theta_1} \times \left\{1 + \left[1 - \sin^2 2\theta_1 \exp\left(4\int_{-\infty}^{2x} \mathrm{d}y\ F_p(y)\right)\right]^{1/2}\right\}. \tag{11.61}$$

A measure of the accuracy of this inversion may be obtained by calculating $\varepsilon_2/\varepsilon_1$ from (11.57) and (11.61), letting x tend to $+\infty$, and using

$$\int_{-\infty}^{\infty} \mathrm{d}x\ F_s(x) = r_s(0), \qquad \int_{-\infty}^{\infty} \mathrm{d}x\ F_p(x) = r_p(0). \tag{11.62}$$

The long wave limits of the Rayleigh approximations for r_s and r_p are given by (5.87); when these are substituted in (11.57) and (11.61) the right-hand sides are equal to $\varepsilon_2/\varepsilon_1$, at all angles. The true long wave limits, given by (5.88), do not give agreement between the left and right hand sides. For example, at normal incidence where $r_s(0) = r_p(0) = (n_1 - n_2)/(n_1 + n_2)$, (11.53), (11.57) and (11.61) both give

$$\frac{n_2}{n_1} \approx \exp 2\left(\frac{n_2 - n_1}{n_2 + n_1}\right), \tag{11.63}$$

which has an error of the third order in $(n_2 - n_1)/(n_2 + n_1)$.

The inversion formulae (11.57) and (11.61) give $\varepsilon(x)$, not the required $\varepsilon(z)$, and thus need to be complemented by a functional relation between the physical coordinate z and the "optical" coordinate x. This is obtained from the given reflection amplitude r_s via its Fourier transform F_s by integrating $q\mathrm{d}z = q_1\mathrm{d}x$ using (11.56):

$$z(x) = \int_0^x \mathrm{d}x_1 \exp\left[2\int_{-\infty}^{2x_1} \mathrm{d}x_2\ F_s(x_2)\right]. \tag{11.64}$$

11.7 Principal Angle of an Absorber

From (11.7) we find that the zero of $\rho_r = \mathrm{Re}(r_p/r_s)$ is given by

$$q_1^2\left(q_r^2 + q_i^2\right) = K^4. \tag{11.65}$$

Since $q^2 = (q_r + iq_i)^2 = (\varepsilon_r + i\varepsilon_i)(\omega/c)^2 - K^2$, we have

$$q_r^2 - q_i^2 = \varepsilon_r(\omega/c)^2 - K^2, \quad 2q_rq_i = \varepsilon_i(\omega/c)^2. \tag{11.66}$$

We square (11.65), and use an identity following from (11.66):

$$\left(q_r^2+q_i^2\right)^2=\left[\varepsilon_r(\omega/c)^2-K^2\right]^2+\varepsilon_i^2(\omega/c)^4 \tag{11.67}$$

The result is a cubic equation for K^2, or equivalently, a cubic for $S=\sin^2\theta_p$:

$$\varepsilon_r^2+\varepsilon_i^2-2\left(\varepsilon_r^2+\varepsilon_i^2+\varepsilon_1\varepsilon_r\right)S+\left(\varepsilon_r^2+\varepsilon_i^2+4\varepsilon_1\varepsilon_r+\varepsilon_1^2\right)S^2-2\varepsilon_1(\varepsilon_r+\varepsilon_1)S^3=0. \tag{11.68}$$

Equation (11.68) was obtained by Humphreys-Owen (1961), who also gives a cubic equation for the angle at which R_p, given by (11.5), is minimum. This is

$$\left(\varepsilon_r^2+\varepsilon_i^2\right)^2-2\left(\varepsilon_r^2+\varepsilon_i^2\right)^2S+\left(\varepsilon_r^2+\varepsilon_i^2\right)\left(\varepsilon_r^2+\varepsilon_i^2-3\varepsilon_1^2\right)S^2+2\varepsilon_1^2\left(\varepsilon_r^2+\varepsilon_i^2+\varepsilon_1\varepsilon_r\right)S^3=0 \tag{11.69}$$

When ε_i is set to zero these cubic equations share a common root, which gives the Brewster angle:

$$S_B=\sin^2\theta_B=\frac{\varepsilon_r}{\varepsilon_1+\varepsilon_r}. \tag{11.70}$$

References

Aspnes DE (1976) Spectroscopic ellipsometry of solids. In: Seraphin BO (ed) Optical properties of solids: new developments. North Holland, Amsterdam (Chapter 15)

Azzam RMA (1979) Direct relation between Fresnel's interface reflection coefficients for the parallel and perpendicular polarisations. J Opt Soc Am 69:1007–1016

Azzam RMA (1983) Ellipsometry of unsupported and embedded thin films. J de Phys C 10:67–70

Gelfand IM, Levitan BM (1951/1955) On the determination of a differential equation by its spectral function. Am Math Soc Tran (Ser. 2) 1:253–304

Hirsch J (1979) An analytic solution to the synthesis problem for dielectric thin-film layers. Optica. Acta 26:1273–1279

Humphreys-Owen SPF (1961) Comparison of reflection methods of measuring optical constants, and proposal for new methods based on the Brewster angle. Proc Phys Soc 77:949–957

Jordan A K (1980) Inverse scattering theory: exact and approximate solutions. In: DeSanto JA, Sáenz AW, Zachary WW (eds) Mathematical methods and applications of scattering theory. Springer Lecture Notes in Physics No. 130

Kay I, Moses HE (1956) Reflectionless transmission through dielectrics and scattering potentials. J Appl Phys 27:1503–1508

Kay I (1960) The inverse scattering problem when the reflection coefficient is a rational function. Commun Pure Appl Math 13:371–393

Kofink W (1947) Reflexion elektromagnetischer Wellen an einer inhomogenen Schicht. Ann Phys 1:119–132

Lekner J (1990) Analytic inversion of ellipsometric data for an unsupported nonabsorbing uniform layer. J Opt Soc Am A: 7:1875–1877

Lekner J (1994a) Inversion of transmission ellipsometric data for transparent films. Appl Opt 33:5108–5110
Lekner J (1994b) Inversion of reflection ellipsometric data. Appl Opt 33:5159–5165
Lekner J (1997) Inversion of the *s* and *p* reflectances of absorbing media. J Opt Soc Am A 14:1355–1358
Moses HE (1956) Calculation of the scattering potential from reflection coefficients. Phys Rev 102:559–567
Potter RF (1969) Pseudo-Brewster angle technique for determining optical constants. In: Nudelman S, Mitra SS (eds) Optical properties of solids. Plenum, New York
Vašiček A (1960) Optics of thin films. North-Holland, Amsterdam (Section 5.2)
Wooten F (1972) Optical properties of solids. Academic Press, New York

Further Readings

Introductory and review articles on inverse problems
Dyson FJ (1976) Old and new approaches to the inverse scattering problem. In: Lieb EH, Simon B, Wightman AS (eds) Studies in mathematical physics, essays in Honor of Valentine Bargmann. Princeton University Press, Princeton, pp 151–167
Kac M (1966) Can one hear the shape of a drum? Am. Math. Monthly 73(Pt.II):1–23
Keller JB (1976) Inverse problems. Am Math Monthly 83:107–118
Newton RG (1970) Inverse problems in physics. SIAM Rev 12:346–355
Papers and books on the inverse problem of scattering and diffraction theory
Agranovich ZS, Marchenko VA (1963) The inverse problem of scattering theory. Gordon and Breach, New York
Bargmann V (1949a) Remarks on the determination of a central field of force from the elastic scattering phase shifts. Phys Rev 75:301–303
Bargmann V (1949b) On the connection between phase shifts and scattering potential. Rev Mod Phys 21:488–493
Chadan K, Sabatier PC (1977) Inverse problems in quantum scattering theory. Springer, New York
Faddeev LD (1963) The inverse problem in the quantum theory of scattering. J Math Phys 4:72–104
Jost R, Kohn W (1952) Equivalent potentials. Phys Rev 88:382–385
Marchenko VA (1955) The construction of the potential energy from the phases of the scattered waves. Dokl Akad Nauk SSSR 104:695–698 Math. Rev. 17:740 (1956))
Nieto-Vesperinas M (2006) Scattering and diffraction in physical optics, 2nd edn. World Scientific, Singapore
Collections of papers on electromagnetic and optical inverse problems
Baltes HP (ed) (1978) Inverse source problems in optics. Springer, Berlin
Baltes HP (ed) (1980) Inverse scattering problems in optics. Springer, Berlin
Boemer WM, Jordan AK, Kay IW (eds) (1981) IEEE transactions on antennas and propagation. Inverse Methods Electromagnet AP-29(2):185–417
Devaney AJ (ed) (1985) Inverse problems in propagation and scattering. J Opt Soc Am A2:1901–2061 (In relation to Section 11-5, see especially the papers by Landouceur HD and Jordan AK, and by Jaggard DL and Kim Y.)
Another important inversion problem arises in the extraction of the electron density as a function of height from the measured times of travel of nearly monochromatic radio pulses which are reflected from the ionosphere. The problem reduces to that of solving Abel's integral equation, and is related to several inverse problems in mechanics (Keller 1976, quoted above). The ionospheric case is considered in detail by
Budden KG (1961) Radio waves in the ionosphere. Cambridge University Press, Cambridge (Chapter 10)
Budden KG (1985) The propagation of radio waves. Cambridge University Press, Cambridge (Chapter 12)

Chapter 12
Matrix and Numerical Methods

The idea of representing an arbitrary stratification by a series of homogeneous layers goes back at least to Rayleigh (1912). The problem of wave propagation through the transition is solved by matching the wave amplitude and derivative at the boundaries of the uniform layers, and letting the number of layers increase and their thickness decrease. (In the case of finite number of homogeneous layers, such as optical coatings, this limiting process is not required.) Rayleigh carried through the necessary algebra without reference to matrices; Weinstein (1947), Herpin (1947) and Abelès (1950, 1967) have shown how matrix algebra simplifies and systematizes this approach. We will give three versions of the matrix method, of which the last (given in Sect. 12.2) is the closest to that currently in use, but differs from it in having all matrix elements real in the absence of absorption. This last method is the one we use in the remainder of this chapter, and in the next.

12.1 Matrices Relating the Coefficients of Linearly Independent Solutions

We consider the reflection problem for waves satisfying

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}z^2} + q^2\psi = 0, \qquad \mathrm{e}^{iq_a z} + r\,\mathrm{e}^{-iq_a z} \leftarrow \psi \rightarrow t\,\mathrm{e}^{iq_b z}. \tag{12.1}$$

(A change of notation from our usual q_1, q_2 designation of the limiting values of the normal components of the wavevector is required here, since the subscripts $1, 2, \ldots, N, N+1$ will be needed for quantities belonging to the N layers and $N+1$ boundaries.) The function $q^2(z) = (\omega^2/c^2)\left[\varepsilon(z) - \varepsilon_a \sin^2\theta_a\right]$ is either given by or approximated by a series of steps. Figure 12.1 shows the corresponding $\varepsilon(z)$.

Let q_n be the value of q in the interval (z_n, z_{n+1}). The general solution of $\mathrm{d}^2\psi/\mathrm{d}z^2 + q^2\psi = 0$ in this interval may be written as

J. Lekner, *Theory of Reflection*, Springer Series on Atomic, Optical, and Plasma Physics 87, DOI 10.1007/978-3-319-23627-8_12

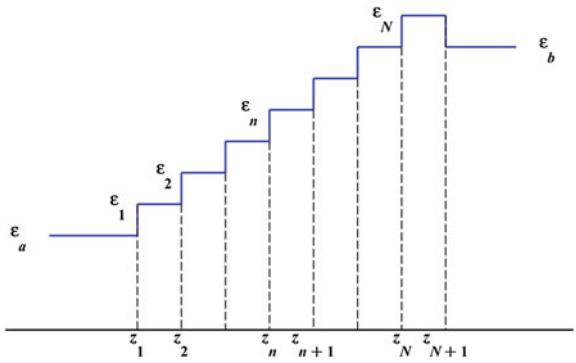


Fig. 12.1 A stack of N homogeneous layers, bounded by media with dielectric constants ε_a and ε_b. The nth layer extends from z_n to z_{n+1} and has dielectric constant ε_n. The layer thicknesses need not be equal

$$\psi_n = \alpha_n \, \mathrm{e}^{iq_n z} + \beta_n \, \mathrm{e}^{-iq_n z}. \tag{12.2}$$

In the interval (z_{n-1}, z_n) the solution is

$$\psi_{n-1} = \alpha_{n-1} \, \mathrm{e}^{iq_{n-1} z} + \beta_{n-1} \, \mathrm{e}^{-iq_{n-1} z}. \tag{12.3}$$

At $z = z_n$ we have $\psi_{n-1} = \psi_n$ and $\mathrm{d}\psi_{n-1}/\mathrm{d}z = \mathrm{d}\psi_n/\mathrm{d}z$; these continuity conditions imply

$$\alpha_{n-1} \, \mathrm{e}^{iq_{n-1} z_n} + \beta_{n-1} \, \mathrm{e}^{-iq_{n-1} z_n} = \alpha_n \, \mathrm{e}^{iq_n z_n} + \beta_n \, \mathrm{e}^{-iq_n z_n}, \tag{12.4}$$

$$q_{n-1}(\alpha_{n-1} \, \mathrm{e}^{iq_{n-1} z_n} - \beta_{n-1} \, \mathrm{e}^{-iq_{n-1} z_n}) = q_n \big(\alpha_n \, \mathrm{e}^{iq_n z_n} - \beta_n \, \mathrm{e}^{-iq_n z_n}\big). \tag{12.5}$$

Solving for α_n and β_n in terms of α_{n-1} and β_{n-1}, we find

$$\begin{pmatrix} \alpha_n \\ \beta_n \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\left(1 + \frac{q_{n-1}}{q_n}\right) \mathrm{e}^{i(q_{n-1} - q_n) z_n} & \frac{1}{2}\left(1 - \frac{q_{n-1}}{q_n}\right) \mathrm{e}^{-i(q_{n-1} + q_n) z_n} \\ \frac{1}{2}\left(1 - \frac{q_{n-1}}{q_n}\right) \mathrm{e}^{i(q_{n-1} + q_n) z_n} & \frac{1}{2}\left(1 + \frac{q_{n-1}}{q_n}\right) \mathrm{e}^{-i(q_{n-1} - q_n) z_n} \end{pmatrix} \begin{pmatrix} \alpha_{n-1} \\ \beta_{n-1} \end{pmatrix} \tag{12.6}$$

The two-by-two matrix in (12.6) will be written as M_n; it gives us the coefficients of the nth layer in terms of the coefficients of the $(n-1)$th layer. Note that the determinant of M_n is

$$\det M_n = \frac{q_{n-1}}{q_n}. \tag{12.7}$$

As we shall see shortly, this value is relevant to the conservation law of Sect. 2.1,

$$q_a\left(1-|r|^2\right)=q_b|t|^2. \tag{12.8}$$

To find the reflection and transmission amplitudes r and t we note from (12.1) that

$$\begin{pmatrix}\alpha_1\\ \beta_1\end{pmatrix}=M_1\begin{pmatrix}1\\ r\end{pmatrix},\qquad M_{N+1}\begin{pmatrix}\alpha_N\\ \beta_N\end{pmatrix}=\begin{pmatrix}\alpha_{N+1}\\ \beta_{N+1}\end{pmatrix}=\begin{pmatrix}t\\ 0\end{pmatrix}. \tag{12.9}$$

Thus

$$\begin{pmatrix}t\\ 0\end{pmatrix}=M\begin{pmatrix}1\\ r\end{pmatrix}=\begin{pmatrix}m_{11} & m_{12}\\ m_{21} & m_{22}\end{pmatrix}\begin{pmatrix}1\\ r\end{pmatrix}, \tag{12.10}$$

where the characteristic or profile matrix M is the product of $N+1$ two-by-two matrices,

$$M=M_{N+1}M_n\ldots M_n\ldots M_2M_1. \tag{12.11}$$

From (12.7) and the fact that the determinant of a product of matrices is the product of their determinants,

$$\det M=q_a/q_b, \tag{12.12}$$

From (12.10) and (12.12) the reflection and transmission amplitudes are given by

$$r=-\frac{m_{21}}{m_{22}},\qquad t=m_{11}-\frac{m_{12}m_{21}}{m_{22}}=\frac{\det\ M}{m_{22}}=\frac{q_a/q_b}{m_{22}}. \tag{12.13}$$

When q is real everywhere the matrix M_n defined by (12.6) has the form

$$M_n=\begin{pmatrix}\mu_{11} & \mu_{12}\\ \mu_{12}^* & \mu_{11}^*\end{pmatrix}, \tag{12.14}$$

that is, its elements are related by $\mu_{22}=\mu_{11}^*$, $\mu_{21}=\mu_{12}^*$. The product of two such matrices will also have this property (as may be verified by direct multiplication), and thus M has this property. Therefore

$$|m_{22}|^2-|m_{21}|^2=m_{11}m_{22}-m_{12}m_{21}=\det M, \tag{12.15}$$

and the conservation law (12.8) is satisfied by the reflection and transmission amplitude given by (12.13).

The above is for waves originating in medium a and transmitted into medium b. The reflection and transmission amplitudes for this case are

$$r_{ab} = -\frac{m_{21}}{m_{22}}, \qquad t_{ab} = \frac{q_a/q_b}{m_{22}}. \tag{12.16}$$

For waves incident from medium b the wavefunction has the limiting forms

$$t_{ba}\,\mathrm{e}^{-iq_a z} \leftarrow \psi \rightarrow \mathrm{e}^{-iq_b z} + r_{ba}\,\mathrm{e}^{iq_b z}, \tag{12.17}$$

and thus, for the same profile matrix M as before,

$$\begin{pmatrix} r_{ba} \\ 1 \end{pmatrix} = M \begin{pmatrix} 0 \\ t_{ba} \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} 0 \\ t_{ba} \end{pmatrix}. \tag{12.18}$$

Thus

$$r_{ba} = \frac{m_{12}}{m_{22}}, \qquad t_{ba} = 1/m_{22}. \tag{12.19}$$

Comparison with (12.16) gives the reciprocity relations

$$q_b t_{ab} = q_a t_{ba}, \tag{12.20}$$

and

$$\frac{r_{ba}}{r^*_{ab}} = -\frac{m_{12}}{m^*_{21}}\frac{m^*_{22}}{m_{22}} = -\frac{m_{12}}{m^*_{21}}\frac{t_{ba}}{t^*_{ba}}. \tag{12.21}$$

The first of these is the same as (2.14), and the second reduces to (2.18) when q is real everywhere (see (12.14) and the lines following it).

We now return to the case of waves incident from medium a, and consider a slightly different formulation which has the advantage of having real matrix elements when q is real. This is obtained by writing ψ_n in the interval (z_n, z_{n+1}) as

$$\psi_n = \alpha_n \cos q_n z + \beta_n \sin q_n z. \tag{12.22}$$

In the interval (z_{n-1}, z_n) the solution is

$$\psi_{n-1} = \alpha_{n-1} \cos q_{n-1} z + \beta_{n-1} \sin q_{n-1} z, \tag{12.23}$$

and continuity of ψ and $\mathrm{d}\psi/\mathrm{d}z$ at z_n now gives

$$\alpha_n c_n + \beta_n s_n = \alpha_{n-1} c_{n-1} + \beta_{n-1} s_{n-1}, \tag{12.24}$$

$$q_n(-\alpha_n s_n + \beta_n c_n) = q_{n-1}(-\alpha_{n-1} s_{n-1} + \beta_{n-1} c_{n-1}), \tag{12.25}$$

where c_n, s_n and c_{n-1}, s_{n-1} stand for the cosines and sines of $q_n z_n$ and $q_{n-1} z_n$. On solving for α_n and β_n in terms of α_{n-1} and β_{n-1} we now have

$$\begin{pmatrix} \alpha_n \\ \beta_n \end{pmatrix} = \begin{pmatrix} c_n c_{n-1} + \frac{q_{n-1}}{q_n} s_n s_{n-1} & c_n s_{n-1} - \frac{q_{n-1}}{q_n} s_n c_{n-1} \\ s_n c_{n-1} - \frac{q_{n-1}}{q_n} c_n s_{n-1} & s_n s_{n-1} + \frac{q_{n-1}}{q_n} c_n c_{n-1} \end{pmatrix} \begin{pmatrix} \alpha_{n-1} \\ \beta_{n-1} \end{pmatrix}. \tag{12.26}$$

The determinant of the two-by-two matrix M_n in (12.26) is again q_{n-1}/q_n, and so the determinant of the profile matrix $M = M_{N+1} M_n \ldots M_n \ldots M_2 M_1$ is q_a/q_b as before. The reflection and transmission amplitudes are now given by

$$\begin{pmatrix} t \\ it \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} 1+r \\ i-ir \end{pmatrix}; \tag{12.27}$$

and thus

$$r = -\frac{m_{11} - m_{22} + i m_{12} + i m_{21}}{m_{11} + m_{22} - i m_{12} + i m_{21}}. \tag{12.28}$$

$$t = \frac{2(m_{11} m_{22} - m_{12} m_{21})}{m_{11} + m_{22} - i m_{12} + i m_{21}} = \frac{2 q_a/q_b}{m_{11} + m_{22} - i m_{12} + i m_{21}}. \tag{12.29}$$

These formulae are not as simple as (12.13), but the advantage of the cosine and sine representation of ψ is that the matrix elements m_{ij} are real when q is real. This advantage is shared by the method introduced in the next section.

12.2 Matrices Relating Fields and Their Derivatives

The matrix methods developed above were based on recurrence relations for the coefficients of two linearly independent solutions within a given layer (chosen to be $\mathrm{e}^{\pm i q_n z}$, or $\cos q_n z$ and $\sin q_n z$). Here we give the matrices which relate the field and its derivative, layer to neighbouring layer. This version of the matrix formulation will be used in the remainder of this chapter, and also in the next chapter (on periodic stratifications). We will consider the s polarization first. The electric field is $\boldsymbol{E} = (0, \mathrm{e}^{iKz} E(z), 0)$, with

$$\frac{\mathrm{d}^2 E}{\mathrm{d}z^2} + q^2 E = 0, \qquad \mathrm{e}^{i q_a z} + r_s \, \mathrm{e}^{-i q_a z} \leftarrow E \rightarrow t_s \, \mathrm{e}^{i q_b z}. \tag{12.30}$$

The second order differential equation for E may be written as a pair of coupled first order differential equations (as in Sect. 5.1):

$$\frac{\mathrm{d}E}{\mathrm{d}z} = D, \qquad \frac{\mathrm{d}D}{\mathrm{d}z} = -q^2 E. \tag{12.31}$$

If $q(z)$ takes the value q_n in (z_n, z_{n+1}) as before, and E_n and D_n are the values of the field and of its derivative at z_n, the solution of (12.31) in $z_n \le z \le z_{n+1}$ is

$$E = E_n \cos q_n(z - z_n) + \frac{D_n}{q_n} \sin q_n(z - z_n), \tag{12.32}$$

$$D = D_n \cos q_n(z - z_n) - E_n q_n \sin q_n(z - z_n). \tag{12.33}$$

Since E and D are continuous at z_{n+1} (continuity follows from the differential equation for E, provided q^2 has no delta function singularities), it follows that

$$E_{n+1} = E_n \cos \delta_n + \frac{D_n}{q_n} \sin \delta_n, \tag{12.34}$$

$$D_{n+1} = D_n \cos \delta_n - E_n q_n \sin \delta_n. \tag{12.35}$$

Here

$$\delta_n = q_n(z_{n+1} - z_n) \tag{12.36}$$

is the phase increment in propagating from z_n to z_{n+1}. The relation between the coefficients at z_{n+1} and those at z_n is thus

$$\begin{pmatrix} E_{n+1} \\ D_{n+1} \end{pmatrix} = \begin{pmatrix} \cos \delta_n & q_n^{-1} \sin \delta_n \\ -q_n \sin \delta_n & \cos \delta_n \end{pmatrix} \begin{pmatrix} E_n \\ D_n \end{pmatrix} = M_n \begin{pmatrix} E_n \\ D_n \end{pmatrix}. \tag{12.37}$$

The matrix in (12.37) is unimodular (has unit determinant). Note that when $\delta_n = m\pi$, m an integer, the matrix equals $(-)^m$ times the unit matrix. This corresponds to the layer thickness $z_{n+1} - z_n$ being an integer (for even m) or half-integer (for odd m) multiple of the effective wavelength for motion in the z direction, $\lambda_z = 2\pi/q_n$.

Before deriving expressions for r_s and t_s in terms of the elements of the profile matrix $M = M_N \ldots M_1$ we will give the corresponding matrix formulation for the p polarization. This has $\mathbf{B} = (0, \mathrm{e}^{iKx} B(z), 0)$ with $B(z)$ satisfying

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{\varepsilon}\frac{\mathrm{d}B}{\mathrm{d}z}\right) + \frac{q^2}{\varepsilon} B = 0, \qquad \mathrm{e}^{iq_a z} - r_p \, \mathrm{e}^{-iq_a z} \leftarrow B \rightarrow \left(\frac{\varepsilon_b}{\varepsilon_a}\right)^{1/2} t_p \, \mathrm{e}^{iq_b z}. \tag{12.38}$$

We again write the second order differential equation for B as a pair of coupled first order equations (as in Sect. 5.3):

$$\frac{1}{\varepsilon}\frac{\mathrm{d}B}{\mathrm{d}z} = C, \qquad \frac{\mathrm{d}C}{\mathrm{d}z} = -\frac{q^2}{\varepsilon}B. \tag{12.39}$$

If B_n and C_n are the values at z_n, within $z_n \le z \le z_{n+1}$ we have

$$B = B_n \cos q_n(z - z_n) + \frac{\varepsilon_n C_n}{q_n} \sin q_n(z - z_n) \tag{12.40}$$

$$C = C_n \cos q_n(z - z_n) - \frac{B_n q_n}{\varepsilon_n} \sin q_n(z - z_n). \tag{12.41}$$

B and C are continuous at a discontinuity in ε; their continuity at z_{n+1} leads to the matrix relation

$$\begin{pmatrix} B_{n+1} \\ C_{n+1} \end{pmatrix} = \begin{pmatrix} \cos\delta_n & Q_n^{-1}\sin\delta_n \\ -Q_n \sin\delta_n & \cos\delta_n \end{pmatrix} \begin{pmatrix} B_n \\ C_n \end{pmatrix} \tag{12.42}$$

where $Q_n = q_n/\varepsilon_n$. The p wave matrix is also unimodular, and equal to $(-)^m$ times a unit matrix when $\delta_n = m\pi$ with integer m.

Note that the *layer* matrices in (12.37) and (12.42) depend on the thickness of the nth layer and on its dielectric properties. This is in contrast to the *boundary* matrices of the last section, where M_n was a function of the boundary position z_n and of the wavenumber components q_{n-1} and q_n on either side of the boundary. The N-layer system of Fig. 12.1 can be characterized by N layer matrices or by $N+1$ boundary matrices.

The profile matrix in the present case is

$$M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} = M_N M_{N-1} \ldots M_n \ldots M_2 M_1. \tag{12.43}$$

The reflection and transmission amplitudes for the s wave are obtained by matching the limiting form (12.3) to (12.32) at z_1 and z_{N+1}. We will use the notation $\alpha = q_a z_1$, $\beta = q_b z_{N+1}$. Then

$$E_1 = \mathrm{e}^{i\alpha} + r_s\,\mathrm{e}^{-i\alpha}, \qquad D_1 = iq_a(\mathrm{e}^{i\alpha} - r_s\,\mathrm{e}^{-i\alpha}), \tag{12.44}$$

$$E_{N+1} = t_s\,\mathrm{e}^{i\beta}, \qquad D_{N+1} = iq_b t_s \mathrm{e}^{i\beta}. \tag{12.45}$$

Since

$$\begin{pmatrix} E_{n+1} \\ D_{n+1} \end{pmatrix} = M_n \begin{pmatrix} E_n \\ D_n \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} E_{N+1} \\ D_{N+1} \end{pmatrix} = M \begin{pmatrix} E_1 \\ D_1 \end{pmatrix},$$

we have

$$\begin{pmatrix} t_s e^{i\beta} \\ iq_b t_s e^{i\beta} \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} e^{i\alpha} + r_s e^{-i\alpha} \\ iq_a(e^{i\alpha} - r_s e^{-i\alpha}) \end{pmatrix}. \tag{12.46}$$

Thus

$$r_s = e^{2i\alpha} \frac{q_a q_b m_{12} + m_{21} - iq_b m_{11} + iq_a m_{22}}{q_a q_b m_{12} - m_{21} + iq_b m_{11} + iq_a m_{22}}, \tag{12.47}$$

$$t_s = e^{i(\alpha-\beta)} \frac{2iq_a}{q_a q_b m_{12} - m_{21} + iq_b m_{11} + iq_a m_{22}}. \tag{12.48}$$

(In the numerator of (12.48) we have replaced $\det M = m_{11}m_{22} - m_{12}m_{21}$ by unity, since the matrix M is a product of unimodular matrices.) These formulae have the same form as (2.25) and (2.26), if a unit Wronskian is assumed in the latter. The matrix elements m_{ij} are real if ε is real, just as in Sect. 2.2 the functions F and G may be taken to be real in the same circumstances. When q_a, q_b and the matrix elements are real, the reflectance and transmittance are given by

$$R_s = |r_s|^2 = \frac{(q_a q_b m_{12} + m_{21})^2 + (q_b m_{11} - q_a m_{22})^2}{(q_a q_b m_{12} - m_{21})^2 + (q_b m_{11} + q_a m_{22})^2}, \tag{12.49}$$

$$T_s = \frac{q_b}{q_a}|t_s|^2 = \frac{4q_a q_b}{(q_a q_b m_{12} - m_{21})^2 + (q_b m_{11} + q_a m_{22})^2}. \tag{12.50}$$

The corresponding formulae for the p wave may be obtained from (12.38) and (12.42). We find

$$-r_p = e^{2i\alpha} \frac{Q_a Q_b m_{12} + m_{21} - iQ_b m_{11} + iQ_a m_{22}}{Q_a Q_b m_{12} - m_{21} + iQ_b m_{11} + iQ_a m_{22}}, \tag{12.51}$$

$$\left(\frac{\varepsilon_b}{\varepsilon_a}\right)^{1/2} t_p = e^{i(\alpha-\beta)} \frac{2iQ_a}{Q_a Q_b m_{12} - m_{21} + iQ_b m_{11} + iQ_a m_{22}}. \tag{12.52}$$

The meaning of α and β is $q_a z_1$ and $q_b z_{N+1}$ as before; $Q_a = q_a/\varepsilon_a$ and $Q_b = q_b/\varepsilon_b$. The matrix elements here are not the same as in (12.47) and (12.48): they are found by taking the product of the N matrices given by (12.42) instead of (12.37). The analogous formulae (2.40) and (2.41) are not of precisely the same form, since they apply to the case of continuous $\varepsilon(z)$ only. For real ε the matrix elements are real, and in the absence of total reflection the p wave reflectance and transmittance are given by

$$R_p = |r_p|^2 = \frac{(Q_a Q_b m_{12} + m_{21})^2 + (Q_b m_{11} - Q_a m_{22})^2}{(Q_a Q_b m_{12} - m_{21})^2 + (Q_b m_{11} + Q_a m_{22})^2}, \tag{12.53}$$

$$T_p = \frac{q_b}{q_a}\left|t_p\right|^2 = \frac{4Q_aQ_b}{\left(Q_aQ_bm_{12} - m_{21}\right)^2 + \left(Q_bm_{11} + iQ_am_{22}\right)^2}. \tag{12.54}$$

An interesting special configuration is a stack of half-wavelength layers. We noted above that when a layer has phase increment equal to an integer times π, the layer matrix is equal to plus or minus times the unit matrix, depending on whether the integer is even or odd. We can define an effective wavelength for propagation in the z direction as $2\pi/q$; for propagation in the zx plane the two component wavelengths are $\lambda_x = 2\pi/K$ and $\lambda_z = 2\pi/q$, with the total wavelength λ given by $\lambda^{-2} = \lambda_x^{-2} + \lambda_z^{-2}$. The condition $q\delta z = m\pi$ is equivalent to

$$\delta z = m\lambda_z/2; \tag{12.55}$$

the layer thickness is an integer times one half of the effective z component wavelength within the layer. When this condition holds all the multiply reflected waves are in phase (see Sect. 2.4, and especially Fig. 2.5b). For thick layers this may happen at several angles of incidence: the condition $q\delta z = m\pi$ is satisfied for angles of incidence

$$\theta_a(m) = \arcsin\left\{\frac{\varepsilon - [m\pi/(\omega/c)\delta z]^2}{\varepsilon_a}\right\}^{1/2}. \tag{12.56}$$

Possible values of m lie in the range

$$\frac{1}{\pi}\frac{\omega}{c}\delta z(\varepsilon - \varepsilon_a)^{\frac{1}{2}} \le m \le \frac{1}{\pi}\frac{\omega}{c}\delta z\varepsilon^{1/2}. \tag{12.57}$$

For example, when $\varepsilon_a = 1$, $\varepsilon = (4/3)^2$, $\varepsilon_b = (3/2)^2$ and $(\omega/c)\delta z = 27$ (as in Fig. 5.4), the possible values of m are 8–11. At the corresponding angles $\theta_a(m)$ the layer is invisible: the reflectivity of the stack is the same as if the layer were absent.

A stack of N homogeneous layers, each with thickness equal to an integer m_n times half of the effective wavelength in the layer, namely with $q_n\delta z_n = m_n\pi$, will have a profile matrix which is equal to $(-)^S$ times the unit matrix, where $S = \sum_1^N m_n$. This stack will be invisible as far as reflectivity measurement is concerned (at the given angle of incidence and frequency). When S is even, the reflection *amplitudes* (given by (12.47) and (12.51)), are also identical to that of a step in the dielectric function from ε_a to ε_b, *located at* z_1. This is remarkable, since no restriction has been placed on the thickness $\Delta z = \sum_1^N \delta z_n$ of the stack. However, actual layer structures will deviate slightly from the assumed conditions.

12.3 Multilayer Dielectric Mirrors at Normal Incidence

High reflectivity mirrors (used, for example, to form the optical cavity of lasers) are made by depositing alternating layers of high and low dielectric constant materials on a substrate, as shown in Fig. 12.2.

These mirrors are wavelength-selective, high reflectivity at a particular frequency being obtained by constructive interference of the waves reflected at each discontinuity in refractive index. Each layer is made a quarter of a wavelength thick (at the design frequency) so as to make all the reflected waves in phase. For example: if in Fig. 12.2 the front of the mirror (on the left) is at $z = 0$, the reflection amplitude off the first face is $(q_a - q_h)/(q_a + q_h)$, which is real and negative. The reflection amplitude off the second (high|low) interface is, from (1.15), a positive fraction times $e^{2i\delta_h}(q_h - q_l)/(q_h + q_l)$, where $\delta_h = q_h \delta z_h$. When $\delta_h = \pi/2$, which at normal incidence amounts to $\delta z_h = \lambda_h/4$ where λ_h is the wavelength in the high-index material, this second reflection amplitude is also real and negative. Similarly, the contribution from the next (low|high) interface will be in phase with the preceding if $\delta z_l = \lambda_l/4$. Thus constructive interference is obtained by making each layer a quarter of a wavelength thick (or in general an odd integer times a quarter wavelength).

The theory for periodically stratified media will be more fully developed in Chap. 13. The matrix for a single period is

$$\begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} = \begin{pmatrix} c_l & s_l/q_l \\ -q_l s_l & c_l \end{pmatrix} \begin{pmatrix} c_h & s_h/q_h \\ -q_h s_h & c_h \end{pmatrix} = \begin{pmatrix} c_l c_h - \frac{q_h}{q_l} s_l s_h & \frac{c_l s_h}{q_h} + \frac{s_l c_h}{q_l} \\ -q_l s_l c_h - q_h s_h c_l & c_l c_h - \frac{q_l}{q_h} s_l s_h \end{pmatrix}. \tag{12.58}$$

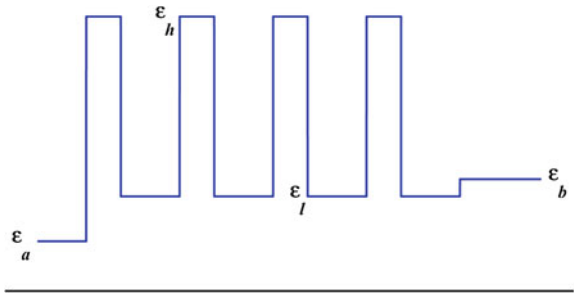


Fig. 12.2 Dielectric function profile of a multilayer dielectric mirror, drawn to scale for an $(HL)^4$ configuration, with the refractive indices for the high and low index materials $n_h = 2.35$, $n_l = 1.38$. These correspond to ZnS and MgF_2, at 633 nm (data from Table l.l of Yariv and Yeh 1984). The substrate is glass, $n_b = 1.5$. For optimum reflectance at a given frequency the layers are a quarter wavelength thick: $\delta z_h = \frac{\lambda_h}{4} = (\pi c/2\omega) n_h^{-1}$, $\delta z_l = \frac{\lambda_l}{4} = (\pi c/2\omega) n_l^{-1}$

Here we consider *normal incidence*, for which, with subscripts h and l applied to the values for the high and low refractive index materials,

$$c = \cos\delta, \qquad s = \sin\delta, \qquad \delta = n\left(\frac{\omega}{c}\right)\delta z. \tag{12.59}$$

The matrix elements are the same for the s and the p waves at normal incidence; Sect. 13.2 deals with the case of general incidence. The reflectivity for an $(HL)^N$ mirror, with light incident from a medium of dielectric constant ε_a, resting on a substrate of dielectric constant ε_b, is obtained by substituting the matrix elements m_{ij} of (12.58) and (12.59) into (12.49).

For a perfect $\lambda/4$ stack at normal incidence and at the design angular frequency ω_0, $\delta_h = \pi/2 = \delta_l$ and from (12.58)

$$m_{11} = -\frac{n_h}{n_l}, \qquad m_{22} = -\frac{n_l}{n_h}, \qquad m_{12} = 0 = m_{21}, \tag{12.60}$$

where $n_h = \sqrt{\varepsilon_h}$ and $n_l = \sqrt{\varepsilon_l}$, are the refractive indices of the alternating layers. In this case the matrix for one period is diagonal, and the profile matrix is equal to

$$M^N = \begin{pmatrix} \left(-\frac{n_h}{n_l}\right)^N & 0 \\ 0 & \left(-\frac{n_l}{n_h}\right)^N \end{pmatrix}. \tag{12.61}$$

The normal incidence reflectivity at the design frequency is thus

$$R(\omega_0) = \left\{ \frac{\frac{n_b}{n_a}\left(\frac{n_h}{n_l}\right)^{2N} - 1}{\frac{n_b}{n_a}\left(\frac{n_h}{n_l}\right)^{2N} + 1} \right\}^2. \tag{12.62}$$

This tends to unity rapidly with N, the number of high|low strata. For example: when $n_a = 1$, $n_b = 1.5$, $n_h = 2.35$ and $n_l = 1.38$ (as in Figs. 12.2 and 12.3) the $N = 1, 2, 3, 4, 5$, and 6 stacks give $R(\omega_0) \approx 0.392,\ 0.728, 0.896, 0.963, 0.987$ and 0.996.

At normal incidence, but away from the design wavelength, the phase changes $\delta_h = n_h(\omega/c)\delta z_h$ and $\delta_l = n_l(\omega/c)\delta z_l$ remain equal (again for the $\lambda/4$ stack) but are no longer equal to $\pi/2$. Let $\delta = (\pi/2)(\omega/\omega_0)$ denote the common value of δ_h and δ_l. The matrix of one high|low period is now

$$\begin{pmatrix} \cos^2\delta - \frac{n_h}{n_l}\sin^2\delta & \frac{c}{\omega}\left(\frac{1}{n_h} + \frac{1}{n_l}\right)\cos\delta\sin\delta \\ -\frac{\omega}{c}(n_h + n_l)\cos\delta\sin\delta & \cos^2\delta - \frac{n_l}{n_h}\sin^2\delta \end{pmatrix}. \tag{12.63}$$

Figure 12.3 shows the frequency dependence of the reflectivity of an $(HL)^4$ stack at normal incidence.

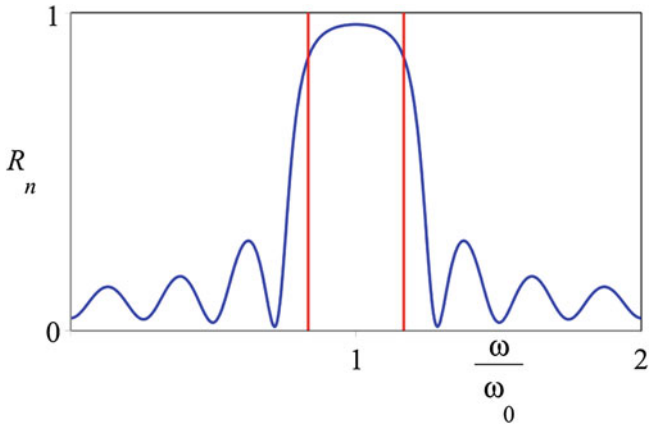


Fig. 12.3 Normal incidence reflectivity of a $(HL)^4$ stack of dielectric layers, as a function of the frequency. The reflectivity is maximum at the design frequency ω_0, where the wavelength in both the high and low index layers is four times the layer thickness. The reflectivity is periodic in ω, with period $2\omega_0$. The refractive indices are as in Fig. 12.2. The vertical lines indicate the infinite-stack stop band boundaries, as given by (12.65)

We shall see in Sect. 13.2 that the reflectivity depends on the trace of the single-period matrix, specifically on a phase angle ϕ defined by

$$\cos\phi = \frac{1}{2}(m_{11} + m_{22}) = \cos^2\delta - \frac{1}{2}\left(\frac{n_h}{n_l} + \frac{n_l}{n_h}\right)\sin^2\delta. \tag{12.64}$$

The cross-over from high to low normal incidence reflectivity takes place as $\cos\phi$ increases through -1 from its minimum value of $-\frac{1}{2}[(n_h/n_l) + (n_l/n_h)]$, the latter being attained at the design frequency. The value $\cos\phi = -1$ occurs at the frequencies $\omega = \omega_0 \pm \Delta\omega$, where from (12.64) and $\delta = (\pi/2)(\omega/\omega_0)$,

$$\frac{\Delta\omega}{\omega_0} = \frac{2}{\pi}\arcsin\left(\frac{n_h - n_l}{n_h + n_l}\right). \tag{12.65}$$

(The high|low stack of Fig. 12.2 has $\Delta\omega/\omega_0 \approx 0.16748$.) Within the band $\omega = \omega_0 \pm \Delta\omega$ the value of $\cos\phi$ is below -1, and the reflectivity has a single maximum. As the number of high|low periods increases, the reflectivity within the band tends to unity. From zero frequency to $\omega_0 - \Delta\omega$, and from $\omega_0 + \Delta\omega$ to $2\omega_0$ (we stay within one frequency period in this characterization), $|\cos\phi| < 1$ and the reflectivity oscillates. Full discussion of the underlying band structure and of these oscillations may be found in Sect. 13.2.

A modification of the $(HL)^N$ quarter-wave stack gives a narrow band of transmission in the middle of the high-reflectivity region: such a structure is symbolically represented by $(HL)^n H^2 (LH)^n$. As the integer n increases, the pass band in the middle of the stop band gets narrower. See, for example, Lipson et al. (2010), Sect. 10.3.5.

12.4 Reflection of Long Waves

We return now to the problem of reflection by an arbitrary profile, as treated by the matrix method of Sect. 12.2, and consider the case where the total profile thickness is small compared to the wavelength of the radiation. The interface is represented by N homogeneous layers. As N increases to infinity the phase increments $\delta_n = q_n \delta z_n$ become infinitesimal. At the start we will keep first and second order terms in δ_n in the s wave matrix of (12.37),

$$M_n = \begin{pmatrix} 1 - \delta_n^2/2 & q_n^{-1}\delta_n \\ -q_n\delta_n & 1 - \delta_n^2/2 \end{pmatrix} + \cdots, \tag{12.66}$$

but it will turn out that only the first order terms of M_n play a role as $N \to \infty$. We write (12.66) as

$$M_n = \left(1 - \frac{1}{2}\delta_n^2\right) I + q_n^{-1}\delta_n J - q_n \delta_n \tilde{J} + \cdots, \tag{12.67}$$

where I is the identity (or unit) matrix, and

$$J = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \tilde{J} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \tag{12.68}$$

The profile matrix

$$M = M_N M_{N-1} \ldots M_n \ldots M_2 M_1 \tag{12.69}$$

may be expanded in powers of δ_n. The fact that

$$J^2 = 0 = \tilde{J}^2 \tag{12.70}$$

simplifies the result, which reads

$$\begin{aligned} M = &\left(1 - \frac{1}{2}\sum_1^N \delta_n^2\right) I + \left(\sum_1^N \delta z_n\right) J - \left(\sum_1^N q_n^2 \delta z_n\right) \tilde{J} \\ &- \sum_{n=1}^{N-1} \sum_{l=n+1}^{N} (q_n^2 \delta z_n \delta z_l J\tilde{J} + \delta z_n q_l^2 \delta z_l \tilde{J} J) + \cdots \end{aligned} \tag{12.71}$$

In the limit as $N \to \infty$ and $\delta z_n \to 0$, $\sum_1^N \delta_n^2 \to 0$ and

$$\sum_{1}^{N} \delta z_n f_n \to \int_a^b \mathrm{d}z\, f(z), \tag{12.72}$$

$$\sum_{n=1}^{N-1} \sum_{l=n+1}^{N} \delta z_n f_n \delta z_l g_l \to \int_a^b \mathrm{d}z f(z) \int_z^b d\zeta g(\zeta), \tag{12.73}$$

where $a = z_1$ and $b = z_{N+1}$ denote the left and right boundaries of the stack (see Fig. 12.1). Since

$$J\tilde{J} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \tilde{J}J = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \tag{12.74}$$

the limit as $N \to \infty$ of (12.71) becomes

$$M = \begin{pmatrix} 1 - \int_a^b \mathrm{d}z\, q^2(z)(b-z) & b-a \\ -\int_a^b \mathrm{d}z\, q^2(z) & 1 - \int_a^b \mathrm{d}z\, q^2(z)(z-a) \end{pmatrix} \tag{12.75}$$

The s wave reflectivity is given by (12.49). After some reduction, the result to second order in the thickness $b - a$ is found to be

$$\begin{aligned} R_s = \left(\tfrac{q_a - q_b}{q_a + q_b}\right)^2 + \tfrac{4q_a q_b}{(q_a + q_b)^4} \Bigg[(q_b^2 - q_a^2) \int_a^b \mathrm{d}z (2z - a - b) q^2 \\ + \int_a^b \mathrm{d}z (q^2 - q_a^2) \int_a^b d\zeta (q^2 - q_b^2) \Bigg] + \cdots \end{aligned} \tag{12.76}$$

The substitution $q^2 = \varepsilon\omega^2/c^2 - K^2$ reduces the square bracket in (12.76) to ω^4/c^4 times the angle-independent term

$$(\varepsilon_b - \varepsilon_a) \int_a^b \mathrm{d}z (2z - a - b)\varepsilon + \int_a^b \mathrm{d}z (\varepsilon - \varepsilon_a) \int_a^b \mathrm{d}\zeta (\varepsilon - \varepsilon_b). \tag{12.77}$$

At this stage the s reflectivity has been reduced to the same form as (5.67). The subsequent analysis of Sect. 5.5 shows that this is equivalent to the second order result of Chap. 3, (3.51).

For the p wave, the profile matrix is the product of matrices of the form (12.42), which to second order in the phase increment $\delta_n = q_n \delta z_n$ is

$$M_n = \begin{pmatrix} 1-\delta_n^2/2 & \delta_n/Q_n \\ -Q_n\delta_n & 1-\delta_n^2/2 \end{pmatrix} + \cdots, \tag{12.78}$$

where $Q_n = q_n/\varepsilon_n$. We again write M_n in terms of the matrices I, J and $\tilde{J}$, and form the product (12.69) to obtain the profile matrix M. The resulting elements of M are, after taking the limit of $N \to \infty$ as before,

$$\begin{aligned} m_{11} &= 1 - \int_a^b \mathrm{d}z\; q^2(z)/\varepsilon(z) \int_z^b \mathrm{d}\zeta\; \varepsilon(\zeta) + \cdots \\ m_{12} &= \int_a^b \mathrm{d}z\; \varepsilon(z) + \cdots \\ m_{21} &= -\int_a^b \mathrm{d}z\; q^2(z)/\varepsilon(z) + \cdots \\ m_{22} &= 1 - \int_a^b \mathrm{d}z\; \varepsilon(z) \int_z^b \mathrm{d}\zeta\; q^2(\zeta)/\varepsilon(\zeta) + \cdots. \end{aligned} \tag{12.79}$$

The p wave reflectivity is given by (12.53); to second order in the total thickness of the inhomogeneity this takes the form

$$\begin{aligned} R_p = \left(\tfrac{Q_a - Q_b}{Q_a + Q_b}\right)^2 + \tfrac{4Q_aQ_b}{(Q_a+Q_b)^4}\big[(Q_b^2 - Q_a^2)(m_{11} - m_{22}) \\ + Q_b^2Q_b^2m_{12}^2 + m_{12}m_{21}(Q_a^2 + Q_b^2) + m_{21}^2\big] + \cdots. \end{aligned} \tag{12.80}$$

A substantial reduction of (12.80) is required in order to regain the invariant form (3.50).

12.5 Absorbing Stratified Media: Some General Results

General theorems for arbitrary stratifications have already been given in Sects. 2.1–2.3; Sect. 10.2 briefly discussed results for absorbing media. Here we give three theorems which follow from the matrix analysis of wave propagation through layered media.

(i) *The transmittance of a stratified medium is independent of the direction of propagation*. The transmittance T is defined as the ratio of the energy leaving a unit area of the interface in unit time to the energy incident on a unit area in unit time. The transmittance for propagation from a non-absorbing medium a, through an arbitrary stratification (which may be absorbing), to a non-absorbing medium b, is

$$T_{ab} = \frac{q_b}{q_a}|t_{ab}|^2 \tag{12.81}$$

(see the discussion following (2.8) and Fig. 2.1). The equality of T_{ab} and T_{ba} follows from (12.16) and (12.19), of which the relevant parts are

$$t_{ab} = \frac{q_a/q_b}{m_{22}}, \qquad t_{ba} = 1/m_{22}. \tag{12.82}$$

(ii) *An arbitrary stratified medium is equivalent to two suitably chosen adjacent homogeneous layers* (Herpin 1947). Equivalence here means that the profile matrix elements m_{ij} of the two systems are the same. A general profile matrix

$$M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \tag{12.83}$$

has four elements (in general complex), which are linked by one constraint, namely the value of the determinant $m_{11}m_{22} - m_{12}m_{21}$. The latter is equal to q_a/q_b for the boundary matrices defined in Sect. 12.1, and to unity for the layer matrices used from Sect. 12.2 onward. A single homogeneous layer has the layer matrix

$$M_1 = \begin{pmatrix} \cos\delta_1 & q_1^{-1}\sin\delta_1 \\ -q_1\sin\delta_1 & \cos\delta_1 \end{pmatrix}.$$

This cannot represent (12.83) since it has its diagonal elements equal, and has only two free parameters (δ_1 and q_1). The profile matrix for two homogeneous layers 1 and 2 is

$$M = M_2M_1 = \begin{pmatrix} c_1c_2 - \frac{q_1}{q_2}s_1s_2 & \frac{s_1c_2}{q_1} + \frac{c_1s_2}{q_2} \\ -c_1q_2s_2 - c_2q_1s_1 & c_1c_2 - \frac{q_1}{q_2}s_1s_2 \end{pmatrix}, \tag{12.84}$$

where $c_1 = \cos\delta_1, \ldots, s_2 = \sin\delta_2$. This has unit determinant and four parameters (δ_1, q_1 and δ_2, q_2); it is thus sufficiently general to represent (12.83). But note that the equivalence, established by making the elements of (12.83) and (12.84) equal, will hold at a given angle of incidence and a given frequency only: as either changes, so do the parameters of the two-layer system.

(iii) For non-absorbing media the reflectance R and transmittance T are related by $R + T = 1$. For an absorbing stratification between two non-absorbing media the conservation law becomes $R + T + A = 1$, where A is the absorptance, a positive quantity for passive media. Thus the ratio $(1 - R)/T = 1 + A/T$ is in general greater than unity. Abelès (1950) has shown that *if an arbitrary non-absorbing layer is inserted in front of the absorbing layer, causing the reflectance to change to* R' *and the transmittance to* T'*, the ratio of* $1 - R$ *to* T *is unaltered:*

$$\frac{1-R}{T} = \frac{1-R'}{T'}. \tag{12.85}$$

Since $(1-R)/T = 1+A/T$, the Abelès result is equivalent to $A/T = A'/T'$: the absorptance to transmittance ratio is unchanged by the insertion of a non-absorbing layer in front of the absorber. The unprimed and primed configurations are illustrated in Fig. 12.4.

Let m_{ij} be the complex elements of the matrix representing the left-hand configuration in Fig. 12.4, and m'_{ij} be those representing the right-hand configuration. Then from (12.47) and (12.48) we have

$$R = \left| \frac{q_a q_b m_{12} + m_{21} - i q_b m_{11} + i q_a m_{22}}{q_a q_b m_{12} - m_{21} + i q_b m_{11} + i q_a m_{22}} \right|^2, \tag{12.86}$$

$$T = \frac{4 q_a q_b}{\left| q_a q_b m_{12} - m_{21} + i q_b m_{11} + i q_a m_{22} \right|^2}, \tag{12.87}$$

so that

$$\frac{1-R}{T} = \mathrm{Re}(m_{22}^* m_{11} - m_{12}^* m_{21}) - \mathrm{Im}(q_b m_{12}^* m_{11} + q_b^{-1} m_{22}^* m_{21}). \tag{12.88}$$

A similar expression holds for $(1-R')/T'$, with m_{ij} replaced by m'_{ij}, the elements of $M' = M\tilde{M}$, where $\tilde{M}$ is the layer matrix for the profile with dielectric function $\tilde{\varepsilon}(z)$. The matrix elements of M' are

$$\begin{aligned} m'_{11} &= m_{11}\tilde{m}_{11} + m_{12}\tilde{m}_{21} \quad & m'_{12} &= m_{11}\tilde{m}_{12} + m_{12}\tilde{m}_{22} \\ m'_{21} &= m_{21}\tilde{m}_{11} + m_{22}\tilde{m}_{21} \quad & m'_{22} &= m_{21}\tilde{m}_{12} + m_{22}\tilde{m}_{22} \end{aligned} \tag{12.89}$$

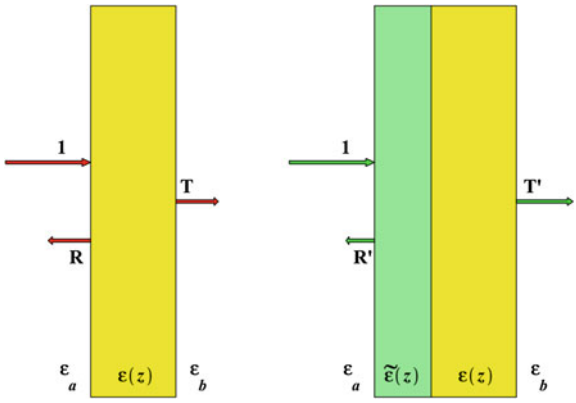


Fig. 12.4 Two configurations which have equal value of the ratio of $1-R$ to T. The dielectric function $\varepsilon(z)$ may be complex, while $\tilde{\varepsilon}(z)$ is real; ε_a and ε_b are real constants

On using the fact that the elements of $\tilde{M}$ are real, and that $\tilde{M}$ has unit determinant, the expression for $(1-R')/T'$ reduces to that for $(1-R)/T$.

12.6 High Transparency of an Absorbing Film in a Frustrated Total Reflection Configuration

In Sect. 10.6 we discussed attenuated total reflection, the phenomenon where a metallic layer or substrate converts a total internal reflection situation into one of low (or even zero) reflection of the p polarized wave. The physical basis of the phenomenon is the excitation of surface waves at the metal|dielectric boundaries, as explained in Sect. 10.6. There we considered the two configurations illustrated in Fig. 10.4. Here we consider the high|low|complex|low|high dielectric function configuration of Fig. 12.5. The dielectric function profile corresponding to this configuration is shown in the lower diagram.

We will calculate R_p and T_p for the symmetric case where the low-index dielectric layers have the same thickness l. The profile matrix for this case is

$$M = M_l M_m M_l \tag{12.90}$$

where, from (12.42),

$$M_l = \begin{pmatrix} \cos\delta_l & Q_l^{-1}\sin\delta_l \\ -Q_l\sin\delta_l & \cos\delta_l \end{pmatrix}, \qquad M_m = \begin{pmatrix} \cos\delta & Q^{-1}\sin\delta \\ -Q\sin\delta & \cos\delta \end{pmatrix}. \tag{12.91}$$

Here $Q_l = q_l/\varepsilon_l$, $\delta_l = q_l l$, $Q = q/\varepsilon$ and $\delta = qd$. (The quantities for the metal film are complex, with $q = q_r + iq_i$ and so on.) Thus the elements of the profile matrix M are

$$\begin{aligned} m_{11} &= (c_l^2 - s_l^2)c - c_l s_l\left(\tfrac{Q_l}{Q} + \tfrac{Q}{Q_l}\right)s = m_{22} \\ m_{12} &= c_l^2 Q^{-1}s + 2Q_l^{-1}c_l s_l c - Q_l^{-2}s_l^2 Qs \\ m_{21} &= Q_l^2 s_l^2 Q^{-1}s - 2Q_l c_l s_l c - c_l^2 Qs, \end{aligned} \tag{12.92}$$

where c, s and c_l, s_l stand for the cosines and sines of δ and δ_l. From (12.51) and (12.52) the p reflectance and transmittance are given by

$$R_p = \left|\frac{Q_h^2 m_{12} + m_{21}}{Q_h^2 m_{12} - m_{21} + 2iQ_h m_{11}}\right|^2, \tag{12.93}$$

$$T_p = \frac{4Q_h^2}{\left|Q_h^2 m_{12} - m_{21} + 2iQ_h m_{11}\right|^2}. \tag{12.94}$$

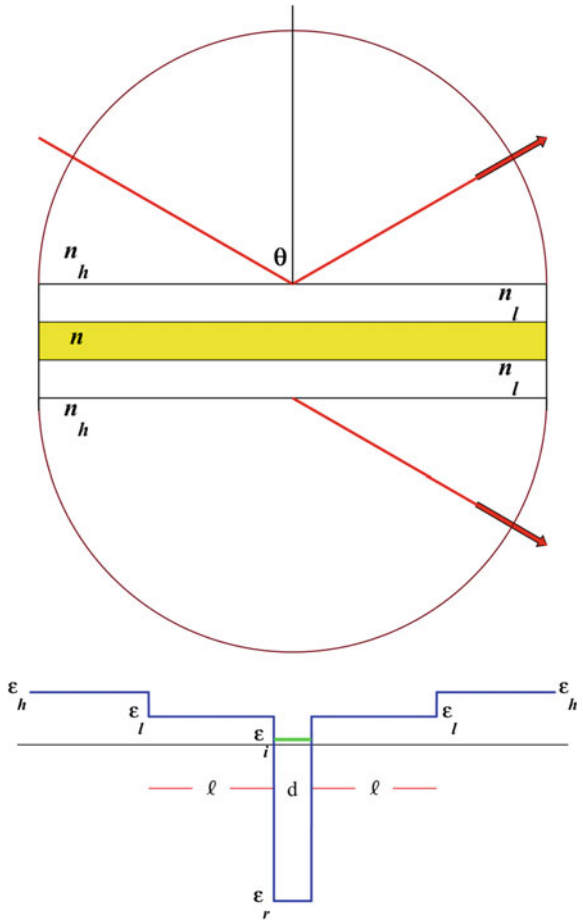


Fig. 12.5 *Upper diagram*: the high|low|complex|low|high refractive index symmetric configuration. *Lower diagram*: the corresponding dielectric function profile, for a metal film with complex dielectric constant $\varepsilon = \varepsilon_r + i\varepsilon_i$, sandwiched between two layers of low dielectric constant ε_l, which in turn are bounded by a material of high dielectric constant ε_h. The profile is drawn to scale for high refractive index glass $(\varepsilon_h = 3.617)$, lithium fluoride $(\varepsilon_l = 1.938)$ and silver $(\varepsilon = -10.755 + 0.361i)$ at 546 nm, as in Figs. 10.5, 10.6, 10.7, 10.8 and 10.9

We are most interested in the attenuated total reflection case, for which the angle of incidence exceeds the critical angle at the high|low interface. Then q_l, Q_l and δ_l are positive imaginary, $c_l = \cosh|\delta_l|$ and $s_l = i\sinh|\delta_l|$. The reflectivity can be zero if the real and imaginary parts of $Q_h^2 m_{12} + m_{21}$ can be made zero simultaneously. For a given set of materials the variables are the angle of incidence, and the thicknesses of the low refractive index dielectric and of the metal film. Figure 12.6 gives an example of the p wave reflectance and transmittance for the configuration shown in Fig. 12.5. Further examples can be found in Otto (1976) (see especially

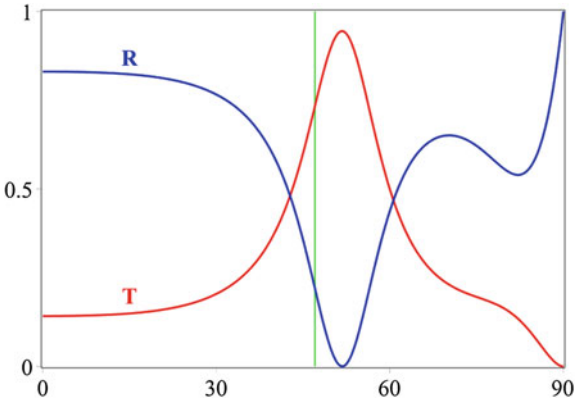


Fig. 12.6 Reflectance and transmittance for the p wave, for the configuration of Fig. 12.5 with $l = 100\,\text{nm}$, $d = 30\,\text{nm}$, at 546 nm vacuum wavelength. The reflectance minimum is near 51.8°, with $R_p \approx 1.3 \times 10^{-4}$. The location of the critical angle $\theta_c = \arcsin(n_l/n_h) \approx 47°$ for the high|low interface is indicated by the vertical line

his Fig. 14b, c), and in Dragila et al. (1985). The physical interpretation of the reflectance minimum in terms of the excitation of surface waves is discussed in these references and in Sect. 10.6.

12.7 Comparison of Numerical Approaches

Approximate analytical results for the reflection amplitudes have been given in the long wave and short wave cases (Chaps. 3 and 6). The long wave region of validity is extended by the perturbation and variational theories in Chap. 4, and the Rayleigh approximation of Chap. 5 is good at all wavelengths provided the reflection is weak. All these analytical methods share the drawback that higher-order approximations rapidly become cumbersome and thus of little practical value. For accurate results at intermediate wavelengths, and for a profile which is not among the few exactly soluble, numerical methods are needed.

The next Section describes numerical methods based on the matrix theory of Sect. 12.2. We do not give details of the direct solution of the wave equation, because the complications of that approach are greater, as the following outline shows. Let $\psi(z)$ satisfy

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}z^2} + q^2\psi = 0, \qquad \mathrm{e}^{iq_a z} + r\,\mathrm{e}^{-iq_a z} \leftarrow \psi \rightarrow t\,\mathrm{e}^{iq_b z}. \tag{12.95}$$

Since r and t are unknown, to integrate the differential equation we change the boundary conditions to

$$\alpha\, e^{iq_a z} + \beta\, e^{-iq_a z} \leftarrow \psi \rightarrow e^{iq_b z}, \tag{12.96}$$

integrate backward from some z_b at which $q(z)$ is close enough to q_b, and extract α and β in the region where $q(z)$ is close enough to q_a. Then the reflection and transmission amplitudes are found from $r = \beta/\alpha$, $t = 1/\alpha$. There are two complications in this method, both avoided by the matrix methods to be given later. The first is that r, t and α, β are in general complex, and thus solutions for both Re ψ and Im ψ are required. The second is that the extraction of the real and imaginary parts of α and β requires matching the real part of ψ to

$$(\alpha_r + \beta_r)\cos q_a z - (\alpha_i - \beta_i)\sin q_a z, \tag{12.97}$$

and the imaginary part of ψ to

$$(\alpha_i + \beta_i)\cos q_a z + (\alpha_r - \beta_r)\sin q_a z. \tag{12.98}$$

For example, by matching at points where $q_a z$ is an even and an odd multiple of $\pi/2$, one can obtain the four quantities $\alpha_r + \beta_r$, $\alpha_r - \beta_r$, $\alpha_i + \beta_i$ and $\alpha_i - \beta_i$, and hence the real and imaginary parts of α and β.

The complications in the direct solution of the differential equation outweigh (in our view) the advantage of ready access to a very large literature on the numerical solution of ordinary differential equations [see, for example, Temme (2010) and the references listed therein]. The matrix methods we will use, in contrast, evaluate only real quantities (in the absence of absorption), and the matching is done automatically: see for example the derivation of the expression for r and t in terms of the elements of the profile matrix in Sect. 12.2. The calculation of the profile matrix involves merely the computation of a product of two-by-two matrices, which is easily programmed.

12.8 Numerical Methods Based on the Layer Matrices

Two kinds of matrices have been introduced: the boundary matrices of Sect. 12.1, and the layer matrices of Sect. 12.2 onward. The latter are more convenient for numerical work and will be used here. Figure 12.7 shows the Rayleigh profile $\varepsilon(z) = \left[n_a^{-1} + (n_b^{-1} - n_a^{-1})(z - z_1)/\Delta z\right]^{-2}$ $(z_1 \le z \le z_{N+1},\ \Delta z = z_{N+1} - z_1)$, approximated by N homogeneous layers.

The nth layer extends from z_n to z_{n+1}, and in the case illustrated is homogeneous, with dielectric constant ε_n. The corresponding layer matrices for the s or p waves are given by (12.37) and (12.42):

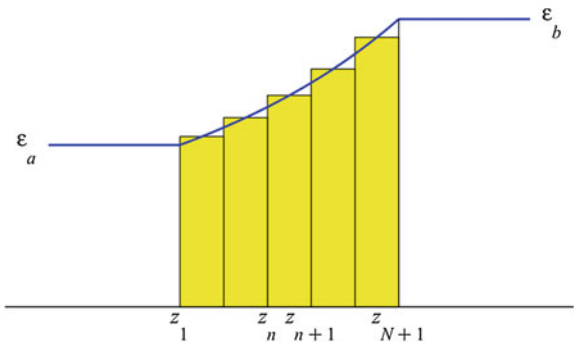


Fig. 12.7 The Rayleigh profile approximated by a set of five homogeneous layers. The figure is drawn for $n_a = 1$, $n_b = 4/3$

$$M_n = \begin{pmatrix} \cos\delta_n & q_n^{-1}\sin\delta_n \\ -q_n\sin\delta_n & \cos\delta_n \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \cos\delta_n & Q_n^{-1}\sin\delta_n \\ -Q_n\sin\delta_n & \cos\delta_n \end{pmatrix}, \tag{12.99}$$

where $\delta_n = q_n(z_{n+1} - z_n) \equiv q_n\delta z_n$, $q_n^2 = \varepsilon_n\omega^2/c^2 - K^2 = \omega^2/c^2(\varepsilon_n - \varepsilon_a\sin^2\theta_a)$, and $Q_n = q_n/\varepsilon_n$. To first order in the layer thickness δz_n these matrices are

$$\begin{pmatrix} 1 & \delta z_n \\ -q_n^2\delta z_n & 1 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 1 & \varepsilon_n\delta z_n \\ -q_n^2\delta z_n/\varepsilon_n & 1 \end{pmatrix}. \tag{12.100}$$

As N gets large, the layer thicknesses δz_n become small, and the matrices in (12.99) are well approximated by (12.100). This approximation for the layer matrices is in fact equivalent to the first order Euler method of solving the differential equation (12.95). To see this, let u be the real part of ψ, and $v = \mathrm{d}u/\mathrm{d}z$. The second order equation for u, $\mathrm{d}^2u/\mathrm{d}z^2 + q^2u = 0$, can be replaced by the pair of coupled first order equations, $\mathrm{d}u/\mathrm{d}z = v$ and $\mathrm{d}v/\mathrm{d}z = -q^2u$. The discretized version of this pair is

$$\frac{u_{n+1} - u_n}{\delta z_n} = v_n, \qquad \frac{v_{n+1} - v_n}{\delta z_n} = -q_n^2u_n. \tag{12.101}$$

In matrix form this reads (compare to the first matrix in (12.100))

$$\begin{pmatrix} u_{n+1} \\ v_{n+1} \end{pmatrix} = \begin{pmatrix} 1 & \delta z_n \\ -q_n^2\delta z_n & 1 \end{pmatrix}\begin{pmatrix} u_n \\ v_n \end{pmatrix}. \tag{12.102}$$

The matrix method with the profile replaced by a stack of homogeneous layers, and with the layer matrices calculated to first order in the layer thickness, is thus equivalent in accuracy to the Euler method.

We will not use this simplest approach, since it is easy to improve on the homogeneous layer approximation without much complication in the matrices and

the consequent programming. The improvement consists in approximating the profile by a set of layers in which the dielectric function varies *linearly* within each layer. This is illustrated in Fig. 12.8.

The variation of $\varepsilon(z)$ in $[z_n, z_{n+1}]$ is approximated by

$$\varepsilon(z) = \varepsilon_n + (z - z_n)\delta\varepsilon_n/\delta z_n, \tag{12.103}$$

where $\delta z_n = z_{n+1} - z_n$ as before, and $\delta\varepsilon_n = \varepsilon_{n+1} - \varepsilon_n$. Let $\Delta z = z_{N+1} - z_1$ be the total thickness of the profile. (When the values ε_a and ε_b are attained at minus and plus infinity, the profile must be truncated at some points z_1 and z_{N+1} as discussed later.) Then if at a given angular frequency ω a large enough number N of the layers is taken so that $(\omega/c)\delta z_n \ll 1$ (or $(\omega/c)\Delta z \ll N$ assuming the δz_n are roughly equal), each layer matrix will be well approximated by its long-wave form as given in Sect. 12.4 to second order in the layer thickness. (Lekner and Dorf (1987) go to third order, and also discuss a cubic fit to the dielectric function profile for each layer.)

For the s wave we find from (12.103) and (12.75) that the elements s_{ij} of the matrix M_n representing the nth layer are given by

$$\begin{aligned} s_{11} &= 1 + (\delta z_n)^2\left[K^2/2 - \frac{\omega^2}{c^2}(2\varepsilon_n + \varepsilon_{n+1})/6\right], \\ s_{12} &= \delta z_n, \qquad s_{21} = \delta z_n\left[K^2 - \frac{\omega^2}{c^2}(\varepsilon_n + \varepsilon_{n+1})/2\right], \\ s_{22} &= 1 + (\delta z_n)^2\left[K^2/2 - \frac{\omega^2}{c^2}(\varepsilon_n + 2\varepsilon_{n+1})/6\right]. \end{aligned} \tag{12.104}$$

The p wave matrix elements p_{ij} to second order in δz_n are found from (12.103) and (12.79): they are

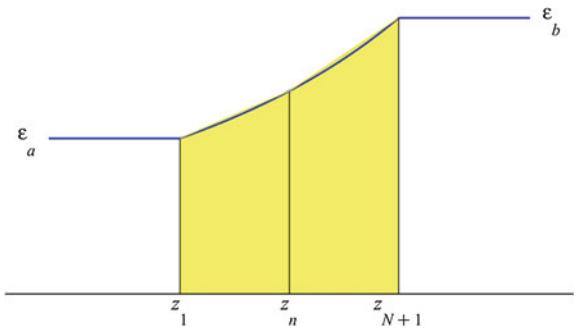


Fig. 12.8 Approximation of a profile by layers with linear variation in ε. The diagram is drawn for the Rayleigh profile, with parameters as in Fig. 12.7

$$p_{11} = 1 + (\delta z_n)^2 \left[\frac{K^2}{4\delta\varepsilon_n} \left\{ \frac{2\varepsilon_{n+1}^2}{\delta\varepsilon_n} \ln\frac{\varepsilon_{n+1}}{\varepsilon_n} - \varepsilon_{n+1} - \varepsilon_n \right\} - \frac{\omega^2}{c^2}(\varepsilon_n + 2\varepsilon_{n+1})/6 \right],$$

$$p_{12} = \delta z_n(\varepsilon_n + \varepsilon_{n+1})/2, \qquad p_{21} = \delta z_n \left[\frac{K^2}{\delta\varepsilon_n} \ln\frac{\varepsilon_{n+1}}{\varepsilon_n} - \frac{\omega^2}{c^2} \right], \tag{12.105}$$

$$p_{22} = 1 + (\delta z_n)^2 \left[\frac{K^2}{4\delta\varepsilon_n} \left\{ \varepsilon_n + \varepsilon_{n+1} - \frac{2\varepsilon_{n+1}^2}{\delta\varepsilon_n} \ln\frac{\varepsilon_{n+1}}{\varepsilon_n} \right\} - \frac{\omega^2}{c^2}(2\varepsilon_n + \varepsilon_{n+1})/6 \right].$$

In computation it is faster to replace the expressions involving $\ln(\varepsilon_{n+1}/\varepsilon_n)$ by the leading terms in their $\delta\varepsilon_n/\varepsilon_n$ expansion. The resulting matrix elements are

$$p_{11} \approx 1 + (\delta z_n)^2 \left[\frac{K^2}{6\varepsilon_n}(2\varepsilon_n + \varepsilon_{n+1}) - \frac{\omega^2}{c^2}(\varepsilon_n + 2\varepsilon_{n+1})/6 \right],$$

$$p_{12} = \frac{\delta z_n(\varepsilon_n + \varepsilon_{n+1})}{2}, \quad p_{21} \approx \delta z_n \left[K^2(1/\varepsilon_n + 1/\varepsilon_{n+1})/2 - \frac{\omega^2}{c^2} \right], \tag{12.106}$$

$$p_{22} \approx 1 + (\delta z_n)^2 \left[\frac{K^2}{6\varepsilon_{n+1}}(\varepsilon_n + 2\varepsilon_{n+1}) - \frac{\omega^2}{c^2}(2\varepsilon_n + \varepsilon_{n+1})/6 \right].$$

For comparison with the linear approximations (12.104) and (12.106) we write down the s and p homogeneous layer matrices (ε constant within each layer) to second order in δz_n. These are, from (12.99),

$$\begin{pmatrix} 1 - (\delta z_n q_n)^2/2 & \delta z_n \\ -\delta z_n q_n^2 & 1 - (\delta z_n q_n)^2/2 \end{pmatrix}, \begin{pmatrix} 1 - (\delta z_n q_n)^2/2 & \varepsilon_n \delta z_n \\ -\delta z_n q_n^2/\varepsilon_n & 1 - (\delta z_n q_n)^2/2 \end{pmatrix} \tag{12.107}$$

and are seen to be the degenerate forms of (12.104) and (12.106), obtained by setting $\varepsilon_{n+1} = \varepsilon_n$ in the linear layer formulae.

The linear layer formulae taken to first order in δz_n will be referred to as $L1$, and those retaining the terms second order in δz_n as $L2$. It is possible to improve on these methods, at a given order, by using *unimodular* matrices (Lekner 1990). We recall from Sect. 12.2 that the exact layer matrices are unimodular (have unit determinant); this guarantees energy conservation $R + T = 1$ (for both polarizations, in the absence of absorption), and reciprocity between the transmission amplitudes, as expressed for example in (2.14). In general, $R + T$ differs from unity by a term with factor $\det(M) - 1$. Approximate matrices which are not unimodular will not give reflection and transmission amplitudes which are energy conserving or which satisfy the reciprocity relations between the direct and inverse transmission amplitudes.

To construct unimodular matrices we can use *symmetrisation*, as detailed in Lekner (1990). The *s*-wave layer matrix given to second order in the layer thickness by (12.104) may be written in general as

$$\begin{pmatrix} 1-I_2 & I_1 \\ -J_1 & 1-J_2 \end{pmatrix} \tag{12.108}$$

$$I_1 = \delta z_n, \quad J_1 = \int_{z_n}^{z_{n+1}} \mathrm{d}z\, q^2(z),$$

$$I_2 = \int_{z_n}^{z_{n+1}} \mathrm{d}z\, q^2(z)(z_{n+1} - z), \tag{12.109}$$

$$J_2 = \int_{z_n}^{z_{n+1}} \mathrm{d}z\, q^2(z)(z - z_n).$$

The following matrix, correct to first-order in the layer thickness, has unit determinant

$$\begin{pmatrix} \frac{1-I_1J_1/4}{1+I_1J_1/4} & \frac{I_1}{1+I_1J_1/4} \\ \frac{-J_1}{1+I_1J_1/4} & \frac{1-I_1J_1/4}{1+I_1J_1/4} \end{pmatrix} \tag{12.110}$$

We note the identity $I_1J_1 = I_2 + J_2$, which follows from (12.109) and $\delta z_n = z_{n+1} - z_n$. (The same identity holds also for the *p*-wave layer matrix whose elements were given in (12.79).) From this identity it follows that the symmetrised second-order matrix

$$\begin{pmatrix} \frac{1-I_2/2}{1+I_2/2} & \frac{I_1}{1+I_2/2} \\ \frac{-J_1}{1+J_2/2} & \frac{1-J_2/2}{1+J_2/2} \end{pmatrix} \tag{12.111}$$

is also unimodular. We shall refer to the numerical methods using linear fit to the profile and the matrices (12.110) and (12.111) as *UL*1 and *UL*2. The results obtained

Table 12.1 Errors in the calculated reflectivities, and values of det(M), for the Rayleigh profile at normal incidence, with $(\omega/c)\Delta z = 1, n_a = 1, n_b = 4/3$ (air|water)

Method	*L*1	*UL*1	*L*2	*UL*2	*C*
Error (ppt)	−4	-8	−2	+1	−9
det(M)	1.14	1	1.0005	1	1

The profile was approximated by ten layers in each calculation, and a constant step size was used, $\delta z_n = \Delta z/10$. The error entries are in parts per thousand, calculated as $1000(R/R_e - 1)$. The letter *L* denotes that a linear variation of dielectric function within each layer is used; 1 or 2 denote that first or second order terms in δz_n are retained in each layer matrix. The exact reflectivity R_e is given in (2.108)

Table 12.2 Calculated normal incidence reflectivities at the *first reflectivity zero* for the Rayleigh profile with parameters as in Fig. 12.7 ($n_a = 1, n_b = 4/3$, $(\omega/c)\Delta z = 2.73295\ldots$, from (2.110))

Method	$L1$	$UL1$	$L2$	$UL2$	C
Error (ppm)	21	1.6	5.1	1.3	0.2
det(M)	2.6	1	1.026	1	1

The notation is as in Table 12.1, and again $N = 10$, but now the error is in parts per million (the table entries are the calculated reflectivity times 10^6)

by the $L1$, $UL1$ and $L2$, $UL2$ methods, for the Rayleigh profile with parameters as in Figs. 12.7 and 12.8, are shown in Tables 12.1 and 12.2. For comparison we also give the results, denoted by C, for the unimodular matrices in (12.37) or (12.99) where each layer has a constant value of the dielectric function.

From these results and similar ones for other profiles we draw the conclusion that the second order method is preferable to the first order method, and that the corresponding symmetrized unitary matrices are generally better in numerical accuracy, and also guarantee energy conservation and reciprocity. The unimodular homogeneous layer matrices denoted by C in Tables 12.1 and 12.2 are accurate, but slower to calculate because of their sines and cosines.

Further improvements are possible, by better than linear approximations to $\varepsilon(z)$ within each layer, and by going to higher order in δz_n. For example, one may approximate $\varepsilon(z)$ by a cubic in $[z_n, z_{n+1}]$ by using the derivatives ε_n' and ε_{n+1}' at the end-points. The formula resulting from matching to ε and ε' at z_n and z_{n+1} is

$$\begin{aligned}\varepsilon(z) \approx \varepsilon_n + (z - n_n)\varepsilon_n' + \left(\tfrac{z-z_n}{\delta z_n}\right)^2 \{3\delta\varepsilon_n - \delta z_n(2\varepsilon_n' + \varepsilon_{n+1}')\} \\ + \left(\tfrac{z-z_n}{\delta z_n}\right)^3 \{\delta z_n(\varepsilon_n' + \varepsilon_{n+1}') - 2\delta\varepsilon_n\}.\end{aligned} \tag{12.112}$$

The method obtained by using (12.112) and calculating the matrix elements to second order in δz_n, was found to be not much better than $L2$. The cubic third order formulae are available for extensive numerical work where high accuracy is needed (Lekner and Dorf 1987).

12.9 Variable Step Size, Profile Truncation, Total Reflection and Tunneling, Absorption, and Calculation of Wavefunctions

We now briefly discuss some further aspects of the numerical application of these matrix methods.

A constant step size $\delta z_n = \Delta z/N$ was chosen in the calculations discussed above. This is convenient, but not necessary; the matrix formulae given here are valid for *variable step size*. However, a constant step size is normally the simplest to

program, and in most cases is just as accurate as (for example) a variable step size chosen to make $\delta\varepsilon_n = \varepsilon_{n+1} - \varepsilon_n$ a constant.

The Rayleigh profile shown in Figs. 12.7 and 12.8, for which the results of Tables 12.1 and 12.2 were calculated, is an example of a profile of strictly finite range. For dielectric functions in which the inhomogeneity extends to infinity, such as the hyperbolic tangent profile

$$\varepsilon(z) = \frac{1}{2}(\varepsilon_a + \varepsilon_b) - \frac{1}{2}(\varepsilon_a - \varepsilon_b)\tanh\frac{z}{2\Delta z} = \frac{\varepsilon_a + \varepsilon_b e^{z/\Delta z}}{1 + e^{z/\Delta z}}, \tag{12.113}$$

profile truncation is necessary for the application of numerical methods. By truncation is meant that ε is set equal to ε_a for $z < a$ and to ε_b for $z > b$, where a and b are chosen so that $\varepsilon(a) - \varepsilon_a$ and $\varepsilon(b) - \varepsilon_b$ are sufficiently small to cause negligible error. For example, suppose we take $a = -7\Delta z$ and $b = 7\Delta z$ for the hyperbolic tangent profile. Since $e^7 \approx 10^3$, truncation at $\pm 7\Delta z$ can be expected to introduce an error of the order of one part per thousand. Larger values of $|a|$ and b will introduce smaller errors, but correspondingly larger numbers of layer matrices will be required to attain convergence to the truncated profile matrix elements. This is illustrated in the following table, calculated using the unimodular homogeneous matrices in (12.37) and (12.99). Truncation at the larger value of $|a|$ and b ultimately gives a more accurate reflectivity, but in the case illustrated the smaller cut-off gives a better reflectivity up to about 40 layers. This is because the smaller effective thickness of the profile is better approximated by a given number of layers (Table 12.3).

The formulae given in this chapter remain valid when $q(z)$ is imaginary and $q^2(z) < 0$, as is the case for a range of z values in *total internal reflection*, and in *tunneling*. No change is required in the calculation of the elements of the profile matrix, which remain real. The reflection and transmission amplitudes are still given by (12.47) and (12.48) in the s wave case, and by (12.51) and (12.52) in the p wave case. In total reflection q_b and Q_b are positive imaginary, and both R_s and R_p are unity. The quantity of interest is the phase of the reflected wave, given by

$$\delta_s = 2q_a z_1 - 2\text{atn}[s_{21} + |q_b| s_{11},\ q_a(s_{12}|q_b| + s_{22})], \tag{12.114}$$

$$\delta_p = 2q_a z_1 - 2\text{atn}[Q_a(p_{12}|Q_b| + p_{22}),\ p_{21} + |Q_b|\, p_{11}], \tag{12.115}$$

Table 12.3 The ratio of the calculated to the exact reflectivity, for the tanh profile truncated at a and b, as a function of the number of layer matrices

N	10	20	30	40
$-a, b = 7\Delta z$	0.965	0.991	0.995	0.997
$-a, b = 9\Delta z$	0.944	0.986	0.993	0.996

The values given are for $(\omega/c)\Delta z = 0.2$, $\varepsilon_a = 1$, $\varepsilon_b = (4/3)^2$, at normal incidence

where $\mathrm{atn}(y,x)$ is the arctangent of y/x placed in the correct quadrant according to the signs of x and y.

In the presence of *absorption* the dielectric function becomes complex. If only the substrate (characterized by ε_b, q_b and Q_b) is absorbing, the matrix elements remain real, and only the calculation of the reflectivity from the reflection amplitude is modified. (The expressions (12.49) and (12.53) for the s and p reflectances no longer apply.) When however the stratification is itself absorbing, the matrix elements are complex, and four multiplications of real matrices are needed in place of one performed in the non-absorbing case: if $R+iS$ and $U+iV$ represent the real and imaginary parts of two matrices, their product is

$$(U+iV)(R+iS) = UR - VS + i(VR+US). \tag{12.116}$$

Thus calculations involving absorption within the interface are roughly four times longer than those which do not.

We turn finally to the problem of the calculation of *wavefunctions* within the stratification. These are obtained, if required, as a by-product of the calculation of the elements of the profile matrix. In the s wave case, for example, we have

$$\begin{pmatrix} E_{n+1} \\ D_{n+1} \end{pmatrix} = M_n \begin{pmatrix} E_n \\ D_n \end{pmatrix} = M_n M_{n-1} \ldots M_2 M_1 \begin{pmatrix} E_1 \\ D_1 \end{pmatrix}. \tag{12.117}$$

Let v_{ij} be the elements of the product of n matrixes in (12.117). Then

$$E_{n+1} = v_{11}E_1 + v_{12}D_1 \tag{12.118}$$

gives the wavefunction at z_{n+1} in terms of the wavefunction and its derivative at z_1. The latter are given by

$$E_1 = \mathrm{e}^{i\alpha} + r_s\,\mathrm{e}^{-i\alpha}, \qquad D_1 = iq_a(\mathrm{e}^{i\alpha} - r_s\,\mathrm{e}^{-i\alpha}) \tag{12.119}$$

where $\alpha = q_a z_1$, and $r_s = r_r + ir_i$ is the reflection amplitude. The latter is found first, by calculating the product up to $n = N$. If the elements v_{11} and v_{12} are stored for all intermediate n, the wavefunction may then be plotted at the completion of the calculation of r_s. From (12.118) and (12.119), we have in the absence of absorption (real v_{ij}) that

$$\mathrm{Re}(E_{n+1}) = v_{11}\{(1+r_r)c + r_i s\} + v_{12}q_a\{r_i c - (1+r_r)s\}. \tag{12.120}$$

$$\mathrm{Im}(E_{n+1}) = v_{11}\{r_i c + (1-r_r)s\} + v_{12}q_a\{(1-r_r)c - r_i s\}, \tag{12.121}$$

where $c = \cos\alpha$ and $s = \sin\alpha$.

References

Abelès F (1950) Recherches sur la propagation des ondes électromagnétiques sinusoidales dans les milieux stratifiés. Application aux couches minces. Annales de Physique 5(596–640):706–782

Abelès F (1967) Optics of thin films In: van Heel ACS (ed) Advanced Optical Techniques, Chap 5, North-Holland

Dragila R, Luther-Davies B, Vukovic S (1985) High transparency of classically opaque metallic films. Phys Rev Lett 55:1117–1120

Herpin A (1947) Calcul du pouvoir réflecteur d'un système stratifié quelconque. Comptes Rendus 225:182–183

Lekner J, Dorf MC (1987) Matrix methods for the calculation of reflection amplitudes. J Opt Soc Am A: 4:2092–2095

Lekner J (1990) Matrix methods in reflection and transmission of compressional waves by stratified media. J Acoust Soc Am 87:2319–2324

Lipson A, Lipson SG, Lipson H (2010) Optical physics (4 ed). Cambridge University Press

Otto A (1976) Spectroscopy of surface polaritons by attenuated total reflection. In: Seraphin BO (ed) Optical properties of solids: new developments, Chap 13. North-Holland

Rayleigh JWS (1912) On the propagation of waves through a stratified medium, with special reference to the question of reflection. Proc R Soc A86:207–266

Temme NM (2010) Numerical methods. In: Olver FWJ et al (ed) NIST handbook of mathematical functions, Chap 3. Cambridge University Press

Weinstein W (1947) The reflectivity and transmissivity of multiple thin coatings. J. Opt. Soc. Amer. 37:576–581

Yariv A, Yeh P (1984) Optical waves in crystals. Wiley

Chapter 13
Periodically Stratified Media

Electron wavefunctions in crystals are modified by interaction of the electrons with the periodic ionic lattice to such an extent that band gaps appear in the spectrum of allowed states. This became clear in the early days of quantum mechanics (see for example Mott and Jones 1936/1958). However, the history of wave propagation in periodic structures extends back to Newton, who considered elastic waves on a one-dimensional lattice of masses connected by springs as a model for sound (Brillouin 1946/1953 gives a brief historical review). Rayleigh (1887, 1917) recognized the possibility of what are now known as stop bands or band gaps for waves in periodic structures, particularly in relation to the high reflection (at certain wavelengths and angles of incidence) by periodically stratified media. The optical aspects are covered in a monograph on photonic crystals (Joannopoulos et al. 1995). An overview of all kinds of waves in locally periodic media is given by Griffiths and Steinke (2001); Kinoshita (2008, 2013) surveys the optics of periodically structured biomaterials. This chapter concentrates on electromagnetic waves. Neutron reflection by period stratifications is discussed in Sect. 16.6.

The modern optics of stratifications was advanced by Abelès (1950); of special utility is his application of matrices to wave propagation, and use of the theorem that the Nth power of a unimodular (one with unit determinant) 2×2 matrix

$$M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \tag{13.1}$$

is given by

$$M^N = \begin{pmatrix} m_{11}S_N - S_{N-1} & m_{12}S_N \\ m_{21}S_N & m_{22}S_N - S_{N-1} \end{pmatrix}, \tag{13.2}$$

$$S_N = \frac{\sin N\phi}{\sin \phi}, \qquad \cos \phi = \frac{1}{2}\text{trace } M = \frac{1}{2}(m_{11} + m_{22}). \tag{13.3}$$

This result is easily proved by induction, on using $m_{11}m_{22} - m_{12}m_{21} = 1$ and the identity (or recurrence relation)

J. Lekner, *Theory of Reflection*, Springer Series on Atomic, Optical, and Plasma Physics 87, DOI 10.1007/978-3-319-23627-8_13

$$2S_N \cos\phi - S_{N-1} = S_{N+1}. \tag{13.4}$$

The matrices used by Abelès link electric and magnetic field components at successive layers of the stratification. For non-absorbing media these matrices are complex, with real diagonal elements and imaginary off-diagonal elements. Matrices which link fields and their derivatives (for example E and $\mathrm{d}E/\mathrm{d}z$ for the electromagnetic s or TE wave) are entirely real for non-absorbing media, as we saw in Sect. 12.2. This is both simpler, and four times faster in numerical work (the matrix product $AB = (A_r + iA_i)(B_r + iB_i)$ requires the evaluation of four products if A and B are complex). The matrices which link fields and their derivatives are also unimodular. In this chapter we shall follow Lekner (1994) to both simplify and generalize the existing theory of light propagation in periodically stratified media. An expression is given for the matrix of a layer with continuous but otherwise arbitrary dielectric function variation. The eigenvalue equation for the Bloch factor in a periodic system is shown to be determined by the trace of the matrix of a unit cell. When the wavelength is long compared to the period of the stratification, the periodic structure is equivalent to a uniaxial homogeneous medium with its optic axis normal to the layers and with the ordinary dielectric constant equal to the average of the dielectric function, while the extraordinary dielectric constant is equal to the reciprocal of the average of the reciprocal of the dielectric function.

13.1 Electromagnetic Waves in Stratified Media

We consider plane electromagnetic waves incident from a medium of index n_1 onto a non-magnetic planar stratification, whose optical properties are contained in the dielectric function $\varepsilon(z) = n^2(z)$ ($n(z)$ is the local value of the refractive index). For isotropic media, with scalar rather than tensor dielectric function, any plane wave can be written as a superposition of an s (or TE) wave and a p (or TM) wave. The s wave has its electric vector perpendicular to the plane of incidence, the p wave has its electric vector in the plane of incidence (and its magnetic vector perpendicular to the plane of incidence: hence its designation as a TM or transverse magnetic). We have assumed, as usual, that the medium is stratified in the z direction (so that $\varepsilon = \varepsilon(z)$), and have taken the plane of incidence to be the zx plane. Then the s wave has electric field vector $E = (0, E_y, 0)$ and the p wave has magnetic vector $B = (0, B_y, 0)$. It follows directly from the Maxwell curl equations that, for monochromatic waves of angular frequency ω,

$$\begin{aligned} E_y(z,x,t) &= \exp[i(Kx - \omega t)]E(z), \\ B_y(z,x,t) &= \exp[i(Kx - \omega t)]B(z), \end{aligned} \tag{13.5}$$

where K (the x component of the wavevector) is a separation-of-variables constant, whose existence derives from the planar nature of the stratification, and whose

constancy implies Snell's law, as we saw in Chap. 1. The functions $E(z)$ and $B(z)$ satisfy the ordinary differential equations (see Sects. 1.1 and 1.2)

$$\frac{\mathrm{d}^2E}{\mathrm{d}z^2} + q^2E = 0, \qquad \varepsilon\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{\varepsilon}\frac{\mathrm{d}B}{\mathrm{d}z}\right) + q^2B = 0, \tag{13.6}$$

where $q(z)$ is the local value of the normal component of the wavevector, given by

$$q^2(z) = \varepsilon(z)\omega^2/c^2 - K^2.$$

If θ_1 is the angle of incidence, and θ_2 is the angle between the wavevector and the normal in the homogeneous substrate of index n_2,

$$K = n_1(\omega/c)\sin\theta_1 = n_2(\omega/c)\sin\theta_2, \tag{13.7}$$

$$q_1 = n_1(\omega/c)\cos\theta_1, \qquad q_2 = n_2(\omega/c)\cos\theta_2. \tag{13.8}$$

It follows from the differential equations (13.6) that $\mathrm{d}E/\mathrm{d}z$ and $\varepsilon^{-1}\mathrm{d}B/\mathrm{d}z$ are continuous at any discontinuity in $\varepsilon(z)$ (otherwise delta-function terms would arise in the second derivatives). A fortiori, E and B are continuous at any discontinuity in $\varepsilon(z)$.

Consider a stratification extending from $z = a$ to $z = b$, bounded by homogeneous media of indices n_1 and n_2, and suppose at first that $\varepsilon(z)$ is continuous for $a < z < b$. For the s wave, let $F(z)$ and $G(z)$ be two linearly independent solutions of the second-order differential equation $\mathrm{d}^2E/\mathrm{d}z^2 + q^2E = 0$. Then $E(z)$ may be written as a linear superposition of F and G:

$$E(z) = fF(z) + gG(z), \tag{13.9}$$

where f and g are constants. We will use a layer matrix $M = \{m_{ij}\}$ which links fields and their derivatives; in the s wave case it is defined by

$$\begin{pmatrix} E_b \\ E_b' \end{pmatrix} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{pmatrix} E_a \\ E_a' \end{pmatrix}, \tag{13.10}$$

where E_a and E_a' represent $E(a+)$ and the derivative of $E(z)$ at $z = a+$, and similarly E_b and E_b' stand for $E(b-)$ and the derivative of E at $z = b^-$. From (13.9) and its derivative we see that

$$\begin{pmatrix} E_a \\ E_a' \end{pmatrix} = \begin{bmatrix} F_a & G_a \\ F_a' & G_a' \end{bmatrix} \begin{pmatrix} f \\ g \end{pmatrix} \equiv A \begin{pmatrix} f \\ g \end{pmatrix}, \tag{13.11}$$

$$\begin{pmatrix} E_b \\ E_b' \end{pmatrix} = \begin{bmatrix} F_b & G_b \\ F_b' & G_b' \end{bmatrix} \begin{pmatrix} f \\ g \end{pmatrix} \equiv B \begin{pmatrix} f \\ g \end{pmatrix}. \tag{13.12}$$

Substitution of (13.11) and (13.12) into (13.10) shows that the layer matrix can be expressed in terms of the fundamental field and derivative values at the boundaries of the layer:

$$M = BA^{-1} = W^{-1}\begin{bmatrix} -(F',G) & (F,G) \\ -(F',G') & (F,G') \end{bmatrix}. \tag{13.13}$$

W is the (constant) Wronksian of the two basic solutions F and G,

$$W = FG' - F'G, \qquad W' = 0, \tag{13.14}$$

and

$$\begin{array}{ll} (F,G) \equiv F_a G_b - G_a F_b, & (F,G') \equiv F_a G'_b - G_a F'_b, \\ (F',G) \equiv F'_a G_b - G'_a F_b, & (F',G') \equiv F'_a G'_b - G'_a F'_b. \end{array} \tag{13.15}$$

The layer matrix M is unimodular: from the identity (2.31) of Chap. 2,

$$\det M = W^{-2}\Big[(F,G)(F',G') - (F,G')(F',G)\Big] = 1. \tag{13.16}$$

An important example is that of a homogeneous layer, for which $\varepsilon(z)$ and $q(z)$ are constant. We can then take $F = \cos qz$, $G = \sin qz$, for which $W = q$ and

$$\begin{array}{ll} (F,G) = \sin\delta, & (F,G') = q\cos\delta, \\ (F',G) = -q\cos\delta, & (F',G') = q^2\sin\delta, \end{array} \tag{13.17}$$

where $\delta = q(b-a)$ is the phase increment across the layer. The layer matrix in this case is

$$M = \begin{bmatrix} \cos\delta & q^{-1}\sin\delta \\ -q\sin\delta & \cos\delta \end{bmatrix} \tag{13.18}$$

(The same matrix is obtained by choosing $F = \exp(iqz)$ and $G = \exp(-iqz)$.)

The matrix elements m_{ij} determine the reflection and transmission amplitudes r_s and t_s of the $n_1|n(z)|n_2$ structure: we have, from the definition of M,

$$\begin{pmatrix} t_s\exp(iq_2b) \\ iq_2t_s\exp(iq_2b) \end{pmatrix} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}\begin{pmatrix} \exp(iq_1a) + r_s\exp(-iq_1a) \\ iq_1[\exp(iq_1a) - r_s\exp(-iq_1a)] \end{pmatrix} \tag{13.19}$$

when $\exp(iq_1z) + r_s\exp(-iq_1z)$ and $t_s\exp(iq_2z)$ are the forms of $E(z)$ in the medium of incidence and in the substrate. It follows from (13.19) that (see Sect. 12.2)

$$r_s = \exp(2iq_1a)\frac{q_1q_2m_{12} + m_{21} + iq_1m_{22} - iq_2m_{11}}{q_1q_2m_{12} - m_{21} + iq_1m_{22} + iq_2m_{11}}, \tag{13.20}$$

$$t_s = \exp[i(q_1a - q_2b)]\frac{2iq_1}{q_1q_2m_{12} - m_{21} + iq_1m_{22} + iq_2m_{11}}. \tag{13.21}$$

The form of the differential equation for $E(z)$ is the same as that of the Schrödinger equation for a particle of energy $\mathcal{E}$ and mass m in a potential $V(z)$, with

$$\varepsilon(z)\omega^2/c^2 \leftrightarrow \frac{2m}{\hbar^2}[\mathcal{E} - V(z)]. \tag{13.22}$$

Thus the results derived for the electromagnetic s wave apply also to quantum particle waves in a z-stratified medium, as discussed in Sect. 1.3.

The p wave layer matrix is defined to link the quantities $B(z)$ and $\varepsilon^{-1}\mathrm{d}B/\mathrm{d}z$ which are continuous at discontinuities of ε. Let B_a and B_b stand for $B(a+)$ and $B(b^-)$, and $\tilde{B}_a$ and $\tilde{B}_b$ represent the values of $\varepsilon^{-1}\mathrm{d}B/\mathrm{d}z$ at $z = a+$ and $z = b^-$. Then

$$\begin{pmatrix} B_b \\ \tilde{B}_b \end{pmatrix} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}\begin{pmatrix} B_a \\ \tilde{B}_a \end{pmatrix}. \tag{13.23}$$

We express $B(z)$ as a linear combination of two independent solutions of the second equation of (13.6), say $C(z)$ and $D(z)$. Then, by the arguments used above, with F, F' and G, G' replaced by $C, \tilde{C}$ and D, $\tilde{D}$,

$$M = U^{-1}\begin{bmatrix} -(\tilde{C}, D) & (C, D) \\ -(\tilde{C}, \tilde{D}) & (C, \tilde{D}) \end{bmatrix} \tag{13.24}$$

where, for example, $(C, \tilde{D}) \equiv C_a\tilde{D}_b - D_a\tilde{C}_b$ and

$$U = C\tilde{D} - \tilde{C}D, \qquad U' = 0 \tag{13.25}$$

(the Wronskian $W = CD' - C'D$ is not constant for the p wave; $U = W/\varepsilon$ is constant). This matrix is also unimodular, since

$$\det M = U^{-2}\left[(C, D)(\tilde{C}, \tilde{D}) - (C, \tilde{D})(\tilde{C}, D)\right] = 1 \tag{13.26}$$

For the homogeneous layer we take $C = \cos qz$ and $D = \sin qz$ (or $\exp(iqz)$ and $\exp(-iqz)$) to find that $U = Q = q/\varepsilon$ and

$$M = \begin{bmatrix} \cos\delta & Q^{-1}\sin\delta \\ -Q\sin\delta & \cos\delta \end{bmatrix} \tag{13.27}$$

The reflection and transmission amplitudes are defined slightly differently for the p wave if one wishes to retain $r_p = r_s$ and $t_p = t_s$ at normal incidence, where there is no physical difference between the s and p waves:

$$\exp(iq_1 z) - r_p \exp(-iq_1 z) \leftarrow B(z) \rightarrow \frac{n_2}{n_1} t_p \exp(iq_2 z). \tag{13.28}$$

Thus the equation analogous to (13.19) reads, from (13.23),

$$\begin{pmatrix} \frac{n_2}{n_1} t_p \exp(iq_2 b) \\ iQ_2 \frac{n_2}{n_1} t_p \exp(iq_2 b \end{pmatrix} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{pmatrix} \exp(iq_1 a) - r_p \exp(-iq_1 a) \\ iQ_1 \left[\exp(iq_1 a) + r_p \exp(-iq_1 a)\right] \end{pmatrix}, \tag{13.29}$$

where $Q_1 = q_1/\varepsilon_1$, and $Q_2 = q_2/\varepsilon_2$. This gives (Sect. 12.2)

$$-r_p = \exp(2iq_1 a) \frac{Q_1 Q_2 m_{12} + m_{21} + iQ_1 m_{22} - iQ_2 m_{11}}{Q_1 Q_2 m_{12} - m_{21} + iQ_1 m_{22} + iQ_2 m_{11}}, \tag{13.30}$$

$$\frac{n_2}{n_1} t_p = \exp[i(q_1 a - q_2 b)] \frac{2iQ_1}{Q_1 Q_2 m_{12} - m_{21} + iQ_1 m_{22} + iQ_2 m_{11}}. \tag{13.31}$$

It is clear from the definition of the layer matrix that a stratification of any number N of layers has the matrix

$$M = M_N M_{N-1} \ldots M_2 M_1. \tag{13.32}$$

The results for the reflection and transmission amplitudes given above thus apply to any isotropic stratification. For nonabsorbing media $\varepsilon(z)$ is real, and the s- and p-wave basic solutions can be taken to be real (if ψ is a solution of a linear differential equation with real coefficients, then ψ^* is also a solution and so is $\psi + \psi^*$). Thus the matrices are real in the absence of absorption. Energy conservation in the absence of absorption is expressed in the algebraic identities

$$R_s + T_s = 1, \qquad R_p + T_p = 1, \tag{13.33}$$

where $R_s = |r_s|^2$, $R_p = |r_p|^2$, $T_s = (q_2/q_1)|t_s|^2$, and $T_p = (q_2/q_1)|t_p|^2$. (The reason for the q_2/q_1 factor is discussed in Sect. 2.1; see especially Fig. 2.1.)

13.2 Periodic Structures, Multilayer Dielectric Mirrors

We now consider periodic stratifications, such as the high-low multilayer mirror configuration shown in Fig. 13.1. The reflection at normal incidence of this multilayer was considered in Sect. 12.3, in the maximum reflectivity configuration (quarter wave stack).

We first discuss propagation of waves in an infinite periodic structure. If one period has matrix M, the fields and their derivatives at a corresponding point one period along are given by

$$\begin{pmatrix} \psi_{n+1} \\ \psi'_{n+1} \end{pmatrix} = M \begin{pmatrix} \psi_n \\ \psi'_n \end{pmatrix}, \tag{13.34}$$

where ψ represents E or B and ψ' represents E' or $\tilde{B} = \varepsilon^{-1} \mathrm{d}B/\mathrm{d}z$. In an infinite structure these positions (one period along from each other) are equivalent, and so the two vectors in (13.34) are proportional:

$$\begin{pmatrix} \psi_{n+1} \\ \psi'_{n+1} \end{pmatrix} = \beta \begin{pmatrix} \psi_n \\ \psi'_n \end{pmatrix}. \tag{13.35}$$

The Bloch factor β is determined from the condition that (13.35) subtracted from (13.34), namely,

$$(M - \beta I) \begin{pmatrix} \psi_n \\ \psi'_n \end{pmatrix} = 0, \tag{13.36}$$

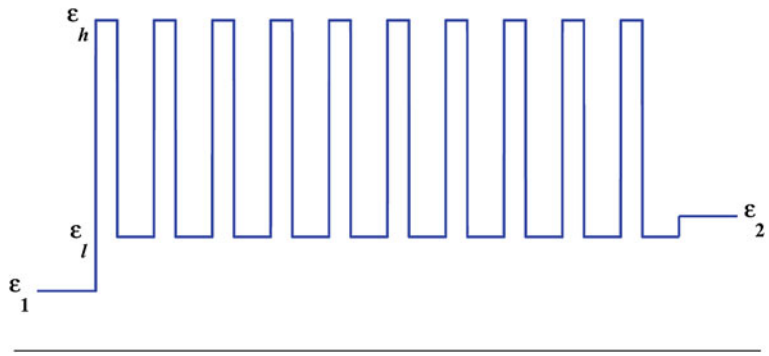


Fig. 13.1 Dielectric function profile for a $(\mathrm{HL})^{10}$ dielectric mirror, drawn to scale with $n_1 = 1$, $n_h = 2.35$ (ZnS), $n_l = 1.38$ (MgF_2), $n_2 = 1.5$ (glass). For maximum reflectivity at normal incidence and vacuum wavelength λ, the layer thicknesses are $d_h = \lambda/4n_h$ and $d_l = \lambda/4n_l$ (a quarter-wave stack)

has a solution other than zero. ($I = \mathrm{diag}(1, 1)$ is the 2×2 identity matrix.) The condition for nonzero solutions is $\det(M - \beta I) = 0$. When $\mathrm{trace}\, M = 2\cos\phi$ and $\det M = 1$ are used this condition reduces to

$$\beta^2 - 2\beta\cos\phi + 1 = 0. \tag{13.37}$$

The quadratic (13.37) has solutions

$$\beta_\pm = \cos\phi \pm (\cos^2\phi - 1)^{1/2} = \exp(\pm i\phi). \tag{13.38}$$

Note that β_+ and β_- both have unit modulus if $\cos^2\phi < 1$, but that if $\cos^2\phi > 1$ the solutions are real (in the absence of absorption) and not equal to unity. If the magnitude of the trace of M exceeds 2, the solutions will grow or decay exponentially: no propagating waves are possible. The condition $\cos^2\phi > 1$ thus gives the *band gaps* or *stop bands* of the structure. The band edges are given by $\cos^2\phi = 1$; they occur when ϕ is a multiple of π. When $|\cos\phi| > 1$, ϕ is complex, with the real part a multiple of π, and the imaginary part $\phi_i = \mathrm{Im}\,\phi$ given by (in the absence of absorption, $\mathrm{trace}\, M = 2\cos\phi$ real)

$$\exp(\phi_i) = |\cos\phi| + (\cos^2\phi - 1)^{1/2}. \tag{13.39}$$

(There is an ambiguity in the sign of ϕ_i (which we can ignore here) since when $\phi_r = n\pi$, $\cos(\phi_r + i\phi_i) = \cos n\pi \cosh\phi_i$. See Hardy (1952, pp. 464–465).)

In infinite periodic stratifications the possibility of wave propagation is entirely determined by the trace of the matrix for a single period. We may expect (and we shall shortly show this to be true) that finite periodic structures reflect strongly in the stop bands. The ϕ_s and ϕ_p values for the high-low stack of Fig. 13.1, repeated to infinity, are shown in Fig. 13.2. They are calculated from the matrix of the unit cell, a high-low bilayer, which for the s wave is given by

$$\begin{aligned} &\begin{pmatrix} c_l & q_l^{-1}s_l \\ -q_l s_l & c_l \end{pmatrix}\begin{pmatrix} c_h & q_h^{-1}s_h \\ -q_h s_h & c_h \end{pmatrix} \\ &= \begin{pmatrix} c_l c_h - q_l^{-1}q_h s_l s_h & q_h^{-1}c_l s_h + q_l^{-1}s_l c_h \\ -q_l s_l c_h - q_h c_l s_h & c_l c_h - q_l q_h^{-1} s_l s_h \end{pmatrix}, \end{aligned} \tag{13.40}$$

where $c_l = \cos\delta_l$, $s_l = \sin\delta_l$, $q_l = (\varepsilon_l\omega^2/c^2 - K^2)^{1/2}$, and the phase increment δ_l is $q_l d_l$ where d_l is the thickness of the low-index layer; the parameters for the high-index layer are defined in the same way. One half of the trace of this unit cell matrix is, for the s wave,

$$\begin{aligned} \cos\phi_s &= c_l c_h - \tfrac{1}{2}s_l s_h(q_l^{-1}q_h + q_h^{-1}q_l) \\ &= \cos(\delta_l + \delta_h) - \tfrac{1}{2}s_l s_h\left[(q_l/q_h)^{1/2} - (q_h/q_l)^{1/2}\right]^2. \end{aligned} \tag{13.41}$$

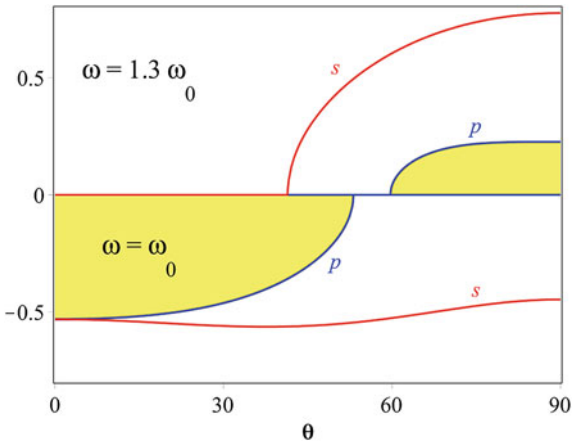


Fig. 13.2 The imaginary parts of ϕ_s and ϕ_p for the ZnS|MgF_2 high|low structure, as a function of the angle of incidence. The *lower part* of the figure is at the design frequency ω_0 for high reflectivity at normal incidence, at which $d_h = \lambda_h/4$ and $d_l = \lambda_l/4$ (the $\lambda/4$ stack), and thus $\delta_h = \pi/2 = \delta_l$. The *upper part* is drawn for $\omega = 1.3\omega_0$. We have chosen opposite signs for Im ϕ for the two frequencies for clarity (see the note below (13.39) regarding the sign of Im ϕ)

For the p wave the arguments of the trigonometric functions remain unchanged, but q_l is replaced by $Q_l = q_l/\varepsilon_l$ and q_h by $Q_h = q_h/\varepsilon_h$ in the matrix elements. Thus

$$\begin{aligned}\cos\phi_p &= c_l c_h - \tfrac{1}{2}s_l s_h(Q_l^{-1}Q_h + Q_h^{-1}Q_l)\\ &= \cos(\delta_l+\delta_h) - \tfrac{1}{2}s_l s_h\left[(Q_l/Q_h)^{1/2} - (Q_h/Q_l)^{1/2}\right]^2.\end{aligned} \tag{13.42}$$

Figure 13.2 shows the imaginary parts of ϕ_s and ϕ_p as a function of the angle of incidence θ_1. When these are non-zero, the corresponding polarizations cannot propagate in the semi-infinite periodic medium. We see from the Figure that at the design frequency ω_0 for high reflectivity, the infinite high-low stack does not permit s wave propagation at any angle of incidence, while the p wave can propagate for $\theta_1 \geq 53.14°$. At $\omega = 1.3\omega_0$ both polarizations can propagate into the stack near normal incidence, but at 41.44° the s polarization begins to reflect totally, as does the p polarization at 59.72°. The band edges at which this happens are given by the locations of $\cos^2\phi = 1$.

The variation of ϕ as a function of frequency is shown in Fig. 13.3, in which we plot the real and imaginary parts of ϕ versus ω at normal incidence. The stop band ($\cos^2\phi > 1$, ϕ complex) is between $\omega_0 - \Delta\omega$ and $\omega_0 + \Delta\omega$, where (Sect. 12.3, (12.66))

$$\frac{\Delta\omega}{\omega_0} = \frac{2}{\pi}\arcsin\left(\frac{n_h - n_l}{n_h + n_l}\right). \tag{13.43}$$

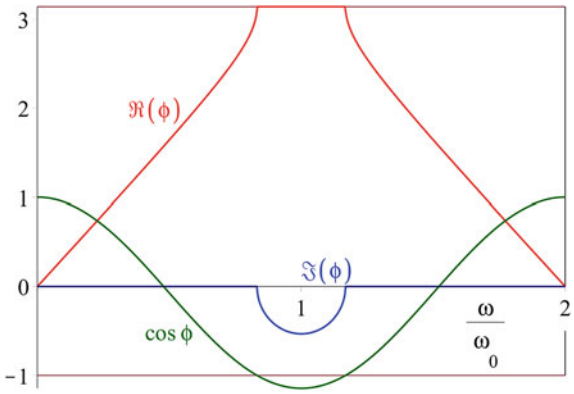


Fig. 13.3 The real and imaginary parts of the band structure parameter $\phi = \arccos\left(\frac{1}{2}\text{trace}\, M\right)$ for a $\lambda/4$ stack, as a function of the frequency, drawn for normal incidence onto the high-low stack of Fig. 13.1. The stop band is centred at the design angular frequency ω_0, with half-width given by (13.43). It corresponds to the regions where $\cos^2\phi > 1$; in this example, $\cos\phi < -1$

(The repeated stack of Fig. 13.1 has $\Delta\omega/\omega_0 \approx 0.16748$.) At normal incidence the phase increments δ_h and δ_l for the quarter-wave stack are both equal to $(\pi/2)(\omega/\omega_0)$. Thus $\cos\phi$ is periodic in ω, with period $2\omega_0$. At oblique incidence the s and p waves have different stop bands, determined by a transcendental equation to be given in Sect. 13.3, equation (13.67).

We now look at the optical properties of *finite* periodic structures. The reflection and transmission amplitudes are given by (13.20) and (13.21) for the s wave, and (13.30) and (13.31) for the p wave, where in each case m_{ij} are the matrix elements of the whole structure. Thus for N periods (for example, N bilayers of the high-low stack) the matrix elements are those of the Nth power of the unit cell matrix, and are given by (13.2). It is convenient to define the quantity

$$\sigma_N = \frac{S_{N-1}}{S_N} = \frac{\sin[(N-1)\phi]}{\sin(N\phi)} = \cos\phi - \sin\phi\cot(N\phi), \tag{13.44}$$

where $\cos\phi$ is half the trace of the unit cell matrix. Then we have, for the s wave,

$$r_s = \frac{q_1 q_2 m_{12} + m_{21} + iq_1(m_{22} - \sigma_N) - iq_2(m_{11} - \sigma_N)}{q_1 q_2 m_{12} - m_{21} + iq_1(m_{22} - \sigma_N) + iq_2(m_{11} - \sigma_N)}, \tag{13.45}$$

$$t_s = \frac{2iq_1 S_N^{-1}}{q_1 q_2 m_{12} - m_{21} + iq_1(m_{22} - \sigma_N) + iq_2(m_{11} - \sigma_N)}. \tag{13.46}$$

We have omitted the phase factors multiplying r_s and t_s; these are the same for the p wave reflection and transmission coefficients, and do not feature in any experiment that does not compare reflection and transmission phases.

For non-absorbing media the reflectance and transmittance are given by

$$R_s = |r_s|^2 = 1 - \frac{4q_1q_2S_N^{-2}}{(q_1q_2m_{12})^2 + m_{21}^2 + q_1^2(m_{22}-\sigma_N)^2 + q_2^2(m_{11}-\sigma_N)^2 + 2q_1q_2S_N^{-2}}, \tag{13.47}$$

$$T_s = \frac{q_2}{q_1}|t_s|^2 = 1 - R_s. \tag{13.48}$$

The p wave has a different unit cell matrix, and thus different ϕ and σ_N. The reflection and transmission amplitudes are, again omitting the phase factors,

$$-r_p = \frac{Q_1Q_2m_{12} + m_{21} + iQ_1(m_{22}-\sigma_N) - iQ_2(m_{11}-\sigma_N)}{Q_1Q_2m_{12} - m_{21} + iQ_1(m_{22}-\sigma_N) + iQ_2(m_{11}-\sigma_N)}, \tag{13.49}$$

$$\frac{n_2}{n_1}t_p = \frac{2iQ_1S_N^{-1}}{Q_1Q_2m_{12} - m_{21} + iQ_1(m_{22}-\sigma_N) + Q_2(m_{11}-\sigma_N)}. \tag{13.50}$$

For non-absorbing media the reflectance is $R_p = 1 - T_p$, where

$$T_p = \frac{q_2}{q_1}|t_p|^2 = \frac{4Q_1Q_2S_N^{-2}}{(Q_1Q_2m_{12})^2 + m_{21}^2 + Q_1^2(m_{22}-\sigma_N)^2 + Q_2^2(m_{11}-\sigma_N)^2 + 2Q_1Q_2S_N^{-2}}. \tag{13.51}$$

The forms for the reflectance and transmittance have been obtained by using $\det M = 1$, $\operatorname{trace} M = 2\cos\phi$ (M is the matrix for a unit cell) and the identity

$$\sigma_N^2 - 2\sigma_N\cos\phi + 1 = \left[\frac{\sin\phi}{\sin N\phi}\right]^2 = S_N^{-2}. \tag{13.52}$$

When $N\phi$ is a multiple of π and $(N-1)\phi$ is not, σ_N is infinite and

$$r_s \to \frac{q_1 - q_2}{q_1 + q_2}, \qquad -r_p \to \frac{Q_1 - Q_2}{Q_1 + Q_2}. \tag{13.53}$$

These are the reflection amplitudes of the bare substrate. When $(N-1)\phi$ is a multiple of π and $N\phi$ is not, σ_N is zero and r_s and r_p are the same as the reflection amplitudes of a single period of the structure (supported by the substrate). Thus for large N there will be many passes of the reflectivity through the bare-substrate and single-period values as the wavelength or angle of incidence varies.

At the band edges, where $\cos\phi = \pm 1$ and ϕ is a multiple of π, $S_N^2 = N^2$. Thus the transmittance goes to zero as N^{-2} at the band edges, and the reflectance is $1 - O(N^{-2})$. Within the band gaps $\cos\phi = \frac{1}{2}\text{trace}\, M$ has magnitude greater than unity, and ϕ is a multiple of π plus an imaginary part given by (13.39):

$$\text{Im}\,\phi = \ln\left[|\cos\phi| + (\cos^2\phi - 1)^{1/2}\right]. \tag{13.54}$$

Then S_N^2 increases exponentially with N,

$$S_N^2 = \left[\frac{\sinh(N\,\text{Im}\,\phi)}{\sinh(\text{Im}\,\phi)}\right]^2, \tag{13.55}$$

and thus the s and p transmittances tend to zero exponentially with the number of periods.

The results (13.45)–(13.55) hold for waves in any finite periodic stratification. In particular the facts that $1 - R = O(N^{-2})$ at the band edges, and that inside the stop bands R approaches unity exponentially with N, are universal. The construction of the matrices does not assume homogeneity within parts of a unit cell (as is assumed in Yeh, Yariv and Hong (1977) and Yariv and Yeh (1977), for example).

Figure 13.4 shows the normal incidence reflectivity for a 10-bilayer high-low stack as a function of frequency, and Fig. 13.5 the s and p reflectivities for the same 10-bilayer stack at $\omega = \omega_0$ and $\omega = 1.3\omega_0$, as a function of the angle of incidence. The stack parameters are as in Figs. 13.1, 13.2 and 13.3. The normal incidence reflectivity for a 4-bilayer stack was shown in Fig. 12.3.

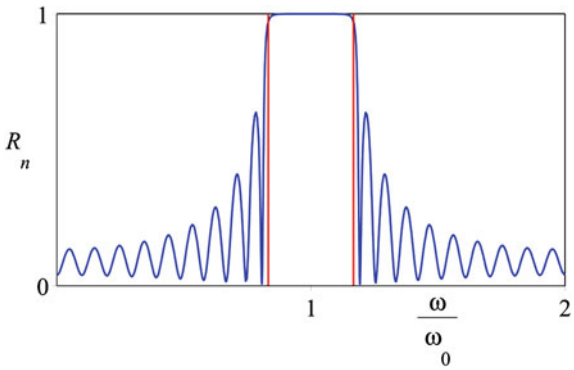


Fig. 13.4 Frequency dependence of the normal incidence reflectivity of a dielectric multilayer, drawn for 10 ZnS|MgF_2 bilayers on glass, as depicted in Fig. 13.1. The multilayer is tuned for high reflectivity at $\omega = \omega_0$ (it is a quarter-wave stack at the design frequency). The *vertical lines* denote the stop-band limits, given in (13.43)

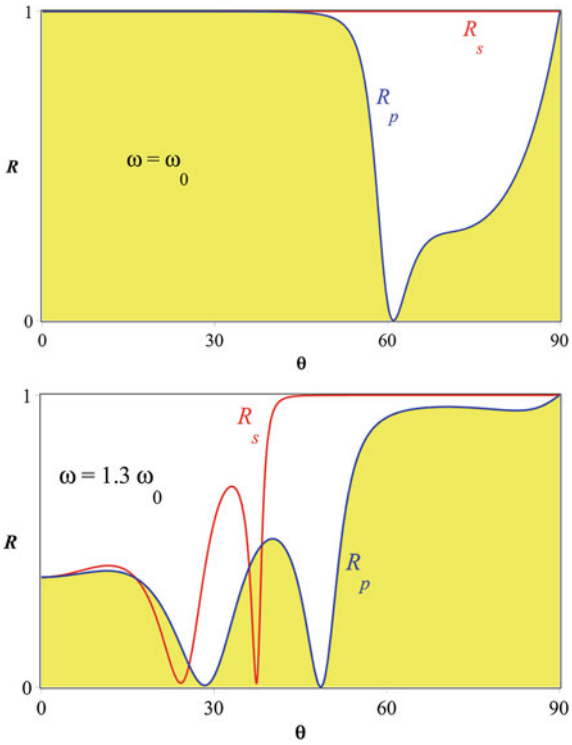


Fig. 13.5 Reflectivities of a 10-bilayer high-low stack as a function of the angle of incidence, at $\omega = \omega_0$ and $\omega = 1.3\omega_0$. The parameters are as in Figs. 13.1, 13.2, 13.3 and 13.4. The stop band edges are at 53.1° for the p waves when $\omega = \omega_0$, and 41.1° for the s wave and 59.7° for the p wave when $\omega = 1.3\omega_0$, from Fig. 13.2

13.3 Omnidirectional Reflection by Multilayer Dielectric Mirrors

We have seen that multilayer dielectric mirrors reflect strongly in *stop bands*, within which light propagation is not possible in an infinite periodic structure. We have also seen that certain aspects of reflection by multilayers are universal: for N periods the transmittance goes to zero as N^{-2} at the band edges, where the reflectance is $1 - O(N^{-2})$. Within the band gaps the reflectance tends to unity exponentially with N. Thus not very many periods of the multilayer structure are needed to give high reflectance. Typical use of dielectric multilayer mirrors has been at normal incidence; the reflectance for a 10 period stack was shown in Fig. 13.4, with the (homogeneous) layers a quarter-wavelength thick at the design frequency:

$$d_h = \frac{\lambda_h}{4} = \frac{\lambda}{4n_h}, \qquad d_l = \frac{\lambda_l}{4} = \frac{\lambda}{4n_l}, \tag{13.56}$$

where λ is the vacuum wavelength. Then the optical paths are $n_h d_h = n_l d_l = \lambda/4$ and maximum reflection (at normal incidence) occurs at angular frequency

$$\omega_0 = \frac{2\pi c}{\lambda} = \frac{\pi}{2}\frac{c}{n_h d_h} = \frac{\pi}{2}\frac{c}{n_l d_l}. \tag{13.57}$$

The edges of the stop band are at $\omega_0 \pm \Delta\omega$, where $\Delta\omega/\omega_0$ was given in (13.43).

The normal incidence results are independent of polarization, but at oblique incidence the reflectances of the s (TE) and p (TM) polarizations are of course different. The question arises: is it possible to design a stack to have strong reflection (perfect reflection, for the infinite stack) for *all angles of incidence*, and *both polarizations?* The answer is *yes* (Winn et al. 1998; Fink et al. 1998; Chigrin et al. 1999a, b; Nusinsky and Hardy 2007). Southwell (1999) has given analytical approximations for an omnidirectional mirror consisting of a quarter-wave dielectric stack. In this section we present improved analytical approximations which give the band edges of the s and p stop bands at any angle of incidence, on a dielectric stack which need not be quarter-wave, based on Lekner (2000).

We know from Sect. 13.2 that strong reflection will occur when the trace of the 2×2 matrix for one period exceeds 2 in magnitude. (These conditions, one for the s polarization and one for the p polarization, locate the band edges of the s and p waves.) For homogeneous layers of high and low refractive indices and n_l, and thicknesses d_h and d_l, these conditions take the form

$$|\cos\delta_l \cos\delta_h - \Lambda \sin\delta_l \sin\delta_h| > 1, \tag{13.58}$$

where

$$\delta_l = \frac{\omega d_l}{c}\sqrt{n_l^2 - n_1^2 \sin^2\theta}, \qquad \delta_h = \frac{\omega d_h}{c}\sqrt{n_h^2 - n_1^2 \sin^2\theta}, \tag{13.59}$$

are the phase shifts of the waves of angular frequency ω in traversing the layers of low and high index, n_1 is the refractive index of the medium of incidence, and θ is the angle of incidence. The function Λ is frequency-independent, and takes different forms for the s and p polarizations:

$$\Lambda_s = \frac{1}{2}\left(x_s + \frac{1}{x_s}\right), \qquad x_s = \frac{q_h}{q_l} = \sqrt{\frac{n_h^2 - n_1^2 \sin^2\theta}{n_l^2 - n_1^2 \sin^2\theta}}, \tag{13.60}$$

$$\Lambda_p = \frac{1}{2}\left(x_p + \frac{1}{x_p}\right), \qquad x_p = \frac{Q_l}{Q_h} = \left(\frac{n_h}{n_l}\right)^2 x_s. \tag{13.61}$$

We assume that $n_h > n_l$, so $x_s > 1$. Note that x_p can be less than unity for angles of incidence greater than θ_p, where $\sin^2\theta_p = n_l^2 n_h^2/n_1^2(n_l^2 + n_h^2)$, provided $n_l^2 n_h^2 < n_1^2(n_l^2 + n_h^2)$.

From (13.56), (13.57) and (13.59), a quarter-wave stack at normal incidence has

$$\delta_l = \frac{\pi}{2}\frac{\omega}{\omega_0} = \delta_h, \tag{13.62}$$

and $\cos\delta_l \cos\delta_h - \Lambda \sin\delta_l \sin\delta_h$ becomes $\cos^2\delta - \frac{1}{2}\left(\frac{n_h}{n_l} + \frac{n_l}{n_h}\right)\sin^2\delta$, which takes the value -1 at $\omega_0 \pm \Delta\omega$, where $\Delta\omega$ is given by (13.43).

13.3.1 Band Edges at Oblique Incidence for a General Stack

At normal incidence the stop band for a quarter-wave stack lies between $\omega_0^- = \omega_0 - \Delta\omega$ and $\omega_0^+ = \omega_0 + \Delta\omega$, where ω_0 and $\Delta\omega$ are given by (13.57) and (13.43) respectively. The quantity $\cos\delta_l \cos\delta_h - \Lambda \sin\delta_l \sin\delta_h$ in (13.58) decreases with frequency from unity at zero frequency, so the first stop bands lie between frequencies ω^- and ω^+ given by solving the transcendental equation

$$\cos\delta_l \cos\delta_h - \Lambda \sin\delta_l \sin\delta_h = -1 \tag{13.63}$$

numerically for $\Lambda = \Lambda_s$ and Λ_p as given by (13.60) and (13.61). At normal incidence $\Lambda_p = \Lambda_s$ and the stop band for both polarizations lies between ω_0^- and ω_0^+. The s polarization stop band typically increases in width as the angle of incidence increases, while the p stop band width decreases. At glancing incidence the p stop band ranges from ω_p^- to ω_p^+, and provided

$$\omega_p^- < \omega_0^+ \tag{13.64}$$

there will be a frequency region from ω_p^- to ω_0^+ where both s and p polarizations are strongly reflected (perfectly reflected, in the case of an infinite stack).

At oblique incidence on a general stack, $\delta_l = (\omega/c)D_l$ and $\delta_h = (\omega/c)D_h$, where

$$D_l = d_l\sqrt{n_l^2 - n_1^2\sin^2\theta}, \qquad D_h = d_h\sqrt{n_h^2 - n_1^2\sin^2\theta}. \tag{13.65}$$

At normal incidence for a quarter-wave stack, the phase increments δ_l and δ_h at the band edges are both $\pi/2 \pm \arcsin\left(\frac{x_0-1}{x_0+1}\right)$, where $x_0 = n_h/n_l$ is the common value of x_s and x_p at $\theta = 0$. At all frequencies the phase increments are in the ratio $\delta_h/\delta_l = D_h/D_l$. We therefore put the phase shifts at the band edges equal to

$$\delta_l^{\pm} = \frac{2D_l}{D_h + D_l}\left(\frac{\pi}{2} \pm \phi^{\pm}\right), \qquad \delta_h^{\pm} = \frac{2D_h}{D_h + D_l}\left(\frac{\pi}{2} \pm \phi^{\pm}\right) \tag{13.66}$$

where the angles $\phi^{\pm}$ are to be found for each polarization. The transcendental equation (13.63) for the band edges now reduces to an equation for $\phi^{\pm}$:

$$\sin\phi^{\pm} = \left(\frac{x-1}{x+1}\right)\cos\left[\left(\frac{D_h - D_l}{D_h + D_l}\right)\left(\frac{\pi}{2} \pm \phi^{\pm}\right)\right]. \tag{13.67}$$

In deriving (13.67) from (13.63) we have used

$$\frac{\Lambda - 1}{\Lambda + 1} = \left(\frac{x-1}{x+1}\right)^2 \tag{13.68}$$

(x stands for x_p or x_s, Λ stands for Λ_p or Λ_s) and taken a square root on the assumption that $x > 1$. No approximation has yet been made. When $D_h = D_l$ we obtain a generalization of (13.43):

$$\phi^{\pm} = \arcsin\left(\frac{x-1}{x+1}\right). \tag{13.69}$$

It follows from (13.67) that the stop band width is greatest when $D_h = D_l$, because then the cosine term then attains its maximum value of unity.

In many cases of interest $(D_h - D_l)/(D_h + D_l)$ is a small quantity. Expansion of the right-hand side of (13.67) gives

$$\sin\phi^{\pm} = \left(\frac{x-1}{x+1}\right)\left\{1 - \frac{1}{2}\left(\frac{D_h - D_l}{D_h + D_l}\right)^2\left(\frac{\pi}{2} \pm \phi^{\pm}\right)^2 + O\left(\frac{D_h - D_l}{D_h + D_l}\right)^4\right\}, \tag{13.70}$$

so that, on using $\arcsin(S+s) = \arcsin(S) + s\sqrt{1-S^2} + O(s^2)$,

$$\begin{aligned}\phi^{\pm} = {} & \arcsin\left(\frac{x-1}{x+1}\right) - \frac{x-1}{4\sqrt{x}}\left(\frac{D_h - D_l}{D_h + D_l}\right)^2\left[\frac{\pi}{2} \pm \arcsin\left(\frac{x-1}{x+1}\right)\right]^2 \\ & + O\left(\frac{D_h - D_l}{D_h + D_l}\right)^4.\end{aligned} \tag{13.71}$$

The band-edge frequencies can then be found from (13.66) and (13.59):

$$\omega^{\pm} = \frac{2c}{D_h + D_l}\left(\frac{\pi}{2} \pm \phi^{\pm}\right). \tag{13.72}$$

As a numerical example, consider the band edges for a tellurium|polystyrene stack which is quarter-wave at normal incidence (Fink et al. 1998).

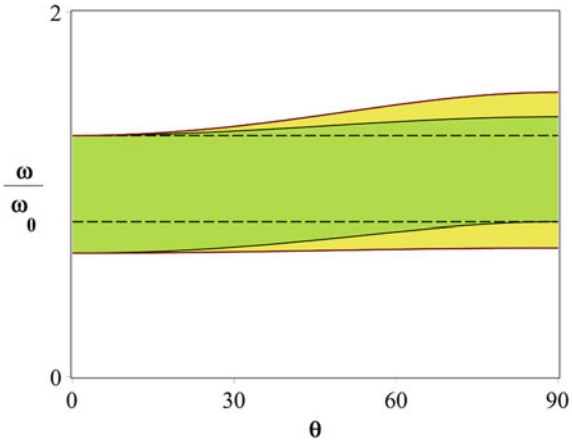


Fig. 13.6 Band edges for a high-low multilayer with $n_1 = 1$, $n_h = 4.6$ and $n_l = 1.6$ (Fink et al. 1998), versus θ. At normal incidence the stop band extends from $\omega_0 - \Delta\omega$ to $\omega_0 + \Delta\omega$, with $\Delta\omega$ given by (13.43). The *outer band* is for s polarization, the *darker inner band* is for the p polarization. Omnidirectional reflection occurs in the frequency range between the *dashed lines*. The calculations are for a "quarter-wave" stack, with $n_h d_h = n_l d_l = \lambda/4$

Equation (13.43) gives the band edges at normal incidence, exactly. Equations (13.71) and (13.72) give the band edges at all angles of incidence to such accuracy that exact and approximate results cannot be differentiated in Fig. 13.6. The errors in $\omega^{\pm}$ at glancing incidence range from 12 to 268 parts per million.

13.3.2 Refractive Indices for Which Omnidirectional Reflection Exists

Omnidirectional reflection requires that the s and p band gaps of the multilayer persist from normal incidence to glancing incidence. This will happen if the criterion (13.64) is satisfied. The region in the (n_l, n_h) plane where (13.64) is satisfied is bounded by the curve where

$$\omega_p^- = \omega_0^+ . \tag{13.73}$$

In terms of the approximate band edge frequencies given in (13.72), this reads

$$\frac{\frac{\pi}{2} - \phi_p^-}{d_h n_h r_h + d_l n_l r_l} = \frac{\frac{\pi}{2} + \phi_0^+}{d_h n_h + d_l n_l} , \tag{13.74}$$

where

$$r_h = \sqrt{1 - n_1^2/n_h^2}, \qquad r_l = \sqrt{1 - n_1^2/n_l^2}, \tag{13.75}$$

and ϕ_0^+ is evaluated at normal incidence with $x_0 = n_h/n_l$, while ϕ_p^- is evaluated at glancing incidence with $x_p = (n_h/n_l)(r_l/r_h)$, $D_h = d_h n_h r_h$ and $D_l = d_l n_l r_l$. For quarter-wave stacks, (13.74) reduces to

$$\frac{\frac{\pi}{2} - \phi_p^-}{r_h + r_l} = \frac{\frac{\pi}{2} + \phi_0^+}{2}. \tag{13.76}$$

Solution of (13.76) gives a curve in the (n_l, n_h) plane which is shown in Fig. 13.7. Above this curve is the omnidirectional reflection region. The minimum value of $n_h (= 2.265\,899\,n_1)$ occurs at $n_l = 1.517\,523\,n_1$. The exact equation (13.73), with ω_p^- found from (13.63) or (13.67) at glancing incidence, has $(n_l, n_h)_{\min}$ at $(1.517\,522\,n_1,\ 2.265\,899\,n_1)$. In contrast, Fig. 13.3 of Southwell (1999) has a minimum for its onset of omnidirectional reflection curve at $n_l \approx 1.45\,n_1$, $n_h \approx 2.24\,n_1$. This is due to the fact that his equation for the edges of the bands is less accurate than the one used here, even for quarter-wave stacks.

It is of interest to widen the search and admit dielectric stacks which are not quarter-wave at normal incidence. For general stacks we have an extra parameter, the ratio of the optical thicknesses of the layers at normal incidence: $\rho = n_h d_h / n_l d_l$. Lekner (2000) finds that the lowest n_h is attained at

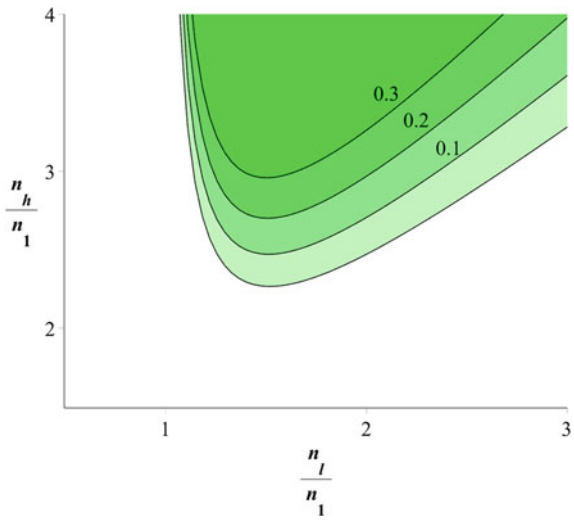


Fig. 13.7 Region of omnidirectional reflection of a high-low multilayer. The *lower curve* shows the limit of omnidirectional reflection for a quarter-wave stack according to (13.76) with ω_p^- approximated by (13.72, 13.73). The contours labelled 0.1, 0.2 and 0.3 show where the quarter-wave stack omnidirectional reflection region is $0.1\omega_0$, $0.2\omega_0$ and $0.3\omega_0$ wide

$$\rho = 1.362\,086, \qquad n_l = 1.492\,045\,n_1, \qquad n_h = 2.246\,763\,n_1. \tag{13.77}$$

These values are to be compared with the quarter wave stack values found above:

$$\rho = 1, \qquad n_l = 1.517\,522\,n_1, \qquad n_h = 2.265\,899\,n_1. \tag{13.78}$$

13.4 Form Birefringence

A pencil of light entering an anisotropic material, such as a crystal of calcite, is in general split into two beams: calcite is doubly refracting, or birefringent. The optical properties of anisotropic materials are characterized by a tensor dielectric function, as we saw in Chap. 8. For a given angle of incidence of a plane electromagnetic wave onto a given crystal face, two plane wave modes are possible within the crystal. For a uniaxial crystal, such as calcite, these are called the ordinary and extraordinary modes. The configuration of interest in relation to planar stratified media is one where the optic axis of a uniaxial material coincides with the surface normal. Then the ordinary and extraordinary modes have wavevectors $(K, 0, q_o)$ and $(K, 0, q_e)$ where (Sects. 7.1 and 8.2)

$$q_o^2 = \varepsilon_o \omega^2/c^2 - K^2, \qquad q_e^2 = \varepsilon_o \omega^2/c^2 - (\varepsilon_o/\varepsilon_e)K^2, \tag{13.79}$$

where $\varepsilon_o = n_o^2$ and $\varepsilon_e = n_e^2$ are the ordinary and extraordinary dielectric constants for the crystal. The electric field vectors of the ordinary and extraordinary modes are along the directions

$$E_o \sim (0,\ 1,\ 0), \qquad E_e \sim [q_e,\ 0,\ -(\varepsilon_o/\varepsilon_e)K]. \tag{13.80}$$

From Maxwell's equation for the curl of $\mathbf{E}$ we find $B_e \sim (0,\ 1,\ 0)$. Thus the normal mode o and e field directions in the crystal correspond to the s and p have characterizations used in isotropic media. (This holds only when the optic axis is normal to the reflecting surface of the crystal.)

When a narrow beam is incident onto the crystal, it is refracted into two beams, whose directions are those of $\boldsymbol{E} \times \boldsymbol{B}$, along the Poynting vector for each mode. The ordinary mode ray direction always coincides with that of the ordinary wavevector $(K,\ 0,\ q_o)$. The extraordinary ray direction does not coincide with $(K,\ 0,\ q_e)$ in general. When the optic axis is normal to the reflecting surface, the ray direction of the extraordinary wave is along $[(\varepsilon_o/\varepsilon_e)K,\ 0,\ q_e]$.

We now consider waves in a periodic stratification made up of isotropic component layers. *Form birefringence* is the name given to the way in which such a structure behaves like an anisotropic homogeneous medium, in the limit when the wavelength is large compared to the period (Born and Wolf (1965), Sect. 14.5.2 and Yariv and Yeh (1984), Sect. 6.8). The equivalent homogeneous medium is uniaxial, with optic axis normal to the stratifications. To see this, we write the Bloch factor β (which according to (13.38) has eigenvalues $\mathrm{e}^{\pm i\phi}$) as $\mathrm{e}^{\pm iqd}$, where d is the thickness of one unit cell of the stratification, and q is interpreted as the normal component of the effective wave vector, which is thus $(K, 0, q_s)$ and $(K, 0, q_p)$ for the s (TE) and p (TM) polarisations. Since q_s and q_p are different (being determined by the trace of M_s and of M_p), we have a one-to-one correspondence with the normal components q_o and q_e of the ordinary and extraordinary waves.

For the high-low stack, $\cos\phi_s$ and $\cos\phi_p$ are given by (13.41) and (13.42). In the long-wave limit we have $q_l d_l \ll 1$ and $q_h d_h \ll 1$; expansion of (13.41) and (13.42) in powers of $q_l d_l$ and $q_h d_h$, with

$$\phi_s = q_s(d_h + d_l) = q_s d, \qquad q_s^2 \equiv \varepsilon_s \omega^2/c^2 - K^2, \tag{13.81}$$

$$\phi_p = q_p(d_h + d_l) = q_p d, \qquad q_p^2 \equiv \frac{\varepsilon_s \omega^2}{c^2} - \left(\frac{\varepsilon_s}{\varepsilon_p}\right) K^2, \tag{13.82}$$

for the TE and TM waves (compare the expressions for q_o^2 and q_e^2 in (13.79)) gives

$$\varepsilon_s = f_h \varepsilon_h + f_l \varepsilon_l, \qquad \varepsilon_p = \frac{\varepsilon_h \varepsilon_l}{f_h \varepsilon_l + f_l \varepsilon_h}. \tag{13.83}$$

In (13.83) f_h, f_l are the fractions of the total volume occupied by the high and low index materials in the medium:

$$f_h = \frac{d_h}{d_h + d_l}, \qquad f_l = \frac{d_l}{d_h + d_l}. \tag{13.84}$$

The expressions (13.83) have been obtained by electrostatic considerations (Born and Wolf 1965), and by Bloch-wave dynamical argument (Yariv and Yeh 1984) as above. Note that the effective anisotropy $\varepsilon_e - \varepsilon_o$ cannot be positive:

$$\varepsilon_p - \varepsilon_s = -\frac{(\varepsilon_h - \varepsilon_l)^2 f_h f_l}{f_h \varepsilon_l + f_l \varepsilon_h}. \tag{13.85}$$

Thus the "ordinary" s or TE wave experiences a larger effective refractive index than does the "extraordinary" p or TM wave. Experimental demonstration of form birefringence may be seen in van der Ziel et al. (1976) and Kitagawa and Tateda (1985), for example.

The previously known results of the last paragraph apply only to the case where the unit cell is composed of two homogeneous layers. The long-wave limit can be generalized to an arbitrary dielectric function profile within the unit cell (Lekner 1994, Sect. 4). To second-order in the cell thickness divided by the wavelength, the single-period matrix for the s wave is (Chap. 12, (12.76))

$$M_s = \begin{bmatrix} 1 - \int_a^b \mathrm{d}z\, q^2(z)(b-z) & b-a \\ -\int_a^b \mathrm{d}z\, q^2(z) & 1 - \int_a^b \mathrm{d}z\, q^2(z)(z-a) \end{bmatrix} \tag{13.86}$$

(the unit cell extends from $z = a$ to $z = b = a + d$). Thus

$$\cos\phi_s = \frac{1}{2}\mathrm{trace}M_s = 1 - \frac{1}{2}(b-a)\int_a^b \mathrm{d}z\, q^2(z) + \cdots \tag{13.87}$$

We expand $\cos\phi_s$ as $1 - \frac{1}{2}q_s^2(b-a)^2 + \cdots$ and put

$$q_s^2 = \varepsilon_s\omega^2/c^2 - K^2, \quad q^2(z) = \varepsilon(z)\omega^2/c^2 - K^2. \tag{13.88}$$

Then (13.87) gives

$$\varepsilon_s = \langle\varepsilon\rangle \equiv \frac{1}{b-a}\int_a^b \mathrm{d}z\, \varepsilon(z). \tag{13.89}$$

For the p wave the unit cell matrix is given by (12.80) of Sect. 12.4 to second-order in the cell thickness:

$$M_p = \begin{bmatrix} 1 - \int_a^b \mathrm{d}z[q^2(z)/\varepsilon(z)]\int_z^b \mathrm{d}\zeta\, \varepsilon(\zeta) & \int_a^b \mathrm{d}z\, \varepsilon(z) \\ -\int_a^b \mathrm{d}z\, q^2(z)/\varepsilon(z) & 1 - \int_a^b \mathrm{d}z\, \varepsilon(z)\int_z^b \mathrm{d}\zeta\, q^2(\zeta)/\varepsilon(\zeta) \end{bmatrix}. \tag{13.90}$$

We expand $\cos\phi_p = \frac{1}{2}\mathrm{trace}M_p$ as $1 - \frac{1}{2}q_p^2(b-a)^2 + \cdots$, and set $q_p^2 = \varepsilon_s\omega^2c^2 - (\varepsilon_s/\varepsilon_p)K^2$, to find the same $\varepsilon_s = \langle\varepsilon\rangle$ as above, and

$$\frac{\varepsilon_s}{\varepsilon_p}(b-a)^2 = \int_a^b \mathrm{d}z/\varepsilon(z)\int_z^b \mathrm{d}\zeta\ \varepsilon(\zeta) + \int_a^b \mathrm{d}z\ \varepsilon(z)\int_z^b \mathrm{d}\zeta/\varepsilon(\zeta), \tag{13.91}$$

which reduces to

$$\frac{1}{\varepsilon_p} = \left\langle \frac{1}{\varepsilon} \right\rangle \equiv \frac{1}{b-a} \int_a^b \frac{\mathrm{d}z}{\varepsilon(z)}. \tag{13.92}$$

The expressions (13.89) and (13.92) for the ordinary and extraordinary dielectric constants of the equivalent homogeneous but anisotropic medium reduce to (13.83) in the special case of a unit cell made up of two homogeneous layers.

Thus, in the long-wave limit, any periodically stratified isotropic medium can be replaced by a homogeneous uniaxial medium, with optic axis normal to the stratification, and $\varepsilon_o = \langle \varepsilon \rangle$, $\varepsilon_e^{-1} = \langle \varepsilon^{-1} \rangle$. Since the harmonic mean of a set of positive quantities is never more than their arithmetic mean, it follows that ε_e will not exceed ε_o, provided $\varepsilon(z)$ is positive everywhere.

The reader may have noticed a curious feature of the proofs given above: the periodicity of the stratification is used to define a Bloch wavevector via $\phi = qd$, where the cosine of ϕ is half of the trace of the matrix for a unit cell, but the thickness d of the unit cell drops out of the expressions for the equivalent ordinary and extraordinary indices of the equivalent homogeneous medium, in the long-wave limit. Could it be that the $\varepsilon_o = \langle \varepsilon \rangle$ and $\varepsilon_e = \langle \varepsilon^{-1} \rangle^{-1}$ results apply also to *disordered* finely layered media? The following argument suggests that they do: consider a stratification which appears disordered on a fine scale (for example the nanometer scale), but is actually periodic on a larger scale (for example the period is in the tens of nanometer range). The above proof then applies, provided the wavelength of the radiation is larger still (for example hundreds of nanometers). It seem plausible that non-periodic finely layered media can be represented in the long-wave limit by an effective uniaxial medium with ε_o and ε_e given by (13.89) and (13.92); the only difference is that disordered media will scatter more: they will show reflection from variations in the dielectric function $\varepsilon(z)$, even in the long-wave limit.

13.5 Absorbing Periodically Stratified Media

This section examines the reflection by *absorbing* periodically stratified media. Stop bands no longer exist in the strict sense, but their remnants influence reflection, as we shall see. The rapid variation with angle of incidence or wavelength is smoothed by absorption. On physical grounds we expect that the strong dependence on N (the number of unit cells, or repetitions of one period of the stratification) will also be smoothed by absorption. This is shown to be true, and indeed we prove that provided N exceeds a number which is inversely proportional to the absorption, the s and p reflectivities attain a universal form, independent of N and of the properties

of the substrate. In this sense, at least, reflection from an absorbing periodically stratified medium is simpler than from its non-absorbing idealization.

The idea that absorption will cause a periodically stratified medium to have the above properties was seen in early papers (Koppelmann 1960; Flannery et al. 1979) on multilayer dielectric reflectors, which found asymptotic (large N) properties for the reflectance. Lakhtakia (2011) showed numerically that with absorption the reflectance of rugate filters approaches a limiting form as N increases, dominated by the stop band structure. General formulae were derived for the reflection amplitudes of both polarizations by Lekner (2014), and will be given here.

13.5.1 *Reflection of* s*-Polarized Plane Waves*

The results of Sects. 13.1 and 13.2 hold for absorbing media also, since the layer matrices giving the transmission and reflection properties of the periodic structure remain unimodular in the presence of absorption. Let the reflecting stratification contain N periods (of, for example, alternating high-index and low-index identical bilayers), and let the outer surface of the stratification be the $z = 0$ plane. Then the reflection amplitude of the s-wave was given by in Sect. 13.2, (13.45):

$$r_s = \frac{q_1 q_2 m_{12} + m_{21} + iq_1(m_{22} - \sigma_N) - iq_2(m_{11} - \sigma_N)}{q_1 q_2 m_{12} - m_{21} + iq_1(m_{22} - \sigma_N) + iq_2(m_{11} - \sigma_N)}. \tag{13.93}$$

Here $q_1 = n_1(\omega/c)\cos\theta_1$ and $q_2 = n_2(\omega/c)\cos\theta_2$ are the normal components of the incident and transmitted wavevectors (θ_1 and θ_2 are the angles of incidence and transmission) and

$$\sigma_N = \frac{\sin(N-1)\phi}{\sin N\phi} = \cos\phi - \sin\phi\cot N\phi \tag{13.94}$$

depends on the angle ϕ defined by

$$\cos\phi = \frac{1}{2}\text{trace } M = \frac{1}{2}(m_{11} + m_{22}), \tag{13.95}$$

where M is the 2×2 matrix of one period,

$$M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}. \tag{13.96}$$

For example, in the special case of a high-low multilayer, consisting of N repetitions of constant refractive index n_h and thickness h followed by constant index n_l and thickness l, the s-wave matrix was given in (13.40):

$$M = \begin{pmatrix} c_l & q_l^{-1}s_l \\ -q_l s_l & c_l \end{pmatrix}\begin{pmatrix} c_h & q_h^{-1}s_h \\ -q_h s_h & c_h \end{pmatrix} = \begin{pmatrix} c_l c_h - q_l^{-1}q_h s_l s_h & c_l q_h^{-1}s_h + q_l^{-1}s_l c_h \\ -q_l s_l c_h - c_l q_h s_h & c_l c_h - q_l s_l q_h^{-1}s_h \end{pmatrix} \tag{13.97}$$

The wave-vector normal components q_l and q_h are given by

$$q_l^2 = \varepsilon_l\,\omega^2/c^2 - K^2, \quad q_h^2 = \varepsilon_h\,\omega^2/c^2 - K^2, \quad K = (\omega/c)n_1 \sin\theta_1, \tag{13.98}$$

and we use the shorthand notation

$$c_l = \cos\, q_l l, \quad s_l = \sin\, q_l l, \quad c_h = \cos\, q_h h, \quad s_h = \sin q_h h. \tag{13.99}$$

In this case, the angle ϕ defined by (13.95) is given by

$$\cos\phi_s = c_l c_h - \frac{1}{2}s_l s_h\left(q_l^{-1}q_h + q_l q_h^{-1}\right). \tag{13.100}$$

For non-absorbing media, the stop bands are the regions where $\cos^2\phi > 1$.

For absorbing media, the dielectric functions ε_l and ε_h are complex, related to the complex indices of refraction by $\varepsilon = \varepsilon_r + i\varepsilon_i = (n_r + in_i)^2 = n_r^2 - n_i^2 + 2in_r n_i$. Thus all of the matrix elements become complex, and ϕ (still defined by (13.3)) is always complex. As noted above, we expect the reflectance to become independent of the number of periods N, if N is large enough (how large is specified below). Let $\phi = \phi_r + i\phi_i$, and assume for the moment that $\phi_i > 0$. In the definition (13.94) of σ_N we have

$$\begin{aligned} \cot N\phi &= \frac{\cos N\phi_r \cosh N\phi_i \;-\; i\sin N\phi_r \sinh N\phi_i}{\sin N\phi_r \cosh N\phi_i \;+\; i\cos N\phi_r \sinh N\phi_i} \\ &= \frac{\cos N\phi_r(1+\xi) \;-\; i\sin N\phi_r(1-\xi)}{\sin N\phi_r(1+\xi) \;+\; i\cos N\phi_r(1-\xi)} \\ &= -i + O(\xi), \quad \xi = e^{-2N\phi_i} \end{aligned} \tag{13.101}$$

Thus, for $N\phi_i$ large and positive, (13.94) and (13.101) give us

$$\sigma_N = \cos\phi + i\sin\phi + O\left(e^{-2N\phi_i}\right) = e^{i\phi} + O\left(e^{-2N\phi_i}\right). \tag{13.102}$$

If ϕ_i is negative, on the other hand, we obtain

$$\sigma_N = \cos\phi - i\sin\phi + O\left(e^{2N\phi_i}\right) = e^{-i\phi} + O\left(e^{2N\phi_i}\right). \tag{13.103}$$

We have thus verified our expectation that the exact number N of the stratifications becomes unimportant, provided it is large enough to make $\exp(-2N|\mathrm{Im}(\phi)|)$ negligible.

Similar physical reasoning also leads us to expect that the substrate properties should become unimportant for the reflectivity, because when there is absorption and N is 'large', the incident wave does not penetrate to the substrate. Hence we expect $q_2 = n_2(\omega/c)\cos\theta_2$ to drop out of the reflection amplitude (13.1) at the same time as σ_N tends to its limit $\exp(\pm i\phi)$, independent of N. What is the condition under which r_s becomes independent of q_2? Let us write the s-wave reflection amplitude as

$$r_s = \frac{\alpha q_2 + \beta}{A q_2 + B} = \frac{\alpha(q_2 + \beta/\alpha)}{A(q_2 + B/A)}. \tag{13.104}$$

It will be independent of q_2 when $\beta/\alpha = B/A$, namely when

$$\frac{m_{21} + i q_1(m_{22} - \sigma)}{q_1 m_{12} - i(m_{11} - \sigma)} = \frac{-m_{21} + i q_1(m_{22} - \sigma)}{q_1 m_{12} + i(m_{11} - \sigma)}. \tag{13.105}$$

Because of the unimodularity of the layer matrix $(m_{11}m_{22} - m_{21}m_{12} = 1)$ the condition (13.105) simplifies to

$$\sigma^2 - (m_{11} + m_{22})\sigma + 1 = 0. \tag{13.106}$$

Since $m_{11} + m_{22} = 2\cos\phi$ from the definition (13.95) of ϕ, the solutions of (13.106) are

$$\sigma_\pm = e^{\pm i\phi}, \tag{13.107}$$

in accord with (13.103, 13.104). The physical root of (13.106) is the one with modulus less than unity. We have thus obtained the result, valid when $N|\phi_i|$ is large, that the s-reflection amplitude takes a value independent of N and of q_2, namely

$$r_s = \frac{\alpha}{A} = \frac{q_1 m_{12} - i(m_{11} - \sigma)}{q_1 m_{12} + i(m_{11} - \sigma)} \tag{13.108}$$

(where σ satisfies (13.106)) or equivalently

$$r_s = \frac{\beta}{B} = \frac{m_{21} + i q_1(m_{22} - \sigma)}{-m_{21} + i q_1(m_{22} - \sigma)}. \tag{13.109}$$

13.5.2 *Reflection of* p-*Polarized Plane Waves*

We can abbreviate the discussion here, because the formalism for p(TM)-polarized incidence is almost the same as for s(TE)-polarized incidence. We shall show only

the differences. The p-wave reflection amplitude for an N-layer stratification is given by (13.49):

$$-r_p = \frac{Q_1 Q_2 m_{12} + m_{21} + iQ_1(m_{22} - \sigma_N) - iQ_2(m_{11} - \sigma_N)}{Q_1 Q_2 m_{12} - m_{21} + iQ_1(m_{22} - \sigma_N) + iQ_2(m_{11} - \sigma_N)}, \tag{13.110}$$

where

$$Q_1 = q_1/\varepsilon_1 = n_1^{-1}\left(\frac{\omega}{c}\right)\cos\theta_1, \quad Q_2 = q_2/\varepsilon_2 = n_2^{-1}\left(\frac{\omega}{c}\right)\cos\theta_2, \tag{13.111}$$

and the matrix elements m_{ij} are to be calculated as specified above (13.42). In the special case of the same high-low multilayer given as an example for the s-wave, the matrix for one period becomes

$$M = \begin{pmatrix} c_l & Q_l^{-1}s_l \\ -Q_l s_l & c_l \end{pmatrix}\begin{pmatrix} c_h & Q_h^{-1}s_h \\ -Q_h s_h & c_h \end{pmatrix}, \tag{13.112}$$

where $c_l = \cos q_l l$, $s_l = \sin q_l l$, $c_h = \cos q_h h$ and $s_h = \sin q_h h$ as before. The effective wavenumbers Q_l and Q_h are defined by

$$Q_l = \frac{q_l}{\varepsilon_l}, \quad Q_h = \frac{q_h}{\varepsilon_h}. \tag{13.113}$$

The angle ϕ defined in (13.95) is thus given by

$$\cos\phi_p = c_l c_h - \frac{1}{2}s_l s_h\left(Q_l^{-1}Q_h + Q_l Q_h^{-1}\right). \tag{13.114}$$

The function $\sigma_N(\phi)$, which transforms the one-period into the N-period reflection amplitude, is defined by (13.94) as before, but for the p-wave reflection we use ϕ_p, which is different from ϕ_s except at normal incidence. Hence we again have $\sigma_N = \mathrm{e}^{\pm i\phi} + O(\mathrm{e}^{-2N|\phi_i|})$, and the reflection amplitude becomes independent of N for $N|\phi_i|$ large. Thus, by the reasoning given in the previous section, the substrate effective wave-number component Q_2 also drops out of the reflection amplitude, which can be written in the equivalent forms

$$-r_p = \frac{Q_1 m_{12} - i(m_{11} - \sigma)}{Q_1 m_{12} + i(m_{11} - \sigma)} = \frac{m_{21} + iQ_1(m_{22} - \sigma)}{-m_{21} + iQ_1(m_{22} - \sigma)}. \tag{13.115}$$

The s and p reflectivities are the absolute squares of the reflection amplitudes:

$$R_s = |r_s|^2, \quad R_p = |r_p|^2 \tag{13.116}$$

13.5.3 Application to an Absorbing Quarter-Wave Stack

Koppelmann (1960) appears to have been the first to show that weakly absorbing multilayer dielectric reflectors have limiting reflectance properties independent of N for large N. For a high-low stack, with refractive indices $n_h + ik_h$ and $n_\ell + ik_\ell$, at the design angular frequency ω_0 and quarter-wave layer thicknesses, such that

$$\frac{\omega_0}{c} = \frac{\pi/2}{n_h h} = \frac{\pi/2}{n_l l}. \tag{13.117}$$

Koppelmann showed that the reflectivity at normal incidence and at $\omega = \omega_0$ is

$$R(\omega_0) = 1 - \frac{2\pi n_1 (k_h + k_l)}{n_h^2 - n_l^2} + O\left(k_h^2, k_h k_l, k_l^2\right). \tag{13.118}$$

In the absence of absorption the (fundamental) stop band extends over the range $\omega_0 - \Delta\omega$ to $\omega_0 + \Delta\omega$, where

$$\Delta\omega/\omega_0 = (2/\pi)\arcsin[(n_h - n_l)/(n_h + n_l)]. \tag{13.119}$$

Within this range the reflectivity is unity for a non-absorbing infinite stack, and approaches unity exponentially with N (Sect. 13.2). The value of $\cos^2\phi$ exceeds unity: or example at $\omega = \omega_0$, $\cos\phi = -(n_h^2 + n_l^2)/2n_h n_l < -1$. Hence ϕ is complex even in the absence of absorption. Figure 13.8 shows the reflectance of an

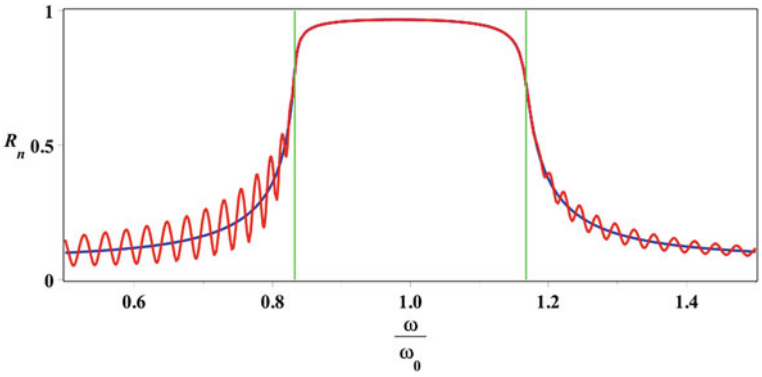


Fig. 13.8 Normal incidence reflectivity of a quarter-wave high-low stack, with $n_1 = 1$, $n_h = 2.35$ (ZnS), $n_l = 1.38$ (MgF_2) and $n_2 = 1.5$ (glass). The absorptive (imaginary) parts of the refractive indices k_h, k_ℓ have both been set to 0.01, about two orders of magnitude larger than actual values, so as to demonstrate the asymmetry in frequency dependence. The frequency range extends from $0.5\omega_0$ to $1.5\omega_0$; the stop band lies between $\omega_0 - \Delta\omega$ and $\omega_0 + \Delta\omega$, with $\Delta\omega$ given by (13.119). The smooth curve is the large N limit; the oscillatory curve is drawn for $N = 30$. The *vertical lines* indicate the limits of the stop band for an infinite non-absorbing stack

absorbing stack, calculated for 30 high-low layers, and also from any of the formulae (13.109, 13.110) or (13.115) (all equivalent at normal incidence).

The matrix elements used are as defined in (13.97). Over the frequency range plotted, the appropriate σ value to be used is σ_+. For example, at $\omega = \omega_0$ we find, to first order in k_h, k_ℓ,

$$\sigma_+(\omega_0) = -n_l/n_h + i(k_h n_l - k_l n_h)/n_h^2, \quad \sigma_-(\omega_0) = -n_h/n_l - i(k_h n_l - k_l n_h)/n_l^2. \tag{13.120}$$

For finite N the reflectivity is oscillatory outside of the stop band region, $\omega_0 - \Delta\omega$ to $\omega_0 + \Delta\omega$. The oscillations increase in number and decrease in amplitude with N. Inside the stop band the difference between the finite N reflectivity and its asymptotic value is exponential in N (Sect. 13.2), and is well below visibility in the example illustrated in Fig. 13.8.

Flannery et al. (1979) have noted that, with absorption, the maximum reflectivity is obtained at a frequency ω_m not equal to ω_0. By expanding the reflectivity obtained from (13.116) to first order in the imaginary parts of the refractive indices, and to second order in $\omega/\omega_0 - 1$, and differentiating the result with respect to ω, we obtain

$$\frac{\omega_m}{\omega_0} = 1 + \frac{4(k_h n_l - k_l n_h)(n_h - n_l)}{\pi^2[k_h(n_h^2 + 2n_l^2 - n_1^2) + k_l(2n_h^2 + n_l^2 - n_1^2)]} + O(k_h, k_l). \tag{13.121}$$

To this order, the frequency shift is homogeneous of degree zero in k_h, k_ℓ. Thus even infinitesimal absorption will shift the maximum away from ω_0 by a finite amount, which seems paradoxical until one remembers that the whole range $\omega_0 - \Delta\omega$ to $\omega_0 + \Delta\omega$ is a maximum for zero absorption. More details about the reflectivity at $\omega = \omega_m$ and at the location of the stop band edges can be found in Lekner (2014).

The theory of this section is not restricted to piecewise constant dielectric functions: an arbitrary periodic dielectric function profile will have (analytically or numerically) a matrix for one period, with elements depending on the profile, frequency, angle of incidence and polarization, and the formulae derived here will give its asymptotic reflection properties.

References

Abelès F (1950) Recherches sur la propagation des ondes électromagnétiques sinusoidales dan les milieux stratifiés. Application aux couches minces. Annales de Physique 5(596–640):706–782

Born M, Wolf E (1965) Principles of optics, Chap. 14, 3rd edn. Pergamon, Oxford

Brillouin L (1946/1953) Wave propagation in periodic structures. McGraw-Hill, New York; Dover, New York

Chigrin DN, Lavrinenko AV, Yarotsky DA, Gaponenko SV (1999a) Observation of total omnidirectional reflection from a one-dimensional dielectric lattice. Appl Phys A 68:25–28
Chigrin DN, Lavrinenko AV, Yarotsky DA, Gaponenko SV (1999b) All-dielectric one-dimensional periodic structures for total omnidirectional reflection and partial spontaneous emission control. J Lightwave Tech 17:2018–2024
Fink Y, Winn JN, Fan S, Chen C, Michel J, Joannopoulos JD, Thomas EL (1998) A dielectric omnidirectional reflector. Science 282:1679–1682
Flannery M, Loh E, Sparks M (1979) Nearly perfect multilayer dielectric reflectors: theory. Appl Opt 18:1428–1435
Griffiths DJ, Steinke CA (2001) Waves in locally periodic media. Am J Phys 69:137–154
Hardy GH (1952) A course of pure mathematics, 3rd edn. Cambridge University Press, Cambridge
Joannopoulos JD, Meade RD, Winn JN (1995) Photonic crystals. Princeton University Press, Princeton
Kinoshita S (2008) Structural colors in the realm of nature. World Scientific, Singapore
Kinoshita S (2013) Bionanophotonics. Taylor and Francis, London
Kitagawa M, Tateda M (1985) Form birefringence of SiO_2/Ta_2O_5 periodic multilayers. Appl Opt 24:3359–3363
Lakhtakia A (2011) Reflection from a semi-infinite rugate filter. J Mod Opt 58:562–565
Lekner J (1994) Light in periodically stratified media. J Opt Soc Am A: 11:2892–2899
Lekner J (2000) Omnidirectional reflection by multilayer dielectric mirrors. J Opt A: Pure Appl Opt 2:349–342
Lekner J (2014) Reflection by absorbing periodically stratified media. J Opt 16:035104 (4 pp)
Mott NF, Jones H (1936/1958) The theory of the properties of metals and alloys. Clarendon Press, Oxford; Dover, New York
Nusinsky I, Hardy AA (2007) Omnidirectional reflection in several frequency ranges of one-dimensional photonic crystals. Appl Opt 46:3510–3517
Rayleigh JWS (1887) On the maintenance of vibrations by forces of double frequency, and on the propagation of waves through a medium endowed with a periodic structure. Phil Mag 24:145–159
Rayleigh JWS (1917) On the reflection of light from a regularly stratified medium. Proc R Soc A93:565–577
Southwell WH (1999) Omnidirectional mirror design with quarter-wave dielectric stacks. Appl Opt 38:5464–5467
Van der Ziel JP, Ilegems M, Mikulyak RM (1976) Optical birefringence of thin GaAs-AlAs multilayer films. Appl Phys Lett 28:735–737
Von Koppelmann G (1960) Theory of thin film layers of weakly absorbing materials and their application as interferometer mirrors. Ann Phys 7:388–396 [in German]
Winn JN, Fink Y, Fan S, Joannopoulos JD (1998) Omnidirectional reflection from a one-dimensional photonic crystal. Opt Lett 23:1573–1575
Yariv A, Yeh P (1977) Electromagnetic propagation in periodic stratified media. II. Birefringence, phase matching, and x-ray lasers. J Opt Soc Am 67:438–448
Yariv A, Yeh P (1984) Optical waves in crystals. Wiley, New York
Yeh P, Yariv A, Hong CS (1977) Electromagnetic propagation in periodic stratified media. I. General theory. J Opt Soc Am 67:423–438

Chapter 14
Rough or Structured Surfaces

We have seen in Chap. 1 that a planar surface, or arbitrary stratification, will give specular reflection of an incident plane wave. No real surface is perfectly planar, and thus in practice there is a diffuse or scattered component, as well as a specular component of the radiation. The rougher the surface, the greater the diffuseness of the re-radiation from it. A rough surface which is planar on average (for example a liquid-vapour interface stabilized by a gravitational field) is characterized by at least two parameters: a length h giving the typical variation in the height of the surface, and another length l giving the scale of correlations between displacements at different points of the surface. The incident plane wave is characterized by its wavelength λ and angle of incidence θ (measured relative to the mean surface, assumed planar); the scattered radiation is characterized by two angles θ' and ϕ'. (We will not consider inelastic scattering by a dynamic surface here, so the wavelength of the scattered radiation is taken to be that of the incident radiation.) The characterization of scattered light is thus in terms of at least three lengths λ, h, l, and three angles θ, θ', ϕ'. In the geometrical optics limit ($\lambda \ll h, l$) the surface may be taken to be locally plane, and thus the scattered light is obtained from the statistical geometry of the surface by assuming specular reflection from each tilted element. This is a good description of the reflection of light from large bodies of water, provided that foam and spray are absent. Cox and Munk (1954), for example, measured the roughness of the sea surface from photographs of the sun's glitter. Longuet-Higgins (1960) has studied in detail the geometry of reflection and refraction at a random moving surface, of light originating at a point source. The reflection of extended objects by gently rippled water is discussed in an illustrated note by Lynch (1985).

This chapter will concentrate on the wave theory of reflection by rough surfaces, specifically including diffraction effects which arise when the wavelength is comparable to the lengths characterizing the surface roughness.

J. Lekner, *Theory of Reflection*, Springer Series on Atomic, Optical, and Plasma Physics 87, DOI 10.1007/978-3-319-23627-8_14

14.1 Reflection from Rough Surfaces: The Rayleigh Criterion

From the wave theory point of view, a variation in height in the reflecting surface by an amount h will be significant if the resulting path difference is comparable to the wavelength λ. Rayleigh (1879) noted that in the specular case the path difference is $2h\cos\theta$ for the simple geometry shown in Fig. 14.1.

Since $(2\pi/\lambda)\cos\theta = q$, the normal component of the wavevector of the incident light, we may write the Rayleigh criterion for the specular reflection off a surface as

$$2\pi h\cos\theta \ll \lambda \text{ or } qh \ll 1. \tag{14.1}$$

The factor π inserted here is arbitrary: a more precise specification comes when one takes a particular model of surface roughness. For the Gaussian model considered in Sect. 14.4 the operative roughness parameter is

$$\mathcal{R} = (q+q')^2\langle\zeta^2\rangle, \tag{14.2}$$

where q and q' are the magnitudes of the normal components of the incident and scattered wavevectors, and $\langle\zeta^2\rangle$ is the mean square surface height variation when $\langle\zeta\rangle = 0$. The intensity of specularly reflected radiation is proportional to $\exp(-\mathcal{R})$ in this model; non-specularly scattered light depends both on $\mathcal{R}$ and on the product of the lateral correlation length l with the change in the lateral component of the wavenumber.

The Rayleigh criterion (14.1) in terms of path or phase differences indicates that with a given roughness long waves may be reflected specularly and short waves diffusely, or that for given roughness and wavelength there may be diffuse scattering near normal incidence and specular reflection near grazing incidence. The

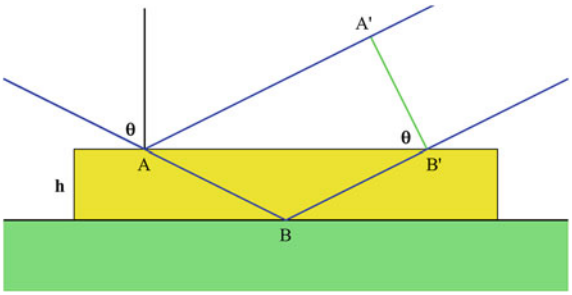


Fig. 14.1 Path difference between rays reflecting specularly in the presence and absence of a step of height h. The path difference between the rays equals $(AB + BB') - AA' = 2h\sec\theta - 2h\tan\theta\sin\theta = 2h\cos\theta$

change from diffuse to specular reflection with angle of incidence is apparent with a plate of ground or smoked glass. Further discussion may be found in Rayleigh's lecture on "Polish" (1901) and in Wood (1934), pp. 39–41.

14.2 Corrugated Surfaces, Diffraction Gratings

The simplest roughness to consider is that of a periodic corrugation of a sharply defined surface at $z = \zeta(x)$:

$$\zeta(x) = \sum_1^\infty (c_n \cos npx + s_n \sin npx) = \frac{1}{2}\sum_1^\infty [(c_n - is_n)e^{inpx} + (c_n + is_n)e^{-inpx}]. \tag{14.3}$$

The period of the corrugations is $d = 2\pi/p$. The method to be presented here was developed by Rayleigh (1896, 1907a) and rests on an assumption ("the Rayleigh hypothesis") which will be stated below and discussed again at the end of this section. Consider a plane wave incident (perpendicularly to the corrugations) at an angle of incidence θ relative to the normal to the averaged surface. If $k = 2\pi/\lambda$ is the total wavenumber in the medium of incidence, the incoming wave is

$$\psi_i = e^{ik(x\sin\theta + z\cos\theta)} = e^{i(Kx+qz)}. \tag{14.4}$$

The specularly reflected wave is

$$\psi_0 = A_0 e^{ik(x\sin\theta - z\cos\theta)} = A_0 e^{i(Kx-qz)}. \tag{14.5}$$

For reflected spectra of the nth order the wave is represented by terms like (14.5) with θ_n, K_n and q_n instead of θ, K and q, where

$$k\sin\theta_n \equiv K_n = K + np, \;\; k\cos\theta_n \equiv q_n. \tag{14.6}$$

The first part of (14.6) is the grating equation giving the direction of the nth order, which may be written as the condition that the path difference (to a distant point) between waves originating a distance d apart on the grating is an integral number of wavelengths:

$$d(\sin\theta_n - \sin\theta) = n\lambda. \tag{14.7}$$

The zeroth order $(n = 0)$ is the specularly reflected wave ψ_0. Since the surface shape $\zeta(x)$ is expressed as a sum over positive integers n, while the grating equations (14.6) or (14.7) may have negative n, it is convenient to define the primed quantities θ'_n, K'_n, q'_n obtained by changing the sign of n in (14.6), and sum over positive n:

$$k\sin\theta'_n \equiv K'_n = K - np, \quad k\cos\theta'_n \equiv q'_n. \tag{14.6'}$$

The spectra of the nth order are represented by

$$\psi_n = A_n e^{ik(x\sin\theta_n - z\cos\theta_n)} + A'_n e^{ik(x\sin\theta'_n - z\cos\theta'_n)} = A_n e^{i(K_n x - q_n z)} + A'_n e^{i(K'_n x - q'_n z)}. \tag{14.8}$$

Since

$$K_n^2 + q_n^2 = k^2 = \left(K'_n\right)^2 + \left(q'_n\right)^2, \tag{14.9}$$

the wavefunction (time dependence e^{-ikct} is understood)

$$\Psi = \psi_i + \psi_0 + \sum_1^\infty \psi_n \tag{14.10}$$

satisfies the wave equation $(\nabla^2 + k^2)\Psi = 0$ outside the reflecting surface. The *Rayleigh hypothesis* is that (14.10), which expresses the total wave as a sum of the incident wave plus all reflected spectral orders, with arbitrary amplitudes A_0, A_n, A'_n, is *complete*. That is, it is assumed that (14.10) has sufficient generality to satisfy the boundary conditions on an arbitrary periodically corrugated surface. More will be said about this hypothesis later; here we note that *evanescent waves,* that is those with

$$q_n^2 = k^2 - (K + np)^2 < 0 \text{ or } \left(q'_n\right)^2 = k^2 - (K - np)^2 < 0 \tag{14.11}$$

are to be included in (14.10). These correspond to orders that have "passed off over the grating horizon", and decay exponentially with $|z|$. From (14.6), (14.6′) and (14.9), we see that at normal incidence ($K = 0$) the maximum visible order is the integer part of d/λ, so that if $d < \lambda$ only the zeroth or specular order will be seen. At oblique incidence, the maximum visible order is the integer part of $2d/\lambda$, since the maximum value of K is k (attained at grazing incidence), and

$$\left(q'_n\right)^2 = k^2 - (K - np)^2 \tag{14.12}$$

will stay positive when $K = k$ for $np < 2k$ or $n < 2d/\lambda$. Thus if $d < \lambda/2$ there will be only specular reflection, at any angle of incidence, and the corrugated surface no longer acts as a diffraction grating. (There are near-field effects, but no spectra are visible in the far-field region.)

The amplitudes of the spectral orders, A_0, A_n and A'_n, are found from the boundary conditions to be imposed on (14.10) at the surface. The simplest case to

consider is that of a perfect reflector with the incident wave polarized so that the electric field lies along the corrugations. The boundary condition for this case is $\Psi = 0$ on $z = \zeta$. All of the incident energy is thrown back, and is distributed between the specular beam and the reflected spectra. (The spectra represented by evanescent waves do not radiate energy.) The total wavefunction

$$\Psi(x,z) = e^{iKx}\left\{e^{iqz} + A_0 e^{-iqz} + \sum_1^\infty (A_n e^{inpx - iq_n z} + A'_n e^{-inpx - iq'_n z})\right\} \tag{14.13}$$

is to be zero on $z = \zeta(x)$.

Two methods, both due to Rayleigh, exist for extracting the spectral amplitudes. We shall give an outline of both. In the first method, Rayleigh expands (14.13) in powers of ζ and equates coefficients of $\exp(\pm inpx)$, $\zeta(x)$ being given by (14.3). If we keep the zeroth and first powers of ζ in (14.13), $\Psi(x, \zeta(x)) = 0$ reads

$$0 = 1 + A_0 + \sum_1^\infty \left(A_n e^{inpx} + A'_n e^{-inpx}\right) + iq\zeta(1 - A_0) + O(\zeta^2). \tag{14.14}$$

To the zeroth power of ζ

$$1 + A_0 = 0, \tag{14.15}$$

and all the A_n and A'_n are zero. To the first power of ζ the value $A_0 = -1$ still holds, and

$$A_n = -iq(c_n - is_n), \quad A'_n = -iq(c_n + is_n) \tag{14.16}$$

To this approximation, the amplitudes of the $\pm n$ orders are given by the nth Fourier coefficients of the corrugation. If the corrugation is purely sinusoidal (only c_1 and s_1 non-zero) only the ± 1 diffraction orders have amplitudes which are non-zero when the calculation is taken to the first power of ζ. Rayleigh (1907a) gives the coefficients to the second power of ζ.

We now turn to consider the other (p or TM) polarization, with the electric field in the incident wave perpendicular to the corrugations. In this case Ψ corresponds to B_y, and the electric field is proportional to $curl\,\mathbf{B}$, that is to $(-\partial_z\Psi, 0, \partial_x\Psi)$. For a perfect reflector, the component of the electric field parallel to the (local) surface is to be zero. The boundary condition is thus

$$(\partial_z\Psi - \partial_x\Psi\, \mathrm{d}\zeta/\mathrm{d}x)_{z=\zeta} = 0. \tag{14.17}$$

Ψ is the total field, again given by (14.13). Expansion in powers of ζ now gives $A_0 = 1$ from the coefficients of the zeroth power, and

$$A_n = i(c_n - is_n)(q^2 - nKp)/q_n,$$
$$A'_n = i(c_n + is_n)(q^2 + nKp)/q'_n, \tag{14.18}$$

from the coefficients of the first power of ζ, A_0 being unchanged to this order. We note that (in this approximation) the coefficients A_n or A'_n diverge when q_n or q'_n go through zero, that is at the passing off of the nth order. Rayleigh shows that the passing off of an order of the spectrum can have an effect on other orders. For example: in the special case of normal incidence, and a corrugation for which only c_1 and c_2 are non-zero in the Fourier expansion (14.3), calculation of A_1 to the second power in ζ gives

$$A_1 = \frac{ik^2c_1}{q_1} - \frac{k^2c_1c_2}{2q_1^2}\left(q_1^2 + 2p^2\right) - \frac{k^2c_1c_2}{2q_1q_2}\left(q_2^2 + 2p^2\right). \tag{14.19}$$

To this approximation, the coefficient of the *first* order can diverge when the *second* order is passing off. According to Rayleigh (1907a), "we may at least infer the probability of abnormalities in the brightness of any spectrum at the moment when one of a higher order is disappearing, abnormalities limited, however, to the case where the electric displacement is perpendicular to the ruling". In a subsequent paper Rayleigh (1907b) used these results to interpret and explain anomalies found by Wood (1902) under precisely these conditions. There are in fact two types of *Wood's anomalies:* a sharp anomaly, appearing as a sudden change of intensity along the spectrum at frequencies and indices corresponding to a passing off of a higher order; and a diffuse anomaly related to resonance in the production of surface waves in the grating. Grating anomalies and electromagnetic surface modes are reviewed by Maystre (1982).

Rayleigh's *second method* (Rayleigh 1896, 1907a) uses the Jacobi expansion

$$\exp(i\alpha\cos\phi) = J_0(\alpha) + 2\sum_1^\infty i^n \cos n\phi J_n(\alpha), \tag{14.20}$$

which follows on substituting $t = ie^{i\phi}$ into the generating function for Bessel functions (Watson 1944, Sect. 2.1)

$$\exp\left(\alpha\left[t - t^{-1}\right]/2\right) = \sum_{-\infty}^{\infty} t^n J_n(\alpha) = J_0(\alpha) + \sum_1^\infty [t^n + (-)^n t^{-n}]J_n(\alpha). \tag{14.21}$$

We will demonstrate the method for the simplest case of normal incidence onto a pure cosine corrugation,

$$\zeta(x) = a\cos px \tag{14.22}$$

($c_1 = a$, all other c_n and s_n are zero in (14.3)). At normal incidence $K = 0, q = k, q_n = q'_n = \sqrt{k^2 - n^2p^2}$, and for a cosine corrugation the diffraction pattern is symmetric, with $A_n = A'_n$. Thus (14.13) becomes

$$\Psi(x,z) = e^{ikz} + A_0 e^{-ikz} + 2\sum_1^\infty A_n \cos npx e^{-iq_n z}. \tag{14.23}$$

If the electric field is parallel to the corrugations, Ψ corresponds to E_y and the boundary condition $\Psi(x, \zeta(x)) = 0$ reads

$$\exp(i\alpha\cos px) + A_0\exp(-i\alpha\cos px) + 2\sum_1^\infty A_n \cos npx \exp(-i\alpha_n \cos px) = 0, \tag{14.24}$$

where $\alpha = ka, \alpha_n = q_n a$. On applying (14.20) and setting the coefficients of $\cos npx$ equal to zero for $n = 0, 1, 2, \ldots$ one obtains an infinite set of linear equations for the coefficients A_n. The first of these, obtained by equating the coefficient of the term independent of x to zero, is

$$(1 + \mathrm{A}_0)J_0(\alpha) + 2\sum_1^\infty A_n(-i)^n J_n(\alpha_n) = 0. \tag{14.25}$$

Rayleigh (1907a) obtained the coefficients in the expansion in powers of the amplitude a of the corrugation up to the third order. The first five are given below up to the fourth order in a:

$$\begin{aligned}
A_0 &= -1 + \alpha\alpha_1 + \frac{1}{8}\alpha\alpha_1(\alpha^2 - 4\alpha\alpha_1 + 2\alpha_1^2 - 2\alpha_1\alpha_2 + \alpha_2^2)\\
A_1 &= -i\alpha + \frac{i\alpha}{8}\left(\alpha^2 + 4\alpha\alpha_1 - 3\alpha_1^2 + 2\alpha_1\alpha_2\right),\\
A_2 &= \frac{1}{2}\alpha\alpha_1 + \frac{\alpha}{48}(3\alpha^2\alpha_1 - \alpha^2\alpha_3 - 12\alpha\alpha_1^2 + 5\alpha_1^3 - 6\alpha_1^2\alpha_2 + 3\alpha_1^2\alpha_3 + 6\alpha_1\alpha_2^2 - 6\alpha_1\alpha_2\alpha_3)\\
A_3 &= \frac{i\alpha}{24}(\alpha^2 - 3\alpha_1^2 + 6\alpha_1\alpha_2)\\
A_4 &= -\frac{\alpha}{48}(\alpha^2\alpha_3 + \alpha_1^3 - 3\alpha_1^2\alpha_3 - 3\alpha_1\alpha_2^2 + 6\alpha_1\alpha_2\alpha_3)
\end{aligned} \tag{14.26}$$

The infinite set of linear equations for the spectral amplitudes may be solved, approximately but without assuming α to be small, by truncating at some $n = N$. That is, all A_n for $n > N$ are set equal to zero, and the linear system is solved for the $N + 1$ unknowns $A_0, A_1, \ldots, A_N$. The solution is checked by increasing N to see if this produces an appreciable difference. Another check is to solve the modified but equally valid systems obtained by multiplying (14.24) by an arbitrary power of

$\exp(i\cos px)$. For example, the terms independent of x in (14.24) multiplied by $\exp(i\alpha\cos px)$ or $\exp(-i\alpha\cos px)$ are respectively

$$J_0(2\alpha) + A_0 + 2\sum_1^\infty A_n i^n J_n(\alpha - \alpha_n) = 0, \tag{14.27}$$

$$1 + A_0 J_0(2\alpha) + 2\sum_1^\infty A_n (-i)^n J_n(\alpha + \alpha_n) = 0. \tag{14.28}$$

The preceding analyses were based on the Rayleigh hypothesis, namely on the assumption that the wave function (14.10) is sufficiently general to satisfy the boundary conditions on an arbitrary corrugated surface. The hypothesis turns out to be true for some cases and false in others, but in a restricted least-squares sense it can always be applied. Petit and Cadilhac (1966) showed that for the pure cosine corrugation $\zeta(x) = a\cos px$, and with $\Psi = 0$ on the boundary, the Rayleigh hypothesis breaks down for $pa > 0.4477432\ldots$, that is when the amplitude a of the corrugation is greater than about 7 % of its period $2\pi/p$. This number comes from solving the transcendental equation

$$e^\beta = \frac{\beta+1}{\beta-1}, \quad pa = (\beta - \beta^{-1})/2, \tag{14.29}$$

which is obtained by considering properties of the solution (14.10) analytically continued across the boundary $z = \zeta(x)$. Millar (1971) later demonstrated that the Rayleigh hypothesis is valid for this problem when pa is smaller than the critical value given above. These papers thus established that there are situations in which the Rayleigh hypothesis is valid and others in which it is not. From the practical point of view, the Rayleigh hypothesis may always be used if the coefficients are determined by satisfying the boundary conditions in the least-squares sense, since it has been shown that there is a linear combination of N elements of the set of plane waves in (14.10) that converges on the boundary to the prescribed boundary values, in the mean-square sense, as $N \to \infty$ (Yasuura 1971; Millar 1973).

The Rayleigh methods can be used to study scattering by non-periodic surfaces $z = \zeta(x)$ or $z = \zeta(x, y)$, since the function ζ can artificially be made periodic by repetition of a large section of the surface, and ζ can then be expanded in a single or double Fourier series. We noted above (see (14.6), (14.11) and the discussion following (14.19)) that evanescent waves, with imaginary components of the normal component of the wavenumber, and corresponding to grating orders that have "passed off", are associated with the production of surface waves. A rough surface thus enables coupling of an incoming plane wave to electromagnetic surface waves. For metal surfaces this effect is important in the study of surface roughness and of surface electromagnetic waves in metals (referred to in solid state physics as surface plasmons or surface polaritons). Reviews of this field may be found in the collection edited by Agranovich and Mills (1982): see in particular the chapters by

Raether (1982) and Maradudin (1982); further references will be given at the end of this chapter. For smooth surfaces, coupling to electromagnetic surface waves is possible via attenuated total reflection, as discussed in Sects. 10.6 and 12.6.

14.3 Scattering of Light by Liquid Surfaces

A clean and undisturbed liquid surface, such as that of mercury or water, gives an impression of perfect smoothness. The liquid surface is however roughened by thermal excitation of surface waves. Mandelstam (1913) calculated the angular distribution of the light scattered by the thermally induced fluctuations in a liquid surface, in the plane of incidence. His calculation is based on expanding the surface distortion in terms of a double Fourier series,

$$\zeta(x,y,t) = \sum_k e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\zeta_{\boldsymbol{k}}(t). \tag{14.30}$$

(In this section $\boldsymbol{k}$ is a two-dimensional wavevector in the plane of the surface, and $\boldsymbol{r} = (x,y)$ a two-dimensional position vector in the plane.) It follows from the equations of continuum hydrodynamics that the Fourier components $\zeta_{\boldsymbol{k}}$ are the normal mode coordinates for the surface vibrations, since the Hamiltonian of the excitations is, to second order in $\zeta_{\boldsymbol{k}}$,

$$H = \frac{1}{2}A\sum_k \left\{\frac{\varrho}{k}|\partial_t\zeta_{\boldsymbol{k}}|^2 + (\rho g + \sigma k^2)|\zeta_{\boldsymbol{k}}|^2\right\}, \tag{14.31}$$

where A is the area of the surface, ρ is the mass density of the liquid, g is the acceleration due to gravity, and σ is the surface tension. Comparison of (14.31) with the harmonic oscillator Hamiltonian $\frac{1}{2}m\left[(\partial_t x)^2 + \omega^2 x^2\right]$ shows that the effective mass m_k and angular frequency ω_k of the mode corresponding to ζ_k are given by

$$m_k = \frac{\rho A}{k}, \quad \omega_k^2 = gk + \frac{\sigma k^3}{\rho}. \tag{14.32}$$

From statistical mechanics of the harmonic oscillator the thermal average of $|\zeta_{\boldsymbol{k}}|^2$ is

$$\left\langle|\zeta_{\boldsymbol{k}}|^2\right\rangle = \frac{\hbar}{2m_k\omega_k}\coth\frac{\hbar\omega_k}{2T}, \tag{14.33}$$

where $\hbar$ is Planck's constant divided by 2π and T is the temperature, expressed in energy units. When $T \gg \hbar\omega_k$, which holds for most of the surface excitations of

classical liquids but not for those of the quantum liquids He^3, He^4 and H_2, (14.33) gives

$$\left\langle |\zeta_{\boldsymbol{k}}|^2 \right\rangle \approx \frac{T}{m_k \omega_k^2} = \frac{T}{A(\rho g + \sigma k^2)}. \tag{14.34}$$

Mandelstam's calculation was based on Rayleigh's theory of gratings, as outlined in the last section. There we assumed for simplicity that the corrugated surface was a perfect reflector. This would be a good approximation for the surfaces of liquid metals, but a very poor approximation if applied to dielectric fluids. We will therefore give a condensed treatment of Rayleigh's calculation for reflection and scattering off a corrugated interface between two media of dielectric functions ε and $\tilde{\varepsilon}$. A plane electromagnetic wave of angular frequency ω with wavevector in the zx plane is incident at angle θ onto a one-dimensionally corrugated surface

$$\zeta(x) = \sum \zeta_n e^{inpx}, \tag{14.35}$$

where the sum is over positive and negative integers n but excluding zero. This Fourier series is equivalent to (14.3) if $\zeta_n = (c_n - is_n)/2$ for positive n and $(c_n + is_n)/2$ for negative n. The incident, specularly reflected, and nth order diffracted waves are again given by

$$\psi_i = e^{i(Kx+qz)}, \quad \psi_0 = e^{i(Kx-qz)}, \quad \psi_n = A_n e^{i(K_n x - q_n z)}, \tag{14.36}$$

where the tangential and normal components of the wavevectors are related by

$$K^2 + q^2 = \varepsilon\omega^2/c^2, \quad K = \sqrt{\varepsilon}(\omega/c)\sin\theta, \quad q = \sqrt{\varepsilon}(\omega/c)\cos\theta, \tag{14.37}$$

$$K_n^2 + q_n^2 = \varepsilon\omega^2/c^2, \quad K_n = \sqrt{\varepsilon}(\omega/c)\sin\theta_n, \quad q_n = \sqrt{\varepsilon}(\omega/c)\cos\theta_n, \tag{14.38}$$

with K_n given by the grating equation (14.6) or (14.6′), $K_n = K + np$. The total wave on the incidence side of the surface is

$$\Psi = \psi_i + \psi_0 + \sum_{n\neq 0} \psi_n. \tag{14.39}$$

The total wave on the other side is

$$\tilde{\Psi} = \tilde{\psi}_0 + \sum_{n\neq 0} \tilde{\psi}_n = B_0 e^{i(Kx+\tilde{q}z)} + \sum_{n\neq 0} B_n e^{i(K_n x + \tilde{q}_n z)}, \tag{14.40}$$

where

$$K^2 + \tilde{q}^2 = \tilde{\varepsilon}\,\omega^2/c^2, \quad K = \sqrt{\tilde{\varepsilon}}(\omega/c)\sin\tilde{\theta}, \quad \tilde{q} = \sqrt{\tilde{\varepsilon}}(\omega/c)\cos\tilde{\theta}, \tag{14.41}$$

$$K_n^2 + \tilde{q}_n^2 = \tilde{\varepsilon}\,\omega^2/c^2, \quad K_n = K + np, \quad \tilde{q}_n = \sqrt{\tilde{\varepsilon}}(\omega/c)\cos\tilde{\theta}_n \tag{14.42}$$

We will consider the case where the electric field vector is along the corrugations. Then Ψ and $\tilde{\Psi}$ represent the electric field, and the boundary conditions are the equality of Ψ and $\tilde{\Psi}$ and of their normal derivatives on $z = \zeta(x)$. The latter condition implies that

$$(\partial_z\Psi - \partial_x\Psi\, d\zeta/dx)_{z=\zeta} = \left(\partial_z\tilde{\Psi} - \partial_x\tilde{\Psi} d\zeta/dx\right)_{z=\zeta} \tag{14.43}$$

(compare (14.17)). The equality of Ψ and of $\tilde{\Psi}$ on $z = \zeta(x)$ gives, after removal of the common factor e^{iKx},

$$e^{iq\zeta} + A_0 e^{-iq\zeta} + \sum_{n\neq 0} A_n e^{i(npx - q\zeta)} = B_0 e^{i\tilde{q}\zeta} + \sum_{n\neq 0} B_n e^{i(npx + \tilde{q}\zeta)}. \tag{14.44}$$

To first order in ζ this reads

$$1 + A_0 - B_0 + \sum_{n\neq 0}(A_n - B_n)e^{inpx} + i\zeta(q - qA_0 - \tilde{q}B_0) = 0. \tag{14.45}$$

The zeroth order part gives $B_0 = 1 + A_0$, and the first order part (on using (14.35) and equating coefficients of e^{inpx}) gives

$$B_n - A_n = i\zeta_n\{q(1 - A_0) - \tilde{q}B_0\}. \tag{14.46}$$

The boundary condition (14.43) may be written as

$$\partial_z\left(\Psi - \tilde{\Psi}\right) = \partial_x\left(\Psi - \tilde{\Psi}\right)\mathrm{d}\zeta/\mathrm{d}x \quad on \;\; z = \zeta(x), \tag{14.47}$$

in which we use

$$\Psi - \tilde{\Psi} = e^{iKx}\left\{e^{iqz} + A_0 e^{-iqz} - B_0 e^{i\tilde{q}z} + \sum_{n\neq 0} e^{inpx}(A_n e^{-iq_n z} - B_n e^{i\tilde{q}_n z})\right\}. \tag{14.48}$$

Thus (14.47) reads, to the first order in ζ,

$$\begin{aligned} &q(1 - A_0) - \tilde{q}B_0 + i\zeta\left[q^2(1 + A_0) - \tilde{q}^2 B_0)\right] - \sum_{n\neq 0} e^{inpx}(q_n A_n + \tilde{q}_n B_n) \\ &\quad = pK(1 + A_0 - B_0)\sum_{n\neq 0} n\zeta_n e^{inpx}. \end{aligned} \tag{14.49}$$

The zeroth order part of (14.49) is $q(1-A_0)=\tilde{q}B_0$, and the first order part gives, on equating coefficients of e^{inpx} and using the fact that $1+A_0=B_0$ (which makes the right-hand side of (14.49) zero),

$$i\zeta_n B_0\left(q^2-\tilde{q}^2\right)=q_nA_n+\tilde{q}_n^2B_n. \tag{14.50}$$

The zeroth order parts $1+A_0=B_0$ and $q(1-A_0)=\tilde{q}B_0$ give the specular coefficients

$$A_0=\frac{q-\tilde{q}}{q+\tilde{q}},\quad B_0=\frac{2q}{q+\tilde{q}}. \tag{14.51}$$

Thus A_0 and B_0 are the Fresnel reflection and transmission amplitudes for a flat surface (Sect. 1.1). Also (14.46) implies $B_n=A_n$ (to this order in ζ) and (14.50) gives

$$A_n=2i\zeta_nq\frac{q-\tilde{q}}{q_n+\tilde{q}_n}=2i\zeta_n\sqrt{\varepsilon}(\omega/c)\frac{\cos\theta\sin\left(\tilde{\theta}-\theta\right)}{\sin\tilde{\theta}\cos\theta_n+\sin\theta\cos\tilde{\theta}_n}. \tag{14.52}$$

Equation (14.52) gives the angular dependence of the amplitude of the non-specularly reflected or transmitted radiation due to the Fourier component ζ_n of the corrugation. The average intensity is proportional to $\left\langle|A_n|^2\right\rangle$ and thus to $\left\langle|\zeta_n|^2\right\rangle$; the latter is found from (14.34) on identifying the wavenumber k of the surface waves with the change in the lateral component of the radiation wavevector. From (14.42) this is $np=K_n-K=\sqrt{\varepsilon}(\omega/c)(\sin\theta_n-\sin\theta)$. The intensity (in the plane of incidence) scattered into the direction θ_n by a thermally roughened surface is thus proportional to

$$\left\langle|A_n|^2\right\rangle=4\varepsilon\frac{\omega^2}{c^2}\frac{T/A}{\rho g+\sigma(K_n-K)^2}\left\{\frac{q(q-\tilde{q})}{q_n+\tilde{q}_n}\right\}^2. \tag{14.53}$$

This, in essence, is the result of Mandelstam (1913). Andronov and Leontovich (1926) and Gans (1926) later obtained formulae for the scattered intensity which are not restricted to the plane of incidence. Bouchiat and Langevin (1978) have extended these results by including scattering by surface structural and orientational fluctuations, but still keeping the assumption of a discontinuous interface. Earlier, in an unpublished thesis, Triezenberg (1973) had in fact constructed a perturbation theory for light scattering from thermal fluctuations in a thin diffuse interface. Triezenberg's results are restricted to the *s* polarization (electric field perpendicular to the plane of incidence), and thus are not applicable to ellipsometric data. Beaglehole (1980) first gave a theoretical expression for the ellipsometric ratio

r_p/r_s, to first order in the distortions of an interface. It takes the form (3.46), namely that obtained for a planar diffuse interface, with

$$I_1 = \frac{(\varepsilon - \tilde{\varepsilon})^2}{\varepsilon + \tilde{\varepsilon}} \frac{Tk_{\max}}{\pi\sigma}, \tag{14.54}$$

where $k_{\max}$ is a wavenumber cut-off for the thermally excited surface waves. This result was deduced using the work of Bedeaux and Vlieger (1973) and Kretschman and Kroger (1975). A more complete theory leading to (14.54) has been given by Zielinska et al. (1981, 1983). Further discussion may be found in Beag1ehole (1982, 1983).

14.4 The Surface Integral Formulation of Scattering by Rough Surfaces

The preceding two sections have introduced and applied the Rayleigh method of calculating scattering by rough surfaces. An alternative method is based on the Kirchhoff formulation of diffraction theory. It is sometimes called the Kirchhoff method, although the approach was developed by Antokolskii, Brekhovskikh, Isakovich, Davis, Beckmann and others. References to the original work may be found in Beckmann and Spizzichino (1963), Shmelev (1972) and Bass and Fuks (1979); the outline given below is based on Beckmann's part of the book by Beckmann and Spizzichino. A more detailed treatment is given in Chap. 4 of Ogilvy (1991) and Chap. 7 of Nieto-Vesperinas (2006).

The method to be described is based on the following theorem of Helmholtz (Baker and Copson 1950, Sect. 4.2): if E is a solution of $(\nabla^2 + k^2)E = 0$ whose first and second partial derivatives are continuous within and on a closed surface S, R is the distance from a fixed point P and ∂_n denotes differentiation along the inward normal to S, then if P lies inside S,

$$E(P) = \frac{1}{4\pi} \iint \mathrm{d}S \left\{ E\partial_n \left(\frac{e^{ikR}}{R} \right) - \frac{e^{ikR}}{R} \partial_n E \right\}. \tag{14.55}$$

Thus the solution of the wave equation $(\nabla^2 + k^2)E = 0$ at an interior point of a region can be found in terms of the values of E and $\partial_n E$ on the boundary of the region.

Consider now the evaluation of the integral in (14.55) in terms of the (presumed known) values of E on a rough surface, when the point P is taken to be far from the surface. If R is the distance to P from a radiating element on the surface, which is at $\boldsymbol{r}$ relative to an origin near the surface, and $\boldsymbol{k}$ and $\boldsymbol{k}'$ are the wavevectors of the incoming and scattered radiation, then $R \approx R_0 - \boldsymbol{k}'.\boldsymbol{r}/k$, where R_0 is the distance

from the origin to P, and k is the magnitude of $\boldsymbol{k}$ and of $\boldsymbol{k}'$. In the surface integral (14.55), the spherical wave radiating from the element at $\boldsymbol{r}$ may be approximated as

$$\frac{e^{ikR}}{R} \approx \frac{e^{ikR_0}}{R_0} e^{-i\boldsymbol{k}'.\boldsymbol{r}}. \tag{14.56}$$

The unknowns in (14.55), namely the field E and its normal derivatives $\partial_n E$, are approximated from the field that would be present at a given point on the surface if the surface were replaced by its tangent plane at that point. If the radiation is excited by an incident plane wave $\exp(i\boldsymbol{k}.\boldsymbol{r})$, and is assumed to have the electric field locally along the surface (the latter can only be true in an average sense for a surface with two-dimensional roughness), then E corresponds to E_y when zx is the plane of incidence. If further the surface is assumed to reflect perfectly, the scattered intensity which follows from the squared modulus of (14.55) is shown to be proportional to

$$F^2 \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y \, e^{ix\Delta K_x + iy\Delta K_y} \chi(\Delta q, \varrho). \tag{14.57}$$

Here ΔK_x and ΔK_y are the x and y components of $\boldsymbol{k} - \boldsymbol{k}'$, and $\Delta q = q + q'$ is the z component of $\boldsymbol{k} - \boldsymbol{k}'$. The function χ is the expectation value of $\exp[i\Delta q(\zeta_1 - \zeta_2)]$, where ζ_1 and ζ_2 are the vertical displacements of the surface (from a mean value of zero) at two points separated on the surface by the distance $\varrho = \sqrt{x^2 + y^2}$. The factor F depends on the angle of incidence θ, and on the direction of the scattered radiation (θ', ϕ') with θ and θ' being measured relative to the normal to the averaged surface, and ϕ' being the azimuthal deviation of the scattered radiation from the specular plane:

$$F(\theta, \theta', \phi') = \frac{1 - \cos\theta\cos\theta' - \sin\theta\sin\theta'\cos\phi'}{\cos\theta(\cos\theta + \cos\theta')}. \tag{14.58}$$

On transforming the integration variables from x, y to ϱ, ϕ via $x = \varrho\cos\phi, y = \varrho\sin\phi$, (14.57) becomes

$$F^2 \int_0^{\infty} \mathrm{d}\varrho \; \varrho \; \chi(\Delta q, \varrho) \int_0^{2\pi} \mathrm{d}\phi \, e^{i(\alpha\cos\phi + \beta\sin\phi)}, \tag{14.59}$$

where $\alpha = \varrho\Delta K_x$ and $\beta = \varrho\Delta K_y$. The integral over ϕ is equal to $2\pi J_0(\sqrt{\alpha^2 + \beta^2})$; this follows from the Jacobi expansion (14.20). Thus the scattered intensity is proportional to

$$2\pi F^2 \int_0^\infty \mathrm{d}\varrho\ \varrho\ \chi(\Delta q, \varrho) J_0(\varrho \Delta K), \tag{14.60}$$

where $\Delta K = \sqrt{\Delta K_x^2 + \Delta K_y^2}$. The scattered intensity is seen to depend on both Δq and ΔK, the magnitudes of the normal and lateral components of the change in wavevector due to the scattering, $\boldsymbol{k} - \boldsymbol{k}'$.

An important special case is that of scattering by a normally distributed surface (Beckmann, Sect. 5.3) for which

$$\chi(\Delta q, \varrho) = \exp\{-\mathcal{R}[1 - C(\varrho)]\}. \tag{14.61}$$

In (14.61) $\mathcal{R}$ is the roughness parameter defined in Sect. 14.1,

$$\mathcal{R} = (q + q')^2 \langle \zeta^2 \rangle = (\Delta q)^2 \langle \zeta^2 \rangle, \tag{14.62}$$

and $C(\varrho)$ is the correlation function for vertical displacements of the surface, characterized by a lateral correlation length l:

$$C(\varrho) = \frac{\langle \zeta(0)\zeta(\varrho) \rangle}{\langle \zeta^2 \rangle} = \exp(-\frac{\varrho^2}{l^2}). \tag{14.63}$$

In this case (14.60) becomes

$$2\pi F^2 e^{-\mathcal{R}} \sum_0^\infty \frac{\mathcal{R}^n}{n!} \int_0^\infty \mathrm{d}\varrho\ \varrho \exp(-n\frac{\varrho^2}{l^2}) J_0(\varrho \Delta K). \tag{14.64}$$

The integral in (14.64) is known as Weber's first exponential integral (Watson 1944, Sect. 13.3), and is equal to $(l^2/2n)\exp[-(l\Delta K)^2/4n]$. The $n = 0$ term is zero, except for exactly specular reflection ($\Delta K = 0$), in which case it diverges. This divergence is a consequence of having calculated the intensity for an infinite and perfectly reflecting surface (see also the discussion following (3.34) in Rice 1951). We omit the $n = 0$ term; the non-specular intensity is thus proportional to

$$\pi F^2 l^2 e^{-\mathcal{R}} \sum_1^\infty \frac{\mathcal{R}^n}{n!n} \exp[-(l\Delta K)^2/4n]. \tag{14.65}$$

The series in (14.65) has the upper and lower bounds

$$\mathcal{R} \exp\left[-\frac{(l\Delta K)^2}{4}\right] < \sum_1^\infty \frac{\mathcal{R}^n}{n!n} \exp\left[-\frac{(l\Delta K)^2}{4n}\right] \leq \int_0^{\mathcal{R}} \mathrm{d}r \frac{(e^r - 1)}{r}. \tag{14.66}$$

The lower bound comes from taking the first term in the series, the upper bound from setting $l\Delta K = 0$. For a slightly rough surface (small $\mathcal{R}$), (14.65) is approximately equal to

$$\pi F^2 l^2 \mathcal{R} \exp[-\mathcal{R} - (l\Delta K/2)^2]. \tag{14.67}$$

For other than specular reflection the intensity depends on both the mean-square height variation $\langle \zeta^2 \rangle$, and on the latera1 correlation length l.

The specular intensity has to be obtained separately. Beckmann shows that it depends on $(\Delta q)^2 \langle \zeta^2 \rangle$ only, being proportional to $e^{-\mathcal{R}}$. Details may be found in Beckmann (Sect. 5.3) and Chandley and Welford (1975); see also Bennett and Porteus (1961) for a discussion and experimental test of the $e^{-\mathcal{R}}$ law at normal incidence, and Nieto-Vesperinas and Garcia (1981) for an analysis of the validity of the surface integral method.

14.5 Absorbing and Rough Surfaces that Are Wet

From early childhood we learn to distinguish wet from dry, not just by touch, but also by sight. Many objects, notably those with rough and absorbing surfaces, are darker when wet: they reflect less light. Ångström (1925) noted that the 'reflection power of the surface of the earth' needs to be known in order to understand 'the heat economy of the earth's atmosphere and the circulation of energy within it'. He suggested why a rough absorbing surface reflects less light when wet: diffuse reflection from the rough surface. If the rough absorbing surface is covered by a

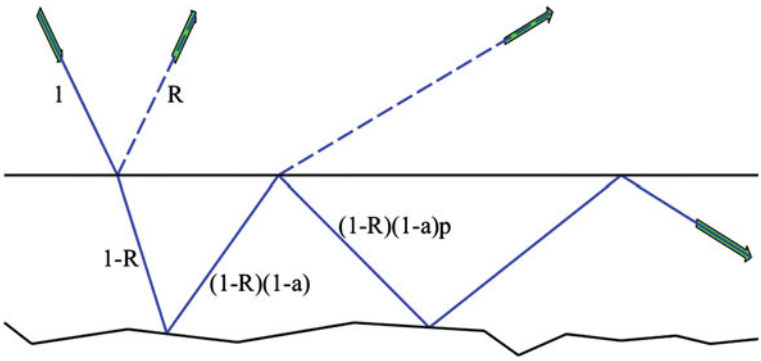


Fig. 14.2 Liquid layer on a rough surface. The fraction of light intensity along paths in the liquid layer is indicated. The dashed lines indicate light which contributes to the albedo

thin layer of water, the diffuse reflection leads to *internal reflection* from the water-air interface, and thus more absorption.

This idea is illustrated in Fig. 14.2: a fraction $1 - R$ (R is the reflectivity at the liquid surface) is transmitted into the liquid layer, a fraction a of this is absorbed at the solid surface, and $(1 - R)(1 - a)$ reflected. At the liquid surface, there is probability p of reflection back to the absorbing solid, and so on. (A planar liquid surface is drawn, but the idea works for an arbitrarily curved or distorted surface).

The probability of absorption by the rough surface is

$$P = (1 - R)\left[a + a(1 - a)p + a(1 - a)^2 p^2 + \ldots\right] = \frac{(1 - R)a}{1 - (1 - a)p}. \qquad (14.68)$$

Ångström evaluated probability p of internal reflection at the liquid surface by assuming that all of the light with internal angle of incidence greater than the critical angle $\theta_c = \arcsin(1/n)$ for the liquid is reflected (as it is) and that none incident at smaller angle is reflected (an approximation). His result, assuming a Lambertian surface, which has intensity reflected at angle θ proportional to its area projection onto the outgoing direction, namely to $\cos\theta$, is

$$p = \cos^2\theta_c = 1 - 1/n^2. \qquad (14.69)$$

The equations (14.68) and (14.69) are together equivalent to Ångström's result, except that he also omitted the $(1 - R)$ factor. The effect of reflection below the critical angle can be calculated (Lekner and Dorf 1988) but the resulting p is only about 10 % bigger for water.

The albedo $A = 1 - P$ for a wetted surface is plotted against the corresponding dry value (no liquid film) in Fig. 3 of Lekner and Dorf (1988). The wet albedo is found to be smaller than the dry albedo except in the limit of very small absorption.

A very different explanation of the wet-dark effect was given by Bohren (1983) and Twomey et al. (1986) in terms of multiple scattering by particulate matter. The

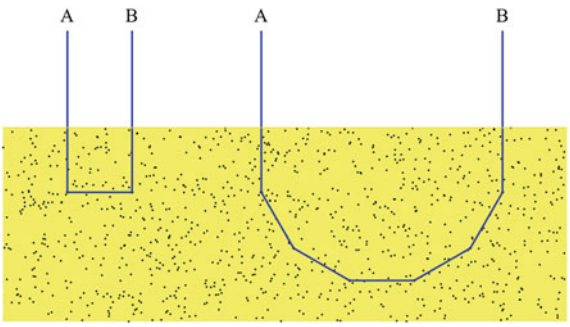


Fig. 14.3 A light ray requires just two scatterings at 90° to re-emerge, but six scatterings at 30° (the shortest possible path is assumed in both cases, with equal distances between scattering). Enhanced forward scattering increases the likelihood of absorption

basic idea is that water or other wetting liquids decrease the relative refractive index and thereby increases the degree of forward scattering. (An extreme case is perfect index matching, in which case there is no scattering, or equivalently, entirely forward scattering.) In wet sand, for example, each sand grain scatters more in the forward direction than in dry sand. This is schematically illustrated in Fig. 14.3, based on Fig. 3 of Bohren's (1983) paper.

One interesting and verifiable consequence of the multiple scattering explanation is that the better the match of the liquid refractive index with that of the scatterer, the more forward bias there is in the scattering, and the darker the wetted substance. The Bohren papers show sand ($n \approx 1.5$) wetted with water ($n \approx 1.3$) and with benzene ($n \approx 1.4$). The latter is much darker, because of the closer index match. The effect of index matching is also in the Ångström theory through the probability of absorption a, which is greater when the solid is wetted, because it reflects less. In fact the multiple scattering approach gives a graph of wet versus dry albedo (Twomey, Bohren and Mergenthaler, Fig. 5) almost the same as Fig. 3 of Lekner and Dorf (1988), based on Ångström's idea and its developments.

The two theories, based on such different approaches, clearly apply to different circumstances. Ångström's idea is best for surfaces such as asphalt or concrete, the Bohren idea of multiple scattering together with enhanced forward scattering applies best to porous granular materials such as sand. Clothing fabric is an intermediate case.

14.6 Coherent Backscattering

Suppose that in Fig. 14.3 the wave follows the paths $A \to B$, and simultaneously the reversed paths $B \to A$ as well. (We assume an incoming wave front wider than AB in both cases.) *The two amplitudes corresponding to the direct and reversed paths will be in phase, and will thus interfere constructively*. Thus we can expect enhanced reflection in the backscattering direction. For perfect coherence and exact reversal the intensity is expected to peak at twice the nearby intensity, since coherent superposition of two waves ψ gives intensity $|2\psi|^2$, whereas the incoherent (random phase) intensity is $2|\psi|^2$. The coherent backscattering effect is generally weak, and may be difficult to measure because of occlusion of the detector by the source. Enhanced backscattering is just visible from the air at high altitude above *dry* ground, in a narrow angle about the shadow point of the aircraft. Wet ground and especially clouds give strong backscattering, the 'glory', but this is due to the large backward scattering amplitude of *individual* dielectric spheres, the water droplets (Bohren and Huffman (1983); Nussenzveig (2002)). Akkermans et al. (1986) discuss coherent backscattering by disordered media, and show theoretically that the ideal backscattering intensity maximum is twice the background, as the simple argument given above suggests.

Measurements dating back to the 1920's showed an anomalously strong reflectance of the lunar surface near full moon; Gehrels et al. (1964) deduced that

reflection from the moon's surface can be enhanced by up to a factor two in the retro-direction. In exploring the possible cause of this enhanced backscattering Oetking (1966) found that volume scatterers can produce backscattering peaks for angles of incidence up to 50°. Egan and Hilgeman (1976) saw similar backscattering effects in photometric standards and paints and Becker et al. (1985) found unusually strong backscattering from soils at infrared wavelengths. O'Donnell and Mendez (1987) studied surfaces whose height fluctuations are approximately Gaussian. They found that when the lateral correlation length is larger than a wavelength and the surface slopes are mild, the Beckmann theory (Sect. 14.4) gave good agreement with measured reflectances, provided the angle of incidence was not too large. However, for stronger slopes, enhanced backscattering was observed. An extensive series of simulations of scattering from rough surfaces have been carried out by Maradudin and collaborators (Simonsen et al. 2010a, b, 2011; Nordam et al. 2014). Well-defined backscattering peaks were found in all of these numerical studies.

References

Agranovich VM, Mills DL (eds) (1982) Surface polaritons: electromagnetic waves at surfaces and interfaces, North-Holland

Akkermans E, Wolf PE, Maynard R (1986) Coherent backscattering of light by disordered media: analysis of peak line shape. Phys Rev Lett 56:1471–1474

Andronov AA, Leontovich MA (1926) Zur Theorie der molekularen Lichtzerstreuung an Flüssigkeitsoberflächen. Z Physik 38:485–501

Ångström AK (1925) The albedo of various surfaces of ground. Geografiska Annaler 7:323

Baker BB, Copson ET (1950) The mathematical theory of Huygen's principle. Clarendon Press, Oxford

Bass FG, Fuks IM (1979) Wave scattering by statistically rough surfaces, Pergamon

Beaglehole D (1980) Ellipsometric study of the surface of simple liquids. Physica 100B:163–174

Beaglehole D (1982) Short-ranged roughness on the near-critical liquid interface. Physica 112B:320–330

Beaglehole D (1983) Ellipsometry of liquid surfaces. J. de Physique C10:147–154

Becker F, Ramanantsizehena P, Stoll M (1985) Angular variation of the bidirectional reflectance of bare solis in the thermal infrared band. Appl Opt 24:365–375

Beckmann P, Spizzichino A (1963) The scattering of electromagnetic waves from rough surfaces. Pergamon, New York

Bedeaux D, Vlieger J (1973) A phenomenological theory of the dielectric properties of thin films. Physica 67:55–73

Bennett HE, Porteus JO (1961) Relation between surface roughness and specular reflectance at normal incidence. J Opt Soc Am 51:123–129

Bohren CF (1983) Multiple scattering at the beach. Weatherwise, August 1983

Bohren CF, Huffman DR (1983) Absorption and scattering of light by small particles. Wiley, New York

Bouchiat MA, Langevin D (1978) Relation between molecular properties and the intensity scattered by a liquid interface. J Colloid Interface Sci 63:193–211

Chandley PJ, Welford WT (1975) A re-formulation of some results of P. Beckmann for scattering from rough surfaces. Opt Quant Electron 7:393–397

Cox C, Munk W (1954) Measurement of the roughness of the sea surface from photographs of the sun's glitter. J Opt Soc Am 44:838–850

Egan WG, Hilgeman T (1976) Retroreflectance measurements of photometric standards and coatings. Appl Opt 15:1845–1849
Gans R (1926) Lichtzerstreuung infolge der molekularen Rauhigkeit der Trennungsfläche zweier durchsichtiger Medien. Ann Phys 79:204–226
Gehrels T, Coffeen T, Owings D (1964) Wavelength dependence of polarization. III The lunar surface. Astron J 69:826–852
Kretschmann E, Kroger E (1975) Reflection and transmission of light by a rough surface, including results for surface-plasmon effects. J Opt Soc Am 65:150–154
Lekner J, Dorf MC (1988) Why some things are darker when wet. Appl Opt 27:1278–1280
Longuet-Higgins MS (1960) Reflection and refraction at a random moving surface: I-pattern and paths of specular points; II-number of specular points in a Gaussian surface; III-frequency of twinkling in a Gaussian surface. J Opt Soc Am 50:838–856
Lynch DK (1985) Reflections on closed loops. Nature 316:216–217
Mandelstam L (1913) Uber die Rauhigkeit freier Flussigkeitsoberflächen. Ann Phys 41:609–624
Maradudin AA (1982) Interaction of surface polaritons and plasmons with surface roughness. Chapter 10 in the volume edited by Agranovich and Mills
Maystre D (1982) General study of grating anomalies from electromagnetic surface modes. In: Boardman AD (ed) Chapter 17 in Electromagnetic surface modes, Wiley, New York
Millar RF (1971) On the Rayleigh assumption in scattering by a periodic surface, 2. Proc Camb Phil Soc 69:217–225
Millar RF (1973) The Rayleigh hypothesis and a related least-squares solution to scattering problems for periodic surfaces and other scatterers. Radio Sci 8:785–796
Nieto-Vesperinas M, Garcia N (1981) A detailed study of the scattering of scalar waves from random rough surfaces. Optica Acta 289:1651–1672
Nieto-Vesperinas M (2006) Chapter 7 of scattering and diffraction in physical optics, 2nd edn. World Scientific, Singapore
Nordam T, Letnes PA, Simonsen I, Maradudin AA (2014) Numerical solutions of the Rayleigh equations for the scattering of light from a two-dimesional randomly rough perfectly conducting surface. J Opt Soc Am A 31:1126–1134
Nussenzveig HM (2002) Does the glory have a simple explanation? Opt Lett 27:1379–1381
O'Donnell KA, Mendez ER (1987) Experimental study of scattering from characterized random surfaces. J Opt Soc Am A 4:1194–1205
Oetking P (1966) Photometric studies of diffusely reflecting surfaces with applications to the brightness of the moon. J. Geophys Res 71:2505–2513
Ogilvy JA (1991) Theory of wave scattering from random rough surfaces. Adam Hilger, Bristol
Petit R, Cadilhac M (1966) Sur la diffraction d'une onde plane par un réseau infiniment conducteur. C R Acad Sci Paris B 263:468–471
Raether H (1982) Surface plasmons and roughness. Chapter 9 in the volume edited by Agranovich and Mills
Rayleigh JWS (1879) On the accuracy required in optical surfaces. Section 5 of Article 62, Scientific papers, vol I (Cambridge 1899, Dover 1964)
Rayleigh JWS (1896) Theory of sound, section 272a (Dover 1954)
Rayleigh JWS (1901) Polish. Article 268 in scientific papers, vol IV
Rayleigh JWS (1907a) On the dynamical theory of gratings. Article 322 in scientific papers, vol V
Rayleigh JWS (1907b) Note on the remarkable case of diffraction spectra described by Prof. Wood. Article 323 in scientific papers, vol V
Rice SO (1951) Reflection of electromagnetic waves from slightly rough surfaces. Commun Pure Appl Math 4:351–378
Shmelev AB (1972) Wave scattering by statistically uneven surfaces. Sov Phys Uspekhi 15: 173–183
Simonsen I, Maradudin AA, Leskova TA (2010a) Scattering of electromagnetic waves from two-dimesional randomly rough perfectly conducting surfaces: the full angular intensity distribution. Phys Rev A 81:013806 (13 pp)

Simonsen I, Maradudin AA, Leskova TA (2010b) Scattering of electromagnetic waves from two-dimesional randomly rough penetrable surfaces. Phys Rev Lett 104:233904 (4 pp)
Simonsen I, Kryvi JB, Maradudin AA, Leskova TA (2011) Light scattering from anisotropic, randomly rough, perfectly conducting surfaces. Comput Phys Commun 182:1904–1908
Triezenberg DG (1973) Capillary surface waves in a diffuse liquid-gas interface. Ph.D. Thesis, University of Maryland
Twomey SA, Bohren CF, Mergenthaler JL (1986) Reflectance and albedo differences between wet and dry surfaces. Appl Opt 25:431
Watson GN (1944) Theory of Bessel functions. Cambridge University Press, Cambridge
Wood RW (1902) On a remarkable case of uneven distribution of light in a diffraction grating spectrum. Phil Mag 4:396–402
Wood RW (1934) Physical optics. Macmillan, New York
Yasuura K (1971) A view of numerical methods in diffraction problems. Progress in radio science 1966–1969, 3:257–270, (International Union of Radio Science, Brussels)
Zielinska BJA, Bedeaux D, Vlieger J (1981, 1983) Electric and magnetic susceptibilities for a fluid-fluid interface: I the ellipsometric coefficient. Physica 107A:81–108; II critical behaviour. Physica **117**A:28–46

Further Readings

Additional references on gratings
Hutley MC (1982) Diffraction gratings. Academic Press, New York
Petit R (ed) (2013) Electromagnetic theory of gratings. Springer, Berlin
Stroke GW (1967) Diffraction gratings. Handbuch der Physik 29:426–754, Springer
Metallic rough surfaces and surface plasmons
Beaglehole D, Hunderi O (1970) Study of the interaction of light with rough metal surfaces. I-experiment, II-theory. Phys Rev B 2(309–321):321–329
Celli V, Marvin A, Toigo F (1975) Light scattering from rough surfaces. Phys. Rev. B 11:1779–1786
Elson JM, Ritchie RH (1971) Photon interactions at a rough metal surface. Phys Rev B 4:4129–4138
Hornauer DL, Raether H (1981) Determination of roughness of LiF films using guided light modes. Opt Commun 40:105–110
Hunderi O (1980) Optics of rough surfaces, discontinuous films and heterogeneous materials. Surf Sci 96:1–31
Keller JM, Fuchs R, Kliever KL (1975) *p*-polarized optical properties of a metal with a diffusely scattering surface. Phys Rev B 12:2012–2029
Kroger E, Kretschmann E (1970) Scattering of light by slightly rough surfaces on thin films including plasmon resonance emission. Z Physik 237:1–15
Marvin A, Toigo F, Celli V (1975) Light scattering from rough surfaces: general incidence angle and polarization. Phys. Rev. B 11:2777–2782
Scattering from liquid surfaces
Vrij A, Joosten JGH, Fijnaut HM (1981) Light scattering from thin liquid films. Adv Chem Phys 48:329–396
Peternelj J (1981) Scattering of light due to the density fluctuations in the interface region of a liquid-vapour system. Can J Phys 59:1009–1019
Meunier J, Langevin D (1982) Optical reflectivity of a diffuse interface. J de Physique Lett 43:185–191
Experimental studies of surface roughness
Elson JM, Bennett HE, Bennett JM (1979) Scattering from optical surfaces. Chapter 7 of vol 7, applied optics and optical engineering, Academic Press

Rufenach CL, Alpers WR (1981) Imaging ocean waves by synthetic aperture radars with long integration times. IEEE Trans AP-29:422–428
Wang Y, Wolfe WL (1983) Scattering from microrough surfaces: comparison of theory and experiment. J Opt Soc Am 73:1596–1602
Bennett JM (1985) Comparison of techniques for measuring the roughness of optical surfaces. Opt Eng 24:380–387

Laser speckle and surface roughness

Rabal HJ, Braga RA (eds) (2009) Dynamic laser speckle and applications. CRC Press, New York
Dainty JC (ed) (1975) Laser speckle and related phenomena. Springer, New York
Welford WT (1980) Laser speckle and surface roughness. Contemp. Phys. 21:401–412

Chapter 15
Particle Waves

In Sect. 1.3 we saw that there is a one-to-one correspondence between the propagation in planar-stratified media of the s polarized electromagnetic wave, and of the non-relativistic particle wave satisfying the Schrödinger equation

$$-\frac{\hbar^2}{2m}\nabla^2\Psi + V\Psi = \mathcal{E}\Psi. \tag{15.1}$$

($\mathcal{E}$ and m are particle energy and mass, V is the potential energy, and $\hbar$ is Planck's constant divided by 2π). In this chapter we give a representative selection from the main results derived in the book for the electromagnetic s wave, translated into quantum mechanical language and notation. The next chapter discusses neutron reflection, and Chap. 19 deals with the reflection of particle wavepackets.

15.1 General Results

Equation (15.1) may be written as $\nabla^2\Psi + k^2\Psi = 0$, where

$$k^2 = \frac{2m}{\hbar^2}(\mathcal{E} - V) \tag{15.2}$$

is the square of wavevector. We assume that V is a function of one spatial coordinate only, $V = V(z)$, and takes the limiting values V_1 and V_2 at $z = -\infty$ and $z = \infty$. Then

$$\frac{2m}{\hbar^2}(\mathcal{E} - V_1) = k_1^2 \leftarrow k^2(z) \rightarrow k_2^2 = \frac{2m}{\hbar^2}(\mathcal{E} - V_2). \tag{15.3}$$

For planar stratification the wave equation separates. For propagation in the zx plane the wavefunction is $\Psi = e^{iKx}\psi(z)$, where K is the x-component of the wave vector,

J. Lekner, *Theory of Reflection*, Springer Series on Atomic, Optical, and Plasma Physics 87, DOI 10.1007/978-3-319-23627-8_15

$$K = k_1 \sin\theta_1 = k_1 \sin\theta_2, \tag{15.4}$$

θ_1 and θ_2 being the angles of incidence and refraction. Equation (15.4) gives Snell's Law for particle waves, and shows that the refractive index is proportional to $(\mathcal{E} - V)^{1/2}$, that is, to the square root of the kinetic energy of the particle.

The function ψ satisfies

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}z^2} + q^2\psi = 0, \quad q^2(z) = k^2(z) - K^2. \tag{15.5}$$

The normal component of the wavevector, $q(z)$, has limiting values $q_1 = \left(k_1^2 - K^2\right)^{1/2}$ and $q_2 = \left(k_2^2 - K^2\right)^{1/2}$. The reflection and transmission amplitudes r and t are defined by the limiting forms of ψ as $z \to \pm\infty$:

$$\mathrm{e}^{iq_1 z} + r\mathrm{e}^{-iq_1 z} \leftarrow \psi(z) \to t\mathrm{e}^{iq_2 z}. \tag{15.6}$$

The left side of (15.6) represents an incident plane wave of unit amplitude, and a reflected plane wave of amplitude r. With x and time dependence included, the incident plane wave is $\exp i(Kx + q_1 z - \mathcal{E}t/\hbar)$. The factor $\exp i(Kx - \mathcal{E}t/\hbar)$ is common to all parts of the wavefunction in the propagation of plane waves through stratified media, and will usually be omitted. The time and x dependence is needed only in the treatment of wavepackets (Sect. 15.9, Chap. 19), and of finite beams (Chap. 20).The reflection and transmission amplitudes are found by solving the wave equation (15.5). In the simplest case of a potential step at (say) z_1, where $V(z)$ changes from V_1 to V_2, ψ is given by

$$\psi = \begin{cases} \mathrm{e}^{iq_1 z} + r_0\mathrm{e}^{-iq_1 z} & (z \le z_1) \\ t_0\mathrm{e}^{iq_2 z} & (z \ge z_1) \end{cases}. \tag{15.7}$$

Continuity of ψ and $\mathrm{d}\psi/\mathrm{d}z$ at z_1 gives the step reflection and transmission amplitudes

$$r_0 = \mathrm{e}^{2iq_1 z_1}\frac{q_1 - q_2}{q_1 + q_2}, \quad t_0 = \mathrm{e}^{i(q_1 - q_2)z_1}\frac{2q_1}{q_1 + q_2}. \tag{15.8}$$

Other exactly solvable potential energy profiles will be considered in the next section. Here we give some results valid for arbitrary profiles. Let $V(z)$ and $\tilde{V}(z)$ be two potential functions with the same limiting values, V_1 at $-\infty$ and V_2 at $+\infty$. Consider the reflection of plane waves by the two profiles, at the same angle of incidence (the limiting values q_1 and q_2 are thus also common to the two reflection problems). In Sect. 2.1 we showed that the corresponding reflection amplitudes r and $\tilde{r}$ are related by the comparison identities

$$r = \tilde{r} - \frac{im}{\hbar^2 q_1} \int_{-\infty}^{\infty} \mathrm{d}z \big(V - \tilde{V}\big)\psi\tilde{\psi}, \tag{15.9}$$

$$q_1(1 - r\tilde{r}^*) - q_2 t\tilde{t}^* = \frac{im}{\hbar^2} \int_{-\infty}^{\infty} \mathrm{d}z \big(V - \tilde{V}\big)\psi\tilde{\psi}^*. \tag{15.10}$$

The second identity holds for real $\tilde{V}$ only. For $\tilde{V} = V$ it gives

$$q_1\Big(1 - |r|^2\Big) = q_2|t|^2, \tag{15.11}$$

which expresses the conservation of particle flux at a non-absorbing barrier, since the probability current density

$$\mathbf{J} = \frac{\hbar}{2im}(\Psi^*\nabla\Psi - \Psi\nabla\Psi^*) = \frac{\hbar}{m}\mathrm{Im}(\Psi^*\nabla\Psi) \tag{15.12}$$

has x-component $\hbar K/m$, and z-component limiting values

$$\frac{\hbar q_1}{m}\Big(1 - |r|^2\Big) \leftarrow J_z \rightarrow \frac{\hbar q_2}{m}|t|^2. \tag{15.13}$$

Other identities can be obtained (as in Sect. 2.1) by comparing the wave incident from medium 1 with that incident from medium 2. If these are denoted by the subscripts 12 and 21 respectively, a comparison identity relating the corresponding transmission amplitudes is

$$q_2\tilde{t}_{12} - q_1 t_{21} = \frac{im}{\hbar^2} \int_{-\infty}^{\infty} \mathrm{d}z \big(V - \tilde{V}\big)\psi_{21}\tilde{\psi}_{12}. \tag{15.14}$$

Setting $\tilde{V} = V$ in (15.14) gives

$$q_2 t_{12} = q_1 t_{21}, \tag{15.15}$$

so that (15.14) may be rewritten as

$$t_{12} = \tilde{t}_{12} - \frac{im}{\hbar^2 q_2} \int_{-\infty}^{\infty} \mathrm{d}z \big(V - \tilde{V}\big)\psi_{21}\tilde{\psi}_{12}. \tag{15.16}$$

Comparison of ψ_{21} with $\tilde{\psi}^*_{12}$ gives, for real $\tilde{V}$,

$$q_2 r_{21}\tilde{t}^*_{12} + q_1 t_{21}\tilde{r}^*_{12} = -\frac{im}{\hbar^2}\int_{-\infty}^{\infty} \mathrm{d}z\big(V-\tilde{V}\big)\psi_{21}\tilde{\psi}^*_{12}. \tag{15.17}$$

For $\tilde{V} = V$ this implies

$$q_2 r_{21} t^*_{12} + q_1 t_{21} r^*_{12} = 0, \tag{15.18}$$

which gives, on using (15.15),

$$r_{21} = -\frac{t_{12}}{t^*_{12}} r^*_{12}. \tag{15.19}$$

The last relation shows that $|r_{12}|^2 = |r_{21}|^2$: the reflectance is the same in either direction for a non-absorbing barrier of arbitrary profile. A particle wave will be reflected by the same fractional amount in going up or down a potential gradient, in the absence of absorption and total reflection.

For a potential energy $V(z)$ which varies only in the interval $z_1 \le z \le z_2$, with $V = V_1$ for $z \le z_1$, and $V = V_2$ for $z \ge z_2$, general expressions for r and t may be written down in terms of two independent solutions $F(z)$ and $G(z)$ of (15.15). In the interval (z_1, z_2), $\psi = \alpha F + \beta G$. Also ψ and ψ' are continuous at z_1 and z_2, assuming that $V(z)$ has no delta function singularities (or worse) at the end-points. The continuity conditions give four equations for r, t, α and β, the solution of which gives

$$r = \mathrm{e}^{2iq_1z_1}\Big\{q_1q_2(F_1G_2 - G_1F_2) + iq_1\Big(F_1G_2' - G_1F_2'\Big) + iq_2\Big(F_1'G_2 - G_1'F_2\Big) - \Big(F_1'G_2' - G_1'F_2'\Big)\Big\}/D$$

$$\begin{aligned} t &= \mathrm{e}^{i(q_1z_1 - q_2z_2)}2iq_1W/D, \\ \alpha &= \mathrm{e}^{iq_1z_1}2iq_1(G_2' - iq_2G_2)/D, \\ \beta &= -\mathrm{e}^{iq_1z_1}2iq_1(F_2' - iq_2F_2)/D. \end{aligned} \tag{15.20}$$

In (15.20) W is the Wronskian $FG' - GF'$ (a constant), and the common denominator D is given by

$$\begin{aligned} D =& q_1q_2(F_1G_2 - G_1F_2) + iq_1\big(F_1G_2' - G_1F_2'\big) \\ & - iq_2\big(F_1'G_2 - G_1'F_2\big) + \big(F_1'G_2' - G_1'F_2'\big). \end{aligned} \tag{15.21}$$

Some general properties of r and t follow directly from (15.20), as shown in Sects. 2.2 and 2.3. For non-absorbing interfaces, $|r| = 1$ when q_2 is imaginary: total reflection occurs for

$$\theta_1 \geq \theta_c = \arcsin\left\{\frac{\mathcal{E} - V_2}{\mathcal{E} - V_1}\right\}^{1/2}. \tag{15.22}$$

As the thickness $\Delta z = z_2 - z_1$ of the profile tends to zero, r and t tend to the step (Fresnel) values given in (15.8). The probability density current conservation law (15.11) follows from (15.20), for non-absorbing interfaces. And finally, at grazing incidence the reflection amplitude tends to -1, for profiles of arbitrary shape, with or without absorption, and even for internally reflected waves. Thus Lloyd's mirror experiment (discussed for neutron reflection in Sect. 16.5) produces diffraction fringes with destructive interference at the mirror's edge, for particle as well as for electromagnetic waves.

Potential energy profiles of the form

$$V(z) = \frac{1}{2}(V_1 + V_2) - \frac{1}{2}(V_1 - V_2)f(z, a), \tag{15.23}$$

where the function f depends on parameters (collectively denoted by a) which are independent of V_1, V_2 and $\mathcal{E}$, have following property. If the reflection amplitude at normal incidence is known as a function of the limiting magnitudes k_1 and k_2 of the wavevector, the same functional form gives the reflection amplitude at oblique incidence, with the normal components of the wavevector q_1 and q_2 replacing k_1 and k_2 (see Sect. 2.5). The homogeneous, linear, and hyperbolic tangent profiles have the scaling property (15.23); the exponential and Rayleigh profiles do not. (These exactly solvable profiles are among those discussed in the next section.)

If $V(z)$ changes monotonically between V_1 and V_2, the reflectivity of this profile cannot be greater than that of the potential step between V_1 and V_2 (at the same energy and angle of incidence):

$$R \leq \left(\frac{q_1 - q_2}{q_1 + q_2}\right)^2. \tag{15.24}$$

This upper bound was obtained in Sect. 5.4.

15.2 Some Exactly Solvable Profiles

Homogeneous layer. For a potential barrier or well with V constant in $(z_1, z_2), z_2 = z_1 + \Delta z$, we have from Sect. 2.4 that

$$r = \mathrm{e}^{2iq_1z_1}\frac{q(q_1-q_2)c + i(q^2-q_1q_2)s}{q(q_1+q_2)c - i(q^2+q_1q_2)s}, \tag{15.25}$$

$$t = \mathrm{e}^{i(q_1z_1-q_2z_2)}\frac{2q_1q}{q(q_1+q_2)c - i(q^2+q_1q_2)s}, \tag{15.26}$$

where $q^2 = k^2 - K^2, \hbar^2k^2/2m = \mathcal{E} - V$, $K = k_1 \sin\theta_1$, $c = \cos q\Delta z$ and $s = \sin\Delta z$. Equivalent formulae may be written down in terms of the reflection amplitudes at the discontinuities in potential,

$$r_1 = \frac{q_1 - q}{q_1 + q}, \quad r_2 = \frac{q - q_2}{q + q_2}. \tag{15.27}$$

These are

$$r = \mathrm{e}^{2iq_1z_1}\frac{r_1 + r_2\mathrm{e}^{2iq\Delta z}}{1 + r_1r_2\mathrm{e}^{2iq\Delta z}}, \tag{15.28}$$

$$t = \mathrm{e}^{i(q_1z_1-q_2z_2)}\frac{(1+r_1)(1+r_2)\mathrm{e}^{iq\Delta z}}{1 + r_1r_2\mathrm{e}^{2iq\Delta z}}. \tag{15.29}$$

Zero reflection is possible if (i) $r_1 = r_2$ and $\mathrm{e}^{2iq\Delta z} = -1$, or if (ii) $r_1 = -r_2$ and $\mathrm{e}^{2iq\Delta z} = 1$. Condition (i) is satisfied if $q^2 = q_1q_2$ and $2q\Delta z$ is an odd multiple of π. At normal incidence $q^2 = q_1q_2$ is satisfied if

$$\mathcal{E} = U \equiv \frac{V^2 - V_1V_2}{2V - V_1 - V_2}. \tag{15.30}$$

At oblique incidence $q^2 = q_1q_2$ can be satisfied only if $\mathcal{E} > U$; it then holds at one angle,

$$\theta_1 = \arcsin\left\{\frac{\mathcal{E} - U}{\mathcal{E} - V_1}\right\}^{1/2}. \tag{15.31}$$

The second condition for zero reflection can be satisfied only when $V_1 = V_2$, and can hold at more than one angle, given by

$$\theta_1 = \arcsin\left\{\frac{\mathcal{E} - V - \frac{\hbar^2}{2m}\left(\frac{n\pi}{\Delta z}\right)^2}{\mathcal{E} - V_1}\right\}^{1/2}, \tag{15.32}$$

where n is an integer. Conditions (i) and (ii) hold when Δz is respectively equal to an odd or an even multiple of $\lambda/4$, where $\lambda = 2\pi/q$ is the effective wavelength

within the layer for propagation in the z direction. Zero reflection is not possible in the tunneling case, $\mathcal{E} < V$.

Provided V lies between V_1 and V_2, the reflectance $R = |r|^2$ of a homogeneous layer does not exceed the reflectance $(q_1 - q_2)^2/(q_1 + q_2)^2$ of a single potential step between the same values V_1 and V_2 (at the same energy and angle of incidence). This is a special case of the inequality (15.24): a monotonic profile cannot reflect more than the corresponding step profile.

From (15.28) and (15.29) we see that for real q, that is when

$$\mathcal{E} - V > (\mathcal{E} - V_1) \sin^2 \theta_1, \tag{15.33}$$

the reflectance R and transmittance $T = (q_2/q_1)|t|^2$ are periodic functions of $q\Delta z$, with period π. The reflectance

$$R = \frac{r_1^2 + 2r_1 r_2 \cos 2q\Delta z + r_2^2}{1 + 2r_1 r_2 \cos 2q\Delta z + (r_1 r_2)^2} \tag{15.34}$$

has extrema with respect to Δz when $\cos 2q\Delta z = \pm 1$; these are

$$R^+ = \left(\frac{q_1 - q_2}{q_1 + q_2}\right)^2, \quad R^- = \left(\frac{q^2 - q_1 q_2}{q^2 + q_1 q_2}\right)^2. \tag{15.35}$$

R^- is less than R^+ when V lies between V_1 and V_2; when V is outside this range, R^+ becomes the minimum value.

Linear profile. A transition region where the classical force $-\mathrm{d}V/\mathrm{d}z$ is constant has the potential energy

$$V(z) = \begin{cases} V_1 & z \le z_1 \\ V_1 + \frac{\Delta V}{\Delta z}(z - z_1) & z_1 < z < z_2 \\ V_2 & z \ge z_2, \end{cases} \tag{15.36}$$

where $\Delta z = z_2 - z_1$ and $\Delta V = V_2 - V_1$. Within (z_1, z_2) the wave equation (15.5) for motion normal to the interface can be written in terms of a dimensionless variable $\xi = q^2 a^2$ as

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}\xi^2} + \xi\psi = 0, \tag{15.37}$$

where the length a is given by

$$a = \left(\frac{\hbar^2}{2m}\left|\frac{\Delta z}{\Delta V}\right|\right)^{1/3}. \tag{15.38}$$

The standard pair of independent solutions of (15.37) are the Airy functions $Ai(-\xi), Bi(-\xi)$ (definitions and elementary properties of the Airy functions are given in Sect. 5.2). The reflection and transmission amplitudes may be obtained from (15.20), care being taken to convert between derivatives with respect to z and ξ via

$$\frac{\mathrm{d}\psi}{\mathrm{d}z} = a^2 \frac{\Delta q^2}{\Delta z} \frac{\mathrm{d}\psi}{\mathrm{d}\xi} = a^2 \frac{2m}{\hbar^2} \left(-\frac{\Delta V}{\Delta z} \right) \frac{\mathrm{d}\psi}{\mathrm{d}\xi}. \tag{15.39}$$

The linear profile has the scaling property (15.23), the function f being given by

$$f(z, z_1, z_2) = \begin{cases} -1 & z \le z_1 \\ (2z - z_1 - z_2)(z_2 - z_1) & z_1 < z < z_2 \\ 1 & z \ge z_2. \end{cases} \tag{15.40}$$

Hyperbolic tangent profile. This may be written in several equivalent ways, the first of which explicitly shows the scaling property (15.23):

$$\begin{aligned} V(z) &= \frac{1}{2}(V_1 + V_2) - \frac{1}{2}(V_1 - V_2) \tanh z/2a \\ &= \frac{V_1 + V_2 \mathrm{e}^{z/a}}{1 + \mathrm{e}^{z/a}} = \frac{V_1}{1 + \mathrm{e}^{z/a}} + \frac{V_2}{1 + \mathrm{e}^{-z/a}}. \end{aligned} \tag{15.41}$$

The solution for this profile in terms of hypergeometric functions is discussed in detail in Sect. 2.5. There is no need to translate the formulae given there into quantum mechanical notation, since they are given in terms of the variables $y_1 = q_1 a$ and $y_2 = q_2 a$ and thus can be applied directly to either electromagnetic or particle waves.

sech^2 profile. The potential energy is given by

$$V(z) = V_0 + \Delta V \operatorname{sech}^2 \frac{z}{a}. \tag{15.42}$$

The Schrödinger equation for the probability amplitude ψ reads

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}z^2} + \left[q_0^2 - \frac{2m\Delta V}{\hbar^2} \operatorname{sech}^2 \frac{z}{a} \right] \psi = 0. \tag{15.43}$$

Here q_0 is the limiting value at large $|z|$ of the normal component of the wavevector, given by

$$q_0^2 = \frac{2m}{\hbar^2}(\mathcal{E} - V_0) - K^2 = k_0^2 \cos^2 \theta, \tag{15.44}$$

where

$$k_0^2 = \frac{2m}{\hbar^2}(\mathcal{E} - V_0), K = k_0 \sin\theta. \tag{15.45}$$

A solution of (15.43) can be found in terms of the hypergeometric function, as discussed for the electromagnetic case in Sect. 4.3. The solution is characterized by two dimensionless parameters,

$$\alpha = -2ma^2\Delta V/\hbar^2, \quad \beta = q_0 a. \tag{15.46}$$

Formulae for the reflection and transmission amplitudes in terms of α and β are given in Sect. 4.3. Tunneling can occur for positive ΔV, when

$$(\mathcal{E} - V_0)\cos^2\theta < \Delta V, \qquad \text{or} \qquad \beta^2 < -\alpha. \tag{15.47}$$

Certain negative ΔV values give zero reflection, at any energy. This remarkable phenomenon is discussed in detail in Sect. 19.2.

Exponential profile. The potential energy is given by

$$V(z) = \begin{cases} V_1 & z \le z_1 \\ \mathcal{E} - (\mathcal{E} - V_1)\exp(z - z_1)/a & z_1 < z < z_2 \\ V_2 & z \ge z_2 \end{cases} \tag{15.48}$$

where the length a depends on V_1, V_2 and $\mathcal{E}$, and can be positive or negative:

$$a = \frac{(z_2 - z_1)}{\ln\left(\frac{\mathcal{E}-V_2}{\mathcal{E}-V_1}\right)}. \tag{15.49}$$

Transformation to a dimensionless independent variable proportional to the local magnitude of the wavevector,

$$u = 2ka = 2a\left\{\frac{2m}{\hbar^2}(\mathcal{E} - V)\right\}^{1/2}, \tag{15.50}$$

converts (15.5) to Bessel's equation

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}u^2} + \frac{1}{u}\frac{\mathrm{d}\psi}{\mathrm{d}u} + \left[1 - \frac{(2Ka)^2}{u^2}\right]\psi = 0. \tag{15.51}$$

The general solution within (z_1, z_2) is thus $\alpha J_s(u) + \beta Y_s(u)$, where $s = 2Ka$. Note that both the order s and the argument u of the Bessel functions are proportional to the thickness of the inhomogeneity. The order also depends on the angle of incidence, increasing from zero at normal incidence to $2k_1 a$ at grazing incidence.

Reflection and transmission amplitudes may be obtained from (15.20), on converting between derivatives with respect to z and u via

$$\frac{\mathrm{d}\psi}{\mathrm{d}z} = k\frac{\mathrm{d}\psi}{\mathrm{d}u}. \tag{15.52}$$

Some reduction of the formulae is possible by using the properties of Bessel functions. Details are given in Sect. 2.5.

Rayleigh profile. Within this potential transition the local wavelength $2\pi/k(z)$ is linear in z. It is therefore useful to work in terms of a dimensionless variable which is linear in z,

$$\eta(z) \equiv (k\Delta z)^{-1} = \eta_1 + (z - z_1)\Delta\eta/\Delta z. \tag{15.53}$$

The interface extends from z_1 to $z_2 = z_1 + \Delta z$, and

$$\Delta\eta = \eta_2 - \eta_1 = \left(k_2^{-1} - k_1^{-1}\right)/\Delta z. \tag{15.54}$$

The potential energy is given by

$$\mathcal{E} - V(z) = \frac{\hbar^2}{2m\eta^2(z)(\Delta z)^2}. \tag{15.55}$$

At normal incidence the wave equation has a simple power-law solution: on changing to the variable η the equation $\mathrm{d}^2\psi/\mathrm{d}z^2 + k^2\psi = 0$ becomes

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}\eta^2} + \frac{\psi}{\eta^2(\Delta\eta)^2} = 0, \tag{15.56}$$

and has the solutions $\psi_\pm = \eta^{1/2\pm\nu}$, where

$$\nu^2 = \frac{1}{4} - (\Delta\eta)^{-2}. \tag{15.57}$$

(The parameter ν is introduced for mathematical convenience, and to increase commonality with the results of Sect. 2.5.) On converting between z and η derivatives via

$$\frac{\mathrm{d}\psi}{\mathrm{d}z} = \frac{\Delta\eta}{\Delta z}\frac{\mathrm{d}\psi}{\mathrm{d}\eta}, \tag{15.58}$$

the reflection and transmission amplitudes at normal incidence are found from (15.20) to be

$$r_n = \mathrm{e}^{2ik_1 z_1} \frac{\frac{1}{2} i\Delta\eta[\varrho^\nu - 1]}{\varrho^\nu - 1 - i\nu\Delta\eta[\varrho^\nu + 1]}, \tag{15.59}$$

$$t_n = -\mathrm{e}^{i(k_1 z_1 - k_2 z_2)} \frac{2i\nu\Delta\eta\varrho^{\frac{\nu}{2}-\frac{1}{4}}}{\varrho^\nu - 1 - i\nu\Delta\eta[\varrho^\nu + 1]}, \tag{15.60}$$

where ϱ is the square of the ratio of the refractive indices,

$$\varrho = \left(\frac{\eta_1}{\eta_2}\right)^2 = \left(\frac{k_2}{k_1}\right)^2 = \frac{\mathcal{E} - V_2}{\mathcal{E} - V_1}. \tag{15.61}$$

The reflectivity at normal incidence takes different forms, depending on whether $(\Delta\eta)^2$ is greater than or less than four. When $(\Delta\eta)^2 > 4$, ν is real, and

$$R_n = \frac{\frac{1}{4}(\Delta\eta)^2(\varrho^\nu - 1)^2}{(\varrho^\nu - 1)^2 + (\nu\Delta\eta)^2(\varrho^\nu + 1)^2}. \tag{15.62}$$

When $(\Delta\eta)^2 < 4$, $\nu = i|\nu|$ and

$$R_n = \frac{\sin^2(\frac{1}{2}|\nu| \ln \varrho)}{4|\nu|^2 + \sin^2(\frac{1}{2}|\nu| \ln \varrho)}. \tag{15.63}$$

At $\nu = 0$ these two forms take the common value

$$R_n(\nu = 0) = \frac{\ln^2 \varrho}{16 + \ln^2 \varrho}. \tag{15.64}$$

From (15.63) we see that the normal incidence reflectivity is zero when $(\Delta\eta)^2 < 4$ and $|\nu| \ln \varrho$ is an integer multiple of 2π. This happens when the interface thickness Δz takes one of the values

$$\Delta z = \frac{|k_1 - k_2|}{k_1 k_2} \left\{ \frac{1}{4} + \left(\frac{n\pi}{ln(k_2/k_1)} \right)^2 \right\}^{1/2}, \quad n = 1, 2, \ldots \tag{15.65}$$

At oblique incidence the wave equation in the η variable reads

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}\eta^2} + \left[\frac{\frac{1}{4} - \nu^2}{\eta^2} - \left(\frac{K\Delta z}{\Delta\eta} \right)^2 \right] \psi = 0, \tag{15.66}$$

and has solutions proportional to $\eta^{1/2}$ times Bessel functions of order ν and imaginary argument $\pm i(K\Delta z/\Delta\eta)\eta$. The reflection and transmission amplitudes

may thus be obtained from (15.20). A reference to explicit expressions obtained in this way is given in Sect. 2.5.

References to listings and systematic generation of solvable profiles are given at the end of Chap. 2.

15.3 Perturbation and Variational Theories

Perturbation theory gives expressions for the unknown function ψ satisfying (15.5) and (15.6) (and thus also for its reflection and transmission amplitudes r and t) in terms of some known function ψ_0, satisfying

$$\frac{\mathrm{d}^2\psi_0}{\mathrm{d}z^2} + q^2\psi_0 = 0, \quad e^{iq_1 z} + r_0 e^{-iq_1 z} \leftarrow \psi_0 \rightarrow t_0 e^{iq_2 z}. \tag{15.67}$$

The solution is given in terms of $\psi_0(z)$ and a Green's function $G(z, \zeta)$, which satisfies

$$\frac{\partial^2 G}{\partial z^2} + q_0^2(z)G = \delta(z - \zeta), \tag{15.68}$$

and has the appropriate limiting forms to make ψ, given by the integral equation

$$\psi(z) = \psi_0(z) - \int_{-\infty}^{\infty} \mathrm{d}\zeta \Delta q^2(\zeta) G(z, \zeta)\psi(\zeta), \tag{15.69}$$

take the limiting forms (15.6). In the above, q and q_0 share the limiting values q_1 at $-\infty$ and q_2 at $+\infty$, and

$$\Delta q^2 = q^2 - q_0^2 = -\frac{2m}{\hbar^2}(V - V_0). \tag{15.70}$$

(V and V_0 also share the limiting values V_1 and V_2, and the equations for ψ and ψ_0 are to be solved at the same particle energy.) Note that the perturbation Δq^2 is independent of energy and of the angle of incidence.

In the long wave case the perturbation is built up from the step potential energy profile

$$V_0(z) = \frac{1}{2}(V_1 + V_2) - \frac{1}{2}(V_1 - V_2)\mathrm{sgn}(z), \tag{15.71}$$

for which the wavefunction ψ_0 and Green's function $G(z, \zeta)$ are given in Sect. 3.1. The perturbation theory expression for r_n (the contribution of order $(V - V_0)^n$ to the

reflection amplitude) is given in (3.12). When $V_0(z)$ is constant $(V_1 = V_2 = V_0), r_0 = (q_1 - q_2)/(q_1 + q_2)$ is zero, and r_1 is equal to the Fourier transform of the deviation of the potential from V_0:

$$r_1 = -\frac{im}{q_0\hbar^2}\int_{-\infty}^{\infty} \mathrm{d}z(V - V_0)e^{2iq_0 z}. \tag{15.72}$$

This is divergent at grazing incidence, when $q_0 \to 0$.

In the short wave case the perturbation theory is constructed from the Liouville-Green functions of Sect. 6.2:

$$\psi^+ = (q_1/q)^{1/2}\exp(i\phi), \quad \psi^- = (q_2/q)^{1/2}\exp(-i\phi), \quad \phi(z) = \int^z \mathrm{d}\zeta q(\zeta), \tag{15.73}$$

the Green's function being given in Sect. 6.5. The corresponding first order perturbation result for the reflection amplitude is

$$r^{(1)} = \frac{1}{4i}\int_{-\infty}^{\infty} \mathrm{d}z\left[\frac{\mathrm{d}\gamma}{\mathrm{d}z} + \frac{1}{2}q\gamma^2\right]e^{2i\phi}, \tag{15.74}$$

where the dimensionless function $\gamma = q^{-2}\mathrm{d}q/\mathrm{d}z$ must be small everywhere for the perturbation result to be accurate. The short wave perturbation theory thus also fails at grazing incidence.

Variational expressions for the reflection amplitude may be derived from the perturbation theories. The general variational principle is $\delta(F^2/S) = 0$, where F and S are first and second order in the unknown ψ:

$$F = \int_{-\infty}^{\infty} \mathrm{d}z\Delta q^2(z)\psi(z)\psi_0(z), \tag{15.75}$$

$$S = \int_{-\infty}^{\infty} \mathrm{d}z\Delta q^2(z)\psi^2(z) + \int_{-\infty}^{\infty} \mathrm{d}z\Delta q^2(z)\psi(z)\int_{-\infty}^{\infty} \mathrm{d}\zeta\Delta q^2(\zeta)\psi(\zeta)G(z,\zeta). \tag{15.76}$$

The variational estimate for the reflection amplitude is

$$r^{\mathrm{var}} = r_0 - \frac{F^2}{2iq_1 S}. \tag{15.77}$$

Further details are given in Chap. 4. Here we note only that the variational theory based on the long-wave perturbation theory removes the grazing incidence divergence which troubles all orders of the perturbation theory when $V_1 = V_2$.

15.4 Long Waves, Integral Invariants

In the long wave limit a given potential energy profile reflects predominantly as a step profile, with a small correction which depends on the deviation of the profile from the step $V_0(z)$ given by (15.71). This correction can be expressed as a series in the ratio of the interface thickness to the wavelength of the wave. The reflection amplitude, written as $r_0 + r_1 + r_2 \ldots$ where the subscript n here refers to the order (power) in the interface thickness, is found from the long-wave perturbation theory of the previous section to be given by

$$r_0 = \frac{q_1 - q_2}{q_1 + q_2}, \quad r_1 = \left(-\frac{2m}{\hbar^2}\right) \frac{2iq_1\mu_1}{(q_1 + q_2)^2}, \tag{15.78}$$

$$r_2 = \left(\frac{2m}{\hbar^2}\right) \frac{2q_1}{(q_1 + q_2)^3} \left\{2(q_1 + q_2)q_2\mu_2 - \frac{2m}{\hbar^2}\mu_1^2\right\},$$

where

$$\mu_n = \int_{-\infty}^{\infty} \mathrm{d}z[V(z) - V_0(z)]z^{n-1}. \tag{15.79}$$

The integrals μ_n depend on the relative positioning of the profiles V and V_0, with the exception of μ_1 in the case where V_0 is a constant ($V_1 = V_2$).

The reflectivity $R = |r|^2$, a measurable quantity, must be independent of the relative positioning of the actual and step profiles. When $q(z)$ is real everywhere, R differs from $R_0 = r_0^2$ by a term which is of second order in the interface thickness:

$$R = R_0 + |r_1|^2 + 2r_0r_2 + \ldots$$

$$= R_0 - \frac{4q_1q_2}{(q_1 + q_2)^4} \left(\frac{2m}{\hbar^2}\right)^2 \left\{2(V_1 - V_2)\mu_2 - \mu_1^2\right\} + \ldots \tag{15.80}$$

The expression in braces, which we call i_2, is invariant to the relative positioning of V and V_0. It is the first in an infinite set of integral invariants which may be constructed from the μ_n (see Sect. 3.3). The result (15.80) shows that the reflectivity takes a universal form in the long wave limit:

$$R = \left(\frac{q_1 - q_2}{q_1 + q_2}\right)^2 - \frac{4q_1q_2}{(q_1 + q_2)^4} \left(\frac{2m}{\hbar^2}\right)^2 i_2 + \ldots \tag{15.81}$$

This holds for all non-absorbing profiles $V(z)$ which do not contain delta function singularities or worse. When $V_1 = V_2$ the μ_2 integral is not needed (to this order)

and μ_1 is separately invariant. The reflectivity is then proportional to μ_1^2, and is independent of the sign of $V - V_0$, to this order. When $V_1 \neq V_2$ and V is real everywhere it is possible to position the profiles V and V_0 to make $\mu_1 = 0$. The second order invariant i_2 may then be put in the form

$$i_2 = (V_2 - V_1) \int_{-\infty}^{\infty} \mathrm{d}z \frac{\mathrm{d}V}{\mathrm{d}z} z^2 \quad (\mu_1 = 0), \tag{15.82}$$

which shows that $R - R_0$ is proportional to $(V_2 - V_1)$ times the second moment of the force $-\mathrm{d}V/\mathrm{d}z$. When V increases or decreases monotonically, i_2 is positive and the second order term decreases the reflectivity from that of a step profile, as may be expected. The last result may be strengthened by writing the second order invariant as

$$i_2 = -\int_{-\infty}^{\infty} \mathrm{d}z_1 \int_{-\infty}^{-\infty} \mathrm{d}z_2 [V(z_1) - V_0(z_1 - z_2)][V(z_2) - V_0(z_2 - z_1)]. \tag{15.83}$$

This form shows that i_2 is positive if $V(z)$ lies between V_1 and V_2 for all z, as can be seen by considering the sign of the integrand for $z_1 < z_2$ and for $z_1 > z_2$. Thus the second order term decreases the reflectivity if

$$\min(V_1, V_2) \leq V(z) \leq \max(V_1, V_2). \tag{15.84}$$

The functional form of i_2 for the profiles considered in Sect. 3.6 (all but one of which, the double exponential profile, were transcribed to the quantum mechanical case in Sect. 15.2) can be obtained from Table 3.1 by the substitution $\varepsilon \to \mathcal{E} - V$. For example, the homogeneous layer of thickness Δz and potential V has

$$i_2/(\Delta z)^2 = (V_1 - V)(V - V_2), \tag{15.85}$$

and the exponential profile of thickness Δz has

$$i_2/(\Delta z)^2 = \left\{ \frac{V_1 - V_2}{\ln \frac{\mathcal{E} - V_2}{\mathcal{E} - V_1}} \right\}^2 - (\mathcal{E} - V_1)(\mathcal{E} - V_2). \tag{15.86}$$

(The energy dependence in (15.86) comes from the energy dependence of the potential for the exponential profile, given by (15.48).)

Profiles of the form (15.23), for example the linear, tanh, error function and double exponential profiles (the last being defined in (3.69)) all have i_2 proportional to $(V_1 - V_2)^2$, provided the function f is continuous.

15.5 Riccati-Type Equations; the Rayleigh Approximation

The second order differential equation (15.5) is equivalent to a pair of coupled first order equations in ψ and $\psi' = \mathrm{d}\psi/\mathrm{d}z$. On setting

$$\psi = F + G, \quad \psi' = iq(F - G), \tag{15.87}$$

we find that F and G satisfy

$$F' = iqF - \frac{q'}{2q}(F - G), \tag{15.88}$$

$$G' = -iqG + \frac{q'}{2q}(F - G). \tag{15.89}$$

From (15.6) we see that when incidence is from medium 1, $F \to \mathrm{e}^{iq_1 z}$ and $G \to r\mathrm{e}^{-iq_1 z}$ as $z \to -\infty$. Thus the ratio $\varrho = G/F$ tends to $\mathrm{e}^{-2iq_1 z}$ times the reflection amplitude r as $z \to -\infty$, and to zero as $z \to \infty$. The equation satisfied by ϱ is of the generalized Riccati type:

$$\varrho' + 2iq\varrho - \frac{q'}{2q}\left(1 - \varrho^2\right) = 0. \tag{15.90}$$

On writing $\varrho = |\varrho|e^{i\theta}$, separating the real and imaginary parts, and integrating the equation for $|\varrho|'$ over all z, one finds

$$\ln\frac{1+|r|}{1-|r|} = -\int_{-\infty}^{\infty} \mathrm{d}z \frac{q'}{q}\cos\theta \tag{15.91}$$

(more detail may be found in Chap. 5). From this follows the inequality of Sect. 5.4,

$$R \le \left(\frac{q_1 - q_2}{q_1 + q_2}\right)^2; \tag{15.92}$$

any non-absorbing monotonic profile cannot reflect more than a step profile between the same limiting values of potential, at the same energy and angle of incidence. An alternative approach is to deal directly with the reflection amplitude, by setting

$$\psi = f\mathrm{e}^{i\phi} + g\mathrm{e}^{-i\phi}, \quad \psi' = iq\left(f\mathrm{e}^{i\phi} - g\mathrm{e}^{-i\phi}\right), \tag{15.93}$$

where ϕ is the phase integral defined in (15.73). On using $\phi' = q$ we find that f and g satisfy the equations

$$f' + \frac{q'}{2q}\left(f - g\mathrm{e}^{-2i\phi}\right) = 0, \tag{15.94}$$

$$g' + \frac{q'}{2q}\left(g - f\mathrm{e}^{2i\phi}\right) = 0. \tag{15.95}$$

If the lower limit in the integral defining ϕ is chosen so as to make $\phi \to q_1 z$ as $z \to -\infty$, the ratio g/f tends to the reflection amplitude r as $z \to -\infty$, and in fact may be interpreted as the reflection amplitude $r(z)$ for a profile truncated at z, as explained in Chap. 5. The equation satisfied by $r(z) = g/f$ is

$$r'(z) = \frac{q'}{2q}\left(\mathrm{e}^{2i\phi} - r^2(z)\mathrm{e}^{-2i\phi}\right). \tag{15.96}$$

The reflection amplitude r of the entire profile is thus

$$r = -\int_{-\infty}^{\infty} \mathrm{d}z \frac{q'}{2q}\left(\mathrm{e}^{2i\phi} - r^2(z)\mathrm{e}^{-2i\phi}\right). \tag{15.97}$$

The Rayleigh or weak reflection approximation is obtained by neglecting the term proportional to $r^2(z)$ in the integrand:

$$r_R = -\int_{-\infty}^{\infty} \mathrm{d}z \frac{q'}{2q}\mathrm{e}^{2i\phi}; \tag{15.98}$$

its long-wave limit is

$$r_R \to \frac{1}{2}\ln\frac{q_1}{q_2}. \tag{15.99}$$

The Rayleigh approximation works well at all wavelengths, provided the reflection is weak. It fails whenever the reflection is strong, for example at grazing incidence. Note that the factor q'/q in the integrand of both the exact and the approximate formulae (15.97) and (15.98) for the reflection amplitude can be written as

$$\frac{q'}{q} = \frac{(q^2)'}{2q^2} = \frac{-\mathrm{d}V/\mathrm{d}z}{\hbar^2 q^2/m}. \tag{15.100}$$

The contributions to the reflection amplitude are thus weighted by the local value of the ratio of the force $-\mathrm{d}V/\mathrm{d}z$ to the kinetic energy of motion normal to the interface, $\hbar^2 q^2/2m$.

15.6 Reflection of Short Waves

In Sect. 6.2 we saw that the Liouville-Green functions

$$\psi^+ = \left(\frac{q_1}{q}\right)^{1/2} \mathrm{e}^{i\phi}, \quad \psi^- = \left(\frac{q_2}{q}\right)^{1/2} \mathrm{e}^{-i\phi}, \quad \phi = \int^z \mathrm{d}\zeta q(\zeta) \tag{15.101}$$

are approximate solutions of the wave equation (15.5). In fact $\psi^\pm$ satisfy

$$\frac{\mathrm{d}^2\psi^\pm}{\mathrm{d}z^2} + q^2\left\{1 + \frac{1}{2}\frac{\mathrm{d}\gamma}{\mathrm{d}\phi} + \frac{\gamma^2}{4}\right\}\psi^\pm = 0, \tag{15.102}$$

where the dimensionless function $\gamma(z)$ is given by

$$\gamma = \frac{\mathrm{d}q}{q^2\mathrm{d}z} = \frac{\mathrm{d}q}{q\mathrm{d}\phi}. \tag{15.103}$$

If $\mathrm{d}\gamma/\mathrm{d}\phi (= q^{-1}\mathrm{d}\gamma/\mathrm{d}z)$ and γ^2 are small compared to unity, the functions $\psi^\pm$, and approximations to r and t resulting from their use, are expected to be accurate. Since

$$\gamma = \frac{1}{2q^3}\frac{\mathrm{d}q^2}{\mathrm{d}z} = \frac{-m\mathrm{d}V/\mathrm{d}z}{\hbar^2 q^3}, \tag{15.104}$$

where $-\mathrm{d}V/\mathrm{d}z$ is the force, $\hbar^2 q^2/2m$ the kinetic energy and $2\pi/q$ the effective wavelength (all relating to change or motion in the z direction),

$$4\pi\gamma = \frac{\text{force} \times \text{wavelength}}{\text{kinetic energy}}. \tag{15.105}$$

Thus $4\pi\gamma$ is the ratio of the potential energy change in one wavelength to the local kinetic energy of motion in the z direction.

Short wavelength approximations all depend on γ^2 and $\mathrm{d}\gamma/\mathrm{d}\phi$ being small, at least for most of the range of z. They generally fail at grazing incidence ($q_1 \to 0$), and special techniques are needed at discontinuities in the slope of the potential, and at turning points (where q^2 passes through zero). We will translate, without proof, some of the results of Chap. 6 into particle-wave language.

The Rayleigh approximation (15.98), which can be written as

$$r_R = -\frac{1}{2}\int_{-\infty}^{\infty} \mathrm{d}\phi\ \gamma e^{2i\phi} = \frac{1}{4}\int_{-\infty}^{\infty} \mathrm{d}z \frac{\mathrm{d}V/\mathrm{d}z}{\mathcal{E} - V - \hbar^2 K^2/2m} e^{2i\phi}, \tag{15.106}$$

turns out to be closely related to the first order perturbation theory result

$$r^{(1)} = -\frac{1}{2}\int_{-\infty}^{\infty} \mathrm{d}\phi \big(\gamma - \gamma^2/4i\big) e^{2i\phi}, \tag{15.107}$$

based on the approximate solutions $\psi^{\pm}$ and a Green's function constructed from them.

When the potential $V(z)$ has discontinuities in its gradient at the profile boundaries, but is otherwise smooth (as is the case for the linear, exponential and Rayleigh profiles), (15.106), (15.107), or (15.20) with $F = \psi^+$ and $G = \psi^-$, all give the reflection amplitude

$$r \approx \frac{1}{4} e^{i(\phi_1 + \phi_2)} \big\{\gamma_1 e^{-i\Delta\phi} - \gamma_2 e^{i\Delta\phi}\big\}, \tag{15.108}$$

where ϕ_1 and ϕ_2 are the values of $\phi(z)$ at z_1 and z_2, $\Delta\phi = \phi_2 - \phi_1$ is the change in the accumulated phase across the profile, and the function γ changes from zero to γ_1 at z_1 and from γ_2 to zero at z_2. Exponentially small terms, originating from the smooth variation in $V(z)$ other than at the end points, are omitted from (15.108). The resulting reflectivity is

$$R \approx \frac{1}{16}\big\{\gamma_1^2 + \gamma_2^2 - 2\gamma_1\gamma_2 \cos 2\Delta\phi\big\}. \tag{15.109}$$

The dominant part of the reflectivity thus depends quadratically on the discontinuities in the potential gradient, and shows oscillatory decay with increasing energy.

Regions where $q^2(z) < 0$, which are classically inaccessible since the kinetic energy of motion in the z direction is then negative, occur where

$$\mathcal{E} - V < (\mathcal{E} - V_1)\sin^2\theta_1. \tag{15.110}$$

The locations where $q^2(z) = 0$ are called classical turning points. When $V_2 > V_1$ and $\theta_1 \geq \theta_c = \arcsin[(\mathcal{E} - V_2)/(\mathcal{E} - V_1)]^{1/2}$, total reflection occurs. Interest then centres on the phase of the reflection amplitude, which determines the time of arrival and the shape of reflected pulses, as discussed in Sect. 19.1. We write $r = \mathrm{e}^{i\delta}$; the phase δ is then given by (in the short wave limit, on reflection from a profile with a single turning point)

$$\delta \approx 2(\phi_0 - \phi_-) - \pi/2, \tag{15.111}$$

where ϕ_0 is the value of the phase integral at the turning point, and ϕ_- is defined by

$$\phi(z) \to q_1 z + \phi_- \quad \text{as} \quad z \to -\infty. \tag{15.112}$$

The result (15.111) is derived in Sect. 6.7. A simpler version follows if one takes the (so far unspecified) lower limit of integration in the definition of ϕ to be the turning point z_0, in which case $\phi_0 = 0$. If also the origin $z = 0$ is chosen such that $q = q_1$ for $z < 0$, (15.111) becomes

$$\delta \approx 2 \int_0^{z_0} \mathrm{d}z \; q(z) - \frac{\pi}{2}. \tag{15.113}$$

In the case of a *potential barrier*, (15.110) may hold in an interval between two turning points, z_1 and z_2. The wave penetrates the classically forbidden region where $q^2 < 0$, and a part of it tunnels through to beyond z_2 where $q^2 > 0$. For well separated turning points, and smooth potential energy barriers with $V_1 = V_2$, the reflection and transmission probabilities are given in Sect. 6.8:

$$R \approx \tanh^2(\Delta\Phi + \ln 2), \quad T \approx \operatorname{sech}^2(\Delta\Phi + \ln 2). \tag{15.114}$$

Here $\Delta\Phi$ is the increment in the imaginary part of the phase between the turning points (and thus gives the exponent of the change in amplitude across the barrier),

$$\Delta\Phi = \int_{z_1}^{z_2} \mathrm{d}z |q(z)|. \tag{15.115}$$

For large $\Delta\Phi$ the results (15.114) have the limiting forms

$$R \to 1 - \mathrm{e}^{-2\Delta\Phi}, \quad T \to \mathrm{e}^{-2\Delta\Phi}. \tag{15.116}$$

15.7 Absorption, the Optical Potential

As we saw in Sect. 1.5, a medium containing scatterers can be approximated by an effective potential $V(z)$. When there is absorption, as for example in the case of neutrons by means of a nuclear reaction, or in the case of electrons by trapping or by "annihilation" with hole quasiparticles, the interaction with the medium can be approximated by a complex potential. This is in close analogy with the electromagnetic case, where absorption is represented by an imaginary part in the

dielectric function or refractive index. Because of the analogy, the complex potential is referred to as the *optical potential* in nuclear and atomic physics. For a medium with z-stratification, the propagation normal to the interface is characterized by the wave equation (15.5), with

$$q^2(z) = \frac{2m}{\hbar^2}(\mathcal{E} - V(z)) - K^2. \tag{15.117}$$

In the presence of absorption the potential has a negative imaginary part, as we shall see. Accordingly we set

$$V = V_r - iV_i, \tag{15.118}$$

with $V_i \geq 0$. The normal component of the wavevector is also complex, $q = q_r + iq_i$. From the real and imaginary parts of (15.117) we get

$$q_r^2 - q_i^2 = \frac{2m}{\hbar^2}(\mathcal{E} - V_r) - K^2, \tag{15.119}$$

$$2q_r q_i = \frac{2m}{\hbar^2} V_i. \tag{15.120}$$

In a homogeneous medium the transmitted wave is proportional to $\exp i(Kx + qz) = \exp i(Kx + q_r z)\exp(-q_i z)$. Both q_r and q_i are non-negative, and hence so is V_i. The real and imaginary components of q, for incidence at angle θ_1 from a medium with real potential V_1, are found from (15.119) and (15.120) to be given by

$$q_r^2 = \frac{m}{\hbar^2}\left\{\mathcal{E} - V_r - (\mathcal{E} - V_1)\sin^2\theta_1 + [(\mathcal{E} - V_r - (\mathcal{E} - V_1)\sin^2\theta_1)^2 + V_i^2]^{1/2}\right\}, \tag{15.121}$$

$$q_i = \frac{mV_i}{\hbar^2 q_r}. \tag{15.122}$$

($V_i = 0$ is a degenerate case; then either $q_i = 0$ or $q_r = 0$, depending on whether θ_1 is less than or greater than the critical angle θ_c given by (15.22).) In a homogenous absorbing medium (a square well or potential in quantum mechanical terms) the surfaces of constant amplitude are planes parallel to the interface, while surfaces of constant real phase are the planes $Kx + q_r z =$ constant. The normal to these planes is inclined at an angle θ_2' to the normal to the interface, where $\theta_2' = \arctan(K/q_r)$, with $\hbar^2 K^2/2m = (\mathcal{E} - V_1)\sin^2\theta_1$. In general the angle of refraction θ_2, defined by Snell's Law (15.4), is complex. It is equal to the real angle θ_2' only at normal incidence, or when V is real.

At a sharp boundary between a medium with real potential V_1 and an absorbing medium with $V_2 = V_r - iV_i$, the reflection amplitude and reflectivity are

$$r = \mathrm{e}^{2iq_1 z_1} \frac{q_1 - q_r - iq_i}{q_1 + q_r + iq_i}, \quad R = \frac{(q_1 - q_r)^2 + q_i^2}{(q_1 + q_r)^2 + q_i^2} \tag{15.123}$$

(z_1 being the location of the interface). The formulae for the homogeneous absorbing potential barrier are obtained from those of Sect. 15.2 by making q complex. The same is true for the hyperbolic tangent profile, for which formulae in the absorbing case are given in Sect. 10.7.

The conservation law (15.11) no longer holds in the presence of absorption, since particles are removed and the probability density current thus decreases into the absorbing medium. However, the reciprocity law (15.15) remains valid, and so the transmittance through an arbitrary inhomogeneous absorbing interface (between two non-absorbing media) is the same in either direction:

$$T_{12} = \frac{q_2}{q_1} |t_{12}|^2 = \frac{q_1}{q_2} |t_{21}| = T_{21}. \tag{15.124}$$

The result that $r \to -1$ at grazing incidence also holds in the presence of absorption.

A non-absorbing film on an absorbing substrate can give zero reflection (and thus total absorption) at an angle given by

$$\left(q^2 - q_1 q_r\right)(q_r - q_1) = q_1 q_i^2, \tag{15.125}$$

where q_1, q, and $q_r + iq_i$ are the normal components of the wavevector in the first medium, the film, and the substrate. This case is discussed in Sect. 10.3. Formulae for the reflectance and transmittance of an absorbing layer on a non-absorbing substrate are given in Sect. 10.4.

Thin absorbing films between two unlike media ($V_1 \neq V_2$) can either decrease or increase the reflectivity, depending on whether the particles go up or down in potential. This follows from the general expression for the reflection amplitude to first order in the film thickness, which, from (15.78), is

$$r = r_0 - \frac{2iq_1}{(q_1 + q_2)^2} \frac{2m\mu_1}{\hbar^2} + \ldots \tag{15.126}$$

The corresponding reflectivity is

$$R = \left(\frac{q_1 - q_2}{q_1 + q_2}\right)^2 - \frac{4q_1(q_1 - q_2)}{(q_1 + q_2)^3} \frac{2m}{\hbar^2} \int_{-\infty}^{\infty} \mathrm{d}z \; V_i(z) + \ldots \tag{15.127}$$

Since V_i is non-negative, we see that an absorbing film will (to first order in the film thickness) decrease the reflectance if $V_1 < V_2$, and increase the reflectance if

$V_1 > V_2$. The transmittance is always decreased by absorption, on the other hand. From (15.16) with $\tilde{V} = V_0$ and $\tilde{\psi} = \psi_0$ we find

$$t = t_0 - \frac{2iq_1}{(q_1+q_2)^2}\frac{2m\mu_1}{\hbar^2} + \ldots, \tag{15.128}$$

where $t_0 = 2q_1/(q_1+q_2)$. Thus the transmittance to first order in the film thickness is

$$T = \frac{q_2}{q_2}|t|^2 = \frac{4q_1q_2}{(q_1+q_2)^2}\left\{1 - \frac{2}{q_1+q_2}\frac{2m}{\hbar^2}\int_{-\infty}^{\infty} \mathrm{d}z\ V_i(z) + \ldots\right\}. \tag{15.129}$$

In the presence of absorption the flux conservation law $R+T=1$ does not hold. One can define an absorptance A such that $R+T+A=1$; A is the probability of a particle being absorbed within the film. From (15.127) and (15.129) we find that, to first order in the film thickness,

$$A = \frac{4q_1}{(q_1+q_2)^2}\frac{2m}{\hbar^2}\int_{-\infty}^{\infty} \mathrm{d}z\ V_i(z) + \ldots \tag{15.130}$$

Reflection at a gradual transition between a non-absorbing medium (potential V_1) and an absorbing medium (potential $V_2 = V_r - iV_i$), which would be total when $V_i = 0$ for $\theta_1 > \theta_c$ in the $V_1 < V_r$ case, is less than total in the presence of absorption. The decrease from unity is greater the thicker the transition region, as a result of the greater probability of particle penetration into the absorbing region. The formulae derived for the tanh profile in Sect. 10.7 apply directly to the particle case.

15.8 Inversion of a Model Reflection Amplitude

An exact inversion, due to Gelfand, Levitan and Marchenko, is possible if the reflection amplitude is known for all wavenumbers. (References are given in Chap. 11.) The exact inversion depends on the solution of an integral equation. An approximate inversion, based on the Rayleigh approximation and Fourier analysis, was given in Sect. 11.3. This is adapted here to the particle case. The Rayleigh approximation to the scattering amplitude is, from (15.98),

$$r \approx -\int_{-\infty}^{\infty} \mathrm{d}z \frac{\mathrm{d}q/\mathrm{d}z}{2q} e^{2i\phi}, \quad \phi = \int^{z} \mathrm{d}\zeta\ q(\zeta). \tag{15.131}$$

This may be written as the Fourier transform $r(q_1)$ of the function $-(\mathrm{d}q/\mathrm{d}x)/2q$, where $x = \phi/q_1$:

$$r(q_1) \approx -\int_{-\infty}^{\infty} \mathrm{d}x \frac{\mathrm{d}q/\mathrm{d}x}{2q} e^{2iq_1 x}. \tag{15.132}$$

The Fourier inverse of (15.132) is

$$-\frac{\mathrm{d}q/\mathrm{d}x}{2q} \approx \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}q_1 \mathrm{e}^{-2iq_1 x} r(q_1) \equiv F(2x), \tag{15.133}$$

where the reflection amplitude is analytically continued to negative q_1 via $r(-q_1) = r^*(q_1)$. Thus, on integrating (15.133) from $-\infty$ to x,

$$q(x) \approx q_1 \exp\left[-2 \int_{-\infty}^{2x} \mathrm{d}y\, F(y)\right]. \tag{15.134}$$

The square of (15.134) gives, on using $q^2 = \left(2m/\hbar^2\right)(\mathcal{E} - V) - K^2$,

$$\frac{\mathcal{E} - V(x)}{\mathcal{E} - V_1} \approx \sin^2\theta_1 + \cos^2\theta_1 \exp\left[-4 \int_{-\infty}^{2x} \mathrm{d}y\, F(y)\right]. \tag{15.135}$$

From the definition (15.133) of F, the long wave limit of the reflection amplitude is

$$\int_{-\infty}^{\infty} \mathrm{d}y\, F(y) = r(q_1 \to 0). \tag{15.136}$$

A check on the accuracy of the solution (15.135) is thus provided in the limit as $x \to \infty$. The left side of (15.135) then tends to $(\mathcal{E} - V_2)/(\mathcal{E} - V_1)$; the right side tends to the same value if the long wave limit (15.99) of the Rayleigh approximation, $r \to 1/2 \ln(q_1/q_2)$, is substituted. But if the correct limit $(q_1 - q_2)/(q_1 + q_2)$ is substituted in the $x \to \infty$ limit of (15.135), the two sides differ by a term of order $(V_1 - V_2)^3$.

The approximate inversion formula (15.135), which can be expected to work well if the reflection is weak at all wavenumbers, must be supplemented by a relationship between the variable x and the physical depth z. This is obtained by integrating $\mathrm{d}\phi = q\mathrm{d}z = q_1 \mathrm{d}x$, using (15.134):

$$z(x) = \int_0^x \mathrm{d}x_1 \frac{q_1}{q(x_1)} \approx \int_0^x \mathrm{d}x_1 \exp\left(2 \int_{-\infty}^{2x_1} \mathrm{d}x_2 F(x_2)\right) \tag{15.137}$$

The inverse relation is

$$x(z) = \phi(z)/q_1 = q_1^{-1} \int_0^z \mathrm{d}\zeta \; q(\zeta)$$

.

15.9 Time Delay in the Reflection of Wavepackets

Sections 19.2 and 19.3 deal in detail with zero, partial or total reflection of quantum particle wavepackets. Here we consider one aspect, namely the time delay in the reflection of a wavepacket made up by superposition of plane wave energy eigenstates, each with time dependence $\mathrm{e}^{-i\mathcal{E}t/\hbar}$, which we write as $\mathrm{e}^{-i\omega t}$. If

$$\psi_i(t) = \int_{-\infty}^{\infty} \mathrm{d}\omega \, f(\omega) \mathrm{e}^{-i\omega t} \tag{15.138}$$

represents the incident wavepacket at some reference plane (say $z = 0$) then, because of the linearity of Schrödinger's equation, the reflected wave at the same plane is made up of a similar superposition, each energy component having reflected with its own reflection amplitude:

$$\psi_r(t) = \int_{-\infty}^{\infty} \mathrm{d}\omega \; r(\omega) f(\omega) \mathrm{e}^{-i\omega t}. \tag{15.139}$$

When the energy distribution of the incident wavepacket is strongly peaked about some value $\mathcal{E}_0$, the wavepacket is nearly sinusoidal, with $\psi_i(t) = A(t)\mathrm{e}^{-i\omega_0 t}$ and an amplitude function $A(t)$ which varies slowly over most of its range. (An example of such a pulse is given in Sect. 19.1.) From the Fourier inverse of (15.138), the energy (or frequency) distribution function $f(\omega)$ is given by

$$f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}t\ A(t) \mathrm{e}^{i(\omega-\omega_0)t}, \tag{15.140}$$

and thus the reflected wavepacket is

$$\psi_r(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\omega\ r(\omega) \mathrm{e}^{-i\omega t} \int_{-\infty}^{\infty} \mathrm{d}\tau\ A(\tau) \mathrm{e}^{i(\omega-\omega_0)\tau}. \tag{15.141}$$

An explicit form for the reflected wavepacket can be obtained on the assumption that $r(\omega) = |r(\omega)|\mathrm{e}^{i\delta(\omega)}$ is well represented (within the dominant range of energies which make up the incident wavepacket) by

$$|r(\omega)| \approx |r(\omega_0)|, \quad \delta(\omega) \approx \delta_0 + (\omega - \omega_0)\delta_0', \tag{15.142}$$

where $\delta_0 = \delta(\omega_0)$ and δ_0' is the derivative $\mathrm{d}\delta/\mathrm{d}\omega = \hbar \mathrm{d}\delta/\mathrm{d}\mathcal{E}$ evaluated at $\mathcal{E}_0$. Then (15.141) gives

$$\psi_r(t) \approx |r(\omega_0)| \mathrm{e}^{i\delta_0} A\big(t - \delta_0'\big) \mathrm{e}^{-i\omega_0 t}. \tag{15.143}$$

The amplitude function of the reflected wavepacket is thus unchanged in shape, but the wavepacket is delayed by

$$\Delta t = \delta_0' = \hbar \left[\frac{\mathrm{d}\delta}{\mathrm{d}\mathcal{E}}\right]_{\mathcal{E}_0}. \tag{15.144}$$

We will give some applications of this time-delay formula. The simplest case is that of partial reflection at a potential step, located at $z = z_1$. At normal incidence the reflection amplitude is

$$r = \mathrm{e}^{2ik_1 z_1} \frac{k_1 - k_2}{k_1 + k_2}, \tag{15.145}$$

and the time-delay formula (15.144) gives

$$\Delta t = 2z_1 \hbar \frac{\mathrm{d}k_1}{\mathrm{d}\mathcal{E}} = \frac{2z_1}{u_1}, \tag{15.146}$$

where $u_1 = \hbar k_1/m$ is the group speed $\mathrm{d}\omega/\mathrm{d}k = \hbar^{-1}\mathrm{d}\mathcal{E}/\mathrm{d}k$ in the first medium. The reflected pulse is delayed by just the time it takes to propagate to the barrier and back, at the group or wavepacket speed.

When $\mathcal{E} < V_2$ there is total reflection. Again considering reflection at a step located at z_1, at normal incidence, we have

$$r = \mathrm{e}^{2ik_1 z_1} \frac{k_1 - i|k_2|}{k_1 + i|k_2|} = \exp 2i\left(k_1 z_1 - \arctan \frac{|k_2|}{k_1}\right), \tag{15.147}$$

where

$$k_1^2 = \frac{2m}{\hbar^2}(\mathcal{E} - V_1), \quad |k_2|^2 = \frac{2m}{\hbar^2}(V_2 - \mathcal{E}). \tag{15.148}$$

The time-delay formula (15.144) now gives

$$\Delta t = \frac{2z_1}{u_1} + \hbar[(\mathcal{E} - V_1)(V_2 - \mathcal{E})]^{-1/2} = \frac{2}{u_1}\left(z_1 + |k_2|^{-1}\right). \tag{15.149}$$

The reflected wavepacket is thus delayed by more than the travel time $2z_1/u_1$ to and from the potential step. The increase can be interpreted in terms of penetration to the depth $|k_2|^{-1}$ into the barrier. (This penetration depth diverges as $\mathcal{E}$ tends to V_2 from below, but (15.144) is not valid as $|k_2| \to 0$ because the square root singularity in δ cannot be approximated by (15.142).)

For total reflection at a gradually rising potential barrier, the phase shift can be approximated by the short wave formula

$$\delta \approx 2\int_0^{z_0} \mathrm{d}z\, k(z, \mathcal{E}) - \pi/2 \tag{15.150}$$

(this is the normal incidence form of (15.113)). The time-delay can then be written in terms of the local value of the wavepacket speed, $u = \hbar^{-1}\mathrm{d}\mathcal{E}/\mathrm{d}k = \hbar k/m$:

$$\Delta t \approx 2\int_0^{z_0} \frac{\mathrm{d}z}{u(z, \mathcal{E})}. \tag{15.151}$$

(The turning point z_0 is also a function of $\mathcal{E}$, but in the differentiation of δ the term $\mathrm{d}z_0/\mathrm{d}\mathcal{E}$ is multiplied by $k(z_0, \mathcal{E})$, which is zero.) The interpretation of (15.151) is that the wavepacket travels up to the classical turning point z_0, where $\mathcal{E} = V(z_0)$, and back, at the group velocity $u(z, \mathcal{E})$. In contrast to (15.149), there is negligible penetration into the classically forbidden region where $\mathcal{E} < V(z)$, because (15.151) was derived using the semiclassical formula (15.150). For a linear variation of V with z in $z_1 \le z \le z_1 + \Delta z$, with V rising between V_1 and $V_2 = V_1 + \Delta V$, (15.151) gives

$$\Delta t \approx \frac{2z_1}{u_1} + 2[2m(\mathcal{E} - V_1)]^{\frac{1}{2}} \frac{\Delta z}{\Delta V}. \tag{15.152}$$

Note that the correction to $2z_1/u_1$ is the same as the classical time delay $2\int_{z_1}^{z_0} \mathrm{d}z/v(z)$, where $\frac{1}{2}mv^2(z) = \mathcal{E} - V(z)$.

The time-delay discussed above is based on the approximation $\delta \approx \delta_0 + (\omega - \omega_0)\delta_0'$. Higher order terms in the Taylor expansion of δ about ω_0 lead to pulse spreading and distortion; references to these are given in Sect. 19.1. (The intrinsic spreading of the wavepacket with time has been neglected here, but is made explicit in Sects. 19.2 and 19.3, where exact solutions of Schrödinger's time-dependent equation are used. This spreading takes place even in a homogeneous medium, but can be made small for highly mono-energetic wavepackets.)

References

Many references to quantum mechanical theory are made in Chaps. 1, 2, 16 and 19. Here we give a few additional references to specific particle reflection and transmission problem, including tunneling.

Barone A, Paternò G (1982/2005) Physics and applications of the Josephson effect. Wiley, New York

Binnig G, Rohrer H (1986) Scanning tunneling microscopy. IBM. J Res Dev (July issue)

Burstein E, Lundquist S (eds) (1969) Tunneling phenomena in solids. Plenum, New York

Edwards DO, Fatouros P, Ihas GG, Mrozinski P, Shen SY, Gasparini FM, Tam CP (1975) Specular reflection of ^{4}He atoms from the surface of liquid ^{4}He. Phys Rev Lett 34:1153–1156

Edwards DO, Fatouros PP (1978) Theory of atomic scattering at the free surface of liquid ^{4}He. Phys Rev B 17:2147–2159

Lu JR, Lee EM, Thomas RK (1996) The analysis and interpretation of neutron and X-ray specular reflection. Acta Cryst A 52:11–41

Solymar L (1972) Superconductive tunnelling and applications. Chapman and Hall, London

Stroscio JA, Kaiser WJ (2013) Scanning tunneling microscopy .Academic Press, Boston

Tersoff J, Hamann DR (1985) Theory of the scanning tunneling microscope. Phys Rev B 31:805–813

Chapter 16
Neutron and X-ray Reflection

All reflection, whether of particle, electromagnetic or acoustic waves, is the result of the constructive interference of many scattered waves originating from scatterers in a planar stratified medium, as we saw in Sect. 1.5. For regular arrays (gratings or lattices), specular reflection can be viewed as a special case of diffraction: it is the zero order diffraction peak, and the *only* one when the wavelength is greater than twice the lattice spacing. When the latter condition holds, an assembly of scatterers can be replaced by a medium characterized by a potential V, or dielectric function ε, or refractive index n. The same can be done for disordered systems, again in an averaged sense, except that incoherent scattering is thereby omitted. For planar-stratified media whose properties depend spatially only on the depth z, reflection properties follow (in principle, at least) from the knowledge of $V(z)$ or $\varepsilon(z)$ or $n(z)$. We know from Sect. 1.3 that there is a one-to-one correspondence between electromagnetic s-wave reflection and particle wave reflection, with an effective dielectric function $\varepsilon = 1 - V/E$, where E is the particle energy. We shall see that because the form of V for neutrons given in Sect. 1.5 and the form of ε for an electron gas given in Sect. 7.6, there is a close correspondence between neutron and X-ray reflection. A major application of reflection experiments is to the study of solid surfaces and liquid-vapour interfaces (Penfold and Thomas 1990; Felcher and Russell 1991; Lu et al. 1996). A review of X-ray and neutron reflection study of polymers at interfaces is given by Russell (1990). Neutron and X-ray reflection has the advantage of much shorter wavelengths compared to reflection at visible wavelengths, and thus the ability to probe down to nanometre scale. A disadvantage is that the reflection is very weak except near glancing incidence.

The topics covered in this chapter all relate to specular reflection. Non-specular reflection, discussed in Chap. 14, occurs for any reflectors without an ideal stratification dependent in its properties only on the depth variable z. X-ray and neutron scattering from rough surfaces has been calculated in the first Born approximation by Sinha et al. (1988); more recent work on this topic includes Felcher et al. (1994), and Sentenac and Daillant (Chap. 2) and Daillant, Mora and Sentenac (Chap. 3) in *X-ray and neutron reflectivity*, edited by Daillant and Gibaud (2009).

A much wider range of neutron optics is covered in the review by Werner and Klein (1986), in the monograph by Sears (1989), and in the compilation *X-ray and neutron reflectivity* noted above.

J. Lekner, *Theory of Reflection*, Springer Series on Atomic, Optical, and Plasma Physics 87, DOI 10.1007/978-3-319-23627-8_16

16.1 Common Features of X-ray and Neutron Optics

The dielectric function for a free-electron plasma is

$$\varepsilon(\omega) = 1 - \omega_p^2/\omega^2, \quad \omega_p^2 = 4\pi n e^2/m, \tag{16.1}$$

where n is the number of electrons per unit volume. The same form holds for X-rays interacting with matter, provided ω is not close to one of the atomic excitation frequencies. Another way of writing (16.1) is as

$$\varepsilon(\lambda) = 1 - \lambda^2/L^2, \quad L^2 = \pi v/r_e, \tag{16.2}$$

where $v = 1/n$ is the volume per electron, and $r_e = e^2/mc^2 \approx 2.818 \times 10^{-15}$ m is the classical electron radius. (Gaussian units were used in the definitions of ω_p and r_e above. In SI units e^2 is to be replaced by $e^2/4\pi\varepsilon_0$ in both definitions, but (16.2) does not change.) For water, $v \approx 2.99$ Å^3, $L \approx 577$ Å and for 1.54 Å wavelength (X-rays of about 8 keV energy), $1 - \varepsilon \approx 7 \times 10^{-6}$. X-ray reflection at normal incidence is tiny: one must go to glancing incidence to get substantial reflection. Since $\varepsilon < 1$, even total reflection is possible, as discussed in the next Section.

Neutron motion is according to Schrödinger's equation

$$-\frac{\hbar^2}{2M}\nabla^2\Psi + V\Psi = E\Psi \tag{16.3}$$

where V is the potential energy and E the total energy. Fermi and others (see Sect. 1.5, Sears 1989, or Lekner 1991 for references) show that the effective potential for neutrons in a medium consisting of particles off which the neutrons scatter coherently with bound scattering length b, and where the volume per scatterer is v, is

$$V = \left(\frac{\hbar^2}{2M}\right)4\pi b/v \equiv \left(\frac{\hbar^2}{2M}\right)4\pi\rho, \quad \rho = b/v \tag{16.4}$$

($\rho = b/v$ is the scattering length density.) The scattering is predominantly due to nuclei: neutron-electron interaction is much weaker in nonmagnetic media. The refractive index for particle waves (see Sect. 1.3) is $(1 - V/E)^{1/2}$. The energy E is equal to $\hbar^2k^2/2M$, where $k = 2\pi/\lambda$, λ being the free-space neutron wavelength. Thus the effective-medium dielectric function for neutrons, the square of the refractive index, can be written as

$$\varepsilon = 1 - b\lambda^2/\pi v = 1 - \lambda^2\rho/\pi, \tag{16.5}$$

and is thus of the same form as (16.2):

$$\varepsilon(\lambda) = 1 - \lambda^2/L^2, \qquad L^2 = \pi v/b = \pi/\rho. \tag{16.6}$$

The neutron wavelength is $\lambda \approx 9.04\,\text{Å}/(E/\text{meV})^{1/2}$, so an 81 meV neutron has a wavelength of about 1 Å. For un-magnetized Fe, $b \approx 9.5\,fm$ and $v \approx 11.8\,\text{Å}^3$ (Werner and Klein 1986, Table II), so $L \approx 625\,\text{Å}$, and $1 - \varepsilon \approx 2.6 \times 10^{-6}$. The effective potential for neutrons in iron is of the order of 0.2 μeV, which is about the same as the work needed to raise a neutron by two metres in the earth's gravitational field.

We have seen that not only can the X-ray and neutron interactions with matter be characterized by the same form of wavelength-dependent dielectric function, but that in both cases, for typical wavelengths, this differs from unity by parts per million. Hence only reflection not too far from glancing incidence is important (in Sect. 2.3 we saw that reflection becomes perfect as the glancing angle tends to zero, for all reflecting profiles).

Further, it is known (Sect. 1.3) that reflection of particle waves by stratified media is mathematically the same as reflection of electromagnetic *s*-waves. Thus we can treat X-ray *s*-wave and scalar-interaction neutron reflection together. X-ray reflection of *p*-waves is nearly the same as that of *s*-waves in the regions of interest, as we shall see shortly.

16.2 Reflection Near the Critical Angle

The dielectric function $\varepsilon(\lambda) = 1 - \lambda^2/L^2$ is less than unity for X-rays, and also for neutron targets with net positive scattering length. Thus total reflection of X-rays or neutrons incident from vacuum onto such materials is possible. The critical angle is close to grazing incidence, and in X-ray and neutron reflection it is usual to work in terms of glancing angle, this being given the same symbol θ used for the angle to the surface normal in optics. Snell's law for refraction at a boundary media 1 and 2, in terms of *glancing* angles θ_1 and θ_2, reads

$$\varepsilon_1 \cos^2\theta_1 = \varepsilon_2 \cos^2\theta_2. \tag{16.7}$$

When $\varepsilon_1 = 1$ and $\varepsilon_2 = 1 - \lambda^2/L_2^2$, with L_2 being the interaction length associated with the substrate, total reflection will occur for $q_2^2 = (2\pi/\lambda)^2(\varepsilon_2 - \cos^2\theta) = (2\pi/\lambda)^2(\sin^2\theta - \lambda^2/L_2^2) \leq 0$, that is when

$$\sin\theta_1 \leq \sin\theta_c = \lambda/L_2. \tag{16.8}$$

From now on we make no use of θ_2, and can drop the subscript 1 on θ_1, so θ will always mean θ_1, the glancing angle of incidence. Waves with $\lambda/\sin\theta > L_2$ will be

totally reflected. For the examples used above, X-rays of wavelength 1.54 Å will totally reflect from water for $\theta < \theta_c \approx 0.153^\circ$, and neutrons of wavelength 1 Å will totally reflect from unmagnetized iron for $\theta < \theta_c \approx 0.092^\circ$.

Any smooth, non-absorbing planar-stratified medium will reflect totally below the critical glancing angle $\theta_c = \arcsin(\lambda/L_2)$, where L_2 means the final value taken by $L(z)$ as the depth z increases into the medium. As in the optical case, we take the zx plane to be the plane of incidence. Then the incident, reflected and transmitted plane waves have $\exp i(Kx - \omega t)$ for the x and t dependence, with $K = k_x = \frac{2\pi}{\lambda}\cos\theta$ (incidence from vacuum is assumed in this chapter).

For the electromagnetic s-wave, and for neutrons, the remaining factor of the probability amplitude is $\psi(z)$, which satisfies

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}z^2} + q^2\psi = 0, \quad q^2(z) = \left(\frac{2\pi}{\lambda}\right)^2 \left[\varepsilon(z) - \cos^2\theta\right]. \tag{16.9}$$

The square of the normal component of the wavevector can also be written as

$$q^2(z) = q_1^2 - [2\pi/L(z)]^2 = \left(\frac{2\pi}{L_2}\right)^2 \left[\left(\frac{\sin\theta}{\sin\theta_c}\right)^2 - \left(\frac{L_2}{L(z)}\right)^2\right]. \tag{16.10}$$

where $q_1 = (2\pi/\lambda)\sin\theta$ is the normal component of the wavevector in the incident wave.

The form of (16.10) shows that, for a fixed profile $L(z)$ on the same substrate, the s-wave or neutron reflection properties will be the same for the same value of $\sin\theta/\sin\theta_c$. Thus

$$S = \sin\theta/\sin\theta_c = q_1/q_c \quad [q_c = (2\pi/\lambda)\sin\theta_c = 2\pi/L_2]$$

is the appropriate variable in plotting the reflectance, since then data taken at different wavelengths and different angles of incidence can be plotted on one curve characteristic of a given profile (Lekner 1991).

For *negative neutron scattering length* (titanium for example) there is no real critical angle, and the above characterization in terms of the angle variable $S = \sin\theta/\sin\theta_c$ has to be modified. For example, we can define $S = \lambda^{-1}\sqrt{\pi v/|b|}\sin\theta$, and then the reflection amplitude for neutrons by a step profile can be written as

$$\begin{aligned} r_0 &= (q_1 - q_2)/(q_1 + q_2) = \left(S - \sqrt{S^2+1}\right)/\left(S + \sqrt{S^2+1}\right) \\ &= -\left(S + \sqrt{S^2+1}\right)^{-2}. \end{aligned}$$

For magnetized cobalt, one neutron polarization has a positive total scattering length, while the other polarization has negative total scattering length. This case is discussed in Sect. 16.7, with the reflectivities shown in Fig. 16.11.

For the electromagnetic p wave, with $\mathbf{B} = (0, B_y, 0)$, we have $B_y(z, x, t) = B(z) \exp i\,(Kx - \omega t)$ as in (2.21), with

$$\varepsilon \frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{\varepsilon}\frac{\mathrm{d}B}{\mathrm{d}z}\right) + q^2(z)B = 0, \tag{16.11}$$

where $q^2(z)$ is given by (16.9) or (16.10) as before. The characteristics of the reflecting profile thus appear in the logarithmic derivative $\varepsilon^{-1}\mathrm{d}\varepsilon/\mathrm{d}z$ as well as in $q^2(z)$, and reflectivities will not fall on one curve when plotted against $S = \sin\theta/\sin\theta_c$. However, we shall see that the difference between the s and p reflectivities is of order $\sin^2\theta_c$ compared to unity, and can be ignored for short-wavelength X-rays.

For reflection at a step (an interface which is sharp on the scale of the wavelength) the reflection amplitudes of the s-wave and of neutrons are given by (1.13), namely $r_{s0} = (q_1 - q_2)/(q_1 + q_2)$. For $\theta \geq \theta_c$ this gives

$$r_{s0} = \frac{\sin\theta - \left(\sin^2\theta - \sin^2\theta_c\right)^{\frac{1}{2}}}{\sin\theta + \left(\sin^2\theta - \sin^2\theta_c\right)^{\frac{1}{2}}} = \frac{S - (S^2-1)^{\frac{1}{2}}}{S + (S^2-1)^{\frac{1}{2}}} = \frac{1}{\left[S + (S^2-1)^{\frac{1}{2}}\right]^2}. \tag{16.12}$$

The reflectivity $R_{s0} = |r_{s0}|^2$ is unity for $\theta \leq \theta_c$, and falls rapidly as θ increases beyond θ_c: at $\sin\theta = 2\sin\theta_c$ the reflectivity is $\left(2-\sqrt{3}\right)^4 \approx 5 \times 10^{-3}$, and for $\sin\theta \gg \sin\theta_c$, $R_{s0} \to (\sin\theta_c/2\sin\theta)^4 = (q_c/2q_1)^4$.

The p-wave reflection amplitude is, from (1.31),

$$r_{p0} = -\frac{\sin\theta\cos^2\theta_c - \left(\sin^2\theta - \sin^2\theta_c\right)^{\frac{1}{2}}}{\sin\theta\cos^2\theta_c - \left(\sin^2\theta - \sin^2\theta_c\right)^{\frac{1}{2}}}. \tag{16.13}$$

Note that at the critical angle $r_{s0} = 1$ and $r_{p0} = -1$. (These are the reverse of the limiting values at grazing incidence, $\theta \to 0$). As discussed following (1.25), the reflection amplitude for E_z is $-r_p$, so (for small θ_c) the p-wave dominant electric field component E_z has the incident and reflected waves in-phase at the critical angle. For the s-wave there is only one component (E_y), and the incident and reflected electric fields are in-phase at θ_c. (For the variation of the phases of r_{s0} and r_{p0} when total reflection exists, see the Appendix of Chap. 20).

The difference between R_{s0} and R_{p0} is of order $\sin^2\theta_c$, and is negligible for most X-ray reflection studies. From (16.12) and (16.13) we find

$$r_{p0} = \frac{\sin^2\theta_c \sin^2\theta_B - r_{s0}}{1 - r_{s0}\sin^2\theta_c \sin^2\theta_B} \tag{16.14}$$

where θ_B is the Brewster angle at which the p reflectivity goes to zero. From (16.13), it given by

$$\sin^2\theta_B = 1/(2-\sin^2\theta_c) \quad \text{or} \quad \tan\theta_B = \sec\theta_c. \tag{16.15}$$

θ_B is close to 45° for X-rays with wavelength small compared to $L = (\pi v/r_e)^{1/2}$. Note also that when angles are measured from the surface normal, the last relation of (16.15) reads $\tan\theta_B = \sin\theta_c$.

Figure 16.1 shows the reflectivity R_{s0} for neutrons and X-rays, at the sharp boundary of any medium with positive neutron scattering length b, and R_{p0} for X-rays of 1.54 Å and 15.4 Å wavelength incident on water $(L \approx 577\ \text{Å})$, for which $\sin\theta_c \approx 0.00267$ and 0.0267, respectively. For the smaller X-ray wavelength the s and p reflectivities are not distinguishable on this scale. For the larger wavelength the difference is apparent for $\sin\theta \geq 5\sin\theta_c$ (or $R \leq 10^{-4}$).

The reflectivity is seen to fall off rapidly as θ increases from θ_c; (16.12) and (16.13) show that the behaviour near θ_c is dominated by the square root of $\theta - \theta_c$ (or equivalently, the square root of $\lambda_c - \lambda$ where $\lambda_c = L_2\sin\theta$). Lekner (1991) shows that this square root singularity is universal for non-absorbing smooth profiles. Absorption or surface roughness will round off the reflectivity drop at the critical angle.

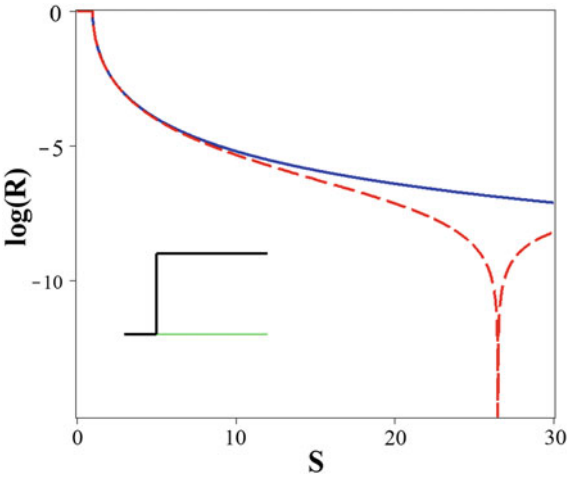


Fig. 16.1 Logarithmic plot (base 10) of the Fresnel reflectivities R_{s0} for neutrons with positive scattering length and for X-rays, and R_{p0} for X-rays reflecting off water. A precisely defined boundary surface is assumed (*inset*). The angle variable is $S = \sin\theta/\sin\theta_c = q_1/q_c$, so the R_{s0} curve is universal for s-polarized X-rays and for neutrons with a real critical angle (positive scattering length). The *solid curve* is R_{s0}, for all neutron or X-ray wavelengths. For $\lambda = 1.54$Å, R_{p0} is indistinguishable from R_{s0} on this graph. The *dashed curve* is R_{p0} for $\lambda = 15.4$Å; this goes through zero at the Brewster angle, at $\sin\theta_B \approx 26.5\sin\theta_c$ (the imaginary part of the X-ray refractive index, about a 1000 times smaller than the difference between the real part and unity, has been set to zero. A nonzero imaginary part will make the reflectivity at θ_B nonzero)

16.3 Reflection by Profiles Without Discontinuities

The following identity is derived in Sect. 5.6:

$$r_s = -\frac{1}{2}\int_{-\infty}^{\infty} \mathrm{d}z\, q^{-1}\frac{\mathrm{d}q}{\mathrm{d}z}[e^{2i\phi} - r^2(z)\, e^{-2i\phi}]. \tag{16.16}$$

Here $q(z)$ is (as always) the normal component of the wavevector in a z-stratified medium, $r(z)$ is the reflection amplitude for a profile truncated at z (see Fig. 5.1 for definition of truncation), and ϕ is the accumulated phase at z:

$$\phi(z) = \int^{z} \mathrm{d}\zeta\phi(\zeta). \tag{16.17}$$

The Rayleigh (1912) approximation for r_s is obtained by dropping the r^2 term in (16.16):

$$r_R = -\frac{1}{2}\int_{-\infty}^{\infty} \mathrm{d}z\, q^{-1}\frac{\mathrm{d}q}{\mathrm{d}z}e^{2i\phi}. \tag{16.18}$$

It can also be called the weak reflection approximation (Sect. 5.7), and works extremely well for smooth profiles which reflect weakly, as seen for example in Figs. 5.4 and 6.3. Lekner (1991) discusses various further approximations that can be obtained from (16.18). Here we give one that is often used in X-ray reflection (see for example Pershan 1990): we rewrite (16.18) as

$$r_R = -\int \mathrm{d}z\frac{\mathrm{d}q^2/\mathrm{d}z}{4q^2}e^{2i\phi}, \tag{16.19}$$

and replace $4q^2$ in the integrand by $(q_1 + q_2)^2$, where $q_1 = (2\pi/\lambda)\sin\theta$, and $q_2^2 = q_1^2 - (2\pi/L_2)^2$ from (16.10). In terms of the variable $S = \sin\theta/\sin\theta_c$, we can write $q_1 = q_c S, q_2 = q_c[S^2 - 1]^{\frac{1}{2}}$. Also, we can define a dimensionless profile shape function $f(z)$ by

$$q^2(z) = \frac{1}{2}(q_1^2 + q_2^2) - \frac{1}{2}(q_1^2 - q_2^2)f(z). \tag{16.20}$$

($f(z)$ tends to -1 in the medium of incidence, normally the vacuum, and to +1 deep in the reflecting substance.) Then (16.19) becomes

$$r_R \approx \frac{q_1 - q_2}{q_1 + q_2} \int \mathrm{d}z \, \frac{1}{2} \frac{\mathrm{d}f}{\mathrm{d}z} e^{2i\phi}. \tag{16.21}$$

Thus the reflection amplitude is written as a product of the Fresnel reflection amplitude for a step profile, $r_{s0} = (q_1 - q_2)/(q_1 + q_2)$, and the Fourier transform of the derivative of the profile shape function in the ϕ variable. (Note that $\int \mathrm{d}f \, e^{2i\phi} = \int \mathrm{d}\phi \frac{\mathrm{d}f}{\mathrm{d}\phi} e^{2i\phi}$.) Since $q^2 = \varepsilon\omega^2/c^2 - K^2$, (16.20) is the same as

$$\varepsilon(z) = \frac{1}{2}(\varepsilon_1 + \varepsilon_2) - \frac{1}{2}(\varepsilon_1 - \varepsilon_2) f(z), \tag{16.22}$$

and therefore

$$f(z) = \frac{\varepsilon_1 + \varepsilon_2 - 2\varepsilon(z)}{\varepsilon_1 - \varepsilon_2}. \tag{16.23}$$

For X-rays $\varepsilon(z) = 1 - (\lambda^2 r_e/\pi) n(z)$ where $n(z)$ is the local electron density, while for neutrons the effective dielectric function is $\varepsilon(z) = 1 - (\lambda^2/\pi)\rho(z)$, where $\rho(z) = b(z)/v(z)$ is the scattering length density, with both $n(z)$ and $\rho(z)$ being zero in the vacuum. Thus

$$f(z) = 2n(z)/n_2 - 1 \quad \text{or} \quad f(z) = 2\rho(z)/\rho_2 - 1 \tag{16.24}$$

in the X-ray and neutron cases. We will use the latter form to avoid possible confusion between the electron density and the refractive index. Equation (16.21) now reads

$$r_R \approx \frac{q_1 - q_2}{q_1 + q_2} \frac{1}{\rho_2} \int \mathrm{d}z \, \frac{\mathrm{d}\rho(z)}{\mathrm{d}z} e^{2i\phi}. \tag{16.25}$$

One further simplifying assumption is to replace 2ϕ by $2q_1 z$, which is a good approximation for X-rays and neutrons where q_1 and q_2 are nearly equal. Thus finally the reflection amplitude is approximated as the product of the Fresnel amplitude and a Fourier transform of the derivative of the scattering length (or electron) density,

$$r_R \approx \frac{q_1 - q_2}{q_1 + q_2} \frac{1}{\rho_2} \int \mathrm{d}z \, \frac{\mathrm{d}\rho}{\mathrm{d}z} e^{2iq_1 z}. \tag{16.26}$$

For a single sharp transition between vacuum and the reflecting material, $\mathrm{d}\rho/\mathrm{d}z \to \rho_2 \delta(z - z_1)$, where z_1 gives the reflecting boundary location. Thus (16.26) correctly gives the Fresnel reflection amplitude in this limit.

As a test of (16.26), consider the reflection by a hyperbolic tangent profile, $f(z) = \tanh(z/2a)$ of Sect. 2.5. The reflection amplitude is, from (2.88),

$$r_s = e^{i\Phi}\frac{\sinh \pi a(q_1 - q_2)}{\sinh \pi a(q_1 + q_2)} \tag{16.27}$$

where Φ is a known phase. From (16.24),

$$\rho(z) = \frac{1}{2}\rho_2\left(1 + \tanh\frac{z}{2a}\right) \tag{16.28}$$

and so

$$\frac{1}{\rho_2}\int_{-\infty}^{\infty} \mathrm{d}z \frac{d\rho}{\mathrm{d}z} e^{2iq_1 z} = \frac{1}{4a}\int_{-\infty}^{\infty} \mathrm{d}z\ \mathrm{sech}^2\left(\frac{z}{2a}\right) e^{2iq_1 z} = \frac{2\pi q_1 a}{\sinh 2\pi q_1 a}. \tag{16.29}$$

Thus for the hyperbolic tangent profile, from (16.27) and (16.26, 16.29),

$$R^{\text{exact}} = \left|\frac{\sinh \pi a(q_1 - q_2)}{\sinh \pi a(q_1 + q_2)}\right|^2, \quad R^{\text{approx}} = \left|\frac{q_1 - q_2}{q_1 + q_2}\right|^2 \left(\frac{2\pi q_1 a}{\sinh 2\pi q_1 a}\right)^2. \tag{16.30}$$

These reflectivities are compared in Fig. 16.2, which shows R^{exact} and R^{approx} for X-rays or neutrons reflecting from an interface with a hyperbolic tangent profile, with two thicknesses of the interface. Note that for this gradual transition between vacuum and bulk matter the approximation (16.30) works quite well, and has the virtue of also

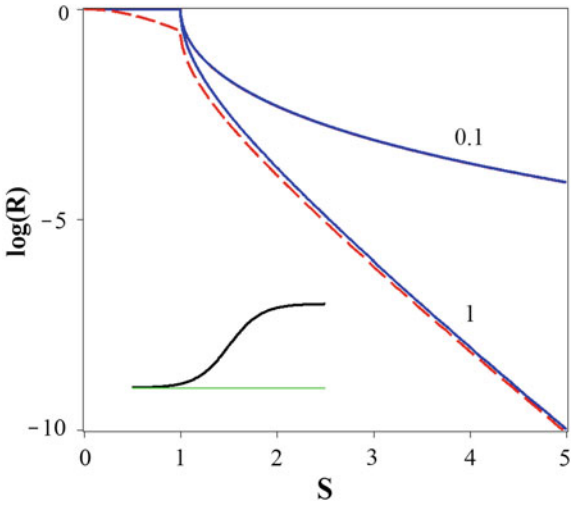


Fig. 16.2 Exact and approximate reflectivities for the tanh profile (*inset*), as a function of the angle-wavelength variable $S = \sin\theta/\sin\theta_c$. The values of the dimensionless parameter $\pi a q_c$ are 0.1 (*upper curve*) and 1 (*lower curves*). The approximate $\log_{10} R$ curve (*dashed*) is not distinguishable from the exact reflectivity for the smaller value $\pi a q_c = 0.1$. The reflectivities apply equally to *s*-polarized X-rays and to neutrons

being exact in the zero-thickness limit. At the critical angle, where $q_1 = q_c, q_2 = 0$, $R^{\text{exact}} = 1$ and $R^{\text{approx}} = (2\pi q_c a/\sinh 2\pi q_c a)^2$, with $q_c = (2\pi/\lambda)\sin\theta_c = 2\pi/L_2$. Thus when the dimensionless parameter $\pi a q_c = 2\pi^2 a/L_2$ is small compared to unity, R^{approx} is accurate near the critical angle. In Fig. 16.2, we have chosen the values 0.1 and 1 for $\pi a q_c$, so for X-rays reflecting from water the corresponding interface thickness parameter has the values $a = L_2/2\pi^2 \approx 2.9$ Å and 29 Å, respectively. The 10–90 thickness (the distance over which the change in the profile passes through 10 and 90 % of the total) of a *tanh* profile is $(2\ln 9)a \approx 4.39a$; the 10–90 thickness corresponding to $a \approx 2.9$ Å is about 12.8 Å.

16.4 Reflection by Profiles with Discontinuities

We have seen that a version of the weak reflection approximation, (16.26), works well for a gradual transition between two media. Consider now a homogeneous layer on a substrate, with

$$\rho(z) = \begin{cases} 0 & z<0 \\ \rho & 0<z<\Delta z \\ \rho_2 & z>\Delta z \end{cases} \tag{16.31}$$

The exact reflection amplitude was given in (2.58), which gives the reflectivity

$$R^{\text{exact}} = \frac{r_1^2 + 2r_1 r_2 \cos 2q\Delta z + r_2^2}{1 + 2r_1 r_2 \cos 2q\Delta z + (r_1 r_2)^2}, \quad r_1 = \frac{q_1 - q}{q_1 + q}, \quad r_2 = \frac{q - q_2}{q + q_2}. \tag{16.32}$$

From the identity $(r_1 + r_2)/(1 + r_1 r_2) = (q_1 - q_2)/(q_1 + q_2)$ it follows that R^{exact} coincides with the Fresnel reflectivity whenever $\cos 2q_1\Delta z = 1$. In terms of the variable $S = \sin\theta/\sin\theta_c$, $q_1 = q_c S, q_2 = q_c[S^2 - 1]^{\frac{1}{2}}$ and $q = q_c[S^2 - \rho/\rho_2]^{\frac{1}{2}}$.

In evaluating the version (16.26) of the weak reflection approximation we assume that the outer boundary of the reflecting surface is at $z = 0$, as specified in (16.31). The two steps in the scattering length density give delta functions in $\mathrm{d}\rho/\mathrm{d}z$ at $z = 0$ and at $z = \Delta z$, of strengths ρ and $\rho_2 - \rho$. Thus (16.26) gives, for real q_2 $(\theta > \theta_c)$,

$$r^{\text{approx}} = \left(\frac{q_1 - q_2}{q_1 + q_2}\right)\left\{\frac{\rho}{\rho_2} + \left(1 - \frac{\rho}{\rho_2}\right)e^{2iq_1\Delta z}\right\}, \tag{16.33}$$

$$R^{\text{approx}} = \left(\frac{q_1 - q_2}{q_1 + q_2}\right)^2\left\{1 - 2\frac{\rho}{\rho_2}\left(1 - \frac{\rho}{\rho_2}\right)(1 - \cos 2q_1\Delta z)\right\}. \tag{16.34}$$

Both R^{exact} and R^{approx} are bounded above by the Fresnel reflectivity, in accord with a general theorem for monotonic profiles (Sect. 5.4). Figure 16.3 compares R^{approx}

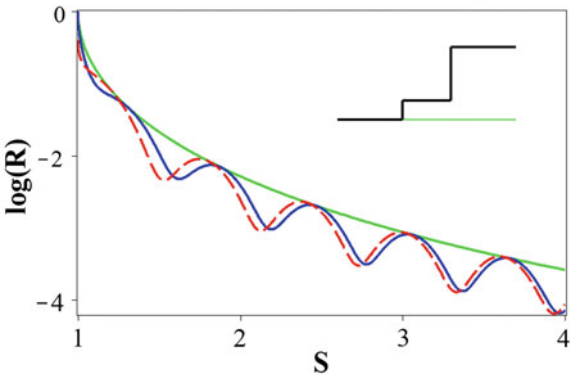


Fig. 16.3 Neutron reflection by a layer of H_2O $\left(\rho = 0.056 \times 10^{-5}\text{Å}^{-2}\right)$ on Si $\left(\rho = 0.215 \times 10^{-5}\text{Å}^{-2}\right)$, as a function of the angle-wavelength variable $S = \sin\theta/\sin\theta_c$. The smooth curve is for bare silicon. The *solid curve* is R^{exact} (16.32), the *dashed curve* is R^{approx} (16.34). The water layer is 1000Å thick. The *inset* shows the scattering length density profile

and R^{exact}, this time for neutrons reflecting from silicon $\left(L = 1208\ \text{Å}\right)$ covered with a layer of water $\left(L = 2363\ \text{Å}\right)$, 1000 Å thick.

We see from Fig. 16.3 that the reflectivity given by the version (16.26) of the weak reflection approximation is in qualitative agreement with R^{exact}, except near the critical angle. The local minima and maxima are off, because q_1 has replaced q in the phase factor.

We can do better: Lekner (1991) has developed a sequence of approximations which have the features that they give *exact results for homogeneous layers*, and are correct at and near the critical angle. The simplest of these is

$$r_0 = e^{2iq_1a}\frac{(q_1q_b - q_2q_a)c + i(q_aq_b - q_1q_2)s}{(q_1q_b + q_2q_a)c - i(q_aq_b + q_1q_2)s} \equiv e^{2iq_1a}\frac{N_0}{D_0}. \tag{16.35}$$

In (16.35) q_a and q_b are the values of q just inside the outer and inner boundaries of a stratification, and $c = \cos\Delta\phi$ and $s = \sin\Delta\phi$, where $\Delta\phi$ is the total phase increment across the stratification:

$$\Delta\phi = \phi(b) - \phi(a) = \int_a^b \mathrm{d}z\, q(z). \tag{16.36}$$

Extrema of the reflectivity, obtained as the absolute square of (16.35), occur when $\Delta\phi$ is a multiple of $\pi/2$. (For a homogeneous layer they occur when $q\Delta z$ is a multiple of $\pi/2$.)

When $q_a = q_b, r_0$ takes the same form as the exact reflection amplitude for a homohogenous layer, within which the normal component of the wavevector is q (a constant);

$$r = e^{2iq_1 a} \frac{q(q_1 - q_2)c + i(q^2 - q_1 q_2)s}{q(q_1 + q_2)c - i(q^2 + q_1 q_2)s}. \tag{16.37}$$

In (16.37), $\Delta\phi = q\Delta z = q(b - a)$, and $c = \cos\Delta\phi, s = \sin\Delta\phi$ as before.

The zeroth approximation (16.35) is based on the simplest wavefunctions,

$$\psi_0^{\pm} = e^{\pm i\phi}, \quad \phi(z) = \int^z \mathrm{d}\zeta q(\zeta). \tag{16.38}$$

The next approximation is based on the Liouville-Green wavefunctions of Sect. 6.2,

$$\psi_1^+(z) = \left(\frac{q_a}{q(z)}\right)^{1/2} \mathrm{e}^{i\phi(z)}, \quad \psi_1^-(z) = \left(\frac{q_b}{q(z)}\right)^{1/2} \mathrm{e}^{-i\phi(z)}. \tag{16.39}$$

The Liouville-Green wavefunctions take into account the variation in normal component of wavenumber q through the dimensionless function γ introduced in Chap. 6:

$$\gamma = \frac{\mathrm{d}q/\mathrm{d}z}{q^2} = -\frac{2\pi}{q^3}\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{b}{v}\right) = \frac{4\pi^2}{(qL)^3}\frac{\mathrm{d}L}{\mathrm{d}z}. \tag{16.40}$$

Let γ_a, γ_b be the values at $z = a+, z = b-$. Then the reflection amplitude takes the form

$$r_1 = e^{2iq_1 a} \frac{N_0 + N_1}{D_0 + D_1}, \tag{16.41}$$

where N_0, D_0 were defined in (16.35), and, from Lekner (1991) or from Sect. 6.3,

$$N_1 = -\frac{1}{2}(q_1 q_b \gamma_b + q_2 q_a \gamma_a)s + \frac{i}{2} q_a q_b [(\gamma_b - \gamma_a)c + \frac{1}{2}\gamma_a \gamma_b s], \tag{16.42}$$

$$D_1 = -\frac{1}{2}(q_1 q_b \gamma_b - q_2 q_a \gamma_a)s - \frac{i}{2} q_a q_b [(\gamma_b - \gamma_a)c + \frac{1}{2}\gamma_a \gamma_b s].$$

Both r_0 and r_1 give the correct result for a homogeneous layer, and for an arbitrary layer correctly give unit reflectivity at the critical angle or critical wavelength when $q_2 \to 0$, and at glancing incidence when $q_1 \to 0$.

According to the formulae for r_0 and r_1, reflection is mainly the result of discontinuities at $z = a$ and b, and of interference between the reflections from these discontinuities. Discontinuities in slope also contribute to r_1, while a gradual variation of the medium enters the formulae only through the phase increment $\Delta\phi$.

The reflectivities R_0 and R_1, obtained by squaring the modulus of (16.35) and (16.41), are compared with the exact reflectivity for a profile with *linear variation* in $\rho = b/v$ in Figs. 16.4 and 16.5. Such profiles have $q^2 = q_1^2 - 4\pi\rho$ linear in z, and the Airy functions of Sect. 5.2 are solutions of (16.9). Specifically, the solutions are $Ai(-\zeta)$, $Bi(-\zeta)$, where

$$\zeta(z) = \left[\frac{1}{4\pi}\left|\frac{\Delta z}{\Delta\rho}\right|\right]^{\frac{2}{3}} q^2(z) = \left|\frac{\Delta z}{\Delta q^2}\right|^{\frac{2}{3}} q^2(z) = \left[\frac{1}{4\pi}\left|\frac{\Delta z}{\Delta\rho}\right|\right]^{\frac{2}{3}} \left\{q_a^2 - (z-a)4\pi\frac{\Delta\rho}{\Delta z}\right\}. \tag{16.43}$$

In (16.43), $\Delta\rho = \rho_b - \rho_a$ is the change in the scattering length density over the extent $\Delta z = b - a$ of the profile. The exact reflectivity is calculated by substituting $F(z) = Ai(-\zeta)$, $G(z) = Bi(-\zeta)$ into $\psi = \alpha F + \beta G$, and matching both ψ and $d\psi/dz$ at the boundaries. At $z = a$ we match to $e^{iq_1 z} + re^{-iq_1 z}$, at $z = b$ to $te^{iq_2 z}$. The resulting equations are (with primes denoting derivatives with respect to z)

$$e^{iq_1 a} + re^{-iq_1 a} = \alpha F_a + \beta G_a, \quad iq_1(e^{iq_1 a} - re^{-iq_1 a}) = \alpha F_a^{'} + \beta G_a^{'}, \tag{16.44}$$

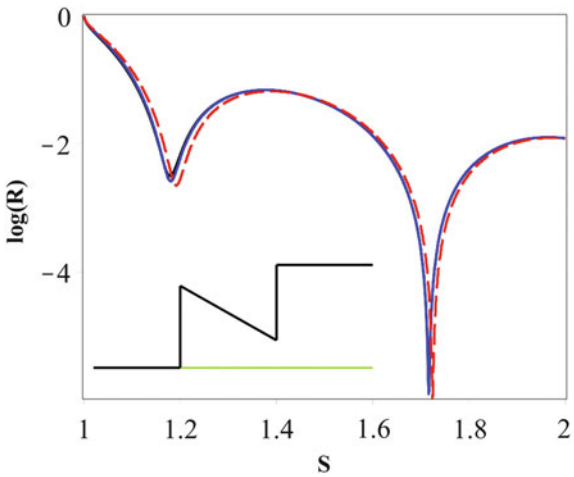


Fig. 16.4 Comparison of the exact and approximate reflectivities R_e, R_0 and R_1 for a profile with discontinuities in the scattering length density ρ at its boundaries, and a linear variation in $\rho(z)$ in between. The curves for R_e and R_1 are barely distinguishable; R_0 is the *dashed curve*. The angle-wavelength variable is $S = \sin\theta/\sin\theta_c = q_1/q_c$. The scattering length densities are $\rho_a = 0.641$, $\rho_b = 0.215$, $\rho_2 = 0.805$ (units of 10^{-5} Å^{-2}), corresponding to D_2O, Si and Fe respectively. The layer thickness is 500 Å

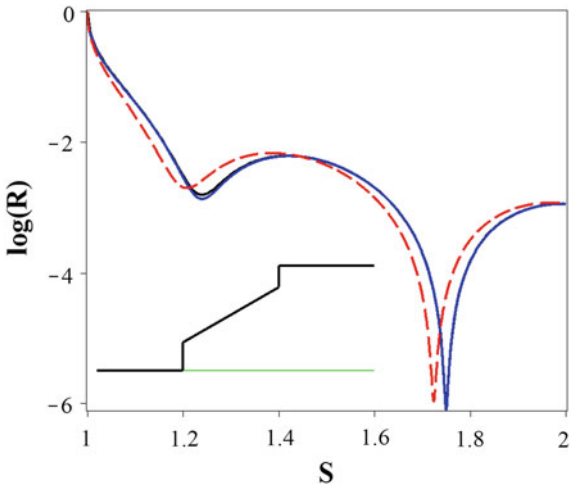


Fig. 16.5 As for Fig. 16.4, with the scattering length densities ρ_a and ρ_b interchanged: $\rho_a = 0.215,\ \rho_b = 0.641,\ \rho_2 = 0.805$ (units of 10^{-5} Å^{-2}); the layer thickness remains 500 Å. Note that R_0 (*dashed curve*) is not as good as in Fig. 16.4, because less of the reflection is now due to the smaller discontinuities in ρ at the boundaries

$$\alpha F_b + \beta G_b = te^{iq_2 b}, \quad \alpha F'_b + \beta G'_b = iq_2 te^{iq_2 b}. \tag{16.45}$$

The symbols have the meaning $F_a = F(a+)$, $F'_a = (\mathrm{d}F/\mathrm{d}z)_{z=a+}, \ldots$, $G_b = G(b-)$, $G'_b = (\mathrm{d}G/\mathrm{d}z)_{z=b-}$. There are four unknowns in the four equations (16.44), (16.45): r, t, α, β. The solution generalizes that given in Sect. 2.2 to profiles with discontinuities at the boundaries. The reflection and transmission amplitudes are

$$r = e^{2iq_1 a}\frac{q_1q_2(F_aG_b - G_aF_b) + iq_1\left(F_aG'_b - G_aF'_b\right) + iq_2\left(F'_aG_b - G'_aF_b\right) - \left(F'_aG'_b - G'_aF'_b\right)}{q_1q_2(F_aG_b - G_aF_b) + iq_1\left(F_aG'_b - G_aF'_b\right) - iq_2\left(F'_aG_b - G'_aF_b\right) + \left(F'_aG'_b - G'_aF'_b\right)} \tag{16.46}$$

$$t = e^{iq_1 a - iq_2 b}\frac{2iq_1\left(F_bG'_b - G_bF'_b\right)}{q_1q_2(F_aG_b - G_aF_b) + iq_1\left(F_aG'_b - G_aF'_b\right) - iq_2\left(F'_aG_b - G'_aF_b\right) + \left(F'_aG'_b - G'_aF'_b\right)}$$

These amplitudes have the same form as those given in (2.25) and (2.26) of Chap. 2, but allow for a jump in the effective potential or dielectric function at $z = a$ and $z = b$.

To calculate the approximate reflectivities R_0 and R_1 we need $\Delta\phi$, the phase increment over the inhomogeneity between $z = a$ and $z = b$. For linear variation in the scattering length density, this is

$$\Delta\phi = \int_a^b \mathrm{d}z\, q(z) = \frac{1}{6\pi}\frac{\Delta z}{\Delta\rho}(q_a^3 - q_b^3). \tag{16.47}$$

We also need the values of γ defined in (16.40) at the boundaries of the profile. These are

$$\gamma_a = -2\pi\frac{\Delta\rho}{\Delta z}q_a^{-3}, \quad \gamma_b = -2\pi\frac{\Delta\rho}{\Delta z}q_b^{-3}. \tag{16.48}$$

We see that R_0 is qualitatively correct, and R_1 is barely distinguishable from the exact reflectivity. The deep minima in the reflectivities are associated with the nearby zero of $\cos\Delta\phi$ at $\sin\theta/\sin\theta_c \approx 1.724$ (which makes the real part of N_0 zero), close to a zero of the imaginary part of N_0 at $\sin\theta/\sin\theta_c \approx 1.832$.

16.5 Total Reflection: Extraction of the Phase in Lloyd's Mirror Experiments

We have so far considered partial reflection, for glancing angle θ greater than the critical angle. When $\theta < \theta_c$, the reflectivity is identically unity, and all the information about the scattering length density profile $\rho(z)$ that can be measured is contained in the phase of the reflected wave. Of course, the absolute phase has no meaning; one has to measure the relative phase of the reflected and incident waves. Lloyd's mirror experiment produces interference fringes between the direct and reflected beams, and Klein and Opat and collaborators have suggested Lloyd's mirror configuration for neutrons (Gudkov et al. 1993), and implemented it for light (Allman et al. 1993a, b). They used the semiclassical short-wave reflected phase expression of Sect. 6.7,

$$\delta \approx 2\int_0^{z_0} \mathrm{d}z\, q(z) - \frac{\pi}{2}, \tag{16.49}$$

It is assumed in the derivation of this expression that the scattering length density profile $\rho(z)$ is restricted to $z \geq 0$. The classical turning point is at z_0 defined by $q(z_0) = 0$. Since $q^2(z) = q_1^2 - 4\pi\rho(z), \rho(z_0) = q_1^2/4\pi$. The approximation (16.49) is known to fail at grazing incidence, where the effective wavelength $2\pi/q$ is large (see Fig. 6.9 and the accompanying discussion). In the limit $\theta \to 0$, (16.49) gives $r = e^{i\delta} \to -i$, whereas the correct limit is always -1, as we know from Sect. 2.3. Lekner (1995, 1996) shows that, for all profiles,

$$\delta = \pm\pi + c\theta + O(\theta^3) = \pm\pi + c'\frac{\sin\theta}{\sin\theta_c} + O\left(\frac{\sin\theta}{\sin\theta_c}\right)^3. \tag{16.50}$$

In total reflection, the normal component of the wavevector in the substrate is imaginary: $q_2 = i|q_2|, |q_2| = \left(4\pi\rho_2 - q_1^2\right)^{\frac{1}{2}} = (2\pi/\lambda)\left[\sin^2\theta_c - \sin^2\theta\right]^{\frac{1}{2}}$, and the exact formula (16.46) gives (we set $a = 0$ for simplicity),

$$r = \frac{iq_1\left[|q_2|(F_aG_b - G_aF_b) + \left(F_aG_b' - G_aF_b'\right)\right] - |q_2|\left(F_a'G_b - G_a'F_b\right) - \left(F_a'G_b' - G_a'F_b'\right)}{iq_1\left[|q_2|(F_aG_b - G_aF_b) + \left(F_aG_b' - G_aF_b'\right)\right] + |q_2|\left(F_a'G_b - G_a'F_b\right) + \left(F_a'G_b' - G_a'F_b'\right)}. \tag{16.51}$$

Equation (16.51) can be written as

$$r = \frac{q_1 + iQ}{q_1 - iQ} = e^{2i\arctan Q/q_1}, \quad Q = \frac{|q_2|\left(F_a'G_b - G_a'F_b\right) + \left(F_a'G_b' - G_a'F_b'\right)}{|q_2|(F_aG_b - G_aF_b) + \left(F_aG_b' - G_aF_b'\right)}. \tag{16.52}$$

In the absence of absorption, Q is real. Thus the phase δ of the reflection amplitude $r = e^{i\delta}$ is given by

$$\delta = 2\arctan Q/q_1 \tag{16.53}$$

As $q_1 = (2\pi/\lambda)\sin\theta_1$ tends to zero, $|q_2| = \left(q_c^2 - q_1^2\right)^{\frac{1}{2}} \to q_c = (4\pi\rho_2)^{\frac{1}{2}} = (2\pi/\lambda)\sin\theta_c$, and Q tends to its glancing incidence value Q_0. Also, for real X,

$$\arctan X = \frac{\pi}{2}\mathrm{sgn}(X) - X^{-1} + O(X^{-3}). \tag{16.54}$$

Hence we have a proof of (16.50) for reflecting profiles of finite thickness:

$$\delta = \pi\,\mathrm{sgn}(Q_0) - \frac{2q_1}{Q_0} + O\left(\frac{q_1}{Q_0}\right)^3. \tag{16.55}$$

Equation (16.55) determines the constants c and c' in (16.50) to be $c = -(4\pi/\lambda)Q_0^{-1}$, $c' = -4\pi/Q_0L_2 = -4\sqrt{\pi\rho_2}/Q_0$. Lekner (1995) evaluates the slope of the reflection phase for the homogeneous layer, the linear profile, and the hyperbolic tangent profile. The last variation of the scattering length density extends continuously without bound, but still has the form (16.50). Analytic forms of c for linear and quadratic profiles are given in Lekner (1996).

For the linear variation of scattering length density we know from Sect. 16.4 that the functions F and G are Airy functions, and the reflection phase can be evaluated exactly from (16.52) and (16.53). Figure 16.6 show the phase for the profile of

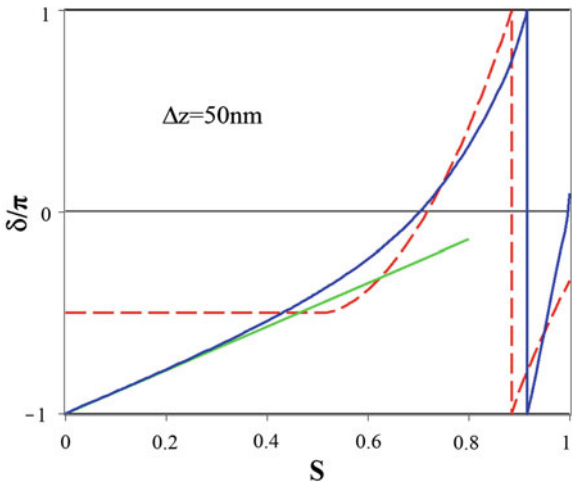


Fig. 16.6 The reflection phase δ plotted as a function of $S = \sin\theta/\sin\theta_c$, for the same material parameters as the profile of Fig. 16.5, with layer thickness 500 Å. The straight line is the small-angle variation predicted by (16.55), the dashed curve is the short-wave approximation (16.49), and the full curve is the exact phase (16.53)

Fig. 16.5, and Fig. 16.7 for the same material parameters, but a thicker layer of 2000 Å depth. These figures also show the high-frequency approximation (16.49), in which the integral is to be taken over the real part of $q(z)$. Since $q^2(z) = 4\pi\rho_2 S^2 - 4\pi\rho(z)$ is negative for $S = \sin\theta/\sin\theta_c < S_a = \sin\theta_a/\sin\theta_c = \sqrt{\rho_a/\rho_2}$, the real part of the phase integral in (16.49) is zero for $S < S_a$. There is also a change in analytic form at $S_b = \sin\theta_b/\sin\theta_c = \sqrt{\rho_b/\rho_2}$. The phase integral thus has three forms (Lekner 1995): zero for $S < S_a$, and

$$\int_0^{z_0} \mathrm{d}z\, q(z) = \frac{1}{6\pi}\frac{\Delta z}{\Delta\rho}\left(q_1^2 - 4\pi\rho_a\right)^{\frac{3}{2}}, \quad S_a < S < S_b,$$

$$\int_0^{z_0} \mathrm{d}z\, q(z) = \frac{1}{6\pi}\frac{\Delta z}{\Delta\rho}\left\{\left(q_1^2 - 4\pi\rho_a\right)^{\frac{3}{2}} - \left(q_1^2 - 4\pi\rho_b\right)^{\frac{3}{2}}\right\}, \quad S_b < S < 1. \tag{16.56}$$

We see that the high frequency approximation (16.49) fails at glancing incidence, as is expected, since $q_1 \to 0$ there, and is not good at the smaller thickness of $\Delta z = 500$ Å in Fig. 16.6, but it is a good approximation (away from glancing incidence) for $\Delta z = 2000$ Å shown in Fig. 16.7. The phase curves are drawn against $S = \sin\theta/\sin\theta_c$, and are thus universal as regards wavelength and angle of incidence. They differ in the thickness and form of the layer, the operative dimensionless constants being $(\Delta z)^2\rho$, where the various scattering length densities

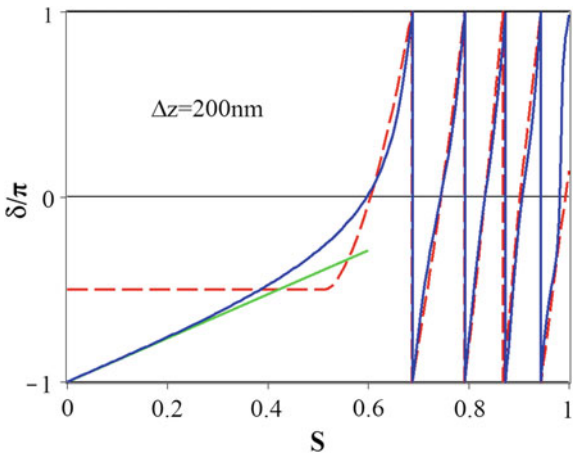


Fig. 16.7 Reflection phase δ for the linear profile. As for Fig. 16.6, but now with four times greater thickness of the linear variation in density, $\Delta z = 2000$ Å

ρ_a, ρ_b, ρ_2 are substituted for ρ. These dimensionless constants are all sixteen times larger when the thickness is quadrupled, improving the agreement with the exact reflection phase.

In the preceding we explored the properties of the reflection phase δ, which relates to the motion in the z-direction, normal to the reflecting stratification. There is also a phase difference between the reflected and direct neutron waves due to motion in the x-direction. Figure 16.8 shows the two paths.

The phase difference between the reflected and direct neutron rays can be broken up into the straight-path contributions (where the neutrons travel in vacuum), and the curved part contribution from within the reflecting stratification. The straight-path phase difference is, with $L = D + \Delta x + d$ being the horizontal distance between the source and the detector, and Δx the horizontal component of the curved path,

$$\begin{aligned}\Delta_s &= \frac{2\pi}{\lambda}\left\{\sqrt{D^2+H^2}+\sqrt{d^2+h^2}-\sqrt{L^2+(H-h)^2}\right\}\\ &= \frac{2\pi}{\lambda}\left\{-\Delta x+\frac{2Hh}{L}+\frac{(H^2+h^2)\Delta x}{2L^2}+O(L^{-3})\right\}.\end{aligned} \tag{16.57}$$

The curved part contributes

$$\Delta_c = K\Delta x + \delta = \frac{2\pi}{\lambda}\Delta x\cos\theta + \delta. \tag{16.58}$$

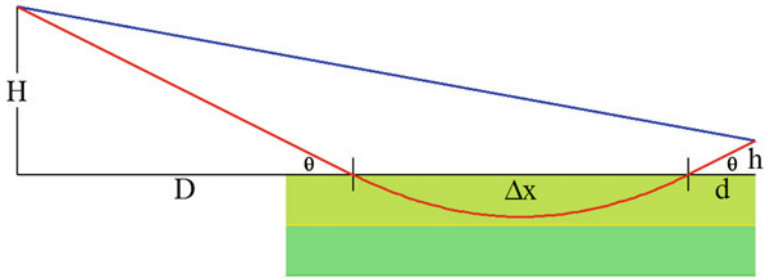


Fig. 16.8 Lloyd fringe formation in total reflection, with *vertical* distances enhanced for clarity. If the total *horizontal* distance between source and detector is $L = D + \Delta x + d$, the direct ray has length $\sqrt{L^2 + (H-h)^2}$. Since $H/D = \tan\theta = h/d$, we have $D = (L - \Delta x)[1 + h/H]^{-1}$, $d = (h/H)D$

The tangential component K of the wave vector is a constant of the motion in a planar stratification. The value of Δx in the ray picture is, from (10.50) of Sect. 10.3,

$$\Delta x \approx 2\int_0^{z_0} \mathrm{d}z \cot\theta(z) = 2\int_0^{z_0} \mathrm{d}z \frac{K}{q(z)} = \frac{K}{\sqrt{\pi}} \int_0^{z_0} \frac{\mathrm{d}z}{\sqrt{\rho(z_0) - \rho(z)}}. \tag{16.59}$$

The last equality follows from $q^2(z) = q_1^2 - 4\pi\rho(z)$ and $0 = q_1^2 - 4\pi\rho(z_0)$, the definition of the classical turning point. The semiclassical value of Δx given in (16.59) also follows from $\Delta x \approx \mathrm{d}\delta/\mathrm{d}K$ ((10.51) of Sect. 10.3) when δ is approximated by (16.49).

The total phase difference between the direct and reflected waves is $\Delta = \Delta_s + \Delta_c$, and the Lloyd mirror fringe intensity is proportional to $|1 + \exp(i\Delta)|^2 = 4\cos^2\frac{\Delta}{2}$. The contribution of the curved part of the reflected ray itself depends on the scattering length density variation with z, as can be seen from (16.59). Lekner (1996) gives details for the linear and quadratic variation with z. The fringe spacing, which correspond to a change of 2π in Δ, is

$$\Delta h = \frac{2\pi}{\mathrm{d}\Delta/\mathrm{d}h}. \tag{16.60}$$

For a linear profile without discontinuities this is constant, equal to $\lambda D/2H$; for a quadratic variation the spacing decreases as h increases. Thus, in principle at least, neutron scattering density profiles can be distinguished in the Lloyd mirror experiment (the feasibility of optical refractive index determination has been demonstrated in Allman et al. 1993a, b). In the neutron Lloyd's mirror experiment, the wavelength is smaller and the fringes are closer together. Photographic film

impregnated with boron 10 or lithium 6 may give better resolution than conventional detectors (Lekner 1995).

Other methods of extracting the phase in neutron reflectometry have been proposed: Majkrzak and Berk (1995) by means of a known reference layer having three tuneable values of the scattering length density, and de Haan et al. (1995) by adding to the unknown layer a known ferromagnetic layer. For liquid surfaces and thin films on liquid surfaces, the phase problem can be treated in the distorted-wave Born approximation (Blasie et al. 2003).

16.6 Reflection of Neutrons by Periodic Stratifications

It is well known that periodic structures diffract waves strongly into certain directions determined by the constructive interference of the scattered waves. Planar periodic structures reflect specularly, as we saw in Chap. 13, and most strongly at wavelengths and angles of incidence which are such as to combine in phase the waves reflected from each of the repeated components of the stratification. These stop bands or Bragg peaks, as they are known, are of finite width (in angle, or wavelength, or frequency), because the periodic structure which produces the reflection modifies the propagating waves to such an extent that, if the structure were infinite, forbidden bands would appear, within which only evanescent non-propagating waves are possible. The infinite and semi-infinite cases have been thoroughly explored, for instance in solid-state physics. In neutron reflection studies, interest lies in finite stratifications with a relatively small number of repetitions of the basic unit, for example nickel-titanium multilayers (Penfold 1991), or lamellar phases of polyolefin diblock copolymer films (Foster et al. 1992). The theory is treated in Lekner (1994) and Sears (1997). Figure 16.9 illustrates a four-fold repetition of a Ni–Ti bilayer.

In Sect. 12.2 we showed that the reflection and transmission amplitudes for a general stratification are given in terms of the stratification matrix $M = \{m_{ij}\}$ by

$$r = e^{2iq_1 a} \frac{q_1 q_2 m_{12} + m_{21} - iq_2 m_{11} + iq_1 m_{22}}{q_1 q_2 m_{12} - m_{21} + iq_2 m_{11} + iq_1 m_{22}}, \tag{16.61}$$

$$t = e^{iq_1 a - iq_2 b} \frac{2iq_1 (m_{11} m_{22} - m_{12} m_{21})}{q_1 q_2 m_{12} - m_{21} + iq_2 m_{11} + iq_1 m_{22}}. \tag{16.62}$$

(We have reinstated the matrix determinant $\det(M) = m_{11} m_{22} - m_{12} m_{21}$, normally unity, in the numerator.) We shall compare these with the general formulae (16.46) in Sect. 16.2, expressed in terms of exact solutions F, G of $\mathrm{d}^2\psi/\mathrm{d}z^2 + q^2\psi = 0$ within the reflecting layer, namely

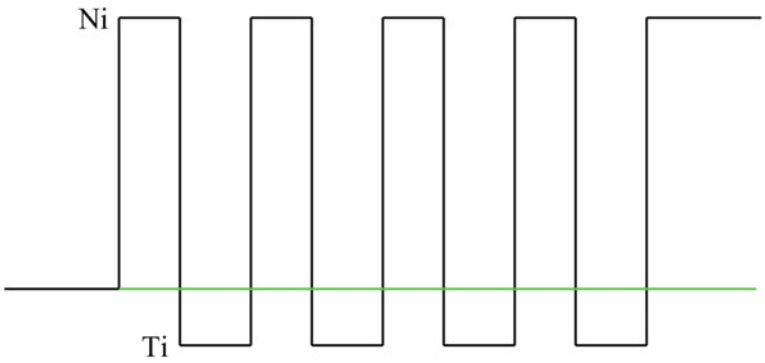


Fig. 16.9 A fourfold repetition of a Ni–Ti bilayer, on a substrate of nickel. The layer thicknesses are $Ni : 47.8$ Å, $Ti : 55.7$ Å. The structure shown is drawn to scale, with bound coherent scattering lengths $b(Ni) = 10.3fm$, $b(Ti) = -3.4fm$, and volumes per atom $v(Ni) = 11.14$ Å^3, $v(Ti) = 17.67$ Å^2. The corresponding scattering length densities are $\rho(Ni) \approx 9.25 \times 10^{-6}$ Å^{-2}, $\rho(Ti) \approx -1.92 \times 10^{-6}$ Å^{-2}, indicated by *Ni* and *Ti* on the diagram

$$r = e^{2iq_1a}\frac{q_1q_2(F,G) + iq_1(F,G') + iq_2(F',G) - (F',G')}{q_1q_2(F,G) + iq_1(F,G') - iq_2(F',G) + (F',G')}, \tag{16.63}$$

$$t = e^{iq_1a - iq_2b}\frac{2q_1(F_bG'_b - G_bF'_b)}{q_1q_2(F,G) + iq_1(F,G') - iq_2(F',G) + (F',G')}. \tag{16.64}$$

In (16.63) and (16.64) we have used the shorthand notation

$$F_aG_b - G_aF_b = (F,G), \quad F_aG'_b - G_aF'_b = (F,G'),$$
$$F'_aG_b - G'_aF_b = (F',G), \quad F'_aG'_b - G'_aF'_b = (F',G'). \tag{16.65}$$

The combination $W = FG' - F'G$ is the Wronskian of the solutions; W is independent of z. Comparison of (16.61) with (16.63) and (16.62) with (16.64) shows that the exact matrix elements are (Lekner 1994)

$$m_{11} = -(F',G)/W, \quad m_{12} = (F,G)/W, \quad m_{21} = -(F',G')/W, \quad m_{22} = (F,G')/W. \tag{16.66}$$

With the common factor W^{-1} in (16.66), the matrix is unimodular:

$$\det(M) = W^{-2}\{(F,G)(F',G') - (F',G)(F,G')\} = 1. \tag{16.67}$$

This result follows from the identity (2.31).

We now consider a periodic stratification, consisting of N repetitions of a fundamental unit cell. In the example of Fig. 16.9, $N = 4$, and the unit cell is a Ni–Ti bilayer. We know from Chap. 13, (16.1) to (16.4), that if $M = \{m_{ij}\}$ is the matrix of one period, the matrix for N periods is

$$M^N = \begin{bmatrix} m_{11}S_N - S_{N-1} & m_{12}S_N \\ m_{21}S_N & m_{22}S_N - S_{N-1} \end{bmatrix}, \tag{16.68}$$

$$S_N = \frac{\sin N\Phi}{\sin \Phi}, \quad \cos\Phi = \frac{1}{2}\text{trace}(M) = (m_{11} + m_{22}). \tag{16.69}$$

In the simplest case of pairs of homogeneous layers, as illustrated in Fig. 16.9, the unit cell matrix is a product of the unimodular matrices of the two homogeneous layers:

$$\begin{aligned} M &= \begin{bmatrix} c_l & q_l^{-1}s_l \\ -q_l s_l & c_l \end{bmatrix} \begin{bmatrix} c_h & q_h^{-1}s_h \\ -q_h s_h & c_h \end{bmatrix} \\ &= \begin{bmatrix} c_l c_h - q_l^{-1}q_h s_l s_h & q_h^{-1}c_l s_h + q_l^{-1}s_l c_h \\ -q_l s_l c_h - q_h c_l s_h & c_l c_h - q_l q_h^{-1} s_l s_h \end{bmatrix}. \end{aligned} \tag{16.70}$$

The notation used in (16.70) is as follows: the subscripts h, l stand for the high and low values of the scattering length density $\rho = b/v$, $c_l = \cos\delta_l, s_l = \sin\delta_l, \delta_l = q_l l$, where q_l is the value of the normal component of the wavevector in the layer where $\rho = \rho_l$, found from $q^2 = q_1^2 - 4\pi\rho$, and likewise for c_h, s_h, δ_h. The matrices for the homogeneous layers follow from (16.66) on setting $F = \cos qz, G = \sin qz$. For this unit cell matrix the phase Φ is given by

$$\begin{aligned} \cos\Phi &= c_l c_h - \frac{1}{2}s_l s_h\left(q_l^{-1}q_h + q_l q_h^{-1}\right) \\ &= \cos(\delta_l + \delta_h) - \frac{1}{2}s_l s_h\left(\sqrt{q_l/q_h} - \sqrt{q_h/q_l}\right)^2. \end{aligned} \tag{16.71}$$

We see that Φ is *approximately* equal to the total phase increment across a single period, $\Phi \approx \delta_h + \delta_l$. The same approximation holds for a general bilayer (Lekner 1994). The Bragg formula $2d\sin\theta = n\lambda$, which applies to radiation of wavelength λ incident at glancing angle θ onto *planes* of scatterers separated by distance d, can be written as $qd = n\pi$, since the normal component of the wavevector is $q = (2\pi/\lambda)\sin\theta$ (if one assumes the planes of scatterers to be separated by vacuum). Thus, according to the Bragg formula, reflection maxima occur when the phase increment qd across one period is a multiple of π. (This begs the question: when the medium is not vacuum, but a composite structure, what q should one put into the Bragg formula?) As we have seen in Sect. 13.2, a semi-infinite periodic stratification reflects totally whenever $\cos\Phi$ lies outside the range $[-1, 1]$. Such regions correspond to band gaps; the band edges are given by $\cos^2\Phi = 1$. When

$\cos^2 \Phi > 1$ the waves cannot propagate within the periodic structure, and reflection is perfect for a non-absorbing semi-infinite medium, and strong for finite or absorbing periodic structures. We conclude that the designation of regions of strong reflectivity as Bragg peaks is a simplification of (or at best a shorthand for) the underlying band structure, even though in neutron reflection the stop bands can be very narrow, as illustrated in the figure below.

Figure 16.10 shows the function $\cos \Phi$ and the reflectivity R for a single bilayer and a 15-bilayer Ni–Ti stack on a nickel substrate, a part of which was shown in Fig. 16.9.

To calculate the reflectivity of a periodic structure with N periods, we substitute the matrix elements of the matrix M^N for the complete stratification as given in (16.68) into the reflection amplitude (16.61). The result is

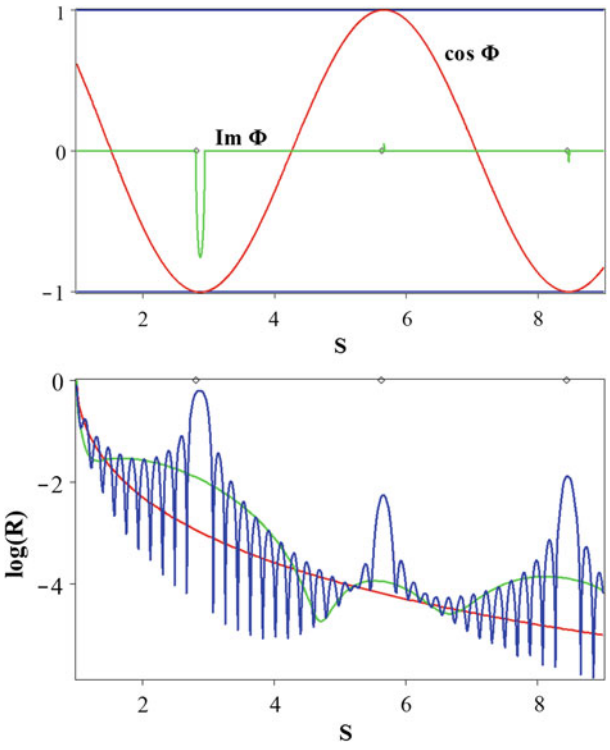


Fig. 16.10 Glancing angle dependence of the function $\cos \Phi = (m_{11} + m_{22})/2$ and the reflectivity R for a 15-bilayer Ni–Ti stack, with parameters as given in Fig. 16.9. Whenever $\cos^2 \Phi > 1$ the infinite periodic structure has a band gap, and reflects totally; the corresponding imaginary part of Φ is shown in the *upper graph*. The non-oscillatory curve on the *lower graph* is the Fresnel reflectivity of the bare substrate, $R_F = (q_1 - q_2)^2/(q_1 + q_2)^2$. The smooth oscillatory curve is the reflectivity of a single period on the same substrate. The graphs are plotted as a function of the angle variable $S = \sin\theta/\sin\theta_c$.The diamonds on both graphs indicate the locations of the Bragg angles, for which S is an integer times $\pi/q_c d$, from $q_1 d = n\pi$

$$r = e^{2iq_1 a} \frac{q_1 q_2 m_{12} + m_{21} - iq_2(m_{11} - \sigma_N) + iq_1(m_{22} - \sigma_N)}{q_1 q_2 m_{12} - m_{21} + iq_2(m_{11} - \sigma_N) + iq_1(m_{22} - \sigma_N)}. \tag{16.72}$$

As in Sect. 13.2, it is convenient to work in terms of the quantity

$$\sigma_N = \frac{S_{N-1}}{S_N} = \frac{\sin(N-1)\Phi}{\sin N\Phi} = \cos\Phi - \sin\Phi \cot N\Phi. \tag{16.73}$$

We noted in Sect. 13.2 that when $N\Phi$ is a multiple of π and $(N-1)\Phi$ is not, σ_N is infinite and $R = |r|^2 \to (q_1 - q_2)^2/(q_1 + q_2)^2$, the Fresnel reflectivity of the bare substrate. On the other hand, when $(N-1)\Phi$ is a multiple of π and $N\Phi$ is not, σ_N is zero, and the reflection amplitude is the same as that of a single period of the structure (supported by the substrate), namely as given by (16.61) with the matrix elements of (16.70). Hence the rapid variation of the reflectivity with angle of incidence for the $N = 15$ stratification seen in Fig. 16.10.

16.7 Neutron Reflection by Magnetic Materials

The neutron magnetic dipole $\boldsymbol{\mu}$ interacts with the magnetic field $\mathbf{B}$, the interaction energy being $-\boldsymbol{\mu}.\mathbf{B}$. Neutron spin is quantized, with equal and opposite magnetic moments parallel and antiparallel to the magnetic field, so the magnetic interaction will add equal and opposite terms to the interaction potential. The result is that two waves of opposite spin travel with different phase velocities within magnetic media, just as waves of opposite circular polarization are characterized by different refractive indices in the case of chiral media (Chap. 18).

It follows from Maxwell's equations in magnetic media that *only the magnetization tangential to the surface of the reflecting medium* is active in the neutron-medium magnetic interaction. Let us suppose, as usual, that the plane of incidence is the zx plane, and that the outer surface of the magnetic reflector is the $z = 0$ plane. The simplest case arises when the magnetization is along the y direction (normal to the scattering plane). Then the magnetic interaction potential is (Sears (1989) Sect. 3.5, Zabel et al. (2007))

$$\frac{2\pi\hbar^2}{Mv} b_m \sigma_y, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \tag{16.74}$$

In (16.74) σ_y is the y-component Pauli spin matrix, which has eigenvalues ± 1; M is the mass of a neutron, and v the volume per scatterer. In this simplest case the total effective medium-neutron interaction is

$$V_{\pm} = V_n \pm V_m = \frac{2\pi\hbar^2}{Mv}(b_n \pm b_m) = \frac{2\pi\hbar^2}{Mv} b_{\pm}. \tag{16.75}$$

The subscripts n, m stand for *nuclear* and *magnetic*, and the plus sign is associated with neutron spin parallel to the ambient field, the minus sign with spin antiparallel. The effective dielectric function replacing (16.5) becomes

$$\varepsilon_{\pm} = 1 - \frac{\lambda^2}{\pi v}(b_n \pm b_m) \tag{16.76}$$

From $q_1 = \frac{2\pi}{\lambda}\sin\theta$ and $q_2^2 = q_1^2 - 4\pi b/v$ we see that reflection will be total for the two neutron spin orientations for glancing angles smaller than the critical angles $\theta_c^{\pm}$ given by

$$\sin\theta_c^{\pm} = \lambda\sqrt{b_{\pm}/\pi v} \tag{16.77}$$

The scattering lengths $b_{\pm}$ can be quite different: for *Fe* for example, $b_n = 9.54,\ b_m = 5.98,\ b_+ = 15.52,\ b_- = 3.56$. (The scattering lengths here and below are in $fm = 10^{-15}$ m, from Table 3.4 of Sears 1989.) Thus a range of angles exists for which one polarization is reflected totally, and the other is not. For neutrons of 10 Å wavelength incident on fully magnetized iron, $\theta_c^+ \approx 1.17°,\ \theta_c^- \approx 0.56°$.

For *Co* the antiparallel polarization has a negative total scattering length, and therefore no critical angle: $b_n = 2.50,\ b_m = 4.64,\ b_+ = 7.14,\ b_- = -2.14$. Figure 16.11 shows the reflectivity of the two neutron spin orientations from magnetized cobalt, assuming a sharp interface. The volume per scatterer in cobalt is 11.1 Å^3, and $\theta_c^+ \approx 0.82°$ for neutrons with 10 Å wavelength. The neutrons with positive spin orientation have the reflectivity obtained by squaring the step profile reflection amplitude (16.12) for $\theta \geq \theta_c^+$ (and unity for $\theta \leq \theta_c^+$):

$$R_+ = \left[S + \left(S^2 - 1\right)^{\frac{1}{2}}\right]^{-4}, \quad S = \frac{\sin\theta}{\sin\theta_c^+} \tag{16.78}$$

For negative spin orientation we can use the formulation of Sect. 16.2, namely

$$R_- = \left[S + \left(S^2 + 1\right)^{\frac{1}{2}}\right]^{-4}, \quad S = \frac{\sin\theta}{\lambda\sqrt{|b_-|/\pi v}} \tag{16.79}$$

The analysis of neutron polarization experiments involves correction for the polarizer and analyser efficiencies, and contamination by unwanted spin states. These aspects are covered by Wildes (1999, 2006).

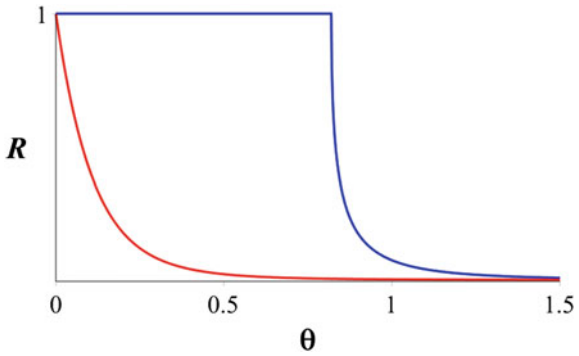


Fig. 16.11 Reflectivity of neutrons with 10Å wavelength by magnetized cobalt, as functions of the angle of incidence (in degrees). Those with spin parallel to the magnetic field (*upper curve*) reflect totally for $\theta \leq \theta_c^+ \approx 0.82°$. Those with spin antiparallel to the magnetic field (*lower curve*) have a reflectivity near θ_c^+ which is small (about 0.4 %), so the magnetized cobalt acts as a reflection polarizer

References

Allman BE, Klein AG, Nugent KA, Opat GI (1993a) Lloyd's mirage: a variant of Lloyd's mirror. Eur J Phys 14:272–276

Allman BE, Klein AG, Nugent KA, Opat GI (1993b) Refractive-index-profile determinations by using Lloyd's mirage. Appl Opt 33:1806–1811

Blasie JK, Zheng S, Strzalka J (2003) Solution to the phase problem for specular x-ray or neutron reflectivity from thin films on liquid surfaces. Phys Rev B 67:224201 (8 pp)

Daillant J, Gibaud A (eds) (2009) X-ray and neutron reflectivity. Springer, Berlin

Daillant J, Mora S, Sentenac A (2009) Diffuse scattering. In: Daillant J, Gibaud A (eds) X-ray and neutron reflectivity. Lecture Notes in Physics vol 770. Springer, Berlin, pp 133–182 (Chapter 4)

De Haan VO, van Well AA, Adenwala S, Felcher GP (1995) Retrieval of phase information in neutron reflectometry. Phys Rev B 52:10831–10833

Felcher GP, Russell TP (1991) Proceedings of the workshop on analysis and interpretation of neutron reflectivity data. Phys B 173:1–120

Felcher GP, Goyette RJ, Anastasiadis S, Russell TP, Foster M, Bates F (1994) Topology of forward scattering of neutrons from imperfect multilayers. Phys Rev B 50:9565–9658

Foster MD, Sikha M, Singh N, Bates FS, Satija SK, Majkrzak CF (1992) Structure of symmetric polyolefin block copolymer thin films. J Chem Phys 96:8605–8615

Gudkov VP, Opat GI, Klein AG (1993) Neutron reflection interferometry: physical principles of surface analysis with phase information. J Phys: Condens Matter 5:9013–9024

Lekner J (1991) Reflection theory and the analysis of neutron reflection data. Physica B 173:99–111

Lekner J (1994) Reflection of neutrons by periodic stratifications. Phys B 202:16–27

Lekner J (1995) Neutron reflection interferometry: extraction of the phase in total reflection from stratified media. Phys B 215:329–336

Lekner J (1996) Analysis of Lloyd's mirror fringes for graded index reflectors. J Opt Soc Am 13:1809–1815

Lu JR, Lee EM, Thomas RK (1996) The analysis and interpretation of neutron and X-ray specular reflection. Acta Cryst A 52:11–41

Majkrzak CF, Berk NF (1995) Exact determination of the phase in neutron reflectometry. Phys Rev B 52:10827–10830

Penfold J (1991) Instrumentation for neutron reflectivity. Physica B 173:1–10 (Figure 6)
Penfold J, Thomas RK (1990) The application of the specular reflection of neutrons to the study of surfaces and interfaces. J Phys: Condens Matter 2:1369–1412
Pershan PS (1990) Structure of surfaces and interfaces as studied using synchrotron radiation. Faraday Discuss Chem Soc 89:231–245
Rayleigh JWS (1912) On the propagation of waves through a stratified medium, with special reference to the question of reflection. Proc Roy Soc A86:207–266
Russell TP (1990) X-ray and neutron reflectivity for the investigation of polymers. Mater Sci Rep 5:171–271
Sears VF (1989) Neutron optics. Oxford University Press, Oxford
Sears VF (1997) Dynamical diffraction in periodic multilayers. Acta Crysr A 53:649–662
Sentenac A, Daillant J (2009) Statistical aspects of wave scattering at rough surfaces. In: Daillant J, Gibaud A (eds) X-ray and neutron reflectivity. Lecture in Notes Physics, vol 770. Springer, Berlin, pp 59–84 (Chapter 2)
Sinha SK, Sirota EB, Garoff S, Stanley HB (1988) X-ray and neutron scattering from rough surfaces. Phys Rev B 38:2297–2311
Werner SA, Klein AG (1986) Neutron optics. In: Methods of experimental physics, vol 23A. Academic Press, New York (Chapter 4)
Wildes A (1999) The polarizer-analyzer correction problem in neutron polarization analysis experiments. Rev Sci Inst 70:4241–4245
Wildes A (2006) Neutron polarization analysis corrections made easy. Neutron News 17:17–25
Zabel H, Theis-Bröhl K, Toperverg BP (2007) Polarized neutron reflectivity and scattering from magnetic nanostructures and spintronic materials. In: Kronmüller H, Parkin S (eds) Handbook of magnetism and advanced magnetic materials, vol **3**. Wiley, Joboken, pp 1–71

Chapter 17
Acoustic Waves

Section 1.4 introduced the basics of sound propagation in isotropic media, and some elementary properties of compressional wave reflection and transmission. An important aspect noted there was the possibility of zero reflection at a sharp boundary between two media at the Green's angle, the acoustic analogue of the Brewster angle. This chapter will summarize known results in the reflection and transmission of sound waves. As for electromagnetic and particle waves, there are relations that must be satisfied by the exact reflection and transmission amplitudes of acoustic compressional waves in an arbitrary planar stratification. These are the conservation and reciprocity theorems. We shall also give low-frequency and high-frequency limiting forms, and an upper bound on the reflectivity. These are compared with analytic solutions for two special stratifications, both having exponential variation of density with depth, and linear or exponential variations of sound speed.

17.1 General Relations for Stratified Media

The linearized equation for the acoustic pressure p is (Bergmann 1946; Brekhovskikh 1960)

$$\nabla^2 p - \rho^{-1}\nabla\rho\cdot\nabla p - v^{-2}\partial_t^2 p = 0, \tag{17.1}$$

where $v^2 = \partial_\rho p_a$ is the adiabatic derivative of the hydrostatic or ambient pressure p_a with respect to the density ρ. For planar stratifications, p and v are functions of the depth z only; $v(z)$ is usually referred to as the local value of the phase velocity, but this is literally true only if the medium changes little in one 'wavelength' = 'speed'/frequency (Gupta 1965). For a plane monochromatic wave propagating in the zx plane, solutions of (17.1) have the form

$$p(z,x,t) = \mathrm{e}^{i(Kx-\omega t)}P(z), \tag{17.2}$$

J. Lekner, *Theory of Reflection*, Springer Series on Atomic, Optical, and Plasma Physics 87, DOI 10.1007/978-3-319-23627-8_17

where ω is the angular frequency of the wave and K is the x component of the wave vector, which is a constant of the motion. The angle $\theta(z)$ between the normal to the wave front and the z axis, and the local sound speed $v(z)$, are related to K via the generalized Snell's law,

$$K = \frac{\omega}{v(z)} \sin\theta(z) = \text{constant}. \tag{17.3}$$

The differential equation for P was given in (1.57):

$$\rho \frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{\rho}\frac{\mathrm{d}P}{\mathrm{d}z}\right) + q^2 P = 0, \quad q^2(z) = \frac{\omega^2}{v^2(z)} - K^2, \tag{17.4}$$

where $q(z)$ is the normal component of the wavevector. Here we consider reflection and transmission by a bounded stratification between $z = a$ and $z = b$, with homogeneous media (for $z < a$ and $z > b$) above and below. Acoustic parameters relating to these bounding media will be labeled by the subscripts a or b; thus $q_a = (\omega/v_a)\cos\theta_a$ is the value of the normal component of the wave vector in medium a. When sound is incident from medium a and transmitted, via the stratification, to medium b, the pressure variable $P(z)$ takes the forms

$$P_{ab} = \mathrm{e}^{iq_a z} + r_{ab}\mathrm{e}^{-iq_a z} (z \le a), \quad P_{ab} = t_{ab}\mathrm{e}^{iq_b z} (z \ge b) \tag{17.5}$$

in media a and b. When sound is incident from below, the forms taken by $P(z)$ in media a and b are

$$P_{ba} = t_{ba}\mathrm{e}^{-iq_a z} (z \le a), \quad P_{ba} = \mathrm{e}^{-iq_b z} + r_{ba}\mathrm{e}^{iq_b z} (z \ge b). \tag{17.6}$$

We shall derive general relations linking the reflection amplitudes r_{ab}, r_{ba} and the transmission amplitudes t_{ab}, t_{ba}, and their relation to the appropriate acoustic reflectances and transmittances.

17.1.1 General Results for the Reflection and Transmission Amplitudes

Let $F(z)$ and $G(z)$ be two solutions of (17.4), and consider their Wronskian

$$W(F, G) = FG' - F'G, \tag{17.7}$$

where the prime denotes differentiation with respect to z. The derivative of W is $W' = FG'' - F''G$. From (17.4), $F'' = (\rho'/\rho)F' - q^2 F$, and likewise for G, so

$$W' = \frac{\rho'}{\rho} W, \quad \ln W = \ln \rho + \text{constant}. \tag{17.8}$$

Thus W/ρ is a constant: the Wronskian of the two solutions is proportional to the local density $\rho(z)$.

Consider first the Wronskian of $P_{ab}(z)$ and $P_{ba}(z)$. From (17.5) and (17.6) we see that in medium a this takes the value $-2iq_a t_{ba}$, and in medium b the value $-2iq_b t_{ab}$. Since W/ρ is a constant, this proves the reciprocity relation

$$Q_a t_{ba} = Q_b t_{ab}, \tag{17.9}$$

where Q denotes the normal component of the wave vector divided by the density:

$$Q(z) = \frac{q(z)}{\rho(z)} = \frac{\omega \cos\theta(z)}{v(z)\rho(z)}, \quad Q_a = \frac{q_a}{\rho_a}, \quad Q_b = \frac{q_b}{\rho_b}. \tag{17.10}$$

The derivation of (17.9) assumes that there is no attenuation in media a or b (otherwise the forms (17.5) and (17.6) would, with complex q_a or q_b, show unacceptable exponential growth).

Total internal reflection is excluded for the same reason. It has *not* been assumed that the stratification between the homogenous media a and b is free from attenuation. The equality (17.9) demonstrates that the (complex) transmission amplitudes t_{ab} and t_{ba} carry the same phase. It also implies that the transmittance (the fraction of acoustic intensity transmitted through the stratification) is the same from above and from below,

$$T_{ab} = T_{ba}, \quad \text{where} \quad T_{ab} = (Q_b/Q_a)|t_{ab}|^2, \quad T_{ba} = (Q_a/Q_b)|t_{ba}|^2. \tag{17.11}$$

To see that the transmittance is $T_{ab} = (Q_b/Q_a)|t_{ab}|^2$, consider the situation in Fig. 17.1, in which a beam incident from medium a insonifies a strip of unit width on the $z = a$ boundary. The energy density of a plane wave with complex acoustic pressure p is proportional to $|p|^2/\rho v^2$, so the energy flux is proportional to $|p|^2/\rho v$. For the case shown, the amount of energy in the primary wave that is incident on a unit area of interface in unit time is proportional to $\cos\theta_a/\rho_a v_a$, while the energy reflected away in unit time is proportional to $|r_{ab}|^2 \cos\theta_a/\rho_a v_a$. The energy transmitted in unit time is proportional to $|t_{ab}|^2 \cos\theta_b/\rho_b v_b$. It follows that

$$R_{ab} = |r_{ab}|^2, \quad T_{ab} = \frac{\rho_a v_a}{\rho_b v_b} \frac{\cos\theta_b}{\cos\theta_a} = \frac{Q_b}{Q_a} |t_{ab}|^2. \tag{17.12}$$

We next consider the Wronskian of P_{ab} and P^*_{ba}, assuming the absence of attenuation everywhere (if P is a solution of (17.4), so is its complex conjugate P^*, provided q^2 is real). This Wronskian takes the value $2iq_a r_{ab} t^*_{ba}$ in medium a, and $-2iq_b t_{ab} r^*_{ba}$ in medium b. Since W/ρ is constant, we have the reciprocity relation

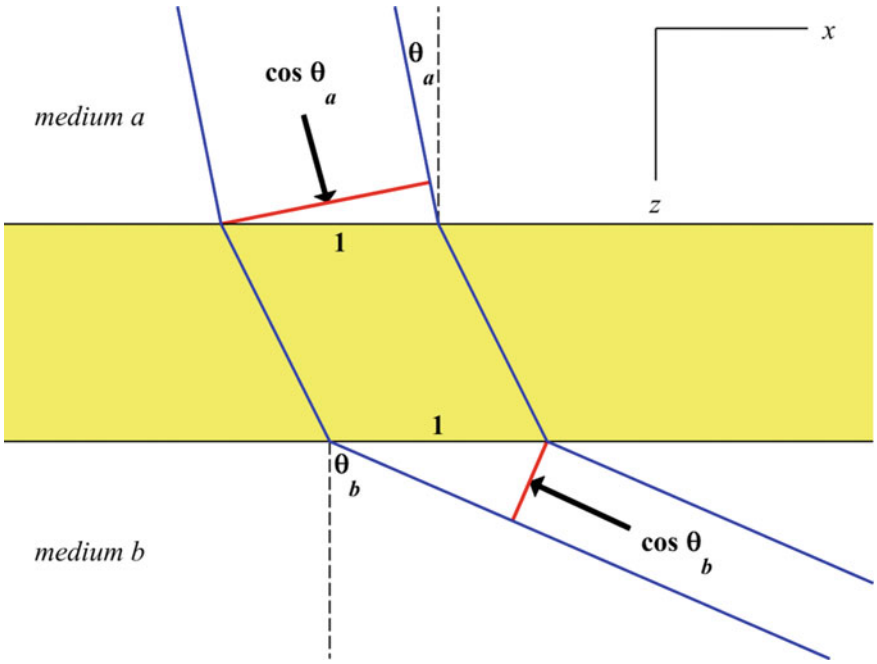


Fig. 17.1 A strip of unit width on the stratification is insonified by a beam of width $\cos\theta_a$; the transmitted beam has width $\cos\theta_b$ (the reflected beams are not shown)

$$Q_a r_{ab} t^*_{ba} = -Q_b t_{ab} r^*_{ba}. \tag{17.13}$$

Together with the reciprocity relation (17.9), this shows that

$$r^*_{ba} = -\left(t^*_{ba}/t_{ba}\right) r_{ab}, \tag{17.14}$$

which in turn implies the equality of the reflectances $R_{ab} = |r_{ab}|^2$ and $R_{ba} = |r_{ba}|^2$. Note that the required reality of q_a and q_b also excludes total internal reflection: for example, if the wave is totally internally reflected from medium b, q_b is imaginary.

Under the same conditions, the Wronskian of P_{ab} and P^*_{ab} is equal to $-2iq_a(1 - |r_{ab}|^2)$ in medium a, and $-2iq_b|t_{ab}|^2$ in medium b; the constancy of the Wronskian divided by the density thus implies

$$Q_a\left(1 - |r_{ab}|^2\right) = Q_b|t_{ab}|^2, \tag{17.15}$$

which is the energy conservation law

$$R_{ab} + T_{ab} = 1. \tag{17.16}$$

(The same equality links R_{ba} and T_{ba}.) The analogous results for electromagnetism and quantum mechanics are derived, by different methods, in Sect. 2.1.

The differential equation (17.4), to be satisfied by the acoustic pressure variable $P(z)$, is linear and of the second order. Thus, in a general stratification, (17.4) has two linearly independent solutions, say $F(z)$ and $G(z)$, and $P(z)$ is a linear combination of these within the stratification:

$$P = uF + vG. \tag{17.17}$$

Consider the reflection-transmission problem, in which sound is incident from medium a. Then P is given by (17.5) in media a and b, while within the inhomogeneous layer it is given by (17.17). The boundary conditions at $z = a$ and $z = b$ are the continuity of P and of $\rho^{-1}\mathrm{d}P/\mathrm{d}z$. (Note that these conditions are implied by the differential equation (17.4), and are not additional physical input.) These boundary conditions give four equations in the four unknown constants u, v, r, and t:

$$\mathrm{e}^{i\alpha} + r\mathrm{e}^{-i\alpha} = uF_a + vG_a, \quad uF_b + vG_b = t\mathrm{e}^{i\beta}, \tag{17.18}$$

$$iQ_a\left(\mathrm{e}^{i\alpha} - r\mathrm{e}^{-i\alpha}\right) = u\overline{F_a} + v\overline{G}_a, \quad u\overline{F_b} + v\overline{G_b} = iQ_b t\mathrm{e}^{i\beta}. \tag{17.19}$$

Here $\alpha = q_a a, \beta = q_b a$; r and t are the reflection and transmission amplitudes for insonification from medium a (we will drop the ab subscript from now on unless it is specifically needed); F_a stands for $F(z_a)$, $\overline{F_a}$ for the derivative of F at $z = a$ divided by the value of ρ just inside the stratification, and so on (This notation allows for possible discontinuity of density at either interface). Solving (17.18, 17.19) we find

$$r = \mathrm{e}^{2i\alpha}\frac{Q_aQ_b(F,G) + iQ_a\left(F,\overline{G}\right) + iQ_b\left(\overline{F},G\right) - \left(\overline{F},\overline{G}\right)}{Q_aQ_b(F,G) + iQ_a\left(F,\overline{G}\right) - iQ_b\left(\overline{F},G\right) + \left(\overline{F},\overline{G}\right)}, \tag{17.20}$$

$$t = \mathrm{e}^{i(\alpha-\beta)}\frac{2iQ_a\left(F_b\overline{G_b} - \overline{F_b}G_b\right)}{Q_aQ_b(F,G) + iQ_a\left(F,\overline{G}\right) - iQ_b\left(\overline{F},G\right) + \left(\overline{F},\overline{G}\right)}, \tag{17.21}$$

$$u = \mathrm{e}^{i\alpha}\frac{2iQ_a\left(\overline{G_b} - iQ_bG_b\right)}{Q_aQ_b(F,G) + iQ_a\left(F,\overline{G}\right) - iQ_b\left(\overline{F},G\right) + \left(\overline{F},\overline{G}\right)}, \tag{17.22}$$

$$v = -\mathrm{e}^{i\alpha}\frac{2iQ_a\left(\overline{F_b} - iQ_bF_b\right)}{Q_aQ_b(F,G) + iQ_a\left(F,\overline{G}\right) - iQ_b\left(\overline{F},G\right) + \left(\overline{F},\overline{G}\right)}. \tag{17.23}$$

In (17.20–17.23), $(F,G) \equiv F_a G_b - G_a F_b$, $\left(F,\overline{G}\right) \equiv F_a \overline{G_b} - G_a \overline{F_b}$, et cetera. Note that there is a common denominator to all four expressions. Corresponding results for electromagnetic waves were given in Sect. 2.2.

If the density ρ is continuous across both the interfaces at $z = a$ and at $= b$, the equations linking the derivatives across the boundaries simplify, and the (17.20–17.23) may be replaced by a set in which $Q_a \to q_a, Q_b \to q_b$ and $\overline{F} \to F', \overline{G} \to G'$.

In the absence of attenuation or total internal reflection, q is real everywhere. Then F and G may be chosen to be real, being the solutions of a real differential equation, and $R = |r|^2$ and $T = (Q_b/Q_a)|t|^2$ are given by

$$R = \frac{\left[Q_a Q_b (F,G) - \left(\overline{F},\overline{G}\right)\right]^2 + \left[Q_a \left(F,\overline{G}\right) + Q_b \left(\overline{F},G\right)\right]^2}{\left[Q_a Q_b (F,G) + \left(\overline{F},\overline{G}\right)\right]^2 + \left[Q_a \left(F,\overline{G}\right) - Q_b \left(\overline{F},G\right)\right]^2}, \tag{17.24}$$

$$T = \frac{4 Q_a Q_b \left(F_b \overline{G_b} - \overline{F_b} G_b\right)^2}{\left[Q_a Q_b (F,G) + \left(\overline{F},\overline{G}\right)\right]^2 + \left[Q_a \left(F,\overline{G}\right) - Q_b \left(\overline{F},G\right)\right]^2}. \tag{17.25}$$

By using the identity

$$(F,G)\left(\overline{F},\overline{G}\right) - \left(F,\overline{G}\right)\left(\overline{F},G\right) = (F_a \overline{G_a} - \overline{F_a} G_a)(F_b \overline{G_b} - \overline{F_b} G_b) \equiv \left(\frac{W}{\rho}\right)_a \left(\frac{W}{\rho}\right)_b, \tag{17.26}$$

and the fact that W/ρ is a constant, the conservation law (17.16) is seen to be satisfied by (17.24) and (17.25). It also follows from (17.24) that $R < 1$, as can be seen by writing it in the form

$$R = 1 - \frac{4 Q_a Q_b (W/\rho)^2}{\left[Q_a Q_b (F,G) + \left(\overline{F},\overline{G}\right)\right]^2 + \left[Q_a \left(F,\overline{G}\right) - Q_b \left(\overline{F},G\right)\right]^2}. \tag{17.27}$$

In total internal reflection (the angle of incidence exceeding the critical angle $\arcsin v_a/v_b$), $q_b^2 = \frac{\omega^2}{v_b^2} - K^2 = \frac{\omega^2}{v_a^2}\left(\frac{v_a^2}{v_b^2} - \sin^2\theta_a\right)$ is negative, q_b is imaginary, and the reflection amplitude takes the form $r = \mathrm{e}^{2i\alpha}(iA - B)/(iA + B)$, so that $|r|^2 = 1$. But note that this is true only in the absence of attenuation, which makes the reflection less than perfect even if q_b is pure imaginary.

At grazing incidence (from medium a) the normal component $q_a = (\omega/v_a)\cos\theta_a$ of the wave vector tends to zero. It follows from (17.20) to (17.21) that

$$r \to -1, \quad t \to 0 \quad \text{as} \quad \theta_a \to 90^\circ. \tag{17.28}$$

(F and G are functionals of $q^2(z) = (\omega^2/v^2(z)) - K^2$, and thus depend on the angle of incidence through $K = (\omega/v_a)\sin\theta_a$; however, this dependence cannot override

the effect of $Q_a \to 0$ in (17.20) and (17.21)). Thus there is perfect reflection and zero transmission at glancing incidence. The reflected wave is then 180° out of phase with the incident wave. These statements hold whether or not there is attenuation in the stratification and/or the bottom medium, and also hold when there is total internal reflection (q_b imaginary).

The reflection and transmission amplitudes for insonification 'from below' may be obtained by applying the boundary conditions to (17.6) with P written in the form (17.17). They are

$$r_{ba} = \mathrm{e}^{-2i\beta} \frac{Q_a Q_b(F,G) - iQ_a\big(F,\overline{G}\big) - iQ_b\big(\overline{F},G\big) - \big(\overline{F},\overline{G}\big)}{Q_a Q_b(F,G) + iQ_a\big(F,\overline{G}\big) - iQ_b\big(\overline{F},G\big) + \big(\overline{F},\overline{G}\big)}, \tag{17.29}$$

$$t_{ba} = \mathrm{e}^{i(\alpha-\beta)} \frac{2iQ_b\big(F_b\overline{G_b} - \overline{F_b}G_b\big)}{Q_a Q_b(F,G) + iQ_a\big(F,\overline{G}\big) - iQ_b\big(\overline{F},G\big) + \big(\overline{F},\overline{G}\big)}. \tag{17.30}$$

Together with r_{ab} and t_{ab} given by (17.20) and (17.21), they satisfy the reciprocity theorems (17.9) and (17.14). The corresponding solution for the constants in $P = uF + vG$ is

$$u = \mathrm{e}^{-i\beta} \frac{2iQ_b\big(\overline{G_a} + iQ_a G_a\big)}{Q_a Q_b(F,G) + iQ_a\big(F,\overline{G}\big) - iQ_b\big(\overline{F},G\big) + \big(\overline{F},\overline{G}\big)}, \tag{17.31}$$

$$v = -\mathrm{e}^{i\beta} \frac{2iQ_b\big(\overline{F_a} + iQ_a F_a\big)}{Q_a Q_b(F,G) + iQ_a\big(F,\overline{G}\big) - iQ_b\big(\overline{F},G\big) + \big(\overline{F},\overline{G}\big)}. \tag{17.32}$$

The general expressions for reflection and transmission amplitudes derived here will be used in the next three sections to obtain low-frequency and high-frequency limiting forms, and exact solutions for certain velocity and density profiles.

17.2 Matrix Methods

The techniques introduced in Chap. 12 may be adapted to acoustic compressional waves. The method involves taking the product of N 2×2 matrices when the stratification is approximated by N layers. These layers can be chosen to have linear variation in the acoustic parameters to best represent the actual stratification without undue complexity in the resulting matrix elements. It is possible to guarantee unimodularity of the matrices, thus making sure that the energy conservation law (17.15) or (17.16) and the reciprocity law (17.13) or (17.14) are automatically satisfied. The accuracy of these methods will be tested in Sects. 17.3 and 17.4 against an exactly solvable model stratification, in which the density and speed of sound both vary exponentially with depth.

The second-order differential equation (17.4) for $P(z)$ may be written as a pair of coupled first-order differential equations in P and its derivative divided by the density:

$$\frac{1}{\rho}\frac{\mathrm{d}P}{\mathrm{d}z} = D, \quad \frac{\mathrm{d}D}{\mathrm{d}z} = -\frac{q^2}{\rho}P. \tag{17.33}$$

We divide the inhomogeneous stratification lying between $z = a$ and $z = b$ into N layers, with $z_n \leq z \leq z_{n+1}$ in the nth layer, and $z_1 = a$, $z_{N+1} = b$. The integral versions of (17.33), incorporating the boundary values at z_n, are

$$P(z) = P_n + \int_{z_n}^{z} \mathrm{d}\zeta \rho(\zeta) D(\zeta), \quad D(z) = D_n + \int_{z_n}^{z} \mathrm{d}\zeta \frac{q^2(\zeta)P(\zeta)}{\rho(\zeta)}. \tag{17.34}$$

These coupled integral equations can be solved by iteration. We set

$$P(z) = \sum_{j=0}^{\infty} P^j(z), \quad D(z) = \sum_{j=0}^{\infty} D^j(z), \tag{17.35}$$

and start with $P^0 = P_n, D^0 = D_n$. The superscript j gives the degree of the correction in the thickness $\delta z_n = z_{n+1} - z_n$. The first iterates are

$$P^1(z) = D_n \int_{z_n}^{z} \mathrm{d}\zeta \rho(\zeta), \quad D^1(z) = -P_n \int_{z_n}^{z} \mathrm{d}\zeta \frac{q^2(\zeta)}{\rho(\zeta)}. \tag{17.36}$$

The second-order iterates, evaluated at z_{n+1}, are

$$P^2(z_{n+1}) = \int_{z_n}^{z_{n+1}} \mathrm{d}z\, \rho(z) D^1(z) = -P_n \int_{z_n}^{z_{n+1}} \mathrm{d}z\, \rho(z) \int_{z_n}^{z} \mathrm{d}\zeta \frac{q^2(\zeta)}{\rho(\zeta)}, \tag{17.37}$$

$$D^2(z_{n+1}) = \int_{z_n}^{z_{n+1}} \mathrm{d}z\, q^2(z) P^1(z)/\rho(z) = -D_n \int_{z_n}^{z_{n+1}} \mathrm{d}z\, q^2(z)/\rho(z) \int_{z_n}^{z} \mathrm{d}\zeta \rho(\zeta). \tag{17.39}$$

To find the matrix relation between P_{n+1}, D_{n+1} and P_n, D_n we evaluate (17.35) at z_{n+1}. To second order in $\delta z_n = z_{n+1} - z_n$ we need the integrals

$$I_1 = \int_{z_n}^{z_{n+1}} \mathrm{d}z\, \rho(z), \quad I_2 = \int_{z_n}^{z_{n+1}} \mathrm{d}z\, \rho(z) \int_{z_n}^{z} \mathrm{d}\zeta \frac{q^2(\zeta)}{\rho(\zeta)}, \tag{17.40}$$

$$J_1 = \int_{z_n}^{z_{n+1}} \mathrm{d}z \frac{q^2(z)}{\rho(z)}, \quad J_2 = \int_{z_n}^{z_{n+1}} \mathrm{d}z\, q^2(z)/\rho(z) \int_{z_n}^{z} \mathrm{d}\zeta \rho(\zeta). \tag{17.41}$$

The second-order matrix which relates P and D at z_{n+1} to the values at z_n is thus

$$\begin{pmatrix} P_{n+1} \\ D_{n+1} \end{pmatrix} = \begin{pmatrix} 1 - I_2 & I_1 \\ -J_1 & 1 - J_2 \end{pmatrix} \begin{pmatrix} P_n \\ D_n \end{pmatrix}. \tag{17.42}$$

We note that by interchange of the order of integration J_2 may be written in the form

$$J_2 = \int_{z_n}^{z_{n+1}} \mathrm{d}z\, \rho(z) \int_{z}^{z_{n+1}} \mathrm{d}\zeta \frac{q^2(\zeta)}{\rho(\zeta)}. \tag{17.43}$$

It follows that

$$I_2 + J_2 = I_1 J_1. \tag{17.44}$$

Thus the determinant of the 2×2 layer matrix in (17.42) is equal to $1 + I_2 J_2$; the matrices obtained by iterating P and D to second order in δz_n have a correction to unimodularity of order $(\delta z_n)^4$. If we had stopped at first order, the determinant would have been $1 + I_1 J_1$, and the correction to unimodularity would have been of second order in δz_n.

The importance of unimodularity is the link with conservation and reciprocity laws, as we shall see shortly. Here we note that symmetrized starting values in the iteration, namely

$$P^0 = \frac{1}{2}(P_n + P_{n+1}), \quad D^0 = \frac{1}{2}(D_n + D_{n+1}), \tag{17.45}$$

improve the unimodularity. Lekner (1990a) shows that, to second order, the matrix for the layer $z_n \le z \le z_{n+1}$ is

$$M_n = \begin{pmatrix} \frac{1-I_2/2}{1+I_2/2} & \frac{I_1}{1+I_2/2} \\ \frac{-J_1}{1+J_2/2} & \frac{1-J_2/2}{1+J_2/2} \end{pmatrix}. \tag{17.46}$$

That this matrix is exactly unimodular follows from the identity (17.44).

The reflection and transmission amplitudes are found by multiplying together the layer matrices. To see this, we note that from (17.5)

$$P_1 = \mathrm{e}^{i\alpha} + r\mathrm{e}^{-i\alpha}, \quad D_1 = iQ_a\left(\mathrm{e}^{i\alpha} - r\mathrm{e}^{-i\alpha}\right), \quad \alpha = q_a z_1 = q_a a, \tag{17.47}$$

$$P_{N+1} = t\mathrm{e}^{i\beta}, \quad D_{N+1} = iQ_b t\mathrm{e}^{i\beta}, \quad \beta = q_b z_{N+1} = q_b b. \tag{17.48}$$

The values of P and D are related by the layer matrices:

$$\begin{pmatrix} P_{N+1} \\ D_{N+1} \end{pmatrix} = M_N \begin{pmatrix} P_N \\ D_N \end{pmatrix} = M_N M_{N-1} \begin{pmatrix} P_{N-1} \\ D_{N-1} \end{pmatrix} = \cdots = M \begin{pmatrix} P_1 \\ D_1 \end{pmatrix}. \tag{17.49}$$

M is the profile matrix, the ordered product of all the layer matrices:

$$M = M_N M_{N-1} \ldots M_n \ldots M_2 M_1 = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}. \tag{17.50}$$

From (17.47), (17.48) and (17.50), we have

$$\begin{pmatrix} t\mathrm{e}^{i\beta} \\ iQ_b t\mathrm{e}^{i\beta} \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} \mathrm{e}^{i\alpha} + r\mathrm{e}^{-i\alpha} \\ iQ_a(\mathrm{e}^{i\alpha} - r\mathrm{e}^{-i\alpha}) \end{pmatrix}. \tag{17.51}$$

Solving for the reflection and transmission amplitudes r and t we find

$$r = \mathrm{e}^{2i\alpha} \frac{Q_a Q_b m_{12} + m_{21} + iQ_a m_{22} - iQ_b m_{11}}{Q_a Q_b m_{12} - m_{21} + iQ_a m_{22} + iQ_b m_{11}}, \tag{17.52}$$

$$t = \mathrm{e}^{i(\alpha-\beta)} \frac{2iQ_a \det M}{Q_a Q_b m_{12} - m_{21} + iQ_a m_{22} + iQ_b m_{11}}. \tag{17.53}$$

We note the close correspondence between the exact expressions (17.20), (17.21) and the matrix expressions (17.52), (17.53). The correspondence becomes equivalence in the limit $N \to \infty$. In (17.53) $\det M = m_{11}m_{22} - m_{12}m_{21}$ is the determinant of the profile matrix; $\det M = 1$ when the matrix is unimodular.

In the absence of dissipation within any part of the system, and also excluding total internal reflection, all q and Q values are real. No absorption within the stratification also implies that all the matrix elements are real. Then the reflectance $R = |r|^2$ and transmittance $T = (Q_b/Q_a)|t|^2$ are given by

$$R = \frac{(Q_a Q_b m_{12} + m_{21})^2 + (Q_a m_{22} - Q_b m_{11})^2}{(Q_a Q_b m_{12} - m_{21})^2 + (Q_a m_{22} + Q_b m_{11})^2}, \tag{17.54}$$

$$T = \frac{4Q_a Q_b (\det M)^2}{(Q_a Q_b m_{12} - m_{21})^2 + (Q_a m_{22} + Q_b m_{11})^2}. \tag{17.55}$$

Since there is no dissipation, the incident intensity must be equal to the sum of the reflected and transmitted intensities, $R+T=1$. From the formulae (17.54) and (17.55),

$$R+T=1-\frac{4Q_aQ_b\det M(1-\det M)}{(Q_aQ_bm_{12}-m_{21})^2+(Q_am_{22}+Q_bm_{11})^2}. \tag{17.56}$$

Thus energy conservation requires $\det M=1$ or $\det M=0$. The latter is excluded, as we now show. In the case of representation by homogeneous layers, M is a product of unimodular matrices: if within $z_n\le z\le z_{n+1}$ the density $\rho=\rho_n$ and the normal component of the wavevector $q=q_n$, and $Q_n=q_n/\rho_n$, the solutions of (17.33) in this layer are

$$P(z)=P_n\cos q_n(z-z_n)+Q_n^{-1}D_n\sin q_n(z-z_n), \tag{17.57}$$

$$D(z)=D_n\cos q_n(z-z_n)-Q_nP_n\sin q_n(z-z_n). \tag{17.58}$$

Hence the homogeneous layer matrix is unimodular:

$$M_n=\begin{pmatrix}\cos\delta_n & Q_n^{-1}\sin\delta_n\\ -Q_n\sin\delta_n & \cos\delta_n\end{pmatrix}. \tag{17.59}$$

M is then the product of unimodular matrices, and $\det M=1$. Since $\det M$ is a continuous function of the matrix elements, a zero determinant is excluded in the general case.

Next we compare the reflection and transmission when the wave is incident 'from below' (from the homogeneous medium b). Equation (17.49) still holds, with the same M as before, but now

$$P_1=t'\mathrm{e}^{-i\alpha},\quad D_1=-iQ_at'\mathrm{e}^{-i\alpha},\quad \alpha=q_az_1=q_aa, \tag{17.60}$$

$$P_{N+1}=\mathrm{e}^{-i\beta}+r'\mathrm{e}^{i\beta},\quad D_{N+1}=-iQ_b(\mathrm{e}^{-i\beta}-r'\mathrm{e}^{i\beta}),\quad \beta=q_bz_{N+1}=q_bb. \tag{17.61}$$

(We use the shortened notation $r=r_{ab},t=t_{ab},r'=r_{ba},t'=t_{ba}$ in this section.) Thus (17.51) is replaced by

$$\begin{pmatrix}\mathrm{e}^{-i\beta}+r'\mathrm{e}^{i\beta}\\ -iQ_b(\mathrm{e}^{-i\beta}-r'\mathrm{e}^{i\beta})\end{pmatrix}=\begin{pmatrix}m_{11} & m_{12}\\ m_{21} & m_{22}\end{pmatrix}\begin{pmatrix}t'\mathrm{e}^{-i\alpha}\\ i-iQ_at'\mathrm{e}^{-i\alpha}\end{pmatrix}. \tag{17.62}$$

As before, we can solve for the reflection and transmission amplitudes. The results are

$$r' = \mathrm{e}^{-2i\beta}\frac{Q_aQ_bm_{12} + m_{21} - iQ_am_{22} + iQ_bm_{11}}{Q_aQ_bm_{12} - m_{21} + iQ_am_{22} + iQ_bm_{11}}, \tag{17.63}$$

$$t' = \mathrm{e}^{i(\alpha-\beta)}\frac{2iQ_b}{Q_aQ_bm_{12} - m_{21} + iQ_am_{22} + iQ_bm_{11}}. \tag{17.64}$$

The reciprocity law (17.9), here written as $Q_at' = Q_bt$, (which implies the important result that the transmittances $T = (Q_b/Q_a)|t|^2$ and $T' = (Q_a/Q_b)|t'|^2$ are equal, even if there is absorption within the stratification) is seen to be valid on comparing (17.53) with (17.64), provided $\det M = 1$. The other reciprocity law (17.14), namely $r' = -t'r^*/t'^*$, which implies that the reflectance is the same from either side, is valid only in the absence of absorption. It can be verified from the equations for r, r' and t' given above, independently of the value of $\det M$.

We have shown that unimodularity of M is necessary for energy conservation and for the reciprocity law $T' = T$. If each layer matrix is unimodular, M will be unimodular, since the determinant of a product of matrices is equal to the product of their determinants. Thus unimodularity of the layer matrix guarantees these laws, and is a desirable characteristic in any approximation scheme. We have seen an example of an approximate layer matrix which is exactly unimodular, given in (17.46). Section IV of Lekner (1990a) gives the elements of this matrix when the variation of ρ and of q within each layer is linearly approximated. The integrals I_1, J_1, I_2, J_2 are then

$$I_1 = \frac{1}{2}\delta z_n\left(\rho_n + \rho_{n+1}\right), \tag{17.65}$$

$$J_1 = \frac{1}{2}\delta z_n\left(q_n^2/\rho_n + q_{n+1}^2/\rho_{n+1}\right) = \frac{1}{2}\delta z_n\left(Q_n^2\rho_n + Q_{n+1}^2\rho_{n+1}\right), \tag{17.66}$$

$$I_2 = \frac{1}{2}I_1J_1 + (\delta z_n)^2\rho_n\rho_{n+1}(Q_n^2 - Q_{n+1}^2)/12, \tag{17.67}$$

$$J_2 = \frac{1}{2}I_1J_1 - (\delta z_n)^2\rho_n\rho_{n+1}(Q_n^2 - Q_{n+1}^2)/12. \tag{17.68}$$

Note that the identity (17.44) is satisfied. Section V of Lekner (1990a) compares seven algorithms, based on constant or linear approximation of the density and normal wavenumber component, with or without unimodularity. The unimodular matrices are numerically superior, as well as avoiding violation of the conservation and reciprocity laws.

17.3 Low-Frequency Reflection and Transmission

The low-frequency regime is attained when the dimensionless parameter $(\omega/v)\Delta z$ is small compared to unity, where $\Delta z = b - a$ is the thickness of the stratification, and v some relevant speed, such as v_a. Since the parameter $(\omega/v)\Delta z$ is equal to $2\pi\Delta z/\lambda$ the low-frequency limit is equally well characterized as the thin-layer or long-wavelength limit. We first note that as $(\omega/v)\Delta z \to 0$ the reflection and transmission amplitudes tend to the sharp-transition values

$$r_0 = \mathrm{e}^{2iq_a a}\frac{Q_a - Q_b}{Q_a + Q_b}, \quad t_0 = \mathrm{e}^{i(q_a - q_b)a}\frac{2Q_a}{Q_a + Q_b}. \tag{17.69}$$

These are obtained by matching P and $\rho^{-1}\mathrm{d}P/\mathrm{d}z$ at the boundaries with the homogeneous media a and b. As before, $Q_a = q_a/\rho_a, Q_b = q_b/\rho_b$. The fact that the equations (17.69) give the low-frequency limits of (17.20) and (17.21) is intuitively plausible: At long wavelengths the wave is mainly affected by the change in the acoustical parameters, and is not sensitive to details in the transition between the two homogeneous media.

An important question is: What are the corrections to (17.69) and to the reflectance and transmittance? It is natural to express the corrections as power series in Δz (more precisely, as power series in a dimensionless parameter such as $(\omega/v)\Delta z$):

$$r = r_0 + r_1 + r_2 + \cdots. \tag{17.70}$$

A variety of techniques for extracting r_1 and the higher-order corrections are developed in Chap. 3 and Sect. 12.4. Here we will make use of the results (17.52) and (17.53), with the matrix elements up to second order in the layer thickness given by the integrals I_1, J_1, I_2, J_2 in the matrices (17.42) and (17.46).

The corrections to the reflectance $R = |r|^2$ depend on whether there is attenuation or not. If attenuation is negligible within the stratification and in the homogeneous media, and in the absence of total internal reflection, all matrix elements and wavevectors are real, the first-order correction to $R_0 = |r_0|^2$ is zero, and to second-order

$$R = R_0 + \frac{4Q_aQ_b}{(Q_a + Q_b)^4}\left\{Q_a^2Q_b^2I_1^2 + J_1^2 - 2Q_a^2J_2 - 2Q_b^2I_2\right\}. \tag{17.71}$$

The low-frequency approximation to the reflectance, namely the expression (17.71), is shown compared to the exact reflectance in Fig. 17.2, for a stratification in which both density and sound speed vary exponentially with depth. For this stratification, all the integrals needed in (17.71) may be found analytically. The exact reflectance is calculated from the formulae of Sect. 17.5. Also shown is the high-frequency approximation of Sect. 17.4.

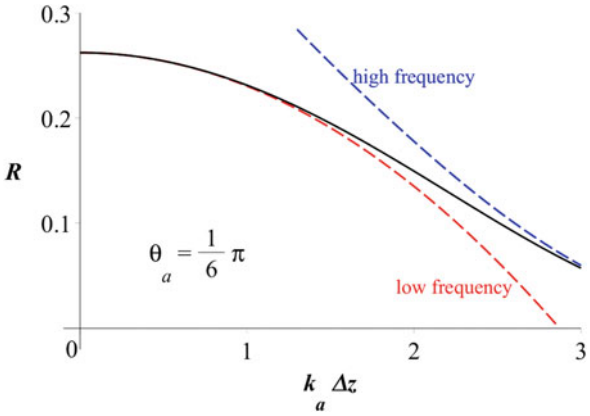


Fig. 17.2 Reflectance of a stratification in which both density and speed vary exponentially, versus the layer thickness Δz times $k_a = \omega/v_a$: exact from Sect. 17.5, low-frequency approximation (17.71) and high-frequency approximation given by (17.93). The parameters used are $\rho_b = 2\rho_a, v_b = \left(\frac{4}{3}\right)v_a$, with ρ and v continuous at $z = a$ and at $z = b = a + \Delta z$. The angle of incidence is $\theta_a = 30^\circ$

Attenuation changes the low-frequency behavior dramatically: whereas the correction to R_0 given in (17.71) is second order in the small parameter $(\omega/v)\Delta z$, attenuation makes one of the first-order contributions to the matrix elements complex: (17.41) gives

$$J_1 = \int_a^b \mathrm{d}z \frac{k^2(z) - K^2}{\rho(z)} = \int_a^b \mathrm{d}z \frac{k_r^2 - k_i^2 - K^2 + 2ik_r k_i}{\rho(z)}. \tag{17.72}$$

The wavenumber has been split into its real and imaginary parts: $\omega/v(z) \equiv k(z) = k_r(z) + ik_i(z)$. The consequence is a first-order correction to the reflectance: $|r|^2 = r_0^2 + 2\mathrm{Re}\left(r_0^* r_1\right) + \cdots$. From (17.52) we find

$$r_0 = \mathrm{e}^{2i\alpha} \frac{Q_a - Q_b}{Q_a + Q_b}, \quad r_1 = \mathrm{e}^{2i\alpha} \frac{2iQ_a}{Q_a + Q_b}(J_1 - Q_b^2 I_1). \tag{17.73}$$

The reflectivity to first order in the thickness parameter $(\omega/v)\Delta z$ is thus

$$R = R_0 - \frac{8Q_a(Q_a - Q_b)}{(Q_a + Q_b)^3} \int_a^b \mathrm{d}z \frac{k_r(z)k_i(z)}{\rho(z)} + \cdots. \tag{17.74}$$

Since k_r and k_i, the real and imaginary parts of $\omega/v(z)$ are both non-negative, so the reflectance is decreased from R_0 if $Q_a > Q_b$, and increased from R_0 if $Q_a < Q_b$. On using $Q = \omega \cos\theta/\rho v$ and the constancy of $K = \omega \sin\theta/v$ (Snell's law), we find that $Q_a > Q_b$ if

$$\tan^2\theta_a > \frac{(\rho_a v_a)^2 - (\rho_b v_b)^2}{\rho_a^2(v_b^2 - v_a^2)}. \tag{17.75}$$

For stratifications in which ρ and v increase together, the right-hand side of (17.75) is negative, and so $Q_a > Q_b$ and attenuation in the stratification decreases the reflectance of long waves from R_0 at all angles of incidence. We note in passing that the equality $Q_a = Q_b$, which makes $R_0 = 0$, requires equality in (17.75). The angle at which this happens is Green's angle θ_G, the acoustical analog of Brewster's angle in optics (Sect. 1.4, (1.61)). At Green's angle (if it exists) the first-order correction to the reflectance vanishes. This is not true of the transmittance $T = (Q_b/Q_a)|t|^2$, which from (17.53) and (17.72) is, to first order in $(\omega/v)\Delta z$,

$$T = \frac{4Q_aQ_b}{(Q_a+Q_b)^2}\left\{1 - \frac{4}{Q_a+Q_b}\int_a^b \mathrm{d}z\,\frac{k_r(z)k_i(z)}{\rho(z)} + \cdots\right\}. \tag{17.76}$$

Thus there is a first-order correction to the transmittance at all angles, and the low-frequency attenuation correction always decreases the transmittance, as expected. The second-order correction to T, in the absence of attenuation and using (17.71) and $R + T = 1$, is

$$T = T_0 - \frac{4Q_aQ_b}{(Q_a+Q_b)^4}\left\{Q_a^2Q_b^2I_1^2 + J_1^2 - 2Q_a^2J_2 - 2Q_b^2I_2\right\}. \tag{17.77}$$

The degree of attenuation required for it to dominate the low-frequency corrections may be estimated from (17.71) and (17.74) or (17.77). The first-order and second-order terms are, respectively, of magnitude $k_i\Delta z$ and $[(\omega/v)\Delta z]^2$, with k_i and v here representing average values within the stratification. Attenuation is correspondingly important in the low-frequency case unless

$$k_i \ll (\omega/v)^2\Delta z. \tag{17.78}$$

If there is attenuation in medium b, this will be important at all frequencies, manifesting itself in the formulae via a complex Q_b.

17.4 High-Frequency Limiting Forms

Reflection and transmission of high-frequency (or short-wavelength) acoustic waves is intrinsically more complicated than at low-frequencies, because short waves are sensitive to details of the stratification while long waves are influenced by average properties as expressed in integrals over the reflecting inhomogeneity. There is no universal form of the high-frequency reflectivity, in contrast to the low frequency case of the previous section. Nevertheless it is possible to give explicit formulae for the dominant terms in the reflection amplitude in some simple cases.

We first transform (17.4) by defining a new dependent variable $p = (\rho_a/\rho)^{\frac{1}{2}}P$ (Bergmann 1946). The differential equation satisfied by $p(z)$ is, with primes denoting differentiation with respect to z,

$$p'' + \left[q^2 + \frac{\rho''}{2\rho} - \frac{3}{4}\left(\frac{\rho'}{\rho}\right)^2\right]p = 0. \tag{17.79}$$

At high frequencies the q^2 term is dominant and approximate solutions of (17.79) are the Liouville-Green functions of Sect. 6.2,

$$p^{\pm}(z) = \left(\frac{q_a}{q}\right)^{\frac{1}{2}} \mathrm{e}^{\pm i\phi}, \quad \phi(z) = \int^z \mathrm{d}\zeta\, q(\zeta). \tag{17.80}$$

The phase integral $\phi(z)$ gives the accumulated phase at depth z; its derivative is $\phi' = q$. The corresponding approximations to the solutions of (17.4) are

$$P^{\pm}(z) = \left(\frac{Q_a}{Q(z)}\right)^{\frac{1}{2}} \mathrm{e}^{\pm i\phi(z)}. \tag{17.81}$$

The Liouville-Green functions $p^{\pm}(z)$ both satisfy

$$p'' + \left[q^2 + \frac{q''}{2q} - \frac{3}{4}\left(\frac{q'}{q}\right)^2\right]p = 0. \tag{17.82}$$

In the case of acoustic waves incident from medium a, the limiting forms of $p(z)$ are

$$\mathrm{e}^{iq_a z} + r\mathrm{e}^{-iq_a z} \leftarrow p(z) \rightarrow \left(\frac{\rho_a}{\rho_b}\right)^{\frac{1}{2}} t\mathrm{e}^{iq_b z}. \tag{17.83}$$

The limiting forms of p^+ are

$$\mathrm{e}^{iq_a z} \leftarrow p^+(z) \rightarrow \left(\frac{q_a}{q_b}\right)^{\frac{1}{2}} \mathrm{e}^{i(q_b z + \phi_+)}. \tag{17.84}$$

We have chosen the lower limit in the integral defining ϕ so as to make $\phi(z) \rightarrow q_a z$ as $z \rightarrow -\infty$; ϕ_+ is a constant phase. We now multiply the differential equation for p by p^+, that for p^+ by p, subtract, and integrate from $-\infty$ to $+\infty$. The result is a *comparison identity* similar to that obtained in Sect. 6.4,

$$r = \frac{1}{4iq_a} \int_{-\infty}^{\infty} \mathrm{d}z \left[\frac{q''}{q} - \frac{\rho''}{\rho} - \frac{3}{2}\left(\frac{q'}{q}\right)^2 + \frac{3}{2}\left(\frac{\rho'}{\rho}\right)^2\right] pp^+. \tag{17.85}$$

This holds for all stratifications; the reflection amplitude is given as an integral over the derivatives of $q = (\omega/v_a)\left[v_a^2/v^2 - \sin^2\theta_a\right]^{\frac{1}{2}}$ and of the density ρ. The function p^+ given by (17.80), while p is unknown. A useful approximation to r at high frequencies is obtained by replacing p by $p^+ = (q_a/q)^{\frac{1}{2}}\mathrm{e}^{i\phi}$ in (17.85):

$$r^{(1)} = \frac{1}{4i} \int_{-\infty}^{\infty} \mathrm{d}z \frac{\mathrm{e}^{2i\phi}}{q} \left[\frac{q''}{q} - \frac{\rho''}{\rho} - \frac{3}{2}\left(\frac{q'}{q}\right)^2 + \frac{3}{2}\left(\frac{\rho'}{\rho}\right)^2\right]. \tag{17.86}$$

A closely related "weak reflection" or Rayleigh approximation is (Lekner 1989)

$$r_R = -\int_{-\infty}^{\infty} \mathrm{d}z\, \mathrm{e}^{2i\phi} \frac{Q'}{2Q}. \tag{17.87}$$

This may be put in a form similar to (17.86) by changing to ϕ as integration variable, integrating by parts, and then changing back:

$$\begin{aligned} r_R &= -\frac{1}{2}\int_{-\infty}^{\infty} \mathrm{d}\phi\, \mathrm{e}^{2i\phi} Q^{-1}\frac{\mathrm{d}Q}{\mathrm{d}\phi} = \frac{1}{4i}\int_{-\infty}^{\infty} \mathrm{d}\phi\, \mathrm{e}^{2i\phi}\frac{\mathrm{d}}{\mathrm{d}\phi}\left(\frac{\mathrm{d}Q}{Q\mathrm{d}\phi}\right) = \frac{1}{4i}\int_{-\infty}^{\infty} \mathrm{d}z\, \mathrm{e}^{2i\phi}\left(\frac{Q'}{qQ}\right)' \\ &= \frac{1}{4i}\int_{-\infty}^{\infty} \mathrm{d}z \frac{\mathrm{e}^{2i\phi}}{q}\left[\frac{q''}{q} - \frac{\rho''}{\rho} - 2\left(\frac{q'}{q}\right)^2 + \left(\frac{\rho'}{\rho}\right)^2 + \frac{\rho' q'}{\rho q}\right]. \end{aligned} \tag{17.88}$$

Both approximations fail if q is zero, as happens at a classical turning point where $q^2 = 0, v^2 = v_a^2/\sin^2\theta_a$; or if q is small, as at grazing incidence, when $q_a \rightarrow 0$.

As an application of (17.86) or (17.87, 17.88), we will consider the high-frequency reflection amplitude from a stratification that is smooth except at a finite number of points z_j where there are discontinuities in the derivatives of ρ or v, or of both. Under these conditions the dominant terms in the integrands of (17.86) or (17.87, 17.88) are delta functions at z_j, arising from the second derivatives of ρ and q.

Let the density derivative ρ' change by $\Delta\rho'_j$ as z passes through z_j. This discontinuity in the derivative contributes $\Delta\rho'_j\delta(z-z_j)$ to ρ''. A discontinuity in the derivative of q gives a delta function whose strength may be calculated from

$$\frac{\mathrm{d}q}{\mathrm{d}z}=\frac{1}{2q}\frac{\mathrm{d}q^2}{\mathrm{d}z}=\frac{\omega^2}{2q}\frac{\mathrm{d}v^{-2}}{\mathrm{d}z}=-\frac{\omega^2}{qv^3}\frac{\mathrm{d}v}{\mathrm{d}z}. \tag{17.89}$$

A change $\Delta v'_j$ in the derivative of v as z passes through z_j thus contributes $-\left(\omega^2/q_jv_j^3\right)\Delta v'_j\delta(z-z_j)$ to q''. The integrand near z_j thus contains the singular delta function term originating from $q''/q^2-\rho''/\rho q$ of the form $-\sigma_j\delta(z-z_j)$. The dimensionless strength σ of the delta function is determined by the discontinuities in the derivatives of density and sound speed:

$$\sigma=\frac{\Delta\rho'}{q\rho}+\frac{\omega^2\Delta v'}{q^3v^3}. \tag{17.90}$$

The phase factor $\exp 2i\phi$ oscillates rapidly in the high-frequency limit. This ensures that smooth parts of the integrand average out to near zero, so that

$$r^{(1)}\approx\frac{i}{4}\sum_j\sigma_j\exp 2i\phi_j \tag{17.91}$$

(ϕ_j being the value of the phase integral at $z=z_j$). Since σ varies with frequency as ω^{-1}, the resulting reflectance is proportional to ω^{-2}, with oscillatory terms due to the phase factors $\exp 2i\phi_j$.

A simple and important special case is that of a stratification that is smooth except for discontinuities in the derivatives of ρ and/or v at the boundaries $z=a$ and b. The formula (17.91) gives

$$r^{(1)}\approx\frac{i}{4}(\sigma_a\exp 2i\phi_a+\sigma_b\exp 2i\phi_b), \tag{17.92}$$

with the reflectance

$$R^{(1)} \approx \frac{1}{16}\left(\sigma_a^2 + \sigma_b^2 + 2\sigma_a\sigma_b \cos 2\Delta\phi\right). \tag{17.93}$$

In (17.93) $\Delta\phi$ is the increment in phase across the stratification,

$$\Delta\phi = \phi_b - \phi_a = \int_a^b \mathrm{d}z\, q(z) = \frac{\omega}{v_a}\int_a^b \mathrm{d}z \left[\frac{v_a^2}{v^2(z)} - \sin^2\theta_a\right]^{\frac{1}{2}}. \tag{17.94}$$

Figure 17.2 showed the high-frequency reflectance (17.93) for the *exp-exp* stratification discussed in the next section, continuous in ρ and v at $z = a$ and $z = b$. For this case

$$\rho(z) = \rho_a \mathrm{e}^{(z-a)/L_\rho}, \quad v(z) = v_a \mathrm{e}^{(z-a)/L_v}. \tag{17.95}$$

The lengths L_ρ, L_v are related by the stratification thickness $\Delta z = b - a$ by

$$L_\rho = \frac{\Delta z}{\ln(\rho_b/\rho_a)}, \quad L_v = \frac{\Delta z}{\ln(v_b/v_a)}. \tag{17.96}$$

The strengths of the delta functions in this case are

$$\sigma_a = q_a^{-1}\left(L_\rho^{-1} + L_v^{-1}\sec^2\theta_a\right), \sigma_b = -q_b^{-1}\left(L_\rho^{-1} + L_v^{-1}\left[1 - \left(v_b^2/v_a^2\right)\sin^2\theta_a\right]^{-1}\right). \tag{17.97}$$

Assuming that the angle of incidence is less than the critical angle $\theta_c = \arcsin(v_a/v_b)$, so that $q(z)$ remains real, the phase increment is

$$\Delta\phi = L_v\{K[\arctan q_b/K - \arctan q_a/K] - (q_b - q_a)\}. \tag{17.98}$$

As usual, $K = (\omega/v_a)\sin\theta_a$ is the tangential component of the wavevector. At normal incidence $K \to 0$, $q_a \to \omega/v_a, q_b \to \omega/v_b$, and (17.97) and (17.98) reduce to

$$\sigma_a = \frac{v_a}{\omega\Delta z}\ln\frac{\rho_b v_b}{\rho_a v_a}, \quad \sigma_b = -\frac{v_b}{\omega\Delta z}\ln\frac{\rho_b v_b}{\rho_a v_a}. \tag{17.99}$$

$$\Delta\phi = \frac{\omega L_v}{v_a}\left[1 - \frac{v_a}{v_b}\right]. \tag{17.100}$$

Figure 17.3 compares the normal incidence exact reflectance with the high frequency limiting form (17.93), to higher values of the parameter $\omega\Delta z/v_a$ than were shown in Fig. 17.2. Note that the contributions from the discontinuities in the derivatives of ρ and v, which give the characteristic oscillatory decay with frequency, become dominant at quite moderate values of $\omega\Delta z/v_a$.

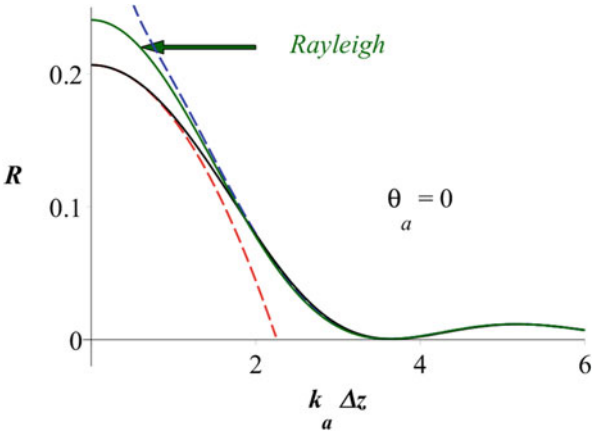


Fig. 17.3 Reflectance of an *exp-exp* stratification. The notation and parameters are the same as in Fig. 17.2, except that the results here are for normal incidence and to higher frequency, and that the Rayleigh approximation (17.101) is also shown

It is interesting to show in Fig. 17.3 the Rayleigh approximation (17.87), which correctly incorporates the high-frequency limiting reflectance, and also is good at low frequencies, provided the reflection is not too strong. From (17.87) to (17.95) we find, at normal incidence,

$$R_R = \frac{1}{4}\left(1 + \frac{L_v}{L_\rho}\right)^2 \left\{[Ci(\alpha) - Ci(\beta)]^2 + [Si(\alpha) - Si(\beta)]^2\right\}, \tag{17.101}$$

In (17.101), $\alpha = 2\omega L_v/v_a$, $\beta = 2\omega L_v/v_b$, and Ci and Si are the standard cosine and sine integrals (Temme 2010),

$$Ci(z) = \ln z + \gamma - \int_0^z \mathrm{d}t\, t^{-1}(1 - \cos t), \quad Si(z) = \int_0^z \mathrm{d}t\, t^{-1} \sin t. \tag{17.102}$$

The low-frequency limit of (17.101) is, in agreement with equation (34) of Lekner (1989),

$$R_R \to \frac{1}{4}\left[\ln \frac{\rho_b v_b}{\rho_a v_a}\right]^2. \tag{17.103}$$

This result may be obtained directly from (17.87), since in the long-wave limit the phase ϕ is nearly constant over the stratification, so at arbitrary angle of incidence

$$r_R \to \frac{1}{2}\ln \frac{Q_a}{Q_b}, \qquad R_R = |r_R|^2 \to \frac{1}{4}\left[\ln \frac{Q_a}{Q_b}\right]^2. \tag{17.104}$$

The Rayleigh approximation (17.103) differs from the exact normal incidence low-frequency limit

$$R_0 = \left(\frac{\rho_b v_b - \rho_a v_a}{\rho_b v_b + \rho_a v_a}\right)^2 \tag{17.105}$$

by an amount of fourth order in the quantity $x = (\rho_b v_b - \rho_a v_a)/(\rho_b v_b + \rho_a v_a)$: The leading term in the difference $R_R - R_0$ is $\frac{2}{3}x^4$, and the fractional difference is $R_R/R_0 - 1 = \frac{2}{3}x^2 + \cdots$. For the case shown in Fig. 17.3, the ratio $x = 5/11$, and there is an appreciable difference between R_0 and the Rayleigh approximation value given by (17.103).

17.5 Exact Solutions for the *exp-lin* and *exp-exp* Stratifications

Several variations of the phase velocity are known for which analytic solution of the acoustic pressure equation (17.4) is possible: exponential decrease of sound speed with depth (Heller 1953), linear variation of speed (Gupta 1965), and linear variation of the reciprocal of sound speed (Morris 1970). Here we will give solutions for exponential variation of density, accompanied by either linear or exponential variation of the sound speed, following Lekner (1990b). Only the case where both sound speed and density vary exponentially will be discussed in detail. For exponential density variation, with densities ρ_1 at $z = a+$ and ρ_2 at $z = b-$,

$$\rho(z) = \rho_1 \mathrm{e}^{(z-a)/L_\rho}, \quad L_\rho = \frac{\Delta z}{\ln \rho_2/\rho_1}. \tag{17.106}$$

When the density is continuous at the boundaries, $\rho_1 = \rho_a$ and $\rho_2 = \rho_b$, and we regain (17.95) and (17.96). The function $p = \rho^{-\frac{1}{2}}P$ satisfies (17.79), which for exponential variation in ρ reduces to

$$p'' + \left[q^2 - \left(2L_\rho\right)^{-2}\right]p = 0. \tag{17.107}$$

If the speed of sound $v(z)$ is linear in the depth z, $\mathrm{d}/\mathrm{d}z = (\Delta v/\Delta z)\mathrm{d}/\mathrm{d}v$, and (17.107) may be written in terms of v as independent variable:

$$\frac{\mathrm{d}^2 p}{\mathrm{d}v^2} + \left(\frac{\Delta z}{\Delta v}\right)^2 \left[\frac{\omega^2}{v^2} - K^2 - \left(2L_\rho\right)^{-2}\right]p = 0. \tag{17.108}$$

Comparison with equation 10.13.1 of Olver and Maximon (2010) shows that solutions of (17.108) are $v^{\frac{1}{2}}(z)M_\mu[sv(z)]$, where M_μ is any of the modified Bessel functions $I_{\pm\mu}, K_\mu$, the order μ and "slowness" parameter s being given by

$$\mu^2 = \frac{1}{4} - \left(\frac{\omega\Delta z}{\Delta v}\right)^2, \quad s = \left|\frac{\Delta z}{\Delta v}\right|\left[K^2 + \left(2L_\rho\right)^{-2}\right]^{\frac{1}{2}}. \tag{17.109}$$

Since the order μ changes from real to imaginary when the angular frequency increases through $\omega_0 = \frac{1}{2}|\Delta v/\Delta z|$, the two frequency ranges $(0, \omega_0)$ and (ω_0, ∞) have to be considered separately. The formulae (17.20) and (17.21) give the reflection amplitudes r and t in terms of two linearly independent solutions F and G of (17.4), of the form $(\rho v)^{\frac{1}{2}}M_\mu(sv)$.

The remainder of this section will be concerned with the *exp-exp* stratification, in which the density varies according to (17.106) and the sound speed according to

$$v(z) = v_1 \mathrm{e}^{(z-a)/L_v}, \quad L_v = \frac{\Delta z}{\ln\, v_2/v_1}. \tag{17.110}$$

It is understood, as for the density variation, that v_1 is the sound speed at $z = a+$, and v_2 at $z = b-$. Equation (17.107) then reads, on transforming to the dimensionless slowness variable $\sigma = v_1/v = \exp(a-z)/L_v$, and noting that $q^2 = \left(\omega^2/v_1^2\right)\sigma^2 - K^2$,

$$\sigma^2 \frac{\mathrm{d}^2 p}{\mathrm{d}\sigma^2} + \sigma \frac{\mathrm{d}p}{\mathrm{d}\sigma} + \left[(\omega L_v/v_1)^2\sigma^2 - (KL_v)^2 - \left(L_v/2L_\rho\right)^2\right]p = 0. \tag{17.111}$$

Two linearly independent solutions of (17.111) are the Bessel functions $J_\mu(y), Y_\mu(y)$, where

$$\mu = L_v\left[K^2 + \left(2L_\rho\right)^{-2}\right]^{\frac{1}{2}}, \quad y = y_1\sigma, \quad y_1 = \omega L_v/v_1. \tag{17.112}$$

The solutions of (17.4) may thus be taken as

$$F = \left(\frac{\rho}{\rho_1}\right)^{\frac{1}{2}} J_\mu(y), G = \left(\frac{\rho}{\rho_1}\right)^{\frac{1}{2}} Y_\mu(y), y = \frac{\omega L_v}{v_1}\mathrm{e}^{(a-z)/L_v}, \frac{\rho}{\rho_1} = \mathrm{e}^{(z-a)/L_\rho}. \tag{17.115}$$

Equations (17.20) and (17.21) then give the reflection and transmission amplitudes. As noted in Sect. 17.1, if the density ρ is continuous across both the interfaces at $z = a$ and at $z = b$, the equations linking the derivatives across the boundaries simplify, and the equations (17.20–17.23) may be replaced by a set in which $Q_a \to q_a, Q_b \to q_b$ and $\overline{F} \to F', \overline{G} \to G'$. This is the case for the profile whose reflectivity is plotted in Figs. 17.2 and 17.3.

17.6 An Upper Bound on the Acoustic Reflectivity

In Sect. 5.4 we derived bounds on the reflection of s and p polarized electromagnetic waves. For the s polarization the result is simple: when the profile dielectric function $\varepsilon(z)$ increases or decreases monotonically between its boundary values $\varepsilon_a, \varepsilon_b$, the reflectivity is always less (at the same angle of incidence) than for a sharp transition between the same values. The p polarization result is more complicated, since the effective normal wavevector component is $Q = q/\varepsilon$. The same complications arise in the reflection of acoustic waves: the effective normal wavevector component is $Q = q/\rho$, and the inequality stated above will hold when Q is monotonic.

From Sect. 5.4 or from Sect. IV of Lekner (1989), we have an identity for the reflection amplitude $r = |r|e^{i\phi}$:

$$\ln\frac{1+|r|}{1-|r|} = -\int_{-\infty}^{\infty} dz\,\frac{Q'}{Q}\cos\phi. \tag{17.116}$$

(A minus sign is missing from the corresponding equation (17) in Lekner 1989.) Let 2α be the dimensionless quantity on the right side of (17.116). Then $|r| = \tanh\alpha$ and

$$R = |r|^2 = \tanh^2\alpha \leq 1. \tag{17.117}$$

(In the absence of absorption, $Q = q/\rho$ is either real or imaginary: when $v_2 > v_1$, it is imaginary for $\theta_1 > \theta_c = \arcsin v_1/v_2$ and there is total internal reflection. Hence Q'/Q is real.)

Suppose that Q'/Q has one sign throughout the stratification, which implies that Q increases or decreases monotonically. Since the cosine of the phase ϕ is bounded above by $+1$ and below by -1, the right side of (17.116) is bounded above by $\ln Q_{\max}/Q_{\min}$, where $Q_{\max}$ and $Q_{\min}$ are the greater and lesser of Q_a and Q_b. Thus, if $Q(z)$ is monotonic,

$$\frac{1+|r|}{1-|r|} \leq \frac{Q_{\max}}{Q_{\min}}, \quad R = |r|^2 \leq \left(\frac{Q_a - Q_b}{Q_a + Q_b}\right)^2. \tag{17.118}$$

Thus a profile for which Q is monotonic (or at least does not increase or does not decrease within the profile), will not reflect more than an abrupt transition between the same two media, at the same angle of incidence, and at any frequency. (The reflectance from a step profile is independent of frequency.)

Under what circumstances is Q monotonic? Since $Q = q/\rho$,

$$Q^2(z) = \left[\frac{\omega^2}{v^2(z)} - K^2\right]\rho^{-2}(z), \quad \frac{\mathrm{d}Q^2}{\mathrm{d}z} = -\frac{2}{\rho^2}\left[q^2\frac{\mathrm{d}\ln\rho v}{\mathrm{d}z} + K^2\frac{\mathrm{d}\ln v}{\mathrm{d}z}\right]. \quad (17.119)$$

At *normal incidence* ($K = 0$), $\mathrm{d}Q^2/\mathrm{d}z$ has the sign of $-\mathrm{d}\ln\rho v/\mathrm{d}z$, so if the product ρv of speed and density increases or decreases monotonically, the reflectivity at normal incidence is never greater than that for a sharp interface.

At *general incidence*, if ρv and v both increase or both decrease monotonically, the reflectivity at any angle will be smaller than the reflectivity (at the same angle) at an abrupt change between the same bounding media. (In this case there is no Green's angle, at which $R_0 = 0$, see (1.61).) If, on the other hand, ρv increases monotonically and v decreases monotonically (or vice versa), the expression within large parentheses in (17.22) may change sign, in which case there is the possibility of greater reflection than from a sharp interface.

The *exp-exp* profile of the previous section is an example in which both ρ and v increase monotonically with the depth z, and thus the reflectivity is less than that of an abrupt transition between the boundary values of density and sound speed, provided there is no discontinuity of either at either boundary.

Another example is that of reflection from a *homogeneous layer* of thickness $\Delta z = b - a$, within which the speed and density are constant. By the methods of Sect. 2.4, the reflection amplitude is

$$r = \mathrm{e}^{2iq_a a}\frac{r_a + r_b\mathrm{e}^{2iq\Delta z}}{1 + r_a r_b\mathrm{e}^{2iq\Delta z}}, \quad r_a = \frac{Q_a - Q}{Q_a + Q}, \quad r_b = \frac{Q - Q_b}{Q + Q_b}. \quad (17.120)$$

The reflectivity is the absolute square of r; for real q_a, q_b and q, this is

$$R = |r|^2 = \frac{r_a^2 + 2r_a r_b\cos 2q\Delta z + r_b^2}{1 + 2r_a r_b\cos 2q\Delta z + r_a^2 r_b^2}. \quad (17.121)$$

For fixed frequency and angle of incidence (fixed q and Q_a, Q_b, Q), this is a periodic function of the layer thickness Δz, with period π/q (equal to $\pi v/\omega$ at normal incidence), provided there is no attenuation and q is real. The extrema of (17.121) occur when $\cos 2q\Delta z = \pm 1$; these extrema take the same form as for the electromagnetic p wave given in (2.73):

$$R_+ = \left(\frac{Q_a - Q_b}{Q_a + Q_b}\right)^2 = R_0, \quad R_- = \frac{Q_a Q_b - Q^2}{Q_a Q_b + Q^2}. \quad (17.122)$$

The theorem proved above states that if $Q(z)$ is monotonic (which includes the case of discontinuities all of the same sign), $R \le R_0$. Applied to the problem at hand, this reads that if Q lies between Q_a and Q_b, the reflectance must be no greater than R_0. The implication is that $R_+ = R_0$ is greater than R_- when the value of Q is between Q_a and Q_b. From (17.122), we find that R_+ is greater than R_- when

$$\left(Q^2 - Q_a^2\right)\left(Q^2 - Q_b^2\right) < 0, \tag{17.123}$$

which is true when Q lies between Q_a and Q_b, in agreement with the theorem.

It is interesting to compare the exact reflectivity for the homogeneous layer with the Rayleigh approximation (17.87), which rewrite as

$$r_R = -\frac{1}{2}\int_{-\infty}^{\infty} \mathrm{d}z\, \mathrm{e}^{2i\phi}\frac{\mathrm{d}}{\mathrm{d}z}\ln Q. \tag{17.124}$$

At z_a and z_b there are discontinuities in $\ln Q$, by the amounts $\ln Q/Q_a$, and $\ln Q_b/Q$. The derivative of a step function is a delta function of strength equal to the magnitude of the step, so for the homogeneous layer the Rayleigh approximation reflection amplitude and reflectance are

$$r_R = -\frac{1}{2}\left\{\mathrm{e}^{2i\phi_a}\ln Q/Q_a + \mathrm{e}^{2i\phi_b}\ln Q_b/Q\right\}, \tag{17.125}$$

$$R_R = \frac{1}{4}\left\{\ln^2 Q/Q_a + \ln^2 Q_b/Q + 2(\ln Q/Q_a)(\ln Q_b/Q)\cos 2q\Delta z\right\}. \tag{17.126}$$

(The argument of the cosine follows from $\phi_b - \phi_a = q\Delta z$.) Figure 17.4 compares the normal incidence exact and approximate reflectivities, as a function of the thickness of the homogenous layer. The parameters are chosen to approximate a layer of sediment on a seafloor or lake bottom (Hamilton 1980). We note that the Rayleigh reflectivity is most accurate where the reflection is smallest: in Sect. 5.7 we also used the alternative name *weak reflection approximation*.

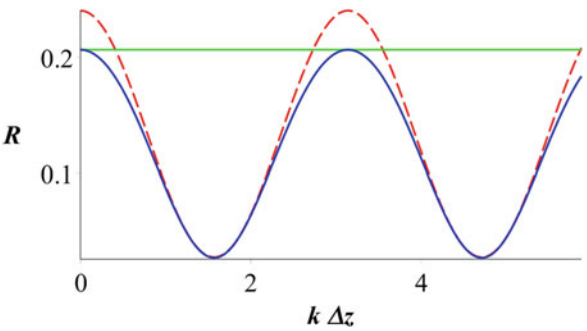


Fig. 17.4 Normal incidence reflectivity from a homogenous layer, as a function of $k\Delta z = \frac{\omega}{v}\Delta z$. The *solid curve* is the exact reflectivity (17.121); the *dashed curve* is the Rayleigh approximation (17.126). The density values used (in g/cm^3) are $\rho_a = 1, \rho = 1.7, \rho_b = 2$; the corresponding sound speeds (in km/s) are $v_a = 1.5, v = 1.7, v_b = 2$. The *horizontal line* shows the upper bound $R_0 = R_+$ of (17.118)

17.7 Profiles with Discontinuities in Density or Sound Speed

The exact results (17.20) and (17.21) for the reflection and transmission amplitudes are general enough to include discontinuities in density and/or sound speed, or in their derivatives with respect to depth z. The previous section provided one example (the homogeneous layer), and compared the exact and the Rayleigh approximation reflectivities. Discontinuities dominate the reflection process, and a perturbation theory or the Rayleigh approximation do not provide an adequate starting point. An alternative formulation was given in Lekner (1990c), which leads to results that are *exact in the high-frequency limit*, and also exact for the homogeneous layer at all frequencies. At low frequencies the formulae derived here fail (except for the homogeneous layer), but there the limiting forms derived in Sect. 17.3 can be used. The problem being discussed is shown schematically in Fig. 17.5. Note that we now need to distinguish the densities and speeds just inside the stratification from those just outside, in the homogeneous media a and b above and below. The densities in the homogeneous media a and b are denoted by ρ_a, ρ_b, those just inside the stratification at $z = a+, z = b-$ by ρ_1, ρ_2, and likewise for the sound speeds v.

In special cases one can find analytic solutions of the equation (17.4) for the acoustic pressure, and then the reflection and transmission amplitudes r and t can be found exactly from (17.20) to (17.21), modified to allow for the discontinuities in density or sound speed at the boundaries of the inhomogeneous layer:

$$r = \mathrm{e}^{2i\alpha}\frac{Q_aQ_b(F,G) + iQ_a\big(F,\overline{G}\big) + iQ_b\big(\overline{F},G\big) - \big(\overline{F},\overline{G}\big)}{Q_aQ_b(F,G) + iQ_a\big(F,\overline{G}\big) - iQ_b\big(\overline{F},G\big) + \big(\overline{F},\overline{G}\big)}, \tag{17.127}$$

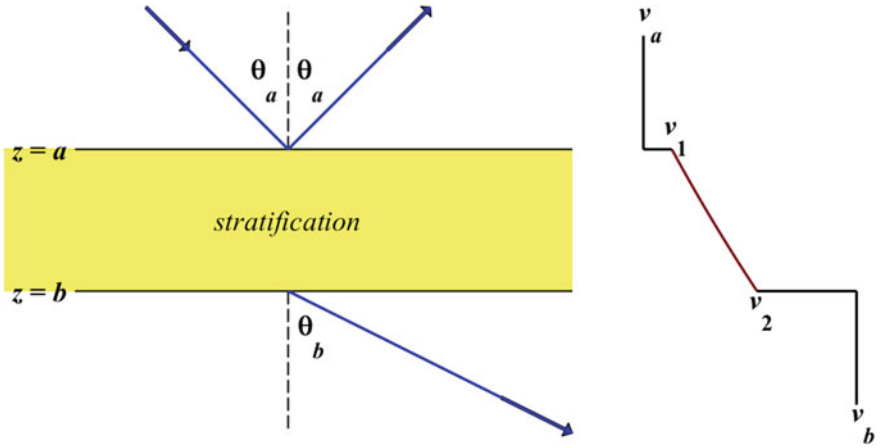


Fig. 17.5 Schematics of reflection and transmission by an inhomogeneous stratification with discontinuities in density ρ and sound speed v at its boundaries. Only the sound-speed profile is shown

$$t = \mathrm{e}^{i(\alpha-\beta)} \frac{2iQ_a\left(F_2\overline{G_2} - \overline{F_2}G_2\right)}{Q_aQ_b(F,G) + iQ_a\left(F,\overline{G}\right) - iQ_b\left(\overline{F},G\right) + \left(\overline{F},\overline{G}\right)}, \tag{17.128}$$

In (17.127–17.128), $\alpha = q_a a, \beta = q_b b$ as before, but now $(F,G) \equiv F_1G_2 - G_1F_2$, $\left(F,\overline{G}\right) \equiv F_1\overline{G_2} - G_1\overline{F_2}$, et cetera.

We are interested in getting approximate r and t for any discontinuous stratification of the type shown in Fig. 17.5. To this end we approximate the solutions F and G by the Liouville-Green waveforms (17.81), except that we now need to insert the values of density and speed just inside the stratification:

$$F(z) \approx \left(\frac{Q_1}{Q(z)}\right)^{\frac{1}{2}} \mathrm{e}^{i\phi(z)}, \quad G(z) \approx \left(\frac{Q_2}{Q(z)}\right)^{\frac{1}{2}} \mathrm{e}^{-i\phi(z)}. \tag{17.129}$$

The phase integral $\phi(z)$ is the accumulated phase at z defined in (17.80), and $Q(z) = q(z)/\rho(z), Q_1 = q_1/\rho_1, \quad Q_2 = q_2/\rho_2$.

The resulting approximate values of (F,G) to $\left(\overline{F},\overline{G}\right)$ are, with $s = \sin\Delta\phi, c = \cos\Delta\phi$,

$$\begin{aligned} &(F,G) \approx -2is, \quad \left(F,\overline{G}\right) \approx iQ_2(-2c+\gamma_2 s), \quad \left(\overline{F},G\right) \approx iQ_1(2c+\gamma_1 s), \\ &\left(\overline{F},\overline{G}\right) \approx iQ_1Q_2\left(-2s + (\gamma_1-\gamma_2)c - \frac{1}{2}\gamma_1\gamma_2 s\right). \end{aligned} \tag{17.130}$$

In (17.130) $\Delta\phi$ is the phase increment on going through the stratification from a to b:

$$\Delta\phi = \int_a^b \mathrm{d}z\, q(z) = \frac{\omega}{v_a}\int_a^b \mathrm{d}z \left[\frac{v_a^2}{v^2(z)} - \sin^2\theta_a\right]^{\frac{1}{2}}, \tag{17.131}$$

and γ_1, γ_2 are the (internal) boundary values of the dimensionless function

$$\gamma(z) = \frac{\mathrm{d}Q/\mathrm{d}z}{qQ} = q^{-1}\left(\frac{1}{q}\frac{\mathrm{d}q}{\mathrm{d}z} - \frac{1}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}z}\right). \tag{17.132}$$

The function $\gamma(z)$, and its derivative divided by q, should be small throughout the stratification if the Liouville-Green functions (17.129) are to be good approximations to the exact solutions of (17.4), since from (17.129) and (17.132) we see that F and G satisfy the equation

$$\rho\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{\rho}\frac{\mathrm{d}F}{\mathrm{d}z}\right) + q^2\left(1 + \frac{1}{2q}\frac{\mathrm{d}\gamma}{\mathrm{d}z} + \frac{\gamma^2}{4}\right)F. \tag{17.133}$$

(The exact acoustic pressure satisfies (17.4), which is (17.133) with γ set to zero.) Thus the approximations based on the Liouville-Green waveforms generally fail at low frequencies (γ is proportional to ω^{-1}) and also whenever q is small, as happens at grazing incidence, and at classical turning points (zeros of q). We note that $q^2(z) = \omega^2/v_a^2[v_a^2/v^2(z) - \sin^2\theta_a]$ stays positive, and classical turning points will not occur, if $v(z) < v_a/\sin\theta_a$. This inequality excludes both total reflection, which occurs for $\sin\theta_a > v_a/v_b$, and the possibility of 'tunneling' through a region of negative q^2 but with $v_b < v_a/\sin\theta_a$.

From (17.130), we find the reflection and transmission amplitudes on substituting into (17.127) and (17.128). We expand these in powers of γ:

$$r = r_0 + r_1 + \cdots, \quad t = t_0 + t_1 + \cdots. \tag{17.134}$$

The zeroth-order amplitudes are, again with $s = \sin\Delta\phi, c = \cos\Delta\phi$,

$$r_0 = \mathrm{e}^{2i\alpha}\frac{(Q_aQ_2 - Q_bQ_1)c - i(Q_aQ_b - Q_1Q_2)s}{(Q_aQ_2 + Q_bQ_1)c - i(Q_aQ_b + Q_1Q_2)s}, \tag{17.135}$$

$$t_0 = \mathrm{e}^{i(\alpha-\beta)}\frac{2Q_a(Q_1Q_2)^{\frac{1}{2}}}{(Q_aQ_2 + Q_bQ_1)c - i(Q_aQ_b + Q_1Q_2)s}. \tag{17.136}$$

When $Q_1 = Q_2 = Q$, these reduce to the homogeneous layer values (r_0 is given in different form in (17.120))

$$r_0 = \mathrm{e}^{2i\alpha}\frac{Q(Q_a - Q_b)c - i(Q_aQ_b - Q^2)s}{Q(Q_a + Q_b)c - i(Q_aQ_b + Q^2)s}, \tag{17.137}$$

$$t_0 = \mathrm{e}^{i(\alpha-\beta)}\frac{2Q_aQ}{Q(Q_a + Q_b)c - i(Q_aQ_b + Q^2)s}. \tag{17.138}$$

Note that an inhomogeneous layer could have $Q_1 = Q_2$; to zeroth order in γ, the approximation used here will give the same reflection and transmission amplitudes as a homogeneous layer, but a correction appears in the first-order terms.

The contributions of first order in γ to the reflection and transmission amplitudes are

$$r_1 = \mathrm{e}^{2i\alpha}\frac{iQ_aQ_1[Q_2^2(\gamma_2 - \gamma_1)c^2 + (Q_b^2\gamma_1 + Q_2^2\gamma_2)s^2 + 2iQ_bQ_2\gamma_1cs]}{[(Q_aQ_2 + Q_bQ_1)c - i(Q_aQ_b + Q_1Q_2)s]^2}, \tag{17.139}$$

$$t_1 = \mathrm{e}^{i(\alpha-\beta)}\frac{Q_a(Q_1Q_2)^{\frac{1}{2}}[(Q_aQ_2\gamma_2 + Q_bQ_1\gamma_1)s - iQ_1Q_2(\gamma_1 - \gamma_2)c]}{[(Q_aQ_2 + Q_bQ_1)c - i(Q_aQ_b + Q_1Q_2)s]^2}. \tag{17.140}$$

When there is no discontinuity in Q at either boundary ($Q_1 = Q_a$ and $Q_2 = Q_b$), r_0 is zero and t_0 reduces to the perfect transmission value $\mathrm{e}^{i(\alpha-\beta)}(Q_a/Q_b)^{\frac{1}{2}}\mathrm{e}^{i\Delta\phi}$, while r_1 takes the value

$$r_1 \to \mathrm{e}^{2i\alpha} i(\gamma_2 \mathrm{e}^{2i\Delta\phi} - \gamma_1)/4, \tag{17.141}$$

which is equivalent to the result (17.92) derived earlier.

The above theory is based on the assumption that γ and its derivative $\mathrm{d}\gamma/\mathrm{d}\phi = q^{-1}\mathrm{d}\gamma/\mathrm{d}z$ are both small. The approximations thus fail at low frequencies (except for the homogeneous layer, for which γ is identically zero). There we have the long-wavelength expansions derived in Sect. 17.3. In particular, the reflectance in the absence of absorption is given by (17.71) in terms of the integrals I_1, J_1, I_2, J_2, which are explicitly written down for the *exp-exp* profile in (17.24) of Lekner (1990c).

We shall compare the results of this section with the exact solution given in Sect. 17.5 for the *exp-exp* profile, in which both density and sound speed vary exponentially within the stratification, according to (17.106) and (17.110). For the high-frequency expressions (17.137–17.140) we need γ_1, γ_2 and $\Delta\phi$. From the defining relation (17.132), and $q^2(z) = \omega^2/v^2(z) - K^2$, we have

$$\gamma(z) = -q^{-1}\left[\left(\frac{\omega}{vq}\right)^2 \frac{1}{v}\frac{\mathrm{d}v}{\mathrm{d}z} + \frac{1}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}z}\right]. \tag{17.142}$$

Into this general expression, we insert the *exp-exp* profile values

$$\frac{1}{v}\frac{\mathrm{d}v}{\mathrm{d}z} = \frac{1}{L_v}, \quad \frac{1}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}z} = \frac{1}{L_\rho}, \tag{17.143}$$

and then obtain γ_1 and γ_2 by substituting the values q_1 and q_2 for q and v_1 and v_2 for v. The phase increment $\Delta\phi$ across the stratification, assuming no absorption and angle of incidence less than $\arcsin(v_a/v_{\max})$ so that q remains real, can be found analytically for the *exp-exp* profile (compare (17.98), which holds for profile continuous in ρ and v with the bounding media):

$$\Delta\phi = L_v\{K[\arctan q_2/K - \arctan q_1/K] + q_1 - q_2\} = \omega L_v\left(\frac{1}{v_1} - \frac{1}{v_2}\right) + O(K^2). \tag{17.144}$$

The high-frequency reflectivity is calculated from (17.137) to (17.139) as

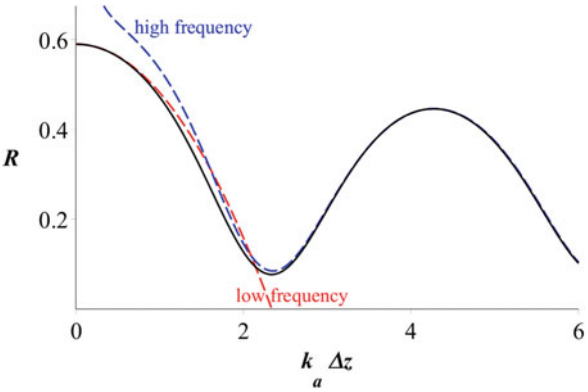


Fig. 17.6 Normal incidence reflectivities for an *exp-exp* stratification, with discontinuities at its boundaries, as a function of frequency. The dimensionless parameter $k_a \Delta z = \omega \Delta z / v_a$ equals 2π times the thickness of the stratification divided by the wavelength in medium a. The *solid curve* is the exact reflectivity, the *dashed curves* are the low- and high-frequency approximations, as indicated. The parameters used are $\rho_a = 1, \rho_1 = 1.5, \rho_2 = 1.7, \rho_b = 2.2$ (g cm^{-3}), and $v_a = 1.5, v_1 = 1.7, v_2 = 2.3, v_b = 5.2$ (km s^{-1})

$$R_{hf} = |r_0 + r_1|^2. \tag{17.145}$$

Figure 17.6 shows the reflectivity from a model *exp-exp* stratification, with acoustic parameters chosen to correspond to the Tufts abyssal plain, as presented by Chapman (1980). The exact reflectivity is obtained from the results of Sect. 17.5, the low-frequency curve from (17.71), and the high-frequency curve from (17.137), (17.139), (17.142), (17.143) and (17.145).

We see that the low-frequency approximation is good up to about $\omega \Delta z / v_a = 1 (\lambda_a \geq 6 \Delta z)$, while the high-frequency results are good from about $\omega \Delta z / v_a = 2 (\lambda_a \leq 3 \Delta z)$. In the intermediate region, the errors can be 20 % or more, and it may be necessary to use the matrix methods of Sect. 17.2 and Lekner (1990a).

Appendix: Universal Properties of Acoustic Pulses and Beams

The results of this Appendix are restricted to pulse and beam propagation within isotropic homogeneous media, within which the acoustic pressure satisfies the wave equation ((17.1) of this chapter with $\nabla \rho = 0$). We also neglect dissipation of energy or momentum, due either to viscous damping, or to scattering by impurities or bubbles.

We begin with a summary of the existing known universal properties of localized acoustic *pulses*, namely (i) the time invariance of the total energy E, momentum P_z and angular momentum J_z of the pulse, and (ii) the inequality $cP_z < E$. (In this Appendix pulse propagation is along the z-direction, and the speed of sound is c, a constant.)

The conservation of energy, momentum and angular momentum in the absence of dissipation is as expected, but the inequality $cP_z < E$ is in contrast to the sound quantum 'phonon', for which the momentum is unidirectional, and $cP = E$. The implication of the inequality $cP_z < E$ seems to be that we cannot model the phonon by any localized pulse wave-function satisfying the wave equation.

Peierls (1983) considered the energy and momentum of localized sound pulses. However, in calculating the energy and momentum, Peierls made approximations that removed the transverse localization, and in the long-wave limit his equation (3.12) gives equality of energy and c times momentum. Lekner (2006a) examined the energy and momentum densities $e(\mathbf{r},t), \mathbf{p}(\mathbf{r},t)$ of a three-dimensionally localized sound pulse, and the total energy and total momentum

$$E = \int \mathrm{d}^3 r\, \mathrm{e}(\mathbf{r},t), \quad \mathbf{P} = \int \mathrm{d}^3 r\, \mathbf{p}(\mathbf{r},t). \tag{17.146}$$

He showed that E and $\mathbf{P}$ are independent of time (are conserved, as expected), and further, that $cP_z < E$ for predominant propagation in the z-direction. Localized pulses are always converging to or diverging from their focal region, hence there is a transverse momentum density, integrating to zero. This is the reason for the inequality, and the prime distinction between pulses and phonons. A consequence of this universal convergence toward or divergence from the focal region is that the pulse pressure gradient, density gradient and particle velocity are not purely longitudinal, as they are for pulses localized in only one dimension. Exact solutions of the wave equation are used to construct specific examples of localized pulses in Lekner (2006b).

We may expect that the total angular momentum of a localized sound pulse should also be constant in time, again in the absence of dissipation. This is indeed true: the angular momentum density $\mathbf{j} = \mathbf{r} \times \mathbf{p}$ integrates to give the total angular momentum

$$\mathbf{J} = \int \mathrm{d}^3 r\, \mathbf{j}(\mathbf{r},t) = \int \mathrm{d}^3 r\, \mathbf{r} \times \mathbf{p}(\mathbf{r},t). \tag{17.147}$$

In the translation of the coordinate system, $\mathbf{r} \to \mathbf{r} - \mathbf{a}$, the total angular momentum transforms to $\mathbf{J} \to \mathbf{J} - \mathbf{a} \times \mathbf{P}$. Thus the component of $\mathbf{J}$ parallel to $\mathbf{P}$ is invariant to the choice of origin. As before, we take direction of $\mathbf{P}$ to define the z-axis; then J_z is the invariant component of interest. Lekner (2006c) has shown that the time derivative of J_z is zero, and gives examples of exact localized solutions of the wave

equation, with analytic expressions for the energy, momentum, and angular momentum.

There are analogous conserved quantities for acoustic *beams*, again in the absence of dissipation. These follow from the conservation of matter, momentum and angular momentum:

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{v}) = 0, \tag{17.148}$$

$$\partial_t (\rho v_i) + \sum_k \partial_k \Pi_{ki} = 0 \quad (i, k = x, y, z), \tag{17.149}$$

$$\partial_t j_i + \sum_k \partial_k \Lambda_{ki} = 0 \quad (i, k = x, y, z). \tag{17.150}$$

The tensors Π, Λ are the momentum flux density and angular momentum flux density tensors (Lekner 2007; Zhang and Marston 2011). Equation (17.148) may be written as $\partial_t \rho + \nabla \cdot \mathbf{p} = 0$, with ρ the mass density and $\mathbf{p} = \rho \mathbf{v}$ the momentum density.

For monochromatic acoustic beams of angular frequency ω the motion everywhere within the sound beam is periodic with period $T = 2\pi/\omega$, and the cycle average of (17.148) gives $\nabla \cdot \overline{\mathbf{p}} = 0$, where the bar denotes average over one or any number of periods, at a fixed point in space. Suppose that the acoustic beam is propagating in the z direction. Integration of $\nabla \cdot \overline{\mathbf{p}} = 0$ over the transverse directions x and y then gives (Lekner 2007)

$$\partial_z \int \mathrm{d}^2 r \, \overline{p_z} = \partial_z P'_z = 0 \quad (\mathrm{d}^2 r = \mathrm{d}x\mathrm{d}y). \tag{17.151}$$

The meaning of (17.151) is that the longitudinal cycle-averaged momentum content within a transverse slice of the beam is constant along the beam: $P'_z = \int \mathrm{d}^2 r \, \overline{p_z}$ is an *invariant*. Note that it was conservation of matter which led to the momentum content beam invariant, not conservation of momentum.

Each component of the conservation of momentum, equation (17.149), leads to another invariant on cycle averaging. The conservation of angular momentum leads to three more. Thus the conservation laws give *seven universal beam invariants*, just as in the electromagnetic case (Sect. 20.1). Perhaps surprisingly, the cycle-averaged energy content in a transverse slice of the beam, $E' = \int \mathrm{d}^2 r \, \overline{\mathrm{e}}$, is not an invariant in general. Neither is the angular momentum content, but both are

constant along the length of the beam for a special class of generalized Bessel acoustic beams (Lekner 2006d). For generalized Bessel beams one finds $cP'_z \leq E'$, and, for beams with azimuthal dependence $e^{im\phi}$, a proportionality between the energy and angular momentum contents per unit length of the beam. Beams with azimuthal dependence $e^{im\phi}$ are termed *acoustic vortex beams* by Zhang and Marston (2011).

References

Bergmann G (1946) The wave equation in a medium with a variable index of refraction. J Acoust Soc Am 17:329–333

Brekhovskikh LM (1960) Waves in layered media. Academic Press, New York

Chapman NR (1980) Low frequency bottom reflectivity measurements in the Tufts abyssal plain. In: Kuperman W, Jensen FB (eds) Bottom-interacting ocean acoustics. Plenum, New York, pp 193–207

Gupta R (1965) "Reflection of plane waves from a linear transition layer in liquid media. Geophysics 30:122–132

Hamilton EL (1980) Geoacoustic modeling of the seafloor. J Acoust Soc Am 68:1313–1340

Heller GS (1953) Reflection of acoustic waves from an inhomogeneous fluid medium. J Acoust Soc Am 25:1104–1106

Lekner J (1989) An upper bound on acoustic reflectivity; and the Rayleigh approximation. J Acoust Soc Am 86:2359–2362

Lekner J (1990a) Matrix methods in reflection and transmission of compressional waves by stratified media. J Acoust Soc Am 87:2319–2324

Lekner J (1990b) Reflection and transmission of compressional waves: some exact results. J Acoust Soc Am 87:2325–2331

Lekner J (1990c) Reflection and transmission of compressional waves by a stratification with discontinuities in density and/or sound speed. J Acoust Soc Am 88:2876–2879

Lekner J (2006a) Energy and momentum of sound pulses. Phys A 363:217–225

Lekner J (2006b) Localized oscillatory acoustic pulses. J Phys: Condens Matter 18:3031–3036

Lekner J (2006c) Angular momentum of sound pulses. J Phys: Condens Matter 18:6149–6158

Lekner J (2006d) Acoustic beams with angular momentum. J Acoust Soc Am 120:3475–3478

Lekner J (2007) Acoustic beam invariants. Phys Rev E 75:036610 (6 pp)

Morris HE (1970) Bottom-reflection-loss model with a velocity gradient. J Acoust Soc Am 48:1198–1202

Olver FWJ, Maximon LC (2010) In: Olver FWJ et al (eds) Chapter 10 in NIST handbook of mathematical functions. NIST and Cambridge

Peierls R (1983) The momentum of a sound pulse in a slightly dispersive medium. Proc Roy Soc A 390:1–12

Temme NM (2010) In: Olver FWJ et al (eds) Chapter 6 in NIST handbook of mathematical functions. NIST and Cambridge

Zhang L, Marston PL (2011) Angular momentum flux of nonparaxial acoustic vortex beams and torques on axisymmetric objects. Phys Rev E 84:06501(5 pp)

Chapter 18
Chiral Isotropic Media

Optical activity is the ability of some crystals, liquids and gases to rotate the plane of polarization of light. Optical activity, or rotatory power, is caused by chirality, either of the molecules making up the substance, or in the helical arrangement of the atomic or molecular constituents in a crystal. (A chiral object is one which cannot be superimposed on its mirror image.) In 1811 Arago found that a plate of quartz produced effects on light polarized by reflection from a pile of glass plates which are now understood to arise from the rotation plane of polarization of the light. In five memoirs presented to the Académie des Sciences from 1812 to 1837, Biot showed that the rotatory power is proportional to the thickness of the quartz plates (propagation is along the optic axis of the crystals), that the rotation depends on the wavelength, approximately as λ^{-2}, and that optical activity appears in liquids and gases, as well as in crystals. Fresnel conjectured in 1822 that on entering an optically active medium light is split into two beams of opposite circular polarization which travel with different phase velocities. In 1848 Pasteur demonstrated that the optical activity of a tartrate solution is related to the form that the crystals of the tartrate take: crystals of opposite handedness dissolve to give solutions with opposite rotatory power. References to these early works and further details may be found in the thorough historical account given by Lowry (1935) in his book on *Optical Rotatory Power*. Other historical outlines may be found in Silverman (1993) and Lindell et al. (1994); a selection of papers on natural optical activity has been compiled by Lahktakia (1990). Modern electromagnetism begins with Maxwell and the electromagnetic theory of light, but although he considered the propagation of light in crystals (Maxwell 1891/1954, Article 794), he did not treat chiral media.

This chapter deals only with *isotropic* chiral media (chiral liquids, or cubic chiral crystals). It is based largely on the work of Silverman and collaborators, and extensions given by Lekner (1996).

J. Lekner, *Theory of Reflection*, Springer Series on Atomic, Optical, and Plasma Physics 87, DOI 10.1007/978-3-319-23627-8_18

18.1 Constitutive Relations

The propagation of light in isotropic non-chiral media is describable in terms of a dielectric function ε and a magnetic permeability μ which relate the fields $\mathbf{D}$ to $\mathbf{E}$ and $\mathbf{B}$ to $\mathbf{H}$ via $\mathbf{D} = \varepsilon\mathbf{E}$ and $\mathbf{B} = \mu\mathbf{H}$. The curl equations of Maxwell in non-chiral media are

$$c\nabla \times \mathbf{E} = -\partial\mathbf{B}/\partial t, \quad c\nabla \times \mathbf{H} = \partial\mathbf{D}/\partial t. \tag{18.1}$$

All research seen by the author agrees that these are to be retained in chiral media. All researchers do not agree on the constitutive relations in chiral media, namely on what replaces $\mathbf{D} = \varepsilon\mathbf{E}$ and $\mathbf{B} = \mu\mathbf{H}$. The results given in this chapter are based on the symmetrized Condon set (Condon 1937)

$$\mathbf{D} = \varepsilon\mathbf{E} - g\partial\mathbf{H}/\partial t, \quad \mathbf{B} = \mu\mathbf{H} + g\partial\mathbf{E}/\partial t, \tag{18.2}$$

as advocated by Silverman (1986). (Condon omitted the μ; relations with it included are called *symmetrized*). Silverman has shown that the constitutive relations in Born's *Optik* (1972), namely

$$\mathbf{D} = \varepsilon_B\mathbf{E} + g_B\nabla \times \mathbf{E}, \quad \mathbf{B} = \mu_B\mathbf{H}, \tag{18.3}$$

lead to reflectances in excess of unity in the vicinity of critical angles. Another choice, known as the Drude–Born–Fedorov relations, is discussed in Chap. 3 of Latkhtakia et al. (1989) and in Sect. 1.2.1 of Lindell et al. (1994):

$$\mathbf{D} = \varepsilon_{\mathrm{DBF}}(\mathbf{E} + b\nabla \times \mathbf{E}), \quad \mathbf{B} = \mu_{\mathrm{DBF}}(\mathbf{H} + b\nabla \times \mathbf{H}). \tag{18.4}$$

From the curl equations (18.1) it is clear that (18.2) and (18.4) are equivalent to first order in g and b.

For monochromatic waves in which the fields have a time dependence given by the factor $\exp(-i\omega t)$, the relations (18.2) become, on introducing the chiral index $\gamma = \omega g$,

$$\mathbf{D} = \varepsilon\mathbf{E} + i\gamma\mathbf{H}, \quad \mathbf{B} = \mu\mathbf{H} - i\gamma\mathbf{E}. \tag{18.5}$$

The Drude–Born–Fedorov relations become, with the use of (18.1) and on setting $\chi = \omega b/c$,

$$\mathbf{D} = \varepsilon_{\mathrm{DBF}}(\mathbf{E} + i\chi\mathbf{B}), \quad \mathbf{B} = \mu_{\mathrm{DBF}}(\mathbf{H} - i\chi\mathbf{D}). \tag{18.6}$$

These relations are equivalent to (18.5) if the dielectric constants and permeabilities differ by a term second order in the chiral index γ:

$$\varepsilon_{\mathrm{DBF}} = \varepsilon - \gamma^2/\mu, \qquad \mu_{\mathrm{DBF}} = \mu - \gamma^2/\varepsilon, \qquad \chi = \gamma/(\varepsilon\mu - \gamma^2). \tag{18.7}$$

The inverse relations are (Lakhtakia et al. 1988; Lindel et al. 1994)

$$\varepsilon = \varepsilon_{\mathrm{DBF}}/\left(1 - \varepsilon_{\mathrm{DBF}}\mu_{\mathrm{DBF}}\chi^2\right), \quad \mu = \mu_{\mathrm{DBF}}/\left(1 - \varepsilon_{\mathrm{DBF}}\mu_{\mathrm{DBF}}\chi^2\right), \tag{18.8}$$

$$\gamma = \varepsilon_{\mathrm{DBF}}\mu_{\mathrm{DBF}}\chi/(1 - \varepsilon_{\mathrm{DBF}}\mu_{\mathrm{DBF}}\chi^2).$$

Bassiri et al. (1988) assume a harmonic time dependence, and use the relations

$$\mathbf{D} = \varepsilon_{\mathrm{BPE}}\mathbf{E} + i\xi\mathbf{B}, \quad \mathbf{H} = i\xi\mathbf{E} + \mathbf{B}/\mu_{\mathrm{BPE}}. \tag{18.9}$$

This form was deduced by them from the work of Jaggard et al. (1979), who calculated the properties of a medium composed of short wire helices. (The effect of the scattered fields of the helices on each other was neglected.)

The Born relations (18.3) can be eliminated on physical grounds: they predict reflectances in excess of unity in the vicinity of critical angles, and also a difference in the normal incidence reflectance of the two circular polarizations which is first order in the chirality parameter g_B, and in disagreement with experiment (Silverman et al. 1988; Lukyanov and Novikov 1990).

As we shall see in Sect. 18.4, the relations (18.5) lead to normal incidence reflectances from an achiral–chiral interface which are independent of the chiral index γ, while the relations (18.9) give reflectances which contain terms of second order in the chiral parameter ξ. In Sect. 18.5 it is shown that the chiral index γ is related to δ, the rotation of the plane of polarization on passing normally through a chiral plate of thickness d, by

$$\gamma = \lambda\delta/2\pi d. \tag{18.10}$$

The specific rotation δ/d for AgGaS is large, for example, 0.95° per μm at $\lambda = 0.485\,\mu\mathrm{m}$, yet even this relatively large value gives $\gamma \approx 1.28 \times 10^{-3}$. The differential reflectance measurement reported by Lukyanov and Novikov (1990) was on α-$LiIO_3$ crystals cut normal to the optic axis, with $\delta/d = 86.8^\circ$ per mm at $\lambda = 0.63\,\mu\mathrm{m}$ and $\gamma \approx 1.52 \times 10^{-4}$. No difference in the normal incidence reflectance of the two circular polarizations was detected to within 10^{-7}, but an effect of order γ^2 is smaller than this. Thus experiment does not yet rule out or confirm normal incidence differential reflectances which are of second order in the chirality index.

The specific rotation δ/d which follows from (18.9) is related to ξ via

$$\xi\mu_{\mathrm{BPE}} = \lambda\delta/2\pi d. \tag{18.11}$$

(Bassiri et al. 1988, (18.72)). Comparison of (18.5) and (18.9) identifies $\xi\mu_{\mathrm{BPE}}$ with γ, and substitution of $\xi = \gamma/\mu_{\mathrm{BPE}}$ into (18.9) gives

$$\mathbf{D} = (\varepsilon_{\mathrm{BPE}} + \gamma^2/\mu_{\mathrm{BPE}})\mathbf{E} + i\gamma\mathbf{H}, \quad \mathbf{B} = \mu_{\mathrm{BPE}}\mathbf{H} - i\gamma\mathbf{E}. \tag{18.12}$$

Thus (18.5) and (18.9) are in agreement if

$$\mu_{\mathrm{BPE}} = \mu, \quad \varepsilon_{\mathrm{BPE}} = \varepsilon - \gamma^2/\mu, \quad \xi = \gamma/\mu, \tag{18.13}$$

that is if the Bassiri, Papas and Engheta dielectric function is made to depend on the square of the chirality index γ . If indeed we set $\varepsilon_{\mathrm{BPE}} = \varepsilon - \gamma^2/\mu$ in the formulae of Bassiri et al. (1988), we find that for the achiral–chiral interface the normal incidence reflection and transmission amplitudes become independent of γ.

We henceforth adopt the constitutive relations (18.2) and (18.5) advocated by Silverman (1986), with ε and μ independent of the chiral index γ. This is also the choice made in the monograph by Lindell et al. (1994), and is consistent with experiment: Silverman et al. (1988) have used optical phase modulation to measure chiral asymmetries in specular reflection from a gyrotropic medium, and have found agreement with the reflection amplitudes calculated by Silverman (1986) using (18.2).

18.2 Reflection and Transmission Amplitudes, Conservation Laws

The optics of stratified chiral and/or anisotropic media can be quantified in terms of four reflection and four transmission amplitudes. These can be of two kinds, depending on whether the wave description is in terms of planar or circular polarization.

In the case of plane polarized states, the electric field components of the incident, reflected and transmitted waves are resolved along the $\mathbf{p}$ and $\mathbf{p}'$ directions which lie in the plane of incidence, and the $\mathbf{s}(= \mathbf{s}')$ direction perpendicular to the plane of incidence. If propagation is in the zx plane (the plane of incidence) and in the direction of positive x, with wavevector components in the (homogeneous) medium of incidence $\mathbf{k}_1 = (K, 0, q_1)$, an s-polarized incident wave of unit electric field magnitude will be

$$\mathbf{E}_s = (0, 1, 0)\exp\, i(Kx + q_1 z). \tag{18.14}$$

If θ_1 is the angle of incidence, the reflected wave electric field will be (by definition of the reflection amplitudes r_{ss} and r_{sp})

$$\mathbf{E}' = (r_{sp}\cos\theta_1, r_{ss}, r_{sp}\sin\theta_1)\exp\, i(Kx - q_1 z). \tag{18.15}$$

For an incident p-polarized wave the incoming and reflected waves are

$$\mathbf{E}_p = (\cos\theta_1, 0, -\sin\theta_1)\exp\, i(Kx + q_1 z), \tag{18.16}$$

$$\mathbf{E}' = (r_{pp}\cos\theta_1, r_{ps}, r_{pp}\sin\theta_1)\exp\, i(Kx - q_1 z). \tag{18.17}$$

(The reflection amplitude for the x component is r_{pp}, while for the z component it is $-r_{pp}$: see equations (1.26) and (1.27) of Sect. 1.2.)

The cause of the reflection is assumed to be a general planar-stratified layer (which may be chiral and anisotropic) resting on a homogeneous achiral isotropic substrate, in which the wavevector of the transmitted wave is $\mathbf{k}_2 = (K, 0, q_2)$. Note that the component of the wavevector along the stratification (the x component K) is a constant of the motion, because of translational invariance in the x direction. The transmitted wave when the incident wave is s-polarized is

$$\mathbf{E}'' = (t_{sp}\cos\theta_2, t_{ss}, -t_{sp}\sin\theta_2)\exp\, i[Kx + q_2(z - d)], \tag{18.18}$$

where θ_2 is the angle of refraction in the substrate and d is the total thickness of the chiral (and possibly anisotropic) layer. The corresponding transmitted electric field when a p-polarized wave is incident is

$$\mathbf{E}'' = (t_{pp}\cos\,\theta_2, t_{ps}, -t_{pp}\sin\theta_2)\exp\, i[Kx + q_2(z - d)]. \tag{18.19}$$

These relations define the transmission amplitudes t_{ss}, t_{sp}, t_{pp} and t_{ps}.

If the chiral layer is non-absorbing, the reflected plus transmitted fluxes of energy must add up to the incident flux. The energy density of a plane electromagnetic wave in a medium with dielectric constant ε and permeability μ is proportional to $\varepsilon|\mathbf{E}|^2$ and the speed is $c/\sqrt{\varepsilon\mu}$; thus the energy flux is proportional to $\sqrt{\varepsilon/\mu}|\mathbf{E}|^2$. The amount of energy in the primary wave which is incident on a unit area of the interface in unit time is proportional to $\sqrt{\varepsilon_1/\mu_1}\cos\theta_1$, the amount reflected to $\sqrt{\varepsilon_1/\mu_1}\cos\theta_1$ times the absolute square of the reflected field, and the amount carried away by the transmitted wave similarly to $\sqrt{\varepsilon_2/\mu_2}\cos\theta_2$ times the absolute square of the transmitted electric field. (See Fig. 2.1 for the geometry leading to the factors $\cos\theta_1$ and $\cos\theta_2$.) Thus energy conservation reads, for incident s and p polarizations,

$$\sqrt{\varepsilon_1/\mu_1}\cos\theta_1\left(1 - |r_{ss}|^2 - |r_{sp}|^2\right) = \sqrt{\varepsilon_2/\mu_2}\cos\theta_2(|t_{ss}|^2 + |t_{sp}|^2),$$

$$\sqrt{\varepsilon_1/\mu_1}\cos\theta_1\left(1 - |r_{pp}|^2 - |r_{ps}|^2\right) = \sqrt{\varepsilon_2/\mu_2}\cos\theta_2(|t_{pp}|^2 + |t_{ps}|^2). \tag{18.20}$$

These relations hold for arbitrary non-absorbing stratifications. When there is absorption, the difference between the left- and right-hand sides gives the absorption in the stratification.

An alternative characterization of polarization states is in terms of positive and negative *helicities* (opposite circular polarizations). For circularly polarized incident light we need the reflection and transmission amplitudes $r_{++}, r_{+-}, r_{-+}, r_{--}$ and $t_{++}, t_{+-}, t_{-+}, t_{--}$, where, for example, r_{+-} gives the complex amplitude of the light reflected with negative helicity when positive helicity light is incident. (We avoid the left and right circular polarization terminology, because two opposite conventions are in use.) Let $(\mathbf{p}, \mathbf{s}, \mathbf{k}_1)$ denote a right-handed triplet of vectors for the incident light, with $\mathbf{p}$ and $\mathbf{s}$ being unit vectors perpendicular to the direction of propagation, and, respectively, parallel and perpendicular to the plane of incidence. Similarly, let $(\mathbf{p}', \mathbf{s}', \mathbf{k}_1')$ be a similar triplet for the reflected light (the choice $\mathbf{s}' = \mathbf{s}$ then implies that $\mathbf{p}' \to -\mathbf{p}$ at normal incidence, and $\mathbf{p}' \to \mathbf{p}$ at glancing incidence). Then $(\mathbf{p} + i\mathbf{s})/\sqrt{2}$ represents an incident wave of positive helicity and unit magnitude. As usual the zx plane is the plane of incidence, with $\mathbf{s} = (0, 1, 0) = \mathbf{s}'$, $\mathbf{p} = (\cos\theta_1, 0, -\sin\theta_1)$ and $\mathbf{p}' = (-\cos\theta_1, 0, -\sin\theta_1)$. From (18.15) and (18.17), an electric field of unit magnitude along $\mathbf{s}$ reflects to $r_{ss}\mathbf{s}' - r_{sp}\mathbf{p}'$, and an electric field of unit magnitude along $\mathbf{p}$ reflects to $-r_{pp}\mathbf{p}' + r_{ps}\mathbf{s}'$. The reflected field is therefore

$$[(-r_{pp} - ir_{sp})\mathbf{p}' + (r_{ps} + ir_{ss})\mathbf{s}']/\sqrt{2}. \tag{18.21}$$

From (18.21) we extract the coefficients of positive and negative helicity, namely of $(\mathbf{p}' \pm i\mathbf{s}')/\sqrt{2}$, to find

$$\begin{aligned} r_{++} &= \tfrac{1}{2}(r_{ss} - r_{pp}) - \tfrac{1}{2}i(r_{sp} + r_{ps}), \\ r_{+-} &= -\tfrac{1}{2}(r_{ss} + r_{pp}) - \tfrac{1}{2}i(r_{sp} - r_{ps}). \end{aligned} \tag{18.22}$$

Similarly, when the incident wave is $(\mathbf{p} - i\mathbf{s})/\sqrt{2}$ (negative helicity), the reflected field is

$$[(-r_{pp} + ir_{sp})\mathbf{p}' + (r_{ps} - ir_{ss})\mathbf{s}']/\sqrt{2}, \tag{18.23}$$

and the corresponding reflection amplitudes are

$$\begin{aligned} r_{-+} &= -\tfrac{1}{2}(r_{ss} + r_{pp}) + \tfrac{1}{2}i(r_{sp} - r_{ps}), \\ r_{--} &= \tfrac{1}{2}(r_{ss} - r_{pp}) + \tfrac{1}{2}i(r_{sp} + r_{ps}). \end{aligned} \tag{18.24}$$

For all the cases considered in this chapter, $r_{sp} = r_{ps}$ and so $r_{+-} = r_{-+}$.

The transmission amplitudes as characterized by helicity are found as follows. Let $(\mathbf{p} \pm i\mathbf{s})/\sqrt{2}$ represent the electric fields of incident waves of positive and

negative helicity, and likewise $(\mathbf{p}'' \pm i\mathbf{s}'')/\sqrt{2}$ for the transmitted helicities. When $(\mathbf{p} + i\mathbf{s})/\sqrt{2}$ is incident, the transmitted field is

$$[(t_{pp} + it_{sp})\mathbf{p}'' + (t_{ps} + it_{ss})\mathbf{s}'']/\sqrt{2} \tag{18.25}$$

The coefficients of positive and negative helicity in (18.25) give

$$\begin{aligned} t_{++} &= \tfrac{1}{2}(t_{pp} + t_{ss}) + \tfrac{1}{2}i(t_{sp} - t_{ps}), \\ t_{+-} &= \tfrac{1}{2}(t_{pp} - t_{ss}) + \tfrac{1}{2}i(t_{sp} + t_{ps}). \end{aligned} \tag{18.26}$$

Similarly, for negative helicity incident we find

$$\begin{aligned} t_{-+} &= \tfrac{1}{2}(t_{pp} - t_{ss}) - \tfrac{1}{2}i(t_{sp} + t_{ps}), \\ t_{--} &= \tfrac{1}{2}(t_{pp} + t_{ss}) - \tfrac{1}{2}i(t_{sp} - t_{ps}). \end{aligned} \tag{18.27}$$

The inverse relations are as follows:

$$\begin{aligned} r_{ss} &= \tfrac{1}{2}(r_{++} + r_{--}) - \tfrac{1}{2}(r_{+-} + r_{-+}), \\ r_{pp} &= -\tfrac{1}{2}(r_{++} + r_{--}) - \tfrac{1}{2}(r_{+-} + r_{-+}), \\ r_{sp} &= \tfrac{1}{2}i(r_{++} - r_{--}) + \tfrac{1}{2}i(r_{+-} - r_{-+}), \\ r_{ps} &= \tfrac{1}{2}i(r_{++} - r_{--}) - \tfrac{1}{2}(r_{+-} - r_{-+}). \end{aligned} \tag{18.28}$$

$$\begin{aligned} t_{ss} &= \tfrac{1}{2}(t_{++} + t_{--}) - \tfrac{1}{2}(t_{+-} + t_{-+}), \\ t_{pp} &= \tfrac{1}{2}(t_{++} + t_{--}) + \tfrac{1}{2}(t_{+-} + t_{-+}), \\ t_{sp} &= -\tfrac{1}{2}i(t_{++} - t_{--}) - \tfrac{1}{2}i(t_{+-} - t_{-+}), \\ t_{ps} &= \tfrac{1}{2}i(t_{++} - t_{--}) - \tfrac{1}{2}i(t_{+-} - t_{-+}). \end{aligned} \tag{18.29}$$

Energy conservation relations may also be written down in terms of the helicity reflection and transmission amplitudes. For incident waves of respectively positive and negative helicity, energy conservation implies

$$\begin{aligned} \sqrt{\varepsilon_1/\mu_1}\cos\theta_1\left(1 - |r_{++}|^2 - |r_{+-}|^2\right) &= \sqrt{\varepsilon_2/\mu_2}\cos\theta_2(|t_{++}|^2 + |t_{+-}|^2), \\ \sqrt{\varepsilon_1/\mu_1}\cos\theta_1\left(1 - |r_{--}|^2 - |r_{-+}|^2\right) &= \sqrt{\varepsilon_2/\mu_2}\cos\theta_2(|t_{--}|^2 + |t_{-+}|^2). \end{aligned} \tag{18.30}$$

18.2.1 Differential Reflectance, Ellipsometry

Chiral media in general reflect opposite circular polarizations differently. When a wave of unit amplitude and positive helicity is incident, the amplitudes of the reflected waves of positive and negative helicities are r_{++} and r_{+-}, respectively. If the detector is polarization-insensitive, the reflected intensity is proportional to $R_+ = |r_{++}|^2 + |r_{+-}|^2$. (This result may be verified by expressing the reflected wave as

$$\frac{r_{++}(\mathbf{p}' + i\mathbf{s}')}{\sqrt{2}} + \frac{r_{+-}(\mathbf{p}' - i\mathbf{s}')}{\sqrt{2}} = \frac{(r_{++} + r_{+-})\mathbf{p}'}{\sqrt{2}} + i\frac{(r_{++} - r_{+-})\mathbf{s}'}{\sqrt{2}}.$$

Then $|\mathbf{E}'|^2 = \frac{1}{2}(|r_{++} + r_{+-}|^2 + |r_{++} - r_{+-}|^2) = |r_{++}|^2 + |r_{+-}|^2$.) Similarly, a wave of unit amplitude and negative helicity produces a reflected intensity proportional to $R_- = |r_{--}|^2 + |r_{-+}|^2$. The differential circular reflectance (DCR) is defined as (Silverman 1986)

$$\mathrm{DCR} = \frac{R_+ - R_-}{R_+ + R_-}. \tag{18.31}$$

If linearly polarized waves of s or p polarization are incident, and the detector is polarization-insensitive, the reflected intensities will be proportional to $R_s = |r_{ss}|^2 + |r_{sp}|^2$ or $R_p = |r_{pp}|^2 + |r_{ps}|^2$, respectively. Differential linear reflectance (DLR) is defined as (Silverman et al. 1988)

$$\mathrm{DLR} = \frac{R_s - R_p}{R_s + R_p}. \tag{18.32}$$

Silverman and collaborators have made measurements on chiral media of both DLR (Silverman et al. 1988) and of DCR (Silverman et al. 1988; Silverman and Badoz 1992; Silverman et al. 1992; Badoz and Silverman 1992).

The reflection amplitudes r_{pp}, r_{ss}, r_{ps} and r_{sp} can be used to calculate the reflection ellipsometric signal, which in the common experimental configurations is one of the ratios ρ_P or ρ_A defined in Chap. 9:

$$\rho_P = \frac{r_{pp} + r_{sp}\tan P}{r_{ps} + r_{ss}\tan P}, \quad \rho_A = \frac{r_{pp} + r_{ps}\tan A}{r_{sp} + r_{ss}\tan A}. \tag{18.33}$$

In (18.33) P is the angle between the polarizer easy axis and the incident $\mathbf{p}$ direction, while A is the angle between the analyser easy axis and the reflected $\mathbf{p}'$ direction. (The $\mathbf{p}$ and $\mathbf{p}'$ TM directions lie in the plane of incidence, the $\mathbf{s}$ TE direction is perpendicular to the plane of incidence; all are perpendicular to the incoming or reflected beams.)

We note in passing that at the polarizing angle defined by $r_{pp}r_{ss} = r_{sp}r_{ps}$ (see Sect. 18.4), ρ_P and ρ_A become independent of the orientations of the polarizer and analyser, and take the respective values r_{pp}/r_{ps} and r_{pp}/r_{sp}. For the cases considered in this paper, $r_{sp} = r_{ps}$, so ρ_P and ρ_A are equal at the polarizing angle.

Ellipsometry of chiral media is discussed in more detail in Sect. 5 of Lekner (1996).

18.3 Wave Propagation in Chiral Media

Let us first consider the general case of chiral *inhomogeneous* media; ε, μ and γ can all be dependent on position. We will henceforth assume the validity of the two curl equations (18.1) and of the constitutive relations (18.2). Since the equations are linear in the fields, we can deal with one Fourier component at a time; we assume a time dependence $e^{-i\omega t}$, so that we can use (18.5) and

$$c\nabla \times \mathbf{E} = i\omega\mathbf{B}, \qquad c\nabla \times \mathbf{H} = -i\omega\mathbf{D}. \tag{18.34}$$

The fields **B**, **H** and **D** can be eliminated from (18.5) and (18.34) by substitution of $\mathbf{B} = (c/i\omega)\nabla \times \mathbf{E}$ into $\mathbf{H} = (\mathbf{B} + i\gamma\mathbf{E})/\mu$, and then of the latter expression into $\mathbf{D} = (c/\omega)\nabla \times \mathbf{H} = \varepsilon\mathbf{E} + i\gamma\mathbf{H}$. The result is a second-order equation for **E**, namely

$$\mu\nabla \times \left(\frac{1}{\mu}\nabla \times \mathbf{E}\right) = \left(\varepsilon\mu - \gamma^2\right)\frac{\omega^2}{c^2}\mathbf{E} + \frac{\omega}{c}\left[\gamma\nabla \times \mathbf{E} + \mu\nabla \times \left(\frac{\gamma}{\mu}\mathbf{E}\right)\right]. \tag{18.35}$$

The equation for **H** has the same form, with $\mathbf{E}, \mathbf{H}$ and ε, μ interchanged.

When the medium is z-stratified, ε, μ and γ are functions of z only. Suppose that there is a plane wave incident on the stratification, propagating in the zx plane. Because of the assumed translational invariance in the x and y directions, there will be no y dependence in any field component, and the x dependence of all field components is contained in the factor $\exp(iKx)$. K is the x-component of the wavevector, and is a constant of the motion because of the translational invariance in the x direction. For a z-stratified chiral medium, the three components E_x, E_y, E_z satisfy the coupled ordinary differential equations

$$E_x'' - \frac{\mu'}{\mu}E_x' + \left(\varepsilon\mu - \gamma^2\right)\frac{\omega^2}{c^2}E_x - \frac{\omega}{c}\left[2\gamma E_y' + \left(\gamma' - \gamma\frac{\mu'}{\mu}\right)E_y\right] - iK\left(E_z' - \frac{\mu'}{\mu}E_z\right) = 0,$$

$$E_y'' - \frac{\mu'}{\mu}E_y' + \left[\left(\varepsilon\mu - \gamma^2\right)\frac{\omega^2}{c^2} - K^2\right]E_y + \frac{\omega}{c}\left[2\gamma E_x' + \left(\gamma' - \gamma\frac{\mu'}{\mu}\right)E_x\right] - 2i\frac{\omega}{c}\gamma KE_z = 0,$$

$$\left[\left(\varepsilon\mu-\gamma^2\right)\frac{\omega^2}{c^2}-K^2\right]E_z-iKE_x'+2i\frac{\omega}{c}\gamma KE_z=0. \tag{18.36}$$

(The primes denote differentiation with respect to z.)

Finally, we specialize to a *homogeneous* chiral medium. In this case ε, μ and γ are constant within the medium. We look for plane wave eigenstates, in which all field components have the variation $\exp(iqz)$, where q is the z-component of the wavevector. The differential equations (18.36) then reduce to the three homogeneous linear algebraic equations

$$\begin{gathered}\left[\left(\varepsilon\mu-\gamma^2\right)\frac{\omega^2}{c^2}-q^2\right]E_x-2i\frac{\omega}{c}\gamma qE_y+qKE_z=0,\\ 2i\frac{\omega}{c}\gamma qE_x+\left[\left(\varepsilon\mu-\gamma^2\right)\frac{\omega^2}{c^2}-K^2-q^2\right]E_y-2i\frac{\omega}{c}\gamma KE_z=0,\\ qKE_x+2i\frac{\omega}{c}\gamma KE_y+\left[\left(\varepsilon\mu-\gamma^2\right)\frac{\omega^2}{c^2}-K^2\right]E_z=0.\end{gathered} \tag{18.37}$$

A solution with non-zero $\mathbf{E}$ is possible only if the determinant of the coefficients of E_x, E_y and E_z in this set of equations is zero. This gives the condition

$$\begin{vmatrix} k_\gamma^2-q^2 & -2i\frac{\omega}{c}\gamma q & qK \\ 2i\frac{\omega}{c}\gamma q & k_\gamma^2-q^2-K^2 & -2i\frac{\omega}{c}\gamma K \\ qK & 2i\frac{\omega}{c}\gamma K & k_\gamma^2-K^2 \end{vmatrix}=0, \tag{18.38}$$

where

$$k_\gamma^2=\left(\varepsilon\mu-\gamma^2\right)\frac{\omega^2}{c^2}. \tag{18.39}$$

A similar eigenvalue equation for q is obtained for anisotropic media, as we saw in Chap. 8. There the matrix of coefficients was symmetric, here it is Hermitian. Equation (18.38) is a quartic in q, with solutions $\pm q_+$ and $\pm q_-$, where

$$q_\pm^2=(\sqrt{\varepsilon\mu}\pm\gamma)^2\frac{\omega^2}{c^2}-K^2\equiv k_\pm^2-K^2. \tag{18.40}$$

The four possible plane waves in the chiral medium have wavevectors

$$(K,0,\pm q_+)\quad\text{and}\quad(K,0,\pm q_-). \tag{18.41}$$

The two with the upper sign are for waves propagating in the $+z$ direction, the two with the lower sign are for waves propagating in the $-z$ direction. The square of the wavevector is

$$K^2 + q_\pm^2 = (\sqrt{\varepsilon\mu} \pm \gamma)^2 \frac{\omega^2}{c^2} \equiv k_\pm^2 \equiv n_\pm^2 \frac{\omega^2}{c^2}. \tag{18.42}$$

Thus there are two effective indices for the chiral medium

$$n_\pm = \sqrt{\varepsilon\mu} \pm \gamma \equiv n \pm \gamma, \tag{18.43}$$

which correspond to waves of positive and negative helicity, as we shall see shortly. The average of the two indices is $n = \sqrt{\varepsilon\mu}$, and their product is $\varepsilon\mu - \gamma^2 = \left(ck_\gamma/\omega\right)^2$.

The electric field eigenstates which correspond to the eigenvalues $q_\pm$ given in (18.42) are obtained from (18.37) by substituting $q_\pm$ for q. We find, for the waves propagating in the $+z$ direction,

$$\mathbf{E}_+ \sim (q_+, ik_+, -K), \qquad \mathbf{E}_- \sim (q_-, -ik_-, -K), \tag{18.44}$$

where $k_\pm = n_\pm \omega/c$. The corresponding wavevectors are $\mathbf{k}_\pm = (K, 0, q_\pm)$, and we see that, for each mode, the electric field eigenstate is perpendicular to the wavevector. The fields given in (18.44) have a phase difference of 90° between their y and z, x components. Also $k_\pm^2 = K^2 + q_\pm^2$, so the two eigenstates correspond to circularly polarized light of positive and negative helicity.

The other fields can be found from $\mathbf{E}$ by means of $\mathbf{B} = (c/\omega)\mathbf{k} \times \mathbf{E}$, $\mathbf{H} = (\mathbf{B} + i\gamma\mathbf{E})/\mu$ and $\mathbf{D} = \varepsilon\mathbf{E} + i\gamma\mathbf{H}$. They are

$$\begin{aligned} \mathbf{B}_+ &= -in_+\mathbf{E}_+, \quad \mathbf{B}_- = in_-\mathbf{E}_-, \\ \mathbf{H}_+ &= -i\sqrt{\varepsilon/\mu}\mathbf{E}_+, \quad \mathbf{H}_- = i\sqrt{\varepsilon/\mu}\mathbf{E}_-, \\ \mathbf{D}_+ &= n_+\sqrt{\varepsilon/\mu}\mathbf{E}_+, \quad \mathbf{D}_- = n_-\sqrt{\varepsilon/\mu}\mathbf{E}_-. \end{aligned} \tag{18.45}$$

The Poynting vectors have the appropriate directions: for example, $\mathbf{E}_+ \times \mathbf{H}_+^*$ is proportional to $\mathbf{k}_+$. The corresponding fields for plane waves propagating in the $-z$ direction are obtained by replacing q_+ by $-q_+$ and q_- by $-q_-$ in $\mathbf{k}_\pm = (K, 0, q_\pm)$ and in (18.44). The helicities are then negative for $\mathbf{E}_+ \sim (-q_+, ik_+, -K)$ and positive for $\mathbf{E}_- \sim (-q_-, -ik_-, -K)$.

18.3.1 Eigenstates of Curl

An elegant alternative approach to propagation in homogeneous chiral media is in terms of two related linear combinations of the $\mathbf{E}$ and $\mathbf{H}$ fields (Bohren 1974; Eftimiu and Pearson 1989; Lindell et al. 1994),

$$\mathbf{F}_\pm = \mathbf{E} \pm i\eta\mathbf{H}, \quad \eta = \sqrt{\mu/\varepsilon}. \tag{18.46}$$

Provided that η is constant in space, the curl equations (18.34) and the constitutive equations (18.5) together imply that $\mathbf{F}_+$ and $\mathbf{F}_-$ are eigenstates of the curl operator

$$\nabla \times \mathbf{F}_\pm = \pm k_\pm \mathbf{F}_\pm, \tag{18.47}$$

where $k_\pm = n_\pm\omega/c = (n \pm \gamma)\omega/c$ as before. Eigenstates of curl are also known as Beltrami fields, discussed in a wider context in Chap. 1 of Lakhtakia (1994).

If we write (18.47) collectively as $\nabla \times \mathbf{F} = k\mathbf{F}$ (with $\mathbf{F} = \mathbf{F}_\pm$ and $k = \pm k_\pm$), plane wave propagation in the zx plane, with $\mathbf{F}$ proportional to $\exp i(Kx + qz)$, is possible if

$$-iqF_y = kF_x, \quad iqF_x - iKF_z = kF_y, \quad iKF_y = kF_z\,. \tag{18.48}$$

These are three homogeneous equations in the field components (F_x, F_y, F_z) and a non-zero solution will exist only if the determinant of their coefficients is zero, namely if

$$\begin{vmatrix} k & iq & 0 \\ -iq & k & iK \\ 0 & -iK & k \end{vmatrix} = 0. \tag{18.49}$$

This determinant factors to $k(k^2 - K^2 - q^2)$. The nonzero values of k are $\pm k_\pm$, and thus we regain (18.42). From (18.45) we see that for the positive helicity wave $\mathbf{H}_+ = \mathbf{E}_+/(i\eta)$ and so $\mathbf{F}_+ = 2\mathbf{E}_+$. For the negative helicity wave $\mathbf{H}_- = \mathbf{E}_-/(-i\eta)$ and $\mathbf{F}_- = 2\mathbf{E}_-$. Thus for plane wave eigenstates in homogeneous chiral media, (18.47) reads

$$\nabla \times \mathbf{E}_+ = k_+\mathbf{E}_+, \quad \nabla \times \mathbf{E}_- = -k_-\mathbf{E}_-, \tag{18.50}$$

or the same equations with $\mathbf{E}_\pm$ replaced by $\mathbf{H}_\pm$.

For inhomogeneous media the positive and negative helicities are coupled: (18.47) is replaced by

$$\nabla \times \mathbf{F}_\pm = \pm[k_\pm\mathbf{F}_\pm + (2\eta)^{-1}\nabla\eta \times (\mathbf{F}_+ - \mathbf{F}_-)]. \tag{18.51}$$

18.3.2 Boundary Conditions

We shall see that the boundary conditions at an interface between chiral media are the continuity of the tangential components of $\mathbf{E}$ and $\mathbf{H}$. For a z-stratified medium and plane waves propagating in the zx plane, the curl of $\mathbf{E}$ is

$$\nabla \times \mathbf{E} = (-E_y', E_x' - iKE_z, iKE_y). \tag{18.52}$$

Since the time derivative of **B**, which by (18.1) is proportional to $\nabla \times \mathbf{E}$, is to be free of singularity at the interface, it follows that E_x' and E_y' are non-singular and thus that E_x and E_y are continuous across the interface. Likewise, since the time derivative of **D**, which is proportional to the curl of **H**, is free of singularity at the interface, H_x and H_y are continuous across the boundary.

The continuity of E_x also follows directly from the last equation in (18.36), while the terms containing derivatives in the first two equations of (18.36) can be written as $i\omega/c$ times

$$\mu H_y' + i\gamma E_y' \quad \text{and} \quad \mu H_x' + i\gamma E_x'. \tag{18.53}$$

Thus the continuity of E_x and H_x follows from the differential equations for the components of **E**, whereas these differential equations allow discontinuities in E_y and H_y across the interface, provided $\mu H_y' + i\gamma E_y'$ remains non-singular. The possibility of discontinuities ΔH_y and ΔE_y satisfying $\mu \Delta H_y + i\gamma \Delta E_y = 0$ is eliminated by the differential equations for H_x, H_y and H_z. These have the same form as those for the components of **E**, with H_x replacing E_x, et cetera, and μ interchanged with ε. They imply that $\varepsilon E_x' + i\gamma H_x'$ and $\varepsilon E_y' + i\gamma H_y'$ are non-singular, and that H_x is continuous at the boundary. Discontinuities in E_y and H_y would then have to satisfy both $\mu \Delta H_y + i\gamma \Delta E_y = 0$ and $\varepsilon \Delta E_y + i\gamma \Delta H_y = 0$. The determinant of the coefficients of ΔE_y and ΔH_y in these two equations is $\varepsilon\mu + \gamma^2$, which we take to be nonzero, thus implying $\Delta E_y = 0$ and $\Delta H_y = 0$. We conclude that the continuity of the tangential components of **E** and **H** follows from the differential equations, provided $\varepsilon\mu + \gamma^2$ is not zero.

18.4 Reflection from an Achiral–Chiral Interface

Reflection at the surface of a chiral medium was considered by Silverman (1986), who derived the reflection and transmission amplitudes on the basis of (a) the symmetric set of constitutive relations (18.2), and (b) the constitutive relations (18.3). He showed that the results are not equivalent, with small violations of energy conservation in case (b) under conditions of total reflection. Here the reflection and transmission amplitudes are derived on the basis of the symmetric set of constitutive relations (18.2). As we saw in Sect. 18.1, the Drude-Born-Fedorov relations (18.5) give equivalent results for time-harmonic waves.

18.4.1 Wavefunctions

Let a plane wave be incident from an optically non-active medium (dielectric and permeability constants ε_1 and μ_1), at an angle θ_1 to the interface normal. We wish to find the four reflection amplitudes r_{ss}, r_{sp}, r_{pp} and r_{ps} which completely characterize the reflection properties of the interface.

Inside the optically active medium the two plane wave eigenstates which propagate in the $+z$ direction have electric field vectors

$$\mathbf{E}_+ = (\cos\theta_+, i, -\sin\theta_+)\exp\, i(Kx + q_+ z), \tag{18.54}$$

$$\mathbf{E}_- = (\cos\theta_-, -i, -\sin\theta_-)\exp\, i(Kx + q_- z).$$

Here $\theta_\pm$ are the angles of refraction for the two plane waves of opposite helicity. Their cosines and sines are given by

$$\cos\theta_\pm = q_\pm/k_\pm, \quad \sin\theta_\pm = K/k_\pm. \tag{18.55}$$

For incident s (TE) plane-polarized waves, the incoming and reflected waves have electric fields given by (18.14) and (18.15). Thus the electric field in medium 1 is

$$\mathbf{E}_1 = \left(r_{sp}\cos\theta_1 \mathrm{e}^{-iq_1 z}, \mathrm{e}^{iq_1 z} + r_{ss}\mathrm{e}^{-iq_1 z}, r_{sp}\sin\theta_1 \mathrm{e}^{-iq_1 z}\right)\mathrm{e}^{iKx}. \tag{18.56}$$

The magnetic field H_1 in medium 1 is given by

$$\mathbf{H}_1 = \mathbf{B}_1/\mu_1 = (c/i\omega\mu_1)\nabla \times \mathbf{E}_1. \tag{18.57}$$

At the boundary $z = 0$ this takes the value

$$\mathbf{H}_1(z=0) = \frac{n_1}{\mu_1}\left(-[1 - r_{ss}]\cos\theta_1, -r_{sp}, [1 + r_{ss}]\sin\theta_1\right)\mathrm{e}^{iKx}. \tag{18.58}$$

Inside the optically active medium, the electric and magnetic fields are

$$\mathbf{E} = t_{s+}\mathbf{E}_+ + t_{s-}\mathbf{E}_-, \qquad \mathbf{H} = -i\sqrt{\varepsilon/\mu}(t_{s+}\mathbf{E}_+ - t_{s-}\mathbf{E}_-), \tag{18.59}$$

where t_{s+} and t_{s-} are the transmission amplitudes for the two circularly polarized waves in the chiral medium.

18.4.2 Reflection and Transmission Amplitudes

The continuity of the tangential components of $\mathbf{E}$ and $\mathbf{H}$ across the interface at $z = 0$ gives four relations, which can be solved for the four unknowns r_{ss}, r_{sp}, t_{s+} and t_{s-}. We find, with

$$c_1 = \cos\theta_1, \qquad c_\pm = \cos\theta_\pm = \sqrt{1 - \left(\frac{n_1}{n_\pm}\sin\theta_1\right)^2}, \tag{18.60}$$

$$D = c_1^2 + \frac{c_1(c_+ + c_-)(m + m^{-1})}{2} + c_+ c_-, \qquad m = \sqrt{\varepsilon\mu_1/\varepsilon_1\mu},$$

that the reflection and transmission amplitudes when s-polarized light is incident are given by

$$\begin{aligned} r_{ss} &= \left[c_1^2 - \frac{c_1(c_+ + c_-)(m - m^{-1})}{2} - c_+ c_-\right]/D, \\ r_{sp} &= -ic_1(c_+ - c_-)/D, \\ t_{s+} &= -ic_1(c_1 + c_-/m)/D, \\ t_{s-} &= ic_1(c_1 + c_+/m)/D. \end{aligned} \tag{18.61}$$

For p polarization incident the incoming and reflected waves have electric fields

$$\begin{aligned} &(\cos\theta_1, 0, -\sin\theta_1)\exp i(Kx + q_1 z), \\ &(r_{pp}\cos\theta_1, r_{ps}, r_{pp}\sin\theta_1)\exp i(Kx - q_1 z). \end{aligned} \tag{18.62}$$

The electric field $\mathbf{E}_1$ is the sum of these; the magnetic field $\mathbf{H}_1$ given by (18.57) takes the value at the $z = 0$ boundary

$$\mathbf{H}_1(z = 0) = \frac{n_1}{\mu_1}\left(r_{ps}\cos\theta_1, 1 - r_{pp}, r_{ps}\sin\theta_1\right)\mathrm{e}^{iKx}. \tag{18.63}$$

The fields inside the chiral medium are now

$$\mathbf{E} = t_{p+}\mathbf{E}_+ + t_{p-}\mathbf{E}_-, \qquad \mathbf{H} = -i\sqrt{\varepsilon/\mu}(t_{p+}\mathbf{E}_+ - t_{p-}\mathbf{E}_-). \tag{18.64}$$

The reflection and transmission amplitudes when p-polarized light is incident are again found from the continuity of the tangential components of $\mathbf{E}$ and $\mathbf{H}$:

$$\begin{aligned} r_{pp} &= -\left[c_1^2 + \frac{c_1(c_+ + c_-)(m - m^{-1})}{2} - c_+ c_-\right]/D, \\ r_{ps} &= -ic_1(c_+ - c_-)/D, \\ t_{p+} &= c_1(c_1/m + c_-)/D, \end{aligned} \tag{18.65}$$

$$t_{p-} = c_1(c_1/m + c_+)/D.$$

The formulae (18.61) and (18.65) are in accord with those of Silverman (1986), but are not identical to those of Bassiri et al. (1988), unless we make the identification (18.13). We note that $r_{sp} = r_{ps}$. When the chirality is zero, r_{sp} and r_{ps} are also zero, while r_{ss} and r_{pp} reduce to the usual Fresnel amplitudes. The diagonal reflection amplitudes satisfy

$$1 - r_{ss}^2 = c_1(c_+ + c_- + 2c_1 m)[c_1(c_+ + c_-)m + 2c_+ c_-]/mD, \tag{18.66}$$

$$1 - r_{pp}^2 = c_1[(c_+ + c_-)m + 2c_1)][c_1(c_+ + c_-) + 2c_+ c_- m]/mD.$$

For non-absorbing chiral media the right-hand sides of (18.66) are non-negative, so the *ss* and *pp* reflectivities cannot exceed unity. Also the magnitude of $r_{sp} = r_{ps}$ is less than unity, by inspection.

At normal incidence $K \to 0$ and the cosines c_1 and $c_\pm$ tend to unity; the reflection and transmission amplitudes then take values independent of the chirality parameter γ:

$$r_{ss}, r_{pp} \to \frac{1-m}{1+m}, \quad r_{sp}, r_{ps} \to 0, \tag{18.67}$$

$$t_{p\pm} \to \frac{1}{1+m}, \quad t_{s\pm} \to \frac{\mp i}{1+m}.$$

The transmission amplitudes for positive and negative helicities follow from the definitions of $t_{s\pm}$ and $t_{p\pm}$ in (18.59) and (18.64). An incident wave with electric field vector $\mathbf{p} \pm i\mathbf{s}$ transmits to $t_{p+}\mathbf{E}_+ + t_{p-}\mathbf{E}_- \pm it_{s+}\mathbf{E}_+ \pm it_{s-}\mathbf{E}_- = \left(t_{p+} \pm it_{s+}\right)\mathbf{E}_+ + \left(t_{p-} \pm it_{s-}\right)\mathbf{E}_-$. Therefore

$$t_{++} = t_{p+} + it_{s+}, \quad t_{+-} = t_{p-} + it_{s-}, \tag{18.68}$$

$$t_{-+} = t_{p+} - it_{s+}, \quad t_{--} = t_{p-} - it_{s-}.$$

The helicity reflection and transmission amplitudes for a sharp achiral–chiral interface located at $z = 0$ are all real when the chiral medium is non-absorbing:

$$r_{+-} = \frac{c_1(c_+ + c_-)(m - m^{-1})}{2D} = r_{-+},$$

$$r_{++} = \frac{(c_1 - c_+)(c_1 + c_-)}{D}, \quad r_{--} = \frac{(c_1 + c_+)(c_1 - c_-)}{D}, \tag{18.69}$$

$$t_{++} = \frac{c_1(c_- + c_1)(1 + m^{-1})}{D}, \quad t_{+-} = \frac{c_1(c_+ - c_1)(1 - m^{-1})}{D},$$

$$t_{--} = \frac{c_1(c_+ + c_1)(1+m^{-1})}{D}, \quad t_{-+} = \frac{c_1(c_- - c_1)(1-m^{-1})}{D}.$$

(D and m are defined in (18.60).) The normal-incidence limiting values are

$$r_{++}, r_{--} \to 0, \quad r_{+-}, r_{-+} \to \frac{m-1}{m+1}, \tag{18.70}$$

$$t_{++}, t_{--} \to \frac{2}{m+1}, \quad t_{+-}, t_{-+} \to 0.$$

At glancing incidence ($c_1 \to 0$) all the transmission amplitudes go to zero, r_{++}, r_{--} tend to -1, r_{+-}, r_{-+} tend to zero, $r_{pp} \to 1$ and $r_{ss} \to -1$. The probability for photon spin-flip is zero at both normal and glancing incidence. (Helicity is reversed at normal incidence because of the reversal of the direction of travel of the light on reflection.)

Reflection near the critical angles $\theta_1^{\pm}$ given by $\sin\theta_1^{\pm} = n_{\pm}/n$ is discussed in Sect. 18.5, together with the chiral layer case. There the off-diagonal reflection amplitudes r_{sp}, r_{ps} and r_{+-}, r_{-+} are proportional to the square root of the chiral index γ, as we shall see.

The energy conservation conditions to be satisfied by the reflection and transmission amplitudes follow from arguments along the lines given in relation to (18.20), and were first written down by Silverman and Badoz (1989). The helicity amplitudes satisfy

$$\sqrt{\varepsilon_1/\mu_1}\cos\theta_1\Big(1 - |r_{++}|^2 - |r_{+-}|^2\Big) = \sqrt{\varepsilon/\mu}\Big(\cos\theta_+|t_{++}|^2 + \cos\theta_-|t_{+-}|^2\Big), \tag{18.71}$$

$$\sqrt{\varepsilon_1/\mu_1}\cos\theta_1\Big(1 - |r_{--}|^2 - |r_{-+}|^2\Big) = \sqrt{\varepsilon/\mu}\Big(\cos\theta_+|t_{-+}|^2 + \cos\theta_-|t_{--}|^2\Big).$$

Since plane-polarized waves are not eigenstates within the chiral medium, the corresponding relations involving r_{ss}, r_{sp}, r_{ps} and r_{pp} are of a hybrid form:

$$\sqrt{\varepsilon_1/\mu_1}\cos\theta_1\Big(1 - |r_{ss}|^2 - |r_{sp}|^2\Big) = 2\sqrt{\varepsilon/\mu}\Big(\cos\theta_+|t_{s+}|^2 + \cos\theta_-|t_{s-}|^2\Big), \tag{18.72}$$

$$\sqrt{\varepsilon_1/\mu_1}\cos\theta_1\Big(1 - |r_{pp}|^2 - |r_{ps}|^2\Big) = 2\sqrt{\varepsilon/\mu}\Big(\cos\theta_+|t_{p+}|^2 + \cos\theta_-|t_{p-}|^2\Big).$$

The reason for the factor of 2 on the right-hand sides of (18.72) lies in our definition of the electric fields and of the transmission amplitudes: from (18.54) we see that $|\mathbf{E}_{\pm}|^2 = 2$, while the incoming s-polarized or p-polarized electric fields are normalized to unity (see (18.14) and (18.16)).

18.4.3 The Angles θ_B, θ_{pp}, and θ_{pol}

At the boundary between two non-chiral media, zero reflection of a p-polarized incident wave occurs at the Brewster angle θ_B, where

$$\tan^2\theta_B = \frac{\varepsilon}{\varepsilon_1}\left(\frac{\varepsilon\mu_1 - \varepsilon_1\mu}{\varepsilon\mu - \varepsilon_1\mu_1}\right) = \frac{\varepsilon\mu(m^2-1)}{\varepsilon\mu - \varepsilon_1\mu_1}, \quad m = \sqrt{\frac{\varepsilon\mu_1}{\varepsilon_1\mu}}. \tag{18.73}$$

The result (18.73) follows from setting to zero the achiral limit of the reflection amplitude r_{pp} given in (18.65). It is a generalization of familiar formula $\tan^2\theta_B = \varepsilon/\varepsilon_1$ of Sect. 1.2, to which it simplifies when $\mu = \mu_1$. A physical zero can occur only when the right-hand side of (18.73) is positive.

For the achiral–chiral interface we can ask for the angle θ_{pp} at which r_{pp} is zero, in analogy with the anisotropic crystal case discussed in Sect. 8.4. From (18.65) we see that this occurs when

$$2(c_+c_- - c_1^2)m = c_1(c_+ + c_-)\left(m^2 - 1\right) \tag{18.74}$$

The squares of the cosines of the angles of incidence and refraction can be expressed in terms of $s_1^2 = \sin^2\theta_1$ (by use of Pythagoras' theorem, and of Snell's law $n_1\sin\theta_1 = n_\pm\sin\theta_\pm$):

$$c_1^2 = 1 - s_1^2, \quad c_\pm^2 = 1 - (n_1/n_\pm)^2 s_1^2 \tag{18.75}$$

If we square both sides of (18.74), isolate the product c_+c_-, and then square again, we will obtain an algebraic equation for s_1^2. This turns out to a quartic in s_1^2, or equivalently a quartic in $t_1^2 = \tan^2\theta_1$, one of the solutions of which gives $\tan^2\theta_{pp}$. When we substitute (18.43) into the quartics, we find that θ_{pp} is given by the right-hand side of (18.73) plus a term of order γ^2. Because γ is small for natural optically active media, the second-order difference between θ_B and θ_{pp} is usually not of experimental interest.

Another angle of interest is the *polarizing angle* at which a monochromatic plane wave of arbitrary polarization becomes linearly polarized on reflection, mentioned in Sect. 8.4. For a linearly polarized wave, the angle ϕ between the electric field vector $\mathbf{E}$ and the $\mathbf{p}$ direction is given by $\tan\phi = \mathbf{s}\cdot\mathbf{E}/\mathbf{p}\cdot\mathbf{E}$, where $\mathbf{p}$ and $\mathbf{s}$ are unit vectors as before. For the reflected wave the azimuthal angle is given by $\tan\phi' = \mathbf{s}'\cdot\mathbf{E}'/\mathbf{p}'\cdot\mathbf{E}'$ where

$$\mathbf{E}' = (\mathbf{p}\cdot\mathbf{E})\left(r_{pp}\mathbf{p}' + r_{ps}\mathbf{s}'\right) + (\mathbf{s}\cdot\mathbf{E})\left(r_{sp}\mathbf{p}' + r_{ss}\mathbf{s}'\right).$$

Thus

$$\tan\phi' = \frac{r_{ps} + r_{ss}\tan\phi}{r_{pp} + r_{sp}\tan\phi}. \tag{18.76}$$

The condition for ϕ' to be independent of ϕ is

$$r_{pp}r_{ss} - r_{ps}r_{sp} = 0. \tag{18.77}$$

The condition (18.77) guarantees that the same polarization azimuth ϕ' will result for all incident azimuths ϕ, namely the ϕ' given by

$$\tan\phi' = \frac{r_{ps}}{r_{pp}} = \frac{r_{ss}}{r_{sp}} \tag{18.78}$$

For isotropic non-chiral media, the condition (18.77) reduces to $r_{pp}r_{ss} = 0$, which for $\mu = \mu_1$ is satisfied by $r_{pp} = 0$ at angle of incidence θ_B given by $\tan^2\theta_B = \varepsilon/\varepsilon_1$, and gives an *s*-polarized reflected wave. For chiral media (18.77) implies

$$2(c_+c_- + c_1^2)m = c_1(c_+ + c_-)\left(m^2 + 1\right). \tag{18.79}$$

The same method that we outlined for the $r_{pp} = 0$ case (which satisfies (18.74)) reduces (18.79) to a quadratic in $\sin^2\theta_1$, or equivalently, to a quadratic in $\tan^2\theta_1$. As we found for $\tan^2\theta_{pp}$, the polarizing angle determined by (18.79) is given by (18.73) plus a term of second order in the chirality parameter γ . Thus measurement of either the Brewster angle or of the polarizing angle is not a viable method of determining a small γ. The full formula for the angle of incidence at which the reflected light is linearly polarized is

$$\tan^2\theta_{pol} = \frac{(m^2-1)\left\{(m^2-1)\left[2n_+^2n_-^2 - n_1^2\left(n_+^2 + n_-^2\right)\right] + 2(m^2+1)n_+n_-\sqrt{\left(n_+^2 - n_1^2\right)\left(n_-^2 - n_1^2\right)}\right\}}{4\,m^2(n_+^2 - n_1^2)(n_-^2 - n_1^2)} \tag{18.80}$$

The symbols have their usual meaning: $n_1 = \sqrt{\varepsilon_1\mu_1}, n_\pm = \sqrt{\varepsilon\mu} \pm \gamma, m = \sqrt{\varepsilon\mu_1/\varepsilon_1\mu}$.

As we noted in relation to $\tan^2\theta_B$, a physical polarizing angle can occur only when the right-hand side of (18.80) is positive.

18.5 Optical Properties of a Chiral Layer

The optical properties of a chiral layer are discussed by Basisri et al. (1988), Jaggard and Sun (1992), Lindell et al. (1994), Silverman and Badoz (1994), among others. Lekner (1996) gives a first-principles derivation of exact *analytic* expressions for the reflection and transmission amplitudes. Here we shall just give an outline of the method and discuss some special cases.

We consider reflection and transmission by an optically active layer of thickness d, between the medium of incidence with index of refraction $n_1 = (\varepsilon_1 \mu_1)^{1/2}$ and the substrate with index $n_2 = (\varepsilon_2 \mu_2)^{1/2}$. The layer lies between $z = 0$ and $z = d$, and is characterized by two indices $n_\pm = (\varepsilon\mu)^{1/2} \pm \gamma$. Because of multiple reflections within the layer, the electromagnetic field inside is made up of four eigenstates or modes, two propagating in the positive z direction, and two in the negative z direction. For each direction of propagation there are two possible helicities. The electric fields have the space and time dependence

$$\mathbf{E}^{\mathbf{f}}_{\pm} \exp i(Kx + q_\pm z - \omega t), \qquad \mathbf{E}^{\mathbf{b}}_{\pm} \exp i(Kx - q_\pm z - \omega t), \tag{18.81}$$

where the superscripts f and b denote forward and backward propagation inside the slab, and $q_\pm$ are given by (18.40).

The reflection and transmission amplitudes are found by applying the continuity of the tangential (x and y) components of $\mathbf{E}$ and $\mathbf{H}$ at the two boundaries $z = 0$ and $z = d$ of the chiral slab. The s wave in the medium of incidence has electric and magnetic fields given by (18.56) and (18.58). The electric field inside the slab is

$$\mathbf{E} = f_+ \mathbf{E}^{\mathbf{f}}_+ + f_- \mathbf{E}^{\mathbf{f}}_- + b_+ \mathbf{E}^{\mathbf{b}}_+ + b_- \mathbf{E}^{\mathbf{b}}_-. \tag{18.82}$$

There are eight equations (given in Lekner 1996) arising from the boundary conditions, and these determine the eight unknowns (for s-polarization incident), namely $r_{ss}, r_{sp}, t_{ss}, t_{sp}, f_+, f_-, b_+, b_-$. Likewise for the p polarization. These simultaneous equations are solved by using mode, phase and layer matrices, as for the anisotropic layer (Sect. 8.9). The general solution is given in the Appendix of Lekner (1996). The conservation laws (18.20) are satisfied by the exact reflection and transmission amplitudes.

18.5.1 Normal Incidence

The simplest special case is that of normal incidence, for which one obtains

$$r_{ss} = r_{pp} = \frac{r + r' Z_+ Z_-}{1 + r r' Z_+ Z_-}, \qquad r_{ps} = r_{sp} = 0, \tag{18.83}$$

$$t_{ss} = t_{pp} = \frac{(1+r)(1+r')(Z_+ + Z_-)/2}{1 + r r' Z_+ Z_-},$$

$$t_{ps} = -t_{sp} = \frac{i(1+r)(1+r')(Z_+ - Z_-)/2}{1 + r r' Z_+ Z_-}.$$

In these formulae r and r' are the normal incidence reflection amplitudes at the first and second interfaces,

$$r = \frac{1-m}{1+m}, \quad r' = \frac{m'-1}{m'+1}, \quad m = \sqrt{\frac{\varepsilon\mu_1}{\varepsilon_1\mu}}, \quad m' = \sqrt{\frac{\varepsilon\mu_2}{\varepsilon_2\mu}}. \tag{18.84}$$

Z_+ and Z_- are the phase factors for waves of positive and negative helicity traversing the layer:

$$Z_+ = \exp(iq_+ d), \quad Z_- = \exp(iq_- d). \tag{18.85}$$

At normal incidence $q_\pm = n_\pm \omega/c$, from (18.40) and (18.42). From (18.22), (18.24) and (18.83) we find that, at normal incidence

$$r_{++} = r_{--} = 0, \qquad r_{+-} = r_{-+} = -r_{ss} = -r_{pp}. \tag{18.86}$$

The normal incidence transmission amplitudes characterized by helicity reduce to (on using (18.26), (18.27) and (18.83))

$$t_{++} = \frac{(1+r)(1+r')Z_+}{1+rr'Z_+Z_-}, \; t_{--} = \frac{(1+r)(1+r')Z_-}{1+rr'Z_+Z_-}, t_{+-} = t_{-+} = 0. \tag{18.87}$$

A chiral slab will thus transmit a normally incident pure circularly polarized wave without mixing in any of the opposite circular polarization. A linearly polarized wave can be regarded as an equal mix of the two opposite circular polarizations (for example, $\mathbf{p} = (\mathbf{p}+i\mathbf{s})/2 + (\mathbf{p}-i\mathbf{s})/2$). On transmission through the slab, the positive and negative helicities are phase-shifted by different amounts, so that a wave of unit amplitude initially linearly polarized along $\mathbf{p}$ will after transmission through the slab have amplitude

$$\begin{aligned}\frac{(\mathbf{p}+i\mathbf{s})t_{++}}{2} + \frac{(\mathbf{p}-i\mathbf{s})t_{--}}{2} &= \frac{(1+r)(1+r')}{1+rr'Z_+Z_-}[\mathbf{p}(Z_+ + Z_-) + i\mathbf{s}(Z_+ - Z_-)] \\ &= \frac{(1+r)(1+r')}{1+rr'Z_+Z_-}\exp(in\omega d/c)[\mathbf{p}\cos\delta - \mathbf{s}\sin\delta].\end{aligned} \tag{18.88}$$

Here $n = (n_+ + n_-)/2 = \sqrt{\varepsilon\mu}$ is the average index in the chiral medium, and

$$\delta = \frac{(n_+ - n_-)\omega d}{2c} = \gamma\frac{\omega d}{c} = \gamma\frac{2\pi d}{\lambda}. \tag{18.89}$$

From (18.88) and (18.89) (which prove (18.10)) we see that the plane of polarization is rotated by δ. For propagation along the optic axis of crystalline quartz, for example, the rotation is 18.8° per mm at $\lambda = 633$ nm, so that $n_+ - n_- \approx 6.6 \times 10^{-5}$ and $\gamma \approx 3.3 \times 10^{-5}$.

We note that at normal incidence the multiple reflections within the slab have no effect on the rotation of the plane of polarization, although they do affect the amount of light transmitted. The situation is more complicated at oblique incidence: the ratio of t_{++} to t_{--} no longer equals Z_+/Z_-, and also t_{+-} and t_{-+} are not zero.

18.5.2 Optical Properties Near the Critical Angles

Enhancement of chirality effects in the vicinity of the critical angles has been noted by Silverman and Badoz (1989, 1992), and by Badoz and Silverman (1992). Here we give the reflection amplitudes for a bulk chiral medium. Section 4.5 of Lekner (1996) gives the reflection amplitudes for a chiral layer.

If the medium of incidence has refractive index n_1 greater than one or both of the indices $n_\pm$, there will be an angle of incidence at which only one of the helicities can propagate within the chiral medium. Suppose that $\gamma > 0$ $(n_+ > n_-)$. Then the negative helicity wave will be the first to decay exponentially within the bulk chiral medium, at angles of incidence greater than the critical angle θ_1^- given by

$$\sin\theta_1^- = n_-/n_1. \tag{18.90}$$

At this critical angle of incidence $c_- = \cos\theta_-$ is zero, and, with $n = (n_+ + n_-)/2 = \sqrt{\varepsilon\mu}$ as before,

$$c_+ = \cos\theta_+ = \frac{2}{n_+}\sqrt{\gamma n} \approx 2\sqrt{\frac{\gamma}{n}}. \tag{18.91}$$

At $\theta_1 = \theta_1^-$ the reflection amplitudes for the bulk medium, (18.61) and (18.65), have the following form:

$$\begin{aligned}
r_{ss} &= \frac{2mc_1 + c_+(1-m^2)}{2mc_1 + c_+(1+m^2)} = 1 - \frac{mc_+}{c_1} + O\left(c_+^2\right),\\
r_{pp} &= -\frac{2mc_1 - c_+(1-m^2)}{2mc_1 + c_+(1+m^2)} = -1 + \frac{c_+}{mc_1} + O\left(c_+^2\right),\\
r_{ps} &= r_{sp} = \frac{-2imc_+}{2mc_1 + c_+(1+m^2)} = -\frac{ic_+}{c_1} + O\left(c_+^2\right).
\end{aligned} \tag{18.92}$$

Thus r_{ps} and r_{sp} are proportional to the square root of the small chirality parameter γ, and measurements at $\theta_1 = \theta_1^-$ may be used to determine γ.

References

Badoz J, Silverman MP (1992) Differential reflection of circularly polarized light from a naturally optically active medium. SPIE 1746:247–258

Bassiri S, Papas CH, Engheta N (1988) Electromagnetic wave propagation through a dielectric–chiral interface and through a chiral slab. J Opt Soc Am A 5:1450–1459

Bohren CF (1974) Light scattering by an optically active sphere. Chem Phys Lett 29:458–462

Born M (1972) Optik. Springer, Berlin, p 412

Condon EU (1937) Theories of optical rotatory power. Rev Mod Phys 9:432–457

Eftimiu C, Pearson LW (1989) Guided electromagnetic waves in chiral media. Radio Sci. 24:351–359

Jaggard DL, Mickelson AR, Papas CJ (1979) On electromagnetic waves in chiral media. Appl. Phys. 18:211–216

Jaggard DL, Sun X (1992) Theory of chiral multilayers. J Opt Soc Am 9:804–813

Lakhtakia A, Varadan VV, Varadan VK (1988) Field equations, Huygen's principle, integral equations, and theorems for radiation and scattering of electromagnetic waves in isotropic chiral media. J Opt Soc Am A 5:175–184

Lakhtakia A, Varadan VK, Varadan VV (1989) Time-harmonic electromagnetic fields in chiral media. Springer-Verlag, New York

Lakhtakia A (ed) (1990) Selected papers on natural optical activity, MS15. SPIE Optical Engineering Press, Bellingham

Lakhtakia A (1994) Beltrami fields in chiral media, World Scientific, Singapore

Lekner J (1996) Optical properties of isotropic chiral media. Pure Appl Opt 5:417–443

Lindell IV, Sihvola AH, Tretyakov SA, Viitanen AJ (1994) Electromagnetic waves in chiral and bi-isotropic media. Artech House, Boston

Lowry TM (1935) Optical rotatory power. Longmans and Green, 1964 Reprinted New York, Dover

Lukyanov AY, Novikov MA (1990) Reflection of light from the boundary of chiral gyrotropic medium. JETP Lett 51:673–674

Maxwell JC (1891/1954) A treatise on electricity and magnetism, section 794. Dover, New York

Silverman MP (1986) Reflection and refraction at the surface of a chiral medium: comparison of gyrotropic constitutive relations invariant or noninvariant under a duality transformation. J Opt Soc Am A 3:830–837

Silverman MP, Ritchie N, Cushman GM, Fisher B (1988) Experimental configurations using optical phase modulation to measure chiral asymmetries in light specularly reflected from a naturally gyrotropic medium. J Opt Soc Am A: 5:1852–1862

Silverman MP, Badoz J (1989) Large enhancement of chiral asymmetry in light reflection near critical angle. Opt Commun 74:129–133

Silverman MP, Badoz J (1992) Multiple reflection from isotropic chiral media and the enhancement of chiral asymmetry. J Electromag Waves Appl 6:587–601

Silverman MP, Badoz J, Briat B (1992) Chiral reflection from a naturally optically active medium. Opt Lett 17:886–888

Silverman MP (1993) And yet it moves: strange systems and subtle questions in physics, section 4.3. Cambridge University Press, Cambridge

Silverman MP, Badoz J (1994) Interferometric enhancement of chiral asymmetries: ellipsometry with an optically active Fabry-Perot interferometer. J Opt Soc Am A 11:1894–1917

Chapter 19
Pulses and Wavepackets

The preceding chapters have dealt with the reflection of monochromatic plane waves from planar interfaces. Here we consider the reflection and transmission of *electromagnetic pulses* and of *quantum particle wavepackets* by stratified media. The theory of pulse reflection is simplest for those still having a plane wave spatial character but bounded in time (or, equivalently, bounded in spatial extent along the direction of propagation at a given time). Such pulses are built up by a superposition of plane waves of differing frequencies. We shall find, accordingly, that the reflection of such pulses is determined by the frequency dependence of the reflection amplitude. Particularly important is the case of total reflection, where all the frequency and angle dependence is contained in the phase of the reflection amplitude, since its modulus is then unity.

Sections 19.2 and 19.3 deal with exact solutions of the time-dependent Schrödinger equation for wavepackets at potential barriers. The Appendix summarizes the known universal properties of electromagnetic wavepackets.

19.1 Reflection of Nearly Monochromatic Pulses: The Time Delay

For simplicity (and in this section only) we consider an electromagnetic pulse which is launched and detected at some fixed position, here taken to be at the plane $z = 0$. The initial pulse $E_i(t)$ may be written as a superposition of monochromatic plane waves by means of the Fourier integral:

$$E_i(t) = \int_{-\infty}^{\infty} \mathrm{d}\omega f(\omega) \mathrm{e}^{-i\omega t}. \tag{19.1}$$

J. Lekner, *Theory of Reflection*, Springer Series on Atomic, Optical, and Plasma Physics 87, DOI 10.1007/978-3-319-23627-8_19

The Fourier inverse of (19.1) is

$$f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}t\, E_i(t) \mathrm{e}^{i\omega t}. \tag{19.2}$$

For example, if the initial pulse is sinusoidal with amplitude $A(t)$,

$$E_i(t) = A(t)\mathrm{e}^{-i\omega_0 t}, \quad f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}t\, A(t) \mathrm{e}^{i(\omega-\omega_0)t}. \tag{19.3}$$

If $A(t)$ varies slowly over most of its range, this represents a wavepacket which is nearly monochromatic. The simplest case is a truncated sine wave, for which $A(t) = A_0$ when $-T/2 < t < T/2$, and $A(t) = 0$ otherwise; then

$$f(\omega) = A_0 \frac{\sin(\omega - \omega_0)T/2}{\pi(\omega - \omega_0)}. \tag{19.4}$$

Each Fourier component reflects with its own reflection amplitude $r(\omega)$. Thus if (19.1) represents the initial pulse at $z = 0$, the reflected pulse at $z = 0$ will be given by

$$E_r(t) = \int_{-\infty}^{\infty} \mathrm{d}\omega\, r(\omega) f(\omega) \mathrm{e}^{-i\omega t}. \tag{19.5}$$

For a sinusoidal pulse with amplitude $A(t)$ this may be written as

$$E_r(t) = \int_{-\infty}^{\infty} \mathrm{d}\omega\, r(\omega) \mathrm{e}^{-i\omega t} \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\tau\, A(\tau) \mathrm{e}^{i(\omega-\omega_0)\tau}. \tag{19.6}$$

We now specialize further to the case where the modulus of $r(\omega)$ is slowly varying compared to the phase, and the frequency variation of the phase is adequately approximated by the first term in its Taylor expansion about $\omega_0 : r(\omega) = |r(\omega)|\mathrm{e}^{i\delta(\omega)}$, with

$$|r(\omega)| \approx |r(\omega_0)|, \quad \delta(\omega) \approx \delta_0 + (\omega - \omega_0)\delta_0^{'}. \tag{19.7}$$

Here $\delta_0 = \delta(\omega_0)$ and $\delta_0^{'}$ is the derivative $\mathrm{d}\delta/\mathrm{d}\omega$ evaluated at ω_0. (The approximation (19.7) is particularly suited to the treatment of total reflection, where $|r(\omega)| = 1$.) On substituting (19.7) in (19.6), $(2\pi)^{-1}$ times the integral over ω becomes a delta function, which selects the time, $\tau = t - \delta_0^{'}$ in the τ integral. Thus

$$E_r(t) \approx |r(\omega_0)| e^{i(\delta_0 - \omega_0 t)} A\left(t - \delta_0'\right). \tag{19.8}$$

The reflected pulse in this approximation is thus decreased in amplitude by $|r(\omega_0)|$ and phase shifted by δ_0. The pulse is unchanged in shape (it has the same time envelope A), but is delayed by the *group delay time*

$$\Delta t \approx \delta_0' = \left(\frac{d\delta}{d\omega}\right)_{\omega=\omega_0}. \tag{19.9}$$

We will consider some examples of the application of (19.9). The simplest case is that of reflection from a discontinuity in the refractive index. If the discontinuity occurs at z_1, the reflection amplitude at normal incidence is, from (1.15),

$$r_n = e^{2in_1(\omega/c)z_1} \frac{n_1 - n_2}{n_1 + n_2}. \tag{19.10}$$

The phase is thus a constant (0 or $\pm\pi$ depending on the sign of $n_1 - n_2$), plus $2n_1 z_1 \omega/c$, and the delay time according to (19.9) is $2n_1 z_1/c$, equal to the distance $2z_1$ travelled from $z = 0$ to z_1 and back, divided by the speed c/n_1.

The above example is special because the medium is homogeneous everywhere except at the discontinuity. In the general case of reflection by a stratified medium we showed in Sect. 6.7 that in the short wave limit, the phase shift on total reflection is approximately

$$\delta \approx 2(\phi_0 - \phi_- - \pi/4), \tag{19.11}$$

where

$$\phi_0 = \int_{z_0}^{z_0} d\zeta q(\zeta), \quad \phi_- = \lim_{z\to-\infty} \left\{ \int_{z_0}^{z} d\zeta q(\zeta) - q_1 z \right\}. \tag{19.12}$$

The lower limit in the phase integral is arbitrary. It is convenient to set it equal to z_0 (the classical turning point, at which $q = 0$). Then $\phi_0 = 0$, and ϕ_- may be written as

$$\phi_- = \lim_{z\to-\infty} \int_{z_0}^{z} d\zeta (q(\zeta) - q_1) - q_1 z_0. \tag{19.13}$$

If $q(z) = q_1$ at and below the observation point $z = 0$, the phase of the reflection amplitude becomes

$$\delta \approx \int_0^{z_0} dz\, q(z, \omega) - \pi/2. \tag{19.14}$$

At normal incidence (vertical propagation in the case of pulses reflected from the ionosphere),

$$\delta \approx \frac{\omega}{c} \int_0^{z_0} \mathrm{d}z\, n(z,\omega) - \pi/2. \tag{19.15}$$

The time delay can be written in terms of the *group velocity* $u(z,\omega)$; it is also estimated by (19.9) as

$$\Delta t = 2 \int_0^{z_0} \frac{dz}{u(z,\omega)} \approx \frac{\mathrm{d}\delta}{\mathrm{d}\omega} \approx \frac{2}{c} \int_0^{z_0} \mathrm{d}z \left[n(z,\omega) + \omega \frac{\partial n}{\partial \omega} \right]. \tag{19.16}$$

(The turning point z_0 is also a function of ω, but its derivative is multiplied by $n(z_0,\omega)$, which is zero.) The approximate equality of integrals in (19.16) shows that the pulse travels to the turning point and back at the group velocity

$$u(z,\omega) = \frac{c}{n(z,\omega) + \omega \partial n/\partial \omega}. \tag{19.17}$$

Hence the name *group delay time* given to Δt. Equation (19.17) is equivalent to the usual definition of group velocity, $u = \mathrm{d}\omega/\mathrm{d}k$, since here $k = n\omega/c$. In the simplest model of the ionosphere,

$$\varepsilon(z,\omega) = n^2(z,\omega) = 1 - \omega_p^2(z)/\omega^2; \tag{19.18}$$

then the group velocity u and the phase velocity $v = c/n$ are related by

$$uv = c^2. \tag{19.19}$$

The above derivation of the group delay time is based, in part, on Ginzburg (1964, Sect. 19.21). An alternative treatment may be found in Budden (1961, Chap. 10, and 1985, Chap. 5); a general discussion of phase, group, signal and energy transport velocities is given by Brillouin (1960).

So far we have considered only the linear term in the Taylor expansion

$$\delta(\omega) \approx \delta_0 + (\omega - \omega_0)\delta_0' + \frac{1}{2}(\omega - \omega_0)\delta_0'' + \ldots. \tag{19.20}$$

The first order term leads to the time delay discussed above; the second and higher order terms cause pulse spreading and distortion. These effects are discussed by Budden and Ginzburg in the limit (common in optics and radio) where the pulse is nearly monochromatic. The opposite extreme is common in underwater acoustics and in seismology, where explosive sources or sudden crust movements give pulses

which are strongly localized, and not at all harmonic. There is then no dominant frequency ω_0, and use of expansions such as (19.20) is not appropriate. A discussion of this case and further references may be found in Brekhovskikh (1980, Sect. 19.15).

Particular examples of exact solutions relating to reflection and nonreflection of highly non-monochromatic particle wavepackets are discussed in the next two sections.

19.2 Nonreflection of Wavepackets by a Subset of the sech^2 Potentials

We saw in Sect. 4.3 that the sech^2 profile did not reflect, at any angle of incidence, when the parameter $\alpha = \Delta\varepsilon(\omega a/c)^2 = n(n+1)$, n a positive integer. In the quantum particle context, the potential well

$$V(z) = -\frac{\hbar^2}{2ma^2}\frac{\mu(\mu+1)}{\cosh^2\frac{z}{a}} \tag{19.21}$$

is reflectionless, *at any energy*, if μ is a positive integer. The potential (19.21), or the equivalent dielectric function profile, was first considered by Epstein (1930) and Eckart (1930), and is treated in the quantum mechanics texts Landau and Lifshitz (1965) and Flügge (1974). The $\mu = 1$ form of (19.21) appears as the simplest of a family of reflectionless profiles (Kay and Moses 1956). We shall show later in this section that positive integer values of μ are special because they give critically bound states, with delocalized wavefunctions.

When μ is a positive integer, there is no reflection at any energy, and thus zero reflection of any wavepacket formed by superposition of positive energy eigenstates. Lekner (2007) has reduced the positive energy eigenstates to elementary form, of which examples are given below.

For positive energies we write $E = \hbar^2k^2/2m$, and the Schrödinger equation with potential energy given by (19.21) reads

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}z^2} + \left[k^2 + \frac{\mu(\mu+1)}{a^2\cosh^2 z/a}\right]\psi = 0. \tag{19.22}$$

The potential is even in z, so parity is a good quantum number, and the two independent solutions of (19.22) can be taken to be the even and odd functions. For $\mu = 0$ these are proportional to the cosine and sine of kz, as expected:

$$\psi_0^e = \cos kz, \qquad \psi_0^o = \frac{\sin kz}{ka}. \tag{19.23}$$

For $\mu = 1$ these are

$$\psi_1^e = \cos kz - \tanh\frac{z}{a}\,\frac{\sin kz}{ka}, \tag{19.24}$$

$$\psi_1^o = \left[1 + (ka)^2\right]^{-1}\left\{ka\sin kz + \tanh\frac{z}{a}\cos kz\right\}. \tag{19.25}$$

The even and odd eigenfunctions for $\mu = 1$ and $\mu = 2$ are shown in Figs. 1 and 2 of Lekner (2007).

We can superpose the even and odd eigenstates to obtain the reflectionless energy eigenstates propagating in either the $+z$ or $-z$ directions, for example

$$\psi_0^+ = \psi_0^e + ika\psi_0^o = \mathrm{e}^{ikz}, \tag{19.26}$$

$$\psi_1^+ = \psi_1^e + \frac{i\left[1 + (ka)^2\right]}{ka}\,\psi_1^o = \left[1 + \frac{i}{ka}\tanh\frac{z}{a}\right]\mathrm{e}^{ikz}. \tag{19.27}$$

19.2.1 Construction of Reflectionless Wavepackets

We wish to solve the time-dependent Schrödinger equation $H\Phi = i\hbar\partial_t\Phi$ to follow the passage of a wave packet through the $\mu = 1$ potential. In the absence of a potential ($\mu = 0$) a Gaussian wave packet, starting at $t = 0$ centred on $z = z_0$,

$$\Phi_0(z,0) = \exp\left\{ ik_0(z - z_0) - (z - z_0)^2/2b^2\right\}, \tag{19.28}$$

is known (Kennard 1927; Darwin 1928) to have the time development

$$\Phi_0(z,t) = \frac{b}{\sqrt{b^2 + \frac{i\hbar t}{m}}}\exp\left\{ ik_0\left(z - z_0 - \frac{1}{2}ut\right) - \frac{(z - z_0 - ut)^2}{2\left(b^2 + \frac{i\hbar t}{m}\right)}\right\}. \tag{19.29}$$

In (19.29) $u = \hbar k_0/m$ is the group velocity of the packet, and b gives the spatial extent of the packet at $t = 0$. Figure 19.1 shows the propagation and spreading of the free space Gaussian packet.

A corresponding wave packet built up from the nonreflecting energy eigenstates $\psi_1^+(k,z)$ is (Lekner 2007)

$$\Phi_1(z,t) = \Phi_0(z,t)\left[\frac{a(z - z_0 - ik_0b^2)}{b^2 + i\hbar t/m} + \tanh\frac{z}{a}\right], \tag{19.30}$$

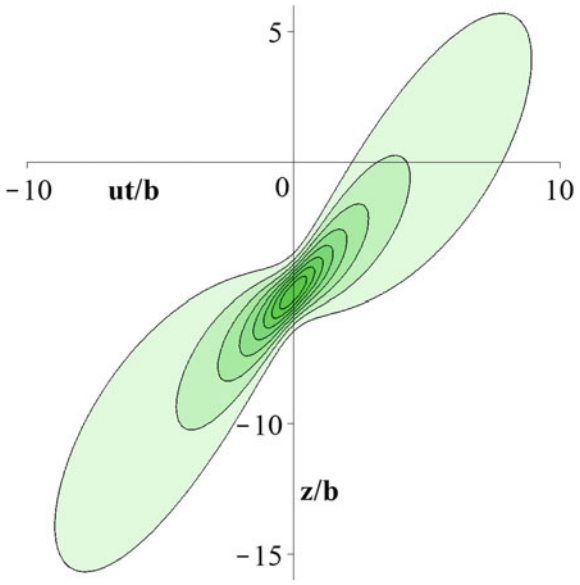


Fig. 19.1 Motion of the free-particle Gaussian wavepacket $\Phi_0(z,t)$ through its focal region. The parameters used are $z_0 = -5b$, $k_0 b = 1$. The time varies from $t = -10b/u$ to $+10b/u$, where $u = \hbar k_0/m$ is the group speed. The position varies from $z = -16b$ to $+6b$. The focal region is centred on $z_0 = -5b$ at $t = 0$. The plot shows contours of the probability density $|\Phi_0|^2$

where Φ_0 is the free-space Gaussian wave packet given in (19.29). The propagation of the packet through the potential well region is illustrated in Fig. 19.2.

Kiriushcheva and Kuzmin (1998) made a numerical study of wave packet propagation in the presence of the $\mu = 1$ potential. They found that a wave packet, constructed to have the form (19.28) at time zero, propagated through the potential region faster than at the group speed $u = \hbar k_0/m$ of the free-space Gaussian solution, and also was narrower after passing through the potential than the free-space Gaussian packet. Lekner (2007) examined the speed and width of the $\Phi_1(z,t)$ wave packet, given in (19.30).

The envelope of this packet is $|\Phi_1(z,t)|$, where

$$|\Phi_1(z,t)|^2 = |\Phi_0(z,t)|^2 \left[\tanh^2 \frac{z}{a} + \frac{2ab^2(z - z_0 - ut)\tanh\frac{z}{a} + a^2\left[(z-z_0)^2 + k_0^2 b^4\right]}{b^4 + (\hbar t/m)^2} \right], \tag{19.31}$$

with

$$|\Phi_0(z,t)|^2 = \frac{b}{\sqrt{b^2 + (\hbar t/mb)^2}} \exp\left\{ \frac{-(z - z_0 - ut)^2}{[b^2 + (\hbar t/mb)^2]} \right\}. \tag{19.32}$$

The Gaussian free-space wave packet has, by inspection of (19.32), group speed $u = \hbar k_0/m$ and width $w_0(t) = \left[b^2 + (\hbar t/mb)^2\right]^{\frac{1}{2}}$. The results for the group speed

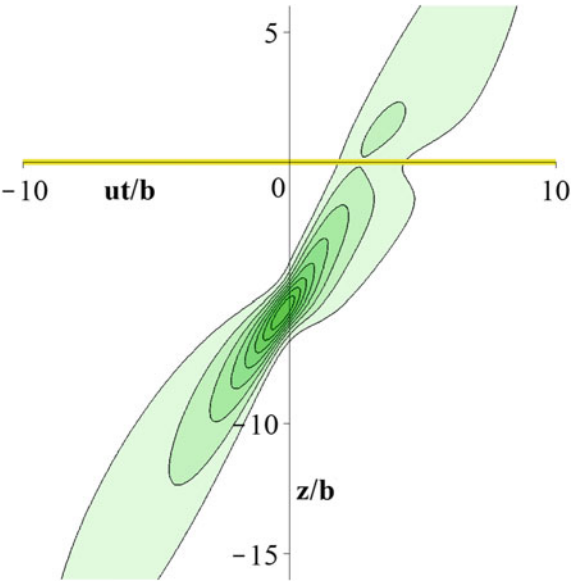


Fig. 19.2 Motion of the non-reflecting wavepacket $\Phi_1(z,t)$ through the potential region; contours of $|\Phi_1|^2$ are plotted. The parameters are as in Fig. 19.1, and $a = b$. The potential $V_1(z)$ is indicated by the band at $z = 0$. Note the constriction in the probability density as the packet passes over the potential well, centred on $z = 0$

and wave packet width of (19.31) are in qualitative agreement with the numerical example given by Kiriushcheva and Kuzmin (1998).

Next we consider the question *what is so special about the integer values of* μ*?* The phenomenon of zero reflection of waves is common in optics, acoustics and quantum mechanics (Lekner 1990). For example antireflection coatings can make reflection zero at one wavelength, or very small over a range of wavelengths. What is rare is zero reflection at *any* wavelength. In optics and acoustics there is zero reflection by a sharp interface at the Brewster and Green angles (Sects. 1.2 and 1.4)

$$\tan^2\theta_B = \frac{\varepsilon_2}{\varepsilon_1}, \qquad \tan^2\theta_G = \frac{(\rho_2 v_2)^2 - (\rho_1 v_1)^2}{\rho_1^2(v_1^2 - v_2^2)}, \tag{19.33}$$

where ε, ρ, v are dielectric constants, densities and sound speeds, respectively. The reflection of the electromagnetic p-wave and of the acoustic wave is however zero only in the limit when the step from ε_1 to ε_2 or (ρ_1, v_1) to (ρ_2, v_2) is very rapid on the scale of the wavelength.

Here we have an example of a potential (or dielectric function profile) with a characteristic length a, and zero reflection for any values of a at any energy, provided μ is an integer. Why?

The potential (19.21) has, for given μ, the bound states (Landau and Lifshitz 1965)

$$E_n = -\frac{\hbar^2}{2ma^2}(\mu - n)^2, \qquad n = 0, 1, 2\ldots[\mu] \tag{19.34}$$

where $[\mu]$ is the integer part of μ. It is shown in Lekner (2007) that the special property of the integer-μ potentials $(-\hbar^2/2ma^2)\mu(\mu+1)\text{sech}^2 z/a$ is that they support a *critically-bound state*: one that has zero energy and a wavefunction of infinite range (Lekner 1972). Systems which are near critical binding have special properties associated with the long range of the wavefunction. This is one reason for the nonreflecting integer-μ sech^2 potentials. Another is the $V(-z) = V(z)$ symmetry of the potential: reflection amplitudes of symmetric profiles automatically have coincident zeros of their real and imaginary parts, whereas general profiles do not (Lekner 1990).

The above association between critical binding strengths of the potentials and zero reflection adds a physical heuristic to the mathematical explanation of supersymmetric quantum mechanics. In the latter the potentials (19.21) and, with integer n,

$$\frac{-\hbar^2}{2ma^2}(\mu-n)(\mu-n+1)\text{sech}^2\frac{z}{a} \tag{19.35}$$

are shown to be partners in supersymmetric algebra. If one of the partners has zero reflection amplitude they all do (Cooper et al. 2001, (3.32)). When μ is an integer one of the potentials will be zero, and a null potential does not reflect, so all the integer-μ potentials are nonreflecting. Cox and Lekner (2008) obtain the nonreflecting eigenstates of the sech^2 potential directly from supersymmetric considerations.

19.3 Exact Solutions of Total and Partial Reflection of Wavepackets

Closed-form solutions of the time-dependent Schrödinger equation can be obtained, describing the propagation of wavepackets in the neighbourhood of the potentials with spatial dependence $1/z^2$ and $\delta(z)$, respectively. The first of these forms an impenetrable barrier and thus causes total reflection. The second gives partial reflection and transmission. Cox and Lekner (2008) obtained the results for the z^{-2} potential from supersymmetric quantum mechanics, and the $\delta(z)$ potential by elementary methods. We shall just state the results to illustrate total and partial reflection of wavepackets.

For the impenetrable potential

$$V(z) = \frac{\hbar^2}{mz^2}, \tag{19.36}$$

which forms a barrier at the origin, we shall consider wavepackets that come up to this barrier from $z = -\infty$ and are totally reflected, with zero probability amplitude

at the $z = 0$ singularity. In the free particle case we build on the stationary wave $e^{ikz} - e^{-ikz}$, which leads to the wavepacket

$$\tilde{\Phi}_0(z,t) = \Phi_0(z,t) - \Phi_0(-z,t). \tag{19.37}$$

Cox and Lekner show that the wavepacket in the presence of the potential (19.36) is

$$\begin{aligned} \Phi(z,t) &= \left(-\partial_z + \tfrac{1}{z}\right)[\Phi_0(z,t) - \Phi_0(-z,t)] \\ &= \left[\tfrac{1}{z} + \tfrac{z - z_0 - ik_0 b^2}{b^2 + i\hbar t/m}\right]\Phi_0(z,t) - \left[\tfrac{1}{z} + \tfrac{z + z_0 + ik_0 b^2}{b^2 + i\hbar t/m}\right]\Phi_0(-z,t). \end{aligned} \tag{19.38}$$

The z^{-1} terms in (19.37) do not cause a singularity at the origin, in fact the leading term is $O(z^2)$. Figure 19.3 shows the propagation and total reflection of this wavepacket.

The term proportional to $\Phi_0(z,t)$ in (19.38) has maximum probability at $z \approx z_0 + ut$ and will be dominant at negative times, while the term proportional to $\Phi_0(-z,t)$ has maximum probability at $-z \approx z_0 + ut$ and will be dominant at positive times (assuming that $|z_0|/b$ is not too large), since the wavepacket is confined to $z < 0$.

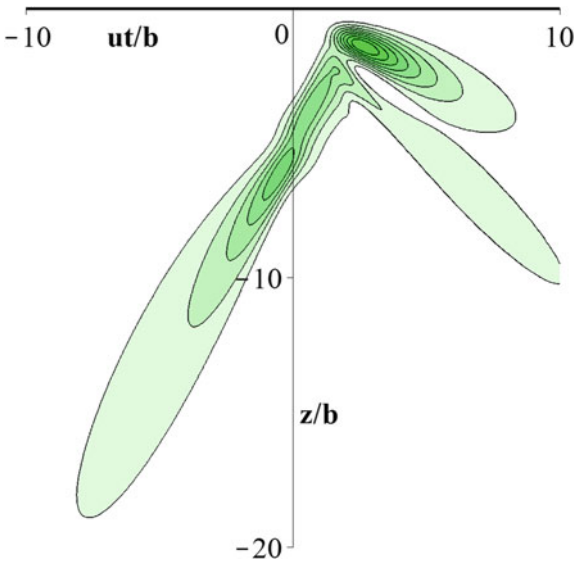


Fig. 19.3 Total reflection of the wavepacket given in (19.38) by the potential $\hbar^2/mz^2$; the parameters are as in Fig. 19.1. Note the large probability $|\Phi|^2$ near $z = -1.5b$, $ut = 2.8b$, greater than at the centre of the focal region because of constructive interference between the incident and reflected parts of the wavepacket. The impenetrable potential is centred on $z = 0$, indicated by the thick horizontal line

Next, we consider the partially reflecting delta function potential

$$V(z) = \frac{\hbar^2 \kappa}{m} \delta(z). \tag{19.39}$$

We characterize the delta function potential by the reciprocal length κ (the magnitude of κ gives the strength, and the sign determines whether the potential is repulsive or attractive):

The delta function causes a discontinuity in the gradient of an energy eigenstate ψ at $z = 0$: setting $E = \hbar^2 k^2/2m$, the time-independent Schrödinger equation becomes

$$\left(-\partial_z^2 + 2\kappa\delta(z)\right)\psi = k^2\psi. \tag{19.40}$$

Integrating across the origin from $-\varepsilon$ to $+\varepsilon$ and letting $\varepsilon \to 0$ gives

$$\psi'(0-) - \psi'(0+) + 2\kappa\psi(0) = 0. \tag{19.41}$$

Let ρ and τ be the plane wave reflection and transmission amplitudes, so that

$$\psi(k,z) = \begin{cases} \mathrm{e}^{ikz} + \rho\,\mathrm{e}^{-ikz} & (z<0) \\ \tau\,\mathrm{e}^{ikz} & (z>0) \end{cases} \tag{19.42}$$

Continuity of ψ at $z = 0$ implies $1 + \rho = \tau$, and the discontinuity in the derivative at $z = 0$ (19.41) gives $ik(1-\rho) - ik\tau + 2\kappa\tau = 0$, so that

$$\rho = \frac{-i\kappa}{k+i\kappa}, \qquad \tau = \frac{k}{k+i\kappa}. \tag{19.43}$$

To construct a wavepacket we superpose the energy eigenstates (19.42) by integrating over k with some Fourier amplitude. In superposing the energy eigenstates we can obtain simple results if, as suggested by the form of (19.43), we use the Fourier amplitude

$$F(k) = (k+i\kappa)F_0(k), \qquad F_0(k) = b\mathrm{e}^{-ikz_0 - \frac{1}{2}(k-k_0)^2 b^2}. \tag{19.44}$$

Then we have (with Φ_0 the free-space Gaussian, as before)

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{d}k\, \mathrm{e}^{ikz - \frac{ik^2\hbar t}{2m}} F_0(k) = \Phi_0(z,t), \tag{19.45}$$

and, by differentiation of (19.45) with respect to z,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{d}k\, \mathrm{e}^{ikz - ik^2\hbar t/2m} ikF_0(k) = \partial_z \Phi_0 = -\frac{z - z_0 - ik_0 b^2}{b^2 + i\hbar t/m}\Phi_0. \tag{19.46}$$

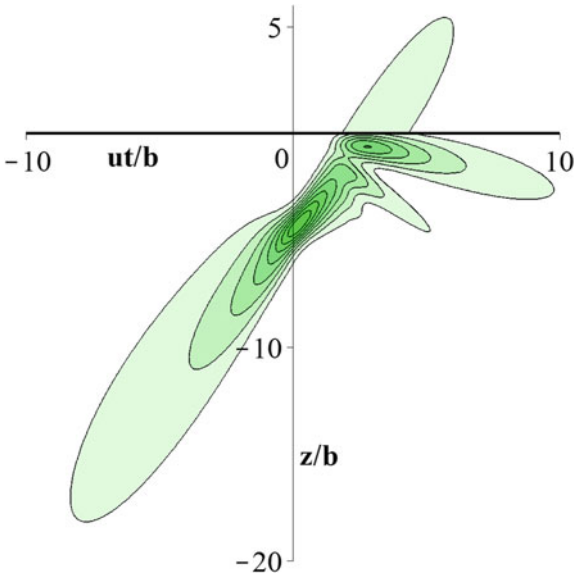


Fig. 19.4 Partial reflection of a wavepacket by the delta function repulsive potential. The parameters are $z_0 = -5b$, $k_0 b = 1$ and $\kappa b = 1$. The delta function potential (at zero z) is indicated by the thick line. The slope in the diagram gives speed; the incident pulse has slope $= \hbar k_0/m$, the group speed. Note that, for this parameter set, the reflected parts separate into a slower and a faster group

The superposition of $\psi(k,z)$ given by (19.42), with Fourier amplitude $(k+i\kappa)F_0(k)$, gives the wavepackets (we remove a factor i)

$$\Phi(z,t) = \begin{cases} (-\partial_z + \kappa)\Phi_0(z,t) - \kappa\Phi_0(-z,t) & z < 0 \\ -\partial_z \Phi_0(z,t) & z > 0 \end{cases} \tag{19.47}$$

Note that one part of the wavepacket, namely $-\partial_z\Phi_0(z,t)$, is the same on both sides of the delta function potential. This part propagates straight through the potential. The other parts at negative z, proportional to the potential strength κ, are the forward-propagating packet $\Phi_0(z,t)$, and the backward-propagating packet $\Phi_0(-z,t)$. The three parts on the left overlap when near $z = 0$, producing an interference maximum, whereas the single transmitted part remains smooth on the right. Figure 19.4 illustrates the process.

There is a large variety of patterns that result from the reflection of a wavepacket by a delta function barrier, as described by (19.47). Four lengths characterize the reflection problem: $b, z_0, k_0^{-1}, \kappa^{-1}$. The example we have given corresponds to a highly localized wavepacket, entering its focal region before the barrier. One can move the focal region, change the localization, and change the barrier strength. For example, highly delocalized wavepackets give rise to fringes due to interference between the incoming and reflected waves.

A reflectance R and transmittance T for *the wavepacket as a whole* can be obtained as follows: at large positive time, the part dominant for negative z is $\Phi = -\kappa\Phi(-z,t)$, while the part for positive z is always $\Phi = -\partial_z\Phi_0(z,t)$. Each of these has a probability density current $J(z,t) = \hbar/m\mathrm{Im}\{\Phi^*(z,t)\partial_z\Phi(z,t)\}$. Integrated over z, at large positive times, these currents are respectively proportional to $-(\hbar k_0/m)\kappa^2, (\hbar k_0/m)[k_0^2 + 3/2b^2]$. From these currents and $R+T=1$ we deduce

$$R = \frac{S}{1+S}, \quad T = \frac{1}{1+S}, \quad S = \frac{\kappa^2}{k_0^2 + 3/2b^2}. \tag{19.48}$$

S is the scattering strength of the delta function potential (which is proportional to κ), for a wavepacket characterized by its mean wavenumber k_0 and minimum width b. For the parameters used in Fig. 19.4, $S = 2/5, R = 2/7, T = 5/7$. In the plane wave limit $k_0 b \gg 1$ the reflectance and transmittance given in (19.48) become the plane-wave expressions obtained by taking the absolute squares of ρ and τ in (19.43), evaluated at the dominant wavenumber k_0.

Notice that z_0, the location of the focal centre of the Gaussian wavepacket at $t = 0$, does not enter into (19.48), which was obtained by local integration over well-separated parts of the wavepacket. In this context we note that the coherence length of wavepackets remains unchanged on propagation (Kaiser et al. 1983; Klein et al. 1983); again this result is independent of the location of the focal region.

The reader may have gained the impression that all quantum particle wavepackets are based on the free-space Gaussian packet Φ_0. This is not necessarily so: other exact solutions exist, for example one based on the Airy function (Lekner 2009), but these have a more complicated space-time dependence. Three-dimensional Gaussian wavepacket solutions of Schrödinger's time-dependent equation are known (Darwin 1928); these can be made to rotate (Lekner 2008). No exact results in the reflection of three-dimensional wavepackets by stratified media are known to the author.

Appendix: Universal Properties of Electromagnetic Pulses

This Appendix surveys the existing known universal properties of electromagnetic pulses, namely (i) the time invariance of the total electromagnetic energy U, momentum $\mathbf{P}$ and angular momentum $\mathbf{J}$ of the pulse, and (ii) the inequality $cP_z < U$. (Net pulse propagation is taken to be along the z direction.) In both (i) and (ii) the theorems follow directly from Maxwell's equations.

The conservation of energy, momentum and angular momentum is no surprise, but the inequality $cP_z < U$ implies that all localized electromagnetic pulses have a zero-momentum frame (not a 'rest' frame, waves are never at rest). The above is of course in contradistinction to Einstein's light quantum, for which the momentum

P is purely in one direction, and $cP = U$ (Einstein 1917). The implication seems to be that we cannot form a model of the photon by any pulse wave-function satisfying Maxwell's equations. If the momentum **P** and energy U formed a four-vector $(c\mathbf{P}, U)$, $U^2 - c^2P^2$ would be a Lorentz invariant. This holds for point particles, but not universally for wavepackets. We show however that $u^2 - c^2p^2$ is universally a Lorentz invariant, non-negative at all space-time points (u and **p** are the energy and momentum densities).

We also discuss the helicity of electromagnetic pulses, and the counter-intuitive relation between the helicity and angular momentum of certain exactly calculable examples.

Maxwell's equations, with the electric and magnetic fields expressed in terms of the vector potential $\mathbf{A}(\mathbf{r}, t)$ and scalar potential $\mathrm{V}(\boldsymbol{r}, t)$ via

$$\mathbf{E} = -\nabla \mathrm{V} - \partial_{ct}\mathbf{A}, \quad \mathbf{B} = \nabla \times \mathbf{A}, \tag{19.49}$$

and with **A** and V satisfying the Lorenz condition $\nabla.\mathbf{A} + \partial_{ct}\mathrm{V} = 0$, lead (in free space) to V and the components of **A** all satisfying the wave equation

$$\nabla^2\psi - \partial_{ct}^2\psi = 0. \tag{19.50}$$

Electromagnetic pulses can then be constructed from solutions of (19.50). For example, the choice $\mathrm{V} = \text{constant}, \mathbf{A} = \nabla \times (0, 0, \psi) = \big(\partial_y, -\partial_x, 0\big)\psi$ gives us the transverse electric (TE) pulse with

$$\mathbf{E} = -\partial_{ct}\mathbf{A} = \big(-\partial_y\partial_{ct}, \partial_x\partial_{ct}, 0\big)\psi, \quad \mathbf{B} = \nabla \times \mathbf{A} = \Big(\partial_x\partial_z, \partial_y\partial_z, -\partial_x^2 - \partial_y^2\Big)\psi. \tag{19.51}$$

The wave equation (19.50) has an infinity of solutions, for example $\psi = f(z - ct)$, with f an arbitrary twice-differentiable function. These solutions, and also the textbook plane wave $\exp i(\mathbf{k}.\mathbf{r} - ckt)$ and spherical waves $r^{-1}\exp ik(r \pm ct)$, are not localized in space-time. The spherical wave solutions generalize to $r^{-1}f(r \pm ct)$, with f again any twice-differentiable function. These spherical wave solutions are singular at the origin.

Bateman (1904) obtained a general solution of the wave equation in integral form. For solutions with axial symmetry (independent of the azimuthal angle ϕ) the Bateman solution is, with $\rho = (x^2 + y^2)^{\frac{1}{2}}$ being the distance from the z-axis,

$$\psi(\rho, z, t) = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\theta f(z + i\rho\cos\theta, ct + \rho\sin\theta). \tag{19.52}$$

We outline a proof (different from Bateman's): the wave equation in cylindrical polars, with no azimuthal dependence, reads

$$\left(\partial_\rho^2 + \frac{1}{\rho}\partial_\rho + \partial_z^2 - \partial_{ct}^2\right)\psi = 0. \tag{19.53}$$

Carrying out the partial differentiations in $(\nabla^2 - \partial_{ct}^2)f$, and comparing with $\partial_\theta^2 f$ shows that

$$\left(\partial_\rho^2 + \frac{1}{\rho}\partial_\rho + \partial_z^2 - \partial_{ct}^2\right)f = -\rho^{-2}\partial_\theta^2 f. \tag{19.54}$$

Operating on (19.52) with $\nabla^2 - \partial_{ct}^2$ therefore gives zero:

$$-2\pi\rho^2\left(\nabla^2 - \partial_{ct}^2\right)\psi = \int_0^{2\pi} \mathrm{d}\theta\, \partial_\theta^2 f = [\partial_\theta f]_0^{2\pi} = 0. \tag{19.55}$$

On the propagation axis $\rho = 0$ the pulse wavefunction becomes

$$\psi(0, z, t) = f(z, ct). \tag{19.56}$$

For example, if the on-axis wavefunction takes the form

$$f(z,t) = \frac{ab}{[a - i(z+ct)][b + i(z-ct)]}\psi_0, \tag{19.57}$$

the corresponding full wavefunction obtained by integrating (19.52) is

$$\psi(\rho, z, t) = \frac{ab}{\rho^2 + [a - i(z+ct)][b + i(z-ct)]}\psi_0. \tag{19.58}$$

This wavefunction has been obtained by other means (see references in Lekner 2003).

Conservation Laws, Energy-Momentum Inequalities

The energy, momentum and angular momentum densities of an electromagnetic field, in free space and in Gaussian units, are (Jackson 1975)

$$u(\mathbf{r}, t) = \frac{1}{8\pi}(E^2 + B^2), \qquad \mathbf{p}(\mathbf{r}, t) = \frac{1}{4\pi c}\mathbf{E} \times \mathbf{B}, \qquad \mathbf{j}(\mathbf{r}, t) = \mathbf{r} \times \mathbf{p}(\mathbf{r}, t). \tag{19.59}$$

$\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{B}(\mathbf{r}, t)$ are the real electric and magnetic fields at position $\mathbf{r}$ and time t. The total energy, momentum and angular momentum at time t of an electromagnetic pulse are

$$U = \int \mathrm{d}^3 r\, u(\mathbf{r}, t), \qquad \mathbf{P} = \int \mathrm{d}^3 r\, \mathbf{p}(\mathbf{r}, t), \qquad \mathbf{J} = \int d^3 r\, \mathbf{j}(\mathbf{r}, t). \tag{19.60}$$

It will come as no surprise that these are all conserved quantities: the integrals in (19.60) are all independent of time. The energy and momenta of electromagnetic pulses based on the solution (19.58) of the wave equation were evaluated in Lekner 2003. Proofs of the constancy of U and of $\mathbf{P}$ were sketched in Lekner (2004b). The conservation of angular momentum was proved in Lekner (2004a). In all cases, the proofs follow from taking the time derivatives of the quantities U, $\mathbf{P}$ and $\mathbf{J}$ defined in (19.60), applying Maxwell's free-space equations

$$\begin{array}{ll} \nabla \cdot \mathbf{B} = 0 & \nabla \cdot \mathbf{E} = 0 \\ \nabla \times \mathbf{E} + \partial_{ct} \mathbf{B} = 0 & \nabla \times \mathbf{B} - \partial_{ct} \mathbf{E} = 0 \end{array} \tag{19.61}$$

and using elementary analytical techniques.

In order for the quantities U, $\mathbf{P}$ and $\mathbf{J}$ to *exist* (let alone be conserved), the electromagnetic pulse has to be localized. The first evaluation of U for any localized pulse was in Feng et al. (1999); later evaluation of energy, momentum and angular momentum for various electromagnetic pulses found (Lekner 2003) that all had $U > cP_z$, with the transverse momenta P_x and P_y zero. Thus these pulses could be Lorentz-transformed into their zero momentum frames, in which the pulse converges onto its focal region and then diverges from it, maintaining zero net momentum at all times. The proof that $U > cP_z$ for all localized electromagnetic pulses is elementary (Lekner 2004a): let the total momentum vector $\mathbf{P}$ point along the z direction, and consider the energy and momentum densities $u(\mathbf{r}, t)$ and $p_z(\mathbf{r}, t)$. From (19.59) we have

$$\begin{array}{c} 8\pi(u - cp_z) = \mathbf{E}^2 + \mathbf{B}^2 - 2(\mathbf{E} \times \mathbf{B})_z \\ = E_x^2 + E_y^2 + E_z^2 + B_x^2 + B_y^2 + B_z^2 - 2(E_x B_y - E_y B_x) \\ = (E_x - B_y)^2 + (E_y + B_x)^2 + E_z^2 + B_z^2 \geq 0 \end{array} \tag{19.62}$$

Equality of U and cP_z would require $u - cp_z$ to be zero everywhere and at all times, which from (19.62) requires $E_z = 0 = B_z$ (purely transverse fields) and also $E_x = B_y$ and $E_y = -B_x$. The divergence equations in (19.63) then give

$$-\partial_x E_y + \partial_y E_x = 0 \qquad \text{and} \qquad \partial_x E_x + \partial_y E_y = 0. \tag{19.63}$$

Thus E_x and $-E_y$ would be a Cauchy-Riemann pair in the variables x and y, and satisfy

$$(\partial_x^2 + \partial_y^2) E_x = 0, \qquad (\partial_x^2 + \partial_y^2) E_y = 0. \tag{19.64}$$

Such harmonic functions cannot have a maximum except at the boundary of their domain, and thus cannot be localized in x and y (for any z and t). For localized

electromagnetic pulses we therefore always have the total energy greater than c times the net total momentum,

$$U > cP_z. \tag{19.65}$$

U and $\mathbf{P}$ are defined by (19.60) as spatial integrals, which have been shown to be independent of time in any given inertial frame. If together they formed the four-vector $(c\mathbf{P}, U)$, $U^2 - c^2P^2$ would be a Lorentz invariant, the same in all inertial frames. Such four-vectors exist for point particles, but cannot be associated (in general) with extended wavepackets. Consider however the squares of the volume densities, $u^2(\mathbf{r}, t)$ and $\mathbf{p}^2(\mathbf{r}, t)$. From (19.59) we have

$$\begin{aligned}(8\pi)^2(u^2 - c^2\mathbf{p}^2) &= (E^2 + B^2)^2 - 4(\mathbf{E} \times \mathbf{B})^2 \\ &= (E^2 + B^2)^2 - 4E^2B^2 + 4(\mathbf{E} \cdot \mathbf{B})^2 \\ &= (E^2 - B^2)^2 + 4(\mathbf{E} \cdot \mathbf{B})^2\end{aligned} \tag{19.66}$$

Hence $u^2 - c^2\mathbf{p}^2$ is everywhere non-negative, and further it is a Lorentz invariant, since $E^2 - B^2$ and $\mathbf{E} \cdot \mathbf{B}$ are Lorentz invariants. We shall return to the Lorentz transformation of pulses at the end of the Appendix.

Angular Momentum, Helicity

We have seen that the energy U, momentum $\mathbf{P}$ and angular momentum $\mathbf{J}$ are all conserved (do not change with time) for any electromagnetic pulse in free space. The energy and momentum are also independent of the choice of origin of the spatial coordinates (which are integrated over, see (19.60)). However, the angular momentum does depend on the choice of origin: in the translation $\mathbf{r} \to \mathbf{r} - \mathbf{a}$, $\mathbf{J} \to \mathbf{J} - \mathbf{a} \times \mathbf{P}$. Textbooks make statements such as (Mezbacher 1998, p. 569) 'the photon has vanishing mass and cannot be brought to rest in any Lorentz frame of reference'. As we have seen, any localized electromagnetic pulse satisfying Maxwell's equations does have a zero momentum frame (not a 'rest' frame). In the frame where $\mathbf{P}$ is zero the angular momentum is independent of the choice of origin, and thus we can associate an *intrinsic angular momentum* with a localized electromagnetic pulse.

Suppose (as we have here) that the net momentum of a pulse is along the z-direction, $\mathbf{P} = (0, 0, P_z)$. A Lorentz boost at speed c^2P_z/U, along the z-axis, will bring the pulse to its zero momentum frame. The component J_z of the angular momentum is unchanged in this Lorentz transformation. This is because the four-tensor of angular momentum $J_{ij} = X_iP_j - X_jP_i$ (X_i and P_i represent components of the space-time and momentum-energy four-vectors) has the same structure as the electromagnetic field four-tensor composed of $\mathbf{E}$ and $\mathbf{P}$ (Landau and Lifshitz 1951, Sect. 2.6)

$$[J_{ij}] = \begin{pmatrix} 0 & J_z & -J_y & J_{14} \\ -J_z & 0 & J_x & J_{24} \\ J_y & -J_x & 0 & J_{34} \\ J_{41} & J_{42} & J_{43} & 0 \end{pmatrix} \tag{19.67}$$

where

$$\begin{aligned} J_{41} &= -J_{14} = i(ctP_x - xU/c) \\ J_{42} &= -J_{24} = i(ctP_y - yU/c) \\ J_{43} &= -J_{34} = i(ctP_z - zU/c) \end{aligned} \tag{19.68}$$

For comparison, the field four-tensor, also in the Minkowski notation, is

$$[F_{ij}] = \begin{pmatrix} 0 & B_z & -B_y & -iE_x \\ -B_z & 0 & B_x & -iE_y \\ B_y & -B_x & 0 & -iE_z \\ iE_x & iE_y & iE_z & 0 \end{pmatrix} \tag{19.69}$$

Since B_z is unchanged by a Lorentz boost along the z-axis, J_z will also be unchanged by such a transformation. Thus we can regard the component of the angular momentum along the momentum (J_z, in this Appendix) as the intrinsic angular momentum of the pulse.

The helicity of the pulse is $+1$ if the sign of J_z is the same as that of P_z (in a frame with $P_z \neq 0$), -1 if the signs are opposite. There is no helicity (or the helicity is zero) if J_z is zero.

We shall give some examples of results for electromagnetic pulses based on the wavefunction (19.10). The first is for the $TE + iTM$ pulse for which

$$\mathbf{A} = \nabla \times (0, 0, \psi) = (\partial_y, -\partial_x, 0)\psi, \tag{19.70}$$

$$\mathbf{B} = \nabla \times \mathbf{A} + i\partial_{ct}\mathbf{A}, \qquad \mathbf{E} = i\mathbf{B}. \tag{19.71}$$

(Here $\mathbf{B}(\mathbf{r}, t)$ and $\mathbf{E}(\mathbf{r}, t)$ are complex; their real and imaginary parts are separately solutions of Maxwell's equations.) The total energy, momentum and angular momentum found in Lekner (2003) are

$$U = \frac{\pi}{8}\frac{a+b}{ab}\psi_0^2, \qquad cP_z = \frac{\pi}{8}\frac{a-b}{ab}\psi_0^2, \quad J_z = 0. \tag{19.72}$$

For this pulse, a Lorentz boost at speed βc, $\beta = cP_z/U = (a-b)/(a+b)$, will bring the pulse to its zero-momentum frame.

If instead one takes the vector potential to be

$$\mathbf{A} = \nabla \times [i\psi, \psi, 0], \tag{19.73}$$

with **B** and **E** defined by (19.71) as before, one finds (Lekner 2003)

$$U = \frac{\pi}{8}\frac{a+3b}{a^2}\psi_0^2, \qquad cP_z = \frac{\pi}{8}\frac{a-3b}{a^2}\psi_0^2, \qquad cJ_z = \frac{\pi}{4}\frac{b}{a}\psi_0^2. \tag{19.74}$$

This example shows that non-zero angular momentum can result from a wavefunction without azimuthal dependence: the curl operator supplies the twist.

More complex exact solutions of the wave equation have been tried, and the energy, momentum and angular momentum evaluated (Lekner 2004c, d). There we find the surprising result that when the wavefunction ψ has an $e^{im\phi}$ azimuthal dependence, the helicity is opposite to the sign of m. Since J_z is represented by the operator $-i\hbar\partial_\phi$ in quantum mechanics, $J_z e^{im\phi} = \hbar m e^{im\phi}$, so there the $e^{im\phi}$ dependence produces $J_z = \hbar m$, the same sign as m. It is not understood physically why electromagnetic pulses do the opposite.

Figure 19.5a, b and c illustrate a time sequence of a pulse based on ψ equal to $\rho\, e^{i\phi}/[b + i(z - ct)]$ times the wavefunction in (19.58), with **A** given by (19.70), V constant, and **E** and **B** given by (19.51). The total energy, momentum and angular momentum of the pulse are (Lekner 2004d)

$$U = \frac{\pi}{16}\frac{3a+b}{b^2}\psi_0^2, \qquad cP_z = \frac{\pi}{16}\frac{3a-b}{b^2}\psi_0^2, \qquad cJ_z = -\frac{\pi}{8}\frac{a}{b}\psi_0^2. \tag{19.75}$$

Lorentz Transformation of Pulses

For point particles of mass M, the energy and momentum are related by $U^2 = M^2c^4 + P^2c^2$, and the combination $(c\mathbf{P}, U)$ is a four-vector, meaning that it transforms in the same way as $(\mathbf{r}, ct)$. It follows that $U^2 - c^2P^2$ is a Lorentz invariant, in this case M^2c^4.

Electromagnetic wavepackets are extended objects, evolving in space-time, and the transformation between inertial frames is more complicated. However, as we have seen in equation (19.66), $u^2 - c^2p^2$ is a non-negative Lorentz invariant, for any electromagnetic pulse.

Consider the transformation of a scalar wavefunction such as (19.58). A Lorentz boost along the direction of motion (here along the z-axis) at speed βc leaves the transverse coordinate ρ unchanged, and changes z and t to z' and t':

$$z = \gamma(z' + \beta ct'), \qquad ct = \gamma(ct' + \beta z'), \qquad \gamma = (1 - \beta^2)^{-\frac{1}{2}}. \tag{19.76}$$

The effect is to change the weight of the $z \pm ct$ components of ψ:

$$z + ct = \sqrt{\frac{1+\beta}{1-\beta}}(z' + ct'), \qquad z - ct = \sqrt{\frac{1-\beta}{1+\beta}}(z' - ct'). \tag{19.77}$$

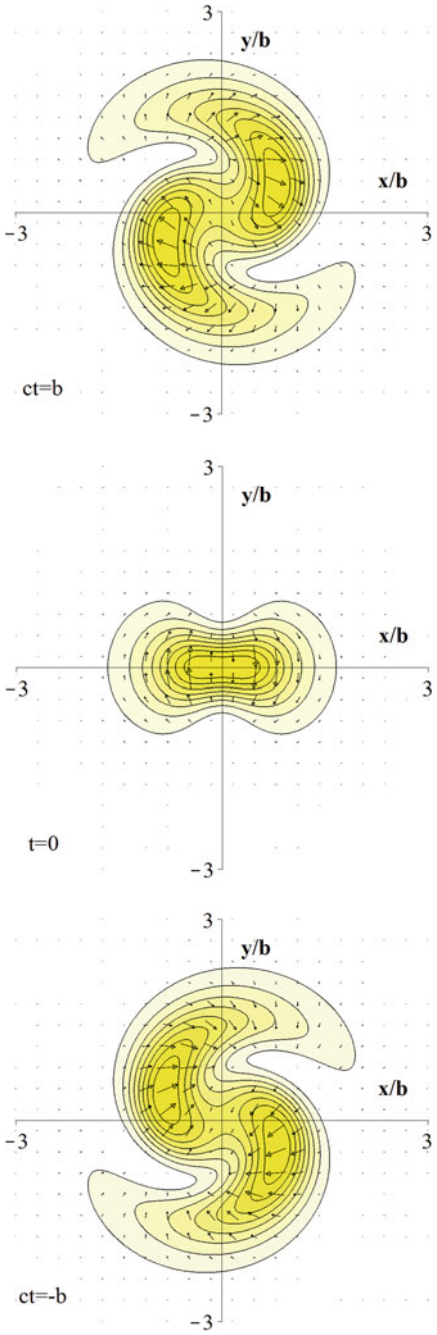


Fig. 19.5 The energy density contours and transverse momentum densities of a helical pulse (wavefunction given in the text), with $a = 2b$. The longitudinal component p_z of the momentum density is not shown. The pulse is shown in its focal plane $z = 0$, at $ct = -b, 0, b$, time increasing upward from the lowest figure. The pulse is travelling out of the page toward the reader. It has negative angular momentum about the propagation direction

For the wavefunction in (19.58), a Lorentz boost with $\beta = (a-b)/(a+b)$ or $(1+\beta)/(1-\beta) = a/b$ transforms ψ to (Lekner 2003)

$$\psi(\mathbf{r}', t') = \frac{ab\psi_0}{\rho^2 + \left[\sqrt{ab} - i(z' + ct')\right]\left[\sqrt{ab} + i(z' - ct')\right]}, \tag{19.78}$$

in which the forward and backward propagations are balanced. Such a choice of β brings the $TE + iTM$ pulse to its zero momentum frame, as we have seen in (19.70)–(19.72). Moreover, the energy in the zero momentum frame, $U_0 = \frac{\pi}{4}\psi_0^2/\sqrt{ab}$, is equal to the square root of $U^2 - c^2P_z^2$, so in this respect the pulse momentum and energy behave as four-vector components. (U and P_z were given in (19.72).)

However, other pulses constructed from the same wavefunction require a different β to bring them to their zero momentum frame, as in the example specified by (19.73) and (19.74) for which $\beta = (a-3b)/(a+3b)$. For this β the wavefunction (19.58) is transformed to

$$\psi(\mathbf{r}', t') = \frac{ab\psi_0}{\rho^2 + \left[\sqrt{3ab} - i(z' + ct')\right]\left[\sqrt{ab/3} + i(z' - ct')\right]}. \tag{19.79}$$

The transformed momentum is zero, and the transformed energy is

$$U_0 = \frac{\pi}{4}\psi_0^2/\sqrt{3ab}. \tag{19.80}$$

This is not (unless $a = 3b$) equal to the square root of $U^2 - c^2P_z^2$, for which the values in (19.74) give

$$\sqrt{U^2 - c^2P_z^2} = \frac{\pi}{4}\psi_0^2\sqrt{\frac{3b}{a^3}}. \tag{19.81}$$

Thus the same solution of the wave equation can lead to pulses for which the energy and momenta may or may not behave like four-vectors. In general, the Lorentz transformation of electromagnetic wavepackets is more complicated than that of point particles, as may be expected.

References

Bateman H (1904) The solution of partial differential equations by means of definite integrals. Proc Lond Math Soc 1:451–458

Brekhovskikh LM (1980) Waves in layered media, 2nd edn. Academic Press, New York

Brillouin L (1960) Wave propagation and group velocity. Academic Press, New York

Budden KG (1961) Radio waves in the ionosphere, Cambridge University Press, Cambridge

Budden KG (1985) The propagation of radio waves, Cambridge University Press, Cambridge

Cooper F, Khare A, Sukhatme U (2001) Supersymmetry in quantum mechanics. World Scientific, Singapore, pp 46–47
Cox T, Lekner J (2008) Reflection and non-reflection of particle wavepackets. Eur J Phys 29: 671–679
Darwin CG (1928) Free motion in the wave mechanics. Proc Roy Soc Lond A117:258–293
Eckart C (1930) The penetration of a potential barrier by electrons. Phys Rev 35:1303–1309
Einstein A (1917) On the quantum theory of radiation. Physik. Zeits. **18**, 121–128. (Eng trans: Laser theory, Barnes FS (ed) New York, IEEE, 1972)
Epstein PS (1930) Reflection of waves in an inhomogeneous absorbing medium,". Proc Nat Acad Sci USA 16:627–637
Feng S, Winful HG, Hellwarth RW (1999) Spatiotemporal evolution of focused single-cycle electromagnetic pulses. Phys Rev E 59:4630–4649
Flügge S (1974) Practical quantum mechanics. Springer, Berlin, Problem 39
Ginzburg VL (1964) The propagation of electromagnetic waves in plasmas. Pergamon, Oxford
Jackson JD (1975) Classical electrodynamics, 2nd edn. Wiley, New York
Kaiser H, Werner SA, George EA (1983) Direct measurement of the longitudinal coherence length of a thermal neutron. Phys Rev Lett 50:560–562
Kay I, Moses HE (1956) Reflectionless transmission through dielectrics and scattering potentials. J Appl Phys 27:1503–1508
Kennard EH (1927) "Zur Quantenmechanik einfacher Bewegungstypen. Zeits Physik 44:326–352
Kiriushcheva N, Kuzmin S (1998) Scattering of a Gaussian wave packet by a reflectionless potential. Am J Phys 66:867–872
Klein AG, Opat GI, Hamilton WA (1983) Longitudinal coherence in neutron interferometry. Phys Rev Lett 50:563–565
Landau L, Lifshitz E (1951) The classical theory of fields. Addison-Wesley, Reading, Section 2:6
Landau LD, Lifshitz EM (1965) Quantum mechanics, 2nd edn. Pergamon, Oxford, Sections 19.23 and 19.25
Lekner J (1972) Critical binding of diatomic molecules. Mol Phys **23**, 619–625 (1972)
Lekner J (1990) Nonreflecting stratifications. Can J Phys 68:738–742
Lekner J (2003) Electromagnetic pulses which have a zero momentum frame. J Opt A: Pure Appl Opt 5:L15–L18
Lekner J (2004a) Angular momentum of electromagnetic pulses. J Opt A: Pure Appl Opt 6:S128–S133
Lekner J (2004b) Energy and momentum of electromagnetic pulses. J Opt A: Pure Appl Opt 6:146–147
Lekner J (2004c) Localized electromagnetic pulses with azimuthal dependence. J Opt A: Pure Appl Opt 6:711–716
Lekner J (2004d) Helical light pulses. J Opt A: Pure Appl Opt 6:L29–L32
Lekner J (2007) Reflectionless eigenstates of the sech^2 potential. Am J Phys **75**, 1151–1157
Lekner J (2008) Rotating wavepackets. Eur J Phys 29:1121–1125
Lekner (2009) Airy wavepacket solutions of the Schrödinger equation. Eur J Phys **30**, L43–L46
Merzbacher E (1998) Quantum mechanics, 3rd edn. Wiley, New York

Further Readings

The appendix is based on

Lekner J (2013) Universal properties of electromagnetic pulses. PIERS Proceedings, Taipei, pp 458–463

Chapter 20
Finite Beams

Chapters 1–18 have dealt with the reflection of monochromatic plane waves from planar interfaces. The previous chapter discussed the reflection of electromagnetic pulses and of quantum particle wavepackets. Here we shall consider transversely finite beams. The simplest beams to consider are those bounded in space but still monochromatic. These can be viewed as a superposition of plane waves of the same frequency but differing propagation directions. We shall find, accordingly, that the reflection of beams depends on the angular dependence of the reflection amplitude. Particularly important is the case of total reflection, where all the frequency and angle dependence is contained in the phase of the reflection amplitude, since its modulus is then unity. The variation of the *s* and *p* phases with angle of incidence is discussed in Appendix 1, and applied to calculation of the lateral beam shift in Sect. 20.2. Section 20.3 gives analytic results for the reflection of Gaussian beams. Appendix 2 summarizes the polarization properties of finite beams. We start by reviewing the properties of finite beams.

20.1 Universal Properties of Scalar and of Electromagnetic Beams

This section is concerned with *universal* properties of scalar and of electromagnetic beams, by which we mean properties that all physical beams must have (or cannot have). We begin with a summary of the existing exact solutions.

For monochromatic beams in free space, in which the time dependence of the complex fields is contained in the factor $e^{-i\omega t}$, the quantum and acoustical scalar amplitudes ψ satisfy the Helmholtz equation

$$\left(\nabla^2 + k^2\right)\psi = 0, \qquad k = \omega/c\,. \tag{20.1}$$

Also all the components of **E** and **B** of an electromagnetic wave satisfy (20.1). This follows from Maxwell's equations by expressing the magnetic and electric fields in terms of the vector and scalar potentials **A** and V,

J. Lekner, *Theory of Reflection*, Springer Series on Atomic, Optical, and Plasma Physics 87, DOI 10.1007/978-3-319-23627-8_20

$$\mathbf{B} = \nabla \times \mathbf{A}, \qquad \mathbf{E} = -\nabla \mathrm{V} - \partial_{ct}\mathbf{A}, \tag{20.2}$$

and choosing the Lorenz gauge

$$\nabla \cdot \mathbf{A} + \partial_{ct}\mathrm{V} = 0. \tag{20.3}$$

In free space, V and all components of $\mathbf{A}$ then satisfy (20.1) (see for example Jackson (1975), pp. 218ff), and so do their derivatives such as $\mathbf{E}$ and $\mathbf{B}$.

The textbook solutions of (20.1) are the plane wave $\exp(i\mathbf{k}\cdot\mathbf{r})$ and the spherical waves $\exp(\pm ikr)/r$. Physical beams are localized transversely to the direction of propagation, in contradistinction to the textbook solutions. Deschamps (1971) noted that a complex shift along the propagating direction (the z-axis, in most of this chapter) gives an exact solution of (20.1) localized transversely:

$$\psi = \frac{\mathrm{e}^{ikR}}{R}, \qquad R^2 = x^2 + y^2 + (z - ib)^2 = \rho^2 + (z - ib)^2. \tag{20.4}$$

This solution is singular on the circle $\{\rho = b,\, z = 0\}$ and so cannot represent a physical beam. One can regularize by subtracting the complex-shifted spherically converging wave $\exp(-ikR)/R$ (Sheppard and Saghafi 1998) to obtain

$$\psi_{00} = \frac{\sin kR}{kR} = j_0(kR), \tag{20.5}$$

and generalize to (Ulanowski and Ludlow 2000)

$$\psi_{\ell m} = j_\ell(kR)\, P_\ell^m\left(\frac{z - ib}{R}\right) \mathrm{e}^{im\phi}, \tag{20.6}$$

but problems remain in the divergence of some invariants (see Lekner (2001) and below), and in the backward-propagating components associated with the terms proportional to $\exp(-ikR)/R$.

The Helmholtz equation (20.1) is separable in cylindrical coordinates (ρ, ϕ, z): it reads

$$\left[\partial_\rho^2 + \frac{1}{\rho}\partial_\rho + \frac{1}{\rho^2}\partial_\phi^2 + \partial_z^2 + k^2\right]\psi = 0. \tag{20.7}$$

This is solved by $J_m(K\rho)\mathrm{e}^{im\phi}\mathrm{e}^{iqz}$ (with $K^2 + q^2 = k^2$), and thus also by the *generalized Bessel beams* (Lekner 2004b)

$$\psi_m(\mathbf{r}) = \mathrm{e}^{im\phi}\int_0^k dK f(K)\, J_m(K\rho)\mathrm{e}^{iqz}, \qquad K^2 + q^2 = k^2. \tag{20.8}$$

Note that K is restricted to the interval $[0, k]$, so $q = \sqrt{k^2 - K^2}$ is a real mapping onto $[0, k]$. These beams are therefore purely forward propagating, by construction. The amplitude function $f(K)$ can be complex; it is constrained by the necessary finiteness of integral invariants (see below). We show first how $\psi_0(\mathbf{r})$ is related to the time-harmonic version of Bateman's integral solution of the wave equation.

20.1.1 Bateman Integral Solution of the Wave Equation

Bateman (1904) considered integral representations of solutions to the wave equation $\left(\nabla^2 - \partial_{ct}^2\right)\psi = 0$. The simplest case is that where the solution is independent of the azimuthal angle ϕ, in which case his solution reduces to

$$\Psi(\rho, z, t) = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\theta\, F(z + i\rho\cos\theta, ct + \rho\sin\theta). \tag{20.9}$$

We can adapt this to find the general solution of the Helmholtz equation (20.1) which is independent of the azimuthal angle. For time-dependence $\mathrm{e}^{-i\omega t} = \mathrm{e}^{-ikct}$, the function F must take the form

$$F(z + i\rho\cos\theta, ct + \rho\sin\theta) = g(z + i\rho\cos\theta)\,\mathrm{e}^{-ik(ct + \rho\sin\theta)}, \tag{20.10}$$

and then the spatial part of Ψ in (20.9) becomes

$$\psi(\rho, z) = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\theta\, g(z + i\rho\cos\theta)\,\mathrm{e}^{-ik\rho\sin\theta}. \tag{20.11}$$

We can verify that this is a solution of the Helmholtz equation as follows. Let $G(\rho, z, \theta) = g(z + i\rho\cos\theta)\,\mathrm{e}^{-ik\rho\sin\theta}$. A short calculation shows that $(\nabla^2 + k^2)G = -\rho^{-2}\partial_\theta^2 G$ and so $2\pi(\nabla^2 + k^2)\psi = [\rho^{-2}\partial_\theta G]_0 - [\rho^{-2}\partial_\theta G]_{2\pi} = 0$. Thus the expression (20.11) is the most general form of the scalar wavefunction corresponding to axially symmetric monochromatic beams. Note that on the beam axis ($\rho = 0$) we get

$$\psi(0, z) = g(z). \tag{20.12}$$

Therefore the amplitude function g in (20.11) given by the axial value of the beam wavefunction.

There is a one-to-one correspondence between (20.11) and the $m = 0$ generalized Bessel beam solution (20.8). Since $K^2 + q^2 = k^2$ we can write $\psi_0(\mathbf{r})$ as an integral over q instead of over K:

$$\psi_0(\mathbf{r}) = \int_0^k \mathrm{d}q\, h(q) J_0\left(\rho\sqrt{k^2-q^2}\right) \mathrm{e}^{iqz}, \quad h(q) = \frac{q}{\sqrt{k^2-q^2}} f\left(\sqrt{k^2-q^2}\right). \tag{20.13}$$

The zero-order Bessel function containing the square root can be rewritten by using Bessel's integral (Watson 1966, Sect. 2.21), which transforms (20.13) into

$$\psi_0(\rho, z) = \frac{1}{2\pi} \int_0^{2\pi} \mathrm{d}\theta\, \mathrm{e}^{-ik\rho\sin\theta} \int_0^k \mathrm{d}q\, h(q) \mathrm{e}^{iq(z+i\rho\cos\theta)}. \tag{20.14}$$

Comparison of (20.11) and (20.14) shows that, for the $m = 0$ generalized Bessel beams, the amplitude function g is given by

$$g(z + i\rho\cos\theta) = \int_0^k \mathrm{d}q\, h(q) \mathrm{e}^{iq(z+i\rho\cos\theta)}. \tag{20.15}$$

The axial value of the beam wavefunction is thus equal to the finite Fourier transform of $h(q)$:

$$\psi(0, z) = g(z) = \int_0^k \mathrm{d}q\, h(q) \mathrm{e}^{iqz}. \tag{20.16}$$

20.1.2 Conservation Laws and Beam Invariants

The energy, momentum and angular momentum densities of an electromagnetic field in free space are, in Gaussian units, given by

$$u(\mathbf{r}, t) = \frac{1}{8\pi}\left(E^2 + B^2\right), \qquad \mathbf{p}(\mathbf{r}, t) = \frac{1}{4\pi c}\mathbf{E} \times \mathbf{B}, \qquad \mathbf{j}(\mathbf{r}, t) = \mathbf{r} \times \mathbf{p}(\mathbf{r}, t). \tag{20.17}$$

Here $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{B}(\mathbf{r}, t)$ are the real fields. For monochromatic fields it is convenient to work in terms of complex fields $\mathbf{E}(\mathbf{r})$ and $\mathbf{B}(\mathbf{r})$ with the real electric field being given by

$$\begin{aligned} \mathbf{E}(\mathbf{r}, t) = \mathrm{Re}\{\mathbf{E}(\mathbf{r})\mathrm{e}^{-i\omega t}\} = \; & \mathrm{Re}\{[\mathbf{E}_r(r) + i\mathbf{E}_i(\mathbf{r})][\cos\omega t - i\sin\omega t]\} \\ & = \mathbf{E}_r(\mathbf{r})\cos\omega t + \mathbf{E}_i(\mathbf{r})\sin\omega t. \end{aligned} \tag{20.18}$$

The average of $u(\mathbf{r},t)$ over one period $2\pi/\omega$ is

$$\bar{u}(\mathbf{r}) = \frac{1}{8\pi}\{\mathbf{E}(\mathbf{r})\cdot\mathbf{E}^*(\mathbf{r}) + \mathbf{B}(\mathbf{r})\cdot\mathbf{B}^*(\mathbf{r})\}. \tag{20.19}$$

Likewise the cycle-averaged momentum density is

$$\bar{\mathbf{p}}(\mathbf{r}) = \frac{1}{16\pi c}[\mathbf{E}(\mathbf{r})\times\mathbf{B}^*(\mathbf{r}) + \mathbf{E}^*(\mathbf{r})\times\mathbf{B}(\mathbf{r})]. \tag{20.20}$$

The conservation of energy equation, $\nabla\cdot\mathbf{S} + \partial_t u = 0$, where $\mathbf{S} = c^2\mathbf{p}$ is the energy flux density, has the cycle-average equal to c^2 times

$$\nabla\cdot\bar{\mathbf{p}} = \partial_x\bar{p}_x + \partial_y\bar{p}_y + \partial_z\bar{p}_z = 0. \tag{20.21}$$

Applying $\int \mathrm{d}^2 r = \int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}y = \int_0^{\infty}\mathrm{d}\rho\,\rho\int_0^{2\pi}\mathrm{d}\phi$ to (20.21) gives, for transversely finite beams propagating in the z direction (Lekner 2004a)

$$\partial_z\int \mathrm{d}^2 r\,\bar{p}_z = 0, \qquad \text{or} \qquad P_z' = \int \mathrm{d}^2 r\,\bar{p}_z = \text{constant}. \tag{20.22}$$

We use the notation P_z', since $\mathrm{d}P_z = P_z'\,\mathrm{d}z$ is the total z-component momentum contained in a transverse slice of the beam, of thickness $\mathrm{d}z$. Equation (20.22) states that the momentum content per unit length, along the direction of net propagation of the beam, is an *invariant*. Note that the invariance of the *momentum* content per unit length is derived from the conservation of *energy* (the energy flux density is proportional to the momentum density).

The conservation of momentum equation is expressed in terms of the stress (or momentum flux density) tensor

$$\partial_t p_i + \sum_j \partial_j\tau_{ij} = 0, \qquad \tau_{ij} = \frac{1}{4\pi}\left[\frac{1}{2}(E^2+B^2)\delta_{ij} - E_iE_j - B_iB_j\right]. \tag{20.23}$$

Taking the cycle average gives $\sum_j \partial_j\bar{\tau}_{ij} = 0$, and operating with $\int \mathrm{d}^2 r$ gives $\partial_z\int \mathrm{d}^2 r\,\bar{\tau}_{iz} = 0 \quad (i = x,y,z)$. Thus momentum conservation leads to three invariants (Lekner 2004a)

$$\begin{aligned}
T_{xz}' &= \int \mathrm{d}^2 r\,\bar{\tau}_{xz} = -\frac{1}{4\pi}\int \mathrm{d}^2 r\left[\overline{E_xE_z} + \overline{B_xB_z}\right],\\
T_{yz}' &= \int \mathrm{d}^2 r\,\bar{\tau}_{yz} = -\frac{1}{4\pi}\int \mathrm{d}^2 r\left[\overline{E_yE_z} + \overline{B_yB_z}\right],\\
T_{zz}' &= \int \mathrm{d}^2 r\,\bar{\tau}_{zz} = \frac{1}{8\pi}\int \mathrm{d}^2 r\left[\overline{E_x^2+E_y^2-E_z^2} + \overline{B_x^2+B_y^2-B_z^2}\right].
\end{aligned} \tag{20.24}$$

Three more invariants follow from the conservation of angular momentum, $\partial_t j_i + \sum_\ell \partial_\ell \mu_{\ell i} = 0$, where the angular momentum flux density tensor $\mu_{\ell i} = \sum_j \sum_k \varepsilon_{ijk} x_j \tau_{k\ell}$ is defined in terms of the momentum flux density tensor τ_{ij} (Barnett 2002). These invariants are

$$\begin{aligned} M'_{zx} &= \int \mathrm{d}^2 r\, \bar{\mu}_{zx} = \int \mathrm{d}^2 r \big[y\bar{\tau}_{zz} - z\bar{\tau}_{yz} \big], \\ M'_{zy} &= \int \mathrm{d}^2 r\, \bar{\mu}_{zy} = \int \mathrm{d}^2 r [z\bar{\tau}_{xz} - x\bar{\tau}_{zz}], \\ M'_{zz} &= \int \mathrm{d}^2 r\, \bar{\mu}_{zz} = \int \mathrm{d}^2 r \big[x\bar{\tau}_{yz} - y\bar{\tau}_{xz} \big]. \end{aligned} \tag{20.25}$$

Thus there are seven universal invariants of electromagnetic beams, arising from the conservation of energy, momentum and angular momentum. Perhaps surprisingly, the energy per unit length of the beam, $U' = \int \mathrm{d}^2 r\, \bar{u}$, is not always an invariant, although it is constant for the types of generalized Bessel beams discussed in Lekner (2004b), as is $J'_z = \int \mathrm{d}^2 r \bar{j}_z$.

The invariants for quantum particle beams and for sound beams also correspond to conservation laws. In both cases they originate from the conservation of particles (continuity equation) and conservation of momentum and of angular momentum (Lekner 2004, 2007).

20.1.3 Non-existence Theorems

In textbooks a light beam is usually represented by a plane wave, with **E**, **B** and the propagation vector **k** everywhere mutually perpendicular. This 'beam' can be everywhere linearly polarized in the same direction, or everywhere circularly polarized in the same plane, and its energy is everywhere transported in a fixed direction at the speed of light. It has been shown (Lekner 2003) that *none* of these properties can hold for a transversely finite beam. We shall just state the theorems, except for the one relating to linear polarization, for which the proof given in Lekner (2003) is incomplete.

(i) Pure TEM beams do not exist.
(ii) Beams of fixed linear polarization do not exist.
(iii) Beams which are everywhere circularly polarized in the same direction do not exist.
(iv) Beams or pulses within which the energy velocity (Lekner 2002) is everywhere in the same direction and of magnitude c do not exist.

Proof of (ii) Suppose $\mathbf{E} = (F(x, y, z), 0, 0)$, so the beam is linearly polarized along $\hat{\mathbf{x}}$, everywhere. Then from the Maxwell curl equations, with e^{-ikct} time dependence,

we have $ik\mathbf{B} = \nabla \times \mathbf{E} = (0, \partial_z, -\partial_y)F$, and $ik\nabla \times \mathbf{B} = \Big(-[\partial_y^2 + \partial_z^2], \partial_x\partial_y, \partial_x \partial_z)F = k^2\mathbf{E}$. Hence $(\partial_y^2 + \partial_z^2 + k^2)F = 0$ and $\partial_x\partial_y F = 0 = \partial_x\partial_z F$. The last two equations imply $F(x, y, z) = f(x) + g(z) + h(y, z)$, which cannot represent a beam localized transversely in the x direction.

Appendix 2 gives more detail about the polarization of finite beams.

20.1.4 Focal Plane Zeros

We have seen that electromagnetic beams can be constructed from solutions of the scalar Helmholtz equation (20.1). In particular the TM, TE, 'LP' and 'CP' beams have their vector potentials proportional (respectively) to

$$(0, 0, \psi), \qquad \big(\partial_y\psi, -\partial_x\psi, 0\big), \qquad (\psi, 0, 0) \quad \text{and} \quad (-i\psi, \psi, 0). \tag{20.26}$$

(The quotation marks indicate that the 'LP' and 'CP' beams are fully linearly and circularly polarized only in the plane wave limit: compare theorems (ii) and (iii) of the previous section.)

What are the universal properties of physically acceptable solutions? We have already seen that seven beam invariants must exist. We also saw that certain textbook properties of plane wave electromagnetic beams cannot hold for laterally finite beams. Here we argue that an infinity of zeros of ψ must occur in the focal plane.

The solutions $\psi(\mathbf{r})$ of the Helmholtz equation are, in general, complex functions of position, $\psi = \psi_r + i\psi_i$. The real and imaginary parts ψ_r and ψ_i are (in free space) smooth functions of position. These functions are zero on surfaces S_r and S_i, and where these surfaces meet (on curves C in space) both ψ_r and ψ_i are zero. If we write

$$\psi(\mathbf{r}) = M(\mathbf{r})\mathrm{e}^{iP(\mathbf{r})} = \big[\psi_r^2 + \psi_i^2\big]^{\frac{1}{2}}\exp\left(i \arctan\frac{\psi_i}{\psi_r}\right) \tag{20.27}$$

we see that, on any such curve C, the modulus $M(\mathbf{r})$ is zero, and the phase $P(\mathbf{r})$ is indeterminate. Nye and Berry (1974) called these curves wave dislocations; Chap. 5 of Nye's (1999) book gives illustrations of such phase singularities.

Lekner (2013) has given a topological argument for the existence of zeros of ψ in the focal plane, on the assumption that the isophase surfaces intersect the focal plane. At the zeros of ψ the phase can be any real number excluding integer multiples of π, as explained below.

The focal plane is a plane of symmetry for an ideal beam, and can be taken as the $z = 0$ plane; we can also take the phase of ψ to be zero at the origin. Then the isophase surfaces correspond to negative $P(\mathbf{r})$ for $z < 0$ and positive $P(\mathbf{r})$ for $z > 0$. The surfaces $P = -n\pi$ and $P = n\pi$ can meet where ψ is not zero, since the phase difference is an integer (n) multiple of 2π. These isophase surfaces are concave

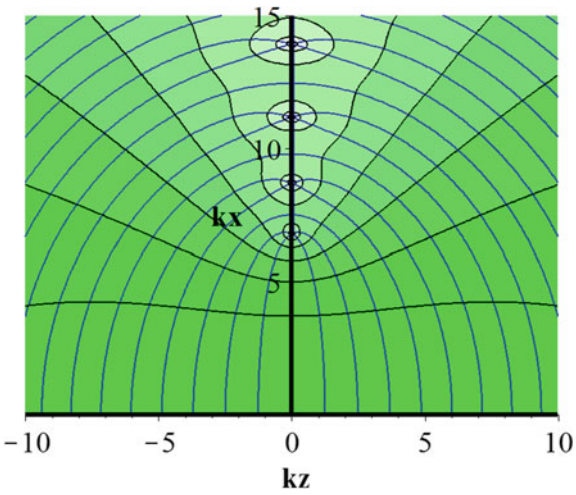


Fig. 20.1 Isophase surfaces and surfaces of constant modulus for the $j_0(kR)$ beam, (20.5), plotted for $kb = 6$. The three-dimensional picture is obtained by rotating about the z-axis. The phase is shown at intervals of $\pi/3$; it has been chosen to be zero at the origin. The surfaces with phase equal to an integer multiple of π converge onto the circles $\rho = \left[(X/k)^2 + b^2\right]^{\frac{1}{2}}$, where $\tan X = X$. The other isophase surfaces converge onto the zeros of $j_0(kR)$ in the $z = 0$ plane, namely on the circles $\rho_n = \left[(n\pi/k)^2 + b^2\right]^{\frac{1}{2}}$. The beam axis and focal plane are indicated by the heavy *horizontal* and *vertical lines*

toward the origin, since a physical beam is converging toward the focal region for $z<0$ and diverging from it for $z > 0$. All other isophase surfaces can only meet on the focal plane if on it there exist curves where ψ is zero. On such curves (circles, in the simplest case) the phase surfaces $P = -\pi/2$ and $P = +\pi/2$ can meet, for example. The surfaces with $0<|P|<\pi$ meet on the first zero curve, $\pi<|P|<2\pi$ meet on the next, and so on. Figure 20.1 illustrates the phenomenon, conjectured to be universal at the focal plane. Because of the topological nature of the above argument, we expect the zeros to persist even when the beam is perturbed (for example, focused by an imperfect lens or mirror). The focal plane would then be distorted to a nearly-planar surface, and the circles of zeros to approximately circular closed curves, where the perturbed phase surfaces $\pm P$ meet.

One counter-example to the above conjecture (of the universality of rings of zeros in the focal plane) appears to be separable spheroidal beams, for which $\psi(\xi, \eta, \phi) = R(\xi)S(\eta)\mathrm{e}^{im\phi}$ with $\rho = b\left[\left(\xi^2+1\right)\left(1-\eta^2\right)\right]^{\frac{1}{2}}, z = b\xi\eta$. The focal plane $z = 0$ corresponds to $\xi = 0$ for $\rho \le b$ and $\eta = 0$ for $\rho \ge b$. Thus if $S(\eta)$ is zero for $\eta = 0$, $\psi = 0$ for $\rho \ge b$ in the focal plane, and the $-P$ and $+P$ isophase surfaces can meet anywhere on the focal plane outside of the central disk $\rho \le b$. However, such spheroidal wavefunctions have been shown to be non-physical (Boyack and Lekner 2011).

20.2 Reflection of Beams: The Lateral Beam Shift

We would expect a lateral beam shift in the case of total reflection from an interface with gradually decreasing refractive index: the semiclassical or geometric optics picture is shown in Fig. 20.2. If the angle between the ray and the z axis is $\theta(z)$, geometric optics gives the lateral shift as

$$\Delta x = 2\int_{x_1}^{x_0} \mathrm{d}x = \int_0^{z_0} \mathrm{d}z \tan\theta(z). \tag{20.28}$$

For a *sharp* transition between media of dielectric constants ε_1 and ε_2, with $\varepsilon_1 > \varepsilon_2$ and $\theta_1 > \theta_c = \arcsin(\varepsilon_2/\varepsilon_1)^{\frac{1}{2}}$, the turning point z_0 and the beginning of the transition coincide, and (20.28) gives zero lateral shift. Goos and Hänchen (1947) however found a *non-zero beam shift* in this case, with a maximum lateral displacement just beyond the critical angle. This phenomenon is referred to as the Goos-Hänchen effect. It is universal for wave phenomena: a comprehensive review, with references to work in optics, acoustics, quantum mechanics and plasma physics has been given by Lotsch (1970); illustrations of acoustic beam displacement may be found in Brekhovskikh (1980). Figure 20.3 illustrates the lateral displacement of a beam at a sharp boundary.

We will show that the beam shift in most cases (excluding the immediate neighbourhood of the critical angle) is well approximated by the formula

$$\Delta x = -\mathrm{d}\delta/\mathrm{d}K\,. \tag{20.29}$$

Here δ is the phase of the reflection amplitude, K is the lateral component of the wavevector ($K = \varepsilon^{\frac{1}{2}}(\omega/c)\sin\theta$) and the derivative is to be evaluated at the dominant K value of the incident beam. The two extremes illustrated in Figs. 20.2 and 20.3 are encompassed by (20.29), except *at* the critical angle where, as is shown in Appendix 1, δ has a square root singularity.

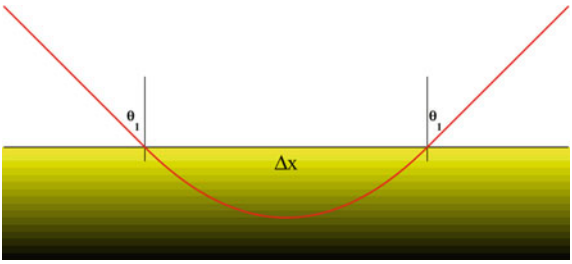


Fig. 20.2 Lateral shift of a ray totally reflected from an inhomogeneous region. The inhomogeneity begins at $z = 0$, and the turning point is at z_0. The *upper* medium is optically denser: *dark* indicates small refractive index. The entry into the inhomogeneous optically less dense medium is at $(x_1, 0)$, the exit point is $(x_1 + \Delta x, 0)$, and the turning point is (x_0, z_0)

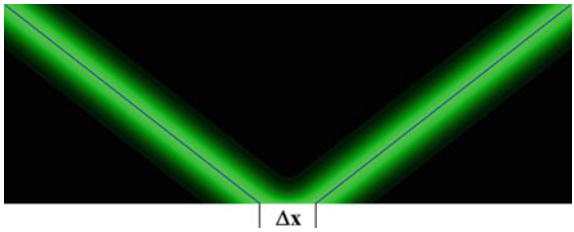


Fig. 20.3 Goos-Hänchen effect: the lateral beam shift in total reflection at a discontinuous transition

A transversely finite beam may be built up by superposing plane waves with a spread of propagation directions, centred on the direction of the beam. We consider only a spread in the angle of incidence: the beam is taken to extend a large distance in the y direction. It is convenient to characterize the plane wave components by their lateral wavenumbers K, following Brekhovskikh (1980, Sect. 14.1). The plane waves $\exp i(Kx \pm qz - \omega t)$, with $K^2 + q^2 = \varepsilon_1 \omega^2/c^2$, are solutions of the wave equation in medium 1, and the incident beam is made up of a superposition of these:

$$E_i(z,x) = \int_{-\infty}^{\infty} \mathrm{d}K f(K) \mathrm{e}^{i(Kx+qz)}. \tag{20.30}$$

When $K^2 > \varepsilon_1 \omega^2/c^2$, q is imaginary, implying evanescent waves. This possibility is excluded here since we will be considering well-collimated beams, with $f(K)$ non-zero only in a narrow range of K about $K_1 = \varepsilon_1^{\frac{1}{2}}(\omega/c)\sin\theta_1$.

The reflected beam is obtained by summing over the reflected component plane waves, each with its own reflection amplitude $r(K) = |r(K)| \exp i\delta(K)$:

$$E_r(z,x) = \int_{-\infty}^{\infty} \mathrm{d}K f(K) r(K) \mathrm{e}^{i(Kx-qz)}. \tag{20.31}$$

If E_i at some reference plane $z = 0$ is given by $E_0(x)$, then

$$E_0(x) = \int_{-\infty}^{\infty} \mathrm{d}K f(K) \mathrm{e}^{iKx}, \qquad f(K) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}x\, E_0(x) \mathrm{e}^{-iKx}. \tag{20.32}$$

The reflected field at $z = 0$ is thus given by

$$E_r(0,x) = \int_{-\infty}^{\infty} \mathrm{d}K\, r(K) \mathrm{e}^{iKx} \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}x'\, E_0(x') \mathrm{e}^{-iKx'} \tag{20.33}$$

We now assume that $|r(K)|$ varies slowly with K compared to the phase $\delta(K)$ (this assumption is exact for $K > K_c = \sqrt{\varepsilon_2}(\omega/c)$, when $|r(K)| = 1$), and also that the variation of $\delta(K)$ is well approximated by the linear term in its Taylor expansion about K_1:

$$|r(K)| \approx |r(K_1)|, \qquad \delta(K) \approx \delta(K_1) + (K - K_1)\delta'(K_1). \tag{20.34}$$

Here $\delta'(K_1)$ stands for $\mathrm{d}\delta/\mathrm{d}K$ evaluated at K_1. The integral over K in (20.33) becomes, in this approximation, 2π times a delta function in x', which selects the value $x' = x + \delta'(K_1)$. Thus the reflected wave at $z = 0$ is

$$E_r(0, x) \approx |r(K_1)|\mathrm{e}^{i(\delta - K_1\delta')}E_0(x + \delta'). \tag{20.35}$$

Thus the beam is reduced in amplitude by $|r(K_1)|$, phase shifted by $\delta - K_1\delta'$, and moved in the x direction by the distance $-\delta'$. The last statement is equivalent to (20.29). Note that under the approximations (20.34) the beam shape does not change: the beam is simply translated. This is analogous to the reflection of pulses considered in Sect. 19.1, where we saw that the corresponding frequency expansion of the reflection amplitude leads to a time delay, with no change in shape of the reflected pulse.

We will now apply (20.29) to the cases illustrated in Figs. 20.2 and 20.3. In the geometrical optics limit the phase is well approximated (except near grazing incidence) by the short wave formula (19.14). On substituting in (20.29) and using $q(z_0) = 0$, and $\mathrm{d}q/\mathrm{d}K = -K/q = -\tan\theta$, we regain (20.28).

In the sharp transition case, we will consider the region of total reflection. The s and p phases are discussed in Appendix 1. They are given by (20.63) and (20.67), which we rewrite in terms of K:

$$\delta_s = -2\arctan\left(\frac{K^2 - k_2^2}{k_1^2 - K^2}\right)^{\frac{1}{2}}, \quad \delta_p = -\pi - 2\arctan\frac{\varepsilon_1}{\varepsilon_2}\left(\frac{K^2 - k_2^2}{k_1^2 - K^2}\right)^{\frac{1}{2}}. \tag{20.36}$$

Here $k_1 = n_1\omega/c$ and $k_2 = n_2\omega/c$ are the magnitudes of the total wavevectors in media 1 and 2. The formula (20.29) gives the beam shifts

$$\Delta x_s = \frac{2K}{q_1|q_2|} = \frac{\lambda_1}{\pi}\frac{\tan\theta_1}{\left(\sin^2\theta_1 - \sin^2\theta_c\right)^{\frac{1}{2}}}, \tag{20.37}$$

$$\Delta x_p = \Delta x_s \frac{\sin^2\theta_c}{\left(1 + \sin^2\theta_c\right)\sin^2\theta_1 - \sin^2\theta_c}, \tag{20.38}$$

λ_1 being the wavelength in the first medium. Near θ_c the p wave beam shift is larger by $1/\sin^2\theta_c = \varepsilon_1/\varepsilon_2$; near grazing incidence it is smaller by the factor $\varepsilon_2/\varepsilon_1$. The formulae are not applicable *at* the critical angle or *at* grazing incidence, since they were derived by using the truncated Taylor expansion (20.34), which fails at a

square root singularity. The fact that δ_s and δ_p always have a term linear in $|q_2| = \left(K^2 - k_2^2\right)^{\frac{1}{2}}$ near θ_c is established in Appendix 1. The behaviour near grazing incidence, with a term linear in $q_1 = \left(k_1^2 - K^2\right)^{\frac{1}{2}}$ is also universal, as may be seen (for example) from (20.89) on letting q_1 tend to zero. Equations (20.36) and Fig. 20.4 show these singularities explicitly in the sharp boundary case.

The simple theory which gives the beam shift as $\Delta x = -\mathrm{d}\delta/\mathrm{d}K$ thus fails at the critical angle and at grazing incidence, the predicted beam shift diverging as $\left(K^2 - k_2^2\right)^{-\frac{1}{2}}$ and $\left(k_1^2 - K^2\right)^{-\frac{1}{2}}$ respectively. (Note however that the shift transverse to the reflected beam direction is $\Delta x \cos\theta_1$, which stays finite as $\theta_1 \to \pi/2$.)

Horowitz and Tamir (1971) have studied the reflection of a Gaussian beam by a sharp interface, without making an approximation equivalent to (20.34). They find that the results given in (20.37) and (20.38) are accurate down to $\theta_1 - \theta_c \approx 60$ millidegrees when the beam width parameter w is one thousand wavelengths, and to about 6 millidegrees when w is ten thousand wavelengths. The definition of w for a Gaussian beam is via the electric field amplitude at the beam waist (see Sect. 20.3):

$$E(\rho) \sim \exp\left[-(\rho/w)^2\right]. \tag{20.39}$$

Here ρ is the distance measured from the beam axis, transversely to the beam propagation direction. Their analysis gives a beam displacement independent of angle in the immediate neighbourhood of θ_c, with magnitude proportional to $(w\lambda)^{\frac{1}{2}}$:

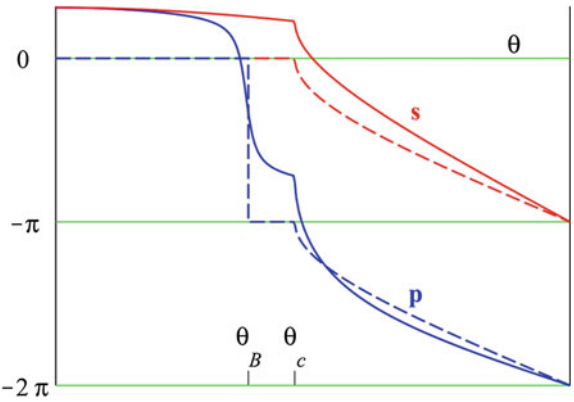


Fig. 20.4 Angular dependence of the phases of the s and p wave reflection amplitudes, for a sharp boundary between glass and air, refractive indices $\frac{3}{2}$ and 1 (*dashed curves*), and for a homogeneous layer of water (refractive index 4/3) between the glass and air, with $(\omega/c)\Delta z = 1/2$ (*solid curves*). The Brewster and critical angles for the glass-air boundary are indicated; $\theta_B \approx 33.7°, \theta_c \approx 41.8°$. Normal incidence is at *left*, glancing incidence at *right*

$$\Delta x_s(\theta_c) = \frac{\Gamma\left(\frac{1}{4}\right)}{2^{\frac{3}{2}}} \frac{(\tan\theta_c)^{\frac{1}{2}}}{\cos\theta_c} \frac{(w\lambda)^{\frac{1}{2}}}{\pi}, \tag{20.40}$$

with the p polarization displacement larger by $\varepsilon_1/\varepsilon_2$. Note that setting $\theta_1 = \theta_c + \lambda_1/w$ in (20.37) and (20.38) (that is, letting the angle of incidence approach the critical angle to within the diffraction-limited broadening of the beam) gives the qualitative features of (20.40) for small λ_1/w:

$$\Delta x_s(\theta_c + \lambda_1/w) \approx \frac{1}{2^{\frac{1}{2}}} \frac{(\tan\theta_c)^{\frac{1}{2}}}{\cos\theta_c} \frac{(w\lambda)^{\frac{1}{2}}}{\pi}. \tag{20.41}$$

Experimental test of the Tamir-Horowitz prediction is difficult, since to approach the critical angle closely one must have a highly collimated beam to obtain the required angular resolution. Laser beams have the required collimation, but the wavelength is then small, and so is the beam shift. Early data of Wolter (1950) (also displayed on page 200 of the Lotsch review) are in good agreement with the simple theory.

The lateral shift on reflection is of importance in waveguides, especially in fibre optics. See for example White and Pask (1977) and Snyder and Love (1983, Chap. 10).

20.3 Reflection of Gaussian Beams

In Sect. 20.1 we saw that electromagnetic beams can be constructed from solutions of the scalar Helmholtz equation (20.1). Equation (20.26) gave the vector potentials corresponding to TM, TE, 'LP' and 'CP' beams. In this section we consider Gaussian beams, which are solutions of the *paraxial equation*, in which one sets $\psi = \mathrm{e}^{ikz}G$, and then neglects the term $\partial_z^2 G$ in the resulting equation for G (to be given below). This amounts to assuming that the dominant z-dependence of the beam lies in the e^{ikz} factor. For axially symmetric solutions we omit the azimuthal derivative, so the Helmholtz equation takes the form (20.7), and the equation for G becomes

$$(\partial_\rho^2 + \rho^{-1}\partial_\rho + 2ik\partial_z + \partial_z^2)G = 0. \tag{20.42}$$

The fundamental solution of (20.42) with the ∂_z^2 term omitted (that is, of the paraxial equation), gives (Zangwill 2013, Sect. 16.7)

$$\psi_G = \mathrm{e}^{ikz}G = \frac{b}{b+iz}\exp\left\{ikz - \frac{k\rho^2}{2(b+iz)}\right\}. \tag{20.43}$$

Alternatively, we can write the Gaussian beam fundamental mode in the modulus times phase factor form:

$$\psi_G = \frac{b}{\sqrt{b^2+z^2}} \exp\left[\frac{-kb\rho^2}{2(b^2+z^2)}\right] \exp i\left\{kz - \arctan\left(\frac{z}{b}\right) + \frac{kz\rho^2}{2(b^2+z^2)}\right\}. \quad (20.44)$$

The arctangent term in the phase causes the phase of the beam to decrease by π relative to the plane wave phase kz as the beam passes through its focal region. This phase lag associated with focusing is universal for waves. It was first noted by Gouy in 1890.

The length b is the Rayleigh or diffraction length: it gives the longitudinal extent of the beam focal region directly, as one can see from the modulus exponential in (20.44). The beam waist size, located at the focal plane (here $z = 0$) is obtained from the modulus in the focal plane, $\exp(-k\rho^2/2b)$. It is usually written as $w_0 = \sqrt{2b/k} = \sqrt{b\lambda/\pi}$. If we define the beam width $w(z)$ by setting the exponential factor in the modulus equal to $\exp[-\rho^2/w^2(z)]$, we get

$$w^2(z) = \frac{2(b^2+z^2)}{kb}\,. \quad (20.45)$$

Thus $w = w_0$ at the beam waist, $w^2 = 2w_0^2$ at $z = \pm b$, and $w^2 \to 2z^2/kb$ when $|z| \gg b$. Away from the focal region the beam spreads as a cone of half-angle $\arctan\sqrt{2/kb}$. For $kb = 2$ and 6 this angle is 45° and 30°, respectively.

We need to consider the validity of the paraxial approximation which leads to the Gaussian beam solution. As noted in Lekner (2001), the quantity $\psi_G^{-1}\nabla^2\psi_G$ should equal $-k^2$, but instead equals $-k^2$ times

$$1 + \frac{2}{k^2(b+iz)^2} - \frac{2\rho^2}{k(b+iz)^3} + \frac{\rho^4}{4(b+iz)^4}\,. \quad (20.46)$$

The errors are thus negligible in the regions where

$$k^2\left(b^2+z^2\right) \gg 1 \quad \text{and} \quad b^2+z^2 \gg \rho^2\,. \quad (20.47)$$

We conclude that if kb is of order unity, the paraxial approximation fails in the focal region $|z| \le b$. Thus tightly focused beams are not well described by wavefunctions based on ψ_G. However, we can hope to describe beam reflection by means of ψ_G for larger kb, and this is what we shall now do.

We shall consider two examples of normal incidence reflection and transmission of a scalar Gaussian beam at a sharp interface between two media. We are looking for *effective reflection and transmission amplitudes for the central part of the beam*. We cannot hope to find these for the whole beam, since the outer parts of the beam will reflect differently from the central part. In any case the second inequality in (20.47) is violated in the focal region when ρ exceeds b, and the paraxial expression (20.43) becomes inaccurate.

20.3.1 Reflection at a Potential Spike (Delta Function)

The simplest example is that of the delta function potential for which exact wavepacket solutions were discussed in Sect. 19.3. The Helmholtz equation for the beam reads

$$[\partial_\rho^2 + \rho^{-1}\partial_\rho + \partial_z^2 + k^2 - 2\kappa\delta(z)]\psi = 0. \tag{20.48}$$

Approximate (paraxial) solutions of (20.48), everywhere except at $z = 0$, are the Gaussian beams

$$\begin{aligned} \psi^+ &= \frac{b}{b + i(z-a)} \exp\left\{ ik(z-a) - \frac{k\rho^2}{2[b + i(z-a)]} \right\}, \\ \psi^- &= \frac{b}{b - i(z+a)} \exp\left\{ -ik(z+a) - \frac{k\rho^2}{2[b - i(z+a)]} \right\}. \end{aligned} \tag{20.49}$$

The first corresponds to a forward-propagating beam with focal plane at $z = a$, the second to a backward-propagating beam with focal plane at $z = -a$ (the image position of the $z = a$ plane with respect to the interface at $z = 0$). We have chosen these forms to make $\psi^+(\rho, 0) = \psi^-(\rho, 0)$ for all ρ. We shall try to find effective reflection and transmission amplitudes r, t such that

$$\psi(\rho, z) = \psi^+ + r\psi^-(z<0), \quad \psi(\rho, z) = t\psi^+(z > 0). \tag{20.50}$$

From the differential equation (20.48), both ψ and $\partial_\rho\psi$ must be continuous at $z = 0$. Because of the delta function, $\partial_z\psi$ is discontinuous at $z = 0$ (compare (19.41)):

$$\partial_z\psi(\rho, 0+) - \partial_z\psi(\rho, 0-) = 2\kappa\psi(\rho, 0)\,. \tag{20.51}$$

Since $\psi^+(\rho, 0) = \psi^-(\rho, 0)$ and also $\partial_\rho\psi^+(\rho, 0) = \partial_\rho\psi^-(\rho, 0)$ for all ρ, the continuity of both ψ and $\partial_\rho\psi$ at $z = 0$ is satisfied by $1 + r = t$. The condition (20.51) can however be satisfied exactly only on the beam axis $\rho = 0$, by

$$r = \frac{-i\kappa}{k + i\kappa - i(a + ib)^{-1}}, \quad t = \frac{k - i(a + ib)^{-1}}{k + i\kappa - i(a + ib)^{-1}}. \tag{20.52}$$

These reflection and transmission amplitudes are accordingly for the central part of the beam; they lose validity for ρ comparable to or greater than $|a + ib|$. When either ka or kb are large (the focal planes far away from the reflecting plane, or wide beams, both on the scale of the wavelength), the amplitudes tend to the plane wave values of (19.43).

20.3.2 Reflection at a Sharp Boundary Between Two Media

We shall take the boundary between the two media in which the wave numbers are k_1, k_2 to be the $z = 0$ plane, and the focal plane of the incident beam to be $z = a_1$. The incident beam is, from (20.49),

$$\psi_i = \frac{b_1}{b_1 + i(z - a_1)} \exp\left\{ ik_1(z - a_1) - \frac{k_1 \rho^2}{2[b_1 + i(z - a_1)]} \right\} \tag{20.53}$$

The reflected beam is obtained by setting $z \to -z$, and taking its focal plane at the image in the reflecting surface of the incident beam focal plane, namely at $z = -a_1$:

$$\psi_r = \frac{rb_1}{b_1 - i(z + a_1)} \exp\left\{ -ik_1(z + a_1) - \frac{k_1 \rho^2}{2[b_1 - i(z + a_1)]} \right\}. \tag{20.54}$$

If a_1 is negative (focal plane of the incident beam to the left of the reflecting surface), the focal plane of the reflected beam will be virtual, to the right of the reflecting surface. The transmitted beam is

$$\psi_t = \frac{tb_2}{b_2 + i(z - a_2)} \exp\left\{ ik_2(z - a_2) - \frac{k_2 \rho^2}{2[b_2 + i(z - a_2)]} \right\}. \tag{20.55}$$

The reflection and transmission amplitudes (for the central part of the beam) are r and t. The boundary conditions are the continuity of $\psi, \partial_z \psi, \partial_\rho \psi$ at $z = 0$, where $\psi = \psi_i + \psi_r (z < 0), \psi = \psi_t (z > 0)$. (If any of these were discontinuous, there would be resultant delta functions in their derivatives at $z = 0$.) For the terms in the exponents proportional in ρ^2 to agree at $z = 0$, we need to set

$$b_2 = (k_2/k_1) b_1, \qquad a_2 = (k_2/k_1) a_1. \tag{20.56}$$

The first equation in (20.56) implies that the beam waist (real or virtual), has the same width $w_0 = \sqrt{2b/k}$ in the incident, reflected and transmitted beams, since $b_2/k_2 = b_1/k_1$. The location of the focal plane of the transmitted beam is scaled by the same factor k_2/k_1. These relations also make, at $z = 0$, the beam prefactors all equal to $b_1/(b_1 - ia_1)$, and the coefficients of $-\rho^2$ in the exponent all equal to $k_1/2(b_1 - ia_1)$. When the equations (20.56) are applied to the transmitted beam (20.55), we find that the continuity of ψ and of $\partial_\rho \psi$ is exactly satisfied at $z = 0$ when

$$1 + r = \mathrm{e}^{i(k_1 a_1 - k_2 a_2)} t = \mathrm{e}^{ik_1 a_1 (1 - k_2^2/k_1^2)} t. \tag{20.57}$$

The other boundary condition (continuity of $\partial_z \psi$ at $z = 0$) is satisfied to order $(kb)^{-1}$ when the reflection and transmission amplitudes take the values

$$r = \left(\frac{k_1 - k_2}{k_1 + k_2}\right) \frac{k_2[b_1 - ia_1] + 1}{k_2[b_1 - ia_1] - 1} = \frac{k_1 - k_2}{k_1 + k_2} \left\{1 + \frac{2}{k_2 b_1} + O(b_1^{-2})\right\}, \tag{20.58}$$

$$t = \mathrm{e}^{-ik_1 a_1 (1 - k_2^2/k_1^2)}(1 + r) = \mathrm{e}^{-ik_1 a_1 (1 - k_2^2/k_1^2)} \frac{2k_1}{k_1 + k_2} \left\{1 + \frac{k_1 - k_2}{k_1 k_2 b_1} + O(b_1^{-2})\right\}. \tag{20.59}$$

The leading expressions in (20.58) and (20.59) make the beam wavefunction and its first derivatives exactly continuous on the axis of the beam, that is on $\rho = 0$. Both amplitude expressions contain the location a_1 of the focal plane of the incident beam. However, we see that the effective reflection amplitude is independent of the location of the focal plane of the incident beam, provided it is not too distant from the reflecting surface (a term proportional to a_1/b_1^2 is in the $O\left(b_1^{-2}\right)$ term in (20.58)). The location of the focal plane does determine the phase of the transmission amplitude, as is made explicit in (20.59). For a plane wave reflecting from an abrupt interface at $z = 0$, the phase would be zero, as we saw in (1.13) or (1.15), which give the normal incidence values

$$r = \frac{k_1 - k_2}{k_1 + k_2}, \qquad t = \frac{2k_1}{k_1 + k_2}. \tag{20.60}$$

The effective reflection and transmission amplitudes for the central part of the beam, given by (20.58) and (20.59), do not exactly satisfy the conservation law $k_1\left(1 - |r|^2\right) = k_2|t|^2$ of Sect. 2.1. There is a difference between the two sides of order b_1^{-1}, arising from the transfer of flux transversely within the beam.

Appendix 1: Total Internal Reflection: The r_s, r_p Phases and Their Difference

In Sect. 20.2 we considered the reflection of bounded beams, with emphasis on the problem of beam shift. The latter depends on the variation of the phase of the reflection amplitude with the angle of incidence, and is greatest near the critical angle where the derivative becomes infinite. We shall give examples of the angular dependence of the phases of r_s and r_p, and then show that a square root singularity at the critical angle is universal for non-absorbing profiles.

Reflection at a sharp boundary. The s and p reflection amplitudes for a step profile located at $z = 0$ are given by (1.13) and (1.31):

$$r_s = \frac{q_1 - q_2}{q_1 + q_2}, \quad r_p = -\frac{Q_1 - Q_2}{Q_1 + Q_2}. \tag{20.61}$$

When medium 1 is optically denser $(\varepsilon_1 > \varepsilon_2), q_1 > q_2$ and r_s has zero phase (all phases are modulo 2π) up to the critical angle θ_c, where $q^2 = (\omega/c)^2(\varepsilon_2 - \varepsilon_1 \sin^2\theta_1)$ passes through zero and q_2 changes from real to imaginary:

$$q_2 = i|q_2| = \frac{\omega}{c}\left(\varepsilon_1 \sin^2\theta_1 - \varepsilon_2\right)^{\frac{1}{2}}, \quad [\theta_1 \geq \arcsin(\varepsilon_2/\varepsilon_1)^{\frac{1}{2}} = \theta_c] \tag{20.62}$$

Beyond this angle of incidence $|r_s| = 1$ and

$$\delta_s = -2\arctan(|q_2|/q_1), \tag{20.63}$$

where

$$\frac{|q_2|}{q_1} = \left\{\cos^2\theta_c \tan^2\theta_1 - \sin^2\theta_c\right\}^{\frac{1}{2}} = \left\{\left(1 - \frac{\varepsilon_2}{\varepsilon_1}\right)\tan^2\theta_1 - \frac{\varepsilon_2}{\varepsilon_1}\right\}^{\frac{1}{2}}. \tag{20.64}$$

We note the square root singularity at θ_c, which leads to an infinite value of $\mathrm{d}\delta_s/\mathrm{d}\theta_1$ at θ_c^+: in terms of $\Delta\theta = \theta_1 - \theta_c$ this is

$$\delta_s = -2\left(\frac{4\varepsilon_2}{\varepsilon_1 - \varepsilon_2}\right)^{\frac{1}{4}}(\Delta\theta)^{\frac{1}{2}} + O(\Delta\theta). \tag{20.65}$$

The s wave phase decreases monotonically from 0 at θ_c to $-\pi$ at grazing incidence, approaching $-\pi$ linearly in the glancing angle $\gamma = \frac{\pi}{2} - \theta_1$:

$$\delta_s = -\pi + 2\left(\frac{\varepsilon_1}{\varepsilon_1 - \varepsilon_2}\right)^{\frac{1}{2}}\gamma + O\left(\gamma^2\right). \tag{20.66}$$

The p wave phase is zero from normal incidence to the Brewster angle $\theta_B = \arctan(\varepsilon_2/\varepsilon_1)^{\frac{1}{2}}$ where $Q_1 = Q_2$ and r_p changes sign. In the interval $\theta_B < \theta_1 < \theta_c$ we can set δ_p equal to $+\pi$ or $-\pi$. We take $\delta_p = -\pi$, this choice being dictated by continuity of the phase as a function of interfacial thickness, as the next example will make clear. Beyond θ_c the p wave phase is (from (20.31) with $Q_2 = i|Q_2|$)

$$\delta_p = -\pi - 2\arctan\frac{|Q_2|}{Q_1} = -\pi - 2\arctan\frac{\varepsilon_1|q_2|}{\varepsilon_2 q_1}. \tag{20.67}$$

The strength of the square root singularity is thus larger for the p phase shift by the factor $\varepsilon_1/\varepsilon_2$:

$$\delta_p = -\pi - 2\frac{\varepsilon_1}{\varepsilon_2}\left(\frac{4\varepsilon_2}{\varepsilon_1 - \varepsilon_2}\right)^{\frac{1}{4}}(\Delta\theta)^{\frac{1}{2}} + O(\Delta\theta). \tag{20.68}$$

The inverse factor applies as δ_p tends to -2π at grazing incidence:

$$\delta_p = -2\pi + 2\frac{\varepsilon_2}{\varepsilon_1}\left(\frac{\varepsilon_1}{\varepsilon_1 - \varepsilon_2}\right)^{\frac{1}{2}}\gamma + O(\gamma^2). \tag{20.69}$$

Figure 20.4 shows δ_s and δ_p for the sharp boundary between two media, and also for a homogeneous layer between the same two media (the latter to be discussed shortly).

The *ellipsometric ratio* r_p/r_s is equal to $\exp i(\delta_p - \delta_s)$ for $\theta_1 > \theta_c$. The phase difference $\Delta = \delta_p - \delta_s$ is given by

$$\Delta = -\pi + 2\left(\arctan\frac{|q_2|}{q_1} - \arctan\frac{\varepsilon_1}{\varepsilon_2}\frac{|q_2|}{q_1}\right). \tag{20.70}$$

The phase difference has an extremum at the angle of incidence

$$\theta_m = \arctan\left(\frac{2\varepsilon_2}{\varepsilon_1 - \varepsilon_2}\right)^{\frac{1}{2}}. \tag{20.71}$$

For comparison we list the tangents and sines of θ_B, θ_c and θ_m:

$$\tan^2\theta_B = \frac{\varepsilon_2}{\varepsilon_1}, \quad \tan^2\theta_c = \frac{\varepsilon_2}{\varepsilon_1 - \varepsilon_2}, \quad \tan^2\theta_m = \frac{2\varepsilon_2}{\varepsilon_1 - \varepsilon_2}, \tag{20.72}$$

$$\sin^2\theta_B = \frac{\varepsilon_2}{\varepsilon_1 + \varepsilon_2}, \quad \sin^2\theta_c = \frac{\varepsilon_2}{\varepsilon_1}, \quad \sin^2\theta_m = \frac{2\varepsilon_2}{\varepsilon_1 + \varepsilon_2}. \tag{20.73}$$

At the extremum the phase difference $\Delta = \delta_p - \delta_s$ is given by

$$\Delta_m = 4\theta_B - 2\pi, \tag{20.74}$$

and the ratio of the reflection amplitudes takes the value

$$\frac{r_p}{r_s} = \frac{\varepsilon_1^2 + \varepsilon_2^2 - 6\varepsilon_1\varepsilon_2 + i4(\varepsilon_1\varepsilon_2)^{\frac{1}{2}}(\varepsilon_1 - \varepsilon_2)}{(\varepsilon_1 + \varepsilon_2)^2} \quad (\theta_1 = \theta_m). \tag{20.75}$$

At θ_m the trajectory of r_p/r_s in the complex plane is farthest to the right on the unit circle. The phase difference $\Delta = \delta_p - \delta_s$ is shown in Fig. 20.5, together with that for a homogeneous layer.

Reflection phases for a homogeneous layer. For a layer of dielectric constant ε and of thickness Δz, the s and p reflection amplitudes are given by (2.52) and (2.68):

$$r_s = \frac{q(q_1 - q_2)c + i(q^2 - q_1 q_2)s}{q(q_1 + q_2)c - i(q^2 + q_1 q_2)s}, \tag{20.76}$$

$$-r_p = \frac{Q(Q_1 - Q_2)c + i(Q^2 - Q_1 Q_2)s}{Q(Q_1 + Q_2)c - i(Q^2 + Q_1 Q_2)s}. \tag{20.77}$$

Here $q^2 = (\omega/c)^2\left(\varepsilon - \varepsilon_1 \sin^2\theta_1\right)$ and $c = \cos q\Delta z, s = \sin q\Delta z, Q = q/\varepsilon$; the film extends from $z = 0$ to Δz. When $\varepsilon_1 > \varepsilon > \varepsilon_2$ we have to consider three ranges of $\theta_1 : \theta_1 \le \theta_c = \arcsin(\varepsilon_2/\varepsilon_1)^{\frac{1}{2}}, \theta_c \le \theta_1 \le \theta_c' = \arcsin(\varepsilon/\varepsilon_1)^{\frac{1}{2}}$, and $\theta_1 \ge \theta_c'$. In the first range q_1, q_2, q are all real, in the second $q_2 = i|q_2|$ and q_1, q are real, and in the third $q_2 = i|q_2|$ and $q = i|q|$. Of particular interest to the beam shift to be discussed in the next section is the behaviour of the phases for θ_1 slightly above θ_c We again find a square root singularity, with

$$\delta_s(\theta_1) = \delta_s(\theta_c) - \frac{2|q_2|/q_{1c}}{1 - \frac{\varepsilon_1 - \varepsilon}{\varepsilon_1 - \varepsilon_2}\sin^2 q_c\Delta z} + O(|q_2|^2), \tag{20.78}$$

$$\delta_p(\theta_1) = \delta_p(\theta_c) - \frac{2|Q_2|/Q_{1c}}{1 - \left(1 - \frac{\varepsilon_1^2(\varepsilon_1 - \varepsilon)}{\varepsilon^2(\varepsilon_1 - \varepsilon_2)}\right)\sin^2 q_c\Delta z} + O(|q_2|^2), \tag{20.79}$$

where

$$q_{1c}^2 = \frac{\omega^2}{c^2}(\varepsilon_1 - \varepsilon_2), \quad q_c^2 = \frac{\omega^2}{c^2}(\varepsilon - \varepsilon_2), \tag{20.80}$$

The phases at the critical angle are given by

$$\delta_s(\theta_c) = 2\arctan\left\{\left(\frac{\varepsilon - \varepsilon_2}{\varepsilon_1 - \varepsilon_2}\right)^{\frac{1}{2}}\tan q_c\Delta z\right\}, \tag{20.81}$$

$$\delta_p(\theta_c) = -\pi + 2\arctan\left\{\frac{\varepsilon_1}{\varepsilon}\left(\frac{\varepsilon - \varepsilon_2}{\varepsilon_1 - \varepsilon_2}\right)^{\frac{1}{2}}\tan q_c\Delta z\right\}. \tag{20.82}$$

The numerators in (20.78) and (20.79) are the sharp boundary values, and have been expressed in terms of $\Delta\theta = \theta_1 - \theta_c$ in (20.65) and (20.68). For the s wave the coefficient of $(\Delta\theta)^{\frac{1}{2}}$ is larger for the homogenous layer than for the Fresnel case; for the p wave it can be larger or smaller, depending on the dielectric constants.

Figure 20.4 shows the s and p phase shifts and Fig. 20.5 their difference for $(\omega/c)\Delta z = \frac{1}{2}$. Note that there is no square root singularity in δ_s or δ_p at θ_c' where q passes through zero, the s and p phases having the variation $\delta = \delta\left(\theta_c'\right) + O(q^2)$ with

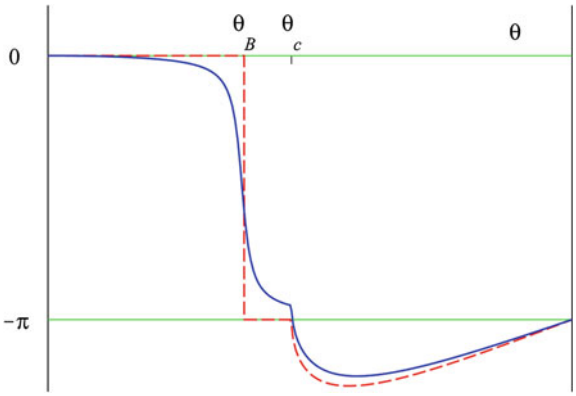


Fig. 20.5 Dependence of $\Delta = \delta_p - \delta_s$ on the angle of incidence, the parameters being as in Fig. 20.4. Normal incidence is at *left*, glancing incidence at *right*. *Dashed curve* glass|air; *solid curve* glass|water|air. The homogeneous film r_p/r_s ratio was shown in Fig. 2.8

$$\delta_s(\theta_c') = -2\arctan\left\{\frac{[(\varepsilon-\varepsilon_2)/(\varepsilon_1-\varepsilon)]^{\frac{1}{2}}}{1+(\varepsilon-\varepsilon_2)^{\frac{1}{2}}(\omega/c)\Delta z}\right\}, \tag{20.83}$$

$$\delta_p(\theta_c') = -\pi - 2\arctan\left\{\frac{\frac{\varepsilon_1}{\varepsilon_2}[(\varepsilon-\varepsilon_2)/(\varepsilon_1-\varepsilon)]^{\frac{1}{2}}}{1+\frac{\varepsilon}{\varepsilon_2}(\varepsilon-\varepsilon_2)^{\frac{1}{2}}(\omega/c)\Delta z}\right\}. \tag{20.84}$$

Thus θ_c' is not a true critical angle, even though q makes a sharp right-angle turn in the complex plane at θ_c', just as q_2 does at θ_c.

Total Reflection by the Hyperbolic Tangent Profile

For $\theta_1 < \theta_c$ the phase for the s polarization is given by (2.89); as θ_c is approached from below the phase tends without singularity to

$$\delta_s(\theta_c) = 2\sum_{n=1}^{\infty}\arctan\left\{\frac{2y_{1c}^3}{n(n^2+3y_{1c}^2)}\right\}, \tag{20.85}$$

where $y_{1c} = q_{1c}a, a$ being the length characterizing the thickness of the profile, and $q_{1c} = (\omega/c)(\varepsilon_1-\varepsilon_2)^{\frac{1}{2}}$. For $\theta_1 > \theta_c$ an analysis based on (2.84) and using the infinite product representation of the gamma function (2.85) gives

$$\delta_s = 2\sum_1^\infty \arctan\left\{\frac{2y_1(y_1^2+|y_2|^2)}{n(n^2+3y_2^2-|y_2|^2)}\right\} - 2\arctan\left(\frac{\tan\pi|y_2|}{\tanh\pi y_1}\right), \tag{20.86}$$

where $y_1 = q_1 a, |y_2| = |q_2|a$. Thus there is again a $|q_2|$ term in the phase just above the critical angle:

$$\delta_s = \delta_s(\theta_c) - \frac{2\pi a|q_2|}{\tanh\pi a q_{1c}} + O(q_2^2). \tag{20.87}$$

The expression (20.86) tends to the sharp profile result (20.63) as $a \to 0$. For large interfacial thickness the coefficient of $|q_2|$ in (20.86) tends to $-2\pi a$, and the strength of the square root singularity is then proportional to the thickness. At grazing incidence $\delta_s \to -\pi$ as before.

The above examples are sufficient to make it plausible that the $(\theta_1 - \theta_c)^{\frac{1}{2}}$ singularity in the phase shift is a universal property. We shall give a proof for the restricted class of finite-ranged profiles, for which the s wave reflection amplitude is given by (2.25), which we write in the form

$$r_s = \frac{q_1 q_2 A + iq_1 B + iq_2 C - D}{q_1 q_2 A + iq_1 B - iq_2 C + D}. \tag{20.88}$$

(We again set $z_1 = 0$: the inhomogeneity extends from $z = 0$ to Δz; the substrate has dielectric function ε_2.) When $\theta_1 > \theta_c$ we have $q_2 = i|q_2|$ and

$$r_s = \frac{-\alpha + i\beta}{\alpha + i\beta}, \quad \alpha = |q_2|C + D, \quad \beta = q_1(|q_2|A + B). \tag{20.89}$$

Thus $\delta_s = 2\arctan(\alpha/\beta)$. The leading terms in α/β near θ_c are

$$\frac{\alpha}{\beta} = \frac{D}{q_1 B} - \frac{|q_2|}{q_1}\left(\frac{W}{B}\right)^2 + O\left(|q_2|^2\right), \tag{20.90}$$

where W is the Wronskian of the solutions of the wave equation; we have used the identity $AD - BC = W^2$ equation (2.31). This shows that all such profiles have a term linear in $|q_2|$, with negative coefficient, leading to a square root singularity:

$$\delta_s = 2\arctan\left(\frac{D}{q_1 B}\right)_c - \frac{2|q_2|}{q_{1c}}\frac{(W/B)_c}{1+(D/q_1 B)_c^2} + O\left(|q_2|^2\right). \tag{20.91}$$

(A similar result may be written down for δ_p, using (2.40), (2.48) and (2.49).) For the homogeneous layer, with solutions $\sin qz$ and $\cos qz$ in $0 \le z \le \Delta z$, we have $W = q, B = q\cos q\Delta z, D = q^2 \sin q\Delta z$, and (20.91) gives the results contained in (20.78) and (20.81).

Appendix 2: Polarization of Electromagnetic Beams

In most of the book we have considered two linearly polarized waves, the *s* and *p* polarizations of Sects. 1.1 and 1.2. By these designations we mean that the electric vector **E** is respectively perpendicular and parallel to the plane of incidence. For plane waves, the corresponding magnetic vector **B** is respectively parallel and perpendicular to the plane of incidence; **E** and **B** are perpendicular to the wavevector, as well as to each other. Plane waves with **E** and **B** both circularly polarized were the eigenstates propagating within chiral media, Chap. 18. Here we shall discuss the most general polarization of a coherent *monochromatic* electromagnetic wave, and gives examples of the polarization properties of electromagnetic beams.

A coherent monochromatic light beam is specified by electric and magnetic vectors varying in space and harmonically in time. In general the polarization properties of the electric and magnetic vectors differ from each other, in contrast to the plane wave idealization. Most polarizers act on the electric field, and most detectors sense the electric field, so it is conventional to refer to the polarization of the electric field as *the* polarization. For monochromatic waves of angular frequency ω we can write

$$\mathbf{E}(\mathbf{r},t) = \mathrm{Re}\left\{\mathbf{E}(\mathbf{r})\mathrm{e}^{-i\omega t}\right\} = \mathbf{E}_r(\mathbf{r})\cos\omega t + \mathbf{E}_i(\mathbf{r})\sin\omega t, \tag{20.92}$$

where $\mathbf{E}_r(\mathbf{r})$ and $\mathbf{E}_i(\mathbf{r})$ are the real and imaginary parts of the complex electric field vector $\mathbf{E}(\mathbf{r})$. The magnetic field is expressed in terms of the real and imaginary parts of the complex vector $\mathbf{B}(\mathbf{r})$ in the same way. For a plane wave in vacuum we have $\mathbf{E}(\mathbf{r}) = \mathbf{E}_0\mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}}, \mathbf{B}(\mathbf{r}) = k^{-1}\mathbf{k}\times\mathbf{E}(\mathbf{r})$ where $k = \omega/c$ and the wavevector $\mathbf{k}$ defines the direction of propagation. If the constant vector $\mathbf{E}_0$ is real (or more generally, if its real and imaginary parts are collinear), it defines the direction of *linear* polarization. If the complex vector $\mathbf{E}_0$ has equal and perpendicular real and imaginary parts, as in the plane wave $\mathbf{E}(\mathbf{r}) = E_0\mathrm{e}^{ikz}(\hat{\mathbf{x}} \pm i\hat{\mathbf{y}})$, the physical electric vector $\mathbf{E}(\mathbf{r},t)$ rotates at any point in space with angular frequency ω, and the wave is *circularly* polarized. The most general case is that of elliptic polarization, in which the endpoint of the vector $\mathbf{E}(\mathbf{r},t)$ describes an ellipse in time $2\pi/\omega$, as we shall now show.

For any $\mathbf{E}(\mathbf{r}) = \mathbf{E}_r(\mathbf{r}) + i\mathbf{E}_i(\mathbf{r})$ one can write

$$\mathbf{E}_r + i\mathbf{E}_i = (\mathbf{E}_1 + i\mathbf{E}_2)\mathrm{e}^{i\gamma}, \tag{20.93}$$

and γ can be chosen so that the real vectors $\mathbf{E}_1$ and $\mathbf{E}_2$ are perpendicular. This value of γ and the components $\mathbf{E}_1$ and $\mathbf{E}_2$ are given by

$$\tan 2\gamma = \frac{2\mathbf{E}_r\cdot\mathbf{E}_\mathbf{i}}{E_r^2 - E_i^2}, \tag{20.94}$$

$$\mathbf{E}_1 = \mathbf{E}_r \cos\gamma + \mathbf{E}_i \sin\gamma, \quad \mathbf{E}_2 = \mathbf{E}_i \cos\gamma - \mathbf{E}_r \sin\gamma. \tag{20.95}$$

Thus the physical electric field can be written as

$$\mathbf{E}(\mathbf{r},t) = \mathrm{Re}\left\{(\mathbf{E}_1 + i\mathbf{E}_2)\mathrm{e}^{i\gamma - i\omega t}\right\} = \mathbf{E}_1(\mathbf{r})\cos(\omega t - \gamma) + \mathbf{E}_2(\mathbf{r})\sin(\omega t - \gamma). \tag{20.96}$$

When γ is given by (20.94), the components $\mathbf{E}_1$ and $\mathbf{E}_2$ are orthogonal, and have magnitudes given by

$$\begin{pmatrix} E_1^2 \\ E_2^2 \end{pmatrix} = \frac{1}{2}\left\{E_r^2 + E_i^2 \pm \left[\left(E_r^2 - E_i^2\right)^2 + 4(\mathbf{E}_r \cdot \mathbf{E}_i)^2\right]^{\frac{1}{2}}\right\}. \tag{20.97}$$

From (20.96), E_1 and E_2 give the lengths of the semiaxes of the polarization ellipse.

For *linear* polarization $E_2 = 0$; the condition for linear polarization is therefore that $\mathbf{E}_r, \mathbf{E}_i$ be collinear,

$$E_r^2 E_i^2 = (\mathbf{E}_r \cdot \mathbf{E}_i)^2. \quad (\textit{linear polarization}) \tag{20.98}$$

For *circular* polarization $E_1^2 = E_2^2$, for which we need $\mathbf{E}_r$ and $\mathbf{E}_i$ to be perpendicular *and* equal in magnitude:

$$\mathbf{E}_r \cdot \mathbf{E}_i = 0 \text{ and } E_r^2 = E_i^2. \quad (\textit{circular polarization}) \tag{20.99}$$

One spatial function can define the local *degree of linear polarization* (Lekner 2003), namely

$$\Lambda(\mathbf{r}) = \frac{E_1^2 - E_2^2}{E_1^2 + E_2^2} = \frac{\left[\left(E_r^2 - E_i^2\right)^2 + 4(\mathbf{E}_r \cdot \mathbf{E}_i)^2\right]^{\frac{1}{2}}}{E_r^2 + E_i^2} = \frac{\left|\mathbf{E}^2(\mathbf{r})\right|}{\left|\mathbf{E}(\mathbf{r})\right|^2}. \tag{20.100}$$

$\Lambda(\mathbf{r})$ is unity when the real and imaginary parts of $\mathbf{E}(\mathbf{r}) = \mathbf{E}_r(\mathbf{r}) + i\mathbf{E}_i(\mathbf{r})$ are collinear (the linear polarization condition), and zero when the circular polarization conditions are met. Equivalently, the eccentricity e of the polarization ellipse provides the same information:

$$\mathrm{e}^2 = 1 - \frac{E_2^2}{E_1^2} = \frac{2\Lambda}{1+\Lambda}. \tag{20.101}$$

This has the same values as Λ of unity and zero for the limiting cases of linear and circular polarizations. Yet more polarization measures exist, namely the Hurwitz (1945) ratio $2E_1E_2/(E_1^2 + E_2^2)$ and the Stokes parameters (Born and Wolf 1999, Sects. 1.4.2, 10.8.3)

$$S_0 = E_r^2 + E_i^2, \quad S_1 = E_r^2 - E_i^2, \quad S_3 = 2\mathbf{E}_r \cdot \mathbf{E}_i, \quad S_4 = 2\left[E_r^2 E_i^2 - (\mathbf{E}_r \cdot \mathbf{E}_i)^2\right]^{\frac{1}{2}}. \tag{20.102}$$

The relation between the degree of linear polarization Λ and the Stokes parameters is $\Lambda^2 = 1 - S_3^2/S_0^2$.

The remainder of this Appendix gives specific examples of polarization properties of finite monochromatic electromagnetic beams. A broader range of topics may be found in Swindell (1975), a collection of reprints of fundamental papers on polarized light with commentary, and in monographs by Collett (1992), Huard (1997) and Brosseau (1998).

Examples of Exactly and Approximately Linearly Polarized Beams

The simplest example is the TM (transverse magnetic) beam, for which the vector potential is given by the first entry in (20.26), $\mathbf{A} = A_0(0, 0, \psi)$. For this beam $\mathbf{B}$ is transverse to the propagation direction (here along the z axis):

$$\mathbf{B} = \nabla \times \mathbf{A} = A_0\big(\partial_x\psi, -\partial_y\psi, 0\big). \tag{20.103}$$

When ψ is independent of the azimuthal angle ϕ, the complex fields are

$$\begin{aligned} \mathbf{B}(\mathbf{r}) &= A_0\big(\sin\phi\partial_\rho\psi, -\cos\phi\partial_\rho\psi, 0\big), \\ \mathbf{E}(\mathbf{r}) &= \frac{iA_0}{k}\left(\cos\phi\partial_\rho\partial_z\psi, \sin\phi\partial_\rho\partial_z\psi, \partial_z^2\psi + k^2\psi\right). \end{aligned} \tag{20.104}$$

If we take A_0 real, and write the complex wavefunction $\psi(\rho, z)$ as $\psi_r + i\psi_i$, the real and imaginary parts of $\mathbf{B}(\mathbf{r})$ are both proportional to $(\sin\phi, -\cos\phi, 0)$, and are thus collinear. The magnetic field is therefore everywhere linearly polarized. The electric field is elliptically polarized, in general.

The dual of the TM beam under the transformation $\mathbf{E} \to \mathbf{B}, \mathbf{B} \to -\mathbf{E}$ (one of a set of duality transformations that leave the free space Maxwell equations unchanged) is the TE beam, transverse and linearly polarized in its electric field. The electric field lines are circles concentric with the beam axis (see Fig. 20.1 of Lekner 2003, for example). However, both the TM and the TE beams disappear in the plane-wave limit: as $\psi \to \exp ikz$ the electric and magnetic fields in both the TM and the TE beams tend to zero.

A beam which does have a plane-wave limit is the 'LP' beam, with vector potential $\mathbf{A} = A_0(\psi, 0, 0)$. The magnetic and electric fields are

$$\mathbf{B} = \nabla \times \mathbf{A} = A_0\big(0, \partial_z\psi, -\partial_y\psi\big), \tag{20.105}$$

$$\mathbf{E} = \frac{i}{k}\nabla(\nabla \cdot \mathbf{A}) + ik\mathbf{A} = \frac{iA_0}{k}(\partial_x^2\psi + k^2\psi, \partial_x\partial_y\psi, \partial_x\partial_z\psi) \tag{20.106}$$

In the plane wave limit $\psi \to \exp ikz, \mathbf{B} \to ikA_0(0,1,0), \mathbf{E} \to ikA_0(1,0,0)$, which is the textbook linearly polarized plane wave with $\mathbf{E}$ and $\mathbf{B}$ transverse and mutually perpendicular. But note the quotes around 'LP': this beam is linearly polarized only in the plane-wave limit, as we shall now see. We again consider beams with ψ is independent of the azimuthal angle ϕ, for simplicity. Then

$$\mathbf{B} = ik\big(0, \partial_z\psi, -\sin\phi\partial_\rho\psi\big), \tag{20.107}$$

$$\begin{aligned}\mathbf{E} = \frac{iA_0}{k}(&\cos^2\phi\partial_\rho^2\psi + \sin^2\phi\rho^{-1}\partial_\rho\psi + k^2\psi,\\ &\sin\phi\cos\phi[\partial_\rho^2\psi - \rho^{-1}\partial_\rho\psi], \cos\phi\partial_\rho\partial_z\psi).\end{aligned} \tag{20.108}$$

Neither $\mathbf{E}$ nor $\mathbf{B}$ have real and imaginary parts collinear in general. The electric field in the $x = 0$ plane ($\cos\phi = 0$) is linearly polarized along the x direction, as can be seen from (20.108). The polarization measure Λ is given in (20.27) and is plotted for the $\psi_{00} = j_0(kR)$ beam in Fig. 20.3 of Lekner (2003) for $kb = 2$, and in Fig. 20.6 for $kb = 6$. The polarization is linear at the beam centre, and $\Lambda \approx 1$ in the

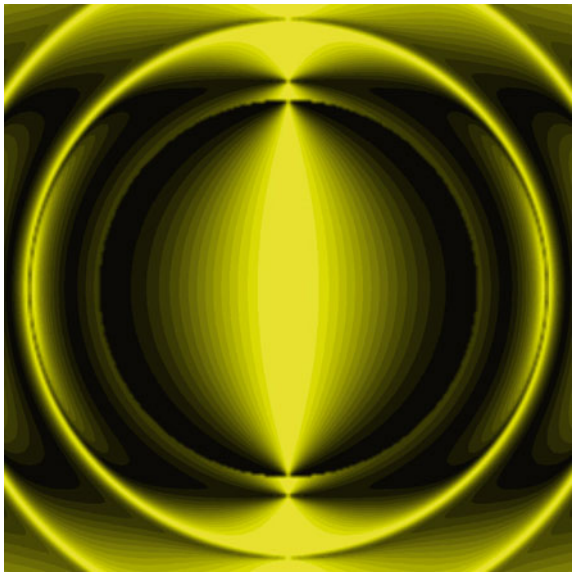


Fig. 20.6 Degree of linear polarization Λ in the focal plane $z = 0$ of an 'LP' beam with $\psi = \psi_{00}, kb = 6$. The light shading corresponds to linear polarization, dark to circular polarization ($\Lambda \to 1$ and $\Lambda \to 0$, respectively). The lateral extent is $|kx| \le 9, |ky| \le 9$

central region $\rho \ll b$, but, remarkably, there are areas of approximately circular polarization in the outer part of the beam.

Approximately Circularly Polarized Beams

We wish to construct beams which in the plane wave limit have the circularly polarized electric field

$$\begin{aligned}&\mathbf{E}(\mathbf{r}) = E_0 \mathrm{e}^{ikz}(1, i, 0),\\ &\mathbf{E}(\mathbf{r}, t) = \mathrm{Re}\big(\mathbf{E}(\mathbf{r})\mathrm{e}^{-i\omega t}\big) = E_0(\cos(kz - \omega t), -\sin(kz - \omega t), 0).\end{aligned} \tag{20.109}$$

The vector potential $\mathbf{A} = k^{-1}E_0(i\psi, -\psi, 0)$ gives the complex fields

$$\begin{aligned}\mathbf{B} &= k^{-1}E_0\big[\partial_z, i\partial_z, -(\partial_x + i\partial_y)\big]\psi,\\ \mathbf{E} &= E_0\big[1 + k^{-2}\partial_x\big(\partial_x + i\partial_y\big), i + k^{-2}\partial_y\big(\partial_x + i\partial_y\big), k^{-2}\partial_z\big(\partial_x + i\partial_y\big)\big]\psi.\end{aligned} \tag{20.110}$$

Fig. 20.7 Degree of linear polarization Λ in the focal plane of a 'CP' beam with $\psi = \psi_{00}, kb = 6$. The light shading corresponds to linear polarization, dark to circular polarization ($\Lambda \to 1$ and $\Lambda \to 0$, respectively). The lateral extent is $|kx| \le 9, |ky| \le 9$. The beam is completely circularly polarized on the axis, and approximately so in the *central dark region*. However, there are *circles* of exactly linear polarization in the outer part

(We have used the fact that ψ satisfies the Helmholtz equation (20.1).) Both the magnetic and the electric fields are therefore circularly polarized, with positive helicity, in the plane wave limit $\psi \to \exp ikz$. Figure 20.7 shows the polarization measure Λ for the electric field of (20.110), with $\psi = \sin kR/kR, kb = 6$, which has the focal plane zeros shown in Fig. 20.1. We note that the dark central part is circularly polarized, but the outer region of the beam (where the intensity is very low) there are circles of linear polarization. More analytic detail may be found in Section 4 and Appendix B of Lekner (2003).

These examples illustrate the theorems (ii) and (iii) of Sect. 20.1, and show that finite beams are quite different from the textbook plane waves, not just in having longitudinal components, but also in their polarization properties.

References

Bateman H (1904) The solution of partial differential equations by means of definite integrals. Proc Lond Math Soc 1:451–458
Barnett SM (2002) Optical angular momentum flux. J Opt B: Quantum Semiclass 4:S7–S16
Born M, Wolf E (1999) Principles of optics, 7edn. Cambridge University Press, Cambridge
Boyack R, Lekner J (2011) Non-existence of separable spheroidal beams. J Opt 13:085701 (3 pp)
Brekhovskikh LM (1980) Waves in layered media, 2nd edn. Academic Press, New York
Brosseau C (1998) Fundamentals of polarized light: a statistical optics approach. Wiley, New York
Collett E (1992) Polarized light, fundamentals and applications. Dekker, New York
Deschamps GA (1971) Gaussian beam as a bundle of complex rays. Electron Lett 7:684–685
Goos F, Hänchen H (1947) Ein neuer und fundamentaler Versuch zur Totalreflexion. Ann Phys 1:333–346
Horowitz BR, Tamir T (1971) Lateral displacement of a light beam at a dielectric interface. J Opt Soc Amer 61:586–594
Huard S (1997) Polarization of light. Wiley, New York
Jackson JD (1975) Classical electrodynamics, 2nd edn. Wiley, New York
Lekner J (2001) TM, TE and 'TEM' beam modes: exact solutions and their problems. J Opt A: Pure Appl Opt 3:407–412
Lekner J (2002) Phase and transport velocities in particle and electromagnetic beams. J Opt A: Pure Appl Opt 4:491–499
Lekner J (2003) Polarization of tightly focused laser beams. J Opt A: Pure Appl Opt 5:6–14
Lekner J (2004a) Invariants of atom beams. J Phys B: At Mol Opt Phys 37:1725–1736
Lekner J (2004b) Invariants of electromagnetic beams. J Opt A: Pure Appl Opt 6:204–209
Lekner J (2004c) Invariants of three types of generalized Bessel beams. J Opt A: Pure Appl Opt 6:837–843
Lekner J (2007) Acoustic beam invariants. Phys Rev E 75:036610 (6 pp)
Lotsch HKV (1970) Beam displacement at total reflection: the Goos-Hänchen effect. Parts I to IV, Optik 32:116–137, 189–204, 299–319 and 553–569
Nye JF (1999) Natural focusing and fine structure of light. Inst Phys Publishing, Bristol
Nye JF, Berry MV (1974) Dislocations in wave trains. Proc R Soc Lond A 336:165–190
Sheppard CJR, Saghafi S (1998) Beam modes beyond the paraxial approximation: a scalar treatment. Phys Rev A 57:2971–2979
Snyder A W and Love J D (1983) Optical waveguide theory. Chapman and Hall, London
Swindell W (ed) (1975) Polarized light. In: Benchmark papers in optics, vol 1. Wiley

Ulanowski Z, Ludlow IK (2000) Scalar field of nonparaxial Gaussian beams. Opt Lett 25: 1792–1794
Watson GN (1966) A treatise on the theory of Bessel functions, 2edn. Cambridge University Press, Cambridge
White IA, Pask C (1977) Effect of Goos-Hänchen shifts on pulse widths in optical waveguides. Appl Opt 16:2353–2355
Wolter H (1950) Untersuchungen zur Strahlversetzung bei Totalreflexion des Lichtes mit der Methode der Minimumstrahlkennzeichnung. Z. Naturforschung 5a:143–153
Zangwill A (2013) Modern electrodynamics. Cambridge University Press, Cambridge

Further Readings

Additional references on beam shifts and related topics

Riesz RP, Simon R (1985) Reflection of a Gaussian beam from a dielectric slab. J Opt Soc Am 2A:1809–1817
Tamir T (1982) The lateral wave (Chapter 13), and Chen WP, Burstein E (1982) Narrow beam excitation of electromagnetic modes in prism configurations (Chapter 14). In: Boardman AD (ed) Electromagnetic surface modes. Wiley
White IA, Snyder AW, Pask C (1977) Directional change of beams undergoing partial reflection. J Opt Soc Am 67:703–705

Section 20.1 is based on

Lekner J (2013) Universal properties of electromagnetic beams. In: PIERS Proceedings Taipei, pp 464–469

Appendix
Reflection and Transmission Formulae

$$n_1 = \sqrt{\varepsilon_1}, \quad n_2 = \sqrt{\varepsilon_2}, \quad n_1 \sin\,\theta_1 = \frac{cK}{\omega} = n_2 \sin\theta_2,$$
$$q_1 = n_1 \frac{\omega}{c}\cos\theta_1, \quad q_2 = n_2 \frac{\omega}{c}\cos\theta_2$$

Critical angle $\theta_c = \arcsin\frac{n_2}{n_1}, \quad q_2(\theta_c) = 0, \quad \frac{cK_c}{\omega} = n_2$

***s*wave**

$$E_y(x,z,t) = \mathrm{e}^{i(Kx-\omega t)}E(z), \quad \frac{\mathrm{d}^2E}{\mathrm{d}z^2} + q^2E = 0, \quad q^2 = \varepsilon\frac{\omega^2}{c^2} - K^2 = k^2 - K^2$$

$$\mathrm{e}^{iq_1z} + r_s\,\mathrm{e}^{-iq_1z} \leftarrow E(z) \rightarrow t_s\,\mathrm{e}^{iq_2t}$$

***p*wave**

$$B_y(x,z,t) = \mathrm{e}^{i(Kx-\omega t)}B(z), \quad \frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{\varepsilon}\frac{\mathrm{d}B}{\mathrm{d}z}\right) + \left(\frac{\omega^2}{c^2} - \frac{K^2}{\varepsilon}\right)B = 0$$

$$\mathrm{e}^{iq_1z} - r_p\,\mathrm{e}^{-iq_1z} \leftarrow B(z) \rightarrow \frac{n_2}{n_1}t_p\,\mathrm{e}^{iq_2t}$$

Step profile

$$r_{s0} = \frac{q_1 - q_2}{q_1 + q_2}, \quad t_{s0} = \frac{2q_1}{q_1 + q_2}$$

$$r_{p0} = \frac{Q_2 - Q_1}{Q_2 + Q_1}, \quad \frac{n_2}{n_1}t_{p0} = \frac{2Q_1}{Q_1 + Q_2}, \quad Q_1 = \frac{q_1}{\varepsilon_1}, Q_2 = \frac{q_2}{\varepsilon_2}$$

$$\theta_B = \arctan\frac{n_2}{n_1}, Q_1^2 = Q_2^2 = \frac{(\omega/c)^2}{\varepsilon_1 + \varepsilon_2} = Q_B^2, \varepsilon_1\varepsilon_2 Q_B^2 = K_B^2 = \frac{\varepsilon_1\varepsilon_2}{\varepsilon_1 + \varepsilon_2}\left(\frac{\omega}{c}\right)^2$$

J. Lekner, *Theory of Reflection*, Springer Series on Atomic, Optical, and Plasma Physics 87, DOI 10.1007/978-3-319-23627-8

General profile

$$q_1\left(1-|r|^2\right)=q_2|t|^2,\quad q_2t_{12}=q_1t_{21},\quad r_{21}=-\frac{t_{12}}{t_{12}^*}r_{12}^*$$

If $F(z)$ and $G(z)$ are solutions of the s-wave equation $\frac{d^2E}{dz^2}+q^2E=0$ in the region $z_1\le z\le z_2$, bounded by homogeneous media 1 and 2,

$$r_s=\mathrm{e}^{2iq_1z_1}\frac{q_1q_2(F_1G_2-G_1F_2)+iq_1\left(F_1G_2'-G_1F_2'\right)+iq_2\left(F_1'G_2-G_1'F_2\right)-\left(F_1'G_2'-G_1'F_2'\right)}{q_1q_2(F_1G_2-G_1F_2)+iq_1\left(F_1G_2'-G_1F_2'\right)-iq_2\left(F_1'G_2-G_1'F_2\right)+\left(F_1'G_2'-G_1'F_2'\right)}$$

$$t_s=\frac{\mathrm{e}^{i(q_1z_1-q_2z_2)}2iq_1(F_2G_2'-G_2F_2')}{q_1q_2(F_1G_2-G_1F_2)+iq_1\left(F_1G_2'-G_1F_2'\right)-iq_2\left(F_1'G_2-G_1'F_2\right)+\left(F_1'G_2'-G_1'F_2'\right)}$$

Glancing incidence: $\theta_1\to\pi/2$, $q_1\to 0$, $r_s\to -1, r_p\to 1$ (all profiles).
[For ε continuous at z_1 and z_2 the p wave reflection and transmission amplitudes are given by (2.40) and (2.41); discontinuities in ε give formulae of the form (17.20) and (17.21).]

Homogeneous layer, thickness Δz:

$$r_s=\mathrm{e}^{2iq_1z_1}\frac{r+r'\mathrm{e}^{2iq\Delta z}}{1+rr'\mathrm{e}^{2iq\Delta z}},r=\frac{q_1-q}{q_1+q},r'=\frac{q-q_2}{q+q_2}$$

$$t_s=\mathrm{e}^{i(q_1z_1-q_2z_2)}\frac{(1+r)(1+r')\mathrm{e}^{iq\Delta z}}{1+rr'\mathrm{e}^{2iq\Delta z}}$$

$$r_s=\mathrm{e}^{2iq_1z_1}\frac{q(q_1-q_2)\cos q\Delta z+i(q^2-q_1q_2)\sin q\Delta z}{q(q_1+q_2)\cos q\Delta z-i(q^2+q_1q_2)\sin q\Delta z}$$

$$-r_p=\mathrm{e}^{2iq_1z_1}\frac{r+r'e^{2iq\Delta z}}{1+rr'e^{2iq\Delta z}},r=\frac{Q_1-Q}{Q_1+Q},r'=\frac{Q-Q_2}{Q+Q_2},Q_i=\frac{q_i}{\varepsilon_i},Q=\frac{q}{\varepsilon}$$

$$\frac{n_2}{n_1}t_p=\mathrm{e}^{i(q_1z_1-q_2z_2)}\frac{(1+r)(1+r')\mathrm{e}^{iq\Delta z}}{1+rr'\mathrm{e}^{2iq\Delta z}}$$

$$-r_p=\mathrm{e}^{2iq_1z_1}\frac{Q(Q_1-Q_2)\cos q\Delta z+i(Q^2-Q_1Q_2)\sin q\Delta z}{Q(Q_1+Q_2)\cos q\Delta z-i(Q^2+Q_1Q_2)\sin q\Delta z}$$

Equations for the reflection amplitudes

$$\phi(z) = \int^{z} \mathrm{d}\zeta q(\zeta), \quad r' = \frac{q'}{2q}\left(\mathrm{e}^{2i\phi} - r^2\mathrm{e}^{-2i\phi}\right), \quad r_s \leftarrow r(z) \rightarrow 0$$

$$r_s = -\int_{-\infty}^{\infty} \mathrm{d}z \frac{q'}{2q}\left(\mathrm{e}^{2i\phi} - r^2\mathrm{e}^{-2i\phi}\right), \quad r_p = \int_{-\infty}^{\infty} \mathrm{d}z \frac{Q'}{2Q}\left(\mathrm{e}^{2i\phi} - r^2\mathrm{e}^{-2i\phi}\right)$$

Rayleigh or weak reflection approximation

$$r_s^R = -\int_{-\infty}^{\infty} \mathrm{d}z \frac{q'}{2q}\mathrm{e}^{2i\phi}, \quad r_p^R = \int_{-\infty}^{\infty} \mathrm{d}z \frac{Q'}{2Q}e^{2i\phi}$$

Absorption

$$q_2 = q_r + iq_i, \quad q_r^2 - q_i^2 = \frac{\omega^2}{c^2}\left(\varepsilon_r - \varepsilon_1 \sin^2\theta_1\right), \quad 2q_r q_i = \frac{\omega^2}{c^2}\varepsilon_i$$

$$\left(\frac{cq_r}{\omega}\right)^2 = \frac{1}{2}\left\{\varepsilon_r - \varepsilon_1 \sin^2\theta_1 + \left[\left(\varepsilon_r - \varepsilon_1 \sin^2\theta_1\right)^2 + \varepsilon_i^2\right]^{\frac{1}{2}}\right\}, \quad \frac{cq_i}{\omega} = \frac{\varepsilon_i/2}{cq_r/\omega}$$

$$Q_r = \frac{\varepsilon_r q_r + \varepsilon_i q_i}{\varepsilon_r^2 + \varepsilon_i^2}, \quad Q_i = \frac{\varepsilon_r q_i - \varepsilon_i q_r}{\varepsilon_r^2 + \varepsilon_i^2}$$

$$R_s = \frac{(q_1 - q_r)^2 + q_i^2}{(q_1 + q_r)^2 + q_i^2}, \quad R_p = \frac{(Q_1 - Q_r)^2 + Q_i^2}{(Q_1 + Q_r)^2 + Q_i^2}$$

Matrix methods

$$\begin{pmatrix} E_{n+1} \\ D_{n+1} \end{pmatrix} = \begin{pmatrix} \cos\delta_n & q_n^{-1}\sin\delta_n \\ -q_n \sin\delta_n & \cos\delta_n \end{pmatrix}\begin{pmatrix} E_n \\ D_n \end{pmatrix} = M_n \begin{pmatrix} E_n \\ D_n \end{pmatrix},$$
$$\delta_n = q_n(z_{n+1} - z_n), \quad D = \frac{dE}{dz}$$

$$\begin{pmatrix} B_{n+1} \\ C_{n+1} \end{pmatrix} = \begin{pmatrix} \cos\delta_n & Q_n^{-1}\sin\delta_n \\ -Q_n \sin\delta_n & \cos\delta_n \end{pmatrix}\begin{pmatrix} B_n \\ C_n \end{pmatrix}, Q_n = \frac{q_n}{\varepsilon_n}, C = \frac{1}{\varepsilon}\frac{dB}{dz}$$

$$M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} = M_N M_{N-1} \ldots M_n \ldots M_2 M_1$$

$$r_s = \mathrm{e}^{2i\alpha}\frac{q_a q_b m_{12} + m_{21} - iq_b m_{11} + iq_a m_{22}}{q_a q_b m_{12} - m_{21} + iq_b m_{11} + iq_a m_{22}}$$

$$t_s = \mathrm{e}^{i(\alpha-\beta)}\frac{2iq_a}{q_aq_bm_{12}-m_{21}+iq_bm_{11}+iq_am_{22}}$$

$$-r_p = \mathrm{e}^{2i\alpha}\frac{Q_aQ_bm_{12}+m_{21}-iQ_bm_{11}+iQ_am_{22}}{Q_aQ_bm_{12}-m_{21}+iQ_bm_{11}+iQ_am_{22}}$$

$$\left(\frac{\varepsilon_b}{\varepsilon_a}\right)^{1/2}t_p = \mathrm{e}^{i(\alpha-\beta)}\frac{2iQ_a}{Q_aQ_bm_{12}-m_{21}+iQ_bm_{11}+iQ_am_{22}}$$

$$\alpha = q_az_1, \quad \beta = q_bz_{N+1}$$

Periodically stratified media

$$M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}, \quad M^N = \begin{pmatrix} m_{11}S_N - S_{N-1} & m_{12}S_N \\ m_{21}S_N & m_{22}S_N - S_{N-1} \end{pmatrix}$$

$$S_N = \frac{\sin N\phi}{\sin\phi}, \qquad \cos\phi = \frac{1}{2}\,\mathrm{trace}\ M = \frac{1}{2}(m_{11}+m_{22}) \quad [\det M = 1]$$

$$\sigma_N = \frac{S_{N-1}}{S_N} = \frac{\sin[(N-1)\phi]}{\sin(N\phi)} = \cos\phi - \sin\phi\cot(N\phi)$$

$$r_s = \frac{q_1q_2m_{12}+m_{21}+iq_1(m_{22}-\sigma_N)-iq_2(m_{11}-\sigma_N)}{q_1q_2m_{12}-m_{21}+iq_1(m_{22}-\sigma_N)+iq_2(m_{11}-\sigma_N)}$$

$$t_s = \frac{2iq_1S_N^{-1}}{q_1q_2m_{12}-m_{21}+iq_1(m_{22}-\sigma_N)+iq_2(m_{11}-\sigma_N)}$$

$$-r_p = \frac{Q_1Q_2m_{12}+m_{21}+iQ_1(m_{22}-\sigma_N)-iQ_2(m_{11}-\sigma_N)}{Q_1Q_2m_{12}-m_{21}+iQ_1(m_{22}-\sigma_N)+iQ_2(m_{11}-\sigma_N)}$$

$$\frac{n_2}{n_1}t_p = \frac{2iQ_1S_N^{-1}}{Q_1Q_2m_{12}-m_{21}+iQ_1(m_{22}-\sigma_N)+Q_2(m_{11}-\sigma_N)}$$

Uniaxial anisotropy

$$q_o^2 = \varepsilon_o\left(\frac{\omega}{c}\right)^2 - K^2$$

q_e values found by solving $\varepsilon_\gamma q^2 + 2\alpha\gamma\Delta\varepsilon Kq + \varepsilon_\alpha K^2 - \varepsilon_o\varepsilon_e\left(\frac{\omega}{c}\right)^2 = 0$

$$q_{\pm} = \pm\bar{q} - \frac{\alpha\gamma K\Delta\varepsilon}{\varepsilon_\gamma}, \quad \bar{q}^2 = \frac{\varepsilon_o}{\varepsilon_\gamma^2}\left[\varepsilon_e\varepsilon_\gamma\left(\frac{\omega}{c}\right)^2 - \varepsilon_{\alpha\gamma}K^2\right]$$

$$\varepsilon_\gamma = \varepsilon_o + \gamma^2\Delta\varepsilon, \qquad \varepsilon_\alpha = \varepsilon_o + \alpha^2\Delta\varepsilon$$
$$\varepsilon_{\alpha\gamma} = \varepsilon_o + (\alpha^2+\gamma^2)\Delta\varepsilon = \varepsilon_o + (1-\beta^2)\Delta\varepsilon = \varepsilon_e - \beta^2\Delta\varepsilon$$

(α, β, γ are the direction cosines of the optic axis, relative to the laboratory coordinate axes) General expressions for reflection and transmission amplitudes for an isotropic|uniaxial boundary are given in Sect. 8.3.

Chiral isotropic media

$$\mathbf{D} = \varepsilon\mathbf{E} + i\gamma\mathbf{H}, \quad \mathbf{B} = \mu\mathbf{H} - i\gamma\mathbf{E}, \quad \gamma = \lambda\delta/2\pi d$$

The plane of polarization rotates by δ on passing normally through a chiral plate of thickness d.

Plane wave eigenstate normal components

$$q_\pm^2 = (\sqrt{\varepsilon\mu} \pm \gamma)^2\frac{\omega^2}{c^2} - K^2 \equiv k_\pm^2 - K^2, \quad n_\pm = \sqrt{\varepsilon\mu} \pm \gamma \equiv n \pm \gamma$$

Reflection amplitudes at an achiral|chiral boundary

$$c_1 = \cos\theta_1, \quad c_\pm = \cos\theta_\pm = \sqrt{1 - (n_1\sin\theta_1/n_\pm)^2}, \quad m = \sqrt{\varepsilon\mu_1/\varepsilon_1\mu}$$

$$D = c_1^2 + \frac{c_1(c_+ + c_-)(m + m^{-1})}{2} + c_+c_-$$

$$r_{ss} = \left[c_1^2 - \frac{c_1(c_+ + c_-)(m - m^{-1})}{2} - c_+c_-\right]/D$$

$$r_{pp} = -\left[c_1^2 + \frac{c_1(c_+ + c_-)(m - m^{-1})}{2} - c_+c_-\right]/D$$

Index

J. Lekner, *Theory of Reflection*, Springer Series on Atomic, Optical, and Plasma Physics 87, DOI 10.1007/978-3-319-23627-8