Multiple principal angles for a homogeneous layer

John Lekner

School of Chemical and Physical Sciences, Victoria University of Wellington, Wellington, New Zealand

Received 9 November 1999, in final form 2 March 2000

Abstract. The ellipsometric Brewster angle (principal angle) is defined by Re $(r_p/r_s) = 0$, where r_p and r_s are the reflection amplitudes for TM and TE waves. We show that there will always be multiple principal angles in reflection from a homogeneous layer on a substrate, if the layer is made thick enough. The principal angle θ_p ranges between the zero-thickness value $\theta_B = \arctan(n_2/n_1)$ (where n_1 and n_2 are the refractive indices of the medium of incidence and of the substrate), and a value θ_m which depends on the refractive index n of the layer. Analytic expressions for θ_m are obtained, for $n^2 < n_1n_2$ and for $n^2 > n_1n_2$. When $n^2 = n_1n_2$, θ_m is zero, and θ_p varies with thickness over the full range of zero to θ_B. In general, rapid variation of the principal angle with layer thickness is expected near odd multiples of $\lambda/4$ times a known function of the refractive indices. Monitoring θ_p during film growth can thus provide information about film thickness and about refractive indices.

Keywords: Ellipsometry, principal angle, Brewster angle, thin films

1. Introduction

In most ellipsometric experiments, the real and imaginary parts of r_p/r_s are monitored separately. A convenient angle of incidence to work at (and to lock onto in automated systems) is the *principal angle*. An arbitrary isotropic interface between two isotropic media will always have at least one principal angle (ellipsometric Brewster angle) at which the real part of the ratio of p to s reflection amplitudes passes through zero. In general, there will be an odd number of principal angles. These results follow from continuity arguments: see [1, section 2.4]. An illustration of triple principal angles for a homogeneous layer is shown in figure 2.8 of [1].

For a homogeneous layer of thickness Δz and dielectric constant $\varepsilon = n^2$ (n is the refractive index), with light incident from medium 1 and with the substrate designated by 2, the principal angle θ_p is a function of ε_1, ε, ε_2 and $\omega\Delta z/c$. In general, it is multivalued as a function of the thickness: at a given value of the thickness parameter $\tau = \omega\Delta z/c = 2\pi\Delta z/\lambda$ (c is the speed of light, ω is its angular frequency, and λ is the vacuum wavelength) there may be, for example, three values of θ_p.

This paper examines in detail the properties of θ_p for a homogeneous layer. One motivation is practical: θ_p is easily measured in polarization–modulation ellipsometry [2–6], and deductions are made from these measurements about the thickness of the layer, whose dielectric function is assumed known, as is that of the substrate. Clearly, there is a problem in extracting the thickness in regions where θ_p is multivalued.

Before considering the homogeneous layer, we note two examples of reflection from a bare substrate in which multiple principal angles can appear:

(i) In reflection from the sharp surface of an absorbing medium with dielectric function $\varepsilon_2 = \varepsilon_r + i\varepsilon_i$, there is a small domain in the $\varepsilon_2/\varepsilon_1$ complex plane within which there are three principal angles [7].

(ii) Even when the substrate is not absorbing, triple principal angles can appear in the total-reflection region, which exists when $\varepsilon_1 > \varepsilon_2$. The trajectory of $\rho = r_p/r_s$ is then the real axis, from +1 at $\theta_1 = 0$ to -1 at $\theta_1 = \theta_c = \arcsin(\varepsilon_2/\varepsilon_1)^{1/2}$, after which ρ climbs out along the unit circle, reaching an extremum at $\arcsin[2\varepsilon_2/(\varepsilon_1+\varepsilon_2)]^{1/2}$, and then retracing its path back to -1 at $\theta_1 = \pi/2$. (See figure 10.2 of [1].) The extremum value of ρ is given in equation (33) of section 10.2 in [1]. The real part of the extremum of ρ will be positive if $\varepsilon_1^2 + \varepsilon_2^2 > 6\varepsilon_1\varepsilon_2$, which happens when $\varepsilon_2/\varepsilon_1 < 3 - \sqrt{8} \approx 0.17$. In this case Re (ρ) will be zero at the three angles θ_B and $\theta_p^{\pm}$, given by

$$\tan^2\theta_B = \frac{\varepsilon_2}{\varepsilon_1},\qquad \tan^2\theta_p^{\pm} = \frac{\varepsilon_1 - \varepsilon_2 \pm \{\varepsilon_1^2 + \varepsilon_2^2 - 6\varepsilon_1\varepsilon_2\}^{1/2}}{2\varepsilon_1}. \tag{1}$$

2. Ellipsometric ratio for a non-absorbing layer

The ellipsometric ratio $\rho = r_p/r_s$ is given by [1, 2]

$$\rho = \frac{p_1 + p_2 Z}{1 + p_1 p_2 Z}\,\frac{1 + s_1 s_2 Z}{s_1 + s_2 Z} \tag{2}$$

where p_1, p_2, s_1 and s_2 are the p and s Fresnel reflection amplitudes at the boundaries of the layer with the medium

1464-4258/00/030239+07$30.00

of incidence and with the substrate (which have dielectric constants ε_1 and ε_2 respectively):

$$s_1 = \frac{q_1 - q}{q_1 + q} \qquad s_2 = \frac{q - q_2}{q + q_2}$$
$$p_1 = \frac{Q - Q_1}{Q + Q_1} \qquad p_2 = \frac{Q_2 - Q}{Q_2 + Q}. \tag{3}$$

Here, q_1, q and q_2 are the normal components of the wavevector in the medium of incidence, the layer and the substrate,

$$q_1^2 = (\omega/c)^2 \varepsilon_1 \cos^2\theta_1,$$
$$q^2 = (\omega/c)^2 (\varepsilon - \varepsilon_1 \sin^2\theta_1), \tag{4}$$
$$q_2^2 = (\omega/c)^2 (\varepsilon_2 - \varepsilon_1 \sin^2\theta_1)$$

and $Q_1 = q_1/\varepsilon_1$, $Q = q/\varepsilon$, $Q_2 = q_2/\varepsilon_2$, where ε is the dielectric constant of the layer. Finally, if Δz is the layer thickness,

$$Z = \exp(2\mathrm{i}q\Delta z). \tag{5}$$

Provided ε_1 is smaller that both ε and ε_2, and both the layer and the substrate are non-absorbing, all the wavevector normal components and all the Fresnel reflection amplitudes are real at all angles of incidence, and the thickness Δz contained in Z can be eliminated by use of $ZZ^* = 1$. The resulting equation defines the locus of ρ in the complex plane as the layer thickness varies [8]. This locus is parametrized by the angle of incidence θ_1 and by the dielectric functions ε_1, ε and ε_2. It is a quartic in $x = \mathrm{Re}\,(\rho)$ and $y = \mathrm{Im}\,(\rho)$, consisting of a closed curve plus an isolated point (acnode) on the real axis. Details are given in the appendix of [8].

At fixed layer thickness, ρ moves continuously in the complex plane from +1 at normal incidence to -1 at grazing incidence, cutting through the imaginary axis ($\mathrm{Re}\,(\rho) = 0$) an odd number of times. Wild behaviour is expected, and is found, near the zeros of r_s (where the layer is an antireflection coating). For the case being considered, s_1 and s_2 are real, and thus r_s can be zero only when Z is real. As discussed in [9], section 1.6.4 and [1], section 2.4, the zeros of r_s occur at $\{Z = +1, s_1 + s_2 = 0\}$ or at $\{Z = -1, s_1 - s_2 = 0\}$. In the first case we have $2q\Delta z$ equal to an even multiple of π, and $q_1 = q_2$, which implies $\varepsilon_1 = \varepsilon_2$, i.e. a substrate optically identical to the medium of incidence. The less restricted second case has $2q\Delta z$ equal to an odd multiple of π and $q_1 q_2 = q^2$. The last equality is possible only if $\varepsilon^2 \leqslant \varepsilon_1\varepsilon_2$, and then holds at the angle of incidence θ_0 given by

$$\sin^2\theta_0 = \frac{\varepsilon_1\varepsilon_2 - \varepsilon^2}{\varepsilon_1(\varepsilon_1 + \varepsilon_2 - 2\varepsilon)}$$
$$\text{or} \qquad \tan^2\theta_0 = \frac{\varepsilon_1\varepsilon_2 - \varepsilon^2}{(\varepsilon - \varepsilon_1)^2}. \tag{6}$$

The corresponding thickness value leading to $r_s = 0$ is given by any odd multiple of $\pi/2q$ evaluated at θ_0, namely by an odd multiple of

$$\tau_0 = \frac{\omega}{c}\Delta z_0 = \frac{\pi}{2}\left\{\frac{\varepsilon_1 + \varepsilon_2 - 2\varepsilon}{(\varepsilon - \varepsilon_1)(\varepsilon_2 - \varepsilon)}\right\}^{1/2}$$
$$\text{or} \qquad \Delta z_0 = \frac{\lambda}{4}\left\{\frac{\varepsilon_1 + \varepsilon_2 - 2\varepsilon}{(\varepsilon - \varepsilon_1)(\varepsilon_2 - \varepsilon)}\right\}^{1/2} \tag{7}$$

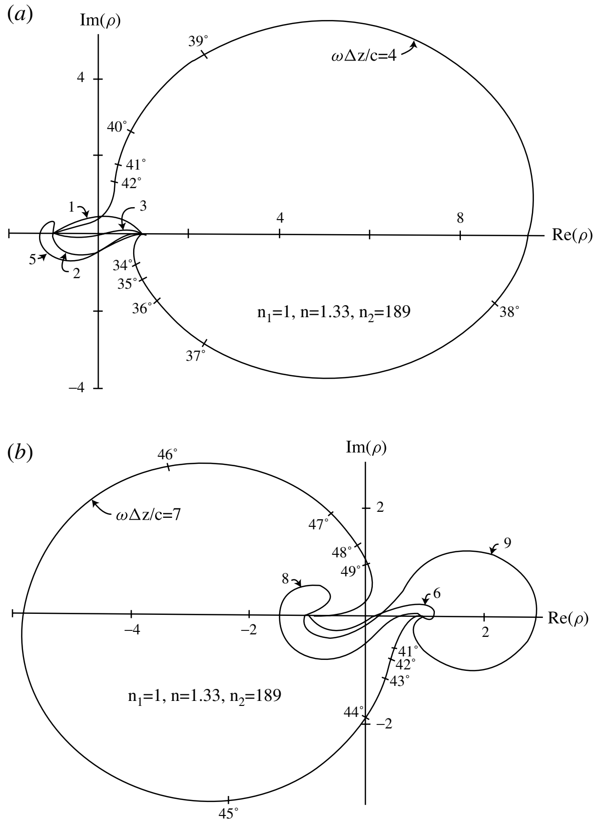

Figure 1. Ellipsometric ratio $\rho = r_p/r_s$ for a layer of water on heaviest flint glass, in air (refractive indices $n_1 = 1.00$, $n = 1.33$ and $n_2 = 1.89$ at 589 nm), drawn for various thicknesses of the water layer. All the trajectories of ρ begin at +1 at normal incidence, and end at -1 at glancing incidence. Since $n^2 < n_1 n_2$, zeros of r_s are possible, and these zeros are responsible for the large excursions of ρ when $(\omega/c)\Delta z$ is near an odd multiple of τ_0. For example, $3\tau_0 \approx 4.07$, hence the behaviour of the $(\omega/c)\Delta z = 4$ trajectory in part (*a*), while $5\tau_0 \approx 6.78$ is close to $(\omega/c)\Delta z = 7$ in part (*b*). Note the rapid excursions near $\theta_0 \approx 40.9^\circ$ at $(\omega/c)\Delta z = 4$ and 7; in this region the angle of incidence is shown with 1° increments. The thickness parameter value $(\omega/c)\Delta z = 7$ also demonstrates the phenomenon of triple principal angles, at 44.0°, 48.9°, and 59.2°. The value $(\omega/c)\Delta z = 4$ does not quite produce triple principal angles: the onset is at 4.057 and the principal angle becomes single-valued again at 4.148. The next range of triple principal angles is between $(\omega/c)\Delta z \approx 6.776$ and 7.123.

where λ is the vacuum wavelength. For a film of water deposited on heaviest flint glass (refractive indices 1.33 and 1.89 at 589 nm), when the dimensionless thickness parameter $\tau = \omega\Delta z/c$ passes through odd multiples of $\tau_0 \approx 1.3567$, r_s will be zero at the angle given by (6), namely at $\theta_0 \approx 40.9^\circ$. Figure 1 shows some trajectories of ρ in the complex plane, as the angle of incidence varies from 0° to 90°. Note how the simple behaviour at small τ values changes drastically as odd multiples of the τ_0 value are approached and exceeded.

More sedate behaviour is shown by layers for which $\varepsilon^2 > \varepsilon_1\varepsilon_2$, as in the ρ trajectories for a layer of water ($n = 1.33$) on glass ($n_2 = 1.50$), illustrated in figure 2. When $\varepsilon^2 > \varepsilon_1\varepsilon_2$, the s wave reflection amplitude cannot be zero, and the curves all lie within a bounded region.

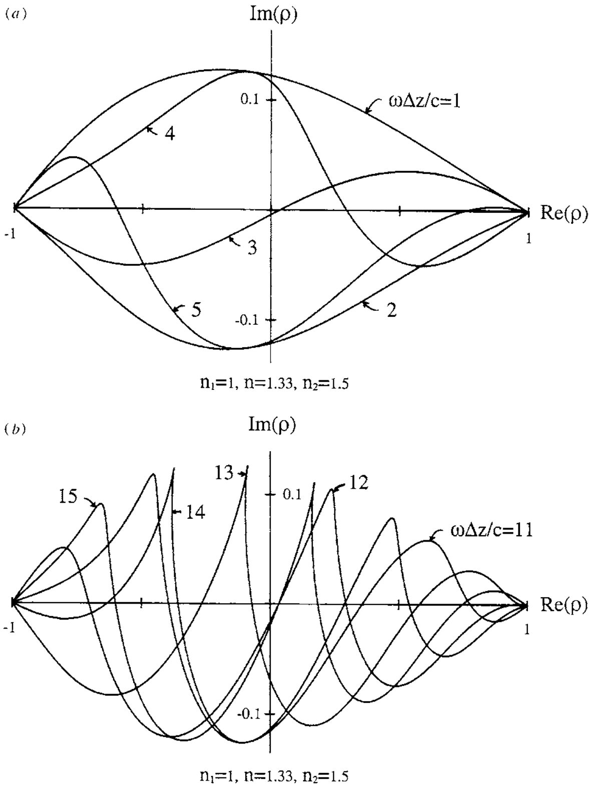

Figure 2. Trajectories of $\rho = r_p/r_s$ as a function of the angle of incidence, at various thicknesses of a water layer on glass ($n_1 = 1.00, n = 1.33, n_2 = 1.50$). In this case $n^2 > n_1 n_2$, zeros of r_s are not possible, and the trajectories are contained within a bounded region. (*a*) Thickness values $(\omega/c)\Delta z$ ranging in unit steps from 1 to 5; (*b*) the values 11–15, again in unit steps. The onset of triple principal angles is near $(\omega/c)\Delta z = 12.7$, when the water layer is about two vacuum wavelengths thick. The first thickness range in which triple principal angles exist is very narrow: 12.699–12.715 in $(\omega/c)\Delta z$, approximately. The next range is a little wider: 15.560–15.657. Thus, no example of triple principal angles appears in this figure.

3. The principal angles

The principal angles are defined by $\mathrm{Re}\,(\rho) = 0$. For the homogeneous layer ρ is given by (2); let us write it as

$$\rho = \frac{p_1 s_1 s_2 + p_2 + s_1 s_2 p_2 Z + p_1 Z^{-1}}{s_1 p_1 p_2 + s_2 + p_1 p_2 s_2 Z + s_1 Z^{-1}} \tag{8}$$

and put $Z = C + \mathrm{i}S$, $Z^{-1} = C - \mathrm{i}S$, where

$$C = \cos 2q\Delta z, \qquad S = \sin 2q\Delta z. \tag{9}$$

Then $\mathrm{Re}\,(\rho)$ is the ratio of two quadratics in C. This result was used in the reduction of the inversion of ellipsometric data to a quintic in ε in [8]. Also, $\mathrm{Im}\,(\rho)$ is S times an expression linear in C, divided by the denominator of $\mathrm{Re}\,(\rho)$. Incidentally, this denominator factors into

$$(1 + 2p_1 p_2 C + p_1^2 p_2^2)(s_1^2 + 2s_1 s_2 C + s_2^2). \tag{10}$$

When the values in equations (3) and (4) are substituted into the expression for $\mathrm{Re}\,(\rho)$, and the result is cleared of fractions, the numerator is a quadratic in C with coefficients which are composed of integral powers of ε_1, ε, ε_2 and $\sin^2\theta_1$. These coefficients are given in the appendix. Let us write the numerator as

$$N = \alpha C^2 + \beta C + \gamma. \tag{11}$$

Then $N = 0$ gives the location of the principal angle or angles. The variation of θ_p with thickness of the layer may be determined in two ways. We can regard $N = 0$ as a transcendental equation for θ_p in terms of $\tau = (\omega/c)\Delta z$ (to be solved numerically), or we can take $N = 0$ to be a quadratic in $C = \cos(2q\Delta z)$, with solutions

$$C_\pm = \frac{-\beta \pm \sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha}. \tag{12}$$

The physical solution must lie in the range $[-1, 1]$; for the examples in figures 3 and 4 the physical solution is C_-. If $\phi = \arccos(C)$, C being the physical solution of (11), then the set $\{\tau_0, \tau_1, \ldots\}$, where

$$\begin{aligned}
\tau_0 &= \frac{\phi}{2\sqrt{\varepsilon - \varepsilon_1\sigma}}, \\
\tau_1 &= \frac{\pi}{\sqrt{\varepsilon - \varepsilon_1\sigma}} - \tau_0, \\
\tau_2 &= \frac{\pi}{\sqrt{\varepsilon - \varepsilon_1\sigma}} + \tau_0, \\
\tau_3 &= \frac{2\pi}{\sqrt{\varepsilon - \varepsilon_1\sigma}} - \tau_0 \ldots
\end{aligned} \tag{13}$$

give the thickness parameter τ as a function of $\sigma = \sin^2\theta_\mathrm{p}$. Thus, by taking σ as plotting variable, $\tau(\sigma)$ is obtained without numerical solution of the transcendental equation $N = 0$.

A particularly simple case obtains when $\varepsilon_1 = \varepsilon_2$, as would be the case for a soap film in air. In that case the principal angle is given by $\sin^2\theta_\mathrm{p} = \varepsilon/(\varepsilon_1 + \varepsilon)$ (this value makes α, β and γ all zero), independent of the film thickness. When ε_1 and ε_2 are close but not equal, the $\sin^2\theta_\mathrm{p}$ curve is flat except near zero and integral multiples of τ_B given by (21), where it drops rapidly to $\sin^2\theta_\mathrm{B} = \varepsilon_2/(\varepsilon_1 + \varepsilon_2)$ (we assume that $\varepsilon_1 \approx \varepsilon_2$ are smaller than ε).

Figures 3 and 4 give the locus of θ_p as the thickness parameter $\tau = \omega\Delta z/c$ varies, for water on heaviest flint glass and on glass, respectively. We see that θ_p oscillates between upper and lower bounds, one of which is the Brewster angle for the bare substrate, given by

$$\sin^2\theta_\mathrm{B} = \frac{\varepsilon_2}{\varepsilon_1 + \varepsilon_2} \qquad \text{or} \qquad \tan^2\theta_\mathrm{B} = \frac{\varepsilon_2}{\varepsilon_1}. \tag{14}$$

The other bound, θ_m, will be discussed in the next section. Here we note only that an indication of which of θ_B and θ_m is the maximum principal angle is given by the general expression of the shift in the principal angle to second-order in $\tau = \omega\Delta z/c$ given in equation (3.53) in [1]. When the values of the integral invariants for the homogeneous layer (given in table 3.1 of [1]) are substituted into this expression, we find

$$\begin{aligned}
\theta_\mathrm{p} - \theta_\mathrm{B} = {} & \frac{(\varepsilon_1\varepsilon_2)^{1/2}\varepsilon_2(\varepsilon - \varepsilon_1)(\varepsilon_2 - \varepsilon)}{2\varepsilon^2(\varepsilon_2^2 - \varepsilon_1^2)^2}[\varepsilon_1\varepsilon_2(\varepsilon_2 - \varepsilon_1) \\
& + (4\varepsilon_1\varepsilon_2 + \varepsilon_1^2 - \varepsilon_2^2)\varepsilon - (3\varepsilon_1 + \varepsilon_2)\varepsilon^2]\tau^2 + \mathrm{O}(\tau^4).
\end{aligned}$$

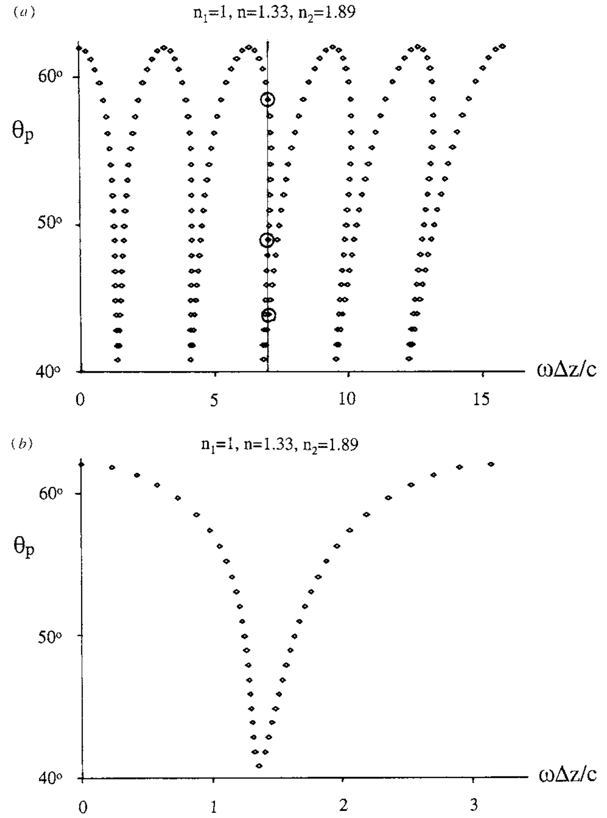

Figure 3. Variation of the principal angle θ_p (at which the real part of the ellipsometric ratio r_p/r_s passes through zero), with the thickness of a layer of water, on heaviest flint glass substrate (the refractive indices are as in figure 1). (*a*) The variation of θ_p with the dimensionless thickness parameter, oscillating between the upper bound $\theta_\mathrm{B} \approx 62.1°$ and the lower bound $\theta_\mathrm{m} \approx 40.9°$. Note the triple principal angles at $(\omega/c)\Delta z = 7$ (for a water layer about 1.1 vacuum wavelengths thick), previously shown on the trajectory of r_p/r_s in figure 1(*b*). (*b*) The variation of θ_p in the fundamental interval $[0, \tau_\mathrm{B}]$.

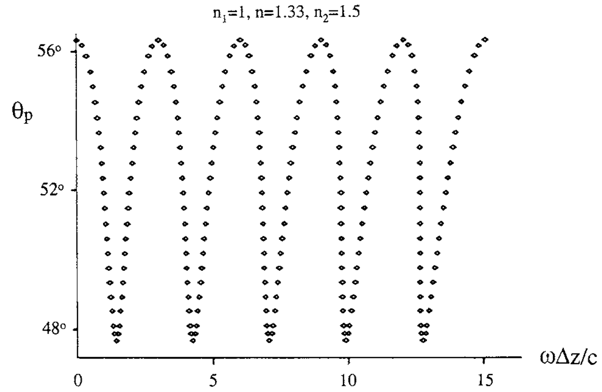

Figure 4. Principal angle θ_p as a function of the thickness Δz of a water layer on glass ($n_1 = 1.00$, $n = 1.33$, $n_2 = 1.50$, as in figure 2). The maxima are at the Brewster angle $\theta_\mathrm{B} = \arctan(n_2/n_1) \approx 56.3°$, the minima at $\theta_\mathrm{m} \approx 47.7°$. The upper period in $(\omega/c)\Delta z$ is $\pi(\varepsilon - \varepsilon_1 \sin^2\theta_\mathrm{B})^{-1/2} \approx 3.03$, the lower period is $\pi(\varepsilon - \varepsilon_1 \sin^2\theta_\mathrm{m})^{-1/2} \approx 2.84$. The onset of triple principal angles is near $(\omega/c)\Delta z = 12.7$.

This expression is negative when $0 < \varepsilon < \varepsilon_2$ except for a small region between $\varepsilon = \varepsilon_1$ and

$$\varepsilon = (\varepsilon_1^2 + 4\varepsilon_1\varepsilon_2 - \varepsilon_2^2 + (\varepsilon_1^4 - 4\varepsilon_1^3\varepsilon_2 + 22\varepsilon_1^2\varepsilon_2^2 - 4\varepsilon_1\varepsilon_2^3 + \varepsilon_2^4)^{1/2})\{2(3\varepsilon_1 + \varepsilon_2)\}^{-1}.$$

It is positive for $\varepsilon > \varepsilon_2$. Thus, we can generally expect θ_B to be the maximum value of θ_p when $\varepsilon < \varepsilon_2$, and to be the minimum for $\varepsilon > \varepsilon_2$. More detail is given in the next section.

We note that the thickness Δz enters into the equation $N = 0$ determining the principal angle θ_p through $\cos(2q\Delta z)$. Thus at given θ_p there is periodicity in the θ_p curves, with period $\Delta z = \pi/q$. The period in the thickness parameter $\tau = \omega\Delta z/c$ is correspondingly $\pi/(\varepsilon - \varepsilon_1 \sin^2\theta_\mathrm{p})^{1/2}$, which varies between $\pi/(\varepsilon - \varepsilon_1 \sin^2\theta_\mathrm{B})^{1/2}$ and $\pi/(\varepsilon - \varepsilon_1 \sin^2\theta_\mathrm{m})^{1/2}$. Thus the upper part of the locus of θ_p has a longer period than the lower part, and eventually the lower part will lag behind sufficiently to produce triple principal angles at a given thickness. As the thickness is increased still further, quintuple principal angles will appear, and so on. This argument shows that multiple principal angles will always be found for sufficiently large layer thickness, except in the case of $\varepsilon_1 = \varepsilon_2$ for which θ_p is constant, as discussed above. An estimate of how thick the layer has to be for triple principal angles to appear is given in section 5.

4. Location of the maxima and minima of θ_p

The locus of the principal angle θ_p is given by the zero of $N = \alpha C^2 + \beta C + \gamma = 0$. As a function of the thickness parameter $\tau = \omega\Delta z/c$, the extrema of θ_p occur where $\partial N/\partial\tau = 0$, i.e. where

$$(2\alpha C + \beta)\frac{\partial C}{\partial \tau} = 0. \qquad (15)$$

(This is because the extrema occur at $\mathrm{d}\sigma/\mathrm{d}\tau = 0$, where $\sigma = \sin^2\theta_\mathrm{p}$ is regarded as a function of $\tau = (\omega/c)\Delta z$. Now, from $N = 0$ we have

$$\mathrm{d}N = \frac{\partial N}{\partial \tau}\,\mathrm{d}\tau + \frac{\partial N}{\partial \sigma}\,\mathrm{d}\sigma = 0 \qquad (16)$$

and thus $\mathrm{d}\sigma/\mathrm{d}\tau = -(\partial N/\partial\tau)/(\partial N/\partial\sigma)$. Since $\partial N/\partial\sigma$ cannot be infinite, the zeros of $\mathrm{d}\sigma/\mathrm{d}\tau$ will be given by the zeros of $\partial N/\partial\tau$.)

When the physical root is in $\partial C/\partial\tau = 0$, $S = \sin 2q\Delta z$ is zero, as is the case for extrema of the reflectances R_p and R_s ([9, section 1.6.4], [1, section 2.4]), which occur when

$$2q\Delta z = 2m\pi \qquad \text{or} \qquad 2q\Delta z = (2m+1)\pi$$

where m is an integer. These correspond to

$$\begin{aligned} & C = 1, \qquad \alpha + \beta + \gamma = 0 \\ \text{or} \quad & C = -1, \qquad \alpha - \beta + \gamma = 0. \end{aligned} \qquad (17)$$

The first of these options, namely $\alpha + \beta + \gamma = 0$, factors to

$$4\varepsilon^2(\varepsilon_1 - \varepsilon_2)^2(\varepsilon - \varepsilon_1 \sin^2\theta_1)^2[\varepsilon_2 - (\varepsilon_1 + \varepsilon_2)\sin^2\theta_1] = 0, \qquad (18)$$

from which we extract the physical root which is the Brewster angle θ_B as already given in equation (14). The second option,

$\alpha - \beta + \gamma = 0$, factors to $4F_1F_2$, where F_1 and F_2 are respectively linear and quadratic in $\sin^2\theta_1$. $F_1 = 0$ gives

$$\sin^2\theta_m = \frac{\varepsilon_1\varepsilon_2 - \varepsilon^2}{\varepsilon_1(\varepsilon_1 + \varepsilon_2 - 2\varepsilon)} \quad \text{or} \quad \tan^2\theta_m = \frac{\varepsilon_1\varepsilon_2 - \varepsilon^2}{(\varepsilon - \varepsilon_1)^2} \tag{19}$$

which we recognize from (6) as the angle θ_0 at which $r_s = 0$, provided $\varepsilon^2 \leqslant \varepsilon_1\varepsilon_2$. When $\varepsilon^2 \geqslant \varepsilon_1\varepsilon_2$ the physical root θ_m lies in the quadratic $F_2 = 0$, which reads

$$\varepsilon_1(\varepsilon^4 - \varepsilon_1^2\varepsilon_2^2)\sin^4\theta_m + \varepsilon[2\varepsilon_1^2\varepsilon_2^2 - (\varepsilon_1 + \varepsilon_2)\varepsilon^3]\sin^2\theta_m + (\varepsilon^2 - \varepsilon_1\varepsilon_2)\varepsilon^2\varepsilon_2 = 0. \tag{20}$$

The cross-over is at $\varepsilon^2 = \varepsilon_1\varepsilon_2$; at this point both F_1 and F_2 give $\theta_m = 0$, so the principal angle oscillates from zero to θ_B as the layer thickness varies.

When $\varepsilon^2 < \varepsilon_1\varepsilon_2$ it is possible for θ_B and θ_m, given by (14) and (19), to coincide, at $\varepsilon = 2\varepsilon_1\varepsilon_2/(\varepsilon_1 + \varepsilon_2)$. Rather large values of substrate refractive index are required for this to happen: for a water layer, the substrate index would have to be about 2.7666 when $n_1 = 1$. (A near example is $n_2 = 2.8$ in figure 5.) When $\varepsilon^2 > \varepsilon_1\varepsilon_2$ it is not possible for θ_m to be equal to θ_B, since substitution of (14) into (20) leads to $\varepsilon = \varepsilon_1$ or $\varepsilon = \varepsilon_2$ (in these cases the layer has no optical effect), or to values of ε inconsistent with $\varepsilon^2 > \varepsilon_1\varepsilon_2$.

Figure 5 shows a set of curves of $\sigma = \sin^2\theta_p$ versus $\tau = (\omega/c)\Delta z$, for $n_1 = 1$, $n = 1.33$ (water) and variable substrate index n_2. We see that for low substrate index θ_m gives the minimum value of θ_p, which drops to zero when $n_2 = n^2/n_1 \approx 1.77$. For larger substrate index, θ_m increases, and eventually θ_m gives a maximum value, with subsidiary minima on either side. The transition from minimum to maximum occurs when $\mathrm{d}\sigma/\mathrm{d}\tau$ and $\mathrm{d}^2\sigma/\mathrm{d}\tau^2$ are both zero. We saw from (16) that $\frac{\mathrm{d}\sigma}{\mathrm{d}\tau} = -\frac{\partial N/\partial\tau}{\partial N/\partial\sigma}$.

We differentiate with respect to τ again, using

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{\partial N}{\partial\tau}\right) = \frac{\partial^2 N}{\partial\tau^2} + \frac{\partial^2 N}{\partial\tau\partial\sigma}\frac{\mathrm{d}\sigma}{\mathrm{d}\tau}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{\partial N}{\partial\sigma}\right) = \frac{\partial^2 N}{\partial\tau\partial\sigma} + \frac{\partial^2 N}{\partial\sigma^2}\frac{\mathrm{d}\sigma}{\mathrm{d}\tau}$$

to obtain

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}\tau^2} = \frac{2(\frac{\partial N}{\partial\tau})(\frac{\partial N}{\partial\sigma})\frac{\partial^2 N}{\partial\tau\partial\sigma} - (\frac{\partial N}{\partial\tau})^2\frac{\partial^2 N}{\partial\sigma^2} - (\frac{\partial N}{\partial\sigma})^2\frac{\partial^2 N}{\partial\tau^2}}{(\frac{\partial N}{\partial\sigma})^3}.$$

Since $\partial N/\partial\sigma$ cannot be infinite, $\mathrm{d}\sigma/\mathrm{d}\tau$ is zero when $\partial N/\partial\tau$ is zero, as noted before. Thus $\mathrm{d}\sigma/\mathrm{d}\tau$ and $\mathrm{d}^2\sigma/\mathrm{d}\tau^2$ are both zero when $\partial N/\partial\tau$ and $\partial^2 N/\partial\tau^2$ are both zero. These imply $\{2\alpha - \beta = 0, C = -1\}$, which in turn are equivalent to $\{\alpha - \gamma = 0, \alpha - \beta + \gamma = 0\}$. The second of these was examined above, leading to (19) or (20) depending on whether ε^2 is less or greater than $\varepsilon_1\varepsilon_2$. For the cases shown in figure 5 for which $n_2 > (1.33)^2 = 1.7689$, $\sin^2\theta_m$ is given by (19), and substitution into $\alpha = \gamma$ gives

$$(3\varepsilon_1 + \varepsilon_2)\varepsilon^4 - 4\varepsilon_1\varepsilon_2\varepsilon^3 + 4\varepsilon_1\varepsilon_2^2\varepsilon^2 - 4(\varepsilon_1\varepsilon_2)^2\varepsilon - (\varepsilon_1\varepsilon_2)^2(\varepsilon_2 - \varepsilon_1) = 0$$

which is quartic in ε and cubic in ε_2. For $\varepsilon_1 = 1$, $\varepsilon = (1.33)^2$, this gives $\varepsilon_2 \approx 5.150$ or $n_2 \approx 2.269$. Above this value of n_2, θ_m becomes a local maximum, and eventually an absolute maximum. Two subsidiary minima appear on either side of θ_m, being given by the first factor in (15), $2\alpha C + \beta = 0$. Together with $N = \alpha C^2 + \beta C + \gamma = 0$ this implies $\beta^2 - 4\alpha\gamma = 0$, which is a sextic in $\sigma = \sin^2\theta_p$, one root of which gives the observed subsidiary minima.

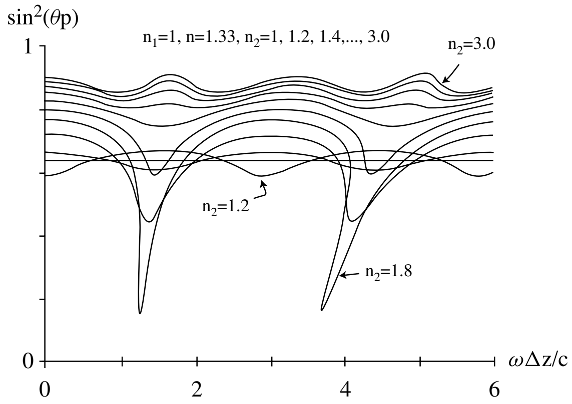

Figure 5. A set of curves of $\sin^2\theta_p$ versus $\tau = (\omega/c)\Delta z$, for variable substrate index n_2, which increases from 1.0 to 3.0 in steps of 0.2. The curves are drawn for $n_1 = 1$ and $n = 1.33$ (water). Note the deep minima near $n_2 = n^2/n_1 = 1.7689$, for which value of n_2 the minima would be at zero angle of incidence. Note also that the minima transform to local maxima as n_2 increases, as discussed in the text. At $n_2 = n_1$ the principal angle is constant at $\sin^2\theta_p = \varepsilon/(\varepsilon_1 + \varepsilon) \approx 0.639$. For $n_2 \neq n_1$ the zero-thickness principal angle takes the Brewster value $\sin^2\theta_B = \varepsilon_2/(\varepsilon_1 + \varepsilon_2)$.

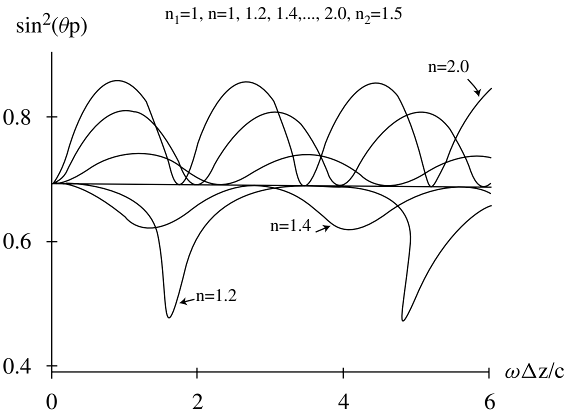

Figure 6. Variation of $\sin^2\theta_p$ with layer thickness, for fixed substrate index $n_2 = 1.5$, and layer index n increasing from 1.0 to 2.0 in 0.2 steps. The Brewster angle is at $\sin^2\theta_B = \varepsilon_2/(\varepsilon_1 + \varepsilon_2) \approx 0.6923$; for $n = 1.2$ and 1.4 the curves have maxima at this value, for $n > 1.5$ they have minima at this value. When $n = n_1$ the layer has no optical effect, and $\sin^2\theta_p$ is constant at the substrate Brewster value (horizontal line).

Figure 6 illustrates the behaviour of $\sin^2\theta_p$ with layer thickness for variable layer refractive index. When n is less than the substrate index n_2 (here 1.5), the $\sin^2\theta_p$ curves have minima at θ_m and maxima at θ_B. For $n > n_2$ the maxima are at θ_m and the minima at θ_B.

We now return to the location of the main maxima and minima of θ_p. One of these is at θ_B, and occurs at zero thickness and then at points when $q_B\Delta z$ is an integral multiple of π (see equation 16), i.e. when $\tau = \omega\Delta z/c$ is an integral multiple of

$$\tau_B = \frac{\omega}{c}\Delta z_B = \frac{\pi}{\sqrt{\varepsilon - \frac{\varepsilon_1\varepsilon_2}{\varepsilon_1+\varepsilon_2}}}. \quad (21)$$

The location of the other extrema depends on whether the layer dielectric function is smaller or greater than the geometric mean of the dielectric functions of the bounding media. When $\varepsilon^2 \leqslant \varepsilon_1\varepsilon_2$, the main extrema occur at odd multiples of $(\omega/c)\Delta z_0$ given in (7) at θ_m given by (19). When $\varepsilon^2 \geqslant \varepsilon_1\varepsilon_2$ the main extrema occur at odd multiples of

$$\tau_m = \frac{\omega}{c}\Delta z_m = \frac{\pi/2}{\sqrt{\varepsilon - \varepsilon_1 \sin^2\theta_m}} \quad (22)$$

where $\sin^2\theta_m$ is the physical solution of equation (20). We note that this equation can be transformed into a quadratic for $\tan^2\theta_m$:

$$\varepsilon_1\varepsilon_2^2(\varepsilon-\varepsilon_1)^2\tan^4\theta_m - \varepsilon[\varepsilon^3(\varepsilon_2-\varepsilon_1) - 2\varepsilon_1\varepsilon_2^2(\varepsilon-\varepsilon_1)]\tan^2\theta_m - \varepsilon^2\varepsilon_2(\varepsilon^2-\varepsilon_1\varepsilon_2) = 0. \quad (23)$$

Write this as $aT^2+bT+c=0$. We see that $a > 0$ and $c \leqslant 0$ (since $\varepsilon^2 \geqslant \varepsilon_1\varepsilon_2$). Thus the discriminant b^2-4ac is positive, and the roots are real. Also the product of the roots is c/a, which is negative, so one root will be positive and the other negative. Thus, there is one and only one possible physical root.

We note finally that at the cross-over point when ε is equal to the geometric mean of ε_1 and ε_2, $\sin^2\theta_m$ is zero, and the non-Brewster extrema occur at odd multiples of

$$\frac{\omega}{c}\Delta z_m = \frac{\pi}{2}(\varepsilon_1\varepsilon_2)^{-1/4}. \quad (24)$$

The period is twice this value, and is smaller than the Brewster period at $\varepsilon = (\varepsilon_1\varepsilon_2)^{1/2}$, which from (21) is

$$\frac{\omega}{c}\Delta z_B = \pi(\varepsilon_1\varepsilon_2)^{-1/4}\left[1 - \frac{(\varepsilon_1\varepsilon_2)^{1/2}}{\varepsilon_1+\varepsilon_2}\right]^{-1/2}. \quad (25)$$

5. Summary and discussion

We have derived formulae for the location of the extrema in the variation of the principal angle θ_p with thickness. One of these is always the Brewster angle. It is interesting that the physical root giving the other main extremum θ_m comes from different factors, depending on whether the layer refractive index is smaller or greater than the geometric mean of the indices of the bounding media. In the former case an angle exists at which r_s goes to zero, so $\rho = r_p/r_s$ is unbounded. In the latter case r_s cannot be zero, and ρ lies within a bounded region of the complex plane. Thus the different mathematical roots correspond to very different physical situations.

The simple qualitative argument at the end of section 3 has shown that multiple principal angles (in general, an odd number of θ_p values for given layer thickness) will always appear if the layer is thick enough. Just how thick can be estimated from the formulae of the previous section: let τ again represent the thickness parameter $\omega\Delta z/c$; then extrema occur at zero and integral multiples of τ_B, and at odd multiples of τ_m, with τ_B and τ_m given by (21) and (22). Let us assume that $\tau_B > 2\tau_m$, for example. Then, when $\ell\tau_B > (2\ell+1)\tau_m$, the maximum at θ_B lies to the right of the minimum θ_m in the θ_p versus τ diagram, so triple principal angles must necessarily occur when the number ℓ of the periods τ_B exceeds $\tau_m/(\tau_B - 2\tau_m)$. In fact, triple principal angles first appear at about half this number of periods, i.e. at a thickness given by

$$\tau = \frac{\omega}{c}\Delta z \approx \frac{1}{2}\frac{\tau_B\tau_m}{\tau_B - 2\tau_m} = \frac{\pi/4}{\sqrt{\varepsilon - \varepsilon_1\sin^2\theta_m} - \sqrt{\varepsilon - \frac{\varepsilon_1\varepsilon_2}{\varepsilon_1+\varepsilon_2}}}. \quad (26)$$

This estimate gives τ values of 4.8 for the example in figure 3(*a*) (actual value about 4.0) and 11.5 for the example in figure 3(*b*) (actual value about 12.7). Equation (26) provides an approximate lower bound for the thickness when triple principal angles are first observed. The range over which triple principal angles appear is quite narrow at first (see figures 1–4), but as the layer thickness increases eventually all thicknesses will have triple principal angles, then quintuple angles will appear, and so on.

The precise thickness at which there is onset into (and exit from) a region of triple principal angles is where $N = \alpha C^2 + \beta C + \gamma$ and $\partial N/\partial\theta_1$ are zero together. (At this point the θ_p versus τ curve is vertical.) The derivative $\partial N/\partial\theta_1$ is linear in $S = \sin 2q\Delta z$, and, by squaring, S and C can be eliminated between the two equations, leaving a quadratic in $[(\omega/c)^2\Delta z/q]^2$, to be solved simultaneously with $N = 0$. I have not been able to extract an analytic expression for the minimum value of $\omega\Delta z/c$ required for multiple principal angles by this method. However, for a given set of refractive indices, the principal angle as a function of the thickness is easily generated from (13), without solution of the transcendental equation $N = 0$.

At a given principal angle θ_p, the thickness dependence is in $\cos 2q\Delta z = \cos(2\frac{\omega}{c}\Delta z\sqrt{\varepsilon - \varepsilon_1\sin^2\theta_p})$. The period in $\tau = \omega\Delta z/c$ is therefore (at given θ_p)

$$\Psi = \frac{\pi}{(\varepsilon - \varepsilon_1\sin^2\theta_p)^{1/2}}. \quad (27)$$

Thus the principal angle need only be calculated in the fundamental interval $[0, \tau_B]$ as given by (21): for greater values of τ we can use

$$\theta_p(\tau + \Psi) = \theta_p(\tau). \quad (28)$$

The period itself satisfies the functional relation

$$\Psi(\tau + \Psi(\tau)) = \Psi(\tau) \quad (29)$$

which confirms that Ψ is independent of τ (as a period must be), since a constant is the only possible solution of (29).

All of the above has been for a non-absorbing layer on a non-absorbing substrate. When either or both are absorbing the analysis is more complicated. In particular, when the layer is absorbing, $Z = \exp(2iq\Delta z)$ no longer lies on the unit circle (except at zero film thickness) and the periodicity with Δz at fixed angle of incidence is lost. The general expression for r_p/r_s given in [1], equation (3.52), to second order in the layer thickness. The shift in the principal angle from the zero-thickness Brewster angle θ_B (for a non-absorbing substrate) now becomes first-order in the layer thickness, in contrast to the transparent-layer case of equation (13) above. The general expression for the principal angle shift is ([1, equation (8.74)])

$$\theta_p - \theta_B = \frac{(\frac{\varepsilon_2}{\varepsilon_1})^{1/2}\frac{\omega}{c}\mathrm{Im}\,(I_1)}{(\varepsilon_1+\varepsilon_2)^{1/2}(\frac{\varepsilon_1}{\varepsilon_2}-\frac{\varepsilon_2}{\varepsilon_1})} + \mathrm{O}\left(\frac{\omega}{c}\Delta z\right)^2 \qquad (30)$$

where I_1 is the integral of $(\varepsilon-\varepsilon_1)(\varepsilon_2-\varepsilon)/\varepsilon$ over the layer, and takes the value

$$I_1 \to \frac{(\varepsilon-\varepsilon_1)(\varepsilon_2-\varepsilon)}{\varepsilon}\Delta z = \left(\varepsilon_1+\varepsilon_2-\frac{\varepsilon_1\varepsilon_2}{\varepsilon}-\varepsilon\right)\Delta z \qquad (31)$$

for a homogeneous layer.

Finally, we note that for anisotropic and/or chiral (optically active) media, reflection ellipsometry measures either

$$\rho_P = (r_{pp}+r_{sp}\tan P)/(r_{ps}+r_{ss}\tan P)$$
$$\text{or} \qquad \rho_A = (r_{pp}+r_{ps}\tan A)/(r_{sp}+r_{ss}\tan A) \qquad (32)$$

where P and A are the polarizer and analyser angles measured from the $\boldsymbol{p}$ direction [10]. (The reflection amplitude r_{sp}, for example, gives the complex amplitude of the reflected electric field s component, when unit electric field is incident, aligned along the $\boldsymbol{p}$ direction.) One can thus define two principal angles by $\mathrm{Re}\,(\rho_P) = 0$, or by $\mathrm{Re}\,(\rho_A) = 0$, both depending on the respective angles P and A, as well as on all four reflection amplitudes.

Appendix A. Values of α, β and γ

The coefficients α, β and γ in $N = \alpha C^2 + \beta C + \gamma$ of (11) are all cubic in $\sin^2\theta_1$ and in ε_2, and sextic in ε. The degrees of α, β and γ in ε_1 are 4, 5 and 5. The coefficients are (with σ standing for $\sin^2\theta_1$)

$$\begin{aligned}\alpha = {} & (\varepsilon-\varepsilon_1)^2(\varepsilon-\varepsilon_2)^2[2\varepsilon_1\sigma^2 \\ & -(\varepsilon_1+\varepsilon)\sigma+\varepsilon][\varepsilon_1(\varepsilon+\varepsilon_2)\sigma-\varepsilon\varepsilon_2] \end{aligned} \qquad (A.1)$$

$$\begin{aligned}\beta = {} & -2(\varepsilon-\varepsilon_1)(\varepsilon-\varepsilon_2)\{\varepsilon_1^2[2\varepsilon^3 \\ & +(\varepsilon_1+\varepsilon_2)\varepsilon^2+(\varepsilon_1^2\varepsilon_2^2)\varepsilon+\varepsilon_1\varepsilon_2(\varepsilon_1+\varepsilon_2)]\sigma^3 \\ & -\varepsilon_1[\varepsilon^4+3(\varepsilon_1+\varepsilon_2)\varepsilon^3+2(\varepsilon_1^2+\varepsilon_1\varepsilon_2+\varepsilon_2^2)\varepsilon^2 \\ & +3\varepsilon_1\varepsilon_2(\varepsilon_1+\varepsilon_2)\varepsilon+(\varepsilon_1\varepsilon_2)^2]\sigma^2 \\ & +\varepsilon(\varepsilon+\varepsilon_1)(\varepsilon+\varepsilon_2)[\varepsilon(\varepsilon_1+\varepsilon_2)+2\varepsilon_1\varepsilon_2]\sigma \\ & -\varepsilon^2\varepsilon_2(\varepsilon+\varepsilon_1)(\varepsilon+\varepsilon_2)\} \end{aligned} \qquad (A.2)$$

$$\begin{aligned}\gamma = {} & 2\varepsilon_1^2[\varepsilon^5+\varepsilon_1\varepsilon^4-(\varepsilon_1^2+2\varepsilon_1\varepsilon_2-\varepsilon_2^2)\varepsilon^3 \\ & -\varepsilon_1(3\varepsilon_2^2-\varepsilon_1^2)\varepsilon^2+\varepsilon_1\varepsilon_2^2(2\varepsilon_2-\varepsilon_1)\varepsilon+\varepsilon_1^3\varepsilon_2^2]\sigma^3 \\ & -\varepsilon_1[\varepsilon^6+(5\varepsilon_1+3\varepsilon_2)\varepsilon^5-(\varepsilon_1^2+3\varepsilon_1\varepsilon_2-3\varepsilon_2^2)\varepsilon^4 \\ & +(3\varepsilon_1^3-7\varepsilon_1^2\varepsilon_2-13\varepsilon_1\varepsilon_2^2+\varepsilon_2^3)\varepsilon^3 \\ & +\varepsilon_1\varepsilon_2(3\varepsilon_1^2-11\varepsilon_1\varepsilon_2+7\varepsilon_2^2)\varepsilon^2 \\ & +(\varepsilon_1\varepsilon_2)^2(5\varepsilon_1+3\varepsilon_2)\varepsilon+(\varepsilon_1\varepsilon_2)^3]\sigma^2 \\ & +\varepsilon[(\varepsilon_1+\varepsilon_2)\varepsilon^5+2(\varepsilon_1^2+3\varepsilon_1\varepsilon_2+\varepsilon_2^2)\varepsilon^4 \\ & +(\varepsilon_1+\varepsilon_2)(\varepsilon_1^2-8\varepsilon_1\varepsilon_2+\varepsilon_2^2)\varepsilon^3 \\ & +4\varepsilon_1\varepsilon_2(\varepsilon_1^2-5\varepsilon_1\varepsilon_2+\varepsilon_2^2)\varepsilon^2 \\ & +5(\varepsilon_1\varepsilon_2)^2(\varepsilon_1+\varepsilon_2)\varepsilon+2(\varepsilon_1\varepsilon_2)^3]\sigma \\ & -\varepsilon^2\varepsilon_2[\varepsilon^4+2(\varepsilon_1+\varepsilon_2)\varepsilon^3+(\varepsilon_1^2-12\varepsilon_1\varepsilon_2+\varepsilon_2^2)\varepsilon^2 \\ & +2\varepsilon_1\varepsilon_2(\varepsilon_1+\varepsilon_2)\varepsilon+(\varepsilon_1+\varepsilon_2)^2]. \end{aligned} \qquad (A.3)$$

Test values: at $\varepsilon_1 = 1$, $\varepsilon_2 = 2$, $\varepsilon = 3$ and $\sigma = \frac{1}{2}$, the coefficients α, β and γ become -21, 337 and $-428\frac{1}{2}$ respectively.

References

[1] Lekner J 1987 *Theory of Reflection* (Dordrecht: Kluwer)
[2] Azzam R M A and Bashara N M 1977 *Ellipsometry and Polarized Light* (Amsterdam: North-Holland)
[3] Aspnes D E 1976 Spectroscopic ellipsometry of solids *Optical Properties of Solids: New Developments* ed B O Seraphin (Amsterdam: North-Holland) chapter 15
[4] Beaglehole D 1980 Ellipsometric study of the surface of simple liquids *Physica* B **100** 163–74
[5] Law B M and Beaglehole D 1981 Model calculations of the ellipsometric properties of inhomogeneous dielectric surfaces *J. Phys. D: Appl. Phys.* **14** 115–26
[6] Evans S D, Allison H, Boden N and Henderson J R 1996 Surface-field induced organization at solid/fluid interfaces *Faraday Discuss.* **104** 37–48
[7] Azzam R M A 1981 Contours of constant principal angle and constant principal azimuth in the complex ε plane *J. Opt. Soc. Am.* **71** 1523–8
[8] Lekner J 1994 Inversion of reflection ellipsometric data *Appl. Opt.* **33** 5159–65
[9] Born M and Wolf E 1965 *Principles of Optics* 3rd edn (Oxford: Pergamon)
[10] Lekner J 1993 Ellipsometry of anisotropic media *J. Opt. Soc. Am.* A **10** 1579–81