Applicability of the Rayleigh hypothesis to real materials

Thomas C. Paulick
245 Sheridan Avenue South, Minneapolis, Minnesota 55405
(Received 31 January 1989; revised manuscript received 1 June 1989)

The validity of Rayleigh's hypothesis with permeable media is investigated. For the two-dimensional transmission problem, the extinction theorem and boundary integral equations are obtained within a common framework, for the surface with periodic corrugations. The surface field functions are found from the integral equations; these functions can be analytically continued to complex values of the coordinates. Then the extinction theorem provides a criterion for the validity of Rayleigh's hypothesis. We find that it is valid for all materials, absorbing or otherwise, when $2\pi h/D < 0.448$. Here h is the depth and D the period of the sinusoidal surface. Criteria of this nature have previously been established only for the Dirichlet problem, with the assumption that the medium of incidence was also nondissipative. Accurate application of Rayleigh's method requires accurate knowledge of both the interface geometry and the material properties. In particular, when "perfect-conductor" formulas are used to measure the dielectric constant and calibrate the corrugation depth of gold gratings, significant errors are introduced into the data. Such distortions may be the true cause of discrepancies observed when Rayleigh-type calculations are applied to light scattering from real metals with rough surfaces.

I. INTRODUCTION

A. The Rayleigh method and Rayleigh hypothesis

Lord Rayleigh was the first, in 1895, to use Fourier series to study wave-scattering processes at rough surfaces.[1] At that time, he also introduced the so-called "Rayleigh hypothesis," which states that the far-field expansions could be extended all the way back to the surface itself in order to satisfy Maxwell boundary conditions. The validity of this supposition was admittedly unknown, but at the time it was the most powerful theoretical tool available for attacking such problems.

Despite the subsequent development of exact differential and integral methods,[2–6] the Rayleigh hypothesis is actually being used more frequently now than in earlier years.[7–11] Part of this current interest is due to the relative simplicity[12] of the Rayleigh method compared with, for instance, the widely used extinction theorem. The original convenience was further enhanced by Toigo *et al.*,[3] who "reduced" the Rayleigh matrix equations to allow reflected and refracted Bragg waves to be calculated separately. From these reduced equations Maradudin[13] has developed an accurate and useful perturbation theory. Greffet[14] has very recently generalized Maradudin's procedure, and also introduced a hybrid method combining the Rayleigh hypothesis with the extinction theorem.

This ongoing activity leads us to consider the conditions under which the Rayleigh hypothesis is valid. Petit and Cadilhac[15] showed that it is *not* valid for Dirichlet boundary data on pure sinusoidal corrugations whenever

$$\Gamma \equiv \kappa h \geq 0.448 , \tag{1.1}$$

where

$$\kappa \equiv \frac{2\pi}{D} \tag{1.2}$$

is the "reciprocal-lattice vector" of the grating whose profile is described by

$$z = h\cos(\kappa x) \equiv \zeta(x) \tag{1.3}$$

as in Fig. 1. In a difficult and very instructive series of papers,[16–20] Millar showed, again for the Dirichlet problem in two dimensions, how to locate the singularities of a local Fourier or Fourier-Bessel series. In the last of these papers, he asked the following: for what corrugation dimensions do all the singularities of the Helmholtz equation lie below the line $z = -h$? In that case, the Fourier analysis of the scattered field Φ_s into plane waves, which constitutes the Rayleigh series, is valid everywhere in $z \geq -h$; the Rayleigh hypothesis is thus confirmed. He found that for this geometry, Rayleigh's hypothesis does hold when $\Gamma < 0.448$.

Millar's result was later obtained for more general surface textures by Hill and Celli[21] using a completely different approach. These authors began with the extinction theorem, which provides an integral expression for the Fourier expansion coefficients of the reflected field, valid at all points $z > h$ above the highest point of the surface described by Eq. (1.3). This expansion is again the Rayleigh series. Hill and Celli then asked under what conditions the series converges at *all* points (even within the grooves) outside of the perfect conductor. This convergence is the Rayleigh hypothesis. They recovered Millar's result in a way that is applicable to a large class of surface profiles in two or three dimensions.

The foregoing research provides the reason and also the tools for the present work. Hill and Celli had to assume certain analytic properties of the boundary data for the electromagnetic fields, which are not at all guaranteed

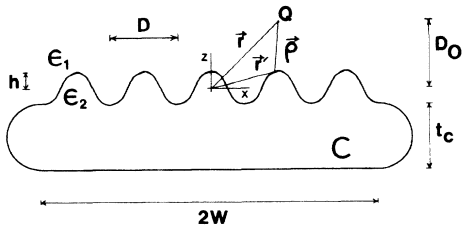

FIG. 1. Geometry and notation for the scattering object.

to hold. Accordingly, they offered their readers only "a high degree of confidence" that the Rayleigh hypothesis is valid under the condition anticipated by Millar. But it is just these analytic properties that were established by Millar's labor, and which his earlier papers showed *how* to establish. More importantly, Hill and Celli considered only the Dirichlet problem (as did Millar), stating that the extension to Neumann or impedance boundary conditions is straightforward.[22] These "textbook" problems can provide only idealized models of reality, in that light of s-wave (p-wave) polarization satisfies Dirichlet (Neumann) boundary data for **E** (**H**) at a perfect conductor. They are also idealized in the mathematical sense because they involve prescribed boundary data: in reality, the surface fields and their normal derivatives are both nonzero and unknown.

The present work considers the conditions under which the Rayleigh hypothesis is valid for a corrugated interface between two permeable media. This question has not been previously addressed. The analyticity of the surface field functions is studied with Millar's methods; these functions are then used in the Hill-Celli procedure (cited from now on as HC) to validate Rayleigh's hypothesis. Implicitly, we thereby render the latter work a true proof of Millar's criterion in the special case of the Dirichlet problem. The new result is that the Rayleigh hypothesis holds for real materials, under the same conditions established for the Dirichlet problem by these earlier authors.

Several authors[12,14] have found that in practice the Rayleigh hypothesis sometimes gives good results even in the regime of inequality (1.1), where it cannot be justified mathematically. Reference 3 showed that the Rayleigh series can be viewed as an expansion in a complete but nonorthogonal set of basis functions. So while the expansion itself is always legitimate, stability of the coefficients is not guaranteed when the series is truncated for computation. Outside of the regime in which convergence is theoretically certain, the series may still be useful in the same sense that a nonconvergent asymptotic series is useful. We note that Glass and Maradudin[7] reported instances in which certain calculated quantities diverged when the truncated Rayleigh series contained too many terms; this behavior is typical of asymptotic expansions.

Before considering the boundary integral equations, we need to look in some detail at that part of the extinction theorem which underlies the HC procedure. In the present work, two quite different mathematical approaches are combined, which must be mutually consistent if the results are to be taken seriously.

B. Assumptions behind the extinction theorem

The extinction theorem formulas for a plane corrugated surface are obtained from Green's second identity in the form of the Helmholtz-Kirchhoff integral. However, the formulas are applied in a form that depends on certain definite approximations, which are sometimes not recognized. These have to be made explicit in order to investigate the Rayleigh hypothesis.

Figure 1 shows the geometry of the scattering object alone. Figure 2 shows the configuration of the scattering problem. No dimension in Fig. 2 is assumed to be infinite: the drawings could serve as a schematic for an experiment. But the problem becomes mathematically tractable if the various dimensions satisfy

$$D_S \gg R \gg W \gg D_O \ , \tag{1.4a}$$

$$\lambda \ll t_c ; \quad \lambda \ll W \ , \tag{1.4b}$$

$$W \gg D \ , \tag{1.4c}$$

where D_S is the source-scatterer distance, R is the radius of a large circle used to close the volume of Green's identity, $2W$ is the cross section of the scattering object, D_O is the observer-scatterer distance, t_c is twice the radius of curvature of the scatterer at the shadow line, λ is the wavelength of the incident light, and D is the period of the corrugation. The Helmholtz-Kirchhoff integral, as it is always applied to large planar objects, requires all of the inequalities (1.4). We consider only two-dimensional geometry, so that the curves of Fig. 2 are cross sections of cylinders.

For this geometry, Green's second identity is

$$\int_V (\Psi \nabla'^2 \Phi - \Phi \nabla'^2 \Psi) dv' = \int_{C+C_R} (-\Psi \nabla' \Phi + \Phi \nabla' \Psi) \cdot \hat{\mathbf{n}}' \, ds' \ , \tag{1.5}$$

where $\hat{\mathbf{n}}'$ is the normal *into* V from its bounding surfaces

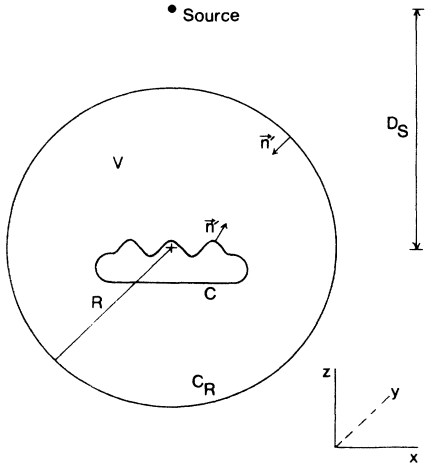

FIG. 2. Configuration and notation for the scattering problem. Two-dimensional scattering is considered: C is the cross section of a cylinder illuminated by a line source. C_R is a large cylinder used to close a volume V for application of Green's identity.

and s' is arclength. Now let $\Phi(\mathbf{r})$ be the total transverse field: E in the case of s-wave polarization, H in the p-wave case. $\Phi(\mathbf{r})$ satisfies

$$(\nabla^2+k^2)\Phi=0 \ . \tag{1.6}$$

Let $\Psi\equiv G(\mathbf{r},\mathbf{r}')$ be the free-space Green's function satisfying

$$(\nabla^2+k^2)G(\mathbf{r},\mathbf{r}')=-4\pi\delta(\mathbf{r}-\mathbf{r}') \ , \tag{1.7}$$

where, for two dimensions,

$$G(\mathbf{r},\mathbf{r}')=i\pi H_0^{(1)}(k\rho); \quad \rho=|\mathbf{r}-\mathbf{r}'| \tag{1.8}$$

and $H_0^{(1)}$ is the usual Hankel function of the first kind.

Equations (1.6)–(1.8) are now put into (1.5) to obtain

$$4\pi\Phi(\mathbf{r})=i\pi\int_{C+C_R}\left[-H_0^{(1)}(k\rho)\frac{\partial\Phi(\mathbf{r}')}{\partial n'}+\Phi(\mathbf{r}')\frac{\partial H_0^{(1)}(k\rho)}{\partial n'}\right]ds' \ . \tag{1.9}$$

If the first inequality of (1.4a) holds, then the incident wave can be considered a plane wave; if the second of (1.4a) holds, then the Hankel function in (1.9) can be replaced by the leading term in its asymptotic series.[23] Carlson and Heins[24] showed rigorously that under these conditions

$$i\pi\int_{C_R}\left[-H_0^{(1)}(k\rho)\frac{\partial\Phi(\mathbf{r}')}{\partial n'}+\Phi(\mathbf{r}')\frac{\partial H_0^{(1)}(k\rho)}{\partial n'}\right]ds'=4\pi\Phi_{\text{inc}}(\mathbf{r}) \ , \tag{1.10}$$

where Φ_{inc} is the incident wave, so that (1.9) becomes

$$\Phi(\mathbf{r})=\Phi_{\text{inc}}(\mathbf{r})-\frac{i}{4}\int_C\left[H_0^{(1)}(k\rho)\frac{\partial\Phi(\mathbf{r}')}{\partial n'}-\Phi(\mathbf{r}')\frac{\partial H_0^{(1)}(k\rho)}{\partial n'}\right]ds' \ . \tag{1.11a}$$

Without further specification about the surface described by C, any analysis that leads to (1.11a) requires that the object have finite size.[25] Taking certain limits to infinity must be the last step in the development.

Equation (1.11a) still involves an integral over the entire surface of the scatterer. But the back of the scatterer can be ignored if the radius of curvature of the sample is large in the region where the shadow begins and if the cross section presented to the source is large, compared to the incident wave length.[26] In this case the surface fields decay exponentially with distance from the shadow boundary into the shadow region.[27] Thus, inequalities (1.4b) allow us to approximate (1.11a) by

$$\Phi(\mathbf{r})=\Phi_{\text{inc}}(\mathbf{r})-\frac{i}{4}\int_{-S}^{S}\left[H_0^{(1)}(k\rho)\frac{\partial\Phi(\mathbf{r}')}{\partial n'}-\Phi(\mathbf{r}')\frac{\partial H_0^{(1)}(k\rho)}{\partial n'}\right]ds' \ , \tag{1.11b}$$

where $2S$ is the arclength of the illuminated surface. From Fig. 1,

$$S=\int_0^W\frac{ds'}{dx'}dx'=\int_0^W[1+(\zeta_{x'})^2]^{1/2}dx' \ . \tag{1.12}$$

For the present work, the critical approximation is the final one: that of extending $S\to\infty$ in Eq. (1.11b). This will be physically realistic for scattering objects that satisfy inequality (1.4c) and the last member of (1.4a). In studies that have been concerned with the mathematical validity of this step, it has been customary[23,28,29] to give k a small positive imaginary part, to ensure convergence of the integrals. The physical fields on the boundary are then assumed to be the limiting values of Φ and $\partial\Phi/\partial n'$ obtained from (1.11b) as first $S\to\infty$ and then $\text{Im}k\to 0$. Since the domain of validity with respect to complex dielectric constant is itself an important part of the question under study here, it is preferable not to let any part of this depend on making specific assumptions about k. In Appendix A we show that for the special case of the periodically corrugated surface, S can be taken to infinity in (1.11b) even when k is real, and the integrals remain bounded. This fortunate result was unexpected: it seems to have been previously unnoticed. It will be used in both Secs. II and III to follow, in which Eq. (1.11b) is the starting point for two separate developments that will then be put together in Sec. IV. In Appendix B we show how this result is related to another surprising fact discovered by Millar, which played a key role in his examination of the Dirichlet problem in Ref. 20.

From here we proceed as follows. Section II outlines the method of Hill and Celli for the surface with pure sinusoidal corrugations. Section III obtains the boundary integral equations that determine the surface field functions. Section IV uses Millar's method to analytically continue the surface field functions, and shows that the continuation is sufficient to validate The HC calculation. In all of this, we use a notation as close as possible to that of the original authors, to make that work immediately accessible to the reader. The new result is that the Rayleigh hypothesis is valid for real materials under the same conditions established for the Dirichlet problem by Millar. In Sec. V we discuss the little available data from this standpoint, and relate Rayleigh's method to two other mathematical treatments of rough surfaces. Section VI summarizes the results.

II. THE HILL-CELLI (HC) METHOD IN TWO DIMENSIONS

We begin by setting $S=\infty$ in Eq. (1.11b) to obtain

$$\Phi_1(\mathbf{r})=\Phi_{\text{inc}}(\mathbf{r})-\frac{i}{4}\int_{-\infty}^{\infty}\left[H_0^{(1)}(k\rho)\frac{\partial\Phi_1(\mathbf{r}')}{\partial n'}-\Phi_1(\mathbf{r}')\frac{\partial H_0^{(1)}(k\rho)}{\partial n'}\right]ds' \ , \tag{2.1}$$

where the subscript "1" indicates that (2.1) is written for medium 1 in Fig. 1. Into Eq. (2.1) we now put the expan-

sion for the Hankel function[30]

$$H_0^{(1)}(k\rho)=\frac{1}{\pi}\int_{-\infty}^{\infty} dq\,(k^2-q^2)^{-1/2}e^{i\mathbf{q}^+\cdot\boldsymbol{\rho}}\ , \tag{2.2}$$

where

$$\boldsymbol{\rho}=\mathbf{r}-\mathbf{r}';\quad \rho=|\boldsymbol{\rho}|;\quad \mathbf{q}^+\equiv\hat{\mathbf{x}}q+\hat{\mathbf{z}}(k^2-q^2)^{1/2}\ .$$

The scattered field and the incident field are expressed as Fourier series:

$$\Phi_1(\mathbf{r})\rightarrow\sum_G F_G e^{i\mathbf{k}_G\cdot\mathbf{r}}\ , \tag{2.3a}$$

$$\Phi_{\text{inc}}(\mathbf{r})=e^{i\mathbf{k}_i\cdot\mathbf{r}}\ . \tag{2.3b}$$

Then (2.1) becomes (after some algebraic manipulations)

$$F_G=\frac{1}{2p_G D}\int_0^D dx\,e^{\gamma_m(x)}[M_1(x)(p_G-K_G\zeta_x)-iL_1(x)]\ , \tag{2.4}$$

$$\gamma_m(x)\equiv -ip_G\zeta(x)-iK_G x\ ,$$

which is the starting point of Ref. 21. (The expression for F_G given in Ref. 21 contains two typographical errors.) In Eqs. (2.3) and (2.4),

$$\begin{aligned}&\mathbf{k}_i\equiv(K_0,-p_0),\quad \mathbf{k}_G\equiv(K_G,p_G)\ ,\\ &k_0\equiv\frac{\omega}{c},\quad K_G\equiv K_0+G\ ,\\ &p_0\equiv(\epsilon_1 k_0^2-K_0^2)^{1/2},\quad p_G\equiv(\epsilon_1 k_0^2-K_G^2)^{1/2}\ ,\\ &\zeta_x\equiv\frac{d\zeta}{dx},\quad G_m\equiv\frac{2\pi m}{D}=m\kappa\quad (m=0,\pm1,\ldots)\ .\end{aligned} \tag{2.5}$$

Here the surface profile is described by $z=\zeta(x)$ as in Eq. (1.3), and the variable of integration has been changed from that in (2.1) using

$$ds'=(1+\zeta_x^2)^{1/2}dx\ . \tag{2.6}$$

Finally, the (unknown) boundary fields in (2.4) are

$$\begin{aligned}&M_1(x)\equiv\Phi_1(\mathbf{r}'\text{ on boundary})\equiv\phi_1(s')\ ,\\ &L_1(x)\equiv(1+\zeta_x^2)^{1/2}\frac{\partial\Phi_1}{\partial n'}(\mathbf{r}'\text{ on boundary})\\ &\qquad\equiv(1+\zeta_x^2)^{1/2}\chi_1(s')\ ,\end{aligned} \tag{2.7}$$

which relate the HC notation to that of Millar. Note that $M_1(x)$ and $L_1(x)$ could be found explicitly as in Ref. 3, but the purpose here is only to find out when (2.3a) converges for all $z>\zeta(x)$, with F_G given by (2.4). For that, it is much more useful to know the general behavior of M_1 and L_1, which is found in Sec. IV.

Convergence is determined by the behavior of F_G when $|G|\rightarrow\infty$. HC evaluate F_G by deforming the integration contour to use the method of steepest descents. For large $|G|$, the exponent in Eq. (2.4) is

$$\gamma_m(x)\cong|m|\kappa\zeta(x)-im\kappa x\ , \tag{2.8}$$

which has saddle points at

$$\frac{d\gamma_m}{dx}=0=|m|\kappa\zeta_x-im\kappa\ . \tag{2.9}$$

For definiteness we consider only $m>0$. Then for the surface profile given by Eq. (1.3) we set $x\equiv x_1+ix_2$ to find that the saddle points are located at

$$\begin{aligned}&\kappa x_{1\text{SP}}=j\pi\quad (j=0,1,2)\\ &x_{2\text{SP}}=(-1)^{j+1}\sinh^{-1}\Gamma\ ,\end{aligned} \tag{2.10}$$

where Γ is given by (1.1). $L_1(x)$ has branch points given by

$$\zeta_x=\pm i\ , \tag{2.11}$$

half of which coincide with the saddle points of the exponent $\gamma_m(x)$. Since $L_1(x_{\text{SP}})=0$ at the saddle points x_{SP}, we simply omit an arbitrarily small neighborhood of each such point from the deformed contour. The analytic structure of $L_1(x)$ in the complex x plane is shown in Fig. 3(a). The cuts are drawn in such a way that the required deformation will be possible.

Although Millar's calculations were completely different from that of HC, his earlier work required the same deformed contour as theirs.[16] As he showed, only

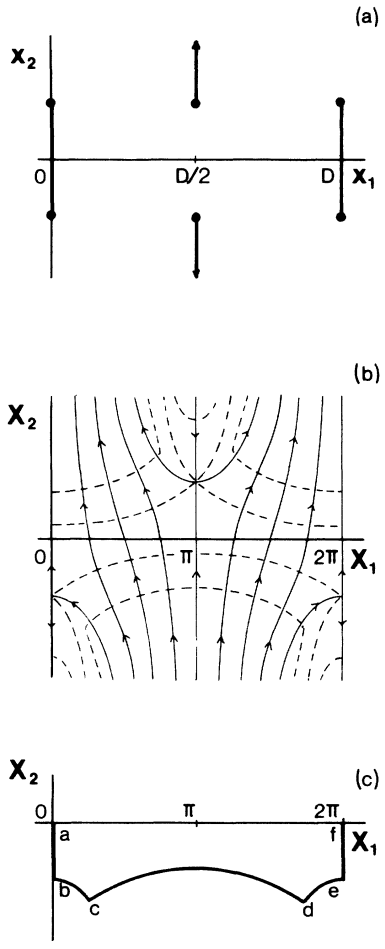

FIG. 3. Geometry of the complex plane as it enters into the calculation of F_G. (a) The branch points and associated cuts of $L_1(x)$. (b) The contours of constant phase (solid lines) and constant magnitude (dashed lines) of $\exp(i\gamma_m)$. (c) Required deformation of the integration path to evaluate F_G by the method of steepest descents.

the two saddle points of γ_m in the lower half-plane are useful for the deformation. With the definitions

$$X_1 \equiv \kappa x_1; \quad X_2 \equiv \kappa x_2 \tag{2.12}$$

the contours of constant phase and magnitude of e^{γ_m} are given by

$$\begin{aligned} &\mathrm{Im}\gamma_m(x) = m(-\Gamma \sin X_1 \sinh X_2 - X_1) = \text{const} , \\ &\mathrm{Re}\gamma_m(x) = m(\Gamma \cos X_1 \cosh X_2 + X_2) = \text{const} , \end{aligned} \tag{2.13}$$

which are plotted in Fig. 3(b). The contours of integration must be deformed as in Fig. 3(c). The contributions from segments *ab* and *ef* cancel because the integrand is periodic. (It is not symmetric because the phase increases with X_1.) The contribution from *cd* is vanishingly small because *cd* is chosen to be a "level curve" of e^{γ_m} along which $\mathrm{Re}\gamma_m(x) \ll \mathrm{Re}\gamma_m(x_{SP})$ is constant. Hence only *bc* and *de* contribute to the integral in Eq. (2.4), and those contributions are found by the familiar method of steepest descents.

Evaluated in this way, F_G has the asymptotic dependence for large m:

$$|F_G| \sim e^{\mathrm{Re}\gamma_m} = \exp[m(\Gamma \cos X_1 \cosh X_2 + X_2)] . \tag{2.14}$$

The exponential that multiplies F_G in (2.3a) behaves for large m as

$$e^{i\mathbf{k}_G \cdot \mathbf{r}} \sim e^{-|m\kappa|z} . \tag{2.15}$$

If (2.3a) is to converge for $z > -h$, then Γ must satisfy

$$\Gamma \cos X_{1SP} \cosh X_{2SP} + X_{2SP} + \Gamma < 0 ,$$

where $X_{SP} \equiv X_{1SP} + iX_{2SP}$; or, using (2.10),

$$(1+\Gamma^2)^{1/2} - \sinh^{-1}\left(\frac{1}{\Gamma}\right) + \Gamma < 0 , \tag{2.16}$$

which requires that

$$\Gamma < 0.448 . \tag{2.17}$$

This estimate is always available if M_1 and L_1 can be analytically continued throughout a strip that includes the axis of real x, such that the saddle points in the lower half-plane lie within the strip of analyticity. In Sec. IV we will show that the required continuation is possible for any ϵ_1 and ϵ_2, so that the criterion (2.17) established by Millar for the Dirichlet problem is also valid for any real materials, independently of their dielectric constants.

III. BOUNDARY INTEGRAL EQUATIONS

The starting point for this section is again Eq. (1.11). The coordinates are defined in Fig. 1. We follow Millar's notation almost without exception. Q is a field point in the volume V of Fig. 2. s' is a value of arclength that serves as the coordinate of a source point on the surface. The boundary values of the field Φ_1 and its normal derivative are denoted by

$$\begin{aligned} &\Phi_1(\mathbf{r}')\Big|_{\mathbf{r}' \text{ on } C} \equiv \phi_1(s'); \quad \frac{\partial \Phi_1(\mathbf{r}')}{\partial n'}\Big|_{\mathbf{r}' \text{ on } C} \equiv \chi_1(s') ; \\ &\Phi_{\mathrm{inc}}(\mathbf{r}')\Big|_{\mathbf{r}' \text{ on } C} \equiv \phi_{\mathrm{inc}}(s'); \quad \frac{\partial \Phi_{\mathrm{inc}}(\mathbf{r}')}{\partial n'}\Big|_{\mathbf{r}' \text{ on } C} \equiv \frac{\partial \phi_{\mathrm{inc}}(s')}{\partial n'} . \end{aligned} \tag{3.1}$$

Let Q approach a point on the boundary C whose arclength coordinate is τ, so that (1.11a) becomes

$$\phi_{\mathrm{inc}}(\tau) - \lim_{Q \to \tau} \frac{i}{4} \int_C \left[H_0^{(1)}(k_1\rho)\chi_1(s') - \frac{\partial H_0^{(1)}(k_1\rho)}{\partial n'} \phi_1(s') \right] ds' = \phi_1(\tau) . \tag{3.2}$$

Now the familiar limiting operation[31] is used to separate out the integrable singularity, resulting in a Fredholm integral equation of the second kind for the boundary function ϕ_1:

$$2\phi_{\mathrm{inc}}(\tau) = \phi_1(\tau) + \frac{i}{2} \int_C^* ds' \phi_1(s') k_1\rho H_1^{(1)}(k_1\rho) \left[\frac{1}{\rho} \frac{\partial \rho}{\partial n'} \right] + \frac{i}{2} \int_C^* ds' \chi_1(s') H_0^{(1)}(k_1\rho) , \tag{3.3}$$

where $\int_C^*$ denotes the Cauchy principal value of the integral. Equation (3.3) is the first boundary integral equation we need.

Next we differentiate Eq. (1.11a) at a point Q very near to C along the normal to C, and let $Q \to \tau$ once again, to obtain

$$\frac{\partial \phi_{\mathrm{inc}}(\tau)}{\partial n_\tau} = \lim_{Q \to \tau} \frac{i}{4} \frac{\partial}{\partial n_Q} \int_C \left[H_0^{(1)}(k_1\rho)\chi_1(s') - \frac{\partial H_0^{(1)}(k_1\rho)}{\partial n'} \phi_1(s') \right] = \chi_1(\tau) . \tag{3.4}$$

The last integral in (3.4) requires a procedure given by Millar;[19] otherwise the limiting process is the same as that which leads to (3.3). Thus Eq. (3.4) becomes

$$2\frac{\partial \phi_{\mathrm{inc}}(\tau)}{\partial n_\tau} = \chi_1(\tau) - \frac{i}{2} \int_C^* ds' \chi_1(s') k_1\rho H_1^{(1)}(k_1\rho) \left[\frac{1}{\rho} \frac{\partial \rho}{\partial n_\tau} \right] + \frac{i}{2} \int_C^* ds' \phi_1(s') (k_1\rho)^2 H_0^{(1)}(k_1\rho) \left[\frac{1}{\rho} \frac{\partial \rho}{\partial n'} \right] \left[\frac{1}{\rho} \frac{\partial \rho}{\partial n_\tau} \right] + \frac{i}{2} \int_C^* ds' \frac{\partial}{\partial n_\tau} \left[\frac{1}{\rho} \frac{\partial \rho}{\partial n'} \right] \left[k_1\rho H_1^{(1)}(k_1\rho)\phi_1(s') + \frac{2i}{\pi} \phi_1(\tau) \right] \tag{3.5}$$

which is the secondary boundary integral equation we need.

All normal derivatives are taken outward from medium 2 into V. The equations for medium 2 analogous to (3.3) and (3.5) can be written by inspection, by everywhere changing the signs of $\partial/\partial n'$, $\partial/\partial n_\tau$, and χ in (3.3) and (3.5), and dropping the incident-field terms. The results are

$$0=\phi_2(\tau)-\frac{i}{2}\int_C^* ds'\,\phi_2(s')k_2\rho H_1^{(1)}(k_2\rho)\left[\frac{1}{\rho}\frac{\partial\rho}{\partial n'}\right]-\frac{i}{2}\int_C^* ds'\chi_2(s')H_0^{(1)}(k_2\rho) \tag{3.6}$$

and

$$0=\chi_2(\tau)+\frac{i}{2}\int_C^* ds'\chi_2(s')k_2\rho H_1^{(1)}(k_2\rho)\left[\frac{1}{\rho}\frac{\partial\rho}{\partial n_\tau}\right]-\frac{i}{2}\int_C^* ds'\phi_2(s')(k_2\rho)^2H_0^{(1)}(k_2\rho)\left[\frac{1}{\rho}\frac{\partial\rho}{\partial n'}\right]\left[\frac{1}{\rho}\frac{\partial\rho}{\partial n_\tau}\right]$$
$$-\frac{i}{2}\int_C^* ds'\frac{\partial}{\partial n_\tau}\left[\frac{1}{\rho}\frac{\partial\rho}{\partial n'}\right]\left[k_2\rho H_1^{(1)}(k_2\rho)\phi_2(s')+\frac{2i}{\pi}\phi_2(\tau)\right]. \tag{3.7}$$

For brevity, we denote

$$K\equiv\frac{1}{\rho}\frac{\partial\rho}{\partial n'};\quad K_\tau\equiv\frac{1}{\rho}\frac{\partial\rho}{\partial n_\tau};\quad N\equiv\frac{\partial}{\partial n_\tau}\left[\frac{1}{\rho}\frac{\partial\rho}{\partial n'}\right] \tag{3.8}$$

for which explicit expressions will be given shortly.

We now add Eq. (3.3) to Eq. (3.6); then add (3.5) to (3.7). ϕ_2 and χ_2 are eliminated using the Maxwell boundary conditions

$$\phi_2=\phi_1;\quad \frac{\chi_1}{\eta_1}=\frac{\chi_2}{\eta_2}, \tag{3.9}$$

where $\eta_1=\eta_2=1$ for TE (s-wave) polarization, and $\eta_1=\epsilon_1$; $n_2=\epsilon_2$ for TM (p-wave) polarization. Here, ϵ_j is the dielectric constant of medium j. These manipulations yield coupled integral equations for ϕ_2 and χ_1:

$$2\phi_1(\tau)+\frac{i}{2}\int_C^* ds'\,\phi_1(s')K[k_1\rho H_1^{(1)}(k_1\rho)-k_2\rho H_1^{(1)}(k_2\rho)]+\frac{i}{2}\int_C^* ds'\,\chi_1(s')\left[H_0^{(1)}(k_1\rho)-\frac{\eta_2}{\eta_1}H_0^{(1)}(k_2\rho)\right]=2\phi_{\rm inc}(\tau) \tag{3.10a}$$

and

$$\chi_1(\tau)\left[1+\frac{\eta_2}{\eta_1}\right]+\frac{i}{2}\int_C^* ds'\,\phi_1(s')KK_\tau[(k_1\rho)^2H_0^{(1)}(k_1\rho)-(k_2\rho)^2H_0^{(1)}(k_2\rho)]$$
$$+\frac{i}{2}\int_C^* ds'\,\phi_1(s')N[k_1\rho H_1^{(1)}(k_1\rho)-k_2\rho H_1^{(1)}(k_2\rho)]$$
$$-\frac{i}{2}\int_C^* ds'\,\chi_1(s')K_\tau\left[k_1\rho H_1^{(1)}(k_1\rho)-\frac{\eta_2}{\eta_1}k_2\rho H_1^{(1)}(k_2\rho)\right]=2\frac{\partial\phi_{\rm inc}(\tau)}{\partial n_\tau}. \tag{3.10b}$$

Kress and Roach[32] showed on purely mathematical grounds that when $\mathrm{Im}k_1\geq 0$ and $\mathrm{Im}k_2\geq 0$, Eqs. (3.10) have a unique solution under very general conditions. One of their specifications implies that k_1 and k_2 are such that ϕ_1 and χ_1 do not correspond to a natural oscillation of the space contained within C. The same assumption is made in the present work.

Following the discussion in Sec. I, the integrals can be evaluated using only the illuminated part of C. The theory of the solubility of equations such as (3.10), whether starting from Fredholm's theorems or from more abstract linear functional analysis,[33] assumes that the integration interval is of finite length. In Appendix A, however, we show that the integrals in Eqs. (3.10) exist even for $S=\infty$, so that no operational distinction exists between "very large" and "infinite." We therefore carry out the same sequence of approximations in (3.10) that underlie the extinction theorem formulas of Sec. II, namely

$$\int_C^* \to \int_{-S}^{S*} \to \int_{-\infty}^{\infty *} \equiv \int^*, \tag{3.11}$$

where again the asterisk indicates the Cauchy principal value.

At this point, ϕ_1 and χ_1 could themselves be analytically continued to complex values of arclength, and then introduced into the HC calculation using (2.7). But the discussion is more direct and easier to follow if we change variables in Eqs. (3.10) themselves from arclength s' to Cartesian coordinate x_s using (2.6), and correspondingly go from τ to the related x_τ. In this case, the explicit expressions for K, K_τ, and N of (3.8) are

$$K\,ds' = \frac{-(x_s-x_\tau)\dfrac{dz_s}{dx_s}+(z_s-z_\tau)}{\rho^2}dx_s\ , \tag{3.12a}$$

$$K_\tau = \frac{(x_s-x_\tau)\dfrac{dz_\tau}{dx_\tau}-(z_s-z_\tau)}{\rho^2}\frac{dx_\tau}{d\tau}\ , \tag{3.12b}$$

$$N\,ds' = -\frac{\left[\dfrac{dz_s}{dx_s}\dfrac{dz_\tau}{dx_\tau}-1\right][(z_s-z_\tau)^2-(x_s-x_\tau)^2]+2(x_s-x_\tau)(z_s-z_\tau)\left[\dfrac{dz_\tau}{dx_\tau}+\dfrac{dz_s}{dx_s}\right]}{\rho^4\left[1+\left[\dfrac{dz_\tau}{dx_\tau}\right]^2\right]^{1/2}}dx_s\ , \tag{3.12c}$$

where

$$\rho^2=(x_s-x_\tau)^2+(z_s-z_\tau)^2$$
$$\frac{dz_i}{dx_i}\equiv\left.\frac{d\zeta}{dx}\right|_{x=x_i},\quad (i=s,\tau)\ . \tag{3.13}$$

Putting (3.11), (3.12), and (3.13) into (3.10), and using (2.7) to explicitly relate Millar's work to HC, we obtain the coupled integral equations in the form

$$2M_1(x_\tau)+\frac{i}{2}\int^* M_1(x_s)\frac{-(x_s-x_\tau)\dfrac{dz_s}{dx_s}+(z_s-z_\tau)}{\rho^2}[k_1\rho H_1^{(1)}(k_1\rho)-k_2\rho H_1^{(1)}(k_2\rho)]dx_s$$
$$+\frac{i}{2}\int^* L_1(x_s)\left[H_0^{(1)}(k_1\rho)-\frac{\eta_2}{\eta_1}H_0^{(1)}(k_2\rho)\right]dx_s=2\phi_{\text{inc}}(x_\tau) \tag{3.14a}$$

and

$$L_1(x_\tau)\left[1+\frac{\eta_2}{\eta_1}\right]-\frac{i}{2}\int^* L_1(x_s)\frac{(x_s-x_\tau)\dfrac{dz_\tau}{dx_\tau}-(z_s-z_\tau)}{\rho^2}\left[k_1\rho H_1^{(1)}(k_1\rho)-\frac{\eta_2}{\eta_1}k_2\rho H_1^{(1)}(k_2\rho)\right]dx_s$$
$$+\frac{i}{2}\int^* M_1(x_s)\frac{\left[(x_\tau-x_s)\dfrac{dz_s}{dx_s}-(z_\tau-z_s)\right]\left[(x_s-x_\tau)\dfrac{dz_\tau}{dx_\tau}-(z_s-z_\tau)\right]}{\rho^4}[(k_1\rho)^2H_0^{(1)}(k_1\rho)-(k_2\rho)^2H_0^{(1)}(k_2\rho)]dx_s$$
$$+\frac{i}{2}\int^* M_1(x_s)\frac{\left[\dfrac{dz_s}{dx_s}\dfrac{dz_\tau}{dx_\tau}-1\right][(x_s-x_\tau)^2-(z_s-z_\tau)^2]-2\left[\dfrac{dz_s}{dx_s}+\dfrac{dz_\tau}{dx_\tau}\right](x_s-x_\tau)(z_s-z_\tau)}{\rho^4}$$
$$\times[(k_1\rho)^2H_0^{(1)}(k_1\rho)-(k_2\rho)^2H_0^{(1)}(k_2\rho)]dx_s=2\frac{\partial\phi_{\text{inc}}(\tau)}{\partial n_\tau}\left[\frac{dx_\tau}{d\tau}\right]^{-1}\ . \tag{3.14b}$$

At this point, a single physically motivated but mathematically justified set of approximations has produced both Eq. (2.1), which in turn leads to the extinction theorem and Eq. (2.4) for the reflected amplitudes; and the boundary integral Eqs. (3.14) from which we can get general information about the functions M_1 and L_1. [Equations (3.14) could also be solved directly for M_1 and L_1, which would correspond to the method of Zaki and

Neureuther[6] for permeable media.]

In his study of the Dirichlet problem, Millar was able to specify the boundary data $\phi_1(s)$ by an analytic function, and proceed from there to the analytic continuation of $\chi_1(\tau)$. In contrast, the transmission problem involves two boundary functions M_1 and L_1 which are themselves determined by Eqs. (3.14) rather than prescribed *a priori*. We could here restrict k_1 to be real (which includes almost all cases of physical interest) and use the known analytic properties of real elliptic differential equations[34] to find that $\Phi_1(r)$ is analytic in V; then proceed to a discussion of L_1 and M_1. Instead, we use a slightly different approach to preserve as much generality as possible.

Reference 32 established that $\Phi_1(r)$ is twice continuously differentiable in V and once continuously differentiable on C, assuming only that the closed contour C is "Lyapunov" (has a continuously turning tangent). When C is analytic, having at every point a Taylor-series description in terms of some parameter, it is easy to show by continuity considerations that the twice-continuous differentiability of $\Phi_1(r)$ also holds on C itself. Then a different argument by Garabedian[35] shows that Φ_1 is actually analytic throughout V. The analyticity of the boundary function $\zeta(x)$ is then used once again to deduce that the boundary function M_1 is analytic. (Garabedian's argument used only the "smoothness" of C, but he assumed *a priori* the twice-continuous differentiability on C.) These facts having been established, L_1 is then analytic for real x by definition.

Therefore, if two functions can be found that are analytic for complex x and that reduce, respectively, to $M_1(x)$ and $L_1(x)$ when x is real, then these extended functions are the unique analytic continuations required by Hill and Celli. Millar's procedure is now used to find $M_1(x)$ and $L_1(x)$ along the deformed contour of Fig. 3(c).

IV. VALIDITY OF THE RAYLEIGH HYPOTHESIS

A. Analytic continuation of M_1 and L_1 (ϕ_1 and χ_1)

The difference between Millar's work and this is that the integrals in Eqs. (3.14) run over the entire real axis, while Millar's work involved integrals over either a closed finite curve or over one cycle of a periodic surface. In this section we let x become complex. Then ρ becomes complex according to (3.13). It has to be kept in mind that the complete x plane is not being used to represent the real x-z plane, a technique common in two-dimensional problems.[36] Rather, (3.14) involves only one coordinate, x, which is now allowed to take complex values as a computational artifice. From now on, the symbol ρ in Eqs. (3.14) must be replaced by $|\rho|$ in the sense of the complex plane. The symbol ρ will be used from now on for a complex variable.

There are two different types of functions here to be analytically continued. The first consists of the set

$$K, K_\tau, N \tag{4.1}$$

which can have poles when $|\rho|^2=0$; the second consists of the set

$$H_0^{(1)}(k_i|\rho|);\quad (k_i|\rho|)^2 H_0^{(1)}(k_i|\rho|);\quad k_i|\rho| H_1^{(1)}(k_i|\rho|)\ , \tag{4.2}$$

which have branch points at $|\rho|^2=0$. Consider first the set (4.1) for both x_s and x_τ on the real axis. Then K and K_τ have a finite limit as $x_s \to x_\tau$, and N diverges as $(x_s-x_\tau)^{-2}$. However, the kernel containing N, i.e.,

$$N[k_1|\rho| H_1^{(1)}(k_1|\rho|) - k_2|\rho| H_1^{(1)}(k_2|\rho|)]$$

has only a logarithmic singularity at $x_s=x_\tau$ for x_τ on the real axis. The corresponding integral is thus well defined. [The solution theorems applied to Eqs. (3.14) by Kress and Roach hold true because such a singularity is "weak."] So the functions (4.1) contribute no singularities to M_1 and L_1 when x_s and x_τ are real.

Next, consider set (4.2). Each of these contributes at worst a weak singularity to a kernel in which it appears. So the only task in analytic continuation is to cancel out the multivaluedness inherent in the Hankel function of complex argument. Then M_1 and L_1 will be regular within a strip containing the axis of real x. Each of (4.2) is actually a function of $|\rho|^2$ rather than $|\rho|$, so that the branch cut of the mapping $\rho=(\rho^2)^{1/2}$ can be drawn (in the ρ^2 plane) along the cut inherent in the Hankel function. This consideration allows us to worry only about continuation in the ρ^2 plane, and ρ^2 is an entire function of x_τ for every x_s. At this point, then x_τ is allowed to be complex.

From (3.13) we have the relations for small $|x_s-x_\tau|$:

$$\rho^2 \cong (x_s-x_\tau)^2\left[1+\left(\frac{dz_s}{dx_s}\right)^2\right],$$
$$\rho \simeq (x_s-x_\tau)\frac{ds'}{dx_s}\ , \tag{4.3}$$

so that $x_s-x_\tau=|x_s-x_\tau|\cdot e^{i\xi}$. Thus, $\xi=0$ when $x_s>x_\tau$

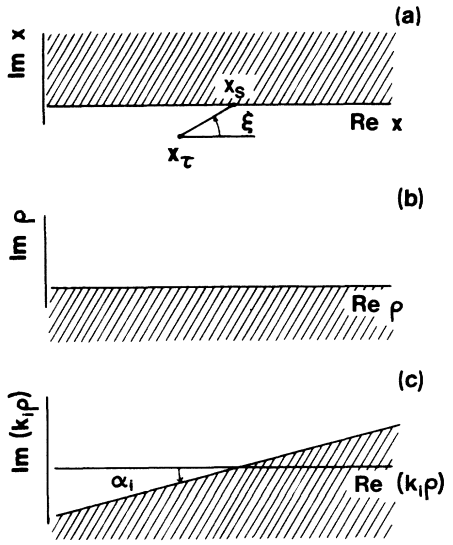

FIG. 4. (a) Domain of x_τ (unshaded) that contains the deformed integration contour for evaluation of F_G. Corresponding domains of (b) ρ and (c) $k_1\rho$.

with x_τ real. Then when x_τ moves to the lower half-plane along the deformed contour, ξ ranges from 0 to π as shown in Fig. 4(a). In particular, when x_τ is real and now $x_s < x_\tau$, $\xi = +\pi$. In that case

$$|\rho| = \rho e^{-i\pi} \quad (x_\tau \text{ real}, \ x_s < x_\tau) . \tag{4.4}$$

Now we can use formula (9.1.37) of Ref. 23:

$$\sin(\nu\pi) H_\nu^{(1)}(ze^{mi\pi}) = -\sin[(m-1)\nu\pi] H_\nu^{(1)}(z) - e^{-\nu i\pi} \sin(m\nu\pi) H_\nu^{(2)}(z) , \tag{4.5}$$

with $m = -1$ and $z = k_i\rho$ from Eq. (4.4): here, $H_\nu^{(2)}$ is the Hankel function of the second kind.

Taking the limit of (4.5) as $\nu \to 0$, we get

$$H_0^{(1)}(k_i|\rho|) = 2H_0^{(1)}(k_i\rho) + H_0^{(2)}(k_i\rho) \tag{4.6a}$$

$$= H_0^{(1)}(k_i\rho) + 2J_0(k_i\rho) \quad (x_\tau \text{ real}, \ x_s < x_\tau) . \tag{4.6b}$$

Then we take the limit of (4.5) as $\nu \to 1$ to get

$$H_1^{(1)}(k_i|\rho|) = -2H_1^{(1)}(k_i\rho) - H_1^{(2)}(k_i\rho)$$

so that

$$k_i|\rho| H_1^{(1)}(k_i|\rho|) = (-k_i\rho)[-2H_1^{(1)}(k_i\rho) - H_1^{(2)}(k_i\rho)]$$

$$= k_i\rho[2H_1^{(1)}(k_i\rho) + H_1^{(2)}(k_i\rho)] \tag{4.7a}$$

$$= k_i\rho[H_1^{(1)}(k_i\rho) + 2J_1(k_i\rho)] \quad (x_\tau \text{ real}, \ x_s < x_\tau) . \tag{4.7b}$$

As a particular example, when (4.7b) is substituted into the first integral on the left-hand side of (3.14a), we get

$$\int_{-\infty}^{\infty *} M_1(x_s) K k_1 |\rho| H_1^{(1)}(k_1|\rho|) dx_s \to \int_{-\infty}^{\infty *} M_1(x_s) K k_1 \rho H_1^{(1)}(k_1\rho) dx_s + 2\int_{-\infty}^{x_\tau} M_1(x_s) K k_1 \rho J_1(k_1\rho) dx_s , \tag{4.8}$$

which was obtained by Millar using a less compact approach involving separation of $k|\rho|H_1^{(1)}(k|\rho|)$ into an entire part and a part with a logarithmic singularity in the complex ρ^2 plane. [See Eqs. (23) and (28) of Ref. 19.] He also continued $k|\rho|H_1^{(1)}(k|\rho|)$ into the lower half ρ^2 plane, that is, into the upper half x plane. By the symmetry principle,[37] the resulting function is the analytic continuation of (4.7a). Since continuation into the lower half-plane is exactly analogous to that into the upper half-plane, and is not needed here, we omit the details.

All the other terms in both Eqs. (3.14) take on a form similar to (4.8). The kernels that appear in these integrals have singularities only for

$$\rho^2 = (x_s - x_\tau)^2 + (z_s - z_\tau)^2 = 0; \quad \mathrm{Im} x_\tau \neq 0 . \tag{4.9}$$

These kernels are analytic functions of complex x_τ near the axis of real x, for all real x_s as a parameter. Therefore Eqs. (3.14), as analytically continued using (4.6a) and (4.7a), formally defined $M_1(x_\tau)$ and $L_1(x_\tau)$ for complex x_τ.

Equations (4.6a), (4.7a), and (4.8) show that the analytic continuation introduces new integrals extending to $x_s = -\infty$, whose kernels contain a Hankel function of the first kind. When x_τ traverses the contour of Fig. 3(c), $0 \leq \arg\rho \leq \pi$, so that

$$\alpha_i \leq \arg(k_i\rho) \leq \alpha_i + \pi , \tag{4.10}$$

where

$$k_i \equiv |k_i| e^{i\alpha_i} \quad (i = 1, 2) \tag{4.11}$$

as illustrated in Figs. 4(b) and 4(c). Because the function $H_\nu^{(1)}(k_i\rho)$ diverges in the lower half-plane,[38] it appears at first sight that these new integrals all diverge whenever $\alpha_i \neq 0$; i.e., for dissipative media. However, for $\arg(k_i\rho) > \pi$ the function $H_\nu^{(2)}(k_i\rho)$ also diverges in the lower half-plane. (This is not obvious from the commonly used asymptotic expansion for $H_\nu^{(2)}$, which holds for a range of arguments that does not include the present case.) It turns out that the divergent parts of $H_\nu^{(1)}$ and $H_\nu^{(2)}$ exactly cancel in (4.6a) and (4.7a). Specifically, in Appendix C we show that

$$2H_0^{(1)}(k_i\rho) + H_0^{(2)}(k_i\rho) \xrightarrow[x_s \to -\infty]{} -3\left[\frac{2}{\pi k_i\rho}\right]^{1/2} e^{-i(k_i\rho - \pi/4)} \quad [\pi \leq \arg(k_i\rho) \leq \pi + \alpha_i] \tag{4.12}$$

and

$$k_i\rho[2H_1^{(1)}(k_i\rho) + H_i^{(2)}(k_i\rho)] \xrightarrow[x_s \to -\infty]{} -3\left[\frac{2k_i\rho}{\pi}\right]^{1/2} e^{i(k_i\rho - 3\pi/4)} \quad [\pi \leq \arg(k_i\rho) \leq \pi + \alpha_i] \tag{4.13}$$

so that the asymptotic forms are well behaved in the lower half $(k_i\rho)$ plane.

From these results we can conclude that the analytic continuations $L_1(x_\tau)$ and $M_1(x_\tau)$ always exist. With (4.6a) and (4.7a), Eqs. (3.14) become a pair of coupled Volterra equation for L_1 and M_1. As an example, term (4.8) can be written,

using (4.7a) rather than (4.7b), in the form

$$\begin{aligned}\int_{-\infty}^{\infty} M_1(x_s)Kk_1|\rho|H_1^{(1)}(k_1|\rho|)dx_s \rightarrow & \int_{-\infty}^{x_\tau} M_1(x_s)Kk_1\rho[2H_1^{(1)}(k_1\rho)+H_1^{(2)}(k_1\rho)]dx_s \\ & + \int_{x_\tau}^{\infty} M_1(x_s)Kk_1\rho H_1^{(1)}(k_1\rho)dx_s \\ = & \int_{-\infty}^{x'} M_1(x_s)Kk_1\rho[2H_1^{(1)}(k_1\rho)+H_1^{(2)}(k_1\rho)]dx_s \\ & + \int_{x'}^{\infty} M_1(x_s)Kk_1\rho H_1^{(1)}(k_1\rho)dx_s \\ & + \int_{x'}^{x_\tau} M_1(x_s)[H_1^{(1)}(k_1\rho)+H_1^{(2)}(k_1\rho)]Kk_1\rho dx_s \\ = & I_C+I_D+I_F ,\end{aligned}$$

where x' is prescribed to be real. Each integral like I_C and I_D can be treated as a known function of x_τ, because M_1 and L_1 are themselves known for real x_s after Ref. 32. Each such integral can therefore be lumped with the "free term" in Eqs. (3.14a) or (3.14b) in which it occurs. On the other hand, the kernel of I_F is analytic within a strip containing the real axis. Thus, for x_τ within the strip, L_1 and M_1 satisfy a pair of coupled Volterra equations of the second kind on finite intervals, whose kernels are bounded and integrable. Such a system always has a solution, which is demonstrated by construction.[39]

It remains to show that when M_1 and L_1 are thus continued, the strip of analyticity that includes the real-x axis also contains the saddle points needed by HC to evaluate F_G by Eq. (2.4). Millar showed that this is true, but his work contains a small error. We correct the mistake and complete the present argument by briefly discussing the roots of Eq. (4.9) in the next few paragraphs.

B. The root loci and validity criterion

Millar called the solutions of (4.9) "root loci." These are most easily studied by factoring (4.9) to get

$$\kappa h\,[\cos(\kappa x_\tau)-\cos(\kappa x_s)]\pm i(\kappa x_\tau-\kappa x_s)=0 . \tag{4.14}$$

We discuss only the root loci having the lower $(-)$ sign in Eq. (4.14). With the definitions

$$\kappa x_\tau \equiv T \equiv T_1+iT_2;\quad \kappa x_s \equiv X \tag{4.15}$$

that member of (4.14) becomes

$$\Gamma(\cos T-\cos X)-i(T-X)=0 , \tag{4.16}$$

where Γ is given by (1.1). This corresponds to two real equations

$$\cos T_1=\frac{\cos X-T_2/\Gamma}{\cosh T_2} \tag{4.17a}$$

$$\Gamma\sinh T_2=\frac{X-T_1}{\sin T_1} . \tag{4.17b}$$

In all of this, $-\infty<X<\infty$ and $0<T_1<2\pi$. [Equations (4.17) are the same as Eqs. (50) and (51) of Ref. 16.] Using (4.17) it is easy to show that the other root loci [with the $(+)$ in (4.14)] are just the reflections of (4.17) through the T_1 axis.

That the root loci are symmetric about the vertical line $T_1=\pi$ can be seen by considering the solutions T' for $X'\equiv 2\pi-X$. We can further compare the root loci centered, respectively, on $T_1=0$ and $T_1=2\pi$ with the locus centered on $T_1=\pi$. In (4.16) we substitute $T\equiv\pi+\Delta_\pi$ to obtain

$$\Gamma[\cos\Delta_\pi-\cos(X-\pi)]+i[\Delta_\pi-(X-\pi)]=0 . \tag{4.18}$$

The solution Δ_π of Eq. (4.18) for $X'\equiv X-\pi$ is the complex conjugate of the solution T of Eq. (4.16) for X. The root loci centered about $T_1=0$ and $T_1=2\pi$ are thus translated (along the T_1 axis) reflections (through the T_1 axis) of the locus centered about $T_1=\pi$, as shown in Fig. 5. This picture is identical to Millar's Fig. 6 of Ref. 16 in the upper half-plane, but it differs from his results in the lower half-plane. Since he did not present the detailed

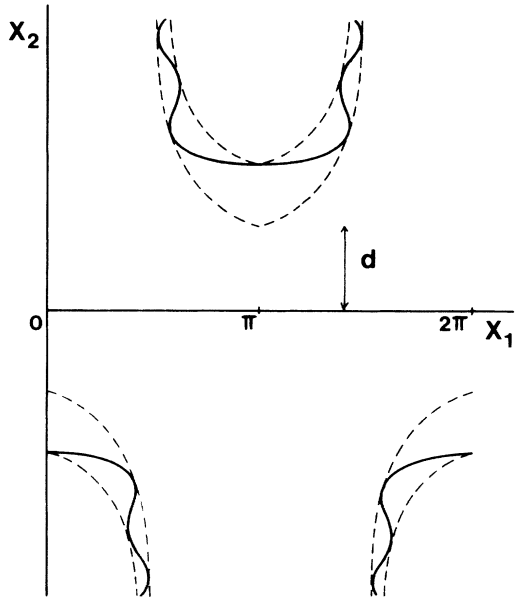

FIG. 5. The "root loci" (solid lines) that demarcate the strip in which $L_1(x)$ and $M_1(x)$ are analytic. The envelopes (dashed lines) pinch together onto the root loci rapidly with decreasing Γ.

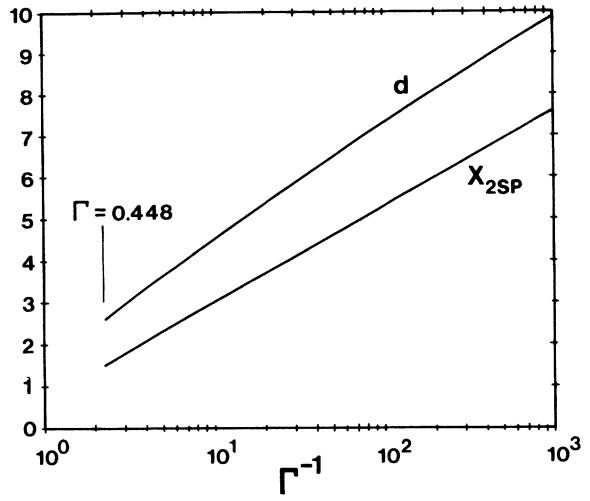

FIG. 6. Saddle-point location X_{2SP}, and the width d of the strip of analyticity of L_1 and M_1 in the lower-half X plane, as functions of Γ.

solutions of his equations [corresponding to (4.16) and (4.17) above], we are unable to find the source of the discrepancy. We also keep in mind that there is another set of root loci determined by taking the upper (+) sign in Eq. (4.14). The complete set is then symmetrical about the T_1 axis and about the T_2 axis.

The envelopes of the root loci are shown as dashed lines in Fig. 5; they are obtained by setting $\cos X = \pm 1$ in Eq. (4.17a). The envelopes approach each other as Γ decreases, but the distance d is always a lower bound on the width of the strip of analyticity. From (4.17a) we obtain

$$\cosh d = 1 + \frac{d}{\Gamma} . \tag{4.19}$$

At the saddle points, X_{2SP} is given by (2.10) and (2.12). Figure 6 shows that $|X_{2SP}| < d$ for all $\Gamma < 0.448$, so that Hill and Celli did prove the validity of the Rayleigh hypothesis subject to this condition, for the Dirichlet problem. For $\Gamma \geq 0.448$, HC showed that the Rayleigh hypothesis would fail even if the saddle points lay within the strip of analyticity.

Figure 6 concludes the theoretical part of this work. We now briefly discuss some published data that are at least consistent with the foregoing conclusions.

V. DISCUSSION

A. Comparison with data

Through the years a number of experimentalists have compared their light- and sound-scattering data with the description offered by Rayleigh's method. There have also been many numerical studies of the convergence of Rayleigh's method for the Dirichlet problem, and also numerical comparisons of this method with other (i.e., extinction theorem, boundary integral equation, etc.) approaches. We know of only one investigation, however, in which identical experiments were conducted on two different metals, with results that invite comparison with Rayleigh's hypothesis.

The experimental configuration of Raether[40] is shown in Fig. 7. He produced gratings of various depth and period in photoresist by holographic exposure, then overlaid these with silver (about 2000 Å) or gold (about 1500 Å). In each case the metal was sufficiently thick to support a bulklike surface-plasmon polariton (SPP) at the metal-air interface.[41] Raether measured the angular position θ and half-width $\theta_{1/2}$ of the reflectance dip associated with the excitation of the SPP by the first-order diffracted wavelet, for each film, at $\lambda_0 = 5682$ Å. These are related to the SPP wave vector and frequency by

$$\sin\theta = \mathrm{Re}s - \frac{\lambda_0}{D} ,$$
$$\theta_{1/2} = 2\frac{\mathrm{Im}s}{\cos\theta} , \tag{5.1}$$

where $s = cq_s/\omega_s$ is the normalized SPP wave vector, q_s the SPP wave vector itself, ω_s the frequency, and λ_0 the vacuum wavelength.

Raether himself fitted his data using the Rayleigh series for the scattered light, and obtained excellent agreement for both metals. His calculations are not useful for our purposes, for two reasons. First, as the corrugation depth goes to zero, (5.1) must give the easily calculated[42] flat-surface values for θ and $\theta_{1/2}$. In fact, Raether's reported grating constants give poor agreement with the data as extrapolated to $h=0$, especially for $D < \lambda_0$, in which case θ is very sensitive to small errors in D. Even if the values of D are adjusted to agree with the extrapolated $\theta(h=0)$, the calculated and extrapolated $\theta_{1/2}(h=0)$ still differ from each other well in excess of the error bars shown on the author's graphs. So we know that the reported data are not precisely correct. Even more importantly, his calculated curves agree beautifully with the data all the way down to $h=0$, so it is certain that he normalized the calculations to the flat-surface values of θ and $\theta_{1/2}$ separately. (See also the discussion of the Kroger-Kretschmann theory below.)

Secondly, it is well known[43,44] that one can often get good results from the Rayleigh method by using sufficiently *few* plane waves, even when the series eventually diverges. Our interest here is in the conditions under which the Rayleigh hypothesis is valid in principle, while Raether's concern was the interpretation of his experiments. We have no information as to how he truncated the series or how it was then behaving.

The determinental equation for the SPP dispersion re-

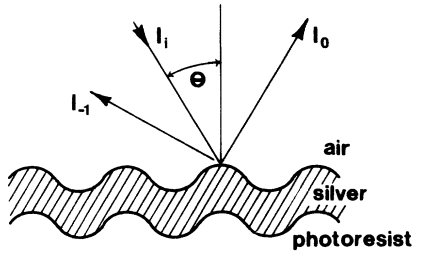

FIG. 7. Raether's experimental configuration.

lation that results from Rayleigh's method is the same one that is obtained from the extinction theorem. The routines for solving such equations require very accurate initial guesses to succeed.[45] But Wirgin[44] has pointed out that finite computer precision will cause a Rayleigh calculation to diverge for any corrugation strength, even one for which the method is valid theoretically, if the matrices are large enough. We see then that a direct convergence test, by summing increasing numbers of Fourier components, is not only very difficult in practice, but actually impossible in principle. There is, however, another way to test Rayleigh's hypothesis, which we now use to analyze Raether's data.

Maradudin[13] developed a perturbation theory that expands the complete sum of the Rayleigh series in powers of the corrugation depth, and which has the correct analytic properties. The resulting SPP dispersion relation, to order h^2, which was also given by Toigo *et al.*,[3] is

$$\epsilon_2 A_1+\epsilon_1 A_2=(\epsilon_2-\epsilon_1)^2 k_0^2 \times \sum_{s'} |\hat{\xi}_{s-s'}|^2 \frac{(ss'-A_1' A_2)(ss'-A_1 A_2')}{\epsilon_2 A_1'+\epsilon_1 A_2'} , \tag{5.2}$$

where s has the same meaning as in (5.1). The other symbols are

$$A_j=(s^2-\epsilon_j)^{1/2}, \quad A_j'=[(s')^2-\epsilon_j]^{1/2} \tag{5.3}$$

$$\hat{\xi}_{s-s'}=\frac{1}{D}\int_0^D e^{-i(s-s')k_0 x}\zeta(x)dx .$$

For the surface of (1.2),

$$\hat{\xi}_{s-s'}=\frac{h}{2}(\delta_{s-s',(\lambda_0/D)}+\delta_{s'-s,(\lambda_0/D)}) \qquad \lambda_0=\frac{2\pi c}{\omega}\equiv\frac{2\pi}{k_0} , \tag{5.4}$$

in which case (5.2) becomes

$$\epsilon_2 A_1+\epsilon_1 A_2=\frac{(\epsilon_2-\epsilon_1)^2}{4}(hk_0)^2(W_+ + W_-) , \tag{5.5}$$

where

$$W_\pm=\frac{(ss_\pm - A_1^\pm A_2)(ss_\pm - A_1 A_2^\pm)}{\epsilon_2 A_1^\pm+\epsilon_1 A_2^\pm}; \quad s_\pm\equiv s\pm\frac{\lambda_0}{D} . \tag{5.6}$$

In previous work[46] we found that to order h^2, (5.5) can be written

$$s^2=\frac{\epsilon_1^2\epsilon_2-\epsilon_2^2\epsilon_1(1-f)^2}{\epsilon_1^2-\epsilon_2^2(1-f)^2} , \tag{5.7}$$

where

$$f\equiv\alpha e^{i\beta}h^2 = \frac{(\epsilon_1-\epsilon_2)^2\sqrt{-\epsilon_1-\epsilon_2}h^2}{4\epsilon_1\epsilon_2}\left[\frac{\omega}{c}\right]^2(W_{+0}+W_{-0}) , \tag{5.8}$$

in which $s_0^2=\epsilon_1\epsilon_2/(\epsilon_1+\epsilon_2)$ describes the flat-surface SPP, and $W_{\pm 0}$ are the values of $W_\pm$ evaluated at $s=s_0$. In that work we obtained f empirically by fitting (5.7) to published data. For this work we use Eq. (5.8) to compute f from information given by Raether. In Ref. 46 we also estimated using Maxwell's equations that (5.7) is useful for $|f|\lesssim 0.1$, roughly. Since (5.7) and (5.5) are identical to $O(h^2)$, we have that the second-order dispersion relation is valid for $h\lesssim h_2$, where

$$h_2=\left[\frac{0.1}{\alpha}\right]^{1/2} . \tag{5.9}$$

Figure 8 shows the magnitude and phase of f/h^2 for silver and gold at $\lambda_0=5682$ Å used by Raether. The sharp peak is due to the degeneracy at the first Brillouin-zone boundary, which can be correctly dealt with only with some sort of degenerate perturbation theory.[45,47]

Because the SPP is a free oscillation, its dispersion cannot depend on the sign of h in (1.2), so the error in (5.5) [and therefore in (5.7)] is $O(h^4)$. For small corrugation depths, this expression is then a very good approximation to the completely summed Rayleigh series if this series converges. Moreover, if the series converges, it represents the unique solution to the physical problem represented by the Helmholtz equation and associated boundary conditions, and hence must describe the corresponding measurements.

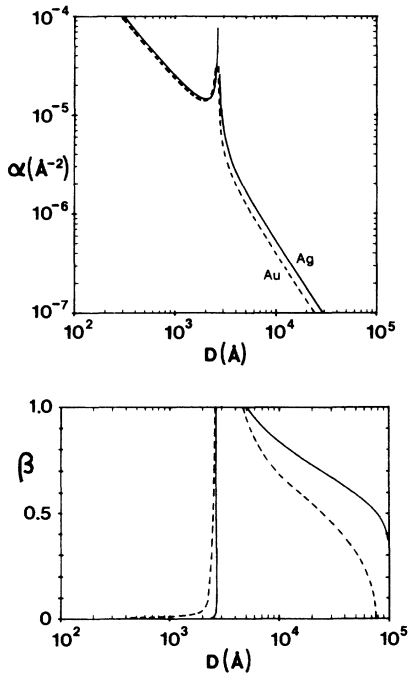

FIG. 8. Magnitude (α) and phase (β) of f/h^2 at $\lambda_0=5682$ Å for simple sinusoidal gratings, as functions of grating period D. Solid curves describe silver; dashed curves describe gold.

TABLE I. Parameters characterizing the four gratings studied by Raether in Ref. 40.

Figure	Metal	Reported period (Å)	Assumed period (Å)	h_{RFM} (Å)	h_2 (Å)	α (10^7 Å^{-2})	β	ϵ
9(a), 10(a)	Ag	4400	4413	313	213	22.1	1.11	$-13.3+i0.49$
9(b), 10(b)	Ag	14910	15000	1065	597	2.81	0.768	$-13.3+i0.49$
11(a), 12(a)	Au	4417	4443	315	247	16.4	1.00	$-6.8+i1.81$
11(b), 12(b)	Au	7965	8003	568	415	5.82	0.744	$-6.8+i1.81$

We evaluate (5.7) using the parameters in Table I. The dielectric constants are Raether's. The "assumed period" is the value of D that brings θ calculated by (5.1) with the flat surface value of s into agreement with Raether's extrapolated value for $h=0$. (In the following subsection we will show that Raether's value for ϵ_{Au} contains a systematic error. In this comparison with the Rayleigh method, however, we need to avoid a wholesale adjustment of constants.) h_{RFM} is Millar's bound, calculated from (2.17), and h_2 is the limit beyond which (5.5) and (5.7) can no longer be trusted. Results are shown in Figs. 9–12. Each theoretical curve has been drawn only as far as h_2. Whenever h_{RFM} lies within the graph, it has been indicated with an arrow. Except for the small changes in

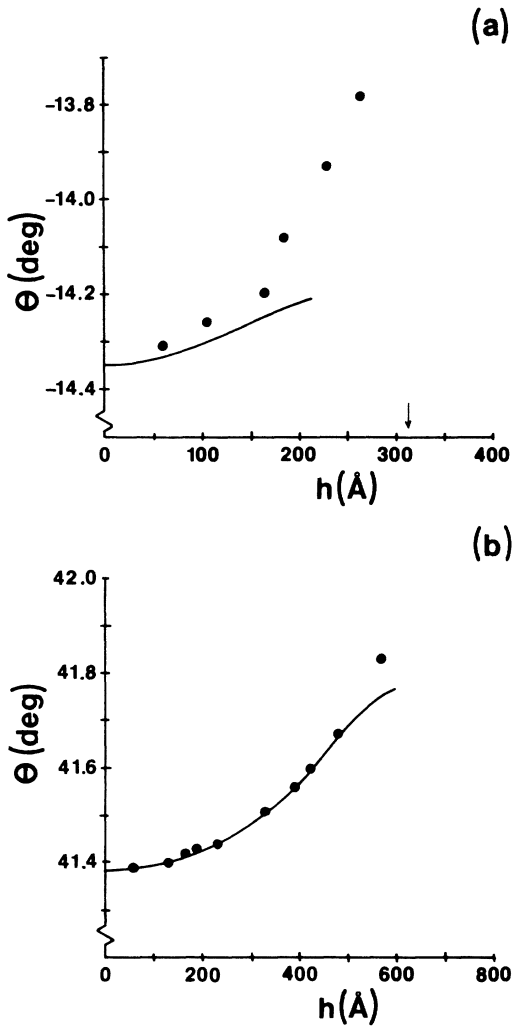

FIG. 9. Angular location of the reflection minimum as a function of grating depth, for silver at $\lambda_0=5682$ Å. When no degeneracy is present, the minimum corresponds to the excitation of a SPP. (a) $D=4413$ Å; (b) $D=15000$ Å. Curves are calculated from Rayleigh theory, out to $h=h_2$, where the second-order perturbation expression is no longer trustworthy. The arrow denotes the theoretical limit of validity of the Rayleigh hypothesis.

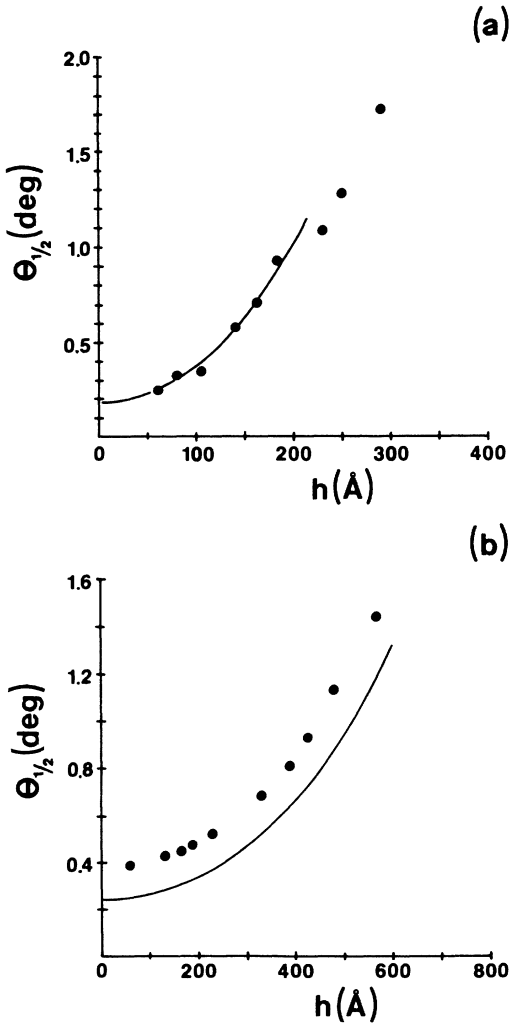

FIG. 10. Angular width of the reflection minimum as a function of grating depth, for silver at $\lambda_0=5682$ Å. When no degeneracy is present, the width is derived from the imaginary part of the SPP wave vector. (a) $D=4413$ Å; (b) $D=15000$ Å.

D (without which the theory would only qualitatively resemble the data in Figs. 9–12), these graphs contain no adjustable parameters. We now consider them from both theoretical and experimental viewpoints.

B. Discrepancies between calculations and data

We first note that each Rayleigh curve shows a slight downturn near $h = h_2$; these confirm (5.9) as an upper bound for the validity of second-order perturbation theory. With that in mind, considering only the data for $h < h_2$, where equation (5.5) can accurately sum the entire Rayleigh series, we see the calculations are considerably closer to the data for silver than for gold. In particular, Fig. 9(b) shows very good agreement, and the translation in Fig. 10(b) can be removed by small changes in D and ϵ_{Ag} that leave Fig. 9(b) essentially unchanged.

We first considered the possibility that higher harmonics in the surface profile function strongly influence the data, and also reduce the value of h_2 above which (5.5) cannot be used. It is very easy to generalize (5.8) to more complicated surface profiles. We found that the curves for both films with $D \approx 4400$ Å could be brought into much better agreement with the data in this way, but only if the second-harmonic amplitudes were 23% (Au) and 17% (Ag) of the fundamental amplitude h of the corrugation. These fractions are far larger than the values measured by Raether for any of the films of Ref. 40. We therefore looked for more likely sources of error.

The minimum prerequisite for a rough-surface light-scattering calculation is that it pass smoothly to the correct flat-surface result as a limit. Figures 9–12 show that some of the given information D, ϵ, and h must be incorrect. In these graphs we suppose ϵ to be correct, but in fact Raether's value for ϵ_{Au} is distorted by the method used to measure it. Kretschmann[48] derived approximate expressions for a nonradiating excitation of a thin metal film on a prism. The formulas give the dielectric constant of the metal from the measured position and linewidth of the reflectance minimum, but they are practical only when $|\epsilon_r| \gg \epsilon_i$, where $\epsilon = \epsilon_r + i\epsilon_i$ is the metal dielectric constant. At $\lambda_0 = 5682$ Å, silver meets this requirement but gold does not. To estimate the error that results, we can ignore the correction term for finite film thickness,

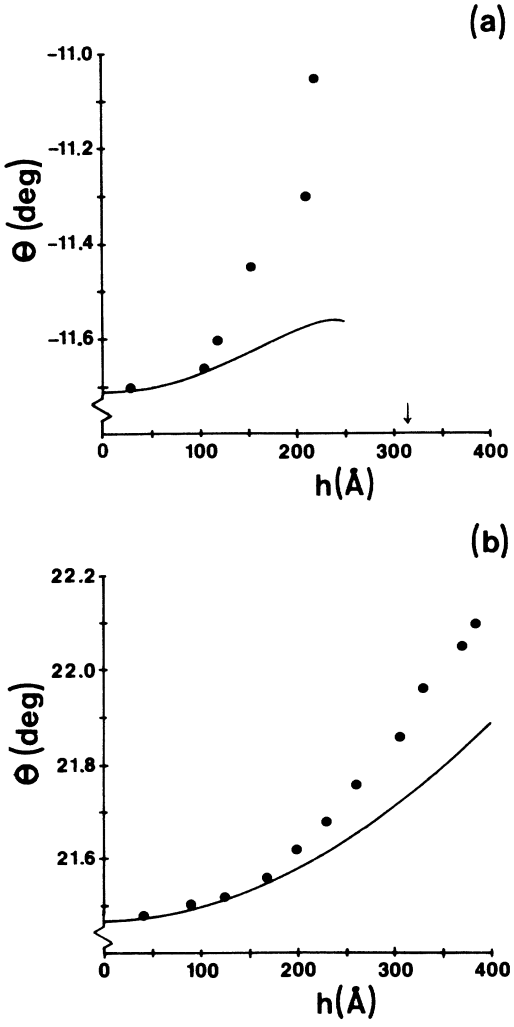

FIG. 11. Same as Fig. 9, but for gold. (a) $D = 4443$ Å; (b) $D = 8003$ Å.

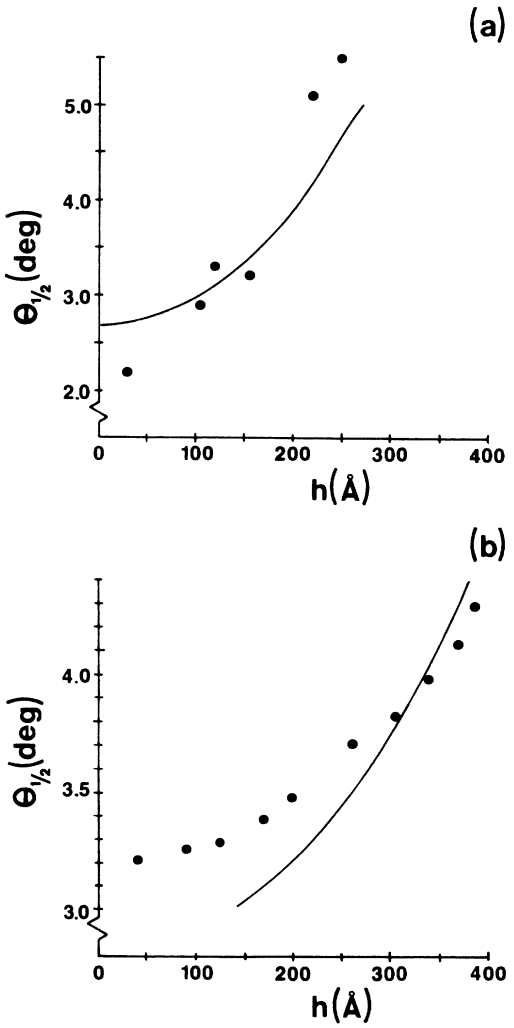

FIG. 12. Same as Fig. 10, but for gold. (a) $D = 4443$ Å; (b) $D = 8003$ Å.

which is very small for films of thickness $d > 500$ Å, as were Raether's.

The SPP dispersion relation for a flat metal surface in air is given by

$$k_\perp \equiv \frac{\omega}{c} s = \frac{\omega}{c} \left[\frac{\epsilon_r + i\epsilon_i}{1+\epsilon_r+i\epsilon_i} \right]^{1/2} . \tag{5.10}$$

When (5.10) is expanded in a power series in ϵ_i, we obtain

$$k_\perp \cong \frac{\omega}{c} \left[\frac{|\epsilon_r|}{|\epsilon_r|-1} \right]^{1/2} \left[1 + \frac{i}{2} \frac{\epsilon_i}{|\epsilon_r|(|\epsilon_r|-1)} + \left[\frac{\epsilon_i}{|\epsilon_r|} \right]^2 \frac{1-4|\epsilon_r|}{8(|\epsilon_r|-1)^2} \right] . \tag{5.11}$$

In Eq. (1.1) of Ref. 48, Kretschmann omitted the second-order term in the square brackets of (5.11) above, and this seems at first to be justified. With gold, $|\epsilon_r| \approx 7$ and $\epsilon_i \approx 1.8$ at $\lambda_0 = 5682$ Å, so that

$$\frac{1}{8} \left[\frac{\epsilon_i}{|\epsilon_r|} \right]^2 \frac{1-4|\epsilon_r|}{(|\epsilon_r|-1)^2} \approx -0.006 .$$

This seemingly small correction turns out to make a big difference in data interpretation. In reality, the measurements for gold give, through Eq. (5.1),

$$\text{Res} \approx \left[\frac{|\epsilon_r|}{|\epsilon_r|-1} \right]^{1/2} (1-0.006) , \tag{5.12}$$

but without the correction term, the investigator interprets

$$\text{Res} \approx \left[\frac{|\bar{\epsilon}_r|}{|\bar{\epsilon}_r|-1} \right]^{1/2} . \tag{5.13}$$

(5.12) and (5.13) show that the correct value $|\epsilon_r|$ is related to the reported value $|\bar{\epsilon}_r|$ by

$$|\epsilon_r| \approx 0.94 |\bar{\epsilon}_r| . \tag{5.14}$$

Likewise, from (5.1), (5.10), (5.11), and (5.14), we find that the true value ϵ_i is related to the reported value $\bar{\epsilon}_i$ by

$$\epsilon_i \approx 0.81 \bar{\epsilon}_i . \tag{5.15}$$

These estimates have been obtained only from the general formula (5.11) and the known bulk dielectric constant of gold. However, we can also take Raether's extrapolated flat-surface values $\theta(h=0)$ and $\theta_{1/2}(h=0)$ as given; then invert (5.1) and (5.10) to find combinations $(D, \epsilon_r, \epsilon_i)$ that simultaneously reproduce $\theta(0)$ and $\theta_{1/2}(0)$. Table II shows the results for the film of Figs. 11(a) and 12(a), for which $D \approx 4400$ Å, $\theta(0) \simeq -11.72°$, and $\theta_{1/2}(0) \approx 2.1°$. For realistic values of D, the corresponding values of ϵ_r and ϵ_i are evidently consistent with (5.14) and (5.15).

This same analysis indicates that for silver at $\lambda_0 = 5682$ Å,

$$|\epsilon_r| \approx 0.99 |\bar{\epsilon}_r| ; \quad \epsilon_i \approx 0.97 |\bar{\epsilon}_i| . \tag{5.16}$$

In less sensitive applications all of these errors would be less significant, since variations in the true dielectric constant of a metal (depending on sample preparation and/or film deposition procedure) are often greater than the errors shown by (5.14)–(5.16).[49] Such natural variation is difficult to eliminate; this is no doubt why Raether normalized his theoretical curves to the flat-surface values. However, we cannot account for the discrepancies in Fig. 12(b) even with (5.14) and (5.15). This film, with $D \approx 8000$ Å, appears to have a special feature, which we will also see from a consideration of the values of h that are used in Figs. 9–12.

TABLE II. Values of D, ϵ_r, and ϵ_i, which correctly reproduce the flat-surface values $\theta(0) \approx -11.72°$ in Fig. 11(a) and $\theta_{1/2} \approx 2.1°$ in Fig. 12(a). The values given in Ref. 40 are $D = 4417$ Å, $\epsilon_r = -6.8$, and $\epsilon_i = 1.81$.

D (Å)	ϵ_r	ϵ_i
4380	−5.89	0.98
4400	−6.20	1.10
4417	−6.50	1.23
4430	−6.74	1.35
4440	−6.95	1.45

The measurements of corrugation depth h for the gold films were also distorted. Raether determined h by measuring the intensity of the low-order diffracted beams of s-polarized light, using a formula derived from the Kirchoff approximation for a perfect conductor.[50] As a specific example, this so-called "scalar theory" gives the relative intensity of the first-order diffracted beam with normally incident light as

$$R_1 = \frac{1}{4} \left[\frac{\omega}{c} h (1+\cos\theta_1) \right]^2 = \left[\frac{2\pi h}{D} \right]^2 \left[\frac{1+[1-(\lambda_0/D)^2]^{1/2}}{2} \right]^2 , \tag{5.17}$$

where θ_1 is the angle at which the first-order beam propagates with respect to the surface normal. Equation (5.17) itself was used earlier by Pockrand and Raether.[51] Besides the assumption of perfect conductivity, (5.17) also requires that $\lambda_0/D \ll 1$.[50,52]

Heitmann[53] derived a much more useful expression from the Kroger-Kretschmann perturbation theory.[54] The general expression is simple, but contains an abundance of symbols. In the special case of the first-order diffracted beam with normally incident light,

$$R_1 = \left[\frac{2\pi h}{D} \right]^2 \left[1 - \left[\frac{\lambda_0}{D} \right]^2 \right]^{1/2} \times \left| \frac{1-\sqrt{\epsilon}}{[1-(\lambda_0/D)^2]^{1/2} + [\epsilon - (\lambda_0/D)^2]^{1/2}} \right|^2 , \tag{5.18}$$

where ϵ is again the dielectric constant of the metal. For the perfect conductor with $\lambda_0/D \ll 1$, (5.18) and (5.17) are equivalent. But (5.18) requires only that $h/D \ll 1$ and $h/\lambda_0 \ll 1$ so that a perturbation theory is valid;

moreover, it does not assume that ϵ is real.

We do not know the details of Raether's configuration. For instance, he could not have used normally incident light at $\lambda_0=5682$ Å, with $D\approx4400$ Å, because no diffracted beams are obtained under those conditions. But the general deficiency of the "scalar theory" is shown in Fig. 13. Here we have evaluated $R_1(h)$ for silver and gold using (5.18), and for the perfect conductor using (5.17), at $\lambda_0=5682$ Å for $D=8000$ Å and $D=15\,000$ Å. We see that as D increases, the scalar theory does very well for silver. It is significantly in error for gold, and the error remains large as D increases. Although (5.17) and (5.18) represent only one experimental configuration, we expect these same discrepancies to occur generally: real gold absorbs, so that a corrugation must be deeper than predicted by (5.17) to produce a given R_1.

Figure 13 suggests that the disagreement between the gold data and Rayleigh calculations is largely due to miscalibration of the h scales in Figs. 11 and 12. It also suggests that the good agreement in Figs. 9(b) and 10(b) is partly due to the large value of D, at which the scalar theory becomes more accurate even for silver.

The linewidth data of Fig. 12(b) are still unexplained, since they give the impression that the measured h values are too large rather than too small. We believe that this is due to a peculiarity of the period D. Figure 14 shows the flat-surface SPP dispersion relation for gold in air, calculated with the optical data of Johnson and Christy.[55] The shaded area on the right-hand side shows the broadening of the state as indicated by Imk. At $\lambda_0=5682$ Å (≈2.18 eV), the SPP has a wave vector $k_{\rm SPP}\approx1.19\times10^5$ cm^{-1}. It is therefore separated from its mirror image (which travels in the opposite direction) by twice this amount, or

$$\Delta k_{\rm deg}\approx2.38\times10^5\ {\rm cm}^{-1}\ .$$

The reciprocal lattice vector of a grating with period 8000 Å is $k_{\rm lat}=7.85\times10^4$ cm^{-1}, and three times this is

$$\Delta k_3=-3k_{\rm lat}=-2.36\times10^5\ {\rm cm}^{-1}\ .$$

Even the flat-surface linewidth is approximately sufficient to let Δk_3 mix the two degenerate states. This quasimomentum could be provided by even a small admixture of third harmonic in the surface profile function. Furthermore, the same argument shows that

$$\Delta k_2=-2k_{\rm lat}=-1.57\times10^5\ {\rm cm}^{-1}$$

couples the incoming radiation directly to the excitation at point B when $k_{\rm lat}$ couples it to the excitation at point A. Raether measured the relative admixture of second harmonic in his surface profiles to be about 6%; this provides the quasimomentum Δk_2. Both of these effects modify the data but are not contained within the second-order nondegenerate perturbation formulas.

A quantitative description of this situation is very complicated, and is beyond the scope of this report. It is

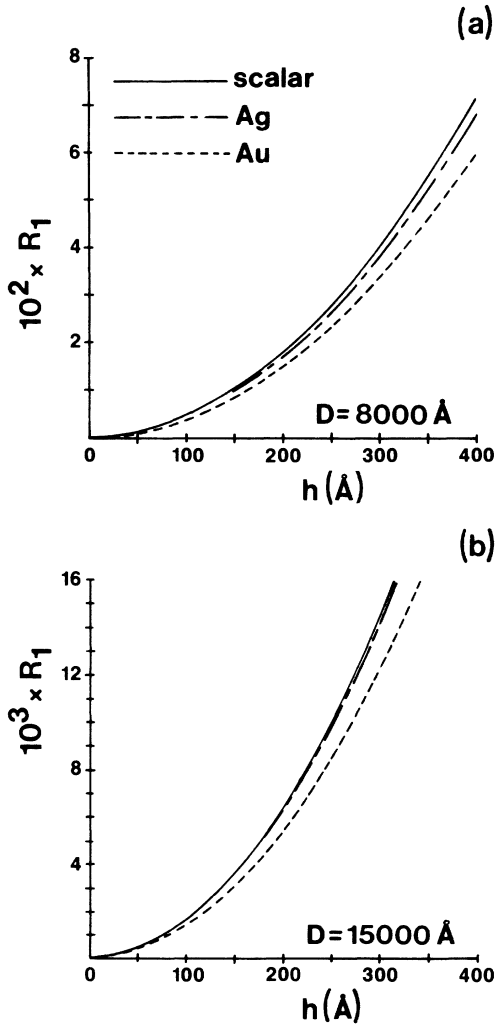

FIG. 13. Relative intensity of s-polarized normally incident light scattered into the first-order diffracted beam. The "scalar theory" is the Kirchoff approximation applied to a perfect conductor. The curves for Ag and Au come from Heitmann's formula for real materials. (a) $D=8000$ Å; (b) $D=15\,000$ Å.

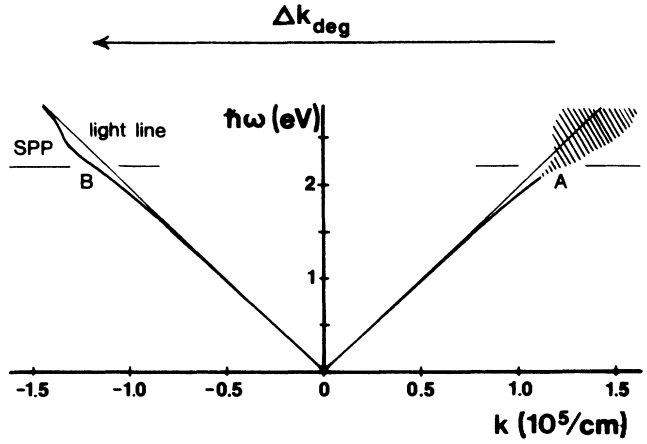

FIG. 14. The flat-surface SPP dispersion relation for gold in air. Light horizontal lines at $\hbar\omega\approx2.18$ eV locate the experimental conditions of Raether's experiments of Ref. 40. The shaded area shows the linewidth.

nonetheless possible to suggest a plausible interpretation of the data of Figs. 11(b) and 12(b). We suppose that at $h=0$, both $|\Delta k_2|$ and $|\Delta k_3|$ are either slightly too large or else are just beginning to mix the two degenerate states at $\hbar\omega=2.18$ eV, as shown in Fig. 15. As h increases, the individual excitations broaden further, and (for fixed $\hbar\omega$) move to larger values of $|k|$, until at $h=h_0$ the excitations have maximum overlap at 2.18 eV. Then $\Delta k_3=-\Delta k_{\text{deg}}$. The level may split, and small gap in either k or ω can occur.[56] Weber and Mills[57] showed that if a ω gap occurs, there can still be a minimum in the film reflectivity within the gap. In measurements that sweep θ at fixed ω, this minimum evolves continuously out of the minima that characterize the true SPP dispersion relation away from the gap. These are just the sorts of measurements done by Raether. However, within the gap, neither the position nor the linewidth of the reflection minimum is associated with the excitations themselves, which (at the point of degeneracy) lie slightly above and below the frequency of the incident light. (See Figs. 1–3 of Ref. 57.) Finally, as h increases further to h_1 in Fig. 15, both $|\Delta k_2|$ and $|\Delta k_3|$ are no longer large enough to couple the excitations at 2.18 eV. Then the degeneracy becomes irrelevant, at least to lower order.

In the reduced zone picture, the dispersion relation can be viewed as a pair of branches above and below $\hbar\omega=2.18$ eV, which have different characteristics.[45] When ω is successively fixed at increasing values, and the reflection minimum found by sweeping θ, the coordinates (k,ω) of the minimum move from the lower branch through the gap to the upper branch. In Raether's experiments ω was fixed, but the reflection surface is translated to smaller ω as h increases. If the topography of the reflection surface remains qualitatively unchanged, then the coordinates (k,ω) of the minimum, here also, move smoothly from the lower to the upper branch as h increases. In view of our uncertainty as to the correct values of D, h, and ϵ_{Au}, we cannot know for certain where (or even whether) the data of Figs. 11(b) and 12(b) exhibit

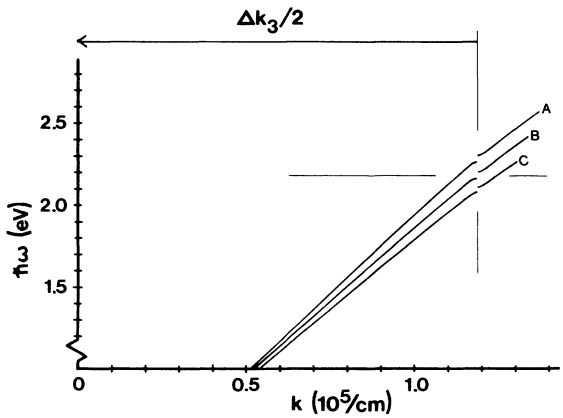

FIG. 15. Possible evolution of the gold-air SPP dispersion relation as corrugation depth increases. Curve A, $h=0$; curve B, $h=h_0$, at which the excitation coordinates coincide simultaneously with the photon frequency and the texture-related wavevector $\Delta k_3/2$; curve C, $h=h_1>h_0$.

these effects. It is possible, however, that the sudden change in slope in Fig. 11(b) occurs either at h_0, as the gap moves through the chosen energy $\hbar\omega=2.18$ eV ($\lambda_0=5682$ Å), or else at some smaller h at which the reflection minimum first departs from the SPP dispersion curve. And if the linewidth data of Fig. 12(b) were taken as $(k\omega)$ crossed the gap, then it has nothing to do with the theoretical curve, and no agreement should be expected.

C. Other perturbation theories

The Kroger-Kretschmann (KK) perturbation theory of rough-surface light scattering also depends on the Rayleigh hypothesis. Those authors began[54] by integrating Maxwell's equations from the true rough surface back to the plane of the average surface. Since the wave fields are well-behaved functions in space, this procedure is generally valid. They then expanded the fields in both media as Fourier series, and used these to express the Maxwell boundary conditions at the average surface. For randomly rough surfaces of small amplitude, they could then diagonalize the matrix that mixed the separate Fourier components.[58] Their specific interest was the dispersion relation for SPP's on a rough surface.

Although KK correctly accounted for the change in the surface normal as the roughness was "deformed" into a plane, their procedure is not quite the same as Rayleigh's, which matches the boundary fields at the true rough surface. But these authors did assume that the boundary conditions can be satisfied by the *same* Fourier series that describe the fields outside of the roughness; i.e., that these series converge everywhere in their respective half-spaces. This is the Rayleigh hypothesis. Pockrand and Raether[59] modified the last step of Ref. 58 to obtain a "two-beam" scattering formula suitable for a sinusoidal rather than stochastic surface. But in applying this formula, they separated real and imaginary parts as if the metal had a real dielectric constant; hence the comparison shown in Ref. 40 seems to have ignored $\text{Im}\epsilon_{\text{Au}}$ completely. Moreover, the authors normalized the theoretical formulas to the flat-surface values of θ and $\theta_{1/2}$ independently. We therefore cannot use their KK curves as an indicator of the validity of Rayleigh's hypothesis.

Finally, we note that the model of Ref. 46 is also based on Rayleigh's hypothesis. In fact, that work is a sort of "poor man's Rayleigh method," which used only the single Fourier component that dominates at a nonradiating mode resonance.

D. Other materials

Almost all the recent applications of Rayleigh's method have been to silver, perhaps because the little available data for another metal (i.e., gold) have not been satisfactorily described by Rayleigh-type methods (which was also remarked in Ref. 45). The present work suggests that the fault lies not with the Rayleigh hypothesis or method, but with the difficulty of accurately determining the geometry and material properties of the metals whose

optics are being studied.

Neither have there been applications of Rayleigh's method to absorbing insulators. In a previous work[60] we mistakenly claimed that Chuang and Johnson[61] used Waterman's original extinction theorem[2] to analyze their sound-scattering data. Toigo *et al.*[3] showed that this method is equivalent to the Rayleigh hypothesis. In fact, however, Ref. 61 used the extinction theorem of Ref. 3, which is exact.

VI. SUMMARY

We have combined the work of Millar with that of Hill and Celli to investigate the validity of the Rayleigh hypothesis for real materials. In doing this we implicitly complete the HC study of the Dirichlet problem for the purely sinusoidally corrugated surface. For that particular geometry we find that the Rayleigh hypothesis is valid for all real materials under the same conditions established by Millar (and by Hill and Celli) for the perfect conductor.

The perturbation theory of Maradudin was used to compare light-scattering data on corrugated gold and silver with calculations using Rayleigh's method. Agreement is better for silver than gold, but the discrepancies with gold appear to be caused by the use of auxiliary measurement techniques, which suppose very small absorption in the metal, and by degeneracy effects accidental to one of the gold films investigated.

APPENDIX A

Here we show that certain integrals occurring in Secs. I–III converge for real k when their limits are taken to infinity. The kernels that appear in Eqs. (2.1) and (3.14) are of five different kinds:

$$\frac{-(x_s-x_\tau)\dfrac{dz_s}{dx_s}+(z_s-z_\tau)}{\rho^2}(k_i\rho)H_1^{(1)}(k_i\rho)\ , \tag{A1}$$

$$H_0^{(1)}(k_i\rho)\ , \tag{A2}$$

$$\frac{(x_s-x_\tau)\dfrac{dz_\tau}{dx_\tau}-(z_s-z_\tau)}{\rho^2}(k_i\rho)H_1^{(1)}(k_i\rho)\ , \tag{A3}$$

$$\frac{(x_\tau-x_s)\dfrac{dz_s}{dx_s}-(z_\tau-z_s)}{\rho^2}\cdot\frac{(x_s-x_\tau)\dfrac{dz_\tau}{dx_\tau}-(z_s-z_\tau)}{\rho^2}(k_i\rho)^2H_0^{(1)}(k_i\rho)\ , \tag{A4}$$

$$\frac{\left[\dfrac{dz_s}{dx_s}\dfrac{dz_\tau}{dx_\tau}-1\right][(x_s-x_\tau)^2-(z_s-z_\tau)^2]-2\left[\dfrac{dz_s}{dx_s}+\dfrac{dz_\tau}{dx_\tau}\right](x_s-x_\tau)(z_s-z_\tau)}{\rho^4}(k_i\rho)H_1^{(1)}(k_i\rho)\ . \tag{A5}$$

Inspection of these functions reveals that (A1)–(A4) have the same limiting forms for large $|x_s-x_\tau|$. (A5) has a limiting form that contains an additional factor of (x_s-x_τ) in the denominator. Therefore, we need only show that an integral with the kernel (A1) is convergent, and the convergence of all the other integrals follows directly. We need this proof only for the solution of the boundary integral equations for the physical fields, before analytic continuation. Hence in (A1)–(A5) we take x_s, x_τ, k_1, and k_2 to be real. For definiteness we consider $x_s>x_\tau$.

We first note that

$$\frac{-(x_s-x_\tau)\dfrac{dz_s}{dx_s}+(z_s-z_\tau)}{\rho^2}\xrightarrow[x_s-x_\tau\to\infty]{}-\frac{1}{x_s-x_\tau}\left[\frac{dz_s}{dx_s}+O\left[\frac{1}{x_s-x_\tau}\right]\right] \tag{A6}$$

and that the Hankel functions have the limiting forms

$$(k_i\rho)H_1^{(1)}(k_i\rho)\xrightarrow[x_s-x_\tau\to\infty]{}C_1\frac{e^{ik_i(x_s-x_\tau)}}{\sqrt{x_s-x_\tau}}(x_s-x_\tau)\times\left[1+O\left[\frac{1}{x_s-x_\tau}\right]\right]\ , \tag{A7}$$

where C_1 is a constant of order unity. Combining (A6) and (A7), we obtain the asymptotic form of the kernel (A1):

$$K(k_i\rho)H_1^{(1)}(k_i\rho)=-C_1\frac{dz_s}{dx_s}\frac{e^{ik_i(x_s-x_\tau)}}{\sqrt{x_s-x_\tau}}\times\left[1+O\left[\frac{1}{x_s-x_\tau}\right]\right] \tag{A8}$$

for large (x_s-x_τ).

When the periodic surface extends to infinity in both directions, there are no "end effects;" then the functions

$L_1(x)$, $M_1(x)$, and products of these with functions of $\zeta(x)$ can all be written in the form

$$F(x_s)=\frac{1}{\sqrt{D}}\sum_{n=-\infty}^{\infty} f_n e^{i(K_0+n\kappa)(x_s-x_\tau)} \tag{A9}$$

for any fixed point x_τ. Here, K_0 is the x component of the incident wave vector as defined in (2.5), and κ is the "reciprocal lattice vector" defined in (1.2). Moreover, the f_n satisfy Parseval's equality

$$\sum_{n=-\infty}^{\infty}|f_n|^2=\int_0^D|F(x)|^2dx\equiv\mathcal{I}\ . \tag{A10}$$

Consider, for example, the first integral that occurs on the left of equation (3.14b). The contribution to that integral for $x_s-x_\tau>A$ can be written, using (A8) and (A9),

$$-\int_A^\infty L_1(x_s)K(k_1\rho)H_1^{(1)}(k_1\rho)=C_1\int_A^\infty\frac{d(x_s-x_\tau)}{\sqrt{x_s-x_\tau}}e^{ik_1(x_s-x_\tau)}\sum_{n=-\infty}^{\infty}f_n e^{i(K_0+n\kappa)(x_s-x_\tau)}\left[1+O\left[\frac{1}{x_s-x_\tau}\right]\right], \tag{A11}$$

where $F(x)=(dz_s/dx_s)L_1(x_s)$ in (A9). Because the "error term" in (A11) is bounded by $C_2/\sqrt{A}$, with $|C_2|=O(1)$, we need only to find an upper bound for

$$U\equiv\int_A^\infty dw\sum_{n=-\infty}^{\infty}\frac{e^{iw(K_0+n\kappa+k_1)}}{\sqrt{w}}f_n$$
$$=\sum_{n=-\infty}^{\infty}f_n\int_A^\infty\frac{e^{iw(K_0+n\kappa+k_1)}}{\sqrt{w}}dw\ . \tag{A12}$$

The integration and summation can be interchanged because the Fourier series (A9) converges uniformly over every interval on which $F(x)$ is continuous.

Now define $S_n\equiv K_0+n\kappa+k_i$. When $S_n=0$, the diffracted wave of index n propagates parallel to the surface. This is a Rayleigh-Woods anomaly,[62] in which we have no interest. Hence we can specify that $S_n\neq0$. Then

$$|U|=\left|\sum_{n=-\infty}^{\infty}f_n\int_A^\infty\frac{e^{iwS_n}}{\sqrt{w}}dw\right|$$
$$\le\left[\sum_{n=-\infty}^{\infty}|f_n|^2\right]^{1/2}\left[\sum_{n=-\infty}^{\infty}\left|\int_A^\infty e^{iwS_n}\frac{dw}{\sqrt{w}}\right|^2\right]^{1/2}$$
$$=\mathcal{I}^{1/2}\left[\sum_{n=-\infty}^{\infty}\left|\int_A^\infty e^{iwS_n}\frac{dw}{\sqrt{w}}\right|^2\right]^{1/2} \tag{A13}$$

using the Schwarz inequality and (A10). But

$$\int_A^\infty\frac{e^{iwS_n}}{\sqrt{w}}dw=\sqrt{\pi/-iS_n}\,\mathrm{erfc}\sqrt{-iAS_n}$$
$$=\frac{e^{iAS_n}}{-iS_n\sqrt{A}}\left[1+O\left[\frac{1}{AS_n}\right]\right] \tag{A14}$$

using formula (7.1.23) of Ref. 23. Putting (A14) into (A13), we obtain

$$|U|\le\mathcal{I}^{1/2}\left[\sum_{n=-\infty}^{\infty}\frac{1}{S_n^2A}\left|1+O\left[\frac{1}{AS_n}\right]\right|\right]^{1/2}$$
$$=\left[\frac{\mathcal{I}}{A}\right]^{1/2}\left[\sum_{n=-\infty}^{\infty}\frac{1}{S_n^2}\right]^{1/2}\left[1+O\left[\frac{1}{A}\right]\right]$$
$$\cong\frac{\pi}{\kappa}\left[\frac{\mathcal{I}}{A}\right]^{1/2}\csc\left[\pi\frac{K_0+k_1}{\kappa}\right]<\infty\ , \tag{A15}$$

where the last step used the identity[63]

$$\csc^2\pi x=\frac{1}{\pi^2}\sum_{n=-\infty}^{\infty}\frac{1}{(x-n)^2}\ .$$

Hence we find that the integrals of Secs. I–III that contain the kernels (A1)–(A5) are bounded even when k_1 and k_2 are real.

APPENDIX B

This section derives a very interesting result of Millar, who gave it without details in Ref. 20. The context into which it fits is given in Ref. 19, to which we refer the reader for particulars. The result to be demonstrated was central to his proof of criterion (2.17) above, and is closely related to that aspect of the present work established in Appendix A. We are able to begin with expression (3) of Ref. 19 and show that it leads to Eq. (9) of Ref. 20, in a self-contained development that uses Millar's notation with only minor changes for clarity.

Figure 16(a) shows Millar's geometry. His principal coordinate system was offset from the origin 0 for reasons that are not important to this calculation; he himself intended to show that the contour S_N could be ignored for any values of the offset coordinates (R_0,θ_0). For simplicity we therefore take $R_0=h$, $\theta_0=\pi/2$, to obtain the coordinate system in Fig. 16(b), from which we work. Figure 16(c) will be explained shortly.

Millar set out to locate the singularities of the scattered wave, denoted here by

$$\Phi(u,v)=\sum_{m=-\infty}^{\infty}a_m e^{-i(k\cos\alpha+m\kappa)u+i\kappa_m v}\ , \tag{B1}$$

where

$$\kappa_m\equiv[k^2-(k\cos\alpha+m\kappa)^2]^{1/2};\quad\kappa\equiv\frac{2\pi}{D}\ . \tag{B2}$$

Here, k is the incident wave vector and α the angle of incidence measured from the u axis. [Our Eq. (B1) is Eq. (6) of Ref. 20; Φ is defined as the scattered field above Eq. (4) of that work, and above Eq. (2) of Ref. 19.]

In connection with (B2), it is convenient to define

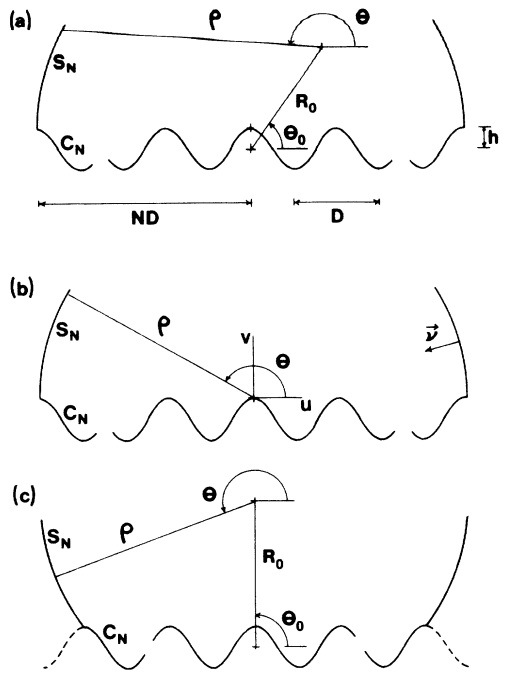

FIG. 16. (a) Geometry used by Millar in his investigation of the Dirichlet problem for the infinite periodic surface. (b) A special case of Millar's geometry: $R_0=h$, $\theta_0=\frac{1}{2}\pi$. (c) The same geometry for large R_0.

$$k\cos\alpha+m\kappa\equiv k\cos B_m\ ,$$
$$\kappa_m\equiv k\sin B_m\ . \tag{B3}$$

The B_m are diffraction angles for Bragg wavelets. κ_m and B_m are positive real for propagating wavelets. κ_m is positive imaginary for evanescent wavelets. It will clarify what follows to have a picture of how the B_m are situated in the complex plane. Setting

$$B_m\equiv B_{m1}+iB_{m2}$$
$$\frac{\kappa_m}{k}=\sin B_{m1}\cosh B_{m2}+i\cos B_{m1}\sinh B_{m2} \tag{B4}$$

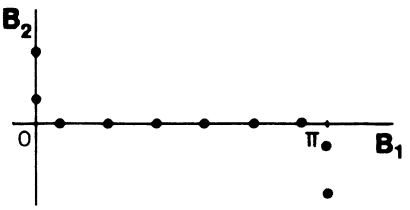

FIG. 17. Distribution of Bragg angles in the complex plane.

we see that when κ_m is real, $0<B_{m1}<\pi$; the imaginary B_m must be located as shown in Fig. 17, in order that κ_m be positive imaginary for evanescent waves.

Millar's approach required him to prove that

$$I\equiv\int_{S_N}\left[\Phi\frac{\partial}{\partial\nu}-\frac{\partial\Phi}{\partial\nu}\right]H_n^{(1)}(k\rho)e^{-in\theta}ds\xrightarrow[N\to\infty]{}O[N^{-1/2}], \tag{B5}$$

where $\partial/\partial\nu=-\partial/\partial\rho$ is the normal derivative, N is a number of periods as shown in Fig. 1(a), and $H_n^{(1)}(k\rho)$ is the Hankel function. n is an integer that labels a Fourier-Bessel coefficient for which Millar wished to find an estimate; here it is a fixed parameter. He terminated S_N at each end on one of the peaks of C_N, so that S_N lay entirely outside the so-called "selvedge" region. Therefore (B1) converges everywhere on S_N. The variable s in (B5) is arclength.

Along S_N, $u=ND\cos\theta$ and $v=ND\sin\theta$; for sufficiently large N and fixed n,

$$H_n^{(1)}(k\rho)\cong\left[\frac{2}{\pi kND}\right]^{1/2}e^{i(kND-n\pi/2-\pi/4)}\ . \tag{B6}$$

With these definitions, (B1), (B2), (B3), and (B6) are put into (B5) to obtain

$$I=\int_0^{\pi}ND\,d\theta\sum_{m=-\infty}^{\infty}a_m e^{iND(-k\cos B_m\cos\theta+k\sin B_m\sin\theta)}e^{-in\theta}$$
$$\times\left[\frac{2}{\pi kND}\right]^{1/2}e^{i(kND-n\pi/2-\pi/4)}(-ik-ik\cos B_m\cos\theta+ik\sin B_m\sin\theta)$$
$$=i\left[\frac{2}{\pi k}\right]^{1/2}(ND)^{1/2}e^{i(kND-n\pi/2-\pi/4)}\sum_{m=-\infty}^{\infty}a_m\int_0^{\pi}e^{NDf_m(\theta)}g_m(\theta)d\theta\left[1+O\left[\frac{1}{N}\right]\right], \tag{B7}$$

where

$$f_m(\theta)=-i(k\cos\alpha+m\kappa)\cos\theta+\kappa_m\sin\theta$$
$$=-ik\cos(\theta+B_m) \tag{B8a}$$

and

$$g_m(\theta)=-k-(k\cos\alpha+m\kappa)\cos\theta+\kappa_m\sin\theta$$
$$=-k+\frac{f_m(\theta)}{i}\ . \tag{B8b}$$

The asymptotic behavior of I for large N is determined by

the integrals that appear in Eq. (B7):

$$I_m \equiv \int_0^{\pi} e^{NDf_m(\theta)} g_m(\theta) d\theta \tag{B9}$$

along with the factor $N^{1/2}$ in (B7). I_m in turn is determined by the behavior of $f_m(\theta)$, which has stationary points at θ_{SP} given by $f_m'(\theta_{SP})=0$, or

$$\sin(\theta_{SP}+B_m)=0 \ . \tag{B10}$$

We will now evaluate the dominant part of $I_m(N)$ for all m.

1. κ_m imaginary

For definiteness we take m such that (B4) becomes

$$B_{m1}=0; \quad B_{m2}=\sinh^{-1}\left[\frac{|\kappa_m|}{k}\right] \tag{B11}$$

for which the solutions to (B10) are $\theta_{SP}\equiv\theta_{SP1}+i\theta_{SP2}$ given by

$$\theta_{SP1}=0,\pi; \quad \theta_{SP2}=-\sinh^{-1}\left[\frac{|\kappa_m|}{k}\right] \tag{B12}$$

and (B8a) becomes

$$f_m(\theta)=-ik\left\{\cos\theta_1\cosh\left[\theta_2+\sinh^{-1}\left[\frac{|K_m|}{k}\right]\right]\right.$$
$$\left.-i\sin\theta_1\sinh\left[\theta_2+\sinh^{-1}\left[\frac{|\kappa_m|}{k}\right]\right]\right\}. \tag{B13}$$

The saddle points, as well as the contours of $\mathrm{Re}f_m$ and $\mathrm{Im}f_m$, are shown in Fig. 18(a). It is possible to deform the integration path as in Fig. 18(b) and generate a complete asymptotic expansion for $I_m(N)$. For the leading term, the method of steepest descents[64] is much more convenient. The appropriately deformed path is shown in Fig. 18(c). Along segments C_1 and C_2, $\mathrm{Im}f_m$ is constant and $\mathrm{Re}f_m$ decreases rapidly. Along segment C_3 the integrand is exponentially small. The important feature here is that the deformed contour does not terminate on the saddle points: therefore, the first derivative f' does not vanish at $\theta=0$ and $\theta=\pi$. Hence, the leading contribution to I_m from C_1 is

$$\int_{C_1} e^{NDf_m(\theta)} g_m(\theta) d\theta \approx \int_0^{\theta_a} e^{ND[f_m(0)+\theta f_m'(0)]} g_m(0) d\theta$$
$$=\frac{k(-1-\cos B_m)}{ND|\kappa_m|} e^{-iNDk\cos B_m}$$
$$\times\int_0^{ND|\kappa_m|\theta_a} e^{-w} dw \tag{B14a}$$

using (B3) and (B11) to simplify. Since $\mathrm{Re}w>0$ along C_1,

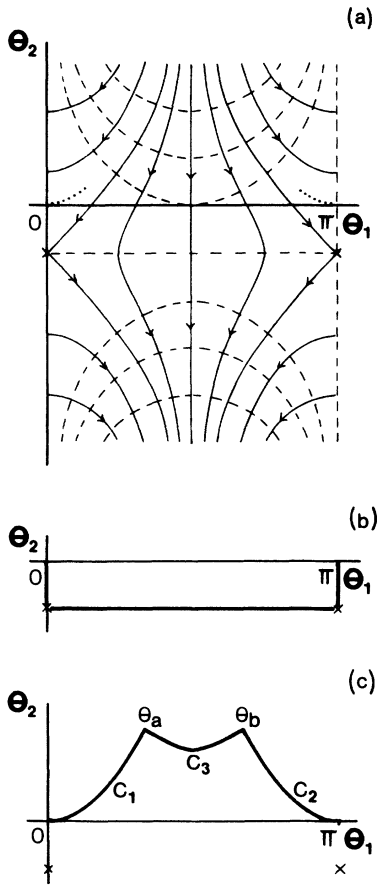

FIG. 18. Geometry of the complex plane for κ_m imaginary. (a) Contours of constant $\mathrm{Im}f_m(\theta)$ (solid lines) and constant $\mathrm{Re}f_m(\theta)$ (dashed lines). Dots show paths of steepest descent from $\theta=0$ and $\theta=\pi$. (b) Deformation used to find an asymptotic series for I_m by stationary phase calculation. (c) Deformation suitable for the method of steepest descents.

the remaining integral converges. In the same way the contribution from C_2 is

$$\int_{C_2} e^{NDf_m(\theta)} g_m(\theta) d\theta \approx \frac{k(-1+\cos B_m)}{ND|\kappa_m|} e^{iNDk\cos B_m}$$
$$\times\int_0^{ND|\kappa_m|(\pi-\theta_b)} e^{-w} dw \ , \tag{B14b}$$

where by (B11), $\cos B_m=(1+|\kappa_m|^2/k^2)^{1/2}$. So for κ_m imaginary, $I_m(N)=O[N^{-1}]$: These I_m contribute to I only to $O[N^{-1/2}]$ as Millar found.

2. κ_m real

In this case (B4) becomes

$$B_{m2}=0; \quad B_{m1}=\sin^{-1}\left[\frac{\kappa_m}{k}\right] \tag{B15}$$

for which the solutions of (B10) are

$$\theta_{SP1}=\pi-\sin^{-1}\left[\frac{\kappa_m}{k}\right];\quad \theta_{SP2}=0 \tag{B16}$$

and (B8a) becomes

$$f_m(\theta)=-ik\left\{\cos\left[\theta_1+\sin^{-1}\left[\frac{\kappa_m}{k}\right]\right]\cosh\theta_2 - i\sin\left[\theta_1+\sin^{-1}\left[\frac{\kappa_m}{k}\right]\right]\sinh\theta_2\right\}. \tag{B17}$$

The corresponding contours are shown in Fig. 19. Comparison of (B13) and (B17) shows that these contours are translations of those in Fig. 18(a). From Eq. (B10) it is clear that the saddle point with $\theta_{SP1}=\pi$ in Fig. 18(a) has moved up and to the left to take its place in Fig. 19. Figure 17 helps to visualize this, and to find the correct root (B16) of (B10).

In this case the method of steepest descents is not available, so we find an asymptotic expansion for $I_m(N)$. This is first expressed in the form

$$I_m(N)=\left[-k+\frac{1}{i}\frac{d}{d\gamma}\right]\int_0^{\pi}e^{\gamma f_m(\theta)}d\theta\Bigg|_{\gamma=ND} \tag{B18}$$

using Eq. (B8b). Now we need to find the simpler integral,

$$I_1\equiv\int_0^{\pi}e^{\gamma f_m(\theta)}d\theta=\int_{-B_1}^{\pi-B_1}dt\,e^{ik\gamma\cos t}, \tag{B19}$$

where we have used (B8a) for $f_m(\theta)$, then substituted $t\equiv\pi-(\theta+B_{m1})$. Now set $\cos t\equiv 1-w^2$ to obtain

$$I_1=\int_{-b_1}^{b_2}dw\frac{2e^{ik\gamma(1-w^2)}}{(2-w^2)^{1/2}} =\sqrt{2}e^{ik\gamma}\sum_{l=0}^{\infty}\left[\frac{-1}{2i}\right]c_l\frac{d^l}{d(k\gamma)^l}\int_{-b_1}^{b_2}e^{-ik\gamma w^2}dw. \tag{B20}$$

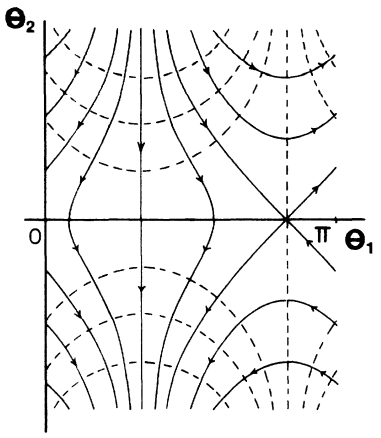

FIG. 19. Geometry of the complex plane for κ_m real.

Equation (B20) is still exact. On the right, the c_l are the coefficients in

$$\frac{1}{\sqrt{1-x}}=\sum_{l=0}^{\infty}c_l x^l \tag{B21}$$

and we have written

$$b_1\equiv\sqrt{1-\cos B_{m1}};\quad b_2\equiv\sqrt{1+\cos B_{m1}}. \tag{B22}$$

The remaining integral in (B20) is now written as a sum of error functions of complex argument. Because these have exponential factors in their asymptotic forms, all the derivatives with respect to $k\gamma$ must be retained in (B20). They produce a simple algebraic power series in the final result. To evaluate (B20), we need only the well-known Fresnel integrals and formula (7.1.23) of Ref. 23. Keeping only the leading terms, the result is

$$\int_{-b_1}^{b_2}e^{-ik\gamma w^2}dw\approx\frac{1}{\sqrt{k\gamma}}\left[\sqrt{\pi}e^{-i\pi/4}+\frac{i}{2\sqrt{k\gamma}}\left[\frac{e^{-ib_1^2k\gamma}}{b_1}+\frac{e^{-b_2^2k\gamma}}{b_2}\right]\right]. \tag{B23}$$

When (B23) is put into (B20), we obtain to leading order in γ

$$I_1=\sqrt{2}e^{ik\gamma}\left[\left[\frac{\pi}{k\gamma}\right]^{1/2}e^{-i\pi/4}+\frac{i}{2k\gamma}\left[\frac{e^{-ib_1^2k\gamma}}{b_1(1-b_1^2/2)^{1/2}}+\frac{e^{-ib_2^2k\gamma}}{b_2(1-b_2^2/2)^{1/2}}\right]\right] =\left[\frac{2\pi}{k\gamma}\right]^{1/2}e^{i(k\gamma-\pi/4)}+\frac{2i\cos(k\gamma\cos B_{m1})}{k\gamma\sin B_{m1}} \tag{B24}$$

where all the derivatives have been removed using (B21). The first term in (B24) is just the asymptotic form of the outgoing Hankel function. When (B24) is put into (B18) to finish this calculation, the leading term is $O(N^{-1})$, so that these I_m also contribute to I only to order $N^{-1/2}$, as Millar found.

However, in Fig. 19 there is another saddle point to the left of the origin, whose leading contribution contains an

incoming Hankel function. This is a property of the plane-wave expansion (B1) used for the scattered field. If the offset coordinate R_0 in Fig. 16(a) is large enough, then S_N in Eq. (B5) encloses an angle greater than π as in Fig. 16(c). In particular, if R_0 is such that $|\sin\theta| > \sin B_{\max}$ anywhere on S_N, where $B_{\max}$ is the Bragg angle of the most extreme diffracted order, then the leading term in I is of order unity, not $N^{-1/2}$. Then I cannot be neglected as Millar's proof required. Such an inconvenience survives only if $R_0/(ND)$ is kept constant as $N\to\infty$. Since this particular limiting process would not ordinarily be used. Millar could in fact conclude that I is "$O(N^{-1/2})$ uniformly with (R_0,θ_0)."[20]

Millar assumed at the outset that his surface was infinitely periodic, then applied Green's identity to a part of it. In this way he eliminated "end effects." Without the strict periodicity, a more-detailed (and difficult) analysis of the fields on S_N would have been needed. Likewise, the result the Appendix A required strict periodicity of the surface. Without making that assumption, we could not have obtained the extinction theorem, the boundary integral equations, and the necessary analytic continuations of the surface fields from one and the same progression of approximations.

APPENDIX C

Here we derive an analytic continuation of a Hankel function needed in Sec. IV A. The commonly used asymptotic expansions for Hankel functions of fixed order and large argument are given by formulas (9.2.3) and (9.2.4) of Ref. 23:

$$H_\nu^{(1)}(z)\sim\left[\frac{2}{\pi z}\right]^{1/2} e^{i(z-\nu\pi/2-\pi/4)} \quad (-\pi<\arg z<2\pi)\ , \tag{C1}$$

$$H_\nu^{(2)}(z)\sim\left[\frac{2}{\pi z}\right]^{1/2} e^{-i(z-\nu\pi/2-\pi/4)} \quad (-2\pi<\arg z<\pi)\ . \tag{C2}$$

But Fig. 4(c) shows that when $\arg k_i>0$, then $\arg(k_i\rho)>\pi$, in which case (C2) does not hold. We therefore have to relate $H^{(2)}(k_i\rho)$ to $H^{(2)}(k_i\rho_A)$, where

$$\rho\equiv\rho_A e^{2\pi i} \quad (x_s\to-\infty)\ . \tag{C3}$$

The limiting form (C2) is valid for $z=k_i\rho_A$.

To do this we need formula (9.1.38) of Ref. 23:

$$\sin(\nu\pi)H_\nu^{(2)}(ze^{m\pi i})=\sin[(m+1)\nu\pi]H_\nu^{(2)}(z) +e^{\nu\pi i}\sin(m\nu\pi)H_\nu^{(1)}(z)\ . \tag{C4}$$

Let $z=k_i\rho_A$ and $m=2$ from Eq. (C3), to obtain

$$H_\nu^{(2)}(k_i\rho)=\frac{\sin(3\nu\pi)}{\sin(\nu\pi)}H_\nu^{(2)}(k_i\rho_A) +\frac{\sin(2\nu\pi)}{\sin(\nu\pi)}e^{\nu\pi i}H_\nu^{(1)}(k_i\rho_A)\ . \tag{C4a}$$

Taking the limit as $\nu\to0$, we find

$$H_0^{(2)}(k_i\rho)=3H_0^{(2)}(k_i\rho_A)+2H_0^{(1)}(k_i\rho_A)\ . \tag{C5}$$

Because $-\pi<\arg(k_i\rho_A)<\pi$, Eq. (C2) now applies, so that (C5) becomes

$$H_0^{(2)}(k_i\rho)\xrightarrow[|k_i\rho|\to\infty]{} 3\left[\frac{2}{\pi k_i\rho_A}\right]^{1/2}e^{-i(k_i\rho_A-\pi/4)}+2\left[\frac{2}{\pi k_i\rho_A}\right]^{1/2}e^{i(k_i\rho-\pi/4)} =-3\left[\frac{2}{\pi k_i\rho}\right]^{1/2}e^{-i(k_i\rho-\pi/4)}-2\left[\frac{2}{\pi k_i\rho}\right]^{1/2}e^{i(k_i\rho-\pi/4)}\ , \tag{C6}$$

where (C3) was used in the last line. When (C6) is used in Eq. (4.6), we get the asymptotic form of $H_0^{(1)}(k_i|\rho|)$ in the third quadrant of Fig. 4(c):

$$H_0^{(1)}(k_i|\rho|)=2H_0^{(1)}(k_i\rho)+H_0^{(2)}(k_i\rho) \quad (x_s\to-\infty) \xrightarrow[|\rho|\to\infty]{} -3\left[\frac{2}{\pi k_i\rho}\right]^{1/2}e^{-i(k_i\rho-\pi/4)}\ . \tag{C7}$$

In the same way we take the limit of (C4a) as $\nu\to1$ to get

$$H_1^{(2)}(k_i\rho)=3H_1^{(2)}(k_i\rho_A)+2H_1^{(1)}(k_i\rho_A)\ , \tag{C8}$$

and when this is put into (4.7), and (C3) is used once again, the result is

$$k_i|\rho|H_1^{(1)}(k_i|\rho|)\xrightarrow[|\rho|\to\infty]{} -3\left[\frac{2k_i\rho}{\pi}\right]^{1/2}e^{-i(k_i\rho-3\pi/4)} \quad (x_s\to-\infty)\ . \tag{C9}$$

[1]J. W. S. Rayleigh, *The Theory of Sound* (MacMillan, London, 1895).
[2]P. C. Waterman, J. Acoust. Soc. Am. **57**, 791 (1975).
[3]F. Toigo, A. Marvin, V. Celli, and N. R. Hill, Phys. Rev. B **15**, 5618 (1977).
[4]A. Wirgin, Opt. Commun. **27**, 189 (1978).
[5]R. Petit, Nouv. Rev. Opt. **6**, 129 (1975).
[6]K. A. Zaki and A. R. Neureuther, IEEE Trans. Antennas Propag. **AP-19**, 208 (1971); **AP-19**, 747 (1971).
[7]N. E. Glass and A. A. Maradudin, Phys. Rev. B **24**, 595 (1981).

[8]N. E. Glass, R. Loudon, and A. A. Maradudin, Phys. Rev. B **24**, 6843 (1981); N. E. Glass, A. A. Maradudin, and V. Celli, *ibid.* **26**, 5357 (1982); N. E. Glass and A. A. Maradudin, J. Appl. Phys. **54**, 796 (1983); N. E. Glass, A. A. Maradudin, and V. Celli, J. Opt. Soc. Am. **73**, 1240 (1983).

[9]M. G. Cavalcante, G. A. Farias, and A. A. Maradudin, J. Opt. Soc. Am. B **4**, 1372 (1987).

[10]S. Dutta Gupta, G. V. Varada, and G. S. Agarwal, Phys. Rev. B **36**, 6331 (1987).

[11]H. J. Simon, C. Huang, J. C. Quail, and Z. Chen, Phys. Rev. B **38**, 7408 (1988).

[12]See, for example, A. Wirgin, J. Acoust. Soc. Am. **68**, 692 (1980).

[13]A. A. Maradudin, J. Opt. Soc. Am. **73**, 759 (1983); E. P. da Silva, G. A. Farias, and A. A. Maradudin, *ibid.* A **4**, 2022 (1987).

[14]J. J. Greffet, Phys. Rev. B **37**, 6436 (1988).

[15]R. Petit and M. Cadilhac, C. R. Acad. Sci. Paris Ser. A and B **262**, 468 (1966).

[17]R. F. Millar, SIAM J. Math. Anal. **1**, 333 (1970).

[18]R. F. Millar, SIAM J. Math. Anal. **1**, 345 (1970).

[19]R. F. Millar, Proc. Cambridge Philos. Soc. **69**, 175 (1971).

[20]R. F. Millar, Proc. Cambridge Philos. Soc. **69**, 217 (1971).

[21]N. R. Hill and V. Celli, Phys. Rev. B **17**, 2478 (1978).

[22]A comprehensive and useful recent review of Rayleigh's hypothesis for the Dirichlet problem was written by B. V. Apelt'sin and A. G. Kyurkchan, Sov. J. Commun. Tech. Elec. **30**, 97 (1985).

[23]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).

[24]J. F. Carlson and A. E. Heins, Q. J. Appl. Math. **4**, 313 (1947).

[25]See, for example, J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941); B. B. Baker and E. T. Copson, *The Mathematical Theory of Huyghens' Principle* (Clarendon, Oxford, 1950); A. E. Heins and S. Silver, Proc. Cambridge Philos. Soc. **51**, 149 (1955); A. Wirgin, Rev. Opt. **44**, 20 (1965).

[26]J. B. Keller, IRE Trans. Antennas Propag. **AP-4**, 312 (1956); N. D. Kazarinoff and R. K. Ritt, Ann. Phys. (N.Y.) **6**, 277 (1959); for the opposite extreme, see R. F. Goodrich and N. D. Kazarinoff, Proc. Cambridge Philos. Soc. **59**, 167 (1963). (1963).

[27]D. Ludwig, Commun. Pure Appl. Math. **20**, 103 (1967); this work is related to the preceding reference 26 by R. M. Lewis, N. Bleistein and D. Ludwig, Commun. Pure Appl. Math. **20**, 295 (1967).

[28]R. L. Holford, J. Acoust. Soc. Am. **70**, 1116 (1981).

[29]J. L. Uretsky, Ann. Phys. (N.Y.) **33**, 400 (1965).

[30]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics, Volume I* (McGraw-Hill, New York, 1953).

[31]W. C. Meecham, J. Rat. Mechn. Anal. **5**, 323 (1956).

[32]R. Kress and G. F. Roach, J. Math. Phys. **19**, 1433 (1978).

[33]*Integral Equations–A Reference Text* edited by P. P. Zabreyko *et al.* (Noordhoff, Leyden, 1975).

[34]R. Courant and D. Hilbert, *Methods of Mathematical Physics, Volume II* (Interscience, New York, 1962).

[35]P. R. Garabedian, *Partial Differential Equations* (Wiley, New York, 1964), pp. 136–145. Garabedian assumed that the elliptic integral equation was real, but used that fact only to draw conclusions about the values of the solution for real values of the independent variable. The analyticity argument proceeds as well for the Helmholtz equation with complex k^2. The argument can, but does not have to be, rephrased using the standard generalization of the adjoint differential operator to the case of equations with complex coefficients. See P. Dennery and A. Krzywicki, *Mathematics for Physicists* (Harper and Row, New York, 1967), Chap. IV.

[36]Millar has used the complex plane for both purposes simultaneously in some of his studies, requiring extra concentration by the reader.

[37]B. A. Fuchs and B. V. Shabat, *Functions of a Complex Variable, Volume I* (Pergamon, New York, 1964).

[38]A helpful treatment of Hankel and Bessel functions is found in R. Courant and D. Hilbert, *Methods of Mathematical Physics, Volume I* (Interscience, New York, 1953).

[39]W. Pogorzelski, *Integral Equations and Their Applications, Volume I* (Pergamon, New York, 1966), pp. 6–13 and 182–185.

[40]H. Raether, Opt. Commun. **42**, 217 (1982).

[41]A. E. Craig, G. A. Olson, and D. Sarid, Opt. Lett. **8**, 380 (1983).

[42]R. J. Bell, R. W. Alexander, Jr., W. F. Parks, and G. Kovener, Opt. Commun. **8**, 147 (1973).

[43]G. R. Jiracek, IEEE Trans. Antennas Propag. **AP-21**, 393 (1973).

[44]A. Wirgin, J. Opt. Soc. Am. **72**, 963 (1982).

[45]N. E. Glass, M. Weber, and D. L. Mills, Phys. Rev. B **29**, 6548 (1984).

[46]T. C. Paulick, J. Appl. Phys. **64**, 1384 (1988).

[47]M. G. Weber, Phys. Rev. B **33**, 909 (1986).

[48]E. Kretschmann, Z. Phys. **241**, 313 (1971).

[49]See, for example, J. H. Weaver, C. Krafka, D. W. Lynch, and E. E. Koch, Appl. Opt. **20**, 1125 (1981); H. A. MacLeod, J. Vac. Sci. Technol. A **4**(3), 418 (1986).

[50]P. Beckmann and A. Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces* (MacMillan, New York, 1963), Chaps. 3 and 4.

[51]I. Pockrand and H. Raether, Opt. Commun. **18**, 395 (1976).

[52]D. Heitmann and V. Permien, Opt. Commun. **23**, 131 (1977).

[53]D. Heitmann, Opt. Commun. **20**, 292 (1977).

[54]E. Kroger and E. Kretschmann, Z. Phys. **237**, 1 (1970); E. Kretschmann and E. Kroger, J. Opt. Soc. Am. **65**, 150 (1975).

[55]P. B. Johnson and R. W. Christy, Phys. Rev. B **6**, 4370 (1972).

[56]Y. J. Chen, E. S. Koteles, R. J. Seymour, G. I. Sonek, and J. M. Ballantyne, Solid State Commun. **46**, 95 (1983); D. Heitmann, N. Kroo, C. Schulz, and Zs. Szentirmay, Phys. Rev. B **35**, 2660 (1987); V. Celli. P. Tran, A. A. Maradudin, and D. L. Mills, *ibid.* **37**, 9089 (1988).

[57]M. G. Weber and D. L. Mills, Phys. Rev. B **34**, 2893 (1986).

[58]E. Kroger and E. Kretschmann, Phys. Status Solidi B **76**, 515 (1976). The reader is cautioned that this paper contains so many typographical errors that the mathematics has to be totally reworked by the user.

[59]I. Pockrand and H. Raether, Appl. Opt. **16**, 1784 (1977).

[60]T. C. Paulick, J. Appl. Phys. **62**, 3016 (1987).

[61]S. L. Chuang and R. K. Johnson, J. Acoust. Soc. Am. **71**, 1368 (1982).

[62]C. H. Palmer, J. Opt. Soc. Am. **42**, 269 (1952).

[63]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic, New York, 1980), p. 36.

[64]G. F. Carrier, M. Krook, and C. E. Pearson, *Functions of A Complex Variable* (McGraw-Hill, New York, 1966).