Perturbation theory for light diffraction with surface-polariton resonances on a bigrating

N. E. Glass
Department of Physics, Naval Postgraduate School, Monterey, California 93943-5100
(Received 7 July 1986)

The diffraction of light of arbitrary polarization and arbitrary plane of incidence from a bigrating surface is described by an existing theory based on Rayleigh's method and the vectorial Kirchhoff integral. Its implementation requires solving large matrix equations (e.g., 162×162). This paper presents an approximation of that theory, to first order in the corrugation amplitude. It yields a 4×4 matrix equation—allowing one to determine the reflectance for any diffracted order, as well as the surface-field enhancement, when there are surface-polariton resonant excitations. The theory retains all the nonresonant amplitudes (to renormalize the surface-polariton frequency) and is valid directly at—or away from—the boundaries of the two-dimensional Brillouin zones defined by the surface periodicity. The complex dispersion relation, for a surface polariton which propagates in an arbitrary direction across the bigrating, is given by a simple algebraic expression in this perturbation theory.

I. INTRODUCTION

Studies of optical interactions, with resonant coupling to surface polaritons, at metallic bigrating surfaces (i.e., those periodic along two directions) have been proceeding in pace with similar studies on classical gratings (those periodic along just one direction), and for many of the same reasons. These reasons include the possibility, which gratings provide, to perform experiments on well-characterized rough surfaces and the possibility to create exact (i.e., nonperturbative) theories—to study, for example, the electromagnetic mechanism, via nonlocalized modes, for the "surface enhancement phenomena" (e.g., surface-enhanced Raman scattering). The grating is thus used, to some degree, as a substitute for random roughness, which is difficult to characterize experimentally and difficult to treat theoretically. The grating is important in its own right in practical applications: for optical couplers, Bragg reflectors, and solar absorbers.

In many instances there are advantages in employing the bigrating over the classical grating. The bigrating can couple incident light of arbitrary polarization and azimuthal orientation to surface polaritons, making for more efficient absorbers of unpolarized (e.g., solar) light. The possibility of simultaneous resonant excitation of two surface polaritons, at the boundaries of the two-dimensional Brillouin zones, also yields an enhanced absorptance. Moreover, having corrugations in two directions allows one to model true surface bumps of arbitrary shape, as, for example, the square array of hemiellipsoidal bumps studied in Ref. 1 and depicted in Fig. 1. That extra degree of freedom in modulating the surface profile can be a very important factor in using the grating as a test of theories of random roughness. Present theories treat the randomly rough surface perturbatively; there are no rigorously exact theories for testing—by way of comparison—the limits of validity of the perturbation theories. Analogies can be drawn, however, to the limits of validity of perturbation theories for gratings, as tested against the exact grating theories. The surface profile provided by the classical grating, unfortunately, contains in its Fourier decomposition no cross terms between the two surface directions: there are no contributions from terms such as $\cos(2\pi x_1/a)\cos(2\pi x_2/a)$, as there must be on the randomly rough surface. There has already been some indication provided by theoretical studies on bigratings that perturbation theory may, in fact, not perform as well where these cross terms are concerned.[2] No such statement, one way or the other, is possible when the purely classical grating is used for comparison.

Several theoretical approaches have been taken to describe nonperturbatively optical diffraction from a bi-

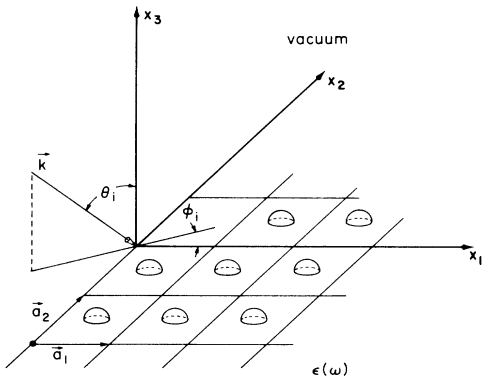

FIG. 1. A bigrating, with translation vectors, $\mathbf{a}_1$ and $\mathbf{a}_2$, on a planar interface, at $x_3=0$, between vacuum and the medium of dielectric constant $\epsilon(\omega)$. Light of wave vector $\mathbf{k}$ is incident at angle θ with respect to the surface-normal and at ϕ with respect to the x_1 axis of the bigrating.

grating with surface-polariton excitation; all involve large-scale computing in their numerical implementation.[1,3–6] For example, the method of Glass, Maradudin, and Celli[1] requires solving matrix equations of dimensions from 98×98 up to 242×242 in order to achieve convergence of 1% to 5% for corrugation strengths (ratio of maximum height to period) of 0.02 to 0.14.

As for experimental work, a recent study of the coupling of incident light to surface polaritons on a square bigrating was carried out by Inagaki *et al.* using a photoacoustical method.[7]

The time-consuming length of the numerical calculations required to implement the exact theories of light scattering from gratings shows very clearly the need for fast and accurate perturbation theories. This is especially true in the case of bigratings, for which the added corrugation dimension necessitates much larger matrix equations than for classical gratings.

Several perturbation theories have been used to study periodic surface roughness. The perturbation theory of Kroger and Kretschmann[8] has been applied quite successfully for describing the position and width of the reflectivity dips, corresponding to surface-polariton resonances, in the case of optical scattering from classical gratings on Ag.[9–12] Also for the classical grating, and appearing at about the same time, were the perturbation theories of Toigo *et al.*[13] and of Mills[14] (both using a two-beam approximation near zone boundaries). The Mills theory allows for an arbitrary angle between the grating and the direction of the surface-polariton propagation, yielding either the dispersion relation or diffracted intensities. Another perturbation theory, that of Elson and Sung,[15] yields the total absorptance on a rough surface by evaluating, to first order in the roughness, the integral of the induced current times electric field. This theory was applied by Inagaki *et al.*[7] to fit their own experimental work for total absorptance on a bigrating.

The perturbation theory of Glass, Weber, and Mills[16] provides a simple algebraic equation for determining the complex frequency $\omega_R+i\omega_I$ as a function of wave vector for surface polaritons propagating in a direction normal to the grooves of a classical grating (thus yielding both the dispersion curves and the lifetimes of the surface polaritons). The theory was specifically cast so as to be valid directly at—as well as away from—the minigaps at the boundaries of the Brillouin zones associated with the grating periodicity. The method, moreover, can be extended to yield reflectivities as well as the dispersion relations.[17]

The general perturbative technique employed by Glass, Weber, and Mills for a classical grating will be applied here to the exact theory of Glass, Maradudin, and Celli for a bigrating, even though the point of departure, the exact theory, in the two cases is quite different. The work of Glass, Weber, and Mills begins with the exact set of linear algebraic equations for the transverse magnetic field and its derivative, which arise from use of the extinction theorem, for a p-polarized wave propagating in a direction normal to the grating grooves. By contrast, the work of Glass, Maradudin, and Celli develops an "exact" set of linear algebraic equations for the p and s components of the Rayleigh coefficients for the electric field, which arises from applying the Rayleigh hypothesis to the vectorial Kirchhoff integral, for arbitrary polarization and orientation of the incident light.

The present application of the Glass-Weber-Mills technique to the Glass-Maradudin-Celli equations allows one to determine the reflectance of each diffracted order, as well as the complex amplitudes of the evanescent waves, and thus to determine the total electric field and the field enhancement in the vacuum above the selvedge region. For determining the fields, one is of course limited by use of the Rayleigh hypothesis to this region outside of the grooves: the hypothesis is not valid within the grooves (see, for example, Mills and Weber[18]). Additionally, a simple algebraic equation is obtained, whose complex solution will directly yield the dispersion curves and lifetimes for surface polaritons propagating in any direction on the bigrating. The theory, moreover, is valid directly on the boundaries of the two-dimensional Brillouin zone. It is thus useful for studying the absorptance peaks which correspond to the simultaneous excitation of two surface polaritons propagating in noncollinear directions (as first measured and described by Inagaki *et al.*[7]).

II. THEORY

The equations which result from the nonperturbative theory developed in Ref. 1 are briefly outlined below. The physical geometry is shown in Fig. 1. The plane $x_3=0$ determines the average position of a surface whose profile is defined by $x_3=\zeta(\mathbf{x}_{||})$ (where $\mathbf{x}_{||}=\hat{\mathbf{x}}_1x_1+\hat{\mathbf{x}}_2x_2$). The region $x_3>\zeta(\mathbf{x}_{||})$ is vacuum; $x_3<\zeta(\mathbf{x}_{||})$ is dielectric, described by the complex dielectric function $\epsilon(\omega)=\epsilon_R(\omega)+i\epsilon_I(\omega)$. The surface profile function is periodic in two directions: $\zeta(\mathbf{x}_{||})=\zeta(\mathbf{x}_{||}+n\mathbf{a}_1+m\mathbf{a}_2)$. Later I shall implement the theory specifically for a square-lattice grating ($|\mathbf{a}_1|=|\mathbf{a}_2|=a$; $\mathbf{a}_1\perp\mathbf{a}_2$). Light of frequency ω is incident (from vacuum), with a wave vector $\mathbf{k}$, whose projection in the $x_3=0$ plane is $\mathbf{k}_{||}$. The angle of incidence is θ, while the azimuthal angle ϕ (the angle between $\mathbf{k}_{||}$ and $\hat{\mathbf{x}}_1$) describes the orientation of the incident wave vector with respect to the grating.

The electric field in the vacuum is

$$\mathbf{E}(\omega,\mathbf{x})=\mathbf{E}^i(\omega,\mathbf{k}_{||})\exp[i\mathbf{k}_{||}\cdot\mathbf{x}_{||}-i\alpha_o(\omega,k_{||})x_3]+\sum_{\mathbf{G}}\mathbf{E}^s(\omega,\mathbf{K}_{\mathbf{G}})\exp[i\mathbf{K}_{\mathbf{G}}\cdot\mathbf{x}_{||}+i\alpha_o(\omega,K_{\mathbf{G}})x_3]\ . \tag{1}$$

Here one has

$$\mathbf{K}_{\mathbf{G}}=\mathbf{k}_{||}+\mathbf{G}\ , \tag{2}$$

where $\mathbf{G}$ is a reciprocal lattice vector defined by the bigrating periodicity—which for a square-lattice grating becomes

$$\mathbf{G}=\frac{2\pi}{a}(m_1\hat{\mathbf{x}}_1+m_2\hat{\mathbf{x}}_2)\quad\text{for } m_j=0,\pm1,\pm2,\ldots, \tag{3}$$

$$\mathbf{k}_{||}=\frac{\omega}{c}\sin\theta(\cos\phi\,\hat{\mathbf{x}}_1+\sin\phi\,\hat{\mathbf{x}}_2)=\mathbf{K}_0\ , \tag{4}$$

$$\alpha_o(\omega,K_G)=\begin{cases}(\omega^2/c^2-K_G^2)^{1/2}, & \text{for } K_G^2<\omega^2/c^2\ , \quad (5a)\\ i(K_G^2-\omega^2/c^2)^{1/2}, & \text{for } K_G^2>\omega^2/c^2\ . \quad (5b)\end{cases}$$

The incident field has an amplitude

$$\mathbf{E}^i(\omega,\mathbf{k}_{||})=\left[\hat{\mathbf{k}}_{||}+\hat{\mathbf{x}}_3\frac{k_{||}}{\alpha_o(\omega,k_{||})}\right]B_{||}+(\hat{\mathbf{x}}_3\times\hat{\mathbf{k}}_{||})B_\perp\ , \tag{6}$$

and the scattered field has an amplitude

$$\mathbf{E}^s(\omega,\mathbf{K}_G)=\left[\hat{\mathbf{K}}_G-\hat{\mathbf{x}}_3\frac{K_G}{\alpha_o(\omega,K_G)}\right]A_{||}(\omega,\mathbf{K}_G)+(\hat{\mathbf{x}}_3\times\hat{\mathbf{K}}_G)A_\perp(\omega,\mathbf{K}_G)\ . \tag{7}$$

The p and s components are expressed in terms of $B_{||}$ and $B_\perp$ for the incident wave and in terms of the (unknown) Rayleigh coefficients $A_{||}$ and $A_\perp$ for the scattered components. The linear equations for the latter [abbreviated as $A_{||}(\mathbf{G})$ and $A_\perp(\mathbf{G})$] are found to be

$$\sum_{G'}\frac{I(\alpha_{GG'}\,|\,\mathbf{G}-\mathbf{G}')}{\alpha_{GG'}}[a_{GG'}A_{||}(\mathbf{G}')+b_{GG'}A_\perp(\mathbf{G}')]=\frac{-I(\beta_G\,|\,\mathbf{G})}{\beta_G}(a_{G0}B_{||}+b_{G0}B_\perp)\ , \tag{8a}$$

$$\sum_{G'}\frac{I(\alpha_{GG'}\,|\,\mathbf{G}-\mathbf{G}')}{\alpha_{GG'}}[c_{GG'}A_{||}(\mathbf{G}')-a_{GG'}A_\perp(\mathbf{G}')]=\frac{-I(\beta_G\,|\,\mathbf{G})}{\beta_G}(e_GB_{||}-a_{G0}B_\perp)\ , \tag{8b}$$

where

$$\beta_G=\alpha(\omega,K_G)+\alpha_o(\omega,k_{||})\ , \tag{9a}$$

$$\alpha_{GG'}=\alpha(\omega,K_G)-\alpha_o(\omega,K_{G'})\ , \tag{9b}$$

$$a_{GG'}=\hat{\mathbf{x}}_3\cdot(\hat{\mathbf{K}}_G\times\hat{\mathbf{K}}_{G'})\ , \tag{9c}$$

$$b_{GG'}=\hat{\mathbf{K}}_G\cdot\hat{\mathbf{K}}_{G'}\ , \tag{9d}$$

$$c_{GG'}=b_{GG'}+\frac{K_GK_{G'}}{\alpha(\omega,K_G)\alpha_o(\omega,K_{G'})}\ , \tag{9e}$$

$$e_G=\hat{\mathbf{K}}_G\cdot\hat{\mathbf{k}}_{||}-\frac{K_Gk_{||}}{\alpha(\omega,K_G)\alpha_o(\omega,k_{||})}\ , \tag{9f}$$

with

$$\alpha(\omega,K_G)=\left[\epsilon(\omega)\frac{\omega^2}{c^2}-K_G^2\right]^{1/2} \tag{9g}$$

and

$$I(\alpha\,|\,\mathbf{G})=\frac{1}{a_c}\int_{\text{cell}}d^2x_{||}\,e^{-i\mathbf{G}\cdot\mathbf{x}_{||}}e^{-i\alpha\zeta(\mathbf{x}_{||})}\ . \tag{10}$$

Here a_c is the area of a unit cell on the bigrating.

One begins the perturbation theory by expanding the integrand of "I" to first order in ζ, which gives

$$I(\alpha\,|\,\mathbf{G})\approx\delta_{G,0}-i\alpha\tilde{\zeta}(\mathbf{G})\ , \tag{11}$$

where

$$\tilde{\zeta}(\mathbf{G})=\frac{1}{a_c}\int_{\text{cell}}d^2x_{||}\,e^{-i\mathbf{G}\cdot\mathbf{x}_{||}}\zeta(\mathbf{x}_{||})\ . \tag{12}$$

Substituting Eq. (11) for "I" into Eqs. (8), and assuming $\tilde{\zeta}(0)=0$, one finds that

$$A_\perp(\mathbf{G})=i\alpha_{GG}\sum_{\substack{G'\\{\neq G}\}}\tilde{\zeta}(\mathbf{G}-\mathbf{G}')[a_{GG'}A_{||}(\mathbf{G}')+b_{GG'}A_\perp(\mathbf{G}')]-\alpha_{GG}\left[\frac{\delta_{G,0}-i\beta_G\tilde{\zeta}(\mathbf{G})}{\beta_G}\right](a_{G,0}B_{||}+b_{G,0}B_\perp)\ , \tag{13a}$$

$$A_{||}(\mathbf{G})=i\frac{\alpha_{GG}}{c_{GG}}\sum_{\substack{G'\\{\neq G}\}}\tilde{\zeta}(\mathbf{G}-\mathbf{G}')[c_{GG'}A_{||}(\mathbf{G}')-a_{GG'}A_\perp(\mathbf{G}')]-\frac{\alpha_{GG}}{c_{GG}}\left[\frac{\delta_{G,0}-i\beta_G\tilde{\zeta}(\mathbf{G})}{\beta_G}\right](e_GB_{||}-a_{G,0}B_\perp)\ . \tag{13b}$$

The degree of validity of this approximation depends on the value of the parameter $\alpha\zeta$, which is assumed to be small.

It will be assumed that the amplitudes of the two components of the specularly diffracted beam, namely, $A_{||}(0)$ and $A_\perp(0)$, dominate the diffracted orders. Notice that I must consider the possibility that both the p and s components of the specular beam are large. This is because I consider arbitrary polarization of the incident wave and also because, whatever the incident polarization, the bigrating can couple an incident p wave to an s wave and vice versa. Next, I assume that a condition of resonance exists among the evanescent waves: one, and possibly two, evanescent wave amplitudes are much larger than all the rest, corresponding to the one, or possibly two, resonantly excited surface polaritons. A particular evanescent wave, of wave vector $\mathbf{K}_{G_r}$, will correspond to a surface polariton which is resonantly excited by the incident light wave, if and when

$$\mathbf{K}_{G_r}\equiv\mathbf{k}_{||}+\mathbf{G}_r=\mathbf{K}_{SP}(\omega)\ , \tag{14}$$

where $\mathbf{K}_{SP}(\omega)$ is a wave vector of a surface polariton of frequency ω (ω being equal to the incident light frequency). On a flat surface the dispersion relation, $\mathbf{K}_{SP}(\omega)$, is given by

$$\alpha(\omega,K_{SP})+\epsilon(\omega)\alpha_0(\omega,K_{SP})=0\ .$$

Since the flat surface is isotropic, and the dispersion relation thus depends only on the magnitude of $\mathbf{K}_{SP}$, all the

vectors $\mathbf{K}_{SP}(\omega)$ corresponding to a given frequency lie on a circle in the k_1-k_2 plane. Here, where I am treating only small roughness, this flat-surface dispersion relation remains approximately true: *to zeroth order* in the roughness, the $\mathbf{K}_{G_r}$ corresponding to the resonantly excited evanescent wave will lie on the circle of constant frequency and will satisfy the relation:

$$\alpha(\omega,K_{G_r})+\epsilon(\omega)\alpha_o(\omega,K_{G_r})=0 . \tag{15}$$

For certain incident azimuthal angles, ϕ, there may be another reciprocal-lattice vector $\mathbf{G}_p$ such that

$$\mathbf{K}_{G_p}\equiv\mathbf{k}_{||}+\mathbf{G}_p=\mathbf{K}'_{SP}(\omega) . \tag{16}$$

One notices that $\mathbf{K}_{G_r}$ and $\mathbf{K}_{G_p}$ here lie on the boundaries of a two-dimensional Brillouin zone (of the grating) which intersects a constant frequency surface. To zeroth order,

$$|\mathbf{K}_{SP}(\omega)|=|\mathbf{K}'_{SP}(\omega)| \Longrightarrow |\mathbf{K}_{G_r}|=|\mathbf{K}_{G_p}| ,$$

and thus $\mathbf{K}_{G_p}$ lies on the same circle as $\mathbf{K}_{G_r}$. For example, for a square bigrating, such a situation occurs when $\phi=45°$, $\mathbf{G}_r=(1,0)2\pi/a$, and $\mathbf{G}_p=(0,1)2\pi/a$ (see Fig. 2). One thus can have two surface polaritons, in possibly noncollinear directions, simultaneously excited. Such a situation corresponds to the two-surface-polariton peaks measured by Inagaki *et al.*[7] Whenever this condition for the simultaneous resonant excitation of two surface polaritons holds, the two amplitudes $A_{||}(\mathbf{G}_r)$ and $A_{||}(\mathbf{G}_p)$ will be the dominant ones. Only these p components become large, not the s components $A_{\perp}(\mathbf{G}_r)$ and $A_{\perp}(\mathbf{G}_p)$, because the surface polariton is itself a p-polarized wave. For normal incidence, it would be possible to excite simultaneously four surface polaritons.

Here I shall therefore assume that the four amplitudes $A_{||}(0)$, $A_{\perp}(0)$, $A_{||}(\mathbf{G}_r)$, and $A_{||}(\mathbf{G}_p)$ can be larger than the others (non-normal incidence). One can use Eq. (13) to write an explicit expression for each of these four coefficients, and in each case pull the corresponding four important terms out of the sum on the right-hand side:

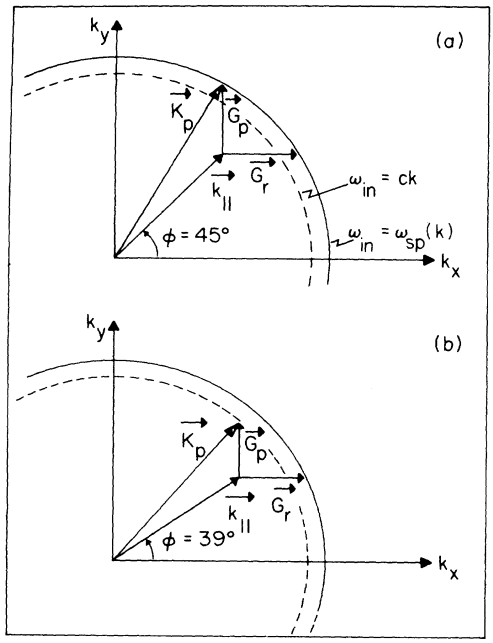

FIG. 2. Schematic showing the solid circle, in the k_x-k_y plane, on which the frequency of surface polaritons, ω_{SP}, equals the frequency of the incident light, ω_{in}. The light line at that frequency, $\omega_{in}=ck$, is the dashed circle. The surface projection of the incident light wave vector is $\mathbf{k}_{||}$. In (a) $\phi=45°$ and $\mathbf{G}_r$ and $\mathbf{G}_p$ each couples to a surface polariton. In (b) $\phi=39°$ and $\mathbf{G}_r$ couples to a surface polariton, but $\mathbf{G}_p$ now represents a diffracted beam ($\mathbf{K}_p=\mathbf{k}_{||}+\mathbf{G}_p$ has crossed inside the light line).

$$A_{||}(\mathbf{G}_r)=i\frac{\alpha_{rr}}{c_{rr}}\left[\tilde{\xi}(r-p)c_{rp}A_{||}(\mathbf{G}_p)+\tilde{\xi}(r)[c_{ro}A_{||}(0)-a_{ro}A_{\perp}(0)]+\tilde{\xi}(r)(e_rB_{||}-a_{ro}B_{\perp})\right.$$
$$\left.+\left[-\tilde{\xi}(r-p)a_{rp}A_{\perp}(\mathbf{G}_p)+\sum_j{}'\tilde{\xi}(r-j)[c_{rj}A_{||}(\mathbf{G}_j)-a_{rj}A_{\perp}(\mathbf{G}_j)]\right]\right] , \tag{17}$$

$$A_{\perp}(0)=i\alpha_{oo}[\tilde{\xi}(-r)a_{or}A_{||}(\mathbf{G}_r)+\tilde{\xi}(-p)a_{op}A_{||}(\mathbf{G}_p)]-\frac{\alpha_{oo}}{\beta_o}B_{\perp}$$
$$+i\alpha_{oo}\left[+\tilde{\xi}(-r)b_{or}A_{\perp}(\mathbf{G}_r)+\tilde{\xi}(-p)b_{op}A_{\perp}(\mathbf{G}_p)+\sum_j{}'\tilde{\xi}(-j)[a_{oj}A_{||}(\mathbf{G}_j)+b_{oj}A_{\perp}(\mathbf{G}_j)]\right] , \tag{18}$$

$$A_{||}(0)=i\frac{\alpha_{oo}}{c_{oo}}[\tilde{\xi}(-r)c_{or}A_{||}(\mathbf{G}_r)+\tilde{\xi}(-p)c_{op}A_{||}(\mathbf{G}_p)]-\frac{e_o\alpha_{oo}}{\beta_o c_{oo}}B_{||}$$
$$+i\frac{\alpha_{oo}}{c_{oo}}\left[-\tilde{\xi}(-r)a_{or}A_{\perp}(\mathbf{G}_r)-\tilde{\xi}(-p)a_{op}A_{\perp}(\mathbf{G}_p)+\sum_j{}'\tilde{\xi}(-j)[c_{oj}A_{||}(\mathbf{G}_j)-a_{oj}A_{\perp}(\mathbf{G}_j)]\right] . \tag{19}$$

The equation for $A_{||}(\mathbf{G}_p)$ is obtained from Eq. (17) by interchanging the indices r and p. To simplify the notation, the vector subscripts $\mathbf{G}_i$ (where $i=0$, r, p, or j) of the terms defined in Eqs. (9) have been replaced by the index "i" alone (e.g., $a_{rp}\equiv a_{\mathbf{G}_r\mathbf{G}_p}$). Similarly, $\tilde{\xi}(\mathbf{G}_j)$ has been written as $\tilde{\xi}(j)$; $\tilde{\xi}(\mathbf{G}_r-\mathbf{G}_p)$ as $\tilde{\xi}(r-p)$; etc. The prime on the summation sign indicates $j\neq 0$, p, or r.

The *nonresonant terms are retained* in Eqs. (17)—(19), and are segregated in large parentheses at the end of each equation on the right-hand side. One can now express each of the nonresonant coefficients with Eq. (13), while dropping the nonresonant scattered amplitudes from the right-hand side: One writes the nonresonant coefficients in terms of only the four important, or resonant, coefficients and of the incident amplitudes as

$$A_{\perp}(\mathbf{G}_j)=i\alpha_{jj}\{\tilde{\xi}(j-r)a_{jr}A_{||}(\mathbf{G}_r)+\tilde{\xi}(j-p)a_{jp}A_{||}(\mathbf{G}_p)+\tilde{\xi}(j)[a_{jo}A_{||}(0)+b_{jo}A_{\perp}(0)]+\tilde{\xi}(j)(a_{jo}B_{||}+b_{jo}B_{\perp})\}\ , \tag{20}$$

$$A_{||}(\mathbf{G}_j)=\frac{i\alpha_{jj}}{c_{jj}}\{\tilde{\xi}(j-r)c_{jr}A_{||}(\mathbf{G}_r)+\tilde{\xi}(j-p)c_{jp}A_{||}(\mathbf{G}_p)+\tilde{\xi}(j)[c_{jo}A_{||}(0)-a_{jo}A_{\perp}(0)]+\tilde{\xi}(j)(e_jB_{||}-a_{jo}B_{\perp})\}\ . \tag{21}$$

One can now use Eqs. (20) and (21) for the nonresonant terms and substitute these back into Eq. (17), giving $A_{||}(\mathbf{G}_r)$ in the form

$$A_{||}(\mathbf{G}_r)=i\frac{\alpha_{rr}}{c_{rr}}F(A_{||}(\mathbf{G}_r),A_{||}(\mathbf{G}_p),A_{||}(0),A_{\perp}(0),B_{||},B_{\perp})\ , \tag{22}$$

where F is a linear function of the six coefficients. One notes, from Eq. (9), that

$$c_{rr}=1+\frac{K_r^2}{\alpha(r)\alpha_o(r)}=\frac{\alpha(r)\alpha_o(r)+K_r^2}{\alpha(r)\alpha_o(r)}\ , \tag{23}$$

where I have introduced the abbreviations: $\mathbf{K}_r=\mathbf{K}_{\mathbf{G}_r}$, $\alpha(r)=\alpha(\omega,K_{\mathbf{G}_r})$, and $\alpha_o(r)=\alpha_o(\omega,K_{\mathbf{G}_r})$. Under resonance conditions on the grating, $K_r=K_{\mathrm{SP}}$, the flat-surface dispersion relation holds to zeroth order: $\alpha(r)\alpha_o(r)+K_r^2=0$; and c_{rr} (or at least its real part) thus vanishes to zeroth order, making $A_{||}(\mathbf{G}_r)$ resonantly large.

One substitutes Eq. (23) into Eq. (22), and multiplies both sides of that equation by

$$\alpha^2(r)\alpha_o^2(r)-K_r^4=[\alpha(r)\alpha_o(r)-K_r^2][\alpha(r)\alpha_o(r)+K_r^2]\ ,$$

so that

$$[\alpha^2(r)\alpha_o^2(r)-K_r^4]A_{||}(\mathbf{G}_r)=i\alpha_{rr}\alpha(r)\alpha_o(r)[\alpha(r)\alpha_o(r)-K_r^2]\times F(\cdots)\ . \tag{24}$$

One sees, by using the definitions for α and α_o in Eqs. (5) and (9g), that

$$\alpha^2(r)\alpha_o^2(r)-K_r^4=\frac{\omega^2}{c^2}\left[\epsilon\frac{\omega^2}{c^2}-(1+\epsilon)K_r^2\right]\ . \tag{25}$$

I define ω_r as the frequency that a surface polariton of wave vector $\mathbf{K}_r$ has on a flat surface, namely,

$$\omega_r^2\equiv\left[\frac{1+\epsilon}{\epsilon}\right]c^2K_r^2\ . \tag{26}$$

In the limit of the flat surface, i.e., as $\zeta\rightarrow 0$, one has $\omega_r\rightarrow\omega$. Expressing K_r on the right-hand side of Eq. (25) in terms of ω_r as defined in Eq. (26), one finds

$$\alpha^2(r)\alpha_o^2(r)-K_r^4=\frac{\omega^2}{c^2}\epsilon\left[\frac{\omega^2-\omega_r^2}{c^2}\right]\ . \tag{27}$$

One can now substitute Eq. (27) into the left-hand side of Eq. (24), yielding the result

$$(\omega^2-\omega_r^2)A_{||}(\mathbf{G}_r)=i\alpha_{rr}\frac{c^4}{\epsilon\omega^2}\times\gamma_rF(A_{||}(\mathbf{G}_r),A_{||}(\mathbf{G}_p),\ldots)\ , \tag{28}$$

where

$$\gamma_r=\alpha(r)\alpha_o(r)[\alpha(r)\alpha_o(r)-K_r^2]\ . \tag{29}$$

The term $A_{||}(\mathbf{G}_r)$ within the function F on the right-hand side arose from keeping the nonresonant terms and is thus proportional to ζ^2. One can remove this term from F and use it on the left-hand side to renormalize the flat surface frequency ω_r; so that,

$$(\omega^2-\tilde{\omega}_{rp}^2)A_{||}(\mathbf{G}_r)=i\alpha_{rr}\frac{c^4}{\epsilon\omega^2}\gamma_rF^1(A_{||}(\mathbf{G}_p),A_{||}(0),A_{\perp}(0),B_{||},B_{\perp})\ , \tag{30}$$

where the F^1 is just F with the term $A_{||}(\mathbf{G}_r)$ removed, and where

$$\tilde{\omega}_{rp}^2=\omega_r^2\left\{1-\frac{c^2}{(1+\epsilon)\omega^2}\left[\frac{\gamma_r}{K_r^2}\right]\alpha_{rr}\left[\alpha_{pp}a_{rp}^2\tilde{\xi}(r-p)\tilde{\xi}(p-r)+\sum_j{}'\tilde{\xi}(r-j)\tilde{\xi}(j-r)\left[\frac{c_{rj}c_{jr}}{c_{jj}}+a_{rj}^2\right]\alpha_{jj}\right]\right\}\ . \tag{31}$$

One finds the analogous equation for $A_{||}(\mathbf{G}_p)$ by interchanging r and p in Eqs. (30) and (31).

For the case in which $A_{||}(\mathbf{G}_r)$ is resonant—that is, when $\mathbf{K}_r=\mathbf{K}_{sp}(\omega)$—one could substitute the zeroth-order, flat surface, approximation (in which $\omega_r\approx\omega$) into the right-hand side of Eq. (28), in which case

$$\alpha(r)\alpha_o(r)\approx -K_r^2 \tag{32}$$

and thus

$$\gamma_r\approx 2K_r^4 . \tag{33}$$

This approximation for γ_r, if used in Eq. (28), would then appear also in Eqs. (30) and (31).

The above procedure is correct in this first-order theory, because the next term in the expansion of $\alpha(r)\alpha_o(r)$ [starting with Eq. (25)] is proportional to $\omega^2-\omega_r^2$ and is thus of the order ζ^2, while the function F on the right-hand side of Eq. (28) is already of the order ζ. Thus, one is neglecting only terms of higher order than ζ^2. To this same order of approximation, one could replace ω by ω_r everywhere on the right-hand side of Eqs. (28), (30), and (31).

In the expansion of $\alpha(r)\alpha_o(r)$—starting with Eq. (25) and expanding the square root—I chose the negative sign for the square root to get Eq. (32), because $\alpha(r)$ is imaginary (ϵ_I gives only a small real part) as is $\alpha_o(r)$ (assuming that K_r corresponds to the resonant evanescent wave).

For the case in which $A_{||}(\mathbf{G}_p)$ is in simultaneous resonance, the approximation $\gamma_p\approx 2K_p^4=2K_r^4$ can be used in the equation for $A_{||}(\mathbf{G}_p)$. Moreover, in this case of simultaneous resonance, $\alpha_{rr}=\alpha_{pp}$, so that the relation

$$\alpha_{rr}\alpha_{pp}=\alpha_{rr}^2=(1+\epsilon)\frac{\omega^2}{c^2} \tag{34}$$

can be used in the expression for $\tilde{\omega}_{rp}^2$ [Eq. (31)] and similarly for $\tilde{\omega}_{pr}^2$.

One must be careful in using these approximations [Eqs. (32)—(34)]. Obviously, they break down as one moves away from resonance conditions, e.g., as one moves far enough away from the minimum in a reflectivity dip. But even directly at a resonance or reflectivity minimum, a serious problem develops as soon as one moves off a Brillouin zone boundary, i.e., off the condition for *two*-polariton simultaneous resonance. One beam, say $\mathbf{K}_r$, is still resonant, but the other, $\mathbf{K}_p$, which is not resonant, can rapidly approach the light line ($\omega=ck$), where $\alpha_o(p)$ has a square-root singularity. For example (see Fig. 2), consider a square-lattice grating, when both $\mathbf{K}_r=\mathbf{k}_{||}+(1,0)2\pi/a$ and $\mathbf{K}_p=\mathbf{k}_{||}+(0,1)2\pi/a$ are resonant at $\phi=45°$. Let ϕ be lowered and let $\mathbf{K}_r$ remain resonant. For $a=2186$ nm and $\lambda=633$ nm, $\mathbf{K}_p$ hits the light line when $\phi\approx 40°$. When ϕ is further lowered, $\mathbf{K}_p$ crosses the light line and becomes a diffracted order, in which case the approximations like Eqs. (32) and (33) are completely inappropriate for the $\mathbf{K}_p$ beam.

Although the above approximations do give some algebraic simplification in the final expressions, in practice when it comes to computing reflectivities—when all the terms are evaluated on the computer anyway (at a given incidence frequency ω)—it is just as easy to determine γ_r, γ_p, α_r, and α_p exactly (at ω) and avoid complication. On the other hand, when solving for the dispersion relation, in which ω is the unknown, one can get a simple expression for ω, only if on the right-hand side of Eq. (28) one uses the approximation $\omega=\omega_r$ *or* $\omega=\omega_p$, depending on which term is resonant, or else $\omega=\omega_r=\omega_p$ for zone boundaries.

Next, one takes Eq. (18) for $A_\perp(0)$ and Eq. (19) for $A_{||}(0)$ and substitutes into them the expressions of Eqs. (20) and (21) for the nonresonant terms. This gives $A_\perp(0)$ and $A_{||}(0)$ in terms of $A_{||}(\mathbf{G}_r)$, $A_{||}(\mathbf{G}_p)$, $B_{||}$, and $B_\perp$.

One now has four, linear, algebraic, inhomogeneous equations for $A_{||}(0)$, $A_\perp(0)$, $A_{||}(\mathbf{G}_r)$, and $A_{||}(\mathbf{G}_p)$ in terms of $B_{||}$ and $B_\perp$. In matrix form these equation are

$$\begin{bmatrix} (\Omega^2-\tilde{\Omega}_{rp}^2) & L_{rp} & M_{rp} & N_{rp} \\ L_{pr} & (\Omega^2-\tilde{\Omega}_{pr}^2) & M_{pr} & N_{pr} \\ O_{rp} & O_{pr} & 1+P_{rp} & Q_{rp} \\ R_{rp} & R_{pr} & S_{rp} & 1+T_{rp} \end{bmatrix} \begin{bmatrix} A_{||}(\mathbf{G}_r) \\ A_{||}(\mathbf{G}_p) \\ A_{||}(\mathbf{0}) \\ A_\perp(\mathbf{0}) \end{bmatrix} = \begin{bmatrix} (U_r-M_{rp})B_{||}-N_{rp}B_\perp \\ (U_p-M_{pr})B_{||}-N_{pr}B_\perp \\ (W+V-P_{rp})B_{||}-Q_{rp}B_\perp \\ (X-S_{rp})B_{||}-\left[\dfrac{\alpha_{00}}{\beta_0}+T_{rp}\right]B_\perp \end{bmatrix} . \tag{35}$$

Here, $\Omega\equiv\omega/(2\pi c/a)=a/\lambda$; $\tilde{\Omega}_{rp}\equiv\tilde{\omega}_{rp}/(2\pi c/a)$; and the other symbols, namely, the letters L through X, are defined in the Appendix. One notes that all these symbols $(\Omega,\tilde{\Omega}_{rp},L,\ldots,X)$ represent dimensionless quantities. One also notes that P_{rp}, Q_{rp}, S_{rp}, and T_{rp} are invariant with respect to interchanging r and p, and are second order in ζ (along with V and X).

The problem is now reduced to one of solving just four, linear, simultaneous equations for the four unknowns: $A_{||}(\mathbf{G}_r)$, $A_{||}(\mathbf{G}_p)$, $A_{||}(0)$, and $A_\perp(0)$. These four amplitudes can then be used back in Eqs. (20) and (21) to calculate the nonresonant amplitudes $A_{||}(\mathbf{G}_j)$ and $A_\perp(\mathbf{G}_j)$. Then with Eq. (7), the scattered-field amplitude $\mathbf{E}^s(\mathbf{K}_G)$ for any of the diffracted orders (and for any of the evanescent waves as well) can be found. One can thus obtain the reflectance for each diffracted order of the bigrating and also find the total field and field enhancement above the bigrating.

In order to determine the dispersion relation for surface polaritons on the bigrating, one sets to zero the right-hand side (i.e., the incident field) of Eq. (35). With zero incident fields, the amplitudes $A_{||}(\mathbf{0})$ and $A_\perp(\mathbf{0})$ represent only the *grating-induced* radiative losses of the leaky surface polariton. One can thus include the $A_{||}(\mathbf{0})$ and $A_\perp(\mathbf{0})$ back in the sums with the other *non*resonant terms, leaving one with the homogeneous 2×2 problem:

$$\begin{bmatrix} \Omega^2-\tilde{\Omega}_{rp}^2 & L_{rp} \\ L_{pr} & \Omega^2-\tilde{\Omega}_{pr}^2 \end{bmatrix} \begin{bmatrix} A_{\|}(\mathbf{G}_r) \\ A_{\|}(\mathbf{G}_p) \end{bmatrix} = 0 , \tag{36}$$

for which a solution exists when

$$(\Omega^2-\tilde{\Omega}_{rp}^2)(\Omega^2-\tilde{\Omega}_{pr}^2)-L_{rp}L_{pr}=0 \tag{37}$$

or

$$\Omega_{\pm}^2=\tfrac{1}{2}(\tilde{\Omega}_{rp}^2+\tilde{\Omega}_{pr}^2)\pm\tfrac{1}{2}[(\tilde{\Omega}_{rp}^2-\tilde{\Omega}_{pr}^2)^2+4L_{rp}L_{pr}]^{1/2} . \tag{38}$$

Not surprisingly, this equation has the same structure as that found for the classical grating by Glass, Weber, and Mills.[16] One must remember that in order to be used in Eq. (38), the definitions of $\tilde{\omega}_{rp}$ in Eq. (31) (needed for $\tilde{\Omega}_{rp}$) and of L_{rp} in Eq. (A1) must be changed to include $j=0$ (i.e., $\mathbf{G}_j=0$) in the summations, $\sum'$.

If one ignores all the nonresonant terms, one finds that $\Omega^2=\Omega_r^2\pm(L_{rp}L_{pr})^{1/2}$. At a Brillouin zone boundary [where $K_r=K_p$ and where Eqs. (32), (33), and (34) are true to lowest order] one finds that

$$L_{rp}L_{pr}=16K_r^2\,|\,\tilde{\zeta}(r-p)\,|^2\sin^4\left[\frac{\theta_{rp}}{2}\right]\left[\frac{|\epsilon|}{(|\epsilon|-1)^2}\right]\Omega_r^4 ,$$

for which θ_{rp} is defined as the angle between $\mathbf{K}_r$ and $\mathbf{K}_p$, and where it is assumed that the bigrating is symmetric, so that $\tilde{\zeta}(p-r)=\tilde{\zeta}^*(r-p)$. In that case, $\omega=\omega_r(1\pm\Delta)^{1/2}\approx\omega_r(1\pm\Delta/2)$, and the gap width is

$$\Delta=4K_r\,|\,\tilde{\zeta}(r-p)\,|\,\sin^2\left[\frac{\theta_{rp}}{2}\right]\left[\frac{|\epsilon|^{1/2}}{(|\epsilon|-1)}\right] . \tag{39}$$

This result is the same as that found by Mills.[14] The more complete result found here, Eq. (38), by taking into account the nonresonant terms [including the (0,0)], allows for the radiative losses, and is thus the appropriate equation for finding the complex solutions in the radiative region. Moreover, it allows for the displacement of the gap center, unlike the earlier result. Its only limitation is at the intersection of two Brillouin zone boundaries, where there may be four resonant waves. A straightforward generalization of the present results, with four resonant amplitudes, yields a 6×6 equation for the normal incidence reflectance and a 4×4 equation for the dispersion relation.

III. NUMERICAL IMPLEMENTATION OF THE THEORY

This section describes some numerical results for reflectances and field enhancements, with various bigrating amplitudes, for the purpose of indicating the limits of validity of the perturbation theory—as determined by comparison with the results of the exact theory.

The reflectance of the Gth diffracted beam [see Eqs. (1)–(7)] is

$$R_{\mathbf{G}}=\frac{|\,\mathbf{E}^s(\omega,\mathbf{K}_{\mathbf{G}})\,|^2\alpha_o(\omega,K_{\mathbf{G}})}{|\,\mathbf{E}^i(\omega,\mathbf{k}_{\|})\,|^2\alpha_o(\omega,k_{\|})} , \tag{40}$$

and the total reflectance is $R=\sum_{\mathbf{G}}R_{\mathbf{G}}$. The field enhancement will be defined as

$$\mathscr{E}=\frac{|\,\mathbf{E}(\omega,\mathbf{x})\,|^2}{|\,\mathbf{E}^i(\omega,\mathbf{k}_{\|})\,|^2} \tag{41}$$

evaluated at $\mathbf{x}$ just above the maximum in the surface profile.

The surface profile function that is employed here has the following form:

$$\begin{aligned}\zeta(\mathbf{x}_{\|})=h_1&\left[\cos\left[\frac{2\pi x_1}{a}\right]+\cos\left[\frac{2\pi x_2}{a}\right]\right]\\&+h_2\left[\cos\left[\frac{4\pi x_1}{a}\right]+\cos\left[\frac{4\pi x_2}{a}\right]\right]\\&+h_{11}\cos\left[\frac{2\pi x_1}{a}\right]\cos\left[\frac{2\pi x_2}{a}\right] .\end{aligned} \tag{42}$$

This ζ represents a square-lattice grating whose profile is a symmetric function in x_1 and in x_2. For this square bigrating, $\mathbf{G}=(m_1,m_2)2\pi/a$ (where m_1 and $m_2=0,\pm1,\ldots$), so that one can write $R_{\mathbf{G}}$ as $R(m_1,m_2)$ and refer to the (m_1,m_2) beam.

Implementation of the exact theory is discussed in detail in Refs. 1 and 2. Its implementation in the present work—to serve as a standard of comparison—will now be described briefly (this description provides further indication of the computational advantages of the perturbation theory).

To implement the exact theory requires evaluating the Fourier transform of the exponential of $\zeta(\mathbf{x}_{\|})$, namely, $I(\alpha\,|\,\mathbf{G})$ in Eq. (10), for each value of $\mathbf{G}$. In the present case (and in general) there is no analytic expression for $I(\alpha\,|\,\mathbf{G})$. Thus, a fast Fourier transform (FFT) must be performed numerically for each value of $\mathbf{G}$. The cross term in the profile function—the h_{11} term—would necessitate a full two-dimensional FFT. Use of the two-dimensional FFT can be avoided, however, if one expands the exponential of the cross term. Then one obtains, for $\mathbf{G}=\hat{\mathbf{x}}_1G_1+\hat{\mathbf{x}}_2G_2$,

$$I(\alpha\,|\,\mathbf{G})=\sum_{n=0}^{\infty}\frac{(-i\alpha h_{11})^n}{n!}I^{(n)}(\alpha\,|\,G_1)I^{(n)}(\alpha\,|\,G_2) , \tag{43}$$

where the integral

$$I^{(n)}(\alpha\,|\,G_v)=\frac{1}{a}\int_{-a/2}^{a/2}e^{-iG_vx}\exp\left\{-i\alpha\left[h_1\cos\left[\frac{2\pi x}{a}\right]+h_2\cos\left[\frac{4\pi x}{a}\right]\right]\right\}\left[\cos\left[\frac{2\pi x}{a}\right]\right]^n dx \tag{44}$$

is evaluated with a *one*-dimensional FFT (32 point mesh used here). The summation is truncated at the required precision. For the cases presented here, just three or four terms in the sum will yield graphical accuracy or better (in some cases, better than one part in 10^4).

With the $I(\alpha \mid \mathbf{G})$ thus determined, the exact method then requires solving Eq. (8), an infinite-dimensional, inhomogeneous, matrix equation. One truncates to a $2N^2 \times 2N^2$ matrix problem, by retaining in Eq. (8) only those reciprocal lattice vectors $\mathbf{G}=(m_1,m_2)2\pi/a$ for which m_1 and $m_2=-(N-1)/2$ to $+(N-1)/2$. Convergence of the results to at least graphical accuracy, and usually much better (in some cases to one part in 10^4) is found here for $N=7$. Therefore, all the exact results presented here are for $N=7$ (i.e., for 98×98 matrices).

In implementing the perturbation theory one solves the 4×4 matrix equation, Eq. (35); and one needs no FFT's, since $I(\alpha \mid \mathbf{G})$ is not needed. All one needs is $\tilde{\zeta}(\mathbf{G})$, which for the profile of Eq. (42) is

$$\tilde{\zeta}(m_1,m_2)=\tfrac{1}{2}\delta_{m_1,0}(h_1\delta_{|m_2|,1}+h_2\delta_{|m_2|,2}) + \tfrac{1}{2}\delta_{m_2,0}(h_1\delta_{|m_1|,1}+h_2\delta_{|m_1|,2}) + \tfrac{1}{4}h_{11}\delta_{|m_1|,1}\delta_{|m_2|,1} .$$

The recent experimental measurements of absorptance by Inagaki *et al.*[7] were carried out for a bigrating on Ag, with a profile of the form given in Eq. (42) and a period of $a=2186$ nm, and for an incident light wavelength of $\lambda=633$ nm. In Ref. 2 the exact theory was fit to the experimental curves of Inagaki *et al.* by using the following height parameters: $h_1=13.7$ nm, $h_2=2.42$ nm, and $h_{11}=9.1$ nm. For that fitting, the value of the dielectric constant was interpolated from the table in Ref. 19 to be $\epsilon=-18.3+i0.479$.

These same parameters have been chosen here as a starting point in testing the perturbation theory. The azimuthal orientation of the incident propagation direction is taken to be $\phi=45°$. In that case, the evanescent waves $\mathbf{K}_r$ and $\mathbf{K}_p$ for $\mathbf{G}_r=(1,0)2\pi/a$ and $\mathbf{G}_p=(0,1)2\pi/a$ can be simultaneously resonant, so that these values are put into the perturbation theory. Figure 3 shows the results for the total reflectance versus angle of incidence, θ, when the incident light is p polarized. The minimum in the reflectivity dip, corresponding to the resonant excitation of the surface plasmon polariton, occurs at an angle 0.14° smaller with the perturbation theory than with the exact theory, and is 11% lower in magnitude. When ϕ is changed to 0.0°, with $\mathbf{G}_r=(1,1)2\pi/a$ and $\mathbf{G}_p=(1,-1)2\pi/a$, the results become worse: The exact theory gives a reflectance minimum of $R=0.90$, while the perturbation theory gives a minimum also roughly 0.10 in depth, but starting from values where $R>1$. The perturbation theory for finding the reflectance is clearly breaking down at this point.

Recall that the perturbation technique employed here requires the expansion of $\exp(-i\alpha_{GG'}\zeta)$. At a surface polariton resonance $|\alpha_o| \approx (2\pi/a)(|\epsilon|-1)^{-1/2}a/\lambda$ and $|\alpha| \approx |\epsilon|\,|\alpha_o|$, so that

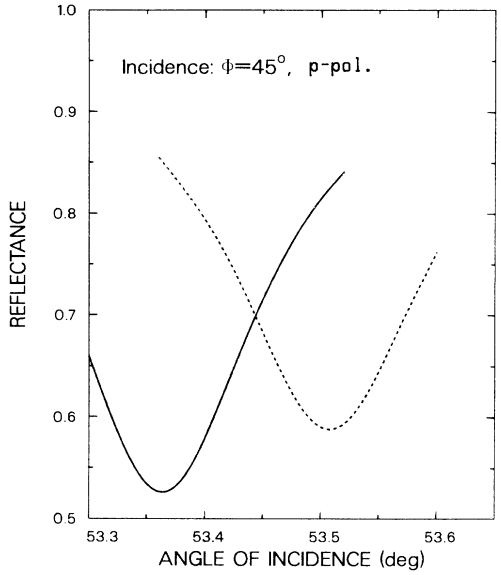

FIG. 3. Total reflectance R versus angle of incidence θ. For p-polarized light ($\lambda=633$ nm) incident at $\phi=45°$ on a square bigrating ($a=2186$ nm) with $h_1=13.7$, $h_2=2.42$, and $h_{11}=9.1$ nm. The solid line is from the perturbation theory; the dashed line is from the exact theory. The dip corresponds to the $(0,1)-(1,0)$ resonant excitation of surface polaritons.

$$|\alpha|\zeta \approx \left[\frac{2\pi|\epsilon|}{(|\epsilon|-1)^{1/2}}\right]\frac{\zeta}{\lambda} .$$

For the physical parameters just considered, one finds that

$$|\alpha(K_G)-\alpha_o(K_G)|\zeta_{max} \approx |\alpha(K_G)|\zeta_{max} \approx 1.8 ,$$

which is greater than 1, thus indicating why the perturbation theory might not have worked well in this instance. This point will be discussed further in Sec. IV.

In order to test the theory for a case in which $\alpha\zeta_{max}<1.0$, the following, otherwise arbitrary, choice of grating amplitudes has been made: $h_1=8.0$ nm, $h_2=1.5$ nm, and $h_{11}=2.0$ nm. The other parameters (λ,a,ϵ) are left unchanged, so that at resonance, $|\alpha|\zeta_{max}=0.9$. The results for the total reflectance are shown in Fig. 4, again for p-polarized incident light at $\phi=45°$. The result of perturbation theory differs from that of the exact theory by only 0.02° for the position of the minimum (or $\Delta k_{||}/k_{||}=0.026\%$), and by 2.7% for its minimum value. The result for the specular reflectance (not shown) looks about the same. The reflectance for two of the off-specular beams, the $(-1,0)$ and $(-1,1)$ are shown in Fig. 5, with good agreement between the perturbation and the exact theories [despite the fact that the off-specular amplitudes come from Eqs. (20) and (21) from which all the other nonresonant terms were dropped]. The complex amplitude of the (1,0)-resonant evanescent wave—that is, the amplitude of one of the two excited surface polaritons—is shown in Fig. 6 [the (0,1) amplitude is identical]. The enhancement of the electric field intensity, evaluated at the position $\mathbf{x}=(0,0,1.005\zeta_{max})$, is shown in

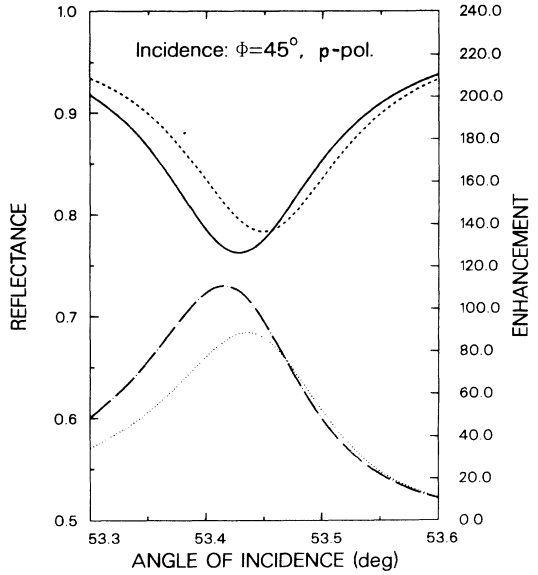

FIG. 4. Total reflectance, R (upper curves). Enhancement of the electric field intensity, $\mathscr{E}$, at $\mathbf{x}=(0,0,1.005\zeta_{\max})$ (lower curves). Both versus angle of incidence θ. For p-polarized light ($\lambda=633$ nm) incident at $\phi=45°$ on a square bigrating ($a=2186$ nm) with $h_1=8$, $h_2=1.5$, and $h_{11}=8$ nm. The solid line shows R from the perturbation theory; the dashed line from the exact theory. The dotted-dashed line shows $\mathscr{E}$ from perturbation theory; the dotted line from exact theory. The dip in R and peak in $\mathscr{E}$ correspond to the $(0,1)-(1,0)$ resonant excitation of surface polaritons.

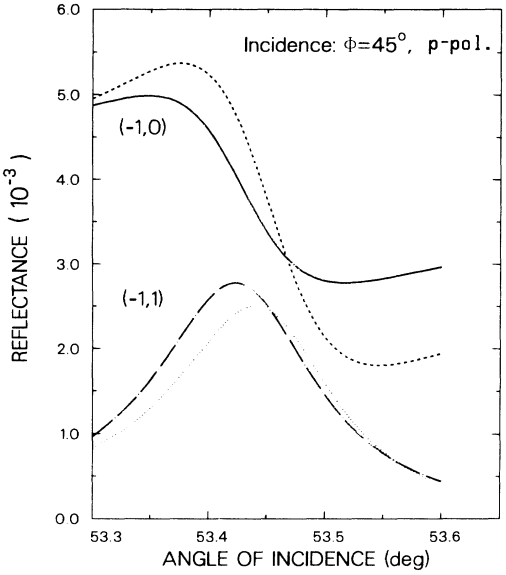

FIG. 5. Reflectance in off-specular beams, R_G, versus angle of incidence, θ. For p-polarized light ($\lambda=633$ nm) incident at $\phi=45°$ on the same bigrating as in Fig. 4. The solid line shows the perturbation theory and the dashed line the exact theory for the $(-1,0)$ beam. The dotted-dashed line shows the perturbation theory and the dotted line the exact theory for the $(-1,1)$ beam. The peak in the $(-1,1)$ reflectance corresponds to the $(1,0)-(0,1)$ resonant excitation of surface polaritons ($\approx$ same θ as the minimum in R in Fig. 4).

Fig. 4. In each of these cases, Figs. 4–6, the curve resulting from perturbation theory is shifted to the left by the same $\approx 0.02°$ with respect to the curve from the exact theory. The greatest error occurs for the field enhancement: perturbation theory overestimates the peak value by 24%. It is not unusual or surprising that a perturbative method based on the Rayleigh hypothesis works less well in the near-field limit (at the surface) than in the far-field limit (for the reflectance).

One notes that in this case of incident p-polarized light, the specularly reflected beam is found to be also p polarized [$A_\perp(0,0)\approx 0.0$], thus conforming with the result in Ref. 1: for specular reflection the p to s coupling is zero at $\phi=0°$ and $\phi=45°$. The nonspecular beams, however, are found to be of mixed polarization: at $\theta=53.45°$, for $B_{||}=1$ and $B_\perp=0$, we find $A_{||}(1,-1)=(1.3+i1.9)\times 10^{-2}$ and $A_\perp=(7.2+i4.0)\times 10^{-2}$.

As expected, the basic assumption—that the largest amplitudes are among our set of four—is confirmed by the exact theory. For the light and grating parameters discussed above (and again $\phi=45°$, $\theta=53.45°$, $B_{||}=1$ and $B_\perp=0$), we find: $A_{||}(0,0)=-0.63-i0.61$ and $A_{||}(1,0)=A_{||}(0,1)=-1.6-i0.60$; whereas the largest nonresonant amplitudes are, for example: $A_{||}(1,1)=-0.18-i0.064$, $A_{||}(2,0)=-0.10-i0.04$, and $A_{||}(-1,0)=-0.012-i0.057$. Similar confirmation of the basic assumption is found in the corresponding cases discussed below.

When the incident light is s polarized, with $\phi=45°$, the result of the perturbation theory, as seen from the total reflectance in Fig. 7, is just as good as for the p polarized case.

The case in which p-polarized light is incident at $\phi=0°$, with $\mathbf{G}_r=(1,1)2\pi/a$ and $\mathbf{G}_p=(1,-1)2\pi/a$, allows one to test the perturbation theory when the cross term h_{11} is responsible for the resonance. I have already noted how in the first trial, with $\alpha\zeta_{\max}>1$, the perturbation results for this $(1,1)-(1,-1)$ resonance are worse than for the $(1,0)-(0,1)$ resonance. What happens in the present case with $\alpha\zeta_{\max}<1$? The results shown in Fig. 8, curves (a), demonstrate the same excellent agreement between the perturbative and exact method as at the $(1,0)-(0,1)$ resonance. Of course, here $h_{11}(=2$ nm$)<h_1(=8$ nm$)$. Curves (b) in Fig. 8 show what happens when one interchanges h_1 and h_{11}, namely, when $h_1=2$ nm, $h_2=1.5$ nm, and $h_{11}=8$ nm (still p polarized at $\phi=0°$); the same excellent agreement between the two methods is seen. This latter case may, nevertheless, still not provide the proper comparison between the $(1,0)-(0,1)$ resonance at $\phi=45°$ and the $(1,1)-(1,-1)$ at $\phi=0°$; because $h_1=8$ nm contributes 16 nm to $\zeta_{\max}$ (i.e., $\zeta_{\max}=2h_1+2h_2+h_{11}=21$ nm for Fig. 4), whereas $h_{11}=8$ nm contributes only 8 nm to $\zeta_{\max}$ [i.e., $\zeta_{\max}=15$ nm for Fig. 8(b)]. There-

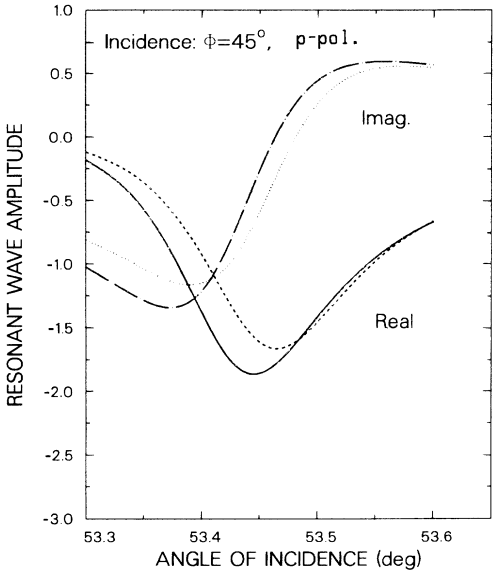

FIG. 6. Amplitude $A_{||}(1,0)$ of the (1,0)-resonant evanescent wave (the surface polariton), normalized with respect to the incident amplitude, versus angle of incidence θ. For p-polarized light ($\lambda=633$ nm) incident at $\phi=45°$ on the same bigrating as in Fig. 4. The solid line shows the perturbation theory and the dashed line the exact theory for the real part, $\mathrm{Re}[A_{||}(1,0)]$. The dotted-dashed line shows the perturbation theory and the dotted line the exact theory for the imaginary part, $\mathrm{Im}[A_{||}(1,0)]$. The $|A_{||}(1,0)|^2$ peaks where R in Fig. 4 has a minimum, corresponding to the $(1,0)-(0,1)$ resonant excitation of surface polaritons.

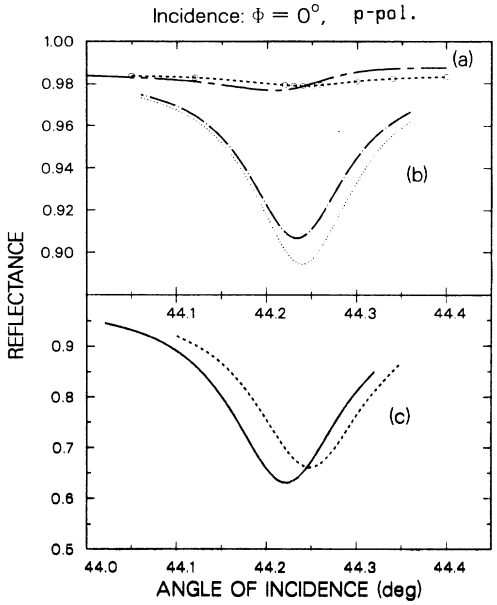

FIG. 8. Total reflectance R versus angle of incidence θ. For p-polarized light ($\lambda=633$ nm) incident at $\phi=0°$ on a square bigrating with: (a) $h_1=8$, $h_2=1.5$, $h_{11}=2$ nm; (b) $h_1=2$, $h_2=1.5$, $h_{11}=8$ nm; and (c) $h_1=1$, $h_2=1.5$, $h_{11}=16$ nm. The results of the perturbation theory are given by the (a) long-dashed chain, (b) dotted-dashed, and (c) solid lines. The results of the exact theory are given by the (a) short-dashed and open circles, (b) dotted, and (c) short-dashed lines. The dips in each correspond to the $(1,1)-(1,-1)$ resonant excitation of surface polaritons.

fore, the case of $h_1=1$ nm, $h_2=1.5$ nm, and $h_{11}=16$ nm was used, with the results shown by curves (c) in Fig. 8. One sees that the perturbation theory gives a minimum whose position is too low by just 0.03° and whose magnitude is too deep by 6%. Although these results are still fairly good, it again seems that the coupling through the h_{11} cross term in the profile is described less accurately than coupling through the h_1 term.

What happens when one changes ϕ such that only a single wave can become resonant, i.e., one moves away from a Brillouin zone boundary? Let us return to the case in which $h_1=8$ nm, $h_2=1.5$ nm, and $h_{11}=2$ nm (λ, a, and ϵ still unchanged), with p-polarized incident light at $\phi=39°$. At this azimuthal angle, when the $\mathbf{G}_r=(1,0)2\pi/a$ wave becomes resonant (at $\theta=51.93°$), the $\mathbf{G}_p=(0,1)2\pi/a$ wave is not only no longer resonant, but it has crossed over the light line and has become a diffracted wave (Fig. 2). Thus, $\phi=39°$ is a good place to test whether the present, coupled-mode, perturbation theory properly goes over into a single-mode, ordinary, perturbation theory. Comparing the results of the perturbation theory to the exact theory for the total reflectance indicates that the minimum is off in position by 0.01° and is too deep by only 1.6%. The good agreement generally carries over even to the nonspecular beams, as shown in Fig. 9 for $R(-1,0)$, $R(1,-1)$, and $R(0,1)$—where the (0,1) beam is the one which has just crossed over the light line.

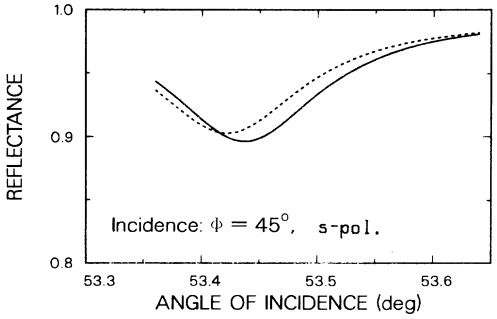

FIG. 7. Total reflectance R versus angle of incidence θ. For s-polarized light ($\lambda=633$ nm) incident at $\phi=45°$ on the same bigrating as in Fig. 4. The solid line is from perturbation theory and the dashed line from the exact theory. The dip corresponds to the $(1,0)-(0,1)$ resonant excitation of surface polaritons.

Under certain conditions, as seen in various grating studies, the dip in the specular reflectance can fall to near zero. Pockrand[9] noted that this optimal coupling of incident photon to surface polariton occurs when there is a matching between the power dissipated in the lossy medium (due to ϵ_I) and the radiative power losses (due to the grating). In Ref. 1, this optimal coupling condition was seen, with our exact theory, to occur on a sinusoidal bi-

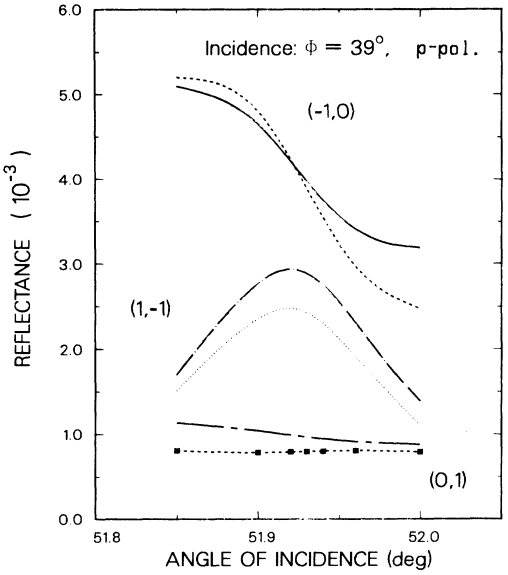

FIG. 9. Reflectance in off-specular beams, R_G, versus angle of incidence, θ. For p-polarized light ($\lambda=633$ nm) incident at $\phi=39°$. The results of the perturbation theory are given by the solid line for $(-1,0)$, the dotted-dashed line for $(1,-1)$, and the long-dashed chain line for $(0,1)$. The results of the exact theory are given by the short-dashed line for $(-1,0)$, the dotted line for $(1,-1)$, and the dashed and solid square line for $(0,1)$. The peak in R for the $(1,-1)$ beam is at the resonant excitation of the (1,0) surface polariton.

grating (with only the h_1 component) of period $a=800$ nm, for p-polarized incident light of $\lambda=514.5$ nm at $\phi=0°$, for the (1,0) resonance. The minimum in the specular reflectance was seen to equal 0.40 at $\theta=23.97°$ for $h_1=8$ nm, and then equal 0.03 at $\theta=24.01°$ for $h_1=16$ nm. What does the present perturbation theory give in this strong-coupling case? For $h_1=8$ nm, it gives $R(0,0)=0.30$ (with about a 3% variation depending on the choice of $\mathbf{G}_p$) at $\theta=23.96°$, thus yielding a 25% error—even though $\alpha\zeta_{\max}\approx 0.68<1$ here. The field enhancement $\mathscr{E}=190$ is overestimated by 30%. When $h_1=16$ nm (in which case $\alpha\zeta_{\max}\approx 1.36>1$), the minimum reflectance is seen to have *increased* to $R(0,0)=0.59$ rather than to have fallen to near zero; and its position is 0.2° too high. However, when $h_1=12$ nm is tried [with $\mathbf{G}_p$ equal to the $(-1,0)$-diffracted beam], perturbation theory does give $R(0,0)=0.003$, for $\theta=24.09°$. Thus, the trend $R(0,0)\to 0$, as predicted for these values of "a" and λ, is indeed found by the perturbation theory; but the value of h_1 at which it occurs is off by around 25%. What is surprising is that the perturbation theory works as well as it does in this instance, since a basic assumption in the theory, that the specular-beam amplitudes are dominant, is breaking down as $R(0,0)\to 0$. The freedom to pick the now nonresonant wave $\mathbf{G}_p$ to correspond to one of the larger amplitude, nonspecular, diffracted beams can partially offset that breakdown.

Just to verify that the above errors are due specifically to the strong-coupling situation, and not due to the fact that this (1,0) resonance at $\phi=0°$ is simply very far from a Brillouin zone boundary, I consider the following. I return to the previous case, in which $a=2186$ nm and $\lambda=633$ nm with $h_1=8$ nm, $h_2=1.5$ nm, and $h_{11}=2$ nm, and look at the (1,0) resonance at $\phi=0°$. Here I find a minimum reflectance of $R(0,0)=0.758$ at $\theta=47.64°$ with perturbation theory, and $R(0,0)=0.777$ at $\theta=47.64°$ with the exact theory. There is an error in the reflectance of only 2%, although one is just as far from any zone boundary as in the previous case (i.e., there is no second wave, $\mathbf{G}_p$, near resonance). The previous large errors thus must be associated with the strong-coupling condition.

IV. DISCUSSION AND CONCLUSIONS

The perturbation theory presented here generally gives excellent results, for both the specular and off-specular reflectance when $\alpha\zeta_{\max}<1$. The exception to this good performance is the special case in which the reflectance minimum drops to near zero (due to a power matching condition). But even then, the perturbation theory can be made to give a reasonable approximation.

When $\alpha\zeta_{\max}>1$, the perturbation theory becomes less reliable. The question arises as to why other perturbation theories, either for the classical grating (e.g., Ref. 17) or for the total absorptance on a bigrating (Ref. 7) seem to work better for similar values of $\alpha\zeta>1$.

First, the classical grating theories do not consider the simultaneous excitation of two surface polaritons propagating in noncollinear directions, as is considered here. Such polaritons can interact via the cross terms in the surface profile (e.g., h_{11}), and have a strong angular dependence in their interaction [see Eq. (39)]. As pointed out by Inagaki *et al.*,[7] in reference to this angular dependence described by Mills,[14] higher-order processes may contribute more strongly for noncollinear propagation.

Second, as far as other perturbation approaches for bigratings are concerned, one notes that the results of the perturbation theory used by Inagaki *et al.*[7] when compared to those of the exact theory [in Ref. (2)], also show at least some signs of difficulty for $\alpha\zeta_{\max}>1$. Fitting the perturbation theory of Elson and Sung[15] to experiment gave $h_{11}=5.14$ nm,[7] whereas fitting the exact theory of Glass, Maradudin, and Celli[1] gave $h_{11}=9.1$ nm.[2] Such discrepancies notwithstanding, the Elson-Sung theory for the total absorptance was employed by Inagaki *et al.* with less severe difficulties than those encountered here with the present theory for the same parameters ($\alpha\zeta>1$).

This brings one to the third and most important point concerning the performance of the present theory: it relies on the Rayleigh hypothesis. That hypothesis is accepted to be valid within a certain domain ($\zeta_{\max}/a<0.072$) for which the Rayleigh series converges. Here $\zeta_{\max}/a$ is well within that limit, and one does, in fact, find rapid numerical convergence in the exact method. The problem exists when one then approximates the Rayleigh method to first order. When $\alpha\zeta>1$ (i.e., when the penetration of the field into the vacuum is much less than the groove depth), then the Rayleigh sum, which is analytically continued down to the surface to satisfy the

boundary conditions, is poorly represented in its first-order approximation of $\exp(-\alpha\zeta)$.

It is important to note, however, that even when $\alpha\zeta > 1$ and the present perturbation theory for determining the *magnitude* of the reflectance is inaccurate (and even unphysical: $R > 1$), the *position* of the reflectance dip can still be well predicted: In the worse case seen in the present study, the minimum was off by 0.4° and the $\Delta K_{SP}/K_{SP} = 0.7\%$. Thus, this perturbation theory, when used for solving the complex dispersion relation [Eq. (38)], should remain valid over a greater range of groove heights than it does for calculating the scattered fields.

Knowing the limits of validity of the present perturbation theory, as just described, should be useful in bigrating studies for the various practical reasons mentioned in the Introduction. But there is yet another use for these results, namely, to give an indication of where to place the limits on some recent theories of the randomly rough surface. Resonant light scattering from a randomly rough surface was recently treated by a method that begins with the same procedures as followed here: the Rayleigh hypothesis was used together with an extinction theorem to eliminate the field in the medium, and equations analogous to Eq. (8) were obtained.[20–22] A procedure that formally includes all orders was presented. It has been implemented to lowest order in ζ.[21,22] The results obtained here, by having an exact theory for comparison, can therefore be useful, by way of analogy, for better understanding the limitations on other such first-order numerical calculations.

ACKNOWLEDGMENT

This work was made possible with the support of the Naval Postgraduate School (NPS) Foundation Research Program.

APPENDIX

The matrix elements in Eq. (35) use the following definitions.

$$L_{rp} = h_r \left[i\tilde{\xi}(r-p)c_{rp} - \sum_j{}' \tilde{\xi}(r-j)\tilde{\xi}(j-p)(f_{rj}c_{jp} - a_{rj}a_{jp})\alpha_{jj} \right], \tag{A1}$$

$$M_{rp} = h_r \left[i\tilde{\xi}(r)c_{ro} + \tilde{\xi}(r-p)\tilde{\xi}(p)\alpha_{pp}a_{rp}a_{po} - \sum_j{}' \tilde{\xi}(r-j)\tilde{\xi}(j)(f_{rj}c_{jo} - a_{rj}a_{jo})\alpha_{jj} \right], \tag{A2}$$

$$N_{rp} = -h_r \left[i\tilde{\xi}(r)a_{ro} - \tilde{\xi}(r-p)\tilde{\xi}(p)\alpha_{pp}a_{rp}b_{po} - \sum_j{}' \tilde{\xi}(r-j)\tilde{\xi}(j)(f_{rj}a_{jo} + a_{rj}b_{jo})\alpha_{jj} \right], \tag{A3}$$

$$O_{rp} = \frac{-\alpha_{oo}}{c_{oo}} \left[i\tilde{\xi}(-r)c_{or} + \tilde{\xi}(p-r)\tilde{\xi}(-p)\alpha_{pp}a_{pr}a_{op} - \sum_j{}' \tilde{\xi}(j-r)\tilde{\xi}(-j)(f_{oj}c_{jr} - a_{rj}a_{jo})\alpha_{jj} \right], \tag{A4}$$

$$P_{rp} = \frac{\alpha_{oo}}{c_{oo}} \left[\tilde{\xi}(r)\tilde{\xi}(-r)a_{or}^2\alpha_{rr} + \tilde{\xi}(p)\tilde{\xi}(-p)a_{op}^2\alpha_{pp} + \sum_j{}' \tilde{\xi}(j)\tilde{\xi}(-j)(f_{oj}c_{jo} + a_{jo}^2)\alpha_{jj} \right], \tag{A5}$$

$$Q_{rp} = \frac{-\alpha_{oo}}{c_{oo}} \left[\tilde{\xi}(r)\tilde{\xi}(-r)a_{or}b_{ro}\alpha_{rr} + \tilde{\xi}(p)\tilde{\xi}(-p)a_{op}b_{po}\alpha_{pp} + \sum_j{}' \tilde{\xi}(j)\tilde{\xi}(-j)(f_{oj} - b_{jo})a_{jo}\alpha_{jj} \right], \tag{A6}$$

$$R_{rp} = -\alpha_{oo} \left[i\tilde{\xi}(-r)a_{or} - \tilde{\xi}(p-r)\tilde{\xi}(-p)a_{pr}b_{po}\alpha_{pp} - \sum_j{}' \tilde{\xi}(j-r)\tilde{\xi}(-j)(f_{jr}a_{oj} + a_{jr}b_{jo})\alpha_{jj} \right], \tag{A7}$$

$$S_{rp} = \alpha_{oo} \left[\tilde{\xi}(p)\tilde{\xi}(-p)a_{po}b_{po}\alpha_{pp} + \tilde{\xi}(r)\tilde{\xi}(-r)a_{ro}b_{ro}\alpha_{rr} + \sum_j{}' \tilde{\xi}(j)\tilde{\xi}(-j)(f_{jo} - b_{jo})a_{oj}\alpha_{jj} \right], \tag{A8}$$

$$T_{rp} = \alpha_{oo} \left[\tilde{\xi}(p)\tilde{\xi}(-p)b_{po}^2\alpha_{pp} + \tilde{\xi}(r)\tilde{\xi}(-r)b_{ro}^2\alpha_{rr} + \sum_j{}' \tilde{\xi}(j)\tilde{\xi}(-j) \left[\frac{a_{oj}^2}{c_{jj}} + b_{jo}^2 \right] \alpha_{jj} \right], \tag{A9}$$

$$U_r = h_r \left[i\tilde{\xi}(r)g_r - \sum_j{}' \tilde{\xi}(r-j)\tilde{\xi}(j)\tau_j c_{rj} \right], \tag{A10}$$

$$V = \frac{\alpha_{oo}}{c_{oo}} \sum_j{}' \tilde{\xi}(j)\tilde{\xi}(-j)\tau_j c_{oj}, \tag{A11}$$

$$W = -\frac{\alpha_{oo}e_o}{c_{oo}\beta_o}, \tag{A12}$$

$$X = \alpha_{oo} \sum_j{}' \tilde{\xi}(j)\tilde{\xi}(-j)\tau_j a_{oj}. \tag{A13}$$

In the above equations the following definitions have been employed:

$$f_{mi}=\frac{c_{mi}}{c_{ii}} , \tag{A14a}$$

$$f_{im}=\frac{c_{im}}{c_{ii}} , \tag{A14b}$$

$$g_i=c_{io}-e_i , \tag{A15}$$

$$h_r=-\left[\frac{\alpha_{rr}c^4}{\epsilon\omega^2}\right]\gamma_r\left[\frac{1}{2\pi c/a}\right]^2 , \tag{A16}$$

$$\tau_i=\frac{g_i\alpha_{ii}}{c_{ii}} . \tag{A17}$$

The summations $\sum'$ are over all j except $j=0$, $j=r$, and $j=p$. The terms a_{ij}, b_{ij}, c_{ij}, and e_i were defined in Eq. (9). They are dimensionless, as are f_{ij} and g_i.

[1]N. E. Glass, A. A. Maradudin, and V. Celli, J. Opt. Soc. Am. **73**, 1240 (1983).
[2]N. E. Glass and A. A. Maradudin, Opt. Commun. **56**, 339 (1986).
[3]D. Maystre and M. Nevière, J. Opt. (Paris) **9**, 301 (1978).
[4]P. Vincent, Opt. Commun. **26**, 293 (1978).
[5]G. H. Derrick, R. C. McPhedran, D. Maystre, and M. Nevière, Appl. Phys. **18**, 39 (1979).
[6]R. C. McPhedran, G. H. Derrick, M. Nevière, and D. Maystre, J. Opt. (Paris) **13**, 209 (1982).
[7]T. Inagaki, J. P. Goudonet, J. W. Little, and E. T. Arakawa, J. Opt. Soc. Am. B **2**, 433 (1985).
[8]E. Kroger and E. Kretschmann, Phys. Status Solidi B **76**, 515 (1976).
[9]I. Pockrand, J. Phys. D **9**, 2423 (1976).
[10]I. Pockrand and H. Raether, Appl. Opt. **16**, 1784 (1977).
[11]W. Rothballer, Opt. Commun. **20**, 249 (1977).
[12]H. Raether, Opt. Commun. **42**, 217 (1982).
[13]F. Toigo, A. Marvin, V. Celli, and N. R. Hill, Phys. Rev. B **15**, 5618 (1977).
[14]D. L. Mills, Phys. Rev. B **15**, 3097 (1977).
[15]J. M. Elson and C. C. Sung, Appl. Opt. **21**, 1496 (1982).
[16]N. E. Glass, M. Weber, and D. L. Mills, Phys. Rev. B **29**, 6548 (1984).
[17]M. G. Weber, Phys. Rev. B **33**, 909 (1986).
[18]D. L. Mills and M. Weber, Phys. Rev. B **27**, 2698 (1983).
[19]P. B. Johnson and R. W. Christy, Phys. Rev. B **6**, 4370 (1972).
[20]G. C. Brown, V. Celli, M. Haller, and A. Marvin, Surf. Sci. **136**, 381 (1984).
[21]G. Brown, V. Celli, M. Haller, A. A. Maradudin, and A. Marvin, Phys. Rev. B **31**, 4993 (1985).
[22]A. R. McGurn, A. A. Maradudin, and V. Celli, Phys. Rev. B **31** 4866 (1985).