Analytic model of plasmonic coupling: Surface relief gratings

Nir Rotenberg* and J. E. Sipe
Department of Physics and Institute for Optical Sciences, University of Toronto, Toronto, Ontario M5S 1A7, Canada
(Received 16 May 2010; revised manuscript received 7 October 2010; published 25 January 2011)

We develop an analytic model to describe the coupling of light to surface plasmon polaritons by surface relief gratings. Plasmonic coupling efficiencies over 0.8 are predicted to within 15%, while the corresponding computational times are more than 2 orders of magnitude shorter than needed in standard numerical approaches. In addition, we are able both to provide simple equations for the physically significant constants of this problem, such as the coupling strength and reradiation terms, and to predict the spatial dynamics of the plasmonic and reflected fields.

DOI: 10.1103/PhysRevB.83.045416 PACS number(s): 73.20.Mf, 78.20.Bh, 78.68.+m

I. INTRODUCTION

While surface relief gratings are among the most commonly used optical components, modeling their interaction with electromagnetic radiation in an accurate and efficient manner has proven to be an enduring challenge. The difficulty stems from the periodic interface between the grating and the overlaying media, across which there exist no analytic solutions to Maxwell's equations. In 1907, Lord Rayleigh first addressed this issue by calculating the diffracted fields in terms of outgoing waves;[1] this became known as the Rayleigh hypothesis, which forms the basis for a host of theoretical treatments of the diffraction of electromagnetic waves by gratings.[2–6] Although these methods can be valid even for deep gratings,[6] the advent of modern computing has led to modeling methods such as the C-Method,[7,8] rigorous coupled-wave analysis (RCWA),[9–11] and finite-domain time-difference (FDTD)[12] calculations that are numerical in nature. Although these methods can deal with a vast array of grating geometries, they are all computationally intensive. Moreover, they offer little physical insight into the process that they model: Due to the nature of these methods, the only way to explore the effect of varying a parameter on the system is to repeat the entire calculation with a range of parameters and then analyze the complete set of results.

An exciting application of gratings that has been integral to numerous studies is the use of grating couplers to convert free-space radiation into surface plasmon polaritons (SPPs).[13–16] SPPs are collective electronic excitations that are coupled to the electromagnetic field and bound to a metal-dielectric interface, along which they propagate. The momentum of these modes along the interface is different from that component of incident free-space radiation, and consequently a grating is often used to provide the extra momentum and couple the two fields. Generally, the presence of SPPs is deduced from the light that is reflected from a grating, where the missing energy has been coupled to the SPP instead of being reflected. Consequently, it is the reflection from the grating that is usually simulated, using one of the standard methods,[7–12] and the SPP coupling efficiency is subsequently extracted from the dip in the reflectivity. It is worth noting that while many Rayleigh expansion–based methods, be they semianalytic[2,4] or fully analytic,[5] are formulated to deal with plasmonic coupling, they are not widely employed in current literature that reports on experimental plasmonic studies. This is particularly true of plasmonic experiments involving nanostructures, where the finite spatial extant of the incident beams must be considered, and a formalism to deal with arbitrarily shaped beams is lacking.

In this work we introduce an approach that allows for analytic calculations of light-grating interactions leading to SPP coupling. Using a Green function–based formalism developed earlier,[17,18] we reduce the problem to a simple differential equation; the parameters of this equation have clear physical significance and can be calculated analytically. In contrast with both the Rayleigh expansion–based method and the numeric model already cited, our model is concerned with the coupling of light to SPPs and not with grating diffraction in general; thus, the model presented herein deals only with the specular reflection and the SPP field and not with the different diffraction orders, as do the previous cited works, though it could be extended to allow for diffraction. Consequently, this approach circumvents the need to solve large eigenvalue problems, as in the C-Method or RWCA-based calculations, the finite-element algorithm of FDTD calculations, or indeed the majority of the Rayleigh expansion techniques, and thus dramatically reduces computation times.

Even where the computational times are not significantly improved, as in the case for the model of Sheng *et al.*,[5] which deals specifically with SPPs excited on square gratings, our model offers a unique blend of advantages: First, we provide a functional link between the parameters of the system, such as the grating amplitude and dielectric constant of the metal, and the SPP coupling properties, such as the strength of the coupling and the spectral width of the resonance. Second, our method is constructed to be very general and can easily be extended to cover many practical situations encountered in the laboratory; in this paper, we present the formalism required to calculate the plasmonic coupling from a *p*-polarized continuous-wave beam of either finite or infinite spatial extant, but this work can be easily extended for arbitrary polarizations or direction of propagation, and even to pulsed excitations. This is in contrast with the previously cited methods, which deal only with continuous and plane-wave excitation. Finally, and most importantly, this method allows for the spatial variation of the associated fields. That is, given an excitation beam of finite extent, we are able to predict the spatial evolution of both the reflected and the SPP field.

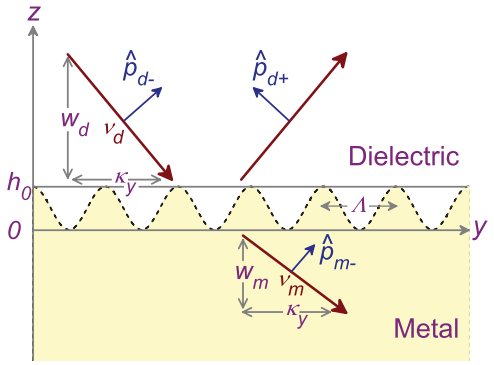

FIG. 1. (Color online) A metallic surface relief grating beneath a dielectric. The grating profile, $h(y)$, shown (dashed line) is sinusoidal but may be any periodic function. Also sketched are the wave vectors and polarization vectors associated with the electric fields. In general these vectors can be complex, and consequently this figure is meant simply as a visualization aid and not an exact representation.

II. STATEMENT OF THE PROBLEM

Our goal is to calculate both the SPP and the reflected fields that can result when light is incident on a metallic grating. We show the surface relief grating and the notation relevant to this problem in Fig. 1. The metallic grating and substrate have a frequency-dependent dielectric constant $\varepsilon_m(\omega)$ and are below a transparent medium with a real dielectric constant, $\varepsilon_d(\omega)$. The grating has amplitude h_0 in the z direction and period Λ in the y direction, with a surface profile $h(y)$ that can be written as a Fourier series,

$$h(y) = \sum_p h_{pG} e^{ipGy}, \tag{1}$$

where $G = 2\pi/\Lambda$ is the grating wave vector.

This system can be thought of as three distinct regions: (1) the region above the grating, $z > h_0$, where only the dielectric is present; (2) the region of the grating, $h_0 \geqslant z \geqslant 0$, where both the dielectric and the metal are present; and (3) the region below the grating, $z < 0$, where only the metal is present. To solve for the electric fields in these regions, given an incident field, we treat the grating as a perturbation. For a planar interface between the metal and the dielectric, the electromagnetic fields are determined by the Fresnel coefficients (see Appendix A); the presence of the grating in region 2 results in a small correction. That is, we begin by treating the material in region 2 as dielectric and then consider the presence of the metal grating as inducing a polarization that then acts like a source term, driving the electric fields.

The electric fields are described both by wave vectors and by polarization vectors. As shown in Fig. 1, the total wave vector of the light in the dielectric (metal) is $\nu_{d(m)} = \tilde{\omega} n_{d(m)}$, where $\tilde{\omega} = 2\pi/\lambda$ is the wave vector of light, of wavelength λ, in a vacuum, and n is the index of refraction of the material. The total wave vector is split into the y component, κ_y, and the z component, $w_{d(m)}$, which is then given by

$$w_i = \sqrt{\tilde{\omega}^2 \varepsilon_i - \kappa_y^2}, \tag{2}$$

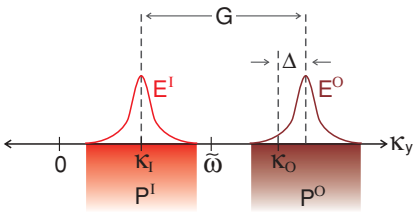

FIG. 2. (Color online) The electric fields and polarizations, in k space, that need to be calculated. Superscripts I and O show that the relevant quantities are evaluated at the incident and SPP wave vectors, respectively.

where $i = d, m$. Here the square root is defined so that $\Im\{w_i\} \geqslant 0$, and $\Re\{w_i\} \geqslant 0$ if $\Im\{w_i\} = 0$. Likewise, we define the unit polarization vectors as

$$\hat{\boldsymbol{p}}_{i\pm} = \frac{\kappa_y \hat{\boldsymbol{z}} \mp w_i \hat{\boldsymbol{\kappa}}}{\nu_i}, \tag{3}$$

where $\hat{\boldsymbol{p}}_{+(-)}$ is the polarization vector for radiation traveling in the positive (negative) z direction. We have restricted ourselves to p-polarized light, as this is most commonly used in plasmonic experiments.[13] It is, however, possible to couple s-polarized light to SPPs when it is obliquely incident on the grating, as long as the wave vectors of the light and the SPP, in the direction of the propagation of the SPP, match. Our approach can be readily extended to this situation by replacing the p-polarized Fresnel coefficients with those for s-polarized light (see Appendix A) and carefully accounting for the component of the vectors in the direction of plasmonic propagation.

We have restricted ourselves to p-polarized light, as s-polarized light has no electric field component in the direction of the grating and can therefore not couple to a SPP[13]. It is easier to develop our model in terms of the Fourier components of the fields (κ space) and obtain the field dependence in real space by a Fourier transform. Using this approach, we show a schematic of our method in Fig. 2. We consider the electric fields, and the induced polarizations, at the incident wave vector, κ_I, separate from those at the SPP wave vector, κ_O. That is, the electric fields and the polarizations are written as

$$\boldsymbol{E} = \boldsymbol{E}^I + \boldsymbol{E}^O, \tag{4}$$

$$\boldsymbol{P} = \boldsymbol{P}^I + \boldsymbol{P}^O, \tag{5}$$

and as shown in Fig. 2, it is assumed that the components do not overlap in κ space.

The SPP wave vector is given by the familiar dispersion relation,

$$\kappa_{\mathrm{SP}} = \tilde{\omega} \left(\frac{\varepsilon_d \varepsilon_m}{\varepsilon_d + \varepsilon_m} \right)^{\frac{1}{2}}, \tag{6}$$

and we define $\kappa_O = \Re\{\kappa_{\mathrm{SP}}\}$ and $\gamma = \Im\{\kappa_{\mathrm{SP}}\}$. As expected, $\kappa_O > \tilde{\omega}$ and a unit of grating momentum G is required to excite a SPP. We allow for a small detuning, which we define as

$$\Delta = \kappa_I - \kappa_O + G. \tag{7}$$

The effect of the grating can be described as follows: (1) An electric field is incident at κ_I and, through grating-assisted momentum transfer, induces a polarization at κ_O. (2) This

polarization drives the SPP field at κ_O. (3) Through reverse scattering, the field at κ_O sets up a polarization at κ_I. (4) The polarization at κ_I drives an electric field at this wave vector that contributes to the reflected electric field. In the following section, we present the mathematical formulation of this scenario.

III. BASIC EQUATIONS

For monochromatic electric fields we have

$$\boldsymbol{E}(\boldsymbol{r},t) = \boldsymbol{E}(\boldsymbol{r})e^{-i\omega t} + \text{c.c.}, \tag{8}$$

where we take

$$\boldsymbol{E}(\boldsymbol{r}) = \int \frac{d\kappa_y}{2\pi} e^{i\kappa_y y}\boldsymbol{E}(\kappa_y,z). \tag{9}$$

Here the spatial components have been explicitly separated, and the y dependence of the field is Fourier transformed, as this allows us to employ the Fourier coefficients of the grating profile [Eq. (1)] instead of its profile.

We proceed by writing general expressions for the electric fields in the three regions of the problem. Using a Green function formalism,[18] the field above the grating, for $z > h_0$, is

$$\begin{aligned}\boldsymbol{E}(\kappa_y,z) = {}& \hat{\boldsymbol{p}}_{d-}E_{\text{inc}}(\kappa_y)e^{-iw_d z} + \hat{\boldsymbol{p}}_{d+}r_{dm}E_{\text{inc}}(\kappa_y)e^{iw_d z} \\ & + \frac{i\tilde{\omega}^2}{2\varepsilon_0 w_d}\hat{\boldsymbol{p}}_{d+}\hat{\boldsymbol{p}}_{d+}e^{iw_d z}\int_0^{h_0} e^{-iw_d z'}\boldsymbol{P}(\kappa_y,z')\,dz' \\ & + \frac{i\tilde{\omega}^2}{2\varepsilon_0 w_d}\hat{\boldsymbol{p}}_{d+}r_{dm}\hat{\boldsymbol{p}}_{d-}e^{iw_d z}\int_0^{h_0} e^{iw_d z'}\boldsymbol{P}(\kappa_y,z')\,dz',\end{aligned} \tag{10}$$

where E_{inc} is the incident electric field, ε_0 is the permittivity of free space, and r_{dm} is the κ_y-dependent Fresnel coefficient of reflection (Appendix A). The polarization vector $\boldsymbol{P}(\kappa_y,z')$ should be understood to be dotted with the preceding unit vector (e.g., $\hat{\boldsymbol{p}}_{d+}$). Each of the four terms in Eq. (10) has a clear physical significance. The first term represents the electric field that is incident on the system; the second term is the reflection from the planar metal dielectric interface at $z = 0$, which occurs as though there was no grating present; and the last two terms are the corrections due to the presence of the grating, where the first corresponds to the field generated by the polarization in the grating that propagates up, away from this region, while the last is due to this induced field, which initially propagates down and is then reflected by the metal.

In the region of the grating, for $h_0 \geqslant z \geqslant 0$, the electric field is

$$\begin{aligned}&\boldsymbol{E}(\kappa_y,z) \\ &= \hat{\boldsymbol{p}}_{d-}E_{\text{inc}}(\kappa_y)e^{-iw_d z} + \hat{\boldsymbol{p}}_{d+}r_{dm}E_{\text{inc}}(\kappa_y)e^{iw_d z} \\ &\quad + \frac{i\tilde{\omega}^2}{2\varepsilon_0 w_d}\hat{\boldsymbol{p}}_{d+}\hat{\boldsymbol{p}}_{d+}\int_0^{h_0}\theta(z-z')e^{iw_d(z-z')}\boldsymbol{P}(\kappa_y,z')\,dz' \\ &\quad + \frac{i\tilde{\omega}^2}{2\varepsilon_0 w_d}\hat{\boldsymbol{p}}_{d-}\hat{\boldsymbol{p}}_{d-}\int_0^{h_0}\theta(z'-z)e^{-iw_d(z-z')}\boldsymbol{P}(\kappa_y,z')\,dz' \\ &\quad - \frac{\hat{z}\hat{z}}{\varepsilon_0\varepsilon_d}\boldsymbol{P}(\kappa_y,z) + \frac{i\tilde{\omega}^2}{2\varepsilon_0 w_d}\hat{\boldsymbol{p}}_{d+}r_{dm}\hat{\boldsymbol{p}}_{d-}\int_0^{h_0} e^{iw_d(z+z')} \\ &\quad \times \boldsymbol{P}(\kappa_y,z')\,dz',\end{aligned} \tag{11}$$

where the Heaviside function used is

$$\theta(z-z') = \begin{cases} 0 & \text{for } z < z', \\ 1 & \text{for } z > z'. \end{cases} \tag{12}$$

While the first two terms, as well as the last, have the same physical significance as those in Eq. (10), the remaining three terms must be interpreted separately. The first represents the electric field, generated in this region, which propagates upward; for this region, we need only consider the field generated below the point of interest (i.e., $z > z'$), which will then propagate to z. Conversely, the second term represents the generated electric field that then propagates downward but has not been reflected yet. Finally, the $\hat{z}\hat{z}$ term describes the part of the field not associated with the aforementioned propagating waves.

Finally, below the grating and in the metal, for $z < 0$, the electric field is

$$\begin{aligned}\boldsymbol{E}(\kappa_y,z) = {}& \hat{\boldsymbol{p}}_{m-}t_{dm}E_{\text{inc}}(\kappa_y)e^{-iw_m z} + \frac{i\tilde{\omega}^2}{2\varepsilon_0 w_d}\hat{\boldsymbol{p}}_{m-}t_{dm}\hat{\boldsymbol{p}}_{d-}e^{-iw_m z} \\ & \times \int_0^{h_0} e^{iw_d z'}\boldsymbol{P}(\kappa_y,z')\,dz'.\end{aligned} \tag{13}$$

Here the first term represents the part of the incident field that is transmitted across the metal-dielectric interface, while the second term provides the correction to this field due to the grating.

It is interesting to consider the nature of the fields in the preceding equations. Since the dielectric function of the metal is complex, so too is w_m, and hence the terms in Eq. (13) all represent evanescent fields. In the dielectric and the region of the grating [Eqs. (10) and (11)], the situation is more complicated. Since, in general, the dielectric function in these regions is real, the nature of w_d depends on the wave vector being considered. For example, at the incident wave vector, w_d is real and the fields are propagating, while at the SPP wave vector, w_d is imaginary and hence the fields are evanescent.

We note that at this point the equations are exact. If the polarization $\boldsymbol{P}(\kappa_y,z)$ were known, then Eqs. (10)–(13) would give the full electric field. We will see that a model we introduce, based on including only κ_y associated with a few center wave numbers, can accurately predict plasmonic coupling. This is in contrast with numerical models that describe grating diffraction, which often require the explicit inclusion of tens if not hundreds of κ_y to converge. The reason for this is that here we are describing a *resonant* phenomenon, in contrast to the general problem of grating diffraction, and hence the exact form of the local, evanescent fields on the surface of the grating is less important. The characterization of the local fields near the surface is an important one for certain applications, and in a future publication we plan to turn to the accuracy of our method in describing them. But it is not central to the coupling problem to which we now turn.

IV. THE MODEL

We adopt the following strategy: We approximate the electric fields at the incident and SPP wave vectors, κ_I and κ_O, respectively, based on the supposition of a small grating height ($w_d h_0 \ll 1$). These fields are coupled by the polarization that

exists in the region of the grating. The polarization itself is expressed in terms of the generating fields, as well as the Fourier coefficients of the surface profile. This self-consistent approach leads to a simple dynamic equation for the evolution of the SPP envelope function, as well as to an equation for the envelope function of the reflected field.

To simplify the calculations we define the following:

$$\boldsymbol{C}(\kappa_y,z) \equiv \boldsymbol{E}(\kappa_y,z) + \frac{\hat{\boldsymbol{z}}\hat{\boldsymbol{z}}}{\varepsilon_o\varepsilon_d}\boldsymbol{P}(\kappa_y,z), \tag{14}$$

$$\boldsymbol{Q}(\kappa_y) \equiv \int_0^{h_0} \boldsymbol{P}(\kappa_y,z)\,dz, \tag{15}$$

where $\boldsymbol{C}(\kappa_y,z)$ will allow us to combine the electric fields in the two regions above the metal, and $\boldsymbol{Q}(\kappa_y)$ will lead to a simplified equation for the grating-induced polarization in k space.

A. Surface plasmon field

We calculate the electric field at κ_O, $\boldsymbol{E}^O$, assuming that the grating perturbation is small, $w_d^O h_0 \ll 1$. Recalling that $E_{\text{inc}}(\kappa_O) = 0$ and making use of Eqs. (14) and (15) allows us to rewrite Eqs. (10)–(13) as

$$\boldsymbol{C}^O(\kappa_y,z) = \frac{i\tilde{\omega}^2}{2\varepsilon_0 w_d^O}\hat{\boldsymbol{p}}_{d+}^O\hat{\boldsymbol{p}}_{d+}^O e^{iw_d^O z}\boldsymbol{Q}^O(\kappa_y) + \frac{i\tilde{\omega}^2}{2\varepsilon_0 w_d^O}\hat{\boldsymbol{p}}_{d+}^O r_{dm}\hat{\boldsymbol{p}}_{d-}^O e^{iw_d^O z}\boldsymbol{Q}^O(\kappa_y),$$

for $z > h_0$. For $h_0 \geqslant z \geqslant 0$,

$$\begin{aligned}\boldsymbol{C}^O(\kappa_y,z) &= \frac{i\tilde{\omega}^2}{2\varepsilon_0 w_d^O}\hat{\boldsymbol{p}}_{d+}^O\hat{\boldsymbol{p}}_{d+}^O e^{iw_d^O z}\int_0^{h_0}\theta(z-z')\boldsymbol{P}^O(\kappa_y,z')\,dz' \\ &+ \frac{i\tilde{\omega}^2}{2\varepsilon_0 w_d^O}\hat{\boldsymbol{p}}_{d-}^O\hat{\boldsymbol{p}}_{d-}^O e^{-iw_d^O z}\int_0^{h_0}\theta(z'-z)\boldsymbol{P}^O(\kappa_y,z')\,dz' \\ &+ \frac{i\tilde{\omega}^2}{2\varepsilon_0 w_d^O}\hat{\boldsymbol{p}}_{d+}^O r_{dm}\hat{\boldsymbol{p}}_{d-}^O e^{iw_d^O z}\boldsymbol{Q}^O(\kappa_y),\end{aligned}$$

and for $z < 0$,

$$\boldsymbol{C}^O(\kappa_y,z) = \frac{i\tilde{\omega}^2}{2\varepsilon_0 w_d^O}\hat{\boldsymbol{p}}_{m-}^O t_{dm}\hat{\boldsymbol{p}}_{d-}^O e^{-iw_m^O z}\boldsymbol{Q}^O(\kappa_y). \tag{16}$$

Since SPPs are resonant modes, they are signaled by poles in the Fresnel coefficients (see Appendix A) at κ_{SP}. Consequently, if the preceding equations are multiplied by $(\kappa_y - \kappa_{\text{SP}})$, only the terms containing Fresnel coefficients do not vanish, and they can be combined to yield

$$(\kappa_y - \kappa_{\text{SP}})\boldsymbol{C}^O(\kappa_y,z) = \frac{i\tilde{\omega}^2\rho_{dm}}{2\varepsilon_0 w_d^O}\boldsymbol{e}^O(z)\hat{\boldsymbol{p}}_{d-}^O\boldsymbol{Q}^O(\kappa_y), \tag{17}$$

for all z, where we have defined a new field:

$$\boldsymbol{e}^O(z) \equiv \hat{\boldsymbol{p}}_{d+}^O e^{iw_d^O z}\theta(z) + \hat{\boldsymbol{p}}_{m-}^O e^{-iw_m^O z}\frac{\tau_{dm}}{\rho_{dm}}\theta(-z). \tag{18}$$

This field decays exponentially both into the metal and into the dielectric; it is the SPP electric field. We introduce an envelope function, f_O, which allows us to express the total field at κ_O as

$$\boldsymbol{C}^O(\kappa_y,z) = \boldsymbol{e}^O(z)\,f_O(\kappa_y - \kappa_O). \tag{19}$$

By inserting this into Eq. (17), we arrive at the equation for the SPP envelope function,

$$(\kappa_y - i\gamma)f_O(\kappa_y) = F(\kappa_y), \tag{20}$$

where

$$F(\kappa_y) = \frac{i\tilde{\omega}^2\rho_{dm}}{2\varepsilon_0 w_d^O}\hat{\boldsymbol{p}}_{d-}^O\cdot\boldsymbol{Q}^O(\kappa_y + \kappa_O). \tag{21}$$

That is, once $F(\kappa_y)$ is known, the SPP envelope function can be calculated. We take the Fourier transform of this equation to arrive at the differential equation for the SPP envelope function in real space,

$$\frac{df_O(y)}{dy} = -\gamma f_O(y) + iF(y). \tag{22}$$

B. Polarization

We proceed by finding the dependence of the polarizations on the electric fields, as this is required to calculate $F(\kappa_y)$ [see Eqs. (15) and (21)]. Using the susceptibility of the metal, $\chi_m = \varepsilon_m - 1$, the polarization induced by an electric field is

$$\boldsymbol{P}(y,z) = \varepsilon_0(\chi_m - \chi_d)\theta[h(y) - z]\theta(z)\boldsymbol{E}(y,z). \tag{23}$$

Using Eq. (14) and grouping by directional unit vectors, this is written as

$$\boldsymbol{P}(y,z) = \varepsilon_0\theta[h(y) - z]\theta(z)\mathrm{L}\cdot\boldsymbol{C}(y,z), \tag{24}$$

where

$$\mathrm{L} = (\varepsilon_m - \varepsilon_d)(\hat{\boldsymbol{x}}\hat{\boldsymbol{x}} + \hat{\boldsymbol{y}}\hat{\boldsymbol{y}}) + \frac{\varepsilon_d(\varepsilon_m - \varepsilon_d)}{\varepsilon_m}\hat{\boldsymbol{z}}\hat{\boldsymbol{z}}. \tag{25}$$

In k space, the electric field requires a single grating scattering event, gaining or losing G, to set up the polarization. Thus, the relevant polarization terms in our equations are

$$\boldsymbol{Q}^O(\kappa_y) = \varepsilon_0 h_G\mathrm{L}\cdot\boldsymbol{C}^I(\kappa_y - G,0), \tag{26}$$

$$\boldsymbol{Q}^I(\kappa_y) = \varepsilon_0 h_{-G}\mathrm{L}\cdot\boldsymbol{C}^O(\kappa_y + G,0). \tag{27}$$

Here we make one of the fundamental assumptions of this model: We explicitly take the electric field and the polarization in the region of the grating to be uniform in z. That is, we assume that $\boldsymbol{C}(\kappa,z) \approx \boldsymbol{C}(\kappa,0)$. A more sophisticated assumption can be made, and the functional form of the field expanded and used in Eq. (24); if some knowledge of the form of the field is known *a priori*, for example, from FDTD simulations, then this could lead to a more accurate, yet complicated, solution.

C. Field at κ_I

To correctly determine the SPP field, the electric field at the incident wave vector, κ_I, must be known, as this is the field that induces $\boldsymbol{P}^O$ [Eq. (26)]. Once energy passes into the metal (i.e., for $z < 0$), it is absorbed and subsequently lost to heat, and therefore only the field above the metal surface needs be

determined. However, the incident field does not vanish here, and consequently all terms need to be considered.

In the region of the grating, at κ_I, using Eq. (3) and noting that $\theta(z-z')+\theta(z'-z)=1$, Eq. (11) can be written as

$$\begin{aligned}\boldsymbol{C}^I(\kappa_y,z) &= \hat{\boldsymbol{p}}^I_{d-}E_{\rm inc}(\kappa_y)+\hat{\boldsymbol{p}}^I_{d+}r^I_{dm}E_{\rm inc}(\kappa_y)+\frac{i}{2\varepsilon_0\varepsilon_d}\\ &\times\left(\frac{\kappa_I^2}{w_d^I}\hat{\boldsymbol{z}}\hat{\boldsymbol{z}}+w_d^I\hat{\boldsymbol{\kappa}}\hat{\boldsymbol{\kappa}}+\frac{\tilde{\omega}^2\varepsilon_d}{w_d^I}\hat{\boldsymbol{p}}^I_{d+}r^I_{dm}\hat{\boldsymbol{p}}^I_{d-}\right)\boldsymbol{Q}^I(\kappa_y)\\ &-\frac{i\kappa_I}{2\varepsilon_0\varepsilon_d}(\hat{\boldsymbol{z}}\hat{\boldsymbol{\kappa}}+\hat{\boldsymbol{\kappa}}\hat{\boldsymbol{z}})\int_0^{h_0}\theta(z-z')\boldsymbol{P}^I(\kappa_y,z')\,dz'\\ &+\frac{i\kappa_I}{2\varepsilon_0\varepsilon_d}(\hat{\boldsymbol{z}}\hat{\boldsymbol{\kappa}}+\hat{\boldsymbol{\kappa}}\hat{\boldsymbol{z}})\int_0^{h_0}\theta(z'-z)\boldsymbol{P}^I(\kappa_y,z')\,dz'.\end{aligned}$$

The sum of the final two terms in this expression is identically 0 if the polarization is assumed to be uniform in z. Even for arbitrary polarizations, the sum of these terms at $z=0$ is the negative value of the sum of the terms at $z=h_0$. Hence, we neglect these terms and simplify the field in the grating region to

$$\begin{aligned}\boldsymbol{C}^I(\kappa_y,z) &= \hat{\boldsymbol{p}}^I_{d-}E_{\rm inc}(\kappa_y)+\hat{\boldsymbol{p}}^I_{d+}r^I_{dm}E_{\rm inc}(\kappa_y)\\ &+\frac{1}{\varepsilon_0\varepsilon_d}\mathrm{N}\cdot\boldsymbol{Q}^I(\kappa_y),\end{aligned}\tag{28}$$

where

$$\mathrm{N}=\frac{i}{2}\left(\frac{\kappa_I^2}{w_d^I}\hat{\boldsymbol{z}}\hat{\boldsymbol{z}}+w_d^I\hat{\boldsymbol{\kappa}}\hat{\boldsymbol{\kappa}}+\frac{\tilde{\omega}^2\varepsilon_d}{w_d^I}\hat{\boldsymbol{p}}^I_{d+}r^I_{dm}\hat{\boldsymbol{p}}^I_{d-}\right).\tag{29}$$

To calculate the field above the grating, we first introduce the envelope function for the incident field,

$$E_{\rm inc}(\kappa_y)=f_{\rm inc}(\kappa_y-\kappa_I),\tag{30}$$

which is taken to be strongly peaked at $\kappa_y=\kappa_I$. We insert this envelope function into Eq. (10), evaluate all other quantities on the right-hand side at κ_I and use Eqs. (7) and (27)to determine the field above the grating,

$$\begin{aligned}\boldsymbol{E}^I(\kappa_y+\kappa_I,z) &= \hat{\boldsymbol{p}}^I_{d-}f_{\rm inc}(\kappa_y)e^{-iw_d^Iz}\\ &+\hat{\boldsymbol{p}}^I_{d+}e^{iw_d^Iz}\left[r^I_{dm}f_{\rm inc}(\kappa_y)+s^Of_O(\kappa_y+\Delta)\right],\end{aligned}\tag{31}$$

where

$$s^O=\frac{i\tilde{\omega}^2h_{-G}}{2w_d^I}\left(\hat{\boldsymbol{p}}^I_{d+}+r^I_{dm}\hat{\boldsymbol{p}}^I_{d-}\right)\cdot\mathrm{L}\cdot\hat{\boldsymbol{p}}^O_{d+}.\tag{32}$$

Hence, the field above the grating, near κ_I, has two components: the first, propagating downward, is the incident field; the second, propagating upward, is the reflected field. This reflection is further split into two terms: the first, with coefficient r^I_{dm}, is the reflection that would occur from a planar interface, while the second, with coefficient s^O, gives the correction due to the presence of the grating. In essence, this second term represents the feedback from the SPP field to the field at κ_I.

D. Self-consistency

Now that the fields at κ_I are known, we can use them to find the closed dynamical equation for the SPP field. As it is the field in the region of the grating that leads to the polarization that leads to the SPP field, we insert Eqs. (7) and (26)–(28) into Eq. (21), resulting in

$$F(\kappa_y)=\Gamma^Xf_{\rm inc}(\kappa_y-\Delta)+\Gamma^Sf_O(\kappa_y),\tag{33}$$

where

$$\Gamma^X=\frac{i\tilde{\omega}^2\rho_{dm}h_G}{2w_d^O}\hat{\boldsymbol{p}}^O_{d-}\cdot\mathrm{L}\cdot\left(\hat{\boldsymbol{p}}^I_{d-}+\hat{\boldsymbol{p}}^I_{d+}r^I_{dm}\right),\tag{34}$$

$$\Gamma^S=\frac{i\tilde{\omega}^2\rho_{dm}h_Gh_{-G}}{2\varepsilon_dw_d^O}\hat{\boldsymbol{p}}^O_{d-}\cdot\mathrm{L}\cdot\mathrm{N}\cdot\mathrm{L}\cdot\hat{\boldsymbol{p}}^O_{d+},\tag{35}$$

which, for convenience, are expanded and simplified in Appendix B. Transforming back into real space,

$$F(y)=\Gamma^Xe^{i\Delta y}f_{\rm inc}(y)+\Gamma^Sf_O(y).\tag{36}$$

These are inserted into Eq. (22) to arrive at the final, dynamic equation for the SPP envelope function,

$$\frac{df_O(y)}{dy}=[-\gamma+i\Gamma^S]f_O(y)+i\Gamma^Xe^{i\Delta y}f_{\rm inc}(y),\tag{37}$$

a one-dimensional ordinary differential equation. Thus, given a grating and excitation geometry, as well as the incident field, $f_{\rm inc}$, the SPP field, f_O, can be calculated by solving Eq. (37); this can be done analytically for a large number of cases and, failing that, numerically.

The physical significance of the three terms in Eq. (37) is as follows: the two terms preceding f_O relate to the losses of the SPP fields. First are the ohmic losses intrinsic to the propagation of the SPP along a planar metal surface, which are given by γ. Second, Γ^S gives a correction that is composed of both the reradiation of the SPP into a propagating mode and a correction due to the redistribution of the field in the presence of a grating. The latter may be either a further loss or even a gain if the electric field distribution is such that the field amplitude in the metal is lower than would be the case without a grating; consequently, the net effect of Γ^S can be either to increase or to decrease the losses. The final term in the equation, Γ^X, gives the strength of the grating-assisted coupling of free-space radiation to the SPP mode. Often, a comparison of the relative magnitude of these three terms will suffice to determine the relative efficiency of the grating coupler, without having to completely solve the system; an example is given in Sec VII.

E. Intensities and plane-wave excitation

In the previous section we developed a theory to calculate the electric fields, both reflected and SPP, associated with a metallic grating coupler. However, it is the relevant intensities, not fields, that are most often measured, or even calculated numerically. Here, we give expressions for the intensities associated with this problem. We also give the form of the solution for the intensities under excitation by a plane wave. This is useful as numerical models, such as the C-Method and RCWA, assume a plane-wave excitation; consequently, for us to test the accuracy of our method, we must do the same.

Given an incident field envelope, $f_{\rm inc}\,(f1y)$, the incident intensity is

$$I_{\rm inc}(y)=|f_{\rm inc}(y)|^2.\tag{38}$$

The reflected intensity can be read directly from the latter part of Eq. (31),

$$I_{\text{ref}}(y) = \left| r^I_{dm} f_{\text{inc}}(y) + s^O e^{-i\Delta y} f_O(y) \right|^2, \tag{39}$$

and the SPP intensity is

$$I_{\text{SP}}(y) = |f_O(y)|^2. \tag{40}$$

It is the coefficient of reflection,

$$R = \frac{I_{\text{ref}}}{I_{\text{inc}}}, \tag{41}$$

predicted by our method that we compare with the numerical models. Plane-wave excitation requires that $f_{\text{inc}}(y) = f_{\text{inc}}$, and with this used in Eq. (37) it is evident that the SPP envelope must have the form

$$f_O(y) = F_O e^{i\Delta y}. \tag{42}$$

Solving the differential equation yields the following SPP envelope function:

$$f_O(y) = \frac{i\Gamma^X e^{i\Delta y}}{\gamma - i\Gamma^S + i\Delta} f_{\text{inc}}, \tag{43}$$

and consequently, from Eq. (39), the coefficient of reflection for the grating coupler, due to plane-wave excitation, is

$$R = \left| r^I_{dm} + \frac{i\Gamma^X s^O}{\gamma - i\Gamma^S + i\Delta} \right|^2. \tag{44}$$

Note that for a planar interface, the coefficient of reflection would be simply, $R_{\text{planar}} = |r^I_{dm}|^2$, and it is the second term in R that accounts for the effect of the grating and the presence of the SPP. Thus, a comparison of R and R_{planar} will yield information into the magnitude of the SPP coupling.

V. USER'S GUIDE

Here we present a brief summary of the method. Some of the equations that we refer to are given in the Appendixes.

1. From the given grating parameters, $h(y)$ and Λ, calculate the grating wave vector, G, and Fourier coefficients, $h_{\pm G}$, of Eq. (1).
2. From the material properties as well as the angle of incidence, calculate $\tilde{\omega}$ and κ_I and use these to find κ_{SP} [Eq. (6)] and Δ [Eq. (7)].
3. Split κ_{SP} into $\kappa_O = \Re\{\kappa_{\text{SP}}\}$ and $\gamma = \Im\{\kappa_{\text{SP}}\}$.
4. Find the w's for both materials and at both κ_I and κ_O using Eq. (2).
5. With these, find the Fresnel coefficients and ρ_{dm} from Eqs. (A1) and (A5), respectively.
6. Calculate Γ^X [Eq. (B1)], Γ^S [Eq. (B2)], and s^O [Eq. (B3)].
7. With these, solve the differential equation [Eq. (37)] to find the SPP envelope, $f_O(y)$. Note that a solution for plane-wave excitation has already been given in Eq. (43).
8. Finally, solve for the reflected intensity [Eq. (39)]. This can then be used to find the coefficient of reflection, as we have done for excitation by a plane wave [Eq. (44)].
9. If required, a discussion of the range of grating amplitudes for which this method is valid is presented at the end of Sec. VI A.

VI. COMPARISON WITH NUMERICAL METHODS

To ensure the validity of our model we compare our results with those obtained via C-Method calculations[8] and RCWA calculations.[11] We first compare simulations of a sinusoidal grating between our method and the C-Method, then compare similar simulations of a square grating, contrasting our results with those from RCWA calculations.

A. Sinusoidal grating

Whereas we treat a grating as a small perturbation to an otherwise flat surface and extract an approximate analytic expression for the generated SPP field, the C-Method relies on a coordinate transformation that maps the periodic grating boundary to a flat interface. Unfortunately, the electromagnetic fields are also transformed, and an analytic solution is still impossible. However, by expanding the incident, reflected, and transmitted modes into a finite number of Bloch modes of the system, a solution can be calculated. Consequently, the C-Method yields very accurate reflection and transmission coefficients, although it is computationally intensive.

To contrast our results with those of the C-Method we must use incident plane waves to excite the SPP. We simulate a sinusoidal gold grating and assume that the adjacent dielectric is air. The grating period is 1200 nm, and we vary the amplitude between 0.1 and 60 nm to determine when our model breaks down. We use an incident wavelength $\lambda = 1000$ nm, at which gold has a dielectric constant[19] $\varepsilon_m = -41.8 + 2.95i$. For this wavelength the SPP resonance occurs at an angle of incidence 10.3°; consequently, we perform our simulations for angles ranging from 8° to 12° to ensure that we capture the entire resonance. Typical results are shown in Fig. 3. In (a) we display the p-polarized coefficient of reflection for a range of incident angles; both R_{planar} (dashed line) and R [solid line; Eq. (44)] are shown. As expected, near 10.3°, there is a dip in R, relative to R_{planar}. This is because R includes the grating effects, and consequently for this angle of incidence the momentum matching conditions are met, $\kappa_I + G = \kappa_O$, and some of the incident radiation is coupled to the SPP. Thus, some energy that would otherwise be present in the reflected beam—as in R_{planar}, which does not include grating effects—is lost.

In Fig. 3(b), we plot $\Delta R = R - R_{\text{planar}}$ for several different grating amplitudes; in essence, what is displayed here is solely the grating-induced adjustment to the reflected intensity and, as such, can be directly interpreted as the intensity transferred to the SPP by the grating. As expected, the higher the grating, the stronger its effect on the incident radiation and the more light that is coupled to the SPP.

We quantify the accuracy of our method by repeating the R and ΔR calculations, employing the C-Method, for the same structure and incident conditions. For the case of the 5-nm grating, we compare the results from the two methods in Fig. 4. Clearly, the magnitudes of the peaks are in excellent agreement: We calculate $(\Delta R)_{\text{max}} = 0.030\,39$, while C-Method calculations yield $(\Delta R)_{\text{max}} = 0.030\,07$, an error of only 1.0%. There are some differences between the curves produced by the two methods, which offer insight into the nature of the error. First, as is evident from the R curve

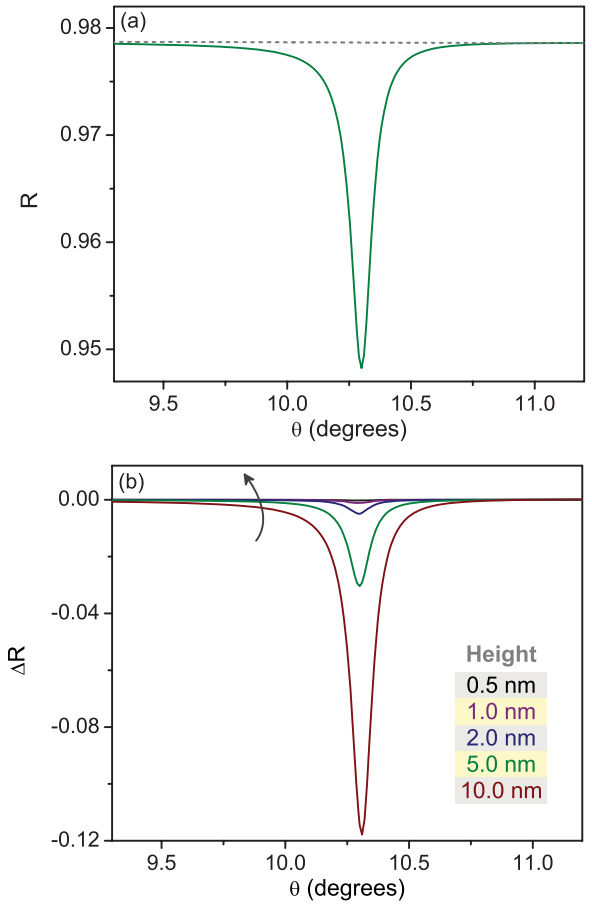

FIG. 3. (Color online) Typical results for our simulations. (a) Reflection spectra as a function of incidence angle for a gold film with a 5-nm sinusoidal grating (solid line) and for a planar gold film (dashed line), for $\lambda = 1000$ nm. (b) Different ΔR curves, as functions of the incident angle, for different amplitude gratings. The arrow points in the direction of decreasing heights, following the inset.

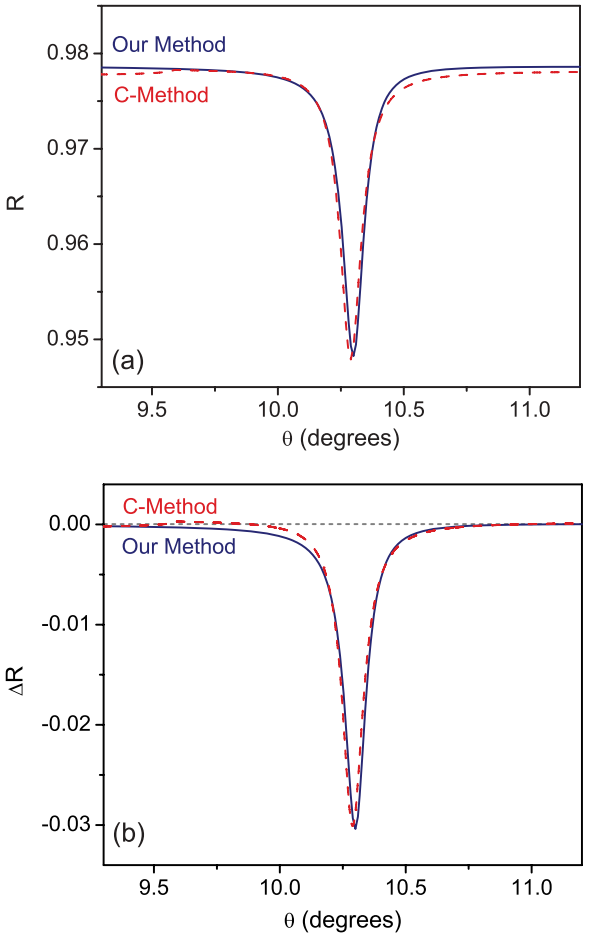

FIG. 4. (Color online) Comparison of (a) R and (b) ΔR curves calculated using both our method and the C-Method, for a 5-nm sinusoidal grating.

[Fig. 4(a)], there is a positive offset between the C-Method curve and ours. That is, we find more energy in the reflection, compared to the C-Method. This is easily explained, as the C-Method solves for all the propagating modes of the system, including the diffracted modes that we omit. Consequently, we attribute energy from the diffracted modes to the reflected mode. Since some energy from these diffracted modes also couples to the SPP, we expect that the error in the plasmonic coupling efficiency calculated with our method will be smaller than the error in the amount of reflected intensity. As we expect, and as we show here, the offset, and hence the error, becomes more significant as the grating amplitude increases and more light is coupled to the diffracted modes.

Second, as is more noticeable in the ΔR plot [Fig. 4(b)], the C-Method curve has features that our calculations do not. Specifically, for smaller incident angles, there is a bump near 9.7° that our work does not reproduce. Again, we explain this as being due to the full modal nature of the C-Method calculation. This bump corresponds to a Wood-Rayleigh anomaly, where a scattered mode propagates along the surface of the grating. Since our model includes only the reflected and SPP mode, this feature is not reproduced.

To determine the range of heights over which our method remains accurate, we repeat the preceding calculations for grating heights reaching 60 nm, comparing the amplitude of the dip as calculated by both methods (Fig. 5). From Fig. 5, it is evident that the two methods agree very well up to an amplitude of ∼50 nm, which, for $\lambda = 1000$ nm, is about four times longer than the skin depth of gold. Further, $(\Delta R)_{\max} > 0.5$ are predicted to within 10%, while $(\Delta R)_{\max} > 0.8$ are predicted to within 15%, which is very good considering the simplicity of our model and that our calculations are more than 250 times faster than those of the C-Method. Taking account of the energy coupled from the higher diffracted orders, we estimate a peak error in the plasmonic coupling efficiency of ∼13%.

For grating amplitudes in excess of 50 nm our model begins to break down. This breakdown occurs due to the linearity in h of the grating scattering [Eqs. (26) and (27)]. We remind the reader that this is a result of the assumption that the local field in the region of the grating is constant, which clearly becomes less valid as the grating amplitude increases. Consequently, as h increases, so do parameters such as Γ^S or Γ^X; where in reality these parameters must have a maximum value, in our model they can grow indefinitely, leading to breakdown. This

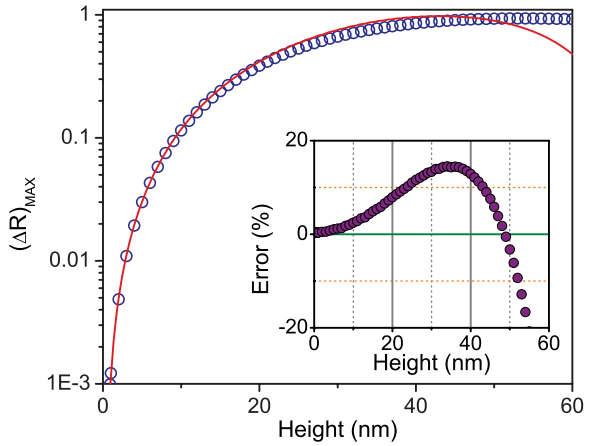

FIG. 5. (Color online) Magnitude of the dip in the reflection spectra calculated both with our method (line) and with the C-Method (circles) for excitation by a 1000-nm plane wave and a sinusoidal grating with a 1200-nm period and amplitudes ranging from 0.1 to 60 nm. Inset: Error between our method and the C-Method.

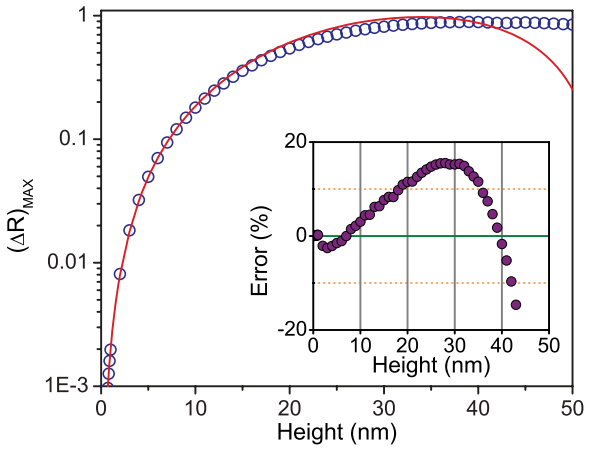

FIG. 6. (Color online) Magnitude of the dip in the reflection spectra calculated both with our method (line) and with the RCWA (circles) for excitation by a 1000-nm plane wave and a square grating with a 1200-nm period and amplitudes ranging from 0.1 to 50 nm. Inset: Error between our method and the RCWA.

manifests in one of two ways, depending on the sign of $\Im\{\Gamma^s\}$ [Eq. (35)].

1. $\Im\{\Gamma^s\} > 0$: This corresponds to grating-induced losses in the system. Since this constant grows like h^2, given our assumptions, at some point it will be comparable in value to the natural losses of the propagating plasmon γ. Since the product $s^O \Gamma^X$ also grows as h^2, near the resonant angle the ratio of this product to Γ^S will dominate the second term in R, [Eq. (44)], which will then approach or even exceed unity.

2. $\Im\{\Gamma^s\} < 0$: This corresponds to a decrease in losses in the system. While this might seem unreasonable, corresponding to gain in the system, this is not the case. In this model we assume that both the electric field and the polarization in the region of the grating are uniform. A consequence of this is that less field can be in the metal than would otherwise be the case for a planar film, and the lessening of the ohmic losses can dominate over the reradiation losses, resulting in what appears to be a gain term. As the grating amplitude increases, the assumption of electric field uniformity becomes increasingly erroneous, until $|\Im\{\Gamma^s\}| > \gamma$, and we obtain net gain and a catastrophic failure of the theory.

B. Square gratings

The previous section shows that our approach to plasmonic coupling calculations agrees with numerical methods for the case of a sinusoidal grating. However, this might not seem surprising, given that we only use the first Fourier coefficient of the grating profile and that a sinusoidal grating only has one component. This suggests changing the profile of the grating to a more complicated structure, one that has higher order Fourier components, and repeating the preceding analysis to see if the first-order coefficient still dominates. An obvious choice for a new surface profile is the square grating.

Since the only difference in our calculations will be in the Fourier components of the grating profile, we do not have to repeat the entire calculation. For a sinusoidal grating $h_{\pm G} = h_0/4$, while for a square grating $h_{\pm G} = h_0/\pi$, and consequently, the results of a calculation for a sinusoidal grating of height $h_0^{(\mathrm{sin})}$ will be identical to those of a square grating of height, $h_0^{(\mathrm{sq})} = \pi/4 \cdot h_0^{(\mathrm{sin})}$. This is a direct consequence of including only the first-order Fourier coefficient and can introduce additional errors for grating profiles with higher harmonics.

This sharp profile presents difficulties for simulations using the C-Method, and consequently we employ the RCWA approach. Results for a square grating, with the same parameters as the sinusoidal grating, are shown in Fig. 6. Clearly, the two theories are in good agreement, even for high coupling efficiencies. That the theory breaks down earlier than for the sinusoidal grating is expected, since the Fourier component for a given grating amplitude is large for the square grating. Further, the sharp edges of this profile tend to result in hot spots, areas where the field is concentrated, and consequently we would expect our assumption of a uniform field (and polarization) to be worse than for a smoothly varying grating. Regardless, the agreement is still very good to upward of 30-nm square gratings, where the coupling efficiency is in excess of 0.8.

Even with the improved algorithm, the RCWA technique does not converge quickly for all grating heights, and as such, Fig. 6 contains data calculated with both truncation number $N = 101$ and $N = 151$. Even for the prior case, the calculation time for our technique is $\sim 3.5 \times 10^3$ times faster than that for the RCWA, while for the latter case the ratio is $\sim 10^4$. For comparison, a calculation with $N = 41$ using the RCWA is still 500 times slower than using our method.

VII. EXAMPLE: SPATIAL EVOLUTION OF THE FIELDS

In the previous calculations we assumed plane-wave excitation, which allowed for direct comparison with numerical methods. Here, we present a more complex example that demonstrates the physical insight that can be gleaned from our approach. We use the same grating as in the previous sections, a sinusoidal gold grating with $\Lambda = 1200$ nm and $h_0 = 20$ nm adjacent to air, and consider illumination with $\lambda = 1000$-nm

radiation. However, instead of plane-wave excitation, we now assume a Gaussian beam with a full-width at half-maximum spot size of 5 μm. This allows for study of the spatial evolution of both the plasmon and the reflected field; $f_O(y)$ is calculated using Eq. (37), and $f_{\text{ref}}(y)$ from Eq. (31).

At $\lambda = 1000$ nm, where the dielectric function of gold[19] is $\varepsilon_m = -41.8 + 2.95i$, the surface plasmon wave vector is $\kappa_{\text{SP}} = (6.36 + 5.46 \times 10^{-3} i)\ \mu\text{m}^{-1}$. Consequently, the angle of incidence that excites an SPP, and that is used in this example, is 10.3°. Further, from the imaginary part of κ_{SP} we calculate a propagation length of 92 μm, while from the real part we calculate the plasmonic wavelength, $\lambda_{\text{SP}} = 988$ nm.

Before calculating $f_O(y)$ and $f_{\text{ref}}(y)$, we must find the constants for this problem. The solution to Eq. (37) in a region with no incident field is

$$f_O(y) \sim e^{-(\gamma - i\Gamma^S)y}, \tag{45}$$

and therefore the SPP intensity is

$$I_{\text{SP}}(y) \sim e^{-2(\gamma + \Im\{\Gamma^S\})y}. \tag{46}$$

Consequently, the effect of the grating on the decay of the SPP is given by Γ^S. For this situation $\Gamma^S = (5.37 \times 10^{-3} + 1.83 \times 10^{-4} i)\ \mu\text{m}^{-1}$, with the positive sign of the imaginary component corresponding to net losses due to reradiation and the field distribution in the grating region. The imaginary component of Γ^S leads to a propagation length of 2.7 mm, which is $\approx$100 times the intrinsic length of 92 μm.

From Eq. (37), it is evident that the coupling between the incident field and the SPP field is described by $i\Gamma^X$. Further, since the coupling from the incident wave to the SPP is linear, we define the coupling length, $\delta^+_{\text{SP}} = 1/\Im\{\Gamma^X\}$, as the distance over which the SPP envelope intensity will build to the same value as a constant incident intensity. Here, $\Gamma^X = (-5.05 \times 10^{-2} - 6.64 \times 10^{-3} i)\ \mu\text{m}^{-1}$, which results in $\delta^+_{\text{SP}} = 151\ \mu$m. It is the interplay between Γ^X and Γ^S that determines the total plasmonic intensity that can be generated, as the incident beam must be large enough to allow sufficient coupling but short enough to avoid unnecessary losses.

Finally, Eq. (31) predicts the reradiation of SPPs back into propagating modes. For this example, the reradiation constant is $s^O = 0.0030 - 0.0265i$, and, consequently, $|s^O|^2 = 7.1 \times 10^{-4}$ of the plasmonic intensity will be reradiated.

We solve Eq. (37) for $f_O(y)$ and use Eq. (31) to find $f_{\text{ref}}(y)$ and show the corresponding intensities in Fig. 7. In Fig. 7(a), we show both the incident, Gaussian intensity and the generated plasmonic intensity. The peak intensity of the SPP is 0.13 of the incident intensity, and as expected, it decays with a constant of 92 μm (see inset). In Fig. 7(b), we show both the incident intensity (solid curve) and the reflected intensity (dashed curve). In the bottom inset, the tail of the reflected intensity is clearly shown and is due to the reradiation of the SPP; from the dip in this curve we conclude that, as expected, the plasmon is radiating out of phase with the incident field. The top inset in Fig. 7(b) displays the difference between the incident and the reflected intensities, at their peaks.

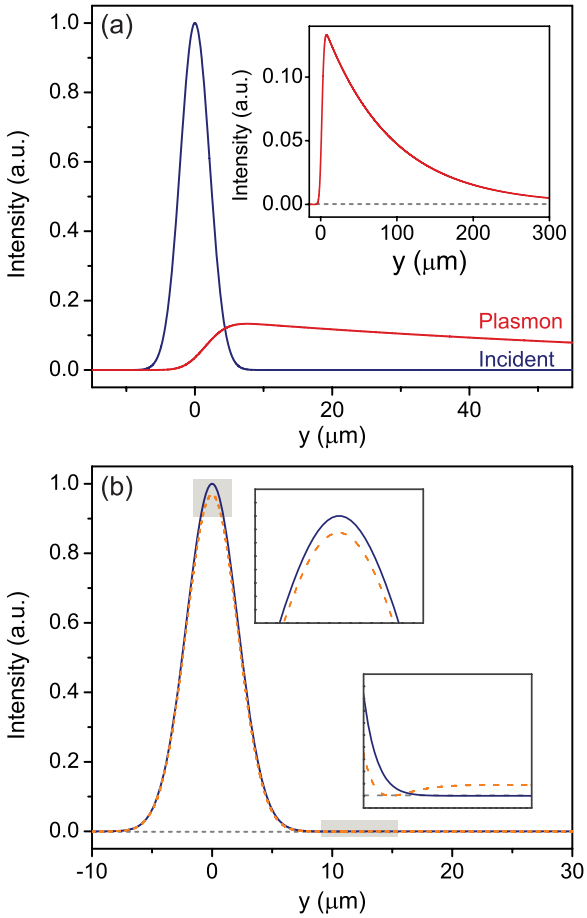

FIG. 7. (Color online) Spatial evolution of the intensities of this problem. (a) SPP intensity built up from the original incident, Gaussian intensity. Inset: Long-range decay of the SPP intensity. (b) Incident (solid) and reflected (dashed) intensities. Top inset: Zoom-in on the peak of the intensities. Bottom inset: Zoom-in on the tail end of the intensities, showing the reradiation of the SPP.

We find the reflectance for the Gaussian excitation according to

$$\Delta I = \frac{I_{\text{ref}} - I_{\text{inc}}}{I_{\text{inc}}}, \tag{47}$$

where

$$I_{\text{ref(inc)}} = \int I_{\text{ref(inc)}}(y)\, dy, \tag{48}$$

and $I_{\text{ref(inc)}}$ are the reflected and incident curves given in Fig. 7, respectively. For our example we find that $\Delta I = 0.034$. It is interesting to note that this value is much smaller than the peak intensity of the SPP. This disparity is readily understood, since the plasmonic mode is more tightly confinement than the incident and reflected modes.

These results do suggest another important consideration that is often neglected: the spatial extent of the coupling beam. In the literature it is often implicitly assumed that the size of the incident beam does not matter, possibly since most of the simulations are performed assuming plane waves. However, for the conditions in this example $R^{\text{PW}} = 0.263$, which is an order of magnitude larger than we calculate for a 5-μm Gaussian beam. Indeed, if a quarter of the energy of the

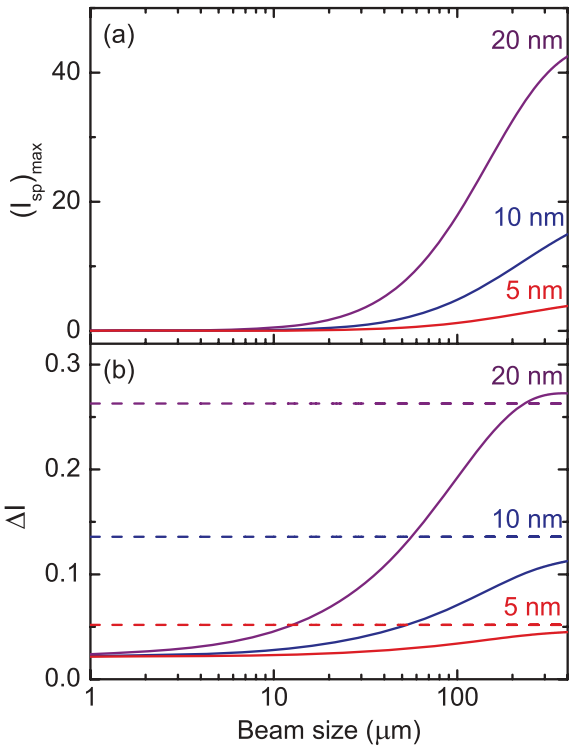

FIG. 8. (Color online) (a) Peak plasmonic intensity (assuming a peak incident intensity of unity) as a function of the incident beam size (FWHM) for grating amplitudes of 5, 10, and 20 μm. (b) ΔI, as defined in Eq. (47) (solid lines), and for plane-wave excitation [unity minus Eq. (44); dashed lines], as functions of the incident beam size for the same gratings as in (a).

incident beam is transferred to the SPP, which has a skin depth of the order of 500 nm, we would expect that the peak plasmonic intensity be *higher* than the peak incident intensity.

The disparity between the plane-wave and the Gaussian results is bridged if we remember that the plasmon intensity is not created instantly, but rather builds up as the plasmon propagated along the surface of the grating [see Eq. (37)], as long as the incident field is *present*. In essence, the beam size in this example is too small for a high plasmonic intensity to build up. To verify this we repeat the preceding calculations for beam sizes (FWHM) ranging from 1 to 400 μm (Fig. 8). We present both [Fig. 8(a)] the peak plasmonic intensity, keeping the peak incident intensity at unity, and [Fig. 8(b)] the R curves, for different amplitude gratings. Clearly, as the beam size grows, so too does the peak plasmonic intensity, increasing to more than 40 times the incident intensity for a 20-μm grating. Correspondingly, ΔI for a Gaussian beam also increases with the beam size, eventually surpassing that of a plane-wave excitation. This should not be too surprising since if the coupling and decoupling constants are sufficiently large, then non-negligible amounts of light can be reradiated in spatial regions where the incident beam is not present. As expected, lower amplitude gratings do not show this effect.

VIII. CONCLUSION

In conclusion, we have developed an analytic model that describes the coupling of light to SPPs via surface relief gratings. Our approach is based on the assumption that the fields in the grating region are uniform, and while this clearly fails to capture the behavior of the local fields as the grating height increases, we can accurately predict plasmonic coupling efficiencies as high as 0.8, to within a 15% error of standard numerical techniques, and with computational times that are more than 2 orders of magnitude shorter. Perhaps more importantly, we describe the process of grating-mediated plasmonic coupling with simple differential equations whose coefficients both have clear physical significance and are expressed in an analytic form in terms of constants that relate to the geometry and material properties of the system; this allows for a quick yet meaningful interpretation of the behavior of the grating coupler, as a function of these fundamental constants. Finally, and in contrast with previous models, we quantitatively describe the spatial growth and decay of the SPP field (and hence the spatial dependence of the reflected field) for arbitrarily shaped incident fields, not just plane waves. This is an important capability for the description of the effects of focused light on nanostructures.

We suggest that this is a powerful tool for future plasmonic research: From a practical perspective, this approach can be used to rapidly optimize a grating-based plasmonic coupler, with the final fine-tuning performed with a numerical method, cutting the total calculation times by orders of magnitude. Conversely, from a more fundamental perspective, this technique allows for quick and efficient investigations of the underlaying physics of grating-assisted plasmonic coupling.

ACKNOWLEDGMENTS

We gratefully acknowledge funding provided by the Natural Sciences and Engineering Research Council of Canada.

APPENDIX A: FRESNEL COEFFICIENTS

Using our notation, the Fresnel coefficients for p-polarized radiation incident on the dielectric-metal interface are

$$r_{ij} = \frac{w_i\varepsilon_j - w_j\varepsilon_i}{w_i\varepsilon_j + w_j\varepsilon_i}, \qquad t_{ij} = \frac{2n_i n_j w_i}{w_i\varepsilon_j + w_j\varepsilon_i}. \tag{A1}$$

Since SPP coupling is a resonant phenomena, the coefficients in Eqs. (A1) must have poles at wave vectors where SPPs are excited. To find these poles we note that the denominator, $f = w_d\varepsilon_m + w_m\varepsilon_d$, must vanish as $\kappa \to \kappa_{SP}$. Recalling Eq. (2), this results in the expected SPP dispersion relation:

$$\kappa_{SP} = \tilde{\omega}\left(\frac{\varepsilon_d\varepsilon_m}{\varepsilon_d + \varepsilon_m}\right)^{\frac{1}{2}}. \tag{A2}$$

Consequently, the Taylor expansion of the denominator about κ_{SP} is, to first order,

$$f(\kappa_y) \approx \left(\frac{\partial f}{\partial \kappa_y}\right)_{\kappa_{SP}} (\kappa_y - \kappa_{SP}), \tag{A3}$$

where

$$\left(\frac{\partial f}{\partial \kappa_y}\right)_{\kappa_{SP}} = -\kappa_{SP}\left(\frac{\varepsilon_m}{w_d} + \frac{\varepsilon_d}{w_m}\right)_{\kappa_{SP}}. \tag{A4}$$

The relevant coefficients to this problem can then be written as

$$r_{dm} \approx \frac{\rho_{dm}}{\kappa_y - \kappa_{\mathrm{SP}}},$$
$$t_{dm} \approx \frac{\tau_{dm}}{\kappa_y - \kappa_{\mathrm{SP}}},$$

where

$$\rho_{dm} = \frac{2\,(w_d \varepsilon_m)_{\kappa_{\mathrm{SP}}}}{\left(\frac{\partial f}{\partial \kappa_y}\right)_{\kappa_{\mathrm{SP}}}},$$
$$\tau_{dm} = \frac{2 n_d n_m\,(w_d)_{\kappa_{\mathrm{SP}}}}{\left(\frac{\partial f}{\partial \kappa_y}\right)_{\kappa_{\mathrm{SP}}}}. \tag{A5}$$

Similarly, this analysis can be repeated with the *s*-polarized Fresnel coefficients:

$$r^s_{ij} = \frac{w_i - w_j}{w_i + w_j},$$
$$t^s_{ij} = \frac{2 w_i}{w_i + w_j}. \tag{A6}$$

APPENDIX B: SIMPLIFIED PARAMETERS

The parameters Γ^X, Γ^S, and s^O are required for the calculation of the SPP field and coupling efficiency. Here we simplify these parameters, expressing them in terms of the fundamental constants of this problem.

To simplify the grating coupling term, Γ^X, we use Eqs. (3) and (25) in Eq. (34), while to simplify the decoupling term for the SPP field, Γ^S, we use Eqs. (3), (25), and (29) in Eq. (35). Finally, to simplify s^O we note the similarity between Eq. (32) and Eq. (34) and the commutability of the polarization vectors:

$$\Gamma^X = \frac{i \rho_{dm} h_G}{2 w_d^O}\left[\frac{\varepsilon_m - \varepsilon_d}{\varepsilon_m}\left(1 + r^I_{dm}\right)\kappa_I \kappa_O + \frac{\varepsilon_m - \varepsilon_d}{\varepsilon_d}\left(1 - r^I_{dm}\right) w^I_d w^O_d\right], \tag{B1}$$

$$\Gamma^S = -\frac{\rho_{dm} h_G h_{-G}}{4}\left[\left(\frac{\varepsilon_m - \varepsilon_d}{\varepsilon_m}\right)^2 \left(1 + r^I_{dm}\right)\frac{\kappa_I^2 \kappa_O^2}{w^I_d w^O_d} - 2\frac{(\varepsilon_m - \varepsilon_d)^2}{\varepsilon_d \varepsilon_m} r^I_{dm} \kappa_I \kappa_O - \left(\frac{\varepsilon_m - \varepsilon_d}{\varepsilon_d}\right)^2 \times \left(1 - r^I_{dm}\right) w^I_d w^O_d\right], \tag{B2}$$

$$s^O = \frac{i h_{-G}}{2 w^I_d}\left[\frac{\varepsilon_m - \varepsilon_d}{\varepsilon_m}\left(1 + r^I_{dm}\right)\kappa_I \kappa_O + \frac{\varepsilon_m - \varepsilon_d}{\varepsilon_d}\left(1 - r^I_{dm}\right) w^I_d w^O_d\right]. \tag{B3}$$

*Corresponding author. nrotenbe@physics.utoronto.ca

[1] Lord Rayleigh (J. W. Strutt), Proc. R. Soc. London Ser. A **79**, 399 (1907).
[2] A. Hessel and A. A. Oliner, Appl. Opt. **4**, 1275 (1965).
[3] P. M. van den Berg, J. Opt. Soc. Am. **71**, 1224 (1981).
[4] H. Numata, J. Phys. Soc. Japan **51**, 2575 (1982).
[5] P. Sheng, R. S. Stepleman, and P. N. Sanda, Phys. Rev. B **26**, 2907 (1982).
[6] A. V. Tishchenko, Opt. Express **17**, 17102 (2009).
[7] J. Chandezon, M. T. Dupuis, G. Cornet, and D. Maystre, J. Opt. Soc. Am. **72**, 839 (1982).
[8] L. Li, J. Chandezon, G. Granet, and J.-P. Plumey, Appl. Opt. **38**, 304 (1999).
[9] M. G. Moharam, E. B. Grann, D. A. Pommet, and T. K. Gaylord, J. Opt. Soc. Am. A **12**, 1068 (1995).
[10] M. G. Moharam, E. B. Grann, D. A. Pommet, and T. K. Gaylord, J. Opt. Soc. Am. A **12**, 1077 (1995).
[11] P. Lalanne and G. M. Morris, J. Opt. Soc. Am. A **13**, 779 (1996).
[12] K. S. Yee, IEEE Trans. Antennas Propag. **14**, 302 (1966).
[13] H. Raether, *Surface Plasmons*, edited by G. Hohler (Springer, Berlin, 1988).
[14] W. L. Barnes, A. Dereux, and T. W. Ebbesen, Nature (London) **424**, 824 (2003).
[15] N. Rotenberg, M. Betz, and H. M. van Driel, Opt. Lett. **33**, 2137 (2008).
[16] K. F. MacDonald, Z. L. Sámson, M. I. Stockman, and N. I. Zheludev, Nat. Photon. **3**, 55 (2009).
[17] J. E. Sipe, Solid State Commun. **39**, 493 (1981).
[18] J. E. Sipe, J. Opt. Soc. Am. B **4**, 481 (1987).
[19] P. B. Johnson and R. W. Christy, Phys. Rev. B **6**, 4370 (1972).