Scattering of surface plasmon polaritons at abrupt surface interfaces: Implications for nanoscale cavities

R. F. Oulton, D. F. P. Pile, Y. Liu, and X. Zhang*
NSF Nano-scale Science and Engineering Center, 3112 Etcheverry Hall, University of California, Berkeley, California 94720, USA
(Received 19 October 2006; revised manuscript received 16 January 2007; published 10 July 2007)

We have developed a rigorous mode matching approach for the analysis of surface wave scattering at nonuniform semi-infinite dielectric-metal interfaces. An analytical derivation of the coupling coefficients between surface waves allows an efficient scattering matrix formulation to describe general structures with multiple interfaces. Using this, we resolve issues of accuracy and convergence of related approaches in the literature. Studies of the reflection, transmission, and radiation of surface plasmons incident on both dielectric and metallic surface discontinuities show a correspondence with an effective Fresnel description. We also model a surface plasmon distributed Bragg reflector (DBR) capable of reflecting between 80% and 90% of incident surface plasmon power. Radiation mode scattering ultimately limits the DBR's reflection performance rather than the intrinsic absorption of the metal. Thus, alternative plasmonic geometries that suppress radiation modes, such as gap and channel structures, could be superior for the design of strongly reflective DBRs for integration in high-Q-factor nanoscale cavities. We anticipate that this method will be a valuable tool for the efficient and intuitive design of plasmonic devices based on structural nonuniformities.

DOI: 10.1103/PhysRevB.76.035408 PACS number(s): 73.20.Mf, 42.25.Bs, 42.70.Qs

I. INTRODUCTION

Surface plasmon polaritons[1,2] have recently re-emerged as unique and promising excitations for efficiently concentrating and delivering electromagnetic energy to the nanoscale: "plasmonics" now offers unique opportunities for subwavelength optical waveguides,[3–5] new sensors and detection techniques based on surface-enhanced Raman scattering,[6,7] and subvacuum wavelength superlens imaging[8] amongst others. Most recently, Miyazaki and Kurokawa[9] experimentally demonstrated a nanoscale metallic gap-plasmon based cavity whose fundamental resonant length was a tenth of the vacuum wavelength. Such observations have reinvigorated new research into nanoscale light emitting devices such as surface plasmon amplification by stimulated emission of radiation[10] and nanolasers.

Many authors have already experimentally and numerically examined components necessary to construct nanoscale cavities. For example, Krenn *et al.* introduced a series of two dimensional components based on periodic arrays of gold nanoparticles for achieving extremely efficient Bragg reflection of surface plasmons.[11] Introducing surface discontinuities in this way naturally points to the potential of building surface wave cavities. An excellent review by Zayats *et al.* discusses such aspects of surface wave optics.[12] A similar principle lies behind the aproach of Miyazaki and Kurokawa:[9] here, the discontinuities at the terminated ends of a gap-plasmon waveguide provide the necessary modal reflection. Most recently, three numerical studies considered nanoscale thin-film[13,14] and channel plasmon waveguide[15] cavities employing some form of surface discontinuity to achieve cavity feedback. However, these past works are predominantly experimental or brute force studies that somewhat disguise important insight into these phenomena.

Plasmonics research today employs a wide variety of numerical techniques for the analysis of highly confined electromagnetic fields. The most popular tools are the finite-difference time-domain (FDTD) and finite-element (FEM) methods. Unfortunately, commercial implementations tend to be inflexible and do not clearly reveal the numerical methods employed, and homemade versions require prodigious programing time and effort for any sophisticated analysis of complex structures. All FDTD and FEM algorithms require vast computational resources to solve even small three-dimensional problems. Moreover, these methods do not lend themselves well to intuitive analysis or clear interpretation of the resultant field profiles they generate. Analytical and semi-analytical methods are still applicable in many plasmonic systems; in these cases, the greater insight and computation efficiency afforded can be crucial for the development of novel devices. The most prevalent in current plasmonics literature are the Green's functions approach[16–18] and discrete dipole approximation approach.[19,20] While providing all the benefits of their analytical nature, these methods are extremely complicated to implement in general cases and are usually applied to the problem of scattering from single or periodic surface defects.

The present paper uses a rapidly converging and accurate semianalytical mode matching[21] method for solving systems of plasmons interacting with multiple abrupt structural metal and/or dielectric half space transitions. The method is amenable to general problems involving both the generation and scattering of surface plasmon waves by employing a scattering matrix formalism. Our mode matching approach demonstrates excellent convergence even for highly confined surface plasmon modes with significant penetration into the metal and directly provides modal coupling coefficients (i.e., reflection, transmission, and radiation scattering). The key aspect of accuracy and correct convergence is addressed through comparisons with commercial software and past work in the literature. Strengths of surface wave reflection, transmission, and radiation scattering at single and multiple abrupt discontinuities of dielectric (Sec. IV A) and metallic (Sec. IV B) permittivities are determined and analyzed. A

surface plasmon distributed Bragg reflector (DBR) capable of reflecting between 80% and 90% of incident surface plasmon power is presented (Sec. IV C) and analyzed.

Schevenchko (see Ref. 22 and references therein) originally laid out analytical methodology for describing open waveguide systems, proving the mathematical completeness of the eigenmode expansion and describing approximate analytical solutions to problems involving continuous surface variations at radio frequencies (rf). Later, Mahmoud and Beal[23] combined these analytical techniques with the numerical mode matching approach of Clarricoats and Slinn.[21] The motivation for their work concerned the detection of abrupt discontinuities on dielectric loaded rf transmission lines, where an analytical solution was not tangible due to the highly coupled nature of the eigenmodes. Unfortunately, their result employed a very low number of radiation modes (the continuous portion of the eigen-mode expansion): with two or three Laguerre polynomials describing the radiation fields, the convergence of their approach is surprising in light of the numerical analysis in this paper. Stegeman *et al.*[24,25] also solved similar problems but did not follow the mode matching approach, instead considering boundary matching of the transverse fields at discrete points along a discontinuity. Their approach suffered two numerical issues: firstly, sufficient accuracy required a large number of boundary matching points, which consequently limited the number of eigenmodes that they could consider to 4; secondly, in order to obtain convergence, the authors had to introduce an artificial grounding plate reducing the problem to essentially a closed waveguide system, which limits the description of scattering loss. The authors also commented on the poor convergence properties of treating the open waveguide system in a similar fashion to Mahmoud and Beal, which, in light of the present study, certainly merits closer attention. Voronko *et al.*[26] also considered a related problem of surface discontinuities of metal and/or dielectric half spaces with an approach similar to the mode matching method. However, the authors made two simplifications: firstly, they neglected inter-radiation mode scattering, which is not a valid approximation for large variations in permittivity at the discontinuity, and secondly, they solved the integral equations for the single interface directly, which limits further adaptation for the description of multiple discontinuities. Finally, as a general observation of the early work on this problem, detailed studies of specific results are missing from the literature and what studies do exist are for the rf band.

II. MODE MATCHING SURFACE WAVES

The mode matching method[21] operates on the premise that a complete set of orthogonal eigenmodes on either side of a discontinuity under study is describable analytically or numerically. The completeness of the set ensures a consistent map between the eigenmode expansions on either side of the discontinuity. Section II A derives the set of eigenmodes of the open metal-dielectric waveguide system for use in the mode matching implementation. (The eigenmodes of this geometry are relatively straightforward to define; however, their normalization and proof of completeness are not. The reader should consult the book of Schevchenko[22] for more information on these issues.) Following this, definition of a scalar product of the orthonormal basis functions provides the mode normalization constants, proof of mutual mode orthogonality, and the coupling coefficients between the eigenmodes of different regions (Sec. II B). Sections II C and II D treat the implementation of the mode matching method, deriving the coupled mode equations and numerical approach to solve the problem using scattering matrices.

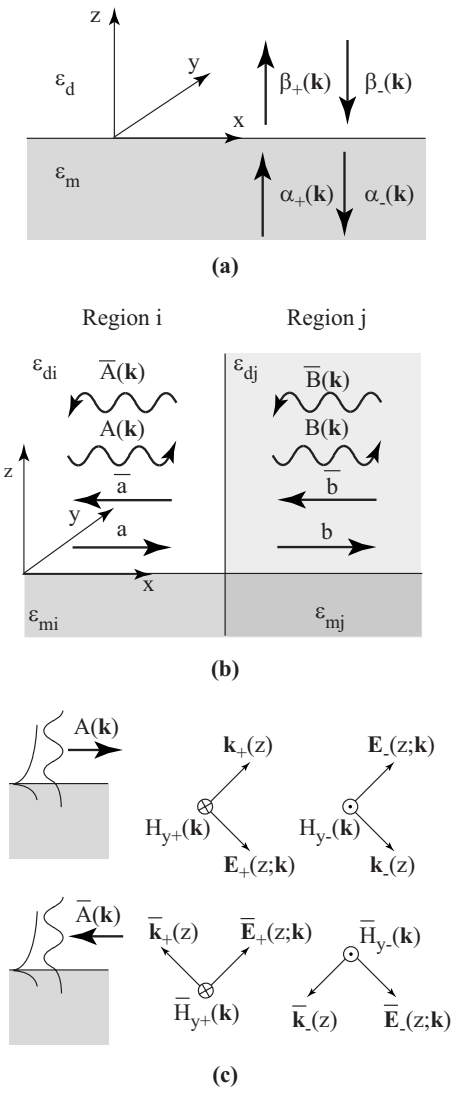

FIG. 1. (a) Schematic of the metal-dielectric open waveguide system. (b) Schematic of the surface discontinuity problem highlighting the required modal amplitudes for both guided surface and radiating volume waves. (c) Convention of coupled plane wave orientations for forward and backward propagations in both the $\hat{z}$ and $\hat{x}$ directions.

A. Eigenmode fields

Consider the geometry in Fig. 1(a) consisting of two material half spaces, one of which is metallic, supporting *surface wave* eigenmodes that propagate in the *x*-*y* plane. The geometry is invariant in the $\hat{y}$ direction. Throughout this pa-

per, the term surface wave applies to any of the set of bound and radiation eigenmodes of the inhomogeneous metal-dielectric open waveguide geometry shown in Fig. 1(a). Note that for brevity, bound surface waves and radiation surface waves are also referred to as *surface plasmons* and *radiation modes*, respectively. Physically, each surface wave is constructed from up to four interfering plane waves [a_+, a_-, b_+, and b_- in Fig. 1(a)]. Later, the paper considers the scattering of these groups of plane waves at discontinuities of the form shown in Fig. 1(b), evaluating parameters such as the reflection, transmission, and scattering. In such problems, it is necessary to consider both forward ($+\hat{x}$) and backward ($-\hat{x}$) propagating surface waves. Since surface waves are groups of up to four plane waves each, there are a possible 16 interacting plane waves for a single surface discontinuity. To help distinguish plane waves and surface waves, Greek letters and Roman letters are used for the amplitudes, respectively. In addition, Fig. 1(c) helps us to simplify the analysis by providing the relative signs of the plane waves that constitute the forward and backward propagating surface waves.

The derivation in this section identifies the relative amplitudes of the plane waves that constitute each surface wave for the transverse magnetic (TM) polarization. It is important to note here that in cases where the incident field is not invariant in the $\hat{y}$ direction (i.e., for surface waves impinging on a surface discontinuity at an angle), interpolarization coupling is important and TE radiation modes are necessary in the subsequent analysis.[25]

Following Schevchenko's prescription for this problem,[22] the set of surface waves consists of a single bound surface mode and a continuous set of radiation modes that form a complete orthogonal set. The derivation assumes an $\exp(-i\omega t)$ time variation of the field. Solving the wave equation independently in both the metallic and dielectric regions yields a continuous set of forward and backward propagating plane waves in the $\pm\hat{z}$ direction. For TM plane waves, the unit magnetic field vector $\hat{\mathbf{H}}_\pm(\mathbf{k})$ is

$$\hat{\mathbf{H}}_\pm(\mathbf{k}) = \pm\left\{-\frac{k_y}{V}, \frac{k_x}{V}, 0\right\}. \tag{1}$$

Here, the choice of sign is related to the direction of wave propagation in the $\hat{z}$ direction, and $V=\sqrt{k_x^2+k_y^2}$. Note that invariance of the structure in the $\hat{x}$ and $\hat{y}$ directions in the current calculation implies conservation of both k_x and k_y, which is equivalent to Snell's law. The total magnetic field $\mathbf{H}(\mathbf{r};\mathbf{k})$, taking into account both forward and backward propagating amplitudes, is

$$\begin{aligned}\mathbf{H}(\mathbf{r};\mathbf{k}) &= [\beta_+(\mathbf{k})\exp(ik_{zd}z)\hat{\mathbf{H}}_+(\mathbf{k}) \\ &\quad + \beta_-(\mathbf{k})\exp(-ik_{zd}z)\hat{\mathbf{H}}_-(\mathbf{k})]f(x,y), \quad z \geq 0, \\ \mathbf{H}(\mathbf{r};\mathbf{k}) &= [\alpha_+(\mathbf{k})\exp(ik_{zm}z)\hat{\mathbf{H}}_+(\mathbf{k}) \\ &\quad + \alpha_-(\mathbf{k})\exp(-ik_{zm}z)\hat{\mathbf{H}}_-(\mathbf{k})]f(x,y), \quad z < 0,\end{aligned} \tag{2}$$

where the subscript labels m and d indicate parameters of the metal ($z<0$) and dielectric ($z\geq 0$) regions, respectively, and $f(x,y)=\exp(ik_x x+ik_y y)$ describes the harmonic variation in the plane. Using the Ampere-Maxwell law, $[\varepsilon(z)k_0/z_0]\mathbf{E}(\mathbf{r};\mathbf{k})=\mathbf{k}\times\mathbf{H}(\mathbf{r};\mathbf{k})$, the vector electric field $\mathbf{E}(\mathbf{r};\mathbf{k})$ is

$$\begin{aligned}\mathbf{E}(\mathbf{r};\mathbf{k}) &= \frac{z_0}{\sqrt{\varepsilon_d}}[\beta_+(\mathbf{k})\exp(ik_{zd}\cdot\mathbf{r})\hat{\mathbf{E}}_+(z;\mathbf{k}) \\ &\quad + \beta_-(\mathbf{k})\exp(-ik_{zd}z)\hat{\mathbf{E}}_-(z;\mathbf{k})]f(x,y), \quad z \geq 0, \\ \mathbf{E}(\mathbf{r};\mathbf{k}) &= \frac{z_0}{\sqrt{\varepsilon_m}}[\alpha_+(\mathbf{k})\exp(ik_{zm}\cdot\mathbf{r})\hat{\mathbf{E}}_+(z;\mathbf{k}) \\ &\quad + \alpha_-(\mathbf{k})\exp(-ik_{zm}z)\hat{\mathbf{E}}_-(z;\mathbf{k})]f(x,y), \quad z < 0,\end{aligned} \tag{3}$$

where the unit electric field vector is

$$\hat{\mathbf{E}}_\pm(z;\mathbf{k}) = \frac{1}{\sqrt{\varepsilon(z)}k_0}\left\{\frac{k_x k_z(z)}{V}, \frac{k_y k_z(z)}{V}, \mp V\right\}. \tag{4}$$

Here, z_0 is the impedance of free space and k_0 is the free-space wave number. Since the metallic region is strongly absorbing, the analysis ignores fields propagating from $-\infty$ within the metal such that $\alpha_+(\mathbf{k})=0$. Continuity of $H_y(\mathbf{r};z=0;\mathbf{k})$ and $E_x(\mathbf{r};z=0;\mathbf{k})$ therefore implies that

$$\begin{aligned}\alpha_-(\mathbf{k}) &= \beta_-(\mathbf{k}) - \beta_+(\mathbf{k}), \\ \alpha_-(\mathbf{k})\frac{k_{zm}}{\varepsilon_m} &= [\beta_+(\mathbf{k}) + \beta_-(\mathbf{k})]\frac{k_{zd}}{\varepsilon_d}.\end{aligned} \tag{5}$$

In this case, the full field functions are

$$\begin{aligned}\mathbf{H}(\mathbf{r};\mathbf{k}) &= -N(\mathbf{k})\psi_-(z)\hat{\mathbf{H}}_+(\mathbf{k})f(x,y), \\ \mathbf{E}(\mathbf{r};\mathbf{k}) &= \frac{N(\mathbf{k})z_0}{\varepsilon(z)k_0}\left\{\psi_+(z)\frac{k_x k_z}{V}, \psi_+(z)\frac{k_y k_z}{V}, \psi_-(z)V\right\}f(x,y),\end{aligned} \tag{6}$$

where

$$\begin{aligned}\psi_\pm(z) &= r\exp(ik_{zd}z) \pm \exp(-ik_{zd}z), \quad z \geq 0, \\ \psi_\pm(z) &= \pm(1-r)\exp(-ik_{zm}z), \quad z < 0.\end{aligned} \tag{7}$$

Here, the amplitudes for the plane wave's components equate to $\beta_-(\mathbf{k})=N(\mathbf{k})$, a normalization factor, and $\beta_+(\mathbf{k})/\beta_-(\mathbf{k})=r(\mathbf{k})$, the reflectivity of the dielectric-metal interface:

$$r(\mathbf{k}) = \frac{\varepsilon_d k_{zm} - \varepsilon_m k_{zd}}{\varepsilon_d k_{zm} + \varepsilon_m k_{zd}}. \tag{8}$$

Equations (6) and (7) along with their associated dispersion relations provide the mode fields for this geometry describing the propagation of surface waves along the $\pm\hat{x}$ and $\hat{y}$ directions. The guided mode has the following dispersion relation arising from the condition that $\beta_+(\mathbf{k})=0$, irrespective of $\beta_-(\mathbf{k})$ and $\alpha_-(\mathbf{k})$: i.e., $r(\mathbf{k})\mapsto\infty$. This provides the following well known dispersion relation for surface plasmons:

$$k_{x,g} = k_0 \sqrt{\frac{\varepsilon_m \varepsilon_d}{\varepsilon_m + \varepsilon_d}}. \tag{9}$$

The surface plasmon field functions arise from Eqs. (6) and (7) by direct substitution for $r(\mathbf{k}) \mapsto \infty$ and $k_x = k_{x,g}$. The dispersion relation for the radiation modes is continuous with $k_{z,r}(z) \in [0, \infty]$. Note that for $k_{zd,r} > \sqrt{\varepsilon_d} k_0$, radiation modes are evanescent in the $\hat{x}$ direction.

The mode matching analysis that follows involves both forward and backward surface waves in the $\pm\hat{x}$. As discussed at the beginning of this section, surface waves consist of up to four interfering plane waves, so some confusion can arise with respect to the relative signs of the various field vectors for the composite surface waves. However, the lengths of the field vectors for forward and backward propagating surface waves in a particular region are identical; only sign changes of some of the field components distinguish the propagation direction. Figure 1(c) highlights these sign changes. The upper two plane wave vector triads describe forward propagating surface waves and the lower two, backward propagating surface waves. Taken individually, the relative signs of the field components of counterpropagating plane waves in the $\pm\hat{z}$ direction can be seen. Taken as upper and lower groups, the relative signs of the field components of counterpropagating surface waves in the $\pm\hat{x}$ direction can be seen. In the analysis that follows, a bar over a field vector or amplitude represents a backward propagating surface wave traveling in the $-\hat{x}$ direction. The barred field components for surface waves are therefore

$$\bar{\mathbf{H}}(\mathbf{r};\mathbf{k}) = \{-H_x(\mathbf{r}), H_y(\mathbf{r}), 0\},$$

$$\bar{\mathbf{E}}(\mathbf{r};\mathbf{k}) = \{E_x(\mathbf{r}), -E_y(\mathbf{r}), -E_z(\mathbf{r})\}. \tag{10}$$

B. Mode orthogonality, normalization, and coupling coefficients

Knowledge of the set of surface waves of the geometry in Fig. 1(a) allows the expansion of any electromagnetic field distribution in that region. In order to describe the field distribution at the interface of two such regions (i,j), as shown in Fig. 1(b), the various amplitudes of the two eigenmode expansions must be determined. The integral expression in Eq. (11) is a generalization of the orthogonality integral[28] for electromagnetic eigenmodes. Here, it is extended to describe coupling between the eigenmodes of the two regions in question: it is essentially a scalar product acting on the union of the two eigenmode spaces. Later on, Eq. (11) defines the coupling coefficients between two modes of regions i and j and naturally reduces to the mode orthogonality integral for $i=j$,

$$\langle \mathbf{E}_i(\mathbf{k}_i), \mathbf{H}_j(\mathbf{k}'_j) \rangle = \int \int_{-\infty}^{\infty} \mathbf{E}_i(\mathbf{r};\mathbf{k}_i) \times \mathbf{H}_j^{\dagger}(\mathbf{r};\mathbf{k}'_j) \cdot \hat{x} dy dz$$

$$= \delta(\mathbf{k}_i - \mathbf{k}'_j) \quad \text{if } i = j. \tag{11}$$

Here, the dagger indicates the adjoint field, which is equivalent to reversing the propagation direction of the field.[28] The use of the adjoint field as opposed to complex conjugate ensures Hermiticity of the coupling coefficients, one of the requirements for a consistent scalar product. The requirement of linearity follows naturally from the definition. Note that both mode coupling and mode orthogonality are independent of each individual mode's harmonic variation in the $\hat{x}$ direction. Evaluating Eq. (11) is straightforward for harmonic variation in the $\hat{y}$ direction and by direct substitution of the field components from Eqs. (6) and (7), and integration over y leaves an integral over z,

$$\langle \mathbf{E}_i(\mathbf{k}_i), \mathbf{H}_j(\mathbf{k}_j) \rangle = \frac{N_i(\mathbf{k}_i) N_j(\mathbf{k}_j) z_0 k_{x,j} V_i}{k_0 V_j} \delta(k_{y,i} - k_{y,j}) I(\mathbf{k}_i, \mathbf{k}_j), \tag{12}$$

where

$$I(\mathbf{k}_i, \mathbf{k}_j) = \int_{-\infty}^{\infty} \frac{\psi_{i,-}(z) \psi_{j,-}(z)}{\varepsilon_i(z)} dz$$

$$= -\frac{[r_i(\mathbf{k}_i) + r_j(\mathbf{k}_j)]\pi}{\varepsilon_{d,i}} \delta(k_{zd,i} - k_{zd,j})$$

$$+ i \frac{[1 - r_i(\mathbf{k}_i)][1 - r_j(\mathbf{k}_j)]}{\varepsilon_{m,i}(k_{zd,i}^2 - k_{zd,j}^2)} \left[k_{zm,i} - \frac{\varepsilon_{d,j}\varepsilon_{m,i}}{\varepsilon_{d,i}\varepsilon_{m,j}} k_{zm,j} - \frac{(k_{zd,i}^2 - k_{zd,j}^2)}{(k_{zm,i} + k_{zm,j})} \right]. \tag{13}$$

It is noteworthy that the mathematical foundation of the following mode matching analysis lies in this single integral. It provides analytical expressions for the coupling between the surface waves of the various regions in a calculation. The full analytical evaluation of the integral leads to a robust and numerically stable implementation, which sets this approach aside from past reports on this problem.

In the case of radiation modes, both $k_{zd,i} \in [0, \infty]$ and $k_{zd,j} \in [0, \infty]$, so that the radiation to radiation mode coupling coefficients retain all parts of Eq. (13). In region i, radiation modes are orthogonal by inspection, such that Eq. (12) is evaluated to

$$\langle \mathbf{E}_i(\mathbf{k}_i), \mathbf{H}_i(\mathbf{k}'_i) \rangle = -\pi \frac{N_i(\mathbf{k}_i) N_i(\mathbf{k}'_i) [r_i(\mathbf{k}) + r_i(\mathbf{k}')] z_0 k'_{x,i}}{\varepsilon_{d_i} k_0}$$

$$\times \delta(k_{y,i} - k'_{y,i}) \delta(k_{z,i} - k'_{z,i}). \tag{14}$$

Therefore, the radiation modes are orthonormal with the normalization constant

$$N_{r,i}(\mathbf{k}_i) = \mathrm{i} \sqrt{\frac{\varepsilon_d k_0}{2\pi r_i z_0 k_{x,i}}}. \tag{15}$$

For the surface plasmon in region i, $r_i(\mathbf{k}_i; k_{xi} \mapsto k_{xg,i}) \mapsto \infty$. From here on, small letters distinguish the surface plasmon's field components from those of the radiation modes to condense: for example, $\{\mathbf{E}_i(\mathbf{k}_i; k_x = k_{xg,i}), \mathbf{H}_j(\mathbf{k}_i; k_x = k_{xg,i})\} = \{\mathbf{e}_i, \mathbf{h}_j\}$. The reader will note that the radiation modes and surface plasmon are orthogonal by inspection of Eq. (13). Evaluating the integral $\langle \mathbf{e}_i, \mathbf{h}_i \rangle$ leads to the following normalization;

$$N_{g,i} = -\sqrt{\frac{k_0}{z_0 k_{x,i}}}\left[\frac{2i\varepsilon_{d,i}\varepsilon_{m,i}^2 k_{zd,i}}{\varepsilon_{d,i}^2 - \varepsilon_{m,i}^2}\right]^{1/2}. \tag{16}$$

In surface plasmon field normalization, the factor $[1-r_i(\mathbf{k}_i)][1-r_j(\mathbf{k}_j)] \mapsto \infty$ appears in the coupling coefficient expression of Eq. (13) but is omitted from the definition of $N_{g,i}$. In fact, $N_{g,i}$ must implicitly eliminate the singularity as $r_i(\mathbf{k}_i) \mapsto \infty$. A consequence of removing the singularity requires an additional minus sign in the definition of $N_{g,i}$.

The set of normalized surface waves and the coupling coefficients can now be applied to the framework of the mode matching technique. Generally, this requires the nontrivial coefficients: radiation to radiation coupling $\langle \mathbf{E}_i(\mathbf{k}_i), \mathbf{H}_i(\mathbf{k}_j)\rangle$, guided to guided coupling $\langle \mathbf{e}_i, \mathbf{h}_j\rangle$, guided to radiation coupling $\langle \mathbf{e}_i, \mathbf{H}_j(\mathbf{k}_j)\rangle$, and radiation to guided coupling $\langle \mathbf{E}_i(\mathbf{k}_i), \mathbf{h}_j\rangle$. In addition, the reader should note the following orthogonality relations: $\langle \mathbf{E}_i(\mathbf{k}_i), \mathbf{H}_i(\mathbf{k}_i')\rangle = \delta(k_{y,i} - k_{y,i}')\delta(k_{z,i} - k_{z,i}')$, $\langle \mathbf{e}_i, \mathbf{h}_i\rangle = 1$, and $\langle \mathbf{e}_i, \mathbf{H}_i(\mathbf{k}_i)\rangle = \langle \mathbf{E}_i(\mathbf{k}_i), \mathbf{h}_i\rangle = 0$.

C. Mode matching equations

The mode matching equations arise from the matching of field components across the discontinuity at $x=0$, shown in Fig. 1(b), followed by use of the coupling coefficients integral of Eq. (12). Using the expressions from Sec. II B, the continuity of the transverse components of the E_z and H_y field components implies

$$(a-\bar{a})e_{z,i}(\mathbf{r}) + \int_0^\infty [A(\mathbf{k}) - \bar{A}(\mathbf{k})]E_{z,i}(\mathbf{r};\mathbf{k})dk_z = (b-\bar{b})e_{z,j}(\mathbf{r}) + \int_0^\infty [B(\mathbf{k}) - \bar{B}(\mathbf{k})]E_{z,j}(\mathbf{r};\mathbf{k})dk_z, \tag{17}$$

$$(a+\bar{a})h_{y,i}(\mathbf{r}) + \int_0^\infty [A(\mathbf{k}) + \bar{A}(\mathbf{k})]H_{y,i}(\mathbf{r};\mathbf{k})dk_z = (b+\bar{b})h_{y,j}(\mathbf{r}) + \int_0^\infty [B(\mathbf{k}) + \bar{B}(\mathbf{k})]H_{y,j}(\mathbf{r};\mathbf{k})dk_z. \tag{18}$$

This is a summation over the set of surface waves with integration over the continuous radiation modes. Choosing to integrate the radiation modes with respect to k_{zd} proves to be useful later on when evaluating the delta function term in the coupling coefficients. Post vector multiplication of Eq. (17) by $\times h_{y,i}(\mathbf{r})$ and $\times H_{y,i}(\mathbf{r};\mathbf{k}_i)$ followed by vector integration over the z-y plane provides two mode matching equations. Similarly, prevector multiplication of Eq. (18) by $e_{z,j}(\mathbf{r})\times$ and $E_{z,j}(\mathbf{r};\mathbf{k}_i)\times$ followed by vector integration over the z-y plane provides the other two,

$$a-\bar{a} = (b-\bar{b})\langle \mathbf{e}_j, \mathbf{h}_i\rangle + \int_0^\infty [B(\mathbf{k}_j) - \bar{B}(\mathbf{k}_j)]\langle \mathbf{E}_j(\mathbf{k}_j), \mathbf{h}_i\rangle dk_{zd,j},$$

$$A(\mathbf{k}_i) - \bar{A}(\mathbf{k}_i) = (b-\bar{b})\langle \mathbf{e}_j, \mathbf{H}_i(\mathbf{k}_i)\rangle + \int_0^\infty [B(\mathbf{k}_j) - \bar{B}(\mathbf{k}_j)] \times \langle \mathbf{E}_j(\mathbf{k}_j), \mathbf{H}_i(\mathbf{k}_i)\rangle dk_{zd,j},$$

$$b+\bar{b} = (a+\bar{a})\langle \mathbf{e}_j, \mathbf{h}_i\rangle + \int_0^\infty [A(\mathbf{k}_i) + \bar{A}(\mathbf{k}_i)]\langle \mathbf{e}_j, \mathbf{H}_i(\mathbf{k}_i)\rangle dk_{zd,i},$$

$$B(\mathbf{k}_j) + \bar{B}(\mathbf{k}_j) = (a+\bar{a})\langle \mathbf{E}_j(\mathbf{k}_j), \mathbf{h}_i\rangle + \int_0^\infty [A(\mathbf{k}_i) + \bar{A}(\mathbf{k}_i)] \times \langle \mathbf{E}_j(\mathbf{k}_j), \mathbf{H}_i(\mathbf{k}_i)\rangle dk_{zd,i}. \tag{19}$$

To allow a numerical solution to these coupled integral equations, consider truncated summations to approximate the integrals over radiation modes. However, in order to improve accuracy and convergence, the integrals are written as a Gaussian quadrature summation. In this case, the quadrature weighting factors require special treatment: the solution to the mode matching problem yields a set of radiation modes that, when summed, gives the total scattered power. So, for example, the backscattered radiation power is approximated by

$$\int |\bar{A}(\mathbf{k}_i)|^2 dk_z \mapsto \sum_m |\bar{A}_m(x)|^2 w_m. \tag{20}$$

It follows that $A(\mathbf{k}_i) \mapsto A_m/\sqrt{w_m}$ gives the correct transformation between the continuous and discrete representations, where w_m are the weighting coefficients of the Gaussian quadrature scheme used.[29] Note that, in principle, any set of abscissa and weight functions may be incorporated into this method. Approximating the integrals in Eq. (19) with Gaussian quadrature summations and substituting for the normalized radiation mode amplitudes, the following mode matching equations result:

$$a-\bar{a} = (b-\bar{b})\langle \mathbf{e}_j, \mathbf{h}_i\rangle + \sum_{m=1} (B_m - \bar{B}_m)\langle \mathbf{E}_j(\mathbf{k}_m), \mathbf{h}_i\rangle \sqrt{w_m},$$

$$A_n - \bar{A}_n = (b-\bar{b})\langle \mathbf{e}_j, \mathbf{H}_i(\mathbf{k}_n)\rangle \sqrt{w_m} + (B_n - \bar{B}_n)\langle \mathbf{E}_j(\mathbf{k}_n), \mathbf{H}_i(\mathbf{k}_n)\rangle + \sum_{m=1} (B_m - \bar{B}_m)\langle \mathbf{E}_j(\mathbf{k}_m), \mathbf{H}_i(\mathbf{k}_n)\rangle w_m,$$

$$b+\bar{b} = (a+\bar{a})\langle \mathbf{e}_j, \mathbf{h}_i\rangle + \sum_{m=1} (A_m + \bar{A}_m)\langle \mathbf{e}_j, \mathbf{H}_i(\mathbf{k}_m)\rangle \sqrt{w_m},$$

$$B_n + \bar{B}_n = (a+\bar{a})\langle \mathbf{E}_j(\mathbf{k}_n), \mathbf{h}_i\rangle \sqrt{w_m} + (A_n + \bar{A}_n)\langle \mathbf{E}_j(\mathbf{k}_n), \mathbf{H}_i(\mathbf{k}_n)\rangle + \sum_{m=1} (A_m + \bar{A}_m)\langle \mathbf{E}_j(\mathbf{k}_n), \mathbf{H}_i(\mathbf{k}_m)\rangle w_m. \tag{21}$$

Arranging these expressions yields two coupled matrix equations,

$$A_n + \bar{A}_n = \sum_{m=0} (B_m + \bar{B}_m)C_{mn},$$

$$B_n - \bar{B}_n = \sum_{m=0} (A_m - \bar{A}_m) C_{mn}^T, \tag{22}$$

where

$$C_{00} = \langle \mathbf{e}_j, \mathbf{h}_i \rangle, \quad m = n = 0,$$

$$C_{0n} = \langle \mathbf{e}_j, \mathbf{H}_i(\mathbf{k}_n) \rangle \sqrt{w_m},$$

$$C_{m0} = \langle \mathbf{E}_j(\mathbf{k}_m), \mathbf{h}_i \rangle \sqrt{w_m},$$

$$C_{mn} = \langle \mathbf{E}_j(\mathbf{k}_m), \mathbf{H}_i(\mathbf{k}_n) \rangle, \quad m = n,$$

$$C_{mn} = \langle \mathbf{E}_j(\mathbf{k}_m), \mathbf{H}_i(\mathbf{k}_n) \rangle w_m, \quad m \neq n. \tag{23}$$

Here, the continuous region labels i and j are replaced with discrete labels m and n, respectively, with $m,n=0$ labeling the surface plasmon. Accordingly, the matrix positions $m,n=0$ represent surface plasmon coupling coefficients and $m,n \geqslant 1$ represent those of the radiation modes. Note that the elements C_{mn} are not trivial to evaluate analytically at the singularity, $k_{zd,i}=k_{zd,j}$. However, since this matrix equation is actually a representation of an integral equation, in cases where $k_{zd,i}=k_{zd,j}$, the principal part of the singular term in $(k_{zd,i}-k_{zd,j})^{-1}$ should be considered in the calculation. The contribution from this term is negligible near $k_{zd,i}=k_{zd,j}$ and only the non-singular components of Eq. (12) contribute to C_{mn}. The weight functions in Eq. (23) indicate the strength of coupling between the guided and radiation modes. Guided to radiation mode coupling is clearly important with a scaling of $\sqrt{w_m}$, whereas scattering between unmatched radiation modes is not as important with a scaling w_m. Again, examining the parts of the C_{mn} elements from Eq. (12) shows that the contribution from the delta function is the most significant term since the other parts scale with w_m. Manipulating the matrix equations further gives the interface scattering matrix relating the modal outputs to the inputs such that

$$\begin{pmatrix} \mathbf{B} \\ \bar{\mathbf{A}} \end{pmatrix} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix} \begin{pmatrix} \bar{\mathbf{B}} \\ \mathbf{A} \end{pmatrix}, \tag{24}$$

where

$$\mathbf{S}_{11} = [1 + \mathbf{C}\mathbf{C}^T]^{-1}(1 - \mathbf{C}\mathbf{C}^T),$$

$$\mathbf{S}_{12} = [1 + \mathbf{C}\mathbf{C}^T]^{-1}\mathbf{C},$$

$$\mathbf{S}_{21} = [1 + \mathbf{C}^T\mathbf{C}]^{-1}\mathbf{C}^T,$$

$$\mathbf{S}_{22} = [1 + \mathbf{C}\mathbf{C}^T]^{-1}(1 - \mathbf{C}\mathbf{C}^T). \tag{25}$$

D. Multiple interface calculations

The scattering matrix method describes general multiple surface discontinuities by defining a propagation scattering matrix for the discretized system of modes from Sec. II C. The scattering matrix relating amplitudes within the same region propagating between the positions x_1 and x_2 has the form

$$\begin{pmatrix} \mathbf{A}(x_2) \\ \bar{\mathbf{A}}(x_1) \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{P} \\ \mathbf{P} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \bar{\mathbf{A}}(x_2) \\ \mathbf{A}(x_1) \end{pmatrix}, \tag{26}$$

where

$$\mathbf{P}_{mn} = \exp\{ik_{xg}(x_2 - x_1)\} \quad m = n = 0$$
$$= \exp\{ik_{xr,m}(x_2 - x_1)\}\delta_{mn}. \tag{27}$$

Combining both interface [Eq. (25)] and propagation [Eq. (26)] scattering matrices provides the description of any system of coplanar discontinuities of the form under investigation. Note that combining scattering matrices requires the usual concatenation method.

III. NUMERICAL ANALYSIS

This section assesses the accuracy and convergence of the model with a series of numerical tests. Firstly, checking the matching of field functions on either side of an open waveguide discontinuity ensures both correct operation of the model as well as self-consistency. Secondly, the method is compared with the commercial finite-element modeling software from Comsol (FEMLAB).

Firstly, consider a comparison of the matched fields at the abrupt discontinuity of two open waveguides with $\varepsilon_{m,1}=\varepsilon_{m,2}=-18.3-0.5\mathrm{i}$ [$\varepsilon_{\mathrm{Ag}}$ at $\lambda=632.8$ nm (Ref. 31)], $\varepsilon_{d,1}=2.25$, and $\varepsilon_{d,2}=1$. To generate an accurate matching of field functions, the model uses 200 radiation modes in addition to the single surface plasmon mode. Studies show that solution convergence requires a large number of evanescent radiation modes; here, the radiation mode truncation is set at $k_{zd,max}=10\sqrt{\varepsilon_{d,max}}k_0$, where $\varepsilon_{d,max}$ is the highest dielectric permittivity in a structure. This prescription works well and only small improvements in convergence occur with $k_{z,max}=100\sqrt{\varepsilon_{d,max}}k_0$. Figures 2(a) and 2(b) show the real and imaginary parts of the H_y and E_z fields, respectively, on either side of the discontinuity; clearly, the correspondence is excellent.

A more detailed examination of the fields very near the metal-dielectric interface highlights the largest deviations from ideal field matching. Using more radiation modes, a larger proportion of which are evanescent eliminates these errors, indicating asymptotic convergence to the true solution. The mean standard deviation between the fields on either side of the discontinuity for the case considered above is slightly larger than 1%. With 500 radiation modes truncated at $k_{z,max}=100\sqrt{\varepsilon_{d,max}}k_0$, the deviation is 0.5%. A high level of convergence clearly requires significant computational expenditure as noted originally for this problem by Stegeman *et al.*[24] However, the reader should note that field matching requires a much higher number of modes of satisfactory convergence compared with the surface plasmon reflection and transmission amplitudes: satisfactory convergence of modal amplitudes occurs for about 50 modes in most of the calculations in this paper.

Upon comparison with Comsol's finite-element (FE) software for this problem, the correspondence is also excellent. Figures 3(a)–3(c) compare the absolute field components $|H_y|$, $|E_z|$, and $|D_x|$ [where $D_x(\mathbf{r})=\varepsilon(z,x)E_x(\mathbf{r})$ is the electric

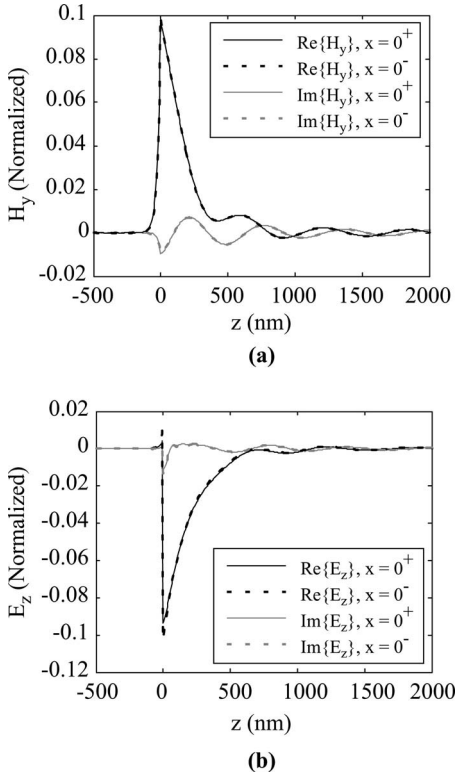

FIG. 2. Comparison of the matched (a) H_y and (b) E_z fields on either side $(x=0^+,0^-)$ of an abrupt open waveguide discontinuity. Here, $\varepsilon_{m,1}=\varepsilon_{m,2}=-18.3-0.5i$ ($\varepsilon_{\mathrm{Ag}}$ at $\lambda=632.8$ nm), $\varepsilon_{d,1}=2.25$, and $\varepsilon_{d,2}=1$.

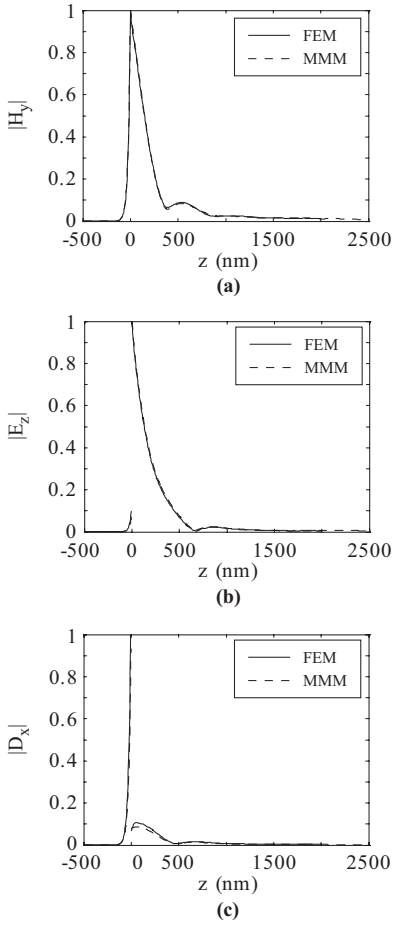

FIG. 3. Comparison of the matched (a) $|H_y|$, (b) $|E_z|$, and (c) $|D_x|$ fields at an abrupt open waveguide discontinuity with FEMLAB simulations. Here, $\varepsilon_{m,1}=\varepsilon_{m,2}=-18.3-0.5i$ ($\varepsilon_{\mathrm{Ag}}$ at $\lambda=632.8$ nm), $\varepsilon_{d,1}=2.25$, and $\varepsilon_{d,2}=1$.

displacement], respectively, at the same open waveguide discontinuity as in Fig. 2 for the FE (solid lines) and semianalytical (broken lines) methods. The discontinuities of the fields at dielectric boundaries in both Figs. 2 and 3 are consistent with the boundary conditions implied by Maxwell's equations.

The FE software uses scattering boundary conditions to minimize interference between the solution and boundary reflections within the calculation domain: this was a rectangular region of 10 μm (x direction) $\times 2$ μm (y direction) divided into quarters with permittivities $\varepsilon_{d,1}$, $\varepsilon_{m,1}$, $\varepsilon_{m,2}$, and $\varepsilon_{d,2}$ in an counterclockwise fashion starting from the top left quarter. Convergence of the FE software for this problem required a pregenerated adaptive mesh with nearly 500 000 elements and a calculation time of approximately 2 min on a 2.4 GHz Pentium IV. The semianalytical result took only 5 s using MATLAB on the same machine.

The accuracy of either method is clearly not in question from the results in Figs. 2 and 3. Although the time saving in using the semianalytical approach is a clear advantage, the principal benefit of this approach is that it provides the modal scattering amplitudes directly. In fact, in generating the field plots in Figs. 2 and 3, the analytical approach must perform additional computations that constitute almost 50% of the overall calculation time. In stark contrast, deconvolving the modal scattering amplitudes from the fields of the FE approach is not a straightforward task.

The numerical analysis has shown that mathematically, the set of surface waves interacts in such a way as to satisfy the boundary conditions dictated by Maxwell's equations. The complexity of the problem arises because there are two independent sets of surface wave eigenmodes where no two eigenmodes from different media have the same field distribution. This is distinct from the case of an infinite planar interface dividing two dielectric regions: this problem is straightforward because there are always two plane waves on either side of the interface that have the same field distribution (below the critical angle). So, in the case of an incident surface plasmon, radiation surface waves are generated at the discontinuity, in addition to the reflected and transmitted surface plasmon waves, to ensure both phase and field matchings. This is evident from the oscillating mode matched field components at the surface discontinuity plotted in Fig. 2.

Physically, the corner dividing all the dielectric regions generates surface waves with $k_{zd,i}\in[0,\infty]$. This includes radiation waves that are both propagating, $k_{zd,i}<\sqrt{\varepsilon_i}k_0$, and evanescent, $k_{zd,i}>\sqrt{\varepsilon_i}k_0$, along the $\hat{x}$ direction. In the case under examination, a surface plasmon mode impinges on the corner, generating both reflected and transmitted surface plasmon waves in addition to some distribution of radial scattering that radiates from the corner. From the numerical analysis in this section, radiation modes are generated with up to $k_{zd,max}=10\sqrt{\varepsilon_{d,max}}k_0$, which indicates the importance

of the evanescent spectrum of radiation modes in these calculations.

IV. RESULTS

This section is split into four parts. The first two parts show the results of transmission, reflection, and scattering calculations for a surface plasmon wave incident on an interface between two dielectric and/or metal half spaces; the first part considers dielectric discontinuities and the second part considers metallic discontinuities. The final part considers the compilation of multiple surface discontinuities into a surface plasmon distributed Bragg reflector (DBR). In all cases, the results show the scattering effect of a single incident surface plasmon wave. Figure 1(b) shows the general case of surface waves that enter and leave a surface discontinuity. In the following calculations, a surface wave impinges from the left such that $a=1$ and $\bar{b}=0$, where $\bar{a}$ is the surface plasmon reflectivity ($R=|\bar{a}|^2$) and b is the surface plasmon transmission ($T=|b|^2$). Radiation modes propagating toward a discontinuity are set to zero, $A_m=\bar{B}_m=0$, such that $\bar{A}_m$ and B_m quantify backward and forward scatterings, respectively. The total scattered power S is a summation over the relevant forward and backward radiation modes, $\Sigma_m|\bar{A}_m|^2$ for $k_{zd,m}\leqslant\sqrt{\varepsilon_{d,i}}k_0$ and $\Sigma_n|B_n|^2$ for $k_{zd,n}\leqslant\sqrt{\varepsilon_{d,j}}k_0$, respectively. Note that the evanescent portion of the radiation modes generated at the discontinuity does not contribute to the radiated power.

In the following calculations, the same model parameters as in Sec. III are used. When not scanning spectrally, results of the following study are for the helium-neon red laser line at 632.8 nm.

A. Dielectric discontinuity

Figure 4 shows the reflection R, transmission T, and scattering S to radiation modes for a surface plasmon wave normally incident on a dielectric discontinuity, yet guided along the same continuous silver surface. The interface calculations satisfy energy conservation by virtue of the fact that $S\approx 1-R-T$ to an acceptable tolerance; note that for the high to low permittivity case of Fig. 4(a), the calculations would converge better with more radiation modes and a higher evanescent wave vector truncation. The results are in general qualitative agreement with those of Stegeman *et al.*,[25] who analyzed this problem for an approximate closed waveguide geometry.

A particularly interesting feature of these calculations is the close correspondence of the surface plasmon reflectivity with the Fresnel value that accounts for the effective phase index of the surface wave, n_i. The dots indicate the Fresnel reflectivity R_F given by

$$R_F=\left|\frac{(n_i-n_j)}{(n_i+n_j)}\right|^2, \tag{28}$$

where $n_i=\sqrt{\varepsilon_{d,i}\varepsilon_{m,i}/(\varepsilon_{d,i}+\varepsilon_{m,i})}$. The correspondence is exact for low to high surface plasmon wave vectors, with a small deviation for the high to low permittivity case. Related reports have identified similar nonreciprocal behavior.[25,27]

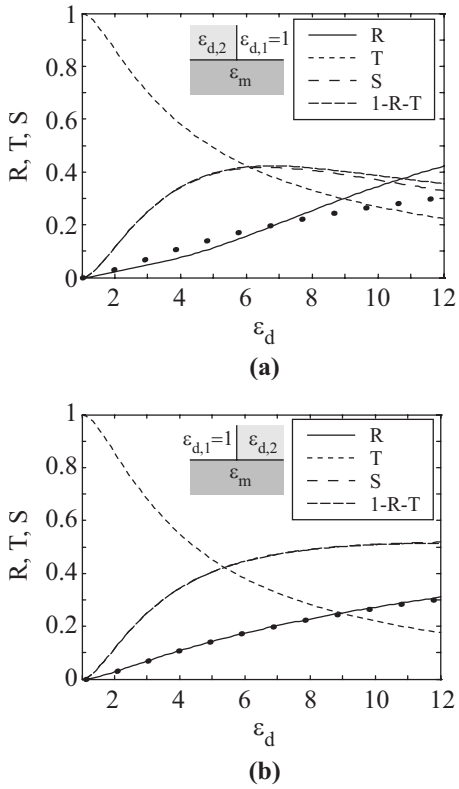

FIG. 4. Reflection, transmission, and scattering of surface waves (propagating left to right) as a function of dielectric permittivity contrasts $\varepsilon_{d,1}$ and $\varepsilon_{d,2}$ for (a) high to low and (b) low to high at a wavelength of 632.8 nm. Dots show the equivalent Fresnel reflection coefficient calculated from Eq. (28).

Stegeman *et al.*[25] also identify the correspondence with the Fresnel coefficients; however, due to the approximate method, they do not identify the close correspondence in the case of low to high permittivity. Although their results are not directly comparable with these, the surface plasmon reflectivity appears to be slightly underestimated.

Analysis of the radiation mode distributions, shown in Fig. 5 for $\varepsilon_{d,i}=2.25$ and $\varepsilon_{d,j}=1$ interfaces with silver, reveals further nonreciprocal behavior. In both cases, most of the scattered power is in the forward direction; however, only in the case of high to low permittivity is there any significant backscattering. This provides an indication of the cause of deviations from the Fresnel coefficient. These results are in basic qualitative agreement with those of Voronko *et al.*;[26] however, quantitatively, there are significant discrepancies indicating a significantly reduced surface plasmon reflectivity for the high to low permittivity case. It is noteworthy that, as a rule of thumb, the reflection is equivalent to the Fresnel value for bulk waves and the transmission and scattering share from the remaining electromagnetic power with proportions that depend on the size of the discontinuity.

B. Metallic discontinuity

Consider now the scattering of surface plasmon waves normally incident on a metallic surface discontinuity, where the dielectric half space is continuous. This study briefly out-

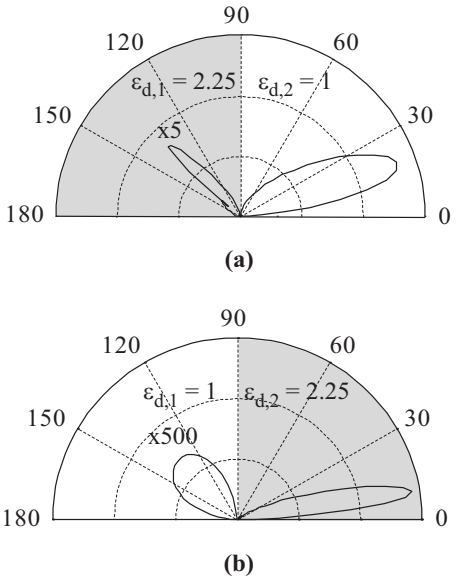

FIG. 5. Scattering of propagating radiation waves at the dielectric-metal surface discontinuity for (a) $\varepsilon_{d,1}=2.25$ to $\varepsilon_{d,2}=1$ on silver and (b) $\varepsilon_{d,1}=1$ to $\varepsilon_{d,2}=2.25$ on silver at a wavelength of 632.8 nm.

lines the effects of a perfect abrupt transition between two metallic half spaces. Figure 6 shows results of calculating the transition between regions of silver/air and aluminum/air. The calculations consider real permittivity data for both aluminum[30] and silver.[31]

Surface plasmon reflection is minimal (<1%) and is only significant near the surface plasma edge of silver, where the shapes and effective indices of the modes on either side of the discontinuity are most different. The small differences in the shape of surface plasmon modes in each region cause a significant amount of scattering, settling to about 10% far from the surface plasmon edge. As with the case of the dielectric discontinuity, forward scattering is dominant with significant nonreciprocal backward scattering for the silver to aluminum and aluminum to silver cases. Again, these results are in qualitative agreement with those of Voronko *et al.*[26] Following a similar approach to the dielectric case in Sec. IV A, the inset of Fig. 6(a) compares the reflection coefficients with the effective Fresnel coefficient calculated using Eq. (28): the effective index for a surface plasmon on silver in the current case is $n_{\mathrm{Ag}}(\lambda)=\sqrt{\varepsilon_{\mathrm{Ag}}(\lambda)/[\varepsilon_{\mathrm{Ag}}(\lambda)+1]}$ (since $\varepsilon_d=1$) and a similar expression holds for $n_{\mathrm{Al}}(\lambda)$. Again, the effective reflection coefficient describes the case of silver to aluminum well. However, the reflectivity of this system is even more asymmetric than in the dielectric case: for aluminum to silver, the results do not closely match at all. Again, the strength of backward scattering in this case is five times higher and indicates that the discrepancy arises due to significant modal mismatch of the surface plasmon fields on either side of the discontinuity.

The various combinations of gold, silver, and aluminum produce similar results, with moderately significant reflection only manifested near to a surface plasmon edge. It seems, therefore, that this system is unappealing for creating a resonant feedback system due to the intrinsic damping of surface

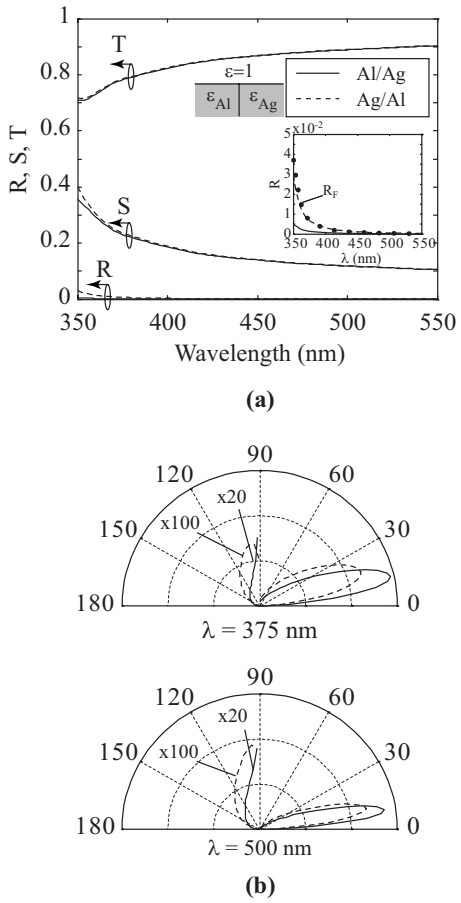

FIG. 6. Surface plasmon scattering at a metallic discontinuity. (a) Reflection, transmission, and scattering coefficients. (b) identifies the direction of radiated scattering. Solid lines represent transition from aluminum to silver and the broken lines represent the reverse. The inset of (a) shows the reflection coefficients compared with the effective Fresnel reflection calculated from Eq. (28).

waves and the relatively week reflectivity of the discontinuity. As the wavelength increases in these systems, the surface plasmon dispersion approaches the light line and therefore takes on more of the characteristics of a bulk wave. Since there is no dielectric discontinuity, scattering and reflection will tend to zero in the low frequency limit.

C. Surface plasmon distributed Bragg reflector

The final example of the current mode matching approach examines the potential to design effective surface wave resonators. One of the key components of nano- and microscale cavities are highly reflective mirrors. Distributed Bragg reflectors (DBRs) are widely used in microcavity physics to create effective mirrors for laser devices. The possibility of using such structures on the nanoscale is extremely appealing in order to enhance the low Q factors of nanoparticulate systems. Figure 7 plots the results of a mode matching analysis of DBRs composed of alternating dielectric layers of polymethyl methacrylate (PMMA) ($\varepsilon_d=2.25$) and air next to a silver metallic interface. The optical thickness of each layer is a quarter of the surface plasmon wavelength at λ_{DBR}

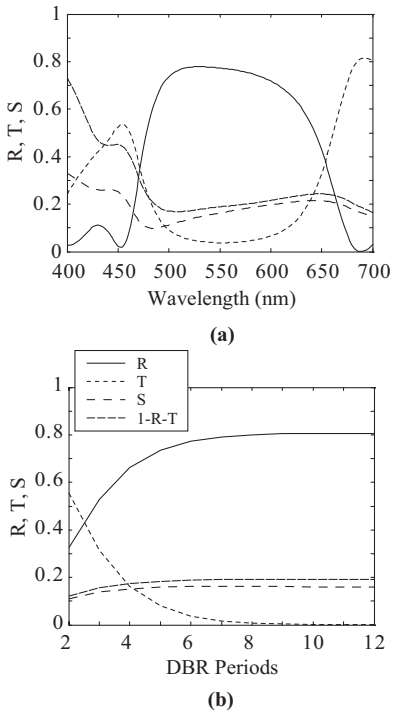

FIG. 7. Surface plasmon reflection, transmission and radiation for a PMMA/air DBRs. (a) Spectral response of modal scattering parameters for a six period DBR. (b) Peak DBR reflectivity as a function of the number of periods.

=550 nm. Under these conditions, the DBR is mainly effective for surface plasmons and a few radiation modes near the light line. Launching a surface plasmon from one side of the structure allows calculation of the effectiveness of a DBR to reflect and transmit surface plasmons as well as scatter radiation waves. The first important feature of the results in Fig. 7(a) is that only moderate surface plasmon reflection, $R \approx 80\%$, is achievable in this type of structure. For a single wavelength cavity formed between two such reflectors, the Q factor could approach $Q=2\lambda_{SP}\sqrt{R}/[\lambda(1-R)] \sim 10$. Figure 7(b) highlights that the limitations are solely from scattering to radiation modes. As the number of DBR periods increases, the scattering only moderately increases, beyond the six period result. The remaining power is shared between surface plasmon transmission and reflection through the DBR. Eventually, the reflectivity saturates and the transmission tends to zero. Although omitted from this paper, optimizing this system to maximize reflectivity with permittivity contrasts brings moderate improvements such that $R=85\%$; this requires balancing intrinsic propagation losses for a given number of periods with the scattering loss of each interface. However, one can see from Fig. 7(a) that the scattered power is larger than that due to intrinsic losses (by comparing $1-R-T$ with the scattered power): in this case, propagation losses comprise about 10% of the scattering losses. This is encouraging since there are other surface plasmon systems that suppress radiation scattering effectively such as for coupled plasmons between the dielectric gap of two coplanar metals.[32,33]

V. CONCLUSIONS

We report an efficient, robust, and semianalytical mode matching approach for modeling the problem of electromagnetic surface wave scattering at single and multiple surface discontinuities. Excellent agreement between this method and a numerical finite-element method was demonstrated for the two-dimensional (2D) problem. However, the semianalytical method solved the field profiles significantly faster (>25 times for typical 2D problems) and provided a wealth of useful information such as modal coupling and scattering coefficients that clearly highlight its applicability for the efficient and intuitive design of plasmonic devices based on structural nonuniformities including nanoplasmonic cavity structures.

Reflectivity, transmission, and scattering of surface plasmons impinging at normal incidence onto single and multiple dielectric and metallic surface discontinuities were analyzed. In the cases studied, the reflectivities for low to high surface plasmon wave vectors followed very closely the Fresnel reflection coefficient for bulk waves, while the transmission and radiation scattering share the remaining power. In the case of dielectric discontinuities, small deviations from the expected Fresnel values occurred for high to low surface plasmon wave vectors due to increased scattering; this effect was much larger in the case of metallic discontinuities. The modal analysis provided an explanation for this behavior: in the case of low to high wave vector surface plasmons, very little backscattered radiation indicated good field matching in the incident medium. On the hand, for high to low wave vector surface plasmons, significantly larger backscattered radiation indicated poor mode matching in the incident medium and therefore a deviation from the effective Fresnel picture. Further analysis of the radiated fields shows that the majority of the scattered power is in the forward direction.

The open guiding geometry (i.e., with dielectric half spaces) readily supports radiating waves and only allows moderate DBR reflection coefficients of 80%–90%. The physical reason lies in the competition between intrinsic and scattering losses: on the one hand, the reflectivity (DBR penetration) of the guided mode increases (decreases) with increasing "strength" (ratio of the guided mode's effective indices) of the nonuniformities, thereby reducing propagation losses; on the other hand, increasing the strength of the nonuniformity results in increasing the scattering into radiation modes. The competition between these two mechanisms results in an optimal (maximum) but nonunity reflection coefficient. Since the main source of loss is through modal scattering, alternative guiding geometries that suppress radiation modes, such as gap and channel structures, would be superior for the design of strongly reflective DBRs.

ACKNOWLEDGMENTS

This work was supported by the U.S. Air Force Office of Scientific Research MURI program under Grant No. FA9550-04-1-0434 and by the NSF Nanoscale Science and Engineering Center (NSEC) under Grant No. DMI-0327077.

APPENDIX: DERIVATION OF $I(\mathbf{k}_i,\mathbf{k}_j)$

The integral from Eq. (13) is first written in terms of the field functions for $z<0$ in the metal region and for $z\geqslant 0$ in the dielectric region as follows:

$$\begin{aligned} I(\mathbf{k}_i,\mathbf{k}_j) &= \int_{-\infty}^{\infty}[1-H(z)]\frac{\psi_{i,-}(z)\psi_{j,-}(z)}{\varepsilon_{m,i}}dz \\ &+ \int_{-\infty}^{\infty}H(z)\frac{\psi_{i,-}(z)\psi_{j,-}(z)}{\varepsilon_{d,i}}dz \\ &= (1-r_i)(1-r_j)\int_{-\infty}^{\infty}[1-H(z)] \\ &\times\frac{\exp(-\mathrm{i}(k_{zm,i}+k_{zm,j})z)}{\varepsilon_{m,i}}dz \\ &+ \int_{-\infty}^{\infty}H(z)\frac{\exp(-\mathrm{i}(k_{zd,i}+k_{zd,j})z)}{\varepsilon_{d,i}}dz \\ &- r_i\int_{-\infty}^{\infty}H(z)\frac{\exp(\mathrm{i}(k_{zd,i}-k_{zd,j})z)}{\varepsilon_{d,i}}dz \\ &- r_j\int_{-\infty}^{\infty}H(z)\frac{\exp(-\mathrm{i}(k_{zd,i}-k_{zd,j})z)}{\varepsilon_{d,i}}dz \\ &+ r_ir_j\int_{-\infty}^{\infty}H(z)\frac{\exp(\mathrm{i}(k_{zd,i}+k_{zd,j})z)}{\varepsilon_{d,i}}dz. \end{aligned} \tag{A1}$$

Note that r_i is short for $r(\mathbf{k}_i)$. Each of the integrals in this expression involves the Fourier transform of the Heaviside step function $H(z)$, where $H(z<0)=0$, $H(z=0)=0.5$, $H(z>0)=1$, and

$$\int_{-\infty}^{\infty}H(z)\exp(\pm\mathrm{i}kz)dz = \pi\delta(k) \mp \frac{\mathrm{i}}{k}. \tag{A2}$$

Substitution into Eq. (A1) gives the general expression for the coupling coefficients integral,

$$\begin{aligned} I(\mathbf{k}_i,\mathbf{k}_j) &= \pi\frac{(1-r_i)(1-r_j)}{\varepsilon_{m,i}}\delta(k_{zm,i}+k_{zm,j}) - \mathrm{i}\frac{(1-r_i)(1-r_j)}{\varepsilon_{m,i}(k_{zm,i}+k_{zm,j})} \\ &+ \pi\frac{(1+r_ir_j)}{\varepsilon_{d,i}}\delta(k_{zd,i}+k_{zd,j}) + \mathrm{i}\frac{(1-r_ir_j)}{\varepsilon_{d,i}(k_{zd,i}+k_{zd,j})} \\ &- \pi\frac{(r_i+r_j)}{\varepsilon_{d,i}}\delta(k_{zd,i}-k_{zd,j}) + \mathrm{i}\frac{(r_i-r_j)}{\varepsilon_{d,i}(k_{zd,i}-k_{zd,j})}. \end{aligned} \tag{A3}$$

Here, note that only one delta function is retained, since the signs of the $\mathbf{k}$ vectors have already been accounted. Therefore, Eq. (A3) simplifies to

$$\begin{aligned} I(\mathbf{k}_i,\mathbf{k}_j) &= -\pi\frac{(r_i+r_j)}{\varepsilon_{d,i}}\delta(k_{zd,i}-k_{zd,j}) + \mathrm{i}\frac{(1-r_ir_j)}{\varepsilon_{d,i}(k_{zd,i}+k_{zd,j})} \\ &+ \mathrm{i}\frac{(r_i-r_j)}{\varepsilon_{d,i}(k_{zd,i}-k_{zd,j})} - \mathrm{i}\frac{(1-r_i)(1-r_j)}{\varepsilon_{m,i}(k_{zm,i}+k_{zm,j})}. \end{aligned} \tag{A4}$$

This expression reduces to the form shown in Eq. (13) by factorizing $(1-r_i)(1-r_j)$ and noting that

$$\frac{1+r_i}{1-r_i} = \frac{\varepsilon_{m,i}k_{zd,i}}{\varepsilon_{d,i}k_{zm,i}}. \tag{A5}$$

*xiang@berkeley.edu

[1] A. D. Boardman, *Electromagnetic Surface Modes* (Wiley, New York, 1982).
[2] H. Raether, *Surface Plasmons on Smooth and Rough Surface and on Gratings* (Springer-Verlag, Berlin, 1986).
[3] S. A. Maier, P. G. Kik, H. A. Atwater, S. Meltzer, E. Harel, B. E. Koeland, and A. A. G. Requicha, Nat. Mater. **2**, 229 (2003).
[4] P. Berini, Opt. Lett. **24**, 1011 (1999).
[5] S. I. Bozhevolnyi, V. S. Volkov, E. Devaux, J.-Y. Laluet, and T. W. Ebbesen, Nature (London) **440**, 508 (2006).
[6] K. Kneipp, Y. Wang, H. Kneipp, L. T. Perelman, I. Itzkan, R. R. Dasari, and M. S. Feld, Phys. Rev. Lett. **78**, 1667 (1997).
[7] K.-H. Su, S. Durant, J. M. Steele, Y. Xiong, C. Sun, and X. Zhang, J. Phys. Chem. B **110**, 3964 (2006).
[8] N. Fang, H. Lee, C. Sun, and X. Zhang, Science **308**, 534 (2005).
[9] H. T. Miyazaki and Y. Kurokawa, Phys. Rev. Lett. **96**, 097401 (2006).
[10] D. J. Bergman and M. I. Stockman, Phys. Rev. Lett. **90**, 027402 (2003).
[11] J. R. Krenn, H. Ditlbacher, G. Schider, A. Hohenau, A. Leitner, and F. R. Aussenegg, J. Microsc. **209**, 167 (2003).
[12] A. V. Zayats, I. I. Smolaninov, and A. A. Maradudin, Phys. Rep. **408**, 131 (2005).
[13] B. Wang and G. P. Wang, Appl. Phys. Lett. **87**, 013107 (2005).
[14] L. Zhou, X.-Q. Yu, and Y.-Y. Zhu, Appl. Phys. Lett. **89**, 051901 (2006).
[15] D. F. P. Pile and G. K. Gramotnev, Appl. Phys. Lett. **86**, 161101 (2005).
[16] F. Pincemin, A. A. Maradudin, A. D. Boardman, and J.-J. Greffet, Phys. Rev. B **50**, 15261 (1994).
[17] W. Braun, V. M. Kaganer, B. Jenichen, and K. H. Ploog, Phys. Rev. B **69**, 165405 (2004).
[18] F. López-Tejeira, F. J. García-Vidal, and L. Martín-Moreno, Phys. Rev. B **72**, 161405(R) (2005).
[19] B. T. Draine, Astrophys. J. **333**, 848 (1988).
[20] A. B. Evlukhin, S. I. Bozhevolnyi, A. L. Stepanov, and J. R. Krenn, Appl. Phys. B: Lasers Opt. **84**, 29 (2006).
[21] P. J. B. Clarricoats and K. R. Slinn, Proc. IEEE **114**, 878 (1967).
[22] V. V. Schevchenko, *Continuous Transitions in Open Waveguides* (Golem, Boulder, CO, 1971).
[23] S. F. Mahmoud and J. C. Beal, IEEE Trans. Microwave Theory Tech. **MTT-23**, 193 (1975).
[24] G. I. Stegeman, A. A. Maradudin, and T. S. Rahman, Phys. Rev. B **23**, 2576 (1981).
[25] G. I. Stegeman, N. Glass, A. A. Maradudin, T. Shen, and R. Wallis, Opt. Lett. **8**, 626 (1983).
[26] A. I. Voronko, L. G. Klimova, and G. N. Shkerdin, Solid State Commun. **61**, 361 (1987).

[27] H. A. Jamid and S. J. Al-Bader, IEEE Photonics Technol. Lett. **9**, 220 (1997).
[28] R. E. Collin, *Field Theory of Guided Waves*, 2nd ed. (Wiley, New York, 1991).
[29] P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration* (Academic, New York, 1975), p. 365.
[30] E. D. Palik, *Handbook of Optical Constants of Solids* (Academic, New York, 1985).
[31] P. B. Johnson and R. W. Christie, Phys. Rev. B **6**, 4370 (1972).
[32] E. N. Economou, Phys. Rev. **182**, 539 (1969).
[33] J. A. Dionne, L. A. Sweatlock, H. A. Atwater, and A. Polman, Phys. Rev. B **73**, 035407 (2006).