

This is the accepted manuscript made available via CHORUS. The article has been published as:

Channeling light into quantum-scale gaps

Rohan D. Kekatpure and Paul S. Davids

Phys. Rev. B **83**, 075408 — Published 7 February 2011

DOI: [10.1103/PhysRevB.83.075408](https://doi.org/10.1103/PhysRevB.83.075408)

Channeling light into quantum-scale gaps

Rohan D. Kekatpure* and Paul S. Davids†

*Applied Photonics and Microsystems,
Sandia National Laboratories,
Albuquerque, New Mexico, USA*

(Dated: December 16, 2010)

We develop a discrete plasmonic mode-matching technique to investigate the ultimate limits to plasmonic light concentration down to length-scales required for observation of quantum-mechanical phenomena, including plasmon-assisted electron tunneling. Our mode-matching calculations, verified by direct numerical solution of Maxwell's equations, indicate achievable coupling efficiencies of $> 20\%$ into symmetric bound gap plasmon modes in sub-10 nm gaps. For a given operating wavelength and a choice of material parameters, we demonstrate the existence of a specific width that maximizes enhancement of the electromagnetic field coupled into the gap. More generally, our calculations establish an intuitive and a computationally efficient framework for determining coupling efficiencies in and out of quantum-scale waveguides.

PACS numbers: 42.60.Da, 42.79.Gn, 42.82.Gw, 37.30.+i

REVIEW COPY
NOT FOR DISTRIBUTION

* rdkekat@sandia.gov

† pdavids@sandia.gov

I. INTRODUCTION

The ability to concentrate light beyond the diffraction limit sets the optical properties of metal nanostructures apart from their dielectric counterparts.¹⁻³ Even when it comes at the cost of increased optical attenuation, the light-concentration ability of metals enables study and application of localized light-matter interactions at length scales unattainable with dielectric materials.^{1,3}

A well-known example of the light-concentrating ability of metals is the phenomenon of extraordinary optical transmission (EOT) through subwavelength metallic holes with sizes on the order of 100 nm.⁴⁻⁶ Squeezing light through these tiny gaps has inspired important applications including advanced optical lithography,⁷ EOT-assisted bio-sensing,⁸ sub-ps switching in metamaterials,⁹ and plasmonic enhanced detectors.¹⁰ These efforts are significant demonstrations of new functionalities enabled by extreme light concentration. Yet, the electronic response of the materials themselves used in these devices is still bulk-like. As a result, it is still mostly the materials that affect the light.

One way to enable light to influence material properties is to make the light-matter interaction occur at the quantum-scale. Quantum-scale here refers to the spatial extent of the electronic wavefunction in metals and bound excitons in semiconductors. For most solids, this length scale is between 1 and 10 nm.¹¹

Light-matter interaction at this length-scale can lead to several interesting quantum-mechanical effects. For example, light-induced absorptive transition in most materials are understood under the electric-dipole approximation¹² which supposes that the spatial extent of the electromagnetic wave causing the transition is much larger than that of the atom/exciton. Shrinking the spatial extent of the driving field to the dimensions of the exciton can be expected to significantly modify the absorption rates and the selection rules governing them. Another example is that of plasmon-assisted electron tunneling across an insulating gap. The requirements for this tunneling process are (a) a gap-size on the order of electron wavefunction¹¹ and (b) existence of high electric field across the gap. The scale mismatch between the wavelength of light and the electron wavefunction once again makes it difficult to achieve optimum tunneling conditions. With an ability to channel light into sub-10 nm gaps, however, one could achieve efficient light-assisted tunneling and envision ways of direct THz frequency rectification (speeds unachievable with electronic devices)¹³. Yet another example is that of broadband (non-resonant) enhancement of radiative transitions achievable in metal-insulator-metal (MIM) gaps.¹⁴ With over 20 \times radiative enhancement already predicted in sub-50 nm gaps, shrinking the gap-size further indicates even further enhancements including the possibility of single-surface plasmon generation through optical¹⁵ and electrical¹⁶ means.

There are several reasons for the difficulty in coupling light into quantum-scale gaps. Primarily, efficient coupling of light from free-space to any metallic or dielectric structure on the quantum-scale is expected to be extremely low.⁵ This fact arises from simple geometric considerations based on overlap area, as well as from a large impedance mismatch between bound propagating modes and free space planewaves. Additionally, there is a practical difficulty in controllable fabrication of sub-10 nm devices with most conventional lithographic patterning methods. Several structures aimed at overcoming the abovementioned difficulties have been proposed. These structures include variations on adiabatic mode transformers^{17,18}, grooved channels^{19,20}, wedges²¹ and, hybrid nanowire waveguides.²²

The ultimate scale of plasmon localization considered in most of the above studies is ≤ 50 nm. We propose a new class of structures that achieve sub 5 nm field localization and are yet amenable to fabrication using simple planar deposition techniques. We consider a strongly confined, bound surface plasmon mode incident on a quantum scale MIM gap structure. The incident surface plasmon mode represents a bound propagating surface charge oscillation that can efficiently couple to the gap MIM provided that the metal interface is continuous and that the insulating dielectric media is uniform. Scattering from the MIM waveguide structure occurs due to the mode mismatch between the incident surface plasmon mode and the modes in the MIM waveguide. Given the range of applications and opportunities for fundamental studies possible with coupling of light into quantum-scale waveguides, we feel it is imperative to accurately quantify the coupling efficiency into such waveguides. Our main objective in the present paper is to provide an analytic and numerical framework for accurate prediction of modal coupling efficiencies into quantum-scale waveguides. This mode matching framework provides physical insight into scattering and coupling processes in quantum-scale waveguides, and lays the ground work for new active plasmonic devices.

Our work differs from the existing body of research in the geometry being considered, the scale of predicted plasmon localization, and the adopted numerical analysis approach. The junctions considered in our paper are between two waveguides: a single metal dielectric interface, and MIM channel having entirely different modal spectra. In contrast, many existing works analyze junctions between structurally identical waveguides that differ only in material parameters (and hence have similar modal spectra).^{23,24} Whereas analyses of waveguide junctions with differing modal spectra certainly exist,^{24,25} our modal analysis method is distinct and provides several independent means for identification and correction of errors in implementation. We can explicitly examine the coupling between modes and evaluate by symmetry modes that are important for scattering and loss. We also augment our rigorous mode-matching results with intuitive and easily calculable effective Fresnel approximations and demonstrate their limitations.

Several noteworthy conclusions emerge from our study. For instance, we find the coupling efficiencies into sub-10 nm channel waveguides to be orders of magnitude higher than those expected from simple scaling arguments based on the ratio of geometrical cross-sections. This large coupling enhancement is due to the strong confinement of the incident single interface plasmon. We also find that, depending on the dielectric material filling the channel, there is usually a specific channel width that maximizes the field-enhancement in the quantum scale gap. Both the field enhancement and the scattering of the incident mode can be qualitatively reproduced with effective Fresnel coefficients, but for accurate numerical prediction the complete mode matching problem needs to be solved. The veracity of our conclusions is confirmed by excellent agreement with direct field solutions performed using finite-difference time domain (FDTD) method. Our numerical studies aid us in forming simple guidelines to improve the coupling efficiencies to quantum-scale waveguides.

We begin in Section II by developing the discrete plasmonic mode-matching theory for coupling from a single-interface plasmon waveguide to a quantum-scale metal-insulator-metal (MIM) waveguide and leading up to the system matrix equations that directly yield scattering coefficients. We numerically illustrate the application of this formalism in Section III by determining coupling efficiency from a single-interface plasmon to a highly confined MIM mode with a 10 nm channel width. In section IV, we assess the accuracy of the mode-matching technique by performing several internal consistency checks and by comparing mode-matching results with finite-difference time-domain (FDTD) method. We study the behavior of the coupling efficiency for varying quantum-scale channel width in Section V. We conclude by providing two simple Fresnel-like approximations for rapid first-order designs of coupling structures to quantum-scale waveguides.

II. PLASMONIC MODE-MATCHING THEORY

In this section, we briefly survey the three semi-analytical methods that are conventionally grouped under mode-matching. Following that, we proceed to describe our method for studying plasmon coupling into quantum-scale gaps. We end this section by obtaining matrix equations for the scattering coefficients of the incident field.

A. Rationale for mode-matching analysis

Most practical nanophotonic devices entail application of numerical methods for their analysis. Contemporary numerical analyses employ either a finite-difference or a finite-element-based subdivision of the geometrical domain of interest. Waveguides geometries like the ones considered in this paper (see Fig. 1) have geometrical features spanning multiple orders of length scales: the quantum-scale gaps are typically sub-10 nm, while the propagation lengths are several micrometers. Finite-difference-based subdivision of such geometries using a uniform grid-size renders the computational analyses memory intensive. Even methods like FEM that allow non-uniform grid sizes tend to be computationally intensive since smooth mesh-size variations are needed to for a reliable analysis (rapid mesh-size variations tend to produce numerical reflections that interfere with physical fields).

Another general feature of field-solvers is the difficulty in extracting the scattering spectra from the calculated field-profiles. This makes it challenging to identify the modes responsible for the electromagnetic behavior of the devices and formulate simple design rules (e.g., the Fresnel-like approximations derived in Section VI).

Mode-matching is a physically intuitive and a computationally efficient technique for analyzing scattering response of electromagnetic structures. This technique is formulated in terms of individual modes of the constituent waveguides at a junction and with the explicit goal of determining coupling between these modes. As a result, scattering coefficients for each participating mode form the natural part of a mode-matching calculation. In particular, scattering coefficients for the modes of interest can simply be read from the list of coefficients generated from the solution of mode-matching equations.

Microwave components are routinely characterized using mode-matching. Recently this technique has gained prominence for analysis of plasmonic structures ranging from planar interfaces²³ to subwavelength waveguide junctions.²⁶ Here we demonstrate the application of this powerful technique to analyze light concentration into quantum-scale waveguides and show that it predicts coupling efficiencies far exceeding the ratio of geometric cross-sections.

B. Brief review of current approaches

With reference to Fig. 1, the algorithm for a typical mode-matching calculation can be enumerated as follows:

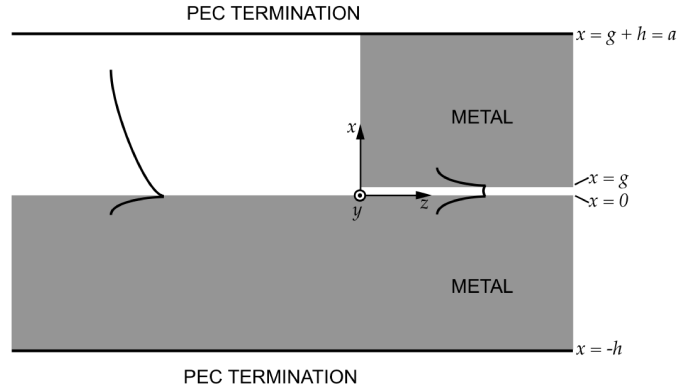


FIG. 1. Schematic of the structure and the co-ordinate system convention. The single-interface plasmon is incident from left on the quantum-scale gap. The structure is assumed to be semi-infinite on both sides in the z -direction and truncated with perfect electric conductor (PEC) termination on top and bottom.

1. First, one determines the set of first N mutually orthogonal modes (mode indices and field profiles) supported individually by the left and the right sides of the junction. This set of modes is expected to be mathematically complete as $N \rightarrow \infty$.
2. In the next step, one calculates the overlap integral of the m^{th} mode on the left side with the n^{th} mode on the right side for all $1 \leq m, n \leq N$.
3. Finally the electric and the magnetic fields continuity at the junction is expressed in terms of the complex scattering coefficients in a linear system of $2N$ equations in terms of the scattering coefficients whose solution yields the desired scattering coefficients.

The choice of modes and the manner of enforcing the field continuity leads to several variants of the basic algorithm presented above. If one chooses to work directly with physically open (unterminated) systems, then the spectrum (i.e., set of modes) consists of a few bound modes and an infinite set of continuous modes. This approach found in Shevchenkov's book²⁷ and has been applied recently by Oulton *et al*²³ to calculate single-interface plasmon reflection at the boundary of two dissimilar metallic/dielectric media. The continuous modes in this approach extend to infinity and carry infinite energy. Considerable care has to be exercised in reinterpreting proper normalization of these modes, both theoretically and computationally. The modes themselves can be obtained simply by a linear variation of the transverse or longitudinal wavevectors and do not require solution of a complex transcendental equation.

An alternative method is to limit the transverse boundaries of the problem by terminating them with a perfect electric conductor (PEC). Because of a forced finite spatial extent, the modes of the PEC terminated structures no longer carry infinite energy and are conveniently normalizable. A PEC termination also leads to *discretization* of the transverse wavevectors that need to be determined by solving for the complex zeros of the dispersion equation. There are two approaches under this method depending on the way \mathbf{E} and \mathbf{H} field continuity conditions are enforced at the junction. In the first approach, field continuities are enforced at a discrete number of points along the junction. More points result in better field continuity as quantified by energy conservation across the junction. Early calculations of Stegeman *et al*^{24,25,28} are based on this approach. The second approach is the formation of equations by taking dot product of the continuity conditions with the set of orthogonal \mathbf{E} and \mathbf{H} fields calculated for the right- and the left-hand sides of the junction. Although continuity at no junction point is ensured by this approach, overall convergence is guaranteed to improve (in a Fourier sense) with inclusion of more modes in the calculation. A detailed discussion of this approach (called the discrete mode-matching technique) is found in the recent article by Kocabas *et al*.²⁶

We adopt the approach of Kocabas *et al*²⁶ and extend it to structures having more than two layers. Our motivation for this approach will become clear in the forthcoming sections. Briefly, the discrete mode-matching approach allows a clear demarcation of the waveguide modes into three distinct classes. Inclusion of modes from each class into the matrix equations is essential for accurate field representation and for obtaining field and Poynting vector continuity at the junction.

C. Discrete mode-matching technique

Since our ultimate goal is to couple light into deep-subwavelength quantum-scale gaps, we examine the waveguide geometry shown in Fig. 1 as our basic structure. On the left ($z < 0$) is a region filled with a uniform dielectric of permittivity ϵ_i (for $x > 0$) and a metal with permittivity ϵ_m (for $x < 0$). The right side of the junction ($z > 0$) is filled with the metal except for a quantum-scale gap opening between $0 \leq x \leq g$. The $z = 0$ plane is the interface between two half-spaces and is called the junction. A surface plasmon traveling on the metal-dielectric interface (termed a single-interface plasmon) is incident from the left toward the junction. Upon encountering the discontinuity at the junction, it undergoes reflection and scattering, and also transfers its energy to the bound mode of the gap (called the gap plasmon or the gap mode). We aim to calculate the field-enhancement for the gap-plasmon mode and the amount of energy transmitted, reflected, and scattered in the coupling process. In order to perform this calculation, we need to determine all the modes supported by the geometries on both sides of the junction and their overlap integrals.

1. Modes of the truncated single-interface waveguide

The left side of the junction supports one bound surface plasmon mode which, in absence of material loss has exponentially decaying field profiles in metal ($x < 0$) and the dielectric ($x > 0$). The single-interface waveguide also supports modes that have oscillating field-profiles in the dielectric, the metal or both. If the structure were not terminated with PEC walls, some of these modes would have undamped oscillations in their field profile and as a result, would carry infinite amount of energy. These modes are conventionally termed ‘continuous’ modes to signify a continuous variation of their transverse wavevector. A PEC termination has the effect of discretizing the transverse wavevector. Moving the PEC discretization boundaries further apart (i.e., increasing h) lessens the magnitude of the discretization bringing the modes closer to those of the semi-infinite waveguide. Although the presence of material loss perturbs this strict classification, the modal spectrum retains the three distinct branches that resemble the spectrum for the lossless waveguides.

The electric and the magnetic field profiles for all three kinds of modes of the single-interface waveguide can be expressed as

$$H_y^S(x) = \begin{cases} \frac{\cosh \kappa_m^S(x+h)}{\cosh \kappa_m^S h}, & -h \leq x \leq 0 \\ \frac{\cosh \kappa_i^S(x-a)}{\cosh \kappa_i^S a}, & 0 < x \leq a \end{cases} \quad (1)$$

$$E_x^S(x) = \begin{cases} \frac{\beta^S Z_0}{\epsilon_m k_0} \frac{\cosh \kappa_m^S(x+h)}{\cosh \kappa_m^S h}, & -h \leq x \leq 0 \\ \frac{\beta^S Z_0}{\epsilon_i k_0} \frac{\cosh \kappa_i^S(x-a)}{\cosh \kappa_i^S a}, & 0 < x \leq a \end{cases} \quad (2)$$

$$E_z^S(x) = \begin{cases} \frac{-i\kappa_m^S Z_0}{\epsilon_m k_0} \frac{\sinh \kappa_m^S(x+h)}{\cosh \kappa_m^S h}, & -h \leq x \leq 0 \\ \frac{-i\kappa_i^S Z_0}{\epsilon_i k_0} \frac{\sinh \kappa_i^S(x-a)}{\cosh \kappa_i^S a}, & 0 < x \leq a \end{cases} \quad (3)$$

The definitions of the various symbols and their expressions in Eqs. (1)–(3) appear in Table I. Superscript ‘S’ denotes the quantities for the single-interface waveguide. Enforcing the continuity of the tangential fields H_y^S and E_z^S at $x = 0$ leads to the following dispersion equation for the truncated single-interface waveguide:

$$\frac{\kappa_m^S}{\epsilon_m} \tanh \kappa_m^S h = -\frac{\kappa_i^S}{\epsilon_i} \tanh \kappa_i^S a \quad (4)$$

Equations (1–4) characterize all possible modes supported by an isolated truncated single-interface waveguide. Note that for $h \rightarrow \infty$, both $\tanh(\kappa_m h)$ and $\tanh(\kappa_i a) \rightarrow 1$ and we recover the familiar single-interface plasmon dispersion

Symbol	Definition	Expression
λ_0	Vacuum wavelength	
k_0	Vacuum wavevector	$2\pi/\lambda_0$
β	Propagation constant	
κ_m	Transverse wavevector in metal	$\sqrt{\beta^2 - k_0^2 \epsilon_m}$
κ_i	Transverse wavevector in dielectric	$\sqrt{\beta^2 - k_0^2 \epsilon_i}$
Z_0	Vacuum impedance	$\sqrt{\mu_0/\epsilon_0}$
p	Polarization factor	ϵ_i/ϵ_m or ϵ_g/ϵ_m
η	$\frac{\kappa_m^G/\epsilon_m}{\kappa_i^G/\epsilon_i} \tanh \kappa_m^G h$	
A	$\cosh \kappa_i^G g + \eta \sinh \kappa_i^G g$	

TABLE I. Definitions of various quantities and their expressions in terms of the material parameters and wavelength.

from Eq. (4). For a lossless metal, these modes can be classified into three categories. Real κ_m and κ_i indicate fields decaying into the metal and the dielectric and thus represent the bound single-interface plasmon mode. Real κ_m and purely imaginary κ_i represents fields decaying into metal and oscillatory in the dielectric. We call these modes type-1 or *T1* modes. The third type of modes have both κ_m and κ_i purely imaginary and their fields are oscillatory in both metal and the dielectric. We call these modes type-2 or *T2* modes. For real metals, values of $\kappa_{m,i}$ (which, for ideal lossless metals were purely real) are now accompanied by a small imaginary part and vice versa. All modes therefore have an oscillatory profile with a decaying envelope. Even so, knowing where the roots lie for the lossless case helps us track their orbits and classify them for the lossy case. Inclusion of *both T1* and *T2* modes in system matrix equations is crucial for accurate representation of the fields.

2. Modes of the truncated MIM waveguide

Modes of the MIM waveguide on the right side of the junction have the same classification as the single-interface case. The field-profiles and the dispersion relation, however, are different due to presence of three layers. The electric and the magnetic field profiles for the truncated MIM waveguide can be written as

$$H_y^G(x) = \begin{cases} \frac{\cosh \kappa_m^G(x+h)}{\cosh \kappa_m^G h}, & -h \leq x \leq 0 \\ (\cosh \kappa_i^G x + \eta \sinh \kappa_i^G x), & 0 < x \leq g \\ A \frac{\cosh \kappa_m^G(x-g-h)}{\cosh \kappa_m^G h}, & g < x \leq a \end{cases} \quad (5)$$

$$E_x^G(x) = \begin{cases} \frac{\beta^G Z_0}{\epsilon_m k_0} \frac{\cosh \kappa_m^G(x+h)}{\cosh \kappa_m^G h}, & -h \leq x \leq 0 \\ \frac{\beta^G Z_0}{\epsilon_g k_0} (\cosh \kappa_i^G x + \eta \sinh \kappa_i^G x), & 0 < x \leq g \\ \frac{A \beta^G Z_0}{\epsilon_m k_0} \frac{\cosh \kappa_m^G(x-g-h)}{\cosh \kappa_m^G h}, & g < x \leq a \end{cases} \quad (6)$$

$$E_z^G(x) = \begin{cases} \frac{-i\kappa_m^G Z_0 \sinh \kappa_m^G(x+h)}{\epsilon_m k_0 \cosh \kappa_m^G h}, & -h \leq x \leq 0 \\ \frac{-i\kappa_i^G Z_0}{\epsilon_g k_0} (\sinh \kappa_i^G x + \eta \cosh \kappa_i^G x), & 0 < x \leq g \\ \frac{-iA\kappa_m^G Z_0 \sinh \kappa_m^G(x-g-h)}{\epsilon_m k_0 \cosh \kappa_m^G h}, & g < x \leq a \end{cases} \quad (7)$$

The superscript ‘G’ denotes that the various quantities belong to the MIM waveguide. Enforcing E_z and H_y field continuity at $x = 0$ and $x = g$ interfaces results in the following dispersion equation for the MIM waveguide:

$$\tanh \kappa_i^G g = \frac{-2p\kappa_i^G \kappa_m^G \tanh \kappa_m^G h}{(\kappa_i^G)^2 + (p\kappa_m^G)^2 \tanh^2 \kappa_m^G h} \quad (8)$$

Equations (5–8) characterize the three types of modes supported by the truncated MIM waveguide.

The success of a mode-matching calculation relies crucially on accurate determination and classification of *all three* complex root types of the transcendental dispersion equations (4) and (8). Appendix A is therefore devoted to transformation and solution of these equations using the fixed-point iteration scheme. Assuming for the moment that we have the set of these roots, we can proceed to compute the mode overlap at the junction (step 2 in the algorithm presented in section II B).

3. Field normalization and overlap integrals

The final step in the mode-matching analysis requires equating the junction fields expressed in terms of the modes on the input (single-interface) side to those expressed in terms of the modes of the output (MIM) side. To convert the continuity equations into those for the undetermined scattering coefficients, we take the inner product of the fields on one side with those on the other side. This operation involves calculation of the field normalization coefficients and the overlap integrals. Non-hermiticity of the Helmholtz operator for metallic waveguides requires a careful redefinition of the inner product. Some recent papers^{26,29} consider these issues in detail and recommend defining the inner product of any two modes m and n as:

$$\langle E_{x,m} | H_{y,n} \rangle \equiv \int_{-\infty}^{\infty} E_{x,m} H_{y,n} dx \quad (9)$$

For any two modes of the single-interface waveguide, we have the orthogonalization condition:

$$\langle E_{x,m}^S | H_{y,n}^S \rangle = \Omega_S(m) \delta_{mn} \quad (10)$$

where Ω_S is the normalization coefficient of the m^{th} single-interface waveguide mode and δ_{mn} is the Kronecker delta function. For MIM waveguide modes we have, analogously,

$$\langle E_{x,m}^G | H_{y,n}^G \rangle = \Omega_G(m) \delta_{mn} \quad (11)$$

We also require two types of overlap integrals defined as:

$$C_{GS}(m, n) \equiv \langle E_{x,m}^G | H_{y,n}^S \rangle \quad (12)$$

$$D_{SG}(m, n) \equiv \langle E_{x,m}^S | H_{y,n}^G \rangle. \quad (13)$$

These overlap integrals define interface transfer matrices that represent the coupling from incident to gap region, and gap to incident region, respectively.

At this point, one has a choice of evaluating the overlap integrals using either a direct numerical integration or substituting the field profiles listed in equations (1), (2), (5), and (6) to obtain analytical expressions for C_{GS} and D_{SG} . In our experience, evaluation using analytical expressions is significantly faster and more accurate than numerical integrations. This is due to the fact that integrands for the higher-order $T1$ and $T2$ modes are highly oscillatory in the domain of integration. This sometimes results in numerical integration routines experiencing instabilities. In Appendix B, we provide explicit expressions for the normalization and overlap integrals for the geometry in Fig. 1.

Equation	Take inner product with			
Eq. (14)	$E_G \times$	$E_S \times$	$E_G \times$	$E_S \times$
Eq. (15)	$\times H_G$	$\times H_S$	$\times H_S$	$\times H_G$

TABLE II. Choices of inner products for forming the scattering matrix of the system.

4. Matrix equations for scattering coefficients

Once the normalization and the overlap integrals are evaluated, the final step is setting up the matrix equations in terms of the scattering coefficients. If we assign $m = 1$ as the single-interface plasmon mode incident from the left, the \mathbf{E} - and \mathbf{H} -field continuity equations at the junction can be expressed as:

$$H_{y,1}^S + \sum_{m=1}^N r_m H_{y,m}^S = \sum_{n=1}^N t_n H_{y,n}^G \quad (14)$$

$$E_{x,1}^S - \sum_{m=1}^N r_m E_{x,m}^S = \sum_{n=1}^N t_n E_{x,n}^G \quad (15)$$

In equations (14) and (15) r_m and t_n are (yet undetermined) amplitude reflection and transmission coefficients, and N is the total number of modes in the calculation and includes the bound plasmon modes, $T1$, and $T2$ modes. Starting from Eqs. (14) and (15), we have four choices for forming matrix equations for r_m and t_n . These choices arise from the field-pairs we choose to take the inner products of Eqs. (14) and (15) with and are depicted in the columns of Table II. We implemented all four choices and found that the first two, involving \mathbf{E} - and \mathbf{H} -fields from the same side, lead to ill-conditioned matrix equations. The other two, involving fields from different sides, lead to well-formed equations and yield nearly identical results. Premultiplying Eq. (14) with $E_{x,k}^G$ and integrating over x yields

$$\begin{aligned} \langle E_{x,k}^G | H_{y,1}^S \rangle + \sum_{m=1}^N r_m \langle E_{x,k}^G | H_{y,m}^S \rangle \\ = \sum_{n=1}^N t_n \langle E_{x,k}^G | H_{y,n}^G \rangle \end{aligned} \quad (16)$$

Using equations (11)–(13), Eq. (16) can be recast as

$$\sum_{m=1}^N r_m C_{GS}(k, m) - t_k \Omega_G(k) = -C_{GS}(k, 1) \quad (17)$$

Postmultiplying Eq. (15) with $H_{y,k}^S$ followed by integration over x yields

$$\begin{aligned} \langle E_{x,1}^S | H_{y,k}^S \rangle - \sum_{m=1}^N r_m \langle E_{x,m}^S | H_{y,k}^S \rangle \\ = \sum_{n=1}^N t_n \langle E_{x,n}^G | H_{y,k}^S \rangle, \end{aligned} \quad (18)$$

which transforms after simplifying using equations (10), (12), and (13) to

$$r_k \Omega_S(k) + \sum_{n=1}^N t_k C_{GS}(n, k) = \Omega_S(1) \delta_{1k} \quad (19)$$

Equations (17) and (19) yield a system of $2N$ equation when k assumes values between $1 \leq k \leq N$.

$$\begin{pmatrix} -\mathbb{C}_{\text{GS}} & \text{diag}(\Omega_{\text{G}}) \\ \text{diag}(\Omega_{\text{S}}) & \mathbb{C}_{\text{GS}}^{\text{T}} \end{pmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{t} \end{bmatrix} = \begin{bmatrix} C_{\text{GS}}(1, 1) \\ C_{\text{GS}}(2, 1) \\ \vdots \\ C_{\text{GS}}(N, 1) \\ \Omega_{\text{S}}(1) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (20)$$

where \mathbb{C}_{GS} is the $N \times N$ transfer matrix with elements $\mathbb{C}_{\text{GS}m,n} = C_{\text{GS}}(m, n)$, $\mathbf{r} = [r_1, r_2, \dots, r_N]^{\text{T}}$, $\mathbf{t} = [t_1, t_2, \dots, t_N]^{\text{T}}$, and $\text{diag}(\Omega_{\text{S,G}})$ is an $N \times N$ matrix having entries $\Omega_{\text{S,G}}(k)$ ($1 \leq k \leq N$) along its diagonal and zero elsewhere. Following a similar procedure with the fourth option in Table II yields an equivalent equation

$$\begin{pmatrix} -\text{diag}(\Omega_{\text{S}}) & \mathbb{D}_{\text{SG}} \\ \mathbb{D}_{\text{SG}}^{\text{T}} & \text{diag}(\Omega_{\text{G}}) \end{pmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{t} \end{bmatrix} = \begin{bmatrix} \Omega_{\text{S}}(1) \\ 0 \\ \vdots \\ 0 \\ D_{\text{SG}}(1, 1) \\ D_{\text{SG}}(1, 2) \\ \vdots \\ D_{\text{SG}}(1, N) \end{bmatrix} \quad (21)$$

Solution of the linear equations (20) or (21) directly provides the amplitude reflection and transmission coefficients. The corresponding energy scattering coefficients for the k^{th} mode, R_k and T_k , may be obtained from r_k and t_k using the relations

$$R_k = |r_k|^2 \frac{S_{\text{S}}(k)}{S_{\text{S}}(1)} \quad (22)$$

$$T_k = |t_k|^2 \frac{S_{\text{G}}(k)}{S_{\text{S}}(1)} \quad (23)$$

where $S_{\text{S}}(k)$ ($S_{\text{G}}(k)$) is the Poynting vector of the k^{th} mode of the single interface (MIM) waveguide. Appendix B 3 provides explicit expressions for $S_{\text{S}}(k)$ and $S_{\text{G}}(k)$.

III. CALCULATING LIGHT TRANSMISSION INTO QUANTUM-SCALE GAPS

The previous section presented the steps in the discrete mode-matching algorithm leading up to direct matrix equations for scattering coefficients. In this section, we illustrate the implementation of the mode-matching algorithm. We do this in steps starting with generation of a complete orthogonal set of modes of structure on both sides of the junction followed by setting up of the matrix equations. The final result of our calculation is transmission efficiency of a single-interface plasmon into a 10 nm gap sandwiched between two symmetric metals.

For our illustrative calculation, we consider a symmetric gold MIM diode with a gap $g = 10$ nm, operating at 850 nm. The permittivity of gold at 850 nm is assumed to be $\epsilon_{\text{m}} = -27 - i0.83$.³⁰ The dielectric insulator on both sides of the junction is silicon dioxide ($\epsilon_{\text{i}} = 2.1025$). We truncate the domain by placing PEC walls at $x = -2h = -1700$ nm and $x = a = 2h + g = 1710$ nm.

A. Modes of the single-interface waveguide ($z < 0$)

Equation (4) governs the various modes of the single-interface waveguide. Any complex root-finding routine should, in principle, be able to locate all the complex roots of Eq. (4). In practice, however, root-finding algorithms are

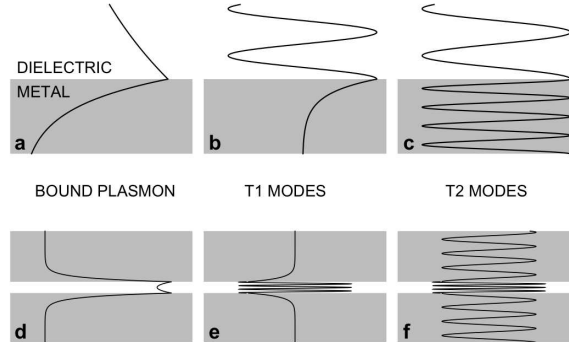


FIG. 2. Magnetic field (H_y) schematics of the three mode types supported by the single-interface and MIM waveguides for an ideal lossless metal. (a) and (d) represent the bound surface/gap plasmon modes with exponentially decaying field profile in the metal and the dielectric. Plots (b) and (e) are the $T1$ modes with field profiles that are oscillatory in the dielectric and exponential in the metal. Finally (c) and (f) represent the $T2$ with oscillatory field profiles in both metal and the dielectric.

sensitive to the initial guess and often return only those roots that are closest to the initial guess. The complex roots of the truncated metallic waveguides are spaced closely together and direct root-searching seldom returns all the required roots. The situation is remedied by solving the dispersion equations with the fixed-point iteration (FPI) method by transforming them into forms that directly yield the roots belonging to a specific branch.³¹ These transformations for a single-interface waveguide are as shown in Appendix A 1.

1. Bound mode: The single-interface plasmon

FPI solution of the Eq. (4) for the bound single-interface plasmon mode yields its propagation constant as $\beta^S/k_0 = 1.5099244 - i0.0019570$. The imaginary part of β^S represents the propagation loss that arises due to the imaginary part of the metal permittivity. We note that this mode is an approximation to the true plasmon mode traveling at the interface of semi-infinite metal and dielectric half-spaces. The propagation constant of the true mode is given by the well-known formula $\beta_\infty^S/k_0 = \sqrt{\epsilon_m \epsilon_i / (\epsilon_m + \epsilon_i)}$. For the material parameters in our example we obtain $\beta_\infty^S/k_0 = 1.5099188 - i0.0019578$. PEC truncation at $\pm 2\lambda$ thus introduces an error of about 0.0414% ($3.6790 \times 10^{-4}\%$) in the imaginary (real) parts of the propagation constants. Pushing the truncation boundary to $\pm 4\lambda$ reduces the error to $2.290 \times 10^{-6}\%$ ($8.920 \times 10^{-9}\%$). Convergence of the imaginary part of β_{sp} is slower with increasing truncation boundary than that of the real part. For all the computations presented in this paper, we have kept the truncation boundary at $\pm 2\lambda$.

2. $T1$ and $T2$ modes

For lossless materials, $T1$ modes are characterized by a real κ_m and an imaginary κ_i . The $T1$ modes are exponentially decaying into the metal but oscillatory in the dielectric media. Eq. (4) can be transformed to reflect this fact and solved using fixed-point iteration using Eq. (A3). For a truncated single-interface waveguide, there are denumerably infinite number of $T1$ modes. Figure 3(a) shows the real and the imaginary parts of the first 200 $T1$ modes. $T2$ modes of a lossless single-interface waveguide are characterized by imaginary values of both κ_m and κ_i indicating purely oscillatory fields in the metal and the dielectric. Once again, a transformation of Eq. (4) reflecting this fact allows us to obtain the $T2$ roots using Eq. (A5). The real and imaginary parts of the first 309 $T2$ roots are plotted in Fig. 3(b).

Whereas the continuous roots of a single-interface waveguide having real metals lie close to the imaginary axis as nominally expected, Fig. 3 reveals the intricate structure of the spectrum. This structure originates from the interplay of metal loss and truncation. For a fixed metal loss, extending the truncation boundary leads to lesser peak deviation from the imaginary axis, but increases the number of oscillations. On the other hand, for a fixed truncation boundary, increasing metal loss leads to damping of the oscillations (in both $T1$ and $T2$ modes). One possible explanation of the oscillatory behavior in Fig. 3 is found in reference²⁶ where it is explained in terms of coupled mode-splitting of a 1-D metallic photonic crystal waveguide.

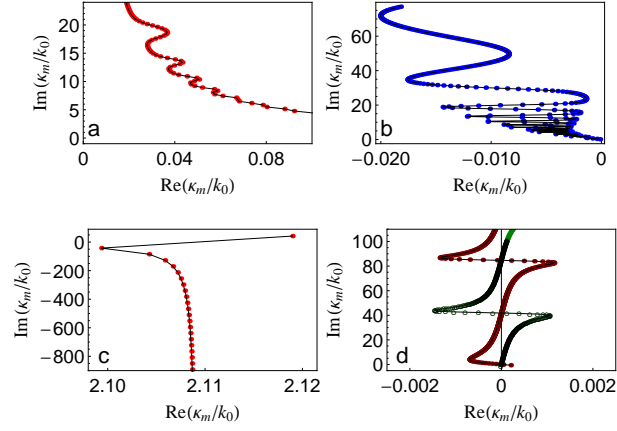


FIG. 3. Real and imaginary parts of the $T1$ and $T2$ continuous roots of the truncated single-interface waveguide [(a) and (b)] and MIM waveguide [(c) and (d)]. In both waveguides, the metal is gold and the dielectric is silicon dioxide.

B. Modes of the MIM waveguide ($z > 0$)

Determining the MIM waveguide modes proceeds in a way similar as those of the single-interface waveguide. The governing dispersion equation (Eq. (8)) has a richer set of modes due the three-layered structure. The MIM waveguide, has a reflection symmetry about the line $x = g/2$ imparting an odd/even symmetry to its modes. Transforming Eq. (8) to reflect this symmetry significantly eases obtaining its roots to a great degree of accuracy. These transformation are specified in Appendix section A 2.

1. Bound mode: The gap plasmon

Using the transformation of Eq. (8) in the form show in Eq. (A7), we solve for the gap plasmon mode of the truncated lossy MIM waveguide to obtain $\beta^G/k_0 = 3.9994922 - i0.0440002$. The gap plasmon mode of the truncated MIM waveguide is an approximation to the bound mode of the gap between two semi-infinite metals (having a propagation constant β_∞^G). Within the 16 digit numerical precision of our computation, β^G was identically equal to β_∞^G . Unlike the single-interface plasmon mode, the gap plasmon mode of the truncated MIM waveguide has a vanishingly small value at both PEC boundary. As a result, its propagation constant is practically unchanged from the that of the true gap plasmon mode.

2. $T1$ and $T2$ modes

Transformations of Eq. (8) appropriate for obtaining $T1$ and $T2$ roots of the MIM waveguide are shown in Appendix sections A 2 b and A 2 c. Figure 3(c) shows first 22 $T1$ continuous roots of the truncated IM structure. The imaginary part of these roots reaches very high values ($\kappa_m \simeq 800k_0$) even for the first few roots. As a result, these modes corresponding to these roots have highly oscillatory field profiles in the metals. To obtain accurate convergence of scattering coefficients in a practical mode-matching calculation, one needs to include roots only up to $\kappa_m \simeq 25k_0$. Consequently, $T1$ continuous modes of quantum-scale gaps beyond the first few do not couple to any modes of the single-interface plasmon. Only first 3–5 need to be included to obtain accurate field and scattering coefficient convergence.

Figure 3(d) plots the first 459 $T2$ continuous roots of the truncated MIM waveguide. The odd and the even roots have distinct orbits and are interlaced with each other. $T2$ continuous modes of the MIM waveguide couple efficiently to both $T1$ and $T2$ modes of the single-interface waveguide and consequently play an important role in the scattering of the single-interface plasmon into the gap mode. For accurate implementation of the mode matching algorithm, it is critical to not miss any $T2$ modes of the MIM waveguide in the range of 0–25 k_0 .

Parameter	Value
λ_0 (Vaccum wavelength)	850 nm
ϵ_m (Metal permittivity)	$-27 - i0.83$
ϵ_i (Insulator permittivity)	2.1025
g (Gap width)	10 nm
h (PEC truncation boundary)	$2\lambda_0$
$T1$ modes of single-interface waveguide	200
$T2$ modes of single-interface waveguide	309
$T2$ modes of MIM waveguide	200
$T2$ modes of MIM waveguide	469

TABLE III. List of parameters used for the mode matching algorithm

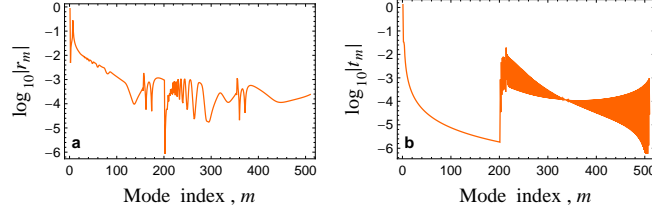


FIG. 4. (a) Amplitude reflection and (b) amplitude transmission coefficients for all modes obtained from solutions for matrix equations (20) and (21) using the parameters in Table III.

C. Orthogonality and completeness of the modes

Mutual orthogonality and completeness of the single-interface and MIM waveguide modes was directly assumed in constructing the matrix equations (17) and (19). A validation of this assumption is critical for ensuring the accuracy of the resulting equations (20) and (21). A rigorous proof of the orthogonality and completeness of the obtained set of modes belongs to the realm of generalized Sturm-Liouville theory and is outside the scope of our paper. Moreover, such a general proof is of limited use in ensuring the completeness of modes obtained numerically. We therefore perform a *numerical* test of modal orthogonality and completeness as described in Appendix C. We emphasize that obtaining the numerical verification of mode orthogonality and completeness is a crucial step before proceeding to enforcing field continuity at the junction and calculation of scattering coefficients.

D. Solution of system matrix equations

The final step in the mode-matching algorithm is enforcement of tangential field continuity (H_y and E_x) at the junction ($z = 0$ plane). This leads to the matrix equation system (20) and (21), either of which can be solved to directly obtain the scattering coefficients. We set up the system using the parameters listed in Table III.

Figure 4 (a) shows the scattering amplitude for the modes on the single-interface waveguide. The coefficient of any given mode is indicative of how strongly it couples to the incident single-interface plasmon. The $m = 1$ is the single interface plasmon mode and it has an amplitude reflection coefficient of $|r_1| = 0.849191$.

Transmission coefficients for the modes of the MIM waveguide are shown in Fig. 4(b). The gap plasmon mode has a transmission coefficient of $|t_1| = 1.32457$ indicating a field enhancement of $\simeq 32.5\%$ upon its transmission into the gap. The transmission coefficients for $m \geq 2$ represents the scattering amplitudes into these higher order modes.

IV. ERROR ANALYSIS OF MODE-MATCHING TECHNIQUE

In this section, we aim to establish the conceptual and numerical validity of the mode-matching calculation presented in the previous section. In order to extract numerical conclusions from the scattering coefficients, it is essential that the convergence and estimates of the numerical accuracy be obtained. We perform this error analysis by applying several

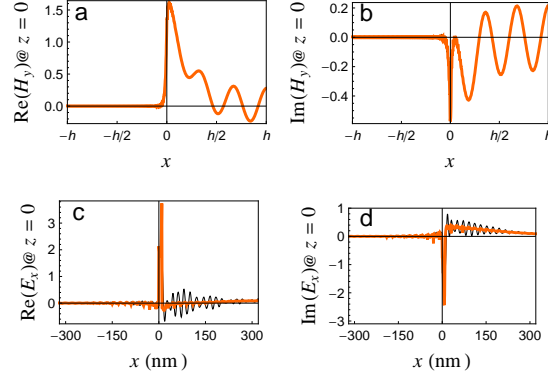


FIG. 5. (a) Real and (b) imaginary parts of the junction magnetic field reconstructed from the scattering coefficients. (c) and (d) show the real and imaginary parts of the reconstructed electric fields. The dashed (black) and continuous (orange) lines are left- and right-handed reconstructions, respectively, from the single-interface waveguide and MIM waveguide modes. Note the different x -axis scale on the electric field plots (c) and (d). Only the electric-field profile between ± 300 nm is shown to highlight the differences in the reconstructed fields.

internal consistency checks to the scattering amplitudes, by verifying the error in continuous/conserved quantities and by comparing mode-matching-obtained field profiles against independent finite-difference time-domain calculations.

1. Verifying upper bound on r_1

The coefficient r_1 can be simply read from Fig. 4(a) and verified to be less than unity. There is no such simple upper bound on the other coefficients. For example, t_1 can have values greater than unity without violating power conservation. This is because t_1 represents the peak field amplitude of the gap plasmon mode. On the other hand, the total power in the mode is related to the integrated area under the field curves. The gap plasmon field can have a higher peak value and still have less power than the incident field.

2. Verifying field continuity

Although electric and magnetic field continuity was enforced in construction of system matrix equations, the matching of the field values is satisfied only in a Fourier sense. A point-by-point field matching can only be achieved by including the infinitely many modes supported by the system. It is therefore a useful check to visualize the quality of field-continuity at the junction. We perform this check by reconstructing the junction fields in two different ways from Eq. (14). Using the left hand side of Eq. (14), we can write the left-hand reconstruction of the junction magnetic fields as:

$$H_y^L(x, z = 0) = \sum_{m=0}^N r_m H_{y,m}^S \quad (24)$$

Similarly, using the right-hand reconstruction can be expressed as

$$H_y^R(x, z = 0) = \sum_{m=0}^N t_m H_{y,m}^G \quad (25)$$

The left- and right-hand reconstructed fields are plotted in Fig. 5. The real and imaginary parts of the reconstructed magnetic fields displayed in Fig. 5(a) and (b) show an excellent match along the entire junction. The electric fields in (c) and (d) match > 150 nm away from the junction but show a poor match in and around the gap. This is a consequence of the large discontinuity in the electric field induced by the significant permittivity mismatch (-27.0 versus 2.1). The quality of the electric field matching improves with increasing number of modes. The close matching of electric-fields, however, is unnecessary for obtaining accurate scattering coefficients as shown in subsection IV 4.

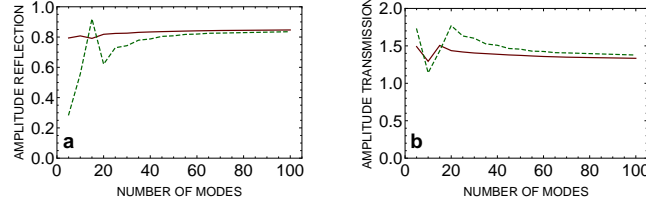


FIG. 6. Convergence of (a) amplitude reflection coefficient $|r_1|$ and (b) amplitude transmission coefficient $|t_1|$ with increasing number of modes in a mode-matching calculation. The x -axis denotes the number of modes included in the calculation from each part of the spectrum for the respective side of the junction. For example $x = 20$ signifies 1 bound mode (single interface plasmon for $z < 0$ and MIM gap plasmon for $z > 0$), 20 $T1$ modes and 20 $T2$ modes. The solid red and the dashed green lines represent $|r_1|$ and $|t_1|$ calculated using the equivalent systems of equations Eq. (20) and Eq. (21) respectively.

3. Verifying power conservation

Truncating the modal expansion at $n = N$ introduces negligible but finite error in the field continuity at the interface. One of the ways to quantify the error is to compute the scattered power in all modes and check for power conservation. For the formalism developed in section II C 4, the power conservation condition can be written as

$$\sum_{k=1}^N R_k + \sum_{k=1}^N T_k = 1 \quad (26)$$

Computing the above sum using Eq. (22) and (23), and r_k and t_k obtained through our solution (Fig. 4), we evaluate the above sum to be 0.9992 which represents about 0.082% error in energy conservation. In Eq. (26), $R_1 = 72.3102\%$ represents reflection into the single-interface plasmon mode and $T_1 = 23.5551\%$ is energy transmission into the MIM mode. Together, R_1 and T_1 account for 95.8656% of the energy. The rest 4.1344% is scattering into single interface and MIM modes $k \geq 2$.

4. Convergence with number of modes

The final consistency check considered for the present mode-matching implementation is convergence of the the scattering coefficients with the number of modes. Whereas convergence, by itself, does not indicate accuracy (i.e., coefficients may converge to wrong value), large variations in the coefficients points to possible errors in finding and including proper roots or the implementation of mode-matching algorithm. Figure 6 shows the how the surface plasmon amplitude reflection and transmission convergence with inclusion of increasing number of modes from the single-interface and MIM sides of the junction. Inclusion of about 100 modes results in the scattering coefficients converging to within $< 0.1\%$ of their final values. Furthermore, the r and the t coefficients calculated using the equivalent matrix equations (20) and (21) approach to the same value with increasing number of modes.

5. Comparison with FDTD calculations

In addition to the internal consistency checks considered in subsections IV 1–IV 4, we compared the results of the mode-matching algorithm against the full-field finite-difference time-domain simulations. As shown in Fig. 7, a surface plasmon is incident from the left side and undergoes reflection and transmission upon encountering the junction. The simulation boundary is terminated with perfectly matched layers (PMLs) that absorbs the outgoing MIM gap plasmon and simulates an ideal semi-infinite MIM waveguide in the $z > 0$ region. The resulting electric field profile is shown in Fig. 7(a). For comparison with mode-matching method, we calculate the electric field along the AA' cutline shown by the dotted line. The electric field plots calculated with FDTD and the mode-matching method are overlaid in Fig. 7(b). We note that no normalization has been applied to the calculated fields—they are plotted *as is*. It is seen that the two fields match closely within the gap and are practically indistinguishable away from it. The slight difference between in the fields within the gap is caused by FDTD mesh discretization. More importantly, the field enhancement in the gap (i.e., the amplitude transmission coefficient) calculated using the two methods is in excellent agreement. A favorable comparison with the FDTD calculations puts the mode-matching algorithm on a firmer footing as a tool for analysis and optimization of light-coupling into quantum-scale gaps.

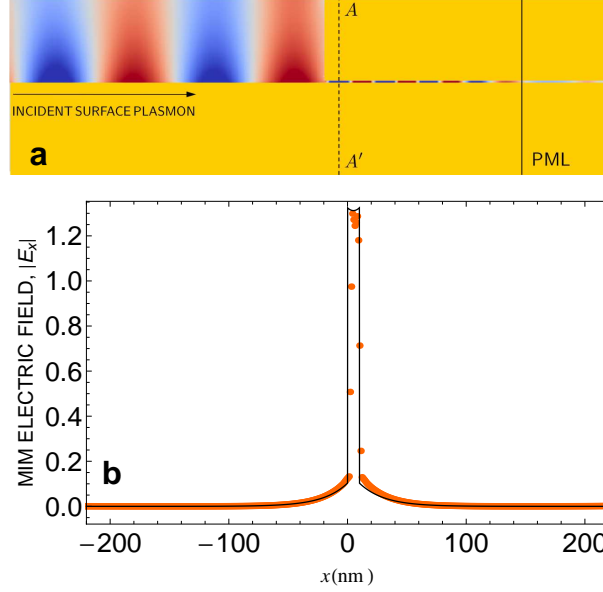


FIG. 7. (a) Electric field profile in the MIM waveguide calculated using finite difference time domain (FDTD) method. The electric field taken along the cutline AA' is used for comparison with mode matching algorithm. (b) Comparison of the electric fields calculated using the mode matching method (black solid line) and FDTD method (orange, filled circles). In (b) the electric field profiles are plotted ‘as is’ (i.e., they are unnormalized).

V. VARIATION OF THE COUPLING EFFICIENCY WITH CHANNEL WIDTH

Previous sections detailed various steps in the implementation of the mode-matching algorithm and provided several verifications of the accuracy of the obtained solutions. Having demonstrated the possibility of obtaining $> 20\%$ coupling into a 10 nm gap, one naturally inquires about the behavior of the coupling efficiency for smaller and larger gap sizes. It is of particular interest to know if decreasing gap dimensions result in progressively larger field enhancements, and if so what are the limits of such field-enhancements. This information would be crucial to designing plasmonic devices that aim to harness the quantum-scale light-matter interactions. The present section is devoted to exploring the above questions by applying the mode-matching theory to examine the behavior of the coupling efficiency as a function of the gap size. The entire mode-matching calculation, including root-searching of the single-interface and MIM waveguide dispersion equations, has to be carried out for every gap size considered. In addition, the accuracy of the calculations has to be verified using the consistency checks detailed in section IV.

The dimension of the gap in the MIM waveguide decides the spatial extent of the waveguide mode. This spatial extent, in turn, impacts the the overlap between the incident and the transmitted modes and determines the coupling efficiency. To analyze the effect of gap size on the coupling, we work with the same parameters listed in Table III but now vary the gap smoothly between 1 nm and 50 nm. These gap sizes cover practically realizable quantum-scale dimensions (using, for example, thin-film deposition techniques) and link up to the familiar regime of EOT ($g \geq 50$ nm). Figure 8 shows the influence of gap variation on the transmission and reflection of the surface plasmon mode.

As shown in Fig. 8(a) and (c) the amplitude and energy reflection increase monotonically toward perfect reflection with decreasing gap-size. This behavior is expected since the single-interface plasmon encounters an increasingly complete metallic wall with decreasing gap size. What is surprising, however, is that $> 10\%$ of the incident energy is coupled even into a gap as small as 3.5 nm as evident in Fig. 8(d). Geometrically, the ratio of the cross-sections of the two waveguides forming the junction is roughly $[\mathbb{R}(\kappa_i^G)] g \simeq 0.87\%$ where $[\mathbb{R}(\kappa_i^G)]^{-1}$ is the decay length of the single-interface plasmon mode in the air and is a measure of its spatial extent. The energy coupling efficiency is therefore nearly *12 times* the ratio of geometrical cross-sections. An alternative way to gauge the efficiency is to compare the fraction of the coupled energy to the ratio of spatial extents of the incoming and the transmitted modes. The spatial extents, L^S and L^G , of the single-interface and the gap plasmon modes are proportional to the sum of the reciprocals of their transverse wavevectors in the metal and the dielectric: $L^S = [\mathbb{R}(\kappa_i^S)]^{-1} + [\mathbb{R}(\kappa_m^S)]^{-1}$ and

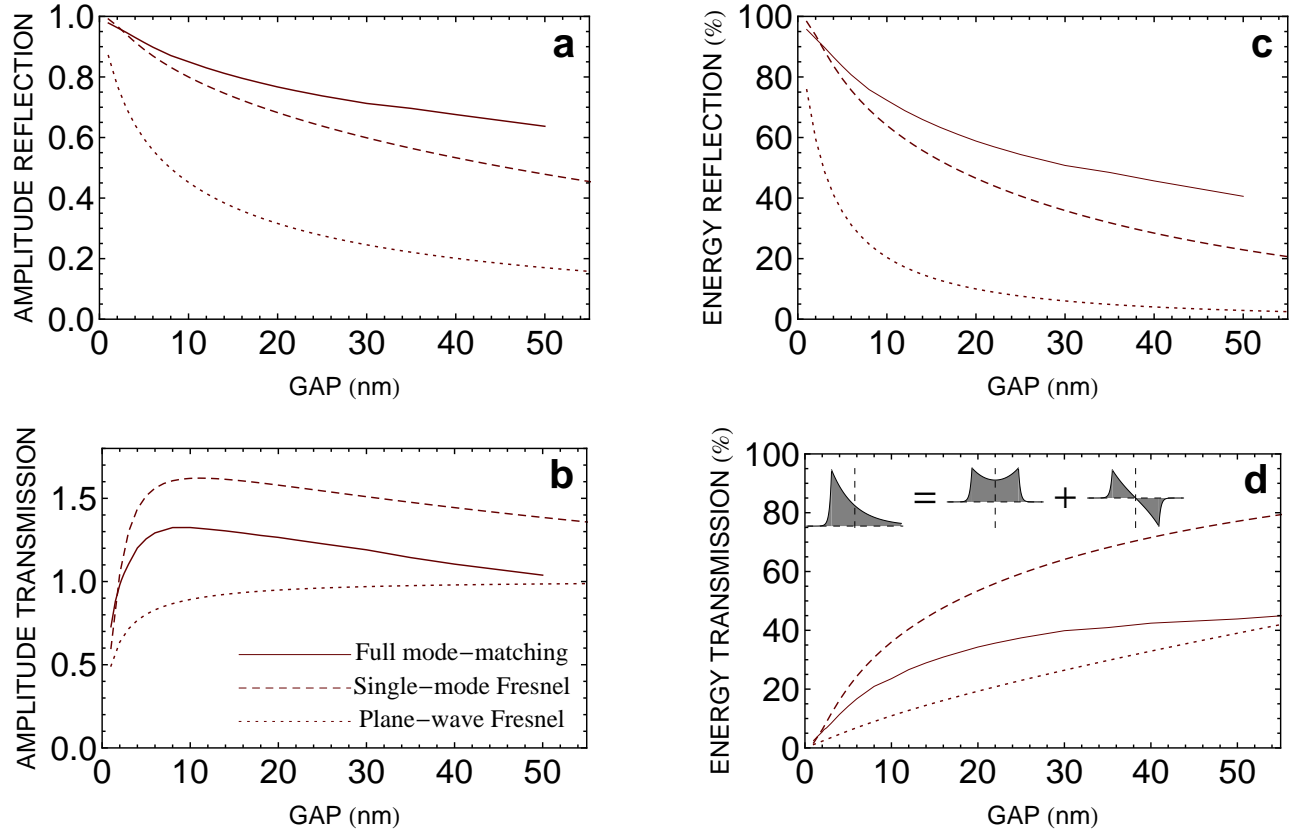


FIG. 8. Influence of gap-size variation on transmission/reflection efficiency. The transmission is into the symmetric gap plasmon mode only. The various parameters used for this calculation are listed in Table III with the exception of the gap size g which is varied smoothly between $1 \text{ nm} \leq g \leq 50 \text{ nm}$. The dashed and dotted line represent the transmission/reflection coefficients calculated by the single-mode and plane-wave Fresnel approximations, respectively.

$L^G = [\Re(\kappa_i^G)]^{-1} + [\Re(\kappa_m^G)]^{-1}$. For $g = 3.5 \text{ nm}$ as considered above, $L^G/L^S \simeq 8.7\%$ and is comparable to the fraction of the energy coupling. The ratio of the modal extents is thus a better predictor of the coupling efficiency than the geometrical cross-sections.

The amplitude transmission coefficient represents the field-enhancement factor and shows a non-monotonic variation with decreasing gap size as shown in Figure 8(c). Decreasing the gap-size from 50 nm results in progressively increasing field-enhancement till about a $g = 10 \text{ nm}$. In the same range of gap sizes, the total energy transmission shows a monotonic *decrease* as seen in Fig. 8(d). Whereas the peak value of the field is indicative of the field enhancement, the power carried by the mode is proportional to its peak value *and* its transverse extent. Thus, the anti-correlation in figures 8(c) and 8(d) for $10 \text{ nm} \leq g \leq 50 \text{ nm}$ implies that the transverse mode extent shrinks *faster* with decreasing gap size than the increase of field enhancement.

For $g < 10 \text{ nm}$, the energy transmission continues to decrease monotonically, but the amplitude transmission reverses its increasing trend and now shows a monotonic *decrease* with decreasing gap size. For the materials and geometric parameters of this problem, the amplitude transmission (field-enhancement) is maximized at $g \simeq 8 \text{ nm}$ with a value close to 33%. The energy transmission corresponding to the maximum field enhancement is $\simeq 20\%$. The existence of such a maximum can be explained by examining the behavior of the effective index of the gap plasmon mode as a function of the gap size shown in Fig. 9(a). The attenuation of the mode (proportional to $\Im(\beta^G)$) increases sharply for $g < 10 \text{ nm}$ causing the field amplitude inside the gap to drop rapidly.

For gap sizes $\geq 50 \text{ nm}$, the amplitude and energy transmissions asymptotically approach $1/\sqrt{2}$ and $1/2$, respectively. The reason for this behavior can be understood by noting that Fig. 8 (d) shows the energy transmission into the *even* gap mode. As the gap-size increases, the effective index of this mode approaches that of the single-interface plasmon leading us to expect perfect (100%) transmission. Yet, it is crucial to remember that the symmetry of MIM waveguide imposes a symmetry on the even mode whereas a single-interface plasmon mode lacks a definite symmetry. Thus, even for an arbitrarily large gap size, the field profiles of the single-interface plasmon and the symmetric gap plasmon do not overlap beyond the plane of symmetry of the MIM waveguide. As a result, the symmetric mode by itself cannot

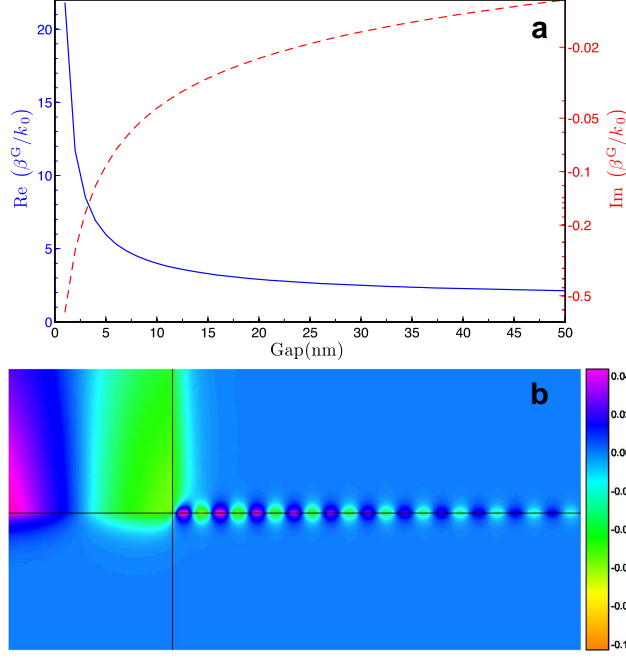


FIG. 9. (a) The variation of the real (blue, solid line) and the imaginary (red, dashed line) of the MIM waveguide propagation constant with decreasing gap size. (b) Full-field calculation of coupling phenomenon into a 1 nm gap verifying the cutoff-like phenomenon predicted by modematching calculations Fig. 8(c).

couple with 100% efficiency regardless of the gap size. The antisymmetric (odd) gap plasmon mode is essential in order to reconstruct the non-symmetric single-interface plasmon and draw out its entire energy. In the limit of large gap size the nearly degenerate even and odd modes carry equal energy. This required superposition is depicted in the inset to Fig. 8(d). The amplitude transmission of the even mode therefore asymptotically approaches $1/\sqrt{2}$ following the limiting value of 50% energy coupling.

The field amplitude inside the gap directly impacts the strength of several quantum optical phenomenon including electron tunneling, absorption, and emission. An existence of a gap size that maximizes the transmitted field amplitude could prove crucial for design of nano-optical devices intending to rely on extreme light concentration for their operation.

VI. FRESNEL FORMULAS FOR LIGHT CHANNELING INTO QUANTUM-SCALE GAPS

Having obtained the coupling and the reflection coefficients via a rigorous mode-matching analysis, it is of interest to see the effectiveness of the well-known Fresnel reflection formulas in reproducing these observations. Such formulas are often useful tools for rapid first-order designs and a quantification of their range of validity is potentially valuable. In this section, we provide two Fresnel-like formulas based respectively on a plane wave and single-mode approximations.

A. Plane-wave approximation

For normal incidence, the reflection and transmission coefficients are given by the well-known formulas:

$$r_{\text{SP}} = \frac{\beta^{\text{G}} - \beta^{\text{S}}}{\beta^{\text{G}} + \beta^{\text{S}}}, \quad (27)$$

$$t_{\text{GP}} = \frac{2\sqrt{\beta^{\text{G}}\beta^{\text{S}}}}{\beta^{\text{G}} + \beta^{\text{S}}}, \quad (28)$$

$$R_{\text{SP}} = |r_{\text{SP}}|^2, \quad (29)$$

$$T_{\text{GP}} = |t_{\text{SP}}|^2 \frac{S_{\text{G}}(1)}{S_{\text{S}}(1)}. \quad (30)$$

In Eqs. (27)–(30), r_{SP} (t_{GP}) is the surface (gap) plasmon amplitude reflection (transmission) coefficient. The subscript ‘P’ signifies the use of the normal-incidence, plane-wave Fresnel approximation. In this approximation the energy transmission is usually obtained as $T_{\text{GP}} = 1 - R_{\text{GP}}$. Use of this formula, however, resulted in over a 100% deviation from the exact mode-matching result. We therefore calculated the amplitude transmission coefficient from the plane-wave formula, but obtained energy transmission by weighing the amplitude transmission by the ratio of the Poynting vectors of the single-interface and the gap plasmon modes.

The approximate amplitude and energy transmission/reflection coefficients calculated using plane-wave Fresnel formulas are plotted by dotted line in Fig. 8. It is seen that the plane-wave approximation significantly underestimates both the energy and amplitude reflection for all gap sizes considered. The amplitude transmission is significantly underestimated for smaller gap sizes, but converges toward the exact result for increasing gap sizes.

B. Single-mode approximation

From the discussion in section IV 3, we note that for a 10 nm gap, modes other than the fundamental bound modes on both sides carry about 4% energy that we term as scattering. Within this energy error, we may construct Fresnel formulas for transmission/reflection coefficients retaining only the fundamental modes in the waveguides. Field continuity equation in terms of fundamental modes may be obtained from Eq. (20) and yield the single-mode Fresnel formulas:

$$r_{\text{SF}} = \frac{\sigma_{\text{S}}\sigma_{\text{G}} - \sigma_{\text{GS}}^2}{\sigma_{\text{S}}\sigma_{\text{G}} + \sigma_{\text{GS}}^2}, \quad (31)$$

$$t_{\text{GF}} = \frac{2\sigma_{\text{S}}\sigma_{\text{GS}}}{\sigma_{\text{S}}\sigma_{\text{G}} + \sigma_{\text{GS}}^2}, \quad (32)$$

$$R_{\text{SF}} = |r_{\text{SF}}|^2, \quad (33)$$

$$T_{\text{GF}} = |t_{\text{SF}}|^2 \frac{S_{\text{G}}(1)}{S_{\text{S}}(1)}, \quad (34)$$

where, r_{SF} (t_{GF}) is the surface (gap) plasmon amplitude reflection (transmission) coefficient obtained using the single-mode Fresnel approximation. For notational convenience, we have defined $\sigma_{\text{S(G)}} \equiv \Omega_{\text{S(G)}}(1)$ and $\sigma_{\text{GS}} \equiv C_{\text{GS}}(1, 1)$.

The transmission/reflection coefficients calculated using the single-mode Fresnel formulas Eqs. (31)–(34) are plotted in Fig. 8 using dashed lines. The single-mode approximation, still underestimates the exact mode-matching results for amplitude and energy reflection but improves the estimate in comparison to the plane-wave approximation. The agreement in the amplitude transmission is good for sub-5-nm gaps but deteriorates thereafter. Unlike the plane-wave Fresnel formulas, the single-mode Fresnel equations, *overestimate* the transmission coefficients. For the energy transmission, the agreement of the full mode-matching results is better with Eq. (30) than Eq. (34).

Both the plane-wave and the single-mode Fresnel approximations reproduce the monotonic decrease in amplitude transmission for $g \leq 10$ nm. This leads us to believe that the cut-off-like phenomenon is roughly related to the maximum of the transmission given by Eq. (28) as a function of gap size. Overall, the single-mode Fresnel equations provide a good first-order approximation to the reflection and transmission coefficients for quantum-scale gaps. The quality of the approximation improves for smaller gap sizes due to the higher order modes carrying a negligible amount of energy.

Parameter	Definition/Relation	Mode type
K	$k_0 \sqrt{\epsilon_i - \epsilon_m}$	All
γ_i	$-i\kappa_i$	$T1$ and $T2$
γ_m	$-i\kappa_m$	$T2$
κ_m	$\sqrt{K^2 + \kappa_i^2}$	Bound plasmon
κ_m	$\sqrt{K^2 - \gamma_i^2}$	$T1$
γ_i	$\sqrt{K^2 + \gamma_m^2}$	$T2$

TABLE IV. Equations and definitions relating the transverse wavevectors in the metal and the dielectric for the mode types indicated in the rightmost column.

VII. CONCLUSIONS

We have numerically demonstrated high-efficiency coupling into sub-10 nm quantum-scale waveguides using computationally efficient framework of plasmonic mode-matching theory. Examining a single-interface to MIM waveguide junction, we predict coupling efficiencies over 10% into < 5 nm gaps and reaching up to 24% for a 10 nm gap size, rivaling typical fiber-to-semiconductor waveguide coupling efficiencies.

These high coupling efficiencies are accompanied by significant enhancements in the electromagnetic fields transmitted into the gaps. Our analysis also demonstrates existence of an optimum gap size that maximizes the field enhancement. The optimum gap size depends on the material and geometrical parameters of the waveguide and, for the numerical values considered in our paper is 8 nm.

Our demonstration of high field enhancements and energy coupling into sub-10 nm waveguides make it possible to envision nanophotonic devices using light matter interaction at scales beyond those achievable in conventional dielectric and plasmonic devices. Applications of such quantum-scale waveguide devices may include direct rectification of optical radiation, engineered artificial optical non-linearities, and novel metamaterial phenomenon unobserved in natural materials.

ACKNOWLEDGEMENTS

We acknowledge the generous support and encouragement of Dr. F. B. McCormick and C. A. Boye. R. D. K thanks Dr. S. E. Kocabaş, A. Chandran, and Prof. M. L. Brongersma for helpful discussions. This work was supported by the Laboratory Directed Research and Development program at Sandia National Laboratories and Air Force Research Laboratory contract F4FBFM9194G001. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energys National Nuclear Security Administration under contract DE-AC04-94AL85000.

Appendix A: Transformation and solution of dispersion equations

In this section we specify the transformations of the dispersion equations (4) and (8) and provide expressions that enable us to solve them using the fixed-point iteration (FPI) method.³¹ In the FPI technique, we rewrite a given algebraic equation $F(z) = 0$ in an alternative form $z = f(z)$ where $f(z)$ is obtained by an algebraic manipulation of $F(z)$. Starting with an initial value z_0 , successive values of z are obtained according to $z_{n+1} = f(z_n)$. If $|f'(z)| < 1$ these values form a convergent sequence whose limit is the solution of the original equation. The transformations outlined below recast the dispersion equations into forms appropriate for an FPI implementation. Different forms are seen to be optimal for locating the roots belonging to a specific branch.

The twelve FPI forms (Eqs. (A1)–(A12)) given in the following are each mathematically equivalent to the corresponding parent dispersion equation (4) or (8) and are specified in terms of the transverse wavevectors κ_i or κ_m . Equations relating κ_i and κ_m for various root types are defined in Table IV. In our experience the FPI technique is relatively insensitive to initial guess and $\kappa_{i,0} = k_0$ can be used in most practical cases. The iteration is understood to be performed in terms of the variable appearing on the left hand sides of Equations (A1)–(A12). The quantities appearing on the right hand side are to be treated as functions of this left-hand-side variable and are to be evaluated using the relations defined in Table IV.

1. Single-interface dispersion equation

a. Bound plasmon mode

To obtain the propagation constant of the bound plasmon mode on the single-interface side ($z < 0$), we rewrite Eq. (4) as

$$\kappa_{i,n+1} = -p\kappa_{m,n} \tanh(\kappa_{m,n}h) \coth(\kappa_{i,n}a) \quad (\text{A1})$$

Equation (A1) is cast to be solved using FPI.

b. T1 modes

For $T1$ modes on the single-interface side, κ_i is imaginary (for lossless metals) and κ_m is real. We rewrite Eq. (4) accounting for this fact by substituting $\kappa_i \rightarrow i\gamma_i$:

$$\frac{\gamma_i}{\epsilon_i} \tan(\gamma_i a) = \frac{\kappa_m}{\epsilon_m} \tanh(\kappa_m h), \quad (\text{A2})$$

which can be transformed into iterative form as

$$\gamma_{i,n+1} = \frac{1}{a} \left\{ M\pi + \tan^{-1} \left[\frac{p\kappa_{m,n}}{\gamma_{i,n}} \tanh(\kappa_{m,n}h) \right] \right\} \quad (\text{A3})$$

Insensitivity of Eq. (A3) to initial guess implies that successive complex $T1$ roots for single-interface side may be simply found by using the initial guess $\gamma_{i,0} = k_0$ and inputting $M = 1, 2, 3, \dots$

c. T2 modes

For $T2$ modes both κ_i and κ_m are imaginary (for lossless metals) and are replaced as $(\kappa_i, \kappa_m) \rightarrow (i\gamma_i, i\gamma_m)$. With this substitution, the dispersion equation (4) assumes the form

$$\frac{\gamma_i}{\epsilon_i} \tan(\gamma_i a) = -\frac{\gamma_m}{\epsilon_m} \tan(\gamma_m h). \quad (\text{A4})$$

To determine the $T2$ roots by FPI, we use the half-angle identity $\tan(z/2) = -\cot z \pm \sqrt{1 + \cot^2 z}$ to express $\tan(\gamma_m h/2)$ in terms of the other quantities in Eq. (A4). We then invert the tangent function and obtain even and odd $T2$ roots by the following FPI formula:

$$\gamma_{m,n+1}^{e,o} = \frac{2}{h} \left[M\pi + \tan^{-1} \left(U_n \pm \sqrt{1 + U_n^2} \right) \right] \quad (\text{A5})$$

where $U_n = (p\gamma_{m,n}^{e,o}/\gamma_{i,n}^{e,o}) \cot(\gamma_{i,n}^{e,o}a)$ with superscripts ‘e’ and ‘o’ referring, respectively, to even (+) and odd (−) $T2$ modes. The mode index starts from $M = 0$ for even modes and $M = 1$ for odd modes.

2. MIM dispersion equation

a. Bound plasmon mode

The dispersion Eq. (8) for a truncated even MIM waveguide separates naturally into even and odd modes as follows

$$\tanh(\kappa_i g/2) = \begin{cases} -\frac{p\kappa_m}{\kappa_i} \tanh(\kappa_m h) & \text{Even modes} \\ -\frac{\kappa_i}{p\kappa_m} \coth(\kappa_m h) & \text{Odd modes} \end{cases} \quad (\text{A6})$$

leading to FPI forms

$$\kappa_{i,n+1} = \begin{cases} -p\kappa_{m,n} \tanh(\kappa_{m,n}h) \coth \frac{\kappa_{i,n}g}{2} & \text{Even modes} \\ -p\kappa_{m,n} \tanh(\kappa_{m,n}h) \tanh \frac{\kappa_{i,n}g}{2} & \text{Odd modes} \end{cases} \quad (\text{A7})$$

For the gap thicknesses considered in this paper, the MIM waveguide supports only the even mode.

b. T1 modes

For *T1* modes of the MIM waveguide, we account for the imaginary nature of κ_i by substituting $\kappa_i \rightarrow i\gamma_i$. We start from Eq. (A6) and craft the following FPI forms for even and odd *T1* modes by inverting the tangent function on the *left*-hand side. We have

$$\gamma_{i,n+1} = \frac{2}{g} \left\{ M\pi + \tan^{-1} \left[\frac{p\kappa_{m,n}}{\gamma_{i,n}} \tanh(\kappa_{m,n}h) \right] \right\} \quad (\text{A8})$$

for the even *T1* modes and

$$\gamma_{i,n+1} = \frac{2}{g} \left\{ M\pi - \tan^{-1} \left[\frac{\gamma_{i,n}}{p\kappa_{m,n}} \coth(\kappa_{m,n}h) \right] \right\} \quad (\text{A9})$$

for the odd *T1* modes of the MIM waveguide. As with the *T1* modes of the single-interface plasmon waveguide, the successive even and odd *T1* modes are calculated by iterating according to Eqs. (A8) and (A9). The mode index M assumes values $M = 0, 1, 2, \dots$ for odd modes and $M = 1, 2, 3, \dots$ for even modes.

c. T2 modes

Substituting $(\kappa_i, \kappa_m) \rightarrow (i\gamma_i, i\gamma_m)$ to reflect the dominantly imaginary nature of κ_i and κ_m for *T2* modes leads to the transformation of Eq. (A6) as

$$\tan(\gamma_i g/2) = \begin{cases} -\frac{p\gamma_m}{\gamma_i} \tan(\gamma_m h) & \text{Even modes} \\ \frac{\gamma_i}{p\gamma_m} \cot(\gamma_m h) & \text{Odd modes} \end{cases} \quad (\text{A10})$$

The *T2* roots of the MIM waveguide are found by inverting the (co)tangent functions on the *right*-hand side resulting in the FPI forms

$$\gamma_{m,n+1} = \frac{1}{h} \left\{ M\pi - \tan^{-1} \left[\frac{\gamma_{i,n}}{p\gamma_{m,n}} \tan \frac{\gamma_{i,n}g}{2} \right] \right\} \quad (\text{A11})$$

for the even *T2* modes and

$$\gamma_{m,n+1} = \frac{1}{h} \left\{ M\pi + \tan^{-1} \left[\frac{\gamma_{i,n}}{p\gamma_{m,n}} \cot \frac{\gamma_{i,n}g}{2} \right] \right\} \quad (\text{A12})$$

for the odd modes. The mode index M assumes integer values starting from $M = 0$ and $M = 1$ for even and odd modes respectively.

Appendix B: Expressions for normalization constants, overlap integrals and Poynting vectors

As mentioned at the end of section II C 3, numerical integration routines experience difficulties in calculating the normalization and overlap integrals for higher-order modes because of their rapidly oscillating field variations making it necessary to obtain closed-form expressions from analytical evaluation of integrals. In the following, we provide closed-form expressions for field normalization constants the single-interface and gap-plasmon sides and the overlap integrals between the two pairs of fields on both sides. We also provide expressions for the power carried by modes on each side (i.e., their Poynting vector) to aid in verifying the power conservation. The normalization constants and overlap integrals are used in section II C 4 to form the matrix equations for the scattering coefficients and Poynting vector integrals are used in Eqs. (22) and (23) to verify power conservation.

1. Normalization Coefficients

The normalization coefficients for the single-interface and MIM fields follow from the orthogonality conditions and are defined by Eqs. (9)–(11). The explicit expressions for $\Omega_S(n)$ and $\Omega_G(n)$ are obtained by straightforward integration

of elementary functions and we provide below only the final expressions. In the expressions below, \mathbb{R} and \mathbb{I} denote the real and the imaginary parts, respectively.

For the single-interface plasmon we have,

$$\Omega_S(n) = \langle E_{x,n}^S | H_{y,n}^S \rangle \equiv V_1 + V_2 \quad (\text{B1})$$

where

$$V_1 = \frac{\beta_n^S Z_0}{2\epsilon_m k_0} \frac{1}{\cosh^2(\kappa_{m,n}^S h)} \left(h + \frac{\sinh(2\kappa_{m,n}^S h)}{2\kappa_{m,n}^S} \right)$$

$$V_2 = \frac{\beta_n^S Z_0}{2\epsilon_i k_0} \frac{1}{\cosh^2(\kappa_{i,n}^S a)} \left(a + \frac{\sinh(2\kappa_{i,n}^S a)}{2\kappa_{i,n}^S} \right)$$

Similarly, for the gap plasmon the normalization integral is

$$\Omega_G(n) = \langle E_{x,n}^G | H_{y,n}^G \rangle \equiv W_1 + W_2 + W_3, \quad (\text{B2})$$

where

$$W_1 = \frac{\beta_n^G Z_0}{2\epsilon_m k_0} \frac{1}{\cosh^2(\kappa_{m,n}^G h)} \left(h + \frac{\sinh(2\kappa_{m,n}^G h)}{2\kappa_{m,n}^G} \right)$$

$$W_2 = \frac{\beta_n^G Z_0}{2\epsilon_g k_0} \left\{ \left(\frac{\sinh(2\kappa_{i,n}^G g)}{2\kappa_{i,n}^G} + g \right) + \eta_n^2 \left(\frac{\sinh(2\kappa_{i,n}^G g)}{2\kappa_{i,n}^G} - g \right) + 2\eta_n \left(\frac{\cosh(2\kappa_{i,n}^G g) - 1}{2\kappa_{i,n}^G} \right) \right\}$$

$$W_3 = \frac{\beta_n^G Z_0}{2\epsilon_m k_0} \frac{A_n^2}{\cosh^2(\kappa_{m,n}^G h)} \left(h + \frac{\sinh(2\kappa_{m,n}^G h)}{2\kappa_{m,n}^G} \right)$$

The quantities η_n and A_n are obtained from η and A defined in Table I by replacing $\kappa_m \rightarrow \kappa_{m,n}^G$ and $\kappa_i \rightarrow \kappa_{i,n}^G$.

2. Overlap Integrals

The overlap integrals between the two field pairs are defined by Eqs. (9), (12), and (13). For the first integral we have

$$C_{GS}(m, n) = \langle E_{x,m}^G | H_{y,n}^S \rangle \equiv U_1 + U_2 + U_3, \quad (\text{B3})$$

where

$$U_1 = \frac{\beta_m^G Z_0}{2\epsilon_m k_0} \frac{1}{\cosh(\kappa_{m,m}^G h) \cosh(\kappa_{m,n}^S h)} \times \left[\frac{\sinh[(\kappa_{m,m}^G + \kappa_{m,n}^S)h]}{\kappa_{m,m}^G + \kappa_{m,n}^S} + \frac{\sinh[(\kappa_{m,m}^G - \kappa_{m,n}^S)h]}{\kappa_{m,m}^G - \kappa_{m,n}^S} \right]$$

$$\begin{aligned}
U_2 = & \frac{\beta_m^G Z_0}{2\epsilon_g k_0} \frac{1}{\cosh(\kappa_{i,n}^S a)} \\
& \times \left[\frac{\sinh[(\kappa_{i,m}^G + \kappa_{i,n}^S)g - \kappa_{i,n}^S a] + \sinh(\kappa_{i,n}^S a)}{\kappa_{i,m}^G + \kappa_{i,n}^S} \right. \\
& + \frac{\sinh[(\kappa_{i,m}^G - \kappa_{i,n}^S)g + \kappa_{i,n}^S a] - \sinh(\kappa_{i,n}^S a)}{\kappa_{i,m}^G - \kappa_{i,n}^S} \\
& + \eta_m \frac{\cosh[(\kappa_{i,m}^G + \kappa_{i,n}^S)g - \kappa_{i,n}^S a] - \cosh(\kappa_{i,n}^S a)}{\kappa_{i,m}^G + \kappa_{i,n}^S} \\
& \left. + \eta_m \frac{\cosh[(\kappa_{i,m}^G - \kappa_{i,n}^S)g + \kappa_{i,n}^S a] - \cosh(\kappa_{i,n}^S a)}{\kappa_{i,m}^G - \kappa_{i,n}^S} \right]
\end{aligned}$$

$$\begin{aligned}
U_3 = & \frac{\beta_m^G Z_0}{2\epsilon_m k_0} \frac{A_m}{\cosh(\kappa_{m,m}^G h) \cosh(\kappa_{i,n}^S a)} \\
& \times \left[\frac{\sinh[(\kappa_{m,m}^G + \kappa_{i,n}^S)h]}{\kappa_{m,m}^G + \kappa_{i,n}^S} + \frac{\sinh[(\kappa_{m,m}^G - \kappa_{i,n}^S)h]}{\kappa_{m,m}^G - \kappa_{i,n}^S} \right]
\end{aligned}$$

For the second overlap integral we have,

$$D_{SG}(m, n) = \langle E_{x,m}^S | H_{y,n}^G \rangle \equiv T_1 + T_2 + T_3, \quad (\text{B4})$$

where

$$\begin{aligned}
T_1 = & \frac{\beta_m^S Z_0}{2\epsilon_m k_0} \frac{1}{\cosh(\kappa_{m,m}^S h) \cosh(\kappa_{m,n}^G h)} \\
& \times \left[\frac{\sinh[(\kappa_{m,m}^S + \kappa_{m,n}^G)h]}{\kappa_{m,m}^S + \kappa_{m,n}^G} + \frac{\sinh[(\kappa_{m,m}^S - \kappa_{m,n}^G)h]}{\kappa_{m,m}^S - \kappa_{m,n}^G} \right]
\end{aligned}$$

$$\begin{aligned}
T_2 = & \frac{\beta_m^S Z_0}{2\epsilon_i k_0} \frac{1}{\cosh(\kappa_{i,n}^S a)} \\
& \times \left[\frac{\sinh[(\kappa_{i,m}^S + \kappa_{i,n}^G)g - \kappa_{i,m}^S a] + \sinh(\kappa_{i,m}^S a)}{\kappa_{i,m}^S + \kappa_{i,n}^G} \right. \\
& + \frac{\sinh[(\kappa_{i,m}^S - \kappa_{i,n}^G)g - \kappa_{i,m}^S a] + \sinh(\kappa_{i,m}^S a)}{\kappa_{i,m}^S - \kappa_{i,n}^G} \\
& + \eta_n \frac{\cosh[(\kappa_{i,m}^S + \kappa_{i,n}^G)g - \kappa_{i,m}^S a] - \cosh(\kappa_{i,m}^S a)}{\kappa_{i,m}^S + \kappa_{i,n}^G} \\
& \left. - \eta_n \frac{\cosh[(\kappa_{i,m}^S - \kappa_{i,n}^G)g - \kappa_{i,m}^S a] - \cosh(\kappa_{i,m}^S a)}{\kappa_{i,m}^S - \kappa_{i,n}^G} \right]
\end{aligned}$$

$$\begin{aligned}
T_3 = & \frac{\beta_m^S Z_0}{2\epsilon_i k_0} \frac{A_n}{\cosh(\kappa_{i,m}^S a) \cosh(\kappa_{m,n}^G h)} \\
& \times \left[\frac{\sinh[(\kappa_{i,m}^S + \kappa_{m,n}^G)h]}{\kappa_{i,m}^S + \kappa_{m,n}^G} + \frac{\sinh[(\kappa_{i,m}^S - \kappa_{m,n}^G)h]}{\kappa_{i,m}^S - \kappa_{m,n}^G} \right]
\end{aligned}$$

3. Poynting Vectors

For the PEC truncated domain considered in this paper, the Poynting vector for the n^{th} mode can be defined in terms of the field quantities as

$$S^{\text{S,G}}(n) = \frac{1}{2} \Re \left[\int_{-h}^{h+g} E_{x,n}^{\text{S,G}} (H_{y,n}^{\text{S,G}})^* dx \right] \quad (\text{B5})$$

where \Re denotes the real part, $(\cdot)^*$ indicates complex conjugation, and the definition holds for either the single-interface (S) or the gap (G) modes. For the single-interface modes the Poynting vector is

$$\begin{aligned} S^{\text{S}}(n) &= \frac{Z_0}{2k_0} \left| \frac{1}{\cosh^2(\kappa_{\text{m},n}^{\text{S}} h)} \right| \\ &\quad \times \left(\frac{\sinh(2\kappa_{\text{mR},n}^{\text{S}} h)}{2\kappa_{\text{mR},n}^{\text{S}}} + \frac{\sin(2\kappa_{\text{mI},n}^{\text{S}} h)}{2\kappa_{\text{mI},n}^{\text{S}}} \right) \Re \left(\frac{\beta_n^{\text{S}}}{\epsilon_{\text{m}}} \right) \\ &\quad + \frac{Z_0}{2k_0} \left| \frac{1}{\cosh^2(\kappa_{\text{i},n}^{\text{S}} a)} \right| \\ &\quad \times \left(\frac{\sinh(2\kappa_{\text{iR},n}^{\text{S}} a)}{2\kappa_{\text{iR},n}^{\text{S}}} + \frac{\sin(2\kappa_{\text{iI},n}^{\text{S}} a)}{2\kappa_{\text{iI},n}^{\text{S}}} \right) \Re \left(\frac{\beta_n^{\text{S}}}{\epsilon_{\text{i}}} \right) \end{aligned} \quad (\text{B6})$$

$$\begin{aligned} S^{\text{G}}(n) &= \frac{Z_0}{2k_0} \left| \frac{1}{\cosh^2(\kappa_{\text{m},n}^{\text{G}} h)} \right| \left(\frac{\sinh(2\kappa_{\text{mR},n}^{\text{G}} h)}{2\kappa_{\text{mR},n}^{\text{G}}} + \frac{\sin(2\kappa_{\text{mI},n}^{\text{G}} h)}{2\kappa_{\text{mI},n}^{\text{G}}} \right) \Re \left(\frac{\beta_n^{\text{G}}}{\epsilon_{\text{m}}} \right) \\ &\quad + \frac{Z_0}{2k_0} \left\{ \frac{\sinh(2\kappa_{\text{iR},n}^{\text{G}} g)}{\kappa_{\text{iR},n}^{\text{G}}} (1 + |\eta_n|^2) + \frac{\sin(2\kappa_{\text{iI},n}^{\text{G}} g)}{\kappa_{\text{iI},n}^{\text{G}}} (1 - |\eta_n|^2) \right. \\ &\quad \left. + 2\Re \left[\eta_n \frac{\cosh(2\kappa_{\text{iR},n}^{\text{G}} g) - 1}{2\kappa_{\text{iR},n}^{\text{G}}} - i\eta_n \frac{\cos(2\kappa_{\text{iI},n}^{\text{G}} g) - 1}{2\kappa_{\text{iI},n}^{\text{G}}} \right] \right\} \Re \left(\frac{\beta_n^{\text{G}}}{\epsilon_{\text{g}}} \right) \\ &\quad + \frac{Z_0}{2k_0} \left| \frac{A_n^2}{\cosh^2(\kappa_{\text{m},n}^{\text{G}} h)} \right| \left(\frac{\sinh(2\kappa_{\text{mR},n}^{\text{G}} h)}{2\kappa_{\text{mR},n}^{\text{G}}} + \frac{\sin(2\kappa_{\text{mI},n}^{\text{G}} h)}{2\kappa_{\text{mI},n}^{\text{G}}} \right) \Re \left(\frac{\beta_n^{\text{G}}}{\epsilon_{\text{m}}} \right) \end{aligned} \quad (\text{B7})$$

Appendix C: Numerical verification of mode orthogonality and completeness

1. Mode orthogonality

Once the modal propagation constants and the field-profiles have been determined and the total number, N , of modes is fixed, it is straightforward to test the set for orthogonality. We simply calculate the overlap integrals $\langle E_{x,m}^{\text{S}} | H_{y,n}^{\text{S}} \rangle$ and $\langle E_{x,m}^{\text{G}} | H_{y,n}^{\text{G}} \rangle$ for $1 \leq m, n \leq N$ and verify that the integrals are numerically significant only for $m = n$ in which case they equal $\Omega_{\text{S}}(m)$ (for single-interface waveguide) or $\Omega_{\text{G}}(n)$ (for MIM waveguide).

Figure 10 pictorially shows the magnitude of the integrals. The bright diagonal represents high value of the integral for $m = n$. For $m \neq n$ the values of the overlap integrals are over 15 orders of magnitude lesser than those on the main diagonal, thereby providing a numerical confirmation of the mode orthogonality.

2. Single-interface waveguide mode completeness

We hypothesize that if the set of modes corresponding to roots depicted in Fig. 3 is complete, then any arbitrary field should be expressible as a linear combination of the set of these modes. In particular, the E_x and the H_y fields of the bound modes on either sides must be expressible as a linear combination of modes of the other side. The coefficients for this expansion are readily expressed in terms of the overlap integrals C_{GS} and D_{SG} already worked out in section II C 3.

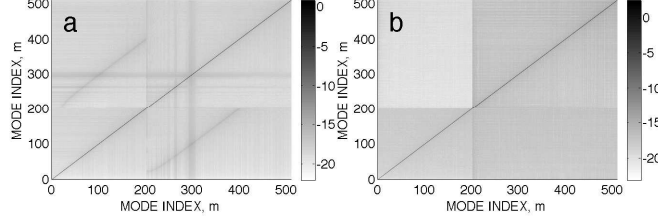


FIG. 10. Numerical verification of mode orthogonality. Logarithm (to base 10) of overlap integrals (a) $\langle E_{x,m}^S | H_{y,n}^S \rangle$ and (b) $\langle E_{x,m}^G | H_{y,n}^G \rangle$ for $1 \leq m, n \leq 510$.

First we try to express the E_x^G and H_y^G fields of the bound gap-plasmon ($n = 1$) mode of the MIM waveguide in terms of the fields E_x^S and H_y^S of the single-interface waveguide fields. We begin by expressing the gap plasmon fields in terms of the single-interface waveguide fields.

$$H_{y,1}^G = \sum_{m=1}^N a_m H_{y,m}^S \quad (C1)$$

$$E_{x,1}^G = \sum_{m=1}^N b_m E_{x,m}^S \quad (C2)$$

Premultiplication of Eq. (C1) with $E_{x,k}^S$ and postmultiplication of Eq. (C2) with $H_{x,k}^S$ followed by integration over all x leads to

$$a_k = \frac{D_{SG}(k, 1)}{\Omega_S(k)} \quad (C3)$$

$$b_k = \frac{C_{GS}(1, k)}{\Omega_S(k)} \quad (C4)$$

Figure 11(a) and (b) show the electric and the magnetic fields of the MIM gap plasmon reconstructed in terms of the single-interface waveguide modes. 200 $T1$ and 309 $T2$ modes were used in this reconstruction in addition to the bound plasmon mode. The reconstructed magnetic field is practically indistinguishable from the true field profile [Eq. (5) and (6)] on the scale of our graph. The electric field has a discontinuities at $x = 0$ and $x = g$ and show the familiar Gibbs phenomenon. We were able to obtain a better reconstruction of the electric field if we increased the number of modes to $N = 800$. This proved, however, to be computationally intensive. As Fig. 6 shows only about 50 modes are required to obtain convergence of the transmission/reflection coefficients. Nevertheless, improved convergence of the reconstructed fields to the true fields with increasing number of modes is a numerical confirmation of the completeness of the obtained modes of the single-interface waveguide. Using the same procedure, we were able to reconstruct any higher-order MIM mode as a linear combination of the single-interface waveguide mode.

3. MIM waveguide mode completeness

We can perform analogous operations to expand the fields of the single-interface plasmon in terms of the MIM waveguide modes. As before, we begin by expressing the single-interface plasmon fields in terms of those of the MIM modes:

$$H_{y,1}^S = \sum_{m=1}^N u_m H_{y,m}^G \quad (C5)$$

$$E_{x,1}^S = \sum_{m=1}^N v_m E_{x,m}^G \quad (C6)$$

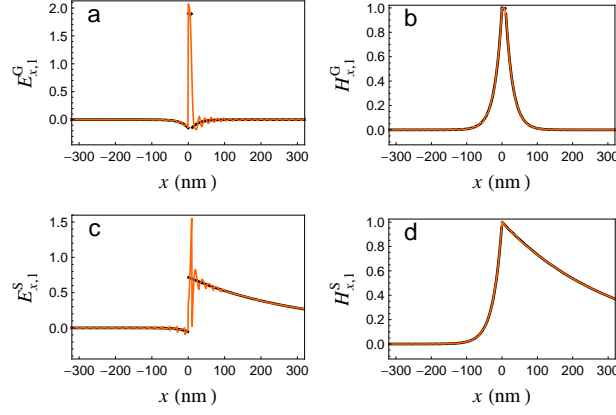


FIG. 11. Real part of the (a) Electric and (b) magnetic fields of the 10 nm gap plasmon reconstructed using 510 modes of the single-interface waveguide. The reconstruction included the bound single-interface plasmon, 200 $T1$ modes and 309 $T2$ modes. Real part of the (c) Electric and (d) magnetic fields of the single-interface plasmon reconstructed using 660 modes of the MIM waveguide. The reconstruction included the bound gap plasmon, 200 $T1$ modes and 459 $T2$ modes. The solid orange lines are the analytical fields plotted using Eq. (5) and (6). The reconstructed fields are plotted using solid black lines.

Premultiplication of Eq. (C5) with $E_{x,k}^G$ and postmultiplication of Eq. (C6) with $H_{x,k}^G$ followed by integration over all x leads to

$$u_k = \frac{C_{GS}(k, 1)}{\Omega_G(k)} \quad (C7)$$

$$v_k = \frac{D_{SG}(1, k)}{\Omega_G(k)} \quad (C8)$$

Reconstructed electric and magnetic fields of the single-interface plasmon are shown in Fig. 11(c) and (d) and show features similar to Fig. 11 (a) and (b). The reconstructed magnetic field is indistinguishable on the scale of the plot, the electric field displays noticeable Gibbs phenomenon. As with the single-interface case, the field reconstruction becomes better with increasing number of modes, N .

Subsections C 2 and C 3 provided a numerical verification of completeness of the set of the obtained modes on both sides of the junction. As seen in Fig. 3 the roots of the governing transcendental equations are spaced very closely that could, in some instances, cause the numerical root-finding routines to miss some of them. A missing root can be visualized directly in field reconstruction where its absence shows up as an oscillatory deviation from the true field profile. This deviation cannot be reduced by increasing the number of modes. We end this subsection by reemphasizing the importance of ensuring the numerical completeness of the modes as shown above, before proceeding to enforcing field continuity at the junction and calculation of scattering coefficients.

-
- ¹ J. A. Schuller, E. S. Barnard, W. Cai, Y. C. Jun, J. S. White, and M. L. Brongersma, *Nat Mater* **9**, 193 (2010).
 - ² H. A. Atwater and A. Polman, *Nat Mater* **9**, 205 (2010).
 - ³ D. K. Gramotnev and S. I. Bozhevolnyi, *Nat Photon* **4**, 83 (2010).
 - ⁴ W. L. Barnes, A. Dereux, and T. W. Ebbesen, *Nature* **424**, 824 (2003).
 - ⁵ C. Genet and T. W. Ebbesen, *Nature* **445**, 39 (2007).
 - ⁶ T. W. Ebbesen, H. J. Lezec, H. F. Ghaemi, T. Thio, and P. A. Wolff, *Nature* **391**, 667 (1998).
 - ⁷ L. Wang, S. M. Uppuluri, E. X. Jin, and X. Xu, *Nano Letters* **6**, 361 (2006).
 - ⁸ R. Gordon, D. Sinton, K. L. Kavanagh, and A. G. Brolo, *Accounts of Chemical Research* **41**, 1049 (2008).
 - ⁹ K. M. Dani, Z. Ku, P. C. Upadhyay, R. P. Prasankumar, S. R. J. Brueck, and A. J. Taylor, *Nano Letters* **9**, 3565 (2009).
 - ¹⁰ P. Neutens, P. Van Dorpe, I. De Vlaminck, L. Lagae, and G. Borghs, *Nature Photonics* **3**, 283 (2009).
 - ¹¹ J. Singh, *Physics of semiconductors and their heterostructures* (Mcgraw-Hill College, 1992).
 - ¹² J. J. Sakurai and J. J. Napolitano, *Modern Quantum Mechanics*, 2nd ed. (Addison-Wesley, 2010).
 - ¹³ D. R. Ward, F. Huser, F. Pauly, J. C. Cuevas, and D. Natelson, *Nat Nano* **5**, 732 (2010).
 - ¹⁴ Y. C. Jun, R. D. Kekatpure, J. S. White, and M. L. Brongersma, *Physical Review B (Condensed Matter and Materials Physics)* **78**, 153111 (2008).
 - ¹⁵ A. V. Akimov, A. Mukherjee, C. L. Yu, D. E. Chang, A. S. Zibrov, P. R. Hemmer, H. Park, and M. D. Lukin, *Nature* **450**, 402 (2007).
 - ¹⁶ R. J. Walters, R. V. A. van Loon, I. Brunets, J. Schmitz, and A. Polman, *Nat Mater* **9**, 21 (2010).
 - ¹⁷ E. Verhagen, M. Spasenović, A. Polman, and L. K. Kuipers, *Phys. Rev. Lett.* **102**, 203904 (2009).
 - ¹⁸ E. Verhagen, L. Kuipers, and A. Polman, *Nano Letters* **7**, 334 (2007).
 - ¹⁹ D. F. P. Pile, T. Ogawa, D. K. Gramotnev, T. Okamoto, M. Haraguchi, M. Fukui, and S. Matsuo, *Applied Physics Letters* **87**, 061106 (2005).
 - ²⁰ S. I. Bozhevolnyi, *Opt. Express* **14**, 9467 (2006).
 - ²¹ E. Verhagen, L. K. Kuipers, and A. Polman, *Nano Letters* **10**, 3665 (2010).
 - ²² R. F. Oulton, V. J. Sorger, D. A. Genov, D. F. P. Pile, and X. Zhang, *Nat Photon* **2**, 496 (2008).
 - ²³ R. F. Oulton, D. F. P. Pile, Y. Liu, and X. Zhang, *Phys. Rev. B* **76**, 035408 (2007).
 - ²⁴ G. I. Stegeman, A. A. Maradudin, T. P. Shen, and R. F. Wallis, *Phys. Rev. B* **29**, 6530 (1984).
 - ²⁵ T. P. Shen, R. F. Wallis, A. A. Maradudin, and G. I. Stegeman, *Appl. Opt.* **23**, 607 (1984).
 - ²⁶ S. E. Kocabas, G. Veronis, D. A. B. Miller, and S. Fan, *Phys. Rev. B* **79**, 035120 (2009).
 - ²⁷ V. V. Schevchenkov, *Continuous transitions in open waveguides* (Golem Press, 1973).
 - ²⁸ G. I. Stegeman, A. A. Maradudin, and T. S. Rahman, *Phys. Rev. B* **23**, 2576 (1981).
 - ²⁹ B. Sturman, E. Podivilov, and M. Gorkunov, *Phys. Rev. B* **76**, 125104 (2007).
 - ³⁰ P. B. Johnson and R. W. Christy, *Phys. Rev. B* **6**, 4370 (1972).
 - ³¹ R. D. Kekatpure, A. C. Hryciw, E. S. Barnard, and M. L. Brongersma, *Opt. Express* **17**, 24112 (2009).