In this paper, we applied the modal expansion method (MEM) to investigate the wave behaviors inside a step-modulated subwavelength metal slit. The physical mechanism of the surface plasmon polariton (SPP) transmission is investigated in detail for slit structures with either dielectric or geometric modulation. The applicability of the effective index method is discussed. Moreover, as a special case of the geometric modulation, the evanescent-wave assisted transmission is demonstrated in a thin-modulated slit. We emphasize that a complete set is necessary in order to expand the wave functions in these kinds of structures. All the calculated results by the MEM are well retrieved by the finite-difference time-domain calculation.

© 2011 Optical Society of America

OCIS codes: (240.6680) Surface plasmons; (260.3910) Metal optics; (230.7380) Waveguides, channeled; (290.5825) Scattering theory.

References and links
3. H. Raether, Surface Plasmons on Smooth and Rough Surfaces and on Gratings (Springer-Verlag, 1988).


29. The commercially available software developed by Rsoft Design Group http://www.rsoftdesign.com is used for the numerical simulations.


33. Since the slit is an infinitely long one, the wave source has to be located in Layer 1. This can be easily done in FDTD by setting the coordinates of the source to $y = H(0)$. A “slab mode” (a normal mode of a slab waveguide with characteristics matching the input waveguide, a feature provided by the RSOFT Fullwave software) is used to excite the SPP mode. However, for the amplitude of the source in FDTD, it should be adjusted to the same value of the normalized SPP mode in Layer 1 as pointed out later. The grid size of the simulation is set as 2.5 nm $\times$ 2.5 nm.


1. Introduction

Surface plasmon polaritons (SPPs) are waves produced by the interaction between the surface charge oscillation and the electromagnetic field, which are able to propagate along a metal-dielectric interface while exponentially decay in both sides [1-3]. In a metal slit structure with the slit width being less than wavelength, there can be more than one mode of the electromagnetic wave propagating along the slit [3-9]. The wavelength of the lowest mode, called as in-slit SPP [9], is less than that in a single metal-dielectric interface, but will approach the latter as the slit width goes to infinity. In this paper we mainly discuss the in-slit SPP and simply call it as SPP. The particular properties of SPPs in subwavelength metallic structures have shown the potential to overcome the diffraction limit in conventional optics, and can be utilized to achieve nanoscaled photonic devices of high integration. Thus, plasmonic waveguides, which have the ability to support and channel SPPs, have attracted much attention. Various waveguides consisted of subwavelength metallic structures have been numerically investigated and/or experimentally fabricated [3-22]. Among them, slit structures, or metal-dielectric-metal waveguides, are believed to possess remarkable advantages, such as strong field localization, simplicity and convenience for fabrication and integration into optical circuits. With appropriate modulations of the slit, a variety of functional plasmonic structures can be designed and fabricated, such as splitters [11], Y-shaped combiners [12], couplers [13], filter [14], and photonic bandgap structures [15-18].

A step modulation [23], either dielectric or geometric, is one of the key elements in photonic engineering that is employed in various photonic devices in order to reduce their size. This kind of modulation allows one to control the waves within the structures. Besides, the step modulation is of important theoretical significance since it is helpful for investigating SPP scattering. With the staircase approximation and transfer matrix technique [24], numerical results from complicated structures can be obtained.

Up to now, a great deal of efforts has been made to calculate the transmission and to model the SPP scattering mechanism in a step-modulated slit [15-22]. It is generally believed that the transmission can be described by a Fabry-Perot-like formula because only the SPP mode survives in a subwavelength slit [15-21]. After taking into account the phase shift caused by the scattering at step-modulated regions, researchers have provided better descriptions about the SPP behavior [19-21]. However, in study of the SPP transmission, the metal has been considered as either a perfect-conducting one [25-28] inside which the electromagnetic fields must be zero, or a real one with the dielectric constants having finite real and imaginary parts [15-22]. Various approximations have been applied such as one-mode approximation (OMA) [15-21], quasistatic approximation [20], and lossless approximation [17, 20]. Recently, Kocabas et al. discussed the eigen problems of a metal slit structure and find a more general and more precise way to investigate the transmission [22].

In this paper, we develop a technique based on model expansion method (MEM), similar to Ref. [22], to calculate the waves inside a step-modulated subwavelength slit constructed by real metal and to investigate the SPP scattering and transmission mechanism. In order to carry out precise calculation, the slit structure has to be confined between two perfect-conducting walls at finite distances, which enables us to find a complete set of orthogonal functions to expand the field components inside the slit. The introduction of the walls will have influence on the numerical results. Of course, when the walls are pushed away from the slit, the influence attenuates. Thus the influence brought by the walls is numerically controllable. When the walls are far enough the calculated results will be very close to those of an unconfined slit structure. As a test all the calculated results are well retrieved by the simulated results from the corresponding unconfined slit by the finite-difference time-domain (FDTD) method [29]. We stress that the introduction of the perfect-conducting walls is necessary to investigate this structure,
otherwise the eigenfunctions would not compose a complete set and the error yielded cannot be ignored.

Two kinds of step modulation, dielectric and geometric modulations, are systematically studied. The physical mechanics of SPP scattering and transmission in both cases are revealed. The positions of the Fabry-Perot-like transmission peaks can be well predicted by taking into account the phase shift of the SPP mode when reflecting in the slit. Moreover, for a thin-modulated case, the multi-reflection of evanescent waves also needs to be considered in the transmission process. Besides, the effects of the two kinds of modulations are compared to each other, and the applicability of the OMA and effective index method (EIM) is discussed.

The paper is arranged as follows: Section 2 sets our model of a subwavelength metal slit structure and presents the MEM formalism. Section 3 analyzes the SPP scattering and transmission mechanism for the dielectric and geometric step modulations, respectively. Finally, conclusions are presented in Section 4.

2. The modal expansion method for metal slit structures

In this section, the model of our single slit structure and the formulas of the MEM are presented. The periodic version of this method can be seen in references [24, 30-32].

Fig. 1. Sketch of a step-modulated metal slit structure confined in x direction with perfect-conducting walls at 

![Fig. 1. Sketch of a step-modulated metal slit structure confined in x direction with perfect-conducting walls at L_l = 0, L_r = 2 μm. The gray areas are silver with dielectric constant ε_Ag = -50.76 + 0.083i; The slit areas marked by slashes are filled with core materials. A TM wave is normally launched at y = Q(0) = -1 μm with wavelength λ_0 = 1 μm. Q(1) = 0.](image)

The model studied in this paper is sketched in Fig. 1. It is infinitely large in yz plane but confined in x direction by perfect-conducting walls at x = L_l and L_r. The structure is divided into three layers along y direction and three regions along x direction. The gray regions in Fig. 1 are silver. The slit areas marked by slashes are filled with some linear, nonmagnetic, and isotropic dielectrics, hereafter called as core materials. The core materials and slit widths can be different from one layer to another, which provides flexibility to design different step modulations. Here we design two kinds of step modulations: one is called as dielectric modulation where the slit widths are the same in all the layers but the core materials are not; another is called as geometric modulation where the core materials are the same in all the layers while the slit widths are not. A transverse magnetic (TM) wave, with magnetic field H being in z direction, is normally launched at y = Q(0) in Layer 1 and propagates upwards [33]. The geometric parameters are given in the caption of Fig. 1. For clarity, the denotations of wave functions and wave vectors
The dielectric constant of silver as a function of the wavelength of incident wave \( \lambda_0 \) is evaluated as \( \varepsilon_{\text{Ag}} = (3.57 - 54.33\lambda_0^2) + i(-0.083\lambda_0 + 0.921\lambda_0^3) \) by fitting the experimental data [34]. In this paper the wavelength is set as \( \lambda_0 = 1 \mu m \), thus, \( \varepsilon_{\text{Ag}} = -50.76 + 0.083i \).

The central region in the second layer is named as scattering region. We investigate in this paper mainly the transmission of SPP through the scattering region.

We apply MEM in this paper. This method is usually used in periodic structure [24, 30-32] and has many derivative versions, such as Fourier modal method [24, 35, 36], RCWA [35] and C method [36]. Nevertheless, no matter the structure is periodic or not, these versions have the same mathematical origin. Simply speaking, the substance of MEM is to expand unknown functions (electromagnetic field distribution in this paper) in terms of a complete set of orthogonal functions. Thus, seeking for an appropriate complete set of orthogonal functions is the starting point of the whole procedure.

By separating variables, we are able to consider the eigenfunctions along the \( x \) direction. The eigenfunctions and their derivatives should be continuous at the boundaries of the metal, and must be zero at the walls. Under these boundary conditions, the eigenvalue equation can be solved. In the \( l \)th layer of Fig. 1, the eigenfunctions along the \( x \) direction are expressed as

\[
\phi_n^{(l)}(x) = \begin{cases} 
A_n^{(l)} \sinh \left( k_{\text{s}1n}^{(l)}(x - L_l) \right), & L_l \leq x < x_1^{(l)} \\
B_n^{(l)} e^{ik_{\text{s}2n}^{(l)}x} + C_n^{(l)} e^{-ik_{\text{s}2n}^{(l)}x}, & x_1^{(l)} \leq x < x_2^{(l)} \\
D_n^{(l)} \sinh \left( k_{\text{s}3n}^{(l)}(x - L_r) \right), & x_2^{(l)} \leq x \leq L_r,
\end{cases}
\]

(1)

where

\[
\left( k_{\text{s}2n}^{(l)} \right)^2 + \left( k_{\text{s}yn}^{(l)} \right)^2 = \left( \frac{\omega}{c} \right)^2 \varepsilon_{\text{s}2}^{(l)}; \quad \left( k_{\text{s}yn}^{(l)} \right)^2 + \left( \frac{\omega}{c} \right)^2 \varepsilon_{\text{s}y}^{(l)} = \left( k_{\text{s}yn}^{(l)} \right)^2, \quad j = 1, 3.
\]

With the boundary conditions at \( x = x_1^{(l)} \) and \( x_2^{(l)} \), we achieve the equation that determines the eigenvalues:

\[
\begin{bmatrix}
\left( \gamma_1^{(l)} \right)^2 & \left( \gamma_2^{(l)} \right)^2 & 1 \\
2 k_{\text{s}2n}^{(l)} & 2 k_{\text{s}2n}^{(l)} & 1 \\
\left( \gamma_3^{(l)} \right)^2 & \left( \gamma_3^{(l)} \right)^2 & 1
\end{bmatrix}
\begin{bmatrix}
1 \left( \gamma_1^{(l)} \right)^2 & 1 \left( \gamma_2^{(l)} \right)^2 & 1 \\
1 \left( \gamma_2^{(l)} \right)^2 & 1 \left( \gamma_3^{(l)} \right)^2 & 1 \\
1 \left( \gamma_3^{(l)} \right)^2 & 1 \left( \gamma_3^{(l)} \right)^2 & 1
\end{bmatrix}
\]

(2)

where

\[
\gamma_1^{(l)} = \frac{1}{\left( k_{\text{s}2n}^{(l)} \right)^2}; \quad \gamma_2^{(l)} = \frac{1}{\left( k_{\text{s}2n}^{(l)} \right)^2}; \quad \gamma_3^{(l)} = \frac{1}{\left( k_{\text{s}2n}^{(l)} \right)^2}.
\]

The coefficients in Eq. (1) can be determined by the orthonormalization condition:

\[
\int_{L_r}^{L_l} \frac{1}{\varepsilon^{(l)}} \phi_{m}^{(l)+}(x) \phi_{n}^{(l)}(x) dx = \delta_{mn},
\]

(3)

where \( \phi \) means the complex conjugate of \( \phi \) and \( \delta_{mn} \) is the Kronecker delta. The equation satisfied by \( \phi^+ \) is the adjoint of that by \( \phi \). The proof of Eq. (3) is straightforward and similar to that given in Ref. [30].

The functions presented by Eq. (1) provide the required complete and orthogonal eigenfunctions of the given structure [37, 22]. Now the magnetic fields and its derivative are expressed
as

\[ H_z(x,y) = \begin{cases} 
\sum_{n=1}^{\infty} \phi_n^{(1)}(x) \left( I_n e^{ik_n^{(1)} y} + R_n e^{-ik_n^{(1)} y} \right), & Q^{(0)} \leq y < Q^{(1)} \\
\sum_{n=1}^{\infty} \phi_n^{(2)}(x) \left( E_n e^{ik_n^{(2)} y} + F_n e^{-ik_n^{(2)} y} \right), & Q^{(1)} \leq y < Q^{(2)} \\
\sum_{n=1}^{\infty} \phi_n^{(3)}(x) T_n e^{ik_n^{(3)} y}, & Q^{(2)} \leq y < \infty,
\end{cases} \]

where \( I_n, R_n, \) and \( T_n \) are the incidence, reflection and transmission coefficients for the \( n \)th mode, respectively. All the coefficients \( I_n, R_n, T_n, E_n, \) and \( F_n \) appearing in Eq. (4) are complex numbers and each is expressed by its amplitude and phase, say, \( T_n = |T_n| e^{i\theta} \). We denote the phase by \( \theta = \arg(T_n) \). The amplitudes and arguments represent the excitation efficiencies and phase shifts caused by the scattering. Thus, both the amplitudes and arguments are closely related to the scattering mechanism and will be discussed in the next section.

The boundary conditions along the \( y \) direction in Fig. 1 are that the magnetic field and its derivative at \( y = Q^{(1)} \) and \( Q^{(2)} \) must be continuous. Applying these conditions, we obtain the coupled-wave equations as follows:

\[
\begin{align*}
&I_n e^{ik_n^{(1)} y^{(1)}} + R_m = \sum_{n=1}^{\infty} \int_{L_1} \frac{1}{\epsilon^{(1)}} \phi_n^{(1)}(x) \left[ I_n e^{ik_n^{(1)} y^{(0)}} - R_n e^{-ik_n^{(1)} y^{(1)}} \right] \frac{1}{\epsilon^{(1)}} \phi_m^{(1)+}(x) \phi_n^{(1)}(x) \, dx \\
&\sum_{n=1}^{\infty} \int_{L_1} \frac{1}{\epsilon^{(1)}} \phi_n^{(1)}(x) \left[ I_n e^{ik_n^{(1)} y^{(1)}} - R_n e^{-ik_n^{(1)} y^{(1)}} \right] \frac{1}{\epsilon^{(1)}} \phi_m^{(1)+}(x) \phi_n^{(1)}(x) \, dx = \sum_{n=1}^{\infty} \int_{L_1} \frac{1}{\epsilon^{(1)}} \phi_n^{(1)}(x) \left[ I_n e^{ik_n^{(1)} y^{(1)}} - R_n e^{-ik_n^{(1)} y^{(1)}} \right] \frac{1}{\epsilon^{(1)}} \phi_m^{(1)+}(x) \phi_n^{(1)}(x) \, dx,
\end{align*}
\]

where \( q^{(1)} = Q^{(1)} - Q^{(0)} \) and \( q^{(2)} = Q^{(2)} - Q^{(1)} \). The coefficients \( R_n, T_n, E_n, \) and \( F_n \) can be numerically calculated by means of Eq. (6) and then are used to compute all the physical quantities.

Before going to present our numerical results, we would like to make some discussion about the eigenvalues solved from Eq. (2) and associated eigenfunctions. First, if the confinement along the \( x \) direction is removed, i.e., \( (x_1^{(f)} - L_r) \to \infty \) and \( (x_2^{(f)} - L_r) \to -\infty \), then \( \tanh(\pm \infty) \to \pm 1 \) and Eq.(2) becomes

\[
\begin{pmatrix}
\gamma_1^{(f)} + 1 \\
\gamma_2^{(f)} - 1
\end{pmatrix}
\begin{pmatrix}
\gamma_1^{(f)} - 1 \\
\gamma_2^{(f)} + 1
\end{pmatrix}
\begin{pmatrix}
e^{2i\phi_3 (x_2^{(f)} - x_1^{(f)})} & \end{pmatrix}
\begin{pmatrix}
1
\end{pmatrix}
\]

which is the frequently used formula to solve the wave vectors of SPPs in a slit [1-9]. Unfortunately, the eigenfunctions corresponding to the wave vectors solved by Eq. (7) are not complete because in the limiting process the wave vectors \( k_{1n}^{(f)} \) and \( k_{2n}^{(f)} \) with positive real part are left and those with negative real part are lost. The exponentially growing terms in the eigenfunctions have to be discarded when the region is infinitely extended, which leads to the loss of the completeness in mathematics. Thus, although Eq. (7) can describe the SPPs in an unconfined slit.
that they dominate the transmission process. Since the guided modes is that they have important contributions to the scattering of the field so decays in metallic region, hereafter called as a "guided mode". The physical meaning of $|\psi_n|\rangle$ has the maximum value at the metal/dielectric interface and decay in both regions. When $k_{x1n}$ has a characteristic that $|\text{Im}[k_{x1n}]/\text{Re}[k_{x1n}]|^2$ is rather small. The associated wave function decays in metallic region, hereafter called as a "guided mode". The physical meaning of the guided modes is that they have important contributions to the scattering of the field so that they dominate the transmission process. Since $|k_{x2n}|$ has an upper limit, the number of the guided mode is finite. The lowest guided mode is just the so-called SPP mode, which has the maximum value at the metal/dielectric interface and decay in both regions. When $|k_{x2n}| < (\omega c)^2 |\varepsilon_2^{(l)} - \varepsilon_1^{(l)}|$, $k_{x1n}$ has a characteristic that $|\text{Im}[k_{x1n}]/\text{Re}[k_{x1n}]|^2$ is large. The eigenfunction associated with such a wave vector oscillates in metallic region, called as an "radiation mode". Because the radiation modes attenuate quickly in the $y$ direction, they influence the scattering but have little contribution to the transmission. Since $|k_{x2n}|$ has no upper limit, the number of the radiation mode is infinite.

To show the above points, we plot the eigenvalues of $k_{x1n}$ and the associated wave functions in

![Figure 2](image_url)
Fig. 2. The lowest six modes are illustrated. Figures 2(a) and (b) show how the slit width affects the distribution of the $k_{s2n}$. For the SPP mode, the real part of $k_{s21}$ is very small and insensitive to the slit width, while the imaginary part, compared with higher modes, is a relatively large, which indicates the evanescent nature of the SPP mode whose field not only decays in metal but also in dielectric. For all the higher modes, their $k_{s2n}$ have a relatively large real part and small imaginary part, reflecting that the fields have oscillating behavior inside the slit. Some higher modes exhibit platform in both Figs. 2(a) and (b). If the $k_{s2n}$ value is in the platform region, it is a radiation mode. Otherwise, it is a guided mode. For example, when a slit width is 0.25 µm, among the lowest six modes, the first four are guided modes and the next two are radiation modes, as shown in Fig. 2(c).

Third, to find the eigenvalues in the present structure is rather difficult because their number is infinite and they may be dense in a small interval. Fortunately, Botten et al. provided a powerful program, based on Cauchy principle, to solve this problem [38]. With minor modifications in eigenequation and division procedure, this program is well applicable in the present case. Two numerical evaluations are implemented to testify if the eigenfunctions are orthogonal and complete. One is to use the orthonormalization condition, Eq.(3), which can rule out the repetitive solutions; and the other is numerical completeness test proposed by Botten and McPhedran, see Eq. (50) in Ref. [32], which can check if there are missing solutions. With these procedures, a set of trustworthy solutions can be obtained. By the way, it is numerically tested that the eigenfunctions corresponding to the wave vectors solved from Eq. (7) are not complete.

Our aim in this paper is to calculate the field inside an unconfined slit. However, since the walls are in a finite distance from the slit, an approximation in the calculated results is inevitably encountered. So, one question needs to be answer: how does this approximation affect the numerical results? Intuitively, the wider the confined region, the more precise results we get. However, because of the oscillating nature of the radiation modes, a wider region will contain a denser distribution of eigenvalues, which will cause problems in finding them. Fortunately, the intrinsic properties of the metal help us to get rid of this dilemma. The guided modes are hardly affected by the walls because they are mainly determined by the slit width, unless the confining walls are very close to the slit edges. This is why Cao et al. could characterize the slit by two thin metal layers on its both sides [28]. As for the radiation modes, although the associated waves in the metallic region are influence by the walls, they make little contribution to the transmission since their components inside the slit are very small and attenuate quickly in the propagation direction. As long as the walls are far enough, the calculated results will be very close to those of an unconfined structure. Thus, the results calculated from the confined slit by MEM can be to arbitrary precision as one wants. Our experience shows that when $L_f - L_l = 2$ µm, the convergence is less than 1% for all the calculations in this paper.

3. Numerical results and analysis

In this section, we investigate the scattering and transmission mechanism inside a step-modulated subwavelength metal slit. The incident wave is launched from Layer 1. The transmission wave in Layer 3 and the reflection wave in Layer 1 are calculated, respectively. Di-electric and geometric modulations are considered here. For both cases, total 101 eigenmodes, including even and odd modes, are used in calculation so as to guarantee that the convergence is less than 1%. Moreover, since the structures considered in this paper (including source) are symmetric with respect to the central line of the slit, all the calculated coefficients of the odd modes are negligibly small and we will ignore them in the following discussion. Besides, the slit width is restricted to allow only the lowest modes, namely, the SPP modes, to propagate in the slit [9], which is the common situation for using the OMA [15-21]. Since the incident wave only contains the lowest mode, we have $I_m = \delta_{m1}$ in Eq. (6).
The criterion for using the OMA is also discussed in this paper. Before showing the numerical proof, we would like to briefly compare the MEM and OMA methods. In scattering process, the field in each layer should be expanded by the complete set of eigenfunctions to one SPP mode; moreover, the excitation efficiency of each mode are determined by the coupled-wave equation, which cannot be precisely calculate by only using the SPP modes or EIM. When the scattering region is short, i.e., a thin-modulated case, the higher modes will transfer energy in despite of their evanescence. In this case, more than one mode participate in transmission, which causes the numerical imprecision of OMA. As for MEM, all the problems mentioned above have been handled by use of a complete set of eigenfunctions to expand the field. Generally speaking, the OMA can be considered as a special case of the MEM.

For example, speaking, the OMA can be considered as a special case of the MEM.

In the dielectric-modulated case, the three layers have the same slit width $w = 0.2 \mu m$ but different core materials. This kind of structure is briefly labeled as $[\epsilon_2^{(1)} - \epsilon_2^{(2)} - \epsilon_2^{(3)}]$. For example, $[1 - 6 - 1]$ means the structure shown in Fig. 3(a), where in the slit the layers 1 and 3 are vacuum and layer 2 is filled with a substance of $\epsilon_2^{(2)} = 6$.

Since the coupled-wave equations Eq. (6) contain all the physical information of the SPP transmission in the given structure, all the work left to do is to decouple these modes and to figure out what roles they play in this process. Figure 3(a) shows the calculated steady field distribution, the $H_z$ amplitude distribution, for $[1 - 6 - 1]$. Along the central axis of the slit, a good fit is obtained between the MEM and FDTD. In the slit, as pointed out already, only the SPP mode propagates and all the other modes attenuate. In Layer 3, the steady distribution is a steady field distribution of $H_z$ at the upper boundary of Layer 1, $y = Q^{(1)}(0)$; (c) The same $H_z$ at the lower boundary of Layer 2, $y = Q^{(1)}(+0)$.

### 3.1. Dielectric modulation

In the dielectric-modulated case, the three layers have the same slit width $w = (l_2^{(1)} - l_1^{(1)}) = 0.2 \mu m$ but different core materials. This kind of structure is briefly labeled as $\epsilon_2^{(1)} - \epsilon_2^{(2)} - \epsilon_2^{(3)}$.

For example, $[1 - 6 - 1]$ means the structure shown in Fig. 3(a), where in the slit the layers 1 and 3 are vacuum and layer 2 is filled with a substance of $\epsilon_2^{(2)} = 6$.

Since the coupled-wave equations Eq. (6) contain all the physical information of the SPP transmission in the given structure, all the work left to do is to decouple these modes and to figure out what roles they play in this process. Figure 3(a) shows the calculated steady field distribution, the $H_z$ amplitude distribution, for $[1 - 6 - 1]$. Along the central axis of the slit, a good fit is obtained between the MEM and FDTD. In the slit, as pointed out already, only the SPP mode propagates and all the other modes attenuate. In Layer 3, the steady distribution is a...
straight line because only the upward propagating SPP mode survives. In Layers 1 and 2, due to the interference of the upward and downward propagating SPP modes, the oscillating period of the steady field distribution is half of the wavelength of corresponding SPP modes; for instance, in Layer 1, the oscillation period is 0.450 μm, and the corresponding SPP wavelength is 0.900 μm.

Although only the SPP wave goes from Layer 1 to Layer 3, the transmission mechanics is closely related to the modes scattering at the layer boundaries. To show this point, we plot in Figs. 3(b) and (c) the total field and its first 10 eigenmodes at two sides of \( y = Q^{(1)} \) subject to the field continuity condition. First, half of the these 10 modes are odd modes and their coefficients are negligibly small since no fields is observed to pass through the central point \((1,0)\) at the cross section. Second, it is seen that the SPP mode is dominate in both layers and all other even modes merely give comparatively very small contributions to the total field. The discrepancy of the curves of SPP mode in Figs. 3(b) and (c) is small, which means that the scattering at the boundary between Layers 1 and 2 is weak, hereafter called as weak scattering. This is originated from that the eigenfunctions are mainly dependent on the slit width. For all the modes, the peak positions at the two sides of the layer boundary are nearly the same, as shown in Figs. 3(b) and (c). In the slit regions the contributions from the radiation modes are negligible. Because the slit widths in Layer 1 and 2 are the same, the transverse components of the SPP wave vector in these two layers are nearly the same, which is called quasi-transversal-phase match. This match is the physical reason of the weak scattering and suppression of the radiation modes.

![Graphs showing SPP transmission and reflection coefficients](image)

Fig. 4. Comparison between the SPP transmission and reflection coefficients in a two-layer structure by MEM (lines) and by EIM-based Fresnel equation (dots). The structure is constructed by setting \( w^{(l)} = 0.2 \) μm and \( Q^{(0)} = Q^{(1)} = Q^{(2)} = 0 \). For waves incident from air to dielectric, \( \varepsilon_1^{(1)} = 1, \varepsilon_2^{(1)} = \varepsilon_2^{(3)} \), (a) amplitude and (b) phase. For waves incident from dielectric to air, \( \varepsilon_2^{(1)} = \varepsilon_2^{(2)}, \varepsilon_2^{(3)} = 1 \), (c) amplitude and (d) phase.

Actually, the weak scattering inside the slit is very similar to the case when light incidents on a semi-infinite dielectric plane. To illustrate this point, we specially study the scattering mechanism of a dielectric-modulated two-layer structure, as shown in Fig. 4. Both cases of the...
wave incident from the dielectric region, where the effective index is $n_{\text{eff}}^{(l)} = k_{\text{eff}}^{(l)} / (2\pi/\lambda_0)$, to air and its opposite are examined. Please note that the Stokes relations are inappropriate for this multimode scattering. The SPP transmission and reflection coefficients are calculated by MEM and compared with the ones calculated by the EIM-based Fresnel equation. It can be seen from Fig. 4 that these two kinds of curves are close to each other both in amplitude and phase; only small differences in height are observed as the dielectric constant becomes larger. Besides, due to the fact that the results obtained by the EIM satisfy the Stocks relations which describe the one-mode scattering, the high resemblance of the curves in Fig. 4 also indicates that the OMA is applicable in the dielectric-modulated slit under weak scattering condition. Another example of the coincidence between EIM and FDTD is that in investigating the band structure of a one-dimensional photonic crystal confined in a subwavelength slit, the results calculated by the two methods were almost identical [15].

Fig. 5. Transmission vs. $q^{(2)}$ calculated by MEM for $[1-3-1]$ (solid line), $[1-6-1]$ (dashed line) and $[3-1-6]$ (dash-dotted line). The results calculated by FDTD (open circles) and EIM-based triple-layer Fresnel equation (dots) are also shown for comparison.

To better illustrate the transmission process, we calculate by the MEM the SPP transmission coefficient as a function of the length of Layer 2, $q^{(2)}$, for three structures $[1-3-1]$, $[1-6-1]$, and $[3-1-6]$. The results are plotted in Fig. 5. For comparison, we use the FDTD method and EIM-based triple-layer Fresnel equation to compute the same quantity, and the results are also plotted in Fig. 5. These three kinds of curves fit quite well, even for a very small $q^{(2)}$, namely, the thin modulation. Note that the EIM-based Fresnel equation only takes into account the non-phase-shift SPP modes. It is known that only the SPP mode propagates and multi-reflects in Layer 2. The phase shift of SPP reflection coefficients is too small (note that the y-axis in Figs. 4 (b) and (d) is less than $1^\circ$) to cause an observable variation in the position of the transmission peaks in Fig. 5. These facts manifest the weak scattering. We conclude that the OMA is applicable in describing the transmission behavior in the dielectric-modulated case. However, neglecting the small contribution of the higher modes will cause the small difference in the height of the peaks between the results of EIM and MEM.

3.2. Geometric modulation

In the geometric-modulated case, the slit regions in the three layers are filled with the same core materials $\varepsilon_2^{(l)} = 1$ but have different slit width. The geometric-modulated slit structures is labeled by $[w^{(1)} \sim w^{(2)} \sim w^{(3)}]$. For example, $[0.2 \sim 0.4 \sim 0.2]$ represents the structure shown in Fig. 6.
When light reaches a metal, its alternatively varying electric field causes the motion of electrons in metal. In turn, this motion emits the magnetic field to the space outside the metal. Therefore, the magnetic field in the corner area is necessarily stronger than that near plane surface. The electron motion is thus concentrated in the corner area, which is one quadrant of the space near the metal surface. The electron motion quickly in metal, and thus the radiation modes have contributions to the expansion (the modal expansion in the different layers are quite different, as shown in Figs. 6(b) and (c)). At the upper side of this boundary where the slit is wider, the total field mainly comes from the contribution of guided modes; while at the lower side where the slit is narrower, the guided modes cannot satisfy the field continuity condition because they evanesce quickly in metal, and thus the radiation modes have contributions to the expansion (the ordinal of each plotted mode in Fig. 6 is indicated by the number of its peaks). Therefore, the magnetic field in the corner area is necessarily stronger than that near plane surface. Therefore, the “bright corner” phenomenon is quite common.

Here we would like to have a few words about the essential origin of the brightness at corners. When light reaches a metal, its alternatively varying electric field causes the motion of electrons near the metal surface. The electron motion in turn emits the magnetic field to the space outside of the metal. In the corner area, the space is one quadrant, less than the space near plane surface. Therefore, the magnetic field in the corner area is necessarily stronger than that near plane surface. Therefore, the “bright corner” phenomenon is quite common.

To see how the SPP scatters at the layer boundary, we turn to two-layer structures sketched in Fig. 7. The amplitudes and phases of the SPP as functions of slit width calculated by the MEM (line) and the FDTD method (dots). (b) Steady field distribution of $H_z$ and its first 10 modes at the upper boundary of Layer 1, $y \rightarrow Q^{(1)}(\pm 0)$; (c) The same $H_z$ at the lower boundary of Layer 2, $y \rightarrow Q^{(1)}(\pm 0)$.

Figure 6(a) shows the steady field distribution in $[0.2 \sim 0.4 \sim 0.2]$ structure. A good fit between the calculated results from the MEM and FDTD method is obtained. The calculation confirms the suppression of the odd modes and the fact that only SPP mode propagates in the slit, as in the dielectric-modulated case. However, due to the difference in slit width, at the boundary of $y = Q^{(1)}$, the modal expansion in the different layers are quite different, as shown in Figs. 6(b) and (c). At the upper side of this boundary where the slit is wider, the total field mainly comes from the contribution of guided modes; while at the lower side where the slit is narrower, the guided modes cannot satisfy the field continuity condition because they evanesce quickly in metal, and thus the radiation modes have contributions to the expansion (the ordinal of each plotted mode in Fig. 6 is indicated by the number of its peaks). Obviously, the contributions from the higher modes cannot be ignored at the layer boundaries, indicating that the scattering is not weak. In addition, it is very interesting that the 2nd even mode in the wider layer (Layer 2) always has a relatively larger share in the total field which causes a strong field distribution at the metal/dielectric interface. This is why there are bright spots at the corners of a geometric-modulated slit, as shown in Fig. 6(a).
MEM and EIM-based Fresnel equations are plotted in Fig. 7. Our further numerical calculations manifest that as long as the core materials are the same in a geometric-modulated slit structure these features still hold. The phases of reflected SPP waves in the two opposite structures show remarkable difference, as shown in Figs. 7(b) and (d). All the remaining quantities are nearly the same for the two structures. This remarkable difference in phases of reflected waves is originated from the mismatch of eigenfunctions at the boundary between the two layers, and indicates that the scattering is not weak and the EIM cannot be employed in the present case.

In order to disclose the mechanism behind the transmission behavior, the SPPs transmission coefficients as a function of the length of Layer 2, \( q^{(2)} \), are calculated for three different structures, \([0.2 \sim 0.4 \sim 0.2], [0.4 \sim 0.2 \sim 0.4] \) and \([0.4 \sim 0.2 \sim 0.6] \).

First, in Fig. 8(a), the space between neighboring peaks is half length of the SPPs wavelength in Layer 2, for example, for \( w^{(2)} = 0.2 \, \mu m \) and \( 0.4 \, \mu m \), \( \lambda_{\text{SPP}} = 0.900 \, \mu m \) and 0.945 \( \mu m \), respectively, which indicates only the SPP mode propagates and resonates in Layer 2. The wave length and the position of the first peak can be used to calculate phase shift of the SPP mode. For \([0.4 \sim 0.2 \sim 0.4] \), the phase shift is \( \Delta \theta = 2 \, (0.386/0.9006) \ast 360^\circ \approx 308.59^\circ \), where the factor 2 is introduced by the absolute value of the fields. To confirm the result, we read the corresponding phase shift of SPP reflection coefficients from Fig. 7(b) to be \( \Delta \theta_{\text{SPP}} = [2 \ast (-154.30^\circ)] = 308.60^\circ \). For \([0.4 \sim 0.2 \sim 0.6] \), the phase shifts calculated from Fig. 8(a) and read from Fig. 7(b) are \( \Delta \theta = 2 \, (0.356/0.9006) \ast 360^\circ \approx 284.60^\circ \), and \( \Delta \theta_{\text{SPP}} = [-154.30^\circ - 129.93^\circ] = 284.23^\circ \), respectively. For \([0.2 \sim 0.4 \sim 0.2] \), the position of the first peak is 0.484 \( \mu m \), larger than the half of resonant SPP wavelength, which means that the first peak should locate at \((0.484 - 0.9449)/2 = 0.01155 \, \mu m \); thus the phase shift \( \Delta \theta = 2 \, (0.01155/0.9449) \ast 360^\circ \approx 8.80^\circ \); this time the phase shift of SPP reflection coef-

Fig. 7. Comparison between the SPPs transmission and reflection coefficients in a two-layer structure by the MEM (lines) and EIM-based Fresnel equation (dots). The structure is constructed by setting \( e_2^{(1)} = 1 \) and \( Q^{(0)} = Q^{(1)} = Q^{(2)} = 0 \). For waves incident from a narrower slit to a wider one, \( w^{(1)} = 0.2 \, \mu m, w^{(2)} = w^{(3)} \), (a) amplitude and (b) phase. For waves incident from a wider slit to a narrower one, \( w^{(1)} = w^{(2)}, w^{(3)} = 0.2 \, \mu m \), (c) amplitude and (d) phase.
and resonates in the modulated region. However, when the fit of the MEM and multi-reflection curves also indicates that only the SPP mode propagates reflection iteration program for revealing the transmission process (here we choose 10 times of scattering SPP transmission and reflection coefficients, read from Fig. 7, to compile a SPP multi-reflection processes, more or less affect the total transmission since they couple with each other at wavelengths metal slit structures are systematically studied. In general, all the modes, including the “guided mode”, oscillating in metal and mainly taking part in the field continuity at the layer boundaries.

On this basis, the transmission processes of the dielectric and geometric step-modulated sub-wavelength structure. This is because higher even eigenmodes in Layer 2 also make contribution to the transmission in such a thin modulation, as illustrated in Fig. 8(b). Although these modes are evanescent wave, it does not attenuate to a negligible value in this short distance, so that the single mode multi-reflection model cannot well describe the transmission. By introducing more modes into the multi-reflection process, the results converge to that of the MEM (the reflection time is set as 50 in Fig. 8(b)). This evanescent-wave assisted transmission in a structure beyond the skin-depth limit is wildly applied in optics [39-42] and recently highlighted in Terahertz domain [43]. We believe that the above mentioned multimode assisted multi-reflection is suitable for those experimental results in spite of a few differences in the geometry. Further numerical proofs will be given in our future paper.

4. Conclusion

In summary, we have applied the MEM to study the wave behaviors inside a metal slit structure with step modulations. The perfect-conducting walls in the model are theoretically necessary and numerically removable. This point is confirmed by simulated results of an unconfined slit by FDTD. The key point is that two kinds of eigenmodes exist in the system: one is the ”guided mode”; attenuating in metal and dominating the transmission process; the other is the ”radiation mode”, oscillating in metal and mainly taking part in the field continuity at the layer boundaries. On this basis, the transmission processes of the dielectric and geometric step-modulated sub-wavelength metal slit structures are systematically studied. In general, all the modes, including radiation modes, more or less affect the total transmission since they couple with each other at

Fig. 8. (a) Transmission vs. \( q^2 \) calculated by MEM for \([0.2 \sim 0.4 \sim 0.2]\) (solid line), \([0.4 \sim 0.2 \sim 0.4]\) (dashed line) and \([0.4 \sim 0.2 \sim 0.6]\) (dash-dotted line). The results calculated by FDTD (open circles) and multi-reflection (dots) are also shown for comparison. (b) Transmissions of a thin modulation in \([0.2 \sim 0.4 \sim 0.2]\) with different number of the mode that used in the multi-reflection process.
the layer boundaries. However, because the model is a symmetric subwavelength structure, only the SPP mode propagates and resonates in the scattering region and all the other modes decay quickly from the layer boundaries, which means that the amplitudes and phase shifts of the SPP coefficients are quite important to the transmission. For the dielectric modulated case, since the slit width remains the same, a quasi-transversal-phase match for the SPP modes, resulted from the boundary conditions, leads to the weak scattering at each layer boundaries, which gives the grant for using OMA to describe transmission procedure and EIM to simplify the calculating process. For the geometric modulation, due to the strong scattering at the layer boundaries, the EIM is not applicable anymore. Numerical results confirm that a phase shift of the resonant SPP reflection coefficient determines the position of the Fabry-Perot-like transmission peaks. In addition, the evanescent-wave assisted multi-reflection is demonstrated in a thin-modulated slit.

Acknowledgments

This work is supported by the 973 Program of China (Grant No.2011CB301801), the National Natural Science Foundation of China (Grant No. 10874124), and Natural Science Foundation of Beijing (No. 1102012).