

Scattering mechanism in a step-modulated subwavelength metal slit: a multi-mode multi-reflection analysis

C. Li^{1,2,a}, Y.S. Zhou¹, and H.Y. Wang³

¹ Center for Theoretical Physics, Department of Physics, Capital Normal University, Beijing 100048, P.R. China

² School of Mathematical Sciences, Capital Normal University, Beijing 100048, P.R. China

³ Department of Physics, Tsinghua University, Beijing 100084, P.R. China

Received 29 August 2011

Published online 11 January 2012

© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract. In this paper, the scattering/transmission inside a step-modulated subwavelength metal slit is investigated in detail. We firstly investigate the scattering in a junction structure by two types of structural changes. The variation of transmission and reflection coefficients depending on structural parameters are analysed. Then a multi-mode multi-reflection model based on ray theory is proposed to illustrate the transmission in the step-modulated slit explicitly. The key parts of this model are the multi-mode excitation and the superposition procedure of the scatterings from all possible modes, which represent the interference and energy transfer occurring at the interfaces. The method we use is an improved modal expansion method (MEM), which is a more practical and efficient version compared with the previous one [C. Li, Y.S. Zhou, H.Y. Wang, F.H. Wang, Opt. Express **19**, 10073 (2011)]. In addition, some commonly used methods including FDTD, scattering matrix method and improved characteristic impedance method are compared with MEM to highlight the accuracy of these methods.

1 Introduction

Subwavelength metal slits, as a kind of metal/insulator/metal waveguide, have attracted much attention in recent years not only because of their ability to guide light beyond the diffraction limit, but also because of several remarkable advantages, such as strong field localisation, simplicity and convenience of fabrication and integration into optical circuits [1–14]. When light (infrared, visible spectrum) propagates along a metal/air interface, it will excite a collective oscillation of free electrons at the surface of the metal, generating a field that decays exponentially away from the interface. This mode is called a surface plasmon polariton (SPP) [2–6]. In a subwavelength metal/air/metal slit the case is somehow different, since the SPP [5] wave decays exponentially in the metals and is flat in the air, which is the lowest eigenmode in the slit structure and the core part of the subwavelength metallic optics.

Step modulation is one of the key elements in photonic engineering that are employed in subwavelength metal structures to design and fabricate functional plasmonic devices, such as filters [7–10], reflectors [11] and photonic bandgap structures [11,12]. In addition step modulation is of important theoretical significance since it is helpful for investigating SPP scattering. An understanding of step modulation, combined with the staircase approximation

and transfer matrix technique, can be used to obtain numerical results describing more complicated structures.

Up until now, a number of methods have been used to calculate the SPP scattering/transmission inside a step-modulated slit. The finite-difference time-domain method (FDTD) is a well-developed simulation method that provides relatively accurate results and has been considered as a standard for testing other theoretical methods [5–14]. The effective index method, on the other hand, is a simplified and direct theoretical method where only the SPP modes are involved in the calculation, a method of one mode approximation; however, this simplification leads to considerable loss of the scattering information and the numerical inaccuracy turns out to be considerable under some conditions [6]. Matsuzaki et al. presented a transmission model and gave a better description of the SPP scattering using the characteristic impedance method [13]. Pannipitiya et al. [14] suggested an improved version of this method, which will be called the improved characteristic impedance method (ICIM) in this paper. Lin et al. presented a similar transmission model and used the scattering matrix method (SMM) to calculate the transmission [7–9]. Although the calculated results from these two methods fit the FDTD results, two approximations are used in the calculation, one mode approximation and quasi-statistic approximation. This limits their application scope and accuracy, as will be discussed in Section 3 below. Recently, we successfully applied the modal expansion method (MEM) to describe the wave behaviour inside

^a e-mail: lichao19840621@163.com

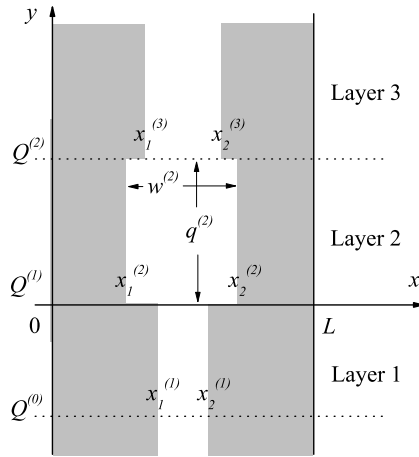


Fig. 1. Sketch of a step-modulated metal slit structure confined in the x direction with perfectly conducting walls at 0 and $L = 2 \mu\text{m}$. A TM wave with wavelength λ_0 is normally launched from $y = Q^{(0)}$. $Q^{(1)} = 0$.

a symmetric step-modulated slit [6]. This method did not involve the two aforementioned approximations and provided more accurate results.

In this paper, the MEM is further improved so as to apply to the asymmetric modulated case in investigating the scattering/transmission mechanisms inside a step-modulated slit. A multi-mode multi-reflection model is proposed to explain the transmission process. A remarkable advantage of MEM is that its accuracy is controllable. This enables us to discuss the accuracy of FDTD and ICIM, by comparing the results from these methods and MEM.

The paper is arranged as follows. Section 2 sets up our model of a step-modulated metal slit and presents the improved MEM formulas. In Section 3, the scattering in a junction structure is studied firstly as a prerequisite for later discussion, and then a multi-mode multi-reflection model is proposed to reveal the transmission mechanism in a step-modulated slit. Comparisons between MEM, FDTD, SMM, and ICIM are also given in this section to highlight the restriction of the one mode approximation and quasi-statistic approximation. Finally, conclusions are presented in Section 4.

2 Model and the improved MEM

In this section, we set up the single-slit structure model and present the formulas of the improved MEM, which are more practical and efficient compared with those in our previous work [6].

The model structure is shown in Figure 1. It is infinitely large in the yz plane but confined in the x direction by perfectly conducting walls at $x = 0$ and $x = L$. The structure is divided into three regions along the x direction, composed of silver/air/silver, and three layers along the y direction. The lower boundary of the l th layer is labeled as $Q^{(l-1)}$, and the interfaces between regions by

$x_1^{(l)}$ and $x_2^{(l)}$. The symbols $q^{(l)}$ and $w^{(l)}$ label layer height and slit width, respectively. A transverse magnetic (TM) wave, with magnetic field $\mathbf{H}$ being in the z direction, is normally launched from $y = Q^{(0)}$ in Layer 1 and propagates upward. The structural parameters are given in the caption of Figure 1.

The dielectric constant of silver as a function of the wavelength of the incident wave λ_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 [15], which is valid for $0.6 \leq \lambda_0 \leq 1.6 \mu\text{m}$. In this paper, the wavelength is mainly set as $\lambda_0 = 1 \mu\text{m}$; thus, $\varepsilon_{\text{Ag}} = -50.76 + 0.083i$.

In the remaining part of this section, we suggest an improved version of the MEM, which has the same output as the previous one [6] but is easier and faster.

The approach of the MEM is to expand the unknown functions (electromagnetic field distribution in the present case) by a complete set of orthogonal functions. This means that the MEM has two steps: one being the eigenvalue problem of the system; and the other to establish and solve the coupled equations subject to boundary conditions. However, the choice of the complete set in the modal expansion is not unique, but depends on the configuration of the system. It can be the eigenfunctions of a specific structure, or other functions such as sine or exponential functions.

In reference [6], the fields in the given structure were handled by separation of variables. The factors containing the x variable were expanded by the eigenmodes $\{\psi_n^{(l)}(x)\}$ between two perfectly conducting walls. The magnetic fields were expressed as

$$H_z(x, y) = \begin{cases} \sum_n \psi_n^{(1)}(x) [I_n e^{ik_{yn}^{(1)}(y-Q^{(0)})} + R_n e^{-ik_{yn}^{(1)}(y-Q^{(1)})}], & Q^{(0)} \leq y < Q^{(1)}, \\ \sum_n \psi_n^{(2)}(x) [E_n e^{ik_{yn}^{(2)}(y-Q^{(1)})} + F_n e^{-ik_{yn}^{(2)}(y-Q^{(2)})}], & Q^{(1)} \leq y < Q^{(2)}, \\ \sum_n \psi_n^{(3)}(x) T_n e^{ik_{yn}^{(3)}(y-Q^{(2)})}, & Q^{(2)} \leq y < \infty, \end{cases} \quad (1)$$

where I_n , R_n , E_n , F_n , and T_n were expansion coefficients which involved the scattering/transmission information of every eigenmode. It was necessary to solve a transcendental equation in order to determine the eigenvalues and eigenfunctions. Even with the assistance of a powerful root-seeking method [16], this procedure was still time-consuming. Moreover, each layer had its own eigenfunctions. At an interface, it was required by the boundary conditions to calculate the overlap between the eigenfunctions at the two sides of the interface, known as coupling integrals. Such integrals brought complexity to the program.

To avoid these difficulties, in this paper the factors containing the x variable were expanded by a sine basis, subject to the perfectly conducting boundary condition. That is to say, the complete set $\{\varphi_n(x)\}$ is chosen as

$$\varphi_n(x) = \sqrt{2/L} \sin(k_{xn}x), \quad k_{xn} = n\pi/L, \quad n = 1, 2, 3, \dots \quad (2)$$

The eigenvalues k_{xn} are solely determined by the distance between the two perfectly conducting walls, independent of the positions $x_1^{(l)}$ and $x_2^{(l)}$, so this basis set is valid for all three layers.

Correspondingly, the magnetic fields and their derivatives in the three layers can be expressed as [17]

$$H_z(x, y) = \begin{cases} \sum_n \varphi_n(x) \sum_m W_{n,m}^{(1)} [i_m e^{ik_{ym}^{(1)}(y-Q^{(0)})} + r_m e^{-ik_{ym}^{(1)}(y-Q^{(1)})}], & Q^{(0)} \leq y < Q^{(1)}, \\ \sum_n \varphi_n(x) \sum_m W_{n,m}^{(2)} [e_m e^{ik_{ym}^{(2)}(y-Q^{(1)})} + f_m e^{-ik_{ym}^{(2)}(y-Q^{(2)})}], & Q^{(1)} \leq y < Q^{(2)}, \\ \sum_n \varphi_n(x) \sum_m W_{n,m}^{(3)} t_m e^{ik_{ym}^{(3)}(y-Q^{(2)})}, & Q^{(2)} \leq y < \infty, \end{cases} \quad (3)$$

where i_m , r_m , e_m , f_m , and t_m are the expansion coefficients. The insertion of equation (3) into the Helmholtz equation yields an eigenvalue problem in each layer expressed by $A^{(l)} W^{(l)} = (ik_y^{(l)})^2 W^{(l)}$ with $ik_y^{(l)}$ and $W^{(l)}$ being the eigenvalues and eigenfunctions, and the operator $A^{(l)}$ being [18]

$$A^{(l)} = -[\tilde{E}^{(l)}]^{-1} \left\{ k_0^2 [I] + [K] [E^{(l)}]^{-1} [K] \right\}, \quad (4)$$

where

$$\begin{cases} [\tilde{E}^{(l)}]_{mn} = \int_0^L \varphi_m(x) \varphi_n(x) / \varepsilon^{(l)}(x) dx \\ [E^{(l)}]_{mn} = \int_0^L \varphi_m(x) \varphi_n(x) \varepsilon^{(l)}(x) dx \\ [K]_{mn} = \int_0^L \varphi_m(x) \frac{\partial}{\partial x} \varphi_n(x) dx \\ [I]_{mn} = \delta_{mn}. \end{cases} \quad (5)$$

In these equations, $[\cdot]$ denotes a $N \times N$ matrix where N is the truncation number, and $k_0 = 2\pi/\lambda_0$ is the wave vector in vacuum.

Here we mention the two advantages of the sine expansion in the x direction. One is that the eigenvalue problem, equation (4), is very easy for computer implementation, which avoids the cumbersome solution-seeking procedure necessary in the eigenmode expansion [6]. The other is that the complete set of sine functions are the same for all layers, so that the coupling integrals at the interfaces become quite simple. These two advantages greatly simplify the calculation program.

Corresponding to equation (3), the derivative of the magnetic field is expressed as

$$\frac{1}{\varepsilon} \frac{\partial}{\partial y} H_z = \begin{cases} \sum_n \frac{\varphi_n(x)}{\varepsilon^{(1)}} \sum_m W_{n,m}^{(1)} ik_{ym}^{(1)} [i_m e^{ik_{ym}^{(1)}(y-Q^{(0)})} - r_m e^{-ik_{ym}^{(1)}(y-Q^{(1)})}], & Q^{(0)} \leq y < Q^{(1)}, \\ \sum_n \frac{\varphi_n(x)}{\varepsilon^{(2)}} \sum_m W_{n,m}^{(2)} ik_{ym}^{(2)} [e_m e^{ik_{ym}^{(2)}(y-Q^{(1)})} - f_m e^{-ik_{ym}^{(2)}(y-Q^{(2)})}], & Q^{(1)} \leq y < Q^{(2)}, \\ \sum_n \frac{\varphi_n(x)}{\varepsilon^{(3)}} \sum_m W_{n,m}^{(3)} ik_{ym}^{(3)} t_m e^{ik_{ym}^{(3)}(y-Q^{(2)})}, & Q^{(2)} \leq y < \infty. \end{cases} \quad (6)$$

Applying the layer boundary conditions, we obtain the coupled equations as follows

$$\begin{cases} \sum_m W_{pm}^{(1)} [i_m e^{ik_{ym}^{(1)} q^{(1)}} + r_m] = \sum_m W_{pm}^{(2)} [e_m + f_m e^{ik_{ym}^{(2)} q^{(2)}}] \\ \sum_n \tilde{E}_{pn}^{(1)} \sum_m W_{nm}^{(1)} k_{ym}^{(1)} [i_m e^{ik_{ym}^{(1)} q^{(1)}} - r_m] = \sum_n \tilde{E}_{pn}^{(2)} \sum_m W_{nm}^{(2)} k_{ym}^{(2)} [e_m - f_m e^{ik_{ym}^{(2)} q^{(2)}}] \\ \sum_m W_{pm}^{(2)} [e_m e^{ik_{ym}^{(2)} q^{(2)}} + f_m] = \sum_m W_{pm}^{(3)} t_m \\ \sum_n \tilde{E}_{pn}^{(2)} \sum_m W_{nm}^{(2)} k_{ym}^{(2)} [e_m e^{ik_{ym}^{(2)} q^{(2)}} - f_m] = \sum_n \tilde{E}_{pn}^{(3)} \sum_m W_{nm}^{(3)} k_{ym}^{(3)} t_m. \end{cases} \quad (7)$$

The incident coefficients i_m are determined by the incident wave. In this paper, the incident wave is always a SPP wave launched in Layer 1, namely, $\psi_1^{(1)}(x)$. Therefore, one naturally has

$$\int_0^L \varphi_p(x) \psi_1^{(1)}(x) dx = \sum_m W_{pm}^{(1)} i_m, \quad (8)$$

which determines the coefficients i_m . After setting the incident coefficients i_m , the four groups of coefficients, r_m , e_m , f_m , and t_m can be obtained from equation (7). Thus all the field quantities are obtained.

In this paper, we will focus on discussing the reflection/transmission mechanisms, which are mainly described by the reflection coefficients R_n and transmission coefficients T_n in equation (1). For example, the amplitudes of the SPP modes in Layer 1 and 3, $|R_n|$ and $|T_n|$, are the reflection and transmission efficiencies of the system, and their arguments, $\arg(R_n)$ and $\arg(T_n)$, are the corresponding phase shifts. Therefore, a projection between the fields calculated by equation (3) and the eigenmodes $\{\psi_n^{(l)}(x)\}$ is implemented for obtaining R_n and T_n . In the following, the absolute values of these coefficients generally corresponds to the excitation efficiency.

To summarise our proposed approach: the field is expanded using sine functions which are complete and uniform in all layers. The corresponding eigenvalue problem reduces to a matrix equation as shown in equation (4), which makes the calculation quite easy. Accordingly, the procedure here is much more practical and efficient compared to the previous one [6]. Finally, the overlaps between the calculated field and the eigenmodes in equation (1) give the reflection and transmission coefficients necessary for physical analysis.

Although we merely study the three-layer structure, the procedure we have developed here is easily applied to more complicated structures by implementing the S matrix algorithm [19] or the enhanced transmittance matrix approach [20].

3 Numerical results and analysis

In this section, we investigate the scattering/transmission mechanisms inside a step-modulated subwavelength metal

slit. To do so, the scattering in a junction structure is first discussed in detail, since the slit comprises more than one junction structure. Then we disclose the multi-mode multi-reflection model in the transmission process in the slit. In the course of this discussion, the accuracy of FDTD and ICIM is considered by comparing the results of these two methods with those obtained using the MEM.

In our calculation, the confinement is set as $L = 2 \mu\text{m}$. We have tested that 800 modes, $N = 800$ in equation (4), are enough to give results with precision up to four significant digits. The convergence test for truncation number N and the accuracy test for confined width L will be carried out later in Figure 7.

3.1 Junction structures

A junction structure is the connection of two half-infinitely long slits with widths being denoted by $w^{(1)}$ and $w^{(3)}$, respectively, which can be easily realised in Figure 1 by setting the height of Layer 2 to 0. In reference [6], some scattering properties of symmetric structures have been revealed. For example, the main component of the fields inside the slits were guided modes which played a very important role in scattering/transmission, and the unimportant components were the radiation modes excited which were necessary to fulfill the boundary conditions, but made little contribution to the transmission. The following discussion will therefore focus on the guided modes. However, the discussion in reference [6] was limited to the symmetric case. Here we present a detailed investigation on how the scattering is affected by asymmetry.

Two types of structural changes are considered. In Type I, the widths of the two slits are fixed and the position of the narrower one can be anywhere between the left and right extremities, as shown in the inner panel of Figures 2a and 2c. In Type II, the left walls of the slits are aligned and the width of the narrower one is fixed, but that of the wider one can vary, as schematically shown in the inner panel of Figures 3a and 3c.

The results of the Type I structure are plotted in Figure 2. The left wall of the wider slit is at $x_1 = 0.6 \mu\text{m}$. For $w^{(1)} = 0.1$ and $w^{(3)} = 0.8 \mu\text{m}$, when the position of the narrower slit is moved from left to right, the excitation efficiencies and their phase shifts are plotted in Figures 2a and 2b, respectively. In Figures 2c and 2d are the excitation efficiencies and their corresponding phase shifts for the structure with $w^{(1)} = 0.8$ and $w^{(3)} = 0.1 \mu\text{m}$. In Figures 2e and 2f are the absolute values of the first three eigenfunctions of the narrower and wider slits, respectively. In Figures 2a, 2c, 2e and 2f the absolute values are plotted because these quantities are complex. The eigenmodes in the narrower and wider slits are denoted by $\psi_n^{(na)}$ and $\psi_n^{(wi)}$, respectively. In Figures 2e and 2f, the lowest modes $\psi_1^{(na)}$ and $\psi_1^{(wi)}$, plotted as black curves, are just SPP modes, and the second modes $\psi_2^{(na)}$ and $\psi_2^{(wi)}$, plotted as red curves, are actually antisymmetric wave functions within the slits. Note that the curve of $|\psi_3^{(wi)}|$

is divided into three parts by two zeros. The sign of the central part of $\psi_3^{(wi)}$ is reversed relative to the other parts.

We note that under our present parameters, only the first two eigenmodes of the narrower slit, $\psi_1^{(na)}$ and $\psi_2^{(na)}$, are within the slit (guided modes), see the black and red lines in Figure 2e. The wave functions of the higher modes are mainly distributed within the metal (radiation modes), see, as an example, the third mode $\psi_3^{(na)}$ in Figure 2e. The behaviour of $|\psi_3^{(na)}|$ with the position of the narrower slit has to be explicitly described as follows. When the slit moves to the right, the width of metal at the right side of the slit becomes thinner, so that the hill is compressed. If the narrower slit is on the right side of the centre position $x = 1 \mu\text{m}$, the hill will appear at the left side of the slit. If the slit is just at, or very near, the centre $x = 1 \mu\text{m}$, there will be two hills at the two sides of the slit, respectively, since the structure in this case is symmetric [6].

Three obvious features of the excitation efficiencies can be seen in Figures 2a and 2c. The first is that all the curves there exhibit a central symmetry, because all the configurations are symmetric with respect to the central line at $x = 1 \mu\text{m}$. The second is, from comparison of the black and red solid lines in Figures 2a and 2c, that the excitation efficiencies of the SPP modes in narrow slits are much larger than those of the second modes. The third is, by inspection of dash-dotted lines in Figures 2a and 2c, that the shapes of the efficiency curves of the modes in wider slits resemble their eigenfunctions $|\psi_n^{(wi)}|$ in Figure 2f. The latter two features can be attributed to the treatment of MEM which involves a mutual expansion between the modes in different layers.

We should keep in mind that the total field at the layer boundary, $H_z(x)$, can be respectively achieved by the linear combination of eigenfunctions in the narrower and wider slit, and the expansion coefficients depend on the position of the narrower slit, subject to boundary conditions. Then, the curves in Figures 2a and 2c can be explained qualitatively.

We first consider the case where the wave is incident from the narrower slit to the wider one, as shown by the inset in Figure 2a. The reflection efficiencies $|R_n|$ are proportional to the projection $\int_0^L \psi_n^{(na)} H_z(x) dx$, where the integration is along the interface between the narrower and wider slits. Since the eigenfunctions in the wider slit or their combination, $H_z(x)$, can be seen as a smooth variation within the narrower slit, the reflection efficiency of the SPP mode in the narrower slit, $|R_1| \propto \int_0^L \psi_1^{(na)} H_z(x) dx \approx \int_{w^{(na)}} \psi_1^{(na)} H_z(x) dx$, is dominant because $\psi_1^{(na)}$ is also smooth within the slit, and $|R_2| \propto \int_0^L \psi_2^{(na)} H_z(x) dx \approx \int_{w^{(na)}} \psi_2^{(na)} H_z(x) dx$ is very small because the second mode is an antisymmetric function within the slit. For the transmitted waves, the transmission efficiencies $|T_n|$ are qualitatively determined by $\int_0^L \psi_n^{(wi)} H_z(x) dx$, which includes $\int_0^L \psi_n^{(wi)} \psi_1^{(na)} dx$, $\int_0^L \psi_n^{(wi)} \psi_2^{(na)} dx$, $\int_0^L \psi_n^{(wi)} \psi_3^{(na)} dx$ and so on. We have already shown that the excitation of the second mode in the narrow slit is very small, so that the contribution

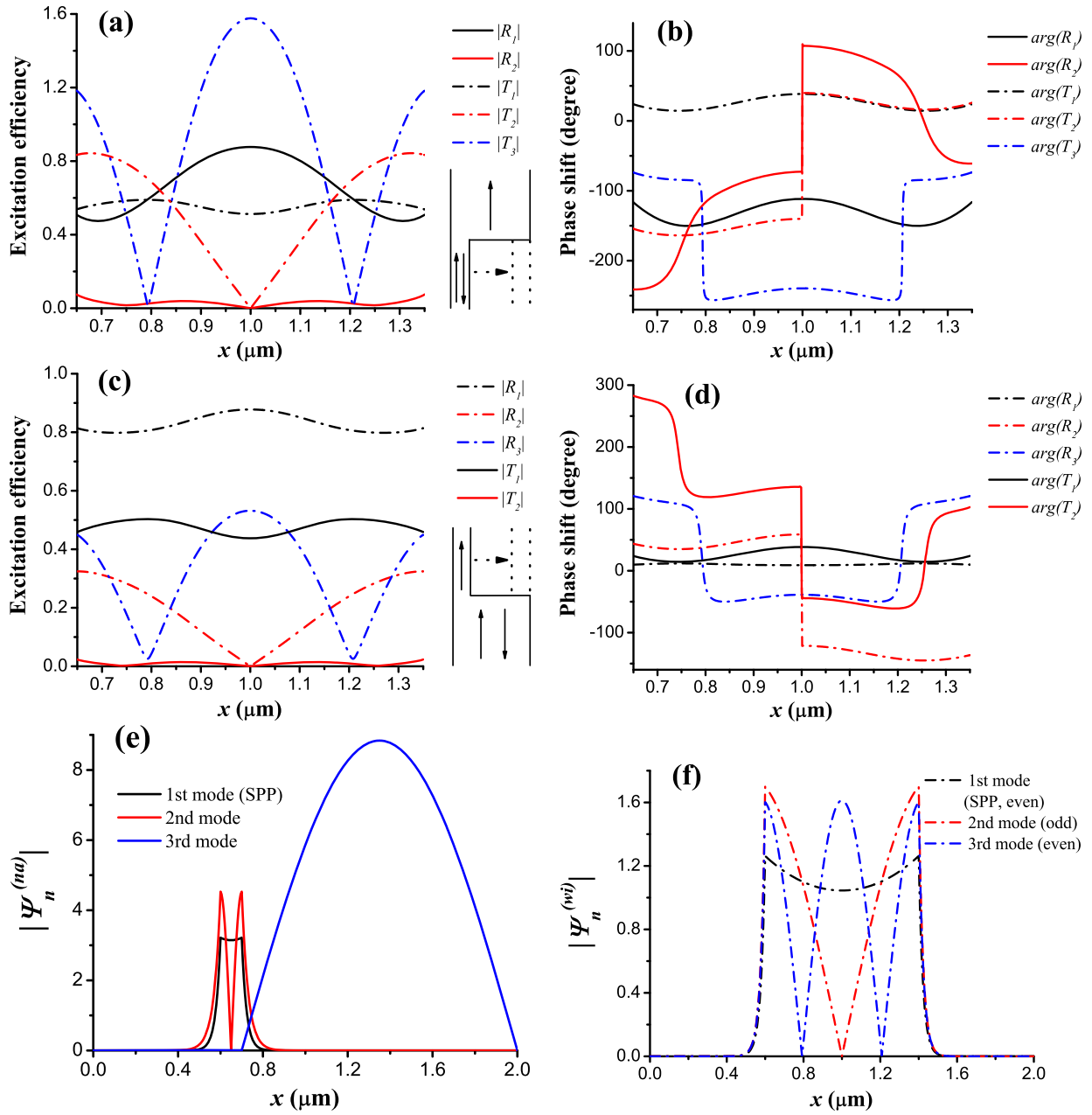


Fig. 2. (Color online) Scattering in Type I junction structures. The left wall of the wider slit is fixed at $x_1 = 0.6$ μm . $Q^{(0)} = Q^{(1)} = Q^{(2)} = 0$. In all the figures, the results of the narrower slit are plotted by solid lines and those of the wider slit by dash-dotted lines. The narrower slit moves from the left to right. In (a) to (d), the x -axes are its central position. For the structure with $w^{(1)} = 0.1$ and $w^{(3)} = 0.8$ μm , (a) excitation efficiency; (b) phase shift. For the structure with $w^{(1)} = 0.8$ and $w^{(3)} = 0.1$ μm , (c) excitation efficiency; (d) phase shift. (e) and (f) depict the absolute value of the eigenfunctions in the narrower and wider slits, respectively, when their left walls are aligned.

of the factor $\int_0^L \psi_n^{(wi)} \psi_2^{(na)} dx$ is negligible. The contribution of the radiation modes is relatively complicated, but unimportant because what happened inside the slit is the key part of the scattering procedure, while the radiation modes localised in the metal are excited to fulfil the boundary condition outside the slit. That is why we try to avoid these modes in the discussion. By employing several numerical tests, it can be shown that the radiation modes do make a contribution to the transmitted waves, but that this contribution is comparatively small. Therefore, the

factor $\int_0^L \psi_n^{(wi)} \psi_1^{(na)} dx \approx \int_{w^{(na)}} \psi_n^{(wi)} \psi_1^{(na)} dx$ mainly determines the transmission efficiencies $|T_n|$. As an example, let us consider the $|T_3|$ curve. $|T_3| \propto \int_{w^{(na)}} \psi_3^{(wi)} \psi_1^{(na)} dx$, where $\psi_1^{(na)}$ is smooth within a narrow region, see, the black line in Figure 2e. When the narrower slit is positioned at the left side with its centre being at $x = 0.65$ μm , $|\psi_3^{(wi)}|$ has a maximum at this position. Therefore the projection of $\psi_1^{(na)}$ onto $\psi_3^{(wi)}$ is at a maximum. As the narrower slit moves to the right, we image that the black

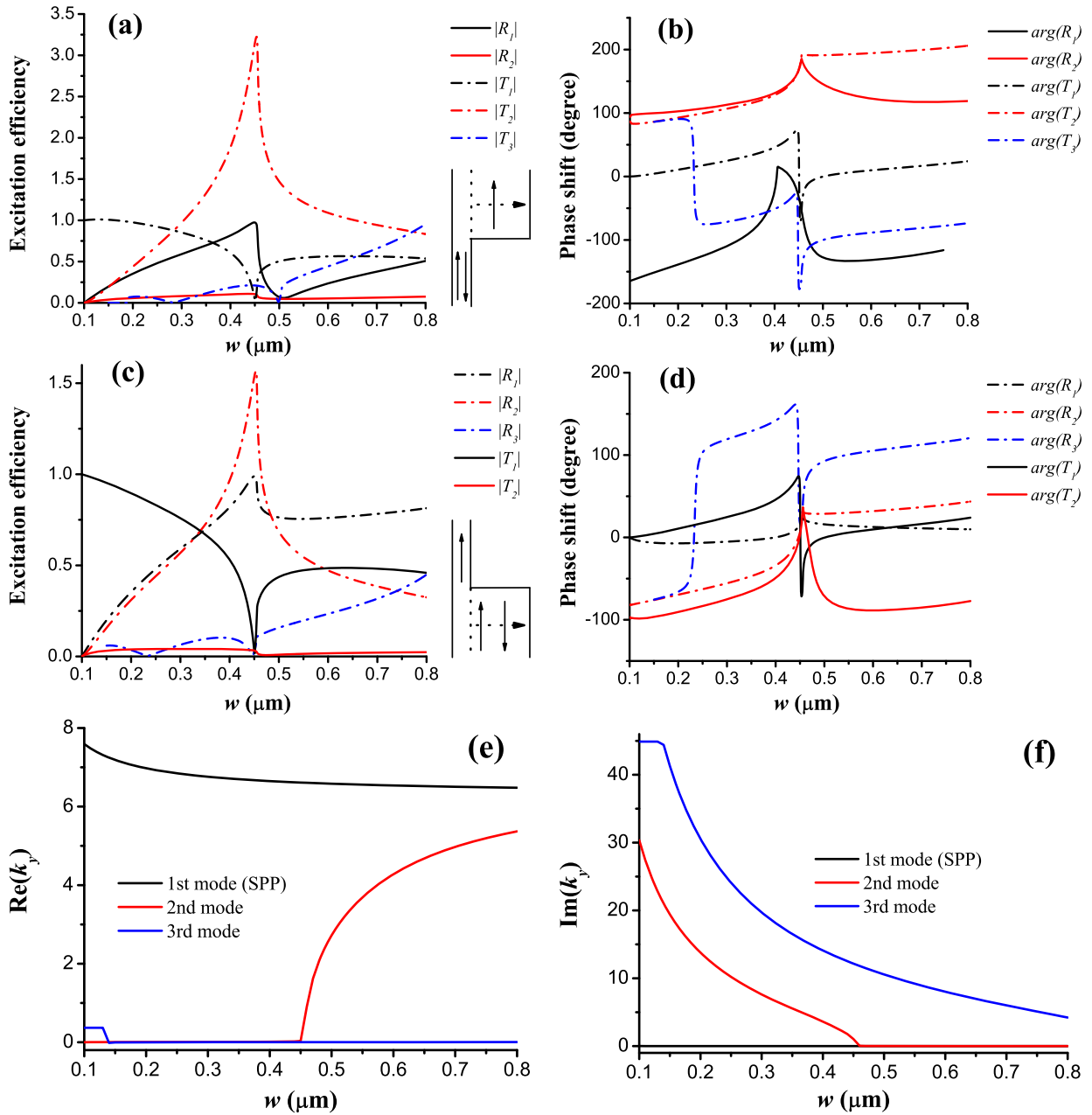


Fig. 3. (Color online) Scattering in Type II junction structures. The left walls of the slits are aligned at $x_1 = 0.6 \mu\text{m}$. $Q^{(0)} = Q^{(1)} = Q^{(2)} = 0$. In all the figures, the results of the narrower slit are plotted by solid lines and those of the wider slit by dash-dotted lines. The width of the narrower slit is $0.1 \mu\text{m}$. The x -axes are the width of the wider slit, denoted as w . For a wave incident from the narrower slit to the wider one, (a) excitation efficiency; (b) phase shift. For a wave incident from the wider slit to the narrower one, (c) excitation efficiency; (d) phase shift. (e) and (f) show the real and imaginary parts of the propagation constant k_y , appearing in equation (1) with the notation $k_y^{(l)}$, of the wider slit.

curve in Figure 2e also shifts to the right. At $x = 0.794 \mu\text{m}$, $|\psi_3^{(wi)}|$ is zero. Accordingly, the projection of $\psi_1^{(na)}$ at this position onto $\psi_3^{(wi)}$, as well as $|T_3|$, becomes zero. Between $x = 0.65$ and $x = 0.794 \mu\text{m}$, $|T_3|$ should drop from the maximum to zero. We notice that around the zero, the phase of T_3 changes by nearly π . At the other zero of $|\psi_3^{(wi)}|$ at $x = 1.206 \mu\text{m}$, $|T_3|$ again reaches zero and its phase changes by nearly π once more. This analysis explains why the shape of $|T_3|$ is similar to $|\psi_3^{(wi)}|$. It is the

narrow and smooth profile of $|\psi_1^{(na)}|$ that results in the similarity between the $|T_3|$ and $|\psi_3^{(wi)}|$ curves. The $|T_2|$ curve in Figure 2a is understood in the same way. $|T_1|$ is mainly determined by $\int_{w^{(na)}} \psi_1^{(wi)} \psi_1^{(na)} dx$, which is a smooth and relatively flat curve due to the smooth variations of both SPP waves.

We next look to the case where a SPP wave is incident from the wider slit to the narrower one, as shown in the inset of Figure 2c, where the incident wave is a

smooth curve within the range of the wider slit width, see the black curve in Figure 2f. We again begin with the waves in the narrower slit. $|T_n|$ is proportional to the integral $\int_0^L \psi_n^{(na)} H_z(x) dx$, where $H_z(x)$ is the combination of eigenfunctions in the wider slit and considered as a smooth varying curve within the range of the narrower slit, so that the transmission efficiency of the SPP mode is dominant and much larger than that of the second mode. For the waves in the wider slit, ignoring the contribution of the second mode and radiation modes, the reflection efficiencies $|R_n|$ are mainly determined by $\int_{w(na)} \psi_n^{(wi)} \psi_1^{(na)} dx$, leading to the fact that the $|R_n|$ curves in Figure 2c have similar shapes to the $|T_n|$ curves in Figure 2a.

The transmission efficiency $|T_1|$ in Figure 2a is exactly the same as that in Figure 2c, and the efficiencies $|R_1|$, $|T_2|$ and $|T_3|$ in Figure 2a have the same behaviour as $|R_1|$, $|R_2|$ and $|R_3|$ in Figure 2c, respectively, although with different values. The reflection efficiencies $|R_1|$ and $|R_2|$ in Figure 2c are higher than those in Figure 2a. This is because Figure 2c represents the case of a wave incident from a wider slit to a narrower one, which needs to squeeze light into a narrower space, so the higher reflection is understandable.

The variations of the scattering phase shifts for the two different incident cases discussed above are plotted in Figures 2b and 2d. It is seen that the phases of T_1 in both these figures are also the same. We note that at the positions where $\psi_n^{(wi)}$ is zero, the corresponding coefficients R_n and T_n have an associated phase change of π .

The results of the Type II structure are plotted in Figure 3. The left walls of the slits are aligned at $x_1 = 0.6 \mu\text{m}$. The width of the narrower slit is $0.1 \mu\text{m}$, but that of the wider one, denoted by w , varies from 0.1 to $0.8 \mu\text{m}$, as shown in the insets in Figures 3a and 3c. The most distinct feature is the drastic changes of the excitation efficiencies over a narrow range of slit width, as shown in Figures 3a and 3c near $w = 0.46 \mu\text{m}$.

When the width of a slit is 0.1 and $0.8 \mu\text{m}$, the first three eigenmodes have been plotted in Figures 2e and 2f, respectively. Now the width w varies. Our calculation shows that the modes resemble those in Figure 2e when $w < 0.15 \mu\text{m}$ and those in Figure 2f otherwise. The first three eigenmodes are denoted ψ_1 , ψ_2 and ψ_3 , and their k_y 's are by k_{y1} , k_{y2} and k_{y3} , respectively.

The SPP mode ψ_1 is obviously a propagation mode since k_{y1} has a negligible imaginary part, and ψ_3 is a decaying mode as demonstrated by the large imaginary component of k_{y3} . When $w < 0.46 \mu\text{m}$, k_{y2} is nearly purely imaginary so that ψ_2 is an evanescent mode, while when $w > 0.46 \mu\text{m}$, k_{y2} becomes nearly purely real so that ψ_2 becomes a propagation mode. A turning point appears at $w = 0.46 \mu\text{m}$ at which ψ_2 transforms its propagation property. This transformation leads to the drastic changes of the excitation efficiencies, similar to the cause of the well-known Wood's anomaly in the grating theory [21].

The analyses of the excitation efficiencies in Figures 3a and 3c are performed in the same way as those in the Type I structure. Therefore, some similar conclusions are

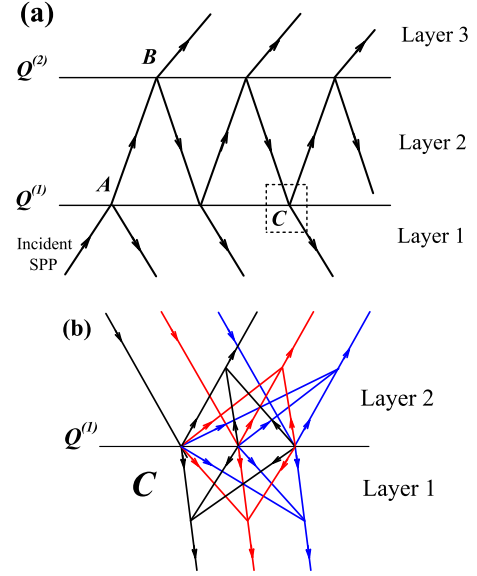


Fig. 4. (Color online) Sketch of the multi-mode multi-reflection model: (a) multi-reflection between interfaces and (b) multi-mode excitations at each point.

obtained, such as the identity of $|T_1|$ curves in Figures 3a and 3c and the similarity between the efficiencies $|R_1|$, $|T_2|$ and $|T_3|$ in Figure 3a and $|R_1|$, $|R_2|$ and $|R_3|$ in Figure 3c, respectively, although with different values.

The drastic changes of excitation efficiencies at the turning point have to be explained from an energy perspective. As an example, let us consider the case where the wave is incident from the narrower slit to wider one, as shown by the inset in Figure 3a. At the start point $w = 0.1 \mu\text{m}$, the two slits are the same, so that $|T_1| = 1$ and all other excitation efficiencies are zero. Close to the turning point, $|T_1|$ reaches a minimum, and $|R_1|$ and $|T_2|$ reach a maximum. Since the second mode in the wider slit ψ_2 now is an evanescent mode, the energy is mainly stored in the reflected SPP wave. Once ψ_2 becomes a propagation mode, it must gain a large portion of energy from the reflection due to its large excitation efficiency, and leads to the rapid drop of $|R_1|$ as shown in Figure 3a. At the same time, the decreasing $|R_1|$ causes a further redistribution of excitation efficiencies by the boundary continuum condition. Therefore, what lies behind the drastic changes of excitation efficiencies is a redistribution of energy between evanescent modes and propagation modes. The explanation of the energy redistribution is also suitable for the case where the wave is incident from the wider slit to the narrower one.

It is seen from Figures 3b and 3d that the phases of both T_1 in these two figures are also the same. At the turning point, T_1 changes its phase by π . The change of π in phase at the turning point also occurs for T_3 in Figure 3b and R_3 in Figure 3d.

So far, the scattering mechanisms of the two types of structures have been investigated. Although the investigation here is restricted to the incident wave with wavelength

$\lambda_0 = 1 \mu\text{m}$, the analysis above is also applicable to the infrared and visible spectrum. Furthermore, the analysis has important practical applications. For example, one can excite/suppress specific modes to control the field distribution inside a slit, or design a high-efficiency reflector, by changing the position or width of the slit.

3.2 The multi-mode multi-reflection model

Having understood the scattering at the interface in a single junction structure, we are ready in this subsection to discuss the transmission in a step-modulated slit, which can be regarded as the combination of two junction structures. In order to reveal the transmission clearly, we present here an analysis of the multi-mode multi-reflection model that combines wave and ray optics.

Figure 4 presents a sketch of the multi-mode multi-reflection model. In a step-modulated subwavelength metal slit, there are two interfaces at $Q^{(1)}$ and $Q^{(2)}$. Let us discuss the wave reflection and transmission in the slit. When the incident SPP launched from $Q^{(0)}$ in Layer 1 impinges on the interface between Layers 1 and 2, $Q^{(1)}$, it generates a reflection wave in Layer 1 and transmission wave in Layer 2. The latter continues going upwards, and when it reaches the other interface, $Q^{(2)}$, yields reflection and transmission waves again. Obviously, there occurs multi-reflection in Layer 2, shown by Figure 4a.

Because the wave in each layer is some linear combination of the eigenmodes of the layer, each ray in Figure 4a can in fact be expanded by eigenmodes, except the primary incident light. For example, at the point A, the reflected wave contains all possible modes in Layer 1 and the transmitted wave contains the modes in Layer 2. That is to say, the scattering excites all the modes in both layers. When all the possible modes in Layer 2 reach point B, each mode again excites all possible eigenmodes in a reflected wave in Layer 2 and in a transmitted wave in Layer 3. This phenomenon is termed multi-mode excitation. To show the phenomenon explicitly, we draw in Figure 4b the multi-mode excitation at point C. Suppose that the waves in Layers 1 and 2 are expanded by three eigenmodes, respectively. Then when the three modes in Layer 2 are incident at point C, as shown in Figure 4b, the first mode yields the reflected and transmitted waves, both containing three eigenmodes in the respective layer, i.e. the incident black line excites the black, red and blue lines in the transmitted waves in Layer 1 and reflected waves in Layer 2, respectively. In the same way, the incident red line also excites the black, red and blue lines in the transmitted waves in Layers 1 and reflected waves in Layer 2, respectively, and so does the incident blue line. Therefore, the total reflected wave at point C includes three eigenmodes in Layer 2, each being in turn the superposition of the reflections from the three incident eigenmodes. Similarly, the total transmitted wave at point C includes three eigenmodes in Layer 1, each being in turn the superposition of the transmissions from the three incident eigenmodes.

In summary, the total transmission and reflection coefficients in Layer 3 and Layer 1 in Figure 4a are obtained by summing up all the single-scattered coefficients, respectively. In addition, it is worth mentioning that the multi-mode multi-reflection model is a generalised form of the single-mode multi-reflection model which occurs in a F-P cavity. The former will be simplified to be the latter if only one mode can be excited.

In this paper, when we say model analysis, we mean the physical explanation of the multi-mode multi-reflection. In order to test this analysis, numerical calculations based on this physical picture were carried out and the results compared with MEM. In Figure 5 we have plotted the transmission efficiencies as a function of the length of Layer 2, $q^{(2)}$, when the incident wave is SPP mode. The solid lines in Figures 5a and 5b are the results of MEM, which necessarily comprise the contributions from all possible eigenmodes. The symbols represent the results of the model analysis. In a slit with width $w^{(2)} = 0.3 \mu\text{m}$, only the first mode, i.e. the SPP mode, can propagate and all other modes are evanescent. Thus, when $q^{(2)}$ is sufficiently long, the higher modes attenuate to a negligible value, and the transmission can be well described by a multi-reflection of only the SPP mode.

In Figure 5a it is seen that the results of the model analysis including the first two modes are the same as the line obtained from MEM. When $q^{(2)} > 0.8 \mu\text{m}$, the circles and crosses are identical, indicating that the contribution from the second mode is negligible. While for $q^{(2)} < 0.8 \mu\text{m}$, crosses deviate from circles, indicating that the second mode should not be omitted since it does not fade out within this distance range. In Figure 5a, the crosses end at $q^{(2)} = 0.09 \mu\text{m}$, because below this distance the multi-reflection of the first two modes diverges. What is the reason for this divergence? Firstly, the divergence is not caused by the propagation mode since the multi-reflection of the SPP mode always converges, as shown by the circles in Figure 5a. Neither is it caused by the exponentially increasing term which originates from improperly handling of the evanescent waves [20], for it occurs only at short distances. Actually, the divergence arises from the coupling between the eigenmodes containing the evanescent modes. With the contribution of the evanescent wave, as shown in Figure 4b, the superposition will result in larger transmission and reflection coefficients after each scattering if the second mode does not decay to a certain value. Thus, there exists a critical distance above which the multi-mode multi-reflection analysis is applicable. For the structure given in Figure 5a, it is $q^{(2)} = 0.09 \mu\text{m}$.

To verify the above conclusion concerning the divergence of the evanescent mode, we suppress the antisymmetric second mode by reforming the slit structure to a symmetric one, as shown in Figure 5b. Then the circles and crosses are identical at any distance. Let us see the contribution from the third mode. The plus signs including contributions from the first three modes are in good agreement with the results of MEM as $q^{(2)} > 0.012 \mu\text{m}$. The crosses and circles are identical when $q^{(2)}$ is above $0.3 \mu\text{m}$, but it is not so when $q^{(2)}$ is below $0.3 \mu\text{m}$.

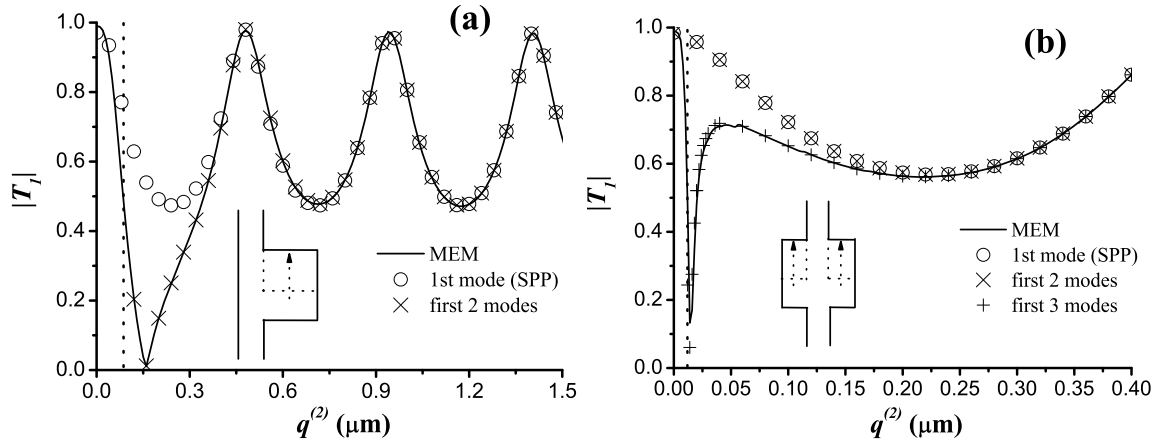


Fig. 5. The SPP transmission vs. $q^{(2)}$ calculated by MEM and model analysis for slit structure with $[Q^{(0)}, Q^{(1)}, Q^{(2)}] = [-1, 0, q^{(2)}] \mu\text{m}$. The solid lines are the results of MEM. The circles, crosses and plus signs are the results including contributions from the first one, two and three modes, respectively, from the model analysis. (a) Slits align to left at $x_1^{(l)} = 0.85 \mu\text{m}$, $[w^{(1)}, w^{(2)}, w^{(3)}] = [0.1, 0.3, 0.1] \mu\text{m}$; (b) the symmetric case of the slit structure in (a).

That is to say, if the distance is less than $0.3 \mu\text{m}$, the evanescent third mode is not negligible. This time the critical height for the third mode is $q^{(2)} = 0.012 \mu\text{m}$, which is much smaller than that of the second mode in the structure shown in Figure 5a. The reason is that the decay of the third mode is faster than that of the second mode, for the propagation constant k_y of the former has a larger imaginary part than the latter, as shown in Figure 3f.

Next, we investigate the coupling between propagation modes. In Figure 5a, only the first mode, the SPP mode, is a propagation mode in Layer 2 with width $w^{(2)} = 0.3 \mu\text{m}$. When the width is enlarged, the second mode can also become propagating. In Figure 6a the transmission efficiencies are plotted in the same structures as in Figure 5a except that the width of Layer 2 is extended to be $0.8 \mu\text{m}$. For this width, the second mode is indeed propagating. It can be seen that even when $q^{(2)}$ goes to zero, the result containing the contributions from the first two modes is not divergent, and agrees with the MEM curve very well when $q^{(2)} > 1.2 \mu\text{m}$, which confirms the statement that the propagation modes do not cause the divergence. When $q^{(2)}$ is below $1.2 \mu\text{m}$, the contribution from the third mode has to be added in order to achieve accurate results. However the cost of this accuracy is the introduction of divergence below $q^{(2)} = 0.181 \mu\text{m}$.

Similar to the treatment in Figure 5b where the second mode in Layer 2 is removed by a structural change, it is also possible to suppress the third mode excited in Layer 2. The way to implement the suppression is to shift the center of the narrower slits to $x = 0.794 \mu\text{m}$, as shown in the inset of Figure 6b. At this position, the excitation efficiency of the third mode is nearly zero, see Figure 2a. The transmission results are plotted in Figure 6b. It is seen from the figure that up to the first five eigenmodes have to be included in the model analysis in order to meet the MEM curve. The divergence in this case is caused by

the fourth mode, and the corresponding critical width is $q^{(2)} = 0.084 \mu\text{m}$.

In summary, the multi-mode multi-reflection model provides a correct description of the transmission inside a step-modulated subwavelength metal slit when the height of modulated layer (Layer 2) is above a critical height. It fails otherwise due to the coupling between propagation modes and evanescent modes.

3.3 Comparison of different methods

In this subsection, the MEM and the other three methods, FDTD, SMM and ICIM, are discussed, and the calculated results of FDTD and ICIM are compared to the MEM results. The accuracy of these methods is investigated and some useful conclusions are obtained. Before presenting the numerical results, we would like to briefly compare these four methods.

FDTD, a commonly used simulation method in optics, is used to calculate field quantities directly from Maxwell's equations. In principle, this method and MEM can both provide accurate and reliable results. Here we would like to point out the three key differences between them. Firstly, the way that they solve Maxwell's equations is different: FDTD uses the finite difference method to evolve fields in space and time domains, while the MEM establishes and solves the coupled equations in the frequency domain by the method of moments. Secondly, the way that they handle outermost boundaries is different: FDTD makes use of, for the outermost boundaries of a system, perfectly matched layers which can totally absorb waves without reflecting them back, while MEM confines the structure with two perfectly conducting walls such as in this paper. The feasibility of the latter is due to the fast attenuation of light (infrared, visible spectrum) in a metal. If the confined width is large enough, the effects introduced by the two perfectly conducting walls are negligible, as shown in

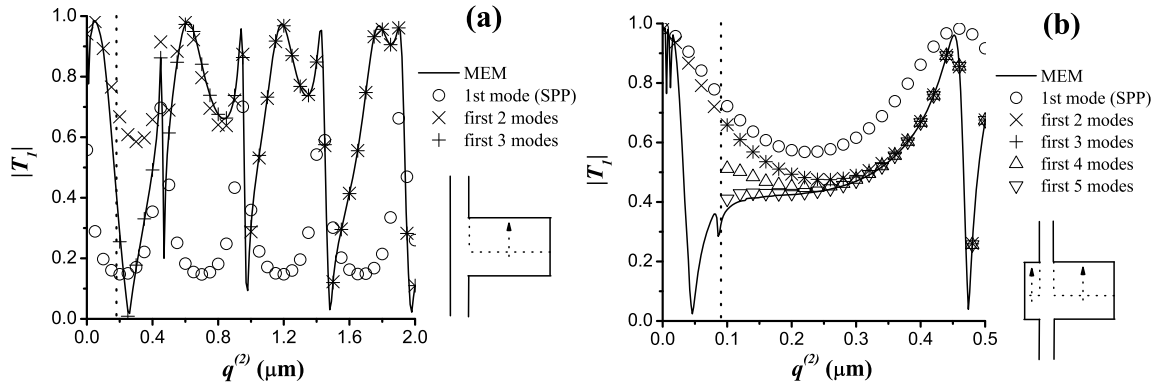


Fig. 6. The SPP transmission vs. $q^{(2)}$ calculated by MEM and model analysis for slit structure with $[Q^{(0)}, Q^{(1)}, Q^{(2)}] = [-1, 0, q^{(2)}] \mu\text{m}$. The solid lines are the results of MEM. The circles, crosses, plus signs, up-triangles and down-triangles are the results including contributions from the first one to five modes, respectively, from the model analysis. (a) Slits align to left at $x_1^{(1)} = 0.6 \mu\text{m}$, $[w^{(1)}, w^{(2)}, w^{(3)}] = [0.1, 0.8, 0.1] \mu\text{m}$; (b) the same structure but the narrower slits are shifted to be $x_1^{(1)} = x_1^{(3)} = 0.744 \mu\text{m}$, a position that totally suppresses the excitation of the 3rd mode, see Figure 2a.

Figure 7a below. Although the perfectly matched layer technique can be introduced to MEM [22], it dramatically complicates the MEM calculation. Thirdly, the way that they converge is different: the convergence of FDTD depends on the size of the Yee cell used in the simulation, while that of MEM on the confined width L and the truncation number N .

SMM [7–9] and ICIM [13,14] are two other frequently used methods which show the following three features. Firstly, according to the two methods, the modulated region, Layer 2, would be divided into a central scattering region and a stub (as shown in Fig. 2 of Ref. [7] and Fig. 4 of Ref. [13]), and it would be assumed that the SPP mode multi-reflection occurred in the stub (although Refs. [13,14] did not mention this point, it could be recognised from the transmission equations, Eq. (4) in Ref. [13] and Eqs. (8) in Ref. [14]). Secondly, both of them use the one mode approximation, which means that only the SPP modes exist in the stub and slits. Thirdly, the phase shifts caused by scattering in the central scattering region could not be calculated properly, so they were ignored by means of the quasi-statistic approximation [13,14,23] (although Refs. [7–9] did not mention the quasi-statistic approximation, it was easily seen by the procedure used to obtain the scattering matrix given in Ref. [8]).

In the following, a numerical comparison between these methods is performed. The convergence comparison of MEM and FDTD in a Type I structure is presented in Figure 7.

Figure 7a shows the calculated results of MEM where the confined width $L = 2, 4 \mu\text{m}$ and the truncation number $N = 400, 800$, respectively. The results plotted as squares, crosses and plus signs in Figure 7a are identical, showing that the boundary effect imposed by the perfectly conducting walls can be completely ignored in such confined widths. As already mentioned at the beginning of Section 3, the parameters $L = 2 \mu\text{m}$ and $N = 800$ ensure that all the calculated results have at least four signifi-

cant digits. A FDTD simulated transmission curve with cell size $2.5 \times 2.5 \text{ nm}^2$ is also plotted in Figure 7a for comparison. Obviously, the FDTD curve is close to the MEM ones, but they do not coincide. In order to investigate the convergence of FDTD, three simulated transmission curves with different cell sizes are plotted in Figure 7b for the same structure as in Figure 7a but with horizontal abscissa being $x_1^{(1)} \in [0.7, 0.8] \mu\text{m}$ to highlight two absorption peaks. A MEM curve is also plotted in the figure for comparison. For a large cell size such as $5 \times 5 \text{ nm}^2$, only one vague dip, instead of two, is observed. The dip will gradually separate into two and approach the MEM curve as the cell size decreases. However, even for $1.25 \times 1.25 \text{ nm}^2$, namely 1/800 of the incident wavelength or 1/80 of the stub width (height of Layer 2), the deviation of the results between FDTD and MEM is still observable, which means that FDTD has a relatively slow convergence. For this reason we do not use FDTD to verify the calculated results of MEM in this paper.

The transmission behaviour of the structure considered in Figure 7 was also investigated by SMM [9]. This method actually utilises some of the results of FDTD to obtain the scattering matrix elements and loses some phase information due to the quasi-statistic approximation, so that its final results could not be better than those of FDTD. This is demonstrated explicitly by the comparison between the results of FDTD and SMM given in references [8,9]. Therefore, it is not necessary to discuss the accuracy of SMM here, because it depends on the simulation results of FDTD which have already been shown in Figure 7b. Besides, since the SMM and ICIM utilise a similar transmission model, their calculation errors ought to have the same order of magnitude. We now investigate the calculation error associated with ICIM.

The SPP transmission behaviour as calculated by ICIM and MEM in a Type II structure is plotted in Figures 8 and 9. This kind of structure was also studied in references [13,14].

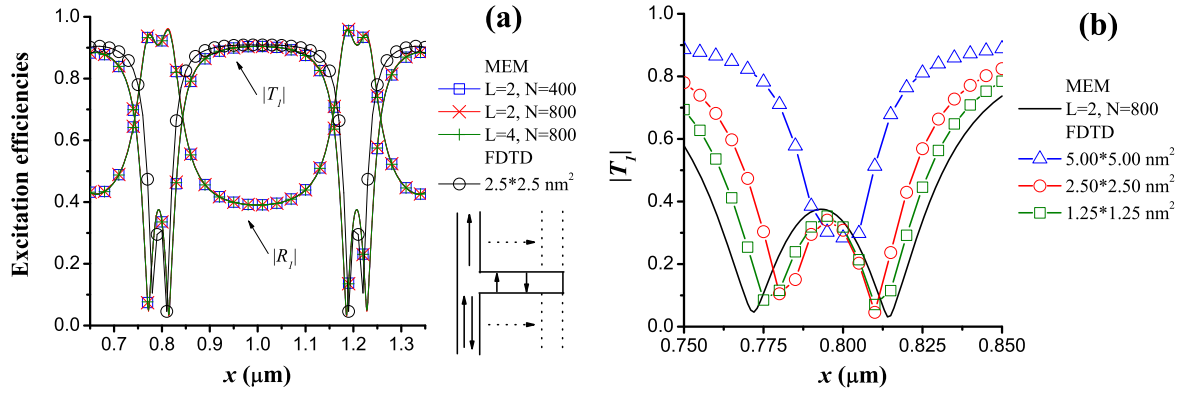


Fig. 7. (Color online) The SPP transmission and reflection in a Type I structure with parameter $x_1^{(2)} = 0.6 \mu\text{m}$, $[w^{(1)}, w^{(2)}, w^{(3)}] = [0.1, 0.8, 0.1] \mu\text{m}$, $[Q^{(0)}, Q^{(1)}, Q^{(2)}] = [-1, 0, 0.1] \mu\text{m}$, see the inset of (a). (a) Convergence test for MEM; (b) convergence test for FDTD relative to MEM where the x -axis is a part of (a) for $x_1^{(1)} \in [0.7, 0.8] \mu\text{m}$.

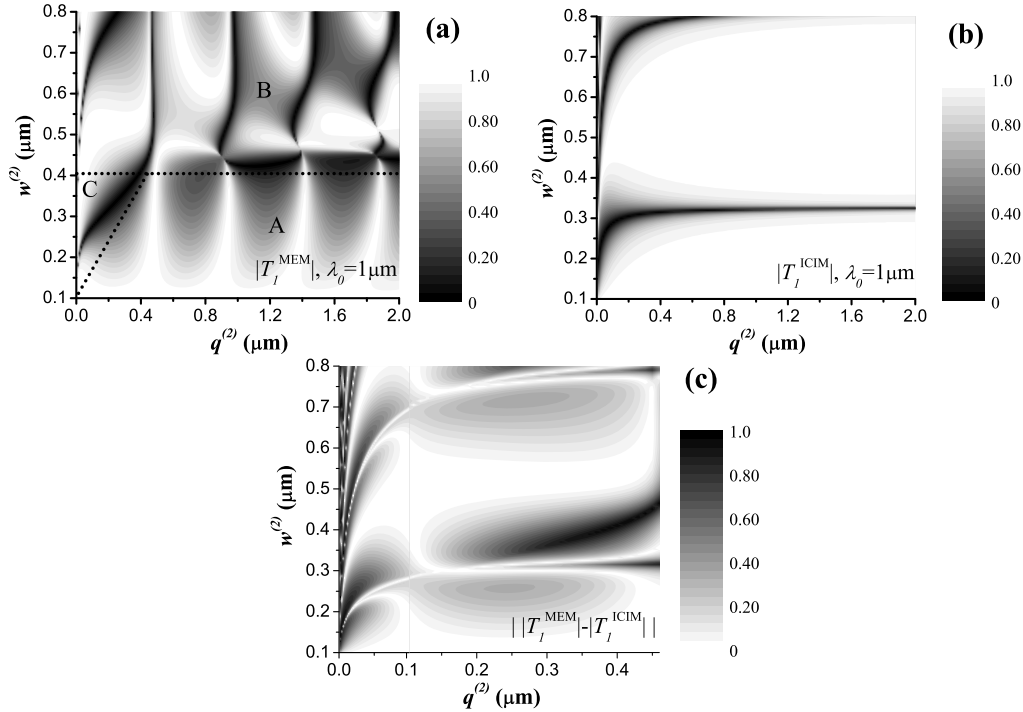


Fig. 8. Comparison of the SPP transmission by MEM and ICIM of a Type II structure under the variation of $w^{(2)}$ and $q^{(2)}$. The left sides of the slits in all layers are aligned to $x_1^{(1)} = 0.6 \mu\text{m}$. $[w^{(1)}, w^{(2)}, w^{(3)}] = [0.1, w^{(2)}, 0.1] \mu\text{m}$, $[Q^{(0)}, Q^{(1)}, Q^{(2)}] = [-1, 0, q^{(2)}] \mu\text{m}$. (a) $|T_1^{\text{MEM}}|$ by MEM; (b) $|T_1^{\text{ICIM}}|$ by ICIM; (c) absolute value of the difference between MEM and ICIM, $||T_1^{\text{MEM}}| - |T_1^{\text{ICIM}}||$ for $q^{(2)} < 0.46 \mu\text{m}$.

In Figure 8 the SPP transmission is plotted as a function of the height and width of Layer 2 under a fixed incident wavelength, $\lambda_0 = 1 \mu\text{m}$. In Figure 8a, the calculated results of MEM can be approximately divided into three areas A, B, and C by dotted lines. In Area A, a regular oscillating pattern is observed because when $w^{(2)} < 0.4 \mu\text{m}$ only one SPP mode propagates in Layer 2 and forms the FP-like oscillation [5,6]. In Areas B and C, however, where the higher modes also contribute to transmission, the transmission pattern becomes complicated. Intuitively, there are two ways for higher modes to trans-

port energy. One is by way of a propagation mode, which is appropriate for $w^{(2)} > 0.46 \mu\text{m}$ because the 2nd mode in Layer 2 becomes propagating. The other is by way of an evanescent mode, which is appropriate for $0.4 < w^{(2)} < 0.46 \mu\text{m}$ and Area C because the 2nd mode in Layer 2 does not attenuate to a negligible value in such a modulated slit and transports energy through Layer 2.

The results from ICIM shown in Figure 8b can be explained according to whether $q^{(2)}$ is larger or smaller than $0.46 \mu\text{m}$. When $q^{(2)} > 0.46 \mu\text{m}$, more than one mode is allowed to propagate in the stub while the ICIM

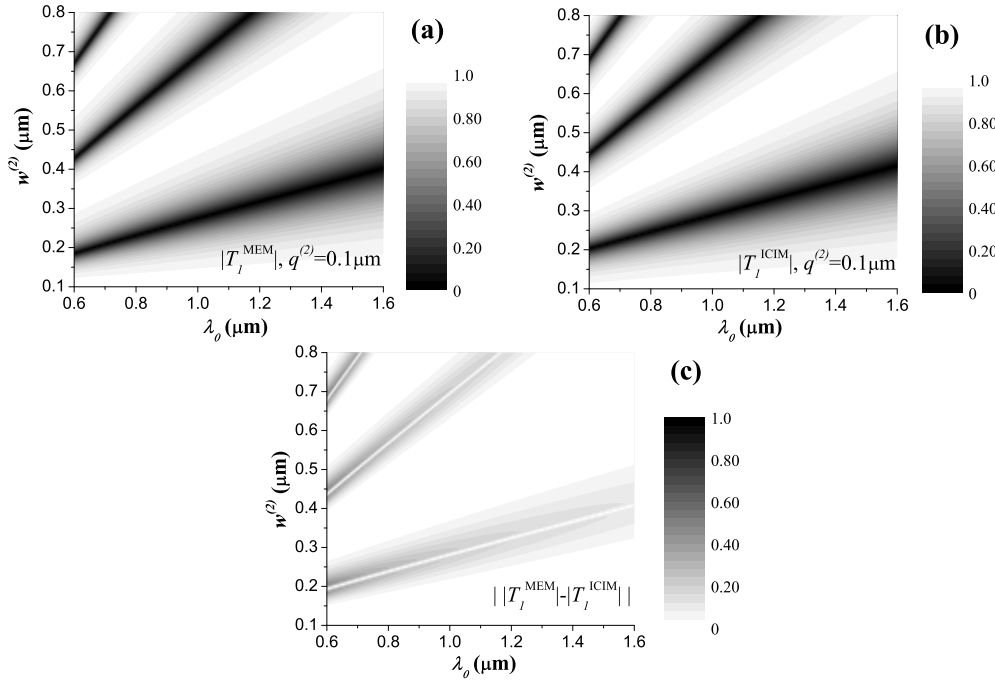


Fig. 9. Comparison of the SPP transmission by MEM and ICIM of a Type II structure under the variation of $w^{(2)}$ and wavelength λ_0 . The left sides of the slits in all layers are aligned to $x_1^{(l)} = 0.6 \mu\text{m}$. $[w^{(1)}, w^{(2)}, w^{(3)}] = [0.1, w^{(2)}, 0.1] \mu\text{m}$, $[Q^{(0)}, Q^{(1)}, Q^{(2)}] = [-1, 0, 0.1] \mu\text{m}$. (a) $|T_1^{\text{MEM}}|$ by MEM; (b) $|T_1^{\text{ICIM}}|$ by ICIM; (c) absolute value of the difference between MEM and ICIM, $||T_1^{\text{MEM}}| - |T_1^{\text{ICIM}}||$.

assumes only the SPP mode, which results in the great difference relative to the MEM results, leading to totally different patterns between Figures 8a and 8b. When $q^{(2)} < 0.46 \mu\text{m}$, the difference $||T_1^{\text{MEM}}| - |T_1^{\text{ICIM}}||$ is plotted in Figure 8c. In this region, the difference is mainly caused by the neglect of the phase shifts in the central scattering region, which indicates that these phase shifts have to be taken into account in the calculation. In Figure 8c, it is seen that when the length of Layer 2 is less than $0.03 \mu\text{m}$, i.e. less than $1/30$ of the incident wavelength, the difference becomes larger due to the energy transported by evanescent modes. This demonstrates that consideration of a very narrow region cannot guarantee accurate results. Thus, the application of the quasi-statistic approximation in a modulated metal slit requires an optimum geometry in order to provide relatively accurate results: the incident wavelength is nearly 10 times larger than the stub width.

In order to test the optimum geometry, the slit width of Layer 2 and the incident wavelength λ_0 are considered as variables and the length of Layer 2 is set as $q^{(2)} = 0.1 \mu\text{m}$, the same as the slit widths in Layers 1 and 3. The MEM results in Figure 9a exhibit a few straight black bands indicating the inverse-proportional relationship between frequency and stub width which was mentioned in references [8,9]. For ICIM, similar results are obtained in Figure 9b. The difference of these two figures is plotted in Figure 9c. Clearly, the calculation error of ICIM in this case is lower than that in Figure 8c. Although a small deviation takes place near the absorption peaks, ICIM

successfully predicts the positions of the peaks. All these numerical results confirm that the ICIM can provide relatively accurate results when the incident wavelength is around 10 times the stub width.

It is worth emphasising that each method has its own advantages in certain respects. For example, FDTD can provide a visualisation of the transmission process in the time domain and ICIM is indubitably the fastest calculation method available. Here we simply point out that these methods should be used with caution and due consideration of the convergence and accuracy.

4 Conclusion

In this paper, the MEM developed in reference [6] has been improved to make it a more practical and efficient method for handling the scattering/transmission. Using this method, the scattering in a juncture structure and the transmission inside a step-modulated slit was investigated.

Firstly, the scattering in a juncture structure was studied for two types of structural changes. For the Type I change where the widths of the two slits are fixed and the position of the narrower one can be anywhere within the wider one, the excitation efficiencies of the modes in the wider slit resemble their eigenfunctions respectively, while in the narrower slit the excitation efficiency of the SPP mode is dominant and much larger than that of the second mode. For the Type II change where the left walls of the slits are aligned and one slit gradually becomes wider, a

wood-anomaly-like drastic change of excitation efficiencies is observed when the propagation property of one mode begins to transform.

Next, the transmission inside a step-modulated slit was studied. In addition to the MEM calculation, we presented explicitly a multi-mode multi-reflection model to reveal the transmission process. The multi-mode excitation and the superposition procedure of the scatterings from all possible modes are the key parts of the model, which represent the interference and energy transfer occurring at layer boundaries. However, it was demonstrated that there exists a critical height of the modulated layer for applying the model due to the coupling between propagation modes and evanescent modes. Above this critical height, the model can provide the same result as the MEM calculation.

Finally, some commonly used methods were compared with MEM. The following useful conclusions about these methods were obtained: for a subwavelength metal slit, MEM is a versatile and fast method that can provide accurate results; FDTD has a relatively slow convergence and needs a very small Yee cell to ensure the accuracy of the simulation; ICIM incorporating the one mode and quasi-statistic approximations provides relatively accurate results when the incident wavelength is around 10 times larger than the stub width.

This work is supported by the 973 Program of China (Grant No. 2011CB301801) and the National Natural Science Foundation of China (Grant No. 10874124), and (Grant No. 11074145).

References

1. W.L. Barnes, A. Dereux, T.W. Ebbesen, *Nature* **424**, 824 (2003)
2. J.A. Dionne, L.A. Sweatlock, H.A. Atwater, *Phys. Rev. B* **73**, 035407 (2006)
3. B. Sturman, E. Podivilov, M. Gorkunov, *Phys. Rev. B* **76**, 125104 (2007)
4. A.R. Zakharian, J.V. Moloney, M. Mansuripur, *Opt. Express* **15**, 183 (2007)
5. C. Li, Y.S. Zhou, H.Y. Wang, F.H. Wang, *J. Opt. Soc. Am. B* **27**, 59 (2010)
6. C. Li, Y.S. Zhou, H.Y. Wang, F.H. Wang, *Opt. Express* **19**, 10073 (2011)
7. X.S. Lin, X.G. Huang, *Opt. Lett.* **33**, 2874 (2008)
8. X.S. Lin, X.G. Huang, *J. Opt. Soc. Am. B* **26**, 1263 (2009)
9. J. Tao, X.G. Huang, X.S. Lin, J.H. Chen, Q. Zhang, X.P. Jin, *J. Opt. Soc. Am. B* **27**, 323 (2010)
10. Q. Zhang, X. Huang, X. Lin, J. Tao, X. Jin, *Opt. Express* **17**, 7549 (2009)
11. A. Hosseini, Y. Massoud, *Opt. Express* **14**, 11318 (2006)
12. Y.S. Zhou, B.Y. Gu, H.Y. Wang, *Phys. Rev. A* **81**, 015801 (2010)
13. Y. Matsuzaki, T. Okamoto, M. Haraguchi, M. Fukui, M. Nakagaki, *Opt. Express* **16**, 16314 (2008)
14. A. Pannipitiya, I.D. Rukhlenko, M. Premaratne, H.T. Hattori, G.P. Agrawal, *Opt. Express* **18**, 6191 (2010)
15. P.B. Johnson, R.W. Christy, *Phys. Rev. B* **6**, 4370 (1972)
16. L.C. Botten, M.S. Craig, R.C. McPhedran, *Comput. Phys. Commun.* **29**, 245 (1983)
17. M.G. Moharam, E.B. Grann, D.A. Pommet, T.K. Gaylord, *J. Opt. Soc. Am. A* **12**, 1068 (1995)
18. L. Li, *J. Opt. Soc. Am. A* **13**, 1870 (1996)
19. L. Li, *J. Opt. Soc. Am. A* **13**, 1024 (1996)
20. M.G. Moharam, D.A. Pommet, E.B. Grann, T.K. Gaylord, *J. Opt. Soc. Am. A* **12**, 1077 (1995)
21. A. Hessel, A.A. Oliner, *Appl. Opt.* **4**, 1275 (1965)
22. Q. Cao, P. Lalanne, *Phys. Rev. Lett.* **88**, 057403 (2002)
23. S. Ramo, J.R. Whinnery, T. Van Duzer, *Fields and Waves in Communication Electronics* (Wiley, New York, 1994)

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.