Theory of the scattering of light and surface plasmon polaritons by finite-size subwavelength metallic defects via field decomposition

To cite this article: Guangyuan Li et al 2011 New J. Phys. 13 073045

View the article online for updates and enhancements.

Related content
- Unidirectional coupling of surface plasmons with ultra-broadband and wide-angle efficiency: potential applications in sensing
  Kun Li, Feng Xiao, Fan Lu et al.
- Analytical description of quasi-normal mode in resonant plasmonic nano cavities
  Fan Yang, Haitao Liu, Hongwei Jia et al.
- Theory on the scattering of light and surface plasmon polaritons by arrays of holes and dimples in a metal film
  F de León-Pérez, G Brucoli, F J García-Vidal et al.

Recent citations
- Experimental solution for scattered imaging of the interference of plasmonic and photonic mode waves launched by metal nano-slits
  Xing Li et al.
- Ultra-broadband and efficient surface plasmon polariton launching through metallic nanoslits of subwavelength period
  Guangyuan Li and Jiasen Zhang
- Coherent-form energy conservation relation for the elastic scattering of a guided mode in a symmetric scattering system
  Haitao Liu
Theory of the scattering of light and surface plasmon polaritons by finite-size subwavelength metallic defects via field decomposition

Guangyuan Li\textsuperscript{1,3}, Feng Xiao\textsuperscript{2}, Lin Cai\textsuperscript{1}, Kamal Alameh\textsuperscript{2} and Anshi Xu\textsuperscript{1}

\textsuperscript{1} State Key Laboratory of Advanced Optical Communication Systems and Networks, School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, People’s Republic of China
\textsuperscript{2} WA Center of Excellence for MicroPhotonic System, Electron Science Research Institute, Edith Cowan University, Joondalup, WA, 6027, Australia
E-mail: gyli2008@hotmail.com

\textit{New Journal of Physics} 13 (2011) 073045 (15pp)
Received 16 December 2010
Published 29 July 2011
Online at \url{http://www.njp.org/}
doi:10.1088/1367-2630/13/7/073045

Abstract. A theoretical model is presented for the scattering of light and surface plasmon polaritons (SPPs) by finite-size subwavelength metallic defects. Based on the decomposition of the scattered fields into SPPs and quasi-cylindrical waves (CWs), an SPP–CW model is developed to depict the multiple scattering of SPPs and CWs in finite-size defects using the elementary scattering processes in a single one. The involved elementary scattering of the CW, as well as the CW-related coefficients, which are difficult or even impossible to define and calculate according to classical scattering theory, is clarified. A close relationship between the scattering coefficients of the SPP and those of the CW has been pointed out and used to simplify the developed model. Compared to the corresponding pure SPP model and the fully vectorial computational data, the SPP–CW model is shown to be versatile and quantitatively accurate for finite-size defects such as grooves, ridges, slits or even hybrid systems of various geometrical parameters, over a broad spectral range from the visible to the thermal infrared regime.

\textsuperscript{3} Author to whom any correspondence should be addressed.
1. Introduction

Surface plasmon polaritons (SPPs) are electromagnetic surface modes confined to and propagating along the interface between a metal and a dielectric. To couple or decouple light to SPPs, finite-size subwavelength defects such as nanoholes [1], air grooves [2] or metallic ridges [3] in or on a metal film have been used widely. By proper design, they have also been used to reflect SPPs in various plasmonic devices, including the unidirectional [4, 5] or bidirectional [6] coupler and nano-cavities [7]. As the propagation length of the SPP mode along the air–metal (gold or silver) interface is tens of micrometers in the visible and near-infrared regimes, the metallic defects should be carefully designed to improve the efficiency of a specific usage. As a result, gaining a quantitative knowledge of the SPP excitation and reflection efficiencies, which is a key issue in engineering-related plasmonic devices, has attracted much interest. Until now, most studies have been carried out using numerical simulation techniques, such as the finite-element method [8], or theoretical methods such as the Green tensor approach [9–12], the Rayleigh expansion [13, 14], the modal expansion technique [15–17] and the multiple multipole method [18]. These methods, although they may be accurate, suffer from high numerical cost, especially when the number of defects is very large. What is worse is that by using these methods the underlying physics is not clear enough.

To circumvent these problems, the interactions involved in subwavelength metallic defects should be clarified first. Under the approximation that SPP fields are dominant in the scattered fields near the dielectric–metal interface, quantitative theoretical models have been developed for extraordinary optical transmission (EOT) through subwavelength hole arrays [19] and a slit surrounded by symmetric [20] or asymmetric [21] periodic grooves, and for the SPP-Bragg reflectors composed of various defects [22]. However, as SPPs are the primary vector of the interactions only at a distance of several wavelengths from a scatterer [23] and at visible frequencies [24], there are pronounced deviations between fully vectorial computational data and the predictions by models incorporating only the SPP-to-SPP scattering [19–22],
especially for scatterers of small distances or for longer wavelengths. This is because the residual field referred to as the quasi-cylindrical wave (CW), which was pointed out in [24], plays an important role near the scatter and even dominates in the long wavelength limit, as confirmed experimentally in [25]. The physical origin and the main characteristics are then clarified analytically in [26]. Recently, the cross-conversion between SPPs and CWs, specifically the CW-to-SPP conversion, was proposed in [27]; the SPP-to-CW conversion as well as the reflection and transmission of the CW was further proposed by some authors [28]. Quite recently, a hybrid-wave model was also proposed and shown to be accurate in predicting the field at the air–metal interface [29] and the EOT through the hole array [30]. However, the underlying interactions between SPPs and CWs are not clear in the hybrid-wave model. Moreover, as CWs are not normal modes (they are a superposition of modes), there are no scattering matrices for them according to classical theories [31], and one may think that it is difficult (or even impossible) to define scattering coefficients for those waves [26, 29]. As a result, it is necessary to clarify the interactions between SPPs and CWs, especially the CW-related scattering coefficients used in our previous work [28].

In this paper, we study the scattering of free-space light and SPPs by finite-size subwavelength metallic defects via field decomposition and interactions between SPPs and CWs. In section 2, we summarize the field decomposition into SPPs and CWs, and elementary scattering processes of the light, SPP and CW by a single defect. The scattering of CWs and the method for calculating the CW-related coefficients via field decomposition are clarified. Based on the field decomposition and interactions between SPPs and CWs, an SPP–CW model for the scattering of the free-space light and SPP mode by finite-size defects is then developed in section 3. The model is exemplified for SPP excitation and reflection efficiencies, as well as the scattered magnetic fields at the surface, and quantitatively validated using exhaustive calculations with various parameters by comparing it with the fully vectorial aperiodic Fourier modal method (a-FMM) [34]. Throughout this paper, the dielectric is assumed to be air with $n_d = 1.0$, and gold is used with the frequency-dependent permittivities tabulated in [32]. Without special specifications, an analysis will be performed with air grooves as the defect at $\lambda = 800 \text{ nm}$.

2. Scattering by a single-subwavelength defect

We focus first on the simplest case, the scattering of the plane wave, SPP mode and CW by a single defect, as shown in figure 1. As we restrict ourselves to the magnetic field at the surface, i.e. $H_y(x, z = 0)$, the outgoing radiations generated by the SPP incidence or CW incidence, as shown in figures 1(b) and (c), respectively, will not be considered in this paper. Although the groove is used as an example, we should emphasize that the defect may be a groove, a metallic ridge or a slit.

To understand the scattering processes and related coefficients, we first briefly introduce the concepts of the zero field and scattered field. In a scattering problem, the zero field is the field when there are no scatterers at all; the scattered field is calculated by subtracting the zero field from the total field with scatterers. For example, when a plane wave is incident on a groove, as shown in figure 1(a), the zero field is the field composed of the incidence and reflection in the dielectric–metal bilayer structure, which can be calculated analytically.
2.1. The field decomposition into surface plasmon polaritons (SPPs) and cylindrical waves (CWs)

When a groove of width \( w \) and the center in the \( x \)-direction \( x_j \) is illuminated by a normally incident TM-polarized plane wave of unity Poynting vector, as shown in figure 1(a), the scattered magnetic field \( H^\text{sc}_y(x) \) launched at the surface on every side of the groove can be accurately described as a combination of an SPP and a CW [27],

\[
H^\text{sc}_y(|x - x_j| > w/2) = \beta_\text{sp} u(|x - x_j|) + \beta_\text{cw} v(|x - x_j|),
\]

where \( \beta_\text{sp} \) and \( \beta_\text{cw} \) are the excitation coefficients of the SPP and CW, respectively, \( u(x) = \exp(ik_0n_{\text{sp}}x) \), \( k_0 = 2\pi/\lambda \), \( n_{\text{sp}} = n_d n_m / \sqrt{n_d^2 + n_m^2} \) is the effective refractive index of the SPP mode along the dielectric–metal interface, with \( n_d \) being the refractive index of the dielectric and \( n_m \) being that of the metal, and \( v(x) \approx W(2\pi\gamma|x|/\lambda) / W(2\pi\gamma)(\lambda/|x|)^{3/2} \exp(ik_0n_{\text{sp}}x) \) with \( W(t) = -2\sqrt{-i/\pi} \int_{-\infty}^{\infty} z^2 \exp(-z^2)/(z - \sqrt{-i})dz \) and \( \gamma = n_{\text{sp}} - n_d \) [26]. Note that here we use the scattered fields instead of the total fields since SPPs do not occur in the zero fields but show up in the scattered fields [33].

For the scattering of an SPP mode by a single groove, as shown in figure 1(b), the zero field is exactly the incident SPP mode, which is expressed as \( H_{\text{ref}}^\text{in}_{\text{SPPin}} = H_0 u(x - x_i) = H_1 u(x - x_i) \), with \( x_i < x_j - w/2 \) and \( H_1 = H_0 u(x_j - x_i) \) being its complex amplitudes at \( x = x_j \). Similarly to the scattering of the plane wave, the generated forward-going and backward-going fields (i.e. the scattered fields) can also be decomposed into SPPs and CWs. The reflected field (i.e. the transmitted field) is the sum of the zero field and the generated forward-going field) at the surface on every side of the groove are expressed as

\[
H_{\text{ref}}^\text{trans}_{\text{SPPin}}(x < x_j - w/2) = r_{\text{sp}} H_1 u(x_j - x) + r_{\text{cw}}\text{ref} H_1 v(x_j - x), \tag{2a}
\]

\[
H_{\text{ref}}^\text{trans}_{\text{SPPin}}(x > x_j + w/2) = t_{\text{sp}} H_1 u(x_j - x) + t_{\text{cw}}\text{ref} H_1 v(x_j - x), \tag{2b}
\]

respectively, where \( r_{\text{sp}} \) and \( t_{\text{sp}} \) are the SPP reflection and transmission coefficients, and \( r_{\text{cw}}\text{ref} \) and \( t_{\text{cw}}\text{ref} \) are defined as cross conversion coefficients associated with the scattering of the incident
SPP into the generated backward- and forward-going CWs. Note that as the transmitted SPP is the total field, the generated forward-going SPP mode (i.e. the scattered field) is expressed as $(t_{sp} - 1)H_t u(x - x_j)$. In other words, the corresponding forward-going SPP-generation coefficient is $t_{sp} - 1$.

For the scattering of a CW by a single groove, as shown in figure 1(c), the zero field is exactly the incident CW originating from $x_i$ and expressed as $H_{y, \text{CWin}}^\text{in}(x) = H_2 v(x - x_j)$ at the surface with $x_j < x_j - w/2$. The generated forward-going and backward-going fields (i.e. the scattered fields) at the surface on every side of the groove can also be decomposed into SPPs and CWs,

$$H_{y, \text{CWin}}^{\text{sc}, \text{ref}}(x < x_j - w/2) = r_{sp}^c H_3 u(x_j - x) + r_{cw}^c H_3 v(x_j - x), \quad (3a)$$

$$H_{y, \text{CWin}}^{\text{sc}, \text{trans}}(x > x_j + w/2) = t_{sp}^c H_3 u(x - x_j) + t_{cw}^c H_3 v(x - x_j), \quad (3b)$$

where $H_3 = H_2 v(x_j - x_j)$, $r_{sp}^c$ and $t_{sp}^c$ (denoted as $r_c$ and $t_c$ in [27]) are cross conversion coefficients associated with the scattering of the incident CW into the backward- and forward-going SPPs, and $r_{cw}$ and $t_{cw}$ are the backward- and forward-going CW-generation coefficients, respectively. Note that the scattering coefficients of the SPP and CW, i.e. $r_{sp}, t_{sp}, r_{cw}, t_{cw}, r_{cw}^c, t_{cw}^c$, are independent of $x_j$.

As the SPP is a normal mode, SPP-related coefficients, i.e. $\beta_{sp}, r_{sp}$ and $t_{sp}$, are easy to calculate according to the mode orthogonality (the details can be found in the appendix). Although CW-related coefficients cannot be calculated similarly, as the CW is not a normal mode, they are obtained through the field decomposition. By setting $|x - x_j| = \lambda$ in (1)–(3) as $v(\lambda) = 1$, $\beta_{cw}, r_{cw}^c$ and $t_{cw}^c$, $r_{cw}$ and $t_{cw}$ are then obtained, respectively. It is clear that the definitions of $r_{cw}^c, t_{cw}^c$ and CW-related coefficients are quite different from those of $\beta_{sp}, r_{sp}$ and $t_{sp}$, which are defined according to the scattering theory and reciprocity [31].

Figure 2 shows the decomposition of the total scattered fields into SPPs and CWs. In this figure, the total scattered fields (red solid lines) are calculated using the a-FMM, the SPP fields (as well as SPP-related coefficients) are calculated according to the mode orthogonality, and the residual fields (CWs in black lines) are then extracted by subtracting SPPs from the total scattered fields. By setting $|x - x_j| = \lambda$, CW-related coefficients are obtained, by using which the generated CWs for the light incidence, SPP incidence and CW incidence are fitted very well (in dots) with (1)–(3), respectively. In other words, the CW-related coefficients calculated via field decomposition are very accurate.

We should emphasize that $t_{cw}$ is quite different from the SPP transmission coefficient $t_{sp}$. The reason for this difference is that the transmitted SPP can be easily expressed in one term that includes the contributions of both the incident and the generated forward-going SPPs, whereas the transmitted CW should be expressed in the form of the sum of the incident and the generated forward-going CWs,

$$H_{y, \text{CWin}}^{\text{trans}}(x > x_j + w/2) = H_{y, \text{CWin}}^{\text{in}}(x) + H_{y, \text{CWin}}^{\text{sc}, \text{trans}}(x) = H_2 v(x - x_j) + t_{cw} H_3 v(x - x_j). \quad (4)$$

Equation (4) cannot be incorporated in one term because of the complex propagation characteristics of the CW.
Figure 2. The magnetic fields at the surface are decomposed into SPPs and CWs. SPPs (blue dashed lines) are obtained using the mode orthogonality, and CWs (black solid lines) are then extracted by subtracting SPPs from the total scattered fields (red solid lines). Green dots show the CWs fitted using CW-related coefficients obtained by setting \(|x-x_j|=\lambda\). ‘Re’ means the real part and ‘Im’ means the imaginary part. Panels (a) and (b) are for the plane wave incidence, (c) is for the SPP incidence and (d) for the CW incidence. Note that for the sake of clarity, the incident SPP is incorporated into the region of \(x<0\) in (c), as it has also been included into the transmitted SPP \((x>0)\). The vertical dot-dashed line indicates the groove region with \(w=0.2\lambda\), \(h=0.2\lambda\), \(x_j=0\).

2.2. Relationship between the scattering coefficients

We note that there is a close relationship between the scattering coefficients of SPPs and CWs,

\[
\begin{align*}
    r_{cw} &\approx r_{sp}, & t_{cw} &\approx t_{sp} - 1, \\
    r_{cw} &\approx r_{sp}^{cw}, & t_{cw} &\approx t_{sp}^{cw},
\end{align*}
\]

where (5a) is exactly the same as (5) in [27]. Equation (5) has been thoroughly validated for various geometries (grooves, metallic ridges and slits) over a broad spectral range from the visible to the thermal infrared regime, as illustrated in figure 3. This approximation is very useful since it eases the calculations of the scattering coefficients considerably. One needs to calculate four coefficients instead of eight for the elementary scattering processes of the SPP and the CW by a single groove. As the CW incidence is slightly more complex than the SPP incidence in the a-FMM, we suggest calculating the four scattering coefficients of the SPP mode, i.e. \(r_{sp}, t_{sp}, r_{sp}^{cw}\) and \(t_{sp}^{cw}\).

This approximation can be explained by the causality principle, just as was done in [27]. The forward- and backward-going SPP-generation (or CW-generation) coefficients for the CW incidence should be equal to their counterparts for the SPP incidence. In other words, the generated SPPs and CWs (i.e. the scattered fields) are identical for the CW incidence and SPP
3. Scattering by finite-size subwavelength defects

In this section, we study the scattering of the plane wave or the SPP mode by finite-size subwavelength metallic defects, as illustrated in figure 4. We should emphasize that the defects may be groove systems, metallic ridge systems, slit systems or a hybrid system composed of the above defects. Based on the elementary scattering processes shown in figure 1, an intuitive SPP–CW model will be developed to present a useful microscopic picture of the physics and for the engineering of the SPP excitation, reflection and transmission involving multiple scattering by finite-size defects. For comparison, a pure SPP model that is reduced from the SPP–CW model by ignoring CWs will also be developed. Exhaustive comparisons among the SPP–CW model predictions, the pure SPP model predictions and the fully vectorial a-FMM computational data will be made in order to validate the quantitative accuracy of the developed model, demonstrating that the SPP–CW model is accurate and versatile to depict the interactions among the plane wave, SPPs and CWs in finite-size defects.

Figure 3. The behaviors of the four coefficient pairs \((r_{sp}, r_{sp}^{cw})\), \((t_{sp} - 1, t_{sp}^{cw})\), \((r_{cw}, r_{cw})\) and \((t_{cw}, t_{cw})\) as functions of the operating wavelength with \(w = 0.2\lambda\), \(h = 0.2\lambda\) (a)–(d) and of the groove width with \(h = 0.4\lambda\), \(\lambda = 0.8 \mu\text{m}\) (e)–(h). Blue lines and symbols are for the real parts, and red for the imaginary.

incidence as SPP and CW share many properties [27]. Note that here we use the generation coefficients instead of the transmission coefficients so that the zero fields are excluded for both the SPP incidence and the CW incidence. With these in mind, it is natural to understand the term \(t_{sp} - 1\) instead of \(t_{sp}\) in (5a). We should emphasize that \(r_{cw}\) (or \(t_{cw}\)) may not be equal to \(r_{sp}^{cw}\) (or \(t_{sp}^{cw}\)). This does not violate the reciprocity according to the classic scattering theory [31] as CWs are not normal modes.
3.1. The SPP–CW model

Let us consider the scattering of the normal incident TM-polarized free-space light of unity Poynting vector by \( N \) defects, as illustrated in figure 4(a). According to the method of field decomposition, the free-space light will generate left- and right-going SPPs and CWs at each defect, which will then be further scattered by other defects. Based on the elementary scattering processes shown in figure 1, we develop an SPP–CW model to clearly depict the involved multiple scattering processes. In the model, \( A_j \) and \( B_j \) are the respective complex amplitudes of the right- and left-going SPPs propagating away from defect ‘\( j \)’, while \( C_j \) and \( D_j \) are those of CWs generated by the defect ‘\( j \)’ \( (j = 1, 2, \ldots, N) \). Note that the definitions of \( A_j \) (or \( B_j \)) and \( C_j \) (or \( D_j \)) are quite different. This difference is due to different propagation characteristics of SPPs and CWs, as has been clarified in (4). The multiple scattering of SPPs and CWs are easily embodied in the following coupled-mode equations,

\[
A_j = \beta \sp{sp}_{j} + t \sp{sp}_{j} U_{j-1} A_{j-1} + r \sp{sp}_{j} U_{j} B_{j+1} + r \sp{cw}_{j} \sum_{m=1}^{j-1} C_{m} V_{m,m} + C_{j} V_{m,j} + \sum_{m=j+1}^{N} D_{m} V_{m,j}, \quad (6a)
\]

\[
B_j = \beta \sp{sp}_{j} + r \sp{sp}_{j} U_{j-1} A_{j-1} + t \sp{sp}_{j} U_{j} B_{j+1} + r \sp{cw}_{j} \sum_{m=1}^{j-1} C_{m} V_{m,m} + C_{j} V_{m,j} + \sum_{m=j+1}^{N} D_{m} V_{m,j}, \quad (6b)
\]

\[
C_j = \beta \sp{cw}_{j} + t \sp{sp}_{j} U_{j-1} A_{j-1} + t \sp{sp}_{j} U_{j} B_{j+1} + r \sp{cw}_{j} \sum_{m=1}^{j-1} C_{m} V_{m,m} + C_{j} V_{m,j} + \sum_{m=j+1}^{N} D_{m} V_{m,j}, \quad (6c)
\]

\[
D_j = \beta \sp{cw}_{j} + r \sp{sp}_{j} U_{j-1} A_{j-1} + r \sp{sp}_{j} U_{j} B_{j+1} + r \sp{cw}_{j} \sum_{m=1}^{j-1} C_{m} V_{m,m} + C_{j} V_{m,j} + \sum_{m=j+1}^{N} D_{m} V_{m,j}, \quad (6d)
\]
where \( U_j = u(x_{j+1} - x_j) \) with \( U_0 = U_N = 0 \), \( V_{j,m} = v(x_j - x_m) \), \( V_{m,j} = v(x_m - x_j) \), \( j = 1, \ldots, N \) with \( A_0 = C_0 = B_{N+1} = D_{N+1} = 0 \). The number of coefficients and the complexity of the model are reduced taking into account the approximation in (5).

The right-hand side of (6a) is interpreted as follows. The first term corresponds to the SPP excitation by the plane wave at defect ‘\( j \)’. The second and third terms mean the transmission and reflection of SPPs propagating away from neighboring defects on the left and right sides, respectively. The fourth (fifth) term is associated with the CW-to-SPP conversion for all the right-going (left-going) CWs generated by all the defects on the left (right) side. Other equations can be understood similarly.

Solving the derived \( 4N \) linear equations is routine and of negligible numerical cost for a practical \( N \) (usually smaller than 30). Once \( A_j, B_j, C_j, \) and \( D_j \) \((j = 1, \ldots, N)\) are obtained, the SPP excitation coefficient is then expressed as \( \beta_N = A_N \) (the SPP excitation efficiency is \( |\beta_N|^2 \)).

The scattered magnetic fields at the surface in the region \( x_{j-1} + w_{j-1}/2 < x < x_j - w_j/2 \) are then expressed as

\[
H_{x}^w(x) = A_{j-1}u(x - x_{j-1}) + B_{j}u(x_j - x) + \sum_{m=1}^{j-1} C_m v(x - x_m) + \sum_{m=j}^{N} D_m v(x_m - x),
\]

where \( 1 \leq j \leq N + 1 \) with \( x_0 + w_0/2 = -\infty \) and \( x_{N+1} + w_{N+1}/2 = \infty \).

For the scattering of the SPP mode by \( N \) defects, as shown in figure 4(b), the incident SPP mode is assumed to be of unity amplitude at \( x = x_1 \), i.e. \( A_0 = 1 \). The corresponding SPP–CW model is developed similarly to the scattering of the plane wave, with \( A_j, B_j, C_j \) and \( D_j \) \((j = 1, 2, \ldots, N)\) of the same meaning as their counterparts in figure 4(a). The coupled-mode equations are obtained by simply deleting \( \beta_{sp,j}, \beta_{cw,j} \), and setting \( U_0 = A_0 = 1 \) in (6). The SPP reflection and transmission coefficients are then expressed as \( r_N = B_1, t_N = A_N \) (the reflectance \( R_N = |r_N|^2 \) and the transmittance \( T_N = |t_N|^2 \)), and the fields at the surface are expressed similarly to (7).

Note that in the SPP–CW model, the contributions of defect ‘\( j \)’, including the influences of the defect’s shape and size, are inherently embodied by a set of coefficients, i.e. \( \beta_{sp,j}, \beta_{cw,j}, r_{sp,j}, t_{sp,j}, r_{cw,j}, t_{cw,j} \). As a result, the developed SPP–CW model is applicable not only to groove systems, but also to metallic ridge systems slit systems and even hybrid systems.

With the SPP–CW model, the involved multiple scattering of SPPs and CWs is clearly depicted via the field decomposition and the elementary scattering processes. The excitation, reflection and transmission of the SPP by \( N \) defects, as well the fields at the surface, are obtained starting from the involved elementary coefficients of each defect. In this way, the computational cost of \( \beta_N, r_N \) and \( t_N \) for \( N \) defects is reduced to that for a single one, because that of solving the \( 4N \) linear equations is negligible.

3.2. The pure SPP model

For comparison, we also developed a pure SPP model in the form of \( 2N \) linear equations simply by ignoring CWs in the corresponding SPP–CW model. For example, (6) for the SPP–CW model is reduced to (6a) and (6b) with \( C_j = D_j = 0 \) for the pure SPP model. Especially, if the defects are periodic with period \( p \) \((p = x_{j+1} - x_j \) for \( j = 1, \ldots, N - 1)\), the pure SPP model
can also be expressed in recursive forms:

$$\beta_N = \beta_{N-1} + \tau_{N-1} U \frac{\beta_1 + \beta_{N-1} \rho_1 U}{1 - \rho_{N-1} \rho_1 U^2},$$

$$\rho_N = \rho_1 + \rho_{N-1} \frac{\tau_1 U^2}{1 - \rho_{N-1} \rho_1 U^2},$$

$$\tau_N = \frac{\tau_{N-1} \tau_1 U}{1 - \rho_{N-1} \rho_1 U^2},$$

where $U = u(p)$, $\beta_1 = \beta_{\text{sp},1}$, $\rho_1 = \rho_{\text{sp},1}$ and $\tau_1 = \tau_{\text{sp},1}$ are the SPP excitation, reflection and transmission coefficients by a single defect, respectively, and $\beta_N$, $\rho_N$, $\tau_N$ are those by $N$ periodic ones. For the deduction of (8), see our previous work on the SPP-Bragg reflector [22], where the equivalence between the linear equation forms and the recursive forms has also been demonstrated. Note that we use $U = \exp(i k_0 n_{\text{sp}} p)$ here instead of $\exp[i k_0 n_{\text{sp}} (p - w)]$ in [22]. This is because the factor $\exp(-i k_0 n_{\text{sp}} w)$ has been incorporated in $r_{\text{sp}}$ and $t_{\text{sp}}$ used here, but not in [22] (see the appendix for details).

3.3. Model validations

To validate the quantitative accuracy and the versatility of the developed SPP–CW model, exhaustive calculations will be performed by varying the defect’s geometry and the operating wavelength. For the sake of readability, we use periodic grooves ($w = w_j$, $h = h_j$ for $j = 1, \ldots, N$) and randomly select some geometrical parameters for the SPP excitation, reflection and transmission coefficients, focusing on the model’s accuracy, and a hybrid subwavelength...
Figure 6. Comparison of (a, d, g) the pure SPP model predictions, (b, e, h) the SPP–CW model predictions, and (c, f, i) the a-FMM computational data on (a)–(c) $|\beta_{10}|$, (d)–(f) $|r_{10}|$ and (g)–(i) $|t_{10}|$ as functions of the groove width $w$ and height $h$. The calculations were performed with $p = \lambda$, $N = 10$.

ridge–groove doublet of various sizes for the scattered fields at the surface, emphasizing the model’s versatility. We should emphasize that this work is intended not to employ the optimized geometrical parameters, but to quantitatively validate the theoretical models, which will pave the way for the structure design and optimization with a clear physical insight as well as a great reduction of the computational cost.

Figure 5 compares the performances of the SPP–CW and pure SPP models. As clearly seen from the figure, the SPP–CW model predicts the SPP excitation and reflection coefficients with a high accuracy, in both amplitudes and phases, even when the periods are very small, whereas the deviations of the pure SPP model predictions from the a-FMM data decrease as the periods increase. This is because for small periods, the generated CW field that is equal to (for the visible regime) or much larger than (for the long wavelength limit) the generated SPP field cannot be neglected, whereas for large periods, the generated SPP largely dominates in the visible and near-infrared regimes. Furthermore, the optimized periods for the peak SPP excitation and reflection coefficients are well predicted by the pure SPP and SPP–CW models, indicating that the pure SPP model is accurate enough to predict the optimal periods, although there are some deviations in the peak values.

The SPP–CW and pure SPP models are capable of capturing all the salient features of the fully vectorial a-FMM computational data, as illustrated in figure 6. There are some
deviations (mainly in values) for the pure SPP model, while the SPP–CW model predicts with a high accuracy on $\beta_N$, $r_N$ and $t_N$ even when $N$ is very large, as will be further discussed later. It is clear that the SPP–CW model is accurate even when the groove is very wide and very deep.

The deviations between the pure SPP model predictions and the a-FMM data become pronounced for a large period number or in the long wavelength limit, whereas the SPP–CW model is accurate for various $N$ over a broad spectral range (from the visible to the thermal infrared), as illustrated in figure 7. We emphasize that the phases of $\beta_N$, $r_N$ and $t_N$ are also accurately predicted by the SPP–CW model, but are not shown due to the space limitations.

Although the accuracy of the SPP–CW model is exemplified for periodic grooves, the model is also applicable to aperiodic grooves, ridges, slits or even hybrid systems, as we have emphasized in context. Here we check the versatility of the SPP–CW model using a subwavelength ridge–groove doublet as an example, which is illuminated by the normal incident TM-polarized free-space light, as shown in figure 8. We compare the scattered magnetic fields at the surface calculated directly by the a-FMM and those obtained using the SPP–CW model with (6) and (7). It is evident that for the ridge–groove doublets of various parameters, the scattered fields are very accurately predicted by the SPP–CW model.

Figure 7. Comparison of the a-FMM computational data (red lines), the pure SPP model predictions (blue circles) and the SPP–CW model predictions (green dots) on (a) $|\beta_N|$ and (b) $|r_N|$ with $\lambda = 800$ nm as functions of $N$, (c) $|\beta_{10}|$ and (d) $|r_{10}|$ with $N = 10$ as functions of $\lambda$ from the visible ($\lambda = 0.5 \mu$m) to the thermal infrared regime ($\lambda = 10 \mu$m). The calculations were performed with $w = 0.7\lambda$, $h = 0.3\lambda$ and $p = \lambda$. 

New Journal of Physics 13 (2011) 073045 (http://www.njp.org/)
Figure 8. Comparison of the a-FMM computational data (red lines) and the SPP–CW model predictions (green circles) on the scattered magnetic fields at the surface ($z = 0$) for the structure composed of a metallic ridge (of width $w_R$ and height $h_R$) and an air groove (of width $w_G$ and height $h_G$). The calculations were performed with (a) $w_R = h_R = w_G = h_G = 0.1\lambda$, (b) $0.2\lambda$, (c) $0.6\lambda$ and $d = 2\lambda$. Blue dashed lines indicate the ridge and the groove regions.

4. Conclusions

In conclusion, the SPP–CW model has been developed for the scattering of light and SPPs by finite-size subwavelength metallic defects based on the decomposition of the scattered fields into SPPs and CWs. Via the field decomposition, the scattering of CW was clarified, and the CW-related coefficients, which were believed to be difficult or even impossible to define and calculate according to classical scattering theory, were obtained. A close relationship between the scattering coefficients of the SPP and those of the CW was pointed out and used to simplify the developed model. The multiple scattering of SPPs and CWs in $N$ defects is clearly depicted using coupled-mode equations and the involved elementary scattering processes in a single one. By comparison with the corresponding pure SPP model and the fully vectorial a-FMM computational data, the quantitative SPP–CW model was shown to be versatile and highly accurate for various parameters, including the periods, the defects’ shapes, sizes and numbers and the operating wavelengths (from the visible to the thermal infrared regime). We believe that the SPP–CW model is useful to pave the way for structure design and optimization with a clear physical insight as well as a great reduction of the computational cost.
Acknowledgments

GL thanks Professor P Lalanne for fruitful discussions. This work was supported by the National Natural Science Foundation of China under grant no. 60772002 and the State Key Laboratory of Advanced Optical Communication Systems and Networks, China.

Appendix. Calculation of SPP-related coefficients

For the calculation of SPP-related coefficients, here we use \( t_{sp} \) as an example; other coefficients, i.e. \( \beta_{sp} \) (which has also been discussed in [26]), \( r_{sp} \), \( r_{cw}^{sp} \) and \( t_{cw}^{sp} \), can be calculated similarly. For the sake of simplicity, we assume that \( x_j = 0 \).

As the right-going SPP mode is expressed as (with \( E_{x,SP}(x, z) \) omitted)

\[
[H^+_{y,SP}(x, z), E^+_{z,SP}(x, z)] = \exp(ik_0\gamma^+_d z)[1, -n_{sp}/n_d^2] \quad \text{for} \quad z \geq 0,
\]

\[
\exp(-ik_0\gamma^+_m z)[1, -n_{sp}/n_m^2] \quad \text{for} \quad z \leq 0,
\]

where \( \gamma^+_d = \sqrt{n_d^2 - n_{sp}^2}, \gamma^+_m = \sqrt{n_m^2 - n_{sp}^2} \), and the left-going one is

\[
[H^-_{y,SP}(x, z), E^-_{z,SP}(x, z)] = \exp(-ik_0n_{sp}x) \left\{ \begin{array}{ll}
\exp(ik_0\gamma^-_d z)[1, n_{sp}/n_d^2] & \text{for} \quad z \geq 0, \\
\exp(-ik_0\gamma^-_m z)[1, n_{sp}/n_m^2] & \text{for} \quad z \leq 0,
\end{array} \right.
\]

the total transmitted field for the SPP incidence, i.e. (2b), is then rewritten as

\[
H_y(x > w/2, z) = t_{sp}H_1H^+_y,SP(x, z) + t_{cw}^spH_1H^+_y,cw(x, z),
\]

\[
E_z(x > w/2, z) = t_{sp}H_1E^+_z,SP(x, z) + t_{cw}^spH_1E^+_z,cw(x, z).
\]

The transmission coefficient is calculated by using the mode orthogonality [35]

\[
t_{0,sp}(x) = (H_1N_{SP})^{-1} \int_{-\infty}^{\infty} H_y(x, z)E^+_{z,SP}(x, z) - E_z(x, z)H^-_{y,SP}(x, z) \, dz,
\]

where the normalized constant \( N_{SP} \) is given by

\[
N_{SP} = \int_{-\infty}^{\infty} [H^+_y,SP(x, z)E^+_{z,SP}(x, z) - E^+_z,SP(x, z)H^-_{y,SP}(x, z)] \, dz,
\]

and can be further written as \( N_{SP} = (i/k_0)(n_d^4 - n_{sp}^4)/(n_m^4n_d^4) \) [26].

Note that the transmitted coefficient \( t_{sp} \) used here is defined as the ratio of the complex amplitude of the generated SPP mode starting from the defect’s center to that of the incident one (i.e. \( H_1 \)), \( t_{sp} = t_{0,sp}(w/2)\exp(-ik_0n_{sp}w/2) \). However, \( t_{sp} \) used in [22] (labeled as \( t'_{sp} \) here for clarity) is defined as the ratio of the complex amplitude of the generated SPP mode starting from the defect’s right edge to that of the incident SPP mode starting from the defect’s left edge (i.e. \( H_1\exp(-ik_0n_{sp}w/2) \), \( t'_{sp} = t_{0,sp}(w/2)/\exp(-ik_0n_{sp}w/2) = t_{sp}\exp(ik_0n_{sp}w) \). As a result, \( t'_{sp}\exp(ik_0n_{sp}(p - w)) = t_{sp}\exp(ik_0n_{sp}p) \), making it easy to understand that \( \exp(ik_0n_{sp}(p - w)) \) was used as \( U \) in [22], whereas \( \exp(ik_0n_{sp}p) \) is used here. As an example, \( t_{sp} = 0.8062 - 0.1925i \) for a groove with \( w = h = 0.2\lambda \) on the gold film (\( n_m = 0.181 + 5.118i \) at \( \lambda = 800 \text{ nm} \).

New Journal of Physics 13 (2011) 073045 (http://www.njp.org/)
References


New Journal of Physics 13 (2011) 073045 (http://www.njp.org/)