1-1-2012

Theory, figures of merit, and design recipe of the plasmonic structure composed of a nano-slit aperture surrounded by surface corrugations

Guangyuan Li

Feng Xiao
Edith Cowan University

Kun Li

Kamal Alameh
Edith Cowan University

Anshi Xu

10.1109/JLT.2012.2199737

Theory, Figures of Merit, and Design Recipe of the Plasmonic Structure Composed of a Nano-Slit Aperture Surrounded by Surface Corrugations

Guangyuan Li, *Member, IEEE, Member, OSA*, Feng Xiao, *Member, IEEE, Member, OSA*, Kun Li, Kamal Alameh, *Member, IEEE, Member, OSA*, and Anshi Xu, *Senior Member, IEEE, OSA*

***Abstract*—We theoretically investigate a widely-used plasmonic structure composed of a nano-slit aperture surrounded by surface corrugations. A systematical semi-analytical theory in form of two nested coupled-mode models is developed to provide intuitive physical pictures. Based on the theory, figures of merit (FoMs) of the structures designed for normal and for oblique incidence/beaming are defined for the first time to incorporate the interlinks among key structural parameters, making global optimization simple and efficient. Both the theory and the FoMs are quantitatively validated with exhaustive calculations and shown to be highly accurate on performance prediction and structural optimization. With the theory and the FoMs, an efficient, effective and standard recipe is introduced for optimal structure design. We believe this work will help to understand the mechanisms of and to facilitate the design of such a structure in various configurations used in various applications.**

***Index Terms*—Aperture antennas, coupled mode analysis, design methodology, modeling, nanoscale devices, optical surface waves, performance evaluation, plasmons.**

I. Introduction

PLASMONICS, as a new field of science and technology that exploits the unique optical properties of metallic nanostructures to manipulate light at nanometre length scales, has aroused increasing interest in fundamental science and device applications during the last decade [1]. One of its most successful applications is the extreme light concentration with various plasmonic antennas. Among these, a plasmonic structure composed of a subwavelength aperture flanked by surface corrugations, which is referred to as three-dimensional bull's eye structure or two-dimensional slit-gratings structure [2], [3] as shown in Fig. 1, has received great attentions because the optical transmittance through it can be largely tailored with respect to that of an isolated aperture [4]. With such a structure, we witnessed many promising applications including plasmon-enhanced photodetectors [5]–[8], sensitive darkfield detection and imaging with bright background suppression (referred to as SWEDA microscopy) [9], [10], plasmonic photon sorters for spectral and polarimetric imaging [11], a plasmonic visible nanosource that could be applied in nanolithography or optical data storage [12], optical beaming [13]–[19], and laser beam collimation [20], [21].

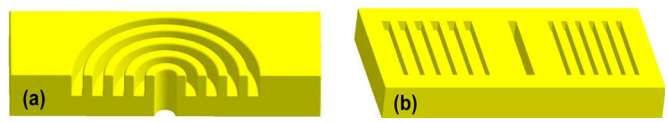

Fig. 1. Schematics of a plasmonic structure composed of a subwavelength aperture flanked by surface corrugations. (a) Three-dimensional bull's eye structure. (b) Two-dimensional slit-grating structure.

To achieve high performance for specific applications with such a structure, great efforts, both theoretical and experimental, have been put on the physical mechanisms and the parameter optimization [2], [4], [15]–[17], [20]–[33]. In most of the previous works, the influence analysis of parameters on the performance was performed by varying only one or two parameters in sequence while keeping the others fixed [2], [4], [16], [17], [22], [23], [26]–[29]. Currently there is no generalized model to understand and design the structures for various applications such as optical concentration, optical beaming and collimation, with normal or oblique incidence/beaming. The design recipe varies according to different applications and different structures, making the device design very complex and empirical. For example, it has been suggested that the optimal groove width should be around half the period [2], [27], [29], and the optimal slit-groove distance should be just over half the surface plasmon polariton (SPP) wavelength $\lambda_{\rm sp}$ [28] or be approximate to $m\lambda_{\rm sp}/2$ with m being an integer [4]. However, empirical recipes are not always applicable. A vivid example is that the optimized aperture-groove distance, which may lead to enhanced or suppressed transmission, has been under debate [28], [31] until recently. Moreover, it is accepted that the effects of many geometrical features are interlinked [28]–[30]. These properties greatly increase the complexity of structure design and aroused a global optimization by varying all the important structural parameters [30], which is very time-consuming and of excessive computational cost. To circumvent these problems, recently some authors developed semi-analytical models on optical transmission through the slit-gratings

Manuscript received October 07, 2011; revised April 04, 2012; accepted May 09, 2012. Date of publication May 16, 2012; date of current version June 20, 2012. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61107065, in part by the China Postdoctoral Science Foundation, and in part by the State Key Laboratory of Advanced Optical Communication Systems and Networks, China.

G. Li, K. Li, and A. Xu are with State Key Laboratory of Advanced Optical Communication Systems and Networks, School of Electronics Engineering and Computer Science, Peking University, Beijing, 100871, China (e-mail: gyli_2008@hotmail.com).

F. Xiao and K. Alameh are with WA Center of Excellence for MicroPhotonic Systems, Electron Science Research Institute, Edith Cowan University, Joondalup, WA 6027, Australia.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JLT.2012.2199737

structures under normal [32] or oblique [33] incidence. In these models, closed-form expressions incorporating the dominant structural influences have been provided, and the horizontal Fabry-Perot (F-P) resonance effect due to the reflection of gratings has been unveiled for symmetrical structure under normal incidence. As a result, we largely resolved the debate on the optimal slit-groove distance, and also suggested a better choice of the groove duty cycle than the choice made in [28] for a special case. However, there are still many problems left to solve. Questions aroused such as why the duty cycle we picked is better than the one for the maximum SPP excitation efficiency, how to facilitate the usage of the models with tens of coefficients or parameters, and are there any figures of merit (FoMs) or standard design recipes to design various structures for various applications.

To answer these questions, we further developed theoretical models on the SPP reflectance [34] and excitation coefficients by periodic surface corrugations under normal incidence [35]. These coefficients were calculated by simulations with corrugations being treated as 'black boxes' and first optimized for maximum SPP excitation efficiency in our previous models on the slit-gratings structure [32], [33]. As the SPP reflectance and excitation coefficients of N corrugations are also accurately and efficiently predicted by quantitative models starting with those of a single one, it is possible to develop a systematical theory incorporating all these models elegantly, and to provide an efficient and standard recipe that could be followed by experimentalists without efforts.

In this paper, we present a systematical semi-analytical theory on the slit-gratings structure designed for any incidence/beaming angle, propose FoMs for such a structure for the first time to the best of our knowledge, and introduce a standard and simple recipe to design such a structure based on the theory and the FoMs. The slit-grooves structure is analyzed as the example, and 'pure' SPP coupled-mode models are set up for the performance prediction and the structure optimization with high accuracy at visible or near infrared regimes [35]. Both the theory and the FoMs will be quantitatively validated with exhaustive examples by comparing with simulation results using the fully vectorial aperiodic Fourier modal method (a-FMM) [36] and finite element method (FEM) [37]. With the FoMs, the above-mentioned problem on better choice of groove duty cycle will be solved elegantly. We will also show that it is easy to obtain better performance following the proposed recipe than a widely-adopted conventional recipe, where corrugations were first optimized for maximum SPP excitation efficiency. The remainder of the paper is organized as follows. We set up the systematical theory in forms of two nested semi-analytical models in Section II, deduce FoMs in Section III, and then validate them in Section IV. In Section V, the recipe will then be introduced and exemplified. Finally, some concluding remarks are summarized in Section VI.

II. A Systematical Theory

In this section, we set up a systematical semi-analytical theory for the slit-grooves structure designed for any incidence/beaming angle θ, where the surrounding grooves are assumed to be asymmetrical without loss of generality. Although the model on optical transmission of TM-polarized plane wave and the one on its reverse process, i.e., optical beaming as illustrated in Figs. 2(a) and (b), respectively, lead to similar results according to the Lorentz reciprocity theorem, here we address the former instead of the latter to make use of our previous works. The theory is expressed in form of two nested models: a global model that treats the grooves on both sides as 'black boxes', and embodies all the key parameters with emphasis on the slit parameters and the slit-groove distances since they may lead to enhanced or suppressed transmission [23], [28], [32], [33]; and a nested model on the SPP excitation and reflectance coefficients of the 'black boxes'.

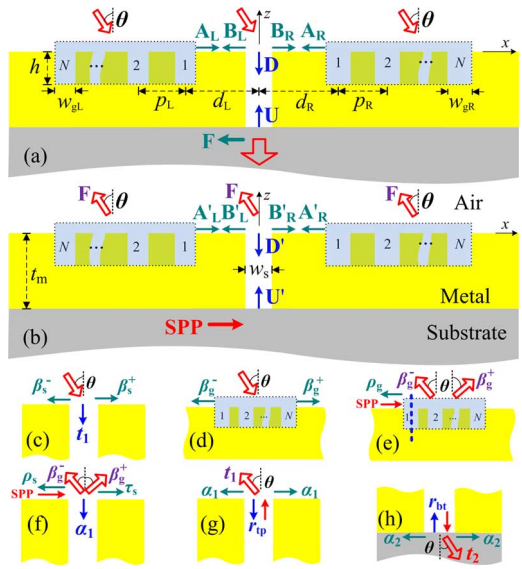

Fig. 2. Schematics of the global model on the slit-gratings structure. (a)Optical transmission under illumination by a TM-polarized (magnetic vector along y axis) plane wave of incidence angle θ. (b) Shows one of its reciprocal problem, i.e., optical beaming of angle θ under an SPP incidence (it should be under normal light incidence from the bottom in reality, but both cases lead to similar results). In the model, the gratings on both sides are treated as "black boxes" that excite and reflect SPPs. The groove widths, periods, and the slit-groove distances on the left and right sides are $w_{\rm L}, p_{\rm L}, d_{\rm L}$ and $w_{\rm R}, p_{\rm R}, d_{\rm R}$, respectively. Grooves are assumed to be of the same depth h to facilitate the manufacturing. The electromagnetic quantities $A_{\rm L}, B_{\rm L}, A_{\rm R}, B_{\rm R}, U, D, F$ are all defined in the text. Note that only F are of the same value in (a) and (b) according to the Lorentz reciprocity theorem. (c)–(h) Involved main elementary scattering processes. They are all associated with the scattering of an electromagnetic field by a slit (c), (f)–(h) or a grating (d), (e) under illumination of plane wave (c), (d), an SPP mode (e), (f), or the slit fundamental mode (g), (h). The vertical blue-dashed line in (e) indicates the zero phase of the radiated plane wave in x axis when calculating $\beta^{+}_{\rm gL}$ and $\beta^{-}_{\rm gR}$.

A. The Global Model on the Whole Structure

The global 'pure' SPP coupled-mode model on optical transmission is shown by Fig. 2(a), where $B_{\rm L}$, $A_{\rm R}$ and $A_{\rm L}$, $B_{\rm R}$ are complex amplitudes of magnetic field H_y of the left- and right-going SPP modes at the air-metal interface, respectively, F is that of the left-going SPP mode at the metal-substrate interface, and D and U are those of the slit fundamental modes propagating downward and upward, respectively. To obtain a closed-form expression for the transmission efficiency, which is defined as the total power transmitted into the far field normalized to the power incident on the slit aperture, we treat

the gratings on both sides as 'black boxes' and start with the knowledge of the elementary-event scattering coefficients shown in Fig. 2(c)–(h): $\beta_{\rm s}^-$ and $\beta_{\rm s}^+$ are the excitation coefficients of the left- and right-going SPP modes by the slit, respectively; $\beta_{\rm g}^-$ and $\beta_{\rm g}^+$ are those by a groove array; t_1 and t_2 are the respective excitation coefficients of the slit fundamental mode under plane wave illumination from the air and the substrate, and vice versa according to the reciprocity theorem; $\rho_{\rm g}$ is the SPP reflectance coefficient of a groove array; $\rho_{\rm s}$ and $\tau_{\rm s}$ are the SPP reflectance and transmittance coefficients of the slit; α_1 (or α_2) is the scattering coefficient from the SPP mode at the air-metal (or metal-substrate) interface to the slit fundamental mode and vice versa; and $r_{\rm tp}$ and $r_{\rm bt}$ are the reflectance coefficients of the slit fundamental mode at the top and the bottom openings, respectively. The coupled-mode equations lead to

$$\begin{cases} A_{\rm L} = w_{\rm L}^{-1}\beta_{\rm gL}^+ + \rho_{\rm gL}u_{\rm L}B_{\rm L} \\ B_{\rm L} = \beta_{\rm s}^- + \rho_{\rm s}u_{\rm L}A_{\rm L} + \tau_{\rm s}u_{\rm R}A_{\rm R} + \alpha_1 vU \\ A_{\rm R} = w_{\rm R}\beta_{\rm gR}^- + \rho_{\rm gR}u_{\rm R}B_{\rm R} \\ B_{\rm R} = \beta_{\rm s}^+ + \rho_{\rm s}u_{\rm R}A_{\rm R} + \tau_{\rm s}u_{\rm L}A_{\rm L} + \alpha_1 vU \\ D = t_1 + \alpha_1 u_{\rm L}A_{\rm L} + \alpha_1 u_{\rm R}A_{\rm R} + r_{\rm tp}vU \\ U = r_{\rm bt}vD \end{cases} \tag{1}$$

where $u_{\rm L} = \exp(ik_0 n_{\rm sp} d_{\rm L})$ and $u_{\rm R} = \exp(ik_0 n_{\rm sp} d_{\rm R})$ with $n_{\rm sp} = [n_{\rm air}^2 n_{\rm m}^2/(n_{\rm air}^2 + n_{\rm m}^2)]^{1/2}$ being the complex effective refractive index of the SPP mode at the flat air-metal interface, and $v = \exp(ik_0 n_{\rm eff}^{\rm s} t_{\rm m})$ with $n_{\rm eff}^{\rm s}$ being the complex effective refractive index of the slit fundamental mode as the slit is very small and of single-mode. $w_{\rm L} = \exp(ik_0 d_{\rm L}\sin\theta)$ and $w_{\rm R} = \exp(ik_0 d_{\rm R}\sin\theta)$ are phase shifts introduced by the incident plane wave. This is because the zero phase of the incidence is assumed to be at the top opening's center of the slit $(x = z = 0)$; whereas it is at the top opening's centers of the grooves nearest to the slit for the calculation of $\beta_{\rm gL}^+$ $(x = -d_{\rm L}, z = 0)$ and $\beta_{\rm gL}^-$ $(x = d_{\rm R}, z = 0)$. Note that the propagation losses of the SPP mode and of the slit fundamental mode have been embodies via complex $n_{\rm sp}$ and $n_{\rm eff}^{\rm s}$, respectively.

For a practical plasmonic structure with optimized grooves, $\beta_{\rm s}^{\pm}$ can be omitted as $|\beta_{\rm s}^{\pm}| \ll |\beta_{\rm gL}^+|$ and $|\beta_{\rm s}^{\pm}| \ll |\beta_{\rm gR}^-|$. By neglecting some trivial terms related to multiple cross conversions between $A_{\rm L;R}, B_{\rm L;R}$ and D, U, we obtain a closed-form expression from (1):

$$D \approx \frac{1}{1 - r_{\rm tp}r_{\rm bt}v^2}\left[t_1 + \frac{\alpha_1\left(\psi_{\rm R}\frac{u_{\rm L}}{w_{\rm L}}\beta_{\rm gL}^+ + \psi_{\rm L}u_{\rm R}w_{\rm R}\beta_{\rm gR}^-\right)}{(1-\rho_{\rm s}\sigma_{\rm L})(1-\rho_{\rm s}\sigma_{\rm R}) - \tau_{\rm s}^2\sigma_{\rm L}\sigma_{\rm R}}\right] \tag{2}$$

where $\psi_{\rm L;R} = 1 + (\tau_{\rm s} - \rho_{\rm s})\sigma_{\rm L;R}$ with $\sigma_{\rm L} = \rho_{\rm gL}u_{\rm L}^2$ and $\sigma_{\rm R} = \rho_{\rm gR}u_{\rm R}^2$. Equation (2) can be further reduced by making some empirical approximations: $\rho_{\rm gL}$ and $\rho_{\rm gR}$ are small as $\beta_{\rm gL}^+$ and $\beta_{\rm gR}^-$ are usually optimized or quasi-optimized; $|\rho_{\rm s}| < 0.15$ and $|\tau_{\rm s} - \rho_{\rm s}| \approx 1$ for so small $w_{\rm s}$ that the slit supports only the fundamental mode. As a result, terms incorporating $\rho_{\rm s}\rho_{\rm gL}$, $\rho_{\rm s}\rho_{\rm gR}$ and $\rho_{\rm gL}\rho_{\rm gR}$ are negligible, and $(1-\rho_{\rm s}\sigma_{\rm L})(1-\rho_{\rm s}\sigma_{\rm R}) - \tau_{\rm s}^2\sigma_{\rm L}\sigma_{\rm R} \approx 1$. Equation (2) is then further reduced into

$$D \approx \frac{t_1 + \alpha_1\left[\left(1+\rho_{\rm gR}u_{\rm R}^2\right)\frac{u_{\rm L}}{w_{\rm L}}\beta_{\rm gL}^+ + \left(1+\rho_{\rm gL}u_{\rm L}^2\right)u_{\rm R}w_{\rm R}\beta_{\rm gR}^-\right]}{1 - r_{\rm tp}r_{\rm bt}v^2}. \tag{3}$$

Specially, for the symmetrical structure designed for normal incidence, $\theta = 0$, $\rho_{\rm gL} = \rho_{\rm gR} = \rho_{\rm g}$, $\beta_{\rm gL}^+ = \beta_{\rm gR}^- = \beta_{\rm g}$, $d_{\rm L} = d_{\rm R} = d$, and $u_{\rm L} = u_{\rm R} = u$. Equation (2) is then reduced into

$$D \approx \frac{1}{1 - r_{\rm tp}r_{\rm bt}v^2}\left[t_1 + \frac{2\alpha_1 u\beta_{\rm g}}{1 - (\tau_{\rm s} + \rho_{\rm s})\rho_{\rm g}u^2}\right]. \tag{4}$$

The optical transmission efficiency η is then expressed as

$$\eta = \int_\theta |Dvt_2(\theta)\cos\theta|^2 {\rm d}\theta \tag{5}$$

and F is expressed as

$$F = Dv\alpha_2. \tag{6}$$

Obviously, $\eta \propto |D|^2$, $|F|^2 \propto |D|^2$, thus $\eta \propto |F|^2$ when the slit width and material (and then v and t_2) are usually given first according to the specific application requirement. As a result, we use $|F|^2$ or η to assess the performance.

B. The Nested Model on the Gratings

Now let us consider the 'black boxes' that excite and reflect SPPs. In previous works [16], [17], [27], [28], the gratings were usually first optimized by simulation scan for optimal $|\beta_{\rm gL}^+|$ and $|\beta_{\rm gR}^-|$. However, for a large groove number N, the optimization suffers from a high computational cost. What is worse, one has to repeat the optimization if N should be increased for better performance. To circumvent these problems, here we develop an efficient theoretical model for $\beta_{\rm gL}^+(\theta)$, $\beta_{\rm gR}^-(\theta)$ and $\rho_{\rm gL;gR}$ for any θ and N.

We take $\beta_{\rm gL}^+$ and $\rho_{\rm gL}$ of grooves on the left side as an example, where the zero phase of the incident plane wave is at the top opening's center of the rightmost groove, as shown in Fig. 3. $\beta_{\rm gR}^-$ and $\rho_{\rm gR}$ are calculated similarly but with a different zero phase position, as mentioned previously. The model may be expressed in two equivalent forms: the linear-equations form and the recursive form. The equivalence of the two forms has been verified for $\rho_{\rm g}$ in [34], thus it will not be discussed here due to space limitations. The model in a linear-equations form for $\beta_{\rm gL}^+$ and $\rho_{\rm gL}$ is shown in Fig. 3(a) and expressed as

$$\begin{cases} B_j = w_{\rm gL}^{j-N}\beta_{1,{\rm L}}^- + \rho_{1,{\rm L}}u_{\rm gL}A_{j-1} + \tau_{1,{\rm L}}u_{\rm gL}B_{j+1} \\ A_j = w_{\rm gL}^{j-N}\beta_{1,{\rm L}}^+ + \tau_{1,{\rm L}}u_{\rm gL}A_{j-1} + \rho_{1,{\rm L}}u_{\rm gL}B_{j+1}, \end{cases} \tag{7}$$

where $w_{\rm gL} = \exp(ik_0 p_{\rm L}\sin\theta)$ and $u_{\rm gL} = \exp(ik_0 n_{\rm sp} p_{\rm L})$ with $p_{\rm L}$ being the groove period, $j = 1, 2, \ldots, N$ with $B_{N+1} = 0$. $\beta_{1,{\rm L}}^{\pm}$, $\rho_{1,{\rm L}}$ and $\tau_{1,{\rm L}}$ are the right-going ('+') and left-going ('−') SPP excitation coefficients, the SPP reflectance and transmittance coefficients of a single groove, respectively, as shown in Figs. 3(c) and (d). To calculate the SPP excitation coefficients, $\beta_{\rm gL}^+ = A_N$, $\beta_{\rm gL}^- = B_1$, one sets $A_0 = 0$; whereas to calculate

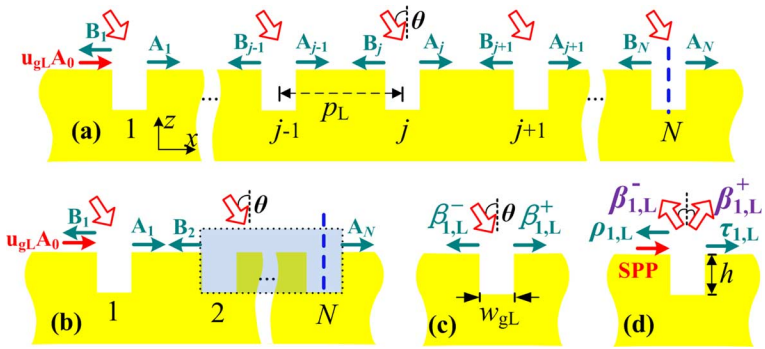

Fig. 3. (a) Schematics of the linear-equations form and (b) the recursive form of the coupled-mode model for the SPP excitation and reflectance coefficients by N periodic grooves on the left side of the central slit. $u_{\mathrm{gL}}A_0, A_1, B_1, \ldots, A_N, B_N$ are complex H_y amplitudes of SPP modes at the air-metal interface. (c) and (d) show the SPP excitation coefficients $\beta^{\pm}_{1,\mathrm{L}}$, reflectance coefficient $\rho_{1,\mathrm{L}}$ and transmittance coefficient $\tau_{1,\mathrm{L}}$ by a single groove, respectively. The vertical blue-dashed lines in (a) and (b) indicate the incidence zero phase at the top opening's center of the rightmost groove.

the SPP reflectance coefficient, $\rho_{\mathrm{gL}} = A_1/(u_{\mathrm{gL}}A_0)$, one sets $\beta^{\pm}_{1,\mathrm{L}} = 0$.

The model may also be expressed in a recursive form (see Append. A for details), as shown in Fig. 3(b)

$$\beta^{-}_{N,\mathrm{L}} = w_{\mathrm{gL}}^{1-N}\beta^{-}_{1,\mathrm{L}} + \tau_{1,\mathrm{L}}u_{\mathrm{gL}}\frac{\beta^{-}_{N-1,\mathrm{L}} + \rho_{N-1,\mathrm{L}}u_{\mathrm{gL}}w_{\mathrm{gL}}^{1-N}\beta^{+}_{1,\mathrm{L}}}{1-\rho_{1,\mathrm{L}}\rho_{N-1,\mathrm{L}}u_{\mathrm{gL}}^2} \quad (8a)$$

$$\beta^{+}_{N,\mathrm{L}} = \beta^{+}_{N-1,\mathrm{L}} + \tau_{N-1,\mathrm{L}}u_{\mathrm{gL}}\frac{w_{\mathrm{gL}}^{1-N}\beta^{+}_{1,\mathrm{L}} + \rho_{1,\mathrm{L}}u_{\mathrm{gL}}\beta^{-}_{N-1,\mathrm{L}}}{1-\rho_{1,\mathrm{L}}\rho_{N-1,\mathrm{L}}u_{\mathrm{gL}}^2} \quad (8b)$$

$$\rho_{N,\mathrm{L}} = \rho_{1,\mathrm{L}} + \rho_{N-1,\mathrm{L}}\frac{\tau_{1,\mathrm{L}}^2u_{\mathrm{gL}}^2}{1-\rho_{1,\mathrm{L}}\rho_{N-1,\mathrm{L}}u_{\mathrm{gL}}^2} \quad (8c)$$

$$\tau_{N,\mathrm{L}} = \frac{\tau_{1,\mathrm{L}}\tau_{N-1,\mathrm{L}}u_{\mathrm{gL}}}{1-\rho_{1,\mathrm{L}}\rho_{N-1,\mathrm{L}}u_{\mathrm{gL}}^2}. \quad (8d)$$

After the recursion, one obtains $\beta^{+}_{\mathrm{gL}} = \beta^{+}_{N,\mathrm{L}}$ and $\rho_{\mathrm{gL}} = \rho_{N,\mathrm{L}}$. Specially, for normal incidence, $w_{\mathrm{gL}} = 1$, and

$$\beta_{N,\mathrm{L}} = \beta_{N-1,\mathrm{L}} + \tau_{N-1,\mathrm{L}}u_{\mathrm{gL}}\frac{\beta_{1,\mathrm{L}} + \rho_{1,\mathrm{L}}u_{\mathrm{gL}}\beta_{N-1,\mathrm{L}}}{1-\rho_{1,\mathrm{L}}\rho_{N-1,\mathrm{L}}u_{\mathrm{gL}}^2}. \quad (9)$$

For fully periodic grooves ($N = \infty$) or large enough N, one has $|\rho_{1,\mathrm{L}}\rho_{N-1,\mathrm{L}}u_{\mathrm{gL}}^2| \ll 1$, $\beta^{-}_{N,\mathrm{L}} \approx w_{\mathrm{gL}}^{-1}\beta^{-}_{N-1,\mathrm{L}}$ and $\rho_{N,\mathrm{L}} \approx \rho_{N-1,\mathrm{L}}$. From (8), we obtain the generalized grating equation and Bragg equation expressed as

$$\arg(\tau_{1,\mathrm{L}}) + k_0\mathrm{Re}(n_{\mathrm{sp}})p \pm k_0n_0p\sin\theta = 2m_0\pi \quad (10a)$$

$$\arg(\tau_{1,\mathrm{L}}) + k_0\mathrm{Re}(n_{\mathrm{eff}})p = m_0\pi \quad (10b)$$

where '+' and '−' correspond to the excitations of the left- and right-going SPP modes, respectively. Compared with the conventional grating equation and Bragg equation, the 'generalized' ones with an additional term $\arg(\tau_{1,\mathrm{L}})$ are versatile for a general grating composed of periodic defects, where the defect may be of various geometries and refractive index profiles, as their influences have been embodied via $\arg(\tau_{1,\mathrm{L}})$. This additional term is quite important for dielectric [14], [16] or metal-dielectric composite [17] surface gratings since $\arg(\tau_{1,\mathrm{L}})$ may be relatively large in these cases. However, for grooves usually used because of manufacturing convenience, $\arg(\tau_{1,\mathrm{L}})$ is negligible. In other words, it is suitable to use the conventional grating equation and Bragg equation for periodic grooves for the sake of simplicity. In this paper, we restrict ourselves to periodic grooves.

By combining (3) for oblique incidence/beaming or (4) for normal incidence/beaming with (8), two theoretical models are nested elegantly, resulting in a systematical semi-analytical theory on the whole structure. Note that one may expect a whole set of linear equations based on the elementary-event scattering coefficients of a slit and a single groove, i.e., β^{+}_{gL}, ρ_{gL}, β^{-}_{gR} and ρ_{gR} in (2) are replaced by linear equations similar to (7). In this case, however, the linear equations are so complex that it is difficult to obtain a closed-form expression for $|D|^2$ (and $|F|^2$, η) and to provide intuitive physical pictures. This is the reason why we express the systematical theory in form of two nested models.

III. Figures of Merit

The systematical theory incorporates the dominant structural parameters. Based on the model, in this section we propose FoMs that lead to global optimization and a standard recipe, which will be introduced in the next section.

Comparing (3) and (4), we notice they share the same denominator, i.e., $1 - r_{\mathrm{tp}}r_{\mathrm{bt}}v^2$. Neglecting the propagation loss of the slit fundamental mode, $|1/(1 - r_{\mathrm{tp}}r_{\mathrm{bt}}v^2)| \le 1/(1 - |r_{\mathrm{tp}}r_{\mathrm{bt}}|)$ with equality holds when

$$2k_0\mathrm{Re}(n^{\mathrm{s}}_{\mathrm{eff}})t_{\mathrm{m}} + \arg(r_{\mathrm{tp}}) + \arg(r_{\mathrm{bt}}) = 2m_1\pi \quad (11)$$

where 'Re' and 'arg' mean the real part and the argument of a complex number, respectively, and m_1 is an integer. Equation (11) attributes to the vertical F-P resonance effect in the cavity formed by the slit's top and bottom openings. As this is well known and widely used to determine the optimized slit parameters [38], [39], we will spare space for groove parameters and the slit-groove distances. With (11), given the slit width w_{s} and material n_{s}, one obtains the optimal thickness of the metal film t_{m} simply by calculating $n^{\mathrm{s}}_{\mathrm{eff}}$, r_{tp}, and r_{bt}, which are determined by w_{s} and n_{s}.

Let us focus on the numerators of D in (3) and (4). For normal incidence/beaming

$$\left|t_1+\frac{2\alpha_1 u\beta_{\rm g}}{1-(\tau_{\rm s}+\rho_{\rm s})\rho_{\rm g}u^2}\right|^2 \tag{12a}$$

$$\approx |t_1|^2+\frac{4|\alpha_1|^2|\beta_{\rm g}|^2}{|1-(\tau_{\rm s}+\rho_{\rm s})\rho_{\rm g}u^2|^2}+\frac{4|\alpha_1||\beta_{\rm g}|\cos\Psi'}{|1-(\tau_{\rm s}+\rho_{\rm s})\rho_{\rm g}u^2|} \tag{12b}$$

$$\le |t_1|^2+\frac{4|\alpha_1|^2|\beta_{\rm g}|^2}{\left[1-|\tau_{\rm s}+\rho_{\rm s}||\rho_{\rm g}|\right]^2}+\frac{4|\alpha_1||\beta_{\rm g}|\cos\Psi'}{1-|\tau_{\rm s}+\rho_{\rm s}||\rho_{\rm g}|} \tag{12c}$$

$$\le \left[|t_1|+\frac{2|\alpha_1||\beta_{\rm g}|}{1-|\tau_{\rm s}+\rho_{\rm s}||\rho_{\rm g}|}\right]^2 \tag{12d}$$

where $\Psi'=\arg(\alpha_1\beta_{\rm g}/t_1)+k_0{\rm Re}(n_{sp})d$. The approximation in (12b) is performed by neglecting the propagation loss of SPP mode. Equality in (12c) holds if and only if the SPP propagation loss is neglected and there is horizontal F-P resonance effect due to the SPP modes' multiple reflections by the surrounding gratings

$$2k_0{\rm Re}(n_{\rm sp})d+\arg(\tau_{\rm s}+\rho_{\rm s})+\arg(\rho_{\rm g})=2m_2\pi. \tag{13}$$

Equality in (12d) holds when the slit fundamental modes excited by the incident light and by the groove-generated SPPs interfere constructively

$$k_0{\rm Re}(n_{\rm sp})d+\arg(\alpha_1/t_1)+\arg(\beta_{\rm g})=2m_3\pi. \tag{14}$$

Because the interference effect dominates when $2|\alpha_1||\beta_{\rm g}|$ is comparable to or smaller than $|t_1|$, it prevails in most of previous works [24]–[28] and only a few works addressed the horizontal F-P resonance effect [4], [32]. As a result, it was natural to consider only the interference effect and determine the optimal aperture-groove distance according to (14). However, as N increases, $2|\alpha_1||\beta_{\rm g}|$ becomes several times of $|t_1|$, and the contribution of the horizontal F-P resonance effect becomes preponderant. Bear these in mind, we make sure that (13) for the horizontal F-P resonance is always satisfied, and define a FoM from (12) by substituting (13) into Ψ'. The FoM is defined as

$${\rm FoM}_{\rm norm}=\frac{|\beta_{\rm g}||\cos\Psi|}{1-|\tau_{\rm s}+\rho_{\rm s}||\rho_{\rm g}|} \tag{15}$$

where $\Psi=\arg(\alpha_1\beta_{\rm g}/t_1)-1/2\arg[(\tau_{\rm s}+\rho_{\rm s})\rho_{\rm g}]$. We emphasize that (15) embodies the interlinks among the slit parameters (via α_1/t_1 and $\tau_{\rm s}+\rho_{\rm s}$), the groove parameters (via $\beta_{\rm g}$ and $\rho_{\rm g}$), and the slit-groove distance (determined by (13) according to the basic assumption). As a result, it is convenient to achieve global optimization via ${\rm FoM}_{\rm norm}$ with only a few parameters.

For oblique incidence/beaming, the numerator in (3)

$$\left|t_1+\alpha_1\left[\left(1+\rho_{\rm gR}u_{\rm R}^2\right)\frac{u_{\rm L}}{w_{\rm L}}\beta_{\rm gL}^{+}+\left(1+\rho_{\rm gL}u_{\rm L}^2\right)u_{\rm R}w_{\rm R}\beta_{\rm gR}^{-}\right]\right| \tag{16a}$$

$$\le |t_1|+|\alpha_1|[(1+|\rho_{\rm gR}|)|\beta_{\rm gL}^{+}|+(1+|\rho_{\rm gL}|)|\beta_{\rm gR}^{-}|]. \tag{16b}$$

This evokes multiple interference among the slit modes excited by the incidence, by groove-generated SPPs and by their first-order reflection by grooves on the other side. Because $|\rho_{\rm gL}|$ and $|\rho_{\rm gR}|$ are relatively small since $|\beta_{\rm gL}^{+}|$ and $|\beta_{\rm gR}^{-}|$ are usually optimized or quasi-optimized, the interference between the slit modes excited by the incidence and by groove-generated SPPs dominates. As a result, we make sure the constructive interference satisfies, i.e., $|t_1+\alpha_1((u_{\rm L})/(w_{\rm L})\beta_{\rm gL}^{+}+u_{\rm R}w_{\rm R}\beta_{\rm gR}^{-})|\le |t_1|+|\alpha_1|(|\beta_{\rm gL}^{+}|+|\beta_{\rm gR}^{-}|)$, where equality holds when

$$k_0\left[{\rm Re}(n_{\rm sp})-\sin\theta\right]d_{\rm L}+\arg(\alpha_1/t_1)+\arg(\beta_{\rm gL}^{+})=2m_4\pi \tag{17a}$$

$$k_0\left[{\rm Re}(n_{\rm sp})+\sin\theta\right]d_{\rm R}+\arg(\alpha_1/t_1)+\arg(\beta_{\rm gR}^{-})=2m_4'\pi. \tag{17b}$$

As N increases, $|\alpha_1|(|\beta_{\rm gL}^{+}|+|\beta_{\rm gR}^{-}|)$ becomes larger than $|t_1|$, and the contribution of the first-order reflection of SPPs by grooves on the other side increases and should not be omitted. From equality in (16b), which is achievable when parameters are properly designed, we define a FoM for the structure designed for oblique incidence/beaming as

$${\rm FoM}_{\rm obl}=(1+|\rho_{\rm gR}|)|\beta_{\rm gL}^{+}|+(1+|\rho_{\rm gL}|)|\beta_{\rm gR}^{-}|. \tag{18}$$

It is clear that two gratings on both sides are interlinked, resulting in a complex optimization. Taking into account the fact that $|\beta_{\rm gL}^{+}|$ and $|\beta_{\rm gR}^{-}|$ are comparable, we redefine the FoM as

$${\rm FoM}_{\rm obl}={\rm FoM}_{\rm gL}+{\rm FoM}_{\rm gR} \tag{19a}$$

$$=(1+|\rho_{\rm gL}|)|\beta_{\rm gL}^{+}|+(1+|\rho_{\rm gR}|)|\beta_{\rm gR}^{-}|. \tag{19b}$$

In such a way, there are two independent FoMs for gratings on each side, ${\rm FoM}_{\rm gL}$ and ${\rm FoM}_{\rm gR}$, greatly facilitating the optimization procedure. After the gratings are optimized, the optimal or quasi-optimal slit-groove distances $d_{\rm L}$ and $d_{\rm R}$ are then determined by (17).

IV. Computational Validations

In this section, we shortly validate the nested theoretical models (referred to as 'model A+B' for clarity) with $\beta_{\rm gL}^{+}(\theta)$, $\beta_{\rm gR}^{-}(\theta)$ and $\rho_{\rm gL;gR}$ of the groove array calculated by the nested model (referred to as 'model B'), by comparing with the sole global model using $\beta_{\rm gL}^{+}$, $\beta_{\rm gR}^{-}$ and $\rho_{\rm gL;gR}$ calculated by simulations (referred to as 'model A'), as done in our previous works [32], [33], and the fully vectorial a-FMM and FEM computations. The effectiveness of FoMs with $\beta_{\rm gL}^{+}$, $\beta_{\rm gR}^{-}$ and $\rho_{\rm gL;gR}$ (or $\beta_{\rm g}$ and $\rho_{\rm g}$ for $\theta=0$) calculated by 'model B' or by simulations are also validated with exhaustive calculations. In the models, t_1 is calculated analytically [40], $\rho_{\rm gL;gR}$, $\rho_{\rm s}$, $\tau_{\rm s}$, α_1, α_2, $\rho_{1,{\rm L;R}}$, $\tau_{1,{\rm L;R}}$, $r_{\rm tp}$ and $r_{\rm bt}$ are all calculated using the completeness theorem of the normal mode set and the mode orthogonality, as done in [35], [41]. To calculate $\beta_{\rm gL}^{+}$, $\beta_{\rm gR}^{-}$ and $\beta_{1,{\rm L;R}}^{\pm}$ of the groove array and F of the slit-grooves structure, we refer to an efficient method that allows us to calculate the SPP excitation coefficients for all incidence angles with a single computation, as has been developed in [42]. This method combines the a-FMM approach with the Lorentz reciprocity theorem. Instead of considering a SPP mode to be excited under illumination of the plane wave (Figs. 2(a) and (d)), we consider the excitation

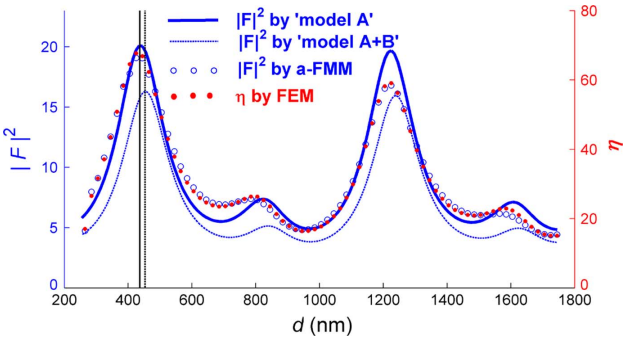

Fig. 4. Model predictions (the blue thick-solid and thin-dashed lines are for 'model A' and 'model A+B', respectively) and the a-FMM computational data using the reciprocal method (blue circles) on $|F|^2$ compared with η calculated by FEM (red dots). The vertical black solid and dashed lines indicate $d = 0.437$ μm and $d = 0.453$ μm, which are determined with (13) using ρ_g calculated by simulation and by 'model B', respectively. The calculations are performed with $\theta = 0°$, $p = \lambda_{sp} = 785$ nm, $w_g/p = 0.56$, $h = 70$ nm, $N = 11$ [28].

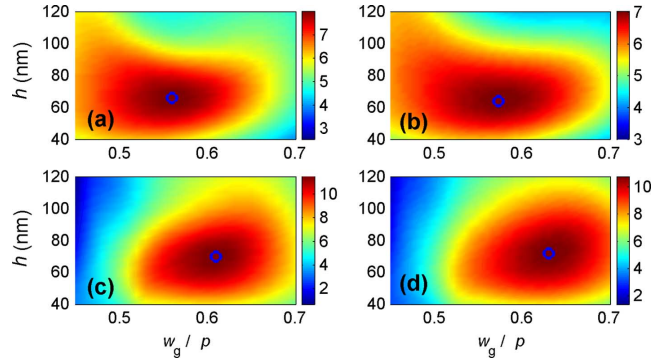

Fig. 5. Comparisons of the a-FMM computational data (a), (c) and 'model B' predictions (b), (d) on $|\beta_g|$ (a), (b) and FoM_{norm} (c), (d) as functions of the groove duty cycle and depth. The blue circles stand for the optimized groove sizes: $(w_g/p, h) = (0.56, 66$ nm) (a), (0.57, 64 nm) (b), (0.61, 70 nm) (c), (0.63, 72 nm) (d). The calculations are performed with $p = 785$ nm, $N = 11$.

of an out-going plane wave by the same structures with the SPP mode incidence (Figs. 2(b) and (e)).

Throughout the paper, we assume that the incident/beaming plane wave is normalized such that its power flow over the slit aperture is unitary, and the SPP mode is normalized such that it power flow along the x direction is also unitary. The analysis will be performed with gold at $\lambda = 800$ nm ($n_m = 0.181 + 5.118i$ [43]), $\theta = 0°$ and $\theta = 20°$, $n_{sub} = 1.46$, $n_s = 1.0$, $w_s = 50$ nm and $t_m \approx 178$ nm according to (11).

A. Normal Incidence/Beaming

For normal incidence/beaming ($\theta = 0°$), we first compare $|F|^2$ calculated by 'model A', by 'model A+B' and by the a-FMM approach combined with the reciprocity theorem, and η calculated by FEM. As shown in Fig. 4, theoretical predictions on $|F|^2$ and especially on the optimal d by 'model A' and by 'model A+B' agree well with a-FMM and FEM calculations, and $\eta \propto |F|^2$. Compared with Fig. 2(a) in [32], where most parameters are the same except $n_{sub} = 1.0$ and thus $t_m \approx 200$ nm according to (11), it is clear that $|F|^2$ and η as functions of the slit-groove distance hold the same features. We emphasize that it is better to determine the optimal d using ρ_g calculated by simulations than by 'model B'. This is because 'model B' as a 'pure' SPP model results in deviations in amplitude as well as in phases [35]. As a result, we hereafter determine the optimal d with (13) using ρ_g calculated by a-FMM.

Fig. 5 compares the a-FMM computational data and 'model B' prediction on $|\beta_g|$, as well as FoM_{norm} using β_g and ρ_g calculated by a-FMM and by 'model B', respectively. It is clear that both $|\beta_g|$ and FoM_{norm}, especially the optimized groove sizes are accurately predicted by 'model B'. As the numerical cost of the recursive equations is negligible compared with that of the calculation of β_1, the cost of the groove optimization for optimal $|\beta_g|$ or FoM_{norm} of N periodic grooves is greatly reduced into that of a single one.

Moreover, there are pronounced shifts between the optimized groove duty cycles for $|\beta_g|$ and for FoM_{norm}, indicating the influences of the horizontal F-P resonance effect. The shift is better illustrated in Fig. 6, where Fig. 6(a) replots Fig. 3(a) in

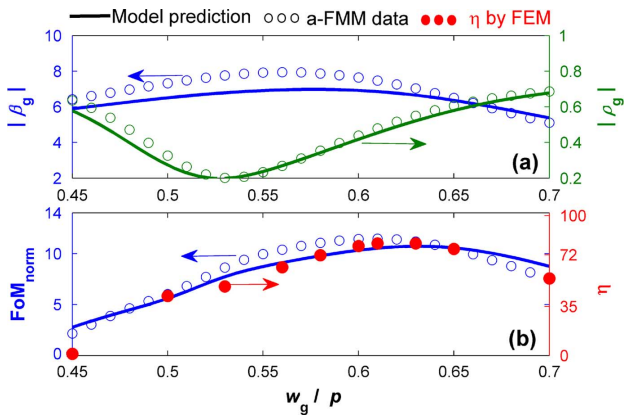

Fig. 6. Comparisons of FoM_{norm} using $|\beta_g|$ and $|\rho_g|$ calculated by a-FMM (circles) and by 'model B' (lines), and η calculated by FEM with d optimized according to (13) (red dots), where ρ_g is calculated by a-FMM. Other parameters are the same with Fig. 4.

[32]: FoM_{norm} is optimized at $w_g/p = 0.61$ (using β_g and ρ_g calculated by a-FMM) or 0.63 (using β_g and ρ_g calculated by 'model B'), compared with $|\beta_g|$ optimized at $w_g/p = 0.56$. η calculated by FEM using the optimal d determined by (13), where ρ_g is calculated by a-FMM, validates the effectiveness of FoM_{norm}. As a result, it is easy to make a better choice on the groove duty cycle with FoM_{norm} rather than with 'luckiness' in [32].

B. Oblique Incidence/Beaming

The model validation on β_{gL}^+, β_{gR}^-, $\rho_{gL;gR}$ and $\mathrm{FoM}_{gL;gR}$ of the groove array is exemplified with $\theta = 20°$, $N = 10$, and β_{gL}^+ and FoM_{gL} with $p_L = 1180$ nm determined by the conventional grating equation. Comparisons of the 'model B' predictions and the a-FMM computational data, as illustrated in Fig. 7, reveal that the model quantitatively captures all the salient features of $|\beta_{gL}^+|$ and FoM_{gL}, especially the optimized groove sizes.

We notice that there is also a pronounced shift on the optimized groove duty cycle for the optimal $|\beta_{gL}^+|$ and for the optimal FoM_{gL}. This shift is introduced by the first-order reflection of grooves-generated SPPs by grooves on the other side, and is better illustrated by Fig. 8, where the optimal FoM_{gL} is achieved when $|\beta_{gL}^+|$ is not optimized while $|\rho_{gL}|$ is relatively

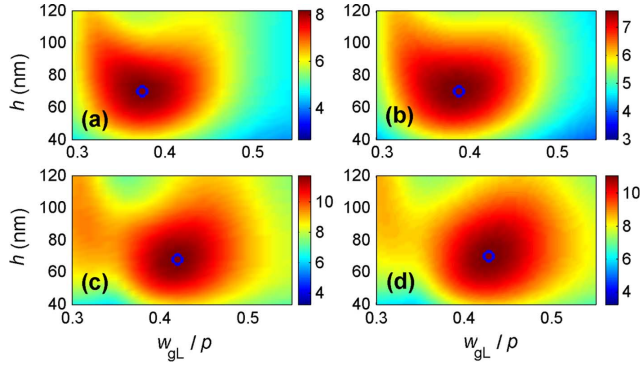

Fig. 7. Comparisons of the a-FMM computational data (a), (c) and 'model B' predictions (b), (d) on $|\beta_{\mathrm{gL}}^{+}|$ (a), (b) and $\mathrm{FoM}_{\mathrm{gL}}$ (c), (d) as functions of the groove duty cycle and depth. The blue circles indicate the optimized parameters: $w_{\mathrm{gL}}/p = 0.38$ (a), 0.42 (b), 0.39 (c), 0.43 (d), and $h = 70$ nm for (a)–(d). The calculations are performed with $\theta = 20°$, $N = 10$, and $p_{\mathrm{L}} = 1180$ nm.

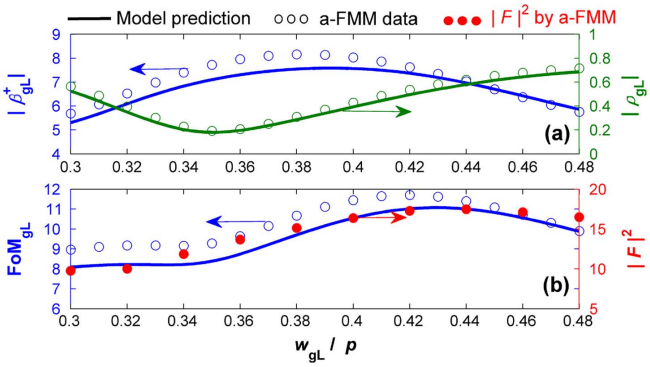

Fig. 8. Comparisons of $\mathrm{FoM}_{\mathrm{obl}}$ using $|\beta_{\mathrm{gL}}^{+}|$ and $|\rho_{\mathrm{gL}}|$ calculated by a-FMM (circles) and by 'model B' (lines), and $|F|^2$ calculated by the a-FMM approach using the reciprocal method (red dots), where d_{L} and d_{R} are optimized by (17) with $m_4 = m_4' = 1$, β_{gL}^{+} and β_{gR}^{-} being calculated by a-FMM. The calculations are performed with $p_{\mathrm{R}} = 588$ nm, $w_{\mathrm{gR}}/p_{\mathrm{R}} = 0.52$, $h_{\mathrm{L}} = 70$ nm, and other parameters same as in Fig. 7.

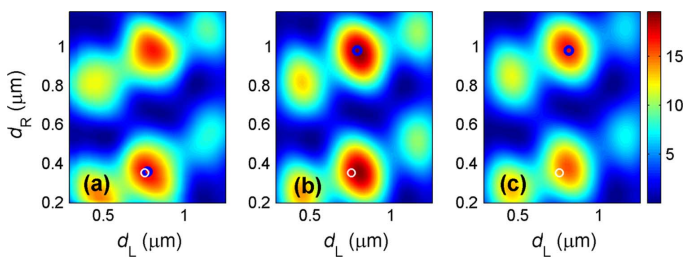

Fig. 9. Comparisons of the a-FMM computational data (a) and theoretical predictions by 'model A' (b) and 'model A+B' (c) on $|F|^2$ as functions of the slit-groove distances. The thick-blue circles indicate the optimized slit-groove distances, $(d_{\mathrm{L}}, d_{\mathrm{R}}) = (0.77\ \mu\mathrm{m}, 0.36\ \mu\mathrm{m})$ (a), (0.79 μm, 0.98 μm) (b), (0.81 μm, 0.98 μm) (c). The thin-white circles indicate $(d_{\mathrm{L}}, d_{\mathrm{R}}) = (0.76\ \mu\mathrm{m}, 0.35\ \mu\mathrm{m})$ determined by (17) with $m_4 = m_4' = 1$, β_{gL}^{+} and β_{gR}^{-} being calculated by a-FMM simulations ($d_{\mathrm{R}} = 0.94\ \mu$m if $m_4' = 2$). The calculations are performed with $p_{\mathrm{L}} = 1180$ nm, $w_{\mathrm{gL}}/p_{\mathrm{L}} = 0.42$, and other parameters same as in Fig. 8.

large. The effectiveness of the $\mathrm{FoM}_{\mathrm{gL}}$ is also well validated by comparing with $|F|^2$ calculated by a-FMM using the optimal slit-groove distances determined with (17), where $m_4 = m_4' = 1$, β_{gL}^{+} and β_{gR}^{-} are calculated by a-FMM.

Fig. 9 compares model predictions and the a-FMM computational data on $|F|^2$. It is shown that both 'model A' and 'model A+B' capture all the salient features. More importantly, the optimal slit-groove distances indicated by a-FMM, by 'model A' and by 'model A+B' are all very close to the ones determined with (17) provided β_{gL}^{+} and β_{gR}^{-} are calculated by simulations. As the numerical cost of 'model A+B' is greatly reduced compared with 'model A' while the prediction accuracy holds, it is more favorable to predict performance with the systematical theory in form of two nested models, and to obtain optimal or quasi-optimized structural parameters with $\mathrm{FoM}_{\mathrm{gL}}$, $\mathrm{FoM}_{\mathrm{gR}}$ and (11) and (17).

V. Standard Design Recipe

With the systematical theory and the FoMs, we now introduce an efficient, standard and simple design recipe for preparing a slit-grooves structure with optimal performance for optical concentration, beaming or collimation at a given wavelength ranging from the visible to the near infrared regime. It takes four steps:

1) The slit width w_{s}, the material filling the slit n_{s}, the metal n_{m}, the substrate n_{sub}, the incidence/beaming angle θ and the operating wavelength λ should be set first according to the specific application requirements.
2) The slit depth (or the metal film thickness) t_{m} is optimized using (11) to satisfy vertical F-P resonance in the slit. This step has been widely accepted and adopted.
3) Given the groove number N, the groove widths and depth are then optimized using $\mathrm{FoM}_{\mathrm{obl}}$ for oblique incidence/beaming or using $\mathrm{FoM}_{\mathrm{norm}}$ for normal incidence/beaming, where the groove periods are determined by the conventional grating equation, β_{gL}^{+} and ρ_{gL}, β_{gR}^{-} and ρ_{gR} (or β_{g} and ρ_{g} for $\theta = 0$) are calculated theoretically using 'model B'.
4) For oblique incidence/beaming, the optimal d_{L} and d_{R} are determined by (17), whereas for normal incidence/beaming, the optimal d is determined by (13), where β_{gL}^{+} and β_{gR}^{-}, or ρ_{g} are calculated by simulations with the optimized groove parameters.

To illustrate the recipe, here we present an example for $\theta = 0$ by comparing with a conventional recipe adopted in [28]: step 3 is replaced by optimizing the groove width and depth for maximum $|\beta_{\mathrm{g}}|$; in step 4, the optimal d is determined by the constructive interference condition, i.e., (14). Table I summarizes the optimized parameters and transmission performance for different N. First of all, we set $w_{\mathrm{s}} = 50$ nm, $n_{\mathrm{s}} = n_{\mathrm{sub}} = 1.0$ (air). We then obtain the optimal metal film thickness $t_{\mathrm{m}} \approx 200$ nm according to (11), where $n_{\mathrm{eff}}^{\mathrm{s}} = 1.4313 + 0.0143i$, $r_{\mathrm{tp}} = r_{\mathrm{bt}} = -0.4317 - 0.4739i$. To optimize the groove width and depth using $\mathrm{FoM}_{\mathrm{norm}}$, we first calculate β_1, ρ_1 and τ_1 of a single groove as functions of w_{g} and h. The groove period is determined by the conventional grating equation, $p = 785$ nm. Given the groove number N, β_{g}, ρ_{g} and $\mathrm{FoM}_{\mathrm{norm}}$ as functions of w_{g} and h are obtained by 'model B' with high efficiency and accuracy, followed by the optimized w_{g} and h. Finally, the optimal slit-groove distance d is determined with (13) using ρ_{g} calculated by a-FMM. Note that $\tau_{\mathrm{s}} + \rho_{\mathrm{s}} = 0.8779 - 0.0782i$ with $\arg(\tau_{\mathrm{s}} + \rho_{\mathrm{s}}) \approx 0$. In other words, the optimal slit-groove distance is determined by the horizontal F-P resonance in the

TABLE I
PERFORMANCE OF THE PROPOSED STANDARD RECIPE, WHERE $w_{\rm gL}/p$, h, $\mathrm{FoM}_{\rm norm}$ AND d ARE OBTAINED THEORETICALLY, $\rho_{\rm g}$ IS CALCULATED BY A-FMM, AND η CALCULATED BY FEM. η^* IS TAKEN FROM FIG. 3 IN [28]

N	$w_{\rm gL}/p$	h(nm)	$\mathrm{FoM}_{\rm norm}$	$\rho_{\rm g}$	$d(\mu$m)	η	η^*
4	0.62	100	5.959	0.422-0.092i	0.411	32	32
6	0.64	88	7.760	0.496-0.122i	0.413	45	42
8	0.64	82	9.142	0.513-0.141i	0.415	59	50
10	0.64	76	10.246	0.528-0.170i	0.417	72	59
12	0.63	70	11.200	0.515-0.189i	0.420	85	75
14	0.63	66	12.062	0.525-0.213i	0.422	96	78
16	0.62	62	12.852	0.504-0.218i	0.423	106	86

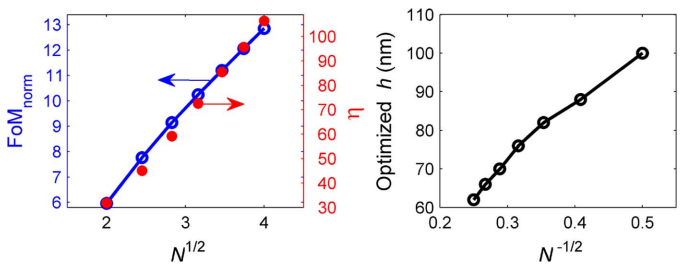

Fig. 10. The optimized $\mathrm{FoM}_{\rm norm}$ calculated by 'model B' (blue line with circles) and the optimized η calculated by FEM (red dots) scale with the number of grooves on each side $N^{1/2}$ (a), while the optimized groove depth obtained following the proposed standard recipe scales with $N^{-1/2}$ (b).

cavity formed by surrounding grooves, and it seems as if there is no slit at all.

It is clear that the standard recipe is very effective, leading to better performance than the reference conventional recipe. Note that the optimized groove duty cycle is about 0.63 instead of 0.5 as suggested in [4]. By plotting some parameters in Fig. 10, we notice the optimized $\mathrm{FoM}_{\rm norm}$ and η scale with $N^{1/2}$, while the optimized groove depth scales with $N^{-1/2}$ (consistent with [28]). As a result, $\mathrm{FoM}_{\rm norm}$ may also used to evaluate the performances of different groove numbers.

The proposed recipe is very efficient and flexible. One only needs to scan $\beta_1^{\pm}(\theta)$, ρ_1 and τ_1 of a single groove instead of $\beta_{\rm gL}^{+}$, $\rho_{\rm gL}$, $\beta_{\rm gR}^{-}$ and $\rho_{\rm gR}$ of N periodic grooves as functions of $w_{\rm g}$ and h. With the information of $\beta_1^{\pm}(\theta)$, ρ_1 and τ_1 as functions of $w_{\rm g}$ and h, $\beta_{\rm gL}^{+}$, $\rho_{\rm gL}$, $\beta_{\rm gR}^{-}$ and $\rho_{\rm gR}$ for various groove numbers or periods are obtained with negligible computational cost using 'model B'. In other words, the computational cost of simulation scan is greatly reduced. This cost reduction is especially remarkable when one needs to increase N to improve the performance, or when metallic, dielectric or metal-dielectric surface gratings are used, in that case the grating periods should also be scanned around the value determined by the conventional grating equation.

We emphasize that the proposed recipe provides clear physical picture and has great generality. It is capable to treat various structure configurations with surface gratings made of various shapes of grooves or dielectric ridges, with normal or oblique incidence/beaming. Furthermore, it provides a routine design procedure without relying on empirical experiences, making it a normalized recipe and easy to follow. We suggest it as a standard recipe for the design and optimization of the subwavelength aperture surrounded by surface corrugations.

VI. CONCLUDING REMARKS

Although we focused on 2D slit-grooves structure, the systematical coupled-mode theory, the FoMs, and the standard recipe are also applicable to 3D bull's-eye structure by making some modifications. This is because there is no transmission mode in the subwavelength hole aperture. In this case, the EM field inside the hole should be expanded with a set of waveguide modes, and the related coefficients such as t_1, t_2, $\rho_{\rm s}$, $\tau_{\rm s}$, α_1, α_2, $r_{\rm tp}$ and $r_{\rm bt}$ should be replaced by a set of corresponding coefficients accordingly [4], [44]. As the optimal aperture-groove distance under normal incidence is determined by the grooves' reflectance coefficient as if there is no slit or hole, it holds for 2D slit-grooves, 3D hole-grooves, and 3D bull's eye patterns, as noted in [19], [24], [25], [28].

Moreover, we should emphasize that the generalized grating equation and Bragg equation are also applicable for normal waveguide modes (simply by replacing $n_{\rm sp}$ with $n_{\rm eff}$). A typical application is in the free-space excitation or the reflection [45] of a dielectric waveguide mode by high-index-contrast gratings, the additional term $\arg(\tau_1)$ should not be omitted.

In conclusion, we have developed a systematical theory in form of two nested 'pure' SPP coupled-mode models for the widely-used plasmonic structure composed of a subwavelength aperture surrounded by surface corrugations. Based on the theory incorporating interlinks among key parameters with clear physical pictures, FoMs of the structures designed for normal and for oblique incidence/beaming have been proposed for the first time, making global optimization simple and efficient. Exhaustive calculations have shown that the theory and the FoMs are highly accurate on performance prediction and structural optimization. A standard recipe making full use of the theory and the FoMs has been introduced to facilitate the structure design and optimization with great flexibility and computational cost reduction.

APPENDIX A
SPP SCATTERING COEFFICIENTS IN RECURSIVE FORM

The model on SPP scattering coefficients $\beta_{N,\rm L}^{\pm}$ and $\rho_{N,\rm L}$ expressed in a recursive form is illustrated by Fig. 3(b). The coupled-mode equations lead to

$$\begin{cases} B_1 = w_{\rm gL}^{1-N}\beta_{1,\rm L}^{-} + \tau_{1,\rm L}u_{\rm gL}B_2 + \rho_{1,\rm L}u_{\rm gL}A_0 \\ A_1 = w_{\rm gL}^{1-N}\beta_{1,\rm L}^{+} + \rho_{1,\rm L}u_{\rm gL}B_2 + \tau_{1,\rm L}u_{\rm gL}A_0 \\ B_2 = \beta_{N-1,\rm L}^{-} + \rho_{N-1,\rm L}u_{\rm gL}A_1 \\ A_N = \beta_{N-1,\rm L}^{+} + \tau_{N-1,\rm L}u_{\rm gL}A_1. \end{cases} \tag{20}$$

To calculate the SPP excitation coefficients, $\beta_{N,\rm L}^{+} = A_N$, $\beta_{N,\rm L}^{-} = B_1$, one sets $A_0 = 0$; whereas to calculate the SPP reflectance and transmittance coefficients, $\rho_{N,\rm L} = A_1/(u_{\rm gL}A_0)$ and $\tau_{N,\rm L} = B_N/(u_{\rm gL}A_0)$, one sets $\beta_{1,\rm L}^{\pm} = 0$. Note that when calculating $\beta_{N,\rm L}^{\pm}$ and $\beta_{N-1,\rm L}^{\pm}$, the zero phases of the incident plane wave are all set to be at the center of the rightmost groove's top opening.

Then the SPP excitation, reflectance and transmittance coefficients are expressed recursively as

$$\beta_{N,\rm L}^{-} = w_{\rm gL}^{1-N}\beta_{1,\rm L}^{-} + \tau_{1,\rm L}u_{\rm gL}\frac{\beta_{N-1,\rm L}^{-} + \rho_{N-1,\rm L}u_{\rm gL}w_{\rm gL}^{1-N}\beta_{1,\rm L}^{+}}{1-\rho_{1,\rm L}\rho_{N-1,\rm L}u_{\rm gL}^2} \tag{21a}$$

$$\beta_{N,\mathrm{L}}^{+} = \beta_{N-1,\mathrm{L}}^{+} + \tau_{N-1,\mathrm{L}} u_{\mathrm{gL}} \frac{w_{\mathrm{gL}}^{1-N}\beta_{1,\mathrm{L}}^{+} + \rho_{1,\mathrm{L}} u_{\mathrm{gL}} \beta_{N-1,\mathrm{L}}^{-}}{1 - \rho_{1,\mathrm{L}}\rho_{N-1,\mathrm{L}} u_{\mathrm{gL}}^{2}} \tag{21b}$$

$$\rho_{N,\mathrm{L}} = \rho_{1,\mathrm{L}} + \rho_{N-1,\mathrm{L}} \frac{\tau_{1,\mathrm{L}}^{2} u_{\mathrm{gL}}^{2}}{1 - \rho_{1,\mathrm{L}}\rho_{N-1,\mathrm{L}} u_{\mathrm{gL}}^{2}} \tag{21c}$$

$$\tau_{N,\mathrm{L}} = \frac{\tau_{1,\mathrm{L}}\tau_{N-1,\mathrm{L}} u_{\mathrm{gL}}}{1 - \rho_{1,\mathrm{L}}\rho_{N-1,\mathrm{L}} u_{\mathrm{gL}}^{2}}. \tag{21d}$$

Specially, for normal incidence, $w_{\mathrm{gL}} = 1$, $\beta_j^{+} = \beta_j^{-} = \beta_j$ for $j = 1, \ldots, N$, (21a) and (21b) are then reduced into

$$\beta_{N,\mathrm{L}} = \beta_{N-1,\mathrm{L}} + \tau_{N-1,\mathrm{L}} u_{\mathrm{gL}} \frac{\beta_{1,\mathrm{L}} + \rho_{1,\mathrm{L}} u_{\mathrm{gL}} \beta_{N-1,\mathrm{L}}}{1 - \rho_{1,\mathrm{L}}\rho_{N-1,\mathrm{L}} u_{\mathrm{gL}}^{2}} \tag{22}$$

which agrees with our previous results [35].

For fully periodic grooves ($N = \infty$) or N is large enough, one has $|\rho_{1,\mathrm{L}}\rho_{N-1,\mathrm{L}} u_{\mathrm{gL}}^{2}| \ll 1$, $\beta_{N,\mathrm{L}}^{-} \approx w_{\mathrm{gL}}^{-1}\beta_{N-1,\mathrm{L}}^{-}$ and $\rho_{N,\mathrm{L}} \approx \rho_{N-1,\mathrm{L}}$. Equations (21a) and (21c) are then reduced into

$$\beta_{N,\mathrm{L}}^{-} \approx w_{\mathrm{gL}}^{1-N} \frac{\beta_{1,\mathrm{L}}^{-} + \tau_{1,\mathrm{L}}\rho_{N-1,\mathrm{L}} u_{\mathrm{g}}^{2}\beta_{1,\mathrm{L}}^{+}}{1 - \tau_{1,\mathrm{L}} u_{\mathrm{gL}} w_{\mathrm{gL}}} \tag{23a}$$

$$\rho_{N,\mathrm{L}} \approx \frac{\rho_{1,\mathrm{L}}}{1 - \tau_{1,\mathrm{L}}^{2} u_{\mathrm{gL}}^{2}} \tag{23b}$$

respectively. From (23a), it is easy to obtain the condition for the constructive interference of left-going SPP modes excited by the groove array:

$$\arg(\tau_{1,\mathrm{L}}) + k_0 \mathrm{Re}(n_{\mathrm{sp}})p + k_0 n_0 p \sin\theta = 2m_0\pi. \tag{24}$$

Similarly, the condition for the constructive interference of right-going SPP modes are also obtained. We refer to these conditions as the generalized grating equation:

$$\arg(\tau_{1,\mathrm{L}}) + k_0 \mathrm{Re}(n_{\mathrm{sp}})p \pm k_0 n_0 p \sin\theta = 2m_0\pi \tag{25}$$

where '$+$' and '$-$' correspond to the excitations of the left- and right-going SPP modes, respectively.

From (23b), the generalized Bragg equation

$$\arg(\tau_{1,\mathrm{L}}) + k_0 \mathrm{Re}(n_{\mathrm{eff}})p = m_0\pi \tag{26}$$

has been introduced and validated in [34].

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for providing valuable comments.

REFERENCES

[1] J. A. Schuller, E. S. Barnard, W. Cai, Y. C. Jun, J. S. White, and M. I. Brongersma, "Plasmonics for extreme light concentration and manipulation," *Nat. Mater.*, vol. 9, pp. 193–204, 2010.

[2] T. Thio, K. M. Pellerin, R. A. Linke, H. J. Lezec, and T. W. Ebbesen, "Enhanced light transmission through a single subwavelength aperture," *Opt. Lett.*, vol. 26, pp. 1972–1974, 2001.

[3] T. Thio, H. J. Lezec, T. W. Ebbesen, K. M. Pellerin, G. D. Lewen, A. Nahata, and R. A. Linke, "Giant optical transmission of sub-wavelength apertures: Physics and applications," *Nanotechnol.*, vol. 13, pp. 429–432, 2002.

[4] S. Carretero-Palacios, O. Mahboub, F. J. García-Vidal, L. Martín-Moreno, S. G. Rodrigo, C. Genet, and T. W. Ebbesen, "Mechanisms for extraordinary optical transmission through bull's eye structures," *Opt. Exp.*, vol. 19, pp. 10 429–10 442, 2011.

[5] T. Ishi, J. Fujikata, K. Makita, T. Baba, and K. Ohashi, "Si nano-photodiode with a surface plasmon antenna," *Jpn. J. Appl. Phys.*, vol. 44, pp. L364–L366, 2005.

[6] Z. Yu, G. Veronis, S. Fan, and M. L. Brongersma, "Design of midinfrared photodetectors enhanced by surface plasmons on grating structures," *Appl. Phys. Lett.*, vol. 89, p. 151116, 2006.

[7] R. D. Bhat, N. C. Panoiu, S. R. Brueck, and R. M. Osgood, "Enhancing the signal-to-noise ratio of an infrared photodetector with a circular metal grating," *Opt. Exp.*, vol. 16, pp. 4588–4596, 2008.

[8] J. A. Shackleford, R. Grote, M. Currie, J. E. Spanier, and B. Nabet, "Integrated plasmonic lens photodetector," *Appl. Phys. Lett.*, vol. 94, p. 083501, 2009.

[9] G. Zheng, X. Cui, and C. Yang, "Surface-wave enabled darkfield aperture for background suppression during weak signal detection," *Proc. Natl. Acad. Sci.*, vol. 107, pp. 9043–9048, 2010.

[10] G. Zheng and C. Yang, "Improving weak-signal identification via predetection background suppression by a pixel-level, surface-wave-enabled dark-field aperture," *Opt. Lett.*, vol. 35, pp. 2636–26 348, 2010.

[11] E. Laux, C. Genet, T. Skauli, and T. W. Ebbesen, "Plasmonic photon sorters for spectral and polarimetric imaging," *Nat. Photonics*, vol. 2, pp. 161–164, 2008.

[12] M. Consonni, J. Hazart, and G. Lerondel, "Fabry-pérot type enhancement in plasmonic visible nanosource," *Appl. Phys. Lett.*, vol. 94, p. 051105, 2009.

[13] H. J. Lezec, A. Degiron, E. Devaux, R. A. Linke, L. Martín-Moreno, F. J. García-Vidal, and T. W. Ebbesen, "Beaming light from a subwavelength aperture," *Science*, vol. 297, pp. 820–822, 2002.

[14] D. Z. Lin, C. K. Chang, Y. C. Chen, D. L. Yang, M. W. Lin, J. T. Yeh, J. M. Liu, C. H. Kuan, C. S. Yeh, and C. K. Lee, "Beaming light from a subwavelength metal slit surrounded by dielectric surface gratings," *Opt. Exp.*, vol. 14, pp. 3503–3511, 2006.

[15] L. Martín-Moreno, F. J. García-Vidal, H. J. Lezec, A. Degiron, and T. W. Ebbesen, "Theory of highly directional emission from a single subwavelength aperture surrounded by surface corrugations," *Phys. Rev. Lett.*, vol. 90, p. 167401, 2003.

[16] S. Kim, H. Kim, Y. Lim, and B. Lee, "Off-axis directional beaming of optical field diffracted by a single subwavelength metal slit with asymmetric dielectric surface gratings," *Appl. Phys. Lett.*, vol. 90, p. 051113, 2007.

[17] H. Kim, J. Park, and B. Lee, "Tunable directional beaming from subwavelength metal slits with metal-dielectric composite surface gratings," *Opt. Lett.*, vol. 34, pp. 2569–2571, 2009.

[18] B. Lee, S. Kim, H. Kim, and Y. Lim, "The use of plasmonics in light beaming and focusing," *Prog. Quant. Electron.*, vol. 34, pp. 47–87, 2010.

[19] Y. C. Jun, K. C. Y. Huang, and M. L. Brongersma, "Plasmonic beaming and active control over fluorescent emission," *Nat. Commun.*, vol. 2, p. 283, 2011.

[20] N. Yu, J. Fan, Q. J. Wang, C. Pflugl, L. Diehl, T. Edamura, M. Yamanishi, H. Kan, and F. Capasso, "Small-divergence semiconductor lasers by plasmonic collimation," *Nat. Photonics*, vol. 2, pp. 564–570, 2008.

[21] N. Yu *et al.*, "Plasmonics for laser beam shaping," *IEEE Trans. Nanotechnol.*, vol. 9, pp. 11–29, 2010.

[22] F. J. García-Vidal, H. J. Lezec, T. W. Ebbesen, and L. Martín-Moreno, "Multiple paths to enhance optical transmission through a single subwavelength slit," *Phys. Rev. Lett.*, vol. 90, p. 213901, 2003.

[23] A. Degiron and T. W. Ebbesen, "Analysis of the transmission process through single apertures surrounded by periodic corrugations," *Opt. Exp.*, vol. 12, pp. 3694–3700, 2004.

[24] P. Lalanne and J. P. Hugonin, "Interaction between optical nano-objects at metallo-dielectric interfaces," *Nat. Phys.*, vol. 2, pp. 551–556, 2006.

[25] G. Gay, O. Alloschery, B. V. D. Lesegno, C. O'Dwyer, J. Weiner, and H. J. Lezec, "The optical response of nanostructured surfaces and the composite diffracted evanescent wavemodel," *Nat. Phys.*, vol. 2, pp. 262–267, 2006.

[26] J. Olkkonen, K. Kataja, and D. G. Howe, "Light transmission through a high index dielectric hole in a metal film surrounded by surface corrugations," *Opt. Exp.*, vol. 14, pp. 11 506–11 511, 2006.

[27] C. M. Wang, C. C. Chao, H. I. Huang, B. Ung, Y. Sheng, and J. Y. Chang, "Transmission enhancement through a metallic slit assisted by low scattering loss corrugations," *Opt. Comm.*, vol. 281, pp. 2996–2999, 2008.
[28] O. T. A. Janssen, H. P. Urbach, and G. W. 't Hooft, "Giant optical transmission of a subwavelength slit optimized using the magnetic field phase," *Phys. Rev. Lett.*, vol. 99, p. 043902, 2007.
[29] O. Mahboub, S. C. Palacios, C. Genet, F. J. García-Vidal, S. G. Rodrigo, L. Martín-Moreno, and T. W. Ebbesen, "Optimization of bull's eye structures for transmission enhancement," *Opt. Exp.*, vol. 18, pp. 11 292–11 299, 2010.
[30] Y. X. Cui, S. L. He, and Y. Okuno, "Giant optical transmission through a metallic nano-slit achieved by the optimization of the groove periodicity and other parameters," in *Proc. 2008 Int. Workshop Metamaterials*, 2008, pp. 236–239.
[31] Y. X. Cui and S. L. He, "A theoretical re-examination of giant transmission of light through a metallic nano-slit surrounded with periodic grooves," *Opt. Exp.*, vol. 17, pp. 13 995–14 000, 2009.
[32] L. Cai, G. Li, Z. Wang, and A. Xu, "Interference and horizontal Fabry-Perot resonance on extraordinary transmission through a metallic nano-slit surrounded with grooves," *Opt. Lett.*, vol. 35, pp. 127–129, 2010.
[33] L. Cai, G. Li, F. Xiao, Z. Wang, and A. Xu, "Theory of enhanced optical transmission through a metallic nano-slit surrounded with asymmetric grooves under oblique incidence," *Opt. Exp.*, vol. 18, pp. 19 495–19 503, 2010.
[34] G. Li, L. Cai, F. Xiao, Y. Pei, and A. Xu, "A quantitative theory and the generalized Bragg condition for surface plasmon Bragg reflectors," *Opt. Exp.*, vol. 18, pp. 10 487–10 499, 2010.
[35] G. Li, F. Xiao, L. Cai, K. Alameh, and A. Xu, "Theory of the scattering of light and surface plasmon polaritons by finite-size subwavelength metallic defects via field decomposition," *New J. Phys.*, vol. 13, p. 073045, 2011.
[36] E. Silberstein, P. Lalanne, J. P. Hugonin, and Q. Cao, "Use of grating theories in integrated optics," *J. Opt. Soc. Amer. A*, vol. 18, pp. 2865–2875, 2001.
[37] X. Wei, A. J. H. Wachters, and H. P. Urbach, "Finite-element model for three-dimensional optical scattering problems," *J. Opt. Soc. Amer. A*, vol. 24, pp. 866–881, 2007.
[38] A. Krishnan, T. Thio, T. J. Kima, H. J. Lezec, T. W. Ebbesen, P. A. Wolff, J. Pendry, L. Martín-Moreno, and F. J. Garcia-Vidal, "Evanescently coupled resonance in surface plasmon enhanced transmission," *Opt. Commun.*, vol. 200, pp. 1–7, 2001.
[39] B. Ung and Y. Sheng, "Interference of surface waves in a metallic nanoslit," *Opt. Exp.*, vol. 15, pp. 1182–1190, 2007.
[40] P. Lalanne, J. P. Hugonin, and J. C. Rodier, "Approximate model for surface-plasmon generation at slit apertures," *J. Opt. Soc. Amer. A*, vol. 23, pp. 1608–1615, 2006.
[41] P. Lalanne, J. P. Hugonin, H. Liu, and B. Wang, "A microscopic view of the electromagnetic properties of sub-λ metallic surfaces," *Surf. Sci. Rep.*, vol. 64, pp. 453–469, 2009.
[42] H. Liu, P. Lalanne, X. Yang, and J. P. Hugonin, "Surface plasmon generation by subwavelength isolated objects," *IEEE J. Sel. Top. Quantum Electron.*, vol. 14, pp. 1522–1529, 2008.
[43] E. D. Palik, *Handbook of Optical Constants of Solids*. New York: Academic Press, 1985.
[44] F. de León-Pérez, G. Brucoli, F. J. García-Vidal, and L. Martín-Moreno, "Theory on the scattering of light and surface plasmon polaritons by arrays of holes and dimples in a metal film," *New J. Phys.*, vol. 10, p. 105017, 2008.
[45] J. W. Mu, H. Zhang, and W. P. Huang, "Design of waveguide Bragg gratings with strong index corrugations," *J. Lightw. Technol.*, vol. 26, pp. 1596–1601, 2008.

Author biographies not included at author request due to space constraints.