# Light diffraction by a slit and grooves with a point source model based on wave dynamics

Jian-Shiung Hong and Kuan-Ren Chen*
*Department of Physics, National Cheng Kung University, Tainan 70101, Taiwan*
(Received 28 April 2017; published 6 October 2017)

A point source model based on wave dynamics is proposed to study the fundamental light diffraction physics by a subwavelength slit and grooves in a metallic film. In this model, two opposite traveling waves are considered in each indentation; the resultant outgoing wave can propagate along the film surface to couple each other or radiate into free space as a point source. With small-system simulations, the tangential electric field at each opening determines its source temporal phase; then the energy conservation of each point source radiation and of the total radiant wave determine the source amplitudes. Besides these, this model reveals more physics regarding the wave interactions. In the strong-wave-coupling case studied, the surface waves created by the grooves flow into the slit and delay the Fabry-Pérot-like resonance. When adding the grooves concentrates the light field into a directional beam, the total transmitted energy through the slit significantly decreases. However, the energy in the original nearby grooves increases so that the groove radiation increasingly shares the transmitted energy. As the total transmitted energy decreases, the slit radiation energy decreases further due to the energy conservation. In the weak-wave-coupling cases, the groove radiation still interferes with that from the slit; as a result, the diffracted light is split into two beams. It is interesting to find that, due to the groove radiation, the slit radiation energy is enhanced to become larger than that transmitted through it. Detailed physical interpretations will be given.

DOI: 10.1103/PhysRevA.96.043813

## I. INTRODUCTION

The interaction between electromagnetic (EM) waves and subwavelength metallic structures has been an active research area in the past two decades since the discovery of the extraordinary optical transmission through metallic holes due to the excitation of surface plasmons (SPs) [1]. The EM wave can be either guided or localized in a predefined structure [2] and can also be scattered into space or coupled to another structure [3]. Hence, besides academic importance, controlling light via the excitation of SPs and radiations with subwavelength metallic structures opens up many potential optical applications, such as plasmonic circuit [4], biosensing [5], plasmonic solar cell [6], optical antennas [7], optical data storage [8], and plasmonic lithography [9].

The diffracted light shape can be manipulated; besides focusing [10–16], structured subwavelength metallic films have been shown to be capable of collimating a light beam [17–21] or bending it at a specific angle [22–26]. An analytical theory [27] studied the wave coupling mechanism in a structured perfect electric conducting (PEC) film, which contains a subwavelength slit and a number of grooves at both sides of the slit exit; the diffracted field at an observation location in free space is then yielded from the superposition of the radiations from each of the indentation openings. A similar theory was applied to study the transmission properties of such a structured film [28]. Both the fundamental mode and high-order modes in the indentations were considered by a mode expansion theory [29]. An analytical modal solution [30] was obtained with a two-layer model and the transfer-matrix method. A finite-difference time-domain (FDTD) simulation method was employed to study the phase and amplitude of the field at the openings and their scattering interactions [31]; the horizontal Fabry-Pérot resonance effect between the indentations was investigated theoretically while the slit modes interfere with the groove-generated surface plasmons [32]. The phases of the fields at the grooves were calculated so that the structure can be designed to focus the diffracted light at a designed location or to generate a light beam at a designed angle by the constructive interference [33–35].

The wave-matter interaction mechanism in the light manipulation area so far has been extensively understood by solving and analyzing the wave-coupling equations, but more underlying physics still awaits to be uncovered beyond the conventional scope. In this paper, we put emphasis on revealing the physics from a straightforward perspective by our point source model. Without loss of generality, a subwavelength structure composed of a slit and a number of identical grooves is employed as our example, where each of the indentations is separated equally [27]. The FDTD simulation shows that, when the incident wavelength is close to the indentation separation distance, the diffracted field converges to a light beam with the increase of the grooves; the result is consistent with that in Ref. [27]. To understand the diffraction physics, we begin with the fundamental wave interaction dynamics. Two opposite traveling waves are assumed in all indentations as inspired by the Fabry-Pérot-like resonance [36–39] and the surface wave coupling mechanism [31]. The resultant outgoing waves can propagate along the film surface to couple with each other or radiate into free space so as to be considered as a point source for developing the point source model. With the help of small-system simulations, the tangential electric field at the openings determines the source temporal phases; then the energy-conservation conditions of each point source radiation and of the total radiant wave determine the source amplitudes. This model is verified by the results of large-system simulations. The cases of longer incident wavelengths that split the diffracted light into two beams are also studied.

*chenkr@mail.ncku.edu.tw

 

In the thorough investigation of this diffraction problem, we will reveal the underlying physics that has not been fully uncovered yet. The previous study [28] indicates that the transmitted energy flux through a similar structured film is changed less than 1% when the groove number is varied from 1 to 10. In contradiction, there is about a 10% decrease in our main case. This discrepancy is fundamental and will be shown to be related to the phase change of the Fabry-Pérot-like resonance in the slit. As indicated by our analysis, the surface waves induced by the added grooves flow backward into the slit and then effectively delay the round-trip traveling waves within the slit. The phase difference between the incident and the round-trip waves thus increases, as happens by increasing the film thickness in the case of a single slit, in which the transmitted energy varies periodically with the film thickness. As will be shown, the energy change due to this phase difference in the previous study is negligible, while that in our case becomes significant. These surface waves also flow back into the nearby grooves closer to the slit to increase their energy fluxes; as a result, the groove radiation increasingly shares the transmitted energy. As the transmitted energy decreases, the radiated energy from the slit into free space decreases further due to the energy conservation.

In the cases of the longer incident wavelengths, the groove coupling is weak, so the surface waves from the grooves are not influential; however, their radiation still interferes with that coming from the slit to cause the angularly distributed diffraction patterns. Our study will show that the beam angles can be determined by the point source location and the temporal phase difference, while those in the previous studies [22–26,33–35] were determined according to the conventional grating diffraction theory. Interestingly, in these weak-wave-coupling cases, we find that the slit radiation energy is enhanced and becomes larger than that transmitted through it; the energy-conservation condition indicates that the contribution from the groove radiation is responsible for the enhancement.

Section II introduces the structure of interest and shows the resultant diffraction patterns in the large simulation system. The wave interaction dynamics and the point source model are proposed in Sec. III, with the simulation results in a small system. Section IV presents the modeling results for the case studied in Sec. III and those for the cases of longer incident wavelengths. The physics of the wave interaction dynamics is investigated and interpreted in Sec. V. Section VI is a summary.

## II. SIMULATION FOR THE LIGHT DIFFRACTION BY A SLIT AND GROOVES

Consider a subwavelength slit and grooves in a metallic film, as shown in Fig. 1. Without loss of generality, the film is a PEC, both the slit and groove widths, $2s$ and $2g$, are 40 nm, the groove depth $d$ is 100 nm, the separation distance between each indentation $p$ is 500 nm, and the groove number $N$ at each side of the slit considered here will vary from 0 to 10. These structural parameters are identical to those used in Ref. [27], which considered only the exit side of the film and ignored the entrance side; for this study, we set the film thickness $h$ to 250 nm. For convenience, the indentations including the slit

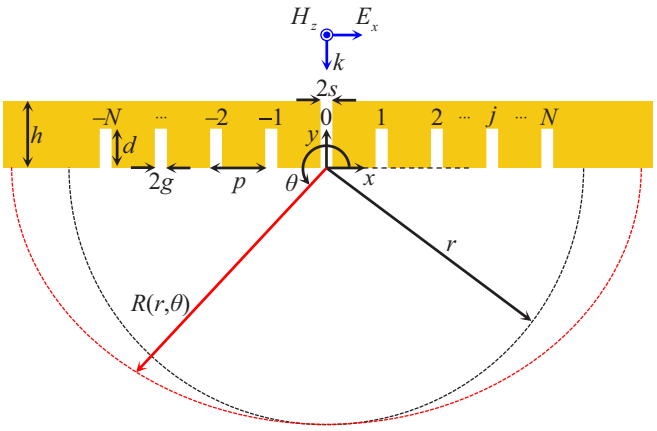


FIG. 1. Schematic of a slit and a number of grooves in a PEC film. The structural parameters are $2s = 2g = 40$ nm, $d = 100$ nm, $p = 500$ nm, and $h = 250$ nm. The groove number at each side of the slit is $N$, where $N$ will vary from 0 to 10; each groove and the slit are numbered in sequence from the left to the right side with an integer $j$ from $-N$ to $N$. A monochromatic $p$-polarized plane wave of wavelength $\lambda = 560$ nm is normally incident. The Cartesian coordinate $(x,y)$ of the system originates at the center of the slit exit and a polar coordinate $(r,\theta)$ is additionally defined. The red dashed curve $R(r,\theta)$ denotes the path of a wave front of the diffracted light at the peak phase that is sought near a semicircle of radius $r$ (black dashed curve) at the angle $\theta$ from 180° to 360°.

are numbered in sequence from the left to the right side with an integer $j$ from $-N$ to $N$. A monochromatic $p$-polarized plane wave of wavelength $\lambda = 560$ nm is normally incident. The origin of the Cartesian coordinate $(x,y)$ is at the center of the slit exit. In addition, we define the polar coordinate $(r,\theta)$, where $r = (x^2 + y^2)^{1/2}$ and $\theta = \tan^{-1}(y/x)$.

In the FDTD simulation, the cell size of both directions is 5 nm and the time step is $0.005T$, where $T$ is the period of the incident wave. The PEC film is simulated with the Drude model with the plasma frequency $\omega_p = 1.0\times10^{30}$ Hz and zero damping coefficient. Perfectly matched layers are employed as the boundary conditions. The wave source plane is fixed at $y = 500$ nm; the incident electric field amplitude is set to 1 V/m and the impedance of vacuum $\eta_0$ is normalized to unity, i.e., the incident-magnetic-field amplitude is also normalized to unity. The system is set to be large enough in order to obtain the diffraction pattern in an area of $40\times20$ $\mu$m$^2$.

When $N = 0$ as the case of a single slit, the diffracted wave is *semicylindrical* in the diffraction region $y < 0$ since the slit width $2s$ is much smaller than the wavelength [27]. The snapshots of the magnetic field $\tilde{\mathbf{H}}(x,y,t) = \hat{z}\tilde{H}_z(x,y,t)$ in the diffraction region taken at $t = 60T$ of the simulation are shown for the cases of $N = 1$, 4, 7, and 10 in Figs. 2(a)–2(d), respectively. In Figs. 2(a)–2(c) the diffracted light is concentrated in the central region $|x| < 5$ $\mu$m as a directional beam when the groove number increases. In Fig. 2(d) the beam remains almost the same when there are more grooves. The simulation results are consistent with the previous theoretical prediction in Ref. [27]. The grooves play a crucial role in determining the shape of the diffracted light.

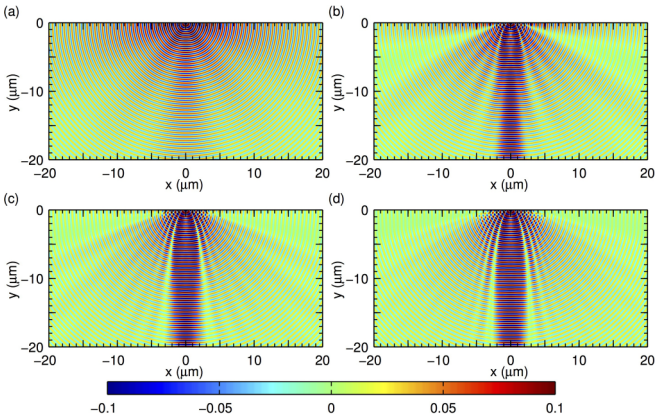


FIG. 2. Snapshot of the magnetic field from the simulation when (a) $N = 1$, (b) $N = 4$, (c) $N = 7$, and (d) $N = 10$.

## III. POINT SOURCE MODELING BASED ON WAVE INTERACTION DYNAMICS

The incident light is transmitted through the slit and the induced surface wave propagates to and is then diffracted by the grooves. The resultant wave pattern after the transmission can be considered as the combined contribution from all the openings including both the slit and the grooves [27]. In our model, we assume that the contribution of each opening can be described by a point source, as shown in Fig. 3(a). The resultant magnetic field of the diffracted wave is summed from the radiation of each point source $\tilde{\mathbf{H}}_\mathbf{j}(x,y,t)$ according to the superposition principle as

$$\tilde{\mathbf{H}}(x,y,t) = \sum_{j=-N}^{N} \tilde{\mathbf{H}}_\mathbf{j}(x,y,t), \tag{1}$$

while the electric field can be obtained by Ampère's law with no currents, i.e., $\nabla\times\tilde{\mathbf{H}} = \partial(\epsilon_0\tilde{\mathbf{E}})/\partial t$.

Because the slit width $2s$ and groove width $2g$ are much smaller than the wavelength, we further assume that each radiant wave originated at each opening center is semicylindrical, as shown in Fig. 3. Thus, for the radiant wave from the $j$th opening (including the slit $j = 0$), a polar coordinate $(r_j,\theta_j)$ is defined to describe the magnetic field radiant from a point source as $\tilde{\mathbf{H}}_\mathbf{j}(x,y,t) = \tilde{\mathbf{H}}_\mathbf{j}(r_j,\theta_j,t)$, where $r_j = [(x - jp)^2 + y^2]^{1/2}$ is the distance between an observation location $(x,y)$ and the coordinate origin located at the opening center $(jp,0)$, and $\theta_j = \tan^{-1}[y/(x - jp)]$ is the angle. Since the metal considered is PEC and the wave is propagating in free space, no wave energy is absorbed, so the semicylindrical wave conserves its energy during the propagation. Therefore, the magnetic field of the radiant wave in the far zone can be expressed as

$$\begin{aligned}\tilde{\mathbf{H}}_\mathbf{j}(r_j,\theta_j,t) &= \hat{z}H_{zj}(r_j)\exp[i(kr_j - \omega t_j)] \\ &= \hat{z}a_j\sqrt{\frac{1}{\pi r_j}}\exp[i(kr_j - \omega t_j)],\end{aligned} \tag{2}$$

where $H_{zj}(r_j) = a_j(\pi r_j)^{-1/2}$ is the amplitude of the magnetic field, $a_j$ is the amplitude of the point source, $k = 2\pi/\lambda$ is the wave number, $\omega$ is the angular frequency, and $t_j = t - \tau_j$ is the time delay related to the temporal phase $\tau_j$ of the radiant wave.

The electric field can be obtained by Maxwell's equations as $\tilde{\mathbf{E}}_\mathbf{j}(r_j,\theta_j,t) = \hat{\theta}_j[i/(2\omega\epsilon_0 r_j) + \eta_0]H_{zj}(r_j)\exp[i(kr_j - \omega t_j)]$, where $\hat{\theta}_j$ is the unit vector in the azimuthal direction and $\eta_0 = (\mu_0/\epsilon_0)^{1/2}$ is the impedance of vacuum. The first (second) term of the electric field is out of phase (in phase) with the magnetic field. In the far zone $r_j \gg \lambda$, the out-of-phase term is negligible. The wave energy is in the in-phase term. The time-averaged Poynting vector becomes $\langle\mathbf{S}_\mathbf{j}(r_j,\theta_j)\rangle = \frac{1}{2}\mathrm{Re}\{\tilde{\mathbf{E}}_\mathbf{j}\times\tilde{\mathbf{H}}_\mathbf{j}^*\} = \hat{r}_j(\eta_0/2)H_{zj}^2(r_j)$, where $\hat{r}_j$ is the unit vector in the radial direction. In other words, the wave energy propagates only in the radial direction. The energy flux of the radiant wave is

$$\begin{aligned}P_j^f &= \int_\pi^{2\pi}\langle\mathbf{S}_\mathbf{j}(r_j,\theta_j)\rangle\cdot\vec{r}_j d\theta_j \\ &= \frac{\eta_0}{2}\int_\pi^{2\pi}H_{zj}^2(r_j)r_j d\theta_j = \frac{\eta_0}{2}a_j^2.\end{aligned} \tag{3}$$

Since $a_j$ is considered a constant, the energy is conserved during the wave propagation. In the theoretical calculation, the radiant field is proportional to the first-kind Hankel function $H_0^{(1)}(kr)$ [27]. When $kr \to \infty$, $H_0^{(1)}(kr) \sim [2/\pi kr]^{1/2}\exp(ikr - i\pi/4)$ [40].

The FDTD simulation of a small system can be utilized to help study the point source radiation: The system is only 1 $\mu$m high, thus putting the wave source and the structure so that we can observe the wave in the structure. The simulation time is $t_f = 25T$; it is long enough to let the wave be at the steady

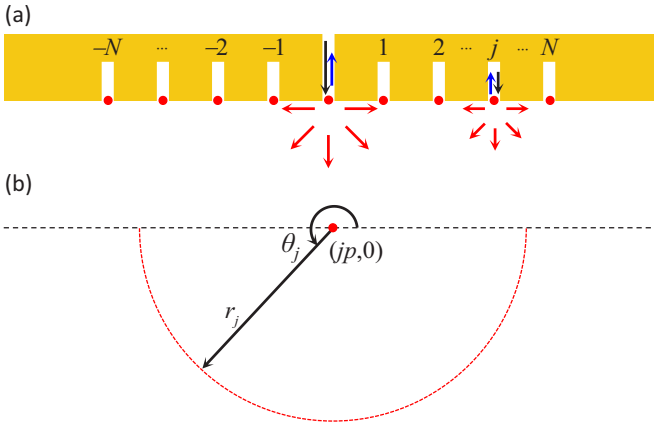


FIG. 3. (a) Schematic of the wave flow for the slit and the $j$th groove for $j \neq 0$. The openings of the slit and the groove radiate waves into free space as point source radiations (red arrows). Point sources (red dots) are then considered at each opening center as the radiations from the openings. A downward traveling wave (black arrow) and an upward traveling wave (blue arrow) wave are assumed to be in both the slit and the groove to obtain the point source radiations. The magnitude of the downward traveling wave is larger than that of the upward traveling wave in the slit, while the magnitudes of the two opposite waves in the groove are equal. (b) Polar coordinate $(r_j,\theta_j)$ originated at the location of the $j$th opening center $(jp,0)$ for all $j$. The $j = 0$ coordinate $(r_0,\theta_0)$ for the slit opening is included, which coincides with the system coordinate $(r,\theta)$.

TABLE I. Floating part of the obtained peak time $t_{j,\text{peak}}$ of the simulated electric field, i.e., $t_{j,\text{peak}} - t_s$, at the location $(jp,0)$ for $j$ from 0 to $N$ and $N = 0,1,\ldots,10$.

| $N$ \ $j$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0.30 | | | | | | | | | | |
| 1 | 0.29 | 0.49 | | | | | | | | | |
| 2 | 0.29 | 0.51 | 0.45 | | | | | | | | |
| 3 | 0.30 | 0.50 | 0.47 | 0.40 | | | | | | | |
| 4 | 0.30 | 0.49 | 0.47 | 0.42 | 0.33 | | | | | | |
| 5 | 0.30 | 0.49 | 0.46 | 0.42 | 0.36 | 0.27 | | | | | |
| 6 | 0.30 | 0.49 | 0.46 | 0.41 | 0.36 | 0.30 | 0.21 | | | | |
| 7 | 0.30 | 0.49 | 0.45 | 0.41 | 0.36 | 0.30 | 0.24 | 0.14 | | | |
| 8 | 0.30 | 0.49 | 0.45 | 0.40 | 0.35 | 0.29 | 0.23 | 0.16 | 0.07 | | |
| 9 | 0.30 | 0.49 | 0.45 | 0.40 | 0.34 | 0.28 | 0.22 | 0.16 | 0.09 | 0.01 | |
| 10 | 0.30 | 0.49 | 0.46 | 0.41 | 0.35 | 0.28 | 0.21 | 0.15 | 0.09 | 0.01 | 0.91 |

state. The diagnostic time is in the time interval $[t_s,t_f]$, the last wave period of the simulation, where $t_s = t_f - T$.

The EM fields in the structure only exist in the slit and the grooves since we use PEC. Therefore, we can define the fields in the slit and in each groove as

$$\tilde{\mathbf{H}}^s_{\mathbf{j}}(x,y,t) = \hat{z}\tilde{H}^s_{zj}(x,y,t), \tag{4a}$$

$$\tilde{\mathbf{E}}^s_{\mathbf{j}}(x,y,t) = \hat{x}\tilde{E}^s_{xj}(x,y,t) + \hat{y}\tilde{E}^s_{yj}(x,y,t), \tag{4b}$$

where $|x| \leqslant s$ and $0 \leqslant y \leqslant h$ for $j = 0$, and $|x - jp| \leqslant g$ and $0 \leqslant y \leqslant d$ for $j \neq 0$.

The tangential electric field at the openings is in theory responsible for the radiations [27]. We use the peak time $t_{j,\text{peak}}$ of the simulated field $\tilde{E}^s_{xj}(jp,0,t)$ for the temporal phase of the point source $j$. Table I lists the floating part of the obtained $t_{j,\text{peak}}$, i.e., $t_{j,\text{peak}} - t_s$, for $j$ from 0 to $N$ and $N = 0,1,\ldots,10$. It shows that $t_{j,\text{peak}}$ varies with the groove number $N$. For instance, $t_{0,\text{peak}} - t_s = 0.30T$ when $N = 0$ and $0.29T$ when $N = 1$ and 2; then, it changes back to $0.30T$ when $N \geqslant 3$. In addition, $t_{1,\text{peak}} - t_s = 0.49T, 0.51T, 0.50T$, and $0.49T$ when $N = 1$, 2, 3, and 4, respectively, and then it remains the same when $N$ increases. For $j \geqslant 2$, $t_{j,\text{peak}} - t_s$ increases and then decreases; in these cases, $t_{j,\text{peak}}$ continues to vary with $N$ when $N$ is larger.

It is also noticed that the difference between $t_{j+1,\text{peak}} + T$ and $t_{j,\text{peak}}$ for $j \geqslant 1$ is roughly close to $pT/\lambda = 0.89T$, a time period for the surface wave to propagate from one groove to the next. Thus, we set the temporal phase of the point source $j$ to be $\tau_j = t_{j,\text{peak}} + lT$, where $l = \lceil (t_{0,\text{peak}} + jpT/\lambda - t_{j,\text{peak}})/T \rceil$ and $\lceil t \rceil$ is the ceiling function that takes the smallest integer greater than or equal to $t$. By setting $\tau_0$ as the reference, Fig. 4 shows the temporal phase difference $\tau'_j = \tau_j - \tau_0$ for $N = 10$ as an example. The result fits well with a linear function $\tau'(j) = 0.95j + 0.20$; the slope of 0.95 indicates the average traveling time between two neighboring grooves. The slightly larger slope should be from the coupling of the wave in the grooves.

Our point source model is based on the energy conservation. The study of the energy flux $P_j$ of the EM fields in the slit ($j = 0$) and those in the grooves ($j \neq 0$) helps determine the amplitudes of the point sources. For the slit, there are two

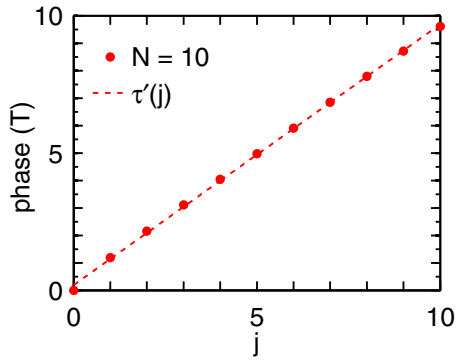


FIG. 4. Temporal phase difference $\tau'_j = \tau_j - \tau_0$ (red circles) for $N = 10$ and the linear fitting function $\tau'(j) = 0.95j + 0.20$ (red dashed line).

methods to determine $P_0$. In the first method, we obtain the time-averaged Poynting vector in the $y$ direction $\langle S^s_{y0}(x,y)\rangle = \langle -\tilde{E}^s_{x0}(x,y,t)\tilde{H}^s_{z0}(x,y,t)\rangle$ and integrate it over the slit area $x = [-s,s]$. Suppose that $P_0$ is positive in the $-y$ direction. We set

$$P_0 = \frac{1}{T}\int_{-s}^{s}\int_{ts}^{t_f} \tilde{E}^s_{x0}(x,h/2,t)\tilde{H}^s_{z0}(x,h/2,t)dt\,dx. \tag{5}$$

The location $y = h/2$ is chosen since only the fundamental mode exists [36] at the location, i.e., $\tilde{E}^s_{x0}(x,h/2,t)$ and $\tilde{H}^s_{z0}(x,h/2,t)$ are independent of $x$, and $\tilde{E}^s_{y0}(x,h/2,t) = 0$. Therefore, Eq. (5) is equal to

$$P_0 = \frac{2s}{T}\int_{ts}^{t_f} \tilde{E}^s_{x0}(0,h/2,t)\tilde{H}^s_{z0}(0,h/2,t)dt. \tag{6}$$

Figure 5(a) shows the yielded $P_0$ for $N$ from 0 to 10 (the red closed circles). We find that $P_0$ is dependent on the groove number: $P_0$ slightly rises from 51.56 to 53.19 when $N$ is from 0 to 1; then it drops to 47.47 (about a $1 - 47.47/53.19 = 11\%$ decrease) when $N$ increases to 5 and it oscillates mildly afterward.

In the second method, we consider the Fabry-Pérot-like resonance in the slit [36–39]. Due to the resonance, the slit is assumed to contain two opposite traveling waves, as indicated

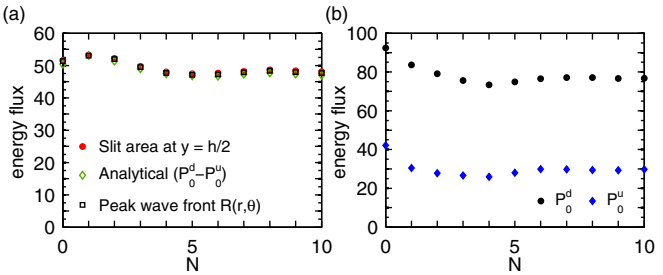


FIG. 5. (a) Energy flux $P_0$ obtained from Eq. (6) inside the slit at the location $(0,h/2)$ (red solid circles) and from the assumption of Eq. (7) for the analytical solution of the two opposite traveling waves inside the slit (green open diamonds) and the energy flux $P$ in the far zone from Eq. (17) (black open squares). (b) Traveling downward $P_0^d$ (black solid circles) and upward $P_0^u$ (dark blue solid diamonds) energy fluxes inside the slit.

in Fig. 3(a). Suppose that the simulated magnetic field in the slit is the superposition of those of the two waves and can be written as

$$\begin{aligned}\tilde{H}^s_{z0}(x,y,t) &= H^s_{z0}(y)\cos\left[\omega t-\phi^s_0(y)\right]\\ &= H^d_{z0}\cos\left(ky+\omega t+\phi^d_0\right)\\ &\quad - H^u_{z0}\cos\left(ky-\omega t+\phi^u_0\right)\end{aligned} \tag{7}$$

for $|x|\leqslant s$ and $0\leqslant y\leqslant h$, where $H^s_{z0}(y)$ and $\phi^s_0(y)$ are the amplitude and phase of the simulated field, respectively, and the constants $H^d_{z0}$ ($H^u_{z0}$) and $\phi^d_0$ ($\phi^u_0$) are the amplitude and phase of the field that travels downward (upward) in the slit. The assumption considers only the fundamental mode, so both $H^s_{z0}(y)$ and $\phi^s_0(y)$ are independent of $x$. As the slit Fabry-Pérot-like resonance, the propagating downward wave is reflected at the slit exit, where its direction switches so that there is a sign change of the magnetic field for the upward traveling wave [41]. The functions $H^s_{z0}(y)$ and $\phi^s_0(y)$ can be observed from the simulation, while $H^d_{z0}$, $H^u_{z0}$, $\phi^d_0$, and $\phi^u_0$ are the four unknowns to be solved. The field amplitude $H^s_{z0}(y)$ for the cases of $N=1$, 4, 7, and 10 is shown in Figs. 6(a)–6(d), respectively, as examples.

The cosine functions in Eq. (7) can be expanded to the linear combination of $\cos(\omega t)$ and $\sin(\omega t)$ and we assume that their respective coefficients on both sides of the equation should be equal. Therefore, one equation can be decomposed into two equations. Since there are four unknowns, the field at two locations is needed for the solution. We choose $y_1=$ 20 nm and $y_2=h-20$ nm, four cells away from the exit and entrance of the slit. With straightforward manipulation, it is not difficult to obtain that

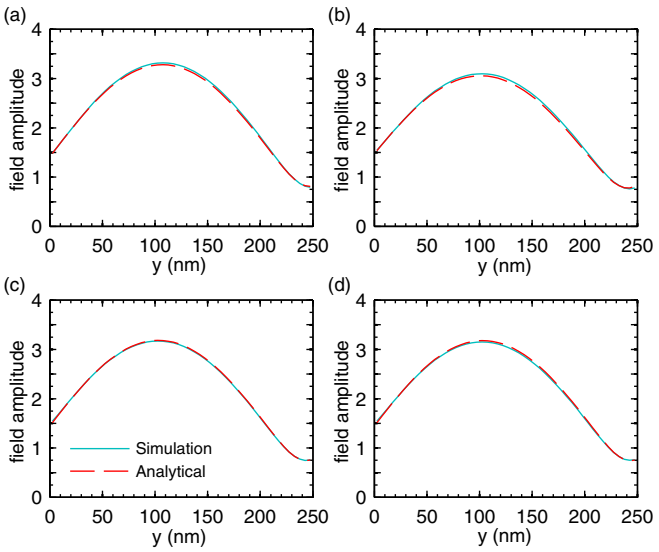


FIG. 6. Field amplitude $H^s_{z0}(y)$ from the simulation (blue solid curve) and that yielded from Eq. (10) (red dashed curve) for (a) $N=1$, (b) $N=4$, (c) $N=7$, and (d) $N=10$.

$$\begin{aligned}H^d_{z0} &= \left(A^2_{ds}+A^2_{dc}\right)^{1/2},\quad \phi^d_0=\tan^{-1}\frac{A_{ds}}{A_{dc}},\\ H^u_{z0} &= \left(A^2_{us}+A^2_{uc}\right)^{1/2},\quad \phi^u_0=\tan^{-1}\frac{A_{us}}{A_{uc}},\end{aligned} \tag{8}$$

where

$$\begin{aligned}A_{ds} &= \frac{H^s_{z0}(y_1)\cos\left[ky_2+\phi^s_0(y_1)\right]-H^s_{z0}(y_2)\cos\left[ky_1+\phi^s_0(y_2)\right]}{2\sin(ky_2-ky_1)},\\ A_{dc} &= \frac{H^s_{z0}(y_1)\sin\left[ky_2+\phi^s_0(y_1)\right]-H^s_{z0}(y_2)\sin\left[ky_1+\phi^s_0(y_2)\right]}{2\sin(ky_2-ky_1)},\\ A_{us} &= \frac{H^s_{z0}(y_1)\cos\left[ky_2-\phi^s_0(y_1)\right]-H^s_{z0}(y_2)\cos\left[ky_1-\phi^s_0(y_2)\right]}{-2\sin(ky_2-ky_1)},\\ A_{uc} &= \frac{H^s_{z0}(y_1)\sin\left[ky_2-\phi^s_0(y_1)\right]-H^s_{z0}(y_2)\sin\left[ky_1-\phi^s_0(y_2)\right]}{-2\sin(ky_2-ky_1)}.\end{aligned} \tag{9}$$

The field amplitude $H^s_{z0}(y)$ from the sum of the two analytical traveling-wave fields with the yielded analytical solutions $H^d_{z0}$, $H^u_{z0}$, $\phi^d_0$, and $\phi^u_0$ is shown in Figs. 6(a)–6(d) for the cases of $N=1$, 4, 7, and 10, respectively, where

$$\begin{aligned}H^s_{z0}(y) = \big[&\left(H^d_{z0}\right)^2+\left(H^u_{z0}\right)^2\\ &-2H^d_{z0}H^u_{z0}\cos\left(2ky+\phi^d_0+\phi^u_0\right)\big]^{1/2}.\end{aligned} \tag{10}$$

The analytical results of $H^s_{z0}(y)$ are in good agreement with the simulation results and thus confirm our assumption for the two opposite traveling waves; the mathematical method is also verified.

With the solutions for the magnetic fields, we can further obtain the electric field $\tilde{E}^s_{x0}(x,y,t)=\eta_0 H^d_{z0}\cos(ky+\omega t+\phi^d_0)+\eta_0 H^u_{z0}\cos(ky-\omega t+\phi^u_0)$ and $\tilde{E}^s_{y0}(x,y,t)=0$. The time-averaged Poynting vector in the $y$ direction $\langle S^s_{y0}(x,y)\rangle=\langle-\tilde{E}^s_{x0}(x,y,t)\tilde{H}^s_{z0}(x,y,t)\rangle$ integrated over the slit area $x=[-s,s]$ as indicated in Eq. (5) yields the analytical energy flux in the slit $P_0=P^d_0-P^u_0$, where

$$P^d_0=\frac{\eta_0}{2}\left(H^d_{z0}\right)^2(2s),\quad P^u_0=\frac{\eta_0}{2}\left(H^u_{z0}\right)^2(2s) \tag{11}$$

are defined as the energy fluxes of the downward and upward traveling waves, respectively. The analytical $P_0$ for $N$ from 0 to 10 is shown in Fig. 5(a) (the green open diamonds). The results are in good agreement with those from the first method.

The grooves can couple the transmitted waves through the slit and reradiate it. The coupling suggests that there are also

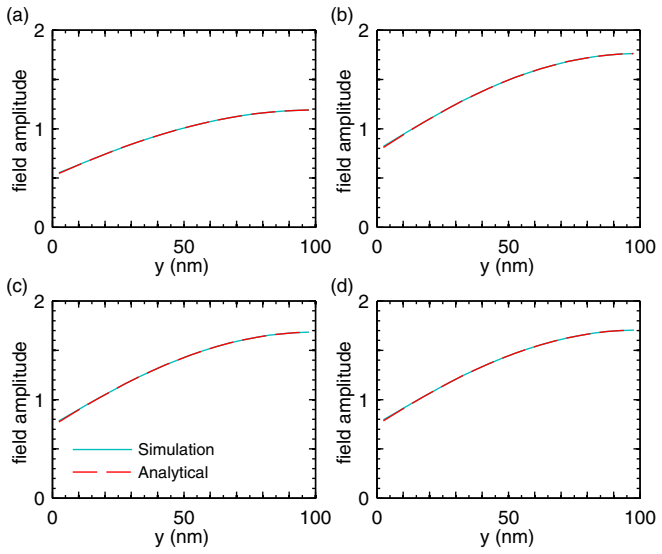


FIG. 7. Field amplitude $H^s_{z1}(y)$ from the simulation (blue solid curve) and that yielded from Eq. (10) altered for the $j = 1$ amplitude and phase (red dashed curve) for (a) $N = 1$, (b) $N = 4$, (c) $N = 7$, and (d) $N = 10$.

two opposite waves for each groove, such as those indicated for the $j$th groove in Fig. 3(a). We use the second method to determine the energy fluxes of the opposite waves. In the $j$th groove, we have the similar assumption for the magnetic field: $\tilde{H}^s_{zj}(x,y,t) = H^s_{zj}(y)\cos[\omega t - \phi^s_j(y)] = H^d_{zj}\cos(ky + \omega t + \phi^d_j) + H^u_{zj}\cos(ky - \omega t + \phi^u_j)$ for $|x - jp| \leqslant g$ and $0 \leqslant y \leqslant d$. There is no minus sign for the magnetic fields of the upward traveling wave; the sign change occurs for the electric field at the ceiling reflection since the tangential electric field vanishes at a PEC surface. The field amplitude $H^s_{z1}(y)$ from the simulation for the cases of $N = 1, 4, 7$, and 10 is shown in Figs. 7(a)–7(d), respectively, as examples.

By setting $y_1 = 20$ nm and $y_2 = d - 20$ nm (still four cells away from the exit and the groove ceiling), the simple alteration of Eqs. (8) and (9) for the $j \neq 0$ cases can lead us to the analytical solutions of the four unknowns $H^d_{zj}$, $H^u_{zj}$, $\phi^d_j$, and $\phi^u_j$. We yield the analytical field amplitude $H^s_{z1}(y)$ by altering Eq. (10) for the $j = 1$ case and show the results in Fig. 7. The simulation is perfectly predicted by the analytical solutions. The assumption and the method for the groove cases are thus verified.

The electric field yielded is $\tilde{E}^s_{xj}(x,y,t) = \eta_0 H^d_{zj}\cos(ky + \omega t + \phi^d_j) - \eta_0 H^u_{zj}\cos(ky - \omega t + \phi^u_j)$ and $\tilde{E}^s_{yj}(x,y,t) = 0$. Similar to the slit case in Eq. (11), the time-averaged Poynting vector in the $y$ direction $\langle S^s_{yj}(x,y)\rangle = \langle -\tilde{E}^s_{xj}(x,y,t)\tilde{H}^s_{zj}(x,y,t)\rangle$ integrated over the groove area $x = [jp - g, jp + g]$ yields the analytical energy flux $P_j = P^d_j - P^u_j$, where

$$P^d_j = \frac{\eta_0}{2}\left(H^d_{zj}\right)^2(2g), \quad P^u_j = \frac{\eta_0}{2}\left(H^u_{zj}\right)^2(2g) \tag{12}$$

are defined as the energy fluxes of the downward and upward traveling waves, respectively. We notice that the yielded $H^d_{zj}$ and $H^u_{zj}$ are identical. This means that the energy fluxes of the two opposite traveling waves are $P^d_j = P^u_j$. The energy flux $P^u_j$ is collectively determined when the wave flows into the $j$th groove and the same amount of energy can only return to the exit as $P^d_j$ after the wave is reflected at the ceiling. Therefore, the net energy flux $P_j$ in the groove is zero. However, when the downward traveling wave continues to be transmitted through the exit, the transmitted wave will have the energy flux $P_j = P^d_j$ due to the energy conservation.

Because each groove is identical and is equidistant from its neighbors, it is physically reasonable to assume that the relation between $P^f_j$ and $P_j$ for each $j \neq 0$ is also identical. Suppose that the relationship is

$$P^f_j = c_f^2 P_j \quad \text{for } j \neq 0, \tag{13}$$

where $c_f$ is a constant factor. As indicated by Eq. (3), the amplitude of the point source for the $j$th groove is then $a_j = (2P^f_j/\eta_0)^{1/2} = (2c_f^2 P_j/\eta_0)^{1/2}$. We estimate that $c_f$ is coincidentally about the ratio $p/\lambda = 500/\lambda$ in the cases studied. This estimation is accurate, as will be evidenced by the numerous accurate modeling results in the next section. A theoretical explanation for the ratio is beyond the scope of this article.

The energy-conservation condition is employed for $P^f_0$ and $a_0$. Let us consider the point source radiation at an observation location represented in the system polar coordinate $(r,\theta)$ for $r \to \infty$. The distance between the location and the $j$th groove opening center can be approximated as $r_j = r - r'_j\cos\theta$, where $r'_j = jp$. Therefore, from Eqs. (1) and (2), the diffracted magnetic field expressed in the polar coordinate is $\tilde{\mathbf{H}}(r,\theta,t) = \hat{z}H_z(r,\theta)\exp[i(kr - \omega t)]$, where

$$H_z(r,\theta) = \sqrt{\frac{1}{\pi r}}\left[a_0 e^{i\phi_0} + 2\sum_{j=1}^{N} a_j e^{i\phi_j}\cos(kr'_j\cos\theta)\right], \tag{14}$$

$\phi_0 = \omega\tau_0$, $\phi_j = \omega\tau_j$, $(\pi r_j)^{-1/2} \simeq (\pi r)^{-1/2}$, the symmetry is utilized, i.e., $a_{-j} = a_j$, and $\phi_{-j} = \phi_j$. We obtain that the wave front of the diffracted wave is also semicylindrical. The electric field yielded is $\tilde{\mathbf{E}}(r,\theta,t) = [\hat{r}E_r(r,\theta) + \hat{\theta}E_\theta(r,\theta)]\exp[i(kr - \omega t)]$, where

$$E_r(r,\theta) = \frac{2i}{\omega\epsilon_0 r}\sqrt{\frac{1}{\pi r}}\sum_{j=1}^{N} a_j e^{i\phi_j} kr'_j \sin\theta\sin(kr'_j\cos\theta), \tag{15a}$$

$$E_\theta(r,\theta) = \left[\frac{i}{2\omega\epsilon_0 r} + \eta_0\right]H_z(r,\theta). \tag{15b}$$

The out-of-phase terms in both $E_r(r,\theta)$ and $E_\theta(r,\theta)$ are negligible when $r \to \infty$. Similar to a single point source radiation, the energy of the diffracted wave is then composed of the in-phase term $\eta_0 H_z(r,\theta)$. The time-averaged Poynting vector becomes $\langle \mathbf{S}(r,\theta)\rangle = \frac{1}{2}\mathrm{Re}\{\tilde{\mathbf{E}}\times\tilde{\mathbf{H}}^*\} = \hat{r}(\eta_0/2)H_z^2(r,\theta)$. That is, the energy propagates only in the radial direction. Suppose that the energy flux of the diffracted wave is $P$. Then

we have

$$P = \int_{\pi}^{2\pi} \langle \mathbf{S}(r,\theta)\rangle \cdot \vec{r}\, d\theta = \frac{\eta_0}{2}\int_{\pi}^{2\pi} H_z^2(r,\theta) r\, d\theta = \frac{\eta_0}{2}\left\{ a_0^2 + a_0 \sum_{j=1}^{N}\left[\frac{4a_j}{\pi}\cos(\phi_j-\phi_0)\int_{\pi}^{2\pi}\cos(kr_j'\cos\theta)d\theta\right]\right.$$
$$\left. + \sum_{j=1}^{N}\sum_{i=1}^{N}\left[\frac{4a_j a_i}{\pi}\cos(\phi_j-\phi_i)\int_{\pi}^{2\pi}\cos(kr_j'\cos\theta)\cos(kr_i'\cos\theta)d\theta\right]\right\}. \quad (16)$$

Due to the energy conservation, $P$ should be equal to the energy flux obtained in the slit $P_0$. We can verify the assumption with the instant diffracted magnetic field $\tilde{H}_z(x,y,t)$ from the simulation of the large system, such as that shown in Fig. 2. By observing the second integral of Eq. (16), we learn that the diffracted energy flux is in fact the integral of the field amplitude squared over a wave front. However, since the system is finite, the wave front of the diffracted wave can only be approximately semicylindrical. Therefore, Eq. (16) can be altered to

$$P = \frac{\eta_0}{2}\int_{\pi}^{2\pi} H_z^2(x_R,y_R) R\, d\theta, \quad (17)$$

where $R(r,\theta)$ is the path of the wave front of the diffracted field at the peak phase that is sought near a semicircle of radius $r \gg 0$ at the angle $\theta$ from $\pi$ to $2\pi$, as indicated in Fig. 1, $x_R = R\cos\theta$, $y_R = R\sin\theta$, and $H_z(x_R,y_R)$ represents the field on $R(r,\theta)$, which is considered the field amplitude. The results measured at $r \simeq 20$ $\mu$m for $N$ from 0 to 10 are shown in Fig. 5(a) (the black open squares). They agree well with those from Eq. (6). Therefore, we obtain $P = P_0$.

In Eq. (16), $P$ can be replaced with $P_0$, the phases $\phi_j$ for all $j$ and the amplitudes $a_j$ for $j \geqslant 1$ are already obtained, and the integrals in the summations can be yielded numerically. Then $a_0$ is the only unknown in Eq. (16) as a quadratic equation and can be readily solved, i.e.,

$$a_0 = \frac{-b + \sqrt{b^2 - 4(c - 2P_0/\eta_0)}}{2}, \quad (18)$$

where $b$ is the coefficient term of $a_0$ and $c$ is the constant term in the curly brackets of Eq. (16). The energy flux $P_0^f = (\eta_0/2)a_0^2$, as a result, can be obtained.

Figure 8 exemplarily demonstrates the obtained $a_j$ for $j =$ 0 to $N$, where $N = 1$, 4, 7, and 10. For the $N = 1$ case, $a_0$ and $a_1$ are 9.25 and 3.36, respectively. When $N$ increases to 4, $a_0$ drops to 8.25 (about a $1 - 8.25/9.25 = 11\%$ decrease) while $a_1$ increases to 4.98 (about a $4.98/3.36 - 1 = 48\%$ increase), and the $j = 2$, 3, and 4 grooves have the point sources of amplitudes $a_2 = 3.50$, $a_3 = 2.42$, and $a_4 = 1.58$, respectively.

For the case of $N = 7$, $a_j$ for $j \leqslant 2$ remains almost the same as those when $N = 4$, but $a_3$ and $a_4$ increase to 2.79 and 2.22 (about 15% and 41% increases), respectively; the amplitudes of the point sources at the $j = 5$, 6, and 7 grooves are $a_5 = 1.70$, $a_6 = 1.24$, and $a_7 = 0.80$, respectively. In the $N = 10$ case, while $a_j$ for $j \leqslant 5$ remains almost the same, $a_6$ and $a_7$ increase to 1.45 and 1.19 (about 17% and 49% increases), respectively, and $a_8 = 0.92$, $a_9 = 0.65$, and $a_{10} = 0.42$.

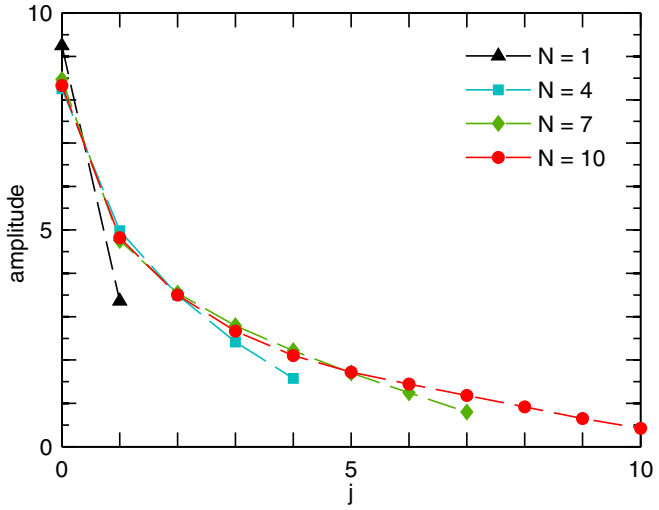


FIG. 8. Amplitude of the point source $a_j$ for $j = 0$–$N$, where $N = 1$ (black triangles), $N = 4$, (blue squares), $N = 7$ (green diamonds), and $N = 10$ (red circles); the colored and dashed lines are to guide the eye.

## IV. MODELING RESULTS

Figure 9 shows the snapshots of the modeled magnetic field by Eq. (1) for $N = 1$, 4, 7, and 10 in comparison to those shown in Fig. 2. Since the model assumes the time-harmonic oscillation $\exp(-i\omega t)$ everywhere in the diffraction region, we set the time $t = 0$. The results are seen as the replicas of those

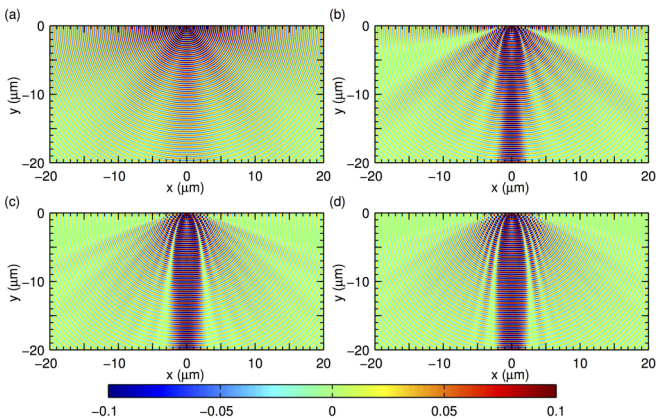


FIG. 9. Snapshot of the magnetic field from the model when (a) $N = 1$, (b) $N = 4$, (c) $N = 7$, and (d) $N = 10$.

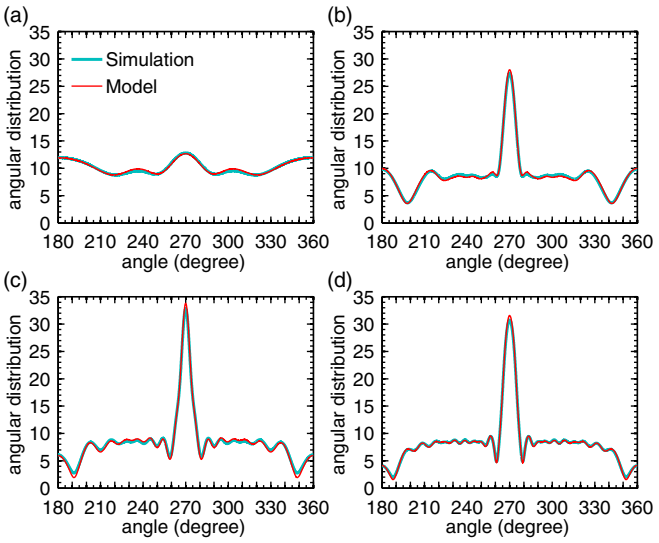


FIG. 10. Angular distribution from the simulation (blue thick curve) and the model (red thin curve) for (a) $N = 1$, (b) $N = 4$, (c) $N = 7$, and (d) $N = 10$.

from the simulation by comparing the distribution patterns and the field magnitudes from the near zone to the far zone.

To examine the results in more detail, we use the path of the wave front $R$ in Eq. (17) to define the angular distribution

$$f(\theta) = \sqrt{\pi R} H_z(x_R, y_R) \tag{19}$$

for both the simulation and the model, where $t = 60T$ for the simulation; its purpose is to eliminate the influence of the distance decay. Figure 10 shows the distributions at $r \simeq 20$ $\mu$m also for $N = 1$, 4, 7, and 10. In each case, the model is in excellent agreement with the simulation; it perfectly reproduces the profiles even in the details. It is worth mentioning that the influence of $N$ on the angular distribution also agrees with the theoretical work [27]; the peak increases and reaches a maximum when $N = 7$ for the convergence and then decreases slightly when $N = 10$.

We further study the diffraction problem for the cases of $\lambda = 680$ and 800 nm. In the simulation, the setting is the same as that used for the field in Fig. 2 except for a minor difference: The simulation time is changed to $45T$ and $40T$ for $\lambda = 680$ and 800 nm, respectively, since their wave periods are proportional to the wavelengths and larger than that for $\lambda = 560$ nm. In addition, the time step of the $\lambda = 800$ nm case is changed to $0.0025T$ to satisfy the Courant condition [42]. Figures 11(a) and 11(b) show snapshots of the simulated magnetic field for the two wavelengths and for $N = 10$. The diffracted light becomes two beams and their angle with respect to each other increases with the wavelength; the results are also consistent with the theoretical work [27].

The small-system simulations for the modeling are performed similarly to those for the cases of $\lambda = 560$ nm, but the corresponding simulation time is changed to $t_f = 21T$ and $18T$ for $\lambda = 680$ and 800 nm, respectively. For $N = 10$, Fig. 12(a) shows the temporal phase difference $\tau'_j = \tau_j - \tau_0$ from the observed peak time $t_{j,\mathrm{peak}}$ of the electric field in the two simulations. The results fit well with linear functions $\tau'(j) = 0.76j + 0.11$ and $0.64j + 0.11$ for the two

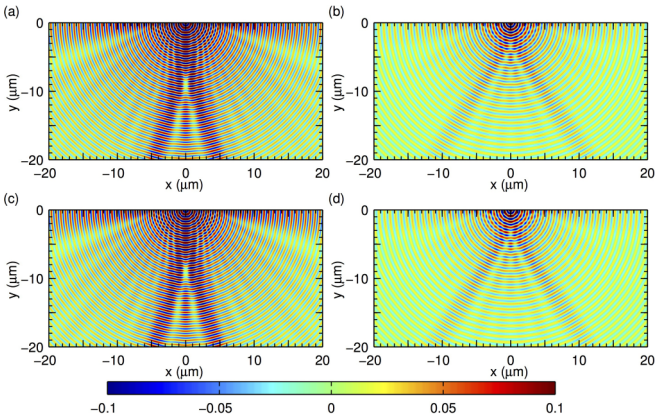


FIG. 11. Snapshot of the magnetic field from the simulation for the cases of (a) $\lambda = 680$ nm and (b) 800 nm for $N = 10$. (c) and (d) Snapshots of the magnetic field from the model for the corresponding wavelengths.

wavelengths, respectively. While the times for the surface wave from one groove to the next are $pT/\lambda = 0.74T$ and $0.63T$, respectively, the slopes verify our assumption of the surface wave time delay for different incident wavelength.

The amplitudes of the point sources for the two wavelengths and for $N = 10$ are shown in Fig. 12(b) followed by the same procedure demonstrated. When $\lambda = 680$ nm, $a_0$ is higher than that when $\lambda = 560$ nm, but $a_1$ drops immediately to one-seventh its value; then $a_j$ decays slower than in the previous case. For the case of $\lambda = 800$ nm, both $a_0$ and $a_j$ for $j > 0$ are the smallest and they have the slowest decaying rates of all the cases. With the two wavelengths, the wave coupling is weak, so both the temporal phases and the amplitudes are almost independent of $N$.

Snapshots of the magnetic field from the model for the two wavelengths are shown in Figs. 11(c) and 11(d), respectively, in comparison with the simulation results. As can be seen, the model can also duplicate the field distribution for different wavelengths. The angular distributions from both the simulation and the model are examined as shown in Fig. 13. The excellent agreement from the comparison of detail verifies the authenticity of the model.

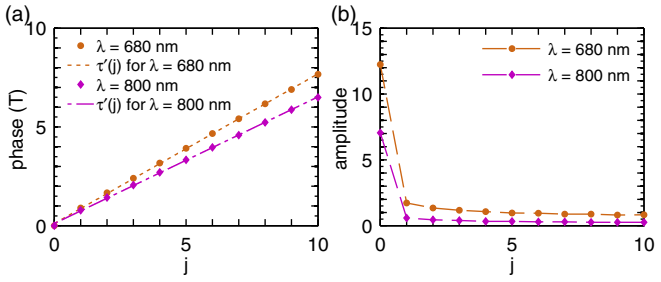


FIG. 12. (a) Temporal phase difference $\tau'_j = \tau_j - \tau_0$ for $\lambda = 680$ nm (brown circles) and 800 nm (purple diamonds) and for $N = 10$, along with the linear fitting functions $\tau'(j) = 0.76j + 0.11$ (brown short-dashed line) and $0.64j + 0.11$ (purple short-dash–long-dashed line). (b) Amplitude of the point source $a_j$ for $\lambda = 680$ nm (brown circles) and 800 nm (purple diamonds) and for $N = 10$; the colored and dashed lines are to guide the eye.

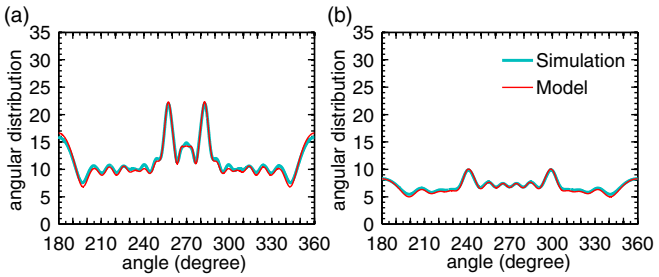


FIG. 13. Angular distribution from the simulation (blue thick curve) and the model (red thin curve) for (a) $\lambda = 680$ nm and (b) 800 nm.

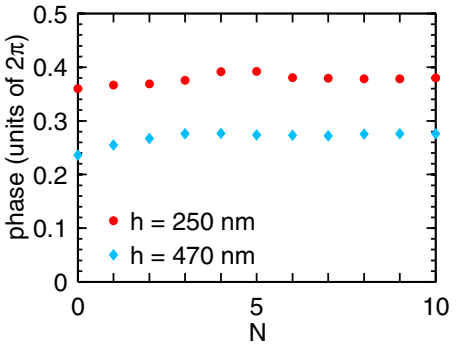


FIG. 14. Phase of the upward traveling wave in the slit $\phi_0^u$ when $h = 250$ nm (red circles) and 470 nm (blue diamonds).

## V. PHYSICS OF THE WAVE INTERACTION DYNAMICS AND DISCUSSION

In the study of the wave interaction dynamics, we find that adding the grooves creates surface waves that can propagate backward to interfere with those at the slit and the already existing grooves. The slight variation of the peak time $t_{j,\text{peak}}$ with the increase of the groove number $N$, as shown in Table I, is an implication.

We believe that this kind of interference at the slit is also able to change the original resonance condition so that the transmitted energy flux $P_0$ is altered accordingly. As we have shown in Fig. 5(a), $P_0$ slightly increases and then decreases more when $N$ increases, while adding the grooves enhances the energy density of the diffracted wave in the central region to the directional light beam as shown in Fig. 2. To analyze further, we show in Fig. 5(b) the traveling downward and upward energy fluxes $P_0^d$ and $P_0^u$, respectively, yielded from Eq. (11) for $P_0$. We obtain that both $P_0^d$ and $P_0^u$ decrease at the beginning. For $N$ from 0 to 1, $P_0^d$ is from 92.44 to 83.63 while $P_0^u$ is from 42.08 to 30.50, respectively. Therefore, since the decrease of $P_0^d$ is less than that of $P_0^u$, $P_0$ increases. As $N$ continues to increase to 4, $P_0^d$ decreases to 73.43 and $P_0^u$ to 25.95; the quicker decrease of $P_0^d$ than of $P_0^u$ results in the decrease of $P_0$. Although both $P_0^d$ and $P_0^u$ slightly increase to 74.87 and 28.06 for $N = 5$, respectively, the smaller increase of $P_0^d$ still causes the decrease of $P_0$. Then $P_0^d$ and $P_0^u$ vary slightly when $N > 5$, so $P_0$ remains almost constant.

The decrease of $P_0^u$ should result from the destructive interference between the wave in the slit and those that propagate from the grooves to flow into it. As indicated in Eq. (7), there is a sign change of the magnetic field of the upward traveling wave. Therefore, while the magnetic field radiated from the slit and those from the grooves are almost in phase in the central diffraction region, the field that is reflected at the exit and those from the grooves are almost out of phase in the slit. Thus, $P_0^u$ decreases at the beginning. Since the phase of the surface wave shifts according to the groove location, the in-phase interference can occur when $N$ is large.

We show the yielded phase of the upward traveling wave in the slit $\phi_0^u$ in Fig. 14 (the red circles). As $N$ increases, $\phi_0^u$ increases. That is, the upward traveling wave is delayed due to the interference of the surface waves from the grooves. After the upward traveling wave is reflected at the slit entrance to propagate downward to the exit as a round-trip, the phase is still delayed. The induced phase delay of the round-trip wave should have an influence on the slit resonance.

The single-slit energy flux $P_0$ ($N = 0$) as a function of the film thickness $h$ is shown in Fig. 15, where $P_0$ is obtained by the first method introduced in Sec. III. The film thickness $h$ determines the phase delay of the round-trip wave due to the Fabry-Pérot-like resonance [36–39]. The single-slit $P_0$ therefore varies periodically with $h$ and the period is $\lambda/2$. The two peaks are at $h_{\text{res1}} = 220$ nm and $h_{\text{res2}} = 500$ nm, respectively. For our case, we choose $h = 250$ nm (30 nm larger than $h_{\text{res1}}$); at this $h$, the single-slit $P_0$ decreases as $h$ slightly increases. Since adding the grooves also increases the phase delay of the round-trip wave, $P_0$ should decrease with the increase of $N$ as the single-slit $P_0$ does with the increase of $h$. This explains the decrease of $P_0$ shown in Fig. 5(a). Since both $P_0$ and $P_0^u$ decrease, $P_0^d$ decreases accordingly.

We notice that, for the case studied in the previous work [28] (for the same configuration but for $h$ being 350 nm), $P_0$ is almost independent of $N$ when $N \geqslant 1$. From our viewpoint, we believe that this is because, at the film thickness, the single-slit $P_0$ is close to the minimum and is not sensitive to $h$, i.e., the phase delay of the round-trip wave does not influence $P_0$.

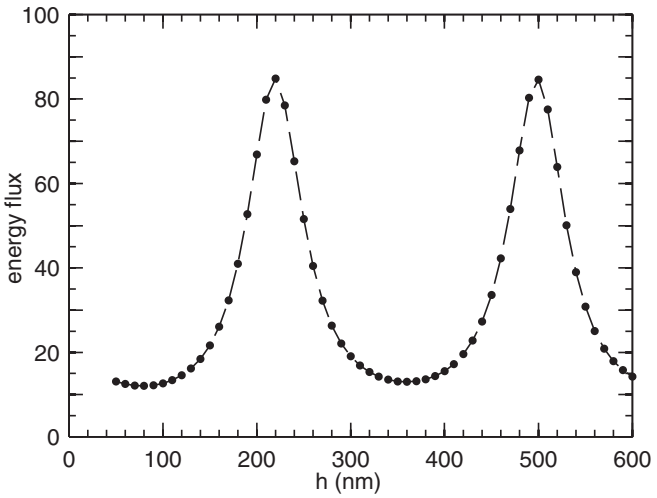


FIG. 15. Energy flux $P_0$ obtained from Eq. (6) as a function of the film thickness $h$ when $N = 0$ (the single slit).

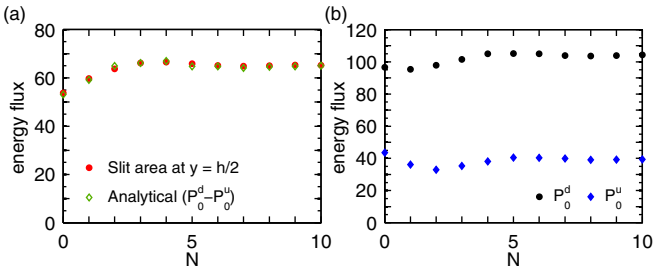


FIG. 16. (a) Energy flux $P_0$ obtained from Eq. (6) (red closed circles) and from the assumption of Eq. (7) (green open diamonds) when $h = 470$ nm, where $y_1$ and $y_2$ are adjusted accordingly. (b) Traveling downward $P_0^d$ (black closed circles) and upward $P_0^u$ (dark blue closed diamonds) energy fluxes inside the slit.

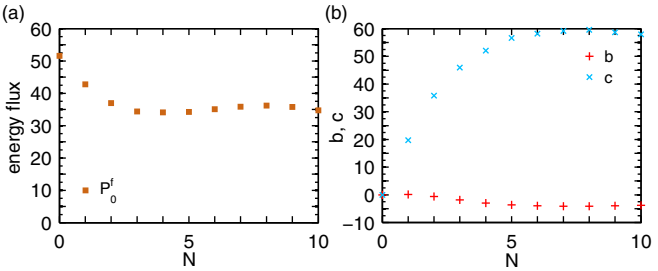


FIG. 17. (a) Energy flux radiated from the slit into free space $P_0^f$. (b) Coefficient term $b$ (red pluses) and the constant term $c$ (blue crosses) in Eq. (18) obtained from Eq. (16).

Therefore, even when adding the grooves delays the round-trip wave, $P_0$ in this case remains unchanged.

The case of $h = 470$ nm (30 nm smaller than $h_{res2}$) is chosen to further study the influence of the wave dynamics on $P_0$. The phase of the upward traveling wave in the slit $\phi_0^u$ is shown in Fig. 14 (the blue diamonds). The increase of the phase indicates that adding the grooves delays the slit round-trip wave, similar to the $h = 250$ nm case. Therefore, $P_0$ should vary with $N$ as the single-slit $P_0$ does with the increase of $h$. At $h = 470$ nm, the single-slit $P_0$ increases as $h$ slightly increases, as shown in Fig. 15. The yielded $P_0$ as a function of $N$ is shown in Fig. 16(a), where $P_0$ is obtained by both the first and second methods in Sec. III. As can be seen, $P_0$ indeed increases in this case, contrary to the $h = 250$ nm case. Our physical interpretation of the wave dynamics for the $P_0$ variation is thus verified. Besides the increase of the film thickness $h$, adding the grooves also increases the phase delay of the round-trip wave in the cases studied; the wave dynamics will result in the change of $P_0$ with $N$ in the way the single-slit $P_0$ varies with $h$.

The yielded $P_0^d$ and $P_0^u$ from the second method are shown in Fig. 16(b). In this $h = 470$ nm case, $P_0^u$ decreases at the beginning as we anticipate, since the condition of the destructive interference between the wave reflected at the exit and those from the grooves are still the same; however, it increases earlier. In addition, $P_0^d$ decreases slightly when $N$ is from 0 to 1; then it increases immediately due to the increase of $P_0$. These results are qualitatively consistent with our physical interpretation.

The surface waves from the grooves also propagate backward to flow into the nearby grooves. The increase of $P_1$ when $N$ increases from 1 to 4, as the increase of $a_1$ indicated in Fig. 8, suggests the influence of the surface waves from the higher-$j$ grooves. Since the separation distance $p$ is almost a wavelength, the surface waves that propagate to the groove have the constructive interference to increase the energy flux $P_1$ [about a $(4.98/3.36)^2 - 1 = 120\%$ increase]. When $N$ is from 4 to 7, the point sources are created farther away from the center. The surface waves correspondingly created have less influence on the first few grooves than those nearby. That is, $P_1$ and $P_2$ remain almost the same, while the increasing ratio of $P_3$ is less than that of $P_4$ [about $(2.79/2.42)^2 - 1 = 33\%$ and $(2.22/1.58)^2 - 1 = 98\%$, respectively].

With the similarity in the $N = 10$ case, the surface waves created are so far that the $P_j$ for $j$ up to 5 are unchanged, while the increasing ratios of $P_6$ and $P_7$ [about $(1.45/1.24)^2 - 1 = 37\%$ and $(1.19/0.80)^2 - 1 = 121\%$ increases, respectively] can be analogous to those of $P_3$ and $P_4$ in the previous $N = 7$ case. These results verify the existence of the surface waves and indicate that they are more influential on the more nearby grooves.

It is important to keep in mind that $P_0^f$ is only the energy flux radiated from the slit into free space. While $P_0$ decreases with $N$ in our original $h = 250$ nm case, $P_0^f$ should further decrease so that the energy is conserved, as indicated in Eq. (18) for the solution of $a_0$ and $P_0^f = (\eta_0/2)a_0^2$. As shown in Fig. 17(a), $P_0^f$ drops from 51.56 to the lowest 34.07 when $N = 4$ (about a $1 - 34.07/51.56 = 34\%$ decrease); the decreasing ratio is much larger than that of $P_0$.

To analyze, we show in Fig. 17(b) the coefficient term $b$ and the constant term $c$ in Eq. (18) that are obtained from Eq. (16). Since $b$ is close to zero, $b$ in the square root term is negligible; the slit source amplitude can be approximated to $a_0 = (2P_0/\eta_0 - c)^{1/2} - b/2$. Then we obtain $P_0^f \cong (\eta_0/2)[(2P_0/\eta_0 - c) - b(2P_0/\eta_0 - c)^{1/2}]$, where the $b^2/4$ term is ignored. The term $c$ increases to become comparable to $2P_0/\eta_0$ so that $P_0^f$ decreases more quickly. That is, since $c$ represents the energy of the superposed waves radiated from the grooves, the grooves share $P_0$ efficiently to reduce $P_0^f$. When $N$ is large, $b$ increases negatively, which indicates that the radiated waves from the grooves in fact contribute to $P_0^f$ so that it slightly increases.

From Eqs. (14) and (16) it is not difficult to obtain that the intensity of the diffracted wave is

$$I(r,\theta) = \frac{\eta_0}{2} H_z^2(r,\theta) = \frac{\eta_0}{2}\frac{1}{\pi r}\left[a_0^2 + 2a_0\sum_{j=1}^{N} a_j B_j(\theta) + \sum_{j=1}^{N}\sum_{i=1}^{N} a_j a_i C_{ji}(\theta)\right], \tag{20}$$

where

$$B_j(\theta) = \cos[kr'_j\cos\theta + (\phi_j - \phi_0)] + \cos[kr'_j\cos\theta - (\phi_j - \phi_0)],$$

$$C_{ji}(\theta) = \cos[kr'_j \cos\theta + kr'_i \cos\theta + (\phi_j - \phi_i)] + \cos[kr'_j \cos\theta + kr'_i \cos\theta - (\phi_j - \phi_i)] + \cos[kr'_j \cos\theta - kr'_i \cos\theta + (\phi_j - \phi_i)] + \cos[kr'_j \cos\theta - kr'_i \cos\theta - (\phi_j - \phi_i)]. \quad (21)$$

That is, the angular distribution is related to the light path difference due to the groove location and the temporal phase difference between each point source. In our case, since the separation distance $p$ is about a wavelength and the temporal phase difference is almost an integer multiple of $T$, we obtain $B_j(\theta) \cong 2\cos(2j\pi\cos\theta)$ and $C_{ji}(\theta) \cong 2\cos[2(j+i)\pi\cos\theta] + 2\cos[2(j-i)\pi\cos\theta]$. At the peak, $B_j(\theta)$ and $C_{ji}(\theta)$ should be at the maximum, i.e., the most constructive interference between the radiated waves of each point source. Therefore, the angle yielded is $\theta = 270^\circ$. The cause of the central directional light beam is explained.

Also, the change of $a_j$ with $N$ may be linked to the convergence mechanism of the diffraction patterns shown in Fig. 2. As indicated in Fig. 8, more point sources cause the more constructive interference in the central diffraction region; eventually, a directional light beam can be formed. However, the quickly decaying $a_j$ means that only the point sources at the first few grooves have the major contribution. Taking the $N = 7$ case for example, the source amplitude that decays from $a_1 = 4.76$ to $1/e$ it value is $a_5 = 1.70$. That is, only the $j \leqslant 5$ grooves are accountable for the main formation of the light beam. Therefore, even when $N$ further increases, the correspondingly created point source does not have an influence, so the diffraction pattern caused by the $N = 10$ case is similar to that by the $N = 7$ case.

For the longer-wavelength cases demonstrated in the preceding section, the surface wave from the slit creates the point sources at the grooves that are not in phase, as the temporal phase relationship indicated in Fig. 12(a). Therefore, we obtain that the amplitudes of the groove point sources, as shown in Fig. 12(b), are much smaller. Besides, the surface waves from the grooves are not in phase with those at the other grooves either, to contribute their energy. In other words, the energy fluxes that flow into and propagate out from the grooves are weak, so the groove source amplitudes are independent of the groove number.

It is thus instinctive to consider that the grooves in the structure have the weak-wave-coupling ability to the longer wavelength. Since the energy from the grooves is weak, it should not influence the slit either. That is, $P_0$ is also anticipated to be independent of $N$. As shown in Figs. 18(a) and 18(b) for the cases of $\lambda = 680$ and 800 nm, respectively, $P_0$ is constant. The yielded $P_0^d$ and $P_0^u$ shown in Figs. 18(c) and 18(d) for the two cases are both constant too. This indicates that the surface waves do not change the amplitude and phase of the upward traveling wave; thus, they do not influence the wave resonance in the slit.

For the diffracted wave, however, the point source radiations from the grooves do have the contribution to cause angular light beams; however, the contribution is mainly from the interaction with the slit radiation. According to Eqs. (20) and (21), since $a_j$ for $j \geqslant 1$ is small, the angular distribution is mostly determined by the second term $B_j(\theta)$. For the peak of

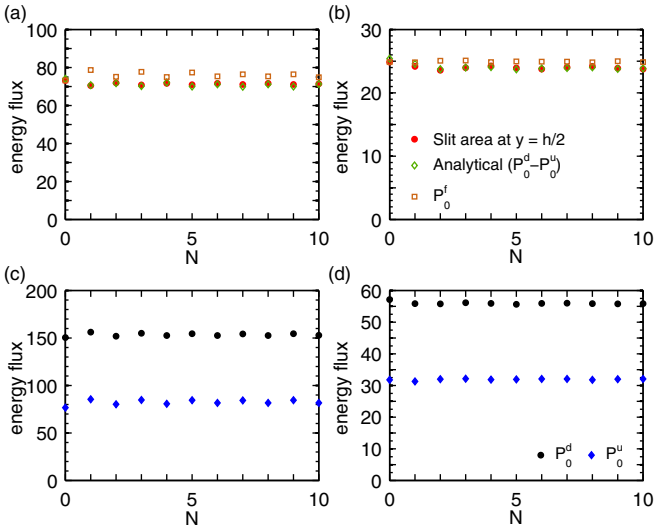


FIG. 18. Energy fluxes $P_0$ obtained from Eq. (6) (red closed circles) and the assumption of Eq. (7) (green open diamonds), where $y_1$ and $y_2$ are adjusted accordingly, and the radiated energy flux from the slit into free space $P_0^f$ (brown open squares) for the cases of (a) $\lambda = 680$ nm and (b) 800 nm, respectively. Also shown are energy fluxes traveling downward $P_0^d$ (black closed circles) and upward $P_0^u$ (dark blue closed diamonds) inside the slit, for the cases of (c) $\lambda = 680$ nm and (d) 800 nm, respectively.

$B_j(\theta)$, it should satisfy the condition

$$kr'_j \cos\theta \pm (\phi_j - \phi_0) = \pm 2j\pi. \quad (22)$$

The condition considers the phase difference $\phi_j - \phi_0$ that includes the wave coupling due to the groove depth, which is in contrast to the previous studies that use the conventional grating theory to study the beaming angle and ignore the groove coupling [22–26,33–35]. Using the temporal phase relationship in Fig. 12(a), we obtain that the angle is $\theta = 270^\circ \pm 19.05^\circ$ and $270^\circ \pm 35.17^\circ$ for the cases of $\lambda = 680$ and 800 nm, respectively. Note that this approximation assumes $r \to \infty$, but the results are still consistent with the peak angles of the distributions shown in Fig. 13. As a matter of fact, our test indicates that, when $r$ increases, these peak angles will converge to the predicted angles.

It is interesting to find that the yielded $P_0^f$ in both cases are larger than $P_0$, as shown in Figs. 18(a) and 18(b) (the brown open squares) for $\lambda = 680$ and 800 nm, respectively. We show in Figs. 19(a) and 19(b) the coefficient term $b$ and the constant term $c$ in Eq. (18) for the two cases. When $N$ is small, the ratio of $2P_0/\eta_0$ and $c$ is large, so we can further have approximately $P_0^f \cong (\eta_0/2)[(2P_0/\eta_0) - b(2P_0/\eta_0)^{1/2}]$ in the previous analysis for the case of $\lambda = 560$ nm. Since $b$ is negative, we obtain $P_0^f > P_0$. Indeed, $c$ increases with $N$, but it is still much smaller than $2P_0/\eta_0$; besides, $b$ also becomes more negative. Therefore, $P_0^f$ is larger than $P_0$ when $N$ continues to increase. In either of the two scenarios, Eqs. (16) and (18) suggest that, while the point sources at the grooves do not share $P_0$ efficiently, their radiated waves still contribute energy to the slit. As a result, the energy flux radiated from the slit into free space is larger than that transmitted through it.

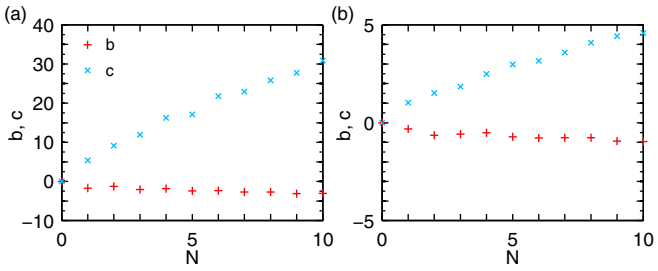


FIG. 19. Coefficient term $b$ (red pluses) and the constant term $c$ (blue crosses) in Eq. (18) obtained from Eq. (16) for (a) $\lambda = 680$ nm and (b) 800 nm.

## VI. SUMMARY

A point source model has been developed to study the physics of the light diffracted by a subwavelength slit and a number of grooves, in which the dynamics of the wave interacting with the metallic structure is emphasized. With the help of small-system simulations, we determined the temporal phases of the point sources by the tangential electric field at the openings, and the amplitudes by the conservation of energy fluxes from the slit and the grooves. Although the far-zone approximation was employed, our modeling results still work in the near zone. Besides the cases of the standard incident wavelength, the cases of the longer incident wavelengths are also in good agreement. The wave analysis based on this model enables us to reveal interesting physics. In the standard case studied, the surface waves created by the added grooves propagate backward into the slit and then delay the reflection. The transmitted energy flux thus decreases with the groove number in a way similar to the single slit case of increasing the film thickness. We also found that the independence of the transmitted energy with the groove number in the previous study [28] results from the choice of the phase whose transmitted energy is not sensitive to the film thickness. This finding was further verified by the study of the case whose transmitted energy is at the increasing phase. These surface waves also flow back into the nearby grooves closer to the slit to increase their energy fluxes. In this strong-wave-coupling case, the superposed waves radiated from the grooves increasingly share the transmitted energy flux. As the total energy decreases with the groove number, the radiated energy flux from the slit decreases further. In the cases of the longer wavelengths, the surface wave coupling is not influential, so the grooves poorly share the transmitted energy, but their radiations still interfere with the slit radiation, so the diffracted light is split into two beams. Applied from our formula of the energy-conservation condition, the angular distribution of the field intensity in the far zone is approximated; this approximation gives the origin of the directional light beaming, which suggests that the beam angles are determined by the point source location and the temporal phase difference that considers the time delay due to the surface wave propagation from the slit and the groove coupling. In these weak-wave-coupling cases, the energy of the slit radiation is in fact larger than that transmitted through it; as indicated by the calculation results for the energy-conservation condition, the contribution from the groove radiation enhances the slit radiation energy.

## ACKNOWLEDGMENTS

The authors are grateful for support from Ministry of Science and Technology, Taiwan (MOST 104-2112-M-006-004-MY3) and computational resources at National Center for High-performance Computing (NCHC) of National Applied Research Laboratories (NARLabs) of Taiwan.

[1] T. W. Ebbesen, H. J. Lezec, H. F. Ghaemi, T. Thio, and P. A. Wolff, Nature (London) **391**, 667 (1998).
[2] S. A. Maier and H. A. Atwater, J. Appl. Phys. **98**, 011101 (2005).
[3] E. Hutter and J. H. Fendler, Adv. Mater. **16**, 1685 (2004).
[4] D. K. Gramotnev and S. I. Bozhevolnyi, Nat. Photon. **4**, 83 (2010).
[5] J. N. Anker, W. P. Hall, O. Lyandres, N. C. Shah, J. Zhao, and R. P. Van Duyne, Nat. Mater. **7**, 442 (2008).
[6] H. A. Atwater and A. Polman, Nat. Mater. **9**, 205 (2010).
[7] L. Novotny and N. van Hulst, Nat. Photon. **5**, 83 (2011).
[8] W. A. Challener, C. Peng, A. V. Itagi, D. Karns, W. Peng, Y. Peng, X. Yang, X. Zhu, N. J. Gokemeijer, Y. T. Hsia, G. Ju, R. E. Rottmayer, M. A. Seigler, and E. C. Gage, Nat. Photon. **3**, 220 (2009).
[9] N. Fang, H. Lee, C. Sun, and X. Zhang, Science **308**, 534 (2005).
[10] F. J. García-Vidal, L. Martín-Moreno, H. J. Lezec, and T. W. Ebbesen, Appl. Phys. Lett. **83**, 4500 (2003).
[11] H. Shi, C. Du, and X. Luo, Appl. Phys. Lett. **91**, 093111 (2007).
[12] K. R. Chen, Opt. Lett. **35**, 3763 (2010).
[13] K. R. Chen, W. H. Chu, H. C. Fang, C. P. Liu, C. H. Huang, H. C. Chui, C. H. Chuang, Y. L. Lo, C. Y. Lin, H. H. Hwung, and A. Y. G. Fuh, Opt. Lett. **36**, 4497 (2011).
[14] Y. Chen, X. Li, Y. Sonnefraud, A. I. Fernández-Domínguez, X. Luo, M. Hong, and S. A. Maier, Sci. Rep. **5**, 8660 (2015).
[15] A. Haddadpour and G. Veronis, Opt. Express **23**, 5789 (2015).
[16] X. Zhang, L. Yan, Y. Guo, W. Pan, B. Luo, and X. Luo, Plasmonics **11**, 109 (2016).
[17] H. J. Lezec, A. Degiron, E. Devaux, R. A. Linke, L. Martin-Moreno, F. J. Garcia-Vidal, and T. W. Ebbesen, Science **297**, 820 (2002).
[18] B. Wang and G. P. Wang, Appl. Phys. Lett. **88**, 013114 (2006).
[19] S. Kang and K. Oh, Opt. Commun. **284**, 5388 (2011).
[20] J.-M. Yi, A. Cuche, E. Devaux, C. Genet, and T. W. Ebbesen, ACS Photon. **1**, 365 (2014).
[21] G. Yuan, E. T. F. Rogers, T. Roy, G. Adamo, Z. Shen, and N. I. Zheludev, Sci. Rep. **4**, 6333 (2014).
[22] D.-Z. Lin, T.-D. Cheng, C.-K. Chang, J.-T. Yeh, J.-M. Liu, C.-S. Yeh, and C.-K. Lee, Opt. Express **15**, 2585 (2007).
[23] Y. Liu, H. Shi, C. Wang, C. Du, and X. Luo, Opt. Express **16**, 4487 (2008).
[24] W. Dai and C. M. Soukoulis, Phys. Rev. B **82**, 045427 (2010).
[25] Y.-j. Su and H.-c. Chang, IEEE Photon. J. **5**, 4801015 (2013).
[26] K. Kim, S.-Y. Lee, H. Yun, J.-B. Park, and B. Lee, Opt. Express **22**, 5465 (2014).

[27] L. Martín-Moreno, F. J. García-Vidal, H. J. Lezec, A. Degiron, and T. W. Ebbesen, Phys. Rev. Lett. **90**, 167401 (2003).
[28] F. J. García-Vidal, H. J. Lezec, T. W. Ebbesen, and L. Martín-Moreno, Phys. Rev. Lett. **90**, 213901 (2003).
[29] D. Y. Na, J. H. Kim, Y. B. Park, and K. Y. Jung, IET Microw. Antennas Propag. **7**, 843 (2013).
[30] Y.-L. Hua and Z.-Y. Li, J. Appl. Phys. **105**, 013104 (2009).
[31] H. Shi, X. Dong, Y. Lv, and C. Du, Appl. Phys. B **95**, 345 (2009).
[32] L. Cai, G. Li, Z. Wang, and A. Xu, Opt. Lett. **35**, 127 (2010).
[33] F. Hao, R. Wang, and J. Wang, Plasmonics **5**, 45 (2010).
[34] F. Hao, R. Wang, and J. Wang, J. Opt. **13**, 015002 (2011).
[35] Y. S. Hwang, J. Kim, and K. Y. Kim, IEEE Photon. Technol. Lett. **26**, 2051 (2014).
[36] Y. Takakura, Phys. Rev. Lett. **86**, 5601 (2001).
[37] J. Bravo-Abad, L. Martín-Moreno, and F. J. García-Vidal, Phys. Rev. E **69**, 026601 (2004).
[38] R. Gordon, Phys. Rev. B **73**, 153405 (2006).
[39] J.-S. Hong, A. E. Chen, and K.-R. Chen, Opt. Express **23**, 9901 (2015).
[40] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables* (Sfetcu, San Francisco, 2014).
[41] R. Gordon, J. Opt. A **8**, L1 (2006).
[42] A. Taflove and S. C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 3rd ed. (Artech House, Boston, 2005).