


Rayleigh-Sommerfeld diffraction on a subwavelength scale: Theories and a resolution criterion

Yan-ming Gao, Jun-peng Xie, and Xiang-yang Yu*

School of Physics, State Key Laboratory of Optoelectronic Materials and Technologies, Sun Yat-Sen University, Guangzhou 510275, China (Received 5 September 2018; revised manuscript received 30 November 2018; published 7 February 2019)

Classical resolution criteria such as Rayleigh's and Sparrow's are based on the fringe structure produced by Fresnel diffraction or Fraunhofer diffraction. However, these two diffraction theories are based on the assumption of large-aperture scale and are thus incapable of describing the diffraction of subwavelength structure. We find a singularity near the incident wave vector k_0 in a subwavelength aperture by considering Rayleigh-Sommerfeld diffraction. The diffraction fringe structure disappears once the aperture scale is smaller than a threshold value. A two-point resolution criterion unrelated to the fringe structure is proposed based on the second-order derivative of the field structure. Numerical results indicate that, for the illumination wavelength of $\lambda = 500$ nm, resolution of two 100-nm rectangular holes under near-field Rayleigh-Sommerfeld diffraction is improved by about 300 and 20% in the x and z directions, respectively, compared with the Sparrow criterion, while the lateral resolution limit is reduced to 35 nm.

DOI: [10.1103/PhysRevA.99.023814](https://doi.org/10.1103/PhysRevA.99.023814)**I. INTRODUCTION**

Classical resolution criteria such as, most famously, the Rayleigh criterion [1] and the Sparrow criterion [2] are based on the fringe structure produced by Fresnel or Fraunhofer diffraction [3]. However, as the Fresnel approximation assumes that the aperture scale is much larger than the wavelength level [4], its applicability on a subwavelength scale is seriously questioned [5]. When considering resolution problems for imaging on a subwavelength scale, say, label-free super-resolution or nanoscale photonic circuits, we should consider the more stringent Rayleigh-Sommerfeld (RS) diffraction, which is proved to be a strict scalar or vector diffraction formula in the half plane behind the screen with given boundary conditions [6,7].

Due to the complexity of the high-frequency oscillation term in the Rayleigh-Sommerfeld integral, calculus is generally performed numerically on a normal micron scale [8–10], while comparison between RS diffraction and other diffraction formulas including Kirchhoff diffraction [11,12], Fresnel diffraction [5,13], and other approaches [14,15] is studied to facilitate the problem. For the diffraction problem on a subwavelength scale, the complex effect of the high-frequency oscillation term on the phase in the RS integral makes it difficult to analyze the diffraction characteristics of the image, and study devoted to such issue is rarely seen [16,17]. Meanwhile, when the aperture scale is reduced into subwavelength level, it is generally recognized that vector diffraction theories [18–21] are needed to explain the entire electromagnetic field in the vicinity of the diffraction screen. Studies with different approaches [20,22,23] have been devoted to such issue for a long time but it was not until the publication of Ref. [22] that a clear picture about the full vectorial behaviors in the near zone ($0 \leq z < d^2/4\lambda$) was achieved.

In this paper, in a farther but more “practical in imaging” zone ($z > d^2/4\lambda$), we propose a theoretical scheme to understand RS diffraction on a subwavelength scale and point out that, unlike what we generally take for granted, due to a neglected singularity of the phase shift of each plane-wave component, there is a threshold effect that, once the aperture scale is smaller than a wavelength, diffraction fringes of the aperture itself disappear and the final pattern will eventually tend to a smooth spot regardless of any finer structure of the screen, within the axial propagation distance of a few wavelengths.

Therefore, the resolution problem of RS diffraction on a subwavelength scale becomes an even harder question. Not only can we not find an analytic function to predict the diffraction pattern but also we face a smooth image without any fringe structure for classical resolution criteria to resolve. Thus, to break through the resolution limit of RS diffraction on a subwavelength scale, we need an alternative resolution criterion. In this paper, we try a step on the higher-order terms of Taylor expansion in RS diffraction and propose a criterion based on the second derivative of the image. Such a criterion, which does not depend on the structure of the image itself, is proven to achieve a better resolution than classical criteria in general optical systems, under both coherent and incoherent illumination. Moreover, numerical results suggest an even greater improvement in the RS diffraction system.

This paper is organized as follows. In Sec. II the basic concepts of scalar RS diffraction and a further analysis with numerical results are given to demonstrate the disappearance of the fringe structure. In Sec. III analysis of the higher-order terms of Taylor expansion of the two-point imaging model and a resolution criterion based on the second derivative of the image are proposed. The numerical result of resolution improvement compared with the Sparrow criterion of a two-fold model under RS diffraction follows in Sec. IV. Finally, in Sec. V we present our conclusions.

*cesyxy@mail.sysu.edu.cn

Calculations in this paper are set in the zone $z > 0.1d^2/4\lambda$ suggested by Ref. [22] in which vectorial effect can be somehow ignored while Kirchhoff boundary conditions are reasonably valid. Since vectorial structure of the light field is beyond the scope of this paper, we start with scalar RS diffraction theory with Kirchhoff boundary conditions. A further discussion about the validity of these assumptions is given at the end of Sec. V.

II. DISAPPEARANCE OF FRINGE STRUCTURE

We now consider scalar RS diffraction under illumination from a monochromatic plane wave parallel to the optical axis. The result of the first and second RS integrals at any point behind the screen is proved to be an exact solution to the scalar wave equation:

$$\nabla^2 u + k_0^2 u = 0, \quad (1)$$

where k_0 is the incident wave vector. The first and second RS integrals can be represented as

$$u(x, z) = \frac{-1}{2\pi} \int_{-\infty}^{\infty} u(x', 0) \frac{\partial}{\partial z} \frac{e^{ik_0 \sqrt{(x-x')^2 + z^2}}}{\sqrt{(x-x')^2 + z^2}} dx', \quad (2)$$

$$u(x, z) = \frac{-1}{2\pi} \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial z} u(x', 0) \right] \frac{e^{ik_0 \sqrt{(x-x')^2 + z^2}}}{\sqrt{(x-x')^2 + z^2}} dx', \quad (3)$$

where x and x' are the transverse coordinates perpendicular to the optical axis on the image plane and the plane right behind the diffraction screen, respectively, and z is the distance from the diffraction screen to the image plane.

The solution of each of the formulas is proved to be equivalent to the plane-wave representation as follows [24]:

$$u(x, z) = F^{-1}\{U(k_x, 0)\exp(i\varphi)\}, \quad (4)$$

where k_x is the corresponding Fourier-transform coordinate of x' , $U(k_x, 0)$ is the Fourier transform of $u(x', 0)$, and F^{-1} is the inverse Fourier transform operation on the light field to its lateral k space. Fig. 1 is a schematic diagram. Equation (4) can be understood as the recombination of the plane-wave components (including the evanescent waves) decomposed right behind the object plane on the image plane. After propagating

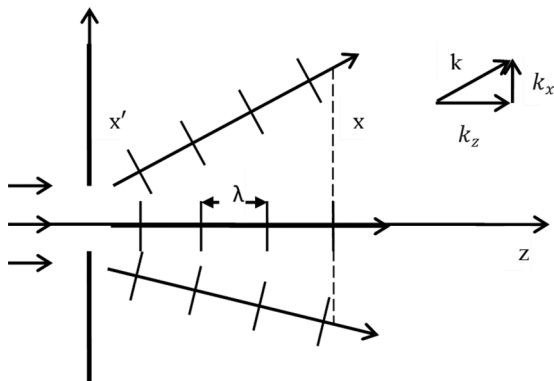


FIG. 1. Plane-wave representation of one-dimensional RS diffraction.

an axial distance z , each plane-wave component obtains a phase shift:

$$\varphi_{rs} = zk_z = z\sqrt{k_0^2 - k_x^2}. \quad (5)$$

Notice here that the Fresnel diffraction formula could be also represented by Eq. (4) but with a different phase shift [5]:

$$\varphi_f = z \left[k_0 - \frac{k_x^2}{2k_0} \right]. \quad (6)$$

The term φ_{rs} reflects the influence of both k_x and k_z of each plane-wave component on the phase superposition on the image plane. For the shape of the final image, consider the norm $|u|$ of Eq. (4), where each point of the image plane is the reflection of the variance of the phase contributed by the different plane waves at that point. It is the variance of this phase shift that determines the shape of the final pattern. If $\varphi(k_x)$ slowly changes in k_x , the contribution of each plane-wave component to the phase superposition on each point on the image plane will be stable, and the final diffraction pattern will be regular and predictable, forming the fringe structure that is familiar to us. Differentiating on both sides of Eq. (5),

$$d\varphi_{rs} = \frac{-zk_x}{\sqrt{k_0^2 - k_x^2}} dk_x. \quad (7)$$

When k_x is small compared with k_0 , corresponding to the normal micron scale, $d\varphi_{rs}/dk_x$ is small and well fitted with the Fresnel approximation:

$$d\varphi_f = -z \frac{k_x}{k_0} dk_x; \quad (8)$$

then contribution of each plane-wave component is stable and predictable.

However, what is neglected in the Fresnel approximation and the previous results is a singularity of $d\varphi_{rs}/dk_x$ at $k_x = k_0$. When the lateral diffraction scale is reduced to the wavelength level, it is this part of the plane-wave components near the singularity k_0 that begins to dominate and produce an anomalous but significant contribution to the final disturbance.

When k_x is large compared with k_0 , say k_x tends to k_0 , the term $\frac{zk_x}{\sqrt{k_0^2 - k_x^2}}$ tends to infinity:

$$\lim_{k_x \rightarrow k_0} \frac{d\varphi_{rs}}{dk_x} = \frac{-zk_x}{\sqrt{k_0^2 - k_x^2}} = -\infty, \quad \text{when } z \neq 0, \quad (9)$$

causing the phase shift of these plane-wave components to vary rapidly in the range of $-\pi$ to π and eventually cover the entire phase interval. The difference of phase variance on each point of the pattern composed by the lower-frequency signals is compensated by such rapid change of phase, at which point the image contrast begins to decrease and the fringe structures begin to disappear.

Further, when the lateral dimension of the object is reduced to the wavelength level, correspondingly in the k space, this part of the plane-wave components begins to dominate, causing a compensation for the phase difference of each point on the image plane that is large enough to smooth out all the fringe structure. Finally, regardless of any finer structure, the pattern tends to be smooth and consistent after propagating a few wavelengths in the z direction [as in Eq. (9),

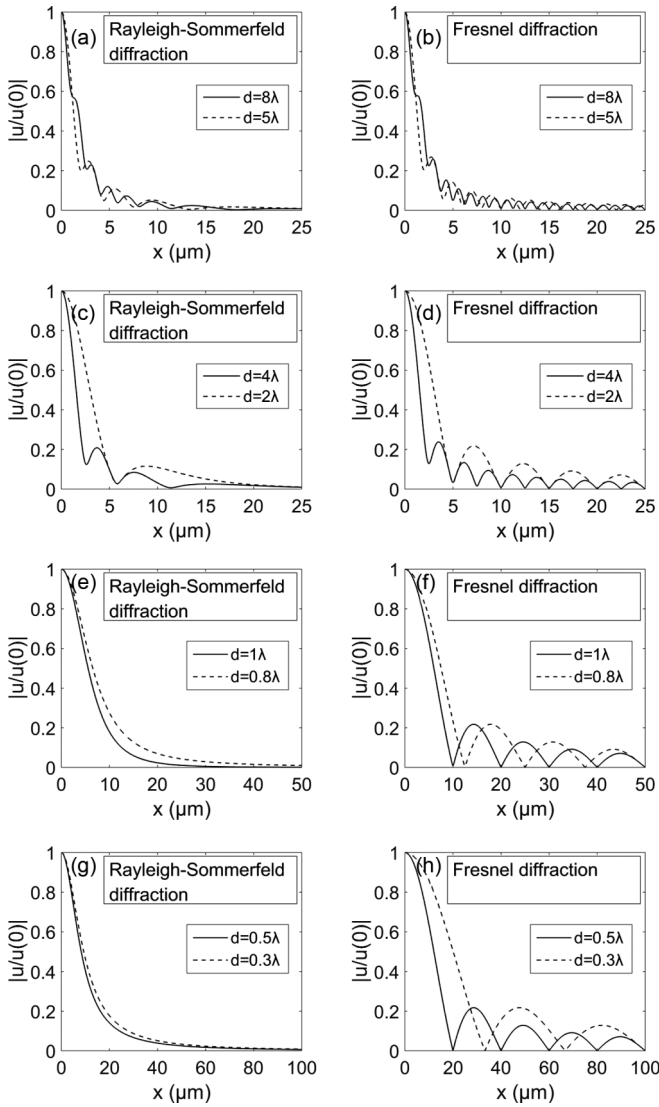


FIG. 2. Comparison of the normalized amplitude $|u/u(0)|$ between Rayleigh-Sommerfeld diffraction and Fresnel diffraction of one-dimensional rectangular apertures with different widths d , calculated by the first Rayleigh integral Eq. (2) and the Fresnel approximation Eq. (10), with boundary condition Eq. (11). $\lambda = 500$ nm, $z = 10$ μm . The sampling interval on x and x' is 5 nm.

where z appears as a multiplier in the numerator]. Such result cannot be seen on a normal micron scale; due to the strong nonlinearity near the singularity, there might be a threshold that only when the lateral diffraction scale is reduced to a certain value or a certain range around a wavelength will this effect start to reveal itself.

To further demonstrate such effect, numerical results follow. We now consider diffraction of a single one-dimensional rectangular aperture under the illumination of a monochromatic plane wave parallel to the optical axis with different aperture widths. Numerical results under RS diffraction and Fresnel diffraction are shown in Fig. 2, calculated, respectively, by the first Rayleigh integral Eq. (2) and the Fresnel approximation:

$$u(x, z) = \frac{e^{ik_0 z}}{iz\lambda} \int_{-\infty}^{\infty} u(x', 0) e^{\frac{ik_0(x-x')^2}{2z}} dx', \quad (10)$$

where the incident wavelength $\lambda = 500$ nm. For different boundary conditions, Ref. [25] has proved that their influence on the final results on the image plane is almost the same when it is not too close to the screen, therefore we choose the simplest Kirchhoff boundary assumption, which is set to be

$$u(x', 0) = \text{rect}\left(\frac{x'}{d}\right) \quad (11)$$

with d the aperture width. Reference [22] also demonstrates well its validity in the region we have chosen. A detailed discussion on this assumption will be given at the end of Sec. V. In order to focus on the lateral scale, we only show the result under the axial propagation distance $z = 10$ μm here. A full result under other propagation distances from $z/(d^2/4\lambda) = 0.1$ –10 is also given in the Appendix to better illustrate the problem.

For a normal micron scale, as shown in Figs. 2(a) and 2(b), when $d = 8\lambda = 4$ μm , in the vicinity of the paraxial region, a fringe structure similar to Fresnel approximation is retained in the RS diffraction, which is similar to the results shown in Refs. [8–13]. However, when the aperture size is reduced to several wavelengths, as what we have predicted, the part of the plane-wave component near the singularity begins to increase, causing the diffraction fringes to begin to disappear. As shown in Fig. 2(c), when the aperture width is reduced to 4λ and 2λ , the number of diffraction fringes of the aperture itself begins to decrease conspicuously. Further, when the aperture size is smaller than one wavelength, as shown in Figs. 2(e) and 2(g), the diffraction fringes disappear completely; with the aperture size further reduced, the diffraction pattern no longer changes significantly with the change of the aperture, reaching the threshold condition for fringe disappearance. Since there are no singularities in the Fresnel approximation, this effect cannot be observed in the result of Fresnel diffraction. As shown in Figs. 2(d), 2(f), and 2(h), the Fresnel diffraction results for different aperture widths retain an infinite number of diffraction fringes.

Another interesting point that should be noticed is, as shown in Figs. 2(a), 2(c), and 2(e), the number of fringes retained by the aperture in RS diffraction on the half axis is directly equal to the ratio of aperture to wavelength d/λ . More results shown in the Appendix further indicate that for any $z > 0.1(d^2/4\lambda)$ the number of fringes retained in the diffraction pattern of the aperture N is equal to

$$N = 2 \times [d/\lambda] + 1, \quad (12)$$

where $[x]$ is the floor function that takes as input a real number x and gives as output the greatest integer $< x$. When the aperture is reduced to one wavelength, $N = 1$, the diffraction fringe structure caused by the aperture disappears completely, leaving only one smooth peak on the image plane. Such threshold effect in Eq. (12) might be related to that of boundary condition Eq. (11), the Fourier transform of which is

$$U\left(\frac{k_x}{2\pi}, 0\right) = d \text{sinc}\left(d \frac{k_x}{2\pi}\right). \quad (13)$$

When $d = n\lambda$, k_0 is exactly the n th zero point of the sinc function on the half axis. For other boundary conditions, although the threshold effect may not conform to Eq. (12)

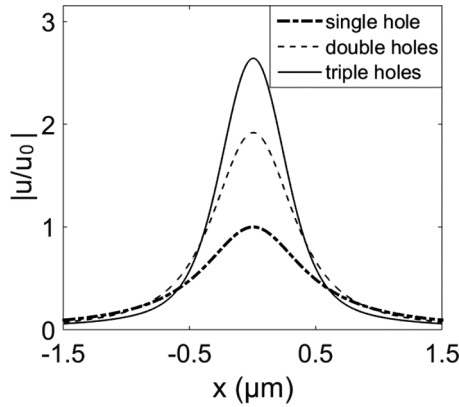


FIG. 3. Normalized amplitude $|u/u_0|$ of different structures below the wavelength level, calculated by the first Rayleigh integral Eq. (2) with boundary condition Eqs. (14). $\lambda = 500$ nm, $z = 1\lambda$. The three curves are shown as single-hole, double-hole, and three-hole diffraction results, respectively. The aperture d of each hole is 0.2λ , and the central distance r between the holes is 0.4λ . The sampling interval on x and x' is 5 nm.

exactly, the disappearance tendency of the fringe should be inevitable, since the singularity of $d\varphi_{rs}/dk_x$ is not dependent on the boundary conditions $u(x', 0)$ but rather on the nature of the RS diffraction itself. Other special properties only related to the aperture to wavelength ratio d/λ on a subwavelength scale were also reported in Refs. [6,22,25], indicating that a specific shape of the boundary condition has little to do with such effect and that it is the ratio d/λ that truly matters.

Results of different finer structures on a subwavelength scale are shown in Fig. 3. Consider a more demanding situation, that is, when the light field propagates only one wavelength in the axial direction, with Kirchhoff boundary conditions

$$\begin{aligned} u_1(x') &= \text{rect}\left(\frac{x'}{d}\right), \\ u_2(x') &= \text{rect}\left(\frac{x' - 0.5r}{d}\right) + \text{rect}\left(\frac{x' + 0.5r}{d}\right), \\ u_3(x') &= \text{rect}\left(\frac{x' - r}{d}\right) + \text{rect}\left(\frac{x' + r}{d}\right) + \text{rect}\left(\frac{x'}{d}\right), \end{aligned} \quad (14)$$

where d is the aperture width and r is the central distance between the holes. When the plane-wave components near k_0 smooth the diffraction fringe structure of the aperture itself, it is conceivable that the fringe reflects a finer structure and the aperture should disappear simultaneously. As shown in Fig. 3, within an axial propagation distance of one wavelength, whether it is a single-hole, double-hole, or three-hole structure, the final result is a smooth bright spot that tends to be uniform after only one wavelength is propagated in the axial direction. Regardless of different finer structure, the final pattern will tend to be uniform and smooth, at which time classical resolution criteria completely fail.

To summarize the above analysis, unlike what we usually see in the Fresnel diffraction or RS diffraction on a larger scale, there is a threshold effect that once the aperture size is reduced to about one wavelength the diffraction fringes will

disappear due to the existence of the singularity of $d\varphi/dk_x$ at $k_x = k_0$. Regardless of different finer structures, the final pattern will eventually tend to a uniform smooth spot after propagating only a few wavelengths in the axial direction, resulting in the failure of classical resolution criteria.

III. ALTERNATIVE CRITERION

Classical resolution criteria, as we have shown in the previous section, are ineffective when considering the RS diffraction results at subwavelengths scale. Due to the threshold effect brought by the singularity, regardless of any finer structure, the diffraction pattern tends to a smooth image spot while fringe structure disappears in a short propagation distance. At this point, whether using the most well-known Rayleigh criterion [1] or the more stringent Sparrow criterion [2], it is impossible to resolve this smooth image formed by two points but without any fringe (or central dip), although in theory, as long as the optical system is sufficiently accurate, we can find a certain mathematical model to “invert” and resolve the so-called calculated images of the system [3,26,27]. However, as we discussed in the previous section, for RS diffraction, we cannot use an analytical function to calculate the imaging results for Fraunhofer diffraction or Fresnel diffraction.

Nevertheless, although the high-frequency components near the singularity quickly smooth out the fringe structure, we can still find a certain rule in the process of the two points approaching each other, so as to continue to “read” the smooth image beyond the classical resolution criteria. Here we try a step on the higher-order terms of Taylor expansion and propose a two-point resolution criterion based on the second derivative of the image. Such a criterion, which is not based on the fringe structure of the image itself, will continue to “recognize” the two-point image in a greater lateral distance after the two points merge into a smooth spot, and also work at a larger axial distance.

For a general two-point model, whether it is RS diffraction or Fresnel diffraction, as long as the point spread function is an even function, the extrema (fringe structure) near the center of the image depends on the second-order derivative value $I''(0) \equiv d^2I(x)/dx^2|_{x=0}$ on its axis. When $I''(0) = 0$, the fluctuation of the first derivative near the center point disappears, so that the first derivative has only one zero point near the center point, and correspondingly the central dip of the image disappears, leaving only one peak at the center. Therefore, the Sparrow criterion requires that the two points approach each other until the second derivative on the axis $I''(0) = 0$, and the image reaches its resolution limit.

However, after the center minima disappears, similar function shapes are still very likely to remain in their higher derivatives. If this shape is different from the case of single-point imaging, then we can still recognize it from its higher-order derivative as the result of two-point imaging. With Taylor expansion near the center of the image function,

$$I(x) = I(0) + I'(0)x + \frac{I''(0)x^2}{2} + \cdots + \frac{I^{(k)}(0)x^k}{k!}, \quad (15)$$

after arbitrary n -order derivation,

$$I^{(n)}(x) = I^{(n)}(0) + I^{(n+1)}(0)x + \cdots + \frac{I^{(k)}(0)x^{k-n}}{(k-n)!}. \quad (16)$$

Obviously, for any $m \leq n$, changing the m -order derivative on-axis value $I^{(m)}(0)$ may affect the position and number of extrema of $I(x)$ itself, but it has no effect on the number and locations of the extrema of $I^{(n)}(x)$. Therefore, looking back at the Sparrow limit, although $I''(0)$ affects the distribution of the extrema near the center of the image $I(x)$ itself, it has no effect on the structure of the second derivative $I''(x)$ [since $I''(0)$ is a constant in $I''(x)$, Eq. (16)]. When the second derivative of its single-point imaging is continuous, if its zero point is not an extremum, the number of extrema of $I''(x)$ does not change in a smaller range of lateral center distance r after $I''(x) = 0$.

This means that now we can resolve a two-point image that is completely blurred to even smooth by such a simple act. In a two-point resolution model, when the two points are close enough to each other and merge into a single smooth spot, although we cannot directly observe the two peaks formed by two points through the image, we can still recognize from the structure of its second derivative that it is the imaging result of two different points. Even in theory, any even-order derivative can continue to retain such information of the last even-order derivative. However, from the point of view of actual measurement accuracy, we only consider to the second order.

Therefore, in order to further resolve a smooth image beyond classical resolution criteria, here we propose a two-point resolution criterion: when the two-point sources are close to each other until the number of extrema of the second derivative of the image is exactly equal to that of the single-point imaging, the image just reaches its resolution limit. For general large-aperture far-field Fraunhofer diffraction imaging systems, analyzing a certain point spread function, this criterion is obviously effective and capable to improve the resolution. Suppose a general rectangular large-aperture system with a point spread function:

$$u = \text{sinc}\left(\frac{x}{a}\right), \quad (17)$$

where a is the Rayleigh limit of the spot. In incoherent diffraction,

$$I = \text{sinc}^2\left(\frac{x-0.5r}{a}\right) + \text{sinc}^2\left(\frac{x+0.5r}{a}\right), \quad (18)$$

where r is the lateral distance between the two point sources, the resolution power is about 42.9% higher than the Rayleigh limit and 18.5% higher than the Sparrow limit (Fig. 4), calculated by

$$\text{RI} = (1/r_s - 1/r_0)/(1/r_0) \times 100\%, \quad (19)$$

where RI stands for the resolution improvement, r_s is the resolution limit under the second derivative, and r_0 corresponds to the Rayleigh resolution limit or the Sparrow resolution limit. For coherent diffraction with different phase differences,

$$I = \text{Re}\left[\text{sinc}\left(\frac{x-0.5r}{a}\right) + e^{i\sigma}\text{sinc}\left(\frac{x+0.5r}{a}\right)\right]^2, \quad (20)$$

where σ is the phase difference, resolution can be further increased by more than 20% over the Sparrow limit (Fig. 5).

From the above analysis, we can see that in the general large-aperture Fraunhofer diffraction system the second

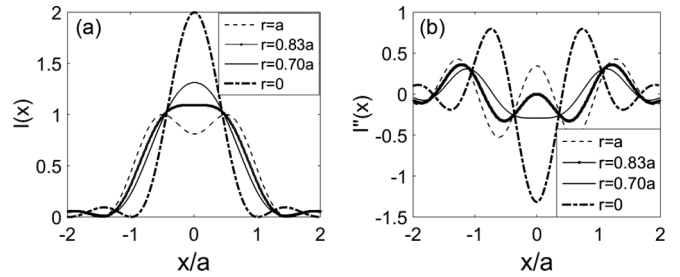


FIG. 4. Comparison of second-order derivative resolution and classical resolution limits in general large-aperture incoherent Fraunhofer diffraction, calculated by Eq. (18). Panel (a) shows the normalized intensity $I(x)$, panel (b) shows its second derivative $I''(x)$ in the x direction, r is the lateral distance between the two point sources, and a is the Rayleigh limit. Dashed, dot-continuous, continuous, and dot-dashed lines correspond to the Rayleigh limit, the Sparrow limit, the second-order derivative limit, and single-point imaging. Although the Sparrow limit appears when r is shortened to around $0.83a$, we cannot see the two-peak structure from the image itself, but the second derivative retains the same extremum structure until r is further reduced to $0.7a$.

derivative criterion can continue to resolve a smooth image beyond the classical resolution limit. Possible applications should be further discussed. For now, to focus back on the resolution problem of RS diffraction on a subwavelength scale, which is a more complicated problem, numerical analysis is followed in the next section to examine the utility of this criterion in RS diffraction.

IV. TWO-HOLE MODEL UNDER RS DIFFRACTION

For RS diffraction, as we discussed in Sec. II, when the lateral dimension is less than one wavelength, the norm of the optical field $|u|$ on the image plane will quickly become a single smooth peak within the axial propagation distance z of several wavelengths. Considering that the two-point incoherent imaging result is a direct superposition of the two envelopes at this time, the performance of the second derivative criterion should be similar to that of the large-aperture far-field Fraunhofer system. However, for two-point coherent imaging, the situation is different. We represent the final light

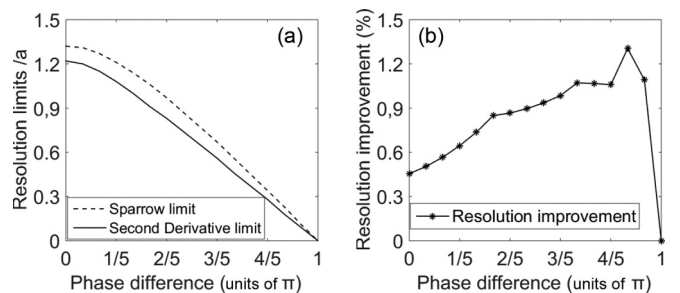


FIG. 5. Relative resolution limits of two-point coherent Fraunhofer diffraction under (a) different criteria (the second derivative limit over the Sparrow limit) and (b) resolution improvement vs phase difference σ , calculated by Eqs. (19) and (20), respectively.

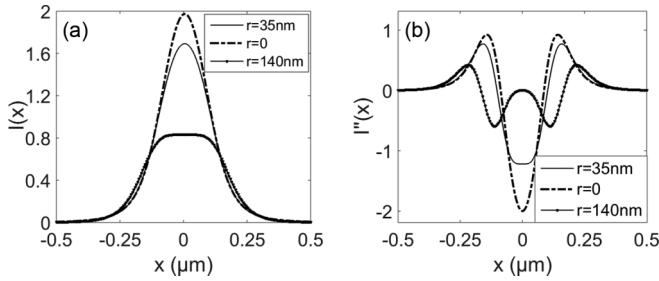


FIG. 6. Normalized intensity $I(x)$ (a) and its second derivative $I''(x)$ (b) of coherent RS diffraction results of two 100-nm-wide rectangular holes at different separate distance r with no phase difference after axial propagation at $z = 140$ nm. Dot-continuous, continuous, and dot-dashed lines correspond to the Sparrow limit, the second-derivative limit, and single-point imaging, respectively.

field as the product of the norm and the angle:

$$u(x, t) = |u(x)|\exp[i\varphi(x, t)]. \quad (21)$$

As we discussed earlier, the plane-wave component near the singularity compensates for the variance of the phase difference at each point on the image plane, so that the norm $|u(x)|$ of each point tends to be evenly distributed. However, when considering a single moment, each point on the image plane will obtain a phase $\varphi(x, t)$ that differs more from the other points. This larger difference of the phase information may be more clearly manifested by the second derivative, making the improvement of resolution of the second derivative criterion versus the classical criterion even more significant under coherent RS diffraction.

We now consider coherent RS diffraction of two one-dimensional rectangular holes. For simplicity, we only consider equal phase coherent diffraction with widths of each hole fixed to 100 nm, and the incident wavelength $\lambda = 500$ nm. For cases with more complex coherence, please see Ref. [16], although it does not study the higher-order derivative of the image, but it gives a very clear ideal diffraction image of multipoint sources under different coherence. We define the distance r between the two holes as the distance between two adjacent sides of the holes, so that when $r = 0$ the two holes just merge into one hole. As the diffraction fringe structure of the aperture itself is completely gone at this point, the Rayleigh criterion is out of our consideration.

First, consider the two-point imaging result of different lateral distances r under a single longitudinal distance with no phase difference. The results of the image intensity at 140 nm in the axial direction are shown in Fig. 6. Similar to the case of large-aperture far-field Fraunhofer diffraction, the second-order derivative criterion continues to resolve the image within a shorter distance r after the Sparrow limit; when $r \in [35 \text{ nm}, 140 \text{ nm}]$, the central dip of the image itself disappears, but its second derivative retains the same structure of the extrema; it is when the central maximum of the second derivative disappears at $r = 35$ nm that the image reaches its second derivative limit. Nevertheless, compared with Fig. 4, we can see that this extra resolution range is much larger than that under the general Fraunhofer diffraction. Under such conditions, resolution limit is reduced to 35 nm, which is 1/4 of the Sparrow limit, and the resolution power is increased

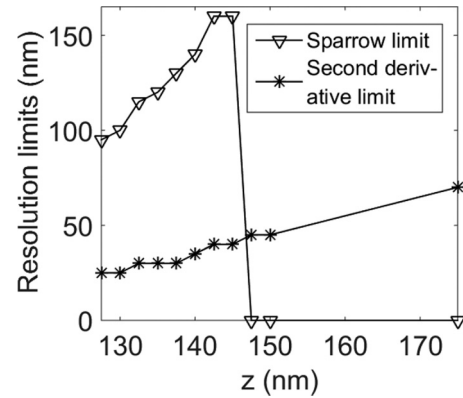


FIG. 7. Comparison of the resolution limits at different axial propagation distances z under RS diffraction. The dotted line corresponds to the Sparrow limit and the asterisked line corresponds to the second derivative limit, defined as the minimum two-hole distance r in accordance with the second-order derivative resolution criterion and the Sparrow criterion. The on-axis sampling interval is 2.5 nm, and the lateral sampling interval is 5 nm.

by 300%. At this point we still have to pay attention to the width of each rectangular hole itself, which is 100 nm, and the wavelength is 500 nm. This resolution limit will be further reduced in a shorter axial distance; results under different axial propagation distances are shown in Fig. 7.

As we can see from Fig. 7, within a certain axial propagation distance z , the lateral resolutions of the second derivative are all increased by about 300% compared to the Sparrow criterion. At the same time, due to the improvement of the lateral resolution, the axial resolution capability is also improved, and the second derivative criterion will work within a longer axial distance. For lateral resolution, the Sparrow resolution limit is up to 95 nm at an axial distance of 127.5 nm, which is the distance of almost one aperture width. With the increase of the propagation distance z , the resolution limit increases to 160 nm at $z = 145$ nm. Since this model considers the equal-phase coherent diffraction of two holes, when the axial distance z is further increased, interference fringes begin to appear on the axis after $r > 160$ nm (at this point the total lateral dimension of the two holes is close to one wavelength). As there is no central minimum while diffraction fringes show up near the center, the Sparrow criterion is completely ineffective. However, since the resolution limit of the second derivative resolution criterion is much smaller than the Sparrow limit, it can continue to be resolved within a certain lateral distance as the axial distance increases. When the axial propagation distance z increase to 175 nm the image cannot be resolved by the second derivative criterion due to the deformation of the interference. In addition to the lateral resolution limit far beyond the Sparrow criterion, the axial resolution depth of the second derivative criterion is also 20.6% higher than the Sparrow criterion. At its axial resolution limit of $z = 175$ nm, its lateral resolution limit of 65 nm is also 46.2% smaller than the Sparrow criterion at the minimum lateral limit of 95 nm at $z = 127.5$. Even in such comparison, the second-order derivative resolution criterion improves the resolving power under RS diffraction far more than that of the general large-aperture Fraunhofer diffraction.

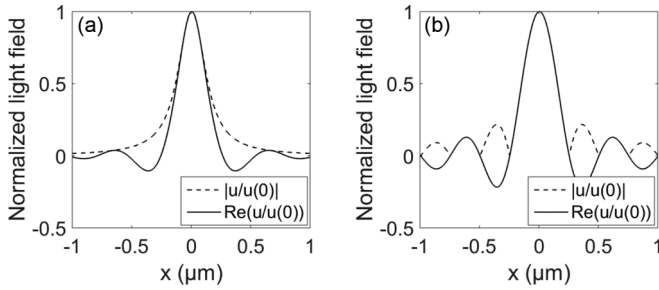


FIG. 8. Comparison of the normalized amplitude $|u/u_0|$ and the real part $\text{Re}(u/u_0)$ at a single moment of RS diffraction (a) and Fraunhofer diffraction (b), calculated by Eqs. (2) and (22), respectively, with boundary condition Eq. (11). The incident wavelength $\lambda = 500$ nm. Aperture width $d = 200$ nm, $z = 140$ nm in (a) and $d = 10$ cm, $z = 5$ cm in (b) (consider a focal lens with large NA), and the sampling interval on x and x' is 5 nm. Although under such conditions the norms of the light field $|u|$ of RS diffraction and Fraunhofer diffraction share a similar size, the real part differs from the norm in panel (a) while it does not in panel (b), indicating that phase information is still stored in RS diffraction on such a scale while it is not in Fraunhofer diffraction.

Numerical results above show a significant improvement in the lateral resolution of two symmetrical rectangular holes with a width of 100 nm under coherent RS diffraction, which is much greater than that under general large-aperture Fraunhofer diffraction. This surprising result may be due to the fact that in subwavelength RS diffraction, although the norm of light field $|u|$ no longer retains the fringe structure, the difference between the phases of each point on the image plane is rather serious, hiding information in the real part [Fig. 8(a)]. Such information might be somehow reflected by the second derivative during the coherent process so that the lateral resolution is significantly improved. For a general large-aperture Fraunhofer diffraction system,

$$u(x) = C e^{\frac{ik_0 x^2}{2z}} F\{u(x')\}_{|x'/(\lambda z)}, \quad (22)$$

phase difference between each point on the image plane is generally independent of position, and the shape of the norm of the light field directly reflects the shape of the real part [Fig. 8(b)]. Therefore, although the second derivative can still improve the resolution beyond the classical criterion, the improvement is not as large as in the subwavelength scale RS diffraction.

A complete theoretical explanation may be obtained by treating the light field behind a small hole ($d < \lambda$) as the point spread function of the first RS integral [Eq. (2)] with a scalar factor c : $\text{PSF}(x/c)$, where

$$\text{PSF}(x) = \frac{\partial}{\partial z} \left(\frac{e^{ik_0 \sqrt{(x)^2 + z^2}}}{\sqrt{(x)^2 + z^2}} \right), \quad (23)$$

since when below the threshold ($d < \lambda$) the diffraction pattern of each hole does not change significantly with the reduction of the aperture width (as discussed in Sec. II and the Appendix). Also, it is similar to the results of the no-width ideal point imaging given in Ref. [16].

With such threshold effect, we also have reasons to believe that the lateral resolution limit of the second derivative

criterion will further reduce as the width of the two holes shrinks; meanwhile, the axial resolution will also increase. With all the results above, it is convincing to conclude that the second derivative criterion might significantly improve the lateral and axial resolution of subwavelength RS diffraction beyond traditional resolution limits.

V. CONCLUSION

In conclusion, we first point out that, due to the presence of a singularity of $d\varphi/dk_x$ [Eq. (3)] at $k_x = k_0$ in RS diffraction, a threshold effect occurs when the lateral dimension of the object is less than one wavelength. At this point, the diffraction fringes of the aperture itself disappear in the norm of the light field, and no matter how fine the structure is the final image will quickly become smooth and consistent within the axial propagation distance of several wavelengths, resulting in the failure of the classical criterion.

To further solve the resolution problem of RS diffraction at subwavelength scale, a two-point resolution criterion based on the structure of the second derivative of the image is then proposed. By Taylor expansion, we theoretically proved that the criterion can continue to resolve a smooth two-point image without any fringes or a central dip beyond the classical resolution limit. For a general large-aperture Fraunhofer diffraction system, this improvement can be accurately calculated.

Moreover, for coherent RS diffraction at subwavelength scale, a much greater improvement in the lateral resolution suggested by the numerical results is reported here, namely, that the lateral resolution of the second derivative resolution criterion is about 300% higher than the Sparrow criterion over different longitudinal distances. Meanwhile, due to the improvement of the lateral resolution, the axial resolution capability is also improved. This surprising result may be due to the fact that the phase difference at each point on the image plane is more severe in RS diffraction at the subwavelength scale and this difference can be reflected by the second derivative of the image.

Calculations in this paper are produced by one-dimensional scalar RS diffraction with Kirchhoff boundary conditions. For the vectorial structure of the light behind a perfectly conducting screen with zero thickness, Refs. [18,22] have given an almost perfect vectorial solution by means of the Hertz vector model, with experimental proof by Refs. [28,29]. Distribution of Poynting vector \mathbf{S} and the validity of Kirchhoff boundary conditions are also examined in detail in Ref. [22]. For a perfectly conducting screen with apertures at subwavelength scale ($d/\lambda \in 0.1 \sim 10$) (incident by a linearly polarized plane wave with \mathbf{E} in the x direction and \mathbf{H} in the y direction), Guha and Gillen have demonstrated [22] that by invoking Kirchhoff boundary conditions calculated fields of light only have a major difference of asymmetry in E_x and S_z in the extreme near zone ($z \ll 0.2d^2/4\lambda$), while H_x and H_y are exact and E_y , E_z , and H_z are different but with negligible magnitude (Secs. 3 and 6.1 in Ref. [22]). When $z > 0.2(d^2/4\lambda)$ (Sec 6.2), asymmetry of the light spot in x and y directions begins to vanish (Fig. 11 in Ref. [22]), and the on-axis value of $|E_x/E_0|^2$ calculated by the Hertz vector and S_z/S_0 calculated by both

the Hertz vector and Kirchhoff boundary conditions no longer hold noticeable difference (Fig. 10 in Ref. [22]).

All these results suggest that for apertures in subwavelength scale, when ignoring the longitude components of the light field and the lateral component of the Poynting vector, using one-dimensional scalar RS diffraction formulas with Kirchhoff boundary conditions is valid enough to presume the shape of the light spot in the zone where $z \gg 0.2d^2/4\lambda$. All calculations in the main part of the paper are set in this zone; in fact, most of them are set in a much farther zone $z > 1(d^2/4\lambda)$, and only a small part of the results of Fig. 9 are set near $z > 0.1(d^2/4\lambda)$ as a supplement.

More computational and experimental data with theoretical analysis that support this view can be found in Ref. [25] and references therein. We believe that using the scalar assumption and Kirchhoff boundary conditions is valid enough for our paper. Though a full vectoral theory is of interest, we believe that our resolution criteria can enable more far-field and near-field structured-light super-resolution systems to further achieve higher resolution. We would like to pursue this goal in future work.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (Grants No. 11274398 and No. 61575223) and the Fundamental Research Funds for the

Central Universities. The authors thank Prof. Jian-ying Zhou, Prof. Xiang-sheng Xie, Mr. Shuang Liu, and two anonymous reviewers for their valuable discussion and kind efforts on this paper.

APPENDIX

Here we supplement the RS diffraction numerical results of a one-dimensional rectangular hole with different aperture sizes d and axial propagation distance z . To show more clearly how the threshold effect is related to the ratio d/λ , we use two dimensionless ratios $x_1 = x/d$ and $z_1 = z/(d^2/4\lambda)$ to replace the spatial coordinates. We can clearly see from Fig. 9 that when $d > \lambda$ the number of diffraction fringes on the half axis is exactly equal to the aperture to wavelength ratio d/λ . Such effect does not change for a larger number of z_1 . When d is reduced to 1λ , the aperture no longer diffracts any fringes, but quickly tends to a smooth bright spot as we discussed in Sec. II. At the same time, we can see that when the aperture is smaller diffraction patterns do not change notably under the ratio coordinates x_1 and z_1 , supporting the conclusion at the end of Sec. IV.

Our paper does not consider the transmission function, that is, the power decreasing function after the light passes through a small hole. For a detailed analysis of this issue, please refer to Ref. [22].

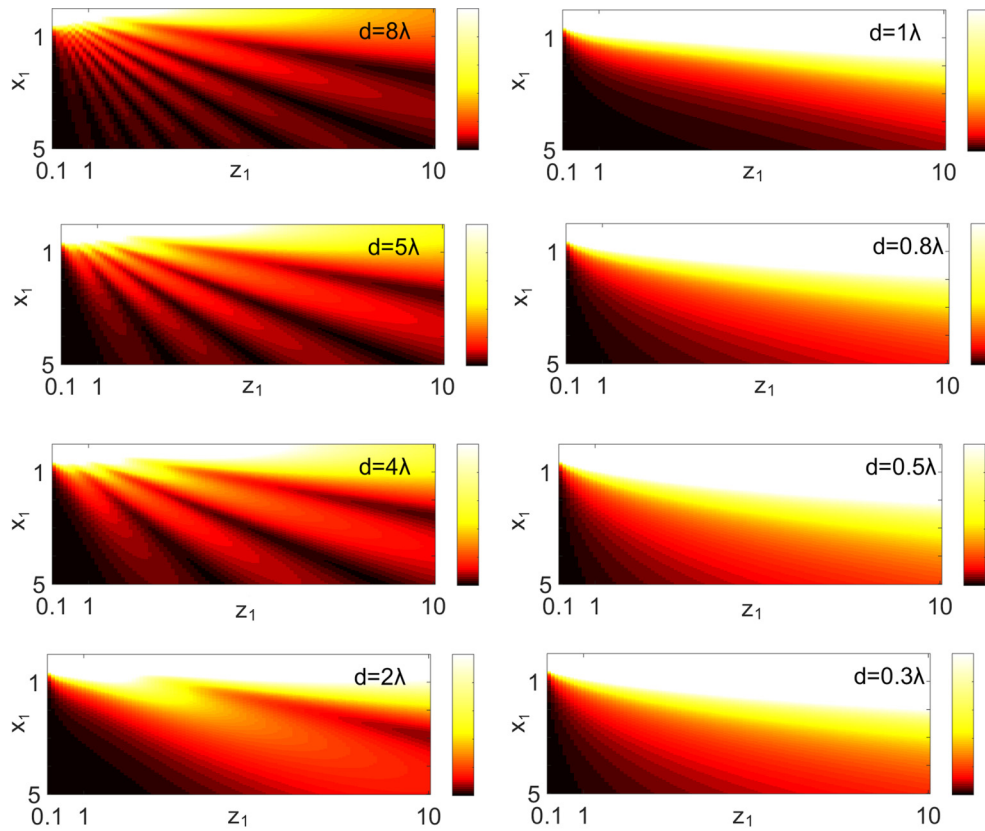


FIG. 9. The norm of light field $|u|$ in the half plane after the diffraction screen of a one-dimensional rectangular hole with different aperture sizes d under the illumination of a monochromatic plane wave parallel to the optical axis, calculated by the first Rayleigh integral Eq. (2) and boundary condition Eq. (11) with dimensionless coordinates x_1 and z_1 . Incident wavelength $\lambda = 500$ nm, the axial sampling interval of z_1 is 0.1, and the lateral sampling interval on either x' or x is $0.01d$. For clarification on axis values and the power transmission problem, please refer to Ref. [22].

- [1] L. Rayleigh, *Collected Papers* (Cambridge University, Cambridge, England, 1902), Vol. 3, p. 84.
- [2] C. Sparrow, *Astrophys. J.* **44**, 76 (1916).
- [3] A. J. den Dekker and A. van den Bos, *J. Opt. Soc. Am. A* **14**, 547 (1997).
- [4] W. H. Southwell, *J. Opt. Soc. Am. A* **71**, 7 (1981).
- [5] G. W. Forbes, *J. Opt. Soc. Am. A* **13**, 1816 (1996).
- [6] X. Jia, Y. Wang, and B. Li, *Opt. Express* **18**, 7064 (2010).
- [7] W. Freude and G. K. Grau, *J. Lightwave Technol.* **13**, 24 (1995).
- [8] J. A. C. Veerman, J. J. Rusch, and H. Paul Urbach, *J. Opt. Soc. Am. A* **22**, 636 (2005).
- [9] F. Shen and A. Wang, *Appl. Opt.* **45**, 1102 (2006).
- [10] J. J. Stamnes, *J. Opt. Soc. Am. A* **6**, 1330 (1989).
- [11] E. Wolf and E. W. Marchand, *J. Opt. Soc. Am.* **54**, 587 (1964).
- [12] J. C. Heurtley, *J. Opt. Soc. Am.* **63**, 1003 (1973).
- [13] M. W. Farn and J. W. Goodman, *J. Opt. Soc. Am. A* **7**, 948 (1990).
- [14] A. S. Marathay and J. F. McCalmont, *J. Opt. Soc. Am. A* **21**, 510 (2004).
- [15] N. A. Ochoa, *Opt. Express* **25**, 12008 (2017).
- [16] R. Castañeda, *Phys. Rev. A* **89**, 013843 (2014).
- [17] K. Yang and X. Xie, *Opt. Commun.* **407**, 33 (2018).
- [18] G. Bekefi, *J. App. Phys.* **24**, 1123 (1953).
- [19] B. Lu and K. Duan, *Opt. Lett.* **28**, 2440 (2003).
- [20] H. A. Bethe, *Phys. Rev.* **66**, 163 (1944).
- [21] L. Carretero, P. Acebal, and S. Blaya, *J. Opt. Soc. Am. A* **30**, 733 (2013).
- [22] S. Guha and G. D. Gillen, *Opt. Express* **13**, 1424 (2005).
- [23] R. Edward English and N. George, *J. Opt. Soc. Am. A* **5**, 192 (1988).
- [24] G. C. Sherman, *J. Opt. Soc. Am.* **57**, 546 (1967).
- [25] Y. Li, *J. Mod. Opt.* **54**, 2723 (2007).
- [26] V. Ronchi, *J. Opt. Soc. Am.* **51**, 458 (1961).
- [27] S. Van Aert and D. Van Dyck, *Opt. Express* **14**, 3830 (2006).
- [28] C. L. Andrews, *J. Appl. Phys.* **21**, 761 (1950).
- [29] M. J. Ehrlich, S. Silver, and G. Held, *J. Appl. Phys.* **26**, 336 (1955).