

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/347409707>

Frequency sampling strategy for numerical diffraction calculation

Article in Optics Express · December 2020

DOI: 10.1364/OE.413636

CITATIONS

0

READS

60

3 authors, including:



Wenhui Zhang

Tsinghua University

32 PUBLICATIONS 79 CITATIONS

[SEE PROFILE](#)



Hao Zhang

Tsinghua University

74 PUBLICATIONS 1,423 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



compressive multiple-angle holographic microscope [View project](#)



Frequency sampling strategy for numerical diffraction calculations

WENHUI ZHANG, HAO ZHANG, * AND GUOFAN JIN

State Key Laboratory of Precision Measurement Technology and Instruments, Department of Precision Instrument, Tsinghua University, Beijing 100084, China

*haozhang274@tsinghua.edu.cn

Abstract: Diffraction calculations play an essential role in Fourier optics and computational imaging. Conventional methods only consider the calculation from the perspective of discrete computation which would either cause error or sacrifice efficiency. In this work, we provide a unified frequency response analysis from the joint physics-mathematics perspective and propose corresponding adaptive frequency sampling strategies for five popular diffraction calculation methods. With the proposed strategies, the calculation correctness is guaranteed and the calculation efficiency is improved. Such an idea of unified frequency response study would help researchers make a do-it-yourself analysis for various diffraction calculation tasks and choose or develop a method for accurate and efficient computations of the diffraction fields.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Diffraction matters in optics, either as a physical phenomenon or as a tool used in engineering [1]. Particularly, diffraction in free space is an important branch. To study it, the mathematical description was used and many theories were proposed, such as Rayleigh-Sommerfeld theory and angular spectrum theory [1,2]. The goal of such theories is to predict an accurate field distribution after the light propagation. It is well known that the diffraction field is an integral transform of the input field. Thus, theoretical works were carried out for a high-accuracy transform, such as a modified Fresnel propagation [3]. However, we can only get the analytical diffraction field of a few simple input fields. On the other hand, in many cases, the input fields have already been discretely sampled, such as the holograms [4,5] or x-ray diffraction patterns [6,7] captured by digital sensors. Therefore, to get the diffraction field, such as the reconstruction of holograms or x-ray diffraction patterns, discrete computation is necessary in practice. In such cases, the discretization process should be implemented correctly based on the sampling theorem.

From the analytical solution to the discrete computation, three points need attention. First, high accuracy should be maintained. Second, computational efficiency should be high. Third, computational flexibility is desired. The first two points are basic requirements of diffraction calculations. The last one requires the flexibility to observe the diffraction field with adjustable propagation distance, adjustable position and size of the observation window, and adjustable sampling interval. To achieve such goals, many methods have been proposed. Band-limited angular spectrum method (ASM) [8], band-extended ASM [9] were proposed for accurate calculation in a large propagation distance range. Shift Fresnel diffraction [10], two-step Fresnel diffraction [11], and scaled ASM [12] were proposed for an adjustable sampling interval of the diffraction field. Shift band-limited ASM [13] and oblique ASM [14] were proposed for an off-axis diffraction calculation. A generalized Debye integral was proposed for a larger focus range [15]. Semi-analytical propagation techniques [16] and homeomorphic Fourier transform [17] were proposed to improve the computational efficiency. A brief review can be found in the introduction of our recent paper [18].

Here, we would like to emphasize that the numerical diffraction calculation via discrete computation is a research field where both physics and mathematics should be considered.

Previous studies only focused on the discrete computation but ignored the physical picture which should be carefully considered. The conventional pure discrete sampling analysis would cause error in the non-convolution-based method. Another representative example is the zero padding operation in the numerical diffraction calculation. Most studies gave the reason of zero padding from the discrete computation perspective: to avoid the circular convolution errors [8,9,19]. Of course, zero padding has this function. However, there are other more important reasons why zero padding is necessary, which are given from the perspective of physics in this paper. Furthermore, with the physical picture, an adaptive zero padding method can be achieved based on needs for a higher efficiency. In short, a unified frequency sampling strategy from a joint physics-mathematics analysis could benefit numerical diffraction calculations a lot.

In this work, from the perspective of frequency response (physics) and discrete computation (mathematics), we propose different sampling strategies for five common-used diffraction calculation methods: single-Fourier transform based Fresnel transform (SFR), ASM/Fresnel transfer function approach (Fres-TF), and Rayleigh-Sommerfeld convolution (RSC)/Fresnel impulse response approach (Fres-IR). Note that Fres-TF is the paraxial version of ASM; and Fres-IR is the paraxial version of RSC. For each method, the characteristic of its frequency response is analyzed associated with the sampling theorem. After understanding the physics-consistent sampling request, the corresponding frequency sampling strategies are then proposed. Such strategies either guarantee the calculation correctness, which was affected in the conventional methods, or improve the calculation efficiency, which was limited by the wasted sampling resources. Besides, a further discussion related to the band-limited ASM (BLASM) [8] and band-extended ASM (BEASM) [9] is given. Here, frequency response refers to the spatial distribution of the diffraction field determined by the spatial frequency of the input field, which will be further explained with schematic diagrams in the following.

This paper is organized as following: part 2 briefly introduces the physical diffraction methods; part 3 discusses the properties of discrete computation; part 4 analyzes the frequency response of above mentioned methods and proposes corresponding strategies; parts 5 analyzes the reason why BLASM fails in far field and introduces the solution of BEASM; part 6 presents a discussion and concludes this work.

2. Physical diffraction methods

The diffraction model is shown in Fig. 1. The input field $u(x)$ is known and the task of diffraction calculation is to obtain the diffraction field $\tilde{u}(X)$.

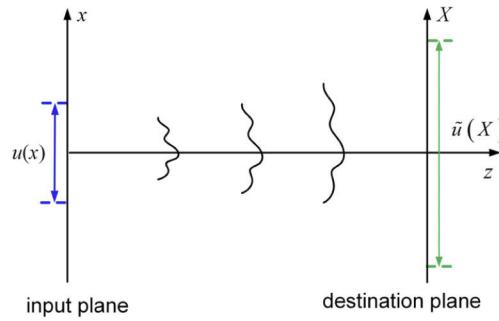


Fig. 1. Schematic diagram of free-space diffraction.

Here, we first summarize the mathematical expressions of five common-used methods in Table 1. Note that all the constants are neglected in the expressions. The first three methods are all based on the Fresnel approximation. Thus, analytically, they are equivalent. Since they

are based on Fresnel approximation, they can be only used in the paraxial case. The last two methods, ASM and RSC, are also analytically equivalent with each other because the transfer function of ASM and the impulse response function of RSC are a pair of Fourier transformations. Because there is no paraxial approximation used in these two methods, they can be used in both paraxial and non-paraxial cases.

Table 1. Summary of several common-used physical diffraction methods

Method	Mathematical expression	Applicable condition
SFR	$\tilde{u}(X) = \exp\left(\frac{i\pi}{\lambda z} X^2\right) \int \left[u(x)\exp\left(\frac{i\pi}{\lambda z} x^2\right)\right] \exp\left(-i2\pi \frac{X}{\lambda z} x\right) dx$ $\tilde{u}(X) = IFT\{FT\{u(x)\} \times H(f_x)\}$	paraxial
Fres-TF	$H(f_x) = \exp(-i\pi \lambda z f_x^2)$ $\tilde{u}(X) = IFT\{FT\{u(x)\} \times FT\{h(x)\}\}$	paraxial
Fres-IR	$h(x) = \exp\left(\frac{i\pi}{\lambda z} x^2\right)$ $\tilde{u}(X) = IFT\{FT\{u(x)\} \times H(f_x)\}$	paraxial
ASM	$H(f_x) = \exp\left(i\frac{2\pi}{\lambda} z \sqrt{1 - (\lambda f_x)^2}\right)$	\
RSC	$\tilde{u}(X) = IFT\{FT\{u(x)\} \times FT\{h(x)\}\}$ $h(x) = \frac{1}{2\pi} \frac{z}{r} \left(\frac{1}{r} - \frac{i2\pi}{\lambda}\right) \frac{\exp(i2\pi r/\lambda)}{r}, r = \sqrt{x^2 + z^2}$	\

Physically, we can observe the diffraction field with any propagation distance z ; inside an observation window with any size and located at any lateral position. However, after the discretization, all these flexibilities are limited in the discrete computation. More seriously, the accuracy of the diffraction field may be affected if the discrete computation is carried out in an incorrect way. Therefore, it is important to know the properties of the discrete computation.

3. Properties of discrete computation

As seen from Table 1, all these five methods need Fourier transform to get the results. In practice, Fourier transform is implemented by the fast Fourier transform (FFT) algorithm for a high computational efficiency.

Note that the latter four methods are all convolution-based. Take RSC as example, the diffraction field $\tilde{u}(X)$ is the linear convolution between the input field $u(x)$ and the impulse response function $h(x)$. To get a fast calculation, the linear convolution is carried out in the spatial frequency domain by multiplying their Fourier transformations because the computational complexity of multiplication is much cheaper than that of convolution. Then, inverse Fourier transform is implemented to the product for calculating the diffraction field. The theoretical basis for this operation is the convolution theorem. However, in discrete computation, the convolution theorem is for circular convolution not the linear convolution [20]. Therefore, to use the convolution theorem for a high computational efficiency, zero padding is required because circular convolution between zero-padded signals is equivalent to linear convolution between the original signals. That is to say padding zeros to $u(x)$ and $h(x)$ is necessary.

Now let us examine the properties of the discrete computation carefully. Suppose the input field $u(x)$ is sampled by N points with the sampling interval of Δ_x ; and the number of padded zeros is N_p ; i.e. $N + N_p$ points for the zero-padded input field $u_p(x)$. According to the transform property of FFT, the sampling interval of its Fourier spectrum $U_p(f_x)$ is

$$\Delta_f = \frac{1}{(N + N_p)\Delta_x}. \quad (1)$$

Equation (1) means that the sampling interval in the spatial frequency domain is determined by the sampling parameters in the space domain. This is the most important property of the FFT-based discrete computation. In the following, we would like to analyze each method in detail and provide corresponding frequency sampling strategy based on this property.

4. Frequency sampling strategies

4.1. SFR

Among the above five methods, SFR is the only one that is not convolution-based. Therefore, in the conventional method, zero padding is not required. However, it has been proven that the conventional method would cause aliasing error of high-frequency components [21]. In addition to the solution given in [21], here, we would like to provide another approach for an aliasing-free SFR calculation. As indicated in Eq. (2), there are two chirp functions used in SFR calculation. We call $\exp(\pi x^2/\lambda z)$ inner chirp and $\exp(\pi X^2/\lambda z)$ outer chirp.

$$\tilde{u}(X) = \exp\left(\frac{i\pi}{\lambda z} X^2\right) \int \left[u(x)\exp\left(\frac{i\pi}{\lambda z} x^2\right)\right] \exp\left(-i2\pi \frac{X}{\lambda z} x\right) dx. \quad (2)$$

To sample them correctly, the sampling theorem should be satisfied. The input field $u(x)$ is sampled by N points with the sampling interval of Δ_x , so is the inner chirp. Therefore, we have

$$\left| \frac{1}{2\pi} \frac{\partial \varphi(x)}{\partial x} \right|_{\max} \leq \frac{1}{2\Delta_x}, \quad (3)$$

where $\varphi(x) = \pi x^2/\lambda z$ is the phase of the inner chirp. From Eq. (3) we can get

$$z \geq \frac{N\Delta_x^2}{\lambda}, \quad (4)$$

which means that the propagation distance should be no smaller than $N\Delta_x^2/\lambda$. Only considering the sampling of the chirp function is the conventional method [22]. However, what should be sampled correctly is not just the separate inner chirp but the product of the inner chirp $\exp(\pi x^2/\lambda z)$ and the input field $u(x)$. Similar analysis was carried out in the calculation of focus fields [23]. Besides the condition given in Eq. (4), we still need

$$\left| \frac{1}{2\pi} \frac{\partial \varphi(x)}{\partial x} + f_x(x) \right|_{\max} \leq \frac{1}{2\Delta_x}, \quad (5)$$

where $f_x(x)$ is the local spatial frequency of the input field $u(x)$. From Eq. (5) and after a simple inequality scaling, we have

$$|f_x(x)|_{\max} \leq \frac{1}{2\Delta_x} - \frac{N\Delta_x}{2\lambda z}, \quad (6)$$

which means only the components with a local spatial frequency less than such a value can be sampled correctly. In other words, what has been correctly transformed of $u(x)$ are those components with a local spatial frequency within the range of $B = [N\Delta_x/2\lambda z - 1/2\Delta_x, 1/2\Delta_x - N\Delta_x/2\lambda z]$. For SFR, the frequency response is directly related with the spatial distribution of the diffraction field as $f_x = X/\lambda z$ given in Eq. (2), therefore, we have

$$\left| \frac{X}{\lambda z} \right|_{\max} \leq \frac{1}{2\Delta_x} - \frac{N\Delta_x}{2\lambda z}, \quad (7)$$

which gives

$$X \in \left[-\frac{L}{2}, \frac{L}{2} \right], L = \frac{\lambda z}{\Delta_x} - N\Delta_x. \quad (8)$$

Equation (8) means only the transform result within this range is correct. According to the transform property of FFT, the whole transform result would spread on the range of

$[-\lambda z/2\Delta_x, \lambda z/2\Delta_x]$, but only the part inside the range given by Eq. (8) is correct because outside parts have aliasing error due to the incorrect sampling. It is evident that the correct range increases with the propagation distance.

Next, we need to calculate the product of the Fourier transformation and the outer chirp. Similarly, for the individual outer chirp,

$$\left| \frac{1}{2\pi} \frac{\partial \Phi(X)}{\partial X} \right|_{\max} \leq \frac{1}{2\Delta_X}, \quad (9)$$

where $\Phi(X) = \pi X^2/\lambda z$ is the phase of outer chirp. Before solving Eq. (9), we need to determine the sampling interval Δ_X . Suppose there are \hat{N} sampling points inside the range of $[-\lambda z/2\Delta_x, \lambda z/2\Delta_x]$, thus

$$\Delta_X = \frac{\lambda z}{\hat{N}\Delta_x}. \quad (10)$$

Readers may ask that why N points before the Fourier transform and \hat{N} points after the Fourier transform. The answer is that we can do the Fourier transform of any number of points by FFT. Note that when calculate Eq. (9), the maximum value of the left-hand side should be taken when $X = \lambda z/2\Delta_x - N\Delta_x/2$, because this is the largest value of correct sampling range.

Combining Eq. (9) and Eq. (10), we have

$$z \leq \frac{(\hat{N} + N)\Delta_x^2}{\lambda}, \quad (11)$$

which means that the propagation distance should be no larger than $(\hat{N} + N)\Delta_x^2/\lambda$.

For the final complex amplitude of the diffraction field, there should be

$$|f_x(X)|_{\max} \leq \frac{1}{2\Delta_X}, \quad (12)$$

where $f_x(X)$ is the local spatial frequency of the output field $\tilde{u}(X)$. Due to the linear property of the free-space propagation, the maximum of $f_x(X)$ is equivalent to the maximum of $f_x(x)$. Therefore, we have

$$\frac{1}{2\Delta_x} - \frac{N\Delta_x}{2\lambda z} \leq \frac{1}{2\Delta_X}. \quad (13)$$

Substituting Eq. (10) into Eq. (13), we can get

$$\hat{N} \geq \frac{\lambda z}{\Delta_x^2} - N. \quad (14)$$

If $\hat{N} < N$, we can set $\hat{N} = N$ since there are N points as the input of the calculation.

Summarizing the above analysis, we can get the following calculation strategy:

- (1) Give the sampling number N and sampling interval Δ_x of the input field; as well as the illumination wavelength λ ;
- (2) Give the propagation distance z . To ensure the correct sampling of the inner chirp function, Eq. (4) should be satisfied;
- (3) Determine the sampling number \hat{N} of the diffraction field according to Eq. (14). If $\hat{N} < N$, set $\hat{N} = N$;
- (4) Do \hat{N} -point FFT to the N -point sampled $u(x)\exp(\pi x^2/\lambda z)$; and multiply the \hat{N} -point sampled $\exp(\pi X^2/\lambda z)$ to the Fourier transformation;

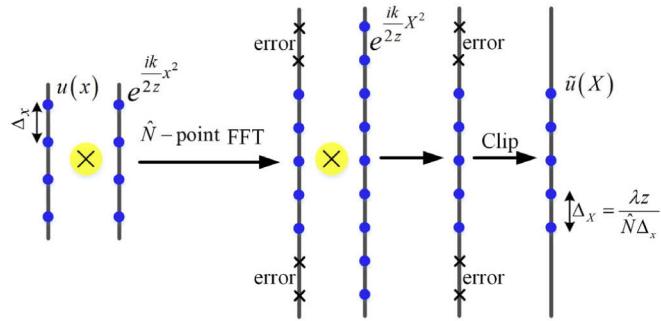


Fig. 2. Flow chart of the proposed aliasing-free calculation strategy for SFR.

(5) Clip the calculated diffraction field to the range given by Eq. (8).

Please note that Eq. (11), which is the condition to correctly sample the outer chirp, is automatically satisfied. This strategy is intuitively shown by a flowchart in Fig. 2.

To verify the proposed strategy, we carry out a numerical demonstration. A linear chirp grating is used as the input field. The reason why using such an input field is because it has rich frequency components so that it can clearly show the response of different frequencies. Furthermore, an input field with rich frequency components is more in line with the actual situation. The analytical expression of the input field and simulation parameters are listed in Table 2; and the propagation distance is $z = 10$ mm. Note that the expression of the input field is in a concise form of $|\cos(400\pi x^2)|\exp[i\varphi(x)]$. Phase $\varphi(x) = 0$ for those x values where $\cos(400\pi x^2) \geq 0$ and $\varphi(x) = \pi$ for those x values where $\cos(400\pi x^2) < 0$. For simplicity, $|\cos(400\pi x^2)|\exp[i\varphi(x)]$ is written as $\cos(400\pi x^2)$. This work is studied from the perspective of holography and diffractive optics. Commonly used sampling intervals are a few microns. Therefore, we choose a $2 \mu\text{m}$ sampling interval which is appropriate.

Table 2. Input field and simulation parameters used in numerical calculations

Input field	$u(x) = \cos(400\pi x^2)$
parameters	$N = 500, \Delta_x = 2 \mu\text{m}, \lambda = 500 \text{ nm}$

As shown in Fig. 3, where the results by the conventional method [22], proposed strategy, and analytical Fresnel integral (as reference) are given, the proposed strategy gives an accurate result by avoiding the aliasing error caused by the incorrect sampling. In contrast, the conventional method gives the aliased result because only considering the sampling of inner chirp cannot satisfy the sampling theorem. For the amplitude, there is aliasing error of the result by the conventional method because only the range given in Eq. (8) is correct. By cutting off the aliasing part, no error occurs in the result by the proposed strategy. For the phase, worse situation happens in the conventional method because the outer chirp cannot be correctly sampled. Most of the output field is wrong. On the contrary, all the output field given by the proposed strategy is accurate thanks to the \hat{N} -point FFT. For a clear comparison, the results by the conventional method within

the corresponding range, which is $[-L/2, L/2]$ with $L = \lambda z/\Delta_x - N\Delta_x = 1.5$ mm, are stretched to align with the results by the proposed strategy. By a detailed frequency analysis, the proposed strategy can give an accurate result, either the amplitude or the phase, whose correlation to that by analytical Fresnel integral is 1. Note that this strategy guarantees the transform correctness of $u(x)$ within the bandwidth of $B = [N\Delta_x/2\lambda z - 1/2\Delta_x, 1/2\Delta_x - N\Delta_x/2\lambda z]$, which is lower than the full bandwidth of $[-1/2\Delta_x, 1/2\Delta_x]$. The strategy for correct full-bandwidth transfer can be found in [21].

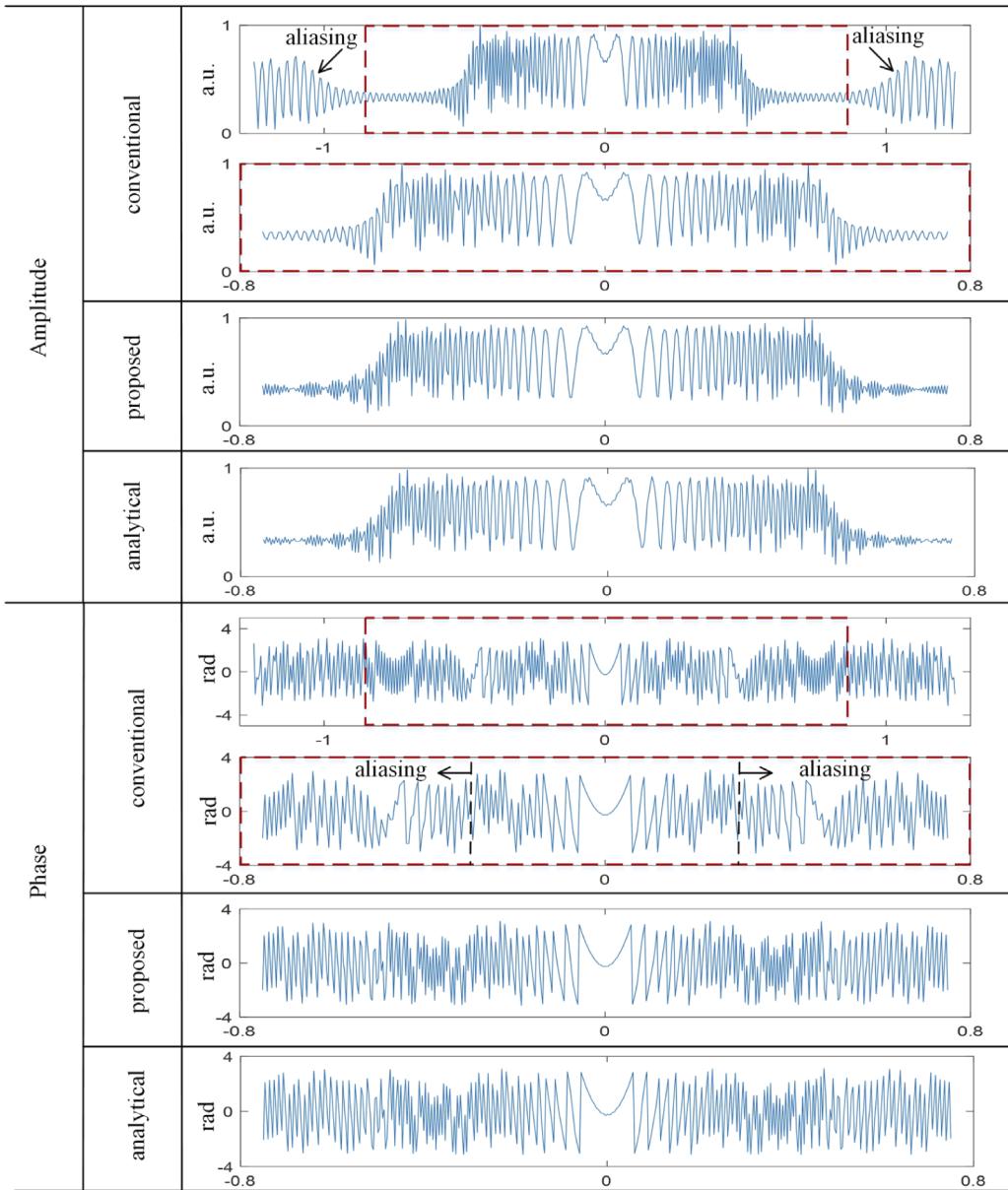


Fig. 3. Calculation results by conventional method, proposed strategy and analytical Fresnel integral. The x -axis unit is mm.

All the above analyses are carried out mainly from the perspective of discrete computation. They should be consistent with the physics. The physical model is shown in Fig. 4, where the frequency response is intuitively illustrated. Before going into the detail, we would like to mention that the free-space propagation has the linear space-invariant property. Thus, conventional method only considers the central point of the input field. The input field is sampled by N points with the interval Δ_x ; and the Fourier transform is implemented by FFT. Therefore, the diffraction range of the output field is determined as $\lambda z/\Delta_x$. Within such a range, all the diffracted light from the central point can be correctly calculated. Problems come with the diffraction light from the edge points. If the spatial frequency of the input field located at the edge point is $1/2\Delta_x$, its diffracted light would exceed the received range, as indicated by the symbol ① in Fig. 4. Due to the circular property of FFT, this part would go into the received range from the other side, as indicated by the symbol ②, which would cause errors. Therefore, the range without errors is

$$L = 2 \left(\frac{\lambda z}{2\Delta_x} - \frac{N\Delta_x}{2} \right) = \frac{\lambda z}{\Delta_x} - N\Delta_x, \quad (15)$$

which is consistent with Eq. (8). Of course, if the spatial frequency of the input field located at the edge point is not so high, the range without error could be larger. However, as a diffraction calculation method where accuracy is essential, we should guarantee its correctness no matter what input field is given.

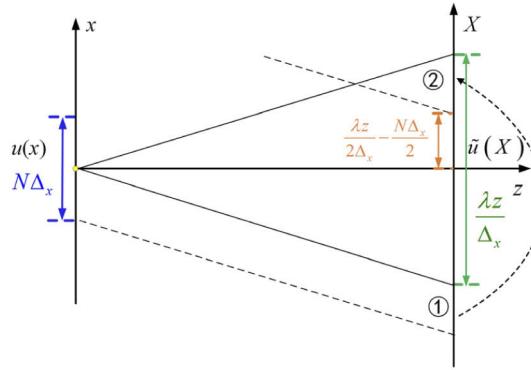


Fig. 4. Physical model used to explain the SFR method.

4.2. ASM and Fres-TF

ASM and Fres-TF model diffraction in the spatial frequency domain, where the transfer function is used, as indicated in Eq. (16). Because the transfer function of Fres-TF is the paraxial approximation version of that of ASM [18], we focus on the analysis of ASM in this section. Fres-TF case can be easily obtained under a paraxial approximation in a similar way.

$$\begin{aligned} \tilde{u}(X) &= IFT\{FT\{u(x)\} \times H(f_x)\} \\ H(f_x) &= \exp\left(i\frac{2\pi}{\lambda}z\sqrt{1-(\lambda f_x)^2}\right). \end{aligned} \quad (16)$$

As a convolution-based method, ASM usually requires zero padding for accurate calculation [8]. Traditionally, the number of padded zeros is equal to the sampling number of the input field, which is $N_p = N$. However, we find that it is not always necessary to pad such many zeros. The number of padded zeros increases from 0 to N with the propagation distance, which would be given with detailed analysis in the following. Due to the symmetrical operation of the ASM

calculation indicated in Eq. (18), one FFT and one IFFT, the sampling interval of the output field is the same as the input field. And in most cases, the size of the output field is set to be the same as the input field. For ASM, the key point is to correctly discretize the transfer function.

According to the transform property of FFT as given in Eq. (1), the Fourier spectrum of the input field is sampled by $N_p + N$ points with the interval

$$\Delta_f = \frac{1}{(N + N_p)\Delta_x}. \quad (17)$$

Thus, the transfer function $H(f_x)$ is also sampled with these parameters. By applying the sampling theorem to the phase kernel $\varphi(f_x) = \frac{2\pi}{\lambda}z\sqrt{1 - (\lambda f_x)^2}$ of $H(f_x)$,

$$\left| \frac{1}{2\pi} \frac{\partial \varphi(f_x)}{\partial f_x} \right|_{\max} \leq \frac{1}{2\Delta_f}, \quad (18)$$

we have

$$\left| \frac{\lambda z f_x}{\sqrt{1 - (\lambda f_x)^2}} \right|_{\max} \leq \frac{(N + N_p)\Delta_x}{2}. \quad (19)$$

Substituting the highest achievable spatial frequency $f_{x_max} = 1/2\Delta_x$ into Eq. (19) and after some simplifications, we can get

$$z \leq \frac{(N + N_p)\Delta_x^2}{\lambda} \sqrt{1 - \left(\frac{\lambda}{2\Delta_x} \right)^2}. \quad (20)$$

Equation (20) means that the transfer function can only be correctly sampled in such a distance range. Generally speaking, the number of padded zeros is no more than the sampling number of the input field, i.e. $N_p \leq N$; because when the propagation distance is larger than the critical distance z_c , which is

$$z > z_c = \frac{2N\Delta_x^2}{\lambda} \sqrt{1 - \left(\frac{\lambda}{2\Delta_x} \right)^2}, \quad (21)$$

we can use RSC or Fres-IR for a more efficient calculation, which will be introduced in the next subsection.

We can see that the upper limit of propagation distance given by Eq. (20) varies with the number of padded zeros. Does this mean the number of padded zeros required at different distances is also different? The answer is yes. To quantitatively study this problem, we need to analyze it from the perspective of frequency response of the discrete ASM diffraction model. As shown in Fig. 5(a), physically, the diffraction range increases with the propagation distance. However, in ASM calculation, the output field has the same size as the input field. Thus, we need to deal with the exceeded part of the diffraction field. Numerically, the Fourier transform in Eq. (16) is implemented by the FFT algorithm, which has the circular property. The exceeded part would get into the output field from the other side, as shown in Fig. 5(a), which would cause calculation error. It is evident that the size of wrong part would increase with the propagation distance. This effect is usually called circular convolution errors. To get rid of it, zero padding is used to enlarge the calculation window, as shown in Fig. 5(b). Since the exceeded part would cause error, we can control the number of padded zeros to prevent the exceeded part from entering

the output field. From Fig. 5(b), the number of padded zeros can be calculated as

$$N_p = 2 \times \frac{z \tan \theta}{2\Delta_x} = \frac{\lambda z}{2\Delta_x^2} \left[1 - \left(\frac{\lambda}{2\Delta_x} \right)^2 \right]^{-\frac{1}{2}}, \quad (22)$$

where

$$\theta = \arcsin \left(\frac{\lambda}{2\Delta_x} \right), \quad (23)$$

is the maximum diffraction angle. Note that the maximum diffraction angle in ASM does not change with the propagation distance because it is a diffraction algorithm based on plane wave model [18]. Therefore, the padded zeros only play a role in avoiding circular convolution errors. It is clear that when the propagation distance increases from 0 to the critical distance z_c , the number of padded zeros increases from 0 to N , as shown in Fig. 6. Therefore, it is not always necessary to pad N zeros. Less number of padded zeros leads to a more efficient calculation.

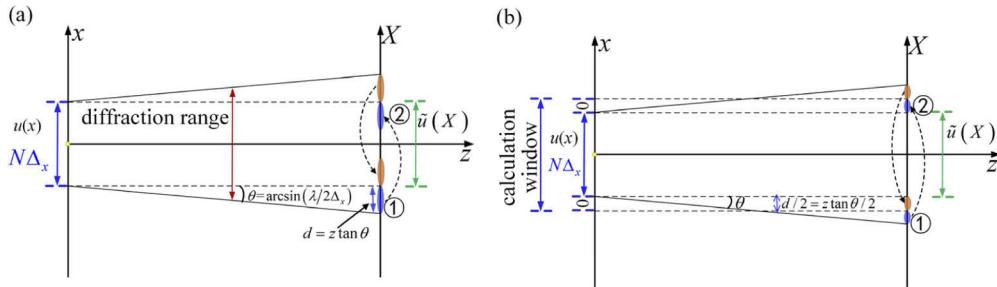


Fig. 5. Schematic diagram to show the effect of circular convolution errors (a) and adaptive zero padding method to avoid it (b).

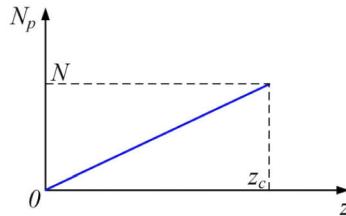


Fig. 6. Number of padded zeros varies with the propagation distance.

To prove that Eq. (22) can guarantee the correctness of ASM, we demonstrate a numerical calculation. The input information are listed in Table 2 and the propagation distance is set as $z = 3$ mm. Figure 7(a) shows the results with $N_p = 190$ padded zeros; and Fig. 7(b) shows the results with $N = 500$ padded zeros. The cross correlation between the amplitude fields by these two methods is 1, so is the cross correlation between the phase fields, which proves the effectiveness of the proposed method.

The procedure of the proposed adaptive zero padding ASM, as flow chart shown in Fig. 8, can be concluded as following:

- (1) Give the sampling number N and sampling interval Δ_x of the input field; as well as the illumination wavelength λ ;
- (2) Give the propagation distance z which is no larger than z_c given in Eq. (21);

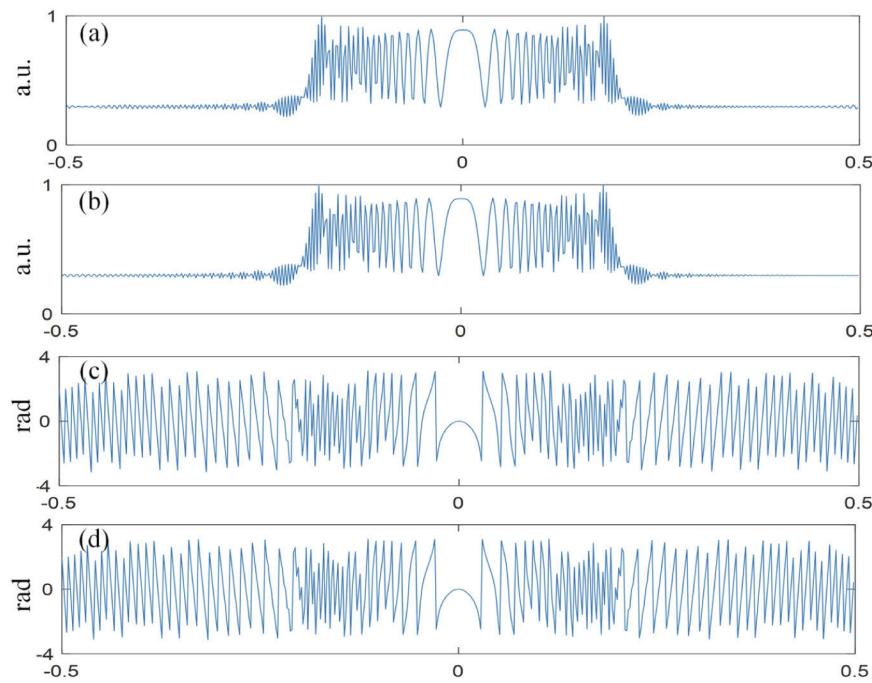


Fig. 7. Calculation results by the proposed adaptive zero padding ASM: (a) amplitude and (c) phase; by the conventional ASM: (b) amplitude and (d) phase. The x -axis unit is mm.

- (3) Calculate the number of padded zeros N_p according to Eq. (22), and pad zeros to the input field;
- (4) Fourier transform the zero-padded input field by FFT to get its Fourier spectrum;
- (5) Model the transfer function with $N + N_p$ points according to Eq. (16) [second equation];
- (6) Inverse Fourier transform the product of the outputs of step 4 and step 5 by IFFT;
- (7) Clip the calculated diffraction field to the size of $N\Delta_x$.

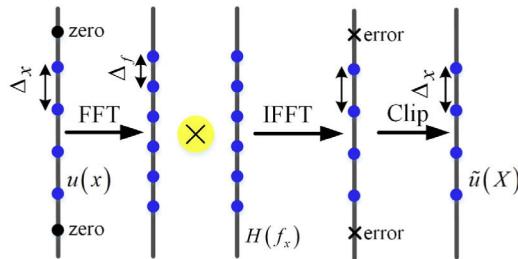


Fig. 8. Flow chart of the proposed adaptive zero padding ASM.

4.3. RSC and Fres-IR

Different from ASM (FRes-TF), RSC (Fres-IR) models the diffraction in the space domain, where the impulse response function is used, as indicated in Eq. (24). Because the impulse response

function of Fres-IR is the paraxial approximation version of that of RSC [18], here, we focus on the analysis of RSC. Fres-IR case can be easily obtained under a paraxial approximation in a similar way.

$$\begin{aligned}\tilde{u}(X) &= IFT\{FT\{u(x)\} \times FT\{h(x)\}\} \\ h(x) &= \frac{1}{2\pi} \frac{z}{r} \left(\frac{1}{r} - \frac{i2\pi}{\lambda} \right) \frac{\exp(i2\pi r/\lambda)}{r}, r = \sqrt{x^2 + z^2}.\end{aligned}\quad (24)$$

Also, as a convolution-based method, RSC needs zero padding for accurate calculation. However, zero padding is not only being used to get rid of circular convolution errors, but also to increase the transfer bandwidth. Let us go into detail of RSC. Due to the symmetrical operation of the RSC calculation indicated in Eq. (24), two FFTs and one IFFT, the sampling interval of the output field is the same with that of the input field. And in most cases, the size of the output field is set to be the same as the input field. For RSC, the key point is to correctly discretize the impulse response function, including the sampling number and sampling interval. There are two options: (1) modeling $h(x)$ by N points and padding N_p zeros; or (2) modeling $h(x)$ by $(N + N_p)$ points. From the perspective of discrete computation, i.e. to get rid of the circular convolution errors, option (1) should be used with $N_p = N$ [20]. However, the appropriate option should be determined through the frequency response of the RSC model.

RSC is a spherical-wave model [18]. The impulse response function $h(x)$ describes a weighted spherical wave from one point of the input field; and the diffraction field is the coherent superposition of these spherical waves from all the points of the input field. Due to the linear space-invariant property of free-space diffraction, we just analyze the response from the central point of the input field and then extend to the edge points. As shown in Fig. 9(a), if option (1) is used, the response range of the central point (green line) cannot cover the full calculation window. Neither can the edge points. Therefore, the response from point A of the input field cannot reach point B of the output field, which limits the transfer bandwidth. In other words, at any point (except the central one), the output field is not the superposition of the response from all the points of the input plane. To achieve such a full-bandwidth transfer, option (2) is used, as shown in Fig. 9(b). Because the impulse response function is modeled by $N + N_p = 2N$ points, the response range of the central point of the input field can cover the full calculation window. However, the diffraction light from the edge point would exceed the calculation window and the exceeded part would get into the calculation window from the other side, which would cause errors in those areas. Fortunately, the output field in concern has a smaller size than the calculation window; and the calculation accuracy of the output field can be guaranteed, as shown in Fig. 9(b).

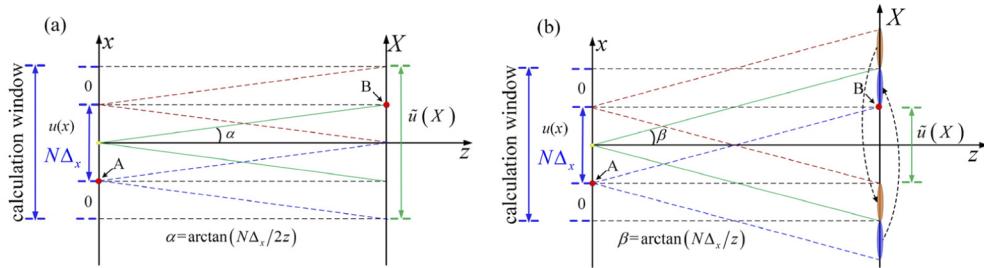


Fig. 9. Schematic diagram to show zero padding. (a) Option (1) is used with a limited transfer bandwidth; (b) option (2) is used with a full transfer bandwidth.

There is one point needing attention. The diffraction angle $\beta = \arctan(N\Delta_x/z)$ shown in Fig. 9(b) should be no more than the highest potential diffraction angle $\theta = \arcsin(\lambda/2\Delta_x)$,

therefore, we have

$$\arctan\left(\frac{N\Delta_x}{z}\right) \leq \arcsin\left(\frac{\lambda}{2\Delta_x}\right). \quad (25)$$

After some simplification, we get

$$z \geq z_c = \frac{2N\Delta_x^2}{\lambda} \sqrt{1 - \left(\frac{\lambda}{2\Delta_x}\right)^2}, \quad (26)$$

which means that RSC can be only used in such a distance range. The same conclusion can be obtained by applying the sampling theorem on the impulse response function of RSC.

To prove the above analysis, we carry out one numerical demonstration with the parameters shown in Table 2 and the propagation distance is set as $z = 50$ mm. Figure 10(a) shows the results by option (1); and Fig. 10(b) shows the results by option (2). It is clear that there are no circular convolution errors of the result by option (1) inside the whole calculation window, but the high-frequency components are lost, which cannot give an accurate physics-consistent output field as shown in Fig. 10(c) given by Rayleigh-Sommerfeld integral (RSI) [2]. In contrast, the result by option (2), a full-bandwidth calculation, has more high-frequency details which is consistent with the real physics. As a price, only the range of the same size as the input field is correct. Readers may notice that there is still a little difference between the result by option (2) and that by RSI. This is because only N discrete points of the input field contribute to the result by option (2); but the result by RSI is analytically calculated which would give a more accurate result. However, such a difference would not have much impact on calculation accuracy. Through the above analysis, we reveal that in RSC, zero padding not only avoids circular convolution errors, but also expands the transfer bandwidth, as comparison between Fig. 10(a) and Fig. 10(b).

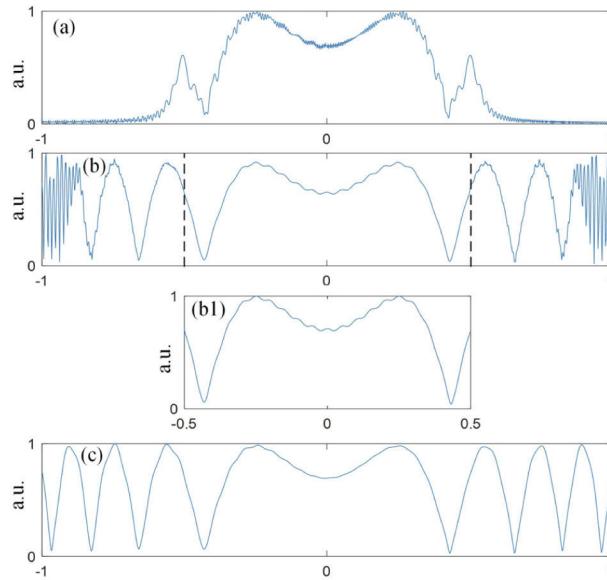


Fig. 10. Calculated amplitude field by option (1): (a); option (2): (b); and RSI: (c). Figure 10(b1) is the zoomed part of Fig. 10(b) with a renormalization. The x -axis unit is mm.

The procedure of option (2), as flow chart shown in Fig. 11, can be concluded as following:

- (1) Give the sampling number N and sampling interval Δ_x of the input field; as well as the illumination wavelength λ ;

- (2) Give the propagation distance z which is no smaller than z_c given in Eq. (26);
- (3) Pad N zeros to the input field;
- (4) Fourier transform the zero-padded input field by FFT to get its Fourier spectrum;
- (5) Model the impulse response function with $2N$ points according to Eq. (24) [second equation], and get its Fourier spectrum by FFT;
- (6) Inverse Fourier transform the product of the outputs of step 4 and step 5 by IFFT;
- (7) Clip the calculated diffraction field to the size of $N\Delta_x$.

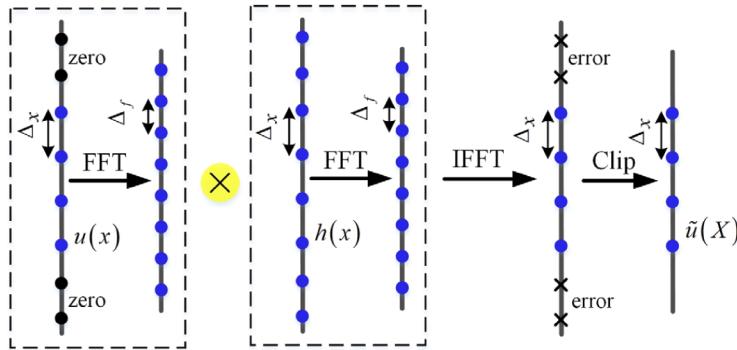


Fig. 11. Flow chart of option (2) of the common-used RSC.

4.4. Summary

In this section, we have analyzed several popular diffraction calculation methods. For each method, we provide a detailed frequency response analysis and give the corresponding sampling strategy. Readers may find that ASM is only applicable for near-field calculation and RSC is only for far-field calculation. To make ASM suitable for far field, some methods have been proposed [8,9,24,25]. One popular method of them is BLASM [8]. Compared with ASM, BLASM can work in a larger distance range. However, the performance of BLASM at long distance is also limited [9,25]. In the next section, we would give the reason why this happens although all the effective bandwidth is maintained in BLASM; and provide a solution of BEASM [9].

5. BLASM and BEASM

BLASM works by limiting the effective bandwidth of the transfer function $H(f_x)$. When the propagation distance is larger than the critical distance given in Eq. (21), $H(f_x)$ is aliased which would cause high-frequency noise [8,9]. By cutting off the aliased part, BLASM works in a larger distance range. It has been proven that the reserved bandwidth without aliasing is exactly what is needed for the output field. Here, we give a brief introduction of BLASM and detailed analysis can be found in [8].

Let us revisit Eq. (19) and set $N_p = N$. By regarding f_x as the variable, we can get

$$|f_x| \leq f_{\max} = \frac{N\Delta_x}{\lambda z \sqrt{1 + (N\Delta_x/z)^2}}, \quad (27)$$

which means only the frequency components inside this range can be correctly calculated. This

range is exactly what is needed for the output field, as shown in Fig. 12. From Eq. (27) or Fig. 12, we can see that the effective bandwidth $B_f = 2f_{max}$ decreases as the propagation distance z increases. Therefore, it is needed to cut off the exceeded part of the transfer function $H(f_x)$. Physically, this operation would give accurate results. However, numerically, due to the shrink of the effective bandwidth, the effective sampling number would also decrease with the propagation distance. When the propagation distance z is relatively large, both the effective bandwidth and

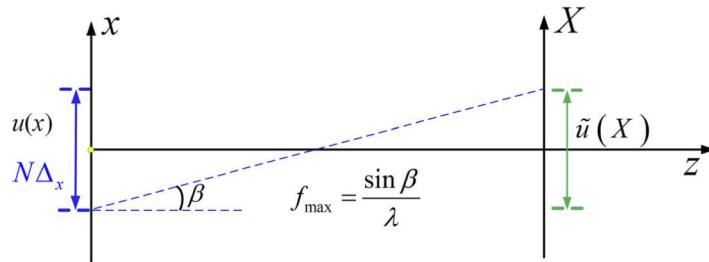


Fig. 12. Schematic diagram of the bandwidth required in BLASM.

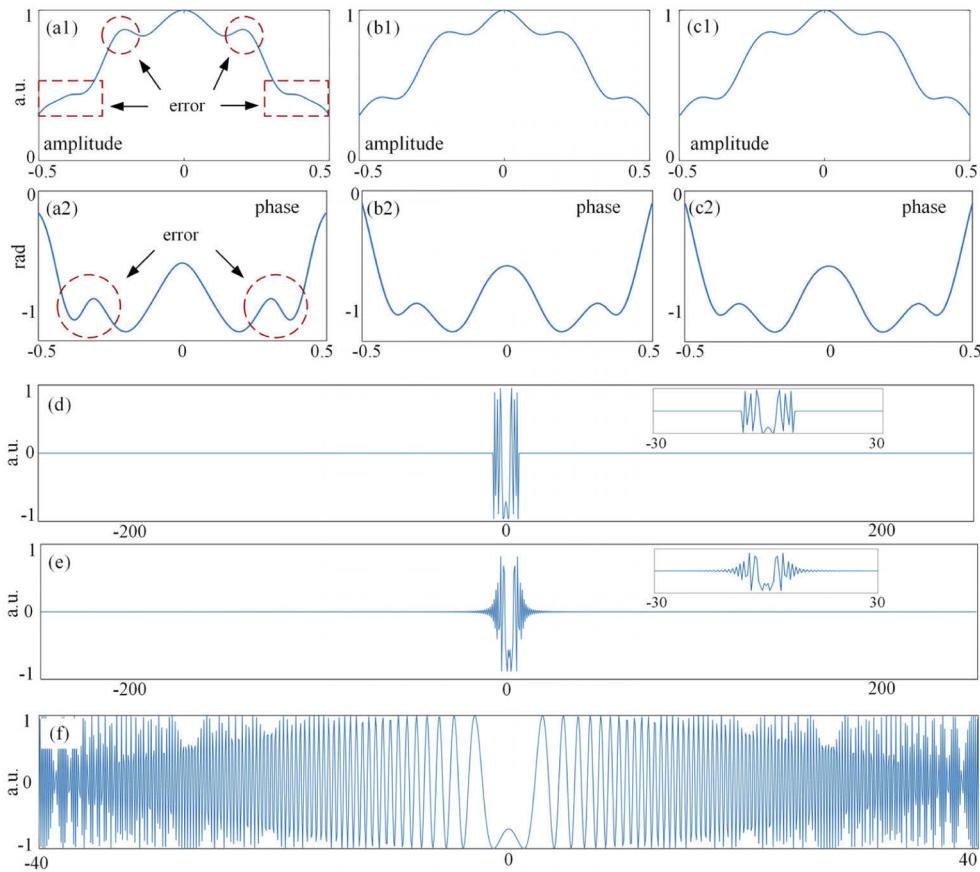


Fig. 13. Calculation results by BLASM (a); by RSC (b) and by BEASM (c). Imaginary part of the transfer function of BLASM (d); RSC (e); and BEASM (f). The x -axis unit for (a-c) is mm and for (d-f) is mm^{-1} .

sampling number are small. With a hard truncation of the aliased transfer function, the calculation accuracy would be affected.

A numerical demonstration is carried out to illustrate this effect. A rectangle aperture with the width of 0.8 mm is used as the input field. The sampling number and interval of the input field are $N = 500$ and $\Delta_x = 2 \mu\text{m}$, respectively. The illumination wavelength is $\lambda = 500 \text{ nm}$ and the propagation distance is $z = 300 \text{ mm}$. Figure 13(a) shows the result by BLASM. As analyzed in Section 4.3, RSC is applicable in far-field calculation. Therefore, we give the result by RSC in Fig. 13(b). From the comparison, we can see that the performance of BLASM is poor. The only difference between BLASM and RSC is the transfer function. In BLASM, the transfer function is directly modeled with $H(f_x)$ and followed with a band limiting operation. In RSC, the transfer function is obtained by the Fourier transformation of the impulse response function $h(x)$. To show the difference, we draw these two transfer functions in Figs. 13(d) and (e). It is evident that BLASM has a hard truncation and RSC has a soft truncation. Therefore, it is the hard truncation that makes BLASM fail in relative far-field calculation.

To overcome this problem, we have previously proposed a BEASM for accurate calculation in a wide propagation distance range [9]. In BEASM, the bandwidth in the calculation is larger than that is physically needed. Therefore, for the needed bandwidth, there is no hard truncation any more, which gives an accurate calculation. For the same calculation task, the result by BEASM is shown in Fig. 13(c); and the transfer function of BEASM is shown Fig. 13(f). It is clear that the BEASM works better than BLASM, which verifies our analysis.

6. Discussion and conclusion

It is obvious that the Nyquist sampling criteria is used in this work. It is worth mentioning that there may be two kinds of numerical diffraction calculations. Diffraction calculation of an analytical signal and diffraction calculation of a discrete signal. For the analytical case, the first step is to discretize the signal with proper sampling parameters. After discretization, the numerical diffraction can be followed. The space bandwidth product (SBP) of the signal is *a prior* knowledge for the discretization and calculation [26]. For the discrete case, the sampling parameters, i.e., the sampling number and sampling interval, are first given, not determined by analytical signals, such as the diffraction calculation used in digital holography. Thus, we need to ensure the calculation correctness by applying the Nyquist criteria for a robust result. Of course, if the SBP of the discrete signal can be evaluated, more efficient calculation could be possible by considering the sampling of the evaluated SBP not the upper bound given by the sampling parameters. In this work, we study the discrete case.

To summarize, we have analyzed several popular diffraction calculation methods from a joint physics-mathematics perspective. For each method, the frequency response has been carefully studied, by which we found (1) the conventional single-Fourier transform based Fresnel diffraction (SFR) cannot deal with the full bandwidth of the input field; (2) there is no need to always pad N zeros in angular spectrum method (ASM); (3) zero padding in Rayleigh-Sommerfeld convolution (RSC) not only avoids circular convolution errors, but also expands the transfer bandwidth. Based on these findings, we have proposed a new aliasing-free SFR method and an adaptive zero-padding ASM. In addition, through the comparison between band-limited ASM (BLASM) and RSC, we found it is the hard truncation that makes BLASM fail in long propagation distance and introduced band-extended ASM as an effective solution. All the Matlab codes of this work can be found in [27].

The key contribution of this work is the idea of frequency response analysis based on the joint perspective of physics and mathematics. With this idea, for any diffraction calculation task, readers can make a do-it-yourself analysis and choose or develop a method for accurate and efficient computation. More flexible calculation methods, such as shifting and scaling, can also be analyzed with this idea.

Funding

National Natural Science Foundation of China (61875105, 62035003).

Disclosures

The authors declare no conflicts of interest.

References

1. M. Born and E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light* (Elsevier, 2013).
2. J. W. Goodman, *Introduction to Fourier optics* (W. H. Freeman, Macmillan Learning, 2017).
3. C. J. R. Sheppard and M. Hrynevych, "Diffraction by a circular aperture: a generalization of Fresnel diffraction theory," *J. Opt. Soc. Am. A* **9**(2), 274–281 (1992).
4. U. Schnars and W. Jüptner, "Direct recording of holograms by a CCD target and numerical reconstruction," *Appl. Opt.* **33**(2), 179–181 (1994).
5. W. Zhang, H. Zhang, and G. Jin, "Resolution-enhanced digital in-line holography by extension of the computational bandwidth," *Opt. Commun.* **472**, 126038 (2020).
6. A. L. Robisch, J. Wallentin, A. Pacureanu, P. Cloetens, and T. Salditt, "Holographic imaging with a hard x-ray nanoprobe: ptychographic versus conventional phase retrieval," *Opt. Lett.* **41**(23), 5519–5522 (2016).
7. D. Paganin, *Coherent X-ray optics* (Oxford University, 2006).
8. K. Matsushima and T. Shimobaba, "Band-Limited Angular Spectrum Method for Numerical Simulation of Free-Space Propagation in Far and Near Fields," *Opt. Express* **17**(22), 19662–19673 (2009).
9. W. Zhang, H. Zhang, and G. Jin, "Band-extended angular spectrum method for accurate diffraction calculation in a wide propagation range," *Opt. Lett.* **45**(6), 1543–1546 (2020).
10. R. P. Muffoletto, J. M. Tyler, and J. E. Tohline, "Shifted Fresnel diffraction for computational holography," *Opt. Express* **15**(9), 5631–5640 (2007).
11. F. Zhang, I. Yamaguchi, and L. P. Yaroslavsky, "Algorithm for reconstruction of digital holograms with adjustable magnification," *Opt. Lett.* **29**(14), 1668–1670 (2004).
12. T. Shimobaba, K. Matsushima, T. Kakue, N. Masuda, and T. Ito, "Scaled angular spectrum method," *Opt. Lett.* **37**(19), 4128–4130 (2012).
13. K. Matsushima, "Shifted angular spectrum method for off-axis numerical propagation," *Opt. Express* **18**(17), 18453–18463 (2010).
14. C. Guo, Y. Xie, and B. Sha, "Diffraction algorithm suitable for both near and far field with shifted destination window and oblique illumination," *Opt. Lett.* **39**(8), 2338–2341 (2014).
15. Z. Wang, O. Baladron-Zorita, C. Hellmann, and F. Wyrowski, "Generalized Debye integral," *Opt. Express* **28**(17), 24459–24470 (2020).
16. D. Asoubar, S. Zhang, F. Wyrowski, and M. Kuhn, "Efficient semi-analytical propagation techniques for electromagnetic fields," *J. Opt. Soc. Am. A* **31**(3), 591–602 (2014).
17. Z. Wang, O. Baladron-Zorita, C. Hellmann, and F. Wyrowski, "Theory and algorithm of the homeomorphic Fourier transform for optical simulations," *Opt. Express* **28**(7), 10552–10571 (2020).
18. W. Zhang, H. Zhang, C. J. R. Sheppard, and G. Jin, "Analysis of numerical diffraction calculation methods: from the perspective of phase space optics and the sampling theorem," *J. Opt. Soc. Am. A* **37**(11), 1748–1766 (2020).
19. W. Zhang, H. Zhang, and G. Jin, "Adaptive-sampling angular spectrum method with full utilization of space-bandwidth product," *Opt. Lett.* **45**(16), 4416–4419 (2020).
20. R. G. Lyons, *Understanding digital signal processing, 3/E* (Pearson Education India, 2004).
21. W. Zhang, H. Zhang, and G. Jin, "Single-Fourier transform based full-bandwidth Fresnel diffraction," *J. Opt.* (under review).
22. D. Voelz and M. Roggemann, "Digital simulation of scalar optical diffraction: revisiting chirp function sampling criteria and consequences," *Appl. Opt.* **48**(32), 6132–6142 (2009).
23. M. Hillenbrand, D. Kelly, and S. Sinzinger, "Numerical solution of nonparaxial scalar diffraction integrals for focused fields," *J. Opt. Soc. Am. A* **31**(8), 1832–1841 (2014).
24. T. Kozacki and K. Falaggis, "Angular spectrum-based wave-propagation method with compact space bandwidth for large propagation distances," *Opt. Lett.* **40**(14), 3420–3423 (2015).
25. Y. Xiao, X. Tang, Y. Qin, H. Peng, and W. Wang, "Wide-window angular spectrum method for diffraction propagation in far and near field," *Opt. Lett.* **37**(23), 4943–4945 (2012).
26. D. Kelly, "Numerical calculation of the Fresnel transform," *J. Opt. Soc. Am. A* **31**(4), 755–764 (2014).
27. W. Zhang, H. Zhang, and G. Jin, "Frequency-sampling-strategy-for-numerical-diffraction-calculation," Github (2020), <https://github.com/thu12zwh/Frequency-sampling-strategy-for-numerical-diffraction-calculation>.