Fast-Fourier-transform based numerical integration method for the Rayleigh–Sommerfeld diffraction formula

Fabin Shen and Anbo Wang

The numerical calculation of the Rayleigh–Sommerfeld diffraction integral is investigated. The implementation of a fast-Fourier-transform (FFT) based direct integration (FFT-DI) method is presented, and Simpson's rule is used to improve the calculation accuracy. The sampling interval, the size of the computation window, and their influence on numerical accuracy and on computational complexity are discussed for the FFT-DI and the FFT-based angular spectrum (FFT-AS) methods. The performance of the FFT-DI method is verified by numerical simulation and compared with that of the FFT-AS method.

OCIS codes: 050.1960, 000.4430, 000.3860, 350.5500.

1. Introduction

The Helmholtz–Kirchhoff and the Rayleigh–Sommerfeld diffraction formulas have been widely used to analyze the propagation and diffraction of light in an isotropic, homogenous linear medium and have proved to be valid when the aperture size of the light field is much larger than the wavelength of light.[1,2] The Rayleigh–Sommerfeld diffraction formula has been proved to yield exact evaluations of the light fields by methods of Fourier treatment[3–5] and gives correct results for both far-field and near-field diffraction.[6,7]

Unfortunately, in most cases, these formulas cannot be solved analytically. Some approximations, such as Fresnel and Fraunhofer, are used to ease the difficulties in using these integral formulas for regions not close to the aperture. Also, paraxial approximations are often used for regions close to the optical axis. The validity of these approximations has been one issue in diffraction theories.[8,9]

The development of high-speed computers makes it possible to calculate these diffraction formulas without the need for Fresnel or Fraunhoffer approximations. In this paper we focus on the numerical calculation of the Rayleigh–Sommerfeld diffraction integral of an aperture under normal chromatic plane-wave illumination. The observation plane is parallel to the aperture plane, as Fig. 1 shows. The numerical solution of the Rayleigh–Sommerfeld integral can be calculated by either of two methods: angular spectrum (AS) and direct integration (DI). The AS method treats the propagation of light as a superposition of plane waves with different wave vectors and uses a Fourier transform to compute the light fields in the spatial-frequency domain.[5,10,11] A fast-Fourier-transform (FFT) based AS (FFT-AS) method can have a high calculation speed and can be used for both parallel and arbitrarily oriented planes.[12] The DI method computes the diffraction integrals in the spatial domain by means of numerical integration, which can be treated as a linear convolution and can be effectively calculated by means of a FFT and an inverse FFT (IFFT).[13–15]

Although numerical calculation of both the AS and the DI methods has been discussed extensively in the literature,[4,6,7,12–16] there is still some confusion to be cleared up. First, the implementation of the diffraction integral by means of convolution and a FFT was not clearly presented. Proper implementation requires that the sampling intervals be properly selected and the discrete Fourier transformation be properly zero padded; improper implementation will cause erroneous results. Second, the selection of the sampling interval for a given diffraction problem was not clearly discussed. Li *et al.* suggested using an energy conservation criterion for validation of the sampling quality, but it is difficult to select a proper sampling interval in advance.[15] Delen and Hooker used a small sampling interval (less than half of a

The authors are with the Center for Photonics Technology, Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061. F. Shen's e-mail address is fashen1@vt.edu.

Received 15 March 2005; revised 9 September 2005; accepted 22 September 2005.

0003-6935/06/061102-09$15.00/0

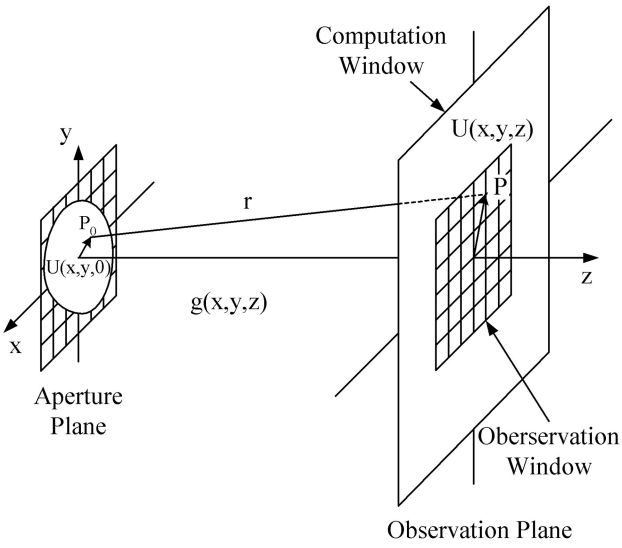

Fig. 1. Illustration of the coordinate system of the Rayleigh–Sommerfeld diffraction theory.

wavelength) for the FFT-AS method but encountered a large computational load.[16] An improper sampling interval will cause either a large aliasing error or a large computation load. Third, the selection of the computation window size in the FFT-AS method and its influence on numerical calculation accuracy, especially for an observation window that is not on the optical axis, was not clearly discussed. Some other confusion includes the applicable regions and computational complexity of the two methods. For example, although the FFT-AS method had been considered suitable for solving full scalar propagation problems,[12] its expected numerical calculation accuracy and applicability to different regions have not been completely verified.

In this paper we present an implementation of a fast-Fourier-transform based DI (FFT-DI) method to calculate the Rayleigh–Sommerfeld diffraction integral and use Simpson's rule to improve the calculation accuracy. The selections of the sampling interval and the computation window and their influence on the calculation accuracy and computational load are discussed. The calculation accuracy, the computational speed, and the applicability regions of the FFT-DI and the FFT-AS methods are compared.

This paper is arranged as follows: In Section 2 a brief review of the AS method and the DI method is given. In Section 3 a method of implementing the FFT-DI method is described with improved accuracy by use of Simpson's rule. In Section 4 the sampling intervals, the computation window size, the calculation accuracy, and the computational load of the two methods are discussed. In Section 5 the simulation results for diffraction behind a circular aperture and a single slit are demonstrated; the calculation accuracy and computational speeds of the two methods are compared. In Section 6 our conclusions are given.

2. Review

The three-dimensional scalar Helmholtz equation for a linear homogenous isotropic medium,

$$\frac{\partial^2 U}{\partial x^2}+\frac{\partial^2 U}{\partial y^2}+\frac{\partial^2 U}{\partial z^2}+k^2U=0, \tag{1}$$

can be solved by either the AS method or the DI method described below.

A. Angular Spectrum Method

In the AS method the initial light field and its propagation are handled in the spatial-frequency domain. The propagation of the light can be given as

$$A(\alpha,\ \beta,\ z)=A(\alpha,\ \beta,\ 0)G(\alpha,\ \beta,\ z), \tag{2}$$

where $A(\alpha, \beta, z)$ is the Fourier transformation of the light field at distance z,

$$\begin{aligned}A(\alpha,\ \beta,\ z)&=F\{U(x,\ y,\ z)\}\\&=\int\int U(x,\ y,\ z)\exp(-j\alpha x-j\beta y)\mathrm{d}x\mathrm{d}y,\end{aligned} \tag{3}$$

and

$$G(\alpha,\ \beta,\ z)=\exp\left(j\sqrt{k^2-\alpha^2-\beta^2}z\right) \tag{4}$$

is the optical transfer function of the linear homogenous isotropic medium. Then the light field $U(x, y, z)$ can be solved as the two-dimensional (2D) inverse Fourier transformation of $A(\alpha, \beta, z)$, given as

$$\begin{aligned}U(x,\ y,\ z)&=F^{-1}\{A(\alpha,\ \beta,\ z)\}\\&=\frac{1}{4\pi^2}\int\int A(\alpha,\ \beta,\ 0)\\&\quad\times\exp\left(j\alpha x+j\beta y+j\sqrt{k^2-\alpha^2-\beta^2}z\right)\mathrm{d}\alpha\mathrm{d}\beta.\end{aligned} \tag{5}$$

The FFT-AS method for calculating $U(x, y, z)$ numerically has been reported and can be given as

$$Q=\mathrm{IFFT2}\{\mathrm{FFT2}\{U(x_m,\ y_n,\ 0)\}\cdot\times G(\alpha_m,\ \beta_n,\ z)\}, \tag{6}$$

where $U(x_m, y_n, 0)$ and $G(\alpha_m, \beta_n, z)$ are samples of $U(x, y, 0)$ and $G(\alpha, \beta, z)$, FFT2 and IFFT2 denote a 2D FFT and a 2D IFFT, and $\cdot\times$ means element-by-element multiplication.[5,10–12]

B. Direct Integration Method

$G(\alpha, \beta, z)$ in Eq. (4) is the optical transfer function of the medium, and its inverse Fourier transformation

will give the impulse response as

$$g(x,\ y,\ z) = \frac{1}{4\pi^2}\int\int G(\alpha,\ \beta,\ z) \times \exp(j\alpha x + j\beta y)\mathrm{d}\alpha\mathrm{d}\beta = \frac{1}{2\pi}\frac{\exp(jkr)}{r}\frac{z}{r}\left(\frac{1}{r} - jk\right), \quad (7)$$

where $r = \sqrt{x^2 + y^2 + z^2}$.[10] Thus $U(x, y, z)$ can be solved as the convolution of $U(x, y, 0)$ and $g(x, y, z)$:

$$U(x,\ y,\ z) = \int\int_A U(\varsigma,\ \eta,\ 0)g(x - \varsigma,\ y - \eta,\ z)\mathrm{d}\varsigma\mathrm{d}\eta = \int\int_A U(\varsigma,\ \eta,\ 0)\frac{\exp(jkr)}{2\pi r}\frac{z}{r} \times \left(\frac{1}{r} - jk\right)\mathrm{d}\varsigma\mathrm{d}\eta, \quad (8)$$

where $r = \sqrt{(x - \varsigma)^2 + (y - \eta)^2 + z^2}$. It is exactly the Rayleigh–Sommerfeld diffraction integral formula, which can be used for both near and far fields without any approximation.

In most cases, the diffraction integral in Eq. (8) has to be calculated by direct numerical integration. On the aperture plane, $U(\varsigma, \eta, 0)$ is sampled to $N \times N$ equidistant grids. For a point on the observation plane (x_m, y_n, z), the integral can be calculated by numerical integration as a Riemann sum:

$$U(x_m,\ y_n,\ z) = \sum_{i=1}^{N}\sum_{j=1}^{N} U(\varsigma_i,\ \eta_j,\ 0)g \times (x_m - \varsigma_i,\ y_n - \eta_j,\ z)\Delta\varsigma\Delta\eta, \quad (9)$$

where $\Delta\varsigma$ and $\Delta\eta$ are sampling intervals on the aperture plane.

3. Fast-Fourier-Transform Based Direct Integration Method

The Riemann sum in Eq. (9) can be regarded as a discrete linear convolution of $U(\varsigma_i, \eta_j, 0)$ and $g(x_m, y_n, z)$, which can be calculated effectively by means of a FFT. We present a FFT-DI implementation for N^2 points located at sampling grids on the observation plane. The discrete convolution in Eq. (9) can be calculated as

$$S = \mathrm{IFFT2}[\mathrm{FFT2}(U) \cdot\times \mathrm{FFT2}(H)]\Delta\varsigma\Delta\eta, \quad (10)$$

where

$$U = \begin{bmatrix} U_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{(2N-1)\times(2N-1)} = \left[\begin{array}{ccc|c} U(\varsigma_1,\ \eta_1,\ 0) & \cdots & U(\varsigma_1,\ \eta_N,\ 0) & \\ \vdots & \ddots & \vdots & 0_{N\times(N-1)} \\ U(\varsigma_N,\ \eta_1,\ 0) & \cdots & U(\varsigma_N,\ \eta_N,\ 0) & \\ \hline 0_{(N-1)\times N} & & & 0_{(N-1)\times(N-1)} \end{array}\right], \quad (11)$$

$$H = \begin{bmatrix} g(X_1,\ Y_1,\ z) & \cdots & g(X_1,\ Y_{2N-1},\ z) \\ \vdots & \ddots & \vdots \\ g(X_{2N-1},\ Y_1,\ z) & \cdots & g(X_{2N-1},\ Y_{2N-1},\ z) \end{bmatrix}_{(2N-1)\times(2N-1)}, \quad (12)$$

$$X_j = \begin{cases} x_1 - \varsigma_{N+1-j} & j = 1,\ \ldots,\ N-1 \\ x_{j-N+1} - \varsigma_1 & j = N,\ \ldots,\ 2N-1 \end{cases}, \quad (13)$$

$$Y_j = \begin{cases} y_1 - \eta_{N+1-j} & j = 1,\ \ldots,\ N-1 \\ y_{j-N+1} - \eta_1 & j = N,\ \ldots,\ 2N-1 \end{cases}. \quad (14)$$

The sampling grid of $U(\varsigma_i, \eta_j, 0)$ is zero padded to $(2N - 1) \times (2N - 1)$ as shown in Eq. (11) because Eq. (10) gives the circular convolution of U and H. The result S is a $(2N - 1) \times (2N - 1)$ complex matrix. The desired light fields in the observation plane can be given by the $N \times N$ lower right submatrix of S:

$$U(x_m,\ y_n,\ z) = S_{m+N,\ n+N}. \quad (15)$$

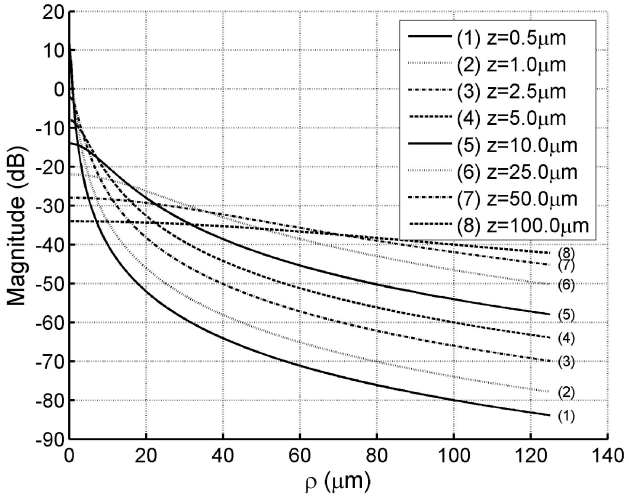

Fig. 2. Magnitude of $g(x, y, z)$ at several observation planes.

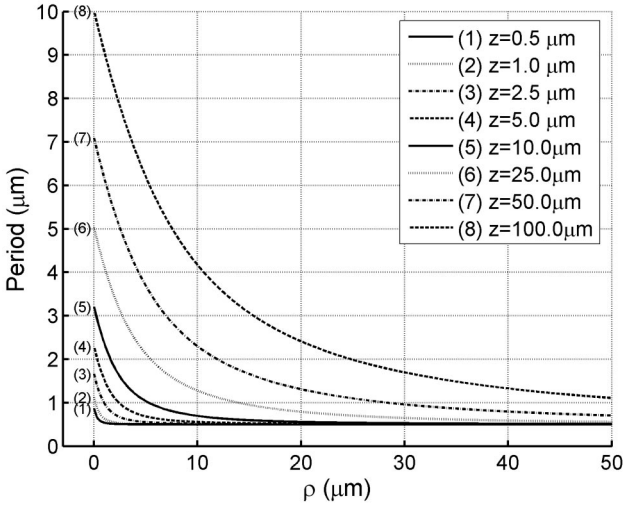

Fig. 3. Oscillating period of $g(x, y, z)$ at several observation planes.

The accuracy of the FFT-DI method depends on the sampling intervals. Higher accuracy can be obtained when smaller sampling intervals are used. The error of the numerical integration in Eq. (9) can be estimated as

$$E = O(\Delta\varsigma^2) + O(\Delta\eta^2). \tag{16}$$

One can use Simpson's rule to improve the accuracy of numerical integration by introducing weight matrix W into Eq. (11) as

$$U = \begin{bmatrix} W \cdot\times U_0 & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}_{(2N-1)\times(2N-1)}, \tag{17}$$

where

$$W = B^T B, \tag{18}$$

$$B = \tfrac{1}{3}[1\ 4\ 2\ 4\ 2\ \ldots\ 2\ 4\ 1] \tag{19}$$

for an odd N.[17] The error of Simpson's rule for 2D numerical integration can be estimated as

$$E = O(\Delta\varsigma^4) + O(\Delta\eta^4), \tag{20}$$

which is smaller than that of Eq. (16).

The FFT-DI method discussed above can also be extended to calculation of the Helmholtz–Kirchhoff integrals by use of different values of $g(x, y, z)$ in Eq. (9).

4. Discussion

A. Sampling

Both the FFT-AS and the FFT-DI methods need discrete $U(\varsigma, \eta, 0)$. If $U(\varsigma, \eta, 0)$ is band limited in the spatial-frequency domain as $A(\alpha, \beta, 0)|_{|\alpha|>\alpha_M \text{ or } |\beta|>\beta_M} = 0$, then, according to the Nyquist sampling theorem, the sampling intervals have to be less than $(2\alpha_M)^{-1}$ and $(2\beta_M)^{-1}$, respectively. However, most apertures are finite in spatial size, and their Fourier transformations are infinite in the spatial-frequency domain. Thus the frequency components greater than the sampling frequencies are truncated. When the sampling frequency is high enough, the total power of the truncated band is usually much lower than that of the low frequencies, and thus the truncation error can be neglected in most cases.

The FFT-DI method in Eq. (10) requires discrete $g(x, y, z)$. Because $g(x, y, z)$ is not band limited; the sampling of $g(x, y, z)$ is always aliased. However, $g(x, y, z)$ is of low frequencies for paraxial points and relatively long propagation distances. Thus, for a given diffraction problem, one can estimate the maximum frequency and select the appropriate sampling interval. The magnitude and the oscillation period of $g(x, y, z)$ at different observations planes are plotted in Figs. 2 and 3, respectively, with a simulation wavelength of 0.5 μm.

The attenuation of light fields with respect to the offset between the point and the optical axis at planes z = 0.5, 1, 2.5, 5, 10, 25, 50, 100 μm are plotted in Fig. 2. It is evident that more energy is diffracted to areas far from the optical axis as z increases, but the attenuation rate gets smaller. This implies that more sampling points are required for small z because of the rapid variations in the light field.

We can estimate the oscillation period of $g(x, y, z)$ by calculating the interval $\Delta\rho$ between points on the observation plane with 2π phase difference, given as

$$k\sqrt{(\rho + \Delta\rho)^2 + z^2} - k\sqrt{\rho^2 + z^2} = 2\pi, \tag{21}$$

where $\rho = \sqrt{x^2 + y^2}$ is the offset of a point from the optical axis. The solution is

$$\Delta\rho = \sqrt{\lambda^2 + \rho^2 + 2\lambda\sqrt{\rho^2 + z^2}} - \rho. \tag{22}$$

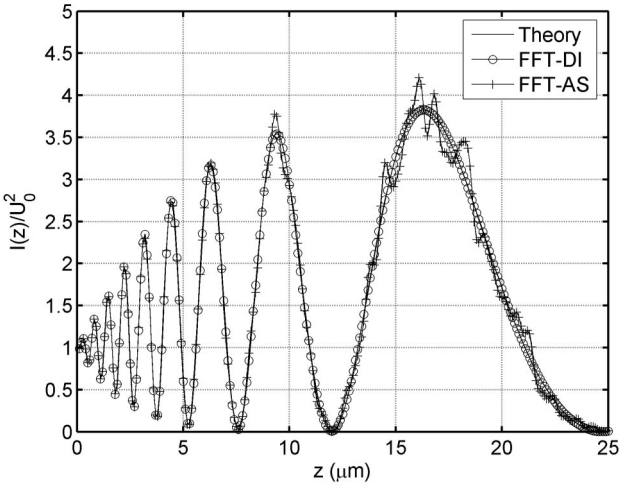

Fig. 4. Axial intensity distribution behind a circular aperture.

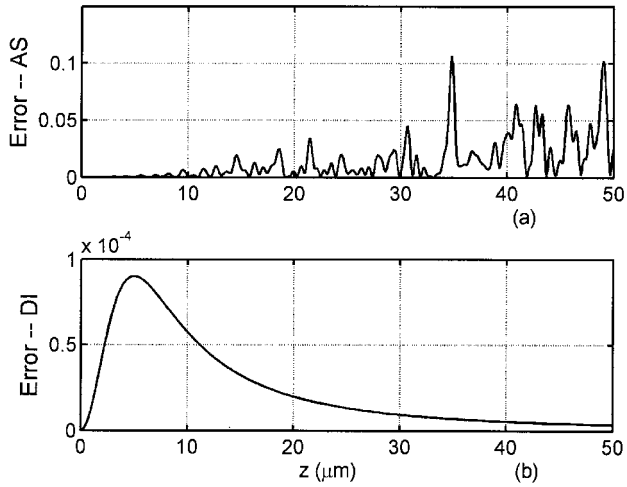

Fig. 5. Calculation errors of the FFT-AS method and the FFT-DI method: (a) FFT-AS and (b) FFT-DI methods.

The relationship between $\Delta\rho$ and ρ on different planes is shown in Fig. 3. When z is large and ρ is small, which is a paraxial case, the oscillation period of $g(x, y, z)$ is large and $g(x, y, z)$ is of low frequency. Thus a large sampling interval can be used without sacrificing calculation accuracy. However, when ρ is large or z is small, which is not the paraxial case, the oscillating period of $g(x, y, z)$ is asymptotic to λ, and the sampling intervals for $g(x, y, z)$ need to be smaller than half of the optical wavelength at least.

The minimum $\Delta\rho$ for a given z and ρ is a measure of the maximum frequency that $g(x, y, z)$ contains. Equation (22) gives a convenient rule for the selection of the sampling interval. For a light diffraction problem to be solved, the minimum propagation distance $z_{\min}$ and the maximum offset from the optical axis $\rho_{\max}$ can be substituted into Eq. (22) to yield a minimum $\Delta\rho_{\min}$. The sampling intervals on the x–y plane should be less than $\Delta\rho_{\min}/2$ to reduce the aliasing error.

B. Accuracy

The ranges of the integration in Eq. (5) and the resultant light fields are infinite. However, for a numerical calculation with the FFT-AS method, the size of the computation window is finite. The result of the IFFT in Eq. (6), $Q(x_m, y_n, z)$, is a replica of $U(x, y, z)$ in the spatial domain, which can be given as the summation of samples of $U(x, y, z)$ at all the grid points with equal distances of the computation window size:

$$Q(x_m, y_n, z) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} U(x_m + iX, y_n + jY, z), \quad (23)$$

where X and Y are dimensions of the computation window.[18]

It is evident that the calculation error of the FFT-AS method depends mainly on the light fields of grid points outside the computation window on the observation plane. Therefore the size of the computation window has to be large enough that the light fields outside the computation window can be neglected. An oversampling of the same computation window with small sampling intervals will not increase the accuracy. The accuracy of the DI method depends on the accuracy of the numerical integration in Eq. (9), which in turn depends on the sampling intervals, as shown in Eqs. (16) and (20). Small sampling intervals will yield high calculation accuracy. A simulation result for comparison of the accuracy of the FFT-AS method and the FFT-DI method is given in Subsections 5.A and 5.B below.

C. Computation Window

The rule for the selection of computation window size in the FFT-AS method is that all the light fields outside the window can be neglected for a given desired accuracy. In practice, owing to the diffraction of light, the size of the computation window has to be selected to be much larger than the aperture size, even if only a small observation window is desired. However, a large computation window will cause a large sampling array and thus a high computational load. Therefore the application of the FFT-AS method is limited to small apertures and near-field cases.

When the observation window is not on the optical axis, a large computation window centered on the optical axis has to be selected. This is not computationally efficient. Delen *et al.* tried to use a shifted Fourier transform to handle this problem with a smaller computation window by moving the center of the computation window to the optical axis.[12] However, because the mathematical shift of the computation window does not change the real light field distribution, a small computation window will result in an incorrect calculation.

Here is a general guideline that we used for selection of the computation window in the FFT-AS method. For a given propagation distance z, the magnitude of $g(x, y, z)$ for light fields outside the computation window should be much smaller than that on the optical axis, given by

$$|g(x, y, z)| < \varepsilon |g(0, 0, z)|, \quad (24)$$

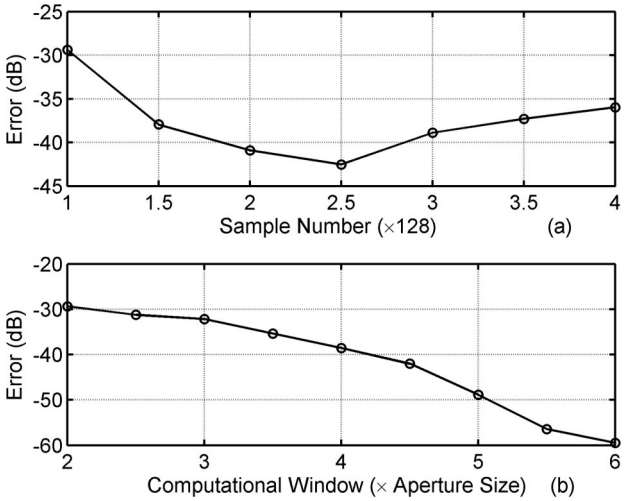

Fig. 6. Calculation errors of the FFT-AS method for the sampling numbers and computation window sizes shown.

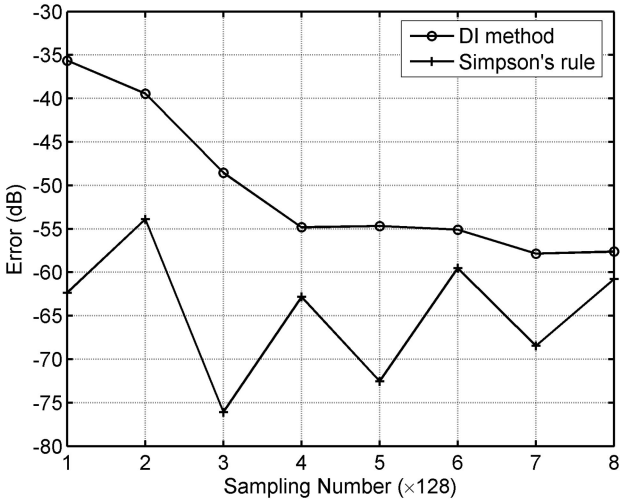

Fig. 7. Calculation errors of the FFT-DI method for several sampling numbers.

where ε is a small number. Inequality (24) can be rewritten as

$$\left|\frac{1}{r^2}\left(\frac{1}{r}-jk\right)\right|<\varepsilon\left|\frac{1}{z^2}\left(\frac{1}{z}-jk\right)\right|, \tag{25}$$

where $r = \sqrt{\rho^2 + z^2}$. In practice, z is usually much larger than several wavelengths; thus inequality (23) can be simplified to

$$\frac{1}{r^2} < \varepsilon \frac{1}{z^2}, \tag{26}$$

and the minimum ρ can thus be solved as

$$\rho > z[(1/\varepsilon) - 1]^{1/2}. \tag{27}$$

The minimum computation window can then be selected to be $(2\rho + a) \times (2\rho + b)$, where $a \times b$ is the size of the aperture.

The FFT-DI method, however, requires that the size of the computation window be the same as that of the aperture window. When the observation window is larger than the aperture, one can either enlarge the aperture window or divide the observation window into subwindows with smaller sizes. In Subsection 4.D below, we show that the latter method is more computationally effective.

D. Computational Complexity

The computational load of the FFT-AS method in Eq. (6) comes from (a) a FFT of U, (b) calculation of G, (c) element-by-element multiplication $A = \mathrm{FFT}(U) \cdot\times G$, and (d) IFFT ($A$). Assuming that the array size is $N \times N$, the computational complexity is

$$\begin{aligned} C_1 &= C_a + C_b + C_c + C_d \\ &= O(N^2 \log_2 N) + O(N^2) + O(N^2) + O(N^2 \log_2 N). \end{aligned} \tag{28}$$

When N^2 points on the observation plane are arbitrarily located, the computational complexity of the traditional DI method is $O(N^4)$ (Ref. 12) because each point has a computational complexity of $O(N^2)$, as shown in Eq. (9). The FFT-DI method calculates the light fields on grid points and uses a FFT and an IFFT to improve the calculation speed. The computational load of the FFT-DI method in Eq. (10) comes from (a) FFT(U), (b) calculation of H, (c) FFT(H), (d) the element-by-element product of FFT(U) and FFT(H), and (e) the IFFT of the product. For maximum efficiency of FFT and IFFT, U and H should be zero padded to have the size $2^m \times 2^m$, where m is an integer. Assuming that the length of the FFT is N_F, the computational complexity of the FFT-DI method is

$$\begin{aligned} C_2 &= C_a + C_b + C_c + C_d + C_e \\ &= O({N_F}^2 \log_2 N_F) + O({N_F}^2) + O({N_F}^2 \log_2 N_F) \\ &\quad + O({N_F}^2) + O({N_F}^2 \log_2 N_F), \end{aligned} \tag{29}$$

which is much lower than $O(N^4)$.

If the $N \times N$ sampling array in the FFT-DI method is zero padded to $2N \times 2N$, then the lengths of the FFTs in the FFT-DI method are the same as in the FFT-AS method when a computation window with 2×2 times the aperture size is used. From Eqs. (28) and (29), one can find that the FFT-DI method needs only one more 2D FFT than the FFT-AS method does, which means that the computational loads of the FFT-DI and FFT-AS methods are comparable.

In the FFT-DI method the size of the computation window is the same as that of the aperture window. One can divide a large observation window into smaller subwindows that have the same size as the aperture, such that the maximum computational efficiency of the FFT algorithm can be obtained. For example, if the desired observation window is 2×2

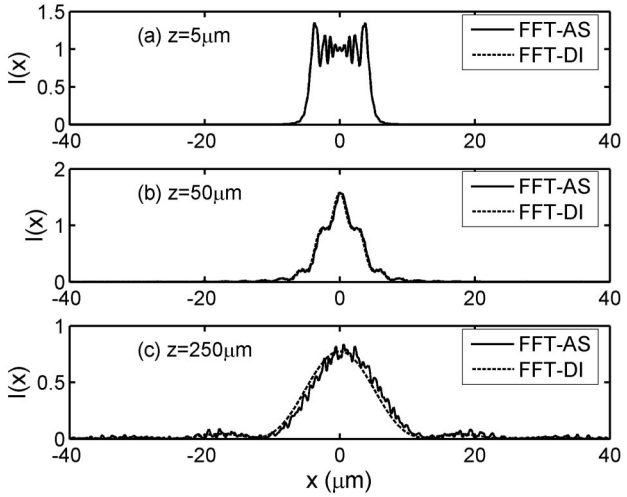

Fig. 8. Simulation results of diffraction pattern of a single slit.

times the aperture, one can divide it into four subwindows with $N \times N = (N_1/2) \times (N_1/2)$ arrays, where N_1 is the sampling number for the large computation window. The computational complexity for the FFT and the IFFT for the small windows is

$$4 \times O(N^2 \log_2 N) = O\left[N_1{}^2(\log_2 N_1 - 1)\right],$$

which is less than that for the large computation windows, whose computational complexity is $O(N_1{}^2 \log_2 N_1)$.

E. Two-Dimensional Diffraction

The impulse response of a 2D linear isotropic homogenous medium is[7,11]

$$h(x, z) = \frac{jkz}{2r} H_1{}^{(1)}(kr), \tag{30}$$

where $H_1{}^{(1)}(kr)$ is the first-order, first-kind Hankel function and $r = \sqrt{x^2 + z^2}$. Thus the propagation of light can be represented as the convolution of $U(x, 0)$ and $h(x, z)$:

$$\begin{aligned} U(x, z) &= \int_A U(\varsigma, 0) h(x - \varsigma, z) \mathrm{d}\varsigma \\ &= \int_A U(\varsigma, 0) \frac{jkz}{2r} H_1{}^{(1)}(kr) \mathrm{d}\varsigma, \end{aligned} \tag{31}$$

where $r = \sqrt{(x - \varsigma)^2 + z^2}$.

The FFT-DI method for 2D diffraction can be given as

$$S = \mathrm{IFFT}[\mathrm{FFT}(U) \cdot\times \mathrm{FFT}(H)]\Delta\varsigma, \tag{32}$$

where

$$U = [U(\varsigma_1, 0) \ \ldots \ U(\varsigma_N, 0) \ 0 \ \ldots \ 0]_{2N-1}, \tag{33}$$

$$H = \frac{jkz}{2}\left[\frac{H_1{}^{(1)}(kr_1)}{r_1} \ \cdots \ \frac{H_1{}^{(1)}(kr_{2N-1})}{r_{2N-1}}\right]_{2N-1}, \tag{34}$$

$$r_j = \begin{cases} \sqrt{(x_1 - \varsigma_{N+1-j})^2 + z^2}, & j = 1, \ \ldots, \ N-1 \\ \sqrt{(x_{j-N+1} - \varsigma_1)^2 + z^2}, & j = N, \ \ldots, \ 2N-1 \end{cases}, \tag{35}$$

and $S(N{:}2N - 1)$ gives the desired light fields. Similarly, Simpson's rule can also be applied to improve the accuracy of the calculation.

5. Simulation Results

A. Circular Aperture

It has been shown that the light fields on the optical axis behind a circular aperture under normal uniform plane-wave illumination have the following exact solution:

$$U(z) = U_0 z\left[\frac{\exp(jkz)}{z} - \frac{\exp\left(jk\sqrt{z^2 + a^2}\right)}{\sqrt{z^2 + a^2}}\right], \tag{36}$$

where U_0 is the magnitude of the incident light and a is the radius of the circular aperture.[19,20]

We used both the FFT-AS and the FFT-DI methods to calculate the light fields of axial points and compared the results with the theoretical solution in Eq. (36) to evaluate the accuracy of the two methods. The parameters used in the simulation are as follows: The wavelength of the chromatic light was $\lambda = 0.5$ μm, the radius of the circular aperture was $a = 10\lambda = 5$ μm, and the sampling interval was 0.1λ. For the FFT-AS method the computation window was set to 2×2 times the aperture window.

The simulation results of the axial intensity distribution are shown in Fig. 4. The FFT-DI method gives consistent simulation results for both small and large z. In contrast, the FFT-AS method gives consistent results only for small z. When z is increased, because more light power is diffracted outside the computation window, a larger error occurs. The square errors of the calculations, $|U - U_{\mathrm{AS}}|^2$ and $|U - U_{\mathrm{DI}}|^2$, are plotted in Fig. 5, where U is the theoretical value and U_{AS} and U_{DI} are the simulation results. The FFT-DI method has a much smaller calculation error than the FFT-AS method for a relatively large z. The error of the FFT-AS method tends to increase with z, whereas the error of the FFT-DI method shows a maximum at a certain z and is asymptotic to zero when z is large.

The errors of the FFT-AS method for several sampling intervals and computation window sizes are shown in Fig. 6. Figure 6(a) shows the errors of the FFT-AS method when the aperture is sampled from 128×128 to 512×512 with the computation window kept at the same size, which is 2×2 of the aperture window. The results show that oversampling on the aperture will not reduce the calculation error. Figure

Table 1. Comparison of Calculation Speeds

FFT-DI[a]		FFT-AS[b]		
Sampling Grid of Aperture	Execution Time (s)	Computation Window	Execution Time (s)	Array Size of FFT
128×128	0.23	2×2	0.19	256×256
256×256	0.95	4×4	0.70	512×512
512×512	3.36	8×8	2.56	1024×1024
1024×1024	13.34	16×16	10.80	2048×2048

[a]The computation window in the FFT-DI method has the same size as the aperture.
[b]The sampling grid of the aperture in FFT-AS method is 128×128.

6(b) shows the errors of the FFT-AS method when the computation window size is selected from 2×2 to 6×6 of the aperture size while the sampling numbers of the aperture are kept at 128×128. The results show that a large computation window will decrease the calculation error.

The errors of the FFT-DI method for several sampling numbers are shown in Fig. 7. The calculation error $|U - U_{\mathrm{DI}}|^2$ at $z = 8$ μm for sampling numbers from 128×128 to 1024×1024 are plotted together with the calculation errors of Simpson's rule. It can be seen that the calculation error decreases when the sampling number increases. However, when the sampling interval is small enough, the rate of decrease in error becomes saturated. Simpson's rule can greatly reduce the calculation error and thus improve the calculation accuracy.

B. Single Slit

We verified the 2D diffraction formulas in Eqs. (30)–(35) by calculating the diffraction patterns of a single infinite slit under uniform plane-wave illumination. The wavelength of the light was 0.5 μm. The width of the slit was 10 μm. The aperture plane was sampled to 1024 points along the x axis. The observation window was set to 80 μm width with the center on the optical axis. In the FFT-AS method the computation window was the same size as the observation window, which was eight times the size of the aperture. The diffraction patterns at observation planes of $z = 5, 50, 250$ μm, calculated by both the FFT-AS and the FFT-DI methods, are plotted in Fig. 8.

When z is small (5 μm), the differences between the calculation results from the FFT-AS and the FFT-DI methods are so small that they cannot be identified, as Fig. 8(a) shows. However, the error of the FFT-AS method increases when z becomes larger (50 μm), as Fig. 8(b) shows. When z is large (250 μm), the FFT-AS method fails to give an accurate result, as shown in Fig. 8(c).

C. Calculation Speed

In our simulation we used a Compaq personal computer with a Pentium 4, 1.4 GHz CPU and 512 Mbytes of RAM to investigate the speed of the FFT-AS and the FFT-DI methods. The simulation was based on the circular aperture diffraction as described in Subsection 5.A. The observation window was set to be 10 μm × 10 μm, the same size as the aperture window. The average execution time of 25 running cycles and the FFT array sizes of the FFT-DI and the FFT-AS methods are compared in Table 1. The simulation code was programmed and executed in Matlab v6.5 on a Windows XP operating system.

It is evident that the FFT-DI method can greatly reduce the computational load compared with the traditional DI method. The execution time of the FFT-DI method for an array size of 512×512 is 3.36 s , which is much less than that of the traditional DI method reported in Ref. 16. The computational loads of the FFT-DI and the FFT-AS methods are comparable when the array sizes in these two methods are the same. The execution times of the FFT-DI method with a 1024×1024 sampling array and of the FFT-AS method with a computation window of 16×16 times the aperture window are 13.34 and 10.80 s, respectively. The lengths of the FFT and the IFFT are 2048×2048 for both cases. The execution time of the FFT-DI method is only 1.24 times that of the FFT-AS method.

The FFT-DI method uses one more FFT than the FFT-AS method and thus needs one more array to save the FFT result. For example, for an array size of 1024×1024 with double precision, the FFT-AS needs 8 Mbytes of memory to save the result of one FFT, while the FFT-DI method needs 16 Mbytes of memory to save the results of the two FFTs.

6. Conclusions

Based on the investigation of the numerical calculation methods for the Rayleigh–Sommerfeld diffraction integral, a fast-Fourier-transform based DI method was implemented to lower the computational load of the numerical integration. Simpson's rule was introduced to improve the calculation accuracy. The parameter selections and the performance of the FFT-DI and the FFT-AS methods were discussed. The sampling of the light fields on the aperture plane needs to meet the requirements of the Nyquist sampling theorem in both methods. The accuracy of the FFT-AS method depends on the error caused by the finite size of the computation window. An oversampling in the aperture window will not increase the accuracy of the FFT-AS method. The accuracy of the FFT-DI method depends on the sampling intervals in the aperture plane. The size of the computation win-

dow in FFT-DI method needs to be the same as that of the aperture window. The computational load of the FFT-DI method is comparable to that of the FFT-AS method but much lower than that of the traditional DI method. Simulation results have shown that the FFT-DI method is accurate and efficient and can be used universally, whereas the FFT-AS method should be restricted to small apertures and near-field cases.

This research was supported in part by the U.S. Department of Energy under grant DE-FC36-01G011050 and by the National Science Foundation under grant CMS-0427951.

References

1. M. Born and E. Wolf, *Principles of Optics,* 6th ed. (Pergamon, 1980), Chap. 8.
2. E. Hecht, *Optics,* 2nd ed. (Addison-Wesley, 1987), Chap. 10.
3. N. Mukunda, "Consistency of Rayleigh's diffraction formulas with Kirchhoff's boundary conditions," J. Opt. Soc. Am. **52,** 336–337 (1962).
4. E. Wolf and E. W. Marchand, "Comparison of the Kirchhoff and the Rayleigh–Sommerfeld theories of diffraction at an aperture," J. Opt. Soc. Am. **54,** 587–594 (1964).
5. J. E. Harvey, "Fourier treatment of near-field scalar diffraction theory," Am. J. Phys. **47,** 974–980 (1979).
6. J. C. Heurtley, "Scalar Rayleigh–Sommerfeld and Kirchhoff diffraction integrals: a comparison of exact evaluations for axial points," J. Opt. Soc. Am. **63,** 1003–1008 (1973).
7. M. Totzeck, "Validity of the scalar Kirchhoff and Rayleigh–Sommerfeld diffraction theories in the near field of small phase objects," J. Opt. Soc. Am. A **8,** 27–32 (1991).
8. A. M. Steane and H. N. Rutt, "Diffractions in the near field and the validity of the Fresnel approximation," J. Opt. Soc. Am. A **6,** 1809–1814 (1989).
9. W. H. Southwell, "Validity of the Fresnel approximation in the near field," J. Opt. Soc. Am. **71,** 7–14 (1981).
10. E. Lalor, "Conditions for the validity of the angular spectrum of plane waves," J. Opt. Soc. Am. **58,** 1235–1237 (1968).
11. J. J. Stamnes, "Focusing of two dimensional waves," J. Opt. Soc. Am. **71,** 15–20 (1981).
12. N. Delen and B. Hooker, "Free-space beam propagation between arbitrarily oriented planes based on full diffraction theory: a fast Fourier transform approach," J. Opt. Soc. Am. A **15,** 857–867 (1998).
13. J. A. Hudson, "Fresnel–Kirchhoff diffraction in optical systems: an approximate computational algorithm," Appl. Opt. **23,** 2292–2295 (1984).
14. C. Kopp and P. Meyrueis, "Near-field Fresnel diffraction: improvement of a numerical propagator," Opt. Commun. **158,** 7–10 (1998).
15. J. Li, Z. Fan, and Y. Fu, "The FFT calculation for Fresnel diffraction and energy conservation criterion of sampling quality," in *Lasers in Material Processing and Manufacturing,* S. Deng, T. Okada, K. Behler, and X. Wang, eds., Proc. SPIE **4915,** 180–186 (2002).
16. N. Delen and B. Hooker, "Verification and comparison of a fast Fourier transform-based full diffraction method for tilted and offset planes," Appl. Opt. **40,** 3525–2531 (2001).
17. C. Pozrikidis, *Numerical Computation in Science and Engineering* (Oxford U. Press, 1998), Chap. 7.
18. W. L. Briggs and V. E. Henson, *The DFT: An Owner's Manual for the Discrete Fourier Transform* (Society for Industrial and Applied Mathematics, 1995).
19. H. Osterberg and L. W. Smith, "Closed solutions of Rayleigh's integral for axial points," J. Opt. Soc. Am. **51,** 1050–1054 (1961).
20. A. Dubra and J. A. Ferrari, "Diffracted field by an arbitrary aperture," Am. J. Phys. **67,** 87–92 (1999).