Digital simulation of scalar optical diffraction: revisiting chirp function sampling criteria and consequences

David G. Voelz[1,*] and Michael C. Roggemann[2]

[1]New Mexico State University, Klipsch School of Electrical and Computer Engineering,
Las Cruces, New Mexico 88003, USA

[2]Michigan Technological University, Department of Electrical and Computer Engineering,
Houghton, Michigan 49931, USA

*Corresponding author: davvoelz@nmsu.edu

Accurate simulation of scalar optical diffraction requires consideration of the sampling requirement for the phase chirp function that appears in the Fresnel diffraction expression. We describe three sampling regimes for FFT-based propagation approaches: ideally sampled, oversampled, and undersampled. Ideal sampling, where the chirp and its FFT both have values that match analytic chirp expressions, usually provides the most accurate results but can be difficult to realize in practical simulations. Under- or oversampling leads to a reduction in the available source plane support size, the available source bandwidth, or the available observation support size, depending on the approach and simulation scenario. We discuss three Fresnel propagation approaches: the impulse response/transfer function (angular spectrum) method, the single FFT (direct) method, and the two-step method. With illustrations and simulation examples we show the form of the sampled chirp functions and their discrete transforms, common relationships between the three methods under ideal sampling conditions, and define conditions and consequences to be considered when using nonideal sampling. The analysis is extended to describe the sampling limitations for the more exact Rayleigh–Sommerfeld diffraction solution.   © 2009 Optical Society of America

OCIS codes: 050.1755, 050.1940, 070.2580, 070.7345.

1. Introduction

Digital simulations of scalar diffraction are used for studies that range from imaging and beam propagation investigations to digital holography and diffractive element design. A common approach for these physical or wave optics simulations is to implement a form of the Fresnel diffraction formula and utilize the two-dimensional discrete Fourier transform (DFT), usually in the form of the efficient fast Fourier transform (FFT) algorithm. A critical issue for these methods is the discrete sampling of the "chirp" function that appears in the Fresnel expression [1–3]. Nonideal sampling can lead to a limiting of the avail-

able source or observation plane support size or source bandwidth. Researchers who work with these propagators [4–8] generally have a good understanding of the parameters that provide acceptable results but specific sampling criterion for the various common simulation approaches are not consolidated in the literature. In this paper we revisit chirp sampling limitations and describe under-, over-, and ideal sampling regimes in the context of convolution-based Fresnel propagation approaches. The DFT of a chirp function is discussed and example propagation results are presented that illustrate consequences of the various sampling regimes. We present sampling criteria for the Fresnel convolution, single-FFT, two-step, and Rayleigh–Sommerfeld propagation approaches. In addition to the inventory of sampling

criteria, new results include the criteria for the Fresnel two-step and Rayleigh–Sommerfeld methods and the suggestion of using a transform/windowed form of the chirp to reduce artifacts in undersampled situations.

2. Fresnel Impulse Response/Transfer Function Approach

A frequently used propagation simulation approach is based on the convolution form of the Fresnel diffraction formula. The approach provides context for our sampling discussion and is typically referred to in the literature as the "angular spectrum method" or the "spectrum method" [4,7]. Here we distinguish between two implementations: one involving the Fresnel impulse response (IR) and the other involving the transfer function (TF). The convolution form of the Fresnel expression is given by [1]

$$U_2(x,y) = U_1(x,y) \otimes h(x,y), \quad (1)$$

where x and y are the transverse spatial coordinates, U_1 is a complex representation of the field at the source plane, U_2 is the complex field at an observation plane, $\otimes$ indicates a two-dimensional convolution, and h is the IR defined as

$$h(x,y) = \frac{e^{jkz}}{j\lambda z} \exp\left(\frac{jk}{2z}(x^2 + y^2)\right), \quad (2)$$

where λ is the wavelength, $k = 2\pi/\lambda$ is the wavenumber, and z is the distance between the centers of the source and observation planes. Applying the Fourier convolution theorem to Eq. (1) yields

$$U_2(x,y) = F^{-1}\{F\{U_1(x,y)\}F\{h(x,y)\}\}, \quad (3)$$

where $F\{\cdot\}$ indicates a two-dimensional Fourier transform and $F^{-1}\{\cdot\}$ is the inverse transform. By definition, the Fourier transform of the impulse $F\{h(x,y)\}$ is known as the TF $H(f_X, f_Y)$, which for Eq. (2) is

$$H(f_X, f_Y) = e^{jkz} \exp(-j\pi\lambda z(f_X^2 + f_Y^2)). \quad (4)$$

f_X and f_Y are the spatial frequency variables in the transform domain. Using the TF notation, Eq. (3) can be written

$$U_2(x,y) = F^{-1}\{F\{U_1(x,y)\}H(f_X, f_Y)\}. \quad (5)$$

A TF propagation simulation can be implemented following Eq. (5) by taking the FFT of a uniformly sampled source field, multiplying the result pointwise by a sampled version of the TF, and then taking an inverse FFT. Alternatively, an IR simulation following Eq. (3) can be implemented where the IR is transformed before multiplication with the transform of the input field. Even though Eqs. (3) and (5) represent identical analytical operations, with

discrete sampling and transforms the two approaches can yield different results.

A characteristic of the IR or TF simulation approaches is that the modeled source and observation planes have the same linear dimension (physical side length). For other simulation approaches considered in this paper, the source and observation plane dimensions can be different. However, in all cases we presume the source and observation planes have the same number of samples. Interpolation and decimation operations can be performed "mid-simulation" to change the sample interval and number of samples but it is more common and convenient to maintain the number of samples between planes. We do not consider interpolation or decimation in this work.

A. Chirp Sampling

Both h and H, Eqs. (2) and (4), respectively, contain a complex exponential term with a phase function whose absolute value increases with the square of either the position or frequency variable. This type of function in the time or spatial domain is commonly referred to as a "chirp" function, although for simplification we also apply this term to the analogous function in the frequency domain. Sampling a chirp function can be problematic because it is not bandlimited. Consider the spatial sampling of the impulse function of Eq. (2) where only the phase of the chirp term is a function of the transverse variables. Extracting the phase function, we have

$$\phi_h(x,y) = \frac{k}{2z}(x^2 + y^2). \quad (6)$$

The sampling criterion for the two orthogonal variables can be evaluated separately so we need only look at one transverse direction as the other follows similarly. For a uniform sample interval of Δx, the criterion for the unaliased representation of the phase when it is encoded in a modulo 2π format, which is the case for a complex exponential term, can be written

$$\Delta x \left|\frac{\partial \phi_h}{\partial x}\right|_{\max} \le \pi, \quad (7)$$

which states the maximum change in the phase must be no more than π between any two adjacent samples. The slope is found to be $\partial \phi_h/\partial x = kx/z$ and since λ and z are constants for a given scenario, the maximum slope occurs when x is a maximum ($x_{\max}$). Inserting this information into Eq. (7) and solving for Δx gives the following criterion for the sample interval:

$$\Delta x \le \frac{\lambda z}{2|x_{\max}|}. \quad (8)$$

A chirp with finite extent or, in effect, a windowed chirp is used in a digital simulation. If L is the

physical side length corresponding to the source plane that is being simulated, then the sample positions in x typically range from $-L/2$ through $L/2 - \Delta x$ in steps of Δx and the number of samples is $N = L/\Delta x$. It is apparent that $|x_{\max}| = L/2$; thus substituting in Eq. (8), the sampling criterion for the x component of the IR h_X becomes

$$\Delta x \leq \frac{\lambda z}{L}. \tag{9}$$

If Eq. (9) is violated, a periodic form of the chirp phase function is created [5]. An example of this behavior is shown in Fig. 1(a), where the argument of a sampled version of h_X [Eq. (2) with $y = 0$] is compared with the analytic phase [Eq. (6) with $y = 0$]. This figure shows the case where h_X is undersampled by a factor of three ($\Delta x = 3\lambda z/L$). The sampled phase has been unwrapped where $\pm 2\pi$ radians is added to the phase when a transition of more than π occurs between adjacent samples. The phase for the sampled function follows the analytic phase over the center part of the curve but aliased values for larger x give the curve a periodic appearance. The full width of the undersampled h_X function where the phase is unaliased is given by

$$D_{\text{Undersampled}h} = \frac{\lambda z}{\Delta x}. \tag{10}$$

It is of interest to also consider the sampling of the TF of Eq. (4) since this function is applied directly in the TF approach. The extracted phase function is

$$\phi_H(f_X, f_Y) = -\pi \lambda z (f_X^2 + f_Y^2), \tag{11}$$

and the sampling criterion for one dimension is given by

$$\Delta f_X \left| \frac{\partial \phi_H}{\partial f_X} \right|_{\max} \leq \pi, \tag{12}$$

where Δf_X is the sample interval in the spatial frequency domain. Taking the derivative and rearranging parameters yields

$$\Delta f_X \leq \frac{1}{\lambda z 2 |f_{X\max}|}, \tag{13}$$

where $f_{X\max}$ is the largest value in the set of sampled frequencies.

Understanding that the sampled versions of h and H should be a DFT pair, we apply the same spatial sampling range and interval defined for the IR so, by definition, the N frequency samples for f_X range from $-1/2\Delta x$ through $1/2\Delta x - \Delta f_X$ in steps of Δf_X, where $\Delta f_X = 1/L$. Substituting $|f_{X\max}| = 1/2\Delta x$ and $\Delta f_X = 1/L$ into Eq. (13) and solving for Δx yields the following space domain criterion for the oversampled H_X function:

$$\Delta x \geq \frac{\lambda z}{L}. \tag{14}$$

When the frequency domain expression of Eq. (13) is violated, or equivalently Eq. (14) is violated, the resulting phase of the undersampled H_X function exhibits the aliased/periodic form. This is illustrated in Fig. 1(b) for the case where H_X is undersampled by a factor of two ($\Delta x = \lambda z/2L$). The full spectral width of the undersampled H_X function where the phase is unaliased is given by

$$B_{\text{Undersampled}H} = \frac{L}{\lambda z}. \tag{15}$$

Equations (9) and (14) reveal the interesting situation that the sense of the spatial sampling criterion is opposite for the sampled pair h and H: when h is undersampled then H is oversampled and vice versa. Futhermore, Eqs. (9) and (14) are only satisfied simultaneously when

$$\Delta x \equiv \frac{\lambda z}{L}. \tag{16}$$

It can be shown rigorously that when Eq. (16) is true the DFT of a sampled version of h results in frequency sample values that exactly match the analytic function H, and in a reciprocal manner, the DFT of a sampled version of H results in exact samples of the analytic h [5,9]. For this reason we refer to Eq. (16) as the "ideal" sampling condition. To help identify the sampling regimes, the criterion can be rearranged to form a sampling regime factor $\lambda z/\Delta x L$, where a
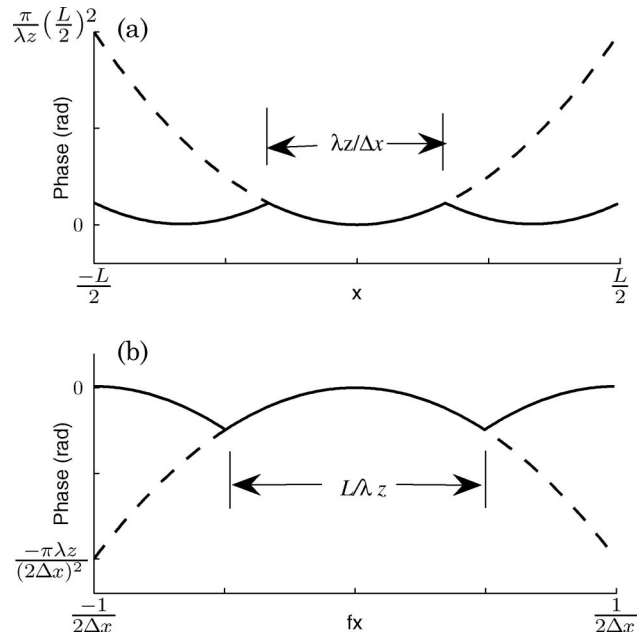

Fig. 1. Sampled phase profiles (solid) for (a) h_X undersampled by a factor of three ($\lambda z/\Delta x L = 0.33$) and (b) H_X undersampled by a factor of two ($\lambda z/\Delta x L = 2$). Analytic phase profiles (dashed) shown for comparison.

value less than 1 indicates an undersampled h (oversampled H), a value of 1 indicates ideal sampling, and a value greater than 1 indicates an oversampled h (undersampled H).

When implementing the IR or TF simulation, a discrete transform (or inverse transform) involving either h or H is computed. One might expect that the simulation result that most closely follows the analytic Fresnel result will be generated using the sampling of Eq. (16). An example is presented in Subsection 2.D that supports this expectation. However, maintaining the ideal sampling condition can be inconvenient. Consider that Eq. (16) can be rearranged to give $N = \lambda z / \Delta x^2 = L^2 / \lambda z$, which defines the ideal sampling criterion in terms of the number of samples that span the linear dimension of the array. For a given scenario, the ideal condition may dictate either too many samples for a practical FFT calculation or too few for adequate sampling of the source or observation planes. Other requirements can also be at odds with the ideal criterion. For example, the phase screens used to simulate propagation through turbulence have their own set of conditions for sample interval and array size [10,11]. Other authors have pointed out that the sampling criterion can always be maintained with the help of interpolation or decimation [12]. But extra computational steps add complexity and run time and still may not solve the practical issues of working with too many or too few samples. Thus, an important question for simulation flexibility is what are the consequences of violating the ideal criterion of Eq. (16)?

B. Discrete Fourier Transform of a Chirp Function

A DFT relationship was assumed for sampled versions of h and H and sampling regimes for these functions were identified. To investigate the consequences of violating the ideal condition for a propagation simulation, we are further concerned with the characteristics of the DFT of these sampled functions. Although the DFT of a finite chirp function is a well-studied problem [1,5,9], for reference purposes we provide the following discussion. First consider the IR function in Eq. (2). Since it is separable, we only examine one dimension:

$$h_X(x) = \frac{1}{\sqrt{j\lambda z}} \exp\left\{ j\frac{k}{2z} x^2 \right\}, \qquad (17)$$

where the constant phase term in Eq. (2) is ignored. The discrete Fourier transform can be written

$$H_X^{\mathrm{DFT}}(p\Delta f_X) = \sum_{n=-N/2}^{N/2-1} h_X(n\Delta x) \exp(-j2\pi p\Delta f_X n\Delta x)\Delta x, \qquad (18)$$

where n is an integer index corresponding to the spatial samples and p is the index corresponding to the frequency samples and both n and p range over $-N/2$

to $N/2 - 1$. As before $\Delta f_X = 1/L = 1/N\Delta x$ and the Δx multiplier at the end of Eq. (18) is included to correctly approximate the analytic transform. Equation (18) can be further manipulated but, in general, it is difficult to simplify significantly, so for illustration we evaluate Eq. (18) numerically for several cases of interest and present normalized results.

Figure 2 shows the magnitude and phase of H_X^{DFT} for three sampling regime cases. The magnitude and phase of the analytic transform $H_X(f_X) = \exp(-j\pi\lambda z f_X^2)$ are shown for comparison. All phase profiles are unwrapped for comparison purposes. Figure 2(a) shows a case where h_X is slightly undersampled ($\lambda z / \Delta x L = 0.75$). The magnitude of the discrete transform result exhibits considerable oscillation. Roughly speaking, "more" undersampling tends to generate "more" magnitude oscillations and further degradation of the phase profile.

Results for ideal sampling ($\lambda z / \Delta x L = 1$) are presented in Fig. 2(b) and both the magnitude and phase of H_X^{DFT} match the analytic result. DFT results involve a periodic extension property where taking
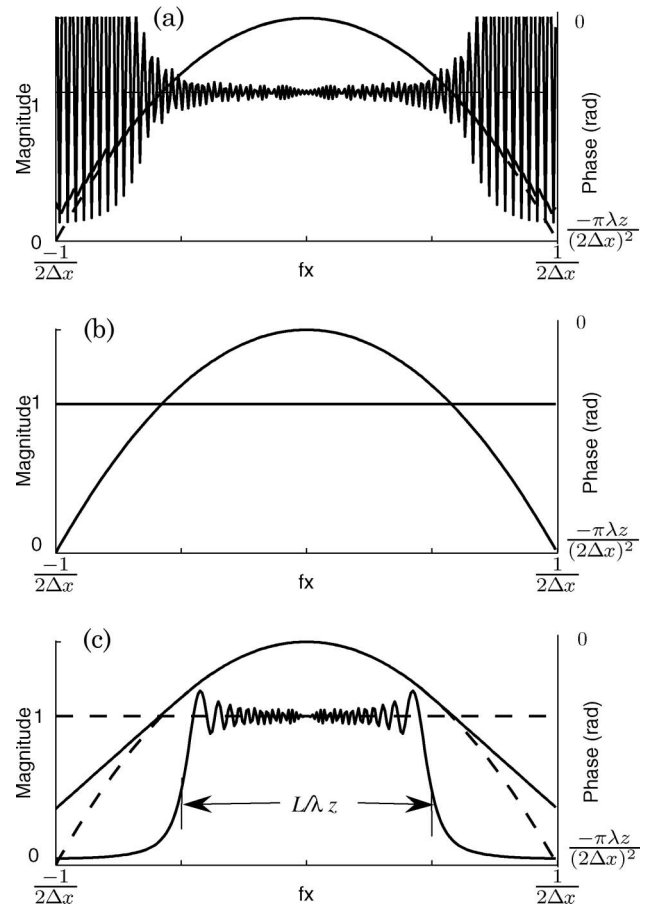

Fig. 2. Magnitude and phase profiles (solid curves) for H_X^{DFT}, the discrete Fourier transform of h_X, where h_X is (a) slightly undersampled ($\lambda z / \Delta x L = 0.75$), (b) ideally sampled ($\lambda z / \Delta x L = 1$), and (c) oversampled by a factor of 2 ($\lambda z / \Delta x L = 2$). Analytic transform profiles (dashed) shown for comparison. Phase profiles are the downward concave curves.

the transform yields a result that is consistent with the analytic transform of a periodic version of the input. We might take the view that under the ideal sampling condition the periodicity of the argument of h and H exactly matches the periodicity inherent in the DFT.

Figure 2(c) shows the case where h is oversampled by a factor of 2 ($\lambda z/\Delta xL = 2$). Here the magnitude exhibits a window-like form and the phase profile has a slope in the wings that saturates at an absolute value of π between samples. Goodman ([1], pgs. 15–18) shows that this magnitude form is expected for a chirp function that has been windowed with a rectangle function. The effective full spectral width of the window form produced by the DFT of the oversampled h_X is approximately

$$B_{\text{DFT\{Oversampled}h\}} \approx \frac{L}{\lambda z}. \qquad (19)$$

This window form represents a reduction of available bandwidth in the frequency domain, and the spectral width identified in Eq. (19) is identical to that in Eq. (15).

As may be expected, similar behavior to that illustrated in Fig. 2 is found when the analytic TF H_X is sampled and the inverse DFT is computed. When H_X is undersampled the transform magnitude exhibits excess oscillations, for ideal sampling the transform matches the analytic result, and when H_X is oversampled the magnitude window-like form appears in the transform result. In this case the window form represents a limitation in spatial extent of the transform result. The effective full width of the window form for the inverse DFT of the oversampled H_X is approximately

$$D_{\text{DFT}^{-1}\{\text{Oversampled}H\}} \approx \frac{\lambda z}{\Delta x}, \qquad (20)$$

which is the same width as indicated in Eq. (10).

C. Source Bandwidth and Size Considerations

The effects of nonideal sampling are not necessarily simple to describe as they depend on interactions between the source size, source bandwidth, array side-length, and the amount of under- or oversampling. Here we show that using either under- or oversampled chirp functions can cause artifacts in the propagation results and, depending on the approach, can limit the area (support) available in the source plane for defining the source field, the available source plane bandwidth, or the available support size in the observation plane.

A well-known criterion for wave optics simulation is that the fields at the source and observation planes should "fit" within the array dimensions. In other words, the support length of the source field D_1 and the support of the observation field D_2 should each be smaller than the side length L, otherwise nonphysical edge effects result. Further limitations

for the IR and TF approaches can be described in terms of the following conditions.

If $\lambda z/\Delta xL \leq 1$, where h is oversampled or H is undersampled, then Eqs. (19) and (15) are of interest. To avoid significant windowing effects (IR approach) or aliased phase values (TF approach) in the frequency domain, the spectral width of the source field B_1 should obey

$$B_1 \leq \frac{L}{\lambda z}. \qquad (21)$$

Here, B_1 is the spectrum full width, spanning both the negative and positive frequencies. For illustration, consider a source field propagating a long distance such that the Fraunhofer pattern appears at the observation plane. The support of the observation plane field is then given by $D_2 = B_1\lambda z$. Combining this information with Eq. (21) leads to the relationship $D_2 \leq L$, which is consistent with the criterion in the previous paragraph.

On the other hand, if $\lambda z/\Delta xL \geq 1$, where h is undersampled or H is oversampled, then because of the periodic nature of h [e.g., Fig. 1(a)], the result for the IR approach is periodic copies of a field in the observation plane separated by the distance of $\lambda z/\Delta x$. For the TF approach, which can be thought of as a convolution of the initial field U_1 with the transform of H, the observation plane result will have a valid support of approximately the initial field width D_1 plus $\lambda z/\Delta x$ [Eq. (20)]. Therefore, the width D_2 of the field in the observation plane that we expect to model accurately is limited by

$$D_2 \leq D_1 + \frac{\lambda z}{\Delta x}. \qquad (22)$$

The field in the observation plane should fit within the width $D_1 + \lambda z/\Delta x$, which is typically smaller than L. In practice, the TF approach tends to produce an observation plane field that "drops off" outside the width of $D_1 + \lambda z/\Delta x$. For illustration, consider a long propagation distance where a simple source field expands through diffraction. If we assume $\lambda z/\Delta x \gg D_1$, then Eq. (20) becomes $D_2 \leq \lambda z/\Delta x$ and the largest value that $\lambda z/\Delta x$ can take in this case is L, which corresponds to ideal sampling. This leads again to the relationship $D_2 \leq L$.

It is also instructive to consider the conditions in Eqs. (21) and (22) for ideal sampling. In this case, Eq. (21) becomes $B_1 \leq 1/\Delta x$, which is simply a statement of the Shannon–Nyquist sampling rate (where B_1 is the full bandwidth). Equation (22) becomes $D_2 \leq D_1 + L$, which is perhaps not expected since D_2 is practically limited by the side length L. This expression actually describes the maximum potential support of a convolution result and the size $D_1 + L$, in one respect, signifies that circular convolution effects are a concern even under ideal sampling conditions.

D. Simulation Example

In practical simulations, Eqs. (21) and (22) are generally not completely satisfied. Pushing the limits of these criteria leads to artifacts in the observation plane such a loss of structure or size limitations. These effects are now illustrated. Consider a problem involving a range of propagation distances where, for comparison reasons, the source/observation plane side length and sampling interval remain constant. Assume a wavelength of $\lambda = 0.5\,\mu$m and source/observation planes with a side length of $L = 0.5$ m. The sample interval is fixed at $\Delta x = 2$ mm (and $\Delta y = 2$ mm) so the number of samples along the linear dimension is $N = L/\Delta x = 250$ [13]. From Eq. (16) the propagation distance for ideal sampling is $z = L\Delta x/\lambda = 2000$ m, so results for other distances will require the ideal criterion to be violated. The source in all cases is a square aperture with a full width of 10.2 cm illuminated with a unit amplitude field [14]. The first column in Table 1 lists the propagation distances z that are considered here and the second column contains values for the sampling regime factor. The sampling regimes for h and H are listed in the last two columns for reference.

FFT-based IR and TF simulations were implemented for the parameters in Table 1. Figure 3 displays irradiance (squared-magnitude of the field) profiles of a slice through the center of the observation plane. The left-hand column panels in Fig. 3 were generated with the IR approach and the right-hand column panels were produced with the TF approach. As expected, the patterns for the ideal sampling are identical for the two approaches [Figs. 3(c) and 3(d)]. But for the closer distance of 1000 m, the irradiance pattern for the IR method exhibits copies of the fundamental diffraction pattern that arise from the convolution with the aliased-periodic phase function of the undersampled h. On the other hand, the TF method produces a single pattern. For the case where the ideal distance is doubled [Figs. 3(e) and 3(f)] both the IR and TF patterns are similar, although some "ticks" appear in the TF result. For the 20,000 m distance [Figs. 3(g) and 3(h)] the IR result looks as might be expected, although with spurious lobes at the array edges, but the TF result shows a number of periodic spike features in the profile. The spikes arise from the aliased phase structure in the undersampled H.

For a few of the cases we provide a detailed comparison of the FFT-based propagation results with results obtained from numerical integration of Eq. (1). For the integration, the source and observa-
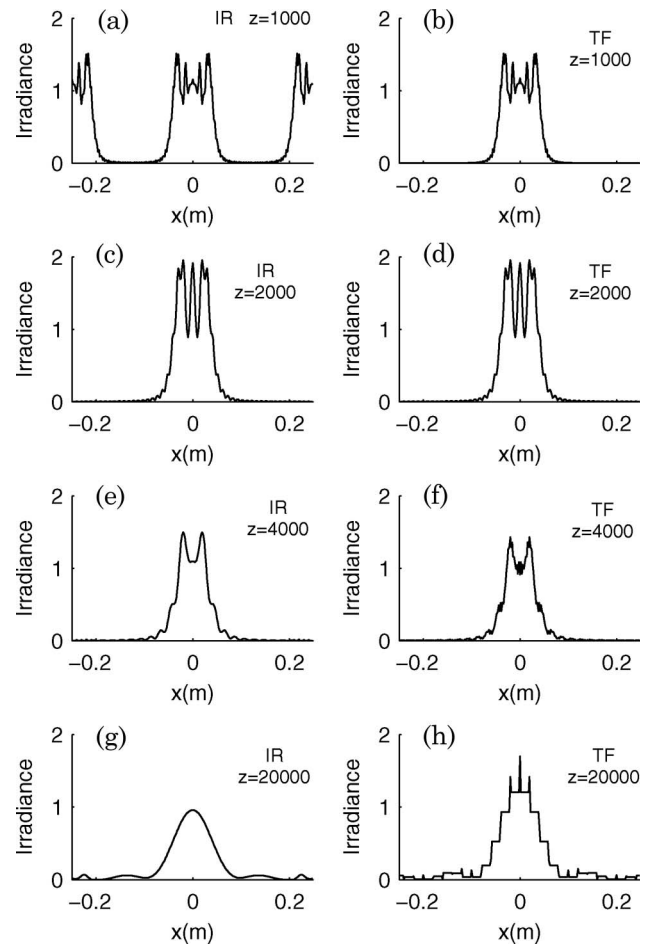

Fig. 3. Irradiance profiles for IR [(a), (c), (e), and (g)] and TF [(b), (d), (f), and (h)] simulations corresponding to cases listed in Table 2.

tion plane coordinates are defined by (x_1, y_1) and (x_2, y_2), respectively. Since the initial field for the square aperture is separable, the integration is only done along one axis. The 1D integral was implemented as

$$U_{2,X}(x_2) = \frac{1}{\sqrt{j\lambda z}} \sum_n U_{1,X}(n\Delta x_1)$$
$$\times \exp\left(\frac{jk}{2z}(n\Delta x_1 - x_2)^2\right)\Delta x_1, \quad (23)$$

where Δx_1 is the sample interval at the source, n is an index for the source samples, and the $\exp(jkz)$ multiplier of Eq. (2) is ignored. 1000 samples along the x_1 axis across the 10.2 cm aperture were used for the source and 1000 points spanning the x_2 positions -0.25 to 0.25 m in the observation plane were computed. Scaling due to the y-field component, where $y_2 = 0$, is included as $U_2(x_2, y_2 = 0) = U_{2X}(x_2)U_{2X}(0)$.

The solid curves in Fig. 4 show magnitude and unwrapped phase profiles through the center of the observed field for the simulation result of Fig. 3(b) (TF approach, $z = 1000$ m). The corresponding numerical

Table 1. Parameters and Sampling Regimes for IR and TF Simulation Example

z (m)	$\frac{\lambda z}{\Delta x L}$	Impulse Response h	Transfer Function H
1000	0.5	Undersampled	Oversampled
2000	1	Ideal Sampling	Ideal Sampling
4000	2	Oversampled	Undersampled
20000	10	Oversampled	Undersampled

$^a\Delta x = 2$ mm, $L = 0.5$ m, $N = 250$, $\lambda = 0.5\,\mu$m

integration results are shown with dashed lines. The magnitude comparison in Fig. 4(a) shows the significant features are essentially the same between the two curves but differences occur in the wings of the profiles where the integration result shows more rapid variation with position. One result of using the oversampled H chirp function in this case is the smoothing of the wings of the magnitude profile. The phase profiles [Fig. 4(b)] are nearly identical except at the edges of the array where the simulation phase slightly lags the integration phase. Equation (22) is of interest in this case. Applying the simulation parameters on the right side of the equation we get $D_2 \leq 0.352\,\text{m}$, which is roughly the apparent width of the magnitude curve before the wings drop abruptly below the analytic curve [Fig. 4(a)]. In this example, the most significant part of the field at the observation plane ($D_2 \approx 0.2$) is well within the limiting width of $0.352\,\text{m}$.

The solid curves in Fig. 5 are the magnitude and unwrapped phase for the ideal sampled case of Fig. 3(c) (IR approach, $z = 2000\,\text{m}$). Here the simulation and integration magnitude curves [Fig. 5(a)] are nearly identical. In particular, the oscillations in the wings are matched. One difference difficult to see in the figure is that the magnitude of the FFT-based simulation result is slightly elevated near the edge of the array. This is a consequence of the circular convolution that occurs when using the DFT for convolution operations. The phase profiles of Fig. 5(b) are essentially identical. This result supports our earlier conjecture that ideal sampling for the FFT-based simulation can provide a close match to the analytic Fresnel result.

The solid curves in Fig. 6 are the magnitude and unwrapped phase for the long-distance case presented in Fig. 3(h) (IR approach, $z = 20,000\,\text{m}$). The simulation/integration magnitude curves [Fig. 6(a)] are similar but the FFT simulation shows the two spurious lobes near the array edges. Considering the 1D spectrum of the source field, B_1 is proportional
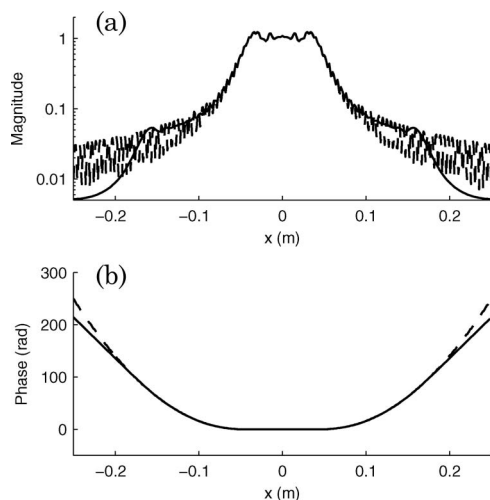
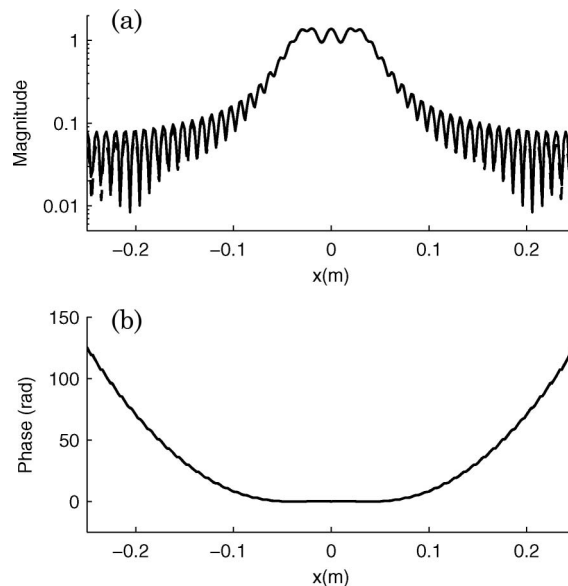

Fig. 5. (a) Field magnitude and (b) phase profiles for numerical integration result (dashed) and IR simulation approach (solid) where $z = 2000\,\text{m}$ ($\lambda z/\Delta x L = 1$, ideal sampling).

to $\text{sinc}(0.102 f_X)$, which corresponds to a central lobe full width of $19.6\,\text{cyc/m}$. For the right side of Eq. (21) we calculate $L/\lambda z = 50\,\text{cyc/m}$, which means the main lobe, the first lobes, and a portion of the second lobes of the source field spectrum satisfy Eq. (21) but higher spatial frequencies are significantly attenuated. In fact, the loss of frequency content in this case has little effect on the propagation result. The most obvious artifacts in the observation plane, the spurious lobes, appear for reasons other than the simple loss of frequency content and are more commonly explained as circular convolution effects.
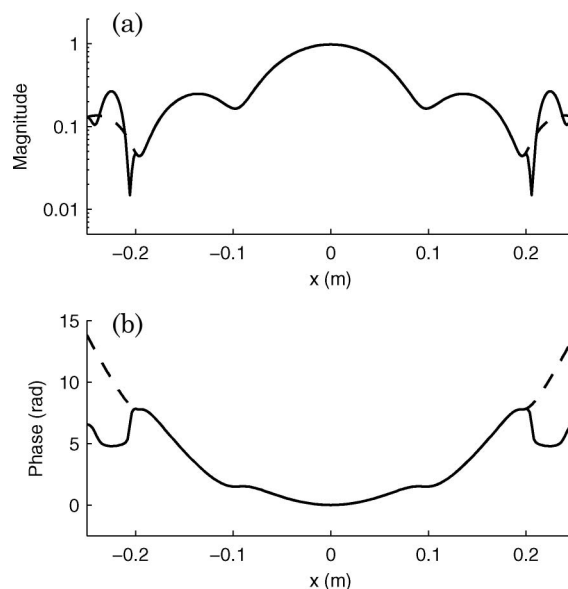

Fig. 4. (a) Field magnitude and (b) phase profiles for numerical integration result (dashed) and TF simulation approach (solid) where $z = 1000\,\text{m}$ ($\lambda z/\Delta x L = 0.5$).

Fig. 6. (a) Field magnitude and (b) phase profiles for numerical integration result (dashed) and IR simulation approach (solid) where $z = 20,000\,\text{m}$ ($\lambda z/\Delta x L = 10$).

Some observations regarding these simulation results: (a) when $\lambda z / \Delta x L < 1$, corresponding to shorter distances, the TF simulation produced results that were more representative of the analytic results; (b) when $\lambda z / \Delta x L > 1$, corresponding to longer propagation distances, the IR simulation produced more representative results; (c) under ideal sampling $\lambda z / \Delta x L = 1$ both approaches gave identical results that were a close match to the analytic result. Consequences of violating the ideal sampling criteria were periodic copies in the observation plane (IR), limiting of observation plane support (TF), smoothing of details (TF), and spiky features (TF). Spurious sidelobes, credited as circular convolution effects, were most noticeable in the long-distance IR case. We further comment that for the long-distance results, where $\lambda z / \Delta x L > 1$, the choice between a TF or IR simulation is a choice between applying an aliased chirp or a windowed form chirp in the spectral domain. In our example the windowed form avoids the problematic spiky features of Fig. 3(h) but leads to extra lobes at the array edges.

3. Fresnel Single Transform Approach

We turn our attention to a simulation approach based on factoring the integral form of Eq. (1) to give [1]

$$U_2(x_2, y_2) = \frac{e^{jkz}}{j\lambda z} \exp\left(j\frac{k}{2z}(x_2^2 + y_2^2)\right)$$
$$\times F\left\{U_1(x_1, y_1)\exp\left(j\frac{k}{2z}(x_1^2 + y_1^2)\right)\right\}. \quad (24)$$

In this expression, a chirp function multiplies the initial field U_1. The result is Fourier transformed where the spatial frequencies f_{X1} and f_{Y1} are evaluated as $x_2/\lambda z$ and $y_2/\lambda z$, and a chirp function, with other constants, multiplies the transform result. Others have referred to the implementation of Eq. (24) in a simulation as the "direct" approach [4,7].

When implementing Eq. (24) in a simulation only a single FFT is required and the side length of the observation plane L_2 can be different than the length of the source plane L_1, although a specific scaling relationship exists between the planes. Assuming in the source plane the sample interval is Δx_1 and the sample positions range from $-L_1/2$ to $L_1/2 - \Delta x_1$ and in the observation plane the sample interval is Δx_2 and the positions range from $-L_2/2$ to $L_2/2 - \Delta x_2$, then diffraction theory and FFT scaling dictate that the observation plane sample interval is related to the source plane parameters by $\Delta x_2 = \lambda z / L_1$ and the side length is $L_2 = \lambda z / \Delta x_1$.

The chirp term sampling for this approach can be derived by applying the approach presented in Subsection 2.A. For the chirp term in the source plane of Eq. (24), the following sampling criterion is obtained:

$$\Delta x_1 \leq \frac{\lambda z}{L_1}. \quad (25)$$

For the chirp term in the observation plane the sampling criterion is $\Delta x_2 \leq \lambda z / L_2$. Substituting $\Delta x_2 = \lambda z / L_1$ and $L_2 = \lambda z / \Delta x_1$ and rearranging, we find the criterion for the observation plane in terms of the source plane variables

$$\Delta x_1 \geq \frac{\lambda z}{L_1}. \quad (26)$$

Thus, the situation is the same as for the IR and TF approaches where the criteria for two chirp functions associated with the propagation are mutually exclusive except for the condition

$$\Delta x_1 \equiv \frac{\lambda z}{L_1}. \quad (27)$$

Furthermore, considering that $\Delta x_2 = \lambda z / L_1$, it is apparent that Eq. (27) is only satisfied when $\Delta x_2 = \Delta x_1$ and $L_2 = L_1$. Thus, the ideal sampling criterion requires the same relationships found for the IR and TF approaches including the fact that the source and observation planes have the same physical dimension.

The utility of this simulation approach is the requirement for only one FFT and the scaling of the observation plane with distance. However, for the planes to be of different sizes the ideal sampling criterion will be violated and either the source chirp or the observation plane chirp will be undersampled. Here we outline several conditions to be considered when using this simulation approach.

1. If $\lambda z / \Delta x_1 L_1 < 1$, the source chirp is undersampled (typically "shorter" propagation distances). Then, from Eq. (10), the requirement for the source field support to avoid the aliased portion of the source chirp phase is $D_1 \leq \lambda z / \Delta x_1$. The observation plane chirp function in this case will be oversampled.

2. If $\lambda z / \Delta x_1 L_1 > 1$, the source chirp is oversampled ("longer" propagation distances). Then, considering that the transform of Eq. (24) produces a convolution of the source field spectrum and the chirp transform with frequencies evaluated as $x_2/\lambda z$, the width D_2 of the field in the observation plane that we expect to model accurately is limited by $D_2 \leq B_1 \lambda z + L_1$. The observation plane chirp in this case is undersampled so, if the phase of the field in the observation plane is also of interest, then avoiding the aliased portion of the observation plane chirp phase requires $D_2 \leq \lambda z / \Delta x_2 = L_1$.

3. If only the irradiance at the observation plane is of interest, then the observation plane chirp can be ignored.

Drawing on insight gained from the previous IR/TF analysis, the effects of an aliased-periodic chirp in this approach can sometimes be moderated by

creating the windowed form of the offending chirp. For example, if the source plane chirp is undersampled, then one can start with its analytic transform and the FFT is taken to get the windowed source form for use in Eq. (24) [15]. Because the scaling between planes is so restrictive, simulations based on Eq. (24) are usually reserved for a few situations such as a single long-distance propagation or to model the field at the focus of a lens [4]. But the approach can be particularly useful when only the irradiance is of interest since the observation plane chirp function can be ignored.

4. Fresnel Two-Step Approach

The ability to independently select the physical side lengths of the source and observation planes in a simulation can be extremely helpful. This allows, for example, the observation plane to expand to contain a diverging beam. A two-step propagation method described by Rydberg and Bangtsson and others provides this feature [16,17]. The method is represented by the expression

$$
U_2(x_2, y_2) = \frac{z_2}{z_1} e^{jk(z_1 - z_2)} \exp\left(-j\frac{k}{2z_2}(x_2^2 + y_2^2)\right)
$$
$$
\times F^{-1}\left\{\exp\left(j\frac{k}{2}\left(\frac{1}{z_1} - \frac{1}{z_2}\right)(x_d^2 + y_d^2)\right)\right.
$$
$$
\left.\times F\left\{U_1(y_1, x_1) \exp\left(j\frac{k}{2z_1}(x_1^2 + y_1^2)\right)\right\}\right\}. (28)
$$

The approach is described by considering two propagations. The first is a propagation of distance z_1 from the source plane (with spatial variables x_1, y_1) to a dummy plane (x_d, y_d). The second propagation is a distance of z_2 from the observation plane (x_2, y_2) to the dummy plane. Equating fields at the dummy plane and rearranging terms yields Eq. (28). The actual propagation distance from the source to observation plane is defined by $z = z_1 - z_2$, and the source plane side length L_1 and the observation plane side length L_2 can be chosen independently. The source sample positions in one dimension range from $-L_1/2$ to $L_1/2 - \Delta x_1$ in steps of Δx_1, the observation sample positions are defined from $-L_2/2$ to $L_2/2 - \Delta x_2$ in steps of Δx_2, and the dummy plane samples range from $-L_d/2$ to $L_d/2 - \Delta x_d$ in steps of Δx_d. Considering the propagation scaling, we have $L_d = \lambda z_1/\Delta x_1 = \lambda z_2/\Delta x_2$ and $\Delta x_d = \lambda z_1/L_1 = \lambda z_2/L_2$. With these definitions the following can be derived:

$$
\frac{\Delta x_1}{\Delta x_2} = \frac{L_1}{L_2} = \frac{z_1}{z_2}, \tag{29}
$$

$$
z_1 = z\left(\frac{L_1}{L_1 - L_2}\right), \qquad z_2 = z\left(\frac{L_2}{L_1 - L_2}\right). \tag{30}
$$

It is acceptable for $L_2 > L_1$, which means z_1 and z_2 become negative and implies the dummy plane is located in front of the source and observation planes rather than after.

Equation (28) involves three chirp functions: one in the source plane, one in the observation plane, and one in the dummy plane. For the source chirp to be oversampled we have $\Delta x_1 \leq \lambda z_1/L_1$, and using the relation for z_1 in Eq. (30) we find

$$
\Delta x_1 \leq \frac{\lambda z}{|L_1 - L_2|}, \tag{31}
$$

where the absolute value is indicated to account for the case where $L_2 > L_1$. For the observation plane the requirement for oversampling is $\Delta x_2 \leq \lambda z_2/L_2$, which leads to $\Delta x_2 \leq \lambda z/|L_1 - L_2|$. To allow for comparison in terms of the source plane sampling, we find by applying Eq. (29)

$$
\Delta x_1 \leq \frac{L_1}{L_2}\frac{\lambda z}{|L_1 - L_2|}. \tag{32}
$$

For the dummy plane chirp, the oversampling requirement is $\Delta x_d \leq \lambda z_d/L_d$, where we define $z_d = (1/z_1 - 1/z_2)^{-1} = z_1 z_2/(z_2 - z_1)$. Applying the relationships between the dummy plane and source plane parameters, we derive the following for oversampling the dummy plane:

$$
\Delta x_1 \geq \frac{\lambda z}{L_2}. \tag{33}
$$

It is actually possible to simultaneously satisfy the oversampling requirements of Eqs. (31)–(33). However, for the ideal criterion, these equations become equalities and, considering Eqs. (31) and (32), we find the requirement $L_2 = L_1$. In this case, z_1 and z_2 become infinite and the source and observation chirp functions become unitary and are of no consequence. Only Eq. (33) is of concern and we end up with the familiar ideal sampling criterion $\Delta x_1 \equiv \lambda z/L_1$. Rydberg and Bangtsson [16] also show that the two-step method reduces to the TF (angular spectrum) approach of Eq. (5) when z_1 and z_2 become large, which is consistent with our result.

The utility of the two-step approach is that L_1 and L_2 can be different sizes, although this requires nonideal sampling. Developing sampling rules is complicated for this method as there are a number of possible combinations depending on simulation parameter choices. Here we choose to examine a common modeling situation where the propagating field is slowly expanding with distance. We assume $D_1 \leq L_1$ and $D_2 \leq L_2$ and suppose that L_2 is larger than L_1 but the difference $|L_1 - L_2|$ is relatively small. These assumptions lead to the following:

1. The source chirp is oversampled ($\Delta x_1 < \lambda z/|L_1 - L_2|$), so in the dummy plane the relationship $B_1\lambda z_1 + L_1 \leq \lambda z_d/\Delta x_d$ can be derived from convolutional considerations. With substitutions for the source and dummy plane variables and

rearranging, we find the condition $B_1 \leq L_1/\lambda z$, which is interestingly the same condition as Eq. (21) with L_1 instead of L.

2. The observation plane chirp will also tend to be oversampled ($\Delta x_2 < \lambda z/|L_1 - L_2|$) and can be ignored if only irradiance is of interest.

We present an example to illustrate the use of the two-step approach. Consider the case in Subsection 2.D, where the unit amplitude square beam of 10.2 cm width and $\lambda = 0.5\,\mu$m propagates a distance of 20,000 m [Fig. 3(g)]. As before, $N = 250$ and the source plane side length is $L_1 = 0.5$ m but now the observation plane is selected to be $L_2 = 0.8$ m. Table 2 shows the relevant parameter values for the two-step approach associated with the source, dummy, and observation planes. The sampling regime factor values in the last column of Table 1 show the source and observation planes are considerably oversampled but the dummy plane chirp is undersampled. The simulation is conducted using the source and observation chirps as presented in Eq. (28) but because of the undersampled situation the windowed form of the dummy plane chirp is employed [15].

The irradiance profile at the observation plane in Fig. 7(a) appears as expected. In fact, the larger observation plane has improved upon the result of Fig. 3(g) because the spurious sidelobes have been reduced in magnitude and are located near the edges of the larger array. The field magnitude and phase plots, Figs. 7(b) and 7(c), are compared with curves (dashed) generated with the numerical integration approach of Eq. (21), where an observation plane length of 0.8 m is used. The magnitude result shows the relatively small spurious lobes near the array edges and the phase comparison shows the FFT-based simulation again slightly lags the integration result near the array edges.

The constraint $B_1 \leq L_1/\lambda z$ derived in 1) above is identical to the constraint considered in Subsection 2.D for the long-distance IR case where the main lobe, the first lobes, and a portion of the second lobes of the source field spectrum obey the criteria. Thus, it is not surprising that the propagation result is similar to that of Fig. 3(g), although the larger observation plane clearly reduces the lobes related to circular convolution and provides a better result. It is also of interest to note that if the phase/aliased form of the chirp is used in the dummy plane rather than the windowed form then the result appears more like that of Fig. 3(h) with spiky features.

Table 2. Parameters for Two-Step Simulation Example

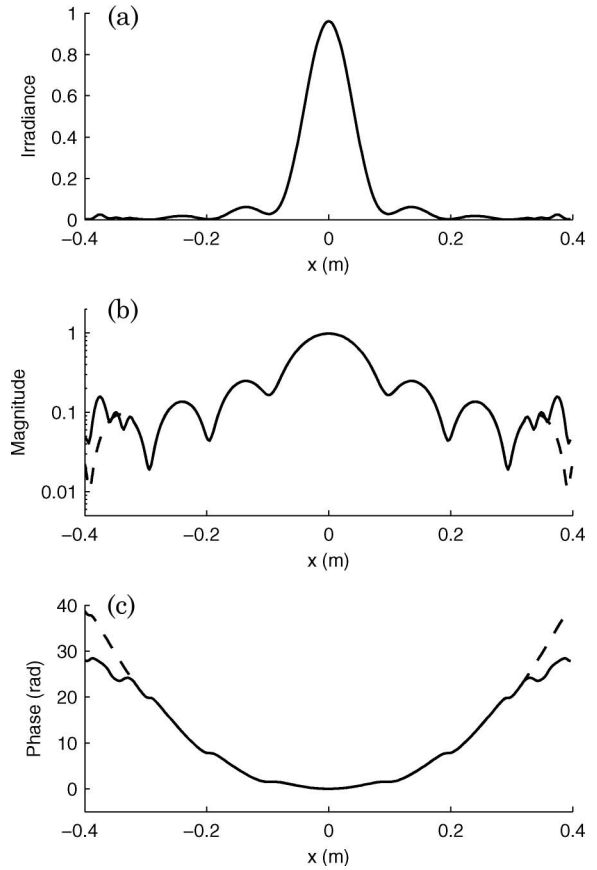| p (plane)	z_p (m)	L_p (m)	Δx_p (mm)	$\left| \frac{\lambda z_p}{\Delta x_p L_p} \right|$
1	-3.33×10^4	0.5	2.0	16.7
d	-8.89×10^4	8.33	33.3	0.16
2	-5.33×10^4	0.8	3.2	10.4

$^a N = 250$, $\lambda = 0.5\,\mu$m

Fig. 7. (a) Irradiance, (b) field magnitude, and (c) phase profiles for numerical integration result (dashed) and two-step simulation approach (solid) where $z = 20,000$ m and $L_1 = 0.5$ m, $L_2 = 0.8$ m.

5. Rayleigh–Sommerfeld Solution

The Fresnel diffraction solution, which is the basis for the approaches discussed to this point, is consistent with the paraxial approximation (small ray angles). However, when this approximation is not valid, application of the more exact Rayleigh–Sommerfeld solution is preferable. The first Rayleigh–Sommerfeld solution for planar source and observation areas is also a convolution described by Eq. (1) but with the following IR [1]:

$$h(x,y) = \frac{z}{j\lambda} \frac{\exp(jk\sqrt{z^2 + x^2 + y^2})}{z^2 + x^2 + y^2}. \quad (34)$$

The associated TF, usually found through an angular spectrum derivation, is given by

$$H(f_X, f_Y) = \exp(j2\pi \frac{z}{\lambda}\sqrt{1 - (\lambda f_X)^2 - (\lambda f_Y)^2}), \quad (35)$$

where $\sqrt{f_X^2 + f_Y^2} < 1/\lambda$ for propagating waves. Following the approach of Subsection 2.A, the sampling requirement for the IR is found to be $\Delta x \leq \lambda(z^2 + (L/2)^2)^{1/2}/L$, and the requirement for the TF is the same but with the opposite sense in the inequality. Thus, the ideal sampling criterion is given by

$$\Delta x \equiv \frac{\lambda (z^2 + (L/2)^2)^{1/2}}{L}. \qquad (36)$$

Under the paraxial approximation, $z \gg L$, Eq. (36) reduces to Eq. (16) as expected. Onural has shown that the undersampled phase of the Rayleigh–Sommerfeld solution has an aliased-periodic nature that is similar to that of the undersampled Fresnel chirp but with a dispersion term in the periodic copies [18]. We find, in practice, the behavior of propagation results for a Rayleigh–Sommerfeld/FFT simulation is essentially the same as described in Section 2 for the Fresnel convolution approach, subject to the ideal sampling criterion given in Eq. (36).

6. Summary

Nonideal sampling of the chirp functions in an FFT-based propagation simulation leads to a reduction in the available source plane support size, source bandwidth, or observation support size, depending on the approach used and which chirp is under- or oversampled. Sampling regimes and criteria were identified in this paper for several popular propagation approaches. When the criteria are not satisfied, the field at the observation plane can contain artifacts that include periodic copies of the pattern, smoothing of oscillating features, or spiky features. Spurious lobes in the observation plane, described in terms of circular convolution, were noted for the long-distance propagations with the IR approach. We found that effects, such as spiky features resulting from an undersampled chirp function, can sometimes be moderated by applying a windowed chirp form.

Ideal sampling allows full use of the source and observation plane support and bandwidth allocation. However, it is not always suitable for practical simulations and it restricts the source and observation planes to the same side length.

Our discussion centered on chirp-related sampling issues, although we noted that certain observation plane artifacts, such as extra lobes, are commonly discussed from the viewpoint of circular convolution. Relationships between sampling and circular convolution were not explored in detail but this consideration is important. For example, consider the propagation simulation presented in Subsection 2.D. To take advantage of ideal sampling, one could implement 10 sequential propagations of 2000 m under ideal sampling conditions rather than a single propagation over a distance of 20,000 m. But in fact the compounding of circular convolution effects for the 10 propagations produces a poor result with spiky features that is similar to Fig. 3(h).

References and Notes

1. J. W. Goodman, *Introduction to Fourier Optics*, 3rd ed. (Roberts and Company, 2005).
2. A. Papoulis, "Pulse compression, fiber communications and diffraction, a unified approach," J. Opt. Soc. Am. A **11**, 3–13 (1994).
3. F. Gori, "Fresnel transform and sampling theorem," Opt. Commun. **39**, 293–297 (1981).
4. D. Mas, J. Garcia, C. Ferreira, L. M. Bernardo, and F. Marinho, "Fast algorithms for free-space diffraction patterns calculations," Opt. Commun. **164**, 233–245 (1999).
5. L. Onural, "Sampling of the diffraction field," Appl. Opt. **39**, 5929–5935 (2000).
6. R. Rao, "Statistics of the fractal structure and phase singularity of a plane light wave propagation in atmospheric turbulence," Appl. Opt. **47**, 269–276 (2008).
7. D. P. Kelly and B. M. Hennelly, "Analytical and numerical analysis of linear optical systems," Opt. Eng. **45**, 088201 (2006).
8. A. Stern and B. Javidi, "Analysis of practical sampling and reconstruction from Fresnel fields," Opt. Eng. **43**, 239–250 (2004).
9. A. Brodzik, "On the Fourier transform of finite chirps," IEEE Signal Process. Lett. **13**, 541–544 (2006).
10. S. M. Flatte and J. S. Gerber, "Irradiance-variance behavior by numerical simulation for plane-wave and spherical-wave optical propagation through strong turbulence," J. Opt. Soc. Am. A **17**, 1092–1097 (2000).
11. A. Belmonte, "Feasibilty study for the simulation of beam propagation: consideration of coherent lidar performance," Appl. Opt. **39**, 5426–5445 (2000).
12. B. M. Hennelly and J. T. Sheridan, "Generalizing, optimizing, and inventing numerical algorithms for the fractional Fourier, Fresnel, and linear canonical transforms," J. Opt. Soc. Am. A **22**, 917–927 (2005).
13. Computation time for the FFT algorithm is minimized when N is a power of 2, although depending on the value, other lengths can be computed nearly as fast. The value of $N = 250$ was chosen here to simplify our illustration.
14. The size is chosen so an odd number of samples spans the aperture and one sample can be placed on the optical axis. In this case 51 samples × 0.2 cm = 10.2 cm width.
15. The analytic Fourier transform of the chirp $q(x,y) = \exp(j(k/2z)(x^2 + y^2))$ is $Q(f_X, f_Y) = j\lambda z \exp(-j\pi\lambda(f_X^2 + f_Y^2))$. In starting with a sampled version of Q, the inverse FFT of Q typically needs to be multiplied by $N^2 \Delta f_X^2$ to reproduce the correct magnitude for the sampled version of q (assumes $1/N^2$ scaling on inverse 2D FFT).
16. C. Rydberg and J. Bengtsson, "Efficient numerical representation of the optical field for the propagation of partially coherent radiation with a specified spatial and temporal coherence function," J. Opt. Soc. Am. A **23**, 1616–1625 (2006).
17. X. Deng, B. Bihari, J. Gan, F. Zhao, and R. T. Chen, "Fast algorithm for chirp transforms with zooming-in ability and its applications," J. Opt. Soc. Am. A **17**, 762–771 (2000).
18. L. Onural, "Exact analysis of the effects of sampling of the scalar diffraction field," J. Opt. Soc. Am. A **24**, 359–367 (2007).