Chapter 5
Propagation Simulation

Now we look at several implementations of the diffraction expressions of Chapter 4 to simulate optical propagation. Although the material is presented as a teaching exercise, these propagation methods are used extensively in research and industry for modeling laser beam propagation. The concentration is on methods that use the FFT and only monochromatic light will be considered here. When designing a simulation there are a variety of issues related to discrete sampling that need to be considered. We will get to that, but let's first get our feet wet with some programming.

5.1 Fresnel Transfer Function (TF) Approach

The Fresnel diffraction expression is often the approach of choice for simulations since it applies to a wide range of propagation scenarios and is relatively straightforward to compute. A common propagation routine is based on (4-18), which is repeated here for reference,

$$U_2(x,y) = \Im^{-1}\{\Im\{U_1(x,y)\}H(f_X, f_Y)\}, \tag{5-1}$$

and uses the transfer function H given by

$$H(f_X, f_Y) = e^{jkz}\exp\left[-j\pi\lambda z\left(f_X^2 + f_Y^2\right)\right]. \tag{5-2}$$

Start a new **M-file** and save it with name "**propTF**". Enter the following function:

```
1  function[u2]=propTF(u1,L1,lambda,z);
2  % propagation - transfer function approach
3  % assumes same x and y side lengths and
4  % uniform sampling
5  % u1 - source plane field
6  % L1 - source plane side length
7  % lambda - wavelength
8  % z - propagation distance
9  % u2 - observation plane field
10 %
```

```
11 [M,N]=size(u1);           %get input field array size
12 dx1=L1/M;                 %sample interval
13 k=2*pi/lambda;            %wavenumber
14 %
15 fx1=-1/(2*dx1):1/L1:1/(2*dx1)-1/L1; %freq coords
16 [FX1,FY1]=meshgrid(fx1,fx1);
17 %
18 H=exp(-j*pi*lambda*z*(FX1.^2+FY1.^2));  %trans func
19 H=fftshift(H);            %shift trans func
20 U1=fft2(fftshift(u1));    %shift, fft src field
21 U2=H.*U1;                 %multiply
22 u2=ifftshift(ifft2(U2));  %inv fft, center obs field
23 end
```

This propagator function takes the source field `u1` and produces the observation field `u2` where the source and observation side lengths and sample coordinates are identical. Here are a few remarks on the routine with associated line numbers:

[11] The `size` function finds the sample dimensions for the input field matrix `u1` (only `M` is used). This helps reduce the number of parameters passed to the **propTF** function.

[16] A line is saved by using `fx1` twice in the `meshgrid` command since `fy1` would be the same.

[18] The transfer function `H` of (5-2) is programmed although the exp(*jkz*) term is ignored. This term doesn't affect the transverse spatial structure of the observation plane result.

[19] `H` is created in the array center but is shifted (`fftshift`) before the FFT.

[20] Similarly, the source field `u1` is assumed to be in the array center so `fftshift` is applied before the 2-D FFT is computed.

[21] `U1` is multiplied by the transfer function `H` and the inverse FFT is computed to complete the convolution.

[22] Finally, `ifftshift` centers `u2` for display.

Note that lower case `u` is used for the spatial field and upper case `U` for Fourier domain quantities, which is not consistent with the use of upper case for the analytic spatial fields, for example in (5-1). But what can you do? Both are established notations, so we live with a little notational mixing.

Before putting this function to use, let's look at another method to simulate propagation.

5.2 Fresnel Impulse Response (IR) Approach

Another propagation approach can be devised based on (4-17), which is repeated here

$$U_2(x,y) = \Im^{-1}\{\Im\{U_1(x,y)\}\Im\{h(x,y)\}\}. \tag{5-3}$$

The impulse response *h* is given by

$$h(x,y) = \frac{e^{jkz}}{j\lambda z}\exp\left[\frac{jk}{2z}\left(x^2+y^2\right)\right]. \tag{5-4}$$

Even though (5-1) and (5-3) represent identical analytical operations, with discrete sampling and transforms the transfer function and impulse response approaches can yield different results. Again, the source and observation planes with the impulse approach have the same side length.

For making the impulse response propagator, some typing can be saved by starting with a copy of **propTF**. For example, with **propTF** in the Editor, click on **File** on the MATLAB® desktop window, select **Save As**, edit the file name to be **propIR**, and click **Save**. The new copy is now ready to be edited (carefully) to match the following:

```
function[u2]=propIR(u1,L1,lambda,z);
% propagation - impulse response approach
% assumes same x and y side lengths and
% uniform sampling
% u1 - source plane field
% L1 - source plane side length
% lambda - wavelength
% z - propagation distance
% u2 - observation plane field
%
[M,N]=size(u1);           %get input field array size
dx1=L1/M;                 %sample interval
k=2*pi/lambda;            %wavenumber
%
x=-L1/2:dx1:L1/2-dx1;   %spatial coords
[X,Y]=meshgrid(x,x);
%
h=1/(j*lambda*z)*exp(j*k/(2*z)*(X.^2+Y.^2)); %impulse
H=fft2(fftshift(h))*dx1^2; %create trans func
U1=fft2(fftshift(u1));    %shift, fft src field
U2=H.*U1;                 %multiply
u2=ifftshift(ifft2(U2));  %inv fft, center obs field
end
```

This code is nearly identical to the **propTF** function. Again, the source and observation planes in this approach have the same side length. Some specific remarks for this code:

[18] `h` is implemented and the FFT is applied to get `H`.

[19] Note the multiplier `dx1^2` for `H`. The FFT of `u1` and FFT^{-1} of `U2` take care of each other's scaling, but the FFT of `h` needs its own scaling.

The IR approach is more limited in terms of the situations where it should be used than the TF approach. But this approach provides a way to simulate propagation over longer distances and is useful for the discussion of simulation limitations and artifacts.

5.3 Square Beam Example

It is time to try out the TF and IR propagators. Consider a source plane with dimensions 0.5 m x 0.5 m (L_1 = 0.5 m). Start a new M-file and use the name **sqr_beam**. Enter the following:

```
% sqr_beam propagation example
%
L1=0.5;               %side length
M=250;                %number of samples
dx1=L1/M;             %src sample interval
x1=-L1/2:dx1:L1/2-dx1; %src coords
y1=x1;
```

The subscript `1` indicates source plane quantities. The code sets up 250 samples across the linear dimension of the source plane and the sample interval `dx1` works out to be 2×10^{-3} m (2 mm).

Assume a square aperture with a half-width of 0.051 m (51 mm) illuminated by a unit-amplitude plane wave from the backside where λ = 0.5 μm. The simulation therefore places 51 samples across the aperture, which provides a good representation of the square opening (Sec. 2.1). Add the following code

```
lambda=0.5*10^-6;     %wavelength
k=2*pi/lambda;        %wavenumber
w=0.051;              %source half width (m)
z=2000;               %propagation dist (m)

[X1,Y1]=meshgrid(x1,y1);
u1=rect(X1/(2*w)).*rect(Y1/(2*w)); %src field
I1=abs(u1.^2);        %src irradiance
%
figure(1)
imagesc(x1,y1,I1);
axis square; set(gca,'YDir','normal')
colormap('gray'); xlabel('x (m)'); ylabel('y (m)');
title('z= 0 m');
```

The propagation distance z is 2000 m and checking the Fresnel number for this case we find $N_F = w^2/\lambda z$ = 2.6, which is reasonable for applying the Fresnel

expression. The source field is defined in the array `u1`. The irradiance is found by squaring the absolute value of the field. Executing this script generates Fig. 5.1, which verifies the source plane arrangement.

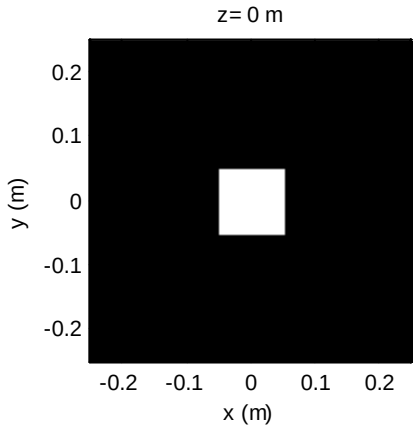

Figure 5.1 Source plane irradiance for **sqr_beam** propagation simulation.

The next part of the script is where the propagation takes place. Add the following:

```
22 u2=propTF(u1,L1,lambda,z); %propagation
23
24 x2=x1;                 %obs coords
25 y2=y1;
26 I2=abs(u2.^2);         %obs irrad
27
28 figure(2)              %display obs irrad
29 imagesc(x2,y2,I2);
30 axis square; set(gca,'YDir','normal')
31 colormap('gray'); xlabel('x (m)'); ylabel('y (m)');
32 title(['z= ',num2str(z),' m']);
33 %
34 figure(3)              %irradiance profile
35 plot(x2,I2(M/2+1,:));
36 xlabel('x (m)'); ylabel('Irradiance');
37 title(['z= ',num2str(z),' m']);
38 %
39 figure(4)              %plot obs field mag
40 plot(x2,abs(u2(M/2+1,:)));
41 xlabel('x (m)'); ylabel('Magnitude');
42 title(['z= ',num2str(z),' m']);
43 %
44 figure(5)              %plot obs field phase
45 plot(x2,unwrap(angle(u2(M/2+1,:))));
46 xlabel('x (m)'); ylabel('Phase (rad)');
```

```
47  title(['z= ',num2str(z),' m']);
```

Observation plane parameters are indicated with a "2″. In lines 24 and 25, the observation coordinates are defined to be the same as the source coordinates. Lines 28 through 47 contain code to display an image of the observation plane irradiance as well as irradiance and field magnitude and phase profiles. The `num2str` function (line 47) is introduced to display the propagation distance in the plot title. The use of the `unwrap` function for the phase profile display (line 45) is discussed in the next section.

Execute **sqr_beam**. The irradiance results are shown in Fig. 5.2 where the constructive and destructive interference of the coherent light, brought on by diffraction, results in peaks that have greater irradiance than the initial source value. These peaks are offset by valleys with less irradiance. Also apparent in the

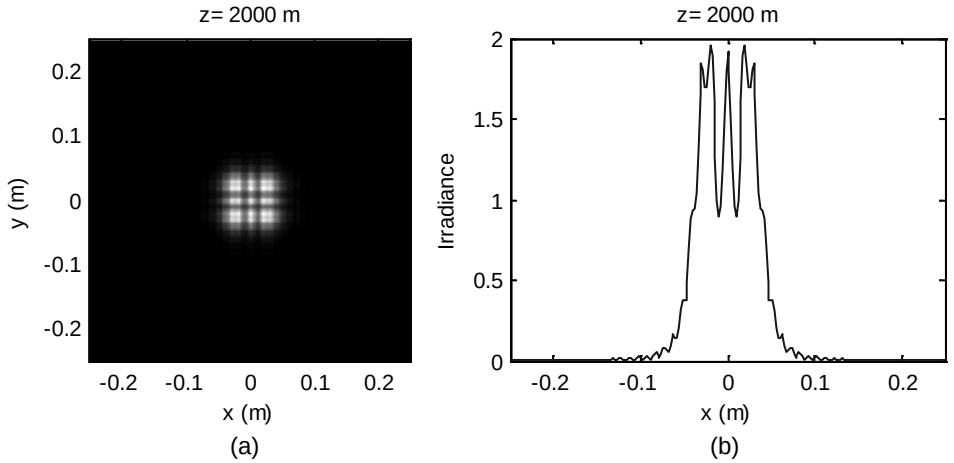

Figure 5.2 Observation plane irradiance (a) pattern and (b) profile.

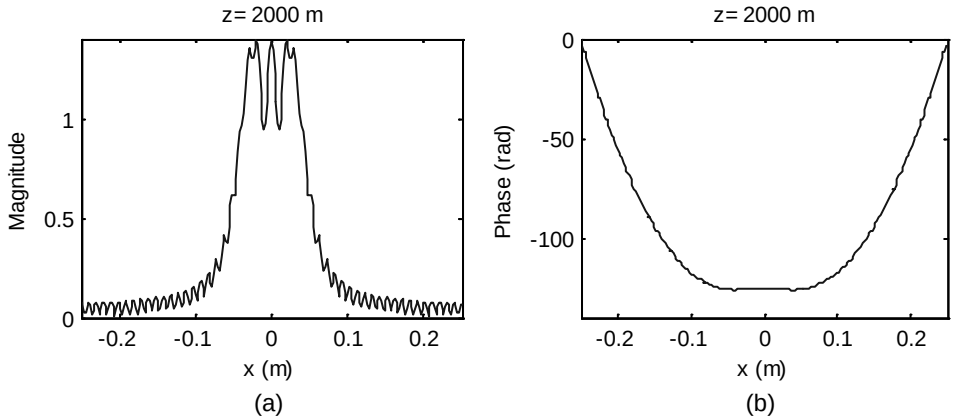

Figure 5.3 Observation plane (a) magnitude and b) phase profiles.

profile plot is a spread of the light into "wings" that exceeds the initial width of the source rectangle.

The field magnitude and phase are shown in Fig. 5.3. The phase is in units of radians. The function `unwrap`, introduced for displaying the phase, adds multiples of $\pm 2\pi$ when jumps between consecutive samples are greater than π. This serves to remove the discontinuous skips that appear in the modulo-2π formatted phase function and helps the underlying form of the phase to be seen. Remove the `unwrap` command to see what the phase looks like with modulo-2π jumps.

In words, Fig. 5.3(b) is representative of a slice through the surface of constant phase for the optical wave at the observation plane. But the important physical interpretation is that it represents the *shape of the optical wavefront* at the observation plane. Recall that the field temporal phasor, defined as $\exp(-j2\pi\nu t)$ in (4-6), is such that as time increases the phase becomes more negative. Therefore, the wavefront profile in Fig. 5.3(b), which is bowl-shaped with a flattened center, can be thought of as progressing "downward" as time carries on. Furthermore, imagine rays projecting normal from the wavefront surface to get an idea of where the energy along the wavefront is headed. The magnitude plot in Fig. 5.3(a) shows that most of the field magnitude is concentrated near the flattened center of the wavefront and the phase indicates much of that energy is still headed normal to the observation plane.

Now try the impulse response IR propagator. Change line 22 to

```
u2=propIR(u1,L1,lambda,z); %propagation
```

Run **sqr_beam**. The results in this case should be identical to those in Figs. 5.2 and 5.3.

5.4 Fresnel Propagation Sampling

5.4.1 Square beam results and artifacts

Now for the bad news. Discrete sampling of the source field, discrete sampling of the transfer function or impulse response and the periodic nature of the FFT can lead to a variety of artifacts in the propagation result. Much of the trouble comes because the chirp functions on the right side of (5-2) and (5-4) are not bandlimited and cannot be adequately sampled. This issue is introduced here with some example results.

In Fig. 5.4 both **propTF** and **propIR** results are shown for the **sqr_beam** routine at propagation distances ranging from 1000 to 20,000 m. At the distance of 1000 m the TF result appears reasonable with some constructive/destructive interference features and slight spreading beyond the initial aperture width. On the other hand, the IR result exhibits periodic copies of the pattern. At 2000 m the TF and IR results are essentially identical. At 4000 m the TF result displays some "ticks" in the profile that are absent from the IR result. For 20,000 m the

TF profile has a "stair-step" appearance whereas the IR result is smooth. At longer distances the irradiance pattern is predicted by Fraunhofer theory to take on a sinc^2 form. This appears to be the case for Fig. 5.4(h). Corresponding irradiance patterns are shown in Fig. 5.5.

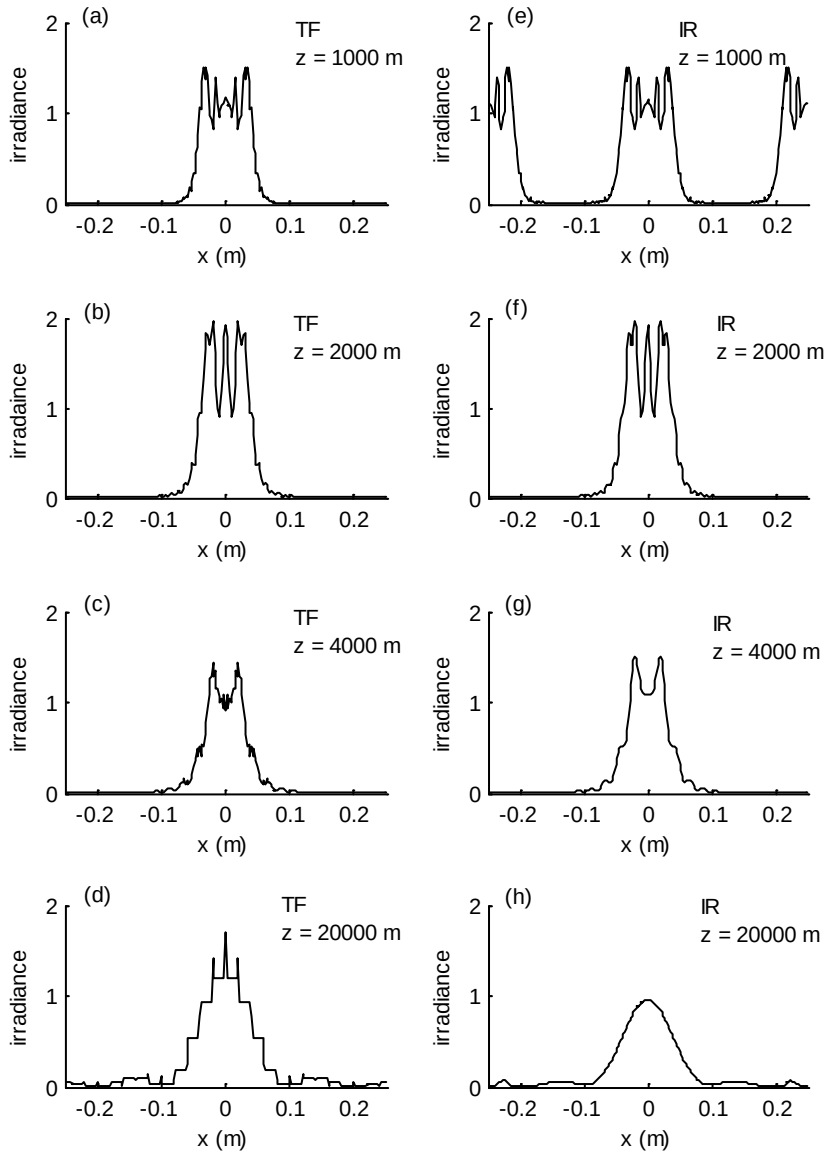

Figure 5.4 Irradiance profiles (x-axis) for the transfer function (TF) and impulse response (IR) propagation approaches for propagation distances ranging from 1000 to 20000 m. "Spiky" and "stair-step" artifacts appear in (c) and (d); periodic copies appear in (e) and spurious sidelobes appear in (h).

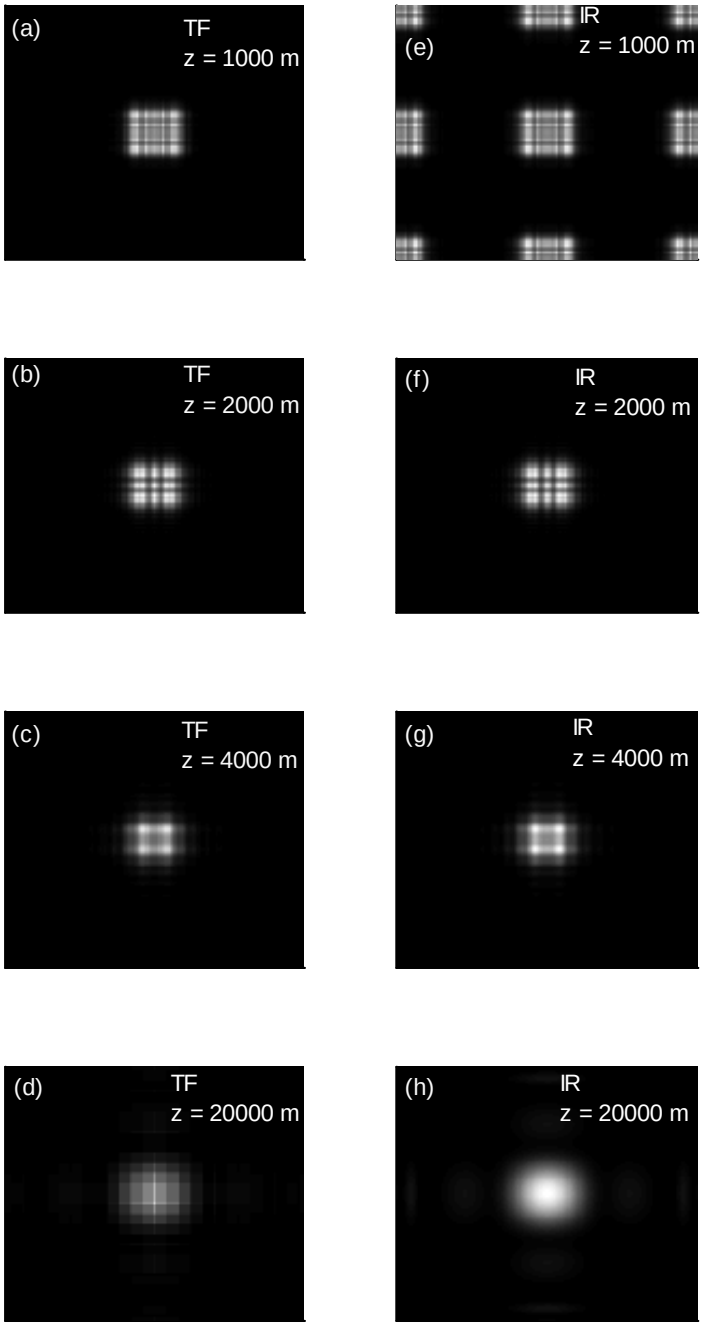

Figure 5.5 Irradiance patterns for the transfer function (TF) and impulse response (IR) propagation approaches corresponding to the profiles in Fig. 5.4.

5.4.2 Sampling regimes and criteria

Several of the panels of Fig. 5.4 display artifacts related to sampling. In this section we examine criteria used to predict when there will be problems. For more details on these and other criteria see Appendix A, references 1 and 2, and the references therein.

An obvious first criterion is that the support of the source field should "fit" within the numerical array. If D_1 is the effective support of the source field (maximum linear width) and L is the array side length then

$$D_1 < L \tag{5-5}$$

is desirable. It is good practice to provide a "guard area" around the source function, for example, $L \approx 2D_1$ or $3D_1$. This helps reduce artifacts at the edges of the array after propagation due to the periodic extension properties of the FFT.

Further criteria are derived by considering the effects of sampling the chirp functions in the Fresnel transfer function H and impulse response h expressions (Appendix A). The H chirp, given as $\exp[-j\pi\lambda z(f_X{}^2 + f_Y{}^2)]$, is adequately sampled (oversampled) when

$$\Delta x \geq \frac{\lambda z}{L}. \tag{5-6}$$

This relationship is derived by considering aliasing of the chirp function in the frequency domain and then converting to the sampled space domain (Appendix A). The h chirp, given as $\exp[jk(2z)^{-1}(x^2 + y^2)]$, is oversampled when

$$\Delta x \leq \frac{\lambda z}{L}. \tag{5-7}$$

Expressions (5-6) and (5-7) reveal the oversampling criteria for the transform pair H and h are opposite. Oversampling is a good thing in general. If (5-6) or (5-7) are violated, an aliased representation of the phase of these respective chirp functions is created that leads to simulation artifacts. Both conditions are only satisfied when

$$\Delta x = \frac{\lambda z}{L}, \tag{5-8}$$

which we call the *critical sampling* condition. More about this in a moment.

Table 5.1 lists the three regimes defined by over-, under-, or critical sampling of the phase term in the H or the h functions. For each regime, a criterion is

TABLE 5.1 PROPAGATOR SAMPLING REGIMES AND CRITERIA

Regime	Chirp Phase Sampling	Source Bandwidth Criterion	Approach	Comments
$\Delta x > \frac{\lambda z}{L}$	TF: Over IR: Under	$B_1 \le \frac{1}{2\Delta x}$	TF: Preferred IR: Periodic copies	Relatively "short" z or small λ. Observation plane limitation. TF: Observation plane field limited to full width $D_1 + \lambda z / \Delta x$
$\Delta x = \frac{\lambda z}{L}$	TF: Critical IR: Critical	$B_1 \le \frac{1}{2\Delta x}$ or $B_1 \le \frac{L}{2\lambda z}$	TF and IR identical	Full use of array space & bandwidth allocation
$\Delta x < \frac{\lambda z}{L}$	TF: Under IR: Over	$B_1 \le \frac{L}{2\lambda z}$	TF: If bandwidth criterion essentially satisfied IR: Perhaps better if bandwidth criterion violated	Relatively "long" z or large λ. Source bandwidth limitation.

Δx sample interval
λ wavelength
z propagation distance
B_1 source bandwidth
D_1 support of field in source plane

described that involves the source field bandwidth B_1. In practice, the source bandwidth criteria of Table 5.1 will never be satisfied because, as described in Sec. 2.2, a source with finite support will have infinite bandwidth. So an effective bandwidth B_1 can be used when considering the criteria. Further comments on the sampling regimes:

$\Delta x > \lambda z/L$: The "short distance" regime. Here, the support size available in the observation plane is limited. For the TF approach the size of the field in the observation plane that can be represented accurately is roughly $D_1+\lambda z/\Delta x$ (see Appendix A). This limitation is usually not a problem as the observation plane field is often negligible beyond the full width of $D_1+\lambda z/\Delta x$. Thus, the TF approach generally provides good results in this regime. The undersampled IR phase function has an aliased, periodic phase representation and using this approach produces periodic copies of the field. The source bandwidth B_1 is only limited in the usual way by the sampling theorem in the source plane.

$\Delta x = \lambda z/L$: This is the critical sampling situation where, remarkably, the sampled H and h functions as an FFT pair, turn out to have values that *exactly match* the analytic functions H and h (Appendix A). Under this condition the full bandwidth

of the sampled array ($1/2\Delta x$) is available for modeling the source and the full area of the array in the observation plane can be used.

$\boldsymbol{\Delta x < \lambda z/L}$: The "long distance" regime. Here, the bandwidth available for the source field becomes limited. The H chirp has an aliased phase form where any significant source bandwidth that extends beyond $L/(2\lambda z)$ leads to artifacts using the TF approach. Applying the IR approach actually corresponds to windowing, or filtering the source frequency content beyond ~ $L/(2\lambda z)$. This leads to "smoother", but not always accurate results.

5.4.3 Criteria applied to square beam example

What do the criteria of the previous section predict for the **sqr_beam** example of Sec. 5.3.1? The source field for that example is given by

$$U_1(x,y) = \text{rect}\left(\frac{x}{2w}\right)\text{rect}\left(\frac{y}{2w}\right), \tag{5-9}$$

where w = 0.051 m. The source support is $D_1 = 2w$ = 0.102 m, which easily fits within the side length L = 0.5 m. This is illustrated in Fig. 5.1. To consider the criteria in Table 5.1, an estimate is needed for the source effective bandwidth. Referring to Sec. 2.2, a reasonable estimate for the effective bandwidth is

$$B_1 \approx \frac{5}{w} = 98 \text{ cyc/m}. \tag{5-10}$$

Table 5.2 presents pertinent sampling information for the four propagation distances in the **sqr_beam** example. Referring to Table 5.2 as well as Fig. 5.4, the following observations can be made:

***z* = 1000 m**: H is oversampled by a factor of 2 relative to critical sampling. The observation plane size limitation has a negligable effect on the TF result [Fig. 5.4(a)] as the significant part of the field fits within $D_1+\lambda z/\Delta x$ = 0.352 m. The IR approach introduces periodic copies of the field separated by $\lambda z/\Delta x$ [Fig. 5.4(e)]

***z* = 2000 m**: Critical sampling. The TF and IR results are identical [Figs. 5.4(b) and (f)].

***z* = 4000 m**: H is undersampled by a factor of 2. The source bandwidth limit for propagation is 125 cyc/m, which is slightly larger than the source effective bandwidth B_1. Thus, most of the significant source spectrum obeys the criterion. But small ticks still creep into the TF approach result [Fig. 5.4(c)]. On the other hand, artifacts are not apparent in the IR result [Fig. 5.4(h)].

***z* = 20,000 m**: H is undersampled by a factor of 10. The available bandwidth of 25 cyc/m is only a fourth of B_1. Thus, the TF approach causes significant stair-

step artifacts [Fig. 5.4(d)]. The IR approach suppresses source frequency components that lie beyond the available bandwidth. This gives a smoother result but with the small spurious sidelobes near the array edge [Fig. 5.4(h)].

TABLE 5.2 SAMPLING REGIMES FOR **sqr_beam** SIMULATION EXAMPLE

z (m)	$\frac{\lambda z}{\Delta x\, L}$	TF (H) Sampling	IR (h) Sampling	Available Obs Plane Size (m)	Source Bandwidth Limit (cyc/m)
1000	0.5	Over	Under	$D_1+\lambda z/\Delta x = 0.352$	$1/2\Delta x = 250$
2000	1	Critical	Critical	$L = 0.5$	$1/2\Delta x = 250$
4000	2	Under	Over	$L = 0.5$	$L/2\lambda z = 125$
20,000	10	Under	Over	$L = 0.5$	$L/2\lambda z = 25$

Δx = 2 mm, L = 0.5 m, λ = 0.5 μm, N = 250, D_1 = 0.102 m

5.4.4 Propagation accuracy

How closely do the results of the **sqr_beam** examples follow the analytic Fresnel propagation results? Three particular cases are studied where the irradiance results appear to be "reasonable": TF, z =1000 m; TF, z = 2000 m; and IR, z = 20,000 m. For comparison purposes, a highly-oversampled, direct numerical evaluation of the Fresnel integral is used to provide a close approximation to the analytic result.[1]

In Fig. 5.6 the field magnitude and unwrapped phase at the observation plane are compared with numerical integration profiles. The TF and IR propagator profiles are displayed with solid lines and the numerical integration results are displayed with dashed lines. There is generally good consistence between the curves. Some comments on the comparison plots:

TF, z = 1000 m: The significant features of the magnitude profiles [Fig. 5.6(a)] match but smoothing is seen in the wings of the propagator result (the highly oscillating wings are the numerical integration result). The phase profiles [Fig. 5.6(b)] are nearly identical except at the edges of the array where the propagator phase slightly lags. $D_1+\lambda z/\Delta x$ = 0.352 m, which is roughly the apparent width of the magnitude curve before the propagator wings drop abruptly below the integration curve. Overall, the propagator result appears reasonably accurate. The primary deviation is in the wings and is of little consequence.

TF, z = 2000 m:The propagator magnitude curve [Fig. 5.6(c)] is nearly identical to the integration curve although the propagator curve is slightly elevated near the edge of the array as forced by periodic extension (difficult to see in the figure). The phase profiles [Fig. 5.6(d)] are essentially identical. In this case critical sampling leads to an extremely close match with the integration result.

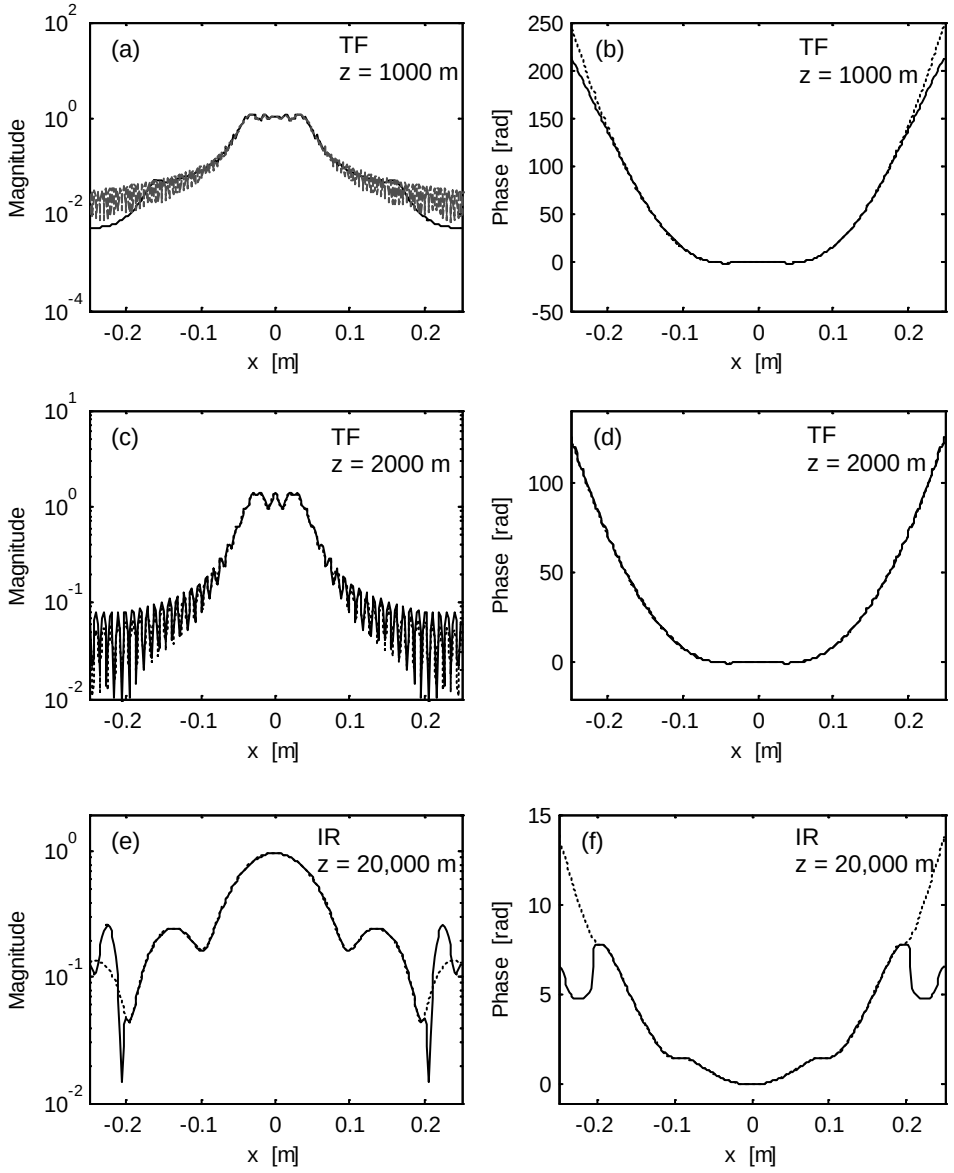

Figure 5.6 Solid curves are irradiance profiles (x-axis) for the transfer function (TF) and impulse response (IR) propagation approaches for propagation distances of 1000, 2000 and 20,000 m. Dashed curves are corresponding numerical integration results.

IR, *z* = 20,000 m: For this long-distance case, the IR propagator magnitude [Fig 5.6(e)] follows the integration result but exhibits the spurious lobes near the array edges. The propagator phase in Fig. 5.6(f) also has errors in the wings. A better result could be found by readjusting the sampling parameters, but this irradiance pattern might be acceptable in a pinch.

5.4.5 Sampling decisions

The steps for designing a propagation simulation are summarized in Table 5.3. As mentioned previously, a field with finite support in the source plane can not be bandlimited, so the bandwidth criteria in Table 5.1 are never satisfied. However, the criteria can still be used to help find reasonable simulation parameters, although, it often becomes something of an art form to juggle sampling and field parameters to get a satisfactory propagation result.

Critical sampling helps minimize artifacts by allowing full use of the array side length and sampling bandwidth. It seems prudent to try and use critical sampling, but maintaining this condition can be inconvenient. Consider that the critical sampling expression $\Delta x=\lambda z/L$ can be rearranged to give

$$M = \frac{\lambda z}{\Delta x^2} = \frac{L^2}{\lambda z} \tag{5-11}$$

which defines the critical sampling criterion in terms of the number of samples that span the array side length. For a given situation, the critical condition may dictate either too many samples for a practical FFT calculation or too few to adequately sample the source or observation planes. Other requirements can be at odds with the critical criterion. For example, phase screens used to simulate propagation through atmospheric turbulence have their own set of sample interval and array size conditions.[2] The sampling criterion can always be maintained with the help of interpolation or decimation. 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.

TABLE 5.3 PROPAGATION DESIGN STEPS

1	Consider source support: $D_1 < L$ by a factor of 2 or 3?
2	Determine sampling regime (Table 5.2).
	$\Delta x > \lambda z/L$: TF approach will often work well with some loss of observation plane support.
	$\Delta x = \lambda z/L$: Critical sampling, TF approach, best use of bandwidth and spatial support.
	$\Delta x < \lambda z/L$: IR approach with loss of available source bandwidth, artifacts.
3	Consider source bandwidth criteria in Table 5.2.
4	Reconsider source sampling depending on bandwidth criteria and resulting artifacts.

In practice, after checking the sampling regime, often the most expedient thing to do next is guess (did I say "guess"?) - I mean, estimate the sampling parameters then run a few trial simulations to see what happens. If there are signs of artifacts, as illustrated in Figs. 5.4 and 5.5, then the problem and simulation parameters need a closer look.

5.4.6 Split-step simulation, windowing, and expanding grids

Some of you might say – "Hey, wait! Let's just break that long distance propagation problem into a sequence of shorter, better behaved, TF propagations where each satisfies $\Delta x \geq \lambda z/L$!" But alas, the result is the same whether a single TF propagation is used or a sequence of shorter TF propagations. This is because a succession of TF propagations is the same as applying the product of the transfer functions to the initial field. For example, if the distance z_N is broken into a series of shorter distances, it is easy to show

$$H(f_X, f_Y; z_N) = H(f_X, f_Y; z_1)H(f_X, f_Y; z_2 - z_1)...H(f_X, f_Y; z_N - z_{N-1}) \tag{5-12}$$

where the z arguments are understood to replace z in (5-2). So even if the shorter propagations are critically sampled, the final result is the same as a single propagation! It is the total propagation distance that is important. But split-step simulations are applied in many situations for reasons such as propagating between a series of atmospheric turbulence phase screens.

Previously it was noted that the reason the IR approach behaved better for the long propagation example is that it effectively suppresses source frequency content where the frequency chirp function is going bad. In fact, the IR approach is mainly introduced to give a quick and relatively painless way to approach longer propagation distances. But there are other ways to handle this issue. Most researchers working with laser beam propagation simulations apply window functions to either suppress the source spectrum or remove energy in the wings of the source field. This combined with multi-step propagation can give good results. This subject is covered in more detail by Schmidt in reference 2.

Suppose a simulation involves some fixed parameters in the source or observation planes such that a single side length and sample interval will not serve for modeling both planes. In this situation the ability to independently select the physical side lengths of the source and observation planes is helpful. The *two-step method* allows the source and observation plane side lengths to be different. It is described and analyzed in Appendix B. While it still suffers from some of the same sampling limitations described for the TF approach, it affords flexibility in the simulation design.

5.5 Fraunhofer Pattern

The expression for the Fraunhofer pattern is repeated here

$$U_2(x_2,y_2)=\frac{\exp(jkz)}{j\lambda z}\exp\left[j\frac{k}{2z}\left({x_2}^2+{y_2}^2\right)\right]$$
$$\cdot\iint U_1(x_1,y_1)\exp\left[-j\frac{2\pi}{\lambda z}\left(x_2x_1+y_2y_1\right)\right]dx_1dy_1, \qquad (5\text{-}13)$$

where the source plane variables are indicted with the subscript 1 and the observation plane with subscript 2. When using the FFT to compute the Fraunhofer field, the source and observation plane side lengths are not generally the same. From (4-31) we have

$$\lambda z f_{X1}\rightarrow x_2, \qquad (5\text{-}14)$$

and using (2-18), we find the observation plane side length and sample interval in terms of the source plane parameters

$$L_2=\frac{\lambda z}{\Delta x_1} \quad \text{and} \quad \Delta x_2=\frac{\lambda z}{L_1}. \qquad (5\text{-}15)$$

So the observation plane coordinates are given as

$$x_2=\left[\frac{-L_2}{2}:\Delta x_2:\frac{L_2}{2}-\Delta x_2\right]=\lambda z\left[\frac{-1}{2\Delta x_1}:\frac{1}{L_1}:\frac{1}{2\Delta x_1}-\frac{1}{L_1}\right] \qquad (5\text{-}16)$$

If critical sampling is used ($\Delta x_1 = \lambda z/L_1$) then (5-15) indicates the side lengths will be equal, $L_2 = L_1$. Otherwise the side lengths are different. The function **propFF** that computes the Fraunhofer pattern follows:

```
function[u2,L2]=propFF(u1,L1,lambda,z);
% propagation - Fraunhofer pattern
% assumes uniform sampling
% u1 - source plane field
% L1 - source plane side length
% lambda - wavelength
% z - propagation distance
% L2 - observation plane side length
% u2 - observation plane field
%
[M,N]=size(u1);           %get input field array size
dx1=L1/M;                 %source sample interval
k=2*pi/lambda;            %wavenumber
%
L2=lambda*z/dx1;          %obs sidelength
dx2=lambda*z/L1;          %obs sample interval
```

```
x2=-L2/2:dx2:L2/2-dx2;  %obs coords
[X2,Y2]=meshgrid(x2,x2);
%
c=1/(j*lambda*z)*exp(j*k/(2*z)*(X2.^2+Y2.^2));
u2=c.*ifftshift(fft2(fftshift(u1)))*dx1^2;
end
```

This function also outputs the side length of the observation plane `L2` so it doesn't have to be computed again in the main script.

To try this out make the following changes in the **sqr_beam** routine,

```
w=0.011;              %source half width (m)

[u2,L2]=propFF(u1,L1,lambda,z);

dx2=L2/M;
x2=-L2/2:dx2:L2/2-dx2; %obs ords
y2=x2;

imagesc(x2,y2,nthroot(I2,3));%stretch image contrast
```

Use **Save as** and give this file a new name, **sqr_beam_FF**. The source half-width `w = 0.011` m with the propagation distance of z = 2000 m gives a Fresnel number of 0.12, which is reasonable for the Fraunhofer approximation. Running **sqr_beam_FF** gives the irradiance results of Fig. 5.7. Stretch the contrast of the irradiance pattern with the `nthroot` function to bring out the sidelobes.

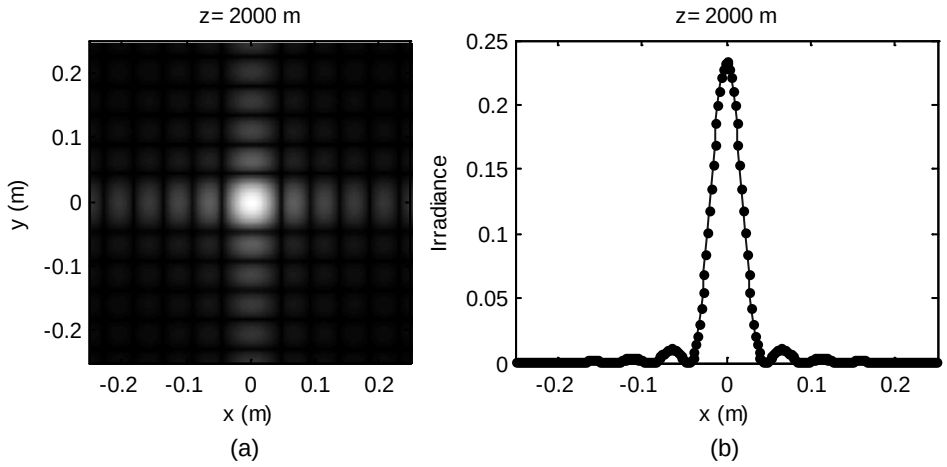

Figure 5.7 Fraunhofer irradiance (a) pattern and (b) x-profile. Points in (b) are analytic values.

In this case the simulation result can be checked against the analytic Fraunhofer result. Take the Fourier transform of the source distribution

$$\Im\left\{\text{rect}\left(\frac{x_1}{2w}\right)\text{rect}\left(\frac{y_1}{2w}\right)\right\} = 4w^2\,\text{sinc}(2wf_{X1})\text{sinc}(2wf_{Y1}). \qquad (5\text{-}17)$$

Substitute $x_2/\lambda z$ for f_{X1} and $y_2/\lambda z$ for f_{Y1} and include the multipliers to give the Fraunhofer field

$$U_2(x_2, y_2) = \frac{\exp(jkz)}{j\lambda z}\exp\left(j\frac{k}{2z}\left(x_2^{\,2} + y_2^{\,2}\right)\right)4w^2\,\text{sinc}\left(\frac{2w}{\lambda z}x_2\right)\text{sinc}\left(\frac{2w}{\lambda z}y_2\right). \qquad (5\text{-}18)$$

The irradiance pattern is $I_2(x_2, y_2) = \left|U_2(x_2, y_2)\right|^2$, which gives

$$I_2(x_2, y_2) = \left(\frac{4w^2}{\lambda z}\right)^2 \text{sinc}^2\left(\frac{2w}{\lambda z}x_2\right)\text{sinc}^2\left(\frac{2w}{\lambda z}y_2\right). \qquad (5\text{-}19)$$

The analytic function is also plotted in Fig. 5.7(b). Try and put (5-19) into the script to get the plot in Fig. 5.7(b).

Usually the irradiance is of interest when calculating the Fraunhofer pattern so the complex exponentials in front disappear. In this case the only worry is the usual source plane sampling condition, $B_1 \le 1/2\Delta x_1$. But suppose the Fraunhofer *field* is of interest, including the chirp term out front. Based on (5-7), the chirp function $\exp[jk(2z)^{-1}(x_2^{\,2} + y_2^{\,2})]$ will be adequately sampled in the observation plane if $\Delta x_2 \le \lambda z / L_2$, or equivalently, by applying (5-10), when the source plane sampling is

$$\Delta x_1 \ge \frac{\lambda z}{L_1} \qquad (5\text{-}20)$$

If (5-20) is not satisfied, the chirp phase will be aliased when tacked on the front of the transform. Furthermore, considering the Fresnel number where $w^2 / \lambda z << 1$ in the Fraunhofer regime and combining this with (5-20) leads to

$$L_1 >> \frac{w^2}{\Delta x_1} \quad \text{or} \quad M >> \frac{w^2}{\Delta x_1^{\,2}}. \qquad (5\text{-}21)$$

This says the number of grid samples M needs to be much greater than the square of the number of samples across the source aperture radius $w^2/\Delta x_1^2$. This means a large M is required. Fortunately the Fraunhofer phase is not often needed.

5.6 Coding Efficiency

Some aspects of the functions and scripts presented in this chapter (and throughout the book) are inefficient for reasons related to ease of use or presentation. For example, `meshgrid` is used in the **sqr_beam** code to define the sample coordinates but then is applied again in the propagation functions to redefine the coordinates. This makes the functions easier to use but is redundant. Speed and efficiency are not a big problem for the examples in this book, but can be an important issue when running many iterations of a propagation code. To help tune-up your code, try out two coding tools provided by MATLAB®: **M-Lint** and the **Profiler**.

M-Lint is an analyzer that checks the code in the Editor for possible problems. Get to **M-Lint** by going to the MATLAB® Desktop toolbar. Select **Tools > Check Code with M-Lint**. You will see a display of potential problems that **M-Lint** finds.

The **Profiler** tracks the execution time of the various statements and function calls in your code. It can help find problems and improve the efficiency of your code. On the Desktop toolbar select **Tools > Open Profiler**. The Profiler window will appear. The file to be profiled is in the menu box next to "Run this code:" Click on **Start Profiling**. The code will execute and a Profile Summary window will give a breakdown on the execution times.

References

1. Voelz, D. G., and M. C. Roggemann, "Digital simulation of scalar optical diffraction: revisiting chirp function sampling criteria and consequences," Appl. Opt. 48, 6132-6142 (2009).
2. Schmidt, J. D., *Numerical Simulation of Optical Wave Propagation with Examples in MATLAB*, SPIE Press Book (2010).

Exercises

5.1 Assume a circular aperture with a radius of w = 0.05 m illuminated by a plane-wave where λ = 0.5 μm. Assume a propagation distance of 1000 m and a simulation array size of 500×500 samples. Assume critical sampling for a Fresnel propagation.

(a) Find the side length L_1, sample interval Δx and the Nyquist frequency F_N.

(b) Determine the source effective bandwidth B_1. Is $B_1 < F_N$? How many samples span the diameter of the circle function?

(c) Determine the Fresnel number. Is the propagation distance within the Fresnel region?

(d) Using the value for L_1 from part (a), simulate Fresnel propagation for distances of 500, 1000 and 2000 m. Try both TF and IR simulations.

5.2 Return to the three apertures of Exercise 4.4. Simulate Fresnel propagation of the source fields in a 500×500 sample array with the following parameters:

(a) L_1 = 2 mm; z = 0.5, 1, and 5 cm

(b) L_1 = 2.5 cm; z = 0.5, 2, and 5 m

(c) L_1 = 2.5 cm; z = 0.5, 2, and 5 m

(d) What is the critical sample distance for the three cases?

5.3 A useful diagnostic for propagation simulations is to compute the power in the source and observation planes. Assuming no absorption or scatter of the light, which is true for the simulations presented in this book, the power (proportional to watts) should be conserved. In other words, the source and observation planes should contain the same optical power. If not, there may be a code error or a sampling problem. The power is the integrated irradiance or

$$P = \int\int I(x, y)dxdy$$

For the **sqr_beam** example in this chapter:

(a) Add code to compute the power at the source and observation planes for the various propagations. Use the MATLAB® function `sum`. Maybe two of these? What about *dx* and *dy*? [Remove the semi-colon from the end of the line with the power calculation so the value displays in the Command Window when the script is executed.]

(b) Are there discrepancies in some of the **sqr_beam** example cases regarding the power in the source and observation planes?

5.4 Fresnel diffraction involves a paraxial (small ray angle) assumption that limits the minimum propagation distance whereas Rayleigh-Sommerfeld diffraction is essentially exact over all distances.

(a) Write transfer function and impulse response propagators for *Rayleigh-Sommerfeld diffraction*.

(b) Test the routines with the **sqr_beam** example for the four distances (1000, 2000, 4000, 20,000 m). Are there differences between the Fresnel and Rayleigh-Sommerfeld results?

(c) Calculate the Fresnel numbers for the **sqr_beam** examples. What can you say about applying Fresnel versus Rayleigh-Sommerfeld propagation in this case?

5.5 Gaussian Beam. Fourier methods are well-suited for simulating laser beam propagation. A laser beam typically obeys the paraxial ray angle approximation, which is valid for the Fresnel expression. Also, the Gaussian function used to describe the beam profile is more forgiving in terms of sampling artifacts than a square or circular aperture beam of similar support. Laser textbooks define the irradiance distribution of a Gaussian laser beam ($TEM_{0,0}$ mode) propagating in the z direction as

$$I(x, y, z) = I_0 \left(\frac{w_0}{w(z)} \right)^2 \exp\left(-2 \frac{(x^2 + y^2)}{w(z)^2} \right), \tag{5-22}$$

where x, y are transverse spatial variables, I_0 is the source irradiance value at beam center (x, y = 0), w_0 = source beam e^{-2} radius (at z = 0) and $w(z)$ is the beam radius at distance z. The beam radius is given by

$$w(z) = w_0 \sqrt{1 + \left(\frac{z}{z_R} \right)^2}, \tag{5-23}$$

where z_R is the Rayleigh range defined by

$$z_R = \frac{\pi w_0^2}{\lambda}, \tag{5-24}$$

For the following questions, assume a source optical field (z = 0) given by

$$U_0(x, y) = A_0 \exp\left[-\frac{(x^2 + y^2)}{w_0^2} \right] \tag{5-25}$$

where w_0 = 1 mm, λ = 0.633 μm, and A_0 = 1 V/m. To be accurate, $I_0 = |A_0|^2/(2\eta)$ W/m^2 where η = 377 Ω in free space.

(a) Assume a side length of 15 mm and an array of 250×250 elements. Simulate Fresnel propagation for distances of 1, 5 and 10 m. Compare these results with analytic result of (5-22).

(b) What is the propagation distance for critical sampling? Test the source bandwidth criterion for the 10 m propagation distance.

(c) Find the Fraunhofer irradiance expression for the U_0 beam. Show your result is consistent with the analytic expression.

5.6 Return to the three apertures of Exercise 4.4. Simulate Fraunhofer propagation of the source fields in a 500×500 sample array with the following parameters:

(a) L_1 = 2 mm; z = 5 m

(b) L_1 = 2.5 cm; z = 50 m

(c) L_1 = 2.5 cm; z = 50 m

Compare discrete and analytic results in an x-axis irradiance profile. (There is no attempt in this exercise to model the Fraunhofer field such that the phase is adequately sampled.)

5.7 For any square aperture illuminated by a plane-wave, find a criterion for the number of linear samples M necessary to adequately sample the Fraunhofer field phase. The result should contain no variables - just a number.

5.8 Code up the two-step propagator function described in Appendix B. Test it with the **sqr_beam** example for the following cases:

(a) L_1 = 0.5 m; z = 2000 m (critical sampling distance). Examine the field magnitude profiles for L_2 = 0.2, 0.4, 0.5, 0.6 and 1 m. The 0.5 m case should be identical to the TF propagation result. What are the apparent artifacts for the other distances?

(b) L_1 = 0.5 m; z = 20,000 m. Does adjusting the L_2 size reduce artifacts in the observation plane result?