Chapter 7 Wave propagation and Fourier optics

Fourier optics describes propagation of light in optical systems using Fourier transform techniques. These techniques are useful since many operations are linear and spatially shift-invariant. They form the basis for analyzing and designing optical imaging and computation systems.

7.1 Propagation of light in the paraxial approximation

Although the classical wave description of light is as a transverse electromagnetic wave, many effects can be studied using a scalar rather than the full vector wave equation. In free space, we have

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} = \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} \tag{7.1}$$

In this equation ψ represents a component of the electric or magnetic field. For monochromatic, coherent light, we can write

$$\psi(x, y, z, t) = \psi(x, y, z, 0) \mathrm{e}^{-\mathrm{j}\omega t}. \tag{7.2}$$

Substituting this into the wave equation, we obtain Helmholtz's equation

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} = -k^2 \psi, \tag{7.3}$$

where $\omega/k = c$.

Consider propagation which is nearly parallel to the z axis, so that

$$\psi(x, y, z, 0) = f_z(x, y) \mathrm{e}^{\mathrm{j}kz}, \tag{7.4}$$

where $f_z(x, y)$ varies slowly with z. (Note, for example that for a plane wave travelling parallel to the z axis, $f_z(x, y)$ is constant).

Substituting (7.4) into Helmholtz's equation (7.3) yields

$$\mathrm{e}^{\mathrm{j}kz} \left[\frac{\partial^2 f_z}{\partial x^2} + \frac{\partial^2 f_z}{\partial y^2} + \frac{\partial^2 f_z}{\partial z^2} + 2\mathrm{j}k \frac{\partial f_z}{\partial z} - k^2 f_z \right] = -k^2 f_z \mathrm{e}^{\mathrm{j}kz}. \tag{7.5}$$

The *paraxial approximation* neglects $\partial^2 f_z / \partial z^2$ since f_z is assumed to vary slowly with z. This yields the paraxial wave equation

$$\boxed{\frac{\partial^2 f_z}{\partial x^2} + \frac{\partial^2 f_z}{\partial y^2} + 2\mathrm{j}k \frac{\partial f_z}{\partial z} = 0} \tag{7.6}$$

If we are given $f_{z_1}(x, y)$, the solution of this partial differential equation will give the amplitude distribution at $z = z_2$.

7.2 Solving the paraxial wave equation

The paraxial wave equation is most easily solved by computing its two-dimensional Fourier transform. The two-dimensional Fourier transform of a function $g(x, y)$ is defined as

$$G(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\exp[-\mathrm{j}2\pi(ux+vy)]\,\mathrm{d}x\,\mathrm{d}y \tag{7.7}$$

The corresponding inverse Fourier transform is

$$g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(u,v)\exp[+\mathrm{j}2\pi(ux+vy)]\,\mathrm{d}u\,\mathrm{d}v \tag{7.8}$$

Exercise: Prove that the above transforms are inverses, using the properties of the usual one-dimensional Fourier transform.

We now consider the paraxial wave-equation (7.6). Let $F_z(u, v)$ denote the two-dimensional Fourier transform of $f_z(x, y)$ so that the Fourier transformed paraxial wave equation is

$$(\mathrm{j}2\pi u)^2 F_z + (\mathrm{j}2\pi v)^2 F_z + 2\mathrm{j}k\frac{\partial F_z}{\partial z} = 0 \tag{7.9}$$

or

$$\frac{\partial F_z}{\partial z} = \left(\frac{2\pi^2}{\mathrm{j}k}\right)(u^2+v^2)F_z(u,v) \tag{7.10}$$

This may be integrated directly, yielding

$$\boxed{F_z(u,v) = F_0(u,v)\exp\left[-\frac{\mathrm{j}2\pi^2}{k}\left(u^2+v^2\right)z\right]} \tag{7.11}$$

Since the (two-dimensional) inverse Fourier transform of $\exp[-\mathrm{j}2\pi^2(u^2 + v^2)z/k]$ is

$$h(x,y) = \frac{1}{\mathrm{j}\lambda z}\exp\left[\frac{\mathrm{j}k}{2z}(x^2+y^2)\right] \tag{7.12}$$

where $\lambda = 2\pi/k$ (check this!), we can take the inverse Fourier transform of the product in (7.11) to obtain the convolutional relationship

$$f_z(x,y) = (f_0 * h)(x,y) \tag{7.13}$$

or, writing out the convolution in full

$$\boxed{f_z(x,y) = \frac{1}{\mathrm{j}\lambda z}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_0(x_0,y_0)\exp\left[\frac{\mathrm{j}k}{2z}\left((x-x_0)^2+(y-y_0)^2\right)\right]\mathrm{d}x_0\mathrm{d}y_0.} \tag{7.14}$$

,This is the *paraxial diffraction integral.* It states that the field amplitudes at the plane at z are related to those at the plane at $z = 0$ by a linear, shift-invariant filtering operation with "impulse response" $h(x, y)$.

Let us now consider the physical interpretation of this result. The impulse response gives the amplitude of the light on a plane at distance z away from a point source of light (e.g. a pinhole in an opaque screen) located at the origin. If we put back the full dependence on z and t we find that in the paraxial approximation, a point source gives

$$\psi(x,y,z,t) = \frac{1}{\mathrm{j}\lambda z} \exp\left[\frac{\mathrm{j}k}{2z}\left(x^2+y^2\right)\right] \exp(\mathrm{j}kz)\exp(-\mathrm{j}\omega t) \tag{7.15}$$

The spatial dependence of the phase term is

$$\exp\left[\mathrm{j}kz\left(1+\frac{x^2+y^2}{2z^2}\right)\right] \tag{7.16}$$

On the other hand, by Huygen's construction we expect it to be

$$\exp\left[\mathrm{j}k\sqrt{x^2+y^2+z^2}\right] \tag{7.17}$$

However if $z \gg x, y$ (which is true in the paraxial approximation),

$$\mathrm{j}k(x^2+y^2+z^2)^{\frac{1}{2}} = \mathrm{j}kz\left(1+\frac{x^2+y^2}{z^2}\right)^{\frac{1}{2}} \tag{7.18}$$

$$\approx \mathrm{j}kz\left[1+\frac{1}{2}\left(\frac{x^2+y^2}{z^2}\right)-\frac{1}{8}\left(\frac{x^2+y^2}{z^2}\right)^2+\ldots\right] \tag{7.19}$$

and so the phasefronts calculated by the paraxial approximation are very close to the true spherical wavefronts.

Note that

- The $1/z$ factor in the amplitude represents spherical spreading (inverse square law for intensity).
- The additional phase term $-\mathrm{j}$ is actually present, although it is not predicted by a simple application of Huygen's construction.

Exercises:

1. Show that if $f_0(x, y) = 1$, then the diffraction formula predicts that $f_z(x, y) = 1$ for all z. This means that a plane-wave propagating along the z axis is a solution.
2. Show that if $f_0(x, y) = \exp(\mathrm{j}ky\sin\theta)$, the diffraction formula predicts that

   $$f_z(x,y) = \exp(\mathrm{j}ky\sin\theta)\exp\left(-\mathrm{j}\frac{kz}{2}\sin^2\theta\right).$$

   Interpret this result physically, and compare it with the exact result (i.e., without the paraxial approximation).

7.3 Fresnel and Fraunhofer Diffraction

Whenever the paraxial diffraction integral is valid, the observer is said to be in the region of *Fresnel diffraction*. The quadratic terms in the exponential of the diffraction integral can be rewritten as

$$\exp\left[\frac{\mathrm{j}k}{2z}\left((x-x_0)^2+(y-y_0)^2\right)\right] = \exp\left[\frac{\mathrm{j}k}{2z}\left(x^2+y^2\right)\right]\exp\left[-\frac{\mathrm{j}k}{z}\left(xx_0+yy_0\right)\right]\exp\left[\frac{\mathrm{j}k}{2z}\left(x_0^2+y_0^2\right)\right] \tag{7.20}$$

Hence we may write

$$f_z(x,y) = \frac{1}{\mathrm{j}\lambda z} P_z(x,y) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ f_0(x_0,y_0) P_z(x_0,y_0) \right\} \exp\left[-\mathrm{j}2\pi \left(\frac{xx_0 + yy_0}{z\lambda} \right) \right] \mathrm{d}x_0 \mathrm{d}y_0 \quad (7.21)$$

where

$$P_z(x,y) = \exp\left[\frac{\mathrm{j}k}{2z} \left(x^2 + y^2 \right) \right] \quad (7.22)$$

is a phase factor with unit modulus. Thus the paraxial diffraction integral for propagating a field through a distance z in free space may be interpreted as a sequence of three operations

- Multiplication of $f_0(x_0, y_0)$ by the phase factor $P_z(x_0, y_0)$,
- Calculation of a two-dimensional Fourier transform with spatial-frequency variables $x/(z\lambda)$ and $y/(z\lambda)$,
- Multiplication of the result by a further phase factor $P_z(x, y)$.

In a diffraction experiment, we usually consider the field at the plane $z = 0$ to be non-zero only over a relatively small region, specified by the aperture or mask placed in that plane. If we suppose that z is chosen to be so large that the phase factor $P_z(x_0, y_0) \approx 1$ over the entire region of the (x_0, y_0) plane in which $f_0(x_0, y_0)$ is non-zero. The equation (7.21) may then be written as

$$f_z(x,y) = \frac{1}{\mathrm{j}\lambda z} P_z(x,y) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_0(x_0,y_0) \exp\left[-\mathrm{j}2\pi \left(\frac{xx_0 + yy_0}{z\lambda} \right) \right] \mathrm{d}x_0 \mathrm{d}y_0 \quad (7.23)$$

In this regime, $f_z(x, y)$ is just the two-dimensional Fourier transform of $f_0(x_0, y_0)$ except for a multiplicative phase factor which does not affect the intensity of the diffracted light. This is called the *Fraunhofer approximation*. It is valid provided that $z \gg k(x_0^2 + y_0^2)_{\max}/2$.

7.4 The diffraction grating

For simplicity, specialize to one dimension so that

$$f_z(x,y) = \frac{1}{\sqrt{\mathrm{j}\lambda z}} \exp\left(\mathrm{j}\frac{kx^2}{2z} \right) \int_{-\infty}^{\infty} f_0(x_0) \exp\left(-\mathrm{j}\frac{2\pi x x_0}{z\lambda} \right) \mathrm{d}x_0. \quad (7.24)$$

Consider a screen placed at $z = 0$ illuminated by a plane wave travelling along the z axis as shown in Figure 7.1

If the transmission function of the screen is $t(x_0)$, this is also the amplitude of the field immediately after the screen for an incident plane wave of unit amplitude. For a a diffraction grating with $2N+1$ rectangular slits of width w separated by distance d,

$$f_0(x_0) = t(x_0) = \left(\sum_{k=-N}^{N} \delta\left(x_0 - kd\right) \right) * \Pi(x_0/w) \quad (7.25)$$

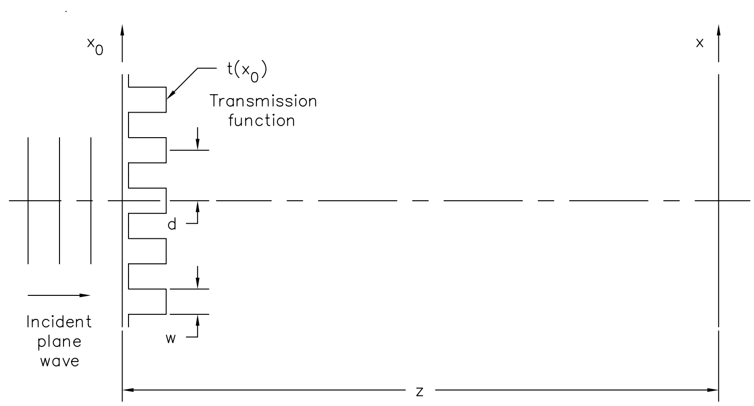


Figure 7.1 Diffraction configuration

where the asterisk denotes convolution and $\Pi(x)$ denotes the "top hat" function which is one for $|x| \leq \frac{1}{2}$ and zero otherwise. The Fourier transform of this is

$$F_0(u) = \left\{ \frac{1}{d} \sum_{k=-\infty}^{\infty} \delta(u - k/d) * (2N+1)\, d \operatorname{sinc}[(2N+1)\, du] \right\} w \operatorname{sinc}(wu) \tag{7.26}$$

$$= (2N+1)\, w \operatorname{sinc}(wu) \sum_{k=-\infty}^{\infty} \operatorname{sinc}[(2N+1)(du - k)]. \tag{7.27}$$

The intensity of the Fourier diffraction pattern is

$$|f_z(x)|^2 = \frac{1}{\lambda z} \left| F_0\left(\frac{x}{\lambda z}\right) \right|^2. \tag{7.28}$$

Figure 7.2 is a plot of $F_0(u)$. The diffraction pattern consists of bright lines positioned at $x_n = n\lambda z/d$. Each line is a sinc function whose first zero is at $\lambda z/[(2N+1)d]$ away from the peak. Hence if N is large, these lines are very sharp. If the incoming light consists of many wavelengths, these components are separated spatially by the grating.

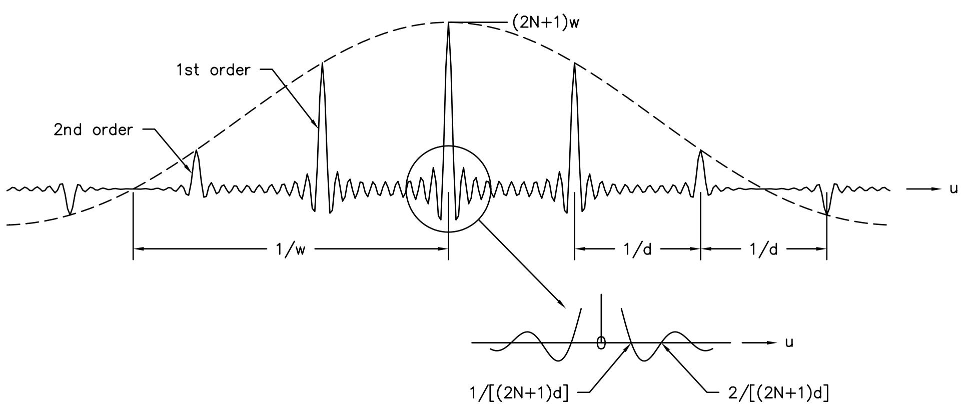


Figure 7.2 Fraunhofer diffraction amplitude

The overall envelope of the diffraction pattern which determines the amount of light diffracted into each order of the spectrum is determined by the width of each slit. Gratings can be designed to diffract most of the light into a particular order of the spectrum or to suppress unwanted orders.

Exercises:

1. Show that for red light ($\lambda \approx 600\,\mathrm{nm}$) and an aperture width of 1 cm, we require $z \gg 120\,\mathrm{m}$ to satisfy the Fraunhofer approximation.

2. Consider a sinusoidal amplitude grating illuminated by a plane wave travelling along the z axis for which the transmission function of the grating is

$$t(x,y) = \frac{1}{2}\left[1 + m\cos\left(2\pi f_0 x\right)\right]\Pi(x/l)\Pi(y/l),$$

where $\Pi(x)$ is the top-hat function which is unity if $|x| < \frac{1}{2}$. Show that the intensity of the Fraunhofer diffraction pattern at z is given by

$$I(x,y) = \left[\frac{l^2}{2\lambda z}\right]^2 \mathrm{sinc}^2\left(\frac{ly_0}{\lambda z}\right) \times \left\{\mathrm{sinc}\left(\frac{lx_0}{\lambda z}\right) + \frac{m}{2}\,\mathrm{sinc}\left(\frac{l\left(x_0 + f_0\lambda z\right)}{\lambda z}\right) + \frac{m}{2}\,\mathrm{sinc}\left(\frac{l\left(x_0 - f_0\lambda z\right)}{\lambda z}\right)\right\}^2$$

Sketch the form of $I(x,y)$.

3. Compute the Fraunhofer diffraction pattern of a circular aperture of diameter d illuminated by a plane wave along the z axis. Find where the first zero of the diffraction pattern occurs.

You may find the following identities useful

$$J_0(z) = \frac{1}{\pi}\int_0^\pi \cos(z\cos\theta)\,\mathrm{d}\theta$$

$$\int_0^z tJ_0(t)\,\mathrm{d}t = zJ_1(z).$$

7.5 Fresnel diffraction – numerical calculation

Numerical techniques for calculating Fresnel diffraction patterns are often the only feasible methods for practical problems. The two analytically equivalent methods turn out to be useful in different regimes.

7.5.1 The convolutional approach

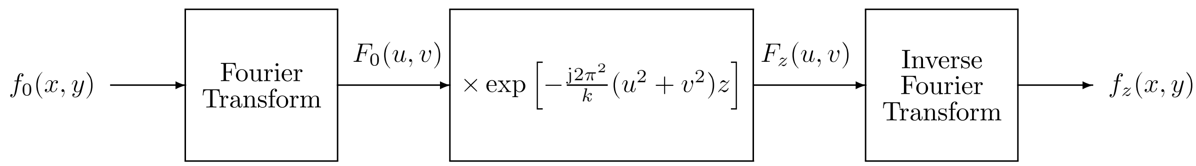
The paraxial diffraction integral is essentially a convolutional relationship

$$f_z(x,y) = \frac{1}{\mathrm{j}\lambda z}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_0(x_0,y_0)\exp\left[\frac{\mathrm{j}k}{2z}\left((x-x_0)^2 + (y-y_0)^2\right)\right]\mathrm{d}x_0\,\mathrm{d}y_0. \tag{7.29}$$

This convolution can be calculated by multiplying together the Fourier transforms

$$F_z(u,v) = F_0(u,v)\exp\left[-\frac{\mathrm{j}2\pi^2}{k}(u^2+v^2)z\right] \tag{7.30}$$

The following block diagram shows the steps involved in the computation. This is useful for small z since the variation in the phase term would cause aliasing if the change in $2\pi^2(u^2+v^2)z/k$ is too great between adjacent sample points. As $z \to 0$, we see that $f_z(x,y) \to f_0(x,y)$ which corresponds to the "geometrical shadow" of the diffracting obstacle.



7.5.2 The Fourier transform approach

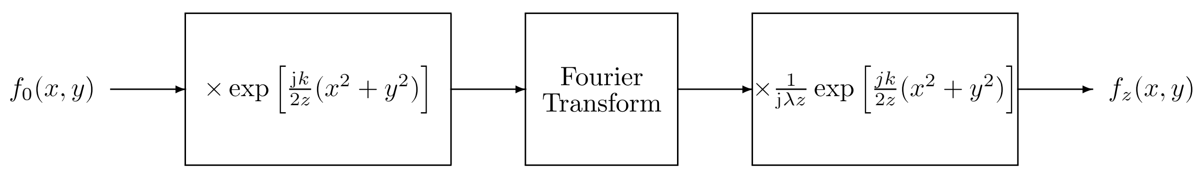
By expanding the quadratic phase term, we saw that the paraxial diffraction integral can be written as

$$f_z(x,y) = \frac{1}{\mathrm{j}\lambda z} P_z(x,y) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{f_0(x_0,y_0)\, P_z(x_0,y_0)\} \exp\left[-\mathrm{j}2\pi\left(\frac{xx_0+yy_0}{z\lambda}\right)\right] \mathrm{d}x_0\,\mathrm{d}y_0 \quad (7.31)$$

where

$$P_z(x,y) = \exp\left[\frac{\mathrm{j}k}{2z}\left(x^2+y^2\right)\right] \quad (7.32)$$

The block diagram shows the steps involved in the computation. This is useful for large z since the phase terms change slowly when z is large. As $z \to \infty$, we see that $f_z(x,y)$ tends to the Fourier transform of $f_0(x,y)$ which is the usual result for Fraunhofer diffraction.



7.6 Matlab code for computing Fresnel diffraction patterns

The following MATLAB function computes a one-dimensional Fresnel diffraction pattern using the algorithms discussed above. A fast Fourier transform is used to compute a discrete approximation to the Fourier transform.

```
function [f1,dx1,x1]=fresnel(f0,dx0,z,lambda)
%
% [fz,dx1,xbase1]=fresnel(f0,dx,z,lambda)
%  computes the Fresnel diffraction pattern at distance
%  "z" for light of wavelength "lambda" and an input
%  vector "f0" (must have a power of two points)
%  in the plane z=0, with intersample distance "dx0".
%  Returns the diffraction pattern in "f1", with
%  intersample separation in "dx1", and a baseline
%  against which f1 may be plotted as "x1".
%
N = length(f0); k = 2*pi/lambda;
%
% Compute the critical distance which selects between
```

```
%  the two methods
%
zcrit = N * dx0^2/lambda;
%
if z < zcrit
   %
   % Carry out the convolution with the Fresnel
   %  kernel by multiplication in the Fourier domain
   %
   du = 1./(N*dx0);
   u = [0:N/2-1 -N/2:-1]*du;        % Note order of points for FFT
   H = exp(-i*2*pi^2*u.^2*z/k);     % Fourier transform of kernel
   f1 = ifft( fft(f0) .* H );       % Convolution
   dx1 = dx0;
   x1 = [-N/2:N/2-1]*dx1;           % Baseline for output
else
   %
   % Multiply by a phase factor, compute the Fourier
   %  transform, and multiply by another phase factor
   %
   x0 = [-N/2:N/2-1] * dx0;         % Input f0 is in natural order
   g = f0 .* exp(i*0.5*k*x0.^2/z);  % First phase factor
   G = fftshift(fft(fftshift(g)));  % Fourier transform
   du = 1./(N*dx0); dx1 = lambda*z*du;
   x1 = [-N/2:N/2-1]*dx1;           % Baseline for output
   f1 = G .* exp(i*0.5*k*x1.^2/z);  % Second phase factor
   f1 = f1 .* dx0 ./ sqrt(i*lambda*z);
end
```

Figure 7.3 shows the results of applying this program to the calculation of the diffraction pattern of a single slit of width 1 mm illuminated with red light of wavelength 600 nm. A grid of 1024 points is used in which the slit occupies the central 50 points. The Fresnel diffraction pattern is calculated at distances of 0.01 m, 0.05 m, 0.2 m, 1 m and 2 m.

7.7 Fourier transforming and imaging properties of lenses

Consider a thin convex lens of focal length f. Plane waves incident on the lens parallel to the axis are converted into spherical wavefronts which collapse onto the focus. The action of the lens is to introduce a position-dependent phase shift. Consider this phase shift as a function of x and y. The portion of the wavefront at a distance r from the centre must travel an additional distance

$$d \approx \frac{r^2}{2f} \tag{7.33}$$

relative to the centre. This corresponds to a phase shift at (x, y) of

$$\exp\left[-\mathrm{j}\frac{k\left(x^2+y^2\right)}{2f}\right] \tag{7.34}$$

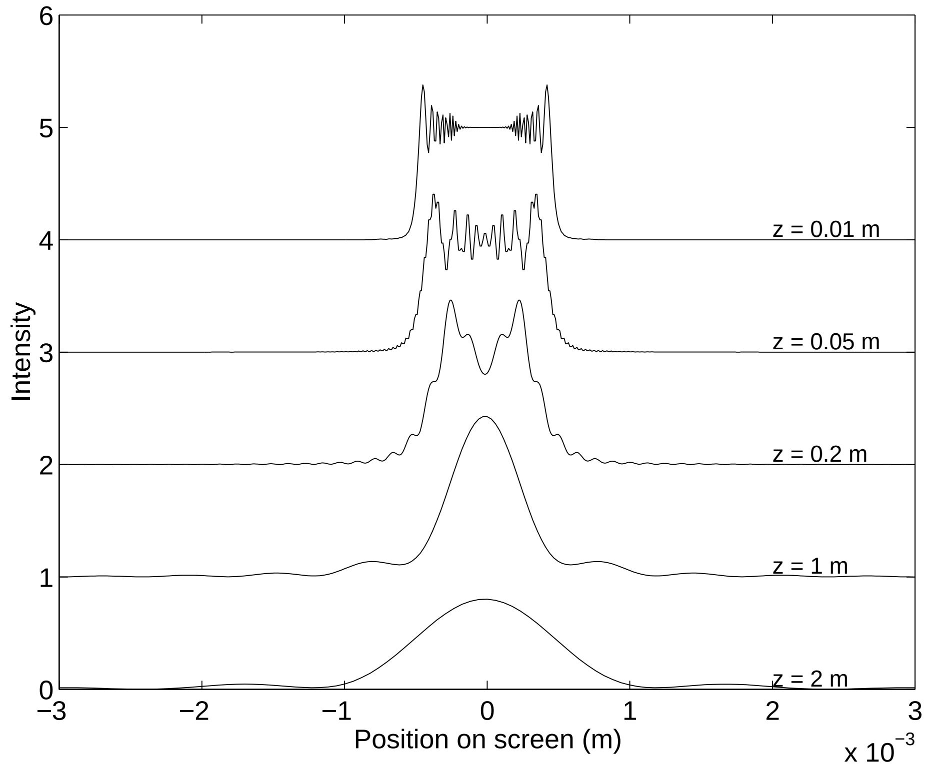


Figure 7.3 One-dimensional Fresnel diffraction from a slit

relative to the centre. Thus the effect of a thin lens is to multiply $f(x, y)$ by $\exp(-\mathrm{j}k(x^2+y^2)/(2f))$. (*Note:* In fact the phase change introduced is $\phi_0 - k(x^2 + y^2)/(2f)$ since we have to retard the centre rather than advance the edges.)

We thus see that:

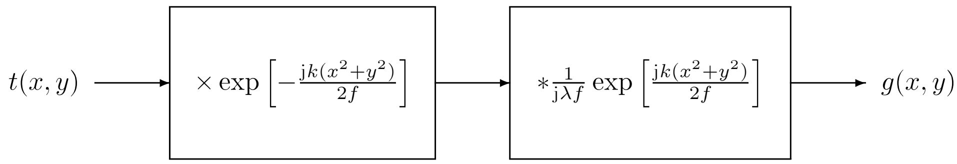
- Propagation through space by a distance z leads us to *convolve* the amplitude $f_z(x, y)$ with the propagation Green's function

$$\frac{1}{\mathrm{j}\lambda z} \exp\left[\mathrm{j}\frac{k\left(x^2+y^2\right)}{2z}\right]. \tag{7.35}$$

- Passing through a thin lens of focal length f leads us to *multiply* the amplitude $f_z(x, y)$ by the transmission function of the lens.

7.7.1 Transparency placed against lens

Consider a transparency with amplitude transmission function $t(x, y)$ placed against a thin convex lens of focal length f. The transparency is illuminated by plane waves travelling along the z axis and we wish to calculate the image $g(x, y)$ on a screen at distance f from the lens. The transformations undergone by the light may be represented by the block diagram:



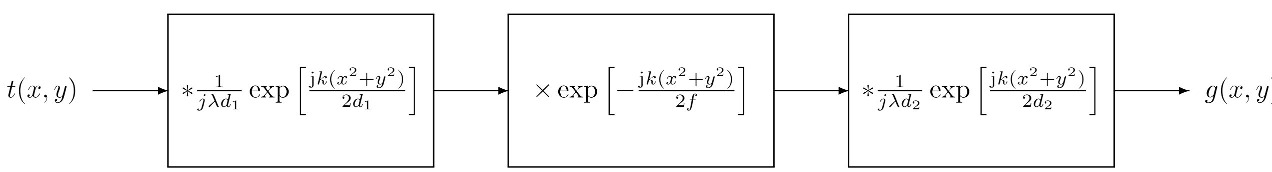
From the block diagram we immediately write down the expression for $g(x,y)$. This is

$$\begin{aligned}
g(x_1,y_1) &= \frac{1}{j\lambda f}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} t(x_0,y_0)\exp\left[-j\frac{k\left(x_0^2+y_0^2\right)}{2f}\right]\exp\left[j\frac{k\left((x_1-x_0)^2+(y_1-y_0)^2\right)}{2f}\right] dx_0\, dy_0 \\
&= \frac{1}{j\lambda f}\exp\left[j\frac{k\left(x_1^2+y_1^2\right)}{2f}\right]\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} t(x_0,y_0)\exp\left[-j2\pi\left(\frac{x_1x_0+y_1y_0}{f\lambda}\right)\right] dx_0\, dy_0 \\
&= \frac{1}{j\lambda f}\exp\left[j\frac{k\left(x_1^2+y_1^2\right)}{2f}\right] T\left(\frac{x_1}{f\lambda},\frac{y_1}{f\lambda}\right)
\end{aligned}$$

where $t \longleftrightarrow T$ form a two-dimensional Fourier transform pair. This configuration allows us to achieve Fraunhofer diffraction conditions with relatively small object-to-screen distances.

7.7.2 Transparency placed in front of lens

Now consider the transparency with amplitude transmission function $t(x,y)$ at a distance d_1 in front of a thin convex lens of focal length f. If we place the screen at distance d_2 behind the lens, the block diagram for this configuration is



From which we find

$$\begin{aligned}
g(x_2,y_2) &= -\frac{1}{\lambda^2 d_1 d_2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left[j\frac{k\left((x_2-x_1)^2+(y_2-y_1)^2\right)}{2d_2}\right]\exp\left[j\frac{k\left(x_1^2+y_1^2\right)}{2f}\right]\times \\
&\quad \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} t(x_0,y_0)\exp\left[\frac{jk}{2d_1}\left((x_1-x_0)^2+(y_1-y_0)^2\right)\right] dx_0\, dy_0\right) dx_1\, dy_1 \\
&= -\frac{1}{\lambda^2 d_1 d_2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} t(x_0,y_0)\exp\left[\frac{jk}{2}\left\{\left(\frac{x_0^2+y_0^2}{d_1}\right)+\left(\frac{x_2^2+y_2^2}{d_2}\right)+\right.\right. \\
&\quad \left.\left.\left(\frac{1}{d_2}+\frac{1}{d_1}-\frac{1}{f}\right)\left(x_1^2+y_1^2\right)-2\left(\frac{x_1x_2+y_1y_2}{d_2}\right)-2\left(\frac{x_1x_0+y_1y_0}{d_1}\right)\right\}\right] dx_0\, dy_0 dx_1\, dy_1
\end{aligned}$$

Two interesting cases arise

1. $d_1 = d_2 = f$.

   In this situation we can carry out the integrals over x_1 and y_1 to obtain

   $$g(x_2,y_2) = \frac{1}{j\lambda f}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} t(x_0,y_0)\exp\left[-j2\pi\left(\frac{x_0x_2+y_0y_2}{f\lambda}\right)\right] dx_0\, dy_0 \qquad (7.36)$$

   This is an exact two-dimensional Fourier transform.

2. $1/d_1 + 1/d_2 = 1/f$.

   Again we can carry out the integrations to obtain

$$g(x_2, y_2) = \frac{1}{\mu} t\left(\frac{x_2}{\mu}, \frac{y_2}{\mu}\right) \exp\left[\frac{\mathrm{j}k}{2d_2}\left(x_2^2 + y_2^2\right)\left(\frac{\mu - 1}{\mu}\right)\right] \tag{7.37}$$

   where $\mu = -d_2/d_1$ is the magnification. Hence

$$|g(x_2, y_2)|^2 = \frac{1}{\mu^2}\left|t\left(\frac{x_2}{\mu}, \frac{y_2}{\mu}\right)\right|^2 \tag{7.38}$$

The distribution of intensity on the screen is a magnified version of that of the transparency function. Since μ is negative (if $d_1 > 0$ and $d_2 > 0$), the image is inverted and the sense of the coordinate system is reversed. The leading term $1/\mu^2$ indicates that the intensity of the image is inversely proportional to the square of the magnification.

Exercise: Carry out the manipulations which lead to the above results from the general expression for $g(x_2, y_2)$.

7.7.3 Vignetting

Since a real lens has finite size, instead of multiplying the amplitude by $\exp[-jk(x^2 + y^2)/(2f)]$, it multiplies by $p(x, y)\exp[-jk(x^2+y^2)/(2f)]$ where $p(x, y)$ is called the *pupil function.* This function is unity within the lens and zero outside.

Exercise: In the imaging configuration $(1/d_1 + 1/d_2 = 1/f)$ show that if the pupil function is not too small,

$$g(x_2, y_2) \approx -\frac{1}{\lambda^2 d_1 d_2}(t * h)\left(\frac{x_2}{\mu}, \frac{y_2}{\mu}\right) \exp\left[\mathrm{j}\frac{k}{2d_2}\left(x_2^2 + y_2^2\right)\left(\frac{\mu - 1}{\mu}\right)\right] \tag{7.39}$$

where

$$h(x, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x_1, y_1) \exp\left[\mathrm{j}2\pi\left(\frac{xx_1 + yy_1}{d_1\lambda}\right)\right] \mathrm{d}x_1\,\mathrm{d}y_1 \tag{7.40}$$

Thus the lens smears each pixel in the original object into a region (called the *point-spread function*) whose shape depends on the Fourier transform of the pupil function. For a large pupil, $h(x, y) = d_1^2\lambda^2\delta(x)\delta(y)$ and so this simplifies to the previous result.

7.8 Gaussian Beams

Fresnel diffraction integrals are often difficult to solve analytically. However an important special case is when $f_z(x, y)$ is Gaussian in x and y – these are called Gaussian beams.

Suppose that at $z = 0$, we have a Gaussian distribution of amplitude

$$f_0(x, y) = A_0 \exp[-B_0(x^2 + y^2)] \tag{7.41}$$

To find the amplitude at $z = z_1$, we convolve the amplitude with $h(x, y)$, the Green's function for free propagation, i.e.,

$$f_1(x, y) = (f_0 * h)(x, y) = f_0(x, y) * \frac{1}{\mathrm{j}\lambda z_1}\exp\left[\mathrm{j}k\left(\frac{x^2 + y^2}{2z_1}\right)\right] \tag{7.42}$$

This may be evaluated using Fourier transforms. We see that

$$F_0(u,v) = \frac{\pi A_0}{B_0}\exp\left[-\frac{\pi^2\left(u^2+v^2\right)}{B_0}\right] \tag{7.43}$$

$$H(u,v) = \exp\left[-\mathrm{j}\frac{2\pi^2 z_1}{k}\left(u^2+v^2\right)\right] \tag{7.44}$$

Hence multiplying these together,

$$F_1(u,v) = \frac{\pi A_0}{B_0}\exp\left[-\frac{\pi^2\left(u^2+v^2\right)}{B_0}\left(1+\frac{\mathrm{j}2B_0 z_1}{k}\right)\right] \tag{7.45}$$

This can be written as

$$F_1(u,v) = \frac{\pi A(z_1)}{B(z_1)}\exp\left[-\frac{\pi^2\left(u^2+v^2\right)}{B(z_1)}\right] \tag{7.46}$$

whose inverse Fourier transform (by analogy with (7.41) and (7.43)) is seen to be

$$f_1(x,y) = A(z_1)\exp[-B(z_1)(x^2+y^2)] \tag{7.47}$$

where we see that

$$B(z_1) = B_0\left(1+\mathrm{j}\frac{2B_0 z_1}{k}\right)^{-1} \tag{7.48}$$

$$A(z_1) = A_0\left(1+\mathrm{j}\frac{2B_0 z_1}{k}\right)^{-1}$$

Thus the Gaussian beam retains its Gaussian form as it propagates. However, even if A_0 and B_0 are real, the functions $A(z_1)$ and $B(z_1)$ become complex away from $z=0$. We need to investigate the physical significance of this. The total field at z is

$$\psi(x,y,z,t) = A(z)\exp[-B(z)(x^2+y^2)]\exp(\mathrm{j}kz)\exp(-\mathrm{j}\omega t) \tag{7.49}$$

Consider the width parameter $B(z)$. We can split it into real and imaginary parts and write

$$B(z) = B_r(z) - \mathrm{j}B_i(z) \tag{7.50}$$

where the negative sign is used for convenience later. The exponentials in the expression for the total field may be written

$$\exp[-(B_r-\mathrm{j}B_i)(x^2+y^2)]\exp(\mathrm{j}kz) = \exp[-B_r(x^2+y^2)]\exp[\mathrm{j}(kz+B_i(x^2+y^2))] \tag{7.51}$$

From which it is clear that

1. The 1/e amplitude points of the Gaussian are at $x^2+y^2 = B_r(z)^{-1}$. We thus define the *half-width* of the beam to be

$$w(z) = \frac{1}{\sqrt{B_r(z)}} \tag{7.52}$$

2. The surfaces of constant phase are where $kz+B_i(x^2+y^2)$ is a constant, or $z=$ constant $-B_i(x^2+y^2)/k$. These wavefronts are curved, indicating that the beam is converging or diverging. The *radius of curvature* of the wavefront is

$$R(z) = \frac{k}{2B_i(z)} \tag{7.53}$$

If $R>0$, this indicates that the beam is diverging, whereas if $R<0$, this indicates that the beam is converging.

7.8.1 Gaussian beams in a homogeneous medium

From the way in which $B(z)$ changes with z as the beam propagates, it is possible to determine how the halfwidth $w(z)$ and radius of curvature of the wavefronts $R(z)$ change. Given that

$$B(z) = B_0\left(1 + \mathrm{j}\frac{2B_0 z}{k}\right)^{-1}$$
$$= B_0\left(1 - \mathrm{j}\frac{2B_0 z}{k}\right)\Big/\left(1 + \frac{4B_0^2 z^2}{k^2}\right)$$

we see from the real part that

$$\frac{1}{w(z)^2} = B_r(z) = \frac{B_0}{1 + 4B_0^2 z^2/k^2} = \frac{1/w_0^2}{1 + (z/z_0)^2} \tag{7.54}$$

or

$$w(z)^2 = w_0^2\left[1 + \left(\frac{z}{z_0}\right)^2\right] \tag{7.55}$$

where $z_0 = \frac{1}{2}kw_0^2$. This gives the evolution of the halfwidth which is readily seen to form a hyperbola. The *waist* of the beam is at the position where $w(z)$ is a minimum. This occurs when $B(z)$ is *purely real*. In the example, this occurs at $z = 0$ and the halfwidth of the waist is w_0.

Similarly, from the imaginary part,

$$R(z) = \frac{k}{2B_i(z)} = \frac{k}{2}\left[\frac{1 + 4B_0^2 z^2/k^2}{2B_0^2 z/k}\right]$$
$$= z_0\left[\left(\frac{z_0}{z}\right) + \left(\frac{z}{z_0}\right)\right] \tag{7.56}$$

This is the evolution of the radius of curvature of the wavefronts. At the waist (where $B(z)$ is real), the radius of curvature is infinite, indicating that the wavefronts are planes. Far away from the waist where $z \gg z_0$, we find that $R(z) \approx z$. Thus the wavefronts become approximately spherical far away from the waist and are centred about the waist. The asymptotic slope of the $w(z)$ graph is w_0/z_0 or $2/(kw_0)$. The semi-divergence angle of the beam is $\theta_0 = \tan^{-1}(2/kw_0)$. The minimum radius of curvature of the wavefronts occurs at $z = z_0$ at which $R(z_0) = 2z_0$ and $w(z_0) = \sqrt{2}\,w_0$.

The evolution of the factor $A(z)$ gives the amplitude and phase shift as the Gaussian beam propagates.

$$A(z) = A_0\left(1 + \mathrm{j}\frac{2B_0 z}{k}\right)^{-1} = A_0\left(1 + \mathrm{j}\frac{z}{z_0}\right)^{-1} \tag{7.57}$$
$$= \frac{A_0}{\sqrt{1 + (z/z_0)^2}}\angle - \tan^{-1}(z/z_0) \tag{7.58}$$

From the amplitude term, we see that there are two regimes. When $z \ll z_0$, the amplitude remains fairly constant at A_0. In this region the beam is said to be *well-collimated*. On the other hand when $z \gg z_0$, the amplitude falls as $A_0 z_0/z$ which is characteristic of spherical spreading (inverse square law in the intensity, amplitude is inversely proportional to distance). The phase shift varies from zero in the well-collimated region to $-\pi/2$ in the region of spherical spreading. As the waist becomes narrower ($w_0 \to 0$), z_0 is reduced, and θ_0 increases. We get a rapid transition into the region of spherical spreading and a rapid divergence of the beam. The phase shift of $-\pi/2$ is consistent with the leading factor of $-\mathrm{j}$ in the paraxial diffraction integral. We now see that it arises from the transition from planar wavefronts to spherical wavefronts.

7.9 Tracing Gaussian beams through optical systems

A Gaussian beam travelling through a homogeneous medium along the z direction is completely characterized by three real numbers

- The position of the waist, i.e., the value z_w at which $B(z_w)$ is real,
- The half-width of the beam at the waist, $w_0 = 1/\sqrt{B(z_w)}$,
- The amplitude of the beam at the waist, A_0.

Given these numbers, we can write down $A(z)$, $B(z)$ and the spatial dependence of the electric field for the beam.

When we trace a gaussian beam through a system, we are usually more interested in the change in the beam width parameter $B(z)$ rather than in $A(z)$. It is possible to devise a catalogue of transformations for various elements in a system. Two simple examples are:

1. Propagation through a homogeneous medium

   If the waist of the beam is at z_w,

$$B(z_1) = B_0 \left[1 + \mathrm{j}\frac{2B_0(z_1 - z_w)}{k}\right]^{-1} \tag{7.59}$$

$$B(z_2) = B_0 \left[1 + \mathrm{j}\frac{2B_0(z_2 - z_w)}{k}\right]^{-1} \tag{7.60}$$

   From these, we can eliminate both B_0 and z_w to obtain

$$\boxed{B(z_2)^{-1} = B(z_1)^{-1} + \frac{2\mathrm{j}}{k}(z_2 - z_1)} \tag{7.61}$$

   This expresses $B(z_2)$ in terms of $B(z_1)$ and may be regarded as the rule for propagation of a Gaussian beam.

2. Passage through a thin convex lens of focal length f

   The effect of the lens is to multiply $f_z(x, y)$ by $\exp[-jk(x^2+y^2)/(2f)]$. For a Gaussian beam,

$$f_z(x, y) = A(z)\exp[-B(z)(x^2 + y^2)]$$

   So after the lens, we have

$$A(z)\exp\left[-\left(B(z) + \frac{\mathrm{j}k}{2f}\right)(x^2 + y^2)\right] \tag{7.62}$$

   In terms of the width parameter, a thin convex lens of focal length f turns

$$\boxed{B(z^-) \quad \text{into} \quad B(z^+) = B(z^-) + \frac{\mathrm{j}k}{2f}} \tag{7.63}$$

   If we are interested in tracking Gaussian beams through media of different refractive indices, it is preferable to use a quantity proportional to $B(z)/k$ rather than $B(z)$. In conventional treatments of Gaussian beams, the parameter $q = -\mathrm{j}k/(2B)$ is often used. Using this notation, the above relationships for free propagation and the effect of a lens become

$$q(z_2) = q(z_1) + (z_2 - z_1) \tag{7.64}$$

and

$$\frac{1}{q(z^+)} = \frac{1}{q(z^-)} - \frac{1}{f} \quad \text{or} \quad q(z^+) = \frac{q(z^-)}{1 - q(z^-)/f} \tag{7.65}$$

These transformations of the q parameter take the form of ratios of two linear terms, namely

$$q_2 = \frac{A_1 q_1 + B_1}{C_1 q_1 + D_1} \tag{7.66}$$

If we consider two such transformations in cascade, namely the above together with

$$q_3 = \frac{A_2 q_2 + B_2}{C_2 q_2 + D_2} \tag{7.67}$$

we find that the combined transformation can be written

$$q_3 = \frac{(A_2 A_1 + B_2 C_1) q_1 + (A_2 B_1 + B_2 D_1)}{(A_1 C_2 + C_1 D_2) q_1 + (B_1 C_2 + D_1 D_2)} \tag{7.68}$$

This is also of the form of the ratio of two linear terms. The coefficients of the ratio are identical to the result of multiplying together the two-by-two matrices

$$\begin{pmatrix} A_2 & B_2 \\ C_2 & D_2 \end{pmatrix} \begin{pmatrix} A_1 & B_1 \\ C_1 & D_1 \end{pmatrix} = \begin{pmatrix} A_2 A_1 + B_2 C_1 & A_2 B_1 + B_2 D_1 \\ A_1 C_2 + C_1 D_2 & B_1 C_2 + D_1 D_2 \end{pmatrix} \tag{7.69}$$

It is thus possible to carry out the various substitutions required for tracing Gaussian beams by multiplying together matrices. These matrices are in fact identical to the $ABCD$ matrices used in geometrical optics as may be confirmed by considering the transformations for free propagation and a thin lens given above. It is thus possible to use the matrix techniques for *ray* optics and finally to interpret the result as applying to Gaussian beams which includes the effects of diffraction!

Example: A Gaussian beam with half-width w_0 at its waist is incident at its waist on a thin convex lens with focal length f. Find the position and halfwidth of the waist of the output beam.

We simply trace the Gaussian beam through each element in turn, remembering that the waist always occurs when B is real. Just before the convex lens, we have that $B = 1/w_0^2$. As a result of passing through the lens,

$$B = \frac{1}{w_0^2} + \mathrm{j}\frac{k}{2f} \tag{7.70}$$

Propagating this through free space for a further distance z yields

$$[B(z)]^{-1} = \left[\frac{1}{w_0^2} + \mathrm{j}\frac{k}{2f}\right]^{-1} + \frac{2\mathrm{j}z}{k} \tag{7.71}$$

The waist occurs where $B(z_w)$ is real. i.e.,

$$-\mathrm{j}\frac{k/(2f)}{w_0^{-4} + k^2/(4f^2)} + 2\mathrm{j}\frac{z_w}{k} = 0 \tag{7.72}$$

or

$$z_w = \frac{k^2 f}{k^2 + 4f^2/w_0^4} = \frac{f}{1 + (f/z_0)^2} \tag{7.73}$$

The half-width of the beam at the waist is

$$[B(z_w)]^{-1/2} = \frac{2f/(kw_0)}{\sqrt{1 + (f/z_0)^2}} \tag{7.74}$$

We see that for a small focused spot, we need to use a small value of f and a large value of w_0.

7.10 One-dimensional propagation in a dispersive medium

Consider a one-dimensional dispersive medium in which the wave velocity c depends on the wave frequency ν. At the position $x = 0$, we set up a disturbance $f(0, t)$ within the medium and wish to consider how this disturbance propagates through the medium in the positive x direction. The disturbance in the the medium (for $x > 0$) is then denoted $f(x, t)$.

The function $f(0, t)$ may be written as a superposition of complex exponentials in terms of its Fourier transform $F(\nu)$. We suppose that

$$f(0,t) = \int_{-\infty}^{\infty} F(\nu) \exp(\mathrm{j}2\pi\nu t)\, d\nu \tag{7.75}$$

Each of the complex exponential components has a well-defined frequency and propagates within the medium at a velocity appropriate to that component. Thus the component which varies as $\exp(\mathrm{j}2\pi\nu t)$ at $x = 0$ propagates towards positive x as the wave $\exp\left[\mathrm{j}2\pi(\nu t - x/\lambda)\right]$.

By linearity, we may write down an expression for the total disturbance

$$f(x,t) = \int_{-\infty}^{\infty} F(\nu) \exp\left[\mathrm{j}2\pi\left(\nu t - \frac{x}{\lambda}\right)\right] \mathrm{d}\nu \tag{7.76}$$

$$= \int_{-\infty}^{\infty} F(\nu) \exp(j2\pi\nu t) \exp\left(-\frac{\mathrm{j}2\pi x}{\lambda}\right) \mathrm{d}\nu \tag{7.77}$$

which is just the inverse Fourier transform relationship with each complex exponential replaced by the contribution that this component would make throughout the medium.

We now consider the function $\lambda^{-1} = \nu/c(\nu)$ which appears in the exponential above. In a non-dispersive medium, we expect this simply to be proportional to ν. In a dispersive medium however, the functional dependence will be more complicated. If the function $f(0, t)$ represents a *narrowband* process containing only frequencies close to ν_0, we can expand $\lambda^{-1} = \nu/c(\nu)$ in a Taylor series around this centre frequency

$$\frac{1}{\lambda} = \frac{\nu}{c\,(\nu)} = a + b(\nu - \nu_0) + d(\nu - \nu_0)^2 + ... \tag{7.78}$$

In the regime of *linear dispersion* we consider only the first two terms. Substituting this into (7.77) we find

$$f\,(x,t) = \int_{-\infty}^{\infty} F\,(\nu) \exp\left(\mathrm{j}2\pi\nu t\right) \exp\left(-\mathrm{j}2\pi\left[a + b\,(\nu - \nu_0)\right] x\right)\, \mathrm{d}\nu \tag{7.79}$$

$$= \exp\left(-\mathrm{j}2\pi\left[a - b\nu_0\right] x\right) \int_{-\infty}^{\infty} F\,(\nu) \exp\left(\mathrm{j}2\pi\nu\left[t - bx\right]\right)\, \mathrm{d}\nu \tag{7.80}$$

$$= \exp\left(-\mathrm{j}2\pi\left[a - b\nu_0\right] x\right) f\,(0, t - bx) \tag{7.81}$$

where we have used the fact that $f(0, t) \leftrightarrow F(\nu)$ in the last equality. In order to interpret this result physically, let us consider a prototypical narrow-band process around ν_0. This has the form

$$f(0,t) = m(t) \exp(\mathrm{j}2\pi\nu_0 t) \tag{7.82}$$

where $m(t)$ is the baseband complex envelope which has a small bandwidth. The multiplication by $\exp(\mathrm{j}2\pi\nu_0 t)$ shifts the spectrum up to be around ν_0.

(*Note:* In this analysis we work with a complex valued $f(0,t)$ so that we can deal with a narrowband process around the single frequency ν_0. In practice, for a real signal, we need to consider frequencies around both ν_0 and $-\nu_0$. The Taylor series expansion discussed above would only work around ν_0, making it necessary to write another expansion around $-\nu_0$. By throwing away the negative frequencies and working with the analytic signal, we need only consider a single Taylor expansion.)

Substituting (7.82) into (7.79) we find

$$f(x,t) = \exp(-\mathrm{j}2\pi[a - b\nu_0]x) m(t-bx) \exp(\mathrm{j}2\pi\nu_0[t-bx]) = m(t-bx)\exp(\mathrm{j}2\pi[\nu_0 t - ax]) \quad (7.83)$$

We see that if we observe the result at position x, the envelope $m(t)$ is delayed (relative to the initial excitation) by bx. The velocity at which the envelope moves through the medium is thus $1/b$. This is called the *group velocity*, denoted c_g.

On the other hand, if we consider the high frequency oscillations beneath the envelope, these propagate at ν_0/a. This is called the *phase velocity*, denoted c_p. Unless $\nu_0/a = 1/b$, the phase velocity and group velocity will be different. Referring back to (7.78), we can write down expressions for a and b in terms of derivatives of $\nu/c(\nu)$ at ν_0. In particular,

$$a = \frac{\nu_0}{c(\nu_0)} \quad (7.84)$$

$$b = \frac{\mathrm{d}}{\mathrm{d}\nu}\left(\frac{\nu}{c(\nu)}\right)\bigg|_{\nu=\nu_0} = \frac{\mathrm{d}}{\mathrm{d}\nu}\left(\frac{1}{\lambda}\right) = \frac{\mathrm{d}k}{\mathrm{d}\omega}$$

where $\omega = 2\pi\nu$ and $k = 2\pi/\lambda$. The phase and group velocities are thus given by

$$c_p = c(\nu_0) \qquad \text{and} \qquad c_g = \frac{\mathrm{d}\omega}{\mathrm{d}k} \quad (7.85)$$

Exercise: Show that

$$c_g = c - \lambda\frac{\mathrm{d}c}{\mathrm{d}\lambda} \quad (7.86)$$

Exercise: Calculate the phase velocity and group velocity (as a function of the particle momentum) for the de Broglie wave of a non-relativistic particle of mass m and momentum p which satisfies the relationships

$$p = \hbar k, \qquad E = \hbar\omega \qquad \text{and} \qquad E = \frac{p^2}{2m} \quad (7.87)$$

How do these relate to the classical particle velocity?

Exercise: If we include the quadratic term in equation (7.78) show that the effect of propagating the narrowband signal $m(t)\exp(\mathrm{j}2\pi\nu_0 t)$ through a distance x is to give

$$f(x,t) = (m * q)(t-bx)\exp(\mathrm{j}2\pi[\nu_0 t - ax]) \quad (7.88)$$

where

$$q(t) = \frac{1}{\sqrt{2\mathrm{j}xd}}\exp\left(\frac{\mathrm{j}\pi t^2}{2xd}\right) \quad (7.89)$$

This is called *quadratic dispersion*. Analytical results are available if $m(t)$ is a Gaussian pulse so that the convolution with q is tractable.