The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

2223-1

Winter College on Optics in Imaging Science

31 January - 11 February, 2011

Digital Holographic Microscopy

D. Kelly
Technische Universität Ilmenau
Germany

Winter College on Optics in Imaging Science

Damien P. Kelly

Feb 2011

Course Outline

- Introduction to holography, wave theory, and diffraction
- Fourier optics view and Digital holography - numerical propagation as an optical system
- Numerical algorithms: Spectral and Direct methods
- Extra topic: Imaging performance of DH systems

Maxwell's equations

To begin we go back to the late 1800s - the time of Maxwell. At this time the phenomena of electricity and magnetism had been investigated by a large number of great scientists and a series of important laws had been discovered, for example Faraday's Law of Induction. Other parallel developments were occurring in optics, the wave theory of light had overcome major obstacles to its acceptance - rectilinear propagation and polarization (calcite crystals) - and significant efforts were made to determine the speed of light from experiment, Fizeau (1849) estimated it to be $c \approx 315\ 000$ km/s [1].

The achievement of Maxwell was to condense empirical knowledge related to electricity and magnetism into a single set of mathematical equations now referred to as Maxwell's equations. He demonstrated that an electromagnetic field could propagate through the aether as a transverse wave using purely theoretical arguments. Solving for the speed of the resulting wave, led to an expression in terms of the electric (ϵ_0) and magnetic (μ_0) constants related to the properties of the medium. These constants were known electrical measurements and yielded a value of $c = 1/\sqrt{\epsilon_0 \mu_0}$. The conclusion was that electricity, magnetism and light were no longer separate problems in physics but rather were intimately interlinked and could be explained using this new framework. At this point we introduce the four equations in derivate form: Gauss's Law

$$\nabla . B = 0 \tag{1}$$

Faraday's Law

$$\nabla \times E = -\frac{\partial B}{\partial t} \tag{2}$$

[1] for further reading try [1, 2, 3]

Columb's Law (can be derived from this)

$$\nabla . D = \rho \tag{3}$$

Ampere and Maxwell's Law

$$\nabla \times B = \mu \left(J + \epsilon \frac{\partial E}{\partial t} \right) \tag{4}$$

For completeness the set of laws governing EM fields we include the Lorentz force law,
Lorentz force Law

$$F = qE + vq \times B. \tag{5}$$

For an uncharged and non-conducting medium (free-space propagation) this equations can be reduced to

$$\nabla^2 E = \mu_0 \epsilon_0 \frac{\partial^2 E}{\partial t^2}. \tag{6}$$

Note that this is a wave equation that has several solutions which we will come to in due course. To highlight that light is vectorial in nature we note that E actually has components (in a Cartesian coordinate system along the x, y and z axes), hence $E = \left\langle E_x \hat{\imath}, E_y \hat{\jmath}, E_z \hat{k} \right\rangle$. Each of these vectorial components must separately satisfy Eq. (6).

Scalar theory

In this course we make a major simplifying assumption: we use scalar diffraction theory which allows us to write to express the actions of electromagnetic fields in terms of a scalar quantity, $u(P,t)$:

$$\nabla^2 u(P,t) = \mu_0\epsilon_0 \frac{\partial^2 u(P,t)}{\partial t^2}. \tag{7}$$

Our scalar quantity is a function of both time, t, and space where P is a spatial coordinate (x, y, z). This transition to a scalar theory will greatly simplify our analysis. It also provides a valid description of the behavior of light for a large range of situations. When is it not valid[2]?

- Diffraction from apertures when the aperture size is comparable to the wavelength,
- Propagation through in-homogenous media,
- Focusing by high numerical apertures,
- Anisotropic media including some crystals.

Eq. (7) is a wave equation and a particular solution for mono-chromatic light (light of one color) is given by

$$U(P,t) = A(P)\cos\left[2\pi vt + \phi(P)\right] \tag{8}$$

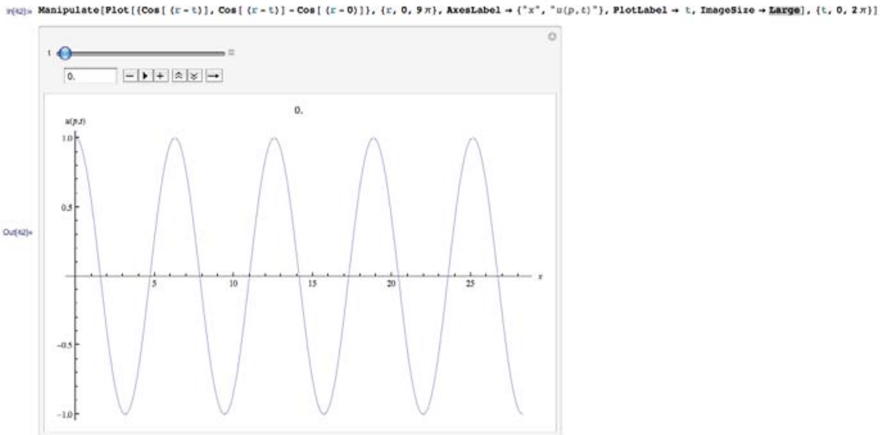
where v is the optical frequency of the electromagnetic wave.

Because it will be helpful later we are going to take a little time now to examine the expression in Eq. (8) in more detail. Setting $A(P) = 1$, allowing our space variable, P, to vary as a function of x only and introducing the wavevector $\vec{k}$ ($k = |\vec{k}| = 2\pi/\lambda$, the optical wavelength), Eq. (8) now becomes

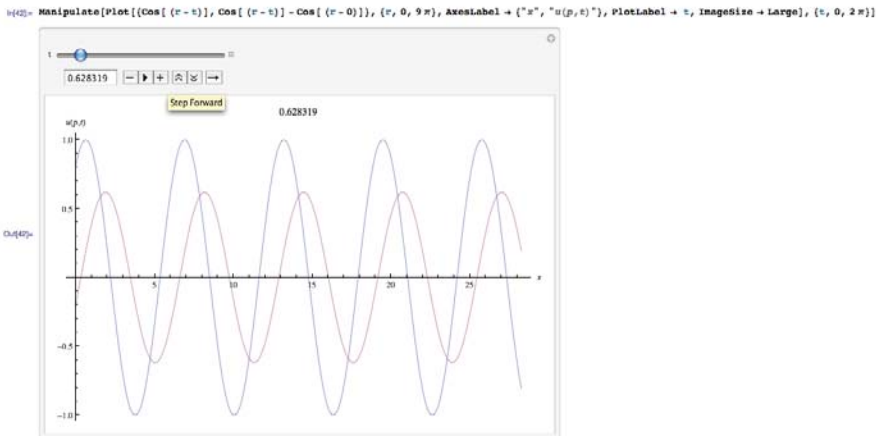
$$U(P,t) = \cos\left(2\pi vt + kx\right). \tag{9}$$

[2]The validity of this approximation has been looked at in great detail. For further reading see: Chapter 3 of [4] and Chapter 8 of [2], and [5, 6]

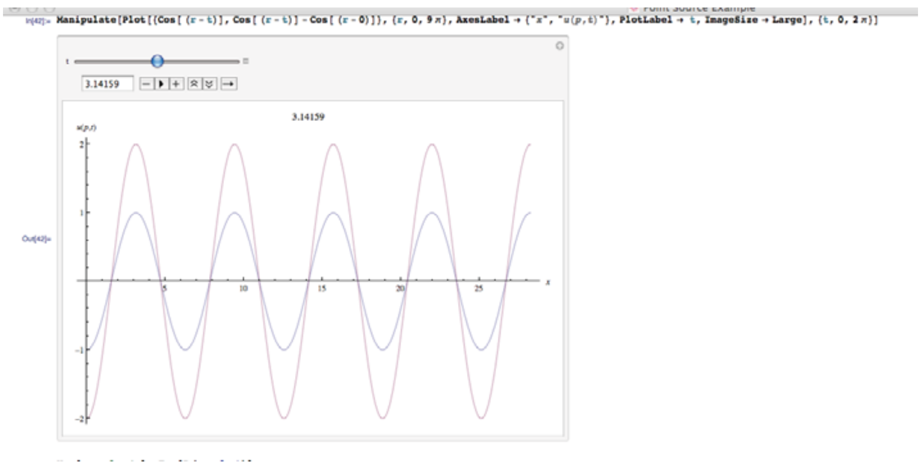
Exercise: Show that Eq. (9) is indeed a solution to Eq. (7)!



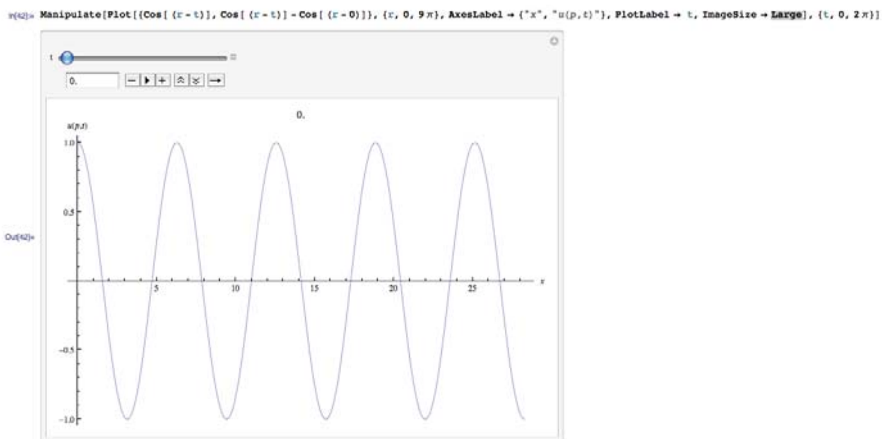
t=0-1.pdf



t = 063.pdf



t =pi.pdf



t=0-1.pdf

Figures showing dependence on time and space

We have just looked at how Eq. (9) varied as a function of time and space. It may be expressed in more convenient notation which we now introduce

$$U(P,t) = \Re\left\{U(P)\exp(-j2\pi vt)\right\}, \tag{10}$$

where $\Re$ signifies the 'real part of' and $U(P)$ is a complex scalar function of position. Note that we have assumed a simple harmonic oscillator, $\exp(j2\pi vt)$, and separated the space and the time variables. This will allow us to examine the spatial variation of our complex wavefield $U(P)$ in isolation. We make several remarks about this step:

- Since we know how the phase will change as a function of time we can include this effect at a later stage. However due to the extremely rapid variation of this field normally only averaged intensity values are considered.
- We are concerned primarily with mono-chromatic light, however light fields with more complicated spectral distributions can be accounted for by noting that any waveform can be expressed by a linear combination of sinusoidal components. The same is true of $U(P)$!

Hence from here we will consider only the scalar field that is a complex function of position[3]. The function, $U(P)$, satisfies the Helmholtz equation:

$$\left(\nabla^2 + k^2\right)U = 0 \tag{11}$$

For the special case of plane wave propagation the function, $U(P)$ takes the form

$$\begin{aligned} U(x,y,z) &= \exp(j\vec{k}.P) \\ &= \exp\left[j\frac{2\pi}{\lambda}\left(\alpha x + \beta y + \gamma z\right)\right] \end{aligned} \tag{12}$$

[3]For further reading see Chapter 3 of [4] and [2]

where the directional cosines α, β, and γ are related through

$$\gamma = \sqrt{1 - \alpha^2 + \beta^2} \tag{13}$$

Thus we see that a function $\exp\left[j2\pi(f_x x + f_y y)\right]$ can be interpreted as a plane wave at $z = 0$, propagating at an angles $(\alpha = \lambda f_x, \beta = \lambda f_y)$. This last observation will be important when we start employing Fourier optics tools to examine how different and arbitrary wavefields propagate. We note in advance that by using Fourier tools we can decompose any physically realizable signal into a weighted superposition of sinusoidal terms, each of which can then be considered as propagating plane waves.

Spherical waves

We saw from the previous sections that waves can add constructively or destructively giving rise to the effect known as interference. (It was not always accepted that the addition of light could result in localized darkness). We want to look at this again but this time for a different but equally important function, the spherical wave,

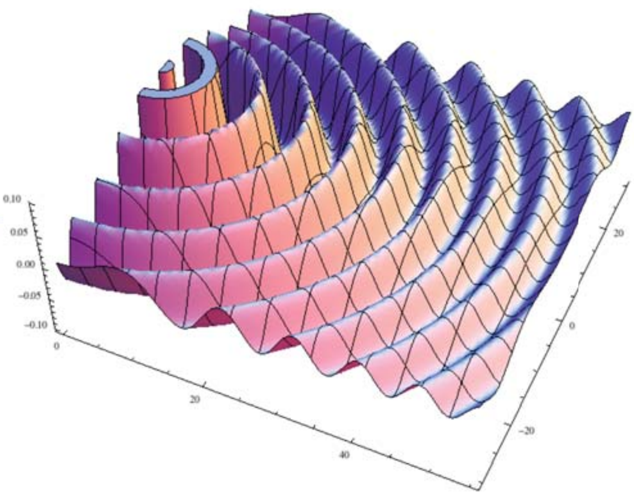
$$U(P) = \frac{1}{r}\exp\left(jkr\right). \tag{14}$$

We see a plot of this figure below. There are several features to note, 'wavefronts' of constant amplitude and phase propagate out from the source (located at $r = 0$) and that the amplitude drops off as a function of distance, r. We proceed by writing Eq. (14) in terms of x, y, z to give

$$U(P) = \frac{1}{z\sqrt{1 + \frac{(x-x_0)^2}{z^2} + \frac{(y-y_0)^2}{z^2}}}\exp\left(jkz\sqrt{1 + \frac{(x-x_0)^2}{z^2} + \frac{(y-y_0)^2}{z^2}}\right), \tag{15}$$

```
In[80]:= Plot3D[Re[Ups[√(z^2 + (x)^2), 0, 0]], {z, 0.1, 55}, {x, -30, 30}, PlotRange → Automatic, PlotPoints → 90, Boxed → False]
```

Out[80]=



where we have used the relation $r = \sqrt{z^2 + (x - x_0)^2 + (y - y_0)^2}$ and assumed that the point source is located at an origin given by $P_0 = (x_0, y_0, z_0 = 0)$.

Diffraction by two pinholes

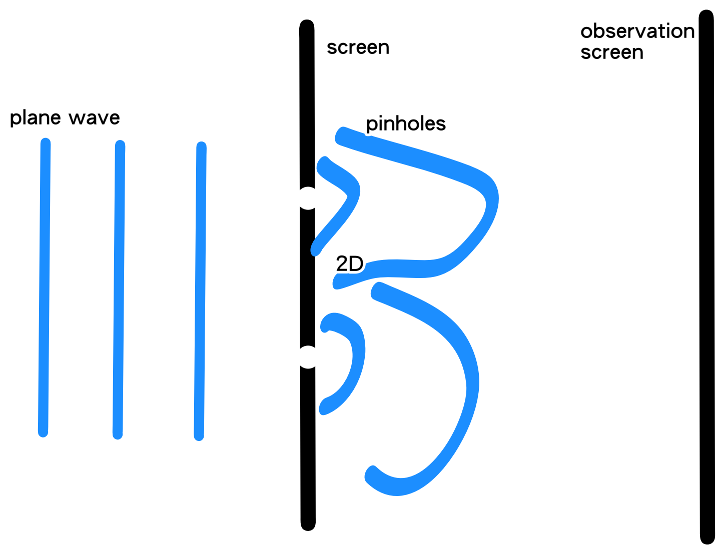
In this lecture i think it is most important to try and explain the underlying reasoning that leads to the mathematical formulation of diffraction integrals. To do this we are for the moment going to sacrifice some mathematical rigor so that the fundamental underlying causes of diffraction can be described. I want you to understand from an intuitive physical perspective the process of diffraction and so we postpone a more rigorous mathematical treatment until the end of lecture 2. We proceed by getting rid of the cumbersome square root operation in Eq. (15) we shall use the binomial expansion

$$\sqrt{1+b} = 1 + \frac{1}{2}b - \frac{1}{8}b^2 + \ldots$$

to approximate it. Ideally we wish to retain as few of the binomial expansion terms as possible while still retaining a reasonable accuracy. Due to the large value of k at optical wavelengths we must retain the second expansion term [4, 6]

$$U(P) = C \exp\left[\frac{jk}{2z}\left[(x-x_0)^2 + (y-y_0)^2\right)\right]. \tag{16}$$

Since the more interesting term in Eq. (15) is the exponential we are going for now to ignore the scaling factor $1/r$ replacing it with a complex constant C (*<u>note</u> this is where we are being somewhat careless with our treatment, although the form of the result is the same we are jumping over some fundamental considerations*). Here again, and in preparation for a major step, we will examine the interference between two spherical wave sources. Consider the following situation depicted in Fig. 3 (for simplicity we shall ignore the y-component). Two coherent point sources are located symmetrically about



the optical axis at a distance $x = \pm D$. The resulting field at a distance

$z = z_P$ can be calculated using Eq. (16) to give

$$\begin{aligned} U(D, z_P) + U(-D, z_P) &= C \exp\left\{\frac{jk}{2z_P}\left[(x-D)^2\right]\right\} \\ &\quad + C \exp\left\{\frac{jk}{2z_P}\left[(x+D)^2\right]\right\} \\ &= C \exp\left[\frac{jk\left(D^2+x^2\right)}{2z_P}\right] \cos\left(\frac{k}{z_P}Dx\right) \end{aligned} \tag{17}$$

Looking at Eq. (17) we note that the apart from a complex constant we have both a quadratic phase factor and a cosine term. Optical detectors are generally sensitive only to the intensity (magnitude squared) of optical field and in practice what is observed is a set of fringes of a single spatial frequency $D/(\lambda z)$. Thus as we vary the distance between the contributing point sources D we observe a different frequency, similarly for changes in both λ and z. For an idea of what this looks like one is referred to B. J. Thompson and E. Wolf - here the authors experimentally measure the spatial coherence of 2 displaced point sources - which has a mathematical form similar to Eq. (17) although the effect they are examining is quite different.

Diffraction

We are now nearly in a position to provide a qualitative description of how light propagates and is diffracted by different apertures. We must first consider another fundamental concept introduced Huygens circa 1678 [4]. Huygens was one of the first scientists to suppose that the behavior of light could be attributed to a wave nature. He expressed an intuitive conviction that if each point on the surface of a wavefront could be considered as a 'secondary' spherical source (disturbance) then the wavefront at a later instant could be found by constructing the envelope of the secondary wavelets, see Chapter 3 Ref. [4].

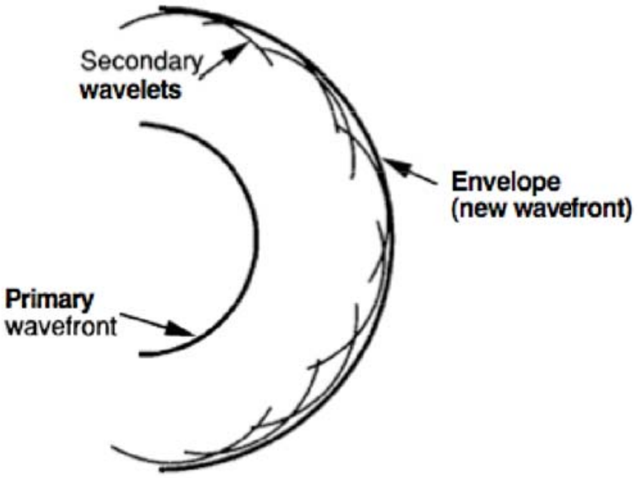


FIGURE 3.4
Huygens' envelope construction.

Figure 1: Taken from Goodman

The Fresnel transform

In the previous sections we have looked at how the problem of diffraction (was) could be attacked (under paraxial conditions) by making a series of assumptions: read → {Interference, Huygens secondary sources, Fresnel and Young}, using a combination of intuitive physical ideas and the resulting mathematical formulation. Now that we understand the underlying operating principle of the Fresnel transform we will accept the formal definition:

$$u_z\left(x, y\right) = \frac{1}{j\lambda z} \int_{\infty}^{\infty} u\left(X, Y\right) \exp\left\{ \frac{j\pi}{\lambda z} \left[\left(x - X\right)^2 + \left(y - Y\right)^2 \right] \right\} dX dY, \tag{18}$$

and use it to examine the diffraction from several different types of apertures. Lets quickly review the main features of Eq. (18):

- it relates the field at a plane: (X, Y) to the field at a different plane: (x, y) located a distance z away,
- the integral limits extend $\pm\infty$,
- a definite scaling factor $1/(j\lambda z)$ has been introduced in place of C

To begin let us consider some examples starting with our old example of the diffraction of two point sources. Previously we had cheated a little and simply wrote the mathematical expression for a spherical wave (approximated with the binominal expansion) a distance from the source location. However if we look at Eq. (14) and allow $r \to 0$ we see that the $1/r \to \infty$. In fact the expression reduces (at $r = 0$) or is defined as a Dirac delta function or impulse function:

$$\begin{array}{ll} \delta(x) = 0, & \text{if } x \neq 0 \\ \int_{\infty}^{\infty} \delta(x) dx = 1 & \end{array} \tag{19}$$

I would recommend reading Chapter 5 of Ref. [7] for more detail on this function. *Indeed in the next lecture we will be looking more closely at different functions that are important in Fourier analysis, and examine whether they are well behaved functions.* With the help of Eq. (19) we find that our input wavefield (1-D) is

$$u(X) = \delta(X - D) + \delta(X + D),$$

which on substitution into Eq. (18) yields

$$u_z(x) = \frac{1}{\sqrt{j\lambda z}} \exp\left[\frac{jk\left(D^2 + x^2\right)}{2z_P}\right] \cos\left(\frac{k}{z_P} Dx\right). \tag{20}$$

<u>Note:</u> that for 1-D analysis the $1/(j\lambda z)$ term becomes $1/\sqrt{j\lambda z}$.

Square aperture: 1-D

We have already seen how to extend our 2 pinhole example to describe diffraction from a square aperture. Now we will look at the more rigorous mathematics for that effect. We begin by writing the expression for a square aperture,

$$\begin{aligned} p_L(X) &= 1, \text{if } |X| < L \\ &= 0, \text{ otherwise.} \end{aligned} \tag{21}$$

Inserting Eq. (21) into Eq. (18)

$$\begin{aligned} u_z(x) &= \frac{1}{\sqrt{j\lambda z}} \int_{\infty}^{\infty} p_L(X) \exp\left\{\frac{j\pi}{\lambda z}\left[(x-X)^2\right]\right\} dX, \\ &= \frac{1}{\sqrt{j\lambda z}} \int_{-L}^{L} \exp\left[\frac{j\pi}{\lambda z}(x-X)^2\right] dX, \\ &= -\frac{1}{2}\left\{\operatorname{erf}\left[(-1)^{\frac{3}{4}}\sqrt{\frac{\pi}{\lambda z}}(L-x)\right] + \operatorname{erf}\left[(-1)^{\frac{3}{4}}\sqrt{\frac{\pi}{\lambda z}}(L+x)\right]\right\}, \end{aligned} \tag{22}$$

The erf functions in Eq. (22) may be calculated quickly using modern computers and so we have an analytical solution to diffraction from a square aperture.

In[7]:= $$\frac{1}{\sqrt{\mathrm{I}\,\lambda\,z}}\int_{-L}^{L}\mathrm{Exp}\left[\frac{\mathrm{I}\,\pi}{\lambda\,z}\,(x-X)^2\right]\mathrm{d}X$$

Out[7]= $$-\frac{(-1)^{1/4}\sqrt{z}\sqrt{\lambda}\left(\mathrm{Erf}\left[\frac{(-1)^{3/4}\sqrt{\pi}\,(L-x)}{\sqrt{z}\sqrt{\lambda}}\right]+\mathrm{Erf}\left[\frac{(-1)^{3/4}\sqrt{\pi}\,(L+x)}{\sqrt{z}\sqrt{\lambda}}\right]\right)}{2\sqrt{i\,z\,\lambda}}$$

In[8]:= `FullSimplify[%]`

In[9]:= $$\mathrm{uz}[x\_,\ z\_,\ \lambda\_,\ L\_] := -\frac{(-1)^{1/4}\sqrt{z}\sqrt{\lambda}\left(\mathrm{Erf}\left[\frac{(-1)^{3/4}\sqrt{\pi}\,(L-x)}{\sqrt{z}\sqrt{\lambda}}\right]+\mathrm{Erf}\left[\frac{(-1)^{3/4}\sqrt{\pi}\,(L+x)}{\sqrt{z}\sqrt{\lambda}}\right]\right)}{2\sqrt{i\,z\,\lambda}}$$

In[10]:=
```
λD = 633 × 10^-9;
zD = 1;
Ld = 1 × 10^-3;
Plot[Abs[uz[x, zD, λD, Ld]], {x, -4 Ld, 4 Ld}, PlotRange → All]
```

Out[13]=

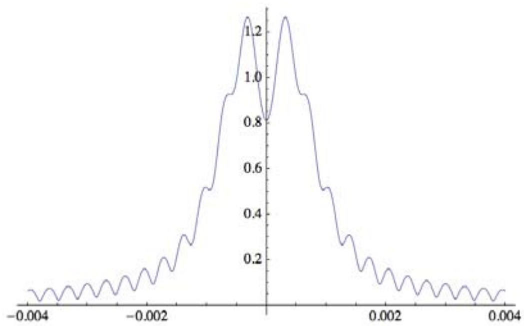


Figure 2: Calculation Example

Plane and Spherical waves

Last time we looked at two solutions to the Helmholtz equation in free space, the plane and spherical wave. Specifically we looked at the diffraction pattern formed a distance z away from a two pinhole screen. We now return to this example to remind ourselves of last lecture's discussion.

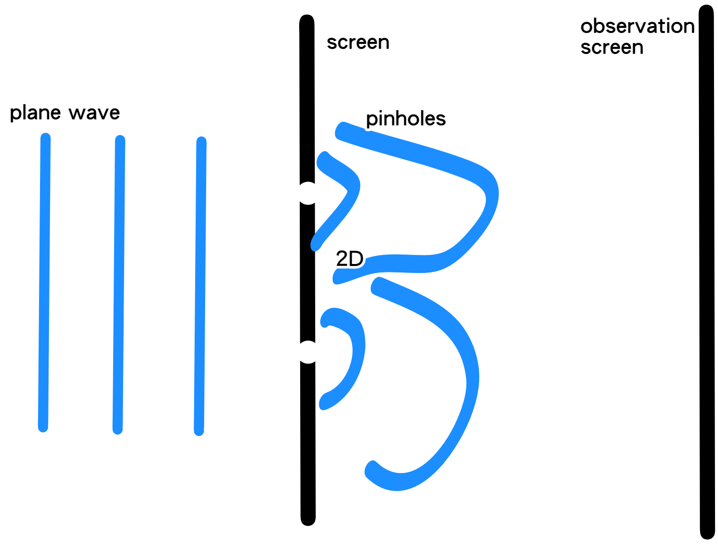


Figure 3: Two pinhole experiment.

This figure highlights two of the characteristics of the wave types we encountered last week:

Plane wave: has a uniform phase along a plane that is perpendicular to the direction of propagation.

$$\begin{aligned} U(x,y,z) &= \exp(j\vec{k}.P) \\ &= \exp\left[j\frac{2\pi}{\lambda}(\alpha x + \beta y + \gamma z)\right] \end{aligned} \tag{23}$$

In Fig. 1, the plane wave is propagating along the z axis and therefore α and β can be set to zero to give

$$U(z) = ?? \tag{24}$$

$$U(z) \quad = \quad \exp\left(j\frac{2\pi}{\lambda}z\right) \tag{25}$$

What does the pinhole screen do? It selects two specific parts of the plane wave, which propagate further as spherical waves!

Spherical waves: originate a specific spatial location; the phase is constant over a spherical surface. How can we represent that graphically? How do we write that mathematically? And in its approximate form?

$$U(x,z) = \frac{1}{z\sqrt{1+\frac{(x-x_0)^2}{z^2}}}\exp\left(jkz\sqrt{1+\frac{(x-x_0)^2}{z^2}}\right), \tag{26}$$

where we have used the relation $r = \sqrt{z^2+(x-x_0)^2}$ and assumed that the point source is located at an origin given by $P_0 = (x_0, z_0 = 0)$, which we approximated as

$$U(x,z) = C\exp\left[\frac{jk(x-x_0)^2}{2z}\right]. \tag{27}$$

What is the expression for the field a distance $z = z_P$ from the pinhole screen? Calculate!

Two coherent point sources are located symmetrically about the optical axis at a distance $x = \pm D$. The resulting field at a distance $z = z_P$ is given by

$$
\begin{aligned}
U(D, z_P) + U(-D, z_P) &= C \exp\left\{\frac{jk}{2z_P}\left[(x-D)^2\right]\right\} \\
&\quad + C \exp\left\{\frac{jk}{2z_P}\left[(x+D)^2\right]\right\} \\
&= C \exp\left[\frac{jk\left(D^2+x^2\right)}{2z_P}\right] \cos\left(\frac{k}{z_P} Dx\right)
\end{aligned}
\tag{28}
$$

The means whereby we have analyzed this classical optics problem has assumed we have perfect plane and spherical waves - an idealization - and therefore the results predicted by Eq. (28) differ from the experimental observation. However the main characteristics are broadly captured.

Thought experiment

What happens if we introduce a piece of glass before one of the pinholes as depicted in Fig. 2 ?

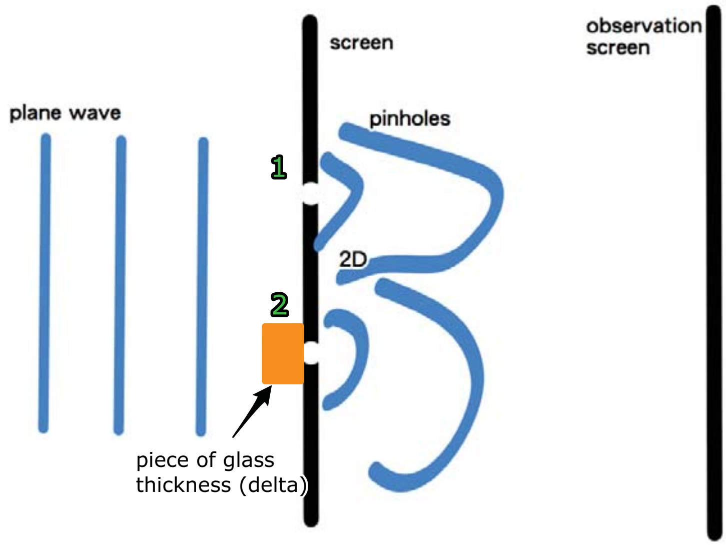


Figure 4: Two pinhole experiment with a piece of glass over one of the pinholes.

Glass has a different refractive index than free space and so a piece of glass will introduce a phase delay, ϕ, by an amount determined by both the thickness and refractive index of the glass section. We depict this graphically in Fig. 3. This phase change is expressed mathematically with the following equation:

$$\phi = kn\Delta(x) \tag{29}$$

where k is still the wavenumber $k = 2\pi/\lambda$, n is the refractive index approx 1.5 for glass, and Δ is the thickness of the glass segment (the dependence on x, indicates that this thickness could vary as a function of spatial location - like a lens!) Hence the spherical wave emerging from Location 2, see Fig. 2, has an additional phase term ϕ. How does this change things? Well we can

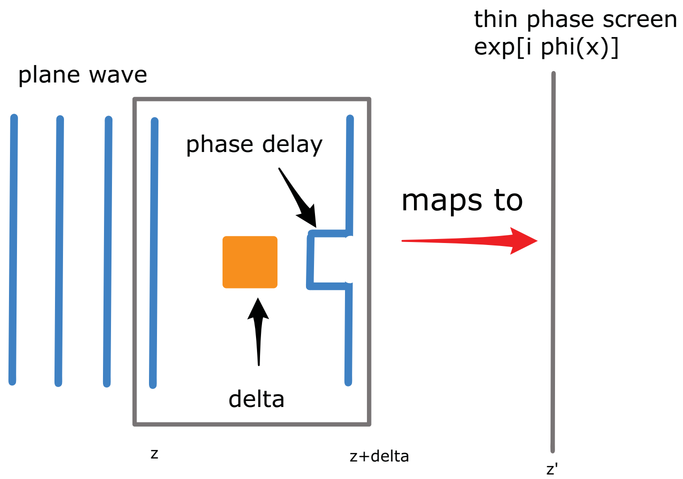


Figure 5: Two pinhole experiment with a piece of glass over one of the pinholes.

re-write our expression given in Eq. (28), to give

$$
\begin{aligned}
U(D, z_P) + U(-D, z_P) \quad &= \quad C \exp\left\{\frac{jk}{2z_P}\left[(x-D)^2\right]\right\} \\
&\quad + C \exp(j\phi) \exp\left\{\frac{jk}{2z_P}\left[(x+D)^2\right]\right\} \qquad (30)
\end{aligned}
$$

Lets see what happens, and turn to Mathematica!

More spherical sources

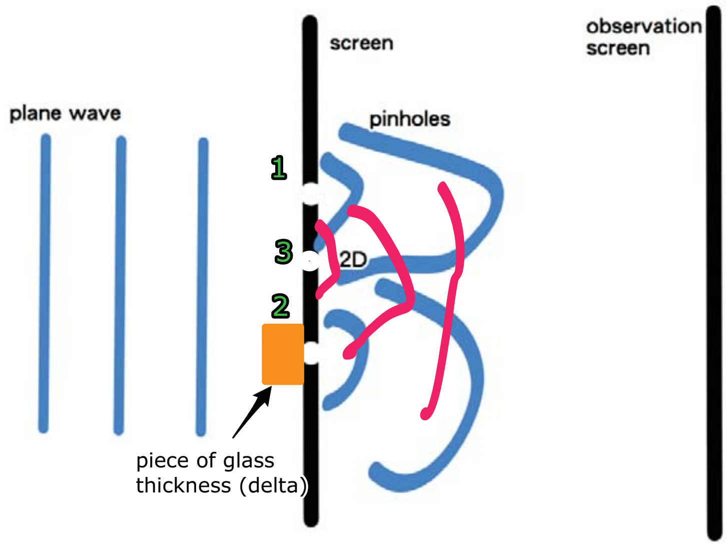


Figure 6: Three pinholes with a piece of glass over one of the pinholes.

An example calculation is shown in the following plot:

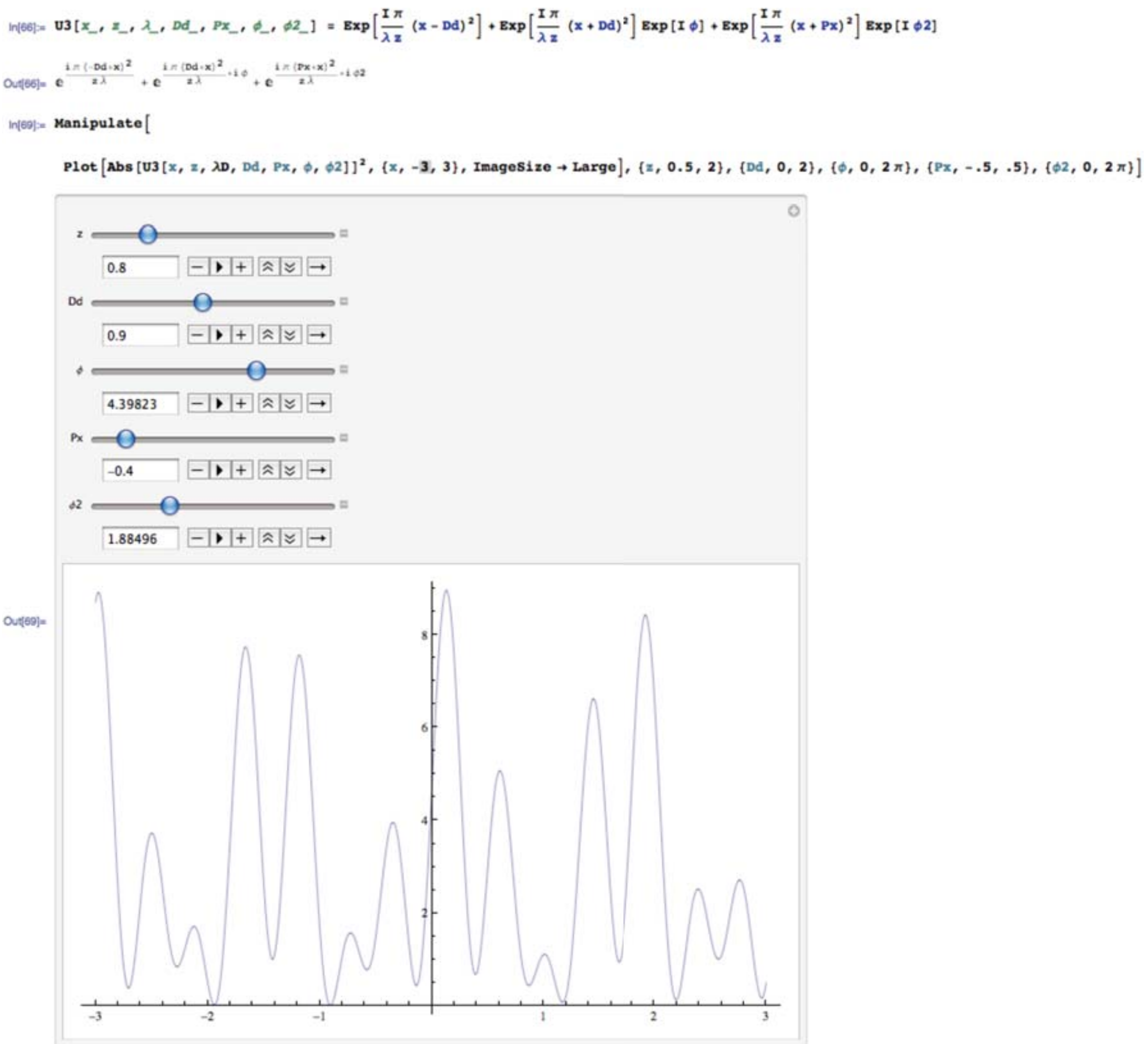


Figure 7: Example calculation in Mathematica

The **main points** that I want to emphasis in this section are the following:

1. Using a mathematical expression for a spherical wave we can describe the diffraction pattern a distance z_P away from our double pinhole grating.

2. By introducing a piece of glass over a particular pinhole, its phase contribution can be changed relative to the other pinholes.

3. If the refractive index and thickness of the glass are known, so is the phase shift that has been introduced.

4. By increasing the number of point sources we can increase the complexity of our diffraction pattern.

In a further generalization we could allow the amplitudes of the different contributing spherical waves to change, i.e. replace C with C_1, C_2 ....

A Conclusion: IF WE KNOW THE PHASE AND AMPLITUDE OF A COLLECTION POINT SOURCES IN A PARTICULAR PLANE WE CAN CALCULATE WHAT THE DIFFRACTION PATTERN WILL LOOK LIKE AT ANOTHER PLANE BY ADDING TOGETHER EACH CONTRIBUTING POINT SOURCE.

Mathematically for N spherical waves, located at spatial locations X_n and with amplitudes and phases C_n, ϕ_n:

$$\begin{aligned} U_s(x) &= |U_s(x)| \exp\left[j\phi_s(x)\right], \\ &= \sum_{n=1}^{n=N} C_n \exp\left(j\phi_n\right) \exp\left[\frac{jk}{2z}(x - X_n)^2\right]. \end{aligned} \tag{31}$$

where $\phi_s(x)$ is the phase of the output optical field.

Course Exercise

Consider the distribution formed a distance $z = 0.5$m, (let $\lambda = 1$) from a 3 pinhole diffracting screen such as that depicted in Fig. 4. Source 1 (S1) is located at $x_0 = D$, with $\phi_{S1} = 0$, S2 is at $x_0 = -D$ and phase lag $\phi_{S2} = \pi$, S3 located at $x_0 = -0.2$ ($\phi_{S3} = 0$), where $D = 0.4$, assume $C_n = 1$ for all point sources. **Write a short report that addresses the following:**

1. Outline how the calculation is performed,
2. Derive a mathematical expression in terms of $z, \lambda, D, \phi_{S1}, \phi_{S2}, \phi_{S3}$ that describes the distribution
3. Calculate and plot the distribution using the values for z, λ... etc above over the range $-3 < x < 3$.
4. Show how this distribution changes when ($\phi_{S3} = 0.63$).
5. Comment on the significance of your results.

Reports should be submitted within 2 weeks (Nov 22) and should include Matlab or Mathematica code in an Appendix. Timely completion of this report will be considered when assessing your final mark!

Intensity of an optical field.

Recording media, in general, are sensitive only to the intensity of the light field incident upon them see Chap 4 in Ref [4] for more detail. That implies that the measurable quantity of the optical field $U_s(x)$ is $I_s(x)$ given by

$$\begin{aligned} I_s(x) &= U_s(x)U_s^*(x) \\ &= |U_s(x)||U_s(x)| \exp\left[j\phi_s(x)\right] \exp\left[-j\phi_s(x)\right], \\ &= |U_s(x)|^2 \end{aligned} \tag{32}$$

and so we see that by recording the intensity of the optical field we loose the important phase information. There are several techniques for recovering this information and we now look at one in more detail.

Digital holography.

The previous two screen grabs were taken from Ref. [8]

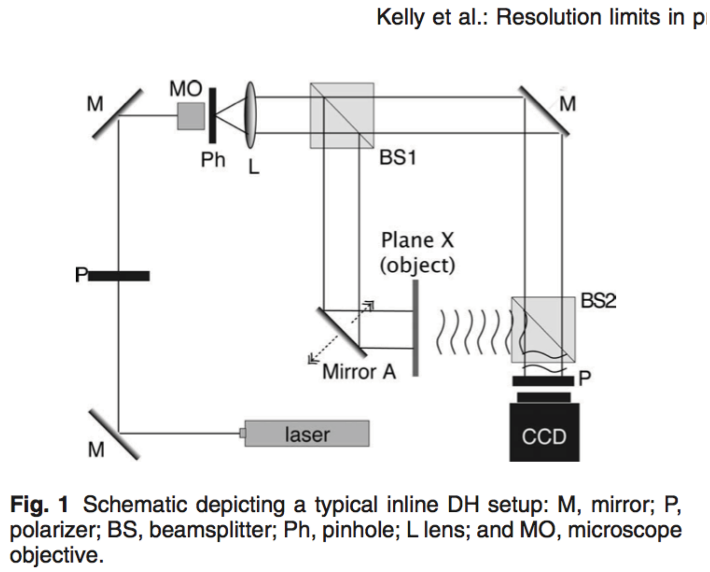


Fig. 1 Schematic depicting a typical inline DH setup: M, mirror; P, polarizer; BS, beamsplitter; Ph, pinhole; L lens; and MO, microscope objective.

Figure 8: Digital holography setup, pic taken from Ref. [8]

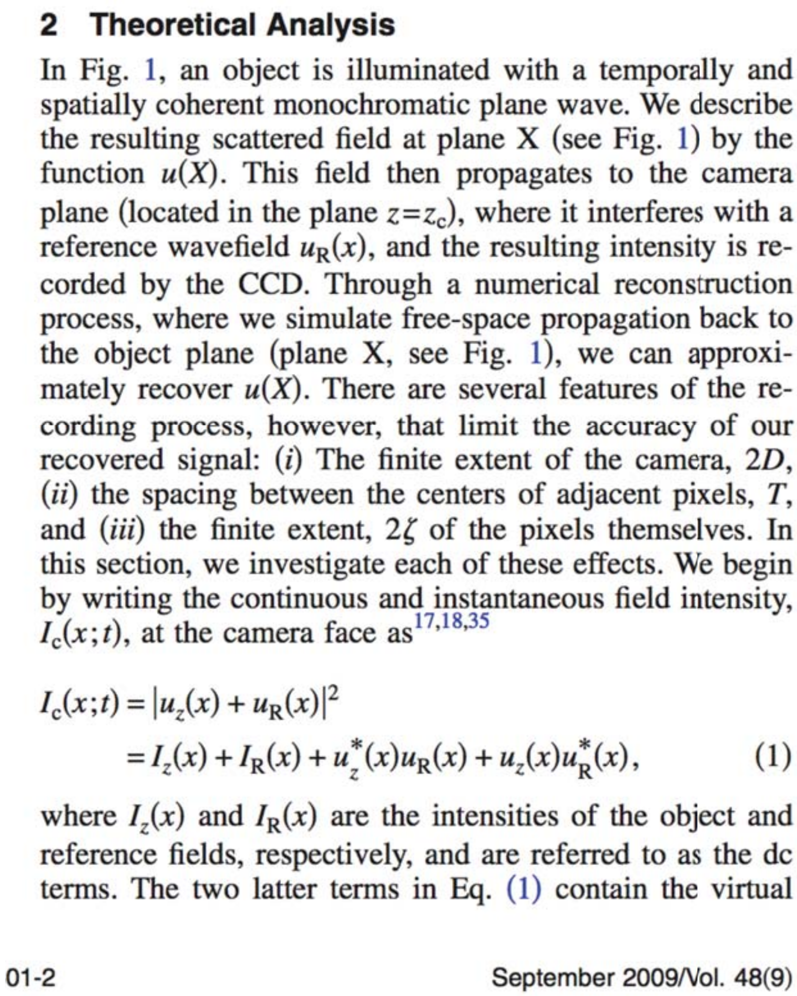
Back to the digital hologram ...

We remember from last week that the Fresnel transform is used to relate an optical wavefield in one plane to the optical distribution (wavefield) in a another plane located a distance z from the first. Again the formal definition is given by:

$$u_z\left(x,y\right) = \frac{1}{j\lambda z}\int_{\infty}^{\infty} u(X,Y)\exp\left\{\frac{j\pi}{\lambda z}\left[(x-X)^2+(y-Y)^2\right]\right\} dXdY, \tag{33}$$

If our input field consists of an array (matrix) of different point sources (spherical waves) then we replace the integral with a sum

$$u_z^D(x,y) = \frac{1}{j\lambda z}\sum_{m=0}^{m=N_y-1}\sum_{n=0}^{n=N_x-1} u_{nm}\exp\left\{\frac{jk}{2z}\left[(x-X_n)^2+(y-Y_m)^2\right]\right\}, \tag{34}$$



2 Theoretical Analysis

In Fig. 1, an object is illuminated with a temporally and spatially coherent monochromatic plane wave. We describe the resulting scattered field at plane X (see Fig. 1) by the function $u(X)$. This field then propagates to the camera plane (located in the plane $z=z_c$), where it interferes with a reference wavefield $u_R(x)$, and the resulting intensity is recorded by the CCD. Through a numerical reconstruction process, where we simulate free-space propagation back to the object plane (plane X, see Fig. 1), we can approximately recover $u(X)$. There are several features of the recording process, however, that limit the accuracy of our recovered signal: (*i*) The finite extent of the camera, $2D$, (*ii*) the spacing between the centers of adjacent pixels, T, and (*iii*) the finite extent, 2ζ of the pixels themselves. In this section, we investigate each of these effects. We begin by writing the continuous and instantaneous field intensity, $I_c(x;t)$, at the camera face as[17,18,35]

$$
\begin{aligned}
I_c(x;t) &= |u_z(x) + u_R(x)|^2 \\
&= I_z(x) + I_R(x) + u_z^*(x)u_R(x) + u_z(x)u_R^*(x),
\end{aligned}
\tag{1}
$$

where $I_z(x)$ and $I_R(x)$ are the intensities of the object and reference fields, respectively, and are referred to as the dc terms. The two latter terms in Eq. (1) contain the virtual

01-2 September 2009/Vol. 48(9)

Figure 9: Digital holography setup.

$$
u_{nm} = C_{nm} \exp(j\phi_{nm}) \tag{35}
$$

is a 2-D matrix of complex numbers and where X and Y are vectors, i.e.

$$
\begin{aligned}
X &= [X_1, X_2, ...X_{N_x}], \\
Y &= \left[Y_1, Y_2, ...Y_{N_y}\right].
\end{aligned}
\tag{36}
$$

We may consider Eq. (34) - (36) as being equivalent to a large number

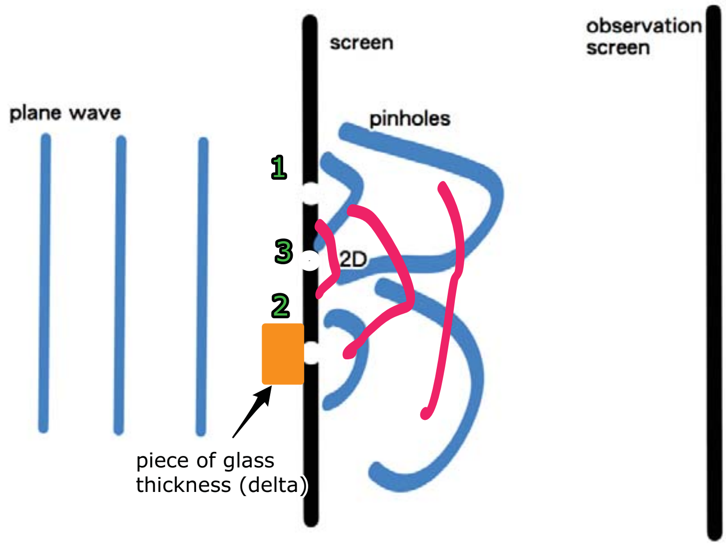


Figure 10: Three pinholes with a piece of glass over one of the pinholes.

$(N_x \times N_y)$ of different spherical waves - i.e. just a straight-forward extension of the theoretical analysis we looked at last week. The next step is to have a look at how we can perform this mathematical operation. Lets expand out Eq. (34),

$$u_z^D(x,y) \quad = \quad \frac{1}{\sqrt{j\lambda z}} \mathrm{e}^{\frac{j\pi}{\lambda z}(x^2+y^2)} \sum_{m=0}^{m=N_y-1} \sum_{n=0}^{n=N_x-1} \left\{ u_{nm} \mathrm{e}^{\frac{j\pi}{\lambda z}(X_n^2+Y_m^2)} \right\} \exp\left[\frac{-j2\pi}{\lambda z}(xX_n + yY_m) \right]. \tag{37}$$

We now make several comments about Eq. (37):

1. Although the distribution is calculated from a discrete number $(N_x \times N_y)$ of values, the output is a continuous function of x and y,
2. To calculate the distribution at any point (x_1, y_1) requires $N_x \times N_y$ which can be very time-consuming,

3. This underlines the need for fast numerical algorithms to calculate Fresnel diffraction integrals,

4. The term in curly brackets may be conveniently expressed and stored in Matlab as a 2-D matrix.

Turn to Mathematica! One calculation on a 1024×1392 takes about 11 secs - way too slow, since to produce a similar size output would take days and days!

The importance of fast algorithms can also be shown by considering this simple example, what is the sum of all the numbers between 1 and 100?

How can we speed up our calculation of Eq. (37)? We proceed with a 1-D analysis for notational simplicity and make the following substitution: $X_n = n\delta X$ so that Eq. (37) may be re-written as

$$u_z^D(x) \;=\; \frac{1}{\sqrt{j\lambda z}} \mathrm{e}^{\frac{j\pi}{\lambda z}(x^2)} \sum_{n=0}^{n=N_x-1} \left\{ u_n \mathrm{e}^{\frac{j\pi}{\lambda z}(n\delta X)^2} \right\} \exp\left[-j2\pi \left(\frac{x}{\lambda z} \right) (n\delta X) \right]. \tag{38}$$

What is δX? It is the step size between neighbouring samples, or in practical digital holographic terms, it is the distance between neighbouring camera pixels, typical values for this distance are $\approx 6\mu m$. We now look at our output spatial variable x. We wish to calculate the output distribution for a finite number of points K in the output plane and it remains to choose how we define the output step size δx. In order to make use of the Fast Fourier Transform algorithm the following choice is imposed: $K\delta x = \lambda z/\delta X$ and furthermore that $K = N$ or that the number of samples in out input domain and output domain are equal. Making these substitutions, i.e. $x = k\delta x$, Eq.

(38) becomes

$$
\begin{aligned}
u_z^k &= \frac{1}{\sqrt{j\lambda z}} \mathrm{e}^{\frac{j\pi}{\lambda z}(k\delta x)^2} \sum_{n=0}^{n=N_x-1} \left\{ u_n \mathrm{e}^{\frac{j\pi}{\lambda z}(n\delta X)^2} \right\} \exp\left[-j2\pi \left(\frac{k\delta x}{\lambda z} \right) (n\delta X) \right], \\
&= \frac{1}{\sqrt{j\lambda z}} \mathrm{e}^{\frac{j\pi}{\lambda z}(k\delta x)^2} \sum_{n=0}^{n=N_x-1} \left\{ u_n \mathrm{e}^{\frac{j\pi}{\lambda z}(n\delta X)^2} \right\} \exp\left[-j2\pi \left(\frac{nk}{N} \right) \right],
\end{aligned}
\tag{39}
$$

where $u_z^k = [u_z^0, u_z^1, ...u_z^{K-1}]$. With this representation we may use the Fast Fourier Transform to calculate the sum in Eq. (39) which reduces the number of calculations from N^2 to $N \log_2(N)$. Hence Eq. (39) can be expressed as

$$
u_z^k = \frac{1}{\sqrt{j\lambda z}} \mathrm{e}^{\frac{j\pi}{\lambda z}(k\delta x)^2} \mathrm{FFT} \left\{ u_n \mathrm{e}^{\frac{j\pi}{\lambda z}(n\delta X)^2} \right\}, \tag{40}
$$

farprop2.m:26

```
%   ud: wavefront after propagation
%   d: distance of propagation
%   ui: wavefront before propagation
%   dx,dy: CCD pixel size
%   lambda is wavelength of light

%   All units in mm

function ud=farprop2(d,ui,dy,dx,lambda)

if nargin<3
    dx = 0.00645;
    dy = 0.00645;
end;

lambda=785e-6;  % wavelength

tic
[M,N]=size(ui);
uo=ui;
[m,n]=meshgrid(-N/2:N/2-1,-M/2:M/2-1);
g=exp(i*pi/lambda/d*((m*dx).^2+(n*dy).^2));
ud=fftshift(ifft2(fftshift(uo.*g))); clear uo g;
ud=ud*exp(i*2*pi/lambda*d).*exp(i*pi*lambda*d*((m/M/dx).^2+(n/N/dy).^2));
clear m; clear n;
toc
```

Figure 11: Fresnel propagation code.

FFT algorithm[4]

The DFT of an N-point sequence is defined as

$$F_m = \sum_{0}^{N-1} f_n \exp\Big(-\frac{j2\pi nm}{N}\Big) \tag{41}$$

$$= \sum_{0}^{N-1} f_n \omega_N^{nm} \tag{42}$$

For, e.g. $N = 4$, this expression constitutes the following matrix multiplication,

$$\begin{pmatrix} F_0 \\ F_1 \\ F_2 \\ F_3 \end{pmatrix} = \begin{pmatrix} \omega_N^0 & \omega_N^0 & \omega_N^0 & \omega_N^0 \\ \omega_N^0 & \omega_N^1 & \omega_N^2 & \omega_N^3 \\ \omega_N^0 & \omega_N^2 & \omega_N^4 & \omega_N^6 \\ \omega_N^0 & \omega_N^3 & \omega_N^6 & \omega_N^9 \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \end{pmatrix} \tag{43}$$

Noting that $\omega_N^{k+pN} = \omega_N^k$,

$$\begin{pmatrix} F_0 \\ F_1 \\ F_2 \\ F_3 \end{pmatrix} = \begin{pmatrix} \omega_N^0 & \omega_N^0 & \omega_N^0 & \omega_N^0 \\ \omega_N^0 & \omega_N^1 & \omega_N^2 & \omega_N^3 \\ \omega_N^0 & \omega_N^2 & \omega_N^0 & \omega_N^2 \\ \omega_N^0 & \omega_N^3 & \omega_N^2 & \omega_N^1 \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \end{pmatrix} \tag{44}$$

Clearly this takes $N \times N$ multiplications. As N grows, this becomes quite slow to compute.

Cooley-Tukey FFT

Recall Eq. (41),

$$F_m = \sum_{0}^{N-1} f_n \omega_N^{nm} \tag{45}$$

[4] I am grateful to John Healy for his contribution to this section.

We can split this into two parts, the even and odd points of the input. (Actually, we can split it into any number of parts providing that number divides N. That number is called the *radix* of the algorithm.)

$$F_m = \sum_{0}^{(N/2)-1} f_{2n}\omega_N^{(2n)m} + \sum_{0}^{(N/2)-1} f_{2n+1}\omega_N^{(2n+1)m} \tag{46}$$

The two sums are very similar. We can emphasis this by bringing the differing part (which we call the *twiddle factor*) outside the sum,

$$F_m = \sum_{0}^{(N/2)-1} f_{2n}\omega_N^{2nm} + \omega_N^m \sum_{0}^{(N/2)-1} f_{2n+1}\omega_N^{2nm} \tag{47}$$

Next, we note the identity $\omega_N^{2nm} = \omega_{N/2}^{nm}$, giving,

$$F_m = \sum_{0}^{(N/2)-1} f_{2n}\omega_{N/2}^{nm} + \omega_N^m \sum_{0}^{(N/2)-1} f_{2n+1}\omega_{N/2}^{nm} \tag{48}$$

Compare these two sums with Eq. (41), and it is evident that we now have the sum of two $N/2$-point DFTs. One more trick,

$$\omega_N^{m+N/2} = -\omega_N^m \tag{49}$$

and we can rewrite Eq. (44) as the following pair of equations,

$$\begin{pmatrix} F_0 \\ F_1 \end{pmatrix} = \begin{pmatrix} \omega_N^0 & \omega_N^0 \\ \omega_N^0 & \omega_N^1 \end{pmatrix} \begin{pmatrix} f_0 \\ f_2 \end{pmatrix} + \begin{pmatrix} \omega_N^0 \\ \omega_N^1 \end{pmatrix} \cdot \begin{pmatrix} \omega_N^0 & \omega_N^0 \\ \omega_N^0 & \omega_N^1 \end{pmatrix} \begin{pmatrix} f_1 \\ f_3 \end{pmatrix} \tag{50}$$

$$\begin{pmatrix} F_2 \\ F_3 \end{pmatrix} = \begin{pmatrix} \omega_N^0 & \omega_N^0 \\ \omega_N^0 & \omega_N^1 \end{pmatrix} \begin{pmatrix} f_0 \\ f_2 \end{pmatrix} - \begin{pmatrix} \omega_N^0 \\ \omega_N^1 \end{pmatrix} \cdot \begin{pmatrix} \omega_N^0 & \omega_N^0 \\ \omega_N^0 & \omega_N^1 \end{pmatrix} \begin{pmatrix} f_1 \\ f_3 \end{pmatrix} \tag{51}$$

The dots above are Hadamard products. We need only perform these matrix multiplications once (e.g. in Eq. (50) and then reuse the results in evaluating Eq. (51)), so instead of N^2 multiplications as before, we only need $2(\frac{N}{2})^2+N$. The extra N comes from the twiddle factors.

The above is a single *iteration* of the algorithm. We can use it again on each of the $N/2$-point DFTs it has given us. If $N = 2^q$ for some integer q, we can do this q times. Then, the only multiplications we need are the twiddle factors, of which there are N at each iteration. Thus we have Nq or $N \log_2 N$ multiplications. Including additions, we say we need $O(N \log_2 N)$ operations to calculate an N-point DFT using the FFT algorithm iteratively. As N grows, this rapidly becomes much less than N^2.

The Fourier Transform

We now turn our attention to the formal definition for the Fourier transform, i.e.

$$U(k) = \int_{-\infty}^{\infty} u(x) \exp(-j2\pi xk) dx \tag{52}$$

We mention first that not every function has a Fourier transform. For example since the integration is over the limits $\pm\infty$, generally we require the power associated with $|u(x)|^2$ be finite over this range. All signals that are physically realizable satisfy this requirement. Surprisingly there are several functions which are so useful mathematically but that nevertheless do not meet this requirement:

1. $\sin(x)$: sine, and cosine functions

2. $\mathrm{H}(x)$: the Heaviside step function

3. $\delta(x)$: the Dirac delta impulse function

None of these functions strictly speaking have a Fourier transform. Indeed none are physically possible signals, see Chapter 2 of Ref. [7] for a more detailed discussion. Nevertheless Fourier transforms for these functions can be defined using limiting arguments. Let us consider the Fourier transforms of some important functions starting with a Gaussian function defined:

$$g(x) = \frac{1}{\sqrt{a}} \exp\left(-\frac{x^2}{a^2}\right). \tag{53}$$

Lets say we want to perform a Fourier transform on Eq. (53) what are the steps? Well, set $g(x)$ in Eq. (53) equal to $u(x)$ in Eq. (52) to give

$$\begin{aligned}
G(k) &= \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{a^2}\right) \exp(-j2\pi xk)dx, \\
&= \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{a^2} - j2\pi xk\right) dx, \\
&= \int_{-\infty}^{\infty} \exp\left[-\left(\frac{x^2}{a^2} - j2\pi xk - \pi^2k^2a^2 + \pi^2k^2a^2\right)\right] dx, \\
&= \int_{-\infty}^{\infty} \exp\left[-\left(\frac{x}{a} - j\pi ka\right)^2 + \pi^2k^2a^2\right] dx, \\
&= \sqrt{\pi a} \exp\left(-a^2k^2\pi\right) \qquad (54)
\end{aligned}$$

For the final step in solving Eq. (54) we note the following relationship $\int \exp(-\pi x^2)dx = 1$.

<u>Exercise:</u> Plot the results of Eq. (53) and compare with the results of Eq. (54) for different values of a, show in Mathematica!

How do we find the FT when $g(x) = rect(x)$, where I define

$$\text{rect}(x) = \begin{cases} 1, & \text{when } |x| < L \\ 0, & \text{otherwise.} \end{cases} \tag{55}$$

Well same as before, we sub Eq. (57) into Eq. (52) to yield

$$G(k) \quad = \quad \int_{-L}^{L} \exp(-j2\pi xk)dx,$$

and noting that $\int \exp(\Theta x)\, dx = (1/\Theta)\exp(\Theta x)$, $G(k)$ becomes

$$\begin{aligned} G(k) &= \frac{-1}{j2\pi k}\exp(-j2\pi xk)\Big\|_{-L}^{L} \\ &= \frac{-1}{j2\pi k}\{\exp(-j2\pi Lk) - \exp(-j2\pi Lk)\} \\ &= \frac{1}{\pi k}\sin(2\pi Lk)\,. \end{aligned} \tag{56}$$

<u>Exercise:</u> Find the Fourier transform of tri(x) where

$$\mathrm{tri}(x) = \begin{cases} L + x, & \text{when } -L < x < 0 \\ L - x, & \text{when } 0 < x < L. \end{cases} \tag{57}$$

References

[1] T. E. Whittaker. *A history of the theories of aether and electricity - Vol.1: The classical theories; Vol.2: The modern theories. 1900-1926.* 1960.

[2] M. Born and E. Wolf. *Principles of Optics Electromagnetic Theory of Propagation, Interference and Diffraction of Light.* 1980.

[3] Eugene Hecht. *Optics (4th Edition).* Addison Wesley, 4 edition, August 2001.

[4] J.W. Goodman. *Introduction to Fourier Optics, 2nd ed.* McGraw-Hill, New York, 1966.

[5] B. Richards and E. Wolf. Electromagnetic Diffraction in Optical Systems. II. Structure of the Image Field in an Aplanatic System. *Royal Society of London Proceedings Series A*, 253:358–379, December 1959.

[6] W. H. Southwell. Validity of the fresnel approximation in the near field. *J. Opt. Soc. Am.*, 71(1):7–14, 1981.

[7] R.N. Bracewell. *The Fourier Transform and its Applications*. McGraw-Hill, New York, 1965.

[8] Damien P. Kelly, Bryan M. Hennelly, Nitesh Pandey, Thomas J. Naughton, and William T. Rhodes. Resolution limits in practical digital holographic systems. *Optical Engineering*, 48(9):095801, 2009.