1 Light Waves

1.1 Maxwell's Equations

The field of optics describes the behavior of light as it propagates through space and materials. To understand the behavior of light, we start with the fundamental classical physics model describing it: Maxwell's equations of electrodynamics. Maxwell's equations show that the electric and magnetic fields can travel as waves. In a source-free region, Maxwell's equations in linear media are[1]

$$\vec{\nabla} \cdot \vec{E} = 0, \tag{1.1}$$

$$\vec{\nabla} \cdot \vec{B} = 0, \tag{1.2}$$

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \tag{1.3}$$

$$\vec{\nabla} \times \vec{B} = \mu\epsilon \frac{\partial \vec{E}}{\partial t}, \tag{1.4}$$

where $\vec{E}$ is the electric field, $\vec{B}$ is the magnetic field, μ is the permeability of the medium, and ϵ is the permittivity of the medium. If we take the curl of both sides of Eq. (1.3), apply the vector identity $\vec{\nabla} \times \vec{\nabla} \times \vec{E} = \vec{\nabla}(\vec{\nabla} \cdot \vec{E}) - \nabla^2 \vec{E}$ to the left-hand side and exchange the order of the time derivative and the curl on the right-hand side, we get

$$\vec{\nabla}(\vec{\nabla} \cdot \vec{E}) - \nabla^2 \vec{E} = -\frac{\partial(\vec{\nabla} \times \vec{B})}{\partial t}. \tag{1.5}$$

Then substitute from Eqs. (1.1) and (1.4) to get

$$\nabla^2 \vec{E} = \mu\epsilon \frac{\partial^2 \vec{E}}{\partial t^2}. \tag{1.6}$$

This is the wave equation in three dimensions where the wave speed is $v = 1/\sqrt{\mu\epsilon}$. Taking the curl of Eq. (1.4) and performing similar algebra shows that the magnetic field also satisfies the wave equation with the same wave speed. Thus Maxwell's equations allow for electromagnetic waves. In vacuum, the speed is $v = 1/\sqrt{\mu_0\epsilon_0} \equiv c$, the speed of light in vacuum. Light is indeed an electromagnetic wave.

We now look for solutions to Eq. (1.6) and its magnetic field counterpart. We actually only need to solve for the electric field because the magnetic field can always be found from $\vec{B} = \frac{1}{c}\hat{k} \times \vec{E}$, where $\hat{k}$ is the direction of travel. (See Exercise 8.1.) We'll assume

[1] A source-free region has no net free charge and no net free current. Linear media include vacuum and dielectrics: air, glass, and so on. For an excellent introduction to electrodynamics, see Griffiths (2017).

a single linear polarization and single-frequency (monochromatic) electromagnetic wave. The electric field should then be in the form

$$\vec{E}(\vec{r}, t) = \hat{n}\, E(\vec{r}) \cos[\omega t + \phi(\vec{r})] = \operatorname{Re}\left\{\hat{n}\, \tilde{E}(\vec{r})\, e^{i(\omega t)}\right\}, \tag{1.7}$$

where $\tilde{E}(\vec{r}) = E(\vec{r})e^{i\phi}$. Generalizing to the complex plane, we look for solutions to Eq. (1.6) of the form

$$\vec{E}(\vec{r}, t) = \hat{n}\, \tilde{E}(\vec{r})e^{i\omega t}, \tag{1.8}$$

with the anticipation that at the end we will take the real part to get the actual physical field. Substituting Eq. (1.8) into Eq. (1.6) allows us to eliminate $\hat{n}$, reducing it to a scalar equation. Also, the time derivatives can be performed explicitly bringing down an ω^2 from the $e^{i\omega t}$. This results in a second-order partial differential equation known as the Helmholtz equation

$$(\nabla^2 + k^2)\tilde{E}(\vec{r}) = 0, \tag{1.9}$$

where $k = \omega/v$ is the wave number. If we can solve Eq. (1.9) for the complex scalar field $\tilde{E}(\vec{r})$, then we can get the actual physical electric field by multiplying our solution by $\hat{n}e^{i\omega t}$ and taking the real part. Since Eq. (1.9) is a second-order partial differential equation in three spatial coordinates (e.g. x, y, z), we will need to specify appropriate boundary conditions for the field on some surface in order to obtain explicit solutions.

1.2 Huygens' Principle

In many cases of interest in optics, Eq. (1.9) is solved to a *good approximation* by Huygens' integral. The field is assumed to be known on a "source plane" S_1 perpendicular to the z-axis and is only nonzero in some finite region of that plane. The values of the field on S_1 serve as a boundary condition for solving Eq. (1.9). The solution is given by Huygens' integral for the complex scalar field at any desired point x, y, z.

$$\boxed{\tilde{E}(x, y, z) = \frac{i}{\lambda} \iint_{S_1} \tilde{E}(x', y', z') \cos\phi\, \frac{e^{-ik\imath}}{\imath} \mathrm{d}S'.} \tag{1.10}$$

The integration over the source plane S_1 is performed using the integration variables x', y', z'. The vector $\vec{\imath}$ joins points in the source plane S_1 with the point (x, y, z) at which we are calculating the field. The angle between $\vec{\imath}$ and the z-axis is ϕ (see Figure 1.1). The solution represented by Huygens' integral is satisfying because it encapsulates an intuitive understanding of how light waves behave that was understood long before the formal mathematics was fully worked out.

The intuitive description of Eq. (1.10) is known as *Huygens' principle*, due to Christiaan Huygens (1629–1695), a Dutch mathematician and scientist. Under Huygens' principle, every point in the source is considered to be emitting light with spherical wavefronts propagating outward – the so-called Huygens' wavelets. These wavefronts are represented by

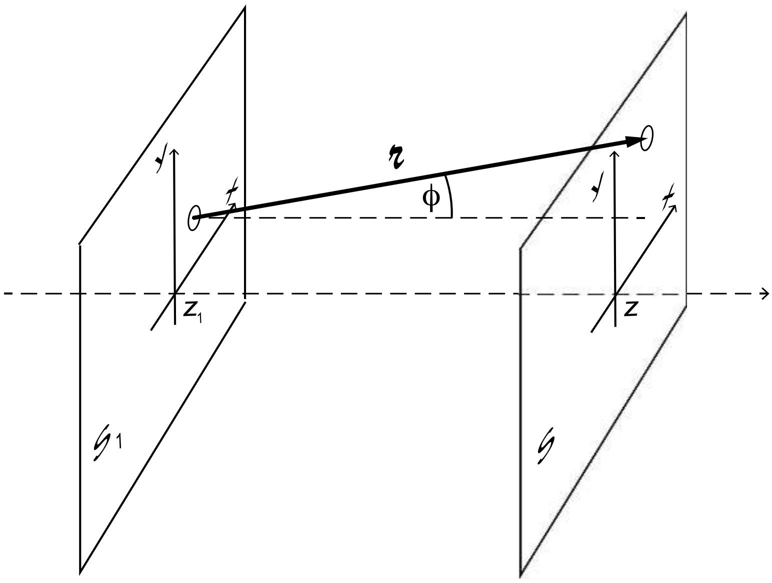
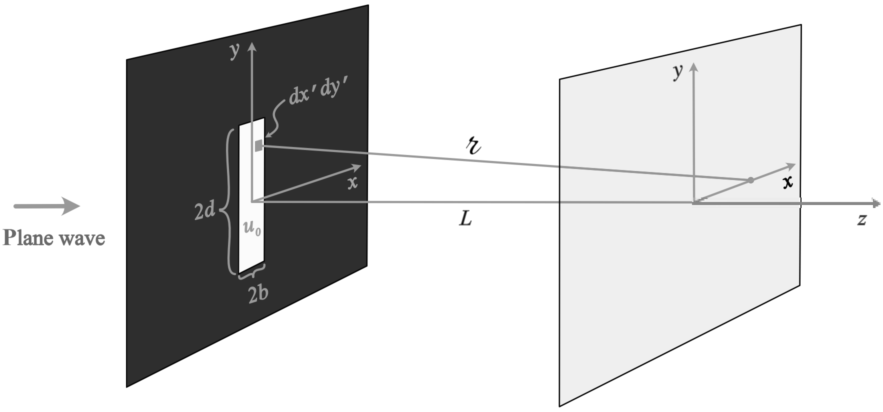


Figure 1.1 The electric field in the source plane $\mathbf{S_1}$ is propagated to the field point in plane $\mathbf{S}$. The source plane is located at $z = z_1$ and the field plane is at z. The circle on the source plane indicates a typical source point involved in the integral. The complex scalar field at this point is $\tilde{E}(x, y, z_1)$. Similarly, the circle on the field plane indicates a typical field point with complex scalar field $\tilde{E}(x, y, z)$.

the factor $\cos\phi \frac{e^{-ik\mathscr{r}}}{\mathscr{r}}$. They are emitted preferentially in the direction perpendicular to the source plane due to the presence of $\cos\phi$. The constant $\frac{i}{\lambda}$ out-front contributes 90° of phase and the λ in the denominator serves to keep the units the same on both sides of the equation. The complex scalar field in the source plane $\tilde{E}(x, y, z_1)$ sets the relative amplitudes and phases of these tiny spherical emitters. The field $\tilde{E}(x, y, z)$ is then simply the *linear superposition* of all the spherical wavefronts emitted from the source.

Example 1.1 Single-Slit Diffraction A typical use of the Huygens' integral solution 1.10 is to find the diffraction pattern from a small aperture of some specific shape. Consider the diffraction of a plane wave from a rectangular aperture of width $2b$ and height $2d$ viewed on a screen at a distance L downstream from the slit. The screen distance is much larger than either dimension of the aperture $b, d << L$. We choose our optic axis z to be perpendicular to the plane of the aperture and centered on the aperture and choose the x and y axes of the source plane and field plane to be parallel to the width and height of the slit, respectively. The slit is located at $z = 0$ and the screen at $z = L$. The complex scalar field in the slit is assumed to be from a monochromatic plane wave impinging on the slit from the left. See the following diagram. So $\tilde{E}(x, y, 0) = u_0$ for $-b < x \leq b$, $-d < y \leq d$ and zero otherwise. We want to calculate the complex scalar field on the screen along the x-axis. Taking the squared magnitude of the complex scalar field gives the irradiance.



Huygens' principle for this case is then

$$\tilde{E}(x, 0, L) = \frac{i}{\lambda} \int_{x'=-b}^{b} \int_{y'=-d}^{d} u_0 \cos\phi \, \frac{e^{-ik\imath}}{\imath} \, \mathrm{d}x' \, \mathrm{d}y'. \tag{1.11}$$

For the $\imath$ in the denominator, it's enough to use the approximation $\imath \approx L$. Since $b, d << L$, we also have $\cos\phi \approx 1$. For the $\imath$ in the exponent, where it's multiplied by $k = \frac{2\pi}{\lambda}$ and therefore causes the integrand to vary rapidly, we need the first-order binomial approximation. Since we're going to start by integrating over y', we approximate $\imath$ in terms of $\frac{y'}{L}$ as

$$\imath \approx \sqrt{L^2 + (x - x')^2} + \frac{y'^2}{2L} + \ldots. \tag{1.12}$$

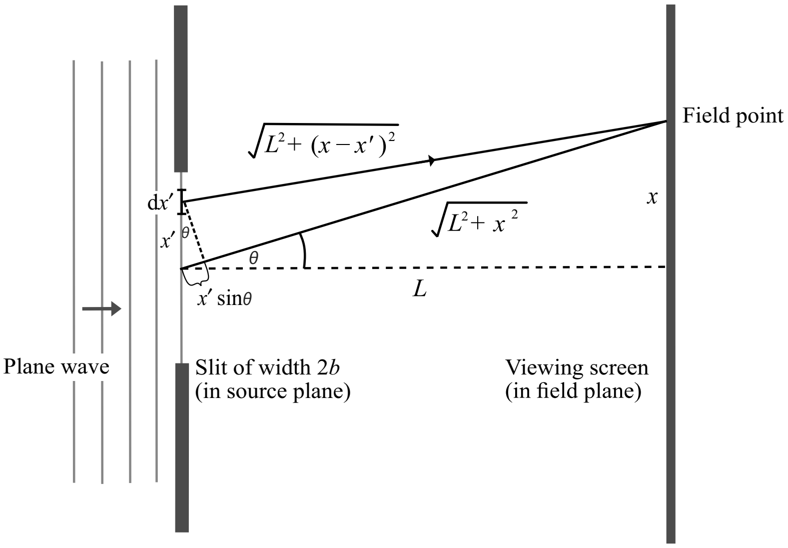
The integrand becomes separable and Eq. (1.11) is then

$$\tilde{E}(x, 0, L) \approx \frac{iu_0}{\lambda L} \int_{x'=-b}^{b} e^{-ik\sqrt{L^2+(x-x')^2}} \, \mathrm{d}x' \int_{y'=-d}^{d} e^{-ik\frac{y'^2}{2L}} \, \mathrm{d}y'. \tag{1.13}$$

The y' integral can be done with the help of tables, computer programs, and so on, but since the integrand depends only on y', the result is just an overall constant, which we shall call A. Then

$$\tilde{E}(x, 0, L) \approx \frac{iA}{\lambda L} \int_{-b}^{b} e^{-ik\sqrt{L^2+(x-x')^2}} \, \mathrm{d}x'. \tag{1.14}$$

We can now redraw the previous figure from the viewpoint of a person looking down onto the arrangement from above. The x' integral is now a sum over ribbons of width $\mathrm{d}x'$ and having the full height of the slit. This is the way the problem is sometimes represented in introductory physics texts and the issue of why it works (i.e. separability of the integrand) is not explained.



This figure should convince you that $\sqrt{L^2+(x-x')^2} = \sqrt{L^2+x^2} - x' \sin\theta$, where θ is the diffraction angle measured from the center of the slit. Making this substitution in Eq. (1.14) and carrying out the integral, we get

$$\tilde{E}(x,0,L) \approx \frac{2iAbu_0}{\lambda L} \frac{\sin(kb\sin\theta)}{kb\sin\theta} e^{-ik\sqrt{L^2+x^2}}. \tag{1.15}$$

The irradiance I is proportional to the magnitude of the complex scalar field squared:

$$I \propto \tilde{E}(x,0,L)^* \tilde{E}(x,0,L) \tag{1.16}$$

$$\propto \frac{sin^2(kb\sin\theta)}{(kb\sin\theta)^2}. \tag{1.17}$$

The maximum irradiance occurs when $\theta = 0$. If the maximum irradiance is I_0 , then

$$I(\theta) = I_0 \frac{sin^2(kb\sin\theta)}{(kb\sin\theta)^2}. \tag{1.18}$$

Note that this expression has minima whenever $kb\sin\theta = \pm n\pi$, where $n = \pm 1, \pm 2, \ldots$. Generally, this condition is written in terms of the full width $a \equiv 2b$ of the slit as

$$a\sin\theta_n = n\lambda. \tag{1.19}$$

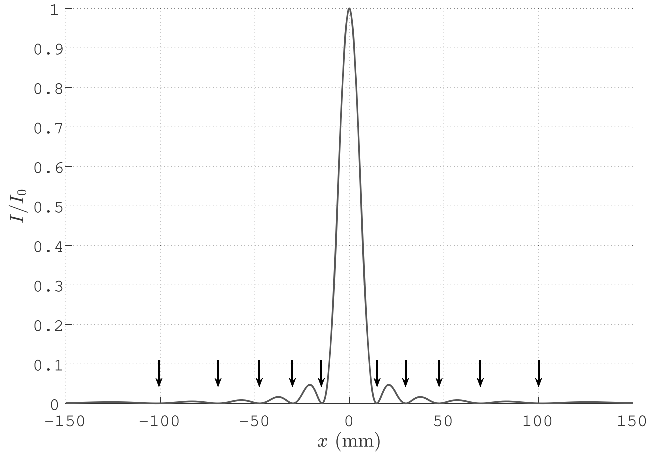
This equation is familiar to students of introductory optics. If we then plug in $\sin\theta = x/\sqrt{x^2+L^2}$, we get the irradiance in terms of the screen position x. This is the form we would want for comparing to actual measurements on a flat screen like a digital camera image chip, and so on.

$$I(x,0,L) = I_0 \left(\frac{\lambda}{\pi a}\right)^2 \frac{x^2+L^2}{x^2} \sin^2\left(\frac{\pi a}{\lambda} \frac{x}{\sqrt{x^2+L^2}}\right). \tag{1.20}$$

The corresponding minima locations are

$$x_n = \pm \frac{n\lambda L}{\sqrt{a^2 - n^2\lambda^2}}, \qquad n = 1, 2, \ldots \tag{1.21}$$

The figure here shows the irradiance distribution from Eq. (1.20) and minima locations from Eq. (1.21) for the diffraction of a HeNe laser, $\lambda = 632.8$ nm at a slit seven wavelengths wide. The screen (field plane) is 1 m from the slit, so you can see from the x-axis that the diffraction pattern isn't very wide. It falls off quickly away from the central maximum.



Huygens' principle in the form of Eq. (1.10) can also be used to get a numerical solution using a finite number of discrete emitters. The integral is thus converted to a sum over the fields emanating from these discrete emitters. As an example, Figure 1.2 shows the result of adding the fields propagating from 100,000 identical spherical emitters in a slit 10 wavelengths wide and 100 wavelengths tall. Although the field near the aperture changes quite rapidly, at large distances from the aperture the diffracted field settles into the uniform pattern characteristic of single-slit diffraction.

It is worth re-emphasizing that the value of Huygens' principle lies not only in the fact that we can evaluate the right-hand side analytically or on a computer. It gives us an intuition for the nature of electromagnetic radiation beyond simple plane waves.

1.3 The Paraxial Approximation

We're going to be most interested in the propagation of "beams" of light.[2] Beams of light propagate mostly in one direction and we choose our axes so that the z-axis lies in the direction of propagation. The paraxial approximation is the assumption that all wavefront normals make small angles with the z-axis. It is appropriate for beams and any situation where light travels mostly in one direction. We consider the propagation between two planes perpendicular to the z-axis as shown in Figure 1.1: a source plane $\mathbf{S_1}$ at $z = z_1$

[2] Lasers are the main source of light beams nowadays. For consistency, the approach in this section corresponds closely to that of classic laser textbooks, which should be consulted for extra details. I recommend Svelto (2010) for a more introductory approach and Siegman (1986) for those who want to fill in all the gaps.

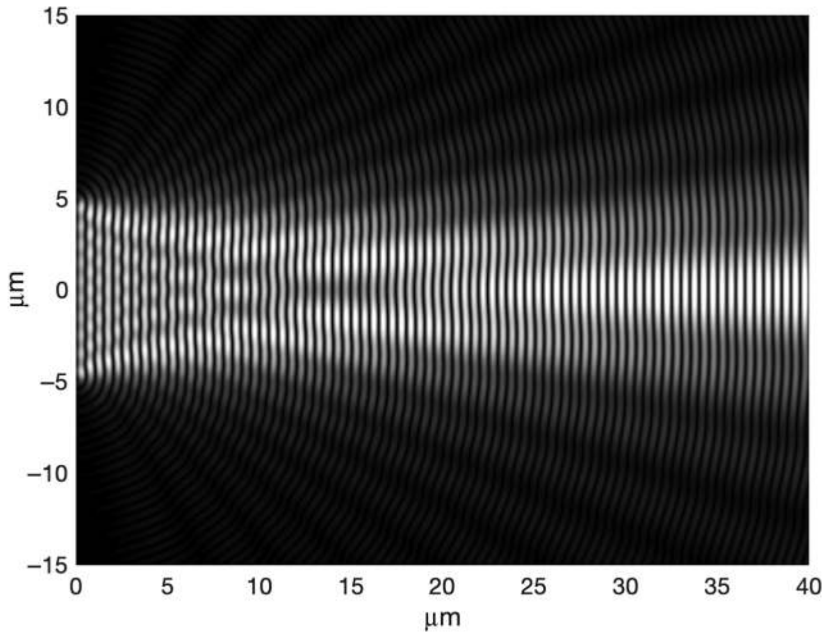


Figure 1.2 Simulation of a plane wave incident from the left, diffracting through an aperture 10 wavelengths across and 100 wavelengths high. The wavelength is 1 μm. The aperture is at the left edge of the image (between −5 and +5 μm). The irradiance was estimated by Huygens' principle using approximately 100,000 spherical Huygens' wavelet emitters in a rectangular grid 10 wavelengths wide and 100 wavelengths high (out of page). The far-field pattern has essentially established itself by the time the light reaches the right-hand edge of the image, 40 μm downstream. A screen placed at the far right of the image, perpendicular to the page, would register the classic single-slit diffraction pattern with the bright central maximum. (The dark and bright bands running mostly vertically are due to the fact that this is a snapshot of the irradiance at a single time. They are the crests, troughs, and zeros of the electromagnetic wave.) To illustrate the fainter features of the diffracted beam, I've plotted the field magnitude (square root of irradiance) rather than the irradiance. This is also closer to the way our eye perceives the irradiance pattern.

and a "downstream" field plane **S** at some unspecified z. As before, we assume that in the source plane, the complex scalar field $\tilde{E}(x, y, z_1)$ is known.

The Helmholtz equation simplifies in the paraxial approximation. Since the light is propagating primarily in the z-direction, it's useful to separate out the rapid phase accumulation in z due to the wave nature of the light by writing

$$\tilde{E}(x, y, z) = u(x, y, z)\, e^{-ikz}. \tag{1.22}$$

The idea is that as long as the light is traveling largely in the z-direction, the $u(x, y, z)$ will vary very little over distances on the order of a wavelength. In other words, our wave can be treated as something close to a plane wave but with a complex amplitude $u(x, y, z)$ that varies *slowly* with position. $u(x, y, z)$ is sometimes known as the complex field amplitude or just the field amplitude. Substituting Eq. (1.22) into Eq. (1.9), we get

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} - 2ik\frac{\partial u}{\partial z} = 0. \tag{1.23}$$

In the common case where the wave occupies only a small region in the source plane and the phasefronts are fairly flat – typical characteristics of what we might call "beams" – then u will vary more slowly in the z-direction than in any other direction, namely

$$\left|\frac{\partial^2 u}{\partial z^2}\right| \ll \left|\frac{\partial^2 u}{\partial x^2}\right| \tag{1.24}$$

$$\left|\frac{\partial^2 u}{\partial z^2}\right| \ll \left|\frac{\partial^2 u}{\partial y^2}\right|. \tag{1.25}$$

Also, the fractional change in the slope $\frac{\partial u}{\partial z}$ should be small over a wavelength λ. That is

$$\left|\frac{\Delta\left(\frac{\partial u}{\partial z}\right)}{\frac{\partial u}{\partial z}}\right| \approx \left|\frac{\frac{\partial^2 u}{\partial z^2}\lambda}{\frac{\partial u}{\partial z}}\right| \ll 1. \tag{1.26}$$

Since $k = \frac{2\pi}{\lambda}$ this implies

$$\left|\frac{\partial^2 u}{\partial z^2}\right| \ll \left|2k\frac{\partial u}{\partial z}\right|. \tag{1.27}$$

So, we can drop $\frac{\partial^2 u}{\partial z^2}$ from Eq. (1.23), leaving

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} - 2ik\frac{\partial u}{\partial z} = 0. \tag{1.28}$$

This equation is generally known as the paraxial wave equation.

A solution to the paraxial wave equation can be obtained from Huygens' integral. In the paraxial approximation, where all propagation is close to the optic axis, the pathlength $\mathscr{r}$ in Figure 1.1 can be approximated as

$$\mathscr{r} = \sqrt{(x - x')^2 + (y - y')^2 + (z - z_1)^2} \tag{1.29}$$

$$\approx L + \frac{(x - x')^2 + (y - y')^2}{2L} + \cdots, \tag{1.30}$$

where, in the second line, L is the propagation distance ($L \equiv z - z_1$). With this, and $\cos\theta \approx 1$, we can rewrite Huygens' integral Eq. (1.10) as

$$u(x, y, z) = \frac{i}{\lambda L}\iint_{\mathbb{R}^2} u(x', y', z_1)\, e^{-ik\frac{(x-x')^2+(y-y')^2}{2L}}\, \mathrm{d}x'\, \mathrm{d}y'\,. \tag{1.31}$$

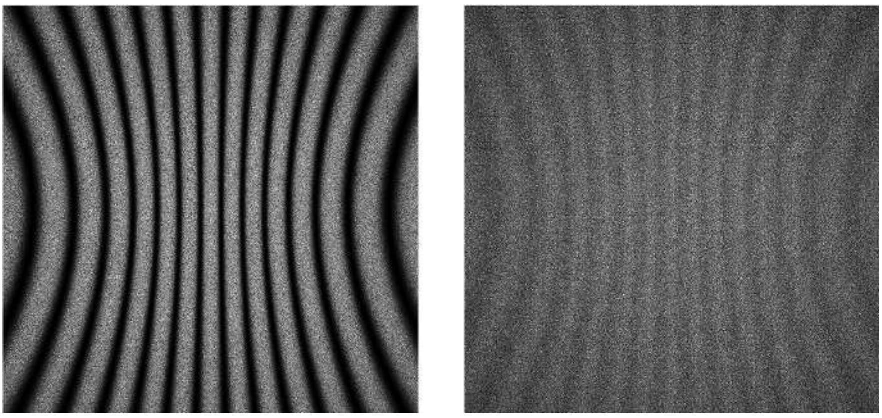
This integral solution to Eq. (1.28) allows us to handle most systems involving beams and is the easiest way to propagate the complex scalar field between two planes. The paraxial approximation made here amounts to what is also called the "Fresnel approximation." Equation (1.31) is therefore referred to as Huygens' integral in the Fresnel approximation.

1.4 Coherence

Huygens' principle also allows us to discuss one of the ways in which we classify the statistical properties of light. Light sources are often discussed in terms of their coherence, which comes in two types: temporal coherence and spatial coherence. In a temporally coherent emitter, all of the Huygens' wavelets are emitting at the same frequency and the phase of each individual emitter remains fixed for a *long time* (e.g. many nanoseconds for a HeNe laser). This is known as the coherence time. Lasers have high temporal coherence compared to other sources of light. As a result, lasers tend to be very narrow-band emitters, emitting in only a very narrow band of wavelengths around some nominal wavelength. In lasers, the bandwidth of the output light is referred to as the "linewidth." For example, HeNe lasers, which have fairly narrow linewidths, may emit wavelengths in the band $\lambda = 632.816 \pm 0.001$ nm. The coherence length of a HeNe is the distance traveled by the beam in the coherence time. For a HeNe, it's typically a few tens of centimeters but can be tens of meters for carefully designed units.

High temporal coherence does not in itself require that all the Huygens' emitters have the same phase, only that the phase of each individual emitter should vary slowly. Spatial coherence describes the phase relationship between the different Huygens' emitters. In a source with high spatial coherence, all the emitters are in phase with one another, or nearly so. For example, Young's double-slit experiment only yields the expected diffraction pattern when the spatial coherence of the incident light is sufficient that parts of the beam separated by the slit distance have similar phase. Sources with high spatial coherence can be focused to very small spot sizes and can be collimated so that they approximate plane waves. Note that high spatial coherence does not require high temporal coherence even though they usually occur together. As long as the phases of all the emitters stay the same, spatial coherence is preserved whether the overall phase changes are fast or slow, random or not. In Chapter 6, we discuss the properties of etendue and radiance, which are closely related to spatial coherence. Sources with high spatial coherence will have high radiance and low etendue. As a rule, both the temporal and spatial coherence of lasers are the highest of all light sources, which is the main reason they're so useful.

Example 1.2 Partial Spatial Coherence Consider a light source like certain LEDs with very small emission regions that possesses partial spatial coherence over its beam. If we consider a transverse cross section of such a beam, then adjacent photons in the cross section will usually be in phase with one another but the greater the transverse distance between the photons, the higher the probability they will have random phase with respect to one another. Imagine taking such a light source, covering its aperture with aluminum foil and poking two small pinholes in the foil. Then you've made a version of Young's double-slit experiment. An interference pattern will be formed wherever the two beams emerging from the holes overlap. The following figure shows how partial spatial coherence reduces the fringe contrast. Both panels show a simulation of the interference pattern formed by illuminating two pinholes ten wavelengths apart and viewed on a flat screen 10 cm away. The panel on the left assumes the source has perfect spatial coherence. The fringes have high contrast. The panel on the right assumes that the fields emanating from the two pinholes are only partially coherent due to an imperfect spatial coherence of the source. The fringe contrast is much lower. The fringe contrast can be characterized by a quantity called the "visibility" which is just the difference between the maximum fringe irradiance and the minimum fringe irradiance, divided by the sum. The visibility of the fringes on the left is clearly higher.



Exercises

1.1 Electromagnetic waves are transverse waves. In what sense are they transverse? Does something actually move up and down *in space*?

1.2 In Section 1.1, we took the curl of Faraday's law, Eq. (1.3), to show that the electric field has wavelike solutions in regions free of charge and current (source-free). Now do the same for the magnetic field by taking the curl of Ampere's law with Maxwell's correction, Eq. (1.4).

1.3 Use the source-free version of Gauss' law, Eq. (1.1), and "Gauss' Law for Magnetic Fields," Eq. (1.2), to show that for a monochromatic plane wave, both the electric and magnetic fields are perpendicular to the direction of propagation. *Hint*: Choose the z-axes so that it lies in the direction of propagation, writing $\vec{E} = \vec{E}_0 \cos(kz - \omega t)$ and $\vec{B} = \vec{B}_0 \cos(kz - \omega t)$.

1.4 Apply Faraday's law to a monochromatic plane wave in a linear medium traveling in the $\hat{z}$ direction to show that $\vec{B}_0 = \frac{1}{v}\hat{z} \times \vec{E}_0$, where v is the speed of light in the medium. $\vec{E}_0$ and $\vec{B}_0$ are the vector-amplitudes of the electric and magnetic fields, respectively. Explain why this relationship implies that the electric and magnetic fields