9 Optics

This chapter will complete our tour of Maxwell's equations by looking at electromagnetic waves with frequencies so high that their wavelength becomes small compared to the structures that manipulate them. The familiar equations of geometrical optics will naturally emerge from matching boundary conditions in this regime, with applications from microscopy to optical information processing. The chapter will close by considering some of the obstacles and opportunities associated with relaxing the assumptions of geometrical optics, through the study of Gaussian optics, nonlinear optics, and metamaterials. Although the focus will be on light (pun not intended), these same ideas are used at lower frequencies with *quasi-optical* RF components, at higher frequencies with reflection *X-ray optics*, and with magnetic lenses in *electron optics*.

9.1 REFLECTION AND REFRACTION

In Chapter 6 we saw that a plane TEM electromagnetic wave in a homogeneous medium can propagate with the following properties:

- $\vec{E} = \vec{E}_0 e^{i(\vec{k}\cdot\vec{r}-\omega t)}$, where the direction of the electric field $\hat{E}_0$ is perpendicular to the direction of travel $\hat{k}$.
- The magnetic field is proportional and transverse to the electric field, $\vec{H} = \sqrt{\epsilon/\mu}\,\hat{k}\times\vec{E}$.
- The velocity $v = 1/\sqrt{\mu\epsilon}$, and the wave number is $k = \omega/v = \omega\sqrt{\mu\epsilon} = 2\pi/\lambda$.

Materials are conveniently described by n, the *index of refraction*, which is the ratio of the speed of light in vacuum to its speed in the material:

$$n = \frac{c}{v} = \frac{\sqrt{\mu_0\mu_r\epsilon_0\epsilon_r}}{\sqrt{\mu_0\epsilon_0}} = \sqrt{\mu_r\epsilon_r} \quad . \tag{9.1}$$

A typical value for glass is $n = 1.5$. Notice that the ratio

$$\frac{k}{n} = \frac{2\pi}{\lambda n} = \frac{\omega}{v}\frac{v}{c} = \frac{\omega}{c} \tag{9.2}$$

is independent of the material and is determined solely by the frequency. This means that as light moves through materials with different indices of refraction the frequency won't change because it must still oscillate at the same rate, but the wavelength will.

In Section 6.3.2 we saw that at an interface between two media a and b the boundary conditions in the absence of free surface charges and currents are

- Normal component of $\vec{D}$ and $\vec{B}$ continuous

$$(\vec{D}_a - \vec{D}_b) \cdot \hat{n} = 0 \qquad (\vec{B}_a - \vec{B}_b) \cdot \hat{n} = 0 \tag{9.3}$$

- Tangential component of $\vec{E}$ and $\vec{H}$ continuous

$$(\vec{E}_a - \vec{E}_b) \times \hat{n} = 0 \qquad (\vec{H}_a - \vec{H}_b) \times \hat{n} = 0 \quad , \tag{9.4}$$

where $\hat{n}$ is normal to the interface.

Now consider a plane wave incident on an interface between two insulating dielectric materials, and let's allow for reflected and transmitted waves that might have different angles and wave vectors as shown in Figure 9.1.

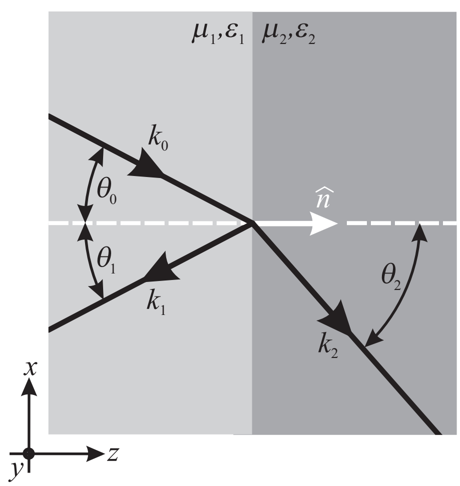

Figure 9.1. Reflection and refraction at an interface.

The continuity of the tangential component of the electric field requires

$$[\vec{E}_0 e^{i(\vec{k}_0 \cdot \vec{r} - \omega_0 t)} + \vec{E}_1 e^{i(\vec{k}_1 \cdot \vec{r} - \omega_1 t)} - \vec{E}_2 e^{i(\vec{k}_2 \cdot \vec{r} - \omega_2 t)}] \times \hat{n} = 0 \quad . \tag{9.5}$$

This must hold for all times, therefore

$$\omega_0 = \omega_1 = \omega_2 \quad , \tag{9.6}$$

and it must hold everywhere on the boundary, which implies that all three waves must have the same periodicity variation along the boundary

$$\vec{k}_0 \cdot \hat{x} = \vec{k}_1 \cdot \hat{x} = \vec{k}_2 \cdot \hat{x} \tag{9.7}$$
$$k_0 \sin\theta_0 = k_1 \sin\theta_1 = k_2 \sin\theta_2 \quad .$$

Since the magnitudes of the wave vectors k_0 and k_1 are equal because their wavelength is the same, we find that the angle of incidence equals the angle of reflection:

$$k_0 \sin\theta_0 = k_0 \sin\theta_1 \tag{9.8}$$
$$\theta_0 = \theta_1 \quad .$$

A second condition comes from $k_1/n_1 = k_2/n_2$ (equation 9.2):

$$\begin{aligned} k_1 \sin\theta_1 &= k_2 \sin\theta_2 \\ \frac{k_1}{k_2} &= \frac{\sin\theta_2}{\sin\theta_1} \\ \frac{n_1}{n_2} &= \frac{\sin\theta_2}{\sin\theta_1} \quad . \end{aligned} \tag{9.9}$$

This is *Snell's Law*, discovered experimentally by Willebrord Snell around 1621. Light rays bend when they cross the interface between media with different indices of refraction. The index of refraction can depend on wavelength and so different colors are bent in different directions; this gives rise to *chromatic dispersion* in a prism, which is usually good, and to *chromatic aberration* in lenses, which is usually bad.

We've found the directions of the rays; now we will find the amplitudes from the continuity equations

$$(\vec{E}_0 + \vec{E}_1) \times \hat{n} = \vec{E}_2 \times \hat{n} \tag{9.10}$$

and

$$(\vec{H}_0 + \vec{H}_1) \times \hat{n} = \vec{H}_2 \times \hat{n} \quad . \tag{9.11}$$

The $\vec{H}$ and $\vec{E}$ components of the wave are related by

$$\begin{aligned} \vec{H} &= \sqrt{\frac{\epsilon}{\mu}}\, \hat{k} \times \vec{E} \\ &= \sqrt{\frac{\epsilon}{\mu}}\, \frac{\vec{k}}{|\vec{k}|} \times \vec{E} \\ &= \sqrt{\frac{\epsilon}{\mu}}\, \frac{1}{\omega\sqrt{\mu\epsilon}} \vec{k} \times \vec{E} \\ &= \frac{1}{\omega\mu}\, \vec{k} \times \vec{E} \\ &\approx \frac{1}{\omega\mu_0}\, \vec{k} \times \vec{E} \end{aligned} \tag{9.12}$$

(since in most dielectric materials $\mu_r \approx 1$). Substituting into equation (9.11),

$$(\vec{k}_0 \times \vec{E}_0 + \vec{k}_1 \times \vec{E}_1) \times \hat{n} = (\vec{k}_2 \times \vec{E}_2) \times \hat{n} \quad . \tag{9.13}$$

An arbitrary incoming wave can be separated into two components that will be analyzed separately: the component with the electric field perpendicular to the plane of incidence ($\vec{E}$ points in the $\hat{y}$ direction in Figure 9.1) and the component with the electric field in the plane of incidence ($E_y = 0$).

- $\vec{E}$ perpendicular to the plane of incidence

  Since all the electric field vectors point in the $\hat{y}$ direction, the continuity equation

$$(\vec{E}_0 + \vec{E}_1 - \vec{E}_2) \times \hat{n} = 0 \tag{9.14}$$

becomes the scalar equation

$$E_0 + E_1 = E_2 \quad . \tag{9.15}$$

Using the BAC–CAB rule $\vec{A} \times (\vec{B} \times \vec{C}) = \vec{B}(\vec{A} \cdot \vec{C}) - \vec{C}(\vec{A} \cdot \vec{B})$, equation (9.13) expands to

$$[\vec{E}_0(\vec{k}_0 \cdot \hat{n}) - \hat{n}(\vec{k}_0 \cdot \vec{E}_0)] + [\vec{E}_1(\vec{k}_1 \cdot \hat{n}) - \hat{n}(\vec{k}_1 \cdot \vec{E}_1)]$$
$$= [\vec{E}_2(\vec{k}_2 \cdot \hat{n}) - \hat{n}(\vec{k}_2 \cdot \vec{E}_2)] \quad . \tag{9.16}$$

Because this is a TEM wave the $\vec{k} \cdot \vec{E}$ terms vanish, and writing out the dot products gives

$$\begin{aligned} E_0(\vec{k}_0 \cdot \hat{n}) + E_1(\vec{k}_1 \cdot \hat{n}) &= E_2(\vec{k}_2 \cdot \hat{n}) \\ E_0 k_0 \cos\theta_0 - E_1 k_1 \cos\theta_1 &= E_2 k_2 \cos\theta_2 \\ E_0 \cos\theta_0 - E_1 \cos\theta_1 &= \frac{k_2}{k_1} E_2 \cos\theta_2 \quad (k_0 = k_1) \\ &= \frac{n_2}{n_1} E_2 \cos\theta_2 \quad . \end{aligned} \tag{9.17}$$

For the incoming wave we know E_0 and θ_0, and therefore θ_1. From Snell's Law we know θ_2, leaving two unknowns (E_1 and E_2) and two equations (9.1 and 9.17). These can be solved to find the desired relationship between the incoming and reflected amplitudes:

$$\begin{aligned} E_0 \cos\theta_0 - E_1 \cos\theta_1 &= \frac{n_2}{n_1} E_2 \cos\theta_2 \\ &= \frac{\sin\theta_1}{\sin\theta_2} E_2 \cos\theta_2 \\ &= \frac{\sin\theta_1}{\sin\theta_2} (E_0 + E_1) \cos\theta_2 \quad \text{(equation 9.15)} \\ E_0 \cos\theta_0 \sin\theta_2 - E_1 \cos\theta_1 \sin\theta_2 &= (E_0 + E_1) \cos\theta_2 \sin\theta_1 \\ E_1 &= \frac{\cos\theta_0 \sin\theta_2 - \cos\theta_2 \sin\theta_1}{\cos\theta_2 \sin\theta_1 + \cos\theta_1 \sin\theta_2} \; E_0 \\ &= \frac{\cos\theta_0 \sin\theta_2 - \cos\theta_2 \sin\theta_0}{\cos\theta_2 \sin\theta_0 + \cos\theta_0 \sin\theta_2} \; E_0 \\ &= \frac{\sin(\theta_2 - \theta_0)}{\sin(\theta_2 + \theta_0)} \; E_0 \quad . \end{aligned} \tag{9.18}$$

From this we find the transmitted amplitude

$$\begin{aligned} E_2 &= E_0 + E_1 \\ &= \left[1 + \frac{\sin(\theta_2 - \theta_0)}{\sin(\theta_2 + \theta_0)}\right] \; E_0 \\ &= \frac{\sin(\theta_2 + \theta_0) + \sin(\theta_2 - \theta_2)}{\sin(\theta_2 + \theta_0)} \; E_0 \\ &= \frac{2 \sin\theta_2 \cos\theta_0}{\sin(\theta_2 + \theta_0)} \; E_0 \quad . \end{aligned} \tag{9.19}$$

Summarizing the results,

$$E_1 = \frac{\sin(\theta_2 - \theta_0)}{\sin(\theta_2 + \theta_0)} \; E_0 \quad ,$$

$$E_2 = \frac{2\sin\theta_2\cos\theta_0}{\sin(\theta_2+\theta_0)}\,E_0 \quad . \tag{9.20}$$

- $\vec{E}$ in the plane of incidence

  Continuity of the tangential component of $\vec{E}$ gives

$$(\vec{E}_0 + \vec{E}_1 - \vec{E}_2)\times\hat{n} = 0 \tag{9.21}$$

or

$$E_0\cos\theta_0 - E_1\cos\theta_1 = E_2\cos\theta_2 \tag{9.22}$$

since the cross products all point in the $\hat{y}$ direction. Similary, equation (9.13) becomes a scalar equation since $\vec{k}\times\vec{E}$ points in the $\hat{y}$ direction, and then $\hat{\vec{y}}\times\hat{n}$ points in the $\hat{x}$ direction:

$$\begin{aligned} k_0E_0 + k_1E_1 &= k_2E_2 \\ E_0 + E_1 &= \frac{k_2}{k_1}\,E_2 \quad (k_0 = k_1) \\ &= \frac{n_2}{n_1}\,E_2 \quad . \end{aligned} \tag{9.23}$$

Once again we have two equations for our two unknowns, which can be solved with a bit more algebra to show that

$$\begin{aligned} E_1 &= \frac{\tan(\theta_0-\theta_2)}{\tan(\theta_0+\theta_2)}\,E_0 \quad , \\ E_2 &= \frac{2\cos\theta_0\sin\theta_2}{\sin(\theta_0+\theta_2)\cos(\theta_0-\theta_2)}\,E_0 \quad . \end{aligned} \tag{9.24}$$

Equations (9.20) and (9.24) are the *Fresnel equations*. Notice that E_1 can vanish in equation (9.24) if the numerator vanishes, which will happen when $\theta_0 = \theta_2$. This is trivial: it says that there is no reflection if the materials are the same. E_1 can also vanish when the denominator diverges, which will happen if $\theta_0 + \theta_2 = \pi/2$, which means that the transmitted and reflected beams are perpendicular. This angle is called *Brewster's angle* θ_B, and may be found from Snell's Law to be

$$\frac{n_2}{n_1} = \frac{\sin\theta_B}{\sin[(\pi/2)-\theta_B]} = \tan\theta_B \quad . \tag{9.25}$$

At this angle incoming radiation with the field pointing in an arbitrary direction will be reflected with no component of the field in the plane of incidence. The reflected radiation will be *linearly polarized* with the field pointing solely parallel to the plane of the interface between the materials. This is how polarizing sunglasses work: since reflected light close to Brewster's angle is nearly linearly polarized, glasses that contain vertically oriented polarizers will block most of the reflected glare [Land, 1951]. We will cover polarization in more detail in Chapter 12.

A second important angle for reflections is the *critical angle* θ_c for which $\theta_2 = \pi/2$:

$$\frac{n_2}{n_1} = \frac{\sin\theta_c}{\sin\theta_2} = \frac{\sin\theta_c}{1} \quad\Rightarrow\quad \theta_c = \sin^{-1}\left(\frac{n_2}{n_1}\right) \quad . \tag{9.26}$$

Since $\sin\theta_2$ can be no larger than 1, if the light arrives at an angle closer to the surface than θ_c then the boundary conditions cannot be satisfied if there is a transmitted wave and so the wave will be completely reflected. This is called *total internal reflection*, and is used to confine light in display panels, light pipes, and multi-mode optical fibers. The light will make multiple bounces inside the dielectric material, but as long as the angle is kept below the critical angle it will not leak out.

9.2 GEOMETRICAL OPTICS

Geometrical optics considers the propagation of light when the wavelength is small compared to the relevant length scales in a problem, so that it can be approximated by the reflection and refraction of plane waves from locally straight interfaces. Consider a ray passing through a spherical lens with a radius of curvature R and index of refraction n, shown in cross-section in Figure 9.2.

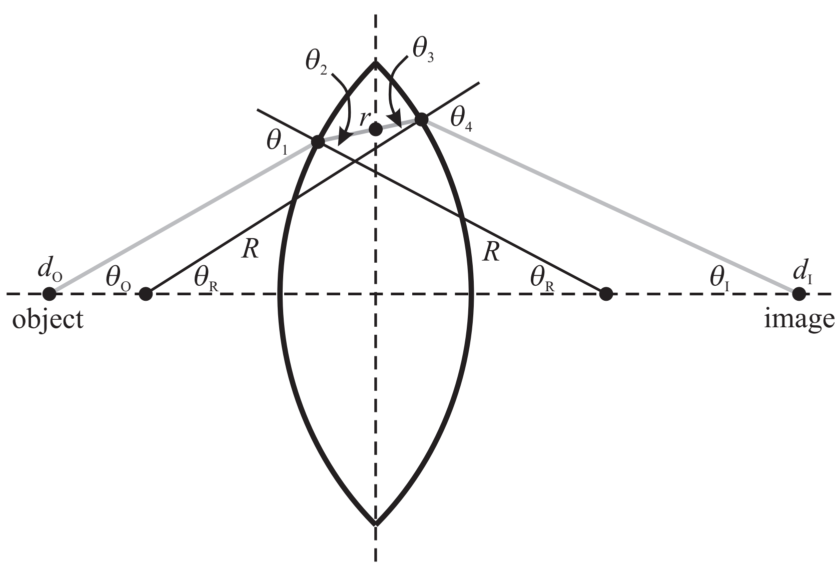

Figure 9.2. A spherical lens.

We will consider only light rays that lie near the axis, called *paraxial* rays, therefore angles from horizontal will be approximated with $\sin\theta \approx \theta$. We will also assume that this is a *thin lens*, so that when a ray passes through the lens its radius r from the axis of the lens is approximately constant. To find out how the lens modifies a ray we must analyze the trigonometry of Figure 9.2 with these approximations. Since the sum of the angles in a triangle must equal 180° we can relate the incoming angle, measured from the surface normal, to the radius of curvature of the lens and the distance from the lens by

$$\begin{aligned}\theta_O + \theta_R + (180^\circ - \theta_1) &= 180^\circ \\ \theta_1 &= \theta_R + \theta_O \\ &\approx \tan^{-1}\left(\frac{r}{R}\right) + \tan^{-1}\left(\frac{r}{d_O}\right) \quad . \end{aligned} \tag{9.27}$$

In the paraxial limit $\tan\theta = \sin\theta/\cos\theta \approx \theta/1 \approx \theta$, and so

$$\theta_1 \approx \frac{r}{R} + \frac{r}{d_O} \quad . \tag{9.28}$$

Similarly, the outgoing angle measured from the surface normal is

$$\theta_4 = \theta_R + \theta_I \approx \frac{r}{R} + \frac{r}{d_I} \quad . \tag{9.29}$$

Adding equations (9.28) and (9.29),

$$\theta_1 + \theta_4 \approx r\left(\frac{2}{R} + \frac{1}{d_1} + \frac{1}{d_2}\right) \quad . \tag{9.30}$$

These angles are related to the angles inside the lens by Snell's Law

$$\frac{\sin\theta_1}{\sin\theta_2} \approx \frac{\theta_1}{\theta_2} = \frac{n_2}{n_1} = n \qquad \frac{\theta_4}{\theta_3} = n \quad , \tag{9.31}$$

taking $n = 1$ outside the lens. Combining these,

$$\theta_1 + \theta_4 = n(\theta_2 + \theta_3) \quad . \tag{9.32}$$

Finally, the internal angles are related by the included angles:

$$\theta_2 - \theta_R = \theta_R - \theta_3 \tag{9.33}$$

and so

$$\theta_2 + \theta_3 = 2\theta_R \approx 2\frac{r}{R} \quad . \tag{9.34}$$

Substituting (9.32) and (9.34) into (9.30) gives the result

$$\begin{aligned} 2n\frac{r}{R} &= r\left(\frac{2}{R} + \frac{1}{d_O} + \frac{1}{d_I}\right) \\ \underbrace{(n-1)\left(\frac{2}{R}\right)}_{\frac{1}{f}} &= \frac{1}{d_O} + \frac{1}{d_I} \quad , \end{aligned} \tag{9.35}$$

where f is the *focal length* of the lens. This is the *lens equation*, giving the relationship between where a ray starts on the axis on one side of the lens and where it crosses the axis on the other side. Notice that the angles have dropped out of this equation: all rays starting at the same distance from the lens on one side in the *object plane* are rejoined in a plane on the other side in the *image plane*. Problem 9.4 looks at the magnification associated with this.

If a ray starts in the *focal plane* $d_O = f$, then the lens equation requires that $d_I = \infty$. This means that the outgoing rays are parallel (*collimated*), meeting only at infinity. Lenses can also be described in terms of the *F number*, which is the ratio of the focal length to the diameter. If the focal length is 10 cm, and the diameter of the lens is 5 cm, then the F number is 10/5=2 and is written $f/2$. Another way to characterize a lens is by the *numerical aperture* (*NA*), the sine of the half-divergence angle times the index of refraction of the space the light is traveling in.

9.2.1 Ray Matrices

The calculations leading up to the lens equation were not too difficult, but they will rapidly become awkward in a system with multiple optical elements. This task can be simplified by introducing *ray matrices* that define how an arbitrary optical element transforms a light ray.

At a point along the axis of an axisymmetric optical system, a geometrical optics ray is characterized by the radius from the axis r and the slope $r' = dr/dz$. The action of a linear optical element can be specified in terms of a matrix operating on this state vector:

$$\begin{bmatrix} r_{\rm out} \\ r'_{\rm out} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} r_{\rm in} \\ r'_{\rm in} \end{bmatrix} \quad . \tag{9.36}$$

The advantage of this approach is that the action of multiple elements can be found by multiplying their ray matrices. As a simple example, a ray that passes through a homogeneous medium of width w will emerge with the same angle but a new radius of $r_{\rm out} = r_{\rm in} + r'w$, and so the corresponding ray matrix is

$$\begin{bmatrix} 1 & w \\ 0 & 1 \end{bmatrix} \quad . \tag{9.37}$$

To find the ray matrix for a lens, first notice that equation (9.35) can be written as

$$\begin{aligned} \frac{1}{f} &= \frac{1}{d_{\rm O}} + \frac{1}{d_{\rm I}} \\ &\approx \frac{1}{r}(\theta_{\rm I} - \theta_{\rm O}) \\ &\approx \frac{1}{r}(r'_{\rm in} - r'_{\rm out}) \quad , \end{aligned} \tag{9.38}$$

with a negative sign because $d_{\rm I}$ is on the opposite side of the lens from $d_{\rm O}$. This can be rearranged as

$$r'_{\rm out} = r'_{\rm in} - \frac{r_{\rm in}}{f} \quad ,$$

therefore the ray matrix for a lens is

$$\begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix} \quad . \tag{9.39}$$

Snell's Law in ray matrices is simply

$$\begin{bmatrix} 1 & 0 \\ 0 & n_1/n_2 \end{bmatrix} \quad . \tag{9.40}$$

As a final example, the ray matrix for a rod of length d with an index of refraction that depends quadratically with radius as $n = n_0[1 - \alpha r^2]$ is [Yariv, 1991]

$$\begin{bmatrix} \cos(\kappa d) & \frac{1}{\kappa}\sin(\kappa d) \\ -\kappa\sin(\kappa d) & \cos(\kappa d) \end{bmatrix} \quad , \tag{9.41}$$

where $\kappa = \sqrt{2\alpha}$. This is called a *GRIN* (GRaded INdex of refraction) lens; by selecting different lengths it can perform a range of useful functions, such as collimating light into or out of an optical fiber.

9.2.2 Optical Transforms

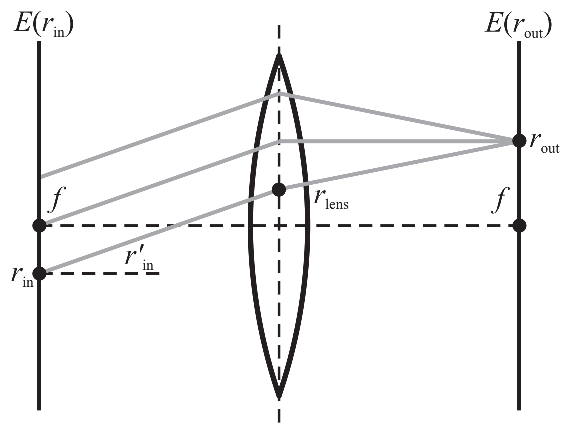

Figure 9.3. A thin lens performing a Fourier transform.

The rays that we have been studying also have a phase that depends on the path lengths and indices of refraction, which leads to interference effects when multiple *coherent* rays are combined that maintain their relative phases. An unexpected consequence of this is shown in Figure 9.3. There is a input field distribution $E(x)$ in the focal plane at a distance f from a thin lens, and we are interested in the resulting distribution in the focal plane on the other side of the lens. The ray matrix for this combined system consists of propagation through free space of width f, the lens, and then more space of width f:

$$\begin{bmatrix} 1 & f \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix} \begin{bmatrix} 1 & f \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & f \\ -1/f & 0 \end{bmatrix} \quad . \tag{9.42}$$

This means that the position of a ray in the output focal plane is $r_{\rm out} = f r'_{\rm in}$, depending solely on the angle of the ray at the input focal plane and not on its position. Therefore, the field at $r_{\rm out}$ is the sum over the input of all rays with an angle $r'_{\rm in} = r_{\rm out}/f$. Because these rays travel over different paths, there will be interference between them that we now calculate.

The position of a ray at the lens for this geometry is

$$\begin{aligned} r_{\rm lens} &= r_{\rm in} + r'_{\rm in} f \\ &= r_{\rm in} + \frac{r_{\rm out}}{f}\, f \\ &= r_{\rm in} + r_{\rm out} \quad . \end{aligned} \tag{9.43}$$

In terms of this, the distance from the input to the lens is

$$\begin{aligned} \sqrt{f^2 + (r_{\rm lens} - r_{\rm in})^2} &= \sqrt{f^2 + r_{\rm out}^2} \\ &= f\sqrt{1 + r_{\rm out}^2/f^2} \\ &\approx f\left(1 + \frac{r_{\rm out}^2}{2f^2}\right) \end{aligned}$$

$$= f + \frac{r_{\rm out}^2}{2f} \quad , \tag{9.44}$$

making the usual paraxial approximation. Similarly, the distance from the lens to the output is

$$\sqrt{f^2 + (r_{\rm lens} - r_{\rm in})^2} \approx f + \frac{r_{\rm in}^2}{2f} \quad . \tag{9.45}$$

In addition to the phase shift associated with traveling over these distances there will also be a phase shift in the lens. If the lens has a thickness $2d_0$ on the optical axis and a radius of curvature R, in the paraxial approximation the thickness of the lens at a distance r from the axis is

$$\begin{aligned}
[(R - d_0) + d]^2 + r^2 &= R^2 \\
R - d_0 + d &= \sqrt{R^2 - r^2} \\
&\approx R - \frac{r^2}{2R} \\
d &= d_0 - \frac{r^2}{2R} \\
2d &= 2d_0 - \frac{r^2}{R} \quad .
\end{aligned} \tag{9.46}$$

Assuming that the index of refraction outside the lens is $n = 1$, the wave vector in the lens is found from

$$\frac{k_{\rm lens}}{k_{\rm air}} = \frac{n}{1} \quad \Rightarrow \quad k_{\rm lens} = nk_{\rm air} \quad . \tag{9.47}$$

The extra phase associated with traveling through the lens, compared to the phase if the lens was not there, is

$$\begin{aligned}
kn2d - k2d &= k(n-1)2d \qquad (9.48) \\
&= k(n-1)\left(2d_0 - \frac{r^2}{R}\right) \quad .
\end{aligned}$$

Since by definition $R = 2f(n-1)$,

$$k(n-1)2d = k(n-1)2d_0 - \frac{kr^2}{2f} \quad . \tag{9.49}$$

We're almost done. The field at the output is found by adding up the phase shifts, integrated over the input:

$$\begin{aligned}
E(r_{\rm out}) &= \int_{-\infty}^{\infty} e^{ik(f+r_{\rm out}^2/2f)} e^{ik[(n-1)2d_0 - r_{\rm lens}^2/2f]} e^{ik(f+r_{\rm in}^2/2f)} \; E(r_{\rm in}) \; dr_{\rm in} \\
&= \int_{-\infty}^{\infty} e^{ik(f+r_{\rm out}^2/2f)} e^{ik[(n-1)2d_0 - (r_{\rm in}+r_{\rm out})^2/2f]} e^{ik(f+r_{\rm in}^2/2f)} \; E(r_{\rm in}) \; dr_{\rm in} \\
&= e^{i2k[f+(n-1)d_0]} \int_{-\infty}^{\infty} e^{ik[r_{\rm out}^2 - (r_{\rm in}+r_{\rm out})^2 + r_{\rm in}^2]/2f} \; E(r_{\rm in}) \; dr_{\rm in} \\
&= e^{i2k[f+(n-1)d_0]} \int_{-\infty}^{\infty} e^{-ikr_{\rm in}r_{\rm out}/f} \; E(r_{\rm in}) \; dr_{\rm in} \quad .
\end{aligned} \tag{9.50}$$

This should look familiar: it is the Fourier transform of the input distribution! Given coherent illumination at its input focal plane, a thin lens produces the Fourier transform at its output focal plane multiplied by an extra phase factor. Because this is computed, literally, at the speed of light, and works just as fast with two-dimensional inputs, optical transforms are appealing for high-speed signal and image processing. Many algorithms that can be expressed in terms of Fourier transforms such as convolution and filtering have been implemented in such optical computers (although "computer" is really a misnomer, because there is no nonlinear interaction among different paths).

9.3 BEYOND GEOMETRICAL OPTICS

9.3.1 Gaussian Optics

According to geometrical optics, a plane wave can be focused down to a spot of infinitesimal size and hence infinite energy density. But this is of course impossible, and does not happen because the plane-wave approximation is no longer justified when dimensions become comparable to the wavelength. It's then necessary to return the wave equation.

Consider a point source emitting a spherical wave e^{ikr}/r. If the wave is run backwards so that it travels towards the source, radiation arrives from all directions. Because of the uniqueness of solutions to partial differential equations, if the radiation is limited to a divergence angle fixed by the finite diameter of a lens then it can't match these boundary conditions.

Instead of a point source, now assume that spherical waves are emitted uniformly over an aperture with a width w. For convenience, this will be analyzed in 2D, corresponding to a distribution of line- rather than point-sources; the conclusion will be unchanged in the full three-dimensional *scalar diffraction theory* [Heald & Marion, 1995]. This is similar to the geometry of Figure 8.1, but by optics convention θ will be measured relative to the surface normal, and x will be the position along the aperture from the z axis. We'll be concerned with the field distribution far from the source, once again making the approximation that the distance from a source at $(z = 0, x)$ to a field location at (r, θ) relative to the origin is $r - x\cos(90 - \theta) = r - x\sin\theta$, and will include the x dependence in the relative phases but not the amplitude over the aperture. Then integrating e^{ikr} over the source gives

$$\begin{aligned}
\int_{-w/2}^{w/2} e^{ikr(x)}\,dx &= \int_{-w/2}^{w/2} e^{ik(r-x\sin\theta)}\,dx \\
&= e^{ikr}\int_{-w/2}^{w/2} e^{-ikx\sin\theta}\,dx \\
&= 2e^{ikr}\frac{\sin\left(k\frac{w}{2}\sin\theta\right)}{k\sin\theta} \\
&\approx e^{ikr} w\cos\left(k\frac{w}{2}\sin\theta\right) \quad .
\end{aligned} \tag{9.51}$$

This last line follows by taking the derivative of the numerator and denominator for the paraxial limit $\sin\theta \to 0$. If we ask for the boundary where the argument of the cosine equals 1,

$$1 = k\frac{w}{2}\sin\theta \approx k\frac{w}{2}\theta = \frac{2\pi n}{\lambda}\frac{w}{2}\theta$$

$$\theta = \frac{\lambda}{\pi n w} \quad . \tag{9.52}$$

The divergence angle of this boundary is proportional to the wavelength, and inversely proportional to the source aperture width and the index of refraction. Problem 9.5 shows that this presents a serious constraint on the resolution of optical devices, and is why microscopes use oil immersion lenses, telescopes are big, and blue lasers are desired for optical storage.

To find the field distribution everywhere, it's necessary to solve Maxwell's equations in cylindrical coordinates for a diverging wave. This is a surprisingly non-trivial calculation [Yariv, 1991], but the result drawn in Figure 9.4 for the fundamental radial and azimuthal mode TEM_{00} is close to what we just found. At the focus, the phase fronts are parallel and the transverse amplitude has a Gaussian dependence with a standard deviation of w, the *beam waist*. That is why this limit is called *Gaussian optics*. Then, after a transition region, the beam diverges with spherical phase fronts in a cone given by equation (9.52). Light with this distribution is said to be *diffraction limited*. It has the narrowest beam waist possible; an imperfect lens will produce a broader distribution. While diffraction-limited optics were once restricted to specialized scientific instruments, the inexpensive plastic lens in a CD player now attains this limit.

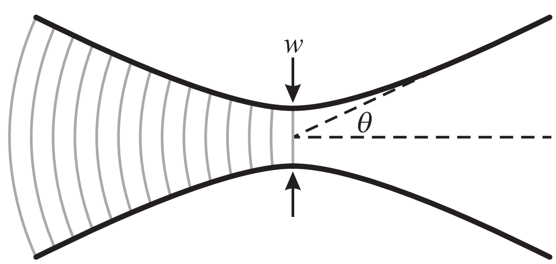

Figure 9.4. Wave fronts around a Gaussian focus.

9.3.2 Nonlinear Optics

Gaussian optics tames the infinite energy density associated with the infinitesimal focus of geometrical optics, but as Problem 6.5 showed the field strength in a tightly-focused optical beam can still be very large. It can in fact exceed the strength of the intra-atomic fields, and can therefore be used to drive atoms into novel states [Weinacht *et al*., 1999]. While reaching this limit does require powerful lasers, well before then the assumption of a linear material response breaks down. The lowest-order correction to the polarization is

$$\vec{D} = \epsilon_0\vec{E} + \vec{P} = \epsilon_0\epsilon_r\vec{E} + \epsilon_0\vec{E}^T \cdot \mathbf{d}_2 \cdot \vec{E} \quad . \tag{9.53}$$

The nonlinear optical coefficient **d** will usually depend on the orientation relative to the axes of the medium. For the simplest scalar case,

$$D = \epsilon_0\epsilon_r E + \epsilon_0 d_2 E^2 \quad . \tag{9.54}$$

Because the optical coefficient is small, on the order of 10^{-12} m/V in typical nonlinear optical materials, it usually can be considered to be a small perturbation on the polarization. Repeating the derivation of the wave equation (6.100) with this nonlinear polarization adds a term

$$\nabla^2 E - \mu_0\epsilon_0\epsilon_r \frac{\partial^2 E}{\partial t^2} = \mu_0\epsilon_0 d_2 \frac{\partial^2 E^2}{\partial t^2} \quad . \tag{9.55}$$

If the electric field has a time dependence $Ee^{i\omega t}$, the right hand side squares this to become $-\omega^2\mu_0\epsilon_0 d_2 e^{2i\omega t}$. This acts as a forcing function, generating a wave at twice the excitation frequency. Microscopically, it corresponds to exciting transitions with two photons rather than one. This is called *second harmonic generation*, and is used to make short-wavelength light from more powerful longer-wavelength sources. A common combination is to use *KDP* (KH_2PO_4) to double the 1064 nm fundamental of a *Nd:YAG* (Nd^{3+}:$Y_3Al_5O_{12}$) laser up to 532 nm. Weaker, higher harmonics can be generated with still shorter wavelengths. And if two laser beams are incident on a nonlinear crystal, the quadratic nonlinearity will generate sum- and difference-frequencies. These can be used for *parametric up*- and *down-conversion* to shift frequencies, and for gain in an *Optical Parametric Amplifier* (*OPA*).

One more restrictive assumption that we've been making is to assume that our optical elements are axisymmetric and passive. Microelectronic fabrication techniques can be used to create lenses with arbitrary shapes, called *binary optics* [Stern, 1996]. For example, a lens can be made with multiple focal lengths by interleaving the profiles for the individual lenses. *Holographic Optical Elements* (*HOE*s) are planar structures that offer the same flexibility for coherent light, using the diffractive effects to be covered in the next chapter.

The performance of a perfect optical system will still be degraded by fluctuations in the ambient environment, which sets a severe limit on ground-based telescopes and long-range optical links. But by analyzing the degradation of a point source it is possible to reconstruct a model of the atmospheric perturbations and then significantly decrease their influence by correcting for them by continuously deforming the shape of *active optical* elements [Bortoletto *et al*., 1999].

Even better, consider what happens in an inhomogeneous medium to an arbitrary paraxial wave, which can be written as a modulated plane wave $E(\vec{x}) = f_+(\vec{x})e^{i(kz-\omega t)}$. Substituting this into the wave equation and separating out the transverse and axial parts of the Laplacian,

$$\begin{aligned} 0 &= \nabla^2 E + \omega^2\mu\epsilon(\vec{x})E \\ &= \nabla_{\mathrm{T}}^2 E + \frac{\partial^2 E}{\partial z^2} + \omega^2\mu\epsilon(\vec{x})E \\ &= \nabla_{\mathrm{T}}^2 f_+ + \frac{\partial^2 f_+}{\partial z^2} + 2ikf_+ - k^2 f_+ + \omega^2\mu\epsilon(\vec{x})f_+ \quad . \end{aligned} \tag{9.56}$$

This has a complex conjugate

$$0 = \nabla_{\rm T}^2 f_+^* + \frac{\partial^2 f_+^*}{\partial z^2} - 2ikf_+^* - k^2 f_+^* + \omega^2 \mu\epsilon(\vec{x}) f_+^* \tag{9.57}$$

(assuming that the material response is real, without gain or loss). If, instead, we started with a solution traveling in the opposite direction given by $E(\vec{x}) = f_-(\vec{x})e^{i(-kz-\omega t)}$, then the wave satisfies

$$0 = \nabla_{\rm T}^2 f_- + \frac{\partial^2 f_-}{\partial z^2} - 2ikf_- - k^2 f_- + \omega^2 \mu\epsilon(\vec{x}) f_- \quad . \tag{9.58}$$

Comparing equations (9.57) and (9.58), f_- and f_+^* satisfy the same governing equations. This means that if it's possible to propagate a beam through the medium, then sending its complex conjugate back through the medium will undo the distortion [Yariv & Pepper, 1977; Yariv, 1987]. This is called *phase conjugate optics* and can be realized through nonlinear mixing.

9.3.3 Metamaterials

metamaterials [Smith *et al.*, 2004]
left-handed materials [Veselago, 1968]
resonance phase shift [Gershenfeld, 1999a]
artificial plasma [Pendry *et al.*, 1998]
negative permeability conductors [Pendry *et al.*, 1999]
split-ring resonator [Marqués *et al.*, 2003]
negative phase, group velocity [Dolling *et al.*, 2006]
causality [Garrett & McCumber, 1970, Chu & Wong, 1982]
"perfect" lens [Pendry, 2000, Tsang & Psaltis, 2008]
cloaking [Pendry *et al.*, 2006, Leonhardt & Tyc, 2009]
photonic band gap structures [Yablonovitch, 1993, Joannopoulos *et al.*, 1997]
negative index, superlens [Luo *et al.*, 2002]
slow light [Phillips *et al.*, 2001, Liu *et al.*, 2001]
electromagnetically induced transparency [Fleischhauer *et al.*, 2005]

9.3.4 Confocal Imaging

Finally, we've seen that lenses can be used to form two-dimensional images, but the world is three-dimensional. Because light that enters a lens from sources away from its focal plane will blur the image, it's not possible to use a lens to see into a three-dimensional object even if light can propagate through it. But consider what happens if the illumination is focused onto a detector through a pinhole in the *confocal* geometry shown in Figure 9.5 [Minsky, 1957; Lichtman, 1994]. Now most of the light that scatters away from the focal plane is deflected away from the pinhole, so that by scanning the sample it is possible to reconstruct a three-dimensional image. The spatial resolution and chemical sensitivity can be further enhanced by using a two-photon process for illumination [Denk *et al.*, 1990]. Such confocal microscopy has become a workhorse for biomedical imaging; Chapter 10 will continue with other kinds of imaging that don't require lenses.

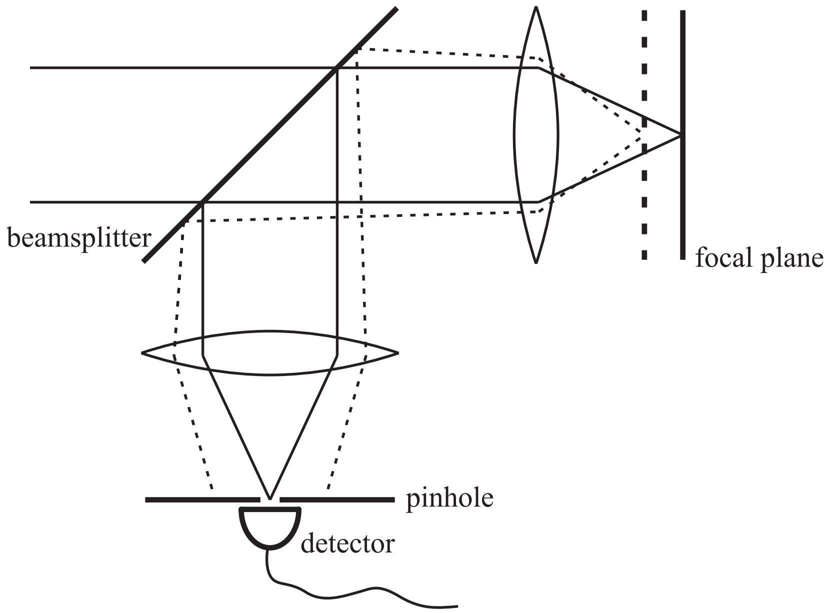

Figure 9.5. Confocal imaging.

9.4 SELECTED REFERENCES

[Born & Wolf, 1999] Born, Max, & Wolf, Emil. (1999). *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. 7th edn. New York: Cambridge University Press.

The classic optics reference.

[Yariv & Yeh, 2006] Yariv, Amnon, & Yeh, Pochi. (2006). *Photonics: Optical Electronics in Modern Communications*. New York: Oxford University Press.

A good source for modern passive and active optical systems.

9.5 PROBLEMS

(9.1) Optics (as well as most of physics) can be derived from a global law as well as a local one, in this case *Fermat's Principle:* a light ray chooses the path between two points that minimizes the time to travel between them. Apply this to two points on either side of a dielectric interface to derive Snell's Law.

(9.2) (*a*) Use Fresnel's equations and the Poynting vectors to find the *reflectivity* and *transmissivity* of a dielectric interface, defined by the ratios of incoming and outgoing energy.

(*b*) For a glass–air interface ($n = 1.5$) what is the reflectivity at normal incidence?

(*c*) What is the Brewster angle?

(*d*) What is the critical angle?

(9.3) Consider a wave at normal incidence to a dielectric layer with index n_2 between layers with indices n_1 and n_3 (Figure 9.6).

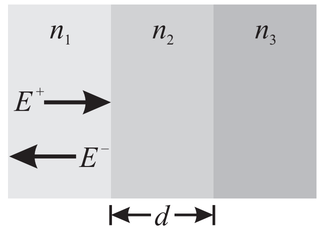

Figure 9.6. Reflection from dielectric interfaces.

(*a*) What is the reflectivity? Think about matching the boundary conditions, or about the multiple reflections.

(*b*) Can you find values for n_2 and d such that the reflection vanishes?

(9.4) Consider a ray starting with a height r_0 and some slope, a distance d_1 away from a thin lens with focal length f. Use ray matrices to find the image plane where all rays starting at this point rejoin, and discuss the magnification of the height r_0.

(9.5) Common CD players use an AlGaAs laser with a 790 nm wavelength.

(*a*) The pits that are read on a CD have a diameter of roughly 1 μm and the optics are diffraction-limited; what is the beam divergence angle?

(*b*) Assuming the same geometry, what wavelength laser would be needed to read 0.1 μm pits?

(*c*) How large must a telescope mirror be if it is to be able to read a car's license plate in visible light ($\lambda \sim 600$ nm) from a *Low Earth Orbit* (*LEO*) of 200 km?