Lecture 33

High Frequency Solutions, Gaussian Beams

When the frequency is very high, the wavelength of electromagnetic wave becomes very short. In this limit, many solutions to Maxwell's equations can be found approximately. These solutions offer a very different physical picture of electromagnetic waves, and they are often used in optics where the wavelength is short. So it was no surprise that for a while, optical fields were thought to satisfy a very different equations from those of electricity and magnetism. Therefore, it came as a surprise that when it was later revealed that in fact, optical fields satisfy the same Maxwell's equations as the fields from electricity and magnetism!

In this lecture, we shall seek approximate solutions to Maxwell's equations or the wave equations when the frequency is high or the wavelength is short compared to the geometry that the wave interacts with. High frequency approximate solutions are important in many real-world applications. This is possible when the wavelength is much smaller than the size of the structure. This can occur even in the microwave regime where the wavelength is not that small, but much smaller than the size of the structure. This is the case when microwave interacts with reflector antennas for instance. It is also the transition from waves regime to the optics regime in the solutions of Maxwell's equations. Often times, the term "quasi-optical" is used to describe the solutions in this regime.

In the high frequency regime, or when we are far away from a source much larger than the Rayleigh distance (see Section 27.2.1), the field emanating from a source resembles a spherical wave. Moreover, when the wavelength is much smaller than the radius of curvature of the wavefront, the spherical wave can be approximated by a local plane wave. Thus we can imagine rays to be emanating from a finite source forming the spherical wave. The spherical wave will ultimately be approximated by plane waves locally at the observation point. This will simplify the solutions in many instances. For instance, ray tracing can be used to track how these rays can propagate, bounce, or "richohet" in a complex environment. In fact, it is now done in a movie industry to give "realism" to simulate the nuances of how light ray will bounce around in a room, and reflect off objects.

Figure 33.1: Ray-tracing technique can be used in the movie industry to produce realism in synthetic images (courtesy of Wikipedia).

33.1 Tangent Plane Approximations

We have learnt that reflection and transmission of waves at a flat surface can be solved in closed form. The important point here is the physics of phase matching. Due to phase matching, we have derived the law of Fresnel reflection, transmission and Snell's law [59].[1]

When a surface is not flat anymore, there is no closed form solution. But when a surface is curved such that the radius of curvature is much larger than the wavelength, an approximate solution can be found. This is obtained by using a local tangent-plane approximation. Hence, this is a good approximation when the frequency is high or the wavelength is short. It is similar in spirit that we can approximate a spherical wave by a local plane wave at the spherical wave front when the wavelength is short compared to the radius of curvature of the wavefront.

When the wavelength is short, phase matching happens locally, and the Fresnel law of reflection, transmission, and Snell's law are satisfied approximately as shown in Figure 33.2. The tangent plane approximation is the basis for the geometrical optics (GO) approximation [33, 209]. In GO, light waves are replaced by light rays. As mentioned before, a light ray is a part of a spherical wave where locally, the spherical wave can be approximated by a plane wave. The reflection and transmission of these rays at an interface is then estimated using the local tangent plane approximation and local Fresnel reflection and transmission coefficients. This is also the basis for lens or ray optics from which lens technology is derived (see Figure 33.3). It is also the basis for ray tracing for high-frequency solutions [210, 211].[2]

Many real world problems do not have closed-form solutions, and have to be treated with approximate methods. In addition to geometrical approximations mentioned above, asymptotic methods are also used to find approximate solutions. Asymptotic methods imply finding a solution when there is a large parameter in the problem. In this case, it is usually the frequency. Such high-frequency approximate methods are discussed in [212–216].

[1] This law is also known in the Islamic world in A.D. 984 [208].

[2] Please note that the tangent plane approximation is invalid near a sharp corner or an edge. The solution has to be augmented by additional diffracted wave coming from the edge or the corner.

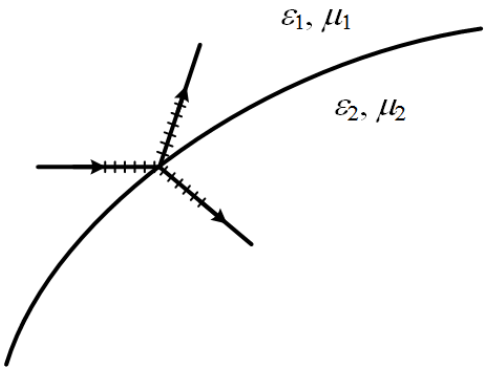

Figure 33.2: In the tangent plane approximation, the surface where reflection and refraction occur is assumed to be locally flat. Thus, phase-matching is approximately satisfied, and hence, the law of reflection, transmittion, and Snell's law are satisfied locally.

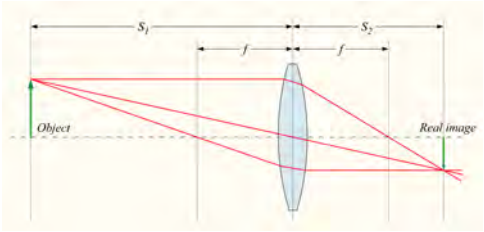

Figure 33.3: Tangent plane approximations can also be made at dielectric interfaces so that Fresnel reflection and transmission coeffients can be used to ascertain the interaction of light rays with a lens. Also, one can use ray tracing to understand the working of an optical lens (courtesy of Wikipedia).

33.2 Fermat's Principle

Fermat's principle (1600s) [59,217] says that a light ray follows the path that takes the shortest time delay between two points.[3] Since time delay is related to the phase shift, and that a light ray can be locally approximated by a plane wave, this can be stated that a plane wave follows the path that has a minimal phase shift. This principle can be used to derive the law of reflection, transmission, and refraction for light rays. It can be used as the guiding principle for ray tracing as well.

[3]This eventually give rise to the principle of least action, which is a wonderful gift of Nature! Nature finds the simplest and most efficient solution in the real world.

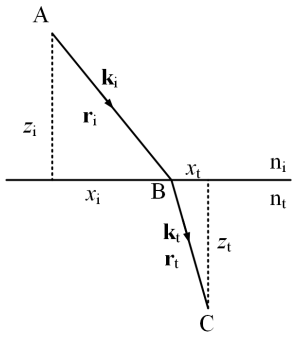

Figure 33.4: In Fermat's principle, a light ray, when propagating from point A to point C, takes the path of least time delay in the time domain, and hence, delay in phase in the frequency domain.

Given two points A and C in two different half spaces as shown in Figure 33.4. Then the phase delay between the two points, per Figure 33.4, can be written as[4]

$$P = \mathbf{k}_i \cdot \mathbf{r}_i + \mathbf{k}_t \cdot \mathbf{r}_t \tag{33.2.1}$$

In the above, $\mathbf{k}_i$ is parallel to $\mathbf{r}_i$, so is $\mathbf{k}_t$ is parallel to $\mathbf{r}_t$. As this is the shortest path with minimum phase shift or time delay, according to Fermat's principle, another other path will be longer giving rise to more phase. In other words, if B were to move slightly to another point, a longer path with more phase shift or time delay will ensue, or that B is the stationary point of the path length or phase shift. Specializing (33.2.1) to a 2D picture, then the phase shift as a function of x_i is stationary. This is shown in Figure 33.4, we have $x_i + x_t = \text{const}$. Therefore, taking the derivative of (33.2.1) or the phase change with respect to x_i, assuming that $\mathbf{k}_i$ and $\mathbf{k}_t$ do not change as B is moved slightly,[5]

$$\frac{\partial P}{\partial x_i} = 0 = k_{ix} - k_{tx} \tag{33.2.2}$$

The above yields the law of refraction that $k_{ix} = k_{tx}$, which is just Snell's law; it can also be obtained by phase matching as have been shown earlier. This law was also known in the Islamic world to Ibn Sahl in A.D. 984 [208].

[4]In this course, for wavenumber, we use k and β interchangeably, where k is prevalent in optics and β is used in microwaves.

[5]One can show that as the separations between A, B, and C are large, and if the change in x_i is Δx_i, the changes in $\mathbf{k}_i$ and $\mathbf{k}_t$ are small. The change in phase shift mainly comes from the change in x_i. Alternatively, we can write $P = k_i r_i + k_t r_t$, and let $r_i = \sqrt{x_i^2 + z_i^2}$, and $r_t = \sqrt{x_t^2 + z_t^2}$, and take the derivative with respect to to x_i, one would also get the same answer.

33.2.1 Generalized Snell's Law

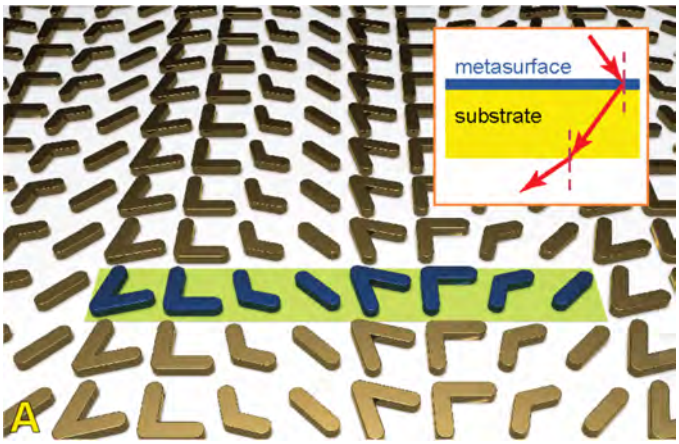

Figure 33.5: A phase screen which is position dependent can be made using nano-fabrication and designed with commercial software for solving Maxwell's equations. In such a case, one can derive a generalized Snell's law to describe the diffraction of a wave by such a surface (courtesy of Capasso's group [218]).

Metasurfaces are prevalent these days due to advances in nano-fabrication and numerical simulation. One of them is shown in Figure 33.5. Such a metasurface can be thought of as a phase screen, providing additional phase shift for the light as it passes through it. Moreover, the added phase shift can be controlled to be a function of position because of advances in nano-fabrication technology and commercial software for numerical simulation.

To model this phase screen, we add an additional function $\Phi(x, y)$ to (33.2.1), namely that

$$P = \mathbf{k}_i \cdot \mathbf{r}_i + \mathbf{k}_t \cdot \mathbf{r}_t - \Phi(x_i, y_i) \tag{33.2.3}$$

Now applying Fermat's principle that there should be minimal phase delay, and taking the derivative of the above with respect to x_i, one gets

$$\frac{\partial P}{\partial x_i} = k_{ix} - k_{tx} - \frac{\partial \Phi(x_i, y_i)}{\partial x_i} = 0 \tag{33.2.4}$$

The above yields that the generalized Snell's law [218] that

$$k_{ix} - k_{tx} = \frac{\partial \Phi(x_i, y_i)}{\partial x_i} \tag{33.2.5}$$

It implies that the transmitted light can be directed to other angles due to the additional phase screen.[6]

[6]Such research is also being pursued by V. Shalaev and A. Boltasseva's group at Purdue U [219].

33.3 Gaussian Beam

We have seen previously that in a source-free medium, using vector and scalar potential formulation, we arrive at

$$\nabla^2 \mathbf{A} + \omega^2 \mu\varepsilon \mathbf{A} = 0 \tag{33.3.1}$$

$$\nabla^2 \Phi + \omega^2 \mu\varepsilon \Phi = 0 \tag{33.3.2}$$

The above are four scalar equations; and the Lorenz gauge

$$\nabla \cdot \mathbf{A} = -j\omega\mu\varepsilon\Phi \tag{33.3.3}$$

connects $\mathbf{A}$ and Φ. We can examine the solution of $\mathbf{A}$ such that

$$\mathbf{A}(\mathbf{r}) = \mathbf{A}_0(\mathbf{r}) e^{-j\beta z} \tag{33.3.4}$$

where $\mathbf{A}_0(\mathbf{r})$ is a slowly varying function while $e^{-j\beta z}$ is rapidly varying in the z direction. (Here, $\beta = \omega\sqrt{\mu\epsilon}$ is the wavenumber.) This is primarily a quasi-plane wave propagating predominantly in the z-direction. We know this to be the case in the far field of a source, but let us assume that this form persists less than the far field, namely, in the Fresnel zone as well. Taking the x component of (33.3.4), we have[7]

$$A_x(\mathbf{r}) = \Psi(\mathbf{r}) e^{-j\beta z} \tag{33.3.5}$$

where $\Psi(\mathbf{r}) = \Psi(x, y, z)$ is a slowly varying envelope function of x, y, and z, whereas $e^{-j\beta z}$ is a rapidly varying function of z when β is large or the frequency is high.

33.3.1 Derivation of the Paraxial/Parabolic Wave Equation

Substituting (33.3.5) into (33.3.1), and taking the double z derivative first in the Laplacian operator ∇^2, we arrive at

$$\frac{\partial^2}{\partial z^2}\left[\underbrace{\Psi(x,y,z)}_{\text{slow}}\underbrace{e^{-j\beta z}}_{\text{fast}}\right] = \left[\frac{\partial^2}{\partial z^2}\Psi(x,y,z) - 2j\beta\frac{\partial}{\partial z}\Psi(x,y,z) - \beta^2\Psi(x,y,z)\right]e^{-j\beta z} \tag{33.3.6}$$

Consequently, after substituting the above into the x component of (33.3.1), making use of the definition of ∇^2, we obtain an equation for $\Psi(\mathbf{r})$, the slowly varying envelope as

$$\frac{\partial^2}{\partial x^2}\Psi + \frac{\partial^2}{\partial y^2}\Psi - 2j\beta\frac{\partial}{\partial z}\Psi + \frac{\partial^2}{\partial z^2}\Psi = 0 \tag{33.3.7}$$

where the last term containing β^2 on the right-hand side of (33.3.6) cancels with the term coming from $\omega^2\mu\varepsilon\mathbf{A}$ of (33.3.1).

[7] Also, the wave becomes a transverse wave in the far field, and keeping the transverse component suffices.

So far, no approximation has been made in the above equation. Since β is linearly proportional to frequency ω, when $\beta \to \infty$, or in the high frequency limit,

$$\left|2j\beta\frac{\partial}{\partial z}\Psi\right| \gg \left|\frac{\partial^2}{\partial z^2}\Psi\right| \tag{33.3.8}$$

where we have assumed that Ψ is a slowly varying function of z within the lengthscale of a wavelength, such that $\beta\Psi \gg \partial/\partial z\Psi$. In other words, (33.3.7) can be approximated by

$$\frac{\partial^2\Psi}{\partial x^2} + \frac{\partial^2\Psi}{\partial y^2} - 2j\beta\frac{\partial\Psi}{\partial z} \approx 0 \tag{33.3.9}$$

The above is called the paraxial wave equation. It is also called the parabolic wave equation.[8] It implies that the $\boldsymbol{\beta}$ vector of the wave is approximately parallel to the z axis, or $\beta_z \cong \beta$ to be much greater than β_x and β_y, and hence, the name.

33.3.2 Finding a Closed Form Solution

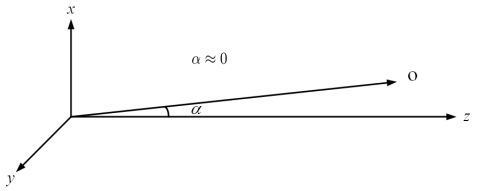

Figure 33.6: Figure showing when the paraxial approximation can be made: when the field is being observed very close to the z axis.

A closed form solution to the paraxial wave equation can be obtained by a simple trick.[9] It is known that

$$A_x(\mathbf{r}) = \frac{e^{-j\beta|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|} \tag{33.3.10}$$

is the exact solution to

$$\nabla^2 A_x + \beta^2 A_x = 0 \tag{33.3.11}$$

as long as $\mathbf{r} \neq \mathbf{r}'$. One way to ensure that $\mathbf{r} \neq \mathbf{r}'$ always is to let $\mathbf{r}' = -\hat{z}jb$, a complex number. Then (33.3.10) is always a solution to (33.3.11) for all $\mathbf{r}$, because $|\mathbf{r}-\mathbf{r}'| \neq 0$ always since $\mathbf{r}'$

[8]The paraxial wave equation, the diffusion equation and the Schrodinger equation are all classified as parabolic equations in mathematical parlance [36, 50, 220, 221].

[9]Introduced by Georges A. Deschamps of UIUC [222].

is complex. Then, we should next make a paraxial approximation to the solution (33.3.10) by assuming that $x^2 + y^2 \ll z^2$. By so doing, it follows that

$$\begin{aligned} |\mathbf{r} - \mathbf{r}'| &= \sqrt{x^2 + y^2 + (z + jb)^2} \\ &= (z + jb)\left[1 + \frac{x^2 + y^2}{(z + jb)^2}\right]^{1/2} \\ &\approx (z + jb) + \frac{x^2 + y^2}{2(z + jb)} + \ldots, \qquad |z + jb| \to \infty \end{aligned} \tag{33.3.12}$$

where Taylor series has been used in approximating the last term. And then using the above approximation in (33.3.10) yields

$$A_x(\mathbf{r}) \approx \frac{e^{-j\beta(z+jb)}}{4\pi(z + jb)} e^{-j\beta \frac{x^2+y^2}{2(z+jb)}} \approx e^{-j\beta z}\Psi(\mathbf{r}) \tag{33.3.13}$$

By comparing the above with (33.3.5), we can identify

$$\Psi(x, y, z) \cong A_0 \frac{jb}{z + jb} e^{-j\beta \frac{x^2+y^2}{2(z+jb)}} \tag{33.3.14}$$

where A_0 is used to absorbed the constants to simplify the expression. By separating the exponential part into the real part and the imaginary part, viz.,

$$\frac{x^2 + y^2}{2(z + jb)} = \frac{x^2 + y^2}{2}\left(\frac{z}{z^2 + b^2} - j\frac{b}{z^2 + b^2}\right) \tag{33.3.15}$$

and writing the prefactor in terms of amplitude and phase gives,

$$\frac{jb}{z + jb} = \frac{1}{\sqrt{1 + z^2/b^2}} e^{j \tan^{-1}(\frac{z}{b})} \tag{33.3.16}$$

We then have

$$\Psi(x, y, z) \cong \frac{A_0}{\sqrt{1 + z^2/b^2}} e^{j \tan^{-1}(\frac{z}{b})} e^{-j\beta \frac{x^2+y^2}{2(z^2+b^2)} z} e^{-b\beta \frac{x^2+y^2}{2(z^2+b^2)}} \tag{33.3.17}$$

The above can be rewritten more suggestively as

$$\Psi(x, y, z) \cong \frac{A_0}{\sqrt{1 + z^2/b^2}} e^{-j\beta \frac{x^2+y^2}{2R}} e^{-\frac{x^2+y^2}{w^2}} e^{j\psi} \tag{33.3.18}$$

where A_0 is a new constant introduced to absorb undesirable constants arising out of the algebra. In the above,

$$w^2 = \frac{2b}{\beta}\left(1 + \frac{z^2}{b^2}\right), \qquad R = \frac{z^2 + b^2}{z}, \qquad \psi = \tan^{-1}\left(\frac{z}{b}\right) \tag{33.3.19}$$

For a fixed z, the parameters w, R, and ψ are constants. It is seen that the beam is Gaussian tapered in the x and y directions, and hence, the name Gaussian beam. Here, w is the beam waist which varies with z, and it is smallest when $z = 0$, or $w = w_0 = \sqrt{\frac{2b}{\beta}}$.

And the term $\exp(-j\beta\frac{x^2+y^2}{2R})$ resembles the phase front of a spherical wave where R is its radius of curvature. This can be appreciated by studying a spherical wave front $e^{-j\beta R}$, and make a paraxial wave approximation, namely, letting $x^2 + y^2 \ll z^2$ to get

$$\begin{aligned} e^{-j\beta R} = e^{-j\beta(x^2+y^2+z^2)^{1/2}} &= e^{-j\beta z\left(1+\frac{x^2+y^2}{z^2}\right)^{1/2}} \\ &\approx e^{-j\beta z - j\beta\frac{x^2+y^2}{2z}} \approx e^{-j\beta z - j\beta\frac{x^2+y^2}{2R}} \end{aligned} \tag{33.3.20}$$

In the last approximation, we assume that $z \approx R$ in the paraxial approximation. We see that the phase of the above wave field, minus the $-j\beta z$ term, is similar in form to the first phase term in (33.3.18). Hence, R in (33.3.18) can be thought of as the radius of curvature of the phase front or wave front.

The phase ψ defined in (33.3.19) changes linearly with z for small z, and saturates to a constant for large z. This underscores the fact that $\Psi(\mathbf{r})$ is a slowly varying function, and also, the phase of the entire wave is due to the $\exp(-j\beta z)$ in (33.3.13) which is rapidly varying when β is large. A cross section of the electric field due to a Gaussian beam is shown in Figure 33.7.

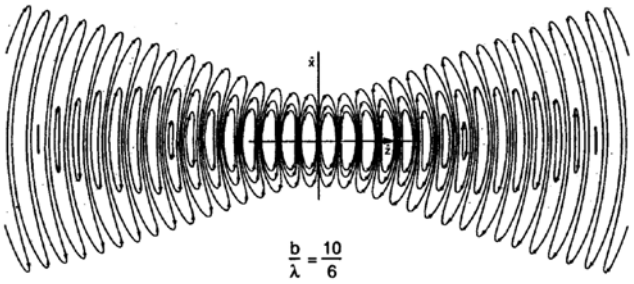

Figure 33.7: Electric field of a Gaussian beam in the x-z plane frozen in time. The wave moves to the right as time increases; here, $b/\lambda = 10/6$ (courtesy of Haus, Electromagnetic Noise and Quantum Optical Measurements [84]). The narrowest beam waist is given by $w_0/\lambda = \sqrt{b/(\lambda\pi)}$.

33.3.3 Other solutions

In general, the paraxial wave equation in (33.3.9) is of the same form as the Schrödinger equation which is of utmost importance in quantum theory. In recent years, the solution of this equation has made use of spill-over knowledge and terms from quantum theory, such as spin angular momentum (SAM) or orbital angular momentum (OAM) even though we are actually in the classical regime. But it is a partial differential equation which can be solved

by the separation of variables just like the Helmholtz wave equation. (Hurrah to the power of separation of variables!) Therefore, in general, it has solutions of the form[10]

$$\Psi_{nm}(x,y,z) \sim \frac{1}{w} e^{-(x^2+y^2)/w^2} e^{-j\frac{\beta}{2R}(x^2+y^2)} e^{j(m+n+1)\tan^{-1}\frac{z}{b}} H_n\left(x\sqrt{2}/w\right) H_m\left(y\sqrt{2}/w\right) \tag{33.3.21}$$

where $H_n(\xi)$ is a Hermite polynomial of order n. The solutions can also be expressed in terms of Laguere polynomials , namely,

$$\Psi_{nm}(x,y,z) \sim \frac{1}{w} e^{-j\frac{\beta}{2R}\rho^2} - e^{-\rho^2/w^2} e^{+j(n+m+1)\tan^{-1}\left(\frac{z}{b}\right)} e^{jl\phi} \left(\frac{\sqrt{2}\rho}{w}\right) L^{n-m}_{\min(n,m)}\left(\frac{2\rho^2}{w^2}\right) \tag{33.3.22}$$

where $L^k_n(\xi)$ is the associated Laguerre polynomial.

These Gaussian beams have rekindled recent excitement in the community because, in addition to carrying spin angular momentum as in a plane wave, they can carry orbital angular momentum due to the complex transverse field distribution of the beams.[11] They harbor potential for optical communications as well as optical tweezers to manipulate trapped nano-particles. Figure 33.8 shows some examples of the cross section (xy plane) field plots for some of these beams. They are richly endowed with patterns implying that they can be used to encode information. These lights are also called structured lights [224].

[10] See F. Pampaloni and J. Enderlein [223]. The author also thanks Bo ZHU for pointing errors in the earlier versions of these equations.

[11] See D.L. Andrew, Structured Light and Its Applications and articles therein [224].

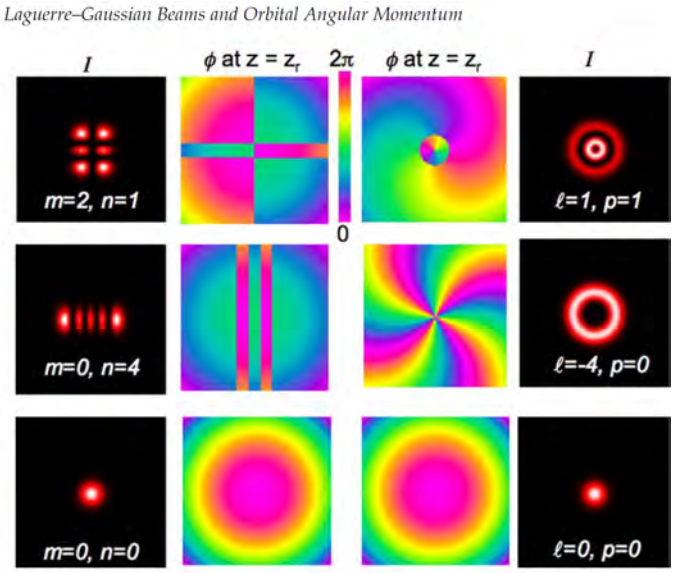

Figure 1.1 Examples of the intensity and phase structures of Hermite–Gaussian modes (*left*) an Laguerre–Gaussian modes (*right*), plotted at a distance from the beam waist equal to the Rayleig range. See color insert.

Figure 33.8: Examples of structured light. It can be used in encoding more information in optical communications (courtesy of L. Allen and M. Padgett's chapter in J.L. Andrew's book on structured light [224]).