Chapter 16. Diffraction. (24 Apr 2021)

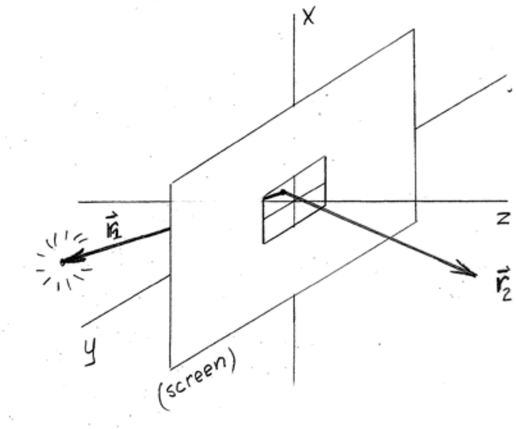
A. Overview.

In diffraction, the wave itself is the source. This is Huyens' Principle, which states that each point of a wavefront is a source for the next wavefront. For plane waves of infinite lateral extent the superposition of these outgoing spherical waves simply generates another plane wave with an appropriate phase delay, so the result (but not the calculation) is trivial. However, if the wavefront is finite, the wave no longer replicates itself but spreads out. This spreading is termed diffraction.

In this chapter, we quantify the above. As with radiation, everything we learn comes from the Green function, although the Green function is used here in a different context. It is used to establish the potential inside a volume of interest in terms of its value on a surface, a context completely familiar from electrostatics. The results are exact, and confirm Huygens' Principle except for deviations in the near-field region. They describe many familiar and less-familiar phenomena: Fraunhofer diffraction, which leads to Rayleigh resolution and synthetic-aperture optics; Fresnel diffraction, which is based on ray tracing and deals with lenses, mirrors, and zone plates; paraxial rays, like those that emerge from your laser pointer; and vortex beams, which introduce the concepts of photon spin, orbital angular momentum, and topological charge, topics that have recently come under intensive investigation.

In contrast to the previous chapter, the wave equation here is homogeneous, and in contrast to electrostatics, the operator is the wave equation, not the Laplacian. We simplify matters by assuming harmonic time dependences throughout. Our approach is distinct from essentially all previous textbook treatments in that we treat diffraction using potentials, rather than fields. By basing the theory on ϕ and $\vec{A}$ instead of $\vec{E}$ and $\vec{H}$, we eliminate the $\nabla(\nabla\cdot)$ term in the wave equation for fields. The results are exact, not approximate, and obtained with far less effort. The approach also resolves Jackson's quandary about which Green function to use: the only viable option is the Dirichlet version. Retardation physics again makes its appearance in that the far-field contribution of the scalar potential includes a term $\hat{k}\cdot\vec{A}$. The calculation of $\vec{E}$ is then reduced to simply multiplying $\vec{A}$ by (ik) and discarding the component parallel to the direction of propagation (note that this is also dimensionally correct.)

The approach provides an excellent opportunity to show how mathematics should proceed, with everything used, nothing left over, and loose ends identified and cleaned up as we go. The result should be increased confidence in applied mathematics.



B. Theory.

The starting point is the configuration shown at the right. An opaque perfectly conducting screen fills the *xy* plane except for an aperture *A*. The aperture is illuminated by a wave incident from the rear ($z < 0$). Our objective is to determine the radiation seen by an observer at $\vec{r}$ in front of the screen. The harmonic time dependence $e^{-i\omega t}$ is understood throughout.

The basic equation is

$$\Box^2(\phi, \vec{A}) = \left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} \right)(\phi, \vec{A}) = 0\,, \tag{16.1}$$

where $\Box^2 = \Box\cdot\Box$ is the d'Alembertian operator, which, as in the three-dimensional case, can be written in divergence form. When dealing with homogeneous equations of this type, the objective is the same as in electrostatics: capitalize on a divergence theorem to convert a "volume" integration of order *n* to a "surface" integration of order $n-1$. Properly done, this allows a quantity, in this case the 4-potential $(\phi, \vec{A})$ in a volume *V*, to be expressed in terms of its value on the closed surface *S* that defines *V*. In Ch. 15 we showed that the original four dimensions of the inhomogeneous version of Eq. (16.1) could be reduced to three by performing the time integration of its Green function exactly, which led to the solution

$$(\phi, \vec{A}) = \frac{1}{c}\int_V d^3r' \left. \frac{(c\rho, \vec{J})}{|\vec{r}-\vec{r}\,'|\left(1 - \frac{\hat{R}}{c}\cdot\vec{v}\right)} \right|_{t'=t_{ret}=t-R/c} \tag{16.2a}$$

$$\cong \frac{1}{c}\int_V d^3r' \left. \frac{(c\rho + \hat{R}\cdot\vec{J}, \vec{J})}{|\vec{r}-\vec{r}\,'|} \right|_{t'=t_{ret}=t-R/c} . \tag{16.2b}$$

$$= (\phi_o + \hat{R}\cdot\vec{A}, \vec{A})_{t_{ret}}\,, \tag{16.2c}$$

where ϕ_o is the electrostatics part of ϕ. The retardation term $\hat{R}\cdot\vec{A}$ in ϕ compensates for the volume distortion of d^3r' as perceived by the observer at $\vec{r}$. It is of the same $1/c$

order as $\vec{A}$, and is essential in far-field radiation, where it eliminates the longitudinal contribution of $\vec{A}$ to $\vec{E}$.

With the assumption of the harmonic time dependence relevant for diffraction, Eq. (16.2b) becomes

$$(\phi, \vec{A}) = \frac{1}{c}\int_V d^3r' \frac{(c\rho + \hat{R}\cdot\vec{J}, \vec{J})}{|\vec{r}-\vec{r}'|} e^{-i\omega t'}\Bigg|_{t'=t_{ret}=t-R/c} \tag{16.3a}$$

$$= \frac{e^{-i\omega t}}{c}\int_V d^3r'(c\rho + \hat{R}\cdot\vec{J}, \vec{J})\frac{e^{ik|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|}, \tag{16.3b}$$

and Eq. (16.1) is reduced to the Helmholtz Equation

$$(\nabla^2 + k^2)(\phi, \vec{A}) = 0. \tag{16.4}$$

Equation (16.3b) is of standard Green-function form, where the Green function is

$$G(\vec{r}, \vec{r}') = \frac{e^{ik|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|}. \tag{16.5}$$

As shown in Ch. 15, this is the Green function of the Helmholtz Equation, satisfying the relation

$$(\nabla_{r'}^2 + k^2)\left(\frac{e^{ik|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|}\right) = -4\pi\delta(\vec{r}-\vec{r}'). \tag{16.6}$$

With this background, we pick up the electrostatics derivation based on Green's First Identity to express $\phi(\vec{r})$ in a volume *V* in terms of its value on a surface *S* that completely encloses *V*. Here, ϕ is the scalar potential or any component of $\vec{A}$, and *S* is the $z = 0$ plane and the hemisphere of radius $R \to \infty$. Consider

$$\int_V d^3r'\nabla\cdot\left(\phi\nabla G\right) = \int_S d^2r'\phi\,\hat{n}\cdot\nabla G = \int_S d^2r'\phi\frac{\partial G}{\partial n}, \tag{16.7}$$

Expanding the divergence yields

$$\int_V d^3r'\nabla\cdot\left(\phi\nabla G\right) = \int_V d^3r'\left(\nabla\phi\cdot\nabla G + \phi\nabla^2 G + k^2\phi G - k^2\phi G\right) \tag{16.8a}$$

$$= \int_V d^3r'\left(\nabla\phi\cdot\nabla G - 4\pi\,\phi\,\delta(\vec{r}-\vec{r}') - k^2\phi G\right) \tag{16.8b}$$

$$= -4\pi\phi(\vec{r}) + \int_V d^3r'\left(\nabla\phi\cdot\nabla G - k^2\phi G\right). \tag{16.8c}$$

In Eq. (16.8a) we accommodate the Helmholtz Equation by adding and subtracting $k^2\phi G$, thereby also allowing the definition of its Green function to be used. Reversing the order of ϕ and *G* and repeating yields

$$\int_V d^3r'\nabla\cdot\left(G\nabla\phi\right)=\int_S d^2r'G\,\hat{n}\cdot\nabla\phi=\int_S d^2r'G\frac{\partial\phi}{\partial n} \tag{16.9a}$$

$$=\int_V d^3r'\left(\nabla G\cdot\nabla\phi+G\nabla^2\phi+k^2G\phi-k^2G\phi\right) \tag{16.9b}$$

$$=\int_V d^3r'\left(\nabla\phi\cdot\nabla G-k^2\phi G\right). \tag{16.9c}$$

Subtracting Eqs. (16.9c) from (16.8c) and doing some algebra gives the result familiar from electrostatics:

$$\phi(\vec{r})=-\frac{1}{4\pi}\int_S d^2r'\left(\phi\frac{\partial G}{\partial n'}-G\frac{\partial\phi}{\partial n'}\right). \tag{16.10}$$

Except for replacing ∇^2 with (∇^2+k^2), the procedure is identical to that done in Ch. 5. With the surface integral involving only the value of ϕ and its normal derivative on *S*, the connection to diffraction is evident. We recall that in Eq. (16.10) ϕ is the scalar potential or any of the three Cartesian components of $\vec{A}$.

The question now is, what is the correct Green function to use? Electrostatics gives us three options: Eq. (16.5), and the Dirichlet and Neumann versions derived from it:

$$G_D(\vec{r},\vec{r}\,')=\frac{e^{ik|R_1}}{R_1}-\frac{e^{ikR_2}}{R_2}; \tag{16.11a}$$

$$G_N(\vec{r},\vec{r}\,')=\frac{e^{ik|R_1}}{R_1}+\frac{e^{ikR_2}}{R_2}; \tag{16.11b}$$

where

$$R_1=|\vec{r}-\vec{r}_1|=\sqrt{(x-x')^2+(y-y')^2+(z-z')^2}; \tag{16.12a}$$

$$R_2=|\vec{r}-\vec{r}_2|=\sqrt{(x-x')^2+(y-y')^2+(z+z')^2}. \tag{16.12b}$$

With any of these three possibilities, both aperture and screen must be in the $z=0$ plane. As usual, $G_D(\vec{r},r')$ and $G_N(\vec{r},r')$ are defined such that the value or normal derivative, respectively, vanish on this plane. Working with fields, Kirchhoff proposed Eq. (16.5). This is clearly the worst possible choice, since it eliminates neither term in Eq. (16.10). Jackson, also working with fields, discusses Eq. (16.5) along with Eqs. (16.11) but draws no conclusions.

We can do better, although an initial assessment is not encouraging. Each of these Green functions has only one singularity in the volume of interest, so this provides no guidance. At the aperture, if either ϕ or $\vec{A}(\vec{r})$ is given, then $(\partial\phi/\partial z)$ or $(\partial\vec{A}/\partial z)$ can easily be calculated, so once again no guidance is provided. Finally, $\partial G/\partial n'$ can be evaluated for all three, again providing no guidance.

The decision is forced by the screen. ϕ is obviously zero on the conductor. $\partial\phi/\partial n$ cannot be specified a priori, but is determined as part of the solution. Thus everywhere except at the aperture we can specify ϕ but not $\partial\phi/\partial n$. Second, from

$$\vec{E} = \frac{i\omega}{c}\vec{A} + \vec{k}\phi, \tag{16.13}$$

the tangential components of $\vec{A}$ are zero on the conductor, since the tangential components of $\vec{E}$ vanish there, and ϕ is already shown to be zero. Finally, the Lorentz-gauge expression $\phi = \hat{k}\cdot\vec{A}$ shows that the normal component A_z of $\vec{A}$ must be zero on the screen as well, since ϕ and the two tangential components of $\vec{A}$ are already zero. Thus if we require ϕ and $\vec{A}$ to be specified on the aperture, then ϕ and $\vec{A}$ are specified everywhere on the $z = 0$ plane. Consequently, the Dirichlet Green function, Eq. (16.11a), is the appropriate choice. We need not consider the infinite hemisphere, since everything vanishes there anyway.

The final step is to evaluate $\partial G/\partial n'$ at $z' = 0$. This is straightforward, and the result is

$$\left.\frac{\partial G_D}{\partial n'}\right|_{z'=0} = -\left.\frac{\partial G_D}{\partial z'}\right|_{z'=0} = \frac{2z}{R}\left(\frac{1}{R} - ik\right)\frac{e^{ikR}}{R} \tag{16.14}$$

where $\hat{n}' = -\hat{z}$, and $R = R_1(z'=0) = R_2(z'=0)$. Therefore, to order $1/c$

$$\phi(\vec{r}) = \frac{1}{2\pi}\int_S d^2r'\left(\hat{R}\cdot\vec{A}\right)\frac{z}{R}\left(\frac{1}{R} - ik\right)\frac{e^{ikR}}{R}, \tag{16.15a}$$

$$\vec{A}(\vec{r}) = \frac{1}{2\pi}\int_S d^2r'\,\vec{A}(\vec{r}\,')\frac{z}{R}\left(\frac{1}{R} - ik\right)\frac{e^{ikR}}{R}, \tag{16.15b}$$

again with $e^{-i\omega t}$ understood. These expressions are completely general. In deriving them we have made *no* approximations beyond restricting the results to order $1/c$. The term (z/R) is called the obliquity factor, and is often written as $(z/R) = \cos\theta$. The term in large brackets contains both near-field $(1/R)$ and far-field (ik) contributions, with the near-field contribution vanishing in the radiation zone.

We have thus arrived at the exact form of Huygens' Principle. The outgoing wave is seen to be a superposition of spherical waves originating from each point on *S*. It differs from the generally accepted version by the obliquity factor z/R and the term $1/R$ in the parentheses. In the next section we show that these expressions indeed reduce properly to $\phi(x, y, 0)$ as $z \to 0$, and for an infinite aperture, convert an incident plane wave into a subsidiary plane wave with an appropriate phase delay. Both calculations show that all terms are necessary. Thus everything is used and nothing is left over, consistent with mathematics done correctly.

To summarize what follows, diffraction is usually treated by approximation. This is the situation except for the next section. In the Fresnel limit, covered in Sec. E, the distance to the observer is typically small or comparable to the size of the aperture or optical element, and the observer is on the z axis. In this case Eq. (16.15b) is approximated as

$$\vec{A}(\vec{r}) \cong \frac{-i}{\lambda} \int_S d^2 r' \, \vec{A}(\vec{r}\,') \frac{e^{ikR}}{R} . \tag{16.16}$$

The $1/R$ term is considered negligible with respect to $(-ik)$, and the obliquity factor is equal to 1. The action occurs in the phase term e^{ikR}, which is evaluated by ray tracing. The results are expressed as the total phase ikR accumulated between source and observer following the different paths. This is the approach commonly used to analyze the optics of mirrors and lenses. Application to the Poisson spot, a historically significant result in the history of electromagnetics, is covered in Sec. F.

Fraunhofer diffraction fits the more usual picture of diffraction, where $|\vec{r}\,'| << |\vec{r}\,|$, the observer is usually on or near the z axis, and far-field approximations are valid. In this case the phase term is expanded as $|\vec{r} - \vec{r}\,'| \cong r - \vec{r} \cdot \vec{r}\,'$, and R in the denominator of the Green function is approximated by r. With these approximations Eq. (16.15b) reduces to

$$\vec{A}(\vec{r}) \cong \frac{-i}{\lambda} \frac{e^{ikr}}{r} \int_S d^2 r' \, \vec{A}(\vec{r}\,') e^{-i\vec{k} \cdot \vec{r}\,'} , \tag{16.17}$$

that is, the diffracted beam is the Fourier transform of the source and aperture. This approximation is covered in Sec. D.

Finally, the paraxial-ray approach discussed in Sec. H involves a new set of assumptions that describe Gaussian beams. These lead to different expressions and are discussed in detail in that section.

In most of what follows, the focus is on $\vec{A}$, not ϕ, because the relevant part of ϕ for diffraction is expressed in terms of $\vec{A}$ anyway. This is an interesting case of role reversal relative to statics and classical treatments of induction, where ϕ is important and $\vec{A}$ serves mainly as a means of solving equations efficiently.

C. Reduction to known special cases.

In any derivation that claims to be general it is always good practice to verify that it reduces to known special cases. This ensures that the derivation contains no structural errors, and generates more confidence in the result.

We consider first the limit $z \to 0$. Placing the origin at $(x, y, 0)$, the structure of Eq. (16.15) shows that the relevant part of the integrand corresponds to $R \to 0$ as well. Using cylindrical coordinates, Eq. (16.15) then reduces to

$$\lim\big|_{z\to 0} \phi(x, y, z) = \lim\big|_{z\to 0} \left(\frac{z}{2\pi} \int_0^\infty \rho' d\rho' \int_0^{2\pi} d\phi' \phi(\vec{r}\,') \frac{1}{R^3} \right) \tag{16.18a}$$

$$= \lim\Big|_{z\to 0}\left(\frac{z}{2}\int_S d(\rho'^2)\,\phi(\vec{r}\,')\frac{1}{(\rho'^2+z^2)^{3/2}}\right) = \lim\Big|_{z\to 0}\left(\frac{2z}{2z}\phi(x,y,z)\right) \tag{16.18b}$$

$$= \phi(x,y,0)\,. \tag{16.18c}$$

An additional calculation shows that in the limit $z \to 0$ the normal derivative vanishes as well. This is consistent with our experience with the Dirichlet Green function from electrostatics. The obliquity factor plays an essential role in Eq. (16.15b), since Eq. (16.18c) cannot be obtained without it.

The solution for plane-wave illumination $\vec{A}(\vec{r},t) = \vec{A}_o e^{ikz-i\omega t}$ of an infinite aperture involves less familiar functions, but also yields the expected result. With the aperture covering the entire $z = 0$ plane no approximations are possible, and Eq. (16.15b) must be treated exactly. Converting Eq. (16.15b) to cylindrical coordinates, we note first that $R = \sqrt{\rho'^2 + z^2}$, so

$$RdR = \frac{2\rho' R\,d\rho'}{2\sqrt{\rho'^2+z^2}} = \rho' d\rho'\,. \tag{16.19}$$

With this substitution for $\rho' d\rho'$ and the time dependence $e^{-i\omega t}$ understood, performing the φ' integration eliminates the numerical prefactor, leaving

$$\vec{A}(x,y,z) = -\vec{A}_o z\int_z^\infty dR\frac{1}{R}\left(ik - \frac{1}{R}\right)e^{ikR} \tag{16.20a}$$

$$= \vec{A}_o kz\int_{zk}^\infty d\xi\left(\frac{i}{\xi}\cos\xi - \frac{\sin\xi}{\xi} - \frac{1}{\xi^2}\cos\xi - i\frac{1}{\xi^2}\sin\xi\right) \tag{16.20b}$$

$$= -\vec{A}_o kz\left(iCi(\xi) - Si(\xi) + \frac{\cos\xi}{\xi} + Si(\xi) + i\frac{\sin\xi}{\xi} - iCi(\xi)\right)_{kz}^{\infty} \tag{16.20c}$$

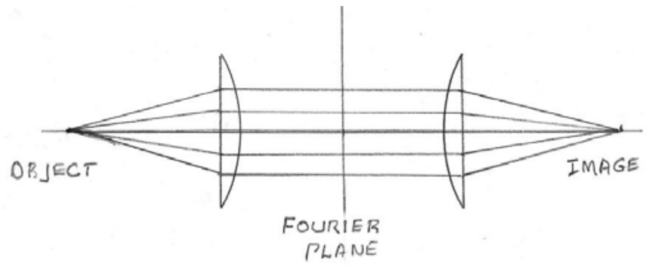
$$= \vec{A}_o e^{ikz}\,, \tag{16.20d}$$

where $\xi = kR$. Thanks to the identical cancellation of the cosine- and sine-integral functions $Ci(\xi)$ and $Si(\xi)$, the result is precisely what we expect for an observer located at z if the phase is referenced to zero at $z = 0$. The exact form of Huygens' Principle is again verified.

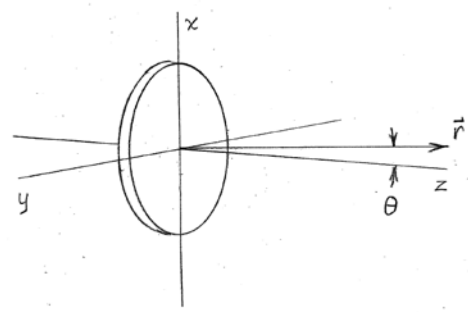
Finally, while Eq. (16.15b) is mathematically exact to within the assumptions used to derive it, any treatment of diffraction that does not include radiative contributions from charges induced at the edges of the aperture is necessarily itself only an approximation. However, calculating the contribution of these charges induced on the edge of the aperture is exceedingly difficult. Born and Wolf give an example of a conducting screen that fills the half-plane $y \le 0$. This configuration can be solved analytically, and takes up a large fraction of their Ch. 11. Those interested in these details are referred there.

D. Fraunhofer diffraction and Rayleigh resolution.

We consider Fraunhofer diffraction, where the observer is far enough away compared to the size of the aperture so the viewing angle can be considered constant and far-field approximations apply. Astronomy is one obvious application. Performing the usual expansion of e^{ikR} leads to Eq. (16.17). This is just an outgoing spherical wave with a multiplicative term that is the product of the obliquity factor, the Fourier transform of the illumination field, and λ^{-1}. With the observer near the *z* axis, the typical situation, the obliquity factor $(\cos\theta)$ can be discarded as well.



The result is a two-dimensional Fourier transform of the potential in the aperture, depending not only on the incident field at the aperture but also on the shape of the aperture itself. Simple examples, i.e., the determination of the diffracted potential $\vec{A}(\vec{r},t)$ for a rectangular aperture illuminated by a plane wave, are left as a homework assignments. In optics, this analog 2D Fourier transform is occasionally used in image processing. If a plano-convex lens is employed to focus the aperture at infinity, as indicated in the diagram, the result in the "Fourier plane" is an optical Fourier transform of the source. This Fourier-plane image can be processed, for example by phase-shifting or deleting specific areas, and the Fourier-processed image reconstructed by a second lens.



Fraunhofer diffraction is also the basis of the Rayleigh definition of resolution of optical systems. The mathematics is somewhat difficult, but manageable. Consider a screen with a circular aperture of radius *a*, as shown at the right. We assume that the observer at $\vec{r}$ is in the *xz* plane and close enough to the symmetry axis so the obliquity factor can be ignored. It is customary to describe the off-axis part of $\vec{r}$ as a polar angle. Thus

$$\vec{r} = r\,(\hat{x}\sin\theta + \hat{z}\cos\theta) \cong r\,(\hat{x}\sin\theta + \hat{z})\,; \tag{16.21a}$$

$$\vec{k} = k\,(\hat{x}\sin\theta + \hat{z}\cos\theta) \cong k\,(\hat{x}\sin\theta + \hat{z})\,; \tag{16.21b}$$

and

$$\vec{r}\,' = \rho'(\hat{x}\cos\varphi' + \hat{y}\sin\varphi')\,. \tag{16.21c}$$

Therefore, the diffracted beam is

$$\vec{A}(\vec{r},t) = -\frac{i}{\lambda}\vec{A}_o\frac{e^{ikr-i\omega t}}{r}\int_0^a \rho'\,d\rho'\int_0^{2\pi} d\varphi' e^{-ik\rho'\sin\theta\cos\varphi'} \tag{16.22a}$$

$$= -\frac{i}{\lambda}\vec{A}_o \frac{e^{ikr-i\omega t}}{r}\int_0^a \rho' d\rho' \int_0^{2\pi} d\varphi' \cos(k\rho' \sin\theta \cos\varphi') . \tag{16.22b}$$

The integral over φ' is standard:

$$\int_0^{2\pi} d\varphi' \cos(z\cos\varphi') = 2\pi\, J_o(z) , \tag{16.23}$$

where $J_o(z)$ is the Bessel function of order zero. The intermediate result is

$$\vec{A}(\vec{r},t) = -\frac{i}{\lambda}\vec{A}_o \frac{e^{ikr-i\omega t}}{r}\int_0^a \rho' d\rho' J_o(k\rho' \sin\theta) . \tag{16.24a}$$

$$= -\frac{i}{\lambda}\vec{A}_o \frac{e^{ikr-i\omega t}}{r(k\sin\theta)^2}\int_0^{ka\sin\theta} u' du' J_o(u') . \tag{16.24b}$$

The last integral is not standard, although the result is well known. We arrive it using the recursion relation for integer Bessel functions, specifically

$$J_o(x) = \frac{J_1(x)}{x} + \frac{dJ_1(x)}{dx} \tag{16.25}$$

(see, e.g., M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions (National Bureau of Standards, Applied Mathematics Series 55, (1964)). This leads to

$$xJ_o(x) = J_1(x) + x\frac{dJ_1(x)}{dx} , \tag{16.26}$$

which in turn leads to

$$\frac{d}{dx}(xJ_1(x)) = J_1(x) + x\frac{dJ_1(x)}{dx} = xJ_o(x) . \tag{16.27}$$

The result is therefore

$$\vec{A}(\vec{r},t) = -\frac{ia^2}{\lambda}\vec{A}_o \frac{e^{ikr-i\omega t}}{r}\frac{J_1(ka\sin\theta)}{ka\sin\theta} . \tag{16.28}$$

Because $J_1(z) \cong (z/2)$ for small z, the expression is regular in the limit that $\theta \to 0$.

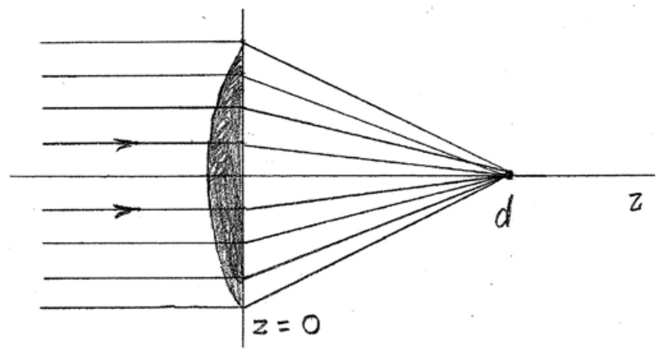
The Rayleigh criterion is defined as the value of $\sin\theta$ at the first zero $x = 3.832$ of $J_1(x)$. This was selected as the minimum separation for two closely spaced absorption lines where the two lines could be distinguished. Specifically, this is the separation where the maximum of one line falls on the zero of the second. With $k = (2\pi/\lambda)$, this is

$$\sin\theta = 0.61\lambda/a . \tag{16.29}$$

While the calculations are done here for monolithic optical elements, this is not an essential restriction. So-called synthetic aperture configurations where the segments are separated, sometimes widely separated, are commonly used in radio astronomy. In principle this allows the entire diameter of the Earth to be the width of the focusing element, and resolution is enhanced accordingly. We have not yet reached the point where we can achieve this extreme resolution enhancement in the optical frequency range, but this technology can be anticipated. A calculation of the diameter needed to resolve Proxima Centauri b, a planet slightly larger than the Earth that orbits the red dwarf Proxima Centauri every 11.2 Earth days, is left as a homework assignment.

E. Fresnel diffraction: lenses, mirrors, and zone plates.

The second general application of Eq. (16.15b) is Fresnel diffraction. In Fresnel diffraction, the size of the aperture is comparable to the distance to the observer, so no small-aperture expansion is possible. In this situation the focus becomes the center of attention, and properties are determined by ray tracing. In these calculations, the accumulated phase along a ray is determined and compared to the accumulated phases of other rays. If all are identical at a given point, the result is constructive interference, and a focus is realized. The lateral (resolution) and longitudinal (depth-of-field) extent of the focus can be determined by determining how far it is possible to move the observer from the focal point before destructive interference sets in.



Rays of course are a geometric concept, with accumulated phase determined with the help (if necessary) of Huygens' Principle. The constructive-interference definition is more general than the Abbé criterion, which defines a focus as that point where all rays leaving a source at a particular time arrive simultaneously. This is in turn a consequence of Fermat's Principle, which states that any ray between source and focus follows the fastest route. The negative-index Veselago lens passes the constructive-interference test but fails the Abbé criterion.

In the following calculations we assume for simplicity normal-incidence, plane-wave illumination of the aperture so $\vec{A}(\vec{r}\,')$ is constant over the aperture, which can also be an optical element such as a lens or mirror. Then for an observer located at $\vec{r} = \hat{z}d$, Fresnel diffraction in this case reduces Eq. (16.16) to

$$\vec{A}(\hat{z}d) = -\frac{i}{\lambda d}\vec{A}_o \int_S d^2r' e^{ikR}, \tag{16.30}$$

where $R = |\vec{r}\,' - \hat{z}d|$ and $e^{-i\omega t}$ is understood. The obliquity factor is generally ignored.

A path-length interpretation of the above integral shows that if $\hat{z}d$ is a focus, then any ray from the source to $\hat{z}d$ must cover the same *optical* distance, that is, exhibit the same phase delay from the source as every other ray. In transparent material such as that in a lens, the optical distance is defined in terms of accumulated phase as

$$\phi = kd = 2\pi n_o d / \lambda = 2\pi d_{opt} / \lambda , \tag{16.31}$$

hence $d_{opt} = n_o d$, where d is the metric distance in the medium and n_o its refractive index.

Consider first a lens illuminated from the left by a plane wave with wave fronts perpendicular to z, as indicated in the diagram on the preceding page. The metric distance for the ray at the z axis is shorter than that for a ray on the perimeter, which we construct by invoking Huygens' Principle. Without the lens, rays along these two paths would be out of phase and a focus would not occur. However, by inserting material with a refractive index $n_o > 1$ to accumulate additional phase in the central ray, the phase delays are made equal and a focus results.

The shape of the lens is easily calculated, at least to lowest order. If the element is circular, the usual case, then the longest metric path in the diagram, measured from $z = 0$, is $d_{\max} = \sqrt{a^2 + d^2}$. Thus the amount of lens material that must be added to a ray arriving at a radius ρ is

$$\sqrt{a^2 + d^2} = \sqrt{\rho^2 + d^2} + (n-1) d_l(\rho) , \tag{16.32}$$

or

$$d_l = \frac{1}{(n-1)} \left(\sqrt{a^2 + d^2} - \sqrt{\rho^2 + d^2} \right) \tag{16.33a}$$

$$\cong \frac{1}{2d(n-1)} \left(a^2 - \rho^2 \right) \tag{16.33b}$$

if a and ρ are both small relative to d. This is the standard parabolic-thickness profile for small plano-convex lenses. Incidentally, to minimize aberrations, the curved surface is placed toward the incident radiation rather than the focus, which for most people is counterintuitive.

By considering extreme rays, the same geometric construction can be used to estimate resolution and depth of field, then establish scales by comparing the resulting phase differences to the Rayleigh criterion. Let the observer be moved off-axis by a distance ρ. The phase difference between two diametrically opposing extremal rays is

$$\Delta\phi = k \left(\sqrt{(a+\rho)^2 + d^2} - \sqrt{(a-\rho)^2 + d^2} \right) \tag{16.34a}$$

$$\cong \frac{2ka\rho}{\sqrt{a^2 + d^2}} . \tag{16.34b}$$

If we define a cutoff at the point where destructive interference begins, then $\Delta\phi = \pi$. In Rayleigh-criterion form, this leads to

$$\rho_F = \frac{\pi}{2k}\frac{\sqrt{a^2+d^2}}{a}\,. \tag{16.35}$$

The Rayleigh equivalent, which assumes $a << d$, is

$$\rho_R = \frac{3.83}{k}\frac{d}{a}. \tag{16.36}$$

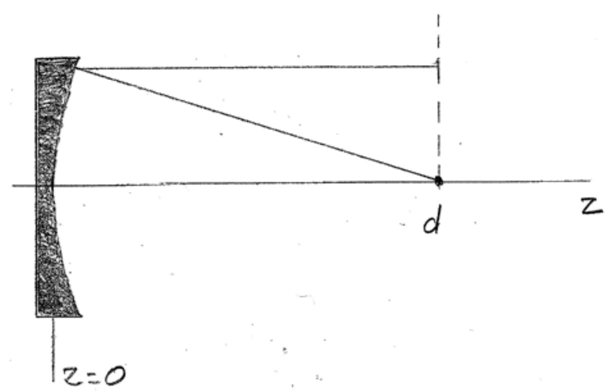
The calculation shows that the Rayleigh criterion is better approximated using 2π instead of π for the phase difference of the extremal rays.

The depth-of-field calculation is similar, although the observer remains on-axis and the difference is now taken between the central ray and an extremal ray for a given displacement Δd down the symmetry axis. This difference is

$$\Delta d = \frac{\pi}{k}\left(1-\frac{d}{\sqrt{a^2+d^2}}\right) \tag{16.37a}$$

$$\cong \frac{\pi}{k}\frac{a^2}{2d^2} \tag{16.37b}$$

in the limit of small *a*. The Rayleigh criterion does not specify depth-of-field, but if we use the same condition here, π would be replaced by 3.83. For large *f*-numbers (d/a ratios), the depth of field is second order in a/d , as opposed to resolution. Therefore, as every photographer knows, depth-of-field tends to be a factor only with lenses of small *f*-numbers.



The second common way to achieve a focus is to use a mirror. The next diagram illustrates the calculation. We assume that the light is incident from infinity on the right. Potential path-length differences must now be accommodated geometrically. Using the perpendicular to the *z* axis passing through *d* as the reference, all path lengths must equal that of the on-axis ray, or 2*d*. Letting Δz be the distance between the $z = 0$ plane and the front surface of the mirror, the length reduction is exactly

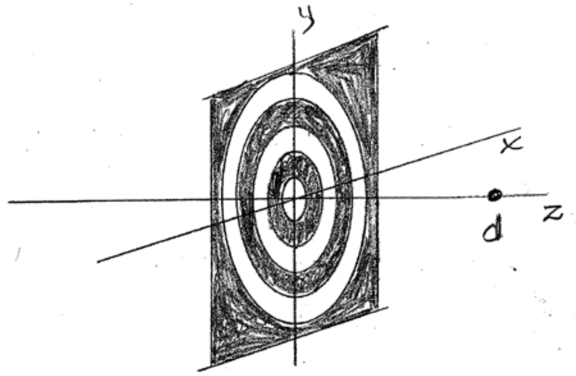
$$2d = (d-\Delta z) + \sqrt{\rho^2 + (d-\Delta z)^2}\,. \tag{16.38}$$

Solving this equation for Δz yields the well-known parabolic shape

$$\Delta z = \frac{\rho^2}{4d}\,. \tag{16.39}$$

This result is exact. Resolution and depth-of-field calculations follow similarly to that of the lens.

The Fresnel zone plate is a third method of achieving a focus. This focus is approximate, and obtained by dividing the screen into radial sectors alternating between open and closed, as shown in the figure, with 2π phase shift between adjacent pairs. The goal is to achieve a phase range of 0 to π or a 2π multiple thereof for all rays transmitted through the screen while blocking all rays with phase shifts of π to 2π or a 2π multiple thereof. The approach is inefficient, but in regions of the spectrum where lenses do not exist and mirrors may be even less efficient, this may be the only alternative.



The formal theory is as follows. Consider a single open band extending from ρ_n to ρ_{n+1}. We wish to maximize

$$\vec{A}(\hat{z}d) = -\frac{i}{\lambda\, d}\vec{A}_o \int_{\rho_n}^{\rho_{n+1}} d^2r' e^{ikR} = -\frac{i}{2\lambda\, d}\vec{A}_o \int_{\rho_n}^{\rho_{n+1}} d\rho'^2 \int_0^{2\pi} d\varphi' e^{ikd} e^{ik\rho'^2/2d}, \tag{16.40a}$$

$$\approx -\frac{i\pi}{\lambda d} e^{ikd}\,\vec{A}_o \int_{\rho_n}^{\rho_{n+1}} d\rho'^2 \left(\cos\frac{k\rho'^2}{2d} + i\sin\frac{k\rho'^2}{2d}\right) \tag{16.40b}$$

$$= -\frac{i\pi}{\lambda d} e^{ikd}\,\vec{A}_o \left(\frac{\sin(k\rho'^2/2d)}{k/2d} - i\,\frac{\cos(k\rho'^2/2d)}{k/2d}\right)_{\rho_n}^{\rho_{n+1}}, \tag{16.40c}$$

where in the second line we performed a first-order expansion of $R = \sqrt{\rho'^2 + d^2}$. Deciding that we can live with the additional 90° phase shift that comes with the last term, the edges of the zones are defined by

$$\frac{k\rho_n^2}{2d} = n\pi, \tag{16.41}$$

where *n* is an integer. The open zones begin on an even *n* and terminate on an odd *n*, with the opaque regions reversing the order. The vector potential at *d* arising from any of the open rings is therefore

$$\vec{A}_n(\hat{z}d) = -\frac{\pi}{\lambda d} e^{ikd}\vec{A}_o\,\frac{4d}{k}$$
$$= 2\vec{A}_o e^{ikd}. \tag{16.42}$$

Thus each ring has the same area, and the vector potential on the *z* axis at a distance *d* from the zone plate per open zone is independent of *d* and is twice that which would occur if the zone plate were not there. Thus *n* open zones result in an enhancement of 2*n*, which when converted to an intensity becomes $4n^2$.

It is worth comparing the above result to a lens with the same area. Using the same small-term expansion, the 2*n*th radius, which is appropriate for a lens of the same radius as a zone plate, is $\rho_{2n}^2 = 2n\lambda d$. Performing the integral with *no* phase shift, which defines a properly designed lens, we obtain

$$\vec{A}_{2n,lens}(\hat{z}d) = -\frac{i\pi}{\lambda d} e^{ikd} \vec{A}_o \pi \rho_{2n}^2 = -\frac{i\pi}{\lambda d} e^{ikd} \vec{A}_o \pi (2n\lambda d) = -i\vec{A}_o e^{ikd} (2\pi^2 n) .$$

$$= -i\vec{A}_o e^{ikd} (2\pi^2 n) \qquad (16.43)$$

Thus the enhancement factor for the lens is π^2 times that of a Fresnel zone plate of the same overall area, nearly an order of magnitude larger. Consequently, zone plates are not about to replace lenses any time soon.

A fourth approach is the Fresnel lens. This is a Fresnel zone plate with no opaque rings but where each ring is itself a lens segment. The advantage of the Fresnel lens is that it is much thinner than a standard lens, although there is some cost in image quality.

Considerable attention is now being directed today to a fifth approach, where deposited thin films are patterned on the scale of the wavelength of light to control optical beams by diffraction in either transmission or reflection. The prototype is obviously the diffraction grating, where the surface of an initially smooth blank is scribed with grooves of a particular spacing and shape to ensure efficient scattering of light in desired directions depending on wavelength. The thin-film versions consist of transparent layers of refractive indices $n > 1$ deposited on smooth substrates. These films are then patterned to achieve the desired behavior. This is an engineering, not a physics, problem: the physics is all contained in the diffraction expression at the start of this section. However, present capabilities go well beyond what could have been envisioned by Fresnel.

F. Poisson's Spot.

Poisson's Spot is an example of Fresnel diffraction with considerable historical significance, and thus justifies its own section. It commemorates a famous incident in optics that resulted from a competition held by the French Academy in 1818. One of the papers submitted was written by Fresnel, who argued rather forcefully that light consisted of waves. Poisson did not agree, and being a mathematician of no small talent, showed that the interpretation proposed by Fresnel predicted a spot of light on the symmetry axis in the shadow cone of an opaque disc inserted in an incoming plane wave. From Poisson's perspective the appearance of such a spot in a shadow cone was obvious nonsense, enough in fact to disqualify the paper from the competition. But this was an easy experiment to run, so Arago, the chair of the committee tried it, and sure enough, the spot was there. This obviously thoroughly embarrassed Poisson, because instead of disproving the wave theory his insight had in fact proved the wave theory to be correct.

To compound Poisson's embarrassment, the spot became known as Poisson's Spot in honor of his successful prediction.

We can evaluate $\vec{A}(\vec{r})$ on the symmetry axis for this configuration using the formalism developed above. This configuration is complementary to the aperture calculation in that the aperture is blocked and the rest of the screen removed. Consequently, far-field approximations do not apply; the phase must be treated exactly. For an incoming plane wave at amplitude $\vec{A}_o$ at $z = 0$, we have

$$\vec{A}(r) = -\frac{i}{\lambda}\vec{A}_o \int_a^{\infty} \rho' d\rho' \int_0^{2\pi} d\varphi' \frac{e^{ikR}}{R}. \tag{16.44}$$

where $R = \sqrt{\rho'^2 + z^2}$. Poisson recognized Eq. (16.19), which enables Eq. (16.44) to be rewritten and solved in closed form:

$$\vec{A}(\vec{r}) = -\frac{i}{\lambda} 2\pi \vec{A}_o \int_a^{\infty} RdR \frac{e^{ikR}}{R} \tag{16.45a}$$

$$= \vec{A}_o e^{ik\sqrt{a^2+z^2}}. \tag{16.45b}$$

It is now evident why Poisson objected so strenuously. The magnitude of the field on the *z* axis is not only independent of *z* but also of *a*! Neither using a bigger disc nor moving *d* further back along the *z* axis changes the on-axis intensity – it only affects the absolute phase, which in an intensity measurement is irrelevant. Of course, the result is totally counterintuitive, so perhaps Poisson can be excused.

However, the above analysis applies only to values *on* the *z* axis, which is a set of measure zero. To evaluate practical spot sizes and intensities, we must eliminate some of the above approximations and integrate the intensity over the plane where the spot is being measured. For off-axis points the values of *a* and *d* clearly become important.

Of course, we know now that this was not the last word on the topic. In 1905 Einstein showed that the photoelectric effect could be interpreted only if the energy in a light beam were carried as particles (photons). Thus the physics came full circle. Of course, we now recognize that electromagnetic radiation can have properties of both waves and particles, although our understanding of why is probably not that much advanced beyond Newton.

G. Gaussian beams: the paraxial-ray approximation.

Gaussian beams are familiar to anyone who has ever worked with lasers, and in particular laser pointers. Unlike plane waves, the lateral extent of these beams is limited. They can be viewed as continuum superpositions that lead to a finite cross section. Gaussian beams have the property that their cross sections and divergences both have minimum values. Higher-order versions include Gauss-Hermite and Gauss-Laguerre beams. A Gauss-Hermite beam is described by a Gaussian base multiplied by a Hermite polynomial, giving it a rectilinear cross-sectional intensity pattern. In Gauss-Laguerre beams the symmetry is circular. Higher-order Gauss-Laguerre beams have zero intensity at their center, and for this reason are also called vortex beams. These are currently under investigation both for their physics and the possibility that they may considerably increase transmission capabilities of optical fibers. Bessel beams form a third category with the singular property that they do not diffract, although this ideal situation can only be approximated in practice.

While the properties of these beams are easy enough to describe qualitatively, their mathematics is formidable. Accordingly, I quote rather than derive solutions, although the basic theory, the paraxial-ray approximation, gets the attention it deserves. This approximation is analogous to the WKB approximation of quantum mechanics, and the resulting differential equation will be recognized as the diffusion equation. The strategy is to isolate the rapidly varying part of the function, the plane-wave factor, and derive a simpler equation for the rest.

The development proceeds as follows. Consider a wave propagating in the *z* direction with an intensity that decreases radially from its center. Our objective is to derive an equation that describes its properties. Because the wave equation is homogeneous, this new term must be compensated by introducing another spatial variation. We let this occur in the *z* direction, so that the component of the potential under investigation no longer depends on *z* as e^{ikz}, but as

$$\phi(\vec{r},t) = u(\rho,z)e^{ikz-i\omega t} . \tag{16.46}$$

Substituting this expression in the wave equation we obtain

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)u(\rho,z)e^{ikz-i\omega t} = \left(\nabla_t^2 + \frac{\partial^2}{\partial z^2} + k^2\right)u(\rho,z)e^{ikz-i\omega t} \tag{16.47a}$$

$$= e^{ikz-i\omega t}\left(\nabla_t^2 + (\frac{\partial^2}{\partial z^2} + 2ik\frac{\partial}{\partial z} - k^2) + k^2\right)u(\rho,z) \tag{16.47b}$$

$$\cong e^{ikz-i\omega t}\left(\nabla_t^2 + 2ik\frac{\partial}{\partial z}\right)u(\rho,z) = 0 , \tag{16.47c}$$

where ∇_t^2 is the transverse part of the Laplacian. The essential approximation is that $u(\rho,z)$ varies slowly enough with *z* so $\partial^2 u/\partial z^2$ can be discarded.

With the k^2 terms cancelling, what remains is the *paraxial wave equation*

$$\left(\nabla_t^2 + 2ik\frac{\partial}{\partial z}\right)u(\rho,z) = 0 . \tag{16.48}$$

This is a second-order differential equation in two variables and a first-order differential equation in a third, and thus has the diffusion-equation structure. This is also the structure of the two-dimensional Schroedinger equation of quantum mechanics, although the second variable in Schroedinger's Equation is *t*, not *z*.

Someone who is significantly better at math than I ever will be obtained an analytic solution by performing algebra on the trial function with a Gaussian lateral dependence. The result is

$$\phi(\vec{r},t) = \frac{4\pi}{\sqrt{1+z^2/z_R^2}} e^{-\rho^2/w_o^2(1+(z^2/z_R^2))} e^{ik\rho^2 z/2(z^2+z_R^2)-i\tan^{-1}(z/z_R)} e^{ikz-i\omega t} . \tag{16.49}$$

The solution contains two parameters, the beam waist w_o and the focal length z_R. These parameters, and the nature of Eq. (16.49), can be best understood by considering special cases.

First, let $z=0$. Then

$$\phi(\vec{r},t) = 4\pi e^{-\rho^2/w_o^2} e^{-i\omega t} . \tag{16.50}$$

The meaning of w_o is now apparent: it describes the minimum diameter, which is realized at $z=0$, and thus the terminology beam waist. For sufficiently large *z*, Eq. (16.49) shows that the diameter expands linearly in *z* thereafter. We also note that at $z=0$ the phase is independent of ρ. Hence $z=0$ is a planar surface of constant phase.

The next simple case occurs at $z=\pm z_R$. To investigate behavior there, let $z=-z_R+z'$, where z' is small. Assume further that ρ^2 is small, so any appearance of z' in the first phase factor can be ignored. Next, if $z_R >> \lambda$, which is almost always the case, the contribution of z' from the inverse tangent term can be ignored relative to ikz'. Then Eq. (16.49) reduces to

$$\phi(\vec{r},t) \simeq 2\sqrt{2}\pi\, e^{-\rho^2/2w_o^2}\, e^{-ik\rho^2/(4z_R)} e^{i\pi/4} e^{ik(-z_R+z')-i\omega t} . \tag{16.51}$$

The interesting terms are those connecting ρ^2 and z', because these define a surface of constant phase. Bringing these terms together, this surface is a parabola defined by

$$k\left(-\frac{\rho^2}{4z_R}+z'\right)=0 . \tag{16.52}$$

Recall that ρ and z' are both small.

To interpret Eq. (16.51), consider the equation of a spherical surface of radius *R* in cylindrical coordinates, centered at $z=-R$. The defining equation is

$$\rho^2+(-R+z')^2 = R^2 . \tag{16.53}$$

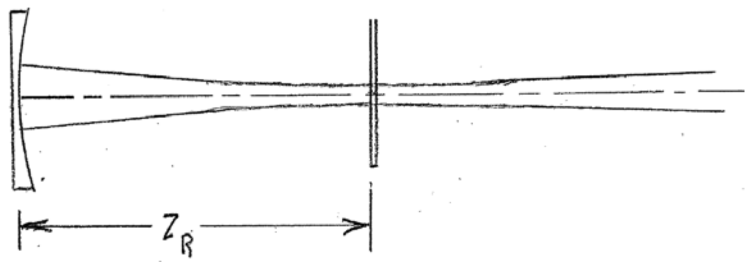
Expanding Eq. (16.53) to first order in *z* yields

$$\rho^2 + 2z'R = 0 . \tag{16.54}$$

But this is identical to Eq. (16.52) if

$$R = 2z_R . \tag{16.55}$$

Thus the surface in Eq. (16.52) is approximately a segment of a sphere with focal length z_R. Treating this surface as a reflector, it superposes source and focus at $z = 0$, as shown in the diagram. The reason that the beam does not converge to a point is diffraction.



Equations (16.50) and (16.52) provide the theory underlying laser cavities that generate Gaussian beams, such as laser pointers. A semitransparent plane mirror is placed at the exit end of the cavity and a spherical mirror at the other. The beam waist occurs at the planar end. A sketch is shown above.

Not surprisingly, a host of solutions exist for the paraxial-ray equation. These are all based on the Gaussian radial dependence, but exhibit different lateral variations. If the paraxial-ray equation is written in Cartesian coordinates, the appropriate functions are Hermite polynomials, of parabolic-quantum-well fame, and the resulting beams are Hermite-Gaussian modes. The basic Hermite-Gaussian functional form (from Wikipedia - Gaussian Beams) is

$$\phi(\vec{r},t) = \frac{w_o}{w(z)} H_l\left(\frac{\sqrt{2}x}{w(z)}\right) H_m\left(\frac{\sqrt{2}x}{w(z)}\right) e^{-(x^2+y^2)/w^2(z)} e^{ik(x^2+y^2)/2R(z)} e^{i\psi(z)} e^{ikz-i\omega t}, \qquad (16.56)$$

where the $H_l(\xi)$ are Hermite polynomials. A set of 12 patterns is shown at the top of the next page.

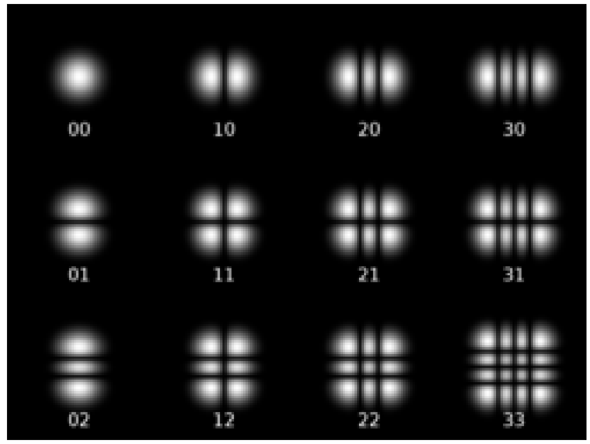
A second class of solutions are cylindrically symmetric and are based on Laguerre polynomials. The basic Laguerre-Gaussian form is

$$\phi(\vec{r},t) = \frac{C_{mp}^{LG}}{w(z)} \frac{\sqrt{2}r}{w(z)} L_m^p\left(\frac{2\rho^2}{w^2(z)}\right) e^{-im\varphi} e^{-\rho^2/w^2} e^{ik\rho^2/2R(z)} e^{i\psi(z)} e^{ikz-i\omega t} \qquad (16.57)$$

where the $L_p^m(\xi)$ are the associated Laguerre polynomials and $|m| \leq p \geq 0$ is an integer or zero. These are characterized as having circularly symmetric envelopes, and for $m \geq 1$ exhibiting zero intensity at the center, as shown below. The lowest modes are shown on the next page, with the first integer being *p* in the above expression and the second being *m*. The $m \geq 1$ modes carry orbital angular momentum, as also discussed below.

The most general class of solutions involve hypergeometric functions, but we'll pass on these.

Laguerre beams of a given *m* are usually generated by passing a (0,0) Gaussian beam through spiral wave plates where the retardation increases around the symmetry axis from zero at $\varphi = 0$ to $2\pi m$ at $\varphi \cong 2\pi$. The phase of the beam is therefore azimuthally dependent, increasing from 0 to $2\pi m$ in the above range. For $m \neq 0$ the result is a vortex beam, that is, one that is forced to have zero intensity on the *z* axis. This follows directly from Eq. (16.57). Letting the incoming light be linearly polarized along *x* and the observer be on the *z* axis, Eq. (16.57) yields

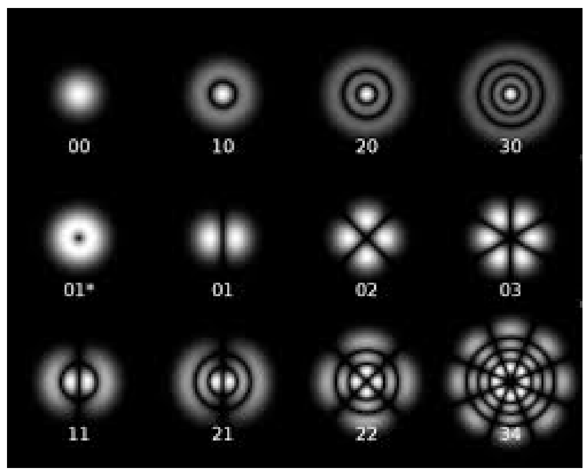


$$\vec{A}(\vec{r}) = \frac{-ie^{ikz}}{\lambda r}\hat{x}A_o \int_0^a \rho' d\rho' \int_0^{2\pi} d\varphi' e^{im\varphi'} . \tag{16.58a}$$

$$= 0 \tag{16.58b}$$

unless $m = 0$. Thus any beam for which $m \neq 0$ is automatically a vortex beam, consistent with the phase being indeterminate at $\rho = 0$.

H. Photon spin, orbital angular momentum, and topological charge.



Spin is defined as the rotation of the electric field vector of a propagating wave that results from polarization. Orbital angular momentum is defined as that which arises from the azimuthal variation of the potential that arises from the term $e^{-im\varphi}$ in Eq. (16.57). Hence orbital angular momentum is associated with specific Laguerre polynomials, and the underlying mechanisms giving rise to spin and orbital angular momentum are different.

Spin and orbital angular momentum have no classical analog, so they are calculated using the quantum-mechanical definition

$$L_z = \frac{\hbar}{i}\frac{\partial}{\partial\varphi} = -i\hbar\frac{\partial}{\partial\varphi}, \tag{16.59}$$

where L_z is the angular momentum operator. Considering orbital angular momentum first, it follows from Eq. (16.57) that the orbital angular momentum of the expression in that equation is $-m\hbar$. This is the origin of the azimuthal variations in the above figure. In terminology in current use, this is known as topological charge. It owes its robustness to the fact that the azimuthal eigenfunction of the wave equation must assume integral values of *m*.

Before considering spin, we examine how orbital angular momentum affects the time evolution of the vector potential. Noting that the observable of any wave is its real projection, consider

$$\vec{A}(\vec{r}) = \hat{x}A_o g(\rho)e^{im\varphi - i\omega t}. \tag{16.60}$$

where $g(\rho)$ represents the Gaussian/Laguerre dependences and the various prefactors. Then

$$L_z\vec{A}(\vec{r},t) = \frac{\hbar}{i}(im)\vec{A}(\vec{r},t)$$
$$= m\hbar\hat{x}A_o g(\rho)\cos(m\varphi - \omega t), \tag{16.61}$$

ignoring possible additive constants to the phase. At a given time t_o, $A(\vec{r},t)$ forms a spatial pattern of the type seen in the above figure, with adjacent regions having opposite signs. As time evolves, the pattern rotates with a speed that can be obtained by considering a fixed-phase point, for example

$$m\varphi = \omega t. \tag{16.62}$$

Thus the rotation speed of the pattern is inversely proportional to *m*.

We now examine spin. Consider a left-hand-circularly polarized plane wave propagating in the *z* direction. Then at $z = 0$

$$\vec{A}(0,t) = A_o(\hat{x} + i\hat{y})e^{-i\omega t} \tag{16.63}$$

Then

$$\mathrm{Re}\left(-i\hbar\frac{\partial}{\partial(\omega t)}\vec{A}(0)e^{-i\omega t}\frac{d(\omega t)}{d\varphi}\right) = \mathrm{Re}\left(-i\hbar(-i)\vec{A}(0)e^{-i\omega t}(-1)\right) \tag{16.64a}$$

$$= \hbar\,\mathrm{Re}\left(A_o(\hat{x} + i\hat{y})e^{-i\omega t}\right) \tag{16.64b}$$

$$= \hbar A_0(\hat{x}\cos\omega t + \hat{y}\sin\omega t). \tag{16.64c}$$

This is not a good derivation, because the sign is determined after the derivative operation rather than part of the derivation itself. This will be fixed in the future. In the present example the sign is positive because the constant-phase calculation of the previous paragraph shows that the left-hand-circularly polarized beam rotates in the direction of increasing angle, and thus has positive helicity. The sign would be negative for right-hand-circularly polarized radiation, and zero for linear polarization.

The more interesting calculation involves the orbital momentum seen by an observer near but not on the *z* axis. For light linearly polarized along *x*, Eq. (16.15b) becomes

$$\vec{A}(\vec{r}) = \frac{-ie^{ikz}}{\lambda r}\hat{x}A_o\int_0^a \rho' d\rho' \int_0^{2\pi} d\varphi' e^{im\varphi'}e^{-i\vec{k}\cdot\vec{r}'} \tag{16.65a}$$

$$= \frac{-ie^{ikz}}{\lambda r}\hat{x}A_o\int_0^a \rho' d\rho' \int_0^{2\pi} d\varphi' e^{im\varphi'} e^{-ik_\rho \rho' \cos(\varphi' - \varphi_o)} , \tag{16.65b}$$

$$= \frac{-ie^{ikr}}{\lambda r}\hat{x}A_o\int_0^a \rho' d\rho' \int_0^{2\pi} d\varphi' e^{im\varphi_o} e^{im(\varphi' - \varphi_o)} e^{-ik_\rho \rho' \cos(\varphi' - \varphi_o)} , \tag{16.65c}$$

where in Eqs. (16.65) we have located the observer at

$$\vec{k} = \vec{k}_o \cong k(\hat{x}\sin\theta_o \cos\varphi_o + \hat{y}\sin\theta_o \sin\varphi_o + \hat{z}) \tag{16.66a}$$

$$= k_\rho(\hat{x}\cos\varphi_o + \hat{y}\sin\varphi_o) + k_z\,\hat{z}\,; \tag{16.66b}$$

and the integration variable is

$$\vec{r}\,' = \rho'(\hat{x}\cos\varphi' + \hat{y}\sin\varphi')\,. \tag{16.66c}$$

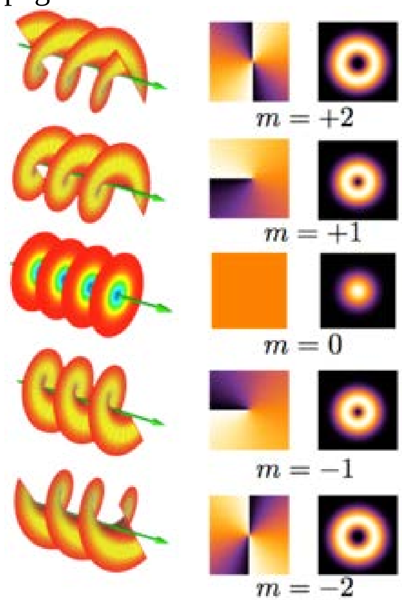
With these definitions Eq. (16.65b) follows directly from Eq. (16.65a). The angular integration in Eq. (16.65b) is nontrivial, but we bypass it by adding and subtracting the phase $im\varphi_o$ in the exponent, part of which is then moved to the prefactor. Next, by multiplying and dividing appropriately by *a*, the integral can be turned into a pure number:

$$\vec{A}(\vec{r}) = -\frac{i}{\lambda}\frac{e^{ikz}}{r}(a^2 e^{im\varphi_k})\hat{x}A_o\left(\int_0^1 z'' dz'' \int_{-\pi}^{\pi} d\varphi'' e^{im\varphi''} e^{-i(k_\rho a)z''\cos\varphi''}\right), \tag{16.67}$$

with all relevant information in the prefactor.

Equation (16.67) shows that the orbital angular momentum of the resulting beam is $L_z = \hbar m$ independent of the distance of the observer from the *z* axis and the polarization state of light. Thus spin and orbital angular momenta remain independent. Examples of different combinations are shown in the figure on the next page.

Technological interest in higher-order Gaussian beams follows from the orthogonality of the cross sections of the different modes. The ability to generate and detect modes independently in principle increases the information-carrying capacity of optical fibers. However, significant technical challenges need to be overcome before this technology is put in widespread use. In particular, the detection of the different independent modes remains a challenge. At this stage higher-order Gaussian beams are mainly used as examples of Gaussian-beam physics.



I. Propagation equations for $\vec{E}$ and $\vec{H}$.

The "fields" analog of the wave equation for potentials gets considerable attention in diffraction in standard textbooks, so we finish by making some comments about these equations. In particular, we show that if the potentials satisfy the wave equation, then the fields satisfy the equivalent equation for fields. We assume that $\varepsilon = \mu = 1$, since for most applications propagation occurs in air and the expressions are easily modified if propagation occurs in a different medium. Because the Poynting vector is written in terms of $\vec{H}$ and we have set $\mu = 1$, in the following we use $\vec{H}$ exclusively instead of $\vec{B}$. Taking the curl of Faraday's Equation and substituting Ampere's Equation for $\nabla \times \vec{H}$ leads to

$$\nabla \times (\nabla \times \vec{E}) + \frac{1}{c}\frac{\partial}{\partial t}(\nabla \times \vec{H}) =$$

$$= \left(\nabla(\nabla \cdot) - \nabla^2 + \frac{1}{c^2}\frac{\partial^2}{\partial t^2} \right) \vec{E} = -\frac{4\pi}{c^2}\frac{\partial \vec{J}}{\partial t}, \tag{16.68}$$

or with the more usual sign convention

$$\left(\nabla^2 - \nabla(\nabla \cdot) - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} \right) \vec{E} = \frac{4\pi}{c^2}\frac{\partial \vec{J}}{\partial t}. \tag{16.69}$$

The corresponding equation for $\vec{H}$ is

$$\left(\nabla^2 - \nabla(\nabla \cdot) - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} \right) \vec{H} = \left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} \right) \vec{H} = \frac{4\pi}{c} \nabla \times \vec{J}\ , \tag{16.70}$$

where we have used $\nabla \cdot \vec{H} = 0$. In the far-field limit, where we can replace ∇ with $(i\,k)$, the equation for $\vec{H}$ is at least consistent with expectations and with $\nabla \cdot \vec{H} = 0$, because $(\nabla \times \vec{J})$ is itself perpendicular to $\vec{k}$ so no longitudinal component is generated.

The same cannot be said for the equation for $\vec{E}$, which remains inconsistent. It is no accident that Jackson for example discusses far-field radiation involving fields by first evaluating $\vec{H}$, then using Ampère’s Law to generate $\vec{E}$. Alternatively, one can work with vector diffraction theory directly. This is the standard approach for eliminating the longitudinal component of $\vec{E}$.

Another possibility is to substitute $\nabla \cdot E = 4\pi\rho$. The result is

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} \right) \vec{E} = -\frac{4\pi}{c^2}\frac{\partial \vec{J}}{\partial t} - 4\pi \nabla \rho. \tag{16.71}$$

Although this now has the wave-equation form, the fact that the source terms contain operators makes the solution in the inhomogeneous case difficult.

We can fix this problem by rewriting these equations in terms of the scalar and vector potentials ϕ and $\vec{A}$. Substituting the definitions of $\vec{E}$ and $\vec{H}$ in terms of ϕ and $\vec{A}$ yields

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} \right) \left(-\frac{1}{c}\frac{\partial \vec{A}}{\partial t} - \nabla \phi \right) = -\frac{4\pi}{c^2}\frac{\partial \vec{J}}{\partial t} - 4\pi \nabla \rho \tag{16.72a}$$

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\nabla \times \vec{A} = -\frac{4\pi}{c}\nabla \times \vec{J}\,. \qquad (16.72b)$$

But these equations are nothing more than wave equations working on the results of differentiation operations performed on the potentials, assuming that $\vec{A}$ originates entirely with $\vec{J}$, and ϕ with ρ. Thus if we solve the inhomogeneous wave equations for the potentials, we are guaranteed that the inhomogeneous wave equations will also be satisfied for the fields.