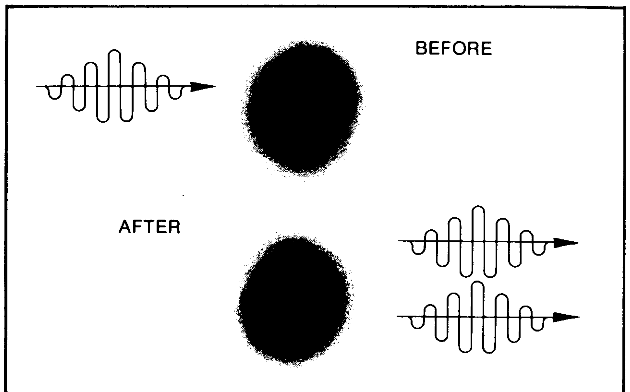
HOW DOES A HELIUM NEON LASER WORK?

All lasers depend for their operation on the phenomenon of stimulated emission of radiation. For this to occur, it is necessary for some atoms to exist in an excited energy state. In an excited atom an electron may decay to a lower energy level, emitting a photon of energy $E = h\nu$, where h is Planck's constant and ν is the photon frequency. This is spontaneous emission of radiation, or luminescence. If the emitted photon passes close to another atom in a similar excited state, it may induce an identical transition. In this case the emitted and stimulating photons will be identical in every respect, having the same frequency, phase, direction, and polarization – they will be spatially and temporally coherent. This is stimulated emission of radiation.

To achieve laser oscillation it is necessary to make the probability of stimulated emission exceed that of spontaneous emission and absorption. This is done by creating a population inversion, in which more atoms exist in an excited energy state than in some lower state to which direct radiative transition is possible. This situation must be maintained or laser action will cease.

In the helium neon system, a mixture of helium and neon gases is contained in a narrow bore tube. These gases are replenished as needed from a reservoir. The mixture is predominantly helium, with about 10% neon. Various isotopic mixtures have been used to improve the power output and stability of lasers. Melles Griot helium neon lasers are filled with a proprietary mixture which has the effect of maximizing the ratio of beam power to plasma volume. An electrical discharge ionizes the gas inside the narrow bore. The plasma so formed contains many helium atoms in metastable states. In these states one of the valence electrons has been raised from the 1s and 2s level, where it will remain for a long time, the downward transition being forbidden by the selection rules of quantum mechanics. Neon has an atomic number of 10, and has 10 electrons normally residing in the $1s^2 2s^2 2p^6$ configuration. Collisions between metastable helium atoms and neon atoms cause one of the neon electrons in the 2p level to be raised to the 4s or 5s levels. The energy match between the 2s helium levels and the 4s and 5s neon levels is very close. It is therefore much more probable that this excitation will occur than one to the 3s or 3p levels. On exciting a neon atom in this way, the helium atom drops back to the ground state. However, because there are many more helium atoms than neon, the metastable collision probability remains high. For neon laser action to occur, it is necessary to maintain a greater population of neon atoms with electrons in the 4s and 5s levels than in the 3p level. This condition is satisfied by the selective excitation process just described, and by the low natural population of the 3p levels. A number of alternative downward transitions are available to the excited neon electrons. The major neon laser line transitions are indicated on the energy level diagram.

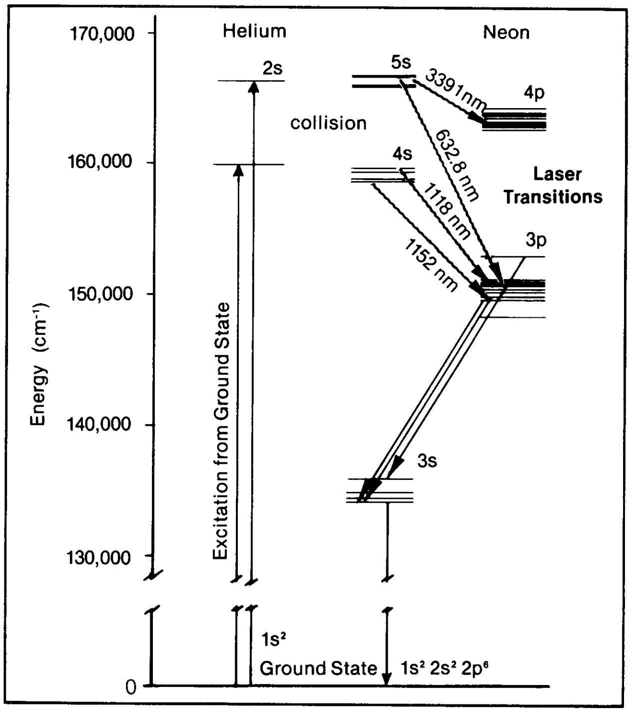


STIMULATED EMISSION, an artist's conception.

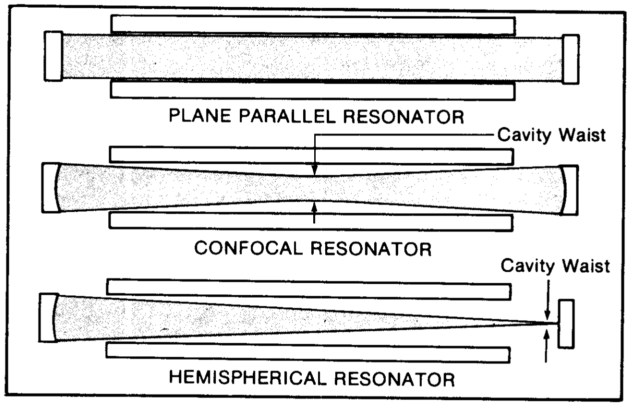
Some of the radiation emitted will be in a direction approximately parallel to the tube bore. The rest will be emitted through the tube walls and does not contribute to the laser operation. For a photon traveling along the bore, there is a chance of passage close to a second excited neon atom and stimulation of the emission of another identical photon. The probability of this occurrence depends on the gas density and length of the bore. For a single pass along the bore the probability is on the order of 10%.

Mirrors are placed at either end of the tube, reflecting photons many times along the length of the bore. In this way the likelihood of stimulated emission is further increased. Each stimulated photon is capable of stimulating another, and each is in perfect coherence with those already present. In this way, light amplification is achieved.

The two mirrors form a resonant optical cavity, and it is this cavity which is responsible for the selective gain characteristics of the laser. Cavity mirrors are coated to have a very high reflectance at the laser wavelength of interest, but to transmit unwanted laser lines. By retaining only the desired wavelength photons in the bore, amplification is restricted to only this wavelength. Similarly, misdirected radiation of the desired wavelength escapes from the cavity and is not amplified, giving rise to the highly directional nature of the laser output. In order to make use of the desired laser radiation, some of it must be allowed to escape from the cavity as a beam. In practice, one of the mirrors is coated to permit about 1% transmission at the laser wavelength. It is this 1% of light which escapes that forms the usable output of the laser. It follows that the internal beam is 100 times more powerful than the output beam.



ENERGY LEVEL DIAGRAMS for helium and neon. Levels are identified in spectroscopic notation, and major laser transitions are shown.

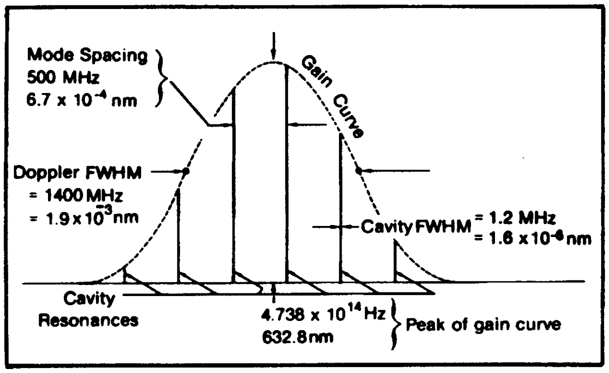


LASER CAVITY TYPES, showing the shape and size of the active plasma volume. The most popular cavity for HeNe lasers is the quasi-hemispherical cavity, because of its stability and short length. To obtain a quasi-hemispherical cavity, the hemispherical cavity is shortened very slightly.

Various shapes of resonant cavity have been used. The plane parallel resonator, perhaps the most obvious choice, is never used. It has the theoretical advantage of fully utilizing the available plasma volume to obtain high efficiency, but is inherently unstable and has large diffraction losses.

The confocal resonator has two equal radius concave spherical mirrors, each placed at the center of curvature of the other. The cavity utilizes a large fraction of the plasma volume and produces high power. It is less sensitive to alignment, but is slightly sensitive to mirror separation. The confocal resonator is rarely used when TEM_{OO} output is required.

The hemispherical resonator has a plane mirror at the center of curvature of a concave spherical mirror. The resonator becomes quasi-hemispherical if this mirror separation is slightly reduced. This quasi-hemispherical configuration is easy to adjust and produces highly coherent output. The disadvantage of this cavity is that it only utilizes about one-third the volume of available plasma due to the focusing effect of the concave mirror. This inefficiency limits the power output. However, its ease of adjustment, and stability once aligned more than outweigh the disadvantage of power limitation. Quasi-hemispherical cavities are, therefore, the most commonly used in commercial lasers.



DOPPLER BROADENED GAIN CURVE supporting 6 longitudinal modes or cavity resonances. The gain curve is of the same shape as the line profile in spontaneous emission.

The presence of an optical cavity leads to longitudinal modes of oscillation. Only the light wavelengths which correspond to standing waves are amplified. Out of phase reflections are quickly lost through destructive interference. In order for a mode or standing wave to exist, the round trip cavity length must be exactly equal to an integer number of wavelengths:

$2L = m\lambda$ (where m is a large integer).

This corresponds to the series of resonant frequencies

$\nu = \frac{mc}{2L}$ (where c = velocity of light).

These frequencies each define an extremely narrow band within which laser oscillation can occur. The frequency separation of these longitudinal modes or resonances is

$$\Delta\nu = \frac{c}{2L},$$

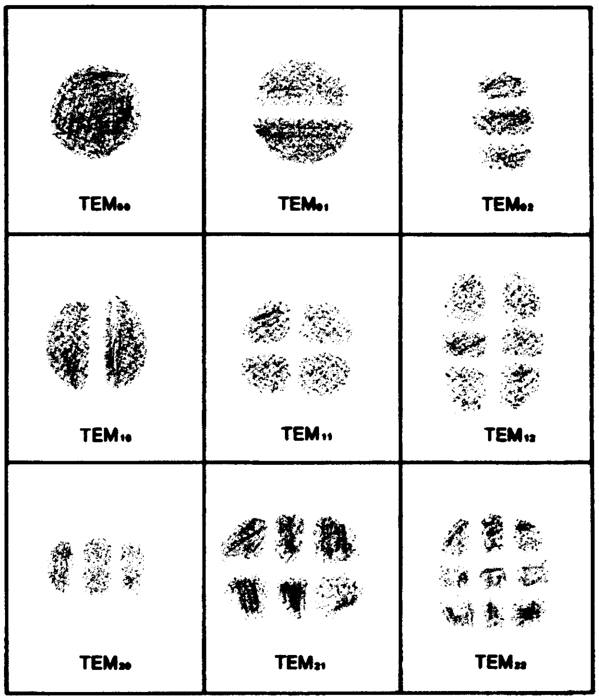
and is inversely proportional to the length of the cavity. Thus, for a 50cm cavity length, the mode spacing will be:

$$\Delta\nu = \frac{3 \times 10^{10}}{2 \times 50} \text{ Hz}$$

$$= 300 \text{ MHz}$$

In order for these modes to exhibit optical gain, their individual frequencies must lie within the finite spectral line width of the natural atomic transition. The net effect of random thermal atomic motion within the plasma is to produce a Doppler broadened spectral line. Modes which fall within the Doppler broadened spectral envelope will be amplified. The linewidth of a single laser mode is extremely narrow compared with the Doppler linewidths ordinarily encountered in atomic-spectroscopy.

In addition to longitudinal modes, the cavity can support a number of Transverse Electric and Magnetic modes (TEM



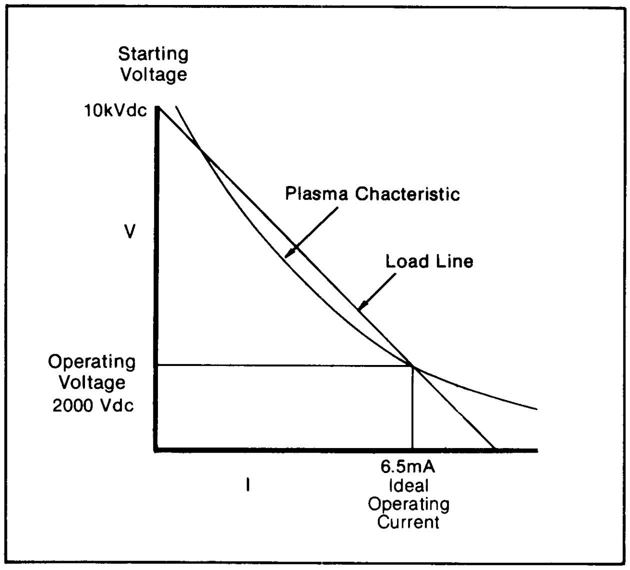
TRANSVERSE ELECTRIC AND MAGNETIC MODES of a laser cavity, with identification, and showing nulls of the irradiance distribution. The solid regions represent the interiors of irradiance contours arbitrarily selected to best reveal distribution shape.

modes). These are modes in which the irradiance distribution along the electric and magnetic field vectors exhibits one or more nulls. At small angles from the cavity axis it is possible to obtain a different series of standing waves, each of which can produce a stable mode of oscillation and gain. Transverse modes are identified by their irradiance distributions and are designated TEM_{pq}, where the subscripts p and q refer to the number of nulls along two orthogonal axes which are perpendicular to the cavity axis. The lowest order TEM_{OO} mode has a Gaussian irradiance distribution. This mode experiences the minimum possible diffraction loss, has minimum divergence, and can be focused to the smallest possible spot. For these reasons, it is often desirable that the laser should be restricted to operation in this fundamental mode. Higher order modes have a larger spread and usually suffer high diffraction losses. A mode often encountered in practice is the "doughnut" mode (TEM^*_{01}), which is a combination of TEM_{01} and TEM_{10}, having a central minimum and a bright ring.

Lasers are sometimes designed to exclude all but the fundamental mode, and sometimes to permit higher order modes. The exact design requirements depend on the type of cavity in use. If higher orders are to be excluded, the bore size must be held very close to the fundamental mode diameter. This may result in diffraction of some wanted radiation out of the beam. With a confocal resonator it is very difficult to exclude high order modes. Most TEM_{oo} mode lasers use quasihemispherical resonators. In this case, beam collimation quality is limited only by diffraction. Most helium neon lasers have an output beam divergence of 0.8 to 1.5 milliradians (beam angular diameter).

The polarization of laser radiation is an important effect which occurs as a direct result of the stimulated emission process. Just as the stimulated photon is in phase, frequency, and directional agreement with the stimulating photon, so it also carries the same state of polarization. The resonant cavity sets up standing waves of constant linear polarization. This results in alternate longitudinal modes having orthogonal states of polarization. Likewise, adjacent transverse modes have orthogonal states of polarization. The resultant polarization depends upon the precise mix of modes present, and this mix may change rapidly with time. So called "randomly polarized" lasers have an indeterminate combination of orthogonally polarized radiation. To eliminate this uncertainty, laser outputs are often deliberately linearly polarized. A single Brewster angle window is placed in the cavity, effectively eliminating one state of polarization, and preventing it from being amplified. The output is more stable and has a known state of polarization, namely the p-polarization defined by the window.

Electrically, helium neon lasers function as negative resistors. They require a high voltage in order to strike the discharge and then as current flows the voltage required to sustain operation is reduced. A ballast resistor is connected in series with the laser to provide safe minimum impedance. The output of the laser is a function of the d.c. discharge current. In order to optimize the laser performance, a particular operating current is chosen, and the power supply is designed to maintain this current with very low ripple.



CHARACTERISTIC OPERATING CURVE OF A HeNe LASER. The curve is typical of other low pressure gaseous discharges, such as advertising signs.

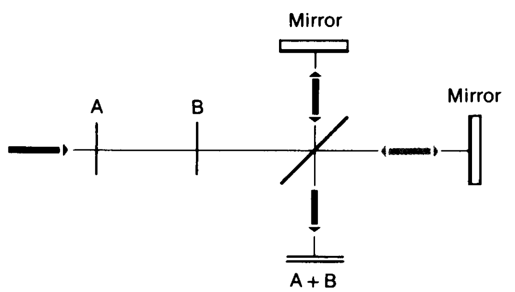
CHARACTERISTICS OF LASER LIGHT

The helium neon laser delivers relatively intense coherent electromagnetic radiation at 632.8nm. It is able to do this reliably, for many hundreds of hours, without variation of output power or beam direction, at modest cost. The temporal coherence of a laser is a measure of the ability of the beam to produce interference effects as a result of differences in path lengths, and is important in interferometry and holography. The spatial coherence is of importance in applications where it is necessary to focus all of the laser's output into an extremely small spot.

Both temporal and spatial coherence have long been sought for industrial and laboratory optical applications. Before the laser was invented, monochromatic sources with various degrees of coherence were in use. The relatively low output levels available, and the Doppler broadening of most atomic sources, were such that only limited applications were ever achieved.

TEMPORAL COHERENCE

If the phase at a particular instant in time along a traveling wavefront is the same as its phase after the wave has traveled a distance x in x/c seconds, for all values of x, the wave is defined to be perfectly temporally coherent. Michelson interference fringes are a direct result of temporal coherence, and they are visible only over the distance in which the temporal coherence is maintained. For conventional sources this distance is rarely more than a few millimeters, whereas for a multimode HeNe laser it is typically 20 to 30 centimeters. That this distance is not infinite demonstrates that lasers are not perfectly monochromatic or perfectly coherent. Off-the-shelf lasers are good enough, however, to permit precise interferometric measurement of distances hundreds of times longer than was formerly possible.



MICHELSON INTERFEROMETER, with path difference adjusted so as to superimpose wavefronts A and B in the output beam, where they will interfere if their separation in the input beam did not exceed the radiation coherence length. The interferometer measures coherence length, and coherence length is the measure of temporal coherence.

Visibility of interference fringes is defined as

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} ,$$

where I_{max} and I_{min} are the maximum and minimum observed fringe intensities. Unity visibility is associated with full temporal coherence. Zero visibility (absence of fringes) corresponds to $I_{max} = I_{min}$, and is associated with complete incoherence.

The minimum path length difference for which V is zero (i.e., fringes vanish) is called the coherence length (S_c). The time delay corresponding to this length is the coherence time (Δt). These are related by $S_c = c\Delta t$. The uncertainty relation between time and frequency leads to $\Delta\nu\, \Delta t = 1$, where $\Delta\nu$ is the line width (frequency spread) of the non-monochromatic laser output line. Lasers operating in a single longitudinal mode have linewidths of approximately 1 MHz. A typical FWHM for a commercial multimode HeNe laser is $\Delta\nu \cong 1500$ MHz, because the Doppler broadened bandwidth provides an envelope for the gain of a number of longitudinal modes.

SPATIAL COHERENCE

Correlation of phase in light beams transverse to the direction of travel is best illustrated by Young-type interference fringes. The light from two transverse portions of the same beam is allowed to interfere, producing fringes. Visibility is defined as above. Visibility of unity will again correspond to complete coherence, whereas visibility of less than unity implies partial coherence.

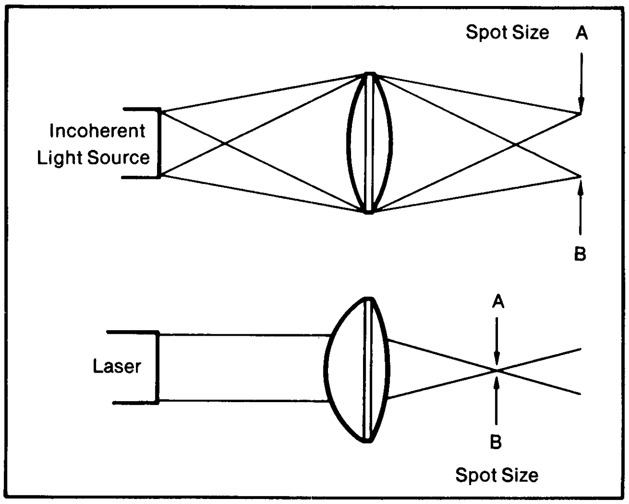
Spatial and temporal coherence are independent of one another; a source may be spatially coherent while temporally incoherent and vice versa. The mode discrimination properties of the laser cavity produce spatial coherence, and the Doppler broadened bandwidth limits temporal coherence.

DIRECTIONALITY AND ENERGY DENSITY

The extreme directionality of laser radiation follows directly from the coherence of the stimulated emission process. Because all of the stimulated photons are alike in every way, they carry no information about the location of the excited neon atoms they come from. Thus all of the excited neon atoms are functionally equivalent, and the laser output radiation behaves as if it all had originated in a tiny imaginary volume, having dimensions of the order of a wavelength, situated precisely on-axis at the center of the cavity waist. This remains true regardless of the size or shape of the active plasma volume.

When an off-axis neon atom is stimulated to emit a photon, the appearance presented to the outside world is that this photon originates on-axis. This behavior is attributable to transverse coherence. When a neon atom far from the waist plane is stimulated to emit, the photon still appears to come from the waist plane. This behavior is attributable to temporal coherence far in excess of the monochromaticity requirements of the beam optics formulas. Because all the coherent energy output of the active medium is effectively funneled by stimulated emission through that tiny imaginary volume, energy densities in focused laser beams can far exceed the highest energy densities actually occurring in the source plasma, a situation impossible with spontaneous emission.

When we look at an extended incoherent source, such as the surface of an incandescent solid or the spontaneous emission of a gaseous discharge, we can plainly see that it is extended.

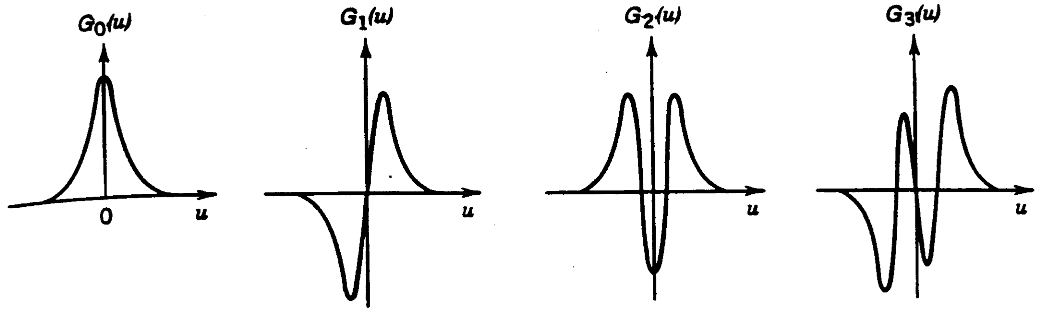


FOCUSABILITY COMPARISON for light from laser and conventional sources. Extended sources have extended images. The laser is the functional equivalent of a true point source situated at the center of the cavity waist.

This is because the photons carry directional information telling us which parts of the source they come from. Laser radiation carries no such information.

SPECKLE

A smooth but diffuse reflector has a surprising appearance under HeNe laser illumination. Though the beam irradiance distribution may be very smooth, the apparent distribution at the surface is granular. The granule size is approximately the size of the resolution element of the eye. Neighboring granules vary enormously in brightness. A bright granule looks bright because of cooperative in-phase reflections from two or more area elements within the granule, directed at the eye. Such interference is possible because coherence length exceeds the scale of surface roughness. Speckle is a great nuisance in laser imaging and holography, but carries information valuable in non-contacting metrology. Radar "glint" is the same effect on a larger scale, and hampers radar imagery.



(a) (b) (c) (d)

Figure 3.3-1 Several low-order Hermite-Gaussian functions: (*a*) $G_0(u)$; (*b*) $G_1(u)$; (*c*) $G_2(u)$; (*d*) $G_3(u)$.

where

$$G_l(u) = H_l(u)\exp\left(\frac{-u^2}{2}\right), \qquad l = 0, 1, 2, \ldots, \tag{3.3-10}$$

is known as the **Hermite–Gaussian function** of order l, and $A_{l,m}$ is a constant.

Since $H_0(u) = 1$, the Hermite–Gaussian function of order 0 is simply the Gaussian function. $G_1(u) = 2u\exp(-u^2/2)$ is an odd function, $G_2(u) = (4u^2 - 2)\exp(-u^2/2)$ is even, $G_3(u) = (8u^3 - 12u)\exp(-u^2/2)$ is odd, and so on. These functions are shown in Fig. 3.3-1.

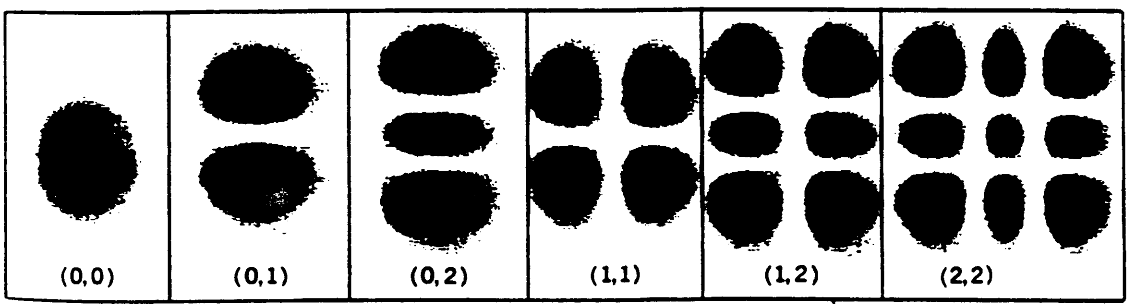
An optical wave with complex amplitude given by (3.3-9) is known as the Hermite–Gaussian beam of order (l, m). The Hermite–Gaussian beam of order (0, 0) is the Gaussian beam.

Intensity Distribution

The optical intensity of the (l, m) Hermite–Gaussian beam is

$$I_{l,m}(x, y, z) = |A_{l,m}|^2\left[\frac{W_0}{W(z)}\right]^2 G_l^2\left[\frac{\sqrt{2}x}{W(z)}\right] G_m^2\left[\frac{\sqrt{2}y}{W(z)}\right]. \tag{3.3-11}$$

Figure 3.3-2 illustrates the dependence of the intensity on the normalized transverse distances $u = \sqrt{2}\,x/W(z)$ and $v = \sqrt{2}\,y/W(z)$ for several values of l and m. Beams of higher order have larger widths than those of lower order as is evident from Fig. 3.3-1.



(a) (b) (c) (d) (e) (f)

Figure 3.3-2 Intensity distributions of several low-order Hermite–Gaussian beams in the transverse plane. The order (l, m) is indicated in each case.