From the above discussion, the eigenvalues $\tilde{\sigma}_{lm}$ are seen to be such that $|\tilde{\sigma}_{lm}|^2$ gives the factor by which the beam intensity is changed as a result of one round trip. Since this change is due to diffraction losses, we must then have $|\tilde{\sigma}_{lm}|^2 < 1_j$ thus the quantity

$$\gamma_{lm} = 1 - |\,\tilde{\sigma}_{lm}\,|^2 \tag{5.2.6}$$

gives the round-trip fractional power loss due to diffraction. One can also see that, according to Eq. (5.2.4), $\Delta\phi_{lm} = -2\,kL + \phi_{lm}$ is the corresponding round-trip phase shift. For the field to be self-reproducing, we must then require that $\Delta\phi_{lm} = -2\pi\,n$, where $n$ is an integer. We thus get $-2kL + \phi_{lm} = -2\pi n$, and, with the substitution $k = 2\pi\nu/c$, we obtain the cavity resonance frequencies as

$$\nu_{lmn} = \frac{c}{2L}\left[n + \frac{\phi_{lm}}{2\pi}\right] \tag{5.2.7}$$

Note that we have indicated explicitly that these frequencies are dependent on the values of the three numbers $l$, $m$, and $n$. The integers $l$ and $m$ represent the order of the eigensolution in Eq. (5.2.5) while the integer number $n$ specifies the total phase shift of the beam, after one round trip, in units of $2\pi$ (i.e. $n = -\Delta\phi_{lm}/2\pi$).

As a conclusion of this section we can say that the eigenmodes and the eigenvalues of our problem can be obtained upon solving the integral equation Eq. (5.2.5). In fact, its eigensolutions, $\tilde{E}_{lm}$, give the field of the eigenmodes at all point in a given plane. For each mode $\tilde{E}_{lm}$, the corresponding eigenvalue $\tilde{\sigma}_{lm}$ then gives: (a) The round-trip diffraction loss, $\gamma_{lm}$, through its magnitude $|\tilde{\sigma}_{lm}|$ [see Eq. (5.2.6)]. (b) The resonance frequency, $\nu_{lmn}$, through its phase, $\phi_{lm}$ [see Eq. (5.2.7)].

## 5.3. PHOTON LIFETIME AND CAVITY Q

Consider a given mode of a stable or unstable cavity and assume, for generality, that some losses other than diffraction losses are also present. For instance one may have mirror losses as a result of mirror reflectivity being smaller than unity. One may also have scattering losses in some optical element within the cavity. Under these conditions we want to calculate the rate of energy decay in the given cavity mode. To this purpose, let $I_0$ be the initial intensity corresponding to the field amplitude $\tilde{E}(x_1,\ y_1,\ 0)$ at a given transverse coordinate $x_1, y_1$. Let $R_1$ and $R_2$ be the (power) reflectivities of the two mirrors and $T_i$ the fractional internal loss per pass due to diffraction and any other internal losses. The intensity $I(t_I)$ at the same point $x_1, y_1$ at a time $t_1 = 2L/c$, i.e. after one cavity round trip, will be

$$I(t_1) = R_1\,R_2(1 - T_i)^2\,I_0 \tag{5.3.1}$$

Note that, since $T_i$ is defined here as the fractional internal loss per pass, the intensity is reduced by a factor $(1 - T_i)$ in a single pass and hence by a factor $(1 - T_i)^2$ in a double pass (round trip). The intensity, at the same transverse co-ordinate, after $m$ round trips, i.e. at time

$$t_m = 2mL/c \tag{5.3.2}$$

is then

$$I(t_m) = [R_1 R_2 (1 - T_i)^2]^m I_0 \tag{5.3.3}$$

Let now $\phi(t)$ be the total number of photons in the given cavity mode at time $t$. Since the mode retains its shape after each round trip, we can set $\phi(t) \propto I(t)$. From Eq. (5.3.3) we can then write

$$\phi(t_m) = \left[R_1\, R_2\, (1 - T_i)^2\right]^m\, \phi_0 \tag{5.3.4}$$

where $\phi_0$ is the number of photons initially present in the cavity. We can also set

$$\phi(t_m) = [\exp(-t/\tau_c)]\phi_0 \tag{5.3.5}$$

where $\tau_c$ is a suitable constant. In fact, a comparison of Eqs. (5.3.5) with (5.3.4) with the help of Eq. (5.3.2) shows that

$$\exp\left(-\, 2mL/c\, \tau_c\right) = \left[R_1\, R_2\, (1 - T_i)^2\right]^m \tag{5.3.6}$$

from which one finds that $\tau_c$ is independent of the number of round trips, $m$, and is given by

$$\tau_c = -2\, L/c \ln\left[R_1\, R_2 (1 - T_i)^2\right] \tag{5.3.7}$$

If we now assume that Eq. (5.3.5) holds, not only at times $t_m$, but also at any time $t$ (>0), we can then write

$$\phi(t) \cong \exp\left(-t/\tau_c\right) \phi_0 \tag{5.3.8}$$

In this way, we justify the assumption Eq. (5.1.1) for the mode field and identify Eq. (5.3.7) as the expression for the cavity photon lifetime. One can notice that Eq. (5.3.7), with the help of Eqs. (1.2.4) and (1.2.6), can readily be transformed to

$$\tau_c = L/c\gamma \tag{5.3.9}$$

We thus see that the cavity photon lifetime is just equal to the transit time $\tau_T = L/c$ of the beam in the laser cavity divided by the (logarithmic) cavity loss $\gamma$.

**Example 5.2.** *Calculation of the cavity photon lifetime.* We will assume $R_1 = R_2 = R = 0.98$ and $T_i \cong 0$. From Eq. (5.3.7) we obtain $\tau_c = \tau_T/[-\ln R] = 49.5\, \tau_T$, where $\tau_T$ is the transit time of the photons for a single-pass in the cavity. From this example we note that the photon lifetime is much longer than the transit time, a result which is typical of low loss cavities. If we now assume $L = 90\,\text{cm}$, we get $\tau_T = 3\,\text{ns}$ and $\tau_c \cong 150\,\text{ns}$.

Having calculated the photon lifetime, the time behavior of the electric field, at any point inside the resonator can, according to Eq. (5.1.1) and within the scalar approximation, be written as $E(t) = \tilde{E} \exp\left[(-t/2\tau_c) + j\omega t\right]$, where $\omega$ is the angular resonance frequency of the mode. The same time behavior then applies for the field of the output wave leaving the cavity through one mirror as a result of finite mirror transmission. If we now take the Fourier transform of this field, we find that the power spectrum of the emitted light has a Lorentzian line shape with linewidth (FWHM) given by

$$\Delta\nu_c = 1/2\pi\, \tau_c \tag{5.3.10}$$

It should be noted that the spectrum of the emitted light, obtained in this way, does not exactly agree with the transmission spectrum shown for a Fabry–Perot interferometer in Sect. 4.5, whose shape is not Lorentzian [see Eq. (4.5.6)]. In particular, the expression for $\Delta\nu_c$ obtained here [see Eq. (5.3.10)], when combined with Eq. (5.3.7) with $T_i \cong 0$, does not coincide with that obtained in Sect. 4.5 [see Eq. (4.5.12) with $L' = L$]. This discrepancy can be traced back to the approximation made in writing Eq. (5.3.8). In numerical terms, however, the discrepancy between the two results is quite small, especially at high values of reflectivity, as can be seen from the following example. From now on we will therefore assume that the cavity line shape is Lorentzian with width given by Eq. (5.3.10) and that the cavity photon lifetime is given by Eq. (5.3.7).

**Example 5.3.** *Linewidth of a cavity resonance.* If we take again $R_1 = R_2 = 0.98$ and $T_i = 0$, from Eqs. (5.3.10) and (5.3.7) we get $\Delta\nu_c \cong 6.4307 \times 10^{-3} \times (c/2L)$, while from Eq. (4.5.12) we get $\Delta\nu_c \cong 6.4308 \times 10^{-3} \times (c/2L)$. For the particular case $L = 90$ cm, we then obtain $\Delta\nu_c \cong 1.1$ MHz. Even at the relatively low reflectivity values of $R_1 = R_2 = 0.5$, the discrepancy is not large. In fact from Eqs. (5.3.10) and (5.3.7) we get $\Delta\nu_c \cong 0.221 \times (c/2L)$, while from Eq. (4.5.12) $\Delta\nu_c \cong 0.225 \times (c/2L)$. Again for $L = 90$ cm we then obtain $\Delta\nu_c \cong 37.5$ MHz. Thus, in typical cases, $\Delta\nu_c$ may range from a few to a few tens of MHz.

Having discussed the cavity photon lifetime, we can now introduce the cavity quality factor, or $Q$ factor, and derive its relation to the photon lifetime. For any resonant system, and in particular for a resonant optical cavity, one defines the cavity $Q$ factor (usually abbreviated to cavity $Q$) as $Q = 2\pi$(energy stored)/(energy lost in one cycle of oscillation). Thus a high value of cavity $Q$ implies low losses of the resonant system. Since, in our case, the energy stored is $\phi h\nu$ and the energy lost in one cycle is $h\nu(-d\phi/dt)(1/\nu) = -hd\phi/dt$, we have

$$Q = -2\pi\nu\phi/(d\phi/dt) \tag{5.3.11}$$

From Eq. (5.3.8) we then get

$$Q = 2\pi\nu\tau_c \tag{5.3.12}$$

which, with the help of Eq. (5.3.10), can be transformed to the more suggestive form

$$Q = \nu/\Delta\nu_c \tag{5.3.13}$$

Thus the cavity $Q$ factor can be interpreted as the ratio between the resonance frequency, $\nu$, of the given mode and its linewidth, $\Delta\nu_c$.

**Example 5.4.** *Q-factor of a laser cavity* According to example 5.2 we will again take $\tau_c \cong 150$ ns and assume $\nu \cong 5 \times 10^{14}$ Hz (i.e. $\lambda \cong 630$ nm). From Eq. (5.3.12) we obtain $Q = 4.7 \times 10^8$. Thus, very high $Q$-values can be achieved in a laser cavity and this means that a very small fraction of the energy is lost during one oscillation cycle

## 5.4. STABILITY CONDITION

Consider first a general two-mirror resonator (Fig. 5.6a) and a ray leaving point $P_0$ of a plane $\beta$ inside the resonator e.g. just in front of mirror 1. This ray, after reflection from mirrors 2 and 1, will intersect the plane $\beta$ at some point $P_1$. If we let $r_0$ and $r_1$ be the transverse

# 16

# WAVE OPTICS AND GAUSSIAN BEAMS

A more accurate treatment of optical beams and laser resonators must take into account diffraction and the wave nature of light. Practical laser beams are almost always well enough collimated even under worst conditions, however, that we can describe their diffraction properties using a scalar wave theory, and working in the paraxial wave approximation.

In this chapter, therefore, we introduce the paraxial wave analysis and the equivalent Huygens-Fresnel integral approach for optical beams in free space. We also introduce the lowest and higher-order gaussian mode solutions of these equations as a widely useful set of "normal modes of free space."

The Hermite-gaussian or Laguerre-gaussian modes which we introduce in this chapter are exact and yet mathematically convenient solutions to the paraxial wave equation in free space. They also provide very close (though not quite exact) approximations for the transverse eigenmodes of stable laser resonators with finite diameter mirrors. Gaussian beams are therefore very widely used in analyzing laser beams and related optical systems. Our approach in this chapter is to focus primarily on the mathematical derivation of these modes, whereas in the following chapter we summarize most of the important practical properties of gaussian beams in considerable detail.

## 16.1 THE PARAXIAL WAVE EQUATION

One fundamental way of analyzing free-space wave propagation, using a differential approach, is through the *paraxial wave equation*, which we can derive once again here in the following fashion.

### Derivation of the Paraxial Wave Equation

Electromagnetic fields in free space (or in any uniform and isotropic medium) are governed in general by the scalar wave equation

$$[\nabla^2 + k^2]\ \tilde{E}(x, y, z) = 0, \tag{1}$$

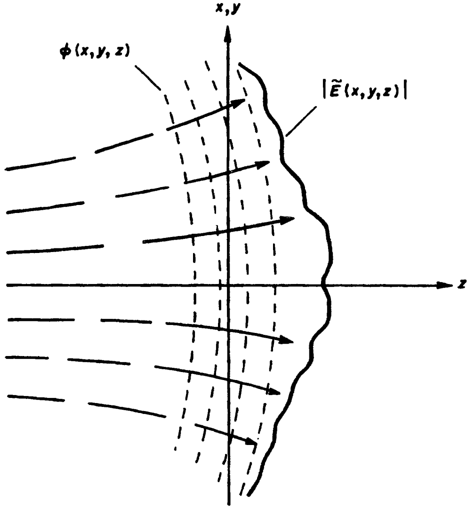

FIGURE 16.1
Transverse amplitude and phase variation of a paraxial optical wave.

where $\tilde{E}(x,y,z)$ is the phasor amplitude of a field distribution that is sinusoidal in time. We will be concerned in this section with optical beams propagating primarily along the $z$ direction, so that the primary spatial dependence of $\tilde{E}(x,y,z)$ will be an $\exp(-jkz)$ variation. This $\exp(-jkz)$ variation has a spatial period of one wavelength $\lambda$ in the $z$ direction.

In addition, for any beam of practical interest the amplitude and phase of the beam will generally have some transverse variation in $x$ and $y$ which specifies the beam's transverse profile, as shown in Figure 16.1; and this transverse amplitude and phase profile will change slowly with distance $z$ due to diffraction and propagation effects. Both the transverse variations across any plane $z$, however, and especially the variation in beam profile with distance along the $z$ axis, will usually be slow compared to the plane-wave $\exp(-jkz)$ variation in the $z$ direction for a reasonably well-collimated beam.

It is then convenient to extract the primary $\exp(-jkz)$ propagation factor out of $\tilde{E}(x,y,z)$, by writing each relevant vector component of the field (such as $E_x$ or $E_y$) in the form

$$\tilde{E}(x,y,z) \equiv \tilde{u}(x,y,z)e^{-jkz}, \tag{2}$$

where $u$ is a complex scalar wave amplitude which describes the transverse profile of the beam. Substituting this into the wave equation 16.1 then yields, in rectangular coordinates, the reduced equation

$$\frac{\partial^2 \tilde{u}}{\partial x^2} + \frac{\partial^2 \tilde{u}}{\partial y^2} + \frac{\partial^2 \tilde{u}}{\partial z^2} - 2jk\frac{\partial \tilde{u}}{\partial z} = 0. \tag{3}$$

Now, we emphasize once again that with the $\exp(-jkz)$ dependence factored out, the remaining $z$ dependence of the wave amplitude $\tilde{u}(x,y,z)$, is caused basically by diffraction effects, and this $z$ dependence will in general be slow compared not only to one optical wavelength, as in $\exp(-jkz)$, but also to the transverse variations due to the finite width of the beam. This slowly varying dependence of $\tilde{u}(x,y,z)$ on $z$ can be expressed mathematically by the *paraxial approximation*

$$\left|\frac{\partial^2 \tilde{u}}{\partial z^2}\right| \ll \left|2k\frac{\partial \tilde{u}}{\partial z}\right| \quad \text{or} \quad \left|\frac{\partial^2 \tilde{u}}{\partial x^2}\right| \quad \text{or} \quad \left|\frac{\partial^2 \tilde{u}}{\partial y^2}\right|. \tag{4}$$

By dropping the second partial derivative in $z$, we thus reduce the exact wave equation 16.3 to the *paraxial wave equation*

$$\frac{\partial^2 \tilde{u}}{\partial x^2} + \frac{\partial^2 \tilde{u}}{\partial y^2} - 2jk\frac{\partial \tilde{u}}{\partial z} = 0. \tag{5}$$

More generally we may write this paraxial wave equation as

$$\nabla_t^2 \tilde{u}(s,z) - 2jk\frac{\partial \tilde{u}(s,z)}{\partial z} = 0, \tag{6}$$

where $s$ refers to the transverse coordinates $s \equiv (x,y)$ or $s \equiv (r,\theta)$, depending on what coordinate system (rectangular or cylindrical) we elect to use, and $\nabla_t^2$ means the laplacian operator operating on these coordinates in the transverse plane. This equation will be the primary governing equation for all the analysis of this and the following several chapters.

### Paraxial Wave Propagation: Finite Difference Approach

The paraxial wave equation can also be turned around and written in the form

$$\frac{\partial \tilde{u}(s,z)}{\partial z} = -\frac{j}{2k}\nabla_t^2 \tilde{u}(s,z). \tag{7}$$

This equation can then be integrated forward in the $z$ direction in order to compute the forward propagation and diffraction spreading of an arbitrary paraxial optical beam. That is, we can employ any suitable numerical differentiation and integration algorithms, first to evaluate the transverse derivative $\nabla_t^2 \tilde{u}(s,z)$ at a given plane $z$, and then to step forward to a new plane $z + \Delta z$. We can thus accomplish numerical forward propagation of an arbitrary optical wavefront, making sure to use adequate numbers of sampling points in both the transverse and longitudinal directions.

This numerical approach, sometimes referred to as the "finite difference approach," has been applied to practical beam propagation problems by several workers. For almost any free-space beam propagation problem that we may consider, however, the integral formulation that we will consider in the next section is probably a better choice for numerical calculations, because of the much greater computational efficiency of fast Fourier transforms that can be employed.

### Validity of the Paraxial Approximation

The paraxial wave equation in either of the above forms is fully adequate for describing nearly all optical resonator and beam propagation problems that arise with real lasers. As perhaps the simplest but most effective way to confirm

## 17

# PHYSICAL PROPERTIES OF GAUSSIAN BEAMS

The previous chapter developed the analytical tools needed for calculating optical-beam propagation in free space. We also need to have, however, a physical and intuitive understanding of the propagation of real optical beams—an understanding which the next two chapters attempt to develop.

In particular, the Hermite-gaussian or Laguerre-gaussian modes which we introduced in the previous chapter are both mathematically convenient, and also provide very good (though not quite exact) approximations to the transverse modes of stable laser resonators with finite diameter mirrors. Gaussian or quasi gaussian beams are therefore very widely used in analyzing laser problems and related optical systems. A good physical as well as mathematical understanding of gaussian beam properties is particularly important. In this chapter we thus review most of the important physical properties of ideal gaussian optical beams in free space.

## 17.1 GAUSSIAN BEAM PROPAGATION

We first look in this section at what the analytic expressions for a lowest-order gaussian beam imply physically in terms of aperture transmission, collimated beam distances, far-field angular beam spread, and other practical aspects of gaussian beam propagation.

### Analytical Expressions

Let us assume a lowest-order gaussian beam characterized by a spot size $w_0$ and a planar wavefront $R_0 = \infty$ in the transverse dimension, at a reference plane which for simplicity we take to be $z = 0$. This plane will henceforth be known for obvious reasons as the *beam waist*, as in Figure 17.1.

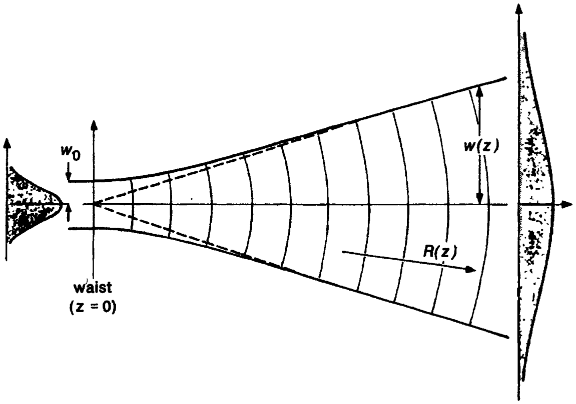

FIGURE 17.1
Notation for a lowest-order gaussian beam diverging away from its waist.

The normalized field pattern of this gaussian beam at any other plane $z$ will then be given by

$$\begin{aligned}\tilde{u}(x,y,z) &= \left(\frac{2}{\pi}\right)^{1/2} \frac{\tilde{q}_0}{w_0\tilde{q}(z)} \exp\left[-jkz - jk\frac{x^2+y^2}{2\tilde{q}(z)}\right] \\ &= \left(\frac{2}{\pi}\right)^{1/2} \frac{\exp[-jkz + j\psi(z)]}{w(z)} \exp\left[-\frac{x^2+y^2}{w^2(z)} - jk\frac{x^2+y^2}{2R(z)}\right],\end{aligned} \tag{1}$$

where the complex radius of curvature $\tilde{q}(z)$ is related to the spot size $w(z)$ and the radius of curvature $R(z)$ at any plane $z$ by the definition

$$\frac{1}{\tilde{q}(z)} \equiv \frac{1}{R(z)} - j\frac{\lambda}{\pi w^2(z)}. \tag{2}$$

In free space this parameter obeys the propagation law

$$\tilde{q}(z) = \tilde{q}_0 + z = z + jz_R, \tag{3}$$

with the initial value

$$\tilde{q}_0 = j\frac{\pi w_0^2}{\lambda} = jz_R. \tag{4}$$

Note that the value of $\lambda$ in these formulas is always the wavelength of the radiation *in the medium in which the beam is propagating.*

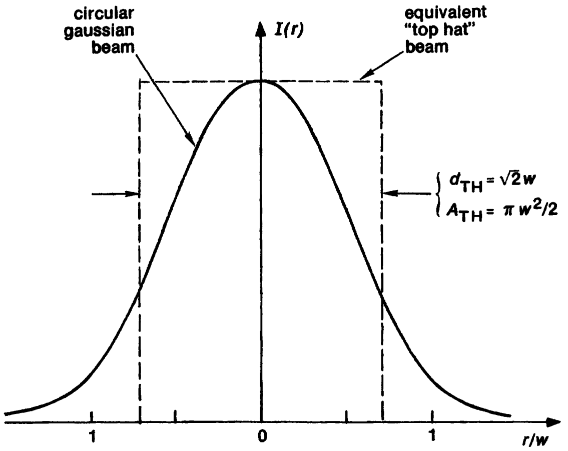

FIGURE 17.2
The equivalent "top hat" radius for a cylindrical gaussian beam.

All the important parameters of this gaussian beam can then be related to the waist spot size $w_0$ and the ratio $z/z_R$ by the formulas

$$w(z) = w_0\sqrt{1 + \left(\frac{z}{z_R}\right)^2},$$

$$R(z) = z + \frac{z_R^2}{z}, \tag{5}$$

$$\psi(z) = \tan^{-1}\left(\frac{z}{z_R}\right).$$

In other words, the field pattern along the entire gaussian beam is characterized entirely by the single parameter $w_0$ (or $\tilde{q}_0$, or $z_R$) at the beam waist, plus the wavelength $\lambda$ in the medium.

### Aperture Transmission

Before exploring the free-space propagation properties of an ideal gaussian beam, we might consider briefly the vignetting effects of the finite apertures that will be present in any real optical system. The intensity of a gaussian beam falls off very rapidly with radius beyond the spot size $w$. How large must a practical aperture be before its truncation effects on a gaussian beam become negligible?

Suppose we define the total power in an optical beam as $P = \int\int |\tilde{u}|^2 \, dA$ where $dA$ integrates over the cross-sectional area. The radial intensity variation of a gaussian beam with spot size $w$ is then given by

$$I(r) = \frac{2P}{\pi w^2} e^{-2r^2/w^2}. \tag{6}$$

The effective diameter and area of a uniform cylindrical beam (a "top hat beam") with the same peak intensity and total power as a cylindrical gaussian beam will

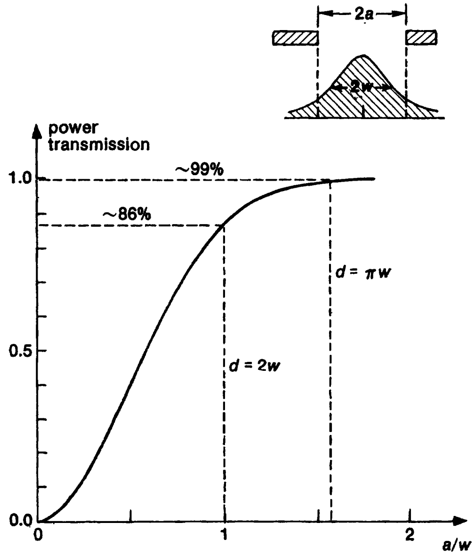

FIGURE 17.3
Power transmission of a cylindrical gaussian beam through a circular aperture.

then be

$$d_{\text{TH}} = \sqrt{2}w \qquad \text{and} \qquad A_{\text{TH}} = \frac{\pi w^2}{2} \tag{7}$$

as shown in Figure 17.2.

An aperture significantly larger than this will be needed, however, to pass a real gaussian beam of spot size $w$ without serious clipping of the beam skirts. The fractional power transfer, for example, for a gaussian beam of spot size $w$ passing through a centered circular aperture of diameter $2a$, as in Figure 17.3, will be given by

$$\text{power transmission} = \frac{2}{\pi w^2}\int_0^a 2\pi r e^{-2r^2/w^2}\,dr = 1 - e^{-2a^2/w^2}. \tag{8}$$

This figure plots this transmission versus aperture radius $a$ normalized to spot size $w$. An aperture with radius $a = w$ transmits $\approx 86\%$ of the total power in the gaussian beam. We will refer to this as the $1/e$ or 86% criterion for aperture size.

A more useful rule of thumb to remember, however, is that an aperture with radius $a = (\pi/2)w$, or diameter $d = \pi w$, will pass just over 99% of the gaussian beam power. We will often use this as a practical design criterion for laser beam apertures, and will refer to it as the "$d = \pi w$" or 99% criterion. (A criterion of $d = 3w$ which gives $\approx 98.9\%$ transmission would obviously serve equally well.) Figure 17.4 illustrates just where some of these significant diameters for a gaussian beam will fall on the gaussian beam profile.

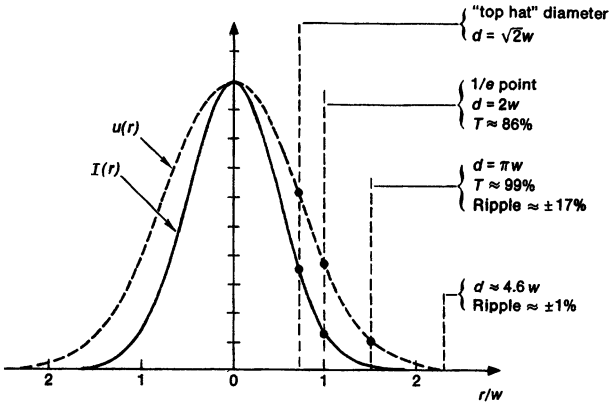

FIGURE 17.4
Significant diameters for hard-edged truncation of a cylindrical gaussian beam. Note that the $d = \pi w$ criterion gives 99% power transmission, but also ±17% intensity ripples and intensity reduction in the near and far fields.

## Aperture Diffraction Effects

Optical designers should take note, however, that sharp-edged apertures, especially circular apertures, even though they may cut off only a very small fraction of the total power in an optical beam, will also produce aperture diffraction effects like those shown in Figure 17.5, which will significantly distort the intensity pattern of the transmitted beam in both the near-field (Fresnel) and far-field (Fraunhofer) regions.

We will show in the following chapter, for example, that the diffraction effects on an ideal gaussian beam of a sharp-edged circular aperture even as large as the $d = \pi w$ criterion will cause near-field diffraction ripples with an intensity variation $\Delta I/I \approx \pm 17\%$ in the near field, along with a peak intensity reduction of $\approx 17\%$ on axis in the far field. We have to enlarge the aperture to $d \approx 4.6w$ to get down to ±1% diffraction ripple effects from a sharp-edged circular aperture.

## Beam Collimation: The Rayleigh Range and the Confocal Parameter

Another important question is how rapidly an ideal gaussian beam will expand due to diffraction spreading as it propagates away from the waist region or, in practical terms, over how long a distance can we propagate a collimated gaussian beam before it begins to spread significantly?

The variation of the beam spot size $w(z)$ with distance as given by Equation 17.5 is plotted in Figure 17.6 for two different waist spot sizes $w_{01}$ and $w_{02} > w_{01}$, with the transverse scale greatly enlarged. The primary point is that as the input spot size $w_0$ at the waist is made smaller, the beam expands more rapidly due to diffraction; remains collimated over a shorter distance in the near field; and diverges at a larger beam angle in the far field.

In particular, the distance which the beam travels from the waist before the beam diameter increases by $\sqrt{2}$, or before the beam area doubles, is given simply

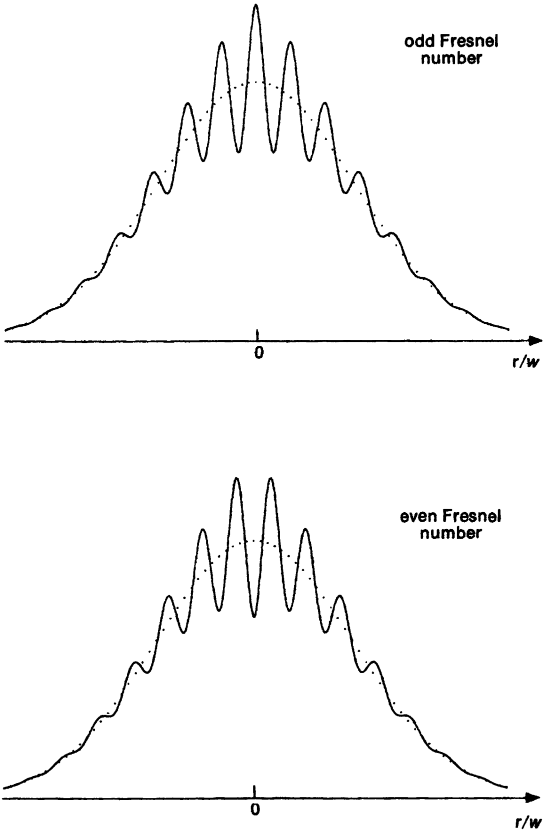

FIGURE 17.5
Near-field Fresnel-diffraction ripples produced by truncation of a gaussian beam.

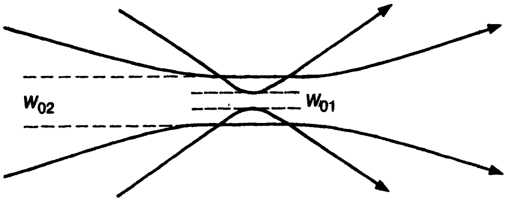

FIGURE 17.6
Diffraction spreading of two gaussian beams with different spot sizes at the waist.

by the parameter

$$z = z_R \equiv \frac{\pi w_0^2}{\lambda} = \text{“Rayleigh range.”} \tag{9}$$

The term *Rayleigh range* is sometimes used in antenna theory to describe the distance $z \approx d^2/\lambda$ that a collimated beam travels from an antenna of aperture diameter $d$ (assuming $d \gg \lambda$) before the beam begins to diverge significantly. We have

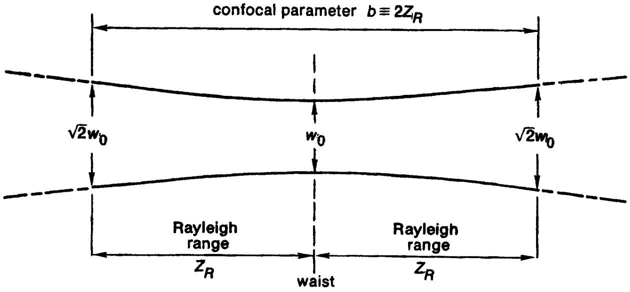

FIGURE 17.7
The collimated waist region of a gaussian beam.

therefore adopted the same term here as a name for the quantity $z_R \equiv \pi w_0^2/\lambda$. The Rayleigh range marks the approximate dividing line between the "near-field" or Fresnel and the "far-field" or Fraunhofer regions for a beam propagating out from a gaussian waist.

To express this same point in another way, if a gaussian beam is focused from an aperture down to a waist and then expands again, the full distance between the $\sqrt{2}w_0$ spot size points is the quantity $b$ given by

$$b = 2z_R = \frac{2\pi w_0^2}{\lambda} = \text{confocal parameter}. \tag{10}$$

This confocal parameter was widely used in earlier writings to characterize gaussian beams. Using the Rayleigh range $z_R \equiv b/2$, as shown in Figure 17.7, seems, however, to give simpler results in most gaussian beam formulas.

### Collimated Gaussian Beam Propagation

Over what distance can the collimated waist region of an optical beam then extend, in practical terms? To gain some insight into this question, we might suppose that a gaussian optical beam is to be transmitted from a source aperture of diameter $D$ with a slight initial inward convergence, as shown in Figure 17.8, so that the beam focuses slightly to a waist with spot size $w_0$ at one Rayleigh range out, and then reexpands to the same diameter $D$ two Rayleigh ranges (or one confocal parameter) out. We will choose the aperture diameter according to the $\pi w$ or 99% criterion, i.e., we will use $D = \pi \times \sqrt{2}\, w_0$ at each end.

The relation between the collimated beam distance and the transmitting aperture size using this criterion is then

$$\text{collimated range} = 2z_R = \frac{2\pi w_0^2}{\lambda} \approx \frac{D^2}{\pi\lambda}. \tag{11}$$

Some representative numbers for this collimated beam range at two different laser wavelengths are illustrated in Figure 17.8 and in Table 17.1. A visible laser with a 1 cm diameter aperture can project a beam having an effective diameter of a few mm with no significant diffraction spreading over a length of 50 meters or more. Such a beam can be used, for example, as a "weightless string" for alignment on a construction project. With the aid of a simple photocell array,

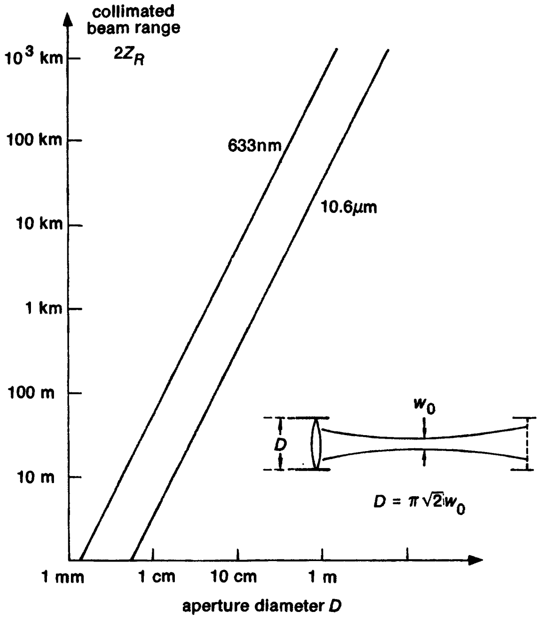

FIGURE 17.8
Collimated gaussian beam ranges versus transmitting aperture diameter $D$, using the $d = \pi w$ criterion.

TABLE 17.1
Collimated Laser Beam Ranges

| *Aperture diameter* $D$ | *Waist spot size* $w_0$ | *Collimated range*, $2z_r$ (10.6 μm) | *Collimated range*, $2r_R$ (633 nm) |
|---|---|---|---|
| 1 cm | 2.25 mm | 3 m | 45 m |
| 10 cm | 2.25 cm | 300 m | 5 km |
| 1 m | 22.5 cm | 30 km | 500 km |

the center of such a beam can easily be found to an accuracy of better than $w/20$, or a small fraction of a mm, over the entire distance.

## Far-Field Beam Angle: The "Top Hat" Criterion

Suppose we next move out into the far field, where the beam size expands linearly with distance, as in Figure 17.9. At what angle does a gaussian beam spread in the far field, that is, for $z \gg z_R$?

From the gaussian beam equations (17.1-17.5), the $1/e$ spot size $w(z)$ for the field amplitude in the far field for a gaussian beam coming from a waist with spot size $w_0$ is given by

$$w(z) \approx \frac{w_0 z}{z_R} = \frac{\lambda z}{\pi w_0} \qquad (z \gg z_R), \tag{12}$$

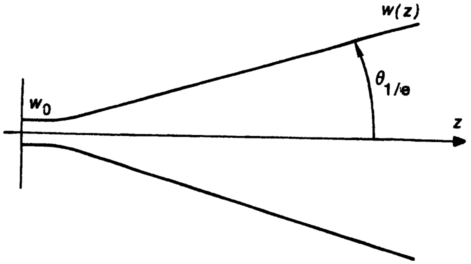

FIGURE 17.9
A gaussian beam spreads with a constant diffraction angle in the far field.

which gives the simple relation

$$w_0 \times w(z) \approx \frac{\lambda z}{\pi} \tag{13}$$

connecting the spot sizes at the waist and in the far field. The far-field angular beam spread for a gaussian beam can then related to the near-field beam size or aperture area in several different ways, depending on how conservative we want to be.

The on-axis beam intensity in the far field, for example, is given by

$$I_{\text{axis}}(z) = \frac{2P}{\pi w^2(z)} \approx \frac{P}{\lambda^2 z^2/2\pi w_0^2}. \tag{14}$$

Hence, the on-axis intensity is the same as if the total power $P$ were uniformly distributed over an area $\pi w^2(z)/2 = \lambda^2 z^2/2\pi w_0^2$. The solid angle for an equivalent "top hat" angular distribution in the far field, call it $\Omega_{\text{TH}}(z)$, is thus given by

$$\Omega_{\text{TH}} = \frac{\pi w^2(z)}{2z^2} = \frac{\lambda^2}{2\pi w_0^2}. \tag{15}$$

At the same time, the "equivalent top hat" definition of the source area at the waist is given from Equation 17.7 by $A_{\text{TH}} = \pi w_0^2/2$. The product of these two quantities is thus given by

$$A_{\text{TH}} \times \Omega_{\text{TH}} = \left(\frac{\lambda}{2}\right)^2. \tag{16}$$

The source aperture size (at the waist) and the far-field solid angular spread thus have a product on the order of the wavelength $\lambda$ squared, although the exact numerical factor will depend on the definitions we choose for the area and the solid angle, as we will see in more detail later.

### Far-Field Beam Angle: The 1/e Criterion

Another and perhaps more reasonable definition for the far-field beam angle is to use the $1/e$ or 86% criterion for the beam diameter, so that the far field half-angular spread is defined by the width corresponding to the $1/e$ point for the $E$ field amplitude at large $z$.

With this definition, the half-angle $\theta_{1/e}$ out to the $1/e$ amplitude points in the far-field beam is given, as shown in Figure 17.9, by

$$\theta_{1/e} = \lim_{z\to\infty} \frac{w(z)}{z} = \frac{\lambda}{\pi w_0}. \tag{17}$$

Twice this angle then gives a full angular spread of

$$2\theta_{1/e} = \frac{2\lambda}{\pi w_0}, \tag{18}$$

which can be interpreted as a more precise formulation, valid for gaussian beams, of the approximate relation $\Delta\theta \approx \lambda/d$ that we gave in Chapter 1. We can then define the gaussian beam solid angle $\Omega_{1/e}$ on this same basis as the circular cone defined by this angular spread, or

$$\Omega_{1/e} = \pi\theta_{1/e}^2 = \frac{\lambda^2}{\pi w_0^2}. \tag{19}$$

This cone will, as noted in the preceding, contain 86% of the total beam power in the far field.

Suppose we use the same $1/e$ criterion to define the effective radius of the input beam at the beam waist (ignoring the fact that an aperture of radius $a = w_0$ at the waist would actually produce some very substantial diffraction effects on the far-field beam pattern). Then the product of the effective source aperture area $A_{1/e} \equiv \pi w_0^2/2$ and the effective far-field solid angle $\pi\theta_{1/e}^2$ using these $1/e$ definitions becomes

$$A_{1/e}\Omega_{1/e} = \pi w_0^2 \times \pi\theta_{1/e}^2 = \lambda^2. \tag{20}$$

This is a precise formulation for gaussian beams of a very general antenna theorem which states that

$$\iint A(\Omega)\, d\Omega = \lambda^2 \tag{21}$$

This theorem says in physical terms that if we measure the effective capture area $A(\Omega)$ of an antenna for plane-wave radiation arriving from a direction specified by the vector angle $\Omega = (\theta, \phi)$, and then integrate these measured areas over all possible arrival angles as specified by $d\Omega$, the result (for a lossless antenna of *any* form) is always just the measurement wavelength $\lambda$. This result is valid for any kind of antenna, at radio, microwave or optical wavelengths.

### Far-Field Beam Angle: Conservative Criterion

Finally, as a still more conservative way of expressing the same points, we might use the $d = \pi w$ or 99% criterion instead of the $1/e$ criterion to define both the effective source aperture size and the effective far field solid angle. We might then say that a source aperture of diameter $d = \pi w_0$ transmitting a beam of initial spot size $w_0$ will produce a far-field beam with 99% of its energy within a cone of full angular spread $2\theta_\pi = \pi w(z)/z$. On this basis the source aperture area, call it $A_\pi$ is $\pi d^2/4$ and the beam far-field solid angle is $\Omega_\pi = \pi\theta_\pi^2$; and these are related by the more conservative criterion

$$A_\pi\Omega_\pi = \left(\frac{\pi}{2}\right)^4 \lambda^2 \approx 6\lambda^2. \tag{22}$$

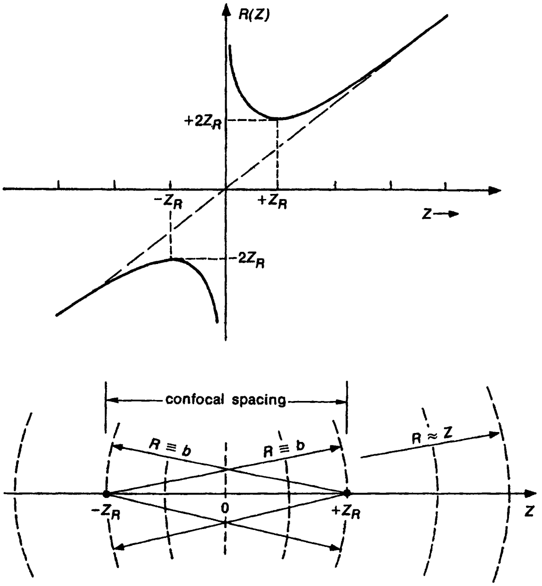

FIGURE 17.10
Radius of curvature for the wavefront of a gaussian beam, versus distance from the waist.

None of the criteria we have introduced here for defining effective aperture size and effective solid angle are divinely ordained, and which of them we use should depend largely on what objective we have in mind.

### Wavefront Radius of Curvature

We can next look at how the wavefront curvature of a gaussian beam varies with distance. The radius of curvature $R(z)$ of a gaussian beam has a variation with distance given analytically by

$$R(z) = z + \frac{z_R^2}{z} \approx \begin{cases} \infty & \text{for} \quad z \ll z_R \\ 2z_R & \text{for} \quad z = z_R \\ z & \text{for} \quad z \gg z_R \end{cases} . \tag{23}$$

This is plotted against normalized distance in Figure 17.10(a).

The wavefront is flat or planar right at the waist, corresponding to an infinite radius of curvature or $R(0) = \infty$. As the beam propagates outward, however, the wavefront gradually becomes curved, and the radius of curvature $R(z)$ drops rather rapidly down to finite values (see Figure 17.10). For distances well beyond the Rayleigh range $z_R$ the radius then increases again as $R(z) \approx z$, i.e., the gaussian beam becomes essentially like a spherical wave centered at the beam waist. What this means in physical terms is that the center of curvature of the wavefront starts out at $-\infty$ for a wavefront right at the beam waist, and then moves monotonically inward toward the waist, as the wavefront itself moves outward toward $z \rightarrow +\infty$.

### Confocal Curvatures

The minimum radius of curvature occurs for the wavefront at a distance from the waist given by $z = z_R$, with the radius value $R = b = 2z_R$. This means that at this point the center of curvature for the wavefront at $z = +z_R$ is located at $z = -z_R$, and vice versa, as illustrated in Figure 17.10.

This particular spacing has a special significance in stable resonator theory. Suppose the curved wavefronts $R(z)$ at $\pm z_R$ are matched exactly by two curved mirrors of radius $R$ and separation $L = R = b = 2z_R$. Since the focal point of a curved mirror of radius $R$ is located at $f = R/2$, the focal points of these two mirrors then coincide exactly at the center of the resonator. The two mirrors are said to form a *symmetric confocal resonator*, thus giving rise to the *confocal parameter* $b \equiv 2z_R \equiv 2\pi w_0^2/\lambda$. Such a resonator has certain particularly interesting mode properties which we will explore later.

## REFERENCES

Further discussion of the concept of the "Rayleigh range" can be found in J. F. Ramsay, "Tubular beams from radiating apertures," in *Advances in Microwaves, Vol. 3*, ed. by L. F. Young (Academic Press, New York, 1968), p. 127.

Earlier considerations of the same ideas by Lord Rayleigh (J. W. Strutt) himself can be found in his papers "On images formed with or without reflection or refraction," *Phil. Mag.* **11**, 214–218 (1881), and "On pinhole photography," *Phil. Mag.* **31**, 87–89 (1891).

## Problems for 17.1

1. *Gaussian beam transmission through a square aperture.* Find the power transmission for a gaussian beam through a *square* aperture with sides of length $2a$, in analogy to the circular aperture results given in the text.

2. *Criteria for centering accuracy of a circular aperture.* Suppose a gaussian beam is transmitted through a circular aperture of diameter $d = \pi w$. How critical is the centering of the laser beam axis with respect to the aperture position (or vice versa)? Attempt to evaluate the decrease in beam transmission versus the displacement between beam and aperture centers, using either approximate mathematical methods or computer evaluation.

3. *Setting tolerances on beam collimation and far-field beam angle.* A laser oscillator is designed to give a collimated beam at its output plane with a specified spot size $w_0$ and a collimated wavefront with radius of curvature $R_0 = \infty$. Due to manufacturing tolerances, however, the actual output wavefront may come out slightly spherical. Suppose we establish as a practical tolerance for good collimation that the far-field beam angle of the laser output beam should not vary by more than 10% from its design value. What is the resulting tolerance on $R_0$ (or $1/R_0$) at the laser's output plane for a fixed value of $w_0$? How much wavefront distortion does this represent, expressed in terms of fractional wavelengths of wavefront distortion at the $1/e$ radius of the output beam?

4. *Simulating an annular beam with positive and negative gaussians.* A circularly symmetric beam with a hole in the center can be simulated by superposing

# 19

# STABLE TWO-MIRROR RESONATORS

The simplest kind of optical resonator consists of just two curved mirrors set up facing each other. If the curvatures of these two mirrors correspond to a stable periodic focusing system, and if their transverse dimensions are large enough so that we can neglect edge-diffraction effects, then these mirrors can in essence trap a set of lowest-order and higher-order gaussian modes or beams that will bounce back and forth between the two mirrors. These trapped Hermite-gaussian modes form, to a first approximation, a set of resonant modes for the two-mirror cavity.

Simple two-mirror cavities such as this are widely used in practical lasers, and the properties of these stable gaussian resonator modes form part of the basic lore of laser physics. In this chapter, therefore, we give a fairly detailed account of these properties and of how they are derived from gaussian beam theory. In addition we give a brief survey of the (usually) small deviations from ideal gaussian beam behavior that occur because of finite mirror sizes, including in particular the finite diffraction losses in finite-diameter resonators.

In later chapters we will discuss the additional complexities that arise in analyzing multielement resonators which contain, for example, intracavity lenses or gaussian apertures, as well as the quite different and nongaussian modes associated with unstable optical resonators. Even in these situations, however, the stable two-mirror gaussian concepts introduced in this section will prove very useful in understanding and explaining the behavior of these more complex resonators.

## 19.1 STABLE GAUSSIAN RESONATOR MODES

Suppose we have a gaussian beam with a certain waist size and waist location, as in Figure 19.1, and suppose that we then fit a pair of curved mirrors to this beam at any two points along the beam, as also illustrated in Figure 19.1. If the radii of curvature of the mirrors are exactly matched to the wavefront radii of the gaussian beam at those two points, and if the transverse size of the mirrors is substantially larger than the gaussian spot size of the beam, each of these mirrors will in essence reflect the gaussian beam exactly back on itself, with exactly reversed wavefront curvature and direction.

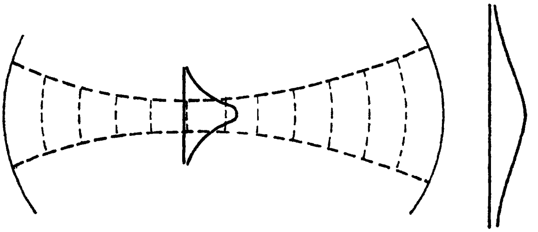

FIGURE 19.1
A gaussian beam can be trapped by two mirrors of the proper curvature and spacing.

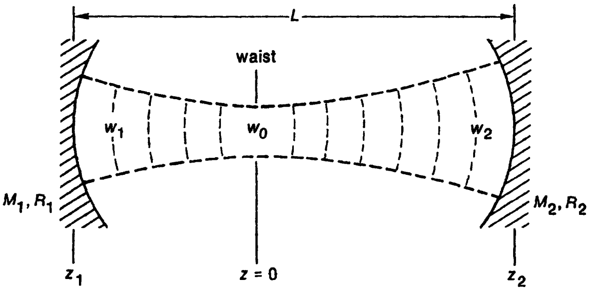

FIGURE 19.2
Notation and analytical model for analyzing a simple stable two-mirror cavity.

These two mirrors can thus trap the gaussian beam as a standing wave between the two mirrors, with, if the mirrors are large enough in size, negligible diffraction or "spillover" losses past the edges of the mirrors. The two mirrors thus form an optical resonator which can support both the lowest-order gaussian mode, and also higher-order Hermite-gaussian or Laguerre-gaussian modes, as resonant modes of the cavity. We will see in this section that this simple description is, in essence, exactly what happens in elementary stable two-mirror gaussian resonators.

### Stable Two-Mirror Resonator Analysis

In practice, instead of being given a gaussian beam and asked to fit mirrors to it, we are much more likely to be given two curved mirrors $M_1$ and $M_2$ with radii of curvature $R_1$ and $R_2$ and spacing $L$, and asked to find the right gaussian beam that will just fit properly between these two mirrors. To analyze this situation we can use the model in Figure 19.2, assuming that the gaussian beam will have an (initially unknown) spot size $w_0$ or Rayleigh range $z_R \equiv \pi w_0^2/\lambda$, and that the mirrors will be located at distances $z_1$ and $z_2$ from the (initially unknown) location of the beam waist.

The essential conditions are then that the wavefront curvature $R(z)$ of the gaussian beam, as given by gaussian beam theory, must match the mirror curvature at each mirror, taking into account the specified mirror spacing $L$. This

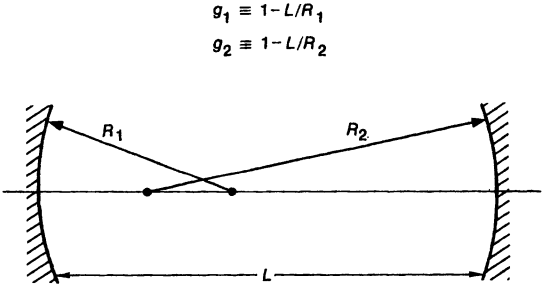

FIGURE 19.3
The resonator $g$ parameters.

provides us with three equations, namely,

$$\begin{aligned} R(z_1) &= z_1 + z_R^2/z_1 = -R_1, \\ R(z_2) &= z_2 + z_R^2/z_2 = +R_2, \end{aligned} \tag{1}$$

and

$$L = z_2 - z_1. \tag{2}$$

The minus sign in the first of these equations arises because of a difference in the sign conventions that we use in in describing beam wavefronts or in describing resonator mirrors. The gaussian wavefront curvature $R(z)$ is usually taken as positive for a diverging beam, or negative for a converging beam, traveling to the right; whereas the mirror curvatures $R_1$ and $R_2$ are usually taken as positive numbers for mirrors that are concave inward, i.e., as seen looking out from within the resonator, and as negative numbers for mirrors that are convex as seen from inside the resonator.

## The g Parameters

We must then invert these three equations in order to find the gaussian beam parameters $z_R$, $z_1$ and $z_2$ in terms of the specified mirror curvatures and spacing $R_1$, $R_2$ and $L$. Before doing this, however, it is customary to define a pair of "resonator $g$ parameters," $g_1$ and $g_2$, which were introduced in the early days of laser theory to describe laser resonators, and have since become standard notation in the field. These parameters are given by

$$g_1 \equiv 1 - \frac{L}{R_1} \quad \text{and} \quad g_2 \equiv 1 - \frac{L}{R_2}. \tag{3}$$

We will see more of their physical significance later.

In terms of these parameters we can then find that the trapped gaussian beam in Figure 19.2 will have a unique Rayleigh range given by

$$z_R^2 = \frac{g_1 g_2 (1 - g_1 g_2)}{(g_1 + g_2 - 2g_1 g_2)^2} L^2, \tag{4}$$

and that the locations of the two mirrors relative to the gaussian beam waist will be given by

$$z_1 = \frac{g_2(1-g_1)}{g_1+g_2-2g_1g_2}L \quad \text{and} \quad z_2 = \frac{g_1(1-g_2)}{g_1+g_2-2g_1g_2}L. \tag{5}$$

(Note that if mirror $M_1$ is located to the left of the beam waist, so that the waist is inside the resonator as in Figure 19.2, then $z_1$ as measured from the waist will be a negative number.)

It is also useful to write out the waist spot size $w_0$, which is given by

$$w_0^2 = \frac{L\lambda}{\pi}\sqrt{\frac{g_1g_2(1-g_1g_2)}{(g_1+g_2-2g_1g_2)^2}}, \tag{6}$$

and the spot sizes $w_1$ and $w_2$ at the ends of the resonator, which are given by

$$w_1^2 = \frac{L\lambda}{\pi}\sqrt{\frac{g_2}{g_1(1-g_1g_2)}} \quad \text{and} \quad w_2^2 = \frac{L\lambda}{\pi}\sqrt{\frac{g_1}{g_2(1-g_1g_2)}}. \tag{7}$$

These quantities depend only on the resonator $g$ parameters defined in the preceding, and on the quantity $\sqrt{L\lambda/\pi}$ which we will discuss in the following.

### Resonator Stability Diagram

It is immediately obvious from Equations 19.4 to 19.7 that real and finite solutions for the gaussian beam parameters and spot sizes can exist only if the $g_1, g_2$ parameters are confined to a stability range defined by

$$0 \le g_1g_2 \le 1. \tag{8}$$

We refer to this as a stability range because this is also exactly the condition required for two mirrors with radii $R_1$ and $R_2$ and spacing $L$ to form a stable periodic focusing system for rays, as analyzed earlier in Chapter 15.

In the early days of gaussian resonator theory this stability criterion was immediately translated into the *resonator stability diagram* shown in Figure 19.4. Every two-mirror optical resonator can then be characterized by the parameters $g_1 = 1 - L/R_1$ and $g_2 = 1 - L/R_2$, and hence represented by a point in the $g_1, g_2$ plane. If this point falls in the shaded stable region, shown in Figure 19.4, the mirrors correspond to a *stable periodic focusing system*, and the resonator (if the mirrors are large enough transversely) will trap a family of lowest and higher-order gaussian modes with gaussian beam parameters given by Equations 19.4 through 19.7. Such a stable resonator will thus have a unique set of gaussian transverse resonator modes.

If the point $g_1, g_2$ instead falls in any of the unstable regions outside the shaded area, the mirrors will correspond to an *unstable periodic focusing system*, and no gaussian beam that will fit properly between the mirrors can be found. These mirror configurations correspond to the very different (but also very useful) *unstable optical resonators* that we will discuss in a later chapter.

Optical ray theory and gaussian mode theory thus have a close connection, which we will study in more detail later on, even though the diffraction effects that are an integral part of gaussian beam theory are entirely neglected in the

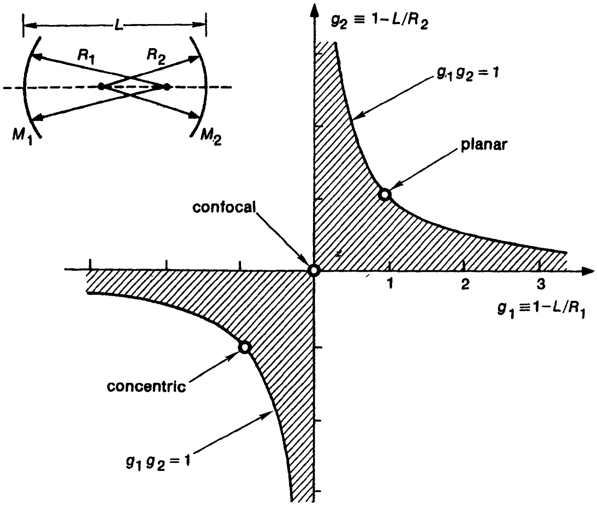

FIGURE 19.4
The stability diagram for a two-mirror optical resonator.

optical ray theory. Note also that these distinctions between stability and instability depend only on the $g$ parameters, and are (to first order) entirely independent of either the optical wavelength or the transverse size or dimensions of the resonator. In the following section we will examine in more detail the various types of resonators that occur in various regions of the stability diagram, and the various practical properties of these resonators.

### Resonator Circle Diagrams

An alternative and less commonly used graphical method for interpreting the gaussian beam parameters in stable two-mirror resonators is the *circle diagram* of Deschamps. Suppose again that two mirrors of radii $R_1$ and $R_2$ are set up with spacing $L$. If we then draw circles with *diameters* $R_1$ and $R_2$ tangent to the concave side of each of these mirrors, as shown in Figure 19.5, the intersection of these two circles is a necessary and sufficient condition for the existence of a stable gaussian mode in the resonator; and moreover the waist location and its relative size in the resonator is determined by the line joining the intersection of these two circles.

## REFERENCES

The standard review article in the journal literature on gaussian resonator modes is H. Kogelnik and T. Li, "Laser beams and resonators," *Appl. Optics* **5**, 1550–1567 (October 1966). See also the many other references cited therein.

Many of the same ideas on gaussian beams as eigenmodes of stable periodic focusing systems developed quite independently of laser resonator theory, and somewhat

(b) $C_2$ located to the left of the left-hand mirror.

(c) $C_2$ located to the right of the right-hand mirror (i.e., a divergent right-hand mirror).

2. *Symmetric cavity with central thin lens.* A stable optical cavity of length $L$ is to be formed by two symmetric mirrors with radius of curvature $R$, plus a thin lens of focal length $f$ placed exactly in in the center of the cavity. Calculate how to fit a stable gaussian beam within this cavity, and then describe the stability limits of the cavity and the profile of the gaussian beam within the cavity for different choices of $f$, $R$, and $L$. Describe in particlar where the waist (or waists) will occur in the cavity under different conditions. Hint: Split the central thin lens into two lenses with half the focal power each, and then put a reference right in the middle between these two thin lenses.

3. *Standing-wave cavity fields.* The standing-wave field $\tilde{u}(x, y, z)$ for a single lowest-order gaussian made inside an ideal stable laser cavity is the sum of two identical but oppositely traveling gaussian beams. Write down this field expression, in terms of the waist spot size $w_0$ characteristic of the resonator and the coordinates $x, y, z$. Do the standing-wave fields oscillate in the same time-phase everywhere inside the laser cavity?

## 19.2 IMPORTANT STABLE RESONATOR TYPES

To gain more insight into the general properties of stable gaussian resonators, let us now survey some of the characteristics associated with resonators at various different points of interest in the stability diagram introduced in the previous section.

### (1) Symmetric Resonators

Perhaps the simplest resonator configurations to analyze are symmetric resonators, which have mirror curvatures $R_1 = R_2 = R$, and hence $g$ parameters $g_1 = g_2 = g = 1 - L/R$. The waist of the gaussian resonant mode is then obviously in the center of the resonator, with waist and end mirror spot sizes given by

$$w_0^2 = \frac{L\lambda}{\pi}\sqrt{\frac{1+g}{4(1-g)}} \quad \text{and} \quad w_1^2 = w_2^2 = \frac{L\lambda}{\pi}\sqrt{\frac{1}{1-g^2}}. \tag{9}$$

All these symmetric resonators obviously lie along the +45° diagonal through the origin in the $g$ plane, with an allowed range from $g = 1$ (planar mirror case), through $g = 0$ (symmetric confocal case), to $g = -1$ (concentric or spherical case).

Figure 19.6 shows how the resonator spot sizes change as the $g$ value is varied along this range, for example, by steadily increasing the mirror curvatures while keeping the mirror spacing fixed.

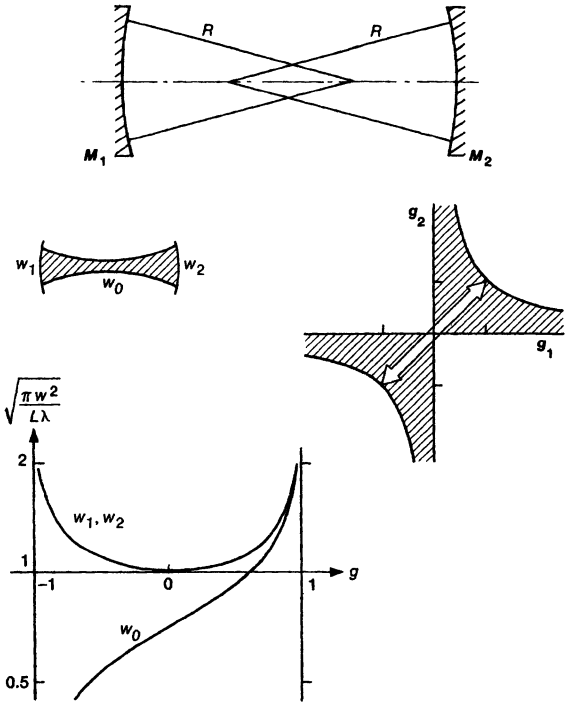

FIGURE 19.6
Symmetric stable resonators lie along the diagonal axis in the $g_1, g_2$ plane.

## (2) Half-Symmetric Resonators

Another elementary system is the half-symmetric resonator of Figure 19.7, in which one mirror is planar, $R_1 = \infty$, and the other curved, so that $g_1 = 1$ and $g_2 = g = 1 - L/R_2$. Such a resonator is obviously equivalent to half of a symmetric system that is twice as long. The waist in this situation will be located on mirror number 1, with spot sizes given by

$$w_0^2 = w_1^2 = \frac{L\lambda}{\pi}\sqrt{\frac{g}{1-g}} \qquad \text{and} \qquad w_2^2 = \frac{L\lambda}{\pi}\sqrt{\frac{1}{g(1-g)}}. \tag{10}$$

The allowed range for $g_2 = g$ is now from +1 to 0, corresponding to a vertical line between the points (1, 1) and (1, 0) in the stability diagram.

## (3) Symmetric Confocal Resonator

The central point in the stability diagram, and in some sense a central type of stable optical resonator, is the *symmetric confocal stable resonator*, which is characterized by the values $R_1 = R_2 = L$ and $g_1 = g_2 = 0$ (Figure 19.8). This is referred to as a confocal resonator because the focal points of the two end mirrors (which are located at $R/2$ out from the mirror) coincide with each other at the center of the resonator. We have already seen in Figure 15.10 that confocality

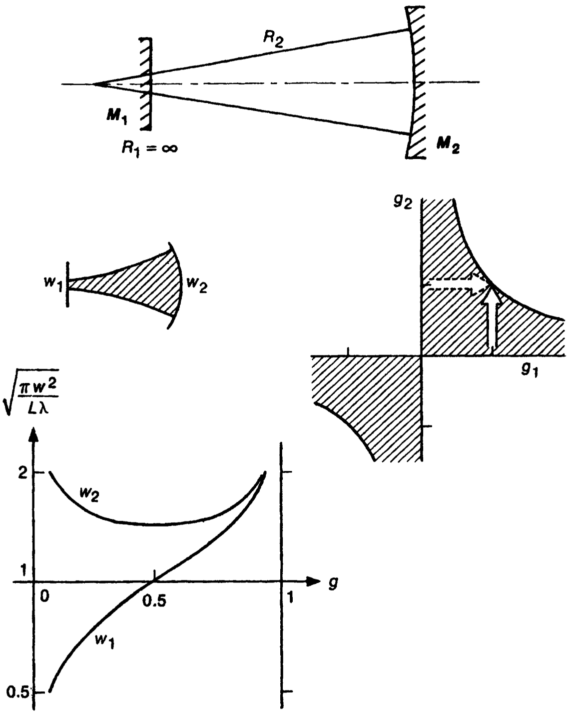

FIGURE 19.7
Half-symmetric resonators have their waist at the plane-mirror end of the cavity.

corresponds to the condition for a gaussian beam in which the center of curvature of each mirror is located exactly on the opposite mirror. The two mirrors are thus spaced from each other by exactly two Rayleigh ranges or by exactly the waist length of the trapped gaussian beam.

The spot sizes at the center and at the end mirrors of a confocal resonator are then given by

$$w_0^2 = \frac{L\lambda}{2\pi} \quad \text{and} \quad w_1^2 = w_2^2 = \frac{L\lambda}{\pi}. \tag{11}$$

The spot sizes on the end mirrors thus correspond exactly to the scale factor $\sqrt{L\lambda/\pi}$ that arises in all types of stable resonators, whereas the spot size at the central waist is smaller by $1/\sqrt{2}$.

Table 19.1 gives some typical values of this spot size for resonators of different lengths at the typical wavelengths of 633 nm for the He-Ne laser and at 10.6 $\mu$m for the $CO_2$ lasers. The Table also shows the laser tube diameter that might be associated with this length of resonator, using the rule of thumb that aperture diameter $d = \pi w$.

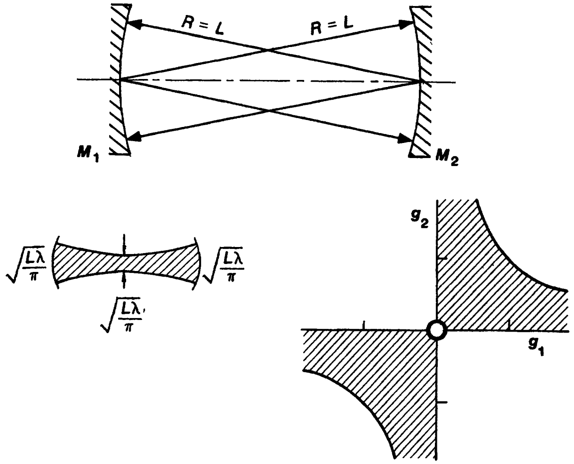

FIGURE 19.8
The symmetric confocal resonator is a special case, located exactly at the origin of the stability diagram.

TABLE 19.1
Confocal Resonator Spot Sizes

| *Cavity length, L* | *10 cm* | *1 m* | *10 m* |
|---|---|---|---|
| Spot size, $w_1 = \sqrt{L\lambda/\pi}$: | | | |
| $\lambda = 633$ nm | 150 $\mu$m | 0.4 mm | 1.5 mm |
| $\lambda = 10.6$ $\mu$m | 600 $\mu$m | 1.7 mm | 6 mm |
| Tube diameter, $\pi w_1$: | | | |
| $\lambda = 633$ nm | 0.4 mm | 1.2 mm | 4 mm |
| $\lambda = 10.6$ $\mu$m | 1.7 mm | 5 mm | 1.7 cm |

These mode diameters are significantly smaller than the diameters of the laser rods or tubes that we might want to use to obtain reasonable laser power outputs at these wavelengths. Finding ways to increase the diameter of the stable gaussian modes (or finding new resonator designs which inherently have larger mode volumes) is one of the primary design objectives in most laser designs.

The confocal resonator in fact has overall the smallest average spot diameter along its length of any stable resonator, although we will see that other resonators may have a smaller waist size at one spot within the resonator. The confocal resonator is also highly insensitive to misalignment of either mirror. Tilting of either mirror still leaves the center of curvature located on the other mirror surface, and merely displaces the optic axis of the resonator by a small amount. The confocal resonator can thus be very useful, for example, as a trial resonator design when we are first attempting to obtain laser oscillation from a laser medium whose gain is small or uncertain. The small mode size then means very small diffraction losses, and the alignment insensitivity means that critical mirror alignment should not be necessary to get the laser to oscillate.

The confocal resonator is also useful for power or energy measurements, in which we simply want to know how much power or energy is available in some laser medium, without consideration of mode control requirements. A confocal

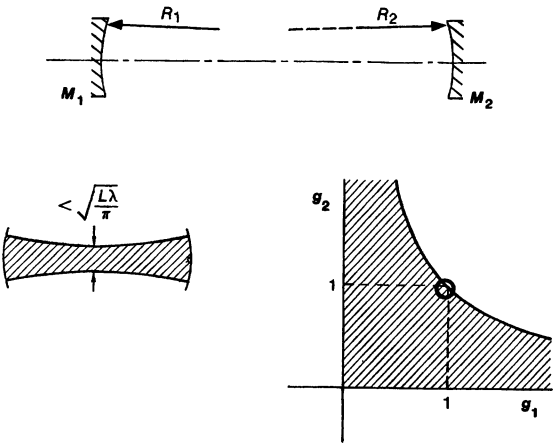

FIGURE 19.9
The long-radius or near-planar resonator can have larger mode volume, but is very sensitive to mirror misalignment.

resonator is then likely to oscillate in a combination of lowest and higher-order modes that will fill the entire volume of the laser medium and extract essentially all the stimulated emission available from the laser medium.

The small average size of the confocal modes, on the other hand, means that the lowest-order or $\text{TEM}_{00}$ confocal mode will not be very effective in extracting power from larger-diameter gain media. Multimode oscillation, as in the power measurement situation, will mean large far-field diffraction spreading of the laser output beam.

### (4) Long-Radius (Near-Planar) Resonators

Another elementary resonator configuration, and one that was used in many of the earliest laser devices, is the *near-planar* or *long-radius stable resonator* of Figure 19.9. A planar or flat-mirror resonator can be regarded as the limiting situation of a long-radius stable resonator as the radii of curvature of the two mirrors go to infinity. The resonator parameters then become $R_1 \approx R_2 \approx \infty$ and $g_1 \approx g_2 \approx 1$. If we let $R_1 = R_2 = R$, the spot sizes in this situation all become large and essentially equal, in the form

$$w_0^2 \approx w_1^2 \approx w_2^2 \approx \frac{L\lambda}{\pi} \times \sqrt{\frac{R}{2L}} \qquad \text{for } R \gg L. \tag{12}$$

In gaussian beam terms the long-radius resonator has a very long and large waist, of which the resonator encompasses only a very short central part. As the mirror radii become infinite the spot sizes become infinite also, though only very slowly, with the radius increasing as $(R/2L)^{1/4}$. The exactly planar resonator occurs right on the stability boundary, at $g_1 = g_2 = 1$, and so the gaussian theory fails at and beyond that point.

Long-radius resonators, although they can have larger mode volumes, are generally avoided in practical laser designs because of their very great alignment difficulties. Since the centers of curvature of the mirrors are cantilevered far out beyond the ends of the resonator, at distances $\pm R$, very delicate angular

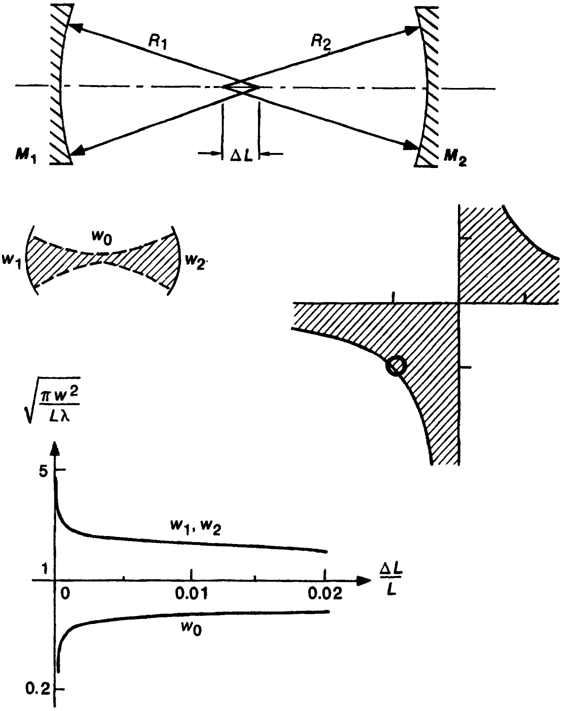

FIGURE 19.10
The near-concentric resonator can have large spot sizes at the ends, but is also very sensitive to misalignment.

alignment of the mirrors becomes necessary if the optical axis of the resonator (which passes through these two centers of curvature) is to be kept aligned within the center of the laser medium itself. Long-radius mirrors are also difficult to manufacture and to test. Note, for example, that for a 2.5 cm diameter mirror with a 50-m radius of curvature, the total sag at the center of the mirror relative to the edges is only $\approx 1.5\ \mu$m. At the same time the spot size enhancement factor for a laser resonator that is $L = 50$ cm long is only $(R/2L)^{1/4} \approx 2.7$.

### (5) Near-Concentric Resonators

The *near-concentric stable resonator* is another design which is on the boundary of the stability region, and which can give large spot sizes at the end mirrors, but now with a vanishingly small spot size in the center of the resonator, as illustrated in Figure 19.10.

For a near-concentric resonator, in which the cavity length $L$ is less than the sum of the two radii $R_1 + R_2$ by the small amount $\Delta L$, the resonator parameters are given by $R_1 \approx R_2 \approx R = L/2 + \Delta L$ and $g_1 \approx g_2 = -1 + \Delta L/R$. The spot size at the central waist is then given by

$$w_0^2 \approx \frac{L\lambda}{\pi} \times \sqrt{\frac{\Delta L}{4L}} \qquad \text{for } \Delta L \ll L, \tag{13}$$

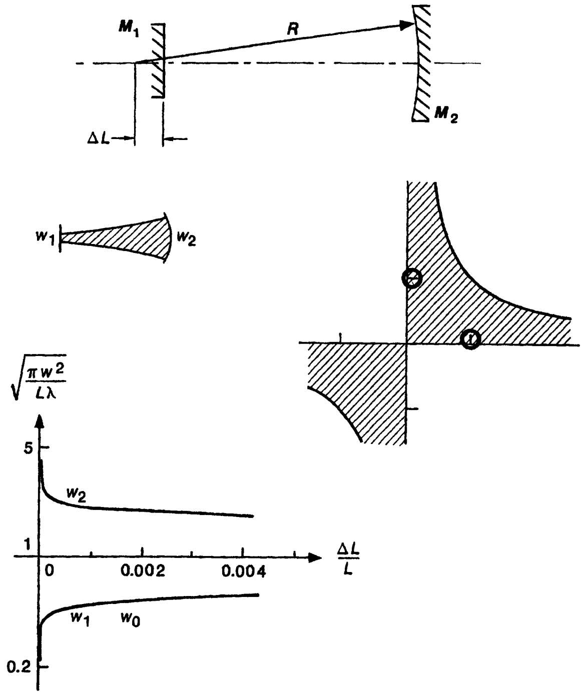

FIGURE 19.11
The near-hemispherical resonator is widely used in practical laser oscillators.

and the end-mirror spot sizes by

$$w_1^2 = w_2^2 \approx \frac{L\lambda}{\pi} \times \sqrt{\frac{4L}{\Delta L}} \qquad \text{for } \Delta L \ll L. \tag{14}$$

The mirror radii are now physically reasonable, and the spot sizes can be adjusted in operation by using translatable end-mirror mounts. The mirrors can then be pulled slowly apart in order to bring the resonator closer to or even across the stability boundary, by making the incremental length $\Delta L$ small or even negative.

The central portion of the resonator, where the spot size becomes very small, is then not very useful, at least for laser power extraction. More seriously, the mirror centers of curvature now become very close to each other at the center of the cavity, as illustrated in Figure 19.10. Hence this resonator again becomes very sensitive to large axis misalignments caused by very small mirror misalignments.

## (6) Hemispherical Resonators

The resonator design that is by far the most commonly used in practical stable-resonator lasers, such as, for example, most medium and low-power gas lasers, is the *near-hemispherical* or *half-concentric stable resonator*, of Figure 19.11, for which the resonator parameters are $R_1 = \infty$ and $R_2 = L + \Delta L$, and hence $g_1 = 1$ and $g_2 = \Delta L/L \approx 0$. This resonator is like half of a near-concentric

## 8.6. MODE-LOCKING

Let us now consider a laser which is oscillating on a rather large number of longitudinal modes. Under ordinary circumstances, the phases of these modes will have random values and, for cw oscillation, the beam intensity will show a random time behavior. As an example, Fig. 8.15 shows the time behavior of the square of the electric field amplitude, $|A(t)|^2$, of the output beam for the case of $N = 31$ oscillating modes, each with the same amplitude $E_0$, and evenly separated in frequency by the frequency difference $\Delta\nu$ between consecutive longitudinal modes. One sees that the output beam consists of a random sequence of light pulses. Despite this randomness, since these pulses arise from the sum of $N$ frequency components which are evenly spaced in frequency, the pulse waveform of Fig. 8.15 has the following general properties which are a characteristic of a Fourier series: (i) The waveform is periodic with a period $\tau_p = 1/\Delta\nu$. (ii) Each light pulse of the random waveform has a duration $\Delta\tau_p$ roughly equal to $1/\Delta\nu_L$ where, $\Delta\nu_L = N\Delta\nu$, is the total oscillating bandwidth. Thus, for lasers with relatively large gain bandwidths, such as solid-state, dye or semiconductor lasers, $\Delta\nu_L$ may be comparable to this gain bandwidth and hence short noise pulses, with durations of picoseconds or less, can be produced. Note that, since the response time of a conventional photodetector is usually much longer than a few picoseconds, one does not resolve this complex time behavior in the detected output of a, random phase, multimode laser, and instead its average value is monitored. This value is simply the sum of powers in the modes and hence is proportional to $NE_0^2$.

Let us now suppose that the oscillating modes, while still having equal or comparable amplitudes, are somehow made to oscillate with some definite relation between their phases. Such a laser is referred to as mode locked, and the process by which the modes are made to adopt a definite phase relation is referred to as *mode locking*.(15) Mode-locked lasers will be considered at some length in this section.

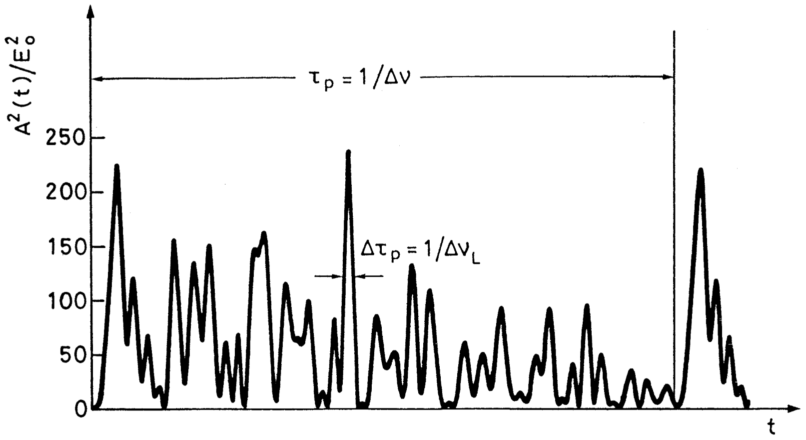

*FIG. 8.15.* Example of time behavior of the squared amplitude of the total electric field, $|A(t)|^2$, for the case of 31 oscillating modes, all with the same amplitude $E_0$ and with random phases.

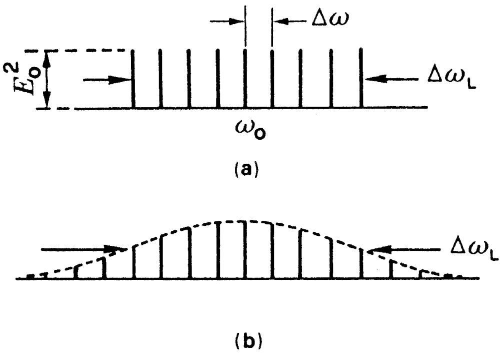

*FIG. 8.16.* Mode amplitudes (represented by vertical lines) versus frequency for a mode-locked laser. (a) Uniform amplitude. (b) Gaussian amplitude distribution over a bandwidth (FWHM) $\Delta\omega_L$.

## 8.6.1. Frequency-Domain Description

We will first describe mode-locking in the frequency domain and consider, as a first example, the case of $2n + 1$ longitudinal modes oscillating with the same amplitude $E_0$ (Fig. 8.16a). We will assume the phases $\varphi_l$ of the modes in the output beam to be locked according to the relation

$$\varphi_l - \varphi_{l-1} = \varphi \tag{8.6.1}$$

where $\varphi$ is a constant. The total electric field $E(t)$ of the e.m. wave, at any given point in the output beam, can be written as

$$E(t) = \sum_{-n}^{+n} {}_lE_0 \exp\{j[(\omega_0 + l\Delta\omega)t + l\varphi]\} \tag{8.6.2}$$

where $\omega_0$ is the frequency of the central mode, $\Delta\omega$ is the frequency difference between two consecutive modes and where the value of the phase for the central mode has, for simplicity, been taken to be zero. According to Eq. (8.6.2), the total electric field of the wave can be written as

$$E(t) = A(t)\exp(j\omega_0 t) \tag{8.6.3}$$

where

$$A(t) = \sum_{-n}^{+n} {}_lE_0 \exp[jl(\Delta\omega t + \varphi)] \tag{8.6.4}$$

Equation (8.6.3) shows that $E(t)$ can be represented in terms of a sinusoidal carrier wave, at the center-mode frequency $\omega_0$, whose amplitude $A(t)$ is time dependent. To calculate the time behavior of $A(t)$, we now change to a new time reference $t'$ such that $\Delta\omega t' = \Delta\omega t + \varphi$. In terms of the new variable $t'$, Eq. (8.6.4) transforms to

$$A(t') = \sum_{-n}^{+n} {}_lE_0 \exp jl(\Delta\omega\, t') \tag{8.6.5}$$

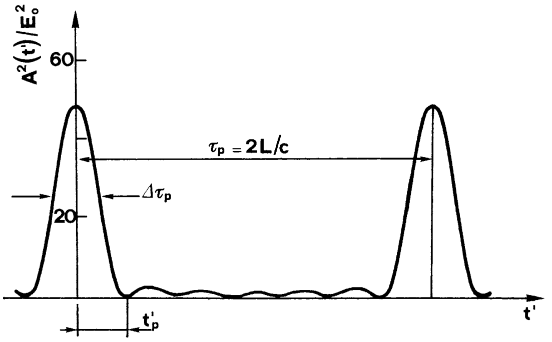

*FIG. 8.17.* Time behavior of the squared amplitude of the electric field for the case of seven oscillating modes with locked phases and equal amplitudes, $E_0$.

and the sum appearing in the right-hand side can be easily recognized as a geometric progression with a ratio $\exp j(\Delta\omega t')$ between consecutive terms. Summation of this progression can then be easily performed and we obtain

$$A(t') = E_0 \frac{\sin[(2n+1)\,\Delta\omega\, t'/2]}{\sin[\Delta\omega\, t'/2]} \tag{8.6.6}$$

To help understanding the physical significance of this expression, Fig. 8.17 shows the quantity $A^2(t')/E_0^2$, $A^2(t')$ being proportional to the beam intensity, versus time $t'$, for $2n+1 = 7$ oscillating modes. It is seen that, as a result of the phase-locking condition Eq. (8.6.1), the oscillating modes interfere so as to produce a train of evenly spaced light pulses. The pulse maxima occur at those times for which the denominator of Eq. (8.6.6) vanishes. In the new time reference $t'$, the first maximum occurs for $t' = 0$. Note that, at this time, the numerator of Eq. (8.6.6) also vanishes and, upon making the approximation $\sin\alpha \cong \alpha$, which holds for small values of $\alpha$, we readily see from Eq. (8.6.6) that $A^2(0) = (2n+1)^2 E_0^2$. The next pulse will occur when the denominator of Eq. (8.6.6) again vanishes and this will happen at a time $t'$ such that $(\Delta\omega t'/2) = \pi$. Two successive pulses are therefore separated by a time

$$\tau_p = 2\pi/\Delta\omega = 1/\Delta\nu \tag{8.6.7}$$

where $\Delta\nu$ is the frequency separation between two consecutive oscillating modes. For $t' > 0$, the first zero for $A^2(t')$ in Fig. 8.17 occurs when the numerator of Eq. (8.6.6) again vanishes. This occurs at a time $t'_p$ such that $\left[(2n+1)\,\Delta\omega t'_p/2\right] = \pi$. Since the width $\Delta\tau_p$ (FWHM) of $A^2(t')$, i.e. of each laser pulse, is approximately equal to $t'_p$, we thus have

$$\Delta\tau_p \cong 2\pi/(2n+1)\Delta\omega = 1/\Delta\nu_L \tag{8.6.8}$$

where $\Delta\nu_L = (2n+1)\Delta\omega/2\pi$ is the total oscillating bandwidth (see Fig. 8.16a).

The mode-locking behavior of Fig. 8.17 can be readily understood if we represent the field components of Eq. (8.6.5) by vectors in the complex plane. The $l$-th amplitude component would thus correspond to a complex vector of amplitude $E_0$ and rotating at the angular velocity $l\Delta\omega$. At time $t' = 0$, all these vectors are seen from Eq. (8.6.5) to have zero phase

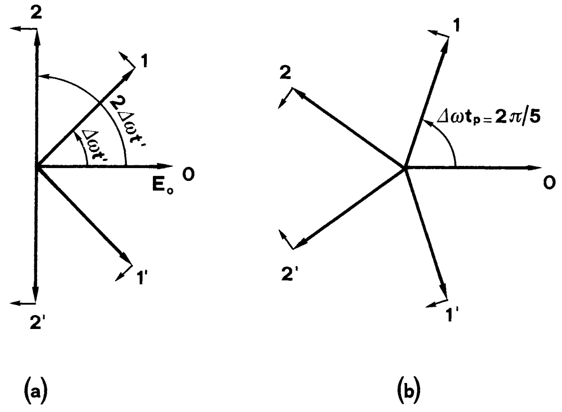

*FIG. 8.18.* Representation of the cavity mode amplitudes in the complex plane for the case of five modes. Figure (a) show the situation at a general time $t' > 0$, while FIG. (b) depicts the time instant at which the sum of the five mode amplitudes is zero.

and, accordingly, they lie in the same direction in Fig. 8.18, which we assume to be the horizontal direction. The total field will, in this case, be $(2n + 1)E_0$. For $t' > 0$, the vector corresponding to the central mode remains fixed, the vectors of the modes with $l > 0$ i.e., with $\omega > \omega_0$, will rotate in one direction (e.g., counterclockwise) while the vectors of the modes with $\omega < \omega_0$ will rotate in the opposite (clockwise) sense. Thus, for the case of e.g., five modes, the situation at some later time $t'$, will be as indicated in Fig. 8.18a. If now the time $t'$ is such that mode 1 has made a $2\pi$ rotation (which occurs when $\Delta\omega t' = 2\pi$), mode −1 will also have rotated (clockwise) by $2\pi$, while modes 2 and −2 will have rotated by $4\pi$. All these vectors will therefore be aligned again with that at frequency $\omega_0$, and the total field amplitude will again be $(2n + 1)E_0$. Thus the time interval $\tau_p$ between two consecutive pulses must be such that $\Delta\omega\tau_p = 2\pi$, as indeed shown by Eq. (8.6.7). Note that, in this picture, the time instant $t'_p$ at which $A(t')$ vanishes (see Fig. 8.17) correspond to the situation where all vectors are evenly spaced around the $2\pi$ angle (Fig. 8.18b). To achieve this condition, mode 1 must have made a rotation of only $2\pi/5$, or, more generally for $(2n + 1)$ modes, of $2\pi/(2n + 1)$. The time duration $t'_p$ and hence the pulse duration $\Delta\tau_p$ thus turn out to be given by Eq. (8.6.8).

Before proceeding further it is worth summarizing and commenting on the main results that have been obtained so far. We have found that, under the mode locking condition given by Eq. (8.6.1), the output beam consists of a train of mode-locked pulses, the duration of each pulse, $\Delta\tau_p$, being about equal to the inverse of the oscillating bandwidth $\Delta\nu_L$. Again this result comes about again from a general property of a Fourier series. Now, since $\Delta\nu_L$ can be of the order of the width of the gain line $\Delta\nu_0$, very short pulses (down to a few picoseconds) can be expected to result from mode-locking of solid-state or semiconductor lasers. For dye or tunable solid-state lasers, the gain linewidth can be at least a factor 100 times larger, so that very much shorter pulsewidths are possible and indeed have been obtained (e.g., ~25 fs for Rhodamine 6G dye laser and ~7 fs for Ti:sapphire laser). In the case of gas lasers, on the other hand, the gain linewidth is much narrower (up to a few GHz) and relatively long pulses are generated (down to ~100 ps). Note also that the peak power of the pulse is proportional to $(2n + 1)^2E_0^2$, while for modes with random phases the average power is simply the sum of powers in the modes and hence is proportional to $(2n + 1)E_0^2$. Therefore, for the same number the $4f$ states into sublevels which are typically separated by $\sim 10{,}000\,\mathrm{cm}^{-1}$. Spin-orbit coupling then splits each term into manifolds typically separated by $\sim 3{,}000\,\mathrm{cm}^{-1}$. Crystal field interaction produces the weakest perturbation (weakened by the screening effect of the $5s^2$ and $5p^6$ orbitals), thus further splitting each sub-level into a manifold with energy separation typically of $200\,\mathrm{cm}^{-1}$. All relevant absorption and emission features are due to transitions between these $4f$ states ($4f$–$4f$ transitions). Electric dipole transitions within the $4f$ shell are parity-forbidden and it needs a mixture of wavefunctions with opposite parity, brought about by the crystal-field interaction, to create non-zero, although still weak, transition probabilities. Thus, one generally finds long (hundreds of $\mu$s) radiative lifetimes. Furthermore, due to the screening from the $5s^2$ and $5p^6$ orbitals, electron-phonon coupling turns out to be very weak. One thus has sharp transition lines and weak non-radiative decay channels for low ion-doping (ion-ion interaction can lead to non-radiative decay at high RE ion concentrations, see Fig. 2.13). From the above considerations one expects large values of the overall lifetime, $\tau$, and of the product $\sigma\tau$, where $\sigma$ is the peak cross-section. This implies a low threshold pump power for laser action since, e.g. for a four level laser, the threshold pump rate is proportional to $1/\sigma\tau$ [see Eq. (7.3.3)].

The electronic configurations for those transition metals, of interest for laser action, are also shown in Table 9.1. Note that the electronic configuration of the most important active species, i.e., Cr, is given by $(\mathrm{Ar})3d^5 4s^1$, while those of Ti, Co, and Ni can be written in the general form $(\mathrm{Ar})3d^N 4s^2$ (with $N = 2$ for Ti, 7 for Co and 8 for Ni). When in an ionic crystal, the $4s^1$ electron and two $3d$ electrons of Cr are used for the ionic binding and Cr is present as a triply ionized ion with 3 electrons left in the $3d$ shell. For Titanium, the two $4s$ electrons and one $3d$ electron are used for the ionic binding and Ti is present as triply ionized ion with only one electron left in the $3d$ shell. For both Co and Ni the two $4s$ electrons are used for the binding and these elements are present as doubly ionized ions. In all cases, the remaining electrons in the $3d$ orbital can arrange themselves in a large number of states (e.g. 24 for $\mathrm{Cr}^{3+}$) and all the absorption and emission features of transition metal ions arise from $3d - 3d$ transitions. Lacking the screening which occurs for RE ions, the $3d$ states interact strongly with the crystal field of the host and, as we shall see later, this is the fundamental reason for the vibronic character, leading to wide absorption and emission bands, for most of the corresponding transitions. Again electric dipole transitions within the $3d$ shell are parity forbidden but, due to the stronger crystal field compared to the RE case, the $3d-3d$ transitions are more allowed and thus the lifetimes are significantly shorter (a few $\mu$s) than those of the $4f - 4f$ transitions of RE ions. Compared to e.g. Nd:YAG, the transition cross sections are somewhat smaller so that the product $\sigma\tau$ is now typically one order of magnitude smaller.

As a conclusion to this section, it is worth noting that ions belonging to the actinide series, notably $\mathrm{U}^{3+}$, were also used in the early days of laser development (actually the $\mathrm{U}^{3+}$ laser was the second solid state laser to be developed, i.e. immediately after the ruby laser). These ions are essentially no longer used but they deserve a mention here for historical reasons.

### 9.2.1. The Ruby Laser

This type of laser was the first to be made to operate (T. H. Maiman, June 1960(2,3)) and still continues to be used in some applications.(4) As a naturally occurring precious stone, ruby

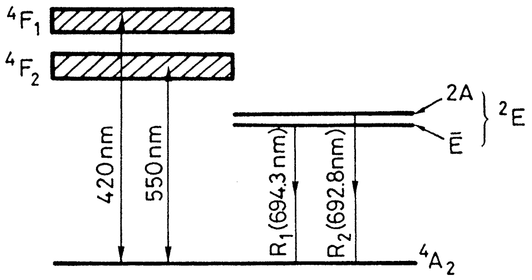

*FIG. 9.1.* Simplified energy levels of ruby.

has been known for at least 2,000 years. It consists of a natural crystal of $Al_2O_3$ (corundum) in which some of the $Al^{3+}$ ions happen to have been replaced by $Cr^{3+}$ ions. For the laser material, one uses artificial crystals obtained by crystal growth from a molten mixture of $Al_2O_3$ with a small percentage of $Cr_2O_3$ (0.05% by weight).(4) Without the addition of $Cr_2O_3$, the crystal that forms is colorless and it is known as sapphire. Due to the strong green and violet absorption bands of $Cr^{3+}$ ions, it only needs the small addition of $Cr_2O_3$ to give the crystal a pink color (pink ruby). In the case of gem stones, the $Cr^{3+}$ concentration is about an order of magnitude larger, giving them a strong red color (red ruby).

The energy levels of ruby are due to one of the three electrons of the $Cr^{3+}$ ion, in the $3d$ inner shell, under the influence of the octahedral field at the Al site in the $Al_2O_3$ lattice. The corresponding levels, of interest for laser action, are shown in Fig. 9.1. The notation used to label the levels is derived from group theory and is not discussed at any length here. We merely limit ourselves to pointing out that the superscript to the left of each letter indicates the multiplicity of the state while the letter indicates the particular rotational symmetry of the state. Thus, as an example, the ground $^4A_2$ state has a multiplicity $(2S + 1) = 4$, i.e., $S = 3/2$ where $S$ is the total spin quantum number of the three $3d$ electrons. This means that the spins of these electrons are all parallel in this case.

Ruby has two main pump bands $^4F_1$ and $^4F_2$ and the peaks of the transitions from the ground $^4A_2$ level to these bands occur at the wavelengths of 550 nm (green) and 420 nm (violet), respectively (see also Fig. 6.7). The two bands are connected by a very fast (ps) nonradiative decay to both $2\bar{A}$ and $\bar{E}$ states, which together form the $^2E$ state. The $2\bar{A}$ and $\bar{E}$ states are themselves connected to each other by a very fast nonradiative decay, which leads to a fast thermalization of their populations, thus resulting in the $\bar{E}$ level being the most heavily populated. Since the total spin of the $^2E$ state is 1/2, the $^2E \rightarrow {}^4A_2$ transition is spin-forbidden. The relaxation time of both $2\bar{A}$ and $\bar{E}$ levels to the ground state is thus very long ($\tau \cong 3$ ms), actually one of the longest among all solid-state laser materials.

From the discussion above it is now apparent that the level $\bar{E}$ accumulates the largest fraction of the pump energy, and is thus a good candidate as the upper laser level. In fact, laser action usually occurs on the $\bar{E} \rightarrow {}^4A_2$ transition ($R_1$ line) at the wavelength $\lambda_1 = 694.3$ nm (red). It should be noted, however, that the frequency separation between $2\bar{A}$ and $\bar{E}$ levels ($\sim$29 cm$^{-1}$) is small compared to $kT/h$ ($\sim$209 cm$^{-1}$ at $T = 300$ K) so that the $2\bar{A}$ population is comparable to, although slightly smaller than, the $\bar{E}$ level population. It then follows that it is also possible to obtain laser action on the $2\bar{A} \rightarrow {}^4A_2$ transition ($R_2$ line, $\lambda_1 = 692.8$ nm). It is anyhow apparent that ruby operates as a three-level laser (actually, together with Er lasers, it represents the most notewhorty example of a three-level laser). As already discussed in

*TABLE 9.2.* Optical and spectroscopic parameters of ruby for room temperature operation

| Property | Values and units |
|---|---|
| $Cr_2O_3$ doping | 0.05 wt. % |
| $Cr^{3+}$ concentration | $1.58 \times 10^{19}$ ions/cm$^3$ |
| Output wavelengths | 694.3 nm ($R_1$ line) |
| | 692.9 nm ($R_2$ line) |
| Upper laser level lifetime | 3 ms |
| Linewidth of $R_1$ laser transition | 11 cm$^{-1}$ |
| Stim. emission cross-section $\sigma_e$ | $2.5 \times 10^{-20}$ cm$^2$ |
| Absorption cross section $\sigma_a$ | $1.22 \times 10^{-20}$ cm$^2$ |
| Refractive index ($\lambda = 694.3$ nm) | $n = 1.763$ ($E \perp c$) |
| | $n = 1.755$ ($E \| c$) |

connection with Fig. 2.10, the $R_1$ transition is, at room temperature, predominantly homogeneously broadened, the broadening arising from the interaction of the $Cr^{3+}$ ions with lattice phonons. The width of the transition (FWHM) is $\Delta\nu_0 \cong 11\,\text{cm}^{-1}$ (330 GHz) at $T = 300$ K. As a summary, Table 9.2 shows some relevant optical and spectroscopic parameters of ruby at room temperature.

Ruby lasers are usually operated in a pulsed regime. For this, the pump configuration of Fig. 6.1 using a medium-pressure (~500 Torr) xenon flashtube, is generally utilized. Typical rod diameters range between 5 and 10 mm with a length between 5 and 20 cm. It should be noted that a helical flashtube surrounding the active rod was used in the earliest ruby lasers. Since this laser operates on a three-level scheme, the threshold pump energy is typically an order of magnitude higher than that of other solid-state lasers operating with four level schemes (e.g. Neodimium lasers). Due to the long upper state lifetime, ruby lasers lend themselves readily to $Q$-switched operation and, due to the relatively broad laser linewidth, they can also produce short pulses (~5 ÷ 10 ps) in mode-locked operation. Both active and passive methods can be used for $Q$-switching and mode-locking. When slow saturable absorbers are used for $Q$-switching, the laser tends to operate on a single transverse and longitudinal mode due to the mode selecting mechanism discussed in Sect. 8.4.2.4. With fast saturable absorbers (usually solutions of cyanine dyes), simultaneous $Q$-switched and mode-locked operation occurs (see Fig. 8.28a). Peak powers of a few tens of MW, for $Q$-switching, and a few GW, when also mode-locked, are typical. Since the gain of the $R_2$ line is somewhat smaller than for the $R_1$ line, laser action on the $R_2$ line can be selected by using, for instance, the dispersive system of Fig. 7.16b. Ruby lasers can also run cw, transversely pumped by a high-pressure mercury lamp or longitudinally pumped by an Ar ion laser.

Ruby lasers, once very popular, are now less widely used since, on account of their higher threshold, they have been superseded by competitors, such as Nd:YAG or Nd:glass lasers. In fact, ruby lasers were extensively used in the past for the first mass production of military rangefinders, an application in which this laser is now completely replaced by other solid-state lasers (Nd:YAG, Nd:glass, Yb:Er:glass). Ruby lasers are, however, still sometimes used for a number of scientific and technical applications where their shorter wavelength, compared to e.g., Nd:YAG, represents an important advantage. This is for instance the case by electrical means, i.e., by passing a sufficiently large current (which may be continuous, at radiofrequency or pulsed) through the gas. The principal pumping mechanisms occurring in gas lasers have been discussed in Sect. 6.4. It should also be pointed out that some gas lasers can in addition be pumped by mechanisms other than electrical pumping. In particular, we mention pumping by gas-dynamic expansion, chemical pumping, and optical pumping by means of another laser (particularly used for far-infrared lasers).

Once a given species is in its excited state, it can decay to lower states, including the ground state, by four different processes: (1) Collisions between an electron and the excited species, in which the electron takes up the excitation energy as kinetic energy (super-elastic collision); (2) near-resonant collisions between the excited species and the same or a different species in the ground state; (3) collisions with the walls of the container; (4) spontaneous emission. Regarding this last case, one should always take into account the possibility of radiation trapping, particularly for the usually very strong UV or VUV transitions. This process slows down the effective rate of spontaneous emission (see Sect. 2.9.1).

For a given discharge current, these various processes of excitation and de-excitation lead eventually to some equilibrium distribution of population among the energy levels. Thus it can be seen that, due to the many processes involved, the production of a population inversion in a gas is a more complicated matter that, e.g., in a solid-state laser. In general we can say that a population inversion between two given levels will occur when either (or both) of the following circumstances occur: (1) The excitation rate is greater for the upper laser level (level 2) than for the lower laser level (level 1); (2) the decay of level 2 is slower than that of level 1. In this regard we recall that a necessary condition for cw operation is that the rate of the $2 \rightarrow 1$ transition be smaller than the decay rate of level 1 [see Eq. (7.3.1)]. If this condition is not satisfied, however, laser action can still occur under pulsed operation provided the condition (1), above, is fulfilled (self-terminating lasers).

### 10.2.1. Neutral Atom Lasers

These lasers make use of neutral atoms in either gaseous or vapor form. Neutral atom gas lasers constitute a large class of lasers and include in particular most of the noble gases. All these lasers oscillate in the infrared (1–10 $\mu$m), apart from the notable exceptions of green and red emission from the He-Ne laser. Metal vapor lasers also constitute a large class of lasers, including, for example, Pb, Cu, Au, Ca, Sr, and Mn. These lasers generally oscillate in the visible, the most important example being the copper vapor laser oscillating on its green (510 nm) and yellow (578.2 nm) transitions. All metal vapor lasers are self-terminating and therefore operate in a pulsed regime.

#### 10.2.1.1. Helium-Neon Lasers(1,2)

The He-Ne laser is certainly the most important of the noble gas lasers. Laser action is obtained from transitions of the neon atom, while helium is added to the gas mixture to greatly facilitate the pumping process. The laser oscillates on many wavelengths, by far the most popular being at $\lambda = 633$ nm (red). Other wavelengths include the green (543 nm) and the infrared ones at $\lambda = 1.15\ \mu$m and $\lambda = 3.39\ \mu$m. The helium-neon laser oscillating on its $\lambda = 1.15\ \mu$m transition was the first gas laser and the first cw laser ever to be operated.(3)

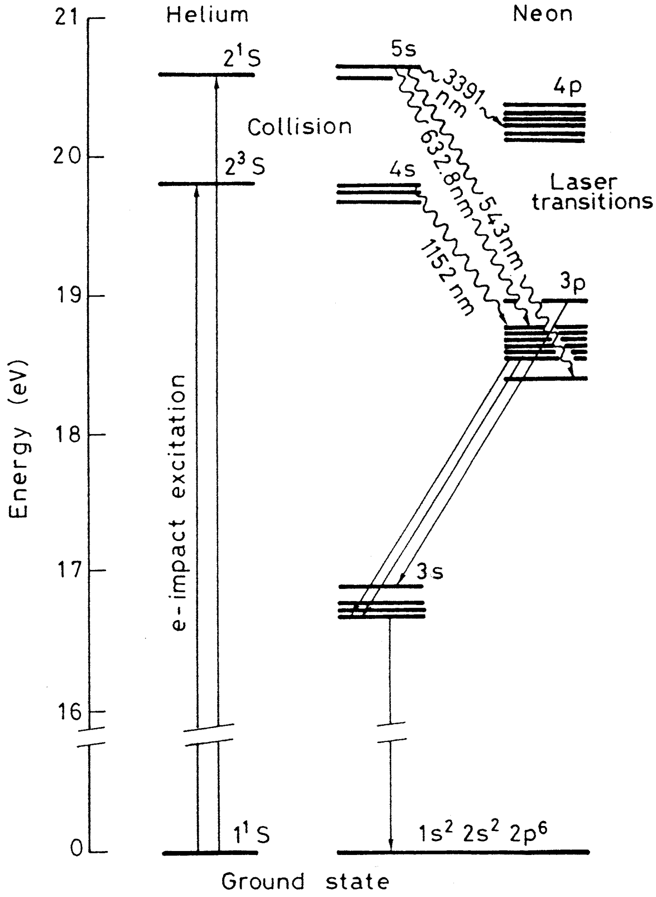

*FIG. 10.1.* Relevant energy levels of the He-Ne laser.

The energy levels of the He-Ne system that are relevant for laser action are shown in Fig. 10.1. The level notation for He is according to Russell-Saunders coupling with the principal quantum number of the given level also indicated as the first number. Thus the $1^1S$ state corresponds to the case where the two electrons of He are both in the $1s$ state with opposite spins. The $2^3S$ and $2^1S$ states correspond to a situation where one of the two electrons is raised to the $2s$ state with its spin either in the same or opposite direction to that of the other electron, respectively. Neon, on the other hand, has an atomic number of ten and a number of ways have been used, such as Paschen or Racah notations, to indicate its energy levels. For simplicity, however, we will limit ourselves here to simply indicating the electron configuration corresponding to each level. So, the ground state is indicated by the $1s^2 2s^2 2p^6$ configuration, while the excited states shown in the figure correspond to the situation where one $2p$ electron is raised to excited $s$ states ($3s$, $4s$, and $5s$) or excited $p$ states ($3p$ and $4p$). One should also notice that, due to the interaction with the remaining five electrons in the $2p$ orbitals, these $s$ and $p$ states are split into four and ten sub-levels, respectively.

It is apparent form Fig. 10.1 that the levels $2^3S$ and $2^1S$ of He are nearly resonant with the $4s$ and $5s$ states of Ne. Since the $2^3S$ and $2^1S$ levels are metastable ($S \to S$ transitions are electric dipole forbidden and, furthermore, the $2^3S \to 1^1S$ transition is also forbidden due to the change of multiplicity, i.e., spin-forbidden), it is found that He atoms in these states prove very efficient at pumping the Ne 4*s* and 5*s* levels by resonant energy transfer. It has been confirmed that this process is the dominant one for producing population inversion in the He-Ne laser, although direct electron-Ne collisions also contribute to the pumping. Since significant population can be built-up in the Ne 4*s* and 5*s* states, they prove suitable candidates as upper levels for laser transitions. Taking account of the selection rules, we see that the possible transitions are those to the *p* states. In addition, the decay time of the *s* states ($\tau_s \cong 100\,\text{ns}$) is an order of magnitude longer than the decay time of the *p* states ($\tau_p \cong 10\,\text{ns}$). So, the condition Eq. (7.3.1) for operation as a cw laser is satisfied. Finally, it should be noted that the electron-impact excitation rates from the ground state to the 3*p* and 4*p* levels are much smaller than the corresponding rates to the 4*s* and 5*s* levels, due to smaller values of cross section involved. As we shall see, however, direct excitation to the 3*p* and 4*p* levels plays an important role in determining the laser performances.

The above discussion indicates that one can expect laser action in Ne to occur between 5*s* or 4*s* levels, as upper levels, and 3*p* and 4*p* levels, as lower levels. Some of the most important laser transitions arising from these levels are also indicated in Fig. 10.1. For transitions differing widely in wavelength ($\Delta\lambda > 0.2\ \lambda$), the actual oscillating transition depends on the wavelength at which the peak reflectivity of the multilayer dielectric mirror is centered (see Fig. 4.9). The laser transitions are predominantly broadened by the Doppler effect. For instance, for the red He-Ne laser transition ($\lambda = 633\,\text{nm}$ in vacuum and $\lambda = 632.8\,\text{nm}$ in air), Doppler broadening leads to a linewidth of ~1.5 GHz (see also example 2.6). By comparison, natural broadening, according to Eq. (2.5.13), can be estimated to be $\Delta\nu_{nat} = 1/2\pi\tau \cong 19\,\text{MHz}$, where $\tau^{-1} = \tau_s^{-1} + \tau_p^{-1}$ and $\tau_s$, $\tau_p$ are the lifetimes of the *s* and *p* states, respectively. Collision broadening contribute even less than natural broadening [e.g., for pure Ne, $\Delta\nu_c \cong 0.6\,\text{MHz}$ at the pressure of $p \cong 0.5\,\text{torr}$, see example 2.2]. Some spectroscopic properties of the 633 nm laser transition are summarized in Table 10.1.

The basic design of a He-Ne laser is shown in Fig. 10.2. The discharge is produced between a ring anode and a large tubular cathode, which can thus withstand the collisions from positive ions. The discharge is confined to a capillary for most of the tube length and high inversion is only achieved in the region where the capillary is present. The large volume of gas available in the tube surrounding the capillary acts as a reservoir to replenish the He-Ne mixture in the capillary. When a polarized output is needed, a Brewster angle plate is also inserted inside the laser tube. The laser mirrors are directly sealed to the two tube ends. The most commonly used resonator configuration is nearly hemispherical since this is easy

*TABLE 10.1.* Spectroscopic properties of laser transitions and gas-mixture composition in some relevant atomic and ionic gas lasers

| Laser type | He-Ne | Copper Vapor | Argon Ion | He-Cd |
|---|---|---|---|---|
| Laser wavelength [nm] | 633 | 510.5 | 514.5 | 441.6 |
| Cross-section [$10^{-14}\,\text{cm}^2$] | 30 | 9 | 25 | 9 |
| Upper-state lifetime [ns] | 150 | 500 | 6 | 700 |
| Lower-state lifetime [ns] | 10 | $\approx 10^4$ | $\sim 1$ | 1 |
| Transition Linewidth [GHz] | 1.5 | 2.5 | 3.5 | 1 |
| Partial pressures of gas mixture [torr] | 4 (He)<br>0.8 (Ne) | 40 (He)<br>0.1–1 (Cu) | 0.1 (Ar) | 10 (He)<br>0.1 (Cd) |

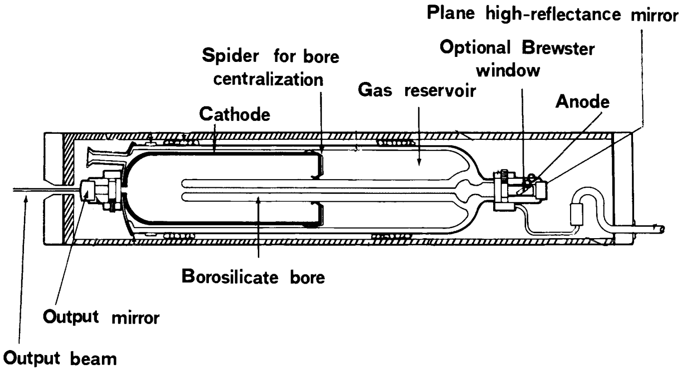

*FIG. 10.2.* Internal design of a hard-sealed helium-neon laser (courtesy of Melles-Griot).

to adjust, is very stable against misalignment, and readily gives $TEM_{00}$ mode operation. The only disadvantage of this configuration is that it does not fully utilize the volume of the plasma discharge since the mode spot size is much smaller at the plane mirror than at the concave mirror. If, however, the left-hand mirror in Fig. 10.2 is chosen to be the plane mirror, the region of smaller spot size for the near-hemispherical $TEM_{00}$ mode will be outside the capillary, i.e., in a region of low inversion.

One of the most characteristic features of the He-Ne laser is that the output power does not increase monotonically with the discharge current but reaches a maximum and thereafter decreases. For this reason, commercially available He-Ne lasers are provided with a power supply designed to give only the optimum current. The fact that there is an optimum value of current, i.e., of current density $J$ within the capillary is because (at least for the 633 nm and 3.39 $\mu$m transitions), at high current densities, de-excitation of the He ($2^3S$ and $2^1S$) metastable states takes place not only by collision at the walls but also by super-elastic collision processes such as

$$\mathrm{He}(2^1S) + e \rightarrow \mathrm{He}(1^1S) + e \qquad (10.2.1)$$

Since the rate of this process is proportional to the electron density $N_e$, and hence to $J$, the overall rate of excitation can be written as $k_2 + k_3J$. In this expression $k_2$ is a constant that represents de-excitation due to collisions with the walls and $k_3J$, where $k_3$ is also a constant, represents the superelastic collision rate of process Eq. (10.2.1). The excitation rate, on the other hand, can be expressed as $k_1J$, where $k_1$ is again a constant. Under steady state conditions we can then write that $N_tk_1J = (k_2 + k_3J)N^*$, where $N_t$ is the ground-state He atom population and $N^*$ is the excited ($2^1S$) state population. The equilibrium $2^1S$ population is then given by

$$N^* = N_t\frac{k_1J}{k_2 + k_3J} \qquad (10.2.2)$$

which can be seen to saturate at high current densities. Since the steady-state population of the 5*s* state of Ne is established by near-resonant energy transfer from He $2^1S$ state, the population of the upper, 5*s*, laser level will also show a similar saturation behavior as $J$ increases (see Fig. 10.3). On the other hand, in the absence of laser action, the population of the lower relatively high efficiency ($\sim$1%). The latter is the result of both the high quantum efficiency of the copper laser ($\sim$55%, see Fig. 10.4) and the high electron-impact cross section for the $^2S \rightarrow {}^2P$ transition. Even higher output powers ($\sim$200 W) and higher efficiencies ($\sim$3%) have recently been obtained with Copper-HyBrID lasers.

Copper vapor lasers are used for some industrial applications (such as high-speed photography, resistor trimming, and more recently, micromachining) and as a pump for dye lasers. In particular, in high-speed flash photography, the short pulse (tens of ns) and high-repetition-rate (10–20 kHz) are exploited in stroboscopic illumination of various, rapidly moving, objects (e.g., a bullet in flight). A large facility based on copper-laser pumped dye lasers (using many copper lasers, each with average power up to 100 W) is currently in use in a pilot plant, in the United States, for $^{235}$U isotope separation.

### 10.2.2. Ion Lasers

In the case of an ionized atom, the scale of energy levels is expanded in comparison with neutral atoms. In this case, in fact, an electron in the outermost orbital(s) experiences the field due to the positive charge $Ze$ of the nucleus ($Z$ being the atomic number and $e$ the electronic charge) screened by the negative charge $(Z-2)e$ of the remaining electrons. Assuming, for simplicity, the screening to be complete, the net effective charge is then $2e$ rather than simply $e$ for the corresponding neutral atom. This expansion in energy scale means that ion lasers typically operate in the visible or ultraviolet regions. As in the case of neutral atom lasers, ion lasers can be divided into two categories: (1) Ion gas lasers, involving most of the noble gases, the most notable example being the $Ar^+$ laser, which we consider below, and the $Kr^+$ laser. Both lasers oscillate on many transitions, the most common being in the green and blue (514.5 nm and 488 nm) for the $Ar^+$ laser and in the red (647.1 nm) for the $Kr^+$ laser. (2) Metal-ion vapor lasers, involving many metals (Sn, Pb, Zn, Cd, and Se), the most notable example being the He-Cd laser, discussed below, and the He-Se laser.

#### 10.2.2.1. Argon Laser(5, 6)

A simplified scheme for the relevant energy levels in an argon laser is shown in Fig. 10.6. The $Ar^+$ ground state is obtained by removing one electron out of the six electrons of the, $3p$, outer shell of Ar. The excited $4s$ and $4p$ states are then obtained by promoting one of the remaining $3p^5$ electrons to the $4s$ or $4p$ state, respectively. As a consequence of the interaction with the other $3p^4$ electrons, both the $4s$ and $4p$ levels, indicated as single levels in Fig. 10.6, actually consist of many sublevels.

Excitation of the Ar ion to its excited states occurs by a two-step process involving collisions with two distinct electrons. The first collision, in fact, ionizes Ar i.e., raises it to the $Ar^+$ ground state, while the second collision excites the Ar ion. Since the lifetime of the $4p$ level ($\sim 10^{-8}$ s, set by the $4p \rightarrow 4s$ radiative transition) is about ten times longer than the radiative lifetime of the $4s \rightarrow 3p^5$ transition, excited Ar ions accumulate predominantly in the $4p$ level. This means that the $4p$ level can be used as the upper laser level, for the $4p \rightarrow 4s$ laser transition, and that, according to Eq. (7.3.1), cw laser action can be achieved. It should be noted that excitation of the Ar ion can lead to ions in the $4p$ state by three distinct processes

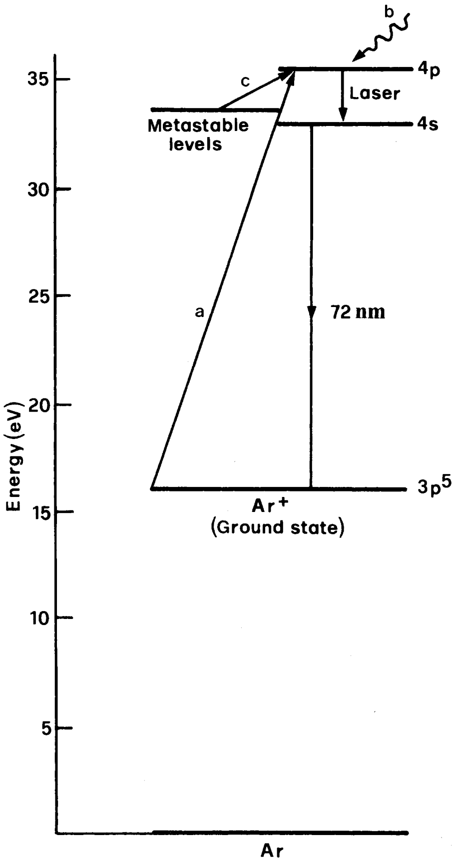

*FIG. 10.6.* Energy levels of Ar relevant for laser action.

(see Fig. 10.6): (a) direct excitation to the $4p$ level starting from the $\mathrm{Ar}^+$ ground level; (b) excitation to higher-lying states followed by radiative decay to the $4p$ level; (c) excitation to metastable levels followed by a third collision leading to excitation to the $4p$ state. Considering, for simplicity, only processes (a) and (b) one can readily see that the pumping process to the upper state is expected to be proportional to the square of the discharge current density. In fact, since processes (a) and (b) involve a two-steps electron collision, the rate of upper state excitation, $(dN_2/dt)_p$, is expected to be of the form

$$(dN_2/dt)_p \propto N_e N_i \cong N_e^2 \tag{10.2.3}$$

where $N_e$ and $N_i$ are the electron and ion density in the plasma ($N_e \cong N_i$ in the positive column of the plasma). Since the electric field of the discharge is independent of the discharge current, the drift velocity, $v_{drift}$, will also be independent of the discharge current. From the standard equation $J = ev_{drift}N_e$, one then see that the electron density $N_e$ is proportional to the discharge current density, and, from Eq. (10.2.3) it follows that $(dN_2/dt)_p \propto J^2$. Laser pumping thus increases rapidly with current density and high current densities ($\sim 1\ \mathrm{kA/cm^2}$) are required if the inherently inefficient two-step processes, considered above, are to pump enough ions to the upper state. This may explain why the first operation of an $\mathrm{Ar}^+$ laser occurred some three-years after the first He-Ne laser.(7)

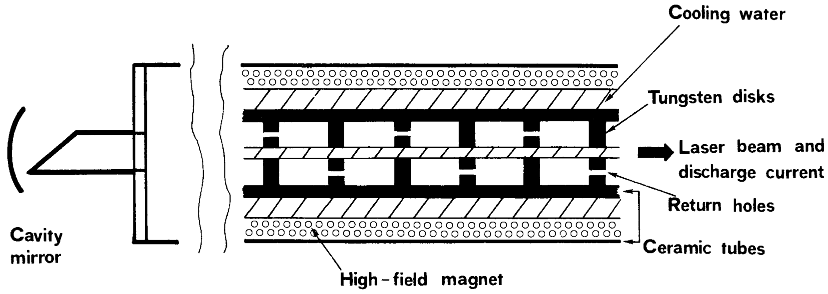

*FIG. 10.7.* Schematic diagram of a high-power water-cooled $Ar^+$ laser tube.

From the discussion above, one expects laser action in an Ar laser to occur on the $4p \rightarrow 4s$ transition. Since both the $4s$ and $4p$ levels actually consist of many sub-levels, the argon laser is found to oscillate on many lines, the most intense being in the green ($\lambda = 514.5\,\text{nm}$) and in the blue ($\lambda = 488\,\text{nm}$). From spectral measurements of the spontaneously emitted light it is found that the Doppler linewidth $\Delta\nu_0^*$, on e.g. the green transition, is about 3500 MHz. From Eq. (2.5.18) it is seen that this implies an ion temperature of $T \cong 3000\,\text{K}$. The ions are therefore very hot, a result of ion acceleration by the electric field of the discharge. Some relevant spectroscopic properties of the Ar ion green laser transition are summarized in Table 10.1.

A schematic diagram of a high power ($>1\,\text{W}$) argon laser is shown in Fig. 10.7. One sees that both the plasma current and the laser beam are confined by metal disks (of tungsten) inserted in a larger bore tube of ceramic material (BeO). The use of this thermally conductive and resistant metal-ceramic combination is necessary to ensure a good thermal conductivity of the tube and, at the same time, to reduce the erosion problems that arise from the high ion temperature. The diameter of the central holes in the disks is kept small ($\sim 2\,\text{mm}$) to confine oscillation to a $TEM_{00}$ mode (long-radius of curvature mirrors are commonly used for the resonator) and to reduce the total current required. A problem that must be overcome in an argon laser is that which arises from the cataphoresis of the argon ions. Due to the high current density, in fact, a substantial migration of Ar ions occurs toward the cathode, where they are neutralized upon combining with electrons emitted by the cathode surface*. Thus an accumulation of neutral atoms tends to build up at this electrode with a corresponding reduction of the Ar pressure in the discharge capillary, below its optimum value. To overcome this problem, off-center holes are also made in the disk to provide return paths for the atoms, from cathode to anode, by diffusion. The holes are arranged in such a way that no current flows along this return path on account of the longer path lengths involved compared to that of the central path. The inner ceramic tube is water cooled to remove the large amount of heat that it is inevitably dissipated in the tube (some kW/m). Note also that a static magnetic field is also applied in the discharge region, parallel to the tube axis, by a solenoid. With this

---

* With reference to the discussion of Sect. 6.4.4., we recall that direct electron ion recombination cannot occur in the discharge volume since the process cannot satisfy, simultaneously, both energy and momentum conservation. Electron-ion recombination can therefore only occur in the presence of a third partner e.g., at the tube walls or at the cathode surface.

result is a continuous flow of metal vapor from the anode to the cathode. Therefore a sufficient supply of Cd ($\sim$1 g per 1,000 h) must be provided for long-term operation of the laser.

He-Cd lasers can give output powers of 50–100 mW, placing them in an intermediate position between red He-Ne lasers (a few mW) and $Ar^+$ lasers (a few W). Thus, He-Cd lasers are used for many applications where a blue or UV beam of moderate power is required (e.g., high-speed laser printers, holography, cell cytometry, fluorescence analysis of e.g., biological specimens).

### 10.2.3. Molecular Gas Lasers

These lasers exploit transitions between the energy levels of a molecule. Depending on the type of transition involved, molecular gas lasers belong to one of the three following categories: (1) Vibrational-rotational lasers. These lasers use transitions between vibrational levels of the same electronic state (the ground state) and the energy difference between the levels falls in the middle- to the far-infrared (2.5–300 $\mu$m). By far the most important example of this category is the $CO_2$ laser oscillating at either 10.6 or 9.6$\mu$m. Other noteworthy examples are the CO laser ($\lambda \cong 5\,\mu$m) and the HF chemical laser ($\lambda \cong 2.7$–$3.3\,\mu$m). (2) Vibronic lasers, which use transitions between vibrational levels of different electronics states: In this case the oscillation wavelength generally falls in the UV region. The most notable example of this category of laser is the nitrogen laser ($\lambda = 337$ nm). A special class of lasers, which can perhaps be included in the vibronic lasers, is the excimer laser. These lasers involve transitions between different electronic states of special molecules (excimers) with corresponding emission wavelengths generally in the UV. Excimer lasers, however, involve not only transitions between bound states (bound-bound transitions) but also, and actually more often, transitions between a bound upper state and a repulsive ground state (bound-free transitions). It is more appropriate therefore to treat these lasers as being in a category of their own. (3) Pure rotational lasers, which use transitions between different rotational levels of the same vibrational state (usually an excited vibrational level of the ground electronic state). The corresponding wavelength falls in the far infrared (25 $\mu$m to 1 mm). Since these pure rotational lasers are relatively less important than the other categories, we shall not discuss them further in the sections that follow. We therefore limit ourselves to pointing out here that laser action is more difficult to achieve in this type of laser since the relaxation between rotational levels is generally very fast. Therefore these lasers are usually pumped optically, using the output of another laser as the pump (commonly a $CO_2$ laser). Optical pumping excites the given molecule (e.g., $CH_3F$, $\lambda = 496\,\mu$m) to a rotational level belonging to some vibrational state above the ground level. Laser action then takes place between rotational levels of this upper vibrational state.

#### 10.2.3.1. The $CO_2$ Laser(8, 9)

The laser utilizes, as active medium, a suitable mixture of $CO_2$, $N_2$, and He. Oscillation takes place between two vibrational levels of the $CO_2$ molecule, while, as we shall see, the $N_2$ and He greatly improve the efficiency of laser action. The $CO_2$ laser is actually one of the most powerful lasers (output powers of more than 100 kW have been demonstrated from a $CO_2$ gas-dynamic laser) and one of the most efficient (15–20% slope efficiency).

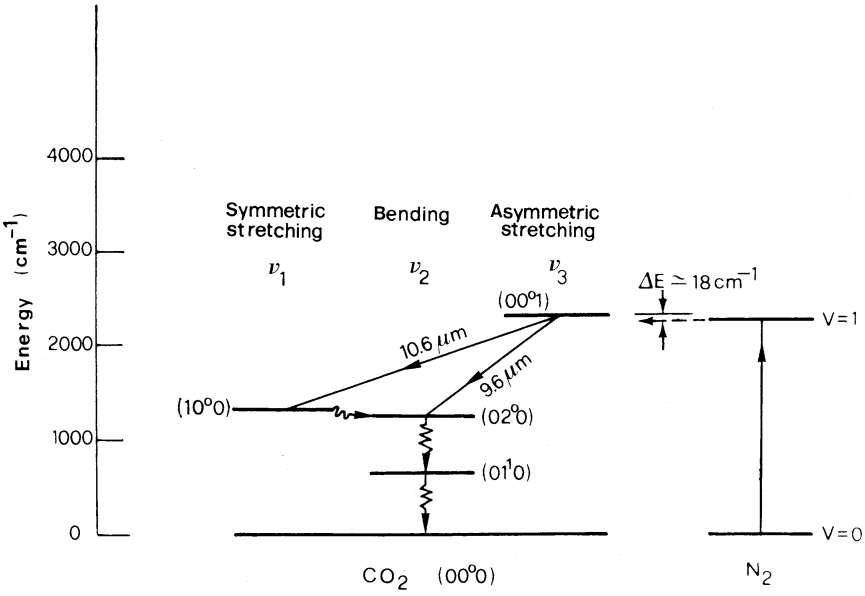

*FIG. 10.9.* The lowest vibrational levels of the ground electronic state of a $CO_2$ molecule and a $N_2$ molecule (for simplicity, the rotational levels are not shown).

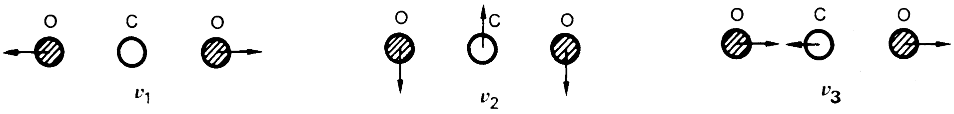

*FIG. 10.10.* The three fundamental modes of vibration for a $CO_2$ molecule: ($\nu_1$) symmetric stretching mode, ($\nu_2$) bending mode, ($\nu_3$) asymmetric stretching mode.

Figure 10.9 shows the relevant vibrational-energy levels for the electronic ground states of the $CO_2$ and $N_2$ molecules. $N_2$, being a diatomic molecule, has only one vibrational mode whose lowest two energy levels ($v = 0,\ v = 1$) are indicated in the figure. The energy levels for $CO_2$ are more complicated since $CO_2$ is a linear triatomic molecule. In this case, there are three nondegenerate modes of vibration (Fig. 10.10): (1) Symmetric stretching mode, (2) bending mode, and (3) asymmetric stretching mode. The oscillation behavior and the corresponding energy levels are therefore described by means of three quantum numbers $n_1$, $n_2$ and $n_3$, which give the number of quanta in each vibrational mode. This means that, apart from zero-point energy, the energy of the level is given by $E = n_1 h\nu_1 + n_2 h\nu_2 + n_3 h\nu_3$, where $\nu_1$, $\nu_2$, and $\nu_3$ are the resonance frequencies of the three modes. For example, the $01^10$ level* corresponds to an oscillation in which there is one vibrational quantum in mode 2. Since mode 2 has the smallest force constant of the three modes (the vibrational motion is transverse), it

---

* The superscript (which we will denote by $l$) on the bending quantum number arises from the fact that the bending vibration is, in this case, doubly degenerate. In fact, it can occur both in the plane of Fig. 10.10 and in the orthogonal plane. A bending vibration therefore consists of a suitable combination of these two vibrations and the superscript $l$ characterizes this combination; more precisely, $l\hbar$ gives the angular momentum of this vibration about the axis of the $CO_2$ molecule. For example, in the $02^00$ state ($l = 0$), the two degenerate vibrations combine in such a way to give an angular momentum $l\hbar = 0$.

follows that this level will have the lowest energy. Laser action takes place between the $00^01$ and $10^00$ levels ($\lambda \cong 10.6\,\mu$m) although it is also possible to obtain oscillation between $00^01$ and $02^00$ ($\lambda \cong 9.6\,\mu$m).

The pumping of the upper $00^01$ laser level is very efficiently achieved by two processes:

*Direct Electron Collisions*. The main direct collision to be considered is obviously as follows: $e + CO_2(000) \rightarrow e + CO_2(001)$. The electron collision cross section for this process is very large and is appreciably larger than those for excitation to both the 100 and 020 levels, probably because the 000 → 001 transition is optically allowed whereas, for instance, the 000 → 100 transition is not. Note also that direct electron impact can also lead to excitation of upper $(0, 0, n)$ vibrational levels of the $CO_2$ molecule. The $CO_2$ molecule, however, rapidly relaxes from these upper states to the (001) state by near resonant collisions of the type*

$$CO_2(0,0,n) + CO_2(0,0,0) \rightarrow CO_2(0,0,n-1) + CO_2(0,0,1) \qquad (10.2.5)$$

This process tends to degrade all excited molecules to the (0, 0, 1) state. Note that, since most molecules in a $CO_2$ laser mixture are in fact in the ground state, collision of an excited with an unexcited molecule constitutes the most likely collisional event.

*Resonant Energy Transfer from $N_2$ Molecule*. This process is also very efficient due to the small energy difference between the excited levels of the two molecules ($\Delta E = 18\,\text{cm}^{-1}$). In addition, the excitation of $N_2$ from the ground level to the $\upsilon = 1$ level is a very efficient process and the $\upsilon = 1$ level is metastable. The 1 → 0 transition is in fact electric-dipole forbidden since, by virtue of its symmetry, a N-N molecule cannot have a net electric dipole moment. Finally the higher vibrational levels of $N_2$ are also closely resonant ($\Delta E < kT$) with the corresponding $CO_2$ levels (up to $00^05$), and transitions between the excited levels, $00n$, and the 001 level of the $CO_2$ molecule occur rapidly through the process indicated in Eq. (10.2.5).

The next point to consider is the decay of both upper and lower laser levels. We note that, although the transitions $00^01 \rightarrow 10^00$, $00^01 \rightarrow 02^00$, $10^00 \rightarrow 01^00$, and $02^00 \rightarrow 01^00$ are optically allowed, the corresponding decay times $\tau_{sp}$ for spontaneous emission are very long (we recall that $\tau_{sp} \propto 1/\nu^3$). The decay of these various levels is therefore determined essentially by collisions. Accordingly, the decay time $\tau_s$ of the upper laser level can be obtained from a formula of the type

$$(1/\tau_s) = \Sigma a_i p_i \qquad (10.2.6)$$

where $p_i$ are the partial pressures and $a_i$ the rate constants that are corresponding to the gases in the discharge. Taking, for example, the case of a total pressure of 15 torr (in a 1:1:8 $CO_2 : N_2 :$ He partial pressure ratio) one finds that the upper level has a lifetime $\tau_s \cong 0.4$ ms. As far as the relaxation rate of the lower level is concerned, we begin by noting that the 100 → 020 transition is very fast and it occurs even in a isolated molecule. In fact the energy difference between the two levels is much smaller than $kT$. Furthermore, a coupling between the two states is present (Fermi resonance) because a bending vibration tends to induce a change of distance between the two oxygen atoms (i.e., induce a symmetric stretching). Levels

---

[*] Relaxation processes in which vibrational energy is given up as vibrational energy of another like or unlike molecule are usually referred to as *VV relaxations*.

## 9.4. SEMICONDUCTOR LASERS(31,32)

Semiconductor lasers represent one of the most important class of lasers in use today, not only because of the large variety of direct applications in which they are involved but also because they have found a widespread use as pumps for solid state lasers. These lasers will therefore be considered at some length here.

Semiconductor lasers require, for the active medium, a direct gap material and, accordingly, the normal elemental semiconductors (e.g., Si or Ge) cannot be used. The majority of semiconductor-laser materials are based on a combination of elements belonging to the third group of the periodic table (such as Al, Ga, In) with elements of the fifth group (such as N, P, As, Sb) (*III–V compounds*). Examples include the best known, GaAs, as well as some ternary (e.g. AlGaAs, InGaAs) and quaternary (e.g., InGaAsP) alloys. The cw laser emission wavelength of these III–V compounds generally ranges between 630–1,600 nm. Quite recently, however, very interesting InGaN semiconductor lasers, providing cw room-temperature emission in the blue ($\sim$410 nm), have been developed and promise to become the best candidates for semiconductor laser emission in the very important blue-green spectral region. Semiconductor laser materials are not limited to III–V compounds, however. For the blue-green end of the spectrum we note that there are wide-gap semiconductors using a combination between elements of the second group (such as Cd and Zn) and of the sixth group (S, Se) (*II–VI compounds*). For the other end of the e.m. spectrum, we mention semiconductors based on some *IV–VI compounds* such as Pb salts of S, Se, and Te, all oscillating in the mid-infrared (4 $\mu$m–29 $\mu$m). Due to the small band gap, these last lasers reuire cryogenic temperatures, however. In the same wavelength range, we thus mention the recent invention of the *quantum cascade* laser,(33) which promises efficient mid infrared sources without requiring cryogenic temperatures.

### 9.4.1. Principle of Semiconductor Laser Operation

The principle of operation of a semiconductor laser can be simply explained with the help of Fig. 9.18, where the semiconductor valence band, $V$, and conduction band, $C$, separated by the energy gap, $E_g$, are indicated. For simplicity, let us first assume that the semiconductor is held at $T = 0$ K. Then, for a non-degenerate semiconductor, the valence band will be completely filled with electrons while the conduction band will be completely empty (see Fig. 9.18a, in which the energy states belonging to the hatched area are completely filled by

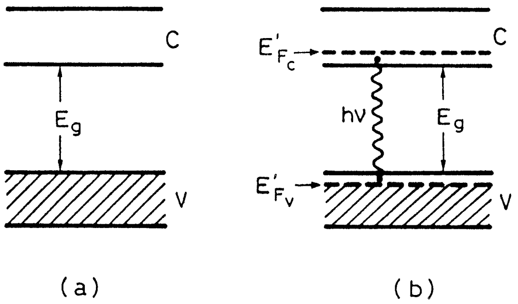

*FIG. 9.18.* Principle of operation of a semiconductor laser.

electrons). Suppose now that some electrons are raised from the valence band to the conduction band by a suitable pumping mechanism. After a very short time ($\sim 1$ ps), the electrons in the conduction band drop to the lowest unoccupied levels of this band, and, meanwhile, any electron near the top of the valence band also drops to the lowest unoccupied levels of this band, thus leaving holes at the top of the valence band (Fig. 9.18b). This situation can be described by introducing the quasi-Fermi levels, $E'_{F_c}$, for the conduction band and, $E'_{F_v}$, for the valence band (see Sect. 3.2.3.). At $T = 0$ K they define, for each band, the energy below which states are fully occupied by electrons and above which, states are empty. Light emission can now occur when an electron, of the conduction band, falls back to the valence band recombining with a hole. This, so-called recombination-radiation process, is the process by which radiation is emitted in light emitting diodes (LED). Given the appropriate conditions, however, a process of stimulated emission of this recombination radiation, thus leading to laser action, can occur. It was shown in Sect. 3.2.5. that the condition for a photon to be amplified rather than absorbed by the semiconductor is simply given by [see Eq. (3.2.39)]

$$E_g \leq h\nu \leq E'_{F_c} - E'_{F_v} \tag{9.4.1}$$

In the simple case where $T = 0$ K, this condition can be readily understood from Fig. 9.18b, since the non-hatched area in the valence band corresponds to states which are empty, and a conduction band electron can only fall into an empty state of the valence band. However, the detailed treatment of Sect. 3.2.5. shows that condition of Eq. (9.4.1) in fact holds for any temperature and simply means that, for the range of transition energy $h\nu$ defined by Eq. (9.4.1), the gain arising from stimulated emission exceeds the absorption. To achieve condition of Eq. (9.4.1) one must, of course, have $E'_{F_c} - E'_{F_v} \geq E_g$. It is important at this point to realize that the values of both $E'_{F_c}$ and $E'_{F_v}$ depend on the intensity of the pumping process, i.e. on the number density, $N$, of electrons raised to the conduction band (see Fig. 3.15). Actually $E'_{F_c} = E'_{F_c}(N)$ increases while $E'_{F_v} = E'_{F_v}(N)$ decreases as $N$ is increased. Thus, to obtain $E'_{F_c} - E'_{F_v} \geq E_g$ i.e., to have gain exceeding absorption losses, the electron density $N$ must exceed some critical value established by the condition

$$E'_{F_c}(N) - E'_{F_v}(N) = E_g \tag{9.4.2}$$

The value of the injected carrier density which satisfies Eq. (9.4.2) is referred to as the carrier density at transparency*, $N_{tr}$. If now the injected carrier density is larger than $N_{tr}$, the semiconductor will exhibit a net gain and, if this active medium is placed in a suitable cavity, laser action can occur if this net gain is sufficient to overcome the cavity losses. Thus, to obtain laser action, the injected carriers must reach some threshold value, $N_{th}$, larger than $N_{tr}$ by a sufficient margin to allow the net gain to overcome the cavity losses.

Semiconductor laser pumping can in principle be achieved, and indeed has been achieved, in a number of ways, e.g., by using either the beam of another laser, or an auxiliary electron beam, to transversely or longitudinally excite a bulk semiconductor. By far the most convenient way of excitation is, however, to use the semiconductor laser in the form of a diode with excitation produced by current flowing in the forward direction of the junction.(34) Laser action in a semiconductor was in fact first observed in 1962 by using a *p-n* junction

---

* Condition (9.4.2) is thus equivalent to the condition $N_2 = N_1$ under which a non-degenerate two level system becomes transparent

diode, the demonstration being made almost simultaneously by four groups,[35–38] three of which were using GaAs. The devices developed during the early stage of semiconductor laser research made use of the same material for both the $p$ and $n$ sides of the junction and are therefore referred to as homojunction lasers. The homojunction laser is now only of historical importance, since it has been essentially superseded by the double heterostructure (DH) laser where the active medium is sandwiched between $p$ and $n$ materials which are different from the active material. Homojunction lasers could in fact operate cw only at cryogenic temperatures ($T = 77$ K), while it was only after the invention of the heterojunction laser that it became possible to operate semiconductor lasers cw at room temperature. This development occurred 7 years after the invention of the homojunction laser (1969)[39–41] and opened up the way to the great variety of applications in which semiconductor lasers are nowadays used. Homojunction semiconductor lasers will nevertheless be discussed briefly in the next section since this discussion helps to understand the great advantages offered by the DH lasers.

## 9.4.2. The Homojunction Laser

In the homojunction laser, the pumping process is achieved in a $p$-$n$ junction where both $p$-type and $n$-type regions, being of the same material (e.g., GaAs), are in the form of a degenerate semiconductor. This means that the donor and acceptor concentrations are so large ($\approx 10^{18}$ atoms/cm$^3$) that the Fermi levels fall in the valence band for the $p$ type, $E_{F_p}$, and in the conduction band for the $n$ type, $E_{F_n}$. When a junction is formed, and if no voltage is applied, the band structure will be as shown in Fig. 9.19a, where the two Fermi energies are seen to be the same. When a forward bias voltage $V$ is applied, the band structure becomes as shown in Fig. 9.19b and the two Fermi levels become separated by $\Delta E = eV$. We see from this figure that, in the junction region, electrons are injected into the conduction band (from the $n$-type region) while holes are injected into the valence band (from the $p$-type region). Thus, under appropriate values of current density, the transparency condition and then the laser threshold condition can be reached. Actually, one of the main limitations of this device comes from the very small potential barrier that an electron, in the conduction band, encounters when it reaches the $p$-side of the junction. The electron can then penetrate into the $p$-type material where it becomes a minority carrier thus recombining with a hole. The penetration

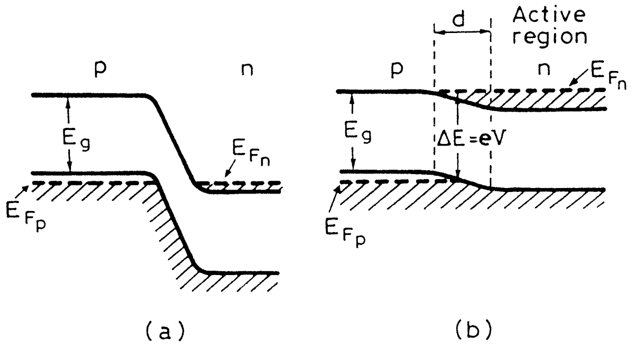

*FIG. 9.19.* Band structure of a $p$-$n$ junction semiconductor laser with zero voltage, (a), and forward voltage, (b), applied to the junction.

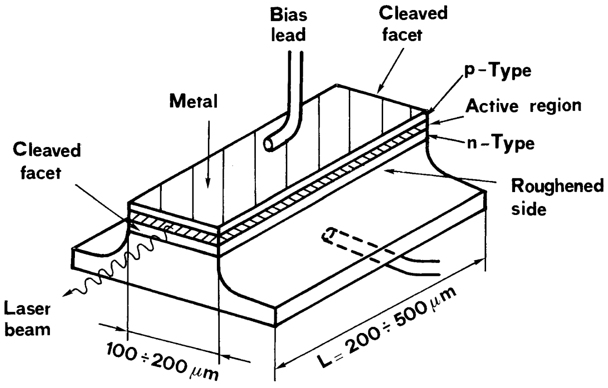

*FIG. 9.20.* Typical broad-area *p-n* homojunction laser.

depth, *d*, of the electron will then be given, according to diffusion theory, by $d = \sqrt{D\tau}$, where $D$ is the diffusion coefficient and $\tau$ is the electron lifetime, as established by electron-hole recombination. In GaAs, $D = 10\,\mathrm{cm}^2/\mathrm{s}$ and $\tau \cong 3\,\mathrm{ns}$ so that $d \approx 1\ \mu\mathrm{m}$ which shows that the active region is quite thick being limited by the diffusion length *d* rather then by the thickness of the depletion layer ($\approx 0.1\ \mu\mathrm{m}$).

A typical configuration of a p-n junction laser is shown in Fig. 9.20, the shaded region corresponding to the active layer. It is seen that the diode dimensions are very small (some hundreds of microns). To provide feedback for laser action, two parallel end faces are prepared, usually by cleavage along crystal planes. Often the two surfaces are not provided with reflective coatings. In fact, since the refractive index of a semiconductor is very large (e.g., $n = 3.6$ for GaAs), there is already a sufficient high reflectivity ($\sim 32\%$ for GaAs) from the Fresnel reflection at the semiconductor-air interface. Note that, as mentioned earlier, the thickness of the active region in the direction perpendicular to the junction is $d \approx 1\ \mu\mathrm{m}$. Because of diffraction, however, the transverse dimension of the laser beam in this direction ($\approx 5\ \mu\mathrm{m}$) is significantly larger than the active region.

A homojunction laser has a very high threshold current density at room temperature ($J_{th} \cong 10^5\ \mathrm{A/cm}^2$) which prevents the laser from operating cw at room temperature (without suffering destruction in a very short time!). There are two main reasons for this high threshold value: (1) The thickness of the active region ($d \approx 1\ \mu\mathrm{m}$) is quite large and the threshold current, being proportional to the volume of the active medium, is proportional to this thickness. (2) The laser beam, owing to its comparatively large transverse dimensions, extends considerably into the *p* and *n* regions, where it is strongly absorbed. Given the above reasons, homojunctions lasers could only operate cw at cryogenic temperatures (typically at liquid nitrogen temperature $T = 77\ \mathrm{K}$). For a given laser transition, in fact, the semiconductor gain, according to Eq. (3.2.37), can be shown to increase rapidly with decreasing temperature and, also, contact of the diode with liquid nitrogen helps to give a very efficient cooling.

### 9.4.3. The Double-Heterostructure Laser

The limitations discussed in the previous section prevented any widespread use of semiconductor lasers until, first, the single-heterostructure, and, immediately after, the

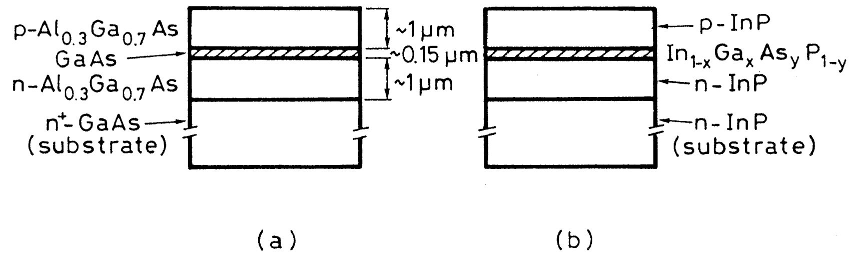

*FIG. 9.21.* Schematic diagram of a double-heterostructure where the active medium (hatched area) consists of GaAs, (a), and InGaAsP, (b).

double-heterostructure lasers were introduced. We will limit our discussion here to the latter type of laser structure since it is the only one that is now in common use.

Two examples of a double-heterostructure, where the active medium is a thin layer (0.1–0.2 $\mu$m) of either GaAs or of the quaternary alloy InGaAsP, are shown in Fig. 9.21a and b respectively. For the two cases considered, the $p$ and $n$ sides are made of $Al_{0.3}Ga_{0.7}As$ and InP, respectively. When properly optimized (see Fig. 9.23), the room-temperature threshold current-density of such a diode structure can be reduced by about two orders of magnitude (i.e., to $\sim 10^3$ A/cm$^2$) compared to the corresponding homojunction devices, thus making cw room temperature operation feasible. This strong reduction of threshold current density is due to the combined effect of three circumstances: (1) The refractive index of the active layer $n_1$ (e.g., $n_1 = 3.6$ for GaAs) is significantly larger than that, $n_2$, of the $p$-side and $n$-side cladding-layers (e.g., $n_2 = 3.4$ for $Al_{0.3}Ga_{0.7}As$), thus providing a guiding structure (see Fig 9.22a). This means that the laser beam will now be mostly confined to the active layer region, i.e., where the gain exists (*photon-confinement*, see Fig. 9.22b). (2) The band gap $E_{g_1}$ of the active layer (e.g., $E_{g_1} \cong 1.5$ eV in GaAs) is significantly smaller than that, $E_{g_2}$, of the cladding layers* (e.g., $E_{g_2} \cong 1.8$ eV for $Al_{0.3}Ga_{0.7}As$). Energy barriers are thus formed at the two junction planes thus effectively confining the injected holes and electrons within the active layer (*carrier-confinement*, see Fig. 9.22c). Thus, for a given current density, the concentration of holes and electrons in the active layer is increased and therefore the gain is increased. (3) Since $E_{g_2}$ is appreciably larger than $E_{g_1}$, the laser beam, which has a frequency $\nu \cong E_{g_1}/h$, is much less strongly absorbed in the tails of the beam profile (see Fig. 9.22b) by the cladding layers, the loss arising, in this case, only from free-carriers (*reduced absorption*).

To form a double heterostructure, thus taking advantage of all its favorable properties, a very important requirement must be fulfilled, namely that the lattice period of the active layer must be equal (to within $\sim 0.1\%$) to that of the cladding layers†. In fact, if this condition is not fulfilled, the resulting strain at the two interfaces will result in misfit dislocations being produced there, each dislocation acting as a rather effective center for electron-hole nonradiative recombination. For the GaAs/AlGaAs structure, this requirement of lattice matching does not constitute a limitation because the lattice periods of GaAs (0.564 m) and AlAs (0.566 m) are

---

* It is a general rule for all III–V compounds that, any change in composition that produces a change, in a given sense e.g. a *decrease*, of band gap also produces a change, in the opposite sense i.e. an *increase*, of the refractive index.

† All III–V compounds crystallize in the cubic structure.

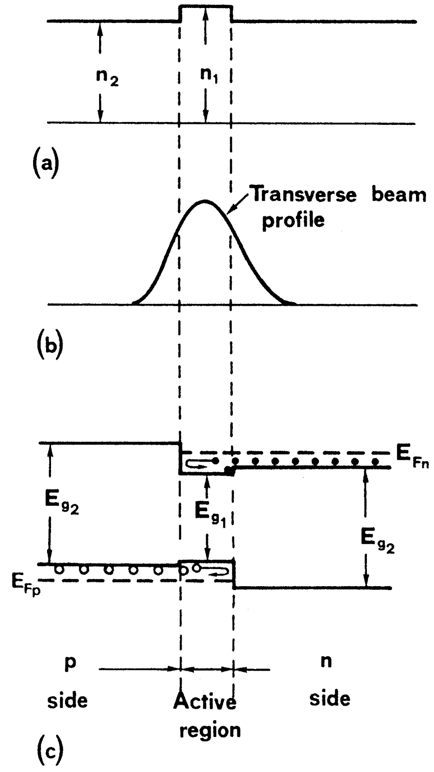

*FIG. 9.22.* (a) Refractive index profile, (b) transverse beam profile, and (c) band structure (very schematic) of a double-heterostructure diode-laser.

very close in value (the atomic radii of Ga and Al are, in fact, almost the same). In the case of the quaternary compound $In_{1-x}Ga_xAs_yP_{1-y}$, the alloy can be lattice matched to InP for a specific $y/x$ ratio, as one can appreciate from the following argument: suppose that, starting with InP for the active layer, some fraction, $x$, of Ga is added, substituting for some In in the lattice (which hence becomes $In_{1-x}$). Since the radius of Ga is *smaller* (by ~19 pm) than that of In, the lattice period of the $In_{1-x}Ga_xP$ will be decreased compared to InP. Suppose now that some fraction, $y$, of As ($As_y$) substitutes for P (thus becoming $P_{1-y}$). Since the radius of As is now *larger* (by ~10 pm) than that of P, this addition will tend to increase the lattice period. So, if the $y/x$ ratio of the two substituents has an appropriate value, the two effects will cancel each other thus resulting in $In_{1-x}Ga_xAs_yP_{1-y}$ being lattice matched to InP. This lattice-matching condition turns out to be given by $y \cong 2.2\,x$. Upon changing $x$, while keeping the $y/x$ ratio equal to the lattice-matching value, the semiconductor band-gap and hence the emission wavelength can be changed. In this way the emission wavelength of $In_{1-x}Ga_xAs_yP_{1-y}$ can be varied between 1,150 and 1,670 nm, for cw room temperature operation, thus encompassing the so-called second (~1300 nm) and third (~1550 nm) transmission windows of silica optical fibers.

Experimental and theoretical plots of the threshold current density, $J_{th}$, vs thickness, $d$, of the active layer, for a DH GaAs laser are shown in Fig. 9.23.(42) Note that, as $d$ decreases, $J_{th}$ first decreases then reaches a minimum value ($J_{th} \cong 1\ \text{kA/cm}^2$ for $d \cong 0.1\ \mu\text{m}$) and thereafter increases. To understand this behavior, we need first to relate the threshold current density, $J_{th}$,

## 15.3 GaAs/$Ga_{1-x}Al_xAs$ LASERS

According to Table 15.1, we note that the energy gap of binary III–V semiconductors covers a broad range (from a fraction of 1 eV to a few eV) for applications. Since the bandgap translates to the emission photon energy of lasers, a wide emission spectrum is covered. In practice, however, operational devices are limited by the availability of substrate materials and the possibility of lattice matching in the crystal growth of the laser structures. The most common substrates that are available with relatively low defect density are GaAs and InP crystals. In the early stage of the crystal growth development, it was realized that a solid solution of two binary compound semiconductors can form ternary or quaternary alloy semiconductors. For example, the solid solution of GaAs and AlAs can form $Ga_{1-x}Al_xAs$ semiconductors. The two most important classes of semiconductor lasers are those that are based on III–V semiconductors that can be grown on GaAs or InP substrates.

The first system is based on GaAs and $Ga_{1-x}Al_xAs$ semiconductor crystals. In this case, the active region is either GaAs or $Ga_{1-x}Al_xAs$. Since an AlAs semiconductor crystal has a larger bandgap than that of GaAs, the ternary compound crystal $Ga_{1-x}Al_xAs$ has a bandgap between those of GaAs ($x = 0$, $E_{gap} = 1.43$ eV) and AlAs ($x = 1$, $E_{gap} = 2.1$ eV). The subscript $x$ indicates the fraction of Ga atoms in GaAs that are replaced by Al. The resulting lasers emit (depending on the active region molar fraction $x$ and its doping) at $0.75\ \mu m < \lambda < 0.88\ \mu m$. This spectral region is convenient for short-haul (2 km) optical communications in silica fibers.

The second system has $Ga_{1-x}In_xAs_{1-y}P_y$ as its active region. The lasers emit in the range $1.1\ \mu m < \lambda < 1.6\ \mu m$ depending on $x$ and $y$. The region near 1.55 μm is especially favorable, since, as shown in Figure 3.21, optical fibers are available with losses as small as 0.15 dB/km at this wavelength, making it extremely desirable for long-distance optical communications. In this section we will consider GaAs/$Ga_{1-x}Al_xAs$ lasers. A generic laser of this type, depicted in Figure 15.10, has a thin (0.1–0.2 μm) region of GaAs sandwiched between two regions of GaAlAs. It is consequently called a *double heterostructure* laser. The basic layered structure

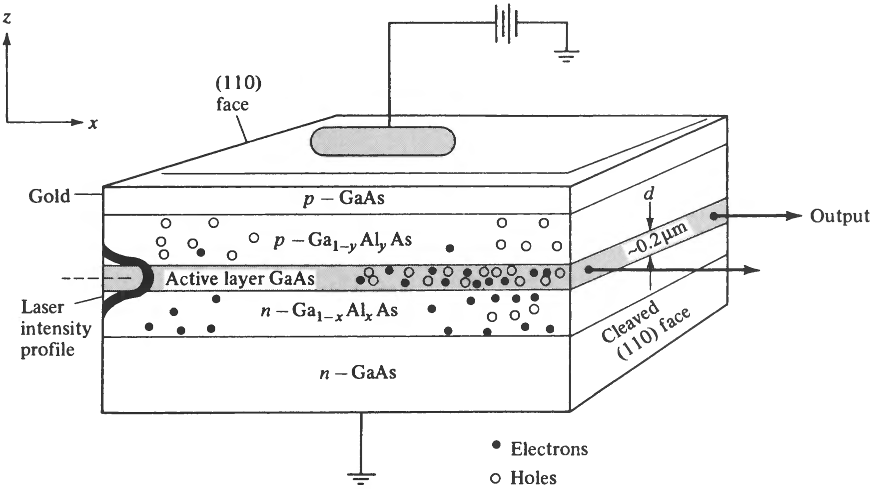

**Figure 15.10** A typical double heterostructure GaAs/GaAlAs laser. Electrons and holes are injected into the active GaAs layer from the $n$ and $p$ GaAlAs. Photons with frequencies near $\nu = E_g/h$ are amplified by stimulating electron–hole recombination.

is grown epitaxially on a crystalline GaAs substrate so that it is uninterrupted crystallographically. The favored crystal growth techniques are liquid-phase epitaxy and chemical vapor deposition using metallo-organic reagents (MOCVD) [11, 13, 14]. Another important technique—molecular beam epitaxy (MBE) [11, 13, 15, 16]—uses atomic beams of the crystal constituents in ultrahigh vacuum to achieve extremely fine thickness and doping control.

The thin active region is usually undoped while one of the bounding $\mathrm{Ga}_{1-x}\mathrm{Al}_x\mathrm{As}$ layers is doped heavily $n$-type and the other $p$-type. The difference

$$n_{\mathrm{GaAs}} - n_{\mathrm{Ga}_{1-x}\mathrm{Al}_x\mathrm{As}} \cong 0.62x$$

between the indices of refraction of GaAs and the ternary crystal with a molar fraction $x$ gives rise to a three-layered dielectric waveguide of the type illustrated in Figure 3.1. At this point the student should review the basic modal concepts discussed in Chapter 3. The lowest-order (fundamental) confined mode has its energy concentrated mostly in the GaAs (high-index) layer. The index distribution and a typical modal intensity plot for the lowest-order mode are shown in Figure 15.11. When a positive bias is applied to the device, electrons are injected from the $n$-type $\mathrm{Ga}_{1-x}\mathrm{Al}_x\mathrm{As}$ into the active GaAs region while a density of holes equal to that of the electrons in the active region is caused by injection from the $p$ side. The density of holes must equal that of the electrons to achieve charge neutrality.

The electrons that are injected into the active region are prevented from diffusing out into the $p$ region by means of the potential barrier $\Delta E_c$ due to the difference $\Delta E_g$ between the energy gaps of GaAs and $\mathrm{Ga}_{1-x}\mathrm{Al}_x\mathrm{As}$. The $x$ dependence of the energy gap of $\mathrm{Ga}_{1-x}\mathrm{Al}_x\mathrm{As}$ is approximated by [13]

$$E_g(x < 0.37) = (1.424 + 1.247x)\ \mathrm{eV}$$

and is plotted in Figure 15.12.

The total discontinuity $\Delta E_g$ of the energy gap at a GaAs/GaAlAs interface is taken up mostly (60%) by the conduction bandedge, that is, $\Delta E_c = 0.6\ \Delta E_g$, while 40% is left to the valence band, $\Delta E_v = 0.4\ \Delta E_g$, so that both holes and electrons are effectively confined to the active region. This double confinement of injected carriers as well as of the optical mode energy to the same region is probably the single most important factor responsible for the successful realization of low-threshold continuous semiconductor lasers [17–19]. Under these conditions, we expect the gain experienced by the mode to vary as $1/d$, where $d$ is the thickness of the active (GaAs) layer, since at a given total current, the carrier density, hence the gain, will be proportional to $1/d$. To further discuss the dependence on the thickness of the active region $d$, we start with the basic definition of the modal gain,

$$g = \frac{\text{Power generated per unit length (in } x)}{\text{Power carried by beam}}$$

$$= \frac{-\int_{-\infty}^{-d/2} \alpha_n |E|^2 dz + \int_{-d/2}^{d/2} \gamma |E|^2 dz - \int_{d/2}^{\infty} \alpha_p |E|^2 dz}{\int_{-\infty}^{\infty} |E|^2 dz} \tag{15.3-1}$$

where $\gamma$ is the gain coefficient experienced by a plane wave in a medium whose inversion density is equal to that of the active medium. The gain coefficient $\gamma$ is given by Equations (15.2-12) and (15.2-17). The parameter $\alpha_n$ is the loss constant of the unpumped $n$-$\mathrm{Ga}_{1-x}\mathrm{Al}_x\mathrm{As}$ and is due mostly to free electron absorption. $\alpha_p$ is the loss (by free holes) in the bounding $p$-$\mathrm{Ga}_{1-y}\mathrm{Al}_y\mathrm{As}$ region. We note that as $d \to \infty$, $g \to \gamma$.

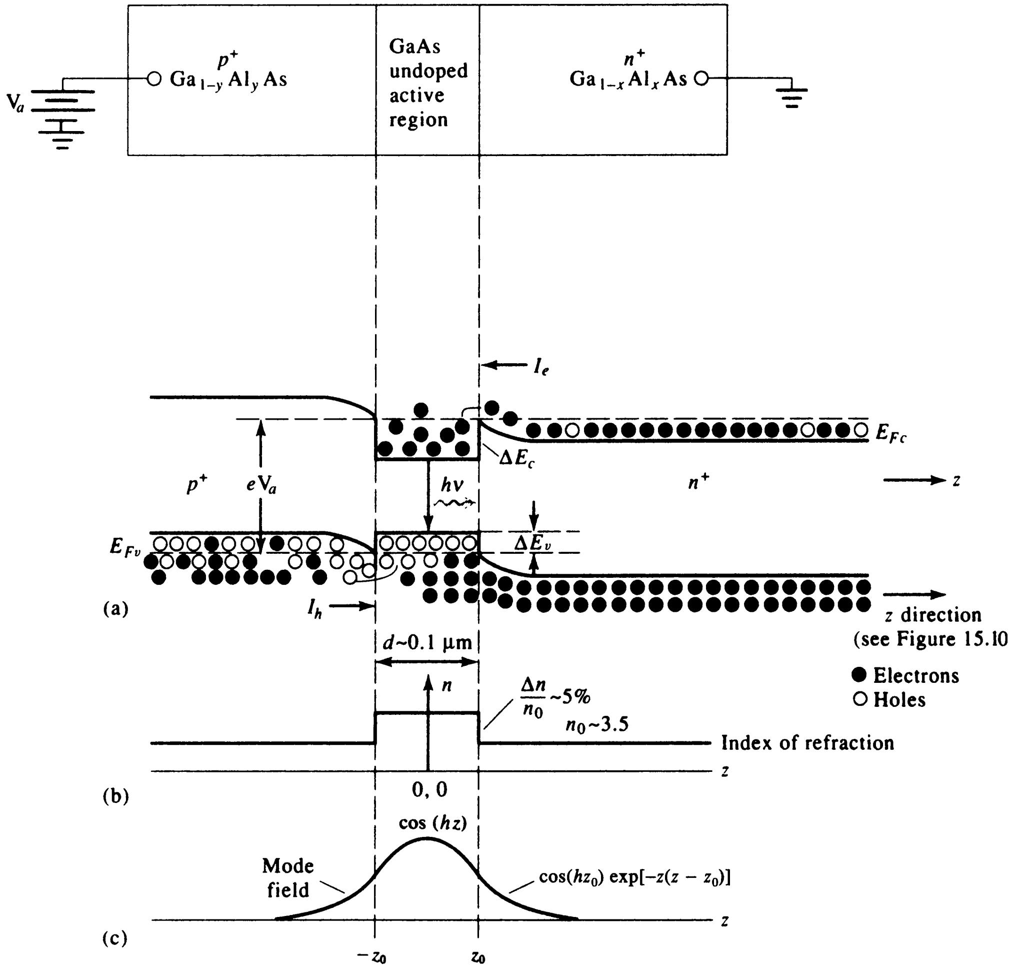

**Figure 15.11** (a) The energy bandedges of a strongly forward-biased (near-flattened) double heterostructure GaAs/GaAlAs laser diode. Note the trapping of electrons (holes) in the potential well formed by the conduction (valence) bandedge energy discontinuity $\Delta E_c (\Delta E_v)$. (b) The spatial ($z$) profile of the index of refraction which is responsible for dielectric waveguiding in the high-index (GaAs) layer. (c) The mode profile of the fundamental optical mode in a slab waveguide.

In the case when $\gamma$ is a constant over $-d/2 < z < d/2$. It is convenient to rewrite Equation (15.3-1) as

$$g = \gamma\Gamma_a - \alpha_n\Gamma_n - \alpha_p\Gamma_p \tag{15.3-2}$$

with

$$\Gamma_a = \frac{\int_{-d/2}^{d/2} |E|^2 dz}{\int_{-\infty}^{\infty} |E|^2 dz}, \quad \Gamma_n = \frac{\int_{-\infty}^{-d/2} |E|^2 dz}{\int_{-\infty}^{\infty} |E|^2 dz}, \quad \Gamma_p = \frac{\int_{d/2}^{\infty} |E|^2 dz}{\int_{-\infty}^{\infty} |E|^2 dz} \tag{15.3-3}$$