CHAPTER 5

Lasers I

The word 'laser' is an acronym for '*l*ight *a*mplification by *s*timulated *e*mission of *r*adiation'. Albert Einstein in 1917 showed that the process of stimulated emission must exist, but it was not until 1960 that T. H. Maiman (ref. 5.1) first achieved laser action at optical frequencies in ruby. The basic principles and construction of a laser are relatively straightforward and it is somewhat surprising that the invention of the laser was so long delayed. A rigorous analysis of the physics of the laser, on the other hand, is quite difficult and the treatment we give below is somewhat simplified. The development of lasers since 1960 has been extremely rapid and although applications for lasers had a very slow start during their first decade, new applications for laser radiation are being found now almost every day (see section 6.5); a selection of texts on laser theory and applications is given in ref 5.2. In view of the increased use of lasers, and as laser radiation is potentially hazardous, some comments on laser safety have been included as Appendix 7.

5.1 Emission and absorption of radiation

It was seen in Chapter 1 that when an electron in an atom undergoes transitions between two energy states or levels it either emits or absorbs a photon, which can be described in terms of a wave of frequency ν where $\nu = \Delta E/h$, ΔE being the energy difference between the two levels concerned. Let us consider the electron transitions which may occur between the two energy levels of the hypothetical atomic system shown in Fig. 5.1. If the electron

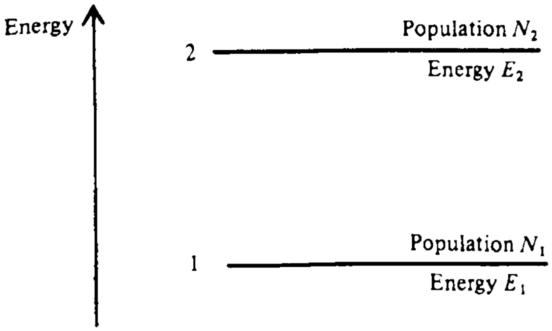


FIG. 5.1 Two-energy-level system.

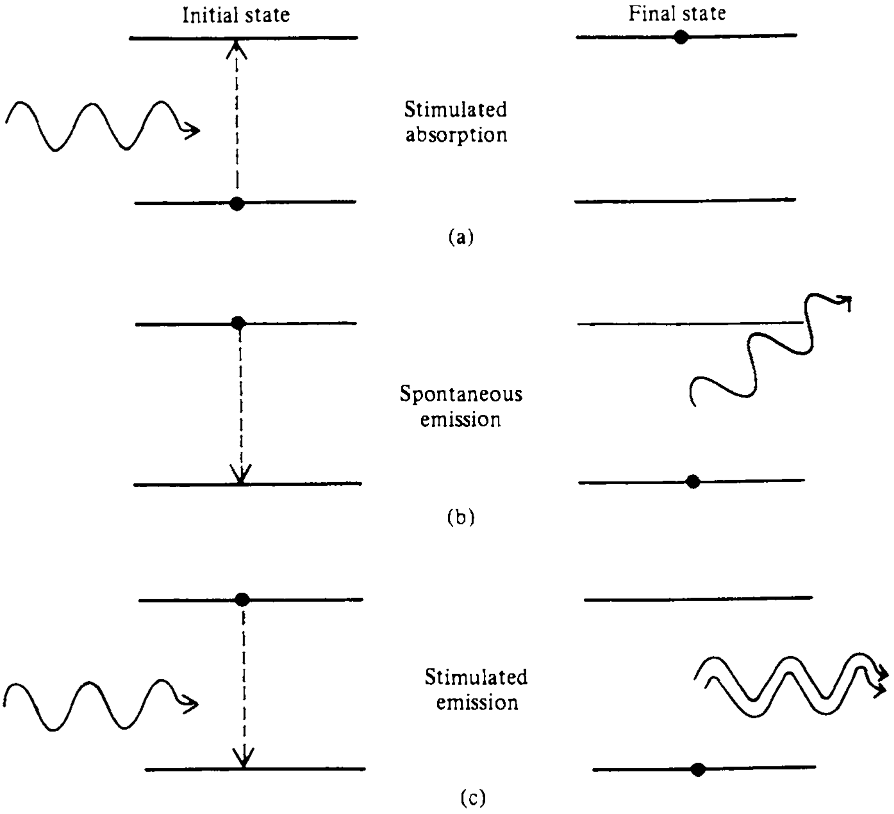


FIG. 5.2 Energy level diagram illustrating (a) absorption, (b) spontaneous emission and (c) stimulated emission. The black dot indicates the state of the atom before and after the transition.

is in the lower level E_1 then in the presence of photons of energy $(E_2 - E_1)$ it may be excited to the upper level E_2 by absorbing a photon. Alternatively if the electron is in the level E_2 it may return to the ground state with the emission of a photon. The emission process may occur in two distinct ways. These are (a) the *spontaneous emission* process in which the electron drops to the lower level in an entirely random way and (b) the *stimulated emission* process in which the electron is 'triggered' to undergo the transition by the presence of photons of energy $(E_2 - E_1)$. There is nothing mystical in this, as the electron would undergo this process sooner or later spontaneously; the transition is simply initiated by the presence of the stimulating photon.

The absorption and emission processes are illustrated in Figs 5.2(a), (b) and (c). Under normal circumstances we do not observe the stimulated process because the probability of the spontaneous process occurring is much higher. The average time the electron exists in the excited state before making a spontaneous transition is called the lifetime τ_{21} of the excited state. The '21' here indicates the energy levels involved. The probability that a particular atom will undergo spontaneous emission within a time interval dt is given by

$A_{21}\,\mathrm{d}t = \mathrm{d}t/\tau_{21}$, where A_{21} is the spontaneous transition rate. Because the spontaneous radiation from any atom is emitted at random, the radiation emitted by a large number of atoms will clearly be incoherent. In contrast to this, the stimulated emission process results in coherent radiation since the waves associated with the stimulating and stimulated photons have identical frequencies (but see section 5.7), are in phase, have the same state of polarization and travel in the same direction. This means that with stimulated emission the amplitude of an incident wave can grow as it passes through a collection of excited atoms in what is clearly an amplification process. As the absorption transition, in common with stimulated emission, can only occur in the presence of photons of appropriate energy, it is often referred to as stimulated absorption. These two processes may be regarded as the inverse of one another.

The above discussion ignores the fact that the emission and absorption processes do not simply involve photons of a precisely defined energy. Consequently there is a range of frequencies associated with these processes. The distribution of energies or frequencies in a given transition is described by the *lineshape function* $g(\nu)$, which is discussed in section 5.7.

5.2 Einstein relations

Einstein (ref. 5.3) showed that the parameters describing the above three processes are related through the requirement that for a system in thermal equilibrium the rate of upward transitions (from E_1 to E_2) must equal the rate of the downward transition processes (from E_2 to E_1). Let us suppose that our simple two-level atomic energy level system is in equilibrium inside a blackbody cavity, where, as we saw in section 1.4, the radiation covers a very wide frequency range. Although we indicated at the end of the previous section that the upward and downward transitions involve a spread of frequencies, we may assume that this will be very small compared with that of a blackbody. Nevertheless in further considering the transitions we must take into account the behaviour of photons within these frequency distributions. This analysis is undertaken in Appendix 4; meanwhile we adopt a simpler, approximate approach.

If there are N_1 atoms per unit volume in the collection with energy E_1, then the upward transition or absorption rate will be proportional to both N_1 and to the number of photons available at the correct frequency. Now ρ_ν, the energy density at frequency ν, is given by $\rho_\nu = \mathcal{N}_\nu h\nu$ where $\mathcal{N}_\nu$ is the number of photons per unit volume having frequency ν. Therefore we may write the upward transition rate as $N_1\rho_\nu B_{12}$ where B_{12} is a constant. Similarly if there are N_2 atoms per unit volume in the collection with energy E_2 then the induced transition rate from level 2 to level 1 is $N_2\rho_\nu B_{12}$, where again B_{21} is a constant. The spontaneous transition rate from level 2 to level 1 is simply N_2A_{21}. The total downward transition rate is the sum of the induced and spontaneous contributions, that is

$$N_2\rho_\nu B_{21} + N_2A_{21}$$

A_{21}, B_{21} and B_{12} are called the *Einstein coefficients*; the relationships between them can be established as follows.

For a system in equilibrium, the upward and downward transition rates must be equal.

Clearly, in equilibrium, we must have:

$$N_1 \rho_\nu B_{12} = N_2 \rho_\nu B_{21} + N_2 A_{21} \tag{6.10}$$

since the total upward and downward transition rates must be equal.

Hence from (6.10):

$$\rho_\nu = \frac{(A_{21}/B_{21})}{(B_{12}N_1/B_{21}N_2) - 1}$$

But we know from the Boltzmann relation (4.13) that:

$$\frac{N_1}{N_2} = \exp\left(-\frac{E_1 - E_2}{kT}\right)$$

and also that $E_2 - E_1 = h\nu_{12}$.

Hence, generalizing from ν_{12} to ν:

$$\rho_\nu = \frac{(A_{21}/B_{21})}{(B_{12}/B_{21})\exp\left(\frac{h\nu}{kT}\right) - 1} \tag{6.11}$$

Now it was shown in section 6.2 that for equilibrium (black-body) radiation (equation 6.8):

$$\rho_\nu = \frac{8\pi h\nu^3}{c^3}\frac{1}{\exp\left(\frac{h\nu}{kT}\right) - 1}$$

Hence it follows, by comparing this with (6.11):

$$B_{12} = B_{21} \tag{6.12a}$$

$$A_{21} = B_{21}\frac{8\pi h\nu^3}{c^3} \tag{6.12b}$$

Relations (6.12) are known as the Einstein relations, and are very important determinants in the relationships between atoms and radiation. For example, it is clear that, under these conditions, the ratio of stimulated to spontaneous emission from E_2 to E_1 is given by:

$$S = \frac{R_{21}}{S_{21}} = \frac{\rho_\nu N_2 B_{21}}{N_2 A_{21}} = \frac{\rho_\nu c^3}{8\pi h\nu^3}$$

and using the expression for ρ_ν from equation (6.8):

$$S = \frac{1}{\exp\left(\frac{h\nu}{kT}\right) - 1}$$

If, for example, we consider the specific case of the He-Ne discharge at a temperature of 370 K with $\lambda = 632.8\,\text{nm}$ ($\nu = 4.74 \times 10^{14}\,\text{Hz}$) then we find

$$\underline{S \approx 2 \times 10^{-27}}$$

Stimulated emission is thus very unlikely for equilibrium systems.

Another point worthy of note is that, for given values of N_2 (density of atoms in upper state E_2) and ρ_ν (density of photons) the rate of stimulated emission (B_{21}) is proportional to $1/\nu^3$. This follows from equation (6.12b) since

$$B_{21} = \frac{A_{21}c^3}{8\pi h\nu^3}$$

and A_{21} is an atomic constant, representing the reciprocal of the spontaneous decay time.

This means that the higher the frequency the more difficult is laser action, for this depends upon stimulated emission (section 4.4.2). Ultraviolet, X-ray and γ-ray lasers present very special problems which, hopefully, will preclude the possibility of 'death-ray' weapons (X-rays and γ-rays are very damaging to living tissues).

However, we do wish to use lasers at lower frequencies, visible and infrared for example, for purposes of communication, display and measurement, and the equation for R_{21} tells us that the way to increase the stimulated emission is to increase the values of N_2 and ρ_ν.

We know that, in equilibrium, $N_2 < N_1$, from the form of the Boltzmann factor, and ρ_ν is given by equation (6.8). Hence we shall have to disturb the equilibrium to achieve significant levels of stimulated emission.

One way in which this can be done is to inject radiation at frequency ν, so that ρ_ν is increased above its equilibrium value. Suppose that this is done until the stimulated emission greatly exceeds the spontaneous emission (which does not, of course, depend upon ρ_ν), i.e., until:

$$N_2\rho_\nu B_{21} \gg N_2A_{21}$$

The condition for this, clearly, is that

$$\rho_\nu \gg \frac{A_{21}}{B_{21}}$$

which, from (6.12), means that:

$$\rho_\nu \gg \frac{8\pi h\nu^3}{c^3}$$

However, increasing ρ_ν does also increase the stimulated absorption. In fact, equation (6.10) becomes, when ρ_ν is large:

$$N_1\rho_\nu B_{12} = N_2\rho_\nu B_{21}$$

But we also know from equations (6.12) that $B_{12} = B_{21}$; hence $N_1 = N_2$ under these conditions. In other words, an incoming photon at frequency ν is just as likely to cause a downward transition (stimulated emission) as it is an upward one (stimulated absorption). Hence we cannot increase the population N_2 above that of N_1 simply by pumping more radiation,

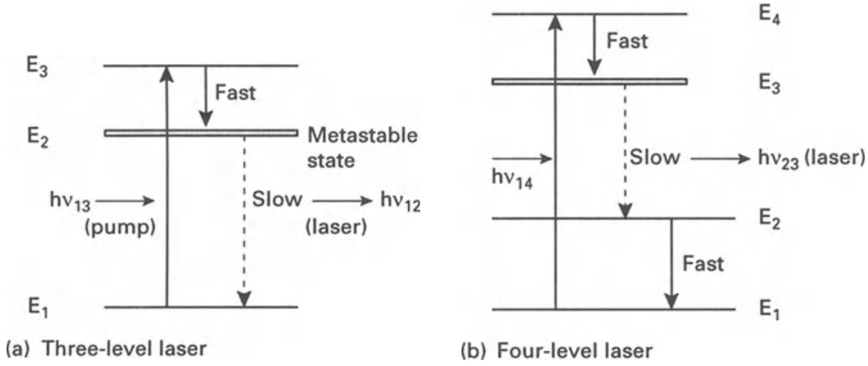


Fig. 6.4 Energy level diagrams for laser action.

at frequency ν, into the system. Clearly, we must change tack if we are to enhance the stimulated emission and produce a laser.

Consider a three-level rather than a two-level system (Fig. 6.4(a)). Suppose that light at frequency ν_{13} is injected into this system, so that there is a large amount of stimulated absorption from E_1 to E_3. Spontaneous decays will occur from E_3 to E_2 and then $E_2 \rightarrow E_1$ with also $E_3 \rightarrow E_1$; but if the levels are chosen appropriately according to the quantum rules, the $E_3 \rightarrow E_2$ decay can be fast and the $E_2 \rightarrow E_1$ relatively much slower. Clearly the result of this will be that atoms will accumulate in level E_2. Now the really important point is that, unlike the previous two-level case, atoms in level E_2 are immune from stimulated emission by photons at frequency ν_{13}. Hence we can now increase the numbers of atoms in level E_2, at the expense of those in E_1, by increasing the intensity of the radiation at frequency ν_{13}. We can thus soon ensure that:

$$N_2 > N_1$$

and we have an 'inverted population' (i.e., more atoms in a higher energy state than a lower one) as a result of the 'pump' at frequency ν_{13}. This inverted population can now be exploited to give optical amplification at frequency ν_{12}.

Let us quantify this amplification via the rate equations we have developed.

5.3 Absorption of radiation

Let us consider a collimated beam of perfectly monochromatic radiation of unit cross-sectional area passing through an absorbing medium. We assume for simplicity that there is only one relevant electron transition, which occurs between the energy levels E_1 and E_2. Then the change in irradiance of the beam as a function of distance is given by

$$\Delta I(x) = I(x + \Delta x) - I(x)$$

For a homogeneous medium $\Delta I(x)$ is proportional both to the distance travelled Δx and to $I(x)$. That is, $\Delta I(x) = -\alpha I(x)\Delta x$, where the constant of proportionality, α, is the *absorption coefficient*. The negative sign indicates the reduction in beam irradiance due to absorption as α is a positive quantity. Writing this expression as a differential equation we have

$$\frac{dI(x)}{dx} = -\alpha I(x)$$

Integrating this equation gives

$$I = I_0 \exp(-\alpha x) \tag{5.10}$$

where I_0 is the incident irradiance.

Let us consider the absorption coefficient in more detail. Clearly, the degree of absorption of the beam will depend on how many atoms N_1 there are with electrons in the lower energy state E_1 and on how many atoms N_2 there are in energy state E_2. If N_2 were zero, then absorption would be a maximum, while if all of the atoms were in the upper state the absorption would be zero and the probability of stimulated emission would be large.

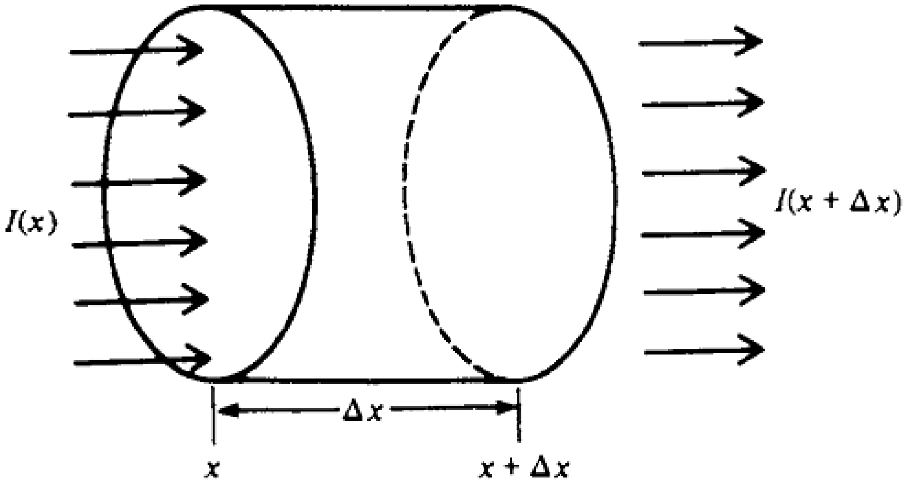


FIG. 5.3 Radiation passing through a volume element of length Δx and unit cross-sectional area.

Considering $-\alpha = g$ in the (5.10) it is possible to demonstrate that

$$g = \frac{h\nu}{c} B_{12}(N_2 - N_1) \tag{6.14a}$$

which is the gain coefficient for the medium (fractional increase in intensity level per unit length) and will be positive (i.e., gain rather than loss) provided that $N_2 > N_1$, as will be the case for an inverted population. Hence this medium is an optical amplifier. The injected radiation at frequency ν_{12} receives gain from the optical pump of amount:

$$G = \frac{I}{I_0} = \exp(gs)$$

so that it increases exponentially with distance into the medium. Clearly g in equation (6.14) is proportional to $(N_2 - N_1)$. In a three-level system such as we are considering the lower level of the amplifying transition is the ground state, which is initially heavily populated. It follows that more than half the atoms must be excited by the pump before population inversion can be achieved ($N_2 > N_1$). It is quite hard work for the pump to excite all these atoms. Consider, however, the *four*-level system shown in Fig. (6.4(b)). Here the pump is at ν_{14}, there is a quick decay to level 3 and a slow one to levels 2 and 1. The decay from 2 to 1 is again fast. Clearly the consequence of this is that it is relatively easy to provide level 3 with an inverted population over level 2, since level 2 was not well populated in the first place (being above the ground state), and atoms do not accumulate there since it decays quickly to ground. Hence we can ensure, fairly easily, that:

$$N_3 \gg N_2$$

with much less pump power than for $N_2 > N_1$ in the three-level case. The amplification at ν_{32} is thus much more efficient, and the four-level system makes for a more efficient amplifier.

5.5 Optical feedback

The laser, despite its name, is more analogous to an oscillator than an amplifier. In an electronic oscillator, an amplifier tuned to a particular frequency is provided with positive feedback and, when switched on, any electrical noise signal of the appropriate frequency appearing at the input will be amplified. The amplified output is fed back to the input and amplified yet again and so on. A stable output is quickly reached, however, since the amplifier saturates at high input voltages, as it cannot produce a larger output than the supply voltage.

In the laser, positive feedback may be obtained by placing the gain medium between a pair of mirrors which, in fact, form an optical cavity (a Fabry – Perot resonator). The initial stimulus is provided by any spontaneous transitions between appropriate energy levels in which the emitted photon travels along the axis of the system. The signal is amplified as it passes through the medium and 'fed back' by the mirrors. Saturation is reached when the gain provided by the medium exactly matches the losses incurred during a complete round trip.

The gain per unit length of many active media is so small that very little amplification of a beam of light results from a single pass through the medium. In the multiple passes which a beam undergoes when the medium is placed within a cavity, however, the amplification may be substantial.

We have tacitly assumed that the radiation within the cavity propagates to and fro between two plane – parallel mirrors in a well-collimated beam. Because of diffraction effects, however, this cannot be the case as a perfectly collimated beam cannot be maintained with mirrors of finite extent; some radiation will spread out beyond the edges of the mirrors. Diffraction losses of this nature can be reduced by using concave mirrors. In practice a number of different mirror curvatures and configurations are used depending on the applications envisaged and the type of laser being used.

A detailed analysis of the effects of different mirror systems requires a rigorous application of diffraction theory and is beyond the scope of this book (see e.g. ref. 5.2a). Using simple ray tracing techniques, however, it is quite easy to anticipate the results of such an analysis in that mirror configurations which retain a ray of light, initially inclined at a small angle to the axis, within the optical cavity after several reflections are likely to be useful (see ref. 5.5). Such cavities are said to be stable.

The commonly used mirror configurations are shown in Fig. 5.7; they all have various advantages and disadvantages. The plane – parallel configuration, for example, is very difficult to align, for if the mirrors are not strictly parallel (to within about 1 second of arc) the optical beam will 'walk off' the mirrors after a few reflections. On the other hand, the radiation beam makes maximum use of the laser medium (we say that it has a large *mode volume* – see also section 5.9) as there is no focusing of the beam within the cavity. In addition, the mirrors need to be flat to within $\lambda/100$. In contrast to the plane – parallel case, the confocal arrangement is relatively easy to align (an accuracy of 1.5 minutes of arc is sufficient) but the use of the active medium is restricted (i.e. the mode volume is small). In gas lasers, if maximum power output is required we use a large radius resonator, while if uniphase operation (i.e. maximum beam coherence) is required we use the hemispherical system.

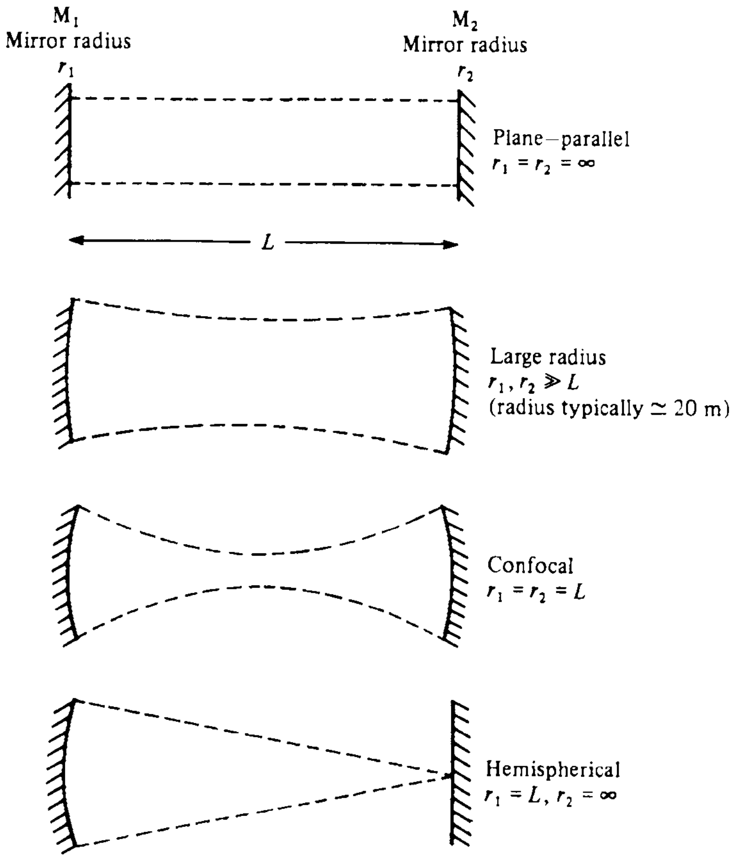


FIG. 5.7 Some commonly used laser cavity mirror configurations (the dashed lines show the extent of the mode volume in each case).

Sometimes mirror configurations are used which give rise to unstable cavities. In these a ray which is initially travelling at a small angle of incidence to the cavity axis will diverge away from the axis after a number of reflections. Such resonators are characterized by high losses, but even so they have some useful properties. In particular they can make efficient use of the mode volume. As unstable resonators have large losses they can be used effectively only with high gain media such as carbon dioxide. As we mentioned earlier, the gain is usually very small so it is essential to minimize all losses in the laser (see section 5.6). One source of loss is absorption in the mirrors. To reduce this, high reflectance multilayer dielectric coatings on the mirrors are used rather than metallic coatings. In these so-called multilayer stacks there is a sequence of quarter-wave (i.e. $\lambda/4$) layers of alternate high and low refractive index dielectric materials on a glass substrate. Because of the phase changes occurring at alternate interfaces, all the reflected waves are in phase and add constructively. More than 20 such layers may be needed to give reflectances in excess of 99.9% – lower reflectances require fewer layers. Clearly, the mirrors will only be effective over a narrow wavelength range. A familiar example of this sort of process is the blooming of camera lenses to reduce unwanted reflections.

We can now derive the minimum pump power required (i.e. the threshold condition) in terms of the parameters of the whole system for laser oscillations to occur.

(ii) *The laser structure*

Having arranged for efficient amplification to take place in a medium it is a relatively straightforward matter to turn it into an oscillator, i.e., a laser source. To do this for any amplifier it is necessary to provide positive feedback, i.e., to feed some of the amplified output back into the amplifier in reinforcing phase.

As has been described in section 4.4.2 this is done by placing parallel mirrors at each end of a box containing the medium, to form a Fabry-Perot cavity (section 2.9). (We should also remember the 'stable' resonator configuration (section 2.11) which is valuable for many types of laser design.) The essential physics of this process is that any given photon at ν_{12} will be bounced back and forth between the mirrors, stimulating the emission of other such photons as it does so, whereas without the mirrors it would make only one such pass.

An important condition for any system to oscillate under these circumstances is that the gain should be in excess of the loss for each cycle of oscillation. The total loss for a photon executing a double passage of the cavity (Fig. 6.5) will depend not only on the loss per unit length in the medium (due to scattering, excitations to other states, wall losses, etc.) but also on the losses at the mirrors, and, it must be remembered, one of the mirrors has to be a partial mirror in order to let some of the light out, otherwise we couldn't use the laser oscillator as a source! Hence the condition for oscillation:

$$\frac{I_f}{I_i} = R_1 R_2 \exp[(g - \alpha)2l] > 1 \qquad (6.14b)$$

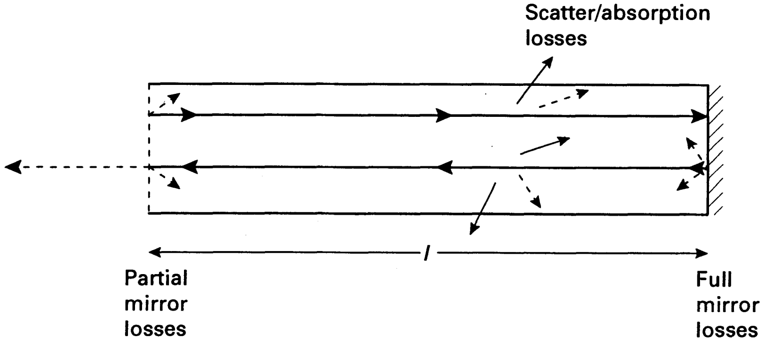


Fig. 6.5 Loss mechanisms in a laser cavity.

where I_f and I_i refer to the final and initial intensities for the double passage of the cavity, R_1 and R_2 are, respectively, the reflectivities for the two mirrors, α is the loss per unit length in the medium and l is the cavity length. (The factor 2 in the exponential refers of course to the double passage of the photon.)

One further word of warning: the value of g must correspond to the population inversion whilst oscillation is taking place, not the value before the feedback is applied. Clearly the value of N_2/N_1 will be very different, once stimulated emission starts to occur, from its value when the system is simply being pumped into its inverted state. This has implications for pumping rates and the balancing of rate equations which we shall not pursue: the principles, hopefully, are quite clear, however.

The simple arrangement of a pumped medium lying between two parallel mirrors (one partial) will, under the correct pump conditions, therefore lead to radiation emerging with the following properties:

(a) narrow linewidth, since only one energy of transition is involved in the laser action; and the mirrors, if wavelength selective, will block any spontaneous light which is emitted in addition.

(b) the output direction of the light will be exactly normal to the (accurately parallel) planes of the mirrors and thus will be highly collimated in one direction.

(c) When a photon is emitted via stimulated emission by another photon, it is emitted with the same phase as the original photon (remember the driving force/resonating system analogy), thus all the laser photons are locked in phase: we have coherent light (within the limitations only of the linewidth of the transition).

(d) The light can be very intense since all the 'light amplification by stimulated emission of radiation' from a long length of medium with small cross-sectional area can be collimated into the one direction.

The above important features summarize the basic properties of laser light:

it is pure (in wavelength and phase), intense, well-collimated light. It is thus easy to control and modulate; it is a powerful tool.

In order to enhance its usefulness as a tool there are two quite simple additions which can be made to the basic design:

The Fabry-Perot cavity formed by the two parallel mirrors will possess defined longitudinal 'modes' as explained in section 2.9. Waves propagating in opposite directions within the cavity, normal to the mirrors, will interfere and reinforce to give rise to an allowable stable mode only when

$$2L = m\lambda$$

where L is the length of the cavity and m is an integer.

From this we can also write

$$\lambda = \frac{2L}{m}; \quad f = \frac{cm}{2L}$$

At all other wavelengths there is destructive interference. Now, of course, the stimulated emission occurs over a small range of wavelengths. This range is determined by the spectral width of the downward transition. The width depends upon a number of factors but primarily (unless cooled to very low temperatures) on the Doppler shift caused by the thermal motion of the molecules. Clearly, at any given time, some molecules will be moving towards the stimulating photon and others away, leading to a spread of Doppler shifts around the central line for the stationary molecule (at absolute zero of temperature!).

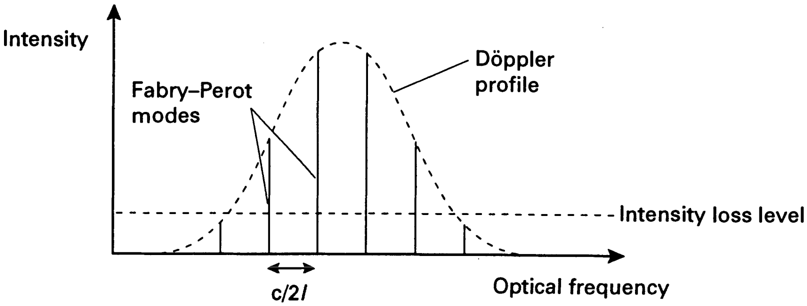


Fig. 6.6 Laser-cavity spectrum.

The output spectrum of the laser light thus is the result of combining these two features, as shown in Fig. 6.6. Here we can see the Fabry-Perot mode structure enveloped by the natural linewidth of the transition.

So far we have dealt only with longitudinal modes; but off-axis rays also may interfere (Fig. 6.7). The reinforcement condition now depends also on the angle which the ray makes with the long axis, and the result is a variation in intensity over the cross-section of the cavity, and thus over the cross-section of the output laser beam (Fig. 6.7). (The notation used to classify these variations will be described in more detail when we deal with wave guiding (Chapter 8) but TEM stands for 'transverse electromagnetic' and the two suffixes refer to the number of minima in the pattern in the horizontal and vertical directions, respectively.)

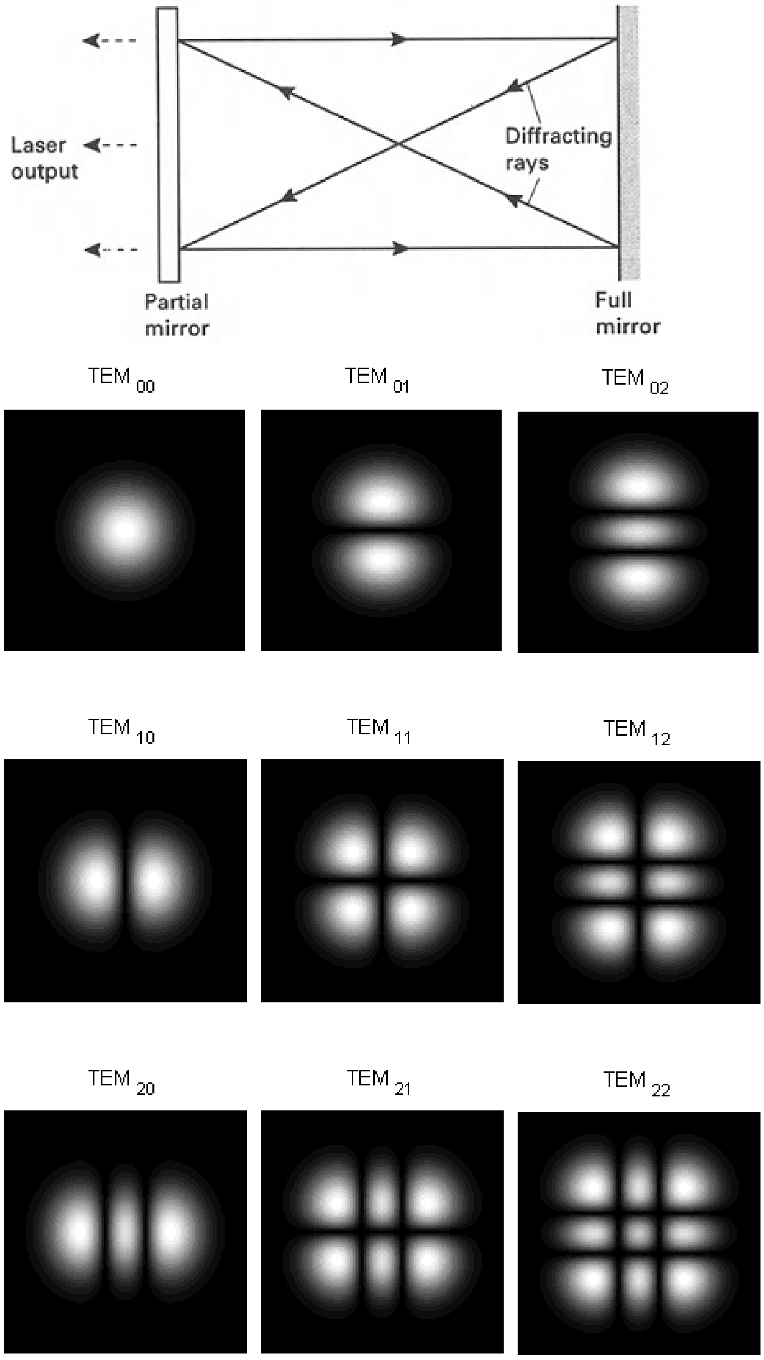


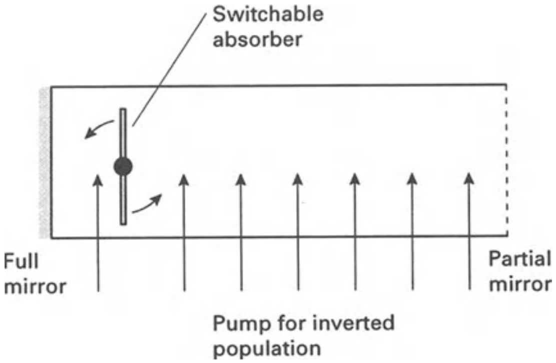
Fig. 6.7 Transverse cavity modes.

(iv) *Q-switching*

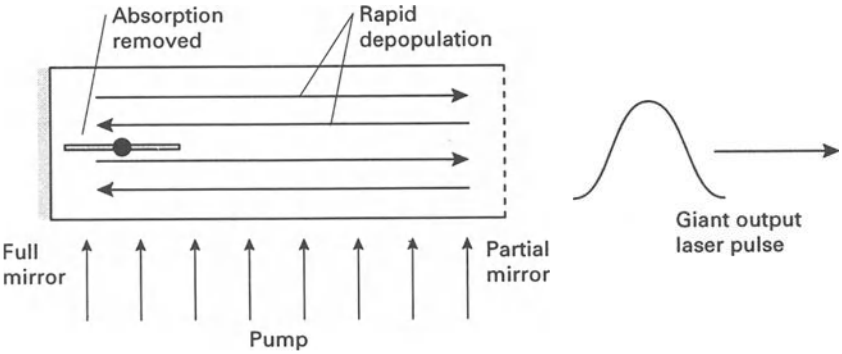
The 'Q' or 'quality factor' of an oscillator refers to its purity, or 'sharpness of resonance'. The lower the loss in an oscillator the narrower is its resonance peak and the longer it will oscillate on its own after a single driving impulse. The equivalent quantity in a Fabry-Perot cavity (an optical oscillator) is the 'finesse' (see section 2.9) and the two quantities are directly related. From these ideas we can readily understand that if the loss in a resonator is varied then so is its 'Q'.

Suppose we have a laser medium sitting in its usual Fabry-Perot cavity but with a high loss; this means that a large fraction of the light power oscillating between the mirrors is lost per pass: we might, for example, have one of the mirrors with very low reflectivity.

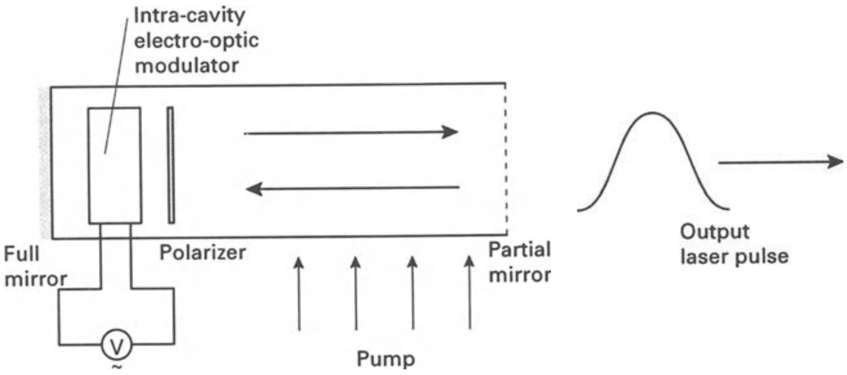
Now the oscillator can only oscillate if the gain which the light receives per double pass between the mirrors exceeds the loss per double pass (section 6.4(ii)), and we shall suppose that the loss is very high, so that as we pump more and more molecules of the medium up into the excited state of the inverted population, the loss still exceeds the gain for as hard as our pump source can work. The result is that the inversion of the population becomes very large indeed, for there are very few photons to cause stimulated emission down to the lower state - they are all being lost by other means (e.g., a poor mirror at one end). Having achieved this very highly inverted population suppose that the loss is now suddenly reduced by means of an intercavity switch ('Q' switch) by, for example, speedily



(a) Inversion of the atomic population before lasing



(b) Rapid depopulation of inverted states on removing absorption



(c) A practical Q-switched laser cavity

Fig. 6.9 Q-switching.

rotating to a high-reflectivity mirror (Fig. 6.9). The result is that there is suddenly an enormous number of photons to depopulate the inverted population, which then rapidly de-excites to emit all its accumulated energy in one giant laser pulse - the Q-switched pulse. In this way is obtained the means by which very large energy, very high intensity pulses can be obtained, albeit relatively infrequently ($\sim$ 25 pps).

Three further points should be noted concerning Q-switching:

(a) At the end of the pulse the lasing action ceases completely, since the large number of photons suddenly available completely depopulates the upper laser state.

(b)The switching to the low loss condition must take place in a time which is small compared with the stimulated depopulation time of the upper state, so as to allow the pulse to build up very quickly.

(c)The pumping rate must be large compared with the spontaneous decay rate of the upper state so as to allow a large population inversion to occur.

Q-switching can produce pulses with several millijoules of energy with only a few nanoseconds duration. Thus, peak powers of several megawatts can result. Such powers take most media into their non-linear regimes (many will be evaporated!), so Q-switching is very useful for studying the non-linear optical effects which will be considered in Chapter 9.

Both mode-locking and Q-switching require intracavity modulation devices. These can take a variety of forms, and in Chapter 7 we shall be considering these.

7.1 INTRODUCTION

Chapter 6 has provided the background necessary to understand a variety of optoelectronic devices, the most important of these being optical sources, modulators and detectors. The source provides the light in the appropriate form, the modulator controls it or impresses information upon it, and the detector receives the light signal and allows that information to be extracted in the form of an electronic signal.

Most optoelectronic devices nowadays are of the solid state variety, for solid devices can be compact, robust, reliable and readily manufacturable (and, therefore, cheap!). However, there are still some important gas lasers, and we must not ignore these.

In this chapter optical sources, modulators and detectors will be dealt with in turn. The emphasis will be on practical aspects of the devices, so that the way in which the basic principles are put into practice will begin to become clear. We shall begin with optical sources.

7.2 OPTICAL SOURCES

The requirements demanded of optical sources by optoelectronic researchers, designers and users are many and varied. The source may be required to be broadband or narrowband; tunable or fixed in frequency; coherent, partially coherent or incoherent; polarized or unpolarized; continuous wave (CW) or pulsed; divergent or collimated. It is hardly surprising that as no single type of source can provide all of these features, a range of sources has been developed for optoelectronic use. One way of providing a quite versatile source is to use a broadband black body, or 'grey' body, source and then to use a variety of external components and devices to manipulate the light in order to provide what is needed. For example, filters can select a limited wavelength range, and these can be made tunable if necessary via frequency-selective components, such as prisms, diffraction gratings or Fabry-Perot interferometers. The polarization can be selected with the aid of polarizing prisms and retardation plates, the collimation controlled with lenses, and pulses provided by fast electro-optic switches.

There are two objections to this approach. Firstly, each one of the selective operations is lossy and significantly reduces the available power; secondly, the source system becomes complex, cumbersome, difficult to align and stabilize, heavy, bulky and expensive.

The efforts, therefore, have been in the direction of developing sources which are both compact and intense but which also have inherent, desirable characteristics, sometimes controllable sometimes not necessarily so, for particular applications. We have already learned that laser sources have many of the required characteristics, as natural features. Hence our study of optical sources should, naturally, begin with lasers.

7.2.1 Laser sources

(i) *Introduction*

Basic laser design was introduced in Chapter 4 but it is necessary to expand somewhat on what was considered there.

Earlier (equation (6.14)) we derived the expression for the gain which was obtained for unit length in a medium pumped to a state of inverted population as:

$$g = \frac{h\nu}{c} B_{12}(N_2 - N_1)$$

This medium can be used as an optical amplifier, since light entering the medium at the frequency corresponding to the inverted transition will cause stimulated emission at the same frequency, and thus will emerge from the medium with increased power. By providing feedback, the amplifier may be made to oscillate and become a laser. All of these ideas were considered in some detail in Chapter 4. In particular, in section 4.4.2, the basic features of He-Ne laser design were considered. This was a three-level system, and it was pointed out that, although the principles of laser action are relatively straightforward, the transition pathways for any particular laser are often complex. We shall now try to cement these ideas by looking at some aspects of practical design for specific lasers.

(ii) *The argon laser: a four-level system*

The first system for consideration is a four-level system. Remember that the advantage of this is that the lower level of the laser transition is not the ground state, hence the level is relatively sparsely populated, inversion

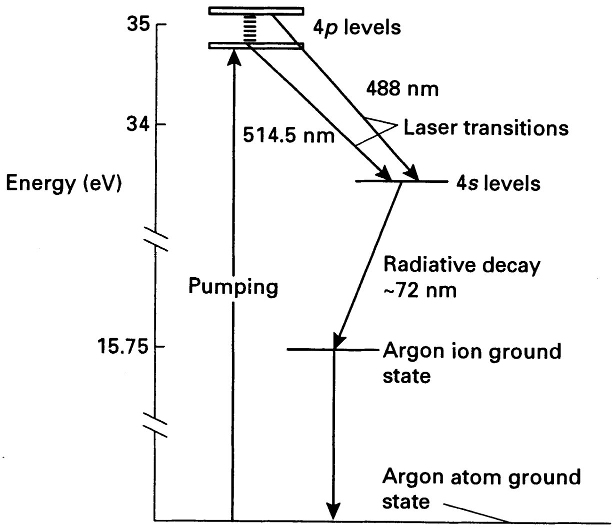


Fig. 7.1 Simplified energy-level diagram for the argon laser.

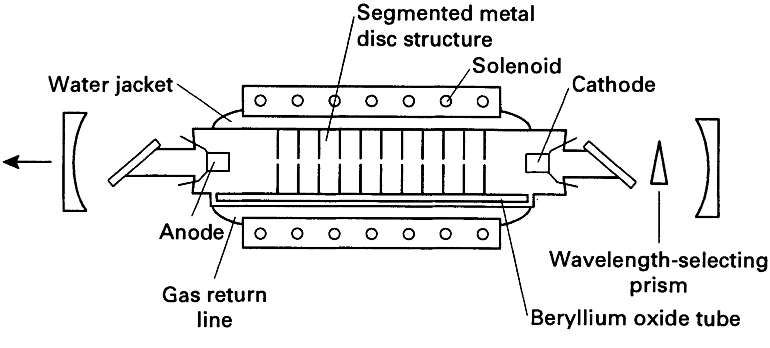


Fig. 7.2 Schematic for the design of an argon laser.

of the population is therefore easier for the pump to achieve, and the laser can thus be more efficient.

The energy level diagram is shown in Fig. 7.1. Argon is an inert gas but can be ionized by passing a large current ($\sim$ 50 A) through it (by striking a gas discharge). In its ionized state the atom is then excited up to the 4p levels by successive electron collisions, and these levels become inverted with respect to the 4s state, which is sparsely populated, being above the Ar^+ ground state, which is itself above the neutral Ar ground state. Thus efficient lasing occurs between 4p and 4s, most strongly at 514.5 nm (green) and 488 nm (blue) wavelengths. The design of the laser is shown in Fig. 7.2. Two main features should be noted: first, since the large discharge current generates a lot of heat, this must be dissipated with the aid of metal-disc heat exchangers and water cooling; secondly, in order to reduce ionic de-excitation via wall collisions (rather than the required laser transitions) the tube is enclosed within a solenoid, which provides a longitudinal magnetic field whose action is to force the ions away from the walls.

Finally, a prism is usually provided inside the cavity, to select the required wavelength.

Argon-ion lasers can provide several watts of CW power and can also be both mode-locked and Q-switched, conditions which were discussed in Chapter 6. Mode-locking provides microsecond pulses at several kilowatts of peak power, whilst Q-switching provides megawatts of peak power (for this particular laser).

Each type of laser presents its own design problems as a result of its own special set of energy-loss pathways.

The argon laser is a valuable high-power source of CW light in the middle of the visible spectrum.

(iv) *The Nd-YAG laser: a solid state system*

There are many types of laser which use solid media. We shall give one example now and another when we have, very shortly, become familiar with some more of the relevant solid state physics.

The present example is the neodymium/yttrium/aluminium garnet, or Nd-YAG, laser.

The big advantage, as has been stated, of a solid state laser is that the large density of atoms means that large outputs per unit volume can be achieved. The laser medium is in the form of a Nd-YAG rod. Nd is a rare earth element. The important characteristic of the rare earth elements is that their atoms all have the same outer electronic structure. As the atomic number increases, moving up the rare earth series, electrons are added to inner, rather than outer, levels. The result is that all the rare earths have very similar chemical properties. From our present point of view, however, the most important thing is that transition between levels can occur deep in the atom, and thus these transitions are shielded, by the outer electrons, from the atom's environment. Hence, the Nd^{3+} ion can sit in a YAG lattice (substituting for the yttrium atoms to the extent of $\sim 1.5\%$) and the transitions to be used for lasing action will be scarcely broadened at all by the fierce fields of the solid lattice structure.

The levels used are shown in Fig. 7.5. The system again is a four-level one and thus is quite efficient. The levels above $4F_{3/2}$ all, conveniently, decay to $4F_{3/2}$, so this level can be readily populated (level 2) by a broad-spectrum pump source, which is usually a krypton or xenon flashtube. This is positioned axially, parallel with the Nd-YAG rod, as shown in Fig. 7.6. The laser transition is from $4F_{3/2}$ (level 2) to $4I_{11/2}$ (level 1) which, being above the ground state (level 0), is very sparsely populated. (It is interesting to note that $F \rightarrow I$ transitions are normally highly forbidden, but the action of the YAG lattice field is to increase their probability in this case). The $1 \rightarrow 0$ transition ($4I_{11/2} \rightarrow 4I_{9/2}$) is allowed and is thus very rapid, as required for efficient population inversion (section 6.2.4). Since the depopulation of the inverted-population state is very rapid (a consequence of the dense material) and, provided that a suitably large

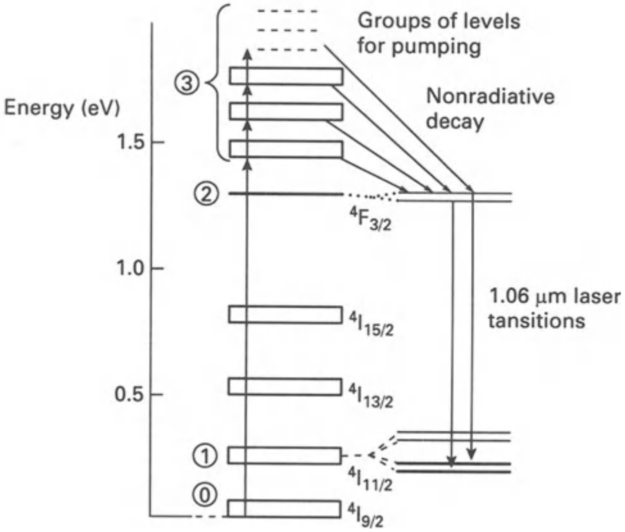
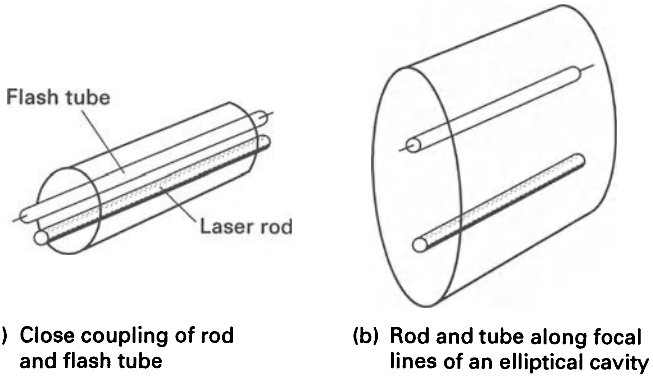


Fig. 7.5 Simplified energy-level diagram for the Nd-Yag laser.



(a) Close coupling of rod and flash tube

(b) Rod and tube along focal lines of an elliptical cavity

Fig. 7.6 Flash-tube pumping geometries for a Nd-Yag laser rod.

amount of pump power is provided, the output can be up to a kilowatt of CW power at 1064 nm ($4F_{3/2} \rightarrow 4I_{11/2}$). The Nd-YAG rod becomes very hot while generating all this power, and must be cooled either with water (for the higher output powers) or with air.

This laser can also readily be Q-switched (section 6.2.4(iv)), in which state it can provide pulses with energies of several millijoules in a time $\sim$ 10 ns, implying peak pulse powers $\sim$megawatts. The Nd-YAG laser is thus a convenient source of high-energy pulses in the near infrared. The pulses are relatively short and thus the peak power of the pulses is very high. We shall see later that this is the prescription for ready entry into the non-linear optical regime where, for example, harmonics of the fundamental optical frequency can be conveniently generated. The Nd-YAG laser also can be used for manufacturing processes such as laser cutting, welding and laser forming of metals.

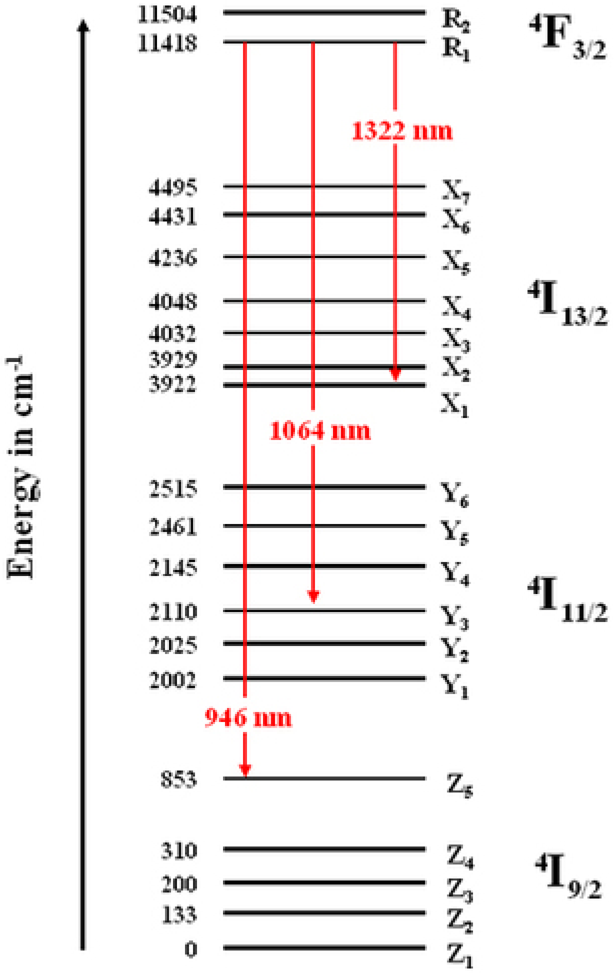
(v) *Other types of laser*

There are many other types of laser: for example, the CO_2 laser which provides large amounts of power in the infrared; the excimer (*exci*ted di*mer*) laser which provides large amounts of power in the ultraviolet; the Er-doped fibre laser which provides a very convenient, lower-power source for optical-fibre communications (see Chapter 10); the Ti-sapphire laser which can provide very short, intense pulses of light ($\sim 100\,\mathrm{ps}$), etc. Each has its own special advantages and disadvantages. Hence each has its own application area.

We have neither the time nor the space to deal with the full range of lasers which are currently in use. However, there is one more which is of paramount importance to our subject: the semiconductor laser. This will be dealt with very soon now (section 7.2.2), just after dealing with its close companion the light-emitting diode (LED).

SOME NOTES ON Nd:YAG LASER LEVELS

The energy levels of the neodymium ion Nd^{3+} are represented by a group of letters and numbers that give the quantum numbers associated with different components: the letter corresponds to the orbital quantum number, the superscript number gives the spin quantum number and the subscript fraction is the angular quantum number. Because of the crystal field (Stark effect), the energy levels are split into sublevels, represented by letters with subscripts (Z_1...R_2).



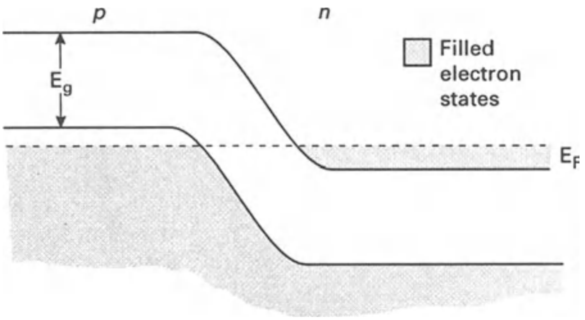
The neodymium ions only stay a long time in the level $^4F_{3/2}$. The lifetime of this level is around 230 μs while it is less than a nanosecond in the other levels. Thus, the ions accumulate in this level and can fall from it by intense stimulated emission.

(iii) *The semiconductor laser diode (SLD)*

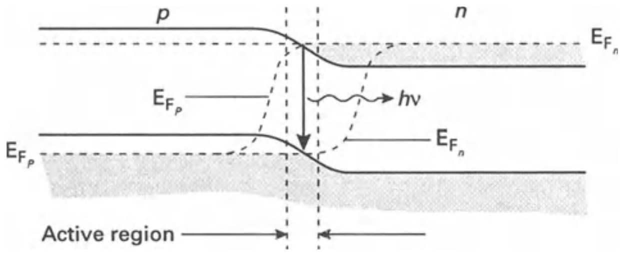
The basic principles on which the semiconductor laser diode is based are very similar to those just discussed for the LED. Forward-carrier injection is again used. However, there are also important differences.

When a $p-n$ junction is created from the two types of extrinsic semiconductor material, the initial condition, with one Fermi level (n-type) lying above the other (p-type), corresponds quite closely to the inverted population which we know is required for optical amplification and, with suitable feedback, lasing action. Obviously, this state is a non-equilibrium one and does not last long (perhaps a few nanoseconds) but we have also learned that it is possible to maintain one Fermi level above the other by applying a forward-bias voltage across the junction. In this case we are, effectively, 'pumping' with the external source of current to maintain an inverted population, of the n-states of the conduction band over the p-states in the valence band, in the volume of material around the junction. The electron energy transitions from the conduction and to the valence band either can occur spontaneously, in which case we have an LED, or can be stimulated to emit by incoming photons of the same frequency E_g/h in which case we have a semiconductor optical amplifier. Applying positive feedback to this latter arrangement can be expected, under the correct circumstances, to produce a semiconductor laser. What are these correct circumstances?

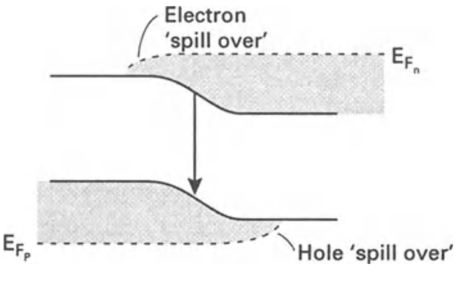
First, it will be necessary to produce a large population inversion in order to provide a high gain (equation 6.14), since the losses in a solid material are expected to be quite high, owing to scattering by the dense medium, and other factors. In order to achieve this we should ensure that the dopant levels are high, so that there are large concentrations of electrons and holes in the depletion region. In fact, for SLD devices, dopant levels (e.g., excesses of Ga or As atoms above 'stoichiometric' values) up



(a) Energy level diagram for an unbiased 'degenerate' p-n junction



(b) Forward-biased degenerate p-n junction leading to stimulated emission



(c) Excess of forward bias, leading to lossy 'spill over'

Fig. 7.8 The degenerate $p - n$ junction.

to ~ 0.01% (equivalent to ~ 10^{24} atoms m^{-3}) are used to give rise to what are known as 'degenerate' semiconductor donor and acceptor levels. So great, in fact, are the donor and acceptor densities that very many of the available levels at the bottom of the conduction band are filled with electrons, and very many of those at the top of valence band are occupied by holes. The consequence of this is that the two Fermi levels can move into the conduction band (n-type) and the valence band (p-type) respectively, since the Fermi level refers to that energy level whose states are just 50% occupied. So great, in other words, are the extra quantities of charge carriers that we are approaching the condition which applies to metallic conductors (i.e., Fermi level halfway up the conduction band). In thermal equilibrium the energy level diagram for the junction will take the form shown in Fig. 7.8(a). If, now, the junction is forward-biased with a voltage of the order of the band-gap energy itself, the diagram changes to that of Fig. 7.8(b). In this case what has happened is that the region in which there is an excess of conduction-band electrons in the n-type material, and the region in which there is an excess of valence-band holes in the p-type material have moved closer to the point of overlap, so that the electrons can fall into the holes and create photons of frequency E_g/h; and they will do this especially readily if a direct-band-gap semiconductor, such as GaAs, is being used.

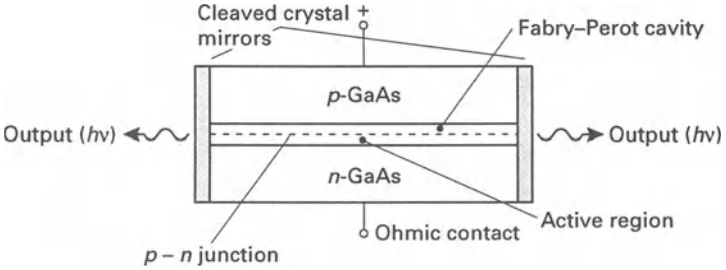


Fig. 7.9 Homojunction GaAs injection laser design.

As the large numbers of electrons and the holes combine they create correspondingly large numbers of photons, and the electrons and holes are replaced constantly by drawing current from the biasing-voltage source (just as for the LED), so light is generated continuously. We have created an effective population inversion in the semiconductor, the inversion being 'pumped' by the injected current.

In order to get the forward-biased $p-n$ junction to lase it is necessary, as we know from section 6.2.4, to provide positive feedback. This is done by arranging the geometry of the junction as shown in Fig. 7.9 and polishing two opposite crystal faces (facets) of the material so that they act as laser mirrors, via the Fresnel reflection coefficient. Provided that the optical gain exceeds the loss, the device will lase by emitting radiation, at the frequency corresponding to the band gap, in a direction normal to the polished facets.

Clearly, the population inversion, and thus the optical gain, can be increased by increasing the bias voltage. This increases the drift current with respect to the diffusion current and thus more electrons combine with holes, leading to more photons. Eq. (6.14b) tells us when the system will lase. If the reflectivities of the two facet mirrors are each equal to R, then lasing occurs when

$$R\exp(\gamma - \alpha)l \geq 1 \tag{7.1}$$

where γ is the gain, α the loss and l the distance between the mirrors. So we have to increase the voltage until lasing occurs. The current density which flows at the onset of lasing is known as the 'threshold' current density and, for the type of geometry just described, is quite high, at $\sim 400\mathrm{A\,mm}^{-2}$. This is due to the fact that the losses are quite high. What are these losses? First, there are of course the losses at the reflecting surfaces, represented by R in equation (7.1). These are moderately high since we are relying on the Fresnel reflection (equation (2.13a)) which results from the difference in refractive index between GaAs and air (at the lasing frequency). This is of the order of 3.35, leading to a value for $[(n_1 - n_2)/(n_1 + n_2)]^2$ of 0.292. This can be increased by using coatings of various forms or by using Bragg reflections (of which more later (Chapter 10)) but equation (7.1) tells us that we shall, in any case, achieve no more than a linear reduction in the threshold current with R, whereas, for a major improvement, we would be better advised to concentrate on the exponential term. It turns out that γ and α are not independent, for, if the bias voltage (and thus γ) is increased, the Fermi levels on each side rise and fall, respectively (Fig. 7.8(c)) to the point where majority carriers on each side can 'spill over' into empty levels in the bulk material, and thus diffuse away from the 'junction' transition region to be lost from the photon-generating process.

Another important source of loss lies in the fact that the difference in refractive index between the junction region and the rest of the semiconductor is quite small. Consequently, the light bouncing back and forth between the facet mirrors within the junction region (Fig. 7.9) is not well confined, laterally, to this region, and when propagating outside the region will simply be absorbed by exciting some valence band electrons up to conduction band levels, a process known as 'free-carrier absorption'.

All of these losses mean that large injection currents are required to overcome them, so the threshold current for lasing is, as has been noted, quite high. The high value of threshold current means, of course, that the device consumes a lot of power, and thus will suffer a potentially ruinous rise in temperature unless operated only in short pulses. This is a severe limitation: we need continuous sources of optical power in addition to pulsed sources. How can the threshold current be lowered?

We have understood the sources of loss in the junction device: this usually means that we should be able to see a way to reduce them. Proper understanding is our most powerful weapon in physics and in optoelectronic design.

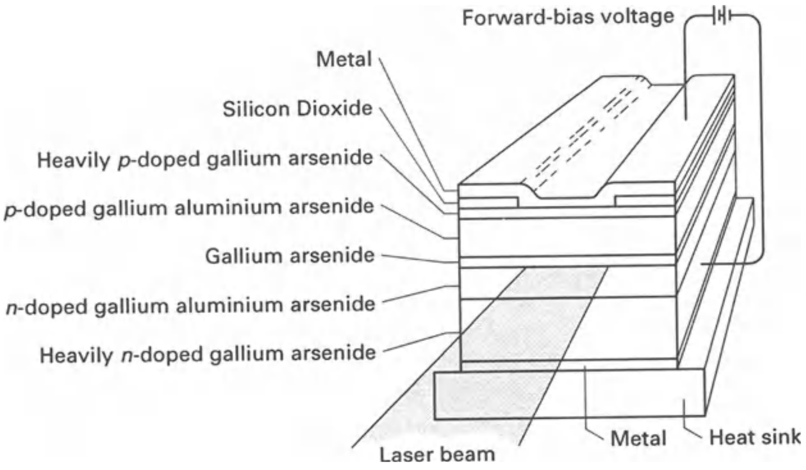
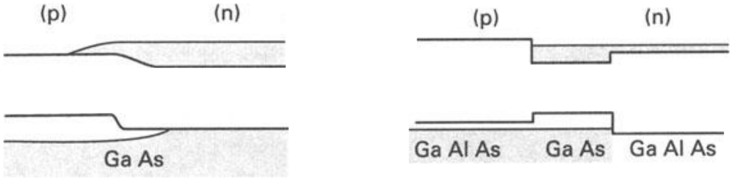


Fig. 7.10 Design for a GaAs heterojunction laser.

Consider, then, the structure shown in Fig. 7.10. Here we have a multilayered structure known as a heterojunction laser, to distinguish it from the type of device we have been considering up to now, which is called a homojunction laser.



(a) Homojunction (b) Heterojunction

Fig. 7.11 Carrier distributions in types of $p - n$ junction.

In the heterojunction laser each layer has a well-defined function (Fig. 7.10). The silicon dioxide simply isolates the semiconductor from the electrical contacts; the heavily p/n-doped layers on each side act as 'ohmic' semiconductors which interface crystallographically with the 'junction' p and n degenerate GaAlAs layers which themselves sandwich an all-important thin layer of GaAs. This layer is only $\sim$ 200 nm thick and, since it has a higher refractive index than the surrounding GaAlAs layer, acts to confine the optical radiation within itself. One of the loss problems is thus alleviated. But the thin GaAs layer performs another important function. A study of Fig. 7.11 illustrates this. Since GaAs has a smaller energy gap than GaAlAs, the majority carriers can no longer spill over into the material on the other side of the gap. Thus loss by carrier diffusion is all but eliminated. Such a design provides laser threshold currents $\sim 10\,\mathrm{A\,mm^{-2}}$, an improvement more than an order of magnitude over the homojunction, and it allows CW operation to be achieved. The heterojunction semiconductor laser was an important breakthrough in optoelectronics for it provided a source of light, pulsed or continuous, which was rugged, compact, coherent, intense, monochromatic, operated with a low voltage, and was easily mass produced (and therefore cheap). Consumer optoelectronics really came of age with this device's availability (which is why it has been considered in some detail).

The GaAs version of the device provided laser light at 840 nm wavelength, but band-gap engineering (see section 6.3.5) allows a range of wavelengths including those suitable for optical-fibre communications. This is a very convenient, versatile and commercially important source of light. It is used, for example, in CD players (see Chapter 10), in supermarket bar-code readers and in many display functions. This most important light source is where we shall finish the discussion of sources of light and move on to consider how it is possible to impress information on the light which they provide.

5.10.2.1 *Threshold current density for semiconductor lasers*

The onset of laser action at the threshold current density is detected by an abrupt increase in the radiance of the emitting region, as shown in Fig. 5.25, which is accompanied by a dramatic narrowing of the spectral width of emission. This is illustrated very clearly in Fig. 5.26 which shows the mode structure below, and at threshold, where the energy has been channelled into a relatively small number of modes. If the current is increased substantially above threshold one mode usually predominates, with a further decrease in the spectral width of the emission.

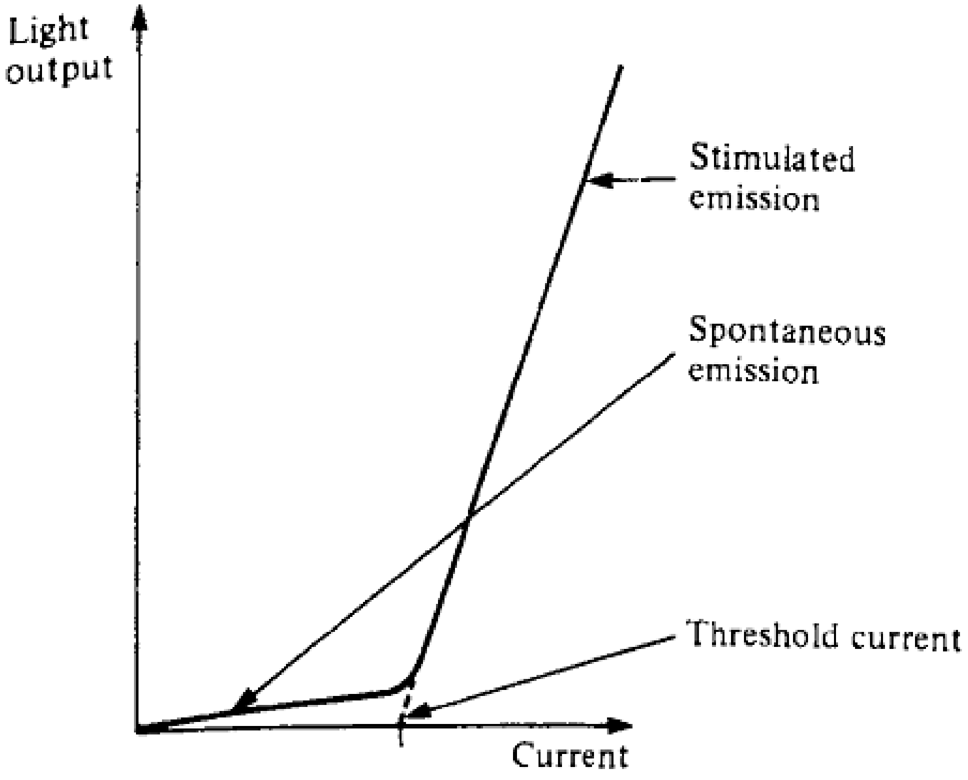


FIG. 5.25 Light output–current characteristic of an ideal semiconductor laser.

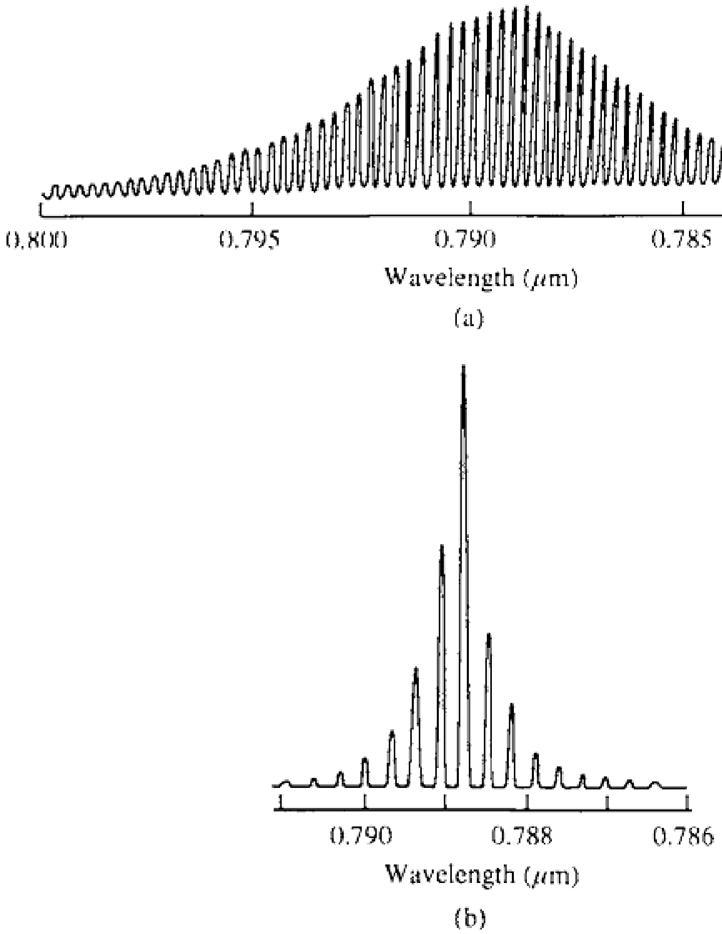


FIG. 5.26 Emission spectrum of a GaAlAs laser diode both just below (a) and just above (b) threshold. Below threshold a large number of Fabry–Perot cavity resonances can be seen extending across a wide LED-type spectrum. Above threshold only a few modes close to the peak of the gain curve oscillate. For the particular laser shown here the threshold current was 37 mA while spectra (a) and (b) were taken with currents of 35 mA and 39 mA, respectively.