4 STIMULATED EMISSION DEVICES: OPTICAL AMPLIFIERS AND LASERS

4.1 STIMULATED EMISSION, PHOTON AMPLIFICATION, AND LASERS

A. Stimulated Emission and Population Inversion

An electron in an atom can be excited from an energy level E_1 to a higher energy level E_2 by the absorption of a photon of energy $h\upsilon = E_2 - E_1$ as shown in Figure 4.1 (a). When an electron at a higher energy level transits down in energy to an unoccupied energy level, it emits a photon. There are essentially two possibilities for the emission process. The electron can undergo the downward transition by itself quite *spontaneously*, or it can be *induced* to do so by another photon.

In **spontaneous emission**, the electron falls down in energy from level E_2 to E_1 and emits a photon of energy $h\upsilon = E_2 - E_1$ in a random direction as indicated in Figure 4.1 (b). The transition is spontaneous provided that the state with energy E_1 is empty, *i.e.,* not already occupied by another electron. In classical physics, when a charge accelerates and decelerates as in an oscillatory motion with a frequency υ it emits an electromagnetic radiation also of frequency υ. The emission process during the transition of the electron from E_2 to E_1 can be classically thought of as if the electron is oscillating with a frequency υ.

In **stimulated emission**[2], an incoming photon of energy $h\upsilon = E_2 - E_1$ stimulates the whole emission process by inducing the electron at E_2 to transit down to E_1. The emitted photon has the *same energy*, since $h\upsilon = E_2 - E_1$, it is *in phase* with the incoming photon, it is in the *same direction*, and it has the *same polarization*, that is, the optical field oscillations in the emitted and in the incoming photon are in the same direction, as illustrated in Figure 4.1 (c). To get a feel of what is happening during stimulated emission one can think of the electric field of the incoming photon coupling to the electron and thereby driving it with the same frequency as the incoming photon. The forced oscillation of the electron at a frequency $\upsilon = (E_2 - E_1)/h$ causes it to emit electromagnetic radiation whose electric field is in total phase with that of the stimulating radiation. When the incoming photon leaves the site, the electron has returned to E_1 because it has emitted a photon of energy $h\upsilon = E_2 - E_1$. Although we considered the

[2] Also called *induced emission.*

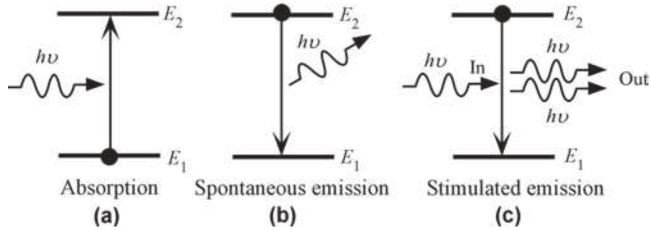


FIGURE 4.1 Absorption, spontaneous (random photon) emission, and stimulated emission.

transitions of an electron in an atom, we could have just well described photon absorption, and spontaneous and stimulated emission in Figure 4.1 in terms of energy transitions of the atom itself, in which case E_1 and E_2 represent the energy levels of the atom.

Stimulated emission is the basis for obtaining photon amplification since one incoming photon results in two outgoing photons which are in phase. How does one achieve a practical light amplifying device based on this phenomenon? It should be quite apparent from Figure 4.1 (c) that to obtain stimulated emission, the incoming photon should *not* be absorbed by another atom at E_1. When we are considering a collection of atoms to amplify light, we must therefore have the majority of the atoms at the energy level E_2. If this were not the case, the incoming photons would be absorbed by the atoms at E_1. When there are more atoms at E_2 than at E_1, we have what is called a **population inversion**. Under normal equilibrium conditions, as a result of Boltzmann statistics, most of the atoms would be at E_1, and very few at E_2. We therefore need to *excite* the atoms, cause population inversion, to obtain stimulated emission. It should be apparent that with *only two* energy levels, we can never achieve population at E_2 greater than that at E_1 because, in the steady state, the incoming photon flux will cause as many upward excitations as downward stimulated emissions. We need at least three energy levels as shown below in Section B.

B. Photon Amplification and Laser Principles

Let us consider the three-energy-level system shown in Figure 4.2 (a), with levels at E_1, E_2, and E_3. There are, of course, other energy levels but they are not involved in the following processes. The present example will roughly represent what happens in a ruby laser. The energy levels

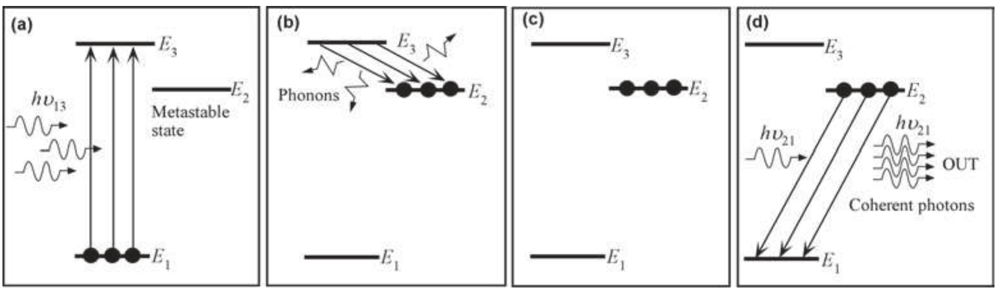


FIGURE 4.2 The principle of the laser, using a ruby laser as an example. (a) The ions (Cr^{3+}) in the ground state are pumped up to the energy level E_3 by photons from an optical excitation source. (b) Ions at E_3 rapidly decay to the long-lived state at the energy level E_2 by emitting lattice vibrations (phonons). (c) As the states at E_2 are long-lived, they quickly become populated and there is a population inversion between E_2 and E_1. (d) A random photon (from spontaneous decay) of energy $h\upsilon_{21} = E_2 - E_1$ can initiate stimulated emission. Photons from this stimulated emission can themselves further stimulate emissions, leading to an avalanche of stimulated emissions and hence coherent photons being emitted.

E_1, E_2, and E_3 correspond to the energies of the chromium ions (Cr^{3+}) in a crystal of Al_2O_3 (a corundum crystal); E_1 is the ground energy of the Cr^{3+} ion. Suppose that an external optical excitation causes the Cr^{3+} ions in the crystal to become excited to the energy level E_3. This is called the pump energy level, and the process of exciting the Cr^{3+} ions to E_3 is called **pumping**. In the present case, **optical pumping** is used, and we need a pump source with a photon energy of $h\upsilon_{13} = E_3 - E_1$.

From E_3, the Cr^{3+} ions decay rapidly to an energy level E_2 by emitting phonons,[3] that is, lattice vibrations (heat), as shown in Figure 4.2 (b). Thus, the transition from E_3 to E_2 is a radiationless transition. The energy level E_2 happens to correspond to a state that does not rapidly and spontaneously decay to the lower-energy state E_1. In other words, the state at E_2 is a **long-lived state**.[4] (The characteristic transition time constant from E_2 to E_1 is about 3 ms, which is very long in the atomic world.) Since the Cr^{3+} ions cannot decay rapidly from E_2 to E_1, they accumulate at this energy level. With ample pumping, we can accumulate sufficient Cr^{3+} ions at E_2 to cause a population inversion between E_2 and E_1, as pumping takes more and more Cr^{3+} ions to E_3 and hence E_2. Population inversion is shown in Figure 4.2 (c).

When a Cr^{3+} ion at E_2 decays spontaneously, it emits a photon (a "random photon") which can go on to a neighboring Cr^{3+} ion and cause that to execute stimulated emission. The photons from the latter can go on to the next Cr^{3+} ion at E_2 and cause that to emit by stimulated emission and so on. The result is an avalanche of stimulated emission processes with all the photons in phase, in the same direction, and with the same polarization, so that the light output is a large collection of *coherent photons* as in Figure 4.2 (d). This is the basic principle of the ruby laser, and hence the three-level laser. At the end of the avalanche of stimulated emission processes, the Cr^{3+} ions at E_2 would have dropped to E_1 and can be pumped again to repeat the stimulated emission cycle. The emission from E_2 to E_1 is called the **lasing emission**.[5] The system we have just described for photon amplification is a laser, an acronym for **Light Amplification by Stimulated Emission of Radiation**.[6]

A more realistic energy diagram for the Cr^{3+} ion in the ruby crystal is shown in Figure 4.3 (a), which is typical of three-level laser systems that involve ions in crystals or glasses. The main difference from the simple schematic in Figure 4.2 is that the higher energy level E_3 is actually a band of energy levels. There is another band of energies around E_3'. The Cr^{3+} ion can be excited by green light to E_3 and by blue light to E_3'. (The ruby crystal is "pink" because it absorbs green and blue light.[7]) Pumping is achieved by using an intense xenon flash light. The ions become excited to the E_3 and E_3' bands and then decay quickly from these bands by emitting phonons. They populate the E_2-level as described above. Once more than half the ions at E_1 have been pumped, there is population inversion between E_1 and E_2 and consequently the pumped medium (the ruby crystal) exhibits **optical gain**, or **photon amplification**.

---

[3] A phonon is a quantum of lattice vibrational energy, just as a photon is a quantum of electromagnetic energy.

[4] We will not examine what causes certain states to be long-lived, but simply accept that these states do not decay rapidly by spontaneous emission to lower-energy states. Put differently, the decay time is very long (typically in milliseconds) on the atomic scale.

[5] *Arthur Schawlow*, one of the coinventors of the laser, was well-known for his humor and has, apparently, said that "Anything will lase if you hit it hard enough." In 1971, Schawlow and Ted Hänsch were able to develop the first edible laser made from Jell-O. (T. Hansch, M. Pernier, A. Schawlow, "Laser Action of Dyes in Gelation," *IEEE J. Quantum Elect.*, *7*, 45, 1971.)

[6] Although it is indeed an acronym, it is now commonly used as a noun that implies a device that emits coherent radiation.

[7] The ruby crystal appears red if the Cr^{3+} ions concentration is large. For the Cr^{3+} concentrations used in the ruby laser, the crystal appears pink.

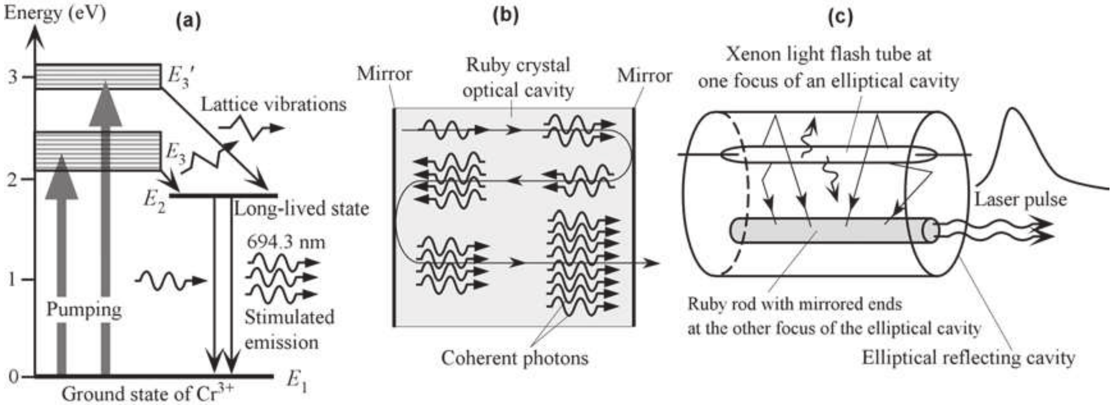


FIGURE 4.3 (a) A more realistic energy diagram for the Cr^{3+} ion in the ruby crystal (Al_2O_3), showing the optical pumping levels and the stimulated emission. (b) The laser action needs an optical cavity to reflect the stimulated radiation back and forth to build up the total radiation within the cavity, which encourages further stimulated emissions. (c) A typical construction for a ruby laser, which uses an elliptical reflector, and has the ruby crystal at one focus and the pump light at the other focus.

To get coherent lasing radiation out from the ruby laser, we need to do more, not just amplify the radiation along the ruby crystal. We can increase stimulated emissions by increasing the number of photons, *i.e.,* the radiation intensity, within the crystal inasmuch as more photons cause more stimulated emissions. The ends of the ruby crystal, which is normally a rod, are silvered to reflect back and forward the stimulated radiation, that is, to form an **optical cavity** with mirrors at the ends, as shown Figure 4.3 (b). As the stimulated photons are reflected back into the crystal, the radiation intensity buildsup inside the crystal, in much the same way we build up voltage oscillations in an electrical oscillator circuit by feedback. The buildup of coherent radiation in the cavity encourages further stimulated emissions, until a large avalanche of stimulated transitions occur and takes most of the ions at E_2 down to E_1. One of the mirrors is partially silvered to allow some of this radiation to be tapped out. What comes out is a pulse of highly coherent radiation that has a high intensity as illustrated in Figure 4.3 (b).

The coherency and the well-defined wavelength of the emitted radiation from a laser are attributes that make it distinctly different from a random stream of different wavelength photons emitted from a tungsten bulb, or randomly phased photons from an LED. The reader might have noticed that the photon energy emitted from the laser system is less than the photon energy we used to pump it, that is, excite it; $h\upsilon_{21} < h\upsilon_{13}$. However, we only needed incoherent radiation to pump the system, and we obtained a fully coherent radiation as output.

Practical ruby lasers need efficient optical pumping, which can be obtained by using an elliptical reflector with the ruby crystal rod at one focus, and the pump light, a xenon flash, at the other focus as shown in Figure 4.3 (c). The early ruby lasers used a helical xenon flash tube surrounding the ruby rod. The lasing emission from the ruby laser is a light pulse, whose duration and intensity depend on the laser construction, and the xenon flash. Ruby lasers are frequently used in interferometry, holography, and hair and tattoo removal, among other applications.

Theodore Harold Maiman was born in 1927 in Los Angeles, son of an electrical engineer. He studied engineering physics at Colorado University, while repairing electrical appliances to pay for college, and then obtained a Ph.D. from Stanford. Theodore Maiman constructed this first laser in 1960 while working at Hughes Research Laboratories (T.H. Maiman, "Stimulated optical radiation in ruby lasers", *Nature*, **187**, 493, 1960). There is a vertical chromium ion doped ruby rod in the center of a helical xenon flash tube. The ruby rod has mirrored ends. The xenon flash provides optical pumping of the chromium ions in the ruby rod. The output is a pulse of red laser light. *(Courtesy of HRL Laboratories, LLC, Malibu, California.)*

C. Four-Level Laser System

The problem with the three-level laser system in Figure 4.2 (a) is that to achieve population inversion, we must pump at least half the Cr^{3+} ions at the ground level E_1 to E_2. It would be much better if we could achieve population inversion as soon as we begin exciting the ions at E_1 to E_3 to E_2. We can quickly and easily achieve population inversion if we use a four-level system as in Figure 4.4, where the ground energy is E_0. Figure 4.4 roughly represents the energy levels of Nd^{3+} ions in a $Y_3Al_5O_{11}$ (yttrium aluminate garnate, YAG) crystal, usually written as Nd^{3+}: YAG. In equilibrium, the population of the ions at these energy levels follow the Boltzmann distribution; and nearly all the ions are at E_0 whereas E_1, E_2, and E_3 are mostly empty.

The ions at E_0 are excited by optical pumping to the energy level E_3. In many cases, this is a band of energy levels and is called the **pump band**. From E_3, the ions decay rapidly by phonon emission to E_2, which is a long-lived state with a lifetime in the milliseconds, *i.e.,* the spontaneous decay from E_2 to E_1 is slow. Clearly, as soon as ions begin to populate E_2 we have population inversion between the level E_1 and E_2 inasmuch as E_1 is initially nearly empty. The lasing emission takes place by stimulated emission as ions drop from E_2 to E_1. From E_1, the ions return back to E_0 by the emission of phonons (or spontaneous photons in some cases). We also need the ions at E_1 to decay quickly back to E_0 so that a buildup of ions at E_1 is avoided, and the ions are quickly returned back to E_0 for repumping. In cases where E_1 is within only a few k_BT to E_0, the four-level system behaves almost like the three-level system in Figure 4.2.

The four-level laser system in Figure 4.4 serves as an excellent way to explain the majority of practical lasers. Indeed, most modern lasers are based on a four-level system, or its equivalent. For example, in the Nd^{3+}: YAG laser, the Nd^{3+} ions are excited by using a flash tube or a laser diode (LD) to E_3 and hence to E_2, which is the *upper* laser level. The lifetime of the ions at E_2 is about 230 μs. The *lower* laser level is E_1. The lasing emission from E_2 to E_1 is in the infrared at 1064 nm. The Nd^{3+}:YAG can be operated in the pulsed mode as well as in CW mode. In contrast, the ruby laser is operated in the pulsed mode because of the large power required for pumping the ruby crystal.

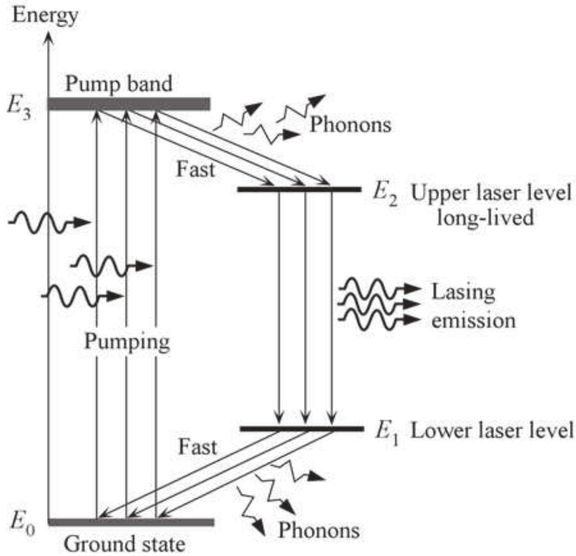


FIGURE 4.4 A four-energy-level laser system. Highly simplified representation of Nd^{3+}:YAG laser.

4.2 STIMULATED EMISSION RATE AND EMISSION CROSS-SECTION

A. Stimulated Emission and Einstein Coefficients

A useful laser medium must have a higher efficiency of stimulated emission compared with the efficiencies of spontaneous emission and absorption. We need to determine the controlling factors for the rates of stimulated emission, spontaneous emission, and absorption. Consider a medium as in Figure 4.5 that has N_1 atoms per unit volume at the energy E_1 and N_2 atoms per unit volume at the energy E_2. Then the rate of upward transitions from E_1 to E_2 by photon absorption will be proportional to the concentration of atoms N_1, and also to the number of photons per unit volume with energy $h\upsilon = E_2 - E_1$. Put differently, this rate will depend on the energy density in the radiation. Thus, the upward transition rate R_{12} is

Upward transitions

$$R_{12} = B_{12}N_1\rho(\upsilon) \qquad \textbf{(4.2.1)}$$

where B_{12} is a proportionality constant termed the **Einstein B_{12} coefficient**, and $\rho(\upsilon)$ is the photon energy density per unit frequency,[8] that is, the energy in the radiation per unit volume

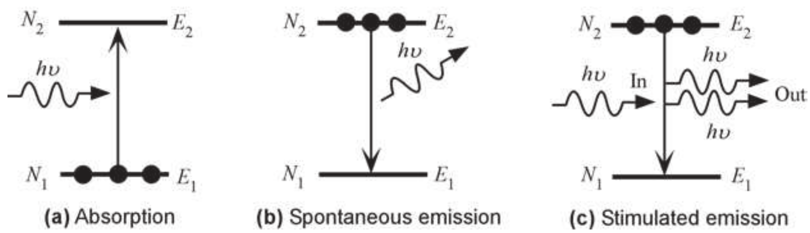


FIGURE 4.5 Absorption, spontaneous emission, and stimulated emission.

[8]Using $\rho(\upsilon)$ defined in this way simplifies the evaluation of the proportionality constants in this section. Further, in the simplified treatise here, we neglected the degeneracy of the energy levels, which introduces numerical factors that multiply N_1 and N_2.

per unit frequency due to photons with energy $h\upsilon = E_2 - E_1$. The rate of downward transitions from E_2 to E_1 involves spontaneous and stimulated emission. The spontaneous emission rate depends on the concentration N_2 of atoms at E_2. The stimulated emission rate depends on both N_2 and the photon concentration $\rho(\upsilon)$ with energy $h\upsilon(= E_2 - E_1)$. Thus, the total downward transition rate is

$$R_{21} = A_{21}N_2 + B_{21}N_2\rho(\upsilon) \tag{4.2.2}$$

Downward transitions

where the first term is due to spontaneous emission [does not depend on the photon energy density $\rho(\upsilon)$ to drive it] and the second term is due to stimulated emission, which requires photons to drive it. A_{21} and B_{21} are the proportionality constants termed the **Einstein coefficients** for spontaneous and stimulated emissions, respectively.

To find the coefficients A_{21}, B_{12}, and B_{21}, we consider the events in equilibrium, that is, the medium in thermal equilibrium (no external excitation). There is no net change with time in the populations at E_1 and E_2 which means

$$R_{12} = R_{21} \tag{4.2.3}$$

Furthermore, in thermal equilibrium, Boltzmann statistics demands that

$$N_2/N_1 = \exp\left[-(E_2 - E_1)/k_BT\right] \tag{4.2.4}$$

Boltzmann statistics in equilibrium

where k_B is the Boltzmann constant and T is the absolute temperature.

Now in *thermal equilibrium*, in the collection of atoms we are considering, radiation from the atoms must give rise to an equilibrium photon energy density, $\rho_{eq}(\upsilon)$, that is given by *Planck's black body radiation distribution law*[9]

$$\rho_{eq}(\upsilon) = \frac{8\pi h\upsilon^3}{c^3\left[\exp\left(\dfrac{h\upsilon}{k_BT}\right) - 1\right]} \tag{4.2.5}$$

Planck's black body radiation in equilibrium

It is important to emphasize that Planck's law in Eq. (4.2.5) applies only in thermal equilibrium; we are using this equilibrium condition to determine the Einstein coefficients. During the laser operation, of course, $\rho(\upsilon)$ is not described by Eq. (4.2.5); in fact it is much larger. From Eqs. (4.2.1) to (4.2.5) we can readily show that

$$B_{12} = B_{21} \tag{4.2.6}$$

Absorption and stimulated emission coefficients

and

$$A_{21}/B_{21} = 8\pi h\upsilon^3/c^3 \tag{4.2.7}$$

Spontaneous and stimulated emission coefficients

Now consider the ratio of stimulated to spontaneous emission

$$\frac{R_{21}(\text{stim})}{R_{21}(\text{spon})} = \frac{B_{21}N_2\rho(\upsilon)}{A_{21}N_2} = \frac{B_{21}\rho(\upsilon)}{A_{21}} \tag{4.2.8}$$

[9] See, for example, any modern physics textbook.

which, by Eq. (4.2.7), can be written as

Stimulated to spontaneous emission

$$\frac{R_{21}(\text{stim})}{R_{21}(\text{spon})} = \frac{c^3}{8\pi h \upsilon^3}\rho(\upsilon) \tag{4.2.9}$$

In addition, the ratio of stimulated emission to absorption is

Stimulated to absorption

$$\frac{R_{21}(\text{stim})}{R_{12}(\text{absorp})} = \frac{N_2}{N_1} \tag{4.2.10}$$

There are two important conclusions. For stimulated photon emission to exceed photon absorption, by Eq. (4.2.10), we need to achieve **population inversion**, that is, $N_2 > N_1$. For stimulated emission to far exceed spontaneous emission, by Eq. (4.2.9) we must have a large photon concentration, which is achieved by building an **optical cavity** to contain the photons.

The rate R_{12} represents the rate at which N_1 is decreasing by absorption, thus, $R_{12} = -dN_1/dt$. The rate R_{21} represents the rate at which N_2 is decreasing by spontaneous and stimulated emission, thus, $R_{21} = -dN_2/dt$. Further, the rate at which N_2 changes by spontaneous emission can be written as

Spontaneous decay lifetime

$$dN_2/dt = -A_{21}N_2 = -N_2/\tau_{\text{sp}} \tag{4.2.11}$$

where, by definition, $\tau_{\text{sp}} = 1/A_{21}$, and is termed the **spontaneous decay time**, or more commonly called the **lifetime** of level E_2.

It is important to point out that the population inversion requirement $N_2 > N_1$ means that we depart from thermal equilibrium. According to Boltzmann statistics in Eq. (4.2.4), $N_2 > N_1$ implies a *negative temperature*! The laser principle is based on **non-thermal equilibrium**.[10]

The number of states an ion or an atom has at a given energy level is called its **degeneracy** g. We assumed that the degeneracy g_1 and g_2 of E_1 and E_2 are the same. We can, of course, redo the treatment above by including g_1 and g_2. The final result is equivalent to replacing N_1 by $N_1(g_2/g_1)$.

EXAMPLE 4.2.1 Minimum pumping power for three-level laser systems

Consider the three-level system as shown in Figure 4.2 (a). Assuming that the transitions from E_3 to E_2 are fast, and the spontaneous decay time from E_2 to E_1 is τ_{sp}, show that the *minimum* pumping power $P_{p\min}$ that must be absorbed by the laser medium per unit volume for population inversion ($N_2 > N_1$) is

Minimum pumping for population inversion for three-level laser

$$P_{p\min}/V = (N_0/2)h\upsilon_{13}/\tau_{\text{sp}} \tag{4.2.12}$$

where V is the volume, N_0 is the concentration of ions in the medium, and hence at E_0 before pumping. Consider a ruby laser in which the concentration of Cr^{3+} ions is 10^{19} cm^{-3}, the ruby crystal rod is 10 cm long and 1 cm in diameter. The lifetime of Cr^{3+} at E_2 is 3 ms. Assume the pump takes the Cr^{3+} ions to the E_3-band in Figure 4.3 (a), which is about 2.2 eV above E_0. Estimate the minimum power that must be provided to this ruby laser to achieve population inversion.

[10] "But I thought, now wait a minute! The second law of thermodynamics assumes thermal equilibrium. We don't have that!" *Charles D. Townes* (born 1915; Nobel Laureate, 1964). The laser idea occurred to Charles Townes, apparently, while he was taking a walk one early morning in Franklin Park in Washington, DC, while attending a scientific committee meeting. Non-thermal equilibrium (population inversion) is critical to the principle of the laser.

Solution

Consider the three-level system in Figure 4.2 (a). To achieve population inversion we need to get half the ions at E_1 to level E_2 so that $N_2 = N_1 = N_0/2$, since N_0 is the total concentration of Cr^{3+} ions all initially at E_1. We will need $[(N_0/2)h\upsilon_{13} \times \text{volume}]$ amount of energy to pump to the E_3-band. The ions decay quickly from E_3 to E_2. We must provide this pump energy before the ions decay from E_2 to E_1, that is, before τ_{sp}. Thus, the *minimum* power the ruby needs to absorb is

$$P_{p\min} = V(N_0/2)h\upsilon_{13}/\tau_{sp}$$

which is Eq. (4.2.12). For the ruby laser

$$P_{p\min} = [\pi(0.5\text{ cm})^2(10\text{ cm})][(10^{19}\text{ cm}^{-3})/2][(2.2\text{ eV})(1.6 \times 10^{-19}\text{ J/eV})]/(0.003\text{ s}) = 4.6\text{ kW}$$

The total pump energy that must be provided in less than 3 ms is 13.8 J.

B. Emission and Absorption Cross-Sections

In many photonic applications, we need to describe absorption and emission in terms of cross-sections. Consider a medium that has been doped with ions, perhaps Nd^{3+} ions in YAG or glass, and consider two energy levels E_1 and E_2, as in Figure 4.6 (a). Suppose that there is a flux of photons Φ_{ph} passing through the medium along a particular direction (x), and the photon energy $h\upsilon = E_2 - E_1$. The flux is the number of photons flowing per unit area along a certain direction so that $\Phi_{ph} = \mathsf{I}/h\upsilon$, where I is the intensity of the beam[11] (radiation energy flow per unit area per unit time). The absorption of radiation, that is, photons, in the medium depends on two factors. First is the number of ions per unit volume, N_1; higher concentration N_1 will mean more absorption. Second is the nature of the interaction of the photon with an individual ion, and is related to the properties of the ion in the medium. The interaction of a photon with an individual ion can be intuitively visualized as the ion having a certain *cross-sectional area*, σ_{ab}, and if the photon

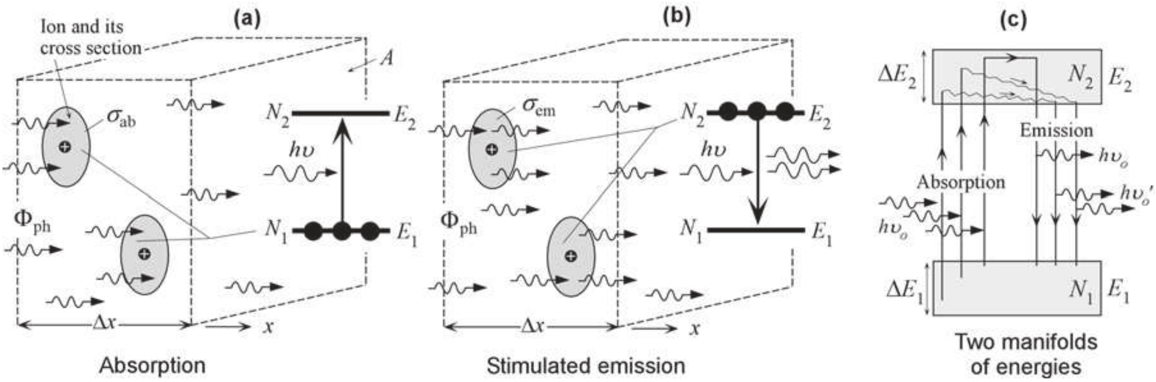


FIGURE 4.6 (a) Absorption of photons by ions in a medium, and the illustration of the absorption cross-section σ_{ab}. (b) Stimulated emission of photons by the same ions in the medium, and the illustration of the emission cross-section σ_{em}. (c) Upper and lower manifolds of energy. Each manifold has numerous closely spaced sublevels. Absorption and emission can involve the same or different levels.

[11] I (san-serif font) is used for the light intensity (optical power flow per unit area, W m^{-2}) and I for the current in this chapter.

crosses this area, it becomes absorbed, as illustrated in Figure 4.6 (a). The optical power absorbed by a single ion depends on the light intensity incident *and* the ion's absorption cross-section, *i.e.,*

Definition of absorption cross-section

$$\text{Optical power absorbed by an ion} = \text{Light intensity} \times \text{Absorption cross-section of ion} = I\sigma_{ab} \quad \textbf{(4.2.13)}$$

which defines the **absorption cross-section** σ_{ab}. For the collection of N_1 ions per unit volume, the total power absorbed per unit volume is $I\sigma_{ab}N_1$.

Consider a small volume $A\Delta x$ of the medium, as in Figure 4.6 (a), where A is the cross-sectional area and the radiation propagates along x. The total optical power ΔP_o absorbed in this volume is then $(I\sigma_{ab}N_1)(A\Delta x)$. Suppose that ΔI is the *change* (negative if absorption) in the intensity of the propagating radiation over the distance Δx. Then

$$A\Delta I = -\text{Total power absorbed in } A\Delta x = -(I\sigma_{ab}N_1)(A\Delta x)$$

i.e.,

Absorption cross-section σ_{ab} and coefficient α

$$-\frac{\Delta I}{I\Delta x} = \sigma_{ab}N_1 = \alpha \quad \textbf{(4.2.14)}$$

where the left-hand side is the fractional change in the light intensity per unit distance, and therefore represents α, the **absorption coefficient** of the medium, due to the absorption of radiation by ions at E_1, which reach E_2. Clearly, $\alpha = \sigma_{ab}N_1$. We could have easily used the latter as the definition of σ_{ab} and derived Eq. (4.2.13) instead.

The stimulated emission case, shown in Figure 4.6 (b), is similar. We can define an **emission cross-section** σ_{em} such that if an incident photon crosses the area σ_{em}, it stimulates the ion to emit a photon as shown in Figure 4.6 (b). In a similar way to Eq. (4.2.13)

Definition of emission cross-section

$$\text{Stimulated optical power emitted by an ion} = \text{Light intensity} \times \text{Emission cross-section of ion} = I\sigma_{em} \quad \textbf{(4.2.15)}$$

As the radiation propagates along x, its intensity I increases due to stimulated emissions from E_2 to E_1. Since only ions at E_2 with a concentration N_2 can be stimulated to emit, the equivalent of Eq. (4.2.14) for the fractional increase in the intensity per unit distance is

Emission cross-section σ_{am} and optical gain

$$\frac{\Delta I}{I\Delta x} = \sigma_{em}N_2 \quad \textbf{(4.2.16)}$$

The net fractional increase in the intensity I per unit distance is defined as the **gain coefficient** g of the medium, so that, in the presence of stimulated emission and absorption

Net optical gain coefficient

$$g = \left[\frac{\Delta I}{I\Delta x}\right]_{\text{net}} = \sigma_{em}N_2 - \sigma_{ab}N_1 \quad \textbf{(4.2.17)}$$

In the presence of just two discrete energy levels, as in Figure 4.6 (a) and (b), we know that $B_{12} = B_{21}$, and similarly, $\sigma_{ab} = \sigma_{em}$.

However, in many practical cases, the two energy levels are actually two manifolds of energies, that is, each manifold consists of narrowly spaced sublevels, as illustrated in Figure 4.6 (c). The lower E_1-manifold has sublevels that cover the range ΔE_1, as in Figure 4.6 (c), and the higher E_2-manifold has sublevels that are within ΔE_2. Suppose we excite the ions with photons of energy $h\upsilon_o$ that are absorbed and take some of ions from the E_1- to the E_2-manifold. Three of the possibilities for absorption are shown in Figure 4.6 (c) that take ions from different sublevels in ΔE_1 and put them at various sublevels in ΔE_2 in which $h\upsilon_o$ is the same for all. However, the downward transitions are more selective because the end level in the E_1-manifold must be empty, and only the higher levels in this manifold (top of ΔE_1) are less populated. Thus, there are less chances for the downward transitions that emit the same $h\upsilon_o$ as shown in Figure 4.6 (c). In addition, the excited ions in the E_2-manifold decay rapidly by emitting lattice vibrations (they thermalize) to the lower levels in E_2 and then emit photons with $h\upsilon_o' < h\upsilon_o$. Clearly, the net effect is that emission at υ_o is less likely than absorption at the same frequency. Consequently, σ_{em} and σ_{ab} are then *not* the same at υ_o; and σ_{ab} is larger than σ_{em} in this example.

We can consider other photon energies besides $h\upsilon_o$ in Figure 4.6 (c), and hence consider σ_{ab} and σ_{em} at other frequencies. We must therefore generalize Eq. (4.2.17) and write σ_{ab} and σ_{em} as frequency dependent, *i.e.,*

$$g(\upsilon) = \sigma_{em}(\upsilon)N_2 - \sigma_{ab}(\upsilon)N_1 \quad \textbf{(4.2.18)}$$

Optical gain coefficient

Although Eq. (4.2.18) has been derived in terms of υ as a variable, it could have just as well be written in terms of λ.

When the net gain in Eq. (4.2.18) is zero, the medium becomes **transparent** to the incoming radiation, which is neither absorbed nor amplified. The **transparency condition** is

$$\sigma_{em}(\upsilon)N_2 - \sigma_{ab}(\upsilon)N_1 = 0 \quad \textbf{(4.2.19)}$$

Transparency condition

To find the overall optical power (or intensity) gain G through a medium that has been pumped or excited, we need to integrate Eq. (4.2.17) along the length of the medium, from $x = 0$ to $x = L$. If N_2 and N_1 remain the same over L and hence g is constant along L, then $G = I(L)/I(0)$ is given by

$$G = \exp(gL) \quad \textbf{(4.2.20)}$$

Net optical power gain

Henceforth we will use the above equations in describing optical amplifiers.

EXAMPLE 4.2.2 Gain coefficient in a Nd^{3+}-doped glass fiber

Consider a Nd^{3+}-doped silica-based glass fiber. Suppose we want to use this fiber as an optical amplifier. The Nd^{3+} concentration in the fiber core is $1 \times 10^{19}\,\text{cm}^{-3}$. The Nd^{3+} ions can be pumped from E_0 to E_3 in Figure 4.4 from which they decay rapidly down to E_2, and the population inversion between E_2 and E_1 gives optical amplification. The E_2 to E_1 emission is approximately at 1.05 μm, and the emission cross-section is about $3 \times 10^{-20}\,\text{cm}^2$. Estimate the maximum gain coefficient with this amount of doping if we could pump all the Nd^{3+} ions. What is gain G in dB if the fiber is 20 cm long?

Solution

The maximum gain will be achieved when all Nd^{3+} ions have been pumped to E_2. From Eq. (4.2.17)

$$g = \sigma_{em}N_2 = (3 \times 10^{-20}\,\text{cm}^2)(1 \times 10^{19}\,\text{cm}^{-3}) = 0.3\,\text{cm}^{-1} \quad \text{or} \quad 30\,\text{m}^{-1}$$

This gain increases with the Nd^{3+} concentration. The maximum theoretical gain G_{max} for a 20 cm fiber is

$$G = \exp(gL) = \exp\left[(0.3\,\text{cm}^{-1}) \times (20\,\text{cm})\right] = 403.4 \quad \text{or} \quad 26.1\,\text{dB}$$

4.3 ERBIUM-DOPED FIBER AMPLIFIERS

A. Principle of Operation and Amplifier Configurations

A light signal that is traveling along an optical fiber over a long distance suffers marked attenuation. It becomes necessary to regenerate the light signal at certain intervals for long-haul communications over several thousand kilometers. Instead of regenerating the optical signal by photodetection, conversion to an electrical signal, amplification, and then conversion back from electrical to light energy by a laser diode, it becomes practical to amplify the signal directly by using an optical amplifier.

One practical **optical amplifier** is based on the **erbium (Er^{3+} ion)-doped fiber amplifier (EDFA).**[12] The core region of an optical fiber is doped with Er^{3+} ions. The host fiber core material is a glass based on silica-aluminate (SiO_3-Al_2O_3) or silica-germania (SiO_3-GeO_2), or both; other glass hosts have also been used. It is easily fused to a single-mode long-distance optical fiber by a technique called *splicing*.

When the Er^{3+} ion is implanted in the host glass material it has the energy levels indicated in Figure 4.7 (a), which appears as though it is a three-level system, but actually mimics a pseudo four-level system.[13] The energies of the Er^{3+} ion in the glass actually fall into *manifolds of energy levels*, which are shown as 1, 2, and 3. Each manifold has several closely spaced energy levels, which are themselves broadened. When an isolated Er^{3+} ion is placed in a glass structure, it finds itself interacting with neighboring host ions. The electric field from these neighboring ions splits the Er^{3+} energy levels in much the same way an applied magnetic field splits the energy levels in the H-atom. In the present case, the energy splitting is due to the local electric field from the ions

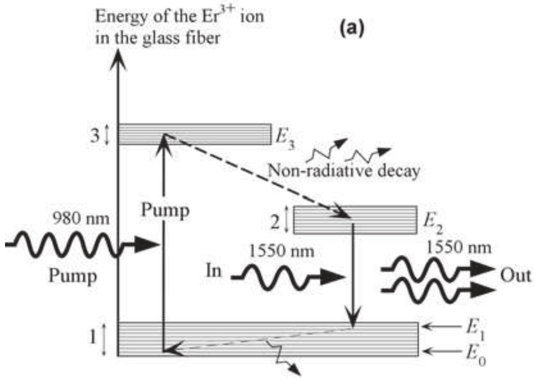


FIGURE 4.7 (a) Energy diagram for the Er^{3+} ion in the glass fiber medium and light amplification by stimulated emission from E_2 to E_1 (features are highly exaggerated). Dashed arrows indicate radiationless transitions (energy emission by lattice vibrations). The pump is a 980-nm laser diode. (b) EDFA can also be pumped with a 1480-nm laser diode.

[12] EDFA was first reported in 1987 by E. Desurvire, J. R. Simpson, and P. C. Becker, and in 1994 AT&T began deploying EDFA repeaters in long-haul fiber communications.

[13] The operation of the Er^{3+}-doped glass fiber laser can be easily explained by using the three-level laser system in Figure 4.2 (a) since the spread between E_1 and E_0 is comparable to k_BT; but a somewhat better explanation would be to treat it as a quasi four-level system as in Figure 4.4. The nonradiative transitions within a manifold are very fast; these are thermalization processes that take typically a few picoseconds.

surrounding the Er^{3+}, and the energy splitting is called the **Stark effect**. Further, the environment of each Er^{3+} ion in the glass is not identical since the glass, unlike a crystal, does not have a well-defined periodic structure; there are local variations in composition and density. The Stark energy levels are therefore broadened as well. The positions and spreads of actual energy levels within a manifold depend on the glass host.

The E_1-manifold is the lowest manifold of energies and is labeled 1 in Figure 4.7 (a); it spreads from E_0 to E_1, and represents the lowest energy levels possible for the Er^{3+} ion.[14] Although most Er^{3+} ions are at E_0, from Boltzmann statistics, there are also Er^{3+} ions at higher levels in this manifold since the manifold is quite narrow, typically about 0.03–0.04 eV (comparable to k_BT at room temperature); the population ratio depends on the spread $E_1 - E_0$ but is very roughly 1 to 4.

There is a higher energy manifold, shown as 3, of narrowly spaced energy levels labeled E_3, as shown in Figure 4.7 (a), that is approximately 1.27 eV above the ground E_1-manifold. The Er^{3+} ions are optically pumped, usually from a laser diode, to excite them from the ground energy manifold 1 to 3. The wavelength for this pumping is about 980 nm. The Er^{3+} ions then rapidly and nonradiatively decay from E_3 to a long-lived energy level E_2 within the manifold 2 of energies in Figure 4.7 (a). The decay from E_3 to E_2 involves energy losses by the emission of phonons, and is very rapid. The Er^{3+} ions at E_2 have a long lifetime ~10 ms. Thus, more and more Er^{3+} ions accumulate at E_2, which is roughly 0.80 eV above the energy level E_1 at the top of the manifold 1. The accumulation of Er^{3+} ions at E_2 eventually leads to a population inversion between E_2 and E_1 inasmuch as E_1 is sparsely populated. Signal photons around 1550 nm have an energy of 0.80 eV, or $E_2 - E_1$, and give rise to *stimulated transitions* of Er^{3+} ions from E_2 to E_1. The transition from E_1 to E_0 is by phonon emission and occurs rapidly (over picoseconds); it corresponds to the thermalization of Er^{3+} ions in this manifold which is narrow. Without pumping, there will be some Er^{3+} ions at E_1 that will *absorb* the incoming 1550 nm photons and reach E_2. To achieve light amplification we must therefore have stimulated emission exceeding absorption. This is only possible if there are more Er^{3+} ions at E_2 than at E_1; that is, if we have population inversion. (As explained below we also need to consider absorption and emission cross-sections.)

It is also possible to pump the Er^{3+} ions to the top of the manifold 2 in Figure 4.7 (a), which is shown in Figure 4.7 (b). Notice that the top of this manifold now acts like a E_3-level, and is labeled E_3'. The pump wavelength is 1480 nm. The Er^{3+} ions decays rapidly by emitting phonons to E_2, and then by stimulated emission down to E_1. Pumping at 980 nm is more efficient than pumping at 1480 nm.

To understand the basic operation of the EDFA we need to consider the absorption and emission cross-sections σ_{ab} and σ_{em} and their spectral dependence, which are shown in Figure 4.8 (a).[15] The net gain $g(\lambda)$ at λ in Eq. (4.2.18) depends on $\sigma_{em}N_2 - \sigma_{ab}N_1$. Under strong pumping, there will be more Er^{3+} in the E_2-manifold than in the E_1-manifold, the emission term $\sigma_{em}N_2$ will dominate over $\sigma_{ab}N_1$ ($N_2 \gg N_1$) and hence the net gain coefficient will be $g(\lambda) \approx \sigma_{em}(\lambda)N_2 \approx \sigma_{em}(\lambda)N_0$, where N_0 is the Er^{3+} concentration in the core of the fiber. Thus, σ_{em} vs. λ also represents the

[14]The valence electrons of the Er^{3+} ion are arranged to satisfy the Pauli exclusion principle and Hund's rule, and this arrangement has the energies lying between E_0 and E_1.

[15]If this were a simple two energy level system, $\sigma_{em} = \sigma_{ab}$ inasmuch as $B_{12} = B_{21}$. It would be impossible to optically pump from E_1 to E_2 and achieve net gain. However, the E_1 and E_2 are manifolds of energy levels. In each manifold we have many levels that have been broadened and overlap.

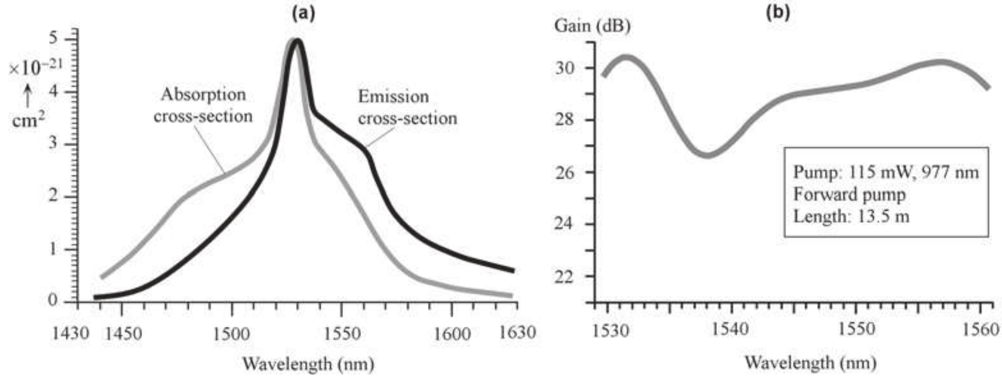


FIGURE 4.8 (a) Typical absorption and emission cross-sections, σ_{ab} and σ_{em}, respectively, for Er^{3+} in a silica glass fiber doped with alumina (SiO_2-Al_2O_3). (Cross-section values for the plots were extracted from B. Pedersen *et al.*, *J. Light Wave Technol.*, *9*, 1105, 1991.) (b) The spectral characteristics of gain, G in dB, for a typical commercial EDF, available from Fibercore as IsoGain™ fiber. Forward pumped at 115 mW and at 977 nm. The insertion losses are 0.45 dB for the isolator, 0.9 dB for the pump coupler and splices.

spectral dependence of the gain coefficient under strong pumping. Clearly, the EDFA response is not "flat"; there is a peak around 1530 nm. The absolute maximum gain with nearly full inversion would be $G = \exp(\sigma_{em}N_0L)$. We can quickly estimate the maximum possible gain at 1550 nm. For an Er^{3+} concentration (N_0) of 10^{19} cm^{-3}, at 1550 nm, using $\sigma_{em} \approx 3.2 \times 10^{-21}$ cm^2 from Figure 4.8 (a), $g \approx \sigma_{em}N_0 = 3.2$ m^{-1}, and the optical gain G for a fiber length of 1 m is 24.5 or 13.9 dB.

If, on the other hand, the EDFA is left unpumped, the Er^{3+} ions would be in the ground state, and the signal will be absorbed with an attenuation coefficient $\alpha = \sigma_{ab}N_0$ as in Eq. (4.2.14). From Figure 4.8 (a), $\sigma_{ab} \approx 2.4 \times 10^{-21}$ cm^2 and hence $\alpha = \sigma_{ab}N_0 = 2.4$ m^{-1} or 10.4 dB m^{-1}. Over the length of the EDFA, ~10 m, this would significantly extinguish the signal. Thus, the EDFA must be pumped at all times.

The spectral shapes of σ_{ab} and σ_{em} in Figure 4.8 (a) depend on the host glass composition and vary significantly between host glasses. It is possible to broaden the bandwidth of the gain curve (*i.e.*, σ_{em} vs. λ) by appropriately choosing the glass composition. Figure 4.8 (b) shows typical spectral characteristics of gain, G in dB, for a commercial erbium-doped fiber that has been pumped with a laser diode at 980 nm with 115 mW. The spectrum is not flat and is normally flattened to obtain uniform gain over the band of wavelengths used in multi channel communications; this is described below.

In practice, the erbium-doped fiber is inserted into the fiber communications line by splicing as shown in the simplified schematic diagram in Figure 4.9, and it is pumped from a laser diode through a coupling fiber arrangement which allows only the pumping wavelength to be coupled; it is a **wavelength-selective coupler**. Some of the Er^{3+} ions at E_2 will decay spontaneously from E_2 to E_1, which will give rise to unwanted noise in the amplified light signal. Further, if the EDFA is not pumped at any time, it will act as an attenuator as the 1550 nm photons will be absorbed by Er^{3+} ions, which will become excited from the ground energy manifold 1 to the E_2-manifold. In returning back to ground energy by spontaneous emission, they will emit 1550 nm photons randomly, which generates noise in the optical link. **Optical isolators** inserted at the entry and exit end of the amplifier allow only the optical signals at 1550 nm to pass in one direction and prevent the 980 pump light from propagating back or forward into the communication system.

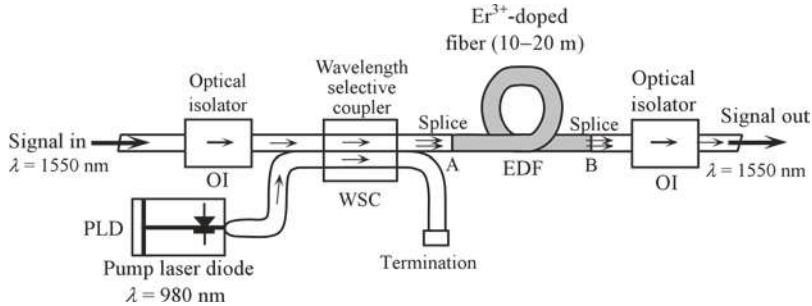
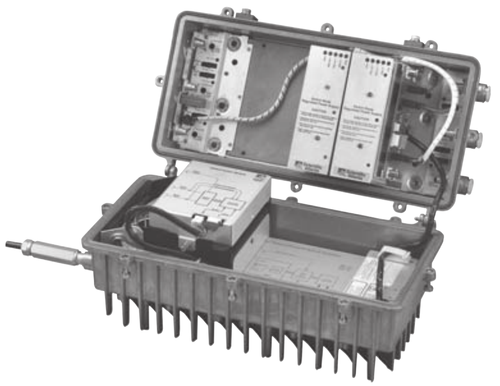


FIGURE 4.9 A simplified schematic illustration of an EDFA (optical amplifier). The Er^{3+}-doped fiber is pumped by feeding the light from a laser pump diode, through a coupler, into the erbium ion-doped fiber. Forward or codirectional pump configuration.



EDFA (Strand Mounted Optical Amplifier, Prisma 1550) for optical amplification at 1550 nm. This model can be used underground to extend the reach of networks, and operates over −40 °C to +65°C. The output can be as high as 24 dBm. *(Courtesy of Cisco.)*

EDFAs (LambdaDriver® Optical Amplifier Modules) with low noise figure and flat gain (to within ± 1 dB) for use in DWDM over 1528–1563 nm. These amplifiers can be used for booster, in-line, and preamplifier applications. *(Courtesy of MRV Communications, Inc.)*

The pumping configuration shown in Figure 4.9 feeds the pump light at A. A simplified fiber optic diagram of this configuration is shown in Figure 4.10 (a). Since the pump light and the signal travel in the same direction, this particular configuration is called **codirectional** or **forward pumping**. Alternatively, one can feed the pump light at B in Figure 4.9, as in Figure 4.10 (b), in which the pump light propagates in the opposite direction to the signal. This configuration is

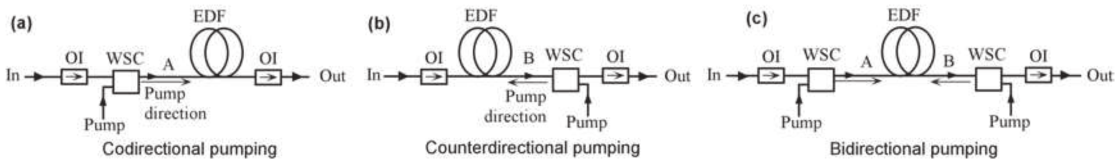


FIGURE 4.10 (a) Codirectional or forward pumping. (b) Counterdirectional or backward pumping. (c) Bidirectional pumping. OI, optical isolator, WSC, wavelength-selective coupler.

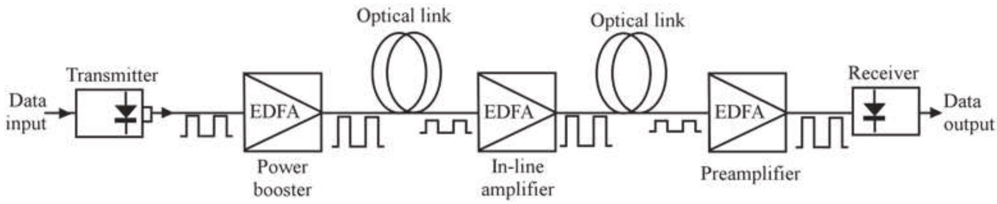


FIGURE 4.11 EDFAs are widely used in telecommunications as a power booster, after the transmitter, in-line amplifier within the optical link, and preamplifier before the receiver. The symbol shown for the EDFA is widely used.

called **counterdirectional** or **backward pumping**. In **bidirectional pumping**, also called **dual pumping**, there are two pump diodes coupled to the EDF at A and B as shown in Figure 4.10 (c). In addition, there is usually a photodetector system coupled to monitor the pump power or the EDFA output power. These are not shown in Figure 4.9 nor in Figure 4.10. Typically EDFAs are used in three types of applications in optical communications. As a power **booster**, the EDFA is placed after the transmitter to boost the signal power fed into the optical network as shown in Figure 4.11. Quite often the signal from a transmitter needs to be boosted if the signal is to be split into N channels through an optical splitter. Booster EDFAs are designed so that the output signal does not easily saturate for large input signals. As an **in-line amplifier**, the EDFA is placed within the optical transmission link, as shown in Figure 4.11, where the signal has become weak and needs to be amplified. The in-line EDFA should have a large gain and add very little noise into the fiber. The noise in the EDFA comes from the amplification of spontaneously (and randomly) emitted photons in the direction of the fiber. As a **preamplifier**, it is placed before the receiver to amplify the signal fed into the detector as illustrated in Figure 4.11. The preamplifier EDFA is designed to provide good gain with low noise.

B. EDFA Characteristics, Efficiency, and Gain Saturation

The gain of an EDFA depends on the launched pump power from the pump diode. The gain coefficient g depends on $\sigma_{em}N_2 - \sigma_{ab}N_1$, and as more pump power is fed into the core of the EDFA, N_2 increases and N_1 decreases given that $N_1 + N_2 = N_0$, the Er^{3+} concentration in the fiber. Figure 4.12 (a) shows typical characteristics of EDFA gain, G, in dB vs. launched pump

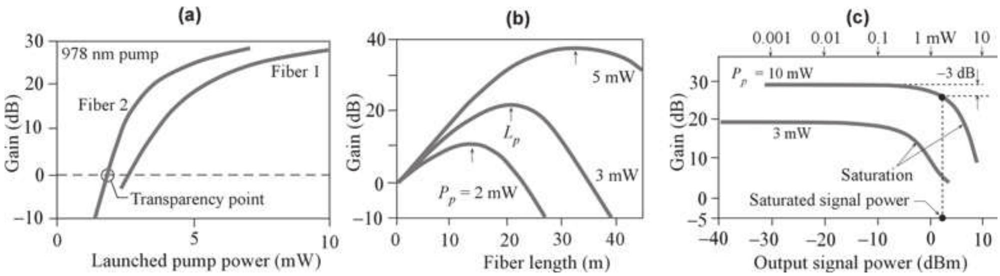


FIGURE 4.12 (a) Typical characteristics of EDFA small signal gain in dB vs. launched pump power for two different types of fibers pumped at 980 nm. The fibers have different core compositions and core diameter, and different lengths (L_1 = 19.9 m and L_2 = 13.6 m). (b) Typical dependence of small signal gain G on the fiber length L at different launched pump powers. There is an optimum fiber length L_p. (c) Typical dependence of gain on the output signal strength for different launched pump powers. At high output powers, the output signal saturates, *i.e.,* the gain drops. (*Source:* Figures were constructed by using typical data from C. R. Jiles *et al., IEEE Photon. Technol. Letts., 3,* 363, 1991; and C. R. Jiles *et al., J. Light Wave Technol., 9,* 271, 1991.)

power P_p for two different types of fibers. As expected, G first increases sharply with P_p. At high pump powers, further increases in N_2 and hence G become constrained due to the diminishing population N_1 with increasing pump power. The gain G shown in Figure 4.12 (a) is for **small signals**, that is, for signal optical powers that do not themselves significantly depopulate N_2 by encouraging stimulated emissions, and hence decrease G.

Since the gain G is given by $\exp(gL)$, G increases with the fiber length L as shown in Figure 4.12 (b) for a given fiber at different pump power levels. However, we cannot indefinitely increase the gain by using a longer fiber. As shown in Figure 4.12 (b), initially G increases with L but then drops with L for long fiber lengths. Put differently, there is an **optimum fiber length** or **pump length** that maximizes the gain at a given launched pump power. Beyond this optimum length, the pump power in the core decreases with length so much that it is unable to maintain the required gain coefficient g. Moreover, further along the fiber, the pump becomes too weak, the gain disappears and the fiber attenuates the signal, which is highly undesirable. Thus, it is important to design the EDFA with a length that is not longer than the optimum length. The optimum fiber length is longer for higher launched pump powers as can be seen from Figure 4.12 (b).

One very important characteristic of an EDFA is the **saturation** of its output signal power, that is, the fall in the gain, under large signals. As the signal is amplified, the optical power at 1550 nm in the fiber core increases, which itself encourages further stimulated emission and thereby depopulates the population N_2; and hence the gain drops. The gain shown in Figures 4.12 (a) and (b) is for small signals, that is, the signal power is much less than the pump power in the core. Normally manufacturers quote the output power that saturates the amplifier, which is the output signal at which the gain drops by 3 dB below its constant small signal value as shown in Figure 4.12 (c). It is simply called the maximum output power and is quoted in dBm. In commercial high power booster EDFAs, the saturated output power is typically around 22–25 dBm (160–300 mW) but at the expense of gain; the corresponding small signal gain is around 10–15 dB (or 10–30). In contrast, small in-line EDFAs typically have a saturation power of 15–20 dBm (30–100 mW) and a small signal gain of 20–30 dB or higher.

The **gain efficiency** of an EDFA is the maximum optical gain achievable per unit optical pumping power and is quoted in dB mW^{-1}. Typical gain efficiencies are around $8-10$ dB mW^{-1} at 980 nm pumping. A 30 dB or 10^3 gain is easily attainable with a few tens of milliwatts of pumping at 980 nm, depending on the fiber core doping and the coupling of the pump into the fiber.

The power conversion efficiency (PCE), η_{PCE}, of an EDFA represents the efficiency with which the amplifier is able to convert power from the pump to the signal. Suppose that $P_{p\text{in}}$ is the pump power into the EDFA as shown in Figure 4.13 (a), where $P_{s\text{in}}$ and $P_{s\text{out}}$ are the input and output powers, respectively. Then

$$\eta_{\text{PCE}} = \frac{P_{s\text{out}} - P_{s\text{in}}}{P_{p\text{in}}} \approx \frac{P_{s\text{out}}}{P_{p\text{in}}} \tag{4.3.1}$$

Power conversion efficiency

Suppose that the flux of photons from the pump is $\Phi_{p\text{in}}$ and the signal input and output photon fluxes are $\Phi_{s\text{in}}$ and $\Phi_{s\text{out}}$. Since optical power is proportional to photon flux $\times\, hc/\lambda$ where hc/λ is the photon energy, we can write Eq. (4.3.1) as

$$\eta_{\text{PCE}} \approx \frac{\Phi_{s\text{out}}}{\Phi_{p\text{in}}} \times \frac{\lambda_p}{\lambda_s} \tag{4.3.2}$$

Power conversion efficiency

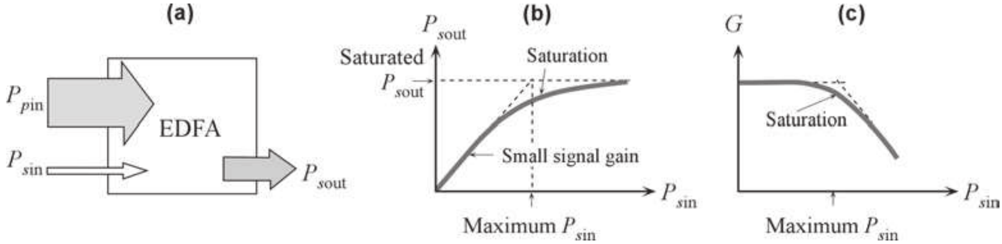


FIGURE 4.13 (a) The EDFA is pumped by P_{pin}. The power from the pump is converted into the output signal power P_{sout}. (b) The P_{sout} vs. P_{sin} behavior is initially linear but saturates at high input signal powers. (c) The output becomes saturated. The gain G drops in the saturation region.

where λ_p and λ_s are the pump and signal wavelengths. Equation (4.3.2) shows that η_{PCE} has a maximum value of λ_p/λ_s, which corresponds to each pump photon yielding one additional signal output photon. Thus, maximum PCE is 63% for 980 nm pumping.

The gain can be defined and written as

EDFA gain

$$G = \frac{P_{sout}}{P_{sin}} = 1 + \eta_{\text{PCE}}\left(\frac{P_{pin}}{P_{sin}}\right) \tag{4.3.3}$$

in which we used Eq. (4.3.1) to substitute for P_{sout}. The maximum gain $G_{\max}$ is for $\eta_{\text{PCE}} = \lambda_p/\lambda_s$ so that $G < G_{\max}$ implies

$$G < 1 + (\lambda_p/\lambda_s)(P_{pin}/P_{sin})$$

or, rearranging, the input signal is

EDFA gain and input signal power

$$P_{sin} < (\lambda_p/\lambda_s)P_{pin}/(G - 1) \tag{4.3.4}$$

The importance of this equation is that it specifies the range of input signals, that is, the maximum input signal power, beyond which the output signal power begins to saturate as indicated in Figure 4.13 (b). Consequently, the gain drops with the input signal power as in Figure 4.13 (c). This phenomenon is called **gain saturation** or **compression**. For example, for an EDFA that has a small signal gain of 30 dB (or 10^3), $P_{pin} = 100$ mW, 980 nm pumping, we can use Eq. (4.3.4) to find P_{sin} must be less than 60 μW, or the output must be less than 60 mW. Thus, for outputs beyond 60 mW, the EDFA will saturate.

For a given pump level P_p, there is a maximum fiber length L_p beyond which the gain is lost rapidly with length, and the fiber attenuates the signal. To find the length of fiber L_p required to absorb the pump radiation, we note that under strong pumping, N_1 is very small. There are very few ions in the E_1-manifold. The absorbed pump power is primarily taken by those excited ions that have spontaneously decayed from E_2 to E_1, because these have to be pumped back up to E_3 and hence E_2. The characteristic time of this natural decay from the E_2 to E_1 is the **spontaneous emission lifetime** $\tau_{\text{sp}}(= 1/A_{21})$. Thus,

$$\begin{aligned}\text{Absorbed pump power} &= \text{Absorbed energy per unit time}\\ &\approx (\text{Volume})(N_2)(\text{Pump photon energy})/(\text{Decay time})\\ &= (AL_p)(N_2)(h\upsilon_p)/\tau_{\text{sp}}\end{aligned}$$

where A is the cross-sectional area of the fiber core that has the Er^{3+} dopants, and L_p is the length of fiber needed to absorb the necessary pump radiation, υ_p is the pump frequency. Thus, the absorbed pump power P_p is

Strong absorption and pump length of EDFA

$$P_p \approx AN_0 h\upsilon_p L_p/\tau_{sp} \tag{4.3.5}$$

where $N_2 \approx N_0$, the Er^{3+} concentration in the core. Equation (4.3.5) essentially determines (very approximately) the maximum fiber length allowed for a given amount of pumping. The actual fiber length L needs to be shorter than L_p for the following reason. If L is longer than L_p, then beyond L_p, the gain will be lost and the signal will become absorbed since this region will have unpumped Er^{3+} ions. We assumed a small signal in our derivation that did not significantly upset the population of the upper level. If the pumping is not strong and/or the signal is not small, then N_1 cannot be neglected. Hence, the actual L is usually less than L_p in Eq. (4.3.5). Further, not all the pump radiation will be confined to the fiber core, whereas we assumed it was. Only some fraction Γ, called **the confinement factor**, of the pump radiation will be guided within the core and hence pump the Er^{3+} ions. Γ is typically 0.6–0.8. Thus, we need to use ΓP_p instead of P_p in Eq. (4.3.5), *i.e.*,

Strong absorption and pump length of EDFA

$$\Gamma P_p \approx AN_0 h\upsilon_p L_p/\tau_{sp} \tag{4.3.6}$$

Given the fiber length L ($<L_p$), the **small signal gain** G is easily found from

Small signal gain

$$G = \exp(gL) \tag{4.3.7}$$

which is maximum for a fiber length $L = L_p$.

EXAMPLE 4.3.1 An erbium-doped fiber amplifier

Consider a 3-m EDFA that has a core diameter of 5 μm, Er^{3+} doping concentration of $1 \times 10^{19}\,\text{cm}^{-3}$, and τ_{sp} (the spontaneous decay time from E_2 to E_1) is 10 ms. The fiber is pumped at 980 nm from a laser diode. The pump power coupled into the EDFA fiber is 25 mW. Assuming that the confinement factor Γ is 70%, what is the fiber length that will absorb the pump radiation? Find the small signal gain at 1550 nm for two cases corresponding to full population inversion and 90% inversion.

Solution

The pump photon energy $h\upsilon = hc/\lambda = (6.626 \times 10^{-34})(3 \times 10^8)/(980 \times 10^{-9}) = 2.03 \times 10^{-19}\,\text{J}$ (or 1.27 eV).

Rearranging Eq. (4.3.6), we get

$$L_p \approx \Gamma P_p \tau_{sp}/ANh\upsilon_p$$

i.e.,

$$L_p \approx (0.70)(25 \times 10^{-3}\,\text{W})(10 \times 10^{-3}\,\text{s})/\left[\pi(2.5 \times 10^{-4}\,\text{cm})^2(1 \times 10^{19}\,\text{cm}^{-3})(2.03 \times 10^{-19}\,\text{J})\right] = 4.4\,\text{m}$$

which is the maximum allowed length. The small signal gain can be rewritten as

$$g = \sigma_{em}N_2 - \sigma_{ab}N_1 = \left[\sigma_{em}(N_2/N_0) - \sigma_{ab}(N_1/N_0)\right]N_0$$

where $N_1 + N_2 = N_0$ is the total Er^{3+} concentration. Let $x = N_2/N_0$, then $1 - x = N_1/N_0$ where x represents the extent of pumping from 0 to 1, 1 being 100%. Thus, the above equation becomes

$$g = \left[\sigma_{em}x - \sigma_{ab}(1 - x)\right]N_0$$

For 100% pumping, $x = 1$,

$$g = \left[(3.2 \times 10^{-21}\,\text{cm}^2)(1) - 0\right](1 \times 10^{19}\,\text{cm}^{-3}) = 3.2\ \text{m}^{-1}$$

and

$$G = \exp(gL) = \exp\left[(3.2\ \text{m}^{-1})(3\ \text{m})\right] = 14{,}765 \quad \text{or} \quad 41.7\ \text{dB}$$

For $x = 0.9$ (90% pumping), we have

$$g = \left[(3.2 \times 10^{-21}\,\text{cm}^2)(0.9) - (2.4 \times 10^{-21}\,\text{cm}^2)(0.1)\right](1 \times 10^{19}\,\text{cm}^{-3}) = 2.64\ \text{m}^{-1}$$

and

$$G = \exp(gL) = \exp\left[(2.64\ \text{m}^{-1})(3\ \text{m})\right] = 2751 \quad \text{or} \quad 34.4\ \text{dB}$$

Even at 90% pumping the gain is significantly reduced. At 70% pumping, the gain is 19.8 dB. In actual operation, it is unlikely that 100% population inversion can be achieved; 41.7 dB is a good indicator of the upper ceiling to the gain.

C. Gain-Flattened EDFAs and Noise Figure

The gain spectrum of an EDFA refers to the dependence of the small signal gain G on the wavelength within the communications band of interest. Figure 4.8 (b) shows how the gain varies as a function of wavelength for an EDFA that has not had its gain flattened. In wavelength division multiplexing, it is essential to have the gain as flat as possible over all the channels, *i.e.,* the signal wavelengths, that are used. Many gain-flattened commercial EDFA for DWDM have their gain variations within ± 0.5 dB. Figure 4.14 (a) shows typical gain spectra for a commercial EDFA for use in DWDM in the C-band. For small input signals (which do not cause any saturation) the gain is reasonably flat. The gain decreases at high input signals and, in addition, the gain becomes "less" flat.

The simplest way to flatten the gain spectrum is to use a filter (or a series of filters) after the EDFA that has an attenuation or a transmission spectrum such that it attenuates

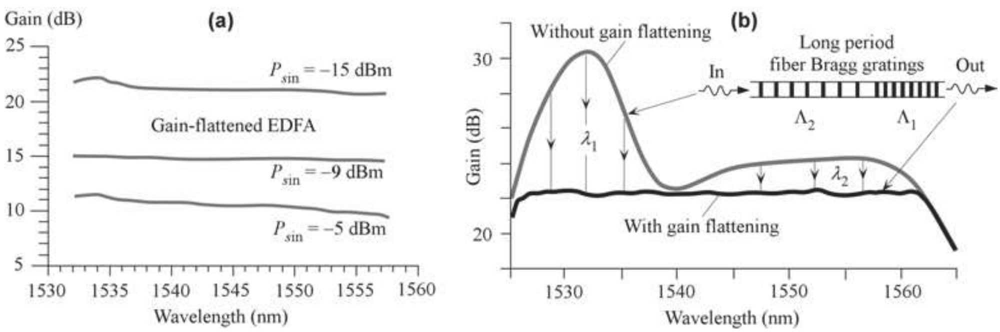


FIGURE 4.14 (a) The gain spectrum of one type of commercial gain-flattened EDFA. The gain variation is very small over the spectrum, but the gain decreases as the input signal power increases due to saturation effects. (*Note*: The corresponding input signal power levels are 0.031, 0.13, and 0.32 mW.) (b) Schematic illustration of gain equalization by using long period fiber Bragg grating filters in series that attenuate the high gain regions to equalize the gain over the spectrum. (An idealized example.)

the higher gain region of the spectrum and not the lower gain region. The high gain regions are attenuated by the right amount to bring these regions down to the same level as the low gain region, as shown for an idealized case in Figure 4.14 (b). The filter **equalizes** the gain. The combination of this filter with the EDFA will then result in a gain spectrum that has been **equalized**. The effect is to attenuate the high gain region more than the low gain region, and hence achieve gain equalization. Typically **long period gratings** (LPGs) are used for the filter part. If the period Λ of a fiber Bragg grating (FBG) is much longer (*e.g.*, by $100\times$) than the wavelength, then the diffraction occurs into the cladding for those wavelengths that satisfy the Bragg diffraction condition. Thus, by choosing the period Λ correctly we can diffract some of the signal in the high gain region into the cladding and hence suppress the signal in this gain region, *i.e.,* equalize the gain. Normally at least two or more properly designed LPGs are needed in series to usefully equalize the gain, as should be apparent from Figure 4.14 (b) where one would bring the gain in the λ_1-region down and the other lowers the gain in the λ_2-region. LPGs have a periodicity (Λ) that is of the order of 100 μm and hence are easier to manufacture than regular FBGs. In addition, the diffracted wave is not simply reflected but is fed into the cladding where it is attenuated and lost; hence the back reflection is low. Commercially available LPGs can equalize the gain down to about ± 0.5 dB. Good gain equalization requires the filter transmission spectrum to have a shape that is the **inverse** of the shape of the gain spectrum of the EDFA. Some manufacturers also provide the transmission spectrum of the filter in their data sheets.

It is also possible to obtain gain equalization by using a regular FBG in which the periodicity Λ is not fixed, but adjusted along the fiber length, *e.g.,* becomes shorter along the fiber, so that the transmission spectrum can be tailored to the need, and can be made to be an almost exact inverse of the EDFA's gain spectrum to yield a flat net gain spectrum. Such FBGs in which the period changes along the fiber are called **chirped FBGs**. Most commercial chirped FBGs can provide a gain flatness down to ± 0.25 dB, which is better than ± 0.5 dB for LPGs; research shows that one can, in fact, do better, down to ± 0.1 dB. Chirped FBGs suffer from high back reflections (attenuated signals are actually back reflected) but they are more stable to temperature variations compared to LPGs.

Most commercial gain-flattened EDFAs have more than just a series of filters placed after the EDFA. Usually a multistage EDFA design is used. A simple broadband gain-flattened two-stage EDFA configuration is shown in Figure 4.15 (a) as reported by Lucent researchers.[16] A 980-nm laser diode is used to forward pump the EDFA1, whose output is passed through a well-designed long period grating for gain flattening. There is a second stage, EDFA2, that is forward pumped from a 1480-nm laser diode. The small signal, that is unsaturated, gain spectrum from such a two-stage gain-flattened EDFA is shown in Figure 4.15 (b), where it can be seen that over a broadband (40 nm), the gain variation is less than 1 dB; or put differently, less than ± 0.5 dB about the mean. By using more stages and filters, excellent gain flattening can be easily obtained. Table 4.1 summarizes the properties of a particular commercial EDFA that has been gain flattened and designed for DWDM applications.

Like all amplifiers, EDFAs also generate noise in addition to the amplification they provide. The noise in an EDFA arises from **amplified spontaneous emission** (ASE), which is the

[16] See P. F. Wysocki *et al.*, *IEEE Photon. Technol. Lett.*, *9*, 1343, 1997; Y. Sun *et al.*, *Bell Labs Tech. J.*, *4*, 187, 1999; and R. P. Espindola and P. F. Wysocki, *U.S. Patent* 5,920,424, June 6, 1999.

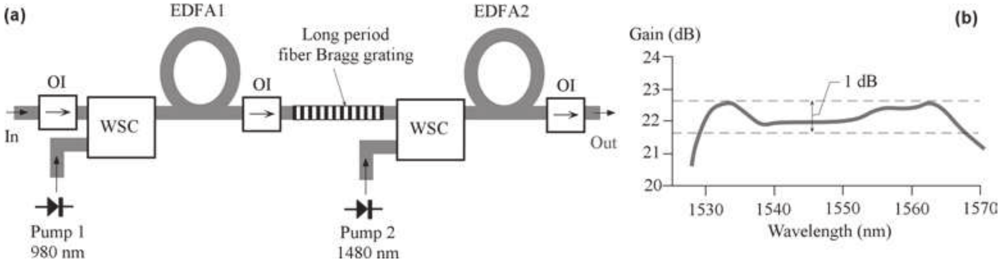


FIGURE 4.15 (a) A gain-flattened EDFA reported by Lucent Technologies[17] uses two EDFAs and a long period grating between the two stages. (A simplified diagram.) The two EDFAs are pumped at 980 nm and 1480 nm. (b) The resulting gain spectrum for small signals is flat to better than 1 dB over a broad spectrum, 40 nm. The length of EDFA1 is 14 m, and that of EDFA2 is 15 m. Pump 1 (980 nm) and Pump 2 (1480 nm) diodes were operated at most at power levels 76 mW and 74.5 mW, respectively. EDFA2 can also be pumped counterdirectionally.

amplification of randomly emitted photons from the upper E_2-manifold to the lower E_1-manifold. This spontaneous emission generates random photons, and as these photos travel along the fiber, they become amplified. Thus, the output from an EDFA not only has the signal but also noise from the ASE as shown in Figure 4.16 (a).

The ratio of the signal power (P_s) to the noise power (P_n) present at the input or at the output of an amplifier is called the **signal-to-noise ratio** (SNR). At the input and the output, the SNR would be designated by SNR_{in} and SNR_{out}, respectively. An ideal amplifier should not add any noise while amplifying the signal so that the SNR_{out} should be the same as SNR_{in} because both the signal and noise at the input would be amplified by the same amount, gain G (and no additional noise would be added). Noise characteristics of amplifiers are generally quantified by their **noise figure** (NF), defined by

Noise figure definition

$$\text{NF} = \frac{\text{SNR}_{\text{in}}}{\text{SNR}_{\text{out}}} \quad \text{or} \quad \text{NF(dB)} = 10\log\left(\frac{\text{SNR}_{\text{in}}}{\text{SNR}_{\text{out}}}\right) \tag{4.3.8}$$

where the second term gives NF in the units of dB, the most commonly used units for NF. For a noiseless amplifier, NF $= 0\,$dB (or 1). Table 4.1 shows a value of 5 dB for the NF for the particular EDFA listed in the table. This value is quite typical for EDFAs in the unsaturated region, but increases in the saturation region, *i.e.,* high input (or output) power levels.

TABLE 4.1 Specifications for a typical gain-flattened EDFA for DWDM (dense wavelength division multiplexing) applications

Operating wavelength (nm)	Output power (dBm)	Number of pump lasers (nm)	Input signal for flat gain (dBm)	Channels	Max gain (dB channel^{-1})	Flatness (dB)	NF (dB)
1529–1561	18–24	2 × 980 2 × 1480	−15 to −7	64	21	±0.5	5

Note: Optilab EDFA-GI series EDFA.

[17] See footnote number 15.

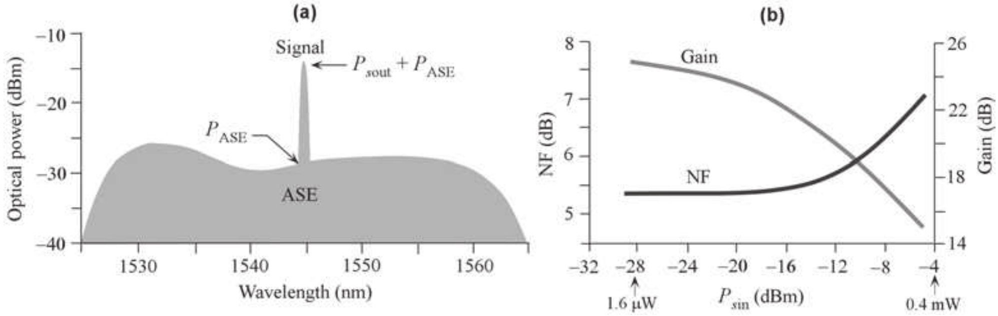


FIGURE 4.16 (a) Amplified spontaneous emission (ASE) noise in the output spectrum and the amplified signal. (b) The dependence of NF and gain (G) on the input signal power level (P_{sin}) for an EDFA under forward (codirectional) pumping. (*Source:* Data for the plots were selectively extracted from G. R. Walker *et al., J. Light Wave Technol., 9,* 182, 1991.)

As shown in Figure 4.16 (b), the NF is initially independent of the input signal power P_{sin} level, but at sufficiently high P_{sin}, it increases with P_{sin}; or put differently, NF increases with falling gain.

4.4 GAS LASERS: THE He-Ne LASER

With the He-Ne laser one has to confess that the actual explanation is by no means simple since we have to know such things as the energy states of the whole atom. We will consider only the lasing emission at 632.8 nm which gives the well-known red color to the He-Ne laser light. The actual stimulated emission occurs from the Ne atoms. He atoms are used to excite the Ne atoms by atomic collisions. Overall, the He-Ne gas laser is like a four-level laser system (Figure 4.4).

Ne is an inert gas with a ground state ($1s^2 2s^2 2p^6$) which will be represented as ($2p^6$) by ignoring the inner closed $1s$ and $2s$ subshells. If one of the electrons from the $2p$ orbital is excited to a $5s$-orbital, then the excited configuration ($2p^5 5s^1$) is a state of the Ne atom that has higher energy. Similarly, He is also an inert gas which has the ground state configuration of ($1s^2$). The state of He when one electron is excited to a $2s$-orbital can be represented as ($1s^1 2s^1$) and has higher energy than the ground state.

The He-Ne laser consists of a gaseous mixture of He and Ne atoms in a gas discharge tube, as sketched schematically in Figure 4.17. The ends of the tube are mirrored to reflect the stimulated radiation and build up intensity within the cavity. In other words, an *optical* cavity is

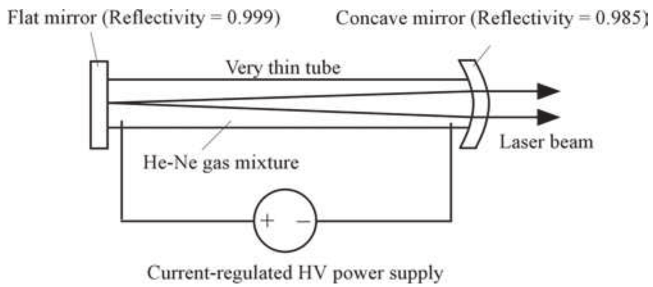


FIGURE 4.17 A schematic illustration of the principle of the He-Ne laser.

formed by the end mirrors so that reflection of photons back into the lasing medium builds up the photon concentration in the cavity, a requirement of an efficient stimulated emission process as discussed above. By using DC or RF high voltage, an electrical discharge is obtained within the tube which causes the He atoms to become excited by collisions with those energetic electrons that have been accelerated by the field. Thus,

$$\text{He} + e^- \rightarrow \text{He}^* + e^-$$

where He* is an excited He atom.

The excitation of the He atom by an electron collision puts the second electron in He into a $2s$ state, so that the excited He atom, He*, has the configuration $(1s^1 2s^1)$, which is *metastable* (that is, a long-lasting state) with respect to the $(1s^2)$ state as shown schematically in Figure 4.18. He* cannot spontaneously emit a photon and decay down to the $(1s^2)$ ground state because the *orbital quantum number* ℓ of the electron must change by ± 1, *i.e.,* $\Delta\ell$ must be ± 1 for any photon emission or absorption process. Thus a large number of He* atoms build up during the electrical discharge because they are not allowed to simply decay back to ground state. When an excited He atom collides with a Ne atom, it transfers its energy to the Ne atom by resonance energy exchange because, by good fortune, Ne happens to have an empty energy level, corresponding to the $(2p^5 5s^1)$ configuration, matching that of $(1s^1 2s^1)$ of He*. Thus the collision process excites the Ne atom and de-excites He* down to its ground energy, *i.e.,*

$$\text{He}^* + \text{Ne} \rightarrow \text{He} + \text{Ne}^*$$

With many He*-Ne collisions in the gaseous discharge, we end up with a large number of Ne* atoms and a population inversion between $(2p^5 5s^1)$ and $(2p^5 3p^1)$ states of the Ne atom as indicated in Figure 4.18. A spontaneous emission of a photon from one Ne* atom falling from $5s$ to $3p$ gives rise to an avalanche of stimulated emission processes which leads to a lasing emission with a wavelength of 632.8 nm in the red.

There are a few interesting facts about the He-Ne laser, some of which are quite subtle. First, the $(2p^5 5s^1)$ and $(2p^5 3p^1)$ electronic configurations of the Ne atom actually have a spread of energies. For example, for Ne$(2p^5 5s^1)$, there are four closely spaced energy levels. Similarly for Ne$(2p^5 3p^1)$ there are ten closely separated energies. We see that we can achieve population

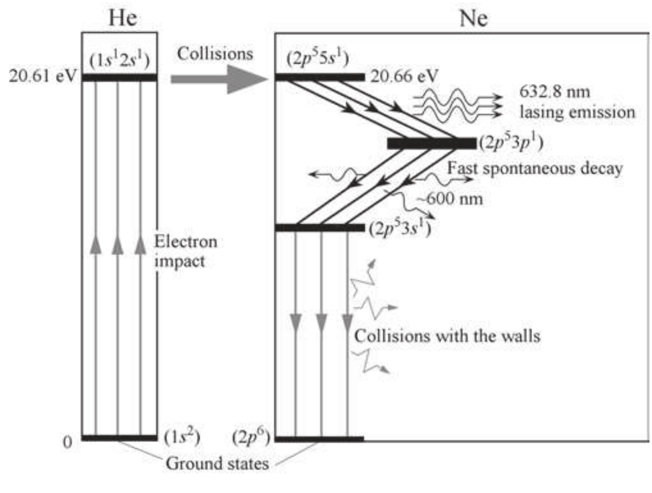


FIGURE 4.18 The principle of operation of the He-Ne laser and the He-Ne laser energy levels involved for 632.8 nm emission.

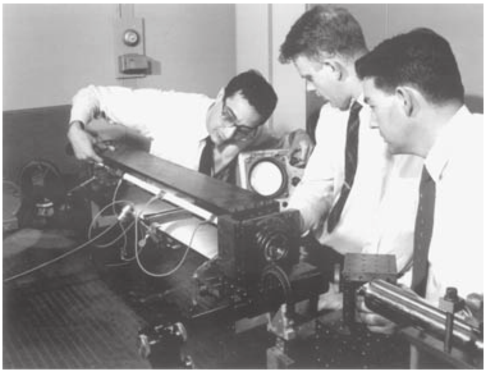
inversion with respect to a number of energy levels and, as a result, the lasing emissions from the He-Ne laser contain a variety of wavelengths. The two lasing emissions in the visible spectrum at 632.8 nm and 543 nm can be used to build a red or a green He-Ne laser. Further, we should note that the energy of the state Ne($2p^5 4p^1$) (not shown) is above Ne($2p^5 3p^1$) but below Ne($2p^5 5s^1$). There will therefore also be stimulated transitions from Ne($2p^5 5s^1$) to Ne($2p^5 4p^1$) and hence a lasing emission at a wavelength of ~3.39 μm (infrared). To suppress lasing emissions at the unwanted wavelengths and to obtain lasing only at the wavelength of interest, the reflecting mirrors can be made wavelength selective. This way, the optical cavity builds up optical oscillations at the selected wavelength.

From the ($2p^5 3p^1$) energy levels, the Ne atoms decay rapidly to the ($2p^5 3s^1$) energy levels by spontaneous emission. Most of Ne atoms with the ($2p^5 3s^1$) configuration, however, cannot simply return to the ground state $2p^6$ by photon emission because the return of the electron in $3s$ requires that its spin is flipped to close the $2p$-subshell. An electromagnetic radiation cannot change the electron spin. Thus, the Ne($2p^5 3s^1$) energy levels are *metastable* states. The only possible return to the ground state (and become pumped again) is by collisions with the walls of the laser tube. We cannot therefore increase the power obtainable from a He-Ne laser by simply increasing the laser tube diameter because that will accumulate more Ne atoms at the metastable ($2p^5 3s^1$) states.

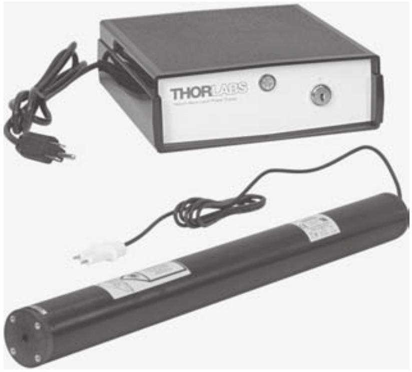
A typical He-Ne laser, as illustrated in Figure 4.17, consists of a narrow glass tube which contains the He and Ne gas mixture—typically He to Ne ratio of 10 to 1 and a pressure of ~1 torr. (The optimum pressure depends on the tube diameter.) The lasing emission intensity increases with the tube length since there are more Ne atoms used in stimulated emissions. The intensity decreases with increasing tube diameter since Ne atoms in the ($2p^5 3s^1$) states can only return to the ground state by collisions with the walls of the tube. The ends of the tube are generally sealed with a flat mirror (99.9% reflecting) at one end and, for easy alignment, a concave mirror (99% reflecting) at the other end to obtain an **optical cavity** within the tube. The outer surface of the concave mirror is shaped like a convergent lens to compensate for the divergence in the beam arising from reflections from the concave mirror. The output radiation from the tube is typically a beam of diameter 0.5–1 mm and a divergence of a few milliradians at a power of few milliwatts. In high power He-Ne lasers, the mirrors are external to the tube. In addition, *Brewster windows* are typically used at the ends of the laser tube to allow only polarized light to be transmitted and amplified within the cavity so that the output radiation is polarized (has electric field oscillations in one plane).

Even though we can try to get as parallel a beam as possible by lining up the mirrors perfectly, we will still be faced with diffraction effects at the output. When the output laser beam hits the end of the laser tube it becomes diffracted so that the emerging beam is necessarily divergent. Simple diffraction theory can readily predict the divergence angle. Further, typically one or both of the reflecting mirrors in many gas lasers are made concave for a more efficient containment of the lasing radiation within the **active medium** and for easier alignment. The beam within the cavity and hence the emerging radiation is approximately a *Gaussian beam.* As mentioned in Chapter 1, a Gaussian beam diverges as it propagates in free space. Optical cavity engineering is an important part of laser design and there are various advanced texts on the subject.

Due to their relatively simple construction, He-Ne lasers are widely used in numerous applications such as interferometry, accurately measuring distances or flatness of an object, laser printing, holography, and various pointing and alignment applications (such as in civil engineering).



Ali Javan and his associates, William Bennett Jr. and Donald Herriott at Bell Labs, were the first to successfully demonstrate a continuous wave (CW) helium-neon laser operation (1960–1962). *(Reprinted with permission of Alcatel-Lucent USA Inc.)*



A modern He-Ne laser with its power supply. This unit provides a linearly polarized TE_{00} output at 633 nm (red) at a power of 10 mW. The beam diameter is 0.68 mm and the divergence is 1.2 mrd. The longitudinal mode separation is 320 MHz. *(Courtesy of Thorlabs.)*

EXAMPLE 4.4.1 Efficiency of the He-Ne laser

A typical low-power, 5-mW He-Ne laser tube operates at a DC voltage of 2000 V and carries a current of 7 mA. What is the efficiency of the laser?

Solution

From the definition of efficiency,

$$\text{Efficiency} = \frac{\text{Output Light Power}}{\text{Input Electrical Power}} = \frac{5 \times 10^{-3}\,\text{W}}{(7 \times 10^{-3}\,\text{A})(2000\,\text{V})}$$

$$= 0.036\%$$

Typically He-Ne efficiencies are less than 0.1%. What is important is the highly coherent output radiation. Note that 5 mW over a beam diameter of 1 mm is 6.4 kW m^{-2}.

4.5 THE OUTPUT SPECTRUM OF A GAS LASER

The output radiation from a gas laser is not actually at one single well-defined wavelength corresponding to the lasing transition, but covers a spectrum of wavelengths with a central peak. This is not a simple consequence of the Heisenberg uncertainty principle, but a direct result of the broadening of the emitted spectrum by the **Doppler effect**. We recall from the kinetic molecular theory that gas atoms are in random motion with an average kinetic energy of $(3/2)k_BT$. Suppose that these gas atoms emit radiation of frequency υ_o, which we label as the source frequency.

Then, due to the Doppler effect, when a gas atom is moving *away* from an observer, the latter detects a lower frequency υ_1 given by

$$\upsilon_1 = \upsilon_o\left(1 - \frac{v_x}{c}\right) \tag{4.5.1}$$

where v_x is the relative velocity of the atom along the laser tube (x-axis) with respect to the observer and c is the speed of light. When the atom is moving *toward* the observer, the detected frequency υ_2 is higher and corresponds to

$$\upsilon_2 = \upsilon_o\left(1 + \frac{v_x}{c}\right) \tag{4.5.2}$$

Since the atoms are in random motion the observer will detect a range of frequencies due to this Doppler effect. As a result, the frequency or wavelength of the output radiation from a gas laser will have a "linewidth" $\Delta\upsilon = \upsilon_2 - \upsilon_1$. This is what we mean by a **Doppler broadened linewidth** of a laser radiation. There are other mechanisms that also broaden the output spectrum but we will ignore these in the present case of gas lasers.

From the kinetic molecular theory we know that the velocities of gas atoms obey the Maxwell distribution. Consequently, the stimulated emission wavelengths in the lasing medium must exhibit a distribution about a central wavelength $\lambda_o = c/\upsilon_o$. Stated differently, the lasing medium therefore has an **optical gain** (or a photon gain) that has a distribution around $\lambda_o = c/\upsilon_o$ as shown in Figure 4.19 (a). The variation in the optical gain with the wavelength is called the **optical gain lineshape**. For the Doppler broadening case, this lineshape turns out to be a Gaussian function. For many gas lasers this spread in the frequencies from υ_1 to υ_2 is 1–5 GHz (for the He-Ne laser the corresponding wavelength spread is ~0.002 nm).

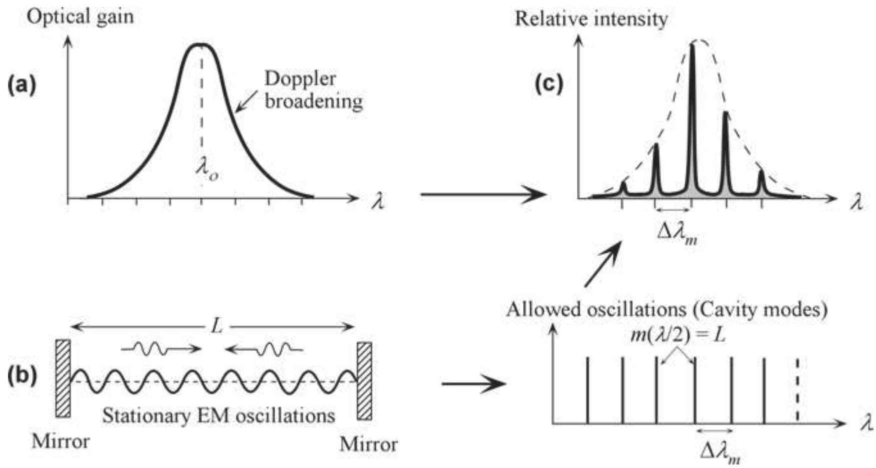


FIGURE 4.19 (a) Optical gain vs. wavelength characteristics (called the optical gain curve) of the lasing medium. (b) Allowed modes and their wavelengths due to stationary EM waves within the optical cavity. (c) The output spectrum (relative intensity vs. wavelength) is determined by satisfying (a) and (b) simultaneously, assuming no cavity losses.

When we consider the Maxwell velocity distribution of the gas atoms in the laser tube, we find that the linewidth $\Delta\upsilon_{1/2}$ between the half-intensity points (full width at half maximum—FWHM) in the output intensity vs. frequency spectrum is given by

Frequency linewidth (FWHM)

$$\Delta\upsilon_{1/2} = 2\upsilon_o\sqrt{\frac{2k_BT\ln(2)}{Mc^2}} \tag{4.5.3}$$

where M is the mass of the lasing atom or molecule (Ne in the He-Ne laser). The FWHM width $\Delta\upsilon_{1/2}$ is about 18% different than simply taking the difference $\upsilon_2 - \upsilon_1$ from Eqs. (4.5.1) and (4.5.2) and using a root-mean-square effective velocity along x, that is, using v_x in $\frac{1}{2}Mv_x^2 = \frac{1}{2}k_BT$. Equation (4.5.3) can be taken to be the FWHM width $\Delta\upsilon_{1/2}$ of the optical gain curve of most gas lasers. It does not apply to solid state lasers in which other broadening mechanisms operate.

Suppose that, for simplicity, we consider an optical cavity of length L with parallel end mirrors as shown in Figure 4.19 (b). Such an optical cavity is called a **Fabry–Perot optical resonator**. The reflections from the end mirrors of a laser give rise to traveling waves in opposite directions within the cavity. These oppositely traveling waves interfere constructively to set up a standing wave, that is, stationary electromagnetic (EM) oscillations as in Figure 4.19 (b). Some of the energy in these oscillations is tapped out by the 99% reflecting mirror to get an output, the same way we tap out the energy from an oscillating field in an LC circuit by attaching an antenna to it. Only standing waves with certain wavelengths, however, can be maintained within the optical cavity just as only certain acoustic wavelengths can be obtained from musical instruments. Any standing wave in the cavity must have an integer number of half-wavelengths, $\lambda/2$, that fit into the cavity length L

Laser cavity modes in a gas laser

$$m\left(\frac{\lambda}{2}\right) = L \tag{4.5.4}$$

where m is an integer that is called the **longitudinal mode number** of the standing wave. The wavelength λ in Eq. (4.5.4) is that within the cavity medium, but for gas lasers the refractive index is nearly unity and λ is the same as the free-space wavelength. Only certain wavelengths will satisfy Eq. (4.5.4), and for each m there will be an EM wave with a certain wavelength λ_m that satisfies Eq. (4.5.4).

Each possible standing wave within the laser tube (cavity) satisfying Eq. (4.5.4) is called a **cavity mode**. The cavity modes, as determined by Eq. (4.5.4), are shown in Figure 4.19 (b). Modes that exist along the cavity axis are called **axial** or **longitudinal modes**. Other types of modes, that is, stationary EM oscillations, are possible when the end mirrors are not flat. An example of an optical cavity formed by confocal spherical mirrors is shown in Figure 1.54 (Chapter 1). The EM radiation within such a cavity is a *Gaussian beam*.

The laser output thus has a broad spectrum with peaks at certain wavelengths corresponding to various cavity modes existing within the Doppler broadened optical gain curve as indicated in Figure 4.19 (c). At wavelengths satisfying Eq. (4.5.4), that is, representing certain cavity modes, we have spikes of intensity in the output. The net envelope of the output spectrum is a Gaussian distribution which is essentially due to the Doppler broadened linewidth. Notice that there is a finite width to the individual intensity spikes within the spectrum which is primarily due to nonidealities of the optical cavity, such as acoustic and thermal fluctuations of the cavity length L and nonideal end mirrors (less than 100% reflection). Typically, the frequency width of an individual spike in a He-Ne gas laser is ~1 MHz.

It is important to realize that even if the laser medium has an optical gain, the optical cavity will always have some losses inasmuch as some radiation will be transmitted through the mirrors, and there will be various losses such as scattering within the cavity. Only those modes that have an optical gain that can make up for the radiation losses from the cavity can exist in the output spectrum (as discussed later in Section 4.6).

EXAMPLE 4.5.1 Doppler broadened linewidth

Calculate the Doppler broadened linewidths $\Delta \upsilon$ and $\Delta \lambda$ (end-to-end of spectrum) for the He-Ne laser transition for $\lambda_o = 632.8\,\text{nm}$ if the gas discharge temperature is about 127°C. The atomic mass of Ne is 20.2 (g mol^{-1}). The laser tube length is 40 cm. What is the linewidth in the output wavelength spectrum? What is mode number m of the central wavelength, the separation between two consecutive modes, and how many modes do you expect within the linewidth $\Delta\lambda_{1/2}$ of the optical gain curve?

Solution

Due to the Doppler effect arising from the random motions of the gas atoms, the laser radiation from gaslasers is broadened around a central frequency υ_o. The central υ_o corresponds to the source frequency. Higher frequencies detected will be due to radiations emitted from atoms moving toward the observer, whereas lower frequency will be result of the emissions from atoms moving away from the observer. We will first calculate the frequency width using two approaches, one approximate and the other more accurate. Suppose that v_x is the root-mean-square (rms) velocity along the x-direction. We can intuitively expect the frequency width $\Delta\upsilon_{\text{rms}}$ between rms points of the Gaussian output frequency spectrum to be[18]

$$\Delta\upsilon_{\text{rms}} = \upsilon_o\left(1 + \frac{v_x}{c}\right) - \upsilon_o\left(1 - \frac{v_x}{c}\right) = \frac{2\upsilon_o v_x}{c} \tag{4.5.5}$$

We need to know the rms velocity v_x along x which is given by the kinetic molecular theory as $\frac{1}{2}Mv_x^2 = \frac{1}{2}k_BT$, where M is the mass of the atom. We can therefore calculate v_x. For the He-Ne laser, it is the Ne atoms that lase, so $M = (20.2 \times 10^{-3}\,\text{kg mol}^{-1})/(6.02 \times 10^{23}\,\text{mol}^{-1}) = 3.35 \times 10^{-26}\,\text{kg}$. Thus,

$$v_x = \left[(1.38 \times 10^{-23}\,\text{J K}^{-1})(127 + 273\,\text{K})/(3.35 \times 10^{-26}\,\text{kg})\right]^{1/2} = 405.9\,\text{m s}^{-1}$$

The central frequency is

$$\upsilon_o = c/\lambda_o = (3 \times 10^8\,\text{m s}^{-1})/(632.8 \times 10^{-9}\,\text{m}) = 4.74 \times 10^{14}\,\text{s}^{-1}$$

The rms frequency linewidth is approximately

$$\begin{aligned}\Delta\upsilon_{\text{rms}} &\approx (2\upsilon_o v_x)/c \\ &= 2(4.74 \times 10^{14}\,\text{s}^{-1})(405.9\,\text{m s}^{-1})/(3 \times 10^8\,\text{m s}^{-1}) = 1.28\,\text{GHz}\end{aligned}$$

The FWHM width $\Delta\upsilon_{1/2}$ of the output frequency spectrum will be given by Eq. (4.4.3)

$$\begin{aligned}\Delta\upsilon_{1/2} &= 2\upsilon_o\sqrt{\frac{2k_BT\ln(2)}{Mc^2}} = 2(4.74 \times 10^{14})\sqrt{\frac{2(1.38 \times 10^{-23})(400)\ln(2)}{(3.35 \times 10^{-26})(3 \times 10^8)^2}} \\ &= 1.51\,\text{GHz}\end{aligned}$$

which is about 18% wider than the estimate from Eq. (4.5.5).

---

[18] The fact that this is the width between the rms points of the Gaussian output spectrum can be shown from detailed mathematics.

To get FWHM wavelength width $\Delta\lambda_{1/2}$, differentiate $\lambda = c/\upsilon$

$$\frac{d\lambda}{d\upsilon} = -\frac{c}{\upsilon^2} = -\frac{\lambda}{\upsilon} \tag{4.5.6}$$

so that

$$\Delta\lambda_{1/2} \approx \Delta\upsilon_{1/2}|-\lambda/\upsilon| = (1.51 \times 10^9\ \text{Hz})(632.8 \times 10^{-9}\ \text{m})/(4.74 \times 10^{14}\ \text{s}^{-1})$$

or

$$\Delta\lambda_{1/2} \approx 2.02 \times 10^{-12}\ \text{m} \quad \text{or} \quad 2.02\ \text{pm}$$

This width is between the half-points of the spectrum. The rms linewidth would be 0.0017 nm. Each mode in the cavity satisfies $m(\lambda/2) = L$ and since L is some 4.7×10^5 times greater than λ, the mode number m must be very large. For $\lambda = \lambda_o = 632.8$ nm, the corresponding mode number m_o is

$$m_o = 2L/\lambda_o = (2 \times 0.4\ \text{m})/(632.8 \times 10^{-9}\ \text{m}) = 1{,}264{,}222.5$$

and actual m_o has to be the closest integer value, that is, 1,264,222 or 1,264,223.

The separation $\Delta\lambda_m$ between two consecutive modes (m and $m + 1$) is

$$\Delta\lambda_m = \lambda_m - \lambda_{m+1} = \frac{2L}{m} - \frac{2L}{m+1} \approx \frac{2L}{m^2}$$

or

Separation between modes

$$\Delta\lambda_m \approx \frac{\lambda_o^2}{2L} \tag{4.5.7}$$

Substituting the values, we find $\Delta\lambda_m = (632.8 \times 10^{-9})^2/(2 \times 0.4) = 5.01 \times 10^{-13}$ m or 0.501 pm.

We can also find the separation of the modes by noting that Eq. (4.5.4), in which $\lambda = c/\upsilon$, is equivalent to

Frequency of a mode

$$\upsilon = \frac{mc}{2L} \tag{4.5.8}$$

so that the separation of modes in frequency, $\Delta\upsilon_m$, is simply

Frequency separation of modes

$$\Delta\upsilon_m = \frac{c}{2L} \tag{4.5.9}$$

Substituting $L = 0.40$ m in Eq. (4.5.9), we find $\Delta\upsilon_m = 375$ MHz. (A typical value for a He-Ne laser.)

The number of modes, *i.e.,* the number of m values, within the linewidth, that is, between the half-intensity points, will depend on how the cavity modes and the optical gain curve coincide, for example, whether there is a cavity mode right at the peak of the optical gain curve as illustrated in Figure 4.20. Suppose that we try to estimate the number of modes by using,

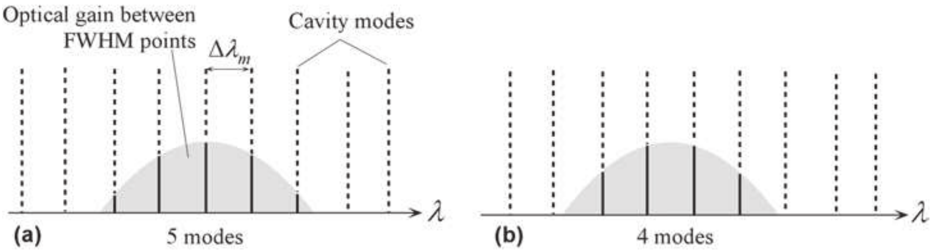


FIGURE 4.20 Number of laser modes depends on how the cavity modes intersect the optical gain curve. In this case we are looking at modes within the linewidth $\Delta\lambda_{1/2}$.

$$\text{Modes} = \frac{\text{Linewidth of spectrum}}{\text{Separation of two modes}} \approx \frac{\Delta\lambda_{1/2}}{\Delta\lambda_m} = \frac{2.02\ \text{pm}}{0.501\ \text{pm}} = 4.03$$

We should expect at most 4 to 5 modes within the linewidth of the output as shown in Figure 4.20. We neglected the cavity losses.

4.6 LASER OSCILLATIONS: THRESHOLD GAIN COEFFICIENT AND GAIN BANDWIDTH

A. Optical Gain Coefficient g

Consider a general laser medium that has an optical gain for coherent radiation along some direction x as shown in Figure 4.21 (a). This means that the medium is appropriately pumped. Consider an electromagnetic wave propagating in the medium along the x-direction. As it propagates, its power (energy flow per unit time) increases due to greater stimulated emissions over spontaneous emissions and absorption across the same two energy levels $E_2 - E_1$ as in Figure 4.21 (a). If the light intensity were decreasing, we would have used a factor $\exp(-\alpha x)$, where α is the attenuation coefficient, to represent the power loss along the distance x. Similarly, we represent the power increase along x by using a factor $\exp(gx)$, where g is a coefficient called the **optical gain coefficient** of the medium. The gain coefficient g is defined as the fractional increase in the light power (or intensity) per unit distance along the direction of propagation. We have already encountered g in Eq. (4.2.17). Optical power along x at any point is proportional to the concentration of coherent photons N_{ph} and the photon energy $h\upsilon$. These coherent photons travel with a velocity c/n, where n is the refractive index.[19] Thus, in time δt they travel a distance $\delta x = (c/n)\delta t$ in the tube. The optical power P is proportional to N_{ph} so that the fractional increase in P per unit distance, which is the gain coefficient g, is given by

$$g = \frac{\delta P}{P\delta x} = \frac{\delta N_{ph}}{N_{ph}\delta x} = \frac{n}{cN_{ph}}\frac{\delta N_{ph}}{\delta t} \qquad \textbf{(4.6.1)}$$

Optical gain coefficient

The gain coefficient g describes the increase in intensity of the lasing radiation in the cavity per unit length due to stimulated emission transitions from E_2 to E_1 exceeding photon

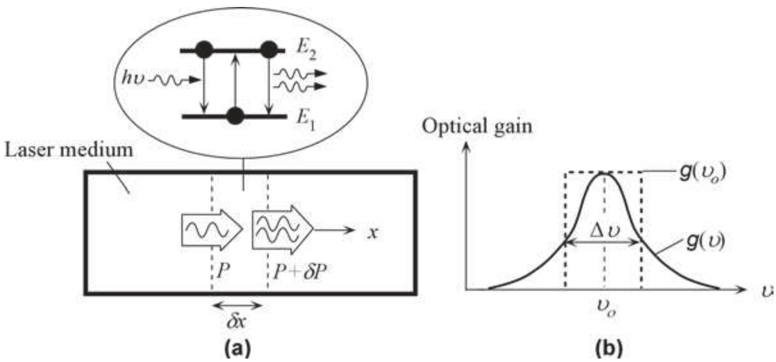


FIGURE 4.21 (a) A laser medium with an optical gain. (b) The optical gain curve of the medium. The dashed line is the approximate derivation in the text.

[19] In semiconductor-related chapters, n is the electron concentration and n is the refractive index; E is the energy and E is the electric field.

absorption across the same two energy levels. We know that the difference between stimulated emission and absorption rates [see Eqs. (4.2.1) and (4.2.2)] gives the *net* rate of change in the coherent photon concentration, that is,

$$\begin{aligned}\frac{dN_{\mathrm{ph}}}{dt} &= \text{Net rate of stimulated photon emission}\\ &= N_2B_{21}\rho(\upsilon) - N_1B_{21}\rho(\upsilon) = (N_2 - N_1)B_{21}\rho(\upsilon)\end{aligned} \tag{4.6.2}$$

It is now straightforward to obtain the optical gain by using Eq. (4.6.2) in Eq. (4.6.1) with certain assumptions. As we are interested in the amplification of coherent waves traveling along a defined direction (x) in Figure 4.21 (a), we can neglect spontaneous emissions which are in random directions and do not, on average, contribute to the directional wave. The reader should have noticed the similarity between Eq. (4.6.2) and Eq. (4.2.18).

Normally, the emission and absorption processes occur not at a discrete photon energy $h\upsilon$, but they would be distributed in photon energy or frequency over some frequency interval $\Delta\upsilon$. The spread $\Delta\upsilon$, for example, can be due to Doppler broadening or broadening of the energy levels E_2 and E_1. There may be other broadening processes. In any event, this means that the optical gain will reflect this distribution, that is, $g = g(\upsilon)$ as illustrated in Figure 4.21 (b). The spectral shape of the gain curve is called the **lineshape function**.

We can express $\rho(\upsilon)$ in terms of N_{ph} by noting that $\rho(\upsilon)$ is defined as the radiation energy density per unit frequency so that at υ_o

$$\rho(\upsilon_o) \approx \frac{N_{\mathrm{ph}}h\upsilon_o}{\Delta\upsilon} \tag{4.6.3}$$

We can now substitute for dN_{ph}/dt in Eq. (4.6.1) from Eq. (4.6.2) and use Eq. (4.6.3) to obtain the *optical gain coefficient*

General optical gain coefficient

$$g(\upsilon_o) \approx (N_2 - N_1)\frac{B_{21}nh\upsilon_o}{c\Delta\upsilon} \tag{4.6.4}$$

Equation (4.6.4) gives the optical gain of the medium at the center frequency υ_o. A more rigorous derivation would have found the optical gain curve as a function of frequency, shown as $g(\upsilon)$ in Figure 4.21 (b), and would derive $g(\upsilon_o)$ from this lineshape.[20]

B. Threshold Gain Coefficient g_{th} and Output Power

Consider an optical cavity with mirrors at the ends, such as the Fabry–Perot (FP) optical cavity shown in Figure 4.22. The cavity contains a laser medium so that lasing emissions build up to a steady state, that is, there is continuous operation. We effectively assume that we have stationary electromagnetic (EM) oscillations in the cavity and that we have reached steady state. The optical cavity acts as an optical resonator. Consider an EM wave with an *initial* optical power P_i starting at some point in the cavity and traveling toward the cavity face 1 as shown in Figure 4.22. It will travel the length of the cavity, become reflected at face 1, travel back the length of the cavity to face 2, become reflected at 2, and arrive at the starting point with a *final* optical power P_f. Under steady state conditions, oscillations do not build up and do not

[20] Commonly known as Füchtbauer–Ladenburg relation, and described in more advanced textbooks.

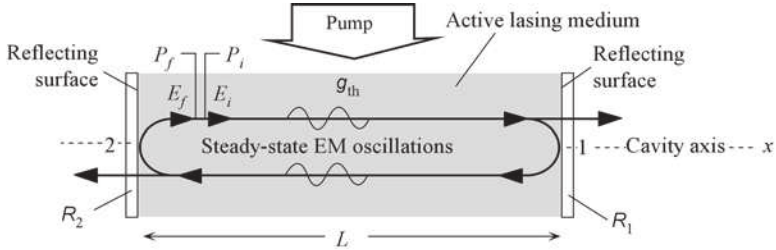


FIGURE 4.22 An optical cavity resonator, and the derivation of the threshold gain condition.

die out, which means that P_f must be the same as P_i. Thus, there should be no optical power loss in the round trip, which means that the **net round-trip optical gain** G_{op} must be unity,

Power condition for maintaining oscillations

$$G_{op} = \frac{P_f}{P_i} = 1 \tag{4.6.5}$$

Reflections at the faces 1 and 2 reduce the optical power in the cavity by the reflectances R_1 and R_2 of the faces. There are other losses such as some absorption and scattering during propagation in the medium. All these losses have to be made up by stimulated emissions in the optical cavity which effectively provides an optical gain in the medium. As the wave propagates, its power increases as $\exp(gx)$. However, there are a number of losses in the cavity medium acting against the stimulated emission gain such as light *scattering* at defects and inhomogeneities, absorption by impurities, absorption by free carriers (important in semiconductors) and other loss phenomena. These internal cavity losses decrease the power as $\exp(-\alpha_s x)$ where α_s is the **attenuation** or **loss coefficient of the medium**, also called the **internal cavity loss coefficient**. α_s represents all losses in the cavity and its walls, *except* light transmission losses though the end mirrors and absorption across the energy levels involved in stimulated emissions (E_1 and E_2), which is incorporated into g.[21]

The power P_f of the EM radiation after one round trip of path length $2L$ (Figure 4.22) is given by

$$P_f = P_i R_1 R_2 \exp\left[g(2L)\right]\exp\left[-\alpha_s(2L)\right] \tag{4.6.6}$$

For steady state oscillations, Eq. (4.6.5) must be satisfied, and the value of the gain coefficient g that makes $P_f/P_i = 1$ is called the **threshold gain coefficient** g_{th}. From Eq. (4.6.6)

Threshold optical gain coefficient

$$g_{th} = \alpha_s + \frac{1}{2L}\ln\left(\frac{1}{R_1R_2}\right) = \alpha_t \tag{4.6.7}$$

where α_t is the **total loss coefficient** that represents all the losses in the second term. Equation (4.6.7) gives the optical gain needed in the medium to achieve a continuous wave lasing emission. The right-hand side of Eq. (4.6.7) represents all the losses, that is, the internal cavity losses, α_s, and end losses in $(1/2L)\ln(R_1R_2)^{-1}$. Threshold gain coefficient g_{th} is that value of the optical gain coefficient that just overcomes all losses, $g_{th} = \alpha_t$.

[21] α_s should *not* be confused with the natural absorption coefficient α. The latter represents the absorption from E_1 to E_2 and is already incorporated into g. Note also that some authors use γ for g and some use γ for α_s.

The necessary g_{th} as required by Eq. (4.6.3), that is, the threshold gain, has to be obtained by suitably pumping the medium so that N_2 is sufficiently greater than N_1. This corresponds to a **threshold population inversion** or $N_2 - N_1 = (N_2 - N_1)_{th}$. From Eq. (4.6.4)

Threshold population inversion

$$(N_2 - N_1)_{th} \approx g_{th} \frac{c\Delta v}{B_{21} n h v_o} \qquad \textbf{(4.6.8)}$$

We can now substitute for B_{21} in terms of A_{21} (which can be determined experimentally) from Eq. (4.2.7), *i.e.,* $A_{21}/B_{21} = 8\pi h v^3/c^3$, and also use $A_{21} = 1/\tau_{sp}$, where τ_{sp} is the spontaneous decay time

Threshold population inversion

$$(N_2 - N_1)_{th} \approx g_{th} \frac{8\pi n^2 v_o^2 \tau_{sp} \Delta v}{c^2} \qquad \textbf{(4.6.9)}$$

Initially the medium must have a gain coefficient g greater than g_{th}. This allows the oscillations to build up in the cavity until a steady state is reached when $g = g_{th}$. By analogy, an electrical oscillator circuit has an overall gain (loop gain) of unity once a steady state is reached and oscillations are maintained. Initially, however, when the circuit is just switched on, the overall gain is greater than unity. The oscillations start from a small noise voltage and become amplified, that is built-up, until the overall round-trip (or loop) gain becomes unity and a steady state operation is reached. The reflectance of the mirrors R_1 and R_2 are important in determining the threshold population inversion as they control g_{th} in Eq. (4.6.9). It should be apparent that the laser device emitting coherent emission is actually a **laser oscillator**.

The examination of the steady state continuous wave (CW) coherent radiation output power P_o and the population difference $(N_2 - N_1)$ in a laser as a function of the pump rate would reveal the highly simplified behavior shown in Figure 4.23. Until the pump rate can bring $(N_2 - N_1)$ to the threshold value $(N_2 - N_1)_{th}$, there would be no coherent radiation output. When the pumping rate exceeds the threshold value, then $(N_2 - N_1)$ remains clamped at $(N_2 - N_1)_{th}$ because this controls the optical gain g, which must remain at g_{th}. Additional pumping increases the *rate of stimulated transition*s and hence increases the photon concentration and the optical output power P_o. Also note that we have not considered how pumping actually modifies N_1 and N_2, except that $(N_2 - N_1)$ is proportional to the pumping rate as in Figure 4.23.[22] Further, the characteristics shown in Figure 4.23 refer to plotting values under steady state or CW operation, and exclude the fact that initially $(N_2 - N_1)$ must be greater than $(N_2 - N_1)_{th}$ for the oscillations to build-up.

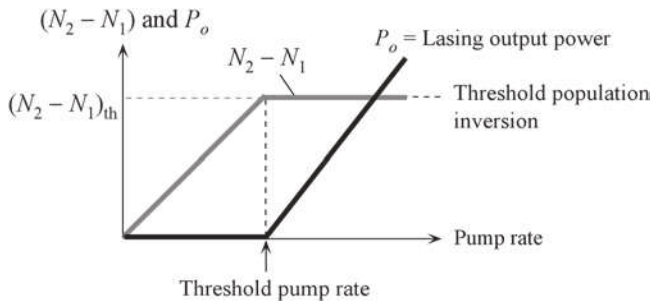


FIGURE 4.23 A simplified description of a laser oscillator. $(N_2 - N_1)$ and coherent output power (P_o) vs. pump rate under continuous wave steady state operation.

[22]To relate N_1 and N_2 to the pumping rate we have to consider the actual energy levels involved in the laser operation (three or four levels) and develop *rate equations* to describe the transitions in the system; see, for example, J. Wilson and J. F. B. Hawkes, *Optoelectronics: An Introduction*, 3rd Edition (Prentice-Hall, Pearson Education, 1998), Ch. 5.

EXAMPLE 4.6.1 Threshold population inversion for the He-Ne laser

Consider a He-Ne gas laser operating at the wavelength 632.8 nm (equivalent to $\upsilon_o = 473.8\,\text{THz}$). The tube length $L = 40$ cm and mirror reflectances are approximately 95% and 100%. The linewidth $\Delta\upsilon$ is 1.5 GHz, the loss coefficient α_s is $0.05\,\text{m}^{-1}$, the spontaneous decay time constant τ_{sp} is roughly 100 ns, and $n \approx 1$. What are the threshold gain coefficient and threshold population inversion?

Solution

The threshold gain coefficient from Eq. (4.6.7) is

$$g_{th} = \alpha_s + \frac{1}{2L}\ln\left(\frac{1}{R_1R_2}\right) = (0.05\text{ m}^{-1}) + \frac{1}{2(0.4\text{ m})}\ln\left[\frac{1}{(0.95)(1)}\right] = 0.114\text{ m}^{-1}$$

The threshold population inversion from Eq. (4.6.9) is

$$\begin{aligned}\Delta N_{th} &\approx g_{th}\frac{8\pi n^2\upsilon_o^2\tau_{sp}\Delta\upsilon}{c^2}\\ &= (0.114\text{ m}^{-1})\frac{8\pi(1)^2(473.8\times10^{12}\text{ s}^{-1})^2(100\times10^{-9}\text{ s})(1.5\times10^9\text{ s}^{-1})}{(3\times10^8\text{ m s}^{-1})^2}\\ &= 1.1\times10^{15}\,\text{m}^{-3}\end{aligned}$$

Note that this is the threshold population inversion for Ne atoms in configurations $2p^55s^1$ and $2p^53p^1$. The spontaneous decay time τ_{sp} is the natural decay time of Ne atoms from E_2 ($2p^55s^1$) to E_1 ($2p^53p^1$). This time must be much longer than the spontaneous decay from E_1 to lower levels to allow a population inversion to be built up between E_2 and E_1, which is the case in the He-Ne laser. Equation (4.6.8) was a simplified derivation that used two energy levels: E_2 and E_1. The He-Ne case is actually more complicated because the excited Ne atom can decay from E_2 not only to E_1 but to other lower levels as well. While the He-Ne is lasing, the optical cavity ensures that the photon density in cavity promotes the E_2 to E_1 transitions to maintain the lasing operation.

C. Output Power and Photon Lifetime in the Cavity

Consider the laser cavity shown in Figure 4.24 that has been pumped so that it is operating under steady state, and emitting CW radiation. Under steady state lasing operation, the coherent photon concentration N_{ph} would neither decay with distance nor with time inside the cavity,

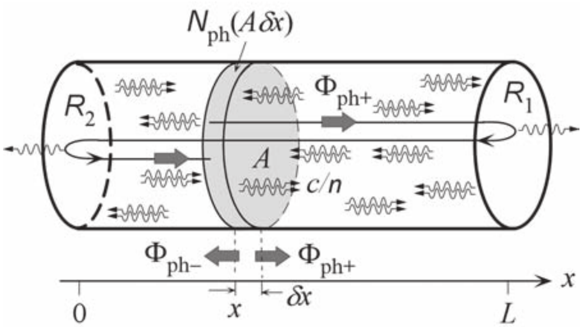


FIGURE 4.24 A pictorial visualization of photons inside a laser optical cavity bouncing back and forth between the cavity ends with some being transmitted. N_{ph} is the *photon concentration* inside the cavity. Φ_{ph+} is the *photon flux* in the $+x$ direction. (Schematic illustration only. In reality, the photons in the cavity are much longer.)

inasmuch as g_{th} just balances the total loss α_t. Further, there will be a coherent photon flux along the $+x$ direction Φ_{ph+} and another flux Φ_{ph-} along $-x$. In the small elemental volume $A\delta x$ shown in Figure 4.24, *half* of $N_{ph}(A\,\delta x)$ is moving toward the $+x$ direction with a velocity c/n. The time δt it takes for photons to travel δx is $\delta t = n\delta x/c$. The photon flux Φ_{ph+} is therefore $(1/2)\,N_{ph}(A\,\delta x)/(A\,\delta t)$ or

Photon flux in $+x$ direction

$$\Phi_{ph+} = \frac{1}{2}N_{ph}(c/n) \tag{4.6.10}$$

The power flowing toward the $+x$ direction is simply $h\upsilon_o A\Phi_{ph+}$. A fraction $(1 - R_1)$ will be transmitted at the right-hand-side mirror so that the output power is

Output power

$$P_o \approx \frac{1}{2}A(1 - R_1)h\upsilon_o N_{ph}(c/n) \tag{4.6.11}$$

Thus, the output power, as expected, depends on the photon concentration N_{ph} in the cavity, and the latter depends on the rate of stimulated transitions.

The total cavity loss coefficient $\alpha_t = \alpha_s + (1/2L)\ln(R_1R_2)^{-1}$ has two parts, in which α_s is the internal loss coefficient and $(1/2L)\ln(R_1R_2)^{-1}$ represents the **end losses**. These losses would cause the photon concentration N_{ph} to decay exponentially with time with a characteristic time constant, called the **photon cavity lifetime** τ_{ph}, that depends inversely on α_t. Consider the photon flux Φ_{ph+} shown in Figure 4.24 that starts at x and is traveling toward the $+x$ direction. It becomes reflected at R_1 and then at R_2, and then it returns back to x with a smaller magnitude as we are considering just the effect of all losses represented by α_t. The round-trip time interval $\delta t = 2nL/c$. The attenuated photon flux after one round trip is $\Phi_{ph}R_1R_2\exp(-2\alpha_sL)$ and the change $\delta\Phi_{ph+}$ is $-\Phi_{ph}\left[1 - R_1R_2\exp(-2\alpha_sL)\right]$, where the negative sign indicates a decrease. Further, we can eliminate α_s and R_1R_2 by using the total attenuation coefficient α_t so that

$$\delta\Phi_{ph+} = -\Phi_{ph}\left[1 - \exp(-2\alpha_tL)\right]$$

The fractional change in the photon concentration per unit time is then

Photon cavity lifetime

$$\frac{\delta N_{ph}}{N_{ph}\delta t} = \frac{\delta\Phi_{ph+}}{\Phi_{ph+}\delta t} = -\frac{1 - \exp(-2\alpha_tL)}{2Ln/c} = -\frac{1}{\tau_{ph}} \tag{4.6.12}$$

Clearly, if we ignore the gain of the medium, the above equation shows that losses would cause the photon concentration N_{ph} to decay exponentially with time as $N_{ph}(t) = N_{ph}(0)\exp(-t/\tau_{ph})$.[23] Assuming small losses, we can expand $\exp(-2\alpha_tL) \approx 1 - 2\alpha_tL$ so that the **photon cavity lifetime** is given by

Photon cavity lifetime

$$\tau_{ph} \approx n/c\alpha_t \tag{4.6.13}$$

Thus, the total attenuation in the optical cavity, as represented by the coefficient α_t, is equivalent to a cavity photon concentration that decays exponentially with time with a time constant τ_{ph} given by Eq. (4.6.13).

[23] If the change δN_{ph} is small over the time δt, then we can use differentials for changes. Further, most books simply use $\tau_{ph} = n/c\alpha_t$ for the photon cavity lifetime.

EXAMPLE 4.6.2 Output power and photon cavity lifetime τ_{ph}

Consider the He-Ne laser in Example 4.6.1 that has a tube length of 40 cm and $R_1 = 0.95$ and $R_2 = 1$. Suppose that the tube diameter is 0.8 mm, and the output power is 2.5 mW. What are the photon cavity lifetime and the photon concentration inside the cavity? (The emission frequency υ_o is 474 THz.)

Solution

Using $L = 40$ cm, $R_1 = 0.95$, $R_2 = 1$, $\alpha_s = 0.05\ \text{m}^{-1}$ gives

$$\alpha_t = \alpha_s + (1/2L)\ln(R_1R_2)^{-1} = 0.05\ \text{m}^{-1} + \left[2(0.4\ \text{m})\right]^{-1}\ln\left[(0.95 \times 1)\right]^{-1} = 0.114\ \text{m}^{-1}$$

and hence from Eq. (4.6.12),

$$\tau_{ph} = \left[(2)(1)(0.4)\right]/\left[(3 \times 10^8)(1 - e^{-2\times0.114\times0.4})\right] = 30.6\ \text{ns}$$

If we use Eq. (4.6.13) we would find 29.2 ns. To find the photon concentration, we use Eq. (4.6.11),

$$P_o = (0.0025\ \text{W}) \approx \frac{1}{2}A(1 - R_1)h\upsilon_o N_{ph}c/n$$

$$= \frac{1}{2}\left[\pi(8 \times 10^{-3}/2)^2\right](1 - 0.95)(6.62 \times 10^{-34})(474 \times 10^{12})N_{ph}(3 \times 10^8)/(1)$$

which gives $N_{ph} \approx 2.1 \times 10^{15}$ photons m^{-3}.

D. Optical Cavity, Phase Condition, Laser Modes

The laser oscillation condition stated in Eq. (4.6.5) which leads to the threshold gain g_{th} in Eq. (4.6.7) considers only the intensity of the radiation inside the cavity. The examination of Figure 4.22 reveals that the initial wave E_i with power P_i attains a power P_o after *one round trip* when the wave has arrived back exactly at the same position as E_f, as shown in Figure 4.22. Unless the total phase change after one round trip from E_i to E_f is a multiple of 2π, the wave E_f cannot be identical to the initial wave E_i. We therefore need the additional condition that the round-trip phase change $\Delta\phi_{\text{round-trip}}$ must be a multiple of 360°

Phase condition for laser oscillations

$$\Delta\phi_{\text{round-trip}} = m(2\pi) \tag{4.6.14}$$

where m is an integer, 1, 2,…. This condition ensures *self-replication* rather than self-destruction. There are various factors that complicate any calculation from the phase condition in Eq. (4.6.14). The refractive index n of the medium in general will depend on pumping (especially so in semiconductors), and the end-reflectors can also introduce phase changes. In the simplest case, we can assume that n is constant and neglect phase changes at the mirrors. If $k = 2\pi/\lambda$ is the free-space propagation constant, only those special k-values, denoted as k_m, that satisfy Eq. (4.6.14) can exit as radiation in the cavity, *i.e.,* for propagation along the cavity axis

Laser cavity modes

$$nk_m(2L) = m(2\pi) \tag{4.6.15}$$

which leads to the usual mode condition

Wavelengths of laser cavity modes

$$m\left(\frac{\lambda_m}{2n}\right) = L \tag{4.6.16}$$

Thus, our intuitive representation of modes as standing waves, as originally described by Eq. (4.5.4) and again by Eq. (4.6.16), is a simplified conclusion from the general phase condition in Eq. (4.6.14). Furthermore, the modes in Eq. (4.6.16) are controlled by the length L of the optical cavity along its axis and are called **longitudinal axial modes**. We can also write the modes in terms of their frequencies υ_m inasmuch as $\lambda_m = c/\upsilon_m$

Frequencies of laser cavity modes

$$\upsilon_m = \frac{mc}{2nL} \quad \textbf{(4.6.17)}$$

and the separation of modes can be found from Eq. (4.6.17), $\Delta\upsilon_m = c/2nL$.

In the discussions of threshold gain and phase conditions, we referred to Figure 4.22 and tacitly assumed plane EM waves traveling inside the cavity between two perfectly flat and aligned mirrors. A plane wave is an idealization as it has an infinite extent over the plane normal to the direction of propagation. All practical laser cavities have a finite transverse size, that is, perpendicular to the cavity axis. Furthermore, not all cavities have flat reflectors at the ends. In gas lasers, one or both mirrors at the tube ends may be spherical to allow a better mirror alignment as illustrated in Figures 4.25 (a) and (b). One can easily visualize off-axis self-replicating rays that can travel off the axis as shown in one example in Figure 4.25 (a). Such a mode would be non-axial. Its properties would be determined not only by the off-axis round-trip distance, but also by the transverse size of the cavity. The greater the transverse size, the more of these off-axis modes can exist. Further, notice that the ray in Figure 4.25 (a) traverses the tube length almost four times before it completes a true round trip, whereas the total round-trip path was $2L$ in the case of flat mirrors.

A better way of thinking about modes is to realize that a mode represents a particular electric field pattern in the cavity that can replicate itself after one round trip. Figure 4.25 (b) shows how a wavefront of a particular mode starts parallel to the surface of one of the mirrors, and after one round trip, it replicates itself. The wavefront curvature changes as the radiation propagates in the cavity and it is parallel to mirror surfaces at the end mirrors. Such a mode has similarities to the Gaussian beam discussed in Chapter 1. Indeed, in many applications, laser beams are modeled by assuming they are Gaussian beams.

More generally, whether we have flat or spherical end mirrors, we can find all possible allowed modes by considering what spatial field patterns at one mirror can self-replicate

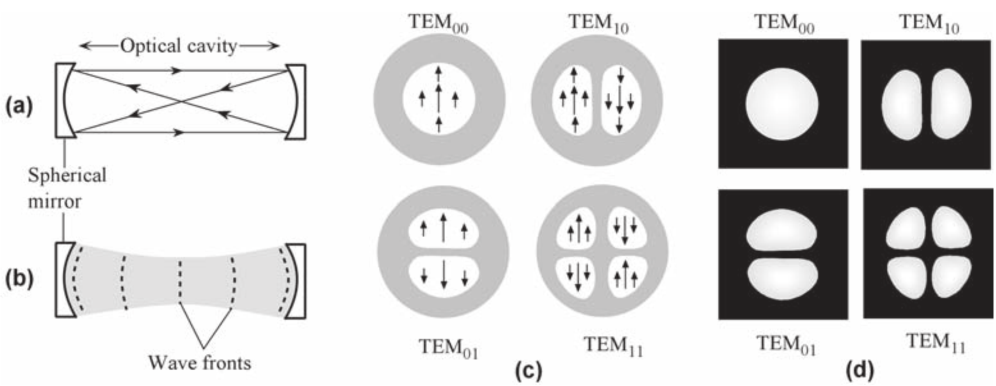


FIGURE 4.25 Laser modes. (a) An off-axis transverse mode is able to self-replicate after one round trip. (b) Wavefronts in a self-replicating wave. (c) Four low-order transverse cavity modes and their fields. (d) Intensity patterns in the modes of (c).

after one round trip[24] through the cavity to the other mirror and back as in the example in Figure 4.25 (b). *A mode is a certain field pattern at a reflector that can propagate to the other reflector and back again and return the same field pattern.* All these modes can be represented by fields **E** and **B** that are nearly normal to the cavity axis; they are referred to as **transverse modes** or **transverse electric and magnetic (TEM) modes**.[25] Each allowed mode corresponds to a distinct spatial field distribution at a reflector. These modal field patterns at a reflector can be described by three integers, p, q, and m, and designated by TEM_{pqm}. The integers p, q represent the number of nodes in the field distribution along the transverse directions y and z to the cavity axis x (put differently, across the beam cross-section). The integer m is the number of nodes along the cavity axis x and is the usual longitudinal mode number. Figures 4.25 (c) and (d) shows the field patterns for four TEM modes and the corresponding intensity patterns for four example TEM modes. Each transverse mode with a given p, q has a set of longitudinal modes (m values) but usually m is very large ($\sim 10^6$ in gas lasers) and is not written, though understood. Thus, transverse modes are written as TEM_{pq} and each has a set of longitudinal modes ($m = 1, 2, \ldots$). Moreover, two different transverse modes may not necessarily have the same longitudinal frequencies implied by Eq. (4.6.7). (For example, n may not be spatially uniform and different TEM modes have different spatial field distributions.)

Transverse modes depend on the optical cavity dimensions, reflector sizes, and other size-limiting apertures that may be present in the cavity. The modes either have **Cartesian** (**rectangular**) or **polar** (**circular**) symmetry about the cavity axis. Cartesian symmetry arises whenever a feature of the optical cavity imposes a more favorable field direction; otherwise, the patterns exhibit circular symmetry. The examples in Figures 4.25 (c) and (d) possess rectangular symmetry and would arise, for example, if polarizing Brewster windows are present at the ends of the cavity.

The lowest order mode TEM_{00} has an intensity distribution that is radially symmetric about the cavity axis, and has a Gaussian intensity distribution across the beam cross-section everywhere inside and outside cavity. It also has the lowest divergence angle. These properties render the TEM_{00} mode highly desirable, and many laser designs optimize on TEM_{00} while suppressing other modes. Such lasers usually require restrictions in the transverse size of the cavity.

4.7 BROADENING OF THE OPTICAL GAIN CURVE AND LINEWIDTH

If the upper and lower energy levels E_2 and E_1 in laser transitions were truly discrete levels, we might think that we would get a stimulated emission at one frequency only at $h\upsilon_o = E_2 - E_1$, and the optical gain curve would be a delta function, a narrow line at υ_o. At least in terms of modern physics, we know that this would be impossible based on the Uncertainty Principle alone. Suppose that we wish to measure the energy E_2 by some perfectly designed instrument. Suppose, further, that lifetime of the atom in this E_2-state is τ_2 after which it ends up at E_1. Then, according to the Uncertainty Principle, if ΔE_2 is the uncertainty in E_2, we must obey $\Delta E_2 \tau_2 > \hbar$, which clearly shows that we cannot think of E_2 as a well-defined discrete single-valued energy level. Consequently the emitted radiation will have this finite minimum spectral width $h\Delta\upsilon = \Delta E_2$.

[24] We actually have to solve Maxwell's equations with the boundary conditions of the cavity to determine what EM wave patterns are allowed. Further we have to incorporate optical gain into these equations (not a trivial task).

[25] Or, *transverse electromagnetic modes.*

Exactly the same Uncertainty Principle argument applies to E_1, and if τ_1 is the lifetime of the atom in a state at E_1, from which it decays to another lower energy state, then we must obey $\Delta E_1 \tau_1 > \hbar$. Thus, the emission spectrum involves transitions from a small region ΔE_2 around E_2 to a small region ΔE_1 around E_1, and consequently has a finite spectral width determined by τ_2 and τ_1. This type of broadening of the emission spectrum, and hence the optical gain curve (its lineshape), is called **lifetime broadening**.

In lifetime broadening, all atoms in the medium will exhibit the same broadening. They will all have the same central frequency υ_o and spectral width $\Delta\upsilon$, as shown in Figure 4.26 (a). This type of spectral broadening in which *all* the atoms in a medium generate the same emission curve with the same υ_o and subjected to the same broadening mechanism is called **homogeneous broadening**. The shape of the emission curve is called a **Lorentzian lineshape**, and its mathematical form is well-known in photonics. (See Question 4.18.) The spectral width due to lifetime broadening is known as the **natural linewidth**, which is the minimum linewidth exhibited by all atoms.

Another homogeneous broadening mechanism that is of particular practical importance is the emission from ions embedded in crystals, for example, in a laser based on Nd^{3+} ions in a YAG crystal, or Ti^{3+} ions in an Al_2O_3 (sapphire) crystal. The Nd^{3+} ions in the YAG crystal ($Y_3Al_5O_{12}$) will be situated at well-defined sites, and their emission characteristics would be well-defined in the absence of any interruptions. However, lattice vibrations, *i.e.,* phonons, will interact, that is *collide*, with the Nd^{3+} ions (phonons will "jolt" the Nd^{3+} ions) and interrupt the emission process. Phonon collisions suddenly change the phase of the EM wave during emission. This type of broadening mechanism is called **collision broadening**, which is also homogeneous, because all Nd^{3+} ions experience the same broadening. Phonons interact with all the ions in the same way. Collision broadening is much stronger than lifetime broadening. The spectral width of the Nd^{3+}:YAG laser is 120 GHz and is due to phonon collision broadening.

Homogeneous collision broadening can also take place in gas lasers in which the atoms collide with each other and interrupt the radiation emission process, either terminate it or suddenly change the phase of the emitted radiation. In the case of gas lasers, collision broadening is usually called **pressure broadening** because the broadening increases with the gas pressure as more and more atoms collide with each other. While for the He-Ne laser, pressure broadening is negligible compared with Doppler broadening, it is nonetheless a contributing factor in certain gas lasers such as the CO_2 laser.

We saw that in the He-Ne laser, individual Ne atoms emitted at different frequencies due to the Doppler effect as illustrated in Figure 4.26 (b). If they were all stationary they would all have the same emission characteristic, *i.e.,* a homogeneously broadened lineshape centered at υ_o; but, they move around randomly in the gas. The Doppler effect shifts the emission curve of each Ne atom an amount that depends on the velocity of the Ne atom and its direction. Since the atoms are

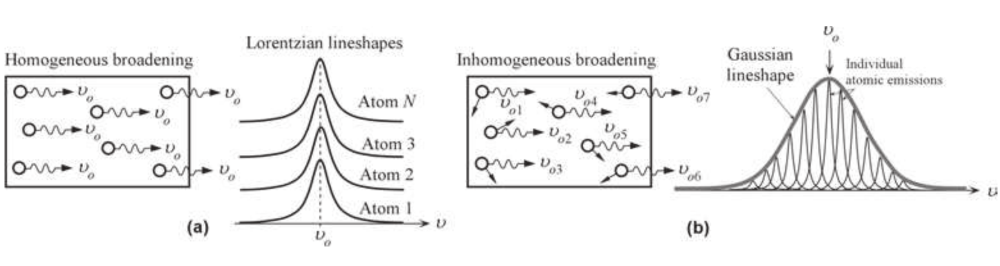


FIGURE 4.26 Lineshapes for (a) homogeneous and (b) inhomogeneous broadening. The lineshape is Lorentzian for homogeneous and Gaussian for inhomogeneous broadening.

moving randomly, the shift can be considered to be random. With a large number of Ne atoms, these different randomly shifted emission curves add to produce an overall emission lineshape that is a **Gaussian**. (The Gaussian lineshape is not unexpected since we have a very large collection of "random processes.") This type of broadening in which each atom emits at a slightly different central frequency is called **inhomogeneous broadening**. The overall spectral width is obviously much wider than the homogeneous broadening inherent in individual atomic emissions as indicated in Figure 4.26 (b). In the case of 632.8 nm emission, homogeneous lifetime broadening is only 14 MHz, whereas inhomogeneous Doppler broadening is 1.5 GHz, two orders of magnitude wider.

Inhomogeneous broadening also occurs in solid state glass lasers, for example, in Nd^{3+}: glass lasers. Each isolated Nd^{3+} ion outside the glass structure would emit the same spectrum. However, within the glass structure, the Nd^{3+} ions find themselves in different local environments, with different neighboring ions. These neighboring ions are not always the same and always exactly at the same place for each Nd^{3+} ion. These variations in the local environment from site to site in the glass result in the energy levels of the Nd^{3+} ion becoming spread. The overall emission characteristic is a Gaussian lineshape because the variations from one Nd^{3+} site to another are random. This type of inhomogeneous broadening is often termed **amorphous structure broadening**. The spectral broadening in Nd^{3+}:glass lasers are typically 5–7 THz, which is much wider than that for Nd^{3+}:YAG above.

Table 4.2 summarizes the basic characteristics of homogeneous and inhomogeneous broadening and provides simple examples.[26] The optical gain lineshape and its width are key parameters in laser design, especially so in mode-locked lasers described in the next section.

The lasing radiation output from a perfect homogeneously broadened laser medium is normally a single mode. On the other hand, the lasing output from an inhomogeneously broadened medium can be multimode or single mode, depending on the overlap of the gain curve with the cavity modes.

Suppose that we manage to pump a homogeneously broadened gain medium above threshold gain g_{th} for an instant as shown in Figure 4.27 (a). Suppose that there are three modes of the cavity that are allowed under the gain curve above g_{th}, marked at υ_o, υ_1, and υ_1' in Figure 4.27 (a).

TABLE 4.2 Comparison of homogeneous and inhomogeneous broadening mechanisms on the emission spectrum, or the optical gain curve

	Main characteristics	Lineshape	Examples of lasers
Homogeneous broadening	All atoms emit the same spectrum with the same center frequency υ_o Single-mode lasing output	Lorentzian function	Nd^{3+}:YAG; phonon collision broadening Pressure broadening in CO_2 lasers
Inhomogeneous broadening	Different atoms emit at slightly differing central frequencies, due to random processes shifting the peak emission frequency Multimode or single-mode lasing output	Gaussian function	He-Ne lasers; Doppler broadening Nd^{3+}: glass lasers; amorphous structure broadening

[26] There are also many practical examples in which the lineshape is neither exactly Lorentzian nor Gaussian, but a combination of the two. When the two functions are suitably combined, the result is a Voigt lineshape.

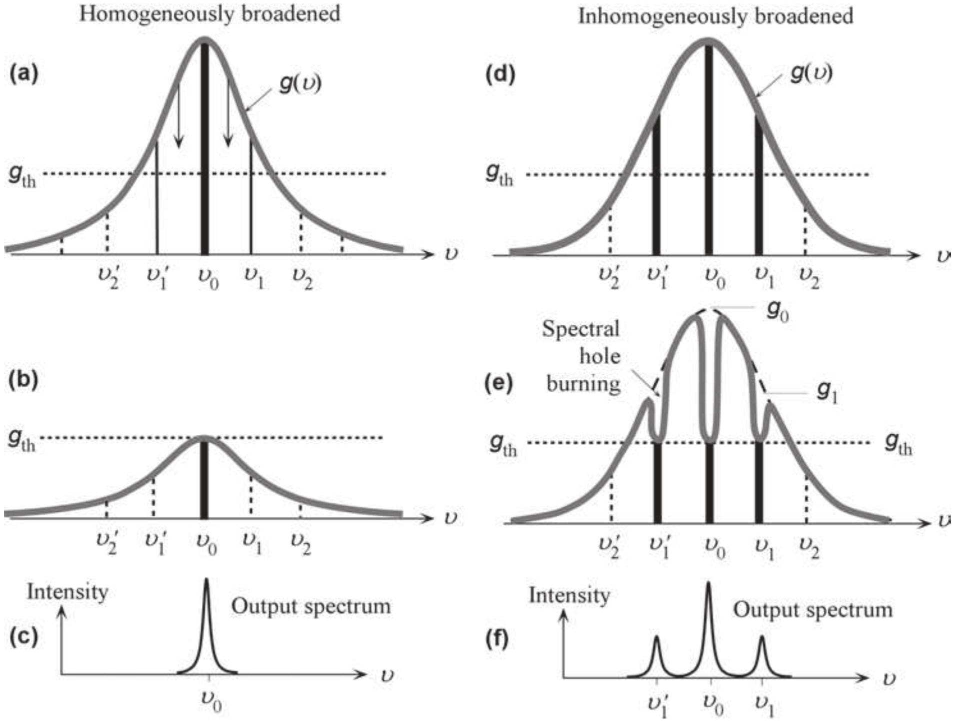


FIGURE 4.27 (a) An ideal homogeneously broadened gain medium that has been pumped above threshold for an instant. The mode at υ_0 is the most intense and de-excites atoms by stimulated emission and reduces the population inversion for all the modes, which reduces the gain. (b) The gain $g(\upsilon)$ is reduced until the threshold is reached, and only the υ_0 mode can oscillate. (c) The output spectrum is a single mode. (d) An ideal inhomogeneously broadened gain medium that has been pumped above threshold. The atoms behave independently and each has its own gain curve. The de-excitation of atoms by stimulated radiation at υ_0 does not affect the atoms at υ_1 and υ_1'. (e) The gain $g(\upsilon)$ must be reduced at υ_0, υ_1, and υ_1' until threshold is reached to sustain steady state oscillations. The gain spectrum thus has dips at υ_0, υ_1, and υ_1'. (f) The output spectrum has three modes at υ_0, υ_1, and υ_1'.

The υ_0-mode has the highest intensity, and the υ_0-radiation de-excites (by stimulated emission) a much greater number of atoms than υ_1- and υ_1'-radiation. Remember that all the atoms have the same gain curve, which means that the υ_0-radiation removes a large number of excited atoms from interacting with υ_1- or υ_1'-radiation. The emissions stimulated by the υ_0-radiation cause even more atoms to become de-excited, leaving very little number of atoms for the υ_1- and υ_1'-radiation. Put differently, the stimulated emissions induced by the highest intensity υ_0-radiation reduce the population inversion $N_2 - N_1$ and hence the gain for all, as illustrated by the vertical downward arrows in Figure 4.27 (a). The modes at υ_1 and υ_1', however, are affected very badly since their gains were already less than that of υ_0. As the gain curve is reduced, in the end, as shown in Figure 4.27 (b), only the υ_0-radiation can sustain itself. The output spectrum is a single mode as shown in Figure 4.27 (c).

Consider now a laser gain medium that has inhomogeneous broadening and suppose that the gain curve has three modes with gains above g_{th} as shown in Figure 4.27 (d). In this case, each atom behaves independently with its own gain curve that is centered away from the others as in Figure 4.26 (b). Consequently, the υ_0-mode radiation only interacts with those atoms that have their individual gain curves overlapping the υ_0-radiation. Similarly, υ_1-radiation interacts only with those atoms that have their individual gain curves overlapping the υ_1-radiation. When the υ_0-radiation de-excites atoms, it does not de-excite those that are at frequencies υ_1 and υ_1'. We know that stimulated emissions de-excite atoms and reduce the population inversion and hence the gain. Consequently, under steady state operation, the gain for the υ_0-mode is reduced from its peak g_0 down to g_{th}, which is the required gain for maintaining steady state oscillations. Similarly, the gain for υ_1-mode is reduced from g_1 down to g_{th}, as shown in Figure 4.27 (e). To maintain steady state laser oscillations, $g(\upsilon)$ must have dipped regions at υ_0, υ_1, and υ_1' where $g(\upsilon)$ becomes g_{th}. The dip in $g(\upsilon)$ at a mode is called **spectral hole burning**. It might seem as though all three modes should have the same intensity as they all have $g = g_{th}$, but this is not the case. Consider the υ_0- and υ_1-modes. The reduction of the gain from its peak at g_0 to g_{th}, and its maintenance at g_{th} involves stimulated emissions for de-excitation. More stimulated emissions are needed to reduce g_0 to g_{th} than for reducing g_1 to g_{th}. Thus υ_0-mode has greater stimulated emissions and more optical power than the υ_1-mode. The output spectrum is illustrated in Figure 4.27 (f). (As one might have surmised, for simplicity only, the mode υ_0 was made coincident with the peak of the gain curve.)

4.8 PULSED LASERS: *Q*-SWITCHING AND MODE LOCKING

A. *Q*-Switching

The *Q*-factor of an optical resonant cavity, like its finesse, is a measure of the cavity's frequency selectiveness. The higher the *Q*-factor, the more selective the resonator is, or the narrower the spectral width. It is also a measure of the energy stored in the resonator per unit energy dissipated, *e.g.,* losses at the reflecting surfaces or scattering in the cavity, per cycle of oscillation.[27] ***Q*-switching** refers to a laser whose optical resonant cavity is switched from a low *Q* to a high *Q* to generate an intense laser pulse, as described below.

A laser operation needs a good optical resonant cavity with low losses, that is, a high *Q*-factor, to allow the gain to reach the threshold gain. The large radiation intensity within the cavity stimulates further emissions. However, by temporarily removing one of the reflectors as in Figure 4.28 (a), we can switch the *Q*-factor of the cavity to some low value. While the optical

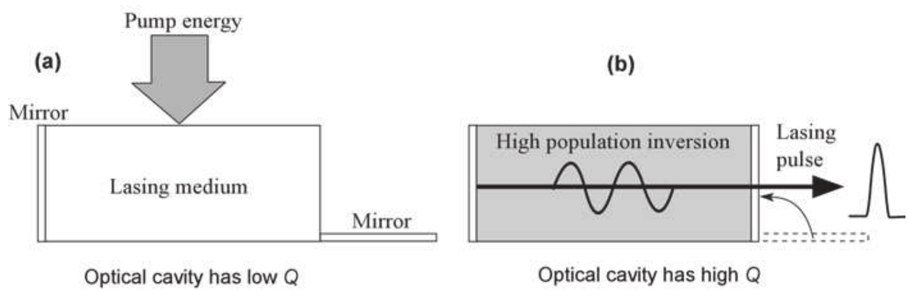


FIGURE 4.28 (a) The optical cavity has a low *Q* so that pumping takes the atoms to a very high degree of population inversion; lasing is prevented by not having a right-hand mirror. (b) The right mirror is "flung" to make an optical resonator, *Q* is switched to a high value which immediately encourages lasing emissions. There is an intense pulse of lasing emission which brings down the excess population inversion.

[27] See Section 1.11 in Chapter 1. All optical resonators have a *Q*-factor, which is another way of expressing their finesse.

cavity has low Q, the lasing is suppressed. Without the reflections from the mirrors, the photon energy density in the cavity is small, and the active medium can be pumped to achieve a large population inversion, greater than what would have normally been achieved had the mirrors been in place. When the Q of the optical cavity is switched to a high value, as in Figure 4.28 (b), an intense lasing emission is generated. While the Q is low, the pumped lasing medium is effectively a very high gain photon amplifier. There is too much loss in the optical resonator (*i.e.*, no optical feedback) to achieve a lasing oscillation. As soon as the Q is switched to a high value, the low loss in the optical resonator allows lasing oscillations to occur, which deplete the population inversion and decrease the gain until the population inversion falls below the threshold value and lasing oscillations cease.

The sequence of events involved in Q-switching a laser is illustrated in Figure 4.29. The pump, a flash tube in many cases, that provides optical excitation, is triggered at time t_1 as shown in Figure 4.29 (a). It must have sufficient intensity to quickly build up (N_2-N_1) to well above the normal threshold value $(N_2 - N_1)_{th}$ as indicated in Figure 4.29 (b). The pumping power needed must be provided before excited ions (or atoms) in the medium can spontaneously decay to the lower laser energy level. The optical cavity's Q is switched at time t_2, before the excited ions decay spontaneously to the lower laser level, as shown in Figure 4.29 (c). The result is a nearly sudden formation of an optical cavity, which enables lasing radiation to be quickly built up inside the cavity from t_2 to t_3 in Figures 4.29 (c)–(d). The radiation buildup from t_2 to t_3 is of the order of nanoseconds and starts with stray photons as in normal CW laser operation. Eventually, the rapid build-up of radiation in the cavity stimulates lasing transitions, and acts as positive feedback, yielding a pulse of coherent radiation as output as in Figure 4.29 (d). The duration of the lasing emission (the width of the output pulse) depends on both the Q-switching time, and the photon cavity lifetime, *i.e.*, how long photons are in the cavity before they leave. The Q-switching speed must be faster than the rate of oscillation buildup; otherwise the lasing oscillation builds up slowly. Table 4.3 summarizes the properties of two commercial Q-switched Nd^{3+}:YAG lasers. Notice that the Q-switched pulses are a few

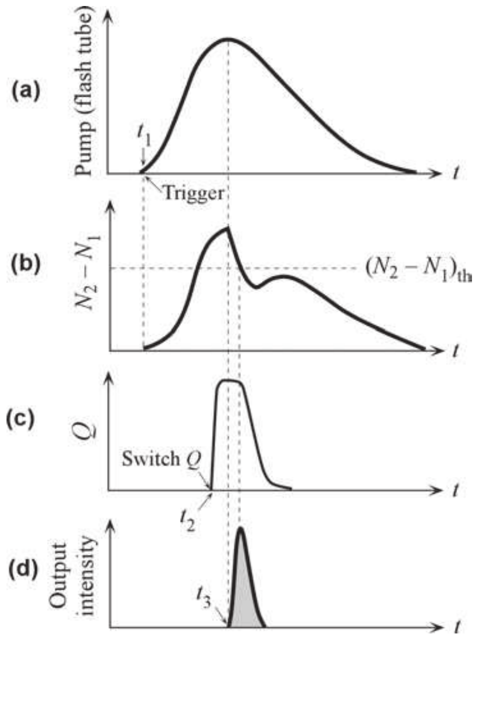


FIGURE 4.29 A simplified schematic of Q-switching in the generation of short laser pulses. (a) The pump profile, *i.e.,* the flash tube output. (b) The evolution of the population difference $N_2 - N_1$ with time. (c) The switching of the optical cavity Q during the pumping procedure while the population inversion is very large and greater than the normal threshold population. (d) The output light pulse.

TABLE 4.3 Selected typical characteristics of two commercial Q-switched Nd^{3+}:YAG lasers

Property	Nd^{3+}:YAG NL220 by EKSPLA	Nd^{3+}:YAG P-1064-150-HE by Alphalas
Emission wavelength (nm)	1064	1064
Output pulse energy (mJ)	10	1.5
Pulse width (ns)	7	1.1
Repetition rate (Hz)	1000	100
Peak power (kW)	~1400	~1400
Q-switch	EO	Passive
Output beam	TEM_{00}	TEM_{00}
Output beam M^2	<1.5	–
Beam divergence (mrad)	<1.5	6
Beam diameter (mm)	2.5	0.3

Note: Usually higher-energy output pulses require operation at a lower repetition frequency.

nanoseconds long, and the actual power in the laser pulse is 1.4–1.8 MW; extremely high compared with CW power levels.

There are a number of ways in which Q-switching can be conveniently implemented as illustrated in three examples in Figures 4.30 (a)–(c). The first example uses a rotating reflector, which in this case is a prism. The cavity Q is switched during the rotation of the prism every time a prism face aligns with the cavity axis, which is shown in Figure 4.30 (a). Another technique is to use a *saturable absorber*, also called a *bleachable dye*, in the optical cavity as in Figure 4.30 (b). A saturable absorber's absorption *decreases* with *increasing* light intensity so that it becomes transparent only at high intensities. When the absorber becomes transparent under a large radiation intensity, the dye is said to be bleached. Initially the absorber is opaque and keeps the cavity Q low by absorbing the radiation, but as the intensity builds up the absorber becomes transparent, which corresponds to switching the cavity Q to a high value. Another technique that is widely employed is the use of an *electro-optic* (EO) *switch* in the optical cavity as illustrated in Figure 4.30 (c). As will be explained in Chapter 6, such an EO switch is essentially an electro-optic (or a piezoelectric) crystal with electrodes, and works in combination with a polarizer. EO switches can be very fast (*e.g.*, less than nanoseconds), which is one of their distinct advantages. For the present purposes it is clear that when the EO switch is turned transparent by the application of a voltage, the cavity Q is switched high, which results in the generation of a lasing pulse.

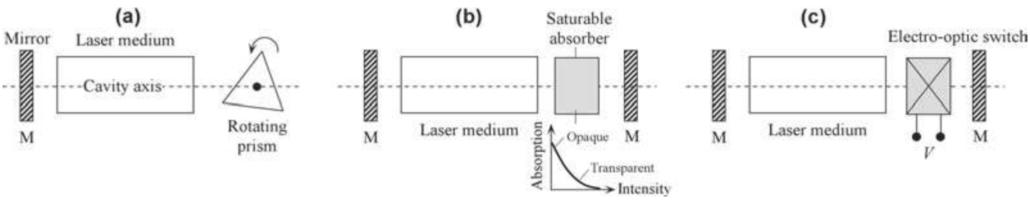


FIGURE 4.30 (a) Q-switching by using a rotating prism. (b) Q-switching by using a saturable absorber. (c) Q-switching by using an electro-optic (EO) switch. Normally a polarizer is also needed before or after the switch but this is part of the EO switch in this diagram.

B. Mode Locking

Mode locking is used to generate short and intense laser light pulses at a certain repetition rate that depends on the laser construction. A **mode-locked laser** is a laser that has been constructed to have one transverse mode and many (N) longitudinal modes, as in Figure 4.31 (a), that have the *same phase*. The N longitudinal modes then *reinforce* each other to generate an intense pulse of lasing emission at fixed time intervals as illustrated in Figure 4.31 (b). The repetition rate depends on the laser cavity length L and the speed of light in the lasing medium, c, assuming that the refractive index n is nearly 1 (*e.g.*, we have a gas laser). Normally, the longitudinal modes of a laser cavity would be "independent" with random relative phases. In such a case, the output intensity from the laser would simply be the sum of the individual mode intensities, *i.e.*, $I_1 + I_2 + \cdots + I_N$. If, on the other hand, the modes have been forced to have the same phase, then these mode oscillations would *reinforce* each other, *the optical fields would add*, in such a way that they generate an intense optical pulse at a certain repetition rate. In this mode-locked case, the intensity in a pulse is proportional to $(E_1 + E_2 + \cdots + E_N)^2$, and can be enormously larger than the case when the modes are not locked, as indicated in Figure 4.31 (b) where the horizontal dashed line, CW, represents the intensity in the unlocked laser. The repetition rate in a mode-locked laser is the reciprocal of the round-trip time of a light pulse in the optical cavity.

Suppose that $T = 2L/c$ is the round-trip time of a light pulse in the optical resonator shown in Figure 4.31 (c), where L is the cavity length and c is the velocity of light. The pulse repeats itself every $T = 2L/c$ seconds. The situation is analogous to taking a fixed number of sine waves with multiple frequencies, adjusting their phases so that one can obtain the maximum amplitude from their summation. The achievement of such an intense light pulse at every T seconds requires that the "modes are locked," that is, the relative phases of the modes have been correctly adjusted and fixed to yield the required maximum output intensity. When modes have been locked, there must be an optical pulse in the resonator that is traveling between the mirrors with exactly the required round-trip time of $T = 2L/c$. (If the refractive index n is not unity, we need to use c/n instead of c.)

The width Δt of an individual light pulse depends on the frequency width $\Delta\upsilon$ of the laser optical gain curve; normally $\Delta\upsilon$ is taken as the full width half maximum width. We know that $\Delta\upsilon$ also determines how many modes N are there in the output. The pulse width $\Delta t \approx 1/\Delta\upsilon \approx T/N$. In practice, it is not very difficult to obtain mode locking since the required output must correspond

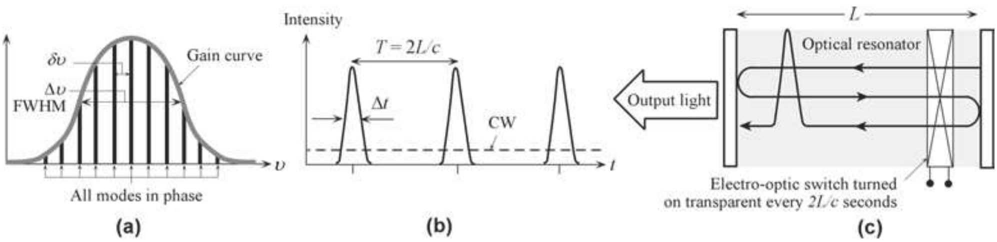


FIGURE 4.31 (a) A mode-locked laser has its N modes all in phase so that the modes add correctly to generate a short laser pulse every T seconds. $\Delta\upsilon$ is the full width at half maximum. (b) The output light intensity from a mode-locked laser is a periodic series of short intense optical pulses that are separated in time by $T = 2L/c$, the round-trip time for the pulse in the resonator. (c) A laser can be mode-locked by using an EO switch in the optical cavity that becomes transparent exactly at the right time, every T seconds. Each time the pulse in the resonator impinges on the left mirror, every $T = 2L/c$ seconds, a portion of it is transmitted, which constitutes the output from a mode-locked laser.

to an intense pulse train of repetition rate $1/T$. This output itself corresponds to a single optical pulse in the resonator bouncing back and forward between the mirrors with a round-trip time T. Each time this pulse impinges on a partially reflecting mirror, a portion of it is transmitted as an output pulse as in Figure 4.31 (c). Suppose that we insert an *electro-optic switch* that is switched to be made transparent at every T seconds (as we did for Q-switching). The switch is on to be transparent only when the pulse is there and only for the duration of the pulse. Then, the only possible situation is where a single optical pulse can bounce back and forward in the resonator and this corresponds to locking the modes. This is an example of **active mode locking**. The difference from the electro-optically Q-switched laser is that in mode locking, the EO switch must be turned on every T seconds for a very short time, so the timing in sequence is critical.

Another possibility is **passive mode locking** where a saturable absorber is used in the optical cavity (again, similar to Q-switching). The optical cavity is lossy for low light intensities. Such an absorber would only allow a high intensity light pulse to exist in the cavity since it is only transparent at high intensities. The latter would correspond to various modes having the right phases to yield an intense pulse, that is, mode locking. The difference from the Q-switching case is that the absorber in a mode-locked laser must be able to respond faster than the time T.

The output pulses from a mode-locked laser are spatially separated by $2L$ because the pulses come out at every T seconds from the output mirror and Tc is $2L$. A mode-locked He-Ne laser typically has a pulse width of roughly 600 ps whereas it is 150 fs for a mode-locked Nd^{3+}: glass laser. It might be thought that we have generated a more powerful output laser beam, a stream of high intensity pulses, than the CW (unlocked) case, but this is not true. The peak power in an individual mode-locked pulse, roughly the energy in the pulse divided by the pulse width Δt, can be extremely high, often in megawatts, but the time-averaged power (over many pulses) is the same as that in the CW unlocked laser.

4.9 PRINCIPLE OF THE LASER DIODE[28]

Consider a degenerately doped direct bandgap semiconductor *pn* junction whose band diagram is shown in Figure 4.32 (a). By degenerate doping we mean that the Fermi level E_{Fp} in the *p*-side is in the valence band (VB) and E_{Fn} in the *n*-side is in the conduction band (CB). All energy levels up to the Fermi level can be taken to be occupied by electrons as in Figure 4.32 (a). In the absence of an applied voltage, the Fermi level is continuous across the diode, $E_{Fp} = E_{Fn}$. The depletion region or the space charge layer (SCL) in such a *pn* junction is very narrow. There is a built-in voltage V_o that gives rise to a potential energy barrier eV_o that prevents electrons in the CB of n^+-side diffusing into the CB of the p^+-side.[29] There is a similar barrier stopping the hole diffusion from p^+-side to n^+-side.

Recall that when a voltage is applied to a *pn* junction device, the change in the Fermi level from end-to-end is the electrical work done by the applied voltage, that is, $\Delta E_F = eV$. Suppose

[28] The first semiconductor lasers used GaAs *pn* junctions, as reported by the American researchers Robert N. Hall *et al.* (General Electric Research, Schenectady), Marshall I. Nathan *et al.* (IBM, Thomas J. Watson Research Center), and Robert H. Rediker *et al.* (MIT), in 1962. Nick Holonyak (General Electric at Syracuse), also in 1962, reported a laser diode based on the compound GaAsP, which emitted in the visible (red).

[29] The potential energy barrier eV_o in a nondegerate *pn* junction is normally taken as the energy required for an electron to move from E_c on the *n*-side to E_c on the *p*-side since electrons are near the bottom of the CB at E_c. In the present case, electrons occupy states from E_c up to E_{Fn}. The exact analysis of the problem is therefore more complicated; however $E_{Fn} - E_c$ is small compared to eV_o.

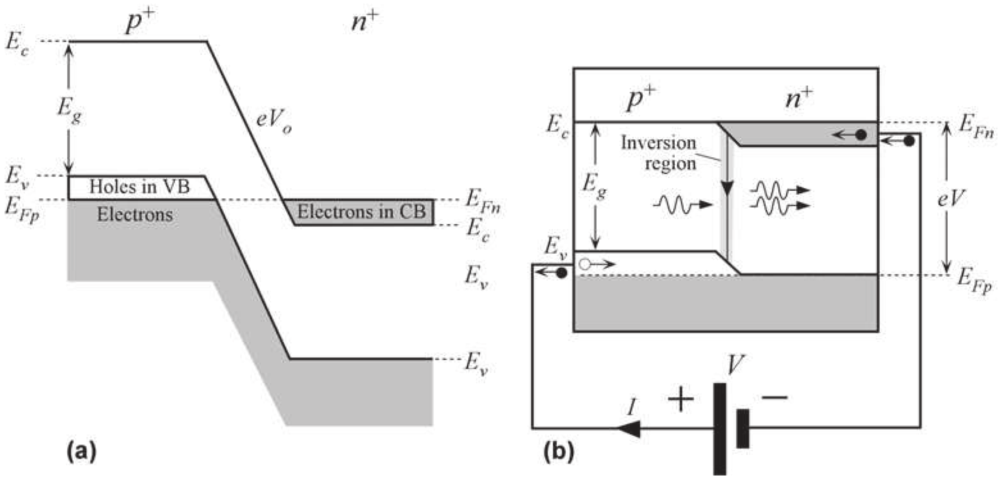


FIGURE 4.32 (a) The energy band diagram of a degenerately doped *pn* with no bias. eV_o is the potential energy barrier against electron diffusion from the n^+-side to the p^+-side. Note that $(E_{Fn} - E_c)$ and $(E_v - E_{Fp})$ are small compared with eV_o. (The diagram is exaggerated.) (b) Band diagram with a sufficiently large forward bias to cause population inversion and hence stimulated emission. (As before, filled black circles are electrons and empty circles are holes. Only electrons flow in the external circuit.)

that this degenerately doped *pn* junction is forward biased by a voltage *V* greater than the bandgap voltage; $eV > E_g$ as shown in Figure 4.32 (b). The separation between E_{Fn} and E_{Fp} is now the applied potential energy or eV. The applied voltage diminishes the built-in potential barrier to almost zero, which means that electrons flow into the SCL and flow over to the p^+-side to constitute the diode current. There is a similar reduction in the potential barrier for holes from p^+- to n^+-side. The final result is that electrons from n^+-side and holes from p^+-side flow into the SCL, and this SCL region is no longer depleted, as apparent in Figure 4.32 (b). If we draw the energy band diagram with $E_{Fn} - E_{Fp} = eV > E_g$ this conclusion is apparent. In this region, there are *more* electrons in the conduction band at energies near E_c than electrons in the valence band near E_v as illustrated by density of states diagram for the junction region in Figure 4.33 (a). In other words, there is a **population inversion** between energies near E_c and those near E_v around the junction.

This population inversion region is a layer along the junction and is called the **inversion layer** or the **active region**. An incoming photon with an energy of $(E_c - E_v)$ cannot excite an

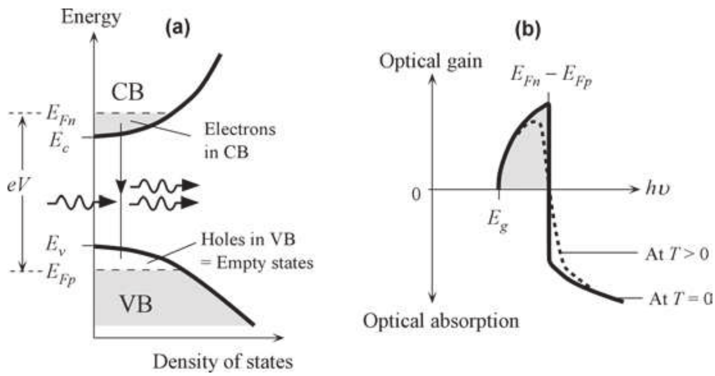


FIGURE 4.33 (a) The density of states and energy distribution of electrons and holes in the conduction and valence bands, respectively, in the SCL under forward bias such that $E_{Fn} - E_{Fp} > E_g$. Holes in the VB are empty states. (b) Gain vs. photon energy ($h\upsilon$). The effect of temperature is also shown.

electron from E_v to E_c as there are almost none near E_v. It can, however, *stimulate* an electron to fall down from E_c to E_v as shown in Figure 4.32 (b). Put differently, the incoming photon stimulates direct recombination. The region where there is population inversion and hence more stimulated emission than absorption, or the active region, has an **optical gain** because an incoming photon is more likely to cause stimulated emission than being absorbed. The optical gain depends on the photon energy (and hence on the wavelength) as apparent by the energy distributions of electrons and holes in the conduction and valence bands in the active layer in Figure 4.33 (a). At low temperatures ($T \approx 0$ K), the states between E_c and E_{Fn} are filled with electrons and those between E_{Fp} and E_v are empty. Photons with energy greater than E_g but less than $E_{Fn} - E_{Fp}$ cause stimulated emissions, whereas those photons with energies greater than $E_{Fn} - E_{Fp}$ become absorbed. Figure 4.33 (b) shows the expected dependence of optical gain and absorption on the photon energy at low temperatures ($T \approx 0$ K). As the temperature increases, the Fermi–Dirac function spreads the energy distribution of electrons in the CB to above E_{Fn} and holes below E_{Fp} in the VB. Put differently, electrons from below E_{Fn} are spread to energies above E_{Fn}. The result is a reduction and also some smearing of the optical gain curve as indicated in Figure 4.33 (b). The optical gain depends on $E_{Fn} - E_{Fp}$, which depends on the applied voltage, and hence on the diode current.

It is apparent that population inversion between energies near E_c and those near E_v is achieved by the injection of carriers across the junction under a sufficiently large forward bias. The pumping mechanism is therefore the forward diode current and the pumping energy is supplied by the external battery. This type of pumping is called **injection pumping**.

In addition to population inversion we also need to have an *optical cavity* to implement a laser oscillator, that is, to build up the intensity of stimulated emissions by means of an optical resonator. This would provide a continuous coherent radiation as output from the device. Figure 4.34 shows schematically the structure of a **homojunction laser diode**. The *pn* junction uses the same direct bandgap semiconductor material throughout, for example GaAs, and hence has the name homojunction. The ends of the crystal are cleaved to be flat and optically polished to provide reflection and hence form an optical cavity. Photons that are reflected from the cleaved surfaces stimulate more photons of the same frequency and so on. This process builds up the intensity of the radiation in the cavity. The wavelength of the radiation that can build up in the cavity is determined by the length L of the cavity because

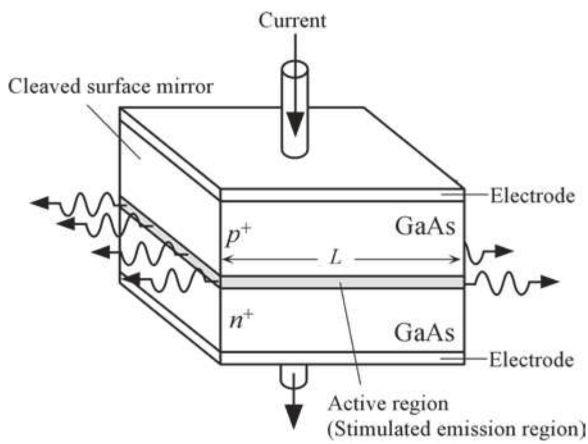
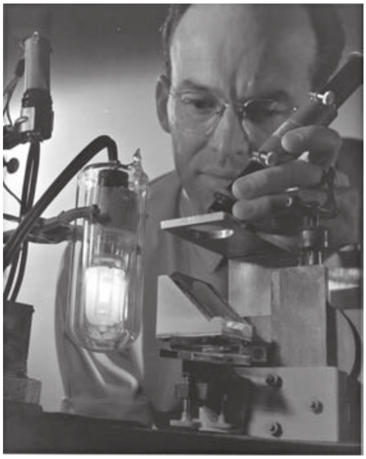


FIGURE 4.34 A schematic illustration of a GaAs homojunction laser diode. The cleaved surfaces act as reflecting mirrors.



Robert Hall and his colleagues, while working at General Electric's Research and Development Center in Schenectady, New York, were among the first groups of researchers to report a working semiconductor laser diode in 1962. He obtained a U.S. patent in 1967, entitled "Semiconductor junction laser diode" for his invention. When Robert Hall retired from GE in 1987, he had been awarded more than 40 patents. *(*R. N. Hall et al., *Phys. Rev. Letts., 9,* 366, 1962; *Courtesy of GE.)*

only multiples of the half-wavelengths can exist in such an optical cavity as explained in Section 4.6D, *i.e.,*

Modes in an optical cavity

$$m\frac{\lambda}{2n} = L \tag{4.9.1}$$

where, as before, m is an integer, called a **mode number**, n is the refractive index of the semiconductor, and λ is the free-space wavelength. Each radiation pattern satisfying the above relationship is essentially a **resonant frequency** of the cavity, that is, a **mode** of the cavity. The separation between possible modes of the cavity (or separation between allowed wavelengths) $\Delta\lambda_m$ can be readily found from Eq. (4.9.1) as in the case of the He-Ne gas laser previously.

The dependence of the optical gain of the medium on the wavelength of radiation can be deduced from the energy distribution of the electrons in the CB and holes in the VB around the junction as in Figures 4.33 (a) and (b). The exact output spectrum from the laser diode depends both on the nature of the optical cavity and the optical gain vs. wavelength characteristics. Lasing radiation is only obtained when the optical gain in the medium can overcome the photon losses from the cavity, which requires the diode current I to exceed a certain threshold value I_{th}. (This is similar to the condition we had for gas lasers where pumping must exceed some threshold pumping required for the threshold gain.) Below I_{th}, the light from the device is due to spontaneous emission and not stimulated emission. The light output is then composed of incoherent photons that are emitted randomly and the device behaves like an LED. For diode currents above I_{th}, the device emits coherent lasing emission.

We can identify two critical diode currents. First is the diode current that provides just sufficient injection to lead to stimulated emissions just balancing absorption. This is called the **transparency current** I_T inasmuch as there is then no net photon absorption; the medium is perfectly *transparent.* Above I_T there is optical gain in the medium though the optical output is not yet a continuous wave coherent radiation. Lasing oscillations occur only when the optical gain in the medium can overcome the photon losses from the cavity, that is, when the optical gain g reaches the threshold gain g_{th}. This occurs at the **threshold current** I_{th}. Those cavity resonant frequencies that experience the threshold optical gain can resonate within the cavity. Some of this cavity radiation is transmitted out from the cleaved ends as these are not perfectly reflecting (typically about

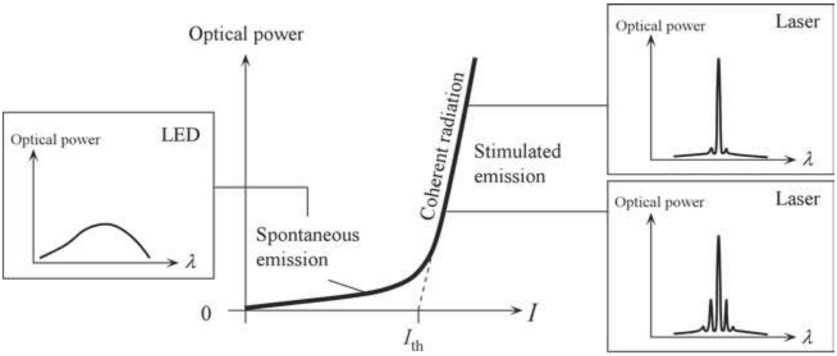


FIGURE 4.35 Typical output optical power vs. diode current (I) characteristics and the corresponding output spectrum of a laser diode. I_{th} is the threshold current and corresponds to the extension of the coherent radiation output characteristic onto the I-axis.

32% reflecting without any antireflection coating). Figure 4.35 shows the output light intensity as a function of diode current. Below the threshold current I_{th}, the light output is incoherent radiation due to the spontaneous recombination of injected electrons and holes. Above I_{th}, the light output becomes coherent radiation consisting of cavity wavelengths (or modes); and the output intensity increases steeply with the current. The number of modes in the output spectrum and their relative strengths depend on the diode current as illustrated in Figure 4.35. Notice that as the current increases, one of the intense modes becomes prominent at the expense of less intense modes. The prominent mode has so much intensity that it ends up using most of the injected electrons; there are insufficient electrons left for other modes to satisfy the threshold gain condition at those frequencies.

The main problem with the homojunction laser diode is that the threshold current density J_{th} is too high for practical uses. For example, the threshold current density is of the order of $\sim 500\,\text{A mm}^{-2}$ for GaAs *pn* junctions at room temperature, which means that the GaAs homojunction laser can be operated continuously only at very low temperatures. However, J_{th} can be reduced by orders of magnitude by using **heterostructure** semiconductor laser diodes.

4.10 HETEROSTRUCTURE LASER DIODES

The reduction of the threshold current I_{th} to a practical value requires improving the rate of stimulated emission and also improving the efficiency of the optical cavity. First, we can confine the injected electrons and holes to a narrow region around the junction. This narrowing of the active region means that less current is needed to establish the necessary concentration of carriers for population inversion. Second, we can build a dielectric waveguide around the optical gain region to increase the photon concentration and hence the probability of stimulated emission. This way we can reduce the loss of photons traveling off the cavity axis. We therefore need both **carrier confinement** and **photon** or **optical confinement.** Both of these requirements are readily achieved in modern laser diodes by the use of heterostructured devices as in the case of high-intensity double heterostructure LEDs. However, in the case of laser diodes, there is an additional requirement for maintaining a good optical cavity that will increase stimulated emission over spontaneous emission.

Figure 4.36 (a) shows a **double heterostructure** (DH) device based on two junctions between different semiconductor materials with different bandgaps. In this case, the semiconductors are $Al_{1-x}Ga_xAs$ (or simply AlGaAs) with $E_g \approx 2$ eV and GaAs with $E_g \approx 1.4$ eV. The *p*-GaAs region is a thin layer, typically about 0.1 μm, and constitutes the **active layer** in which

Izuo Hayashi (left) and Morton Panish (1971) at Bell Labs were able to design the first semiconductor laser that operated continuously at room temperature. The need for semiconductor heterostructures for efficient laser diode operation was put forward by Herbert Kroemer in the United States and Zhores Alferov in Russia in 1963. (Notice the similarity of the energy band diagram on the chalkboard with that in Figure 4.36.) *(Reprinted with permission of Alcatel–Lucent USA Inc.)*

lasing recombination takes place. Both p-GaAs and p-AlGaAs are heavily p-type doped and are degenerate with E_F in the valence band. When a sufficiently large forward bias is applied, E_c of n-AlGaAs moves above E_c of p-GaAs, which leads to a large injection of electrons from the CB of n-AlGaAs into the CB of p-GaAs as shown in Figure 4.36 (b). These electrons, however, are *confined* to the CB of p-GaAs since there is a potential barrier ΔE_c between p-GaAs and p-AlGaAs due to the change in the bandgap (there is also a small change in E_v but we ignore this). Inasmuch as p-GaAs is a thin layer, the concentration of injected electrons in the p-GaAs layer can be increased quickly even with moderate increases in forward current. This effectively reduces the threshold current for population inversion or optical gain. Thus, even moderate forward currents can inject sufficient number of electrons into the CB of p-GaAs to establish the necessary electron concentration for population inversion in this layer.

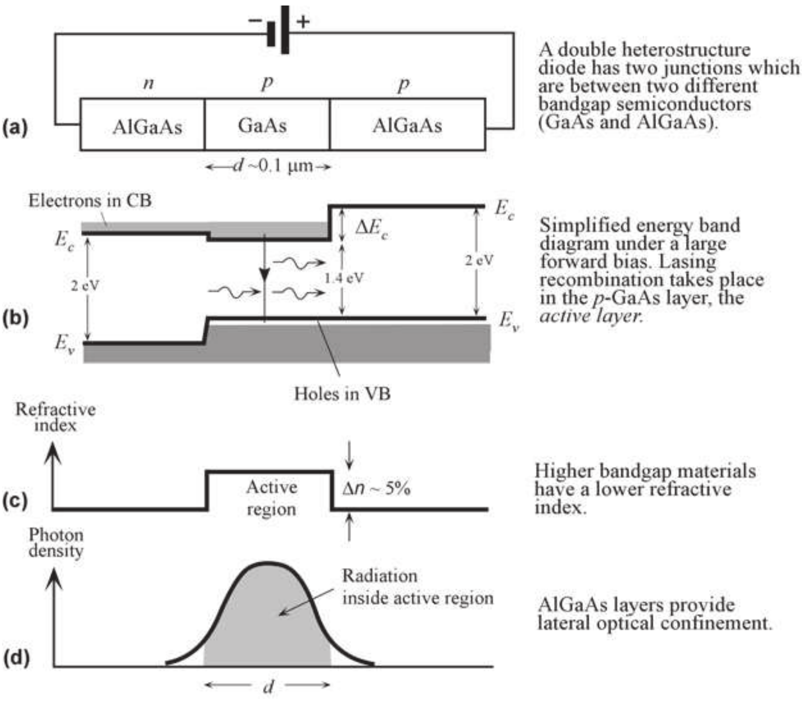


FIGURE 4.36 The basic principle of operation of a double heterostructure laser.

A wider bandgap semiconductor generally has a lower refractive index. AlGaAs has a lower refractive index than that of GaAs. The change in the refractive index defines an optical dielectric waveguide, as illustrated in Figure 4.36 (c), that confines the photons to the active region of the optical cavity and thereby increases the photon concentration. The photon concentration across the device is shown in Figure 4.36 (d). This increase in the photon concentration increases the rate of stimulated emissions. Thus, both carrier and optical confinement lead to a reduction in the threshold current density. Without double heterostructure devices we would not have practical solid state lasers that can be operated continuously at room temperature.

A typical structure of a **double heterostructure laser diode** is similar to a double heterostructure LED and is shown schematically in Figure 4.37. The doped layers are grown epitaxially on a crystalline substrate which in this case is *n*-GaAs. The double heterostructure described above consists of the first layer on the substrate, *n*-AlGaAs, the active *p*-GaAs layer, and the *p*-AlGaAs layer. There is an additional *p*-GaAs layer, called **contacting layer**, next to *p*-AlGaAs. It can be seen that the electrodes are attached to the GaAs semiconductor rather than AlGaAs. This choice allows for better contacting, that is, smaller contact resistance. The *p*- and *n*-AlGaAs layers provide carrier and optical confinement in the vertical direction by forming heterojunctions with *p*-GaAs. The active layer is *p*-GaAs, which means that the lasing emission will be in the range 870–900 nm depending on the doping level. This layer can also be made to be $Al_yGa_{1-y}As$ but of different composition than the confining $Al_xGa_{1-x}As$ layers, and still preserve heterojunction properties. By modifying the composition of the active layer, we can control the wavelength of the lasing emission over 650–900 nm. The advantage of AlGaAs/GaAs heterojunction is that there is only a small lattice mismatch between the two crystal structures and hence negligible strain induced interfacial defects (*e.g.*, dislocations) in the device. Such defects invariably act as nonradiative recombination centers and hence reduce the rate of radiative transitions.

An important feature of this laser diode is the **stripe geometry**, or stripe contact on *p*-GaAs. The current density J from the stripe contact is not uniform laterally. J is greatest

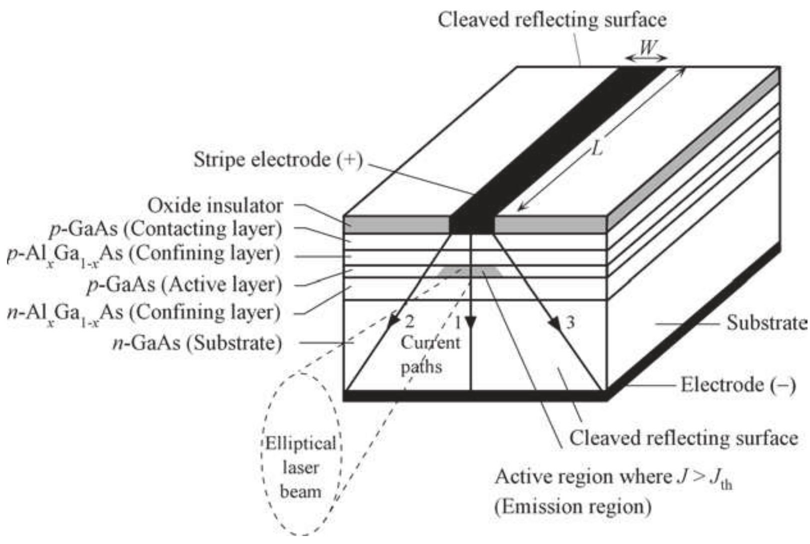


FIGURE 4.37 Schematic illustration of the structure of a double heterojunction stripe contact laser diode.

along the central path, 1, and decreases away from path 1, toward 2 or 3. The current is confined to flow within paths 2 and 3. The current density paths through the active layer where J is greater than the threshold value J_{th}, as shown in Figure 4.37, define the **active region** where population inversion and hence optical gain takes place. The lasing emission emerges from this active region. The width of the active region, or the optical gain region, is therefore defined by the current density from the stripe contact. Optical gain is highest where the current density is greatest. Such lasers are called **gain guided**. There are two advantages to using a stripe geometry. First, the reduced contact area also reduces the threshold current I_{th}. Second, the reduced emission area makes light coupling to optical fibers easier. Typical stripe widths (W) may be as small as a few microns leading to typical threshold currents that may be tens of milliamperes.

The simple double heterostructure laser diode structure in Figure 4.37 has an optical cavity defined by the end surfaces, *i.e.,* the facets, of the crystal. The lasing emission emerges from one of the crystal facets, that is, from an area on a crystal face perpendicular to the active layer. The radiation is emitted from the *edge* of the crystal. These DH devices are usually marketed as **Fabry–Perot cavity, edge emitting** laser diodes, or simply FP laser diodes, and edge emission is implied. Since the refractive index of GaAs is about 3.6, the reflectance is 0.32 (or 32%). The laser efficiency can therefore be further improved and the finesse of the FP cavity enhanced by reducing the reflection losses from the crystal facets by, for example, suitably coating the end surfaces. Further, by fabricating a dielectric mirror (as explained in Chapter 1) at the rear facet, that is, a mirror consisting of a number of quarter-wavelength semiconductor layers of different refractive index, it is possible to bring the reflectance close to unity and thereby improve the optical gain of the cavity. This corresponds to a reduction in the threshold current.

The width, or the lateral extent, of the optical gain region in the stripe geometry DH laser in Figure 4.37 is defined by the current density, and changes with the current. More importantly, the lateral optical confinement of photons to the active region is poor because there is no marked change in the refractive index laterally. It would be advantageous to laterally confine the photons to the active region to increase the rate of stimulated emissions. This can be achieved by shaping the refractive index profile in the same way the vertical confinement was defined by the heterostructure. Figure 4.38 illustrates schematically the structure of such a DH laser diode where the active layer, p-GaAs, is bound both vertically and laterally by a wider bandgap semiconductor, AlGaAs, which has lower refractive index. The active layer (GaAs) is effectively buried within a wider bandgap material (AlGaAs) and the structure is hence called a **buried double heterostructure** laser diode. Since the active layer is surrounded by a lower index material (AlGaAs), it behaves as a dielectric waveguide and ensures

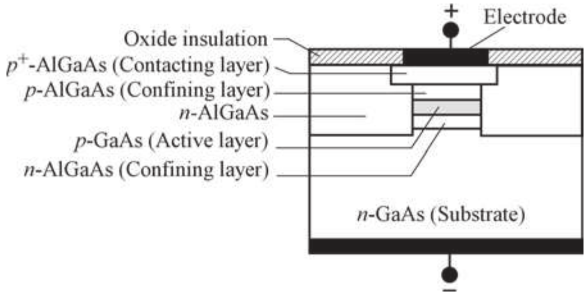


FIGURE 4.38 A simplified schematic diagram of a double heterostructure semiconductor laser device that has its active region (p-GaAs) *buried* within the device in such a way that it is surrounded by a low refractive index material (AlGaAs) rendering the active region as a waveguide.

that the photons are confined to the active or optical gain region. The photon confinement increases the rate of stimulated emission and hence the efficiency of the diode. Inasmuch as the optical power is confined to the waveguide defined by the refractive index variation, these diodes are called **index guided**. Further, if the buried heterostructure has the right lateral dimensions compared with the wavelength of the radiation, then only the **fundamental lateral mode** can exist in this waveguide structure as in the case of dielectric waveguides. There may, however, be several longitudinal modes.

A laser diode in which the lasing emission occurs only in one mode, both longitudinally and laterally, is called a **single-mode laser diode**. Lateral confinement, and hence a single lateral mode (*i.e.*, TE_{00}) can be easily obtained in buried DH laser diodes by ensuring that the buried active layer has a sufficiently small lateral size, *e.g.,* a few microns. In the stripe geometry LDs, a single lateral mode can be obtained by keeping the width of the stripe (W) narrow in Figure 4.37. On the other hand, operation at a single longitudinal mode, as shown in Example 4.10.1, typically needs the cavity length to be sufficiently short to allow only one mode to exit within the optical gain bandwidth. There are several techniques for generating single-mode lasers. One commonly used technique is to restrict the allowed wavelengths in the optical cavity by using wavelength-selective reflectors, as discussed later in Section 4.14.

The laser diode heterostructures based on $Al_{1-x}Ga_xAs$ alloys are suitable for emissions from about 680 nm (red) to about 900 nm (IR). For operation in the optical communication wavelengths of 1.3 μm and 1.55 μm, typical heterostructures are based on InGaAsP ($In_{1-x}Ga_xAs_{1-y}P_y$) alloys grown on InP substrates. InGaAsP quaternary alloys have a narrower bandgap than InP, and a greater refractive index. The composition of the InGaAsP alloy is adjusted to obtain the required bandgap for the active and confining layers. Figure 4.39 shows a highly simplified schematic structure of a buried (index guided) DH LD laser diode for use in optical communications. The active layer (InGaAsP) is surrounded by wider bandgap, lower refractive index InP. Notice that the active layer with the right InGaAsP composition is surrounded by lower refractive index, higher bandgap InP both horizontally and vertically. Layers are grown on an InP substrate. The contact covers the whole surface so that the current must be limited to the central active layer region. The latter is achieved by using InP *np* junctions around the active layer that are reverse biased and therefore prevent the current flow outside the central active region. There is a *p*-InGaAs layer between the electrode and the *p*-InP that serves as a contacting layer; that is, this layer facilitates current injection into the *p*-InP by reducing the contact resistance.

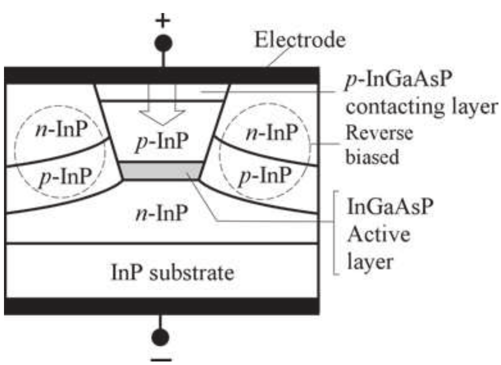
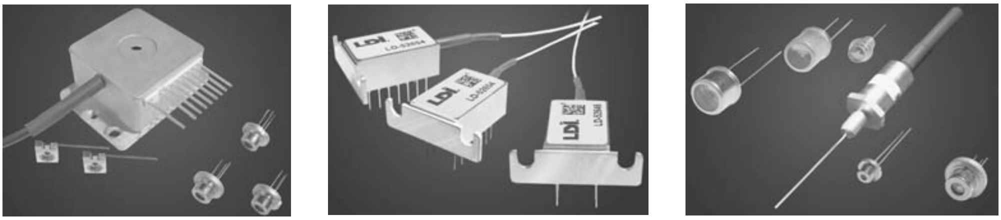


FIGURE 4.39 A highly simplified schematic sketch of a buried heterostructure laser diode for telecom applications. The active layer (InGaAsP) is surrounded by the wider bandgap, lower refractive index InP material. Layers are grown on an InP substrate. The InP *np* junction is reverse biased and prevents the current flow outside the central active region.



Left: High power (0.5–7 W) CW laser diodes with emission at 805 nm and a spectral width of 2.5 nm. Applications include medical systems, diode pumped lasers, analytical equipment, illuminators, reprographics, laser initiated ordnance, *etc.* Center: Typical pigtailed laser diodes for telecom. These are Fabry–Perot laser diodes operating at peak wavelengths of 1310 nm and 1550 nm with spectral widths of 2 nm and 1.3 nm, respectively. The threshold currents are 6 mA and 10 mA, and they can deliver 2 mW of optical power into a single-mode fiber. Right: High power 850 nm and 905 nm pulsed laser diodes for use in range finders, ceilometers, weapon simulation, optical fuses, surveying equipment, *etc.* *(Courtesy of OSI Laser Diode Inc.)*

EXAMPLE 4.10.1 Modes in a semiconductor laser and the optical cavity length

Consider an AlGaAs-based heterostructure laser diode that has an optical cavity of length 200 μm. The peak radiation is at 870 nm and the refractive index of GaAs is about 3.6. What is the mode integer m of the peak radiation and the separation between the modes of the cavity? If the optical gain vs. wavelength characteristics has a FWHM wavelength width of about 6 nm, how many modes are there within this bandwidth? How many modes are there if the cavity length is 20 μm?

Solution

Figure 4.19 schematically illustrates the cavity modes, the optical gain characteristics, and a typical output spectrum from a laser. The wavelength λ of a cavity mode and length L are related by Eq. (4.9.1),

$$m\frac{\lambda}{2n} = L$$

where n is the refractive index of the semiconductor medium, so that

$$m = \frac{2nL}{\lambda} = \frac{2(3.6)(200 \times 10^{-6})}{(870 \times 10^{-9})} = 1655.1 \quad \text{or} \quad 1655 \text{ (integer)}$$

The wavelength separation $\Delta\lambda_m$ between the adjacent cavity modes m and $(m + 1)$ in Figure 4.19 is

$$\Delta\lambda_m = \frac{2nL}{m} - \frac{2nL}{m+1} \approx \frac{2nL}{m^2} = \frac{\lambda^2}{2nL}$$

where we assumed that the refractive index n does not change significantly with wavelength from one mode to another. Thus, the separation between the modes for a given peak wavelength increases with decreasing L. When $L = 200$ μm

$$\Delta\lambda_m = \frac{(870 \times 10^{-9})^2}{2(3.6)(200 \times 10^{-6})} = 5.26 \times 10^{-10}\text{ m} \quad \text{or} \quad 0.526 \text{ nm}$$

If the optical gain has a bandwidth of $\Delta\lambda_{1/2}$, then there will be $\Delta\lambda_{1/2}/\Delta\lambda_m$ number of modes, or (6 nm)/(0.526 nm), that is, 11 modes.

When $L = 20$ μm, the separation between the modes becomes

$$\Delta\lambda_m = \frac{(870 \times 10^{-9})^2}{2(3.6)(20 \times 10^{-6})} = 5.26 \text{ nm}$$

Then $(\Delta\lambda_{1/2})/\Delta\lambda_m = 1.14$ and there will be one mode that corresponds to about 870 nm. In fact, m must be an integer so that choosing the nearest integer, $m = 166$, gives $\lambda = 867.5$ nm (choosing $m = 165$ gives 872.7 nm). It is apparent that reducing the cavity length suppresses higher modes. Note that the optical bandwidth depends on the diode current.

4.11 QUANTUM WELL DEVICES

As in the case of light-emitting diodes (LEDs), quantum wells (QWs) are also widely used in modern laser diodes. A typical **quantum well**-based laser is essentially a heterostructure device, as in Figure 4.36, in which the thin GaAs layer becomes ultrathin, typically less than 20 nm. We now have an ultrathin narrow bandgap GaAs sandwiched between two wider bandgap semiconductors, which results in a QW as illustrated in Figure 4.40 (a). The confinement length d, the size of the QW, is so small that we can treat the electron as in a one-dimensional potential energy (PE) well in the x-direction and as if it were free in the yz plane. The energy of the electron in the QW is quantized along the x-direction of confinement, which is shown in Figure 4.40 (b) as $E_1, E_2, \ldots$, corresponding to the quantum number n being $1, 2, \ldots$, respectively. Similarly, the hole energy is also quantized as E_1', E_2', *etc.,* with a corresponding quantum number n' being 1, 2, *etc.,* respectively. The electron is free in the yz plane (along y and z), which means that we must add this kinetic energy to the quantized energy levels. The electrons in the conduction band (CB) form a 2D (two-dimensional) free electron gas. As explained in Chapter 3,

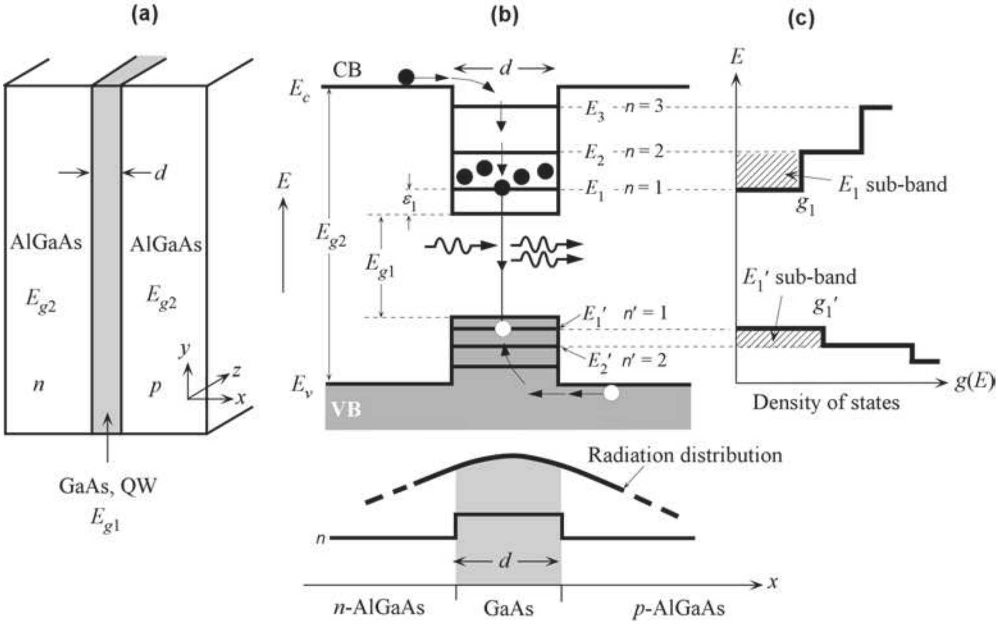


FIGURE 4.40 (a) A single quantum well of bandgap E_{g1} sandwiched between two semiconductors of wider bandgap E_{g2}. (b) The electron energy levels, and stimulated emission. The electrons and holes are injected from n-AlGaAs and p-AlGaAs, respectively. The refractive index variation tries to confine the radiation to GaAs but d is too thin, and most of the radiation is in the AlGaAs layers rather than within d. (c) The density of sates $g(E)$ is a step-like function, and is finite at E_1 and E_1'. The E_1 sub-band for electrons and E_1' sub-band for holes are also shown. The electrons in the E_1 sub-band have kinetic energies in the yz plane.

the density of states $g(E)$[30] for a 2D electron gas is a step-like function with energy as shown in Figure 4.40 (c). Notice that the region between E_1 and E_2 has been hatched in Figure 4.40 (c) to indicate that the energies between E_1 and E_2 are due to the kinetic energy of the electron in the yz plane (and almost continuous). We can therefore identify a sub-band of energies starting at E_1 and another sub-band starting at E_2, and so on. The density of states at E_1, and up to E_2, is g_1, which is large and finite. Similarly for holes, the density of states at E_1' is g_1', large and finite.

A **single quantum well** (SQW) device has only one quantum well surrounded by wider bandgap semiconductors as shown in Figures 4.40 (a) and (b) for a thin GaAs layer (thickness d) sandwiched between two thick AlGaAs layers. Usually the QW layer is undoped, whereas the AlGaAs confining layers are heavily doped. Under a sufficient forward bias, the electrons will be injected from n-AlGaAs, and holes from p-AlGaAs into the QW's CB and VB, respectively. These injected electrons very quickly thermalize and start filling states at and near E_1. Since at E_1 there is a finite and substantial density of states (g_1), the electrons in the conduction band do not have to spread far in energy to find states. In the bulk semiconductor on the other hand, the density of states at the bottom of the band (at E_c) is zero and increases slowly with energy (as $E^{1/2}$), which means that the electrons are spread more deeply into the conduction band in search for states. Thus, in a QW, a large concentration of electrons can easily occur at E_1, whereas this is not the case in the bulk semiconductor at E_c. Similarly, the majority of holes in the valence band will be around E_1' since there are sufficient states at this energy. Thus, under a forward bias, the injected electrons readily populate the ample number of states at E_1, which means that the electron concentration at E_1 increases rapidly with the current and hence, population inversion occurs quickly without the need for a large current to bring in a great number of electrons. Stimulated transitions of electrons from E_1 to E_1' leads to a lasing emission as illustrated in Figure 4.40 (b). In practice, we should also include the kinetic energy of the electron before and after the transition, which means that the radiative transitions occur from the E_1 sub-band to the E_1' sub-band. When the injected electron concentration is high (under a large current), E_2 and E_2' sub-bands will also be involved.

There are two distinct advantages to a QW laser diode. First is that the threshold current for population inversion and hence lasing emission should be markedly reduced with respect to that for bulk semiconductor devices. Second, since the majority of the electrons are at or near E_1 and holes are at or near E_1', the range of emitted photon energies should be very close to $E_1 - E_1'$. Consequently, the optical gain curve, and hence the bandwidth, in a SQW laser diode is narrower than its bulk counterpart. We note that the allowed radiative transitions in a QW follow a **selection rule** that requires $\Delta n = n - n'$ to be zero. Thus, the transition E_1 to E_1' is allowed, as is E_2 to E_2', but E_1 to E_2' has a very low probability.

It is clear that the SQW provides a higher probability of radiative transitions due to the large densities of states at E_1 and E_1', and the confinement of the electrons and holes: a distinct advantage. However, the SQW can still be further improved. The radiation is not well confined to the QW region because the QW is too thin as shown in Figure 4.40 (b); the radiation spreads out into the neighboring layers in a similar fashion to the spread of radiation into the cladding in an optical fiber that has a very thin core (a very small V-number). The optical confinement can be improved by having **optical confinement layers**, or **cladding layers**, surrounding the QW. The second problem is that at high currents, the QW can be flooded with electrons and lose its function as a QW.

The above two problems with the SQW are easily overcome by using **multiple quantum wells** (MQWs) as shown Figure 4.41. In MQW lasers, the structure has alternating

[30] Since g (san serif) is used for optical gain, g (times roman) is used for the density of states function in this chapter.

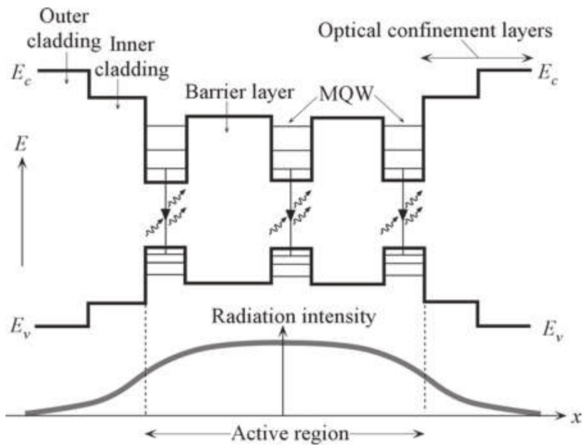


FIGURE 4.41 A simplified schematic diagram of multiple quantum well heterostructure laser diode. Electrons are injected by the forward current into quantum wells. The light intensity distribution is also shown. Most of the light is in the active region.

ultrathin layers of wide and narrow bandgap semiconductors as schematically sketched in Figure 4.41. The smaller bandgap layers are the individual QWs where electron confinement and lasing transitions take place, whereas the wider bandgap layers are the **barrier layers**. The active region now has multiple quantum wells. The active layer is cladded by two optical confinement layers, labeled as inner and outer cladding layers, with wider bandgaps, hence lower refractive indices. Most of the radiation is now within the active region containing the QWs as shown in Figure 4.41, which is the desired goal for increasing the stimulated emission rate. QW flooding is eliminated because the injected electrons are shared by the QWs. The only drawback is that the design must ensure that all the QWs are injected with sufficient electrons to exhibit optical gain.

The green, blue, and violet laser diodes in the market are based on MQWs using $In_xGa_{1-x}N\,(E_{g1})$ and $GaN\,(E_{g2})$ thin semiconductor layers. The overall LD structure is similar to the blue LEDs described in Chapter 3 [see Figure 3.39 (c)], but there is an optical cavity in the present case to ensure lasing operation.

As in the case of LEDs, there are major advantages to incorporating QWs into the laser diode structure. First, the threshold current can be lowered. Second is the overall higher efficiency of the device. Third, the optical gain curve is narrower, which should allow single-mode operation to be achieved more readily. Although the optical gain curve is narrower than the corresponding bulk device, the output spectrum from a quantum well device is not necessarily a single mode. The number of modes depends on the individual widths of the quantum wells. It is, of course, possible to combine an MQW design with an optical cavity using dielectric mirrors (*i.e.*, wavelength-selective reflectors) to generate only one mode. Many commercially available LDs are currently MQW devices.

EXAMPLE 4.11.1 A GaAs quantum well

Consider a very thin GaAs quantum well sandwiched between two wider bandgap semiconductor layers of AlGaAs ($Al_{0.33}Ga_{0.67}As$ in the present case). The QW depths from E_c and E_v are approximately 0.28 eV and 0.16 eV, respectively. Effective mass m_e^* of a conduction electron in GaAs is approximately $0.07m_e$, where m_e is the electron mass in vacuum. Calculate the first two electron energy levels for a quantum well of thickness 10 nm. What is the hole energy in the QW above E_v of GaAs if the hole effective mass $m_h^* \approx 0.50m_e$? What is the change in the emission wavelength with respect to bulk GaAs, for which $E_g = 1.42\,\text{eV}$? Assume infinite QW depths for the calculations.

Solution

As we saw in Chapter 3 (Section 3.12), the electron energy levels in the QW are with respect to the CB edge E_c in GaAs. Suppose that ε_n is the electron energy with respect to E_c in GaAs, or $\varepsilon_n = E_n - E_c$ in Figure 4.40 (b). Then, the energy of an electron in a one-dimensional infinite potential energy well is

$$\varepsilon_n = \frac{h^2 n^2}{8m_e^* d^2} = \frac{(6.626 \times 10^{-34})^2(1)^2}{8(0.07 \times 9.1 \times 10^{-31})(10 \times 10^{-9})^2} = 8.62 \times 10^{-21}\ \text{J} \quad \text{or} \quad 0.0538\ \text{eV}$$

where n is a quantum number, 1, 2, ..., and we have used $d = 10 \times 10^{-9}$ m, $m_e^* = 0.07m_e$, and $n = 1$ to find $\varepsilon_1 = 0.054$ eV. The next level from the same calculation with $n = 2$ is $\varepsilon_2 = 0.215$ eV.

The hole energy levels below E_v in Figure 4.40 (b) are given by

$$\varepsilon_{n'} = \frac{h^2 n'^2}{8m_h^* d^2}$$

where n' is the quantum number for the hole energy levels above E_v. Using $d = 10 \times 10^{-9}$ m, $m_h^* \approx 0.5m_e$, and $n' = 1$, we find $\varepsilon_1' = 0.0075$ eV.

The wavelength of emission from bulk GaAs with $E_g = 1.42$ eV is

$$\lambda_g = \frac{hc}{E_g} = \frac{(6.626 \times 10^{-34})(3 \times 10^8)}{(1.42)(1.602 \times 10^{-19})} = 874 \times 10^{-9}\ \text{m (874 nm)}$$

In the case of QWs, we must obey the selection rule that the radiative transition must have $\Delta n = n' - n = 0$. Thus, the radiative transition is from ε_1 to ε_1' so that the emitted wavelength is

$$\lambda_{QW} = \frac{hc}{E_g + \varepsilon_1 + \varepsilon_1'} = \frac{(6.626 \times 10^{-34})(3 \times 10^8)}{(1.42 + 0.0538 + 0.0075)(1.602 \times 10^{-19})}$$
$$= 838 \times 10^{-9}\ \text{m (838 nm)}$$

The difference is $\lambda_g - \lambda_{QW} = 36$ nm. We note that we assumed an infinite PE well. If we actually solve the problem properly by using a finite well depth,[31] then we would find $\varepsilon_1 \approx 0.031$ eV, $\varepsilon_2 \approx 0.121$ eV, $\varepsilon_1' \approx 0.007$ eV. The emitted photon wavelength is 848 nm and $\lambda_g - \lambda_{QW} = 26$ nm.

4.12 ELEMENTARY LASER DIODE CHARACTERISTICS

The output spectrum from a laser diode depends on two factors: the nature of the optical resonator used to build the laser oscillations and the optical gain curve (lineshape) of the active medium. The optical resonator is essentially a Fabry–Perot cavity as illustrated in Figure 4.42 (a) which can be assigned a length L, width W, and height H. The height is the same as the layer thickness d of the active layer in the heterostructure LD as shown in Figure 4.36 (a). The length L determines the **longitudinal mode** separation, whereas the width W and height H determine the **transverse modes**, or **lateral modes** in LD nomenclature. If the transverse dimensions (W and H) are sufficiently small, only the lowest transverse mode, TEM_{00} mode, will exit. This TEM_{00} mode, however, will have longitudinal modes whose separation depends on L. Figure 4.42 (a)

[31] Clear explanations and the proper calculations for QWs can be found in Mark Fox, *Optical Properties of Solids*, 2nd Edition (Oxford University Press, 2010), Ch. 6.

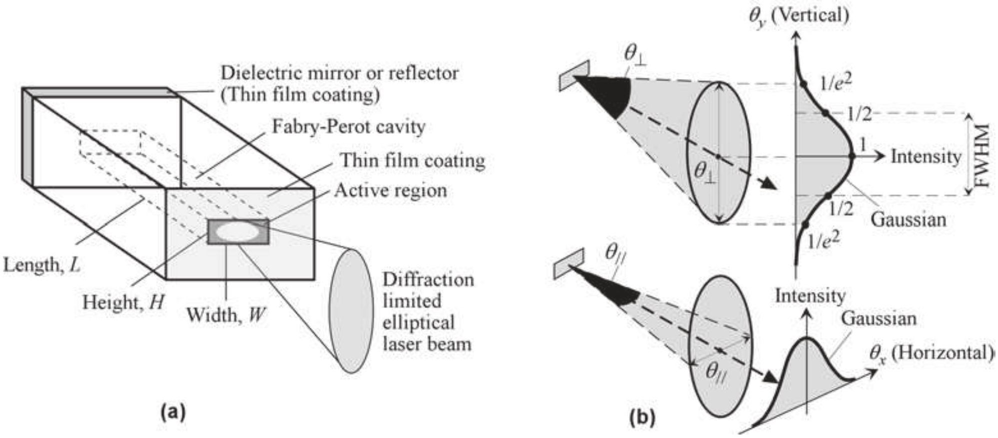


FIGURE 4.42 (a) The laser cavity definitions and the output laser beam characteristics. (b) Laser diode output beam astigmatism. The beam is elliptical, and is characterized by two angles, $\theta_\perp$ and $\theta_{//}$.

also shows that the emerging laser beam exhibits divergence. This is due to the diffraction of the waves at the cavity ends. The smallest aperture (*W* in the figure) causes the greatest diffraction.

The emerging beam from an edge emitting LD has an elliptical cross-section and exhibits astigmatism as shown in Figure 4.42 (b). The angular spread in the vertical (y) and horizontal (x) directions (perpendicular or parallel to the plane of the diode junction) are not the same which leads to a nearly elliptical beam cross-section. Typically, the active layer cavity height H is smaller than the cavity width W. Since the angular spread of the emerging laser beam is diffraction controlled by the emission aperture, smaller the diffracting height H, the wider is the diffracted output beam in the vertical direction. The vertical angular spread $\theta_\perp$ is more than the horizontal spread $\theta_{//}$ as illustrated in Figure 4.42 (b).

The far field intensity distributions (far from the crystal facet) in the y and x directions are nearly Gaussian in the angular displacements θ_y and θ_x. The spreads $\theta_\perp$ and $\theta_{//}$ refer to the angular separation of the e^{-2} points on the Gaussian curve as indicated in Figure 4.42 (b). In some specifications, $\theta_\perp$ and $\theta_{//}$ refer to the angular separation of half-intensity points (full width at half maximum, FWHM); the two are related by $\theta(e^{-2}) = 1.7\theta$ (FWHM). The output divergence is normally quoted as $\theta_\perp \times \theta_{//}$. The output beam from the LD in most applications is either put through a lens-prism system to collimate the beam, or it is coupled, using a suitable (an anamorphic) lens, into a fiber. The manipulation of the LD output beam through optics into a well-defined collimated beam with a circular cross-section ($\theta_\perp = \theta_{//}$) is usually referred to as the **collimation** and **circularization** of the output beam. (Circularization should not be confused with circularly polarized light.) The collimation and circularization can be quite complicated depending on the exact requirements and space available for the optics.

The actual modes that exist in the output spectrum of a LD will depend on the optical gain these modes will experience. The term output spectrum strictly refers to the spectral power vs. λ characteristic of the emitted radiation. The **spectral power** $P_{o\lambda}$ is defined as $dP_o/d\lambda$, that is, optical power emitted per unit wavelength. A typical example is shown in Figure 4.43 for an index guided LD. The area under the $P_{o\lambda}$ vs. λ spectrum curve represents the total power P_o emitted. The spectrum, is either multimode or single mode depending on the optical resonator structure and the pumping current level. As apparent from Figure 4.43, there is a shift in the peak emission wavelength with the current. In most LDs, the shift in the peak wavelength is due to the Joule heating of the semiconductor

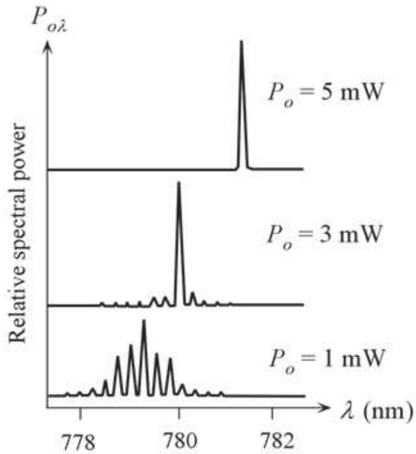


FIGURE 4.43 Output spectra (spectral power density) of lasing emission from an index-guided edge emitting LD. At sufficiently high diode currents corresponding to high optical power, the operation becomes single mode. (*Note*: Relative spectral power scale applies to each spectrum individually and not between spectra.)

at high currents, which changes the properties of the semiconductor and the laser cavity. High stability LDs have a thermoelectric cooler to maintain the temperature at the desired level. The multimode spectrum at low output power typically becomes single mode at high output powers. The shift in the peak wavelength is partly due to the Joule heating of the semiconductor at high currents. In contrast, the output spectrum of most gain-guided LDs tends to remain multimode even at high diode currents. For many applications such as optical communications, the LD has to be designed so that the emission is single mode and the shift in the emission wavelength with increasing output power is small.

The LD output characteristics are temperature sensitive. Figure 4.44 shows the changes in the optical output power vs. diode current characteristics with the case temperature. As the temperature increases, the threshold current increases steeply, typically as the exponential of the absolute temperature, that is,

Threshold current and temperature

$$I_{th} = A\exp(T/T_o) \tag{4.12.1}$$

where A and T_o are constants that have the units of current (A) and temperature (K), respectively; T_o is sometimes called the characteristic temperature for the I_{th} vs. T dependence.

The output spectrum also changes with the temperature. In the case of a single-mode LD, the peak emission wavelength λ_o exhibits jumps at certain temperatures as apparent in Figures 4.45 (a) and (b). A jump corresponds to a **mode hop** in the output. That is, at the new

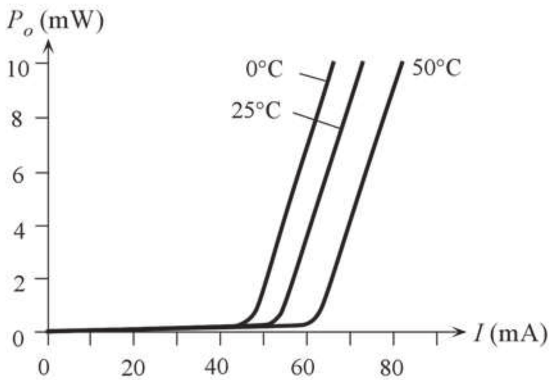


FIGURE 4.44 Output optical power vs. diode current at three different temperatures. The threshold current shifts to higher temperatures.

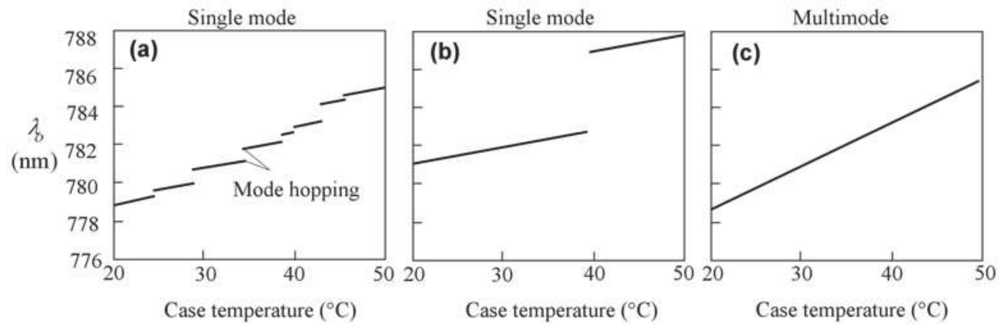


FIGURE 4.45 Peak wavelength λ_o vs. case temperature characteristics. (a) Mode hops in the output spectrum of a single-mode LD. (b) Restricted mode hops and none over the temperature range of interest (20−40°C). (c) Output spectrum from a multimode LD.

operating temperature, another mode fulfills the laser oscillation condition, which means a discrete change in the laser oscillation wavelength. Between mode hops, λ_o increases slowly with the temperature due to the slight increase in the refractive index n (and also the cavity length) with temperature. If mode hops are undesirable, then the device structure must be such to keep the modes sufficiently separated. In contrast, the output spectrum of a gain-guided laser has many modes, so that λ_o vs. T behavior tends to follow the changes in the bandgap (the optical gain curve) rather than the cavity properties as illustrated in Figure 4.45 (c). Highly stabilized LDs are usually marketed with thermoelectric coolers integrated into the diode package to control the device temperature. Many high power laser diodes have a photodetector within the packaging to monitor the output intensity. The output from this photodetector is called the **monitoring current**, whose value (a fraction of a milliamp) is specified by the manufacturer for a given output optical power from the laser diode.

One commonly stated important and useful laser diode parameter is the slope efficiency, which determines the optical power P_o in the lasing emission in terms of the diode current I above the threshold current I_{th}. The **slope efficiency** (SE) η_{slope} is the slope of P_o vs. I above threshold

$$\eta_{slope} = \left(\frac{\Delta P_o}{\Delta I}\right)_{\text{above threshold}} \approx \frac{P_o}{I - I_{th}} \tag{4.12.2}$$

Slope efficiency

and is measured in W A^{-1} or mW mA^{-1}. The slope efficiency depends on the LD structure as well as the semiconductor packaging, and typical values for commonly available LDs are close to 1 W/A. There are several laser diode efficiency definitions as follows:

The **external quantum efficiency** η_{EQE} of a laser diode is defined as

$$\eta_{EQE} = \frac{\text{Number of output photons from the diode per second}}{\text{Number of injected electrons into the diode per second}} = \frac{P_o/h\upsilon}{I/e} \approx \frac{eP_o}{E_g I} \tag{4.12.3}$$

External quantum efficiency

The term *differential* is normally used to identify efficiencies related to changes above the threshold. The **external differential quantum efficiency** η_{EDQE} is defined for operation above threshold as

External differential quantum efficiency

$$\eta_{EDQE} = \frac{\text{Increase in number of output photons from diode per second}}{\text{Increase in number of injected electrons into diode per second}} \tag{4.12.4a}$$

External differential quantum efficiency

$$\eta_{\text{EDQE}} = \frac{\Delta P_o/h\upsilon}{\Delta I/e} = \eta_{\text{slope}}\frac{e}{h\upsilon} \approx \left(\frac{e}{E_g}\right)\frac{P_o}{I - I_{\text{th}}} \quad \textbf{(4.12.4b)}$$

The **internal quantum efficiency** (IQE) η_{IQE}, as in the case of LEDs, measures what fraction of injected electrons recombines radiatively. If τ_r is the radiative recombination time and τ_{nr} is the nonradiative recombination time, then

Internal quantum efficiency

$$\eta_{\text{IQE}} = \frac{\text{Rate of radiative recombination}}{\text{Rate of all recombination processes}} = \frac{1/\tau_r}{1/\tau_r + 1/\tau_{nr}} \quad \textbf{(4.12.5)}$$

Equivalently, we could have defined IQE as the number of generated photons divided by the number of injected electrons.

The **internal differential quantum efficiency** (IDQE) represents the increase in the number of photons generated inside the gain medium per unit increase in the injected electrons, *i.e.*, the current, above threshold. Thus, *above threshold*,

Internal differential quantum efficiency

$$\eta_{\text{IDQE}} = \frac{\text{Increase in number of photons generated internally per second}}{\text{Increase in number of injected electrons into diode per second}} \quad \textbf{(4.12.6)}$$

Put differently, if the current increases by ΔI above threshold, increase in the injected electrons is $\Delta I/e$. The increase in the number of photons generated *internally* is then $\eta_{\text{IDQE}} \times \Delta I/e$. However, these photons have to escape the cavity into the output for which we would need the **extraction efficiency** (EE) η_{EE}.

Suppose that we are interested in the output from one of the mirrors, say R_1. The *loss* at R_1 is the transmitted output. If there were no losses at the other mirror ($R_2 = 1$) and no internal losses ($\alpha_s = 0$), then the radiation in the cavity would eventually be coupled out from R_1 and the EE efficiency η_{EE} would be unity. Thus, what is important is the relative ratio of the loss due to R_1 to all losses in the cavity. The **extraction efficiency** is defined as the ratio of loss coefficient due to R_1 to that due to total losses, α_t,

$$\text{Extraction efficiency} = \frac{\text{Loss from the exit cavity end}}{\text{Total loss}} \quad \textbf{(4.12.7)}$$

i.e.,

Extraction efficiency

$$\eta_{\text{EE}} = (1/2L)\ln(1/R_1)/\alpha_t$$

where $\alpha_t = \alpha_s + (1/2L)\ln(1/R_1) + (1/2L)\ln(1/R_2)$ is the total loss.

The **power conversion efficiency**, η_{PCE}, or the **external power efficiency**, gauges the overall efficiency of the conversion from the input of electrical power to the output of optical power, *i.e.*,

External power efficiency

$$\eta_{\text{PCE}} = \frac{\text{Optical output power}}{\text{Electrical input power}} = \frac{P_o}{IV} \approx \eta_{\text{EQE}}\left(\frac{E_g}{eV}\right) \quad \textbf{(4.12.8)}$$

Although this is not generally quoted in data sheets, it can be easily determined from the output power at the operating diode current and voltage. In some modern LDs this may be as high as 30%.

The expressions on the right-hand side of Eqs. (4.12.2) to (4.12.4) and (4.12.8) give the corresponding formulas in terms of P_o, I, V, and E_g for each definition and are approximate due to the assumptions used in deriving these equations. It is left as an exercise to show that these equations follow directly from the definitions. (Assume $h\upsilon \approx E_g$.)

Table 4.4 shows the main characteristics of a few commercial laser diodes emitting in the red (639−670 nm) and violet (405 nm) from low to high output powers. All are MQW laser diodes. The violet emission uses InGaN/GaN MQWs, whereas the red LDs are based on AlGaInP/GaInP alloys. Higher output power diodes have larger geometries and require larger threshold currents as can be seen from the table. Power conversion efficiencies of the LDs in the table are in the range 6–30%, which are typical values. Figure 4.46 shows typical values for the threshold current I_{th}, slope efficiency, and power conversion efficiency for a selection of commercial red LDs (36) with different optical output powers from 3 mW to 500 mW. The figure also highlights that there can be significant variations in I_{th} even at a given power level between different LDs, as can be seen for the 10-mW output. The trend in higher threshold current for higher output power, however, is quite clear. The η_{slope} values in Table 4.4 range from 0.9 to 1.7, which are typical for commercial LDs.

TABLE 4.4 Typical characteristics for a few selected red and violet commercial laser diodes. All LDs are MQW structures and have FP cavities

LD	P_o (mW)	λ (nm)	I_{th} (mA)	I (mA)	V (V)	$\theta_\perp$	$\theta_{//}$	η_{slope} (W A^{-1})	η_{PCE} (%)
Red	500	670	400	700	2.4	21°	10°	1.0	30
Red	100	660	75	180	2.5	18°	9°	1.0	22
Red	50	660	60	115	2.3	17°	10°	0.90	19
Red	10	639	30	40	2.3	21°	8°	1.0	11
Violet	400	405	160	390	5.0	45°	15°	1.7	21
Violet	120	405	45	120	5.0	17°	8°	1.6	20
Violet	10	405	26	35	4.8	19°	8.5°	1.1	6

Note: Violet lasers are based on InGaN/GaN MQW, and red LEDs use mainly AlGaInP/GaInP MQW.

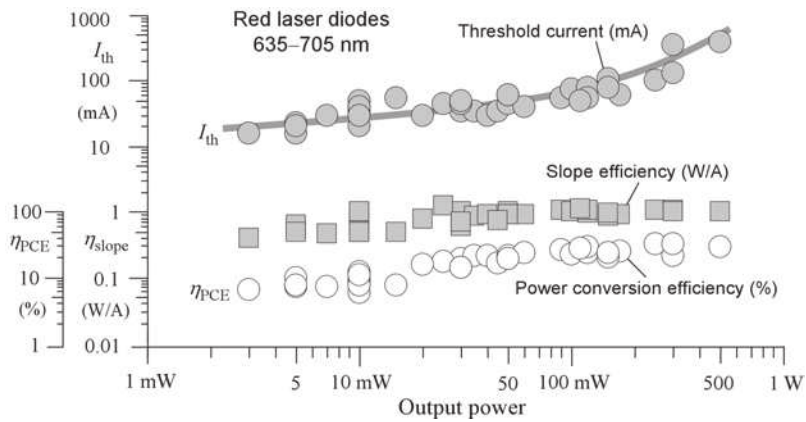
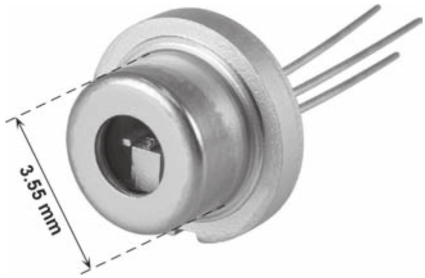


FIGURE 4.46 Typical values for the threshold current I_{th}, slope efficiency (η_{slope}), and power conversion efficiency (η_{PCE}) for 36 commercial red LDs with different optical output powers from 3 mW to 500 mW.



A red emitting (642 nm) high power laser diode that can provide CW laser beam at 150 mW. The threshold current is 110 mA, operating current and voltage are 280 mA and 2.6 V respectively, for 150 mW output power. The device structure is based on AlGaInP MQWs. *(Opnext-Hitachi laser diode. Courtesy of Thorlabs.)*

EXAMPLE 4.12.1 Laser output wavelength variation with temperature

The refractive index n of GaAs is approximately 3.6 and it has a temperature dependence $dn/dT \approx 2.0 \times 10^{-4}\ \mathrm{K}^{-1}$. Estimate the change in the emitted wavelength at around 870 nm per degree change in the temperature for a given mode.

Solution

Consider a particular given mode with wavelength λ_m,

$$m\left(\frac{\lambda_m}{2n}\right) = L$$

If we differentiate λ_m with respect to temperature,

$$\frac{d\lambda_m}{dT} = \frac{d}{dT}\left[\frac{2}{m}nL\right] \approx \frac{2L}{m}\frac{dn}{dT}$$

where we neglected the change in the cavity length with temperature. Substituting for L/m in terms of λ_m,

$$\frac{d\lambda_m}{dT} = \frac{\lambda_m}{n}\frac{dn}{dT} = \frac{870\ \mathrm{nm}}{3.6}(2 \times 10^{-4}\ \mathrm{K}^{-1}) = 0.048\ \mathrm{nm\ K}^{-1}$$

Note that we have used n for a passive cavity, whereas n above should be the effective refractive index of the *active* cavity which will also depend on the optical gain of the medium, and hence its temperature dependence is likely to be somewhat higher than the dn/dT value we used. It is left as an exercise to show that the changes in λ_m due to the expansion of the cavity length with temperature is much less than that arising from dn/dT. The linear expansion coefficient of GaAs is $6 \times 10^{-6}\ \mathrm{K}^{-1}$.

EXAMPLE 4.12.2 Laser diode efficiencies for a sky-blue LD

Consider a 60-mW blue LD (Nichia SkyBlue NDS4113), emitting at a peak wavelength of 488 nm. The threshold current is 30 mA. At a forward current of 100 mA and a voltage of 5.6 V, the output power is 60 mW. Find the slope efficiency, PCE, EQE, and EDQE.

Solution

From the definition in Eq. (4.12.2),

$$\eta_{\text{slope}} = P_o/(I - I_{\text{th}}) = (60\ \mathrm{mW})/(100 - 30\ \mathrm{mA}) = 0.86\ \mathrm{mW\ mA}^{-1}$$

From Eq. (4.12.8), PCE is

$$\eta_{\text{PCE}} = P_o/IV = (60\text{ mW})/\left[(100\text{ mA})(5.6\text{ V})\right] = 0.11 \quad \text{or} \quad 11\%$$

We can find the EQE from Eq. (4.12.3) but we need $h\upsilon$, which is hc/λ. In eV,

$$h\upsilon\,(\text{eV}) = 1.24/\lambda\,(\mu\text{m}) = 1.24/0.488 = 2.54\text{ eV}$$

EQE is given by Eq. (4.12.3)

$$\begin{aligned}\eta_{\text{EQE}} &= (P_o/h\upsilon)/(I/e) = \left[(60\times10^{-3})/(2.54\times1.6\times10^{-19})\right]/\left[(100\times10^{-3})/(1.6\times10^{-19})\right]\\ &= 0.24 \quad \text{or} \quad 24\%\end{aligned}$$

Similarly, η_{EDQE} is given by Eq. (4.12.4b) above threshold,

$$\begin{aligned}\eta_{\text{EDQE}} &= (\Delta P_o/h\upsilon)/(\Delta I/e) \approx (P_o/h\upsilon)/\left[(I-I_{\text{th}})/e\right]\\ &= \left[(60\times10^{-3})/(2.54\times1.6\times10^{-19})\right]/\left[(100\times10^{-3}-30\times10^{-3})/(1.6\times10^{-19})\right]\\ &= 0.34 \quad \text{or} \quad 34\%\end{aligned}$$

The EDQE is higher than the EQE because most injected electrons above I_{th} are used in stimulated recombinations. EQE gauges the total conversion efficiency from all the injected electrons brought by the current to coherent output photons. But, a portion of the current is used in pumping the gain medium.

EXAMPLE 4.12.3 Laser diode efficiencies

Consider an InGaAs FP semiconductor laser diode that emits CW radiation at 1310 nm. The cavity length (L) is 200 μm. The internal loss coefficient $\alpha_s = 20\text{ cm}^{-1}$, $R_1 = R_3 \approx 0.33$ (cleaved ends). Assume that the internal differential quantum efficiency, IDQE, is close to 1. The threshold current is 5 mA. What is the output power P_o at $I = 20$ mA? The forward voltage is about 1.3 V. What are the EDQE and conversion efficiency?

Solution

From the definition of IDQE in Eq. (4.12.6), the number of internal coherent photons generated per second above threshold is $\eta_{\text{IDQE}}(I - I_{\text{th}})/e$. Thus,

$$\text{Internal optical power generated} = h\upsilon \times \eta_{\text{IDQE}}(I - I_{\text{th}})/e$$

The extraction efficiency η_{EE} then couples a portion of this optical power into the output radiation. The output power P_o is then $\eta_{\text{EE}} \times h\upsilon \times \eta_{\text{IDQE}}(I - I_{\text{th}})/e$. Thus,

$$P_o = \eta_{\text{EE}}\eta_{\text{IDQE}}h\upsilon(I - I_{\text{th}})/e \qquad \textbf{(4.12.9)}$$

Output power vs. current

The slope efficiency from Eq. (4.12.2) is

$$\eta_{\text{slope}} = \Delta P_o/\Delta I = \eta_{\text{EE}}\eta_{\text{IDQE}}(h\upsilon/e) \qquad \textbf{(4.12.10)}$$

Slope efficiency

Further, from the definition of EDQE and Eq. (4.12.9) is

$$\eta_{\text{EDQE}} = (\Delta P_o/h\upsilon)/(\Delta I/e) = (P_o/h\upsilon)/\left[(I - I_{\text{th}})/e\right] = \eta_{\text{EE}}\eta_{\text{IDQE}} \qquad \textbf{(4.12.11)}$$

External differential quantum efficiency

We can now calculate the quantities needed. The total loss coefficient is

$$\alpha_t = \alpha_s + (1/2L)\ln(1/R_1R_2) = 2000 + (2\times200\times10^{-6})^{-1}\ln(0.33\times0.33)^{-1} = 7543\text{ m}^{-1}$$

The extraction efficiency is

$$\eta_{EE} = (1/2L)\ln(1/R_1)/\alpha_t = (2 \times 200 \times 10^{-6})^{-1}\ln(1/0.33)/(7543) = 0.37 \quad \text{or} \quad 37\%$$

Thus, using $I = 20$ mA and $h\upsilon = hc/\lambda$ in Eq. (4.12.9),

$$P_o = (0.37)(1)\left[(6.62 \times 10^{-34})(3 \times 10^8)/(1310 \times 10^{-9})\right]\left[(0.02 - 0.005)/(1.6 \times 10^{-19})\right]$$
$$= 5.2 \text{ mW}$$

The slope efficiency from Eq. (4.12.10) is

$$\eta_{\text{slope}} = \Delta P_o/\Delta I = (5.2\text{ mW} - 0)/(20\text{ mA} - 5\text{ mA}) = 0.35\text{ mW mA}^{-1}$$

The EDQE from Eq. (4.12.11) is

$$\eta_{EDQE} = \eta_{EE}\eta_{IDQE} = 0.37 \quad \text{or} \quad 37\%$$

The power conversion efficiency $\eta_{PCE} = P_o/IV = 5.2\text{ mW}/(20\text{ mA} \times 1.3\text{ V}) = 0.20$ or 20%.

4.13 STEADY STATE SEMICONDUCTOR RATE EQUATIONS: THE LASER DIODE EQUATION

A. Laser Diode Equation

Consider a double heterostructure lLD under forward bias as in Figures 4.47 (a) and (b). The current I injects electrons into the active region d and the radiative recombination of these electrons with holes in this layer generates the coherent radiation in the cavity. Suppose that n is the injected electron concentration and N_{ph} is the coherent photon concentration in the active layer. We have to be careful since not all the coherent radiation will be within the active region as illustrated in Figure 4.47 (c) and in Figure 4.36 (d), even if there is a waveguide formed by the refractive index profile between the active layer and the neighboring confining layers as shown in Figure 4.36 (c). A **radiation confinement factor** Γ is used to represent the fraction of the

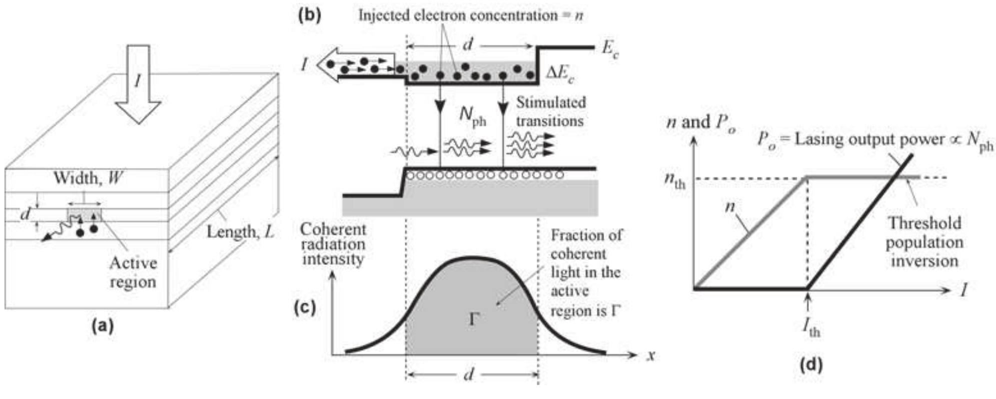


FIGURE 4.47 A highly simplified and idealized description of a semiconductor laser diode for deriving the LD equation. (a) The heterostructure laser diode structure. (b) The current I injects electrons in the conduction band, and these electrons recombine radiatively with the holes in the active region. (c) The coherent radiation intensity across the device; only a fraction Γ is within the active region where there is optical gain. (d) Injected electron concentration n and coherent radiation output power P_o vs. diode current I. The current represents the pump rate.

coherent radiation within the active layer, which is unity for perfect optical confinement. In the following we simply assume $\Gamma = 1$.

The current carries the electrons into the active layer where they recombine with holes radiatively as indicated in Figure 4.47 (b). Under steady state operation, the rate of electron injection into the active layer by the current I is equal to their rate of recombination by spontaneous emission *and* stimulated emission (neglecting nonradiative recombinations). Thus,

$$\text{Rate of electron injection by current } I = \text{Rate of spontaneous emissions} + \text{Rate of stimulated emissions}$$

Thus, per unit volume, we have

$$\frac{I}{eLWd} = \frac{n}{\tau_r} + CnN_{\text{ph}} \tag{4.13.1}$$

where τ_r is the average time for spontaneous radiative recombination, simply called the **radiative lifetime**, and C is a proportionality constant for stimulated emissions (a bit like B_{21}) that, among other factors, includes the probability per unit time that a photon stimulates an electron to undergo a radiative transition. The second term on the right represents the stimulated emission rate, which depends on the concentration of electrons in the conduction band, n, and the coherent photon concentration N_{ph} in the active layer. N_{ph} includes only those coherent photons encouraged by the optical cavity, that is, a mode of the cavity. As the current increases and provides more pumping, N_{ph} increases (helped by the optical cavity), and eventually the stimulated term dominates the spontaneous term (as illustrated in Figure 4.35). The output light power P_o is proportional to N_{ph}.

Consider the coherent photon concentration N_{ph} in the cavity. Under *steady state conditions*,

$$\text{Rate of coherent photon loss in the cavity} = \text{Rate of stimulated emissions}$$

that is,

$$\frac{N_{\text{ph}}}{\tau_{\text{ph}}} = CnN_{\text{ph}} \tag{4.13.2}$$

where τ_{ph} is the *average time* for a photon to be lost from the cavity due to transmission through the end-faces, and also by absorption and scattering in the semiconductor cavity. It is the **photon cavity lifetime** as calculated in Example 4.6.2. If α_t is the total attenuation coefficient representing all these loss mechanisms, inside the cavity and also at the ends of the cavity, then the power in a light wave, in the absence of amplification, decreases as $\exp(-\alpha_t x)$, which is equivalent to a decay in time as $\exp(-t/\tau_{\text{ph}})$, where $\tau_{\text{ph}} = n/(c\alpha_t)$ and n is the refractive index.

In semiconductor laser science, the threshold electron concentration n_{th} and threshold current I_{th} refer to that condition when the stimulated emission just overcomes the spontaneous emission and the total loss mechanisms inherent in τ_{ph}. This occurs when the injected electron concentration n reaches n_{th}, the **threshold electron concentration**. From Eq. (4.13.2), this is when

$$n_{\text{th}} = \frac{1}{C\tau_{\text{ph}}} \tag{4.13.3}$$

Threshold concentration

This is the point when coherent radiation gain in the active layer by stimulated emission just balances all the cavity losses (represented by τ_{ph}) plus losses by spontaneous emission which is random. When the current exceeds I_{th}, the output optical power increases sharply with the current as apparent in Figure 4.35, so we can just as well take $N_{ph} \approx 0$ when $I = I_{th}$ in Eq. (4.13.1), which gives

Threshold current

$$I_{th} = \frac{n_{th} eLWd}{\tau_r} \tag{4.13.4}$$

Clearly the threshold current decreases with d, L, and W which explains the reasons for the heterostructure and stripe geometry lasers and the avoidance of the homojunction laser. Further, τ_{ph} can be made longer, and hence n_{th} and I_{th} smaller, by reducing the cavity losses.

When the current exceeds the threshold current, the excess carriers above n_{th} brought in by the current recombine by stimulated emission. The reason is that above threshold, the active layer has optical gain and therefore builds up coherent radiation quickly, and stimulated emission rate depends on N_{ph}. The steady state electron concentration remains constant at n_{th}, though the rates of carrier injection and stimulated recombination have increased. Above threshold, we therefore have $n = n_{th}$ in Eq. (4.13.1), and we can substitute for n_{th}/τ_r from Eq. (4.13.4)

$$\frac{I}{eLWd} = \frac{n_{th}}{\tau_r} + Cn_{th}N_{ph} \tag{4.13.5}$$

which shows that above threshold, increasing I, as expected, increases N_{ph}. We can tidy up Eq. (4.13.5) by first substituting for C from Eq. (4.13.3) and then substituting for n_{th} from Eq. (4.13.4) to find

Coherent photon concentration

$$N_{ph} = \frac{\tau_{ph}}{eLWd}(I - I_{th}) \tag{4.13.6}$$

To find the optical output power P_o, consider the following. It takes $\Delta t = nL/c$ seconds for photons to cross the laser cavity length L. Only half of the photons, $1/2\,N_{ph}$, in the cavity would be moving toward the output face of the crystal at any instant. Only a fraction $(1 - R)$ of the radiation power will escape. Thus, the output optical power P_o is

$$P_o = \frac{\left(\frac{1}{2}N_{ph}\right)(\text{Cavity Volume})(\text{Photon energy})}{\Delta t}(1 - R)$$

and using N_{ph} from Eq. (4.13.6), we finally obtain the **laser diode equation**

Laser diode equation

$$P_o = \left[\frac{hc^2\tau_{ph}(1 - R)}{2en\lambda L}\right](I - I_{th}) \tag{4.13.7}$$

We can also express Eq. (4.13.7) in terms of the output coherent radiation intensity $I_o(= P_o/Wd)$ and the current densities $J(= I/WL)$ and $J_{th}(= I_{th}/WL)$,

Laser diode equation

$$I_o = \left[\frac{hc^2\tau_{ph}(1 - R)}{2en\lambda d}\right](J - J_{th}) \tag{4.13.8}$$

We neglected the small spontaneous radiation that will also be present in the output in Eq. (4.13.7) in our derivation. This spontaneous radiation in a laser operating above the threshold would be small.

Equation (4.13.7) represents a linear coherent light output power vs. diode current behavior as illustrated in Figure 4.47 (d). Note the similarity between the case for semiconductors and that for gas lasers in Figure 4.23. The plot in Figure 4.47 (d) represents steady state operation. When the current is above I_{th}, the injected electrons rapidly recombine and emit coherent photons, and n_{th} does not change. n becomes clamped at $n = n_{th}$ for $I > I_{th}$. The coherent radiation appears as output only when $I > I_{th}$. At a given current level above the threshold, upon applying the bias, initially the gain in the active regions is actually more than the threshold gain, and n exceeds the threshold concentration n_{th} until oscillations are built and steady state is reached; at that point $n = n_{th}$ and $g = g_{th}$. The semiquantitative steady state approach above is a special case of the more general semiconductor rate equations described in more advanced textbooks where the time response of the laser diode is also analyzed.

B. Optical Gain Curve, Threshold, and Transparency Conditions

The simple rate equation analysis above does not allow us to determine n_{th} and hence J_{th} without knowing the threshold gain. We know that the threshold population inversion is determined by the threshold gain as in Eq. (4.6.8). The latter equation was derived for a two-level system, and we need its corresponding equation applicable for semiconductors; this can be found in more advanced textbooks. There is one further modification. The threshold gain g_{th} must be modified to account for the less than 100% confinement of radiation into the active region. This is easily done by rewriting the g_{th}-condition as

$$\Gamma g_{th} = \alpha_t = \alpha_s + \frac{1}{2L}\ln\left(\frac{1}{R_1 R_2}\right) \qquad \textbf{(4.13.9)}$$

Threshold optical gain

where Γ represents the fraction of the coherent optical radiation within the active region. The gain g works on the radiation within the cavity, which means that we must multiply g with Γ to account for less than perfect optical confinement. In Eq. (4.13.9), α_s represents various losses in the semiconductor active region as well as the cladding.[32] The right-hand term in Eq. (4.13.9) represents the total losses inside and at the ends of the cavity. The α_s losses in semiconductor lasers are free carrier absorption in the active region, scattering in both the active and cladding regions, and absorption in the cladding (which has no net optical gain). As Γ decreases, resulting in less coherent radiation confinement, g_{th} has to be larger so that Eq. (4.13.9) can be satisfied and the gain Γg_{th} can make up for the losses on the right-hand side of Eq. (4.13.9). The active region formed by the heterostructure is normally sufficiently small ($d \sim 0.1$ μm) that there is only one mode (*i.e.*, a single lateral mode) and this mode has a mode field diameter that can substantially penetrate the confining (cladding) layers. The optical confinement factor Γ can easily be as low as 0.2.

The optical gain spectrum $g(\upsilon)$ under sufficiently large forward bias has a frequency dependence that is sketched in Figure 4.33 (b), when the LD has been forward biased to place the quasi-Fermi levels E_{Fn} in the CB and E_{Fp} in the VB as in Figure 4.33 (a). To find the optical gain curve $g(\upsilon)$ we need to consider the density of states in the CB and the VB, their occupation statistics through the Fermi–Dirac function, and positions of E_{Fn} and E_{FP}; the latter depend on the injected carrier concentrations n. Figure 4.48 (a) shows the optical gain curve $g(\upsilon)$ as a function

[32] In the case of semiconductors, some books use α_i to represent internal losses, instead of α_s, but the meaning is the same. α_t in Eq. (4.13.9) is an *effective distributed attenuation coefficient* through the cavity length that represents all losses, that is, both internal and end losses.

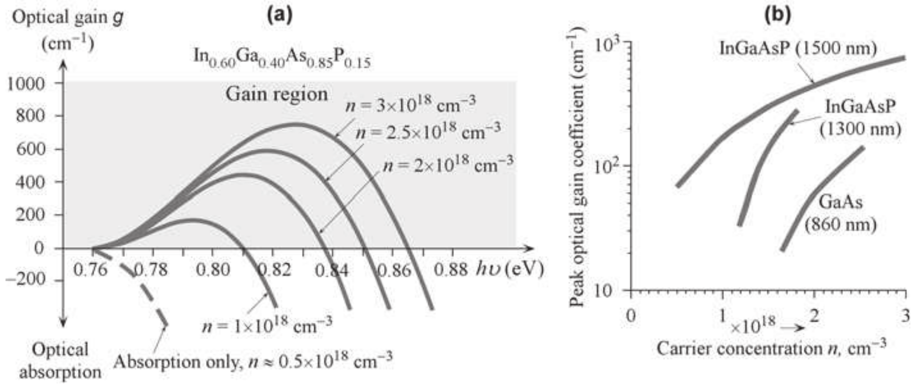


FIGURE 4.48 (a) Optical gain g vs. photon energy for an InGaAsP active layer (in a 1500-nm LD) as a function of injected carrier concentration n from $1 \times 10^{18}\,\text{cm}^{-3}$ to $3 \times 10^{18}\,\text{cm}^{-3}$. (The model described in Leuthold *et al.*, *J. Appl. Phys.*, *87*, 618, 2000, was used to find the gain spectra at different carrier concentrations.) (b) The dependence of the peak gain coefficient (maximum g) on the injected carrier concentration n for GaAs (860 nm), $In_{0.72}Ga_{0.28}As_{0.6}P_{0.4}$ (1300 nm), and $In_{0.60}Ga_{0.40}As_{0.85}P_{0.15}$ (1500 nm) active layers. (*Source:* Data combined from J. Singh, *Electronic and Optoelectronic Properties of Semiconductor Structures,* Cambridge University Press, 2003, p. 390; N. K. Dutta, *J. Appl. Phys., 51,* 6095, 1980; J. Leuthold *et al., J. Appl. Phys., 87,* 618, 2000.)

of photon energy for an $In_{0.60}Ga_{0.40}As_{0.85}P_{0.15}$ (1500 nm) active layer with different amounts of injected carrier concentration n. Notice also the shift in the peak frequency (where the gain is maximum) and the increase in the bandwidth with increasing injection. What is important is the injected carrier concentration that makes the gain g equal to the threshold gain g_{th}, which would allow laser oscillations to be maintained in the cavity. As the LD current and hence the injected carrier concentration n increase, the peak gain becomes larger as shown in Figure 4.48 (b) a$In_{0.60}Ga_{0.40}As_{0.85}P_{0.15}$ (1500 nm). The optical gain in InGaAsP alloys is greater than that in GaAs.

There is one further definition of importance in laser diode engineering. The threshold current and carrier concentration I_{th} and n_{th} correspond to the gain coefficient g_{th} just balancing all the cavity losses, including end losses, *i.e.,* $g_{th} = \alpha_t$, so that a steady state laser oscillation can be maintained. **Transparency**, on the other hand, refers to the condition that makes the gain just balance the absorption within the active medium so that the net optical gain is unity within the active medium. Thus, the active medium provides no net gain, and neither does it attenuate so that it becomes **transparent** to the radiation through it. The diode current and the injected carrier concentration for attaining transparency are denoted by I_T and n_T, which are lower than the threshold I_{th} and n_{th}.

EXAMPLE 4.13.1 Threshold current and optical output power from a Fabry–Perot heterostructure laser diode

Consider GaAs DH laser diode that lases at 860 nm. It has an active layer (cavity) length L of 250 μm. The active layer thickness d is 0.15 μm and the width W is 5 μm. The refractive index is 3.6, and the attenuation coefficient α_s inside the cavity is $10^3\,\text{m}^{-1}$. The required threshold gain g_{th} corresponds to a threshold carrier concentration $n_{th} \approx 2 \times 10^{18}\,\text{cm}^{-3}$. The radiative lifetime τ_r in the active region can be found (at least approximately) by using $\tau_r = 1/Bn_{th}$, where B is the direct recombination coefficient, and assuming strong injection as will be the case for laser diodes [see Eq. (3.8.7) in Chapter 3]. For GaAs, $B \approx 2 \times 10^{-16}\ \text{m}^3\ \text{s}^{-1}$. What are the threshold current density and threshold current? Find the output optical power at $I = 1.5I_{th}$, and the external slope efficiency η_{slope}. How would $\Gamma = 0.5$ affect the calculations?

Solution

The reflectances at the each end are the same (we assume no other thin film coating on the ends of the cavity) so that $R = (n-1)^2/(n+1)^2 = 0.32$. The total attenuation coefficient α_t and hence the threshold gain g_{th}, assuming $\Gamma = 1$ in Eq. (4.13.9), is

$$g_{th} = \alpha_t = (10\ \text{cm}^{-1}) + \frac{1}{(2 \times 250 \times 10^{-4}\ \text{cm})} \ln\left[\frac{1}{(0.32)(0.32)}\right] = 55.6\ \text{cm}^{-1}$$

From Figure 4.48 (b), at this gain of 56 cm^{-1}, $n_{th} \approx 2 \times 10^{18}\ \text{cm}^{-3}$. This is the threshold carrier concentration that gives the right gain under ideal optical confinement, with $\Gamma = 1$.

The radiative lifetime $\tau_r = 1/Bn_{th} = 1/\left[(2 \times 10^{-16}\ \text{m}^3\ \text{s}^{-1})(2 \times 10^{24}\ \text{m}^{-3})\right] = 2.5\ \text{ns}$

Since $J = I/WL$, the threshold current density from Eq. (4.13.4) is

$$J_{th} = \frac{n_{th}ed}{\tau_r} = \frac{(2 \times 10^{24}\ \text{m}^{-3})(1.6 \times 10^{-19}\ \text{C})(0.15 \times 10^{-6}\ \text{m})}{(2.5 \times 10^{-9}\ \text{s})}$$
$$= 1.9 \times 10^7\ \text{A m}^{-2} \quad \text{or} \quad 1.9\ \text{kA cm}^{-2} \quad \text{or} \quad 19\ \text{A mm}^{-2}$$

The threshold current itself is

$$I_{th} = (WL)J_{th} = (5 \times 10^{-6}\ \text{m})(250 \times 10^{-6}\ \text{m})(1.9 \times 10^7\ \text{A m}^{-2}) = 0.024\ \text{A} \quad \text{or} \quad 24\ \text{mA}$$

The photon cavity lifetime depends on α_t, and is given by

$$\tau_{ph} = n/(c\alpha_t) = 3.6/\left[(3 \times 10^8\ \text{m s}^{-1})(5.56 \times 10^3\ \text{m}^{-1})\right] = 2.16\ \text{ps}$$

The laser diode output power is

$$P_o = \left[\frac{hc^2\tau_{ph}(1-R)}{2en\lambda L}\right](I - I_{th}) = \frac{(6.626 \times 10^{-34})(3 \times 10^8)^2(2.16 \times 10^{-12})(1 - 0.32)}{2(1.6 \times 10^{-19})(3.6)(860 \times 10^{-9})(250 \times 10^{-6})}(I - I_{th})$$

that is,

$$P_o = (0.35\ \text{W A}^{-1})(I - I_{th}) = (0.35\ \text{mW mA}^{-1})(I - 24\ \text{mA})$$

When $I = 1.5I_{th} = 36\ \text{mA}$, $P_o = (0.35\ \text{mW mA}^{-1})(36\ \text{mA} - 24\ \text{mA}) = 4.2\ \text{mW}$.

The slope efficiency is the slope of the P_o vs. I characteristic above I_{th}, thus,

$$\eta_{slope} = \frac{\Delta P_o}{\Delta I} = \left[\frac{hc^2\tau_{ph}(1-R)}{2en\lambda L}\right] = 0.35\ \text{mW mA}^{-1}$$

We can now repeat the problem, say for $\Gamma = 0.5$, which would give $\Gamma g_{th} = \alpha_t$, so that $g_{th} = 55.6\ \text{cm}^{-1}/0.5 = 111\ \text{cm}^{-1}$. From Figure 4.48 (b), at this gain of 111 cm^{-1}, $n_{th} \approx 2.5 \times 10^{18}\ \text{cm}^{-3}$. The new radiative lifetime $\tau_r = 1/Bn_{th} = 1/\left[(2.0 \times 10^{-16}\ \text{m}^3\ \text{s}^{-1})(2.5 \times 10^{24}\ \text{m}^{-3})\right] = 2.0\ \text{ns}$.

The corresponding threshold current density is

$$J_{th} = n_{th}ed/\tau_r = (2.5 \times 10^{24}\ \text{m}^{-3})(1.6 \times 10^{-19}\ \text{C})(0.15 \times 10^{-6}\ \text{m})/(2.0 \times 10^{-9}\ \text{s}) = 30\ \text{A mm}^{-2}$$

and the corresponding threshold current I_{th} is 37.5 mA.

There are several important notes to this problem. First, the threshold concentration $n_{th} \approx 2 \times 10^{18}$ cm^{-3} was obtained graphically from Figure 4.48 (b) by using the g_{th} value we need. Second is that, at best, the calculations represent rough values since we also need to know how the mode spreads into the cladding where there is no gain but absorption and, in addition, what fraction of the current is lost to nonradiative recombination processes. We can increase α_s to account for absorption in the cladding, which would result in a higher g_{th}, larger $n_{th,}$ and greater I_{th}. If τ_{nr} is the nonradiative lifetime, we can replace τ_r by an effective recombination time τ such that $\tau^{-1} = \tau_r^{-1} + \tau_{nr}^{-1}$, which means that the threshold current will again be larger. We would also need to reduce the optical output power since some of the injected electrons are now used in nonradiative transitions. Third is the low slope efficiency compared with commercial LDs. η_{slope} depends on τ_{ph}, the photon cavity lifetime, which can be greatly improved by using better reflectors at the cavity ends, *e.g.,* by using thin film coating on the crystal facets to increase R.

4.14 SINGLE FREQUENCY SEMICONDUCTOR LASERS

A. Distributed Bragg Reflector LDs

Ideally, the output spectrum from a laser device should be as narrow as possible, which generally means that we have to allow only a single mode to exist in the cavity. There are a number of device structures that operate with an output spectrum that has high modal purity and hence a very narrow spectral width. Such LDs are often called **single frequency lasers**, even though it would be, in principle, impossible to get a perfect single frequency. (Obviously, a single-mode operation is implied.)

One method of ensuring a single mode of radiation in the laser cavity is to use frequency selective dielectric mirrors at the cleaved surfaces of the semiconductor. The **distributed Bragg reflector** (**DBR**), as shown in Figure 4.49 (a), is a mirror that has been designed like a reflection-type diffraction grating; it has a periodic corrugated structure at the end of the active region. Intuitively, partial reflections of waves from the corrugations interfere constructively (*i.e.*, they reinforce each other) to give a reflected wave only when the wavelength inside the medium corresponds to twice the corrugation period as illustrated in Figure 4.49 (b). For simplicity, the corrugations have been approximated by step changes in the refractive index from n_1 to n_2 to n_1 and so on. For example, two partially reflected waves at two n_1/n_2 interfaces in Figure 4.49 (b) such as A and B have an optical path difference of 2Λ, where Λ is the corrugation period. They can only interfere constructively if 2Λ is a multiple of the wavelength within the medium.

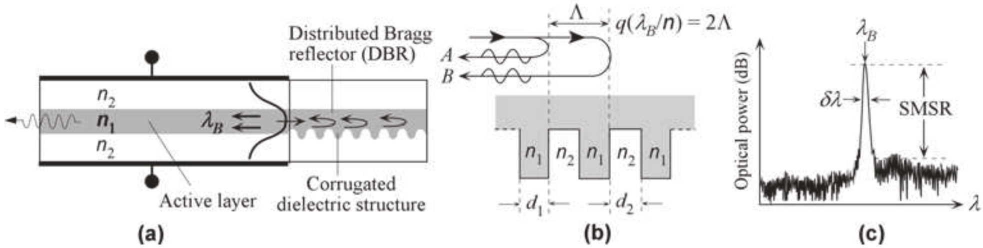


FIGURE 4.49 (a) The basic principle of the distributed Bragg reflection (DBR) laser. (b) Partially reflected waves at the corrugations can constitute a reflected wave only when the wavelength satisfies the Bragg condition. Reflected waves A and B interfere constructively when $q(\lambda_B/n) = 2\Lambda$. (c) Typical output spectrum. SMSR is the side mode suppression ratio.

Each of these wavelengths is called a **Bragg wavelength** λ_B and given by the familiar condition for in-phase interference[33]

$$q\frac{\lambda_B}{n} = 2\Lambda \tag{4.14.1}$$

Bragg wavelength

where n is the *effective* refractive index of the corrugated medium (explained below) and $q = 1, 2, \ldots$ is an integer called the **diffraction order**. The DBR therefore has a high reflectance around λ_B but low reflectance away from λ_B. The result is that only the particular optical cavity mode that is closest to the Bragg wavelength in Eq. (4.14.1), within the optical gain curve, can lase and exist in the output. The refractive index n in Eq. (4.14.1) must be viewed as an *effective* or *average refractive index* of the corrugated medium. The waves *A* and *B* in Figure 4.49 (b) involve reflections from n_1/n_2 interfaces, but there are also reflections from n_2/n_1 interfaces. If we do a proper interference accounting of all the reflections from the DBR under normal incidence, then the result would be the familiar diffraction condition in Eq. (4.14.1), but with n representing an effective or average index given by $n_{\text{av}} = (n_1d_1 + n_2d_2)/\Lambda$, where d_1 and d_2 are the thicknesses of the n_1 and n_2 regions in the grating. The effective index n_{av} would be different if the refractive index variation profile is different than that shown. (Exact calculations of the mode frequency is beyond the scope of this book.) Moreover, while the simplest DBR laser shown in Figure 4.49 (a) has a grating only at one end of the cavity, it is possible to have a grating at both ends to enhance the spectral selectivity and radiation intensity in the cavity; the latter would result in a lower threshold current. The output spectrum from a typical DBR laser is a very narrow mode centered at (or very close to) λ_B as indicated in Figure 4.49 (c). The spectral width, depending on the design, can be less than 1 MHz. Table 4.5 summarizes some of the important properties of a commercial DBR LD. There are various modifications and improvements to the basic DBR design illustrated in Figure 4.49 (a) that can be found in more advanced texts.

There is one important quantity, called the **side mode suppression ratio** (SMSR), that quantifies the intensity of the single-mode peak emission with respect to that of suppressed modes, which are all away from λ_B as illustrated in Figure 4.49 (d). SMSR is normally measured in dB, and typically reported values are usually 45 dB, that is, a suppression ratio of 3×10^4.

B. Distributed Feedback LDs

In a simple Fabry–Perot cavity-type laser, the crystal facets provide the necessary optical feedback into the cavity to build up the photon concentration. In the **distributed feedback (DFB)** laser, as shown in Figure 4.50 (a), there is a corrugated layer, called the **guiding layer**, next to the active layer; radiation spreads from the active layer to the guiding layer. The mode field diameter of the radiation in the active layer is such that it covers the guiding layer. (Remember that the thickness *d* of the active layer is small.) These corrugations in the refractive index act as optical feedback over the length of the cavity by producing partial reflections. Thus optical feedback is *distributed* over the cavity length. Intuitively, we might infer that only those Bragg wavelengths λ_B related to the corrugation periodicity Λ as in Eq. (4.14.1) can interfere constructively and thereby exist in the "cavity" in a similar fashion to that illustrated in Figure 4.49 (b).

[33] We saw in Chapter 1 that the reflectance from alternating layers of high and low refractive index structures (as in a dielectric mirror) is maximum when the free-space wavelength is λ_B, where λ_B is given by Eq. (4.14.1) in which n is an effective (average) index given by $n_{\text{av}} = (n_1d_1 + n_2d_2)/\Lambda$ if the modulation consists of periodic step changes in the refractive index from n_1 to n_2 to n_1, *etc.*

TABLE 4.5 Selected properties of DBR, DFB, and external cavity (EC) laser diodes

LD	λ_o (nm)	$\delta\upsilon$, $\delta\lambda$	SMSR (dB)	P_o (mW)	I (mA)	η_{slope} (mA)	Comment
DBR[a]	1063	2 MHz, 8 fm	45	80	200	0.8	GaAs DBR LD for spectroscopy and metrology, includes monitor current, TEC, and thermistor.
DFB[b]	1063	2 MHz, 8 fm	45	80	190	0.2	GaAs DFB LD for spectroscopy and metrology, includes monitor current, TEC, and thermistor.
DFB[c]	1550	10 MHz, 0.08 pm	45	40	300	0.3	Pigtailed to a fiber, includes monitor current, TEC, and thermistor. CW output for external modulation. For use in long-haul DWDM.
DFB[d]	1653	0.1 nm	35	5	30	0.23	Pigtailed to a single-mode fiber, includes monitor current, TEC, and thermistor. Mainly for fiber optic sensing.
EC[e]	1550	50 kHz, 0.4 fm	45	40	300	0.2	Pigtailed. Tunable over $\Delta\upsilon = 3$ GHz. Mainly for communications.

Notes: λ_o is the peak emission wavelength (λ_o for the DFB LD); I is typical operating current; fm is 10^{-15} s; $\delta\upsilon$ and $\delta\lambda$ are spectral widths; SMSR is the side mode suppression ratio; TEC is thermoelectric cooler. [a]Eagleyard, EYP-DBR-1080-00080-2000-TOC03-0000; [b]Eagleyard, EYP-DFB-1083-00080-1500-TOC03-0000; [c]Furukawa-Fitel, FOL15DCWD; [d]InPhenix, IPDFD1602; [e]Covega SFL1550S, marketed by Thorlabs.

However, the operating principle of the DFB laser is totally different. Radiation is fed from the active into the guiding layer along the whole cavity length, so that the corrugated medium can be thought of as possessing an optical gain. Partially reflected waves experience gain, and we cannot simply add these without considering the optical gain and also possible phase changes [Eq. (4.14.1) assumes normal incidence and ignores any phase change on reflection]. A left-traveling wave in the guiding layer experiences partial reflections, and these reflected waves are optically amplified by the medium to constitute a right-going wave.

In a Fabry–Perot cavity, a wave that is traveling toward the right becomes reflected to travel toward the left. At any point in the cavity, as a resulting of end-reflections, we therefore have these right- and left-traveling waves interfering, or being "coupled." These oppositely traveling waves, assuming equal amplitudes, can only set up a standing wave, a mode, if they are *coherently coupled*, which requires that the round-trip phase change is 2π. In the DFB structure,

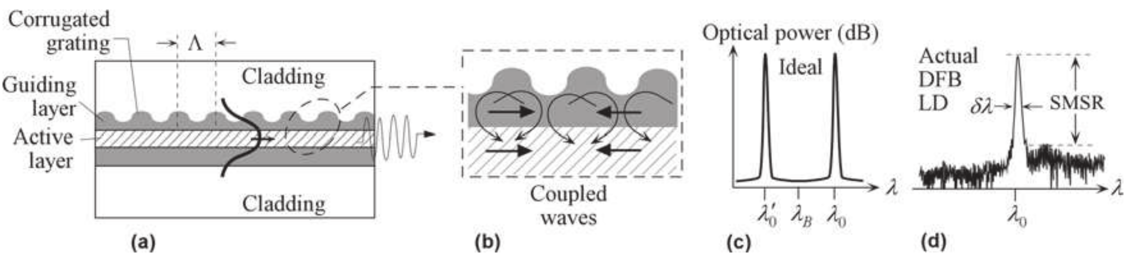


FIGURE 4.50 (a) Distributed feedback laser structure. The mode field diameter is normally larger than the active layer thickness and the radiation spreads into the guiding layer. (b) There are left and right propagating waves, partial reflections from the corrugation, and optical amplification within the cavity, which has both the active layer and the guiding layer. (c) Ideal lasing emission output has two primary peaks above and below λ_B. (d) Typical output spectrum from a DFB laser has a single narrow peak with a $\delta\lambda$ typically very narrow and much less than 0.1 nm.

traveling waves are reflected partially and *periodically* as they propagate in the cavity as illustrated in Figure 4.50 (b). The left- and right-traveling waves can only coherently couple to set up a mode if their frequency is related to the spatial corrugation period Λ, taking into account that the medium alters the wave-amplitudes via optical gain.[34] The allowed DFB modes are not exactly at Bragg wavelengths but are symmetrically placed about λ_B. If λ_m is an allowed DFB lasing mode then

$$\lambda_m = \lambda_B \pm \frac{\lambda_B^2}{2nL}(m + 1) \qquad \textbf{(4.14.2)}$$

DFB laser wavelengths

where m is a mode integer, 0, 1, 2,…, and L is the effective length of the diffraction grating, corrugation length (same as the cavity length in this very simple example). The index n in Eq. (4.14.2) is again the *effective* refractive index. The relative threshold gain for higher modes is so large that only the $m = 0$ mode effectively lases. A perfectly symmetric device has two equally spaced modes, at λ_0 and λ_0', placed around λ_B as in Figure 4.50 (c), corresponding to $m = 0$. In reality, either inevitable asymmetry introduced by the fabrication process, or asymmetry introduced on purpose, leads to only one of the modes to appear as shown in Figure 4.50 (d). Further, typically the corrugation length L is so much larger than the period Λ that the second term in Eq. (4.14.2) is very small and the lasing wavelength is very close to λ_B. Commercial DFB lasers also have thin optical coatings at the ends of the crystal facets to modify the reflectance, *i.e.,* the transmittance. For example, the facet for the radiation output would have an antireflection coating to encourage the transmission of the lasing cavity mode through this facet, and the other facet would have an optical coating to increase the reflectance.

There are numerous commercially available single-mode DFB lasers in the market for various applications such as telecom, instrumentation, metrology, interferometry, and spectroscopy. Typically the spectral widths of telecom DFB LDs are less than 0.1 nm at the communications channel of 1.55 μm. There are high-end DFB LDs that have exceptionally narrow spectral widths, a few MHz (corresponding to a few femtometers of linewidth $\delta\lambda$ in wavelength), and can be used in various spectroscopic applications where spectral purity is important. Table 4.5 highlights the properties of a few commercial DFB lasers.

One technical drawback of using DFB LDs in instrumentation is the shift of the peak wavelength λ_0 with small changes in the temperature, that is, $d\lambda_0/dT$; both Λ and n change with temperature. The increase in λ_0 is ~0.1 nm °C^{-1}. Most DFB LDs come with thermoelectric coolers (TEC) that control the temperature and limit $d\lambda_0/dT$ below certain acceptable limits. Some DFB diode manufacturers, in fact, use the $d\lambda_0/dT$ shift as a means to *tune* the laser output over a range of wavelengths, usually a few nanometers; obviously the temperature has to be very closely controlled.

When the DFB LD is modulated by its current, as in optical communications, the frequency of the optical field in the output optical pulse should remain the same at the operating frequency ω_0 throughout the optical pulse duration. However, in practice, the light pulse exhibits what is called a **frequency chirp**, that is, the frequency of the optical field changes over the time duration of the optical pulse. The reason is that the refractive index n depends on the gain of the active medium, which depends on the electron concentration, n. But, n depends on the current

[34] These partially reflected waves travel in a medium that has the refractive index modulated periodically (with periodicity Λ). We have to consider how left- and right-propagating waves in this corrugated structure are coupled. The rigorous theory is beyond the scope of this book and is called "Coupled Mode Theory."

(increases n) and coherent radiation intensity (decreases n through stimulated emissions).[35] Any change in n causes a phase change, say $\Delta\phi$, in the emitted radiation. The rate of change of this phase change during modulation, *i.e.,* $d\Delta\phi/dt$, is equivalent to an instantaneous frequency shift $\Delta\omega$, which is the **frequency chirp** defined above. Further, the increase in the current during modulation would also result in some additional Joule heating of the semiconductor. However, the TEC is not sufficiently fast to maintain the temperature constant, and hence prevent the shift in ω_0. The narrow spectral widths quoted in MHz (or less) normally refer to *instantaneous* spectral widths inasmuch as fluctuations during operation will cause small shifts in λ_0, and these shifts would appear as though $\delta\lambda$ is broader. Some manufacturers use an effective $\delta\lambda$ to account for the latter effect. The output spectrum would appear as though broadened, which would not be the same spectral width under stabilized CW operation.

C. External Cavity LDs

The main problem with the DFB laser is that the output radiation suffers from frequency chirping. It would be highly desirable to have LDs that are immune to frequency chirping and also possess stability against any drift. By having the optical cavity external to the laser diode, in so-called **external cavity diode lasers** (ECDL), these problems can be overcome, though at the expense of size. In addition, one can increase the photon cavity lifetime (τ_{ph}) and narrow the spectral width further. The basic idea involved in ECDLs is quite simple. The light from one of the facets of the LD is coupled into an external optical resonator, and a wavelength-selective component, such as a diffraction grating or an interference filter, is used within the external cavity to select the wavelength of interest (λ_o), within the gain bandwidth of the LD.

Figure 4.51 is a schematic diagram of one of the many types of ECDL that are available commercially. The radiation from the left facet B of the LD is collected, expanded, and collimated by the lens L_2. The collimated beam passes through an angled interference filter (IF).

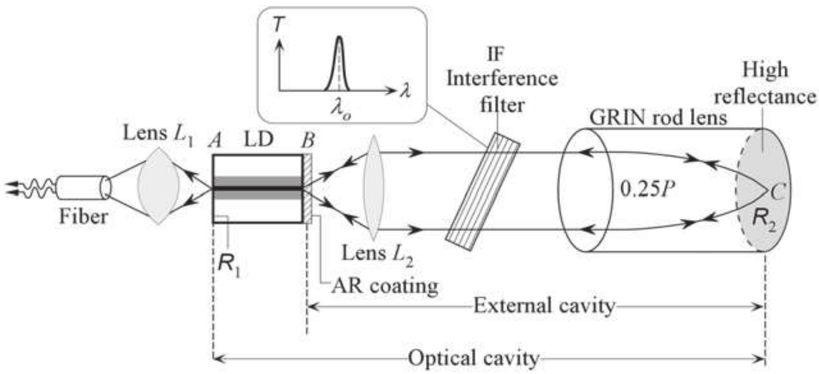


FIGURE 4.51 A simplified diagram of an external cavity diode laser (ECDL), which uses an angled interference filter (IF) to select the wavelength λ_o (depends on the tilt angle of the IF), and the optical cavity has a GRIN rod lens with one end coated for full reflection back to the LD. The output is taken from the left facet of the LD.

[35] It might be thought that the electron concentration should be clamped at n_{th}, but this is only true under CW operation. We need to solve the full rate equations to obtain the transient response, which is normally done in advanced textbooks, *e.g.,* see A. Yariv and P. Yeh, *Photonics*, 6th Edition (Oxford University Press, 2007), Ch. 15.

A commercial external cavity diode laser using the basic principle shown in Figure 4.51 (U.S. Patent 6,556,599, Bookham Technology). The output is a single mode at 785 nm ($\pm$ 1.5 pm) with a linewidth less than 200 kHz, and coupled into a fiber. The output power is 35 mW, and the SMSR is 50 dB. *(ECDL, SWL-7513-P. Courtesy of Newport, USA.)*

The latter has excellent transmittance T at one particular wavelength λ_o, within the gain bandwidth of the LD, with a very narrow spectral width $\delta\lambda$. It is somewhat similar to a dielectric mirror, but the alternating dielectric layers are chosen in such a way to transmit at one wavelength rather than reflect. The transmitted wavelength λ_o depends on the angle of the IF to the incident radiation. Further, IFs need the incident light to be collimated for the full utilization of their properties (such as wavelength selectivity and a narrow $\delta\lambda$). The radiation that passes the IF is at λ_o and enters a GRIN rod lens (a graded index glass rod that acts like a lens) whose end C has been coated to obtain high reflectance, R_2. The light reflected from the right face C of the GRIN rod travels back, passes through the IF again, and is fed back into the LD by the lens L_2. The outside surface of the LD's facet B is coated with an antireflection (AR) coating to prevent any of this light being reflected; we need to couple the light from the external cavity back into the LD. The light then propagates along the active region of the LD, experiences optical gain, and becomes reflected at the left facet A of the LD (reflectance R_1) to complete a round trip. Some of the light, $(1 - R_1)$ fraction, escaping from the left end A is collected by the lens L_1, and fed into a fiber, or collimated as an output beam. The true optical cavity is actually between A and C. The external cavity is between B and C, and much longer than the LD's cavity AB. The AR coating on the LD's facet B suppresses the role of LD's cavity AB. Since the optical path AC is much longer than the path AB, perturbations to the optical path due, for example, to changes in the refractive index in the active region along AB, do not cause shifts in the wavelength. Properly designed ECDLs do not suffer from frequency chirping, which is their advantage. Further, by using a highly wavelength-selective element in the external cavity, such as an interference filter or a suitable diffracting grating, the output radiation can be made to have very narrow spectral width $\delta\upsilon$, *e.g.,* less than 200 kHz for the ECDL in Figure 4.51.

Many ECDLs are **tunable lasers** over a range of wavelengths; that is, the center emission wavelength λ_o can be varied over the gain bandwidth of the LD. Tuning is done most easily by changing the transmission wavelength (λ_o) through the wavelength-selective element in the external cavity. In Figure 4.51, one can change the tilt angle of the IF to tune the external cavity. There are several other types of ECDLs, most of which use a diffraction grating. If we were to replace the IF in Figure 4.51 by a tilted reflection-type diffraction grating, and eliminate the GRIN rod, we would obtain what is called a **Littrow configuration** ECDL. The EC would be between the grating and the LD's facet B. The orientation of the diffraction grating would select the output wavelength. The Littrow grating-based ECDLs are widely used since they are quite simple to construct. As shown in Table 4.5, ECDLs can have very narrow linewidths and high SMSRs.

EXAMPLE 4.14.1 DFB LD wavelength

Consider a DFB laser that has a corrugation period Λ of 0.22 μm and a grating length of 400 μm. Suppose that the effective refractive index of the medium is 3.5. Assuming a first-order grating, calculate the Bragg wavelength, the mode wavelengths, and their separation.

Solution

The Bragg wavelength from Eq. (4.14.1) is

$$\lambda_B = \frac{2\Lambda n}{q} = \frac{2(0.22\ \mu\text{m})(3.5)}{1} = 1.5400\ \mu\text{m}$$

and the symmetric mode wavelengths about λ_B are

$$\lambda_m = \lambda_B \pm \frac{\lambda_B^2}{2nL}(m+1) = 1.5400 \pm \frac{(1.5400\ \mu\text{m})^2}{2(3.5)(400\ \mu\text{m})}(0+1)$$

so that the $m = 0$ mode wavelengths are

$$\lambda_0 = 1.53915 \quad \text{or} \quad 1.54085\ \mu\text{m}$$

The two are separated by 0.0017 μm or 1.7 nm. Due to a design asymmetry, only one mode will appear in the output and for most practical purposes the mode wavelength can be taken as λ_B. *Note*: The wavelength calculation was kept to five decimal places because λ_m is very close to λ_B.

4.15 VERTICAL CAVITY SURFACE EMITTING LASERS[36]

Figure 4.52 shows the basic concept of the vertical cavity surface emitting laser (VCSEL). A VCSEL has the optical cavity axis along the direction of current flow rather than perpendicular to the current flow, as in conventional laser diodes. The active region length is very short compared with the lateral dimensions so that the radiation emerges from the "surface" of the cavity rather than from its edge. The reflectors at the ends of the cavity are **dielectric mirrors**, that is, **distributed Bragg reflectors** (DBRs) made from alternating high and low refractive index quarter-wave thick multilayers. Such dielectric mirrors provide a high degree of wavelength-selective reflectance at the required free-space wavelength λ. If the thicknesses of alternating layers are d_1 and d_2 with refractive indices n_1 and n_2, then the condition

$$n_1 d_1 + n_2 d_2 = \frac{1}{2}\lambda \tag{4.15.1}$$

leads to the constructive interference of all partially reflected waves at the interfaces, and the alternating layers function as a dielectric mirror (as explained in Chapter 1). Since the wave is reflected because of a periodic variation in the refractive index, as in a grating, the dielectric mirror is essentially a **distributed Bragg reflector**. Equation (4.15.1) is equivalent to the Eq. (4.14.1) for the distributed Bragg reflector in which the effective index n is $(n_1 d_1 + n_2 d_2)/\Lambda$.

[36] The VCSEL was conceived by Kenichi Iga (Tokyo Institute of Technology) in 1977, and proposed at a conference in 1978 with the first device implementation in 1979. (See K. Iga, *Jpn. J. Appl. Phys.*, *47*, 1, 2008.)

This VCSEL diode provides a single transverse mode emission 795 nm. The spectral width is less than 100 MHz, and the output power is 0.15 mW at 2 mA. *(Courtesy of Vixar Inc.)*

Sketch of the VCSEL in Kenichi Iga's laboratory book (1997). Professor Iga was at the Tokyo Institute of Technology at the time. *(See K. Iga,* Jpn J. Appl. Phys., 47, *1, 2008; Courtesy of Kenichi Iga.)*

Kenichi Iga, currently (2012) the President of the Tokyo Institute of Technology, was first to conceive the VCSEL, and played a pioneering role in the development of VCSELs. *(Courtesy of Kenichi Iga.)*

The wavelength in Eq. (4.15.1), that is, the Bragg wavelength λ_B, is chosen to lie within the optical gain of the active layer. High reflectance end mirrors are needed because the short cavity length L reduces the optical gain of the active layer inasmuch as the optical gain is proportional to $\exp(gL)$, where g is the optical gain coefficient. There may be 30–60 or so pairs of layers in the dielectric mirrors to obtain the required reflectances; roughly ~99% for the bottom and ~95% for the top

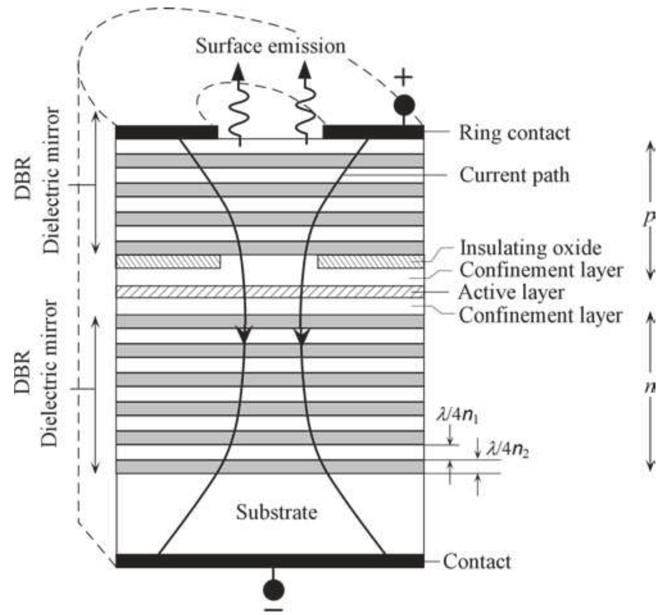
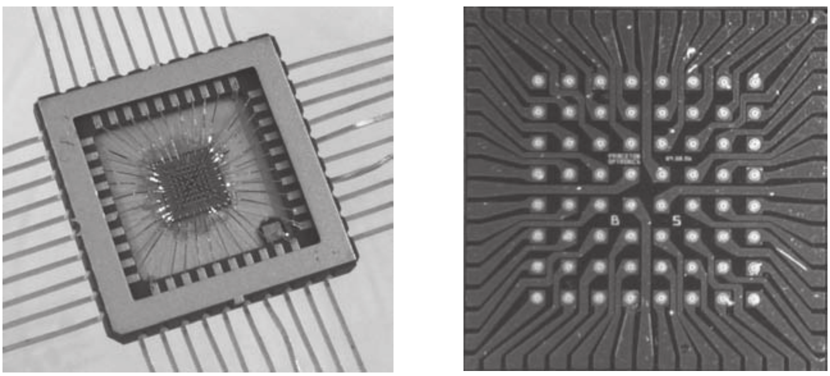


FIGURE 4.52 A simplified schematic illustration of a vertical cavity surface emitting laser. The cross-section is circular.

emitting side. The whole optical cavity in Figure 4.52 looks "vertical" if we keep the current flow the same as in a conventional laser diode cavity.

The active layer is generally very thin (<0.1 μm) and is likely to be a multiple quantum well for reduced threshold current. The active layer is sandwiched between two confinement layers. There is a thin insulating oxide layer that covers the region outside the central part, and hence confines the current flow to the central region to increase the current density flowing into the active region, and lower the threshold current for lasing. The required semiconductor layers are grown by epitaxial growth on a suitable substrate. The emission can be through the top dielectric mirror, as in Figure 4.52, or, as in some VCSELs, through the GaAs substrate. The latter are called **substrate** or **bottom emitting** VCSELs. For example, a 980-nm emitting VCSEL device has InGaAs as the active layer to provide the 980-nm emission, and a GaAs crystal is used as a substrate which is transparent at 980 nm. The dielectric mirrors are then alternating layers of AlGaAs with different compositions and hence different bandgaps and refractive indices. In practice, the current flowing through the dielectric mirrors give rise to an undesirable voltage drop and methods are used to feed the current into the active region more directly, for example, by depositing "peripheral" contacts close to the active region. There are various sophisticated VCSEL structures, and Figure 4.52 is only one very simplified example.

The vertical cavity is generally circular in its cross-section so that the emitted beam has a **circular cross-section**, which is a distinct advantage. The height of the vertical cavity may be as small as several microns. Therefore, the longitudinal mode separation is sufficiently large to allow only one longitudinal mode to operate. However, there may be one or more lateral or transverse modes depending on the lateral size of the cavity. In practice there is only one single transverse mode (STM) (and hence one mode) in the output spectrum for cavity diameters less than ~10 μm. The spectral line width from a VCSEL that has single transverse mode is typically much less than 0.1 nm. Specially designed VCSELs can have spectral widths ($\Delta\upsilon$) as low as 100 MHz, that is, a linewidth ($\Delta\lambda$) of 0.2 pm. Many VCSELs in the market have several lateral modes, and the spectral width is about ~0.5 nm, significantly less than a conventional



Left: A packaged addressable VCSEL array with 8 × 8 individually addressable laser devices. The chip is 3 mm × 3 mm. Right: A closer view of the chip. *(Courtesy of Princeton Optronics, USA.)*

longitudinal multimode Fabry–Perot (FP) laser diode. There is, however, a shift in the emission wavelength with temperature as in the case of DFB and DBR LDs. For VCSELs, the wavelength increases by about 0.05 nm per 1°C increase, which is slightly better than typical values quoted for DBR LDs (0.06−0.1 nm °C^{-1}). Table 4.6 summarizes some of the properties of selected single and multi-transverse mode VCSELs. Compared to conventional FP and DBR LDs, two features should be immediately apparent. First is the lower threshold current; and hence a lower overall operating current. Second is the lower optical power emitted. However, we should compare emitted intensities, power emitted per unit area, which is ~ 1 kW cm^{-2}, and higher than edge emitting FP LDs.

With cavity dimensions in the microns range, such a VCSEL laser is referred to as a **microlaser**. One of the most significant advantages of microlasers is that they can be arrayed to construct a **matrix emitter** that is a broad area surface emitting laser source. Such laser arrays have important potential applications in optical interconnect and optical computing technologies. Further, such laser arrays can provide a higher optical power than that available from a single conventional laser diode. Powers reaching a few watts are easily obtained from such matrix lasers.

VCSELs are used in many diverse applications such as short-haul optical communications (local area networks, data communications up to 10 Gb s^{-1}), sensors, encoders, bar code readers, printing, optical mice, spectroscopy, and smart cables. Graded index fibers have now widely replaced multimode step index fibers in short-haul communications. These graded index fibers need narrow linewidth emitters to avoid the material dispersion (which is significant at 850 nm) overwhelming the intrinsic (and optimized) intermode dispersion of the fiber[37]; VCSELs are highly suitable for such use in gigabit Ethernet (*e.g.*, 2.5−10 Gb s^{-1}). When many VCSELs are integrated onto a single chip as arrays, the emission power from the chip can be quite large. There are commercial array chips, and modules with several chips, that can easily emit 100 W or more. In addition, there are chips with VCSEL arrays that are addressable so that each VCSEL diode on the array can be externally controlled.

VCSELs offer a number of distinct advantages in terms of much easier manufacturability (with high yield and volume), testing, packaging, coupling into optical fibers, and

TABLE 4.6 Selected characteristics of a few commercial VCSELs

λ (nm)	$\Delta\lambda$ (nm)	$d\lambda/dT$ (nm K^{-1})	$\Delta\theta$	I_{th} (mA)	I (mA)	V (V)	P_o (mW)	η_{slope} (mW mA^{-1})	VCSEL
680	<0.1	0.046	21°	0.4	3	2.8	0.95	0.32	STM[a]
680	0.5	0.04	15°	2.5	8	2.3	4	0.7	MTM[a]
850	0.0002	0.06	17°	0.5	2	2.2	0.75	0.5	STM[b]
850	<0.65	0.06	26°	1	6	1.9	2	0.4	MTM[c]

Notes: STM and MTM refer to single transverse mode and multi-transverse mode devices. [a]Vixar, 680-000-x001, single- and multimode devices; [b]Laser Components (ULM Photonics) ULM775-03-TN; [c]Oclaro, 850 nm 8.5 Gb/s Multimode VCSEL Chip. *I* and *V* are typical operating current and voltage values.

[37]The material dispersion is quite significant at 850 nm, so that, unless a very narrow linewidth light source is used (*e.g.*, a VCSEL), the material dispersion would be greater than the intermodal dispersion characteristic of the graded index fiber.

the implementation of 2D arrays, which can be addressable as well. Advances on VCSELs are likely to make this laser diode technology one of the leaders in photonics over the next decade.[38]

4.16 SEMICONDUCTOR OPTICAL AMPLIFIERS

A semiconductor laser structure can also be used as an optical amplifier that amplifies light waves passing through its active region as illustrated in Figure 4.53 (a). The wavelength of radiation to be amplified must fall within the optical gain bandwidth of the laser. Such a device would not be a laser oscillator, emitting lasing emission without an input, but an optical amplifier with input and output ports for light entry and exit. In the **traveling wave (TW) semiconductor optical amplifier (SOA)**, the ends of the optical cavity have antireflection (AR) coatings so that the optical cavity does not act as an efficient optical resonator, a condition for laser oscillations. Laser oscillations are therefore prevented. Light, for example, from an optical fiber, is coupled into the active region of the laser structure. As the radiation propagates through the active layer, optically guided by this layer, it becomes amplified by the induced stimulated emissions, and leaves the optical cavity with a higher intensity. Obviously the device must be pumped to achieve optical gain (population inversion) in the active layer. Random spontaneous emissions in the active layer feed "noise" into the signal and broaden the spectral width of the passing radiation. This can be overcome by using an optical filter at the output to allow the original light wavelength to pass through. Typically, such laser amplifiers are buried heterostructure devices and have optical gains of ~20 dB depending on the efficiency of the AR coating.

The **Fabry–Perot semiconductor optical amplifier**, as shown in Figure 4.53 (b), is similar to the conventional laser oscillator, but is operated below the threshold current for lasing oscillations. The active region has an optical gain but not sufficient to sustain a self-lasing output, *i.e.,* the gain coefficient g is below the threshold gain g_{th}. Light passing through such an active region will be amplified by stimulated emissions but, because of the presence of an optical resonator, there will be internal multiple reflections as indicated in Figure 4.53 (b). These multiple reflections lead to the gain being highest at the resonant frequencies υ_m of the cavity within the optical gain bandwidth of the active region material. Optical frequencies around the cavity resonant

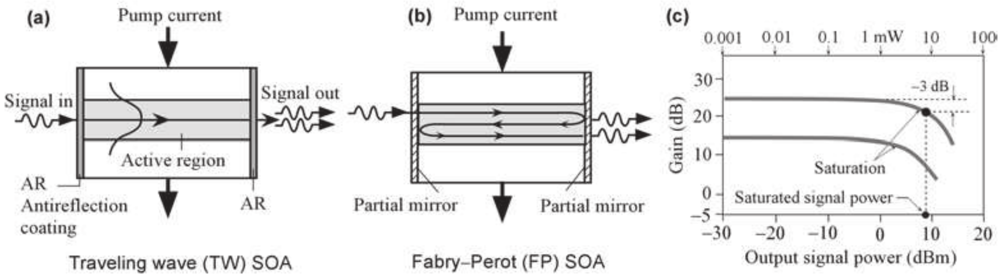


FIGURE 4.53 (a) Traveling wave (TW). (b) Fabry–Perot semiconductor optical amplifier (SOA). (c) Gain vs. output power for a 1500-nm TW SOA (GaInAsP). SOAs exhibit saturation at high output powers; saturation lowers the gain. Higher gain (25 dB) is achieved by passing a higher current through the SOA. (*Source:* Selected data used from T. Saitoh and T. Mukai, *IEEE J. Quantum Electron., QE23,* 1010, 1987.)

[38] See, for example, the excellent review by J. S. Harris *et al.*, *Semicond. Sci. Technol.*, *26*, 1, 2011.

frequencies will experience higher gain than those away from resonant frequencies. The FP SOA is basically a **regenerative amplifier** in which positive feedback (reflections back into the cavity from the end mirrors) has been used to boost the overall gain. Although the FP amplifier can have a higher gain than the TW amplifier, due to the positive feedback, it is less stable; if the gain is not properly controlled, it can oscillate—lase.

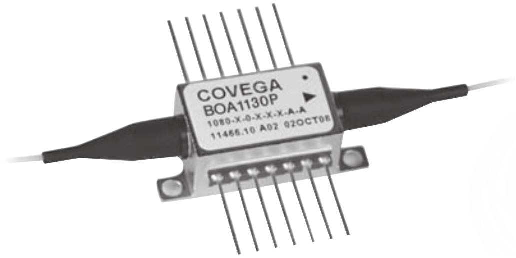
At high output signal power, the gain exhibits saturation, just as in EDFA, which is shown in Figure 4.53 (c). The reason is that at high output powers, there is a high concentration of photons in the active region, which increases the stimulated recombination rate. The increase in rate of stimulated transitions decreases the population inversion, and hence reduces gain. SOA specifications normally also quote values for the saturation output signal power $P_{o\text{sat}}$ at which the gain would have dropped by 3 dB as illustrated in Figure 4.53 (c). For a given SOA, $P_{o\text{sat}}$ depends on the gain, being lower for smaller gain, because a lower gain implies a smaller population inversion, which can be depleted more easily.

The gain of SOAs also exhibit some polarization dependence, *i.e.,* dependence on the direction of the optical field. We can view the active layer of the SOA as a planar dielectric waveguide, which implies that there will be two types of traveling waves, TE and TM polarized. These will be amplified by different amounts along the active layer. **Polarization-dependent gain**, ΔG_p, is defined as the difference between maximum and minimum gain for different orientations of the optical field (TE and TM waves), and is quoted in dB. Table 4.7 summarizes some of the properties of a few selected TW SOAs. Most SOAs come coupled (pigtailed) to a fiber.

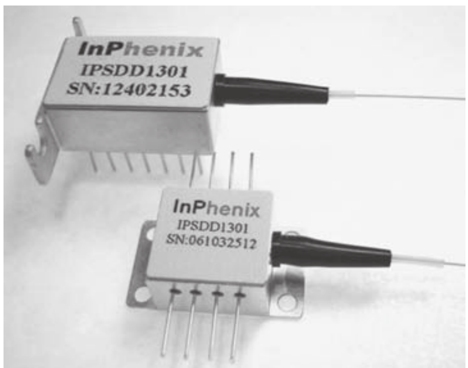
TABLE 4.7 Various properties of three selected TW SOAs

λ (nm)	Gain (dB)	BW (nm)	I_F (mA)	V_F (V)	$P_{o\text{sat}}$ (dBm)	NF (dB)	ΔG_p (dB)	Comment
850	25	40	150	2.2	8		7	Pigtailed (fiber coupled)[a]
1310	22	45	250	~2	10	7	0.5	Pigtailed, in-line amplifier[b]
1550	20	35	200	~2	11	6	0.5	Pigtailed preamplifier[c]

Notes: BW is the bandwidth of the amplifier, wavelengths over which it amplifies the optical signal; I_F and V_F are the forward current and voltage under normal operation with the given gain; ΔG_p is the polarization-dependent gain. [a]Superlum Diodes, SOA-372-850-SM; [b]InPhenix, IPSAD1301; [c]Amphotonix (Kamelian) C-band, preamp.



Covega semiconductor optical amplifier (SOA) for use as a booster amplifier in the O-band (around 1285 nm). The small signal gain is 27 dB and the NF is 7 dB. *(Courtesy of Thorlabs.)*



Two superluminescent light-emitting diodes; each has been pigtailed to a fiber for operation at 1310 nm. Their spectral width is 40–45 nm. *(Courtesy of InPhenix.)*

Additional Topics

4.17 SUPERLUMINESCENT AND RESONANT CAVITY LEDs: SLD AND RCLED

Superluminescent light-emitting diodes, simply called **superluminescent diodes**[39] (SLDs) are very bright LEDs that have an active region with gain to amplify the spontaneously emitted photons. The basic principle is the amplification of spontaneous emissions (ASE). They are not lasers because there are no lasing oscillations, and the light output is just amplified spontaneous emission, *i.e.,* incoherent photons. They operate in a domain that is between LEDs and lasers in the sense that they are based on spontaneous emission, but then the diode structure has an optical cavity with gain to amplify the spontaneously emitted light. A typical stripe geometry SLD structure is shown in Figure 4.54 (a). The structure is based on heterostructures, as in the stripe geometry LD, but there is one important difference. The back-end, shown as L_a, of the SLD shown in Figure 4.54 (a) is not pumped. The top electrode does not cover this region. Consequently the back-end of the cavity (L_a) is lossy, and there are no reflections from the rear end. Thus, laser oscillations are suppressed, but the active region, shown as L_g, is still pumped to achieve high optical gains. Spontaneous photons that are emitted toward the front facet form forward guided waves. As the waves cross the pumped active region, L_g, they are amplified. Since the optical cavity is suppressed, we can pump the LD well above its usual threshold gain g_{th} (which no longer applies as the rear-end mirror no longer reflects the waves). The front end typically has an antireflection coating to allow the photons to escape.

Another structure for the SLD is the tilted cavity shown in Figure 4.54 (b). In normal FP LDs, the cavity has cleaved crystal facets at the ends that reflect the radiation back into the cavity. However, if the cavity is tilted, as in Figure 4.54 (b), then the reflections back into the cavity are suppressed, and lasing oscillations are avoided. Tilted cavity SLDs are also commercially available.

Since the gain inside L_g in the SLD increases with the current, the increase in the output power P_o with the current is more than that in the conventional LED, due to this extra gain. Unlike conventional LEDs, the SLD does not normally suffer from saturation effects at high currents. The second most important advantage is the high optical powers that are emitted from the SLD, which leads to higher powers being launched into a fiber. SLDs have found good use in short-haul optical communications such as local area networks. The spectral width of an SLD used in

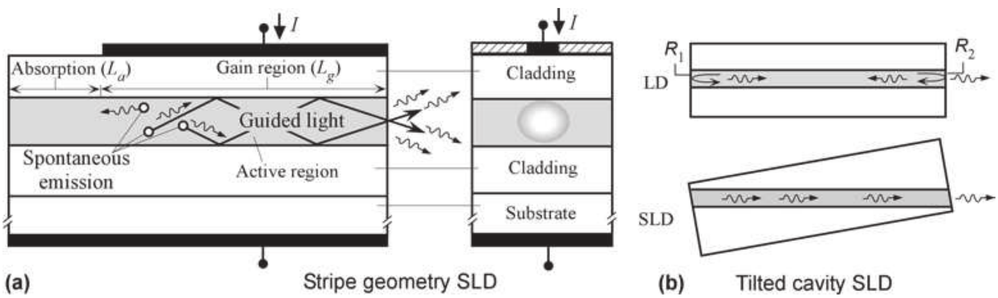


FIGURE 4.54 Superluminescent LED principles. (a) Stripe geometry SLD. (b) Tilted cavity SLD.

[39] The superluminescent LED was invented at RCA Laboratories (Sarnoff Corporation) by Gerard Alphonse in 1986.

communications tends to be narrower than that of a corresponding conventional LED. For example, typically an SLD operating at 1310 nm would have a spectral width approximately in the 35–70 nm range, whereas the conventional LED would have 60–130 nm. However, there are also many SLDs used for their high brightness with significantly wider spectral widths than corresponding LEDs at high currents, and the spectral width increases with the injection current. The output spectrum from an SLD often has small ripples, tiny spikes, corresponding to the resonant frequencies of the optical cavity, which is not totally suppressed. Manufacturers normally quote the extent of this ripple in the output spectrum, called **gain** or **Fabry–Perot ripples** and quoted in dB.

Another important LED technology that is related to lasers is the **resonant cavity LED**[40] (RCLED). RCLED is almost identical to the VCSEL, with a structure very similar to that in Figure 4.52. The distributed Bragg reflector on the upper emission side is made to have a reflectance that is 40–60% whereas the other DBR at the lower side is nearly fully reflecting (95–99%). The poor reflectance of the upper DBR means that the threshold gain is high and the device can be operated without going into lasing oscillations. The spontaneously emitted photons are amplified by the thin active region but then they are reflected back and forth and amplified many times before being emitted. It is reminiscent to the FP SOA which operates below oscillation but has regenerative gain due to reflections from the ends feeding the radiation back into the cavity. The cavity modes matter because these are the modes that are reflected back and forth and become amplified. The emitted radiation from an RCLED therefore has a spectrum that represents a cavity mode with a spectral width that is determined by the cavity losses (or the cavity finesse), primarily the lossy upper DBR. The spectral width of the emitted radiation is broader than that of the corresponding VCSEL (where both DBRs are highly reflecting), but narrower than conventional and superluminescent LEDs; typical values being a few nanometers. Since the LED operation is based on amplifying photons emitted into a mode of the resonant cavity, the device is called a **resonant cavity LED**.

4.18 DIRECT MODULATION OF LASER DIODES

The modulation of the optical output power from a laser diode is achieved by modulating the LD drive current. The modulation can be from an analog signal or digital as in the case of LEDs. We first consider small signal AC modulation as in Figure 4.55 (a). A DC current I_1 is applied first to bias the LD and obtain a DC (steady) optical output power P_{o1}. The LD is then modulated by a sinusoidal signal $\Delta I_m \sin(2\pi ft)$ superimposed on I_1 as in Figure 4.55 (a), where ΔI_m is the maximum amplitude of the signal and f is the modulation frequency. The modulation of the diode input current about I_1 gives rise to the output optical power varying about its DC value of P_{o1}, which can be represented by $\Delta P_m \sin(2\pi ft + \phi)$, as in Figure 4.55 (b), where ΔP_m is the maximum amplitude of the modulated output and ϕ is a phase difference between the input signal (modulating current) and the output signal (modulated optical power).

Under AC operation, ΔP_m, of course, depends on ΔI_m and increases with ΔI_m. However, we are interested in the frequency response, that is, $\Delta P_m(f)$ vs. the modulation frequency f when ΔI_m is kept constant. Suppose that the output signal power at very low frequencies (ideally DC) is $\Delta P_m(0)$. What is of interest is $\Delta P_m(f)/\Delta P_m(0)$ as a function of frequency f, which is shown in Figure 4.55 (c) with some typical values. As can be seen, $\Delta P_m(f)/\Delta P_m(0)$ vs. f characteristic is

[40] RCLEDs were first realized by Fred Schubert and coworkers at AT&T Bell Laboratories, Murray Hill, NY in 1992, and reported in *Appl. Phys. Letts.*, *60*, 921, 1992.

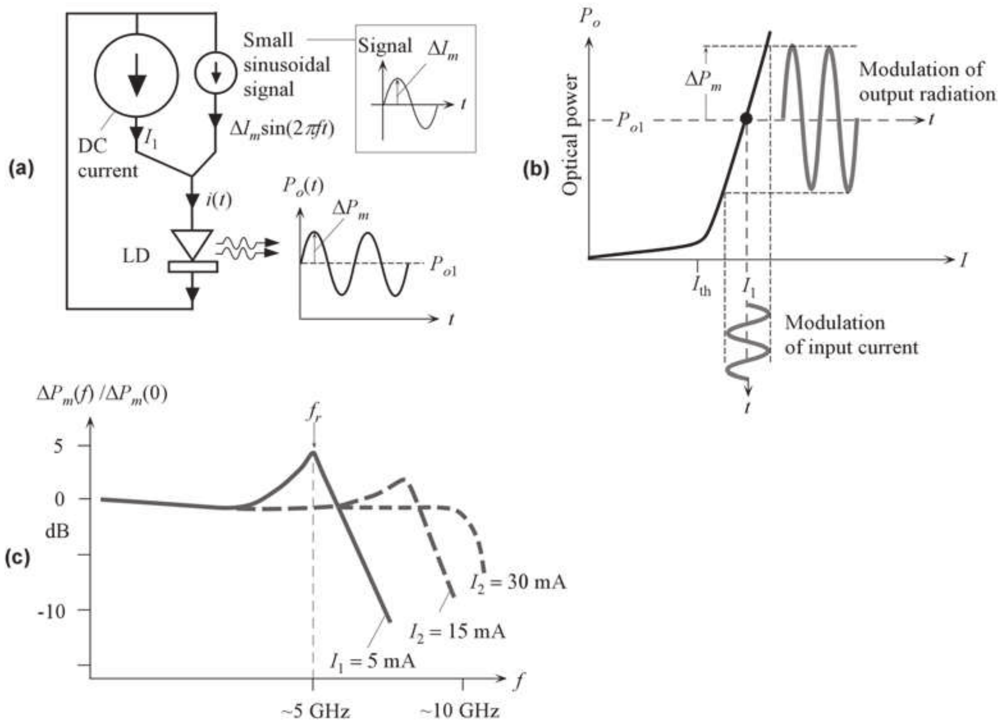


FIGURE 4.55 (a) The LD is biased by passing a DC current I_1 and a sinusoidal current modulation $\Delta I_m \sin(2\pi ft)$ is superimposed on the DC current. (b) P_o vs. I characteristics of a LD showing how a superimposed AC current on top of a DC current I_1 modulates the output optical power. The optical power output has a DC value of P_{o1} about which the power varies sinusoidally with a maximum amplitude ΔP_o. (c) The frequency response $\Delta P_m(f)/\Delta P_m(0)$ vs. frequency has a relaxation oscillation peak at a certain frequency f_r, which depends on the diode current or the output power.

relatively flat until high frequencies, and then exhibits a prominent peak at a certain frequency called the **relaxation oscillation frequency** f_r, as marked in Figure 4.55 (c). f_r depends on the LD characteristics, and is given by[41]

Relaxation oscillation frequency

$$f_r \approx \frac{1}{2\pi(\tau\tau_{ph})^{1/2}}\left(\frac{I_1}{I_{th}} - 1\right)^{1/2} \qquad \textbf{(4.18.1)}$$

where τ is the effective carrier recombination time, which represents the radiative and nonradiative processes acting in parallel; τ_{ph}, as before, is the *photon cavity lifetim*e, the average time it takes a photon to be lost from the laser diode cavity (by absorption inside the cavity and emission through the cavity facets); I_1 is the operating diode bias current; and I_{th} is the usual threshold current. If nonradiative recombination is negligible, then $\tau = \tau_r$, the radiative lifetime. The characteristic times τ and τ_{ph} are typically in the nano- and picosecond range so that for I_1, roughly twice I_{th}, f_r is a few GHz. LDs are usually modulated below f_r.

The frequency response of a LD also depends on the operating current I_1 or the DC output power P_{o1}. As I_1 increases, the bandwidth increases and the magnitude of the peak at f_r decreases, as apparent in Figure 4.55 (c).

---

[41] The derivations for the equations in this section can be found in P. Bhattacharya, *Semiconductor Optoelectronic Devices*, 2nd Edition (Prentice-Hall, New York, 1993), Ch. 7. Both Eqs. (4.18.1) and (4.18.2) are reasonable approximations.

The peak in the frequency response at f_r is typical of a system that has two energy storage elements, such as an inductor L and a capacitor C that are coupled. We can view the LD as made up of two subsystems. One subsystem is the electrical subsystem that has the current and the injected electrons, and the other, the optical subsystem, that has the optical cavity and the coherent radiation. Near f_r, the energy associated with the injected electrons in the active region in the electrical subsystem is transferred back and forth to and from the radiation (photons) within the cavity in the optical subsystem. This energy transfer is mediated by stimulated emissions because both photons and electrons are involved in stimulated transitions. The electron and photon concentrations are coupled through the stimulated emission process. The phase difference ϕ between the output and the input signals arises from the delay in getting the coherent photon output keeping up with the electron injection. There is a delay associated with the recombination process, which is represented by τ, and then there is another delay associated with getting the photons out from the cavity τ_{ph}. The radiation output lags the input current.

Consider a LD that is excited by a step current pulse that changes from I_1 ($<I_{th}$) to I_2 ($>I_{th}$) as shown in Figure 4.56 (a). Assume initially I_1 is very small. The lasing radiation output begins only after a **delay time** t_d, and reaches a steady state value after a few damped oscillations as illustrated in Figure 4.56 (b). The delay time t_d is the time it takes to inject sufficient electrons into the active region to achieve population inversion, and sufficient gain to start the lasing oscillations to build up. Suppose that initially the bias on the LD is I_1, less than I_{th}, as in Figure 4.56 (a). Thus, the diode current changes from I_1 to I_2. The delay time t_d before any coherent radiation output appears is approximately given by

$$t_d \approx \tau \ln\left[\frac{I_2}{I_2 - (I_{th} - I_1)}\right] \qquad \textbf{(4.18.2)}$$

Diode laser output delay time

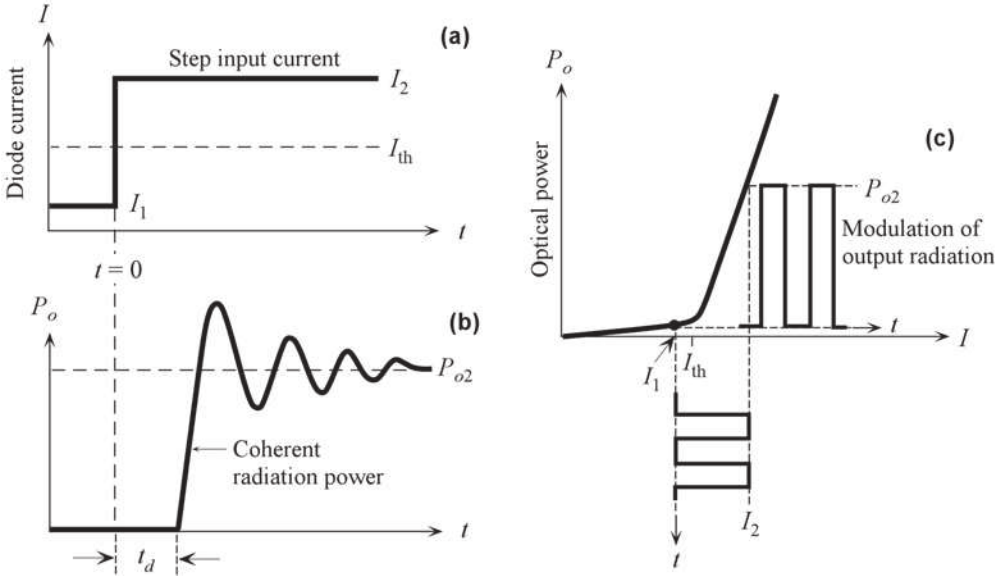


FIGURE 4.56 (a) The Diode current is applied suddenly at $t = 0$ as a step from I_1 (below I_{th}) to I_2 (above I_{th}). (b) The transient response of the output power, P_o, as a function of time. There is a delay t_d before any output appears. P_{o2} is the steady state output, after the relaxation oscillations have died out, after a few cycles. (c) The laser diode is biased at I_1 near I_{th} for faster switching. The input current is a pulse of magnitude ($I_2 - I_1$), which generates an output P_{o2}. (Ideal response in which time delays have been neglected, as would be the case under sufficiently low repetition rates.)

where τ is the overall effective recombination time in the recombination region around the threshold operating conditions and $(I_{th} - I_1)$ is the current needed to reach I_{th}. If the LD is not biased, the delay time will be longer. We can reduce the delay time by biasing the LD at I_1 just below I_{th}, as shown in Figure 4.56 (c), so that $(I_{th} - I_1)$ is negligible, then t_d is as short as possible and roughly the same as τ, the recombination time. The damped relaxation oscillations in the output in Figure 4.56 (b) have the frequency f_r, which should not be surprising, given the above explanation for f_r in terms of two energy storage subsystems in the LD. Usually, the output reaches a steady state P_{o2} after a few oscillations.

Generally, under direct modulation, LD bandwidths are wider than conventional LEDs, and LDs are the obvious choice in high-speed optical communications, given also their much narrower spectral width. There is also a disadvantage to directly modulating LDs by the diode current. The modulation processes modifies the refractive index of the gain medium, and hence causes the radiation frequency to shift, called **frequency chirping**, during the modulation period, over the time of a digital bit. The reason is that the refractive index n, as mentioned above, depends on the injected electron concentration, which is controlled by the current. The frequency chirp can be eliminated by operating the LD as a CW laser, and using a high-speed external modulator. Some electro-optic external modulators, such as Mach–Zhender modulators, as discussed in Chapter 6, have very fast switch times and can easily operate at 40 Gb s^{-1} or above.

4.19 HOLOGRAPHY

Holography is a technique of reproducing three-dimensional optical images of an object by using a highly coherent radiation from a laser source. Although holography was invented by Denis Gabor in 1948, prior to the availability of highly coherent laser beams, its practical use and popularity increased soon after the commercial development of lasers.

The principle of holography is illustrated in Figures 4.57 (a) and (b). The object, a cat, is illuminated by a highly coherent beam as normally would be available from a laser[42] as in Figure 4.57 (a). The laser beam has both *spatial* and *temporal coherence.* Part of the coherent wave is reflected from a mirror to form a **reference beam** E_{ref} and travels toward a fine-grained photographic plate. The waves that are reflected from the cat, E_{cat}, will have both amplitude and phase variations that represent the cat's surface (topology). If we were looking at the cat, our eyes would register the wavefront of the reflected waves E_{cat}. Moving our head around, we would capture different portions of the reflected wave and we would see the cat as a three-dimensional object.

The reflected waves from the cat (E_{cat}) is made to interfere with a reference wave (E_{ref}) at the photographic plate and give rise to a complicated *interference pattern* that depends on the magnitude and phase variation in E_{cat}. The recorded interference pattern in the photographic film (after processing) is called a **hologram**. It contains all the information necessary to reconstruct the wavefront E_{cat} reflected from the cat and hence produce a three-dimensional image.

To obtain the three-dimensional image, we have to illuminate the hologram with the reference beam E_{ref} alone as in Figure 4.57 (b). Most of the beam goes right through but some of it becomes *diffracted* by the interference pattern in the hologram. One diffracted beam is an exact replica of the original wavefront E_{cat} from the cat. The observer sees this wavefront as if the waves were reflected from the original cat and registers a three-dimensional image of the cat. This is the **virtual image**. We know that first-order diffraction from a grating has to satisfy the

[42] A short pulse of laser light, such as from a ruby laser, will not harm the cat; this is how holograms of people are taken. The first successful three-dimensional laser holography was carried out by Emmett Leith and Juris Upatneiks (University of Michigan) in 1960. In 1979 President Jimmy Carter awarded Emmett Leith with the National Medal of Science for his contributions to optics.

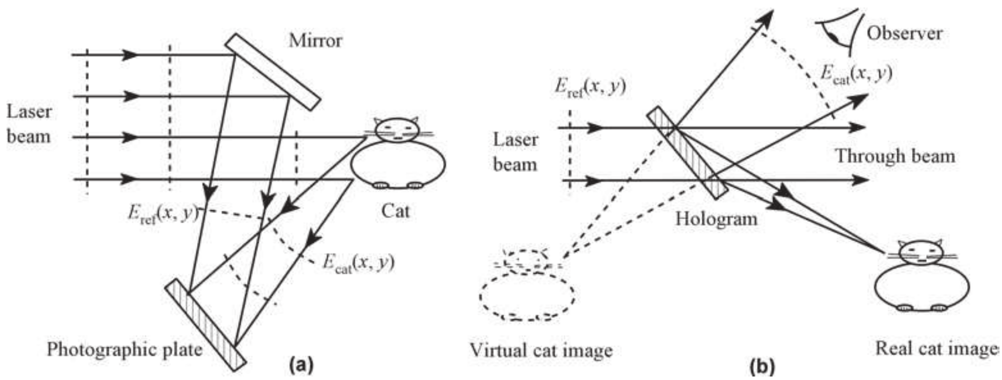


FIGURE 4.57 A highly simplified illustration of holography. (a) A laser beam is made to interfere with the diffracted beam from the subject to produce a hologram. (b) Shining the laser beam through the hologram generates a real and a virtual image.

Bragg condition, $d\sin\theta = m\lambda\,(m = \pm 1)$ where λ is the wavelength and d is the periodicity of the grating, which we took as the separation of slits in a ruled grating in Chapter 1. We can qualitatively think of a diffracted beam from one locality in the hologram as being determined by the local separation d between interference fringes in this region. Since d changes in the hologram, depending on the interference pattern produced by E_{cat}, the whole diffracted beam depends on E_{cat} and the diffracted beam wavefront is an exact scaled replica of E_{cat}. Just as normally there would be another diffracted beam on the other side of the zero-order (through) beam, there is a second image, called the **real image**, as in Figure 4.57 (b), which is of lower quality. (It may help to imagine what happens if the object consists of black and white stripes. We would then obtain periodic dark and bright interference fringes in the hologram. This periodic variation is just like a diffraction grating; the exact analysis is more complicated.)

Holography can be explained by the following highly simplified analysis. Suppose that the photographic plate is in the xy plane in Figure 4.57. Assume that the reference wave can be represented by

$$E_{ref}(x, y) = U_r(x, y)e^{j\omega t} \tag{4.19.1}$$

where $U_r(x, y)$ is its amplitude, generally a complex number that includes magnitude and phase information.

The reflected wave from the cat will have a complex amplitude that contains the magnitude and phase information of the cat's surface, that is,

$$E_{cat}(x, y) = U(x, y)e^{j\omega t} \tag{4.19.2}$$

where $U(x, y)$ is a complex number. Normally we would look at the cat and construct the image from the wavefront in Eq. (4.19.2). We thus have to reconstruct the wavefront in Eq. (4.19.2), *i.e.,* obtain $U(x, y)$.

When E_{ref} and E_{cat} interfere, the "brightness" of the photographic image depends on the intensity I and hence on

$$I(x, y) = |E_{ref} + E_{cat}|^2 = |U_r + U|^2 = (U_r + U)(U_r^* + U^*)$$

where we have used the usual definition that the square of the magnitude of a complex number is the product with its complex conjugate (indicated by an asterisk). Thus,

$$I(x, y) = UU^* + U_rU_r^* + U_r^*U + U_rU^* \tag{4.19.3}$$

This is the pattern on the photographic plate. The first and second terms are the intensities of the reflected and reference waves. It is the third and fourth terms that contain the information on the magnitude and phase of $U(x, y)$.

We now illuminate the hologram, $I(x, y)$, with the reference beam, which has an amplitude U_r as in Eq. (4.19.1), so that the transmitted light wave has a complex magnitude $U_t(x, y)$ proportional to $U_r I(x, y)$

$$U_t \propto U_r I(x, y) = U_r(UU^* + U_r U_r^* + U_r^* U + U_r U^*)$$

i.e.,

$$U_t \propto U_r(UU^* + U_r U_r^*) + (U_r U_r^*)U + U_r^2 U^* \tag{4.19.4}$$

or

$$U_t \propto a + bU(x, y) + cU^*(x, y)$$

where a, b, and c are appropriate constants. The first term $a = U_r(UU^* + U_r U_r^*)$ is the through beam. The second and third terms represent the diffracted beams.[43] Since $b = (U_r U_r^*)$ is a constant (intensity of the reference beam), the second term is a scaled version of $U(x, y)$ and represents the original wavefront amplitude from the cat, *i.e.,* includes the magnitude and phase information. We have effectively reconstructed the original wavefront and the observer will see a three-dimensional view of the cat within the angle of original recording; $bU(x, y)$ is the **virtual image**. The third term $cU^*(x, y)$ is the **real image** and notice that its amplitude is the complex conjugate of $U(x, y)$; it is called the **conjugate image**. It is apparent that holography is a method of **wavefront reconstruction**. It should be mentioned that the observer will always see a positive image whether a positive or a negative image of the hologram is used. A negative hologram simply produces a 180° phase shift in the field of the transmitted wave and the eye cannot detect this phase shift.

Dennis Gabor (1900–1979), inventor of holography, was a Hungarian-born British physicist who published his holography invention in *Nature* in 1948 while he was at Thomson-Houston Co. Ltd, at a time when coherent light from lasers was not yet available. He was subsequently a professor of applied electron physics at Imperial College, University of London. He received the Nobel Prize in Physics in 1971 for his invention of holography. His Nobel Lecture given in 1971 provides an excellent insight into the development of holography, and is accessible from the Nobel website. *(© Linh Hassel/AGE Fotostock.)*

[43] This is a simple factual statement as the mathematical treatment above has not proved this point.

Questions and Problems

4.1 Pumping of three-level laser systems The 3-level Ruby laser system is shown in Figure 4.2 (a). If transitions from E_3 to E_2 are very fast, and the spontaneous decay time from E_2 to E_1 is τ_{sp}, obtain an expression for the minimum pumping power needed (the power that must be absorbed by the laser medium) for population inversion ($N_2 > N_1$).

Consider a 60 mm long ruby crystal with 6 mm diameter, doped with 2×10^{20} Cr^{3+} ions cm^{-3}. The spontaneous decay time is 3 ms. Using the above-obtained expression, find the minimum power that must be absorbed if the excitation is to the band of energies in the blue around (~3.1 eV) in Figure 4.3 (a).

4.2 Neodymium-doped glass laser Consider a phosphate glass rod, with a refractive index $n = 1.50$, that has been doped with 3×10^{20} Nd^{3+} ions cm^{-3}. The glass rod is 10 cm long and 1 cm in diameter. We can use Figure 4.4 to represent the operation of the Nd^{3+}:glass laser. The lasing emission E_2 to E_1 is at 1054 nm. The pump wavelength is 808 nm, and takes the Nd^{3+} from E_0 to E_3. E_1 is roughly 0.26 eV above E_0. The linewidth of the output spectrum $\Delta\lambda$ is about 28 nm. E_2 to E_1 emission cross-section is 3×10^{-20} cm^2 and the spontaneous decay lifetime is 300 μs. (a) Calculate the maximum possible gain coefficient in the medium. (b) What percentage of the input energy is lost as heat? (c) What is the spectral spread (linewidth) in terms of photon energy?

4.3 Erbium doped fiber amplifier A glass fiber has been doped with Er^{3+} ions with doping concentration 8×10^{18} Er^{3+} ions cm^{-3}. Emission and absorption cross-sections of this EDFA at a wavelength of 1550 nm are found to be 3×10^{-21} cm^2 and 2×10^{-21} cm^2, respectively. The Er^{3+} doped fiber is pumped from a 1480 nm laser diode. Obtain the maximum gain coefficient in terms of dB m^{-1} of this medium at 1550 nm. You may assume that the maximum gain coefficient is obtained when there is full population inversion ($N_1 = 0$).

4.4 Erbium-doped fiber amplifier Consider an EDFA that is pumped at 1480 nm. Let N_1 and N_2 be the concentrations of Er^{3+} at E_2 and E_1 manifolds in Figure 4.7 (b). $N_1 + N_2 = N =$ Concentration of Er^{3+} ions, taken as 10^{19} cm^{-3}. When transparency is achieved (no net absorption nor gain), $g = 0$, and $N_1 = N_{1T}$ and $N_2 = N_{2T}$. Show that

$$N_{2T} = N\sigma_{ab}(v)/\left[\sigma_{ab}(v) + \sigma_{em}(v)\right]$$

where $\sigma_{ab}(v)$ and $\sigma_{em}(v)$ are the absorption and emission cross-sections that depend on the frequency (wavelength) of interest. The 1480-nm pump generates N_{2T} of Er^{3+} in the manifold 2 in Figure 4.7 (b). The spontaneous decay time from E_2 to E_1 is τ_{sp}. The pump must accumulate the Er^{3+} ions at E_2 before they can decay. Show that the **minimum pump power needed for transparency** is

$$P_{pump} \approx ALhv_pN\sigma_{ab}/\tau_{sp}(\sigma_{ab} + \sigma_{em})$$

where v_p is the pump frequency, A the pump core area, and L is the fiber length. Calculate the minimum pump power needed for an EDFA if $N = 10^{19}$ cm^{-3}, $\tau_{sp} \approx 10$ ms, core diameter is 4 μm, the fiber length is 7 m, and the transparency is at 1550 nm where emission and absorption cross-sections are 4×10^{-21} cm^2 and 3×10^{-21} cm^2, respectively. What would be the power needed for pumping *all* Er^{3+} ions to E_2?

4.5 Erbium-doped fiber amplifier Consider an EDFA with an Er doping concentration of 2×10^{19} cm^{-3} and pumped at 1480 nm. The fiber length is 3 m. Find the small signal gain G at two wavelengths at 1550 and 1570 nm corresponding to 100% population inversion.

4.6 Erbium-doped fiber amplifier Consider a 5 m long EDFA with Er^{3+} doping concentration of 1×10^{19} cm^{-3}, core diameter of 8 μm. The spontaneous decay time τ_{sp} from E_2 to E_1 is 10 ms. The fiber is pumped at 980 nm from a laser diode. The pump power coupled into the EDFA fiber is 40 mW. What is the fiber length that will absorb the pump radiation when the confinement factor is 80%? Find the small signal gain at 1550 nm corresponding to 80% population inversion.

4.7 Erbium-doped fiber amplifier What is the maximum input signal power beyond which saturation effects take place for an EDFA having a small signal gain (G) of 30 dB, pumped at 80 mW at 980 nm?

4.8 He-Ne Laser The He-Ne laser system energy levels can be quite complicated, as shown in Figure 4.58. There are a number of lasing emissions in the laser output in the red (632.8 nm), green (543.5 nm), orange (612 nm), yellow (594.1 nm), and in the infrared regions at 1.52 μm and 3.39 μm, which give the He-Ne laser its versatility. The pumping mechanism for all these lasers operations is the same, energy transfer from excited He atoms to Ne atoms by atomic collisions in the gas discharge tube. There are two excited states for the He atom in the configuration $1s^12s^2$. The configuration with electron spins parallel has a lower energy than that with opposite electron spins as indicated in Figure 4.58. These two He states can excite Ne atoms to either the $2p^54s^1$ or $2p^55s^1$. There is then a population inversion between these levels and the $2p^53p^1$ and between $2p^55s^1$ and $2p^54p^1$, which leads to the above lasing transitions. Generally, we do not have a single discrete level for the energy of a many-electron atom of given n,ℓ configuration. For example, the atomic configuration $2p^53p^1$ has ten closely spaced energy levels resulting from various different values of m_l and m_s that can be assigned to the sixth excited electron ($3p^1$) and the remaining electrons ($2p^5$) within quantum mechanical rules. Not all transitions to these energy levels are allowed as photon emission requires quantum number selection rules to be obeyed.

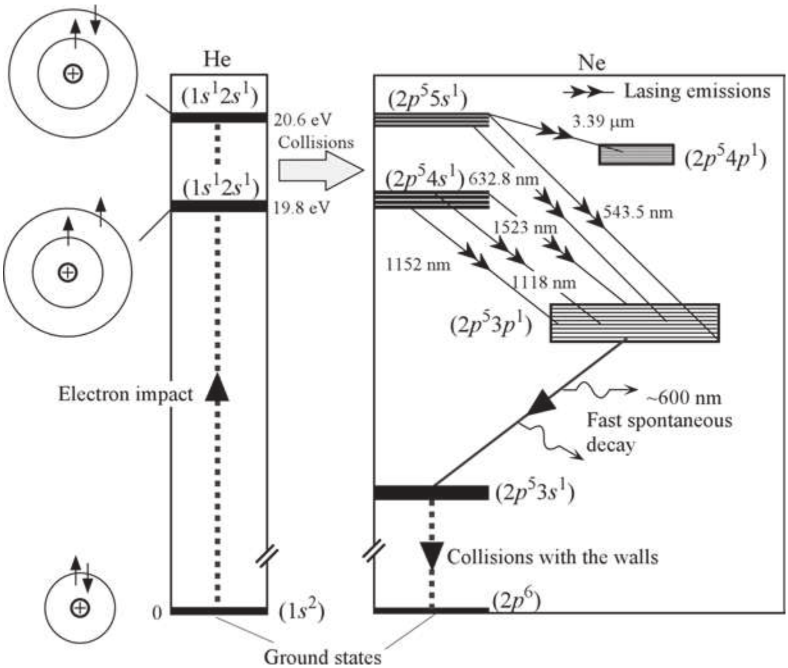


FIGURE 4.58 Various lasing transitions in the He-Ne laser.

(a) Table 4.8 shows some typical commercial He-Ne laser characteristics (within 30–50%) for various wavelengths. Calculate the overall efficiencies of these lasers.

TABLE 4.8 Typical commercial He-Ne laser characteristic

Wavelength (nm)	543.5	594.1	612	632.8	1523
	Green	Yellow	Orange	Red	Infrared
Optical output power (mW)	1.5	2	4	5	1
Typical current (mA)	6.5	6.5	6.5	6.5	6
Typical voltage (V)	2750	2070	2070	1910	3380

(b) The human eye is at least twice as sensitive to orange color light than for red. Discuss typical applications where it is more desirable to use the orange laser.

(c) The 1523-nm emission has potential for use in optical communications by modulating the laser beam externally. By considering the spectral line width ($\Delta\upsilon \approx 1400\,\text{MHz}$), typical powers, and stability, discuss the advantages and disadvantages of using a He-Ne laser over a semiconductor diode.

4.9 The Ar-ion laser The argon-ion laser can provide powerful CW visible coherent radiation of several watts. The laser operation is achieved as follows: The Ar atoms are ionized by electron collisions in a high current electrical discharge. Further multiple collisions with electrons excite the argon ion, Ar^+, to a group of $4p$ energy levels ~35 eV above the atomic ground state as shown in Figure 4.59. Thus a population inversion forms between the $4p$ levels and the $4s$ level which is about 33.5 eV above the Ar atom ground level. Consequently, the stimulated radiation from the $4p$ levels down to the $4s$ level contains a series of wavelengths ranging from 351.1 nm to 528.7 nm. Most of the power, however, is concentrated, approximately equally, in the 488 nm and 514.5 nm emissions. The Ar^+ ion at the lower laser level ($4s$) returns to its neutral atomic ground state via a radiative decay to the Ar^+ ion ground state, followed by recombination with an electron to form the neutral atom. The Ar atom is then ready for "pumping" again.

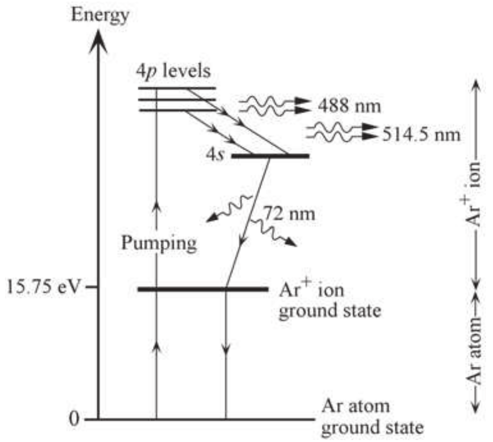


FIGURE 4.59 The Ar^+-ion laser energy levels.

(a) Calculate the energy drop involved in the excited Ar^+ ion when it is stimulated to emit the radiation at 514.5 nm.
(b) The Doppler broadened linewidth of the 514.5-nm radiation is about 3500 MHz ($\Delta\upsilon$) and is between the half-intensity points. Calculate the Doppler broadened width $\Delta\lambda$ in the wavelength.
(c) Estimate the operation temperature of the argon ion gas; give the temperature in °C.
(d) In a particular argon-ion laser the discharge tube, made of Beryllia (Beryllium Oxide), is 30 cm long and has a bore of 3 mm in diameter. When the laser is operated with a current of 40 A at 200 V DC, the total output power in the emitted radiation is 3 W. What is the efficiency of the laser?

4.10 He-Ne laser A particular commercial He-Ne laser operating at 632.8 nm has a tube that is 40 cm long. The operating temperature is 130°C.
(a) Estimate the Doppler broadened linewidth ($\Delta\lambda$) in the output spectrum.
(b) What are the mode number m values that satisfy the resonant cavity condition? How many modes are therefore allowed?
(c) What is the separation $\Delta\upsilon_m$ in the frequencies of the modes? What is the mode separation $\Delta\lambda_m$ in wavelength?
(d) Show that if during operation, the temperature changes the length of the cavity by δL, the wavelength of a given mode changes by $\delta\lambda_m$,

$$\delta\lambda_m = (\lambda_m/L)\delta L$$

Given that typically a glass has a linear expansion coefficient $\alpha \approx 1 \times 10^{-6}\ \mathrm{K}^{-1}$, calculate the change $\delta\lambda_m$ in the output wavelength (due to one particular mode) as the tube warms up from 20°C to 130°C, and also per degree change in the operating temperature. Note that $\delta L/L = \alpha\delta T$, and $L' = L\left[1 + \alpha(T' - T)\right]$. Change in mode wavelength $\delta\lambda_m$ with the change δL in the cavity length L is called **mode sweeping**.
(e) How do the mode separations $\Delta\upsilon_m$ and $\Delta\lambda_m$ change as the tube warms up from 20°C to 130°C during operation?
(f) How can you increase the output intensity from the He-Ne laser?

4.11 Doppler broadening Consider a He-Ne laser with Doppler broadened linewidth of 2000 MHz. What should be the length of the resonator cavity so that a single mode is found to oscillate? You may assume that there is an oscillating mode at the line centre.

4.12 Einstein coefficients and critical photon concentration $\rho(\upsilon)$ is the energy of the electromagnetic radiation per unit volume per unit frequency due to photons with energy $h\upsilon = E_2 - E_1$. Suppose that there are N_{ph} photons per unit volume. Each has an energy $h\upsilon$. The frequency range of emission is $\Delta\upsilon$. Then, $\rho(\upsilon) = (N_{ph}h\upsilon)/\Delta\upsilon$. In case of an Ar-ion laser system with the emission wavelength 488 nm and the linewidth in the output spectrum is 6×10^9 Hz (between half intensity points), estimate the photon concentration necessary to achieve more stimulated emission than spontaneous emission.

4.13 Photon concentration in a gas laser The Ar-ion laser has a strong lasing emission at 488 nm. The laser tube is 1 m in length, and the bore diameter is 3 mm. The output power is 1 W. Assume that most of the output power is in the 488-nm emission. Assume that the tube end has a transmittance T of 0.1. Calculate the photon output flow (number of lasing photons emitted from the tube per unit time), photon flux (number of lasing photons emitted per unit area per unit time), and estimate the order of magnitude of the steady state photon concentration (at 488 nm) in the tube (assume that the gas refractive index is approximately 1).

4.14 He-Ne Laser and the photon cavity lifetime A He-Ne laser operating at 632.8 nm has a tube that is 40 cm long and end mirrors with reflectances 99.0% and 99.9%. The tube diameter is 0.80 mm and the output power is 1 mW. The linewidth $\Delta\upsilon$ is 1 GHz and the loss coefficient α_s is $0.06\,\text{m}^{-1}$. The spontaneous emission lifetime is 125 ns. (a) Calculate the threshold gain. (b) Calculate the threshold population inversion. (c) Calculate the photon cavity lifetime. (d) Calculate the photon concentration in the cavity.

4.15 Threshold population Nd:YAG laser The emission wavelength of a Nd:YAG laser is 1064 nm, the optical gain linewidth is about 100 GHz, the spontaneous lifetime (τ_{sp}) of the upper laser level is 200 μs. The loss coefficient as is $0.2\,\text{m}^{-1}$. A YAG rod of 10 cm in length and has its ends coated to achieve reflectances of 99.9% and 92% (emission end). What is the threshold concentration of pumped Nd-ions (Nd^{3+})?

4.16 Threshold gain and population inversion in an Ar-ion laser The emission wavelength of an Ar-ion gas laser is 488 nm. The tube length $L = 80\,\text{cm}$, tube mirror reflectances are approximately 99.0% and 96.0%. The linewidth $\Delta\upsilon = 2\,\text{GHz}$, the loss coefficient is $\alpha_s \approx 0.15\,\text{m}^{-1}$, spontaneous decay time constant $\tau_{sp} = 1/A_{21} \approx 10\,\text{ns}$, $n \approx 1$. Estimate the threshold population inversion.

4.17 Fabry–Perot optical resonator in gas lasers An idealized He-Ne laser optical cavity of length 50 cm has 99% reflectance $R_1 = R_2$ at both ends. Find out the separation $\Delta\lambda$ of the modes in wavelength and the spectral width $\Delta\lambda$ for this laser.

4.18 Homogeneous broadening

(a) Consider the emission of radiation from an isolated atom in which an electron transits down from E_2 to E_1. Suppose that the lifetime of the atom at E_2 is τ_2. We will assume that the lifetime of the state E_1 is very long and neglect it. The emitted radiation's optical field is not simply sinusoidal oscillations, $\sin(\omega_o t)$ type, that would go on forever since we know this is not possible. The optical field's amplitude is decayed exponentially to give the radiation a finite temporal length, which results in a finite spectral width. Suppose that the field follows $E(t) = E_o \exp(-t/2\tau_2) \times \cos(\omega_o t)$ for all $t > 0$, and zero for $t < 0$. In practice, $1/\tau_2 \ll \omega_o$. The Fourier transform (FT) $E_{FT}(\omega)$ of $E(t)$ is approximately given by

$$E_{FT}(\omega) = \text{FT of } E(t) \approx \frac{E_o/4\pi}{(2\tau_2)^{-1} + j(\omega - \omega_o)}$$

where we neglected the additional small term that has $j(\omega + \omega_o)$ in the denominator. Show that the light intensity is maximum at $\omega = \omega_o$ and that the spectrum has an FWHM width $\Delta\omega_{FWHM} = 1/\tau_2$. What is your conclusion? (The spectrum is called a *Lorentzian lineshape*.)

(b) Consider the two-energy-level system shown in Figure 4.5. The rate equation for the change in the concentration N_2 of atoms at E_2 from Eqs. (4.2.1) and (4.2.2) is

$$dN_2/dt = -A_{21}N_2 - B_{21}\rho(\upsilon)N_2 - B_{12}\rho(\upsilon)N_1$$

Since $E_2 > E_1$, we can take the equilibrium (unpumped) value of N_2 as zero. Suppose we suddenly pump the atoms to the energy level E_2 and we assume the radiation density $\rho(\upsilon)$ remains small. From the rate equation, what is the time evolution of $N_2(t)$? What does A_{21} represent? What is the time evolution of the total emitted optical power? How do your results relate to part (a)? What is your conclusion?

4.19 Homogeneous and inhomogeneous broadening The expressions for homogeneous and inhomogeneous broadening in photonics are given by the *Lorentzian* and *Gaussian* lineshape functions, *i.e.,*

$$g_L(\upsilon) = \frac{\gamma/\pi}{(\upsilon - \upsilon_o)^2 + \gamma^2} \quad \text{and} \quad g_G(\upsilon) = \frac{1}{\pi^{1/2}\Gamma}\exp\left[-(\upsilon - \upsilon_o)^2/\Gamma^2\right]$$

where γ and Γ are characteristic parameters that control the width of the spectrum, for example, for lifetime or natural broadening $1/\gamma = \tau_2$, lifetime of upper state E_2. Further, these functions are such that when integrated over all frequencies they are unity, *i.e.*, $\int g(v)dv = 1$. (a) Show that for Lorentzian broadening, the FWHM spectral width $\Delta v_{\text{FWHM}} = 2\gamma$. (b) Show that for Gaussian broadening, $\Delta v_{\text{FWHM}} = 2\Gamma\left[\ln(2)\right]^{1/2}$. (c) Suppose that widths Δv_{FWHM} of the two distributions where the same, and both have the same central frequency v_o. Plot and compare the two distributions for the following: $v_o = 100\,\text{THz}$ and $\Delta v_{\text{FWHM}} = 1\,\text{GHz}$ from $v = 0$ to $v = v_o \pm 2\Delta v_{\text{FWHM}}$.

4.20 Natural broadening in gas lasers In a He-Ne laser, the Doppler broadened linewidth of the lasing emission is 2.0 GHz. For natural broadening, the spectral width $\Delta v_{1/2}$ is given by $\Delta v_{1/2} \approx (2\pi\tau_2)^{-1}$ where τ_2 is the lifetime of the upper laser level, and we can take $A_{21} = 1/\tau_2$. Given $A_{21} = 3.5 \times 10^{-6}\ \text{s}^{-1}$ for the 632.8 nm emission, calculate the natural linewidth of this emission.

4.21 Q-switching The energy of the pulse in a Q-switched Nd^{3+}: YAG laser is 20 mJ, and the pulse duration is 10 ns. The pulse repetition rate is 200 kHz. Calculate the peak power in the pulse and the average power emitted by the laser.

4.22 Mode-locked lasers (a) Consider a mode-locked Nd^{3+}:glass laser with emission wavelength 1060 nm. The spectral width of the optical gain curve is 8 THz. The refractive index of the Nd^{3+}:glass laser medium is 1.46. Find the mode-locked pulse width, and the repetition rate, for a 10 cm long the Nd^{3+} glass rod. Find the separation of the pulses in space. (b) Consider a He-Ne laser emitting at 632.8 nm, and mode locked. The tube length is 60 cm. The spectral width of the optical gain curve is 2 GHz. Find the mode-locked pulse width, and the repetition rate. What is the separation of pulses in space?

4.23 Photon cavity lifetime and total attenuation Consider a semiconductor Fabry–Perot optical cavity of length L, end mirrors with reflectances R_1 and R_2, and attenuation inside the cavity given by $\alpha_s = 20\ \text{cm}^{-1}$. The total attenuation coefficient for the optical cavity is denoted by α_t. Find out the photon cavity time for an $In_{0.60}Ga_{0.40}As_{0.85}P_{0.15}$ FP optical cavity that has a refractive index of 3.7, and a length of 400 μm.

4.24 Population inversion in a GaAs homojunction laser diode Consider the energy diagram of a forward biased GaAs homojunction LD as shown in Figure 4.32 (b). For simplicity we assume a symmetrical device ($n = p$) and also assume that population inversion has been just reached when the conduction band on the n^+-side overlaps the valence band on the p^+-side around the center of the depletion region, as illustrated in Figure 4.32 (b), and results in $E_{Fn} - E_{Fp} = E_g$. Estimate the minimum carrier concentration $n = p$ for population inversion in GaAs at 300 K. The intrinsic carrier concentration in GaAs is appropriately $2 \times 10^6\ \text{cm}^{-3}$. Assume for simplicity that

$$n = n_i\exp\left[(E_{Fn} - E_{Fi})/(k_BT)\right] \quad \text{and} \quad p = n_i\exp\left[(E_{Fi} - E_{Fn})/(k_BT)\right]$$

(*Note*: The analysis will only be an order of magnitude as the above equations do not hold in degenerate semiconductors. A better approach is to use the Joyce–Dixon equations as can be found in advanced textbooks. The present analysis is an order of magnitude calculation.)

4.25 InGaAsP-InP laser The optical cavity length is 200 μm for a given Fabry–Perot InGaAsP-InP laser diode with emission wavelength 1550 nm and the refractive index of InGaAsP is 3.7. The internal cavity loss (α_s) is $25\ \text{cm}^{-1}$. The optical gain bandwidth (as measured between half intensity points) will normally depend on the pumping current (diode current) but for this problem assume that it is 3 nm. (a) What is the mode integer m of the peak radiation? (b) What is the separation between the modes of the cavity? (c) How many modes are there in the cavity? (d) What is the reflectance at the ends of the optical cavity mode of the InGaAsP crystal?

4.26 SQW laser Consider a SQW (single quantum well) laser which has an ultrathin active InGaAs of bandgap 0.70 eV and thickness 8 nm between two layers of InAlAs which has a bandgap of 1.44 eV. Effective mass of conduction electrons in InGaAs is about $0.04m_e$ and that of the holes in the valence band is $0.44m_e$ where m_e is the mass of the electron in vacuum. Calculate the first and second electron energy levels above E_c and the first hole energy level below E_v in the QW. What is the lasing emission wavelength for this SQW laser? What is this wavelength if the transition were to occur in bulk InGaAs with the same bandgap?

4.27 Semiconductor lasers and wavelength shift with temperature Consider a Fabry–Perot semiconductor laser diode operating at 1550 nm. The active region is a III–V quaternary semiconductor alloy of InGaAsP (but quite close to being InGaAs). The LD has a cavity length of 200 μm. The refractive index of InGaAsP is approximately 3.60 and $dn/dT \approx 2.5 \times 10^{-4}\ \text{K}^{-1}$, the linear thermal expansion coefficient is $5.8 \times 10^{-6}\ \text{K}^{-1}$. The bandgap of InGaAs follows the Varshi equation $E_g = E_{go} - AT^2/(B + T)$ with $E_{go} = 0.950\ \text{eV}$, $A = 5 \times 10^{-4}\ \text{eV K}^{-1}$, $B = 300\,\text{K}$. Find the shift in the optical gain curve, and the emission wavelength for a given mode per unit temperature change. What is your conclusion?

4.28 Fabry–Perot optical resonator in semiconductor lasers Consider a semiconductor Fabry–Perot optical cavity of length 250 μm having end-mirrors of reflectance R. Calculate the cavity mode nearest to the free space wavelength of 1550 nm. If the refractive index of semiconductor is 3.67, calculate the separation of the modes. Calculate the spectral width for $R = 0.95$ and 0.8. What is your conclusion?

4.29 GaAs DH laser diode Consider a GaAs DH laser diode that lases at 850 nm. It has an active layer (cavity) length L of 250 μm. The active layer thickness d is 0.2 μm and the width (W) is 5 μm. The refractive index is 3.6, and the attenuation coefficient α_s inside the cavity is $20\,\text{cm}^{-1}$. The radiative lifetime τ_r in the active region is 3 ns. Find the threshold gain g_{th}, carrier concentration n_{th}, current density J_{th}, and current I_{th}. Find the output optical power at $I = 2I_{\text{th}}$, and the external slope efficiency η_{slope}.

4.30 Threshold current and power output from a 1310-nm FP laser diode Consider a double heterostructure InGaAsP semiconductor laser operating at 1310 nm. The cavity length $L \approx 200\,\mu\text{m}$, width $W \approx 8\,\mu\text{m}$, and $d \approx 0.2\,\mu\text{m}$. The refractive index $n \approx 3.6$. The loss coefficient $\alpha_s \approx 25\,\text{cm}^{-1}$ and the direct recombination coefficient $B \approx 2 \times 10^{-16}\,\text{m}^3\,\text{s}^{-1}$. Assume that the optical confinement factor is 1. Find the threshold gain g_{th}, carrier concentration n_{th}, current density J_{th} and the threshold current I_{th}. Find the output optical power at $I = 1.5I_{\text{th}}$, and the external slope efficiency η_{slope}.

4.31 Laser diode efficiencies Consider a particular commercial AlGaInP power laser diode with an emission wavelength of 630 nm (red). The threshold current at 30°C is 50 mA. The typical operating voltage for this device is 2.5 V. At $I = 100\,\text{mA}$. The output optical power is 40 mW.

(a) Calculate the external QE, external differential QE, external power efficiency, and the slope efficiency of the laser diode. What is the current required for an output of 30 mW?

(b) The threshold current at 60°C is measured to be 80 mA. Assuming that $I_{\text{th}} = A\exp(T/T_o)$ where T_o is a characteristic temperature, and given I_{th} at two temperatures, find the characteristic T_o temperature parameter. What is the threshold current at 0°C?

4.32 Laser diode efficiencies Consider an InGaAsP FP laser diode operating at $\lambda = 1330\,\text{nm}$ for optical communications. The threshold current is 6 mA. At $I = 25\,\text{mA}$, the output optical power is 6 mW and the voltage across the diode is 1.2 V. Calculate external quantum efficiency (QE), external differential QE, power conversion efficiency, and the slope efficiency of the diode. What is the forward diode current that gives an output optical power of 3 mW?

4.33 Laser diode efficiencies Consider a commercial InGaAsP FP laser diode operating at $\lambda = 1550$ nm for primary use in fiber-to-home communications. The threshold current is 8 mA. At $I = 25\,\text{mA}$, the output optical power is 5 mW, and the voltage across the diode is 1.2 V. Calculate external quantum efficiency (QE), external differential QE, external power efficiency (conversion efficiency), and the slope efficiency.

4.34 Laser diode extraction efficiency Consider an optical cavity with end mirrors R_1 and R_2 and an internal loss coefficient of α_s. The total loss coefficient α_t is given by

$$\alpha_t = \alpha_s + (1/2L)\ln(1/R_1) + (1/2L)\ln(1/R_2)$$

where the second term on the right represents the loss at the cavity end 1, associated with R_1, and the third is the loss at the cavity end 2, associated with R_2. Suppose that we are interested in how much light is coupled out from the cavity end 1. The extraction efficiency is then given by

$$\text{Extraction efficiency} = \frac{\text{Loss from cavity end 1}}{\text{Total loss}}$$

that is

$$\eta_{\text{EE}} = (1/2L)\ln(1/R_1)/\alpha_t$$

A semiconductor laser made of GaAs as its active layer has refractive index n is 3.6, the cavity length L is 300 μm, and the internal loss coefficient $\alpha_s = 20\ \text{cm}^{-1}$. (a) What is η_{EE}? (b) Explain the effect when there are no internal losses? (c) Suppose the end 2 had a 100% reflecting dielectric mirror and $\alpha_s = 20\ \text{cm}^{-1}$. What is η_{EE}? (d) Suppose in (c) you set the internal losses to 2. What is η_{EE}?

4.35 Temperature dependence of laser characteristics The threshold current of a laser diode increases with temperature because more current is needed to achieve the necessary population inversion at higher temperatures. For a particular commercial laser diode, the threshold current, I_{th}, is 60 mA at 27 °C, 45 mA at 0°C, and 90 mA at 57°C. Using these three points, explain the dependence of I_{th} on the absolute temperature. Obtain an empirical expression for it.

4.36 UV laser diode Consider a GaN-based laser diode that emits at a peak wavelength of 360 nm. What is the transition energy ΔE in eV that corresponds to 360 nm? The threshold current is 50 mA and the emitted power is 25 mW at a forward current of 70 mA and a voltage of 5 V. Find the slope efficiency, EQE, EDQE, PCE. What is the power emitted at 70 mA?

4.37 Single DFB frequency lasers An InGaAsP DFB laser operates at 1550 nm. Suppose that the effective refractive index $n \approx 3.5$ and the cavity length is 60 μm. What should be the corrugation period Λ for a first-order grating, $q = 1$? What is Λ for a second-order grating, $q = 2$? How many corrugations are needed for a first-order grating? How many corrugations are there for $q = 2$?

4.38 Single frequency DFB lasers A DFB Laser Diode emits at 1063 nm, and has $d\lambda/dT = 0.07\,\text{nm K}^{-1}$ and $d\lambda/dI = 0.004\,\text{nm mA}^{-1}$. What should be change in the temperature, δT, that gives a wavelength shift equal to the width $\delta\lambda$ of the single-mode output spectrum? What should be the change δI in the current that gives a wavelength shift equal to $\delta\lambda$? What percentage of the operating current, 195 mA, is this change?

4.39 VCSEL Consider the VCSELs in Table 4.6. (a) For each, calculate the power conversion efficiency η_{PCE}, and then average these values. What is the average PCE? (b) For each, using P_o, I, and I_{th}, calculate the slope efficiency and compare with the quoted (observed) value. (c) The emitted power from an STM (single transverse mode) VCSEL is about 1 mW. The diameter is roughly ~10 μm as this is a single transverse mode device. Estimate the emitted intensity in kW cm^{-2}.

4.40 VCSEL and DBR reflectances Consider a VCSELs that has 55 pairs of quarter-wave alternating layers of high and low refractive index layers for the lower DBR, and 35 pairs in the DBR on the emission side. Assume that the refractive indices of the DBR, layers are $n_1 = 3.60$ and $n_2 = 3.42$ (4% lower). Take $n_0 = 1$ and $n_3 = 3.6$, and calculate the reflectance of each DBR. (See Example 1.7.2 in Chapter 1.) *Note*: For N pairs of layers, $R = \left[n_1^{2N} - (n_0/n_3)n_2^{2N}\right]^2/\left[n_1^{2N} + (n_0/n_3)n_2^{2N}\right]^2$; see Eq. (1.7.3).

4.41 SOA Consider a TW SOA for use at 1550 nm. It has an active region that is 400 μm long. The gain of the SOA is quoted as 20 dB. Suppose that the confinement factor Γ is 0.4, and the loss coefficient α_s is 20 cm^{-1}. What is the gain coefficient g of the active region at this amplification? *Note*: The gain $G = \exp\left[(\Gamma g - \alpha_t)L\right]$, and assume perfect AR coatings at the facets of the semiconductor, that is, $R_1 = R_2 = 0$.

4.42 Laser diode modulation A GaAs Laser Diode that has the following typical properties. The radiative lifetime $\tau_r \approx 3\,\text{ns}$, non radiative lifetime $\tau_{nr} \approx 60\,\text{ns}$, total attenuation coefficient $\alpha_t \approx 5000\,\text{m}^{-1}$ for the GaAs active layer of length of about 250 μm having refractive index n of 3.6. Obtain the relaxation oscillation frequency and hence the bandwidth at bias currents $I_1 = 2I_{\text{th}}$ and $3I_{\text{th}}$? What is the delay time if the LD is biased at $0.8I_{\text{th}}$ and switched by a current to $2I_{\text{th}}$? What is this delay time if the LD is not biased at all?

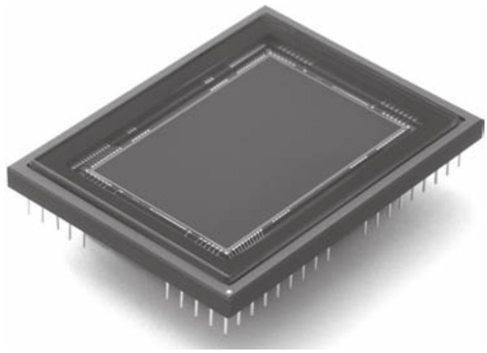
4.43 Laser diode modulation A laser diode has the radiative recombination time $\tau_r = 3\,\text{ns}$, nonradiative lifetime $\tau_{nr} = 40\,\text{ns}$, and the photon cavity lifetime $\tau_{\text{ph}} = 6\,\text{ps}$. Find the resonance frequency for this laser diode at forward currents $I_1 = 2I_{\text{th}}$ and $3I_{\text{th}}$.

Theodore Maiman's ruby laser. The helical xenon flash surrounds the ruby crystal rod (an Al_2O_3 crystal doped with Cr^{3+} ions) and provides optical pumping of the chromium ions in the ruby rod. *(Courtesy of HRL Laboratories, LLC, Malibu, California.)*

The detector is like the journalist who must determine what, where, when, which, and how? What is the identity of the particle? Exactly where is it when it is observed? When does the particle get to the detector? Which way is it going? How fast is it moving?

—Sheldon L. Glashow[1]

The inventors of the CCD (charge-coupled device) image sensor at AT&T Bell Labs: Willard Boyle (left) and George Smith (right). The CCD was invented in 1969, the first CCD solid state camera was demonstrated in 1970, and a broadcast quality TV camera by 1975. (W. S. Boyle and G. E. Smith, "Charge Coupled Semiconductor Devices," *Bell Syst. Tech. J., 49,* 587, 1970; *Courtesy of Alcatel-Lucent Bell Labs.*)



A CCD image sensor. The FTF6040C is a full-frame color CCD image sensor designed for professional digital photography, scientific and industrial applications with 24 megapixels, and a wide dynamic range. Chip imaging area is 36 mm $\times$ 24 mm, and pixel size is 6 μm $\times$ 6 μm. *(Courtesy of Teledyne-DALSA.)*

[1] Sheldon L. Glashow, *Interactions* (Warner Books, 1988), p. 101.