Chapter 3

Optical Transmitters

The role of the optical transmitter is to convert an electrical input signal into the corresponding optical signal and then launch it into the optical fiber serving as a communication channel. The major component of optical transmitters is an optical source. Fiber-optic communication systems often use semiconductor optical sources such as light-emitting diodes (LEDs) and semiconductor lasers because of several inherent advantages offered by them. Some of these advantages are compact size, high efficiency, good reliability, right wavelength range, small emissive area compatible with fiber-core dimensions, and possibility of direct modulation at relatively high frequencies. Although the operation of semiconductor lasers was demonstrated as early as 1962, their use became practical only after 1970, when semiconductor lasers operating continuously at room temperature became available [1]. Since then, semiconductor lasers have been developed extensively because of their importance for optical communications. They are also known as laser diodes or injection lasers, and their properties have been discussed in several recent books [2]–[16]. This chapter is devoted to LEDs and semiconductor lasers and their applications in lightwave systems. After introducing the basic concepts in Section 3.1, LEDs are covered in Section 3.2, while Section 3.3 focuses on semiconductor lasers. We describe single-mode semiconductor lasers in Section 3.4 and discuss their operating characteristics in Section 3.5. The design issues related to optical transmitters are covered in Section 3.6.

3.1 Basic Concepts

Under normal conditions, all materials absorb light rather than emit it. The absorption process can be understood by referring to Fig. 3.1, where the energy levels $E_1$ and $E_2$ correspond to the ground state and the excited state of atoms of the absorbing medium. If the photon energy $h\nu$ of the incident light of frequency $\nu$ is about the same as the energy difference $E_2 - E_1$, the photon is absorbed by the atom, which ends up in the excited state. Incident light is attenuated as a result of many such absorption events occurring inside the medium.
The excited atoms eventually return to their normal “ground” state and emit light in the process. Light emission can occur through two fundamental processes known as spontaneous emission and stimulated emission. Both are shown schematically in Fig. 3.1. In the case of spontaneous emission, photons are emitted in random directions with no phase relationship among them. Stimulated emission, by contrast, is initiated by an existing photon. The remarkable feature of stimulated emission is that the emitted photon matches the original photon not only in energy (or in frequency), but also in its other characteristics, such as the direction of propagation. All lasers, including semiconductor lasers, emit light through the process of stimulated emission and are said to emit coherent light. In contrast, LEDs emit light through the incoherent process of spontaneous emission.

3.1.1 Emission and Absorption Rates

Before discussing the emission and absorption rates in semiconductors, it is instructive to consider a two-level atomic system interacting with an electromagnetic field through transitions shown in Fig. 3.1. If $N_1$ and $N_2$ are the atomic densities in the ground and the excited states, respectively, and $\rho_{ph}(\nu)$ is the spectral density of the electromagnetic energy, the rates of spontaneous emission, stimulated emission, and absorption can be written as

\[
R_{\text{spon}} = AN_2, \quad R_{\text{stim}} = BN_2\rho_{\text{em}}, \quad R_{\text{abs}} = B'N_1\rho_{\text{em}},
\]  

(3.1.1)

where $A$, $B$, and $B'$ are constants. In thermal equilibrium, the atomic densities are distributed according to the Boltzmann statistics [18], i.e.,

\[
N_2/N_1 = \exp(-E_g/k_BT) \equiv \exp(-h\nu/k_BT),
\]  

(3.1.2)

where $k_B$ is the Boltzmann constant and $T$ is the absolute temperature. Since $N_1$ and $N_2$ do not change with time in thermal equilibrium, the upward and downward transition rates should be equal, or

\[
AN_2 + BN_2\rho_{\text{em}} = B'N_1\rho_{\text{em}}.
\]  

(3.1.3)

By using Eq. (3.1.2) in Eq. (3.1.3), the spectral density $\rho_{\text{em}}$ becomes

\[
\rho_{\text{em}} = \frac{A/B}{(B'/B)\exp(h\nu/k_BT) - 1}.
\]  

(3.1.4)
In thermal equilibrium, $\rho_{em}$ should be identical with the spectral density of blackbody radiation given by Planck's formula [18]

$$\rho_{em} = \frac{8\pi \hbar^3}{c^3} \frac{\nu^3}{\exp(h\nu/k_BT) - 1}. \quad (3.1.5)$$

A comparison of Eqs. (3.1.4) and (3.1.5) provides the relations

$$A = (8\pi \hbar^3/c^3)B; \quad B' = B. \quad (3.1.6)$$

These relations were first obtained by Einstein [17]. For this reason, $A$ and $B$ are called Einstein’s coefficients.

Two important conclusions can be drawn from Eqs. (3.1.1)–(3.1.6). First, $R_{spon}$ can exceed both $R_{stim}$ and $R_{abs}$ considerably if $k_BT > h\nu$. Thermal sources operate in this regime. Second, for radiation in the visible or near-infrared region ($h\nu \sim 1$ eV), spontaneous emission always dominates over stimulated emission in thermal equilibrium at room temperature ($k_BT \approx 25$ meV) because

$$R_{stim}/R_{spon} = \left[\exp(h\nu/k_BT) - 1\right]^{-1} \ll 1. \quad (3.1.7)$$

Thus, all lasers must operate away from thermal equilibrium. This is achieved by pumping lasers with an external energy source.

Even for an atomic system pumped externally, stimulated emission may not be the dominant process since it has to compete with the absorption process. $R_{stim}$ can exceed $R_{abs}$ only when $N_2 > N_1$. This condition is referred to as population inversion and is never realized for systems in thermal equilibrium [see Eq. (3.1.2)]. Population inversion is a prerequisite for laser operation. In atomic systems, it is achieved by using three- and four-level pumping schemes [18] such that an external energy source raises the atomic population from the ground state to an excited state lying above the energy state $E_2$ in Fig. 3.1.

The emission and absorption rates in semiconductors should take into account the energy bands associated with a semiconductor [5]. Figure 3.2 shows the emission process schematically using the simplest band structure, consisting of parabolic conduction and valence bands in the energy–wave-vector space ($E-k$ diagram). Spontaneous emission can occur only if the energy state $E_2$ is occupied by an electron and the energy state $E_1$ is empty (i.e., occupied by a hole). The occupation probability for electrons in the conduction and valence bands is given by the Fermi–Dirac distributions [5]

$$f_c(E_2) = \{1 + \exp[(E_2 - E_{fc})/k_BT]\}^{-1}, \quad (3.1.8)$$

$$f_v(E_1) = \{1 + \exp[(E_1 - E_{fv})/k_BT]\}^{-1}, \quad (3.1.9)$$

where $E_{fc}$ and $E_{fv}$ are the Fermi levels. The total spontaneous emission rate at a frequency $\omega$ is obtained by summing over all possible transitions between the two bands such that $E_2 - E_1 = E_{em} = \hbar\omega$, where $\omega = 2\pi\nu$, $\hbar = h/2\pi$, and $E_{em}$ is the energy of the emitted photon. The result is

$$R_{spon}(\omega) = \int_{E_1}^{\infty} A(E_1, E_2) f_c(E_2)[1 - f_v(E_1)]\rho_{cv} dE_2, \quad (3.1.10)$$
Figure 3.2: Conduction and valence bands of a semiconductor. Electrons in the conduction band and holes in the valence band can recombine and emit a photon through spontaneous emission as well as through stimulated emission.

where $\rho_{cv}$ is the joint density of states, defined as the number of states per unit volume per unit energy range, and is given by

$$\rho_{cv} = \frac{(2m_r)^{3/2}}{2\pi^2\hbar^3} (\hbar \omega - E_g)^{1/2}. \quad (3.1.11)$$

In this equation, $E_g$ is the bandgap and $m_r$ is the reduced mass, defined as $m_r = m_e m_h / (m_e + m_h)$, where $m_e$ and $m_h$ are the effective masses of electrons and holes in the conduction and valence bands, respectively. Since $\rho_{cv}$ is independent of $E_2$ in Eq. (3.1.10), it can be taken outside the integral. By contrast, $A(E_1, E_2)$ generally depends on $E_2$ and is related to the momentum matrix element in a semiclassical perturbation approach commonly used to calculate it [2].

The stimulated emission and absorption rates can be obtained in a similar manner and are given by

$$R_{\text{stim}}(\omega) = \int_{E_1}^{\infty} B(E_1, E_2) f_c(E_2)[1 - f_v(E_1)] \rho_{cv} \rho_{em} dE_2, \quad (3.1.12)$$

$$R_{\text{abs}}(\omega) = \int_{E_1}^{\infty} B(E_1, E_2) f_v(E_1)[1 - f_c(E_2)] \rho_{cv} \rho_{em} dE_2, \quad (3.1.13)$$

where $\rho_{em}(\omega)$ is the spectral density of photons introduced in a manner similar to Eq. (3.1.1). The population-inversion condition $R_{\text{stim}} > R_{\text{abs}}$ is obtained by comparing Eqs. (3.1.12) and (3.1.13), resulting in $f_c(E_2) > f_v(E_1)$. If we use Eqs. (3.1.8) and (3.1.9), this condition is satisfied when

$$E_{fc} - E_{fv} > E_2 - E_1 > E_g. \quad (3.1.14)$$
Since the minimum value of \( E_2 - E_1 \) equals \( E_g \), the separation between the Fermi levels must exceed the bandgap for population inversion to occur \([19]\). In thermal equilibrium, the two Fermi levels coincide (\( E_{fc} = E_{fv} \)). They can be separated by pumping energy into the semiconductor from an external energy source. The most convenient way for pumping a semiconductor is to use a forward-biased \( p-n \) junction.

### 3.1.2 \( p-n \) Junctions

At the heart of a semiconductor optical source is the \( p-n \) junction, formed by bringing a \( p \)-type and an \( n \)-type semiconductor into contact. Recall that a semiconductor is made \( n \)-type or \( p \)-type by doping it with impurities whose atoms have an excess valence electron or one less electron compared with the semiconductor atoms. In the case of \( n \)-type semiconductor, the excess electrons occupy the conduction-band states, normally empty in undoped (intrinsic) semiconductors. The Fermi level, lying in the middle of the bandgap for intrinsic semiconductors, moves toward the conduction band as the dopant concentration increases. In a heavily doped \( n \)-type semiconductor, the Fermi level \( E_{fc} \) lies inside the conduction band; such semiconductors are said to be degenerate. Similarly, the Fermi level \( E_{fv} \) moves toward the valence band for \( p \)-type semiconductors and lies inside it under heavy doping. In thermal equilibrium, the Fermi level must be continuous across the \( p-n \) junction. This is achieved through diffusion of electrons and holes across the junction. The charged impurities left behind set up an electric field strong enough to prevent further diffusion of electrons and holds under equilibrium conditions. This field is referred to as the built-in electric field. Figure 3.3(a) shows the energy-band diagram of a \( p-n \) junction in thermal equilibrium and under forward bias.

When a \( p-n \) junction is forward biased by applying an external voltage, the built-in electric field is reduced. This reduction results in diffusion of electrons and holes across the junction. An electric current begins to flow as a result of carrier diffusion. The current \( I \) increases exponentially with the applied voltage \( V \) according to the well-known relation \([5]\)

\[
I = I_s [\exp(qV/k_BT) - 1],
\]

where \( I_s \) is the saturation current and depends on the diffusion coefficients associated with electrons and holes. As seen in Fig. 3.3(a), in a region surrounding the junction (known as the depletion width), electrons and holes are present simultaneously when the \( p-n \) junction is forward biased. These electrons and holes can recombine through spontaneous or stimulated emission and generate light in a semiconductor optical source.

The \( p-n \) junction shown in Fig. 3.3(a) is called the **homojunction**, since the same semiconductor material is used on both sides of the junction. A problem with the homojunction is that electron–hole recombination occurs over a relatively wide region (\(~1–10 \mu m\)) determined by the diffusion length of electrons and holes. Since the carriers are not confined to the immediate vicinity of the junction, it is difficult to realize high carrier densities. This carrier-confinement problem can be solved by sandwiching a thin layer between the \( p \)-type and \( n \)-type layers such that the bandgap of the sandwiched layer is smaller than the layers surrounding it. The middle layer may or may
not be doped, depending on the device design; its role is to confine the carriers injected inside it under forward bias. The carrier confinement occurs as a result of bandgap discontinuity at the junction between two semiconductors which have the same crystalline structure (the same lattice constant) but different bandgaps. Such junctions are called heterojunctions, and such devices are called double heterostructures. Since the thickness of the sandwiched layer can be controlled externally (typically, $\sim 0.1 \, \mu m$), high carrier densities can be realized at a given injection current. Figure 3.3(b) shows the energy-band diagram of a double heterostructure with and without forward bias.

The use of a heterostructure geometry for semiconductor optical sources is doubly beneficial. As already mentioned, the bandgap difference between the two semiconductors helps to confine electrons and holes to the middle layer, also called the active layer since light is generated inside it as a result of electron–hole recombination. However, the active layer also has a slightly larger refractive index than the surrounding $p$-type and $n$-type cladding layers simply because its bandgap is smaller. As a result of the refractive-index difference, the active layer acts as a dielectric waveguide and supports optical modes whose number can be controlled by changing the active-layer thickness (similar to the modes supported by a fiber core). The main point is that a heterostructure confines the generated light to the active layer because of its higher refractive index. Figure 3.4 illustrates schematically the simultaneous confinement of charge carriers and the optical field to the active region through a heterostructure design. It is this feature that has made semiconductor lasers practical for a wide variety of applications.

Figure 3.3: Energy-band diagram of (a) homostructure and (b) double-heterostructure $p$–$n$ junctions in thermal equilibrium (top) and under forward bias (bottom).
3.1.3 Nonradiative Recombination

When a $p$–$n$ junction is forward-biased, electrons and holes are injected into the active region, where they recombine to produce light. In any semiconductor, electrons and holes can also recombine nonradiatively. Nonradiative recombination mechanisms include recombination at traps or defects, surface recombination, and the Auger recombination [5]. The last mechanism is especially important for semiconductor lasers emitting light in the wavelength range 1.3–1.6 $\mu$m because of a relatively small bandgap of the active layer [2]. In the Auger recombination process, the energy released during electron–hole recombination is given to another electron or hole as kinetic energy rather than producing light.

From the standpoint of device operation, all nonradiative processes are harmful, as they reduce the number of electron–hole pairs that emit light. Their effect is quantified through the internal quantum efficiency, defined as

$$\eta_{\text{int}} = \frac{R_{\text{rr}}}{R_{\text{tot}}} = \frac{R_{\text{rr}}}{R_{\text{rr}} + R_{\text{nr}}}$$  \hspace{1cm} (3.1.16)$$

where $R_{\text{rr}}$ is the radiative recombination rate, $R_{\text{nr}}$ is the nonradiative recombination rate.
rate, and $R_{\text{tot}} = R_{\text{rr}} + R_{\text{nr}}$ is the total recombination rate. It is customary to introduce the recombination times $\tau_{\text{rr}}$ and $\tau_{\text{nr}}$ using $R_{\text{rr}} = N/\tau_{\text{rr}}$ and $R_{\text{nr}} = N/\tau_{\text{nr}}$, where $N$ is the carrier density. The internal quantum efficiency is then given by

$$\eta_{\text{int}} = \frac{\tau_{\text{nr}}}{\tau_{\text{rr}} + \tau_{\text{nr}}}.$$  

(3.1.17)

The radiative and nonradiative recombination times vary from semiconductor to semiconductor. In general, $\tau_{\text{rr}}$ and $\tau_{\text{nr}}$ are comparable for direct-bandgap semiconductors, whereas $\tau_{\text{nr}}$ is a small fraction ($\sim 10^{-5}$) of $\tau_{\text{rr}}$ for semiconductors with an indirect bandgap. A semiconductor is said to have a direct bandgap if the conduction-band minimum and the valence-band maximum occur for the same value of the electron wave vector (see Fig. 3.2). The probability of radiative recombination is large in such semiconductors, since it is easy to conserve both energy and momentum during electron–hole recombination. By contrast, indirect-bandgap semiconductors require the assistance of a phonon for conserving momentum during electron–hole recombination. This feature reduces the probability of radiative recombination and increases $\tau_{\text{rr}}$ considerably compared with $\tau_{\text{nr}}$ in such semiconductors. As evident from Eq. (3.1.17), $\eta_{\text{int}} \ll 1$ under such conditions. Typically, $\eta_{\text{int}} \sim 10^{-5}$ for Si and Ge, the two semiconductors commonly used for electronic devices. Both are not suitable for optical sources because of their indirect bandgap. For direct-bandgap semiconductors such as GaAs and InP, $\eta_{\text{int}} \approx 0.5$ and approaches 1 when stimulated emission dominates.

The radiative recombination rate can be written as $R_{\text{rr}} = R_{\text{spon}} + R_{\text{stim}}$ when radiative recombination occurs through spontaneous as well as stimulated emission. For LEDs, $R_{\text{stim}}$ is negligible compared with $R_{\text{spon}}$, and $R_{\text{rr}}$ in Eq. (3.1.16) is replaced with $R_{\text{spon}}$. Typically, $R_{\text{spon}}$ and $R_{\text{nr}}$ are comparable in magnitude, resulting in an internal quantum efficiency of about 50%. However, $\eta_{\text{int}}$ approaches 100% for semiconductor lasers as stimulated emission begins to dominate with an increase in the output power.

It is useful to define a quantity known as the carrier lifetime $\tau_{\text{c}}$ such that it represents the total recombination time of charged carriers in the absence of stimulated recombination. It is defined by the relation

$$R_{\text{spon}} + R_{\text{nr}} = N/\tau_{\text{c}},$$  

(3.1.18)

where $N$ is the carrier density. If $R_{\text{spon}}$ and $R_{\text{nr}}$ vary linearly with $N$, $\tau_{\text{c}}$ becomes a constant. In practice, both of them increase nonlinearly with $N$ such that $R_{\text{spon}} + R_{\text{nr}} = A_{\text{nr}}N + BN^2 + CN^3$, where $A_{\text{nr}}$ is the nonradiative coefficient due to recombination at defects or traps, $B$ is the spontaneous radiative recombination coefficient, and $C$ is the Auger coefficient. The carrier lifetime then becomes $N$ dependent and is obtained by using $\tau_{\text{c}}^{-1} = A_{\text{nr}} + BN + CN^2$. In spite of its $N$ dependence, the concept of carrier lifetime $\tau_{\text{c}}$ is quite useful in practice.

### 3.1.4 Semiconductor Materials

Almost any semiconductor with a direct bandgap can be used to make a $p$–$n$ homojunction capable of emitting light through spontaneous emission. The choice is, however, considerably limited in the case of heterostructure devices because their performance
Figure 3.5: Lattice constants and bandgap energies of ternary and quaternary compounds formed by using nine group III–V semiconductors. Shaded area corresponds to possible InGaAsP and AlGaAs structures. Horizontal lines passing through InP and GaAs show the lattice-matched designs. (After Ref. [18]; ©1991 Wiley; reprinted with permission.)

depends on the quality of the heterojunction interface between two semiconductors of different bandgaps. To reduce the formation of lattice defects, the lattice constant of the two materials should match to better than 0.1%. Nature does not provide semiconductors whose lattice constants match to such precision. However, they can be fabricated artificially by forming ternary and quaternary compounds in which a fraction of the lattice sites in a naturally occurring binary semiconductor (e.g., GaAs) is replaced by other elements. In the case of GaAs, a ternary compound Al$_x$Ga$_{1-x}$As can be made by replacing a fraction $x$ of Ga atoms by Al atoms. The resulting semiconductor has nearly the same lattice constant, but its bandgap increases. The bandgap depends on the fraction $x$ and can be approximated by a simple linear relation [2]

$$E_g(x) = 1.424 + 1.247x \quad (0 < x < 0.45),$$

where $E_g$ is expressed in electron-volt (eV) units.

Figure 3.5 shows the interrelationship between the bandgap $E_g$ and the lattice constant $a$ for several ternary and quaternary compounds. Solid dots represent the binary semiconductors, and lines connecting them corresponds to ternary compounds. The dashed portion of the line indicates that the resulting ternary compound has an indirect bandgap. The area of a closed polygon corresponds to quaternary compounds. The
The shaded area in Fig. 3.5 represents the ternary and quaternary compounds with a direct bandgap formed by using the elements indium (In), gallium (Ga), arsenic (As), and phosphorus (P). The horizontal line connecting GaAs and AlAs corresponds to the ternary compound Al$_x$Ga$_{1-x}$As, whose bandgap is direct for values of $x$ up to about 0.45 and is given by Eq. (3.1.19). The active and cladding layers are formed such that $x$ is larger for the cladding layers compared with the value of $x$ for the active layer. The wavelength of the emitted light is determined by the bandgap since the photon energy is approximately equal to the bandgap. By using $E_g \approx h \nu = h c / \lambda$, one finds that $\lambda \approx 0.87 \, \mu m$ for an active layer made of GaAs ($E_g = 1.424$ eV). The wavelength can be reduced to about 0.81 $\mu m$ by using an active layer with $x = 0.1$. Optical sources based on GaAs typically operate in the range 0.81–0.87 $\mu m$ and were used in the first generation of fiber-optic communication systems.

As discussed in Chapter 2, it is beneficial to operate lightwave systems in the wavelength range 1.3–1.6 $\mu m$, where both dispersion and loss of optical fibers are considerably reduced compared with the 0.85-$\mu m$ region. InP is the base material for semiconductor optical sources emitting light in this wavelength region. As seen in Fig. 3.5 by the horizontal line passing through InP, the bandgap of InP can be reduced considerably by making the quaternary compound In$_{1-x}$Ga$_x$As$_y$P$_{1-y}$ while the lattice constant remains matched to InP. The fractions $x$ and $y$ cannot be chosen arbitrarily but are related by $x/y = 0.45$ to ensure matching of the lattice constant. The bandgap of the quaternary compound can be expressed in terms of $y$ only and is well approximated by [2]

$$E_g(y) = 1.35 - 0.72y + 0.12y^2,$$

where $0 \leq y \leq 1$. The smallest bandgap occurs for $y = 1$. The corresponding ternary compound In$_{0.45}$Ga$_{0.45}$As emits light near 1.65 $\mu m$ ($E_g = 0.75$ eV). By a suitable choice of the mixing fractions $x$ and $y$, In$_{1-x}$Ga$_x$As$_y$P$_{1-y}$ sources can be designed to operate in the wide wavelength range 1.0–1.65 $\mu m$ that includes the region 1.3–1.6 $\mu m$ important for optical communication systems.

The fabrication of semiconductor optical sources requires epitaxial growth of multiple layers on a base substrate (GaAs or InP). The thickness and composition of each layer need to be controlled precisely. Several epitaxial growth techniques can be used for this purpose. The three primary techniques are known as liquid-phase epitaxy (LPE), vapor-phase epitaxy (VPE), and molecular-beam epitaxy (MBE) depending on whether the constituents of various layers are in the liquid form, vapor form, or in the form of a molecular beam. The VPE technique is also called chemical-vapor deposition. A variant of this technique is metal-organic chemical-vapor deposition (MOCVD), in which metal alkalis are used as the mixing compounds. Details of these techniques are available in the literature [2].

Both the MOCVD and MBE techniques provide an ability to control layer thickness to within 1 nm. In some lasers, the thickness of the active layer is small enough that electrons and holes act as if they are confined to a quantum well. Such confinement leads to quantization of the energy bands into subbands. The main consequence is that the joint density of states $\rho_{cv}$ acquires a staircase-like structure [5]. Such a modification of the density of states affects the gain characteristics considerably and improves
the laser performance. Such quantum-well lasers have been studied extensively [14]. Often, multiple active layers of thickness 5–10 nm, separated by transparent barrier layers of about 10 nm thickness, are used to improve the device performance. Such lasers are called multiquantum-well (MQW) lasers. Another feature that has improved the performance of MQW lasers is the introduction of intentional, but controlled strain within active layers. The use of thin active layers permits a slight mismatch between lattice constants without introducing defects. The resulting strain changes the band structure and improves the laser performance [5]. Such semiconductor lasers are called strained MQW lasers. The concept of quantum-well lasers has also been extended to make quantum-wire and quantum-dot lasers in which electrons are confined in more than one dimension [14]. However, such devices were at the research stage in 2001. Most semiconductor lasers deployed in lightwave systems use the MQW design.

### 3.2 Light-Emitting Diodes

A forward-biased $p$–$n$ junction emits light through spontaneous emission, a phenomenon referred to as electroluminescence. In its simplest form, an LED is a forward-biased $p$–$n$ homojunction. Radiative recombination of electron–hole pairs in the depletion region generates light; some of it escapes from the device and can be coupled into an optical fiber. The emitted light is incoherent with a relatively wide spectral width (30–60 nm) and a relatively large angular spread. In this section we discuss the characteristics and the design of LEDs from the standpoint of their application in optical communication systems [20].

#### 3.2.1 Power–Current Characteristics

It is easy to estimate the internal power generated by spontaneous emission. At a given current $I$ the carrier-injection rate is $I/q$. In the steady state, the rate of electron–hole pairs recombining through radiative and nonradiative processes is equal to the carrier-injection rate $I/q$. Since the internal quantum efficiency $\eta_{\text{int}}$ determines the fraction of electron–hole pairs that recombine through spontaneous emission, the rate of photon generation is simply $\eta_{\text{int}}I/q$. The internal optical power is thus given by

$$P_{\text{int}} = \eta_{\text{int}}(\hbar \omega / q)I,$$

(3.2.1)

where $\hbar \omega$ is the photon energy, assumed to be nearly the same for all photons. If $\eta_{\text{ext}}$ is the fraction of photons escaping from the device, the emitted power is given by

$$P_e = \eta_{\text{ext}} P_{\text{int}} = \eta_{\text{ext}} \eta_{\text{int}}(\hbar \omega / q)I.$$

(3.2.2)

The quantity $\eta_{\text{ext}}$ is called the external quantum efficiency. It can be calculated by taking into account internal absorption and the total internal reflection at the semiconductor–air interface. As seen in Fig. 3.6, only light emitted within a cone of angle $\theta_c$, where $\theta_c = \sin^{-1}(1/n)$ is the critical angle and $n$ is the refractive index of the semiconductor material, escapes from the LED surface. Internal absorption can be avoided by using heterostructure LEDs in which the cladding layers surrounding the
active layer are transparent to the radiation generated. The external quantum efficiency can then be written as

$$\eta_{\text{ext}} = \frac{1}{4\pi} \int_{0}^{\theta_c} T_f(\theta)(2\pi \sin \theta) d\theta,$$  \hspace{1cm} (3.2.3)

where we have assumed that the radiation is emitted uniformly in all directions over a solid angle of $4\pi$. The Fresnel transmissivity $T_f$ depends on the incidence angle $\theta$. In the case of normal incidence ($\theta = 0$), $T_f(0) = 4n/(n + 1)^2$. If we replace for simplicity $T_f(\theta)$ by $T_f(0)$ in Eq. (3.2.3), $\eta_{\text{ext}}$ is given approximately by

$$\eta_{\text{ext}} = n^{-1}(n + 1)^{-2}.$$  \hspace{1cm} (3.2.4)

By using Eq. (3.2.4) in Eq. (3.2.2) we obtain the power emitted from one facet (see Fig. 3.6). If we use $n = 3.5$ as a typical value, $\eta_{\text{ext}} = 1.4\%$, indicating that only a small fraction of the internal power becomes the useful output power. A further loss in useful power occurs when the emitted light is coupled into an optical fiber. Because of the incoherent nature of the emitted light, an LED acts as a Lambertian source with an angular distribution $S(\theta) = S_0 \cos \theta$, where $S_0$ is the intensity in the direction $\theta = 0$. The coupling efficiency for such a source [20] is $\eta_c = (\text{NA})^2$. Since the numerical aperture (NA) for optical fibers is typically in the range 0.1–0.3, only a few percent of the emitted power is coupled into the fiber. Normally, the launched power for LEDs is 100 $\mu$W or less, even though the internal power can easily exceed 10 mW.

A measure of the LED performance is the total quantum efficiency $\eta_{\text{tot}}$, defined as the ratio of the emitted optical power $P_e$ to the applied electrical power, $P_{\text{elec}} = V_0I$, where $V_0$ is the voltage drop across the device. By using Eq. (3.2.2), $\eta_{\text{tot}}$ is given by

$$\eta_{\text{tot}} = \eta_{\text{ext}} \eta_{\text{int}} (\hbar \omega/qV_0).$$  \hspace{1cm} (3.2.5)

Typically, $\hbar \omega \approx qV_0$, and $\eta_{\text{tot}} \approx \eta_{\text{ext}} \eta_{\text{int}}$. The total quantum efficiency $\eta_{\text{tot}}$, also called the power-conversion efficiency or the wall-plug efficiency, is a measure of the overall performance of the device.
Another quantity sometimes used to characterize the LED performance is the responsivity defined as the ratio $R_{\text{LED}} = P_e / I$. From Eq. (3.2.2),

$$R_{\text{LED}} = \eta_{\text{ext}} \eta_{\text{int}} \left( \frac{\bar{h} \omega}{q} \right).$$

(3.2.6)

A comparison of Eqs. (3.2.5) and (3.2.6) shows that $R_{\text{LED}} = \eta_{\text{tot}} V_0$. Typical values of $R_{\text{LED}}$ are $\sim 0.01 \text{ W/A}$. The responsivity remains constant as long as the linear relation between $P_e$ and $I$ holds. In practice, this linear relationship holds only over a limited current range [21]. Figure 3.7(a) shows the power–current ($P$–$I$) curves at several temperatures for a typical 1.3-µm LED. The responsivity of the device decreases at high currents above 80 mA because of bending of the $P$–$I$ curve. One reason for this decrease is related to the increase in the active-region temperature. The internal quantum efficiency $\eta_{\text{int}}$ is generally temperature dependent because of an increase in the nonradiative recombination rates at high temperatures.

3.2.2 LED Spectrum

As seen in Section 2.3, the spectrum of a light source affects the performance of optical communication systems through fiber dispersion. The LED spectrum is related to the spectrum of spontaneous emission, $R_{\text{spont}}(\omega)$, given in Eq. (3.1.10). In general, $R_{\text{spont}}(\omega)$ is calculated numerically and depends on many material parameters. However, an approximate expression can be obtained if $A(E_1, E_2)$ is assumed to be nonzero only over a narrow energy range in the vicinity of the photon energy, and the Fermi functions are approximated by their exponential tails under the assumption of weak
CHAPTER 3. OPTICAL TRANSMITTERS

injection [5]. The result is

\[ R_{\text{spont}}(\omega) = A_0 (\hbar \omega - E_g)^{1/2} \exp\left[ -(\hbar \omega - E_g) / k_BT \right], \]  \hspace{1cm} (3.2.7)

where \( A_0 \) is a constant and \( E_g \) is the bandgap. It is easy to deduce that \( R_{\text{spont}}(\omega) \) peaks when \( \hbar \omega = E_g + k_BT / 2 \) and has a full-width at half-maximum (FWHM) \( \Delta \nu \approx 1.8k_BT / \hbar \). At room temperature \( (T = 300 \text{ K}) \) the FWHM is about 11 THz. In practice, the spectral width is expressed in nanometers by using \( \Delta \nu = (c/\lambda^2) \Delta \lambda \) and increases as \( \lambda^2 \) with an increase in the emission wavelength \( \lambda \). As a result, \( \Delta \lambda \) is larger for InGaAsP LEDs emitting at 1.3 \( \mu \text{m} \) by about a factor of 1.7 compared with GaAs LEDs.

Figure 3.7(b) shows the output spectrum of a typical 1.3-\( \mu \text{m} \) LED and compares it with the theoretical curve obtained by using Eq. (3.2.7). Because of a large spectral width (\( \Delta \lambda = 50–60 \text{ nm} \)), the bit rate–distance product is limited considerably by fiber dispersion when LEDs are used in optical communication systems. LEDs are suitable primarily for local-area-network applications with bit rates of 10–100 Mb/s and transmission distances of a few kilometers.

3.2.3 Modulation Response

The modulation response of LEDs depends on carrier dynamics and is limited by the carrier lifetime \( \tau_c \) defined by Eq. (3.1.18). It can be determined by using a rate equation for the carrier density \( N \). Since electrons and holes are injected in pairs and recombine in pairs, it is enough to consider the rate equation for only one type of charge carrier. The rate equation should include all mechanisms through which electrons appear and disappear inside the active region. For LEDs it takes the simple form (since stimulated emission is negligible)

\[ \frac{dN}{dt} = \frac{I}{qV} - \frac{N}{\tau_c}, \]  \hspace{1cm} (3.2.8)

where the last term includes both radiative and nonradiative recombination processes through the carrier lifetime \( \tau_c \). Consider sinusoidal modulation of the injected current in the form (the use of complex notation simplifies the math)

\[ I(t) = I_b + I_m \exp(i \omega_m t), \]  \hspace{1cm} (3.2.9)

where \( I_b \) is the bias current, \( I_m \) is the modulation current, and \( \omega_m \) is the modulation frequency. Since Eq. (3.2.8) is linear, its general solution can be written as

\[ N(t) = N_b + N_m \exp(i \omega_m t), \]  \hspace{1cm} (3.2.10)

where \( N_b = \tau_c I_b / qV \), \( V \) is the volume of active region and \( N_m \) is given by

\[ N_m(\omega_m) = \frac{\tau_c I_m / qV}{1 + i \omega_m \tau_c}. \]  \hspace{1cm} (3.2.11)

The modulated power \( P_m \) is related to \( |N_m| \) linearly. One can define the LED transfer function \( H(\omega_m) \) as

\[ H(\omega_m) = \frac{N_m(\omega_m)}{N_m(0)} = \frac{1}{1 + i \omega_m \tau_c}. \]  \hspace{1cm} (3.2.12)
3.2. LIGHT-EMITTING DIODES

Figure 3.8: Schematic of a surface-emitting LED with a double-heterostructure geometry.

In analogy with the case of optical fibers (see Section 2.4.4), the 3-dB modulation bandwidth $f_{3\text{dB}}$ is defined as the modulation frequency at which $|H(\omega_m)|$ is reduced by 3 dB or by a factor of 2. The result is

$$f_{3\text{dB}} = \sqrt{3} (2\pi \tau_c)^{-1}. \quad (3.2.13)$$

Typically, $\tau_c$ is in the range 2–5 ns for InGaAsP LEDs. The corresponding LED modulation bandwidth is in the range 50–140 MHz. Note that Eq. (3.2.13) provides the optical bandwidth because $f_{3\text{dB}}$ is defined as the frequency at which optical power is reduced by 3 dB. The corresponding electrical bandwidth is the frequency at which $|H(\omega_m)|^2$ is reduced by 3 dB and is given by $(2\pi \tau_c)^{-1}$.

3.2.4 LED Structures

The LED structures can be classified as surface-emitting or edge-emitting, depending on whether the LED emits light from a surface that is parallel to the junction plane or from the edge of the junction region. Both types can be made using either a $p$–$n$ homojunction or a heterostructure design in which the active region is surrounded by $p$- and $n$-type cladding layers. The heterostructure design leads to superior performance, as it provides a control over the emissive area and eliminates internal absorption because of the transparent cladding layers.

Figure 3.8 shows schematically a surface-emitting LED design referred to as the Burrus-type LED [22]. The emissive area of the device is limited to a small region whose lateral dimension is comparable to the fiber-core diameter. The use of a gold stud avoids power loss from the back surface. The coupling efficiency is improved by
etching a well and bringing the fiber close to the emissive area. The power coupled into the fiber depends on many parameters, such as the numerical aperture of the fiber and the distance between fiber and LED. The addition of epoxy in the etched well tends to increase the external quantum efficiency as it reduces the refractive-index mismatch. Several variations of the basic design exist in the literature. In one variation, a truncated spherical microlens fabricated inside the etched well is used to couple light into the fiber [23]. In another variation, the fiber end is itself formed in the form of a spherical lens [24]. With a proper design, surface-emitting LEDs can couple up to 1% of the internally generated power into an optical fiber.

The edge-emitting LEDs employ a design commonly used for stripe-geometry semiconductor lasers (see Section 3.3.3). In fact, a semiconductor laser is converted into an LED by depositing an antireflection coating on its output facet to suppress lasing action. Beam divergence of edge-emitting LEDs differs from surface-emitting LEDs because of waveguiding in the plane perpendicular to the junction. Surface-emitting LEDs operate as a Lambertian source with angular distribution $S_e(\theta) = S_0 \cos \theta$ in both directions. The resulting beam divergence has a FWHM of 120° in each direction. In contrast, edge-emitting LEDs have a divergence of only about 30° in the direction perpendicular to the junction plane. Considerable light can be coupled into a fiber of even low numerical aperture ($< 0.3$) because of reduced divergence and high radiance at the emitting facet [25]. The modulation bandwidth of edge-emitting LEDs is generally larger ($\sim 200$ MHz) than that of surface-emitting LEDs because of a reduced carrier lifetime at the same applied current [26]. The choice between the two designs is dictated, in practice, by a compromise between cost and performance.

In spite of a relatively low output power and a low bandwidth of LEDs compared with those of lasers, LEDs are useful for low-cost applications requiring data transmission at a bit rate of 100 Mb/s or less over a few kilometers. For this reason, several new LED structures were developed during the 1990s [27]–[32]. In one design, known as resonant-cavity LED [27], two metal mirrors are fabricated around the epitaxially grown layers, and the device is bonded to a silicon substrate. In a variant of this idea, the bottom mirror is fabricated epitaxially by using a stack of alternating layers of two different semiconductors, while the top mirror consists of a deformable membrane suspended by an air gap [28]. The operating wavelength of such an LED can be tuned over 40 nm by changing the air-gap thickness. In another scheme, several quantum wells with different compositions and bandgaps are grown to form a MQW structure [29]. Since each quantum well emits light at a different wavelength, such LEDs can have an extremely broad spectrum (extending over a 500-nm wavelength range) and are useful for local-area WDM networks.

3.3 Semiconductor Lasers

Semiconductor lasers emit light through stimulated emission. As a result of the fundamental differences between spontaneous and stimulated emission, they are not only capable of emitting high powers ($\sim 100$ mW), but also have other advantages related to the coherent nature of emitted light. A relatively narrow angular spread of the output beam compared with LEDs permits high coupling efficiency ($\sim 50\%$) into single-mode
fibers. A relatively narrow spectral width of emitted light allows operation at high bit rates (~ 10 Gb/s), since fiber dispersion becomes less critical for such an optical source. Furthermore, semiconductor lasers can be modulated directly at high frequencies (up to 25 GHz) because of a short recombination time associated with stimulated emission. Most fiber-optic communication systems use semiconductor lasers as an optical source because of their superior performance compared with LEDs. In this section the output characteristics of semiconductor lasers are described from the standpoint of their applications in lightwave systems. More details can be found in Refs. [2]–[14], books devoted entirely to semiconductor lasers.

### 3.3.1 Optical Gain

As discussed in Section 3.1.1, stimulated emission can dominate only if the condition of population inversion is satisfied. For semiconductor lasers this condition is realized by doping the $p$-type and $n$-type cladding layers so heavily that the Fermi-level separation exceeds the bandgap [see Eq. (3.1.14)] under forward biasing of the $p$–$n$ junction. When the injected carrier density in the active layer exceeds a certain value, known as the transparency value, population inversion is realized and the active region exhibits optical gain. An input signal propagating inside the active layer would then amplify as $\exp(gz)$, where $g$ is the gain coefficient. One can calculate $g$ by noting that it is proportional to $R_{\text{stim}} - R_{\text{abs}}$, where $R_{\text{stim}}$ and $R_{\text{abs}}$ are given by Eqs. (3.1.12) and (3.1.13), respectively. In general, $g$ is calculated numerically. Figure 3.9(a) shows the gain calculated for a 1.3-µm InGaAsP active layer at different values of the injected carrier density $N$. For $N = 1 \times 10^{18} \text{ cm}^{-3}$, $g < 0$, as population inversion has not yet occurred. As $N$ increases, $g$ becomes positive over a spectral range that increases with $N$. The peak value of the gain, $g_p$, also increases with $N$, together with a shift of the peak toward higher photon energies. The variation of $g_p$ with $N$ is shown in Fig. 3.9(b). For $N > 1.5 \times 10^{18} \text{ cm}^{-3}$, $g_p$ varies almost linearly with $N$. Figure 3.9 shows that the optical gain in semiconductors increases rapidly once population inversion is realized. It is because of such a high gain that semiconductor lasers can be made with physical dimensions of less than 1 mm.

The nearly linear dependence of $g_p$ on $N$ suggests an empirical approach in which the peak gain is approximated by

$$g_p(N) = \sigma_g(N - N_T),$$

(3.3.1)

where $N_T$ is the transparency value of the carrier density and $\sigma_g$ is the gain cross section; $\sigma_g$ is also called the differential gain. Typical values of $N_T$ and $\sigma_g$ for InGaAsP lasers are in the range 1.0–1.5×$10^{18}$ cm$^{-3}$ and 2–3×$10^{-16}$ cm$^2$, respectively [2]. As seen in Fig. 3.9(b), the approximation (3.3.1) is reasonable in the high-gain region where $g_p$ exceeds 100 cm$^{-1}$; most semiconductor lasers operate in this region. The use of Eq. (3.3.1) simplifies the analysis considerably, as band-structure details do not appear directly. The parameters $\sigma_g$ and $N_T$ can be estimated from numerical calculations such as those shown in Fig. 3.9(b) or can be measured experimentally.

Semiconductor lasers with a larger value of $\sigma_g$ generally perform better, since the same amount of gain can be realized at a lower carrier density or, equivalently, at a
lower injected current. In quantum-well semiconductor lasers, $\sigma_g$ is typically larger by about a factor of two. The linear approximation in Eq. (3.3.1) for the peak gain can still be used in a limited range. A better approximation replaces Eq. (3.3.1) with $g_p(N) = g_0 \left[1 + \ln \left(\frac{N}{N_0}\right)\right]$, where $g_p = g_0$ at $N = N_0$ and $N_0 = eN_T \approx 2.718N_T$ by using the definition $g_p = 0$ at $N = N_T$ [5].

### 3.3.2 Feedback and Laser Threshold

The optical gain alone is not enough for laser operation. The other necessary ingredient is optical feedback—it converts an amplifier into an oscillator. In most lasers the feedback is provided by placing the gain medium inside a Fabry–Perot (FP) cavity formed by using two mirrors. In the case of semiconductor lasers, external mirrors are not required as the two cleaved laser facets act as mirrors whose reflectivity is given by

$$R_m = \left(\frac{n-1}{n+1}\right)^2,$$

(3.3.2)

where $n$ is the refractive index of the gain medium. Typically, $n = 3.5$, resulting in 30% facet reflectivity. Even though the FP cavity formed by two cleaved facets is relatively lossy, the gain is large enough that high losses can be tolerated. Figure 3.10 shows the basic structure of a semiconductor laser and the FP cavity associated with it.

The concept of laser threshold can be understood by noting that a certain fraction of photons generated by stimulated emission is lost because of cavity losses and needs to be replenished on a continuous basis. If the optical gain is not large enough to compensate for the cavity losses, the photon population cannot build up. Thus, a minimum amount of gain is necessary for the operation of a laser. This amount can be realized
3.3. SEMICONDUCTOR LASERS

Figure 3.10: Structure of a semiconductor laser and the Fabry–Perot cavity associated with it. The cleaved facets act as partially reflecting mirrors.

only when the laser is pumped above a threshold level. The current needed to reach the threshold is called the \textit{threshold current}.

A simple way to obtain the threshold condition is to study how the amplitude of a plane wave changes during one round trip. Consider a plane wave of amplitude $E_0$, frequency $\omega$, and wave number $k = n\omega/c$. During one round trip, its amplitude increases by $\exp[(g/2)(2L)]$ because of gain ($g$ is the power gain) and its phase changes by $2kL$, where $L$ is the length of the laser cavity. At the same time, its amplitude changes by $\sqrt{R_1R_2}\exp(-\alpha_{\text{int}}L)$ because of reflection at the laser facets and because of an internal loss $\alpha_{\text{int}}$ that includes free-carrier absorption, scattering, and other possible mechanisms. Here $R_1$ and $R_2$ are the reflectivities of the laser facets. Even though $R_1 = R_2$ in most cases, the two reflectivities can be different if laser facets are coated to change their natural reflectivity. In the steady state, the plane wave should remain unchanged after one round trip, i.e.,

$$E_0\exp(gL)\sqrt{R_1R_2}\exp(-\alpha_{\text{int}}L)\exp(2ikL) = E_0. \quad (3.3.3)$$

By equating the amplitude and the phase on two sides, we obtain

$$g = \alpha_{\text{int}} + \frac{1}{2L}\ln\left(\frac{1}{R_1R_2}\right) = \alpha_{\text{int}} + \alpha_{\text{mir}} = \alpha_{\text{cav}}, \quad (3.3.4)$$

$$2kL = 2m\pi \quad \text{or} \quad \nu = \nu_m = mc/(2nL), \quad (3.3.5)$$

where $k = 2\pi\nu c$ and $m$ is an integer. Equation (3.3.4) shows that the gain $g$ equals total cavity loss $\alpha_{\text{cav}}$ at threshold and beyond. It is important to note that $g$ is not the same as the material gain $g_m$ shown in Fig. 3.9. As discussed in Section 3.3.3, the
optical mode extends beyond the active layer while the gain exists only inside it. As a result, \( g = \Gamma g_m \), where \( \Gamma \) is the confinement factor of the active region with typical values <0.4.

The phase condition in Eq. (3.3.5) shows that the laser frequency \( \nu \) must match one of the frequencies in the set \( \nu_m \), where \( m \) is an integer. These frequencies correspond to the longitudinal modes and are determined by the optical length \( nL \). The spacing \( \Delta \nu_L \) between the longitudinal modes is constant \( (\Delta \nu_L = c/(2nL)) \) if the frequency dependence of \( n \) is ignored. It is given by \( \Delta \nu_L = c/(2n_L) \) when material dispersion is included [2]. Here the group index \( n_g \) is defined as \( n_g = n + \omega (dn/d\omega) \). Typically, \( \Delta \nu_L = 100–200 \) GHz for \( L = 200–400 \) μm.

A FP semiconductor laser generally emits light in several longitudinal modes of the cavity. As seen in Fig. 3.11, the gain spectrum \( g(\omega) \) of semiconductor lasers is wide enough (bandwidth \( \sim 10 \) THz) that many longitudinal modes of the FP cavity experience gain simultaneously. The mode closest to the gain peak becomes the dominant mode. Under ideal conditions, the other modes should not reach threshold since their gain always remains less than that of the main mode. In practice, the difference is extremely small (\( \sim 0.1 \) cm\(^{-1}\)) and one or two neighboring modes on each side of the main mode carry a significant portion of the laser power together with the main mode. Such lasers are called multimode semiconductor lasers. Since each mode propagates inside the fiber at a slightly different speed because of group-velocity dispersion, the multimode nature of semiconductor lasers limits the bit-rate–distance product \( BL \) to values below 10 (Gbps)-km for systems operating near 1.55 μm (see Fig. 2.13). The \( BL \) product can be increased by designing lasers oscillating in a single longitudinal mode. Such lasers are discussed in Section 3.4.

### 3.3.3 Laser Structures

The simplest structure of a semiconductor laser consists of a thin active layer (thickness \( \sim 0.1 \) μm) sandwiched between \( p \)-type and \( n \)-type cladding layers of another semi-
3.3. SEMICONDUCTOR LASERS

Figure 3.12: A broad-area semiconductor laser. The active layer (hatched region) is sandwiched between $p$-type and $n$-type cladding layers of a higher-bandgap material.

Figure 3.12 shows such a structure. The laser light is emitted from the two cleaved facets in the form of an elliptical spot of dimensions $\sim 1 \times 100 \, \mu m^2$. In the direction perpendicular to the junction plane, the spot size is $\sim 1 \, \mu m$ because of the heterostructure design of the laser. As discussed in Section 3.1.2, the active layer acts as a planar waveguide because its refractive index is larger than that of the surrounding cladding layers ($\Delta n \approx 0.3$). Similar to the case of optical fibers, it supports a certain number of modes, known as the transverse modes. In practice, the active layer is thin enough ($\sim 0.1 \, \mu m$) that the planar waveguide supports a single transverse mode. However, there is no such light-confinement mechanism in the lateral direction parallel to the junction plane. Consequently, the light generated spreads over the entire width of the laser. Broad-area semiconductor lasers suffer from a number of deficiencies and are rarely used in optical communication systems. The major drawbacks are a relatively high threshold current and a spatial pattern that is highly elliptical and that changes in an uncontrollable manner with the current. These problems can be solved by introducing a mechanism for light confinement in the lateral direction. The resulting semiconductor lasers are classified into two broad categories.

Gain-guided semiconductor lasers solve the light-confinement problem by limiting current injection over a narrow stripe. Such lasers are also called stripe-geometry semiconductor lasers. Figure 3.13 shows two laser structures schematically. In one approach, a dielectric ($SiO_2$) layer is deposited on top of the $p$-layer with a central opening through which the current is injected [33]. In another, an $n$-type layer is deposited on top of the $p$-layer [34]. Diffusion of Zn over the central region converts the $n$-region into $p$-type. Current flows only through the central region and is blocked elsewhere because of the reverse-biased nature of the $p-n$ junction. Many other variations exist [2]. In all designs, current injection over a narrow central stripe ($\sim 5 \, \mu m$ width) leads to a spatially varying distribution of the carrier density (governed by car-
Figure 3.13: Cross section of two stripe-geometry laser structures used to design gain-guided semiconductor lasers and referred to as (a) oxide stripe and (b) junction stripe.

rier diffusion) in the lateral direction. The optical gain also peaks at the center of the stripe. Since the active layer exhibits large absorption losses in the region beyond the central stripe, light is confined to the stripe region. As the confinement of light is aided by gain, such lasers are called gain-guided. Their threshold current is typically in the range 50–100 mA, and light is emitted in the form of an elliptic spot of dimensions $\sim 1 \times 5 \, \mu m^2$. The major drawback is that the spot size is not stable as the laser power is increased [2]. Such lasers are rarely used in optical communication systems because of mode-stability problems.

The light-confinement problem is solved in the index-guided semiconductor lasers by introducing an index step $\Delta n_L$ in the lateral direction so that waveguides are formed in a way similar to the waveguide formed in the transverse direction by the heterostructure design. Such lasers can be subclassified as weakly and strongly index-guided semiconductor lasers, depending on the magnitude of $\Delta n_L$. Figure 3.14 shows examples of the two kinds of lasers. In a specific design known as the ridge-waveguide laser, a ridge is formed by etching parts of the $p$-layer [2]. A SiO$_2$ layer is then deposited to block the current flow and to induce weak index guiding. Since the refractive index of SiO$_2$ is considerably lower than the central $p$-region, the effective index of the transverse mode is different in the two regions [35], resulting in an index step $\Delta n_L \sim 0.01$. This index step confines the generated light to the ridge region. The magnitude of the index step is sensitive to many fabrication details, such as the ridge width and the proximity of the SiO$_2$ layer to the active layer. However, the relative simplicity of the ridge-waveguide design and the resulting low cost make such lasers attractive for some applications.

In strongly index-guided semiconductor lasers, the active region of dimensions $\sim 0.1 \times 1 \, \mu m^2$ is buried on all sides by several layers of lower refractive index. For this reason, such lasers are called buried heterostructure (BH) lasers. Several different kinds of BH lasers have been developed. They are known under names such as etched-mesa BH, planar BH, double-channel planar BH, and V-grooved or channelled substrate BH lasers, depending on the fabrication method used to realize the laser structure [2]. They all allow a relatively large index step ($\Delta n_L \sim 0.1$) in the lateral direction and, as
3.4. CONTROL OF LONGITUDINAL MODES

We have seen that BH semiconductor lasers can be designed to emit light into a single spatial mode by controlling the width and the thickness of the active layer. However, as discussed in Section 3.3.2, such lasers oscillate in several longitudinal modes simultaneously because of a relatively small gain difference ($\sim 0.1 \, \text{cm}^{-1}$) between neighboring modes of the FP cavity. The resulting spectral width (2–4 nm) is acceptable for lightwave systems operating near 1.3 $\mu$m at bit rates of up to 1 Gb/s. However, such multimode lasers cannot be used for systems designed to operate near 1.55 $\mu$m at high bit rates. The only solution is to design semiconductor lasers [36]–[41] such that they emit light predominantly in a single longitudinal mode (SLM).

The SLM semiconductor lasers are designed such that cavity losses are different for different longitudinal modes of the cavity, in contrast with FP lasers whose losses are mode independent. Figure 3.15 shows the gain and loss profiles schematically for such a laser. The longitudinal mode with the smallest cavity loss reaches threshold first as a result, permit strong mode confinement. Because of a large built-in index step, the spatial distribution of the emitted light is inherently stable, provided that the laser is designed to support a single spatial mode.

As the active region of a BH laser is in the form of a rectangular waveguide, spatial modes can be obtained by following a method similar to that used in Section 2.2 for optical fibers [2]. In practice, a BH laser operates in a single mode if the active-region width is reduced to below 2 $\mu$m. The spot size is elliptical with typical dimensions $2 \times 1 \, \mu$m$^2$. Because of small spot-size dimensions, the beam diffracts widely in both the lateral and transverse directions. The elliptic spot size and a large divergence angle make it difficult to couple light into the fiber efficiently. Typical coupling efficiencies are in the range 30–50% for most optical transmitters. A spot-size converter is sometimes used to improve the coupling efficiency (see Section 3.6).

Figure 3.14: Cross section of two index-guided semiconductor lasers: (a) ridge-waveguide structure for weak index guiding; (b) etched-mesa buried heterostructure for strong index guiding.
Figure 3.15: Gain and loss profiles for semiconductor lasers oscillating predominantly in a single longitudinal mode.

and becomes the dominant mode. Other neighboring modes are discriminated by their higher losses, which prevent their buildup from spontaneous emission. The power carried by these side modes is usually a small fraction (\(< 1\%) of the total emitted power. The performance of a SLM laser is often characterized by the mode-suppression ratio (MSR), defined as [39]

\[
MSR = \frac{P_{\text{mm}}}{P_{\text{sm}}},
\]

where \(P_{\text{mm}}\) is the main-mode power and \(P_{\text{sm}}\) is the power of the most dominant side mode. The MSR should exceed 1000 (or 30 dB) for a good SLM laser.

3.4.1 Distributed Feedback Lasers

Distributed feedback (DFB) semiconductor lasers were developed during the 1980s and are used routinely for WDM lightwave systems [10]–[12]. The feedback in DFB lasers, as the name implies, is not localized at the facets but is distributed throughout the cavity length [41]. This is achieved through an internal built-in grating that leads to a periodic variation of the mode index. Feedback occurs by means of Bragg diffraction, a phenomenon that couples the waves propagating in the forward and backward directions. Mode selectivity of the DFB mechanism results from the Bragg condition: the coupling occurs only for wavelengths \(\lambda_B\) satisfying

\[
\Lambda = m(\frac{\lambda_B}{2n}),
\]

where \(\Lambda\) is the grating period, \(n\) is the average mode index, and the integer \(m\) represents the order of Bragg diffraction. The coupling between the forward and backward waves is strongest for the first-order Bragg diffraction \((m = 1)\). For a DFB laser operating at \(\lambda_B = 1.55\ \mu m\), \(\Lambda\) is about 235 nm if we use \(m = 1\) and \(n = 3.3\) in Eq. (3.4.2). Such gratings can be made by using a holographic technique [2].

From the standpoint of device operation, semiconductor lasers employing the DFB mechanism can be classified into two broad categories: DFB lasers and distributed
3.4. CONTROL OF LONGITUDINAL MODES

Bragg reflector (DBR) lasers. Figure 3.16 shows two kinds of laser structures. Though the feedback occurs throughout the cavity length in DFB lasers, it does not take place inside the active region of a DBR laser. In effect, the end regions of a DBR laser act as mirrors whose reflectivity is maximum for a wavelength $\lambda_B$ satisfying Eq. (3.4.2). The cavity losses are therefore minimum for the longitudinal mode closest to $\lambda_B$ and increase substantially for other longitudinal modes (see Fig. 3.15). The MSR is determined by the gain margin defined as the excess gain required by the most dominant side mode to reach threshold. A gain margin of 3–5 cm$^{-1}$ is generally enough to realize an MSR > 30 dB for DFB lasers operating continuously [39]. However, a larger gain margin is needed (> 10 cm$^{-1}$) when DFB lasers are modulated directly. Phase-shifted DFB lasers [38], in which the grating is shifted by $\lambda_B/4$ in the middle of the laser to produce a $\pi/2$ phase shift, are often used, since they are capable of providing much larger gain margin than that of conventional DFB lasers. Another design that has led to improvements in the device performance is known as the gain-coupled DFB laser [42]–[44]. In these lasers, both the optical gain and the mode index vary periodically along the cavity length.

Fabrication of DFB semiconductor lasers requires advanced technology with multiple epitaxial growths [41]. The principal difference from FP lasers is that a grating is etched onto one of the cladding layers surrounding the active layer. A thin $n$-type waveguide layer with a refractive index intermediate to that of active layer and the substrate acts as a grating. The periodic variation of the thickness of the waveguide layer translates into a periodic variation of the mode index $\bar{n}$ along the cavity length and leads to a coupling between the forward and backward propagating waves through Bragg diffraction.
Figure 3.17: Longitudinal-mode selectivity in a coupled-cavity laser. Phase shift in the external cavity makes the effective mirror reflectivity wavelength dependent and results in a periodic loss profile for the laser cavity.

A holographic technique is often used to form a grating with a $\sim 0.2$-$\mu$m periodicity. It works by forming a fringe pattern on a photoresist (deposited on the wafer surface) through interference between two optical beams. In the alternative electron-beam lithographic technique, an electron beam writes the desired pattern on the electron-beam resist. Both methods use chemical etching to form grating corrugations, with the patterned resist acting as a mask. Once the grating has been etched onto the substrate, multiple layers are grown by using an epitaxial growth technique. A second epitaxial regrowth is needed to make a BH device such as that shown in Fig. 3.14(b). Despite the technological complexities, DFB lasers are routinely produced commercially. They are used in nearly all 1.55-$\mu$m optical communication systems operating at bit rates of 2.5 Gb/s or more. DFB lasers are reliable enough that they have been used since 1992 in all transoceanic lightwave systems.

### 3.4.2 Coupled-Cavity Semiconductor Lasers

In a coupled-cavity semiconductor laser [2], the SLM operation is realized by coupling the light to an external cavity (see Fig. 3.17). A portion of the reflected light is fed back into the laser cavity. The feedback from the external cavity is not necessarily in
phase with the optical field inside the laser cavity because of the phase shift occurring in the external cavity. The in-phase feedback occurs only for those laser modes whose wavelength nearly coincides with one of the longitudinal modes of the external cavity. In effect, the effective reflectivity of the laser facet facing the external cavity becomes wavelength dependent and leads to the loss profile shown in Fig. 3.17. The longitudinal mode that is closest to the gain peak and has the lowest cavity loss becomes the dominant mode.

Several kinds of coupled-cavity schemes have been developed for making SLM laser; Fig. 3.18 shows three among them. A simple scheme couples the light from a semiconductor laser to an external grating [Fig. 3.18(a)]. It is necessary to reduce the natural reflectivity of the cleaved facet facing the grating through an antireflection coating to provide a strong coupling. Such lasers are called external-cavity semiconductor lasers and have attracted considerable attention because of their tunability [36]. The wavelength of the SLM selected by the coupled-cavity mechanism can be tuned over a wide range (typically 50 nm) simply by rotating the grating. Wavelength tunability is a desirable feature for lasers used in WDM lightwave systems. A drawback of the laser shown in Fig. 3.18(a) from the system standpoint is its nonmonolithic nature, which makes it difficult to realize the mechanical stability required of optical transmitters.

A monolithic design for coupled-cavity lasers is offered by the cleaved-coupled-cavity laser [37] shown in Fig. 3.18(b). Such lasers are made by cleaving a conventional multimode semiconductor laser in the middle so that the laser is divided into two sections of about the same length but separated by a narrow air gap (width $\sim 1 \mu m$). The reflectivity of cleaved facets ($\sim 30\%$) allows enough coupling between the two sections as long as the gap is not too wide. It is even possible to tune the wavelength of such a laser over a tuning range $\sim 20$ nm by varying the current injected into one of the cavity sections acting as a mode controller. However, tuning is not continuous, since it corresponds to successive mode hops of about 2 nm.

### 3.4.3 Tunable Semiconductor Lasers

Modern WDM lightwave systems require single-mode, narrow-linewidth lasers whose wavelength remains fixed over time. DFB lasers satisfy this requirement but their wavelength stability comes at the expense of tunability [9]. The large number of DFB lasers used inside a WDM transmitter make the design and maintenance of such a lightwave system expensive and impractical. The availability of semiconductor lasers whose wavelength can be tuned over a wide range would solve this problem [13].

Multisection DFB and DBR lasers were developed during the 1990s to meet the somewhat conflicting requirements of stability and tunability [45]–[52] and were reaching the commercial stage in 2001. Figure 3.18(c) shows a typical laser structure. It consists of three sections, referred to as the active section, the phase-control section, and the Bragg section. Each section can be biased independently by injecting different amounts of currents. The current injected into the Bragg section is used to change the Bragg wavelength ($\lambda_B = 2n\Lambda$) through carrier-induced changes in the refractive index $n$. The current injected into the phase-control section is used to change the phase of the feedback from the DBR through carrier-induced index changes in that section. The laser wavelength can be tuned almost continuously over the range 10–15 nm by con-
Figure 3.18: Coupled-cavity laser structures: (a) external-cavity laser; (b) cleaved-coupled-cavity laser; (c) multisection DBR laser.

trolling the currents in the phase and Bragg sections. By 1997, such lasers exhibited a tuning range of 17 nm and output powers of up to 100 mW with high reliability [51].

Several other designs of tunable DFB lasers have been developed in recent years. In one scheme, the built-in grating inside a DBR laser is chirped by varying the grating period $\Lambda$ or the mode index $\bar{n}$ along the cavity length. As seen from Eq. (3.4.2), the Bragg wavelength itself then changes along the cavity length. Since the laser wavelength is determined by the Bragg condition, such a laser can be tuned over a wavelength range determined by the grating chirp. In a simple implementation of the basic idea, the grating period remains uniform, but the waveguide is bent to change the effective mode index $\bar{n}$. Such multisection DFB lasers can be tuned over 5–6 nm while maintaining a single longitudinal mode with high side-mode suppression [47].

In another scheme, a superstructure grating is used for the DBR section of a multisection laser [48]–[50]. A superstructure grating consists of an array of gratings (uniform or chirped) separated by a constant distance. As a result, its reflectivity peaks at several wavelengths whose interval is determined by the spacing between the individual gratings forming the array. Such multisection DBR lasers can be tuned discretely
3.4. CONTROL OF LONGITUDINAL MODES

over a wavelength range exceeding 100 nm. By controlling the current in the phase-control section, a quasi-continuous tuning range of 40 nm was realized in 1995 with a superstructure grating [48]. The tuning range can be extended considerably by using a four-section device in which another DBR section is added to the left side of the device shown in Fig. 3.18(c). Each DBR section supports its own comb of wavelengths but the spacing in each comb is not the same. The coinciding wavelength in the two combs becomes the output wavelength that can be tuned over a wide range (analogous to the Vernier effect).

In a related approach, the fourth section in Fig. 3.18(c) is added between the gain and phase sections: It consist of a grating-assisted codirectional coupler with a superstructure grating. The coupler has two vertically separated waveguides and selects a single wavelength from the wavelength comb supported by the DBR section with a superstructure grating. The largest tuning range of 114 nm was produced in 1995 by this kind of device [49]. Such widely tunable DBR lasers are likely to find applications in many WDM lightwave systems.

3.4.4 Vertical-Cavity Surface-Emitting Lasers

A new class of semiconductor lasers, known as vertical-cavity surface-emitting lasers (VCSELs), has emerged during the 1990s with many potential applications [53]–[60]. VCSELs operate in a single longitudinal mode by virtue of an extremely small cavity length (∼ 1 µm), for which the mode spacing exceeds the gain bandwidth (see Fig. 3.11). They emit light in a direction normal to the active-layer plane in a manner analogous to that of a surface-emitting LED (see Fig. 3.8). Moreover, the emitted light is in the form of a circular beam that can be coupled into a single-mode fiber with high efficiency. These properties result in a number of advantages that are leading to rapid adoption of VCSELs for lightwave communications.

As seen in Fig. 3.19, fabrication of VCSELs requires growth of multiple thin layers on a substrate. The active region, in the form of one or several quantum wells, is surrounded by two high-reflectivity (> 99.5%) DBR mirrors that are grown epitaxially on both sides of the active region to form a high-Q microcavity [55]. Each DBR mirror is made by growing many pairs of alternating GaAs and AlAs layers, each λ/4 thick, where λ is the wavelength emitted by the VCSEL. A wafer-bonding technique is sometimes used for VCSELs operating in the 1.55-µm wavelength region to accommodate the InGaAsP active region [58]. Chemical etching or a related technique is used to form individual circular disks (each corresponding to one VCSEL) whose diameter can be varied over a wide range (typically 5–20 µm). The entire two-dimensional array of VCSELs can be tested without requiring separation of lasers because of the vertical nature of light emission. As a result, the cost of a VCSEL can be much lower than that of an edge-emitting laser. VCSELs also exhibit a relatively low threshold (∼ 1 mA or less). Their only disadvantage is that they cannot emit more than a few milliwatts of power because of a small active volume. For this reason, they are mostly used in local-area and metropolitan-area networks and have virtually replaced LEDs. Early VCSELs were designed to emit near 0.8 µm and operated in multiple transverse modes because of their relatively large diameters (∼ 10 µm).
In recent years, the VCSEL technology have advanced enough that VCSELS can be designed to operate in a wide wavelength range extending from 650 to 1600 nm [55]. Their applications in the 1.3- and 1.55-µm wavelength windows require that VCSELs operate in a single transverse mode. By 2001, several techniques had emerged for controlling the transverse modes of a VCSEL, the most common being the oxide-confinement technique in which an insulating aluminum-oxide layer, acting as a dielectric aperture, confines both the current and the optical mode to a < 3-µm-diameter region. Such VCSELs operate in a single mode with narrow linewidth and can replace a DFB laser in many lightwave applications as long as their low output power is acceptable. They are especially useful for data transfer and local-loop applications because of their low-cost packaging. VCSELS are also well suited for WDM applications for two reasons. First, their wavelengths can be tuned over a wide range (> 50 nm) using the micro-electro-mechanical system (MEMS) technology [56]. Second, one can make two-dimensional VCSEL arrays such that each laser operates at a different wavelength [60]. WDM sources, containing multiple monolithically integrated lasers, are required for modern lightwave systems.

3.5 Laser Characteristics

The operating characteristics of semiconductor lasers are well described by a set of rate equations that govern the interaction of photons and electrons inside the active region. In this section we use the rate equations to discuss first both the continuous-wave (CW) properties. We then consider small- and large-signal modulation characteristics of single-mode semiconductor lasers. The last two subsections focus on the intensity noise and spectral bandwidth of semiconductor lasers.
3.5. LASER CHARACTERISTICS

3.5.1 CW Characteristics

A rigorous derivation of the rate equations generally starts from Maxwell’s equations together with a quantum-mechanical approach for the induced polarization (see Section 2.2). The rate equations can also be written heuristically by considering various physical phenomena through which the number of photons, $P$, and the number of electrons, $N$, change with time inside the active region. For a single-mode laser, these equations take the form [2]

\[
\frac{dP}{dt} = GP + R_{sp} - \frac{P}{\tau_p},
\]

(3.5.1)

\[
\frac{dN}{dt} = \frac{I}{q} - \frac{N}{\tau_c} - GP,
\]

(3.5.2)

where

\[ G = \Gamma v_g g_m = G_N(N - N_0). \]

(3.5.3)

$G$ is the net rate of stimulated emission and $R_{sp}$ is the rate of spontaneous emission into the lasing mode. Note that $R_{sp}$ is much smaller than the total spontaneous-emission rate in Eq. (3.1.10). The reason is that spontaneous emission occurs in all directions over a wide spectral range ($\sim 30–40$ nm) but only a small fraction of it, propagating along the cavity axis and emitted at the laser frequency, actually contributes to Eq. (3.5.1). In fact, $R_{sp}$ and $G$ are related by $R_{sp} = n_{sp}G$, where $n_{sp}$ is known as the spontaneous-emission factor and is about 2 for semiconductor lasers [2]. Although the same notation is used for convenience, the variable $N$ in the rate equations represents the number of electrons rather than the carrier density; the two are related by the active volume $V$. In Eq. (3.5.3), $v_g$ is the group velocity, $\Gamma$ is the confinement factor, and $g_m$ is the material gain at the mode frequency. By using Eq. (3.3.1), $G$ varies linearly with $N$ with $G_N = \Gamma v_g \sigma_g / V$ and $N_0 = N_T V$.

The last term in Eq. (3.5.1) takes into account the loss of photons inside the cavity. The parameter $\tau_p$ is referred to as the photon lifetime. It is related to the cavity loss $\alpha_{cav}$ introduced in Eq. (3.3.4) as

\[
\tau_p^{-1} = v_g \alpha_{cav} = v_g (\alpha_{mit} + \alpha_{int}).
\]

(3.5.4)

The three terms in Eq. (3.5.2) indicate the rates at which electrons are created or destroyed inside the active region. This equation is similar to Eq. (3.2.8) except for the addition of the last term, which governs the rate of electron–hole recombination through stimulated emission. The carrier lifetime $\tau_c$ includes the loss of electrons due to both spontaneous emission and nonradiative recombination, as indicated in Eq. (3.1.18).

The $P$–$I$ curve characterizes the emission properties of a semiconductor laser, as it indicates not only the threshold level but also the current that needs to be applied to obtain a certain amount of power. Figure 3.20 shows the $P$–$I$ curves of a 1.3-µm InGaAsP laser at temperatures in the range 10–130°C. At room temperature, the threshold is reached near 20 mA, and the laser can emit 10 mW of output power from each facet at 100 mA of applied current. The laser performance degrades at high temperatures. The threshold current is found to increase exponentially with temperature, i.e.,

\[ I_{th}(T) = I_{th}(0) \exp(T/T_0), \]

(3.5.5)
where $I_0$ is a constant and $T_0$ is a characteristic temperature often used to express the temperature sensitivity of threshold current. For InGaAsP lasers $T_0$ is typically in the range 50–70 K. By contrast, $T_0$ exceeds 120 K for GaAs lasers. Because of the temperature sensitivity of InGaAsP lasers, it is often necessary to control their temperature through a built-in thermoelectric cooler.

The rate equations can be used to understand most of the features seen in Fig. 3.20. In the case of CW operation at a constant current $I$, the time derivatives in Eqs. (3.5.1) and (3.5.2) can be set to zero. The solution takes a particularly simple form if spontaneous emission is neglected by setting $R_{sp} = 0$. For currents such that $G\tau_p < 1$, $P = 0$ and $N = \tau_c I / q$. The threshold is reached at a current for which $G\tau_p = 1$. The carrier population is then clamped to the threshold value $N_{th} = N_0 + (G N \tau_p)^{-1}$. The threshold current is given by

$$I_{th} = \frac{qN_{th}}{\tau_c} = \frac{q}{\tau_c} \left( N_0 + \frac{1}{G N \tau_p} \right). \quad (3.5.6)$$

For $I > I_{th}$, the photon number $P$ increases linearly with $I$ as

$$P = \left( \tau_p / q \right)(I - I_{th}). \quad (3.5.7)$$

The emitted power $P_e$ is related to $P$ by the relation

$$P_e = \frac{1}{2} (v_g \alpha_{mir}) \hbar \omega P. \quad (3.5.8)$$

The derivation of Eq. (3.5.8) is intuitively obvious if we note that $v_g \alpha_{mir}$ is the rate at which photons of energy $\hbar \omega$ escape from the two facets. The factor of $\frac{1}{2}$ makes $P_e$...
3.5. LASER CHARACTERISTICS

the power emitted from each facet for a FP laser with equal facet reflectivities. For FP lasers with coated facets or for DFB lasers, Eq. (3.5.8) needs to be suitably modified [2]. By using Eqs. (3.5.4) and (3.5.7) in Eq. (3.5.8), the emitted power is given by

$$P_e = \frac{\hbar \omega}{2q} \frac{\eta_{int} \alpha_{mir}}{\alpha_{mir} + \alpha_{int}} (I - I_{th}),$$  \hspace{1cm} (3.5.9)

where the internal quantum efficiency $\eta_{int}$ is introduced phenomenologically to indicate the fraction of injected electrons that is converted into photons through stimulated emission. In the above-threshold regime, $\eta_{int}$ is almost 100% for most semiconductor lasers. Equation (3.5.9) should be compared with Eq. (3.2.2) obtained for an LED.

A quantity of practical interest is the slope of the $P-I$ curve for $I > I_{th}$; it is called the slope efficiency and is defined as

$$\frac{dP_e}{dI} = \frac{\hbar \omega}{2q} \eta_{d} \quad \text{with} \quad \eta_{d} = \frac{\eta_{int} \alpha_{mir}}{\alpha_{mir} + \alpha_{int}}. \hspace{1cm} (3.5.10)$$

The quantity $\eta_{d}$ is called the differential quantum efficiency, as it is a measure of the efficiency with which light output increases with an increase in the injected current. One can define the external quantum efficiency $\eta_{ext}$ as

$$\eta_{ext} = \frac{\text{photon-emission rate}}{\text{electron-injection rate}} = \frac{2P_e}{\hbar \omega \frac{I}{q}} = \frac{2q}{\hbar \omega} \frac{P_e}{I}. \hspace{1cm} (3.5.11)$$

By using Eqs. (3.5.9)–(3.5.11), $\eta_{ext}$ and $\eta_{d}$ are found to be related by

$$\eta_{ext} = \eta_{d} \left(1 - \frac{I_{th}}{I}\right). \hspace{1cm} (3.5.12)$$

Generally, $\eta_{ext} < \eta_{d}$ but becomes nearly the same for $I \gg I_{th}$. Similar to the case of LEDs, one can define the total quantum efficiency (or wall-plug efficiency) as

$$\eta_{tot} = \frac{\hbar \omega}{qV_0} \eta_{ext} \approx \frac{E_g}{qV_0} \eta_{ext}, \hspace{1cm} (3.5.13)$$

where $E_g$ is the bandgap energy. Generally, $\eta_{tot} < \eta_{ext}$ as the applied voltage exceeds $E_g/q$. For GaAs lasers, $\eta_{d}$ can exceed 80% and $\eta_{tot}$ can approach 50%. The InGaAsP lasers are less efficient with $\eta_{d} \sim 50\%$ and $\eta_{tot} \sim 20\%$.

The exponential increase in the threshold current with temperature can be understood from Eq. (3.5.6). The carrier lifetime $\tau_c$ is generally $N$ dependent because of Auger recombination and decreases with $N$ as $N^2$. The rate of Auger recombination increases exponentially with temperature and is responsible for the temperature sensitivity of InGaAsP lasers. Figure 3.20 also shows that the slope efficiency decreases with an increase in the output power (bending of the $P-I$ curves). This decrease can be attributed to junction heating occurring under CW operation. It can also result from an increase in internal losses or current leakage at high operating powers. Despite these problems, the performance of DFB lasers improved substantially during the 1990s [10]–[12]. DFB lasers emitting $>100$ mW of power at room temperature in the 1.55 µm spectral region were fabricated by 1996 using a strained MQW design [61]. Such lasers exhibited $<10$ mA threshold current at $20^\circ$C and emitted $\sim 20$ mW of power at $100^\circ$C while maintaining a MSR of $>40$ dB. By 2001, DFB lasers capable of delivering more than 200 mW of power were available commercially.
### 3.5.2 Small-Signal Modulation

The modulation response of semiconductor lasers is studied by solving the rate equations (3.5.1) and (3.5.2) with a time-dependent current of the form

\[
I(t) = I_b + I_m f_p(t),
\]

where \(I_b\) is the bias current, \(I_m\) is the current, and \(f_p(t)\) represents the shape of the current pulse. Two changes are necessary for a realistic description. First, Eq. (3.5.3) for the gain \(G\) must be modified to become

\[
G = G_N (N - N_0)(1 - \varepsilon_{NL}P),
\]

where \(\varepsilon_{NL}\) is a nonlinear-gain parameter that leads to a slight reduction in \(G\) as \(P\) increases. The physical mechanism behind this reduction can be attributed to several phenomena, such as spatial hole burning, spectral hole burning, carrier heating, and two-photon absorption [62]–[65]. Typical values of \(\varepsilon_{NL}\) are \(\sim 10^{-7}\). Equation (3.5.15) is valid for \(\varepsilon_{NL}P \ll 1\). The factor \(1 - \varepsilon_{NL}P\) should be replaced by \(\left(1 + P/P_s\right)^{-b}\), where \(P_s\) is a material parameter, when the laser power exceeds far above 10 mW. The exponent \(b\) equals \(\frac{1}{2}\) for spectral hole burning [63] but can vary over the range 0.2–1 because of the contribution of carrier heating [65].

The second change is related to an important property of semiconductor lasers. It turns out that whenever the optical gain changes as a result of changes in the carrier population \(N\), the refractive index also changes. From a physical standpoint, amplitude modulation in semiconductor lasers is always accompanied by phase modulation because of carrier-induced changes in the mode index \(\bar{n}\). Phase modulation can be included through the equation

\[
\frac{d\varphi}{dt} = \frac{1}{2} \beta_c \left[ G_N (N - N_0) - \frac{1}{\tau_p} \right],
\]

where \(\beta_c\) is the amplitude-phase coupling parameter, commonly called the linewidth enhancement factor, as it leads to an enhancement of the spectral width associated with a single longitudinal mode (see Section 3.5.5). Typical values of \(\beta_c\) for InGaAsP lasers are in the range 4–8, depending on the operating wavelength [66]. Lower values of \(\beta_c\) occur in MQW lasers, especially for strained quantum wells [5].

In general, the nonlinear nature of the rate equations makes it necessary to solve them numerically. A useful analytic solution can be obtained for the case of small-signal modulation in which the laser is biased above threshold \((I_b > I_{th})\) and modulated such that \(I_m \ll I_b - I_{th}\). The rate equations can be linearized in that case and solved analytically, using the Fourier-transform technique, for an arbitrary form of \(f_p(t)\). The small-signal modulation bandwidth can be obtained by considering the response of semiconductor lasers to sinusoidal modulation at the frequency \(\omega_m\) so that \(f_p(t) = \sin(\omega_m t)\). The laser output is also modulated sinusoidally. The general solution of Eqs. (3.5.1) and (3.5.2) is given by

\[
P(t) = P_b + |p_m| \sin(\omega_m t + \theta_m),
\]

\[
N(t) = N_b + |n_m| \sin(\omega_m t + \psi_m),
\]
where \( P_b \) and \( N_b \) are the steady-state values at the bias current \( I_b \), \( |p_m| \) and \( |n_m| \) are small changes occurring because of current modulation, and \( \theta_m \) and \( \psi_m \) govern the phase lag associated with the small-signal modulation. In particular, \( p_m \equiv |p_m| \exp(i\theta_m) \) is given by [2]

\[
p_m(\omega_m) = \frac{P_b G_N I_m / q}{(\Omega_R + \omega_m - i\Gamma_R)(\Omega_R - \omega_m + i\Gamma_R)}. \tag{3.5.19}
\]

where

\[
\Omega_R = [G_G P_b - (\Gamma_P - \Gamma_N)^2/4]^{1/2}, \quad \Gamma_R = (\Gamma_P + \Gamma_N)/2, \tag{3.5.20}
\]

\[
\Gamma_P = R_p q / P_b + \epsilon_{NL} G_P b, \quad \Gamma_N = \tau_c^{-1} + G_N P_b. \tag{3.5.21}
\]

\( \Omega_R \) and \( \Gamma_R \) are the frequency and the damping rate of relaxation oscillations. These two parameters play an important role in governing the dynamic response of semiconductor lasers. In particular, the efficiency is reduced when the modulation frequency exceeds \( \Omega_R \) by a large amount.

Similar to the case of LEDs, one can introduce the transfer function as

\[
H(\omega_m) = \frac{p_m(\omega_m)}{p_m(0)} = \frac{\Omega_R^2 + \Gamma_R^2}{(\Omega_R + \omega_m - i\Gamma_R)(\Omega_R - \omega_m + i\Gamma_R)}. \tag{3.5.22}
\]

The modulation response is flat \( |H(\omega_m)| \approx 1 \) for frequencies such that \( \omega_m \ll \Omega_R \), peaks at \( \omega_m = \Omega_R \), and then drops sharply for \( \omega_m \gg \Omega_R \). These features are observed experimentally for all semiconductor lasers [67]–[70]. Figure 3.21 shows the modulation
response of a 1.55-µm DFB laser at several bias levels [70]. The 3-dB modulation bandwidth, \( f_{3\text{dB}} \), is defined as the frequency at which \(|H(\omega_m)|\) is reduced by 3 dB (by a factor of 2) compared with its direct-current (dc) value. Equation (3.5.22) provides the following analytic expression for \( f_{3\text{dB}} \):

\[
f_{3\text{dB}} = \frac{1}{2\pi} \left[ \Omega_R^2 + \Gamma_R^2 + 2(\Omega_R^4 + \Omega_R^2 \Gamma_R^2 + \Gamma_R^4)^{1/2} \right]^{1/2}.
\]

For most lasers, \( \Gamma_R \ll \Omega_R \), and \( f_{3\text{dB}} \) can be approximated by

\[
f_{3\text{dB}} \approx \sqrt{3} \Omega_R \frac{\Omega_R}{2\pi} = \left( \frac{3G_N P_b}{4\pi^2 \tau_p} \right)^{1/2} = \left( \frac{3G_N}{4\pi^2 q} (I_b - I_{th}) \right)^{1/2},
\]

(3.5.24)

where \( \Omega_R \) was approximated by \((GG_N P_b)^{1/2}\) in Eq. (3.5.21) and \( G \) was replaced by \( 1/\tau_p \) since gain equals loss in the above-threshold regime. The last expression was obtained by using Eq. (3.5.7) at the bias level.

Equation (3.5.24) provides a remarkably simple expression for the modulation bandwidth. It shows that \( f_{3\text{dB}} \) increases with an increase in the bias level as \( \sqrt{P_b} \) or as \((I_b - I_{th})^{1/2}\). This square-root dependence has been verified for many DFB lasers exhibiting a modulation bandwidth of up to 30 GHz [67]–[70]. Figure 3.21 shows how \( f_{3\text{dB}} \) can be increased to 24 GHz for a DFB laser by biasing it at 80 mA [70]. A modulation bandwidth of 25 GHz was realized in 1994 for a packaged 1.55-µm InGaAsP laser specifically designed for high-speed response [68].

### 3.5.3 Large-Signal Modulation

The small-signal analysis, although useful for a qualitative understanding of the modulation response, is not generally applicable to optical communication systems where the laser is typically biased close to threshold and modulated considerably above threshold to obtain optical pulses representing digital bits. In this case of large-signal modulation, the rate equations should be solved numerically. Figure 3.22 shows, as an example, the shape of the emitted optical pulse for a laser biased at \( I_b = 1.1I_{th} \) and modulated at 2 Gb/s using rectangular current pulses of duration 500 ps and amplitude \( I_m = I_{th} \). The optical pulse does not have sharp leading and trailing edges because of a limited modulation bandwidth and exhibits a rise time \( \sim 100 \) ps and a fall time \( \sim 300 \) ps. The initial overshoot near the leading edge is a manifestation of relaxation oscillations. Even though the optical pulse is not an exact replica of the applied electrical pulse, deviations are small enough that semiconductor lasers can be used in practice.

As mentioned before, amplitude modulation in semiconductor lasers is accompanied by phase modulation governed by Eq. (3.5.16). A time-varying phase is equivalent to transient changes in the mode frequency from its steady-state value \( \nu_0 \). Such a pulse is called chirped. The frequency chirp \( \delta \nu(t) \) is obtained by using Eq. (3.5.16) and is given by

\[
\delta \nu(t) = \frac{1}{2\pi} \frac{d \phi}{dt} = \frac{\beta_c}{4\pi} \left[ G_N (N - N_0) - \frac{1}{\tau_p} \right].
\]

The dashed curve in Fig. 3.21 shows the frequency chirp across the optical pulse. The mode frequency shifts toward the blue side near the leading edge and toward the red
3.5. LASER CHARACTERISTICS

Figure 3.22: Simulated modulation response of a semiconductor laser to 500-ps rectangular current pulses. Solid curve shows the pulse shape and the dashed curve shows the frequency chirp imposed on the pulse ($\beta_c = 5$).

side near the trailing edge of the optical pulse [71]. Such a frequency shift implies that the pulse spectrum is considerably broader than that expected in the absence of frequency chirp.

It was seen in Section 2.4 that the frequency chirp can limit the performance of optical communication systems, especially when $\beta_2 C > 0$, where $\beta_2$ is the dispersion parameter and $C$ is the chirp parameter. Even though optical pulses emitted from semiconductors are generally not Gaussian, the analysis of Section 2.4 can be used to study chirp-induced pulse broadening [72] if we identify $C$ with $-\beta_c$ in Eq. (2.4.23). It turns out that 1.55-µm lightwave systems are limited to distances below 100 km even at a bit rate of 2.5 Gb/s because of the frequency chirp [71] when conventional fibers are used ($\beta_2 \approx -20$ ps$^2$/km). Higher bit rates and longer distances can only be realized by using a dispersion management scheme so that the average dispersion is close to zero (see Chapter 7).

Since frequency chirp is often the limiting factor for lightwave systems operating near 1.55 µm, several methods have been used to reduce its magnitude [73]–[77]. These include pulse-shape tailoring, injection locking, and coupled-cavity schemes. A direct way to reduce the frequency chirp is to design semiconductor lasers with small values of the linewidth enhancement factor $\beta_c$. The use of quantum-well design reduces $\beta_c$ by about a factor of about 2. A further reduction occurs for strained quantum wells [76]. Indeed, $\beta_c \approx 1$ has been measured in modulation-doped strained MQW lasers [77]. Such lasers exhibit low chirp under direct modulation. The frequency chirp resulting from current modulation can be avoided altogether if the laser is continuously operated, and an external modulator is used to modulate the laser output [78]. In practice, lightwave systems operating at 10 Gb/s or more use either a monolithically...
integrated electroabsorption modulator or an external LiNbO$_3$ modulator (see Section 3.6). One can even design a modulator to reverse the sign of chirp such that $\beta_2 C < 0$, resulting in improved system performance.

Lightwave system designed using the RZ format, optical time-division multiplexing, or solitons often require mode-locked lasers that generate short optical pulses (width $\sim 10$ ps) at a high repetition rate equal to the bit rate. External-cavity semiconductor lasers can be used for this purpose, and are especially practical if a fiber grating is used for an external mirror. An external modulator is still needed to impose the data on the mode-locked pulse train; it blocks pulses in each bit slot corresponding to 0 bits. The gain switching has also been used to generate short pulses from a semiconductor laser. A mode-locked fiber laser can also be used for the same purpose [79].

### 3.5.4 Relative Intensity Noise

The output of a semiconductor laser exhibits fluctuations in its intensity, phase, and frequency even when the laser is biased at a constant current with negligible current fluctuations. The two fundamental noise mechanisms are spontaneous emission and electron–hole recombination (shot noise). Noise in semiconductor lasers is dominated by spontaneous emission. Each spontaneously emitted photon adds to the coherent field (established by stimulated emission) a small field component whose phase is random, and thus perturbs both amplitude and phase in a random manner. Moreover, such spontaneous-emission events occur randomly at a high rate ($\sim 10^{12}$ s$^{-1}$) because of a relatively large value of $R_p$ in semiconductor lasers. The net result is that the intensity and the phase of the emitted light exhibit fluctuations over a time scale as short as 100 ps. Intensity fluctuations lead to a limited signal-to-noise ratio (SNR), whereas phase fluctuations lead to a finite spectral linewidth when semiconductor lasers are operated at a constant current. Since such fluctuations can affect the performance of lightwave systems, it is important to estimate their magnitude [80].

The rate equations can be used to study laser noise by adding a noise term, known as the Langevin force, to each of them [81]. Equations (3.5.1), (3.5.2), and (3.5.16) then become

\[
\frac{dP}{dt} = \left( G - \frac{1}{\tau_p} \right) P + R_p + F_p(t),
\]

\[
\frac{dN}{dt} = \frac{I}{q} - \frac{N}{\tau_c} + GP + F_N(t),
\]

\[
\frac{d\phi}{dt} = \frac{1}{2} \beta_c \left[ G_N (N - N_0) - \frac{1}{\tau_p} \right] + F_\phi(t),
\]

where $F_p(t)$, $F_N(t)$, and $F_\phi(t)$ are the Langevin forces. They are assumed to be Gaussian random processes with zero mean and to have a correlation function of the form (the Markoffian approximation)

\[
\langle F_i(t) F_j(t') \rangle = 2D_{ij} \delta(t - t'),
\]
Figure 3.23: RIN spectra at several power levels for a typical 1.55-\(\mu\)m semiconductor laser.

where \(i, j = P, N, \text{ or } \phi\), angle brackets denote the ensemble average, and \(D_{ij}\) is called the diffusion coefficient. The dominant contribution to laser noise comes from only two diffusion coefficients \(D_{PP} = R_{sp}P\) and \(D_{\phi\phi} = R_{sp}/4P\); others can be assumed to be nearly zero [82].

The intensity-autocorrelation function is defined as

\[
C_{pp}(\tau) = \langle \delta P(t) \delta P(t + \tau) \rangle / \bar{P}^2,
\]

where \(\bar{P} \equiv \langle P \rangle\) is the average value and \(\delta P = P - \bar{P}\) represents a small fluctuation. The Fourier transform of \(C_{pp}(\tau)\) is known as the relative-intensity-noise (RIN) spectrum and is given by

\[
\text{RIN}(\omega) = \int_{-\infty}^{\infty} C_{pp}(\tau) \exp(-i\omega \tau) d\tau.
\]

The RIN can be calculated by linearizing Eqs. (3.5.26) and (3.5.27) in \(\delta P\) and \(\delta N\), solving the linearized equations in the frequency domain, and performing the average with the help of Eq. (3.5.29). It is given approximately by [2]

\[
\text{RIN}(\omega) = \frac{2R_{sp} \{ (\Gamma_R^2 + \omega^2) + G_N \bar{P}\{1 + N/\tau R_{sp} \bar{P} - 2\Gamma_N\} \} }{ \bar{P}\{2\Omega_R^2 - \omega^2\} + 2\Gamma_R^2 \} }{ \langle \Omega_R^2 + \omega^2 + 2\Gamma_R^2 \rangle },
\]

where \(\Omega_R\) and \(\Gamma_R\) are the frequency and the damping rate of relaxation oscillations. They are given by Eq. (3.5.21), with \(P\), replaced by \(\bar{P}\).

Figure 3.23 shows the calculated RIN spectra at several power levels for a typical 1.55-\(\mu\)m InGaAsP laser. The RIN is considerably enhanced near the relaxation-oscillation frequency \(\Omega_R\) but decreases rapidly for \(\omega \gg \Omega_R\), since the laser is not able to respond to fluctuations at such high frequencies. In essence, the semiconductor laser
acts as a bandpass filter of bandwidth $\Omega_R$ to spontaneous-emission fluctuations. At a given frequency, RIN decreases with an increase in the laser power as $P^{-3}$ at low powers, but this behavior changes to $P^{-1}$ dependence at high powers.

The autocorrelation function $C_{pp}(\tau)$ is calculated using Eqs. (3.5.31) and (3.5.32). The calculation shows that $C_{pp}(\tau)$ follows relaxation oscillations and approaches zero for $\tau > \Gamma_R^{-1}$ [83]. This behavior indicates that intensity fluctuations do not remain correlated for times longer than the damping time of relaxation oscillations. The quantity of practical interest is the SNR defined as $\overline{P}/\sigma_p$, where $\sigma_p$ is the root-mean-square (RMS) noise. From Eq. (3.5.30), $\text{SNR} = [C_{pp}(0)]^{-1/2}$.

At power levels above a few milliwatts, the SNR exceeds 20 dB and improves linearly with the power as

$$\text{SNR} = \left( \frac{\epsilon_{\text{NL}}}{R_{sp} \tau_p} \right)^{1/2} \tilde{P},$$

(3.5.33)

The presence of $\epsilon_{\text{NL}}$ indicates that the nonlinear form of the gain in Eq. (3.5.15) plays a crucial role. This form needs to be modified at high powers. Indeed, a more accurate treatment shows that the SNR eventually saturates at a value of about 30 dB and becomes power independent [83].

So far, the laser has been assumed to oscillate in a single longitudinal mode. In practice, even DFB lasers are accompanied by one or more side modes. Even though side modes remain suppressed by more than 20 dB on the basis of the average power, their presence can affect the RIN significantly. In particular, the main and side modes can fluctuate in such a way that individual modes exhibit large intensity fluctuations, but the total intensity remains relatively constant. This phenomenon is called mode-partition noise (MPN) and occurs due to an anticorrelation between the main and side modes [2]. It manifests through the enhancement of RIN for the main mode by 20 dB or more in the low-frequency range 0–1 GHz; the exact value of the enhancement factor depends on the MSR [84]. In the case of a VCSEL, the MPN involves two transverse modes. [85]. In the absence of fiber dispersion, MPN would be harmless for optical communication systems, as all modes would remain synchronized during transmission and detection. However, in practice all modes do not arrive simultaneously at the receiver because they travel at slightly different speeds. Such a desynchronization not only degrades the SNR of the received signal but also leads to intersymbol interference. The effect of MPN on the system performance is discussed in Section 7.4.3.

### 3.5.5 Spectral Linewidth

The spectrum of emitted light is related to the field-autocorrelation function $\Gamma_{EE}(\tau)$ through a Fourier-transform relation similar to Eq. (3.5.31), i.e.,

$$S(\omega) = \int_{-\infty}^{\infty} \Gamma_{EE}(t) \exp[-i(\omega - \omega_0)\tau] d\tau,$$

(3.5.34)

where $\Gamma_{EE}(t) = \langle E^*(t)E(t+\tau) \rangle$ and $E(t) = \sqrt{\overline{P}} \exp(i\phi)$ is the optical field. If intensity fluctuations are neglected, $\Gamma_{EE}(t)$ is given by

$$\Gamma_{EE}(t) = \langle \exp[i\Delta\phi(t)] \rangle = \exp[-\langle \Delta\phi^2(\tau) \rangle/2],$$

(3.5.35)
where the phase fluctuation $\Delta \phi(\tau) = \phi(t + \tau) - \phi(t)$ is taken to be a Gaussian random process. The phase variance $\langle \Delta \phi^2(\tau) \rangle$ can be calculated by linearizing Eqs. (3.5.26)–(3.5.28) and solving the resulting set of linear equations. The result is [82]

$$\langle \Delta \phi^2(\tau) \rangle = \frac{R_{sp}}{2P} \left[ (1 + \beta_c^2 b) \tau + \frac{\beta_c^2 b}{2\Gamma_R \cos \delta} [\cos(3\delta) - e^{-\Gamma_R \tau} \cos(\Omega_R \tau - 3\delta)] \right],$$

(3.5.36)

where

$$b = \Omega_R/(\Omega_R^2 + \Gamma_R^2)^{1/2} \quad \text{and} \quad \delta = \tan^{-1}(\Gamma_R/\Omega_R).$$

(3.5.37)

The spectrum is obtained by using Eqs. (3.5.34)–(3.5.36). It is found to consist of a dominant central peak located at $\omega_0$ and multiple satellite peaks located at $\omega = \omega_0 \pm m \Omega_R$, where $m$ is an integer. The amplitude of satellite peaks is typically less than 1% of that of the central peak. The physical origin of the satellite peaks is related to relaxation oscillations, which are responsible for the term proportional to $b$ in Eq. (3.5.36). If this term is neglected, the autocorrelation function $\Gamma_{EE}(\tau)$ decays exponentially with $\tau$. The integral in Eq. (3.5.34) can then be performed analytically, and the spectrum is found to be Lorentzian. The spectral linewidth $\Delta \nu$ is defined as the full-width at half-maximum (FWHM) of this Lorentzian line and is given by [82]

$$\Delta \nu = \frac{R_{sp}(1 + \beta_c^2)}{(4\pi P)},$$

(3.5.38)

where $b = 1$ was assumed as $\Gamma_R \ll \Omega_R$ under typical operating conditions. The linewidth is enhanced by a factor of $1 + \beta_c^2$ as a result of the amplitude-phase coupling governed by $\beta_c$ in Eq. (3.5.28); $\beta_c$ is called the linewidth enhancement factor for this reason.

Equation (3.5.38) shows that $\Delta \nu$ should decrease as $P^{-1}$ with an increase in the laser power. Such an inverse dependence is observed experimentally at low power levels ($< 10$ mW) for most semiconductor lasers. However, often the linewidth is found to saturate to a value in the range $1–10$ MHz at a power level above $10$ mW. Figure 3.24 shows such linewidth-saturation behavior for several 1.55-μm DFB lasers [86]. It also shows that the linewidth can be reduced considerably by using a MQW design for the DFB laser. The reduction is due to a smaller value of the parameter $\beta_c$ realized by such a design. The linewidth can also be reduced by increasing the cavity length $L$, since $R_{sp}$ decreases and $P$ increases at a given output power as $L$ is increased. Although not obvious from Eq. (3.5.38), $\Delta \nu$ can be shown to vary as $L^{-2}$ when the length dependence of $R_{sp}$ and $P$ is incorporated. As seen in Fig. 3.24, $\Delta \nu$ is reduced by about a factor of 4 when the cavity length is doubled. The 800-μm-long MQW-DFB laser is found to exhibit a linewidth as small as $270$ kHz at a power output of $13.5$ mW [86]. It is further reduced in strained MQW lasers because of relatively low values of $\beta_c$, and a value of about $100$ kHz has been measured in lasers with $\beta_c \approx 1$ [77]. It should be stressed, however, that the linewidth of most DFB lasers is small enough that it is not a limiting factor for lightwave systems.

Figure 3.24 shows that as the laser power increases, the linewidth not only saturates but begins to rebroaden. Several mechanisms have been invoked to explain such behavior; a few of them are current noise, $1/f$ noise, nonlinear gain, sidemode interaction, and index nonlinearity [87]–[94]. The linewidth of most DFB lasers is small enough that it is not a limiting factor for lightwave systems.
3.6 Transmitter Design

So far this chapter has focused on the properties of optical sources. Although an optical source is a major component of optical transmitters, it is not the only component. Other components include a modulator for converting electrical data into optical form (if direct modulation is not used) and an electrical driving circuit for supplying current to the optical source. An external modulator is often used in practice at bit rates of 10 Gb/s or more for avoiding the chirp that is invariably imposed on the directly modulated signal. This section covers the design of optical transmitters with emphasis on the packaging issues [95]–[105].

3.6.1 Source–Fiber Coupling

The design objective for any transmitter is to couple as much light as possible into the optical fiber. In practice, the coupling efficiency depends on the type of optical source (LED versus laser) as well as on the type of fiber (multimode versus single mode). The coupling can be very inefficient when light from an LED is coupled into a single-mode fiber. As discussed briefly in Section 3.2.1, the coupling efficiency for an LED changes with the numerical aperture, and can become < 1% in the case of single-mode fibers. In contrast, the coupling efficiency for edge-emitting lasers is typically 40–50% and can exceed 80% for VCSELs because of their circular spot size. A small piece of fiber (known as a pigtail) is included with the transmitter so that the coupling efficiency can
be maximized during packaging; a splice or connector is used to join the pigtail with the fiber cable.

Two approaches have been used for source–fiber coupling. In one approach, known as direct or butt coupling, the fiber is brought close to the source and held in place by epoxy. In the other, known as lens coupling, a lens is used to maximize the coupling efficiency. Each approach has its own merits, and the choice generally depends on the design objectives. An important criterion is that the coupling efficiency should not change with time; mechanical stability of the coupling scheme is therefore a necessary requirement.

An example of butt coupling is shown in Fig. 3.25(a), where the fiber is brought in contact with a surface-emitting LED. The coupling efficiency for a fiber of numerical aperture NA is given by [96]

\[
nc = (1 - R_f)(NA)^2,
\]

(3.6.1)

where \(R_f\) is the reflectivity at the fiber front end. \(R_f\) is about 4% if an air gap exists between the source and the fiber but can be reduced to nearly zero by placing an index-matching liquid. The coupling efficiency is about 1% for a surface-emitting LED and roughly 10% for an edge-emitting LED. Some improvement is possible in both cases.
by using fibers that are tapered or have a lensed tip. An external lens also improves the coupling efficiency but only at the expense of reduced mechanical tolerance.

The coupling of a semiconductor laser to a single-mode optical fiber is more efficient than that of an LED. The butt coupling provides only about 10% efficiency, as it makes no attempt to match the mode sizes of the laser and the fiber. Typically, index-guided InGaAsP lasers have a mode size of about 1 µm, whereas the mode size of a single-mode fiber is in the range 6–9 µm. The coupling efficiency can be improved by tapering the fiber end and forming a lens at the fiber tip. Figure 3.25(a) shows such a butt-coupling scheme for a commercial transmitter. The fiber is attached to a jewel, and the jewel is attached to the laser submount by using an epoxy [97]. The fiber tip is aligned with the emitting region of the laser to maximize the coupling efficiency (typically 40%). The use of a lensed fiber can improve the coupling efficiency, and values close to 100% have been realized with an optimum design [98]–[100].

Figure 3.25(b) shows a lens-coupling approach for transmitter design. The coupling efficiency can exceed 70% for such a confocal design in which a sphere is used to collimate the laser light and focus it onto the fiber core. The alignment of the fiber core is less critical for the confocal design because the spot size is magnified to match the fiber’s mode size. The mechanical stability of the package is ensured by soldering the fiber into a ferrule which is secured to the body by two sets of laser alignment welds. One set of welds establishes proper axial alignment, while the other set provides transverse alignment.

The laser–fiber coupling issue remains important, and several new schemes have been developed during the 1990s [101]–[105]. In one approach, a silicon optical bench is used to align the laser and the fiber [101]. In another, a silicon micromirror, fabricated by using the micro-machining technology, is used for optical alignment [102]. In a different approach, a directional coupler is used as the spot-size converter for maximizing the coupling efficiency [103]. Coupling efficiencies >80% have been realized by integrating a spot-size converter with semiconductor lasers [105].

An important problem that needs to be addressed in designing an optical transmitter is related to the extreme sensitivity of semiconductor lasers to optical feedback [2]. Even a relatively small amount of feedback (< 0.1%) can destabilize the laser and affect the system performance through phenomena such as linewidth broadening, mode hopping, and RIN enhancement [106]–[110]. Attempts are made to reduce the feedback into the laser cavity by using antireflection coatings. Feedback can also be reduced by cutting the fiber tip at a slight angle so that the reflected light does not hit the active region of the laser. Such precautions are generally enough to reduce the feedback to a tolerable level. However, it becomes necessary to use an optical isolator between the laser and the fiber in transmitters designed for more demanding applications. One such application corresponds to lightwave systems operating at high bit rates and requiring a narrow-linewidth DFB laser.

Most optical isolators make use of the Faraday effect, which governs the rotation of the plane of polarization of an optical beam in the presence of a magnetic field: The rotation is in the same direction for light propagating parallel or antiparallel to the magnetic field direction. Optical isolators consist of a rod of Faraday material such as yttrium iron garnet (YIG), whose length is chosen to provide 45° rotation. The YIG rod is sandwiched between two polarizers whose axes are tilted by 45° with
3.6. TRANSMITTER DESIGN

Figure 3.26: Driving circuit for a laser transmitter with feedback control to keep the average optical power constant. A photodiode monitors the output power and provides the control signal. (After Ref. [95]; ©1988 Academic Press; reprinted with permission.)

Figure 3.26 shows a simple driving circuit that controls the average optical power through a feedback mechanism. A photodiode monitors the laser output and generates the control signal that is used to adjust the laser bias level. The rear facet of the laser is generally used for the monitoring purpose (see Fig. 3.25). In some transmitters a front-end tap is used to divert a small fraction of the output power to the detector. The bias-level control is essential, since the laser threshold is sensitive to the operating respect to each other. Light propagating in one direction passes through the second polarizer because of the Faraday rotation. By contrast, light propagating in the opposite direction is blocked by the first polarizer. Desirable characteristics of optical isolators are low insertion loss, high isolation (> 30 dB), compact size, and a wide spectral bandwidth of operation. A very compact isolator can be designed if the lens in Fig. 3.25(b) is replaced by a YIG sphere so that it serves a dual purpose [111]. As light from a semiconductor laser is already polarized, a signal polarizer placed between the YIG sphere and the fiber can reduce the feedback by more than 30 dB.

3.6.2 Driving Circuitry

The purpose of driving circuitry is to provide electrical power to the optical source and to modulate the light output in accordance with the signal that is to be transmitted. Driving circuits are relatively simple for LED transmitters but become increasingly complicated for high-bit-rate optical transmitters employing semiconductor lasers as an optical source [95]. As discussed in Section 3.5.2, semiconductor lasers are biased near threshold and then modulated through an electrical time-dependent signal. Thus the driving circuit is designed to supply a constant bias current as well as modulated electrical signal. Furthermore, a servo loop is often used to keep the average optical power constant.
The threshold current also increases with aging of the transmitter because of gradual degradation of the semiconductor laser.

The driving circuit shown in Fig. 3.26 adjusts the bias level dynamically but leaves the modulation current unchanged. Such an approach is acceptable if the slope efficiency of the laser does not change with aging. As discussed in Section 3.5.1 and seen in Fig. 3.20, the slope efficiency of the laser generally decreases with an increase in temperature. A thermoelectric cooler is often used to stabilize the laser temperature. An alternative approach consists of designing driving circuits that use dual-loop feedback circuits and adjust both the bias current and the modulation current automatically [112].

3.6.3 Optical Modulators

At bit rates of 10 Gb/s or higher, the frequency chirp imposed by direct modulation becomes large enough that direct modulation of semiconductor lasers is rarely used. For such high-speed transmitters, the laser is biased at a constant current to provide the CW output, and an optical modulator placed next to the laser converts the CW light into a data-coded pulse train with the right modulation format.

Two types of optical modulators developed for lightwave system applications are shown in Fig. 3.27. The electroabsorption modulator makes use of the Franz–Keldysh effect, according to which the bandgap of a semiconductor decreases when an electric field is applied across it. Thus, a transparent semiconductor layer begins to absorb light when its bandgap is reduced electronically by applying an external voltage. An extinction ratio of 15 dB or more can be realized for an applied reverse bias of a few volts at bit rates of up to 40 Gb/s [113]–[120]. Although some chirp is still imposed on coded pulses, it can be made small enough not to be detrimental for the system performance.

An advantage of electroabsorption modulators is that they are made using the same semiconductor material that is used for the laser, and thus the two can be easily inte-
3.6. TRANSMITTER DESIGN

Integrated on the same chip. Low-chirp transmission at a bit rate of 5 Gb/s was demonstrated as early as 1994 by integrating an electroabsorption modulator with a DBR laser [114]. By 1999, 10-Gb/s optical transmitters with an integrated electroabsorption modulator were available commercially and were used routinely for WDM lightwave systems [119]. By 2001, such integrated modulators exhibited a bandwidth of more than 50 GHz and had the potential of operating at bit rates of up to 100 Gb/s [120]. An electroabsorption modulator can also be used to generate ultrashort pulses suitable for optical time-division multiplexing (OTDM). A DFB laser, integrated monolithically with a MQW modulator, was used as early as 1993 to generate a 20-GHz pulse train [113]. The 7-ps output pulses were nearly transform-limited because of an extremely low chirp associated with the modulator. A 40-GHz train of 1.6 ps pulses was produced in 1999 using an electroabsorption modulator; such pulses can be used for OTDM systems operating at a bit rate of 160 Gb/s [116].

The second category of optical modulators makes use of the LiNbO$_3$ material and a Mach–Zehnder (MZ) interferometer for intensity modulation [121]–[126]. Two titanium-diffused LiNbO$_3$ waveguides form the two arms of a MZ interferometer (see Fig. 3.27). The refractive index of electro-optic materials such as LiNbO$_3$ can be changed by applying an external voltage. In the absence of external voltage, the optical fields in the two arms of the MZ interferometer experience identical phase shifts and interfere constructively. The additional phase shift introduced in one of the arms through voltage-induced index changes destroys the constructive nature of the interference and reduces the transmitted intensity. In particular, no light is transmitted when the phase difference between the two arms equals $\pi$, because of destructive interference occurring in that case. As a result, the electrical bit stream applied to the modulator produces an optical replica of the bit stream.

The performance of an external modulator is quantified through the on–off ratio (also called extinction ratio) and the modulation bandwidth. Modern LiNbO$_3$ modulators provide an on–off ratio in excess of 20 and can be modulated at speeds up to 75 GHz [122]. The driving voltage is typically 5 V but can be reduced to below 3 V with a suitable design [125]. LiNbO$_3$ modulators with a bandwidth of 10 GHz were available commercially by 1998, and the bandwidth increased to 40 GHz by 2000 [126].

Other materials can also be used to make external modulators. For example, modulators have been fabricated using electro-optic polymers. Already in 1995 such a modulator exhibited a modulation bandwidth of up to 60 GHz [127]. In a 2001 experiment, a polymeric electro-optic MZ modulator required only 1.8 V for shifting the phase of a 1.55-µm signal by $\pi$ in one of the arms of the MZ interferometer [128]. The device was only 3 cm long and exhibited about 5-dB chip losses. With further development, such modulators may find applications in lightwave systems.

3.6.4 Optoelectronic Integration

The electrical components used in the driving circuit determine the rate at which the transmitter output can be modulated. For lightwave transmitters operating at bit rates above 1 Gb/s, electrical parasitics associated with various transistors and other components often limit the transmitter performance. The performance of high-speed trans-
mitters can be improved considerably by using monolithic integration of the laser with the driver. Since optical and electrical devices are fabricated on the same chip, such monolithic transmitters are referred to as optoelectronic integrated-circuit (OEIC) transmitters. The OEIC approach was first applied to integration of GaAs lasers, since the technology for fabrication of GaAs electrical devices is relatively well established [129]–[131]. The technology for fabrication of InP OEICs evolved rapidly during the 1990s [132]–[136]. A 1.5-$\mu$m OEIC transmitter capable of operating at 5 Gb/s was demonstrated in 1988 [132]. By 1995, 10-Gb/s laser transmitters were fabricated by integrating 1.55-$\mu$m DFB lasers with field-effect transistors made with the InGaAs/InAlAs material system. Since then, OEIC transmitters with multiple lasers on the same chip have been developed for WDM applications (see Chapter 8).

A related approach to OEIC integrates the semiconductor laser with a photodetector [137]–[139] and/or with a modulator [117]–[120]. The photodetector is generally used for monitoring and stabilizing the output power of the laser. The role of the modulator is to reduce the dynamic chirp occurring when a semiconductor laser is modulated directly (see Section 3.5.2). Photodetectors can be fabricated by using the same material as that used for the laser (see Chapter 4).

The concept of monolithic integration can be extended to build single-chip transmitters by adding all functionality on the same chip. Considerable effort has been directed toward developing such OEICs, often called photonic integrated circuits [6], which integrate on the same chip multiple optical components, such as lasers, detectors, modulators, amplifiers, filters, and waveguides [140]–[145]. Such integrated circuits should prove quite beneficial to lightwave technology.

### 3.6.5 Reliability and Packaging

An optical transmitter should operate reliably over a relatively long period of time (10 years or more) in order to be useful as a major component of lightwave systems. The reliability requirements are quite stringent for undersea lightwave systems, for which repairs and replacement are prohibitively expensive. By far the major reason for failure of optical transmitters is the optical source itself. Considerable testing is performed during assembly and manufacture of transmitters to ensure a reasonable lifetime for the optical source. It is common [95] to quantify the lifetime by a parameter $t_F$ known as mean time to failure (MTTF). Its use is based on the assumption of an exponential failure probability $[P_F = \exp(-t/t_F)]$. Typically, $t_F$ should exceed $10^5$ hours (about 11 years) for the optical source. Reliability of semiconductor lasers has been studied extensively to ensure their operation under realistic operating conditions [146]–[151].

Both LEDs and semiconductor lasers can stop operating suddenly (catastrophic degradation) or may exhibit a gradual mode of degradation in which the device efficiency degrades with aging [147]. Attempts are made to identify devices that are likely to degrade catastrophically. A common method is to operate the device at high temperatures and high current levels. This technique is referred to as burn-in or accelerated aging [146] and is based on the assumption that under high-stress conditions weak devices will fail, while others will stabilize after an initial period of rapid degradation. The change in the operating current at a constant power is used as a measure of device degradation. Figure 3.28 shows the change in the operating current of a 1.3-$\mu$m
InGaAsP laser aged at 60°C under a constant output power of 5 mW from each facet. The operating current for this laser increases by 40% in the first 400 hours but then stabilizes and increases at a much reduced rate indicative of gradual degradation. The degradation rate can be used to estimate the laser lifetime and the MTTF at the elevated temperature. The MTTF at the normal operating temperature is then extrapolated by using an Arrhenius-type relation $t_F = t_0 \exp(-E_a/k_BT)$, where $t_0$ is a constant and $E_a$ is the activation energy with a typical value of about 1 eV [147]. Physically, gradual degradation is due to the generation of various kinds of defects (dark-line defects, dark-spot defects) within the active region of the laser or LED [2].

Extensive tests have shown that LEDs are normally more reliable than semiconductor lasers under the same operating conditions. The MTTF for GaAs LEDs easily exceeds 10⁶ hours and can be $> 10^7$ hours at 25°C [147]. The MTTF for InGaAsP LEDs is even larger, approaching a value $\sim 10^{10}$ hours. By contrast, the MTTF for InGaAsP lasers is generally limited to 10⁶ hours at 25°C [148]–[150]. Nonetheless, this value is large enough that semiconductor lasers can be used in undersea optical transmitters designed to operate reliably for a period of 25 years. Because of the adverse effect of high temperatures on device reliability, most transmitters use a thermoelectric cooler to maintain the source temperature near 20°C even when the outside temperature may be as high as 80°C.

Even with a reliable optical source, a transmitter may fail in an actual system if the coupling between the source and the fiber degrades with aging. Coupling stability is an important issue in the design of reliable optical transmitters. It depends ultimately on the packaging of transmitters. Although LEDs are often packaged nonhermetically, an hermetic environment is essential for semiconductor lasers. It is common to package the laser separately so that it is isolated from other transmitter components. Figure 3.25 showed two examples of laser packages. In the butt-coupling scheme, an epoxy
is used to hold the laser and fiber in place. Coupling stability in this case depends on how epoxy changes with aging of the transmitter. In the lens-coupling scheme, laser welding is used to hold various parts of the assembly together. The laser package becomes a part of the transmitter package, which includes other electrical components associated with the driving circuit. The choice of transmitter package depends on the type of application; a dual-in-line package or a butterfly housing with multiple pins is typically used.

Testing and packaging of optical transmitters are two important parts of the manufacturing process [149], and both of them add considerably to the cost of a transmitter. The development of low-cost packaged transmitters is necessary, especially for local-area and local-loop applications.

### Problems

3.1 Show that the external quantum efficiency of a planar LED is given approximately by \( \eta_{ext} = n^{-1}(n + 1)^{-2} \), where \( n \) is the refractive index of the semiconductor–air interface. Consider Fresnel reflection and total internal reflection at the output facet. Assume that the internal radiation is uniform in all directions.

3.2 Prove that the 3-dB optical bandwidth of a LED is related to the 3-dB electrical bandwidth by the relation \( f_{3dB}^{\text{optical}} = \sqrt{3} f_{3dB}^{\text{electrical}} \).

3.3 Find the composition of the quaternary alloy InGaAsP for making semiconductor lasers operating at 1.3- and 1.55-\( \mu \)m wavelengths.

3.4 The active region of a 1.3-\( \mu \)m InGaAsP laser is 250 \( \mu \)m long. Find the active-region gain required for the laser to reach threshold. Assume that the internal loss is 30 cm\(^{-1}\), the mode index is 3.3, and the confinement factor is 0.4.

3.5 Derive the eigenvalue equation for the transverse-electric (TE) modes of a planar waveguide of thickness \( d \) and refractive index \( n_1 \) sandwiched between two cladding layers of refractive index \( n_2 \). (Hint: Follow the method of Section 2.2.2 using Cartesian coordinates.)

3.6 Use the result of Problem 3.5 to find the single-mode condition. Use this condition to find the maximum allowed thickness of the active layer for a 1.3-\( \mu \)m semiconductor laser. How does this value change if the laser operates at 1.55 \( \mu \)m? Assume \( n_1 = 3.5 \) and \( n_2 = 3.2 \).

3.7 Solve the rate equations in the steady state and obtain the analytic expressions for \( P \) and \( N \) as a function of the injection current \( I \). Neglect spontaneous emission for simplicity.

3.8 A semiconductor laser is operating continuously at a certain current. Its output power changes slightly because of a transient current fluctuation. Show that the laser power will attain its original value through an oscillatory approach. Obtain the frequency and the damping time of such relaxation oscillations.

3.9 A 250-\( \mu \)m-long InGaAsP laser has an internal loss of 40 cm\(^{-1}\). It operates in a single mode with the modal index 3.3 and the group index 3.4. Calculate the
photon lifetime. What is the threshold value of the electron population? Assume that the gain varies as \( G = G_N(N - N_0) \) with \( G_N = 6 \times 10^3 \text{ s}^{-1} \) and \( N_0 = 1 \times 10^8 \).

3.10 Determine the threshold current for the semiconductor laser of Problem 3.9 by taking 2 ns as the carrier lifetime. How much power is emitted from one facet when the laser is operated twice above threshold?

3.11 Consider the laser of Problem 3.9 operating twice above threshold. Calculate the differential quantum efficiency and the external quantum efficiency for the laser. What is the device (wall-plug) efficiency if the external voltage is 1.5 V? Assume that the internal quantum efficiency is 90%.

3.12 Calculate the frequency (in GHz units) and the damping time of the relaxation oscillations for the laser of Problem 3.9 operating twice above threshold. Assume that \( G_P = -4 \times 10^4 \text{ s}^{-1} \), where \( G_P \) is the derivative of \( G \) with respect to \( P \). Also assume that \( R_{sp} = 2/\tau_p \).

3.13 Determine the 3-dB modulation bandwidth for the laser of Problem 3.11 biased to operate twice above threshold. What is the corresponding 3-dB electrical bandwidth?

3.14 The threshold current of a semiconductor laser doubles when the operating temperature is increased by 50°C. What is the characteristic temperature of the laser?

3.15 Derive an expression for the 3-dB modulation bandwidth by assuming that the gain \( G \) in the rate equations varies with \( N \) and \( P \) as

\[
G(N, P) = G_N(N - N_0)(1 + P/P_s)^{-1/2}.
\]

Show that the bandwidth saturates at high operating powers.

3.16 Solve the rate equations (3.5.1) and (3.5.2) numerically by using \( I(t) = I_b + I_m f_p(t) \), where \( f_p(t) \) represents a rectangular pulse of 200-ps duration. Assume that \( I_b/I_b = 0.8 \), \( I_m/I_b = 3 \), \( \tau_p = 3 \text{ ps} \), \( \tau_c = 2 \text{ ns} \), and \( R_{sp} = 2/\tau_p \). Use Eq. (3.5.15) for the gain \( G \) with \( G_N = 10^4 \text{ s}^{-1} \), \( N_0 = 10^9 \), and \( \epsilon_{NL} = 10^{-7} \). Plot the optical pulse shape and the frequency chirp. Why is the optical pulse much shorter than the applied current pulse?

3.17 Complete the derivation of Eq. (3.5.32) for the RIN. How does this expression change if the gain \( G \) is assumed of the form of Problem 3.15?

3.18 Calculate the autocorrelation \( C_{pp}(\tau) \) by using Eqs. (3.5.31) and (3.5.32). Use it to derive an expression for the SNR of the laser output.

References

REFERENCES


REFERENCES