INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
34100 TRIESTE (ITALY) - P.O.B. 586 - MIRAMARE - STRADA COSTIERA 11 - TELEPHONE: 2240-1
CABLE: CENTRATOM - TELEX 460392 - I

H4.SMR/285 - 13

WINTER COLLEGE ON
LASER PHYSICS: SEMICONDUCTOR LASERS
AND INTEGRATED OPTICS

(22 February - 11 March 1988)

LONG-WAVELENGTH
SEMICONDUCTOR LASERS

H. Melchior
Swiss Federal Institute of Technology
Zürich, Switzerland

LONG-WAVELENGTH SEMICONDUCTOR LASERS

G. P. Agrawal and N. K. Dutta

AT&T Bell Laboratories
Murray Hill, New Jersey

Van Nostrand Reinhold Electrical/Computer Science and Engineering Series

VNR VAN NOSTRAND REINHOLD COMPANY
New York

INTRODUCTION

1.1 HISTORICAL PERSPECTIVE

The advent of the laser dates back to 1958, the year in which the seminal paper of Schawlow and Townes[1] appeared. It was followed by the successful operation of a solid-state ruby laser[2] in May 1960 and of a He-Ne gas laser[3] in December 1960. The feasibility of stimulated emission in semiconductor lasers was considered during this period[4-6], and in 1962 several groups[7-10] reported the lasing action in semiconductors. The device consisted of a forward-biased GaAs *p-n* junction.[7-9] Electron-hole recombination in the depletion region of the *p-n* junction provided the optical gain, and the polished facets perpendicular to the junction plane provided the optical feedback (by forming a resonant cavity)—the two necessary ingredients for any laser. Soon *p-n* junctions of other direct-band-gap semiconductor materials such as InAs, InP, GaAsP, GaInAs, and InPAs were used to obtain semiconductor lasers at different wavelengths. Practical utility of these earlier devices was, however, limited since a large value of the threshold current density ($J_{th} \geq 50$ kA/cm^2) inhibited their continuous operation at room temperature.

As early as 1963 it was suggested[11,12] that semiconductor lasers might be improved if a layer of one semiconductor material were sandwiched between two cladding layers of another semiconductor that has a relatively wider band gap. Such a device consisting of two dissimilar semiconductors is commonly referred to as a *heterostructure laser*, in contrast to the single-semiconductor devices, which are labeled as *homostructure lasers*. Both of these structures are shown schematically in Fig. 1.1, which also indicates their typical physical dimensions. Heterostructure lasers are further classified as *single-heterostructure* or *double-heterostructure* devices depending on whether the active region, where lasing occurs, is surrounded on one or both sides by a cladding layer of higher band gap. The use of a heterostructure, however, requires a careful matching of the lattice constants of the two semiconductors. It was only in 1969 that the successful room-temperature operation of a

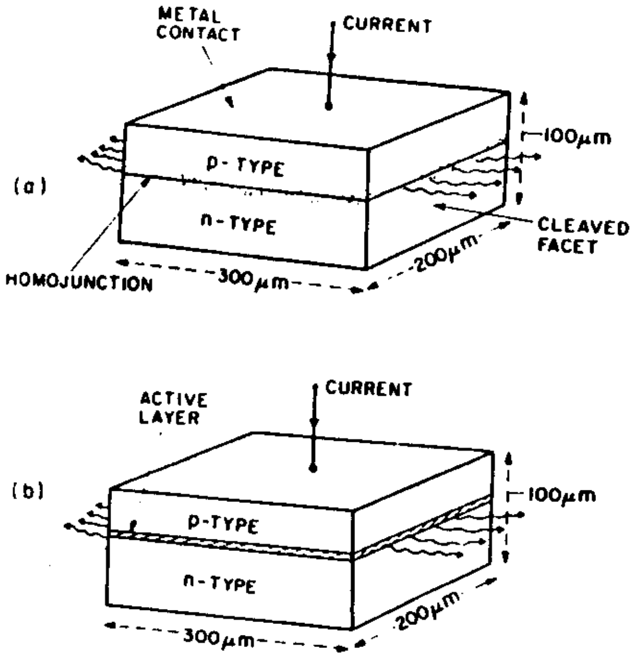


Fig. 1.1 Schematic illustration of (a) homostructure and (b) double-heterostructure semiconductor lasers with their typical physical dimensions. The dotted area represents the depletion region in the vicinity of the homojunction. The hatched area shows the thin (~0.2 μm) active layer of a semiconductor material whose band gap is slightly lower than that of the surrounding cladding layers.

heterostructure laser was demonstrated[13-15] using the liquid-phase epitaxial technique[16] for the growth of GaAs and $Al_xGa_{1-x}As$ layers. However, these lasers operated in the pulsed mode. Further work led in 1970 to heterostructure lasers operating continuously at room temperature.[17,18] Notation such as (Ga,Al)As or AlGaAs/GaAs is often used to emphasize the heterostructure nature of these GaAs lasers. However, since homostructure lasers are no longer used, we shall simplify the notation in this book, whenever no confusion is likely to arise, by denoting a heterostructure laser only by the composition of its active layer.

Already in 1969 double-heterostructure GaAs lasers with a room-temperature value of $J_{th} \cong 5$ kA/cm² were reported.[15] This value was reduced[17] to about 1.6 kA/cm² in 1970, and by 1975 AlGaAs lasers with $J_{th} \cong 0.5$ kA/cm² were demonstrated using thin (~0.1 μm thick) active layers.[19] This was an improvement by more than two orders of magnitude over the simple homostructure lasers first made in 1962. It converted the semiconductor

laser from a laboratory curiosity to a practical, compact, coherent light source useful for numerous applications.

The physical reason for the reduction in the threshold current density with the use of a heterostructure device is twofold.[11,12] The cladding layers surrounding the active layer have a higher band gap and at the same time a lower refractive index compared with those of the active layer (see Fig. 1.2). The band-gap difference helps to confine electrons and holes to the active layer, where they recombine to produce the optical gain. At the same time the refractive-index difference confines the optical mode close to the active layer, which acts as a dielectric waveguide. The optical-mode confinement significantly reduces the internal loss[13] that would otherwise occur in the absence of index guiding due to the spreading of the optical mode in the lossy regions.

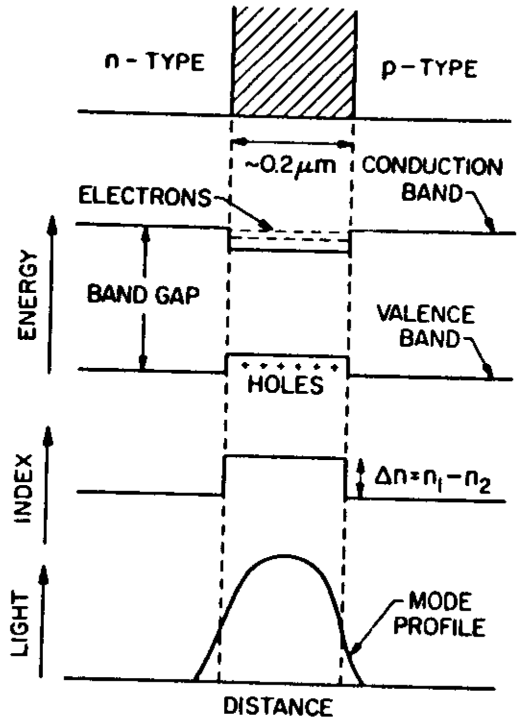


Fig. 1.2 Schematic illustration of the simultaneous confinement of the charge carriers and the optical mode to the active region occurring in a double-heterostructure semiconductor laser. The active layer has a lower band gap and a higher refractive index than those of the cladding layers. (After Ref. 24)

A double-heterostructure semiconductor laser such as shown in Fig. 1.1 is sometimes called a *broad-area laser* since it does not incorporate any mechanism for the lateral (parallel to the junction plane) confinement of the injected current or the optical mode. As early as 1967, stripe-geometry homostructure lasers were proposed[20] to limit the lateral spread of the injected carriers inside the active layer. In these lasers the current is injected over a narrow (~10 μm) central region using a stripe contact. The stripe geometry was adopted for heterostructure lasers[21] in 1971. Such lasers are also referred to as gain-guided since it is the lateral variation of the optical gain that confines the optical mode to the stripe vicinity.[22] By contrast, heterostructure lasers where the optical mode confinement occurs mainly through lateral variations of the refractive index are termed index-guided. A large number of index-guided structures have been proposed and demonstrated over the past decade. The literature is too vast to cite them individually, and a discussion of their relative merits can be found in several books[23-25] that also describe the research and development effort for AlGaAs lasers up to 1980.

GaAs lasers are of continued interest. Using the growth techniques of vapor-phase epitaxy[26] and molecular-beam epitaxy,[27] multiquantum-well laser structures have been developed.[28] In these devices the active region is not a single GaAs layer but rather consists of several ultrathin (~0.01 μm) layers composed alternatively of GaAs and AlGaAs materials. Recently the emphasis has shifted toward high-power GaAs lasers and phased-array semiconductor lasers.[29,30] In one approach, multiple stripes are used to generate distinct regions of optical gain in the junction plane and the near field consists of several spots. However, since the stripes are not widely separated, the optical field in the gain region of each emitter overlaps with that of the neighboring emitter. Such a coupling leads to a phase-locked array of emitters providing well-collimated high power output. Whereas the output power from a conventional GaAs laser is usually below 50 mW, more than 2 W of power has been obtained[30] by this technique of phase-locked arrays.

So far we have followed the development of GaAs lasers operating usually in the wavelength range of 0.8–0.9 μm. Long-wavelength semiconductor lasers in the range of 1.1–1.6 μm are of considerable interest for optical fiber communications.[31] Although several material systems were considered, the combination InGaAsP-InP turned out to be most suitable in view of its nearly perfect lattice match.[32] The active layer is composed of the $In_{1-x}Ga_xAs_yP_{1-y}$ quaternary alloy. By varing the mole fractions x and y, almost any wavelength in the 1.1–1.6-μm range can be selected. The cladding layers in this heterostructure laser (see Fig. 1.1) consist of either InP or InGaAsP itself with different mole fractions x and y.

Room-temperature operation of a 1.1-μm InGaAsP laser in the pulsed mode[33] was reported in 1975. The adoption of stripe geometry led to continuous operation[34] of such lasers in 1976. In 1977 the wavelength was extended[35] to 1.3 μm. Since low-loss dispersion-free fibers at 1.3 μm were already available,[36] considerable attention was focused to develop a practical InGaAsP laser at this wavelength. Motivated by the realization[37] of an ultra-low-loss (~0.2 dB/km) fiber at the 1.55-μm wavelength, several groups[38-43] in 1979 reported on InGaAsP lasers operating in the vicinity of 1.55 μm. Since then the development effort for InGaAsP lasers operating in the wavelength range of 1.3–1.6 μm has proceeded at an enormous pace.[44-46] The primary motivation is due to their application in optical fiber communications, and the use of InGaAsP lasers in long-haul optical communication systems has already reached the commercial stage.[47]

1.2 SEMICONDUCTOR MATERIALS

The previous section has followed the development of heterostructure lasers based on two semiconductor materials, AlGaAs and InGaAsP. However, in view of their potential application in such diverse fields as optical fiber communication,[31] optical data recording,[48] high-speed printing,[49] and molecular spectroscopy,[50] the list of semiconductor materials that have exhibited lasing action has continued to grow. Figure 1.3 shows the range of emission wavelengths for various semiconductor lasers. Taken together, these materials cover the optical spectrum from near ultraviolet to far infrared. Recently a semiconductor laser operating near 100 μm has been reported.[51]

The most important criterion in selecting the semiconductor material for a specific heterostructure laser is related to the quality of the heterojunction interface between the two semiconductors of different band gaps. To reduce the formation of lattice defects, the lattice constants of the two materials should match to better than 0.1%. Figure 1.4 shows the interrelationship between the band gap E_g and the lattice constant a for several ternary and quaternary compounds. Solid dots represent the binary compounds and solid lines correspond to the ternary compounds. The clear region bounded by the polygon (whose edges represent ternary compounds) denote the possible values of E_g and a for the quaternary solid solution of $In_{1-x}Ga_xAs_yP_{1-y}$ obtained by varying the mole fractions x and y. The dotted line shows the range of band-gap values that can be achieved by varying the compositions x and y to obtain a quaternary material that is lattice-matched to the binary InP. Figure 1.5 shows the constant band-gap contours (solid curves) in the x–y compositional plane. The dashed lines correspond to the fixed values of the lattice constant. For given values of E_g and a, the intersection of solid and dashed curves provides the compositional values x and y used to obtain the active-layer quaternary material. Since the photon energy $E = h\nu$ is approximately equal to the band-gap energy, the lasing wavelength λ

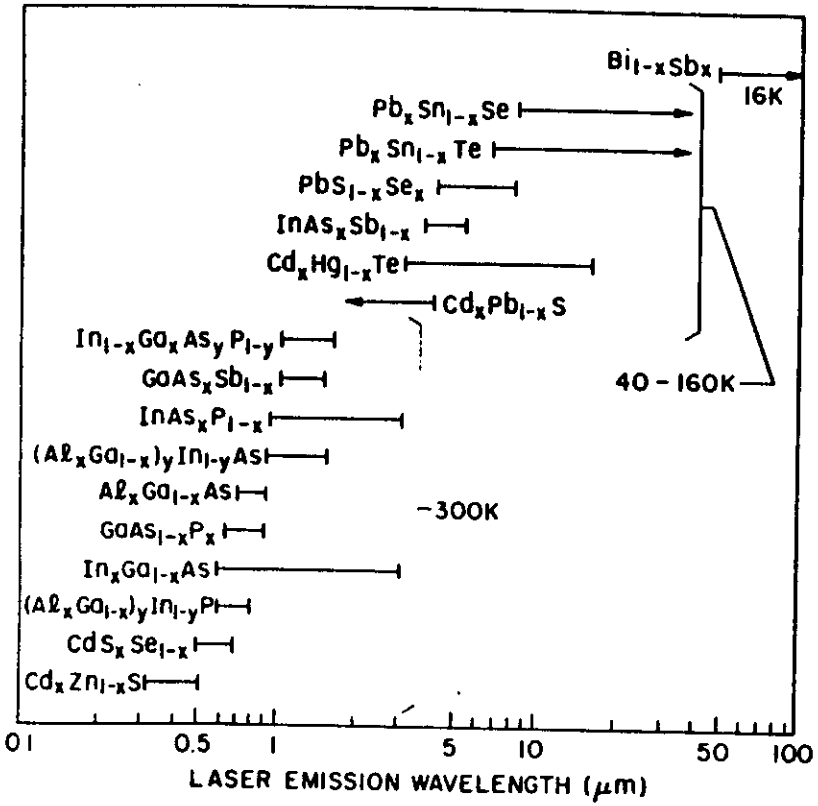


Fig 1.3 Wavelength range of semiconductor lasers covered by different material systems. Semiconductor lasers emitting at $\lambda > 3$ μm usually require low-temperature operation.

is obtained using $E_g = hc/\lambda$, where h is the Planck constant and c is the speed of light in vaccum. If E_g is expressed in electron volts, the lasing wavelength λ in micrometers is given by

$$\lambda \cong \frac{1.24}{E_g} \tag{1.2.1}$$

For $In_{1-x}Ga_xAs_yP_{1-y}$ lasers a wavelength range of 1.1–1.65 μm can be covered by choosing x and y according to Fig. 1.5 such that the active layer is lattice-matched to InP (a = 5.87 Å). Semiconductor lasers emitting at 1.3-μm and 1.55-μm wavelengths are of particular interest because of their application in optical fiber communications. From Fig. 1.5 the active-layer composition at 1.3 μm ($E_g \cong 0.95$ eV) corresponds to $x = 0.28$ and

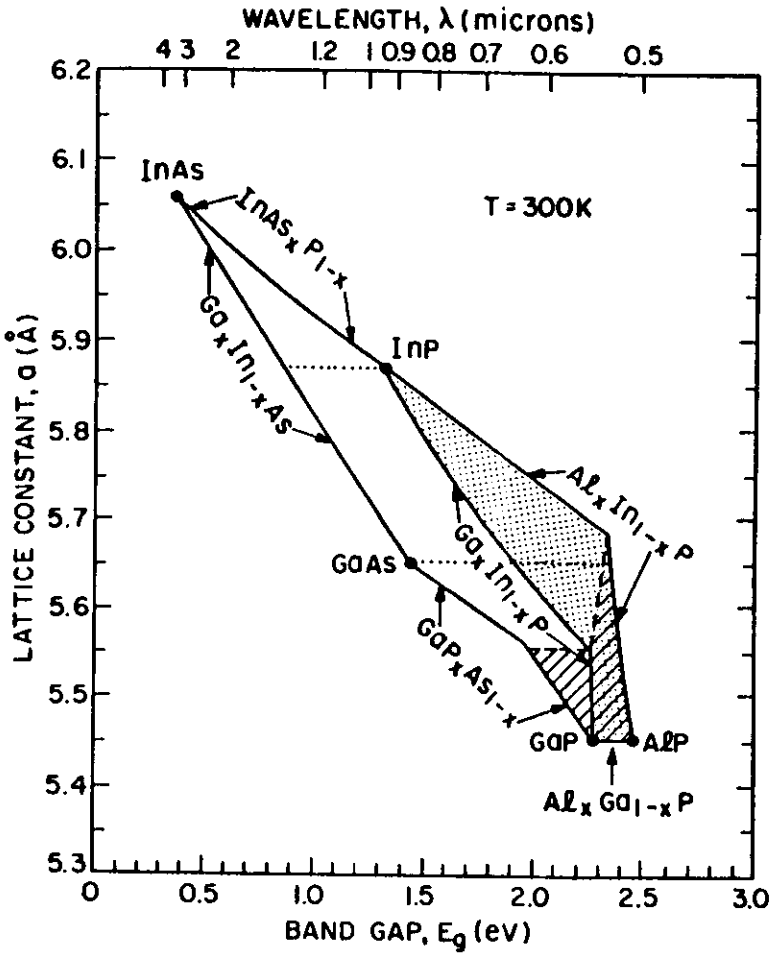


Fig. 1.4 Band gap and lattice constant for $In_{1-x}Ga_xAs_{1-y}P_y$ (clear region) and $(Al_xGa_{1-x})_yIn_{1-y}P$ (shaded regions) obtained by varying compositions x and y. Dashed lines separate indirect-band-gap regions (shown hatched). Dotted lines show the wavelength range (top scale) for a semiconductor laser whose quaternary active layer is lattice-matched to the binary compound. (After Ref. 24)

$y = 0.6$. The cladding-layer composition can also be chosen from Fig. 1.5; the only requirement is that its band gap should be somewhat larger than that of the active layer. Diagrams similar to Figs. 1.4 and 1.5 have been constructed for other heterostructure materials[24] in order to provide guidance for the lattic-matched growth of the active and cladding layers.

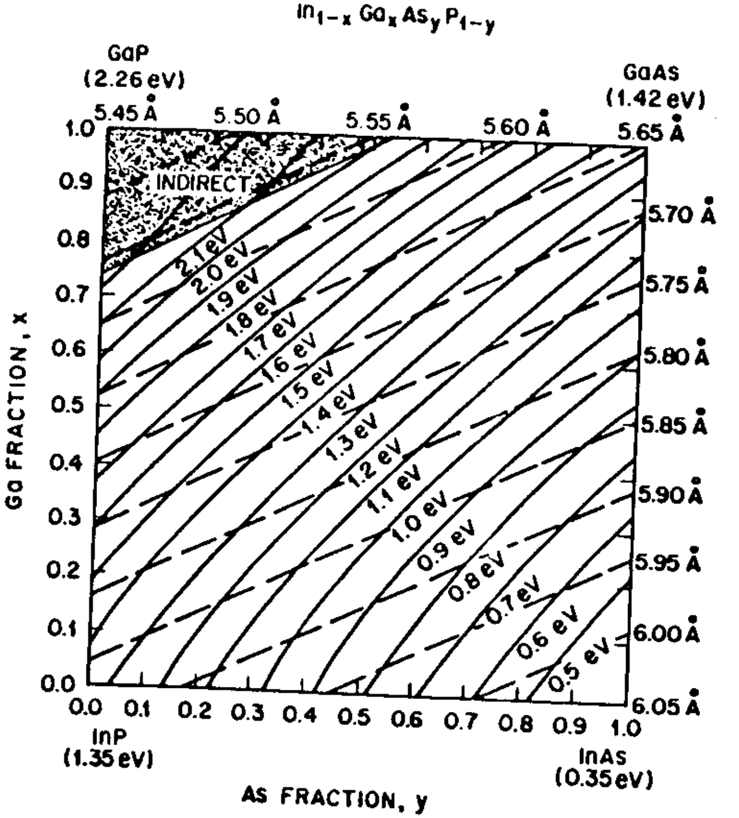


Fig. 1.5 Contours of constant band gap (solid lines) and constant lattice spacing (dashed lines) in the *x–y* compositional plane for $In_{1-x}Ga_xAs_yP_{1-y}$. The composition values *x* and *y* can be chosen to obtain a particular band gap (or laser wavelength) for a given lattice constant. Band gap is indirect in the shaded area. (After Ref. 32)

To cover the longer wavelength region (λ > 1.6 μm), other material systems shown in Fig. 1.3 have been successfully exploited. Lattice-matched active layers of the quaternary $In_xGa_{1-x}As_ySb_{1-y}$ material have been grown[52] on GaSb substrate and can cover the wavelength range of 1.7–4.4 μm. Another important class of materials is lead salts, which has been used to make semiconductor lasers emitting in the far-infrared region ranging from 3–34 μm,[53-57] although low-temperature operation is required. Two material systems of particular interest are PbTe-$Pb_{1-x}Sn_xTe$ and PbS-$PbS_{1-x}Se_x$. The longest wavelength of about 100 μm is obtained using the semiconductor material $Bi_{1-x}Sb_x$.[51]

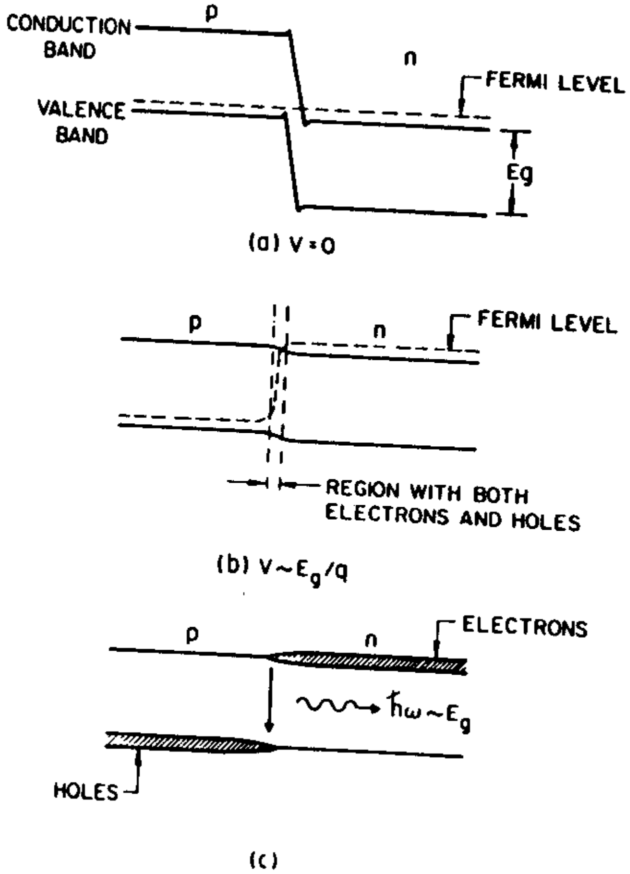
1.3 OPERATING PRINCIPLES

The purpose of this section is to provide a qualitative understanding of the physics behind the semiconductor laser. Most of the concepts introduced here are discussed in detail in subsequent chapters. Two things are required to operate a laser: (i) a gain medium that can amplify the electromagnetic radiation propagating inside it and provide the spontaneous-emission noise input and (ii) a feedback mechanism that can confine the electromagnetic field through the well-defined optical modes. As the name itself implies, the *gain medium* for a semiconductor laser consists of a semiconductor material (see Sec. 1.2). The optical feedback is obtained using the cleaved facets that form a Fabry-Perot (FP) cavity, and the mode confinement is achieved through dielectric waveguiding. To provide the optical gain, a semiconductor laser needs to be externally pumped, and both electrical and optical pumping techniques have been used for this purpose. A simple, practical, and most commonly used method employs current injection through the use of a forward-biased *p-n* junction. Such semiconductor lasers are sometimes referred to as injection lasers or laser diodes.

1.3.1 *p-n* Junction

At the heart of a semiconductor injection laser is the *p-n* junction.[58] A *p-n* junction is formed by bringing a *p*-type and an *n*-type semiconductor into contact with each other. When they first come in contact, their quasi-Fermi levels do not match since the two are not in equilibrium. An equilibrium is, however, quickly established through diffusion of electrons from the *n* side to the *p* side, while the reverse occurs for holes. These diffusing electrons and holes recombine in the junction region. Eventually a steady state is reached in such a way that further diffusion of electrons and holes is opposed by the built-in electric field across the *p-n* junction arising from the negatively charged acceptors on the *p* side and the positively charged donors on the *n* side. The Fermi level is then continuous across the *p-n* junction, as shown in Fig. 1.6 where the energy-band diagram of the *p-n homojunction* (junction between two similar semiconductors) is shown.

When a *p-n* junction is forward-biased by applying an external voltage, the built-in electric field is reduced, making possible a further diffusion of electrons and holes across the junction. As Fig. 1.6b shows, in a narrow depletion region both electrons and holes are present simultaneously and can recombine either radiatively or nonradiatively. Photons of energy $h\nu \cong E_g$ are emitted during radiative recombination. However, these photons can also be absorbed through a reverse process that generates electron-hole pairs. When the external voltage exceeds a critical value, a condition known



(a) $V = 0$

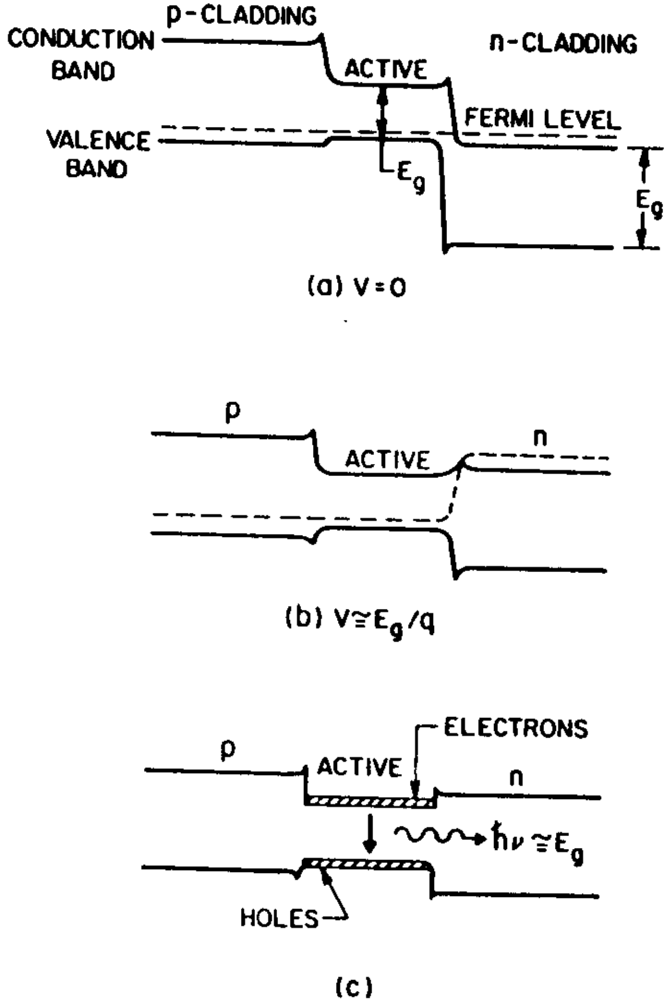
(b) $V \sim E_g/q$

(c)

Fig. 1.6 Energy-band diagram of a *p-n* junction at (a) zero bias and (b) forward bias. (c) Schematic representation of the electron and hole densities under forward bias. Radiative recombination of electrons and holes in the narrow overlapping region generates light.

as *population inversion* is achieved, in which the rate of photon emission exceeds that of absorption. The *p-n* junction is then able to amplify the electromagnetic radiation, whose wavelength satisfies Eq. (1.2.1), and is said to exhibit *optical gain*. However, for a homojunction the thickness of the region where gain is sufficiently high is very small (~0.01 μm) since there is no mechanism to confine the charge carriers.

The carrier-confinement problem is solved through the use of a *p-n heterojunction*. Figure 1.7 shows the energy-band diagram for a double-heterostructure laser wherein the thin *p*-type active region has a lower band gap compared to that of the two *p*-type and *n*-type cladding layers. Electrons and holes can move freely to the active region under forward bias. However, once there, they cannot cross over to the other side because of the potential barrier resulting from the band-gap difference. This allows for a substantial build-up of the electron and hole populations inside the active region, where they can recombine to produce optical gain. The width of the gain region is determined by the active-layer thickness, typically 0.1–0.3 μm. As mentioned



(a) $V = 0$

(b) $V \simeq E_g/q$

(c)

Fig. 1.7 Energy-band diagram of a double-heterostructure semiconductor laser at (a) zero bias and (b) forward bias. (c) The band-gap discontinuities at the two heterojunctions help to confine electrons and holes inside the active region, where they recombine to produce light.

earlier, it was the adoption of the heterostructure scheme that resulted in significantly lower threshold current densities (compared with a homojunction) and led to the room-temperature operation of semiconductor lasers.

1.3.2 Dielectric Waveguide

The successful operation of a laser requires that the generated optical field should remain confined in the vicinity of the gain region. In double-heterostructure semiconductor lasers the optical confinement occurs by virtue of a fortunate coincidence. The active layer with a smaller band gap also has a higher refractive index compared with that of the surrounding cladding layers (see Fig. 1.2). Because of the index difference, the active layer in effect acts as a dielectric waveguide. The physical mechanism behind the confinement is total internal reflection, as illustrated in Fig. 1.8. When a ray traveling at an angle θ (measured from the interface normal) hits the interface, it is reflected back if the angle θ exceeds the critical angle given by

$$\theta = \sin^{-1}\frac{\mu_1}{\mu_2}, \tag{1.3.1}$$

where μ_1 and μ_2 are the refractive indices of the cladding and active layers respectively. Thus, rays traveling nearly parallel to the interface are trapped

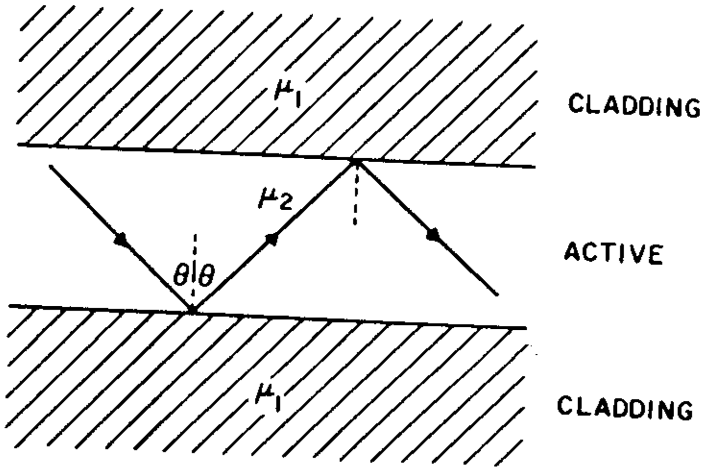


Fig. 1.8 Dielectric waveguiding in a heterostructure semiconductor laser. The relatively higher refractive index ($\mu_2 > \mu_1$) of the active layer allows total internal reflection to occur at the two interfaces for angles such that $\sin\theta > \mu_1/\mu_2$.

and constitute the waveguide mode. A more detailed discussion of waveguide modes requires the use of Maxwell's equations and is given in Chapter 2.

1.3.3 Recombination Mechanisms

When the current flowing through a semiconductor laser is increased, charge carriers (electrons and holes) are injected into the thin active region, where they recombine through radiative or nonradiative mechanisms. As one may expect, nonradiative recombinations are not helpful for laser operation, and attempts are made to minimize their occurrence by controlling point defects and dislocations. However, a nonradiative recombination mechanism, known as the *Auger process*, is intrinsic and becomes particularly important for long-wavelength semiconductor lasers operating at room temperature and above. It is described in detail in Chapter 3. Physically speaking, during the Auger process the energy released by the electron-hole recombination is taken by a third charge carrier and is eventually lost to lattice phonons.

During a radiative recombination, the energy E_g released by the electron-hole pair appears in the form of a photon whose frequency ν or wavelength λ satisfies the energy conservation relation $E_g = h\nu = hc/\lambda$. This can happen through two optical processes known as spontaneous-emission and stimulated emission. These are shown schematically in Fig. 1.9. In the case of *spontaneous*

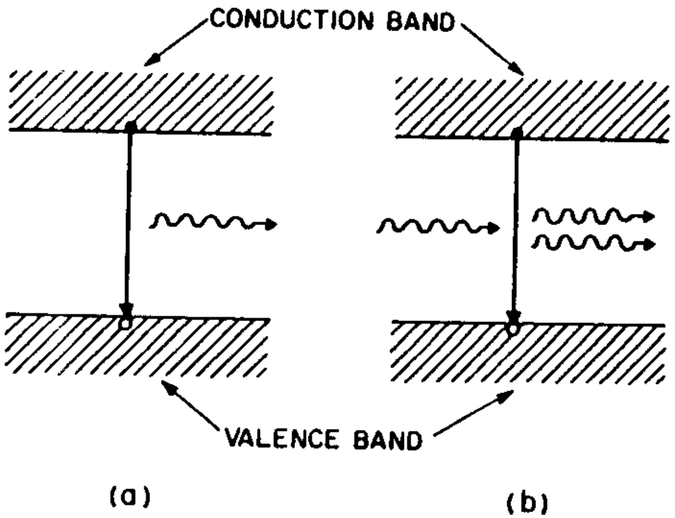


Fig. 1.9 Schematic illustration of (a) spontaneous-emission and (b) stimulated-emission processes wherein an electron-hole pair recombines to generate a photon. In the case of stimulated emission the two outgoing photons match in their frequency and direction of propagation.

emission, photons are emitted in random directions with no phase relationship among them. *Stimulated emission*, by contrast, is initiated by an already existing photon. The remarkable feature is that the emitted photon matches the original photon not only in its wavelength but also in direction of propagation. It is this relationship between the incident and emitted photons that renders the light emitted by a laser coherent.

1.3.4 Laser Threshold

Although stimulated emission can occur as soon as current is applied to the semiconductor laser, the laser does not emit coherent light until the current exceeds a critical value, known as the *threshold current* (I_{th}). This is so because stimulated emission has to compete against the absorption processes during which an electron-hole pair is generated at the expense of an absorbed photon. Since the electron population in the valence band generally far exceeds that of the conduction band, absorption dominates. At a certain value of the external current, a sufficient number of electrons are present in the conduction band to make the semiconductor optically transparent. With a further increase in current, the active region of the semiconductor laser exhibits optical gain and can amplify the electromagnetic radiation passing through it. Spontaneously emitted photons serve as the noise input for the amplification process.

However, optical gain alone is not enough to operate a laser. The other necessary ingredient is *optical feedback*. In semiconductor lasers it is provided by the cleaved facets that form an *FP cavity*. The role of the FP cavity is twofold. First, it provides a direction selectivity for the process of stimulated emission, since only photons traveling along its axis are reflected back and forth. Second, it provides a wavelength selectivity since the feedback is strongest for wavelengths corresponding to the longitudinal modes of the FP cavity.

Because of the optical feedback, the number of photons traveling perpendicular to the facets increases when the current is large enough to satisfy the condition of net stimulated emission. However, some photons are lost through the partially transmitting facets and some get scattered or absorbed inside the cavity. If the loss exceeds the gain, stimulated emission cannot sustain a steady supply of photons. This is precisely what happens below threshold, when the laser output consists of mainly spontaneously emitted photons. At threshold, gain equals loss and stimulated emission begins to dominate. Over a narrow current range in the vicinity of the threshold current, the output power jumps by several orders of magnitude and the spectral width of the emitted radiation narrows considerably because of the coherent nature of stimulated emission.

In the above-threshold regime, laser output increases almost linearly with the current. Almost all electrons and holes injected into the active region now recombine through stimulated emission, and the internal quantum efficiency approaches 100%. The performance of a semiconductor laser is governed by a large number of emission characteristics related to the static, dynamic, and spectral behavior of the light output. These are discussed in Chapter 6.

RECOMBINATION MECHANISMS IN SEMICONDUCTORS

3.1 INTRODUCTION

This chapter describes the electron-hole recombination mechanisms in a direct-band-gap semiconductor. Recombination mechanisms can in general be classified into two groups, radiative and nonradiative. *Radiative recombination* occurs when an electron in the conduction band recombines with a hole in the valence band and the excess energy is emitted in the form of a photon. Radiative recombination is thus the radiative transition of an electron in the conduction band to an empty state (hole) in the valence band. The optical processes associated with radiative transitions are (i) spontaneous emission, (ii) absorption or gain, and (iii) stimulated emission. Stimulated emission, in which the emitted photon has nearly the same energy and momentum as the incident photon, forms the basis for laser action. The concept of stimulated emission dates back to the work of Einstein[1] in 1917.

In thermal equilibrium, a direct-band-gap semiconductor (e.g., GaAs, InP, or GaSb) has a few electrons in the conduction band and a few holes (empty electron states) in the valence band. When a photon of energy greater than the band gap passes through such a semiconductor, the photon has a high probability of being absorbed, giving its energy to an electron in the valence band, thereby raising the electron to the conduction band. In principle, such a photon could stimulate the emission of an identical photon with the transition of an electron from the conduction to the valence band. The emitted photon derives its energy from the energy lost by the electron. In thermal equilibrium the number of electrons in the conduction band is very small ($\sim 10^6\ \mathrm{cm}^{-3}$ for GaAs), so the probability of stimulated emission is negligible compared to the probability for absorption. However, external excitation, can sufficiently increase the number of electrons in the conduction band such that the probability of stimulated emission eventually becomes higher than the probability of absorption. This situation corresponds to population

inversion in a laser medium and is necessary for optical gain. The external excitation which generates a high density of electron-hole pairs in a semiconductor is usually provided by current injection. It can also be achieved by optical pumping (absorption of radiation higher in energy than the band gap). In Section 3.2 we discuss the absorption and emission rates. These are then used to study the dependence of optical gain on the injected current and the carrier density.

Nonradiative recombination of an electron-hole pair, as the name implies, is characterized by the *absence* of an emitted photon in the recombination process. This of course makes the experimental study, and hence identification, of such processes very difficult. In indirect-band-gap semiconductors such as Ge or Si, the probability of nonradiative recombination dominates that of radiative recombination by several orders of magnitude. The measurable quantities associated with nonradiative recombinations are the internal quantum efficiency and the carrier lifetime. The variation of these quantities with parameters like temperature, pressure and carrier concentration is, by and large, the only way to identify a particular nonradiative recombination process. One of the effects of nonradiative recombination on the performance of injection lasers is to increase the threshold current.

The nonradiative recombination processes that affect the performance of long-wavelength semiconductor lasers are Auger recombination, surface recombination, and recombination at defects. The Auger recombination mechanism involves four particle states (three electrons and one hole) and is believed to be important at high temperatures and for low band-gap semiconductors. Section 3.3 gives a detailed discussion of the various Auger recombination mechanisms. Section 3.4 presents the experimental results on the measurement of radiative and nonradiative recombination coefficients. Section 3.5 considers the threshold current density and the contribution of carrier leakage to it. Finally, in Sec. 3.6 we discuss the temperature dependence of the threshold current.

3.2 RADIATIVE RECOMBINATION

The processes associated with the radiative recombination of electron-hole pairs in semiconductors are spontaneous emission, optical absorption or gain, and stimulated emission. The rates of these processes are related to each other by the Einstein relations. These relations were first derived for electronic transitions between two discrete energy levels (i.e., in a gaseous medium). Similar relations also hold for a semiconductor where the recombining electrons and holes can occupy a continuous band of energy eigenstates.[2]

Figure 3.1 shows a simplified energy-versus-wave-vector diagram for a direct-band-gap semiconductor. An accurate description of the band structure

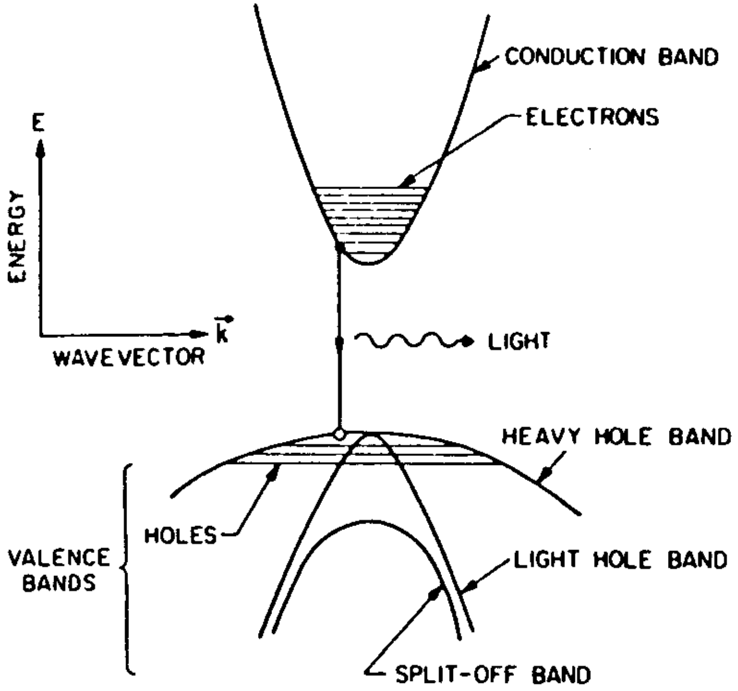


Fig. 3.1 Energy versus wave-vector diagram of a direct-band-gap semiconductor showing schematically the conduction and valence bands. Three valence bands are required to model band-to-band transitions realistically. Horizontal lines show the filled energy states. Radiative recombination of electrons and holes generates photons.

requires sophisticated numerical techniques. A commonly used approximation of the exact band structure in a direct-gap semiconductor is the parabolic band model. In this model the energy-versus-wave vector (E versus **k**) relation is assumed to be parabolic, this is,

$$E_c = \frac{\hbar^2 k^2}{2m_c} \quad \text{for electrons} \tag{3.2.1a}$$

$$E_v = \frac{\hbar^2 k^2}{2m_v} \quad \text{for holes} \tag{3.2.1b}$$

where m_c and m_v are the effective masses of electrons and holes, respectively, and k is the magnitude of the wave vector **k**.

In a direct-gap semiconductor, the minimum in the conduction band curve and the maximum in the valence band curve occur at the same value of the wave vector **k** (**k** = 0 in Fig. 3.1). Since a photon carries negligible momentum compared with the carrier momentum $\hbar\mathbf{k}$, radiative transitions occur between free electrons and free holes of essentially identical wave vectors.

At a given temperature T, the available number of electrons and holes are distributed over a range of energies. The occupation probability f_c of an electron with energy E_c is given by the Fermi-Dirac statistics, and is

$$f_c(E_c) = \frac{1}{\exp[(E_c - E_{fc})/k_B T] + 1} \tag{3.2.2}$$

where E_{fc} is the quasi-Fermi level for the conduction band. Similarly for the valence-band holes, the occupation probability of a hole with energy E_v is

$$f_v(E_v) = \frac{1}{\exp[(E_v - E_{fv})/k_B T] + 1} \tag{3.2.3}$$

where E_{fv} is the quasi-Fermi level for the valence band. The notations used here assume that E_c and E_{fc} are measured from the conduction-band edge and are positive into the conduction band. On the other hand, E_v and E_{fv} are measured from the valence-band edge and are positive into the valence band. Note that f_v represents the occupation probability of the hole and not of the electron.

Consider the transition shown in Fig. 3.1 in the presence of an incident photon whose energy $E = h\nu = E_c + E_v + E_g$, where E_g is the band gap. The photon can be absorbed creating an electron of energy E_c and a hole of energy E_v. The absorption rate is given by

$$R_a = B(1 - f_c)(1 - f_v)\rho(E) \tag{3.2.4}$$

where B is the transition probability, $\rho(E)$ is the density of photons of energy E, and the factors $1 - f_c$ and $1 - f_v$ represent the probabilities that the electron and hole states of energy E_c and E_v are not occupied. The stimulated-emission rate of photons is given by

$$R_e = Bf_cf_v\ \rho(E) \tag{3.2.5}$$

where the Fermi factors f_c and f_v are the occupation probabilities of the electron and hole states of energy E_c and E_v respectively. The stimulated-emission process is accompanied by the recombination of an electron-hole pair. The condition for net stimulated emission or optical gain is

$$R_e > R_a \tag{3.2.6}$$

Using Eqs. (3.2.4) and (3.2.5), this condition becomes

$$f_c + f_v > 1$$

With the use of f_c and f_v given by Eqs. (3.2.2) and (3.2.3), this becomes

$$E_{fc} - E_{fv} > E_c + E_v \tag{3.2.7}$$

By adding E_g to both sides and noting that $E_c + E_v + E_g = h\nu$, it follows that the separation of the quasi-Fermi levels must exceed the photon energy in order for the stimulated-emission rate to exceed the absorption rate. This is the necessary condition for net stimulated emission or optical gain.[3]

3.2.1 Absorption and Emission Rates for Discrete Levels

The spontaneous-emission rate and the absorption rate for photons in a semiconductor can be calculated using time-dependent perturbation theory and summing over the available electron and hole states. We first consider the case of two discrete energy levels of energy E_i and E_f. The electron makes a transition from an initial state of energy E_i to a final state of energy E_f, and in the process a photon of energy $E = E_i - E_f$ is emitted. The transition probability or the emission rate W of such a process is given by Fermi's golden rule[4]:

$$W = \frac{2\pi}{\hbar}|H'_{if}|^2\, \rho(E_f)\, \delta(E - E_i + E_f) \tag{3.2.8}$$

where H'_{if} is the matrix element $\langle i|H'|f\rangle$ of the time-independent part of the perturbation Hamiltonian H_i between the initial state $|i\rangle$ and final state $|f\rangle$, and $\rho(E_f)$ is the density of the final states. In obtaining Eq. (3.2.8), the perturbation Hamiltonian is assumed to be of the form

$$H_i = 2H' \sin(\omega t)$$

The time-dependent sinusoidal part gives rise to the delta function of Eq. (3.2.8) arising from the energy conservation during the emission process.

For photons interacting with free electrons of mass m_0, the perturbation Hamiltonian is given by[5]

$$H_i = -\frac{q}{m_0}\mathbf{A}\cdot\mathbf{p} \tag{3.2.9}$$

where $\mathbf{A}$ is the vector potential of the electromagnetic field and $\mathbf{p}$ is the electron-momentum vector. The vector potential is related to the electric field $\mathcal{E}$ by the equation

$$\mathcal{E} = -\frac{\partial \mathbf{A}}{\partial t} \tag{3.2.10}$$

For a traveling plane wave with the angular frequency ω, the electric field is written as

$$\mathcal{E} = E_0\hat{\epsilon}\cos(\omega t - \mathbf{k}\cdot\mathbf{r}) \tag{3.2.11}$$

where $\hat{\epsilon}$ is a unit vector determining the polarization of the field, E_0 is its amplitude, and $\mathbf{k}$ denotes the propagation vector such that $\mathbf{k}\times\hat{\epsilon} = 0$. The magnitude of the electric field E_0 associated with one photon can be obtained as follows. The magnitude of the Poynting vector that represents the energy flux is given by

$$S = |\langle \mathcal{E} \times \mathcal{H} \rangle| \tag{3.2.12}$$

where angular brackets denote time averaging and $\mathcal{H}$ is the magnetic field, which can be obtained using Maxwell's equations (see Sec. 2.2). Using them, we obtain

$$S = \frac{1}{2}E_0^2\bar{\mu}\epsilon_0 c \tag{3.2.13}$$

where $k = \bar{\mu}\omega/c$ and $\bar{\mu}$ is the refractive index of the medium. The factor 1/2 arises from time averaging over one optical cycle. The energy flux per photon is also given by the product of the photon energy $\hbar\omega$ and the group velocity $c/\bar{\mu}$. Hence it follows that

$$S = \hbar\omega c/\bar{\mu} \tag{3.2.14}$$

From Eqs. (3.2.10), (3.2.13), and (3.2.14), the vector potential $\mathbf{A}$ is given by

$$\mathbf{A} = -\hat{\epsilon}\left(\frac{2\hbar}{\epsilon_0\bar{\mu}^2\omega}\right)^{1/2}\sin(\omega t - \mathbf{k}\cdot\mathbf{r}) \tag{3.2.15}$$

and using Eq. (3.2.9) the perturbation Hamiltonian becomes

$$H_I = \frac{q}{m_0}\left(\frac{2\hbar}{\epsilon_0\bar{\mu}^2\omega}\right)^{1/2}(\hat{\epsilon}\cdot\mathbf{p})\sin(\omega t - \mathbf{k}\cdot\mathbf{r}) \tag{3.2.16}$$

The square of the transition matrix element H'_{if} is now given by

$$|H'_{if}|^2 = \frac{q^2}{m_0^2}\frac{2\hbar}{\epsilon_0\bar{\mu}^2\omega}\frac{1}{4}|\langle i|\hat{\epsilon}\cdot\mathbf{p}|f\rangle|^2 \tag{3.2.17}$$

where we have made the usual assumption that the electron wave functions spread over a linear dimension shorter than the wavelength ($r \ll \lambda$) so that $\mathbf{k}\cdot\mathbf{r} \ll 1$.

For stimulated emission or absorption, $\rho(E_f) = 1$, and the absorption or emission rate at a photon energy $E = \hbar\omega$ for transitions between two discrete levels obtained from Eqs. (3.2.8) and (3.2.17) is given by

$$W = \frac{\pi q^2}{m_0^2\epsilon_0\bar{\mu}^2\omega}|M_{if}|^2\,\delta(E_i - E_f - \hbar\omega) \tag{3.2.18}$$

where $M_{if} = \langle i|\hat{\epsilon}\cdot\mathbf{p}|f\rangle$ is the momentum matrix element between the initial and final electron states. In the case of absorption, the number of photons absorbed per second is W. Since a photon travels a distance $c/\bar{\mu}$ in 1 s, the number of photons absorbed per unit distance or the absorption coefficient $\alpha = \bar{\mu}W/c$. Using Eq. (3.2.18), the absorption coefficient for photons of energy $E = \hbar\omega$ becomes

$$\alpha(E) = \frac{q^2\hbar}{2\epsilon_0 m_0^2 c\bar{\mu}E}|M_{if}|^2\,\delta(E_i - E_f - E) \tag{3.2.19}$$

For spontaneous emission, the quantity $\rho(E_f)$ in Eq. (3.2.8) equals the number of states for photons of energy E per unit volume per unit energy. It is given by

$$\rho(E_f) = (2)\frac{4\pi k^2}{(2\pi)^3}\frac{dk}{dE} \tag{3.2.20}$$

where the factor 2 arises from two possible polarizations of the electromagnetic field. Since $E = \hbar\omega$ and $k = \bar{\mu}\omega/c$, Eq. (3.2.20) reduces to

$$\rho(E_f) = \frac{\bar{\mu}^3\omega^2}{\pi^2c^3\hbar} \tag{3.2.21}$$

Thus using Eqs. (3.2.8), (3.2.17), and (3.2.21), the spontaneous-emission rate per unit volume at the photon energy E is given by

$$r_{sp}(E) = \frac{4\pi q^2\bar{\mu}E}{m_0^2\epsilon_0 c^3\hbar^2}|M_{if}|^2\,\delta(E_i - E_f - E) \tag{3.2.22}$$

3.2.2 Absorption and Emission Rates in Semiconductors

We now calculate the spontaneous-emission rate and the absorption coefficient for direct-gap semiconductors. These quantities are obtained by integrating Eqs. (3.2.19) and (3.2.22) over the occupied electron and hole states in a semiconductor.[6] A realistic band model for a III–V direct-gap semiconductor is the four-band model[7] shown in Fig. 3.1. In this model the valence band is represented by three subbands: the heavy-hole band, the light-hole band, and the spin-split-off-hole band. The light- and heavy-hole bands are degenerate at $\mathbf{k} = 0$. Usually the split-off energy Δ is large compared with the thermal energy k_BT; hence the split-off band is full of electrons, that is, no holes are present. The quasi-Fermi levels for the conduction and valence bands can be obtained from the known effective masses and the density of states in each band. For the conduction band, E_{fc} is obtained using

$$n = \int \frac{\rho_c(E)\,dE}{1 + \exp[(E - E_{fc})/k_BT]} \tag{3.2.23}$$

where $\rho_c(E)$ is the conduction-band density of states and n is the number of electrons in the conduction band. The density of states for electrons of energy E can be obtained using Eq. (3.2.20) with the dispersion relation (3.2.1a) and becomes

$$\rho_c(E) = (2)\frac{4\pi k^2}{(2\pi)^3}\frac{dk}{dE} = 4\pi\left(\frac{2m_c}{h^2}\right)^{3/2}E^{1/2} \tag{3.2.24}$$

where $E = \hbar^2k^2/2m_c$, and m_c is the conduction-band effective mass. The factor 2 in Eq. (3.2.24) arises from the two electronic spin states. Equation (3.2.23) can be rewritten as

$$n = N_c\frac{2}{\pi^{1/2}}\int\frac{\epsilon^{1/2}d\epsilon}{1 + \exp(\epsilon - \epsilon_{fc})} \tag{3.2.25}$$

where $N_c = 2(2\pi m_c k_B T/h^2)^{3/2}$ and $\epsilon_{fc} = E_{fc}/k_B T$. Equation (3.2.25) can be used to calculate ϵ_{fc} for a given electron concentration. An useful approximation for ϵ_{fc} is given by Joyce and Dixon.[8] They show that ϵ_{fc} can be represented by a convergent series of the form

$$\epsilon_{fc} = \ln\left(\frac{n}{N_c}\right) + \sum_{i=1}^{\infty} A_i \left(\frac{n}{N_c}\right)^i \tag{3.2.26}$$

where the first few coefficients are

$$\begin{aligned} A_1 &= 3.53553 \times 10^{-1} \\ A_2 &= -4.95009 \times 10^{-3} \\ A_3 &= 1.48386 \times 10^{-4} \\ A_4 &= 4.42563 \times 10^{-6} \end{aligned}$$

Another approximation, although not as elegant but showing faster convergence, is the Nilson approximation.[9]

For a nondegenerate electron gas ($n << N_c$), all terms except the first in Eq. (3.2.26) can be neglected and the Fermi energy is given by

$$E_{fc} = k_B T \ln\left(\frac{n}{N_c}\right) \tag{3.2.27}$$

Using Eq. (3.2.27) in Eq. (3.2.2) and neglecting 1 in the denominator, the occupation probability becomes

$$f_c(E) \cong \frac{n}{N_c} \exp\left(\frac{-E}{k_B T}\right) \tag{3.2.28}$$

This is often referred to as the *Boltzmann approximation*. The use of Eq. (3.2.28) simplifies considerably the calculation of recombination rates in semiconductors. The nondegenerate approximation $n << N_c$, however, does not always hold for semiconductor lasers and should be used with caution.

The quasi-Fermi level for the holes in the valence band can be similarly calculated. The hole density is given by

$$p = \sum_{i=l,h} \int \frac{\rho_{vi}(E)\, dE}{1 + \exp[(E - E_{fv})/k_B T]} \tag{3.2.29}$$

where p is the density of holes and the summation is over the light- and heavy-hole bands. Using an expression similar to Eq. (3.2.24) for the density of states, we obtain

$$p = N_r \frac{2}{\pi^{1/2}} \int \frac{\epsilon^{1/2} d\epsilon}{1 + \exp(\epsilon - \epsilon_{fr})} \tag{3.2.30}$$

where

$$N_r = 2(2\pi k_B T/h^2)^{3/2}(m_{lh}^{3/2} + m_{hh}^{3/2})$$

and m_{lh} and m_{hh} are the effective masses of light and heavy holes, respectively. For the nondegenerate case where $p << N_r$, Eq. (3.2.30) can be simplified to obtain the following expressions for the quasi-Fermi energy and the hole-occupation probability:

$$E_{fv} = k_B T \ln\left(\frac{p}{N_v}\right) \tag{3.2.31a}$$

$$f_v(E) = \frac{p}{N_v} \exp\left(\frac{-E}{k_B T}\right) \tag{3.2.31b}$$

So far we have considered the idealized case of parabolic bands with the density of states given by Eq. (3.2.24). Under high injected carrier densities or for high doping levels, the density of states is modified. In the case of heavy doping, this occurs because randomly distributed impurities create an additional continuum of states near the band edge. These are called *band-tail states*[10,11] and are schematically shown in Fig. 3.2.

Several models of band-tail states exist in the literature. Principal among them are the Kane model[10] and the Halperin-Lax model.[11] Fig. 3.3 shows a comparison of the calculated density of states for the two models. The curves were calculated by Hwang[12] for p-type GaAs with a hole concentration of 3×10^{18} cm^{-3}. The material was heavily compensated with acceptor and donor concentrations N_A and N_D of 6×10^{18} cm^{-3} and 3×10^{18} cm^{-3}, respectively. Figure 3.3 shows that for the Kane model the band tails extend much more into the band gap than for the Halperin-Lax model. The absorption measurements on heavily compensated samples indicate that the Halperin-Lax model is more realistic. However, this model is not valid for states very close to the band edge. To remedy the situation, Stern[13] approximated the density of states by fitting a density-of-state equation of the Kane form (which is Gaussian) to the Halperin-Lax model of band-tail states. The fit is done at *one* energy value in the band tail, which lies within the validity of the Halperin-Lax model. The Kane form of band tail merges with the parabolic band model above the band edge as seen in Fig. 3.3

The importance of band-tail states in the operation of semiconductor lasers arises from two considerations. First, band-tail states can contribute signifi-

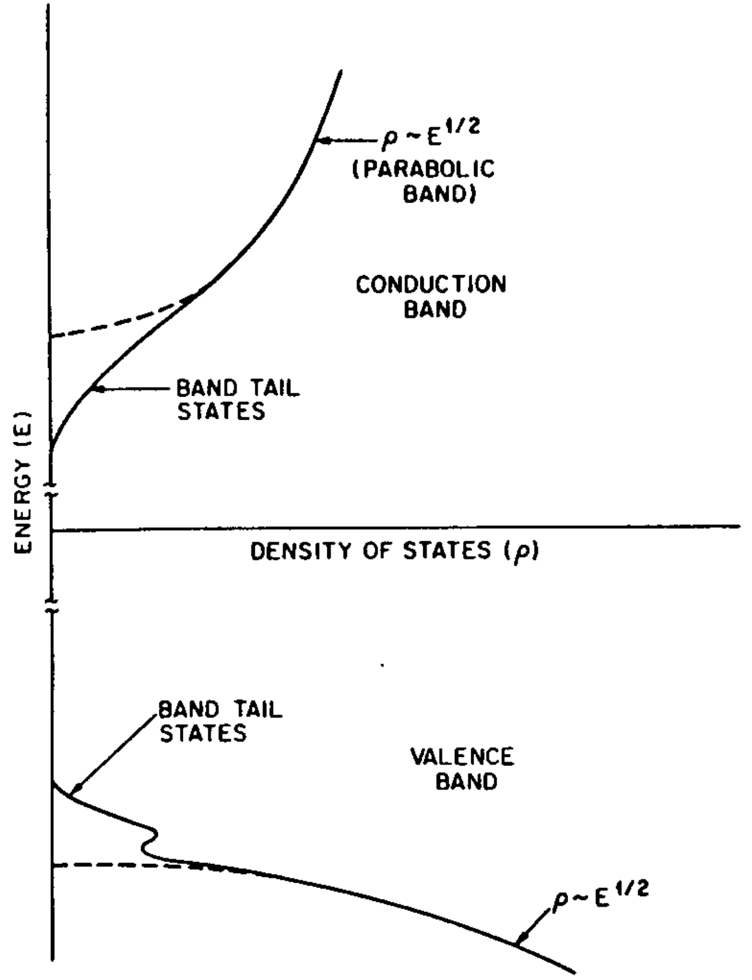


Fig. 3.2 Energy for the conduction and valence bands plotted against density of states. Dashed curves are for the parabolic-band model, while solid curves show schematically the effect of band-tail states.

cantly to the total spontaneous-emission rate and hence to the injection current needed to reach threshold. Second, band-tail transitions can have significant optical gain so that the laser can lase on these transitions, especially for heavily doped semiconductors at low temperatures.

The maxtrix element for transitions between band-tail states differs from that involving free electron and hole states in one important aspect. The former does not satisfy momentum conservation; i.e., since the electron and holes states are not states of definite momentum, the k-selection rule does not apply. By contrast, for transitions involving parabolic band states, the initial and final particle states obey the k-selection rule. Thus the exact matrix element for optical transitions in heavily compensated semiconductors must

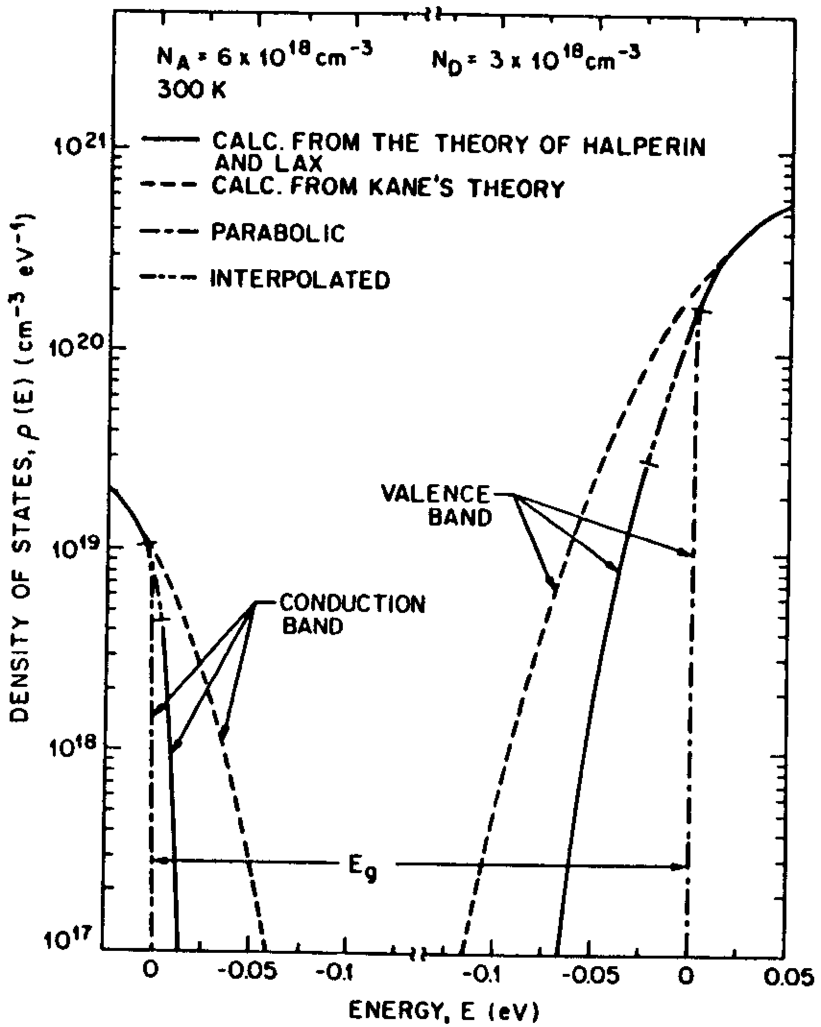


Fig. 3.3 Calculated density of states versus energy in the conduction and valence bands for p-type GaAs with a net carrier concentration of 3×10^{18} cm^{-3}. The acceptor concentration $N_A = 6 \times 10^{18}$ cm^{-3}, and the donor concentration $N_D = 3 \times 10^{18}$ cm^{-3}. (After Ref. 12)

not invoke the k-selection rule for transitions involving band-tail states and, at the same time, should extrapolate to that obeying the k-selection rule for above-band-edge transitions involving parabolic band states. The model for such a matrix element has been considered by Stern.[14-16] Before discussing his model in the next section, we consider the simpler case where the k-selection rule holds and obtain the absorption coefficient and the spontaneous emission rate.

The k-Selection Rule When the k-selection rule is obeyed, $|M_{if}|^2 = 0$ unless $\mathbf{k}_c = \mathbf{k}_v$. If we consider a volume V of the semiconductor, the matrix element $|M_{if}|^2$ is given by

$$|M_{if}|^2 = |M_b|^2 \frac{(2\pi)^3}{V}\, \delta(\mathbf{k}_c - \mathbf{k}_v). \tag{3.2.32}$$

The δ function accounts for the momentum conservation between the conduction-band and valence-band states. The quantity $|M_b|$ is an average matrix element for the Bloch states. Using the Kane model,[7] $|M_b|^2$ in bulk semiconductors is given by [7,17]

$$|M_b|^2 = \frac{m_0^2 E_g (E_g - \Delta)}{12 m_c (E_g - 2\Delta/3)} = \xi m_0 E_g \tag{3.2.33}$$

where m_0 is the free-electron mass, E_g is the band gap, and Δ is the spin-orbit splitting. For GaAs, using E_g = 1.424 eV, Δ = 0.33 eV, m_c = 0.067 m_0, we get $\xi = 1.3$.

We are now in a position to calculate the spontaneous-emission rate and the absorption coefficient for a bulk semiconductor. Equations (3.2.22) and (3.2.32) can be used to obtain the total spontaneous-emission rate per unit volume. Summing over all states in the band, we obtain

$$r_{sp}(E) = \frac{4\pi\bar{\mu}q^2E}{m_0^2\epsilon_0 h^2 c^3} |M_b|^2 \frac{(2\pi)^3}{V} (2) \left(\frac{V}{(2\pi)^3}\right)^2 \frac{1}{V} \times \Sigma \int\cdots\int f_c(E_c) f_v(E_v)\, d^3\mathbf{k}_c\, d^3\mathbf{k}_v\, \delta(\mathbf{k}_c - \mathbf{k}_v)\delta(E_i - E_f - E) \tag{3.2.34}$$

where f_c and f_v are the Fermi factors for electrons and holes. The factor 2 arises from the two spin states. In Eq. (3.2.34), Σ stands for the sum over the three valence bands (see Fig. 3.1). For definiteness, we first consider transitions involving electrons and heavy holes. The integrals in Eq. (3.2.34) can be evaluated with the following result:

$$r_{sp}(E) = \frac{2\bar{\mu}q^2E|M_b|^2}{\pi m_0^2 \epsilon_0 h^2 c^3} \left(\frac{2m_r}{h^2}\right)^{3/2} (E - E_g)^{1/2} f_c(E_c) f_v(E_v) \tag{3.2.35}$$

where

$$E_c = \frac{m_r}{m_c}(E - E_g)$$

$$E_v = \frac{m_r}{m_{hh}}(E - E_g)$$

$$m_r = \frac{m_c m_{hh}}{m_c + m_{hh}}$$

and m_{hh} is the effective mass of the heavy hole. Equations (3.2.35) gives the spontaneous-emission rate at the photon energy E. To obtain the total spontaneous-emission rate, a final integration should be carried out over all possible energies. Thus the total spontaneous-emission rate per unit volume due to electron-heavy-hole transitions is given by

$$R = \int_{E_g}^{\infty} r_{sp}(E)\, dE = A|M_b|^2 I \tag{3.2.36}$$

where

$$I = \int_{E_g}^{\infty} (E - E_g)^{1/2} f_c(E_c) f_v(E_v)\, dE$$

and A represents the remaining constants in Eq. (3.2.35). A similar equation holds for the electron–light-hole transitions if we replace m_{hh} by the effective light-hole mass m_{lh}.

The absorption coefficient $\alpha(E)$ can be obtained in a similar way using Eq. (3.2.19) and (3.2.32) and integrating over the available states in the conduction and valence bands. The resulting expression is

$$\alpha(E) = \frac{q^2 h}{2\epsilon_0 m_0^2 c\bar{\mu}E} |M_b|^2 \frac{(2\pi)^3}{V} (2) \left(\frac{V}{(2\pi)^3}\right)^2 \left(\frac{1}{V}\right) \times \int\cdots\int (1 - f_c - f_v)\, d^3\mathbf{k}_c d^3\mathbf{k}_v\, \delta(\mathbf{k}_c - \mathbf{k}_v)\, \delta(E_i - E_f - E) \tag{3.2.37}$$

$$= \frac{q^2 h|M_b|^2}{4\pi^2\epsilon_0 m_0^2 c\bar{\mu}E} \left(\frac{2m_r}{h^2}\right)^{3/2} (E - E_g)^{1/2} [1 - f_c(E_c) - f_v(E_v)] \tag{3.2.38}$$

For the case of a constant matrix element, the assumption is that $|M_{if}|^2$ is independent of the energy and wave vector of the initial and final states. In such a case, the total spontaneous emission rate and absorption coefficient are obtained by summing over the initial and final states using Eqs. (3.2.19) and (3.2.22). The calculation is similar to that presented above; the only difference is that in the latter case the delta function representing momentum conservation is absent.

3.2.3 Absorption Coefficient and Optical Gain

We now describe the calculation of the spontaneous-emission rate and the absorption coefficient using the Gaussian Halperin-Lax band-tail model and Stern's matrix element. The results in this section are presented for the quaternary InGaAsP alloy emitting near 1.3 μm or 1.55 μm. The gain calculations for the InGaAsP material have been given in Ref. 18–21.

The optical absorption (or gain) coefficient for transitions between the valence and conduction bands is obtained by integrating Eq. (3.2.19) over all states, and is given by

$$\alpha(E) = \frac{q^2 h}{2\epsilon_0 m_0^2 c \bar{\mu} E} \int_{-\infty}^{\infty} \rho_c(E')\rho_v(E'')|M_{if}|^2[1 - f_v(E'') - f_c(E')]\, dE' \quad (3.2.39)$$

where $E'' = E' - E$ and ρ_c and ρ_v are, for the conduction and valence bands respectively, the densities of states per unit volume per unit energy. The integral is evaluated numerically and summed for both light-hole and heavy-hole bands. The matrix element M_{if} can be expressed as a product of two terms, $M_{if} = M_b M_{env}$, where M_b is the previously defined contribution [Eq. (3.2.33)] from the band-edge Bloch functions and M_{env} is the matrix element of the envelope wave functions representing the effect of band-tail states. The envelope wave function is a plane wave above the band edge and takes the form of the ground state of a hydrogen atom for impurity states in the band tail. The calculation described here uses the envelope matrix element of Stern.[14,15] The square of the Bloch-function matrix element $|M_b|^2$ is given by Eq. (3.2.33) for the Kane model. Correction to the value of $|M_b|^2$ can arise from the contribution of other conduction bands.[22,23] These corrections would simply scale the gain values calculated using Eq. (3.2.39).

Figure 3.4 shows the calculated absorption curves for p-type InGaAsP ($\lambda = 1.3$ μm) with different doping levels. The parameters used in the calculation are listed in Table 3.1. The heavy-hole mass, the light-hole mass, the dielectric constant, and the spin-orbit splitting are obtained from a linear extrapolation of the measured binary values. The conduction-band effective

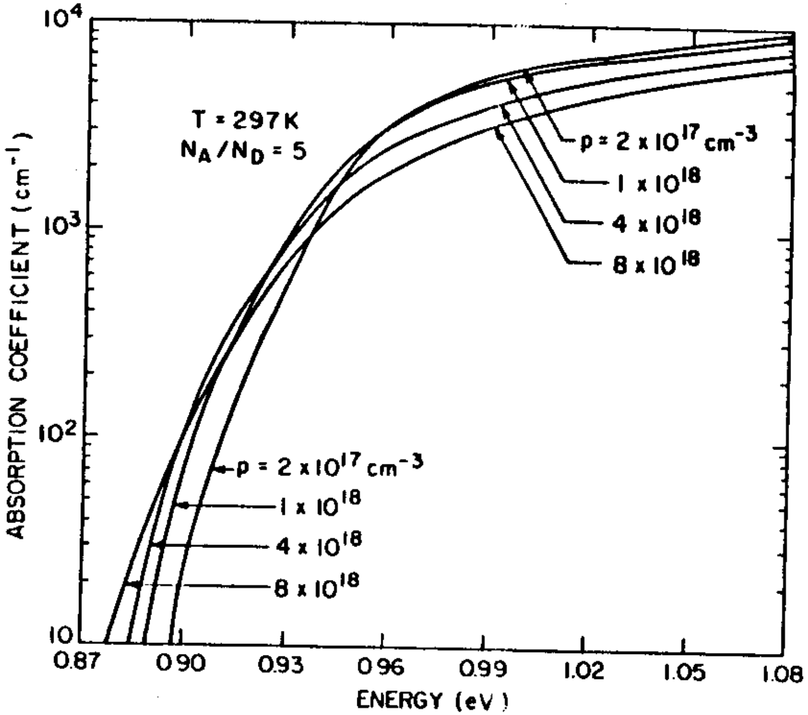


Fig. 3.4 Calculated absorption curves for p-type InGaAsP ($\lambda = 1.3$ μm) for different majority carrier concentrations. The majority carrier density $p = N_A - N_D$. (After Ref. 18)

Table 3.1 Band-Structure Parameters of $In_{1-x}Ga_xAs_yP_{1-y}$ Lattice Matched to InP [$x = 0.4526y/(1 - 0.031y)$]

PARAMETER	DEPENDENCE ON THE MOLE FRACTIONS x AND y
Energy gap at zero doping	E_g(in eV) = $1.35 - 0.72y + 0.12y^2$
Heavy-hole mass	$m_{hh}/m_0 = (1 - y)[0.79x + 0.45(1 - x)] + y[0.45x + 0.4(1 - x)]$
Light-hole mass	$m_{lh}/m_0 = (1 - y)[0.14x + 0.12(1 - x)] + y[0.082x + 0.026(1 - x)]$
Dielectric constant	$\epsilon = (1 - y)[8.4x + 9.6(1 - x)] + y[13.1x + 12.2(1 - x)]$
Spin-orbit splitting	Δ(in eV) = $0.11 + 0.31y - 0.09y^2$
Conduction-band mass	$m_c/m_0 = 0.080 - 0.039y$

mass is obtained from the data in Ref. 24. Since we are dealing with a composition that is lattice-matched to InP, the energy gap E_{g0} (for zero doping) is obtained from the data in Ref. 25. The reported experimental results are in good agreement with the extrapolated light-hole mass.[26] The carrier-concentration-dependent energy-gap reduction is assumed to be the same as for GaAs,[16]

$$\Delta E_g(\text{in eV}) = (-1.6 \times 10^{-8})(p^{1/3} + n^{1/3}) \qquad (3.2.40)$$

where n and p are in units of cm^{-3}. The effect of ΔE_g is only to shift the position of the absorption edge or the gain peak, and the magnitude of the maximum gain at a given current density remains unchanged.

The results shown in Fig. 3.4 assume that acceptors are fully ionized, i.e. $N_A - N_D = p$ where N_A and N_D are the acceptor and the donor concentrations, respectively and p is the majority carrier concentration. The ratio $N_A/N_D = 5$ was chosen on the basis of the experimental studies on Zn-doped GaAs. The calculation shows that with an increase in the doping level, the effect of band-tail states becomes more pronounced and the absorption edge is less steep. For the same reason, the absorption on the high-energy side is reduced with an increase in the doping level.

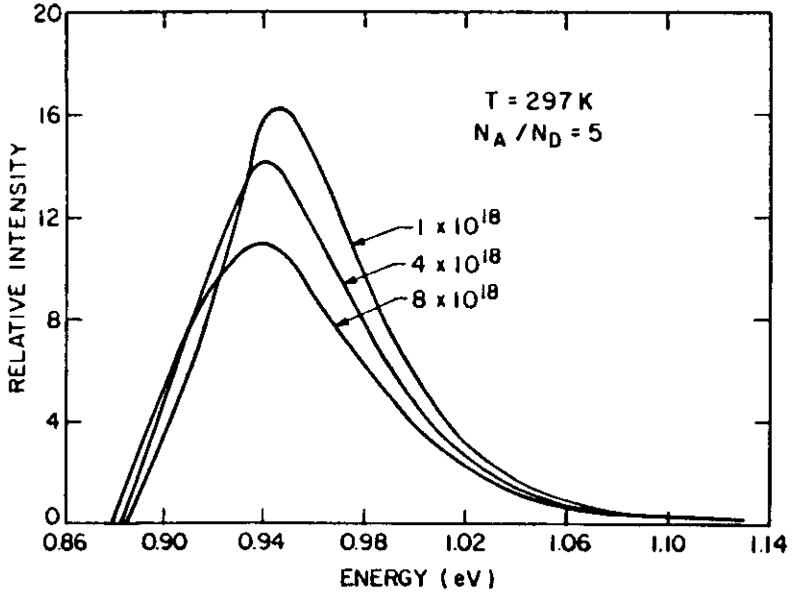


Fig. 3.5 Spontaneous emission spectra for p-type InGaAsP (λ = 1.3 μm). The parameters used are the same as those in Fig. 3.4. (After Ref. 18)

The total spontaneous-emission rate per unit volume can be calculated in a similar manner. Integrating Eq. (3.2.22) over all states, it is given by

$$R(E) = \frac{4\pi q^2 \bar{\mu} E}{m_0^2 \epsilon_0 c^3 h^2} \int_{-\infty}^{\infty} \rho_c(E')\rho_v(E'')f_c(E')f_v(E'')|M_{if}|^2 \, dE' \qquad (3.2.41)$$

where $E'' = E' - E$. The integral is evaluated numerically. Equation (3.2.41) yields the spontaneous-emission spectrum if R is calculated as a function of the photon energy E.

Figure 3.5 shows the calculated spontaneous-emission spectra for different hole concentrations. Note that with increasing carrier concentration the peak of the spontaneous emission shifts to lower energies while its height (maximum emission intensity) decreases. Also note that the width of the emission spectrum increases with increasing carrier concentration as a result of the band-tail states.

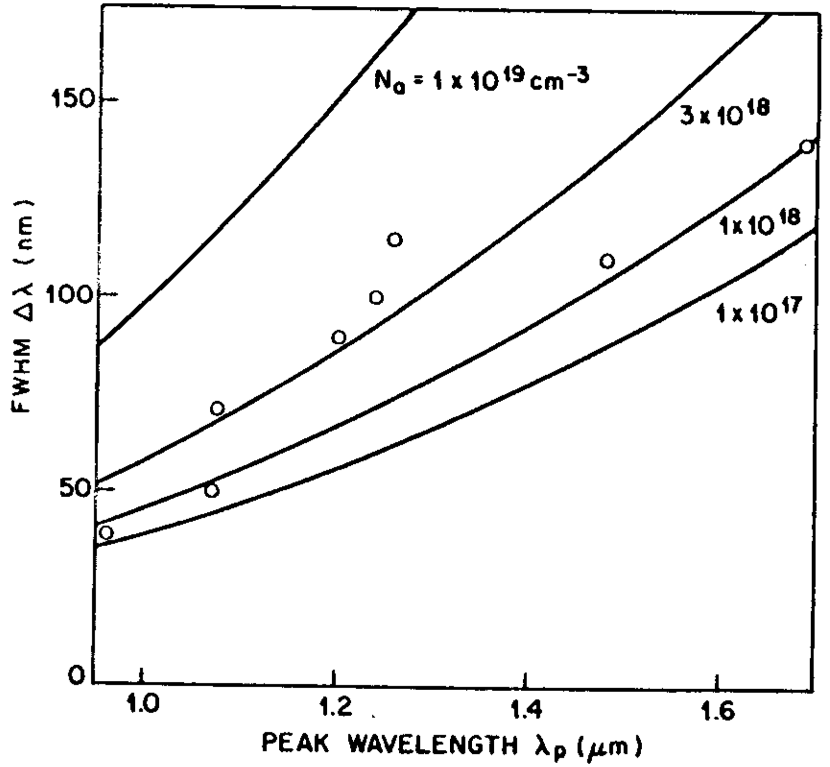


Fig. 3.6 Effect of acceptor concentration on the width of the spontaneous-emission spectrum for p-type InGaAsP (λ = 1.3 μm). Circles are experimental data points. (After Ref. 20)

Figure 3.6 shows the effect of acceptor concentration on the width (FWHM) of the spontaneous emission spectrum. The measured data of Ref. 20 are also shown. The absorption edge shifts to lower energies due to the formation of band-tail states with increasing impurity concentration. The emission and absorption spectra of *n*-type semiconductors have essentially the same features as those of *p*-type semiconductors.

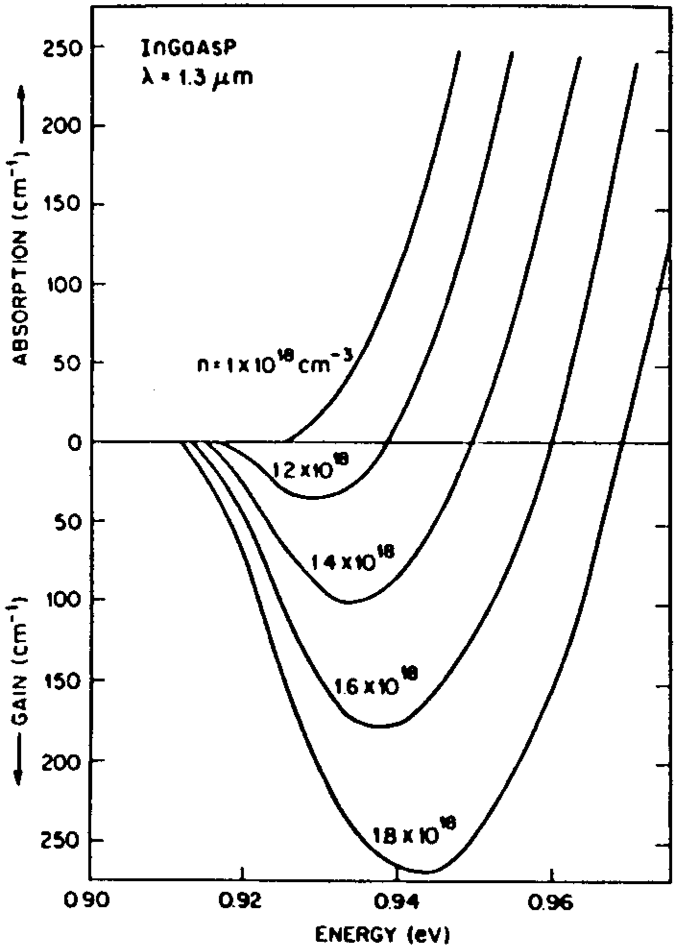


Fig. 3.7 Calculated gain spectra at different injected carrier densities for undoped InGaAsP (E_g = 0.96 eV, λ = 1.3 μm). The absorption coefficient is plotted as a function of the photon energy. (After Ref. 18)

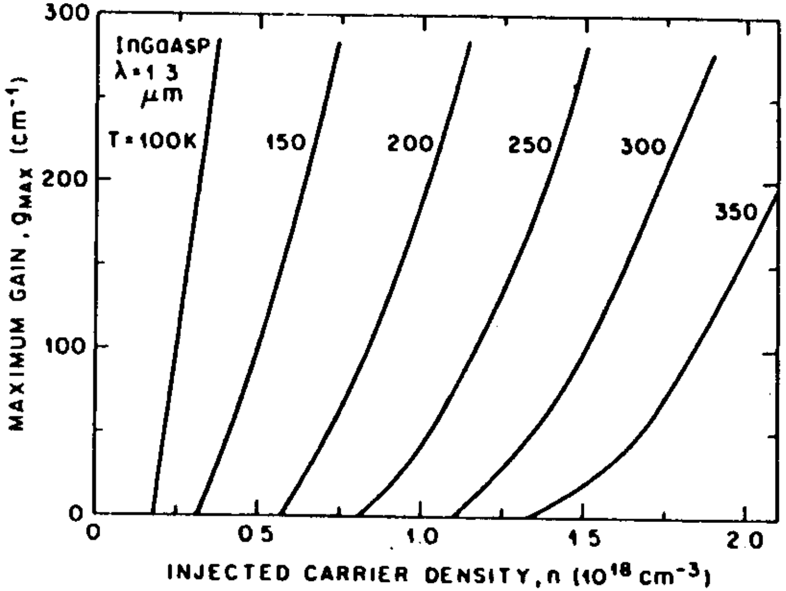


Fig. 3.8 Calculated maximum gain (g_{max}) as a function of injected carrier density (*n*) in undoped InGaAsP (λ = 1.3 μm) at various temperatures. (After Ref. 27)

We now calculate the gain in InGaAsP excited by external pumping. Under sufficient optical excitation or current injection, the number of electrons and holes may be large enough to satisfy Eq. (3.2.7); the semiconductor material would then exhibit optical gain. The rate of excitation can be expressed in terms of the injected carrier density or by the nominal current density defined below by Eq. (3.2.42). Figure 3.7 shows the gain spectra at various injected carrier densities. For injected carrier densities less than 10^{18} cm^{-3}, there is no net gain at any photon energy. As the carrier density is further increased, the absorption coefficient α becomes negative over a limited energy range and gain occurs. The photon energy at which the maximum gain occurs shifts to higher energy with increasing injection while, at the same time, the gain appears at a lower energy with increasing injection. This effect is due to a shift in the band edge given by Eq. (3.2.40). Figure 3.8 shows the maximum gain (peak value in Fig. 3.7) plotted as a function of the injected carrier density at different temperatures.[27] The calculation is done for undoped material with a residual concentration of 2×10^{17} cm^{-3} acceptors and donors. Note that the gain is higher at a lower temperature for the same excitation. The temperature plays a role through the Fermi factors and affects the distribution of electrons and holes in the conduction and valence bands. At higher temperatures the carriers are distributed over a wider energy range, so the maximum gain is less.

A convenient way to express the excitation rate is to use the nominal current density J_n, which is defined as the total spontaneous-emission rate per unit volume and equals the injected current density at unit quantum efficiency. The nominal current density is usually expressed in A/(cm²· μm) and is given by

$$J_n = qR \tag{3.2.42}$$

where R is the total spontaneous-emission rate per unit volume given by Eq. (3.2.41). Figure 3.9 shows the calculated maximum gain as a function of nominal current density for InGaAsP at 1.3 μm. The nominal current density equals the current lost to radiative recombination in a double-heterostructure laser. Figure 3.9 shows that to achieve the same gain, higher current density is needed at higher temperatures. Figures 3.8 and 3.9 show the basis of the increase in threshold current of injection lasers as the temperature is increased. Section 3.6 discusses the temperature dependence in detail.

Results of a similar calculation for the variation of the maximum gain with the injected carrier density and the nominal current density at 1.55 μm are given in Ref. 27. The results are very similar to those of Figs. 3.8 and 3.9 except that a smaller injected carrier density and a smaller nominal current density are needed at the longer wavelength to achieve the same gain. This arises from the smaller conduction-band effective electron mass at the longer wavelength, which allows the gain condition (3.2.7) to be satisfied at a lower carrier density.

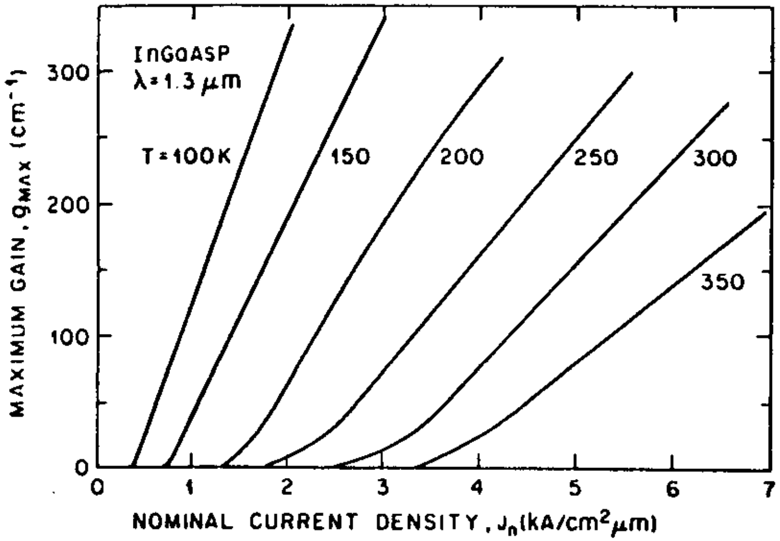


Fig. 3.9 Calculated maximum gain (g_{max}) as a function of nominal current density (J_{nom}) for undoped InGaAsP (λ = 1.3 μm) at various temperatures. (After Ref. 27)

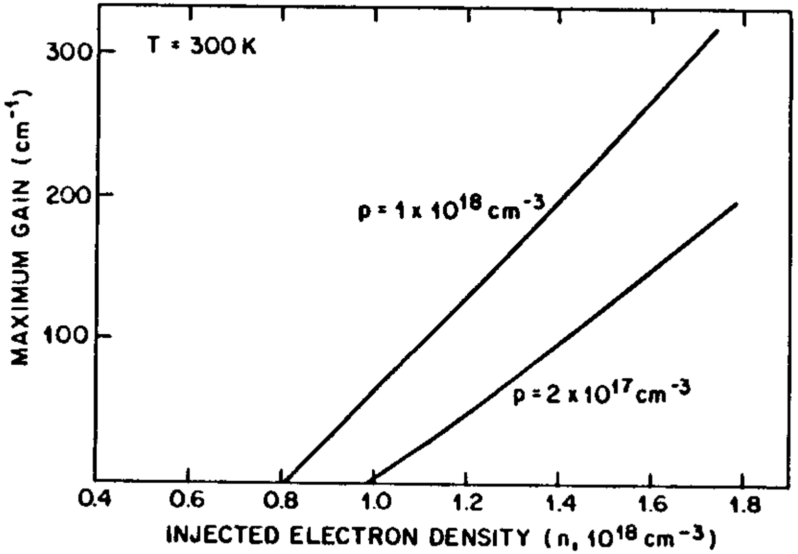


Fig. 3.10 Calculated maximum gain as a function of injected carrier density in p-type InGaAsP for two different majority carrier (hole) concentrations.

We next consider the effect of doping on the optical gain. Figure 3.10 shows the calculated maximum gain as a function of the injected electron density in p-type InGaAsP (λ = 1.3 μm) with different hole concentrations. The calculation assumes $N_A/N_D = 5$. Figure 3.11 shows the result for n-type InGaAsP as a function of the injected hole density. The two figures show that the excitation rate required for transparency ($g = 0$) decreases with higher doping both for p- and n-type materials. The decrease is more rapid for n-type than for p-type materials. The difference in behavior between the n- and p-type materials is due to the smaller effective mass of electrons compared with that of holes. This may be seen as follows. The condition for net stimulated emission or gain is given by Eq. (3.2.7). Increasing the doping level increases both E_{fc} and E_{fv}; hence fewer additional injected carri-

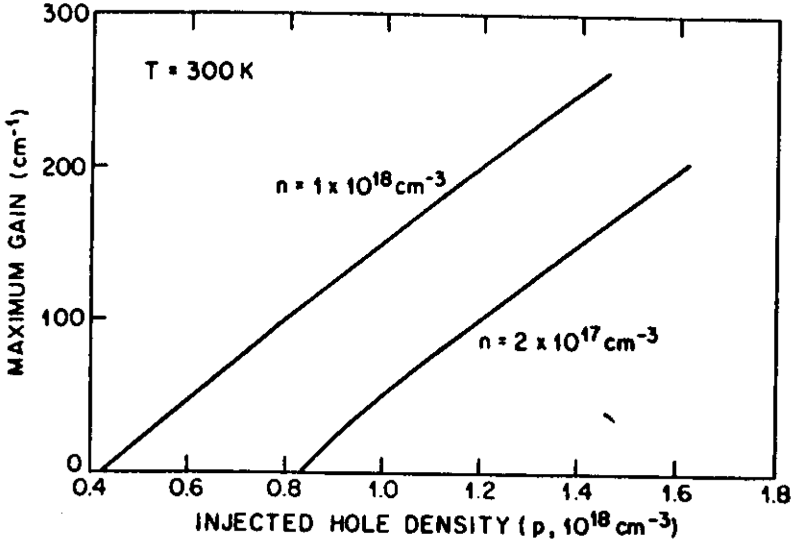


Fig. 3.11 Calculated maximum gain as a function of injected hole density in n-type InGaAsP for two different majority carrier (electron) concentrations.

ers are needed to satisfy the condition for gain. Since the effective mass of the conduction band is smaller than that of the valence band,

$$\left|\frac{\partial E_{fc}}{\partial n}\right| > \left|\frac{\partial E_{fv}}{\partial p}\right|$$

This explains why the effect of doping is more pronounced in n-type than in p-type material.

Figures 3.8–3.11 show that the optical gain varies almost linearly with the injected carrier density and can be approximately written as

$$g = a(n - n_0) \tag{3.2.43}$$

where the slope a is called the gain coefficient and n_0 is the injected carrier density required to achieve transparency (i.e., $g = 0$ when $n = n_0$). From Eq. (3.2.7), the condition for transparency is given by

$$E_{fc} + E_{fv} = 0 \tag{3.2.44}$$

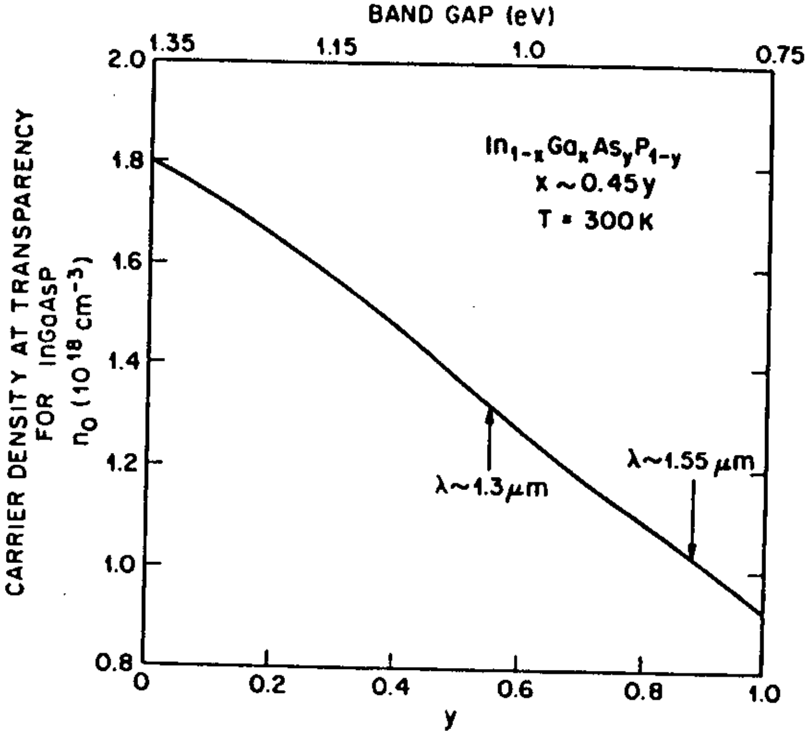


Fig. 3.12 Injected carrier density n_0 needed for transparency ($g = 0$) for InGaAsP lattice-matched to InP. Arrows indicate values for 1.3- and 1.55-μm InGaAsP lasers.

The quantity n_0 can be calculated from Eq. (3.2.25) using the approximation (3.2.26) for parabolic bands. Figure 3.12 shows the calculated n_0 for undoped InGaAsP lattice-matched to InP. The smaller n_0 for low band gap is due to a smaller effective mass.

The variation of the gain with the carrier density is approximately linear. As such, it is useful for a simple but fairly accurate description of semiconductor lasers as was presented in Chapter 2. Equation (3.2.43) was used in the phenomenological model introduced in Sec. 2.4. The gain parameters a and n_0 can be determined either using numerical results presented in Figs. 3.8–3.12 or deduced from the experimental data. Typical values for InGaAsP lasers are in the range of $1.2–2.5 \times 10^{-16}$ cm^2 for a and $0.9–1.5 \times 10^{18}$ cm^{-3} for n_0 depending on the laser wavelength and doping levels. It should be stressed that because of uncertainties in the band-structure parameters, the numerical results are only accurate to within a factor of 2.

The total spontaneous emission rate R given by Eq. (3.2.41) can be approximated by

$$R = Bnp \qquad (3.2.45)$$

where B is the radiative recombination coefficient and n and p are the electron density and the hole density respectively. For undoped semiconductors, $R = Bn^2$. The calculated $B = 1.2 \times 10^{-10}$ cm^3/s for InGaAsP ($\lambda = 1.3$ μm) at a carrier concentration of 1×10^{18} cm^{-3}. Calculation of the radiative recombination rate shows that B decreases with increasing carrier density. This has been confirmed by Olshansky et al.[28] using carrier-lifetime measurements.

3.3 NONRADIATIVE RECOMBINATION

As mentioned in Sec. 3.1, electrons and holes in a semiconductor can also recombine nonradiatively. Nonradiative mechanisms include recombination at defects, surface recombination, and Auger recombination, among others. For long-wavelength semiconductor lasers, however, the Auger process is generally the predominant nonradiative mechanism, and most of this section is devoted to its study. The surface- and defect-recombination processes are discussed briefly in Sec. 3.3.4.

The Auger recombination process involves four particle states (three electrons and one hole, two electrons and two holes, and so forth). In this process, the energy released during the electron-hole recombination is transferred to another electron (or hole), which gets excited to a high energy state in the band. This electron or hole then relaxes back to achieve thermal equilibrium by losing its energy to lattice vibrations or phonons. The Auger recombination rate R_a may be approximately written as

$$R_a = Cn^3 \qquad (3.3.1)$$

where n is the injected carrier density in an undoped semiconductor. It is useful to define the carrier lifetime τ_A for the Auger process as

$$\tau_A = n/R_a = (Cn^2)^{-1} \qquad (3.3.2)$$

The quantity C is called the Auger coefficient. The inverse-square dependence of carrier lifetime on carrier density is used to identify the Auger effect experimentally. Departures from the simple inverse-square relation are discussed later in the section.

It is generally accepted that Auger recombination can be a major nonradiative mechanism in narrow-gap semiconductors.[29-33] The observed strong temperature dependence of threshold current in long-wavelength semiconductor lasers has focused attention on the Auger effect. There are several different types of Auger recombination processes. The three major types are (i) band-to-band processes.[27,32-37] (ii) phonon-assisted Auger processes[38-41] and (iii) trap-assisted processes.[42-45]

3.3.1 Band-to-Band Auger Processes

The band-to-band Auger processes in a direct-gap semiconductor are shown in Figure 3.13. The three processes are labeled CCCH, CHHS, and CHHL where C stands for the conduction band and H, L, and S stand for heavy-hole, light-hole, and split-off valence bands, respectively. The CCCH process involves three electrons and one hole and is dominant in n-type material.

The Auger recombination rate for this process was first derived in Ref. 29. The basis of the Auger effect is the Coulomb interaction between the two electrons 1 and 2 in the conduction band. In the CCCH mechanism, electron 1 recombines with hole 1′ (i.e., electron 1 makes a transition to empty state 1′) and the excess energy is transferred to electron 2, which is excited to state 2′. Electron 2′ loses its energy to opitcal phonons when it relaxes back to thermal equilibrium. The CHHS process involves one electron, two heavy holes, and a split-off-band hole. CHHL is similar to CHHS except that it involves a light hole.

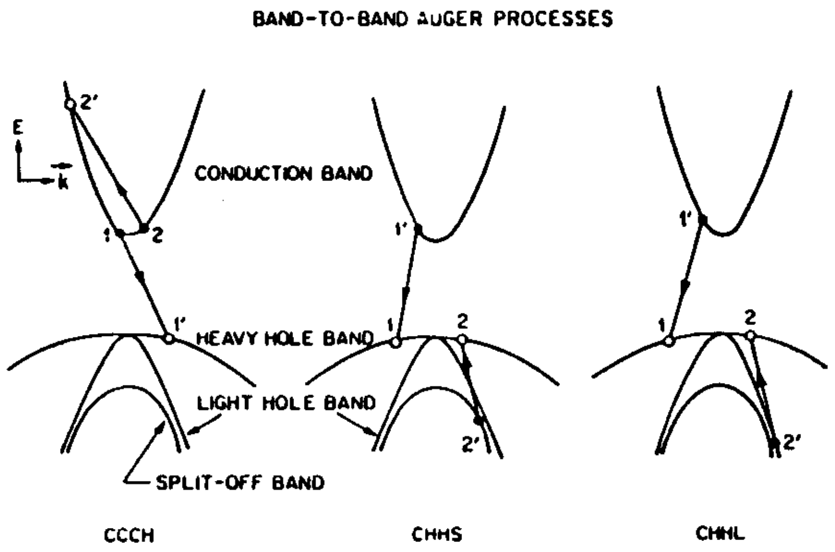


Fig. 3.13 Three different band-to-band Auger processes shown schematically. Electrons are represented by closed circles and holes by open circles.