Light-Emitting Diodes (LED)

Injection Luminescence

The basic structure giving rise to injection luminescence is that of a p–n junction diode operated under forward bias which was discussed in Chapter 2. Under forward bias, majority carriers from both sides of the junction cross the depletion layer and enter the material at the other side, where they are then the minority type of carrier and cause the local minority carrier population to be larger than normal. This situation is described as *minority carrier injection*. The excess minority carriers diffuse away from the junction recombining with majority carriers as they do so. Using eq. (2.42), we may write the excess electron concentration $\Delta n(x)$ in the p material as a function of distance x from the edge of the depletion region as

$$\Delta n(x) = \Delta n(0)\exp(-x/L_e)$$

The process is illustrated in Fig. 4.12. Ideally, in a light-emitting diode (LED) every injected electron takes part in a radiative recombination process and hence gives rise to an emitted photon. In practice this is not so, and the efficiency of the device may be described in terms of the quantum efficiency, which is defined as the rate of emission of photons divided by the rate of supply of electrons. In reverse bias, no carrier injection takes place and consequently no light is emitted. The current–voltage (i, V) relationship for a diode can usually be written (see eq. 2.51 and section 2.8.5) as

$$i = i_0\left[\exp\left(\frac{eV}{\beta kT}\right) - 1\right] \tag{4.4}$$

where i_0 is a constant (the reverse saturation current).

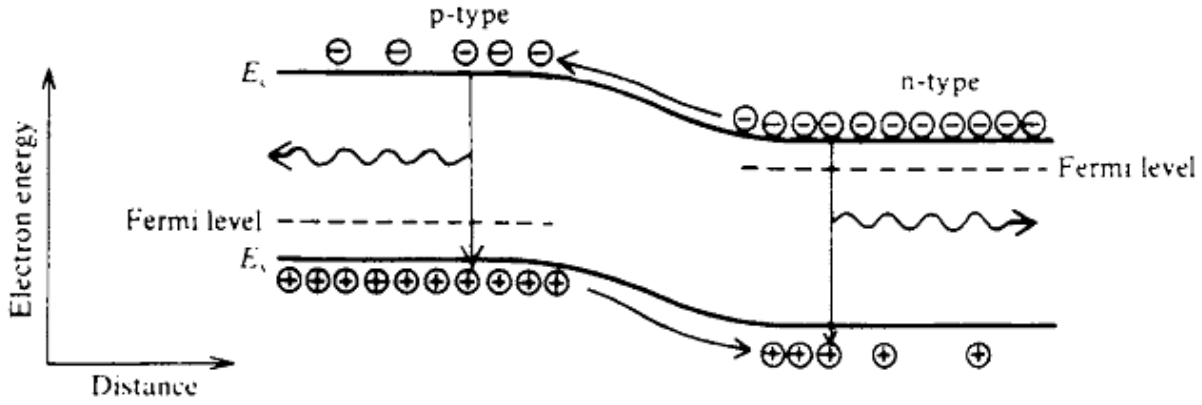


FIG. 4.12 Injection of minority carriers and subsequent radiative recombination with the majority carriers in a forward-biased p–n junction.

The number of radiative recombinations that take place is usually proportional to the carrier injection rate and hence to the total current flowing. If the transitions take place directly between states at the bottom of the conduction band and at the top of the valence band, then the emission wavelength λ_g is given by (see eq. 4.1)

$$hc/\lambda_g = E_c - E_v = E_g$$

Therefore

$$\lambda_g = hc/E_g \tag{4.5}$$

For example, GaAs has an energy bandgap of 1.43 eV, which corresponds to a value for λ_g of 0.86 μm. In fact, because of thermal excitation, electrons in the conduction band have a most probable energy which is $kT/2$ above the bottom of the conduction band (see Problem 2.20). Band-to-band transitions therefore result in a slightly shorter emission wavelength than given by eq. (4.5), and self-absorption can further distort the situation. However, as we shall see later, most transitions involve energy levels within the energy gap, and for these eq. (4.5) represents the shorter wavelength limit. We now consider the transmission process in more detail.

B1.1.2.1 Optical properties of LEDs

Radiative recombination

Figure B1.1.3 shows a number of possible radiative and non-radiative transitions in a simplified band diagram. The direct band–band transition is shown in figure B1.1.3*a*. Because two particles, electrons n and holes p, are involved, this process is called bimolecular recombination. The recombination rate, R (recombinations, per unit volume, per second), is proportional to the density of carriers in the upper state (electrons) and the density of empty lower states (holes) and is given by:

$$R = -\frac{dn}{dt} = -\frac{dp}{dt} = Bnp \tag{B1.1.1}$$

where n and p are the electron and hole concentrations. The proportionality factor B is a measure of the probability of radiative recombination and is called the bimolecular recombination coefficient. It can be calculated from a few basic material parameters such as the bandgap energy, the absorption coefficient and the refractive index [3] which are all experimentally accessible. For most III–V semiconductors, the bimolecular recombination coefficient is in the order of $1-10 \times 10^{-10}\,\mathrm{cm^3\,s^{-1}}$. At very high carrier concentrations, as for instance in laser diodes, B starts to decrease and equation (B1.1.1) has to be modified [4].

4.6.1.3 *Exciton recombination*

Exciton states exist within the energy gaps of even pure semiconductor materials (they must not be confused with impurity donor or acceptor states). We may visualize such states as being akin to Bohr-like states in which an electron and a hole circle round their common centre of gravity at relatively large distances (Fig. 2.12). The electron and hole are relatively weakly bound and the exciton states are situated just below the bottom of the conduction band. In Chapter 2 we showed (eq. 2.28a) that we may write the exciton binding energy E_e as

$$E_e = 13.6 \frac{m_r^*}{m} \left(\frac{1}{\varepsilon_r}\right)^2 \text{eV} \tag{4.9}$$

where m_r^* is the reduced mass. For example, in GaAs we have $\varepsilon_r = 11.5$, $m_e^* = 0.068m$ and $m_h^* = 0.47m$, whence $m_r^* = 0.06m$ and $E_e = 5.9$ meV. Observed experimental values are in reasonable agreement with this simple model calculation: in GaAs the exciton binding energy is found to be 4.8 meV. The exciton is capable of movement through the lattice, although, since exciton energies can be affected by the presence of impurities, in some circumstances an exciton may remain 'bound' in the vicinity of the impurity. If the impurities are neutral donors or acceptors then the exciton binding energy is usually about one-tenth of that of the centres to which they are bound. (The binding energy may be much larger than this at isoelectronic traps, which are discussed in section 4.6.3.) Bound exciton states may be sufficiently well localized so that electron–hole recombination can take place in indirect bandgap semiconductors via these states without the need for phonon intervention to conserve the wavevector.

Appendix:

2.4.3 Excitons

We have just seen that the introduction of suitable impurities into intrinsic semiconductor material can result in the formation of electron energy levels situated just below the bottom of the conduction band. However, electron energy levels similarly situated can also appear in *intrinsic* material. These arise because the Coulombic attraction of an electron for a hole can result in the two being bound together; such a bound electron–hole pair is called an *exciton*. We may picture the exciton as an electron and a hole orbiting about their common centre of gravity with orbital radii which are inversely proportional to their effective masses as shown in Fig. 2.12. The Bohr model is readily adapted to this situation with the electron mass being replaced by the reduced mass m_r^* of the electron and hole; m_r^* is given by

$$\frac{1}{m_r^*} = \frac{1}{m_e^*} + \frac{1}{m_h^*}$$

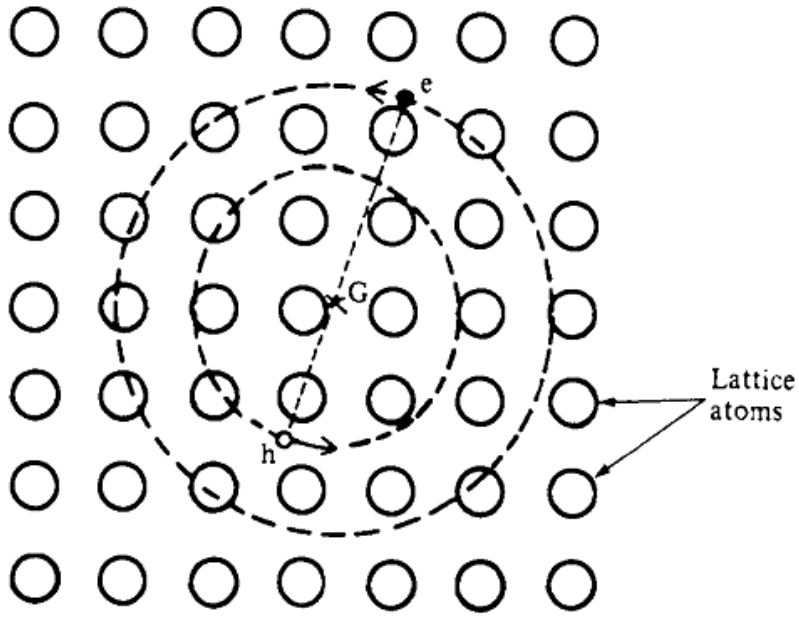


FIG. 2.12 Model of an exciton; the electron e and hole h may be regarded as being bound together and orbiting around their common centre of gravity G with radii which are inversely proportional to their effective masses (we have assumed $m_h^* > m_e^*$).

The binding energy E_e is then obtained by modifying eq. (2.28) to give

$$E_e = 13.6 \frac{m_r^*}{m} \left(\frac{1}{\varepsilon_r}\right)^2 \text{eV} \tag{2.28a}$$

Since m_r^* will be the same order of magnitude as m_e^* and m_h^*, we see that exciton energy levels will be similarly placed to those of donor levels in doped semiconductors.

Excitons may move through a crystalline lattice and thus provide an important means of transferring energy from one point in the material to another. They play an important role in the luminescence of solids and will be discussed further in Chapter 4.

2.6 Non-radiative recombination in the bulk

There are two basic recombination mechanisms in semiconductors, namely *radiative* recombination and *non-radiative* recombination. In a radiative recombination event, one photon with energy equal to the bandgap energy of the semiconductor is emitted, as illustrated in Fig. 2.5. During non-radiative recombination, the electron energy is converted to vibrational energy of lattice atoms, i.e. phonons. Thus, the electron energy is converted to heat. For obvious reasons, non-radiative recombination events are unwanted in light-emitting devices.

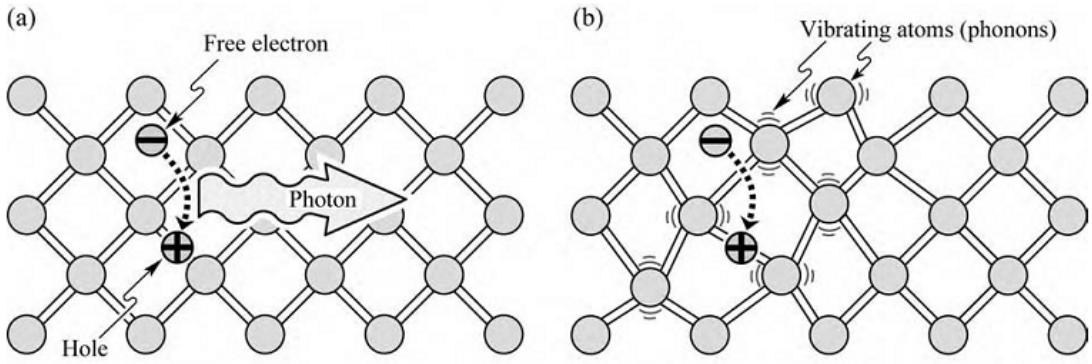


Fig. 2.5. (a) Radiative recombination of an electron–hole pair accompanied by the emission of a photon with energy $h\nu \approx E_g$. (b) In non-radiative recombination events, the energy released during the electron–hole recombination is converted to phonons (adapted from Shockley, 1950).

There are several physical mechanisms by which non-radiative recombination can occur. Defects in the crystal structure are the most common cause for non-radiative recombination. These defects include unwanted foreign atoms, native defects, dislocations, and any complexes of defects, foreign atoms, or dislocations. In compound semiconductors, native defects include interstitials, vacancies, and antisite defects (Longini and Greene, 1956; Baraff and Schluter, 1985). All such defects have energy level structures that are different from substitutional semiconductor atoms. It is quite common for such defects to form one or several energy levels within the forbidden gap of the semiconductor.

Energy levels within the gap of the semiconductor are efficient recombination centers; in particular, if the energy level is close to the middle of the gap. The recombination of carriers via a trap level is shown schematically in Fig. 2.6. Owing to the promotion of non-radiative processes, such deep levels or traps are called ***luminescence killers***.

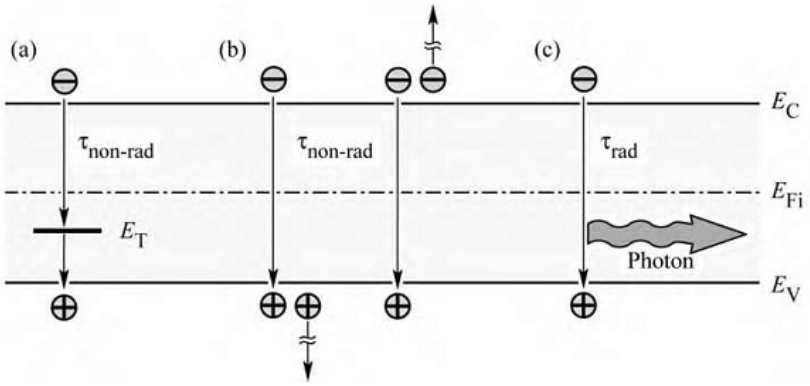


Fig. 2.6. Band diagram illustrating recombination: (a) non-radiative via deep level, (b) non-radiative via Auger process and (c) radiative.

The most relevant non-radiative process for optoelectronic devices is recombination via states related to crystal defects. Examples of such defects are dislocations, pores, grain boundaries, vacancies, inclusions or precipitates. Carriers within a diffusion length of the defect will usually be trapped by the defect states and recombine there. The recombination rate scales linearly with the carrier concentration:

$$R_{nr} = A^* n \tag{B1.1.11}$$

where the proportionality factor A^* increases with the density of non-radiative centres.

Usually a defect deforms the bandstructure, either by trapping charges or by deforming the lattice, thereby inducing local strain. In either case, the deformation produces a potential barrier of the height, E_{act}, which has to be overcome by the carriers. Therefore, the process recombination has a thermal dependence of the form:

$$R_{nr}(T) = R^* \exp\left(-\frac{E_{act}}{kT}\right) \tag{B1.1.12}$$

where $R_{nr}(T)$ is the non-radiative recombination rate and R^* a temperature independent coefficient. The temperature dependence of the radiative transition can be neglected. Therefore, as the temperature is reduced, the non-radiative recombination rate decreases exponentially and the radiative processes become more dominant.

Instead of generating a photon, the energy released during carrier recombination can also be transferred to other carriers (electrons or holes) and then be dissipated as phonons. This non-radiative process is called Auger recombination. Two examples for Auger processes are shown in figure B1.1.3*d*, but many other processes are possible, depending on the nature and occupation of the involved electronic states. Since two carriers of one type and one carrier of the opposite type are required for the Auger process, the recombination rate is proportional to either np^2 or pn^2. The proportionality factor C is called the Auger coefficient and has typical values for III–V materials of 10^{-28}–10^{-29} cm^6 s^{-1}. It is described as

$$R_{Auger} = C_p np^2 + C_n pn^2 \tag{B1.1.14}$$

or, in the intrinsic case:

$$R_{Auger} = Cn_i^3. \tag{B1.1.15}$$

The Auger recombination thereby scales as the cube of the carrier concentration. For LEDs where typical carrier densities are low compared to lasers, the Auger recombination plays a minor role in the device efficiency. However, the Auger effect becomes more prominent in materials with small bandgap energies such as infrared emitting devices.

4.6.1.4 *Emission linewidths*

At a finite temperature electrons in the conduction band of a semiconductor are distributed amongst the energy levels with a probability of occupancy given by the Fermi distribution function. At the end of section 2.5 it is shown that if we denote the number of electrons per unit volume which have energies between E and $E + \mathrm{d}E$ by $n(E)\,\mathrm{d}E$ then

$$n(E) = \frac{4\pi}{h^3}(2m_e^*)^{3/2}(E - E_c)^{1/2}\exp\left(-\frac{E - E_F}{kT}\right)$$

The function $n(E)$ is sketched in Fig. 2.14(c), and it can be shown (Problem 2.20) that it has a halfwidth of about $2kT$ (the actual value is close to $1.8kT$). A similar result can be obtained for holes in the valence band. Thus assuming direct band-to-band transitions we would expect a range of emission frequencies, $\Delta\nu$, which have a halfwidth of about $2kT/h$. This may be converted to the equivalent wavelength spread $\Delta\lambda_0$ by use of the relationship $\lambda_0\nu = c$, so that

$$\frac{\mathrm{d}\nu}{\mathrm{d}\lambda_0} = -\frac{c}{\lambda_0^2}$$

and hence

$$\Delta\lambda_0 = \frac{2kT\lambda_0^2}{hc} \tag{4.10}$$

At a wavelength of 900 nm and a temperature of 300 K, eq. (4.10) predicts a halfwidth of 38 nm which agrees reasonably well with what is observed in practice (see e.g. Fig. 4.15). The result remains reasonably valid even when the transitions involve impurity states rather than direct band-to-band transitions.

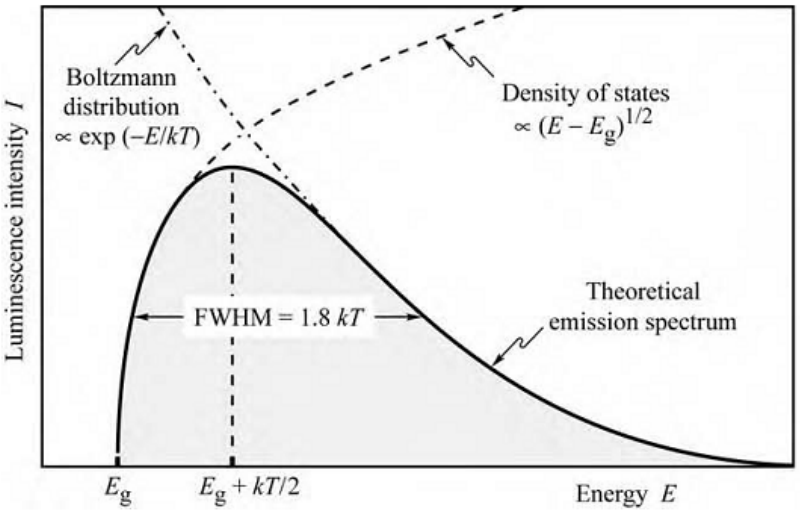


Fig. 5.2. Theoretical emission spectrum of an LED. The full-width at half-maximum (FWHM) of the emission line is $1.8kT$.

4.6.4 LED construction

A typical LED construction is shown in Fig. 4.16. It is obviously advantageous if most of the radiative recombinations take place from the side of the junction nearest the surface, since then the chances of reabsorption are lessened. We may ensure this by arranging that most of the current flowing across the diode is carried by those carriers that are injected into the surface layer.

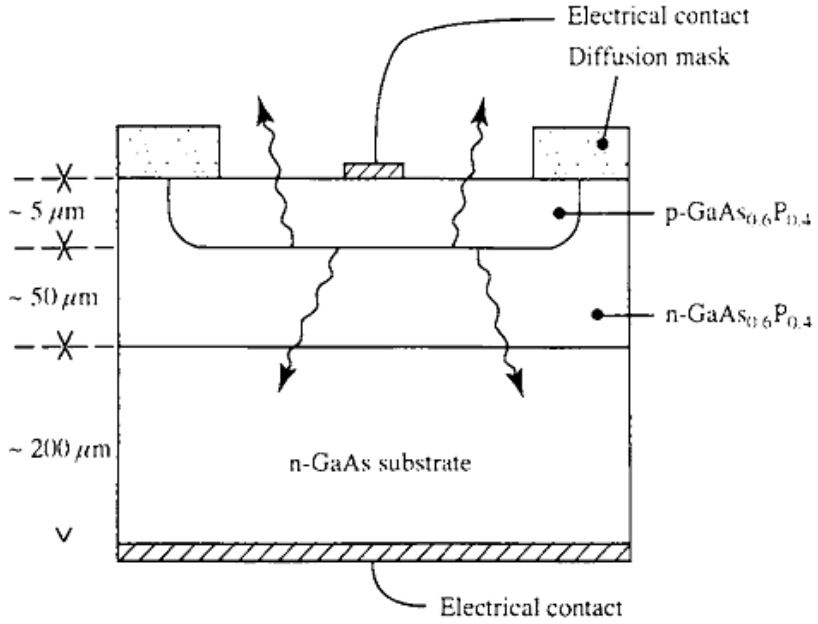


FIG. 4.16 Cross-section of a standard red-emitting LED chip based on GaAsP. A layer of n-type $GaAs_{0.6}P_{0.4}$ (using tellurium as a dopant) is deposited by vapour phase epitaxy on a GaAs substrate. A p–n junction is then formed by diffusing in Zn through a surface mask. The small aluminium contact on the upper surface allows as much of the radiation to escape as possible; any radiation flowing downwards is almost completely absorbed by the GaAs.

Light Extraction

Although the internal quantum efficiencies of some LED materials can approach 100%, the external efficiencies are much lower. The main reason for this is that most of the emitted radiation strikes the material interface at an angle greater than the critical angle and so remains trapped. Unfortunately, the high refractive indices of the III–V materials discussed here give rise to small critical angles. Consider, for example, radiation from a point source within a medium of refractive index n_1 impinging on a plane interface with another medium of refractive index n_2, where $n_2 < n_1$, as shown in Fig. 4.17. Only those rays (e.g. beam 1) that have an angle to the normal less than the critical angle (θ_c) enter the second medium. Those with angles greater than θ_c (e.g. beam 3) are reflected back into the first medium.

From eq. (1.14), we have that the critical angle θ_c is given by

$$\theta_c = \sin^{-1}(n_2/n_1) \tag{4.13}$$

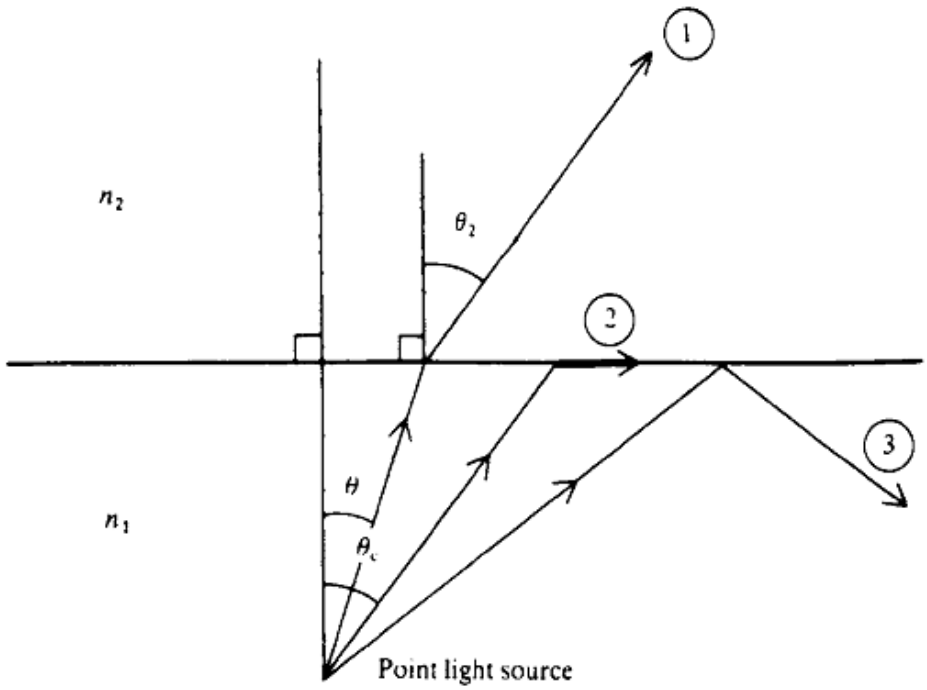


FIG. 4.17 Phenomenon of total internal reflection. When a beam of light is incident at an angle θ_1 onto an interface between two media of refractive indices n_1 and n_2 ($n_2 < n_1$) then the refracted beam (beam 1) makes an angle θ_2 with the normal to the surface where $n_1 \sin \theta_1 = n_2 \sin \theta_2$ (Snell's law). At the critical angle ($\theta_1 = \theta_c$), $\theta_2 = 90°$ (beam 2) and the refracted beam emerges along the interface. At angles of incidence greater than θ_c (beam 3), the beam is totally reflected back into the first medium.

Light originating at recombination centres near the p–n junction will be radiated isotropically, whereas only that within a cone of semiangle θ_c will escape. In Problem 4.1, it is shown that the fraction F of the total generated radiation that is actually transmitted into the second medium is

$$F \simeq \frac{1}{4}\left(\frac{n_2}{n_1}\right)^2\left[1-\left(\frac{n_1-n_2}{n_1+n_2}\right)^2\right] \tag{4.14}$$

There are two obvious ways to increase F; the first is to ensure that most rays strike the surface at less than the critical angle. This may be achieved by shaping the semiconductor/air interface into a hemisphere, as shown in Fig. 4.18(a). However, although this technique is used occasionally in high power diodes, it is too difficult and expensive for most situations. The second, and much commoner, technique is to encapsulate the junction in a transparent medium of high refractive index. This is usually a plastic material with a refractive index of about 1.5. Using eq. (4.14) with $n_1 = 3.6$ and $n_2 = 1.5$ we obtain $F = 0.036$, giving a nearly threefold increase in light output over the simple semiconductor/air interface (Example 4.2). Of course there will be some losses at the plastic/air interface, but these are easily minimized by mouding the plastic into an approximately hemispherical shape (see Fig. 4.18b).

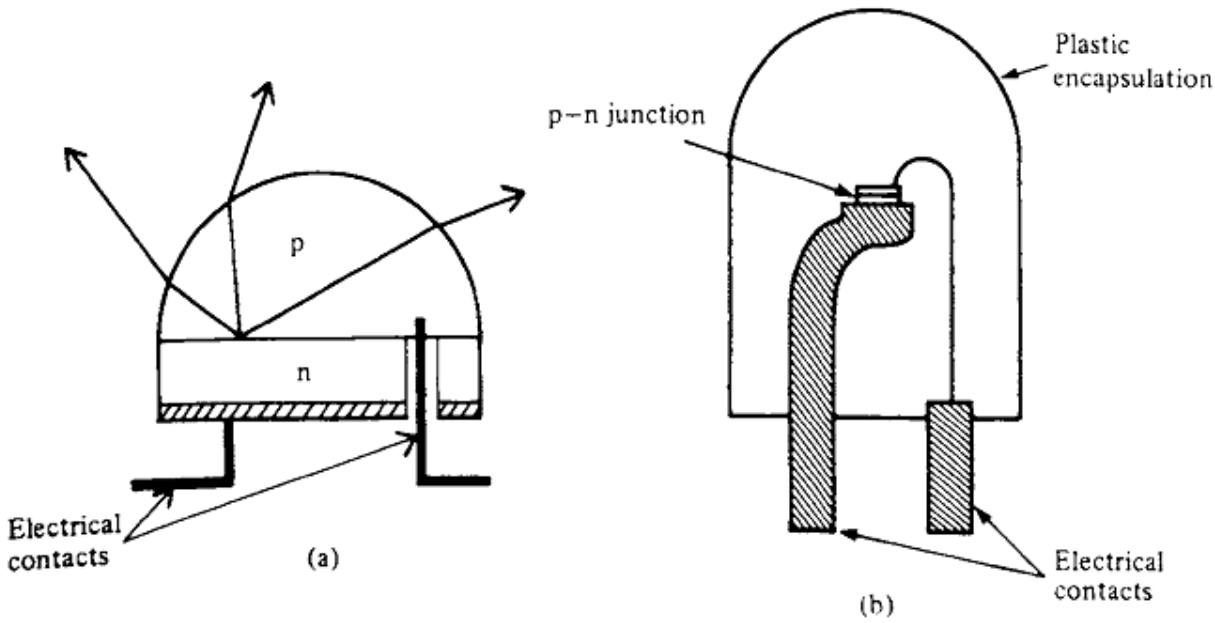


FIG. 4.18 Two methods used to reduce reflection losses in LEDs. In (a) the p material is made into a hemispherical dome. More radiation then strikes the semiconductor/air interface at less than the critical angle than would otherwise be the case. In (b) the p–n junction is surrounded by plastic encapsulation. Losses at the plane semiconductor/plastic interface are then less than for a corresponding semiconductor/air interface.

5.4 Radiation pattern

All LEDs have a certain ***radiation pattern*** or ***far-field pattern***. The intensity, measured in W/cm^2, depends on the longitudinal and azimuth angle and the distance from the LED. The total optical power emitted by the LED is obtained by integration over the area of a sphere.

$$P = \int_A \int_\lambda I(\lambda)\, d\lambda\, dA \tag{5.25}$$

where $I(\lambda)$ is the ***spectral light intensity*** (measured in W per nm per cm^2) and A is the surface area of the sphere. The integration is carried out over the entire surface area.

5.5 The lambertian emission pattern

The index contrast between the light-emitting material and the surrounding material leads to a non-isotropic emission pattern. For high-index light-emitting materials with a planar surface, a lambertian emission pattern is obtained.

The light intensity in air can then be inferred from Eqs. (5.28), (5.30), (5.31), and (5.32). One obtains the ***lambertian emission pattern*** given by

$$\boxed{I_{\text{air}} = \frac{P_{\text{source}}}{4\pi r^2} \frac{\bar{n}_{\text{air}}^2}{\bar{n}_{\text{s}}^2} \cos\Phi} \,. \tag{5.33}$$

The lambertian emission pattern follows a cosine dependence on the angle Φ. The intensity is highest for emission normal to the semiconductor surface, i.e. for $\Phi = 0°$. At an angle of $\Phi = 60°$, the intensity decreases to half of its maximum value. The lambertian emission pattern is shown schematically in Fig. 5.5.

Several other surface shapes are also shown in Fig. 5.5. These non-planar surfaces exhibit various emission patterns. An isotropic emission pattern is obtained for hemispherically shaped LEDs, which have the light-emitting region in the center of the sphere. A strongly directed emission pattern can be obtained in LEDs with parabolically shaped surfaces. However, both hemispherical as well as parabolic surfaces are difficult to fabricate.

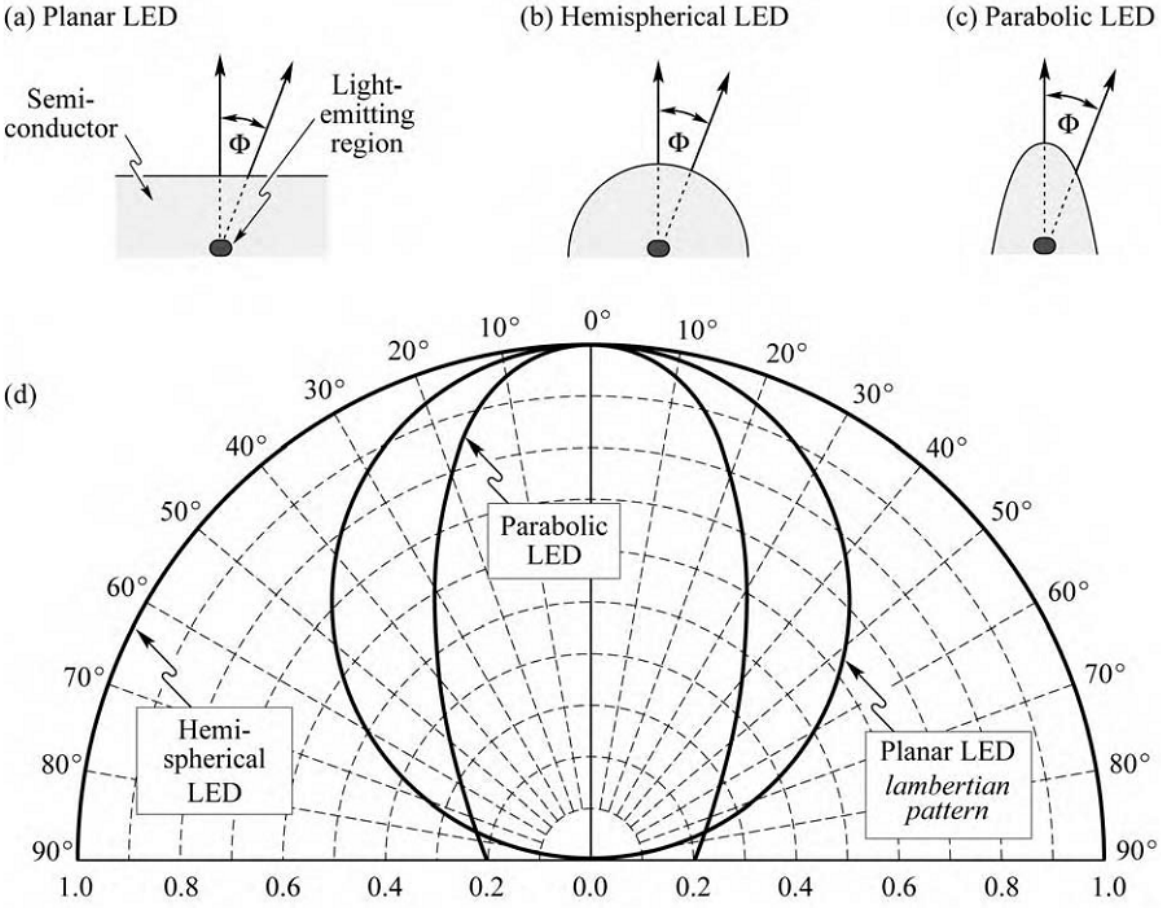


Fig. 5.5. Light-emitting diodes with (a) planar, (b) hemispherical, and (c) parabolic surfaces. (d) Far-field patterns of the different types of LEDs. At an angle of $\Phi = 60°$, the lambertian emission pattern decreases to 50% of its maximum value occurring at $\Phi = 0°$. The three emission patterns are normalized to unity intensity at $\Phi = 0°$.

LED basics: Electrical properties

4.1 Diode current–voltage characteristic

The electrical characteristics of p-n junctions will be summarized, however, a detailed derivation of the results will not be provided in this chapter. We consider an *abrupt p-n junction* with a donor concentration of N_D and an acceptor concentration of N_A. All dopants are assumed to be fully ionized so that the free electron concentration is given by $n = N_D$ and the free hole concentration is given by $p = N_A$. It is further assumed that no *compensation* of the dopants occurs by unintentional impurities and defects.

In the vicinity of an unbiased p-n junction, electrons originating from donors on the n-type side diffuse over to the p-type side where they encounter many holes with which they recombine. A corresponding process occurs with holes that diffuse to the n-type side. As a result, a region near the p-n junction is depleted of free carriers. This region is known as the ***depletion region***.

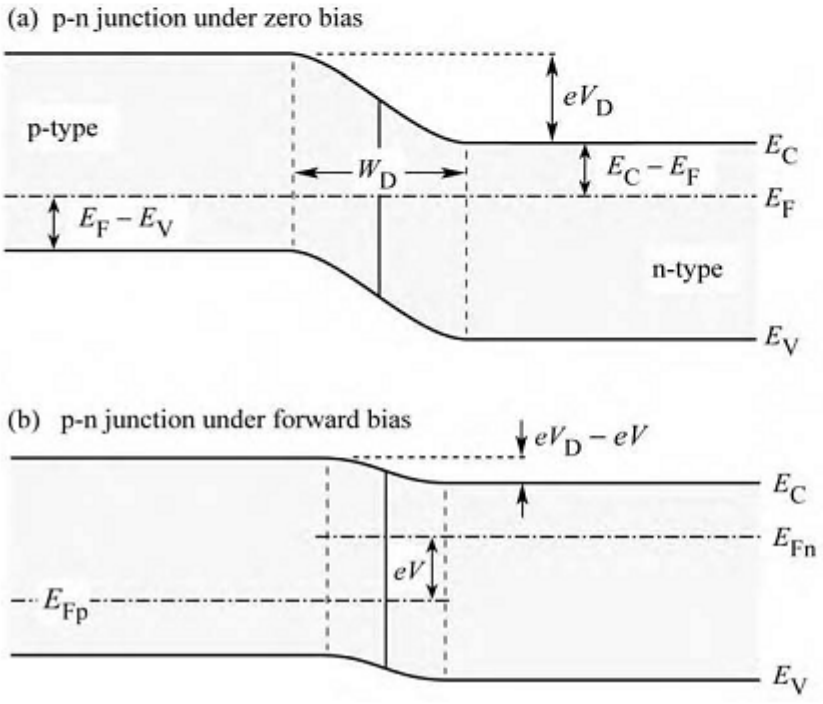


Fig. 4.1. P-n junction under (a) zero bias and (b) forward bias. Under forward-bias conditions, minority carriers diffuse into the neutral regions where they recombine.

In the absence of free carriers in the depletion region, the only charge in the depletion region is from ionized donors and acceptors. These dopants form a space charge region, i.e. donors on the n-type side and acceptors on the p-type side. The space charge region produces a potential that is called the ***diffusion voltage***, V_D. The diffusion voltage is given by

$$V_D = \frac{kT}{e} \ln \frac{N_A N_D}{n_i^2} \tag{4.1}$$

where N_A and N_D are the acceptor and donor concentrations, respectively, and n_i is the intrinsic carrier concentration of the semiconductor. The diffusion voltage is shown in the band diagram of Fig. 4.1. The diffusion voltage represents the barrier that free carriers must overcome in order to reach the neutral region of opposite conductivity type.

The Shockley equation for a diode with cross-sectional area A is given by

$$\boxed{I = eA\left(\sqrt{\frac{D_p}{\tau_p}}\,\frac{n_i^2}{N_D} + \sqrt{\frac{D_n}{\tau_n}}\,\frac{n_i^2}{N_A}\right)\left(e^{eV/kT} - 1\right)} \tag{4.3}$$

where $D_{n,p}$ and $\tau_{n,p}$ are the electron and hole diffusion constants and the electron and hole minority-carrier lifetimes, respectively.

Under typical forward-bias conditions, the diode voltage is $V >> kT/e$, and thus $[\exp(eV/kT) - 1] \approx \exp(eV/kT)$. Using Eq. (4.1), the Shockley equation can be rewritten, for forward-bias conditions, as

$$I = eA\left(\sqrt{\frac{D_p}{\tau_p}}\,N_A + \sqrt{\frac{D_n}{\tau_n}}\,N_D\right) e^{e(V-V_D)/kT} \,. \tag{4.5}$$

The exponent of the exponential function in Eq. (4.5) illustrates that the current strongly increases as the diode voltage approaches the diffusion voltage, i.e. $V \approx V_D$. The voltage at which the current strongly increases is called the ***threshold voltage*** and this voltage is given by $V_{th} \approx V_D$.

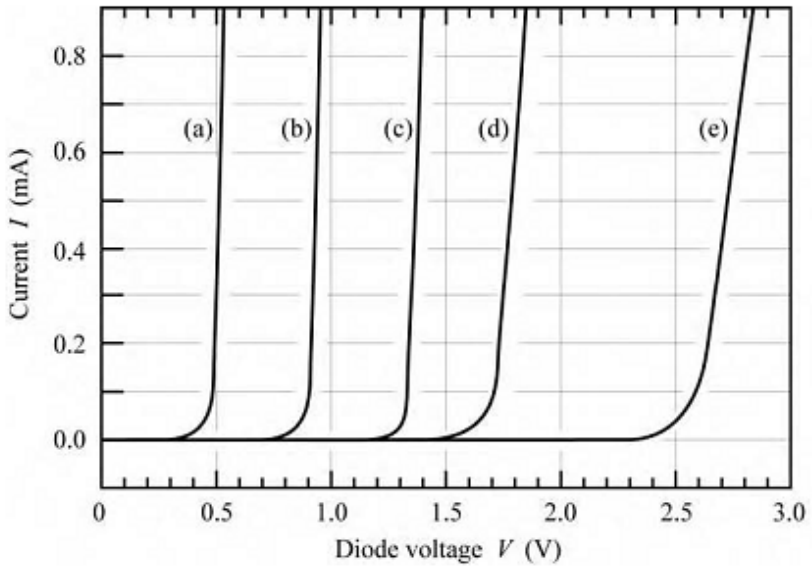
The band diagram shown in Fig. 4.1 illustrates that the following sum of energies is zero:

$$eV_D - E_g + (E_F - E_V) + (E_C - E_F) = 0 \,. \tag{4.8}$$

In highly doped semiconductors, the separation between the band edges and the Fermi level is small compared with the bandgap energy, i.e. $(E_C - E_F) << E_g$ on the n-type side and $(E_F - E_V) << E_g$ on the p-type side. Furthermore, these quantities depend only weakly (logarithmic dependence) on the doping concentration as inferred from Eqs. (4.6) and (4.7). Thus, the third and fourth summand of Eq. (4.8) can be neglected and the diffusion voltage can be approximated by the bandgap energy divided by the elementary charge

$$V_{th} \approx V_D \approx E_g/e \,. \tag{4.9}$$

Several diode *I–V* characteristics of semiconductors made from different materials are shown in Fig. 4.2 along with the bandgap energy of these materials. The experimental threshold voltages shown in the figure, and the comparison with the bandgap energy of these materials, indicates that the energy gap and the threshold voltage indeed agree reasonably well.



$T = 295$ K

(a) Ge $E_g \approx 0.7$ eV
(b) Si $E_g \approx 1.1$ eV
(c) GaAs $E_g \approx 1.4$ eV
(d) GaAsP $E_g \approx 2.0$ eV
(e) GaInN $E_g \approx 2.9$ eV

Fig. 4.2. Room-temperature current–voltage characteristics of p-n junctions made from different semiconductors.

The forward diode voltage at a diode current of 20 mA versus bandgap energy for LEDs emitting in the ultraviolet, visible, and infrared wavelength range is shown in Fig. 4.3 (Krames *et al.*, 2000; Emerson *et al.*, 2002). The solid line illustrates the expected forward diode voltage. The line equals the bandgap energy divided by the elementary charge. Inspection of the figure reveals that most semiconductor LEDs follow the solid line, except for LEDs based on III–V nitrides. This peculiarity is due to several reasons. Firstly, large bandgap discontinuities occur in the nitride material system, which cause an additional voltage drop. Secondly, the contact technology is less mature in the nitride material system, which causes an additional voltage drop at the ohmic contacts. Thirdly, the p-type conductivity in bulk GaN is generally low. Lastly, a parasitic voltage drop can occur in the n-type buffer layer.

Assuming a chip area of 250 μm × 250 μm and a current of 20 mA, the current density used in Fig. 4.3 to characterize the forward voltage is 32 A/cm^2. Typical current densities in LEDs range from 30 A/cm^2 in low-power devices to 100 A/cm^2 in high-power devices.

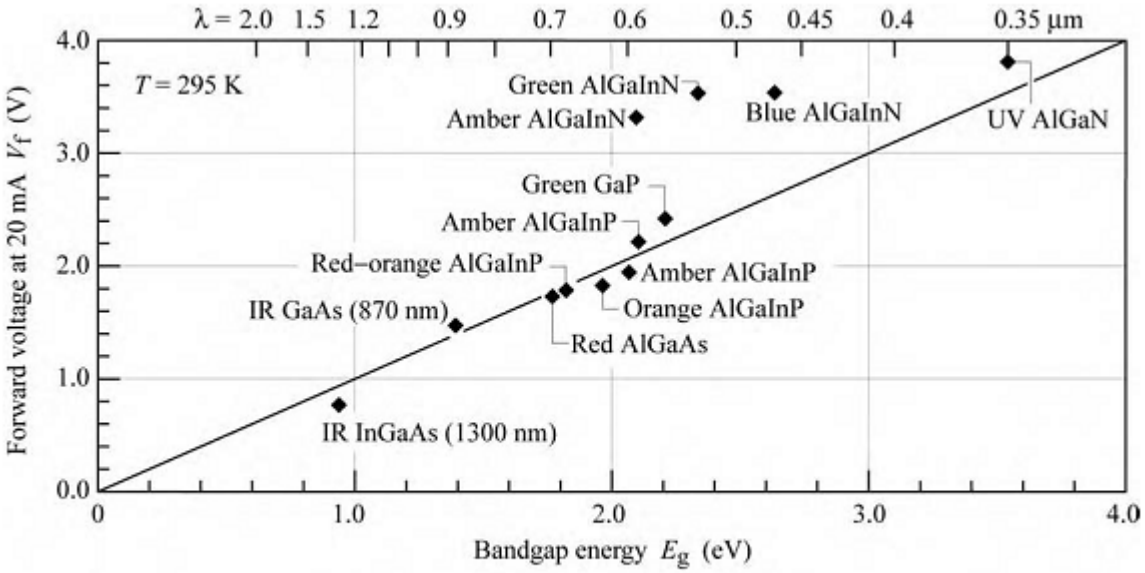


Fig. 4.3. Diode forward voltage versus bandgap energy for LEDs made from different materials (after Krames *et al*., 2000; updated with UV LED data of Emerson *et al*., 2002).

4.2 Deviations from the ideal I–V characteristic

The Shockley equation gives the expected theoretical *I–V* characteristic of a p-n junction. To describe experimentally measured characteristics, the following equation is used:

$$I = I_s \, e^{eV/(n_{ideal} kT)} \tag{4.10}$$

where n_{ideal} is the ***ideality factor*** of the diode. For a perfect diode, the ideality factor has a value of unity ($n_{ideal} = 1.0$). For real diodes, the ideality factor assumes values of typically $n_{ideal} = 1.1$–1.5. However, values as high as $n_{ideal} = 2.0$ have been found for III–V arsenide and phosphide diodes. Values as high as $n_{ideal} = 7.0$ have been found for GaN/GaInN diodes. For a detailed analysis of the diode ideality factor see, for example, Rhoderick and Williams (1988), and Shah *et al*. (2003).

Frequently a diode has unwanted or ***parasitic resistances***. The effect of a series resistance and a parallel resistance is shown in Fig. 4.4 (a). A series resistance can be caused by excessive contact resistance or by the resistance of the neutral regions. A parallel resistance can be caused by any channel that bypasses the p-n junction. This bypass can be caused by damaged regions of the p-n junction or by surface imperfections.

Occasionally, the diode turn-on is distributed over a range of voltages rather than occurring abruptly at the threshold voltage. Both types of turn-on are shown in Fig. 4.4 (b). The non-abrupt turn-on is referred to as ***sub-threshold turn-on*** or ***premature turn-on***. The sub-threshold current

can be caused by carrier transport through surface states or deep levels in the bulk of the semiconductor.

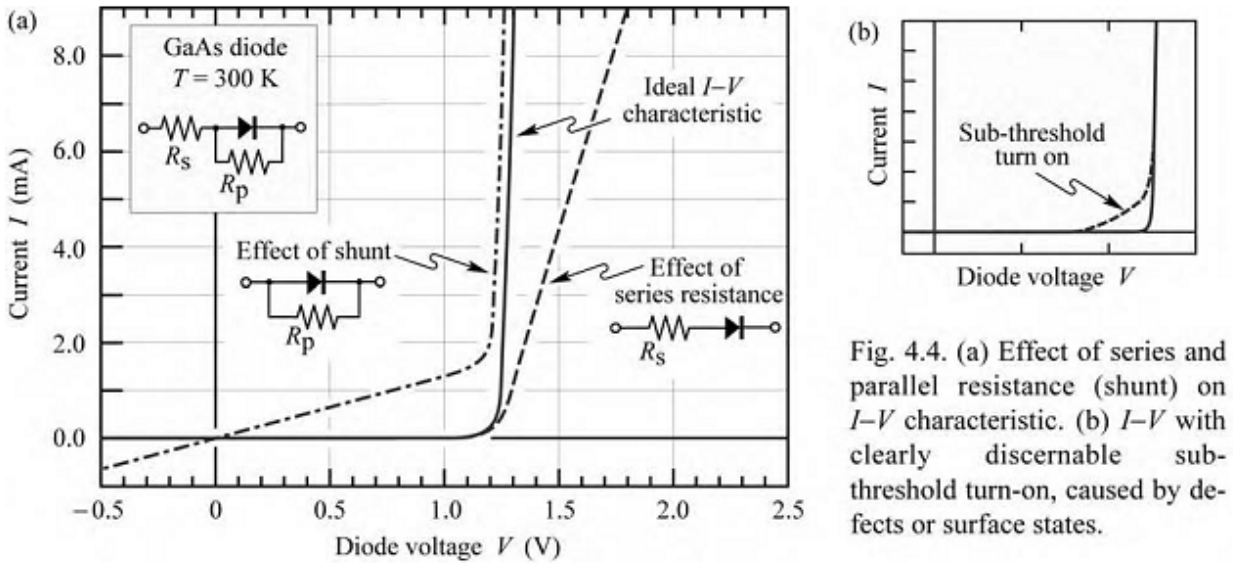


Fig. 4.4. (a) Effect of series and parallel resistance (shunt) on *I–V* characteristic. (b) *I–V* with clearly discernable sub-threshold turn-on, caused by defects or surface states.

9.6.1 Light-Current Characteristics

When a current I is passing through the forward bias diode, a certain fraction of the current is converted to light. If η_{Tot} represents the total efficiency of this conversion,

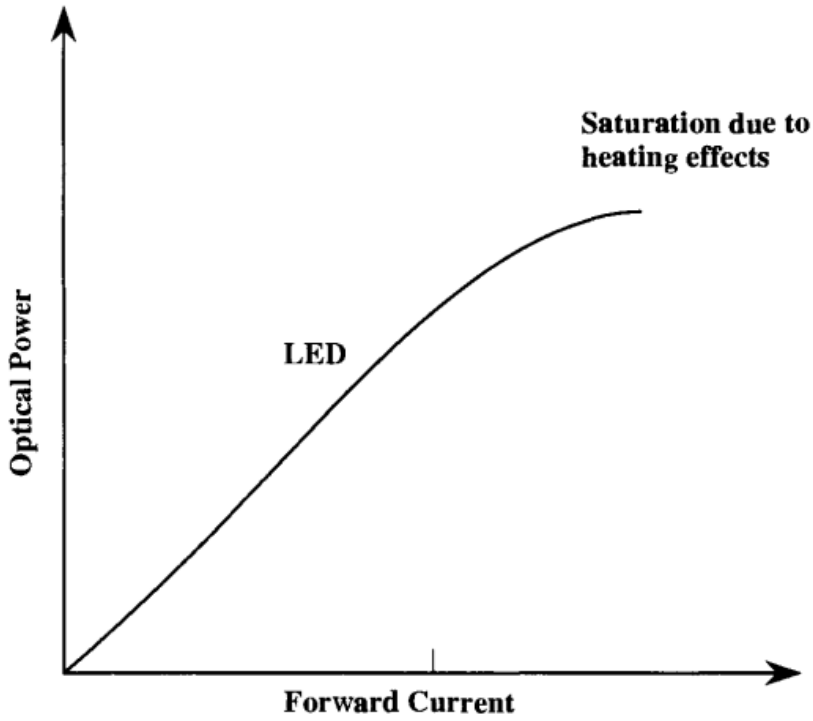


Figure 9.12: The output power of an LED is essentially linear with the injected currents.

the photon current that emerges from the diode is

$$\begin{aligned} I_{ph} &= \text{Number of photons per second} \\ &= \eta_{Tot} \cdot \frac{I}{e} \end{aligned} \tag{9.29}$$

In general, η_{Tot} depends upon the injected current since the carrier radiative lifetime τ_r depends upon carrier injection level. However, in an LED this dependence is quite weak so that the $I_{ph} - I$ characteristics are essentially linear as shown in Fig. 9.12. At very high injection the light output starts to saturate as the device heats and the radiative recombination efficiency decreases.

In surface emitting LEDs, it is often found that the light output decreases sublinearly with current, an effect that cannot be simply explained by heating effects. It appears that at high drive current the photon density in the LED becomes large enough that stimulated emissions start to occur. This emission is in the plane of the LED so that the photons, emitted perpendicular to the surface, decrease. Such LEDs are called superluminescent LEDs and behave similar to the laser diode discussed in the next chapter.

4.6.6 LED drive circuitry

As we have seen, the electrical characteristics of LEDs are essentially those of ordinary rectifying diodes. Typical operating currents are between 20 mA and 100 mA, whilst the forward

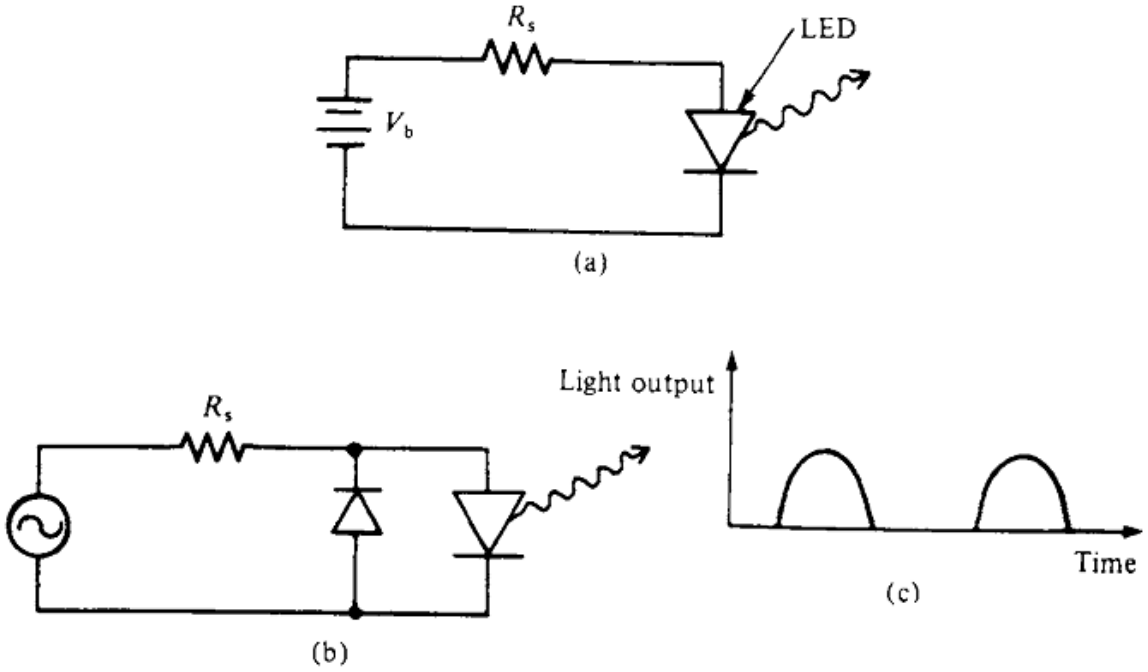


FIG. 4.19 Simple LED drive circuits for (a) d.c. operation and (b) a.c. operation. In both cases a series resistor R_s limits the maximum current flow. In the a.c. circuit a diode is placed with reversed polarity across the LED to prevent damage from excessive reverse bias voltages; (c) shows the light output obtained with the circuit of (b).

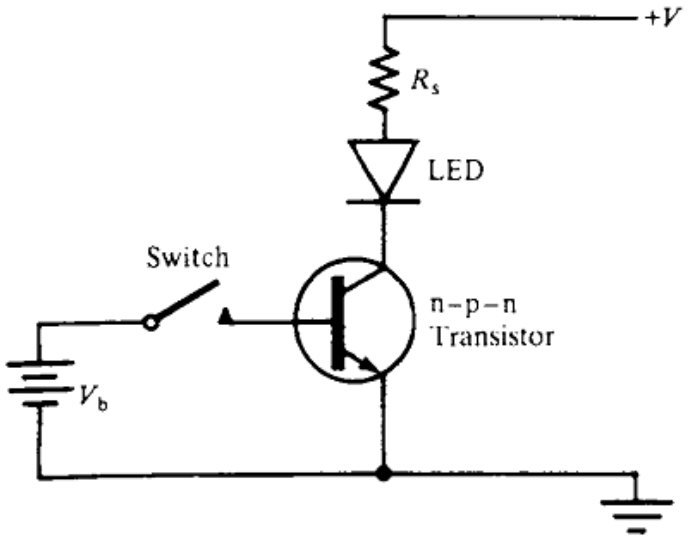
voltages vary from 1.2 V for GaAs to 2 V for GaP. (The operating voltage is approximately equal to the built-in diode potential, which in turn is slightly less than the energy gap expressed in eV.) Simple drive circuits for d.c. and a.c. voltage operation are shown in Figs 4.19(a) and (b) respectively. The current through the diode is limited by a series resistance R_s whose value may be calculated from

$$R_s = \frac{V_b - V_d}{i_d} \tag{4.20}$$

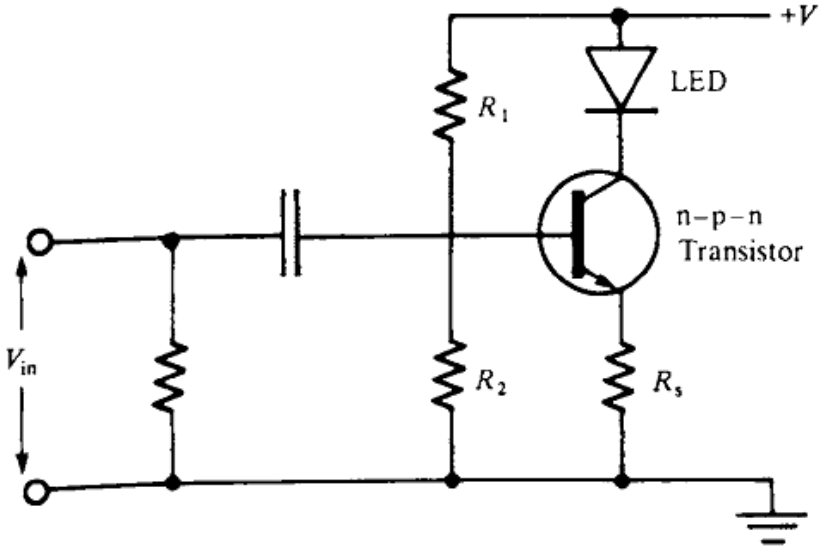
where V_b is the power source voltage, V_d the diode operating voltage and i_d the desired diode current. In the a.c. circuit, a rectifying diode is placed across the LED to protect it against reverse bias breakdown.

These two circuits provide for continuous 'on' operation. If it is desired to switch the diode on or off, or to modulate the output, then the circuits shown in Figs 4.20(a) and (b) respectively may be used. In Fig. 4.20(a) the transistor is used as a simple switch. With no voltage applied to the base, the transistor has a very high impedance between the collector and emitter and hence no current flows through the LED. If a large enough base voltage is then applied so that the emitter–base junction becomes heavily forward biased, the transistor has a rel-

atively low impedance between emitter and collector and a substantial current can flow, resulting in the LED being turned on. In Fig. 4.20(b) the transistor is biased so that the quiescent diode current is about half its peak value and both the transistor and the LED are biased well into their linear regions. Changes in the current flowing through the LED are then directly proportional to changes in the input voltage.



(a)



(b)

FIG. 4.20 LED modulation circuits: (a) provides for a simple on/off modulation via a switch. The voltage V_b is sufficient to switch the transistor on. There is then a low impedance path between the collector and emitter terminals, and the current flowing through the LED is determined by the voltage V and the series resistance R_s. In (b) the diode output may be modulated by voltage V_{in}. The resistors R_1 and R_2 bias the transistor so that the average current through the transistor, and hence through the LED, is about half the maximum value. Both the transistor and the LED are then biased well into their linear regions.

B1.1.2.3 Efficiencies

In an ideal LED, every injected electron generates one photon with an energy of $h\nu = E_g$. In this case, the number of injected carriers I/q equals the number of generated photons and the efficiency of this process is 1. However, in reality, not all injected electrons reach the active region, not all electrons in the active region can generate photons and not all photons can escape from the semiconductor. Therefore, all involved processes like carrier injection, carrier recombination or photon extraction are more or less efficient. Their efficiencies can be expressed either as quantum efficiency in terms of the number of photons and electrons involved or as power efficiency in terms of the optical or electrical power of all carriers involved.

In real LEDs, some of the electrical current is lost, e.g. due to conductive channels along the surfaces of the die or at crystal defects in the vicinity of the pn-junction. The fraction of current which is actually reaching the active region is:

$$I_a = \eta_{inj} I_0. \tag{B1.1.29}$$

The ratio $\eta_{inj} = I_a/I_0$ is called the injection efficiency.

At the pn-junction, the carriers recombine, either via radiative or non-radiative transitions. The internal quantum efficiency η_{qi} is defined as the number of generated photons per unit time over the number of electrons injected per unit time. In some cases, it is more convenient to express η_{qi} in terms of radiative and non-radiative recombination rates (equation (B1.1.13)) or the corresponding carrier lifetimes

$$\eta_{qi} = \frac{R_r}{R_r + R_{nr}} = \frac{\tau_{nr}}{\tau_r + \tau_{nr}}. \tag{B1.1.30}$$

Only photons that leave the semiconductor chip are useful. Since only a small portion of photons generated in the die can be extracted, we have to define an extraction efficiency $\eta_{extraction}$ as the number of extracted photons over the total number of generated photons.

The external quantum efficiency of an LED is the product of the internal quantum efficiency and the extraction efficiency:

$$\eta_{ext} = \eta_{qit}\eta_{extraction}. \tag{B1.1.31}$$

This is the ratio of the number of photons emitted from the LED per unit time to the number of injected electrons into the LED per unit time. The number of emitted photons from the LED is obtained by dividing the measured optical energy by the photon energy $h\nu$. The number of electrons per unit time is the injected electrical current divided by the elementary electron charge q. Thus, the external quantum efficiency, η_{ext}, can be calculated according to

$$\eta_{ext} = \frac{\left(\frac{P_{out}}{h\nu}\right)}{\left(\frac{I_0}{q}\right)} = \frac{P_{out}q}{h\nu I_0}. \tag{B1.1.32}$$

In an ideal diode, the forward voltage, V_f, of the device equals the bandgap energy divided by the elementary electron charge E_g/q. In real LEDs, however, the device structure contains additional series resistances and potential barriers which will increase the forward voltage. Therefore, we can define an external power efficiency or 'wall-plug efficiency' as

$$\eta_{wp} = \frac{P_{out}}{I_0 V_f}. \tag{B1.1.33}$$

The 'wall-plug' efficiency is the most important parameter for LED applications. It can be expressed in terms of several efficiencies defined above:

$$\eta_{\mathrm{wp}} = \eta_{\mathrm{inj}}\,\eta_{\mathrm{qi}}\,\eta_{\mathrm{extraction}}\,\frac{h\nu}{qV_{\mathrm{f}}}. \tag{B1.1.34}$$

The different definitions are summarized in table B1.1.1.

Table B1.1.1. Summary of different definitions of internal and external efficiencies. The last expression for the extraction efficiency is valid for LEDs with a planar structure.

	Quantum efficiency	Power efficiency
Internal efficiency	$\eta_{\mathrm{qi}} = \dfrac{\#\text{ of generated photons}}{\#\text{ of injected electrons}}$	$\eta_{\mathrm{pi}} = \dfrac{\text{power of one photon}}{\text{power of one electron}}$
	$\eta_{\mathrm{qi}} = \dfrac{R_{\mathrm{r}}}{R_{\mathrm{r}} + R_{\mathrm{nr}}} = \dfrac{\tau_{\mathrm{nr}}}{\tau_{\mathrm{r}} + \tau_{\mathrm{nr}}}$	$\eta_{\mathrm{pi}} = \dfrac{h\nu}{qV}$
External efficiency	$\eta_{\mathrm{ext}} = \dfrac{\#\text{ of extracted photons}}{\#\text{ of injected electrons}}$	$\eta_{\mathrm{wp}} = \dfrac{\text{extracted optical power}}{\text{injected electrical power}}$
	$\eta_{\mathrm{ext}} = \dfrac{P_{\mathrm{out}}q}{h\nu I_0}$	$\eta_{\mathrm{wp}} = \dfrac{P_{\mathrm{out}}}{I_0 V_{\mathrm{f}}} = \eta_{\mathrm{inj}}\,\eta_{\mathrm{qi}}\,\eta_{\mathrm{extraction}}\,\dfrac{h\nu}{qV_{\mathrm{f}}}$
Injection efficiency	$\eta_{\mathrm{inj}} = \dfrac{\#\text{ of carriers at pn junction}}{\#\text{ of injected carriers}} = \dfrac{I_{\mathrm{a}}}{I_0}$	
Extraction efficiency	$\eta_{\mathrm{extraction}} = \dfrac{\#\text{ of extracted photons}}{\#\text{ of generated photons}}$	
	$\eta_{\mathrm{extraction}} = \frac{1}{2}(1 - \cos\vartheta_{\mathrm{c}})\left(1 - \dfrac{(n_{\mathrm{m}} - n_{\mathrm{s}})^2}{(n_{\mathrm{m}} + n_{\mathrm{s}})^2}\right)$	

I_0, total injected current; I_{a}, current flow into the pn-junction; R_{rad}, radiative recombination rate; R_{nr}, non-radiative recombination rate; τ_{r}, radiative carrier lifetime; τ_{nr}, non-radiative carrier lifetime; P_{out}, optical power emitted from the LED; V, applied voltage; ϑ_{c}, angle of total reflection; n_{s}, refractive index of the semiconductor material; n_{m}, refractive index of the surrounding medium; #, number.

HIGH-EFFICIENCY LEDS

1. Structures for high-efficiency LEDs

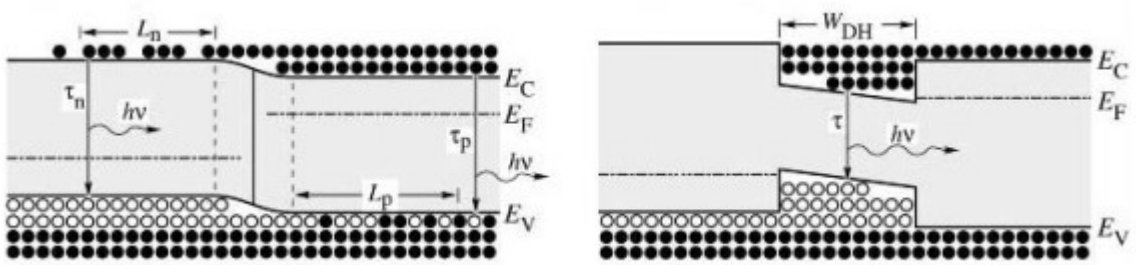
Contrary to semiconductors used in purely electronic circuits, for LEDs it is desirable that all the injected carriers recombine in the active region to form photons. The active region is usually the lowest bandgap region within the depletion region of a *p–i–n* diode. The carrier distribution in *p–n* homojunctions, i.e. *p–n* junctions consisting of a single semiconductor material, is governed by the carrier diffusion properties. In the absence of an external electric field, the minority carriers diffuse into the region with opposite conductivity with a mean distance corresponding to the diffusion length, L_n and L_p for electrons and holes, respectively. These values are typically of the order of several micrometers. This means that the minority carriers are distributed over a large region. The spontaneous radiative recombination rate R_{sp} is given by the bimolecular recombination equation

$$R_{sp} = Bnp \tag{6.1}$$

which means that the rate is proportional to the carrier concentration in the active region (B is the radiative bimolecular recombination coefficient). However if the carriers are only limited by the diffusion their concentration will never be very high, not even for large current injection levels. The carriers therefore need to be confined by other means. In addition in a homojunction the emitted photons tend to be reabsorbed in the semiconductor before arriving at the surface.

1.1 Double heterostructure

The efficiency of LEDs can be drastically improved with the use of a heterojunction, i.e. a junction between different semiconductor materials. By injecting the carriers from a larger bandgap semiconductor into a narrow bandgap active region they are confined to the low bandgap region as the band offsets act as barriers. Thus with this so-called ***double heterostructure*** (DH) higher carrier concentrations can be achieved, as is depicted schematically in figure 1. Furthermore the photons emitted are not absorbed in the wider bandgap confinement layers, as the photon energy is smaller than the bandgap of the barriers.

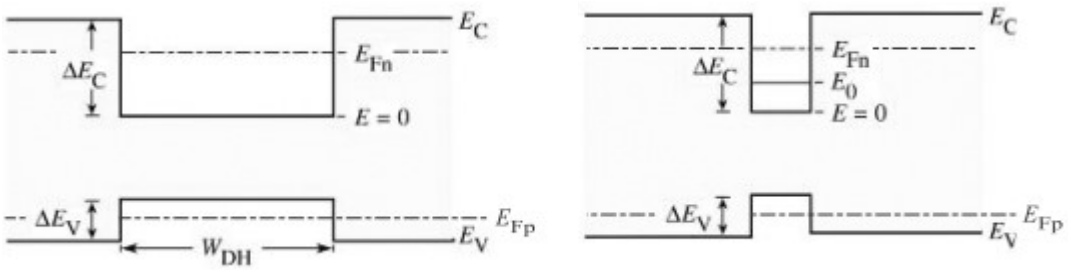


(*a*) (*b*)

Fig. 1 – *Free carrier distribution in a homojunction (left) and a heterojunction (right) under forward bias. In homojunctions carriers are distributed over the diffusion length whereas in heterojunctions they are confined to the well region with thickness* W_{DH}

1.2 Quantum wells (QWs)

A ***quantum well*** (QW) is a double heterojunction structure with a very thin (≤ 50nm) narrow bandgap layer. Under forward bias, approximately rectangular quantum wells appear in the conduction and valence band as shown schematically in figure 2. As the active layer thickness comes close to the *De-Broglie wavelength* (about 10 nm for semiconductor laser devices) quantum effects become apparent with the introduction of other *inter-band levels* (e.g. E_0 level in figure 2.b).



(*a*) (*b*)

Fig. 6.2 – *Fermi level* (E_{Fn}) *and subband level* (E_0) *in a double heterostructure (a) and a quantum well structure (b) at high injection levels*

The size quantization can be used to increase the emission energy. Due to their small thickness QWs can be strained by being lattice-mismatched without introducing any undesired defects. This further increases the accessible wavelength range compared to bulk emission. The thickness reduction leads to high carrier densities. As a result, the carrier lifetime for radiative recombination is reduced and the radiative efficiency is increased. In addition due to the reduced active region thickness the re-absorption of emitted photons is drastically reduced.

However quantum wells are saturated at lower injection levels compared to bulk active regions and hence for high power devices ***multi-quantum well*** (MQW) structures are employed. In that case the barriers between the different wells need to be sufficiently transparent, i.e. low or thin, in order to allow for efficient transport between the wells and to ensure a homogeneous carrier distribution.

1.3 Separate-confinement heterostructure (SCH)

The lower bandgap region in a double heterostructure usually also has a higher index of refraction. Therefore not only the carriers but the photons as well are confined to the narrow bandgap active region and the structure acts like a transverse dielectric optical waveguide or a vertical optical cavity. However with the use of a quantum well this optical confinement effect is sacrificed. Therefore different heterostructures are used for the electrical and the optical confinement. Such ***separate-confinement heterostructures*** (SCH) consist of thin quantum well carrier-confining active regions and a surrounding intermediate bandgap separate photon confinement region. By a grading of the index in the outer heterobarriers the overlap between the optical standing wave and the quantum well regions can be further improved, this configuration being called ***graded-index separate-confinement heterostructure*** (GRINSCH). It is shown in figure 3.

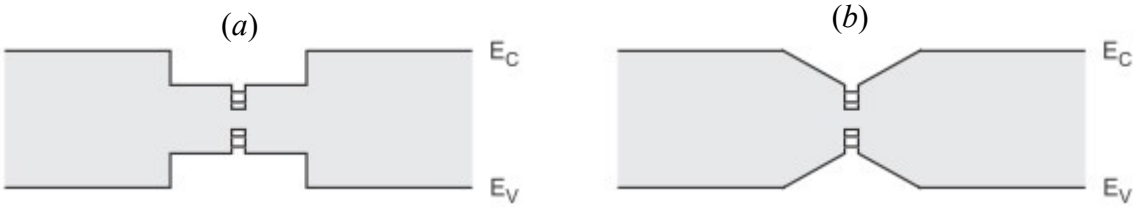


Fig. 3 – *Transverse band structure for two different separate-confinement heterostructures (SCHs): standard SCH (left) and graded-index SCH (GRINSCH) (right). The electric field (photons) is confined by the outer step or graded heterostructure; the central quantum well confines the electrons.*

1.4 Carrier loss

Ideally the injected carriers are confined to the active region by the barrier layers adjoining the active region. This way a high carrier concentration is attained resulting in a high radiative efficiency. For an efficient confinement the barriers must be much larger than the thermal energy of the carriers, which is equal to kT. However, the energy distribution of the free carriers in the active region is given by the Fermi–Dirac distribution. Thus a certain fraction of the carriers has a higher energy than the barriers. These carriers can diffuse into the barrier layers and recombine there instead of in the active region. The mechanism of carrier loss in the active region is illustrated in figure 4.

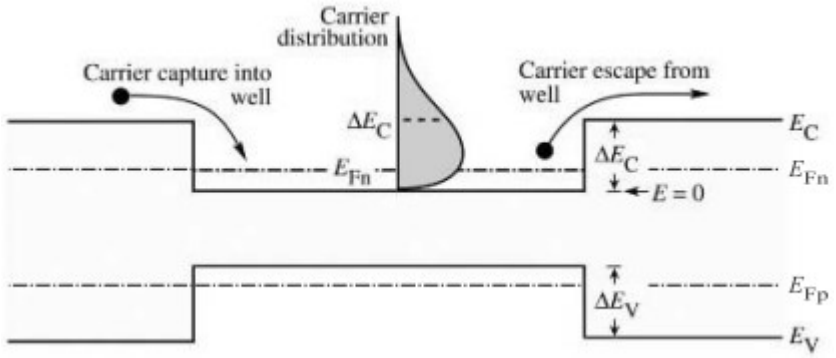


Fig. 4 – *Carrier capture and escape in a double heterostructure. Also shown is the energy distribution of the free carriers in the active layer*

For the AlGaAs/GaAs and the InGaAs/AlGaAs material system the barriers are relatively high, but in case of GaInP/AlGaInP they are significantly lower, both for electrons and holes. This results in significantly higher leakage currents for this material system.

As the carrier energy distribution strongly depends on the temperature, the carrier leakage increases exponentially with temperature.

In addition to temperature, another mechanism can negatively affect the concentration of charges in the active region. At high injection levels carrier start to spill over from the active region into the confinement region. With increasing injection current density the carrier concentration in the active region increases and the Fermi energy rises. For sufficiently high current densities, the Fermi energy will rise to the top of the barrier. At that point the active region is flooded with carriers ("*carrier overflow*") and a further increase in injection current density will not increase the carrier concentration in the active region. As a result the optical intensity saturates.

The problem of carrier overflow is more severe in QW structures. For a QW-based LED structure it has been shown that the saturation level of the optical intensity is proportional to the number of quantum wells, as shown in figure 5.

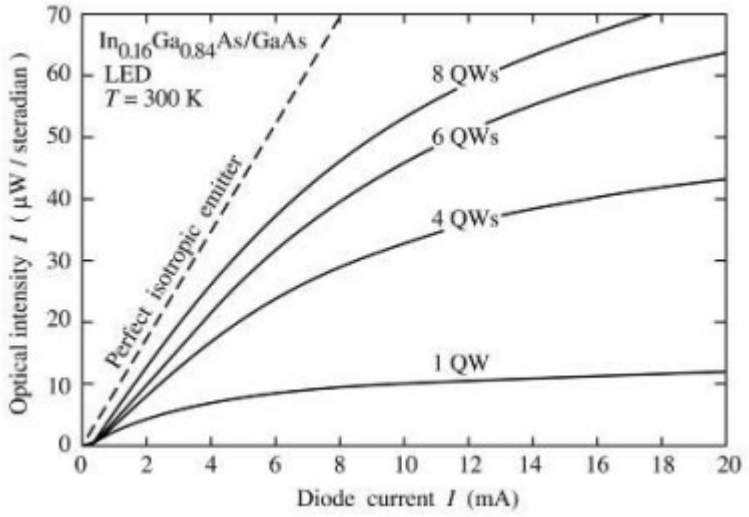


Fig. 5 – *Optical intensity emitted by $In_{0.16}Ga_{0.84}As/GaAs$ LEDs with active regions consisting of 1, 4, 6 and 8 QWs and theoretical intensity of a perfect isotropic emitter (dashed line)*

Eventually, carrier overflow in high-current devices can be avoided by employing either thick DH active regions, multi quantum well (MQW) structures or a large injection area, which means a large contact size.

1.5 Electron-blocking layers

We have seen that the charges tend to escape from the active region of the LED towards the confinement layers. Charge leakage can be substantial in double heterostructures with low confinement barriers. Furthermore, the high temperatures favor the loss of charges from the active region due to the increase in thermal energy.

The electron loss current is greater than the hole loss current as the diffusivity of the electrons is usually greater than that of the holes in group III-V semiconductors. In order to reduce the loss of charges from the active region, layers with a wide bandgap of energy can be effectively used, acting as ***electron-blocking layers***. These layers are deposited at the interface between the active region and the confinement one; thanks to their high bandgap they block the flow of electrons.

The band diagram of a GaInN LED having an electron barrier layer is shown in figures 6. The LED has two AlGaN confinement layers and an active MQW region in GaInN/GaN. An electron-blocking layer in AlGaN (but with a higher percentage of Al) is placed between the *p*-type confinement layer and the active region. In particular, figure 6.a, which represents the band structure in the case of non-doped confinement layers, shows how the electron-blocking layer acts as a barrier for the current both in the conduction band and in the valence band. However, if the structure is doped, as illustrated in figure 6.b, the barrier in correspondence with the valence band is shielded by the free charges, thus resulting lower; in this way there is no barrier to the flow of holes in the *p*-type confinement layer.

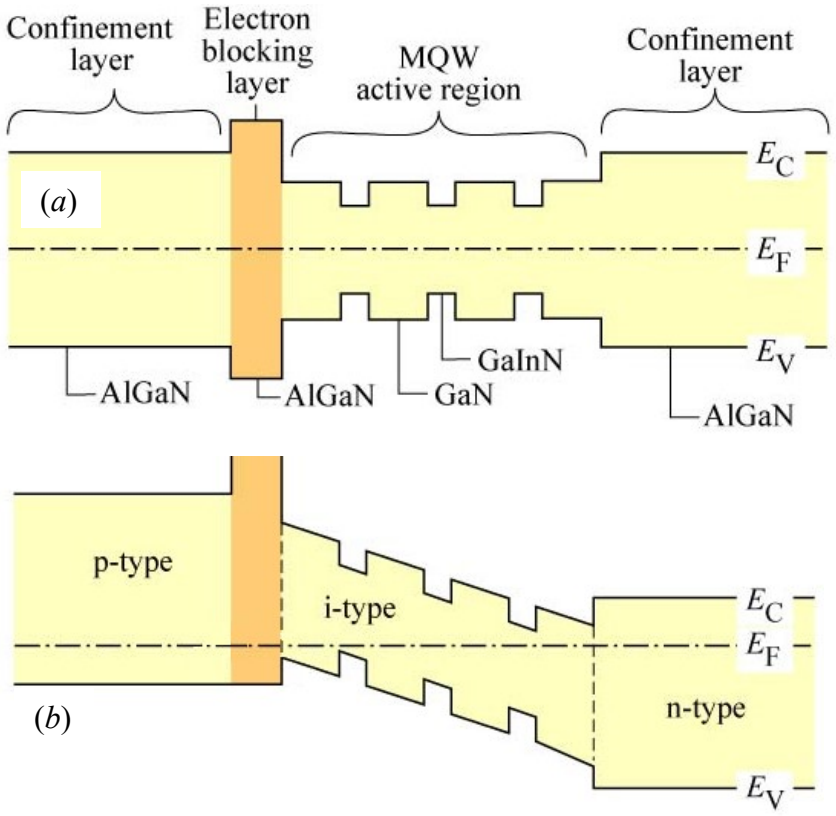


Fig. 6 – *LED structures with electron-blocking layer in the case of: a) undoped, b) doped confinement layers*

2. Radiative and non-radiative recombination mechanisms

Consider a forward biased *p–i–n* junction with a single quantum well of thickness t_{qw} and volume V_{qw} in the center of a nominally undoped intrinsic region. The electron density in the quantum well can then be described by the following rate equation

$$\frac{dn}{dt} = G - R - R_l \tag{6.2}$$

where G is the rate of injected electrons, R the rate of recombining electrons per unit volume in the active region, neglecting *photon recycling*, and R_l the carrier leakage rate. The carrier leakage rate includes vertical carrier spill-over out of the active region and lateral diffusion out of the cavity region.

The generation rate G corresponds to the fraction of the current being injected into the active region,

$$G = \frac{\eta_{\text{inj}} I}{e V_{qw}} \tag{6.3}$$

with I being the applied current and η_{inj} the injection efficiency. The recombination rate is the sum of the ***radiative recombination rate***, R_{rad}, and the ***non-radiative recombination rate***, R_{nr}. For lasers the ***stimulated recombination rate***, R_{st}, would have to be considered as well, but can usually be neglected for LED structures.

$$R = R_{\text{rad}} + R_{\text{nr}} + R_{\text{st}}\,. \tag{6.4}$$

Under steady-state conditions ($\mathrm{d}n/\mathrm{d}t = 0$)

$$\frac{\eta_{\text{inj}} I}{e V_{qw}} = R_{\text{rad}} + R_{\text{nr}} + R_l \tag{6.5}$$

The radiative efficiency η_{rad} is defined as the ratio of the number of photons generated in the active region per unit time per unit volume to the number of electrons injected in the active region per unit time per unit volume. Thus, it can be written as

$$\eta_{\text{rad}} = \frac{R_{\text{rad}}}{R_{\text{rad}} + R_{\text{nr}} + R_l} \tag{6.6}$$

The band-to-band radiative recombination rate R_{rad} is:

$$R_{\text{rad}} = Bnp = Bn^2 \tag{6.7}$$

where n and p are the electron and hole concentrations in the active region, respectively. At high injection levels charge neutrality requires $n = p$ in the active region. B is the radiative bimolecular recombination coefficient and is of the order of 10-10 cm^3/s for InGaAs/GaAs structures.

The non-radiative recombination rate R_{nr} includes ***deep-level defect*** and ***impurity recombination*** in the depletion layer, ***surface and interface recombination*** and ***Auger recombination***, which can

be an important mechanism at very high injection levels. The three mechanisms are represented in figure 7.

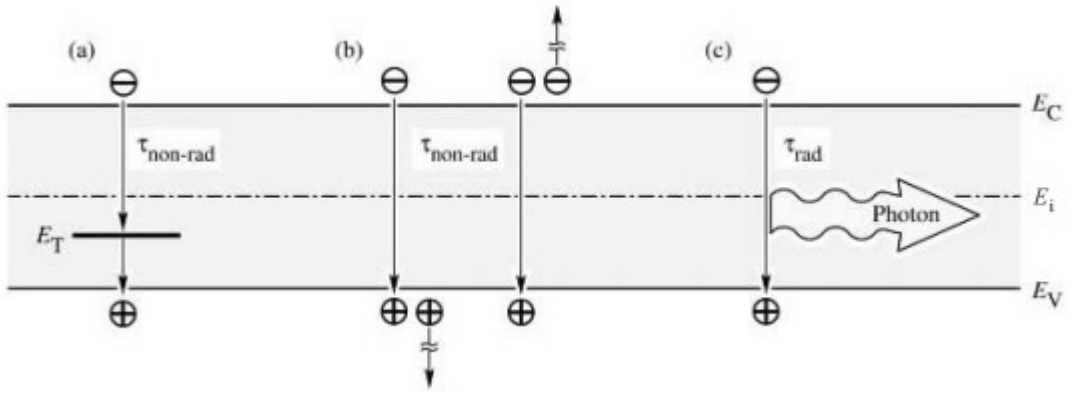


Fig. 7 – *Band diagram illustrating recombination mechanisms: via a deep level (a), via an Auger process (b) and radiative recombination (c) The first two ones are non-radiative.*

The non-radiative recombination rate R_{nr} can be expressed as follows:

$$R_{nr} = An + Cn^3; \tag{6.8}$$

A corresponds to the non-radiative recombination coefficient (in units of [s−1]) and C to the Auger coefficient. Typical values for C are 10^{-28}-10^{-29} cm^6/s for III-V semiconductors. For InGaAs/GaAs devices at room temperature the Auger recombination can be neglected

In the GaAs/AlGaAs and InGaAs/GaAs systems surface recombination can be quite severe for small devices. Assuming a uniform carrier distribution over the whole mesa the non-radiative recombination coefficient can be expressed as

$$A = A_0 + \frac{a_s v_s}{V_{qw}} = A_0 + \frac{4v_s}{d} \tag{6.9}$$

where A_0 stands for defect and impurity recombination, as is the exposed quantum well surface area, V_{qw} the volume of the active region, v_s the surface recombination velocity and d the device diameter. For *strained* InGaAs/GaAs QW lasers the surface recombination velocity has been found to be of the order of 2×10^5 cm/s.

Expression (6.6) for the radiative efficiency can therefore be rewritten in the following way for InGaAs/GaAs LEDs:

$$\eta_{\text{rad}} = \frac{Bn^2}{Bn^2 + An + R_l} \tag{6.10}$$

with A including surface and interface recombination. The deep-level defect and impurity recombination in the depletion layer can be minimized with the use of high purity and low defect density substrates and epitaxial structures. With an appropriate device design the injection efficiency η_{inj} and the carrier density in the active region can be maximized and the carrier leakage rate can be minimized at the same time. With all these measures internal quantum efficiencies close to unity are possible in modern devices. The external quantum efficiencies of LEDs are therefore essentially limited by their extraction efficiencies.

3. Light Extraction Problem

The ***extraction efficiency*** is defined as the fraction of the light generated in the active region which is extracted out of the device. It depends on the optical properties of the device, on its layer structure and on its geometry. Limiting factors are *re-absorption in the semiconductor* and *internal reflection at the interface semiconductor–air*. In addition for top emitting devices the extraction efficiency can be significantly reduced by *shadowing of the emission by the top contact*. Therefore the extraction efficiency can be expressed as the product of an absorption factor γ_{abs}, a reflection factor γ_{refl}, and a shadowing factor γ_{sh}

$$\eta_{\text{extr}} = \gamma_{\text{abs}}\gamma_{\text{refl}}\gamma_{\text{sh}} \,. \tag{6.11}$$

The amount of re-absorption depends on the layer structure and can usually be minimized by using materials with a bandgap energy larger than the photon energy or by limiting their thicknesses. The top contact shadowing depends on the device design, often there exists a trade-off between improved current spreading and reduced extraction for a more dense top contact pattern.

In case of an isotropic internal emission of the active region, the amount of reflection at the interface semiconductor–air is dictated by the device geometry and by the refractive index of the semiconductor layer at the interface. Due to the high refractive index of most semiconductors, ***total internal reflection*** (TIR) at the interface semiconductor–air drastically limits their extraction efficiency. In the following the limits for γ_{refl} for different device geometries are estimated. Since γ_{abs} and γ_{sh} are assumed to be equal to unity, $\eta_{\text{extr}} \approx \gamma_{\text{refl}}$.

3.1 The Light Escape Cone

As shown in figure 8, at the semiconductor-external medium interface, an incident ray is divided into two: a reflected and a refracted (or transmitted) ray.

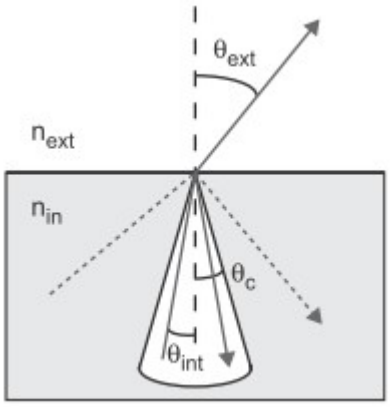


Fig. 8 – *Transmission and reflection at the interface semiconductor–air; θ_c corresponds to the critical angle of total internal reflection, defining the light escape cone*

As well-known, the transmitted ray obeys ***Snell's law of refraction***

$$n_{\text{int}} \sin\theta_{\text{int}} = n_{\text{ext}} \sin\theta_{\text{ext}} . \tag{6.12}$$

where n_{int} and θ_{int} are the refractive index and the incident angle from the normal in the high index semiconductor, n_{ext} is the refractive index of the low index outside medium (typically air) and θ_{ext} the angle of the refracted beam in the outside medium. Total internal reflection occurs for $\theta_{\text{int}} \geq \theta_c$, with θ_c being the ***critical angle*** of total internal reflection (obtained by (6.12) for $\theta_{\text{ext}} = 90°$)

$$\theta_c = \arcsin\left(\frac{n_{\text{ext}}}{n_{\text{int}}}\right). \tag{6.13}$$

For larger angles of incidence, $\theta_{\text{int}} > \theta_c$, θ_{ext} becomes complex and the entire electromagnetic energy is reflected back, apart from an evanescent wave in the low index material which decreases exponentially away from the interface.

The critical angle defines the ***light escape cone*** in the semiconductor. The critical solid angle Ω_c corresponding to θ_c is

$$\Omega_c = 2\pi(1-\cos\theta_c) \tag{6.14}$$

while the total solid angle is 4π. For $\theta_{\text{int}} \leq \theta_c$ the reflected and transmitted intensities of a ray are equal to

$$I_t = T I_i = (1-R) I_i \tag{6.15.a}$$

$$I_r = RI_i \tag{6.15.b}$$

where I_i, I_t and I_r are the intensities of the incident, transmitted and reflected ray, respectively. T is the power transmission coefficient and R the Fresnel power reflection coefficient. In case of TE polarization (named also "*s polarization*") R is equal to

$$R = \left(\frac{n_{\text{int}} \cos\theta_{\text{int}} - n_{\text{ext}} \cos\theta_{\text{ext}}}{n_{\text{int}} \cos\theta_{\text{int}} + n_{\text{ext}} \cos\theta_{\text{ext}}} \right)^2 . \tag{6.16}$$

For normal incidence ($\theta_i = 0$) this expression simplifies to

$$R_0 = \left(\frac{n_{\text{int}} - n_{\text{ext}}}{n_{\text{int}} + n_{\text{ext}}} \right)^2 . \tag{6.17}$$

Assuming an isotropic internal emission and $R \approx R_0$ for all angles $\theta \leq \theta_c$, the fraction of the light emitted in a semiconductor that can be extracted through the top surface is therefore given by

$$\eta_{\text{extr}} \approx \gamma_{\text{refl}} = \frac{2\pi(1-\cos\theta_{\text{c}})}{4\pi}(1-R_0) \tag{6.18}$$

For the interface GaAs–air, the (6.18) formula leads to an extraction efficiency between 1 and 2% only. By encapsulating the LED in a hemispherical epoxy dome with a refractive index of approximately 1.5 the light extraction can be significantly enhanced. Due to the larger index of the outside medium (that, now, is not the air but the epoxy), the critical angle is increased and the reflection at normal incidence R_0 is reduced. Furthermore, the critical angle for the interface epoxy–air is large due to the small index difference.

If the LED is placed at the center of the hemispherical epoxy dome, all the light rays hit the interface epoxy–air at normal incidence and the device shows a quasi-isotropic emission. Alternatively the LED can be placed at the ***Weierstrass point***, i.e. at a distance of r_s/n_{epoxy} below the center of the sphere with radius r_s, which leads to a directional emission. This geometry is called a ***Weierstrass sphere***.

Thus extraction efficiencies between 3 and 4% can be achieved with an epoxy dome.

The different numerical values for GaAs are summarized in table 1 and table 2 for emission into air and epoxy, respectively. The values are calculated both at 650 and 970 nm wavelengths.

Tab. 1 – *Extraction properties for interface GaAs–air at 650 and 970 nm*

Wavelength [nm]	n_{GaAs}	n_{air}	θ_c [°]	$\Omega_c/4\pi$ [%]	R_0 [%]	γ_{refl} [%]
650	3.83	1	15.2	1.7	34	1.1
970	3.52	1	16.5	2.1	31	1.4

Tab. 2 – *Extraction properties for interface GaAs–epoxy at 650 and 970 nm*

Wavelength [nm]	n_{GaAs}	n_{epoxy}	θ_c [°]	$\Omega_c/4\pi$ [%]	R_0 [%]	γ_{refl} [%]
650	3.83	1.5	23.1	4.0	19	3.2
970	3.52	1.5	25.2	4.8	16	4.0

The reflection at the interface epoxy–air is about 4% for normal incidence. By taking this into account the extraction efficiencies for devices encapsulated in a hemispherical epoxy dome diminish slightly to 3.1 and 3.8% at 650 and 970nm, respectively. Even with an epoxy dome the maximum attainable extraction efficiencies are therefore extremely low for planar structures with a single facet emission.

3.2 Optimization Standard Geometry

The standard LED geometry is a rectangular parallelepiped, fabricated by *cleaving* the wafer along its crystallographic axes. The active region is assumed to be close to the surface and the substrate absorbing, due to the fact that it has a lower bandgap than the active region (figure 9.a). Such a device has a total of six light escape cones, two of them perpendicular to the wafer surface and four of them parallel to it. The light emitted into the bottom escape cone is entirely absorbed in the substrate. Apart from a small area close to the edges, the emission in the bottom part of the four in-plane escape cones is absorbed in the substrate too, as well as a large amount of the top part, which is totally internally reflected at the top surface and redirected towards the substrate. The light in the top escape cone is in part obstructed by the top contact. Thus the extraction efficiency of this device geometry is very low, as its reflection factor is close to the value calculated above for a single escape cone. Its external quantum efficiency can be expressed as

$$\eta_{ext} = \eta_{int}\gamma_{refl}\gamma_{sh}\,. \tag{6.19}$$

Even for an internal efficiency of 100 % and no shadowing, the theoretically achievable external quantum efficiency for this geometry is smaller than 2 % for emission into air and below 4 % for emission into epoxy.

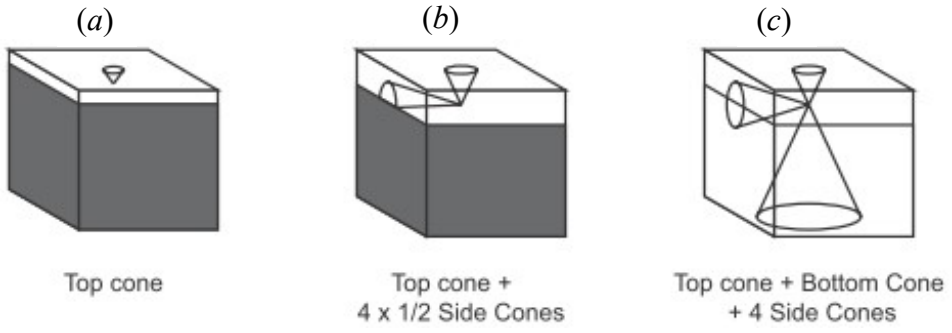


Fig. 9 – *Simplified schematic illustration of light extraction for various LED designs based on a standard rectangular parallelepiped; absorbing substrate with thin window layer (a), absorbing substrate with thick window layer (b) and transparent substrate with thick window layer (c). To simplify matters only one of the four side cones is shown.*

A first approach to increase the extraction efficiency of the standard LED structure is to add a ***thick window layer*** on top of the epitaxial structure. This window layer needs to be optically transparent and electrically conductive. With increasing thickness of this layer the extraction of the top part of the four in-plane escape cones increases until the entire half-cones are extracted, as depicted schematically in figure 9.b. In addition the shadowing of the light in the top escape cone by the top contact can be reduced thanks to an improved current spreading. With this device design the extraction efficiency can therefore be increased by a factor of 3, as in addition to the top cone four more half cones can be extracted.

In order to extract all cones, a ***transparent substrate*** needs to be used in addition, as shown in figure 9.c. The bottom escape cone can then either be extracted via the substrate side or reflected to the top by a bottom mirror. With this design a near 6-fold increase in extraction efficiency can be achieved because it is possible to extract all the six cones.

The maximum achievable extraction efficiencies are summarized in table 3. These are simplified approximations that do not take into account internal absorption or multiple pass reflection effects such as photon recycling. The actual limits might be a few percent higher. The shadowing by the top contact is not included either in these estimations.

Tab. 3 – *Extraction improvements for standard LED geometry (neglecting absorption, shadowing and multiple pass reflection effects)*

Design	Extracted cones	γ_{refl}(air) [%]	γ_{refl}(epoxy) [%]
standard geometry	1	1-2	3-4
thick window layer	3	3-6	9-12
transparent substrate	6	6-12	18-24

For a cylindrical shape the four in-plane escape cones are replaced by an escape ring. Therefore cylindrical LEDs show higher extraction efficiencies than comparable cubic LEDs. However they can only be fabricated by *etching* and not by *cleaving*, so it not generally used.

The ideal device geometry would be a sphere (figure 10.a) with a point source as active region in the center of the LED, or a hemisphere with a perfect bottom mirror (figure 10.b). No total internal reflection would occur in such a device as all the rays are incident at a normal angle at the interface semiconductor–air. However, unless the sphere is coated with an anti-reflection coating, the rays would still undergo Fresnel reflection at this interface.

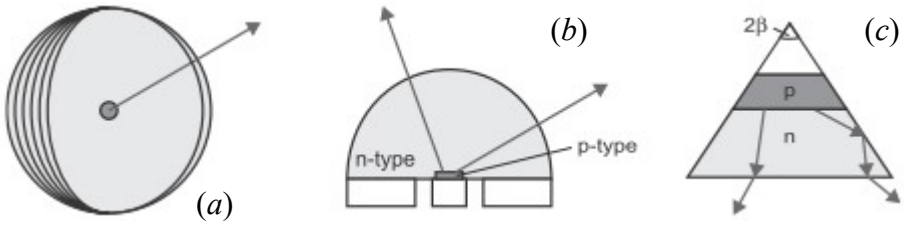


Fig. 10 – *Cross-section schematics of ideal geometries; sphere with point source (a), hemisphere (b) and a truncated cone (c)*

Comparable extraction efficiencies can be expected from the already-mentioned ***Weierstrass spheres***, or **truncated ellipsoids** or ***truncated cones***, in case of the latter for an optimum apex half-angle of $\beta = (\pi / 2 - \theta_c) / 2$.

The problem of light extraction had been realized already in the very beginning of LEDs in the 1960s. Hemispherical shapes and truncated cones had been demonstrated to improve the extraction efficiency over conventional cubic or cylindrical designs. However these device geometries are not very compatible with semiconductor fabrication technologies, optimized for planar processes in view of the flat substrates used in epitaxial growth and were therefore not developed further for a long time.

4. Alternative Types of High-Efficiency LEDs

The evolution of high efficiency LEDs can be divided into three eras. First of all the device design was optimized. Thanks to the introduction of double heterostructures, quantum wells and separate confinement heterostructures the injection efficiencies in modern devices can often be assumed to be close to 100%. With the evolution of the growth methods the crystal quality was constantly optimized, leading to drastically improved radiative efficiencies. Together with the increased carrier densities in the active regions and the reduced leakage rates, internal quantum efficiencies close to unity could be achieved. The problem of light extraction, finally, was realized already in the 1960s but was not addressed until the 1990s.

The evolution of the extraction efficiency can be shown by means of the chronological improvements of the external quantum efficiency for AlGaInP LEDs by Hewlett-Packard (now Lumileds), which follows the approaches presented in section 3.2. In agreement with the simple model the implementation of a thick GaP window layer led to a three-fold increase in efficiency. By transferring the epitaxial structure on a transparent GaP substrate values of the order of 20 % could be achieved. The increase from 24 to 32 % is due to an improved device design, the double heterostructure being replaced by a SCH with MQWs as active region.

This leads to an increased internal quantum efficiency as well as an increased extraction efficiency due to a reduced re-absorption in the thinner active region.

Tab. 4 – *Chronological external quantum efficiency improvements for AlGaInP LEDs obtained in the Hewlett-Packard (now Lumileds) Labs. The η_{ext} value is calculated considering the LED immersed in an epoxy dome*

Anno	Design	η_{ext} (%)
1990	DH on GaAs substrate (standard LED)	≈ 2
1992	thick GaP window layer	≥ 6
1994	transparent GaP substrate	17,6
1996	idem	23,7
1999	idem + MQW	32,0

Figure 11 reproduces a photo of an amber DH AlGaInP-based LED with a thick window in GaP (*a*) on an absorbing GaAs substrate and (*b*) another one on a transparent GaP substrate.

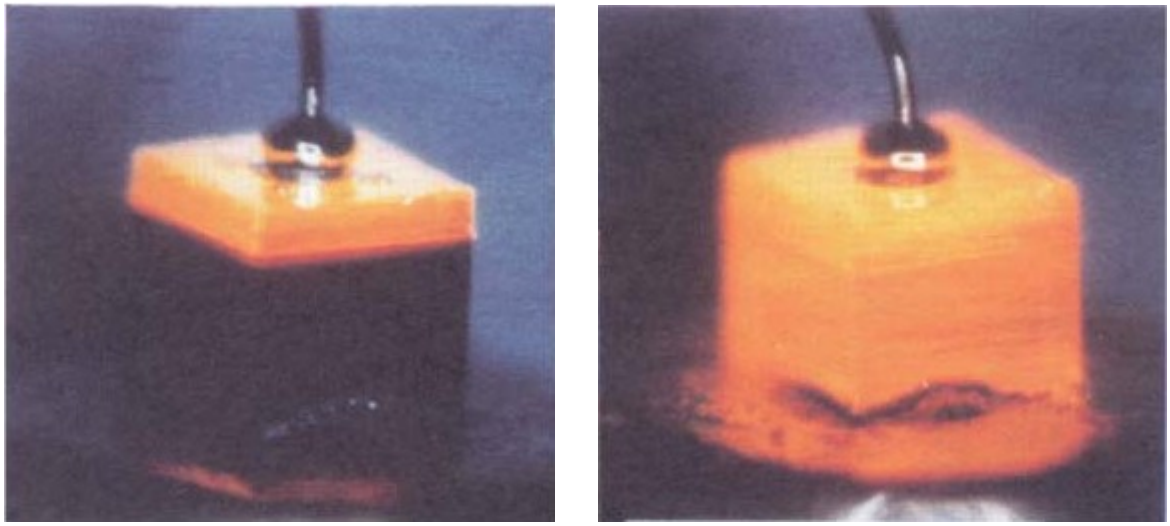


Fig. 11 – *Two different generations of amber AlGaInP LEDs by Hewlett-Packard (now Lumileds): absorbing GaAs substrate LED with thick GaP window layer (a); and wafer-bonded transparent GaP substrate (TS) LED (b)*

The values reported on Table 4 probably represents an ultimate limit to the external quantum efficiency achievable with a rectangular parallelepiped device geometry. In an ideal structure with an internal quantum efficiency of unity and with zero internal optical losses, even for a cubic geometry all the light would be eventually coupled out, after a sufficient number of internal reflection and re-absorption and re-emission steps. However in reality $\eta_{int} < 100\%$ and the internal optical losses, including the finite mirror reflectivity, re-absorption in the active region and free-carrier absorption, are non-negligible and the probability of the light being absorbed in the device increases with the number of internal reflections. For AlGaInP LEDs, a trade-off exists between active layer re-absorption and electron confinement, which results in an optimum active layer thickness. Hence in order to further optimize the extraction efficiency the number of internal reflections before extraction, corresponding to the ***mean photon path length for extraction***, needs to be reduced. This can be done either by modifying the device geometry in order to increase the light escape cones, or by altering the internal emission spectrum in order to force the emission into the existing vertical escape cones of the standard planar geometry.

As shown in section 3, the ideal geometry would be a hemisphere or a Weierstrass sphere, depending on whether a uniform or directed emission is desired. However until today the fabrication of such device geometries is difficult due to technological limitations. More feasible geometries have been realized and led to drastic increases in external quantum efficiency, such as ***truncated inverted pyramids*** (TIP) and ***radial outcoupling tapers***. The ***substrate-less thin–film***

LEDs, for which the extraction is optimized either by a surface roughening or by the implementation of pyramidal dice geometries, represent an alternative successful approach.

In parallel to these geometrical approaches another type of LEDs was developed by modifying the spontaneous emission pattern while retaining a planar geometry. By placing the active region inside an optical cavity the emission intensity in the vertical escape cones can be increased by means of interference effects. This type of light– emitting diode is hence ***called resonant cavity*** LED (RCLED) or ***microcavity LED*** (MCLED).

4.1 Truncated Inverted Pyramid (TIP)

AlGaInP LEDs are generally grown on GaAs substrates, although GaAs is absorbing in the visible wavelength range, making it impossible to couple out light efficiently trough the substrate. In the 1990s Hewlett-Packard (now Lumileds Lighting) continuously increased the extraction efficiency of AlGaInP LEDs by optimizing the device design for an optimal light outcoupling.

The implementation of a thick transparent GaP window layer and the *wafer bonding* of the epitaxial structure to a transparent GaP substrate lead to a drastic improvement in extraction efficiency. In particular, the afore mentioned “wafer bonding” is a method where the epitaxial structure is detached from the GaAs substrate by chemical etching and is then reattached to a transparent GaP substrate. The reason why the epitaxial structure of AlGaInP is not directly grown on GaP is obviously that the lattice mismatch between film and substrate would be too high and would lead to high densities of dislocations and interfacial defects.

However in these structures the extraction is still limited to the six light escape cones of a rectangular parallelepiped. It was realized already in the 1960’s that hemispherical-domed LEDs as well as truncated cones would have optimum extraction properties, however these solutions were considered as not very practical due to the high costs associated with the chip shaping.

Finally, in 1999 a young researcher of the Hewlett-Packard Labs, Micheal R. Krames, presented the realization of a practical shaped AlGaInP/GaP LED chip with a TIP geometry. The TIP geometry reduces the mean photon path length for extraction. This leads to a reduction of the internal optical losses and therefore to an increased external quantum efficiency.

TIP devices are sawn with a beveled dicing blade to provide chip sidewall angles of 35° with respect to the vertical. With red-emitting TIP LEDs ($\lambda_p \sim 650$ nm) record external quantum efficiencies of 32.6 % for emission into air and 55 % for emission into epoxy (60.9% under pulsed operation) were achieved. These devices show luminous efficiencies exceeding 100 lm/W.

A geometry used for nitride-based LEDs similar to TIP is the ***pedestal-shaped*** (GaInN on SiC) shown in figures 12.a.b, commercialized by Osram in 2001 with the trade name “Aton”. The

increase in efficiency of pedestal-shaped devices is about a factor of 2 compared to rectangular-parallelepiped-shaped devices

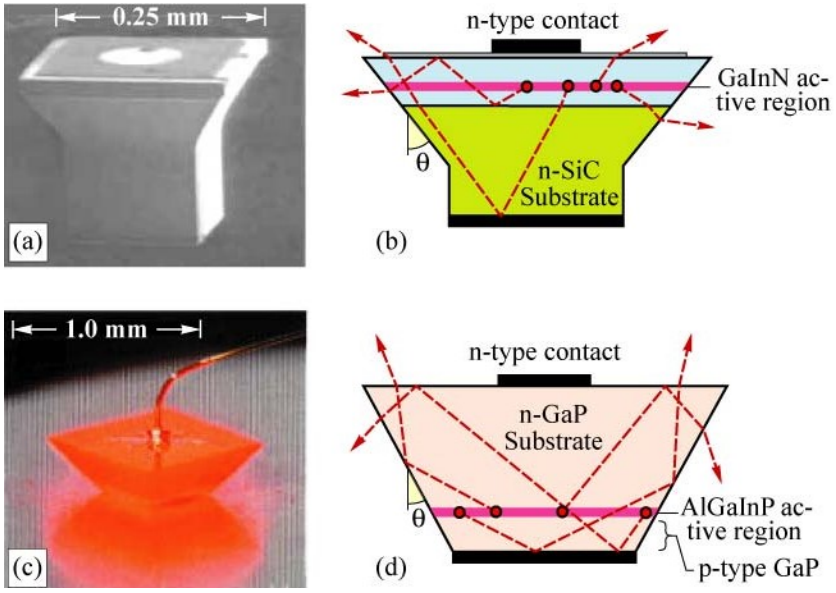


Fig. 12 – *Die-shaped devices: (a) Microphotograph of a blue pedestal-shaped GaInN emitter on SiC substrate with trade name "Aton". (b) Schematic ray traces illustrating enhanced light extraction. (c) Microphotograph of TIP AlGaInP/GaP LED. (d) Schematic ray diagram illustrating enhanced extraction*

4.2 Different types of thin-film LEDs

a) LED with reflecting mirror

By placing the epitaxial structure on a highly reflective mirror, instead of bonding it to a transparent substrate, the process technology is largely simplified; in fact, wafer bonding from semiconductor to semiconductor requires the occurrence of a series of very stringent conditions, such as ultra-flat surfaces, perfect correspondence between the crystal lattices (*lattice matching*) and high-bonding temperatures, which in this case can be avoided.

The first such structure was built in 1993 and was optically pumped; it exhibited an external quantum efficiency of 72%. The technique used was that of the ***epitaxial lift-off*** (ELO), where the multilayer epitaxial structure is detached from the substrate on which it has grown and fixed to the mirror (typically a dielectric layer coated with gold) by van der Waals surface tension forces (or possibly bonded in another way). To detach the epitaxial structure from the substrate, the ELO technique exploits the properties of the high selectivity of etching of $Al_xGa_{1-x}As$ alloys in

hydrofluoric acid (HF). When the composition of the alloy varies from x = 40% to x = 100%, the selectivity of the chemical attack is higher than 100 million to 1!

In this type of structure, the photons initially emitted outside the emission cone are directed back into the structure by internal reflection; they are also reabsorbed and partially re-emitted in the active region. This last mechanism is called *photon recycling*. However, with this approach, the external quantum efficiency is very sensitive to internal optical losses, as well as to the internal quantum efficiency, given that the average path of photons for extraction is extremely high.

b) Surface-textured LEDs

A more practical approach for the angular randomization of the totally reflected light is to scatter it from ***textured semiconductor surfaces***, making the extraction less susceptible to the material quality and internal optical losses. Experiments had shown that a surface texture on the scale of half an optical wavelength produces complete internal angular randomization of light rays in a semiconductor film. The first realization (1993) of an ***electrically pumped surface-textured thin–film LED*** – also called *non-resonant cavity LED* (NRC-LED) – led to an external quantum efficiency of approximately 30%. The surface texturing is done by depositing polystyrene spheres on the surface and using them as an etch mask for ion-beam etching, a process called ***natural lithography*** [70]. The substrate is removed by ELO and the thin film is van der Waals bonded onto large area dielectric coated Au mirrors.

Wet etching is instead used to corrugate the GaN. GaN can be easily etched by KOH and H_3PO_4 along several crystallographic planes (*a* and *m*), while it is not etched in another direction (*c*). The result is that pyramidal structures with extremely smooth surfaces can be easily created (also suitable for laser fabrication), as shown in figure 13.

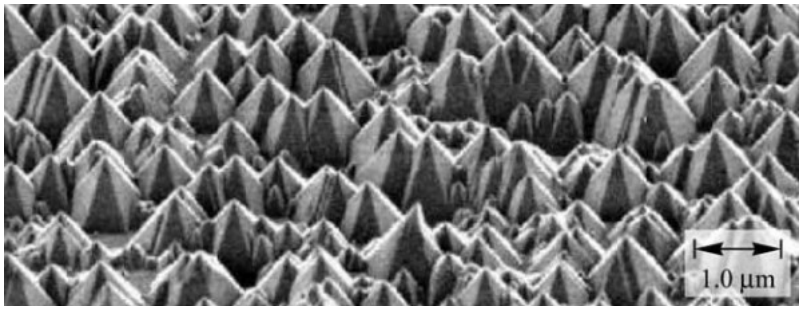


Fig. 13 – *Scanning Electron Microscope picture of a textured GaN surface*

Supplementary optimizations of the structures made it possible to further increase the external quantum efficiency; with 850 nm LEDs, external quantum efficiency values of 46% and 54% were

achieved, respectively for non-encapsulated and encapsulated devices. The structure is shown below in figure 14.

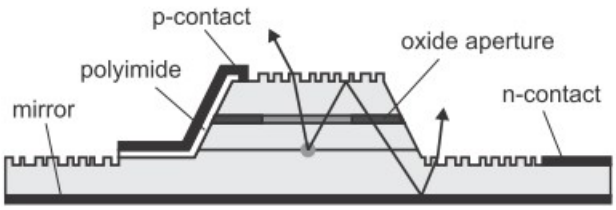


Fig. 14 – *Cross section of a textured surface LED with oxide ring for current confinement and lower metal mirror*

The LED is a ***mesa*** structure, i.e. pyramid, and uses a circular opening in an oxide layer for the ***confinement of the current*** (*oxide aperture*) in order to avoid the generation of light below the top contact (*p*-contact). In order to extract the portion of light reflected laterally and below the mesa, not only the top surface but also the area between mesa and lower contact (*n*-contact) is corrugated.

c) Buried micro-reflector LEDs (BMR LEDs)

Another approach for an efficient outcoupling of internally trapped light is the *structuring* of the backside facing the rear reflector in regular patterns of geometrical structures. The geometries that have been used to shape dies in order to enhance the light extraction, as shown previously, can in general be scaled down as well and still demonstrate their beneficial effect. This has been shown for top emitting AlGaInP LEDs with a bottom reflector on an absorbing GaAs substrate by OSRAM Opto Semiconductors. These so-called ***micro-reflectors*** can be buried at the bonded interface of a thin–film LED as well. After being covered with an insulator and a metal mirror, these structures can then be soldered to a carrier wafer with an intermediate metal layer. Current is injected through small openings in the dielectric layer of the BMRs.

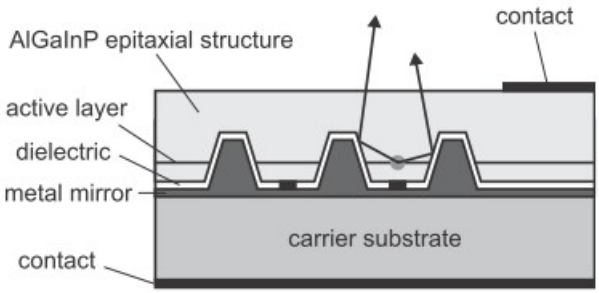


Fig. 15 – *Cross section of a BMR LED*

The performance of BMR LEDs can be further improved by corrugating the surface from which the light is extracted (***buried micro-reflector LEDs with surface texturing***), thus combining the advantages of the two structures seen previously.

d) Tapered LEDs

The concept of the tapered LED is similar to the one of the TIP LED. Light generated in a small center active region not falling in the escape cone is redirected to the bottom surface at an outer radial tapered output coupler. The taper has the shape of a shallow truncated cone and is covered with an insulator and a gold mirror. The fabrication of these devices is less cost-intensive than for the TIP LEDs, however the mean photon path length for extraction is estimated to be longer, leading to slightly lower efficiencies.

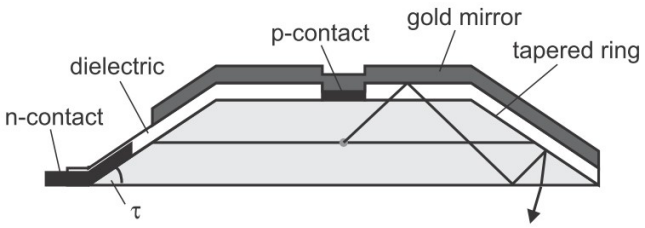


Fig. 16 – *Cross section of a tapered LED*

In order for light with any initial direction to be coupled out, the taper angle τ needs to be equal or smaller than the critical angle for total internal reflection, θ_c, and the azimuthal wave vector component is kept small by choosing appropriate device dimensions in order to avoid guided modes. The taper is fabricated by *photoresist reflow* and mask shape transfer using ion beam etching. After the contact deposition, the structure is covered with polyimide and the gold mirror is evaporated on top.

The photoresist reflow technique allows one to obtain 3-D photoresist patterns with a hemispherical cross-section. The hemispherical structure is formed simply by heating the photoresist to temperatures generally not higher than 130°C. If the photoresist has been previously deposited and developed, as soon as it heats up (beyond the glass transition temperature) it begins to melt and the walls of the structures previously obtained begin to round off. Figures 17 show the rounding of the photoresist starting from a cubic geometry (figure 17.a) and from a mesa type (figure 17.b).

The tapered structure is finally fabricated using the reflow photoresist, creating the appropriate mask shape by ion-beam etching directly on the semiconductor.

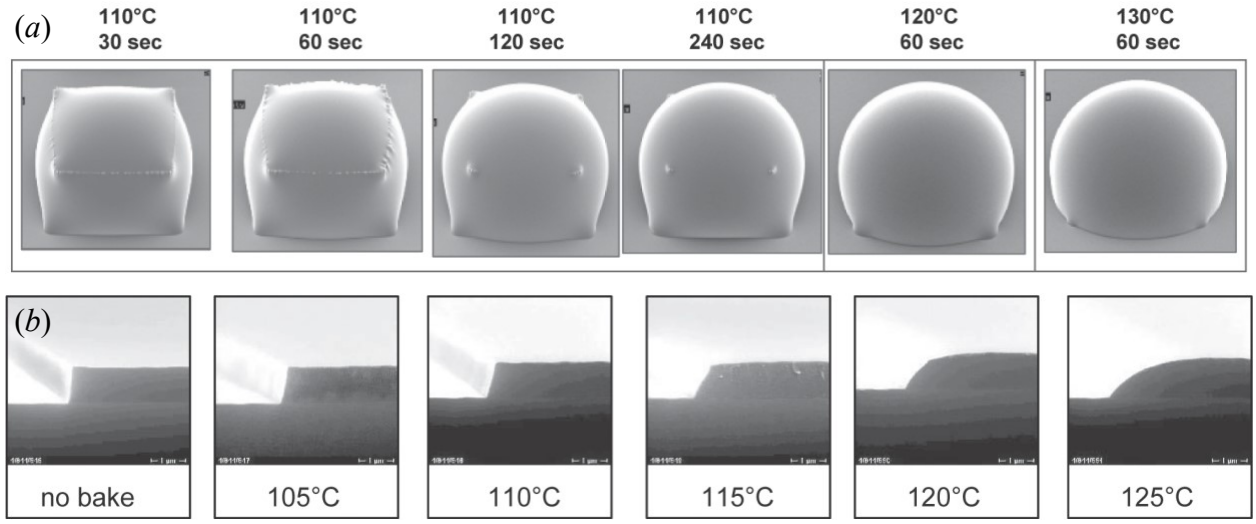


Fig. 17 – *Photoresist reflows at different temperatures of (a) a cube, (b) a mesa*