# Metal-Insulator-Semiconductor and Metal-Insulator-Metal Structures. Part V. Heterojunctions and Photonic Devices

*Alexander Gabovich*, **KPI**,

Lecture 12

# Heterojunctions

Silicon is not the only semiconductor used in the electronics industry. Beside elements from the fourth column of the periodic table and compounds thereof (Si, Ge, C, SiC and SiGe), a whole range of semiconductors can be synthesized using elements from columns III and V, such as GaAs, InP, $Ga_xAl_{1-x}As$, etc. In addition, it is also possible to fabricate semiconductors using elements from other columns of the periodic table, such as CdS and HgCdTe.

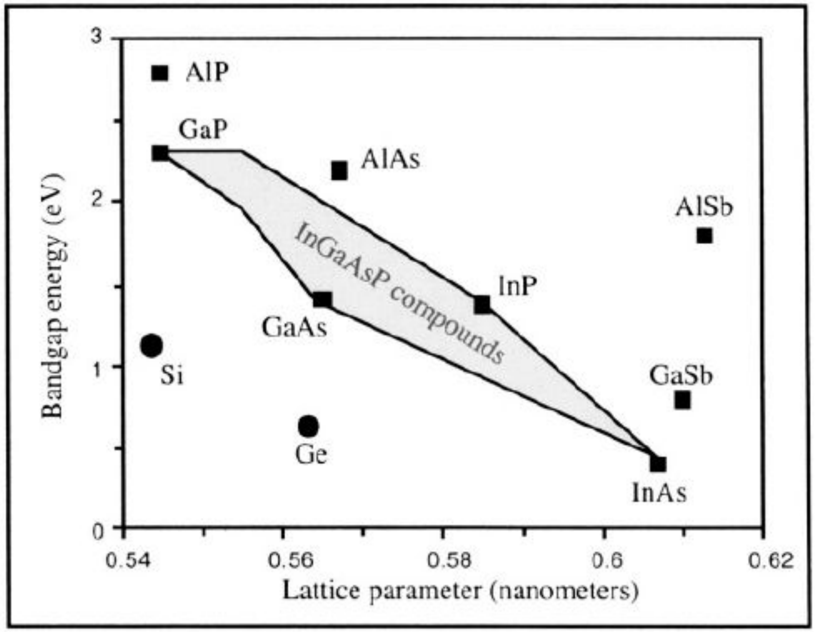


**Figure 9.1**: Energy bandgap of Si, Ge, and III-V compound semiconductors.[1]

# Heterojunctions

The main parameter characterizing the electrical properties of these materials is the width of the bandgap. Figure 9.1 shows the bandgap energy for silicon, germanium, and different III-V compounds. Arbitrary values of the bandgap energy can be obtained using ternary or quaternary compounds, such as $Ga_xAl_{1-x}As$ and $Ga_xIn_{1-x}As_yP_{1-y}$. The desired bandgap energy can be reached by adjusting the x and y coefficients during the fabrication of the material.

A PN junction that encompasses two different semiconductors is called a *heterojunction.* The most distinctive feature of such junctions is that the P and the N region have different energy band gaps. A junction containing only one semiconductor, such as a classical silicon PN junction, is called a *homojunction.*

The presence of two materials with different bandgap energies introduces an additional level of difficulty in the energy band diagram of heterojunctions, when compared to homojunctions. Combining different semiconductor materials within a single device and the art of tailoring the shape of energy bands to achieve properties which could otherwise not be attained is often referred to as "bandgap engineering".

# Heterojunctions. Energy band diagram

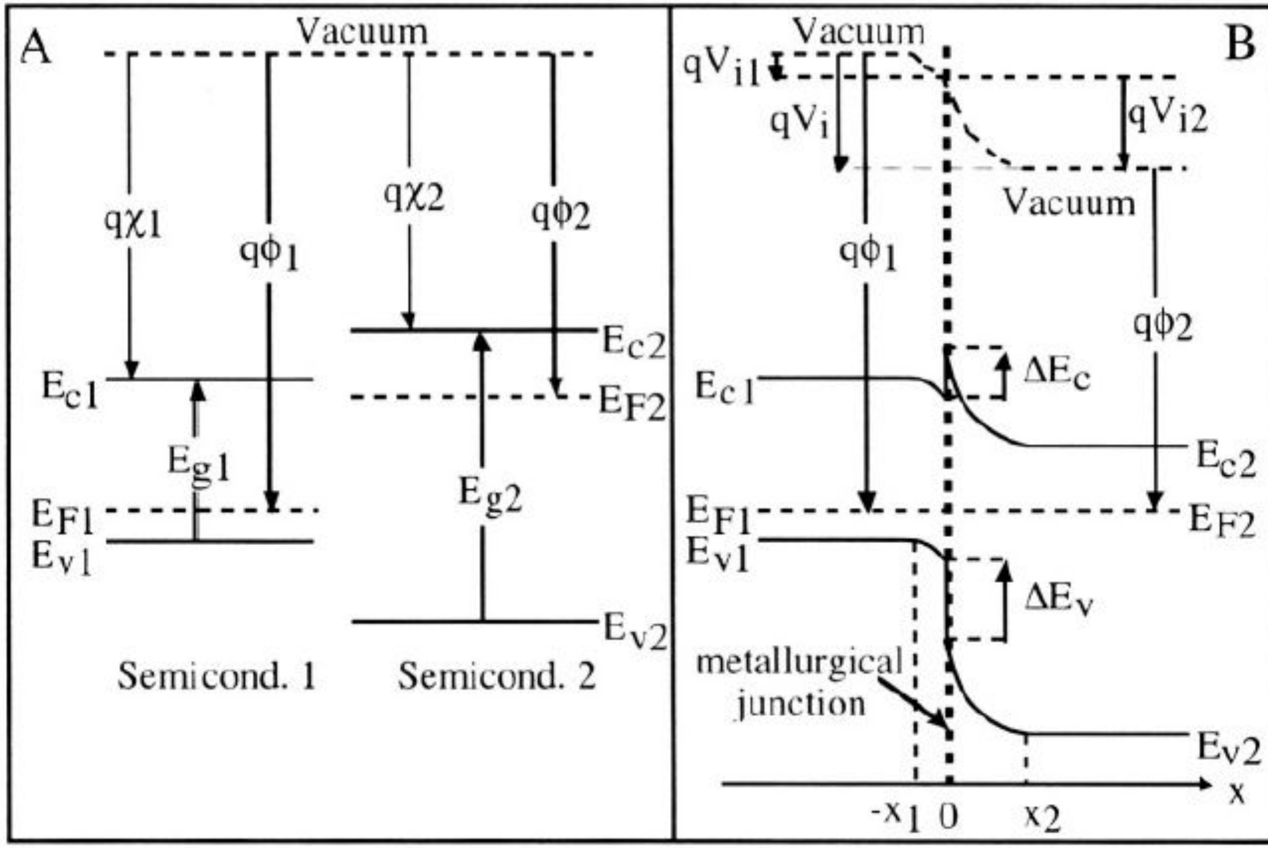


**Figure 9.2**: A: Energy band diagram of the two semiconductors taken separately; B: Energy band diagram of the materials connected (heterojunction).[2,3]

# Heterojunctions. Energy band diagram

Consider the example of Figure 9.2 which illustrates how the energy band diagram of a heterojunction can be drawn. Two different semiconductor materials are combined. Let Semiconductor #1 be P-type and have an energy band gap, a work function, and an electron affinity equal to $E_{g1}$, $q\phi_1$, and $q\chi_1$, respectively. The work function is the energy difference between the vacuum level and the Fermi level; it represents the energy required to remove an electron of energy $E_F$ from the semiconductor. The electron affinity is the energy needed to remove an electron in the conduction band to the vacuum level, as previously explained in Section 5.1.1. Similarly, we will suppose Semiconductor #2 is N-type, and its energy band gap, work function, and electron affinity are $E_{g2}$, $q\phi_2$, and $q\chi_2$, respectively.

# Heterojunctions. Energy band diagram

The procedure for drawing the energy band diagram is the following:

1- Under equilibrium conditions the Fermi level in the two semiconductors is equal and constant. Far from the junction the semiconductor materials will be neutral and their energy band diagram will be the same as when the two materials are taken separately.

2- The work functions $q\phi_1$ and $q\phi_2$ remain unchanged in the neutral zones. This enables us to draw the vacuum levels, far from the junction.

3- The vacuum levels of the two semiconductors are connected by a smooth, continuous curve. The exact shape of the curve is at present unknown and will be calculated later. It is, however, a good idea to assume that it will have a shape similar to the band bending in the transition region of a homojunction. The vacuum level bends only within the transition region, thus between $-x_1$ and $x_2$.

4- During the junction formation electrons will diffuse from the N-type semiconductor into the P-type material since $q\phi_1 > q\phi_2$, and holes will diffuse in the opposite direction from the N-type into the P-type semiconductor. The resulting charge distribution gives rise to a depletion region, an internal junction potential, and therefore, to a curvature of the energy bands. This curvature is parallel to that of the vacuum level. Knowing that the electron affinities, $q\chi_1$ and $q\chi_2$ remain constant in the transition region enables us to draw $E_{V1}$, $E_{V2}$, $E_{C1}$ and $E_{C2}$ in the transition region.

5- Finally the valence ($E_{V1}$ and $E_{V2}$) and conduction ($E_{C1}$ and $E_{C2}$) levels are connected using vertical line segments, at the metallurgical junction ($x = 0$). This feature constitutes what is called a "band discontinuity".

# Heterojunctions. Energy band diagram

The junction potential, $V_i$, is given by:

$$V_i = V_{i1} + V_{i2} = \phi_1 - \phi_2 \quad (9.1.1)$$

where $qV_{i1}$ and $qV_{i2}$ is the band curvature in semiconductors 1 and 2, respectively.

Since both $E_{C1}$ and $E_{C2}$ are parallel to the vacuum level there will be a discontinuity of the energy bands at the metallurgical junction. The discontinuity is equal to:

$$\Delta E_C = q(\chi_1 - \chi_2) \quad (9.1.2)$$

and

$$\Delta E_v = (E_{g2} - E_{g1}) - q(\chi_1 - \chi_2) \quad (9.1.3)$$

The sum of the two band discontinuities is equal to the bandgap difference between the two semiconductors:

$$\Delta E_C + \Delta E_v = E_{g2} - E_{g1} \quad (9.1.4)$$

The exact curvature of the energy bands within the transition region can be obtained by solving Poisson's equation in both semiconductor materials and using the depletion approximation.

# Heterojunctions. Charge distribution

| **Semiconductor #1** | **Semiconductor #2** |
|---|---|
| P-type, doping concentration = $N_a$ | N-type, doping concentration = $N_d$ |
| permittivity = $\varepsilon_1$ | permittivity = $\varepsilon_2$ |
| width of depletion region = $-x_1$ | width of depletion region = $x_2$ |

Poisson's equation is integrated to calculate the electric field:

$$\frac{d^2\phi}{dx^2} = -\frac{\rho}{\varepsilon_1} = \frac{qN_a}{\varepsilon_1} \qquad \frac{d^2\phi}{dx^2} = -\frac{\rho}{\varepsilon_2} = -\frac{qN_d}{\varepsilon_2}$$

$$\mathcal{E}_1 = -\frac{d\phi}{dx} = -\frac{qN_a}{\varepsilon_1}(x+x_1) \qquad \mathcal{E}_2 = -\frac{d\phi}{dx} = -\frac{qN_d}{\varepsilon_2}(x_2-x)$$

Using Gauss' theorem at the metallurgical junction ($x=0$) yields:

$$\mathcal{E}_1\,\varepsilon_1 = \mathcal{E}_2\,\varepsilon_2 \;\Rightarrow\; N_a\,x_1 = N_d\,x_2 \qquad (9.1.5)$$

which expresses charge neutrality in the transition region ($-x_1 \le x \le x_2$).

A second integration of Poisson's equation yields the potential:

$$\phi = A + \frac{qN_a}{2\varepsilon_1}(x+x_1)^2 \qquad \phi = B - \frac{qN_d}{2\varepsilon_2}(x_2-x)^2$$

and the band curvature:

$$\phi(x=0) - \phi(-x_1) = \frac{qN_a}{2\varepsilon_1}x_1^2 \qquad \phi(x_2) - \phi(x=0) = \frac{qN_d}{2\varepsilon_2}x_2^2$$

The sum of the two latter equations is equal to the junction potential, $V_i$:

$$\phi(x_2) - \phi(-x_1) = \frac{qN_a}{2\varepsilon_1}x_1^2 + \frac{qN_d}{2\varepsilon_2}x_2^2 = V_i = V_{i1} + V_{i2} = \phi_1 - \phi_2 \qquad (9.1.6)$$

# Heterojunctions. Charge distribution

Eliminating $x_2$ between 9.1.5 and 9.1.6, we obtain the built-in potential in semiconductor:

$$\frac{qN_a}{2\varepsilon_1} x_1{}^2 + \frac{qN_d}{2\varepsilon_2} \frac{N_a^2}{N_d^2} x_1^2 = V_i \qquad (9.1.7)$$

from which the depletion width in semiconductor #1 can be extracted:

$$x_1 = \sqrt{\frac{2\varepsilon_1\varepsilon_2 N_d V_i}{q N_a (\varepsilon_2 N_d + \varepsilon_1 N_a)}} \qquad (9.1.8a)$$

Using 9.1.5 and 9.1.8a we find the depletion width in semiconductor #2:

$$x_2 = \sqrt{\frac{2\varepsilon_1\varepsilon_2 N_a V_i}{q N_d (\varepsilon_2 N_d + \varepsilon_1 N_a)}} \qquad (9.1.8b)$$

Knowing $x_1, x_2$, and $\phi(x)$ the energy band curvature can now be drawn with accuracy.

# Heterojunctions. Charge distribution. Currents

When an external bias, $V_a$, is applied to the diode, the electron and hole diffusion current densities injected respectively at $x = -x_1$ in the P-type and at $x = x_2$ in the N-type material are given by Equations 4.4.23 (where $l_n = x_2$) and 4.4.24 (where $-l_p = -x_1$) which, in the case of a heterojunction, becomes:

$$J_n(-x_1) = \frac{qD_n n_{i1}^2}{N_a L_n}\left[exp\left(\frac{qV_a}{kT}\right) - 1\right] \text{ and } J_p(x_2) = \frac{qD_p n_{i2}^2}{N_d L_p}\left[exp\left(\frac{qV_a}{kT}\right) - 1\right]$$

where $n_{i1}^2$ and $n_{i2}^2$ are the intrinsic carrier concentrations in semiconductor #1 and #2, respectively. The influence of the heterojunction on the diffusion currents is best illustrated by taking the $J_n/J_p$ ratio at the edges of the depletion region:

$$\frac{J_n}{J_p} = \frac{D_n L_p N_d n_{i1}^2}{D_p L_n N_a n_{i2}^2} = \frac{D_n L_p N_d N_{v1} N_{c1} exp(-E_{g1}/kT)}{D_p L_n N_a N_{v2} N_{c2} exp(-E_{g2}/kT)}$$

$$= \frac{D_n L_p N_d \,(m_{n1}^* m_{p1}^*)^{3/2}}{D_p L_n N_a (m_{n2}^* m_{p2}^*)^{3/2}}\, exp\left(\frac{E_{g2}-E_{g1}}{kT}\right) \quad (9.1.9)$$

where $N_{v1}$, $N_{c1}$, $m_{n1}^*$, $m_{p1}^*$, $N_{v2}$, $N_{c2}$, $m_{n2}^*$, $m_{p2}^*$, are the effective density of states in the valence and conduction band, and the effective electron and hole masses in semiconductor #1 and #2, respectively. An important conclusion can be drawn from Equation 9.1.9: the ratio of electron to hole current in the PN heterojunction is exponentially proportional to the difference of energy bandgaps between the two semiconductors.

# Photonic devices. Light-emitting diode

When a recombination event takes place in a direct-bandgap semiconductor a photon can be emitted. This phenomenon is called "radiative recombination". The wavelength of this photon depends on the bandgap energy of the semiconductor according to the relationship: $E_g = h\nu$. Radiative recombination is observed in many semiconductor materials such as SiC, GaN, GaAsP, AlInGaP and AlGaAs. Furthermore, the bandgap energy in semiconductor compounds can be tailored to produce devices capable of emitting photons with a specific desired color. There is a whole variety of solid-state devices that can emit and collect photons, but we will only focus here on the laser diode. However, it is necessary to understand the operation of the light-emitting diode before beginning the study of the laser diode.

The **L**ight-**E**mitting **D**iode, or LED, is a simple PN junction made in a semiconductor material which exhibits radiative recombination properties. This PN junction can either be a heterojunction or a homojunction. The energy bandgap of the semiconductor material determines the frequency of the emitted light, according to the relationship $E_g = h\nu$. Some examples of semiconductor materials used for the fabrication of LEDs, and the color of the emitted light are: GaN

(blue), SiC (blue), GaP (green), $GaAs_{0.14}P_{0.86}$ (yellow), $GaAs_{0.35}P_{0.65}$ (orange), $GaAs_{0.6}P_{0.4}$ (red), and GaAs (infrared). In this section we will focus on the operation of a homojunction (single material) LED.

# Photonic devices. Light-emitting diode

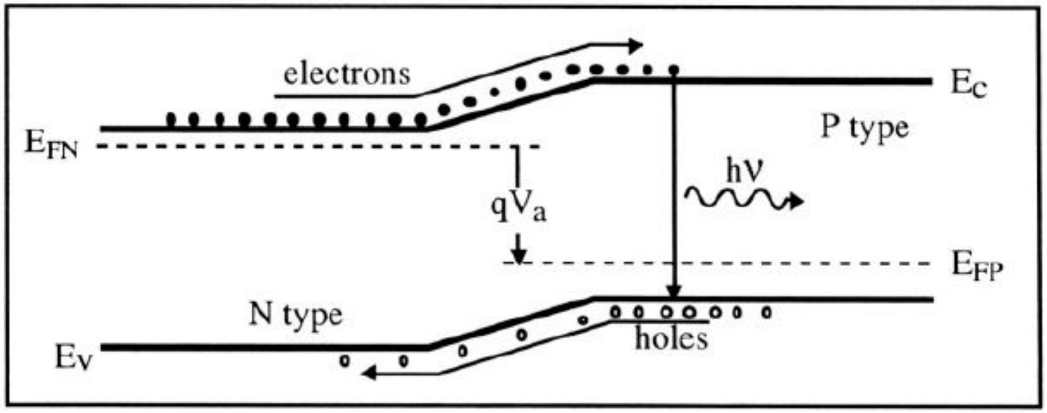


**Figure 9.7**: Operation of a light-emitting diode. [7]

Figure 9.7 illustrates a homojunction (single bandgap) **PN**$^+$ junction in the forward bias mode. Light emission is produced by the radiative recombination of electrons injected into the P-type material. Because $N_d >> N_a$ the electron current is much larger than the hole current. Electrons are injected when the PN junction is forward biased by $V_a$. The injection efficiency $\eta_i$ relates the current of "useful" carriers (the electrons injected in the P-type region) to the total current in the junction:

$$\eta_i = \frac{I_n}{I_n + I_p + I_{rec}}$$

where $I_n$ is the electron current injected into the P-type region, $I_p$ is the hole current injected in the N-type region, and $I_{rec}$ is the current of carriers recombining in a non-radiative way.

# Photonic devices. Light-emitting diode

Usually, $\eta_i$ reaches values of 30 to 60%. As mentioned in Section 3.2 radiative recombination must satisfy conservation of momentum. This criterion is automatically met in direct-bandgap semiconductors since the momentum of electrons at the conduction band minimum is equal to that of holes at the valence band maximum. Light emission is, however, observed in indirect-bandgap semiconductors such as GaP and SiC. The only way radiative recombination can take place in these semiconductors is for the interaction to produce a particle, or something capable of acting like a particle, that can dissipate the initial electron momentum. Fortunately, an appropriate "particle" exists which is a quantum of vibrational energy in the crystal lattice, called a phonon. Phonons produce heat transfer to (or from) the lattice, which acts to reduce electron momentum and thereby enables radiative recombination. The interaction of concern is one in which an electron in the conduction band recombines with a hole

in the valence band, and produces both a photon and a phonon. The combined energy of the photon and the phonon is equal to $E_g$ and the sum of the initial electron momentum and the momentum of the phonon equals zero. This process is much more complex, and therefore, more unlikely to happen than radiative recombination as in direct bandgap semiconductors. As a result, the performance (in terms of brightness) of indirect bandgap LEDs is much lower than that of direct bandgap materials. The luminous intensity of indirect bandgap devices has, however, been substantially increased using the following "trick". The approach is to add an isoelectronic impurity, *i.e.* an impurity from the same column of the periodic table as the element it replaces. An example is nitrogen in GaP, designated GaP:N. Each nitrogen atom creates a localized strain in the crystal that can trap an electron. The electrons are bound so tightly to those traps that there is little uncertainty as to their position. But there is, according to the Heisenberg uncertainty principle, a large statistical uncertainty in their momentum. The uncertainty is large enough for each electron to have a significant probability of having zero momentum and undergoing radiative recombination. This quantum-mechanical "trick" raises the radiative recombination rate, but to date, not enough to rival the rate in direct bandgap semiconductors. [[8]]

# Photonic devices. Laser diode

The laser diode is a PN junction which can emit a laser beam. Laser light is coherent (*i.e.* the emitted photons are in all phase) and monochromatic (*i.e.* the emitted photons all have the same wavelength). Describing in detail how a laser works is beyond the scope of this book. It is, however, necessary to briefly describe the conditions required for a lasing effect to take place. The word "laser" means "**L**ight **A**mplification by **S**timulated **E**mission of **R**adiation". The key word in this definition is "stimulated emission". Stimulated emission is a phenomenon falling into the same category as generation and recombination, but in which an incident photon with an energy $E_g = h\nu$ triggers the recombination of an excited electron (an electron in the conduction band, in the case of a semiconductor laser). During the recombination event a new photon is emitted. This photon has the same wavelength as the incident photon and is in phase with it. This is why laser light is monochromatic (all photons have the same wavelength, fixed by the energy bandgap) and coherent (all photons have the same phase). This photon generation can of course be repeated, and the original photon can be amplified by 2, 4, 8, etc. as shown in Figure 9.8, resulting in a light amplification effect. If two parallel mirrors (which can reflect light) are placed at both sides of the semiconductor crystal light can travel back and forth inside the crystal and undergo significant amplification. Such a structure constitutes a Fabry-Pérot cavity. In practice, one of the mirrors is semitransparent, such that some of the laser light can escape from the crystal. Emitted photons which do not travel perpendicular to the mirrors exit the semiconductor and are lost (Figure 9.8).

# Photonic devices. Laser diode

A photon with energy $h\nu = E_g$ can not only stimulate the emission of another photon, but it can be absorbed by the semiconductor material and generate an electron-hole pair. This effect is highly undesirable in a laser diode since we do not want to see photons absorbed. Unfortunately, photon absorption is unavoidable. It is, however, possible to favor stimulated emission with respect to absorption. This can be achieved if the number of electrons in the excited state (*i.e.*, in the conduction band) is larger than the number of electrons in the ground state (*i.e.*, in the valence band). This condition is called "population inversion". It can be realized if an external source of energy "pumps" a large quantity of electrons from the fundamental state into the excited state. In a laser diode population inversion is obtained by injecting a large amount of electrons into a PN junction.

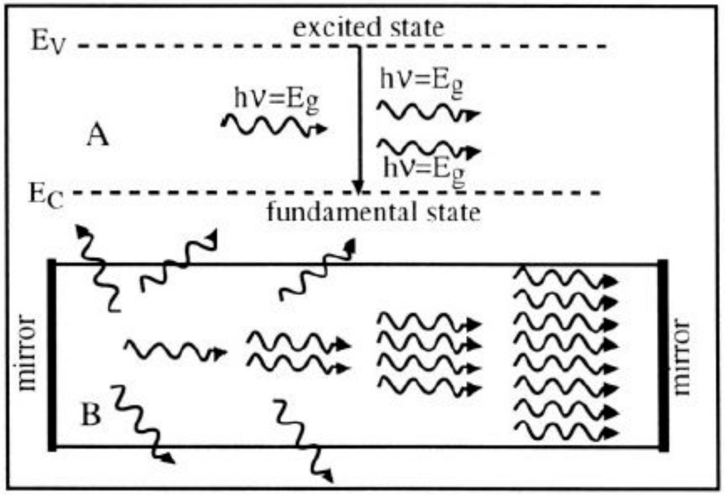


**Figure 9.8**: A: Principle of stimulated emission, B: Light amplification by stimulated emission.

# Photonic devices. Laser diode

Figure 9.9 shows a laser PN homojunction. The $\mathbf{N}^{+}$ and $\mathbf{P}^{+}$-type regions are degenerately doped and the Fermi level in the $\mathbf{N}^{+}$ and $\mathbf{P}^{+}$-type material is above the conduction band minimum and below the valence band maximum, respectively. When a forward bias is applied to the junction a thin region is formed which, instead of being depleted, is in population inversion. In that region there is a strong electron population in the conduction band and a high density of empty states (or holes) in the valence band. Under these conditions laser light is emitted through stimulated emission within the transition region.

A complete laser diode is presented in Figure 9.10. Two semi-transparent, parallel mirrors are obtained by cleaving the semiconductor along a natural crystal direction (*e.g.* (100)). Since the refractive index of the semiconductor material is larger than that of the surrounding air, the cleaved surfaces act as mirrors which reflect the light back into the crystal. These mirrors do not have a 100% reflectivity, however, which allows some of the laser light to be emitted from the device.

# Photonic devices. Laser diode

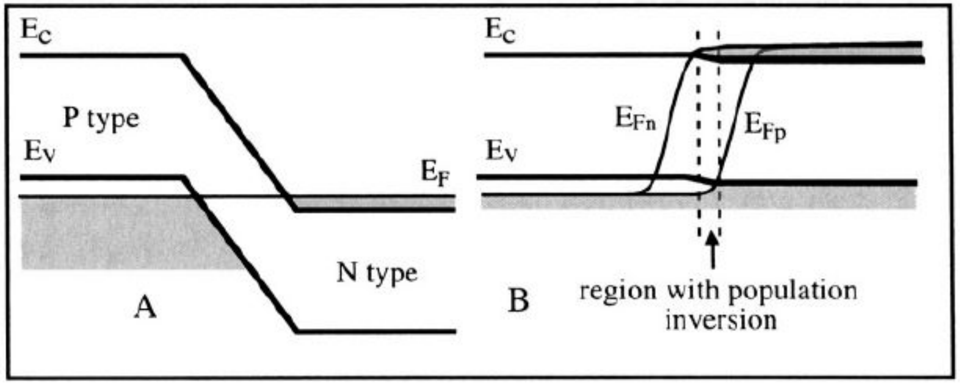


**Figure 9.9**: Laser PN junction. A: at equilibrium, B: under forward bias. [9]

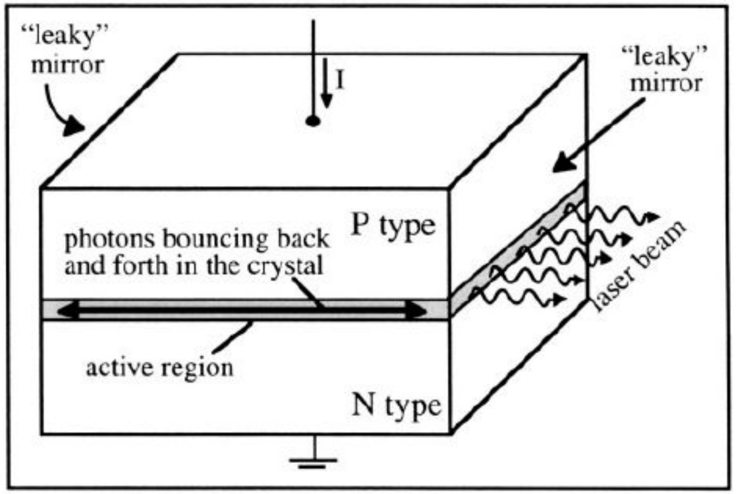


**Figure 9.10**: Laser diode in forward bias mode showing photon path within the transition region with population inversion.

# Photonic devices. Laser diode

The output light power of a laser diode is presented in Figure 9.11 as a function of the current injected into the diode. Below a given threshold, population inversion is not reached, however light is emitted because of radiative recombination. This light is incoherent and is similar to the light emitted by a LED. Above this threshold, population inversion takes place, and laser light is emitted. The light intensity then increases sharply as a function of the current in the diode. Because of the Fabry-Pérot cavity the spectrum of emitted light is compressed into one single

spectral line. The emitted laser light is, therefore, monochromatic. Beside the "useful recombination" of electrons by stimulated emission in the population inversion region, a large quantity of electrons are injected into the P-type semiconductor where they can also recombine and emit either photons which do not take part in the lasing process, or phonons (heat). This renders homojunction laser diodes quite inefficient, and only a fraction of the electrical power supplied to the device is converted into laser light.

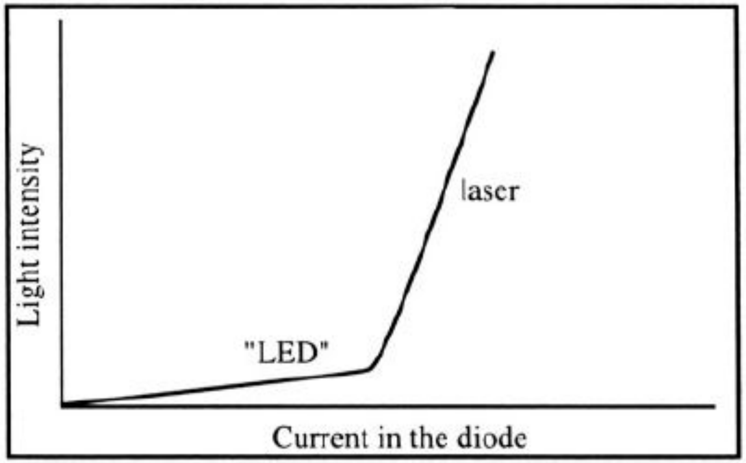


**Figure 9.11**: Emission of incoherent light (LED) and coherent light (laser) as a function of the current intensity. [10]

# Photonic devices. Laser diode

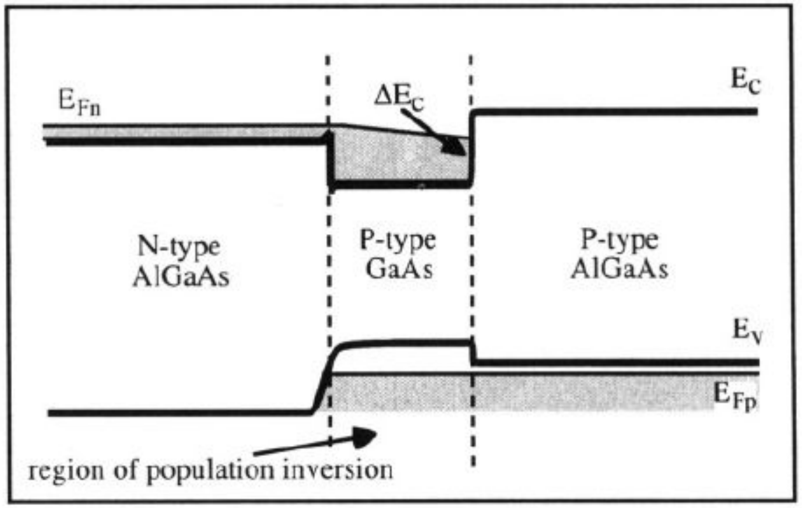


**Figure 9.12**: AlGaAs/GaAs/AlGaAs heterojunction laser diode.[11]

This problem can be solved by the use of a heterojunction structure. Let us take the example of the AlGaAs/GaAs/AlGaAs heterojunction laser diode shown in Figure 9.12: the electrons in the conduction band which are injected from a forward bias from the N-type AlGaAs into the P-type GaAs cannot spill over the potential barrier $\Delta E_C$ created by the P-type GaAs / P-type AlGaAs junction. These electrons are thus confined in the GaAs layer where the inversion population, and thus laser light emission, is produced. In addition, the refractive index of AlGaAs is lower than that

of GaAs, which causes the junctions to act as mirrors. This helps confine the photons in the GaAs layer and limits the leakage of light into the AlGaAs layers. As a result the laser light emission efficiency is greatly enhanced and the current threshold for laser light emission is reduced.