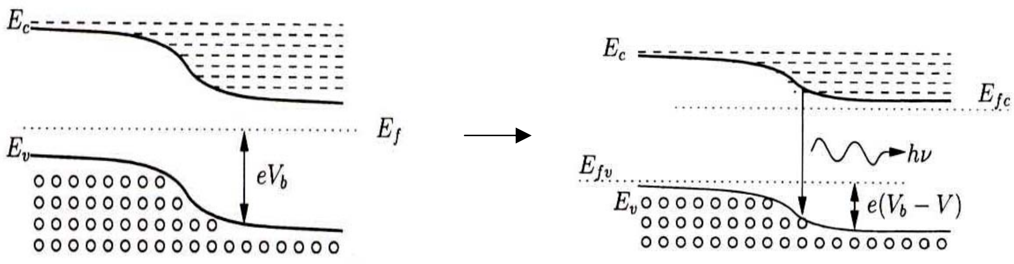
Semiconductor Devices: Lectures 3&4

7. Light Emitting Diodes

There are two distinct optical emission processes in semiconductors: spontaneous emission and stimulated emission. The former is used in light emitting diodes, the latter gives rise to coherent amplification which is the basis of laser action.

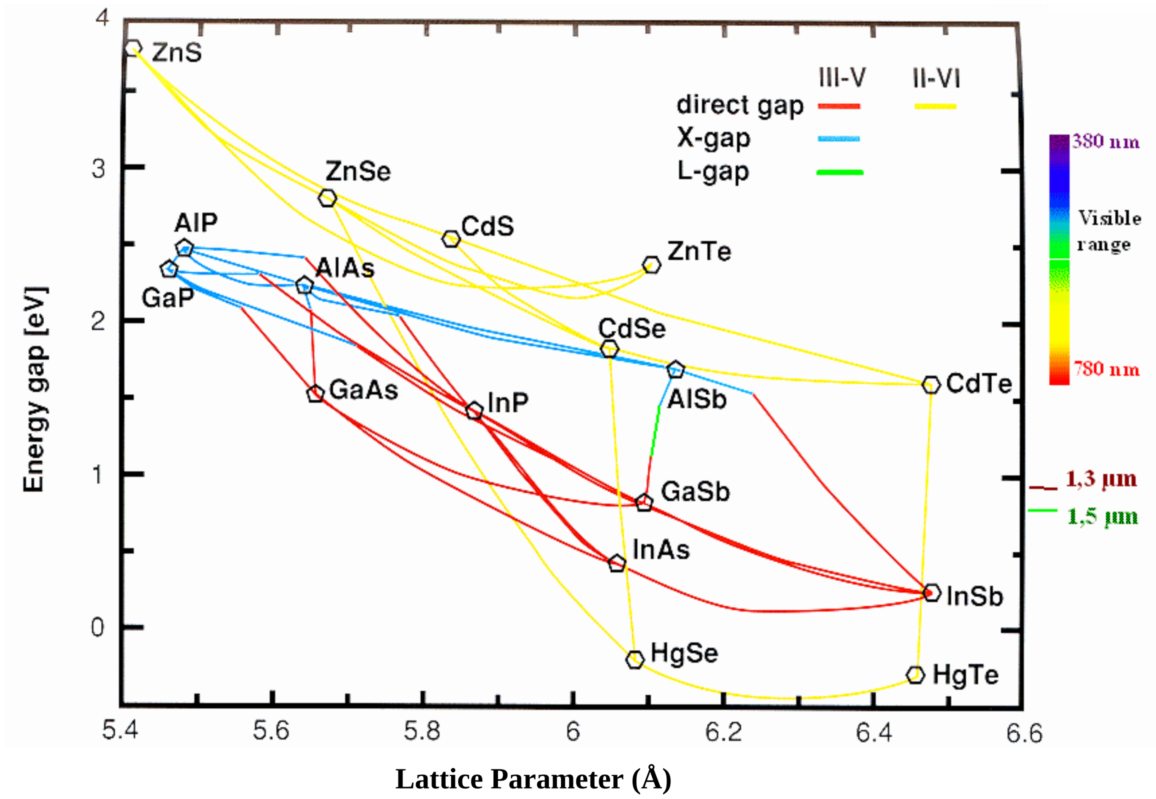


An LED is "simply" a moderately doped *p-n* junction: under forward bias the band offset decreases and the Fermi levels separate out. Holes from the *p*-region and electrons from the *n*-region are swept into the depletion region where they recombine radiatively, which constitutes a forward current through the junction. Note that the quasi Fermi level separation is less than the bandgap energy. The optical power generated by an LED with a current $\boldsymbol{i}$ is :

$$P_{opt} = \eta_i \frac{i}{q} \frac{hc}{\lambda}$$

where η_i is the internal quantum efficiency; however, as we will see below, not all of this power exits the device, and the external efficiency $\eta_e << \eta_i$.

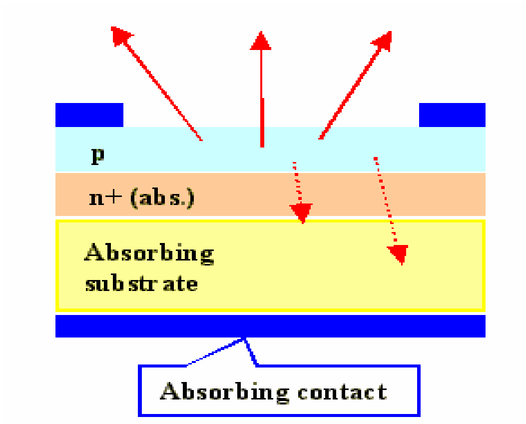
The wavelength of the light emitted depends on the bandgap. Many III-V and II-VI semiconductors and their alloys can be used to provide emission over the range from ~350nm to the mid-infrared. Note that direct gap semiconductors are most commonly used – efficient radiative recombination requires “vertical transitions”. Indirect materials can also be used, but they are normally heavily doped with isoelectronic impurities (e.g. GaP:N)

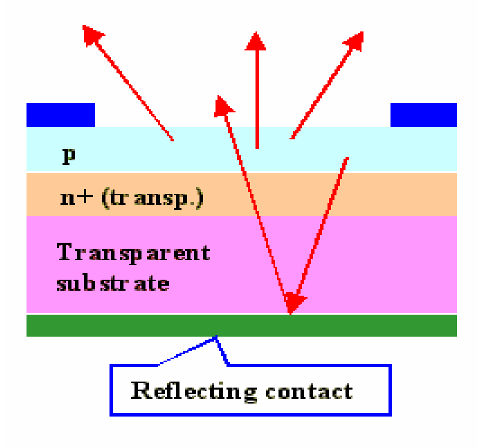


LEDs comes in many different forms depending on the application, e.g. for displays or optical communications:

(i) Surface emitting LEDs

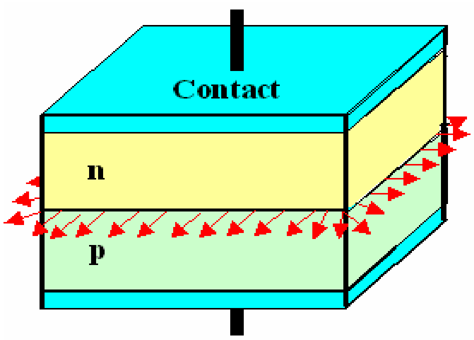
In the surface emitting LED electrons are injected into the top p-layer, and only the photons that manage to escape will be seen. The top layer therefore needs to be as thin as possible. A somewhat better device uses the light emitted to the backside by reflecting it back to the front side. If the light has below-bandgap energy because it arises from recombination at impurities (eg GaP:N), there is no absorption in the substrate, and so it pays to make the back contact reflective and keep the layers thin.

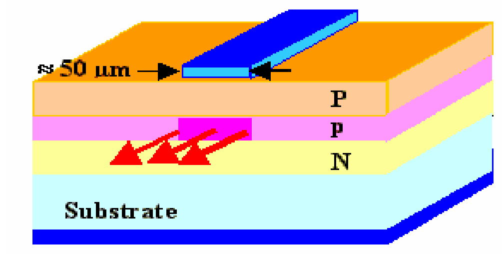


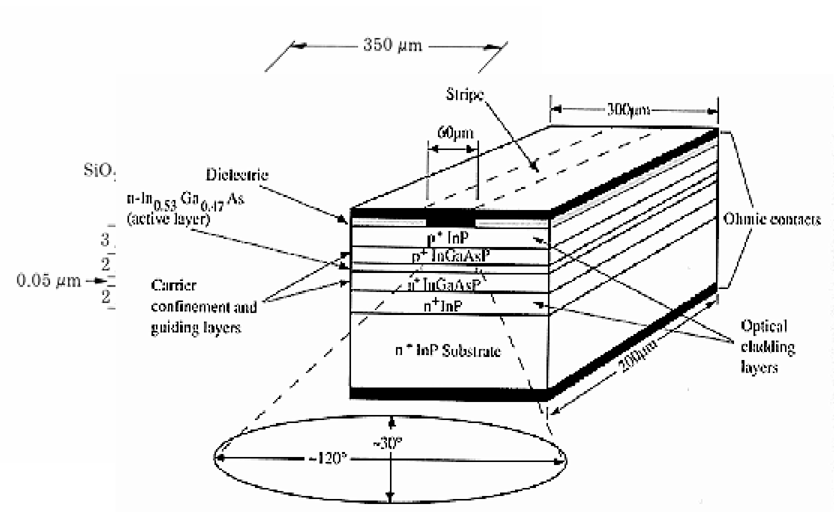


(ii) Edge emitting LEDs

If high intensities are required. i.e. not just a lot of photons but a lot of photons per unit area, the light emission must be confined to a small area where a high injection ratio can be achieved. This is particularly important if the emitted light is to be coupled to an optical fiber or semiconductor waveguide for optical communications purposes. This can be done with an edge-emitting LED. The use of a stripe injector concentrates the current.



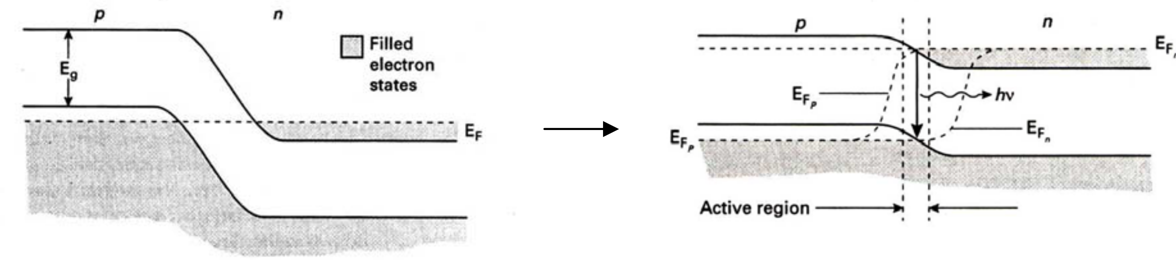




The figure show the structure of a conventional InGaAs LED used for optical communications. Heterojunctions provide an added degree of optical confinement which is crucial for semiconductor diode lasers as we will see later.

8. Semiconductor Laser Diodes

Now consider the behaviour of a heavily doped, forward biased *p-n* diode. The doping is typically $>10^{24}m^{-3}$ and the Fermi levels lie in the valence and conduction bands for the *p-* and *n*-regions respectively. Under forward bias there are very high electron and hole populations in the depletion region; radiative recombination occurs as in the LED, but stimulated emission now dominates.



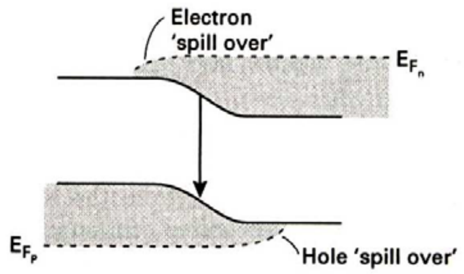
The condition for amplification is that emission is greater than absorption. Since the probabilities for both processes are proportional to the product of availability of initial and final states:

$$P_A \propto \left(1-\frac{1}{e^{(E_p-E_{F_p})/k_BT}+1}\right)\left(1-\frac{1}{e^{(E_n-E_{F_n})/k_BT}+1}\right)$$
$$P_E \propto \left(\frac{1}{e^{(E_n-E_{F_n})/k_BT}+1}\right)\left(\frac{1}{e^{(E_p-E_{F_p})/k_BT}+1}\right)$$

The condition $P_E > P_A$ simplifies to $E_n+E_p = h\nu < E_{F_n}+E_{F_p}$, and since $h\nu > E_G$:

$$E_{F_n}+E_{F_p} > E_G$$

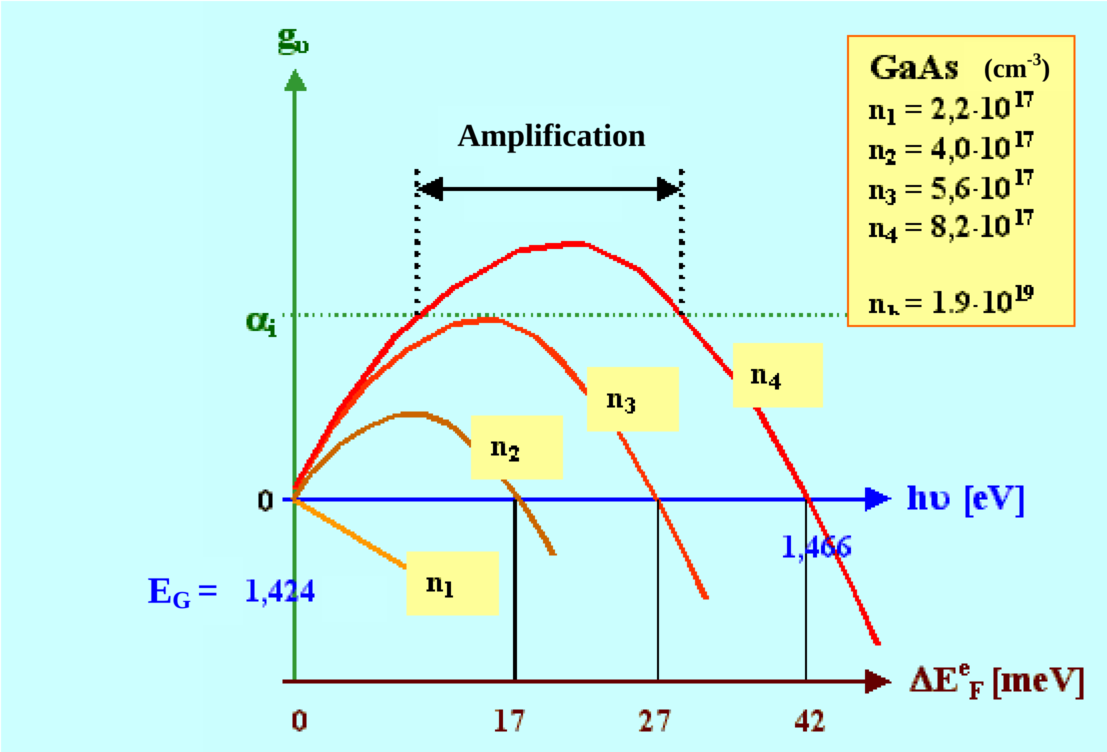
This condition is analogous to the requirement for *population inversion* in an atomic system.



Increasing the forward bias does not necessarily increase the gain indefinitely; a point is reached where "spill over" occurs - carriers diffuse through into the bulk material and are lost from the active region. Confining carriers (and photons) to the active region is a major consideration in semiconductor laser design.

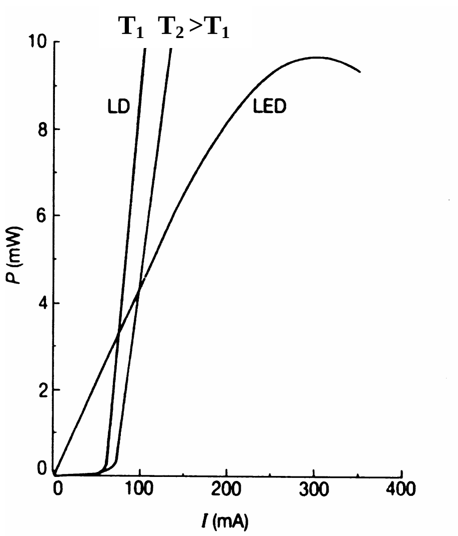
Optical Gain

The rate of stimulated emission depends on the intensity of the incident light, and so the net intensity increases exponentially with distance: $I_\nu(z) = I_0 exp(g_\nu z)$, where g_ν is the gain coefficient. Light is also absorbed, so that the overall result is $I_\nu(z) = I_0 exp((g_\nu - \alpha)z)$. Light amplification occurs when $g_\nu = \alpha$. The results presented below show the variation of the gain bandwidth with electron density assuming a fixed hole density of $1.9\times10^{19}cm^{-3}$. Amplification occurs when the gain increases above the intrinsic loss which includes all nonradiative recombination processes.



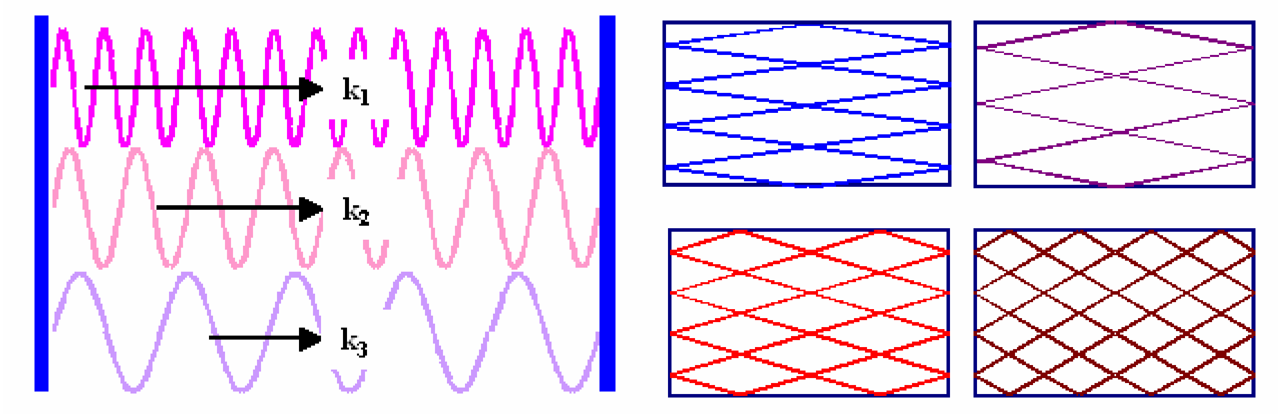
Lasing Threshold

The laser output power is zero until the gain equals the losses, and then it increases rapidly with increasing current. If the temperature of the diode increases, nonradiative recombination increases, and so the threshold current rises.
This behaviour contrasts with that of an LED, where there is no threshold behaviour: the output power increases almost linearly with current, but it eventually saturates.

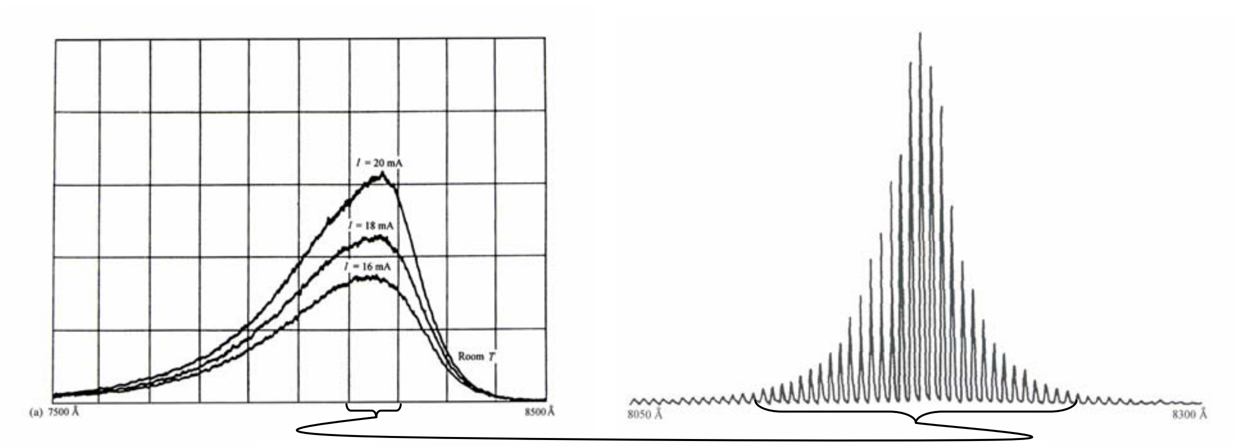


Laser Cavity

The laser cavity is formed by cleaving the semiconductor crystal. It is essentially a Fabry-Perot interferometer; the refractive index n_r is ~3.6 which gives a reflectance of ~30%. Many wavelengths fit in the *longitudinal* direction, which we want to be the direction of the emission. Transverse modes must be avoided because they distort the beam profile, making it difficult to collimate and focus the beam. The standing wave condition is: $k_m = m\pi / L \Rightarrow L = m\lambda / 2n_r$ with m = 1,2,3,... The frequency separation of the modes is: $\Delta\nu = c / 2n_r L$, i.e. the reciprocal of the round trip time. For a GaAs laser with L=500μm, $\Delta\nu \approx$100GHz. The wavelength must also satisfy the condition: $E_G \approx hc / \lambda$; this condition is not precise since the condition for amplification is: $E_G < h\nu < (E_{F_c} - E_{F_v})$.



(a) longitudinal modes **(b) transverse modes**



(a) **(b)**

This figure shows (a) the broad spontaneous emission spectrum of a GaAs laser below threshold, and (b) the longitudinal mode structure that appears above threshold.

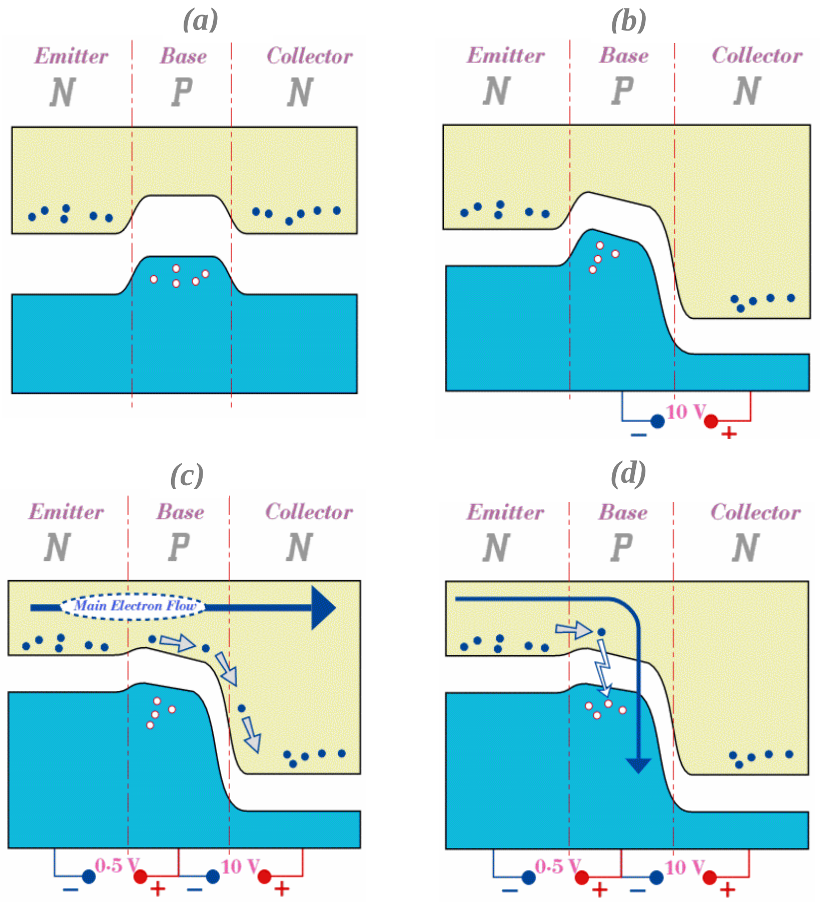
9. Bipolar Junction Transistors

The devices described so far are two-terminal devices. We now turn attention to the bipolar junction transistor which is a three-terminal device. There are two alternative arrangements: *pnp* and *npn*: the function is the same except that the signs of currents and voltages are reversed. The device is bipolar in the sense that both electrons and holes are involved in the carrier transport – unlike the FET described below where only the majority carriers participate

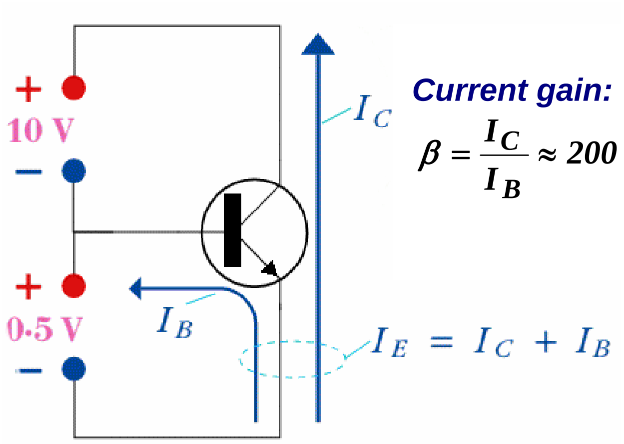
The BJT is characterised by:

- Two back-to-back *p-n* junctions sharing a common base region,
- Carriers are injected into the base from the emitter by forward biasing the E-B junction, and captured by the collector by reverse biasing the B-C junction
- The base is much shorter than the diffusion length, so that carriers can cross the base to reach the collector,
- The base remains neutral except in the depletion regions.

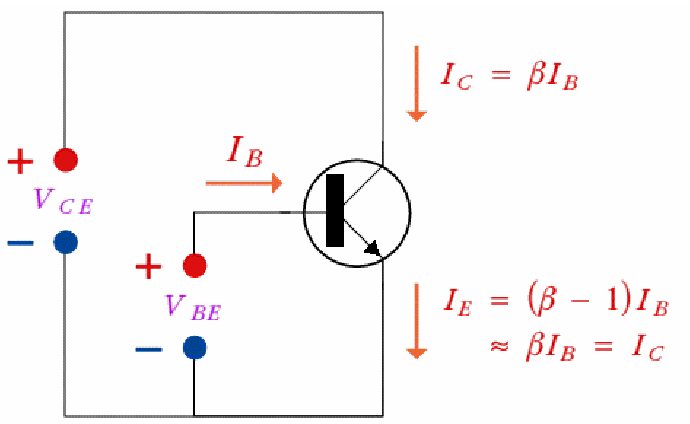
We can see in the diagram that the majority carriers in the base of an *npn* transistor (holes) are trapped in a potential well. Consequently the useful current through the base is carried mostly by minority carriers (electrons). This current is strongly dependent on the height of the potential barrier in the conduction band of the base, which is controlled by the external bias and signal voltages:



(a) in the absence of a bias there is no current,
(b) with the collector reverse biased the BC depletion layer width increases,
(c) the emitter is forward biased by 0.5-1.0V, electrons flow into the base and are immediately attracted to the collector; the collection efficiency $\alpha \approx 0.995$. Too high a bias can produce a large current and can easily damage the device.
(d) typically about 1% of electrons recombine in the base: they must be extracted - otherwise a potential would build up and repel electrons from the emitter - and so they appear as a weak base current.



The device therefore acts as a current amplifier - we put into the base a current I_B and a much larger current I_C is obtained at the collector (NB electron currents are shown here).

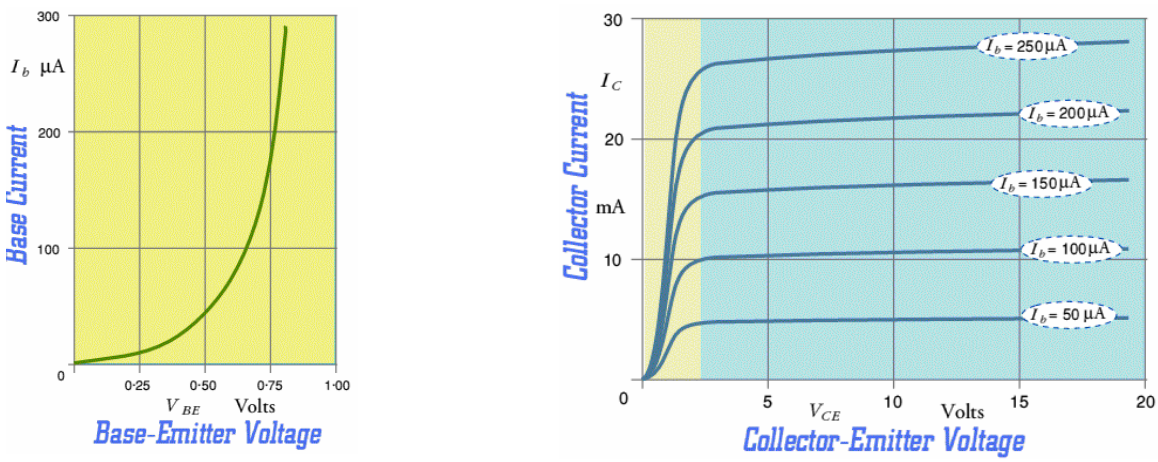


This is the conventional picture of a *npn* transistor. The collector voltage is referred to the emitter (V_{CE}), and the current is in the conventional positive-to-negative direction. This arrangement with a load resistor in the collector circuit is used in a common-emitter amplifier.

[More accurately, $I_E = I_C + I_B$, but we also have $I_C = I_S + \alpha I_E$ where I_S is the collector junction reverse current. Therefore:

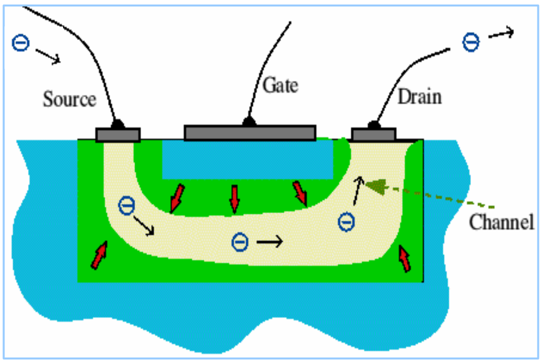
$I_C = \frac{I_S}{1-\alpha} + \frac{\alpha I_B}{1-\alpha}$, and we define $\beta = \frac{\alpha}{1-\alpha}$; we can usually ignore I_S <<I_B.]

10 Transistor Characteristics

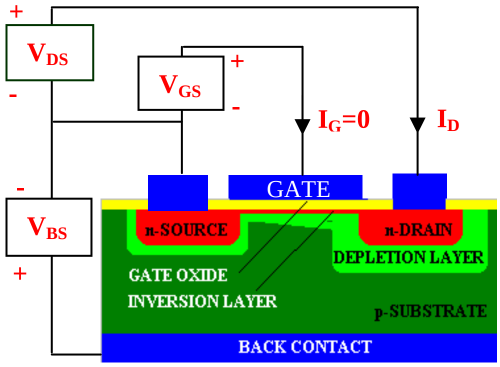


The base current I_B varies with V_{BE} very much as we would expect for a diode: it increases roughly exponentially. The collector current is determined largely by V_{BE} - but only when V_{BC} is large enough to collect all electrons emitted into the base. Thus I_C becomes independent of V_{CE} above a few volts.

11. Field Effect Transistor

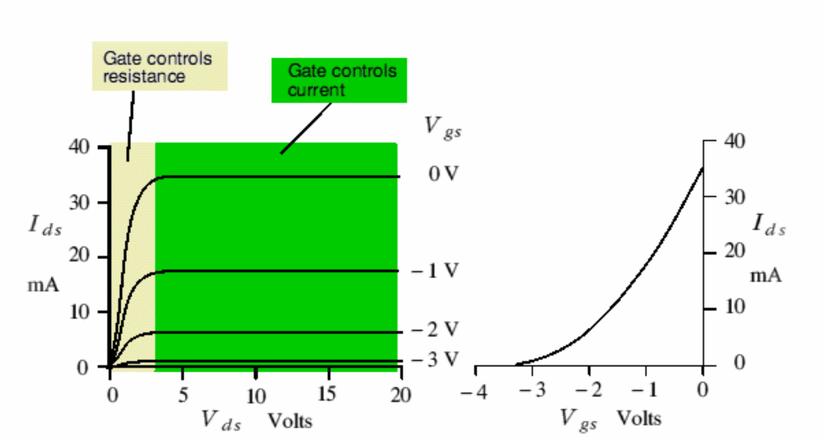


Junction Field Effect Transistor



Metal Oxide Semiconductor
Field Effect Transistor

Field effect transistors (FETs) differ from BJTs in several important respects:
(i) they depend on majority carrier transport - unipolar
(ii) simpler to fabricate, smaller volume - large scale integration of MOSFETs
(iii) can act as a resistor - a digital system can be made entirely of MOS devices
(iv) input impedance ~ $10^{10}\Omega$ - high fan-out possible
(v) small internal capacitance - can function as a memory device
(vi) they are less noisy than BJTs



The principle of operation of a FET is the control of the channel current by application of a reverse potential to the gate. In the JFET the depletion layer becomes wider, the conducting channel gets narrower and the drain- source current I_{ds} decreases. Provided V_{ds} is above some minimum value (about 2V) we find that the current doesn't depend very much on the actual drain-source voltage. When it is less than this ‘turn on’ value the current does vary with the drain-source voltage.

This behaviour can be explained by the cancellation of two opposing effects at high V_{ds}. The higher voltage along the channel speeds up the electrons, increasing the current; but the channel gets "pinched off" near the drain because of the large voltage difference in this region (the gate-source voltage is constant so that the depletion layer near the source is nearly constantat). In the region of low V_{ds} the behaviour is approximately ohmic - the device behaves as a voltage controlled resistor.