The Diode

The diode is the simplest type of semiconductor device. In generally, it is a two-terminal electronic device that permits current flow predominantly in only one direction.

- The current they pass depends upon the voltage between the leads.
- They do not obey Ohm's law!

Most diodes are **semiconductor devices**; diode electron tubes, now uncommon, are also available.

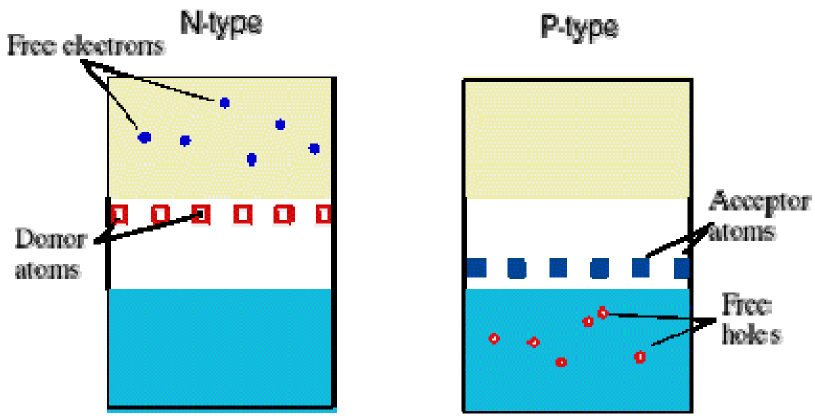
There are two main types of semiconductor materials:

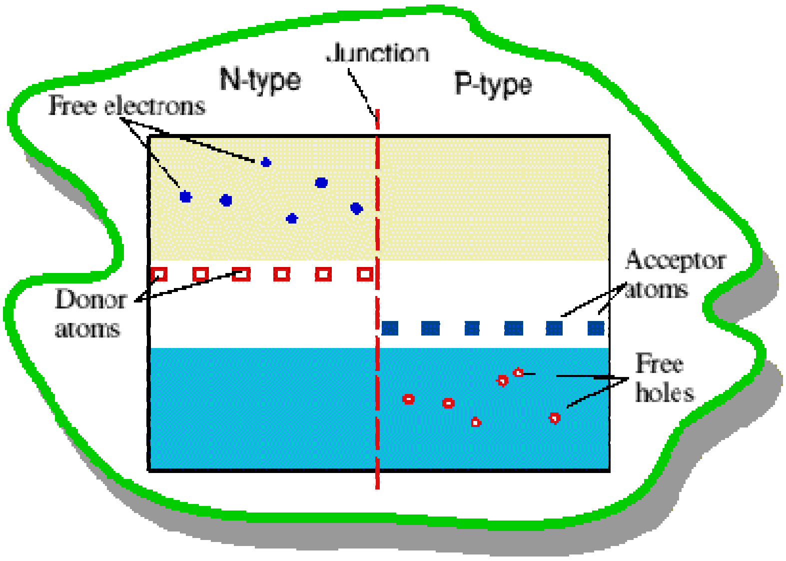
- intrinsic - where the semiconducting properties of the material occur naturally i.e. they are intrinsic to the material's nature.
- extrinsic - they semiconducting properties of the material are manufactured, by us, to make the material behave in the manner which we require.

Nearly all the semiconductors used in modern electronics are extrinsic. This means that they have been created by altering the electronic properties of the material. The two most common methods of modifying the electronic properties are:

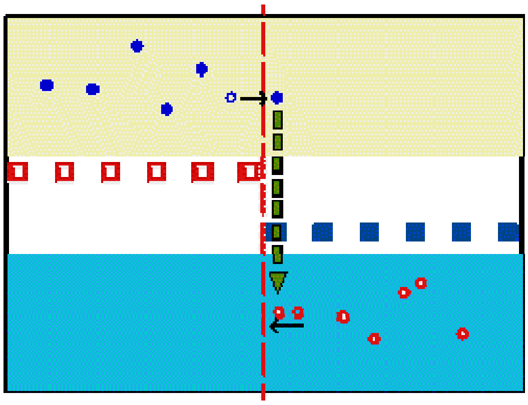
- **Doping** - the addition of 'foreign' atoms to the material.
- Junction effects - the things that happen when we join differing materials together.

To understand how a p-n junction diode works, begin by imagining two separate bits of semiconductor, one n-type, the other p-type.

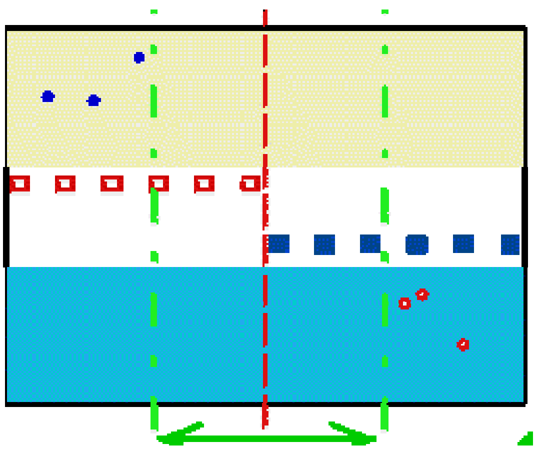




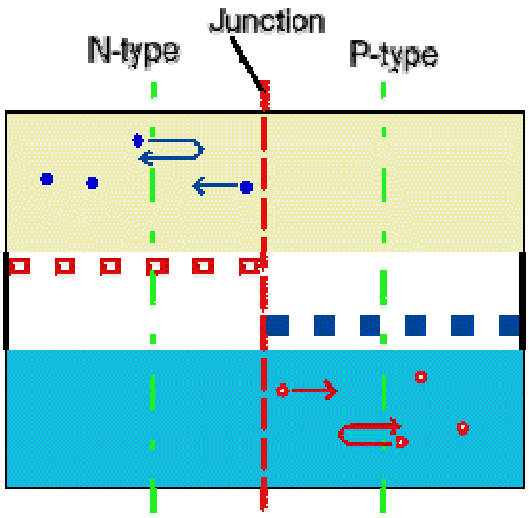
Bring them together and join them to make one piece of semiconductor which is doped differently either side of the junction.



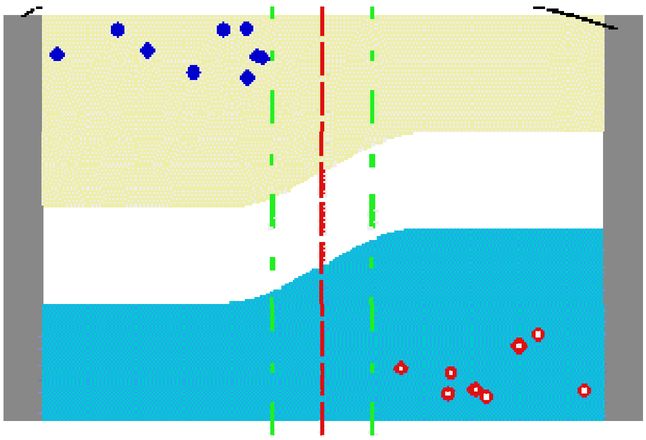
Free electrons on the n-side and free holes on the p-side can initially wander across the junction. When a free electron meets a free hole it can 'drop into it'. So far as charge movements are concerned this means the hole and electron cancel each other and vanish.



As a result, the free electrons near the junction tend to eat each other, producing a region depleted of any moving charges. This creates what is called the depletion zone.



Now, any free charge which wanders into the depletion zone finds itself in a region with no other free charges. Locally it sees a lot of positive charges (the donor atoms) on the n-type side and a lot of negative charges (the acceptor atoms) on the p-type side. These exert a force on the free charge, driving it back to its 'own side' of the junction away from the depletion zone.
The acceptor and donor atoms are 'nailed down' in the solid and cannot move around. However, the negative charge of the acceptor's extra electron and the positive charge of the donor's extra proton (exposed by it's missing electron) tend to keep the depletion zone swept clean of free charges once the zone has formed. A free charge now requires some extra energy to overcome the forces from the donor/acceptor atoms to be able to cross the zone. The junction therefore acts like a barrier, blocking any charge flow (current) across the barrier.

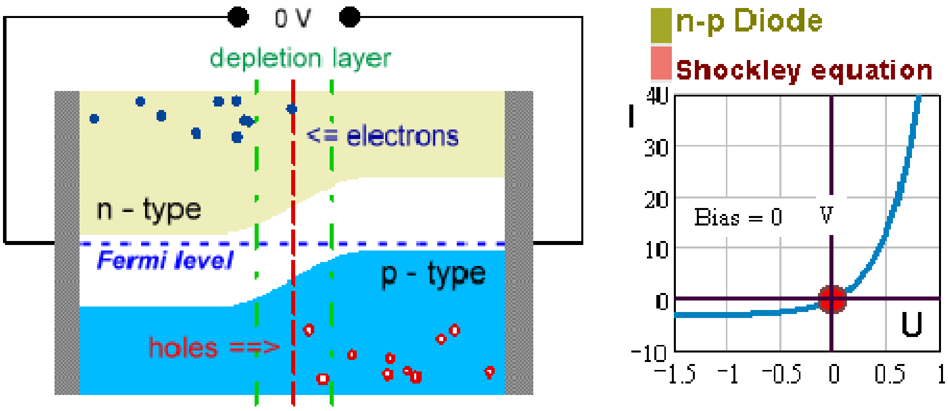


Usually, we represent this barrier by 'bending' the conduction and valence bands as they cross the depletion zone. Now we can imagine the electrons having to 'get uphill' to move from the n-type side to the p-type side. For simplicity we tend to not bother with drawing the actual donor and acceptor atoms which are causing this effect!

The holes behave a bit like balloons bobbing up against a ceiling. On this kind of diagram you require energy to 'pull them down' before they can move from the p-type side to the n-type side. The energy required by the free holes and electrons can be supplied by a suitable voltage applied between the two ends of the pn-junction diode. Notice that this voltage must be supplied the correct way around, this pushes the charges over the barrier. However, applying the voltage the 'wrong' way around makes things worse by pulling what free charges there are away from the junction!

This is why diodes conduct in one direction but not the other.

We create a **p-n junction** by joining together two pieces of semiconductor, one doped n-type, the other p-type. This causes a *depletion zone* to form around the junction (the join) between the two materials. This zone controls the behaviour of the diode. The animation shows the general behaviour of a p-n junction.



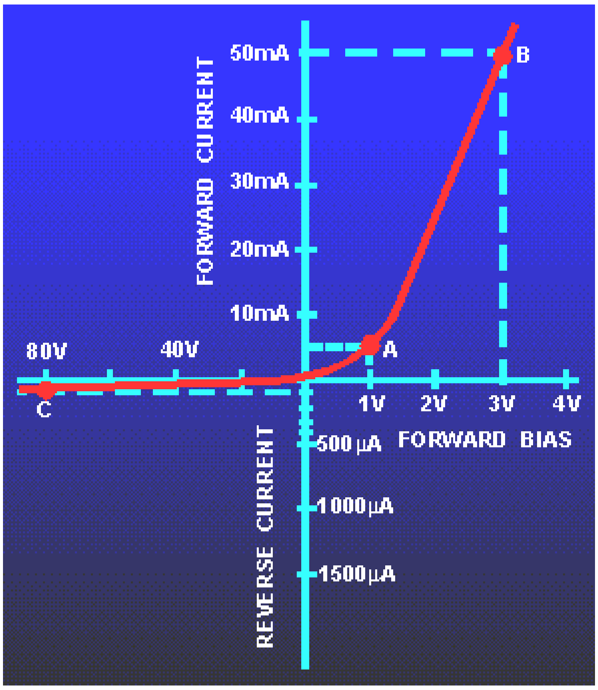
When we apply a potential difference between the two wires in one direction we tend to pull the free electrons and holes away from the junction. This makes it even harder for them to cross the depletion zone.

When we apply the voltage the other way around we push electrons and holes towards the junction, helping to give them extra energy and giving them a chance to cross the junction.

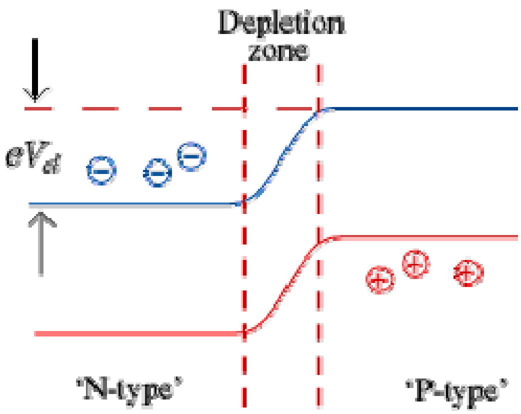
Therefore, when a p-n junction is reverse biased, there will be no current flow because of majority carriers but a very small amount of current because of minority carriers crossing the junction. However, at normal operating temperatures, this small current may be neglected.

In summary, the most important point to remember about the p-n junction diode is its ability to offer very little resistance to current flow in the forward-bias direction but maximum resistance to current flow when reverse biased. Figure 20 shows a plot of this voltage-current relationship (characteristic curve) for a typical p-n junction diode.

Figure 20. - p-n junction diode characteristic curve.



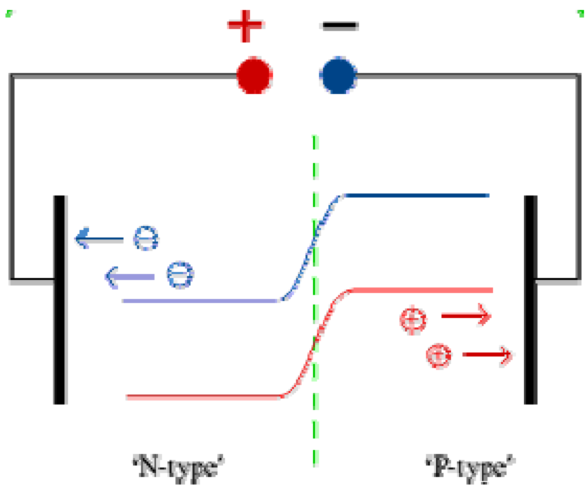
The free electrons inside the n-type material need some extra energy to overcome the repulsion of the p-type's acceptor atoms. If they do manage to get past this energy barrier some of their kinetic energy will have been converted into potential energy, but — once well clear of the depletion zone — they can move around OK unless they 'fall into a hole'.



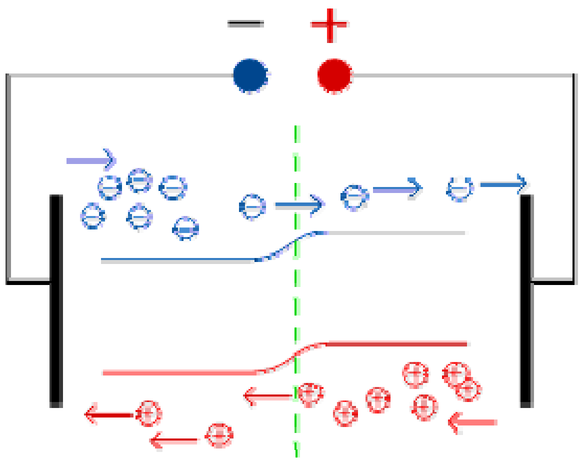
This action is usually described using a conventional 'energy level diagram' of the sort shown in figure 3b. The electrons can roll around the 'flat' parts of the energy diagram, but need extra energy to roll up the step and move from n-type to p-type across the junction. Coming the other way they'd 'drop down' and zip into the n-type material with extra kinetic energy. The size of this energy barrier can be defined in terms of a junction voltage $\mathbf{V_d}$.
This means the amount of energy converted from kinetic to potential form (or vice versa) when an electron crosses the depletion zone is $\mathbf{eV_d}$ where **e** is the charge on a single electron.

A similar argument applies to the free holes in the p-type material. However, this is more complicated to understand because holes have the 'strange' property that their energy increases as they go 'down' the energy level diagram. Their behaviour is similar in some ways to a bubble underwater. This is because a hole is the absence of an electron. (To create a hole at a lower level we have to lift an electron up to the conduction band from further down — this takes more energy.) As a result, the moveable holes in the p-type material find it difficult to 'roll down' the barrier and get into the n-type material. They keep trying to 'bob up' to the top of the band.

The electrical properties of the diode can now be understood as consequences of the formation of this energy barrier & depletion zone around the p-n junction. The first thing to note is that the depletion zone is free of charge carriers & the electrons/holes find it difficult to cross this zone. As a result, we can expect very little current to flow when we apply a small potential difference. (Here, 'small' mean small compared with V_d, which an electron requires to get over the potential barrier.)



The effect of applying a bigger voltage depends upon which way around it's connected. When we make the n-type side more $+^{ve}$ (i.e. drag some of the free electrons out of it) & the p-type more $-^{ve}$ (drag holes out of it) we increase the difference in potential across the barrier. This makes it even harder for a stray electron or hole to cross the barrier.

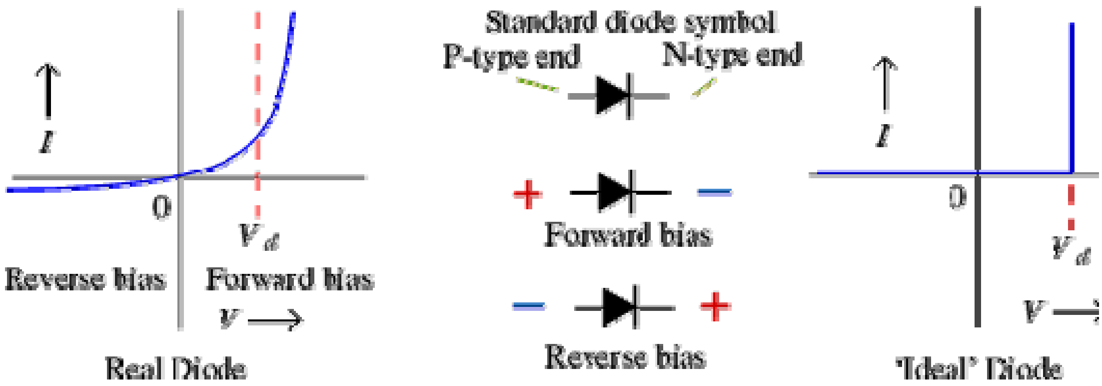


When we make to n-type side more $-^{ve}$ & the p-type side more $+^{ve}$ we force lots of extra electrons & holes into junction. To understand the effect of this, let's concentrate on the electrons.

The electrons in the n-type region near the junction are repelled by the fixed acceptor atoms in the p-type. However, they're also repelled by the other electrons drifting around inside the n-type material.
When we shove extra electrons into the n-type material we increase the number of electrons 'pushing' from this side. This works a bit like pressure in a gas. Electrons near the junction are helped across the junction by being shoved from behind. The effect is to reduce the amount of extra energy required to cross the barrier — i.e. the barrier height reduces. As a

result, the barrier is reduced or removed and free electrons/holes can move freely from one material to the other.

On the basis of the explanation given above we might expect no current to flow when the diode is reverse biassed. In reality, the energies of the electrons & holes in the diode aren't all the same. A small number will have enough energy to overcome the barrier. As a result, there will be a tiny current through the diode when we apply reverse bias. However, this current is usually so small we can forget about it.



The polarity which produces a significant current is called Forward bias. When we raise the forward bias voltage we reduce the diode's energy barrier. This essentially reduces the diode's resistance.

Diode's don't obey Ohm's Law! A more accurate analysis of the physics of a diode shows that the IV curve is a an exponential relationship of the general form

$$I\{V_f\} = I_d(\mathrm{Exp}\{\alpha V_f\} - 1)$$

where the values of α and $\mathbf{I_d}$ depend upon the diode. In fact, the actual IV relationship of a real diode will depend on the details of how it was made.
Most modern diodes are made of silicon. For this semiconductor material, $\mathbf{V_d}$ is about 0,5 V. We should really remember that a forward biassed diode always passes some current, but this current is small unless the applied voltage is around half a volt.

THEORY

We will now present some of the mathematics used to describe semiconductor doping levels, by further investigating a mathematical model used to determine doping profiles called The Inverse Problem:

Consider n-type semiconductor with the following definitions:

$\mathbf{k_B}$ = Bolzmann's constant
ε_s = The dielectric constant
q = A positive elementary charge.
T = Temperature (**^{o}K)**
n = The number of electrons
p = The number of holes
$\mathbf{n_i}$ = The intrinsic concentration of carriers in the material **($n_i^2 = n\,p$)**
For a pure crystal **$n=p=n_i$**

ϕ = The potential
ψ = The electrostatic potential
φ_n = The quasi-Fermi potential
Where $\phi = \varphi_n - \psi$
$\mathbf{N_{D,A}}$ = The concentration of donors **(N_D** for electrons; **N_A** for holes)
ρ = The total charge **($p - n + N_D - N_A$)**
Λ_Δ = The Debye length

We have for the concentration of donors which is equal to the initial number of carriers multiplied by an exponential growth term:

$$N_D = n_i \exp (q\, \varphi_n / (k_B T))$$

We now consider the Poisson Equation for n-type semiconductor:

$$d^2 j/dx^2 = - \rho/\varepsilon_s = - q /\varepsilon_S (p - n + N_D),$$

where $\rho = p - n + N_D - N_A$, and N_A is negligible.

Also:

$$n = n_i \exp (q (\varphi_n - \psi) / (k_B T)), \quad p = n_i \exp (-q (\varphi_n - \psi) / (k_B T))$$

Assuming equilibrium conditions with $p << n, N_D$, and φ_n constants, we get:

$$d^2\psi/dx^2 = q/\varepsilon_s * N_D - n_i \exp(q (\varphi_n - \psi) / (k_B T)), \text{ since } \psi = \varphi_n - \phi$$

$$= q/\varepsilon_s * N_D [1- \exp (-q\psi / (k_B T))] = \sim q^2 N_D / (\varepsilon_s k_B T)*\psi$$

Or, in words, the potential is proportional to the square of the charge, the carrier concentration (donors, in this case), and inversely proportional to the temperature. We set L_D, which is the Debye Length equal to the following:

$$L_D = (\varepsilon_s k_B T / (q^2 N_D))$$

Hence we have:

$$d^2\psi/dx^2 = \psi / L_D^2$$

Where $L_D =\sim 10^2$ Angstroms, 10^3 Angstroms, 10^4 Angstroms

For $N_D = 10^{17}$ cm^{-3}, 10^{15} cm^{-3} , 10^{13} cm^{-3} respectively.

Solving the second order differential equation we get:

$$y = C_1 \exp [(1/ L_D) x]+ C_2 \exp [-(1/ L_D) x]$$

Since the electrostatic potential (ψ) is finite, and because DAVE said so, we disregard the exponentially growing portion and consider the decaying portion alone.

$$y = C \exp [-(1/ L_D) x]$$

Hence, we find that by the simplified Inverse Problem model, the electrostatic potential decays as we move across the semiconductor. Note that for x smaller than the Debye Length, the function approaches one, not zero. The model is insensitive to changes in the doping profile that occur in a distance less than the Debye length. These lengths are significant for very small devices used in present technology.

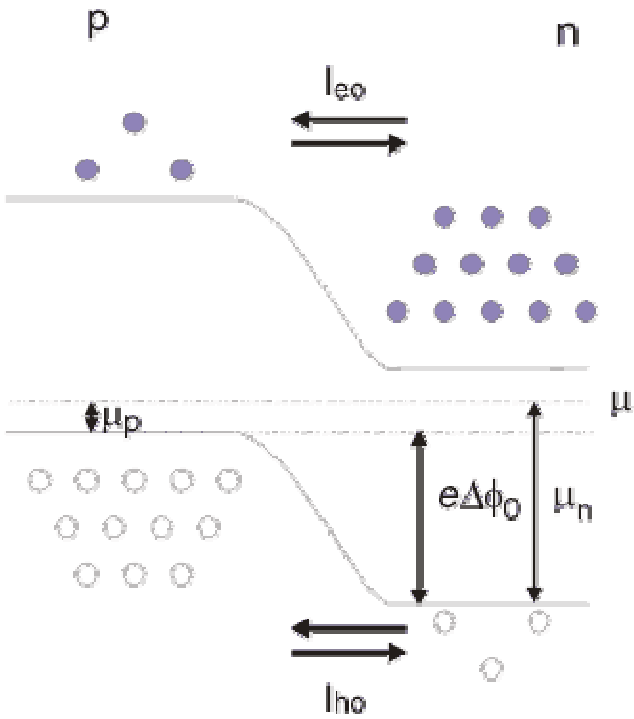
In order to properly treat the situation we would need to consider the complete inverse problem for the semiconductor device, it is a five non-linear second order differential equation system. Two of the equations are series and boundary value problems, which consider the change in the number of carriers as a function of the change of carrier current, and the recombination rate (a non-linear function of n, p). The third equation considers contributions to the potential from the total charge $\rho\,(p\text{-}n+ N_D\text{-} N_A)$. The last two equations in the system consider the electron and hole current relations.

For further information about the complete Inverse Problem see Friedman's book (in the reference page)

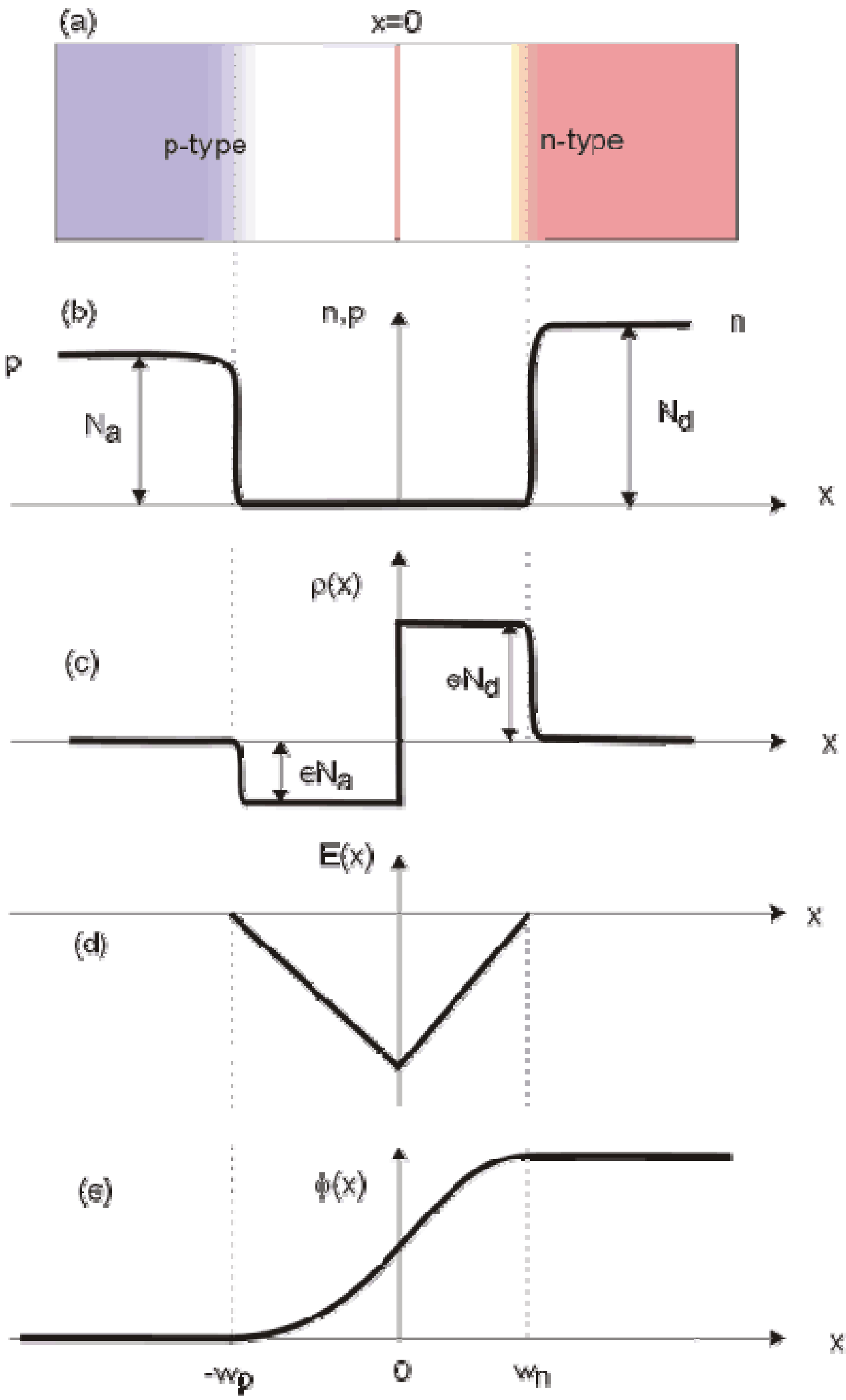
The p-n Junction

Junctions are crucial to many semiconductor applications.
The oldest method of making a p-n junction is by *diffusion*. The dopant diffuses in under heating so that the surface acceptor concentration exceeds the donor concentration. A junction appears when N_d=$N_{a.}$
Another doping technique is ion *implantation*. The starting material, n-type only, is bombarded with the required species of ions, say acceptors. This produces sharper junctions, but causes damage to the crystal lattice structure increasing the number of dislocations and interstitial atoms.
Epitaxial deposition techniques are now widely established. The starting material is a single crystal in all cases, so it is possible to grow further crystal layers which are in register with the starting crystal. The most precise although most expensive way of achieving this is *Molecular Beam Epitaxy* (MBE). Ions of the semiconductor together with dopants are fired at the crystal surface. This technique can produce very sharp junctions and there is no counter-doping, i.e. no donors in the p-type region.
We are supposed to be discussing p-n junctions!

We now consider a p-n junction in the absence of voltage bias, so that it is in thermodynamic equilibrium. This means that the chemical potential μ must be constant across the junction. Since μ is near the valence band edge in a p-type region and near the conduction band in an n-type region, the bands must bend through the junction as shown below:



In the region where the bands are bending, $\boldsymbol{\mu}$ is near the middle of the gap and therefore $n<n_0$, $p<p_0$, where n_0 and p_0 are the concentrations deep inside the n and p type regions. This is therefore called the depletion region. If we assume the junction is sharp, with N_a and N_d changing abruptly at the junction, and assume $N_d=N_a$, $p_0=N_a$, thus we see that within the depletion region $p<N_A$ on the left and $n<N_d$ on the right. Since N_A and N_D are the densities of ionized acceptor and donor ions, this means that there is a net negative charge on the p-type side of the junction, and a net positive charge on the n-type side. These separated charges generate an electric field, which is the physical cause of the band bending. The overall picture is summarized below.



Depletion Layer

We now calculate charge, electric field and potential. First we find the band offset. $e\Delta\phi_0$, i.e. the difference in the height of μ above the valence band on the two sides. In the n-type region, the value of μ relative to the local valence band is given by

$$\mu_n = E_g - k_b T \ln\left(\frac{N_c}{N_d}\right) \qquad (2.1)$$

assuming N_A=0 on the n-side. On the p-type side, we have

$$\mu_p = k_B T \ln\left(\frac{N_v}{N_a}\right) \qquad (2.2)$$

assuming N_D=0 on this side. These give

$$e\Delta\phi_0 = \mu_n - \mu_p = E_g + k_B T \ln\left(\frac{N_d N_a}{N_c N_v}\right) \qquad (2.3)$$

Using the definition of the intrinsic density n_i (1.26), gives

$$\Delta\phi_0 = \frac{k_B T}{2} \ln\left(\frac{N_d N_a}{n_i^2}\right) \qquad (2.4)$$

This is the difference in the electrostatic potential between the two sides since ***e$\Delta\phi_0$*** is the energy difference between electrons at the bottom of the conduction band on the two sides, as may be seen from the diagram.

The variation of ***$\phi(x)$*** and the electric field

$$E = -\frac{d\phi}{dx} \qquad (2.4.1)$$

across the junction can be calculated as long as the variations with x of N_D and N_A are known. If we assume an abrupt junction then the charge distribution has the form:

$$\rho(x) = \begin{cases} -N_a e & -w_p < x < 0 \\ N_d e & 0 < x < w_n \end{cases} \qquad (2.5)$$

where w_p and w_n are the semi-widths of the depletion region on the p-side and n-side; values for them will be found later.

The electric field satisfies Gauss' law:

$$\nabla \mathbf{D} = \rho \qquad (2.6)$$

since $\mathbf{D} = \varepsilon\varepsilon_0 \mathbf{E}$ with ε constant, and in the present 1-d case $\nabla\cdot\mathbf{E} = \frac{dE}{dx}$, this gives

$$\frac{dE}{dx} = \begin{cases} -N_a e/\varepsilon_0\varepsilon & -w_p < x < 0 \\ N_d e/\varepsilon_0\varepsilon & 0 < x < w_n \\ 0 & \text{otherwise} \end{cases} \qquad (2.7)$$

The boundary conditions are E=0 for x<-w_p and x>w_n since the junction is in equilibrium. The solution of (2.7) is therefore,

$$E = \begin{cases} -\dfrac{N_a e}{\varepsilon_0 \varepsilon}(x + w_p) & -w_p < x < 0 \\ \dfrac{N_d e}{\varepsilon_0 \varepsilon}(x - w_n) & 0 < x < w_n \end{cases} \qquad (2.8).$$

Furthermore, E must be continuous at x=0, which gives

$$N_a e w_p = N_d e w_n \qquad (2.9)$$

As seen from the figure, this is simply the condition of electrical neutrality of the whole depletion region. The variation of E with x given is shown in (d)

Equations (2.4.1) and (2.7) together give ***ϕ (x)***

$$\phi = \begin{cases} \dfrac{eN_a}{2\varepsilon_0 \varepsilon}(x + w_p)^2 & -w_p < x < 0 \\ \Delta\phi - \dfrac{eN_d}{2\varepsilon_0 \varepsilon}(x - w_n)^2 & 0 < x < w_n \end{cases} \qquad (2.10)$$

ϕ must be continuous at x=0; this gives a second relation between w_p and w_n:

$$\Delta\phi_0 = \frac{e}{2\varepsilon_0 \varepsilon}\left(N_a w_p^2 + N_d w_n^2\right) \qquad (2.11)$$

Recall that $\Delta\phi_0$ is already known from (2.4). The variation of ϕ with x given by (2.10) and is shown in (c). As stated, (2.4), (2.10) can be solved for the values of w_p and w_n; they are,

$$w_n = \left(\frac{2\varepsilon_0 \varepsilon N_a \Delta\phi_0}{eN_d(N_a + N_d)}\right)^{1/2} \qquad (2.12)$$

$$w_p = \left(\frac{2\varepsilon_0 \varepsilon N_d \Delta\phi_0}{eN_a(N_a + N_d)}\right)^{1/2} \qquad (2.14)$$

The dependence on T and on the doping deserves comment. Assuming for simplicity $N_A=N_D$ (equal doping on both sides), so that the factors $\frac{N_a}{N_d}$ and $\frac{N_d}{N_a}$ are both one, substitution of (2.11) for $\Delta\phi_0$ gives:

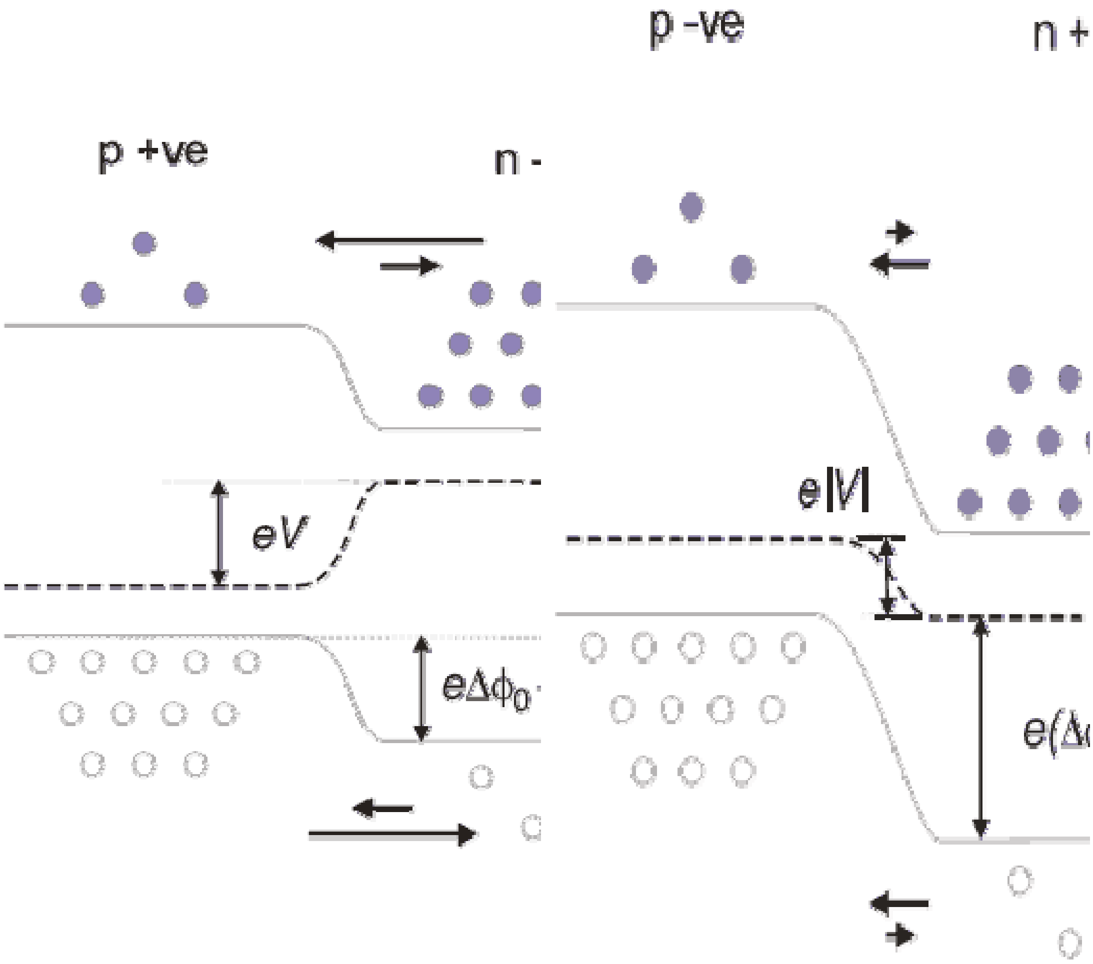
$$w = w_n + w_p \propto \left(\frac{T \ln\left(N_a^2\right)}{N_a + N_d}\right)^{1/2} \qquad (2.15)$$

The logarithmic dependence is very weak compared with the denominator, so:

w increases as T increases
w decreases as doping N_A+N_D increases

The application of p-n junctions depends on having an applied voltage so that a current flows through the junction. With a voltage V applied a difference eV appears between the values of μ on the p and n side. We distinguish between forward and reverse bias, as in the diagram below:

Forward Bias **Reverse Bias**



The calculation of $\Delta\phi_0$ etc. go through much as before, and give:

$$w_n = \left(\frac{2\epsilon_0 \epsilon N_a(\Delta\phi_0 - V)}{eN_d(N_a+N_d)}\right)^{1/2} \qquad (2.16)$$

$$w_p = \left(\frac{2\epsilon_0 \epsilon N_d(\Delta\phi_0 - V)}{eN_a(N_a+N_d)}\right)^{1/2} \qquad (2.17)$$

There is a charge separation in the depletion region. Thus the depletion region behaves like a capacitor, and the capacitance is given by

$$C = \frac{A\sigma}{V} = \frac{AeN_a w_a}{V} = \frac{A}{V}\left(\frac{2\varepsilon_0 \varepsilon N_a N_d (\Delta\phi_0 - V)}{N_a + N_d}\right)^{1/2} \qquad (2.18)$$

Where A is the area of the junction (in the y-z plane). This has the useful property that the capacitance can be varied by applied voltage. In practice, reverse bias is needed so that the current flow is small. A p-n junction device used as a voltage-variable capacitor is known as a ***varactor*** diode.

The important property of a p-n junction is the current-voltage characteristics. A derivation requires discussion of diffusion and recombination of carriers. The result is

$$I = I_0 \, Exp\left[(eV/k_BT) - 1\right] \qquad (2.19)$$

where the predominant temperature dependence of I_0 is given by

$$I_0 \propto n_i^2 \quad (2.20)$$

The first semiconductor lasers where made from heavily doped p-n junctions. Under conditions of forward bias the electrons and holes would recombine at the barrier junction producing some laser emission at high currents. These devices were inefficient and had high threshold currents as the majority carriers tended to drift away from the junction interface. It was soon discovered that more efficient lasers could be produced by the implementation of a heterostructure design.