9.1 Semiconductor Devices 3225: Ian Robinson

Devices are made by doping of semiconductors. Deliberate addition of impurities strongly affects the electrical properties.

"1ppm" material is usually considered "ultrapure"

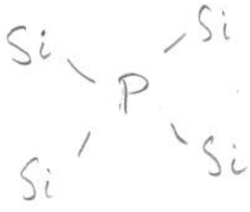
"5N" = 99.999% pure is highest grade attainable

1 part in 10^{11} is world record purity (for Ge)

Therefore, if impurities are electrically active (contribute or remove electrons) they override all of the intrinsic effects. This realisation delayed the start of the semiconductor industry.

Most common doping method is to use impurities which can be incorporated in the host lattice:

- same size, nearby in periodic table
- same chemical config, forming covalent bonds.



Phosphorus atom substitutes for Si without disturbing lattice (not quite true) Donor

Si $3s^2 3p^2$ $Z = 14$

P $3s^2 3p^3$ $Z = 15$

One more nuclear charge + one more electron.

Concentration: 1 in 10^5 typ = dilute limit

Extra electron does not contribute to covalent bonding, but does contribute to counting of electron states. So it goes in Conduction band.

System stays charge-neutral

$\sim \frac{10^{22}}{10^5} \sim 10^{17}$ electrons in conduction band $\gg n_{\text{intrinsic}} \sim 10^{10}$.

W. Shockley (1950).

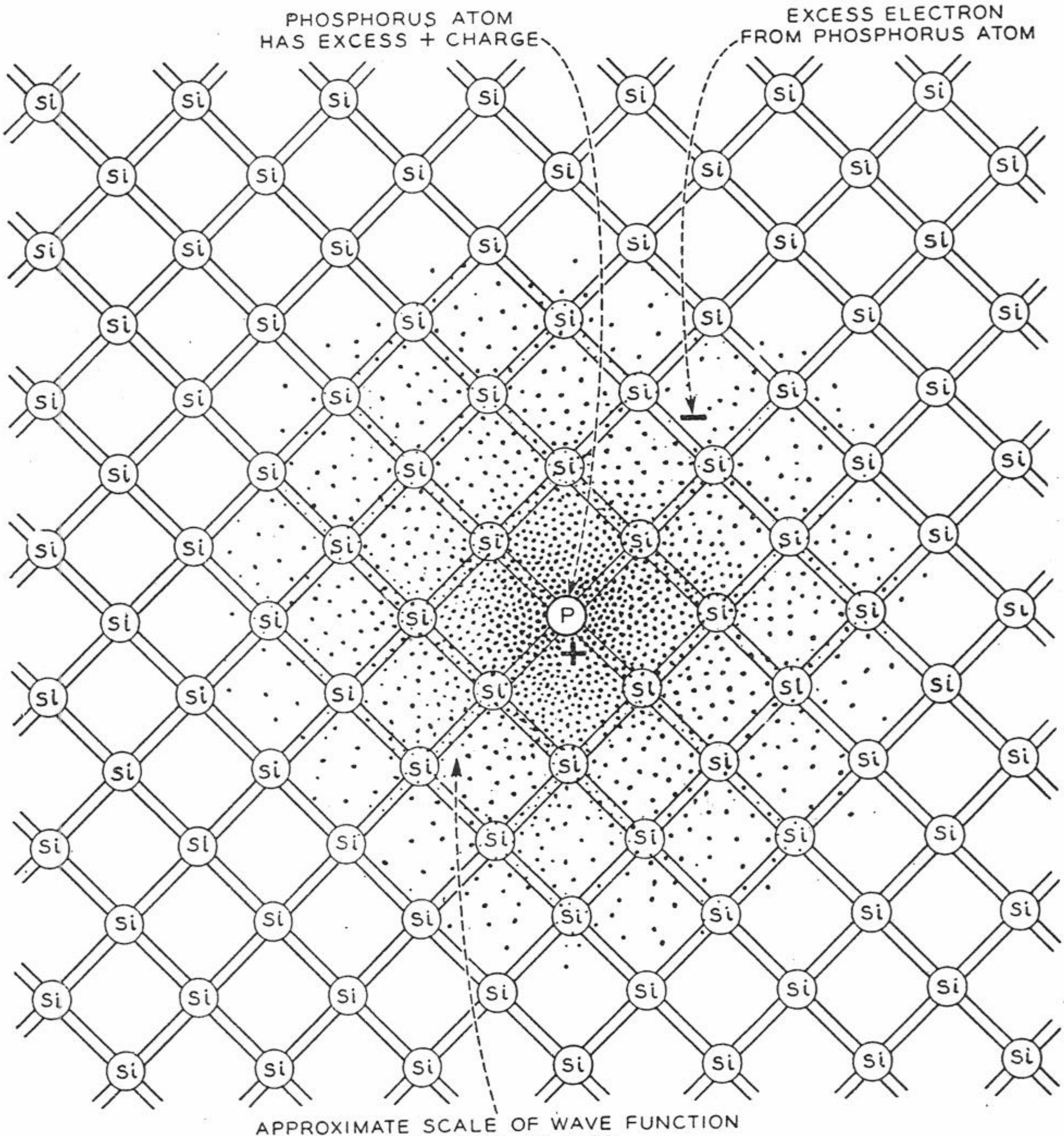
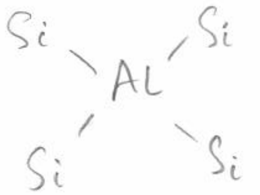


FIG. 1–14—Wave Function of Electron Bound to Phosphorous Atom in Silicon.

Similarly, doping with Al (or B) <u>Acceptor</u>



$Si\ 3s^2 3p^2 \to (sp^3)$ $Z = 14$

$Al\ 3s^2 3p^1$ $Z = 13.$

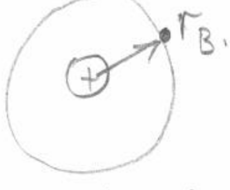
To form the fourth bond, electron is taken from the valence band, leaving a hole.

10^{17} holes >> P intrinsic dominates conductivity.

As, P doping makes n-type (Negative carriers)
B, Al " " P-type (Positive carriers)

9.2 Bound Donor States

We can consider the additional electron and the extra nuclear charge from the Bohr model. Assume filled band of all other electrons is inert.



Bohr model of hydrogen atom

$$\frac{e^2}{4\pi\varepsilon\varepsilon_0 r^2} = \frac{mv^2}{r}$$ classical forces.

Energy levels had to match spectrum (Balmer) $mvr = \hbar$ [angular momentum quantised]

$$E_B = \frac{e^4 m}{2(4\pi\varepsilon_0\hbar)^2} = 13.6\,eV \qquad r_B = \frac{4\pi\varepsilon_0\hbar^2}{me^2} = 0.53\,Å$$

All we have to change is $\varepsilon_0 \to \varepsilon\varepsilon_0$ ε = relative perm.
$m \to m_e$ = effective mass

5.03.07 II

$\varepsilon = 11.7$	(Si)	$m_e = 0.19\,m$
15.8	(Ge)	0.08 m
13	(GaAs)	0.07 m.

Both act to increase the Bohr radius by a factor >100 and reduce the energy by a factor >1000.

Orbits have radius ~ 50 Å = 5nm
- opportunities for quantum computers

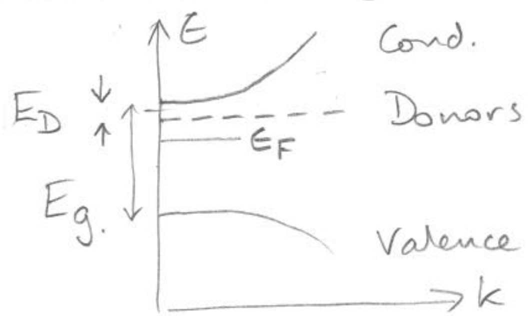
Energies reduced to 45 meV P in Si [~ same for all donors]
13 meV As in Ge

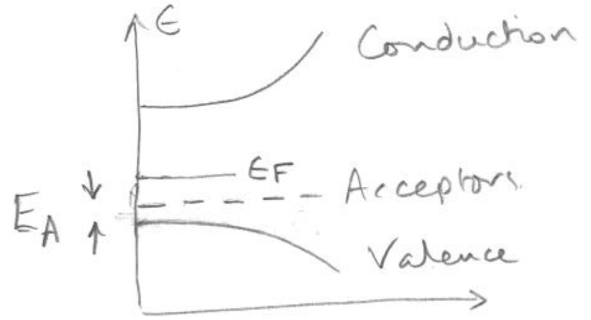
9.3 Acceptor States.

Exact analogy for holes in valence band.

E_A = 45 meV B in Si } compares with donors.
57 meV Al in Si }

To see how these affect the Fermi level, we have to study the band locations:





6.03.06 II

[Donors and Acceptors can form a genuine band if the concentration gets high enough, especially because the wavefunctions are so large].

At T = 300K, a significant number of Donors/Accs are ionized to pump electrons/holes into the nearby band. $k_BT = 25$ meV $E_D \sim 40$ meV

The Fermi level falls near to the small gap between the impurity and the nearby band.

$np = (n_i)^2$ Law of mass action $\left[\begin{array}{l} n(n-\Delta n) = n_i^2 \\ n^2 - n\Delta n - n_i^2 = 0 \end{array}\right.$

$n - p = \Delta n$ Doping, positive or negative

$$\left.\begin{array}{l} n \\ p \end{array}\right\} = \left(\left(\frac{\Delta n}{2}\right)^2 + n_i^2\right)^{1/2} \pm \frac{1}{2}\Delta n \quad \left\{\begin{array}{ll} \to n_i & n_i \gg \Delta n \\ \to \Delta n & n_i \ll \Delta n \end{array}\right.$$

If we define $\mu_i = \mu$ intrinsic which is the mid-point of the gap, adjusted for effective mass

$$n = n_i e^{(\mu-\mu_i)/kT}$$

$$p = n_i e^{-(\mu-\mu_i)/kT}$$

$$\Rightarrow \Delta n = 2n_i \sinh((\mu-\mu_i)/kT)$$, positive or negative.

This allows us to see how far away from mid gap does the Fermi level move when we introduce doping: Eg Si : $n_i = 4.5 \times 10^9$ cm^{-3}

$n_D = \Delta n = 10^{16}$ cm^{-3} (1 part per million)

10^{22} states
10^{16} donors.
10^{10} n_i

$$\frac{\Delta n}{n_i} = 10^6 \Rightarrow \frac{\mu - \mu_i}{kT} = 14.5$$

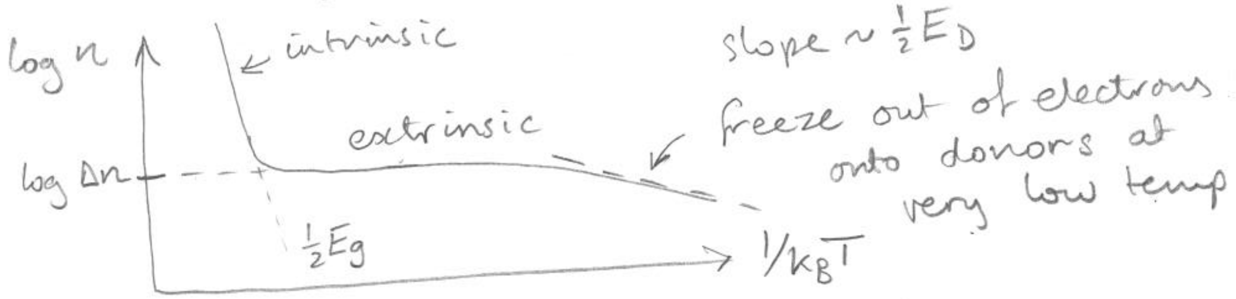
$$\mu - \mu_i = 0.36\text{ eV}$$

Since the band gap is 1.1 eV, the Fermi level moves much of the way up to the donor level.

16.03.09 I There is a crossover from intrinsic to extrinsic behaviour seen when we vary temperature.

Δn is fixed by doping.

$$n_i = 2\left(\frac{kT}{2\pi\hbar^2}\right)^{3/2}(m_e m_h)^{3/4} e^{-E_g/2kT} = \sqrt{np}$$



See this mainly in the conductivity:

10.03.08 I

New keywords:

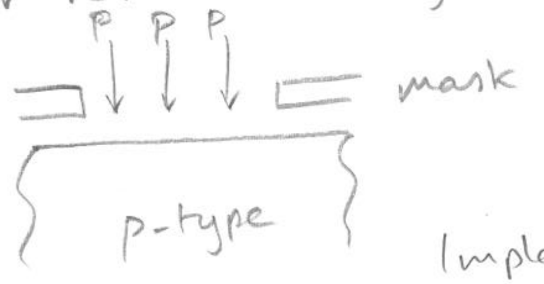
Carrier
Extrinsic / Intrinsic
Donor / Acceptor level
n/p-type doping.
mobility.

9.5 p-n Junction

Force p-type and n-type material together.

Grow crystal in presence of impurities, then switch.

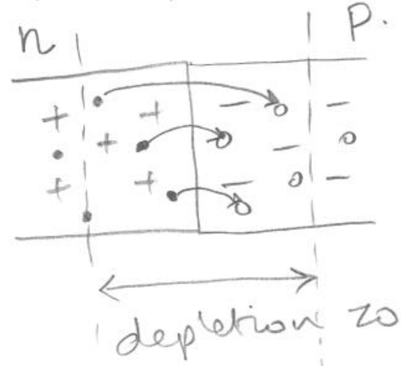
Or reverse doping afterwards:



<u>Implant</u> with donors to a higher concentration than used in p-type.

Implant energy → depth > 100nm to 1μm.

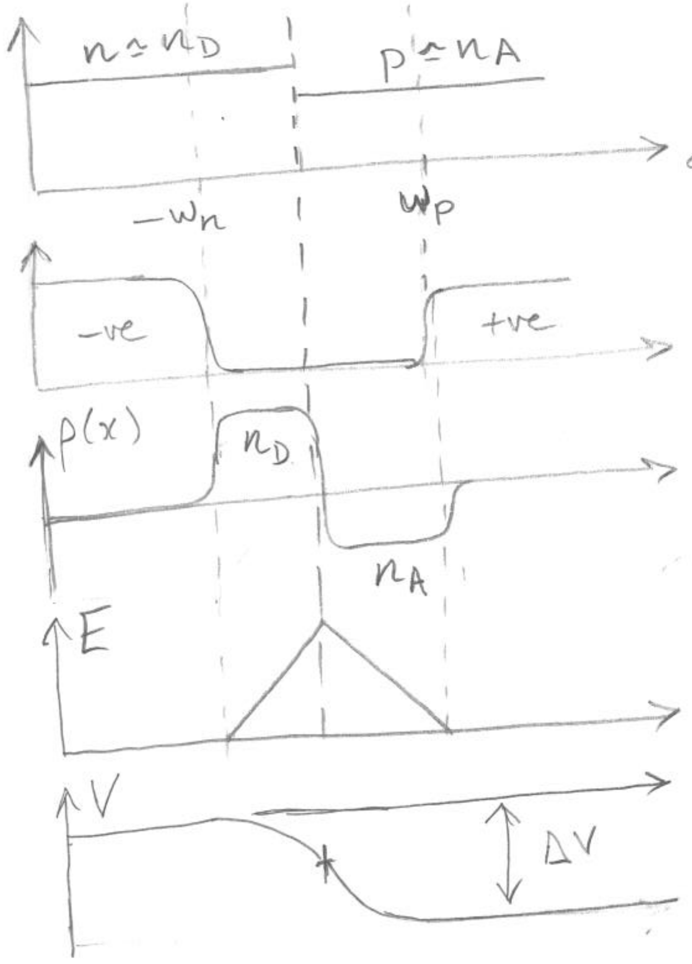
Nearby electrons and holes annihilate, but they leave behind the charged donors and acceptors, which are fixed to the lattice



The reaction stops because a voltage (potential difference) builds up across the D.Z.

= space charge.

Introduce notation:



Before reaction

After reaction (distances not equal)

Net charge density left over.

Electric field

$$\vec{\nabla}\cdot\vec{E} = \rho/\varepsilon\varepsilon_0$$

Potential

$$\vec{E} = -\vec{\nabla} V$$

07.03.06

HW → To keep the equations simple, we model the boundaries as <u>sharp</u>, although there will be partial ionization of the dopants near the sides of the Depletion Zone.

We also ignore intrinsic effects, assuming $n_i \ll n_D$ or n_A, even though there will be a very small carrier concentration in the D.Z.

Electrostatics to determine w_n and w_p:

One dimension $\nabla^2 \rightarrow \frac{d^2}{dx^2}$.

positive quantity

$$\frac{d^2}{dx^2}V(x) = \begin{cases} -n_D e/\varepsilon\varepsilon_0 & -w_n < x < 0 \\ +n_A e/\varepsilon\varepsilon_0 & 0 < x < w_p \\ 0 & \text{elsewhere} \end{cases}$$

constant.

Set $V=0$ at $x=-\infty$ and integrate across:

$$V(x) = \frac{-n_D e}{\varepsilon\varepsilon_0}\frac{1}{2}(x+w_n)^2 \qquad -w_n < x < 0.$$

$$V(0) = \frac{-n_D e}{\varepsilon\varepsilon_0}\frac{w_n^2}{2}$$

$$V(x) = -\Delta V + \frac{n_A e}{\varepsilon\varepsilon_0}\frac{1}{2}(x-w_p)^2 \qquad 0 < x < w_p$$

$$V(0) = -\Delta V + \frac{n_A e}{\varepsilon\varepsilon_0}\frac{w_p^2}{2} \qquad \text{from right.}$$

Matching $V(0)$ sets the value of ΔV:

$$\Delta V = \frac{e}{2\varepsilon\varepsilon_0}\left(n_D w_n^2 + n_A w_p^2\right)$$

Matching <u>slopes</u> $dV/dx\,|_{x=0}$ (no charges at interface)

$$\frac{-n_D e}{\varepsilon\varepsilon_0}w_n = \frac{-n_A e}{\varepsilon\varepsilon_0}w_p \Rightarrow n_D w_n = n_A w_p$$

same condition as equal areas under $\rho(x)$

So $$\Delta V = \frac{e}{2\varepsilon\varepsilon_0}\left(n_D w_n^2 + n_A\left(\frac{n_D}{n_A}\right)^2 w_n^2\right)$$

$$= \frac{e}{2\varepsilon\varepsilon_0}\,\frac{n_D(n_A+n_D)}{n_A}\,w_n^2$$

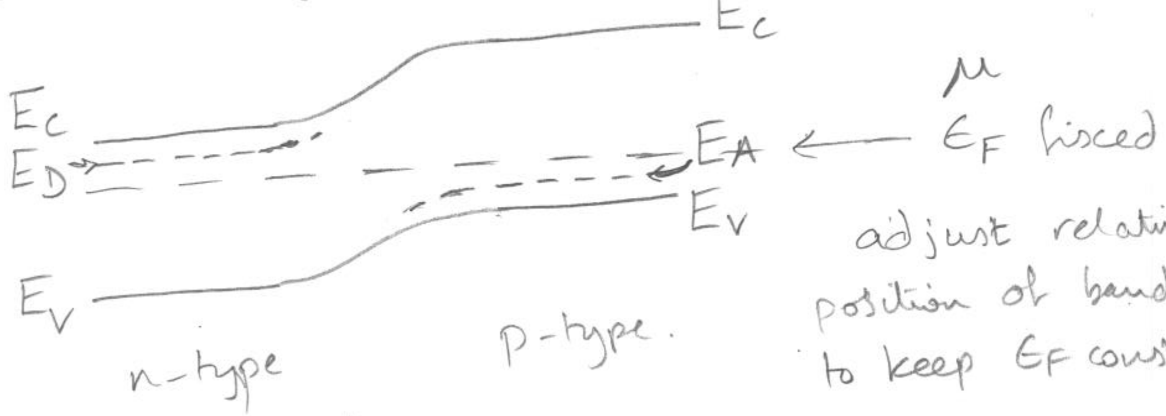
$$\Rightarrow w_n = \left(\frac{2\varepsilon\varepsilon_0 n_A \Delta V}{e n_D(n_A+n_D)}\right)^{1/2}$$ and analogous expression for w_p.

So there is a direct coupling between the potential drop across the junction and the widths of the depletion zone on the two sides.

9.6 Zero Bias

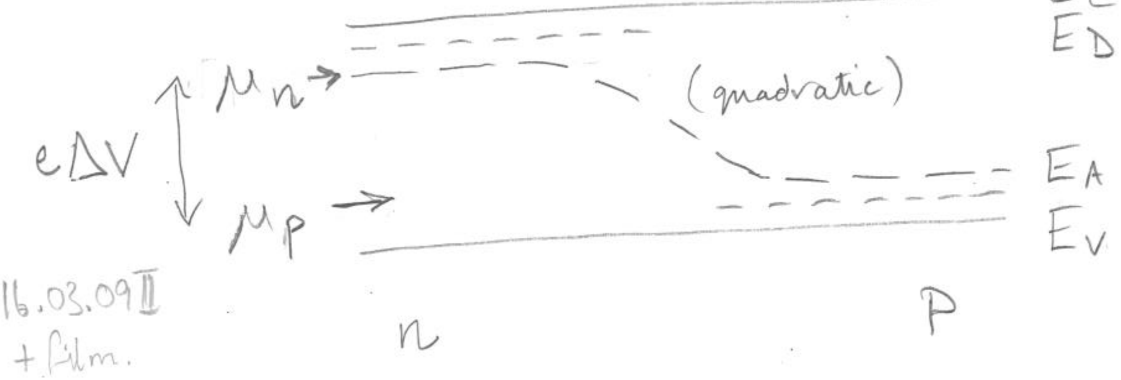
If nothing is connected to the junction, the potential difference across it will be set by the change in chemical potential (Fermi level) caused by the doping.

Two analogous pictures:



adjust relative position of bands to keep ϵ_F const.

or alternatively:



bend Fermi level to match up bands.

$\mu_n - \mu_p = e\Delta V$

16.03.09 II
+ film.

So long as $n_D \gg n_i$, $n_A \gg n_i$ we can take the two limits of the sinh function for Δn:

$$n_D = n_i e^{(\mu_n - \mu_i)/kT} \quad \text{n-type.}$$

$$n_A = n_i e^{-(\mu_p - \mu_i)/kT} \quad \text{p-type.}$$

$$n_D n_A = n_i^2 e^{(\mu_n - \mu_p)/kT} = n_i^2 e^{e\Delta V/kT}$$

$$\Rightarrow e\Delta V = k_B T \ln \frac{n_A n_D}{n_i^2}$$

12.03.07 I

This allows us to calculate ΔV, hence $w_n\, w_p$

Typical numbers for Si at 300K.

$$\left.\begin{array}{l} n_i^2 = 2.1 \times 10^{19}\ \text{cm}^{-3} \\ \text{1 ppm: } n_A = n_D = 10^{16}\ \text{cm}^{-3} \leftarrow \text{1 ppm } (10^{-6} \text{ fraction}) \\ \qquad (10^{22}\ \text{m}^{-3}) \\ k_B T = 25\ \text{meV} \end{array}\right\} \Delta V = 0.73\ \text{eV}$$

$$\left.\begin{array}{l} \varepsilon_0 = 8.9 \times 10^{-12} \\ \varepsilon = 11.7 \\ e = 1.6 \times 10^{-19}\ \text{C} \end{array}\right\} w_n = w_p = 2.2 \times 10^{-7}\ \text{m} = \underline{220\ \text{nm}}$$

Typical depletion region is 200nm in size.

→ Increase doping to 1 in 10^5; ΔV hardly changes

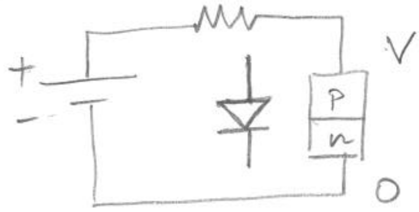
W becomes smaller as square root → 70nm

These are good numbers for engineering: need to be able to construct devices with junctions that are well-defined (flat) on this scale.

10.03.08 II

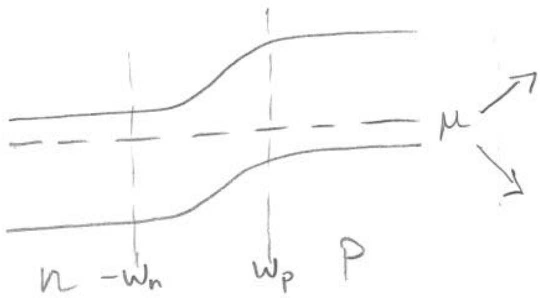
9.7 Applied Field.

If we create a circuit containing a p-n junction. Apply voltage V to "p" side



$V > 0$ Forward bias
$V < 0$ Reverse bias

What happens to the band alignment?
Voltage drop appears <u>entirely</u> at the D.Z. because both doped regions conduct well; DZ is insulating.



increases when $V < 0$

decreases when $V > 0$

NB: $\epsilon(k)$ is electron energy, opposite sign to potential V.

First thing that happens is the DZ width changes in response to the applied potential.

$$W_n = \left(\frac{2 \varepsilon \varepsilon_0 n_A (\Delta V - V)}{e n_D (n_A + n_D)} \right)^{1/2}$$

Forward bias, $V > 0$, DZ gets smaller
Reverse $<$ larger

This turns out to be relatively unimportant for current unless the DZ touches the walls. Reverse breakdown would result.

The key to understanding the p-n junction is to examine the current, which is rather subtle, but changes exponentially with V.

13.03.06 I

Units: $\sigma = n . e . \mu .$

$cm^{-3} . C . \frac{cm^2}{Vs} = (\Omega cm)^{-1}$

9.4 Conductivity.

We saw earlier that the electrical conductivity of a metal depends on carrier density:

$$\sigma = \frac{n e^2 \tau}{m}$$

This works for semiconductors too, but we usually separate electrons and holes.

$$\sigma = ne\underbrace{\left(\frac{e\tau_e}{m_e}\right)}_{\mu_e} + pe\underbrace{\left(\frac{e\tau_h}{m_h}\right)}_{\mu_h}$$

μ_h = hole <u>mobility</u>

Do not confuse mobility with Fermi level!

The mobility [cm^2/Vs] is the carrier velocity per unit electric field: $\mu = |v|/E$

It is mainly determined by τ = lifetime

Purest materials have highest mobility:

Si	1300 cm^2/Vs
Ge	3600 cm^2/Vs
GaAs	8000 cm^2/Vs
PbTe	5×10^6 at low temperature.

$(\Omega cm)^{-1}$

$\overset{10^{10}}{}\ \overset{1.6\times10^{-19}}{}$

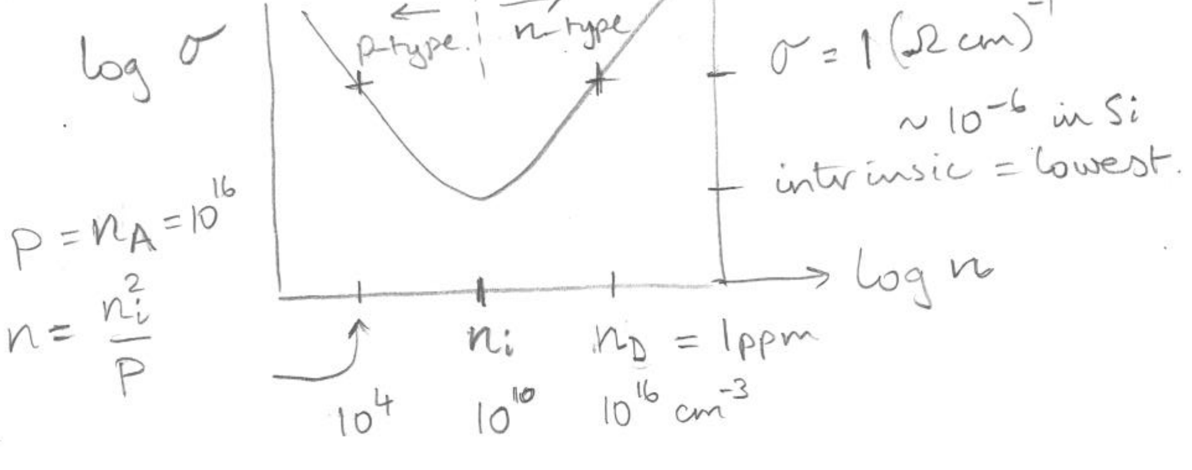
$\sigma_i = n e \mu_e \simeq 10^{-6}$ at RT.

$\Rightarrow \rho_i = 10^6\ \Omega cm$

Record is 10^7 in GaAs heterostructures used for Quantum Hall-effect devices. (2D)

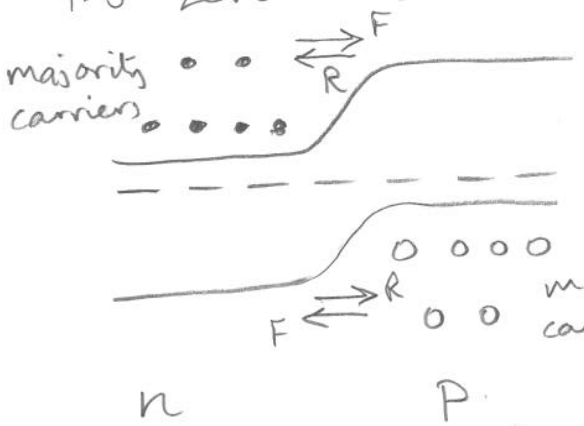
If $\mu_e \simeq \mu_h$ then $\sigma = (n+p)e\mu_e$

so this is an easy way to estimate n:



(6.03.07)

9.8 Zero-bias Currents.



Electron currents
→ forward ← reverse.

Hole currents
← forward → reverse.

All hole contributions will contribute just the same as the electrons, so we ignore for now.

Rich set of names used for two currents:

Forward, j_F	Reverse, j_R.
Uphill	Downhill
Diffusion	Field / drift
→ Recombination	→ Generation

The last two names are the most important

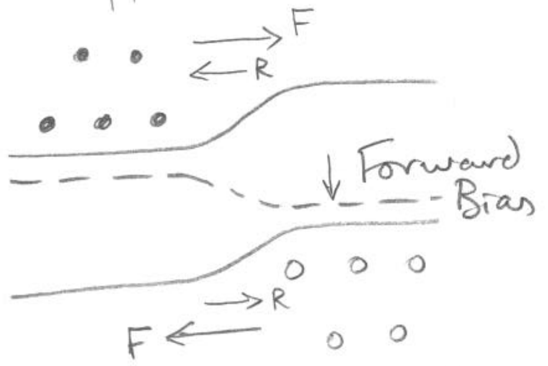
By definition, zero bias ⇒ thermal equilibrium

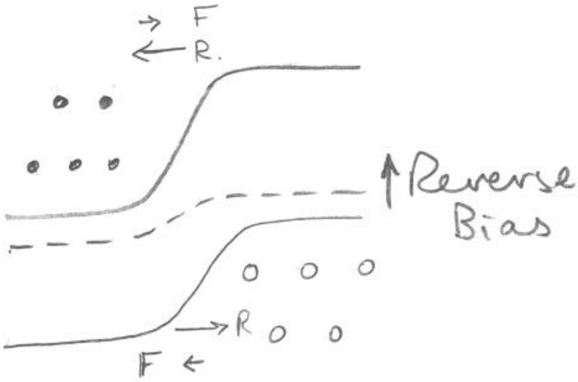
⇒ $j_F = j_R$ [or electron waits for hole to climb]

"Recombination" current is due to the outliers of the FD distribution getting enough energy to climb the hill. Many carriers because of doping; few with $\epsilon > (\Delta V - V)$, but V-dependent. Called "recombination" because when they reach the p-side, they recombine with a hole, immediately emitting a photon.

"Generation" current is spontaneous creation of minority carriers on the opposite side. Once created, they are immediately attracted by the field to the other side of the junction. Depends only on E_g, so independent of V, ΔV.

9.9 Applied field.





23.03.09 I Forward current changes with field.
Reverse current does not (Generation)

Estimate Generation current first: Assume:

i) only carriers generated within one diffusion length at the edge of the DZ contribute, $x < L$
Inside DZ: more carriers but $w \ll L$.
Outside $x = L$: scattered before arrival at edge.

ii) Generation rate $= n_{\text{minority}} / \tau$ (lifetime)
τ = lifetime before electron finds hole and recombines.

$$j_R \text{ (current/area)} = e \underbrace{L n_{\text{min}}}_{\text{no. per unit area}} / \tau \qquad \left[= e v_F n_{\text{min}} \right] \quad 10^{-19} \; 10^{4} \; 10^{10}$$

$$n_{\text{min}} = \frac{n_i^2}{n_A}$$ for electrons on p-side. Small ~ 10^4 cm^{-3} 10^{10} m^{-3}

L = mean free path
τ = survival time
$\frac{L}{\tau}$ = velocity $= v_F \sim 10^4$ m/s
so $j_R \sim 10^{-5} \text{ A m}^{-2} = 10^{-11} \text{ A mm}^{-2}$

12.03.07 II Forward current varies exponentially:

$$j_F = A e^{+eV/kT}$$

$j_F(V=0) = A = j_R$, so j_R determines j_F

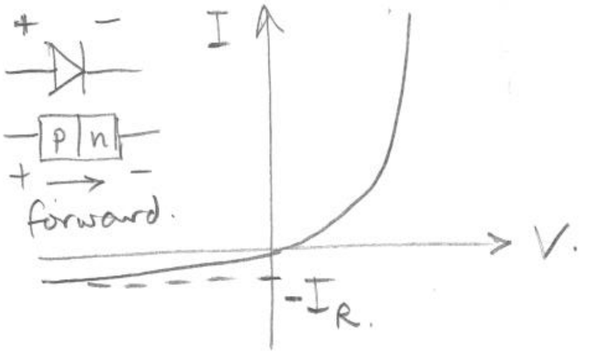
Total current $j = (j_F - j_R) = j_R (e^{eV/kT} - 1)$

Electrons and holes together:

$$j_{tot} = (j_R^e + j_R^h)(e^{eV/kT} - 1)$$

11.03.08
(fire alarm)

Diode characteristic:



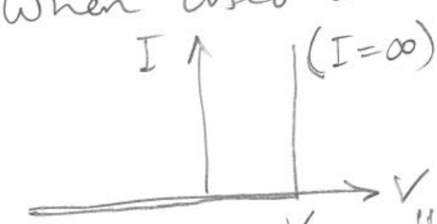
Only one number sets everything, the reverse-bias current:

j_R theoretical density

I_R = practical measurement $\propto$ Area of device.

This shows the rectifier property: current only flows in one direction.

When used in circuits, we use approx form:



V_D = "diode drop"

The diode drop, V_D is an important design parameter.

Useful device needs $j_R \sim 10^{-11}\,A\,mm^{-2}$, unmeasurably small

So to get a current of 10 mA, we need:

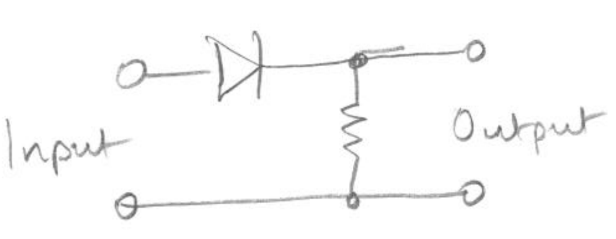
$$\frac{eV}{kT} = \ln\left(\frac{10^{-2}A}{10^{-11}A}\right) = 20.7 \text{ or } V = 0.52V$$

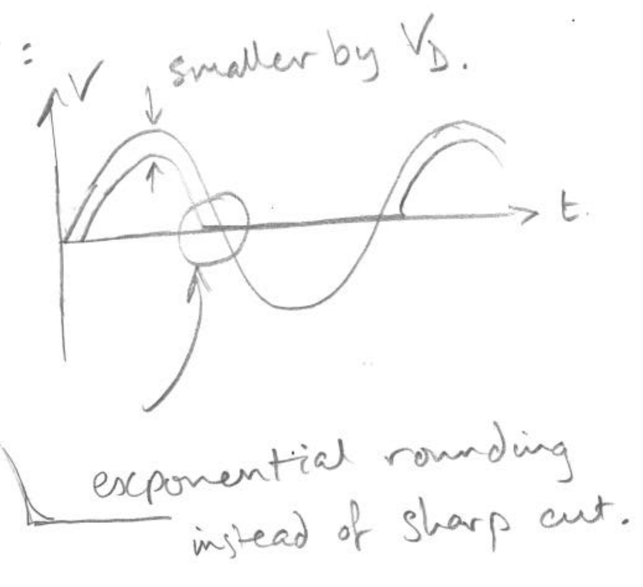
But to increase current to 100mA needs $V = 0.58V$

Hence the rule of thumb V_D = constant.

9.10 Circuit example

Single-wave rectifier:





9.11 Field-effect Transistor (FET)

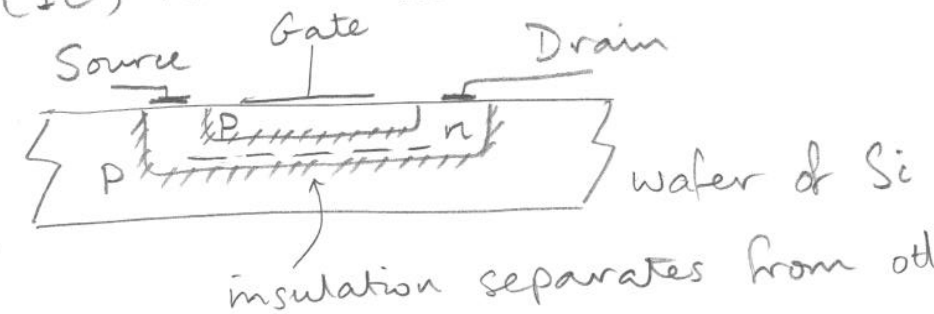
We have learned the two important properties of the p-n junction:

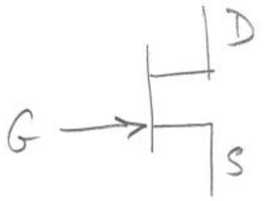
i) variable width depletion zone (insulating)
ii) exponential I-V characteristic.

In the 1950's there was a big push to create a "3-terminal device" with "gain": low power signal on input controls much higher power on output. Solid-state amplifier!

13.03.06 II Original proposal (Shockley) was FET but they could not get it to work; instead the Bipolar Junction Transistor (BJT) came first.

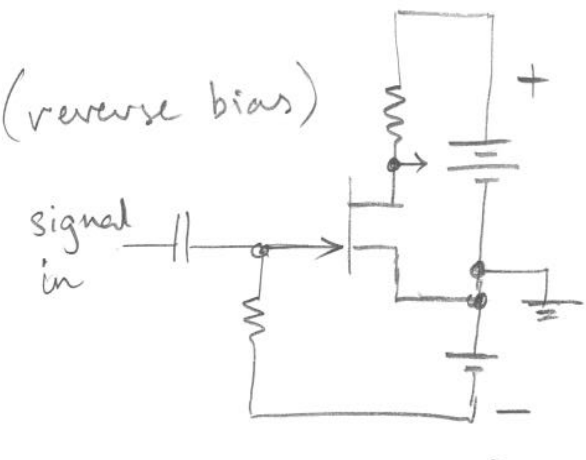
FET is easy to make using Integrated Circuit (IC) technology. ~10^6 to 10^7 per chip wired together.





symbol for n-channel JFET.

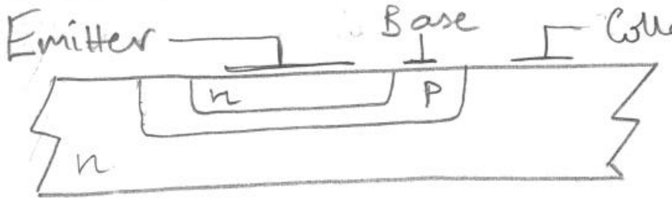
Gate is "biased" negative (reverse bias)
Thickness of channel is changed by the width of the depletion layer, eventually "pinches off" the drain current.

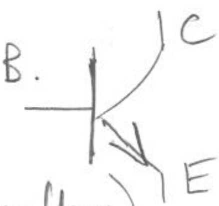


$$I_D = I_0 (V_{DS} - V_P)^2$$

9.12 Bipolar Junction Transistor (BJT)

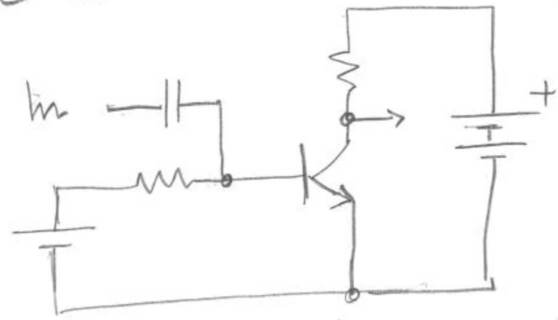
Most of the semiconductor industry used these until about 1990's. Still used a lot.



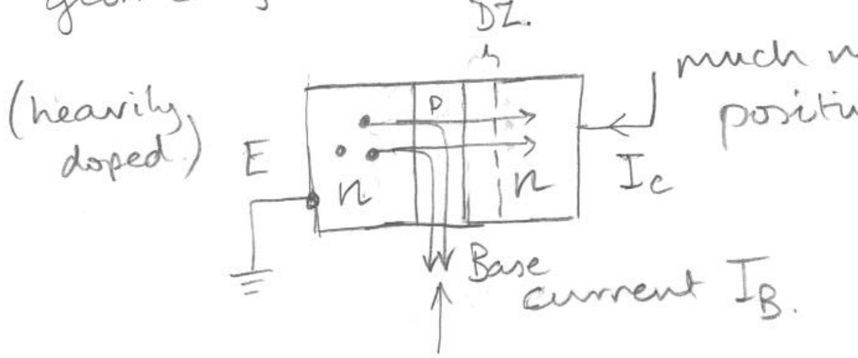


Symbol for npn BJT

BE is forward biased (current flows)
BC is reverse biased (no direct current)



Amplification of currents results from the geometry of the base region:



I_C

current I_B.

$\frac{I_C}{I_B} \simeq$ constant $\beta \sim 100$

set by geometry.

E is more heavily doped than B, so I_B is mostly electrons (not holes). The base is wide and thin so the current travels sideways. The electrons "see" the positively charged collector and are picked off in a large proportion.

Since the proportion is fixed (by geometry) the device is a (linear) current amplifier.

Photovoltaic

9.13 Solar Cells + LED's [Kittel Ch 17]

Photons are involved with all interband processes:

photon $\rightleftharpoons$ electron + hole pair

Normally there is thermal equilibrium with the light field, but we can attach an external circuit to tip the equilibrium one way or the other. Creates or absorbs photons.

Better to use direct-gap semiconductors to avoid waiting for the phonons to couple.

Wider gap than Si & Ge couples to visible light:

GaAs (1.4eV) or AlAs (2.2eV) or alloy mixture.

Lifetime of carriers, once generated, is quite long:

$\tau \sim 10^{-7}$ sec

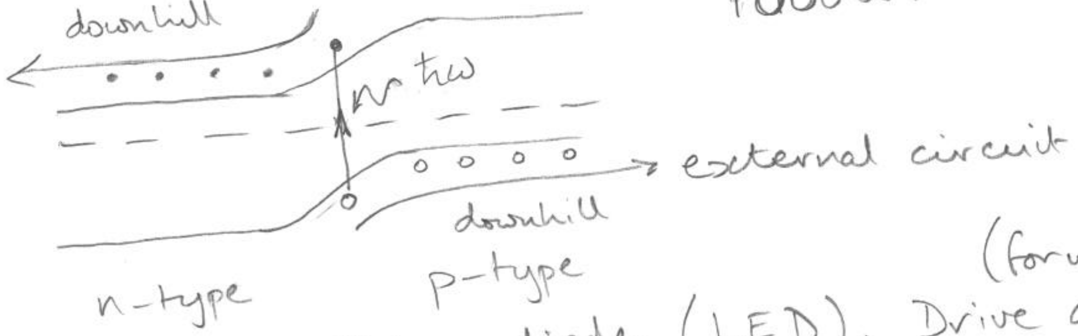
Diffusion length before recombination also long:

$L \sim 10\,\mu m \gg w$ (depletion zone)

So there is plenty of opportunity for carriers to cross the DZ region, if the field is favorable.

Solar Cell (photovoltaic effect)

$1000\,W\,m^{-2}$ sun?



23.03.09 II

Light-emitting diode (LED). Drive current (forward) increases band offset as needed.