Fundamentals of Electronic Components

Universidade do Algarve

2^{o} semester 2002-2003

curso ESC

Peter Stallinga

Contents

Faro, 20 February 2003
Peter Stallinga

Books / Literature

Sze : S.M. Sze, "Physics of semiconductor Devices".
* 2nd ed. Wiley Interscience, ISBN 0-471-05661-8

Shur : M. Shur, "Physics of semiconductor Devices".
Prentice Hall, ISBN 0-13-666496-2

Singh : J. Singh, "Semiconductor Devices, Basic Principles".
Wiley Interscience, ISBN 0-471-36245-X

Blakemore : J.S. Blakemore, "solid state physics", 2nd ed.
* ISBN 0521 31391 0, Saunders

Kittel : C. Kittel, "Introduction to solid state physics",
4th ed. Wiley, ISBN 0-471-49021-0

A & M : N.W. Ashcroft / N.D. Mermin, "Solid State physics", intern.
* ed., Saunders ISBN 0-03-049346-3

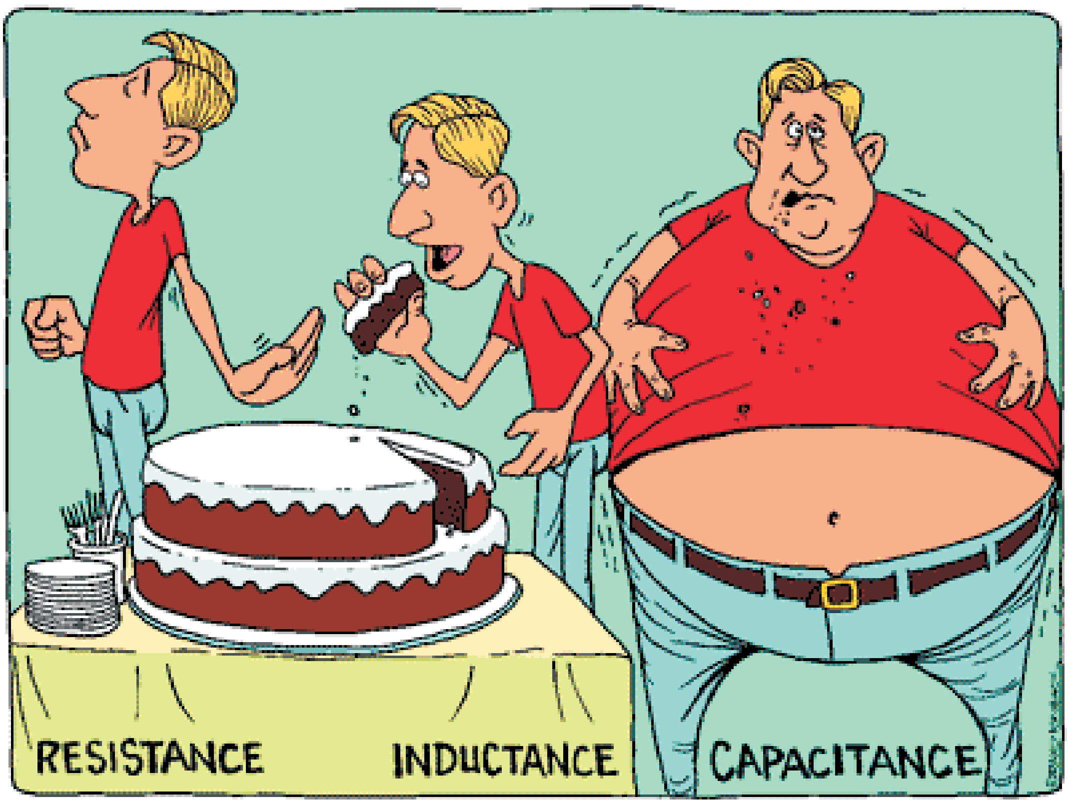
Fromhold "Quantum Mechanics for Applied Physics and Engineering"
A.T. Fromhold J.R.

http://www.ualg.pt/fct/adeec/optoel/theory/
P.S. online

For who wants to know more or doesn't understand
what I am saying.

* : exists a copy in the library

Introduction / Electronics

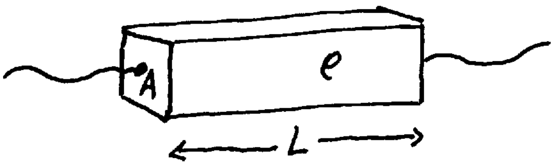


Electronics / Introduction

In our electronics lectures we learned concepts such as resistance (R), capacitance (C) and inductance (L). Although these parameters are good for explaining most phenomena in electronics, they are insufficient for explaining the workings of electronic components on a lower level. As an example, resistance is a device property. What is more relevant in the scope of the current lecture is the material properties. All material properties are intrinsic to the material, meaning that they do not depend on the size of the device made from it.

In this way, a material has a resistivity ρ, while a device made from it has a resistance, R. One can be calculated from the other if the device dimensions are known.

As an example: consider a bar of material with resistivity ρ and dimensions: A: area of cross-section and L: length of bar



L : length of bar

A : cross-section of bar

In this case $R = \rho \cdot \frac{L}{A}$ (I)

note the units : $[R] = \Omega$

$[\rho] = \Omega \, m$ *

$[L] = m$

$[A] = m^2$

The resistance can be measured by measuring the current (I) and the voltage (V) over the sample and using ohm's law:

$$R = \frac{V}{I} \qquad \left(\text{units: } \Omega = \frac{V}{A}\right) \qquad (\text{II})$$

Instead of resistivity, we can also work with conductivity $\sigma = \frac{1}{\rho}$, which is the material equivalent of the device conductance $G = \frac{1}{R}$:

$$G = \frac{I}{V} \qquad \left(\text{units : } \overset{\text{siemens}}{S} = \frac{A}{V} = \frac{1}{\Omega}\right)$$

$$= \sigma \frac{A}{L}$$

These equations still contain (macroscopic)

---

* in literature you normally find Ωcm

device parameters such as current (I) and voltage (V). In a more general description we will only talk about current density $j = \frac{I}{A}$, which is the current per square meter of area. In the same way, it is more convenient to talk about the electric field, which is equal to the voltage-drop per meter:

$$E = \frac{V}{L}$$

with this, Ohm's law $I = \frac{V}{R}$ becomes (I+II)

$$A \cdot j = \frac{E \cdot L}{\rho \cdot L/A}$$

$$j = \frac{E}{\rho} = \sigma E \qquad \text{(III)}$$

in other words:

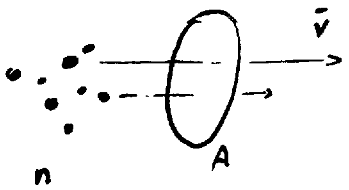
"current density is proportional to the electric field. The proportionality factor is called conductivity".

Quietly we have entered into the field of physics. We started with the electronics sentence

"current is proportional to voltage - The proportionality factor is called conductance".

From now on we will only talk about material properties.

Since current is "charge per second passing through the device" (ampere is coulomb per second, $A = \frac{C}{s}$), current density must be "charge per second passing through an area A":



As we know, the charges passing are electrons; each electron carries a charge of $-q$ ($q = 1.6\ 10^{-19}\,C$) and the above sentence translates to "current density is number of electrons passing per second through the area A, multiplied by their charge $-q$"

The number of electrons passing point A depends on

- the density of electrons (n)
  (number of electrons per cubic meter)
- their average speed ($\bar{v}$)

(look at it as a highway with cars with average speed $\bar{v}$ and density n. The total number of cars passing per hours is $n \times \bar{v}$)

$$j = q \cdot n \cdot \bar{v} \qquad \text{(IV)}$$

In electronic materials the average speed of electrons is proportional to the electric field (E). We will later see why.

The proportionality is called mobility (μ_n):

$$\bar{v} = \mu_n E$$

(The unit of this parameter must be "speed per field", or $\frac{m}{s}$ per $\frac{V}{m}$, thus

$$[\mu_n] = m^2/Vs$$)

Putting this in equation (IV) gives:

$$\boxed{j = q \cdot \mu_n \cdot n \cdot E} \qquad (V)$$

Which summarizes the material propertyes of electronic materials:

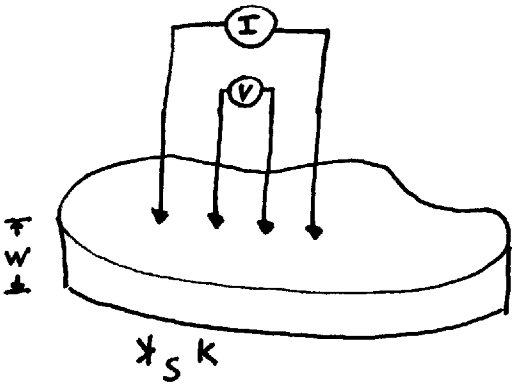
The current density is is proportional to

1) The charge density $q \cdot n$
2) The charge mobility μ_n
3) The electric field E

How to measure these material properties n and μ_n?

Four-point-probe:

- current source on outside two contacts
- measure voltage on inside two contacts

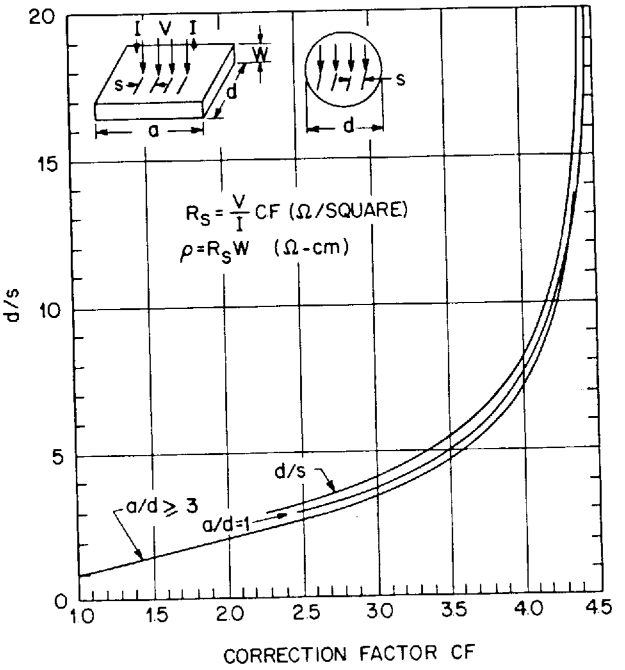


Then

$$\rho = R_s W$$

$$R_s = \frac{V}{I} \cdot CF$$

- correct for device dimensions : correction factor CF

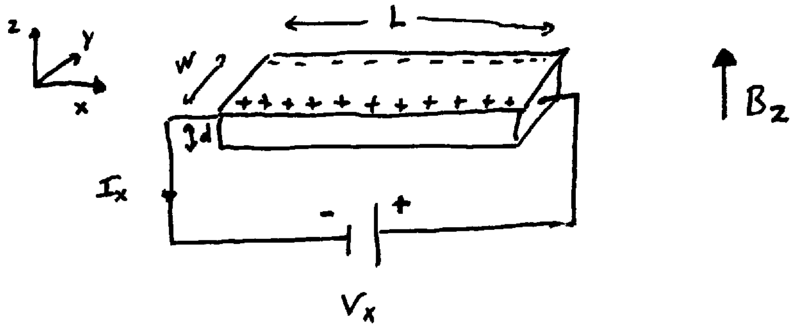


see p. 31 of Sze

This will measure the resistivity / conductivity of the material. ($\sigma = q \mu_n n$). The problem is that we would like to know n and μ_n separately. As we will see later, n can be measured in Schottky diodes and p-n junctions. μ_n can be measured in a Hall measurement:

(see ps online) :

In a Hall experiment the sample is placed in an electric field (external voltage) <u>and</u> a magnetic field. The electric field causes a current I_x.



The charges moving feel the magnetic field and Lorentz forces ($F_L = B \times I$) and they are pushed to the side of the sample, thereby building up charge at the edges. These charges cause a second electric field along y which compensate the magnetic forces. In steady state the forces balance: $F_y = 0 = F_y^B + F_y^E$ VI

$$F_y^B = q \cdot B_z \cdot v_x \qquad \text{VII}$$

$$F_y^E = -q E_y = -q \frac{V_y}{W} \qquad \text{VIII}$$

Remember

$$j_x = q \cdot n \cdot v_x$$

$$I_x = W \cdot d \cdot j_x$$

$$= q W d n v_x \qquad \text{IX}$$

and $v_x = \mu_n \cdot E_x = \mu_n \dfrac{V_x}{L}$ X

combining VI .. X :

$$\mu_n = \left(\frac{V_y}{V_x}\right) \cdot \left(\frac{L}{W}\right) \cdot \left(\frac{1}{B_z}\right)$$

$\Rightarrow$ measures μ_n without having to know n!

$q = 1.6\ 10^{-19}$ C

B_z = magn. field along z

v_x = average speed of charges along x

E_y = electric field along y

V_y = voltage along y

V_x = V along x

W = width of sample

d = thickness of sample

L = length of sample

Summarizing :

- Resistance (R) and conductance ($G = 1/R$) of the devices are related the resistivity (ρ) and conductivity ($\sigma = 1/\rho$) of the materials via the device dimensions (length L, width W, thickness d, area A, etc).

- The conductivity is linearly proportional to
  - density of charges (n)
  - their mobility (μ)

- The current density (j) is linearly proportional to
  - the conductivity (σ)
  - the electric field ($E = \frac{V}{L}$)

- The current density is proportional to
  - the density of charges (n)
  - their average speed (v), which is proportional to E:

    $v = \mu_n \cdot E$

..... it all boils down to calculating the number of electrons and their speed, etc.

DRUDE MODEL

simple model of conduction (beginning of 20 th century) (ch 1 of A&M)

"conductivity is proportional to density of free electrons". Drude

- solid is collection of atoms
- atom has electrons in shells core / valence
- core electrons do nothing
- valence electrons contribute to conduction
- these free electrons scatter from nuclei and reach a terminal velocity in a field (like raindrops in air)

Example : aluminum

13 electrons :
1S = 2 closed shell
2S = 2 closed shell
2P = 6 closed shell
3S = 2 "free electron"
3P = 1 "free electron"

density of aluminum solid: 2700 kg m^{-3}

1 atom of aluminum weighs : 26.98 u (g/mol)

1 mol = N_A atoms = $6.023\ 10^{23}$

density of free electrons : $3\times \dfrac{2700\ \text{kg m}^{-3}}{26.98\ 10^{-3}\ (\text{kg}\cdot\text{mol}^{-1})} \times 6.023\ 10^{23}\ /\text{mol}$

$= 1.8\ 10^{29}\ /m^3$

The link with classical mechanical physics

Kinetic Energy of free particle

$$E_k = \frac{1}{2} m v^2$$

example, free electron in space.

$$E_k = \frac{1}{2} m_e v^2$$

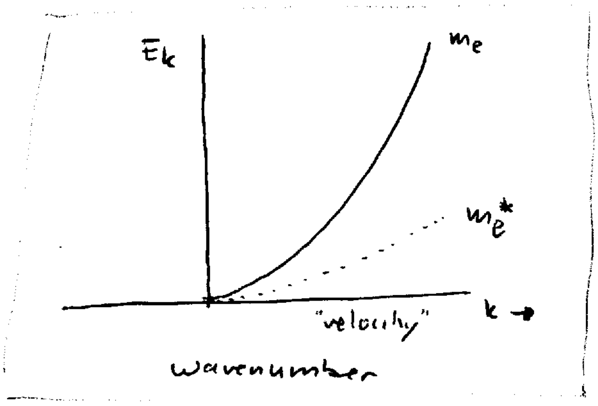
$m_e = 9.1 \; 10^{-31}$ kg. Energy of electron at 100 km/h

$= \frac{1}{2} \cdot (9.1 \; 10^{-31} \text{kg}) (27.8 \frac{m}{s})^2 = 3.5 \; 10^{-28}$ J

(compare to car 0.39 MJ)

It is also possible to express it in <u>momentum</u> (mass x velocity, $p = m \cdot v$) $p = \hbar k$ (quantum mechanics)

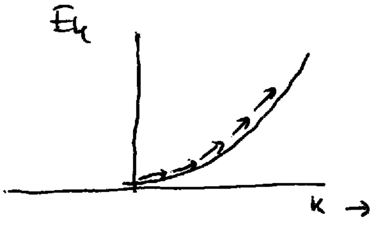
$$E_k = \frac{1}{2}\hbar^2 k^2 / m_e \qquad (k = 2\pi/\lambda) \qquad \hbar = \frac{h}{2\pi}$$



Curve also shows how it would be if the (effective) mass were higher ....

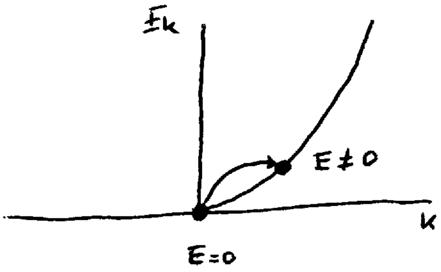
When placed in an electric field ($E = \frac{V}{d}$), the electron will accelerate according to Newtons $F = m \cdot a$.

$$\frac{dp}{dt} = qE = \left(\frac{dk}{dt}\right)$$



electron accelerates and gains kinetic energy.

In a solid state crystal this is different. The strong interaction with the host lattice makes that the electron accelerates to a certain point and then saturates. Moreover, the time it takes to reach this value is very small compared to other processes taking place.

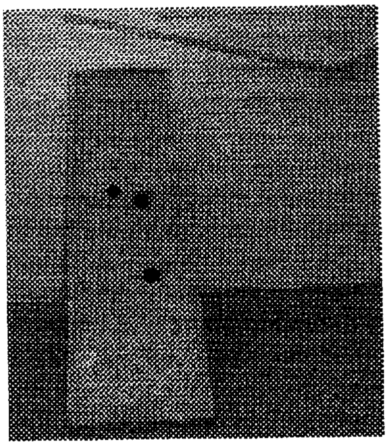


Another important aspect is that the levels are quantized, meaning "not continuous" in a crystal.

This is caused by the fact that we are dealing with a crystal as we will see later.

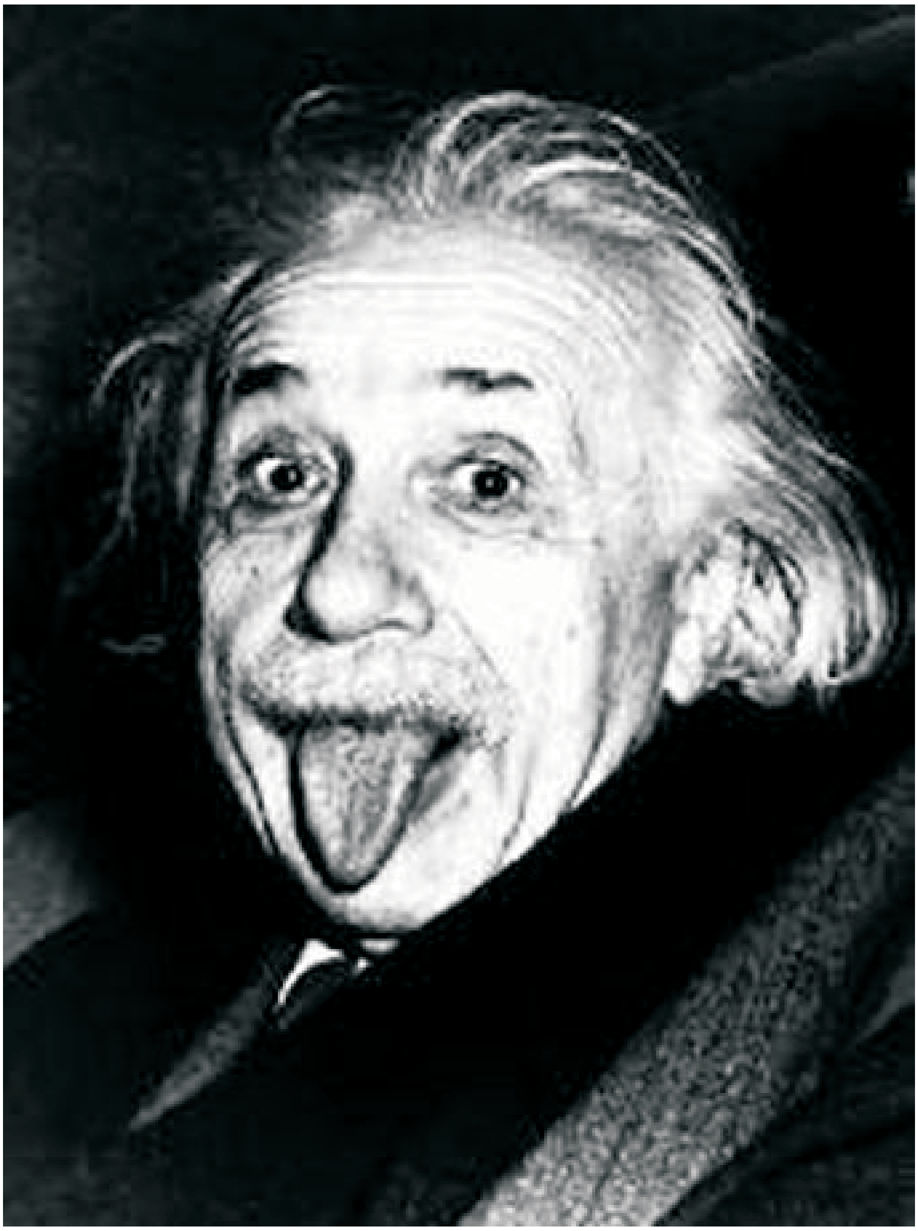
This is like raindrops in air. Even if the gravity force keeps working on them, the raindrops no longer accelerate

Or like the balls in the contraption here on the right: Their velocity is constant due to strong interaction with the lattice.



www.phys.unt.edu/~klittler/demo_room/e&m_demos/5d10_40.html

Physics



Basic equations for device operation (Sze, p50)

Maxwell equations:

$$\nabla \times \bar{E} = -\frac{\partial \bar{B}}{\partial t} \qquad (M1)$$

$$\nabla \times \bar{H} = \frac{\partial \bar{D}}{\partial t} + \bar{J} \qquad (M2)$$

$$\nabla \cdot \bar{D} = \rho \qquad (M3)$$

$$\nabla \cdot \bar{B} = 0 \qquad (M4)$$

$$\bar{D} = \int_{-\infty}^{t} \varepsilon_s (t-t') \, \bar{E}(t') \, dt' \qquad (M5)$$

$$\bar{D} = \varepsilon \bar{E} \, , \quad \bar{B} = \mu_0 \bar{H}$$

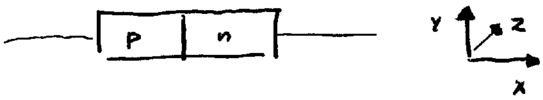
we will use "steady state" (no changes in time, $\frac{\partial}{dt} = 0$) and no external magnetic field ($\bar{B} = 0$). No need to remember all. The only equation needed is M3. In the case of planar symmetry, $\rho(x,y,z) = \rho(x)$ we get

$$\boxed{\frac{dE_x}{dx} = \frac{1}{\varepsilon} \rho(x)} \qquad (P1)$$

thus

$$\boxed{\frac{d^2 V(x)}{dx^2} = \frac{1}{\varepsilon} \rho(x)} \qquad (P2)$$

which are called Poisson's equations. Most devices are planar, we imagine them stretched to infinity in two dimensions, for example the p-n diode:



BAND STRUCTURE

Band structure is the energy distribution of electrons in the presence of a crystal. Without the crystal, the electrons would only have their kinetic energy.
The interactions with the crystal (positively charged regular network) will change their energy by also adding a potential energy.
"Like a person is attracted by the earth in it's gravitanional field, so the electron is attracted by the positively charged nucleus"

Placing the free electron in a crystal will change its energy dispersion relation

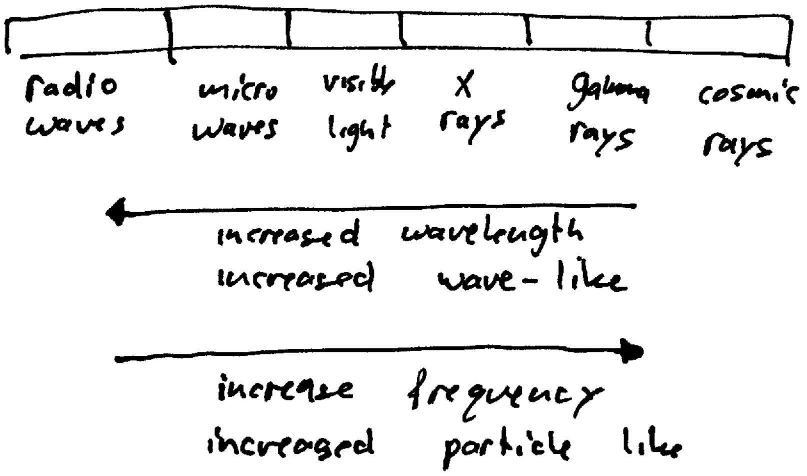
$$E_k = \frac{\hbar^2 k^2}{2m_e}$$

we will
- still assume electrons are individual (don't feel eachother)
- start feeling the crystal: a lattice with periodically placed, positively charged nuclei.

Quantum Mechanics: De Broglie:

particle with momentum p has wavelength
$\lambda = \frac{h}{p}$ (de Broglie relation)

we already know this from light. The Huygens (older) theory: light has wavelength, Later: it can also be seen as a particle; a "foton", especially for shorter wavelengths



New idea: every object has wave length. $\lambda = \frac{h}{p}$

Example: person with weight of 80 km walks with a speed of 6 km/h

$$p = m \cdot v = 133 \text{ kg m s}^{-1}$$
$$\lambda = \frac{h}{p} = 5 \cdot 10^{-36} \text{ m}$$

check units: $[h] = Js$, note einstein $E = mc^2$
$J = kg\, m^2 s^{-2}$

$$[h] = kg \frac{m^2}{s^2} s$$
$$[\lambda] = kg \frac{m^2}{s^2} \cdot s \,/\, kg\, m s^{-1} = m$$ correct.

Example 2 : electron with speed 1% of speed of light ($c = 10^8$ m/s)

$$p = m_e \cdot \frac{1}{100} \cdot c$$

$$= 9.1\ 10^{-31}\,\mathrm{kg} \cdot 0.01 \cdot 3 \cdot 10^8\ \mathrm{m/s} = 2.73\ 10^{-24}\ \mathrm{kg\,m/s}$$

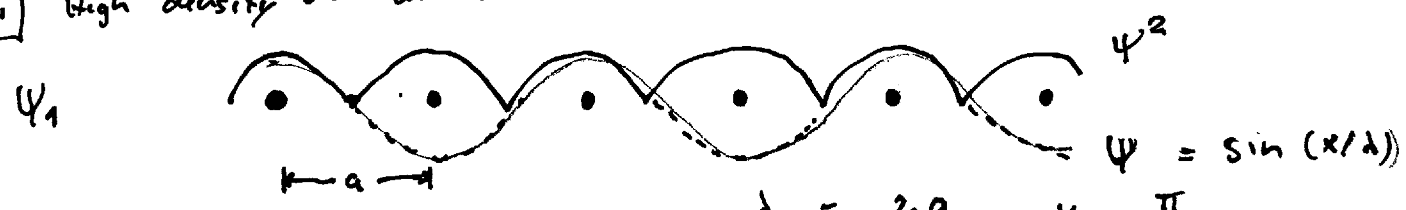
$$\lambda = \frac{h}{p} = \frac{6.6\ 10^{-34}\ \mathrm{Js}}{2.73\ 10^{-24}\ \mathrm{kg\,m/s}} = 2.4\ 10^{-10}\ \mathrm{m} = 2.4\ \mathring{A}$$

This is comparable to the distance between two atoms in a crystal, as we will see later. Thus we can expect some effects.

Quantum Mechanics :
probability of finding particle at $\bar{x}$ is proportional to $|\psi(\bar{x})|^2$

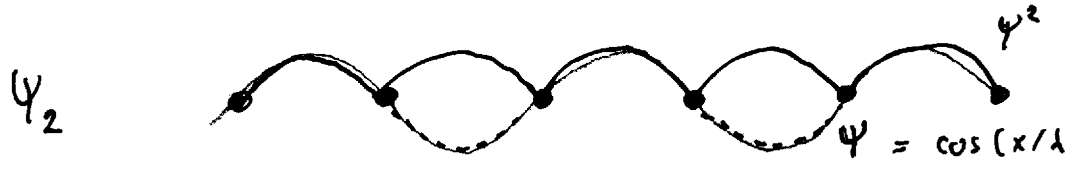
If we have an electron with wavelength equal to the lattice periodicity; there are still two possibilities (other possibilities are superposition of these two)

$\boxed{1:}$ High density on atoms



$\lambda_1 = 2 \cdot a$, $k_1 = \frac{\pi}{a}$

$\boxed{2:}$ high density between atoms



$\lambda_1 = 2 \cdot a$, $k_1 = \frac{\pi}{a}$

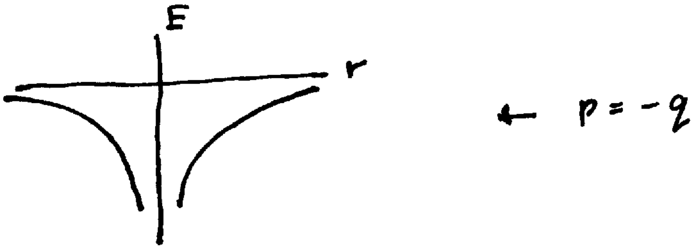
a is interatomic distance

For the first possibility the electron has higher wave density at the positively charged nuclei, or, in other words, it spends more time close to the nuclei.

The potential energy of a charge p at a distance of a second charge q is given by Coulombs equation

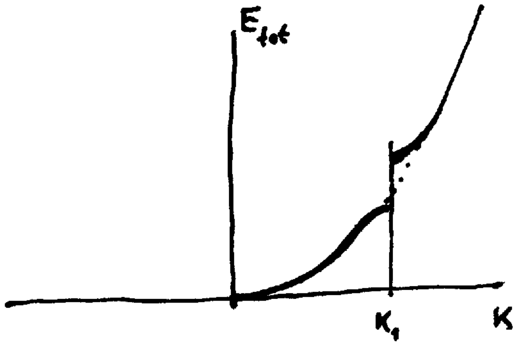
$$E_{coulomb} = \frac{p \cdot q}{4\pi\varepsilon r^2}$$

(see http://hyperphysics.phy-astr.gsu.edu/hbase/electric/elefor.html)



This means that oppositely charged particles attract eachother; the energy is lowest when they are close together.

The wavefunction with high density of electron at the nuclei has lower energy:
the total energy (kinetic + potential) of Ψ_1 is therefore lower than of Ψ_2, although they have the same wavelength (momentum)



k_1 : $\frac{\lambda}{2}$ = atomic distance

$k_1 = \frac{\pi}{a}$

a is intratomic distance

we have intuitively described the formation of bands.

Looking again at the picture on the previous page we see that there are certain energies that the electron cannot have when placed in the crystal.

To conclude our simple introduction to band calculations we have to make two observations

① since the lattice is periodic, we can put all the information of the bands in a limited range of k $-k_1 < k < +k_1$ called the (p 139 of A&M) first Briouillin zone

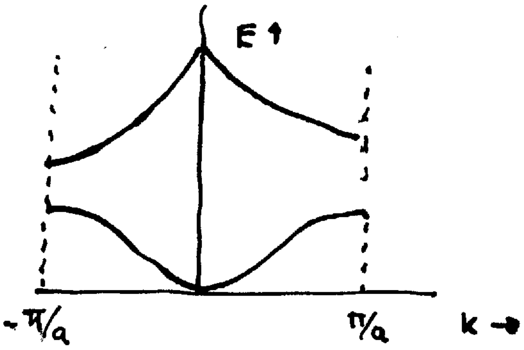
---

because the wave function

$$\psi(k) = e^{ikx}$$

is independent of a translation

$$\psi(k+K) = e^{ikR}\, e^{ikx} = 1 \cdot e^{ikx}$$



② The wave number k is quantized, meaning that it can take only discreet values.

The distance between two k-values is

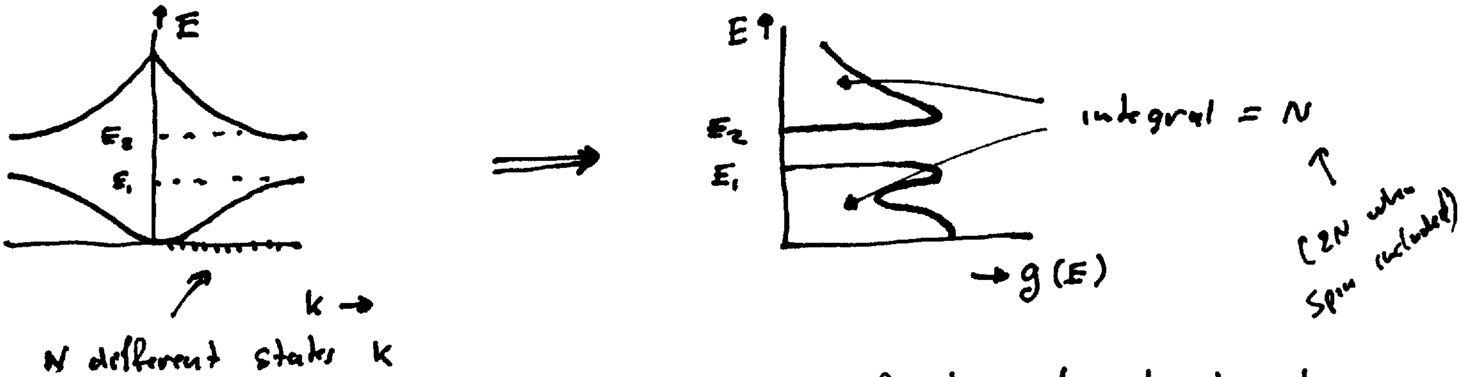
$$\Delta k = \frac{2\pi}{L}$$ in one dimension

with L the length of the crystal $(= N \cdot a)$

The number of states in a branch is therefore $\left(\frac{2\pi}{a}\right) / \left(\frac{2\pi}{L}\right) = \left(\frac{L}{a}\right) = N$, the number of atoms in the crystal!

DENSITY OF STATES

To calculate the number of states $g(E)$ at every energy E, we have to count the number of possible states there:



and we see the formation of the band structure

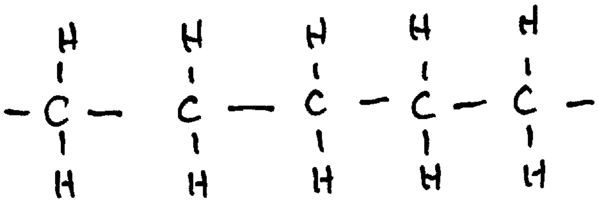
PAULI

> Quantum Mechanics : Pauli's Exclusion Principle
> every state can be occupied by only one (1) electron at the same time

with this information we can start filling the bands with our electrons.

Lets put on every lattice site a $-\overset{H}{\underset{H}{C}}-$ unit, so we get a chain

one dimensional
linear crystal
poly ethylene



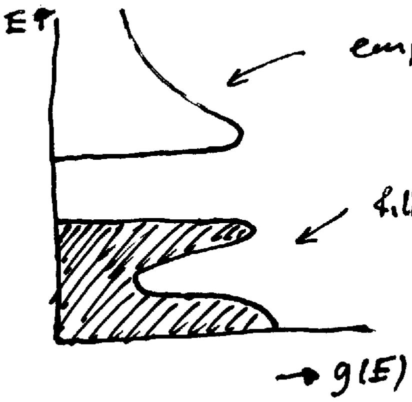
$$-\overset{\overset{H}{|}}{\underset{\underset{H}{|}}{C}}-\overset{\overset{H}{|}}{\underset{\underset{H}{|}}{C}}-\overset{\overset{H}{|}}{\underset{\underset{H}{|}}{C}}-\overset{\overset{H}{|}}{\underset{\underset{H}{|}}{C}}-\overset{\overset{H}{|}}{\underset{\underset{H}{|}}{C}}-$$

every carbon - H_2 unit contributes 2 electrons to the conduction electron gas. ("The valence of CH_2 is 2").

The total number of electrons to store in the states is therefore 2N.

Before we can start filling the states as given in the previous page, we have to remember that apart from velocity and potential energy, every electron also has spin. Therefore, every state can be occupied by <u>two</u> electrons, one with spin up and one with spin down.

To reach the lowest energy distribution we will start filling up the states with lowest energy first. We reach the following situation:



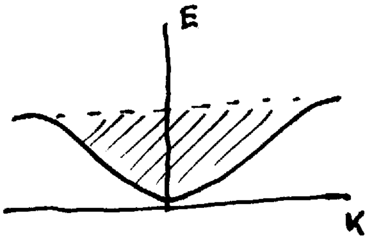
empty

filled with our 2N electrons

situation at T = 0 K

This crystal therefore does not conduct any electricity. at T=0 K. Why?

To arrive at the picture above we filled all the states from the first band, both all the states with positive k and negative k. Remember $k = \frac{p}{\hbar}$ or $k = \frac{m_e v}{\hbar}$. So, for every electron with speed v, we also have one with speed $-v$



The average speed therefore is always 0, even if we add an electric field.
$\Rightarrow$ currents are 0

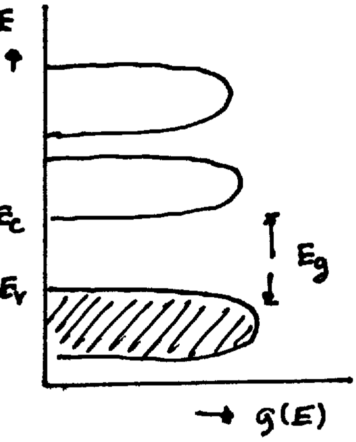
As we will see later, heating the sample will allow for electrons to jump to the next band, thereby creating free electrons and leaving behind free holes. (p 246 of Blakemore)

in reality $v_x = \frac{1}{\hbar}\frac{dE}{dk_x}$

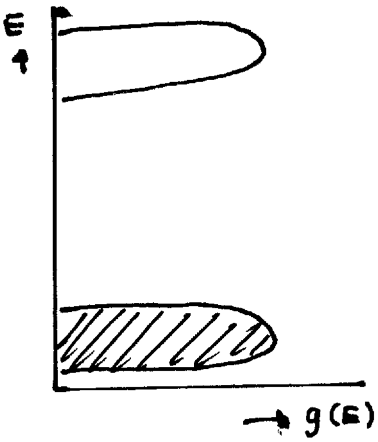
$$\frac{dk_x}{dt} = -e\frac{E_x}{\hbar}$$

INSULATORS - SEMICONDUCTORS - METALS

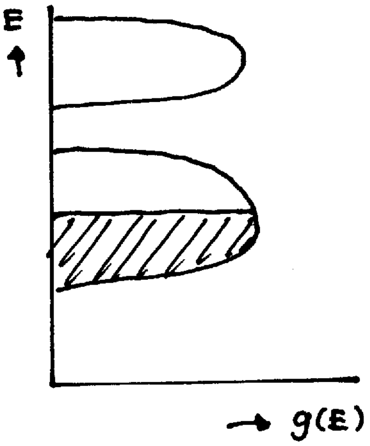
- As we have seen, in a semiconductor, all the bands are completely full or completely empty (at T= 0K) and conduction can only be induced by heating the sample. The highest occupied band is called the valence band, the top of which has energy Ev, while the first unoccupied band is the conduction band with bottom energy Ec. The difference between



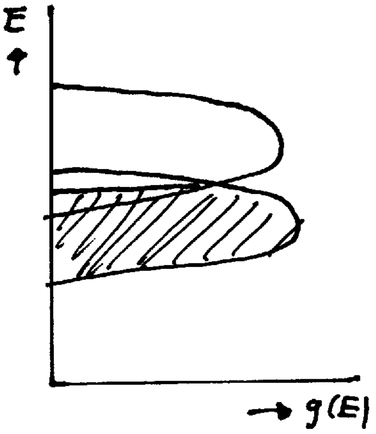
semi conductor



insulator



metal



semi metal

these two values is the forbidden energy gap, $E_g = E_c - E_v$.

with energies that the electron cannot have.

Examples :

Si : $E_g = 1.12$ eV

Ge : $E_g = 0.66$ eV

GaAs : 1.42 eV

GaN : 3.36 eV

C (diamond) : 5.47 eV

polymers : ~ 2.5 eV

at T = 300 k

1 eV = q · 1V J

= $1.6 \; 10^{-19}$ J

- when the gap is wider, the conduction can no longer be induced by temperature and we have an insulator. Examples :

SiO_2 : $E_g = 9$ eV

Si_3N_4 : $E_g = 5$ eV

Diamond is in fact also an insulator.

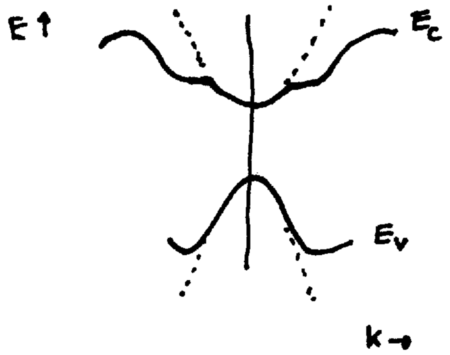
- when the distance between the bands is reduced, they might even start overlapping. This causes that, even at T = 0 K, there is room for current in the material. These materials are called semimetals.

For the above three cases, semiconductors, insulators and semimetals we always have that the number of valence electrons per atom ν is even and they can completely fill a number of bands (remember, each band has 2N states available)

- For metals, the number of electrons per atom is odd. In this case, there is no way to completely fill a band and such materials always conduct well.

EFFECTIVE MASS

In the free-electron case, the energy is given by the parabola $E_k = \frac{\hbar^2 k^2}{2m_e}$. We can now define an effective mass m_e^* that best describes the top of the valence band and the bottom of the ~~vale~~ conduction band

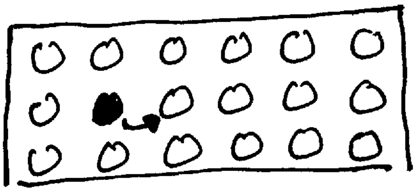


$$E_c^* = \frac{\hbar^2 k^2}{2m_e^*}$$

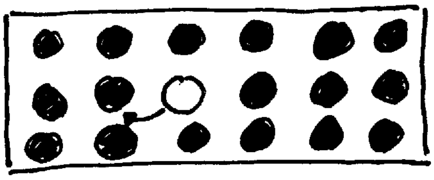
$$E_v^* = -\frac{\hbar^2 k^2}{2m_p^*}$$

ELECTRONS AND HOLES

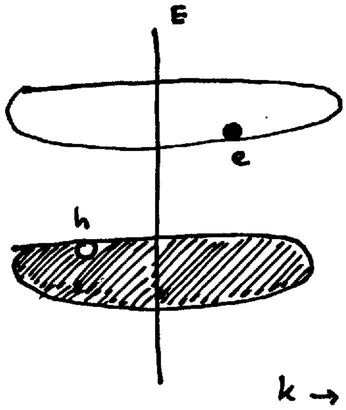
By increasing the temperature electrons can be promoted from the valence band to the conduction band. This electron is then free to assume whatever velocity and thus can contribute to conduction.
At the same time the electron leaves a hole behind in the valence band. The electron is missing from the otherwise full valence band. This leaves room for the remaining electrons in the valence band to assume an average speed ≠ 0. and thus external conduction.
In jargon, "a hole can also contribute to conduction"



electron conduction

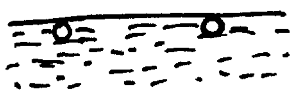


hole conduction



← hole and electron contributing to conduction. In both cases, the average velocity of the electrons is > 0.

We can visualize the bandstructure as a sea with clouds.



The sea represents the valence band with many many electrons, all occupying the lowest possible energy states.

Above the sea, seperated by a layer where there is nothing, the forbidden gap, we find clouds which represents the few electrons that have been thermally activated into a higher state.

In the sea we have some air bubbles which represent the holes. Transport of water ((current) can occur by moving the clouds (by wind) or moving the bubbles.

On a day without clouds (T=0 K) we have neither clouds nor airbubbles and conduction cannot occur.

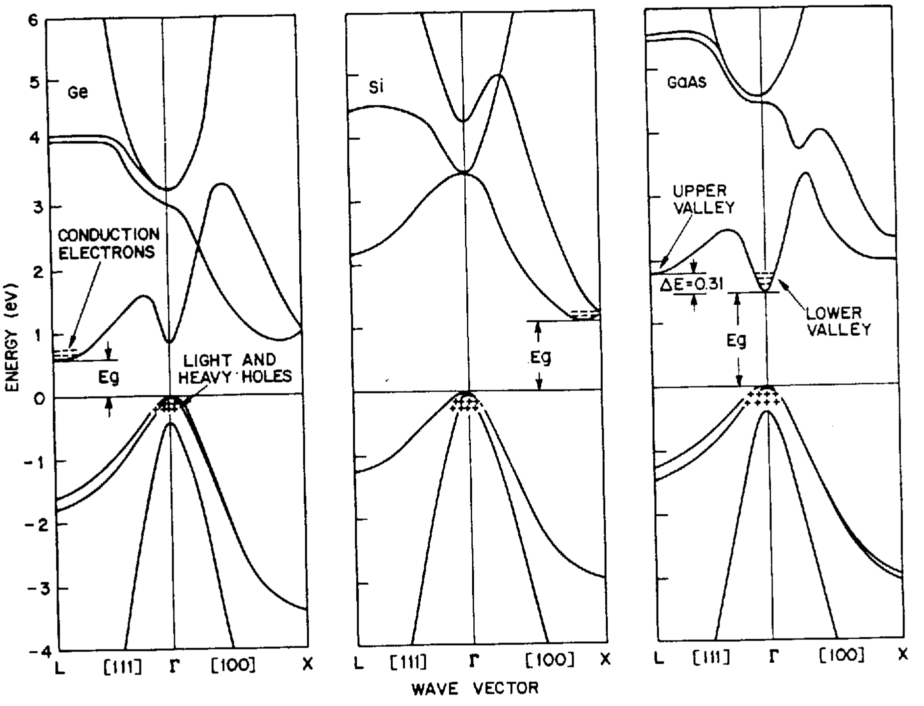


Fig. 5 Energy-band structures of Ge, Si, and GaAs, where E_g is the energy bandgap. Plus (+) signs indicate holes in the valence bands and minus (−) signs indicate electrons in the conduction bands. (After Chelikowsky and Cohen, Ref. 17.)

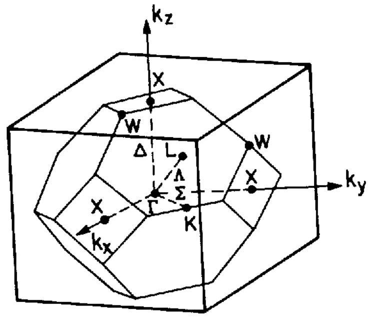
← Sze, p. 13

$\Gamma \Rightarrow k = (0,0,0)$

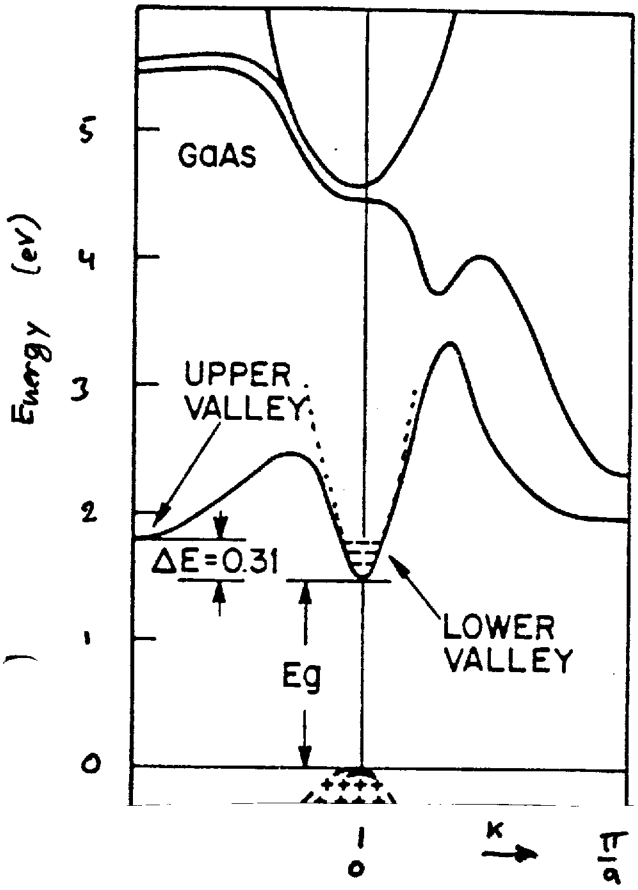
$X \Rightarrow k = \frac{2\pi}{a}(1,0,0)$

$L \Rightarrow k = \frac{2\pi}{a}(\frac{1}{2},\frac{1}{2},\frac{1}{2})$

$K \Rightarrow k = \frac{2\pi}{a}(\frac{3}{4},\frac{3}{4},0)$



Example: on basis of figure above: Determine (estimate) a value for the effective mass of electrons in GaAs.



Approximation .... to minimum of conduction band.

at $k = \frac{\pi}{5a}$ reaches ~ 1eV

$$\frac{\hbar^2 \left(\frac{\pi}{5a}\right)^2}{2\, m_e^*} = 1.6\ 10^{-19}\ J$$

$$\hbar^2 = \left(\frac{h}{2\pi}\right)^2$$

$$a_{GaAs} = 5.6533\ \text{Å}$$

$$\Longrightarrow\ m_e^* = 0.047\ m_e$$

((compares well to 0.067 of table in Sze)

In many cases we do not have to worry about the shape of the bands. The only thing we need to know, what is relevant, is the position of the maxima, E_c and E_v and the <u>effective</u> number of states N_c and N_v

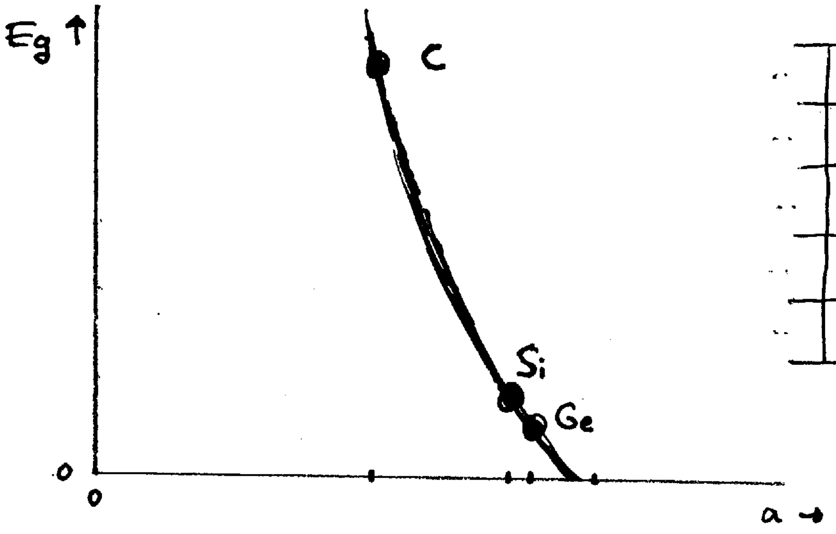
	Ge	Si	GaAs
N_c (cm^{-3})	1.04×10^{19}	2.8×10^{19}	4.7×10^{17}
N_v (cm^{-3})	6.0×10^{18}	1.04×10^{19}	7.0×10^{18}

A comparison of materials

summary of similar materials. The following table show group-IV elements C, Si, Ge, Sn, ordered in increasing mass

at 300K

	a	E_g
C	3.56683 Å	5.47 eV
Si	5.43095 Å	1.12 eV
Ge	5.64613 Å	0.66 eV
Sn	6.48920 Å	metal



III	IV	VI
B	C	N
Al	Si	P
Ga	Ge	As
In	Sn	Sb
Tl	Pb	Bi

The smaller the size of the atom, (C), the closer they can be together. This increases the interactions and the formation of bands and bandgaps.

A similar effect we can expect when <u>warming</u> up the material. The crystal will expand (increasing a) and the bandgap will decrease, see Fig 8 on p. 15 of Sze.

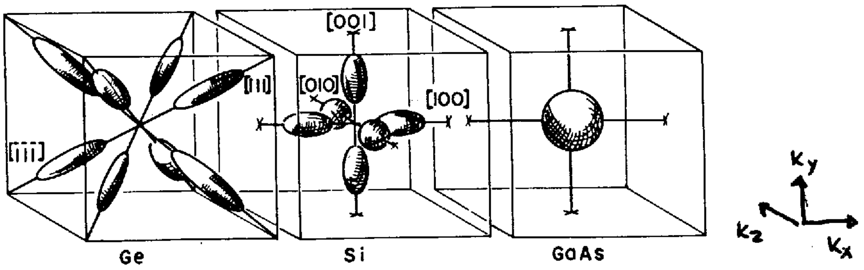
Direct - indirect bandgap semiconductors

Direct bandgap semiconductors have the maximum of the valence band at the same k-value as the minimum of the conduction band, whereas indirect semiconductors don't have this property.

The maximum of the valence band is always at Γ (k = (0,0,0)).

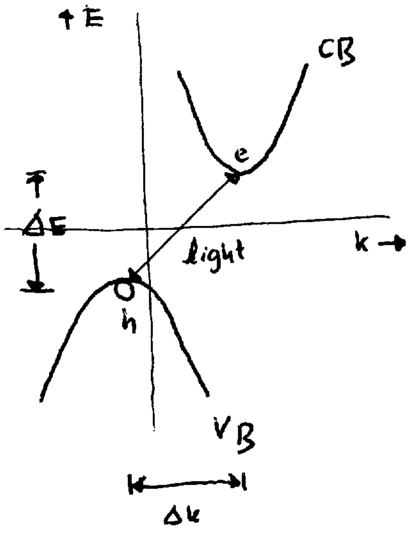
- Direct semiconductors therefore have their CB minimum at Γ, for instance GaAs.

See the Figure on p. 25



← See, p 14 "position of minimum of CB"

- The result is that only GaAs (of these three) can be used for optical applications.



To excite an electron from the top of the VB to the bottom of the CB the process has to

- conserve energy

$$h\nu = \Delta E = E_{CB} - E_{VB}$$

- conserve momentum

$$\Delta k = P_{photon}$$

- The first condition is easily met by choosing the appropriate frequency (wavelength) of the incident light:

$$h\nu = E_{CB} - E_{VB} = \Delta E_g$$

for example: GaAs :

$$\Delta E_g = 1.424 \text{ eV} \quad (2.28\ 10^{-19} \text{ J})$$

$$\Rightarrow \quad \nu = \frac{\Delta E_g}{h} = \frac{2.28\ 10^{-19} \text{ J}}{6.626176\ 10^{-34} \text{ J/Hz}} =$$

$$= 3.44\ 10^{14} \text{ Hz}$$

$$\lambda = \frac{c}{\nu} = 871 \text{ nm} \quad \text{(vermelho)}$$

- The second condition is more dificult to realize. The momentum of the photon has to be equal to Δk. As an example: silicon:

minimum of CB at $\sim (\frac{1}{2}, 0, 0)\ \frac{2\pi}{a}$, $a = 5.65$ Å

maximum of VB at $(0,0,0)$

$$\Delta k \approx 5.6\ 10^{9} \text{ /m}$$

$$\Delta p = \hbar \cdot \Delta k = 5.9\ 10^{-25} \text{ kg ms}^{-1}$$

momentum of photon of $\lambda = 871$ nm $= \frac{h}{\lambda} = 7.6\ 10^{-28}$ kg ms^{-1}

The momentum of the photon is <u>by far</u> not sufficient to supply the needed momentum. (1000×).

Effectively, all <u>optical</u> transitions are "vertical" in the band diagram:

- The first condition is easily met by choosing the appropriate frequency (wavelength) of the incident light:

$$h\nu = E_{CB} - E_{VB} = \Delta E_g$$

for example: GaAs :

$$\Delta E_g = 1.424 \text{ eV} \quad (2.28\ 10^{-19} \text{ J})$$

$$\Rightarrow \quad \nu = \frac{\Delta E_g}{h} = \frac{2.28\ 10^{-19} \text{ J}}{6.626176\ 10^{-34} \text{ J/Hz}} =$$

$$= 3.44\ 10^{14} \text{ Hz}$$

$$\lambda = \frac{c}{\nu} = 871 \text{ nm} \quad \text{(vermelho)}$$

- The second condition is more dificult to realize. The momentum of the photon has to be equal to Δk. As an example: Silicon:

minimum of CB at $\sim (\frac{1}{2}, 0, 0)\frac{2\pi}{a}$, $a = 5.65$ Å

maximum of VB at $(0,0,0)$

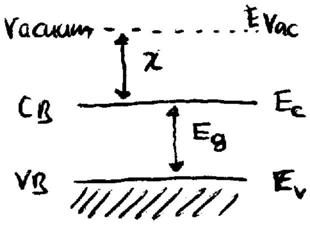
$$\Delta k \approx 5.6\ 10^{9} \text{ /m}$$

$$\Delta p = \hbar \cdot \Delta k = 5.9\ 10^{-25} \text{ kg m s}^{-1}$$

momentum of photon of $\lambda = 871$ nm $= \frac{h}{\lambda} = 7.6\ 10^{-28}$ kg m s^{-1}

The momentum of the photon is by far not sufficient to supply the needed momentum. (1000 ×).
Effectively, all optical transitions are "vertical" in the band diagram:

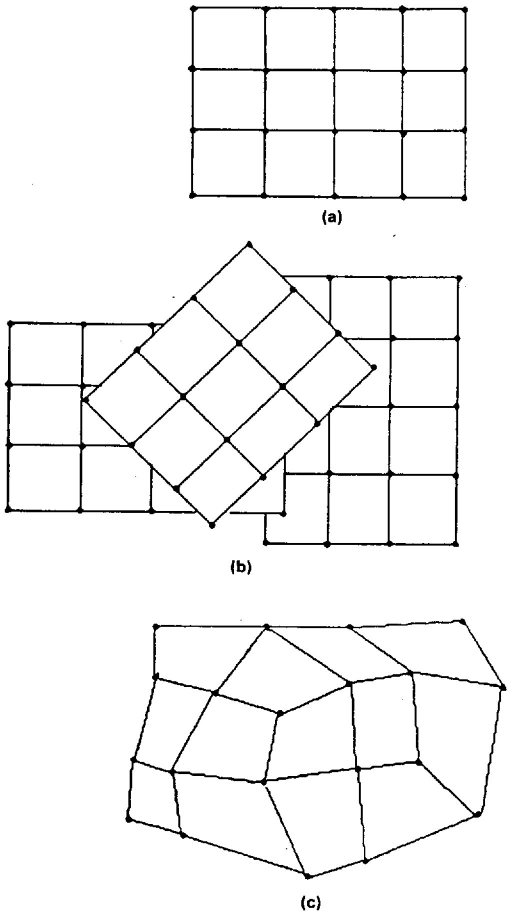
This latter is called the vacuum level



← Figure of material parameters

CRYSTALS

An atom itself doesn't conduct any current. Only when placed together do they conduct. There are several ways in which a material can solidify. If the atoms are placed randomly the material is amorphous. On the other side of the spectrum we have the crystal, where the ~~every~~ atoms are ordered regularly in the lattice. Inbetween we have polycrystaline forms, where locally the atoms are ordered, but on a long range they are disordered.


(a)

(b)

(c)

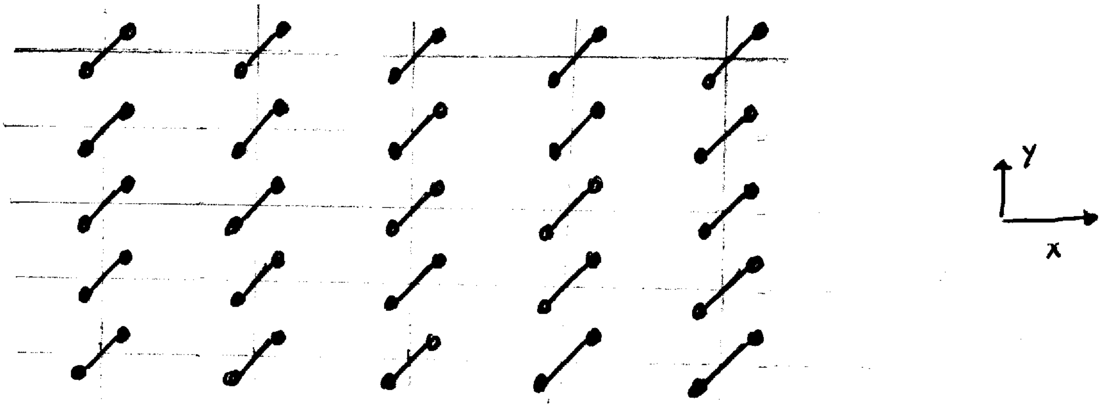
Figure from ← Shur, p. 17

We will focus our attention on crystals only.

Fig. 1-4-1. Schematic representation of atomic structure in single crystal (a), polycrystalline (b), and amorphous material (c).

crystals have (see chapter 1.2 of Blakemore)

- a <u>lattice</u>: a grid of coordinates. The property of this lattice is that "at every point of this lattice, it looks exactly the same".
- a <u>basis</u>. At every point of the lattice we put this object.



Example of a crystal consisting of a rectangular lattice with on every point a hydrogen molecule basis.

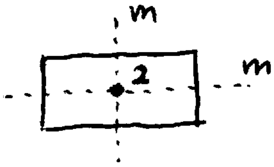
LATTICE

The lattice is defined by its translational symmetry. For example, shifting the above lattice by 2cm horizontal, or 1cm vertical leaves the lattice invariant.

(we assume the lattice stretches to infinity)

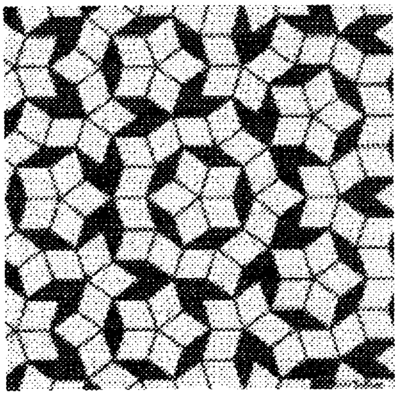
This puts some restrictions on to the <u>symmetry</u> of the crystal. The lattice above, for example (without basis) has

- a 2-fold axis (rotations over $\frac{360^\circ}{2}$ leaves lattice invariant)
- two mirrorplanes



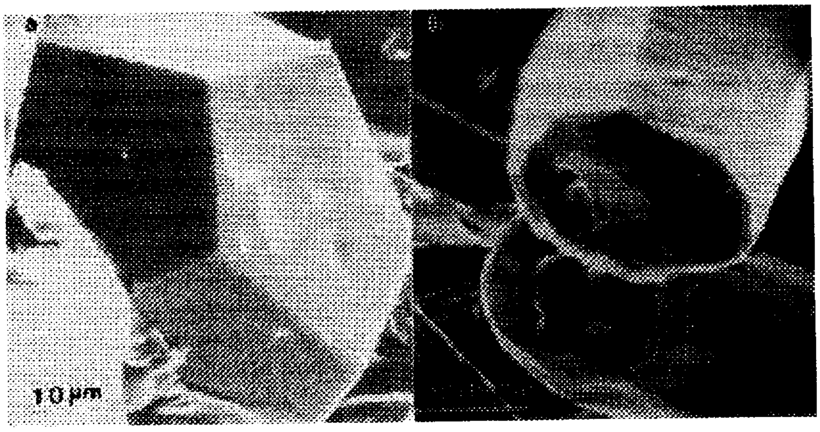
and is therefore called 2 mm.

The number of different lattices is limited. They are called Bravais lattices. For instance, it is not possible to have a lattice with 5-fold symmetry, because it is not possible to cover the entire space regularly with pentagons; the lattice will not be the same everywhere. Crystals of 5-fold symmetry will therefore not exist!



← "Penrose tiling". Attempt to cover a plane with pentagons in a regular way

http://www.josleys.com/TN_Pen011.JPG



← Quasi crystal showing macroscopic 5-fold symmetry

http://www.cmp.caltech.edu/~lifshitz/images/Tsai.jpg

Table : number of different Bravais lattices

1-dimensional	1	
2-dimensional	5	see exercise.
3-dimensional	14	see below

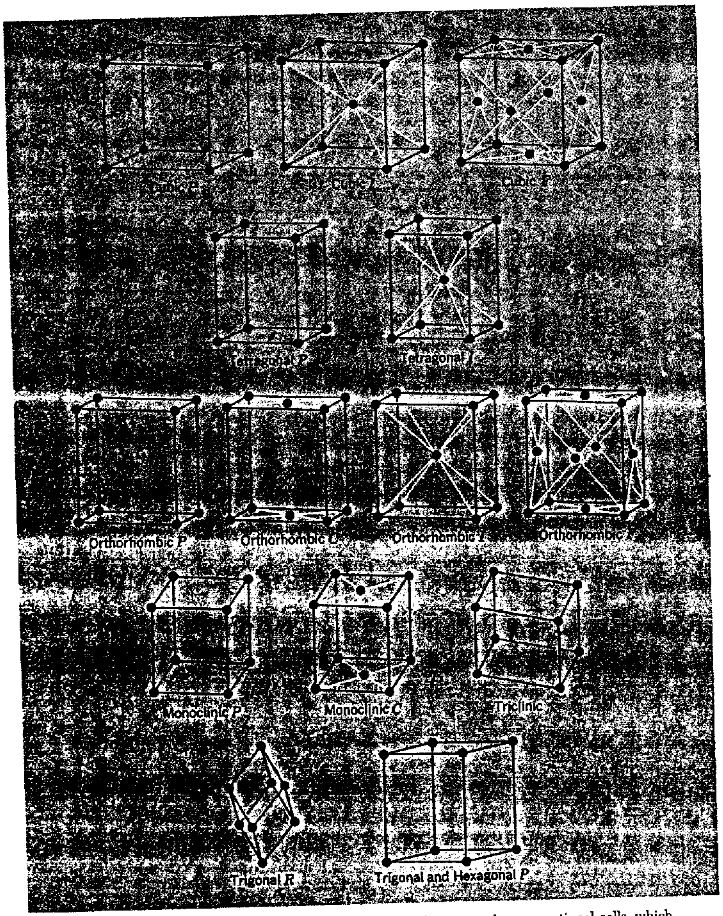
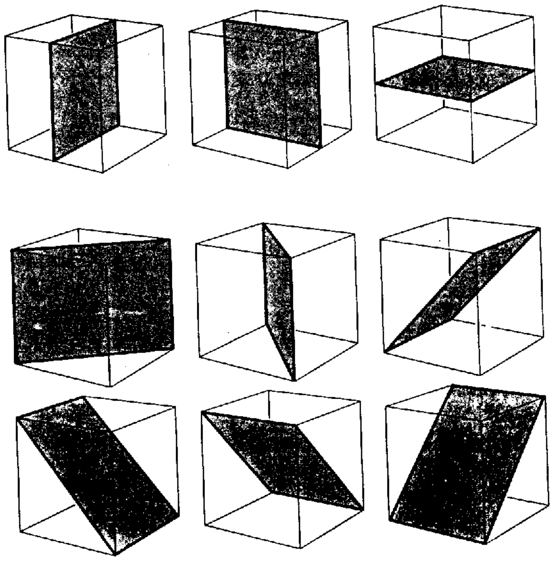


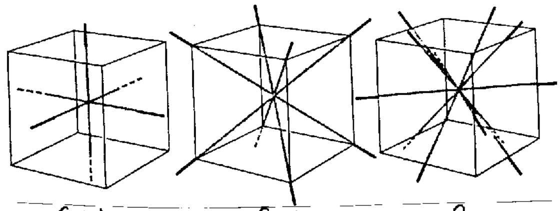
Figure 14 The fourteen Bravais or space lattices. The cells shown are the conventional cells, which are not always the primitive cells.

← 14 Bravais lattices in space.
(Kittel, p. 19)

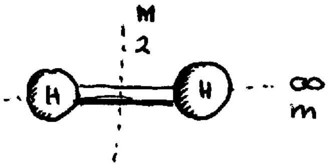
possible symmetry operations shown for a cube →

- mirror planes ▨
- rotation axes ⟋
- inversion (not shown)
- improper rotation axes ⟋
  (rotation + inversion)





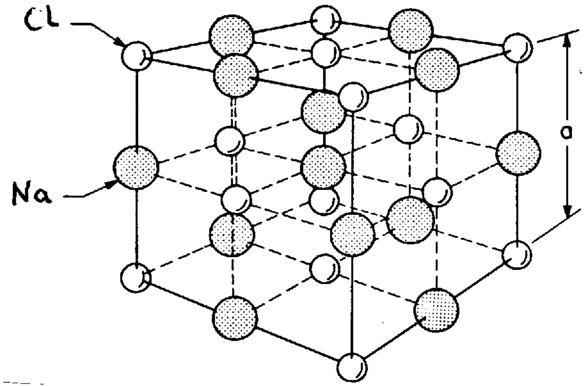
The BASIS is what is placed on every lattice point, for example the hydrogen molecule of the Figure before



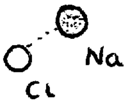
The basis itself also has a symmetry.

The total symmetry of the lattice + basis determines the symmetry of the crystal and this can be lower than the lattice or the basis individually.

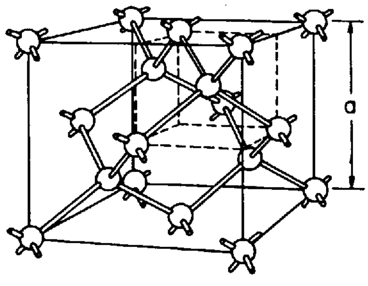
Some examples



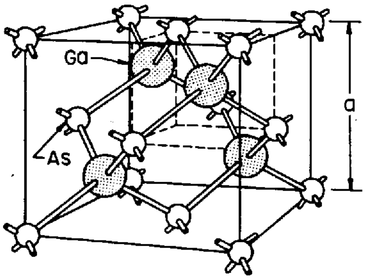
← The rocksalt crystal (NaCl, "kitchen salt") fig 2 . p 9 of Sze

The NaCl is a face-centered-cubic lattice (FCC) with a basis of , a Cl atom at every lattice point and a Sodium atom $(\frac{1}{2}a, 0, 0)$ displaced from that.

A much more common crystal structure of semiconductors is the diamond structure in which Si, Ge, C, GaAs, etc are crystallizing

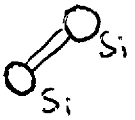


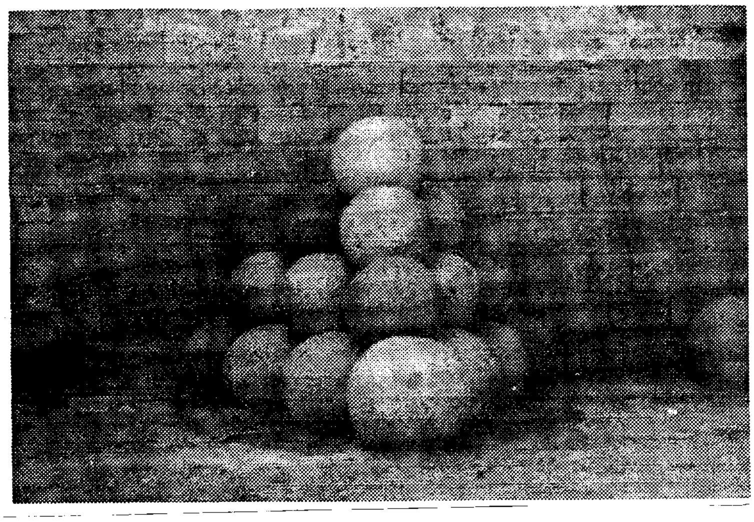
DIAMOND
(C, Ge, Si, etc)



ZINCBLENDE
(GaAs, GaP, etc)

Fig 1, P. 8 of Sze

They have an FCC lattice with on every lattice point a  pair (or Ge-Ge, or C-C). For the similar zincblende structure, every lattice point has a  pair (or Ga-P, etc).



← Stacked canon balls

←

http://www.home.unix-ag.org/scholl/pics/canon.jpg

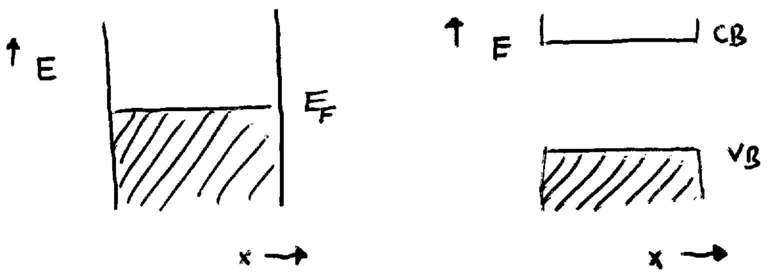
These crystals can be constructed by placing spheres as close as possible to eachother, like stacking canonballs (see image above). It is therefore also called hexagonal close packed (HCP).

In total, there are 30 different crystals possible in space (3 dimensions).

more info on symmetry:

http://www.otterbein.edu/home/fac/dnhjhns/symmetry/index.html

Effect of temperature



← metal (left) and semiconductor at $T = 0\,K$

▨ occupied

☐ empty

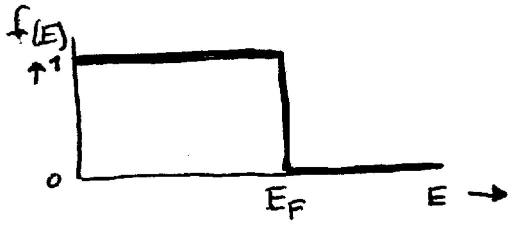
- electrons like to be at lowest energy (like "apples like to be close to the ground")
- 2 electrons cannot be in the same state.

$$\Downarrow$$

States are occupied up-to a certain level. we call this the <u>Fermi level</u> E_F.

below E_F : all states occupied $(T = 0K)$

above E_F : all states empty



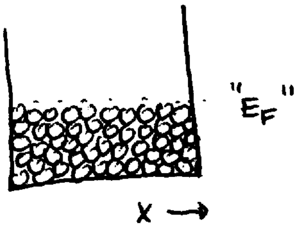
← occupation of levels as function of energy E.

Note : the number of occupied levels $n(E)$ is equal to the number of levels $g(E)$ multiplied by this function

$$n(E) = f(E) \times g(E)$$

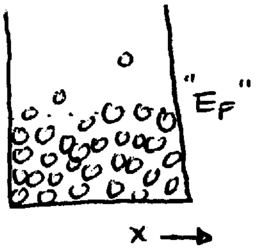
with $g(E)$ the density of states found before.

We can visualize this as a box with marbles. At $T = 0\,K$, all marbles are silently lying at the bottom of the box. What happens

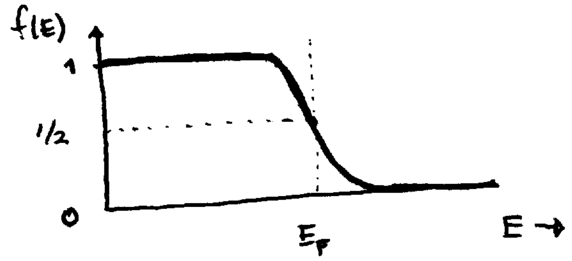


when we increase the temperature $T > 0K$. This can be seen as shaking the box with marbles.

They will jump around. Some of them will have higher energy. The average energy will also be higher.



The effect of temperature on our "box" of electrons will be very similar. Some electrons will have higher energy:



← Fermi-Dirac distribution

This is called the Fermi-Dirac distribution. Without derivation, this is given as (Sze, p. 17)

$$\boxed{f(E) = \frac{1}{1 + \exp\left(\frac{E-E_F}{kT}\right)}}$$

Notes:

- * at the Fermi level, E_F, $f(E) = 0.5$
- * $\int_0^\infty f(E)\,dE = 1$

The first one tells us that the Fermi level is there where the levels are half full (or half empty). The second observation tells us that we are not losing any electrons by heating up, whatever T.

With this information we can calculate the number of electrons $n(E)$ in the conduction band. The equation of p 38, $n(E) = f(E) \cdot g(E)$ becomes

$$n(E_c) = \frac{1}{1 + \exp\left(\frac{E_c - E_F}{kT}\right)} \cdot N_c$$

with E_c the energy of the bottom of the conduction band and N_c the effective number of states of the conduction band. For normal temperatures

$\exp\left(\frac{E_c - E_F}{kT}\right) >> 1$. thus

$$n(E_c) = \exp\left(-\frac{(E_c - E_F)}{kT}\right) \cdot N_c$$

In the same way we can calculate the number of occupied states in the valence band

$$n(E_v) = \frac{1}{1 + \exp\left(\frac{E_v - E_F}{kT}\right)} \cdot N_v$$

Normally $E_F > E_v$ and $\exp\left(\frac{E_v - E_F}{kT}\right) << 1$, thus $n(E_v) \sim 1$. It is much more useful to talk about the number of missing electrons in the valence band, or holes:

$$p(E_v) = N_v - n(E_v)$$

$$= \left(1 - \frac{1}{1+\exp\left(\frac{E_v - E_F}{kT}\right)}\right) \cdot N_v$$

$$= \exp\left(\frac{-(E_F - E_v)}{kT}\right) \cdot N_v$$

"Charge neutrality" tells us that the number of holes in the valence band must be equal to the number of electrons in the conduction band:

$$n \equiv p$$

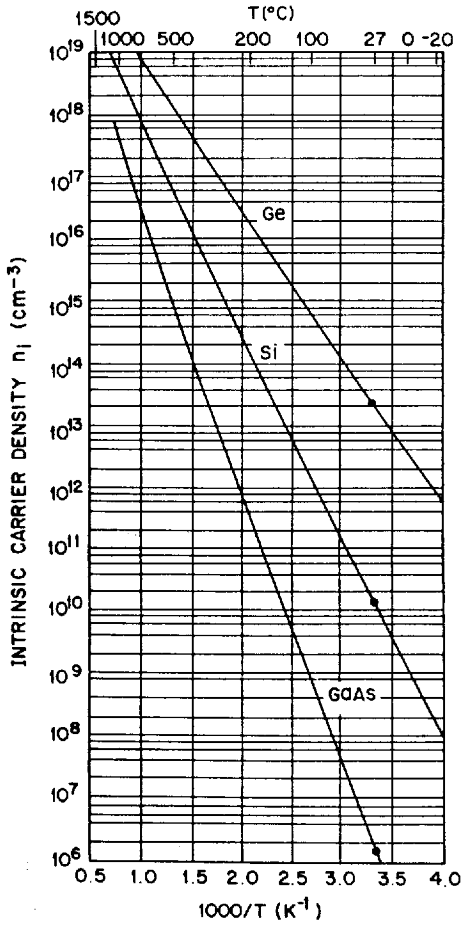
Then

$$n^2 = n \cdot p = N_C \cdot N_v \cdot \exp\left(-\frac{E_c - E_v}{kT}\right)$$

or

$$n = \sqrt{N_C N_v} \; \exp\left(-\frac{E_c - E_v}{2kT}\right)$$

This means that we know the number of "free" electrons in the conduction band (and free holes in the valance band) as a function of temperature: they rise exponentially with temperature. Both N_c and N_v also depend on temperature $(\sim T^{3/2})$, but this is less important. The same applies to dependence on T of $E_g = E_c - E_v$



← Sze. p.19

Intrinsic carrier densities of Ge, Si, and GaAs as a function of reciprocal temperature.

This picture shows the intrinsic free-carrier concentration. This means that we consider a pure crystal, without any impurities. As we will see later, impurities will have a dramatic effect on the free carrier concentration, which gives the materials properties needed for electronic applications.

position of Fermi level: Going back to the equations for $n(E_c)$ and $p(E_v)$. Because they should be equal, we can calculate the position of the Fermi level:

$$\exp\left(-\frac{(E_c-E_F)}{kT}\right)\cdot N_c = \exp\left(-\frac{(E_F-E_v)}{kT}\right)\cdot N_v$$

or

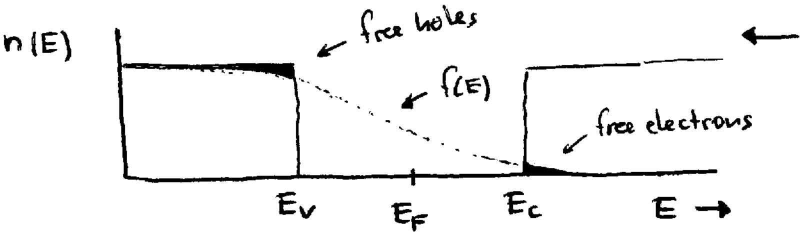
$$\exp\left(\frac{2E_F}{kT}\right) = \frac{N_v}{N_c}\exp\left(\frac{E_v+E_c}{kT}\right)$$

or

$$E_F = \frac{kT}{2}\ln\left(\frac{N_v}{N_c}\right) + \left(\frac{E_v+E_c}{2}\right)$$

This tells us that the Fermi level for intrinsic semiconductors is alway very close to the midgap $\frac{1}{2}(E_v + E_c)$, because N_v is of the same order as N_c.

Example Si : $N_c = 2.8\ 10^{19}$, $N_v = 1.04\ 10^{19}$ $\Rightarrow$ at $T = 300$ K the Fermi level is just 13 meV below the middle of the gap



← Schematic representation of bands for T > 0 K

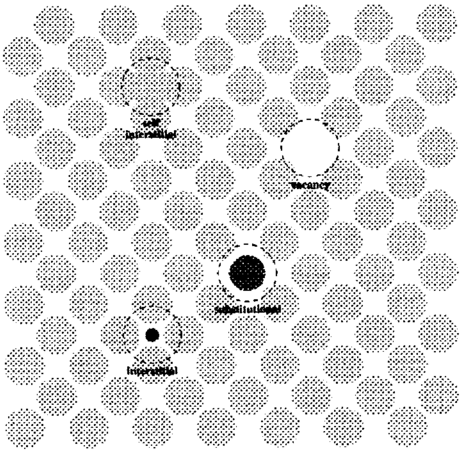
Note : the width of the distribution f(E) is much smaller than shown above. In reality, the width is approximately $kT = 26$ meV (300 K), rather than the ~1000 meV shown above.

Crystal imperfections

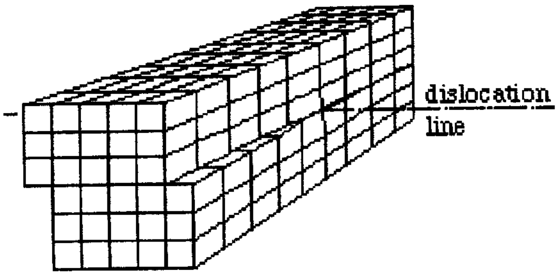
Blakemore ch 1.6
A&M ch 30

Defects in the crystal can have dramatic effect on the electronic properties of the material. They fall into two categories

* Self defects. They consist of extra, missing or misplaced atoms of the same kind as the host crystal. They tend to be neutral (not changing the electric properties), because such atoms have the same valence.



← some point defects



↑ extended defect

* impurities, consisting of the incorporation of foreign atoms into the crystal. Since they normally have a different valence they can change the electric properties of the material. There are two types of impurities

  - donors
  - acceptors

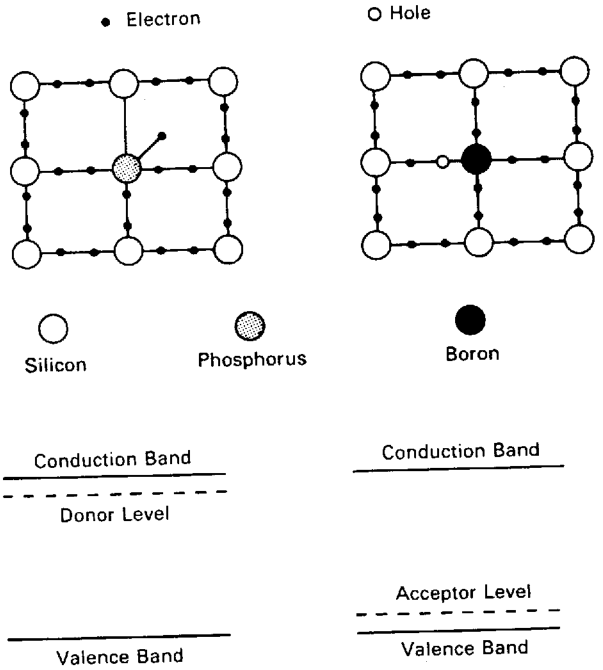
DONORS AND Acceptors (p20, Sze)

[A donor is an atom that has one valence electron more than the host crystal atom, whereas an acceptor has one valence electron less.]

As an example. Phosphorous has 5 valence electrons (in the periodic table of elements of Mendeleev it is in group V). Thus, placed in Silicon (with 4 valence electrons) it will act as a donor. On the other hand, Boron is a group III atom and will be an acceptor in Silicon.

The extra electron is very losely bound to the donor and can easily be removed and placed in the conduction band.

The "binding energy" of the donor electron can be calculated using a hydrogen-model or effective-mass model.

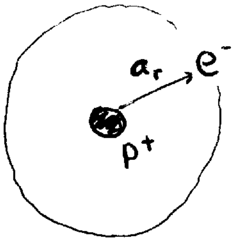


Schematic diagram of a silicon crystal structure with a silicon atom replaced by a phosphorus atom and a boron atom.

Shur, p.56

Effective mass theory

The energy of an electron at a hydrogen atom is given by (no derivation) (Gasiorowicz)



$$E_H = \frac{m_e q^4}{32 \pi \hbar^2 \varepsilon_r^2 \varepsilon_0^2}$$

$q = 1.6 \times 10^{-19}$ C, $\hbar = \frac{h}{2\pi}$, $h = 6.626\ 10^{-34}$ Js, ε_0 is permittivity of vacuum $= 8.85\ 10^{-12}$ F/m, ε_r is relative permittivity ($\equiv 1$ for vacuum), m_e is mass of electron $= 9.1\ 10^{-31}$ kg. This gives (for hydrogen in vacuum)

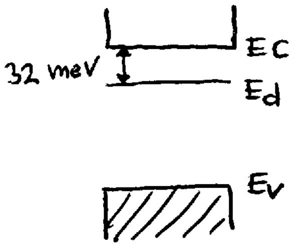
$$E_H = 13.6 \text{ eV}$$

Effective mass theory says that the binding energy is equal to this energy but scaled with

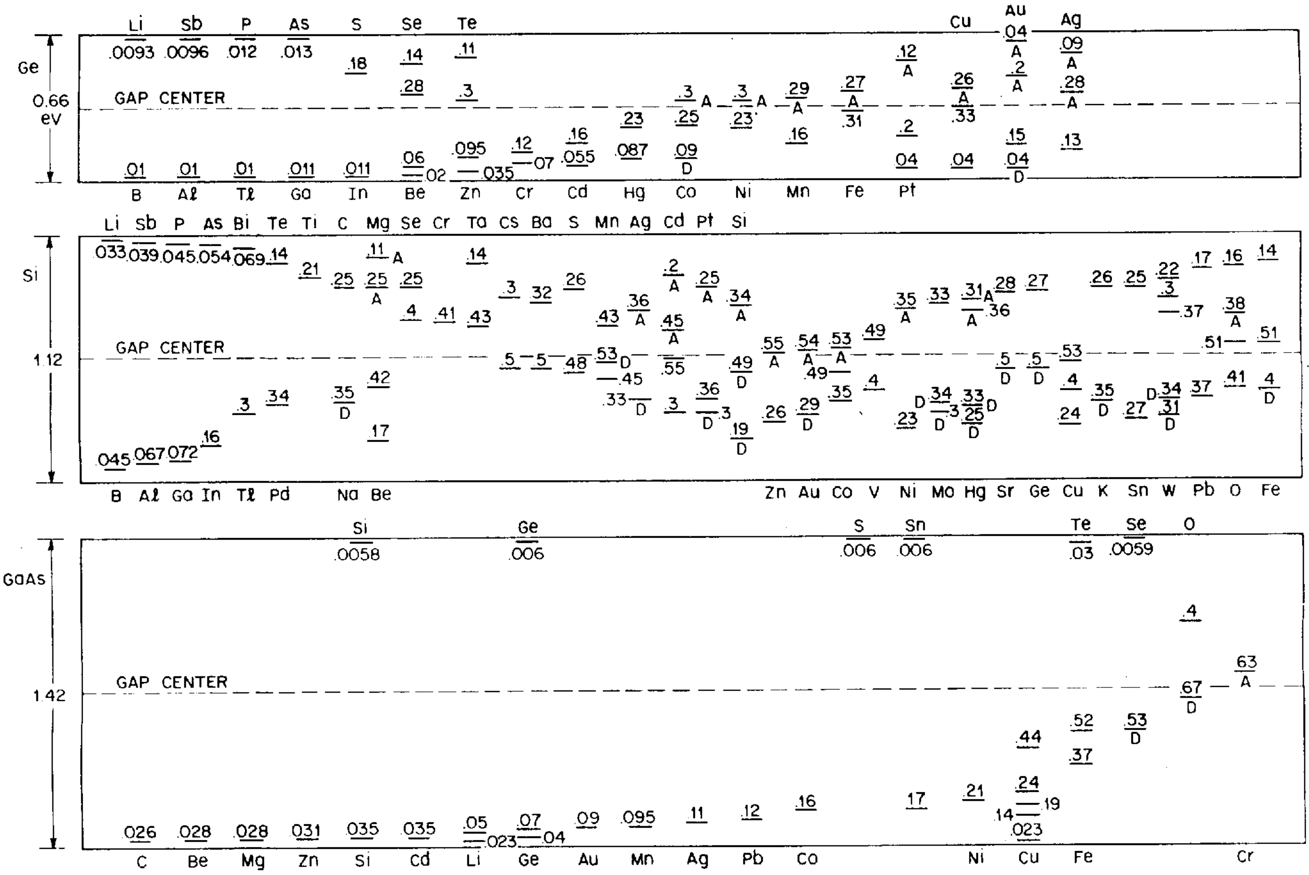
- the effective mass $m_e \rightarrow m_e^*$
- the permittivity of the host material $\varepsilon_r \neq 1$.

Example: donor in silicon, $m_e^* = 0.33\ m_e$, $\varepsilon_r = 11.9$:
Thus $E_d = E_H \cdot \frac{m_e^*}{m_e} \cdot \frac{1}{\varepsilon_r^2} = 32$ meV

Effective mass theory therefore predicts every donor level to be 32 meV below the conduction band in silicon, independent of the type of donor. It gives us therefore only a good estimate for the donor level. The next figure gives a summary of levels in Ge, Si and GaAs.



p 310-313 of Blakemore



Measured ionization energies for various impurities in Ge, Si, and GaAs. The levels below the gap centers are measured from the top of the valence band and are acceptor levels unless indicated by D for donor level. The levels above the gap centers are measured from the bottom of the conduction-band level and are donor levels unless indicated by A for acceptor level. The bandgaps at 300 K are 0.66, 1.12, and 1.42 eV for Ge, Si, and GaAs, respectively. (After Conwell, Ref. 27; Sze and Irvin, Ref. 28; Milnes, Ref. 24.)

← p21 of Sze

The calculations of the acceptor levels are similar. They are also given in the figure above. An acceptor steals an electron from the valence band, or "donates a hole" to the valence band.

Fermi level in doped semiconductors

The next task is to calculate the Fermi level and the number of free electrons and free holes in the doped material. As an example we will calculated this for silicon doped with $N_D = 10^{16}$ cm^{-3} donors ~~at~~ with ionization energy of 32 meV.

(section 1.4.3. of Sze)

The basis of these calculations is "charge neutrality"; for every donor ionized (N_D^+) or hole created in the valence band (p) there must be a free electron in the conduction band:

$$n = N_D^+ + p$$

Now we can use our Ferm- Dirac statistics

$$N_D^+ = N_D \left[1 - \frac{1}{1 + \exp\left(\frac{E_D - E_F}{kT}\right)} \right]$$

$$n = N_C \exp\left(-\frac{E_C - E_F}{kT}\right)$$

$$p = N_V \exp\left(-\frac{E_F - E_V}{kT}\right)$$

This set of equations determines E_F, n and p. A nice way to do this graphically is as follows:

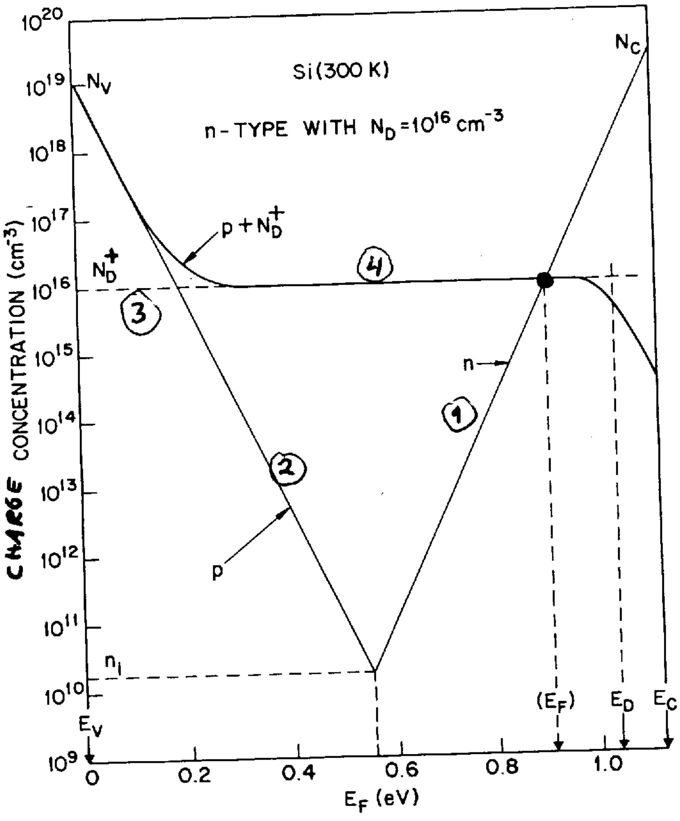


Fig. 15 Graphical method to determine the Fermi energy level E_F. (After Shockley, Ref. 31.)

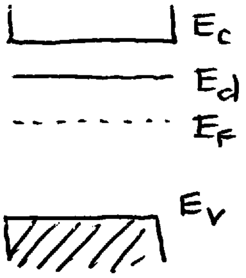
plot as function of E_F:

1. $n(E_F)$
2. $p(E_F)$
3. $N_D^+(E_F)$
4. $p + N_D^+(E_F)$

find crossing point of 1 with 4. This will be E_F. Now read n, N_D^+ and p from curves

↑ Sze, p. 24

In the above example we see that (nearly) all donors are ionized because the Fermi level is well below the donor level

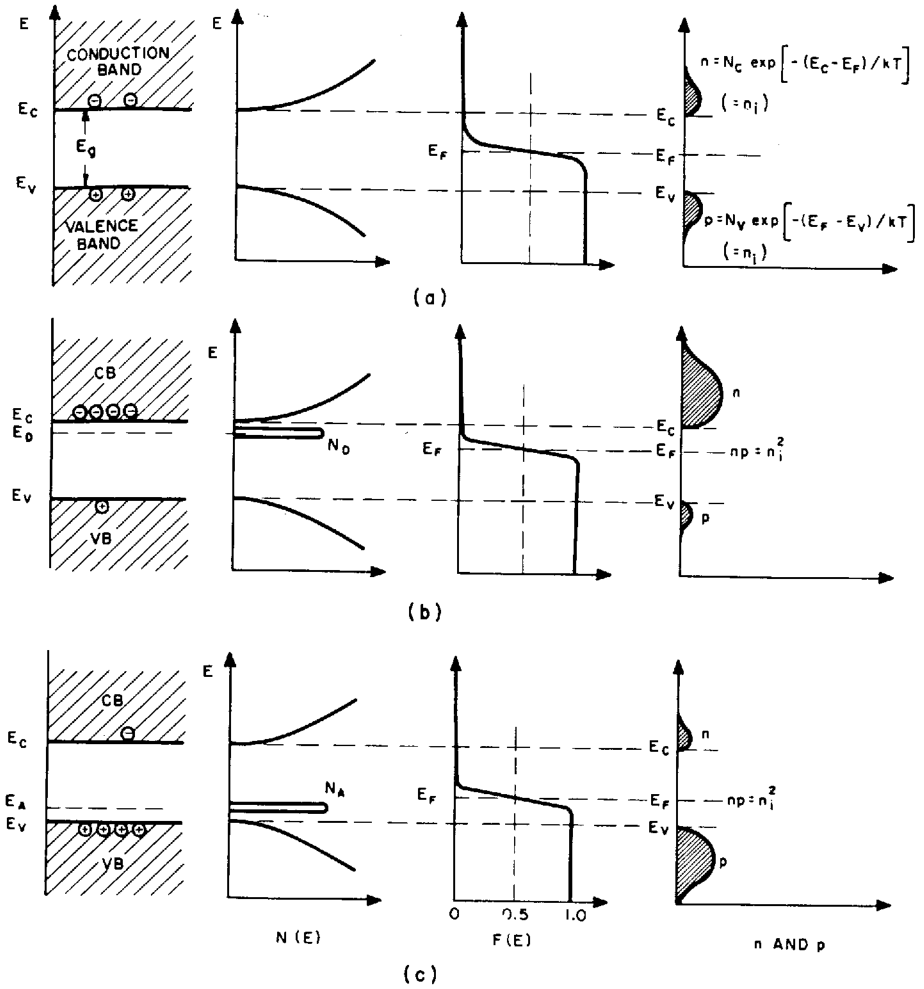


$$E_D - E_F \gg kT$$

Moreover, the number of ionized donors does not depend so much on the donor level depth E_d, nor does the number of free electrons, which is approximately equal to the number of donors N_D:

[Under normal conditions, the number of free electrons only depends on the number of donors.]

On basis of this we can build gas detectors. If the gas molecules act as donors or acceptors, we will observe an increase in conductivity in the presence of these molecules. The sensitivity can be very high. The figure on the previous page shows that the intrinsic density of electrons is only $2 \times 10^{10} / cm^3$ and this is also the order we can detect.



← Sze, p 23

Schematic band diagram, density of states, Fermi–Dirac distribution, and the carrier concentrations for (*a*) intrinsic, (*b*) *n*-type, and (*c*) *p*-type semiconductors at thermal equilibrium. Note that $pn = n_i^2$ for all three cases.

The figure above summarizes the effect of doping on the electron and hole distributions.

One final thing to note is that, independent of the donor or acceptor concentration, the product $n \times p$ is given by

$$N_c \exp\left(-\frac{E_c - E_F}{kT}\right) \times N_v \exp\left(-\frac{E_F - E_V}{kT}\right)$$

$$= N_c N_V \exp\left(-\frac{E_g}{kT}\right)$$

This may come in handy later.

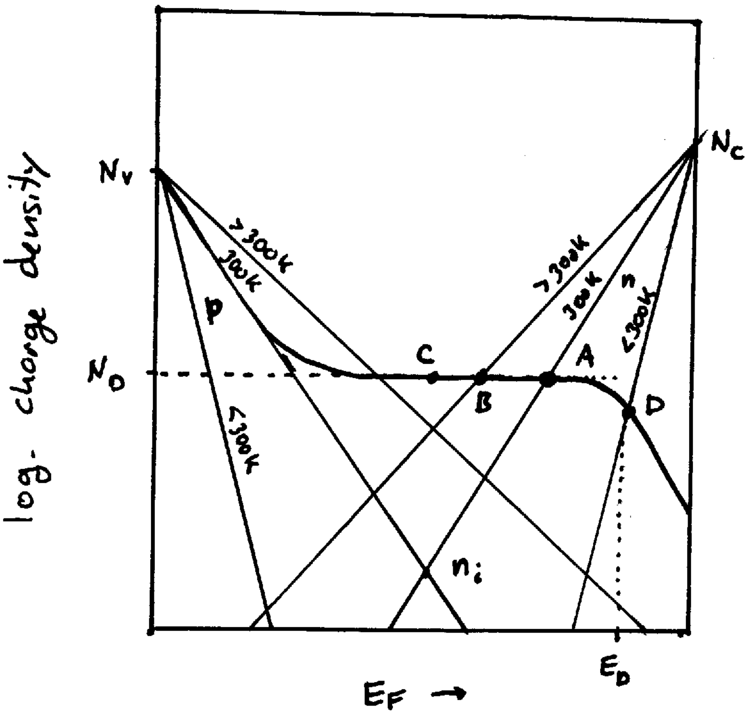
Temperature effects:

The slope of the distributions

$$n \sim N_c \exp\left(-\frac{E_F}{kT}\right)$$

$$p \sim N_v \exp\left(\frac{E_F}{kT}\right)$$

in a log plot depend on the temperature



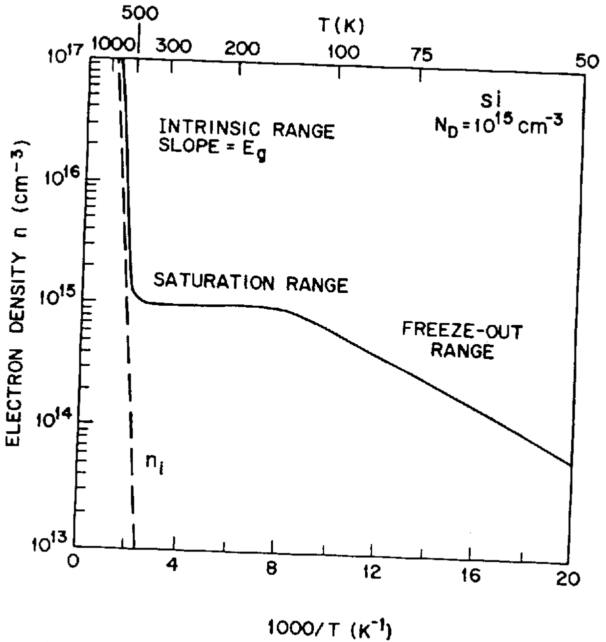
← Schematic effect of temperature change

- For increased temperatures the intrinsic carrier concentration (n_i) will increase. For certain temperatures this n_i can be larger than the donor concentration. (C) For large T we return to the intrinsic case:

$$n = \sqrt{N_c N_v} \exp\left(-\frac{E_g}{2kT}\right)$$

- For low temperatures the donor will no longer be fully ionized (D)

$$n \approx \exp\left(-\frac{E_c - E_D}{kT}\right)$$



← Sze, p. 26

Summary of temperature effects

Electron density as a function of temperature for a Si sample with donor impurity concentration of 10^{15} cm^{-3}.

INTRINSIC RANGE : $n_i > N_D$

SATURATION RANGE : n independent of T, see points $A \rightarrow B \rightarrow C$ on previous page.

FREEZE OUT RANGE : E_F above E_D

see on-line simulation of OptoEL/CEOT :

http://www.ualg.pt/fct/adeec/optoel/Java

current density equation (CDE)

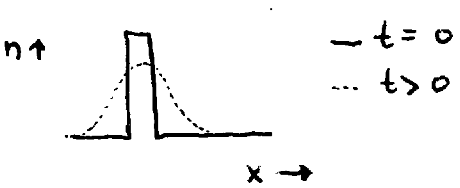
electron current density ($J_n = I_n/A$, "current per area")

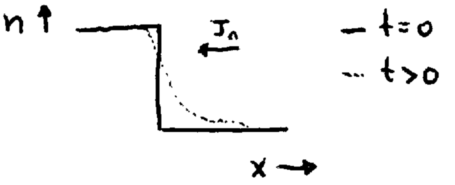
$$\boxed{J_n = q \mu_n n E_x + q D_n \frac{dn}{dx}} \qquad \text{(CDE)}$$

The first term we have already seen in the beginning of the lectures. It says that the drift current is proportional to the density of free electrons (n) and the electric field (E_x). The proportionality is the mobility of the electrons (μ_n). To get current density instead of electron flow density (number of electrons per second per square meter) we have to multiply by the charge of each electron (q).

The second term states that there will be a net flow of electrons, and hence diffusion current, wherever there exists a gradient in electron density ($\frac{dn}{dx}$). This is called diffusion. The proportionality is the diffusion constant D_n. Again, we have to multiply it with the elementary charge, q.

See the two examples on the left. It shows how the distribution of electrons evolves over time. The second example results in a net current (flow of electrons) in the positive x direction.





For "ohmic" devices (bars of homogeneous silicon) the second term is not important because the density of electrons is everywhere the same. However, for inhomogeneous devices, such as p-n junctions or metal-semiconductor interfaces, this term becomes important.

(normally the Einstein relationship $D_n = \frac{kT}{q}\mu_n$ holds and we can write the current-density equation as

$$J_n = q\mu_n\left(nE + \frac{kT}{q}\frac{dn}{dx}\right)$$)

The CDE for holes is similar:

$$\boxed{J_p = q\mu_p p E_x - q D_p \frac{dp}{dx}}$$

note the negative sign in the second term which is caused by the positive sign of the charge of holes.

Charge continuity equations (CCE)

charges cannot be created or disappear (except for electron-hole pair creation discussed in the optical chapters), therefore:

$$\frac{dn}{dt} = \frac{1}{q}\frac{dJ_n}{dx}$$

normally we consider steady state, which means: no longer anything depends on time. $\frac{dn}{dt} = 0$, thus

$\frac{dJ_n}{dx} = 0$, or: "current density is everywhere the same!".

we can talk about "the current of a device".

The same applies to the hole current J_p in steady state :

$$\frac{dp}{dt} = 0 \text{ and } J_p \text{ is constant. in space}$$

$$\boxed{J_n = \mu_n n \frac{dE_F}{dx}}$$

"The current is caused by the gradient of the Fermi level". To prove this, use

$$J_n = q\mu_n \left(nE + \frac{kT}{q}\frac{dn}{dx} \right)$$

with $E = \frac{dv}{dx}$, for example $E = \frac{1}{q}\frac{dE_c}{dx}$

and $n = N_c \exp\left(\frac{E_F - E_c}{kT} \right)$,

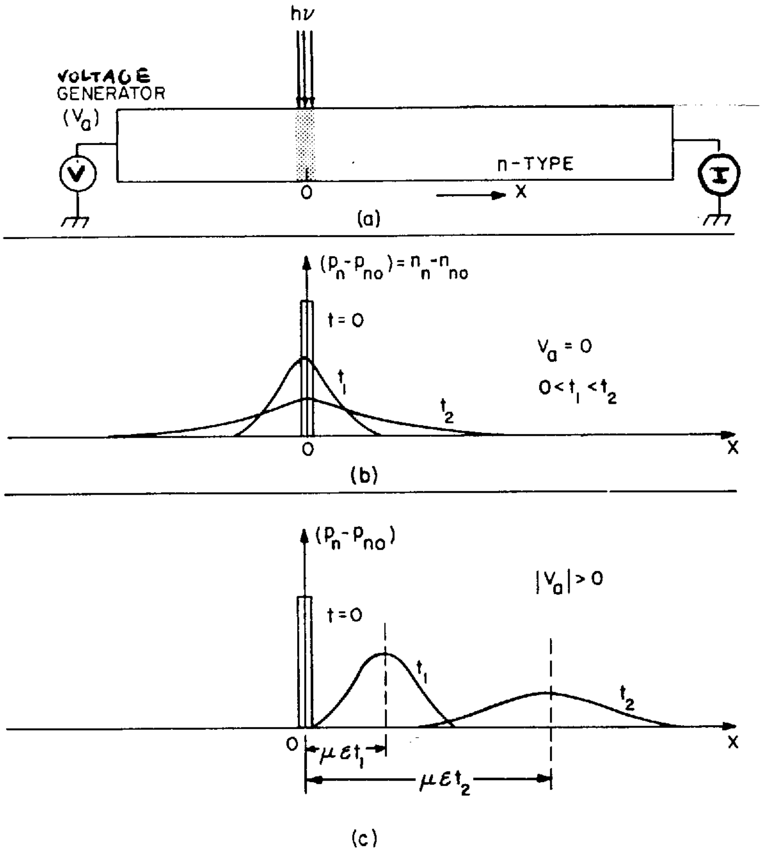
$$\frac{dn}{dx} = \frac{d}{dx}\left(N_c \exp\left(\frac{E_F - E_c}{kT} \right) \right) = N_c \exp\left(\frac{E_F - E_c}{kT} \right) \cdot \frac{1}{kT}\left(\frac{dE_F}{dx} - \frac{dE_c}{dx} \right)$$

$$= n \cdot \frac{1}{kT} \cdot \left(\frac{dE_F}{dx} - \frac{dE_c}{dx} \right).$$

If there is no current, the Fermi level is flat. (constant in space).

← p. 55, Sze

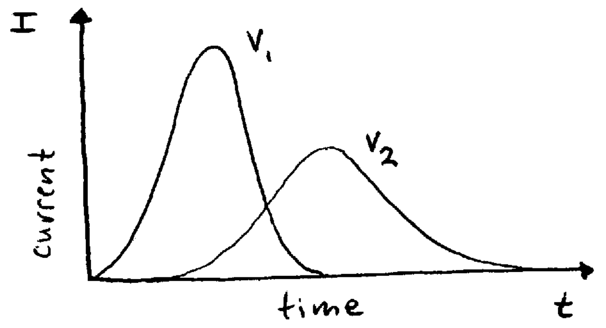
Example that shows diffusion and drift currents



Transient and steady-state carrier diffusion. (*a*) Experimental setup. (*b*) Without applied field. (*c*) With applied field.

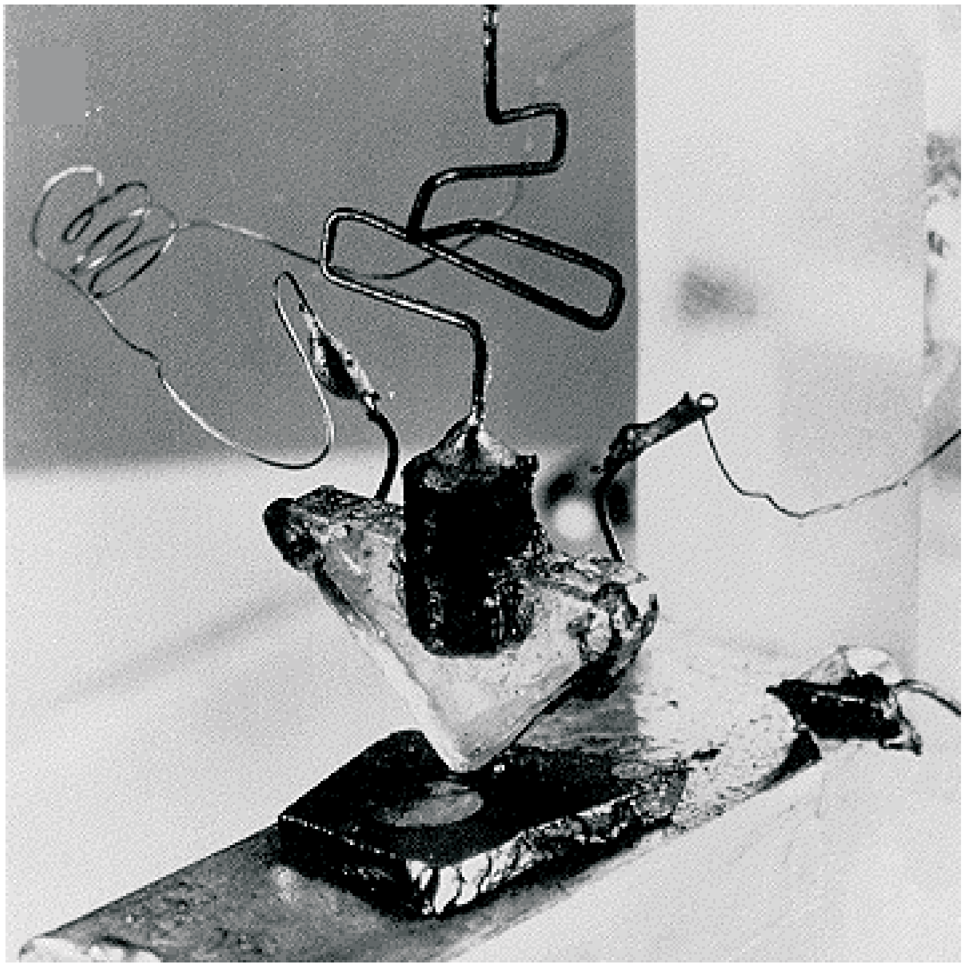
In the above experiment, electrons and holes are created by a light pulse. (a). Initially we start with a rectangular distribution, but diffusion $\left(\frac{dp}{dt} = -2 D_p \cdot \frac{dp}{dx} \right)$ widens the distribution (b). If we also switch on the field this distribution is moved to the right $\left(\frac{dp}{dt} = q \cdot \mu_p \cdot p \cdot E_x \right)$.

The current we might expect is the following:



$V_1 > V_2$

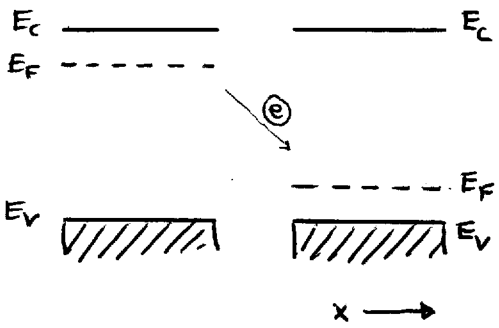
Devices I



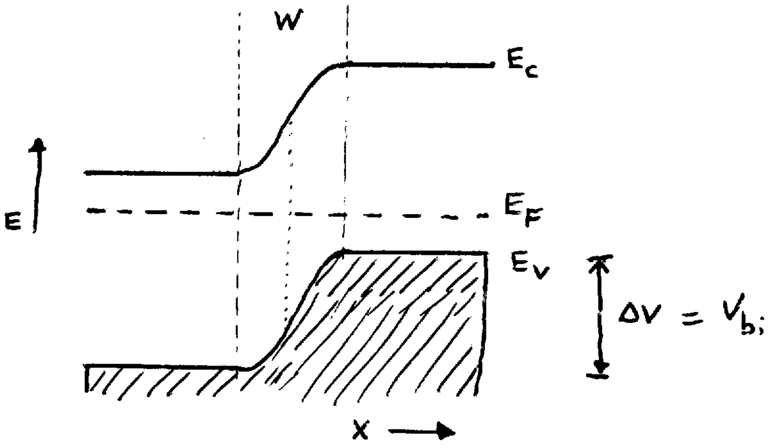
Bipolar (p-n junction) diode

A p-n junction is formed by the connection of p-type and n-type of the same material. As we know, the Fermi level in n-type and p-type material is different. Thus, when we connect the two pieces, electrons can gain energy by moving from the n-type side to the p-type side and we will get a current $(J_n = \mu_n n \frac{dE_F}{dx})$ and holes will move to the n-type side : $(J_p = \mu_p p \frac{dE_F}{dx})$.

The charges that jumped to the other side will create a field and a potential drop accross the interface that counteracts the force of difference in Fermi level. After a while the new device will reach steady state and the currents have stopped.



p-n junction
before contact



pn junction
after contact

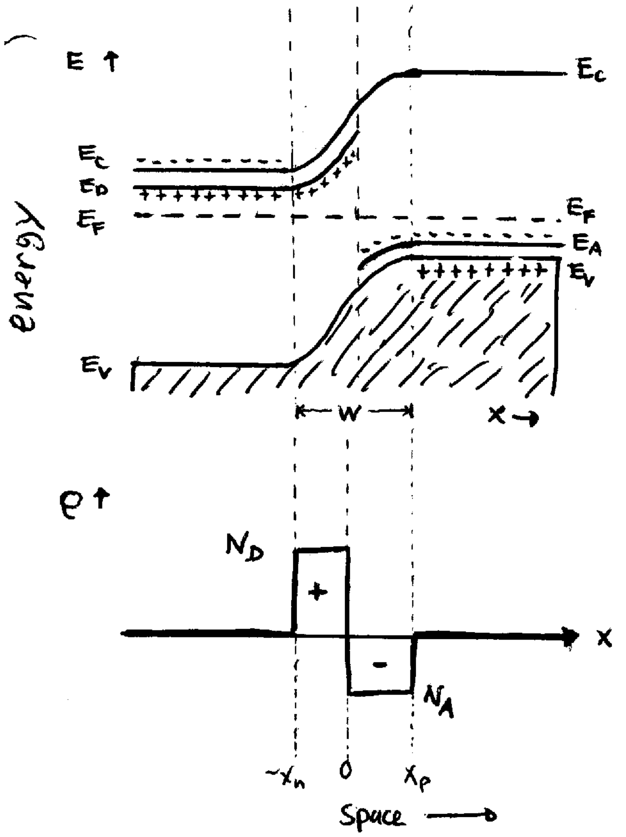
Since we no longer have currents, the Fermi level must be equal everywhere! Far from the interface the Fermi levels are still the same as in the bulk.

The voltage drop caused by the charges, ΔV, must therefore be equal to the difference in Fermi level before contact; this is called the built-in voltage

$$qV_{bi} = \Delta E_F$$

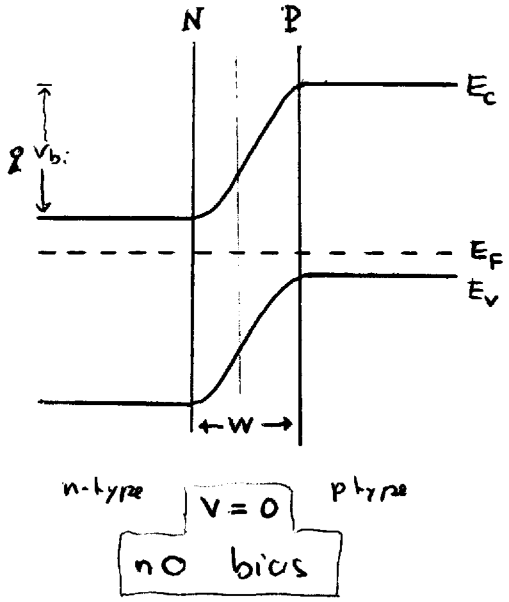
which is a parameter determined by the materials used and depends on donor and acceptor concentrations and bandgaps of the material used.

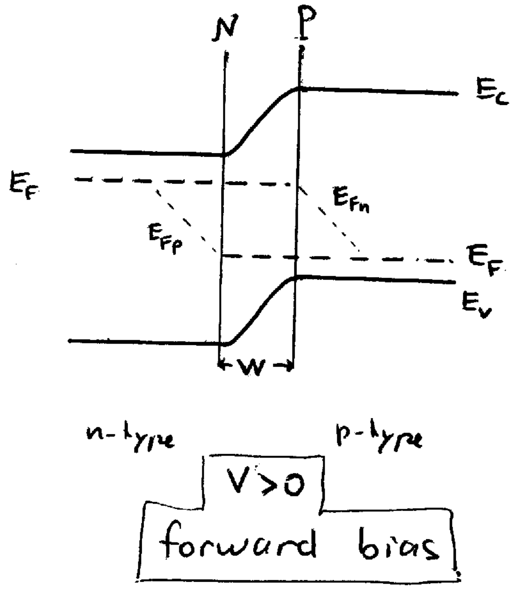
The electrons have moved to the p- side of the device, leaving behind uncompensated, ionized (positively charged) donors in a region close to the interface. This charge, with a density equal to the donor concentration ($\rho = N_d$) is the cause for the field and the voltage drop. On the other side of the interface we have a region where the holes have moved out and left behind uncompensated, ionized (negatively charged) acceptors.



← Schematic picture of the space-charge region or "depletion region" W inside which no free carriers exist and a net charge occurs due to uncompensated donors and acceptors.

It is easy to calculate the width of the depletion zone, W, but we will leave this for the easier case of the Schottky diode. Suffices to say that W and the banddiagram depend on the external voltage, V:





no bias:

The Fermi level is constant inside the depletion width. As a result, the bands move away from this Fermi level and the number of free carriers rapidly decreases ($n = N_c \exp(-\frac{E_c - E_F}{kT})$). As an example: on the left side of W, there is a free electron concentration $n_0 = N_c \exp(-\frac{E_c - E_F}{kT})$, on the right edge (P) there is

$$n = n_0 \exp\left(\frac{-qV_{bi}}{kT}\right)$$

with bias

The picture on the right shows the situation for forward bias. The voltage drop accross the interface has diminished (and is now $V_{bi} - V$). For an external voltage equal to the built-in voltage we would reach a flat-band condition.

Because holes and electrons are injected into the depletion zone, they do no longer follow the thermal equilibrium equations. Quasi - Fermi levels can be defined for electrons an hole seperately

$$n = N_C \exp\left(-\frac{E_C - E_{Fn}}{kT}\right)$$

$$p = N_V \exp\left(-\frac{E_{Fp} - E_V}{kT}\right)$$

Note that these Fermi levels are different.

Although there is a current now, the Fermi levels inside the depletion zone are still flat. This is because the currents are very small and thus

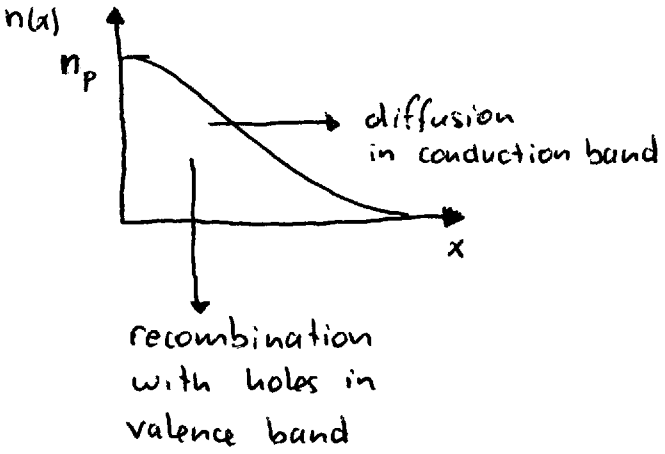
$$J_n = \mu_n n \frac{dE_{Fn}}{dx} = 0 \quad \text{and} \quad J_p = \mu_p p \frac{dE_{Fp}}{dx} = 0 .$$

Beyond the depletion zone, the excess carriers rapidly recombine ($e + h \rightarrow 0 \ (+ E_g)$) and the quasi Fermi levels join.

At the edge (I) of the depletion zone there are

$$n_I = n_0 \exp\left(\frac{-q(V_{bi} - V)}{kT}\right)$$

Current of a p-n junction diode



The electrons that are injected and reach the other side of the depletion zone

$$n_p = n_0 \exp\left(\frac{-q(V_{bi}-V)}{kT}\right)$$

will move away from there, either by diffusion or by recombination with abundant holes from the valenceband. They have a limited lifetime (τ_n). A distribution as given in the figure above is reached and the current

$$J_n = q\,(n_p - n_{p0}) \cdot \sqrt{\frac{D_n}{\tau_n}}$$

$$= q\, n_0 \sqrt{\frac{D_n}{\tau_n}} \left[\exp\left(-\frac{q(V_0-V)}{kT}\right) - 1\right]$$

In the same way, the hole current (on the left edge of the depletion width) is given by

$$J_p = q\, p_0 \sqrt{\frac{D_p}{\tau_p}} \left[\exp\left(\frac{-q(V_0-V)}{kT}\right) - 1\right]$$

where p_0 is the hole concentration in the bulk of the p-type side, D_p is the diffusion coefficient of holes and τ_p is the recombination lifetime. Because the holes have to recombine with electrons, necessarily we have the equality $\tau_n = \tau_p = \tau$.

The total current $J(V) = J_n + J_p$:

$$\boxed{J(V) = J_s \left[\exp\left(\frac{qV}{kT} \right) - 1 \right]}$$

with

$$\boxed{J_s = q n_0 \sqrt{\frac{D_n}{\tau}} + q p_0 \sqrt{\frac{D_p}{\tau}}}$$

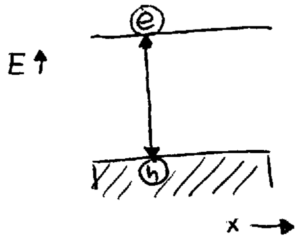
This is the famous Schockley equation, named after W. Schockley of Bell Labs (1950).

- J_s is the reverse-bias ($V = -\infty$) saturation current, which we already know from the electronics lectures. Now we know that it depends on

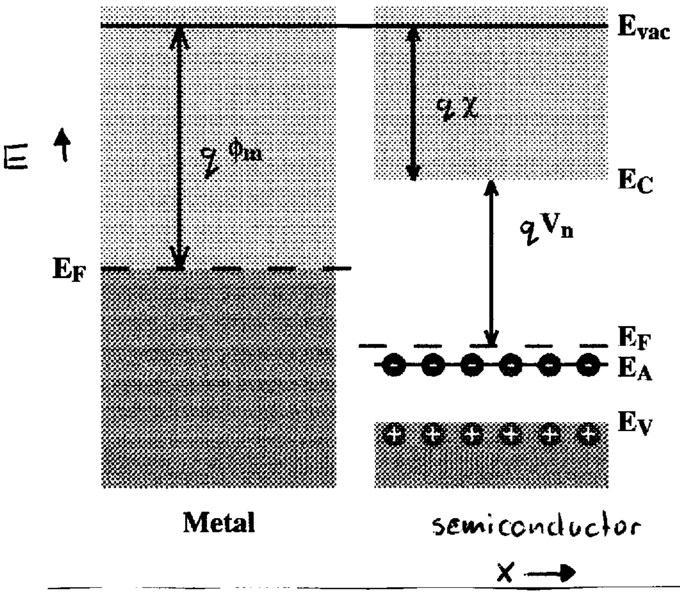
  1. The doping concentration ($n_0 \approx N_D$, $P_0 \approx N_A$)
  2. Diffusion speed of electrons and holes
  3. Recombination lifetime, τ, of process

$$e + h \rightarrow \text{nothing} \quad (+E_g)$$

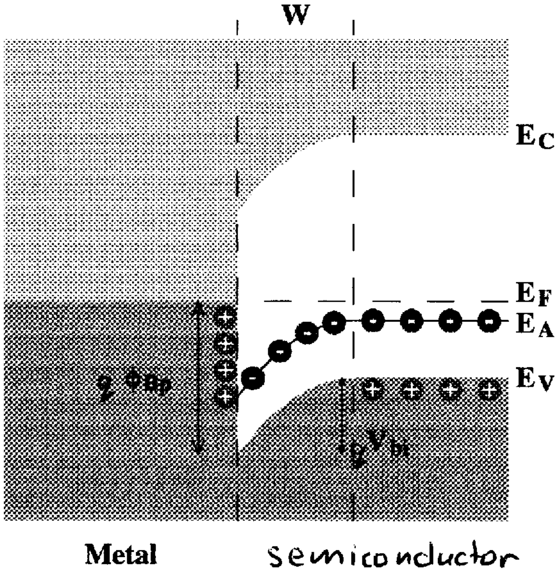
-



Schottky Barrier (Sze, ch. 5) (P.S. online)



Schottky barrier before contact



← P.S. online

Schottky barrier after contact

A Schottky barrier is a contact between a metal and a semiconductor and this will result in a diode ; the device conducts current only in one direction.

The same reasoning can be applied that we have used for the p-n junction : a difference in Fermi level will cauce a transfer of charge and hence the building up of a space charge region.
The figure above shows an example for a Fermi level of the metal higher than the Fermi level of the semiconductor. Electrons will move from the metal to the semiconductor, thereby winning energy.

These charges will build up a field and a voltage drop according to Poisson's equations

$$\frac{dE_x}{dx} = \frac{1}{\varepsilon}\rho(x) \Rightarrow E(x) = \int \frac{1}{\varepsilon}\rho(x)\,dx$$

$$V(x) = \iint \frac{1}{\varepsilon}\rho(x)\,dx^2$$

The electrons will continue flowing to the semiconductor until the voltage drop is equal to the built in voltage V_{bi} determined by the difference in Fermi level

$$qV_{bi} = \Delta E_F$$

Looking at the picture at the previous page we see that the Fermi level before contact is

- In the metal $E_F = E_{vac} - q\phi_m$

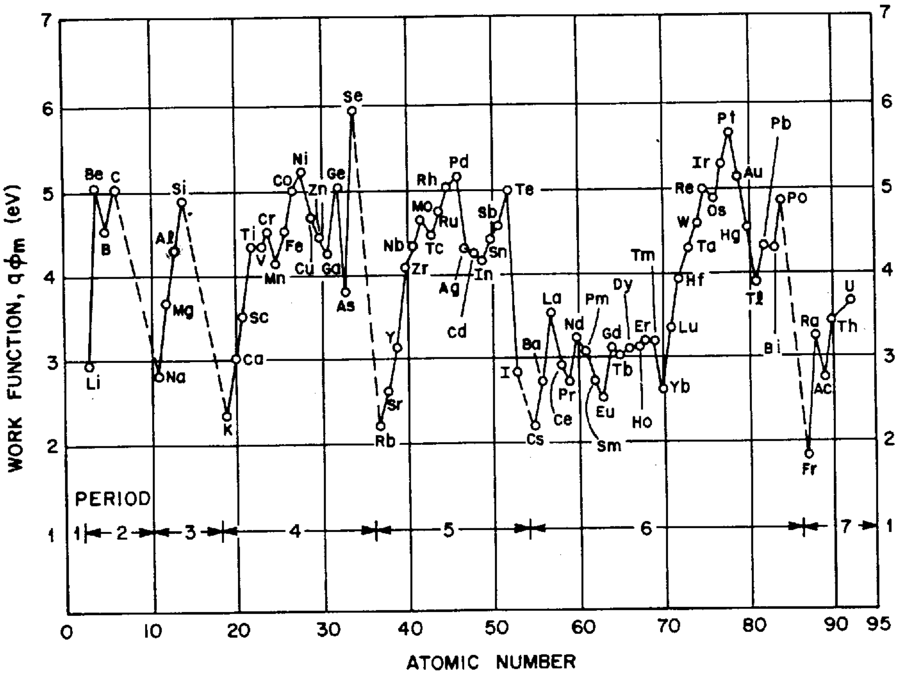
  with E_{vac} the vacuum level and ϕ_m the workfunction of the metal. This is the energy it costs to take an electron out of the metal. This is different for every metal. The figure on the next page summarizes the workfunctions for metals.

- In the semiconductor

  $$E_F = E_{VAC} - q\chi - qV_n$$

  with χ the electron affinity of the semiconductor. This is how much energy it costs to take an electron from the bottom of the conduction band and remove it from the crystal.

  V_n determines the Fermi level and is determined by the doping levels.



← Sze, p. 251

Metal work function for a clean metal surface in a vacuum versus atomic number.

The "band bending" or "built-in voltage" is therefore given by

$$V_{bi} = \chi + V_n - \phi_m$$

χ: electron affinity of semiconductor $((E_{VAC} - E_c)/q)$

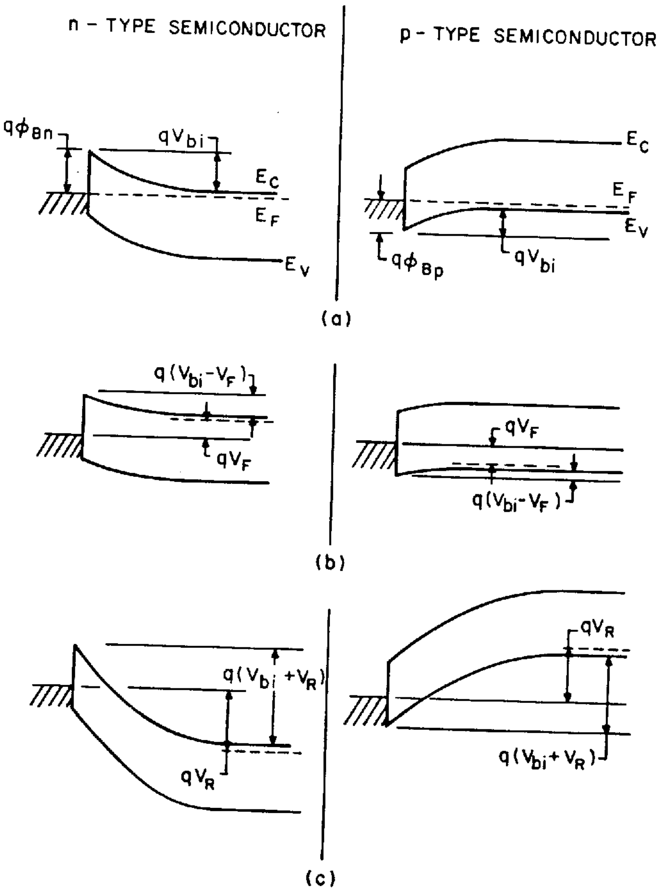
V_n: Fermi level depth $((E_c - E_F)/q)$

ϕ_m: metal work function.

The barrier height as seen from the metal side is called Φ_{Bp} and it can be shown to be $E_{F\,metal} - E_{V\,semicon}$:

$$q\,\Phi_{Bp} = E_g - q\,(\phi_m - \chi)$$

(The best way to visualize it is that we see the point where E_v meets the metal is glued there and the bands in the semiconductor are made of rubber and can be adjusted, "bend")

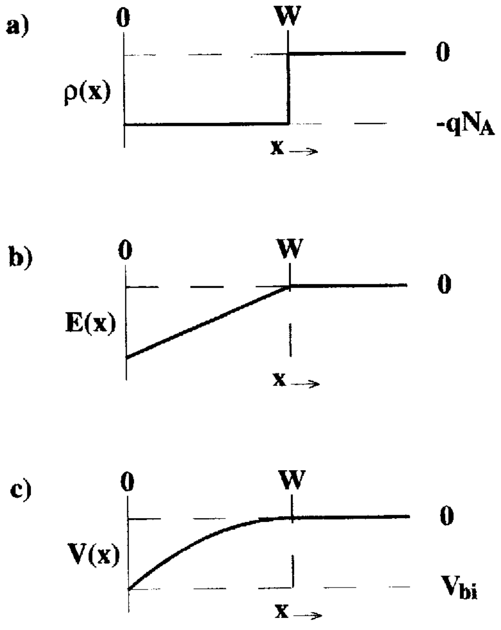


← Sze, p. 249

Energy-band diagram of metal *n*-type and metal *p*-type semiconductors under different biasing conditions (a) Thermal equilibrium. (b) Forward bias. (c) Reverse bias.

when a bias is applied, the total band bending changes to $(V_{bi} - V)$, see Figure above.

Depletion width of Schottky barrier: (p.s. online)



↑
p.s. online

Inside the depletion zone all ionized acceptors (N_A) are uncompensated by holes. The space charge density is therefore

$$\rho(x) = -qN_A$$

see figure a).

To get the electric field we use Poisson's equation

$$E(x) = \int_0^W \frac{1}{\varepsilon}\rho(x)\,dx = -\int_0^W \frac{N_A}{\varepsilon}\,dx$$

This becomes

$$E(x) = \frac{qN_A}{\varepsilon}(x-W)$$

To get the voltage we integrate the field:

$$V(x) = \int_0^W E(x)\,dx$$

$$= \left(\frac{qN_A}{2\varepsilon}\right)(x-W)^2$$

At $x=0$ this must be equal to the band bending $(V_{bi}-V)$:

Therefore

$$\boxed{W = \sqrt{\frac{2\varepsilon(V_{bi}-V)}{qN_A}}}$$

The amount of charge Q inside the depletion zone now follows $Q = A\cdot\int_0^W -qN_A\,dx$

$$= -\sqrt{2\varepsilon(V_{bi}-V)qN_A}\cdot A$$

with A the area of the Schottky barrier.

The capacitance can be defined as

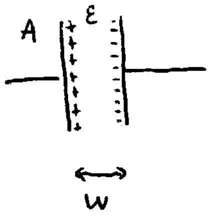
$$C \equiv \frac{dQ}{dV}$$

$$= A\sqrt{\frac{q\varepsilon N_A}{2(V_{bi}-V)}}$$

or

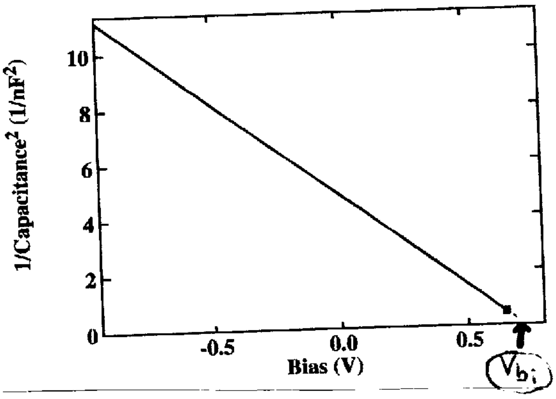
$$C = \frac{\varepsilon A}{W}$$

In other words, the Schottky diode kapacitance is equal to a system of two metal plates with area A and distance W, filled with dielectric ε.



From the same equation we can derive the Mott-Schottky equation:

$$\frac{1}{C^2} = \frac{2\,(V_{bi}-V)}{A^2\, q\, \varepsilon\, N_A}$$



The plot of $\frac{1}{C^2}$ vs. V will be linear and the slope of the plot is proportional to $\frac{1}{N_A}$. In fact, this is the standard way to determine doping concentrations. Moreover, the concentration we are measuring is the one at the edge of the depletion width.

$$N_A(W) = -\frac{2}{q A^2 \varepsilon \;\; d(\frac{1}{C^2})/dV}$$

and

$$W = \frac{\varepsilon A}{C}$$

In this way, the doping profile can be measured, $N_A(x)$.

Current of a Schottky diode (Sze p.254)

The current transport in metal-semiconductor contacts is mainly due to majority carriers (holes in p-type materials, electrons in n-type materials), in contrast to p-n junctions, where minority carriers are responsible.

The current through a Schottky diode is given by

$$J = J_{ST} \left[\exp\left(\frac{qV}{kT} \right) - 1 \right]$$

$$J_{ST} = A^* T^2 \exp\left(- \frac{q\phi_{Bp}}{kT} \right)$$

very similar to the p-n diode equation.

The saturation current now depends on the temperature T and the barrier height as seen from the metal, ϕ_{Bp}. A^* is the effective Richardson constant $(= 4\pi q m^* k^2 / h^3)$

Ohmic contacts (Sze, p 304)

In many cases, rectifying contacts (diodes) are not wanted. Instead, low resistance, ohmic contacts are preferred.

To create a low resistance Schottky barrier of a metal and a semiconductor, the choice of metal is important:

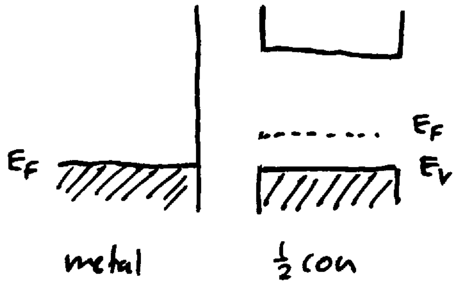
- The resistance of the contact at low voltages can be defined as

$$R_c = \left(\frac{\partial J}{\partial V}\right)^{-1}_{V=0}$$

$$= \left[A^* \cdot T^2 \exp\left(-\frac{q\phi_{Bp}}{kT}\right) \cdot \frac{q}{kT}\right]^{-1}$$

$$= \frac{k}{qA^* \cdot T} \exp\left(\frac{q\phi_{Bp}}{kT}\right)$$

- Therefore, a low barrier height ϕ_{Bp} is needed.

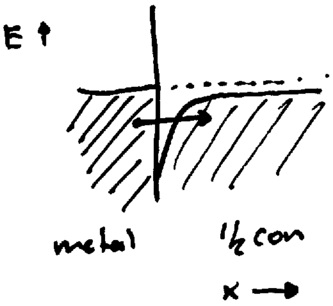


$\phi_{Bp} = 0$

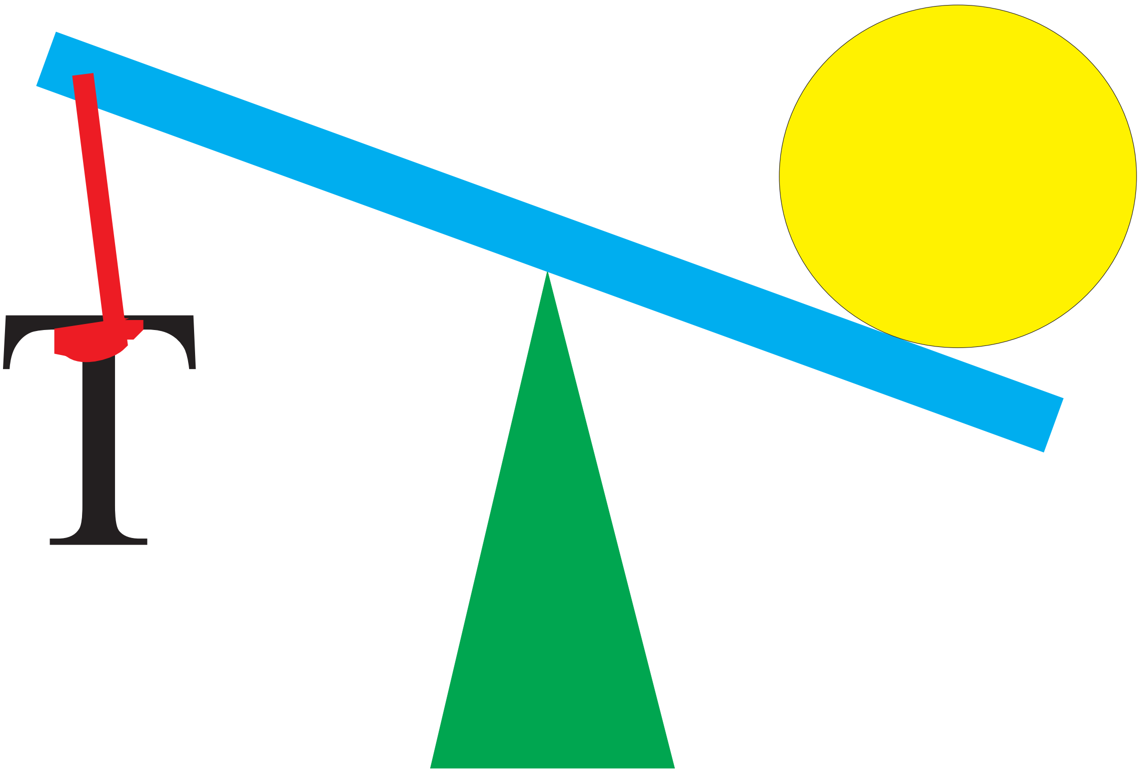
The optimal situation is when the Fermi level of the metal is aligned with the (majority corrier) band of the semiconductor. See the Figure on the left.

? Question: what metal gives a good contact for p-type silicon? And n-type silicon?

If a good metal is not available, an ohmic contact can be made by creating a highly doped layer at the interface. The depletion width, W, then becomes so narrow that carriers can tunnel through this layer directly into the carrier band in the semiconductor, see the Figure below.

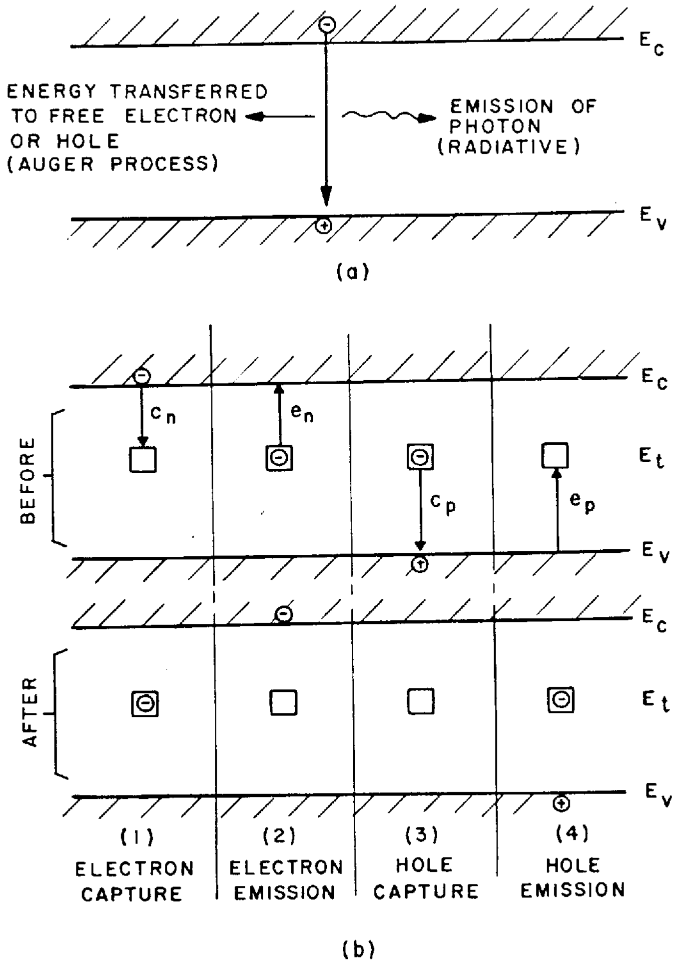


Non (Thermal) Equilibrium



(psonline) RECOMBINATION

Every time the distribution of electrons is disturbed, the system will restore to the normal situation of thermal equilibrium. Imagine a light-pulse generating electron-hole pairs. Slowly these electrons will recombine with the holes, thereby restoring equilibrium. The figure shows the processes of recombination. The top figure shows band-to-band recombination. The energy thus released can be emitted in the form of light ("radiative recombination") or can be absorbed by the electrons and or crystal ("non-radiative recombination"). The bottom picture shows capture and emission of electrons and holes on impurity levels.



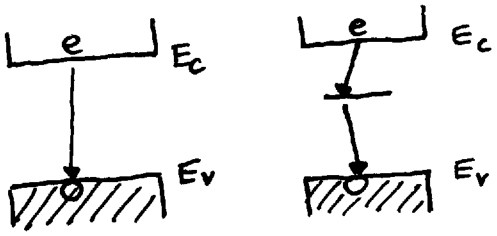
Recombination processes. (a) Band-to-band recombination (b) Single-level recombination.

(See. p. 36) ↑

Relaxation time

The time it takes to restore equilibrium depends on the temperature and on the depth of the level. For higher temperatures it becomes faster and, in general, shallow levels are faster than deep levels.

This is the reason why, to make electronics faster, often impurities are added to the semiconductor that create deep levels. Recombination via these levels (two-step recombination) make is faster than band-to-band recombination



← slow band-to-band versus fast two-step recombination.

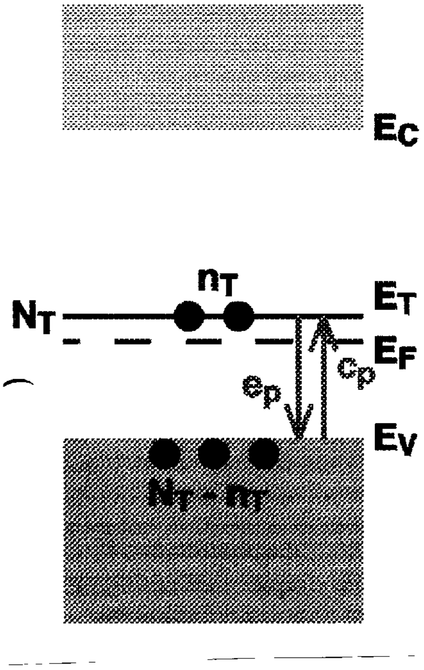
Such added impurities are called life-time killers and can make electronics substantially faster.

When the system is brought off equilibrium, holes and electrons will be captured and emitted from the levels. The speed at which this happens is given by (no derivation. see ps online):

http://www.ualg.pt/fct/adeec/optoel/theory/#emis

$$\boxed{e_p = \gamma T^2 \sigma_{pa} \exp\left(-\frac{E_{pa}}{kT}\right)}$$

e_p the emission rate (of holes in this example)



↑ psonline

γ and σ_{pa} constants

E_{pa} the level depth, $E_{pa} = E_T - E_V$

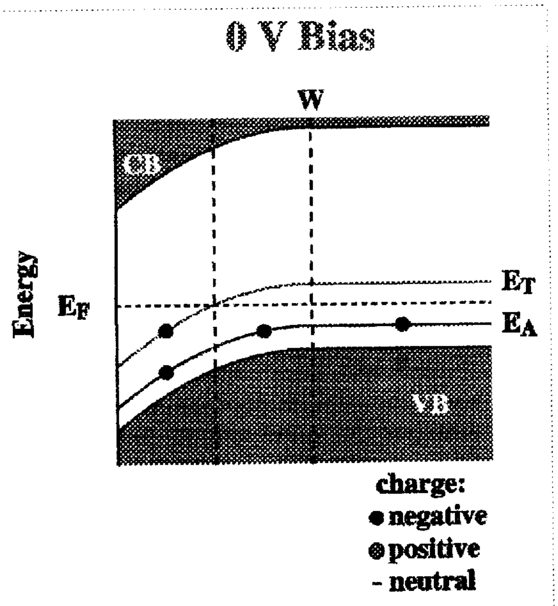
e_p is the fraction of holes trapped at E_T ($=n_T$) that are emitted per second.

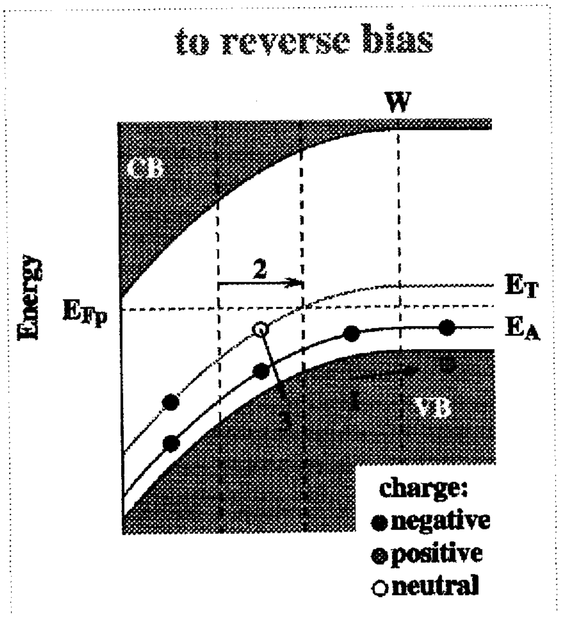
The capture of holes (c_p) is given by the relation

$$\frac{c_p}{e_p} = \exp\left(\frac{E_T - E_F}{kT}\right)$$

In normal conditions the capture of charges is much faster than the emission. Filling a level costs much less time than emptying it !

Capacitance-Transient spectroscopy (predecessor of DLTS)





In DLTS (deep level transient spectroscopy) the capacitance is monitored over time after a sudden change of bias. The above figures clarify what happens to the capacitance after a switch of bias. Easiest is it to understand if we remember that the capacitance is directly linked to the depletion width:

$$C = A\varepsilon / W$$

So, if for instance for some reason the depletion width increases, the capacitance decreases. Assume that the device is completely in thermal equilibrium at 0 V bias (left Figure). All the levels, even the deep ones, have their thermal equilibrium populations and a certain depletion width and capacitance is attained. (See the section on Calculation of the depletion width if you want to read again how to find the depletion width via double integration of Poisson's Equation $d^2V/dx^2 = \rho$). The depletion width W is determined by

$$\iint_W^0 \rho(x) d^2x = V_{bb}$$

The integration is done from $x=W$ to $x=0$. V_{bb} is the total band bending or voltage drop, $V_{bb} = V_{bi} - V$. Note that in the above Figures, the charge density is not constant; inside the depletion width there exists a smaller region where the deep level is below the Fermi level and where the space charge is larger.

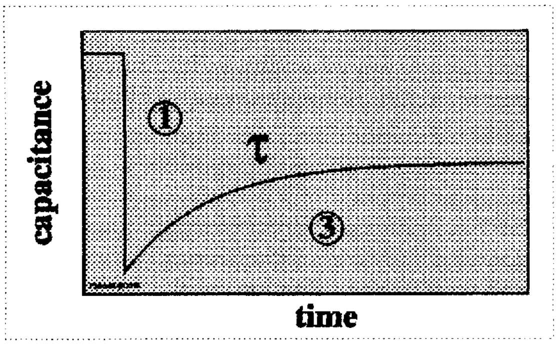
When the voltage is switched to reverse bias many things will happen. A different voltage drop V_{bb} occurs at the interface. This implies a new depletion width and a new capacitance according to the above equations. The new equilibrium capacitance takes time to reach. The following things will happen:

- 1: The free carriers (holes in the valence band) will rapidly move out of the interface region. The space charge increases.
- 2: A new, larger, depletion width is reached.
- Larger depletion width means smaller capacitance.
- Inside the new depletion width the deep level, which cannot respond so fast, is off equilibrium close to the interface.
- 3: Charges (holes) are slowly emitted from the deep level to the valence band at a time scale τ.

http://www.ualg.pt/fct/adeec/optoel/theory/

These charges are immediately swept away by the electric field.

- Because the space charge density increases closer to the interface, less depletion width is needed. (Visualize: the integration of Poisson's Equation reaches V_{bb} a little earlier). W shrinks.
- If W shrinks, C increases. This then happens at a time scale of the hole emission life time τ.



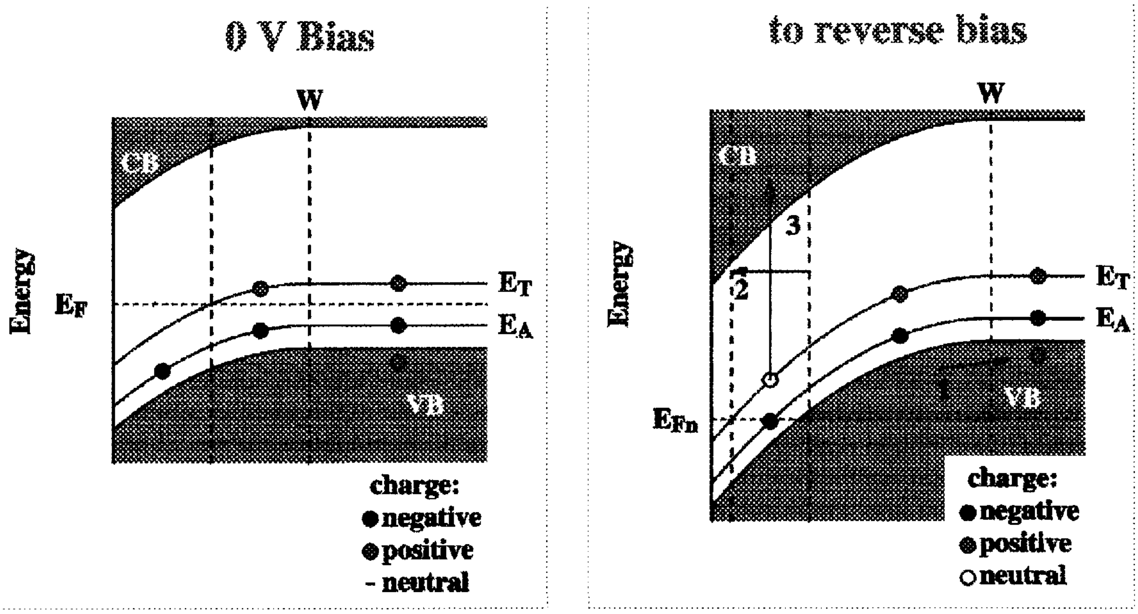
To summarize: after a switch of the voltage to reverse bias we first observe an immediate decrease of the capacitance (1) and after that a slow (partly) recovery of the capacitance at time scale τ (3). The explicit form of the time dependence of the capacitance is

$$C(t) = C_0 + \Delta C\exp[-t/\tau]$$

To find the trap activation energy, E_{pa}, we can use the equation for the emission rate e_p (=$1/\tau$) found in the section Emission and capture of carriers. $e_p = \gamma T^2\sigma_{pa}\exp(-E_{pa}/kT)$. If we monitor the characteristic decay times of the transients as a function of temperature and plot this in the form $\ln(\tau T^2)$ vs. $1000/T$, the resulting straight line will yield the activation energy.

Note: to truly measure the capacitance of the interface, C_d, we must chose a frequency low enough to be below the cut-off frequency of the bulk, see the section on admittance spectroscopy. For higher frequencies we would measure only the bulk properties, which have no transient effects.

capacitance transient of minority traps

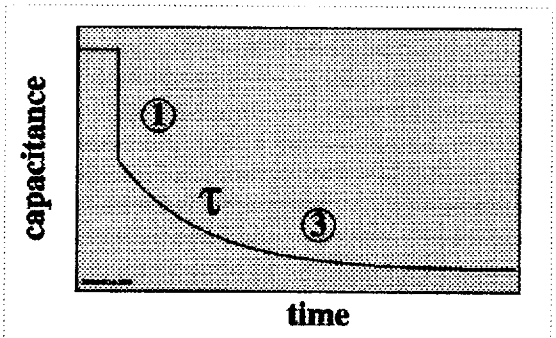


For minority traps the idea is similar. We have to keep in mind, though, that a minority trap is a trap that more readily communicates with the minority-carrier band than with the majority carrier band. For instance, an electron trap will thermalize with the conduction band rather than the valence band as described in the previous section. In other words, we have to draw the picture with the **minority** quasi Fermi level E_{Fn} rather than the majority quasi Fermi level E_{Fp}. As explained before, the minority quasi Fermi level moves in the opposite direction compared to the majority quasi Fermi level.

http://www.ualg.pt/fct/adeec/optoel/theory/

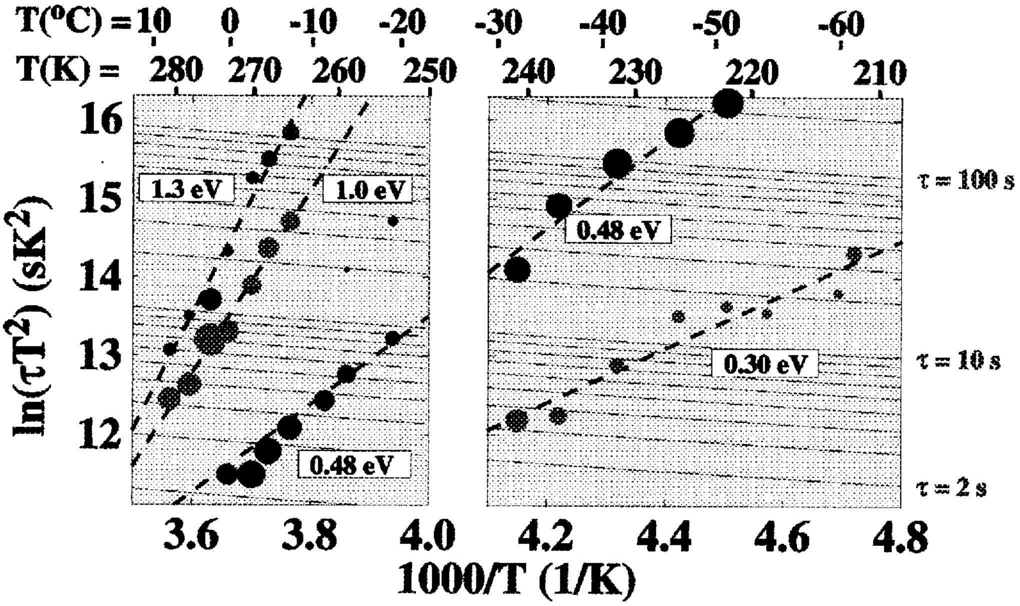
When the voltage is switched to reverse bias, the following things happen (the differences for minority carrier traps are indicated in a **bold** font):

- 1: The free carriers (holes in the valence band) will rapidly move out of the interface region. The space charge increases.
- 2: A new, larger, depletion width is reached.
- Larger depletion width means smaller capacitance.
- Inside the new depletion width the deep level, which cannot respond so fast, is off equilibrium close to the interface.
- 3: Charges (**electrons**) are slowly emitted from the deep level to the **conduction band** at a time scale τ. These charges are immediately swept away by the electric field.
- Because the space charge density **decreases** closer to the interface, **more** depletion width is needed. (Visualize: the integration of Poisson's Equation reaches V_{bb} a little **later**). ***W* increases**.
- If *W* **increases**, *C* **decreases**. This then happens at a time scale of the hole emission life time τ.



To summarize: after a switch of the voltage to reverse bias we first observe an immediate decrease of the capacitance (1) and after that a slow further decrease of the capacitance at time scale τ (3). This is very similar to majority carrier traps, but the sign of the transient is reversed; ΔC is positive here. We can still determine the trap activation energy by monitoring the characteristic transient decay time τ as a function of temperature.

As an example the following graph of energy levels in MEH-PPV on silicon substrates as determined by capacitance-transient spectroscopy (taken from P. Stallinga *et al.*, Synthetic Metals 1999; see reference list). The blue dots are majority-carrier traps with upward-trend transients and the red dots are minority-carrier traps with downward-trend transients. The size of a dot represents the transient amplitude (in the right panel it is multiplied by 5 for better visibility). The plot reveals several trap levels and shows the result of the first successful capacitance transient experiment in MEH-PPV.



DLTS (deep level transient spectroscopy[1])

Historical background

Deep level transient spectroscopy is probably the most used technique in the electrical characterization of semiconductors. This is based on the fact that it is fast, cheap in terms of equipment, and very sensitive. Commercially sold set ups exist in abundance.
The inventor of DLTS is D.V. Lang who wrote an article how to analyze capacitance transients in a systematic way, at that time enabled by the advent of modern computers with their data-crunching and experiment-steering powers. Looking back at it now, it can be said that "it was time" for such a technique to be invented. Lang did it and is now one of the most cited people in literature. His first article in the Journal of Applied Physics is still the best source for understanding the technique. Most things are described very elegantly there. I will do my best to describe it here. When the computers became more powerful more modern versions of DLTS were invented that use more of the data and use more and more-complicated calculations. The result is an increased resolution and sensitivity. The Laplace-DLTS technique is a good example. This uses the entire transient in a Laplace transformation to yield a much higher resolution. Obviously, for this a lot more calculating power is needed.

Basic principles

How DLTS works is the following. Capacitance-transients are recorded in the same way as described above. The device is placed in the emptying voltage V_E and the system is allowed to reach thermal equilibrium. For a short time the device is placed under the filling voltage V_F. In this times the traps will fill with charges. When the voltage is switched back to V_E a transient is observed as described in the previous sections. From this transient two samples are taken at times t_1 and t_2 after switching the bias. The DLTS spectrum is now the difference in capacitance at these two times as a function of temperature:

$$S(T) = C(t_1) - C(t_2)$$

This means (substituting the explicit form of the transient, see the section on capacitance-transient spectroscopy)

$$S(T) = [\exp(-t_1/\tau) - \exp(-t_2/\tau)]$$

The maximum of this signal occurs when the relaxation time τ reaches the value τ_{max} which can be found by differentiating $S(T)$ with respect to τ:

$$\tau_{max} = (t_2 - t_1) / [\ln(t_2/t_1)]$$

The explicit temperature dependence of τ is (see section on emission rates)

$$\tau(T) = \tau_0 T^{-2} \exp(E_{pa}/kT)$$

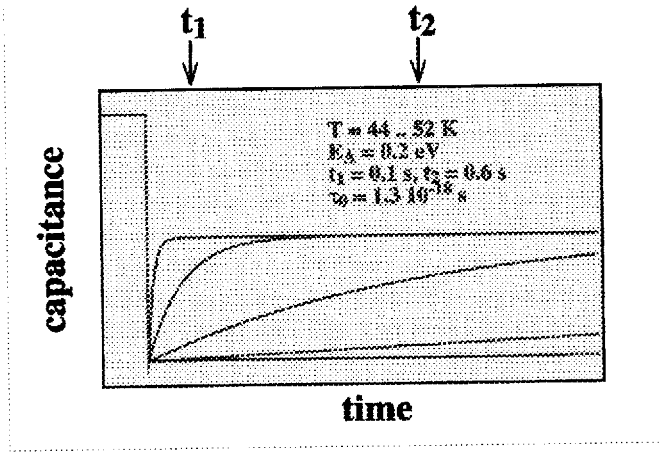
From this it is clear that the temperature at which the maximum DLTS signal occurs does not reveal the activation energy directly, unless τ_0 is known. If not, at least two measurements have to be made with different time windows (t_1, t_2) and the two points (T_{max}, τ_{max}) entered into the above equation will yield the trap activation energy E_{pa}.

Parameters of DLTS are:

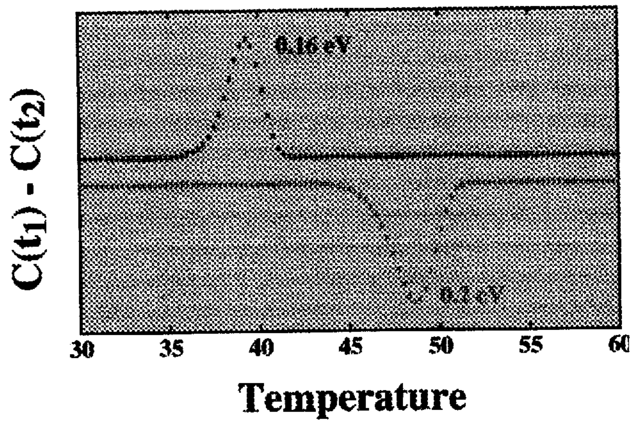
- Repetition rate: how many times per second a transient is recorded. The time between two transients should be long enough to ensure that the system is in thermal equilibrium.
- Time window: The values of the sampling-point times t_1 and t_2.
- Filling voltage V_F and emptying voltage V_E.

http://www.ualg.pt/fct/adeec/optoel/theory/

- Filling time: how long the device is placed in the filling voltage V_F before switching back to reverse voltage V_E.



Capacitance transients at several temperatures



Resulting DLTS signal for majority traps (down) and minority traps (up)

Example

The Figures show a simulation of DLTS. The left figure shows simulations of transients caused by a 0.2 eV deep majority carrier trap for various temperatures (44 K to 52 K in 2 K steps). For the lowest temperatures the transients are very slow (bottom trace) and the difference between the capacitance at t_1 and at t_2 is small, hence the DLTS signal $S(T)$ is very small. When the temperature is increased the transients become faster and the DLTS signal increases. Note that the signal has negative sign; for majority carrier traps, the transient has an upward trend and therefore $C(t_1)$ is smaller than $C(t_2)$. When the temperature is further increased, the transients become so fast that they already died out before the time window (t_1-t_2) and the DLTS signal has vanished again. The figure on the right summarizes this for the majority carrier trap (red dots; lower trace). In the same figure a comparison is given with a minority carrier trap (blue dots; upper trace) whose DLTS signal has the opposite sign because the related transients have downward trends. Also, in this simulation the minority-carrier trap is chosen to be a little more shallow (0.16 eV). Assuming that the pre-factor τ_0 is identical (1.3 10^{-18} s in both cases) it means that the peak has shifted towards lower temperatures.

Comments

The advantage of this scheme is that it can be done completely automatically with a minimum of data processing. The disadvantage is that of the entire transient only two points are used; most of the data are thrown away and the duty cycle of the measurement is very low. In Laplace DLTS the duty cycle is much higher because the entire transient is used in a Laplace transformation. The cost is a more elaborate calculation; only modern computers can be used.

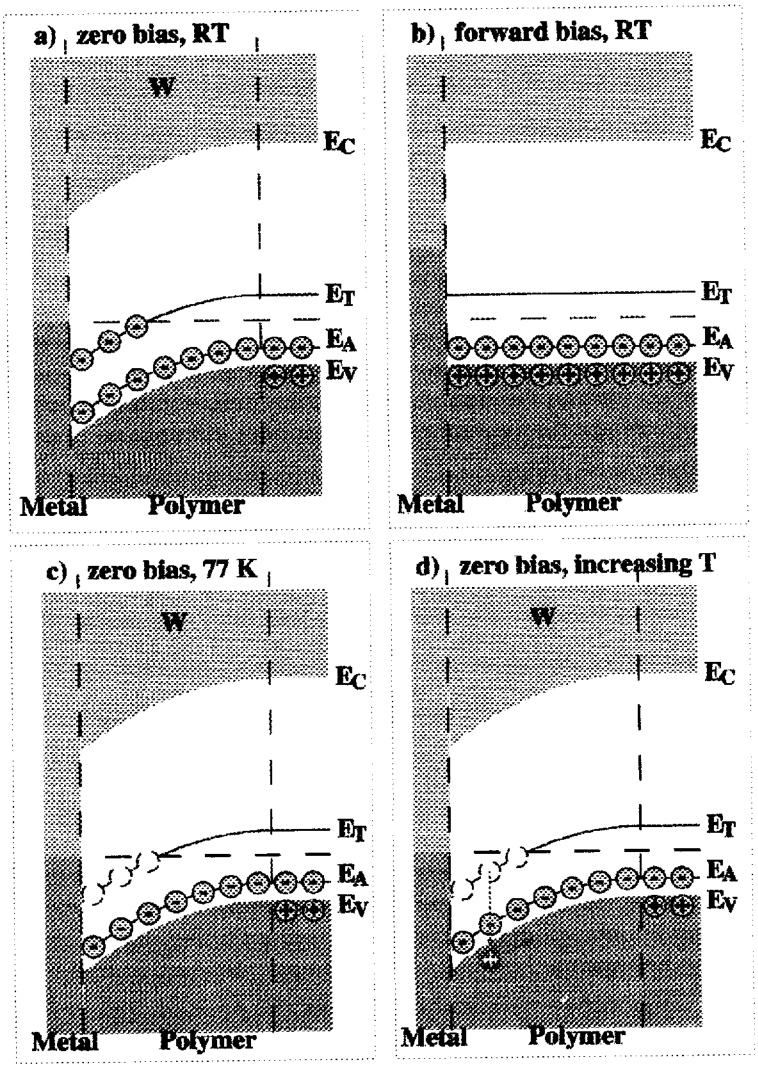
In reality, DLTS is more often used not to determine the trap activation energies but more to show the presence of certain impurities through their fingerprint spectra and to determine their densities through the intensity of the DLTS spectra and through the underlying amplitude of the transients. This is based on the fact that the amplitude of the transient is linearly proportional to the density of the deep defect relative to the density of the dopant (assuming that this ratio is small).

$$\Delta C/C = N / 2(N_A - N_D)$$

1: To be precise, the word spectroscopy is misplaced, because in a spectroscopy experiment an experimental value is monitored while the frequency is scanned; a spectrum is the visualization of an observable as a function of frequency. In DLTS no frequency is scanned, so the name is slightly inadequate. In DLTS (normally) the temperature is scanned and the DLTS-signal intensity is plotted. The end plot resembles a spectrum in that it shows peaks and that it is a "fingerprint" of the defect in the material. It would have been better if "scanning" was used for the S in the acronym DLTS.

TSC (Thermally Stimulated Current)

With temperature stimulated current experiments trap level depths and concentrations can be found. The underlying idea is that the current that is needed for restoring thermal equilibrium can be monitored. In practice this means that the sample is cooled down and the thermal equilibrium is disturbed in one way or another. This can be done either by applying a strong bias (forward or reverse, while cooling or afterwards) or by illumination. At low temperatures, the relaxation time of the sample is so large that it cannot relax back to its ground state. When the sample is warmed up, the charge emission rates become appreciable. The emitted carriers flow out of the interface region and combine into a minute, but observable, external current, to so-called TSC. When the temperature is further increased, the emission rates become very fast and the current increases. When thermal equilibrium is reached again, the charges are no longer emitted from the traps and the related current drops back to zero.



The Figures above show the procedure of Temperature Stimulated Current: a) The device in zero-bias at room temperature, b) The device is placed in forward bias to fill all the traps. c) The sample is cooled down and the bias is removed; all the traps are still frozen out. The dotted circles show the place where the traps are off-equilibrium. (In this simple picture, the Fermi level is considered to be independent of temperature; the shallow acceptors are not frozen out yet). d) The temperature is increased and at a certain temperature the charges (holes) are emitted from the traps. These charges are rapidly pulled out of the depletion zone and a current is monitored This continues until all the traps (within a certain region indicated by the dotted circles) are emptied.

From TSC, the activation energy of the trap can be determined in the following way (p-type

material):
In the depletion zone, there are no free carriers and the capture of carriers does not take place. Holes are emitted from the traps at a rate e_p and the number of filled traps n_t therefore changes with a speed

$$dn_t \,/\, dt = -e_p n_t \textbf{ Eq.1}$$

If we assume that all the charges that are emitted immediately drift towards the bulk, without being recaptured, all contribute to the current:

$$J = e_p n_t$$

When we make a temperature scan of this current, one important parameter is the temperature at which the maximum of the current occurs. At this place, the derivative of the current, $dJ/dT=0$. According to the above equation, the derivative of the current is equal to

$$dJ \,/\, dT = n_t\,(de_p/dT) + e_p\,(dn_t/dT) \quad \textbf{Eq.3}$$

Note that

$$dn_t \,/\, dT = (dn_t/dt)(dt/dT) = -e_p n_t/\beta$$

where Equation 1 was used and the scanning rate $\beta = dt/dT$ was defined. This can be put into Equation 3 and to find the maximum, the result should be set to zero.

$$n_t\,(de_p/dT - e_p^2/\beta) = 0$$

One trivial solution is $n_t=0$. This occurs at the end of the scan, when thermal equilibrium is restored. The other temperature, T_m, can be found by substituting the emission rate found before (see the section on emission rates), resulting in

$$(\sigma\gamma T_m^4/\beta)\exp(-E_a/kT_m) = (2T_m + E_a/k)$$

For $kT_m << E_a$

$$\ln(T_m^4/\beta) = E_a/kT_m + \ln(E_a/\sigma\gamma k)$$

To determine the trap activation energy E_a, a set of scans should be made with different scanning speeds β. Each time the temperature T_m at which maximum current occurs is noted and the slope of a plot of $\ln(T_m^4/\beta)$ vs. $1/T_m$ is then proportional to E_a/k.

The activation energy can also be found directly from a single TSC scan. A peak in a TSC scan follows the equation

$$I = \frac{A\exp(\Theta)}{[\,1 + B\exp(-\Theta)\Theta^{-2}\,]^2}$$

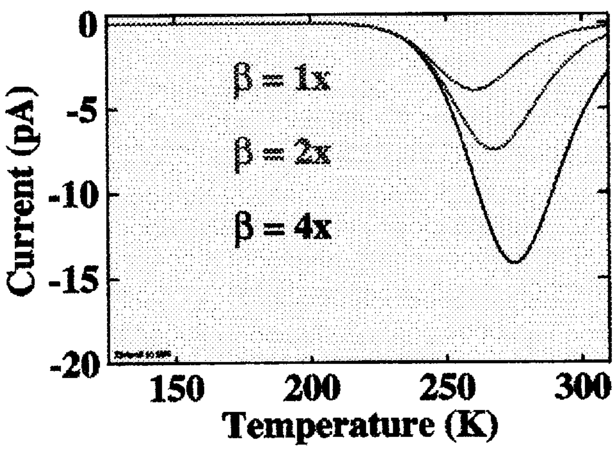
with A and B parameters depending on the properties of the carriers and the trap and the scanning rate β, and $\Theta = E_A/kT$ (see Karg *et al.*, 1999 ref). A is independent of scanning rate, while $B \sim 1/\beta$. The interdependence of the parameters makes the fitting very difficult, though.

In either case, the integrated current over time reveals the number of defects emptying their charge and this can then be related to the defect density if we know the dimensions (depletion width W^* and electrode area A) of the active region.

$$N_A = (1/eWA) \int I\,dt$$

The figure shows an example of a TSC measurement with parameters: $E_a = 0.5318$ eV, $A = -0.269$ A, $B = 9.49\ 10^{12}$ (β=1x), $4.74\ 10^{12}$ (β=2x), $2.37\ 10^{12}$ (β=4x). Note that the integrated curve (over **time!**) *is equal in all three cases. The biggest curve looks larger, but* *it is scanned at higher speed.*

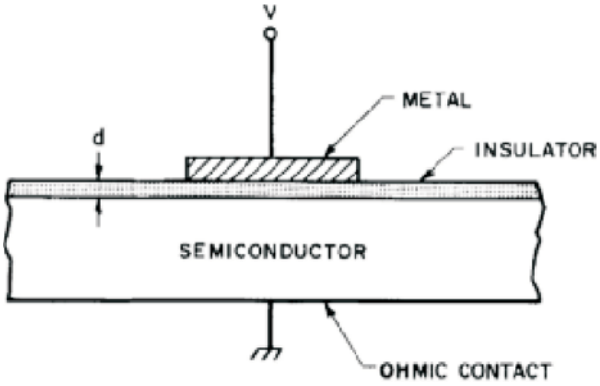
$$\int I\,\mathrm{d}t = \int I\,\mathrm{d}T\,\mathrm{d}t/\mathrm{d}T = (1/\beta)\int I\,\mathrm{d}T$$

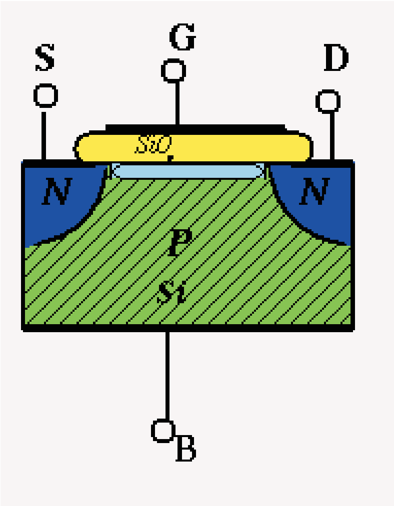


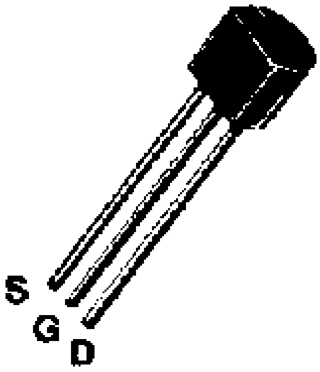
---

[*] We make here the assumption that all defects in the depletion width release their charge in a TSC experiment. More precise would be to use a value ΔW instead of W, where ΔW is the part of the depletion width where the deep level is forced above the Fermi level with the bias (and is below is without the bias!), see the figures a and b in the beginning of this section.

Devices II


V
d
METAL
INSULATOR
SEMICONDUCTOR
OHMIC CONTACT


S
G
D
SiO2
N
N
P
Si
B


S
G
D

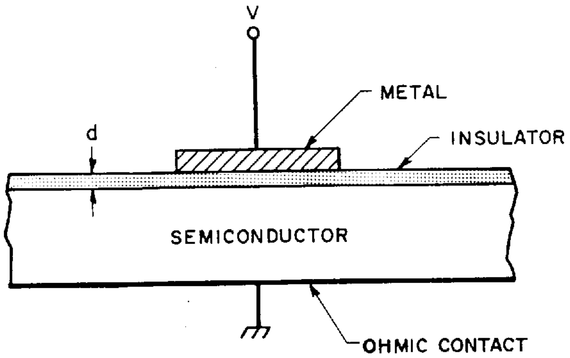
MIS diode (ch. 7 of Sze)

An MIS diode or metal-insulator-semiconductor is equal to a Schottky barrier with an insulating layer between the metal and the semiconductor.

The ideal MIS diode has no current. It's application lies in the fact that

- It is a precursor for the FET (field-effect transistor).
- It is an ideal device for studying interfaces.
- It is the basis for CCD devices (charge-coupled devices)

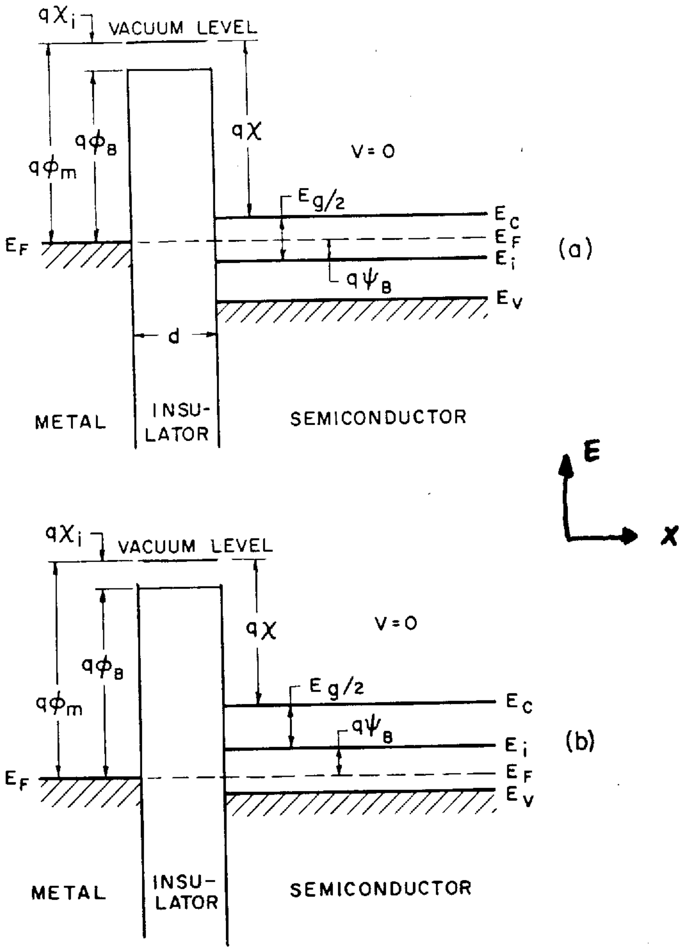
← (Sze, p. 363)



Metal–insulator–semiconductor (MIS) diode.

In the next analysis, it is assumed that the Fermi-levels in the metal and semiconductor are equal in the absence of bias. see next page. This implies no band bending and no space charge at $V = 0$ bias.

← Sze. p. 364



Energy-band diagrams of ideal MIS diodes at $V = 0$. (a) *n*-type semiconductor. (b) *p*-type semiconductor.

In the above figure.

ϕ_m : metal work function

ϕ_B : barrier height as seen from metal

χ : electron affinity of ½ con

E_i : midgap level $= (E_c + E_v)/2$

E_g : energy gap

ψ_B : $|E_i - E_F|$

d : insulator thickness

E_F : Fermi level

E_c : conduction band bottom

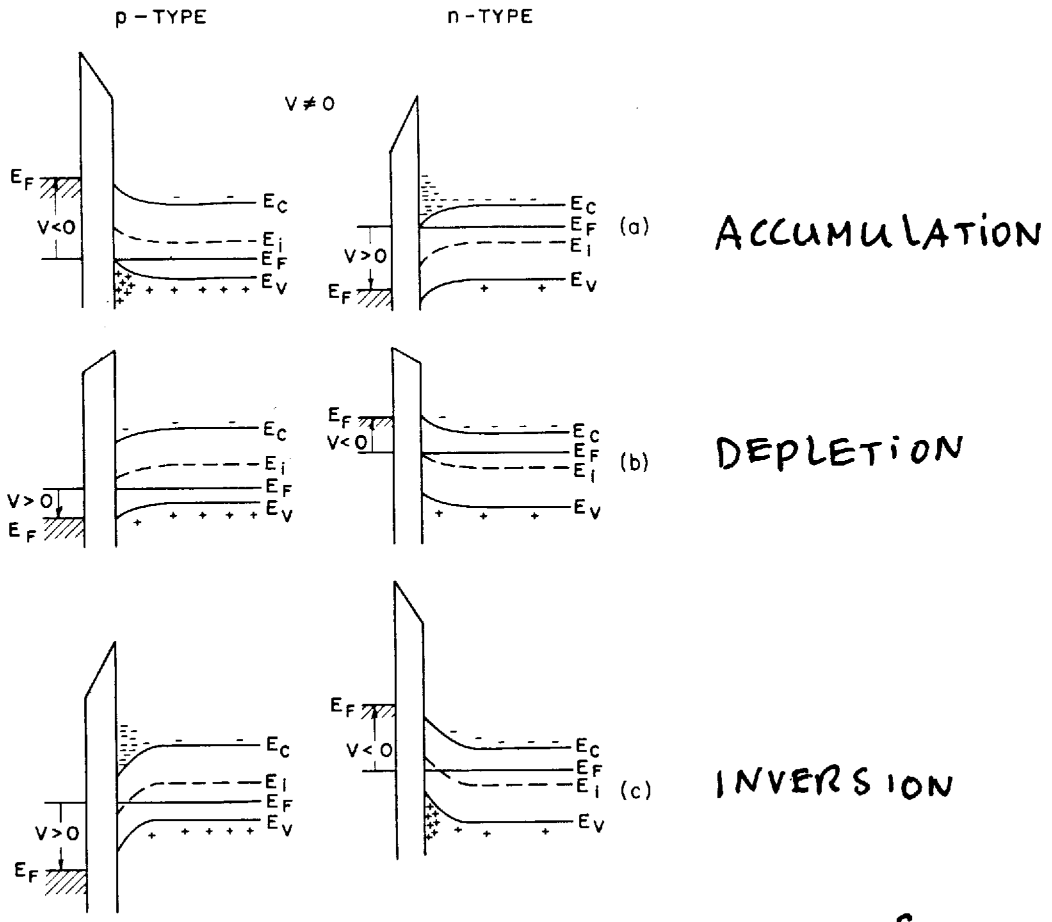
E_v : valence band top

For p-type semiconductor:

when a negative voltage is applied to the metal with respect to the semiconductor, holes are attracted to the interface resulting in a band bending as show in the figure below (a).

This is called accumulation. For a band bending larger than $E_F - E_V$, the Fermi level crosses the valence band at the interface and we have strong accumulation, with a large number of free holes at the interface.

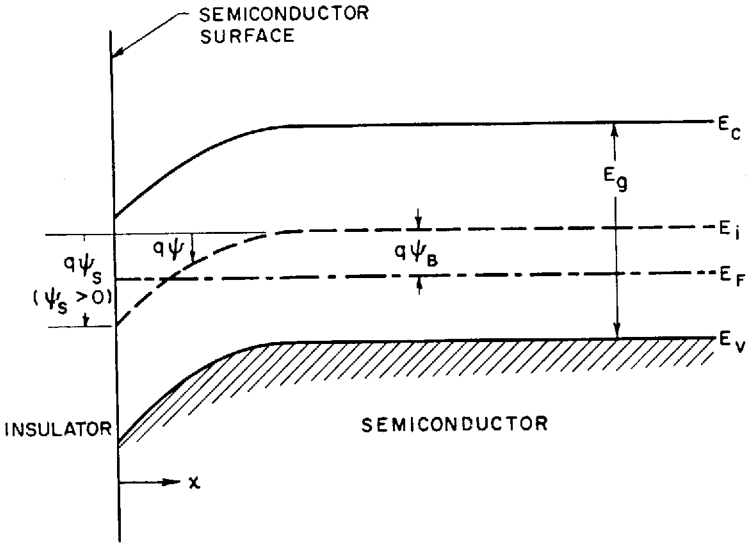
On the other hand, for a small positive voltage, the free holes are pushed away from the interface and the interface is in depletion (b). For stronger



Energy-band diagrams for ideal MIS diodes when $V \neq 0$, for the following cases: (a) accumulation; (b) depletion; (c) inversion. ← Sze, p. 365

voltages the Fermi level will be closer to the conduction band than to the valence band. Hence $n > p$ and this is called inversion. (c). For very strong voltages the Fermi level can enter the conduction band at the interface and a channel of electrons is formed. This effect is used in FETs, as we will see later.

← Sze, p. 366



Energy-band diagram at the surface of a *p*-type semiconductor. The potential ψ, defined as zero in the bulk, is measured with respect to the intrinsic Fermi level E_i. The surface potential ψ_s is positive as shown. (a) Accumulation occurs when $\psi_s < 0$. (*b*) Depletion occurs when $\psi_B > \psi_s > 0$. (c) Inversion occurs when $\psi_s > \psi_B$.

ψ_s is the amount of band bending

$\psi_s = 0$: flat band

$\psi_s = \psi_B$: intrinsic : $n = p$

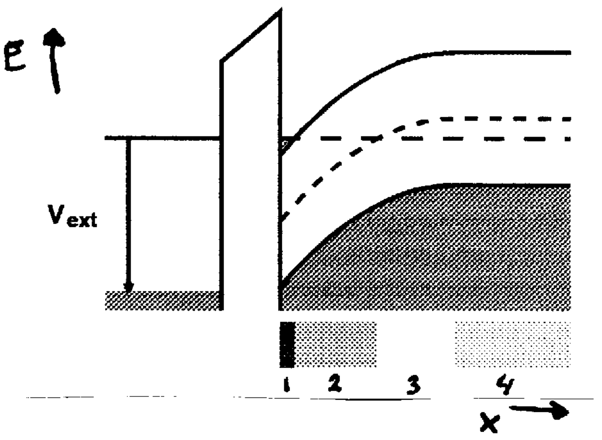
$\psi_B > \psi_s > 0$: depletion

$\psi_s < 0$: accumulation

$\psi_s > \psi_B$: inversion (shown above)

$\psi_s > 2\psi_B$: strong inversion

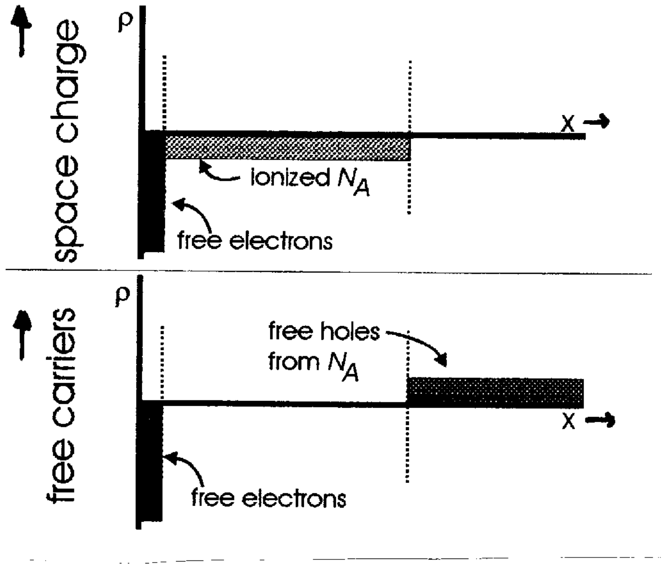
An example : p type semiconductor



The first image shows the voltage drop in the device. Note that there is also a voltage drop in the insulator.

In the semiconductor 4 zones are distinguished starting from the interface there is

1 : strong inversion
2 : inversion
3 : depletion
4 : bulk



The second figure shows the distribution of charge (top) and free carriers (bottom).

Capacitance of an MIS diode

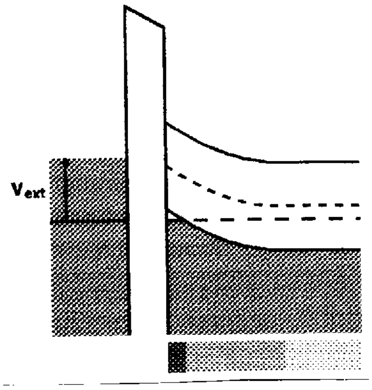
The MIS diode consists of various regions in series. The total capacitance follows the reciprocal rule $\left(\frac{1}{C_{tot}} = \frac{1}{C_1} + \frac{1}{C_2} \text{ , etc.} \right)$

The oxide capacitance is

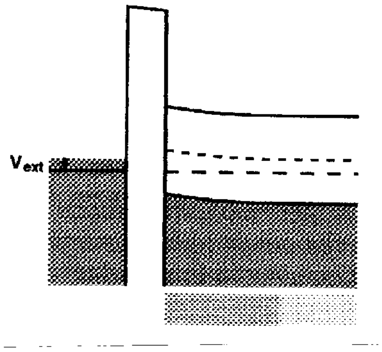
$$C_{ox} = \frac{\varepsilon_{ox} \cdot A}{d_{ox}} .$$

The depletion capacitance : $C_W = \frac{\varepsilon_s \cdot A}{W}$

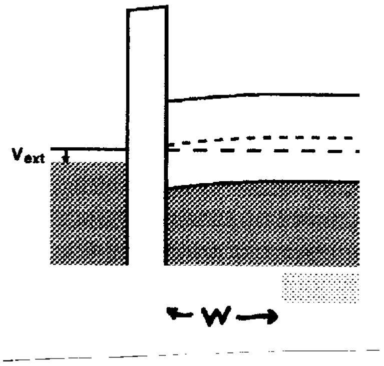
forward bias



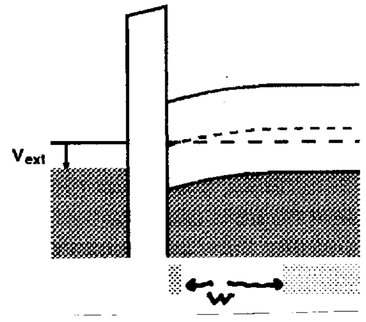
- **strong accumulation**
  The free charges reach all the way to the interface.
  $C = C_{ox}$



- **accumulation:**
  Equal to the situation above
  $C = C_{ox}$

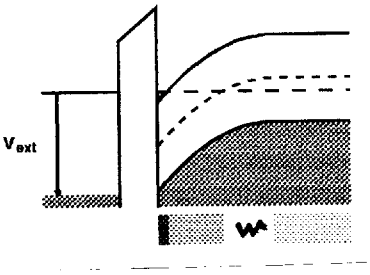


- **depletion:**
  the oxide capacitance in series with the depletion capacitance
  $$\frac{1}{C} = \frac{1}{C_{ox}} + \frac{1}{C_W}$$
  $\Rightarrow$ C decreases



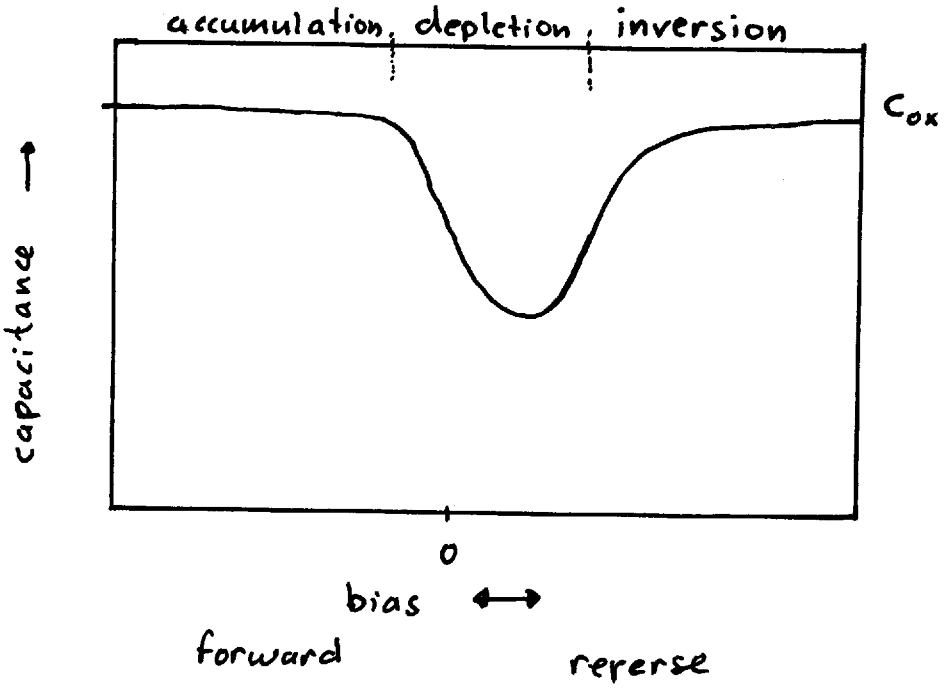
- **inversion**
  Similar to the situation above, but W shrinks, therefore C_W increases and C increases.

reverse bias



- **strong inversion**
  the depletion width shrinks to 0 and C returns to C_{ox}

* note: regions with free carriers do not contribute to C_{tot} (or $C = \infty$)

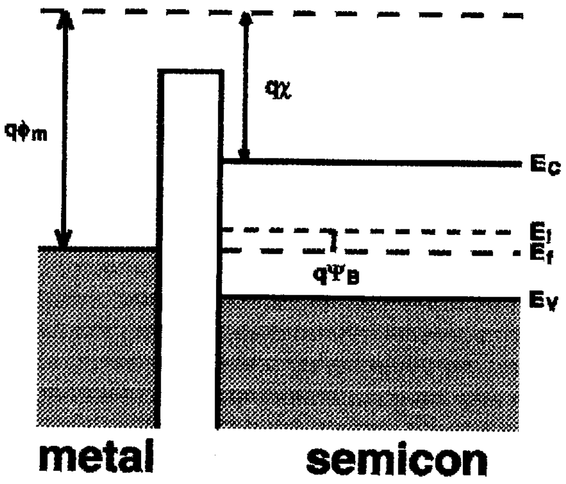

accumulation
depletion
inversion
C_ox
capacitance
0
bias
forward
reverse

← Sze, p 371

MIS junction

To understand how an FET (of the MOSFET type) works, we first have to analyze what a MIS structure is.
In a simple approximation, a MOSFET is a planar device of semiconducting material whose free carrier concentration *n*, and hence it conductivity (via $g = \mu_e n$) is controlled by the gate.

The first figure shows an energy diagram of an MIS device. The device consists of a metal and a semiconductor separated by an insulator. Here we assume that the insulator is not conducting any current.
Standard textbooks start with an ideal MIS device, which means that the Fermi levels in the metal and the semiconductor are aligned even before contact. In other words, the work functions of the metal and of the semiconductor are equal, $\phi_m = \chi + E_C - E_F$. This implies that we have no band bending in the absence of external voltages. Remember the Schottky barrier, where charge was flowing from one side to the other due to the misalignement of the Fermi levels. In the ideal MIS junction such a flow of charge is not needed (which saves us the trouble of explaining how it can pass through the oxide :-)



ϕ_m : the metal workfunction. How much it costs to take an electron from the top of the sea of electrons in the metal to vacuum. The "workfunction" of the semiconductor can be defined as $\chi + E_C - E_F$.

ϕ_b : (not shown, beause it doesn't enter into the calculations): the barrier height of the oxide as seen from the metal; the difference between the "conduction band" of the oxide and the Fermi level of the metal..

χ : the electron affinity of the semiconductor. The energy it takes to take an electron from the *conduction* band to the vacuum level.

E_G : energy gap of the semiconductor.

Ψ_B : the energy from the midgap level to the Fermi level. This is approximately equal to $E_G/2$

E_C : bottom of conduction band.

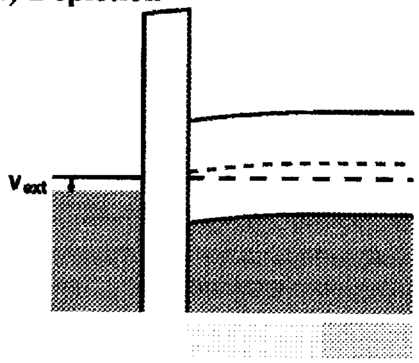
E_V : top of valence band.

The following strip explains what happens when we connect an external voltage to the metal (gate). For that we use the legenda as shown here on the right.

Note that the (horizontal) SCALE of the pictures is changing. Just like in Schottky barriers, the depletion width depends on the voltage. Here the depletion width is shown as constant.

- ■ **strong accumulation**
- ▨ **accumulation**
- ░ **normal (bulk)**
- □ **depletion**
- ▨ **inversion**
- ■ **strong inversion**
- ▨ **filled band**

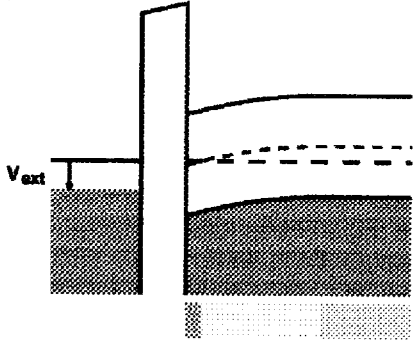
a) Depletion



When a small (positive) voltage is connected to the metal (gate), free holes are pushed out of the interface region and a band bending results on the other side of the insulator. This is the same effect as seen in a Schottky barrier. Free carriers (holes) flow out of the interface region and a **depletion** zone is formed. The uncompensated (negatively) ionized acceptors cause an electric field and - via Poissons equation - a parabolic bending of the bands.

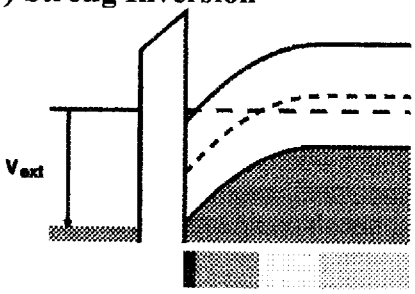
Note that there is also a voltage drop in the oxide. In the oxide no charges can reside and hence the field is constant and the voltage drop is linear in space. The total voltage drop in the oxide plus the band bending is equal to the external voltage.

b) Inversion



When the voltage is further increased, at the interface an **inversion** region is created. The semiconductor becomes here n-type, albeit not very conductive, yet.

c) Strong Inversion

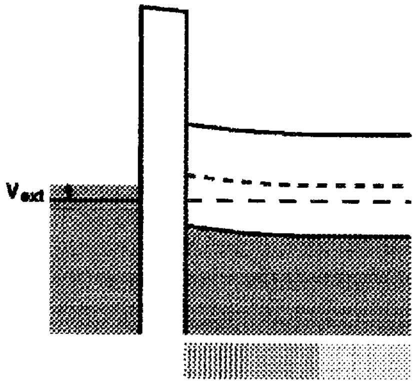


For very strong bias, the Fermi level crosses the conduction band close to the interface. This is **strong inversion**. Free electrons are in a so-called "channel" next to the oxide. This channel is therefore highly conductive.

The ample availability of states in the conduction band means that further increases in the gate voltage will not extend this strong inversion region into space, but rather will increase the *density* of electrons in the channels. The channel is always infinitisimally thin.The huge amounts of free carriers can easily cause a large voltage drop (band bending) and only a thin layer is needed to "absorb" the external voltage.

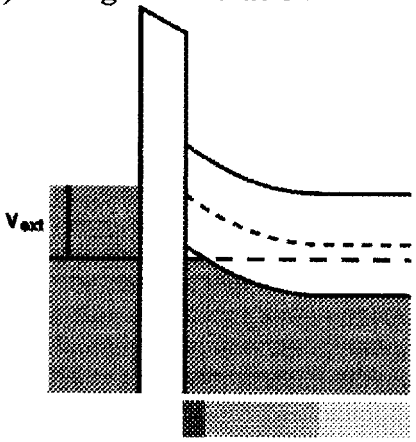
d) Accumulation

Going in the other direction, a negative voltage at the gate metal will attract some free holes to the interface. In this **accumulation** region, an increased number of free holes is created, although it



is not yet dramatic; the interface is still high-ohmic.

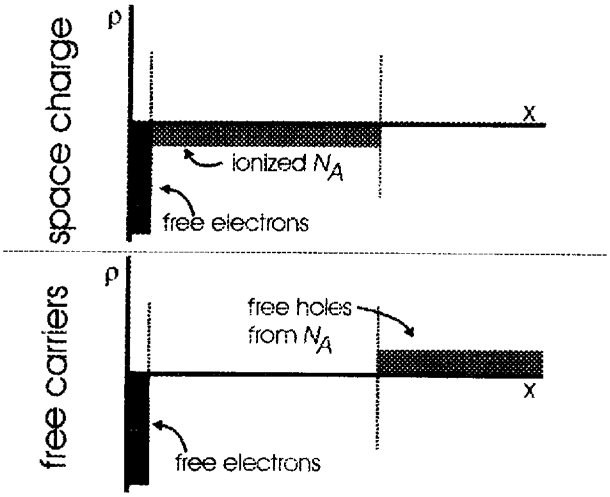
e) Strong accumulation



For stronger negative voltages at the gate the Fermi level is forced below the valence band at the interface. A channel with high density free holes is created. This is called **strong acummulation**.

Again, as for the strong inversion case, further increasing the voltage will result in an increase of the density of the free holes rather than an increase of the channel in space. In the picture the width of the channel is exaggerated. In reality it is only a monolayer thick.

To complete the story, in **strong inversion** we have the following distribution of space charge and free carriers:



The MIS junction is an ideal device for studying interface states and deep levels using the same techniques as demonstrated for the Schottky barrier. The advantage lies in the fact that there is no DC current which might obscure our measurements.

Capacitance of an MIS junction:

The device can consist of several regions in series, each with it's own capcitance. The capacitance of the metal and of any part of the semiconductor with free carriers is zero. For the other two parts of the device, the insulator and the depletion region we can calculate:

$C_{ox} = \varepsilon_{ox} / d$

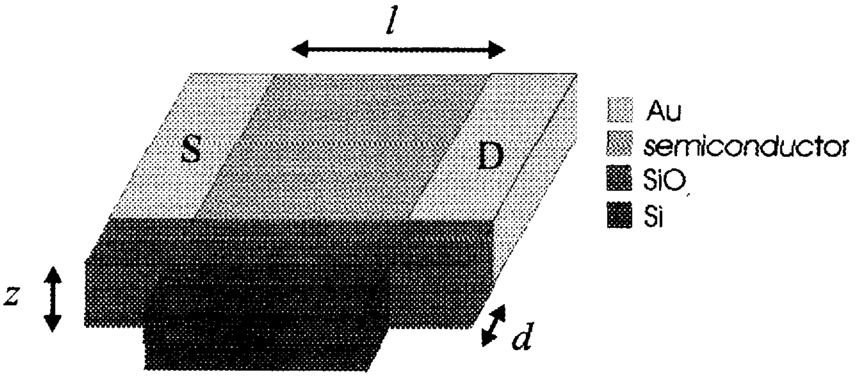
$C_W = \varepsilon_s / W$

The total capacitance is caluclated by taking the serial sum of the capacitances:

$$\text{depletion:} \quad C_d = \frac{C_{ox} \times C_W}{C_{ox} + C_W} \qquad \text{accumulation:} \quad C_a = C_{ox}$$

It is easy to see that the maximum capacitance is equal to the oxide capcitance and occurs for accumulation and strong inversion. In the latter case, the width of the zone with depletion can be so small that the it's capacitance is infinite and the total capacitance reduces to C_{ox}. See Fig. 7 on p. 371 of Sze.

Field Effect Transistor (FET)



Directions:
x: from oxide to surface of film

y: from source to drain
z: along an electrode

Z: electrode width
L: electrode distance
d: oxide thickness

An FET is nothing more than an MIS device with electrodes connected on two lateral sides of the semiconductor. Imagine connecting an electrode above and one below the images above. The first one we will call the source, the latter the drain.

The total current through this device is then linearly proportional to the number of free carriers in the semiconductor. The other parameters are the mobility μ, the external electrical field E_x: and the device dimensions Z and L:

$$I_{DS} = Z|Q|\mu E_x = Z|Q|\,\mu V_y/L \qquad (1)$$

with I_{DS} the total current, μ the mobility of the carriers, V_y the external voltage (is equal to V_{DS}), Z the width of an electrode, L the distance from source to drain, and $|Q|$ the amount of *free* charge in a line perpendicular to the oxide surface (along x) in C/m^2. This is equal to integrating the electron and hole density (ignoring the sign) from the edge of the oxide to infinity (or to the limits of the semiconductor film), see the figure with free carriers above.

In case we have substantial contributions from both electrons *and* holes, and they have different mobilities, we have to replace the part $|Q|\mu$ with $(|Q_p|\mu_p + |Q_n|\mu_n)$.

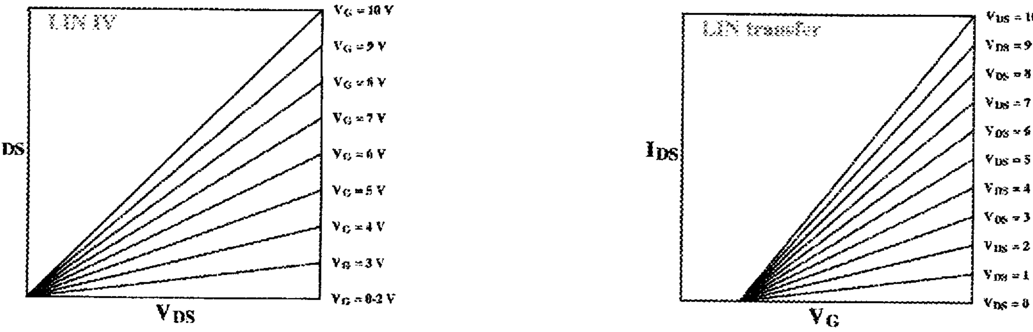
We will see that this can explain the linear region of an FET. To calculate the IV curves in the linear region, we only have to calculate the number of free carriers. For the saturation region the equations are a little different, as will be shown later.

Linear region

As a first order we can say that we only have free carriers when we are in strong inversion or strong accumulation. As can be seen from the strip of figures above, we need to supply a certain voltage to the gate to induce either strong inversion or strong accumulation.This is the so-called threshold voltage V_T. This doesn't mean that we don't have free carriers and conduction below this threshold voltage, but that the conduction beyond this voltage is much larger.

Under the assumption that all the free carriers come from charges in the strong accumulation or depletion zone, it is very easy to calculate the amount of charge on the semiconductor side and hence the current. In the section above we have seen that the capacitance of an MIS junction under these conditions is equal to the oxide capcitance, C_{ox}. Assuming no other capacitances, all the charges go to the interface and therefore, since $C = Q/V$ we find $Q = C_{ox} V_G$. As written above, we need V_T to bring it into strong accumulation or depletion, therefore $Q = C_{ox} (V_G - V_T)$. With this in mind, and with the idea that the conduction is proportional to the charge as seen in equation I the current in the **linear region** becomes

$$I_{DS} = C_{ox} (V_G - V_T) \mu V_{DS} Z/L$$



Examples of IV curves (I_{DS} vs. V_{DS}; left) for different gate voltages and transfer curves (I_{DS} vs. V_G; right) for different drain-source voltages. $V_T = 2$ V.

Remember that we arrived at this by the assumption that all the free carriers are located in the strong inversion or accumulation zone close to the interface and that the rest of the device is not conducting. Especially in materials with bulk Fermi levels close to the band edges, the background conductivity (the "off current") can be very large. In principle, the purer the material, the closer the Fermi level will be to mid gap and the less off-current we will have. Making the films thinner also helps.

Threshold voltage

Following the above discussion and the MIS strip of figures, it is clear that the band bending cannot be larger than the bandgap before either strong accumulation or strong inversion occurs. In other words, at the onset of strong inversion, the band bending V_{bb} is exactly equal to the difference between the conduction band and the Fermi level, $E_C - E_F$.The depletion width is then (see section on Schottky barrier) $W = (2\varepsilon_s V_{bb} / qN_A)^{1/2}$. The electrical field at the interface is the integral of the space charge inside this depletion region: $E_{max} = (qN_A/\varepsilon_s)W = (qN_A 2V_{bb} / \varepsilon_s)^{1/2}$.
Maxwell's equations tells us that, in the absence of any charge, the displacement D ($=\varepsilon E$) is continous accross the interface. On the semiconductor side of the junction we have $D_s = \varepsilon_s (qN_A 2V_{bb}/\varepsilon_s)^{1/2}$. In the oxide we therefore have an electrical field of $E_{ox} = (1/\varepsilon_{ox})(q\varepsilon_s N_A 2V_{bb})^{1/2}$. This field is constant inside the oxide because there are no net charges there. We therefore have an extra voltage drop of $\Delta V = d_{ox} E_{ox} = (d_{ox}/\varepsilon_{ox})(q\varepsilon_s N_A 2V_{bb})^{1/2}$ accross the oxide. The total external voltage then becomes $(d_{ox}/\varepsilon_{ox})(q\varepsilon_s N_A 2V_{bb})^{1/2} + V_{bb} = V_T$.. With V_{bb} equal to $\frac{(E_C - E_F)}{q}$ this becomes

$$V_T = (d_{ox}/\varepsilon_{ox})[2q\varepsilon_s N_A \frac{(E_C-E_F)}{q}]^{1/2} + (E_C-E_F)/q$$

Remember that $C_{ox} = \varepsilon_{ox}/d_{ox}$, and, in normal cases, the Fermi level is very close to the valence band, so that $E_C - E_F = E_G$ which we can call $2\psi_B q$. With this in mind, the above equation becomes the following text book equation for an **n-type inversion channel:**

$$V_T = (4q\varepsilon_s\psi_B N_A)^{1/2} / C_{ox} + 2\psi_B$$

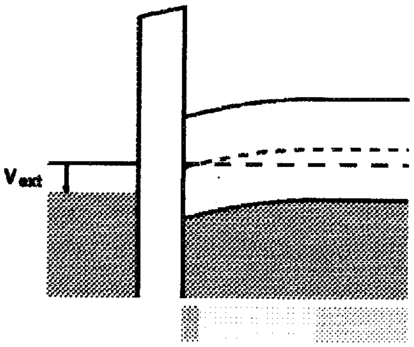
As an example: for $N_A = 1\text{x}10^{16}$ cm^{-3}, $C_{ox} = 50$ µF/m^2, $E_G = 2.5$ eV ($\psi_B = 1.25$ V), $\varepsilon_s = 5.0\varepsilon_0$ we get $V_T = 11.9$ V + 2.5 V = 14.4 V; since the first term in the equation comes from the oxide and the second term comes from the semiconductor we can see that most of the external voltage (83%) is absorbed by the oxide (remarkably).

The Fermi level, of course, also depends on the acceptor concentration, so in total we get a complex dependence on N_A. In textbooks, it is assumed that all acceptors are ionized. In fact, for organic materials, where the acceptor level can be very deep (but abundant) this is not necessarily the case. In fact, we should read N_A as "ionized levels" and this can be a fraction of the real number of acceptors, unlike in classical materials such as Si and GaAs. Moreover, the fraction of levels ionized can be changed by changing the band bending (by applying a gate voltage) and thus, the threshold voltage can change when putting the device in operation. Depending on the depth of the levels, these changes can be in the order of seconds, to minutes and hours and even days.

Note that high levels of current have no effect on this trapping of charges, since they do not attribute to a different band bending or extra charges in the interface.

Note that for an accumulation type FET this calculation does not work because we do not have a space-charge region caused by ionized acceptor levels; V_T does then not depend on N_A., but only on the Fermi level and the number of valence band states.

Subthreshold region



The substreshold currents are controlled by the barriers we have at the source and at the drain. When the channel opens, the barrier at the source disappears, as we have seen before, while the barrier at the drain is only a barrier for holes and not for our electrons. In fact, before threshold voltage, the device resembles an npn bipolar transistor (with floating base, E=S and C=D) and we

Saturation

If we increase the drain-source voltage eventually we will go into saturation where the current becomes independent of the drain-source voltage It is very easy to understand why this should happen. Imagine an FET with a threshold voltage of $V_T = 1$ V. We will apply a gate voltage well beyond this value, let's say 10 V. The channel is therefore open. For low drain-source voltages, both the drain and the source and hence the entire bulk region of the device is at 0 V, or close to 0 V at all places.

At the drain, the gate-drain voltage is 10 V and this is well beyond the threshold voltage. Here we have a conducting channel.

At the source (which is always at 0 V), the gate-source voltage is also 10 V and here the channel is also open. Actually, at all places do we have a conductive channel.

Let us now set the drain voltage to 10 V. At the source we still have a gate-source voltage of 10 V and the band bending here will still induce a channel. However, at the drain, the gate and the drain are at the same voltage; the gate-drain voltage is 0 V and this is below the threshold voltage for creating a conductive channel. The channel is closed here. This we call **pinch-off**.

If we assume a linear voltage drop from source to drain from 0 to 10 V, we can exactly calculate where the channel closes, namely at 10% distance from the drain. We have to bear in mind, though, that a closed channel has less free carriers and thus has higher resistivity. The law of continuity tells us that the current at every place of the device must be equal. Therefore, since voltage drop is current times resistivity, the voltage drops much faster in a closed-channel region than in a open-channel region. Moreover, if we assume that the free-carrier density in the closed-channel region is much smaller than the open-channel region, the pinch-off region is very small.

The voltage of start of saturation is easy to calculate. At this voltage the gate-drain voltage is exactly equal to the threshold voltage. $V_G - V_D = V_T$. Thus ($V_S = 0$, $V_D = V_{DS}$)

$$\text{onset of saturation: } V_{DS} = V_G - V_T$$

Beyond this voltage, the current is constant. We will have a region from the source to near the drain with a voltage drop of $V_G - V_T$ and a very thin region close to the drain with the rest of the voltage drop of VDS. The length of the first region is nearly independent of the total external voltage. It therefore has a contant length and voltage drop, and the current through it is therefore independent of V_{DS}. Again, continuity tells us that the current at every point is equal, therefore, the saturation current is independnet of V_{DS}.

With this we can calculate the free carrier density $Q(x)$ and the voltage $V(x)$ of every point in the channel at the onset of saturation. Along the way we will also calculate the current which is the current in saturation.
Imagine putting up a screen at a distance x from the source, peprpendicular to the source-drain direction, stretching as long as the electrode lengths (Z) and as high as the film thickness. We can calculate the current I through such a screen.
The current I at a crossection at a certain point x is equal to the free-charge density at that point $Q(x)$, the carrier mobility μ and the local field $E(x) = \mathrm{d}V(x)/\mathrm{d}x$:

$$I(x) = Z\mu Q(x)\, \mathrm{d}V(x)/\mathrm{d}x$$

Because of continuity, this current has to be constant along x and equal to the drain-source current I_{DS}. The above equation then reduces to

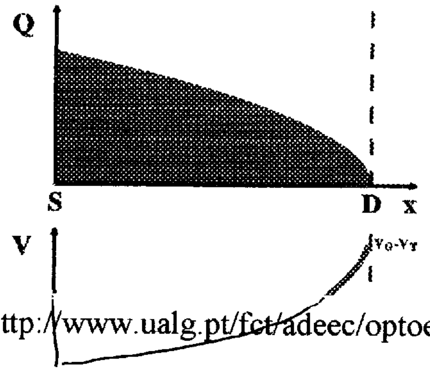
$$\mathrm{d}V(x) / \mathrm{d}x = I_{DS}/Z\mu Q(x) \qquad \text{(eq.I)}$$

The free-carrier density is a function of the local field $V_G - V(x)$, as seen before:

$$Q(x) = C_{ox}[V_G - V(x) - V_T] \qquad \text{(eq.II)}$$

The solution to this system of diferential equations is easy (take the derivative of the second equation and put this in the first), to which we add the following boundary condition:

1a: the charge at the drain is zero: $Q(L) = 0$.
1b: The voltage at the drain is just enough to close the channel: $V(L) = V_G - V_T$.



The soulution is then

$$Q(x) = [(2C_{ox}I_{DS}/Z\mu)(L-x)]^{1/2}$$

$$V(x) = (V_G - V_T) - [(2I_{DS}/\mu Z C_{ox})(L-x)]^{1/2}$$

Using the next boundary condition

2a: The voltage at the source is 0: $V(0) = 0$
2b: The charge at the source is $Q(0) = C_{ox}\,(V_G -$

we can calculate the current

$$I_{DS} = (1/2)\mu(Z/L)\, C_{ox}\, (V_G - V_T)^2$$

which is the current at the onset of **saturation** and, since, as we have shown before, the current beyond this point is independent of V_{DS}, this is the current <u>in</u> saturation (as long as $V_G < V_{DS} + V_T$).
The total amount of charge stored in the device in saturation can be found by integrating the equation for the charge distribution above and multiplying by the device width Z:

at saturation, $V_{DS} = V_G - V_T$: $Q = (2/3)\, Z\, (V_G - V_T)\, C_{ox}$

in the same way:

at $V_{DS} = 0$ V : $Q = Z\, (V_G - V_T)\, C_{ox}$

Note: when we increase the bias, charge is coming out of the device. We have to be careful in our IV measurements that we do not measure the current of these charges going into and coning out of the device. See the section on displacement current.

We can also use the same method to calculate the current before saturation. Going back to the general solution of the system of differential equations:

$V(x) = (V_G - V_T) - [(2I_{DS}/\mu Z C_{ox})(x_0 - x)]^{1/2}$

As boundary conditions we now take:

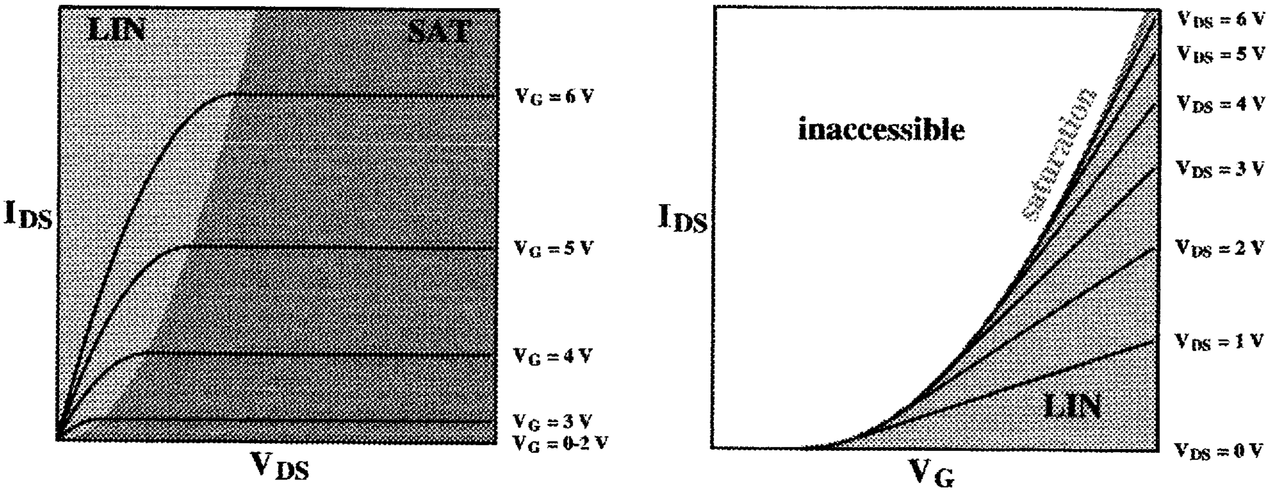
1: The voltage at the source is 0; the charge is $C_{ox}\, (V_G - V_T)$

2: The voltage at the drain is $V(L) = V_{DS}$

The result is

$$I_{DS} = \mu(Z/L)\, C_{ox}\, [(V_G - V_T)V_{DS} - (1/2)\, V_{DS}^2\,]$$

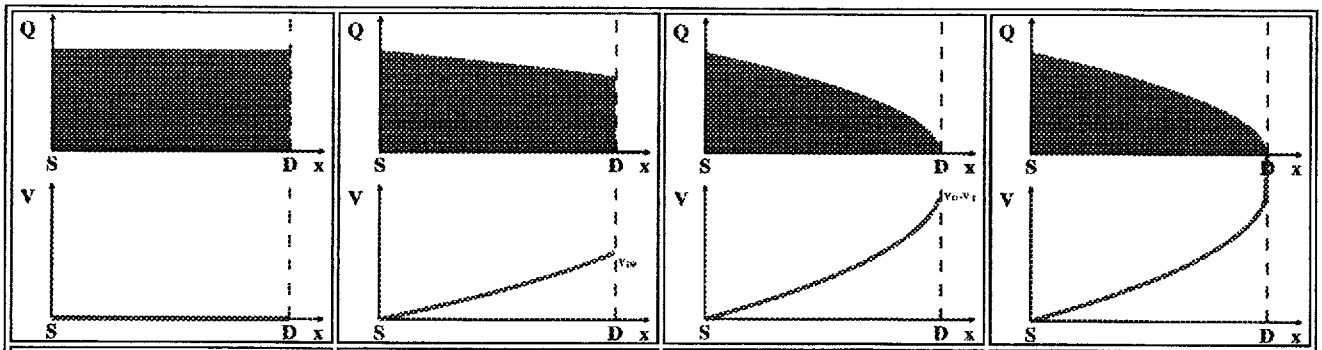
which is equal to what we found for the **linear** region, except for the term V_{DS}^2 which becomes important only when we approach saturation. When we susbtitute $V_{DS} << V_G - V_T$ we get the equation for the linear region. When we substitute the saturation condition $V_{DS} = V_G - V_T$ we reproduce exactly the equation for saturation currents.



Examples of IV curves (I_{DS} vs. V_{DS}; left) for different gate voltages and transfer curves (I_{DS} vs. V_G;

right) for different drain-source voltages. $V_T = 2$ V.

For completeness sake and for later reference, here are the charge distributions and voltage curves for the different regimes, ranging from 1) equilibrium, 2) linear regime, 3) onset of saturation, 4) in saturation.



$V_{DS} = 0$ V	$V_{DS} < V_G - V_T$ LIN	$V_{DS} = V_G - V_T$ onset of SAT	$V_{DS} > V_G - V_T$ SAT

Organic FETs

Why doesn't this work for our organic materials?

- The organic FETs are of the **accumulation** p-channel type. In the accumulation operation there is no space charge region caused by uncompensated ionized acceptors. All the space charge and band bending must come from intrinsic carriers - unbalanced holes-electrons which still follow $pn = N_C N_V \text{Exp}(-E_G/kT)$, but $p>n$ - and carriers in the strong accumulation channel. The result will be that we will still have the equations
$$I_{DS} = C_{ox}(V_G - V_T)\,\mu_p V_{DS} Z/L$$
$$I_{DS} = (1/2)\mu(Z/L)\,C_{ox}(V_G - V_T)^2$$
for currents through the strong accumulation channel in the linear and saturation regions, but that the threshold voltage doesn't follow the same equation anymore as given before. The threshold voltage now becomes (see Horowitz)
$$V_T = qN_A d/C_{\text{ox}}$$
- The FETs are **thin-film** transistors. These don't have a bulk region that can accomodate the band bendings. The result is that very thin films of pure materials will not show any FET behavior, because for no gate voltage can we induce a band bending bthat will open a channel.
- Organic materials have a large sub-threshold voltage current. This also makes that the saturation currents become voltage dependent because the assumption that the size of the region where the voltage drops $V_G - V_T$ is constant is no longer valid. In the discussion above it was assumed that when the device is in saturation, the rest of the voltage is absorbed in a very thin region. Such a strong field can be sustained becuase of the low carrier density. In organic FETs we have a situation where the free-carrier density below threshold is already substantial. The pinch-off region therefore has to be much larger and the rest of the voltage is then absorbed in a smaller region. This gives larger electric fields and larger currents. The currents still rise for increasing V_{DS}.
- The **mobilities** are so **low**, that the above equations don't apply. For instance, we can have

charge built-up at the electrodes which will limit the currents. In this case, the **diffusion** component of the current will take over in importance. See p. 443 of Sze.

The assumption that the current is only depending on the electrical field is wrong. Instead of

$$I = q\mu_e nE$$

we will have

$$I = q\mu_e nE + qD_n\, dn/dx$$

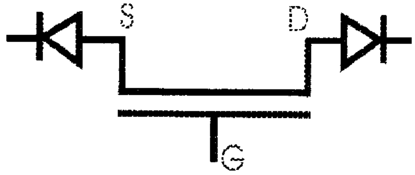
For low mobilities, the first term becomes small and the second dominating. Obviously, this becomes more important when we approach saturation because then we have the largest gradient in carrier concentrations, especially in the pinch-off region close to the drain (see the figures above).

For low-mobility materials it is therefore advisable to measure in the linear region.

- **Charge injection**. In the above discussion it is assumed that the injection of carriers in the channel is not a limiting factor. In the classic n-type inversion-channel FET the source and drain electrodes are, in fact, made of n-type silicon. This means that, by the time an n-channel is created, the Fermi levels in the electrodes and in the channel have exactly lined up (by definition) and there is no barrier at the contact.

  Moreover, where there is no Fermi level alignement (outside the channel) there is a pn-junction that will prohibit conduction through anything else but the channel. We only have to worry about the charges inside the channel.

  In organic FETs the situation is different. When the p-channel is created, we still are not sure if we have Fermi level alignment between the gold electrodes and the p-channel. Maybe we have to overcome a Schottky barrier. This can be simulated by a diode in series with the FET. It implies that for low V_{DS} the current is zero until $V_{DS} = V_{bi}$.

  The proposed structure is the one here on the left. This, however would never conduct (always one of the two diodes is closed). The best way to analyze this is with an MSM device (metal-semiconductor-metal, see p. 613 of Sze). The idea is that the depletion widths on opposite two Schottky barriers can start overlapping for certain voltages. When this happens, at VST, the currents rapidly rise and the barriers effectively disappear. The voltage at which the device opens is therefore NOT the built-in voltage of a diode, but rather the voltage at which breakdown occurs. See Fig. 40 on page 618 of Sze.

  Of course, this is not really an MSM device, because we also have the gate. How does this relate to the FETs we are using?

Short-channel effects

For short channels we can expect the following:

Non saturating saturation currents. We assumed that the voltage drop along the channel is comprised of two parts, a gradual drop $V_G - V_T$ independent of V_{DS} over nearly the entire channel and the last part $[V_{DS} - (V_G - V_T)]$ completely absorbed by an infinitesimal small region. When the channel becomes shorter, the last part takes relatively more space and we cannot assume anymore that the voltage drop $V_G - V_T$ occurs over a constant amount of space independent of V_{DS}. In fact, this region becomes ever smaller and the field will proportionally increase, together with the currents. In other words, the current is, in saturation no longer independent of V_{DS}. See Fig. 41 on p.478 of Sze.

We can easiliy explain this. At the source and the drain we have a depletion region (a zone without free holes in an n-p-n inversion channel FET). This zone will absorb the entire voltage drop $V_{DS} - (V_G - V_T)$. Normally these zones are thin compared to the bulk. It is easy to calculate the width of these depletion zones (see the section on Schottky barriers):

$$y_s = [2\varepsilon_s/qN_A\ (V_{bi} - \psi_s)]^{1/2}$$

$$y_d = [2\varepsilon_s/qN_A\ (V_{bi} - \psi_s + V_D)]^{1/2}$$

for the depletion width at the source and drain respectively. Note that they also depend on the gate; once the channel has formed, the depletion width at the source disappears, but we still keep a depletion width at the drain which will absorp the excess V_{DS}. Note that this "barrier" at the drain doesn't hinder the current, because it is in the forward direction for "minority carriers" (electrons for npn device).

When this depletion width becomes comparable to the channel length we can expect short-channel effects. This also depends on the acceptor concentration.

Non-zero subthreshold currents. Before threshold, we have pn-junctions on either side of the semiconductor. The depletion width W of these follows the standard calculations. When the channel length is decreased, these depletion regions can start overlapping. This will cause that the currents increase. This is easy to see why. When a carrier is injected by diffucsion over the first barrier, into the "channel", it immediately starts feeling the fields of the second barrier on the other side and is pulled away to that electrode; it has a significant chance of making it to the other side. In fact, the diffusion (space-charge limited current) will dominate and the currents will be of the form $J = V_{DS}^2$.

Threshold voltage shift.

Appendix A

List of Symbols

Symbol	Description	Unit
a	Lattice constant	Å
$\mathscr{B}$	Magnetic induction	Wb/m^2
c	Speed of light in vacuum	cm/s
C	Capacitance	F
$\mathscr{D}$	Electric displacement	C/cm^2
D	Diffusion coefficient	cm^2/s
E	Energy	eV
E_C	Bottom of conduction band	eV
E_F	Fermi energy level	eV
E_g	Energy bandgap	eV
E_V	Top of valence band	eV
$\mathscr{E}$	Electric field	V/cm
$\mathscr{E}_c$	Critical field	V/cm
$\mathscr{E}_m$	Maximum field	V/cm
f	Frequency	Hz (cps)
$F(E)$	Fermi-Dirac distribution function	
h	Planck constant	J-s
$h\nu$	Photon energy	eV
I	Current	A
I_C	Collector current	A
J	Current density	A/cm^2
J_t	Threshold current density	A/cm^2
k	Boltzmann constant	J/K
kT	Thermal energy	eV
L	Length	cm or μm
m_0	Electron rest mass	kg
m^*	Effective mass	kg
$\bar{n}$	Refractive index	
n	Density of free electrons	cm^{-3}
n_i	Intrinsic density	cm^{-3}
N	Doping concentration	cm^{-3}
N_A	Acceptor impurity density	cm^{-3}
N_C	Effective density of states in conduction band	cm^{-3}
N_D	Donor impurity density	cm^{-3}
N_V	Effective density of states in valence band	cm^{-3}
p	Density of free holes	cm^{-3}
P	Pressure	N/m^2
q	Magnitude of electronic charge	C
Q_{it}	Interface–trap density	charges/cr
R	Resistance	Ω
t	Time	s
T	Absolute Temperature	K
v	Carrier velocity	cm/s
v_s	Saturation velocity	cm/s
v_{th}	Thermal velocity	cm/s
V	Voltage	V
V_{bi}	Built-in potential	V
V_{EB}	Emitter–base voltage	V
V_B	Breakdown voltage	V
W	Thickness	cm or μm
W_B	Base thickness	cm or μm
x	x direction	
∇	Differential operator	
∇T	Temperature gradient	K/cm
ϵ_0	Permittivity in vacuum	F/cm
ϵ_s	Semiconductor permittivity	F/cm
ϵ_i	Insulator permittivity	F/cm
ϵ_s/ϵ_0 or ϵ_i/ϵ_0	Dielectric constant	
τ	Lifetime or decay time	s
θ	Angle	rad
λ	Wavelength	μm or Å
ν	Frequency of light	Hz

Symbol	Description	Unit
μ_0	Permeability in vacuum	H/cm
μ_n	Electron mobility	cm^2/V-s
μ_p	Hole mobility	cm^2/V-s
ρ	Resistivity	Ω-cm
ϕ	Barrier height or imref	V
ϕ_{Bn}	Schottky barrier height on n-type semiconductor	V
ϕ_{Bp}	Schottky barrier height on p-type semiconductor	V
ϕ_m	Metal work function	V
ω	Angular frequency ($2\pi f$ or $2\pi\nu$)	Hz
Ω	Ohm	Ω

Appendix B

International System of Units

Quantity	Unit	Symbol	Dimensions
Length	meter	m	
Mass	kilogram	kg	
Time	second	s	
Temperature	kelvin	K	
Current	ampere	A	
Frequency	hertz	Hz	1/s
Force	newton	N	$kg\text{-}m/s^2$
Pressure	pascal	Pa	N/m^2
Energy	joule	J	N-m
Power	watt	W	J/s
Electric charge	coulomb	C	A-s
Potential	volt	V	J/C
Conductance	siemens	S	A/V
Resistance	ohm	Ω	V/A
Capacitance	farad	F	C/V
Magnetic flux	weber	Wb	V-s
Magnetic induction	tesla	T	Wb/m^2
Inductance	henry	H	Wb/A

Appendix C

Unit Prefixes[a]

Multiple	Prefix	Symbol	Multiple	Prefix	Symbol
10^{18}	exa	E	10^{-1}	deci	d
10^{15}	peta	P	10^{-2}	centi	c
10^{12}	tera	T	10^{-3}	milli	m
10^{9}	giga	G	10^{-6}	micro	μ
10^{6}	mega	M	10^{-9}	nano	n
10^{3}	kilo	k	10^{-12}	pico	p
10^{2}	hecto	h	10^{-15}	femto	f
10	deka	da	10^{-18}	atto	a

[a] Adopted by International Committee on Weights and Measures. (Compound prefixes should not be used; e.g., not $\mu\mu$ but p.)

Appendix D

Greek Alphabet

Letter	Lowercase	Uppercase	Letter	Lowercase	Uppercase
Alpha	α	A	Nu	ν	N
Beta	β	B	Xi	ξ	Ξ
Gamma	γ	Γ	Omicron	o	O
Delta	δ	Δ	Pi	π	Π
Epsilon	ϵ	E	Rho	ρ	P
Zeta	ζ	Z	Sigma	σ	Σ
Eta	η	H	Tau	τ	T
Theta	θ	Θ	Upsilon	υ	Υ
Iota	ι	I	Phi	ϕ	Φ
Kappa	κ	K	Chi	χ	X
Lambda	λ	Λ	Psi	ψ	Ψ
Mu	μ	M	Omega	ω	Ω

Appendix E

Physical Constants

Quantity	Symbol	Value
Angstrom unit	Å	1 Å = 10^{-4} μm = 10^{-8} cm
Avogadro constant	N_{AVO}	6.02204×10^{23} mol^{-1}
Bohr radius	a_B	0.52917 Å
Boltzmann constant	k	1.38066×10^{-23} J/K (R/N_{AVO})
Elementary charge	q	1.60218×10^{-19} C
Electron rest mass	m_0	0.91095×10^{-30} kg
Electron volt	eV	1 eV = 1.60218×10^{-19} J = 23.053 kcal/mol
Gas constant	R	1.98719 cal mol^{-1} K^{-1}
Permeability in vacuum	μ_0	1.25663×10^{-8} H/cm ($4\pi \times 10^{-9}$)
Permittivity in vacuum	ϵ_0	8.85418×10^{-14} F/cm ($1/\mu_0 c^2$)
Planck constant	h	6.62617×10^{-34} J-s
Reduced Planck constant	$\hbar$	1.05458×10^{-34} J-s ($h/2\pi$)
Proton rest mass	M_p	1.67264×10^{-27} kg
Speed of light in vacuum	c	2.99792×10^{10} cm/s
Standard atmosphere		1.01325×10^{5} N/m^2
Thermal voltage at 300 K	kT/q	0.0259 V
Wavelength of 1-eV quantum	λ	1.23977 μm

Appendix F

Lattice Constants

	Element or Compound	Name	Crystal[a] Structure	Lattice Constant at 300 K (Å)
Element	C	Carbon (diamond)	D	3.56683
	Ge	Germanium	D	5.64613
	Si	Silicon	D	5.43095
	Sn	Grey Tin	D	6.48920
IV–IV	SiC	Silicon carbide	W	$a = 3.086$, $c = 15.117$
III–V	AlAs	Aluminum arsenide	Z	5.6605
	AlP	Aluminum phosphide	Z	5.4510
	AlSb	Aluminum antimonide	Z	6.1355
	BN	Boron nitride	Z	3.6150
	BP	Boron phosphide	Z	4.5380
	GaAs	Gallium arsenide	Z	5.6533
	GaN	Gallium nitride	W	$a = 3.189$, $c = 5.185$
	GaP	Gallium phosphide	Z	5.4512
	GaSb	Gallium antimonide	Z	6.0959
	InAs	Indium arsenide	Z	6.0584
	InP	Indium phosphide	Z	5.8686
	InSb	Indium antimonide	Z	6.4794
II–VI	CdS	Cadmium sulfide	Z	5.8320
	CdS	Cadmium sulfide	W	$a = 4.16$, $c = 6.756$
	CdSe	Cadmium selenide	Z	6.050
	CdTe	Cadmium telluride	Z	6.482
	ZnO	Zinc oxide	R	4.580
	ZnS	Zinc sulfide	Z	5.420
	ZnS	Zinc sulfide	W	$a = 3.82$, $c = 6.26$
IV–VI	PbS	Lead sulfide	R	5.9362
	PbTe	Lead telluride	R	6.4620

---

[a] D = Diamond, W = Wurtzite, Z = Zincblende, R = Rock salt.

Appendix G

Properties of Important Semiconductors

Semiconductor		Bandgap (eV)		Mobility at 300 K (cm^2/V-s)[a]		Band[b]	Effective Mass m^*/m_0		ϵ_s/ϵ_0
		300 K	0 K	Elec.	Holes		Elec.	Holes	
Element	C	5.47	5.48	1800	1200	I	0.2	0.25	5.7
	Ge	0.66	0.74	3900	1900	I	1.64[c]	0.04[e]	16.0
							0.082[d]	0.28[f]	
	Si	1.12	1.17	1500	450	I	0.98[c]	0.16[e]	11.9
							0.19[d]	0.49[f]	
	Sn		0.082	1400	1200	D			
IV–IV	α-SiC	2.996	3.03	400	50	I	0.60	1.00	10.0
III–V	AlSb	1.58	1.68	200	420	I	0.12	0.98	14.4
	BN	~7.5				I			7.1
	BP	2.0							
	GaN	3.36	3.50	380			0.19	0.60	12.2
	GaSb	0.72	0.81	5000	850	D	0.042	0.40	15.7
	GaAs	1.42	1.52	8500	400	D	0.067	0.082	13.1
	GaP	2.26	2.34	110	75	I	0.82	0.60	11.1
	InSb	0.17	0.23	80000	1250	D	0.0145	0.40	17.7
	InAs	0.36	0.42	33000	460	D	0.023	0.40	14.6
	InP	1.35	1.42	4600	150	D	0.077	0.64	12.4
II–VI	CdS	2.42	2.56	340	50	D	0.21	0.80	5.4
	CdSe	1.70	1.85	800		D	0.13	0.45	10.0
	CdTe	1.56		1050	100	D			10.2
	ZnO	3.35	3.42	200	180	D	0.27		9.0
	ZnS	3.68	3.84	165	5	D	0.40		5.2
IV–VI	PbS	0.41	0.286	600	700	I	0.25	0.25	17.0
	PbTe	0.31	0.19	6000	4000	I	0.17	0.20	30.0

---

[a] The values are for drift mobilities obtained in the purest and most perfect materials available to date.

[b] I = indirect, D = direct.

[c] Longitudinal effective mass.

[d] Transverse effective mass.

[e] Light-hole effective mass.

[f] Heavy-hole effective mass.

Appendix H

Properties of Ge, Si, and GaAs at 300 K

Properties	Ge	Si	GaAs
Atoms/cm^3	4.42×10^{22}	5.0×10^{22}	4.42×10^{22}
Atomic weight	72.60	28.09	144.63
Breakdown field(V/cm)	$\sim 10^5$	$\sim 3 \times 10^5$	$\sim 4 \times 10^5$
Crystal structure	Diamond	Diamond	Zincblende
Density (g/cm^3)	5.3267	2.328	5.32
Dielectric constant	16.0	11.9	13.1
Effective density of states in conduction band, N_C (cm^{-3})	1.04×10^{19}	2.8×10^{19}	4.7×10^{17}
Effective density of states in valence band, N_V (cm^{-3})	6.0×10^{18}	1.04×10^{19}	7.0×10^{18}
Effective Mass, m^*/m_0			
Electrons	$m_l^* = 1.64$ $m_t^* = 0.082$	$m_l^* = 0.98$ $m_t^* = 0.19$	0.067
Holes	$m_{lh}^* = 0.044$ $m_{hh}^* = 0.28$	$m_{lh}^* = 0.16$ $m_{hh}^* = 0.49$	$m_{lh}^* = 0.082$ $m_{lh}^* = 0.45$
Electron affinity, χ(V)	4.0	4.05	4.07
Energy gap (eV) at 300 K	0.66	1.12	1.424
Intrinsic carrier concentration (cm^{-3})	2.4×10^{13}	1.45×10^{10}	1.79×10^6
Intrinsic Debye length (μm)	0.68	24	2250
Intrinsic resistivity (Ω-cm)	47	2.3×10^5	10^8
Lattice constant (Å)	5.64613	5.43095	5.6533

Properties of Ge, Si and GaAs at 300K

Properties	Ge	Si	GaAs
Linear coefficient of thermal expansion, $\Delta L/L\Delta T$ (°C^{-1})	5.8×10^{-6}	2.6×10^{-6}	6.86×10^{-6}
Melting point (°C)	937	1415	1238
Minority carrier lifetime (s)	10^{-3}	2.5×10^{-3}	$\sim 10^{-8}$
Mobility (drift) (cm^2/V-s)	3900	1500	8500
	1900	450	400
Optical-phonon energy (eV)	0.037	0.063	0.035
Phonon mean free path λ_0 (Å)	105	76 (electron) 55 (hole)	58
Specific heat (J/g-°C)	0.31	0.7	0.35
Thermal conductivity at 300 K (W/cm-°C)	0.6	1.5	0.46
Thermal diffusivity (cm^2/s)	0.36	0.9	0.24
Vapor pressure (Pa)	1 at 1330°C 10^{-6} at 760°C	1 at 1650°C 10^{-6} at 900°C	100 at 1050°C 1 at 900°C

Appendix I

Properties of SiO_2 and Si_3N_4 at 300 K

Insulator:	SiO_2	Si_3N_4
Structure	Amorphous	Amorphous
Melting point (°C)	~1600	—
Density (g/cm^3)	2.2	3.1
Refractive index	1.46	2.05
Dielectric constant	3.9	7.5
Dielectric strength (V/cm)	10^7	10^7
Infrared absorption band (μm)	9.3	11.5–12.0
Energy gap (eV)	9	~5.0
Thermal-expansion coefficient (°C^{-1})	5×10^{-7}	—
Thermal conductivity (W/cm-K)	0.014	—
dc resistivity (Ω-cm)		
at 25°C	10^{14}–10^{16}	$\sim 10^{14}$
at 500°C	—	$\sim 2\times10^{13}$
Etch rate in buffered HF[a] (Å/min)	1000	5-10

[a] Buffered HF: 34.6% (wt.) NH_4F, 6.8% (wt.) HF, 58.6% H_2O.