ENG1004 - Electronics & Sensors

Part 1 - Semiconductors

S. Le Goff, School of Engineering, Newcastle University

Solid materials may be divided, with respect to their electrical properties, into three categories: conductors, insulators, and semiconductors.

Conductors, semiconductors, and many insulators are formed by ordered, crystalline arrays of atoms in which nearest neighbours are joined by covalent bonds. The structure of an actual crystalline solid, called the *crystal lattice*, is three-dimensional.

1. Conductors

Conductors (e.g., copper, aluminium, silver, gold) have a cloud of free electrons at all temperature above absolute zero. This is formed by the weakly bound “valence” electrons in the outermost orbits of their atoms. If an electric field is applied across such a material, electrons will flow, causing an electric current.

2. Insulators

In insulating materials (e.g., porcelain, glass, most plastics), the valence electrons are tightly bound to the nuclei of the atoms and very few of them are able to break free to conduct electricity. The application of an electric field does not cause a current to flow as there are no mobile charge carriers.

3. Semiconductors

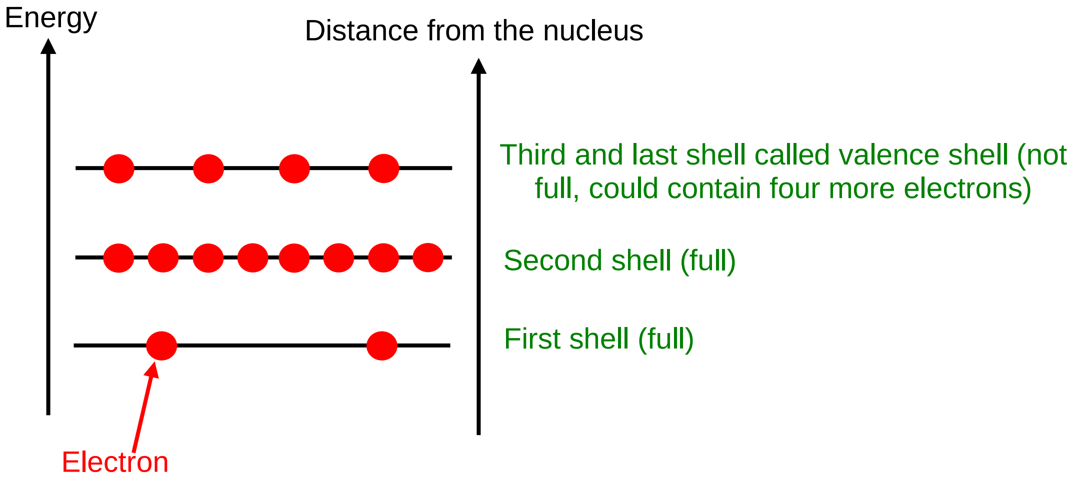
At very low temperatures, semiconductors have the properties of an insulator. However, at higher temperatures, some electrons are free to move and the materials take on the properties of a conductor (albeit a poor one). Nevertheless, semiconductors have some useful characteristics that make them distinct from both insulators and conductors.

1. Energy Bands

In solid-state physics, the energy band theory is often used to make the distinction between conductors, insulators, and semiconductors. The energy band diagram shows all bands of energy admitted for the electrons within a material.

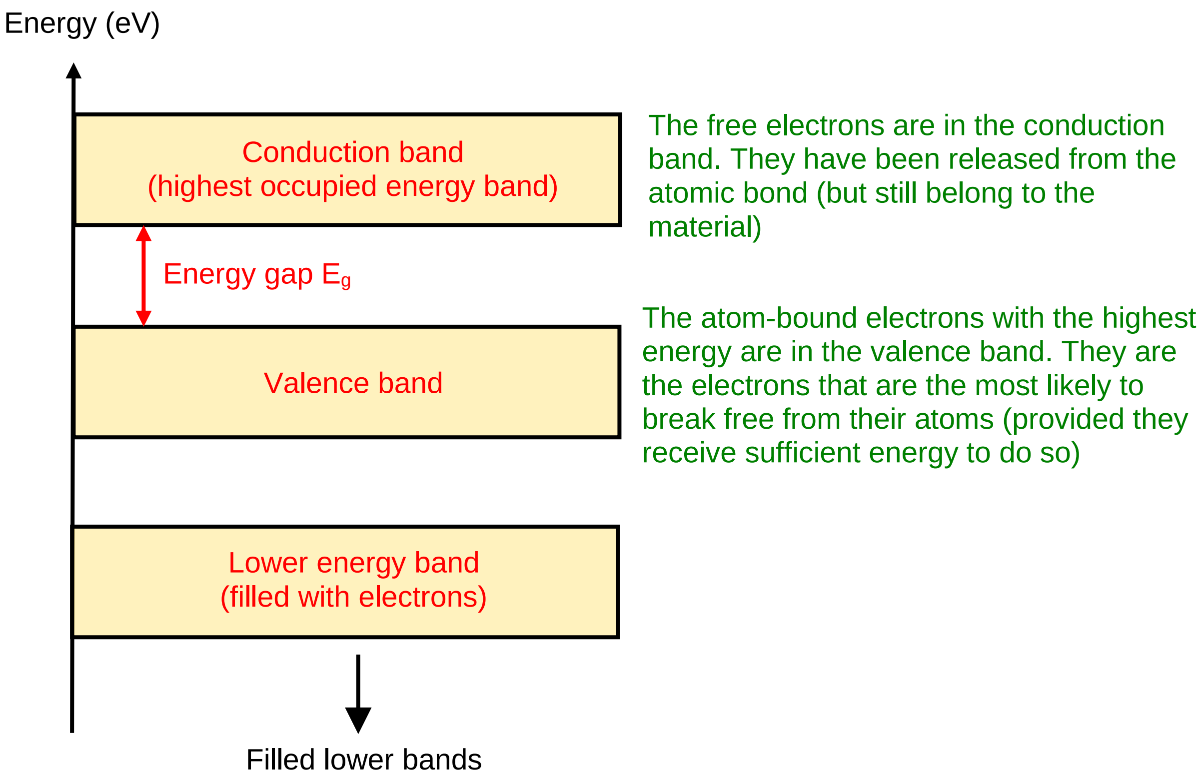
Electrons in an atom occupy discrete energy levels, also known as energy shells. When many atoms are brought into close proximity as in a crystal, the discrete energy levels are replaced with bands of energy states separated by gaps between the bands. One may think of an energy band as a semi-continuum of a very large number of energy states.

Example of a silicon atom with its 14 electrons and three discrete energy levels (energy shells)



The highest occupied energy band in a crystalline solid is called the *conduction band* and the next highest is called the *valence band*. The jumps in energy between the discrete energy levels of a single atom transform into *energy gaps* in the crystalline form of the element. The energy gap just below the conduction band is called the *band gap*.

Naturally, the electrons tend to fill up the low energy bands first. The lower the energy, the more completely a band is filled. In a semiconductor, the valence band is nearly filled with electrons, while the conduction band is nearly empty.



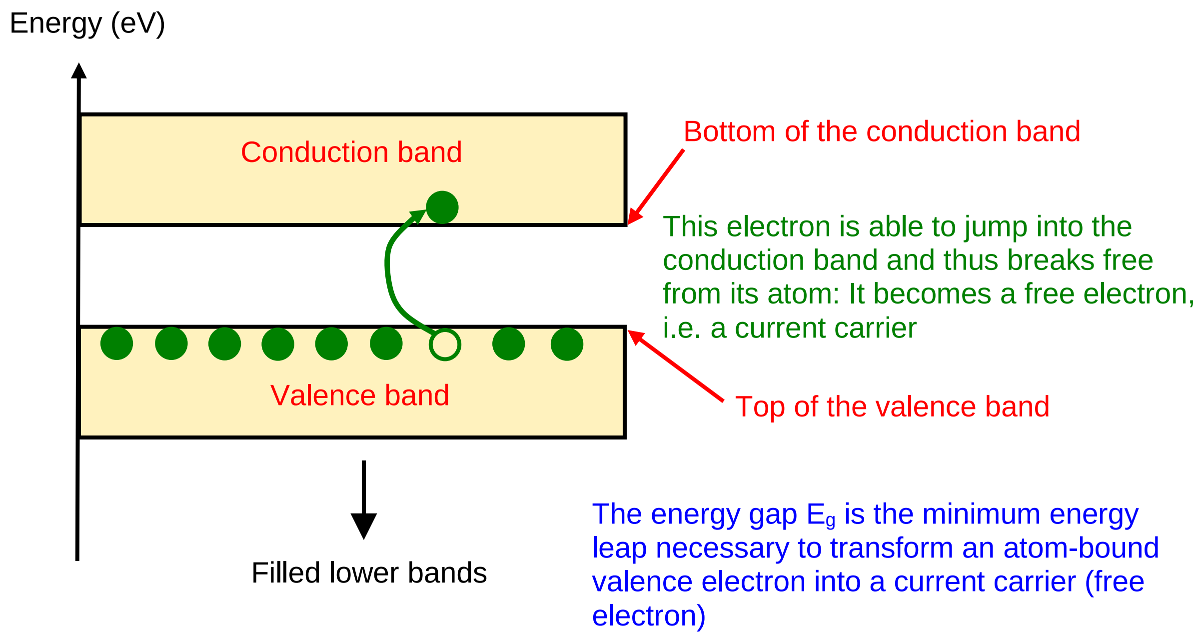
The free electrons are in the conduction band. They have been released from the atomic bond (but still belong to the material)

The atom-bound electrons with the highest energy are in the valence band. They are the electrons that are the most likely to break free from their atoms (provided they receive sufficient energy to do so)

The electrons in a totally filled band do not have a net velocity and do not conduct current, just as the water in a totally filled bottle does not slosh about. Similarly, a totally empty band cannot contribute to current conduction.

These are the reasons the valence band and the conduction band are the only energy bands that contribute to current flows in a semiconductor.

Hence, we only need to focus on those two bands that actually correspond to the two highest energy levels that electrons can have in a material.



Note that the electrons that determine how an atom reacts chemically are those that travel farthest from the nucleus, that is, those with the highest energy.

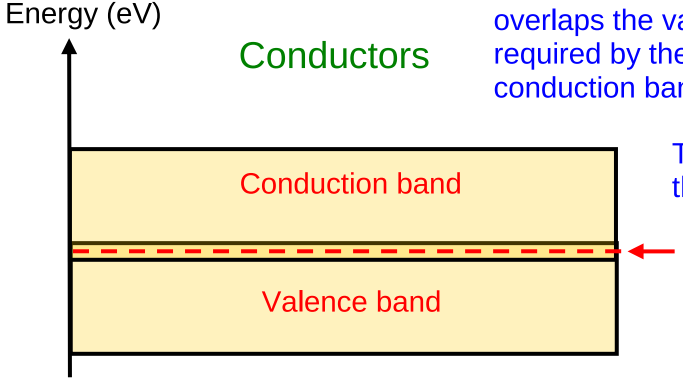
The band structure of a crystal can also be described in terms of chemical bonds. The valence band contains those electrons bound up in covalent bonds. The conduction band contains those electrons that are not bound and are free to move about the crystal. The band gap represents the energy required to break the covalent bonds of the valence electrons.

It is the value of the energy gap E_g that categorizes the materials as insulators, conductors, and semiconductors.

Note that there is in fact no clear boundary between insulators and semiconductors. For instance, diamond exhibits semiconductor characteristics despite having a band gap $E_g \approx 6$ eV.

Conductors

In conductors, such as metals, the conduction band overlaps the valence band → Very little or no energy is required by the valence electrons to “jump” into the conduction band and become current carriers.

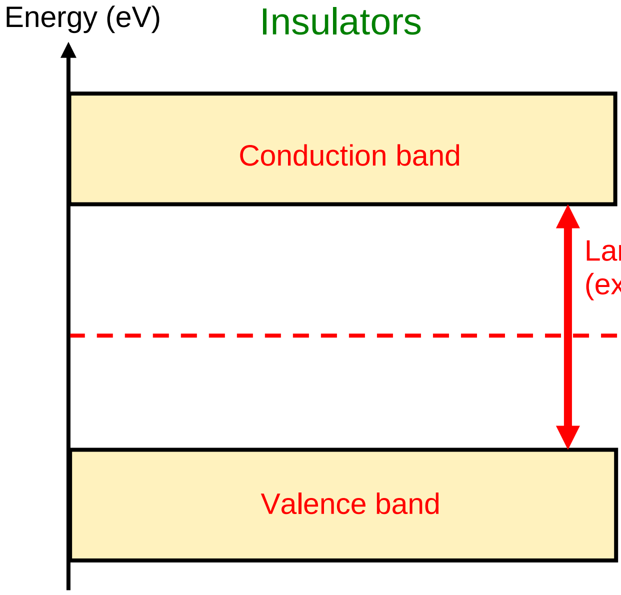


The fact that there is no band gap implies that the valence electrons are essentially free.

Fermi level E_F at the bottom of the conduction band → An electron has a very high probability to be in the conduction band.

"Fermi level" is the term used to describe the top of the collection of electron energy levels at absolute zero temperature. At absolute zero, electrons pack into the lowest available energy states and build up a "Fermi sea" of electron energy states. The Fermi level is the surface of that sea at absolute zero where no electrons will have enough energy to rise above the surface.
When you study the Fermi function at a later stage, you will see that an electron has less than 50% probability to be in an energy level above the Fermi level and more than 50% probability to be in an energy level below the Fermi level. The latter is thus an indication of the likely distribution of electrons.

Insulators



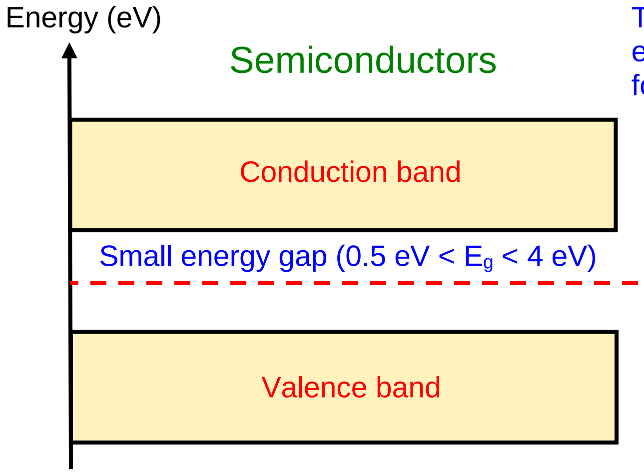
The energy required to transfer the valence electrons to the conduction band is so high that (almost) none of these electrons can reach the conduction band.

Large energy gap $E_g > 4$ eV
(ex: $E_g = 9$ eV for silicon dioxide, $Si0_2$)

Fermi level E_F very far from the conduction band → An electron has a very low probability to be in the conduction band.

As (almost) all electrons stay bonded to their respective atoms, a number of carriers that is sufficient to obtain a significant current cannot be reached.

Even the doping of insulators is not enough to overcome the large band gap to make them good conductors of electricity.



The energy required to transfer the valence electrons to the conduction band is higher than for conductors but lower than for insulators.

Thermal energy may be sufficient for valence electrons to “jump” into the conduction band, and thus to break free from their respective atoms.

Fermi level E_F located halfway between both bands, rather far from the conduction band → An electron has a low probability to be in the conduction band.

At room temperature, $E_g \approx 0.67$ eV for germanium, $E_g \approx 1.12$ eV for silicon, $E_g \approx 1.35$ eV for indium phosphide, and $E_g \approx 1.42$ eV for gallium arsenide

In a semiconductor, the valence band is (almost) completely filled with electrons, while the conduction band is essentially empty. Although no conduction occurs at 0 K, at higher temperatures a finite number of electrons can reach the conduction band and provide some current.

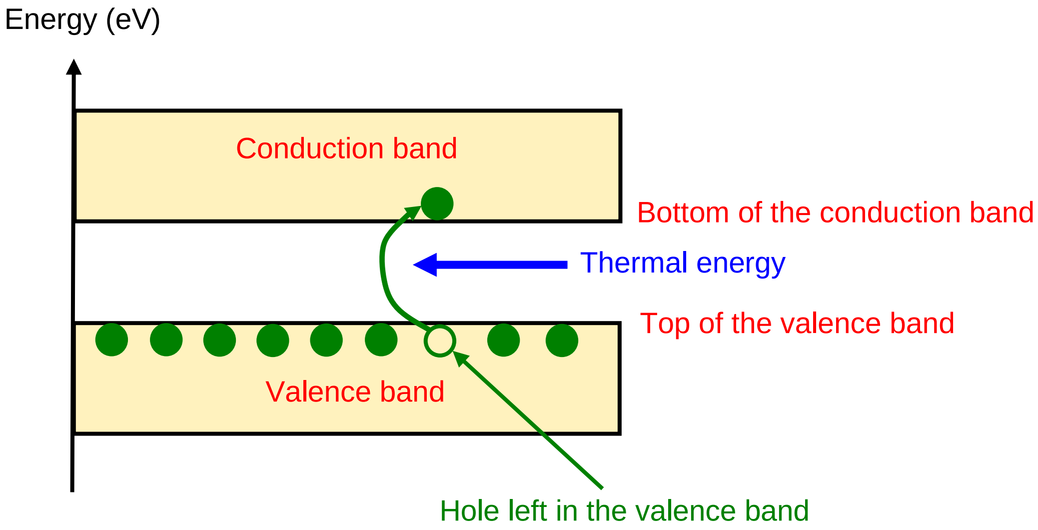
In doped semiconductors, extra energy levels are added. Due to the small gap, the presence of a small percentage of a doping material can increase conductivity dramatically.

In a semiconductor, energy cannot easily be added to valence-band electrons for electric conduction processes because almost all available “spaces” within the valence band are already filled. Adding extra energy to valence-band electrons requires that they be excited all the way up through the band gap to the conduction band. This process is only possible if the added amount of energy is quite significant.

In spite of this, some electrons in a semiconductor material still occasionally manage to “jump” to the conduction band thanks to the energy provided by random thermal vibrations. This is especially true at higher temperatures.

The thermally-excited electrons move spontaneously from the valence band to the conduction band. They thus become current carriers that can contribute to the conduction of current in the crystal if an external voltage is applied. This thermal-excitation mechanism is responsible for the conductivity of a pure semiconductor.

When an electron moves to the conduction band, it leaves behind an “empty space” in the valence band, called a *hole*. This is why the whole process is often referred to as an *electron-hole pair generation*.



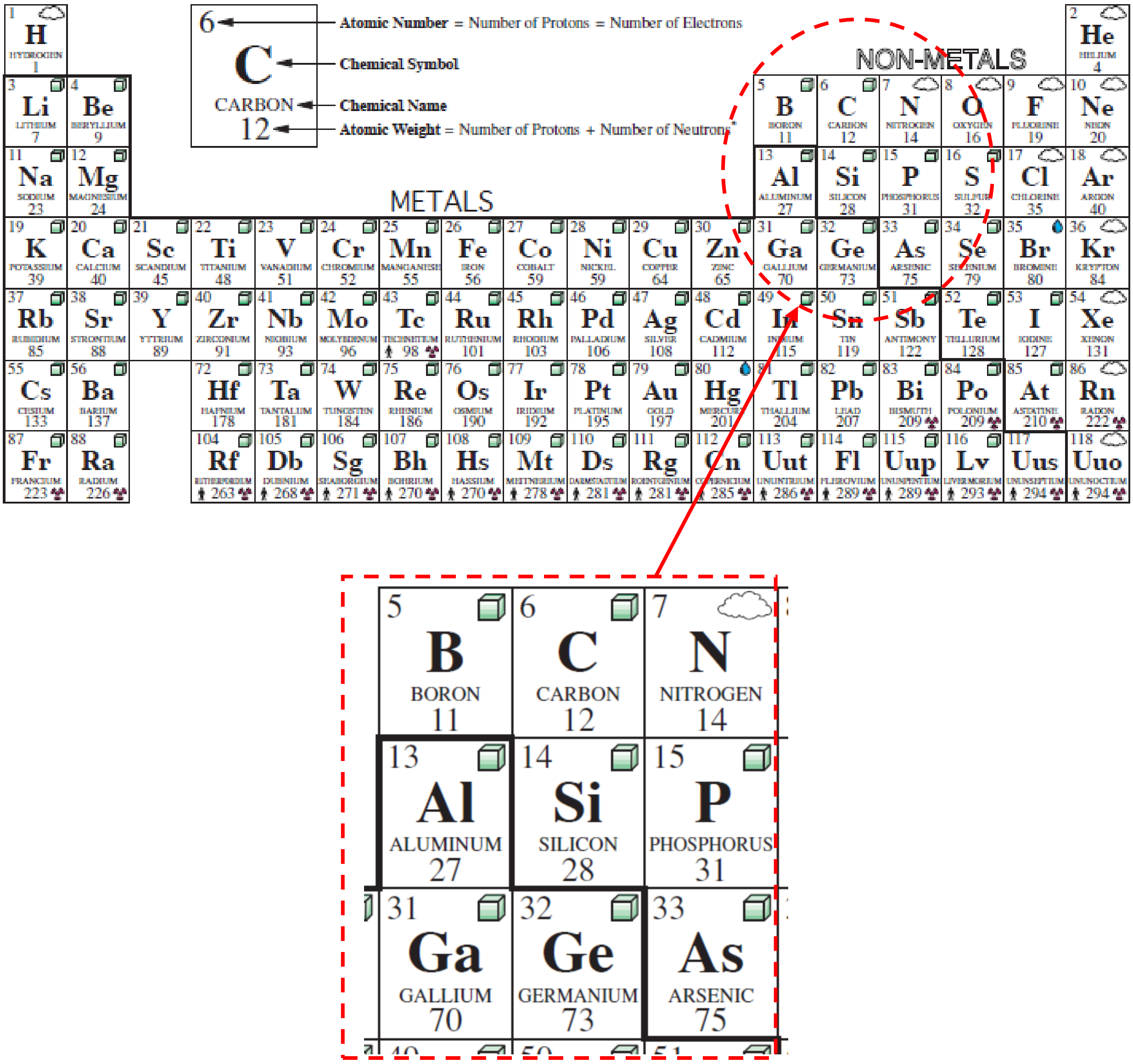
We are now going to focus our attention on the basic structure of semiconductors, as these materials are used to design and manufacture most electronic devices such as diodes and transistors.

A few common semiconductor materials: silicon (Si), germanium (Ge), gallium arsenide (GaAs), indium phosphide (InP), silicon carbide (SiC), silicon-germanium (SiGe). Carbon (C).

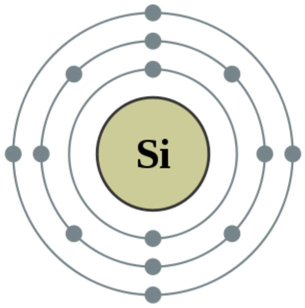
The first transistors were made from germanium (Ge). Silicon (Si) types currently predominate but certain advanced microwave and high-speed versions employ the compound semiconductor material gallium arsenide (GaAs) and the semiconductor alloy silicon germanium (SiGe).

Silicon and germanium fall in column IVa of the Periodic Table. This is the carbon family of elements. The characteristic of these elements is that each atom has four electrons to share with adjacent atoms.

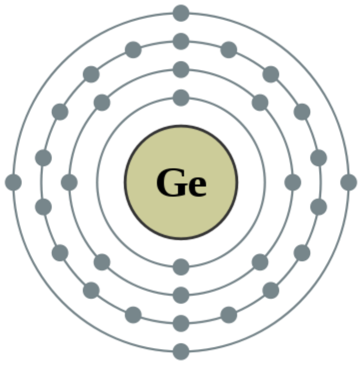
The Periodic Table of Elements



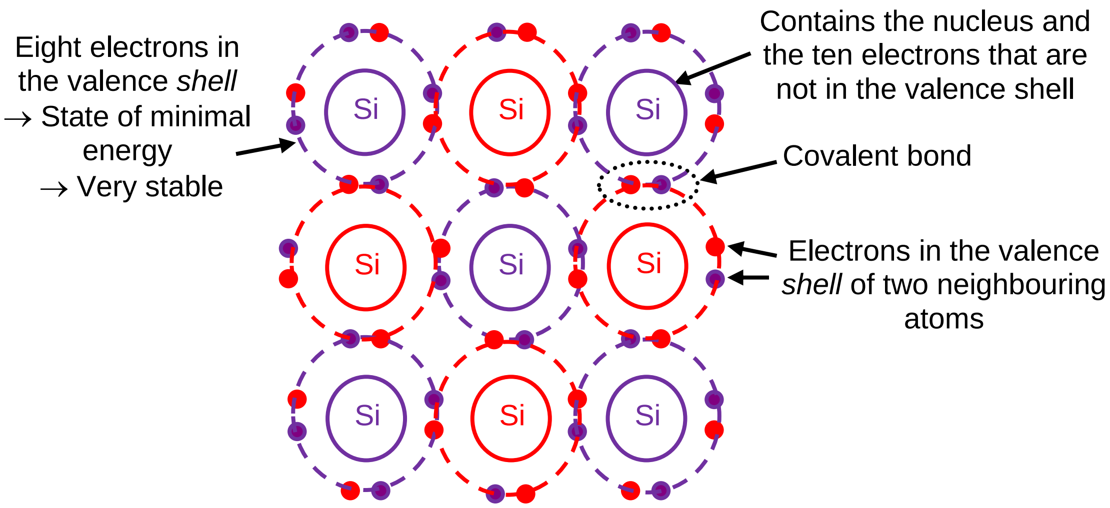
14: Silicon **2,8,4**



32: Germanium **2,8,18,4**



Let us have a closer look at silicon. The crystal structure of silicon is represented below.



The nature of a bond between two silicon atoms is such that each atom provides one electron to share with the other. The two electrons thus shared between atoms form a *covalent bond*. Such a bond is very stable and holds the two atoms together very tightly. It requires a lot of energy to break this bond.

The covalent bond originates from the so-called *octet rule*: an atom with eight electrons in its *valence shell* is in its lowest state of energy. The valence shell is the outer shell of the atom that contains the electrons with the highest energy, i.e. the electrons that are most likely to be found at the periphery of the atom.

Remember that the valence shell represents at the atom level what the valence band represents at the material level. In other words, one should use the term valence shell when reasoning at the atom level, whereas one should use the term valence band when reasoning at the scale of the material.

Everything in nature likes to occupy its lowest possible energy level. As an example, one can think of the atoms of noble gases which do not bind with other atoms because they have eight electrons in their valence shell and are thus already in their state of minimal energy.

For the atoms of other elements, this implies that they naturally tend to combine with other atoms, thus forming molecules, in order to reach a state of minimal energy. The carbon dioxide molecule shown below is another illustration of the octet rule. The bonding between atoms is such that the each atom is surrounded by eight electrons.

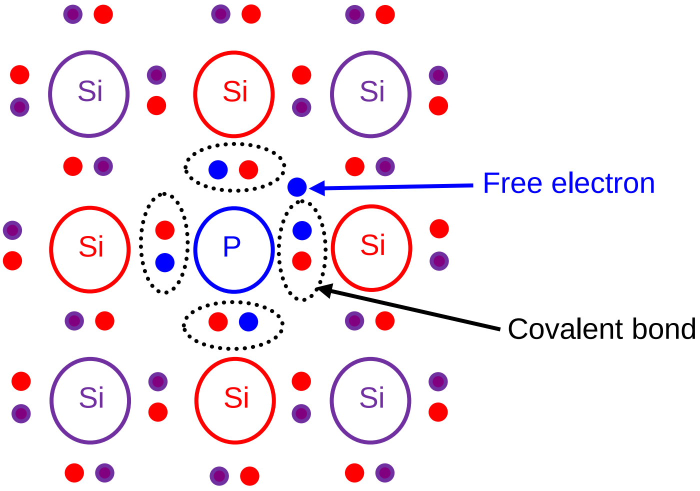
All the electrons on the outer shells of all silicon atoms are used to make covalent bonds with other atoms. There are no electrons available to move from place to place as an electrical current. Thus, a pure silicon crystal is quite a good insulator.

Increasing the temperature results in some electrons breaking free from their covalent bonds and this improves the conductivity of the silicon crystal.

2. N-Type Doping

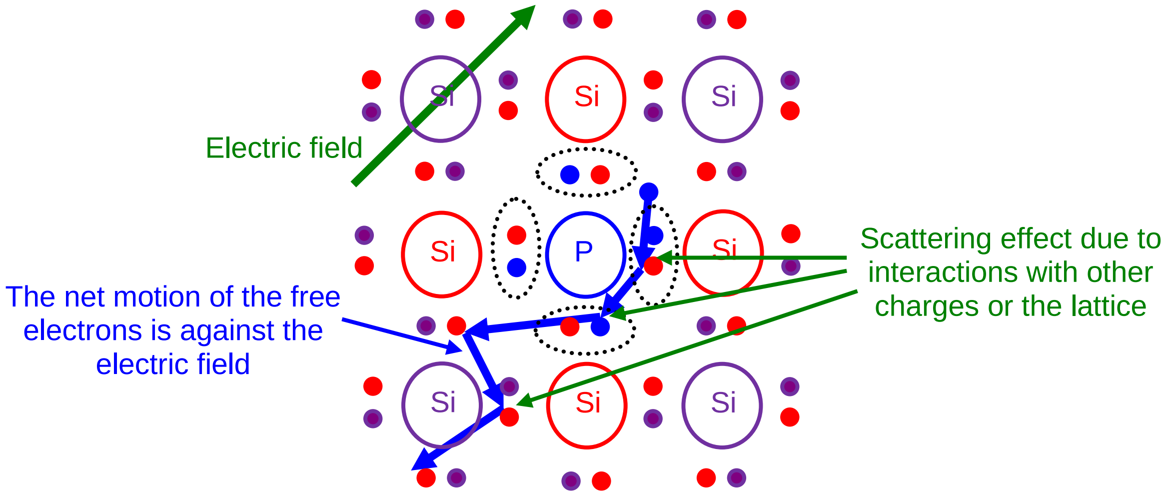
To allow a silicon crystal to conduct electricity without having to increase the temperature, we must find a way to allow some electrons to move from one place to the other within the crystal despite the covalent bonds between atoms.

One way to accomplish this is to introduce an *impurity* such as arsenic, phosphorus, or antimony into the crystal structure. Such process is called *doping*. These elements are from column Va of the Periodic Table, and have five electrons in their valence shell to share with other atoms.



Four of these five electrons bond with adjacent silicon atoms as before, but the fifth electron cannot form a bond and is thus left “alone”. As a result, it can easily break away from its atom through thermal agitation and thus becomes a free electron.

This free electron is in fact in the conduction band of the material and can easily be moved with only a small applied electrical voltage. Because the resulting crystal has an excess of current carriers (free electrons), each with a negative charge, it is known as *N-type silicon*.



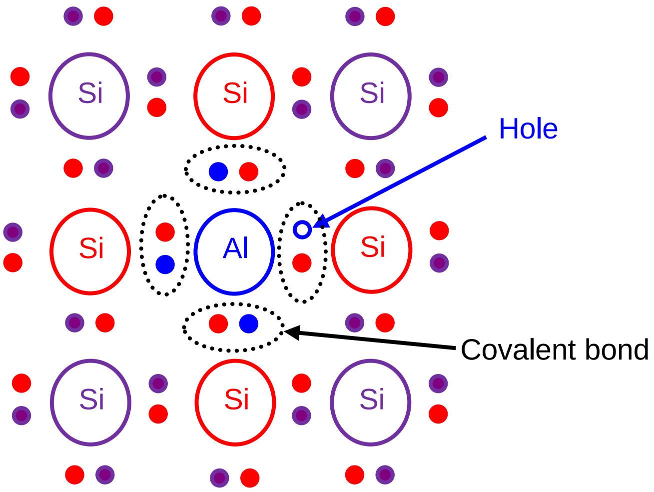
Such construction does not conduct electricity as easily as, say, copper or silver since it does exhibit some resistance to the flow of electricity. It cannot be called a conductor, but at the same time it is no longer an insulator.

3. P-Type Doping

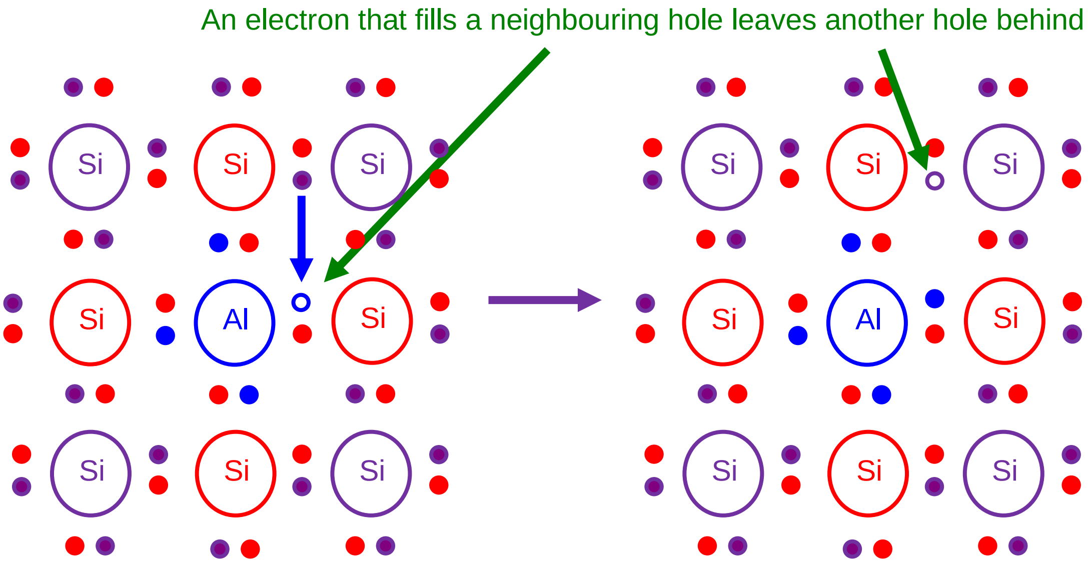
We have significantly improved the conductivity of a semiconductor material by introducing a five-electron impurity into a matrix of four-electron atoms. We can also do the opposite and introduce a three-electron impurity into such a crystal. Suppose we introduce some boron, aluminium or gallium atoms (all taken from column IIIa in the Periodic Table).

These elements only have three valence electrons available to share with other atoms. Those three electrons do indeed form covalent bonds with adjacent silicon atoms, but the expected fourth bond cannot be formed. A complete connection is impossible here, leaving a *hole* in the structure of the crystal.

A hole is an empty place where an electron should naturally go, and often an electron will try to move into that space to fill it. However, the electron filling the hole has to leave a covalent bond behind to fill this empty space, and therefore leaves another hole behind as it moves. Yet another electron may move into that hole, leaving another hole behind, and so on.



In this manner, holes appear to move as positive charges through the crystal. Therefore, this type of semiconductor material is known as *P-type silicon*.



Note that holes can only exist in the valence band, unlike electrons which can exist both in the valence and conduction bands.

The main difference between the current flowing in a conductor and the one flowing in a semiconductor is the existence in the latter of a motion of charges within the crystal lattice. Of course, in both cases, the movement of the free electrons in the conduction band contributes to the current. Less intuitive, but equally fundamental, is the fact that, in semiconductors, the holes also contribute to the overall electric current.

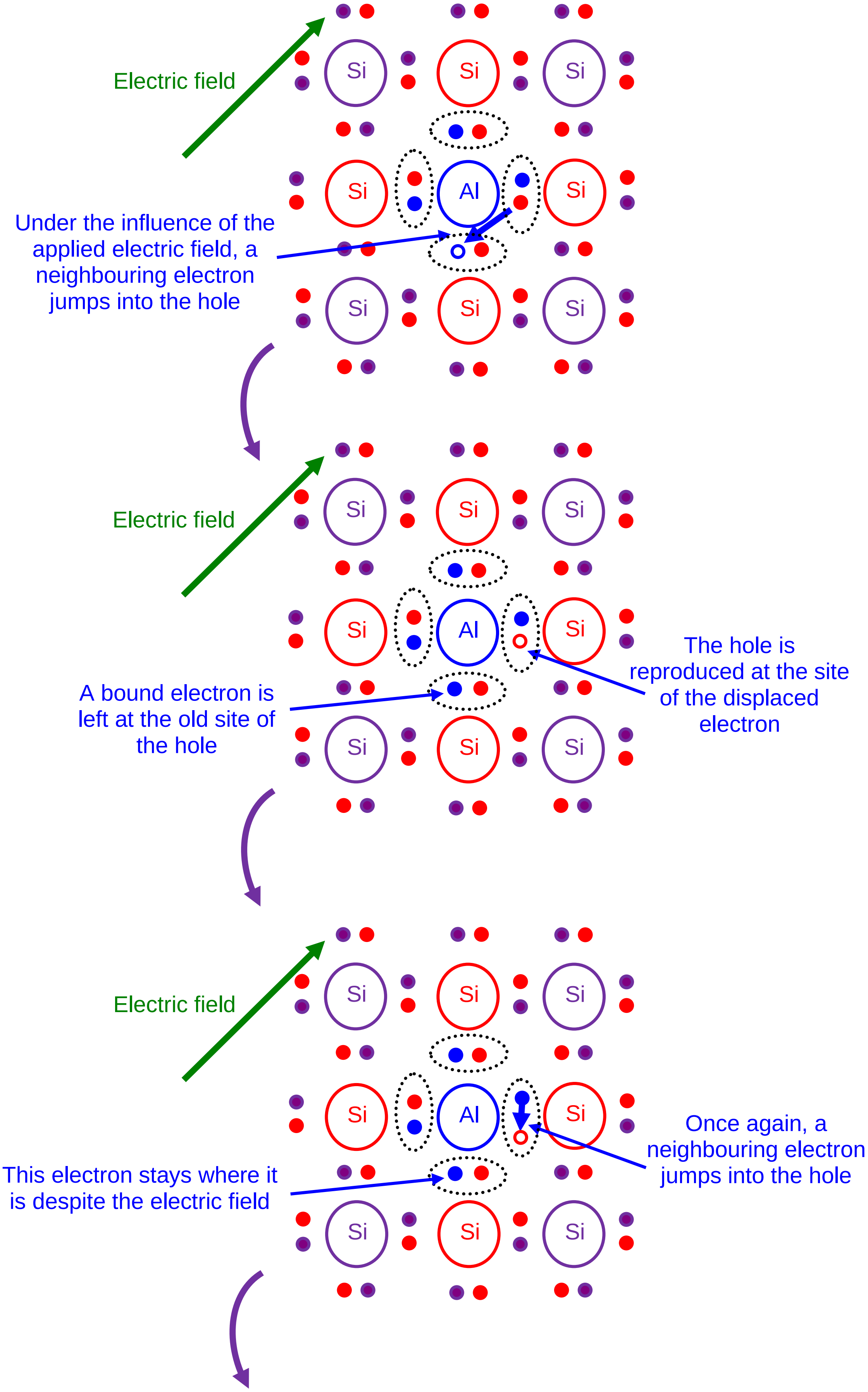
Although it is not a physical particle in the same sense as an electron, a hole can be considered to pass from atom to atom in the opposite direction an electron moves, as if it was a positive charge. Effectively, in semiconductors, the total current is the sum of the current due to the movement of free electrons in the conduction band and the current due to the movement of electrons and holes within the lattice, i.e. in the valence band.

If an electron from the conduction band goes down to the valence band (because it has lost energy for some reason), this motion cannot be considered current, and we have one electron-hole pair less. This event is known as *electron-hole pair recombination*.

Why did we have to introduce the concept of holes? Why not simply consider electrons moving in the opposite direction within the valence band?

The answer is as follows: Instead of analysing the movement of an empty space in the valence band as the movement of many separate electrons, it is more convenient to introduce a single equivalent imaginary particle called a "hole".

In an applied electric field, the electrons move in one direction, corresponding to the hole moving in the other.



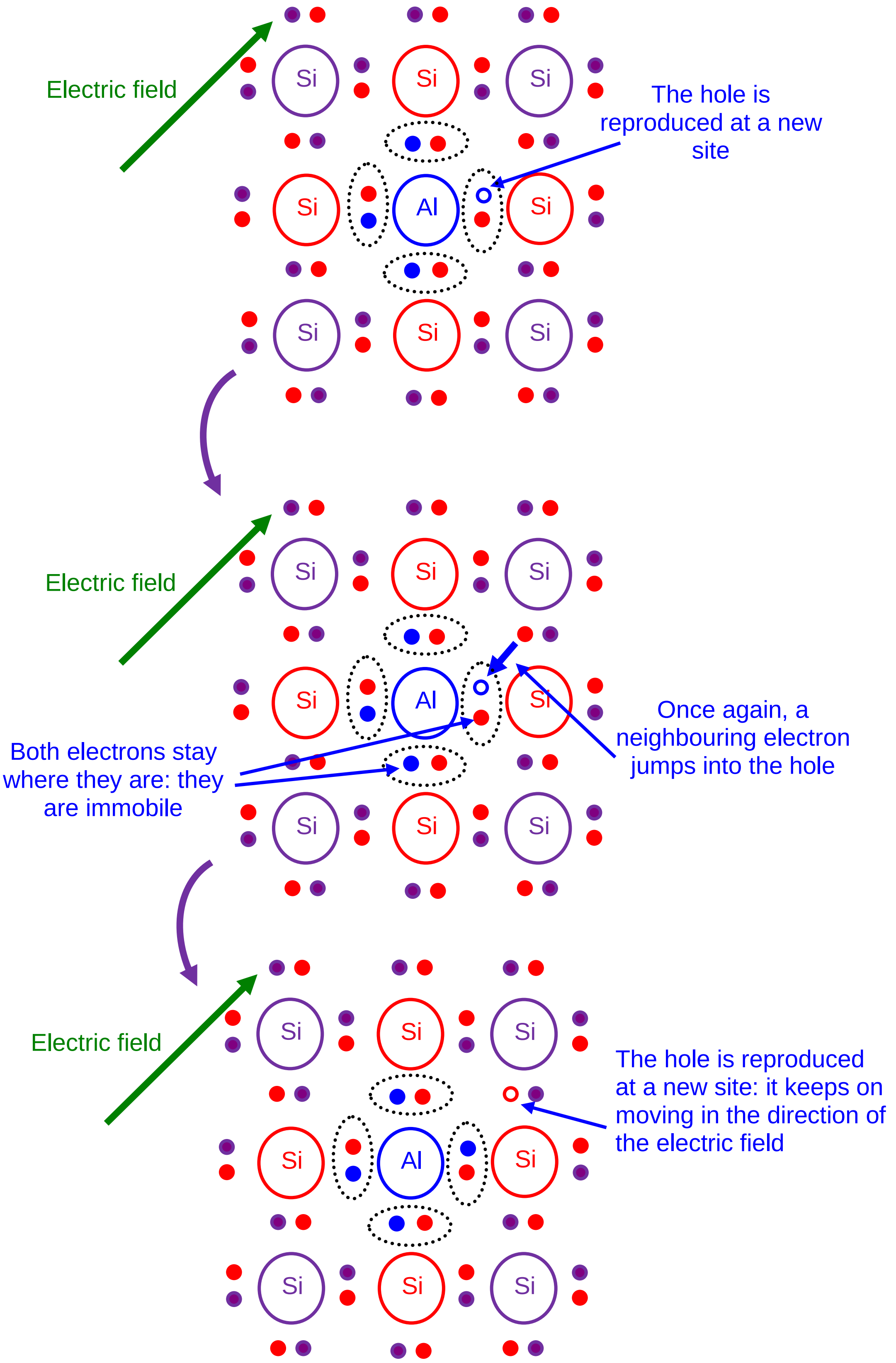
Under the influence of the applied electric field, a neighbouring electron jumps into the hole

A bound electron is left at the old site of the hole

The hole is reproduced at the site of the displaced electron

This electron stays where it is despite the electric field

Once again, a neighbouring electron jumps into the hole



The hole is reproduced at a new site

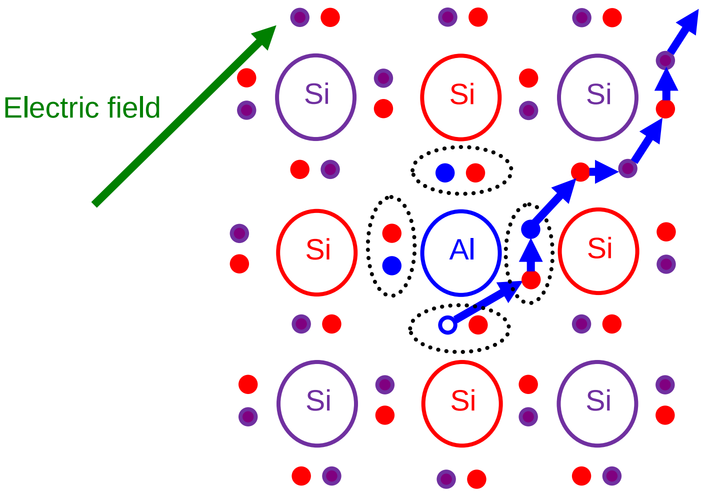
Once again, a neighbouring electron jumps into the hole

Both electrons stay where they are: they are immobile

The hole is reproduced at a new site: it keeps on moving in the direction of the electric field

And so on… The hole continues to migrate in the direction of the electric field, like any positive charge would do, as its vacancy continues to be filled by the next adjacent electron, reproducing the hole at the former site of this electron.

We can understand from this illustration that the individual, bound electrons involved in filling the vacancy of the propagating hole do not exhibit any continuous large-scale motion. Each of these electrons moves only once during the migration process. In contrast, a free electron moves continuously in the direction opposite the electric field.



The hole never stops moving because there is always a neighbouring electron that jumps into it → Electric current

An electron moves against the electric field to fill a neighbouring hole, and then stops moving unless a new hole appears next to it. The reason is that the electric field strength is not sufficient to break the covalent bond it has formed with another electron

There are clearly two types of electric conduction in a semiconductor:

- a first type of conduction, seen as a current of electrons, is due to the motion of free electrons that are not stuck in the crystal lattice;
- a second type of conduction, seen as a current of holes, is due to the motion of electrons that “jump” into adjacent covalent sites, thus remaining at all times stuck in the crystal lattice.

Therefore, the real distinction between these two conduction mechanisms only resides in the way electrons move, not in the particles that are involved (because, as we have seen, holes are only imaginary particles anyway).

4. Thermal Motion of Electrons and Holes

Even without an applied electric field, current carriers (electrons and holes) are not at rest, but possess finite kinetic energies. The average kinetic energy that an electron or hole can have is given by

$$E = \frac{3}{2}kT\ ,$$

where k and T designate the Boltzmann's constant ($\approx 1.38\times10^{-23}$ J/K) and temperature in degrees kelvin, respectively. Using this equation, we could show that the thermal velocity of a current carrier can be expressed as

$$v_{th} = \sqrt{\frac{3kT}{m}}\ ,$$

where m is the effective mass of the carrier (in silicon, $m \approx 2.37\times10^{-31}$ kg for electrons and $m \approx 3.55\times10^{-31}$ kg for holes).

At room temperature (T = 300 K), in silicon, the thermal velocity of an electron is thus

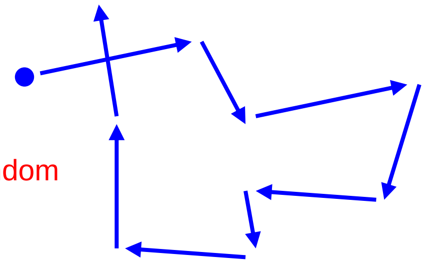
$$v_{th} = \sqrt{\frac{3\times1.38\times10^{-23}\ \mathrm{J\cdot K^{-1}}\times300\ \mathrm{K}}{2.37\times10^{-31}\ \mathrm{kg}}} \approx 2.3\times10^{5}\ \mathrm{m\cdot s^{-1}},$$

whereas the thermal velocity of a hole is equal to

$$v_{th} = \sqrt{\frac{3\times1.38\times10^{-23}\ \mathrm{J\cdot K^{-1}}\times300\ \mathrm{K}}{3.55\times10^{-31}\ \mathrm{kg}}} \approx 1.9\times10^{5}\ \mathrm{m\cdot s^{-1}}.$$

Electrons travel (a bit) faster than holes. It is worth noting that these velocities are about 1000 times lower than the speed of light ($c \approx 3\times10^{8}\ \mathrm{m\cdot s^{-1}}$).

Electrons and holes move at the thermal velocity but not in a simple straight-line fashion. Their directions of motion change frequently due to collisions or scattering with imperfections in the crystal. The carriers move in a zigzag fashion as shown below.



All current carriers are constantly moving in a random fashion, without going anywhere!

The *mean free time* between two successive collisions is typically around 10^{-13} second (0.1 picosecond), and the distance between collisions is a few tens of nano-meters. The *net* thermal velocity (averaged over time or over a large number of carriers at any given time) is zero. Thus, thermal motion does not create a steady electric current, but it does introduce a thermal noise.

5. Current Carriers in Intrinsic Semiconductors

In any semiconductor, both electrons and holes are present and contribute to the overall current. This is why they are generically called carriers. When a semiconductor is pure (not doped), it is called *intrinsic* and the concentration of its electrons equals that of its holes.

The presence of these electrons and holes is due to the thermal energy that keeps on breaking some electron-hole pairs all the time. In other words, every thermally-excited electron moves to the conduction band while leaving one hole behind in the valence band. At the same time, holes also recombine with free electrons, thus destroying hole-electron pairs.

At equilibrium, we can write the following equation:

$$n_0 \cdot p_0 = n_i^2 ,$$

where n_0 and p_0 are the equilibrium electron and hole concentrations, respectively, and n_i represents the semiconductor's intrinsic carrier concentration.

For an intrinsic semiconductor, we can write

$$n_0 = p_0 = n_i.$$

At room temperature (T = 300 K), we have $n_i \approx 10^{10}$ cm^{-3} for silicon, $n_i \approx 2\times10^{13}$ cm^{-3} for germanium, and $n_i \approx 10^{7}$ cm^{-3} for gallium arsenide.

The intrinsic carrier concentration n_i is a strong function of the energy band gap E_g and the temperature T, and can be expressed as

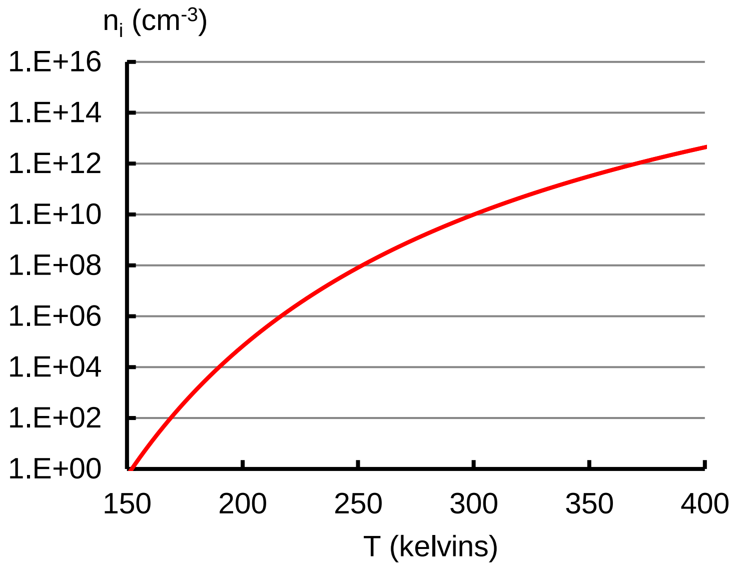
$$n_i = C \cdot T^{1.5} \cdot \exp\left(-\frac{E_{g,0}}{2kT}\right),$$

where C is a constant, $E_{g,0}$ is the zero-temperature energy band gap ($E_{g,0} \approx 1.17$ eV for silicon, 0.74 eV for germanium, 1.52 eV for gallium arsenide), and k denotes the Boltzmann constant expressed here in $eV.K^{-1}$ ($k \approx 8.62\times10^{-5}$ $eV.K^{-1}$).

In semiconductor materials with a larger band gap $E_{g,0}$, it is more difficult for electrons to move to the conduction band. This results into a lower intrinsic carrier concentration at a given temperature.

As the temperature is raised, more thermal energy is provided to electrons to move to the conduction band. This explains why the intrinsic carrier concentration increases with temperature.

As an illustration, the plot below shows the variation of the intrinsic carrier concentration n_i as a function of temperature T, in the case of silicon.



6. Current Carriers in Extrinsic (Doped) Semiconductors

When a material is intentionally doped, it is called *extrinsic* and the balance of electrons/holes density is then altered. However, at equilibrium, the product $n_0 \cdot p_0$ remains equal to n_i^2. This equation is valid for any dopant concentration.

N-type semiconductors

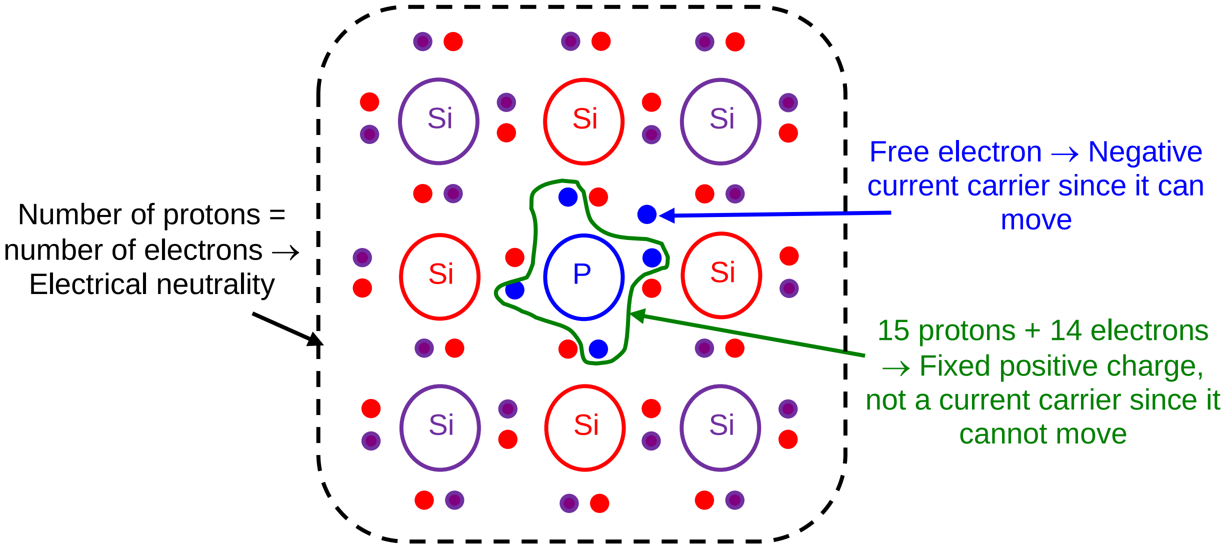
Consider first the case of an N-type doped material. One can say that it contains electrons in excess since there are more electrons than holes. In this case, the electrons are the *majority carriers*, while the holes are the *minority carriers* because they are actually also present but at a much lower concentration.

The concentration N_D of donor atoms (i.e., atoms that “donate” an electron such as arsenic, phosphorus, and antimony atoms) introduced in the crystal lattice by doping typically ranges from 10^{14} to 10^{20} cm^{-3}.

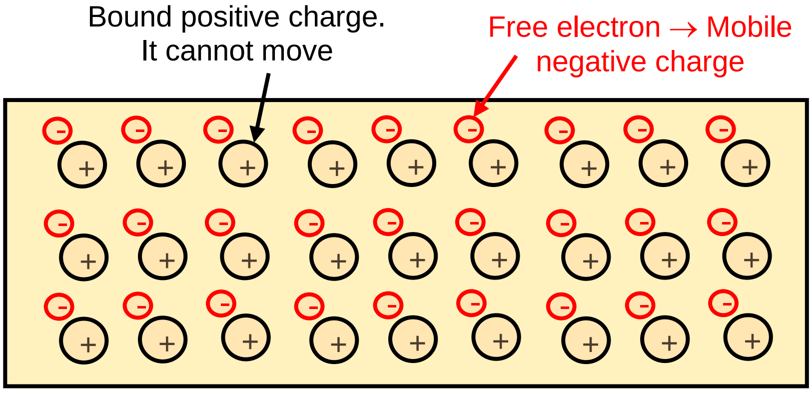
Each donor atom generates a free electron, and there is very little thermal generation of additional carriers so that $N_D >> n_i$ (this is true provided that the temperature is not too high). Therefore, the concentration of electrons in an N-type semiconductor is simply given by $n_0 \approx N_D$ cm^{-3}.

For instance, in an N-type silicon material with $N_D = 10^{15}$ cm^{-3}, we have $n_0 \approx 10^{15}$ cm^{-3} and $p_0 = n_i^2 / n_0 \approx n_i^2 / N_D = 10^5$ cm^{-3}. Clearly, electrons are the majority carriers, while holes are the minority carriers.

The fact that there is an excess of electrons in an N-type semiconductor does not imply that the material contains more negative charges than positive charges. In fact, the material remains electrically neutral. Having an excess of electrons means that there are more negative current carriers than positive current carriers, while there are more fixed positive charges than fixed negative charges.



Remember that each donor atom (arsenic, phosphorus, antimony) that has lost an electron on its valence shell becomes a fixed positive charge. It cannot move because it is “stuck” in the crystal lattice. We will later see that fixed charges play a crucial role in the operation of semiconductor devices because, despite being unable to carry any electric current themselves, their mere presence may be sufficient to generate electric fields which influence the behaviour of current carriers.



N-type silicon
(minority carriers are not shown as there are so few of them)

P-type semiconductors

Everything we have just said about an N-type material is of course applicable to a P-type material by simply replacing the word “electrons” by the word “holes” and vice versa.

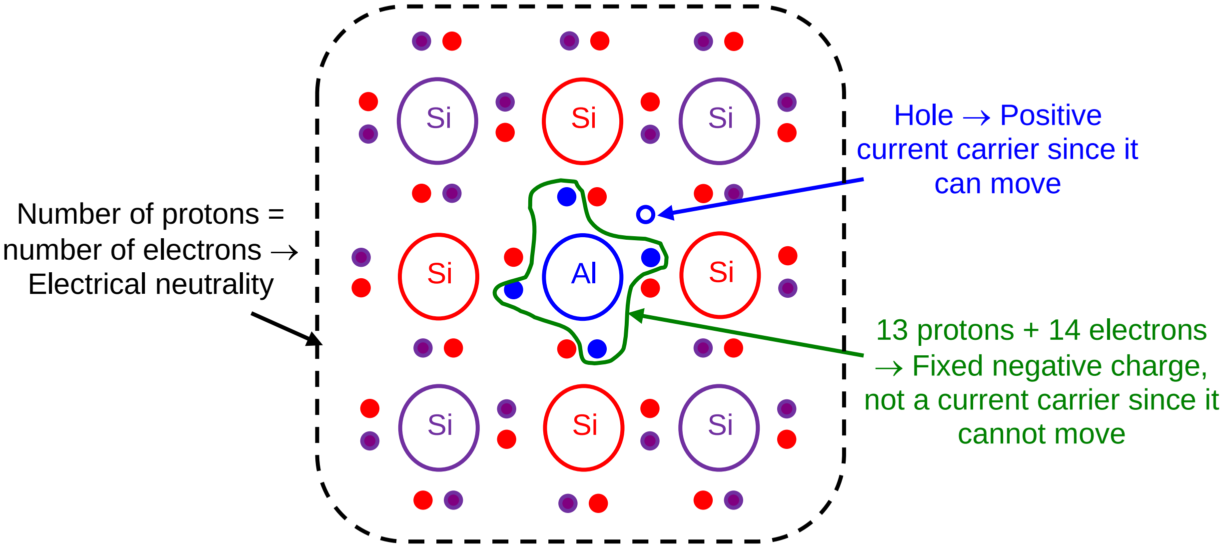
In a P-type doped material, there is an excess of holes and these holes are the majority carriers. The electrons are now the minority carriers because they are actually also present but at a much lower concentration than the holes.

The acceptor atoms (e.g., boron, aluminium, and gallium atoms) are atoms that can “accept” an electron, i.e. provide a hole. Their concentration N_A typically ranges from 10^{14} to 10^{20} cm^{-3}.

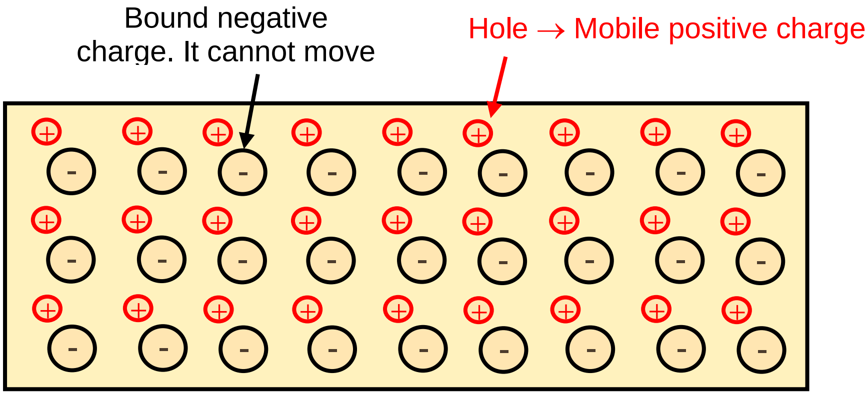
At “reasonable” temperatures, the concentration of holes is approximately equal to that of the acceptor atoms that have been introduced in the crystal lattice through doping: $p_0 \approx N_A$ cm^{-3}. This approximation is valid provided that the number of thermally generated electron-hole pairs is such that $n_i << N_A$, i.e. the temperature does not reach very high values.

For example, in a P-type silicon material with $N_A = 10^{17}$ cm^{-3}, we have $p_0 \approx 10^{17}$ cm^{-3} and $n_0 = n_i^2 / p_0 \approx n_i^2 / N_A = 10^3$ cm^{-3}. Clearly, holes are the majority carriers, while electrons are the minority carriers.

In a P-type semiconductor, holes are said to be in excess although the material remains electrically neutral. This means that there are more positive current carriers than negative current carriers, while there are more fixed negative charges than fixed positive charges.



Once again, we recall that each acceptor atom (boron, aluminium, gallium) that has accepted an electron on its valence shell, and thus generated a hole, becomes a fixed negative charge. This immobile charge cannot directly carry a current, but can create an electric field that has an influence on the motion of current carriers.



P-type silicon
(minority carriers are not shown as there are so few of them)

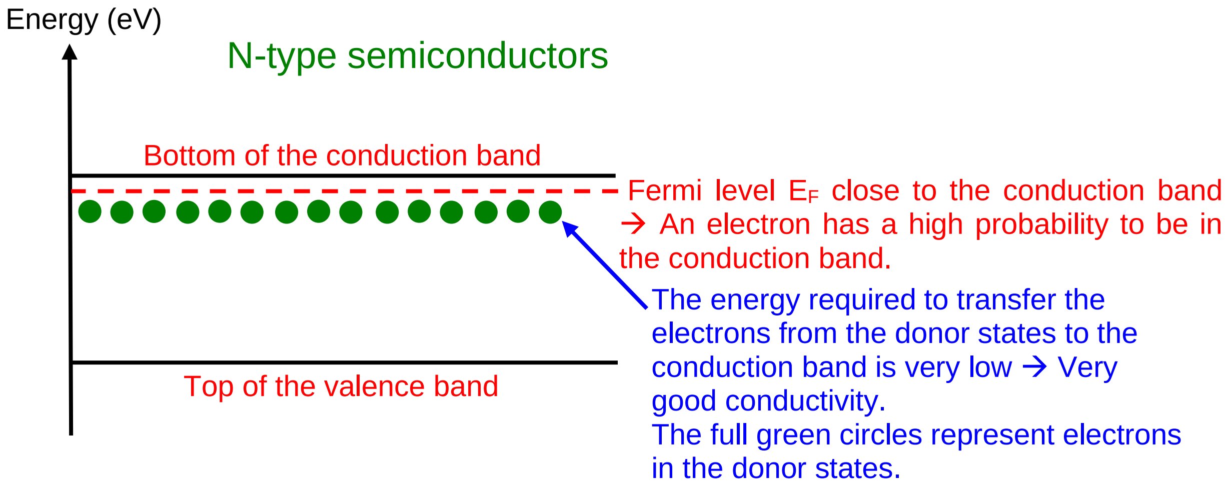
The role played by the minority charge carriers in both N-type and P-type semiconductors can often be ignored for the sake of simplicity and clarity.

Energy bands for doped semiconductors

When a semiconductor is doped, energy states are introduced in the band gap.

If it is doped with donor atoms (N-type doping), the energy states are called donor states. Because it takes very little energy, much less than the band gap energy, to free the electron that inhabits the donor state, the states are shown close to the conduction band.

Adding donor atoms, therefore, adds more electrons to the conduction band (without adding holes to the valence band) making the semiconductor more conductive.



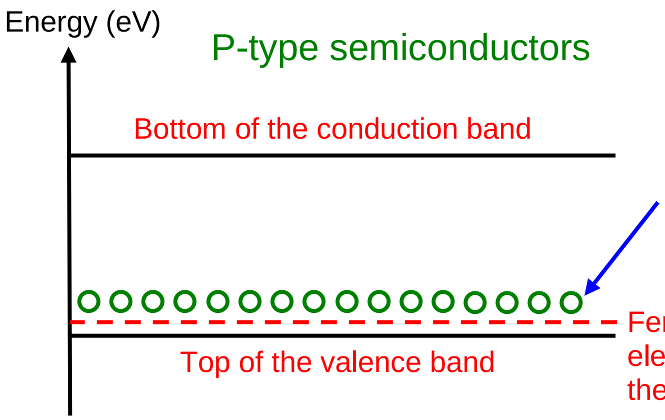
N-type semiconductors

Fermi level E_F close to the conduction band → An electron has a high probability to be in the conduction band.

The energy required to transfer the electrons from the donor states to the conduction band is very low → Very good conductivity.
The full green circles represent electrons in the donor states.

Acceptor states are introduced into the forbidden gap and close to the valence band if the semiconductor is doped with acceptor atoms (P-type doping). These initially empty states readily accept an electron to complete its bonds with the four nearest neighbours in the crystal.

When an electron from the valence band transitions to an acceptor state, it leaves behind a hole. The energy required for an electron to move to an acceptor state is much less than the band gap energy so these states are shown close to the valence band.



The energy required to transfer the electrons from the valence band to the acceptor states is very low → Very good conductivity.
The presence of extra holes (shown using empty green circles) in the band gap allows excitation of valence band electrons, leaving mobile holes in the valence band.

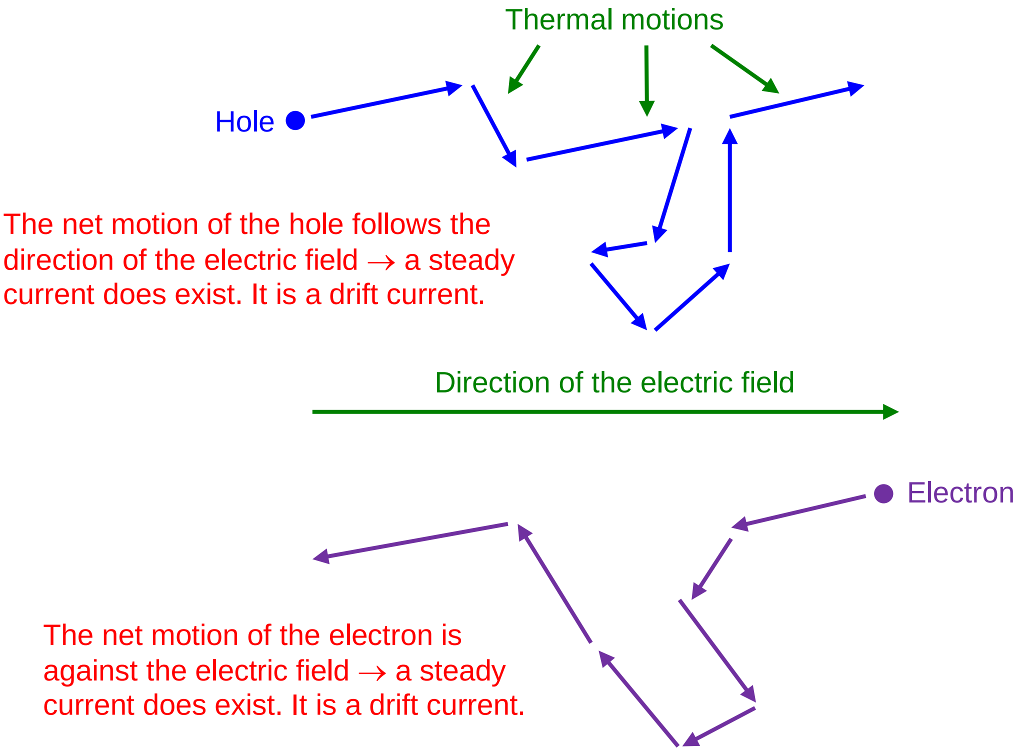
Fermi level E_F close to valence band → An electron has a very low probability to be in the conduction band.

7. Drift Current in Semiconductors

Drift is the motion of charge carriers caused by an electric field. The average velocity of the carriers is no longer zero when an electric field is applied to the semiconductor.

This nonzero velocity is called the *drift velocity*. The drift velocity is superimposed on the thermal motion as shown below.

A drift current is thus due to the movement of current carriers (electrons and holes) under the influence of an electric field E.



The net motion of the hole follows the direction of the electric field → a steady current does exist. It is a drift current.

The net motion of the electron is against the electric field → a steady current does exist. It is a drift current.

Mobility of carriers

For low to moderate values of the electric field E, the drift velocity of current carriers, expressed in $m.s^{-1}$, is proportional to the field strength:

$$v_d = \mu E,$$

where the quantity μ denotes the mobility of the current carriers (electron or hole). The mobility is clearly a metric of how mobile the current carriers are. A faster carrier velocity v_d is desirable as it allows a semiconductor device or circuit to operate at a higher speed.

Most engineers actually specify the mobility in $cm^2 \cdot V^{-1} \cdot s^{-1}$, although the SI unit of mobility is $m^2 \cdot V^{-1} \cdot s^{-1}$. They are related by $1\ m^2 \cdot V^{-1} \cdot s^{-1} = 10^4\ m^2 \cdot V^{-1} \cdot s^{-1}$. We recall that the unit of electric field is $V \cdot m^{-1}$.

Mobility values at room temperature (T = 300 K) for some lightly-doped semiconductors:

- Electrons:

- $\mu_n \approx 1400\ cm^2 \cdot V^{-1} \cdot s^{-1}$ for silicon,

- $\mu_n \approx 3900$ cm^2·V^{-1}·s^{-1} for germanium,
- $\mu_n \approx 8500$ cm^2·V^{-1}·s^{-1} for gallium arsenide.

- Holes:

- $\mu_p \approx 450$ cm^2·V^{-1}·s^{-1} for silicon,
- $\mu_p \approx 1900$ cm^2·V^{-1}·s^{-1} for germanium,
- $\mu_p \approx 400$ cm^2·V^{-1}·s^{-1} for gallium arsenide.

For instance, for the holes in silicon, with E = 1 kV·m^{-1}, we obtain $v_d \approx 450$ cm^2·V^{-1}·s^{-1} × 1 kV·cm^{-1} = 4.5×10^5 cm·s^{-1} = 4.5×10^3 m·s^{-1}.

It can be observed that the electrons in GaAs have a much higher mobility than those in Si. Thus, higher-speed devices (diodes and transistors) can potentially be made using GaAs rather than Si.

Semiconductor mobility depends on several parameters, such as the impurity concentrations (including donor and acceptor concentrations), defect concentration, temperature, and electron and hole concentrations.

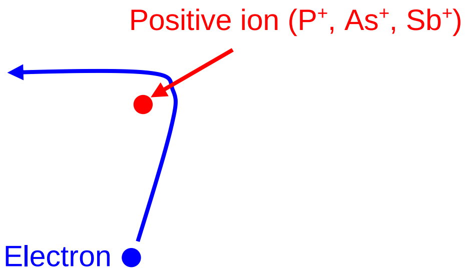
The main factor determining drift velocity is scattering time, i.e. how long the carrier is accelerated by the electric field until it scatters (collides) with something that changes its direction and/or energy. The two most important sources of scattering in typical semiconductor materials are *ionized impurity scattering* and *phonon scattering* (also called *lattice scattering*).

Ionized impurity scattering

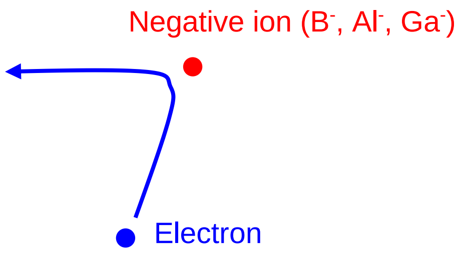
Semiconductors are doped with donors or acceptor atoms which then become positive or negative ions (charged atoms) embedded in the crystal lattice. The Coulombic forces deflect an electron or hole approaching the ionized impurity. This is known as ionized impurity scattering. The amount of deflection depends on the speed of the carrier and its proximity to the ion.

The more heavily a material is doped, the higher the probability that a carrier will collide with an ion in a given time, and the smaller the average time between two successive collisions. This leads to a reduction in carrier mobility.

The effect of ionized impurity scattering becomes less significant with increasing temperature because the average thermal velocities of carriers (associated with the motions due to thermal energy) are increased. Thus, the carriers spend less time near an ionized impurity as they pass and the scattering effect of the ions is thus reduced.



The electron is attracted by the positive ion and thus forced to change direction → Ionized impurity scattering event



The electron is repelled by the negative ion and thus forced to change direction → Ionized impurity scattering event

We conclude that, as far as ionized impurity scattering mechanisms are concerned, increases in temperature and decreases in dopant concentration are beneficial for the mobility of current carriers.

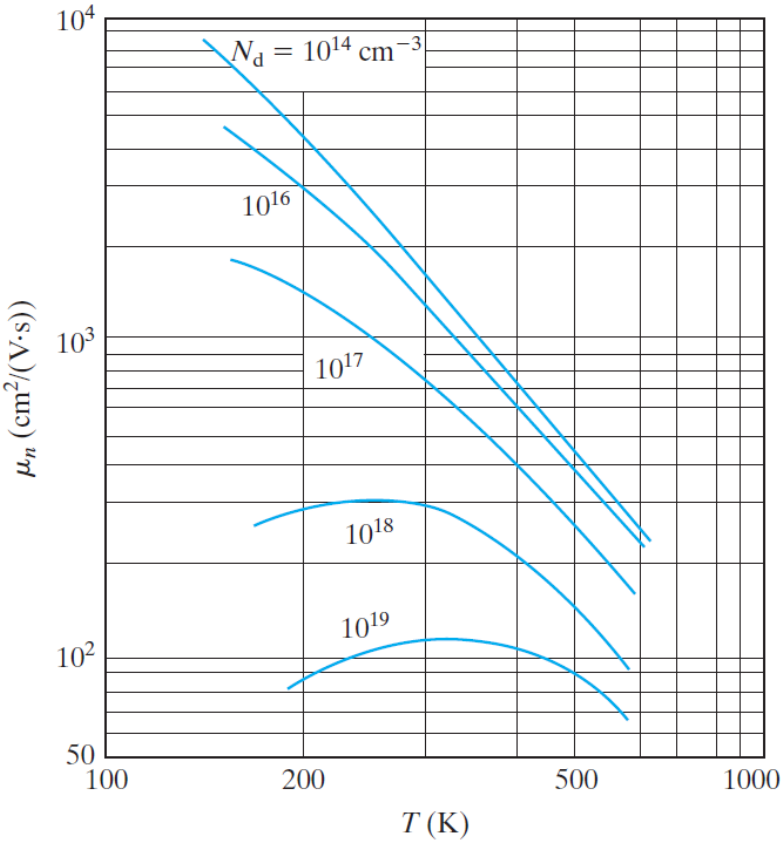
Lattice (phonon) scattering

At any temperature above absolute zero, the vibrating atoms create pressure waves in the crystal which are termed *phonons*. Phonons are the particle representation of the vibration of the atoms in the crystal. A phonon can interact (collide) with an electron or hole and scatter it.

With increasing temperature, phonon concentration increases and causes increased scattering. Thus, lattice scattering lowers the carrier mobility more and more at higher temperature.

Ionized impurity scattering and phonon scattering operate simultaneously on the carriers. At lower temperatures, ionized impurity scattering dominates, while at higher temperatures, phonon scattering dominates. Hence, the actual mobility can be expected to reach a maximum value at an intermediate temperature.

To confirm these statements, we show below the variation of the measured electron mobility in an N-type silicon material as a function of temperature, for several concentrations N_D of donor atoms.



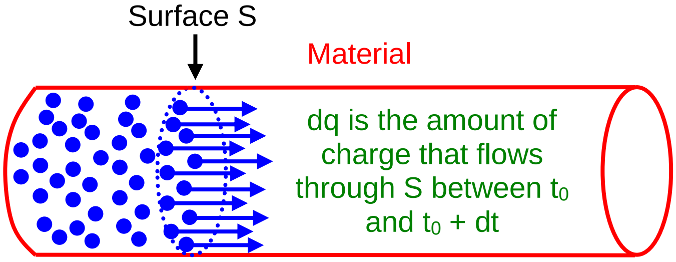
Drift current

We know that, under the influence of an electric field E, current carriers move with a net velocity given by $v_d = \mu E$. But, what is the expression of the drift current corresponding to the motion of charges?

We remember the basic definition of electric current: the current i represents the time rate at which electrical charge flows through a surface (e.g., the cross-section of a wire or a bar of semiconductor material), and is therefore expressed as

$$i = \frac{dq}{dt},$$

where dq is the total amount of charge that flows through the surface between times t_0 and $t_0 + dt$. The unit of current is the ampere (A) which is equivalent to a coulomb per second ($C{\cdot}s^{-1}$).

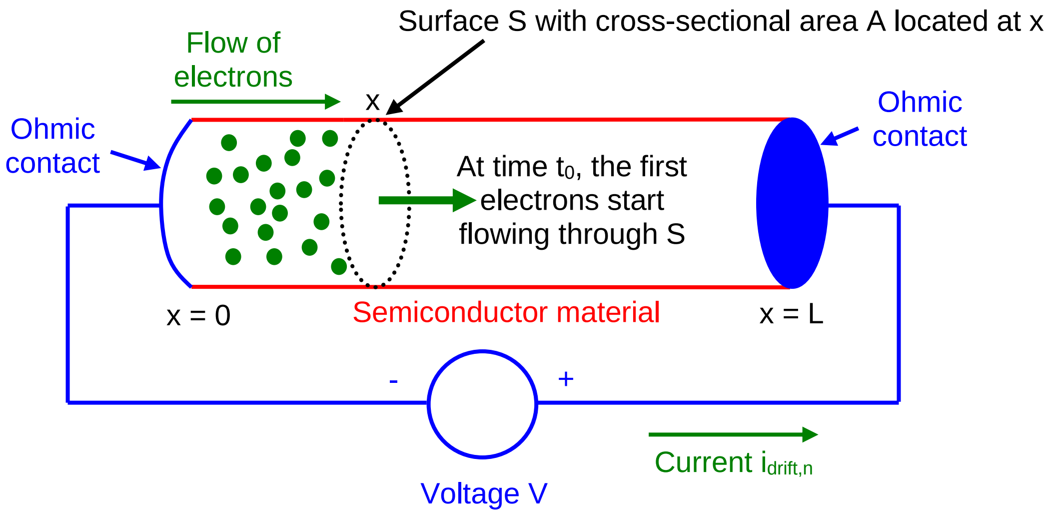


Consider a bar of semiconductor material connected to a voltage source as shown below.

Let us initially focus on the flow of electrons, while ignoring for now the flow of holes flowing in opposite direction inside the material.

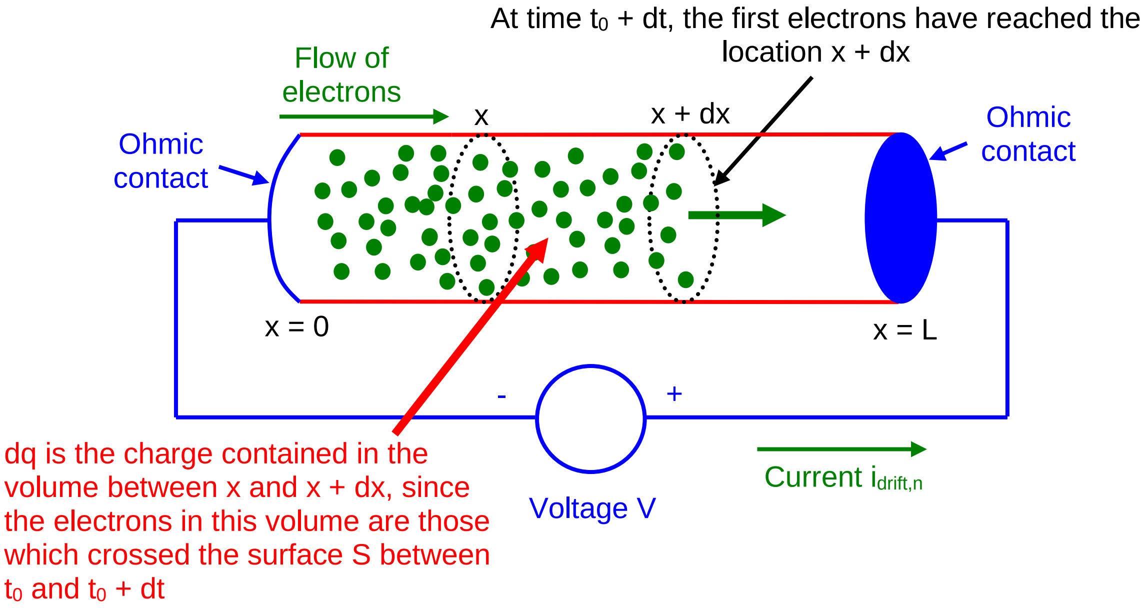
Consider hereafter a particular location x in the material. The index x ranges from x = 0 to x = L, where L is the length of the bar.

The charge dq is the charge contained in a volume of semiconductor material equal to A·dx, where A is the cross-sectional area of the semiconductor bar and dx is the distance travelled by the electrons during dt.



Surface S with cross-sectional area A located at x

At time t_0, the first electrons start flowing through S



At time t_0 + dt, the first electrons have reached the location x + dx

dq is the charge contained in the volume between x and x + dx, since the electrons in this volume are those which crossed the surface S between t_0 and t_0 + dt

Therefore, we can write the expression of the drift current due to the motion of electrons:

$$i_{\text{drift},n} = \frac{dq}{dt} = \frac{-qn_0Adx}{dt},$$

where -q is the charge of an electron ($q \approx 1.6\times10^{-19}$ coulomb) and n_0 is the equilibrium electron concentration.

The ratio dx/dt is simply the speed at which the electrons travel at the particular location x, i.e. the drift velocity $v_{d,n}(x)$ of these electrons. Therefore, we can write

$$i_{\text{drift},n} = -qn_0A\frac{dx}{dt} = -qn_0Av_{d,n}(x) = -qn_0A\mu_nE(x),$$

where E(x) is the electric field at location x inside the semiconductor bar. This electric field exists because of the voltage applied across the material. We can use the well-known expression linking both quantities:

$$E(x) = -\frac{dV(x)}{dx},$$

where $dV(x)$ denotes the voltage difference measured between locations x and x + dx. The expression of the drift current is thus given by

$$i_{\text{drift},n} = qn_0A\mu_n\frac{dV(x)}{dx},$$

which can be re-written as $i_{\text{drift},n}dx = qn_0A\mu_ndV(x)$.

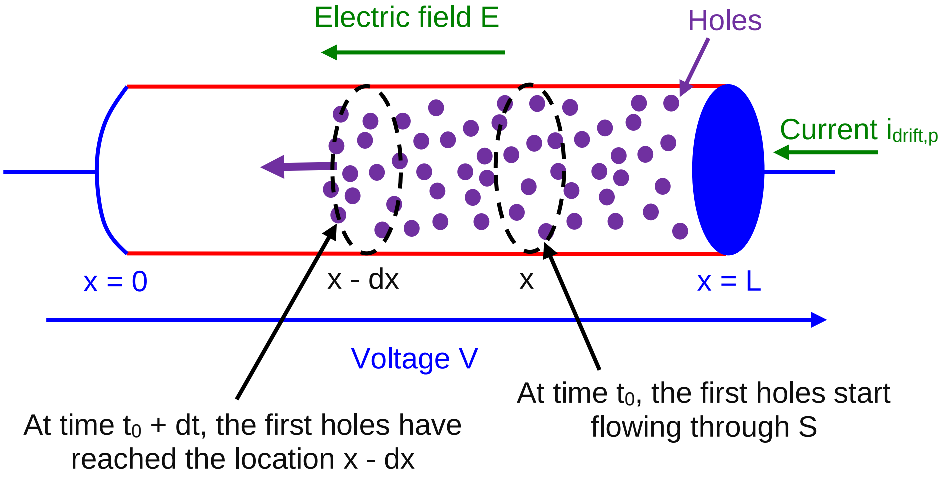
By performing the integration of this equation over the length L of the bar, we obtain

$$\int_0^L i_{\text{drift},n}dx = \int_0^V qn_0A\mu_ndV(x),$$

which is equivalent to $i_{\text{drift},n}\int_0^L dx = qn_0A\mu_n\int_0^V dV(x)$.

Finally, we obtain $i_{\text{drift},n} = \frac{qn_0A\mu_n}{L}V$, where V is the voltage applied across the material.

This result is obtained by only taking into account the drift current due to electrons. We must however also consider the drift current created by the motion of holes within the semiconductor material.



The drift current due to holes is given by

$$i_{\text{drift},p} = qp_0A\left(-\frac{dx}{dt}\right) = -qp_0Av_{d,p}(x) = -qp_0A\mu_pE(x),$$

where the quantities q, p_0, $v_{d,p}$, and μ_p denote the charge of a hole (q ≈ 1.6×10^{-19} coulomb), the equilibrium hole concentration, the drift velocity of holes, and the hole mobility, respectively.

Using the expression

$$E(x) = -\frac{dV(x)}{dx},$$

we obtain $i_{\text{drift},p} dx = q p_0 A \mu_p dV(x)$, which, after integration over the length of the semiconductor bar, leads to $i_{\text{drift},p} = \frac{q p_0 A \mu_p}{L} V$.

The total drift current flowing through the semiconductor bar is given by the sum of both drift currents:

$$i_{\text{drift}} = i_{\text{drift},n} + i_{\text{drift},p} = \left(n_0\mu_n + p_0\mu_p\right)\frac{qA}{L}V.$$

This expression is actually the well-known Ohm's law: the voltage across the semiconductor bar is proportional to the (drift) current flowing through it. The resistance R of this bar, expressed in ohms (Ω), is given by

$$R = \frac{V}{i_{\text{drift}}} = \frac{L}{qA\left(n_0\mu_n + p_0\mu_p\right)}.$$

The resistance R depends on the physical dimensions A and L of the piece of material. To study the characteristics of the resistance thus created without having to worry about its physical dimensions, we generally prefer focusing on the expression of the resistivity ρ defined as

$$\rho = R\frac{A}{L} = \frac{1}{q\left(n_0\mu_n + p_0\mu_p\right)}.$$

We recall that electrical resistivity is an intrinsic property that quantifies how strongly a given material opposes the flow of electric current. For instance, a low resistivity indicates a material that readily allows the flow of electric current.

Since the parameters R, L, and A, are expressed in ohms (Ω), meters (m), and square meters (m^2), respectively, the unit of resistivity is the ohm·meter (Ω·m) or sometimes ohm·centimetre (Ω·cm).

As an illustration, the resistivity ρ of silicon doped with $N_D = 10^{17}$ cm^{-3} of arsenic is given by

$$\rho = \frac{1}{q(n_0\mu_n + p_0\mu_p)} \sim \frac{1}{qN_D\mu_n} \sim 4.5 \times 10^{-2}\ \Omega \cdot \mathrm{cm}.$$

For example, the resistance R of a piece of this silicon material with L = 1 μm and A = 0.1 μm^2 would be equal to $R = \rho L/A \approx 4.5\times10^{-2}$ Ω.cm × 10^{-4} cm / 0.1×10^{-8} cm^2 = 4.5 kΩ.

It is worth noting that the reciprocal of the resistivity ρ is the conductivity σ. This latter quantity, expressed in siemens per meter ($S \cdot m^{-1}$) or sometimes siemens per centimetre ($S \cdot cm^{-1}$), measures a material's ability to conduct an electric current. Here, we obtain the following expression for the conductivity of a semiconductor bar:

$$\sigma = \frac{1}{\rho} = q\left(n_0\mu_n + p_0\mu_p\right).$$

Drift velocity saturation

This equation $v_d = \mu E$ states that one can make a carrier move as fast as one likes just by increasing the electric field. In fact, this is not possible. As the electrical field increases, the kinetic energy of the carriers rises. However, at some stage, the kinetic energy cannot be increased any further.

The culprit is *optical phonon scattering*. Optical phonons are high-energy phonons that interact strongly with electrons and holes. When the kinetic energy of a carrier exceeds the energy of an optical phonon, it generates an optical phonon and loses much of its velocity as a result. Consequently, the kinetic energy and therefore the drift velocity cannot exceed a certain value. The limiting velocity is called the *saturation velocity* and is denoted as v_{sat}.

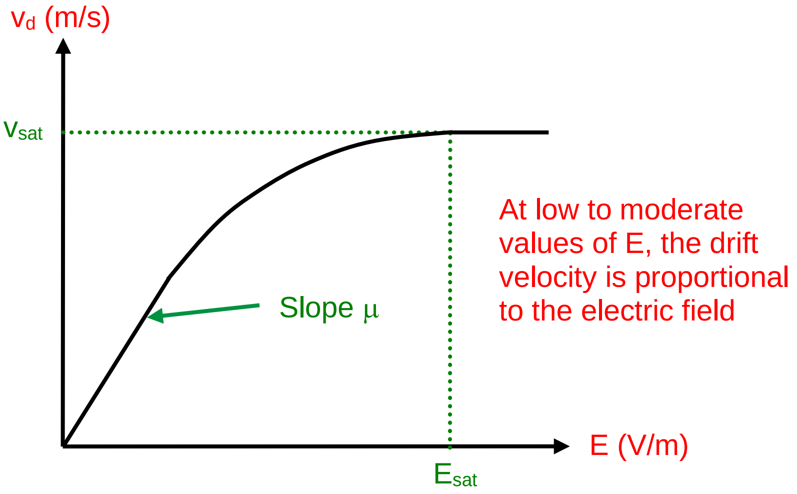
Once the carrier velocity saturates, the expression $v_d = \mu E$ is no longer valid. Instead, one can use the more general following expression for the drift carrier velocity:

$$v_d = \frac{\mu E}{1 + \frac{E}{E_{sat}}} \text{ for } E \le E_{sat},$$

and $$v_d = v_{sat} \text{ for } E > E_{sat},$$

where E_{sat} is the electric field at which the drift velocity of current carriers becomes constant and v_{sat} is the saturated drift velocity of the charge carriers under consideration (electrons or holes). In this equation, the quantity E_{sat} is actually given by

$$E_{sat} = \frac{2v_{sat}}{\mu}.$$



Note that the saturation velocity depends on the temperature. However, unlike the mobility, it is not dependent on the doping level.

Saturation velocity values for some semiconductors (at T = 300 K)

- Electrons:

- $v_{sat} \approx 10^5$ m·s^{-1} for silicon,
- $v_{sat} \approx 0.7\times10^5$ m·s^{-1} for germanium and gallium arsenide.

- Holes:

- $v_{sat} \approx 0.7\times10^5$ m·s^{-1} for silicon,
- $v_{sat} \approx 0.6\times10^5$ m·s^{-1} for germanium,
- $v_{sat} \approx 0.9\times10^5$ m·s^{-1} for gallium arsenide.

As an illustration, in silicon, the electric field E_{sat} at which the drift velocity saturates is given by

$$E_{sat} \approx \frac{2 \times 10^{7}\ \mathrm{cm \cdot s^{-1}}}{1400\ \mathrm{cm^{2} \cdot V^{-1} \cdot s^{-1}}} \approx 14\ \mathrm{kV \cdot cm^{-1}} \text{ for electrons,}$$

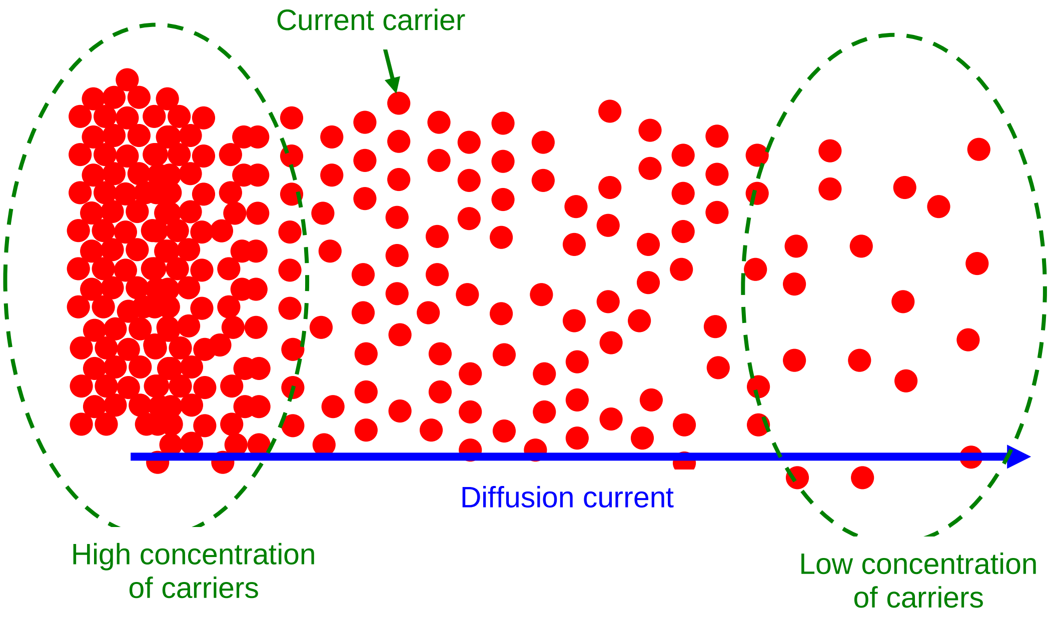
and

$$E_{sat} \approx \frac{2 \times 0.7 \times 10^{7}\ \mathrm{cm \cdot s^{-1}}}{470\ \mathrm{cm^{2} \cdot V^{-1} \cdot s^{-1}}} \approx 30\ \mathrm{kV \cdot cm^{-1}} \text{ for holes.}$$

8. Diffusion Current in Semiconductors

In addition to the drift current, there is a second component of current that can flow in a semiconductor. It is known as *diffusion current*.

This current can exist without any electric field applied across the material. It is due to the movement of current carriers from an area where there is a high concentration of carriers to another area with a lower concentration of carriers, as illustrated in the drawing shown below.



Diffusion current is not a relevant process in metals because there is no mechanism by which a density gradient can originate. Since only negative charge carriers (electrons) are present in a metal, any carrier gradient that might form would upset charge neutrality. The resulting local electric field would cause a drift current of electrons that would instantly eliminate the gradient before any diffusion process could take place.

In contrast, a semiconductor has both positive and negative charge carriers (holes and electrons); hence it is possible for a density gradient of holes and electrons to exist while charge neutrality is maintained. In fact, the ease of creating non-uniform carrier densities makes diffusion an important process in semiconductors.

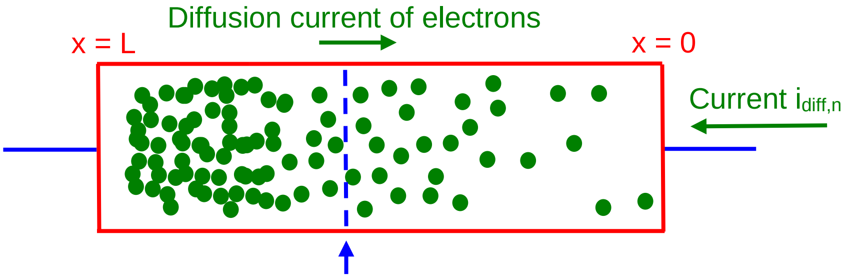
The intensity of the diffusion current depends on the gradient of concentration. In other words, if the concentration of current carriers is not uniform, there will be creation of a corresponding diffusion current proportional to the gradient of the carrier concentration:

$$i_{diff,n} = qD_nA\frac{dn(x)}{dx} \text{ for electrons,}$$

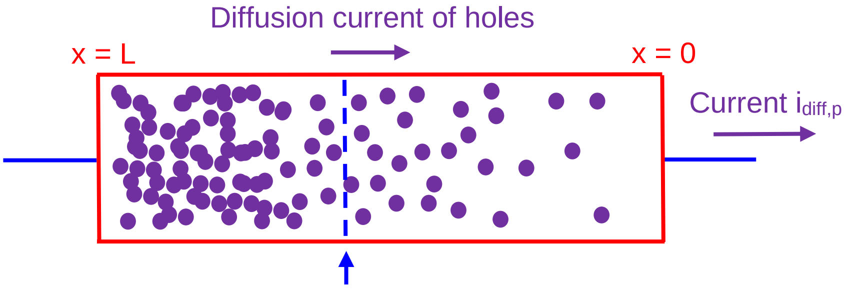
and

$$i_{diff,p} = -qD_pA\frac{dp(x)}{dx} \text{ for holes.}$$

In these expressions, q represents the elementary charge ($q \approx 1.6\times10^{-19}$ coulomb), D_n and D_p are the diffusion constants for electrons and holes, respectively, A is the cross-sectional area, and n(x) and p(x) denote the concentrations of electrons and holes, respectively, at location x along a semiconductor bar.



n(x) is the electron density at a particular location x.
Note that dn(x)/dx > 0 here.



p(x) is the hole density at a particular location x.
Note that dp(x)/dx > 0 here.

The larger the D_n and D_p constants are, the faster the electrons and holes diffuse.

The diffusion and mobility constants for a given carrier species are related by a basic equation of quantum statistics called the Einstein relation:

$$D_n = \frac{kT}{q}\mu_n \text{ and } D_p = \frac{kT}{q}\mu_p,$$

where T denotes the temperature in degrees Kelvin (K) and k designates the Boltzmann's constant ($k \approx 1.38\times10^{-23}$ J/K). This close relationship between mobility and diffusion constants should not come as a surprise since all scattering mechanisms (e.g., phonon and impurity scatterings) that impede carrier drift would also impede carrier diffusion. In fact, it is rather easy to understand that the higher the mobility of a current carrier, the faster it diffuses within a semiconductor material.

The total diffusion current can be expressed by the equation

$$i_{diff} = i_{diff,n} + i_{diff,p} = qD_nA\frac{dn(x)}{dx} - qD_pA\frac{dp(x)}{dx}.$$

Total current flow

Drift and diffusion currents can exist together in a semiconductor, and be in the same or opposite direction. The total current of either carrier in a semiconductor is equal to the sum of its own drift and diffusion components, as expressed by the electron and hole *transport equations*:

$$i_n = i_{drift,n} + i_{diff,n} = q\,A \cdot \left(n_0\,\mu_n\,E + D_n\,\frac{d\,n(x)}{d\,x} \right)$$ for electrons,

and $$i_p = i_{drift,p} + i_{diff,p} = q\,A \cdot \left(p_0\,\mu_p\,E - D_p\,\frac{d\,p(x)}{d\,x} \right)$$ for holes.

- END -