Semiconductor Technology from A to Z

Fundamentals

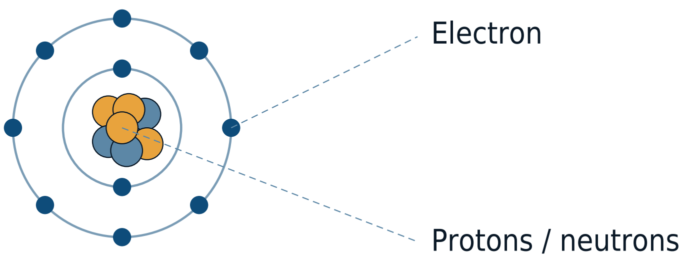
The atomic structure

The Atomic Model

An atom [Gr. átomos: indivisible] is the smallest building block of matter that cannot be further divided by chemical means. There are various types of atoms, each composed of a specific number of electrons, protons, and neutrons. Positively charged protons and uncharged neutrons form the atomic nucleus, which is orbited by negatively charged electrons at certain distances. An atom occurring in nature is electrically neutral, meaning it contains exactly as many protons as electrons. The number of neutrons, however, can vary.

According to the Bohr atomic model, the electrons are located on so-called shells, which represent different energy levels and are arranged concentrically around the atomic nucleus. There are a maximum of seven shells, which can hold varying numbers of electrons; the electrons in the outermost shell are also referred to as valence electrons.

Highly simplified representation of a neon atom

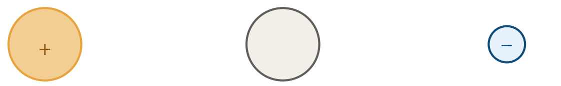


By exchanging electrons with other atoms, atoms can achieve a more stable state, which is why different types of bonds form between atoms.

Orders of Magnitude

Mass
: The mass of an atom is determined mainly by the atomic nucleus, since the mass of protons and neutrons, at $1.67 \cdot 10^{-27}$ kg, is roughly 1800 times greater than that of the electrons in the electron shell ($9.11 \cdot 10^{-31}$ kg).

Building blocks of the atom

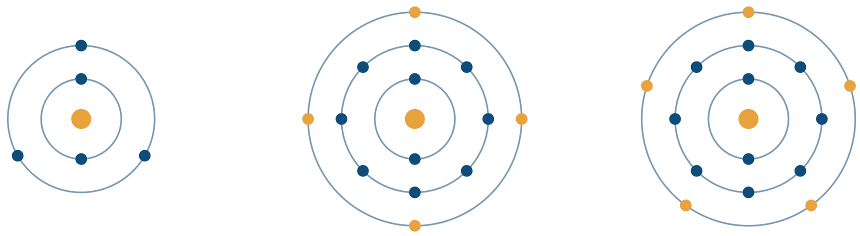


Particle	Proton	Neutron	Electron
Charge	+1	0	−1
Mass	$1.672 \cdot 10^{-27}$ kg	$1.675 \cdot 10^{-27}$ kg	$0.0009 \cdot 10^{-27}$ kg

Dimensions
: The diameter of the electron shell is 0.1–0.5 nm (0.0000000001 m, or 1 ångström); the diameter of the atomic nucleus is smaller by a further factor of 100,000. To illustrate this, imagine a pinhead at the center of a football field representing the atomic nucleus. The distance to the corner flags would then correspond to the distance at which the electrons orbit the nucleus.

Density
: Within the nucleus of an atom, protons and neutrons are packed extremely tightly. If Earth were compressed to the same density, its radius would shrink from its original 6,700,000 m to just 100 m.

Atoms of major importance in the semiconductor industry



The elements, the periodic table

The Elements

An element consists of many identical atoms (meaning atoms with the same number of protons, i.e. the same atomic number) and is a substance that cannot be broken down further by chemical means. The mass of an element is determined solely by the number of protons and neutrons, since the mass of the electrons is negligibly small. Hydrogen, with one proton and no neutrons, has a mass number of 1; the next heavier element, helium, has a mass number of 4 (2 protons + 2 neutrons). The mass number gives the number of particles in the nucleus. Multiplying it by the atomic mass unit u = $1.660 \cdot 10^{-27}$ kg gives approximately the mass of an atom – for helium, $6.64 \cdot 10^{-27}$ kg.

Chemical elements are usually named after the initial letters of their Latin or Greek names (hydrogen [H] from Latin hydrogenium, lithium [Li] from Greek lithos).

The Periodic Table of the Elements

The periodic table of the elements arranges all chemical elements in order of increasing proton number (nuclear charge, atomic number), grouped according to their chemical properties into periods as well as main and subgroups.

The period indicates the number of electron shells, while the main group indicates the number of electrons in the outermost shell (1 to 8 electrons). Groups 1 and 2, as well as 13–18, form the main groups; groups 3–12 form the subgroups.

The first element with one shell (period 1) and one outer electron (group 1) is hydrogen, H. The next element, helium, He, has only one electron shell just like hydrogen and is therefore also in period 1. Since the first shell is already completely filled with two electrons, helium is not in group 2 but in group 18 (the noble gas group).

To accommodate further electrons, a new shell must be started. Lithium is therefore found in group 1, period 2 (two electrons in the first shell, one valence electron in the second shell). A shell can hold a maximum of $2n^2$ electrons, where n stands for the period.

After the first two valence electrons in the outermost shell have been filled in groups 1 and 2, starting from the fourth period further inner shells are first completed with electrons before the outermost shell is fully filled with electrons across groups 13 to 18.

The Periodic Table of the Elements

Tap an element for details.

Important elements in semiconductor technology

Element	Particles	Note
B: Boron	5p, 6n, 5e	3 outer electrons: used for p-type doping of silicon
N: Nitrogen	7p, 7n, 7e	Stable N_2 molecule: protective and purge gas, protective layers on the wafer

O: Oxygen	8p, 8n, 8e	Highly reactive: oxidation of silicon, insulating layers (SiO_2), among others
F: Fluorine	9p, 10n, 9e	The most reactive element: used together with other substances for etching (e.g. HF, CF_4)
Si: Silicon	14p, 14n, 14e	Base material in semiconductor technology
P: Phosphorus	15p, 16n, 15e	5 outer electrons: used for n-type doping of silicon

Elements on the left side of the periodic table are metals. These tend to give up valence electrons in order to reach a stable electron configuration (the noble gas configuration). On the right side of the periodic table are the nonmetals, which take up additional electrons to reach the noble gas configuration. Between them are the metalloids, such as silicon and germanium.

Chemical bonds

Tendency to Form Bonds

Electrons in the outermost shell can detach from the atom (for example, by adding energy in the form of heat) and be exchanged with other atoms. Compounds formed from several elements are called molecules. The reason for this tendency to form bonds is to reach an outermost shell fully occupied with eight electrons: the so-called noble gas configuration. Substances that have already reached a full outer shell do not form compounds (a few exceptions exist, such as xenon–fluorine compounds).

A distinction is mainly made between three different types of bonds, which are explained in more detail below.

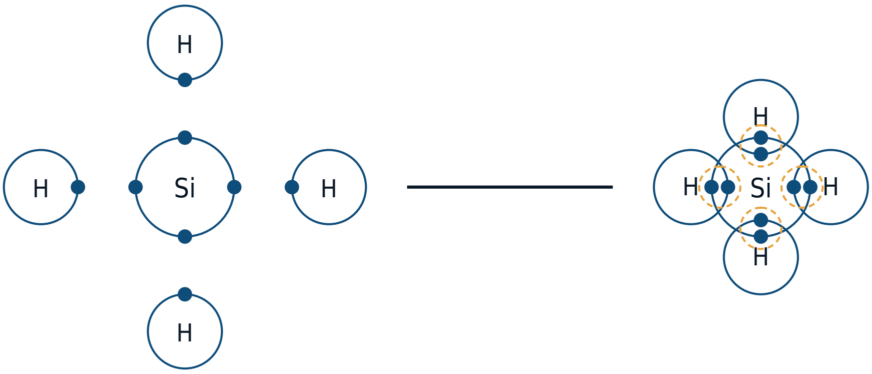
The Covalent Bond

Nonmetals form this type of bond to complete their electron octet. For example, two fluorine atoms (each with seven outer electrons) can fill their electron octet by mutually sharing one electron. The distance between the two atomic nuclei represents a compromise between the attraction of the nuclei to the bonding electrons and the repulsion between the two nuclei. The reason for covalent bonds is nature's tendency to reach the lowest possible energy state. Because combining several atoms into a molecule gives the electrons "more room", which corresponds to a lower energy, covalent bonding occurs in the first place.

The tendency to reach a fully occupied outer shell means that elemental fluorine never occurs as single atoms, but always as fluorine molecules: F_2. The same is true for nitrogen (N_2), oxygen (O_2), chlorine (Cl_2), bromine (Br_2) and iodine (I_2). Because of the shared electron pairs, this bond is also called an electron pair bond or covalent bond.

The following figure shows a covalent bond using silane (SiH_4) as an example (only the outermost shell of the silicon atom is shown). Through this bond, the silicon atom reaches a full outer shell of eight electrons, while the hydrogen atoms reach their first shell, already fully occupied with two electrons.

Covalent bond



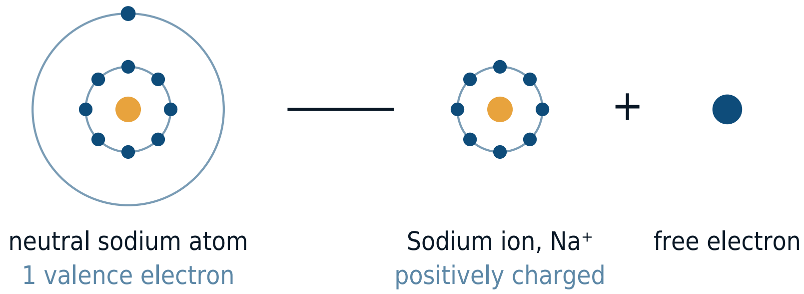
Four shared electron pairs complete the bond

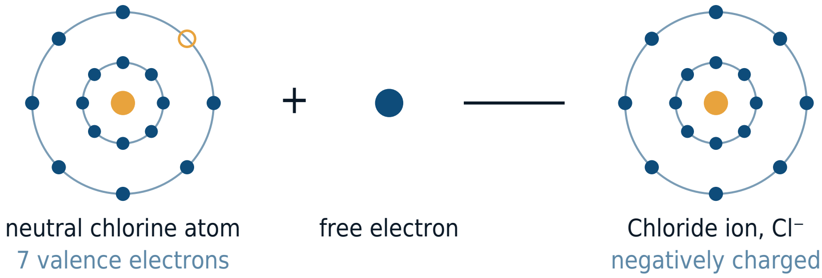
The Ionic Bond

Ionic bonds form when metals combine with nonmetals. Metals tend to give up electrons in order to reach a completely filled outer shell, while nonmetals can take up additional electrons. An example of an ionic bond is sodium chloride (NaCl, table salt).

The sodium atom gives up its valence electron (leaving it with more protons than electrons, so it becomes positively charged), while chlorine takes up an electron and becomes singly negatively charged. The two atoms attract each other because of their opposite charges. A charged atom is called an ion; a distinction is made between cations (positive charge) and anions (negative charge).

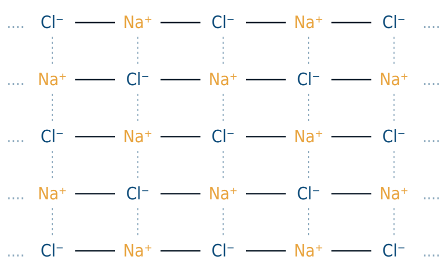
Principle of the ionic bond





Since atoms always occur in very large numbers, the attractive and repulsive forces cause them to arrange themselves into a regular ionic lattice. Substances that form such a lattice in the solid state are called salts.

Crystal lattice of the ionic bond in sodium chloride

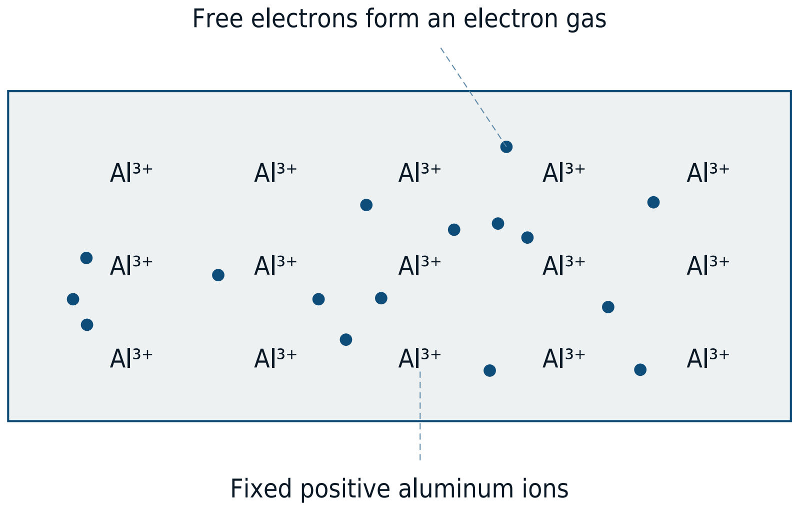


The Metallic Bond

Metals form this type of bond to reach the stable noble gas configuration. To do so, each metal atom gives up its outer electrons: positively charged metal ions (atomic cores) are formed along with free electrons, between which strong attractive forces exist. The metal ions repel one another, just as the electrons repel one another.

Because the attractive and repulsive forces act in all directions in space, the atomic cores arrange themselves into a regular lattice. The freely moving electrons occupy the spaces in between as a so-called electron gas, which holds the positive metal ions together. Because of these freely moving electrons, metals conduct electric current very well.

Metallic bond



The physical and chemical properties of compounds depend on the type of bond. Stronger attractive forces, for instance, mean higher melting and boiling points, while the number of free electrons affects conductivity.

Weak Bonds

Weak bonds are not true bonds within molecules, but rather interactions between molecules that are usually caused by shifts in charge within a molecule. This creates positive and negative centers of charge, through which molecules attract one another.

The strongest of the weak bonds is the hydrogen bond; in the water molecule (H–O–H) it results from the bond angle between the oxygen and hydrogen atoms. This causes the electric charge within the molecule to be distributed asymmetrically (negative at the oxygen, positive at the hydrogen), so that neighboring molecules attract one another. There are also other bonds, such as the van der Waals bond, which are weaker still.

The bond energy of chemical bonds ranges from a few hundred to a few thousand kilojoules per mole (kJ/mol). For hydrogen bonds it is up to 100 kJ/mol, and for van der Waals forces it is 0.5–5 kJ/mol.

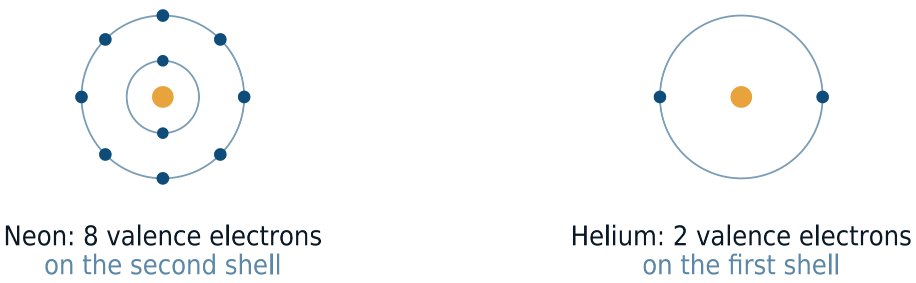
Noble gases

Noble Gases and the Electron Octet

Noble gases are the elements in the eighth main group of the periodic table: helium, neon, argon, krypton, xenon, radon (from top to bottom).

What makes these elements special is that they have eight electrons in their outermost shell. This electron octet represents an energetically very stable state that all elements strive to reach. The chapter on bonds explains how elements reach a full outer shell of eight electrons. Because of the stable noble gas configuration, these elements undergo almost no reactions (a few xenon–fluorine compounds, among others, do exist).

Noble gases



The neon atom has 10 electrons: two in the first shell and eight valence electrons in the second, outermost shell. Exception: the helium atom, which has only one shell, already reaches the noble gas state with two electrons (electron duet).

As inert gases [Lat. inert: idle, sluggish], nitrogen (e.g. as a purge gas in furnace processes) and argon (e.g. in sputtering processes), among others, are used in semiconductor manufacturing.

Conductors – Insulators – Semiconductors

Conductors

Conductors are, in general, substances that have the property of transmitting different forms of energy. In the following, the focus is on electrical conductivity.

Metals

The conductivity of metals is based on the free electrons that exist as an electron gas within the metallic bond. Even a small amount of energy is enough to detach sufficient electrons from the atoms to achieve conductivity.

Metallic bond: fixed atomic cores and free valence electrons (electron gas)

Conductivity depends, among other things, on temperature. As temperature rises, the atomic cores vibrate more strongly, hindering the movement of the electrons and consequently increasing resistance. The best conductors, gold and silver, are used relatively rarely due to their high cost (gold, for example, is used to contact finished chips). The alternatives used in semiconductor technology for wiring the individual components of a chip are aluminium and copper.

Salts

In addition to metals, salts can also conduct electric current. Here, however, there are no free electrons. Conductivity is instead based on ions, which detach from the lattice structure and become freely mobile when a salt melts or dissolves (see the chapter on bonds).

Insulators

Insulators have no free charge carriers in the form of electrons or ions. Insulators are also referred to as non-conductors.

Covalent bond

The covalent bond is based on shared electron pairs between nonmetals. Elements with nonmetallic character all tend to take up electrons, meaning no free electrons are available that could enable conductivity.

Ionic bond

In the solid state, ions are arranged within a lattice structure. The particles are held together by electrical forces. There are no free charge carriers available for current flow. Consequently, substances composed of ions can be either conductors (when dissolved) or insulators.

Semiconductors

Semiconductors are solids whose conductivity lies between that of conductors and insulators. Through the exchange of electrons between atoms of the same kind, in order to complete the electron octet, these atoms arrange themselves into a lattice structure. Unlike metals, conductivity increases with rising temperature – up to a certain point.

As the temperature rises, bonds break and electrons are released. At the location where the electron used to be, a so-called defect electron (also called a hole) remains.

Section of a silicon crystal

Ihr Browser unterstützt keine eingebetteten SVG-Animationen.

The flow of electrons is based on the intrinsic conductivity of semiconductors. The so-called band model illustrates why semiconductors behave this way.

The Band Model

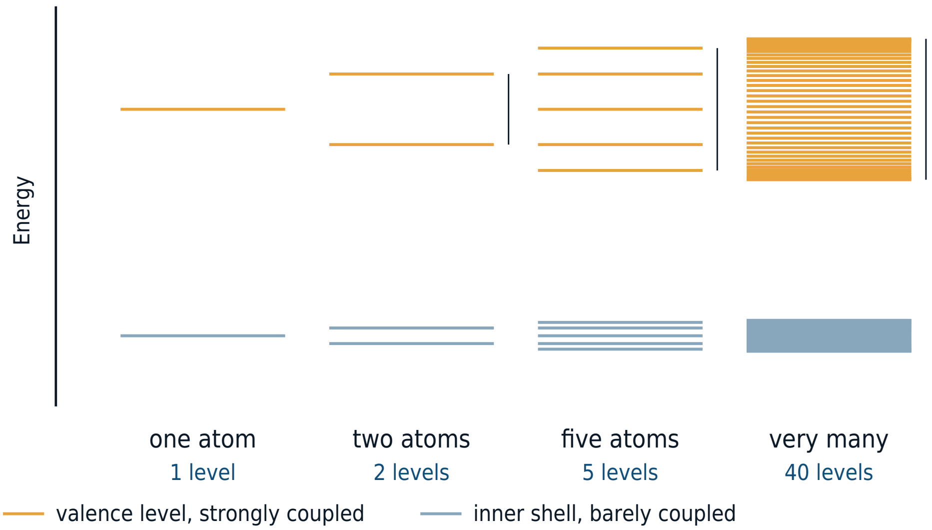
The band model is an energy diagram used to describe the conductivity of conductors, insulators, and semiconductors. The model consists of two energy bands (the valence band and the conduction band) and the band gap. The valence electrons – which serve as charge carriers – are located in the valence band; in the ground state, the conduction band is not occupied by electrons. Between the two energy bands lies the band gap, whose width, among other things, influences conductivity.

The Energy Bands

Looking at a single atom, according to the Bohr atomic model there are sharply separated energy levels that can be occupied by electrons. When several atoms are located next to one another, they interact with each other and the discrete energy levels fan out. In a silicon crystal there are approximately 5×10^{22} atoms per cubic centimetre, so that the individual energy levels are no longer distinguishable from one another and instead form broad energy ranges.

Splitting of Energy Levels

The more atoms interact, the more closely spaced the levels become

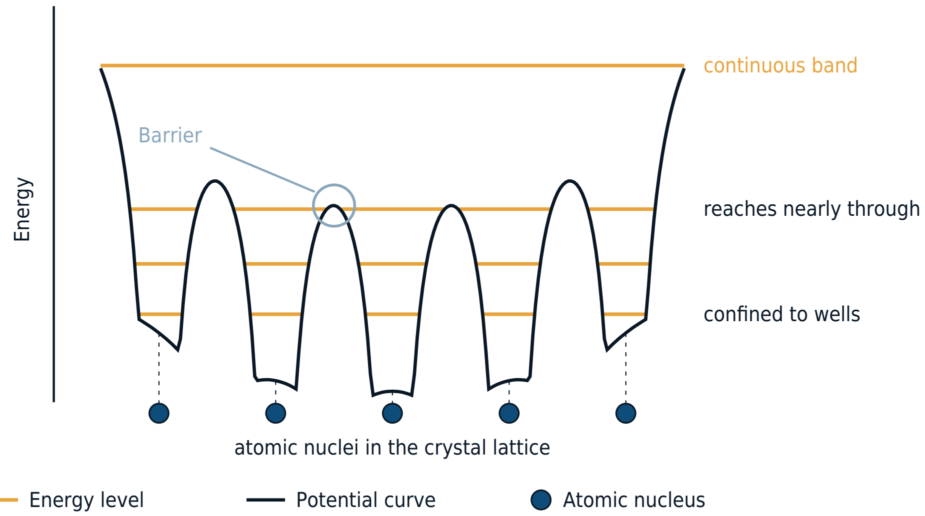


Each additional atom adds one more level. The overall width barely grows any further, so the spacing keeps shrinking – at 10^{22} atoms per cubic centimeter, the levels are no longer distinguishable.

The width of the energy bands depends on how strongly the electrons are bound to the atom. The valence electrons at the highest energy level interact strongly with those of neighbouring atoms and can be detached from the atom relatively easily; with a very large number of atoms, a single electron can no longer be assigned to a specific atom. As a result, the energy bands of the individual atoms merge into a continuous band, the valence band.

Energy Bands from Interacting Atoms

The potential wells of neighboring atoms overlap



Deeply bound electrons stay trapped in their well – their levels remain separate. The valence electrons, by contrast, sit above the barriers between the nuclei and no longer belong to a single atom: a band forms.

The Band Model in Conductors

In conductors, the valence band is either not completely filled with electrons, or the filled valence band overlaps with the empty conduction band. As a rule, both conditions apply simultaneously, so electrons can move either within the partially filled valence band or within the two overlapping bands. In conductors, there is no band gap between the valence band and the conduction band.

The Band Model in Insulators

In insulators, the valence band is completely filled with electrons due to the bonds between atoms. The electrons cannot move because they are "locked in" between the atoms. To achieve conductivity, electrons would have to move from the fully occupied valence band into the conduction band. This is prevented by the band gap that lies between the valence band and the conduction band.

This gap can only be overcome with a very large amount of energy (if at all) – according to the laws of quantum physics, no electron is allowed to reside within the band gap itself.

The Band Model in Semiconductors

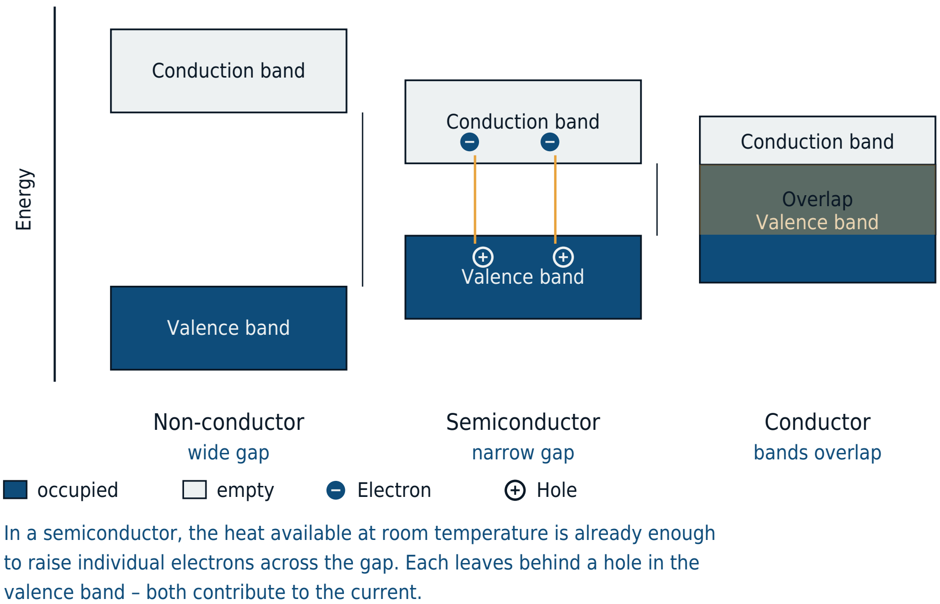
Semiconductors also have this band gap, but compared to insulators it is so small that, even at room temperature, electrons pass from the valence band into the conduction band. These electrons can then move freely and act as charge carriers. In addition, each electron leaves behind a hole in the valence band, which can be filled by other electrons in the valence band. This results in migrating holes in the valence band, which can be regarded as positive charge carriers.

Electron–hole pairs always occur together, meaning there are just as many negative as positive charges, and the semiconductor crystal as a whole remains neutral. A pure, undoped semiconductor is referred to as an intrinsic semiconductor; there are approximately 10^{10} free electrons and holes per cubic centimetre (at room temperature).

Since electrons always adopt the energetically most favourable state, they fall back into the valence band without any energy supply and recombine with holes. At a given temperature, an equilibrium is established between electrons being raised into the conduction band and electrons falling back. As temperature increases, the number of electrons able to cross the band gap increases as well. Thus, the conductivity of semiconductors increases with rising temperature.

The Band Model

The width of the band gap determines conductivity



In a semiconductor, the heat available at room temperature is already enough to raise individual electrons across the gap. Each leaves behind a hole in the valence band – both contribute to the current.

Since the width of the band gap corresponds to a specific energy and therefore a specific wavelength, attempts are made to specifically alter the band gap in order to obtain certain colours in light-emitting diodes. This can be achieved, among other methods, by combining different materials. Gallium arsenide (GaAs) has a band gap of 1.4 electron volts (eV, at room temperature) and therefore emits red light.

The intrinsic conductivity of silicon is of little interest for the functioning of devices, since it depends very strongly on the amount of energy supplied. It therefore also changes with operating temperature, and a conductivity comparable to that of metals is only reached at very high temperatures (several hundred degrees Celsius). In order to specifically influence the conductivity of semiconductors, foreign atoms can be incorporated into the regular silicon lattice, thereby adjusting the concentration of charge carriers – electrons and holes. This process is called *doping*.

Doping: n- and p-semiconductors

Doping

Doping means introducing foreign atoms into a semiconductor crystal in order to specifically alter its conductivity. Two of the most important substances used to dope silicon are boron (3 valence electrons = trivalent) and phosphorus (5 valence electrons = pentavalent). Others include aluminium and indium (trivalent), as well as arsenic and antimony (pentavalent).

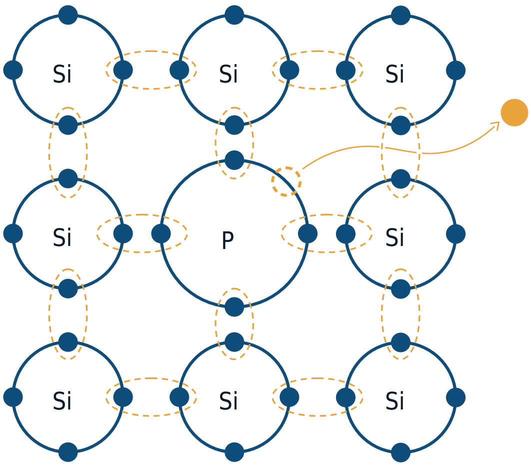
The dopant elements are incorporated into the lattice structure of the semiconductor crystal; the number of outer electrons determines the type of doping. Elements with 3 valence electrons are used for p-type doping, while pentavalent elements are used for n-type doping. The conductivity of a deliberately impurified silicon crystal can be increased in this way by a factor of 10^6.

n-Type Doping

The pentavalent dopant element has one more outer electron than the silicon atoms. Four of its outer electrons can bond with a silicon atom each, while the fifth remains freely mobile and serves as a charge carrier. This unbound electron requires far less energy to be raised from the valence band into the conduction band than the electrons responsible for the intrinsic conductivity of silicon. The dopant element that releases an electron is referred to as an electron donor (from Latin donare = to give).

By releasing negative charge carriers, the dopant elements become positively charged and remain fixed in the lattice; only the electrons are able to move. Doped semiconductors whose conductivity is based on free (negative) electrons are called n-conducting or n-doped. While holes (along with an equal number of electrons) can be spontaneously generated within the crystal at any time, the number of free electrons introduced by the donors now predominates, which is why these are referred to as majority charge carriers. Holes, on the other hand, are called minority charge carriers.

n-type doping with phosphorus



Four bonding electrons, one electron as a free charge carrier

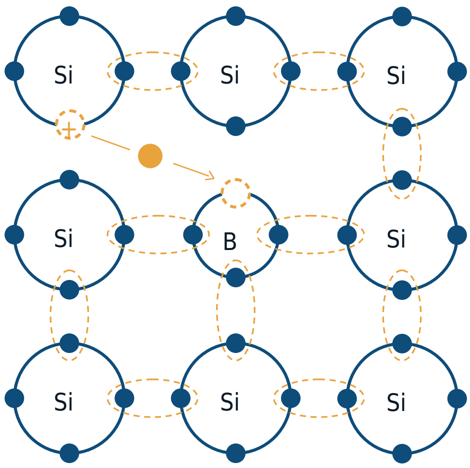
Arsenic is used as an alternative to phosphorus, since its diffusion coefficient is lower. This means that the dopant diffuses less during subsequent process steps, so the doping remains in the location where it was originally introduced.

p-Type Doping

In contrast to the free electron produced by doping with phosphorus, trivalent dopant elements have exactly the opposite effect. They can accept an additional outer electron, thereby leaving a hole behind in the valence band of the silicon atoms. As a result, the electrons in the valence band become mobile. The holes move in the direction opposite to the movement of the electrons. When indium is used as the dopant element, the energy required for this is only 1 % of the energy needed to raise the valence electrons of silicon atoms into the conduction band.

By accepting an electron, the dopant element becomes singly negatively charged; such dopant atoms are called electron acceptors (from Latin acceptare = to accept). Here too, the dopant element remains fixed within the crystal lattice; only the positive charge moves. These semiconductors are referred to as p-conducting or p-doped because their conductivity is based on positive holes. Analogous to n-doped semiconductors, holes are the majority charge carriers here, while free electrons are the minority charge carriers.

p-type doping with boron



Four bonding electrons, one electron as a free charge carrier

Externally, doped semiconductors are electrically neutral. The terms n-type and p-type doping refer only to the majority charge carriers.

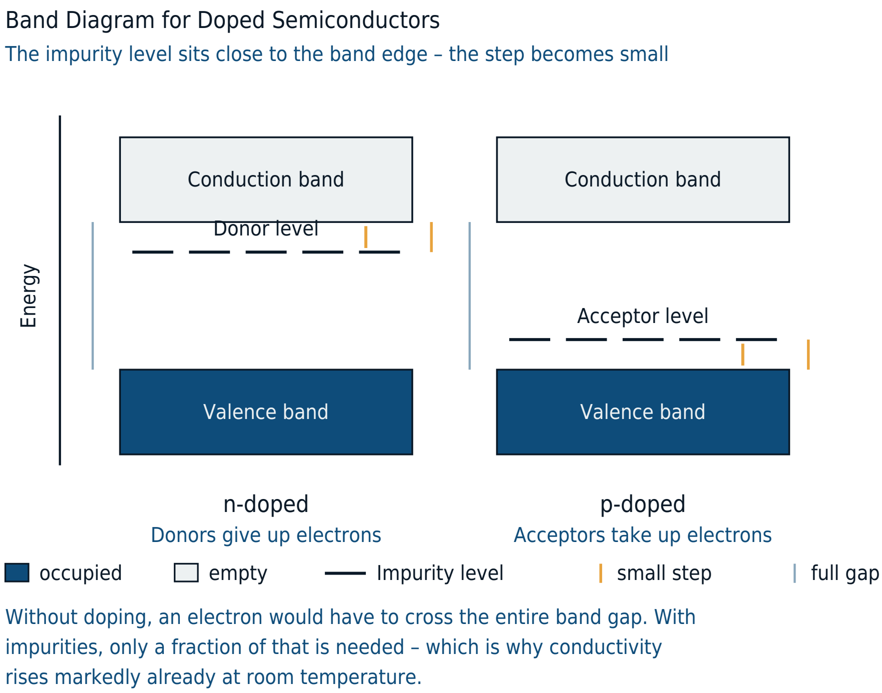
N-doped and p-doped semiconductors behave in an approximately similar way with respect to current flow. As the number of dopant elements increases, so does the number of charge carriers in the semiconductor crystal. Even a very small amount of dopant elements is sufficient for this. Weakly doped silicon crystals contain only 1 foreign atom per 1,000,000,000 silicon atoms, whereas at the highest doping levels the ratio of foreign atoms to silicon atoms is, for example, 1 to 1,000.

Band Diagram of Doped Semiconductors

In n-doped semiconductors, introducing a dopant element with five outer electrons makes an unbound electron available within the crystal, which can therefore be raised into the conduction band with comparatively little energy. As a result, n-doped semiconductors exhibit a donor level close to the conduction band edge, and the band gap that needs to be overcome is very small.

By introducing a trivalent dopant element, p-doped semiconductors provide a vacant site that can already be filled by an electron from the valence band with only a small amount of energy. In p-doped semiconductors, an acceptor level is therefore located close to the valence band edge.

Band diagram of doped semiconductors



The p-n junction

The p-n Junction in Equilibrium

The p-n junction is the transition region where n-doped and p-doped semiconductor crystals meet. In this region there are no free charge carriers, because the free electrons of the n-conductor and the free holes of the p-doped crystal recombine with one another near the contact surface of the two crystals, i.e. the electrons fill the vacant holes. This movement of charge carriers (diffusion) occurs as a result of a concentration gradient: since there are only a few electrons in the p-crystal and only a few holes in the n-crystal, the majority charge carriers (electrons in the n-region, holes in the p-region) migrate into the oppositely doped semiconductor crystal. The crystal lattice at the interface must not be interrupted; simply "pressing together" a p-doped and an n-doped silicon crystal does not result in a functional p-n junction.

Due to the migrated free charge carriers, the regions near the interface become positively charged (n-crystal) or negatively charged (p-crystal). The more charge carriers recombine, the larger this depletion or space-charge region (SCR) becomes, and with it the voltage difference between the n-crystal and the p-crystal. At a certain level of this potential difference, the recombination of holes and electrons comes to a stop, since the charge carriers can no longer overcome the electric field. In silicon, this limit is approximately 0.7 V (see band model of a p-n junction).

p-n junction with no applied voltage

Ihr Browser unterstützt keine eingebetteten SVG-Animationen.

A p-n junction corresponds to an electrical component that, when a voltage is applied, conducts current in one direction (forward direction) and blocks it in the other direction (reverse direction): a diode.

The p-n Junction with an Applied Voltage

If a positive voltage is applied to the n-crystal and a negative voltage to the p-crystal, the internal electric field and the field generated by the voltage source point in the same direction. The field at the p-n junction is thereby reinforced. The oppositely charged free charge carriers are attracted by the poles of the voltage source, which enlarges the depletion layer and prevents any current from flowing.

If the applied voltage across the semiconductor crystals is reversed, the electric field generated by the voltage source opposes the internal field and weakens it. Once the internal field has been completely cancelled out by the external field, new charge carriers continuously flow from the voltage source to the depletion layer and are able to recombine there without interruption: current flows.

p-n junction with an applied voltage

Ihr Browser unterstützt keine eingebetteten SVG-Animationen.

Because of this behaviour, the diode can be used as a rectifier: to convert alternating current into direct current. Regions where p-doped and n-doped semiconductor crystals are in contact with one another occur in many electronic components within semiconductor technology.

Transistors and Memory

MOSFET

A transistor is an electronic semiconductor device used to switch or amplify current. In a field-effect transistor (FET), a gate electrode controls the current flow between the source and drain terminals without being part of the current path itself. The most common design is the MOSFET: with no voltage applied to the gate, the p-doped substrate separates source and drain like a blocked p-n junction; the transistor is off. Only a positive gate voltage attracts electrons to the surface and forms a conducting n-channel there – the same mechanism already described under doping, here made electrically controllable. How the MOSFET is actually fabricated is covered in the chapter Construction of a n-channel FET.

Bipolar Transistor

The second fundamental transistor type is the bipolar transistor. It consists of two p-n junctions connected back to back, with the layer sequence n-p-n or p-n-p; its terminals are called emitter, base and collector. Unlike the MOSFET, both types of charge carriers – electrons and holes – contribute to the current flow here, hence "bipolar". A small voltage at the thin base layer opens the junction to the emitter and allows a much larger current to flow from emitter to collector. Bipolar transistors are faster than MOSFETs but require more area. Details are covered in the chapter Construction of an NPN bipolar transistor.

FinFET

As feature sizes shrink, a flat, planar gate eventually can no longer reliably shut off the channel. In a FinFET, the channel is instead built as a thin, upright silicon "fin" that the gate wraps around on three sides – source, drain and the basic switching function remain identical to the MOSFET. This multi-sided access reduces leakage currents and allows the threshold voltage to be defined more sharply. How a FinFET is built and fabricated in detail is described in the chapter Construction of a FinFET.

DRAM Cell

Transistors are used not only for switching but also for storing data. The DRAM cell combines a single transistor with a capacitor for this purpose (1T1C cell): the capacitor stores one bit as an electric charge, and the transistor connects it to the bitline via the wordline when reading or writing. Because the charge slowly leaks away, it must be refreshed every few milliseconds – hence "dynamic". How the capacitor and access transistor are fabricated is shown in the chapter DRAM Fabrication: The Trench Capacitor Process.