

Questions

- (a) How are semiconductors doped?
- (b) Why are the source and drain regions of a MOSFET heavily doped?
- (c) Why is the emitter of a bipolar transistor heavily doped?
- (d) Why is the collector of a bipolar transistor more lightly doped than the base?

Answers

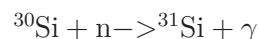
(a) How are semiconductors doped?

Doping of semiconductors means the modulation of the electrical properties by purposely adding impurities, called dopants, to the intrinsic semiconductor. These impurities can either be donors and provide electrons (n-doping) or acceptors and provide holes (p-doping) to the semiconductor.

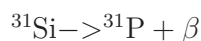
One has to differ between the initial doping of raw semiconductor material and the doping during wafer processing. The methods illustrated in this section are all related to the production of silicon-based semiconductor devices. The processes for the doping of other semiconductors (especially compound semiconductors) might differ somewhat from the processes used for silicon manufacturing.

There are two major ways for the production of raw silicon, the Czochralski- and the float zone method (see figure 1)

The adjustment of the doping level takes place during the Czochralski single crystal production by adding the dopant directly to the crystal melt. However dopants are removed during zone refining, so FZ-Si is doped using neutron transmutation doping. In this technique, the silicon ingots are bombarded with neutrons. Silicon appears as the isotopes ^{28}Si , ^{29}Si and ^{30}Si in an abundance of 92.23 %, 4.67 % and 3.1 %. The neutrons only affect ^{30}Si and transmute it in the reactions



under the release of a γ -quantum and



under the release of an electron (β -quantum) to ^{31}P . The half life of this decay is 2.63 h. The process of neutron transmutation doping takes several weeks but the doping distribu-

tion that is received is very homogeneous.

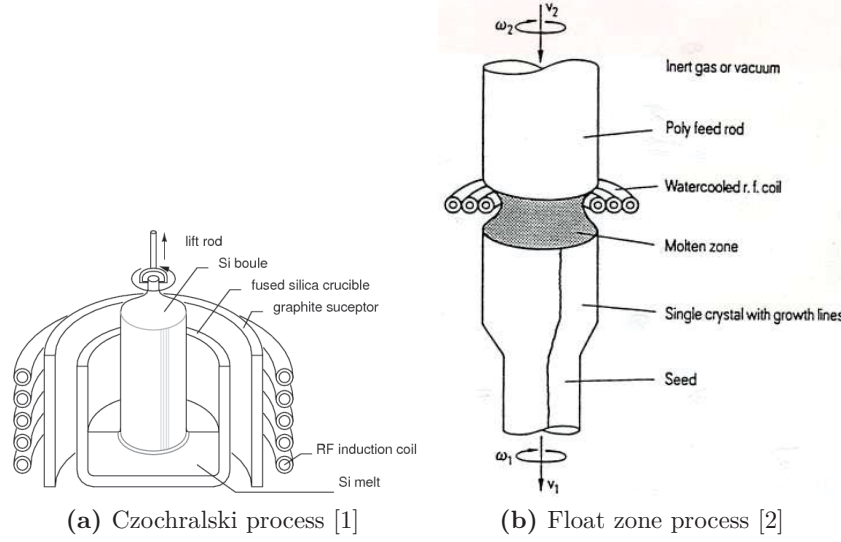


Figure 1: Major methods of raw silicon production

A semiconductor device usually consists of several different doping regions which are produced during the wafer processing. For the production of doped layers close to the surface ion implantation is used (see figure 2). In this method the dopant atoms are ionised and accelerated in an electric field. The focused ion beam is then scanned over the wafer. Due to scattering at the lattice atoms and deceleration in the electron shells of the lattice atoms, the implanted ions get stuck in a certain depth. The implantation depth is controlled by the acceleration energy of the implanted ions and the doping concentration is regulated by the implantation dose. In a subsequent annealing step the radiation damage caused by the implantation is cured. Furthermore the dopant atoms are electrically activated.

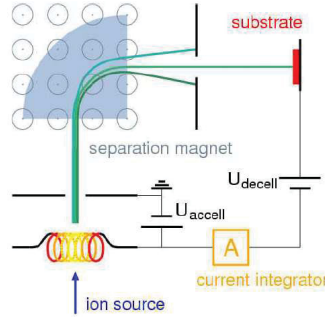


Figure 2: Scheme of an ion implanter [3]

Table 1 shows different dopants for different semiconductor materials.

Table 1: Dopants used for different semiconductor materials. E_c-E_D is the energy difference between the conduction band energy and the dopant energy. E_A-E_v is the difference between the acceptor energy and the valence band energy. [4]

Semiconductor	Donor	E_c-E_D [meV]	Acceptor	E_A-E_v [meV]
Silicon	Li	33	B	45
	Sb	39	Al	67
	P	45	Ga	72
	As	54	In	160
Germanium	Li	9.3	B	10
	Sb	9.6	Al	10
	P	12	Ga	11
	As	13	In	11
Gallium-Arsenide	Si	5.8	C	26
	Ge	6.0	Be	28
	S	6.0	Mg	28
	Sn	6.0	Si	35

(b) Why are the source and drain regions of a MOSFET heavily doped?

If a metal and a semiconductor are brought into contact electrons will flow from a high work-function material to a low work-function material. In thermal equilibrium, this charge transport causes an electric field, which pushes the electrons back into the low work-function material. In the band diagram this can be observed as bending of the valence and conduction bands (see figure 3).

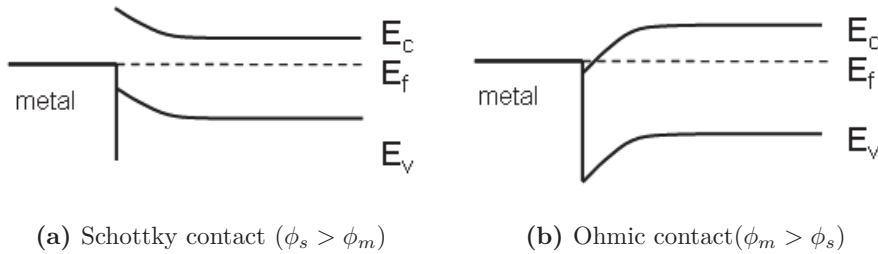


Figure 3: Band diagram of the interface of a metal with a work function ϕ_m and an n-doped semiconductor with a workfunction ϕ_m

Figure 3a shows the band diagram of a contact between a small work function metal and an n-doped semiconductor. The difference in work function causes the bands to bend up and hence, a Schottky diode is formed. If the workfunction of the metal is higher than the work function of the n-doped semiconductor, as shown in figure 3b, an ohmic contact is formed. Unfortunately, this dependence on $\Delta\phi$ is only valid for perfect metal-semiconductor interfaces.

Real interfaces between metals and semiconductors, though, are never perfect. In fact there is a large number of defects at the interface which have energy states close to the middle of the bandgap. These interface states pull the Fermi energy about into the middle of the bandgap and a Schottky contact is formed, no matter what the difference of the workfunctions is.

Hence, to form a good contact between a semiconductor and a metal, the region of the semiconductor close to the metal is degenerately doped. Figure 4 shows the band diagram at a contact between a degenerately doped semiconductor and a metal.

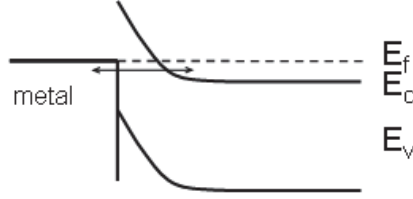


Figure 4: Band diagram of the interface of a metal and a degenerately doped semiconductor. The arrow indicates where the electrons can tunnel between the materials.

Also in a degenerately doped semiconductor, the interface states pull the Fermi energy to the middle of the bandgap causing the bands to bend up. As the depletion width is related to the doping concentration as shown in equation ??, it is very small for high doping concentrations. This enables tunneling of charge carries between the metal and the semiconductor. The tunneling current drops exponentially when the depletion width grows. Nevertheless, a constant width can be seen as a linear resistor and hence, a contact between a metal and a degenerately doped semiconductor shows a linear resistance.

The source/drain regions of a MOSFET (see figure 5 are, as a consequence, heavily doped to provide a good contact between the source/drain region on the semiconductor and the source/drain metallization (black areas in the image) and to avoid unwanted Schottky junctions.

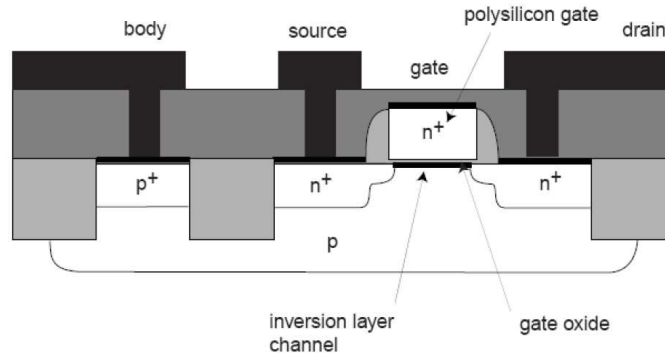


Figure 5: Scheme of an n-channel MOSFET device [5]

(c) Why is the emitter of a bipolar transistor heavily doped?

If a bipolar transistor (see figure 6) is forward biased, minority carriers from the emitter are injected across the forward biased pn-junction into the base. As the width of the base is far smaller than the diffusion length of the charge carriers, they can diffuse across the base to the reverse biased base-collector junction, where they are swept into the collector due to the electric field at this junction. Figure 7 shows the minority carrier concentrations in a forward biased bipolar (npn) transistor.

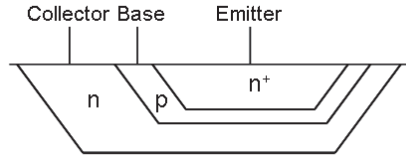


Figure 6: Scheme of an npn-bipolar transistor

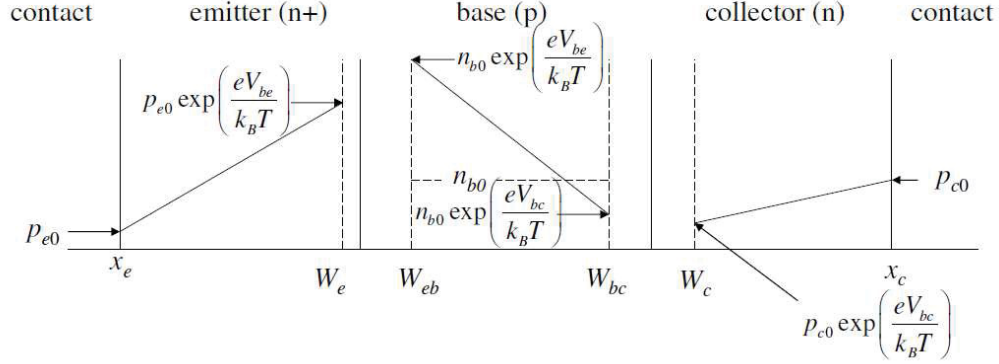


Figure 7: Minority carrier concentration in a forward biased npn bipolar transistor [6].

The emitter current can be calculated as the sum of the hole diffusion current I_{ep} in the emitter and the electron diffusion current I_{en} in the base.

The hole diffusion current in the emitter is

$$I_{ep} = eA_{eb}D_p \frac{p_{e0} \left(e^{eV_{be}/k_B T} - 1 \right)}{W_e - x_e} \quad (1)$$

where A_{eb} is the area of the emitter-base interface, D_p is the diffusion constant for holes, p_{e0} is the concentration of holes in the emitter at thermal equilibrium, V_{be} is the bias voltage between the emitter and the base and $W_e - x_e$ is the width of the neutral emitter.

The electron diffusion current in the base is

$$I_{en} = -eA_{eb}D_n \frac{n_{b0} \left(e^{eV_{be}/k_B T} - e^{eV_{bc}/k_B T} \right)}{W_{bc} - W_{eb}} \quad (2)$$

where D_n is the diffusion constant for electrons, n_{b0} is the concentration of electrons in the base at thermal equilibrium, V_{bc} is the bias voltage between the base and the collector and $W_{bc} - W_{eb}$ is the width of the neutral base.

The emitter efficiency γ_e is the ratio between the electron current in the base and the total emitter current.

$$\gamma_e = \frac{I_{en}}{I_{en} + I_{ep}} \quad (3)$$

To get a high emitter efficiency I_{en} should be much higher than I_{ep} . Hence p_{e0} has to be very small compared to n_{b0} . As

$$p_{e0} = n_i^2 / N_D, \quad (4)$$

the number of donors in the emitter and, hence, the emitter doping has to be high.

(d) Why is the collector of a bipolar transistor more lightly doped than the base?

The width of the neutral base in a bipolar transistor decreases as the reverse bias between the base and the collector is increased. If this width goes to zero, all gain of the transistor would be lost. If W is the total width of the depletion region between the base and the collector, the depletion width in the base W_b can be calculated as

$$W_b = \frac{N_c W}{N_c + N_b}, \quad (5)$$

where N_c is the doping concentration in the collector and N_b is the doping concentration in the base. If N_b is higher than N_c , A higher reverse bias can be applied to the base-collector junction without losing gain.

References

- [1] URL: http://www.allaboutcircuits.com/vol_3/chpt_2/12.html.
- [2] URL: http://people.seas.harvard.edu/~jones/es154/lectures/lecture_2/materials/materials.html.
- [3] URL: <http://lamp.tu-graz.ac.at/~hadley/psd/lectures12/oct16/slide30.html>.
- [4] URL: <http://lamp.tu-graz.ac.at/~hadley/psd/lectures12/oct23/slide4.html>.
- [5] URL: <http://lamp.tu-graz.ac.at/~hadley/psd/lectures12/oct23/slide14.html>.
- [6] URL: <http://lamp.tu-graz.ac.at/~hadley/psd/lectures12/dec18/slide16.html>.

.