Semiconductor Technology

from A to Z
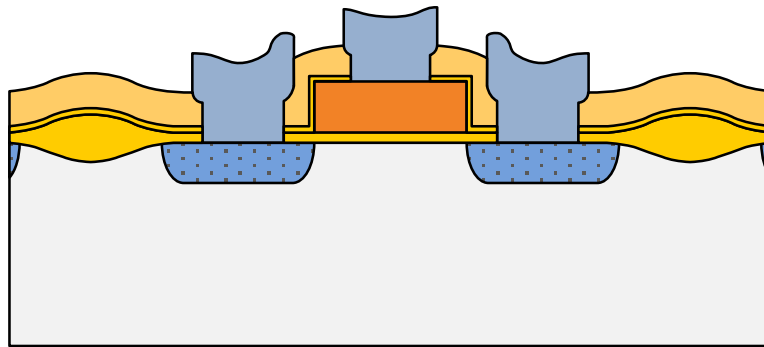
www.halbleiter.org

Contents

List of Figures

List of Tables

1 Fundamentals

1.1 The atomic structure

1.1.1 The atomic model

An atom is the smallest chemically not further divisible component of matter. Depending on the atom, it is composed of a certain number of electrons, protons, and neutrons. The positively charged protons and the neutrons form the nucleus, which is encircled by the electrons in certain intervals. A naturally occurring atom is electrically neutral, there are just as many positive protons as negatively charged electrons inside an atom. While the number of neutrons can vary. The simplest atom is the hydrogen atom, with only one electron, one proton, and no neutron. The next heavier atom, the noble gas helium, consists of two electrons, two protons, and two neutrons.

According to the Bohr model of the atom the electrons are assigned to so-called shells, which represent different energy levels and therefore are arranged concentrically to the nucleus. There is a maximum of seven shells, which can hold a different number of electrons, the electrons assignet to the outermost shell are known as valence electrons.
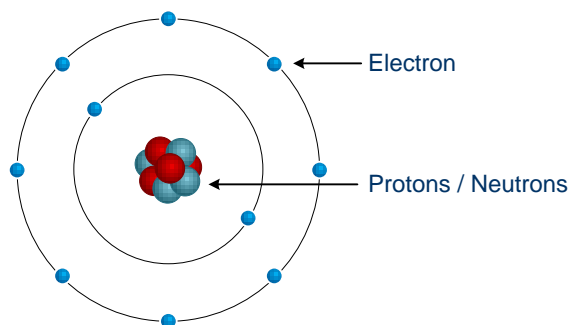
Fig. 1.1: Simplified illustration of a neon atom

The efforts of all the atoms is to fulfill their outermost shell, with eight electrons they reach the so-called noble gas configuration (also electron octet). Elements with few

outer electrons can donate electrons, elements with many outer electrons can accept additional valence electrons (see chapter chemical bonds for details).

1.1.2 Properties of atoms

Mass: The mass of an atom is determined mainly from the nucleus, since the masses of protons and neutrons ($1.67 \cdot 10^{-27}$ kg) are about 1800 times larger than the mass of the electrons in the atomic shell ($9.11 \cdot 10^{-31}$ kg).

Dimensions: The diameter of the atomic shell is 0.1 to 0.5 nm, the diameter of the nucleus is even lower by a factor of 100,000. To illustrate: when a pinhead in the middle of a soccer field represents the atomic nucleus, the distance to the corner flags corresponds to the distance at which the electrons orbiting the nucleus.

Density: In the nucleus of an atom protons and neutrons are packed extremely densely. If one were to compress the earth to the same density, its radius would be reduce from 6,700,000 m initially to only 100 meters.
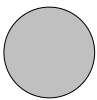
Particle	Proton	Neutron	Electron
Charge	+1	0	-1
Mass	$1.672 \cdot 10^{-27}$ kg	$1.675 \cdot 10^{-27}$ kg	$0.0009 \cdot 10^{-27}$ kg

Fig. 1.2: Elementary particles

Boron: 5p 6n 5e   **Silicon:** 14p 14n 14e   **Phosphorus:** 15p 15n 15e

3 valence electrons   4 valence electrons   5 valence electrons
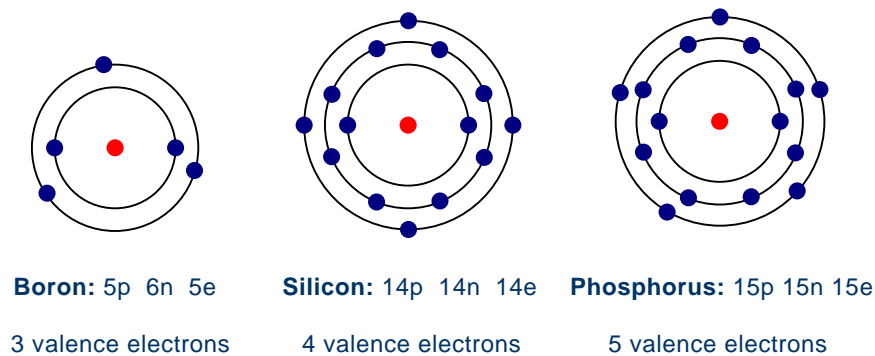
Fig. 1.3: Important elements in semiconductor industry

1.2 The elements, the periodic table

1.2.1 Elements

An element consists of several identical atoms and is a substance that can not be further decomposed by chemical means. The mass of elements is determined only by the number of protons and neutrons, since the electron mass is negligible. Hydrogen with one proton and no neutron has the mass number 1, the next heavier element, helium, has the mass number 4 (2 protons + 2 neutrons).

The elements are usually named with the initials of their Latin or Greek names (from latin hydrogenium: hydrogen H, from greek lithos: lithium Li).

1.2.2 The periodic table of the chemical elements

The periodic table of the chemical elements (periodic table) lists all the chemical elements with increasing proton number (atomic number) and according to their chemical properties, divided into periods as well as main and subgroups.

The period represents the number of electron shells, the main group, the number of electrons in the outermost shell (1 to 8 electrons). Group 1 and 2 and 13-18 are the maingroups, the group of 3-12 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 and is therefore in period 1 as well. Since the first shell is completely filled with only two electrons, helium is not in group 2, but in group 18 (group of the noble gases).

To add more electrons, one needs to begin a new electron shell. Thus we find lithium Li in group 1, period 2 (two electrons on the first shell, one valence electron on the second shell). A shell can hold a maximum of $2n^2$ electrons, where n stands for the period.

From the fourth period there exist the groups (3-12) of the transition elements. After the first two valence electrons were added in maingroup 1 and 2 to the outermost shell, more inner shells are filled in the transitional groups for energetic reasons, before the outer shell is completely filled with electrons in maingroup 13-18.

Elements that occur on the left side in the periodic table are metals. These elements have an aspiration to donate valence electrons to achieve the noble gas configuration.

Main gr.		Sub groups										Main groups					
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18

Period

1	H																	He
2	Li	Be											B	C	N	O	F	Ne
3	Na	Mg											Al	Si	P	S	Cl	Ar
4	K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
5	Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
6	Cs	Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Ti	Pb	Bi	Po	At	Rn
7	Fr	Ra	Ac	Rf	Db	Sg	Bh	Hs	Mt	Ds	Rg	Uub	Uut	Uuq	Uup	Uuh	Uus	Uuo

Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
Th	Ph	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr

Alkali metal	Metals
Alkaline earth metal	Semiconductors
Transition metals	Insulators

Halogens	Lanthanides
Noble gases	Actinides

Fig. 1.4: The periodic table of the elements

On the right side there are the nonmetals, which are trying to accept additional electrons to achieve the noble gas configuration. In between there are the semimetals such as silicon and germanium.

1.3 Chemical bonds

1.3.1 Chemical bonds

Electrons located on the outermost shell, can escape from their atoms by supplying energy (e.g. in the form of heat) and be exchanged with other atoms. Compounds of several atoms are called molecules. The reason for the bond effort is the so-called noble gas configuration - a fulfilled valence shell - which represents an energetically stable state. Substances which have reached a full outer shell, will not form bonds (a few exceptions such as xenon-fluorine compounds are possible).

Element	Particle	Characteristic, application
B Boron	5p, 6n, 5e	3 valence electrons: used for p-doping of silicon
N Nitrogen	7p, 7n, 7e	Stable N_2 molecule: inert gas, cover layer on top of the wafer
O Oxygen	8p, 8n, 8e	Very reactive: oxidation of silicon, insulating layers (SiO_2)
F Fluorine	9p, 10n, 10e	Most reactive element: used for etching in combination with other elements (ie HF, CF_4)
Si Silicon	14p, 14n, 14e	Bulk material in semiconductor industry
P Phosphorous	15p, 16n, 15e	5 valence electrons: used for n-doping of silicon

Tab. 1.1: Important elements in semiconductor industry

There are mainly three different types of bonds, which are discussed below.

1.3.2 The atomic bond

Nonmetals form this bond to achieve an octet of electrons. For example two fluorine atoms (each with seven outer electrons) can fill their octet of electrons by mutual exchange of one electron. The distance between the two nuclei represents a compromise between the attraction of the nuclei and the bonding electrons and the repulsion of the two nuclei and the electrons as well. The reason for the atomic bond is that nature is always striving to reach the lowest energy state. Since the electrons have more space through the bond of several atoms, which corresponds to a lower energy, the atomic bond is formed.

Since atoms try to achieve the fulfilled outer shell, elemental fluorine atoms never appear as a single atom, but always as a fluorine molecule F_2. Also nitrogen N_2, oxygen O_2, chlorine Cl_2, bromine Br_2, and iodine I_2. The atomic bond also is called covalent bond.

Illustration of the atomic bond of silane (SiH_4, for silicon only the outermost shell is drawn). The silicon atom reaches the full outermost shell, the hydrogen atoms reach the filled first shell with only two electrons.
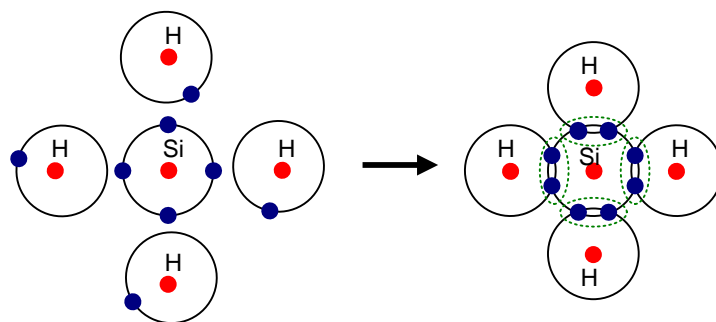
Fig. 1.5: The atomic bond of silane

1.3.3 The ionic bond

Ionic bonds are formed by the fusion of metals and nonmetals. While metals donate electrons to achieve a completely filled outer shell, nonmetals accept additional electrons. An example of an ionic bond is NaCl, sodium chloride.

The sodium atom gives off its valence electron (so it has more protons than electrons and is positively charged), while chlorine accepts one electron and thus is negatively charged. Due to the different charges the two atoms attract each other. A charged atom is known as an ion, while a positive ion is called cation and a negative ion anion, respectivley.

Since atoms always occur at a very large number, they form grids due to the attraction and repulsion forces from ions. Substances which form such a grid in solid state, are known as salts.

1.3.4 The metallic bonding

Metals are forming that bond to attain the stable noble gas configuration. Each metal atom gives off its outer electrons: thus positively charged metal ions and free electrons are genereted with strong attractive forces in between. The metal ions repel each other as well as the electrons.

Since the attraction and repulsion forces act in all dimensions of space, the atoms arrange themselves in a regular lattice. The free electrons are the so-called Fermi gas, which holds the positive metal ions together. Due to the free electrons, metal's electric conductivity is excellent.

(a) Natrium acts as cation

(b) Chloride acts as anion

(c) NaCl ion grid
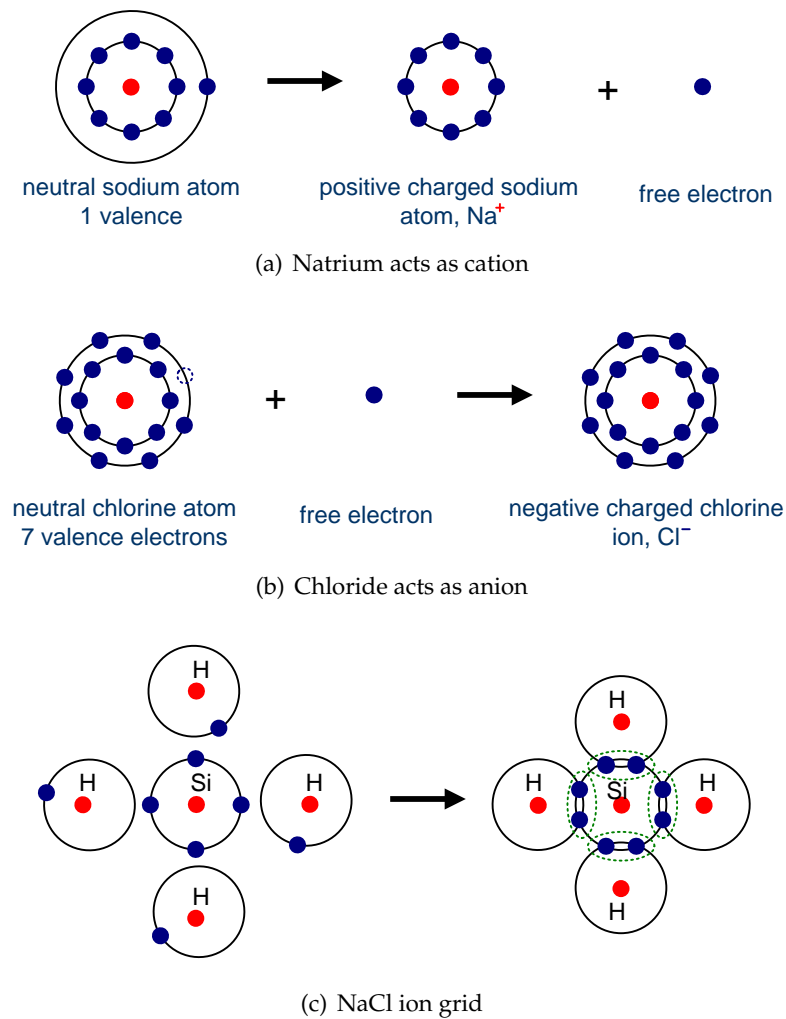
Fig. 1.6: Principle of the ion bonnding

The physical and chemical properties of compounds are dependent on the type of bonding. Thus greater attraction forces mean higher melting and boiling points, the number of free electrons affects the conductivity.

1.3.5 Intermolecular bondings

Intermolecular bondings differ from chemical bondings in that way that these are only forces between two ore more molucules, ions or atoms. Intermolecular bonds are caused mainly by load displacement which result in attracting or repulsing of particles.

Free electrons form an electron gas

Al^{3+} Al^{3+} Al^{3+} Al^{3+}

Al^{3+} Al^{3+} Al^{3+} Al^{3+}

Al^{3+} Al^{3+} Al^{3+} Al^{3+}

Fixed aluminum atoms
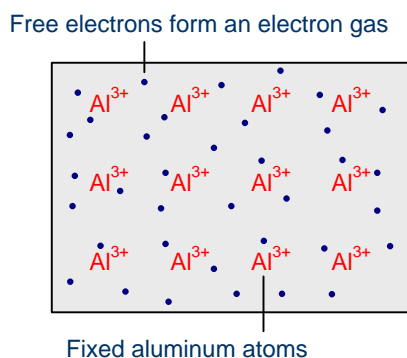
Fig. 1.7: Metallic bonding

The strongest of these forces is the hydrogen bond. Due to its bonding angle the loads inside a water molecule (H−O−H) are distributed asymmetrical, so that the oxygen atom has a negative partial charge while the hydrogen atoms have a positive partial charge. This load displacement leads to attracting and repulsing of other water molecules. Besides the hydrogen bond there are even weaker forces like the van der Waals interaction.

The bond energy of the chemical bonds is in the range of several hundreds to several thousands kilojoule/mol (kJ/mol). The bond energy of hydrogen bonds is up to a hundred kJ/mol and the bond energy of van der Waals forces is in the range of 0.5 to 5 kJ/mol.

1.4 Noble gases

Noble gases are the elements in the eighth maingroup of the periodic table: helium, neon, argon, krypton, xenon, radon (top down).

The specificity of these elements is that they have eight electrons in their outer shell. This octet represents a very stable energetic state, which all elements would like to achieve. In the chapter chemical bonds it is explain how the elements reach a full outer shell. Due to the stable state of the noble gas configuration, these elements will not go into reaction with other elements (a few xenon-fluorine compounds and others are known).

Neon consists of ten electrons: two on the first shell and eight valence electrons on the

second/outer most shell. Exception: helium, which consists of only one shell, reaches the noble gas configuration already with two electrons (electron duet instead of octet).

Neon: 8 valence electrons
on the 2nd shell

Helium: 2 valence electrons
on the 1st shell
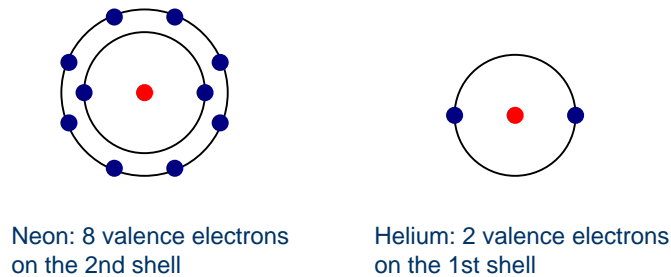
Fig. 1.8: Noble gases

As inert gases [lat. inert: inactive, neutral] oxygen (eg as purge gas), argon (sputtering), and others are used in the semiconductors industry.

1.5 Conductors - Insulators - Semiconductors

1.5.1 Conductors

Conductors are generally substances which have the property to pass different types of energy. In the following, the conductivity of electricity is the value of interest.

Metals:
The conductivity of metals is based on the free electrons (so-called Fermi gas) due to the metal bonding. Already with low energy electrons become sufficiently detached from the atoms and a conductivity is achieved.

The conductivity depends, inter alia, on the temperature. If the temperature rises, the metal atoms swing ever stronger, so that the electrons are constrained in their movements. Consequence, the resistance increases. The best conductors, gold and silver, are used relatively rare because of the high costs (gold e.g. for the contacting of the finished chips). The alternatives in the semiconductor technology for the wiring of the individual components of microchips are aluminum and copper.

Salts:
In addition to metals, salts can also conduct electricity. There are no free electrons, so the conductivity depends on ions which can be solved when a salt is melting or dissolving, so that the ions are free to move (see chapter chemical bonds for details).
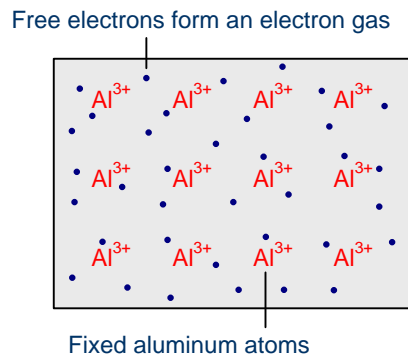
Fig. 1.9: Metallic bonding: fixed ions and free valence electrons (Fermi gas)

1.5.2 Insulators

Insulators possess no free charge carriers and thus are non-conductive.

The atomic bond:
The atomic bond is based on shared electron pairs of nonmetals. The elements which behave like nonmetals have the desire to catch electrons, thus there are no free electrons which might serve as charge carriers.

The ionic bond:
In the solid state, ions are arranged in a grid network. By electrical forces, the particles are held together. There are no free charge carriers to enable a current flow. Thus substances composed of ions can be both conductor and insulator.

1.5.3 Semiconductors

Semiconductors are solids whose conductivity lies between the conductivity of conductors and insulators. Due to exchange of electrons - to achieve the noble gas configuration - semiconductors arrange as lattice structure. Unlike metals, the conductivity increases with increasing temperature.

Increasing temperatures leads to broken bonds and free electrons are generated. At the location at which the electron was placed, a so-called defect electron ("hole") remains.

The electron flow is based on the conductivity properties of semiconductors. The electronic band structure illustrates why semiconductors behave like this.

An electron ● detaches from an atom, a positive charged hole ✚ is generated.

Another electron fills the hole. Thereby a new hole is generated.

This process repeats everywhere in the lattice. While electrons move to the right, holes move to the left.
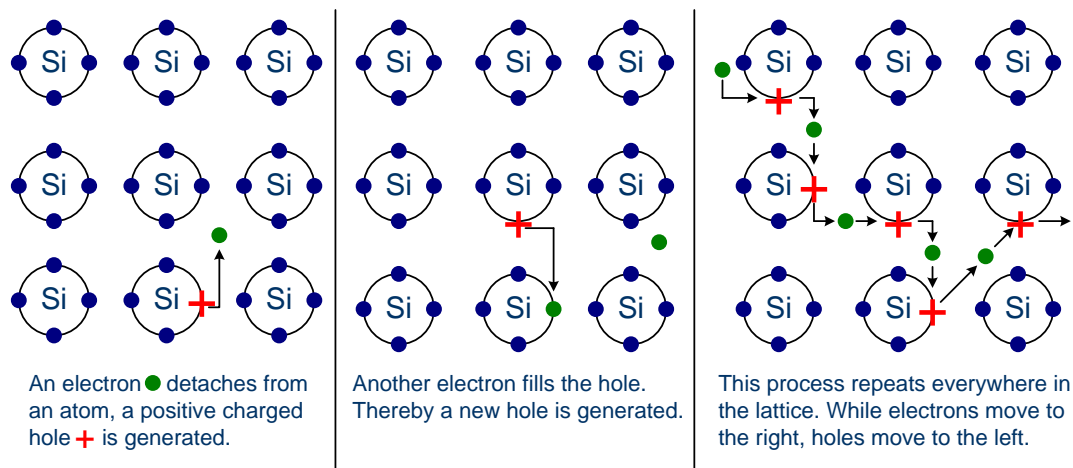
Fig. 1.10: Cut-out of a silicon lattice

1.5.4 The band model

The electronic band structure is an energy schema to describe the conductivity of conductors, insulators, and semiconductors. The schema consists of two energy bands (valence and 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 occupied with no electrons. Between the two energy bands there is the band gap, its width affects the conductivity of materials.

The energy bands

If we consider a single atom, there are according to the Bohr model of atoms sharply distinct energy levels, which may be occupied by electrons. If there are multiple atoms side by side they are interdependent, the discrete energy levels are fanned out. In a silicon crystal, there are approximately 10^{23} atoms per cubic centimeter, so that the individual energy levels are no longer distinguishable from each other and thus form broad energy ranges.

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

The band model of conductors
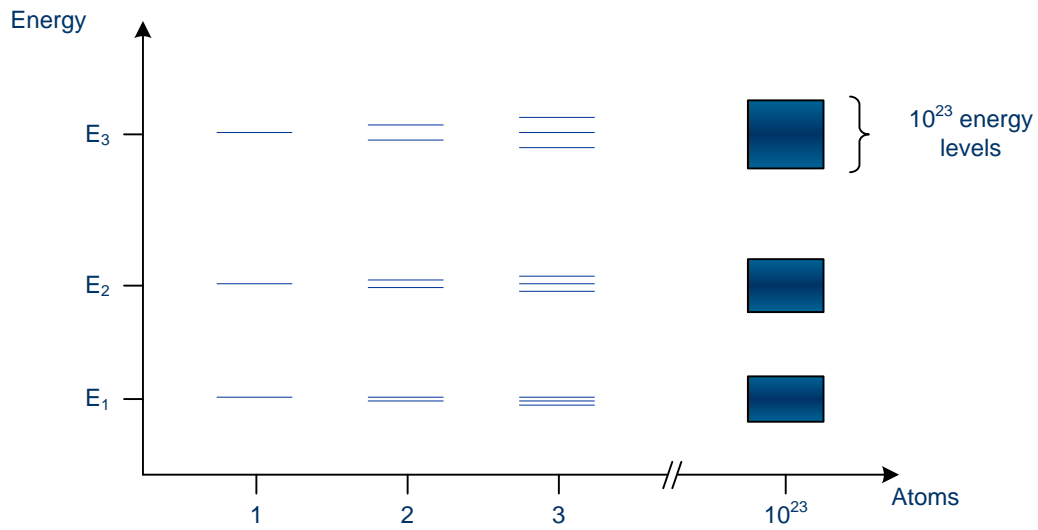
Fig. 1.11: Energy levels of atoms which are in interdependency with other atoms
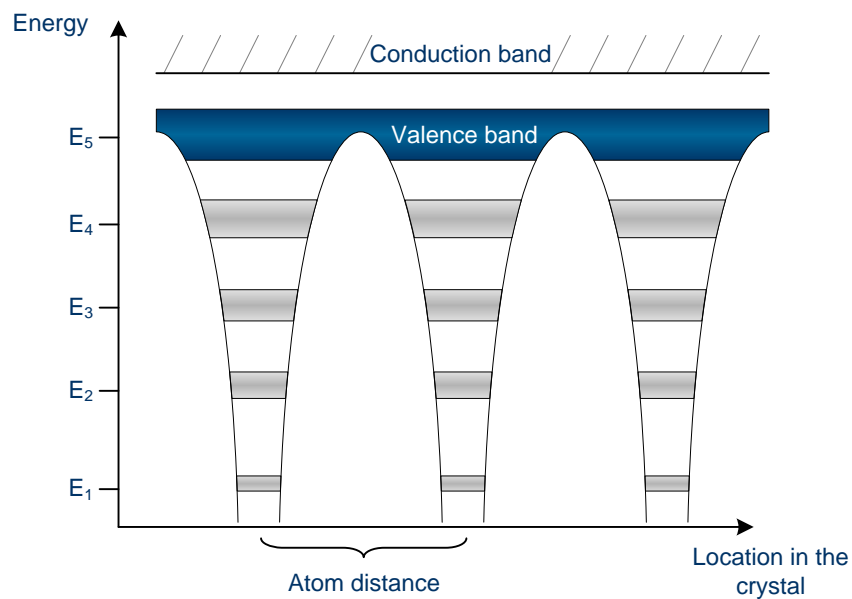
Fig. 1.12: Energy bands of atoms which are in interdependency with other atoms

In conductors, the valence band is either not fully occupied with electrons, or the filled valence band overlaps with the empty conduction band. In general, both states occure at the same time, the electrons can therefore move inside the partially filled valence band or inside the two overlapping bands. In conductors there is no band gap between the valence band and conduction band.

The band model of insulators

In insulators the valence band is fully occupied with electrons due to the covalent bonds. The electrons can not move because they are "locked up" between the atoms. To achieve a conductivity, electrons from the valence band have to move into the conduction band. This prevents the band gap, which lies in-between the valence band and conduction band.

Only with considerable energy expenditure (if at all possible) the band gap can be overcome; thus leading to a negligible conductivity.

The band model of semiconductors

Even in semiconductors, there is a band gap, but compared to insulators it is so small that even at room temperature electrons from the valence band can be lifted into the conduction band. The electrons can move freely and act as charge carriers. In addition, each electron also leaves a hole in the valence band behind, which can be filled by other electrons in the valence band. Thus one gets wandering holes in the valence band, which can be viewed as positive charge carriers.

There are always pairs of electrons and holes, so that there are as many negative as positive charges, the semiconductor crystal as a whole is neutral. A pure undoped semiconductor is known as intrinsic semiconductor. Per cubic centimeter there are about 10^{10} free electrons and holes (at room temperature).

Since the electrons always assume the energetically lowest state, they fall back into the valence band and recombine with the holes if there is no energy supply. At a certain temperature an equilibrium is arranged between the electrons elevated to the conduction band and the electrons falling back. With increasing temperature the number of electrons that can leap the band gap is increased, and thus increasing the conductivity of semiconductors.

Since the width of the band gap represents a certain energy corresponding to a particular wavelength, one tries to alter the width selective in order to obtain certain colors of light emitting diodes (LED). This may be achieved by combining different materials. Gallium arsenide (GaAs) has a band gap of 1.4 eV (electron volts, at room temperature) and thus emits red light.

The intrinsic conductivity of silicon is of no interest for the functioning of components, since it depends, inter alia, on the supplied energy. Which means that it changes with the temperature; in addition a conductivity comparable to metals is only possible at
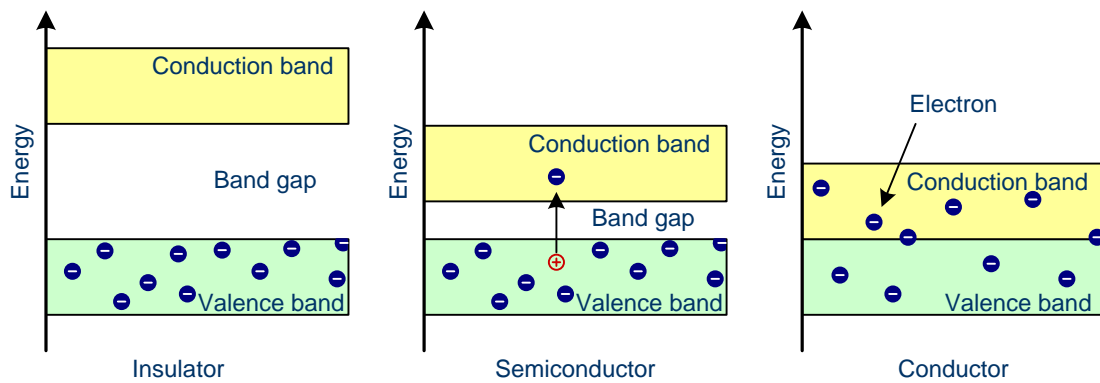
Fig. 1.13: The band model

very high temperatures (several hundred degrees Celsius). In order to deliberately influence the conductivity of semiconductors, impurity atoms can be introduced into the regular silicon lattice to alter the number of free electrons and holes.

1.6 Doping: n- and p-semiconductors

1.6.1 Doping

Doping means the introduction of impurities into a semiconductor crystal to the defined modification of conductivity. Two of the most important materials silicon can be doped with, are boron (3 valence electrons = 3-valent) and phosphorus (5 valence electrons = 5-valent). Other materials are aluminum, indium (3-valent) and arsenic, antimony (5-valent).

The dopant is integrated into the lattice structure of the semiconductor crystal, the number of outer electrons define the type of doping. Elements with 3 valence electrons are used for p-type doping, 5-valued elements for n-doping. The conductivity of a deliberately contaminated silicon crystal can be increased by a factor of 10^6.

1.6.2 n-doping

The 5-valent dopant has an outer electron more than the silicon atoms. Four outer electrons combine with ever one silicon atom, while the fifth electron is free to move and serves as charge carrier. This free electron requires much less energy to be lifted

from the valence band into the conduction band, than the electrons which cause the intrinsic conductivity of silicon. The dopant, which emits an electron, is known as an electron donor (donare, lat. = to give).

The dopants are positively charged by the loss of negative charge carriers and are built into the lattice, only the negative electrons can move. Doped semimetals whose conductivity is based on free (negative) electrons are n-type or n-doped. Due to the higher number of free electrons those are also named as majority charge carriers, while free mobile holes are named as the minority charge carriers.
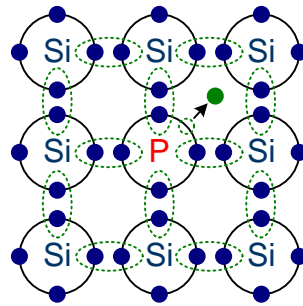
Fig. 1.14: n-doping with phosphorus

Arsenic is used as an alternative to phosphorus, because its diffusion coefficient is lower. This means that the dopant diffusion during subsequent processes is less than that of phosphorus and thus the arsenic remains at the position where it was introduced into the lattice originally.

1.6.3 p-doping

In contrast to the free electron due to doping with phosphorus, the 3-valent dopant effect is exactly the opposite. The 3-valent dopants can catch an additional outer electron, thus leaving a hole in the valence band of silicon atoms. Therefore the electrons in the valence band become mobile. The holes move in the opposite direction to the movement of the electrons. The necessary energy to lift an electron into the energy level of indium as a dopant, is only 1 % of the energy which is needed to raise a valence electron of silicon into the conduction band.

With the inclusion of an electron, the dopant is negatively charged, such dopants are called acceptors (acceptare, lat. = to add). Again, the dopant is fixed in the crystal lattice, only the positive charges can move. Due to positive holes these semiconductors

are called p-conductive or p-doped. Analog to n-doped semiconductors, the holes are the majority charge carriers, free electrons are the minority charge carriers.
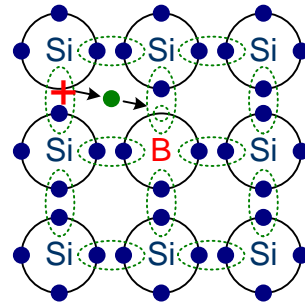
Fig. 1.15: p-doping with boron

Doped semiconductors are electrically neutral. The terms n- and p-type doped do only refer to the majority charge carriers. Each positive or negative charge carrier belongs to a fixed negative or positive charged dopant.

N- and p-doped semiconductors behave approximately equal in relation to the current flow. With increasing amount of dopants, the number of charge carriers increases in the semiconductor crystal. Here it requires only a very small amount of dopants. Weakly doped silicon crystals contain only 1 impurity per 1,000,000,000 silicon atoms, high doped semiconductors for example contain 1 foreign atom per 1,000 silicon atoms.

1.6.4 Electronic band structure in doped semiconductors

Through the introduction of a dopant with five outer electrons, in n-doped semiconductors there is an electron in the crystal which is not bound and therefore can be moved with relatively little energy into the conduction band. Thus in n-doped semiconductors one finds a donator energy level near the conduction band edge, the band gap to overcome is very small.

Analog, through introduction of a 3-valent dopant in a semiconductor, a hole is available, which may be already occupied at low-energy by an electron from the valence band of the silicon. For p-doped semiconductors one finds an acceptor energy level near the valence band.
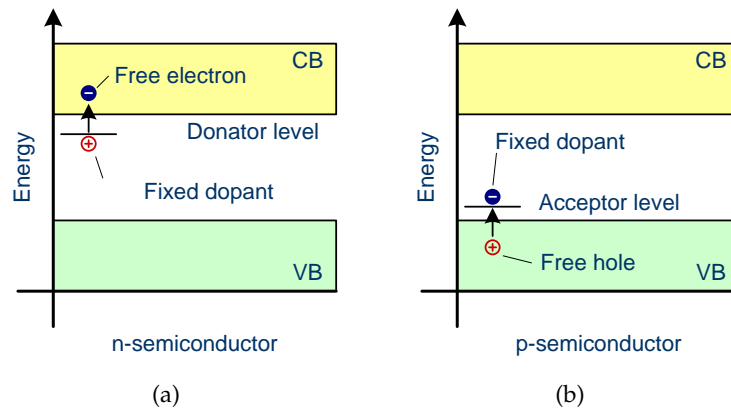
Fig. 1.16: Band model of p- and n-type doped semiconductors

1.7 The p-n junction

1.7.1 p-n junction at thermal equilibrium

The p-n junction is the transition area between two n- and p-doped semiconductor crystals. In this area there are no free charge carriers, since the free electrons of the n-conductor, and the holes of the p-doped crystal in the vicinity of the interface recombine with each other, which means that the electrons fill the holes. This charge movement (diffusion) is obtained in consequence of a concentration gradient: since there is only a few number of electrons in the p-area and only a few number of holes in the n-region, the majority charge carriers (electrons in the n-crystal, holes in the p-crystal) move into the contrary doped semiconductor. The crystal lattice at the interface must not be interrupted, a simple "pressing together" of a p-type and a n-doped silicon crystal does not allow a functional p-n junction.

The regions near the interface are loaded due to the loss of free charge carriers (positive charge in the n-crystal, negative charge in the p-crystal). The more charge carriers recombine, the greater the depletion zone and thus the voltage difference of n- and p-crystal. With a certain amount of this potential gap, the recombination of holes and electrons comes to a complete standstill, the charge carriers can no longer overcome the electric field. In silicon this limit is at about 0.7 V.

A p-n junction represents an electrical component with the function to allow an electric current in one direction (called the forward biased condition) and to block the current
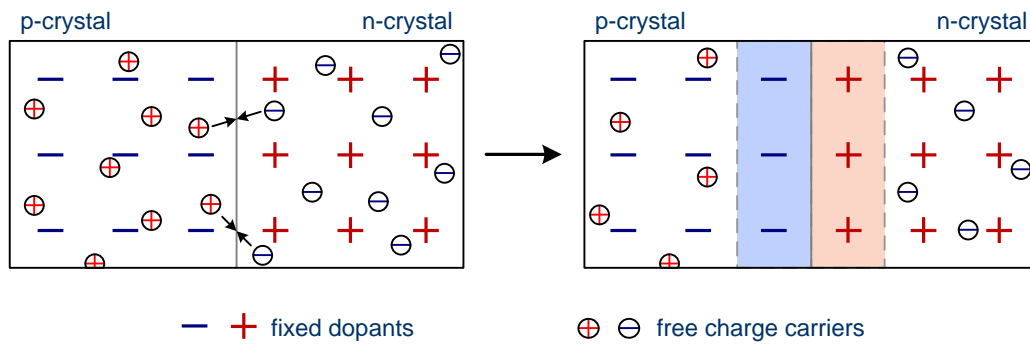
Fig. 1.17: p-n junction without an external applied voltage

in the opposite direction (the reverse biased condition): a diode.

1.7.2 p-n junction with external applied voltage

If the n-type crystal is applied to a positive and the p-crystal to a negative voltage, the electric field inside the semiconductor and the field of the voltage source are in the same direction. Thus the electric field at the p-n junction is reinforced. The oppositely charged free carriers are attracted by the poles of the voltage source, thus the barrier layer is increased and a current flow is inhibited.

If the external voltage is applied in the reverse direction, the external and internal electric field are in the opposite direction and the inner filed is weakened. If the inner field is completely eliminated from the outer field, a constantly flow of free charge carriers from the power source to the interface is possible and the carriers can recombine continuously: there is electric current.

The diode can be used as a rectifier: to convert alternating current into direct current. Areas where p- and n-doped semiconductor crystals are in contact, are found in many electrical devices in the semiconductor technology.

(a) reverse direction (b) forward direction
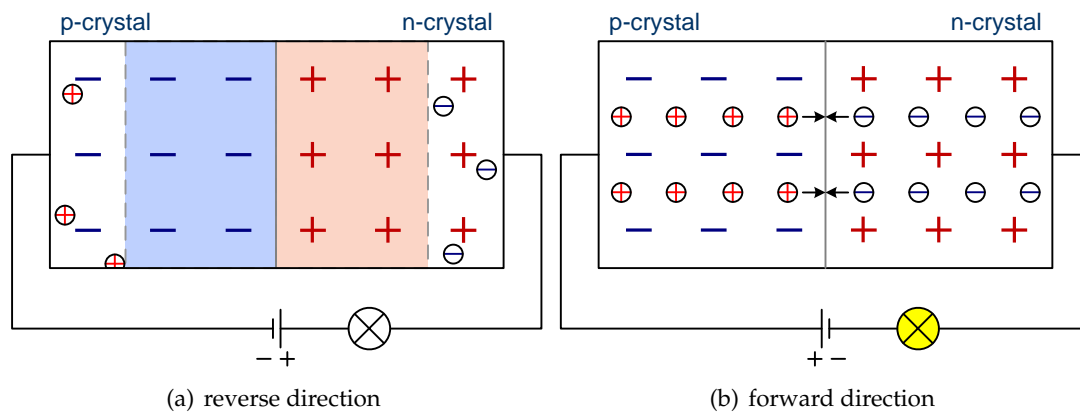
Fig. 1.18: p-n junction with an external applied voltage

1.8 Field-effect transistors

1.8.1 General layout

A transistor is an electronic semiconductor device for switching or amplifying electricity. The current can flow through two junctions - called drain and source -, while the third (gate electrode) is used for control. In addition to the field-effect transistor (FET) which is described here, there is another basic transistor, the bipolar transistor. The operation of the bipolar transistor is based on charge carriers of both polarities (thus bipolar), holes and electrons. Field-effect transistors, also known as unipolar transistors, use either electrons or holes for the transport of electricity.

The transistor is the basic component in semiconductor manufacturing, in modern microchips there are found several millions to billions of transistors. Through the combination of multiple transistors, all logic gates can be implemented in order to obtain logic output signals of corresponding input signals. Thus transistors form the heart of every microprocessor, memory chip, etc. The transistor is the most abundant object made by mankind, and thus became indispensable in today's life.

The transistor is built up layer by layer. This article describes the basic structure of a simple field-effect transistor, the various possibilities to realize the miscellaneous layers will follow in the later chapters.

1.8.2 Construction of a n-channel FET

1. Substrate: Basis for a n-channel field-effect transistor is a p-doped (boron) silicon substrate.

p-Si

2. Oxidation: On top of the substrate, a thin layer of silicon dioxide SiO_2 (the gate oxide) is created via thermal oxidation. It is used for insulation of the later deposited gate and the substrate.

p-Si
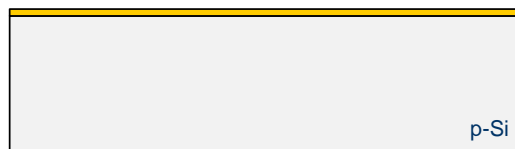
3. Deposition: In a LPCVD process nitride is deposited, it is used later as an masking during the field oxidation.

p-Si
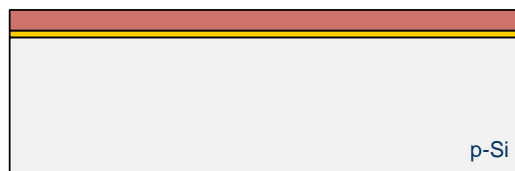
4. Photolithography: On top of the nitride a photoresist is spun on, exposed and developed. Thus a structured coating layer is fabricated which serves as an etching mask.

p-Si
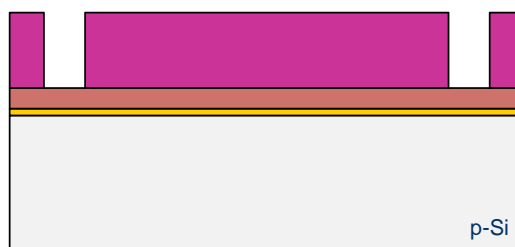
5. Etching: Only at resist free sites nitride is removed using reactive ion etching.
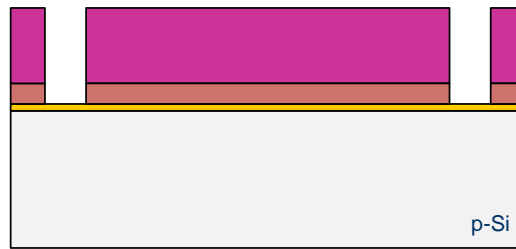
6. Resist removal: Subsequent the resist mask is removed in a wet-chemical developer solution.
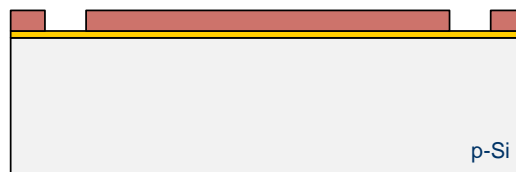
7. Oxidation: During field oxidation, the nitride serves as a mask layer, the thermal wet oxidation takes place only on the bare gate oxide. The grown field oxide is used for lateral isolation to adjacent devices.
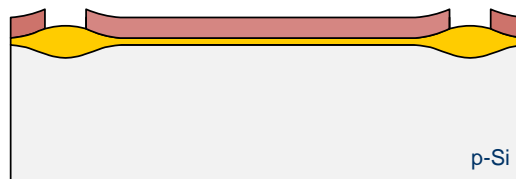
8. Etching: Subseuqent to the oxidation, the nitride is removed in a wet chemical etching process.
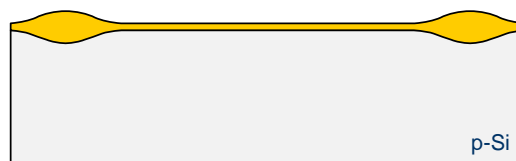
9. Deposition: Via low pressure CVD, polycrystalline silicon is deposited which represents the gate electrode.
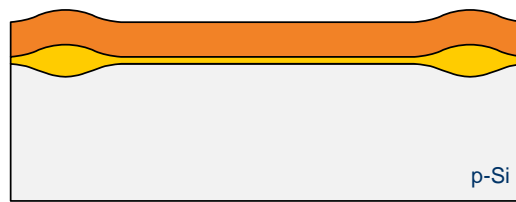
10. Photolithography: Again a resist layer on top of the polysilicon is patterned.
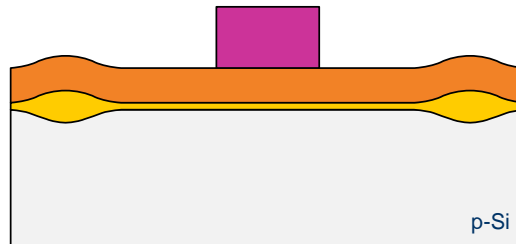
11. Etching: The photoresist in turn serves as a mask layer, via reactive ion etching the gate is patterned.
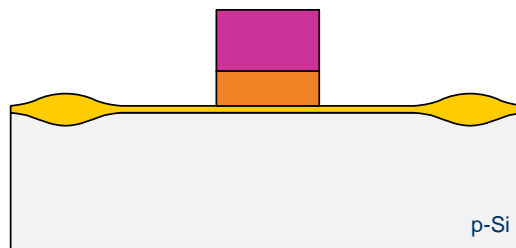
12. Resist removal: The resist is removed via wet-chemical etching.
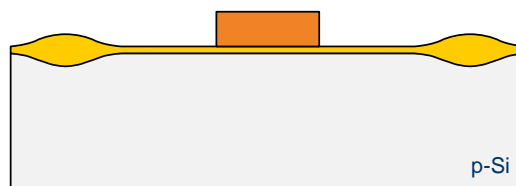
13. Oxidation: A thin oxide (post oxide) is deposited as an insulating layer for the gate electrode as well as a spacer for the subsequent source and drain implantation.
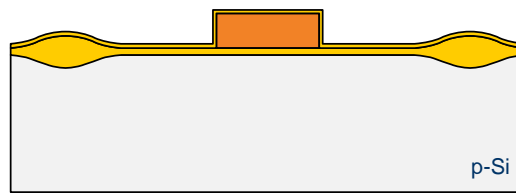
14. Ion implantation: Via ion implantation with phosphorus, the source and drain regions are introduced (n-type). Since the gate electrode acts as an mask during implantation, the width of the n-channel between the source and drain is preset. This is called self alignment.
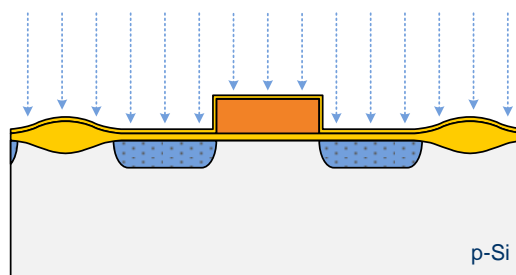
15. Oxidation: As an isolation a nonmetal is deposited (e.g. oxide). This happens in a LPCVD process with TEOS, which provides a good step coverage.
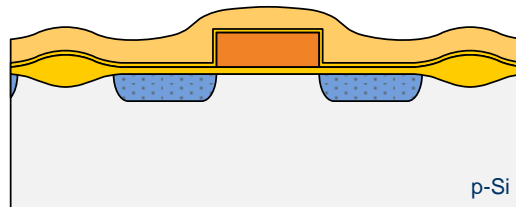
16. Photolithography and etching: In a further step a resist layer is structured and the edges of the contact holes are rounded in an isotropic etch process.
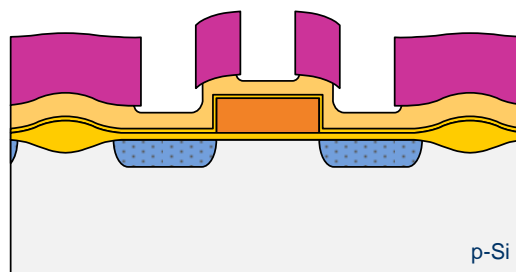
17. Etching: Subsequently the contact holes are opened in a highly anisotropic etch process.
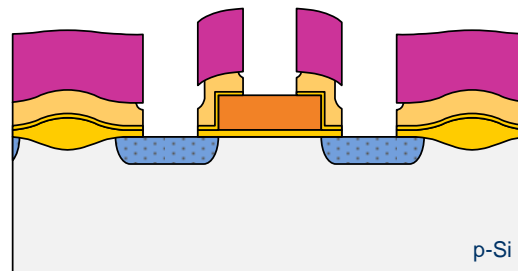
18. Metallization: The contact holes are filled with aluminum via sputtering.
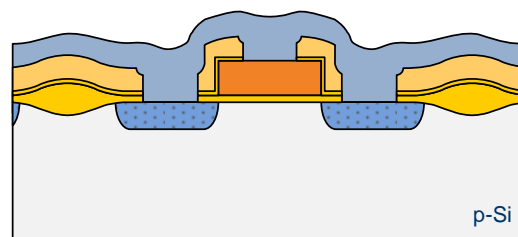
19. Photolithography: In a final lithography step a new resist mask is patterned.
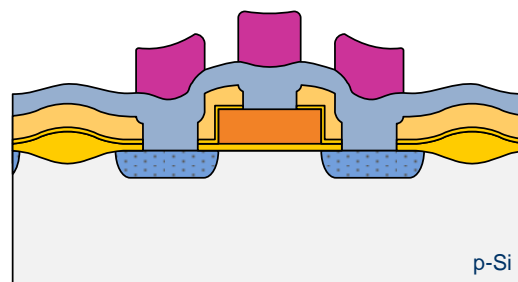
20. Etching: The pattern is transferred into the underlying metallization in an anisotropic dry etch process.
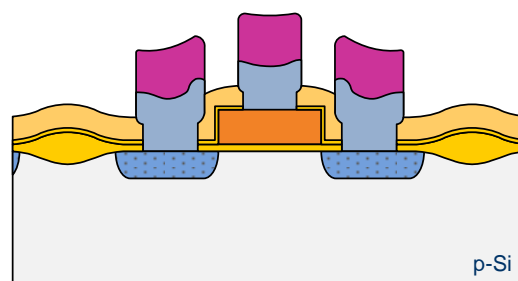
21. Resist removal: Finally, the resist is removed and aluminum conductors remain to actuate the transistor.
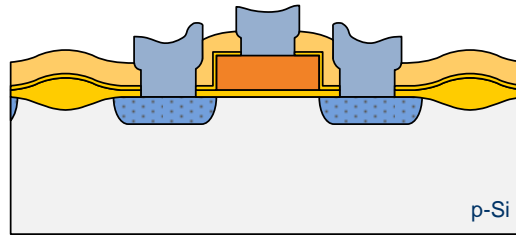
Actual the construction of a transistor is much more complex, since additional planarization layers for photolithography are neccessary or secondary drain and source implantations have to be done to adjust the threshold voltage accurately. On the slopes of the gate elektrode additional (side wall) spacers can be formed to set the exact length of the channel or fine tune the doping profile respectively.

1.8.3 Mode of operation

Enhancement-mode:
Without a positive voltage applied to the gate there are no electrons available, which could act as free charge carriers between source and drain, since the substrate is p-doped. In steady state holes in the substrate act as majority charge carriers, while the electrons are minority charge carriers.

A positive voltage applied to the gate attracts electrons in the substrate, while holes are pushed away. Thus forming a conducting n-channel beneath the gate electrode and in-between source and drain, respectively. The insulating silicon dioxide layer prevents a current flow between the substrate and the gate.

Since the transistor blocks the current flow without an applied voltage, the transistor is also called self-locking.
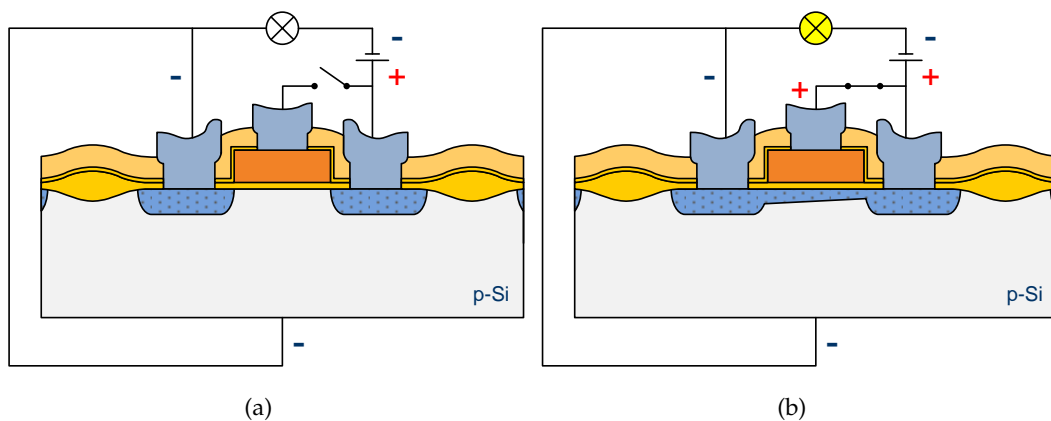
Fig. 1.19: Enhancement nFET

Depletion-mode:

With a light n-type doping between the source and drain, a conductivity is even possible without a gate voltage (a voltage between source and drain is sufficient). So-called depletion FETs, or self-conducting transistors block the current flow only, if the voltage applied to the gate electrode is lower than the voltage at the source port. If the gate voltage is decreased, the electrons that are located beneath the gate, are pushed away - the conductive electron-channel is lost.

1.9 Bipolar transistors

1.9.1 General layout

The second important type of transistors, next to the field-effect transistor, is the bipolar transistor. Its mode of operation is based on both charge carriers (thus bipolar), electrons and holes. Bipolar transistors are faster than field-effect transistors, however, they require more space and are therefore more expensive in mass production.

Bipolar transistors consist essentially of two mutually connected p-n junctions with the layer sequence n-p-n or p-n-p. The connections of the bipolar transistor are emitter (E), base (B) and collector (C). While emitter and collector have the same doping, the very thin base layer in-between is doped contrarily.

In this article an NPN transistor in standard buried collector (SBC) construction is described, the mode of operation of PNP transistors is analog (the signs of the applied

voltages have just to be reversed).

1.9.2 Construction of an NPN bipolar transistor

1. Substrate: Basis for an NPN bipolar transistor is a p-doped (boron) silicon substrate, a thick oxide layer (e.g. 600 nm) is deposited on top.

p-Si
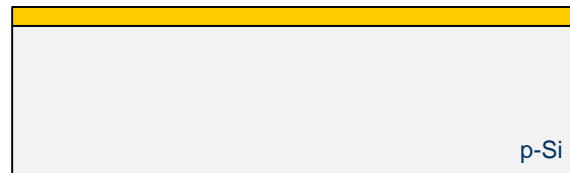
2. Buried Layer Implantation: The oxide serves as an implantation mask. As dopant antimony (Sb) is used, since its diffusion coefficient is lower than of phosphorus, and therefore the dopant won't diffuse as much in subsequent processes. The highly n^+-doped buried collector serves as a low-resistance contact surface for the collector port.

p-Si
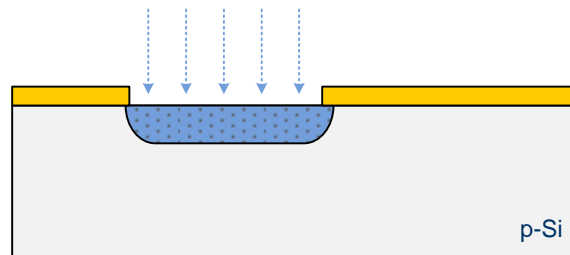
3. Homoepitaxy: In an epitactical process a high-impedance (low n^--doped) collector layer is deposited (typically 10 microns).

n^--Si

p-Si
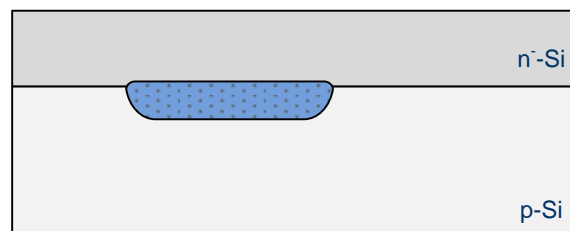
4. Base implantation: With boron the p-doped base is introduced, a subsequent diffusion step magnifies its dimensions.
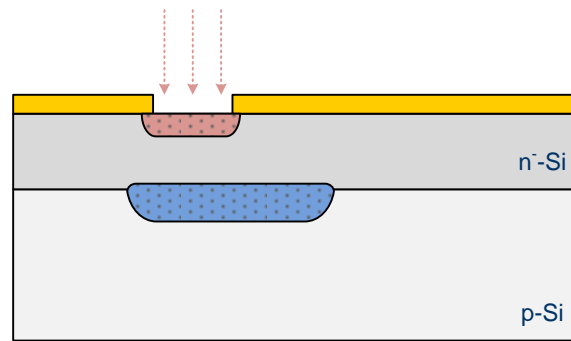
5. Emitter and collector implantation: With phosphorus both highly n^+-doped emitter and collector junctions are introduced.
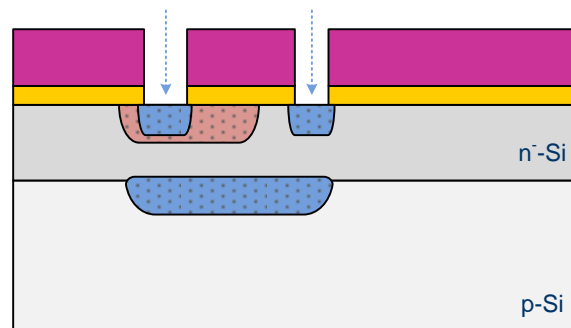
6. Metallization and photolithography: Aluminum is deposited in a sputtering process for contacting and a resist layer is patterned on top of it.
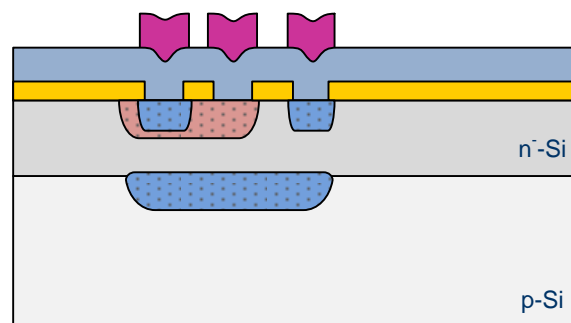
7. Etching: Finally, the connectors for emitter, base, and collector are structured in an anisotropic dry etch process.
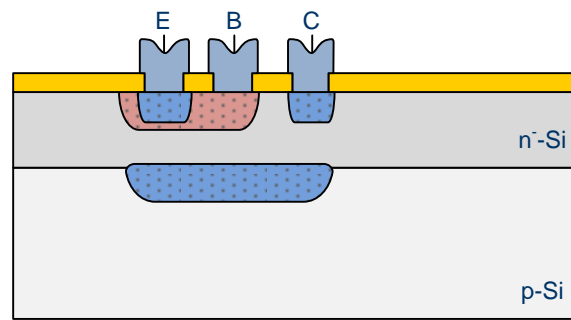
Due to many improvements, bipolar transistor do have more than three layers (npn or pnp). Nowadays the collector region consists out of at least two variable doped zones. The terms npn and pnp just describe the active area, not the actual filmstack.

1.9.3 Mode of operation

The two p-n junctions are hereinafter named as EB (emitter-base) and CB (collector-base). Without an external voltage a depletion zone forms at the interfaces of EB and CB. If a negative voltage is applied to the emitter and a positive voltage is applied to the collector, the depletion zone at EB decreases, while the depletion zone at CB increases. If a positive voltage is now applied to the base, EB becomes conductive - electrons can reach the base layer. As this layer is very thin, the charge carriers can be injected into the collector, where they are extracted due to the positive external voltage. Thus, a current flow from emitter to collector is established. Almost all electrons (>95 %) can reach the collector if only a small voltage is applied to the base, which means that with a relatively small base current (E to B) a very large collector current (E to C) is possible.
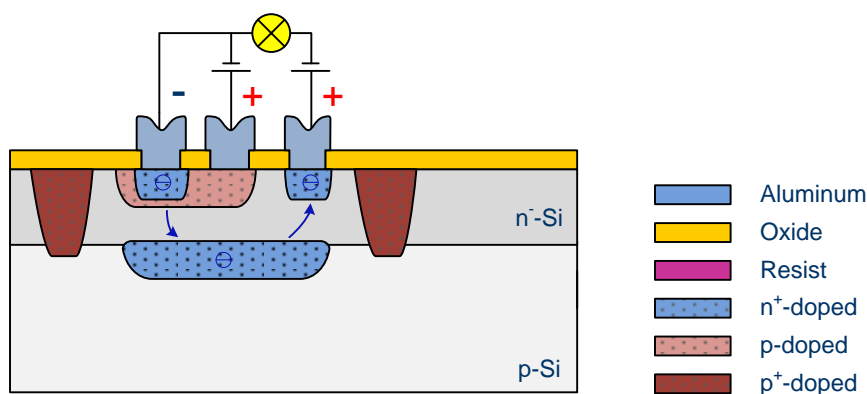
Fig. 1.20: SBC bipolar transistor

The two deep p^+-doped regions are used for lateral isolation from other components. In addition to the transistor a resistance (not in graphics) is needed, since bipolar transistors can not be controlled currentless.

1.10 Construction of a FinFET

1.10.1 General layout and mode of operation

The basic electrical layout and the mode of operation of a FinFET does not differ from a traditional field effect transistor. There is one source and one drain contact as well as a gate to control the current flow.

In contrast to planar MOSFETs the channel between source and drain is build as a three dimensional bar on top of the silicon substrate, called fin. The gate electrode is then wrapped around the channel, so that there can be formed several gate electrodes on each side which leads to reduced leakage effects and an enhanced drive current.

The manufacture of a bulk silicon-based multi gate transistor with three gates (tri gate) is described below.

1.10.2 Construction of a bulk silicon-based FinFET

1. Substrate: Basis for a FinFET is a lightly p-doped substrate with a hard mask on top (e.g. silicon nitride) as well as a patterned resist layer.
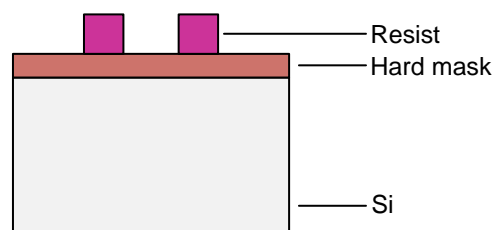
2. Fin etch: The fins are formed in a highly anisotropic etch process. Since there is no stop layer on a bulk wafer as it is in SOI, the etch process has to be time based. In a 22 nm process the width of the fins might be 10 to 15 nm, the height would ideally be twice that or more.
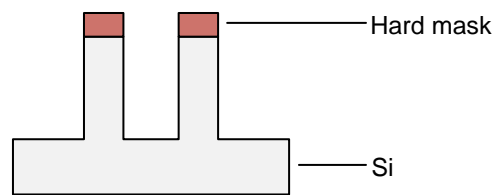
3. Oxide deposition: To isolate the fins from each other a oxide deposition with a high aspect ratio filling behavior is needed.
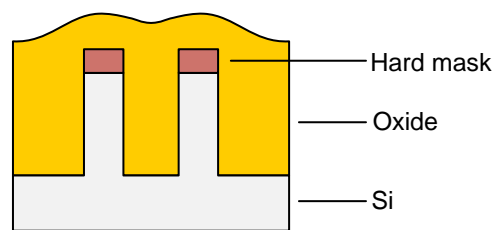
4. Planarization: The oxide is planarized by chemical mechanical polishing. The hard mask acts as a stop layer.
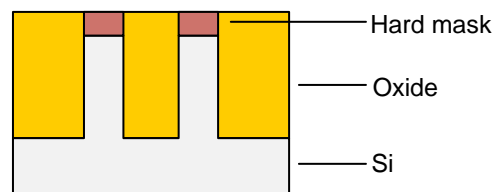
5. Recess etch: Another etch process is needed to recess the oxide film to form a lateral isolation of the fins.
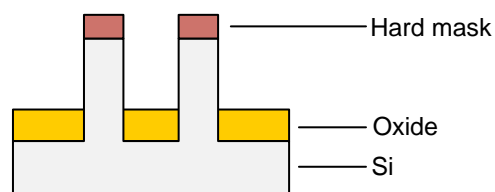
6. Gate oxide: On top of the fins the gate oxide is deposited via thermal oxidation to isolate the channel from the gate elctrode. Since the fins are still connected underneath the oxide, a high-dose angled implant at the base of the fin creates a dopant junction and completes the isolation (not illstrated).

— Gateoxide

— Oxide

— Si
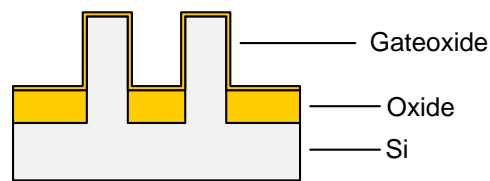
7. Deposition of the gate: Finally a highly n^+-doped poly silicon layer is deposited on top of the fins, thus up to three gates are wrapped around the channel: one on each side of the fin, and - depending on the thickness of the gate oxide on top - a third gate above.

— Gate

— Oxide

— Si
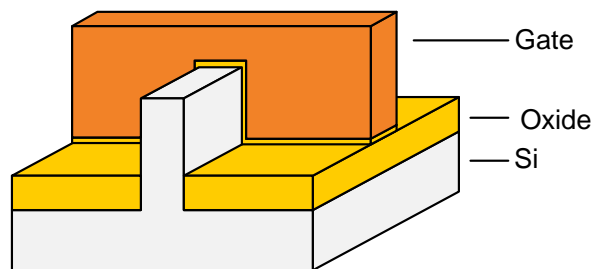
Fig. 1.21: FinFET in bulk process

The influence of the top gate can also be inhibited by the deposition of a nitride layer on top of the channel.

Since there is an oxide layer on an SOI wafer, the channels are isolated from each other anyway. In addition the etch process of the fins is simplified as the process can be stopped on the oxide easily.

— Resist
— Hard mask

— Si

— Oxide

— Si

— Gate

— Oxide

— Si
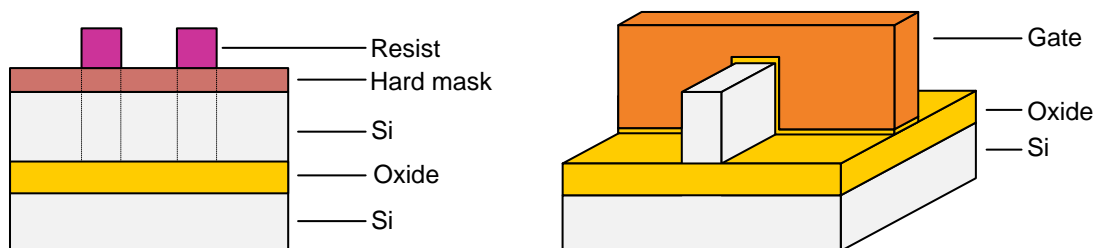
Fig. 1.22: FinFET on SOI

2 Wafer fabrication

2.1 Properties of silicon

Silicon is the chemical element with the atomic number 14 in the periodic table of the elements. Silicon is a classic semiconductor, its conductivity lies between that of conductors and dielectrics. Naturally silicon (from the latin silex/silicis: pebbles) occurs only as oxide: silicon dioxide (SiO_2) in form of sand, quartz, or silicate (compounds of silicon with oxygen, metals and others). Thus silicon is a very cheap starting material, whose value is determined with further processing. Other semiconductors such as germanium or gallium arsenide compound semiconductors offer substantially improved electrical properties than silicon: the charge carrier mobility and thus the resulting switching speeds are significantly higher in germanium and GaAs. However, silicon has significant advantages in contrast to other semiconductors.

On a silicon crystal oxide layers can be produced very easily, the resulting silicon dioxide is an insulator of highest quality which can be fabricated precisely on the substrate. To create similar insulators on germanium or gallium arsenide is very expensive. The possibility to change the conductivity by doping silicon is another big advantage. Other substances are in part very toxic, and compounds with these elements are not as durable and stable as in silicon. Requirement for the use of silicon in semiconductor manufacturing is that the silicon is present in an ultrapure form as single crystal. This means that the silicon atoms in the crystal lattice are regularly arranged and there are absolutely no undefined impurities in the material.

In addition to the single crystal, there is polysilicon (poly = many) and amorphous silicon (a-Si). While the single crystal silicon is the basis for microelectronics in form of circular wafers, the polycrystalline silicon is suitable to fulfill specific tasks (e.g. masking, gate electrode, ...). Polysilicon is made up of many individual irregularly arranged single crystals, and can be deposited and patterned very easily. Amorphous silicon does not have a regular but a disordered lattice structure and plays no role in semi-

conductor manufacturing, but amorphous silicon offers advantages over other forms
of silicon in the manufacture of thin film solar cells.

2.2 Raw silicon

2.2.1 Production of raw silicon

Silicon as it is used in semiconductor manufacturing, is made up of quartz. Oxygen
which reacts very fast with silicon even at room temperature, and which is present in
quartz associated with silicon as silicon dioxide SiO_2, must be removed. This is done
just above the melting point of silicon (1414 °C) in furnaces using carbon. At 1460 °C
oxygen cleaves of the silicon and reacts with carbon C to carbon monoxide CO:

$$SiO_2 + 2\,CSi \longrightarrow 2\,CO$$

Iron prevents the reaction of silicon and carbon to form silicon carbide. At these tem-
peratures the carbon monoxide is in gaseous state and can be separated from the
molten silicon easily. However, the raw silicon is still heavily polluted. There are up
to 5 % impurities, such as for example iron, aluminum, phosphorus, and boron. These
substances must be removed in additional processes.

2.2.2 Purification of the raw silicon

Using a trichlorosilane process many impurities are filtered out by distillation. The raw
silicon and hydrogen chloride HCl react at about 300 °C to form gaseous hydrogen H_2
and trichlorosilane $SiHCl_3$:

$$Si + 3\,HCl \longrightarrow SiHCl_3 + H_2$$
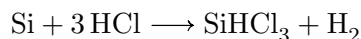
The contaminants which also react with the chlorine, need higher temperatures to
transfer in the gaseous state. This allows separation of the trichlorosilane. Only car-

bon, phosphorus, and boron, which have similar condensation temperatures, can not be filtered out in this process.

The trichlorosilane process can be reversed, so that the purified silicon condeses in polycrystalline form. This is done at approximately 1100 °C by adding hydrogen inside a quartz chamber, in which thin silicon rods are placed:

$$SiHCl_3 + H_2 \longrightarrow Si + 3\,HCl$$
$$4\,SiHCl_3 \longrightarrow Si + 3\,SiCl_4 + 2\,H_2$$
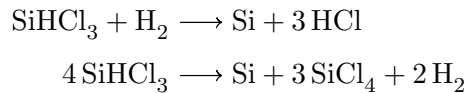
The silicon reflects on the silicon rods which grow to bars with a diameter of more than 30 mm. This polysilicon could already be transformed into a single crystal using the Czochralski process, however, the degree of purity for semiconductor manufacturing is still not high enough.
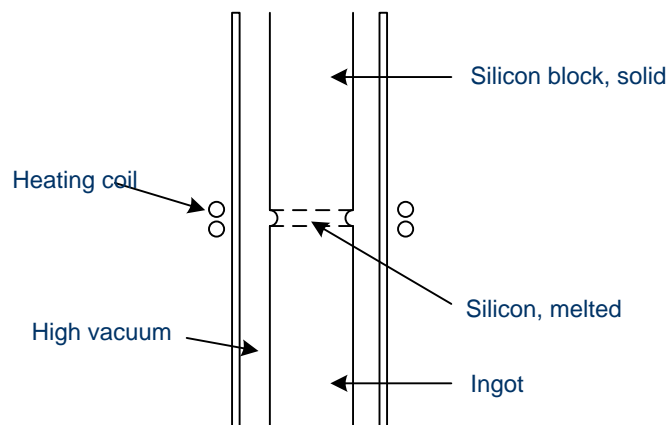
Fig. 2.1: Illustration of the zone cleaning process

2.2.3 Zone cleaning

To increase the purity once more a cleaning process is used. Thereby a high frequency coil is placed around the silicon rod to melt the silicon bars, and therefore the contaminations accumulate at the bottom due to higher solubility in the liquid phase; the surface tension of the silicon prevents the melt to flow out. By multiple repetition of this procedure, the content of impurities in silicon is further reduced and thus it can be

used for fabrication of the single crystal. To prevent further contamination, all of the processes are made under a vacuum atmosphere.

At the end of these processes the silicon has a purity of more than 99,9999999 %, which means that there is less than 1 foreign atom per 1 billion silicon atoms.

2.3 Fabrication of the single crystal

2.3.1 The single crystal

A single crystal (monocrystal), as it is required in semiconductor manufacturing, is a regular arrangement of atoms. There are polycrystalline (composition of many small single crystals) and amorphous silicon (disordered structure). Depending on the orientation of the lattice, silicon wafers have different surface structures which impact various properties as the charge carrier mobility or the behaviour in wet-chemical anisotropic etching of silicon.

| 100 | 110 | 111 |
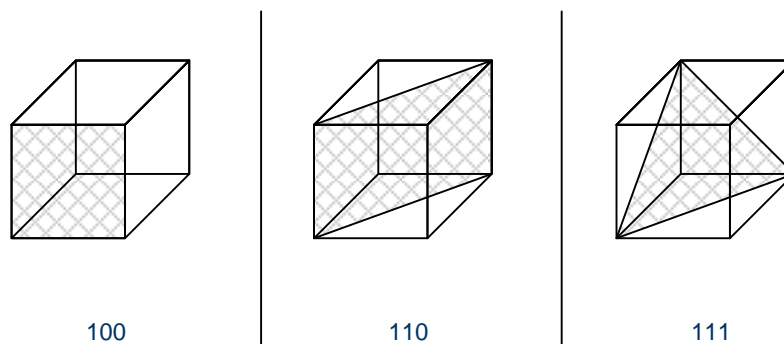
Fig. 2.2: Crystal orientation

In micromechanics the crystal orientation is of particular importance. It allows microchannels with perpendicular walls on (110) silicon, whereas flank angles of 54.74° are possible on (100) orientation.

2.3.2 Czochralski process

The polycrystalline silicon, as it is present after the zone cleaning, is melted in a quartz crucible nearly above the melting point of silicon. Now dopants (e.g. boron or phos-

phorus) can be added to the melt to achieve appropriate electrical characteristics of the single crystal.

A seed crystal (a perfect single crystal) on a rotating rod is brought to the surface of the silicon melt. This seed crystal pretends the orientation of the silicon crystal. In contact with the seed crystal, the melt overtakes its crystal structure. The fact that the crucible temperature is only slightly above the melting point of silicon, the melt solidifies immediately on the seed and the crystal grows.
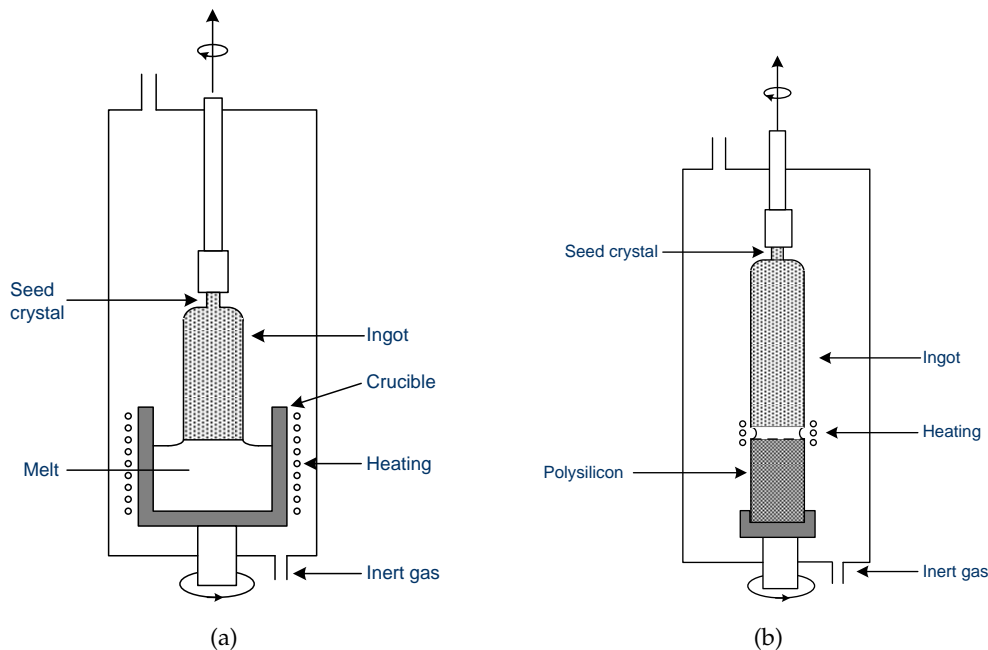
Fig. 2.3: Illustration of (a) the Czochralski and (b) the float-zone process

The seed is slowly pulled upward with constant rotation, while there is constant contact with the melt. The crucible turns in the opposite direction of the seed crystal. A constant temperature of the melt is essential to ensure a steady growth. The diameter of the single crystal is determined by the drawing speed, which provides 2 to 25 cm/h. The higher the drawing speed, the thinner the crystal. The entire apparatus is located in a controlled atmosphere, so that no oxidation of silicon can take place.

The disadvantage of this procedure is that the melt is accumulated with dopants during the process, since the dopants are more solubly in the melt than in the solid state. Thus the dopant concentration along the silicon rod is not constant. Also impurities or metals can dissolve from the crucible and built into the crystal.

The advantages of this method are the lower costs, and the ability to produce larger wafer sizes as in flot-zone processes.

2.3.3 Float-zone silicon

In contrast to the Czochralski process the polysilicon is not entirely molten, but, as in the zone cleaning, only a small area (a few millimeters).

Again, a seed crystal, which will be introduced to the end of the polycrystalline silicon rod, sets the crystal structure. The polycrystal is molten and assumes the structure of the seedling. The heated region is slowly guided along the rod, the polycrystalline silicon rod slowly transforms into a single crystal.

Since only a small portion of the polycrystalline silicon is molten, it can hardly be polluted (impurities accumulate at the bottom since their higher solubility). The doping is done by additions of dopants into the inert gas (eg with diborane or phosphine) which flows around the apparatus.

2.4 Wafer fabrication

2.4.1 Wafer separation and surface refinement

At first the single crystal is turned to a desired diameter and then bedight with one or two flats. The larger, first flat allows an precise alignment of the wafer during manufacturing. The second flat is used to detect the type of the wafer (crystal orientation, p-/n-type doped), but is not always used. Wafers with a diameter of 200 mm or above use a notch instead. This tiny notches on the edge of the disk also provide an alignment of the wafer, but take up much less costly wafer surface.

Sawing:
With an annular saw, whose cutting edge is filled with diamond splinters, the single crystal is sawn into thin discs = wafer. The saw provides a high accuracy during sawing without bumps. Up to 20 % of the crystal rod is lost due to the width of the saw blade. However, nowadays more often wire saws are used, in which multiple wafers can be cut at once from the staff. Therefore a long wire, which is wetted with a suspension of silicon carbide grains and a carrier (glycol or oil), is lead through rotating rollers. The
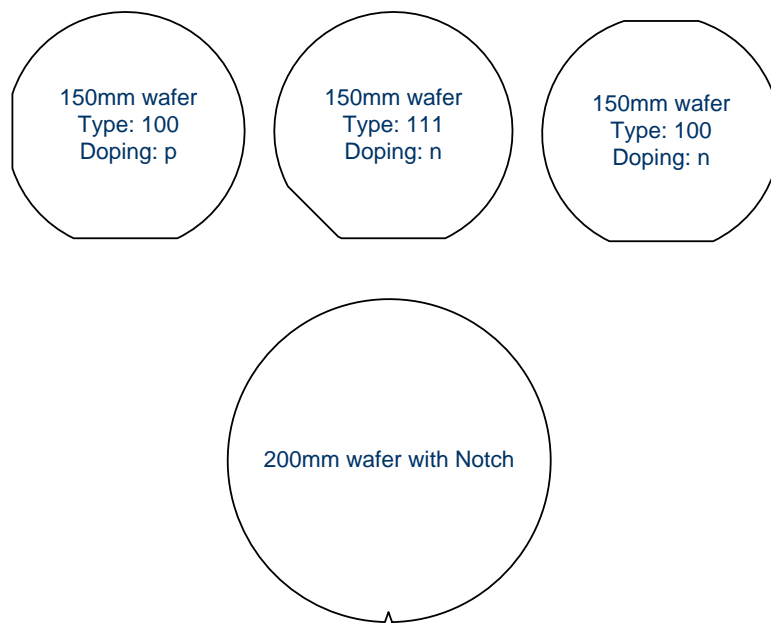
Fig. 2.4: Different types of wafers

silicon crystal is drained into the wire grid and thus cut into single wafers. The wire moves in counterstep with about 10 m/s and has a typically thickness of 0.1-0.2 mm.

(a)   (b)
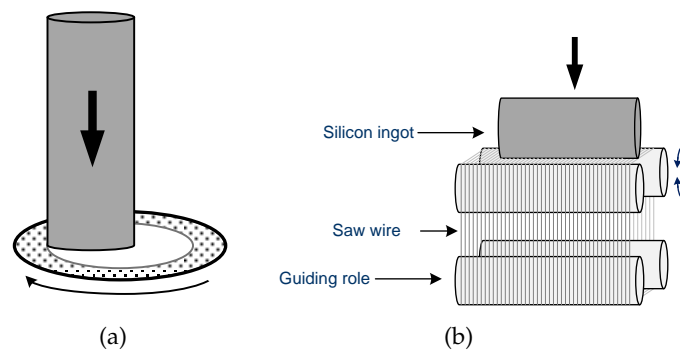
Fig. 2.5: Annular and wire saw

After sawing, the slices have a rough surface, and due to mechanical stress damages in the crystal lattice. For finishing the surface, the wafers pass several process steps.

Lapping:
Using granular abrasives (e.g. aluminum) 50 microns (0.05 mm) of the wafer surface are removed on a rotating steel disc. The grain size is reduced in stages, but the surface is re-injured due to the mechanical treatment. The flatness after lapping is about 2

microns.

Beveling of the edge:
In subsequent processes, the discs must have no sharp edges, as deposited layers may flake off otherwise. Therfore the bevel of the wafers are rounded with a diamond cutter.

Wafer

Milling cutter
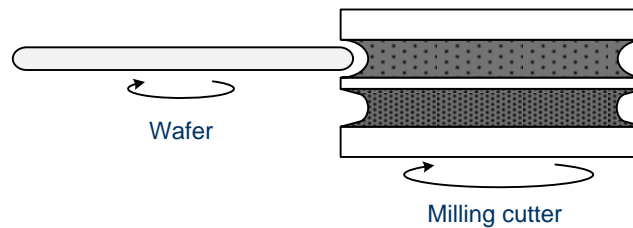
Fig. 2.6: Beveling

Etching:
In an additional wet etch process, with a mixture of hydrofluoric, acetic, and nitric acid, 50 microns are removed. Because this is a chemical process, the surface is not damaged. Crystal defects are permanently resolved.

Polishing:
This is the final step of surface refinement. At the end of the polishing step, the wafers do not have a bump of more than 3 nm (0.000003 mm). The wafers are treated with a mixture of sodium hydroxide NaOH, water, and silicon oxide grains. The oxide removes additional 5 microns from the surface, the hydroxide removes machining marks caused by the oxide grain.

2.4.2 Historical development of the wafer size

The manufacture of integrated circuits on silicon wafers started in the mid 1960s on wafers with a diameter of 25 mm. Nowadays, in modern semiconductor manufacturing wafers with a diameter of 150-300 mm are used. By 2012 the mass production of microchips on wafers with a diameter of 450 mm is expected; prototypes have already been produced for research purposes. The wafer surface is then increased by more than 300-fold of the tiny 1-inch wafer 50 years ago.

With larger wafers, the throughput rate increases significantly in the manufacture of microchips, whereby the cost is reduced accordingly in the production. Thus, with identical structure sizese more than twice as many chips can be produced on a 300 mm

Type	Diameter [mm]	Thickness [μm]	1st flat [mm]	Bowing [μm]
6 inch	150±0,5	≈700	55-60	25
8 inch	200±0,5	≈700	55-60	25
12 inch	300±0,5	≈700	55-60	25

Tab. 2.1: Typical data of wafers

wafer as on a 200 mm wafer. In addition, with increasing diameter the wafer's edge is less curved and thus the cut-off minimized (since chips are off rectangular shape).
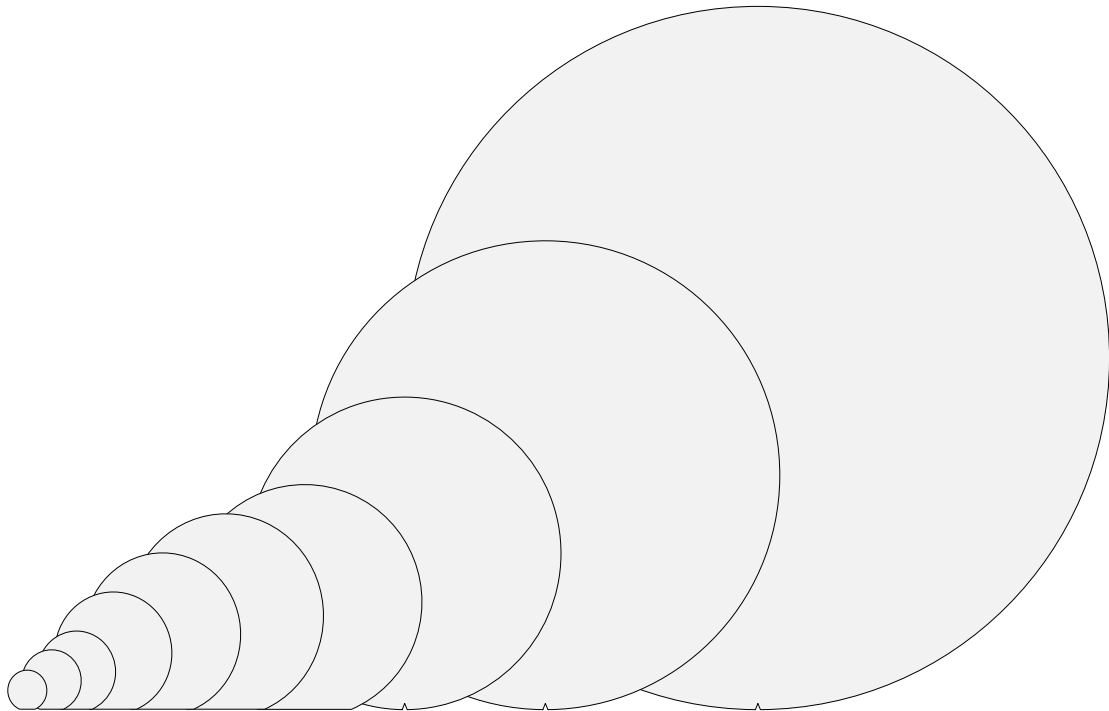
Fig. 2.7: Different wafer sizes: 25, 38, 51, 75, 100, 125, 150, 200, 300, 450 [mm] (drawn to scale)

2.5 Doping techniques

2.5.1 Definition

Doping means the introduction of impurities into the semiconductor crystal to deliberately change its conductivity due to deficiency or excess of electrons. In contrast to the doping during the wafer fabrication, where the entire wafer is doped, this article

describes the partial doping of silicon. The introduction of foreign substances can be achieved by diffusion, ion implantation (or alloy).

2.5.2 Diffusion

Molecular diffusion, often called simply diffusion, is a net transport of molecules from a region of higher concentration to one of lower concentration by random molecular motion. The result of diffusion is a gradual mixing of materials. To illustrate: a drop of ink in a glass of water is evenly distributed after a certain amount of time. In a silicon crystal, one finds a solid lattice of atoms through which the dopant has to move. This can be done in different ways:

- **Empty space diffusion:** the impurity atoms can fill empty places in the crystal lattice which are always present, even in perfect single crystals.

- **Inter lattice diffusion:** the impurity atoms move in-between the silicon atoms in the crystal lattice.

- **Changing of places:** the impurity atoms are located in the crystal lattice and are exchanged with the silicon atoms.

The dopant can diffuse as long as either a concentration gradient is balanced, or the temperature was lowered, so that the atoms can no longer move. The speed of the diffusion process depends on several factors:

- Dopant

- Concentration gradient

- Temperature

- Substrate

- crystallographic orientation of the substrate

Diffusion with an exhaustible source:
Diffusion with an exhaustible source means that the dopant is available in a limited amount only. The longer the diffusion process continues, the lower the concentration at the surface, and therefore the depth of penetration into the substrate increases. The diffusion coefficient of a substance indicates how fast it moves in the crystal. Arsenic
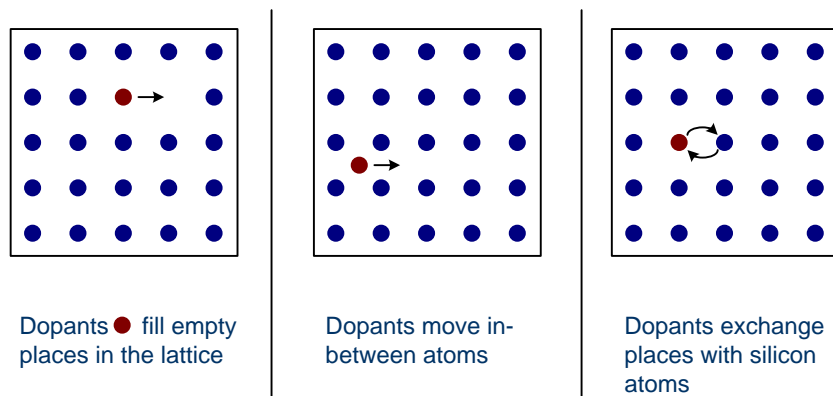
Fig. 2.8: Diffusion process

with a low diffusion coefficient penetrates slower into the substrate, as for example phosphorus or boron.

Diffusion with an inexhaustible source:
In diffusion processes with an inexhaustible source the dopants are available in unlimited amount, and therefore the concentration at the surface remains constant during the process. Particles that have penetrated into the substrate are continually replenished.

2.5.3 Diffusion methods

In the subsequent processes the wafers are placed in a quartz tube that is heated to a certain temperature.

Diffusion from the gas phase:
A carrier gas (nitrogen, argon, ...) is enriched with the desired dopant (also in gaseous form, e.g. phosphine PH_3 or diborane B_2H_6) and led to the silicon wafers, on which the concentration balance can take place.

Diffusion with solid source:
Slices which contain the dopants are placed in-between the wafers. If the temperature in the quartz tube is increased, the dopant from the source discs diffuses into the atmosphere. With a carrier gas, the dopant will be distributed uniformly, and thus reaches the surface of the wafers.

Diffusion with liquid source:
As liquid sources boron bromide BBr_3 or phosphoryl chloride $POCl_3$ can be used. A

carrier gas is led through the liquids and thus transporting the dopant in gaseous state. Since not the entire wafers should be doped, certain areas can be masked with silicon dioxide. The dopants can not penetrate through the oxide, and therefore no doping takes place at these locations. To avoid tensions or even fractions of the discs, the quartz tube is gradually heated (e.g. +10 °C per minute) till 900 C. Subsequent the dopant is led to the wafers. To set the diffusion process in motion, the temperature is then increased up to 1200 C.

Characteristic:

- since many wafers can be processed simultaneously, this method is quite favorable

- if there already are dopants in the silicon crystal, they can diffuse out in later processes due to high process temperatures

- dopants can deposit in the quartz tube, and be transported to the wafers in later processes

- dopants in the crystal are spreading not only in perpendicular orientation but also laterally, so that the doped area is enlarged in a unwanted manner
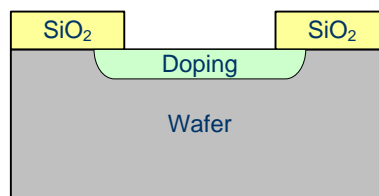
Fig. 2.9: Diffusion with an oxide mask

2.5.4 Ion implantation

In the ion implantation charged dopants (ions) are accelerated in an electric field and irradiated onto the wafer. The penetration depth can be set very precisely by reducing or increasing the voltage needed to accelerate the ions. Since the process takes place at room temperature, previously added dopants can not diffuse out. Regions that should not be doped, can be covered with a masking photoresist layer.

An implanter consists of the following components:

- **ion source:** the dopants in gaseous state (e.g. boron trifluoride BF$_3$) are ionized

- **accelerator:** the ions are drawn with approximately 30 kiloelectron volts out of the ion source

- **mass separation:** the charged particles are deflected by a magnetic field by 90 degrees. Too light/heavy particles are deflected more/less than the desired ions and trapped with screens behind the separator

- **acceleration lane:** several 100 keV accelerate the particles to their final velocity (200 keV accelerate bor ions up to 2.000.000 m/s)

- **Lenses:** lenses are distributed inside the entire system to focus the ion beam

- **distraction:** the ions are deflected with electrical fields to irradiate the desired location

- **wafer station:** the wafers are placed on large rotating wheels and held into the ion beam
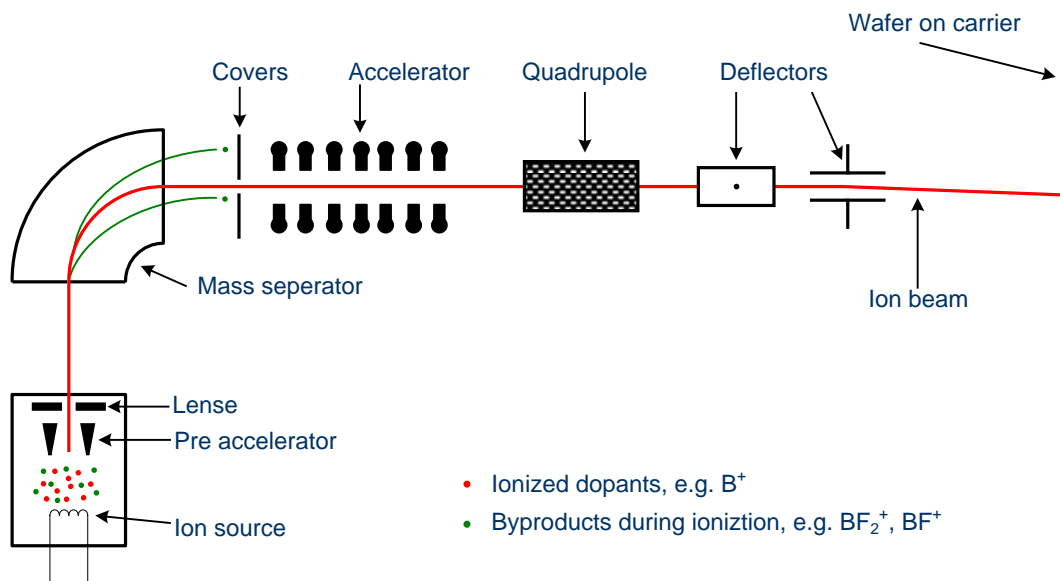
Fig. 2.10: Illustration of an ion implanter

Penetration depth of ions in the wafer:

In contrast to diffusion processes the particles do not penetrate into the crystal due to their own movement, but because of their high velocity. Inside the crystal they are

slowed down by collisions with silicon atoms. The impact causes damage to the lattice since silicon atoms are knocked from their sites, the dopants themselves are mostly placed interstitial. There, they are not electrically active, because there are no bonds with other atoms which may give rise to free charge carriers. The displaced silicon atoms must be re-installed into the crystal lattice, and the electrically inactive dopants must be activated.

Recovery the crystal lattice and activation of dopants:
Right after the implantation process, only about 5 % of the dopants are bond in the lattice. In a high temperature process at about 1000 °C, the dopants move on lattice sites. The lattice damage caused by the collisions have already been cured at about 500 °C. Since the dopants move inside the crystal during high temperature processes, these steps are carried out only for a very short time.

Channeling:
The substrate is present as a single crystal, and thus the silicon atoms are regularly arranged and form "channels". The dopant atoms injected via ion implantation can move parallel to these channels and are slowed only slightly, and therefore penetrate very deeply into the substrate. To prevent this, there are several possibilities:

Channeling: the dopants move parallel to the silicon atoms

By twisting the wafer the dopants collide with the silicon atoms

An oxide layer causes deflection of the dopants
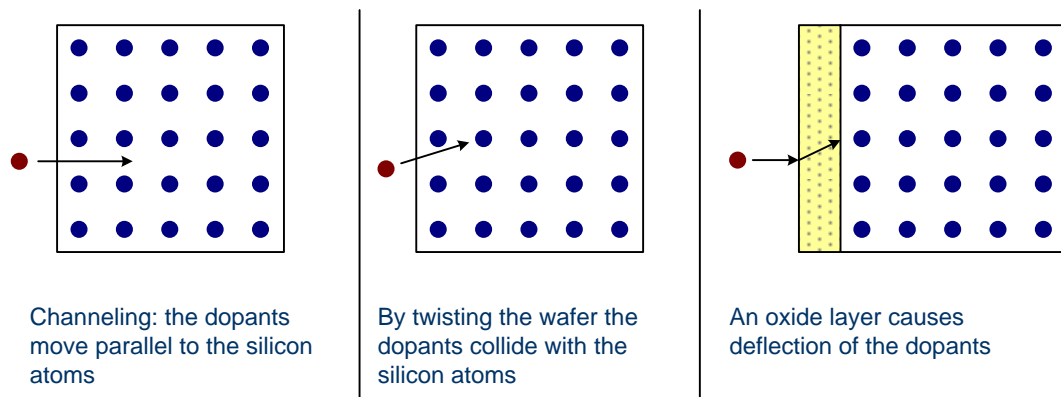
Fig. 2.11: Channeling effect in the ingot

- **Wafer alignment:** the wafers are deflected by about 7° with respect to the ion beam. Thus the radiation is not in parallel direction to the channels and the ions are decelerated by collisions immediately.

- **Scattering:** on top of the wafer surface a thin oxide is applied, which deflects the ions, and therefore prevents a parallel arival

Characteristic:

- the reproducibility of ion implantation is very high

- the process at room temperature prevents the outward diffusion of other dopants

- spin coated photoresist as a mask is sufficient, an oxide layer, as it is used in diffusion processes, is not necessary

- ion implanters are very expensive, the costs per wafer are relatively high

- the dopants do not spread laterally under the mask (only minimally due to collisions)

- nearly every element can be implanted in highest purity

- previous used dopants can deposit on walls or screens inside the implanter and later be carried to the wafer

- three-dimensional structures (e.g. trenches) can not be doped by ion implantation

- the implantation process takes place under high vacuum, which must be produced with several vacuum pumps

There are several types of implanters for small to medium doses of ions (10^{11} to 10^{15} $ions/cm^2$) or for even higher doses of 10^{15} to 10^{17} $ions/cm^2$.

The ion implantation has replaced the diffusion mostly due to its advantages.

Doping using alloy

For completeness it should be mentioned that besides ion implanation and diffusion there is an alternative process: doping using alloy. Since this procedure has disadvantages such as cracks in the substrate, it is not used in today's semiconductor technology any more.

3 Oxidation

3.1 Overview

3.1.1 Application

Oxides in semiconductor industry are used for multiple reasons:

- isolation (interlayer dielectric, ...)

- scatter oxide (ion implantation)

- adaption layers (locos technology, ...)

- planarization (edge rounding, ...)

- mask layer (diffusion, ...)

- alignment marks (lithography)

- cover layer (to prevent corrosion, ...)

3.1.2 Properties of oxide layers

In combination with silicon, oxide appears as silicon dioxide SiO_2. It can be deposited in very thin, electric stable, and uniform layers.

Silicon dioxide, or just dioxide, is very resistant and can only be etched by hydrofluoric acid HF. Water or other acids don't affect oxide; because of metal ions, alkaline lye (KOH, NaOH) can't be used (both KOH and NaOH in contrast are important for anisotropic wet etch in micromechanics). The chemical reaction of dioxide and HF is as follows:

$$SiO_2 + 6\,HF \longrightarrow H_2SiF_6 + 2\,H_2O$$
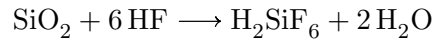
Beside this, oxide is applicable for integrated circuits because it meets the electric requirements (e.g. gate oxide, interlayer dielectric, field oxide).

3.2 Fabrication of oxide layers

3.2.1 Thermal oxidation

Abstract:

In thermal oxidation, silicon wafers are oxidized in furnaces at about 1000 °C. The furnaces consist of a quartz tube in which the wafers are placed on a carrier made of quartz glass. For heating there are several heating zones and for chemical supply multiple pipes. Quartz glass has a very high melting point (above 1500 °C) and thus is applicable for high temperature processes. To avoid cracks or warping, the quartz tube is heatened slowly (e.g. +10 °C per minute). The tempering of the tube can be done very accurate via individual heating zones.
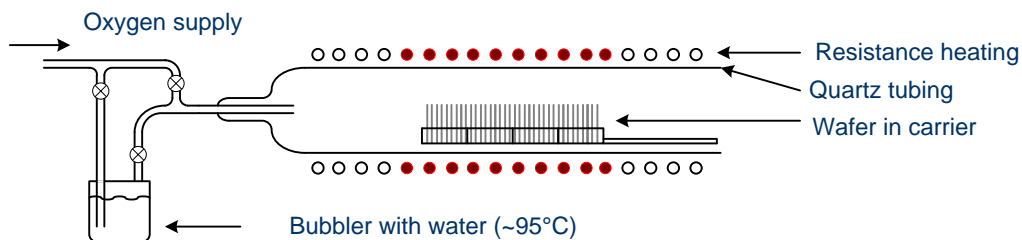
Fig. 3.1: Illustration of a furnace for therma oxidation

The oxygen is led to the wafers in gaseous state and reacts at the wafer surface to form silicon dioxide. A film of glass with amorphous structure is formed. Depending on the gases different oxidations occur (a thermal oxidation has to take place on a bare silicon surface). The thermal oxidation can be devided into the dry and wet oxidation, while the latter can be devided anew into the wet oxidation and the H_2-O_2 combustion.

Dry oxidation:
The oxidation takes place under pure oxygen atmosphere. The silicon and oxide react to form silicon dioxide:

$$\mathrm{Si} + \mathrm{O_2} \longrightarrow \mathrm{SiO_2}$$
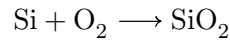
This process is done at 1000 to 1200 °C actually. To create a very thin and stable oxide the process can be done at even lower temperatures of about 800 °C.

Characteristic of the dry oxidation:

- slow growth of oxide

- high density

- high breakdown voltage

Wet oxidation:
In wet thermal oxidation, the oxygen is led through a bubbler vessel filled with heated water (about 95 °C), so that in addition to oxygen water is present in the quartz tube as steam. The oxidation is given by:

$$\mathrm{Si} + 2\,\mathrm{H_2O} \longrightarrow \mathrm{SiO_2} + 2\,\mathrm{H_2}$$
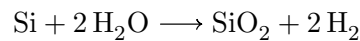
This process is done by 900 to 1000 °C. The characteristics if wet thermal oxidation are:

- fast growth even on low temperatures

- less quality than dry oxides

Temperature	Dry oxidation	Wet oxidation
900 °C	19 nm/h	100 nm/h
1000 °C	50 nm/h	400 nm/h
1100 °C	120 nm/h	630 nm/h

Tab. 3.1: Comparison of the growth rate of wet and dry oxidation of silicon

H_2O_2 combustion:
In the H_2O_2 combustion, pure hydrogen is added to oxygen. The gases are led into the quartz tube and burned in a cold combustion at above 500 °C to avoid Knallgas reaction. This process allows the fabrication of fast growing and only low impurified

films, so that thick oxide layers as well as thin films at moderate temperatures (900 °C) can be procuded. The low temperature also allows fabrication of wafers which were alread doped.

In all thermal oxidation processes, the growth rate on (111) substrates is higher than on (100) substrates. In addition dopants inside the substrate increase the growth rate by far.

Process flow of the oxidation:
In the beginning, the oxygen and silicon react to form silicon dioxide. Now the oxide layer at the surface has to be surpassed by other oxygen atoms which have to diffuse through the dioxide layer to react with the silicon crystal beneath. For this reason the growth rate primarily depends on the reaction time of oxygen and silicon, while at a certain thickness the oxidation rate is mainly determined by the velocity of diffusion of the oxygen through the silicon dioxide. With increasing thickness of the dioxide the growth rate decreases. Since the layer is amorphous, not all bonds in the silicon dioxide are intact. Partial there are dangling bonds (free electrons and holes) at the interface of silicon and SiO2, and therefore there is a slightly positively charged zone at the interface. Since this charges affect the integrated circuit in a negativ manner, one tries to reduce this charges. This can be done with a higher tenmperature during oxidation or by using the wet oxidation which causes only a light charge. Of course wet and dry oxidation can not be exchanged arbitrarily, since electrical properties of gate oxides for example can only be fulfilled by oxides grwon in dry processes.

Segregation:
In thermal oxidation with silicon, the silicon reacts with oxygen to form silicon dioxide. The ratio of the grown oxide layer and of used up silicon is 2.27, which means that the dioxide is growing into the silicon substrate by 45 % of the total thickness of the dioxide.

Dopants which exist within the substrate can be accumulated in the oxide or in the silicon as well. This depends on the solubility of the dopants which can be higher in silicon (e.g. phosphorus) or in silicon dioxide (e.g. boron). The behavior can be calculated as follows, *k* is the coefficient of segregation:

$$k = \frac{\text{Solubility of the dopant in silicon}}{\text{Solubility of the dopant in SiO}_2}$$
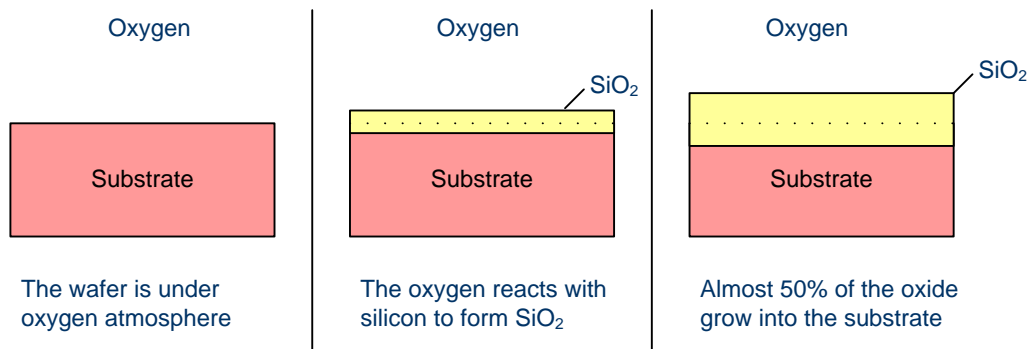
Fig. 3.2: Growth of silicon dioxide on top of silicon

If k is greater than 1 the dopants accumulate at the surface of the substrate, if k is less than 1 the dopants accumulate in the silicon dioxide.

3.2.2 Oxidation by vapor deposition

In thermal oxidation silicon is used up to form oxide. If the silicon surface is covered by other films, the oxide layer has to be created in deposition processes since thermal oxidation needs a bare silicon surface in either case. In deposition processes, oxygen and silicon are added in gaseous states. There are two important processes for oxidation by vapor deposition: the silane pyrolysis and the TEOS deposition. A detailed description of these processes can be found in the chapter deposition.

Silane pyrolysis:
Pyrolysis means a cleavage of chemical compounds - in this case the gas silane SiH_4 and highly purified oxygen O_2 - by heat. Since the toxic silane tends to self-ignition at ambient air above a concentration of 3 % it has to be delutet with nitrogen or argon below 2 %. At about 400 °C silane reacts with oxygen to form silicon dioxide and hydrogen which is exhausted:

$$SiH_4 + O_2 \longrightarrow SiO_2 + 2\,H_2$$
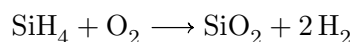
The quality of the dioxide is low, as an alternative a high frequency stimulation at 300 °C can be used. Thus a slightly stabilized oxide is generated.

TEOS deposition:

The liquid TEOS ($SiO_4C_8H_{20}$) which is used in this process contains the required elements silicon and oxygen. Under a vacuum the liquid transforms into gas and is led into a tempered quartz tube with the wafers at about 750 °C where it is cleaved.

$$SiO_4C_8H_{20} \longrightarrow SiO_2 + \text{byproducts}$$

The silicon dioxide deposits on the wafer, byproducts (like H_2O in gaseous state) are exhausted. The uniformity of this oxide layer depends on the pressure and the process temperature. The film has a high electric strength and is very pure.

3.3 LOCOS process

3.3.1 Very large-scale integration

In microfabrication structures are created with photolithography and etch processes. Thus steps are generated on which photoresist can accumulate and reduce the resolving capacity. Due to isotropic etch processes, the resist masks have to be adjusted to ensure the correct size of transfered structures.

At those steps also difficulties during metallization can occur, since conductors are narrowed, and therefore damage due to electromigration is the consequence.

To realize high packing density, in other words to create many devices next to each other in an area as small as possible, steps and other uneven surfaces have to be avoided. This can be done with the LOCOS process: short for LOCal Oxidation of Silicon.

3.3.2 Bird's beak

The LOCOS process utilizes the different rates of oxidation of silicon and silicon nitride, which is used for local masking.

The silicon nitride masks regions where no oxidation should occur, the oxide only growths on the bare silicon. Since silicon and silicon nitride have different coefficients

of thermal expansion, a thin oxide layer - the pad oxide - is deposited between the silicon and the silicon nitride to prevent strain due to temperature changes.
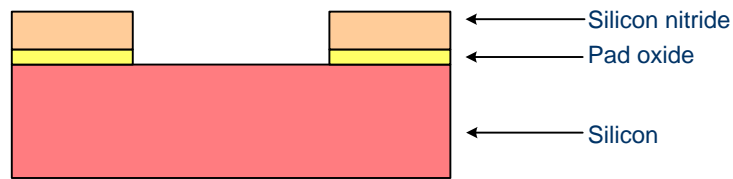
Fig. 3.3: Basic film layer stack

For lateral isolation of transistors, a so-called field oxide (FOX) is deposited on the bare silicon surface. While the oxidation on the bare silicon takes place, the pad oxide causes a lateral diffusion of oxide beneath the silicon nitride and thus a slight growth of oxide at the edge of the nitride mask. This extension has the shape of a bird's beak whose length depends on the length of the oxidation process and the thickness of the pad oxide and the nitride as well.

(a)

(b)
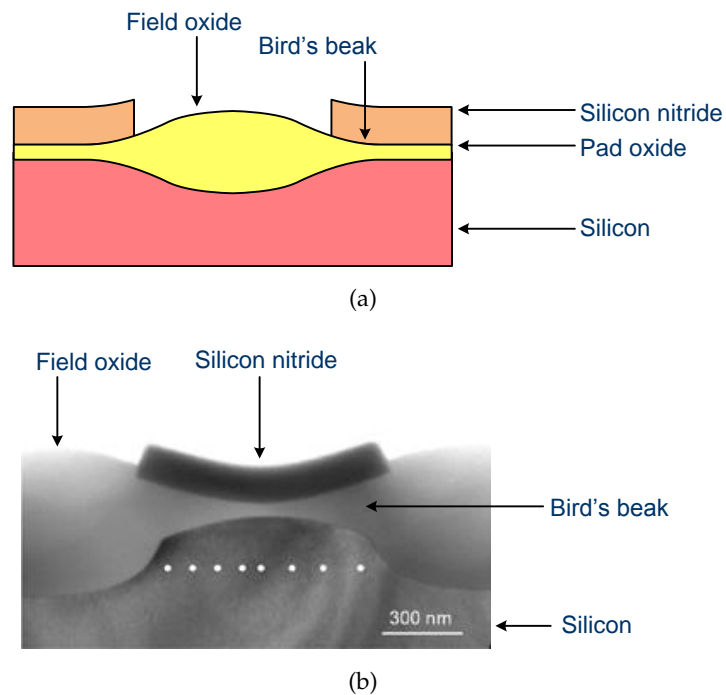
Fig. 3.4: LOCOS profile after oxidation

Beside this effect, the white ribbon or kooi effect can occur during wet oxidation processes. Thereby nitride of the masking and hydrogen used for the wet oxidation react to form ammonia NH3 which can diffuse to the silicon surface and cause a nitridation.

This nitride has to be removed before the gate oxide is deposited because it acts as a masking and prevents oxide growth.

Despite this negative effects, the LOCOS process is an appropriate method to facilitate very large-scale integration. In addition, due to the better uniformity without steps and edges as occur on etched structures, the resolving capacity is improved. The field oxide can be etched back slightly, thus the deposited oxide is removed a bit but the length of the bird's beak is decreased and therefore the surface is even more flattened. This is called fully recessed LOCOS.
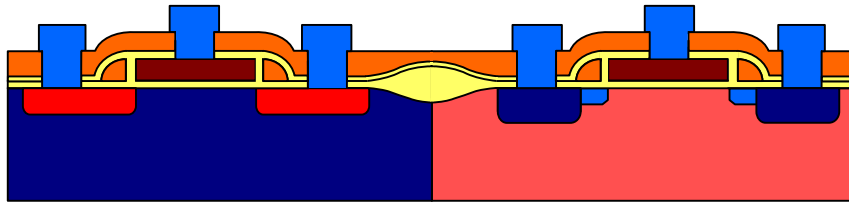
Fig. 3.5: Example of the LOCOS process for lateral isolation of transistors

3.3.3 Alternative

Instead of the local oxidation in modern processes trenches are etched into the silicon substrate and filled with an oxide film to insulate adjacent devices. This so called shallow trench isolation (STI) requires much less space than the locos process, however, more process steps have to be done. Since the trenches are used to isolate transistors the STI process is one of the first process in wafer fabrication. Besides the shallow trench isolation there is a deep trench isolation which is mainly used for anlog devices.

In the 45 nm technology node the trenches for STI are about 100 nm (SOI) or 300 nm (bulk) respectively.

3.4 Film thickness measurement

3.4.1 Metrology

Oxide layers are transparent films. If light is irradiated onto the wafer and reflected, various properties of the light wave are changed which can be detected with metering

devices. If there are multiple layers stacked on each other they must differ in optical properties to allow a determination of the materials.

To monitor the film thickness across the wafer, several measuring points are quantified (e.g. 5 points on 150 mm, 9 points on 200 mm, 13 on 300 mm wafers). Thereby not only the absolute thickness is relevant but also the uniformity across the wafer, because the uniformity is important in subsequent processes. If the deposited layer is too thick or too thin material has to be removed (e.g. by etching) or deposited again.

3.4.2 Interferometry

If light waves interfere with each other the individual waves can be amplified, weakened or annihilated. This phenomenon can be used in semiconductor industry for measuring translucent films.
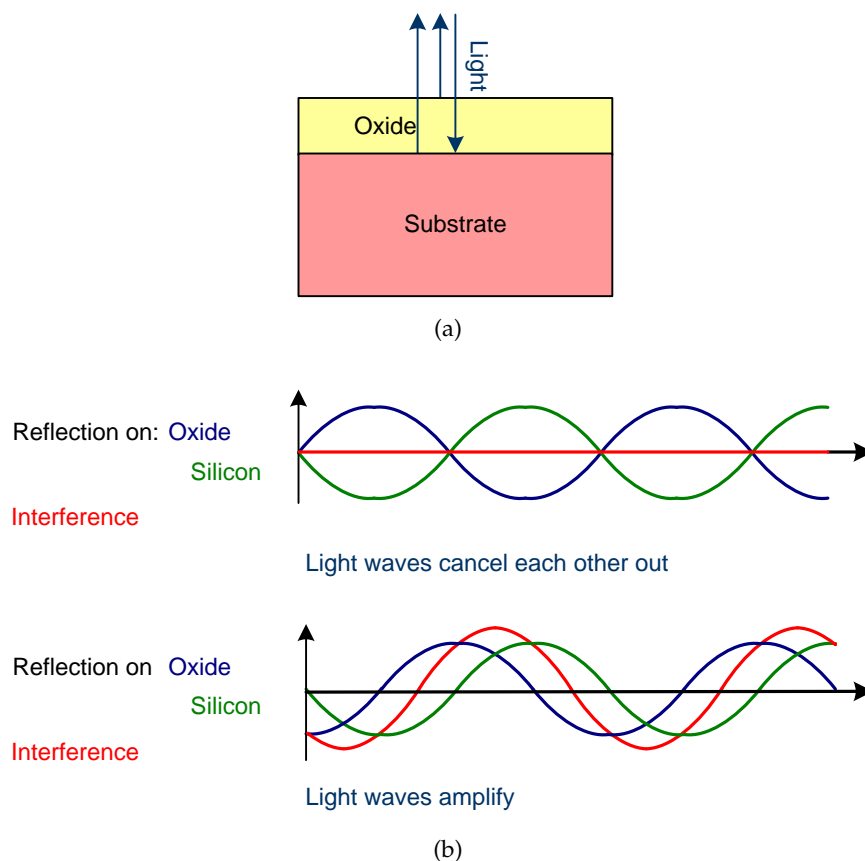
Fig. 3.6: (a) Interferometry, (b) destructive and constructive superimposition of lightwaves

If light is irradiated onto a wafer some beams of light are reflected on top of it and some penetrate into the film. The latter will be reflected on bottom of this layer or penetrate into another layer beneath and so on.

A range of different wavelengths is irradiated onto the wafer and depending on the film thickness of the radiographed layer, the light waves interefere in different ways. Thus results in characteristic interference. A photometer can analyze the reflected light and calculate the film's thickness.

Interferometry can be used on films which are thicker than one fourth of the irradiated light.

3.4.3 Ellipsometry

Ellipsometry is the determination of the optical polarization of light. In metrology for example linear polarized light is irradiated in a fixed angle onto the wafer.
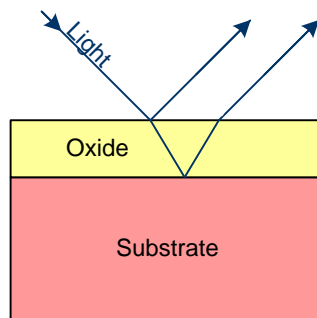
Radiated light is reflected and refracted, also the polarization changes

Fig. 3.7: Ellipsometry

During reflection of light on top of the wafer or on interfaces of two layers, the light's polarization is changed. This change can be detected with an analyzer. By use of known optical properties of the film (e.g. angle of refraction, absorbance), of the light and of the polarization as well, the film thickness can be calculated.

In contrast to the interferometry, elliptical measurements can be used for films with a thickness less than one fourth of the irradiated light.

3.4.4 Appraisal of the measurement

With these technics the thickness measurement can only be carried out indirect, therefor the optical parameters of the measured layers have to be known. Based on these parameters a model is constructed and a measurement is simulated. Subsequent the simulation is compared to the real measurement and the parameters of the model are varied until the measurement and the simulation do fit best.

The more parameteres are varied, the easier a fit can be achieved but the more uncertain is the result. Often a parameter, called goodness of fit (GOF), gives information about how well the simulation and the measurement do match (0 to 100 %).
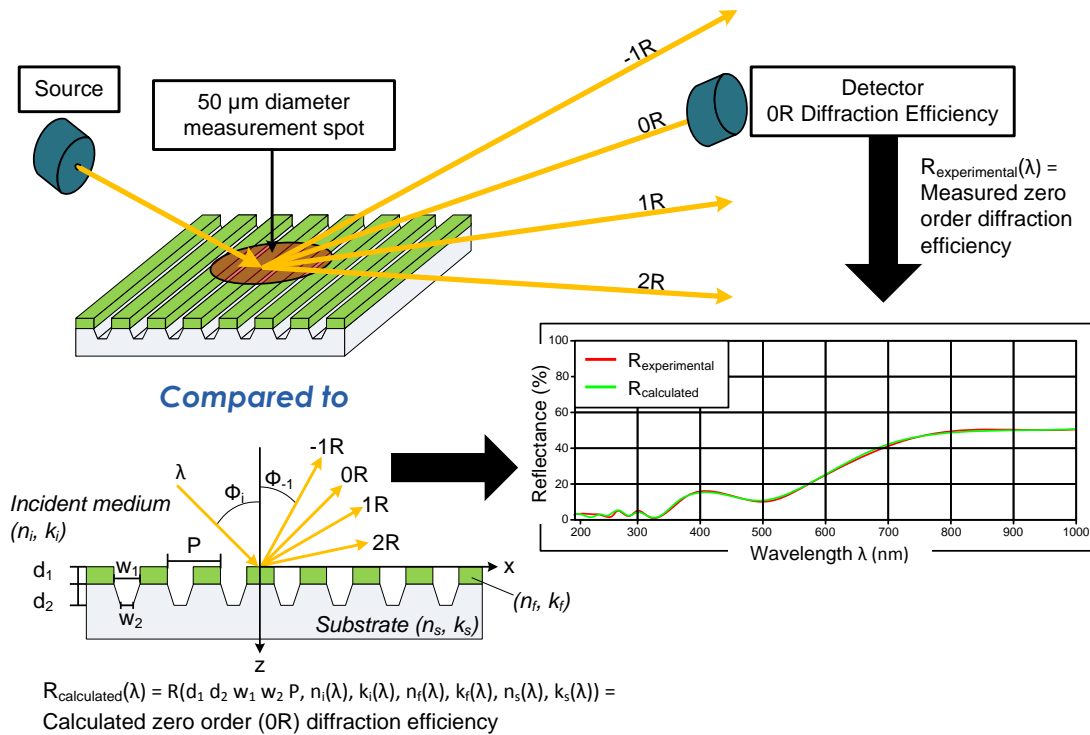
Fig. 3.8: Simulated and real measurement

4 Deposition

4.1 Plasma, the fourth aggregation state of a material

4.1.1 Plasma state

In many semiconductor manufacturing processes, a plasma is used, e.g. in sputtering, deposition or in dry etch prcesses. An important point here is that the plasma is not heated. Therefore wafers, which were already metallized, can be processed in plasma processes.

Plasma is also called the fourth state of matter or fourth aggregate state. An aggregate state is a qualitative condition of materials, which depends on the temperature and pressure. The three states solid, liquid and gaseous one encounters in everyday's life. If temperature is low, every atom in a solid is fixed at one point. Attractive forces prevent them from moving. At absolute zero (-273.15 °C) substances enter into no reaction. With increasing temperature, the particles start to oscillate, and the bonds of the atoms are unstable. If the melting point is reached, a substance transforms from the first to the second aggregate state: ice (solid) is transformed into water (liquid).

The gravitational forces in liquids are still present, but the particles are able to move and have no fixed places like in solid state, the particles are adapted for example to a predetermined shape. If the temperature increases further, the bonds are completely broken, the particles move independently of one another. At the boiling point of a substance it transforms from second to third state: water (liquid) turns into water vapor (gaseous).

While the volume of solids and liquids is constant, gaseous substances take the existing space completely, the particles are distributed evenly throughout the room.

Each substance has a specific melting and boiling point. Silicon melts at 1414 °C and passes into gaseous state at about 2900 °C. If one introduces even more energy to a sub-

stance, the collisions between the particles will strike out electrons from the outermost electron shells. Now there are free electrons and positively charged ions: the plasma state is reached.

4.1.2 Plasma generation

A plasma in semiconductor technology is usually generated by high frequency voltage, for example, argon serves as a gas. The gas is located in a high-frequency field between two charged plates (electrodes) and here it is ionized. Electrons are necessary to strike out electrons from the argon atom's outer shells. These initial electrons can be generated in different ways:

- Electrons are emitted from a thermionic cathode

- By a very high voltage electrons can be pulled out from the negative electrode

- In each gas there always are temporarily free electrons by collisions of the particles

Since the electrons are much lighter than the ions, they are immediately attracted to the positively charged electrode, and the heavy ions moving slowly to the negative electrode. Before they will achieve it, however, the polarity of the electrodes is reversed, the electrons are drawn to the other electrode and on their trajectory they will strike out more electrons from the atoms due to collisions. Typical frequencies for the plasma generation process is 13.56 megahertz and 2.45 gigahertz, so the voltage across the electrodes will be reversed 13.56 million or 2.45 billion times per second.

The electrons are located mainly on the electrodes, while in between the positively charged ions, the plasma, oscillate back and forth, because they can not follow the rapid voltage changes.

The plasma production takes place under vacuum, the produced plasma is not heated, which is important for many processes. The plasma can be used in deposition, sputtering, etching or ion implantation. Due to the rapid oscillations of the positive ions in the high-frequency field they are very energetic. There are not only positive ions and free electrons in the plasma, as other particles are created by collisions: the condition of the plasma changes constantly. Electrons are captured by the ions partially and ejected again, these additional particles, however, do not play a matter in the further use of the plasma. The degree of ionization is 0.001-10 %, depending on the particle density
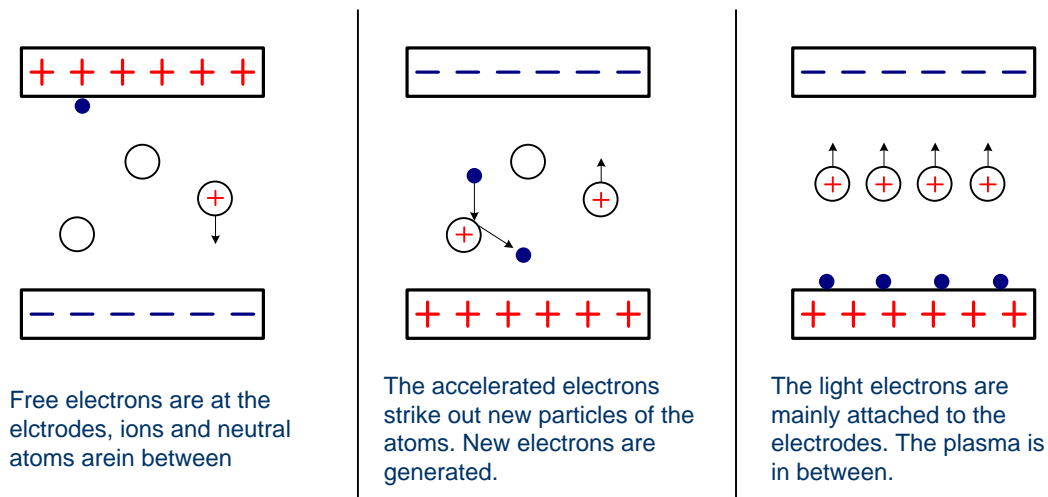
Fig. 4.1: Illustration of plasma generation

in the process chamber ($10^8 - 10^{12}$ particles per cm^3); so the majority of the particles is unloaded.

4.2 Chemical vapor deposition

4.2.1 Silicon vapor phase epitaxy

Epitaxy means "on top" or "assigned to", and represents a process in which a layer is created on top of another layer and inherits its crystal structure. If the deposited layer is of the same material as the substrate one speaks of homoepitaxy, if it's another material it's so-called heteroepitaxy. The most significant process in the homoepitaxy is the deposition of silicon on silicon, in heteroepitaxy usually a silicon layer is deposited on an insulator such as oxide (Silicon On Insulator: SOI).

Homoepitaxy:
Depending on the process, the wafers can be delivered from the wafer manufacturer with an epitaxial layer (e.g. for CMOS technology), or the chip manufacturer has to make it himself (for example in the bipolar technology).

As a gas for generating the epitactical layer, pure hydrogen is used in conjunction with silane (SiH_4), dichlorosilane (SiH_2Cl_2) or silicon tetrachloride ($SiCl_4$). At about 1000 °C, the gases cleave off the silicon, which is deposited on the wafer surface. The silicon

inherits the structure of the substrate and is growing, for energy reasons, layer by layer successively on. To not grow up a polycrystalline silicon, one must always prevail a shortage of silicon atoms, e.g. it is always slightly less silicon available as material could actually grow up. When silicon tetrachloride is used, the reaction proceeds in two steps:

$$SiCl_4 + H_2 \longleftrightarrow SiCl_2 + 2\,HCl$$
$$2\,SiCl_2 \longleftrightarrow Si + SiCl_4$$
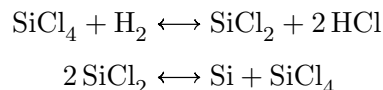
In order to inherit the substrate's orientation the surface must be absolutely clear. So one can utilize the equilibrium reaction. Both reactions can occur in the other direction, depending on the ratio of the gases. If there is only few hydrogen in the atmosphere, as in the trichlorosilane process for the purification of raw silicon, material is removed from the silicon wafer surface due to the high chlorine concentration. Only with increasing concentration of hydrogen growth is achieved.

With $SiCl_4$ the deposition rate is approximately 1 to 2 microns per minute. Since the monocrystalline silicon grows only on the bare surface, certain areas can be masked with oxide where the silicon grows as polycrystalline silicon. This polysilicon, however, is etched very easily compared to single-crystalline silicon through the backward-running reaction. Diborane (B_2H_6) or phosphine (PH_3) are added to the process gases, to create doped layers, since the doping gases decompose at high temperatures and the dopants are incorporated in the crystal lattice.

The process to create home-epitactical layers is realized under vacuum atmosphere. Therefor the process chamber is heated to 1200 °C to remove the native oxide, which is always present on the silicon surface. As mentioned above, due to a low hydrogen concentration there occurs a back etch on the silicon surface. This can be used to clean the surface before the actual process starts. If the gas concentration is varied post this cleaning the deposition begins.

Due to the high process temperatures there's a diffusion of dopants in the substrate or impurities, which have been used in earlier processes, can move to the substrate. If SiH_2Cl_2 or SiH_4 are used there's no need for such high temperatures, so these gases are used primarily. To achieve the etch back process to clean the surface, HCl has to be added separately. The disadvantage of this silanes is that they form germs in the
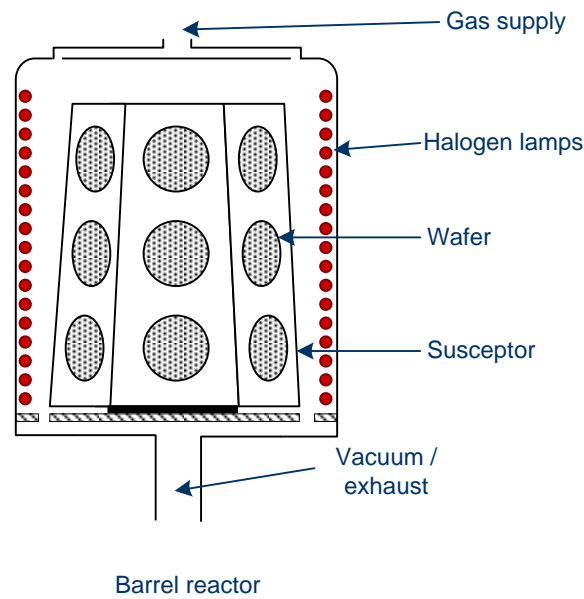
<div align="center">Barrel reactor</div>

Fig. 4.2: Illustration of a barrel reactor for epitactical processes

atmosphere right before deposition, and thus the quality of the layer is not as good as with $SiCl_4$.

4.2.2 CVD process: Chemical Vapor Deposition

There is often a need of layers which can't be created right from the substrate. To deposit layers of silicon nitride or silicon oxynitride one has to use gases which contain all necessary components. The gases are decomposed via thermal energy. That's the principle of the chemical vapor phase deposition: CVD. The wafer surface doesn't react with the gases but serves as bottom layer. Depending on the process parameters - pressure, temperature - the CVD method can be devided in different methods whose layers differ in density and coverage. If the growth on horizontal surfaces is as high as on vertical surfaces the deposition is conform.

The conformity K is the ratio of vertical and horizontal growth, $K = R_v/R_h$. If the deposition is not ideal, the conformity is less than 1 (eg $R_v/R_h = 1/2 \Rightarrow K = 0.5$). A high conformity can only be achieved by high process temperatures.
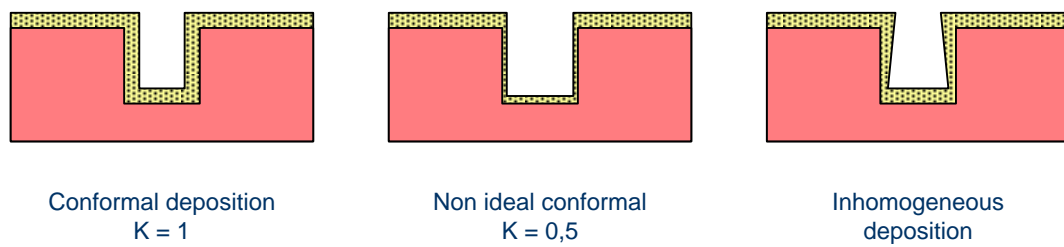
Fig. 4.3: Deposition profiles with different conformity

4.2.3 APCVD: Atmospheric Pressure CVD

APCVD is a CVD method at normal pressure (atmospheric pressure) which is used for deposition of doped and undoped oxides. The deposited oxide has a low density and the coverage is moderate due to a relatively low temperature. Because of improved tools, the APCVD undergoes a renaissance. The high wafer throughput is a big advantage of this process.

As process gases silane SiH_4 (highly deluted with nitrogen N_2) and oxygen O_2 are used. The gases are decomposed thermal at about 400 °C and react with each other to form the desired film.

$$SiH_4 + O_2 \longrightarrow SiO_2 + 2\,H_2 (T = 430\,°C, p = 105\,Pa)$$
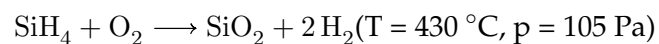
Added ozone O_3 can cause a better conformity because it improves the movability of the accumulated particles. The oxide is porous and electrical instable and can be densified by a high temperature process.

To avoid edges which can result in difficulties at the deposition of additional layers, phosphorus silicate glass (PSG) is used for interlayers. Therefor phosphine is added to SiH_4 and O_2, so that the deposited oxide contains 4 to 8 % phosphorus. A high amount of phosphorus leads to a high increase of the flow properties, however, phosphoric acid can be formed which corrodes aluminum (conductor paths).

Because annealing affects earlier processes (e.g. doping) only short tempering is done with powerful argon lamps (several hundrets kW, less than 10 s, T = 1100 °C) instead of annealing in longsome furnace processes. Analog to PSG boron can be added simul-

taneously (boron phosphorus silicate glass, BPSG, 4 % B and 4 % P).
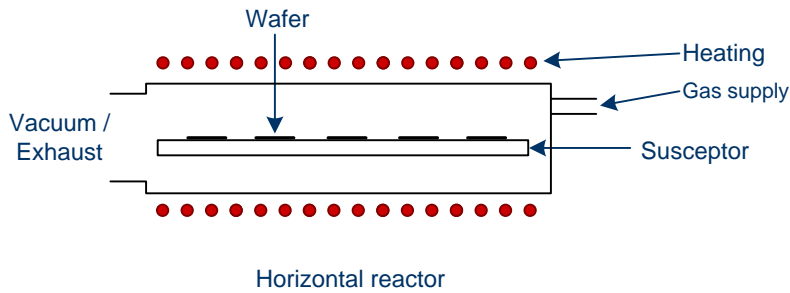
Fig. 4.4: Illustration of a horizontal APCVD reactor

4.2.4 LPCVD: Low Pressure CVD

In LPCVD a vacuum is used. Thin films of silicon nitride (Si_3N_4), silicon oxynitride (SiON), SiO_2 und tungsten (W) can be created. LPCVD processes enable a high conformity of almost 1. This is because of the low pressure of 10 to 100 Pa (atmospheric pressure = 100.000 Pa) which leads to a non-uniform movement of the particles. The particles dispread due to collisions and cover vertical surfaces as well as horizontal ones. The conformity is supported by a high temperature of up to 900 °C. Compared to APCVD the density and stability is very high.

The reactions for Si_3N_4, SiON, SiO_2 and tungsten are as follows:

- **Si_3N_4 (850 °C):** $4\,NH_3 + 3\,SiH_2Cl_2 \longrightarrow Si_3N_4 + 6\,HCl + 6\,H_2$

- **SiON (900 °C):** $NH_3 + SiH_2Cl_2 + N_2O \longrightarrow Si_3N_4 + \text{byproducts}$

- **SiO_2 (700 °C):** $SiO_4C_8H_{20} \longrightarrow SiO_2 + \text{byproducts}$

- **Tungsten (400 °C):** $WF_6 + 3\,H \longrightarrow 2\,W + 6\,HF$

In contrast to gaseous precursors which are used for Si_3N_4, SiON and tungsten, liquid tetraethyl orthosilicate is used for SiO_2. Besides there are other liquid sources like DTBS ($SiH_2C_8H_{20}$) or tetramethylcyclotetrasiloxane (TMTCS, $Si_4O_4C_4H_{16}$).

A tungsten film can only be fabricated on bare silicon. Therefore silane has to be added if there is no silicon substrate.
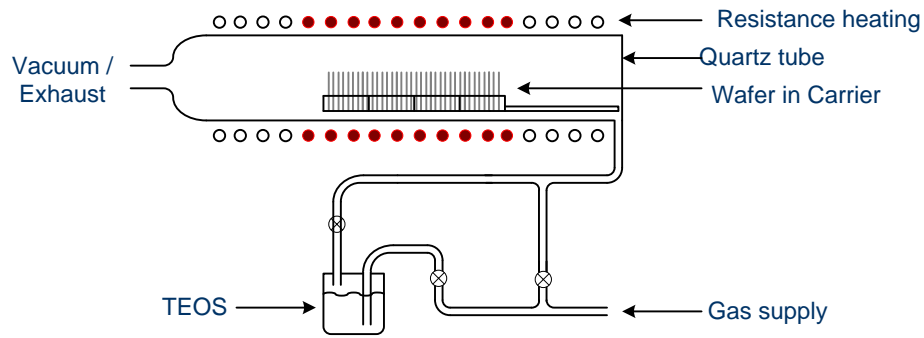
Fig. 4.5: Illustration of a LPCVD reactor for TEOS films

4.2.5 PECVD: Plasma Enhanced CVD

The PECVD takes place at 250 to 350 °C. Due to low temperatures the process gases can not be decomposited thermal. With a high frequency voltage, the gas is transformed into a plasma state. The plasma is energetic and disposes on the surface. Because metallization, such as aluminum, can not be exposed to high temperatures, the PECVD is used for SiO_2 and Si_3N_4 deposition on top of metal layers. Instead of SiH_2Cl_2 silane is used because it decomposes at lower temperature. The conformity is not as good as in LPCVD (0.6 to 0.8), however, the deposition rate is much higher (0.5 microns per minute).
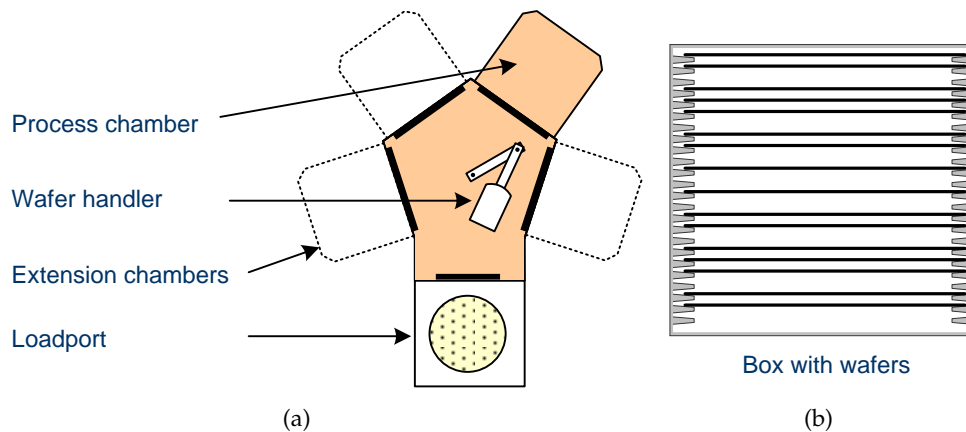
Fig. 4.6: Illustration of a PECVD reactor.

4.2.6 ALD: Atomic Layer Deposition

Atomic Layer Deposition (ALD) is a modified CVD process to manufacture thin films. The process uses several gases which are led into the process chamber alternating. Each gas reacts in such a way that the current surface is saturated, and therefore the reaction comes to a standstill. The alternative gas is able to react with this surface in the same way. Between the reactions of these gases the chamber is purged with an inert gas, like nitrogen or argon. A simple ALD process could look like this:

1. self-limiting reaction at the surface with first gas

2. purging with an inert gas

3. self-limiting reaction at the surface with second gas

4. purging with an inert gas

A specific example for an ALD process is the deposition of aluminum oxide, this can be realized with trimethylaluminum (TMA, C_3H_9Al) and water (H_2O).
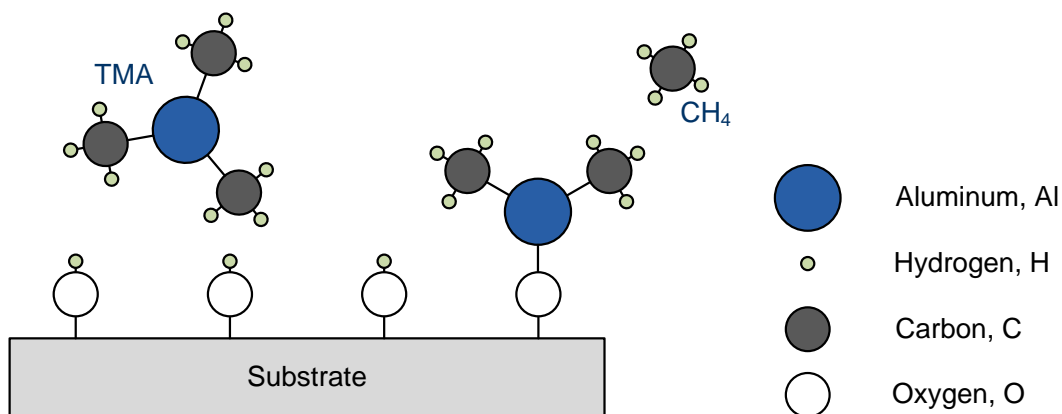
Fig. 4.7: Introduction of TMA into the process chamber

First step is the elimination of hydrogen atoms which are bound to oxygen at the wafer surface. The methyl groups (CH_3) of TMA can react with the hydrogen to form methane (CH_4). The remaining molecules bond with the unsaturated oxygen.

If these atoms are saturated, no more TMA molecules can react at the surface.

The chamber is purged and subsequent water steam is led into the chamber. Ever one hydrogen atom of the H_2O molecules now can react with the former deposited surface
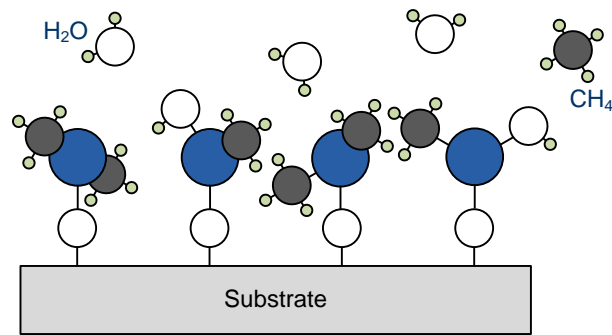
Fig. 4.8: Adhesion of TMA at the surface atoms

atoms to form methane, while the hydroxyl anion is bond to the aluminum atoms.
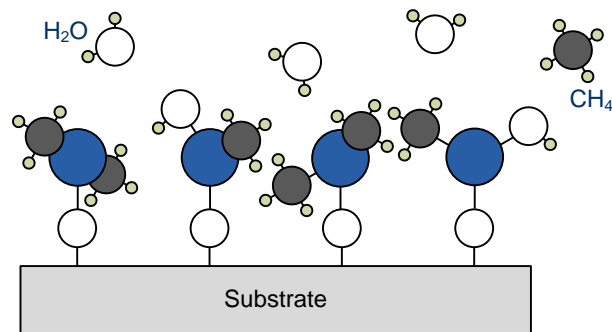
Fig. 4.9: Introduction of water into the process chamber

Hence, there are new hydrogen atoms at the surface which can react in a afterwards step with TMA like in the beginning.

The atomic layer deposition provides significant advantages over other deposition techniques, and therefore it's a very important process to manufacture thin films. With ALD even 3-dimensional structures can be deposited very uniform. Insulating films are possible as well as conductive ones, which can be created on differet substrates (semiconductors, polymers, ...). The film thickness can be controlled very precise by the number of cycles. Since the reactive gases are not led into the chamber simultaneously, they can not form germs right before the actual deposition. Thus the quality of the films is very high.
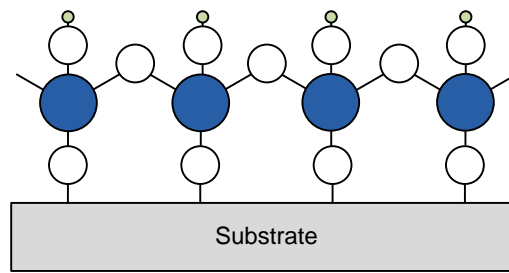
Fig. 4.10: Saturated surface at the end of one cycle

4.3 Physical deposition methods

4.3.1 Molecular beam epitaxy

The deposition characteristics are similar to the silicon vapor phase epitaxy (see CVD), but the deposition technique differs.

The process takes place under ultra-high vacuum (UHV, 10^{-8} Pa), the wafer is held upside down at the top of the chamber, the native oxide is removed at 600 to 800 °C.

With an electron beam highly purified silicon is evaporated and deposes on the wafer. Dopants can be evaporated as well and reach the surface together with the silicon. By selective temperature control and with covers the particle beam can be controlled precisely. This process allows layers of different materials whose dimensions are different and thus can't be deposited with other methods. E.g. a layer of silicon and germanium can be created which is necessary for high frequency devices or in bipolar techniques.

However, the effort to create the ultra-high vacuum is very high. To achieve this pressure, which is less than 10^{-12} of atmospheric pressure, one needs several vacuum pumps and a long time to pump down. Only one wafer can be processed simultaneously, the growth is only about 1 micron per hour.

4.3.2 Evaporating

Also metallic layers, such as aluminum, can be deposed on the wafer. The material is placed in a crucible, made of a hardly meltable metal like tantalum, and heated till evaporation. The vapor reaches the wafer in perpendicular orientation, therefore edges are not covered well, the film is polycrystalline. Alternatively the metal can be evaporated using an electron beam instead of a crucible. Compared with the thermal
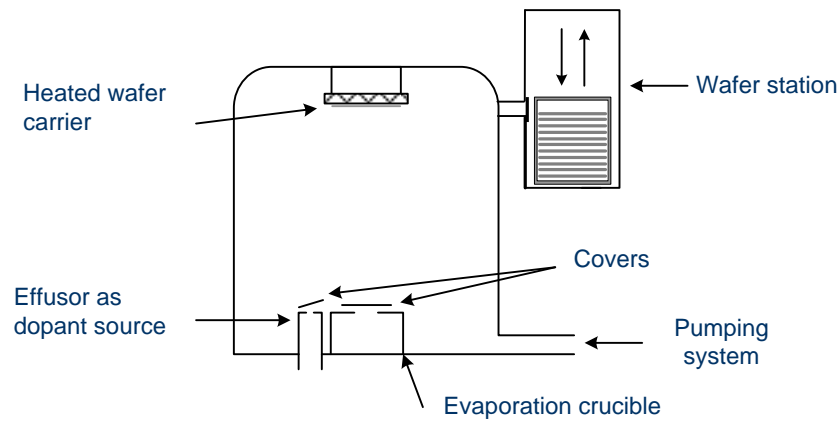
Fig. 4.11: Illustration of a MBE chamber

process, the electron beam allows a precise growth. Because of the bad edge coverage both methods are mostly used for backside deposition for final contacting.
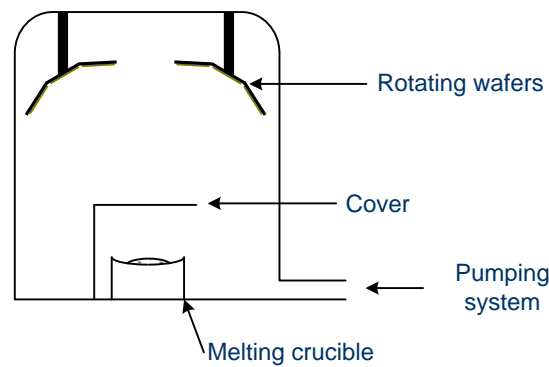
Fig. 4.12: Illustration of an evaporating chamber

4.3.3 Sputtering

In sputtering ions are accelerated onto a target to strike out atoms or molecules. The target consists of the material which shall be deposited. The mean free path is a few millimeters which means that the particles often collide with each other and therefore also vertical surfaces are covered well. Primarily the noble gas argon is ionized by gas discharge. The disposed particles form a porous film which can be densified by annealing. Sputtering can be devided into passive (inert) and reactive sputtering.

By passive sputtering only the material of the target is disposed on the wafers, accord-
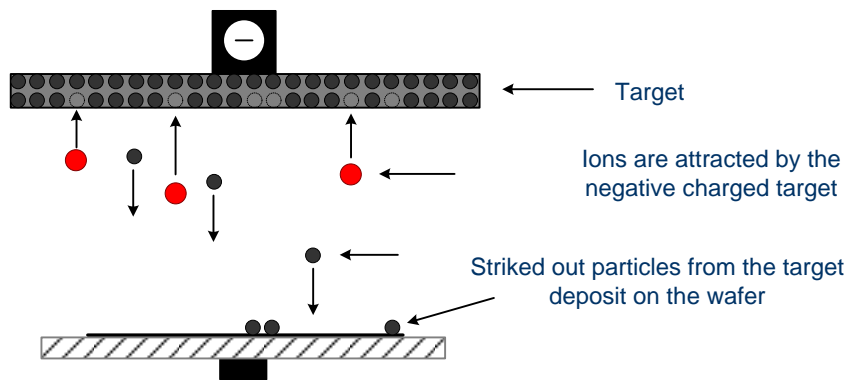
Fig. 4.13: Illustration of the sputter process

ing to the material of the target high-purity films can be created as the mix ratio of the substances in the target can be chosen precisely. In the reactive sputtering an additional reactive gas (e.g. oxygen O2) is added, which reacts with the particles of the target and deposits on the wafer. If one uses a metallic target (e.g. aluminum Al), non-metallic layers like the insulator aluminium oxide are possible:

$$4\,Al + 3\,O_2 \longrightarrow 2\,Al_2O_3$$
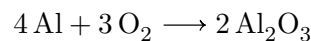
To create metallic layers, the DC sputtering is used. Thereby the ions are accelerated with up to 3 kV onto the target where they are discharged. To dissipate the charges, only a conductive target can be used. For non-metallic layers the reactive sputtering has to be used. If one wants to create an insulating layer right out of the target the RF sputtering is used.

In RF sputtering a voltage is applied to both electrodes behind both the target (cathode) and the wafer (anode). During the positive half-wave on the target, the electrons were attracted to it, thus the target gets negatively charged. The negatively charged target attracts ions which strike out particles from it. To increase the deposition rate one can attach magnets behind the target to deflect the electrons into a circular path. Thus more ions will be ionized and strike out additional particles from the target. Because the anode is connected to the process chamber its potential difference compared to the plasma is much less than the potential difference of the cathode to the plasma. Thats why the ions are accelerated to the target only and not onto the wafer.
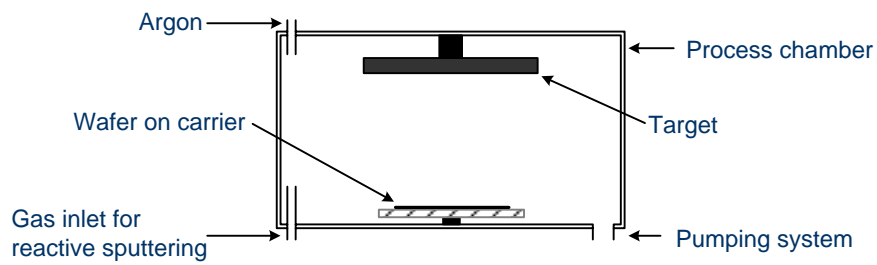
Fig. 4.14: Illustration of a sputter chamber

To increase edge coverage the BIAS sputtering is used. A negative voltage is applied to the substrate, so that here particles are striked out at well, which planishes the surface. However, one has to take care that there is no abrasion of the substrate. This so called back-etch process is the principle of most dry etch processes.

Sputtering is suitable to create metallic films with high conformity and very good reproducibility. The effort is little, the low pressure (5 Pa) can be achieved easily.

5 Metallization

5.1 Requirements on metallization

The metallization realizes contacts to the doped regions in semiconductor manufacturing with conducting paths. From here the connectors are lead to the edge of the microchip to finally connect it to the package or for testing reasons.

Following requirements are essential for metallizations in integrated micro electronic devices:

- excellent adhesion on silicon oxides (insulators between metallization layers)

- high ampacity, low resistance

- low contact resistance between metal and semiconductor

- simple processes to deposit the metallization layers

- low susceptibility to corrosion for long life times

- excellent contacting with wire bonds

- possibility to integrate stacked layers

- high purity of the metal

Aluminum meets many of this requirements and therefore has been the material of choice for many years. However, since structures are getting smaller and smaller, aluminum can't fulfill the requirements any longer. For this reason, copper will replace aluminum in the future.

5.2 Aluminum technology

5.2.1 Aluminum and aluminum alloy

Because of its properties aluminum and its alloys are widely used for wiring in microchips:

- excellent adhesion on SiO2 and interlayers as BPSG or PSG

- excellent contacting with wire bonds (ie gold and aluminum wires)

- low electrical resistance (3 $\mu\Omega \cdot cm$)

- simple to structure in dry etch processes

Aluminum fulfills the requirements in electrical toughness and resistance against corrosion only partial. Metals like silver or copper have better properties, however, these metals are more expensive and cannot be etched in dry etching this easily.

5.2.2 Diffusion in silicon

The use of pure aluminum leads to a diffusion of silicon into the metal. The semiconductor reacts with the metallization at only 200-250 °C. This diffusion of silicon causes cavities at the interface of both materials which are then filled by aluminum. Thus leads to spikes which can cause short circuits if they reach through the doped regions into the silicon crystal beneath.
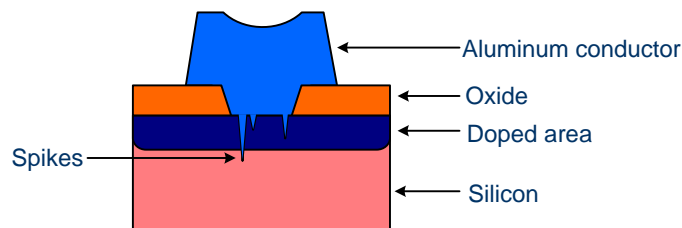
Fig. 5.1: Spikes

The size of these spikes depends on the temperature at which the aluminum was deposited onto the wafer. To avoid spikes there are several possibilities. A deep ion implantation - contact implantation - can be introduced at the location of the vias. Thus the spikes do not reach into the substrate.
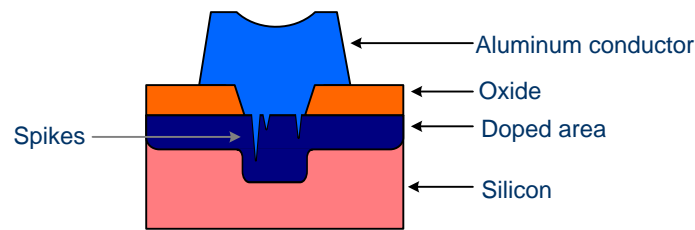
Fig. 5.2: Contact implantation

The disadvantage is that there has to be an additional process step, furthermore the electrical properties change because the doped regions are enlarged.

Instead of pure aluminum an alloy of aluminum and silicon can be used (silicon 1-2 %). Because the aluminum now already contains silicon there will be no diffusion out of the substrate. However, if the vias are very small, the silicon can drop out at the contact area and result in an increased resistance.

For contacts with a high quality a separation of aluminum and silicon is essential. A barrier of different materials (e.g. titan, titan nitride or tungsten) is deposited. To avoid an increased contact resistance at the interface of titan in silicon a thin layer of titan silicide is used.
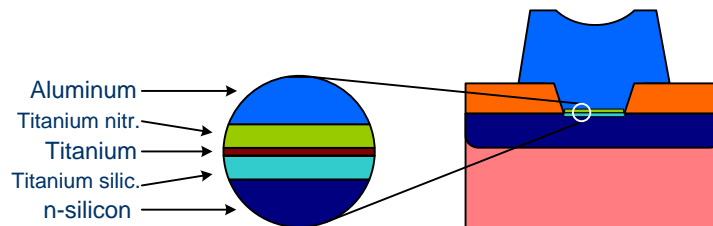
Fig. 5.3: Barrier layer between aluminum and silicon

5.2.3 Electromigration

A high current density results in a friction of electrons and fixed metal ions. The ions are moved due to collisions with the electrons. At locations with small cross sections the current density increases, thus more ions are moved and the cross section decreases which leads to a higher current density. In extreme case the aluminum wires can break.

5.2.4 Hillocks

Electromigration leads to moved material which is accumulated on locations with a lower current density. This hillocks can break through adjacent layers and cause short circuits. In addition moisture can penetrate into the material and cause corrosion. Another reason for hillocks are different coefficients of thermal expansion. Layers expand in different ways due to heating which causes stress. To minimize strain additional layers with an adjusted coefficient of expansion can be deposited between the other layers (e.g. titan, titan nitride).

Further problems which can occure during metallization:

- **Diffracted expose:** the metal layer can reflect light in that way that adjacent regions are exposed. To avoid reflections an anti reflective coating can be deposited

- **Bad edge coverage:** on edges there can be increased aluminum growth, while in corners there is a decreased growth. Therefore edges have to be rounded:
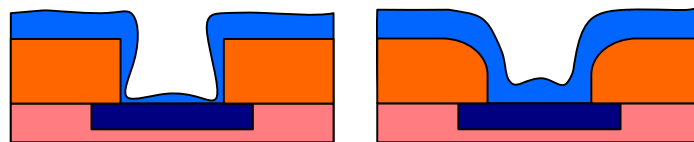
Fig. 5.4: Edge rounding

The layout of the wires has to be planned exactly to avoid these issues. A small additive of copper in the aluminum can increase the life time by far. However, the structuring of the aluminum-copper conductors is much more difficult. To avoid corrosion, the surface is sealed with layers of silicon oxide, silicon tetranitride or silicon nitride. The material of the packages for microchips is some kind of ceramics because synthetic materials are not as resistant.

5.3 Copper technology

5.3.1 Copper technology

Copper has significant advantages compared to aluminum and thus is a good alternative for ever smaller structures. The metal has a much lower resistance than aluminum and is much more efficient in view of power consumption; beacuse of the needs of

smaller and smaller feature sizes, aluminum does not fulfill the electrical requirements any longer. Also the electromigration of copper is much less than of aluminum. A change-over can not be averted.

Copper, however, has the disadvantage that it contamintes alomst everything which gets in contact to it. Therefore areas and equipment on which copper is processed, have to be separated from others. In addition, copper is susceptible for corrosion as well as aluminum and has to be covered with a passivation layer. Another advantage of copper is, that there is no need of tungsten to connect the individual copper layers with each other, and therefore additional process steps are omitted as well as thermal issues at the interface of different materials. One of the biggest disadvantages in semiconductor device fabrication is, that copper can not be structured as easy as aluminum in dry etch processes.

The traditional subtractive process for structuring - as it is used for aluminum and other materials - is done as follows:

- deposition of the layer which has to be structured

- resist deposition, exposure and development

- transfer of the resist mask into the layer beneath via dry etching

- resist removal

- passivation

In copper technology one has to use not a subtractive but an additive method: the so called damascene process.

5.3.2 Damascene process

The damascene process makes use of existent interlayer dielectrics in which the vias and trenches for conduction paths are etched. Subsequent, copper is deposited by CVD, PVD + reflow, or in electrochemical/galvanically processes. Finally, the copper is planished by chemical mechanical polishing (CMP).

The damascene process can be seperated into the single and dual damascene process and the latter can be seperated further into the VFTL (VIA First Trench Last) and the TFVL approach (Trench First VIA Last). In the following, both the TFVL as well as the VFTL process are described.

Dual damascene: Trench First VIA Last:

On top of the wafer (in this example on an existing copper layer) different layers are deposited which act as protection, isolation or passivation layers. As an etch stop and protection against gaseous molecules, silicon nitride (SiN) or silicon carbide (SiC) can be used. As interlayer dielectric (ILD), materials with a low relative static permittivity are used, like silicon dioxide SiO_2. Upon a resist mask is patterned.

1. The wafer is coated with a resist layer which is structured in photolithography (Fig. 5.5).
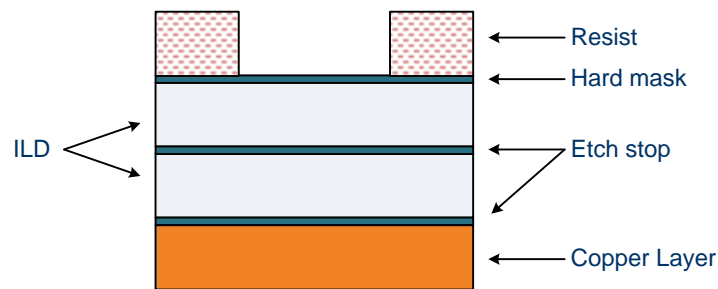
Fig. 5.5: Basic film stack of a damascene approach

2. The hardmask (SiN) and the ILD are etched in a anisotropic dry etch process until the first etch stop layer (SiN) is reached. The resist is removed and the trench for the conduction path is finished. (Fig. 5.6(a)).
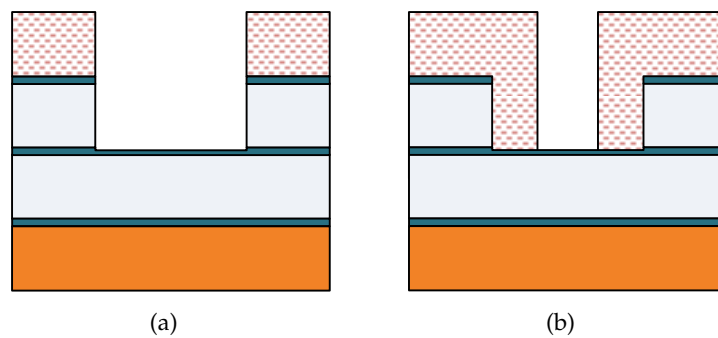
Fig. 5.6: Trench etch and via resist mask

The hardmask on top protects the ILD during the resist ash. This is necessary since the ILD has a similar composition as the resist and therefore is affected by the same process gases. In addition the hardmask acts as a barrier layer during terminal CMP.

3. Next a new resist layer is deposited and structured (Fig. 5.6(b)).

4. Finally the vias are opened in an anisotropic etch process.

With a low energetic etch process, the bottom etch stop is opened to avoid sputtering of the copper beneath which could deposit on the sidewalls and diffuse into the ILD. The resist is removed and a thin layer of tantalum is deposited as a barrier which prevents later deposited copper from diffusing into the ILD (fig. 5.7(a)).

(a)   (b)
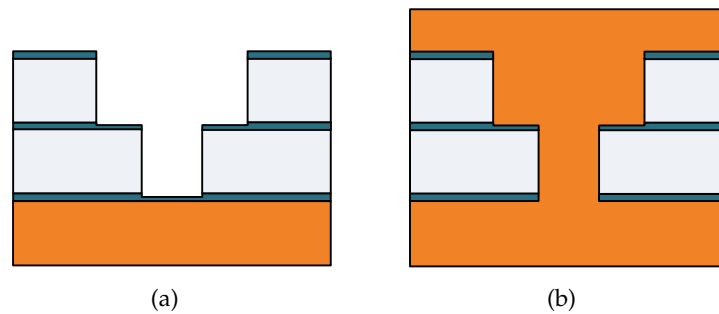
Fig. 5.7: Final structure after etch and copper deposition

5. A thin copper layer acts as a seed layer, so that the vias and trenches can be filled in a galvanically process (Fig. 5.7(b)).

6. The deposited copper is planished in a CMP process (Fig. 5.8).

Metallization (lines)

Vertical interconnect access (VIA)

Metallization (lines)
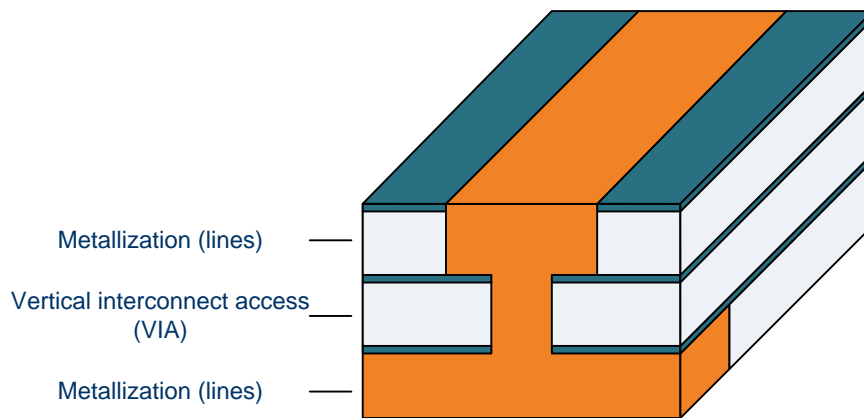
Fig. 5.8: Damascene stack after metallization and CMP

The big disadvantage of this process is the thick resist layer which is deposited after the trenches have been etched (3). To etch the tiny vias in such a thick resist layer is very difficult. For this reason the TFVL approach is done at larger structures only.

Dual damascene: VIA First Trench Last:
The VFTL approach is alike the TFVL process but now the vias are created first.

1. A resist layer to form the vias is structured and the vias transferred into the ILD by an anisotropic etch process till the bottom etch stop layer is reached. To prevent copper from being sputtered out of the metallization beneath, the etch stop must not be opened (Fig. 5.9).
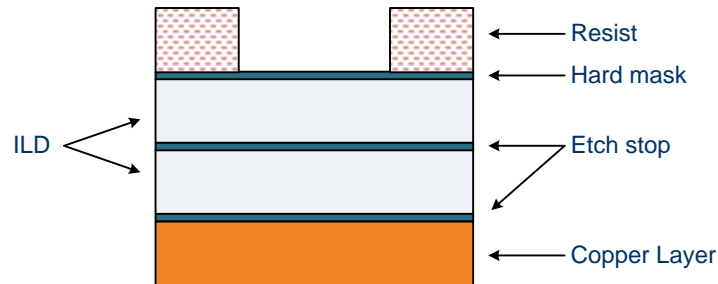
Fig. 5.9: Damascene stack after VIA etch

2. Subsequent the resist is removed, a new resist layer is patterend which represents the trenches; also the previously opened vias are filled with resist (Fig. 5.10(a)).

(a)   (b)
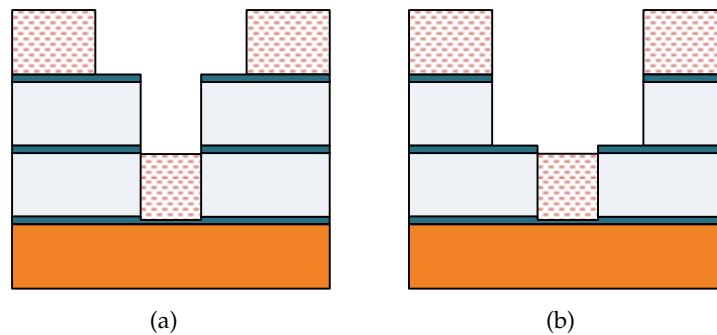
Fig. 5.10: Photolithography and trench etch

3. During the trench etch, the bottom etch stop is covered by resist. Next the bottom etch stop is opened in a low energetic process and a tantalum barrier and a copper seed layer are deposited (Fig. 5.10(b)).

4. After the copper deposition, the metal is planished in a CMP process (Fig. 5.11).

In the single damascene process, the via layer and the trench layer are deposited and structured one after each other, so that there are more process steps needed (ILD deposition $\Rightarrow$ VIA structuring $\Rightarrow$ copper deposition $\Rightarrow$ planarization $\Rightarrow$ ILD deposition $\Rightarrow$ trench structuring $\Rightarrow$ copper deposition $\Rightarrow$ planarization).
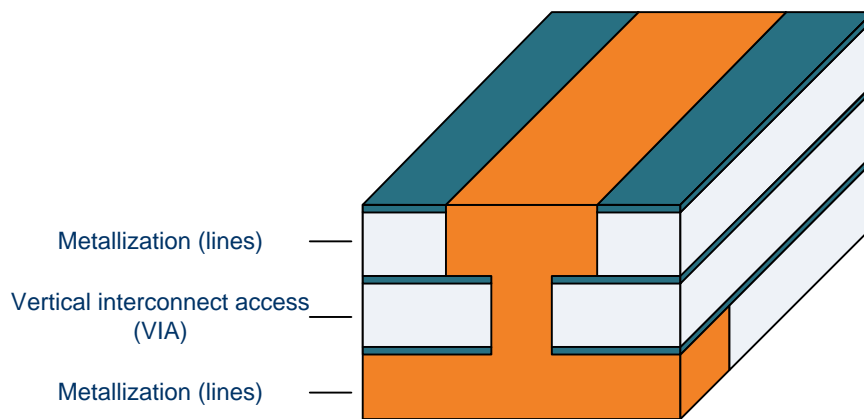
Fig. 5.11: Damascene stack after metallization and CMP

5.3.3 Low-k technology

Since there is a proceeding miniaturization of the structures on microchips to increase packing density, reduce power consumption, and increase switching speeds, the conductors for wiring are moving closer and closer together in vertical and horizontal direction. To isolate the conductors from each other additional films like silicon dioxide SiO_2 have to be deposited as an interlayer dielectric (ILD).

If conductors run parallel or cross each other on different layers upon another, parasitic capacities are created. The conductors represent the electrodes while the SiO_2 in-between is the dielectric.
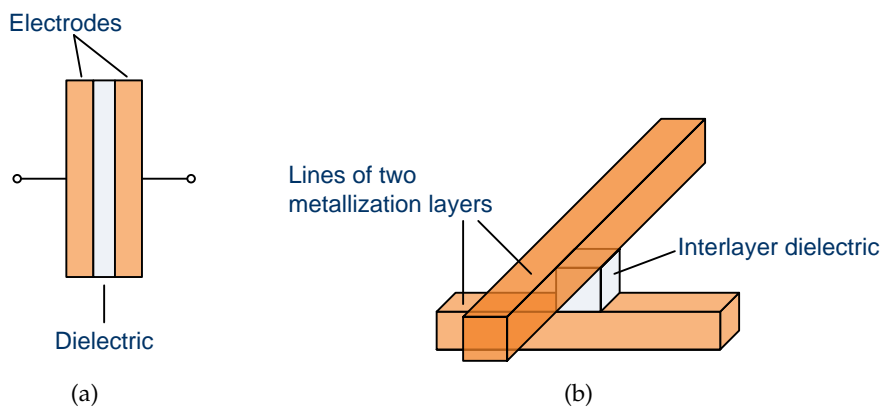
Fig. 5.12: (a) Traditional capacitor and (b) crossing of two conductors

The capacity C is given by:

$$C = \frac{e_0 e_r A}{d}$$

Where d stands for the distance of the electrodes, A is the area of the electrodes, ϵ_0 the vacuum permittivity and ϵ_r (often κ (Kappa) or simplified k) the relative static permittivity of the ILD.

The value of the parasitic capacity influences the electric properties such as the switching speed or the power consumption of a chip, and therefore one tries to decrease C. Theoretical this can be done if ϵ_0, ϵ_r and A are decreased or if d is increased. However, as mentioned above d is getting smaller and smaller, A is preset by electrical requirements and ϵ_0 is a physical constant. Thus the capacity can only be reduced by decreasing ϵ_r.

To sum it up, one needs dielectrics with a *lowϵ_r*: *low-k*.

The traditional dielectric, SiO_2, has a relative permittivity of about 4. Low-k referes to materials whose ϵ_r is less than that of silicon dioxide. Beyond that there will be Ultra-Low-k materials with an ϵ_r of less than 2.4. The permittivity referes to the polarization (dislocation of charge carriers in the insulator) in the dielectric and is the factor by which the charge of a capacitor is increased relative to vacuum or by which the electric field inside the capacitor is weakened.

To reduce the permittivity there are basically two possibilities:

- the polarizability of bonds inside the dielectric has to be decreased

- the quantity of bonds has to be reduced by introduction of porosity in the dielectric

The polarizability can be decreased by materials with less polar groups. Candidates are fluorined (FSG, ϵ_r 3.6) or organic (OSG) silicon oxides. However, for smaller and smaller structures this approach isn't sufficient, thus porous films have been introduced. By introduction of porosity there is "empty space" inside the ILD which has - in case of air - a permittivity of about 1, and therefore ϵ_r is reduced for the entire layer. The pores can be introduced by adding polymers which are expelled later by thermal annealing. In case of silicon dioxide one needs about 50 % of pores in the material to achieve a permittivity of 2. If a dielectric is used whose permittivity is 2.5 without porosity, only 22 % pores have to be introduced to achieve a permittivity of 2.
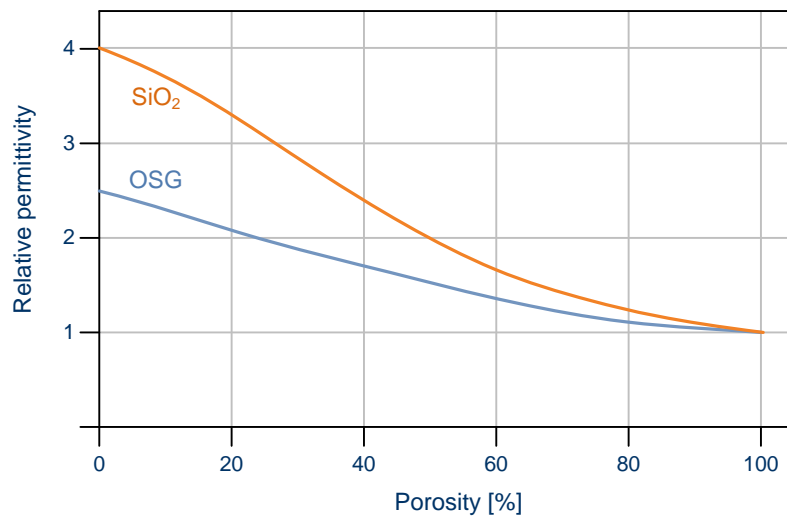
Fig. 5.13: Relative static permittivity against porosity

However, there are several issues to overcome, if one wants to bring in such new materials into semiconductor fabrication.

Due to porosity the density is reduced which results in a lower mechanical resistance. In addition process gases or copper can diffuse into the ILD and cause damage and thus increasing the permittivity or leakage. To counteract this issues, the pores have to be distributed evenly and must not be in contact with each other. To avoid a diffusion of copper, a thin barrier layer has to be deposited in a seperate process or the pores at the surface have to be closed by ion bombardment.

Like the photoresist used for manufacturing, the organic ILD is composed of hydrocarbon. If the resist is stripped in an ash process, the ILD is affected as well. To avoid this issue, additional layers (like silicon nitride as hardmask) have to be introduced in the film stack.

Chemical formula	k value
SiO_2	4,0
$SiO_{1,5}CH_3$	3,0
$SiO(CH_3)_2$	2,7
$SiO_{0,5}(CH_3)_3$	2,55

Tab. 5.1: Overview of organic silicon oxides

5.4 Metal semiconductor junction

5.4.1 Metal semiconductor junction

Subsequent to the manufacturing of transistors in the silicon substrate, the devices have to be connected to each other to realize an integrated circuit. The gate has to be contacted to control the current through the transistor, while the doped source and drain electrodes have to be contacted as well. This results in problems because of the doping. Both source and drain are impurified with dopants and thus contain additional charge carriers; electrons in n-doped crystals, holes in p-doped crystals.

Here the Fermi level is of interest. The Fermi level represents the highest energy level at which electrons can occur at absolute zero (-273,15 °C). In conductors there are electrons in the valence band and in the energetic higher conduction band, thus the Fermi level is at the level of the conduction band. For illustration: a see has a surface - Fermi energy - with water molecules - electrons - beneath it.
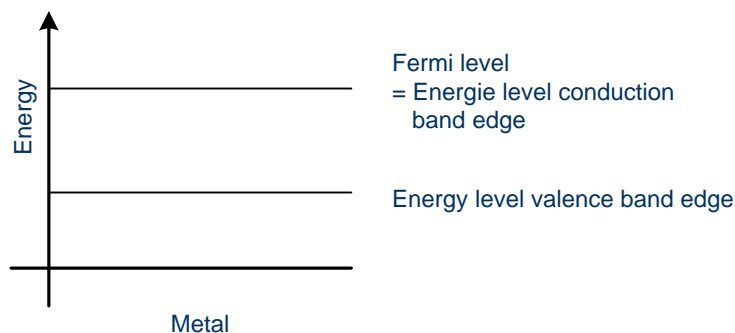
Fig. 5.14: Fermi level in metals

In doped semiconductors there are impurities as donors or acceptors in the lattice. In a p-doped semiconductor the Fermi level is next to the valence band, since electrons from the valence band can easily be lifted into the energy level of the dopant. According to this, the Fermi level in n-doped material lies next to the conduction band since electrons of donors can easily be lifted into the conduction band of the silicon crystal.

If one connects a metal to a semiconductor the Fermi levels of both materials have to equal, next to the interface the Fermi level is constant.

Because the conduction band in the semiconductor is energetic higher than the Fermi level, electrons flow into the metal since they always want to achieve the lowest energy
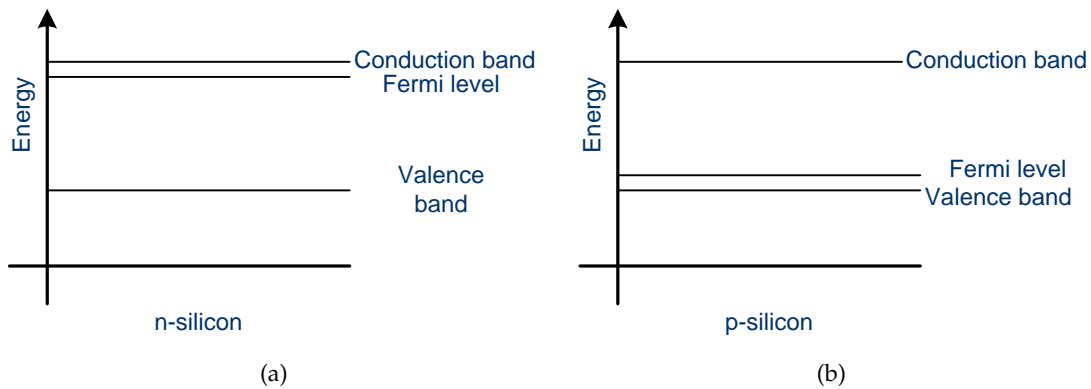
Fig. 5.15: Fermi level in doped semiconductors
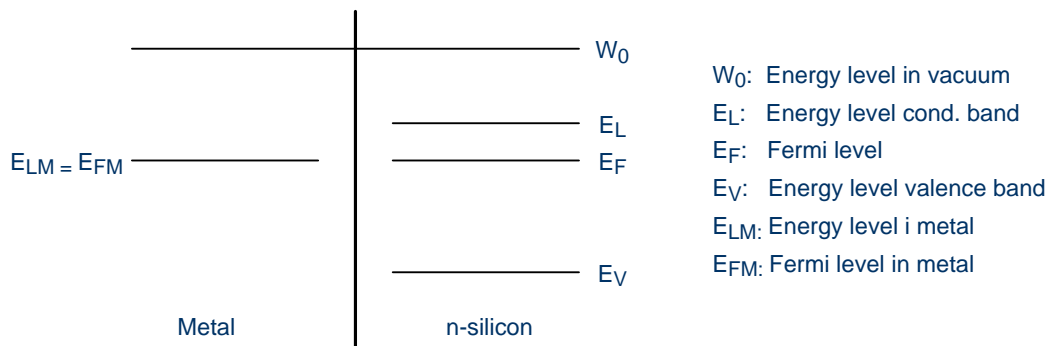
Fig. 5.16: Band model before the contact

state. Thus the probability density of electrons in the conduction band of the semiconductor decreases and therefore the distance between the conduction band and the Fermi level increases (the Fermi level represents the highest energy state at which electrons occur, and those flowed off). The electrons leave positively charged ions behind, and therefore a depletion zone remains. The bending of the energy band illustrates the potential barrier (Schottky barrier) which remaining electrons have to surpass to flow into the metal.

The width w of the depletion zone depends on the intensity of the doping. The migrated electrons lead to a negatively charged region in the metal which is limited to the surface.

This metal semiconductor junction results in a nonlinear current-voltage characteristic, a so-called Schottky diode. This barrier can be surpassed by electrons due to temperature or by tunneling due to an electric field.
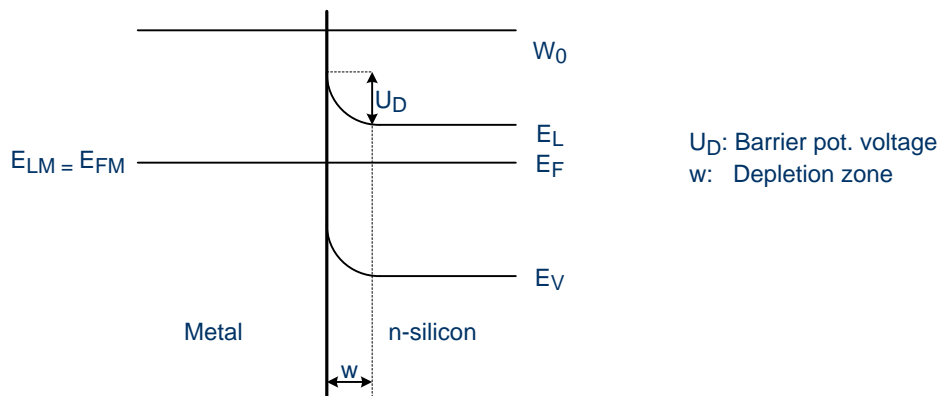
Fig. 5.17: Band model after the contact

Depending on the application this diode effect either is wanted or not. To achieve an ohmic contact (which means a contact without this potential barrier) the contact can be doped with high intensity, thus the width of the depletion zone decreases and the contact has a linear current-voltage characteristic because of tunneling.
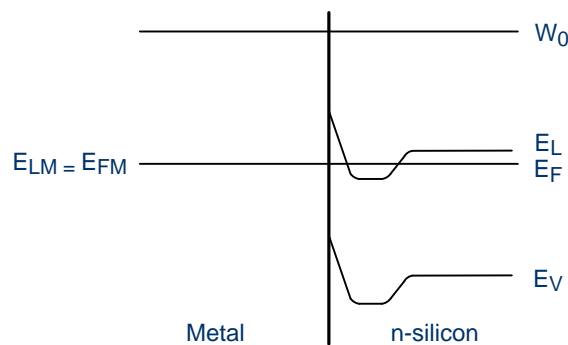
Fig. 5.18: Band model after n^+-doping

Because aluminum is integrated as an acceptor (picks up electrons) a p-doped interface occurs, leading to an ohmic contact in case of a p-doped semiconductor. In case of a n-doped semiconductor, however, aluminum leads to an inverted doping which results in a p-n junction: a diode. To avoid this there are two possibilities:

- the n-doped region is doped with such a high intensity that aluminum can only decrease the doping but not invert it

- an interlayer made of titan, chrome, or palladium prevents the inversion of the n-doped region

To enhance the contact, silicides (silicon on combination with metals) can be deposited

at the contact area.

In metal p-semiconductor junctions there is a band bending downwards, due to the exchance of charge carriers of the metal and the semiconductor. The problems mentioned above do not appear in this case because electrons from the energetic higher conduction band can flow into the metal continuously.
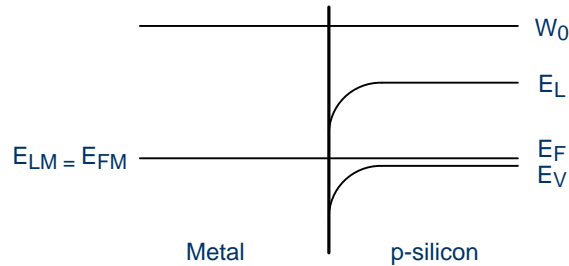
Fig. 5.19: Band model after the contact of metal and p-semiconductors

In contrast to the diode in the p-n junction, whose switching speed depends on the diffusion of electrons and holes, Schottky diodes have a very high switching speed. Thus they are suitable as protective diodes to inhibit voltage peaks.

5.4.2 Band model of p-n junctions

Because the Fermi level has to be constant, there is a band bending in p-n junctions as well. This bending illustrates the depletion zone which occurs because of the migrated charge carriers which is the potential barrier that prevents a further diffusion of electrons and holes in equilibrium (without an applied voltage). In silicon this potential barrier is about 0.7 V.
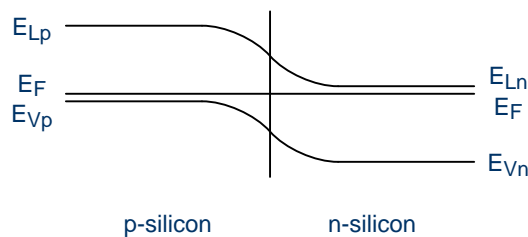
Fig. 5.20: Band model at the interface of n- and p-doped semiconductors

5.5 Wiring

5.5.1 Wiring

The wiring of an integrated circuit can take up to 80 % of the chip's surface, that's why techniques habe been developed to stack the wiring on top of the wafer in multiple layers. The amount of wires with only one additional layer can be reduced about 30 %.

Between the wires, isolation layers (oxide) are deposited, the metal layers are connected through vertical interconnect accesses (via). In today's microchips there are seven or more layers integrated. Edges and steps have to be rounded since the conformity of the metallization layers is not very good. This leads to bottlenecks in which current densities are increased so that electromigration occurs. To remove edges and steps there are several possibilities for planarization.

5.5.2 BPSG reflow

The reflow technique uses doped glasses like phosphorus silicate glass (PSG) or boron phosphorus silicate glass (BPSG). In a high temperature process the glasses melt and result in an uniform surface. Due to high temperature this technique can't be used for planarization of a metallization layer.

5.5.3 Reflow back etching

On top of the wafer a layer of silicon dioxide is deposited which is at least as thick as the highest step on the wafer. Next the oxide is coated with a resist or polyimide layer which is thermal treated for stabilization. In dry etching, the resist/polyimide and the silicon dioxide are removed with identical etch rates (selectivity of 1), thus resulting in a planished surface.

Besides the resist/polyimide, a so-called spin on glass (SOG) can be deposited on the wafer. Thus a planished layer can be produced which is stabilized during a post anneal step. An additional oxide layer is not necessary. However, all of these techniques can planish local steps only and are not sufficient for total leveling.
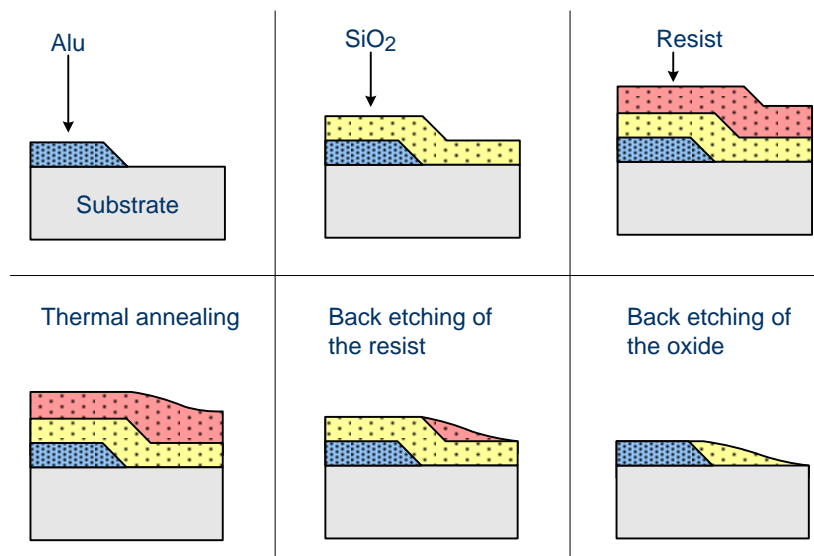
Fig. 5.21: Reflow back etching process

5.5.4 Chemical mechanical polishing

The chemical mechanical polishing/planarization (CMP) provides an uniform surface of the entire wafer. For this, an oxide is deposited on the wafer which is as thick as the highest step. The wafer is held upside down and pressed onto a polish plate. The wafer as well as the plate rotate in opposite directions and also move in horizontal directions. To support the process a slurry is used which contains abrasives and chemicals.
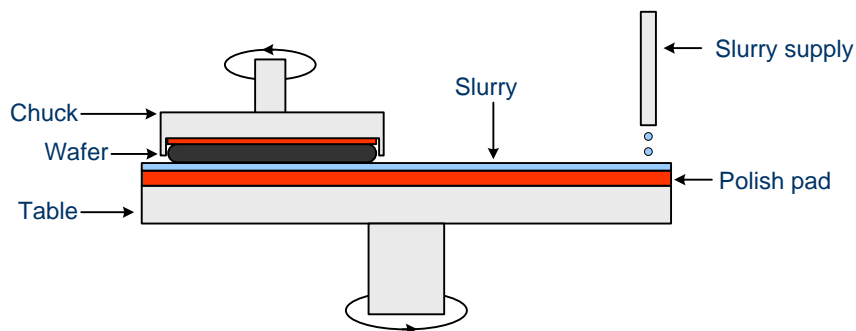
Fig. 5.22: Illustration of a polisher

Even if this process seems to be very rough it allows a surface which has an irregularity of only a few nanometers and thus is the optimal process for planarization.

5.5.5 Contacting

To contact the metallization layers, vias are etched into the isolation layers with high anisotropy. The vias have to be filled in this way, that an optimal contact is realized and the surface is not affected in a bad way.

For filling the vias, tungsten is the material of choice. With silane as additive a thin layer of tungsten is deposited as a seed layer in a CVD process with tungsten hexafluoride; byproducts as silicon tetrafluoride and hydrogen fluoride are exhausted:

$$4\,\mathrm{WF_6} + 3\,\mathrm{SiH_4} \longrightarrow 4\,\mathrm{W} + 3\,\mathrm{SiF_4} + 12\,\mathrm{HF}$$
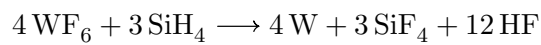
With hydrogen as an additive to the tungsten hexafluoride the vias are filled thereafter:

$$\mathrm{WF_6} + 3\,\mathrm{H_2} \longrightarrow \mathrm{W} + 6\,\mathrm{HF}$$
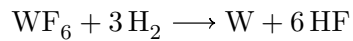
On top if it the next metallization layer can be deposited, structured and planarized. If copper is used for wiring, tungsten will only be needed for the contact to the silicon substrate. The connection of the individual copper layers is done with copper itself.

6 Photolithographie

6.1 Exposure and resist coating

6.1.1 Overview

In the manufacturing of semiconductors, structures are created on wafers by means of lithographic methods. A light sensitive film, primarily a resist layer, is coated on top of the wafer, patterned, and transfered into the layer beneath.

Photolithography consists the following process steps:

- adding adhesives and removing moisture from the surface

- resist coating

- stabilization of the resist layer

- exposure

- development of the resist

- curing of the resist

- inspection

In some processes, as the ion implantation, the resist serves as a mask to cover certain areas which should not be doped. In this case there is no transfer of the patterned resist layer into the layer beneath.

6.1.2 Adhesives

First of all the wafers are cleaned and anneald (pre-bake) to remove adhesive particles and adsorbed moisture. The wafer surface is hydrophilic and has to be hydrophobic before deposition of the photoresist. For this reason adhesives, hexamethyldisilazane

(HMDS) in general, are added to the surface. The wafers are exposed to the vapor of this liquid and are dampened.

Because of moisture in the atmosphere even after the pre-bake there are hydrogen H oder hydroxyl groups OH$^-$ attached to the surface. The HMDS decomposits into trimethylsilyl groups Si-3CH$_3$ and removes the hydrogen by forming ammonia NH$_3$.
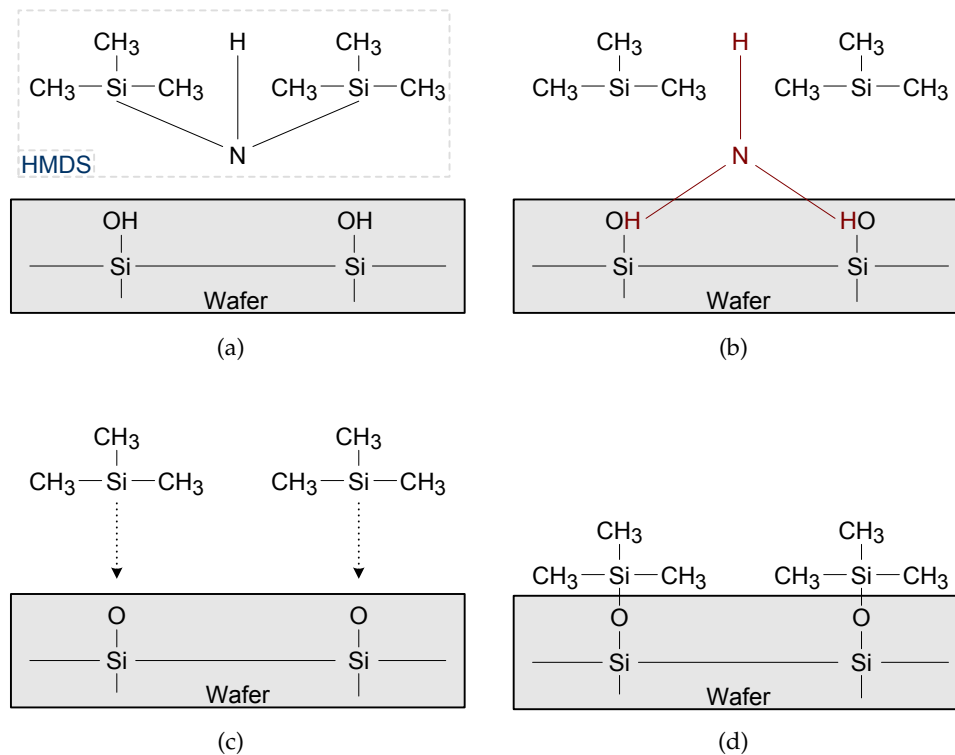
Fig. 6.1: Surface modifictation with HMDS

6.1.3 Coating

The coating of the wafer is done by spin on methods on a rotating chuck. On low rotation the resist is spun on and then planished at for example 2000 to 6000 rpm. Depending on the subsequent process the thickniss of the resist layer can be up to 2 microns. The thickness depends on the rpm and the viscosity of the resist.

To enable a homogeneous layer, the resist contains water and solvents which soften it. For stabilization reasons the wafer is annealed afterwards at about 100 °C (post-/soft-bake). Water and solvents are vaporized partially, some moisture has to remain
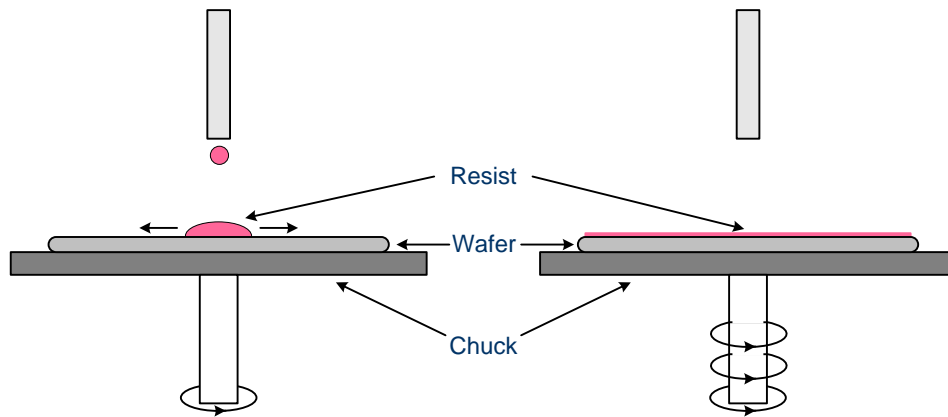
Fig. 6.2: Spin on coating

for subsequent exposure.

6.1.4 Exposure

In a lithographic exposure tool, there is a glass mask which is fractional covered with chrome to partial expose areas of the resist.
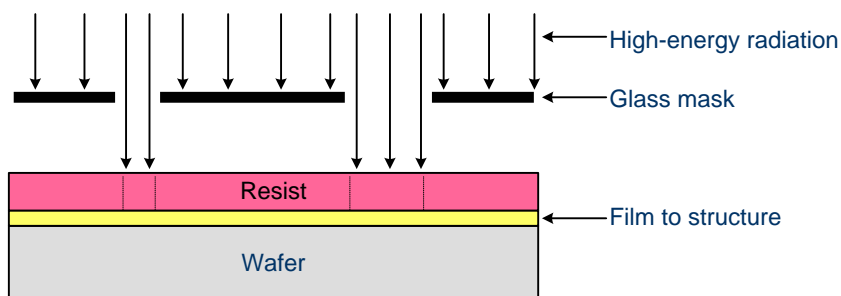
Fig. 6.3: Optical lithography

Depending on the type of the resist, exposured areas are solubly or insolubly. With a wet-chemical developer the solubly parts are removed, so that a patterned resist layer remains. The exposure time is a very important value to achieve the correct dimensions of the structures. The longer the wafers are exposed to the radiation, the larger the radiated area is. Due to fluctuating ambient temperatures a precise determination of the correct exposure time has to be investigated with one or more dummy wafers, because the characteristics of the resist can change with temperature.

An overexposure causes smaller resist patterns, and therefore smaller structures be-

neath, in contrast vias will be enlarged. With a too short exposure time the vias are not opened correctly, conductors are to wide or even in contact to each other (short circuit). In addition, a bad focusing leads to unexposed areas, so that vias can not be opened and conductors are in contact as well.

| Correct exposure | Bad focus:

Resist remains between lines; holes are not opened | Underexposure:

Lines are too wide or connected to each other; Holes are not opened | Overexposure:

Lines are too thin or disappear; holes are too big |
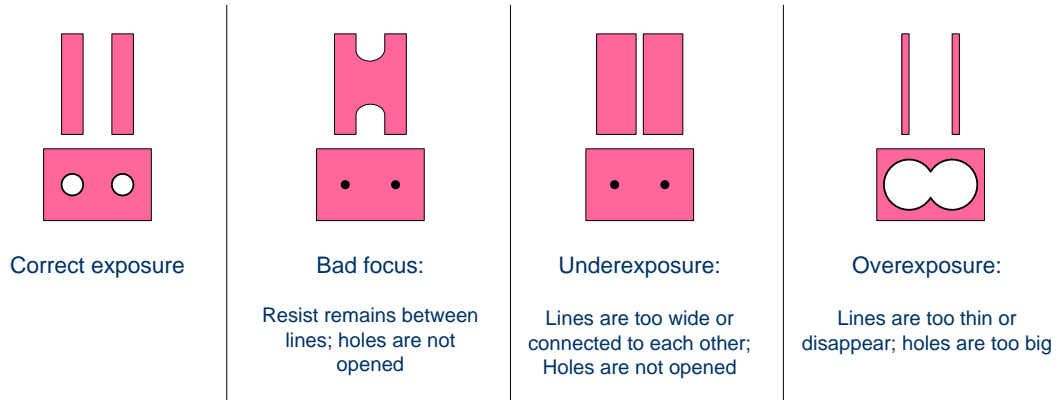
Fig. 6.4: Bad exposure due to focus issues, over exposure, or under exposure

Depending on the subsequent process, the width of the resist patterns or the diameter of the vias, respectively, has to be adjusted. In isotropic etch processes (etching in vertical and horizontal orientation) the resist mask is not transferred 1:1 into the layer beneath.

6.1.5 Exposure methods

For exposure there are different radiation sources, depending on the demands: ultraviolet radiation, electron beam, x-ray, and ion beams. The shorter the wavelength the smaller the possible structures.

For structures of 120 nm an argon fluoride laser is used, for ever smaller structures a nitrogen laser (wavelength 157 nm) or extreme ultra-violet radiation (EUV, wavelength 13 nm) is conceivable. X-rays do have a wavelength of 0.2 to 0.4 nm, electron beam writers about 0.02 nm and ion beams - in case of protons (hydrogen ions) - 0.0001 nm.

That one can fabricate structures with a width of less than x nm with a radiation of x nm is possible due to special photomasks which utilize the phase shift of light. In addition the resolution can be increased with liquid films inside the optical system (immersion lithography). These techniques made it possible to fabricate today's structure sizes of 32 nm still with a radiation wavelength of 193 nm.

While ultra-violet radiation (generated with mercury bulbs) and gas lasers are used for the exposure of the wafer, x-ray or ion beam lithography is used in research. Electron beams are used for photomask manufacturing.

6.2 Exposition methods

6.2.1 Overview

There are different types of lithographic methods, depending on the radiation used for exposure: optical lithography (photolithography), electron beam lithography, x-ray lithography and ion beam lithography.

In optical lithography patterned photomasks (reticles) with partial opaque and partial translucent areas are used. The exposure with ultra-violet radiation or gas lasers is done in a scale of 1:1 or reducing in a scale of 4:1 or 10:1 for instance.

6.2.2 Contact exposure

Contact exposure is the oldest used method. The photomask is in direct contact to the resist layer, the structures are transfered in a scale of 1:1. Thus disruptive scattering or diffraction effects only appear at the edges of the structures. This method allows only moderate feature sizes. Because all chips are exposed simultaneously the wafer throughput is very high, the construction of the lithographic unit is simple.

However, the disadvantages are obvious: the mask is contaminated because of its contact to the resist and can be scratched as well as the resist layer can be damaged. If there are particles between the mask and the resist, the optical imaging is degraded.

6.2.3 Proximity exposure

In proximity exposure there is no direct contact of the photomask and the resist. Thus only a shadow image is projected onto the wafer which results in a much worse resolution of the structures, contact issues are prevented therefore.
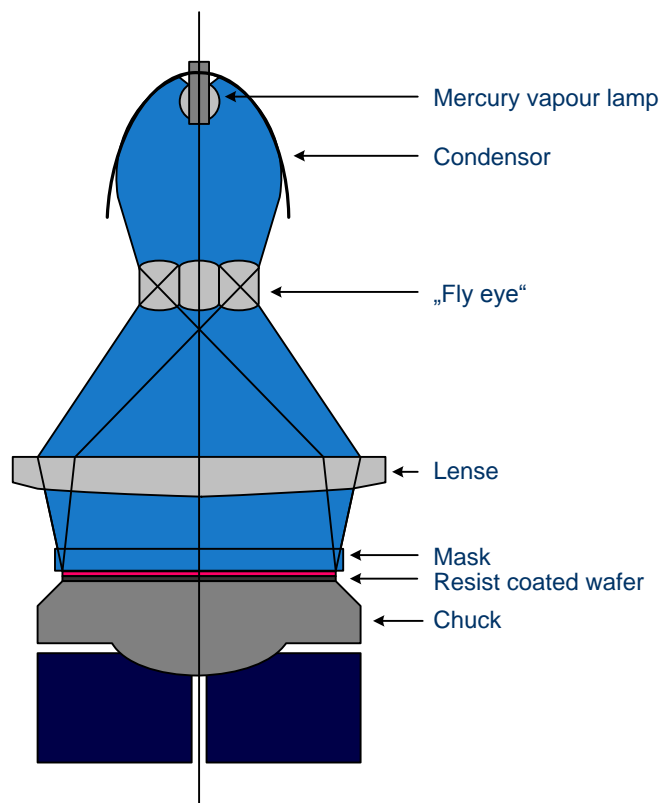
Fig. 6.5: Contact exposure

6.2.4 Projection

The exposure via projection uses the so-called step-and-repeat technique. Thereby only one or a few dies were projected onto the wafer at a time. The entire wafer is exposed step by step - die by die.

The advantage of this method is that the structures on the reticle are enlarged 4-fold or 10-fold. If the structures are projected onto the wafer in reduced scale, also defects, like particles, are reduced. In contrast to other exposure methods the resolution is improved.

Fig. 6.6: Proximity exposure
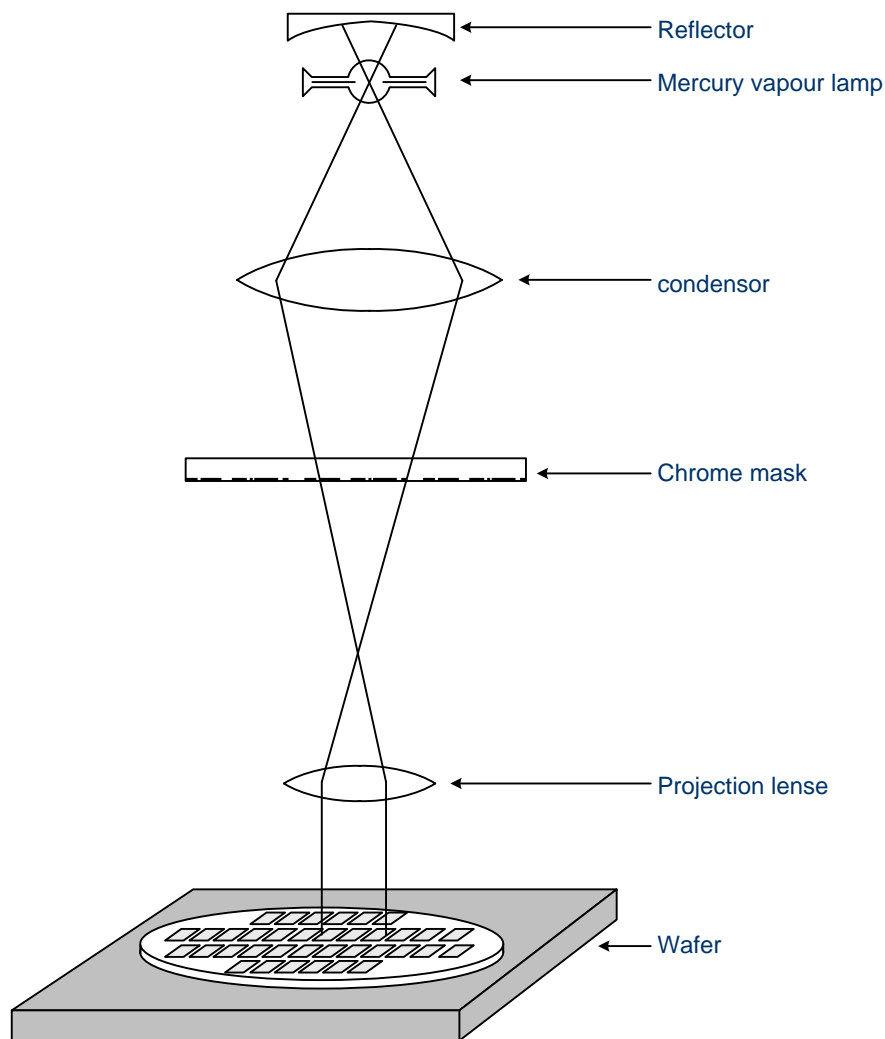
Fig. 6.7: Step and repeat technique

In addition a thin foil - pellicle - is attached to the mask, and therefore particles will be held off the mask and are out of focus during projection.

Besides the projection with lenses also a projection with a complex system of mirrors can be used (scale 1:1). Compared to lenses there are no color aberrations and thermal expansion of the photomask can be adjusted. However, mirror images can be distorted or warped. The resolution is limited due to the scale of 1:1.

6.2.5 Electron beam lithography

As in photomask manufacturing a focused ion beam is scanned across the wafer, the scan is done line by line. Each structure has to be written one after the other which takes a lot of time. The advantage is that there is no need of photomasks which reduces the costs. The process is done under a vacuum.
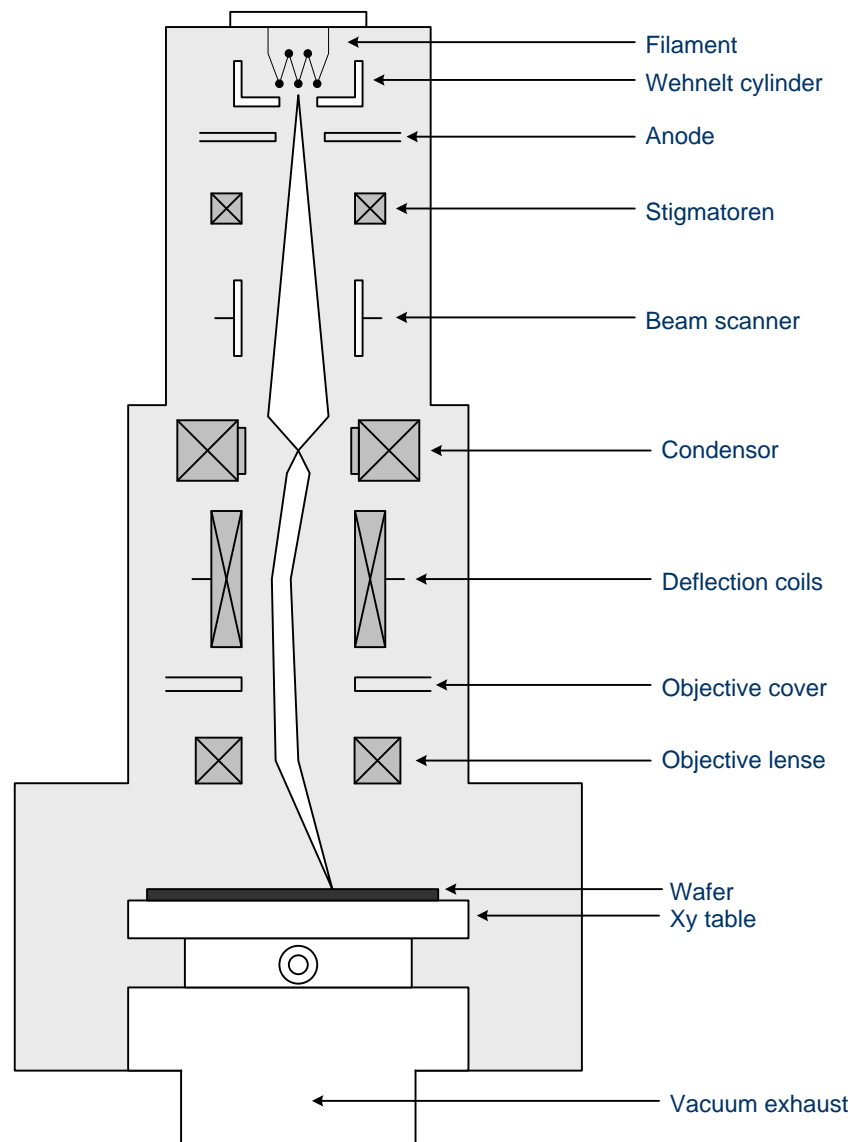
Fig. 6.8: Electron beam lithography

6.2.6 X-ray lithography

The resolution of x-ray lithography is about 40 nm. The imaging is done in step-and-repeat technique in a scale of 1:1 in atmospheric pressure or low pressure in helium atmosphere (about 10.000 Pa). The x-ray source can be a plasma or a synchrotron.

Instead of chrome coated glass masks, thin foils made of beryllium or silicon are used. To adsorb x-rays heavy elements like gold are necessary. The facility as the photomasks are very expensive.
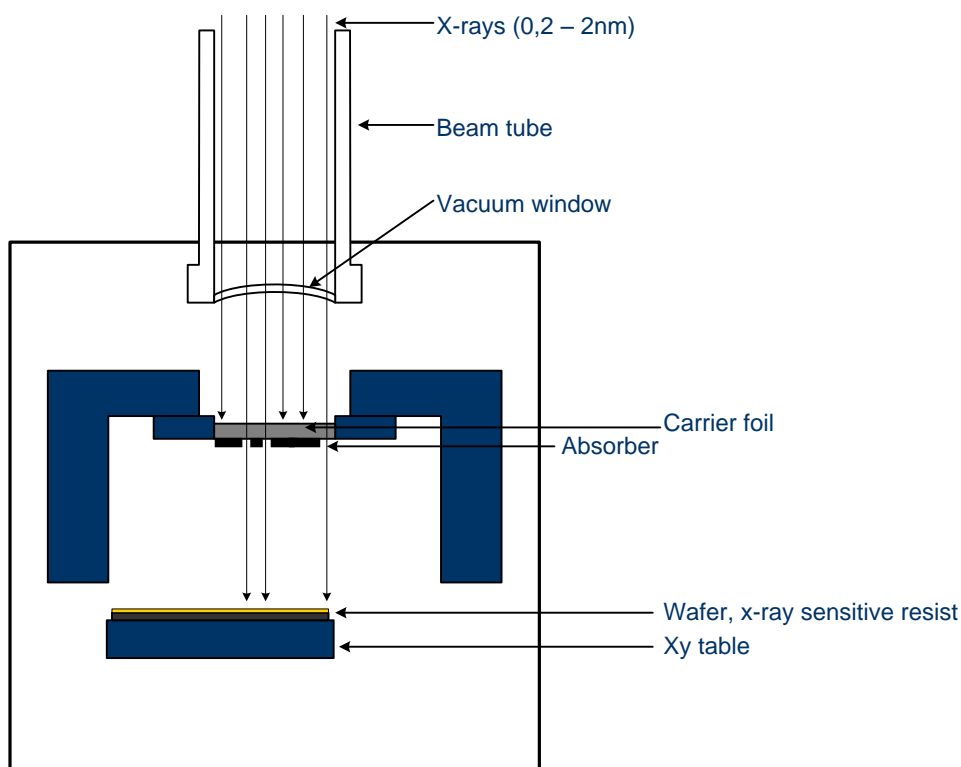
Fig. 6.9: X-ray lithography

6.2.7 Additional methods

An additional method of lithography is the use of ion beams. The wafer can be exposed with a photomask or, like in electron beam lithography, without a mask. In case of hydrogen ions the wavelength is about 0.0001 nm. Other elements allow a direct doping of the wafer without the use of masking layers.

6.3 Photoresist

6.3.1 Photoresist

There are two types of photoresist, positive and negative resist, which are used in different applications. In positive resist, the exposed areas are solubly, in negative resist the exposed areas are insolubly for wet chemical development.

Characteristics of positive resists:

+ excellent resolution

+ stable against developers

+ can be developed in aqueous developers

- bad resistance in etching or implantation processes

- bad adhesion on the wafer

Characteristics of negative resists:

+ high sensitiveness

+ fair adhesion

+ excellent resistance against etch or implantation processes

+ cheaper than positive resists

- lower resolution

- organic developers are needed (toxic)

For patterning of wafers in manufacturing, almost only positive resits are used. Negative resists were primarily used as a passivation which can be cured by ultra-violet radiation. If there is no specification in the text, a positive resist is the subject.

6.3.2 Chemical composition

Photoresists are composed of adhesive agents, sensitizers and solvents.

- **Binders (20%):** As a binder Novolac is used, which is a synthetic resin to control the thermal characteristics of the resist.

- **Sensitizer (10%):** Sensitizers define the photosensitivity of the resist. Sensitizers are composed of molecules which affect the solubility of the resist if it is exposed to energetic radiation. Thus the lithography has to take place in areas with ambient light which has a low energy.

- **Solvents (70%):** Solvents define the viscosity of the resist. By annealing, the solvent is vaporized and the resist is stabilized.

A resist, as it is provided from vendors, has a defined surface tension, density and viscosity. For this reasons the thickness of the resist layer in wafer fabrication depends on the temperature and the rpm of the coating tool.

6.4 Development and inspection

6.4.1 Development

The exposed wafers are developed in dipping baths or in spray processes. While dipping baths allow the development of multiple wafers at a time, in spray development one wafer is processed after another. As in resist coating processes the wafer is placed on a chuck and sprayed with developing chemicals while rotating at low rpm. Subsequent the wafer development process is stopped with water.

Some advantages of the spray development in contrast to dipping baths are:

- smallest structures can be developed

- the chemical is renewed steadily

- the amount of chemicals is much less

Depending on the type of resist - negative/positive - exposed areas are solubly or insolubly in developing chemicals and a patterned wafer remains after development. The exposure causes a chemical reaction in the resist thus that the sensitizer forms an acid which is neutralized by the developer like follows:

$$R-COOH + NaOH \longrightarrow (R-COO)^- + Na^+ + H_2O \text{ (R = uninvolved agents)}$$
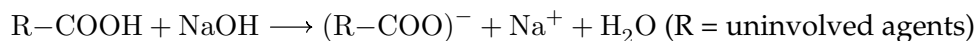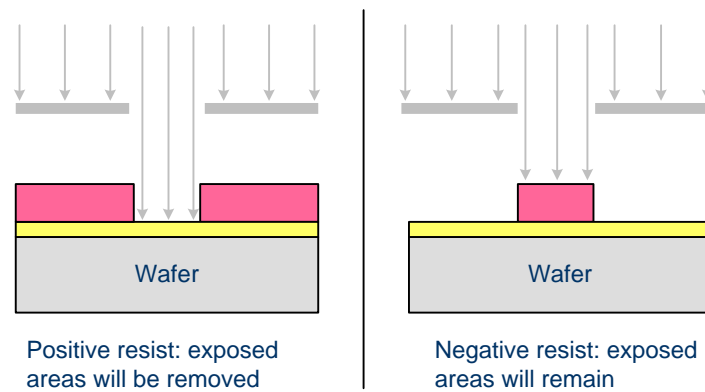
Fig. 6.10: Illustration of positive and negative resist after development

Because potassium hydroxide or sodium hydroxide leave residuals on the wafer, chemicals without metal ions, like TMAH (tetramethylammonium hydroxide), are used. An additional annealing (hard-bake) hardens the resist to be resistant to subsequent etch processes or ion implantation.

6.4.2 Inspection

The resist has to be inspected afterwards. In angular incidence of light the uniformity of the layer can be inspected as well as bad focusing or agglomeration of resist. If structures are too thin or too wide the resist has to be removed and the process has to be repeated. The resist pattern has to be adjusted precisely to the layer beneath or the process has to be repeated as well. Different alignment marks are used to investigate the adjustment and the line width.

The line width is inspected with a microscope: light rays incidence in perpendicular direction onto the wafer and will not be reflected into the objective from edges. Thus the edges appear as dark lines which can be used to calculate the line width and distance to adjacent lines.

6.4.3 Resist removal

After the pattern has been transferred into the layer beneath in etch processes, or after the resist mask was used in ion implantation, the resist has to be removed. This is done with abrasive chemicals (remover), in a dry etch step or with solvents. As solvent acetone can be used since it does not corrode other layers on the wafer. Due to a dry
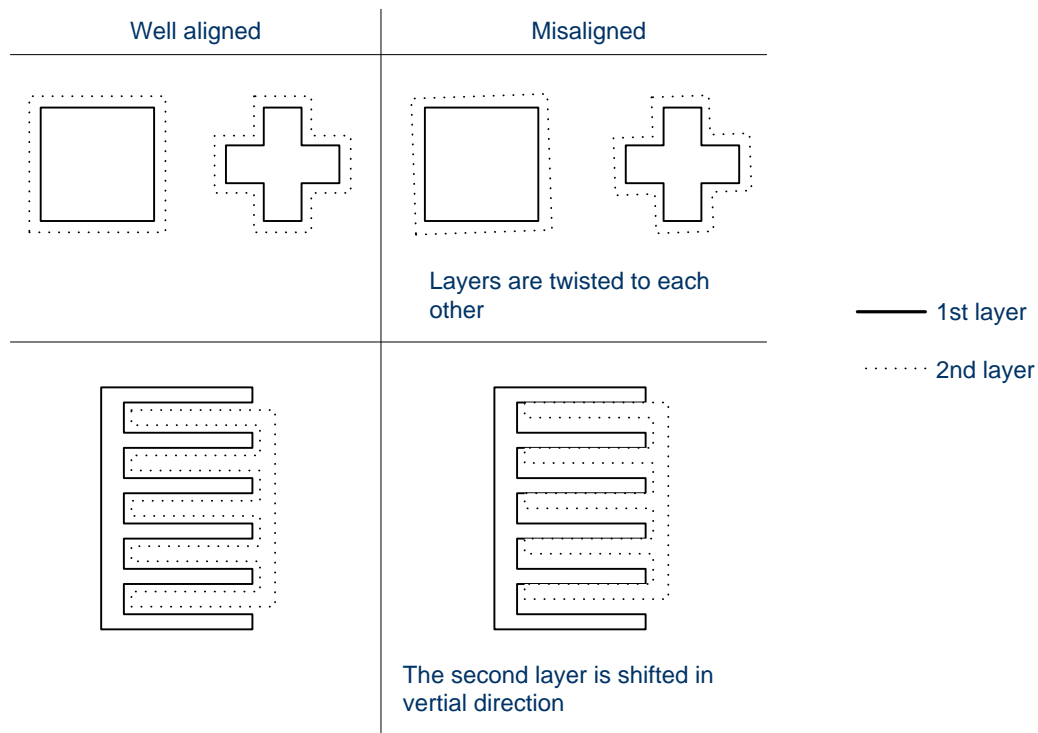
Fig. 6.11: Illustration of alignment marks

etch process or ion implantation the resist could be hardened, so that solvents can't remove it.

In this case the resist can be removed with a remover dilustion at about 80 °C in a dipping bath. If the resist has been heatened above 200 °C even the remover can't remove it. In this case the resist has to be removed in an ashing process.

Under oxygen ambient a gas discharge is ignited by high frequency, thus energetic oxygen atoms are generated. These atoms can ash the resist residue-free. However, charged particles can be accelerated in the electric field an cause damage to the wafer surface.

6.5 Photomasks

6.5.1 Introduction

Photomasks used for optical lithography contain the pattern of the integrated circuits. The basis is a so called *blank*: a glass substrate which is coated with a chrome and a resist layer. The resist is sensitive to electron beams and can be transferred into the chrome layer via etch processes. The chrome represents opaque areas on the photomask which are responsible for the casting of shadow during exposure of the silicon wafers.

The photomasks are directly exposed with electron beams under hard vacuum. With this method a resolution far below 100 nm is possible.

Due to wave optics (e.g. diffraction) there can be aberration during the exposure of wafers. Thus the optical proximity correction (OPC) has been introduced in semiconductor manufacturing, which can eliminate or reduce image defects.

OPC means to modify the structures on the mask in such a way that the shape of the image on the wafer looks like desired. Furthermore there can be additional structures just for minimizing aberration which do not have any function for the integrated device itself.

6.5.2 Photomask manufacture

The manufacturing of photomasks is basically equal to the wafer fabrication. The difference is the exposure of the resist which is done by electron beams (photomasks) or with optical lithography (wafer).

1. First step is the exposure of photoresist with electron beams (or laser).
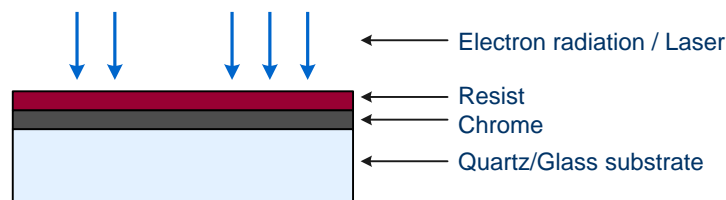
Fig. 6.12: Basic film stack of a chrome on glass mask (COG)

2. Subsequent the resist is developed to form a pattern.
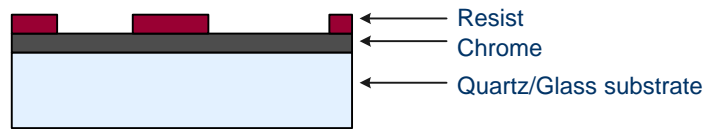
Fig. 6.13: Development of the resist layer

3. The resist acts as a masking to transfer the pattern into the chrome layer by etch processes.
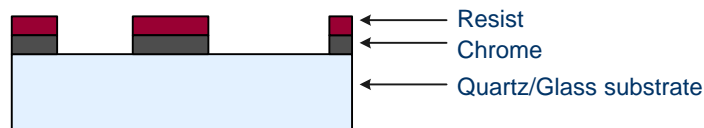
Fig. 6.14: Chrome etch

4. Subsequent the resist is removed.
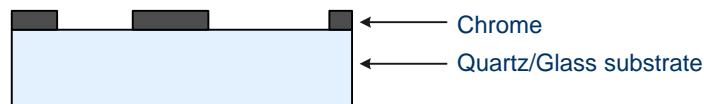
Fig. 6.15: Resist strip

5. Finally a pellicle is montaged on top to prevent contamination of the glass/chrome
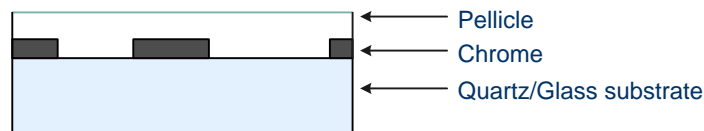
Fig. 6.16: Attached pellicle

6.5.3 Photomasks

Besides the traditional chrome on glass mask (COG) there are various types of photomasks which enhance the optical resolution of the structures. The central issue of COG masks is the diffraction of the light on edges. Thus the light will not only impact in perpendicular direction but will be deflected into areas which must not be exposed.

With different means which are described below the intensity of the diffracted light should be reduced.
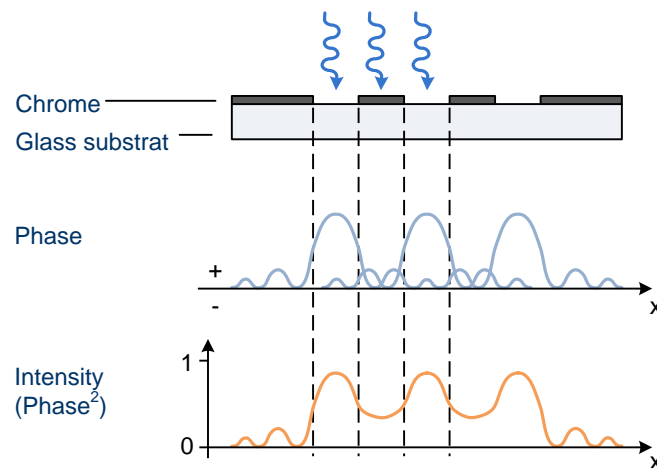
Fig. 6.17: Spectral intensity of a COG mask

Attenuated Phase Shift Mask (AttPSM):

The attenuated phase shift mask (AttPSM; also half tone mask) uses a patterned layer of molybdenum silicide (MoSi) which represents the structures of the circuit. The molybdenum silicide has a thickness which causes a phase shift of the transmitted light of 180°. Thus the phase shifted light and the radiation which transmits through glass only interfere destructively. In addition the molybdenum silicide is dense (6 % or 18 % @ 193 nm wavelength). On the one hand the light is attenuated and on the other hand the light waves which are in opposite phase erase each other almost completely, this results in a higher contrast. A chrome layer can be added to areas which are not used for exposure to mask unused regions. This photomasks are named tritone masks.

(a)

(b)
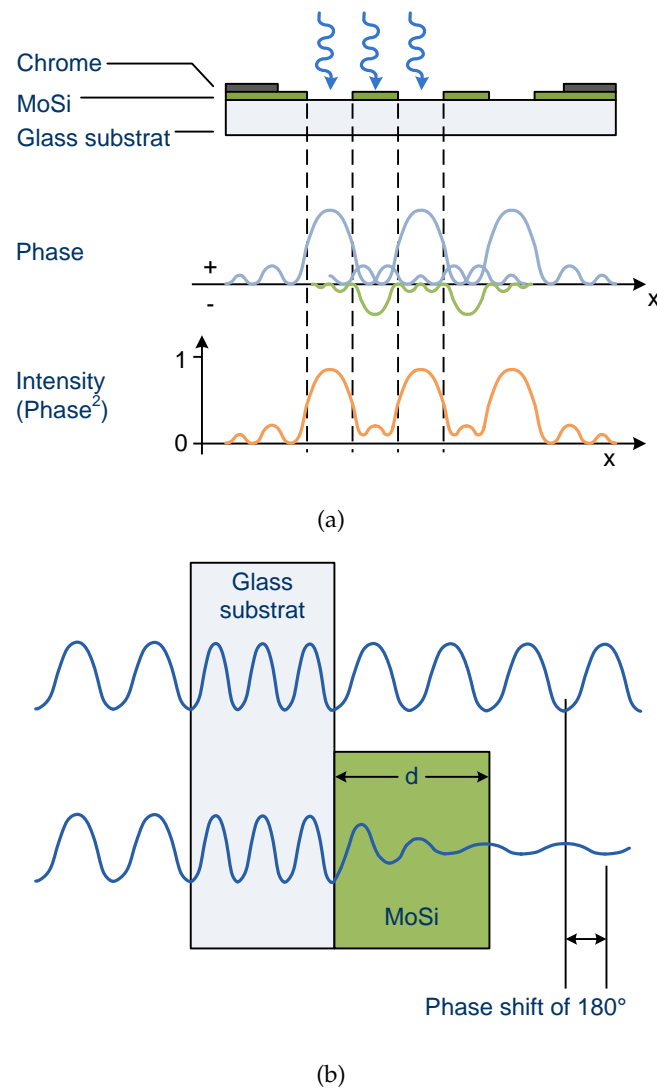
Fig. 6.18: (a) Spectral intensity of an Attenuated Phase Shift Mask, (b) principle of the phase
shift with molybdenum silicide

Chrome less Phase shift mask:

Chrome less phase shift masks don't use opaque films. The phase shift is achieved
by trenches which are directly etched into the glass substrate. The manufacturing of
these masks is difficult, since the etch approach has to be stopped in the middle of the
glass. In contrast to etch processes where one layer is completely etched till the layer
beneath is reached - which causes changes in the etch plasma, so that one knows when
the process is finished -, there is no indication when the exact depth in the substrate is
reached.

Alternating Phase Shift Mask (AltPSM):

The alternating phase shift mask also uses trenches which are etched into the glass substrate alternating to non-etched areas. In addition there are areas which are covered with a chrome layer to decrease the intensity of radiation in this regions.
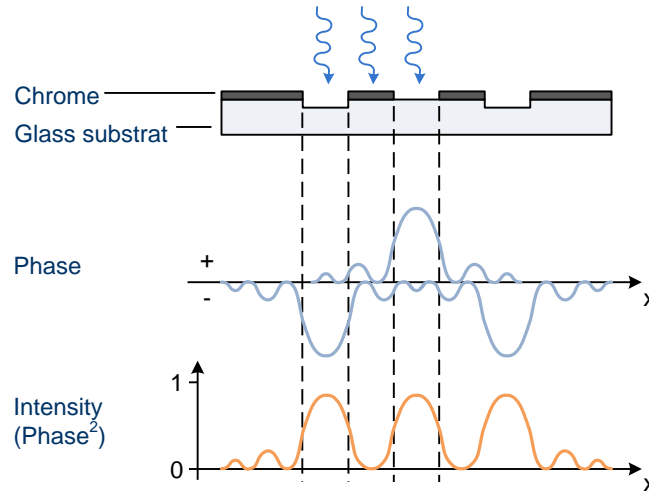
Fig. 6.19: Spectral intensity of an Alternating Phase Shift Mask

However, there are regions with an undefined phase shift, so that one has to exposure twice with different masks. One mask contains the structures which run in x-direction, while the second mask contains the patterns which are orientated in y-direction.
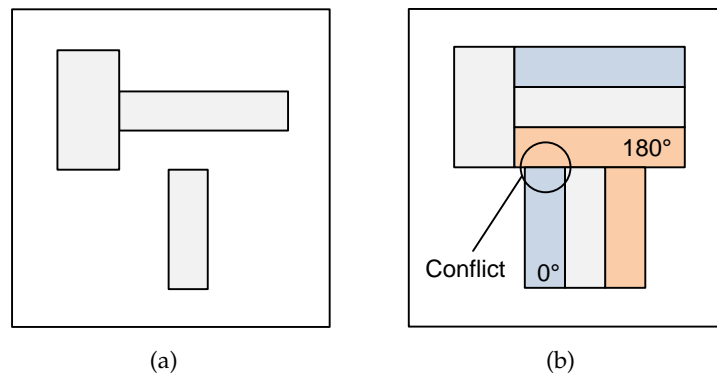
Fig. 6.20: Desired structures on the wafer and conflict on the mask

6.5.4 Next generation lithography

In the future there has to be a complete change-over which means, that the traditional lithography tools and the photomasks have to be replaced. Next generation lithography is expected to use extreme ultra-violet radiation (EUV, wavelength 13.5 nm) which is absorbed in normal atmosphere as well as in glass. For this reason EUV processes have to take place under vacuum and instead of optical lenses mirrors have to be used for focusing. The photomasks will have a reflective surface instead of translucent glass.

Also beacuse of the tremendous effort (technical as well as of money matters) the traditional lithography is kept alive as long as possible, and new inventions as phase shift masks or immersion lithography have pushed the conventional lithography. Thus today's feature sizes of only 32 nm can still be exposed with a wavelength of 193 nm. Sometimes there was a thought of the introduction of radioation sources with 155 nm, however, because of the immense costs it is foreseeable that subsequent to argon fluoride lasers (193 nm) EUV sources will be established instantly.

7 Wet chemistry

7.1 Etch processes

In the fabrication of semiconductor microdevices various materials have to be etched. Either for removing an entire layer from the surface or to transfer a resist pattern into a layer beneath. Etch processes can be devided into wet and dry etching while there is a further separation into isotropic and anisotropic processes and also a separation in chemical and physical etch characteristics.

In an isotropic etch process the etching occurs in lateral and vertical direction. Thereby layers are removed not only in thickness but also in their circumference. In anisotropic processes the layer is only removed in vertical direction. Depending on the demands an isotropic process can be desired as well as an anisotropic.

Isotropic etch process,
particles move in each direction:
Under etch of the resist mask,
high selectivity

Anisotropic etch process,
perpendicular orientated:
Exact transfer of the resist mask,
low selectivity
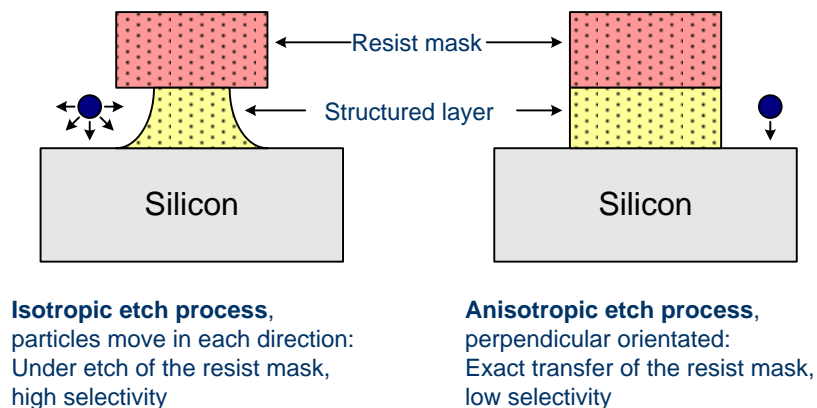
Fig. 7.1: Isotropic and anisotropic etch processes

An important value of etch processes is the selectivity. The selectivity is the ratio of abrasion of the layer which should be etched (e.g. an oxide film) and of the other layer (e.g. a resist mask). If the selectivity is 2:1 the oxide would be etched twice as fast as the resist.

Wet chemistry processes are not only applicable for etching but also for other needs:

- **wet etching:** removal of doped or undoped oxide layers from the entire wafer

- **wafer cleaning**

- **photoresist removal**

- **backside processing:** to remove layers which were deposited as a byproduct during other processes (e.g. thermal oxidation)

- **polymer removal:** to remove byproducts which occur during dry etching

The wet etching is only used very rare for structuring because of its (generally) isotropic etch profile. An exception are micromechanical devices. Due to the atomic structure of silicon crystals, well defined profiles with flank angles of 90° or 54.74° can be produced using wet chemistry.

7.2 Wet etching

7.2.1 Principle

The principle of wet etching processes is the conversion of solid materials into liquid compounds using chemical solutions. The selectivity is very high since the used chemicals can be adapted very precisely to the individual films. For most solutions the selectivity is greater than 100:1.

7.2.2 Requirements

The following requirements have to be fulfilled by liquid chemistry:

- the mask layer must not be attacked

- the selectivity has to be high

- the etch process has to be able to be stopped by dilution with water

- reaction products must not be gaseous because they could shadow other regions

- constant etch rates all along the process

- the reaction products must be solubly to avoid particles

• environmental safety and ease of disposal are neccessary

7.2.3 Batch etching

In batch etching multiple wafers can be etched simultaneously, filters and circulating pumps prevent particles from reaching the wafers. Since the concentration of the chemistry is decreased with each processed wafer it has to be renewed often.

The etch rate, in other words the abrasion per time, has to be well known to ensure a reproducible process. A precise tempering is essential since etch rates increase with increasing temperature.
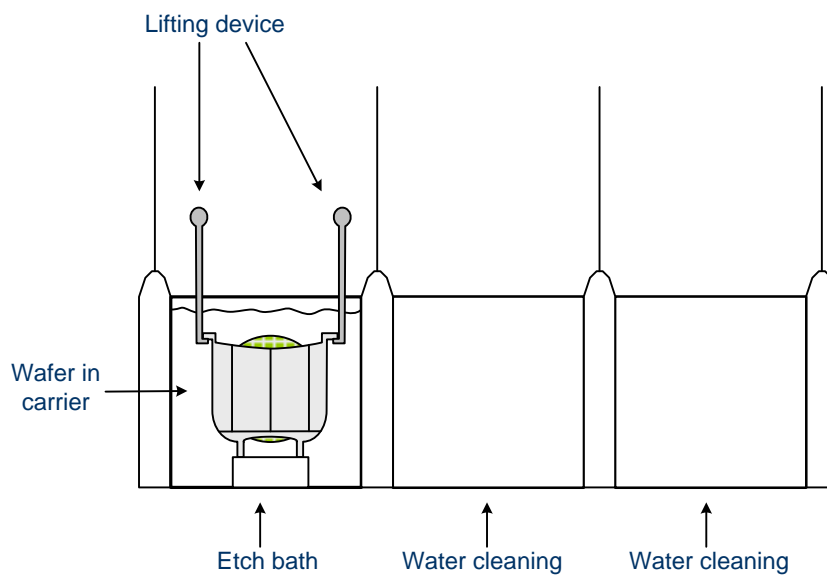
Fig. 7.2: Batch etching

A lever can transfer the wafers in horizontal and vetical direction. After the wafers have been etched, the etch process is stopped by purging with water in seperate baths. Subsequent the moisture is removed in spin-dryers.

The advantage of the batch etching is the high throughput and the simple construction of the etch tools. However, the uniformity is low.

7.2.4 Spray etching

The spray etching is comparable to the spray development in lithography. Due to the rotation of the wafer simultaneously to steadily renewed etch chemistry the uniformity is very good. Bubbles can't emerge because of the fast rotation, however, each wafer has to be processed separately.

As an alternative to the single wafer process the spray etch can be done on multiply wafers at a time. In a spin etcher the wafers are placed around spray nozzles and revolve concentrically. Afterwards the wafers are dried in a hot nitrogen atmosphere.
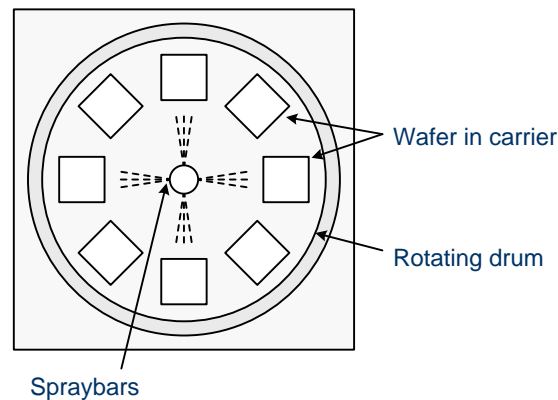
Fig. 7.3: Spray etching

7.2.5 Anisotropic etching of silicon

Although the molecules in liquids can move in every direction there are wet etch processes to create an almost anisotropic etch profile. For this approach the unequal etch rates on different crystal orientations are utilized. (100) and (110) oriented crystal faces can be etched much faster than (111) oriented. Thus 'V' shaped trenches (100 silicon) or trenches with perpendicular sidewalls can be fabricated. The etching is either done with potassium, soda or lithium lye (KOH, NaOH, LiOH) or with an EDP dilution (a mixture of water, pyrazine, catechol, and ethylenediamine). Responsible for the reaction is in either case the OH$^-$ ion (hydroxyl):

$$\mathrm{Si} + 2\,\mathrm{H_2} + 2\,\mathrm{OH^-} \longrightarrow \mathrm{SiO_2(OH)_2^{2-}} + 2\,\mathrm{H_2}$$
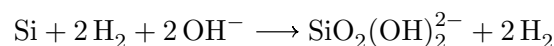
However, anisotropic dilutions are not applicable for microelectronic devices but for micromechanics.

7.2.6 Etching solutions for isotropic etching

There are individual dilutions for all the different materials. For example silicon dioxide is etched by hydrofluoric acid (HF):

$$SiO_2 + 6\,HF \longrightarrow H_2SiF_6 + 2\,H_2O$$
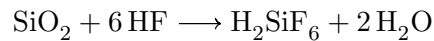
The dilution is buffered with NH_4F to maintain the concentration of HF (so-called buffered HF, BHF). In a mixture of 40 percent NH_4F and 49 percent HF (ratio 10:1) the etch rate on thermal oxide is 50 nm/min. TEOS (CVD) oxides and PECVD oxides are etched much faster (150 nm/min and 350 nm/min, respectively). The selectivity compared to crystalline silicon, silicon nitride and polysilicon is much greater than 100:1.

Silicon nitride is etched by hot phosphoric acid (H_3PO_4). The selectivity in contrast to silicon dioxide is low (10:1). In polysilicon the selectivity compared to silicon nitride is primarily defined by the concentration of the phosphoric acid.

Crystalline or polycrystalline silicon are at first oxidized with nitric acid (HNO), afterwards the oxide is etched with HF.

$$3\,Si + 4\,HNO_3 \longrightarrow 3\,SiO_2 + 4\,NO + 2\,H_2O$$
$$SiO_2 + 6\,HF \longrightarrow H_2SiF_6 + 2\,H_2O$$
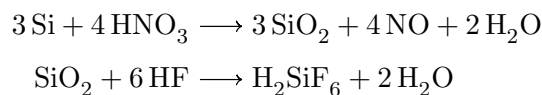
Aluminum can be etched at 60 °C with a mixture of nitric and phosphoric acid, titan is etched with a mixture of ammonia water (NH_4OH), hydrogen peroxide (H_2O_2), and water (ratio 1:3:5). Because this mixture can attack silicon as well its lifetime is low.

In general wet etching is suitable to remove entire layers of the wafer. The selectivity is very high for most materials, and therefore there is no risk to etch the wrong film. In addition the etch rate is very good, in bath etching many wafers can be processed at

a time. However, for small structures the wet etching can't be used since its isotropic character causes lateral etching of the masked films. For this approach layers are removed by dry etching with anisotropic etch profiles.

7.3 Wafer cleaning

7.3.1 Cleanroom

The fabrication of semiconductor microdevices takes place in cleanrooms to protect the complex circuits from contaminations which can impact operability. The cleanrooms are classified by size of the particles and their amount per cubic foot (cu.ft.) or per cubic meter (ISO standard):
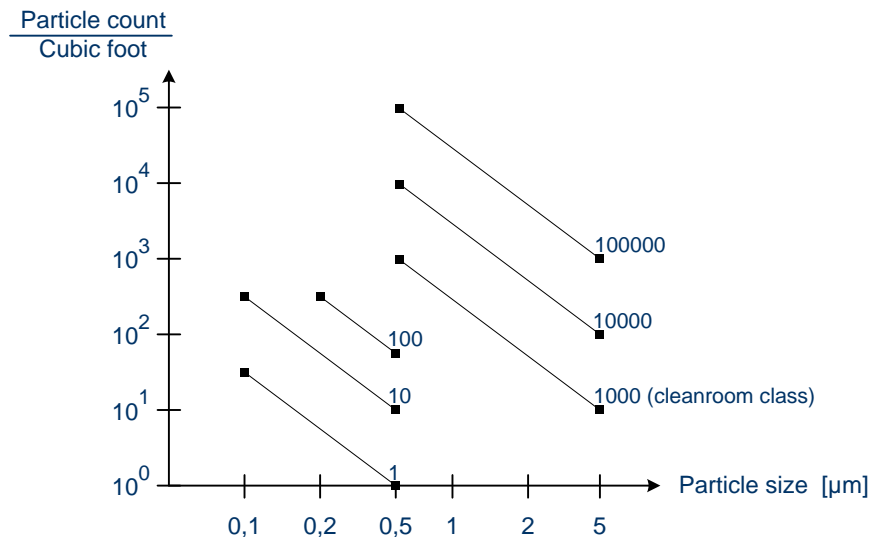
Fig. 7.4: Cleanroom classes

In a class 3 cleanroom, for example, there may only be 1,000 particles with a diameter of ≥ 0.1 microns, 237 with a diameter of ≥ 0.2 microns and so on. In an operating room the cleanroom class is generally 2 or 3. In a volume of $1\ m^3$ air in a city there are e.g. 400 million particles with a size of 5 microns.

In cleanrooms for the semiconductor industry the air is cleaned with super fine filters which is blown through the ceiling and exhausted through holes in the floor. This laminar flow forwards particles top down. To avoid contaminations from the outside

there's always some over pressure inside the cleanroom. Since the production line heatens the air, the air rises and can cause turbulences. To avoid this, the laminar flow can be in reversed direction, from the floor to the ceiling. This is primarily done if the wafers are transported inside of sealed boxes which dock directly to the facility so that there is no possibility for contaminations to reach the wafers. This boxes are called FOUPs (Front Opening Unified Pod).

Since the wafers are completely separated from the ambient air in the cleanroom, there can be another cleanroom class inside the FOUP - a so-called mini environment. For this reason the class inside the boxes can be 1 or 2 to protect the wafers from particles, while the class in the cleanroom itself is only 5, for instance. This is much more efficient since not the entire cleanroom needs to be super clean but only the relatively tiny transport boxes.

In the illustration above the production line is only the region between the light red colored ventilating system and the yellow colored bottom area. Beneath there is the basement with supply units (pumps, generators etc.), on top there are the air filters. To minimize turbulences the huge cleanroom area often is separated into individual corridors. Many cleanrooms are surrounded by a so-called greyroom (service room) in which the production tools are standing. In this case only the locks to bring in the wafers and the control panels are accessible from inside the cleanroom.

The staff wears special cleanroom suits (often called bunny suits) which do not emit particles. Depending on the requirements the suit covers the complete body including head and face. In addition there are special boots, underclothing and gloves. Next to the entrance there are sometimes air showers to blow everything clear.

7.3.2 Types of contamination

Despite the cleanroom there are different types of contamination which are mainly emitted by the staff, the ambient air, by chemicals (gases, dilutions), and by the facility:

- **microscopic contamination:** e.g. particles out of the air or from gases

- **molecular contamination:** e.g. hydrocarbon from oil in pumps

- **ionic contamination:** e.g. sweat

- **atomic contamination:** e.g. heavy metal from dilutions, abrasion of solids

7.3.3 Microscopic contamination

Microscopic contaminations are particles which adsorb at the wafer surface. Sources of these contaminations are the ambient air, clothing, abrasion of moving parts, insufficiently filtered liquids (for cleaning, etching, development, ...), or residuals after dry etching.

Microscopic contaminations cause shadowing at the surface, e.g. during lithography. Relativley big particles reduce resolution in contact exposure if they are located between the photomask and the resist layer. In addition they can hold of accelerated ions of the surface during ion implantation or during dry etch.
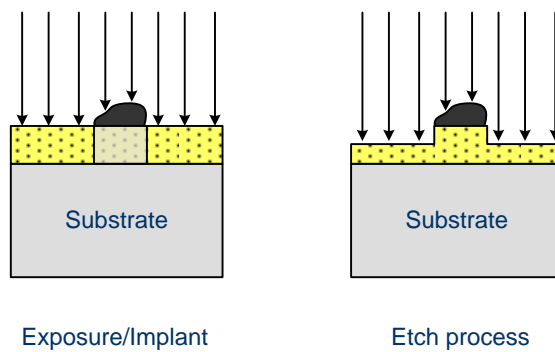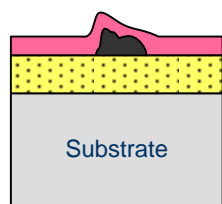

Exposure/Implant   Etch process

Fig. 7.5: Shadowing in etch and ion implant processes due to particles

Particles can also be enclosed by films so that there is uneveness. Subsequent deposited layers can crack at these locations or resist accumulates which causes problems during exposure.

Unevenness due to
embedded particle

Fig. 7.6: Embedded particle

7.3.4 Molecular contamination

Molecular contaminations are a result of resist residuals and solvent residuals on the wafer or of oil mist from vacuum pumps. This contaminations can accumulate at the surface as well as diffuse into the material. Superficial contaminations can handicap the adhesion of later deposited films which is vastly in case of metallization. If the contaminations diffuse into an oxide layer the electric toughness of the film is degraded.

7.3.5 Alkaline and metallic contamination

The key source of this contamination is the human who steadily emits salts through skin and breath. But also (alkaline) ions of sodium or potassium from insufficient deionized water can reach the wafers as well as heavy metals of etch dilutions. By radiation in facilities (ion implantation, dry etching) material can be sputtered of walls which deposits on the wafer.

Ionic contaminations for example affect the electric behavior of MOS transistors since their charges change the threshold voltage (the voltage at which the transistor gets conductive). Heavy metals, like iron or copper, provide electrons so that the power consumption of diodes increases. Metals can also act as recombination centers for free charge carriers and on account of this there are not enough free charge carriers for correct mode of operation.

7.3.6 Cleaning techniques

The wafers are cleaned subsequent to each wet-chemical treatment with ultrapure water, but residuals are also removed after other processes by cleaning techniques. For this reason there are various cleaning techniques for individual contaminations. The consumption of ultrapure water is immense and is several hundrets of million liters per annum. In ultrapure water there are almost no contaminations left, 1-2 parts per million (ppm; contaminations per water molecules) are allowed (state of 2002). Generally the water is purified on location.

One possibility for wafer cleaning is the ultrasonic bath in which the wafers are placed with a dilution of water, ultrasonic cleansers, and surfactants. Particles are dissoluted from the surface by ultrasonic stimulation, metals and molecular contaminations are

partial bound by the cleansers. Not to strong attached particles can also be blown off by nitrogen.

To remove organic contaminations such as grease, oil, or danders, solvents like acetone or ethanol are applicable. However, these can leave carbon residuals.

Ionic contaminations (ions of sodium, potassium etc.) are removed by purging with deionized water. Also the cleaning with rotating brushes and cleansers is possible, but particles will be accumulated at edges and the brushes can damage the surface. Vias or other cavities can be purged with high pressure cleaning (50 bar), however, with this treatment ionic contaminations can not be removed.

Cleaning with water or different cleansers often is not suffiencent, and therefore contaminations have to be removed with aggressive etch dilutions or the surface is milled slightly. Mixtures of hydrogen peroxide and ammonia or Caro's acid (sulfuric acid with hydrogen peroxide) can remove organic contaminations through oxidation at about 90 °C.

A mixture of hydrochloride acid and hydrogen peroxide form alkaline metals into soluble chlorides (salts), heavy metals form complexes and get dissolved. Native oxide can be removed with hydrofluoric acid, in contrast a defined deposited oxide layer can also cover the surface for cetain reasons.

Besides the rinse of the wafers after each individual wet-chemical process, they do pass entire cleaning processes with several different cleaning steps. Thereby the sequence of the cleaning techniques is important, since they can affect each other. A cleaning sequence, for example, could look like this:

- blowing off particles with nitrogen

- cleaning with ultrasonic

- removing organic contaminations with Caro's acid (H_2SO_4 - H_2O_2)

- removing of even finer organic contaminations with a mixture of NH_4 and H_2O_2

- removing of metallic contaminations with hydrochloric acid and hydrogen peroxide

- drying with a spin-dryer in hot nitrogen atmosphere

Subsequent to every cleaning step the wafers are rinsed with ultrapure water, native oxide is removed in a hydrofluoric acid dip. Depending on the wafer surface the clean-

ing sequence is different since some cleansers can attack films. At ever smaller structures the cleaning is hindered, not only because of smaller openings which can not be accessed as easily but also because of surface tension and capillary force which can cause collapse of structures.

8 Dry etching

8.1 Overview

8.1.1 Abstract

In wet etching film stacks are generally etched isotropic, however, at small feature sizes an anisotropic profile is neccesary. Therefore dry etch processes are applicable which have a sufficient selectivity. Such methods allow reproducible uniform etching of almost all materials which are used in semiconductor manufacturing. Besides anisotropic processes, isotropic processes are possible as well. Despite the expensive production facilities and the single wafer processes, dry etching has established itself.

8.1.2 Major values in dry etching

Etch rate r**:** the etch rate is the abrasion per time and for example can be specified in nanometers per minute or Angstrom per second.

$$r = \frac{\text{Etch removal } \Delta z}{\text{Time } \Delta t}$$

Anisotropism f**:** The anisotropism gives the ratio of horizontal etch rate r_h and vertical etch rate r_v.

$$f = 1 - \frac{r_h}{r_v}$$

For structuring high anisotropic processes are desired, which means an etching only in vertical direction, so that the resist mask is not underetched. For anisotropic etch processes $f \to 1$, and accordantly for isotropic processes $f \to 0$.

Selectivity S_{jk} of layer j and layer k: the selectivity is the ratio of the etch rates of two films, e.g. of the layer which should be structured (j; e.g. oxide) and the layer which should not (k; e.g. resist).

$$S_{jk} = \frac{r_j}{r_k}$$

The value of the selectivity depends on the process. For structuring the value should be as high as possible, which means that the layer one wants to structure is etched much faster than a masking resist layer. In reflow back etching the selectivity should be about 1 to ensure a uniform surface.

8.1.3 Dry etch processes

In dry etching gases are stimulated by high frequency, which is primarily 13.56 MHz or 2.45 GHz. At a pressure of 1 to 100 Pa the mean free path is some millimeters to centimeters.

Primarily there are three types of dry etching:

- **physical dry etching:** physical abrasion of the wafer surface by accelerated particles

- **chemical dry etching:** chemical reaction of gas and wafer surface

- **chemical physical dry etching:** physical etch process with a chemical character

8.2 Dry etch processes

8.2.1 Ion Beam Etching

The ion beam etching (IBE) is a physical dry etch process. Thereby argon ions are radiated onto the surface as an ion beam with about 1 to 3 keV. Because of the energy of the ions, they strike out material of the surface. The wafer is held perpendicular or tilted into the ion beam, the etch progress is absolute anisotropic. The selectivity is low because there is no differentiation of the individual layers. The gas and the striked out material are exhausted by vacuum pumps, however, particles can deposit on the wafer or on chamber walls since the reaction products are not in gaseous state.
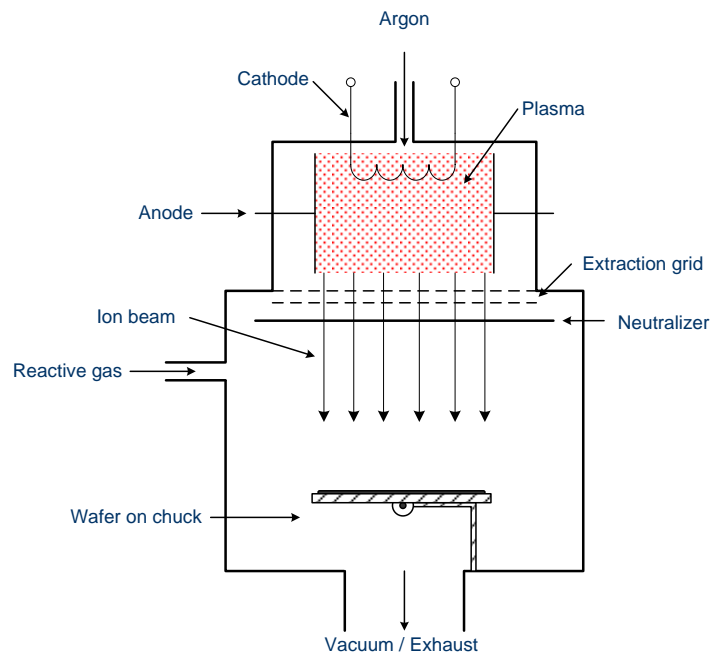
Fig. 8.1: Illustrationn of an ion beam etch reactor

To avoid particles a second gas is led into the chamber. This gas reacts with the argon ions and causes a physical chemical etch process. Partial the gas reacts with the surface but also with the striked out particles to form gaseous byproducts.

Almost every material can be etched with this method. Due to the perpendicular radiation the abrasion on vertical walls is very low (high anisotropism). However, because of the low selectivity and the low etch rate, this process is only used rare in todays semiconductor fabrication.

8.2.2 Plasma Etching

The plasma etching (PE) is an absolute chemical etch process (chemical dry etching, CDE). The advantage is that the wafer surface is not damaged by accelerated ions. Due to the movable particles of the etch gases the etch profile is isotropic, thus this method is used to remove entire film layers (e.g. back side clean after thermal oxidation).

One reactor type for plasma etching is the down stream reactor. Thereby a plasma is ignited at high frequency of 2.45 GHz through impact ionization, the location of the impact ionization is separated from the wafer.
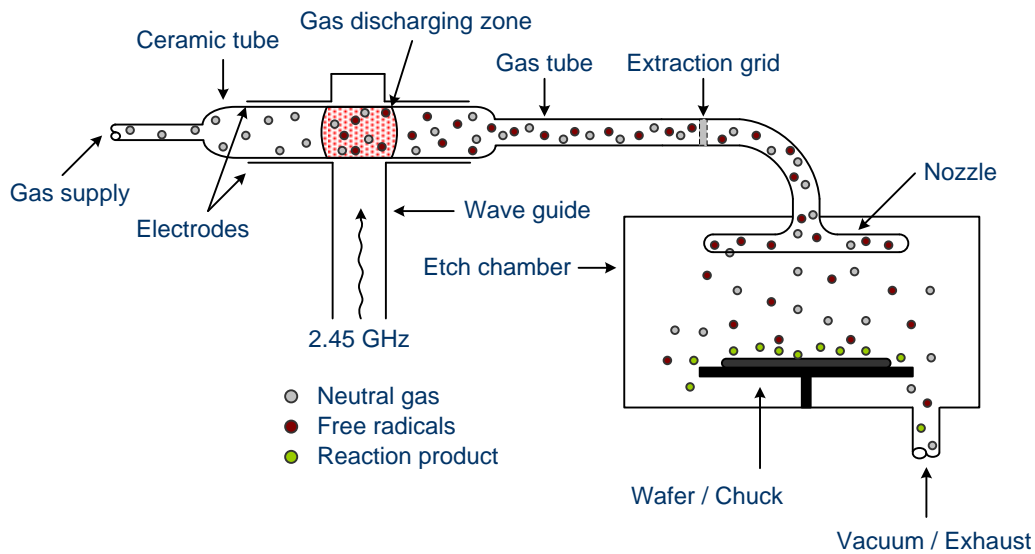
Fig. 8.2: Illustration of a CDE reactor

In the region of the gas discharge there are various particles due to impacts, amongst others there are radicals. Radicals are neutral atoms or molecules with an unsaturated electron which are therefore very reactive. As a neutral gas e.g. tetrafluoromethane CF_4 is led in the gas discharge zone and separated into CF_2 as well as fluorine molecules F_2. Similarly fluorine can be splitted from CF_4 by adding oxygen O_2:

$$2\,CF_4 + O_2 \longrightarrow 2\,COF_2 + 2\,F_2$$
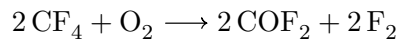
The fluorine molecules can now be splitted into two individual fluorine atoms by energy at the gas discharging zone: each fluorine atom is a fluorine radical, since each atom has seven valence electrons and wants to achieve the noble gas configuration.

Fig. 8.3: Generation process of radicals

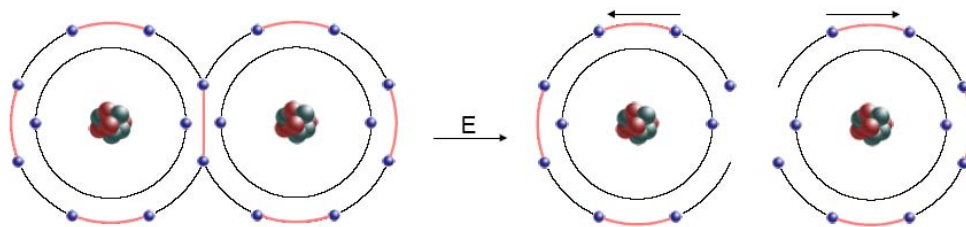Besides neutral radicals there are several, in part charged particles (CF_4^+, CF_3^+, CF_2^+, ...). All the particles, radicals etc. are then led through a ceramics pipe into the etching chamber. Charged particles can be hold off from the etching chamber by an extraction grating or are recombining on their way to form neutral molecules. Also the fluorine radicals are partial recombining, but there are enough to reach the etching chamber which can react at the wafer surface and cause a chemical abrasion. Other neutral particles are not part of the etching process and are exhausted as well as the reaction products.

Examples for films that can be etched in plasma etching:

- **Silicon:** $Si + 4\,F \longrightarrow SiF_4$

- **Siliconn dioxide:** $SiO_2 + 4\,F \longrightarrow SiF_4 + O_2$

- **Silicon nitride:** $Si_3N_4 + 12\,F \longrightarrow 3\,SiF_4 + 2\,N_2$

8.2.3 Reactive Ion Etching

The etching characteristic - selectivity, etch profile, etch rate, uniformity, reproducability - can be controlled very precisely in the reactive ion etching (RIE). An isotropic etch profile is possible as well as an anisotropic. Therefore the RIE process, a chemical physical etch process, is the most important process in semiconductor manufacturing for structuring various films.

Inside the process chamber the wafer is placed on a high frequency electrode (HF electrode). By impact ionization a plasma is generated in which free electrons as well as positively charged ions occur. If the HF electrode is at a positive voltage the free electrons accumulate on it and cannot leave the elctrode again because of their electron affinity. Thus the electrode charges up to -1000 V (BIAS voltage). The slow ions which could not follow the fast alternating field are now moving towards the negatively charged electrode.

If the mean free path of the ions is high, the particles impact on the wafer surface in almost perpendicular direction. Thus material is striked out of the surface by the accelerated ions (physical etching), in addition some of the particles are reacting chemically with the surface. Lateral sidewalls are not affected, so that there is no abrasion and the etch profile remains anisotropic. The selectiviy is not too small, however, due to the
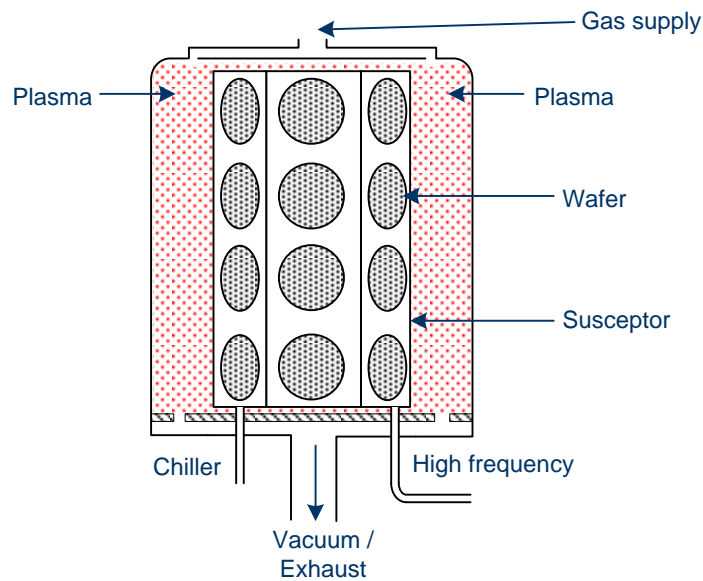
Fig. 8.4: Illustration of a RIE reactor in hexagonal construction

physical etch progress it's not to large either. In addition the wafer surface is damaged by the accelerated ions and has to be cured by thermal annealing.

The chemical part of the etch process is done by the reaction of free redicals with the surface and also with the physical milled out material in such a way that it can not redeposit onto the wafer or the chamber walls as in ion beam etching. By increasing the pressure in the etching chamber the mean free path of the particles is reduced. Therefore there are much more collisions and thus the particles are heading into various directions. This causes a less directed etching, the etch process gets a more chemical character. The selectivity increases, the etch profile is more isotropic.

By a passivation of the sidewalls during silicon etching, an anisotropic etch profile is achieved. Thereby oxygen inside the etch chamber reacts with milled out silicon to form silicon dioxide which deposits an vertical sidewalls. An oxide film on horizontal areas is removed due to the ion bombardement so that the etch progress in lateral direction proceeds.

The etch rate depends on the pressure, the power of the HF generator, the process gases, the real gas flow and on the wafer temperature.

The anisotropism increases with increasing HF power, deacresing pressure and decreasing temperature. The uniformity of the etch process depends on the gases, the
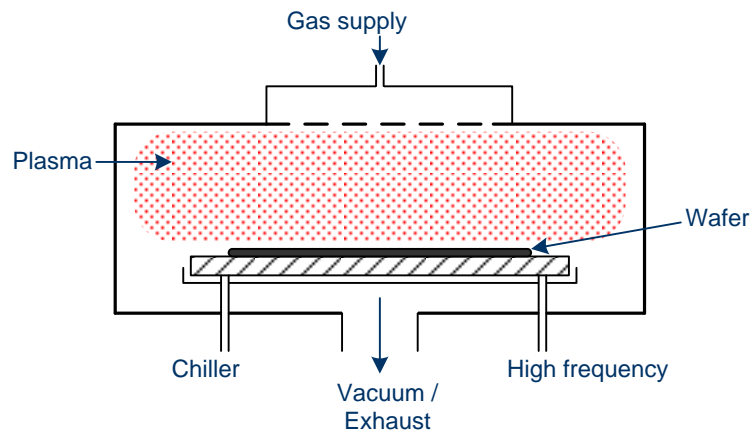
Fig. 8.5: Illustration of a RIE reactor

Gases in the process chamber: HCl, O_2, BCl_3

SiO$_2$ mask

Substrate

1. Generation of $SiCl_x H_y$ through ion bombardment
2. Adsorption of $SiCl_x H_y$ at the sidewalls
3. Oxidation of $SiCl_x H_y$ to a SiO_2 layer
4. Evaporation of Cl_2 and HCl

• Accelerated ions   ○ Oxide
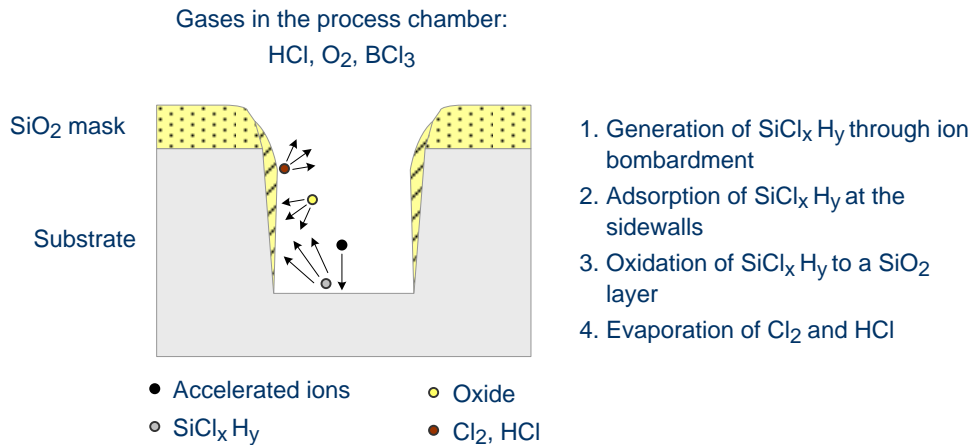○ $SiCl_x H_y$ • Cl_2, HCl

Fig. 8.6: Sidewall passivation during RIE etching

distance of the two electrodes and on the material the electrodes are made off. If the distance is to small, the plasma cannot be dispersed unfiormly and thus leads to inhomogeneity. If the distance of the electrodes is increased, the etch rate decreases because the plasma is distributed in an enlarged volume. For the electrodes carbon has proven itself as the material of choice. Since fluorine and chlorine gases attack carbon as well, the electrode causes an uniform strained plasma and thus the wafer edge is affected in the same way as the wafer center.

The selectivity and etch rate depend very strong on the process gases. For silicon and silicon compounds fluorine and chlorine gases are used primarily.

An etch process is not limited to one gas, a mixture of gases or to fixed process pa-

Tab. 8.1: Overview of process gasses used in dry etch processes

Material	Process gasses	Annotation
SiO_2, Si_3N_4	CF_4, O_2	F etches Si, O_2 removes carbon (C)
	CHF_3, O_2	CHF_3 acts as polymere, enhanced selectivity on Si
	CHF_3, CF_4	
	CH_3F	enhanced selectivity of Si_2N_4 on SiO_2
	C_2F_6 / SF_6	
	C_3F_8	enhanced etch rate compared to CF_4
Poly-Si	BCl_3, Cl_2	no contamination with (C)
	$SiCl_4$, Cl_2	
	HCl, O_2	
	$SiCl_4$, HCl	
	O_2 / $SiCl_4$, HCl	
	HBr / CL_2 / O_2	enhanced selectivity on Fotolack and SiO_2
	SF_6	high etch rate, fair selectivity on SiO_2
	NF_3	high etch rate, isotropic
	HBr, CL_2	
monokrist.	HBr, NF_3, O_2 / CF_3Br	enhanced selectivity on SiO_2
Silicium	BCl3, CL2 / HBr, NF3	
Aluminium-	Cl_2	isotropic etch process
Legierungen	BCl_3	low etch rate
	BCl_3 / CL_2 / CF_4	anisotropic etch process
	BCl_3 / CL_2 / CHF_3	enhanced sidewall passivation
	BCl_3 / Cl_2 / N_2	enhanced aetch rate, no contamination with

rameters. For example native oxide on polysilicon can be removed at first with a high etch rate and low selectivity, while the polysilicon is etched subsequent with a higher selectivity against the layer beneath.