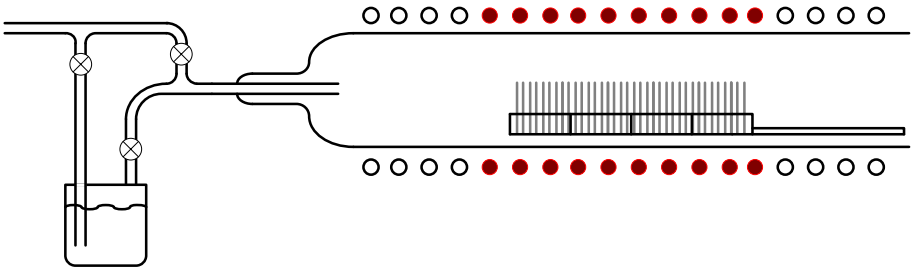
Semiconductor Technology

from A to Z



Oxidation

www.halbleiter.org

Contents

List of Figures

List of Tables

1 Oxidation

1.1 Overview

1.1.1 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, ...)

1.1.2 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).

1.2 Fabrication of oxide layers

1.2.1 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.

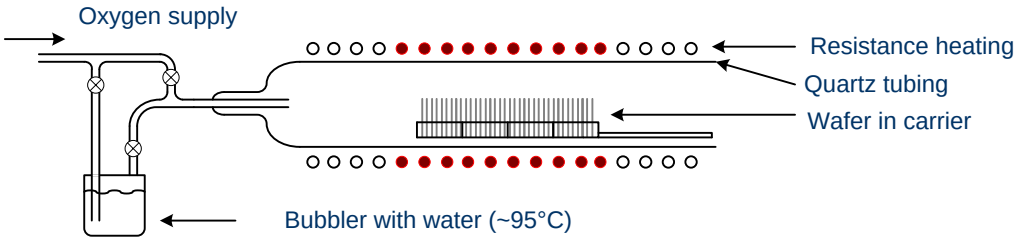


Fig. 1.1: Illustration of a furnace for therma oxidation

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. 1.1: 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:

$$k = \frac{\text{Solubility of the dopant in silicon}}{\text{Solubility of the dopant in } SiO_2}$$

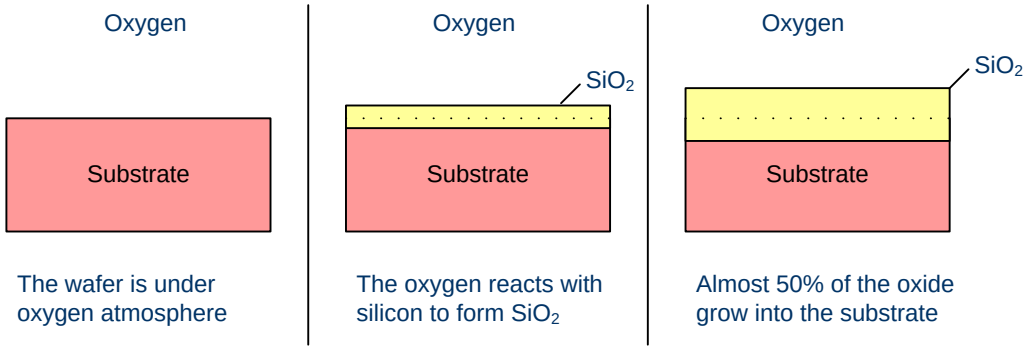


Fig. 1.2: Growth of silicon dioxide on top of silicon

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.

1.2.2 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:

$$\mathrm{SiH_4 + O_2 \longrightarrow SiO_2 + 2\,H_2}$$

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.

1.3 LOCOS process

1.3.1 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.

1.3.2 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.

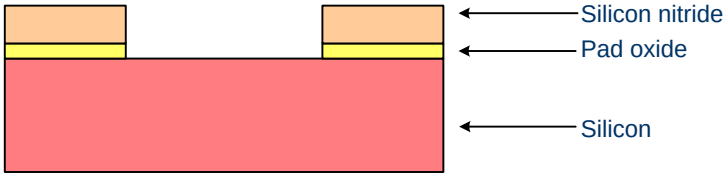
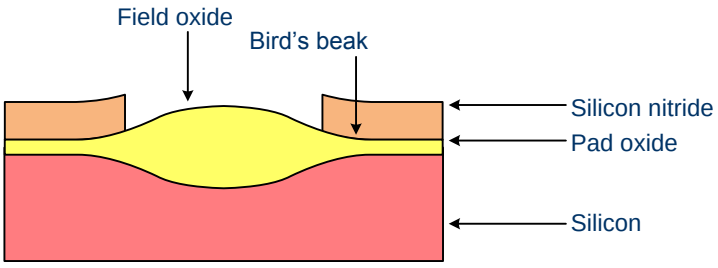
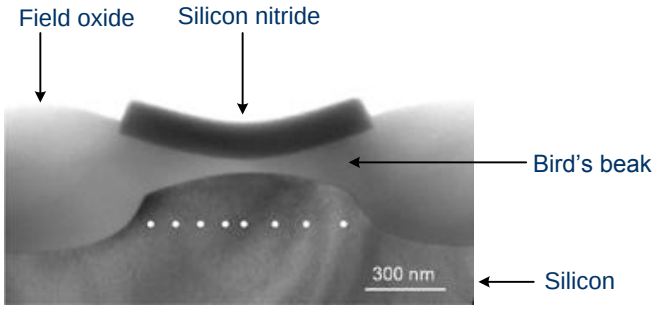


Fig. 1.3: Basic film layer stack

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. 1.4: 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.

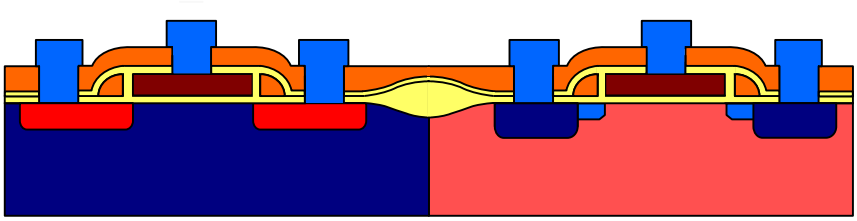


Fig. 1.5: Example of the LOCOS process for lateral isolation of transistors

1.3.3 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.

1.4 Film thickness measurement

1.4.1 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.

1.4.2 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.

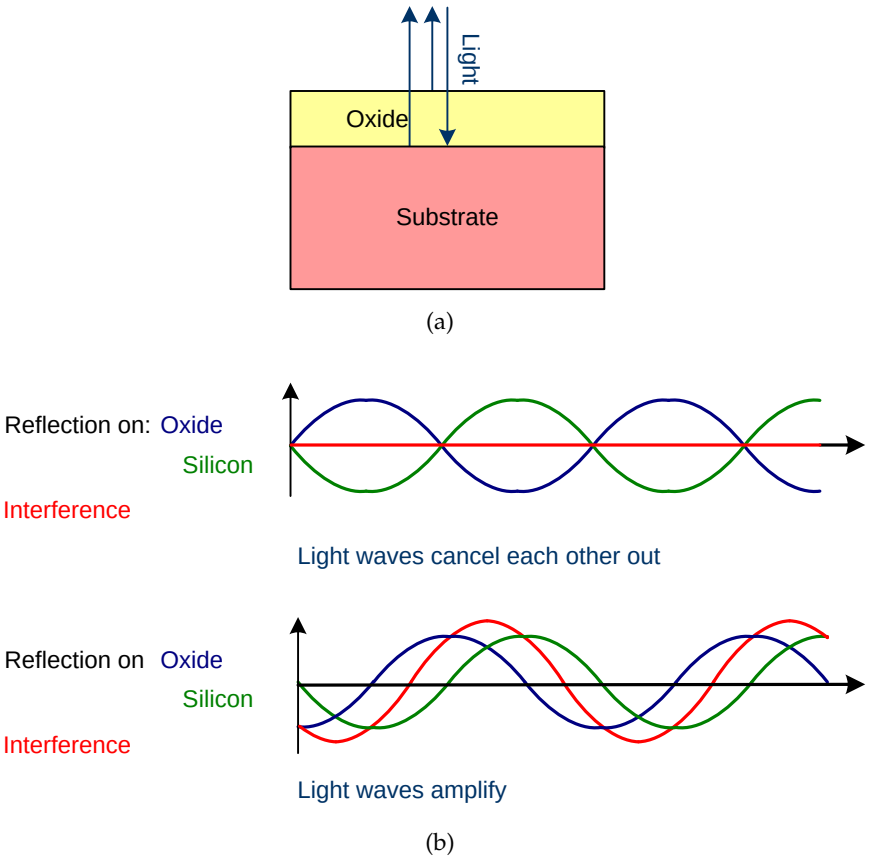


Fig. 1.6: (a) Interferometry, (b) destructive and constructive superimposition of lightwaves

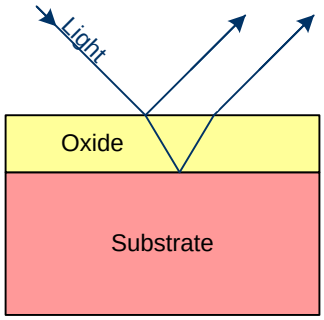
If light is irradiated onto a wafer some beams of light are reflected on top of it and some penetrate into the film. The latter will be reflected on bottom of this layer or penetrate into another layer beneath and so on.

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.

1.4.3 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. 1.7: 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.

1.4.4 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 %).

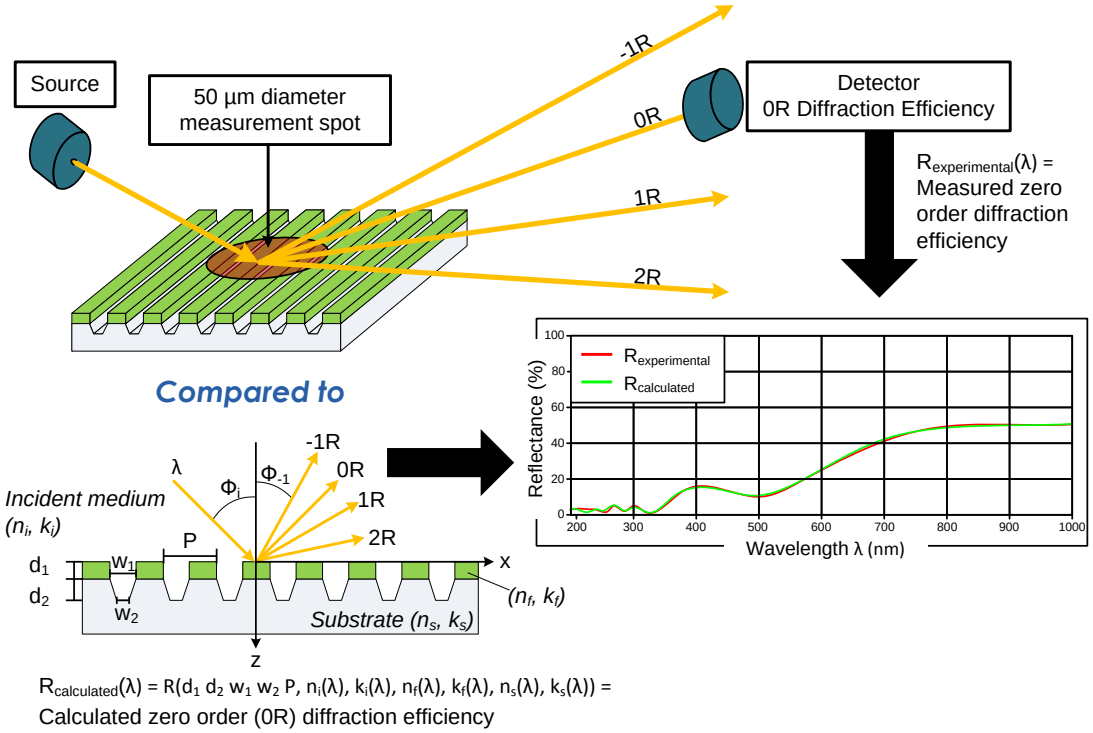


Fig. 1.8: Simulated and real measurement