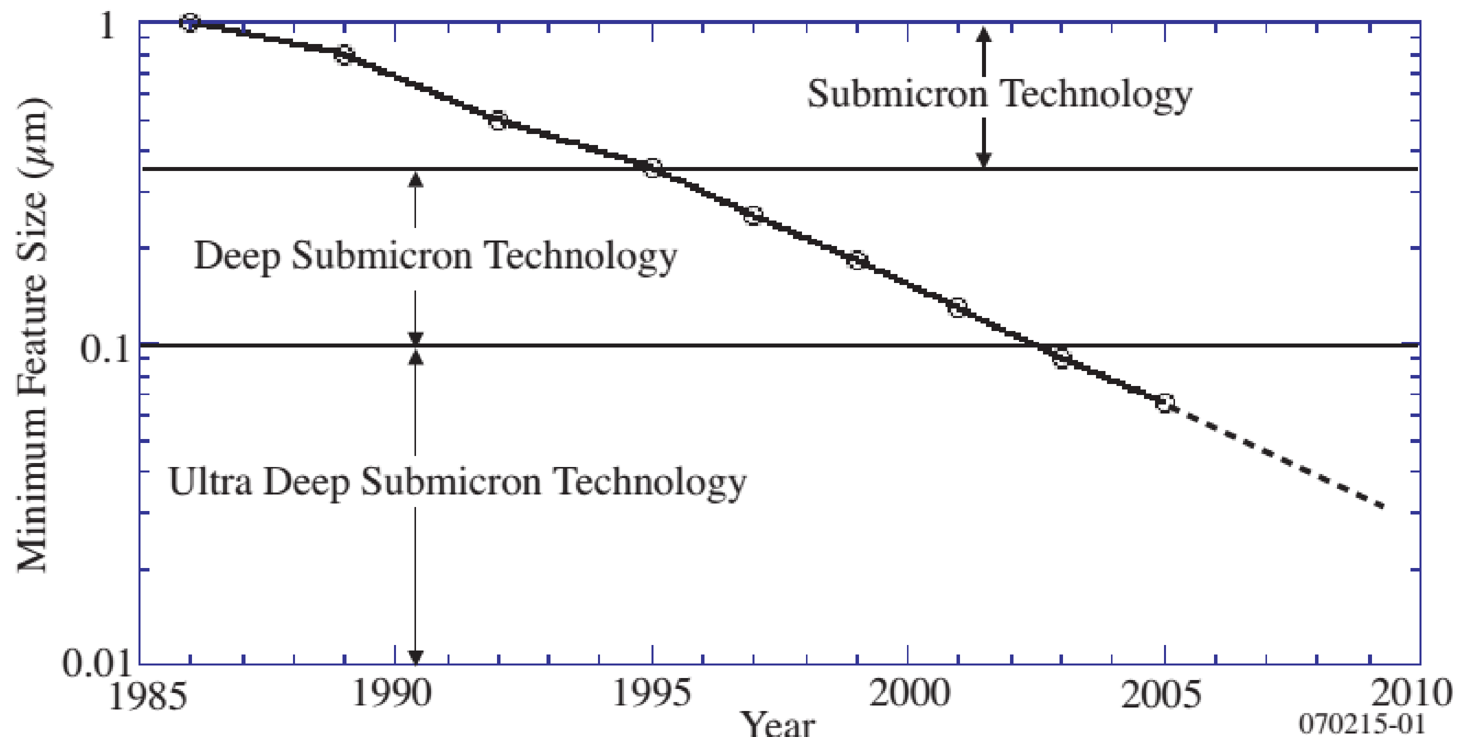
Deep and UltraDeep sub-micrometer MOS

Categorization of CMOS Technology

- **Minimum feature size as a function of time:**

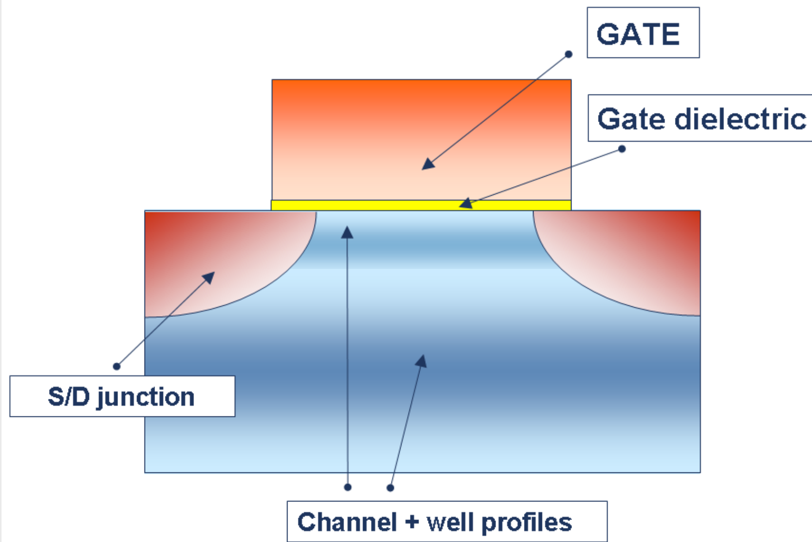


Categories of CMOS technology:

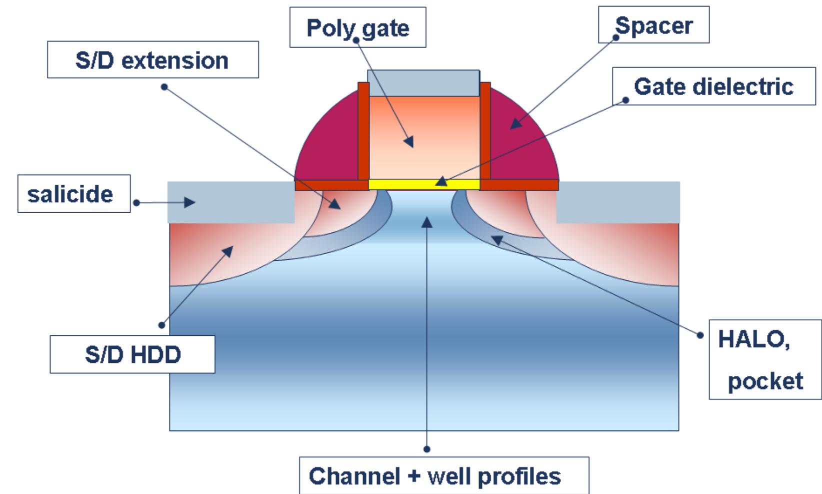
- Submicron technology – Lmin > 0.35 microns
- Deep Submicron technology (DSM) – 0.1 μm ≤ Lmin ≤ 0.35 μm
- Ultra-Deep Submicron technology (UDSM) – Lmin ≤ 0.1 μm

DSM MOS layout

'classical=texbook' MOSFET structure

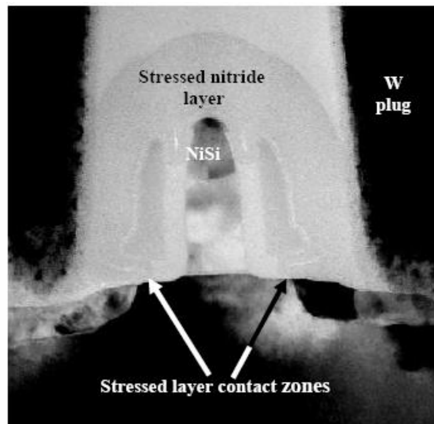


typical 'advanced' MOSFET Structure

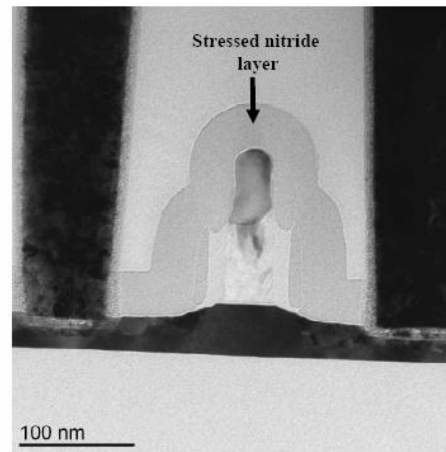


DSM MOS layout

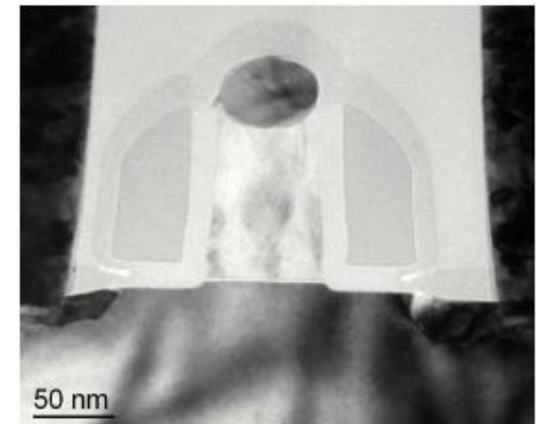
Different foundry = different device process!



Intel



IBM



TSMC

CMOS process flow :

front end (FE) and back end (BE)

DSM MOS process

Two different process flows are required:

- front end of line (FEOL) process flow
- back end of line (BEOL) process flow

FEOL the transistors are fabricated on the substrate

BEOL all the contacts and metallizations are generated

DSM MOS process

Advanced CMOS process flow

Active area definition
Well & channel doping
Gate electrode
S/D HALO / extensions
Spacers
HDD junctions
Silicides

→ **CMOS Front End of Line process**

PMD
Contact
Metal1
IMD1
Via1
Metal2
....

→ **CMOS Back End of Line process**

Techniques: oxidation

Oxidation is the process by which a layer of silicon dioxide is grown on the surface of a silicon wafer.

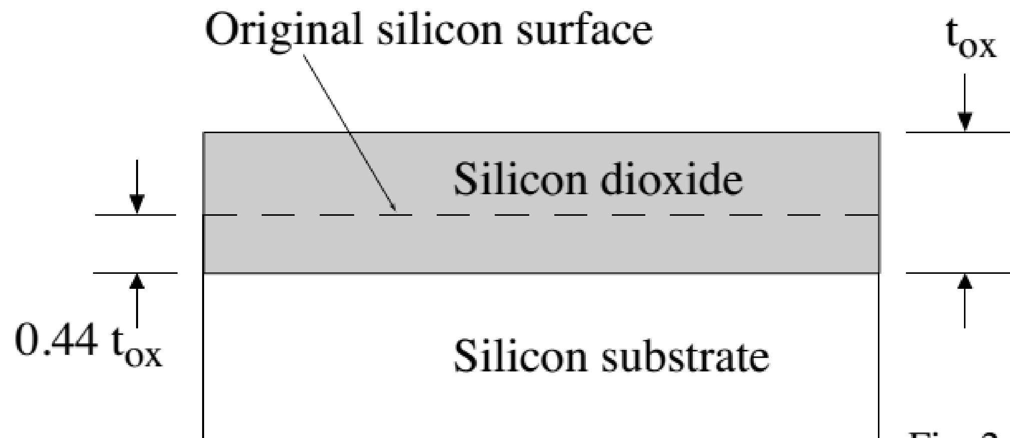


Fig. 2.1-2

Uses:

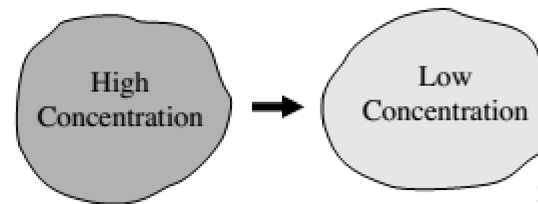
- Protect the underlying material from contamination
- Provide isolation between two layers.

Very thin oxides (100Å to 1000Å) are grown using dry oxidation techniques.

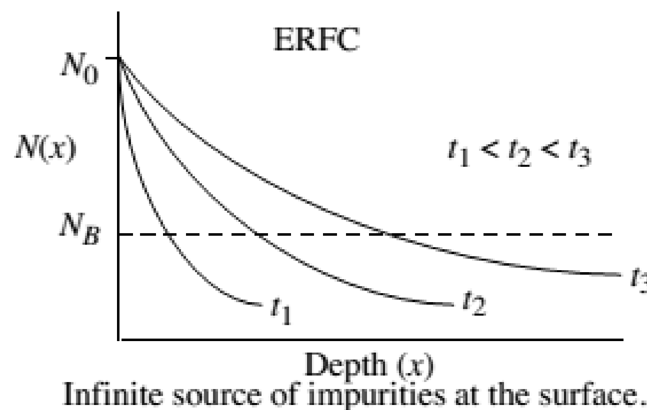
Thicker oxides (>1000Å) are grown using wet oxidation techniques.

Techniques: diffusion

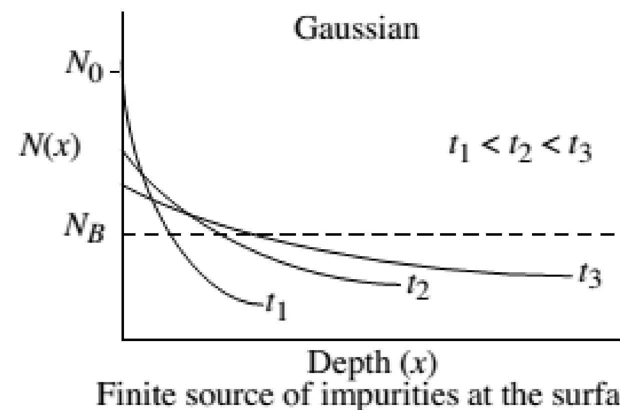
Diffusion is the **movement of impurity atoms** at the surface of the silicon into the bulk of the silicon. Always in the direction **from higher concentration to lower concentration.**



Diffusion is typically done at high temperatures: 800 to 1400°C



Infinite source of impurities at the surface.



Finite source of impurities at the surface.

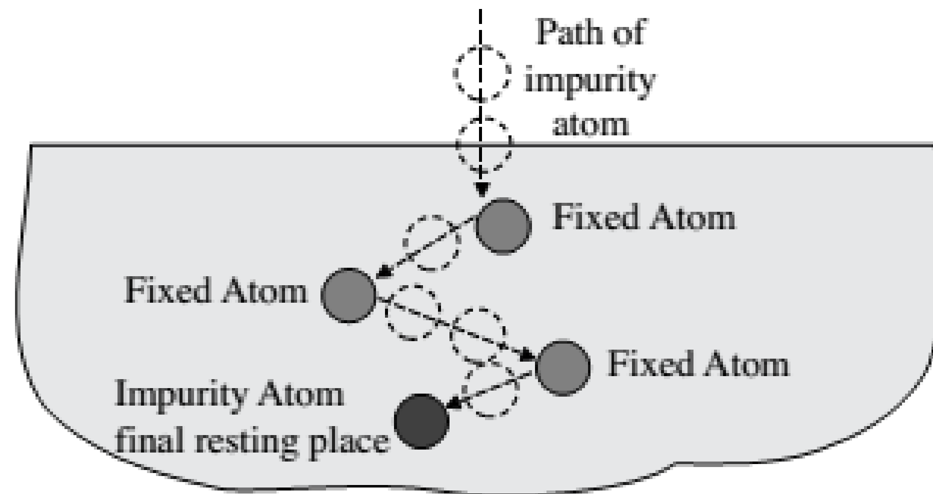
Fig. 150-05

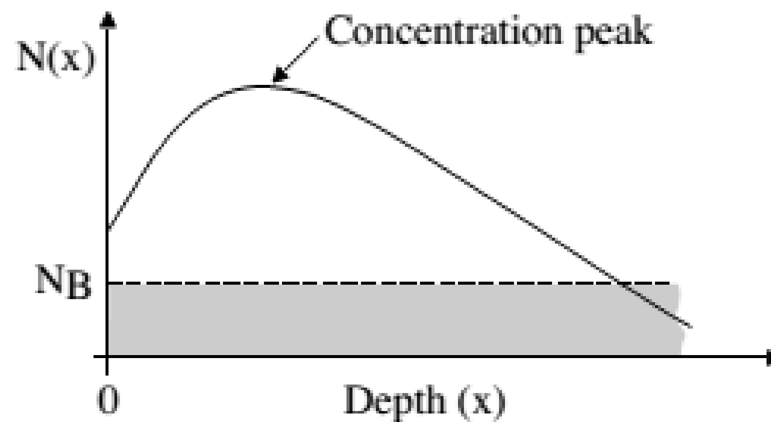
Techniques: ion implantation

Ion implantation is the process by which **impurity ions are accelerated** to a high velocity and physically lodged into the target material.

- **Annealing is required** to activate the impurity atoms and repair the physical damage to the crystal lattice. This step is done at 500 to 800°C.
- Ion implantation is a **lower temperature** process compared to diffusion.
- **Can implant through surface layers**, thus it is useful for field-threshold adjustment.
- **Can achieve unique doping profile** such as buried concentration peak.

Techniques: ion implantation





Techniques: depositions

Deposition is the means by which various materials are deposited on the silicon wafer.

- Silicon nitride (Si3N4)
- Silicon dioxide (SiO2)
- Aluminum
- Polysilicon

There are various ways to deposit a material on a substrate:

- Chemical-vapor deposition (CVD)
- Low-pressure chemical-vapor deposition (LPCVD)
- Plasma-assisted chemical-vapor deposition (PECVD)
- Sputter deposition

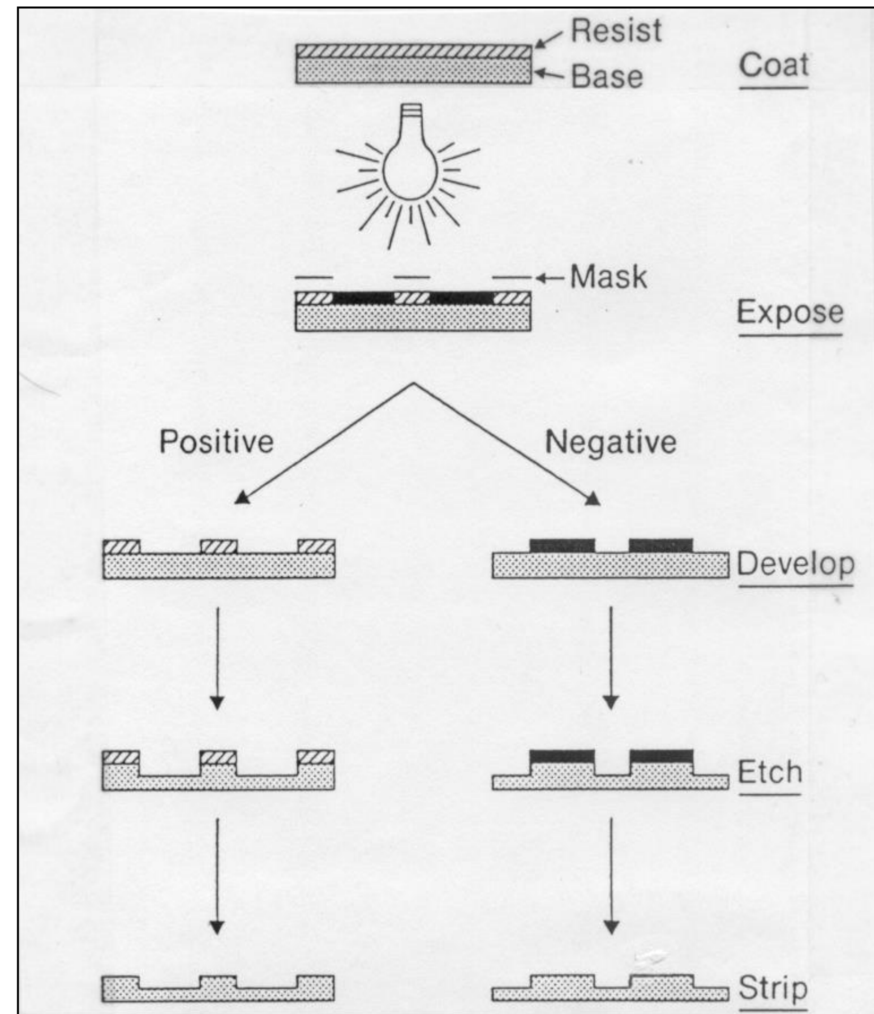
Material that is being deposited using these techniques covers the entire wafer and requires no mask.

Techniques: photolithography

CMOS processing requires only 3 'steps' !
(but repeated a lot of time…)

- deposition
- pattern definition (photo-lithography)
- remove material (etch)

$$resolution = k_1 . \frac{\lambda}{NA}$$



Techniques: etching

Purpose: remove material

Wet etch:

mostly **isotropic**, for ‘large’ patterns, **uncompatible with device scaling**

Chemical action, dip wafers in wet bath

Dry etch:

anisotropic trenches possible, compatible with high density,

plasma reactor, both physical and chemical etch involved

Techniques: dry etching

Dry etching techniques are used in micro-engineering to face the main drawbacks of **wet (purely chemical) etching**, which are:

- the presence of irreproducible disturbances from bubbles, flow pattern, etc.,
- handing with corrosive and toxic materials,
- limitation of the minimum achievable dimension,
- the impossibility to obtain vertical trenches.

Techniques: dry etching

Dry etching may be defined as the use of radical species and/or energetic ions from a cold plasma (or ion beam source) to **remove material where there is no mask material**.

The action of plasma etch is **both physical** (increased anisotropy and decreased selectivity) and **chemical** (more selective and isotropic).

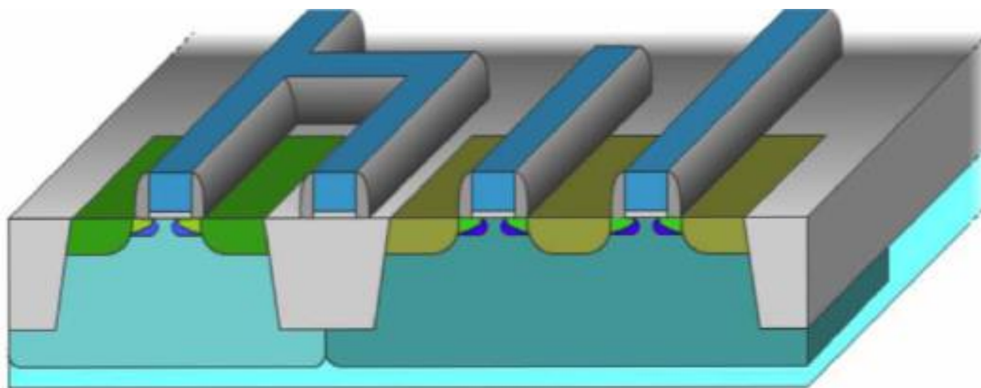
The main process variables are:

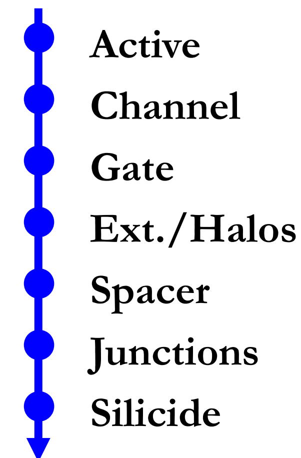
- gas mixture (affecting etch rate and profile shape),
- the flow rate (effect on etch rate through residence time),
- pressure (etch rate, profile shape, selectivity),
- RF power (etch rate through dislocation rate),
- wafer temperature (profile shape and chemistry),
- self-bias of the chuck (selectivity, etch rate and shape profile)

Front End Of Line (FEOL)

These are the 7 FEOL modules that will be introduced (and later correlated with electrical performance)

1. Active area module
2. Channel doping module
3. Gate electrode module
4. Source/drain extensions module
5. Spacer module
6. Junctions module
7. Silicide module





Front End Of Line (FEOL)

The active area module is the **first module** in the FEOL process flow.

The purpose of the active area module is twofold :

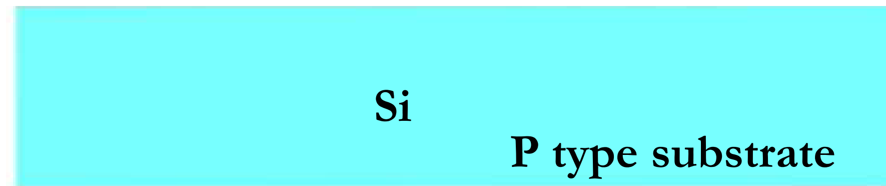
- **It defines the active regions.** These are the regions in which the (active) transistors will be realized.
- **It realizes good lateral insulation between the different active regions.** These lateral insulation areas are called the **field regions or field oxides, FOX** (SiO_2 dioxide, the oxide of silicon). They are necessary to insulate transistors from each other.

Shallow trench isolation

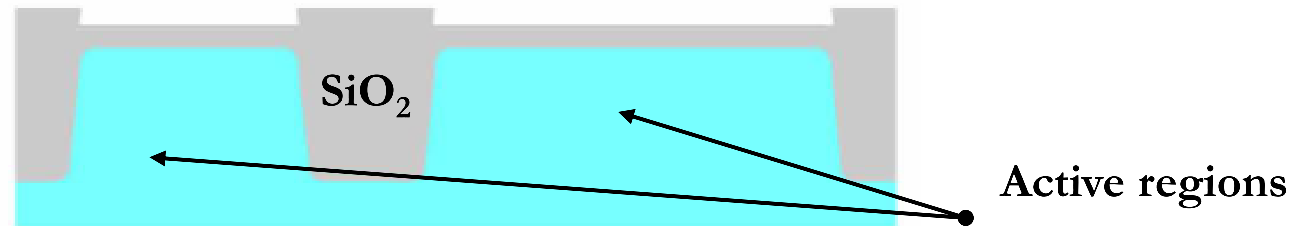
Purpose:

- Define Active regions where transistors are located.
- Good isolation between transistors & parasitic transistors formed by interconnecting poly-Si/metal stripes.

First



Last



Shallow trench isolation

Pad Oxide Growth

It all starts with the silicon wafer on which **a thin (15-20 nm) oxide is thermally grown**. This oxide layer is called the pad oxide and is a **buffer material.**

Depositing the **silicon nitride (hard material) immediately** on the silicon wafer **would cause internal stress, damaging the wafer.** Hence the purpose of **pad oxide is to absorb most of the stress.**

Afterwards a protective silicon nitride (SiN) layer is deposited on the wafer using a technique called "LPCVD" (LOW Pressure Chemical Vapour Deposition). **Silicon nitride is a strong and very hard ceramic.**

Shallow trench isolation

To create the trenches, the position where they will be located has to be defined

This is done using a **photolithography step.**

After the litho step, photo resist remains on top of the future active areas; these are the regions where the nitride should remain after the next step

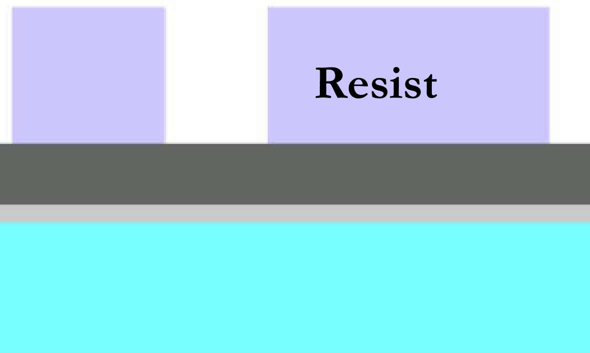
Therefore this step is called the active area litho step.

The dry etch step first removes the unprotected - there is no photo resist on it – nitride above the field regions.

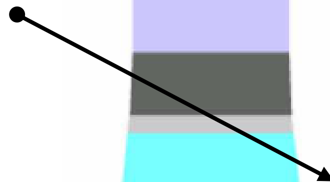
It etches through the pad oxide into the silicon wafer to form trenches

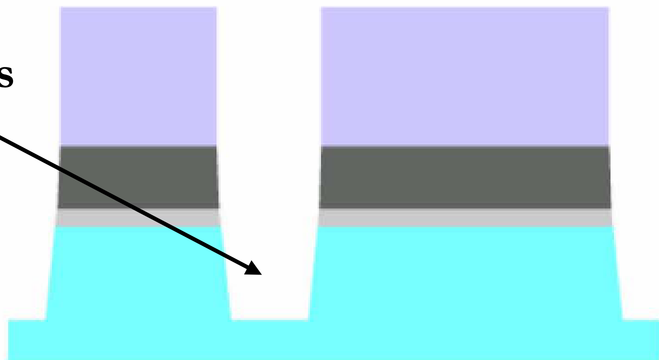
Shallow trench isolation

Pad oxide growth + nitride deposition

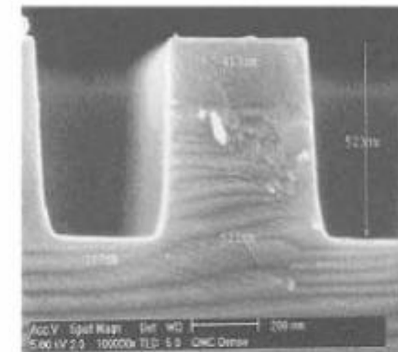


Active area lithography step





Etch



Shallow trench isolation

The resist is then removed. Depending on the fab, this can be done using various etch techniques: dry etch and wet etch or even a combination.

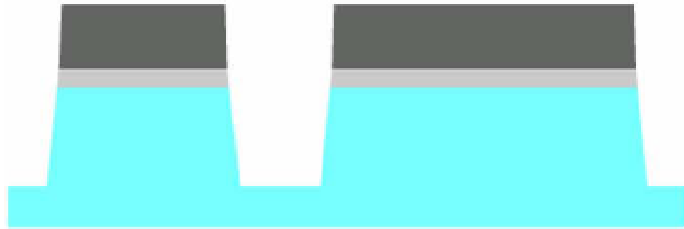
Re-Oxidation and Corner Rounding steps:

Rounded corners: HF (hydrofluoric acid) is a strong acid. A short dip in it followed by a **re-oxidation will remove the sharp corners.**

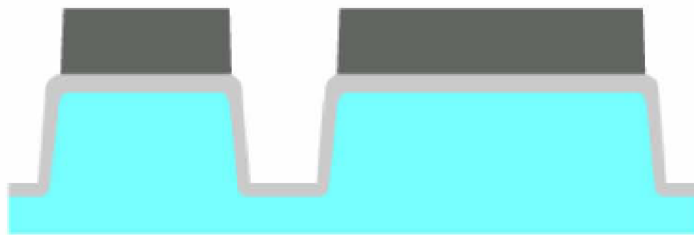
Damage repair: The silicon in the trenches is damaged by the dry-etch, causing imperfections in the lattice. **To get rid of the imperfections, the silicon is converted to oxide during the re-oxidation**.

Shallow trench isolation

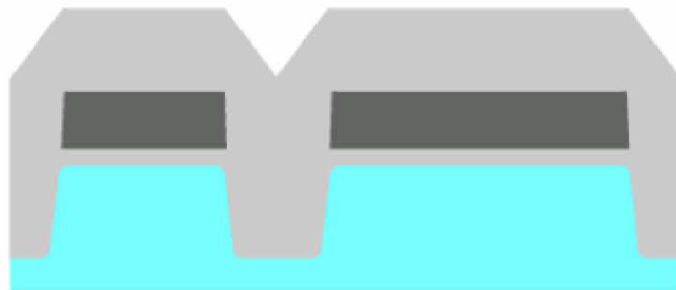
1) Resist removal



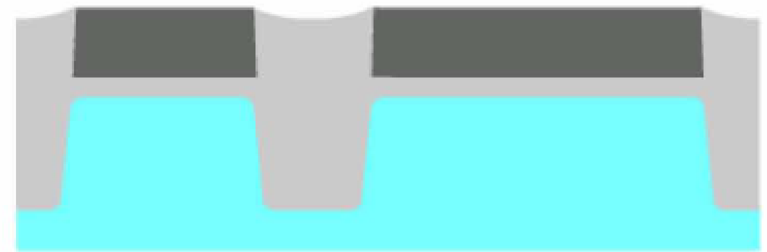
2) Sidewall oxidation, corner rounding



3) Deposition of trench filling oxide (HDP CVD)

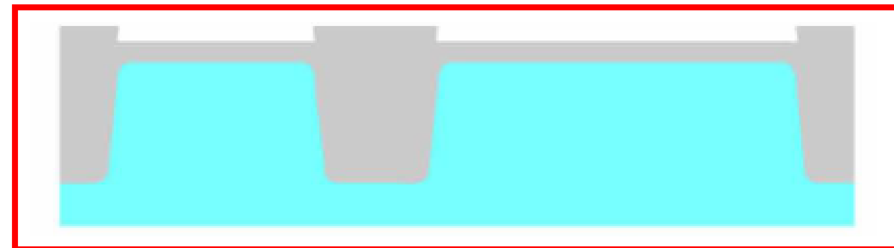


4) Oxide CMP polish step with stop on nitride



5) Field oxide recess

Removal of nitride layer



Channel doping module

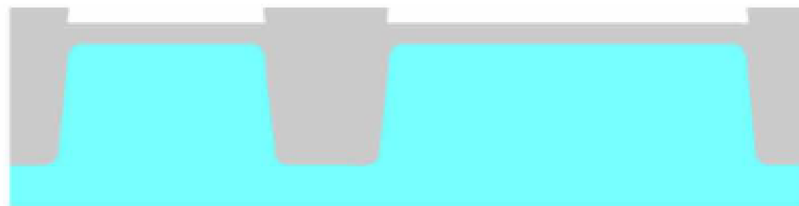
Purpose:

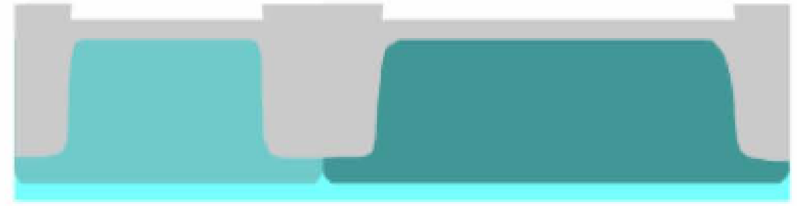
High Energy Implant

- Define P- & N-type regions for NMOS and PMOS
- Set high doping concentration underneath the field-oxide

Low Energy Implant

- threshold voltage adjustment
- leakage current suppression
- drive current optimization
- junction capacitance





Channel doping module

First Litho step

First Implants (n-type)

With the photo resist in place, **two successive n-type (phosphorous) implants are done.**

The first one is a *high-energy implant*.

These dopants have a high-energy and therefore they **can penetrate quite deep** into the silicon material.

- The high-energy implant results in the **bulk doping of the transistor** and in a high doping profile under the field oxides
- **Increasing the threshold voltage of the parasitic transistors.**

Channel doping module

The second one is a low-energy implant

Because of the low-energy, **the dopants will not penetrate deep into the wafer**

- The low-energy implant is required to **optimize the doping level in the channel region**, close to the surface.
- The photo resist protects the future p-regions from the n-dopants.

Channel doping module

Resist Strip

After the n-well is implemented, the photo resist is removed. Depending on the fab, this can be done using various etch techniques: both dry etch and wet etch or even a combination. The removal of the photo resist is called the **resist strip .**

The channel doping module is quite simple.

Second Litho step

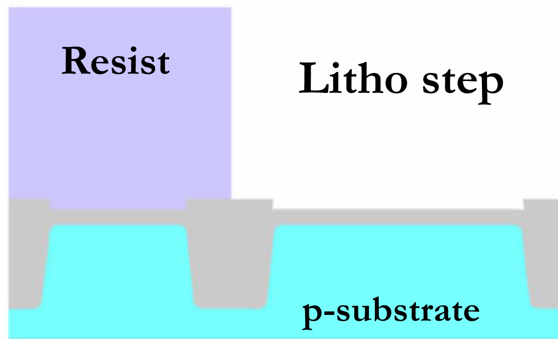
Second Implants (p-type)

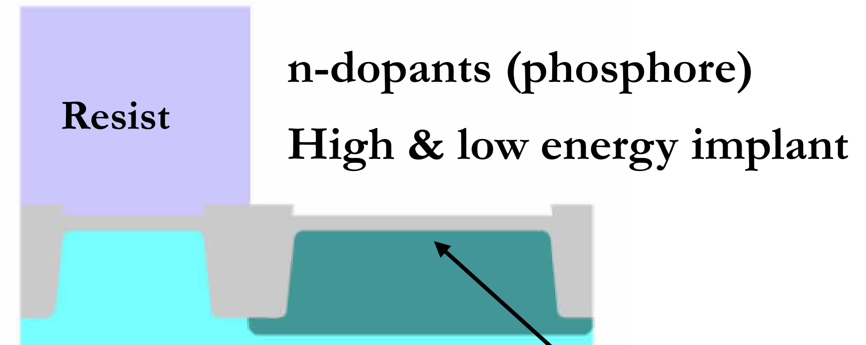
The implementation of the p-wells is **almost the same** as the implementation of the n-wells, except that this time the dopant is p-type, namely **boron (B)**

Again this is done in 2 steps: a high-energy and a low-energy implant.

Channel doping module

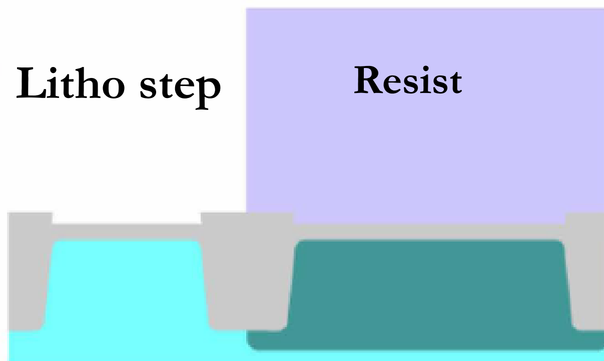
First Implants (n-type)



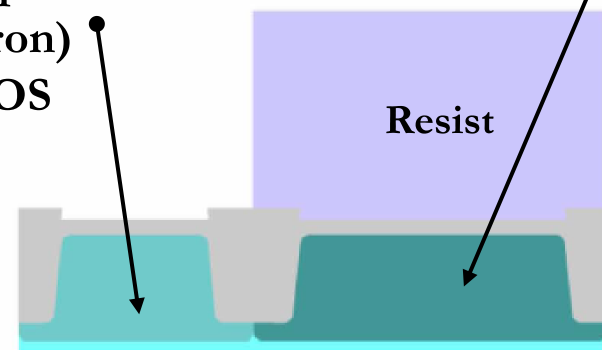


N type well
pMOS

Second Implants (p-type)



P type well
(boron)
nMOS



Gate stack module: gate oxide

Purpose:

Gate oxide growth

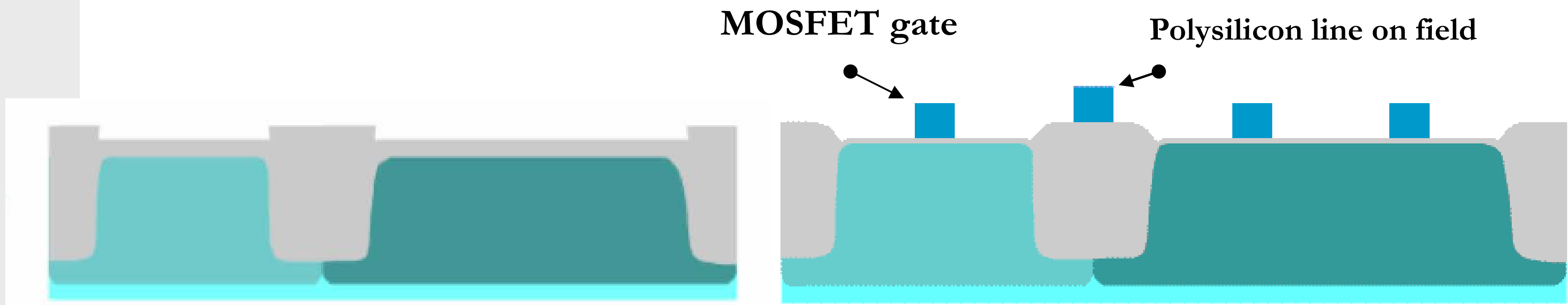
The gate oxide serves as the insulation between the gate electrode and the channel.

The gate oxide is often called gate dielectric or gate insulator.

Implementation of the gate electrodes

The gate electrode is the conductive part of the gate.

Nowadays it is **made of polysilicon** but it will be replaced by metal in future transistor generations



Gate stack module: gate oxide

In the gate electrode module a gate insulator made of oxide is grown and the gate electrode made of polysilicon is patterned.

The consecutive steps in the gate electrode module are:

- Re-oxidation
- Deposition of polysilicon
- Litho step
- Dry etch
- Resist strip

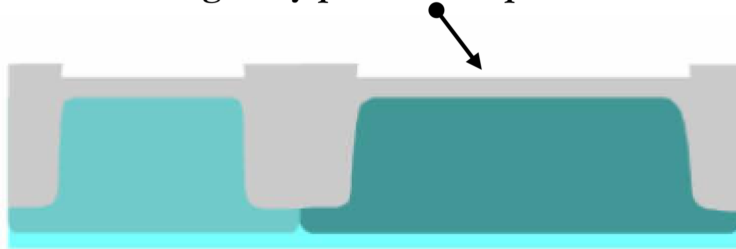
The litho step is quite critical:

Small variation in the dimensions of the photo resist will result in transistors with different gate length and thus different electrical characteristics

Gate stack module: gate oxide

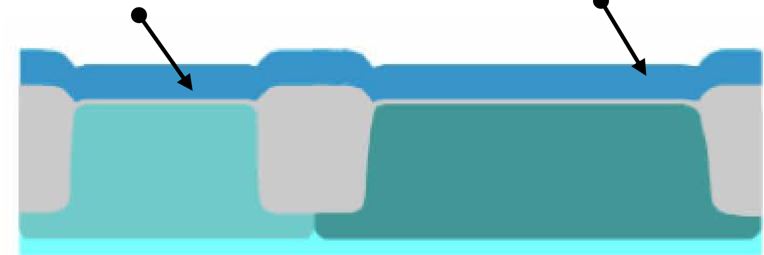
Re-oxidation

Oxide damaged by previous implantation

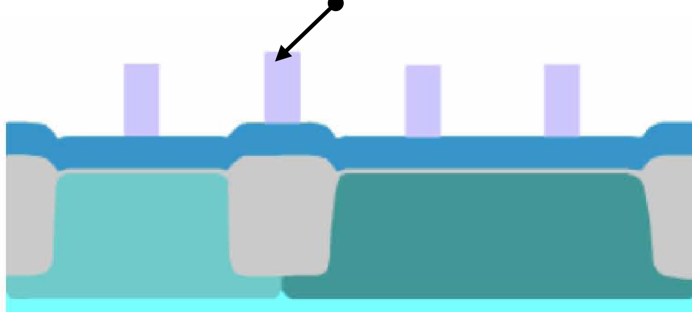


Removed by HF dip and new oxide is
Re-grow by thermal oxidation (high quality)

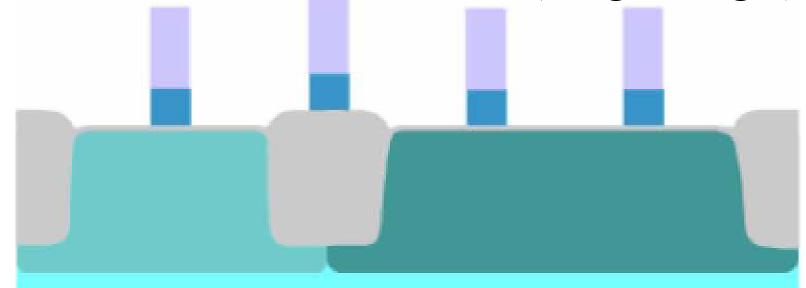
Polysilicon layer deposition



Litho



Dry etch



Gate stack module: gate oxide

Re-Oxidation

Re-oxidation is the first step when processing the gate electrode.

As there is a layer of oxide (the pad oxide) already present

What is the purpose of re-oxidation?

In the previous processing module dopants were "shot" into the wafer using an ion implanter → the pad oxide is damaged

However, **the quality of the gate insulator is very important**, It has a large influence on the **threshold voltage/gate leakage** current and reliability of the transistors

The damaged pad oxide is removed doing a HF (HF = hydrofluoric acid)

A thermal oxidation is done in order to grow an imperfection-free gate insulator.

Gate stack module: gate electrode

Deposition of Polysilicon

After the re-oxidation, **LPCVD** is used to deposit a layer of polysilicon on the wafer.

Afterwards a **dry etch** is done, to define the electrode

(remove the polysilicon and only the polysilicon gate electrodes remain)

N.B. the etch chemistry used during the dry etch must have an **as high as possible selectivity towards oxide** to prevent damaging the gate insulator.

Resist Strip

After the dry etch step, the photo resist is removed (stripped) and the gate electrode module is finished

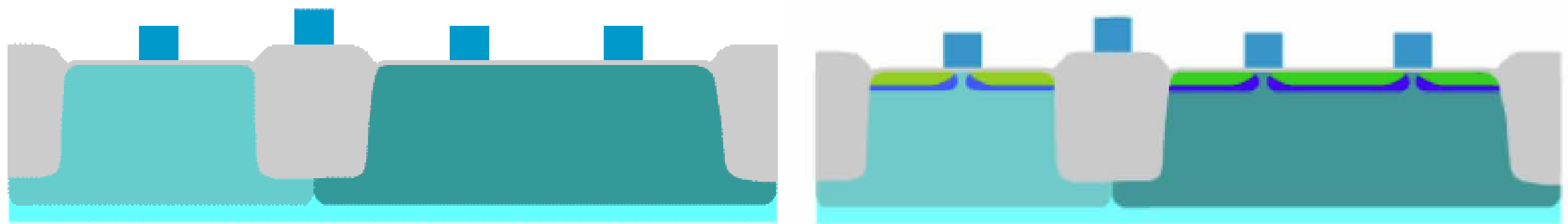
Source and drain extensions

Purpose :

- Minimize Short channel effects
- Optimize drive current/transistor performance (Rs, Rd,..)

How?

- Control maximum electrical field in the channel
- Optimize a 2-dimensional doping profile = trade off between SCE, series resistance, leakage current, drive current



Source and drain extensions

In the source/drain extensions module both source/drain extensions and **HALOs** (also called pockets) are implemented.

They are needed to suppress certain physical effects that start to play role when scaling down the transistor size.

Sometimes an **anti-TED (Temperature enhanced diffusion) anneal** is required in order to repair the damage caused by the previous implant.

The implant induced silicon defects (dislocations, interstitials, Vacancies..) are ideal paths for diffusion, especially during heating steps.

In recent technology research for low-temperature processes or fast anneal, in order to avoid diffusion (spike, laser,…).

Source and drain extensions

LDD implant: lightly doped extension and S/D regions

- ❑ limit series resistance between Source/Drain and channel region, optimize drive current
- ❑ limit maximum electrical field (hot carrier degradations)

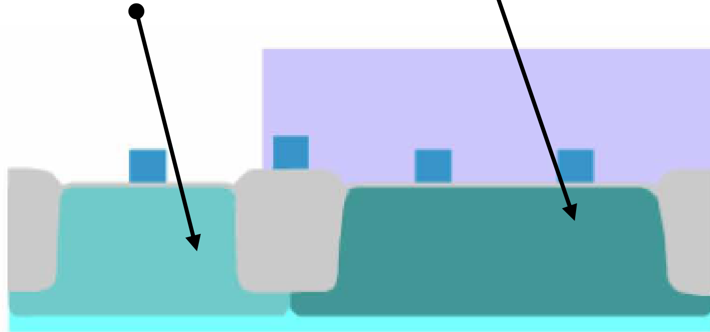
HALO or pocket implant:

- ❑ local increase of channel doping concentration around tip of LDDs.
- ❑ Suppress punch-through current
- ❑ Form an abrupt junction in length direction = stable effective channel length L_{eff}
- ❑ Reduced Vth roll-off but increased Reverse Short channel effects (RSCE)

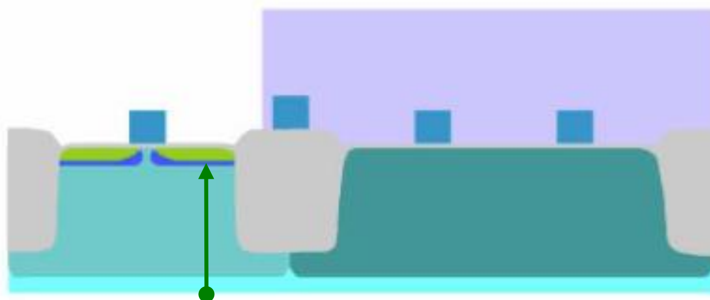
Source and drain extensions

P type well (boron)

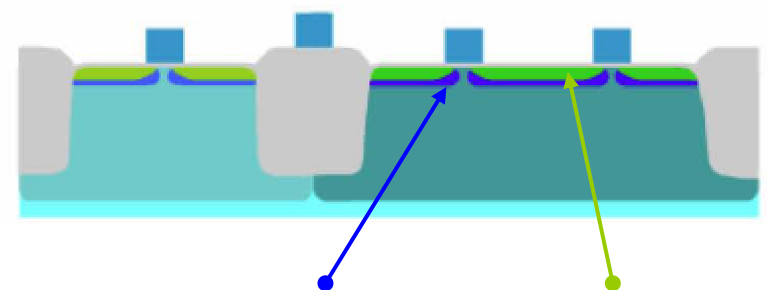
N type well (phosphore)



HALO P type implant (Boron) 23 deg < angle < 45deg)



Extension implant (Arsenic, n- type, angle~7deg)



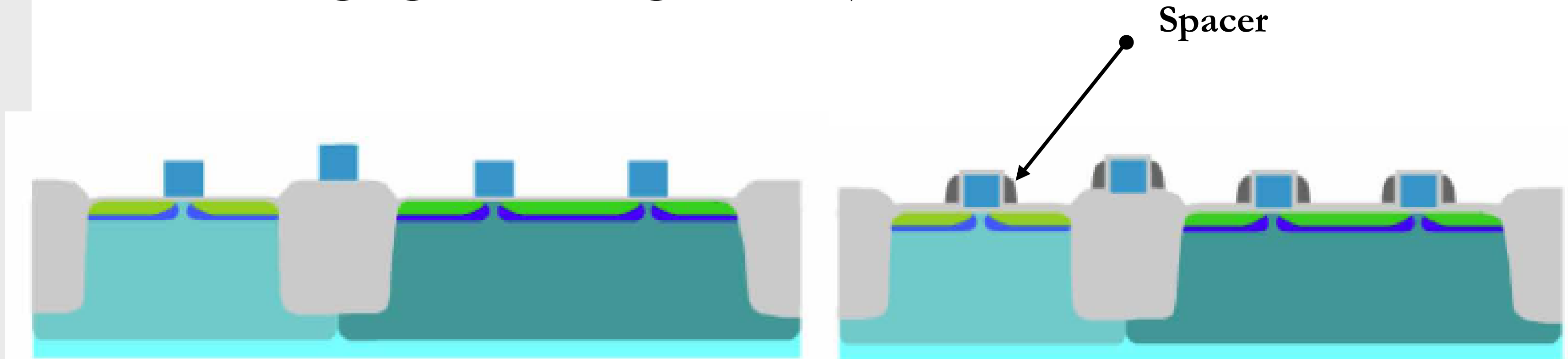
pMOS :HALO implant (Arsenic) Extention (Boron)

Space module

Purpose:

- Realize offset for highly doped junctions

- Avoid bridging between gate and junctions

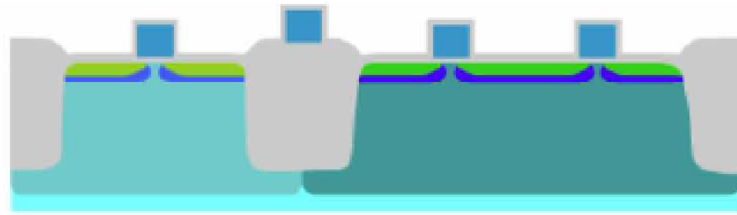


Spacers prevent dopants from diffusing under the gate during the implantation of the highly doped source/drain regions, overruling the source/drain extensions.

Spacers avoid bridging (short-circuiting) of the silicides of the gate electrodes and the source/drain regions.

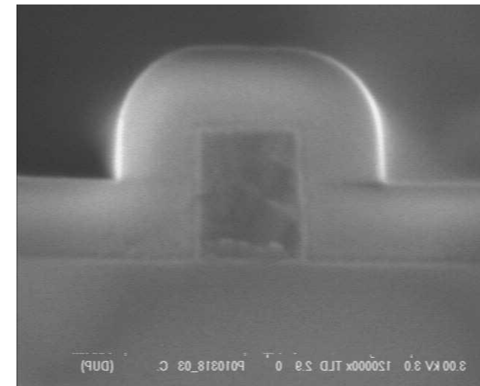
The silicides are implemented in the last module of the FEOL process flow: the silicide module

Space module

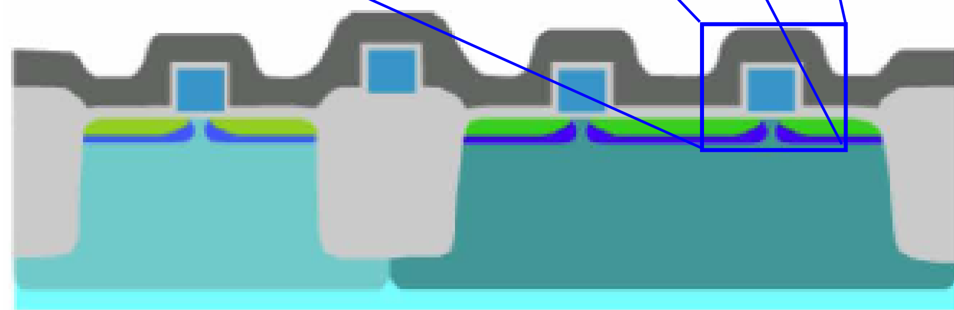


Oxide LPCV Deposition

(TEOS: stopping layer for nitride etch)

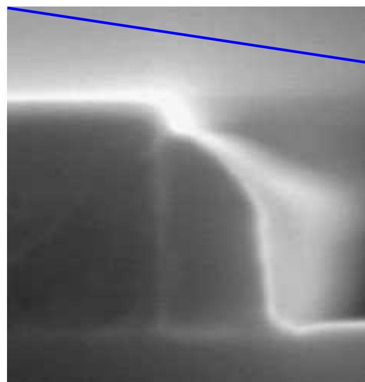


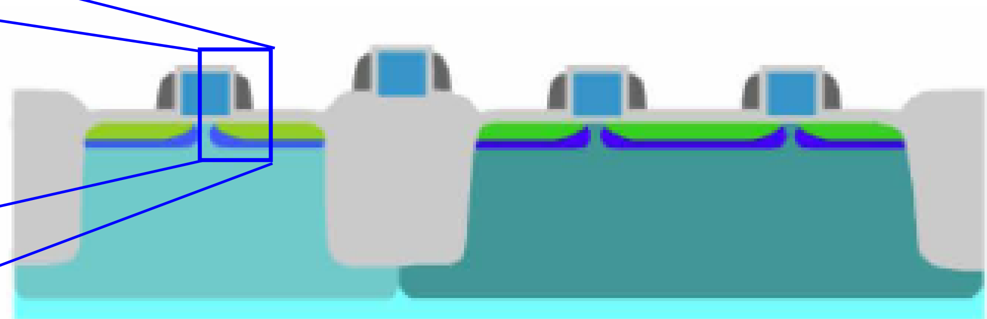
Nitride Deposition



Nitride etch

(TEOS protection and stopping layer)





Space module

LPCVD (Low Pressure Chemical Vapour Deposition) is a technique that uses a gas to deposit a layer on a surface

In this case LPCVD is used **to deposit a layer of oxide**

The **main component of the gas used is a chemical called TEOS** (Tetraethylorthosilicate), TEOS reacts with oxygen on the heated surface forming oxide.

This oxide is also needed to act as a stopping layer during dry etch later on.

Nitride Deposition

Again using LPCVD, a nitride layer is deposited. Of course this time the chemistry is different from the LPCVD of oxide.

Space module

Dry etch step

The key element of a dry etch is that **it is anisotropic: the nitride is etched much faster in the vertical direction than in the lateral direction.**

Stopped once the TEOS oxide layer is reached

Detecting when the TEOS oxide is reached during dry etch is relatively easy because TEOS oxide and nitride are very different materials

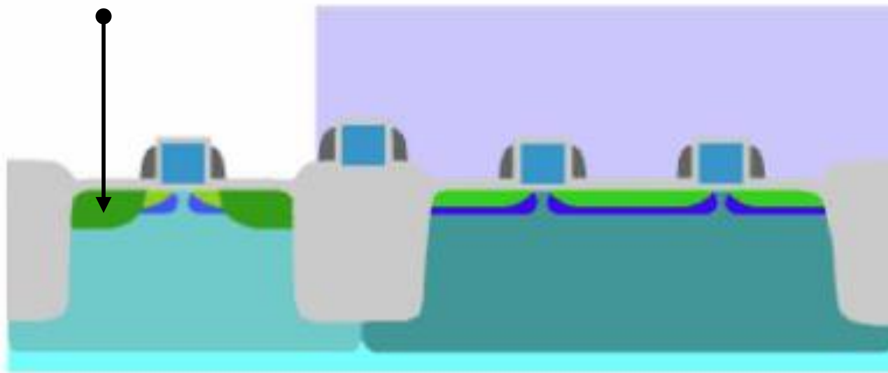
When the TEOS oxide is reached, there is still nitride present next to the gate

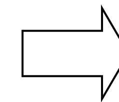
And here is a second advantage: **thanks to the TEOS Dry etch oxide, the silicon beneath is not damaged and thus no re-oxidation has to be done!**

A third advantage of nitride spacers is that, since this **dry etch step is selective to oxide**, **the field regions are not damaged**

Junction module

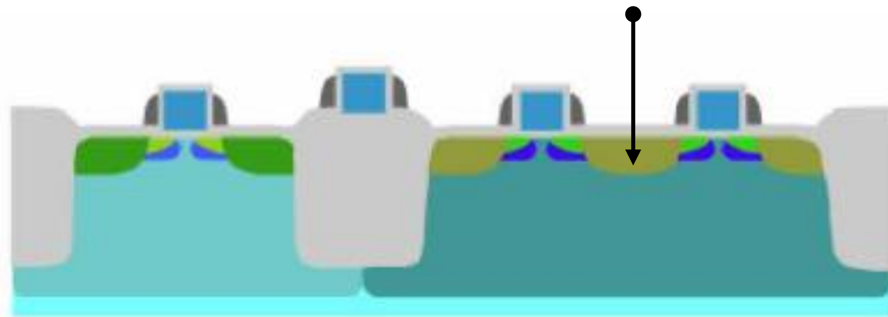
n- type implantation
Arsenic, phosphor

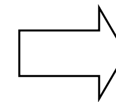




Resist strip
Oven anneal

p- type implantation
Boron





Resist strip
Rapid Thermal Process

Junction module

Step 1: **Litho Step**

Step 2: **implant NMOS**

With the photo resist in place, the n- type implantation is done using an **ion implanter**

The ion implanter shoots the dopants into the wafer. Both **arsenic and phosphor** are n- type elements and can both be used, depending on the transistor optimization.

The dopants are stopped in the thick field oxides and in the spacers.

The spacers create an offset between the gate edge and the source/drain regions.

The dopants are also implanted in the polysilicon gate electrode of the nMOS transistors.

Step 3: **Resist Strip**

Junction module

Step 4**: Anneal (oven) to activate the dopants**

Because the ion implanter shoot the dopants into the wafer, the dopants will occupy random spaces between the silicon atoms in the lattice.

However, **to become electrically active the dopants have to become part of the lattice**, replacing silicon atoms.

The same sequence for p-MOSFETs, junction implant is used.

The p-dopants (boron) are implanted using the ion implanter.

Again, some dopants are stopped in the photo resist, in the field oxides and also in the polysilicon gate electrode of the pMOS transistors.

After the resist strip heating is done for a very short time with an RTP - Rapid Thermal *Processing - step* .

An oven anneal would cause problems: p type dopants are very small (certainly compared to n-type dopants) and therefore diffuse easily, especially when heated.

Salicide module

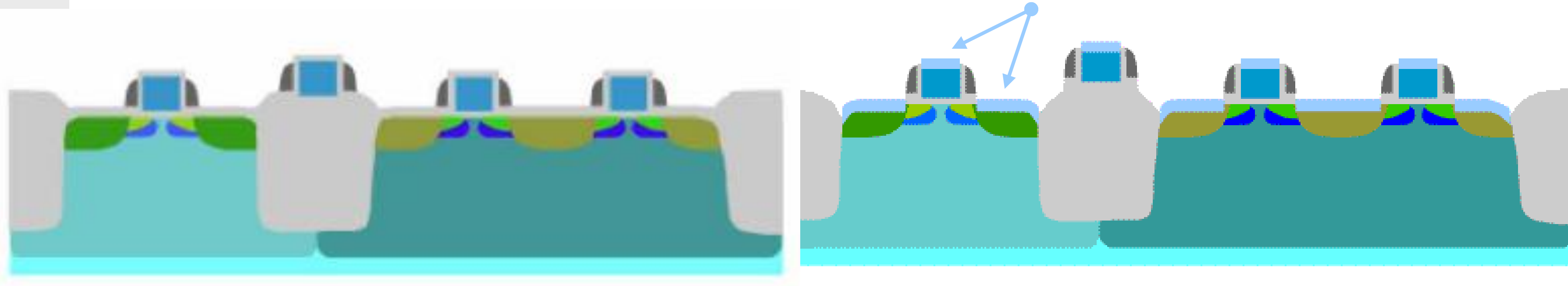
SALICIDE = Self-Aligned Silicides

- Silicide has to be self-aligned to gate and S/D area
- no silicide growth on spacers and isolation regions (= short circuits)

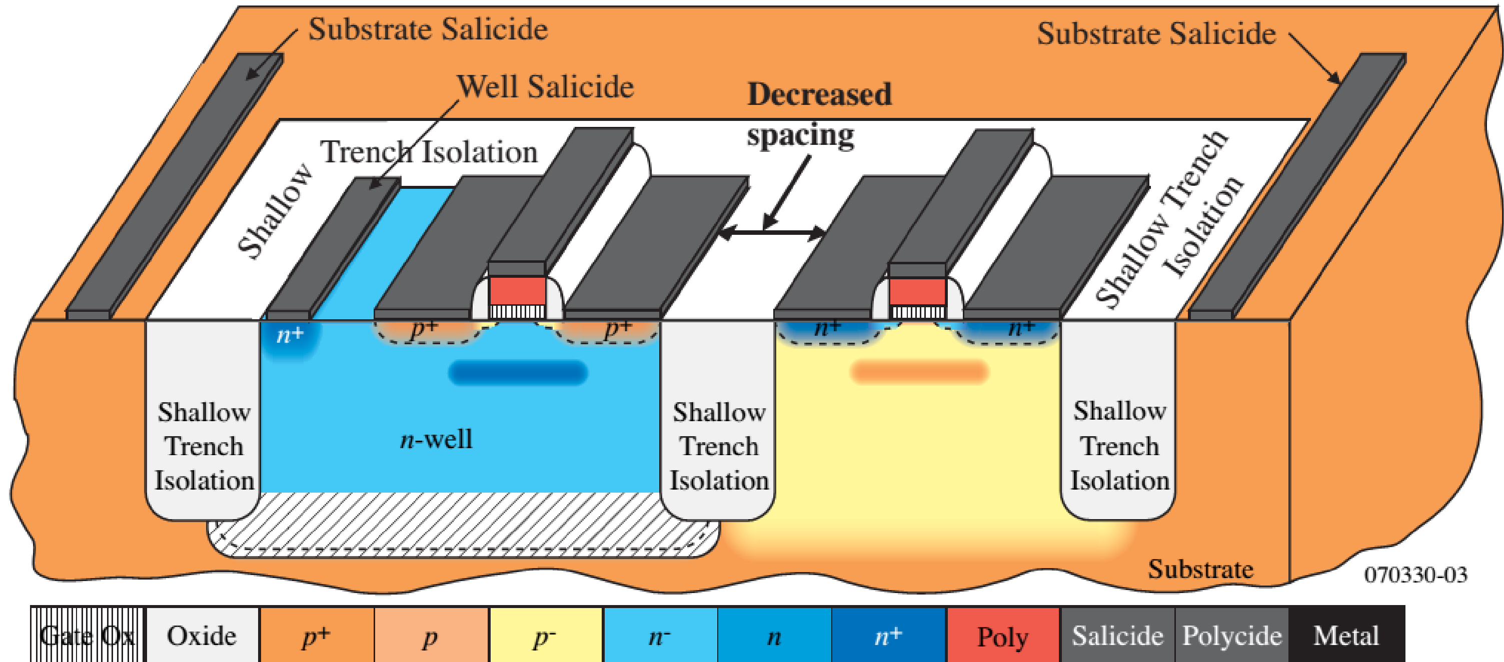
The silicide module serves as a means to lower source/drain/gate resistance

Silicide also avoid current crowding

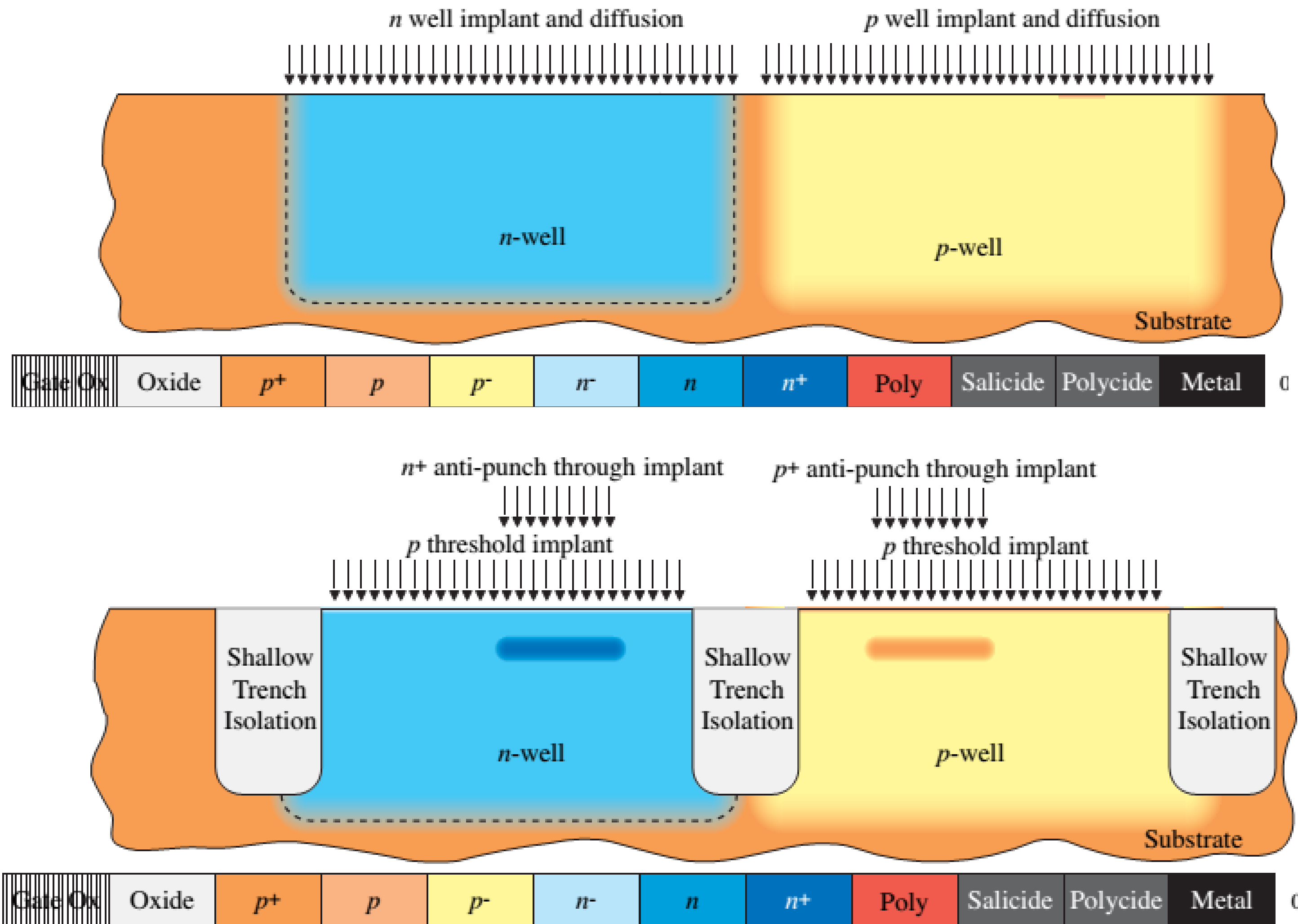
Silicide on the source, drain and gate



Anti-punch through implants

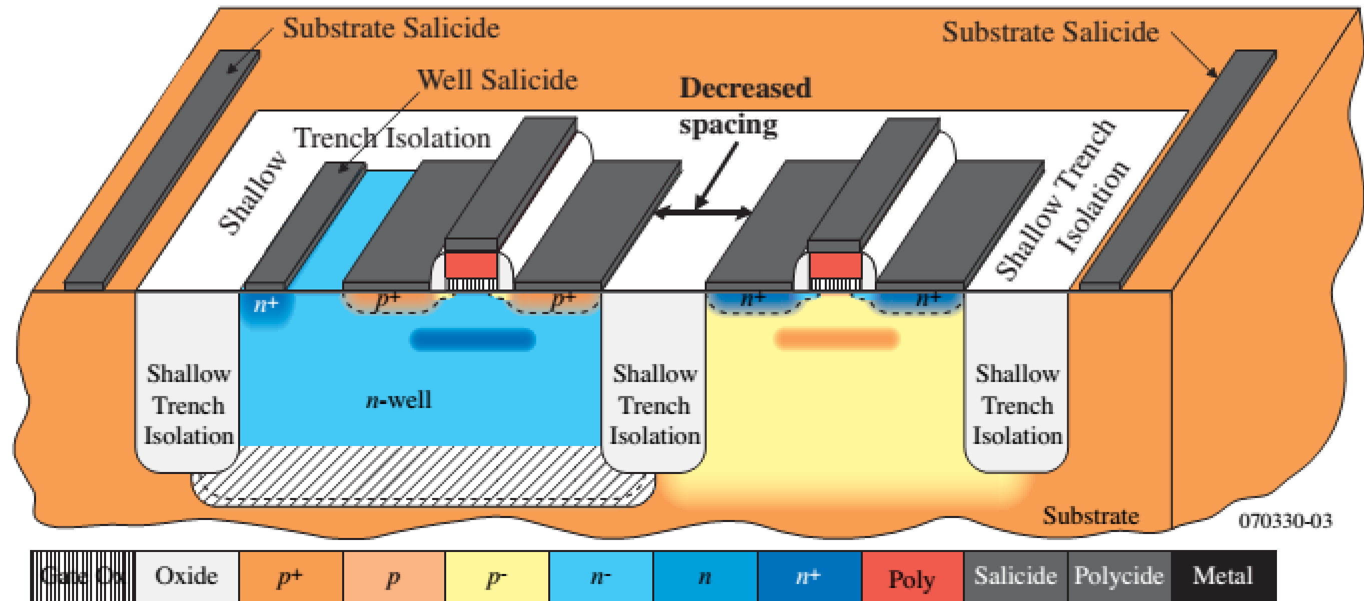


Anti-punch through implants



Also an implant can be applied to create a higher-doped region beneath the channels to prevent punch-through from the drain depletion region extending to source depletion region

Anti-punch through implants



Also an implant can be applied to create a higher-doped region beneath the channels to prevent punch-through from the drain depletion region extending to source depletion region

Why Salicide?

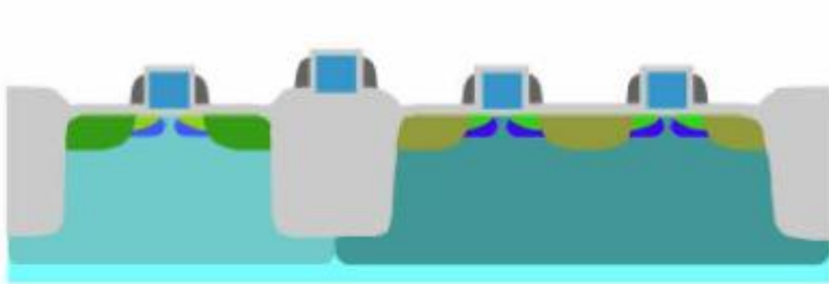
Silicides **lower the contact resistance** between the transistor and the contacts implemented in the BEOL.

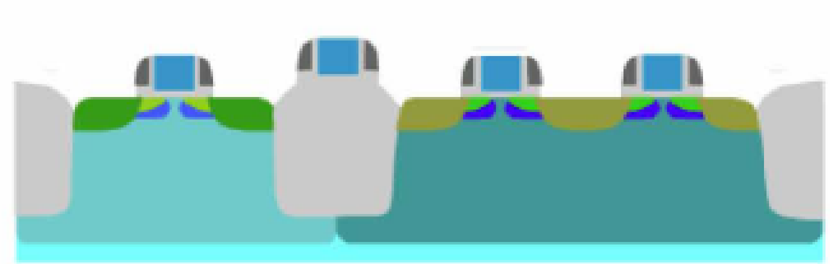
The salicide stands for self aligned silicide: the silicide is formed at the gate source and drain areas at the same time and does not require any litho step.

A silicide is an allow of silicon and a metal (Ti, Co, Ni..) obtained by a high temperature step.

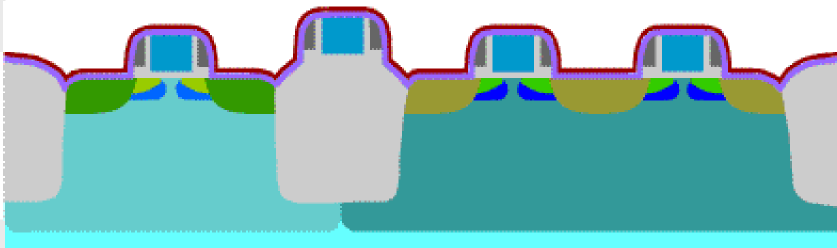
The silicide has a much lower resistivity then the doped silicon (typically a few Ω/sq)

Salicide module

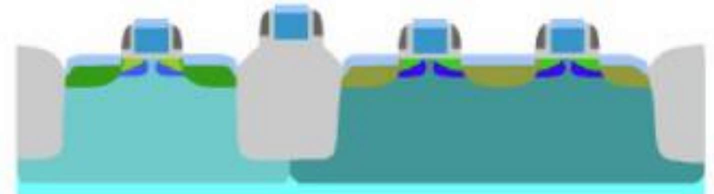




Wet Etch (HF)= Removal of oxide on top of gate, source and drain area



Metal (Co or Ni) is sputtered



Rapid thermal process (RTP1) to allow the reaction metal silicon and selective etch to remove the un-reacted metal

Salicide module

Fabrication steps:

First of all **a HF (hydrofluoric) dip** is done to remove the damaged oxide above the source/drain regions.

It is unavoidable that also a small part of the field oxide is removed.

It is especially a serious problem if too much of the spacers is removed because this can result in bridging.

Nitride spacers do not have this problem and this is one of the reasons why they are used nowadays for small dimensions.

After the oxide is removed, a metal is sputtered onto the wafer surface.

In older technology nodes (before 250 nm) this was usually **titanium and cobalt (130nm).**

For current technology nodes, **nickel** gives better results.

The metal will - in a later process step - react with the silicon beneath, resulting in nickel silicides.

Salicide module

RTP 1

Bridging effect. After sputtering, a first rapid thermal processing (RTP) step is required.

Due to the heat, **a chemical reaction will occur between the first sputtered metal layer and the silicon beneath**.

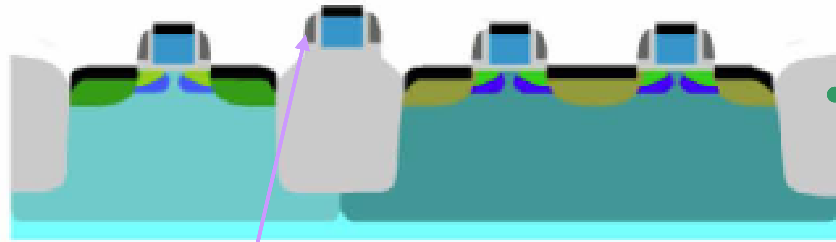
If there is too much heating, the silicided regions will grow into each other. **They will bridge.**

Wet etch

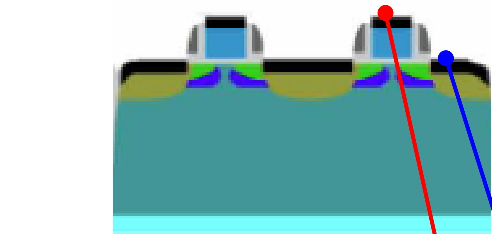
Now both the un-reacted metal and the cap can be removed using a wet etch.

A dry etch is necessary when the etching has to go in only one direction. But in this case, etching has to happen in all directions. Therefore this is a wet etch.

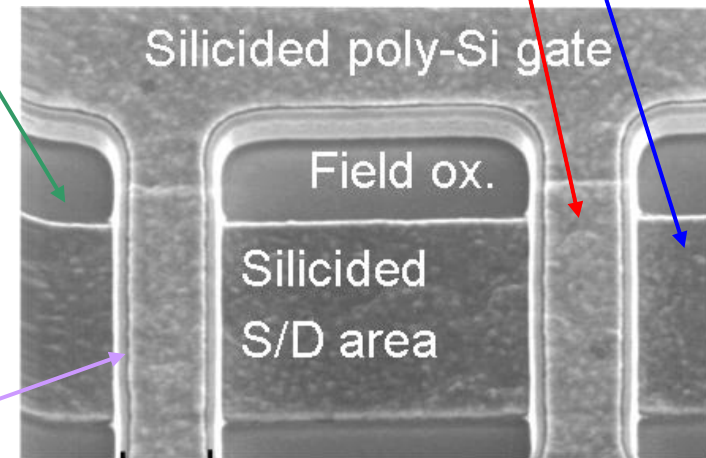
Salicide module



Rapid thermal process (RTP2)



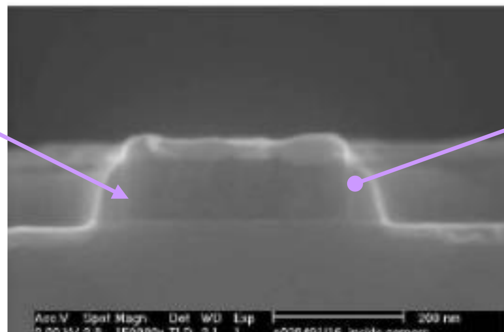
Top view



Sidewall spacers

Cross section

Spacer



Salicide module

RTP2

In order to avoid bridging, the duration of the RTP 1 step was limited. However, as a result, the structure of the silicides is far from optimal.

Therefore a second RTP treatment, called **RTP2, is done when the silicide region are far apart after the selective etch.**

The RTP2 step **improves the conductance** of the silicided regions significantly.

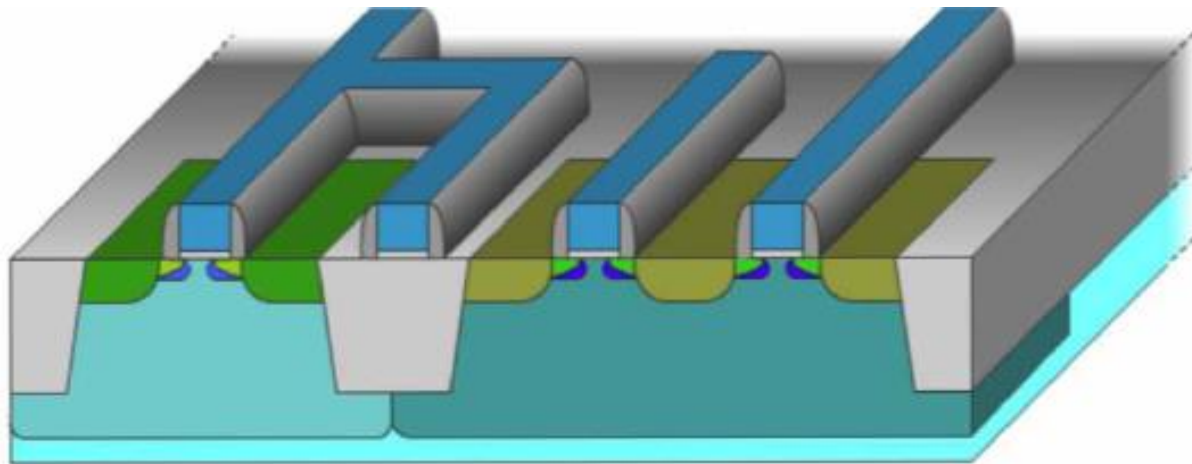
Salicide module

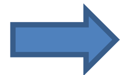
Summary

- Silicides are implemented for various reasons.
- The most important one being a decrease in resistance.
- Several materials have been used to create these silicides. Nowadays cobalt silicides are the most common.
- An important phenomenon that can occur is bridging.
- Implementing 2 heating steps instead of one avoids this.

FEOL - the end

Front end of Line



Next step  **Back end of line**