

Libre Silicon process specification

David Lanzendörfer

March 10, 2018

Abstract

Copyright © 2017 LANCEVILLE TECHNOLOGY GROUP CO., LIMITED. All rights reserved.

This process is licensed under the Libre Silicon public license; you can redistribute it and/or modify it under the terms of the Libre Silicon public license as published by the Libre Silicon alliance, either version 1 of the License, or (at your option) any later version.

This design is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the Libre Silicon Public License for more details.

This is the specification of the free silicon manufacturing standard for manufacturing the LibreSilicon standard logic cells¹ and related free technology nodes from the LibreSilicon project.

For this initial revision 0.1 a gate-first approach has been chosen which led to the choice of polysilicon as the gate electrode material because of the simplicity of the gate alignment. For better isolation properties of the transistors and gates in overall a box-isolation approach has been chosen. All of these choices have been made with the future scale down from the recent $1\mu m$ to smaller structure sizes. **This process is for manufacturing $1\mu m$ only!** But further releases which will have been tested with smaller structure sizes can be expected.

¹<https://github.com/chipforge/StdCellLib>

Contents

1 CMOS in a nutshell	4
2 Physics	6
2.1 Getting doping from resistance	6
2.2 Infusion	7
2.3 Constant source diffusion (Predeposition)	8
2.4 Ion implant	10
2.5 Drive-in (limited source diffusion)	10
2.6 Vertical diffusion and junction formation (Well formation)	11
2.7 MOS Capacitance	12
2.8 Threshold voltage with metal gate (V_T)	12
2.9 Threshold voltage with polysilicon gate(V_T)	14
2.10 Threshold voltage (V_T) adjustment	17
3 Chemistry	18
3.1 Etching silicon dioxide	18
3.2 Etching silicon nitride	19
3.3 Growing silicon nitride	19
4 Process design	20
4.1 Substrate	21
4.2 Isolation	22
4.3 Interconnect	22
4.4 MOS gate material	23
4.5 MOS gate thickness	23
4.5.1 Subthreshold leakage	23
4.5.2 Gate tunneling current	24
4.6 NMOS threshold	25
4.7 PMOS threshold	27
4.8 MOS gate verification	28
4.9 Alignment strategy	29
5 Simulation with parameters	31
6 Process steps	32
6.1 Shallow trench isolation	33
6.1.1 Initial cleaning	34
6.1.2 Sulfuric Cleaning	34
6.1.3 HF dip	34
6.1.4 Pad oxide	34
6.1.5 Nitride layer	34
6.1.6 Patterning positive	35
6.1.7 Nitride etching	35
6.1.8 Resist removal	35
6.1.9 Silicon etching	35
6.1.10 Deep oxidation	35
6.1.11 Oxide deposition	36
6.1.12 Hard mask removal	36
6.2 N-well	37
6.2.1 Mask dioxide layer	38
6.2.2 Patterning	38
6.2.3 Etching	38
6.2.4 Cleaning	39
6.2.5 Injection	39
6.2.6 Oxide for drive-in	39
6.2.7 Drive-in	39
6.2.8 Oxide mask removal	40
6.3 P-well	41
6.3.1 Mask dioxide layer	42
6.3.2 Patterning	42

6.3.3	Etching	42
6.3.4	Cleaning	43
6.3.5	Injection	43
6.3.6	Oxide for drive-in	43
6.3.7	Drive-in	43
6.3.8	Oxide mask removal	44
6.4	Gate	45
6.4.1	Gate oxide deposition	45
6.4.2	Polysilicon deposition	45
6.4.3	Patterning	45
6.4.4	Etching	46
6.4.5	Cleaning	46
6.5	n+ Implant	47
6.5.1	Mask dioxide layer	47
6.5.2	Pattering	47
6.5.3	Etching	48
6.5.4	Cleaning	48
6.5.5	Injection	48
6.5.6	Oxide removal	48
6.6	p+ Implant	49
6.6.1	Mask dioxide layer	49
6.6.2	Pattering	49
6.6.3	Etching	50
6.6.4	Cleaning	50
6.6.5	Injection	50
6.6.6	Oxide removal	50
6.7	Silicification	51
6.7.1	Oxide deposition	52
6.7.2	Silicide block patterning (optional)	52
6.7.3	Sputter etching(Spacers)	52
6.7.4	Titanium deposition	53
6.7.5	First reaction step	53
6.7.6	Etch	53
6.7.7	Second reaction step	53
6.8	First vias	54
6.9	First metal layer	55
6.10	Additional vias	56
6.11	Additional metal layer	57
7	Testing	58
7.1	Diodes	59
7.2	Bipolar transistors	60
8	Design rules	61
8.1	Lithographic limitations	61
9	Machines	62
9.1	Intertech ISI-2808 Laser Direct Write System	62
9.2	ASML Stepper (PHT-S1)	62
9.3	DRIE Etcher #1 (DRY-Si-1)	62
9.4	Buehler Polisher (CMP)	63
10	List of Footnotes	64

1 CMOS in a nutshell

This basic initial project is dedicated to the CMOS Technology only and for this reason two types of metal–oxide–semiconductor field-effect transistors (MOSFET) are required.

Historically, the first chips with MOSFETs on the mass market were p-channel MOSFETs in enhancement-mode.

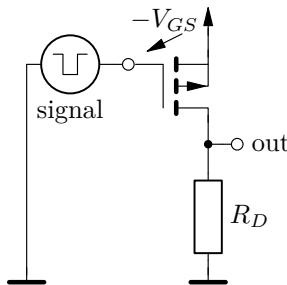


Figure 1: enhancement-mode PMOS transistor use-case

The sectional view of a PMOS transistor in silicon is shown below

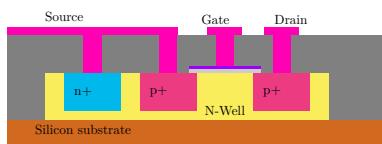


Figure 2: Sectional view of a PMOS transistor

Historically later, faster chips with MOSFETs on the mass market were marked as n-channel MOSFETs in enhancement mode also.

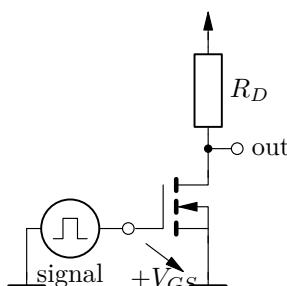


Figure 3: enhancement-mode NMOS transistor use-case

The sectional view of a NMOS transistor in silicon is shown here also.

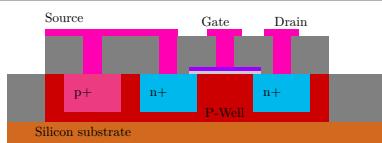


Figure 4: Sectional view of a NMOS transistor

Both technologies, the older PMOS as the newer NMOS, have the same disadvantage. Every time, the transistor is switched on, the current between drain and source of the transistor is limited by the resistor on drain only. Higher currents here means higher power consumption for the chip where the transistors are integrated as well. If the transistors are switched off, no current flows between drain and source anymore, the power consumption of the chip also goes low. Et violá, the US-Patent with Number 3356858² changed the world and combines both technologies to the new complementary metal-oxide-semiconductor (CMOS) technology. Instead of every transistor working against a weak resistor, the transistor works against a complementary switched-off transistor. With the eyes of our antecessor CMOS doubles the transistor count, but contemporary chips all are built in CMOS.

²<https://www.google.com/patents/US3356858>

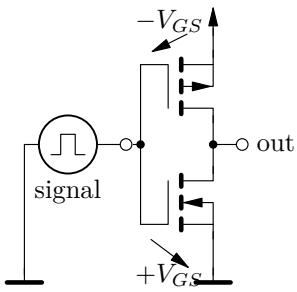


Figure 5: complementary PMOS and NMOS transistor couple use-case

Below the sectional view of the inverter circuitry can be seen. For the run through of this process we will use this cross section diagram as reference.

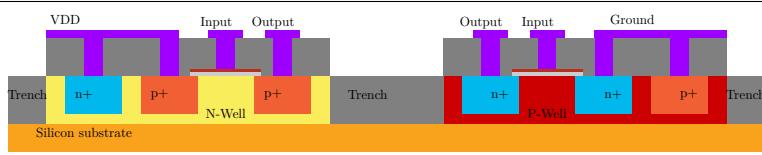


Figure 6: Sectional view of a NMOS-PMOS transistor circuit

2 Physics

In this chapter we deal with all the physics related to solid state device manufacturing. In case there is anything unclear, please look up this chapter and its sub-chapters.

2.1 Getting doping from resistance

In many cases the supplier will only provide the resistance per length specification for their substrate and won't give you the dopant concentration numbers. In this case you will have to find these numbers out yourself by converting it from the numbers they've provided.

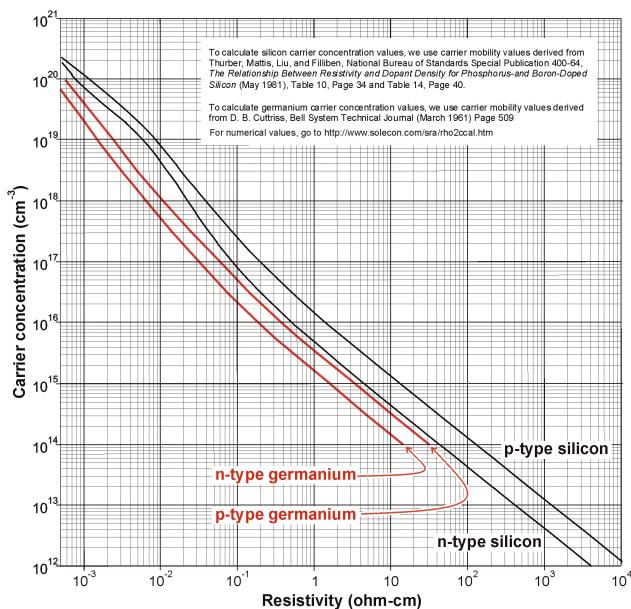


Figure 7: R-L-dopant relation

You can either use the graphics from Figure 7 and determine the dopant concentration graphically, which is very very imprecise or use a online tool like the one from Solecon³

Germanium is being included in this graphics just in case someone is going to fork this process based on Germanium substrate.

³<http://www.solecon.com/sra/rho2ccal.htm>

2.2 Infusion

The redistribution process depends on the ratio of the solubility of the doping material in silicon and SiO_2 . At the Si/SiO_2 interface the dopants are redistributed by segregation until the ratio of their concentration at the interface is the same as the ratio of their solubility in both materials. The ratio of dopant solubility is expressed by the segregation coefficient m which is

$$m = \frac{\text{solubility in silicon}}{\text{solubility in } \text{SiO}_2} \quad (1)$$

As listed in [Table 1](#) below there are dopant species which solubilize better in SiO_2 than in silicon ($m < 1$) and species which have a reversed behavior ($m > 1$). In case of $m < 1$, as for Boron, the dopant concentration is enhanced at the SiO_2 side, whereas beneath the interface, there is a dopant depletion at the silicon surface. For reversed solubility ratios ($m > 1$, like Phosphorus), only few dopant atoms penetrate the interface. In order to obtain the by m determined concentration ratio at the interface, dopant atoms from deeper silicon zones diffuse back to the surface zone. Therefore, the dopant concentration at the silicon surface is enhanced, as illustrated in [Figure 8b](#). In [Figure 8](#), C_c denotes the dopant concentration in the silicon surface zone before oxidation. x is the distance from the silicon surface.

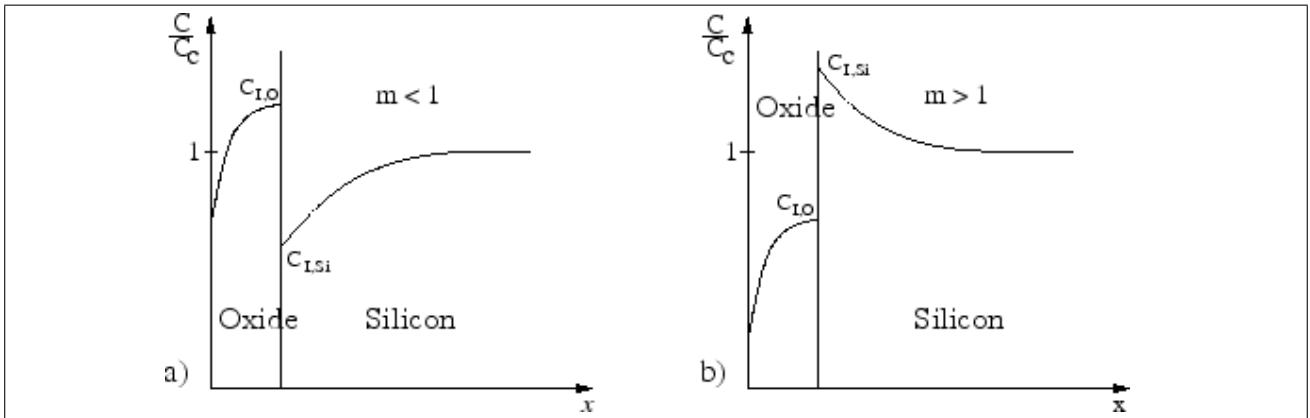


Figure 8: Schematic illustration of dopant redistribution

Dopant species	Boron	Phosphor	Antimon	Arsen	Gallium
m	0.1-0.3	10	10	10	20

Table 1: Segregation coefficients m for important dopant species in silicon

2.3 Constant source diffusion (Predeposition)

Although the diffusion process of donors and acceptors into the silicon crystal is a three dimensional process for simplicity we first only discuss the one dimensional mathematics for it in order to get a "simple" equation for the depth-time-temperature relation.

This is only valid for a constant source of dopants on the surface of the wafer (gas, for instance). These equations are used for predicting the pre-deposition step (in case this process would be adapted by someone for predeposition instead of ion implant)

We start with Ficks⁴ law (for all German speakers: Yes that's his name) where the dopant concentration N is coupled with time and place

$$\frac{\partial N}{\partial t} = D \cdot \frac{\partial^2 N}{\partial x^2} \quad (2)$$

The diffusion coefficient is as well material as well as temperature dependent and can be calculated with the following equation:

$$D = D_0 \cdot \exp\left(-\frac{E_a}{k \cdot T}\right) \quad (3)$$

With $k = 8.62 \cdot 10^{-5} \frac{eV}{K}$ being the Boltzman constant and in table 5 we can see the D_0 and E_a values for the most common materials⁵ which we can use within the further calculations for our well dimensioning phases. The temperature usually is in the area of $1000^\circ C$ or $1273.15 K$.

Element	$D_0 \left[\frac{cm^2}{s} \right]$	$E_a [eV]$
P	10.50	3.69
As	0.32	3.56
Sb	5.60	3.95
B	10.50	3.69
Al	8.00	3.47
Ga	3.60	3.51
Cu	0.0025	0.65

Table 2: D_0 and E_a values for boron and phosphorus

The law stated above

$$\frac{\partial N}{\partial t} = D \cdot \frac{\partial^2 N}{\partial x^2} \quad (4)$$

has the same form as the temperature conductivity equation (Laplace) for which we already have a general solution

$$\frac{\partial u}{\partial t} = a^2 \cdot \frac{\partial^2 u}{\partial x^2} \quad (5)$$

Which means that we can map the general solution for the temperature conductivity equations after Laplace

$$u(x, t) = \frac{1}{2 \cdot a \cdot \sqrt{\pi \cdot t}} \cdot \int_{-\infty}^{\infty} f(a) \cdot \exp\left(\frac{-(x-a)^2}{4 \cdot a^2 \cdot t^2}\right) da \quad (6)$$

to our Ficks law with $a = \sqrt{D}$ and $u = N$

$$N(x, t) = \frac{1}{2 \cdot \sqrt{D} \cdot \sqrt{\pi \cdot t}} \cdot \int_{-\infty}^{\infty} f(\sqrt{D}) \cdot \exp\left(\frac{-(x-\sqrt{D})^2}{4 \cdot D \cdot t^2}\right) da \quad (7)$$

with the edge conditions

$$N(x = 0, t > 0) = N_0 \quad (8)$$

$$N(x \geq 0, t = 0) = 0 \quad (9)$$

we get the resulting function from the solving process for the Laplace temperature conduction equations

$$u(x, t) = u_0 \cdot \text{erfc}\left(\frac{x}{2 \cdot a \cdot \sqrt{t}}\right) \quad (10)$$

⁴https://en.wikipedia.org/wiki/Fick%27s_laws_of_diffusion

⁵ISBN 3-8023-1588: Hoppe Bernhard, Mikroelektronik 2, Page 24, Table 2.1

with the error function being an integral of the form

$$\operatorname{erfc}(z) = \left(1 - \frac{2}{\sqrt{\pi}}\right) \cdot \int_0^z e^{-a^2} da \quad (11)$$

Or in case of our dopant concentration equation we can replace a with the square root of the diffusion coefficient in order to get the error function for our dopant density equation:

$$\operatorname{erfc}(z) = \left(1 - \frac{2}{\sqrt{\pi}}\right) \cdot \int_0^z e^{-D} d\sqrt{D} \quad (12)$$

$$N(x, t) = N_0 \cdot \operatorname{erfc}\left(\frac{x}{2 \cdot \sqrt{D \cdot t}}\right) = N_0 \cdot \operatorname{erfc}\left(\frac{x}{x_l(t)}\right) \quad (13)$$

Now we can extract the layer thickness and the depth of the well in dependency of the time and the temperature, respectively:

$$x_l(t) = 2 \cdot \sqrt{D \cdot t} \quad (14)$$

$$x_l(t) = 2 \cdot \sqrt{D_0 \cdot \exp\left(-\frac{E_a}{k \cdot T}\right) \cdot t} \quad (15)$$

And plot the result for multiple different drive in times

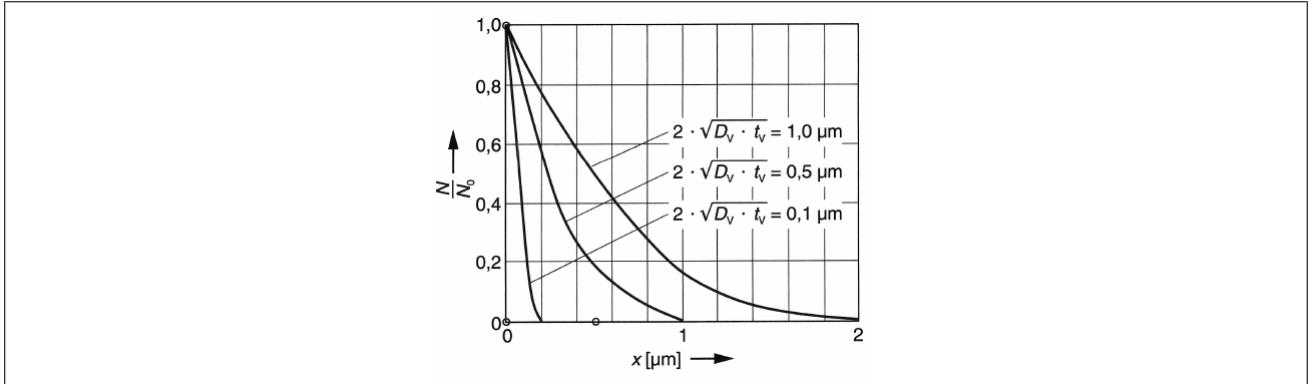


Figure 9: Different predeposition times

We can now describe the dosage based on the time and temperature of the diffusion

$$Q = \frac{2}{\sqrt{\pi}} \cdot N_0 \cdot \sqrt{D \cdot t} \quad (16)$$

Where N_0 (concentration at the surface) equals the maximum solubility of a given element (e.g. boron) within the given medium (e.g. silicon).⁶

⁶If someone really wants to do this in his basement he can google these values and make a pull request

2.4 Ion implant

We can use the following equation to calculate the carrier distribution after implantation:

$$N(x) = N_p \exp\left(-\frac{(x - R_p)^2}{2\Delta R_p^2}\right) = \frac{Q}{\sqrt{2\pi}\Delta R_p} \exp\left(-\frac{(x - R_p)^2}{2\Delta R_p^2}\right) \quad (17)$$

Where the projected range (R_p) and the projected straggle (ΔR_p) need to be looked up in tables ⁷ or looked up using an online tool like the one linked in the footnote⁸

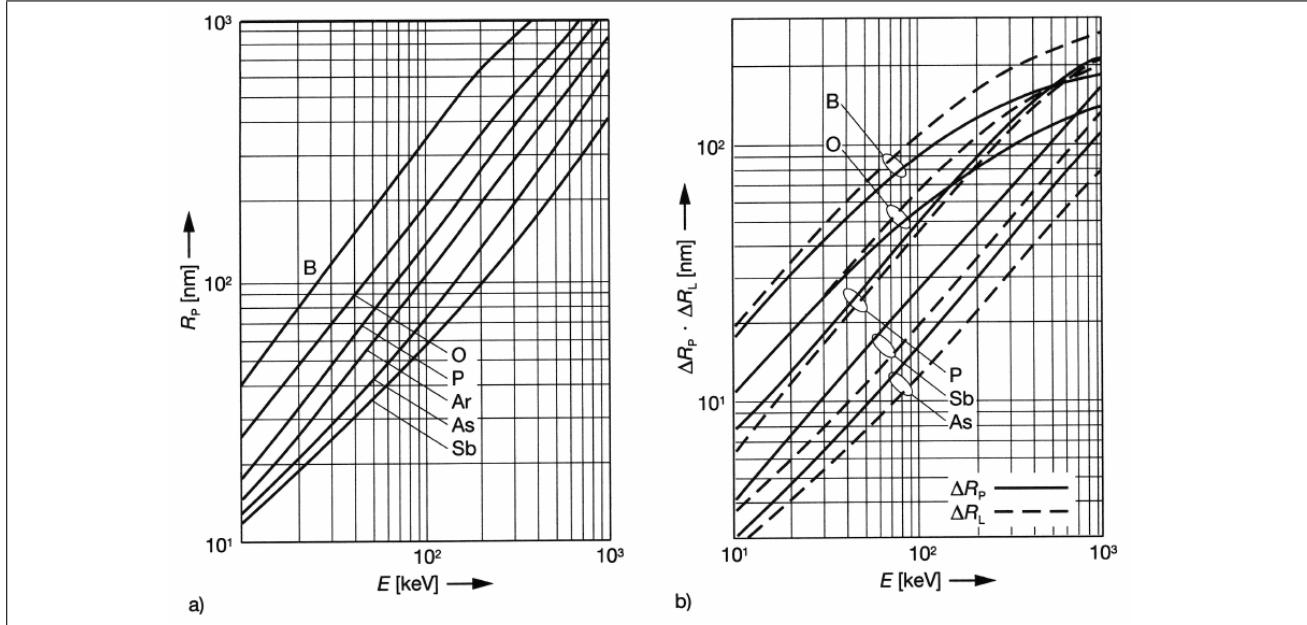


Figure 10: R_p and ΔR_p in silicon

If you do implant before the diffusion just set $x_v = R_p$

2.5 Drive-in (limited source diffusion)

After pre-deposition or ion implant of the initial dosage we need to drive in the ions deeper into the crystal. In order to prevent back-diffusion into the gas we seal off the oxide window with another layer of oxide in order to make sure that all the dopants stay inside the silicon crystal.

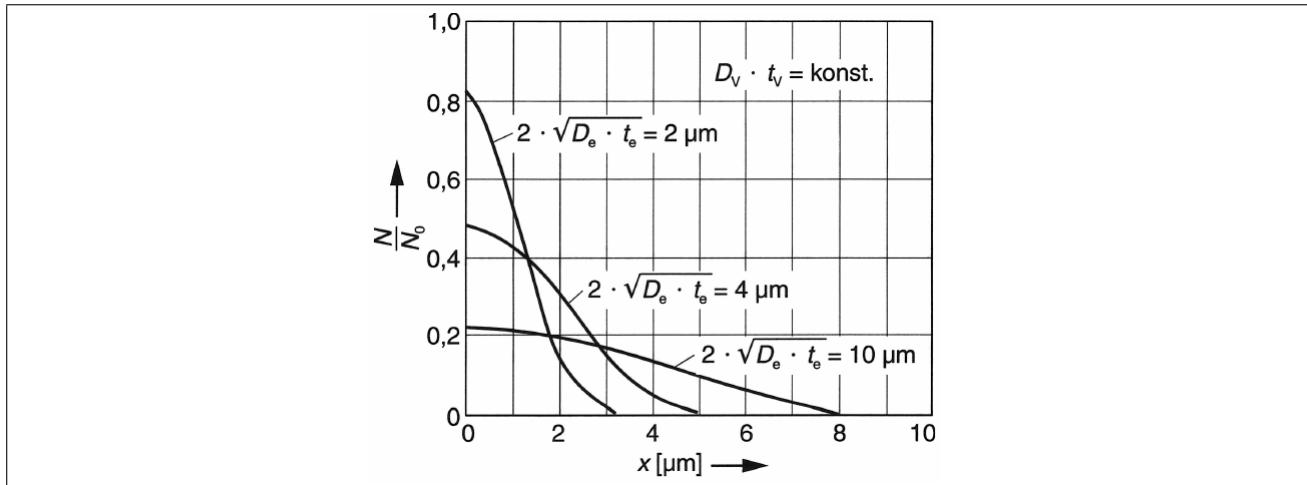


Figure 11: Drive-in well depths and concentrations

⁷ISBN 3-8023-1588: Hoppe Bernhard, Mikroelektronik 2, Page 48, Table 3.2

⁸<http://cleanroom.byu.edu/rangestrallge>

We set the condition that the pre-deposition/implant depth is much lower than the depth of the final diffused volume with the following inequation:

$$x_e = 2 \cdot \sqrt{D_e \cdot t_e} \gg 2 \cdot \sqrt{D_v \cdot t_v} = x_v \quad (18)$$

Where x_v is the the depth of the predeposition/implant step.

By neglecting the distribution thickness of the original implantation dosage and assuming that it's comparably thin compared to the medium thickness we can replace $f(a) \approx \delta(a)$ within [Equation 7](#) which makes

$$N(x, t) = \frac{1}{2 \cdot \sqrt{D} \cdot \sqrt{\pi} \cdot t} \cdot \int_{-\infty}^{\infty} f(\sqrt{D}) \cdot \exp\left(\frac{-(x - \sqrt{D})^2}{4 \cdot D \cdot t}\right) da \quad (19)$$

become

$$N(x, t) = \frac{1}{2 \cdot \sqrt{D} \cdot \sqrt{\pi} \cdot t} \cdot \int_{-\infty}^{\infty} \delta(\sqrt{D}) \cdot \exp\left(\frac{-(x - \sqrt{D})^2}{4 \cdot D \cdot t}\right) da \quad (20)$$

and finally

$$N(x, t) = \frac{Q}{\sqrt{\pi} \cdot D_e \cdot t} \cdot \exp\left(\frac{-x^2}{4 \cdot D_e \cdot t}\right) \quad (21)$$

2.6 Vertical diffusion and junction formation (Well formation)

The goal of most diffusions is to form pn junctions by converting p-type material to n-type material or vice versa. In [Figure 12](#), for example, the wafer is uniformly doped n-type material with a concentration indicated by N_B , and the diffusing impurity is boron. The point at which the diffused impurity profile intersects the background concentration is the metallurgical junction depth (x_j). The net impurity concentration at x_j is zero. Setting $N(x)$ equal to the background concentration N_B at $x = x_j$ yields⁹ for a fixed source

$$x_j = 2 \cdot \sqrt{D \cdot t \cdot \ln\left(\frac{N_0}{N_B}\right)} \quad (22)$$

and for a continuous source

$$x_j = 2 \cdot \sqrt{D \cdot t} \cdot \operatorname{erfc}^{-1}\left(\frac{N_B}{N_0}\right) \quad (23)$$

for the Gaussian and complementary error function distributions, respectively.

In [Figure 12](#), the boron concentration N exceeds N_B to the left of the junction, and this region is p-type. To the right of x_j , N is less than N_B , and this region remains n-type.

To calculate the junction depth, we must know the background concentration N_B of the original wafer. Look at [Figure 7](#) for this purpose.

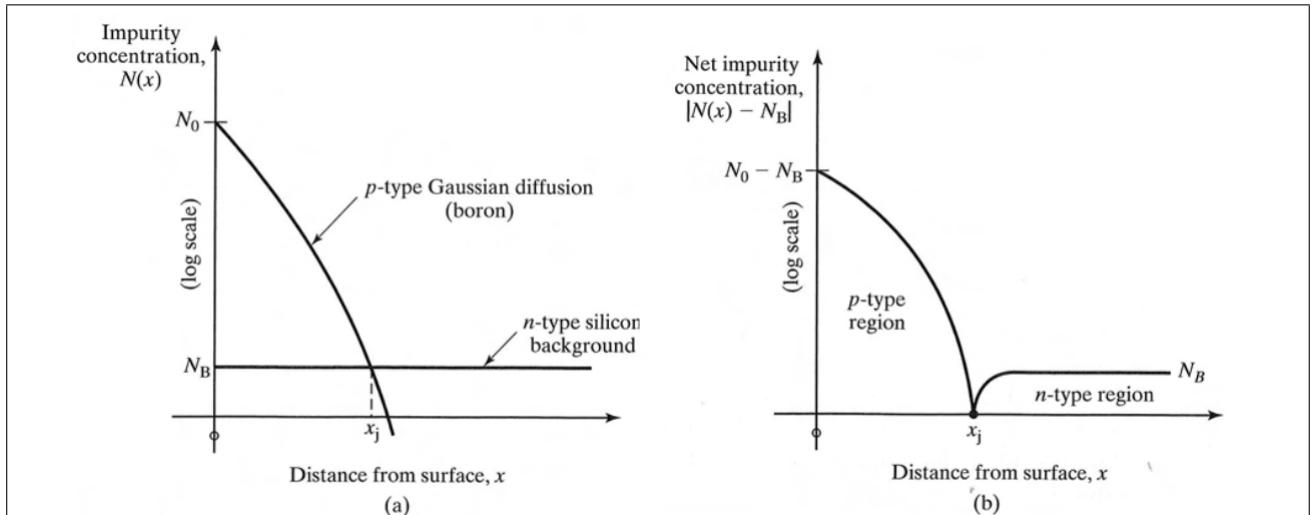


Figure 12: Formation of a pn junction by diffusion: (a) An example of a p-type Gaussian diffusion into a uniformly doped n-type wafer; (b) net impurity concentration in the wafer.

⁹Gerold W. Neudeck and Robert F. Pierret, Modular series on solid state devices, Volume V, Chapter 4

2.7 MOS Capacitance

https://ecee.colorado.edu/~bart/book/book/chapter6/ch6_3.htm

2.8 Threshold voltage with metal gate (V_T)

The formula for calculating the threshold voltage of a MOS device is the following:

$$V_T = V_{t-mos} + V_{FB} \quad (24)$$

where V_{t-mos} is the threshold voltage of an ideal MOS capacitor, V_{FB} is the flat-band voltage and V_{t-mos} is the threshold. The MOS threshold voltage, V_{t-mos} is calculated by considering the MOS capacitor structure that form the gate of the MOS transistor.

The ideal threshold voltage may be expressed as:

$$V_{t-mos} = 2\phi_F + \frac{Q_b}{C_{ox}} \quad (25)$$

$$Q_b = \sqrt{2\epsilon_{Si} \cdot q \cdot N_{implant} \cdot (|2\phi_F| + V_{SB})} \quad (26)$$

where C_{ox} is the oxide capacitance and Q_b which is called the bulk charge term.

The bulk potential is given by:

$$\phi_F = V_{th} \cdot \ln\left(\frac{p}{N_i}\right) = V_{th} \cdot \ln\left(\frac{N_i}{n}\right) \quad (27)$$

V_{th} is the thermal voltage.¹⁰

$$V_{th} = \frac{kT}{q} \approx 0.026 \frac{J}{C} = 0.026V = 26mV \quad (28)$$

With the variables being:

- $k = 1.38064852 \cdot 10^{-23} \frac{J}{K}$ is the Boltzmann constant
- $q = 1.602 \cdot 10^{-19} C$ is the elementary charge
- $T = 300K$ the temperature, which we assume to be the room temperature for simplicity further on in this document as well.

We can directly switch $\frac{J}{C}$ with Volts because these two units are equal!^a Also V_{th} will be treated as a constant for any further calculations within this document.

The same goes for the eV to V conversion, wherever we have work functions to potentials because (e.g. Φ_M for Aluminum): $4.1eV \approx 6.5689241452810^{-19} J$

$$\Phi_M = \frac{E_M}{q} = \frac{4.1eV}{q} = \frac{6.5689241452810^{-19} J}{q} = \frac{6.5689241452810^{-19} J}{1.60217663410^{-19} C} \approx 4.099999966220953 \frac{J}{C} = 4.1V$$

^a<https://en.wikipedia.org/wiki/Volt>

Since we connect bulk and source $V_{SB} = 0$ we can simplify the equation to become

$$Q_b = \sqrt{2 \cdot \epsilon_{Si} \cdot q \cdot N_{implant} \cdot (|2 \cdot \phi_F|)} \quad (29)$$

$$Q_b = 2 \cdot \sqrt{\epsilon_{Si} \cdot q \cdot N_{implant} \cdot |\phi_F|} \quad (30)$$

V_{FB} , is given by:

$$V_{FB} = \phi_{MS} - \frac{Q_{SS}}{C_{ox}} - \frac{1}{C_{ox}} \int_0^{t_{ox}} \frac{x}{x_{ox}} \rho(x) dx \quad (31)$$

Because we're not yet dealing with non-volatile memory devices which contain an oxide surface state charge we can just $\rho(x) = 0$. Q_{SS} is a value which has to be measured.

$$V_{FB} = \phi_{MS} - \frac{Q_{SS}}{C_{ox}} \quad (32)$$

¹⁰https://en.wikipedia.org/wiki/Boltzmann_constant#Role_in_semiconductor_physics:_the_thermal_voltage

with

$$V_{FB} = \phi_{MS} - \frac{Q_{SS}}{C_{ox}} = \phi_M - \phi_S - \frac{Q_{SS}}{C_{ox}} = \phi_M - \left(\chi + \frac{E_g}{2q} + \phi_F \right) - \frac{Q_{SS}}{C_{ox}} \quad (33)$$

And because of the simplifications we did to F_{FB} which essentially led to $F_{FB} = \phi_{MS}$ we get to:

$$V_T = V_{t-mos} + \phi_{MS} \quad (34)$$

$$V_T = 2\phi_F + \frac{Q_b}{C_{ox}} + \phi_{MS} \quad (35)$$

$$V_T = 2\phi_F + \frac{2\sqrt{\epsilon_{Si} \cdot q \cdot |\phi_F| \cdot N_{implant}}}{C_{ox}} + \phi_{MS} \quad (36)$$

$$V_T = 2\phi_F + \frac{2\sqrt{\epsilon_{Si} \cdot q \cdot |\phi_F| \cdot N_{implant}}}{C_{ox}} + \phi_M - \left(\chi + \frac{E_g}{2q} + \phi_F \right) - \frac{Q_{SS}}{C_{ox}} \quad (37)$$

$$V_T = 2\phi_F + \frac{2\sqrt{\epsilon_{Si} \cdot q \cdot |\phi_F| \cdot N_{implant}}}{C_{ox}} + \phi_M - \chi - \frac{E_g}{2q} - \phi_F - \frac{Q_{SS}}{C_{ox}} \quad (38)$$

$$V_T = \phi_F + \frac{2\sqrt{\epsilon_{Si} \cdot q \cdot |\phi_F| \cdot N_{implant}}}{C_{ox}} + \phi_M - \chi - \frac{E_g}{2q} - \frac{Q_{SS}}{C_{ox}} \quad (39)$$

With the variables and constants being the following we now can put the formula together:

- N_i is the carrier concentration in intrinsic (undoped) silicon. N_i is equal to $1.45 \times 10^{10} \text{ cm}^{-3} = 1.45 \times 10^{16} \text{ m}^{-3}$ at 300°K
- Q_{SS} depends on the process and is measured. Usually it's between $10^9 \frac{1}{\text{cm}^2}$ and $10^{10} \frac{1}{\text{cm}^2}$ ergo $Q_{SS} = q \cdot 10^{10} \frac{1}{\text{cm}^2} = 1.6 \cdot 10^{-5} \frac{\text{C}}{\text{m}^2}$
- $E_M = q \cdot \phi_M = 4.1\text{eV}$ is the "work function" of our metal at the gate (Aluminum)
- $E_g = E_g(300)[\text{eV}]$
 $E_g(T) = E_g(0) - \frac{\alpha T^2}{T+\beta} = 1.166 - 4.73 \cdot 10^{-4} \cdot \frac{T^2}{T+636} [\text{eV}]$ is the band gap energy of silicon at a given temperature¹¹ for which the parameters can be taken from [Table 4](#)

	Germanium	Silicon	GaAs
$Eg(0)[\text{eV}]$	0.7437	1.166	1.519
$\alpha[\text{eV}/\text{K}]$	4.77×10^{-4}	4.73×10^{-4}	5.41×10^{-4}
$\beta[\text{K}]$	235	636	204

Table 3: Band cap energy parameters

- $C_{ox} \left[\frac{\text{F}}{\text{m}^2} \right]$ is the capacity of the gate oxide
- $\epsilon_0 = 8.85 \cdot 10^{-14} \frac{\text{F}}{\text{cm}}$, $= 8.85 \cdot 10^{-12} \frac{\text{F}}{\text{m}}$ is the electric permittivity in vacuum
- $\epsilon_{Si} = 11.68 \cdot \epsilon_0$ is the relative permittivity of silicon
- $\epsilon_{ox} = 3.9 \cdot \epsilon_0$ is the relative permittivity of silicon oxide
- $t_{ox}[\text{cm}]$ is the thickness of the oxide layer in cm
- $E_{ef} = q \cdot \chi = 4.05\text{eV}$ is the electron affinity of a silicon crystal surface¹²
- $q = 1.602 \cdot 10^{-19} \text{ C}$ is the elementary charge

The contact potential from the Aluminum contact to the surface of the gate (silicon below the oxide) is fixed for $T = 300^\circ\text{K}$:

$$\phi_M - \chi - \frac{E_g}{2q} = 4.1V - 4.05V - \frac{1.12\text{eV}}{2q} = 4.1V - 4.05V - 0.56V = -0.51V \quad (40)$$

From that we get

$$V_T = \phi_F + \frac{2\sqrt{\epsilon_{Si} \cdot q \cdot |\phi_F| \cdot N_{implant}}}{C_{ox}} - 0.51V \quad (41)$$

¹¹<https://ecee.colorado.edu/~bart/book/eband5.htm>

¹²https://en.wikipedia.org/wiki/Electron_affinity

Now we can calculate the thresholds for P substrate (V_{Tp}) and N substrate (V_{Tn}), respectively the wells we build on unpredoped substrated, which makes the equation for single-doped substrate valid for both wells with

$$\phi_{Fn} = V_{th} \cdot \ln \left(\frac{N_i}{N_{implant}} \right) \quad (42)$$

$$\phi_{Fp} = V_{th} \cdot \ln \left(\frac{N_{implant}}{N_i} \right) \quad (43)$$

Which brings us to the equations for the N-channel and P-channel thresholds:

(N-Channel MOSFETs are built on p-substrate)

$$V_{Tn} = \phi_{Fp} + \frac{2\sqrt{\epsilon_{Si} \cdot q \cdot |\phi_{Fp}| \cdot N_{implant}}}{C_{ox}} - 0.51V \quad (44)$$

(P-Channel MOSFETs are built on n-substrate)

$$V_{Tp} = \phi_{Fn} + \frac{2\sqrt{\epsilon_{Si} \cdot q \cdot |\phi_{Fn}| \cdot N_{implant}}}{C_{ox}} - 0.51V \quad (45)$$

The threshold voltage dependence on the doping density is illustrated with [Figure 14](#) for both n-type and p-type MOSFETs with an aluminum gate metal.

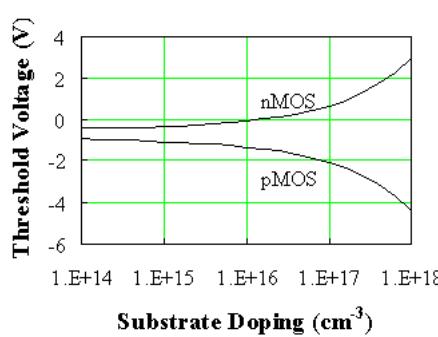


Figure 13: Threshold voltage of n-type (upper curve) and p-type (lower curve) MOSFETs versus substrate doping density.

This equation will be used further on to find the optimum gate oxide thickness for our transistors.

2.9 Threshold voltage with polysilicon gate(V_T)

The formula for calculating the threshold voltage of a MOS device is the following:

$$V_T = V_{t-mos} + V_{FB} \quad (46)$$

where V_{t-mos} is the threshold voltage of an ideal MOS capacitor, V_{FB} is the flat-band voltage and V_{t-mos} is the threshold. The MOS threshold voltage, V_{t-mos} is calculated by considering the MOS capacitor structure that form the gate of the MOS transistor.

The ideal threshold voltage may be expressed as:

$$V_{t-mos} = 2\phi_F + \frac{Q_b}{C_{ox}} \quad (47)$$

$$Q_b = \sqrt{2\epsilon_{Si} \cdot q \cdot N_{implant} \cdot (|2\phi_F| + V_{SB})} \quad (48)$$

where C_{ox} is the oxide capacitance and Q_b which is called the bulk charge term.

The bulk potential is given by:

$$\phi_F = V_{th} \cdot \ln \left(\frac{p}{N_i} \right) = V_{th} \cdot \ln \left(\frac{N_i}{n} \right) \quad (49)$$

V_{th} is the thermal voltage.¹³

$$V_{th} = \frac{kT}{q} \approx 0.026 \frac{J}{C} = 0.026V = 26mV \quad (50)$$

With the variables being:

¹³https://en.wikipedia.org/wiki/Boltzmann_constant#Role_in_semiconductor_physics:_the_thermal_voltage

- $k = 1.38064852 \cdot 10^{-23} \frac{J}{K}$ is the Boltzmann constant
- $q = 1.602 \cdot 10^{-19} C$ is the elementary charge
- $T = 300 K$ the temperature, which we assume to be the room temperature for simplicity further on in this document as well.

We can directly switch $\frac{J}{C}$ with Volts because these two units are equal!^a Also V_{th} will be treated as a constant for any further calculations within this document.

The same goes for the eV to V conversion, wherever we have work functions to potentials because (e.g. Φ_M for Aluminum): $4.1eV \approx 6.5689241452810^{-19} J$

$$\Phi_M = \frac{E_M}{q} = \frac{4.1eV}{q} = \frac{6.5689241452810^{-19} J}{q} = \frac{6.5689241452810^{-19} J}{1.60217663410^{-19} C} \approx 4.099999966220953 \frac{J}{C} = 4.1V$$

^a<https://en.wikipedia.org/wiki/Volt>

Since we connect bulk and source $V_{SB} = 0$ we can simplify the equation to become

$$Q_b = \sqrt{2 \cdot \epsilon_{Si} \cdot q \cdot N_{implant} \cdot (|2 \cdot \phi_F|)} \quad (51)$$

$$Q_b = 2 \cdot \sqrt{\epsilon_{Si} \cdot q \cdot N_{implant} \cdot |\phi_F|} \quad (52)$$

V_{FB} , is given by:

$$V_{FB} = \phi_{MS} - \frac{Q_{SS}}{C_{ox}} - \frac{1}{C_{ox}} \int_0^{t_{ox}} \frac{x}{x_{ox}} \rho(x) dx \quad (53)$$

Because we're not yet dealing with non-volatile memory devices which contain an oxide surface state charge we can just $\rho(x) = 0$. Q_{SS} is a value which has to be measured.

$$V_{FB} = \phi_{MS} - \frac{Q_{SS}}{C_{ox}} \quad (54)$$

with

$$V_{FB} = \phi_{MS} - \frac{Q_{SS}}{C_{ox}} = \phi_M - \phi_S - \frac{Q_{SS}}{C_{ox}} = \phi_M - \left(\chi + \frac{E_g}{2q} + \phi_F \right) - \frac{Q_{SS}}{C_{ox}} \quad (55)$$

And because of the simplifications we did to F_{FB} which essentially led to $F_{FB} = \phi_{MS}$ we get to:

$$V_T = V_{t-mos} + \phi_{MS} \quad (56)$$

$$V_T = 2\phi_F + \frac{Q_b}{C_{ox}} + \phi_{MS} \quad (57)$$

$$V_T = 2\phi_F + \frac{2\sqrt{\epsilon_{Si} \cdot q \cdot |\phi_F| \cdot N_{implant}}}{C_{ox}} + \phi_{MS} \quad (58)$$

$$V_T = 2\phi_F + \frac{2\sqrt{\epsilon_{Si} \cdot q \cdot |\phi_F| \cdot N_{implant}}}{C_{ox}} + \phi_M - \left(\chi + \frac{E_g}{2q} + \phi_F \right) - \frac{Q_{SS}}{C_{ox}} \quad (59)$$

$$V_T = 2\phi_F + \frac{2\sqrt{\epsilon_{Si} \cdot q \cdot |\phi_F| \cdot N_{implant}}}{C_{ox}} + \phi_M - \chi - \frac{E_g}{2q} - \phi_F - \frac{Q_{SS}}{C_{ox}} \quad (60)$$

$$V_T = \phi_F + \frac{2\sqrt{\epsilon_{Si} \cdot q \cdot |\phi_F| \cdot N_{implant}}}{C_{ox}} + \phi_M - \chi - \frac{E_g}{2q} - \frac{Q_{SS}}{C_{ox}} \quad (61)$$

With the variables and constants being the following we now can put the formula together:

- N_i is the carrier concentration in intrinsic (undoped) silicon. N_i is equal to $1.45 \times 10^{10} cm^{-3} = 1.45 \times 10^{16} m^{-3}$ at $300^\circ K$
- Q_{SS} depends on the process and is measured. Usually it's between $10^9 \frac{1}{cm^2}$ and $10^{10} \frac{1}{cm^2}$ ergo $Q_{SS} = q \cdot 10^{10} \frac{1}{cm^2} = 1.6 \cdot 10^{-5} \frac{C}{m^2}$
- $E_M = q \cdot \phi_M = 4.1eV$ is the "work function" of our metal at the gate (Aluminum)

- $E_g = E_g(300)[eV]$
 $E_g(T) = E_g(0) - \frac{\alpha T^2}{T+\beta} = 1.166 - 4.73 \cdot 10^{-4} \cdot \frac{T^2}{T+636}[eV]$ is the band gap energy of silicon at a given temperature¹⁴ for which the parameters can be taken from [Table 4](#)

	Germanium	Silicon	GaAs
$Eg(0)[eV]$	0.7437	1.166	1.519
$\alpha[eV/K]$	4.77 x 10-4	4.73 x 10-4	5.41 x 10-4
$\beta[K]$	235	636	204

Table 4: Band cap energy parameters

- $C_{ox} \left[\frac{F}{m^2} \right]$ is the capacity of the gate oxide
- $\epsilon_0 = 8.85 \cdot 10^{-14} \frac{F}{cm}$. $= 8.85 \cdot 10^{-12} \frac{F}{m}$ is the electric permittivity in vacuum
- $\epsilon_{Si} = 11.68 \cdot \epsilon_0$ is the relative permittivity of silicon
- $\epsilon_{ox} = 3.9 \cdot \epsilon_0$ is the relative permittivity of silicon oxide
- $t_{ox}[cm]$ is the thickness of the oxide layer in cm
- $E_{ef} = q \cdot \chi = 4.05eV$ is the electron affinity of a silicon crystal surface¹⁵
- $q = 1.602 \cdot 10^{-19} C$ is the elementary charge

The contact potential from the Aluminum contact to the surface of the gate (silicon below the oxide) is fixed for $T = 300^\circ K$:

$$\phi_M - \chi - \frac{E_g}{2q} = 4.1V - 4.05V - \frac{1.12eV}{2q} = 4.1V - 4.05V - 0.56V = -0.51V \quad (62)$$

From that we get

$$V_T = \phi_F + \frac{2\sqrt{\epsilon_{Si} \cdot q \cdot |\phi_F| \cdot N_{implant}}}{C_{ox}} - 0.51V \quad (63)$$

Now we can calculate the thresholds for P substrate (V_{Tp}) and N substrate (V_{Tn}), respectively the wells we build on unpredoped substrated, which makes the equation for single-doped substrate valid for both wells with

$$\phi_{Fn} = V_{th} \cdot \ln \left(\frac{N_i}{N_{implant}} \right) \quad (64)$$

$$\phi_{Fp} = V_{th} \cdot \ln \left(\frac{N_{implant}}{N_i} \right) \quad (65)$$

Which brings us to the equations for the N-channel and P-channel thresholds:
(N-Channel MOSFETs are built on p-substrate)

$$V_{Tn} = \phi_{Fn} + \frac{2\sqrt{\epsilon_{Si} \cdot q \cdot |\phi_{Fn}| \cdot N_{implant}}}{C_{ox}} - 0.51V \quad (66)$$

(P-Channel MOSFETs are built on n-substrate)

$$V_{Tp} = \phi_{Fp} + \frac{2\sqrt{\epsilon_{Si} \cdot q \cdot |\phi_{Fp}| \cdot N_{implant}}}{C_{ox}} - 0.51V \quad (67)$$

The threshold voltage dependence on the doping density is illustrated with [Figure 14](#) for both n-type and p-type MOSFETs with an aluminum gate metal.

¹⁴<https://ecee.colorado.edu/~bart/book/eband5.htm>

¹⁵https://en.wikipedia.org/wiki/Electron_affinity

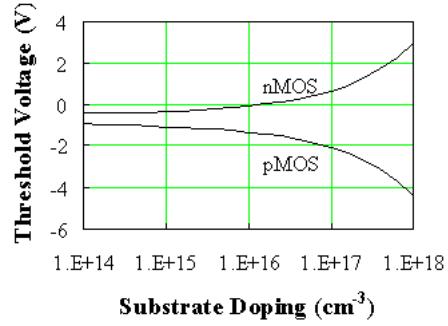


Figure 14: Threshold voltage of n-type (upper curve) and p-type (lower curve) MOSFETs versus substrate doping density.

This equation will be used further on to find the optimum gate oxide thickness for our transistors.

2.10 Threshold voltage (V_T) adjustment

At some point in the future this will be of very high relevance, because the lower the size of the transistors becomes, the higher the offset to V_{Tp} and V_{Tn} needs to be in order to stay on TTL 5V logic level, or at least compensate for the lowered voltages in order to reach the 3.3V CMOS logic levels.

Adjustment of the threshold voltage can be achieved by:

- A relatively small dose N_I (units: ions/cm^2) of dopant atoms is implanted into the near-surface region of the semiconductor.
- When the MOS device is biased in depletion or inversion, the implanted dopants add to (or subtract from) the depletion charge near the oxide-semiconductor interface

The formula to calculate the voltage offset is:

$$\Delta V_T = -\frac{qN_I}{C_{ox}} \begin{cases} N_I > 0 \text{ for donor atoms (Phosphorus/N)} \\ N_I < 0 \text{ for acceptor atoms (Boron/P)} \end{cases} \quad (68)$$

3 Chemistry

3.1 Etching silicon dioxide

A very "selective" chemical for SiO_2 - i.e. does not etch silicon at all - is hydrofluoric acid (HF). If used directly such etchant has a too fast and aggressive action on the oxide, making very difficult the undercut and the linewidth control. For such reason, HF is universally used as a "buffered" solution, which can keep the etch rate low and constant, by moderating the PH level of the bath. This allows the etching time to be reliably correlated to the etching depth.

The industry standard buffered hydrofluoric acid solution (BHF) has the following formulation:

- 6 volumes of ammonium fluoride (NH_4F , 40% solution)
- 1 volume of HF.

This can be prepared, for example, by mixing 113 g of NH_4F in 170 ml of H_2O , and adding 28 ml of HF. The etch rate at room temperature can range from 1000 to 2500 Å/min (100-250nm/min). This depends on the actual density of the oxide which, as an amorphous layer, can have a more compact structure (if thermally grown in oxygen) or less compact (if grown by CVD). The following etching reaction holds:



where H_2SiF_6 is water soluble.

A common buffered oxide etch solution comprises a 6:1 volume ratio of 40% NH_4F in water to 49% HF in water. This solution will etch thermally grown oxide at approximately 2 nanometres per second at 25 degrees Celsius. ¹⁶

Another popular etching formulation is the P-etch:

60 volumes of H_2O + 3 vol. of HF + 2 vol. of HNO_3 , that is: 300 ml of H_2O + 15 ml of HF + 10 ml of HNO_3 .

The P-etch action is strongly dependent on oxide density, as it results from the growth technique. An example is reported in the literature¹⁷, indicating 120 Å/min for thermal oxide and 250-700 Å/min for sputtered oxide. A slow etching bath is preferred for opening mask windows for a silicon substrate. However, the etching process could be used just for removing the oxide film from the whole surface. In this case the etching speed is not critical, and a fast solution can be used, such as HF diluted 1:10 in water. The etching time can be easily evaluated by visually inspecting the surface. Once the oxide film is removed, the metal-grey color of the silicon surface appears.

Sometimes a very light etch is required, for removing just a few atomic layers. This is the case of surface cleaning and decontamination. HF diluted 1 : 50 in water can be used. The etching speed will be around 70 Å / min. For example, a typical 50 Å "native" oxide on silicon can be removed with a 45 - 50 sec light etch.

¹⁶Wolf, S.; R.N. Tauber (1986). Silicon Processing for the VLSI Era: Volume 1 - Process Technology. pp. 532-533. ISBN 978-0-9616721-3-3

¹⁷A. Pliskin, J.Vac.Sci Technol., vol. 14, p.1064, 1977

3.2 Etching silicon nitride

Thin films made of amorphous silicon nitride (Si_3N_4) are usually deposited by chemical vapour deposition from silane (SiH_4) and ammonia (NH_3). Since they act as a barrier for water and sodium, they have a major role as passivation layers in microchip fabrication. Patterned nitride layers are also used as a mask for spatially selective silicon oxide growth, and as an etch mask when SiO_2 masks cannot be used.

One example of the latter situation is given by the anisotropic etching of silicon in KOH. The etching rate of SiO_2 in KOH is nearly 1000 times slower than the etching rate of silicon, and in most cases a SiO_2 mask can be used successfully. However, a very deep selective etch may require a long etching time, and the 1000:1 etching rate ratio may result still too small to prevent the SiO_2 mask from being etched off before the process is completed. In this circumstance Si_3N_4 , thanks to its reduced etched rate, can successfully replace the oxide mask layer.

The wet etching of nitride films is often performed in concentrated hot orthophosphoric acid (H_3PO_4). The bath temperature can range from 150°C to 180°C (boiling point) with a corresponding etch rate between 10 and 100 Å/min. It is good practice to bring the vapours into contact with a cold surface and to drive the condensed liquid back into the etching bath. This technique is referred to as "reflux".

The etching rates of silicon nitride, silicon oxide, and silicon in H_3PO_4 are respectively in the 50 : 5 : 1 ratio.

3.3 Growing silicon nitride

In order to grow a high quality layer of silicon nitride on top of a silicon wafer which is adapted to be patterned and to serve as a mask for diffusion or implantation of selected impurities, the wafer is best put into a chamber evacuated to a pressure less than about 1 Torr and heated between 650 and 900 °C. A gaseous mixture comprising primarily of ammonia and a silicon compound, having a ratio of relative concentrations in the range on 4:1 and 20:1¹⁸, is flooded into that chamber with a silicon compound flow rate of greater than approximately 12 cubic centimeters per minute. The growth rate will be around 50 Angstroms per minute. That setup is called Low-Pressure Chemical Vapor Deposition (LPCVD), which is commonly available in basically any semiconductor manufacturing plant or laboratory.

¹⁸<http://www.freepatentsonline.com/4395438.html>

4 Process design

We need to optimize our process to be TTL compatible (5V logic levels) and at the same time being as fast and power efficient as possible. In order to have a good propagation delay with a technology node of around $1\mu m$ we will have to have gates with up to four stacked MOS transistors.

Acceptable input signal voltages range from 0 volts to 0.8 volts for a low logic state, and 2 volts to 5 volts for a high logic state. Acceptable output signal voltages shall range from 0 volts to 0.5 volts for a low logic state, and 2.7 volts to 5 volts for a high logic state¹⁹

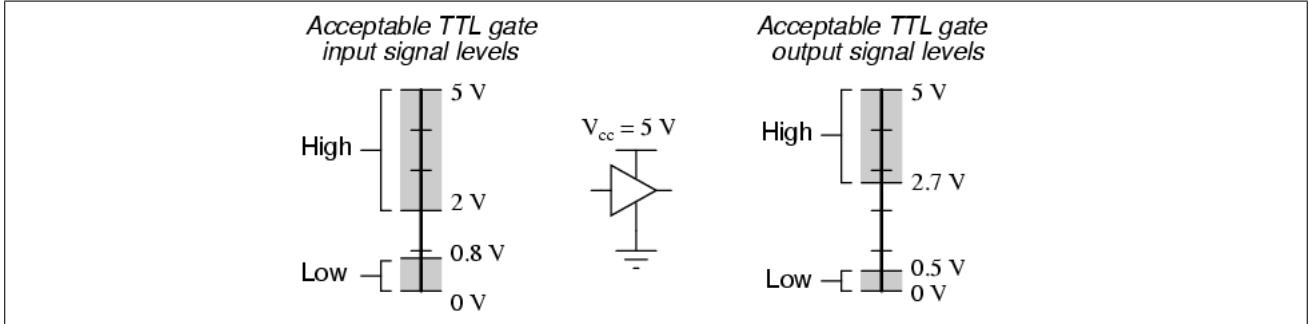


Figure 15: TTL logic levels

As shown in [Figure 15](#) we have some margin to make our PMOS and NMOS transistors work with each other in order to form a CMOS circuit which is actually working without getting warm.

Or more clearly defined

$$V_{off} \leq 0.8V \quad (70)$$

and

$$V_{on} \geq 2V \quad (71)$$

which are limits, elementary to our design.

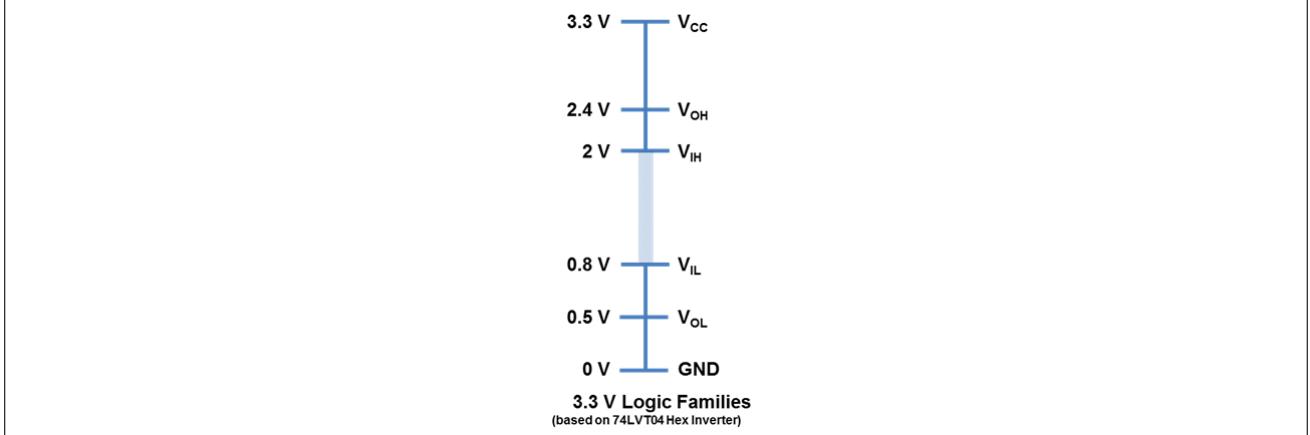


Figure 16: CMOS 3.3V logic levels

This means that we also will be compatible to CMOS logic level output pins since their ON/OFF levels are within our tolerance range²⁰ as it is shown in [Figure 16](#).

We target threshold voltages of $V_{Tn} \approx 0.8V$ and $V_{Tp} \approx -0.8V$ which should be enough. We can internally always shift the voltage supply levels to compensate for threshold variations.²¹

¹⁹<https://www.allaboutcircuits.com/textbook/digital/chpt-3/logic-signal-voltage-levels>

²⁰<https://learn.sparkfun.com/tutorials/logic-levels/33-v-cmos-logic-levels>

²¹Hagen! Please explain this part here

4.1 Substrate

The Hong University of science and technology (short HKUST) provides us with two types of wafers.

- Prime Grade Silicon Wafer, [100] N-type
 - Front-side polished, backside etched
 - Dopant: Phosphorus
 - Thickness: $525\mu m \pm 25\mu m$
 - Resistivity: 4 to 7 ohm-cm
 - Growth Method: CZ
 - Diameter: 100mm +/- 0.5 mm
 - Primary & secondary flat locations: (In compliance with the SEMI)
 - * Carbon concentration $< 2.5 \times 10^{16} \text{ atm/cc}$
 - * Oxygen concentration $< 9.0 \cdot 10^{17} \frac{\text{atm}}{\text{cc}}$
 - * TTV $< 10\mu m$
 - * TIR $< 6\mu m$
 - * Bow/Warp $< 40\mu m$
- Prime Grade Silicon Wafer, [100] P-type
 - Front-side polished, backside etched
 - Dopant: Boron
 - Thickness: $525\mu m \pm 25\mu m$
 - Resistivity: 15 to 25 ohm-cm
 - Growth Method: CZ
 - Diameter: 100mm +/- 0.5 mm
 - Primary & secondary flat locations: (In compliance with the SEMI)
 - * Carbon concentration $< 2.5 \times 10^{16} \text{ atm/cc}$
 - * Oxygen concentration $< 9.0 \cdot 10^{17} \frac{\text{atm}}{\text{cc}}$
 - * TTV $< 10\mu m$
 - * TIR $< 6\mu m$
 - * Bow/Warp $< 40\mu m$

For this process the p-doped mono crystalline silicon substrate is being used, but forks and modifications will be very well possible based on a Graphene substrate or alike, still under the LSPL. The starting material is a p-doped $<100>$ oriented mono crystalline silicon wafer

Reasons for using p-doped substrate:

- We can't use two different substrates for our design because in the design both PMOS and NMOS is present. We have to choose which is more beneficial from fabrication point of view. In general or say it's true that NMOS devices are always more in the Semiconductor Industry in comparison to PMOS devices. For your reference-SRAM requires 6 transistors (4 NMOS, 2 PMOS).
- Another reason for more number of NMOS is because of difference of mobility of electron and holes. Electron mobility is almost twice of holes mobility and because of this ON-RESISTANCE of n-channel device is half of p-channel device with the same geometry and under the same operating conditions. That means to achieve same impedance size of n-channel transistors is almost half of p-channel devices. Same thing I can say in the different way that for same size of wafer, we can have more number of NMOS (means can perform more logical operation) in comparison to PMOS.
- Since we only have the choice between P and N doped substrate, we use P doped substrate, because of the carrier mobility

Using the method from [Figure 7](#) we get a doping concentration between $8.76 \cdot 10^{14} \frac{1}{\text{cm}^3}$ and $5.23 \cdot 10^{14} \frac{1}{\text{cm}^3}$. The average of this range is $N_B = \frac{8.76+5.23}{2} \cdot 10^{14} \frac{1}{\text{cm}^3} \approx 7 \cdot 10^{14} \frac{1}{\text{cm}^3}$

4.2 Isolation

For the isolation ([subsection 6.1](#)) in this design the STI approach is being chosen. Shallow trench isolation (STI), also known as box isolation technique, is an integrated circuit feature which prevents electric current leakage between adjacent semiconductor device components.²² STI is generally used on CMOS process technology nodes of 250 nanometers and smaller.

Reasons for using box isolation:

- We want to be forward compatible to future LibreSilicon nodes with a size of 100nm or smaller
- It simplifies the construction of the gate and allows us to use Aluminum instead of Polysilicon for the gate contact

Issues we have to keep in mind is that the depth is not uniform and can variate strongly within a $2\mu m$ range! This means we have to make the well at least "deep enough" at the shallowest place, so that it provides adequate isolation between the transistors everywhere on the die.

One way to reduce the variation in depth is to have a uniform width of the isolation.

Also the non-uniform thickness of the oxide is a problem.

CMP (Chemical Mechanical Planarization) for evening the oxide out is being chosen, because the hard mask can be removed in the very same process step.

4.3 Interconnect

The interconnects and the gate electrode are being made using Aluminum which is a very commonly used material to do interconnects in low-frequency and low-resolution applications

Reasons for using Aluminum:

- It's a well explored material for interconnect with a lot of literature on how to process it
- Aluminum is easy to etch compared to copper
- Aluminum isn't contaminating everything like copper does and doesn't require special separated setup for handling
- There are many technology nodes even down to 180nm which still are using Aluminum and it's not having a big impact on the clock frequencies, so it should be good enough for us as well.

As soon as we've got CMOS all figured out, we will tackle copper interconnect in release 2.0

²²<https://www.google.com/patents/US7985656>

4.4 MOS gate material

We decided to use the gate-first approach because realizing the gate self-alignment is much easier this way. We further on decided to use the best-practice material polysilicon for the gate electrode, because it is easy to deposit and etch and virtually every manufacturer out there has at least one machine in their lab to deposit it. Because of its high resistance however, we had to throw in another layer of silicide in order to reduce the gate resistivity from these $100\Omega - m$ or so to a few Ohms per meter. A nice side effect is the better etch-stop properties of this low-resistance film adding to the gate, source and drain contacts.

A down side is that we will have to get "a hang on" the reaction times of silicides because a lot of the details of the reactivity between silicon and titanium to form titanium-silicide seem to be under NDA and secrets of the diverse factories running their own CMOS processes.

4.5 MOS gate thickness

As the continuous down-scaling of the device size has lead to very thin gate oxides, the leakage current that can flow from the channel to the gate comes into the order of the subthreshold leakage current and the gate cannot be considered as an ideally insulated electrode anymore. This affects the circuit functionality and increases the standby power consumption due to the static gate current. For dynamic logic concepts the gate leakage drastically reduces the maximum clock cycle time²³. Two tunneling mechanisms are responsible for the gate leakage, Fowler-Nordheim tunneling and direct tunneling²⁴. The gate leakage increases exponentially as the oxide thickness is reduced. This limits the down-scaling of the oxide thickness to about 1.5-2 nm when looking at the total standby power consumption of a chip²⁵. To further decrease the effective oxide thickness alternative high dielectric constant materials can be used²⁶. On the other hand, a thin gate oxide reduces the short-channel effect and improves the driving capabilities of a MOS transistor. However, a tradeoff between this benefit and the gate leakage is necessary.

With $1\mu m$ we don't have to worry about this leakage yet because our gate oxide thickness is too high for these effects to actually become a problem, but we want to do our home work already in preparation of scale-down and also for curiosity.

We for now just use 40 nm. That's still doable with a precision high enough when using dry oxidation and a temperature of $1000^\circ C$ Celsius.

4.5.1 Subthreshold leakage

The sub-threshold leakage current can be calculated with²⁷

$$I_{sub} = I_0 \cdot \left(1 - \exp\left(-\frac{V_{ds}}{V_{th}}\right)\right) \cdot \exp\left(\frac{V_{gs} - V_T}{n \cdot V_{th}}\right) \quad (72)$$

where

$$I_0 = \frac{W}{L} \mu_0 V_{th}^2 \sqrt{\frac{N_A \cdot q \cdot \epsilon_{Si}}{2 \cdot \phi_{sub}}} \quad (73)$$

$V_{th} = 26mV$ is the thermal voltage, V_T is the threshold voltage, V_{ds} and V_{gs} are the drain-to-source and gate-to-source voltages respectively. W and L are the effective transistor width and length, respectively. C_{ox} is the gate oxide capacitance, μ_0 is the carrier mobility and $n = 1 + \frac{C_{dep}}{C_{ox}}$ is the subthreshold swing coefficient.

First of all, lets say $W = L$ which leads to a square:

$$I_0 = \mu_0 V_{th}^2 \sqrt{\frac{N_A \cdot q \cdot \epsilon_{Si}}{2 \cdot \phi_{sub}}} \quad (74)$$

With

- $\epsilon_0 = 8.85 \cdot 10^{-14} \frac{F}{cm}$. is the electric permittivity in vacuum
- $\epsilon_{ox} = 3.9 \cdot \epsilon_0$ is the relative permittivity of silicon dioxide

²³N. Wang, Digital MOS Integrated Circuits, Prentice-Hall, Englewood Cliffs, NJ, 1989

²⁴A. Schenk and G. Heiser, "Modeling and Simulation of Tunneling through Ultra-Thin Gate Dielectrics" J.Appl.Phys., vol. 81, no. 12, pp. 7900, 1997

²⁵Y. Taur, "The Incredible Shrinking Transistor," IEEE Spectrum, pp. 25-29, July 1999.

²⁶S. Thompson, P. Packan, and M. Bohr, "MOS Scaling: Transistor Challenges for the 21st Century," Intel Technology Journal, vol. Q3, 1998

²⁷http://ecee.colorado.edu/~bart/book/book/chapter3/ch3_4.htm#3_4_2

- $\epsilon_{Si} = 11.68 \cdot \epsilon_0$ is the relative permittivity of silicon

The carrier mobility μ_0 can be calculated with²⁸

$$\mu(N) = \mu_{min} + \frac{\mu_{max} - \mu_{min}}{1 + \left(\frac{N}{N_r}\right)^\alpha} \quad (75)$$

using the fitting parameters from [Table 5](#)

	Arsenic	Phosphorus	Boron
$\mu_{min} [\frac{cm^2}{Vs}]$	52.2	68.5	44.9
$\mu_{max} [\frac{cm^2}{Vs}]$	1417	1414	470.5
$N_r [\frac{1}{cm^3}]$	$9.68 \cdot 10^{16}$	$9.20 \cdot 10^{16}$	$2.23 \cdot 10^{17}$
α	0.68	0.711	0.719

Table 5: Parameters for calculation of the mobility as a function of the doping density

4.5.2 Gate tunneling current

The tunneling of electrons (or holes) from the bulk and source/drain overlap region through the gate oxide potential barrier into the gate (or vice-versa) is referred as gate oxide tunneling current. This phenomenon is related with the MOS capacitance concept. There are three major gate leakage mechanisms in a MOS structure. The first one is the electron conduction-band tunneling (ECB), where electrons tunneling from conduction band of the substrate to the conduction band of the gate (or vice versa). The second one is the electron valence-band tunneling (EVB). In this case, electrons tunneling from the valence band of the substrate to the conduct band of the gate. The last one is known as hole valence-band (HVB) tunneling, where holes tunneling from the valence band of the substrate to the valence band of the gate (or vice- versa)

Each mechanism is dominant or important in different regions of operation for NMOS and PMOS transistors. For each mechanism, gate leakage current can be modeled by

$$I = W \cdot L \cdot A \cdot \left(\frac{V_{ox}}{t_{ox}} \right)^2 \exp \left(\frac{-B \left(1 - \left(1 - \frac{V_{ox}}{\phi_{ox}} \right)^{\frac{3}{2}} \right)}{\frac{T_{ox}}{t_{ox}}} \right) \quad (76)$$

²⁸https://ecee.colorado.edu/~bart/book/book/chapter2/ch2_7.htm#2_7_2

4.6 NMOS threshold

First we take a look at the pull down network, which is being formed by the NMOS transistors. We have to make sure that the pull down network will effectively tie our output signal to ground, latest when reaching the voltage defined as high signal (inverter!)

As shown in [Figure 15](#) our acceptable voltages for our CMOS "ON" state range from 0V to 2V

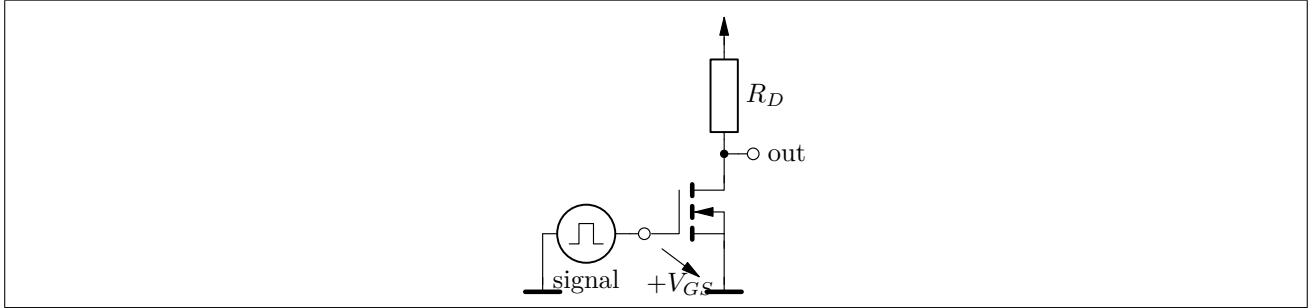


Figure 17: enhancement-mode NMOS transistor use-case

$$V_{off} \leq 0.8V \quad (77)$$

With the values derived from [subsection 4.5](#) which gives us the thickness of the gate ($\approx 40nm$) we target a threshold voltage of $0.8V$, so that we switch the pull down circuit as soon as the low-signal is on.

We target a concentration of $N_p = 10^{16} \frac{1}{cm^3} = 10^{22} \frac{1}{m^3}$.

The depletion zone thickness at its peak will be $W_{dmax} \approx 2.73 \cdot 10^{-7} m = 273nm$

With an implantation (or constant source diffusion step), we can now set a range/energy and dosage in order to cover the depletion zone area.

For getting the energy and dose we look at [Figure 10](#) or use the web tool linked in the implant chapter.

The depth of the p-well $\approx 2\mu m$ comes mainly from the need to fulfill the condition from [subsection 2.5](#)

$$x_e = 2 \cdot \sqrt{D_e \cdot t_e} \gg 2 \cdot \sqrt{D_v \cdot t_v} = x_v \quad (78)$$

We already got the background ($N_B \approx 7 \cdot 10^{14} \frac{1}{cm^3} = 7 \cdot 10^{20} \frac{1}{m^3}$) concentration from the specs of the basis substrate.

$$N_p - N_B = 10^{22} \frac{1}{m^3} - 7 \cdot 10^{20} \frac{1}{m^3} = 9.3 \cdot 10^{21} \frac{1}{m^3} \quad (79)$$

We use a drive in temperature of $1150^\circ C$ which is $T = 1423.15^\circ K$ in Kelvin which gives us the diffusion coefficient $D = 9.1 \cdot 10^{-17} \frac{m^2}{s}$

Now using

$$N(x, t) = \frac{Q}{\sqrt{\pi \cdot D \cdot t}} \cdot \exp\left(\frac{-x^2}{4 \cdot D \cdot t}\right) \quad (80)$$

We set the conditions and get the required diffusion time as well as the initial dosage in one shot:

$$N(0, t) = \frac{Q}{\sqrt{\pi \cdot D \cdot t}} = N_p - N_B = 7 \cdot 10^{20} \frac{1}{m^3} \quad (81)$$

$$x = 2 \cdot \sqrt{D \cdot t \cdot \ln\left(\frac{N_T}{N_B}\right)} = 2 \mu m = 2 \cdot 10^{-6} m \quad (82)$$

$$\Rightarrow t \approx 4259 s \approx 70 min \quad (83)$$

$$\Rightarrow Q = 7 \cdot 10^{20} \frac{1}{m^3} \cdot \sqrt{\pi \cdot D \cdot t} \approx 1.02 \cdot 10^{16} \frac{1}{m^2} \quad (84)$$

If we plot the functions from our calculation we can yield the below graphics²⁹

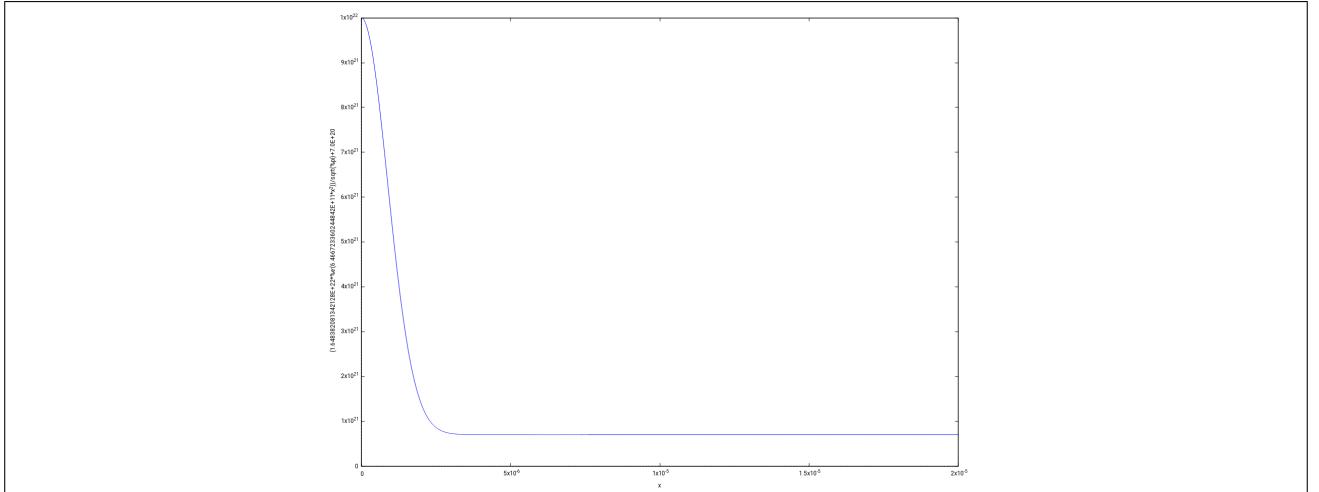


Figure 18: Dopant concentration after around 70 minutes

In [Figure 18](#) we can see that after roughly an hour we already have the desired even gradient and deep penetration of dopants, which will give us a low R_{DS} .

²⁹see simulation/diffusion_pwell.wxml

4.7 PMOS threshold

Now we take a look at the worst case of 4 stacked PMOS transistors, which is the highest stacking amount which will be possible in technologies relying on this process.

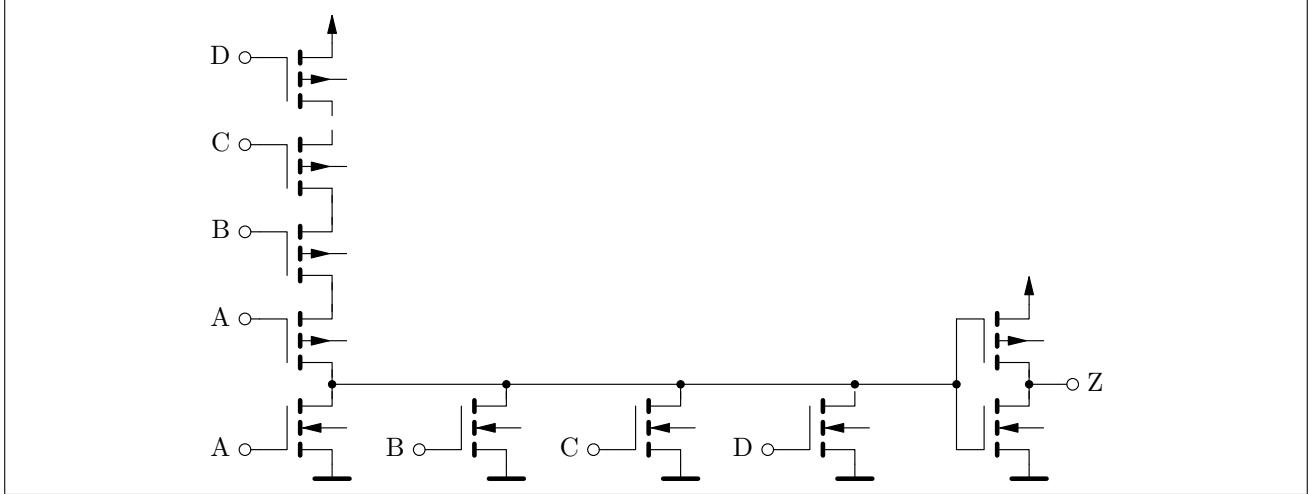


Figure 19: OR4 gate

$\approx 4\mu m$ come mainly from the need to fulfill the condition from subsection 2.5

$$x_e = 2 \cdot \sqrt{D_e \cdot t_e} \gg 2 \cdot \sqrt{D_v \cdot t_v} = x_v \quad (85)$$

We already got the background ($N_B \approx 7 \cdot 10^{14} \frac{1}{cm^3} = 7 \cdot 10^{20} \frac{1}{m^3}$) concentration from the specs of the basis substrate.

We use a drive in temperature of $1150^\circ C$ which is $T = 1423.15^\circ K$ in Kelvin which gives us the diffusion coefficient $D = 9.1 \cdot 10^{-17} \frac{m^2}{s}$

Now using

$$N(x, t) = \frac{Q}{\sqrt{\pi \cdot D \cdot t}} \cdot \exp\left(\frac{-x^2}{4 \cdot D \cdot t}\right) \quad (86)$$

We set the conditions and get the required diffusion time as well as the initial dosage in one shot:

$$N(0, t) = \frac{Q}{\sqrt{\pi \cdot D \cdot t}} = N_p - N_B = 7 \cdot 10^{20} \frac{1}{m^3} \quad (87)$$

$$x = 2 \cdot \sqrt{D \cdot t \cdot \ln\left(\frac{N_T}{N_B}\right)} = 4\mu m = 4 \cdot 10^{-6} m \quad (88)$$

$$\Rightarrow t \approx 16162 s \approx 269 min \approx 4h30min \quad (89)$$

$$\Rightarrow Q = 7 \cdot 10^{20} \frac{1}{m^3} \cdot \sqrt{\pi \cdot D \cdot t} = 7 \cdot 10^{20} \frac{1}{m^3} \cdot \sqrt{\pi} \cdot 2 \cdot 10^{-6} m \approx 2.48 \cdot 10^{15} \frac{1}{m^2} \quad (90)$$

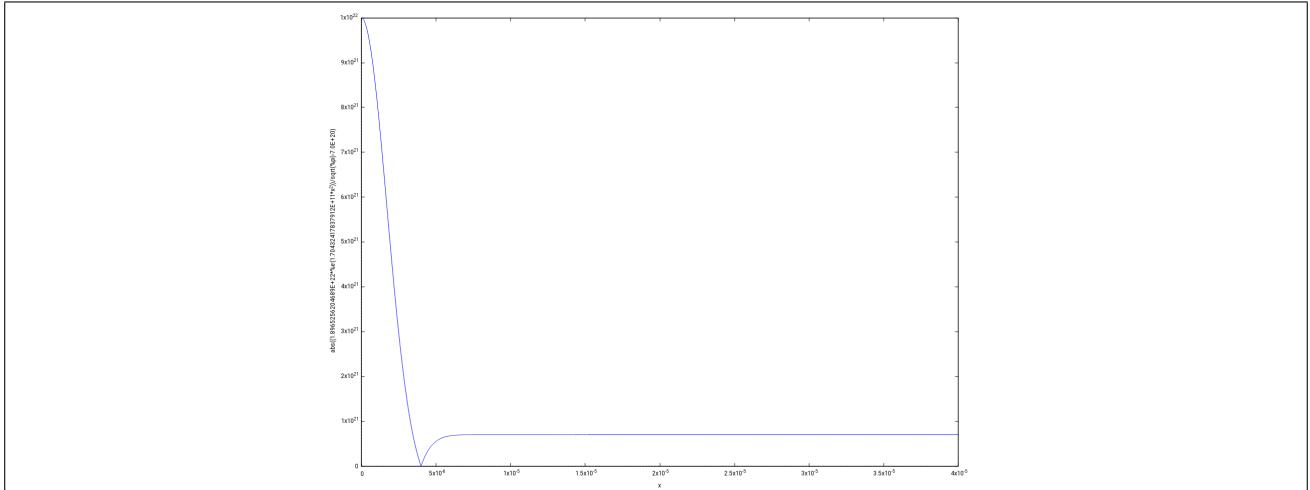


Figure 20: Dopant concentration after 4 hours 30 minutes

4.8 MOS gate verification

Now we have to verify that we don't have relevant leakage on the gate with the thresholds and doping values we've set.

For that we use the formulas from [subsubsection 4.5.1](#)

We can now plot multiple leakages for N- and P-channel transistors with a gate oxide thickness^{[30](#)} and with a surface concentration of $1e16 \frac{1}{cm^3} = 1e22 \frac{1}{m^3}$ each

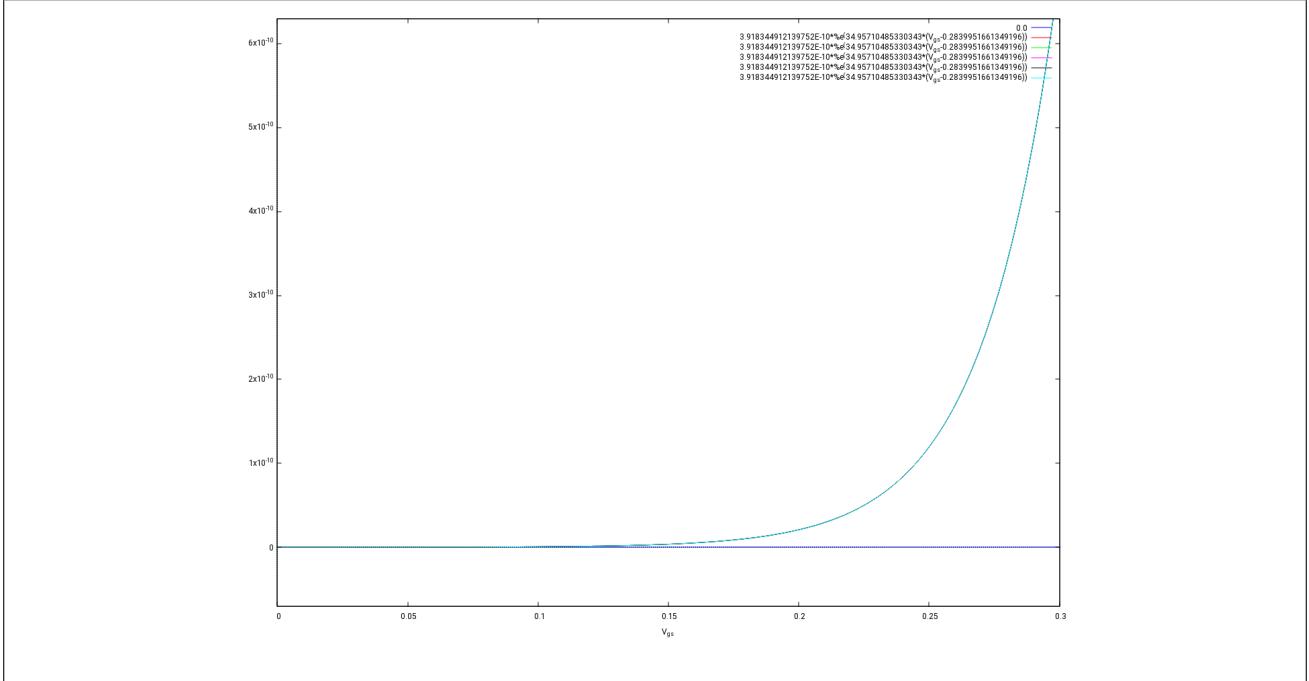


Figure 21: Subthreshold leakage plot(in Ampere)

In [Figure 21](#) we see that with our gate oxide thickness this is really no problem, as we had expected. From $0V$ up to $5V$ and further there is basically no leakage on the gate from the sub threshold current with $V_{Tn} \approx 0.39V$ and $V_{Tp} \approx -0.30V$. That's good enough, as we will see in [subsection 4.6](#) and [subsection 4.7](#). There is actually a reduction of current when reaching the threshold because of the inversion of the capacity in the depletion zone^{[31](#)}, but I didn't include this into the calculation, because "TL;DR". It's a TODO for release 2.1 of this process which will go sub $1\mu m$

³⁰See simulation/gate.wxmlx

³¹https://people.eecs.berkeley.edu/~hu/Chenming-Hu_ch5.pdf

4.9 Alignment strategy

When having multiple layers exposed after each other, there is the problem on how to make sure that for instance vias are actually making contact with the below wire and the wire above. For this purpose we have to align the masks, in order to avoid issues like shown in [Figure 22](#)

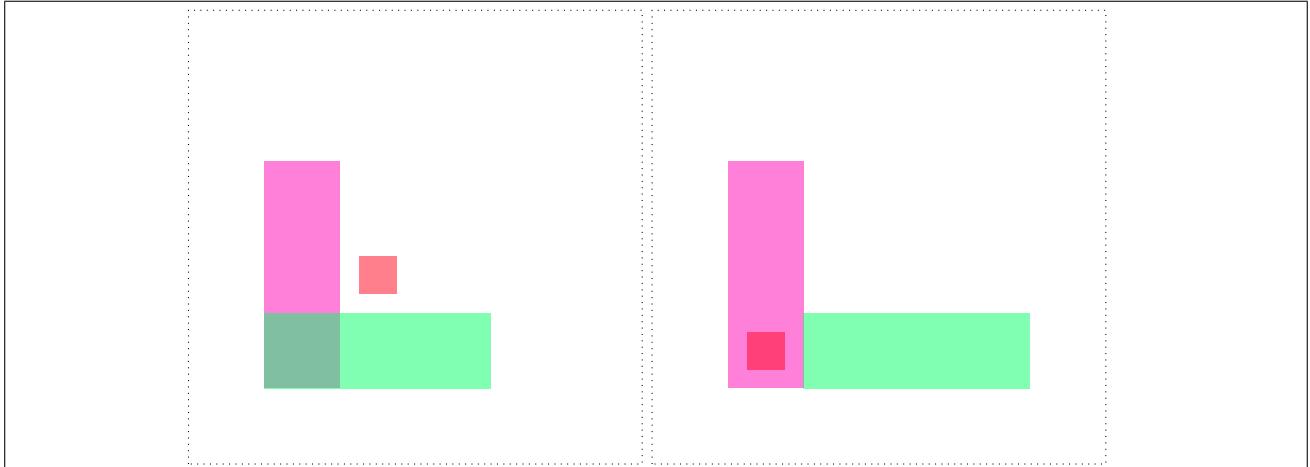


Figure 22: What can go wrong with alignment

In [Figure 22](#) we can see how the wires and vias are missing each other because of an exposure offset and the via is going nowhere in the best case and creates an (of course) undesired short circuit in the worst case. This has to be avoided by using alignment.

We have decided to use backside alignment because shining through the wafer from behind with infrared and finding an orientation marker isn't a problem with our simple CMOS process.³²

The stepper machine at HKUST³³ does that for us.

We need to add orientation cross hairs onto the layout edge in order to identify them during alignment

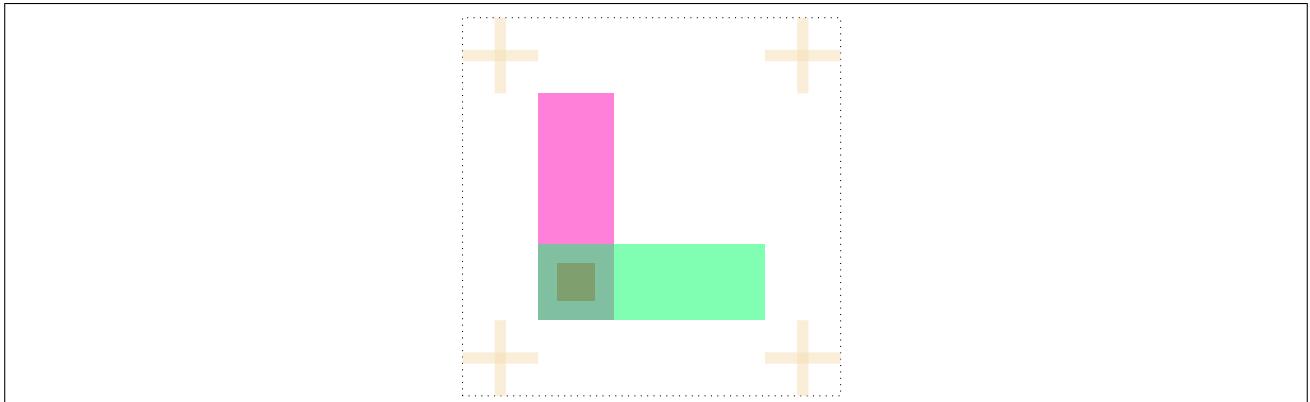


Figure 23: Mask alignment with markers

Using the STI (active) mask for the cross hair structure is a good choice, because it's the lowest layer, will never be covered by any additional material and gives us a good contrast because it's silicon next to silicon dioxide.

³²<https://patents.google.com/patent/US6376329>

³³<http://www.nff.ust.hk/en/equipment-and-process/equipment-list/photolithography-module.html>

The "DRIE Etcher" can't etch more precise than $0.5\mu m$ (Minimum Line/Space: $0.5\mu m$) which means we will have to give the cross hair structure with the dimensions as shown in [Figure 24](#)

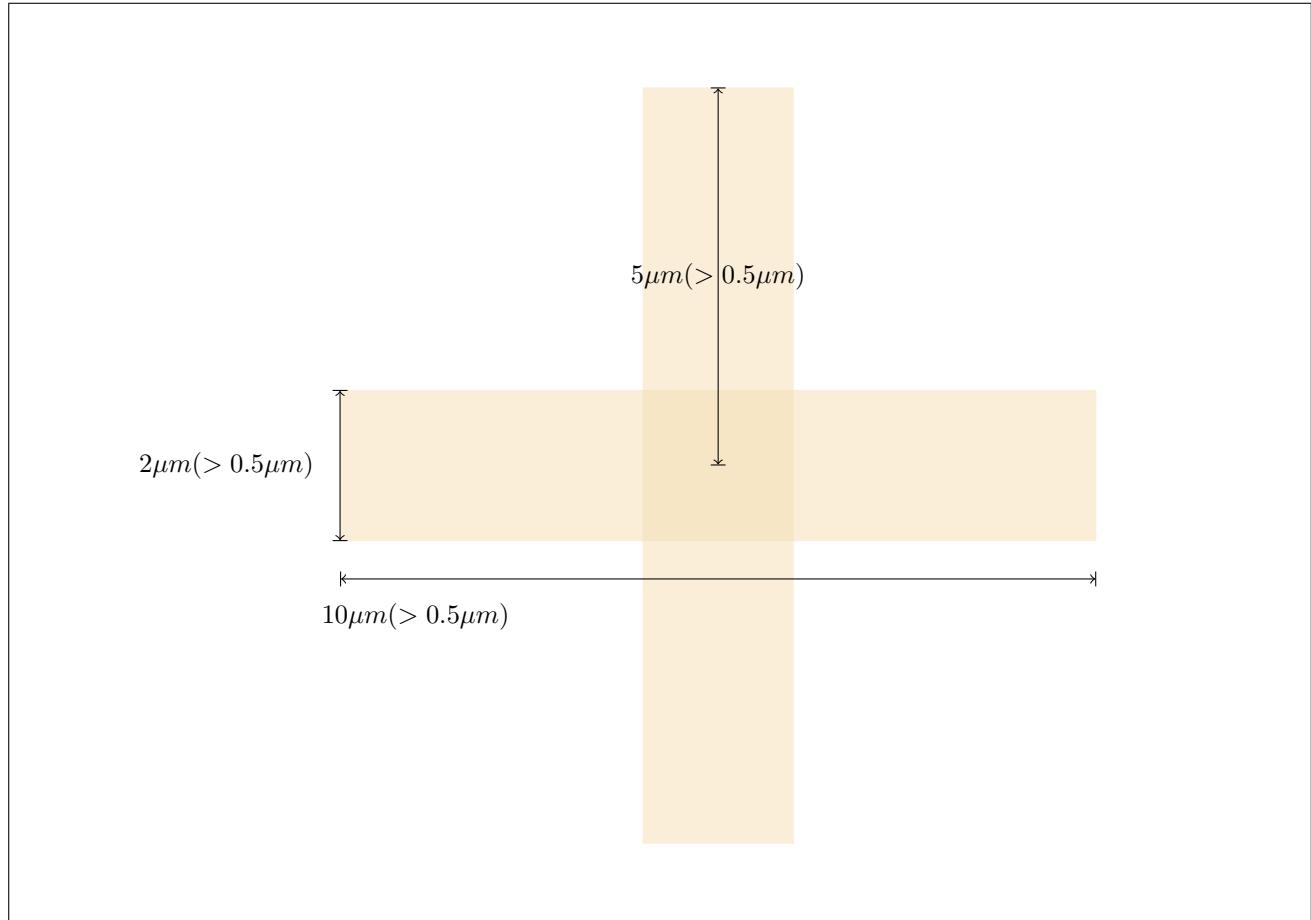


Figure 24: Cross hair dimension

We can now manufacture masks as shown in [Figure 25](#) and align them using the infrared light from the stepper mask alignment and the provided electronic microscope.

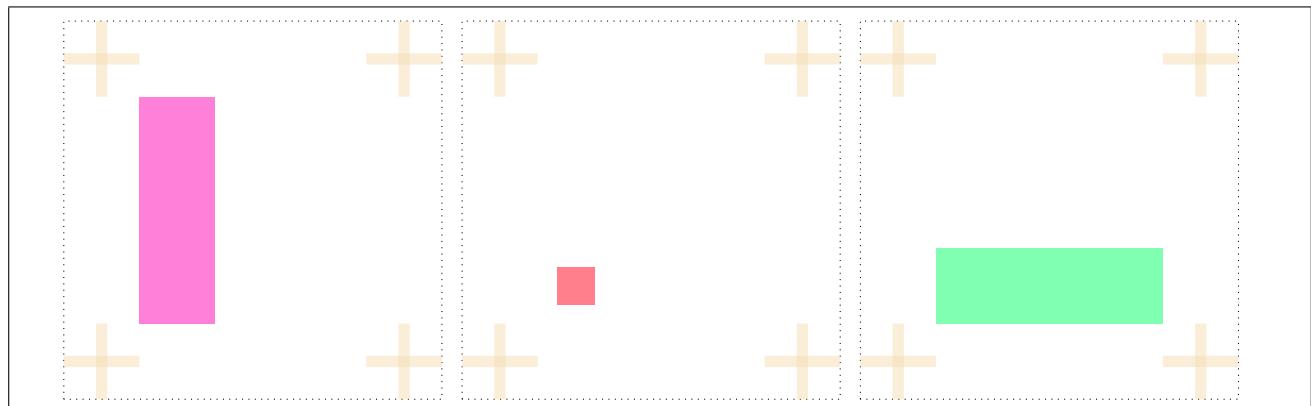


Figure 25: Layout masks with alignment markers

5 Simulation with parameters

6 Process steps

The general flow chart of the overall process flow can be seen in [Figure 26](#). These process steps will be discussed within the following sections.

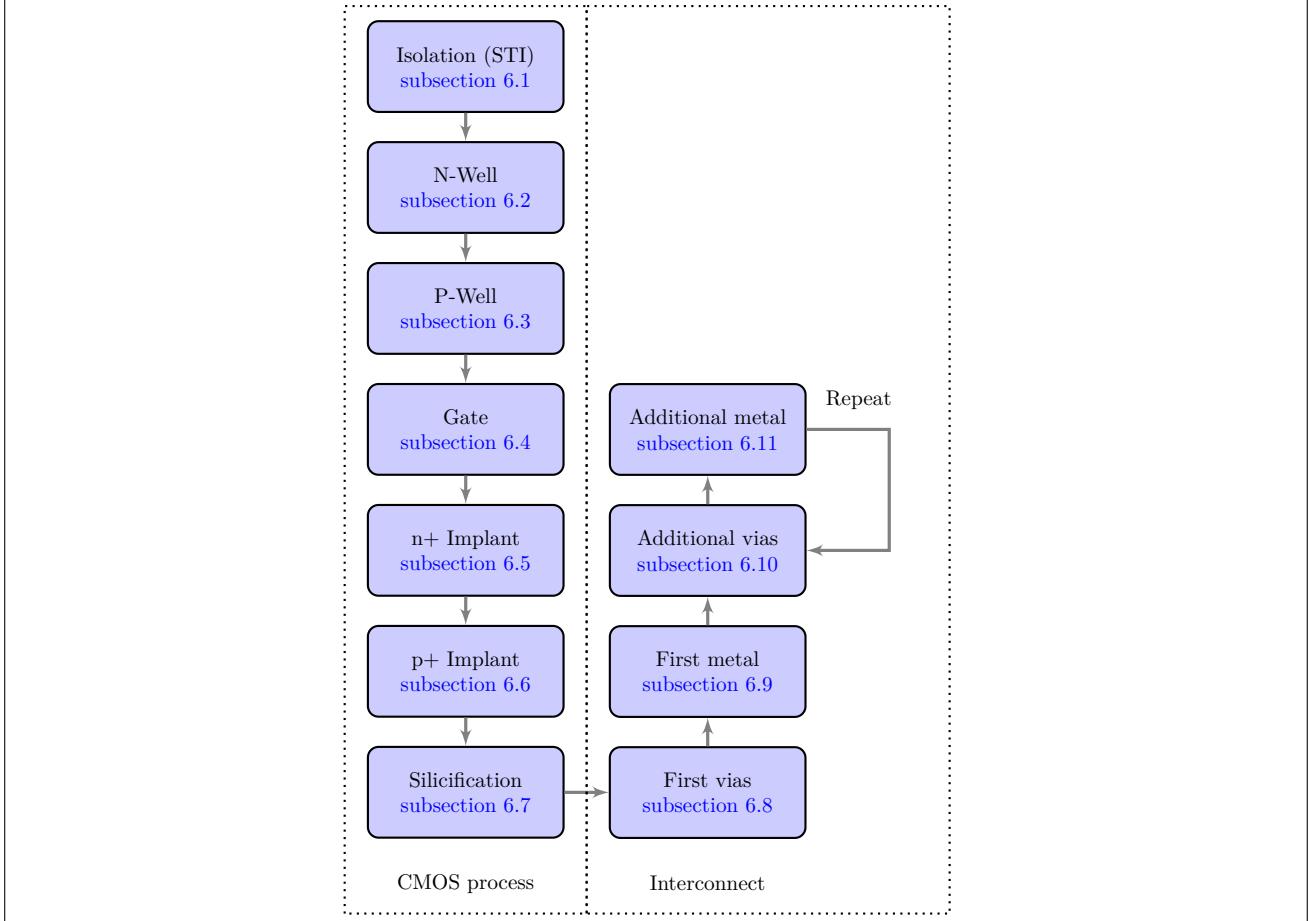


Figure 26: Frontend and backend process flow

The six overall process steps are part of an active part of the technology, while the final metal (respectively contact) layers will be used for making a contact between the logic gates and macro cells and making them available to the exterior world.

For this process p-substrate is the required basic substrate, but forks and modifications will be very well possible based on a Graphene substrate or alike, still under the LSPL. The starting material is a p-type, $<100>$ oriented silicon with a doping concentration of $\approx 9 \times 10^{14} cm^{-3}$.

6.1 Shallow trench isolation

The geometry of a substrate with STI implemented can be seen in [Figure 27](#).

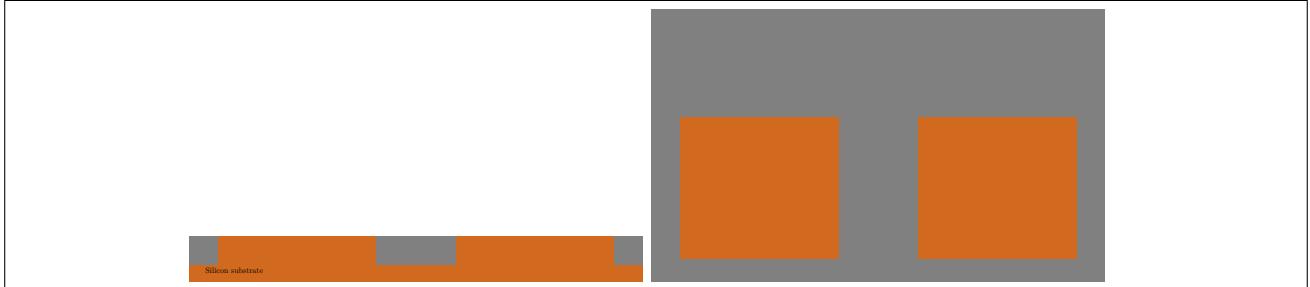


Figure 27: Shallow trench isolation target geometry

As can be seen in [section 1](#), the n-well and the STI trench are supposed to have approximately the same depth. Because the n-well will be $\approx 4\mu m$ in depth we have to match this with our trench depth.

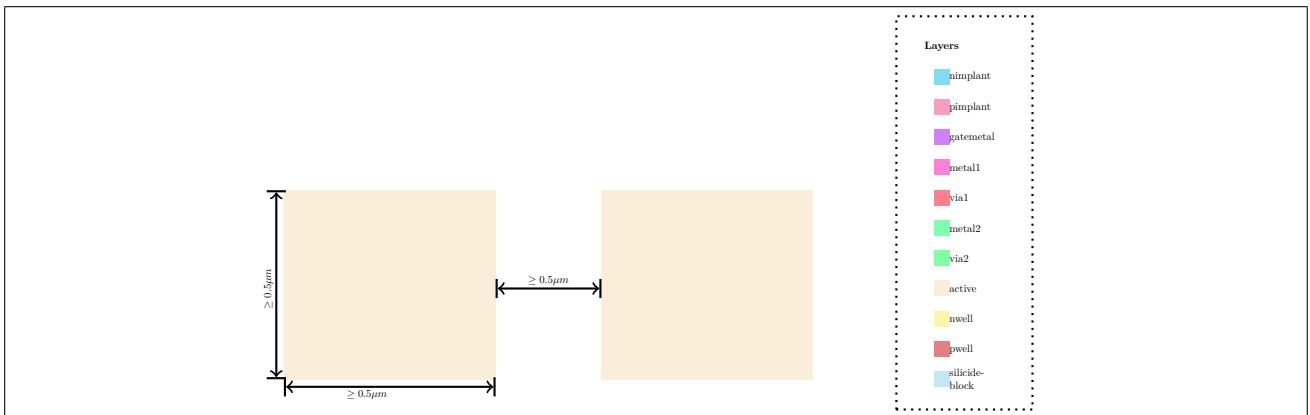


Figure 28: Shallow trench isolation layout

In [Figure 28](#) we can see the layout for the STI area. The STI area will be everywhere, where no active areas are. The deep isolating oxide needs to be grown out of trenches which can't been etched out of the silicon by using resist as a mask. For that reason we will have to resort to a protective mask made from a nitride layer which has to be etched before hand. So the mask will be exposed onto positive resist on top of the nitride in order to form a protective mask covering the active areas from having etched trenches into them as show in [subsubsection 6.1.7](#). After that we will use a dry etching method for cutting into the silicon substrate and making the active area become islands with trenches in between, as shown in [subsubsection 6.1.9](#). After these steps we have to remove the nitride mask, for which we expose the same mask again, only this time to a layer of inverted resist.

Our minimum width and height as well as the space between the active areas comes from the line space constrain of the silicon etcher ([subsection 9.3](#)).

6.1.1 Initial cleaning

In order to remove the initial naturally grown silicon dioxide from the wafer, acid is being applied to the wafer which leads to a pure silicon substrate as in the process illustration shown in [Figure 29](#).

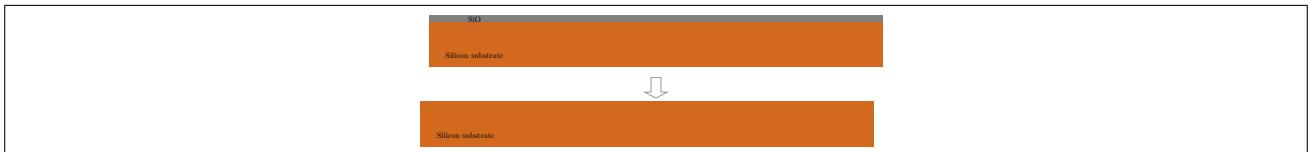


Figure 29: Initial cleaning

This needs to be done because the naturally grown initially existing silicon oxide is not pure and may contain contamination which may render the final product unusable.

6.1.2 Sulfuric Cleaning

The sulfuric acid mixture, $H_2SO_4 + H_2O_2$ is being applied to the wafer for 10 minutes at a temperature of 120 °C.

6.1.3 HF dip

After the sulfuric cleaning a HF (HF: H_2O ,1:50) dip is being performed for one minute.

Hydrofluoric acid (HF) is used to remove native silicon dioxide from wafers. Since it acts quickly, one needs to only expose the wafer for a short time ("dip").

After that the wafer needs to be dried and quickly processed further before new uncontrolled natural oxide can build up on the wafer through the contact with air.

6.1.4 Pad oxide

We need a thin layer of oxide as surface to grow our protective nitride layer on top.

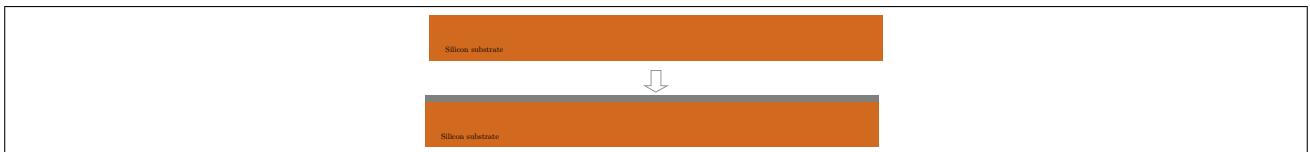


Figure 30: Pad oxide growth

The thin layer of "pad" oxide (around 300nm) is grown in dry ambient for 45 minutes at 1000°C.³⁴

6.1.5 Nitride layer

We need a protective nitride layer for dry etching the trenches into the silicon. This nitride will be grown in this step.

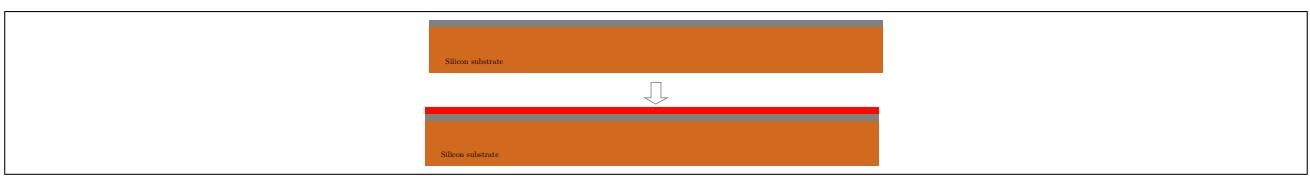


Figure 31: Nitride growth

The required thickness of this layer is not that critical, it can very well variate between 6nm and 10nm.³⁵ For this reason we can put it into the LPCVD for around one or two minutes as explained in [subsection 3.3](#).

³⁴<http://cleanroom.byu.edu/OxideTimeCalc>

³⁵<https://www.google.com/patents/US7985656>

6.1.6 Patterning positive

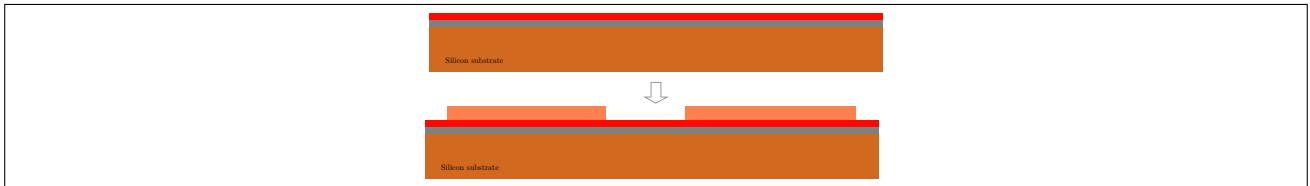


Figure 32: Patterning with positive resist

6.1.7 Nitride etching

We open the access to the silicon outside of the active areas in order to etch the trenches.

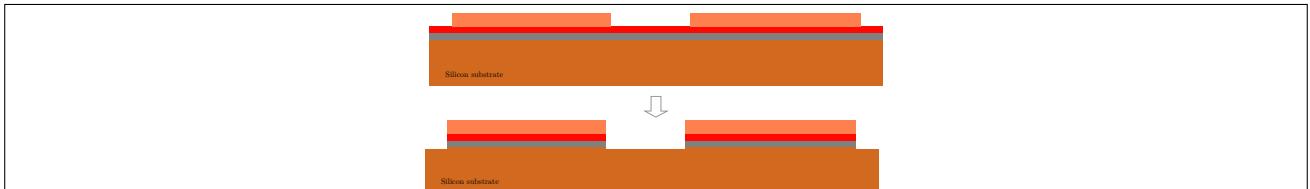


Figure 33: Nitride mask etching

We use the reflux method as described in [subsection 3.2](#)

6.1.8 Resist removal

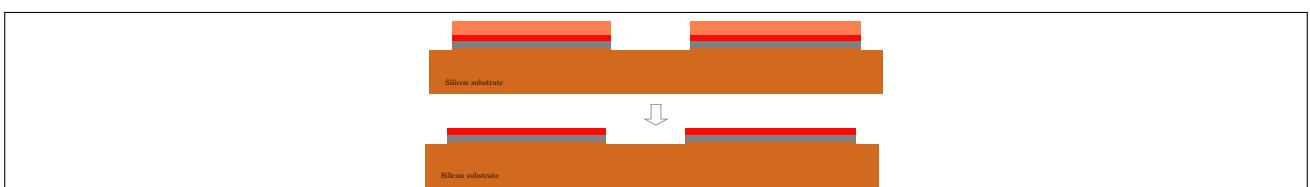


Figure 34: Resist removal

6.1.9 Silicon etching

Silicon can only be etched by a very aggressive chemical cocktail of KOH and TMAH (25%) or by plasma etching.

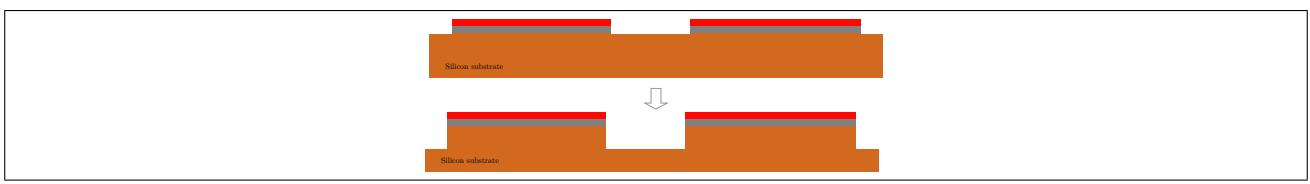


Figure 35: Trench etching

We take the machine "DRIE Etcher #1" from HKUST([subsection 9.3](#)) as reference here, which has a normal etching rate of up to $2 \frac{\mu m}{min}$.

6.1.10 Deep oxidation

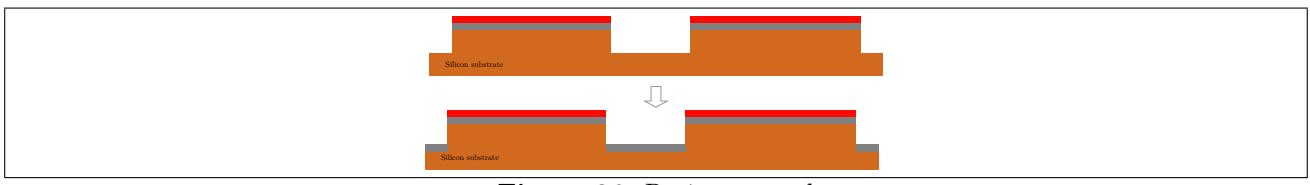


Figure 36: Resist removal

6.1.11 Oxide deposition

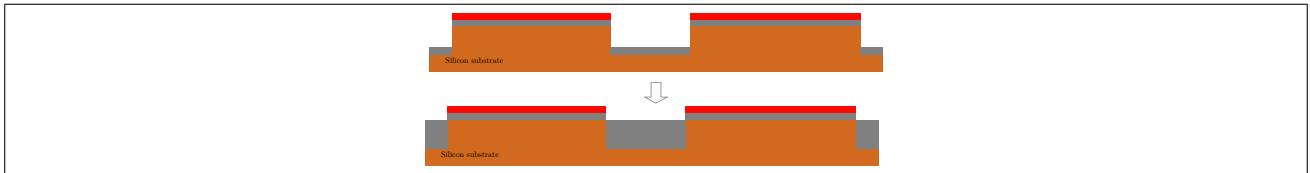


Figure 37: Resist removal

6.1.12 Hard mask removal

Now we have to remove the nitride mask for further processing.

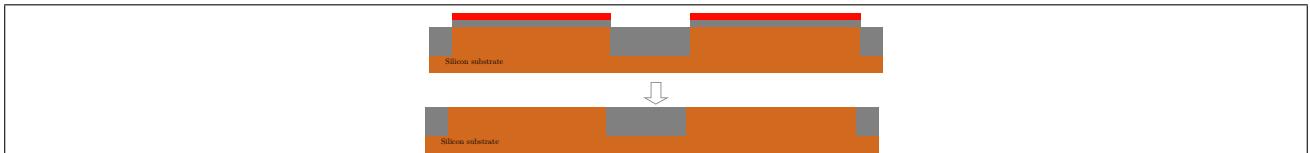


Figure 38: Trench etching

We use a CMP machine. The HKUST lab provides multiple "Buehler Polisher" machines(??), which allow polishing away the hard mask **and** evening out the uneven oxide deposition in one single step!

6.2 N-well

In order to build CMOS on the same substrate, an n-well is required for building the complementary P-channel transistor for a n-p-channel logic circuitry as shown above in the example section. The cross section as well as the top view of the targeted geometry are shown in [Figure 39](#)

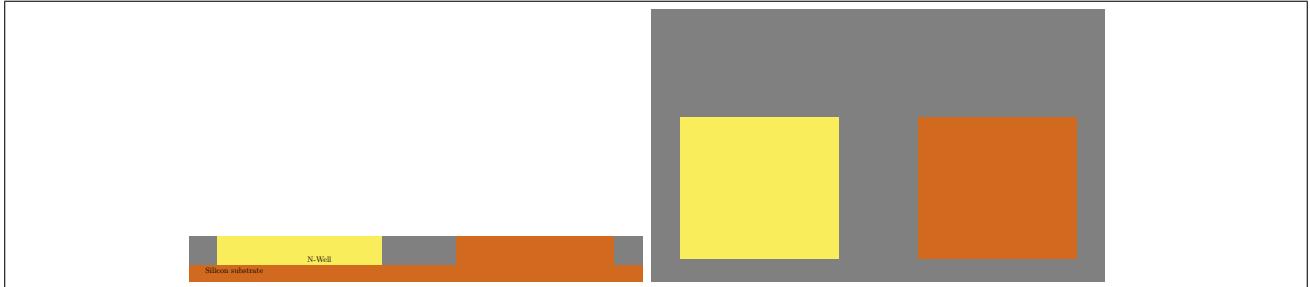


Figure 39: N-well target geometry

The n-well will serve us as an island of n-doped substrate within the p-doped basis substrate. The dopant dose will be $2.5 \times 10^{12} \text{ cm}^{-2}$ as calculated in [subsection 4.7](#).

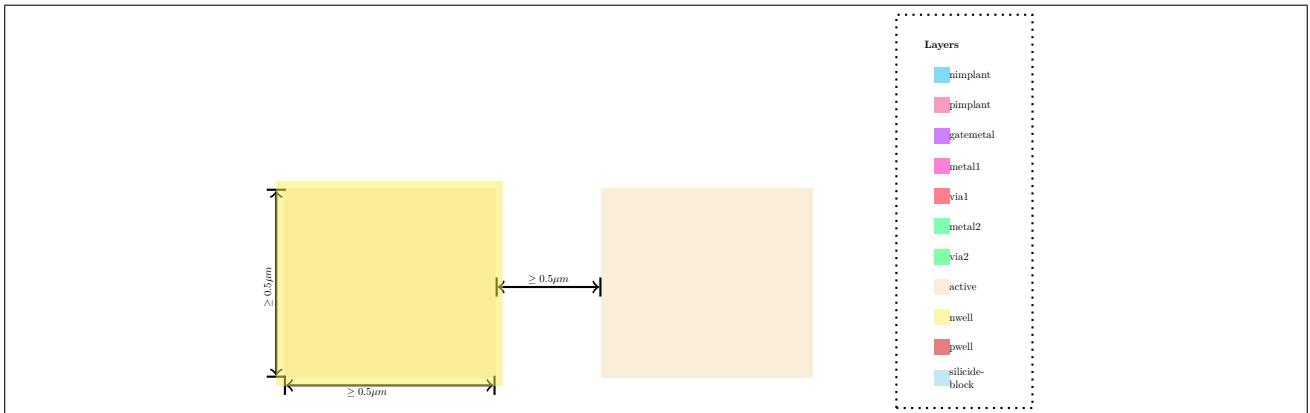


Figure 40: N-Well layout

In [Figure 40](#) the layout of the n-well region on top of the active area region can be seen. The n-well is being fit into the active area. It should even be a little bit bigger than the active area, because of possible alignment offsets

6.2.1 Mask dioxide layer

In order to selectively inject charge carrying atoms into the crystalline structure a protective dioxide (SiO_2) layer needs to be grown on top of a p-type substrate.

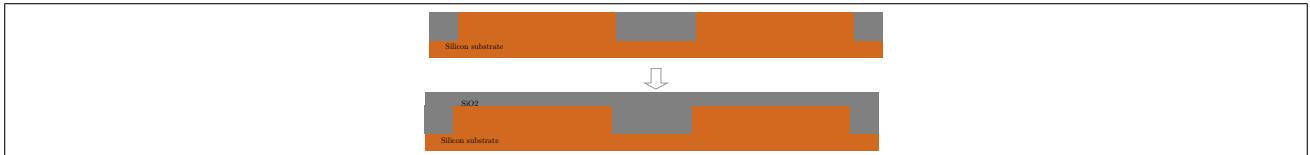


Figure 41: Dioxide layer growth

The industrial best practice is a layer of around (500nm≈5000Å) thickness or more. For this purpose the wafer is being oxidized for at least 90 minutes at 1000°C using wet oxidation which results in a dioxide layer at least 500nm(≈5000Å) in thickness.

6.2.2 Patterning

The resist is being deposited using spin coating and then baked depending on the baking time for the specific resist. The layout for being exposed onto the resist is being extracted from the "nwell" layer within the GDS2 file.

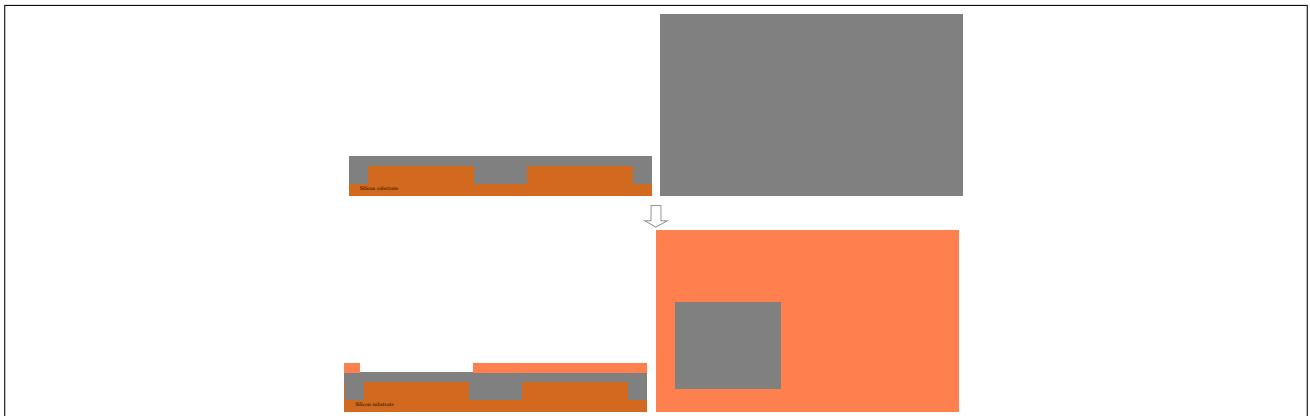


Figure 42: Cross/top view of n-well layout on resist

The thickness of the resist layer and the baking duration will variate depending on the specific equipment for which this process will be implemented with.

6.2.3 Etching

We now need to open a window in the dioxide layer, through which we will inject carrier atoms into the silicon crystal structure.

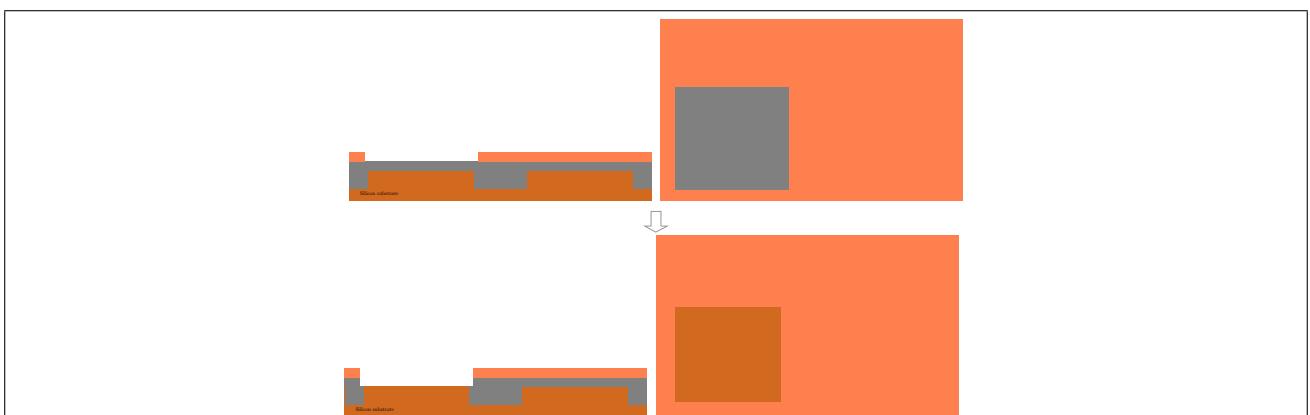


Figure 43: Cross/top view of n-well oxide window

Since the silicon dioxide layer is 500nm thick and we wanna reach the silicon below we can use wet etching as described in the chemistry chapter. Using BHF (6:1) (Equation 3.1) we can etch with a speed of approximately

2 nm/s at 25 °C, we can calculate the etching time to be $\frac{500\text{nm}}{2\text{nm/s}}=250\text{s}=4 \text{ minutes } 10 \text{ seconds}$ (or make it rather 30 seconds instead of 10)

6.2.4 Cleaning

In order to avoid contamination of the machines we need to make sure all the resist has been stripped off from the wafer.

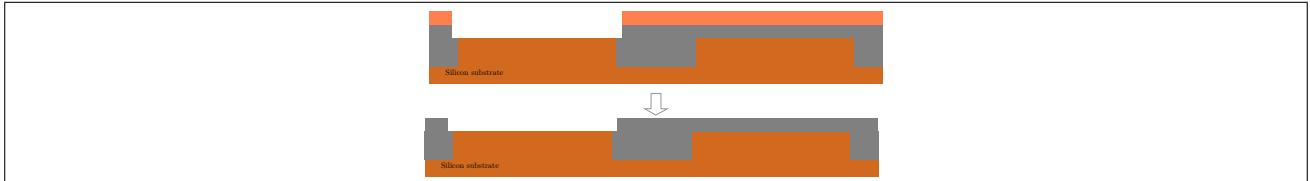


Figure 44: Resist removal

Please just use the solvent for the specific resist.

6.2.5 Injection

We now need to inject the carriers into the upper level of the n-channel area so that we can later on drive them into the crystal during the drive-in step.

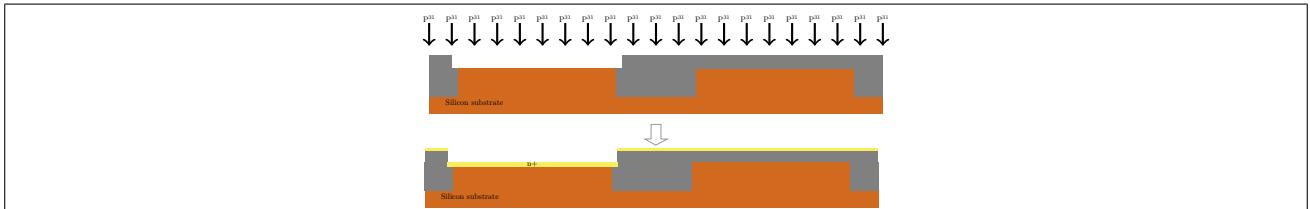


Figure 45: Doping process

The n-well is implanted with a Phosphorus (P^{31}) dose of $2.5 \times 10^{12} \text{ cm}^{-2}$ at an energy of 100 KeV. The n-well is then annealed.

6.2.6 Oxide for drive-in

Now we need to cover the now doped and annealed areas with an oxide layer for the drive-in phase.

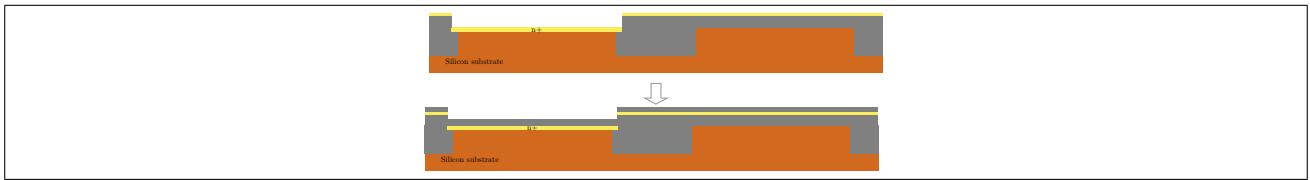


Figure 46: Oxide growth

The wafer is being oxidized for 32 minutes at 1000°C in order to achieve a cover silicon layer of 250nm thickness ($\approx 2500\text{\AA}$).

6.2.7 Drive-in

In order to drive the carrier atoms deeper into the crystalline structure the wafer needs to be driven in after predeposition.

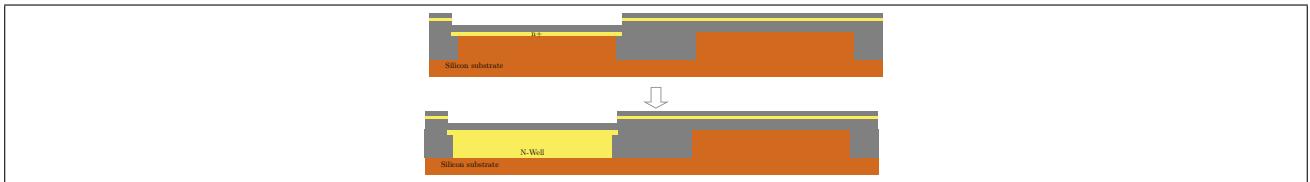


Figure 47: Drive-in process

In this step the wafer is driven-in for 96 minutes at 1150°C in an inert ambient.

6.2.8 Oxide mask removal

We want to remove the silicon mask from the wafer so that the n-well becomes accessible for the further process steps but we don't want to etch "way too much" of the trench material.

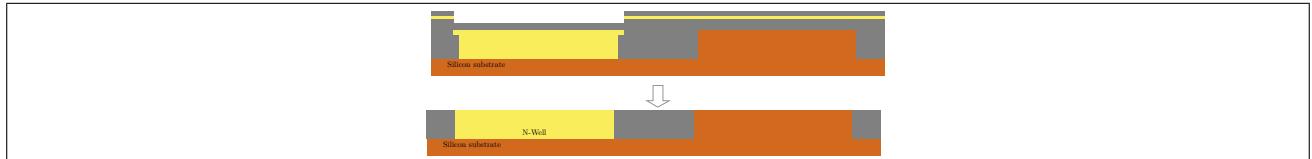


Figure 48: Oxide removal

Since the silicon dioxide layer is 750nm (500nm+250nm) thick and we wanna reach the silicon below we can use wet etching as described in the chemistry chapter. Using BHF (6:1) ([Equation 3.1](#)) we can etch with a speed of approximately 2 nm/s at 25 °C. We can calculate the etching time to be $\frac{750\text{nm}}{2\text{nm/s}}=375\text{s}=6 \text{ Minutes and } 15 \text{ Seconds}$.

Etching away a "little bit too much" of the oxide isn't that bad, because the oxide within the trenches will be "filled up" again during the later steps.

6.3 P-well

In order to build CMOS on the same substrate, an P-well is required for building the complementary P-channel transistor for a n-p-channel logic circuitry as shown above in the example section. The cross section as well as the top view of the targeted geometry are shown in [Figure 39](#)

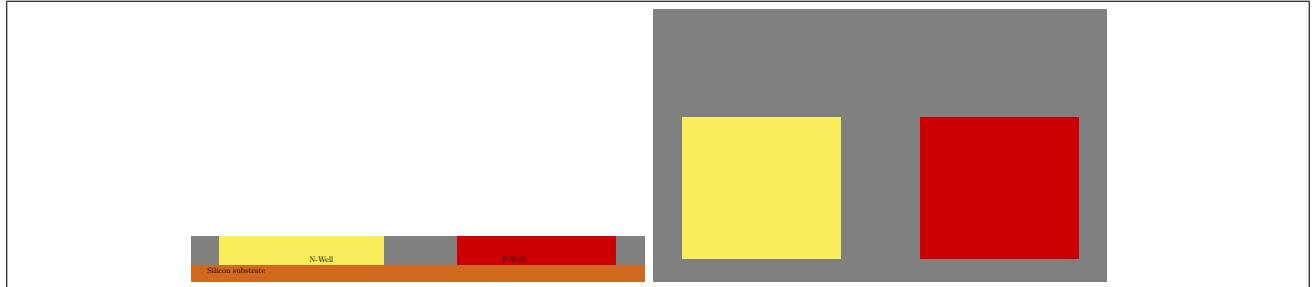


Figure 49: P-well target geometry

The P-well will serve us as an island of higher p-doped substrate within the slightly p-doped basis substrate. The dopant dose will be: $2.5 \times 10^{12} cm^{-2}$

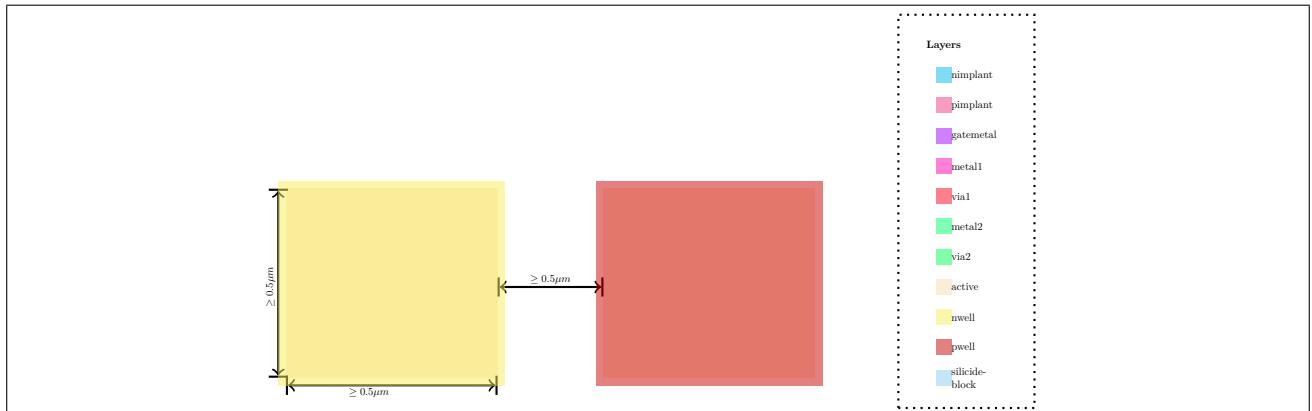


Figure 50: P-Well layout

In [Figure 50](#) the layout of the P-well region on top of the active area region can be seen.

The p-well is being fit into the active area.

It should even be a little bit bigger than the active area, because of possible alignment offsets

6.3.1 Mask dioxide layer

In order to selectively inject charge carrying atoms into the crystalline structure a protective dioxide (SiO_2) layer needs to be grown on top of a p-type substrate.

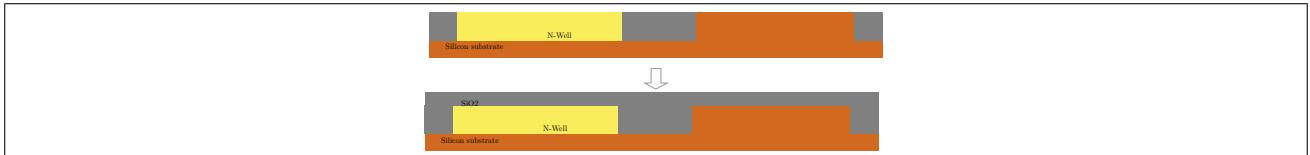


Figure 51: Dioxide layer growth

The industrial best practice is a layer of around (500nm≈5000Å) thickness or more. For this purpose the wafer is being oxidized for at least 90 minutes at 1000°C using wet oxidation which results in a dioxide layer at least 500nm(≈5000Å) in thickness.

6.3.2 Patterning

The resist is being deposited using spin coating and then baked depending on the baking time for the specific resist. The layout for being exposed onto the resist is being extracted from the "nwell" layer within the GDS2 file.

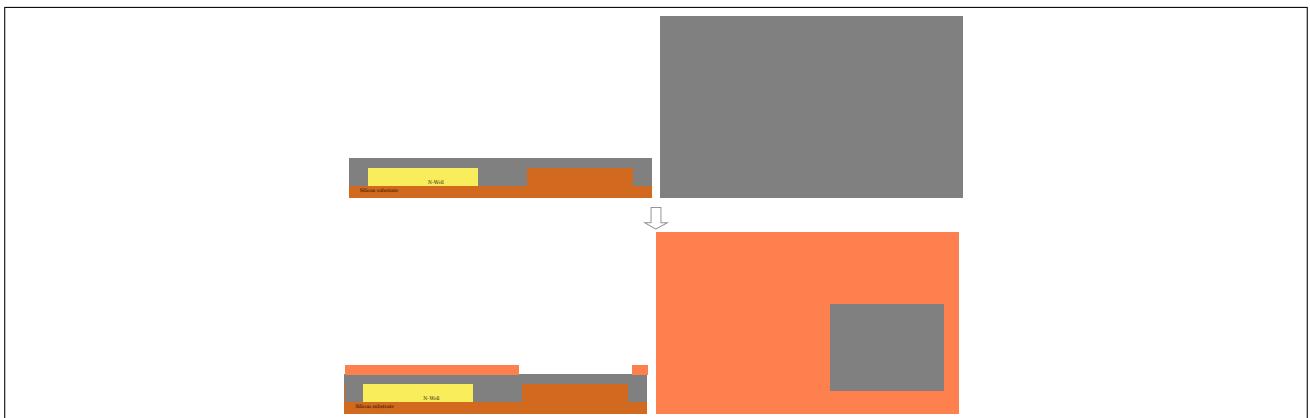


Figure 52: Cross/top view of P-well layout on resist

The thickness of the resist layer and the baking duration will variate depending on the specific equipment for which this process will be implemented with.

6.3.3 Etching

We now need to open a window in the dioxide layer, through which we will inject carrier atoms into the silicon crystal structure.

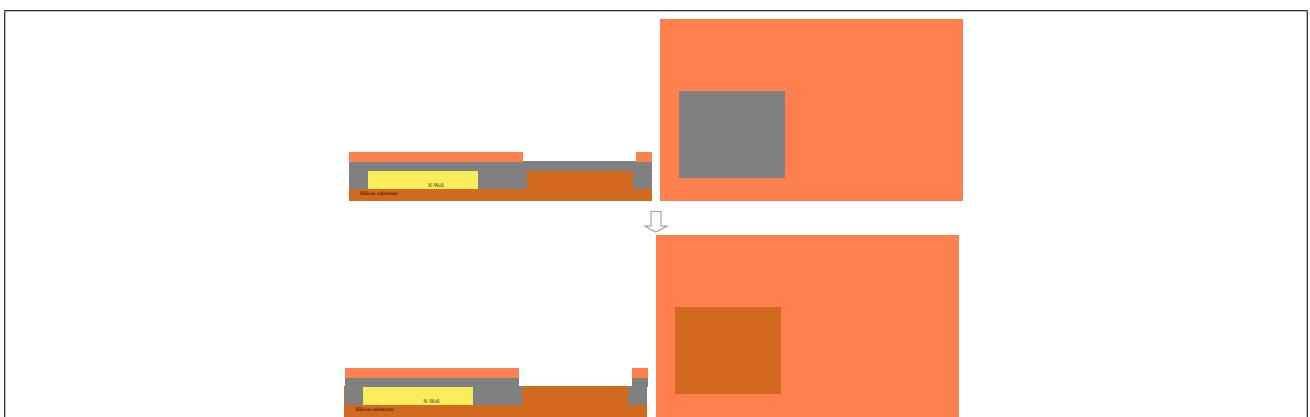


Figure 53: Cross/top view of P-well oxide window

Since the silicon dioxide layer is 500nm thick and we wanna reach the silicon below we can use wet etching as described in the chemistry chapter. Using BHF (6:1) (Equation 3.1) we can etch with a speed of approximately

2 nm/s at 25 °C, we can calculate the etching time to be $\frac{500\text{nm}}{2\text{nm/s}}=250\text{s}=4 \text{ minutes } 10 \text{ seconds}$ (or make it rather 30 seconds instead of 10)

6.3.4 Cleaning

In order to avoid contamination of the machines we need to make sure all the resist has been stripped off from the wafer.



Figure 54: Resist removal

Please just use the solvent for the specific resist.

6.3.5 Injection

We now need to inject the carriers into the upper level of the n-channel area so that we can later on drive them into the crystal during the drive-in step.

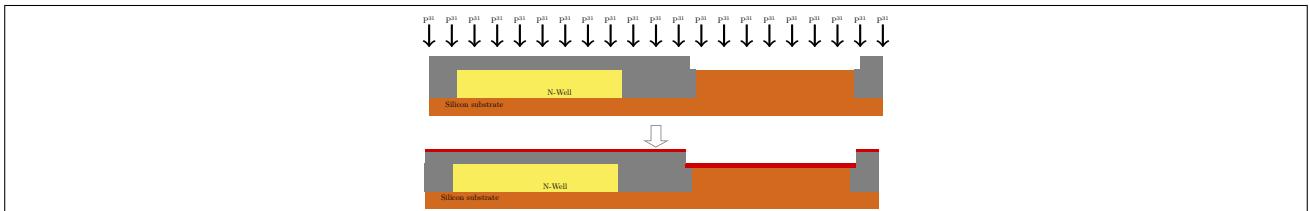


Figure 55: Doping process

The P-well is implanted with a Phosphorus (P^{31}) dose of $2.5 \times 10^{12} \text{cm}^{-2}$ at an energy of 100 KeV. The P-well is then annealed.

6.3.6 Oxide for drive-in

Now we need to cover the now doped and annealed areas with an oxide layer for the drive-in phase.

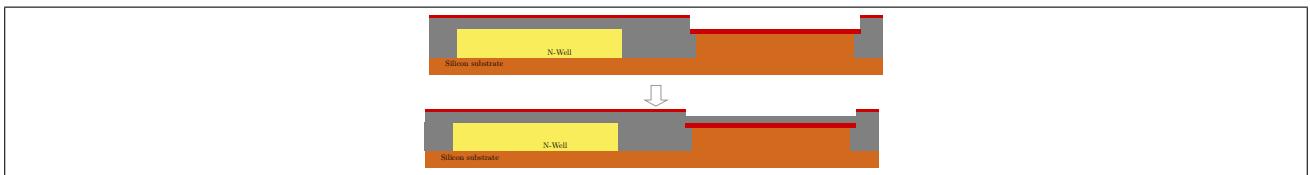


Figure 56: Oxide growth

The wafer is being oxidized for 32 minutes at 1000°C in order to achieve a cover silicon layer of 250nm thickness ($\approx 2500\text{\AA}$).

6.3.7 Drive-in

In order to drive the carrier atoms deeper into the crystalline structure the wafer needs to be driven in after predeposition.

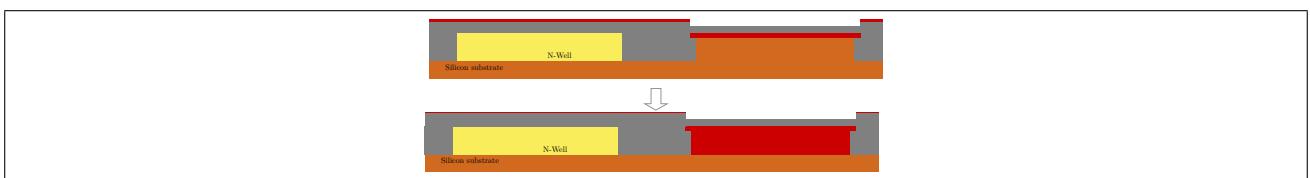


Figure 57: Drive-in process

In this step the wafer is driven-in for 96 minutes at 1150°C in an inert ambient.

6.3.8 Oxide mask removal

We want to remove the silicon mask from the wafer so that the P-well becomes accessible for the further process steps but we don't want to etch "way too much" of the trench material.

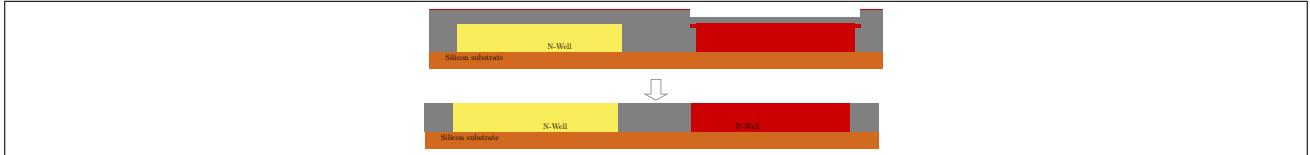


Figure 58: Oxide removal

Since the silicon dioxide layer is 750nm (500nm+250nm) thick and we wanna reach the silicon below we can use wet etching as described in the chemistry chapter. Using BHF (6:1) ([Equation 3.1](#)) we can etch with a speed of approximately 2 nm/s at 25 °C. We can calculate the etching time to be $\frac{750\text{nm}}{2\text{nm/s}}=375\text{s}=6 \text{ Minutes and } 15 \text{ Seconds}$.

Etching away a "little bit too much" of the oxide isn't that bad, because the oxide within the trenches will be "filled up" again during the later steps.

6.4 Gate

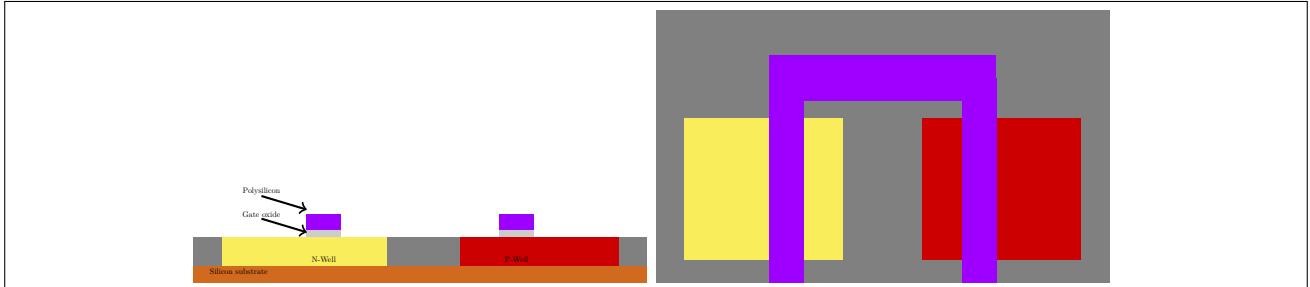


Figure 59: Aluminum gate contacts with gate oxide

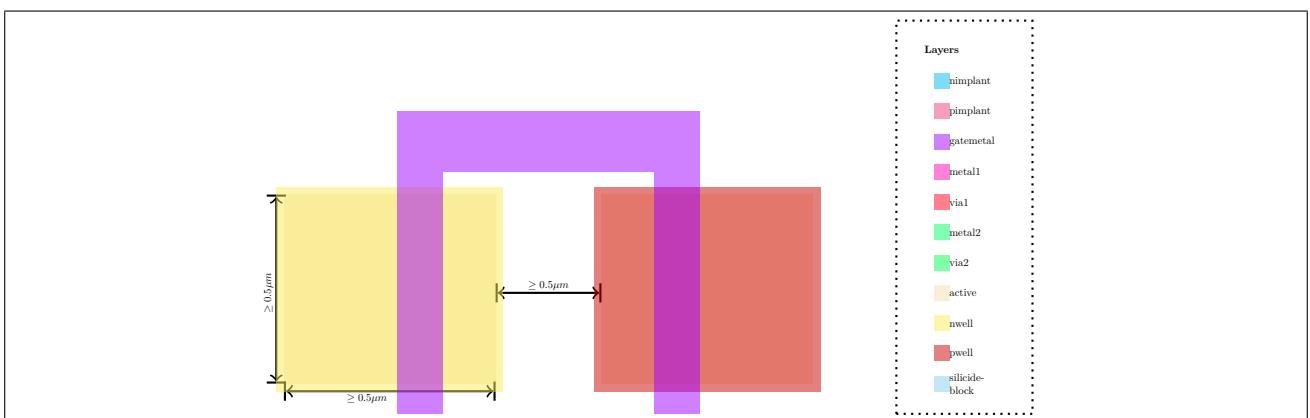


Figure 60: Gate layout

6.4.1 Gate oxide deposition



Figure 61: Thin oxide

6.4.2 Polysilicon deposition



Figure 62: Polysilicon

6.4.3 Patterning

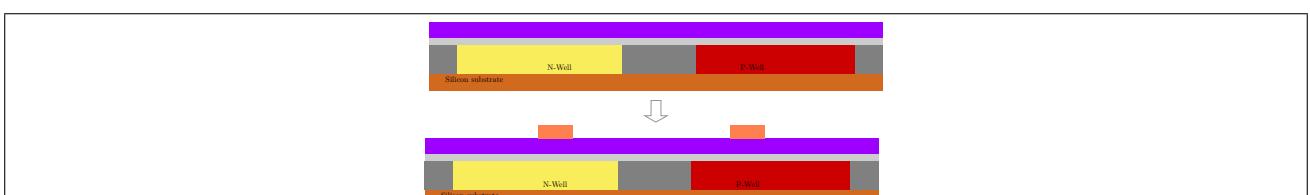


Figure 63: Resist

6.4.4 Etching

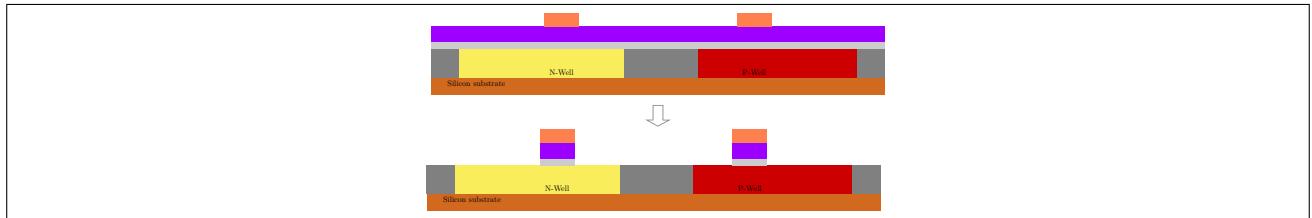


Figure 64: Resist

6.4.5 Cleaning

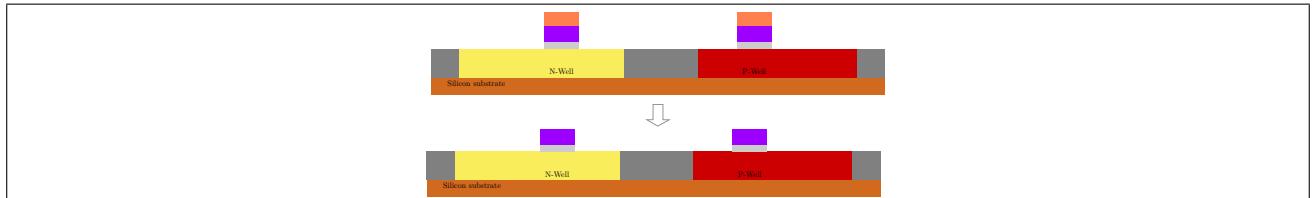


Figure 65: Resist

6.5 n+ Implant

For the bulk of the PMOS transistors and for the source and drain of the NMOS transistors highly doped n+ areas are required. In this step we're going to build these.

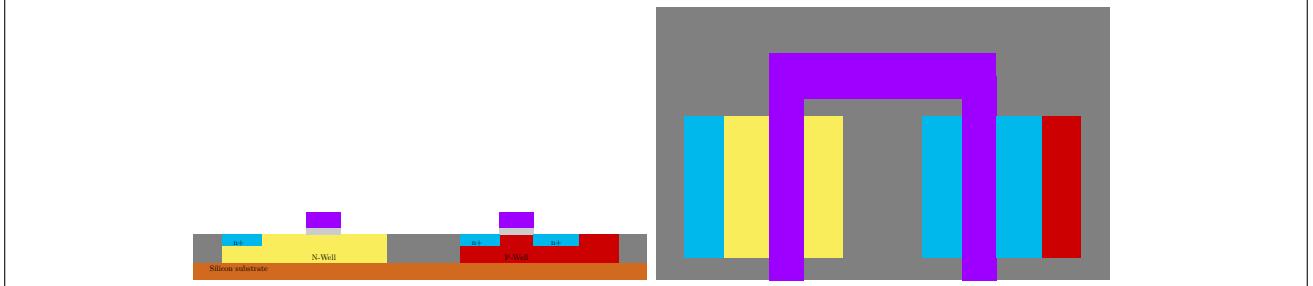


Figure 66: N+ implant geometry target

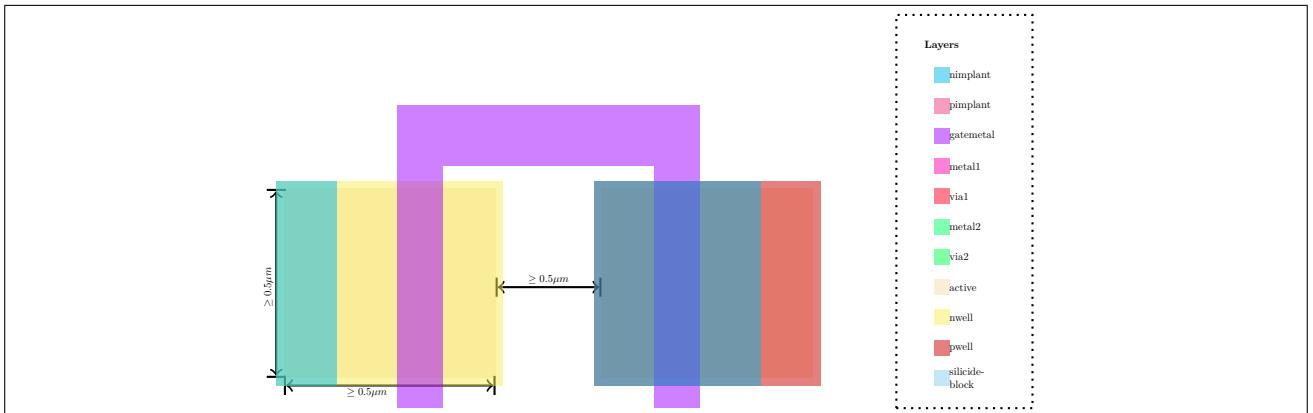


Figure 67: N+ layout

6.5.1 Mask dioxide layer

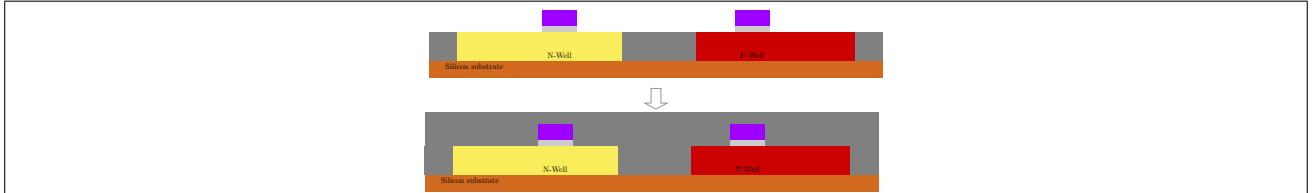


Figure 68: Oxide layer

6.5.2 Patterning

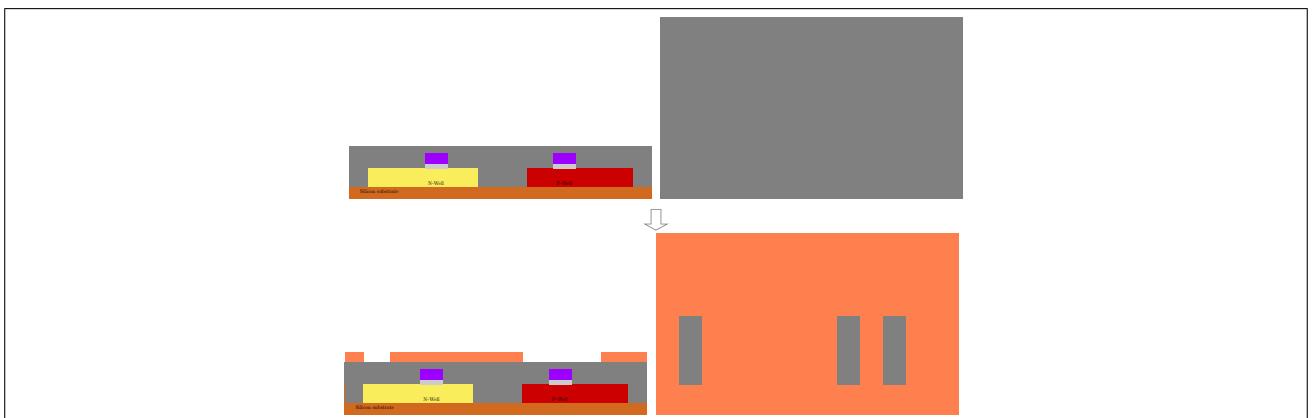


Figure 69: N+ region resist mask

6.5.3 Etching

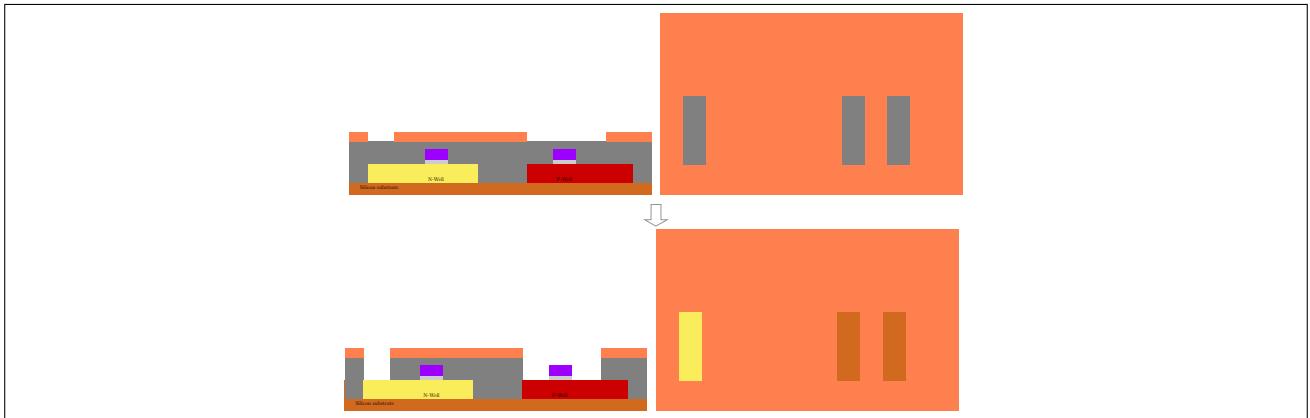


Figure 70: N+ region opened

6.5.4 Cleaning

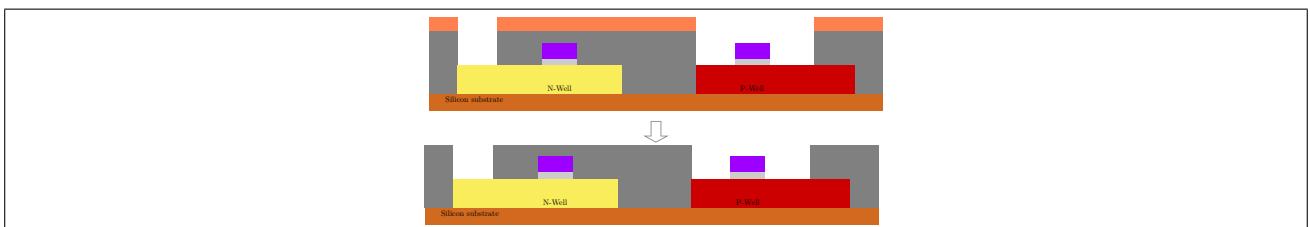


Figure 71: Resist removal

6.5.5 Injection

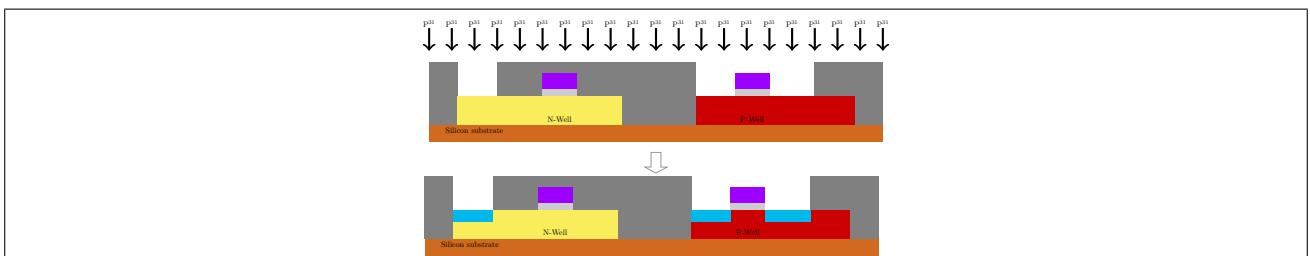


Figure 72: N+ injection process

6.5.6 Oxide removal

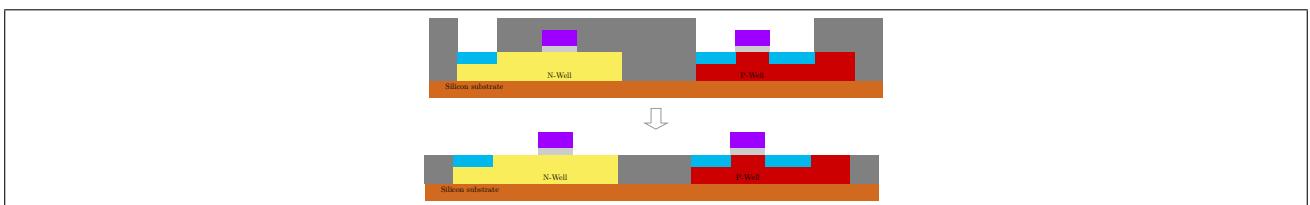


Figure 73: Oxide removal

6.6 p+ Implant

For the bulk of the NMOS transistors and for the source and drain of the PMOS transistors highly doped p+ areas are required. In this step we're going to build these.

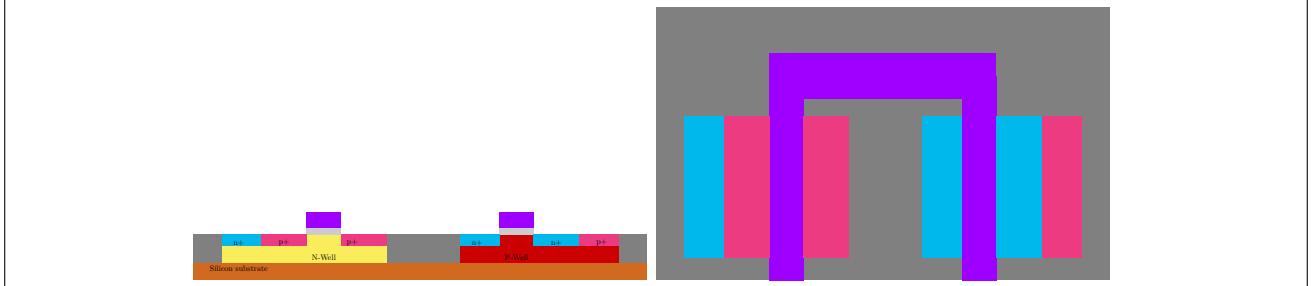


Figure 74: P+ implant geometry target

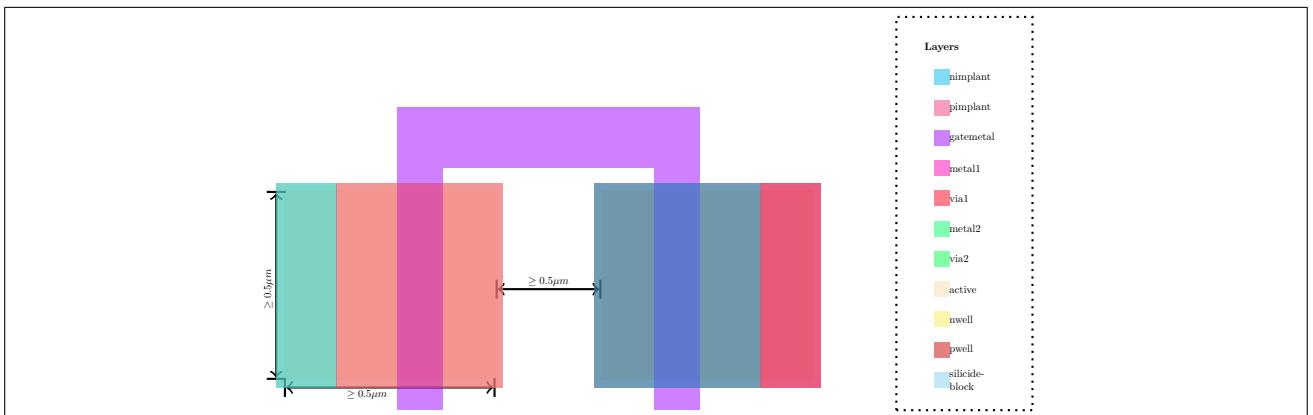


Figure 75: P+ layout

6.6.1 Mask dioxide layer

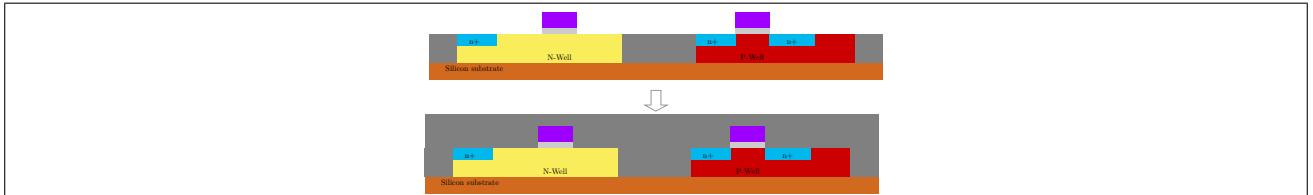


Figure 76: Oxide layer

6.6.2 Pattering

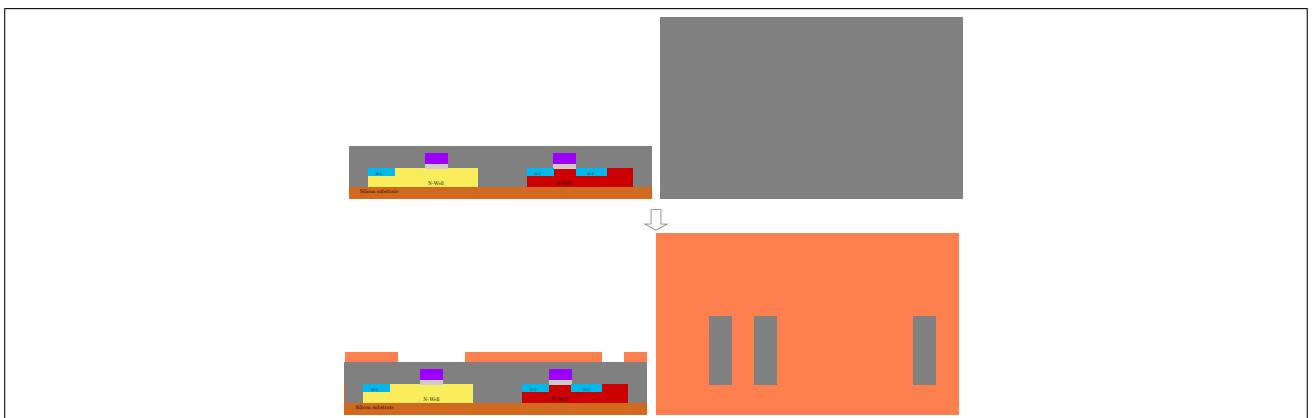


Figure 77: P+ region resist mask

6.6.3 Etching

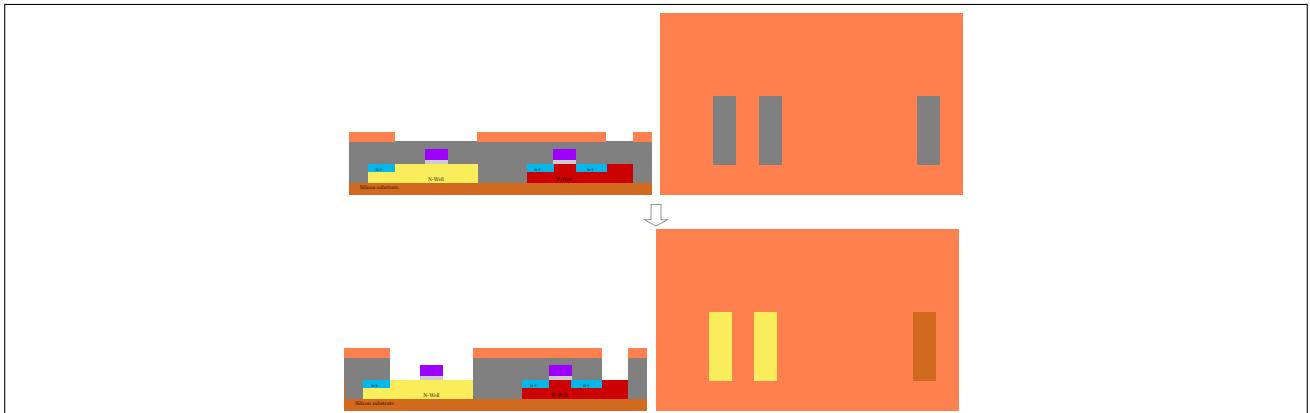


Figure 78: P+ region opened

6.6.4 Cleaning

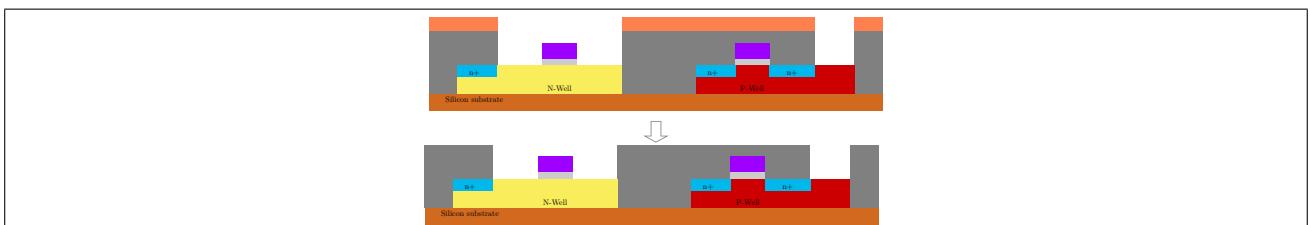


Figure 79: Resist removal

6.6.5 Injection

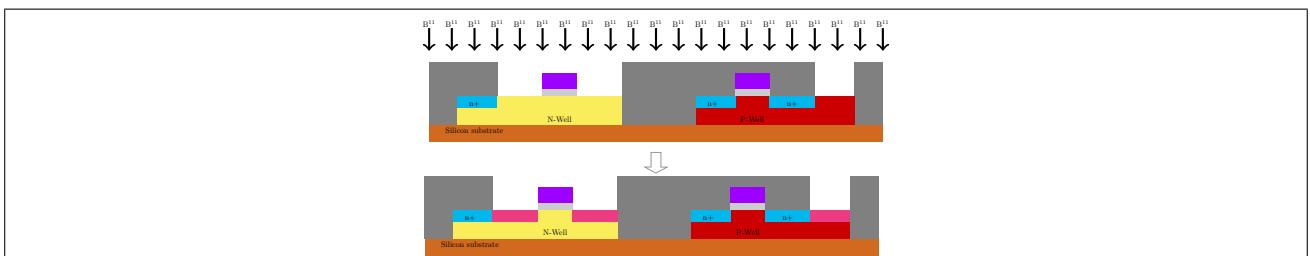


Figure 80: P+ injection process

6.6.6 Oxide removal

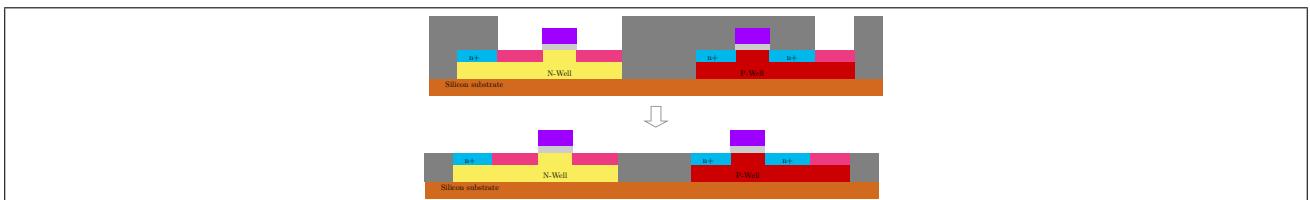


Figure 81: Oxide removal

6.7 Silicification

Titanium silicide is one of the first SALICIDE material introduced in ULSI devices owing to its low resistivity, high thermal stability, ease in deposition and compatibility with silicon processes. Titanium has been one of the familiar materials in ULSI productions, which is also an important advantage in practical use of titanium SALICIDE.³⁶

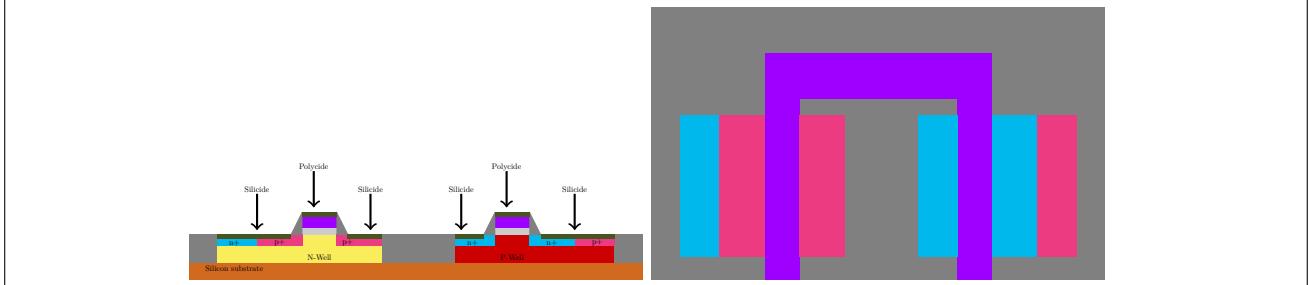


Figure 82: Silicide geometry target

In order to reduce the gate contact resistance as well as the source and drain resistance and in order to provide a more effective etch stop when plasma etching the contact windows to drain, source and gate, silicide/polycide is being added to the wafer as shown in Figure 82.

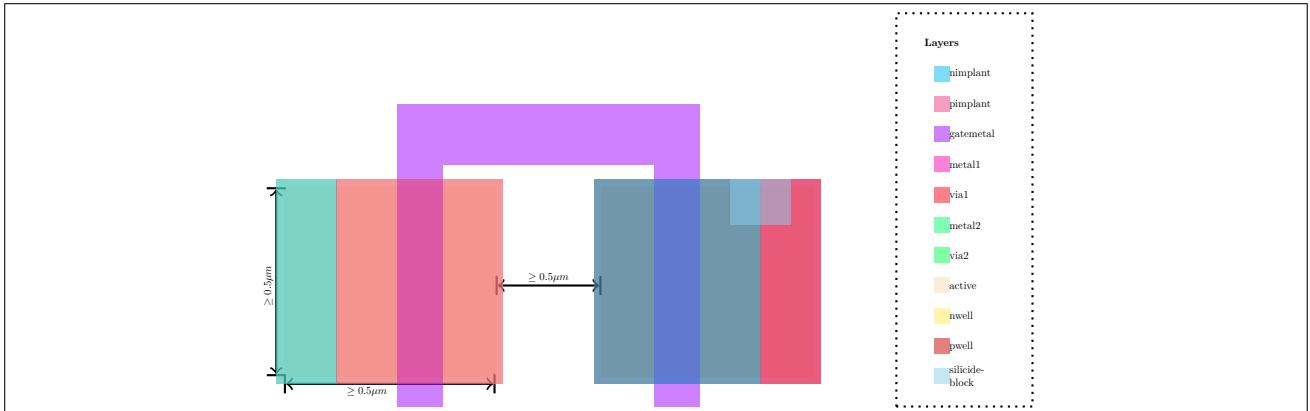


Figure 83: Silicification layout

When titanium and silicon are brought into contact and heated at temperatures above 500 °C (in the presence of excess silicon) the higher-resistivity C49- $TiSi_2$ phase forms before the low-resistivity phase.

The C49- $TiSi_2$ phase has an orthorhombic base-centered structure with 12 atoms per unit cell and a resistivity of $60 - 90 \mu\Omega - cm$.

The C54- $TiSi_2$ phase has an orthorhombic face-centered structure having 24 atoms per unit cell and a significantly lower resistivity ($12 - 20 \mu\Omega - cm$) than the C49- $TiSi_2$.

The basic formation process of titanium SALICIDE is as follows:

A thin titanium film with 20-60 nm thickness is deposited on an entire wafer with MOSFETs structure. The deposited Ti film reacts with the exposed silicon areas such as the source/drain area and polysilicon gate electrodes by the first anneal at 600-700°C in N_2 ambient. In first anneal, C49- $TiSi_2$ phase is formed. Then, the unreacted titanium film on the dielectric layer such as SiO_2 or SiN is selectively etched by APM (Ammonia and Hydrogen Peroxide Mixture) solution. The final step is second anneal at 800°C or above to transform high-resistivity C49- $TiSi_2$ phase to low-resistivity C54- $TiSi_2$ phase at the gate electrodes and source/drain areas.

³⁶A Study on Formation of High Resistivity Phases of Nickel Silicide at Small Area and its Solution for Scaled CMOS Devices, 07D53437, Ryuji Tomita

6.7.1 Oxide deposition

The thickness of this CVD deposited oxide layer will be the width of the spacer after having used highly anisotropic etching in the next few steps, for this reason the thickness of the oxide decides over the distance between the silicide and the gate oxide.

We make the oxide layer 50nm thick.

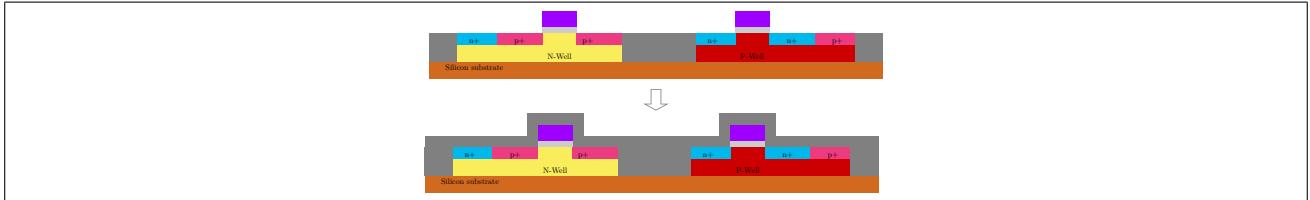


Figure 84: Oxide layer

We use the machine 310PC PECVD (CVD-P1) from HKUST³⁷ and deposit around 50nm of silicon dioxide.

6.7.2 Silicide block patterning (optional)

We now have to pattern the mask for the silicide block layer which will produce oxide wherever no silicide is not desired within active areas.

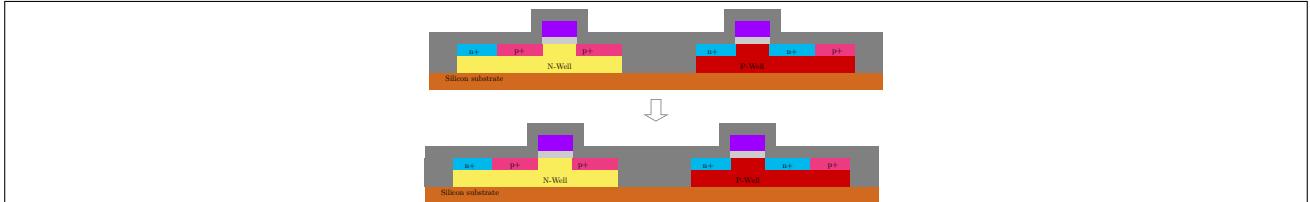


Figure 85: Patterning (silicide block)

6.7.3 Sputter etching(Spacers)

Now we have to etch our oxide as anisotropic as possible. This means that the etching mostly only comes "from above with a few to nearly none horizontal etching. That means the etching process only "sees" the sidewall as a "thicker layer" and starts etching downward. With an etching speed of 35 nm/min for thermal oxide and an oxide thickness of around 50nm and given that the polysilicon is much higher than 50nm we will have our desired spacer geometry forming as well as any potentially resist covered are (given silicide block is being used) with sharp etches.

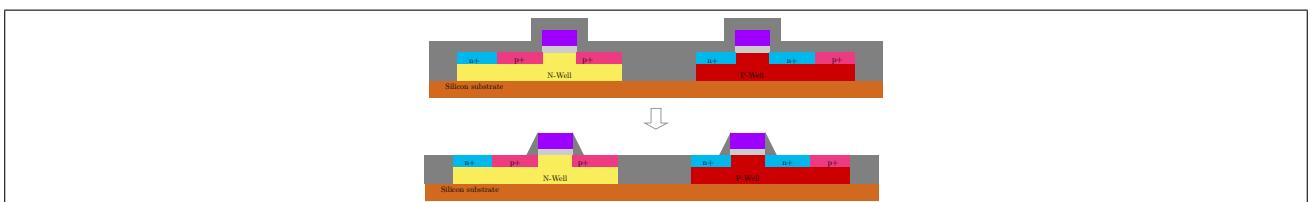


Figure 86: Anisotropic etching

The above mentioned machine is the "Oxford RIE Etcher (DRY-RIE-1)" at the NFF HKUST lab³⁸. The etching process runs on the oxide for 2 minutes.

³⁷<http://www.nff.ust.hk/en/equipment-and-process/equipment-list/thermal-diffusion-and-ion-implantation-module.html>

³⁸<http://www.nff.ust.hk/en/equipment-and-process/equipment-list/dry-etching-and-sputtering-module.html>

6.7.4 Titanium deposition

We deposit a layer of titanium with a thickness of around 20-60nm which will then be reacted into titanium-silicide and titanium-polycide respectively in the further steps.

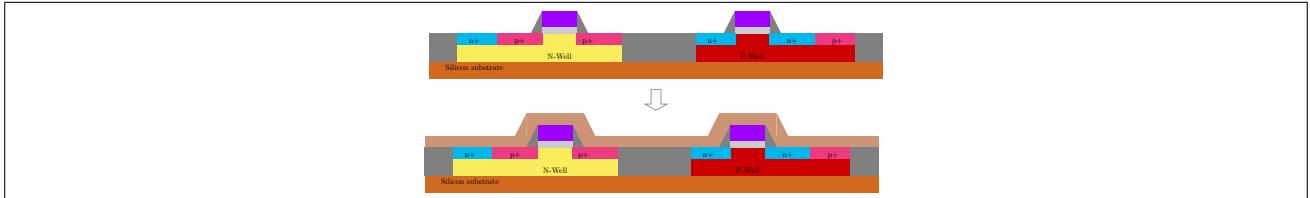


Figure 87: Titanium deposition

For this purpose we use the "Denton Sputter (SPT-Denton)" at HKUST NFF lab³⁹ which has a sputter rate of around 8.8 nm/min for titanium. This means we run the deposition process for around 5 minutes.

6.7.5 First reaction step

The deposited Ti film reacts with the exposed silicon areas such as the source/drain area and polysilicon gate electrodes by the first anneal at 600-700°C in N_2 ambient. In this first anneal, the C49- $TiSi_2$ phase is formed.

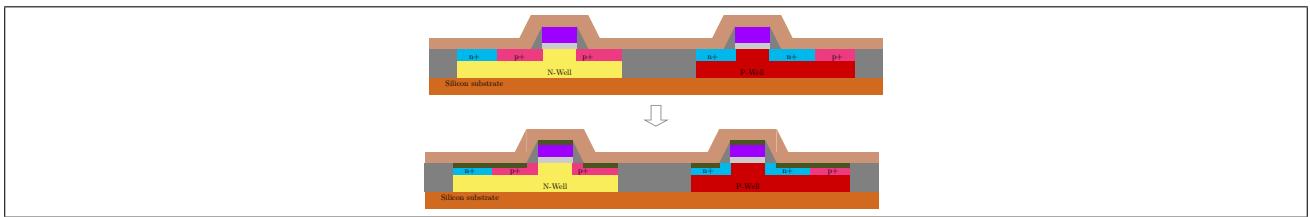


Figure 88: Reaction 1

We use the "AG610 RTP (DIF-R2)" from the HKUST⁴⁰ at 700°C for 240 seconds.

6.7.6 Etch

The unreacted titanium film on the dielectric layer such as SiO_2 or SiN is selectively etched by APM (Ammonia and Hydrogen Peroxide Mixture) solution.

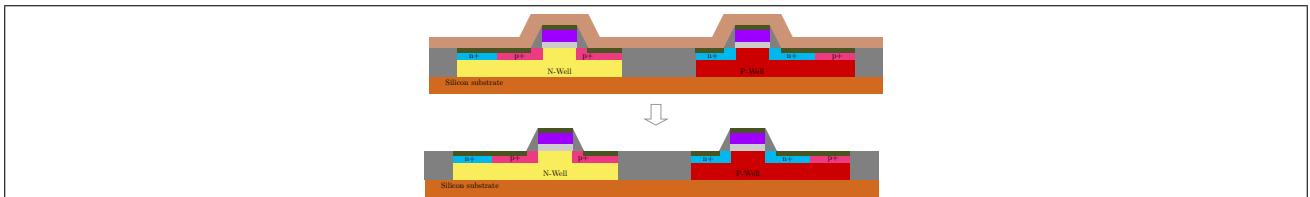


Figure 89: Titanium etch

6.7.7 Second reaction step

The final step is a second anneal at 800 °C or above to transform the high-resistivity C49- $TiSi_2$ phase to the low-resistivity C54- $TiSi_2$ phase at the gate electrodes and source/drain areas.

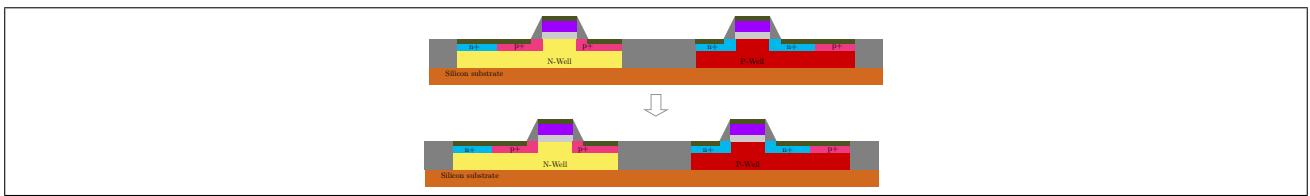


Figure 90: Reaction 2

We use the "AG610 RTP (DIF-R2)" again at 800°C for 240 seconds.

³⁹<http://www.nff.ust.hk/en/equipment-and-process/equipment-list/dry-etching-and-sputtering-module.html>

⁴⁰<http://www.nff.ust.hk/en/equipment-and-process/equipment-list/thermal-diffusion-and-ion-implantation-module.html>

6.8 First vias

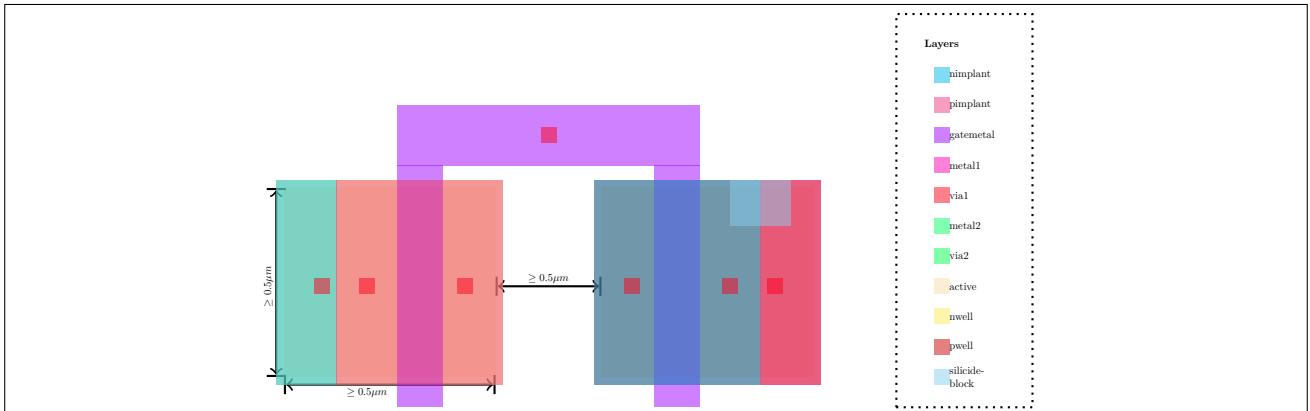


Figure 91: First via layout

6.9 First metal layer

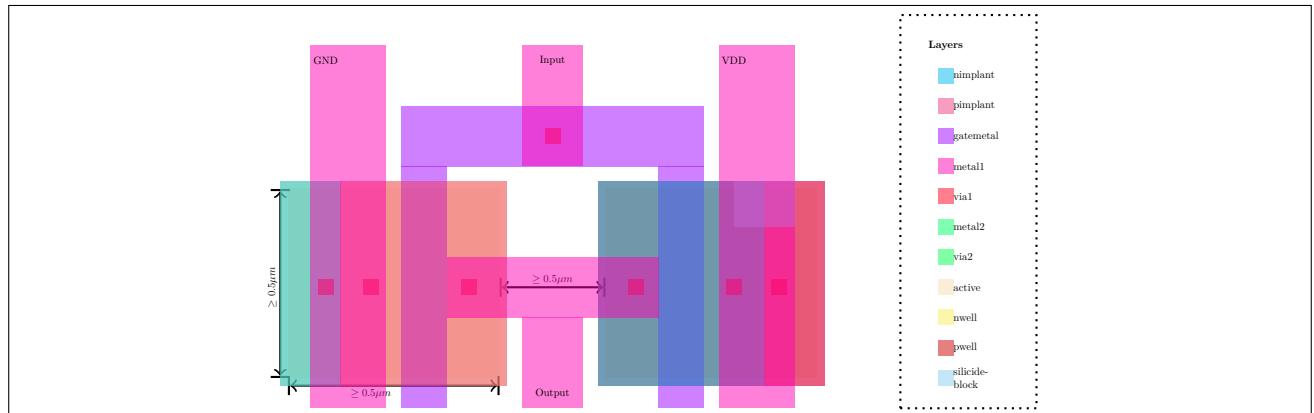


Figure 92: First metal layout

6.10 Additional vias

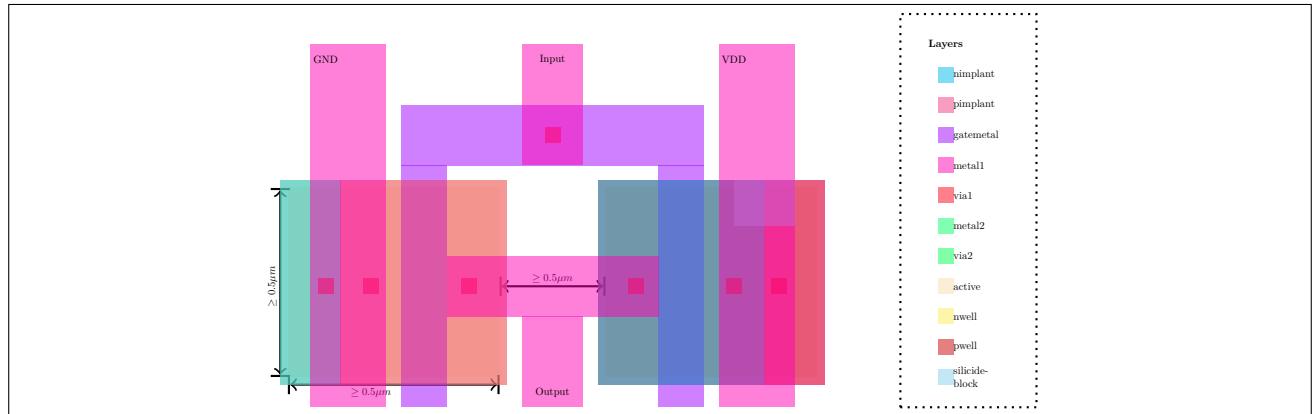


Figure 93: Additional via layout

6.11 Additional metal layer

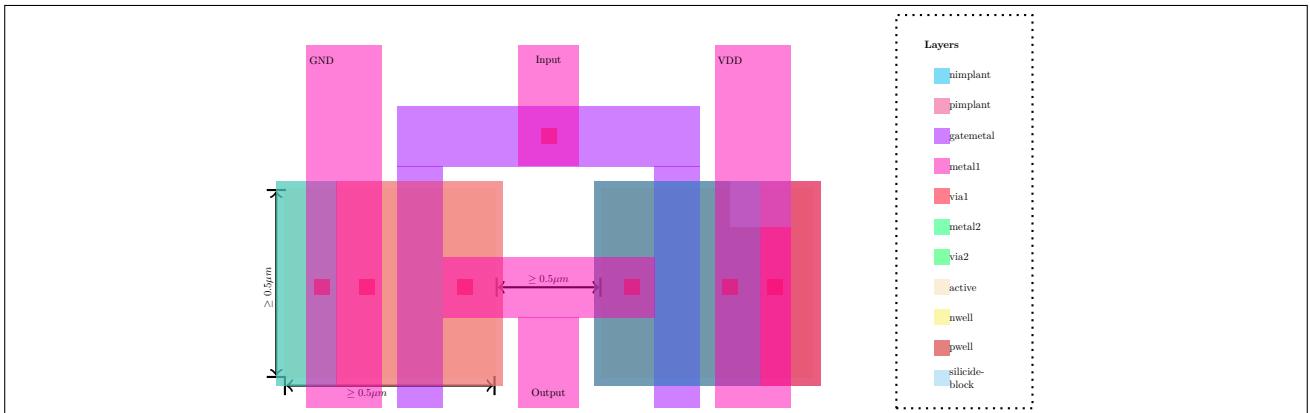


Figure 94: Additional metal layout

7 Testing

In order to get an idea on how strongly the actual product differs from the mathematical models being used before hand to define the initial parameters, one has to run a bunch of test wafers through the process over and over again, tweak parameters and measure out all kinds of aspects of the device.

This is also important for gathering the data which will finally end up within the data sheet.

In this chapter we will tackle the multiple different measurement and test objects we will need to put into the seal area during production in order to ensure the consistent quality of the chips we're going to sell in the end.

Testing parameters:

- Passive properties:
 - The capacity per m^2 from the n-well to the gate electrode
 - The capacity per m^2 from the p-well to the gate electrode
 - The actual resistance/m of the p-well
 - The actual resistance/m of the n-well
 - The actual resistance/m of the p-implant layer
 - The actual resistance/m of the n-implant layer
 - The actual resistance/m of the poly layer
 - The actual resistance/m of the silicide+p-implant layer
 - The actual resistance/m of the silicide+n-implant layer
 - The actual resistance/m of the silicide+poly layer
- Diodes:
 - ESD structure diodes (Voltage protection)
 - Lateral diodes
 - Vertical diodes
- Bipolar transistors:
 - Vertical bipolar transistor (to test latchup conditions)
 - Lateral bipolar transistors (to see what the beta is)
- NMOS as well as PMOS:
 - Frequency characteristics (Transient curve)
 - Drain-source resistance vs. gate-source voltage
 - Actual threshold voltage

Each of these values will variate based on the location on the wafer, because of diffraction towards the edge and also because of the uneven nature of the wafer.

Measuring the same test structure multiple times placed at multiple locations on the wafer will allow us to calculate an effective average value and tolerance ranges for taking into account for further dimensioning of circuits..

7.1 Diodes

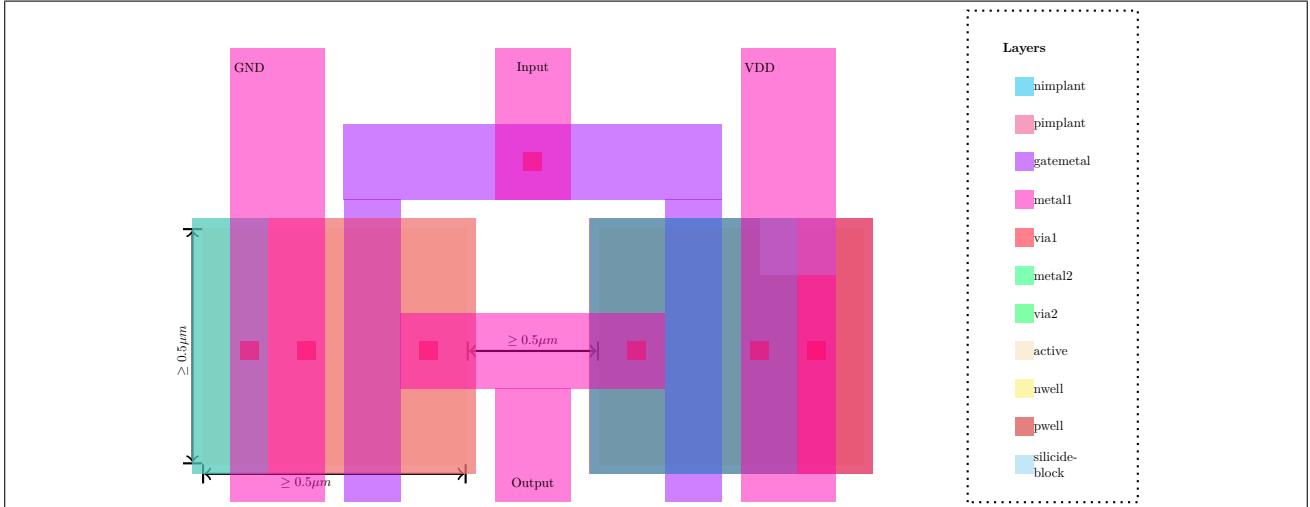
There are multiple different diode types possible which could form, which can categorized into two categories: Lateral and vertical.

Lateral diodes can form between two wells or the well and the junction, the vertical diodes can form between the n-well and the p-substrate.

7.2 Bipolar transistors

8 Design rules

In overall the lambda rules from MOSIS are sufficient for keeping it manufacturable but just for completion we accumulate all the edge parameters arising from using the HKUST equipment which we are using for this open process in this chapter.



8.1 Lithographic limitations

When looking at the mask making module "Intertech ISI-2808 Laser Direct Write System" from HKUST ([subsection 9.1](#)) we see the following limitations for the lithographic masks:

- The Minimum Feature Size of the Laser system is $1.5\mu m$. Any pattern size less than $1.5\mu m$ may not come out.
- The grid of the Laser system is $0.25\mu m$, any off grid pattern will be round off or up to the grid location

So the smallest structure **on the mask** shall be bigger than $1.5\mu m$

If we look at the stepper "ASML Stepper (PHT-S1)" from HKUST([subsection 9.2](#)) we see it has a resolution of $0.5\mu m$ and a reduction ratio of 5:1

This mean that our feature sizes **on the wafer** can be around $0.5\mu m$ (for sizes below $0.5\mu m$ it will require some trickery with multiple overlapping masks in the future)

9 Machines

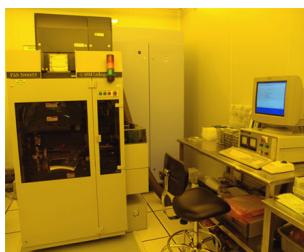
This chapter describes all the machines which are going to be used to verify our parameters in detail. The description of the machines can be found on the HKUST NFF website⁴¹, but all the info is being copied over for the benefit of future reproducibility, in case it is decided to by HKUST in the future to take the page down or restrict the access to the information.

9.1 Intertech ISI-2808 Laser Direct Write System



Laser Source	Helium-Cadmium blue laser (20 mW) 5000 hours life
X - Y Stage	Laser Interferometer Stage
Maximum Writing Area	150 mm x 150 mm (6" x 6")
Laser Spot Size	0.7 μm
Minimum Feature Size	1.5 μm
Overlay Accuracy	0.3 μm
Maximum Butting Error	0.3 μm
Line Uniformity	$\pm 0.1\mu\text{m}$
Wring Grid	0.1, 0.2, 0.25, 0.5 μm
Writing Speed	1.28 $\frac{\text{mm}^2}{\text{sec}}$ @ 0.5 μm grid, 0.64 $\frac{\text{mm}^2}{\text{sec}}$ @ 0.25 μm grid
Writing Method	Raster Scanning, width of swath 256 μm
Depth of Focus	1.8 μm

9.2 ASML Stepper (PHT-S1)



Light source illumination	i-line (365 nm)
Resolution	0.5 μm
Maximum writing area	$\pm 0.1\mu\text{m}(3\sigma)$
Wafer size	4" or 6"
Field size	15 mm x 15 mm or 10 mm x 10 mm (on wafer)
Reduction ratio	5:1
Photomask size	5" square

9.3 DRIE Etcher #1 (DRY-Si-1)



Gases available	C4F8, SF6, O2, N2, He & Ar
RF power source	1x 1000W(max) at 13.56MHz for Coil electrode, 1x 300W(max) at 13.56MHz for Platen electrode
Electrode coolant system	5 to 30 °C
High speed turbo molecular pump	pumping speed of 1000 L/s at 36000 rpm Fully automatic loadlock transfer system
Substrate size	4" wafer
Silicon etch	
Minimum Line/Space	0.5 μm
Low Rate Silicon Etch E/R	From 50nm/cycle
Normal Rate Silicon Etch E/R	Up to 2 μm /min
Selectivity to Photoresist	>50:1
Selectivity to Oxide	>80:1
Uniformity	7%

⁴¹<http://www.nff.ust.hk/en/equipment-and-process/equipment-list.html>

9.4 Buehler Polisher (CMP-4)



- Polished for
 - Silicon
 - Silicon Oxide
 - Silicon Nitride
- $> 5\text{mm}^2$ to 4" wafer size
- $100\text{-}800\mu\text{m}$ wafer thickness

9.5 Buehler Polisher (CMP-5)



- Polished for
 - Copper
 - Carbon nano tubes
 - Silicon
 - Silicon Oxide
 - Silicon Nitride
- $> 5\text{mm}^2$ to 4" wafer size
- $100\text{-}800\mu\text{m}$ wafer thickness

10 List of Footnotes

https://github.com/chipforge/StdCellLib	1
https://www.google.com/patents/US3356858	4
http://www.solecon.com/sra/rho2ccal.htm	6
https://en.wikipedia.org/wiki/Fick%27s_laws_of_diffusion	8
ISBN 3-8023-1588: Hoppe Bernhard, Mikroelektronik 2, Page 24, Table 2.1	8
If someone really wants to do this in his basement he can google these values and make a pull request	9
ISBN 3-8023-1588: Hoppe Bernhard, Mikroelektronik 2, Page 48, Table 3.2	10
http://cleanroom.byu.edu/rangestraggle	10
Gerold W. Neudeck and Robert F. Pierret, Modular series on solid state devices, Volume V, Chapter 4	11
https://en.wikipedia.org/wiki/Boltzmann_constant#Role_in_semiconductor_physics:_the_thermal_voltage	12
https://en.wikipedia.org/wiki/Volt	12
https://ecee.colorado.edu/~bart/book/eband5.htm	13
https://en.wikipedia.org/wiki/Electron_affinity	13
https://en.wikipedia.org/wiki/Boltzmann_constant#Role_in_semiconductor_physics:_the_thermal_voltage	14
https://en.wikipedia.org/wiki/Volt	15
https://ecee.colorado.edu/~bart/book/eband5.htm	16
https://en.wikipedia.org/wiki/Electron_affinity	16
Wolf, S.; R.N. Tauber (1986). Silicon Processing for the VLSI Era: Volume 1 - Process Technology. pp. 532–533. ISBN 978-0-9616721-3-3	18
A. Pliskin, J.Vac.Sci Technol., vol. 14, p.1064, 1977	18
http://www.freepatentsonline.com/4395438.html	19
https://www.allaboutcircuits.com/textbook/digital/chpt-3/logic-signal-voltage-levels	20
https://learn.sparkfun.com/tutorials/logic-levels/33-v-cmos-logic-levels	20
Hagen! Please explain this part here	20
https://www.google.com/patents/US7985656	22
N. Wang, Digital MOS Integrated Circuits, Prentice-Hall, Englewood Cliffs, NJ, 1989	23
A. Schenk and G. Heiser, "Modeling and Simulation of Tunneling through Ultra-Thin Gate Dielectrics" J.Appl.Phys., vol. 81, no. 12, pp. 7900, 1997	23
Y. Taur, "The Incredible Shrinking Transistor," IEEE Spectrum, pp. 25-29, July 1999.	23
S. Thompson, P. Packan, and M. Bohr, "MOS Scaling: Transistor Challenges for the 21st Century," Intel Technology Journal, vol. Q3, 1998	23
http://ecee.colorado.edu/~bart/book/book/chapter3/ch3_4.htm#3_4_2	23
https://ecee.colorado.edu/~bart/book/book/chapter2/ch2_7.htm#2_7_2	24
see simulation/diffusion_pwell.wxmlx	26
See simulation/gate.wxmlx	28
https://people.eecs.berkeley.edu/~hu/Chenming-Hu_ch5.pdf	28
https://patents.google.com/patent/US6376329	29
http://www.nff.ust.hk/en/equipment-and-process/equipment-list/photolithography-module.html	29
http://cleanroom.byu.edu/OxideTimeCalc	34
https://www.google.com/patents/US7985656	34
A Study on Formation of High Resistivity Phases of Nickel Silicide at Small Area and its Solution for Scaled CMOS Devices, 07D53437, Ryuji Tomita	51
http://www.nff.ust.hk/en/equipment-and-process/equipment-list/thermal-diffusion-and-ion-implantation-module.html	52
http://www.nff.ust.hk/en/equipment-and-process/equipment-list/dry-etching-and-sputtering-module.html	52
http://www.nff.ust.hk/en/equipment-and-process/equipment-list/dry-etching-and-sputtering-module.html	53
http://www.nff.ust.hk/en/equipment-and-process/equipment-list/thermal-diffusion-and-ion-implantation-module.html	53
http://www.nff.ust.hk/en/equipment-and-process/equipment-list.html	62