

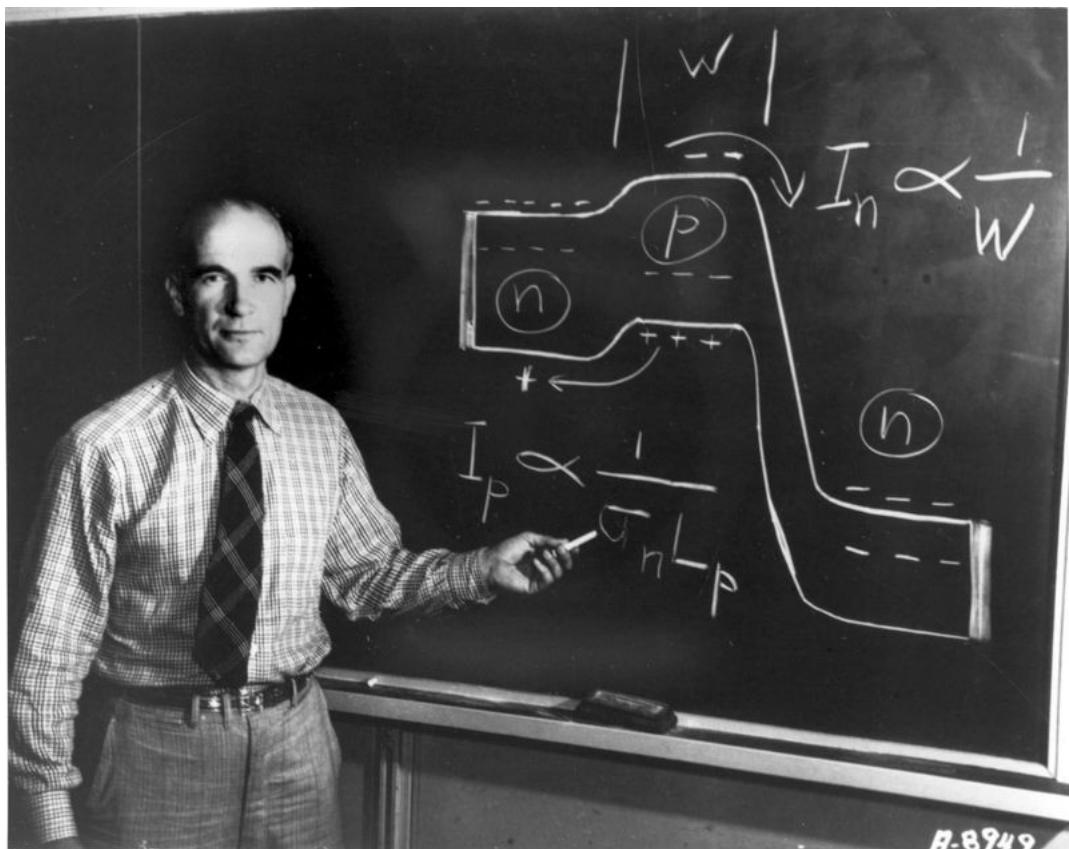


POLITECNICO
MILANO 1863

Electron Devices

Prof. Christian Monzio Compagnoni

Notes by Mattia Marinoni



Department of Electronics

Academic Year 2022-2023

Contents

I Lectures	3
1 Semiconductor generalities	4
1.1 Energy GAP	4
1.2 Density of states	5
1.3 Fermi-Dirac statistics	6
1.4 Intrinsic Si	8
1.5 Extrinsic Si	9
1.6 Current transport	11
1.7 Electrostatic potential and band bending	13
1.8 Spatially variable doping concentration	15
1.9 Non-equilibrium conditions	16
1.10 Continuity equation	17
1.11 Shockley-Read-Hall theory	19
1.12 Main dependences of R	20
2 Diode (p-n junction)	23
2.1 Basics of the p-n junction	23
2.2 Solution of the Poisson equation	25
2.3 Forward and reverse bias	27
2.4 Qualitative behavior of E_{F_n} and E_{F_p}	29
2.5 Minority carriers diffusion in quasi-neutral regions	31
2.6 Wide-base and narrow-base diodes	33
2.7 Spatial profiles for n, p, J_n, J_p	34
2.8 Temperature dependence of the $J - V$ curve	35
2.9 High/low-current regimes	36
2.10 Small signal model	39
3 M-S junction	42
3.1 Basics of the M-S junction	42
3.2 M-S junction as a Schottky diode	44
3.2.1 Current transport: Schottky's model	44
3.2.2 Current transport: Bethe's model	45
3.2.3 Current transport: Thermionic/diffusion model	47
3.2.4 Schottky effect	49
3.3 M-S junction as an ohmic contact	51
3.4 Interface states	54
4 MOS capacitor	57
4.1 Basics of the MOS capacitor	57
4.2 Gate bias	59
4.3 Reformulation of the Poisson equation	60
4.4 Calculation of Q_s as a function of V_s	62
4.5 $Q_s - V_s$ curve	62
4.6 Dependence of V_s and Q_s on V_G	64
4.7 Small signal model	66
4.8 $C - V$ curve under low and high frequencies	68
4.9 Deep-depletion condition	71
4.10 Basics of the nMOS capacitor with ring	72
4.11 Impact of the ring bias on the electrostatics	74
4.12 Dependence of V_s on V_G/V_R	75
4.13 Impact of fixed oxide charge and of interface states on the silicon/oxide interface	77
4.14 Polysilicon gate	79

5 MOS transistor	81
5.1 Basics of the MOS transistor	81
5.2 Electrostatics in the channel region	83
5.3 Subthreshold and Onstate regimes	83
5.4 Continuity equation in the channel	85
5.5 Ohmic/parabolic regime	86
5.6 Saturation regime	88
5.7 Band diagram along the channel in the onstate regime	91
5.8 Body effect	92
5.9 Small signal model	93
5.10 Subthreshold operations	98
5.11 Impact of different parameters on the transcharacteristics	101
5.12 Short-channel regime	104
5.13 Velocity saturation	106
5.14 Constant-field scaling	110
5.15 Generalized scaling	111
5.16 Dielectric imperfections	113
5.17 Design rules	117
6 Bipolar transistor (BJT)	120
6.1 Basics of the BJT	120
6.2 Calculation of I_C	122
6.3 Calculation of I_B	124
6.4 High/low-current regimes	125
6.5 $I_C - V_{CE}$ curve	127
6.6 Small signal model	129
II Exercise classes	132
III Simulation labs	147
IV Experimental labs	181

Part I

Lectures

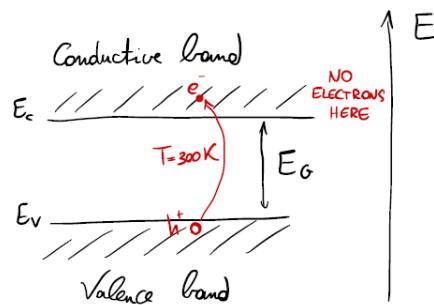
Chapter 1

Semiconductor generalities

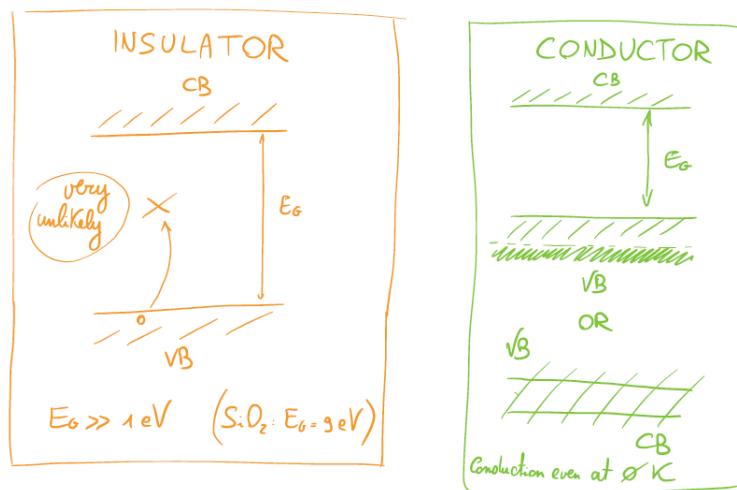
1.1 Energy GAP

For this course we are considering just semiconductors made of **Si**, an element located in the IV group of the periodic table and thus presenting the configuration $3s^2\ 3p^2$ on its most external orbital. We already know that electrons in periodic potentials present quantified levels of energy, so their conduction properties can be easily represented through the *band diagrams*. In particular, we want to focus our study in particular on the **Energy GAP**, which is the most important energy level in semiconductors as electrical and optical properties depend on that. The energy GAP can be described as the distance between the conductive band E_c , a band completely devoid of electrons, and the valence band E_v , a band filled entirely with electrons (both considerations are taken at $T = 300\ K$). To sum it up

$$E_c - E_v = E_G$$



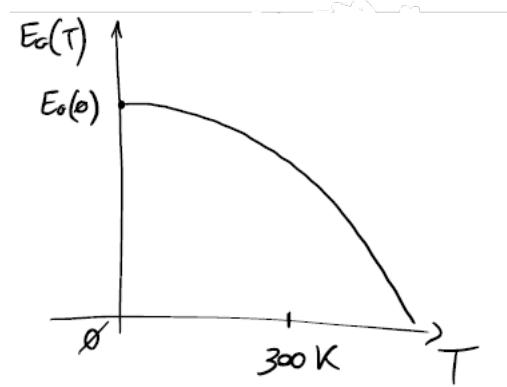
For semiconductors in general $E_G \approx 1\ eV$, while in other materials, like insulators or metals, it is either higher or lower, determining the conductivity of said material (close to none for insulators, very high for metals).



Semiconductors on the other hand have a particular behavior referring to conductivity, because their energy GAP depends

on the temperature according to this relation

$$\begin{cases} E_G(T) = E_G(0) - \frac{\alpha T^2}{\beta + T} \\ \left. \frac{\partial E_G}{\partial T} \right|_{T=300 K} = \frac{-2\alpha T(\beta + T) + \alpha T^2}{(\beta + T^2)^2} \Big|_{T=300 K} = -2,6 \cdot 10^{-4} \text{ eV/K} \end{cases} \quad (1.1)$$



As we can see, the GAP decreases when temperature rises, improving electronic jumps as they now require less energy to be performed. For Si in particular we have these values:

$$\begin{aligned} E_G(0) &= 1,169 \text{ eV} \\ \alpha &= 4,9 \cdot 10^{-4} \frac{\text{eV}}{\text{K}} \implies E_G(300 \text{ K}) = 1,12 \text{ eV} \\ \beta &= 655 \text{ K} \end{aligned} \quad (1.2)$$

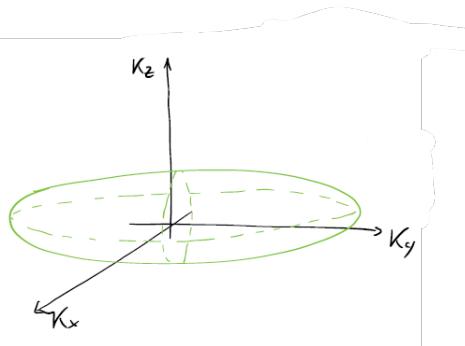
It is also important to remember that missing electrons can be considered as positive-charged carriers: the **holes**.

1.2 Density of states

We want to obtain a description of the energy dispersion relation approximated to the edges; first we are going to consider the conduction band, therefore

$$E - E_c = \frac{\hbar^2 k_x^2}{2m_x} + \frac{\hbar^2 k_y^2}{2m_y} + \frac{\hbar^2 k_z^2}{2m_z} \quad (1.3)$$

where E is a generic energy state.



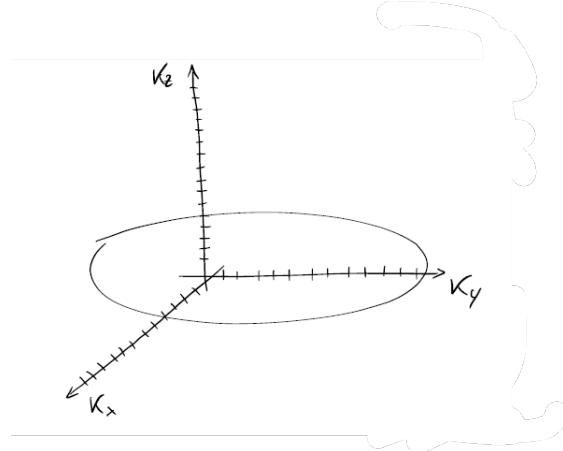
As we can see, this is the equation of an ellipsoid; if we were to find the axis intersections then we would simply reverse the ellipsoid equation and get

$$\bar{k}_x = \pm \sqrt{\frac{2m_x(E - E_c)}{\hbar^2}}; \quad \bar{k}_y = \pm \sqrt{\frac{2m_y(E - E_c)}{\hbar^2}}; \quad \bar{k}_z = \pm \sqrt{\frac{2m_z(E - E_c)}{\hbar^2}} \quad (1.4)$$

Effective mass for Si As we have already noticed, we can define different values for the mass according to the direction we are currently heading to.¹ For Si, looking at the ellipsoid function, we can distinguish a transverse mass from a longitudinal mass:

$$\begin{aligned} m_x = m_z &= 0,19m_0 = m_t \\ m_y &= 0,98m_0 = m_l \end{aligned} \quad (1.5)$$

Now let's consider the discretization of the k -space.



We can define a volume

$$\frac{4}{3}\pi\bar{k}_x\bar{k}_y\bar{k}_z = \frac{4}{3}\pi\frac{\sqrt{8m_xm_ym_z}}{\hbar^3}(E - E_c)^{3/2}$$

and a number of points

$$N = \frac{V}{\left(\frac{2\pi}{L}\right)^3} \frac{1}{L^3} = \frac{4}{3}\pi\frac{\sqrt{8m_xm_ym_z}}{h^3}$$

We then want to see how everything changes according to a small energy variation

$$\frac{dN}{dE} = 2\pi\frac{\sqrt{8m_xm_ym_z}}{h^3}\sqrt{E - E_c}$$

Eventually we obtain the density of states for the conductive band as

$$g_c(E) = \frac{dN}{dE} 2g_{\text{DEG}} \quad \xrightarrow{Si} \quad g_c(E) = \frac{48\pi}{h^3}\sqrt{2m_tm_l}\sqrt{E - E_c}$$

Same procedure can be easily repeated for the density of states in the valence band, thus

$$g_v(E) = \frac{8\pi}{h^3} \left(\sqrt{2m_{hh}^3} + \sqrt{2m_{lh}^3} \right) \sqrt{E_v - E}$$

where m_{hh} and m_{lh} are heavy hole and light hole masses.

1.3 Fermi-Dirac statistics

Under thermodynamic equilibrium it is easy to compute band occupation as we are in a perfectly balanced situation (no net processes are involved). Statistically, that is described through the Fermi-Dirac distribution

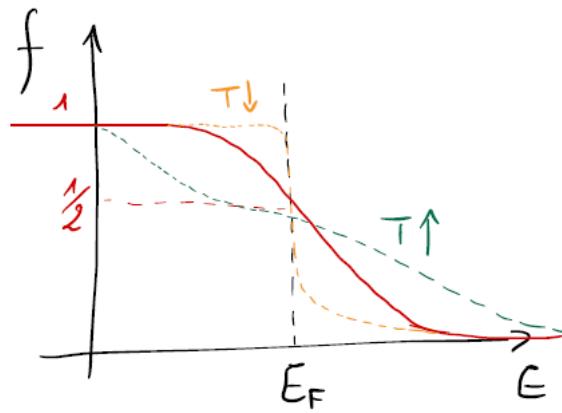
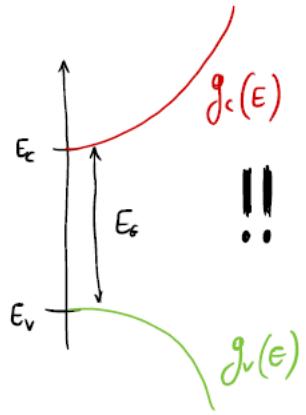
$$f(E) = \frac{1}{1 + \exp\left(\frac{E - E_F}{KT}\right)} \quad (1.6)$$

defined as the probability that the generic state E is occupied by electrons; E_F is the Fermi level, which represents the last filled level.

As we can see, KT determines the transition smoothness, in particular if $E - E_F \gg KT$ we can use the Maxwell-Boltzmann approximation

$$f(E) \approx \exp\left(-\frac{E - E_F}{KT}\right) \quad (1.7)$$

¹Remember that the effective mass in Quantum Physics is related to the energy levels like $m^* = \frac{\hbar^2}{\frac{\partial^2 E}{\partial k^2}}$



Same reasoning can be repeated for the holes, this time considering the non-occupations of the energy levels though, so

$$1 - f(E) = \frac{1}{1 + \exp\left(-\frac{E - E_F}{KT}\right)} \quad (1.8)$$

Now we have all the ingredients to find the carriers concentrations, and we do it like that

$$n = \int_{E_c}^{+\infty} g_c(E)f(E)dE \quad (1.9)$$



Electron concentration for Si Substituting the values we found before for the density of states we get

$$n = \int_{E_c}^{+\infty} \frac{48\pi}{h^3} \sqrt{2m_t^2 m_l} \sqrt{E - E_c} \frac{1}{1 + \exp\left(\frac{E - E_F}{KT}\right)} dE \quad (1.10)$$

with some simplifications, such as $\frac{E - E_c}{KT} = x$ and $\frac{E_F - E_c}{KT} = \eta$ resulting in $dE = KTdx$ and $x - \eta = \frac{E - E_F}{KT}$, we get

$$n = \frac{48\pi}{h^3} \sqrt{2m_t^2 m_l} \int_0^{+\infty} \frac{\sqrt{x} \sqrt{KT}}{1 + e^{x-\eta} KT} dx$$

and rearranging further the equation

$$n = \frac{48\pi}{h^3} \sqrt{2m_t m_l} (KT)^{3/2} \frac{2}{\sqrt{\pi}} \int_0^{+\infty} \frac{\sqrt{x}}{1 + e^{x-\eta}} = N_c F_{1/2}(\eta) \quad (1.11)$$

where N_c is the effective density of states and $F_{1/2}(\eta)$ is the Fermi-Dirac integral, which summarizes the information about E_F position with respect to the conduction band. If $x - \eta \gg 1$, then

$$F_{1/2}(\eta) \approx \exp\left(-\frac{E_c - E_F}{KT}\right) \quad (1.12)$$

for Maxwell-Boltzmann². Therefore

$$n = N_c e^{-\frac{E_c - E_F}{KT}} \quad (1.13)$$

For the valence band, same procedure is repeated

$$p = \int_{-\infty}^{E_v} g_v(E)(1 - f(E))dE = N_v F_{1/2}(\eta) \quad (1.14)$$

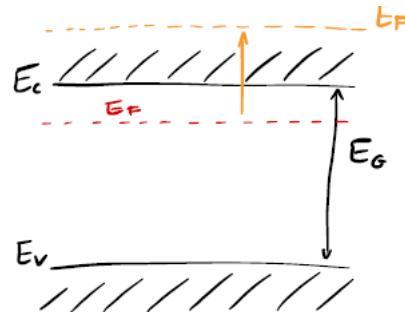
so

$$p = N_v e^{-\frac{E_v - E_F}{KT}} \quad (1.15)$$

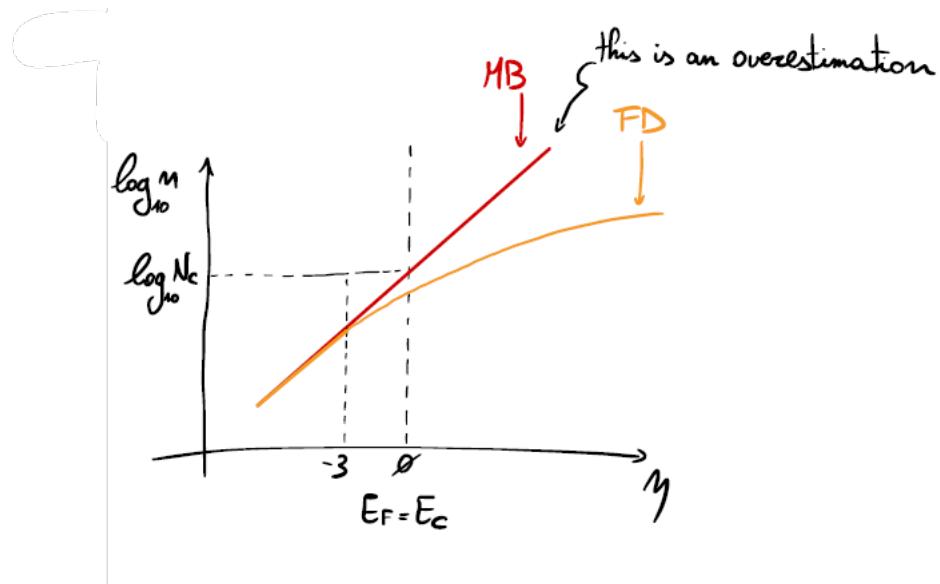
Now we are left to see what is the position of E_F in the band diagram.

1.4 Intrinsic Si

Let's see what happens when $E_c - E_F \ll KT$ or E_F is even higher than E_c :



We would get that, using Maxwell-Boltzmann approximation, the occupation probability is well over 1, which doesn't make any sense. That's because, in a logarithmic scale, Maxwell-Boltzmann statistics is a straight line, while the Fermi-Dirac one has a more like parabolic behavior



²Only valid if $E_c - E_F \gg 3KT$

Now we want to find exactly where E_F is on the band diagram. First we have to make a very important assumption: **the semiconductor is intrinsic**, meaning that every Si atom occupies the position it is expected to have in the lattice, so

$$p = n \quad (1.16)$$

Everything is perfectly balanced; we can develop 1.16 even further:

$$\begin{aligned} N_c e^{-\frac{E_c - E_F}{KT}} &= N_v e^{\frac{E_v - E_F}{KT}} \\ \frac{N_c}{N_v} = e^{\frac{E_c - E_F}{KT}} e^{\frac{E_v - E_F}{KT}} &= e^{\frac{E_c + E_v - 2E_F}{KT}} \end{aligned}$$

thus obtaining

$$E_F = \frac{E_c + E_v}{2} - KT \log \left(\frac{N_c}{N_v} \right) = E_i \quad (1.17)$$

where E_i is known as the intrinsic Fermi level; doing the calculations would result more or less in $E_F \approx \frac{E_G}{2}$. We can also mix 1.16 with 1.17:

$$\begin{aligned} n = N_c e^{-\frac{E_c - E_i}{KT}} &= N_c e^{-\frac{E_c - E_c/2 - E_v/2 + KT/2 \log(N_c/N_v)}{KT}} = N_c e^{-\frac{E_c - E_v}{2KT}} e^{-\frac{1}{2} \log \left(\frac{N_c}{N_v} \right)} = \\ &= N_c e^{-\frac{E_G}{2KT}} e^{\log \left(\frac{N_c}{N_v} \right)^{1/2}} = \sqrt{N_c N_v} e^{-\frac{E_G}{2KT}} = p = n_i \end{aligned} \quad (1.18)$$

where n_i is called *intrinsic concentration*³ (highlighted terms are strongly dependent on temperature). We can rearrange 1.18 to get

$$n = N_c e^{-\frac{E_c - E_F - E_i + E_i}{KT}} = N_c e^{-\frac{E_c - E_i}{KT}} e^{\frac{E_F - E_i}{KT}}$$

which is resulting in

$$n = n_i e^{\frac{E_F - E_i}{KT}} \quad (1.19)$$

$$p = n_i e^{\frac{E_i - E_F}{KT}} \quad (1.20)$$

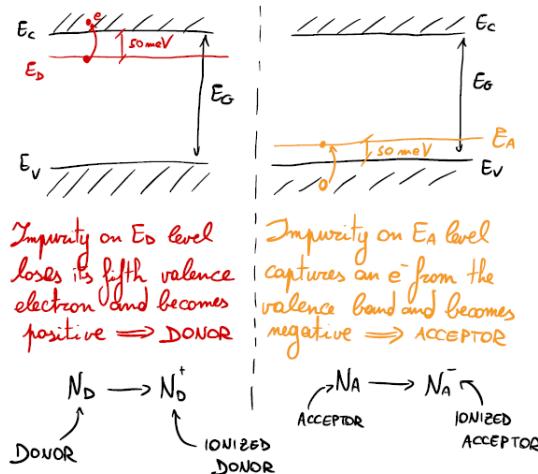
That confirms what we have been doing since in intrinsic semiconductors $E_i = E_F$ so $n = p = n_i$. We can extend this concept by introducing the **law of mass action**

$$n_i^2 = pn \quad (1.21)$$

which is going to come in handy in the next sections.

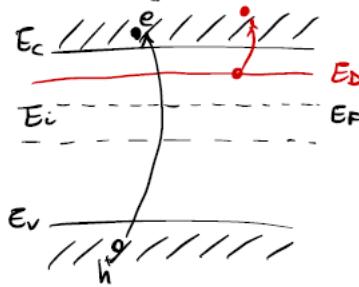
1.5 Extrinsic Si

Intrinsic Si is a perfect crystal, but it can be doped with either **donors** (V group) or **acceptors** (III group) in order to increase respectively n and p concentrations.



³For Si, $n_i(300K) = 1,45 \cdot 10^{-10} \text{ cm}^{-3}$

The question we are asked is always the same: where is E_F ? Beginning from a n -doped semiconductor, it is reasonable to assume that $n \neq p$ no more, but it would have been increased by $n = p + N_D^+$ due to the introduced donor level.



As usual, Fermi-Dirac statistics can be exploited

$$N_c F_{1/2}(\eta) = N_c e^{-\frac{E_v - E_F}{KT}} + N_D \frac{1}{1 + 2 \exp\left(-\frac{E_D - E_F}{KT}\right)} \quad (1.22)$$

where the term highlighted in red is the probability $1 - f(E)$ that E_D is **not** occupied by electrons and the factor 2 is introduce to take into account spin degeneracy. If we assume that $E_D - E_F \gg KT$, then 1.22 can be greatly simplified as E_F would be so below E_D that this would be completely devoid of carriers (**complete ionization**), thus $1 - f(E) \approx 1$ and Maxwell-Boltzmann would be applicable. On the overall, we obtain

$$N_c e^{-\frac{E_c - E_F}{KT}} = \mathfrak{N} + N_D \implies E_c - E_F = KT \log\left(\frac{N_c}{N_D}\right) \quad (1.23)$$

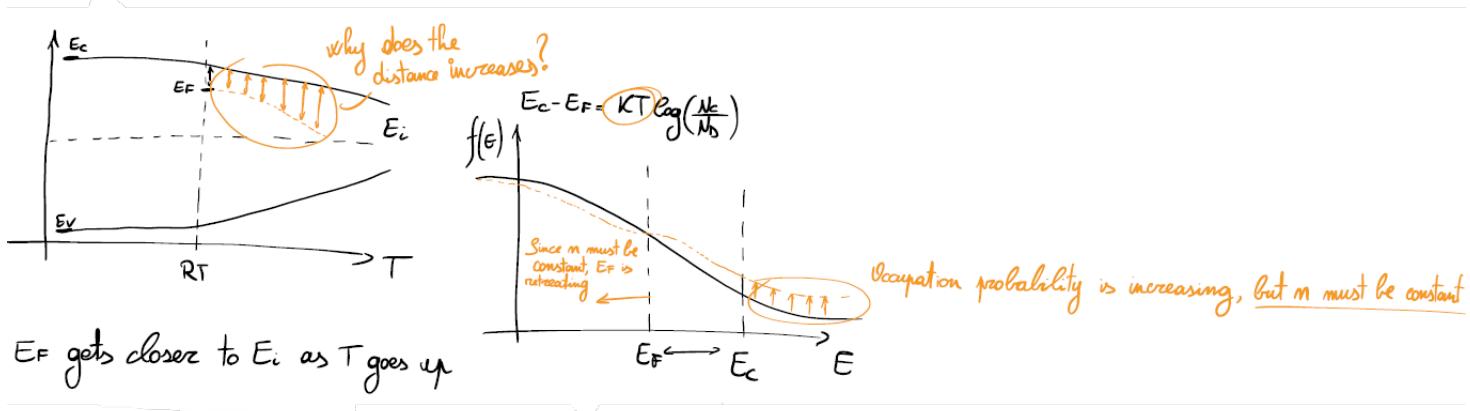
Note that p at $T = 300K$ is negligible as $p \ll n_i$. Eventually we can write

$$n \approx N_D \quad (1.24)$$

$$p \approx \frac{n_i^2}{N_D} \quad (1.25)$$

E_F goes upwards when N_D increases; if N_D is very high, a degeneracy occurs as $E_F \approx E_D$ and Maxwell-Boltzmann is no longer valid. In this case, we have a n^+ -doped semiconductor.

Temperature dependence Empirically, it is observable that whenever T rises, $E_c - E_F$ actually increases; that's because it is true that the occupation probability is getting higher, but n is forced to be constant, therefore E_F must recede until it comes very close to E_i . **Hole concentration must be revisited as well**. In fact p is no longer negligible, but becomes even dominant



over N_D : we enter *intrinsic regime*⁴.

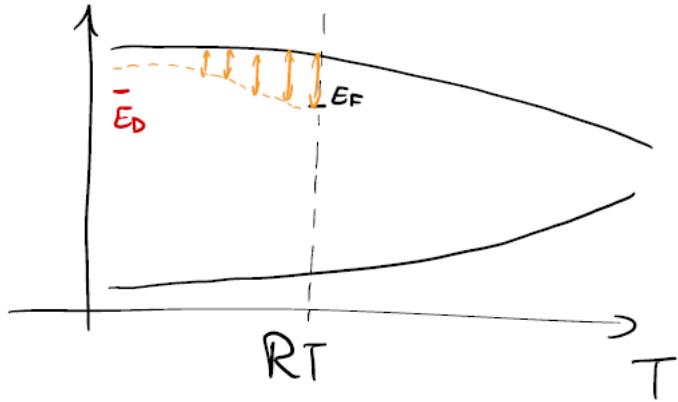
$$n = p + N_D^+ \approx N_D^+ < N_D \quad (1.26)$$

On the contrary, whenever T falls, $E_c - E_F$ keeps shrinking. At a certain point, E_F becomes so close to E_D that we are not in complete ionization anymore, meaning that not every donor is ionized

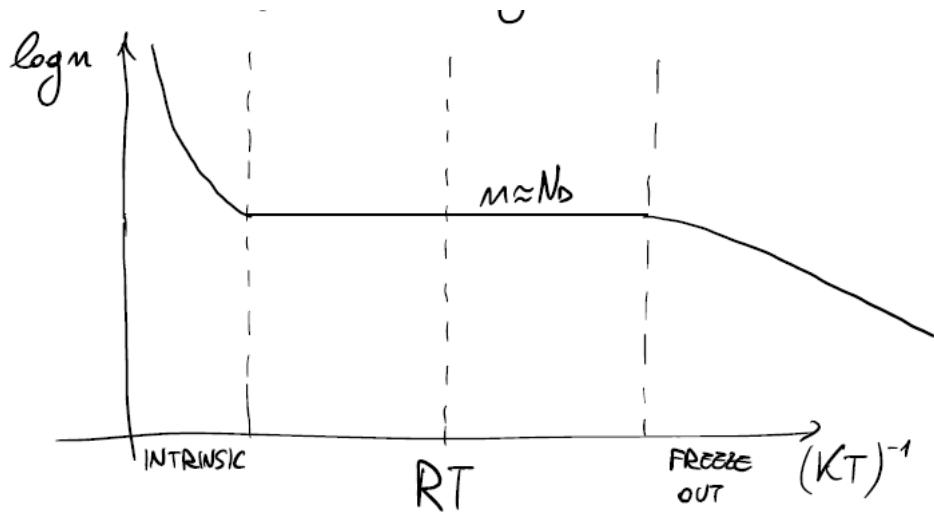
$$n = \mathfrak{N} + N_D^+ \approx N_D^+ < N_D \quad (1.27)$$

⁴Generally, intrinsic regime begins around $T = 600K$

We enter *freeze-out regime*⁵.



All those informations are summarized in the Arrhenius plot

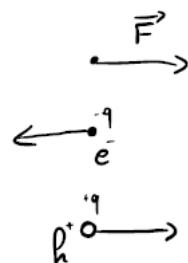


Note that for a *p*-doped semiconductor it is the same

$$\begin{cases} p \approx N_A \\ n \approx \frac{n_i^2}{N_A} \\ E_F - E_v = KT \log \left(\frac{N_v}{N_A} \right) \end{cases} \quad (1.28)$$

1.6 Current transport

It is important to point out that there is not a single transport mechanism, but many different ones; in this case, we are interested just in *drift* and *diffusion*. We begin from the first. If we apply an electric field \vec{F} to our semiconductor, we will notice that the carriers are going to be surely affected: in particular, holes will move in the same verse as \vec{F} , electrons in the opposite.



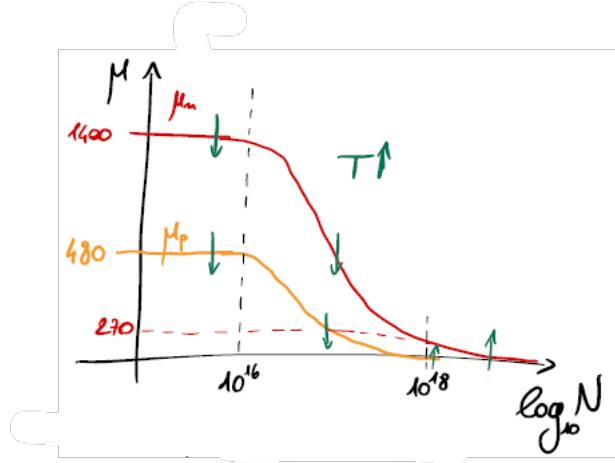
⁵Generally, freeze-out regime begins at around $T = 150K$

Velocity and electric field are linked together through the mobility μ

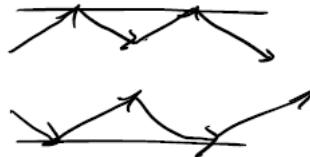
$$\vec{v}_{d_n} = -\mu_n \vec{F} \quad (1.29)$$

$$\vec{v}_{d_p} = +\mu_p \vec{F} \quad (1.30)$$

If carriers were free, acceleration would be constant, but since it is not the case they have to face **scattering**, intended as the collision among carriers and impurities; we can observe that mobility drops when the doping concentration increases, while saturates due to phononic scattering when it decreases.



When temperature rises, phononic scattering gets stronger while impurity scattering is slightly less effective. Mobility is also dependent on dimensionality; for example, in a 2D semiconductor there are more scattering centers because its surfaces are closer to one another.



After mobility, we can introduce current density, defined as

$$\vec{J}_n = -qn\vec{v}_{d_n} = +qn\mu_n \vec{F} \quad (1.31)$$

$$\vec{J}_p = +qp\vec{v}_{d_p} = +qp\mu_p \vec{F} \quad (1.32)$$

As we can see, both vectors are equiangular to \vec{F} , so the total current density would be

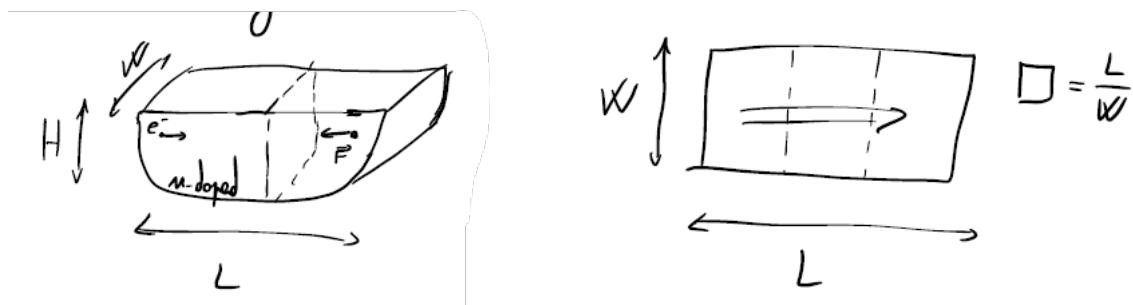
$$\vec{J} = \vec{J}_n + \vec{J}_p = (qn\mu_n + qp\mu_p)\vec{F} = \sigma\vec{F} \quad (1.33)$$

where σ is the conductivity and is related to the resistivity through

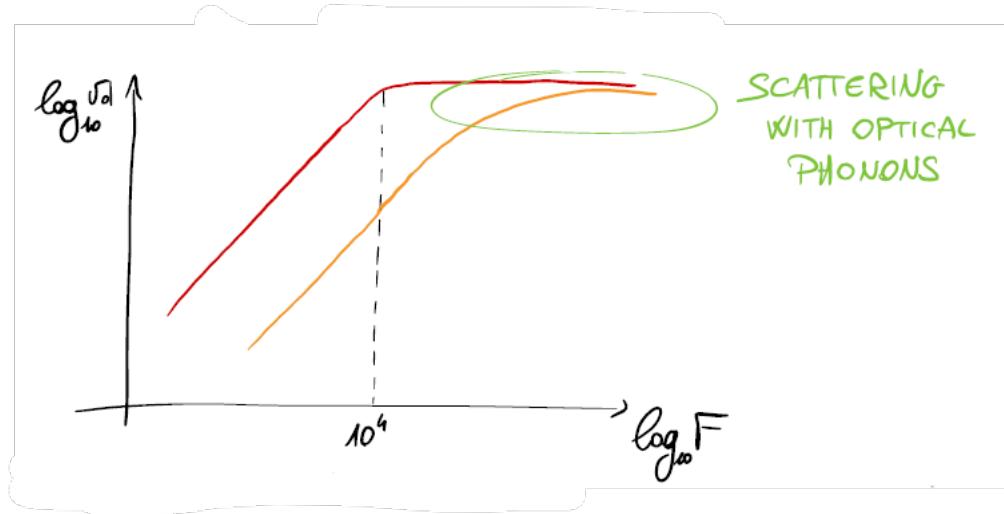
$$\rho = \frac{1}{\sigma} \quad (1.34)$$

For design reasons, sheet resistance is also important, described as the ratio between resistivity and semiconductor thickness

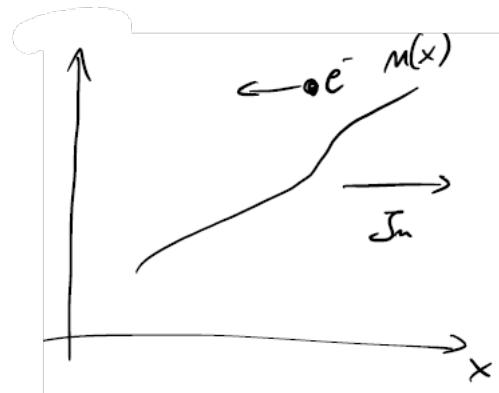
$$R = \rho \frac{L}{HW} = \rho_{sh} \frac{L}{W} \quad (1.35)$$



Finally, if F gets stronger, velocity saturates due to the scattering with optical phonons, which are highly energetic.



Now let's talk about the diffusion process; there aren't external forces moving carriers in this case, thus they simply adjust their position in the semiconductor according to their concentrations, going from high to low.



Current density is then

$$J_n = +qD_n \frac{dn}{dx} \quad (1.36)$$

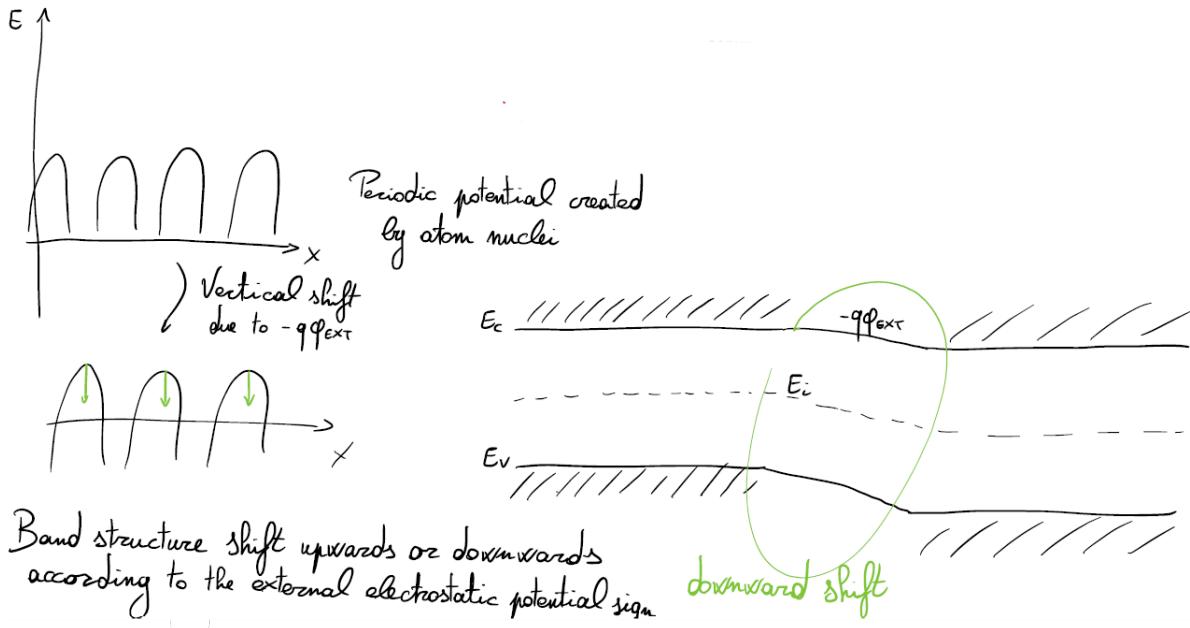
$$J_p = -qD_p \frac{dp}{dx} \quad (1.37)$$

where D_n and D_p are the diffusion coefficients, defined through Einstein's relations

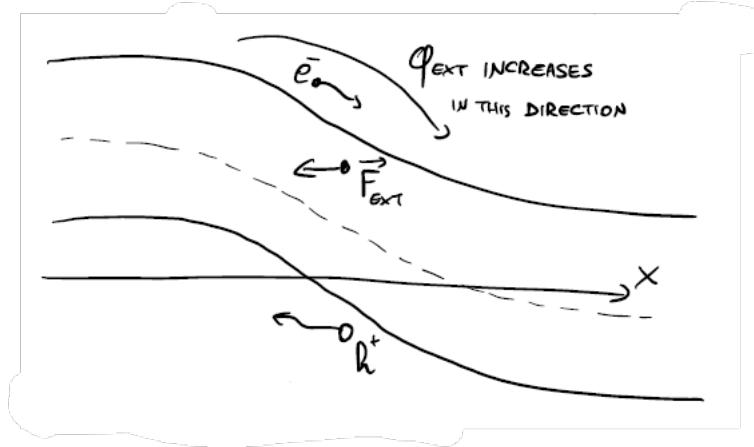
$$\begin{cases} D_n &= \mu_n \frac{KT}{q} \\ D_p &= \mu_p \frac{KT}{q} \end{cases} \quad (1.38)$$

1.7 Electrostatic potential and band bending

Until now we have been considering semiconductors under thermodynamic equilibrium, but what if an external electrostatic potential was applied?



The semiconductor periodic potential is shifted, and so are its energy bands.



Bands bend downwards in the direction where ϕ_{ext} rises; electrons follow the same direction. It is useful to underline this relation

$$\phi_{ext}(x) = -\frac{E_i}{q} \quad (1.39)$$

$$F_{ext}(x) = -\frac{d\phi_{ext}}{dx} \quad (1.40)$$

For Quantum Physics laws we can say that

$$-qF_{ext} = \frac{d(\hbar k)}{dt} \quad (1.41)$$

Overall analysis can be further optimized by introducing the effective mass approximation, so that we can actually remove the *ext* pedices from the equations, leaving just

$$\hbar k = m_x v_x \implies -qF = m_x \frac{dv_x}{dt} \quad (1.42)$$

Now if we were to combine 1.40 with Gauss law

$$\frac{dF}{dx} = \frac{\tilde{\rho}}{\epsilon_{si}} \quad (1.43)$$

we would obtain

$$\frac{d^2\phi}{dx^2} = -\frac{\tilde{\rho}}{\epsilon_{si}} \quad (1.44)$$

which is commonly known as the Poisson equation. For a semiconductor, 1.44 can be generalized into

$$\frac{d^2\phi}{dx^2} = -\frac{q}{\epsilon_{si}}(p - n + N_D^+ - N_A^-) \quad (1.45)$$

To have a better idea about carriers concentrations dependence on ϕ , we can rewrite 1.19 and 1.20 as

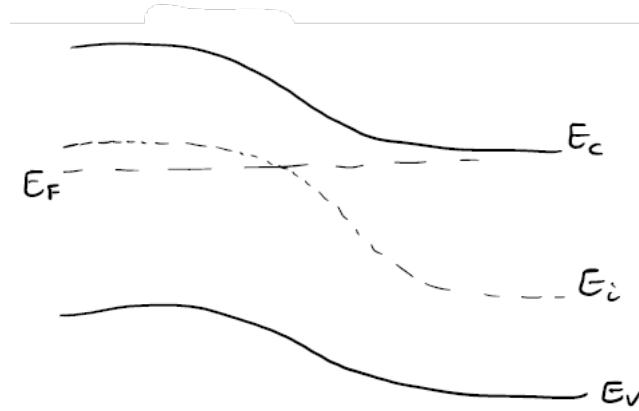
$$n = n_i e^{-\frac{q(\phi - \phi_F)}{KT}} \quad (1.46)$$

$$p = n_i e^{-\frac{q(\phi_F - \phi)}{KT}} \quad (1.47)$$

where $\phi_F = -\frac{E_F}{q}$ is the Fermi potential. We have just found out that n and p are strongly dependent on ϕ .

1.8 Spatially variable doping concentration

In the last section we found out what's the effect of an external voltage applied to the semiconductor, but how can we achieve such a result?

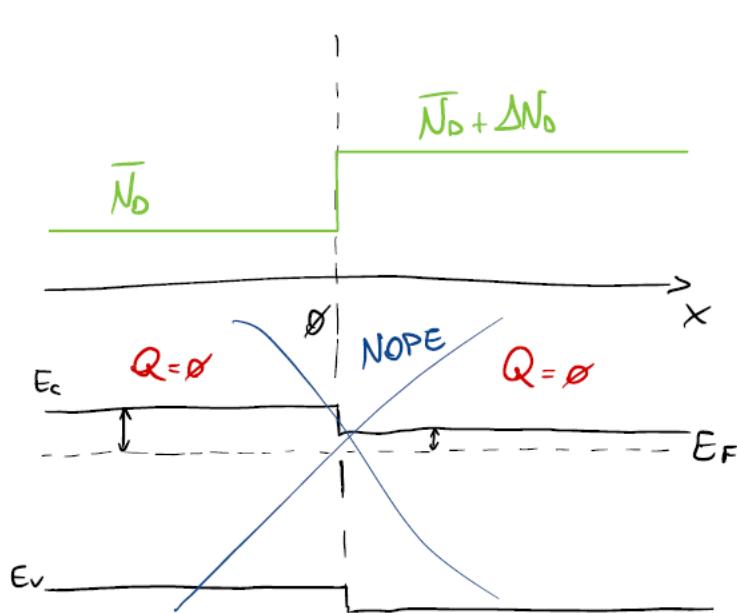


Answer is very simple: we dope the semiconductor, **but only in a certain region** so that

$$n(x) \approx N_D(x) \quad (1.48)$$

We can, for example, have a doping concentration looking like a Heaviside function (step)

$$N_D(x) = \bar{N}_D + \Delta N_D H(x) \quad (1.49)$$



If 1.48 is true, even the band bending should look like a step, but it actually doesn't since at $x = 0$, $F \rightarrow +\infty$ while $F = 0$ everywhere else. **We must rely on the Poisson equation** (in a condition where p is negligible and there are no acceptors and $\Delta N_D \ll \bar{N}_D$):

$$\begin{cases} \frac{d^2\phi(x)}{dx^2} = -\frac{q}{\epsilon_{\text{Si}}}(N_D - n) \\ \phi(x) = \bar{\phi} + \Delta\phi(x) \end{cases} \quad (1.50)$$

Adjusting 1.50 with 1.46 we have

$$\begin{cases} n = n_i e \frac{q(\bar{\phi} - \phi_F)}{KT} = \bar{N}_D \\ n = n_i e \frac{q(\bar{\phi} + \Delta\phi - \phi_F)}{KT} = n_i e \frac{q(\bar{\phi} - \phi_F)}{KT} e \frac{q\Delta\phi}{KT} = \bar{N}_D e \frac{q\Delta\phi}{KT} \approx \bar{N}_D \left(1 + \frac{q\Delta\phi}{KT}\right) \end{cases} \quad (1.51)$$

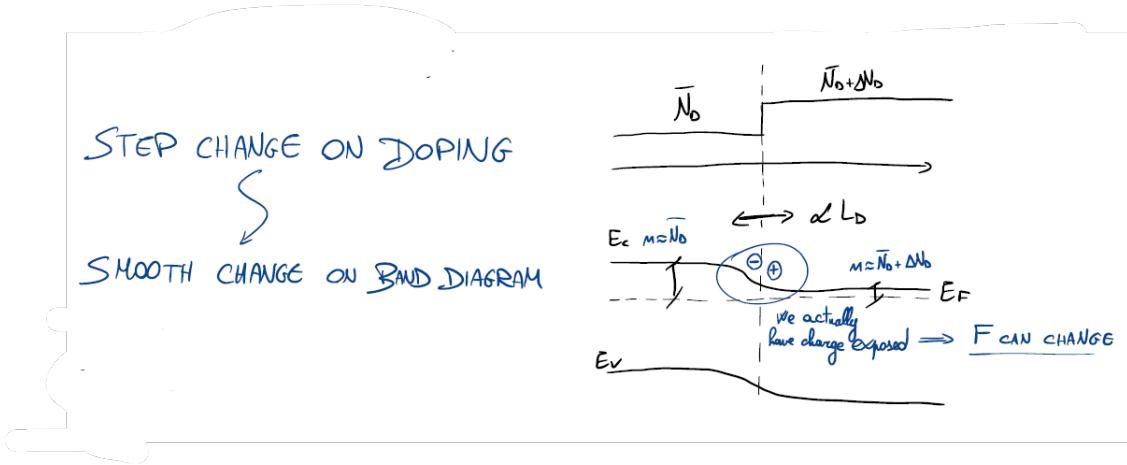
thus

$$\frac{d^2\Delta\phi}{dx^2} = -\frac{q}{\epsilon_{\text{Si}}} \left(-\bar{N}_D - \frac{q\bar{N}_D}{KT} \Delta\phi + \bar{N}_D + \Delta N_D H(x) \right) = \frac{q^2 \bar{N}_D \Delta\phi}{\epsilon_{\text{Si}} K T} - \frac{q \Delta N_D}{\epsilon_{\text{Si}}} H(x)$$

At this point, solution is trivial:

$$\Delta\phi(x) \propto e^{\pm \frac{x}{L_D}}$$

where $L_D = \sqrt{\frac{\epsilon_{\text{Si}} K T}{q^2 \bar{N}_D}}$ is the Debye length⁶. **Band profile changes exponentially.**



1.9 Non-equilibrium conditions

Let's rewrite the total current density, including both drift and diffusion contributions, for each carrier in a more compact way, beginning from J_n :

$$\begin{aligned} J_n &= qn\mu_n F + qD_n \frac{dn}{dx} = -qn\mu_n \frac{d\phi}{dx} + q\mu_n \frac{KT}{q} \frac{d}{dx} \left(n_i e \frac{q(\phi - \phi_F)}{KT} \right) = \\ &= -qn\mu_n \frac{d\phi}{dx} + q\mu_n \frac{KT}{q} n_i e \frac{q(\phi - \phi_F)}{KT} \frac{q}{KT} \left(\frac{d\phi}{dx} - \frac{d\phi_F}{dx} \right) = -qn\mu_n \frac{d\phi}{dx} + qn\mu_n \left(\frac{d\phi}{dx} - \frac{d\phi_F}{dx} \right) = \\ &= -qn\mu_n \frac{d\phi_F}{dx} \end{aligned} \quad (1.52)$$

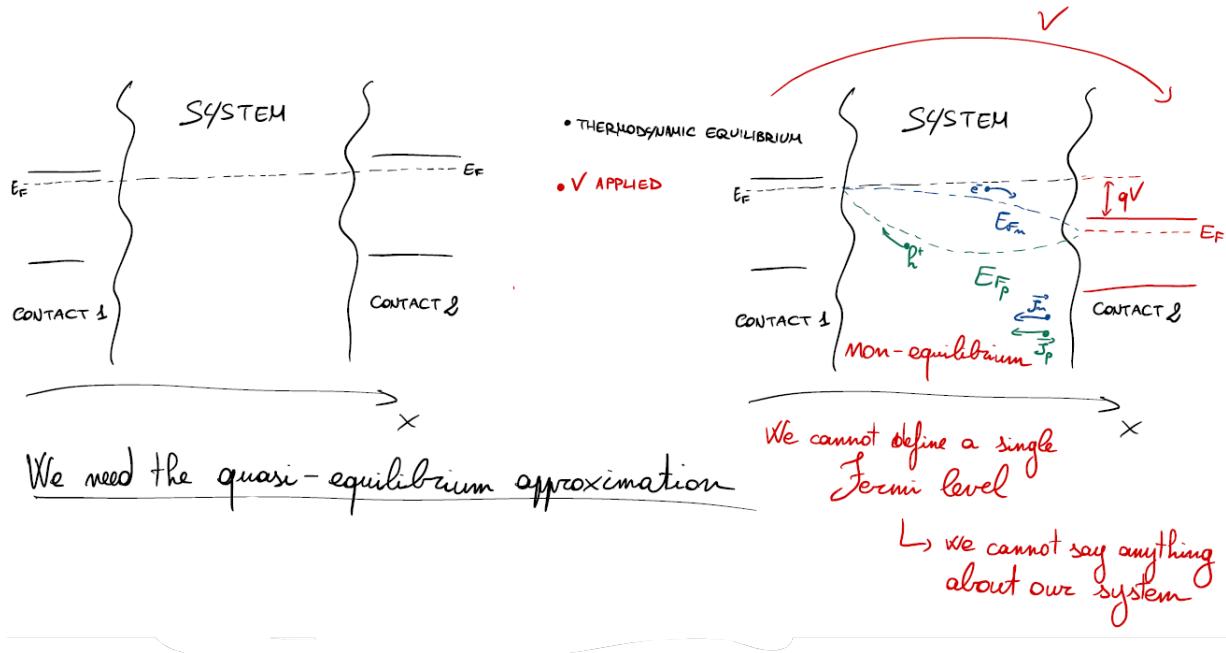
and thus

$$J_p = -qp\mu_p \frac{d\phi_F}{dx} \quad (1.53)$$

Remember that under thermodynamic equilibrium E_F is constant, so $\frac{d\phi_F}{dx} = 0 \Rightarrow J_n = J_p = 0$; net flowing carriers are null as drift and diffusion contributions perfectly balance each other. In order to achieve some net processes we have to perturb the equilibrium state by applying a voltage difference between the contacts⁷.

⁶Usually $L_D \sim 10^{-9} \text{ m}$

⁷We are considering ideal contacts for this analysis



If the perturbation is weak, carriers distribution follows the equilibrium statistics, meaning that electrons are close to the bottom of E_c and holes are close to the top of E_v ; quasi-Fermi levels E_{F_n} and E_{F_p} are also introduced, so 1.19 and 1.20 have to be rewritten as

$$n = n_i e^{\frac{E_{F_n} - E_i}{KT}} \quad (1.54)$$

$$p = n_i e^{\frac{E_i - E_{F_p}}{KT}} \quad (1.55)$$

PAY ATTENTION TO THE PRODUCT

$$pn = n_i^2 e^{\frac{E_{F_n} - E_{F_p}}{KT}} \quad (1.56)$$

Law of mass action is no longer valid! Also 1.52 and 1.53 present a big change as the gradient is not null

$$J_n = -qn\mu_n \frac{d\phi_{F_n}}{dx} = n\mu_n \frac{dE_{F_n}}{dx} \quad (1.57)$$

$$J_p = -qp\mu_p \frac{d\phi_{F_p}}{dx} = p\mu_p \frac{dE_{F_p}}{dx} \quad (1.58)$$

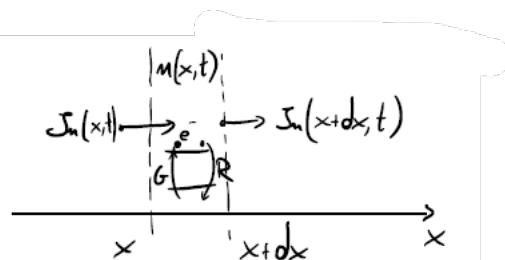
Contacts force the quasi-Fermi levels to merge into the original Fermi level, while in the system they can change. A new problem appears, as Poisson equation is no longer sufficient to study our system.

1.10 Continuity equation

When the system is perturbed, Poisson equation is not enough: we have to add equations for the quasi-Fermi levels

$$\begin{cases} \phi = \phi(x) \\ E_{F_n} = E_{F_n}(x) \\ E_{F_p} = E_{F_p}(x) \end{cases} \quad (1.59)$$

We introduce the **continuity equations**



$$\frac{\partial n}{\partial t} dx = -\frac{J_n(x, t)}{q} + \frac{J_n(x + dx, t)}{t} + (G - R)dx \quad (1.60)$$

We can simplify even further exploiting Taylor expansion $J(x + dx, t) \approx J_n(x, t) + \frac{\partial J_n}{\partial x} dx$ so

$$\begin{cases} \frac{\partial n}{\partial t} = \frac{1}{q} \frac{\partial J_n}{\partial x} + (G - R)_e \\ \frac{\partial p}{\partial t} = -\frac{1}{q} \frac{\partial J_p}{\partial x} + (G - R)_h \\ \frac{d^2\phi}{dx^2} = -\frac{q}{\epsilon_{si}}(p - n + N_D^+ - N_A^-) \end{cases} \quad (1.61)$$

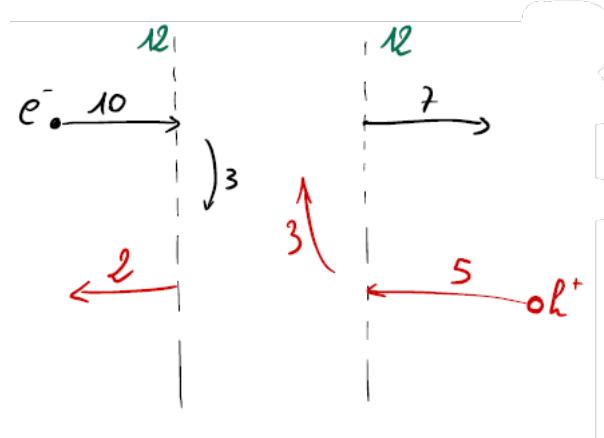
We got everything we need to have. Now let's assume we are in stationary conditions (nothing changes over time), the result is interesting:

$$\frac{\partial n}{\partial t} = 0 \implies \frac{1}{q} \frac{\partial J_n}{\partial x} + (G - R) = 0 \quad (1.62)$$

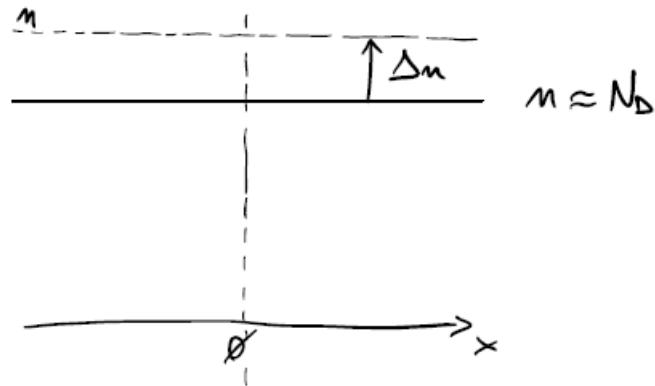
$$\frac{\partial p}{\partial t} = 0 \implies \frac{1}{q} \frac{\partial J_p}{\partial x} + (G - R) = 0 \quad (1.63)$$

therefore we can say that

$$\frac{1}{q} \frac{\partial(J_n + J_p)}{\partial x} = 0 \implies J_n + J_p \text{ is constant} \quad (1.64)$$

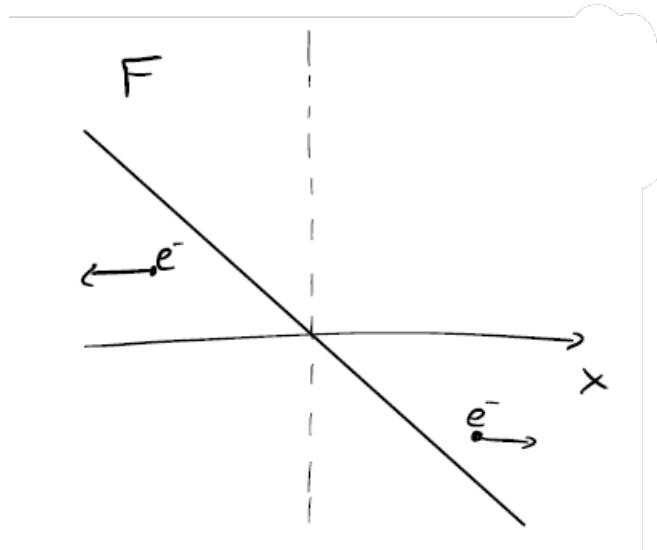


What if we were to increase the doping concentration of the majority carriers?



We now have some charge exposed, so an electric field is generated

$$\frac{d^2\phi}{dx^2} = -\frac{q}{\epsilon_{si}} (\cancel{p} - n + N_D^+ - \cancel{N_A^-}) = -\frac{q}{\epsilon_{si}} (-\cancel{N_D^-} - \Delta n + \cancel{N_D^-}) = \frac{q}{\epsilon_{si}} \Delta n = -\frac{dF}{dx} \quad (1.65)$$



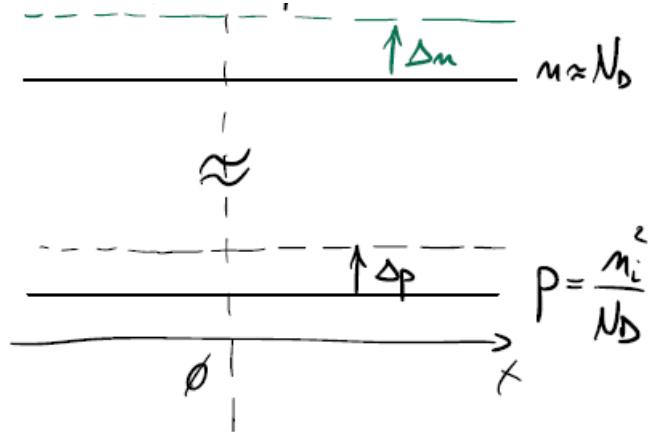
Electrons move away from the material due to the involved electric field: the system is trying to reach thermodynamic equilibrium (without diffusion)

$$\frac{\partial n}{\partial t} = \frac{1}{q} \frac{dJ_n}{dx} + (\cancel{G} - \cancel{R}) = \frac{1}{q} (qn\mu_n) \frac{\partial F}{\partial x} \quad (1.66)$$

then

$$\frac{\partial \Delta n}{\partial t} = -\frac{1}{q} \sigma \frac{d\Delta n}{\epsilon_{Si}} = -\frac{\Delta n}{\rho \epsilon_{Si}} \quad \Rightarrow \quad \Delta n(t) = \Delta n(0) e^{-\frac{t}{\tau_R}} \quad (1.67)$$

We can define $\tau_R = \rho \epsilon_{Si}$ as the dielectric relaxation time⁸. *Equilibrium is restored very quickly.* But what if we perturb the minority carriers now?



As before, we have an electric field pushing away holes, **but since they are a minority, resistivity is higher and electrons are attracted into the material because they are a majority.** Now two steps are taken into account:

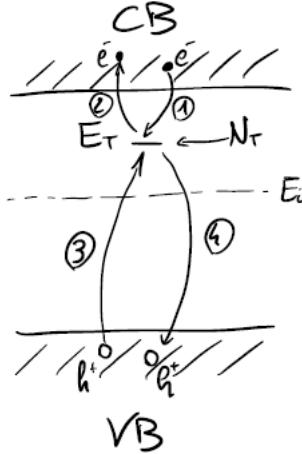
1. Process ends when $\Delta p \approx \Delta n$; a quasi-neutral region is formed.
2. Excessive electrons and holes recombine.

Since process 2 is way faster than 1, we can skip directly to this in our analysis.

1.11 Shockley-Read-Hall theory

We can describe generation and recombination processes through Shockley-Read-Hall theory for defect assisted processes

⁸Usually $\tau_R \approx 10^{-12} \text{ s}$



If the defect is empty, it can either capture an electron (1) or release a hole (4), while if it is filled, it can either release an electron (2) or capture a hole (3).

$$R : (1) + (3) \quad (1.68)$$

$$G : (2) + (4) \quad (1.69)$$

Remember that direct jump through the band gap is very unlikely. Now it is time to find G-R rates:

1.

$$r_1 = N_T(1-f)n v_{\text{TH}} \sigma_n \quad [\text{cm}^{-3}\text{s}^{-1}] \quad (1.70)$$

where these are the empty defects in Si and this is the probability that an electron is captured c_n .

2.

$$r_2 = N_T f e_n \quad (1.71)$$

where e_n is the emission rate; unfortunately, we cannot say anything because free states are way more than electrons, but if we assume thermodynamic equilibrium f is 1.6 and $r_1 = r_2$ as there cannot be net processes, so

$$e_n = \frac{1-f}{f} n v_{\text{TH}} \sigma_n = e \frac{E_T - E_F}{KT} n_i e \frac{E_F - E_i}{KT} v_{\text{TH}} \sigma_n = n_i v_{\text{TH}} \sigma_n e \frac{E_T - E_i}{KT} \quad (1.72)$$

Emission depends exponentially on $E_T - E_i$ and so is energy gain, while energy loss does not.

3.

$$r_3 = N_T f p v_{\text{TH}} \sigma_p \quad (1.73)$$

4.

$$r_4 = N_T(1-f)e_p \quad \Rightarrow \quad e_p = n_i v_{\text{TH}} \sigma_p e \frac{E_i - E_T}{KT} \quad (1.74)$$

In the end $R = r_1 - r_2 = r_3 - r_4 \rightarrow \mathcal{F}$ where \mathcal{F} is the Fermi-Dirac statistics only when $R = 0$. Doing some calculations, we eventually get

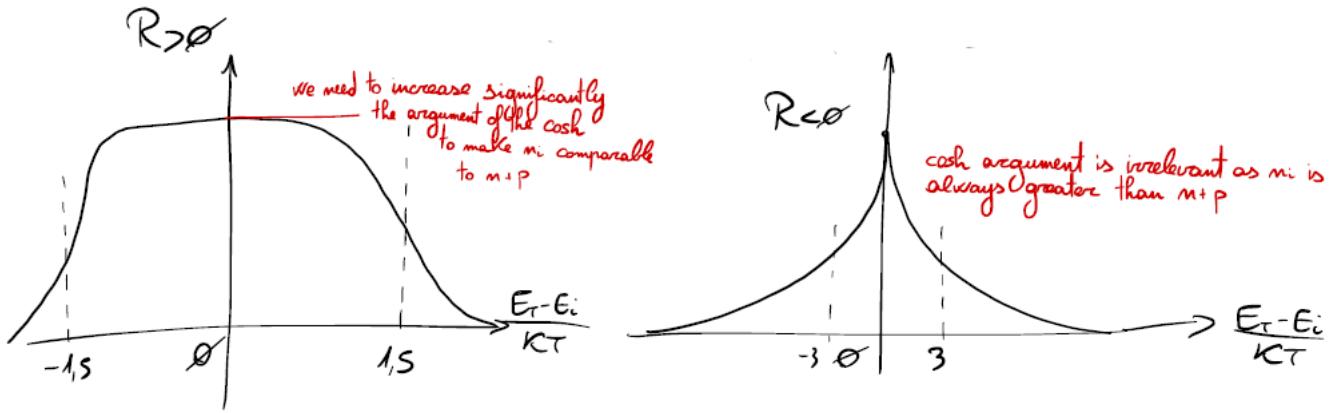
$$R = \frac{pn - n_i^2}{\tau_0 \left[p + n + 2n_i \cosh \left(\frac{E_T - E_i}{KT} \right) \right]} \quad (1.75)$$

where $\tau_0 = \frac{1}{N_T v_{\text{TH}} \sigma}$ assuming $\sigma_n = \sigma_p = \sigma$.

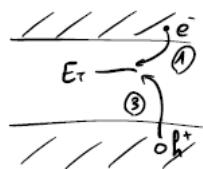
1.12 Main dependences of R

First of all it is important to assess the sign of R : since the denominator is always positive, we have to look at the numerator:

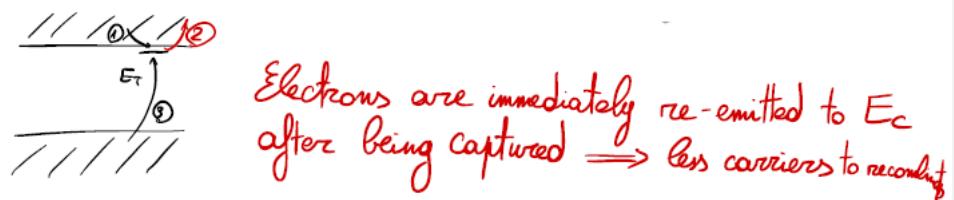
- $pn = n_i^2$ and $E_{F_n} = E_{F_p}$: $R = 0$, no G/R;
- $pn > n_i^2$ and $E_{F_n} > E_{F_p}$: $R > 0$, net R;
- $pn < n_i^2$ and $E_{F_n} < E_{F_p}$: $R < 0$, net G.



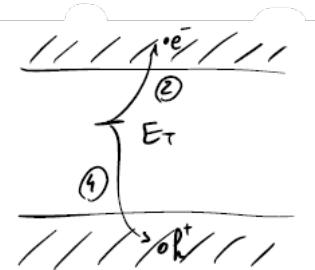
since \cosh is an even function, E_T can be either above or below E_i without any difference. Now we consider the situation where we have just (1) and (3)



In this case $R = r_1 = r_3$, therefore $f = \frac{n}{p+n}$ and then $R = \frac{N_T p n v_{TH} \sigma}{p+n} = \frac{p n}{\tau_0 (p+n)}$ as expected from before (net recombination does not depend on $E_T - E_i$). The drop after $\pm 1,5$ in the general formula is determined by the fact that E_T is very getting very close to E_c where (2) happens.



What if we have just (2) and (4)?



$$As we can imagine R \approx -\frac{n_i}{\tau_0 2 \cosh \left(\frac{E_T - E_i}{kT} \right)}$$

Quasi-neutral region with low level of injection is a common case:

$$\begin{cases} n = n_0 + \Delta n \\ p = p_0 + \Delta p \end{cases} \quad (1.76)$$

where $\Delta n = \Delta p << n_0 + p_0$; we can now simplify 1.75:

$$R = \frac{(p_0 + \Delta n)(n_0 + \Delta n) - n_i^2}{\tau_0 \left[p_0 + \Delta n + n_0 + \Delta n + 2n_i \cosh \left(\frac{E_T - E_i}{KT} \right) \right]} = \frac{p_0 n_0 + \Delta n n_0 + \Delta n p_0 + \Delta n^2 - \cancel{n_i^2}}{\tau_0 \left[p_0 + n_0 + 2\Delta n + 2n_i \cosh \left(\frac{E_T - E_i}{KT} \right) \right]} =$$
(1.77)

$$= \frac{\Delta n(p_0 + n_0 + \cancel{\Delta n})}{\tau_0 \left[p_0 + n_0 + 2\cancel{\Delta n} + 2n_i \cosh \left(\frac{E_T - E_i}{KT} \right) \right]} = \frac{\Delta n}{\tau_n}$$

$$p_0 + n_0 + 2n_i \cosh \left(\frac{E_T - E_i}{KT} \right)$$

where $\tau_n = \tau_0 \frac{p_0 + n_0 + 2n_i \cosh \left(\frac{E_T - E_i}{KT} \right)}{p_0 + n_0}$; if $E_T = E_i$ and low injection is confirmed, $\tau_n \approx \tau_0$ ⁹, otherwise it is not a constant. Finally we get

$$\frac{\partial n}{\partial t} = \frac{1}{g} \frac{\partial J_n}{\partial x} - \frac{\Delta n}{\tau_n} \quad \Rightarrow \quad \Delta n(t) = \Delta n(0) e^{-\frac{t}{\tau_n}}$$
(1.78)

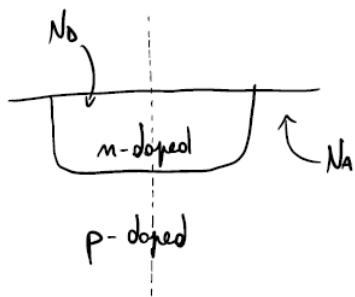
⁹High quality materials have a very long τ_n

Chapter 2

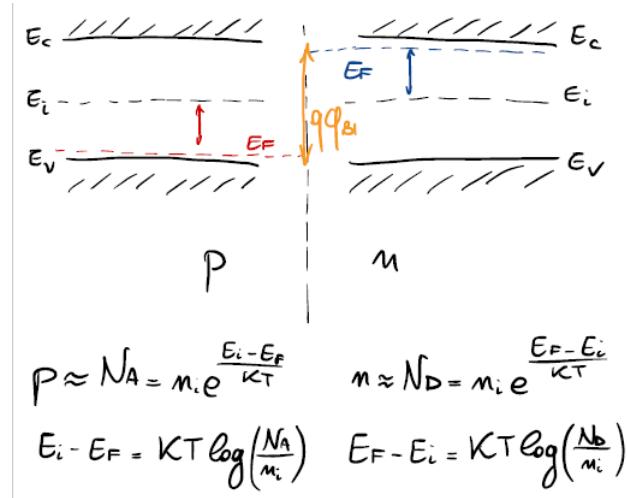
Diode (p-n junction)

2.1 Basics of the p-n junction

We enter into the topic of the p-n junction, one of the most important electronic components. Basically it is nothing more than two semiconductor layers, respectively n-doped and p-doped, joined together. For this analysis, we will always assume complete ionization, so $\tilde{\rho} = q(p - n + N_D - N_A)$.

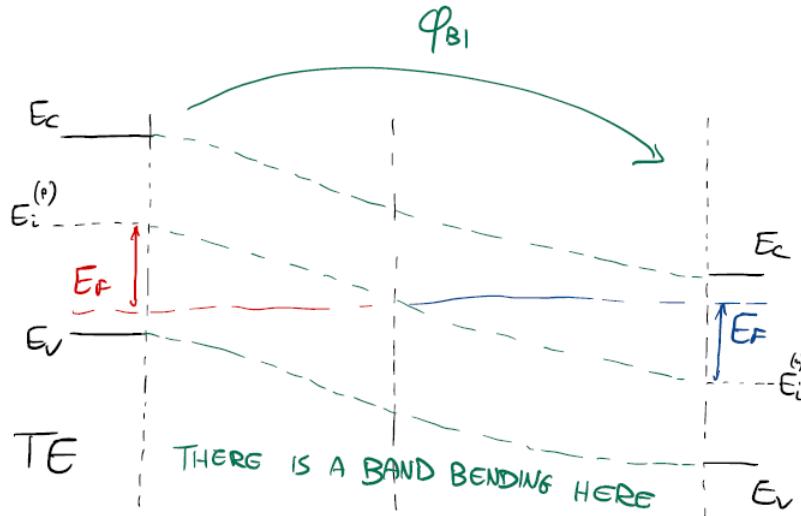


It is vital to underline that mobility is affected by the total doping level. First we are considering the two layers separately



then we put them in contact, keeping in mind that E_F must be unique; we expect a band bending¹.

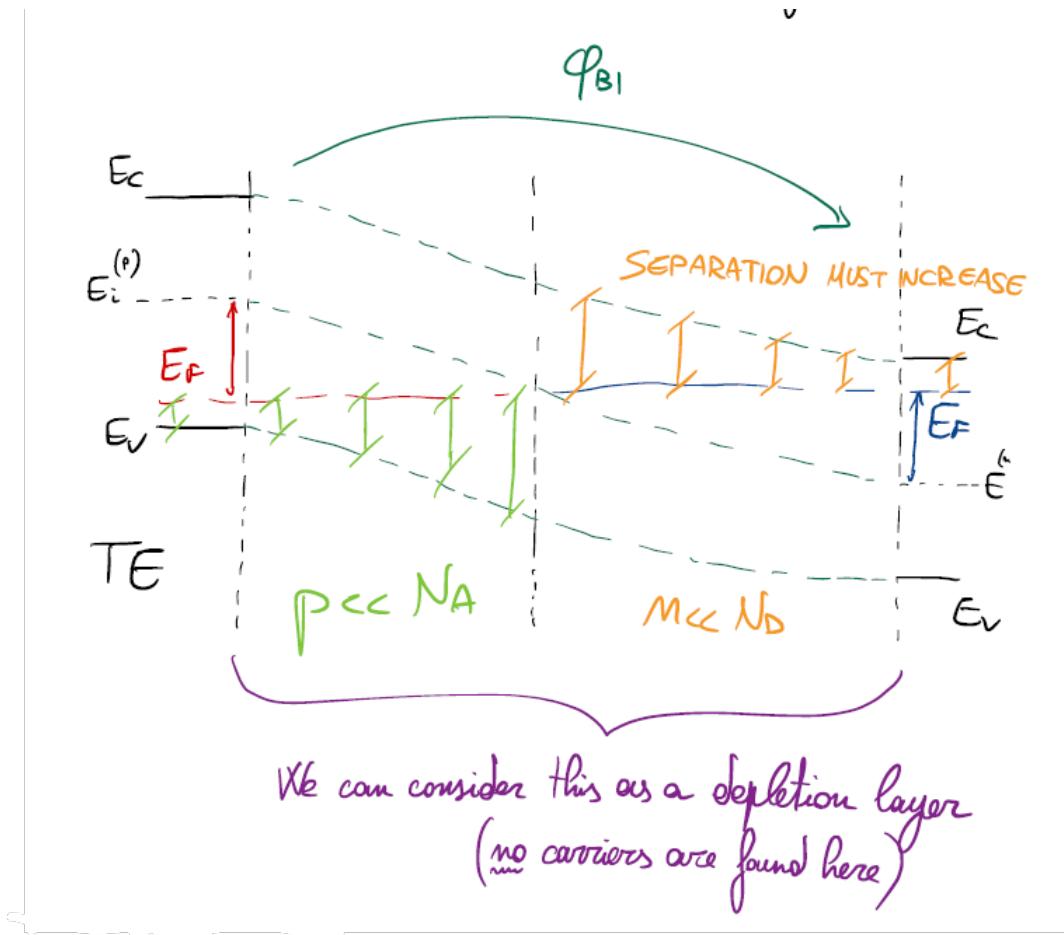
¹It is reasonable to assume charge neutrality at the extremes



ϕ_{BI} is the built-in tension, a inner characteristic of the component² and it is found as

$$\phi_{BI} = \frac{E_i^{(p)} - E_i^{(n)}}{q} \Big|_{TE} = \frac{E_F^{(n)} - E_F^{(p)}}{q} \Big|_{TE} = \frac{KT}{q} \log \left(\frac{N_A N_D}{n_i^2} \right) \quad (2.1)$$

We must exploit Poisson equation once again to seek for more informations about the bending; reasonably, carriers concentrations in the middle region are close to none compared to the doping concentrations



therefore 1.44 is simplified as

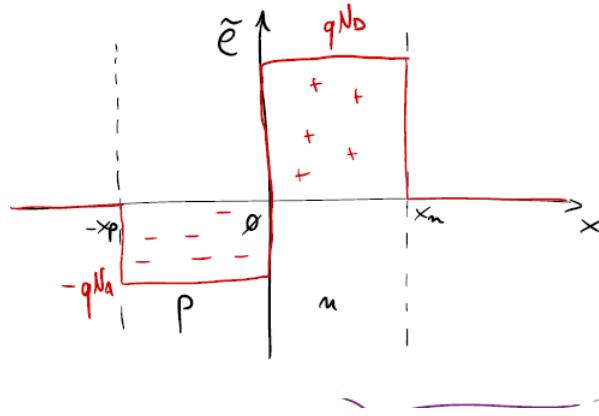
$$\frac{d^2\phi}{dx^2} \approx -\frac{q}{\epsilon_{Si}} (N_D - N_A) \quad (2.2)$$

even though it is not valid in general.

²That would be the total voltage drop across the transition region

2.2 Solution of the Poisson equation

Starting from 2.2 we can focus our analysis just on the depletion layer; considering charge density



we can write for the n -region interval $0 \leq x \leq x_n$:

$$\begin{aligned} \frac{d^2\phi}{dx^2} &= -\frac{q}{\epsilon_{\text{Si}}} N_D \\ \frac{d}{dx} \left(\frac{d\phi}{dx} \right) &= -\frac{q}{\epsilon_{\text{Si}}} N_D \\ d \left(\frac{d\phi}{dx} \right) &= -\frac{q}{\epsilon_{\text{Si}}} N_D dx \\ \int_{\frac{d\phi}{dx}(x)}^{\frac{d\phi}{dx}(x_n)} d \left(\frac{d\phi}{dx} \right) &= \int_x^{x_n} -\frac{q}{\epsilon_{\text{Si}}} N_D dx \\ \left. \frac{d\phi}{dx} \right|_x^0 - \left. \frac{d\phi}{dx} \right|_x &= -\frac{q}{\epsilon_{\text{Si}}} N_D (x_n - x) \end{aligned}$$

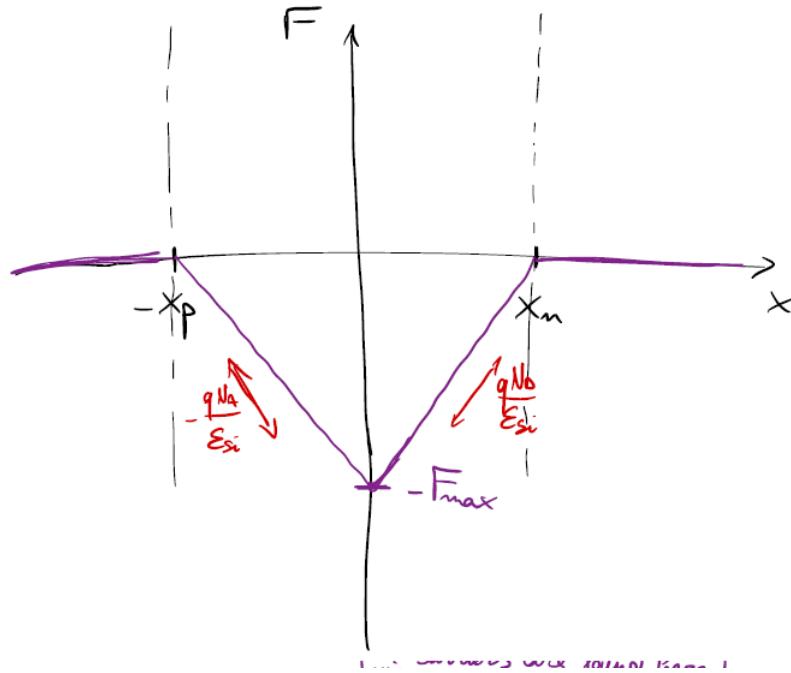
and thus

$$\left. \frac{d\phi}{dx} \right|_x = \frac{qN_D}{\epsilon_{\text{Si}}} (x_n - x) = -F \quad (2.3)$$

For the other side $-x_p \leq x \leq 0$:

$$\left. \frac{d\phi}{dx} \right|_x = \frac{qN_A}{\epsilon_{\text{Si}}} (x + x_p) = -F \quad (2.4)$$

Not very surprisingly we have just discovered that $F = F(x)$: now its behavior can be represented (keep in mind that for Gauss law $\frac{dF}{dx} = \frac{\tilde{\rho}}{\epsilon_{\text{Si}}}$ we already know the slopes)



Still not very surprisingly we acknowledge that no discontinuity is possible as ϵ_{Si} does not change, so when $F = F_{\text{MAX}}$

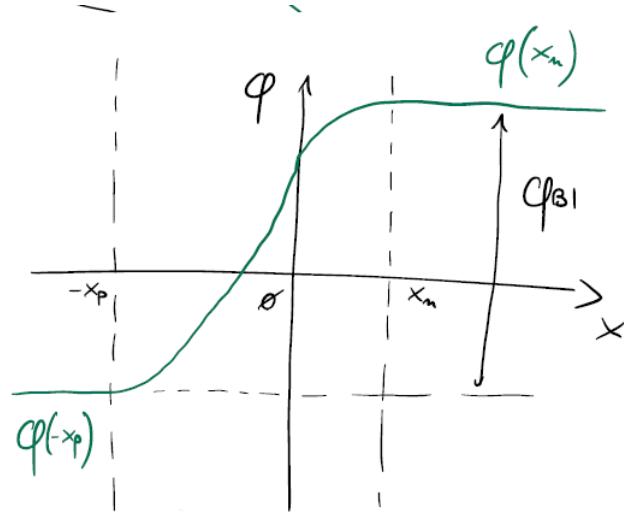
$$\frac{qN_D}{\epsilon_{\text{Si}}} x_n = \frac{qN_A}{\epsilon_{\text{Si}}} x_p \implies N_D x_n = N_A x_p \quad (2.5)$$

Total charge on the two sides must be null still due to Gauss law. It is time to find the potential, simply by integrating 2.3 between $\phi(x)$ and $\phi(x_n)$

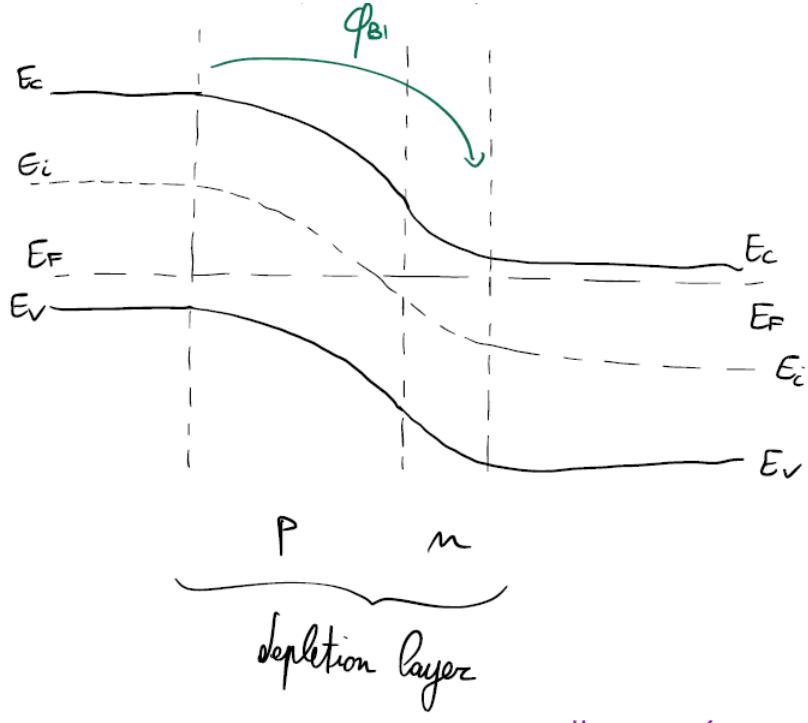
$$\phi(x) = \phi(x_n) - \frac{qN_D}{2\epsilon_{\text{Si}}}(x_n - x)^2 \quad (2.6)$$

and for the other side

$$\phi(x) = \phi(-x_p) + \frac{qN_A}{2\epsilon_{\text{Si}}}(x + x_p)^2 \quad (2.7)$$



Total electrostatic potential is nothing else but the union between two parabolas: band bending is then parabolic.



Depletion layer extension We are ready to compute the actual extension of the depletion layer; first we must assure that ϕ and F are continuous in $x = 0$

$$\begin{cases} \phi(x_n) - \phi(x_p) = \frac{qN_D}{2\epsilon_{Si}}x_n^2 + \frac{qN_A}{2\epsilon_{Si}}x_p^2 = \phi_{BI} \\ N_D x_n = N_A x_p \end{cases} \quad (2.8)$$

then we explicit x_n and x_p

$$\begin{cases} x_n = \sqrt{\frac{2\epsilon_{Si}}{q} \left(\frac{1}{N_A} + \frac{1}{N_D} \right) \phi_{BI}} \frac{N_A}{N_A + N_D} \\ x_p = \sqrt{\frac{2\epsilon_{Si}}{q} \left(\frac{1}{N_A} + \frac{1}{N_D} \right) \phi_{BI}} \frac{N_D}{N_A + N_D} \end{cases} \quad (2.9)$$

and eventually we sum the two terms

$$W_D = x_n + x_p = \sqrt{\frac{2\epsilon_{Si}}{q} \left(\frac{1}{N_A} + \frac{1}{N_D} \right) \phi_{BI}} \quad (2.10)$$

2.3 Forward and reverse bias

We start this section with an interesting consideration: W_D is determined by the less doped region, in fact

$$W_D \approx \sqrt{\frac{2\epsilon_{Si}}{qN_{Low}} \phi_{BI}} \quad (2.11)$$

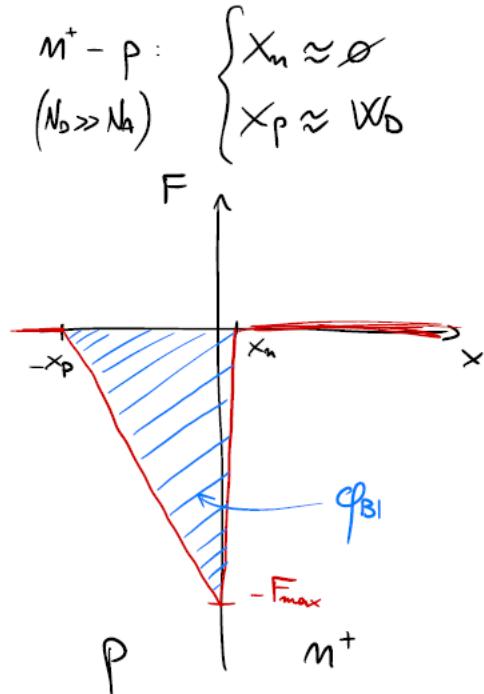
also it is true that

$$\begin{cases} x_n = W_D \frac{N_A}{N_A + N_D} \\ x_p = W_D \frac{N_D}{N_A + N_D} \end{cases} \quad (2.12)$$

Some values of W_D are reported in a table for the sake of better understanding the order of magnitude according to the doping concentration

$N_{Low} [cm^{-3}]$	$W_D [nm]$
10^{16}	300
10^{17}	100
10^{18}	30
10^{19}	10

When $N_{\text{HIGH}} \gg N_{\text{LOW}}$ it is clear that the depletion layer is almost equal to the extension of the less doped region



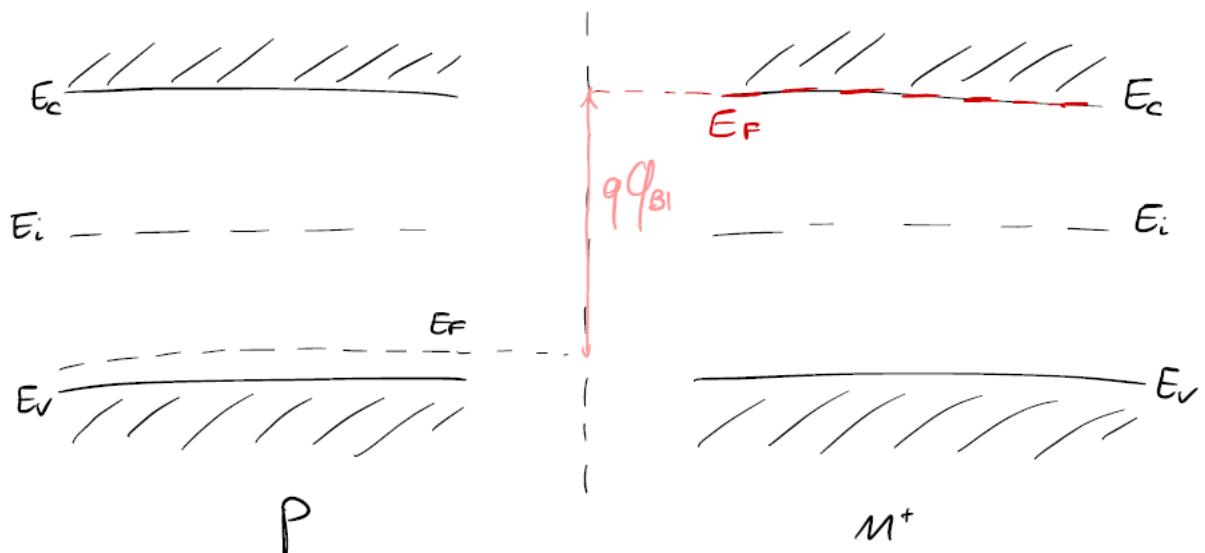
ϕ_{BI} is the area of the triangle:

$$\begin{cases} \phi_p = \frac{qN_A}{2\epsilon_{\text{Si}}} x_p^2 \\ \phi_n = \frac{qN_D}{2\epsilon_{\text{Si}}} x_n^2 \end{cases} \quad (2.13)$$

and of course

$$\phi_{\text{BI}} = \phi_p + \phi_n = \frac{q}{2\epsilon_{\text{Si}}} (N_A x_p^2 + N_D x_n^2) \quad (2.14)$$

When we have a n^+ region, Fermi level is degenerate

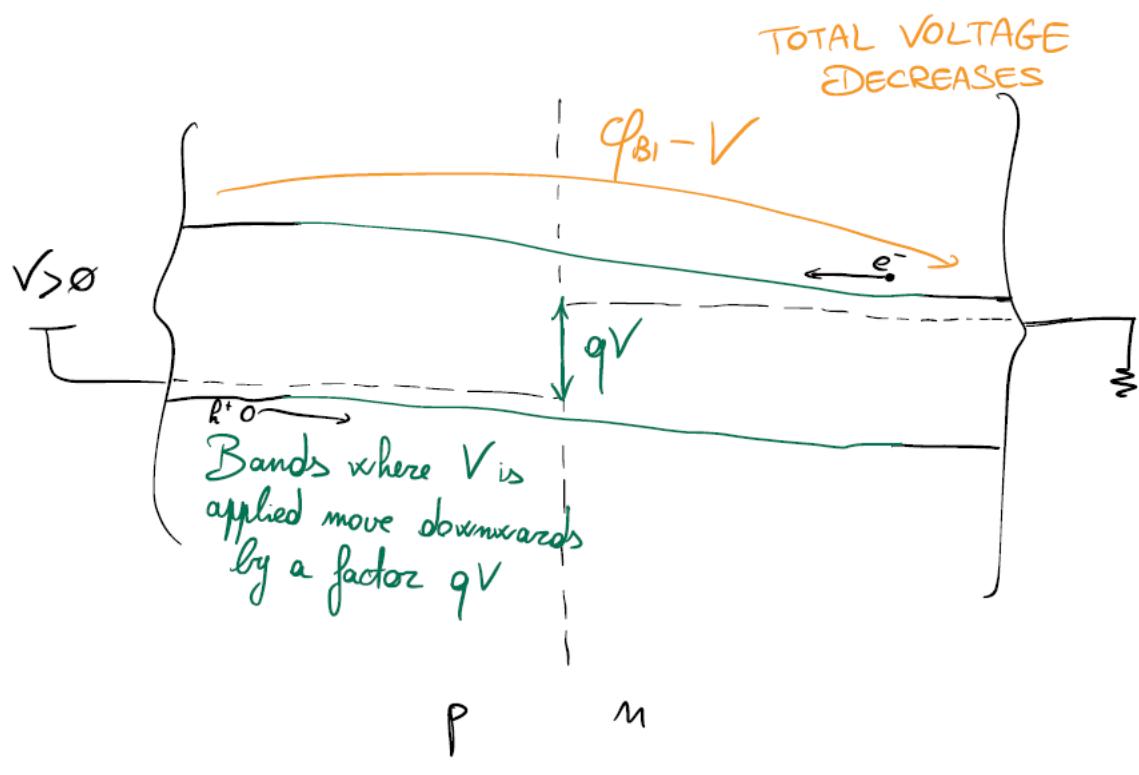


Maxwell-Boltzmann cannot be used, so

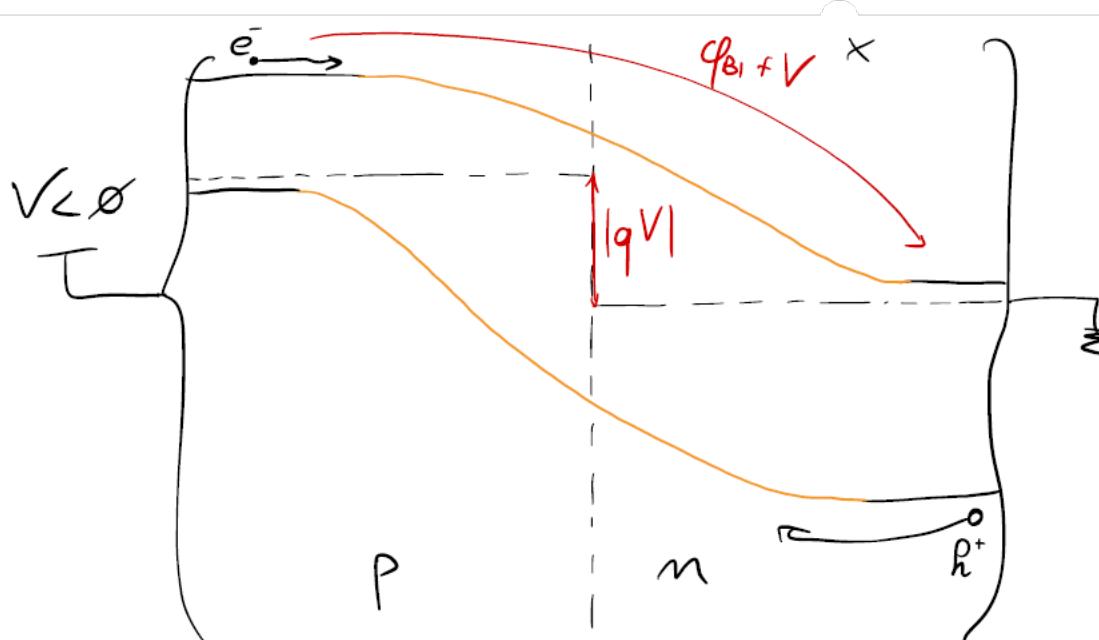
$$\phi_{\text{BI}} = \frac{E_G}{2q} + \frac{KT}{q} \log \left(\frac{N_A}{n_i} \right) \quad (2.15)$$

this is called *unilateral p-n junction*.

Considering the contacts ideal, how does this all change if we apply a bias?



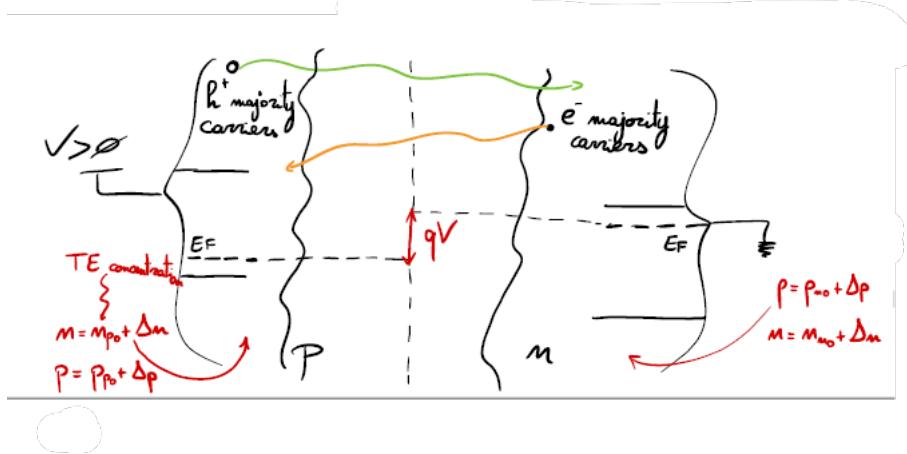
First the case where $V > 0$: bands flatten, bringing more carriers in the depletion layer. Carriers flow then becomes stronger and current density increases.



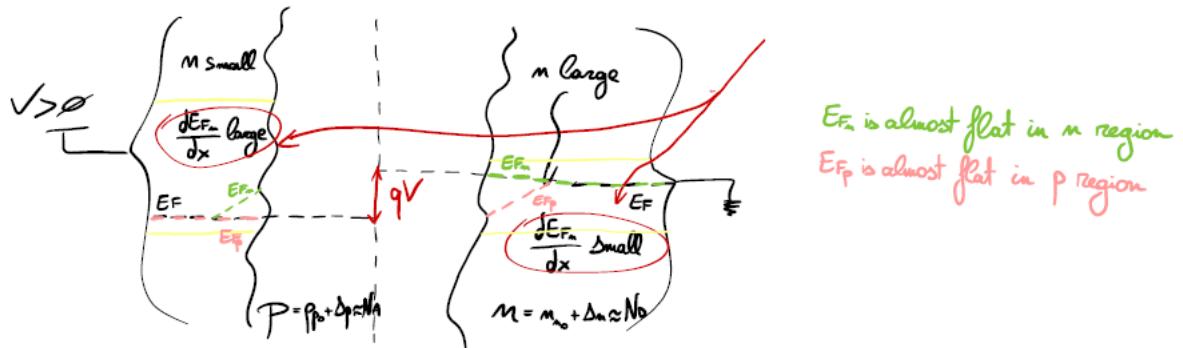
Instead, if $V < 0$, minority carriers flow to the region where they are majority, reducing the current density. In the overall, we can say that the diode is a **rectifying device**.

2.4 Qualitative behavior of E_{F_n} and E_{F_p}

Thermodynamic equilibrium is no more, so Fermi level is splitted as we have seen previously.



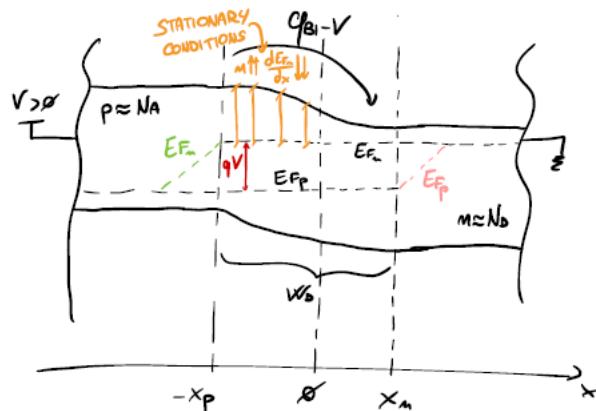
Our goal once again is to simplify 1.59, always considering quasi-neutral regions outside the depletion layer with low injection ($\Delta p = \Delta n \ll p_{n0} + n_{n0}$). Assuming stationary conditions (G/R absent), we know that $J_n = n\mu_n \frac{dE_{F_n}}{dx}$ must be constant.



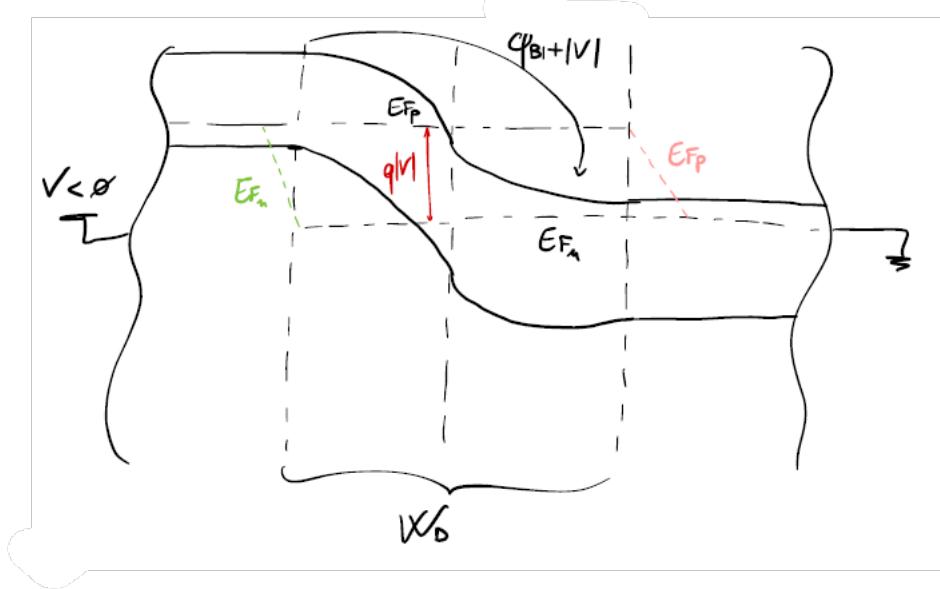
Bands are still behaving parabolically as if they were in thermodynamic equilibrium, but the depletion layer shrinks

$$W_D = \sqrt{\frac{2\epsilon_{Si}}{q} \left(\frac{1}{N_A} + \frac{1}{N_D} \right) (\phi_{Bi} - V)} \quad (2.16)$$

Quasi-Fermi levels drops are necessary to have a unique E_F at the contacts.

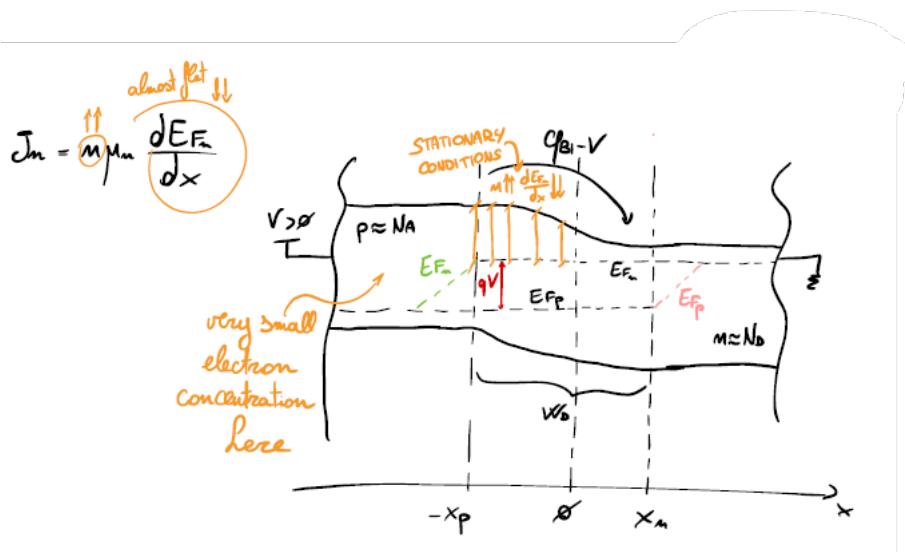


Symmetrically we can consider the reverse bias situation:



2.5 Minority carriers diffusion in quasi-neutral regions

Let's start this section commenting the forward biased band diagram.



If the gradient was clearly visible, electrons flow would be too high, but for now we assume E_{F_n} perfectly flat so that we know its relative position. Now we have to consider both depletion layer edges.

- $x = -x_p$:

$$\begin{cases} p \approx N_A \\ pn = n_i^2 e \frac{E_{F_n} - E_{F_p}}{KT} = n_i^2 e \frac{qV}{KT} \end{cases} \implies n = \frac{n_i^2}{N_A} e \frac{qV}{KT} = n_{p_0} e \frac{qV}{KT} \quad (2.17)$$

- $x = x_n$:

$$\begin{cases} n \approx N_D \\ p = p_{n_0} e \frac{qV}{KT} \end{cases} \quad (2.18)$$

Current increases exponentially; since quasi-neutral regions are constraining minority carriers current density, they have to be

studied carefully with the continuity equations.

$$\begin{cases} \frac{\partial n}{\partial t} = \frac{1}{q} \frac{\partial J_n}{\partial x} + (G - R) \\ J_n \approx q D_n \frac{dn}{dx} \\ (G - R) = -\frac{\Delta n}{\tau_n} \end{cases}$$

We substitute each member and get

$$\frac{\partial \Delta n}{\partial t} = D_n \frac{\partial^2 \Delta n}{\partial x^2} - \frac{\Delta n}{\tau_n} \quad (2.19)$$

also known as the **time-dependent diffusion equation**. In stationary conditions that becomes

$$\frac{d^2 \Delta n}{dx^2} - \frac{\Delta n}{D_n \tau_n} = \frac{d^2 \Delta n}{dx^2} - \frac{\Delta n}{L_n^2} = 0 \quad (2.20)$$

where $L_n = \sqrt{D_n \tau_n}$ is the diffusion length. Solution is straightforward

$$\Delta n(x) = A e^{\frac{x}{L_n}} + B e^{-\frac{x}{L_n}}$$

and considering the proper initial conditions

$$\begin{aligned} \Delta n(0) &= n_{p_0} \left(e^{\frac{qV}{KT}} - 1 \right) = A + B \\ \Delta n(W_p) &= A e^{\frac{W_p}{L_n}} + B e^{-\frac{W_p}{L_n}} = 0 \end{aligned}$$

that results in

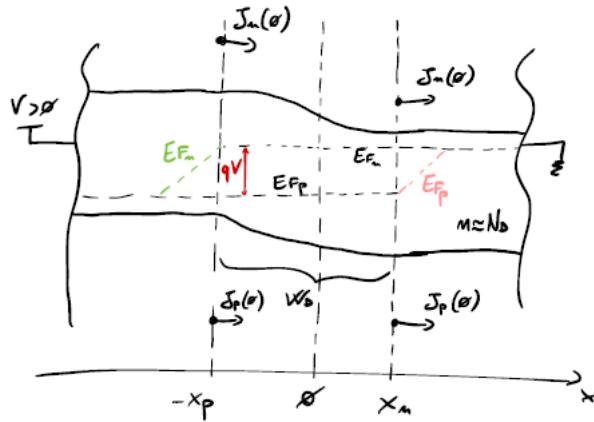
$$\Delta n(x) = \Delta n(0) \frac{-e^{-\frac{W_p-x}{L_n}} + e^{\frac{W_p-x}{L_n}}}{e^{\frac{W_p}{L_n}} - e^{-\frac{W_p}{L_n}}} = \Delta n(0) \frac{\sinh\left(\frac{W_p-x}{L_n}\right)}{\sinh\left(\frac{W_p}{L_n}\right)} \quad (2.21)$$

and thus in

$$J_n(0) = q D_n \frac{d\Delta n}{dx} = -q D_n \frac{\Delta n(0) \cosh\left(\frac{W_p-x}{L_n}\right) \Big|_{x=0}}{L_n \sinh\left(\frac{W_p}{L_n}\right)} = \frac{q D_n n_{p_0} \left(e^{\frac{qV}{KT}} - 1 \right)}{L_n \tanh\left(\frac{W_p}{L_n}\right)} \quad (2.22)$$

also for the p -region

$$J_p(0) = \frac{q D_p p_{n_0} \left(e^{\frac{qV}{KT}} - 1 \right)}{L_p \tanh\left(\frac{W_p}{L_n}\right)} \quad (2.23)$$



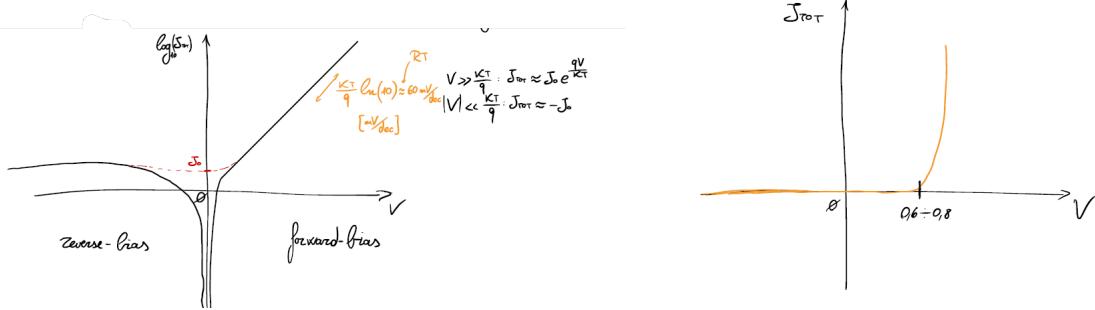
We conclude by saying that the total current density is obviously the sum of the two contributions

$$J_{\text{TOT}} = J_n(0) + J_p(0) = \left[\frac{q D_n n_{p_0}}{L_n \tanh\left(\frac{W_n}{L_n}\right)} + \frac{q D_p p_{n_0}}{L_p \tanh\left(\frac{W_p}{L_n}\right)} \right] \left(e^{\frac{qV}{KT}} - 1 \right) = J_0 \left(e^{\frac{qV}{KT}} - 1 \right) \quad (2.24)$$

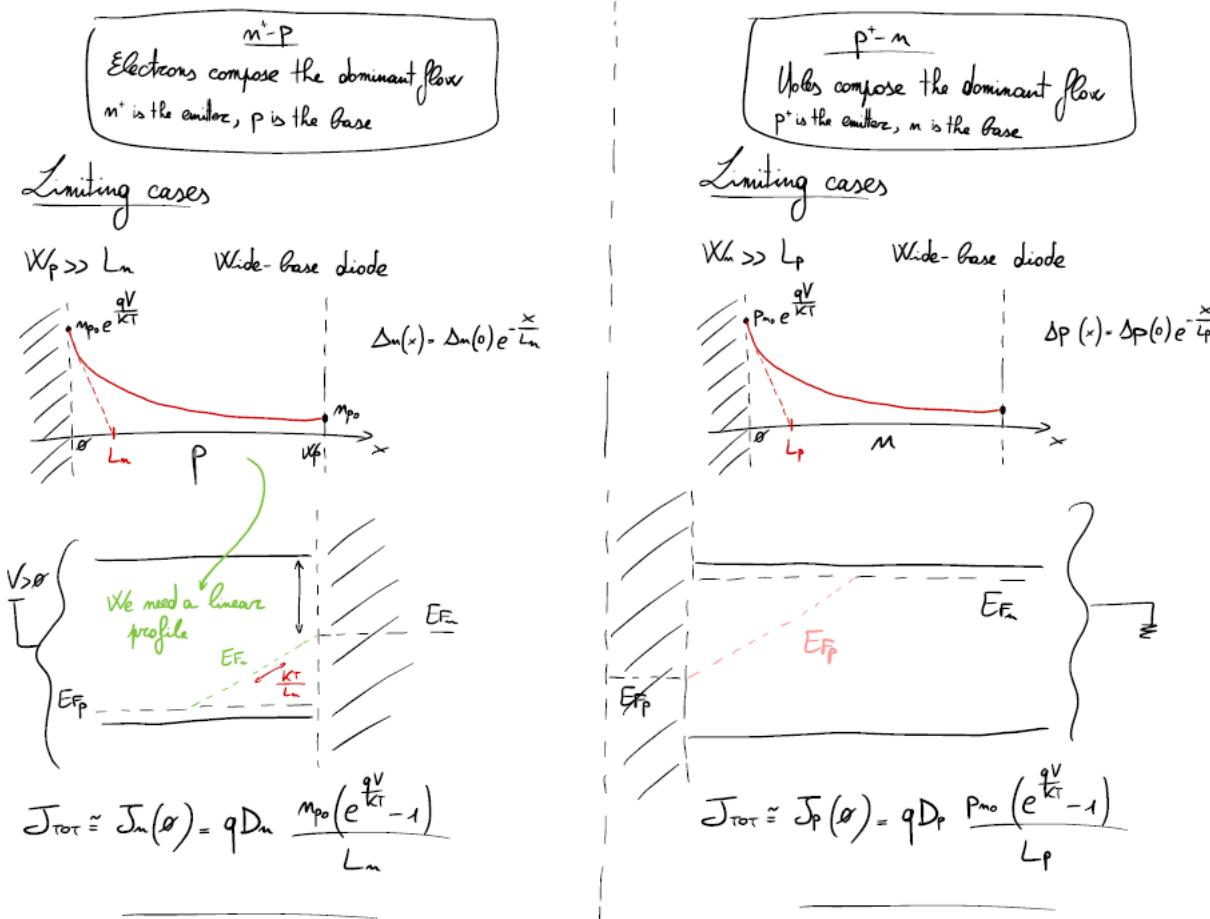
2.24 is known as the **Shockley's ideal diode equation**.

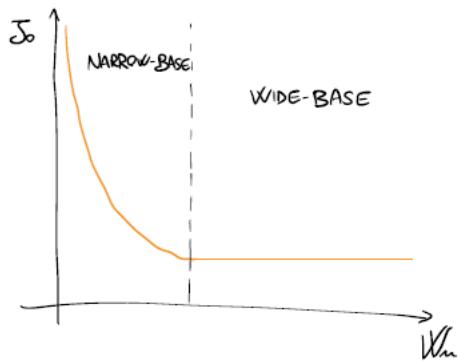
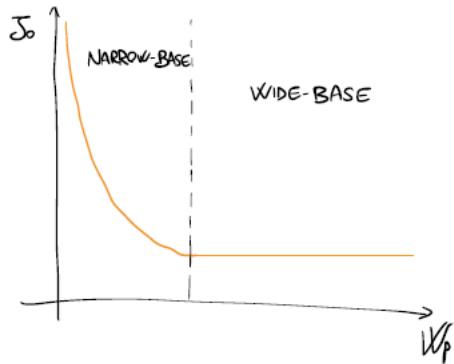
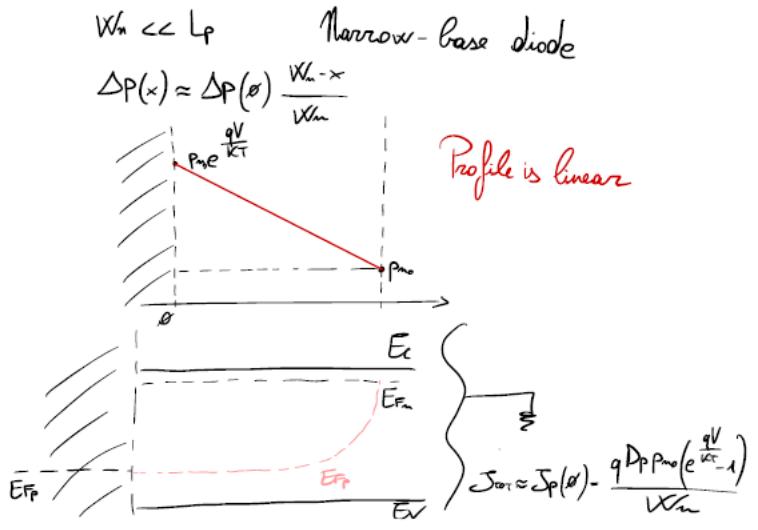
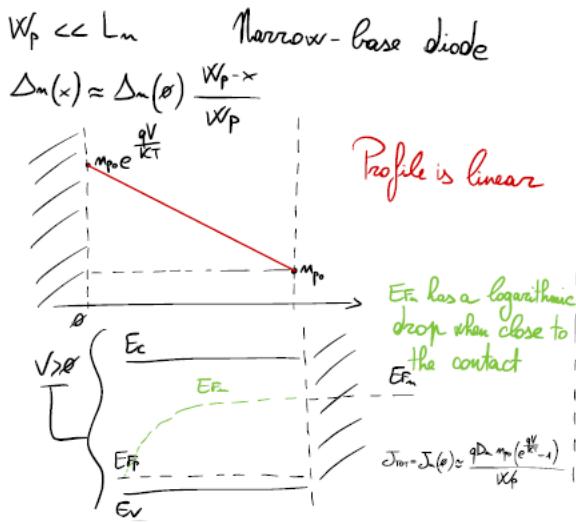
2.6 Wide-base and narrow-base diodes

From 2.24, the rectifying behavior of the diode appears very clearly: let's confront the $J - V$ curves in both logarithmic and linear scales



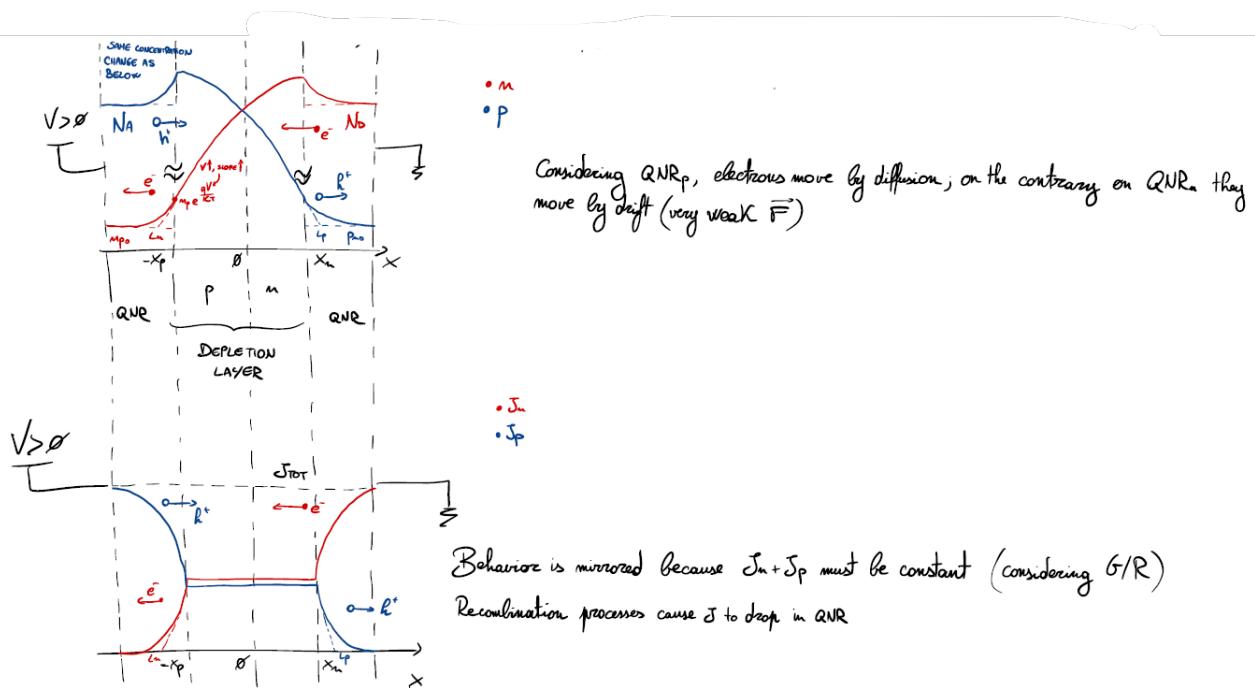
Total current density depends on both doping concentrations: if they are very different, one term is dominant over the other (in particular, less doped region is prevalent). Now here is a confront between two approximations for the diode according to the depleted layer depth.



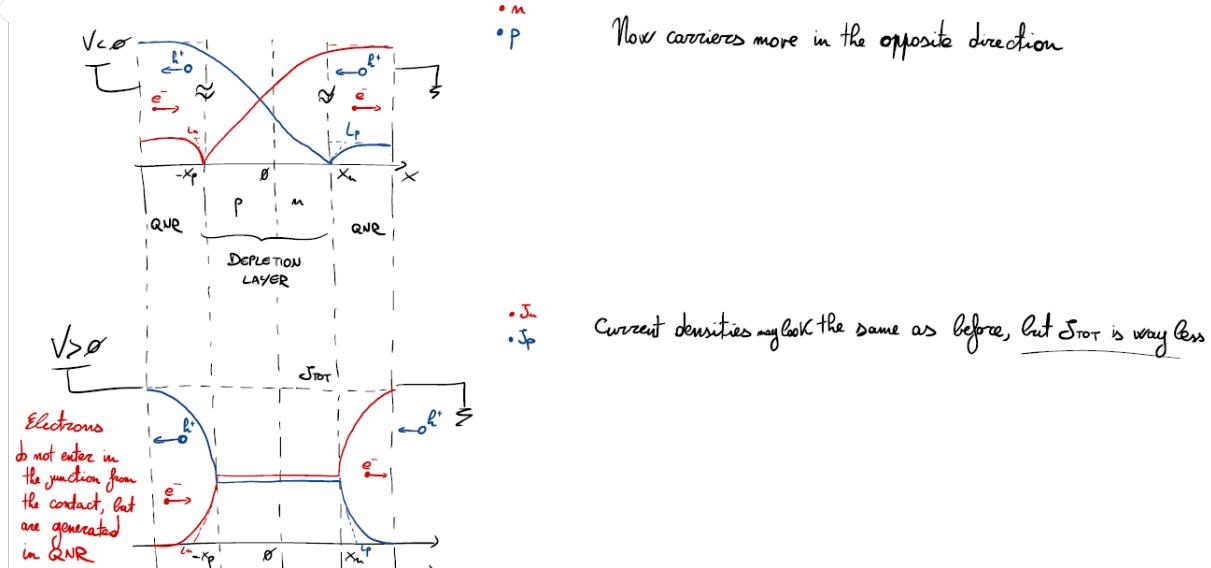


2.7 Spatial profiles for n, p, J_n, J_p

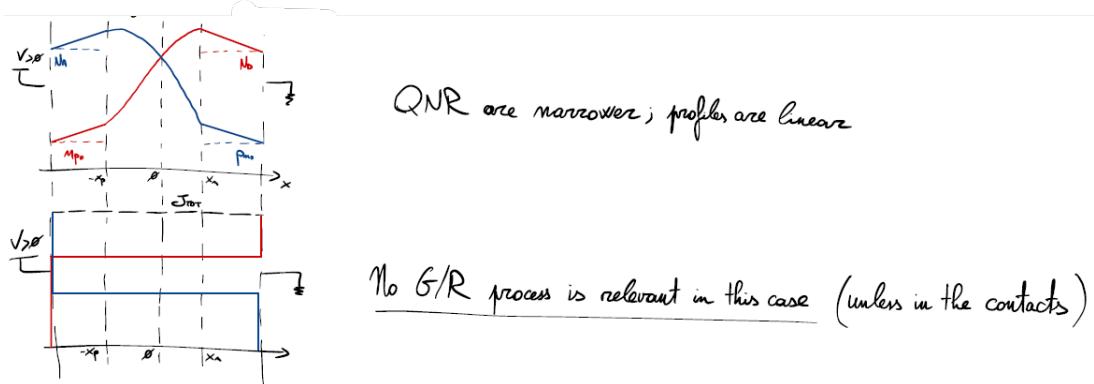
We studied the electrostatics, now we move onwards to the spatial profiles of carriers and current densities.



This happens for forward bias; in reverse bias on the other hand



If we consider a narrow-base diode, finally



2.8 Temperature dependence of the $J - V$ curve

Our starting point is still 2.24; first we would like to make some considerations about some orders of magnitude

$$N_A = 10^{16} \text{ cm}^{-3} \rightarrow \mu_n = 1250 \text{ cm}^2/\text{Vs} \rightarrow \tau_n = 30 \mu\text{s} \rightarrow D_n = 32,5 \text{ cm}^2/\text{s} \rightarrow L_n = 311 \mu\text{m}$$

$$N_D = 10^{17} \text{ cm}^{-3} \rightarrow \mu_p = 470 \text{ cm}^2/\text{Vs} \rightarrow \tau_p = 10 \mu\text{s} \rightarrow D_p = 12,2 \text{ cm}^2/\text{s} \rightarrow L_p = 110 \mu\text{m}$$

As we can see, those two lengths are very long; diodes are more likely to be narrow-base, thus G/R processes are weak in modern technologies.

$$\text{Wide-base: } J_0 = \frac{qD_n n_{p0}}{L_n} + \frac{qD_p p_{n0}}{L_p} = 3,87 \cdot 10^{-12} \text{ A/cm}^2$$

$$\text{Narrow-base: } J_0 = \frac{qD_n n_{p0}}{W_p} + \frac{qD_p p_{n0}}{W_n} = 10^{-9} \text{ A/cm}^2$$

Now it is time to investigate J_0 dependence on T : we can surely say that

$$J_0 \propto \frac{qD_n n_i^2}{N_A L_n \tanh\left(\frac{W_p}{L_n}\right)} = a T^\gamma e^{-\frac{E_G}{kT}} \quad (2.25)$$

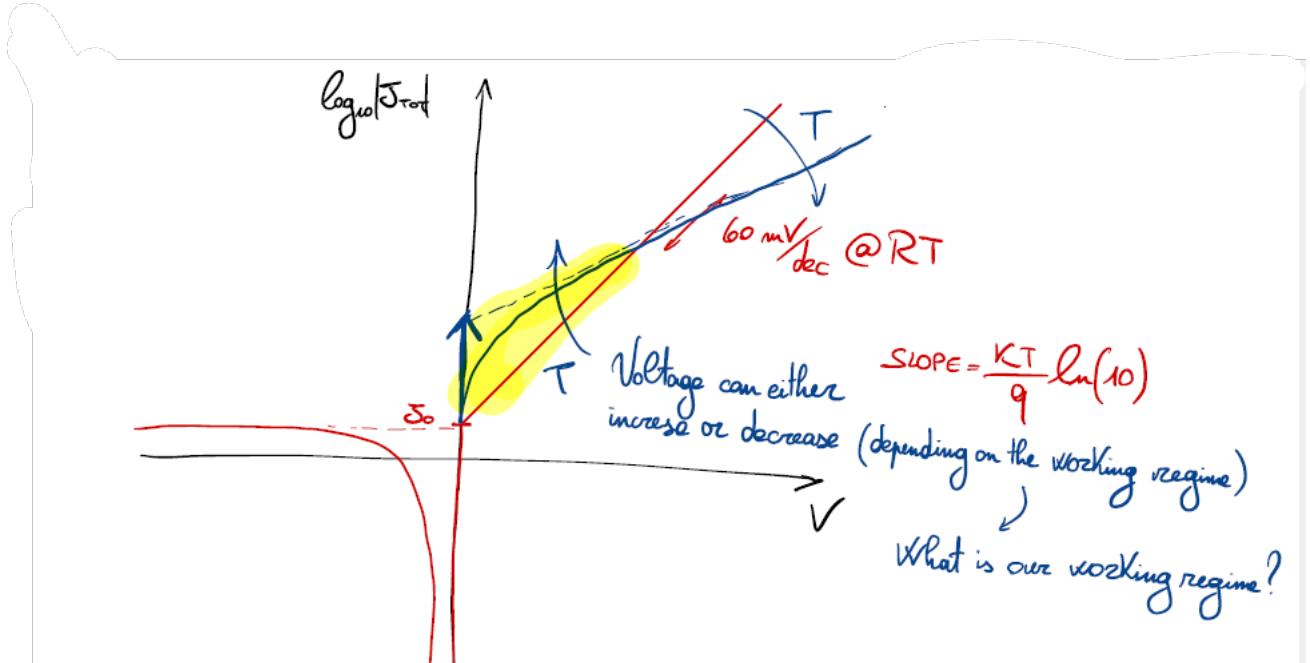
Therefore, always considering 2.24, we can safely say

$$\begin{cases} V = \frac{KT}{q} \log \left(\frac{\bar{J}_{\text{TOT}}}{J_0} \right) \\ \frac{dV}{dT} = \frac{K}{q} \log \left(\frac{\bar{J}_{\text{TOT}}}{J_0} \right) \frac{T}{T} + \frac{KT}{q} \cancel{\frac{J_0}{\bar{J}_{\text{TOT}}} \frac{-\bar{J}_{\text{TOT}}}{J_0^2} \frac{dJ_0}{dT}} = \frac{V}{T} - \frac{KT}{q} \frac{1}{J_0} \frac{dJ_0}{dT} \\ \frac{dJ_0}{dT} = a\gamma T^{\gamma-1} e^{-\frac{E_G}{KT} \frac{T}{T}} + aT^\gamma e^{-\frac{E_G}{KT}} \frac{-\frac{E_G}{dT} KT + E_G K}{(KT)^2} = \frac{\gamma J_0}{T} + J_0 \left[-\frac{dE_G}{dT} \frac{1}{KT} + \frac{E_G}{KT^2} \right] \end{cases}$$

By inserting the third into the second we obtain

$$\frac{dV}{dT} = \frac{V}{T} - \frac{KT}{q} \frac{1}{J_0} \left[\frac{\gamma J_0}{T} - \frac{J_0}{KT} \frac{dE_G}{dT} + \frac{J_0 E_G}{KT^2} \right] = \frac{V - \frac{E_G}{q}}{T} - \frac{K\gamma}{q} + \frac{1}{q} \frac{dE_G}{dT} \quad (2.26)$$

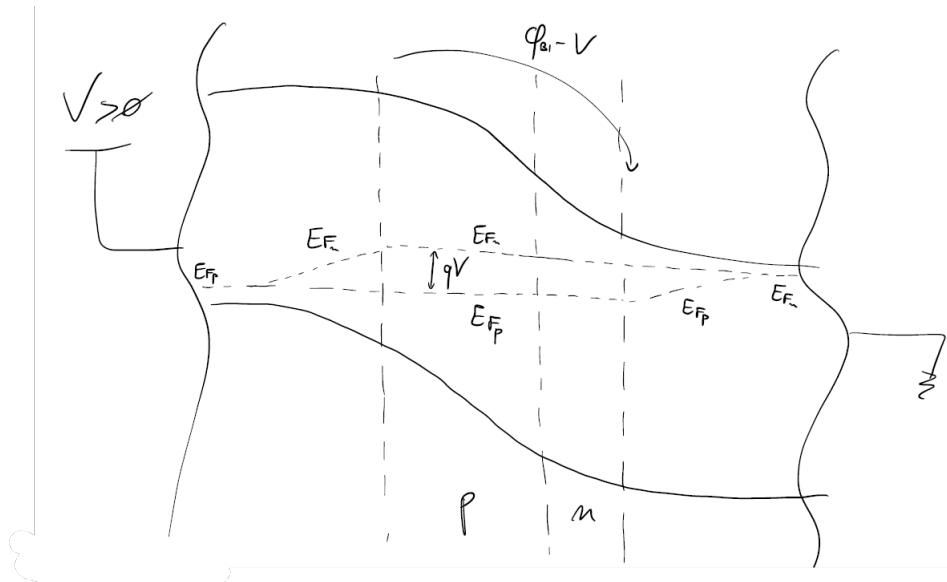
where the highlighted terms are smaller than 0.



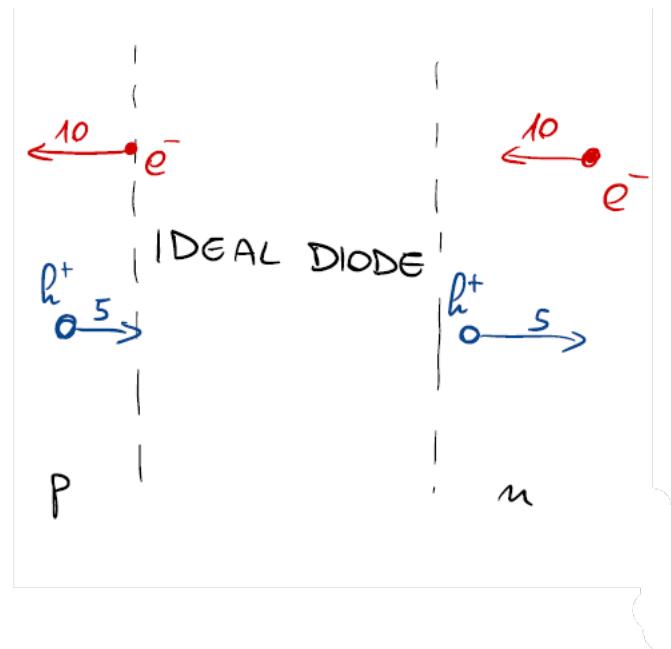
If we consider that $\gamma = 3$ and $V = 0,7 \text{ V}$ we have that $\frac{dV}{dT} = -1,9 \text{ mV/K}$

2.9 High/low-current regimes

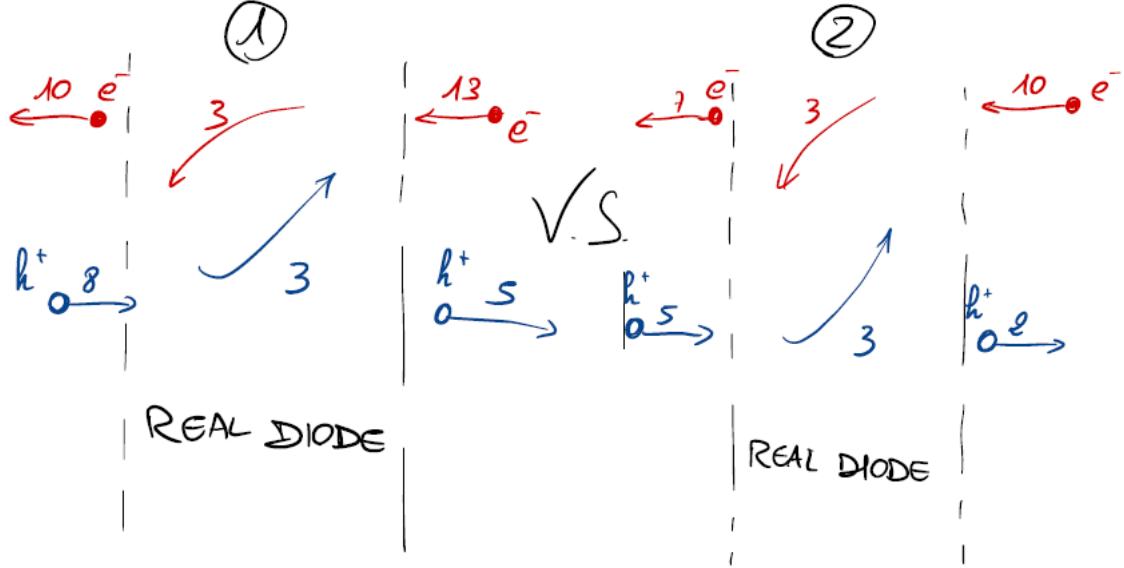
It is time we reconsider our ideal diode analysis by removing some of the approximations we have assumed; from now on we are going to assert G/R processes inside the depletion layer, which are present for sure since $E_{F_n} > E_{F_p}$.



In the ideal diode we would have the following scheme



while in a real one we are in front of two possibilities, which would be



1. Exiting carriers are constant
2. Entering carriers are constant

Just one is correct though.

Low-current regime To determine which is the one, we have to recall a consideration already computed previously: the interface between the depletion layer and the quasi neutral region is a bottleneck for minority carriers, as they must obey continuity equations. Consequentially, **exit flow has to remain constant** as more carriers are requested from the region they are a majority; therefore, (1) is the correct choice and the recombination rate inside the depletion layer, where $E_T = E_i$, looks like

$$R = \frac{n_i^2 \left(e^{\frac{qV}{KT}} - 1 \right)}{\tau_0(p + n + 2n_i)} \quad (2.27)$$

where $p = p(x)$ and $n = n(x)$. In order to have a net recombination, $V > 0$: we are searching for R_{\max} , so we need to find $\min\{p, n\}$:

$$\min\{p, n\} = \min\left\{ p + \frac{n_i^2 e^{\frac{qV}{KT}}}{p} \right\} \implies p = n_i e^{\frac{qV}{2KT}} = n \quad (2.28)$$

Thus

$$R_{\max} = \frac{n_i^2 e^{\frac{qV}{KT}}}{\tau_0 \left[2n_i e^{\frac{qV}{KT}} + 2n_i \right]} = \frac{n_i}{2\tau_0} e^{\frac{qV}{2KT}} \quad (2.29)$$

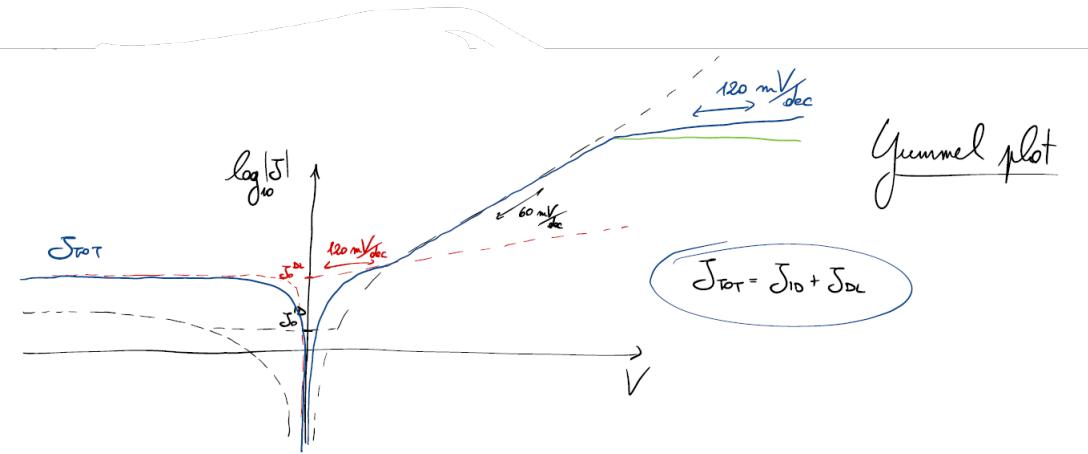
and assuming that is constant inside the depletion layer we get some additional current density

$$J_{DL} = qR_{\max}W_D = \frac{qn_i W_D}{2\tau_0} e^{\frac{qV}{2KT}} \quad (2.30)$$

If $V < 0$, $R \approx \frac{n_i}{2\tau_0}$, so $J_{DL} = \frac{qn_i W_D}{2\tau_0}$; we can safely conclude that

$$J_{DL}^{TOT} = \frac{qn_i W_D}{2\tau_0} \left(e^{\frac{qV}{2KT}} - 1 \right) \quad (2.31)$$

Dependence on V seems to be weaker.



When T reaches higher values, J_{ID} wins over the depletion layer contribution due to its dependence on n_i^2 : **this is known as the low-current regime**.

High-current regime If V is very high, we must consider a high injection condition in the quasi neutral regions, so

$$n = n_0 + \Delta n \approx \Delta n$$

$$p = p_0 + \Delta p \approx \Delta p$$

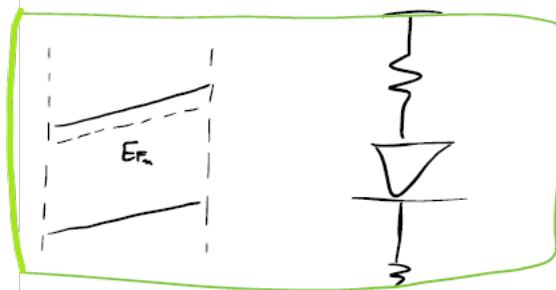
1.21 then becomes

$$pn = \Delta n_i^2 = n_i^2 e^{\frac{qV}{KT}}$$

and so

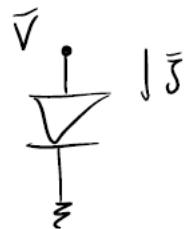
$$\Delta n = n_i e^{\frac{qV}{2KT}} \implies J \propto e^{\frac{qV}{2KT}}$$

Current density behavior resembles the one seen for low voltage; also parasitic resistance becomes relevant because E_{F_n} doesn't have a quasi-flat behavior anymore ($J_n = n\mu_n \frac{dE_{F_n}}{dx}$).



2.10 Small signal model

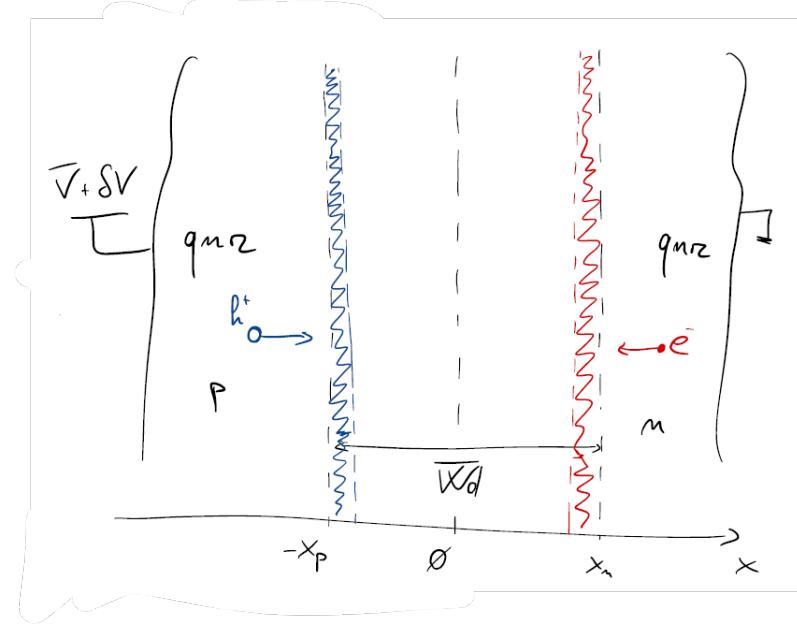
Consider a simple diode biased with \bar{V} and \bar{J} as in figure.



What would happen if we slightly change the voltage? First, we have to asses the *small signal conductance per unit area*

$$g' = \frac{\partial J}{\partial V} = \frac{J}{\frac{KT}{q}} \quad (2.32)$$

Then, if we increase the bias voltage by δV , W_D shrinks.

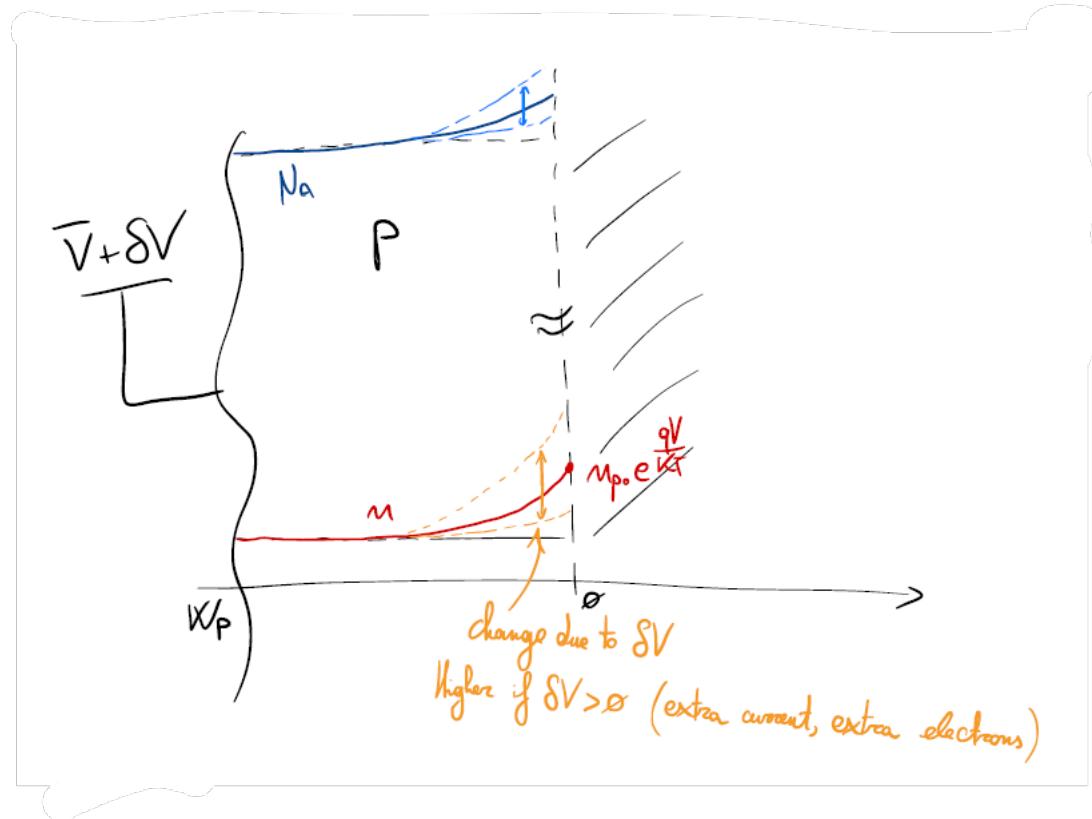


This is a capacitive effect, therefore

$$\begin{aligned}
 C_{\text{DEP}} &= -\frac{dQ_n}{dV} = -\frac{d(qN_D x_n)}{dV} = -qN_D \frac{dx_n}{dV} = -qN_D \frac{N_A}{N_A + N_D} \frac{dW_D}{dV} = \\
 &= -q \frac{1}{\frac{1}{N_A} + \frac{1}{N_D}} \sqrt{\frac{2\epsilon_{\text{Si}}}{q} \left(\frac{1}{N_A} + \frac{1}{N_D} \right) \frac{1}{2} \left(-\frac{1}{\sqrt{\phi_{\text{BI}} - V}} \right)} = \frac{q}{\frac{1}{N_A} + \frac{1}{N_D}} \frac{\sqrt{\frac{2\epsilon_{\text{Si}}}{q} \left(\frac{1}{N_A} + \frac{1}{N_D} \right)}}{\sqrt{2\sqrt{\phi_{\text{BI}} - V}}} \frac{\sqrt{\frac{2\epsilon_{\text{Si}}}{q} \left(\frac{1}{N_A} + \frac{1}{N_D} \right)}}{\sqrt{\frac{2\epsilon_{\text{Si}}}{q} \left(\frac{1}{N_A} + \frac{1}{N_D} \right)}} = \\
 &= \frac{\epsilon_{\text{Si}}}{W_D}
 \end{aligned} \tag{2.33}$$

A second capacitive term is related to the quasi-neutral region: we can define a diffusion charge as

$$Q_{\text{diff}} = \int_0^{W_p} qn(x) dx \approx \int_0^{W_p} q\Delta n(x) dx \tag{2.34}$$



Then we adjust the analysis for both wide-base and narrow-base cases.

- Wide-base:

$$Q_{\text{diff}} = \int_0^{W_p} q\Delta n(0)e^{-\frac{x_n}{L_n}} dx = q\Delta n(0)L_n \frac{L_n D_n}{L_n D_n} = J_n(0)\tau_n \quad (2.35)$$

- Narrow-base:

$$Q_{\text{diff}} = \int_0^{W_p} q\Delta n(0) \frac{W_p - x}{W_p} dx = q\Delta n(0) \frac{W_p}{2} \frac{D_n W_p}{D_n W_p} = J_n(0) \frac{W_p^2}{2D_n} = J_n(0)t_p \quad (2.36)$$

where t_p is the transit time; to make sure that this is correct, first we find the diffusion velocity

$$J_n = \frac{qD_n\Delta n(0)}{W_p} = q\Delta n(x)v_{\text{diff}} = q\Delta n(0) \frac{W_p - x}{W_p} v_{\text{diff}} \implies v_{\text{diff}} = \frac{D_n}{W_p - x} \quad (2.37)$$

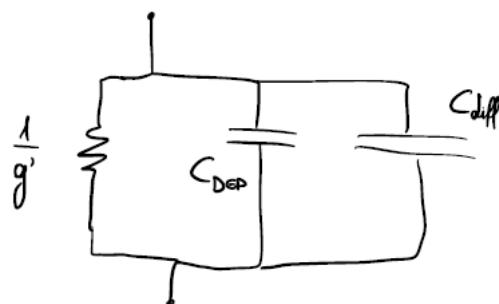
then

$$t_p = \int_0^{W_p} \frac{dx}{v_{\text{diff}}} = \int_0^{W_p} \frac{W_p - x}{D_n} dx = -\frac{(W_p - x)^2}{2D_n} \Big|_0^{W_p} = \frac{W_p^2}{2D_n} \quad (2.38)$$

Those are both balance equations, as in the first we have just rewritten G/R processes in charge terms, while in the second we have pointed out electrons travel time. Now, if we want to define a capacitance, we should consider

$$C_{\text{diff}} = \frac{dQ_{\text{diff}}}{dV} = g'\tau_n \quad (= g't_p \text{ for narrow-base diode}) \quad (2.39)$$

Our model in forward bias would then appear like a resistance in parallel with two capacitances.



Under reverse bias though, C_{DEP} would be dominant over the others.

Chapter 3

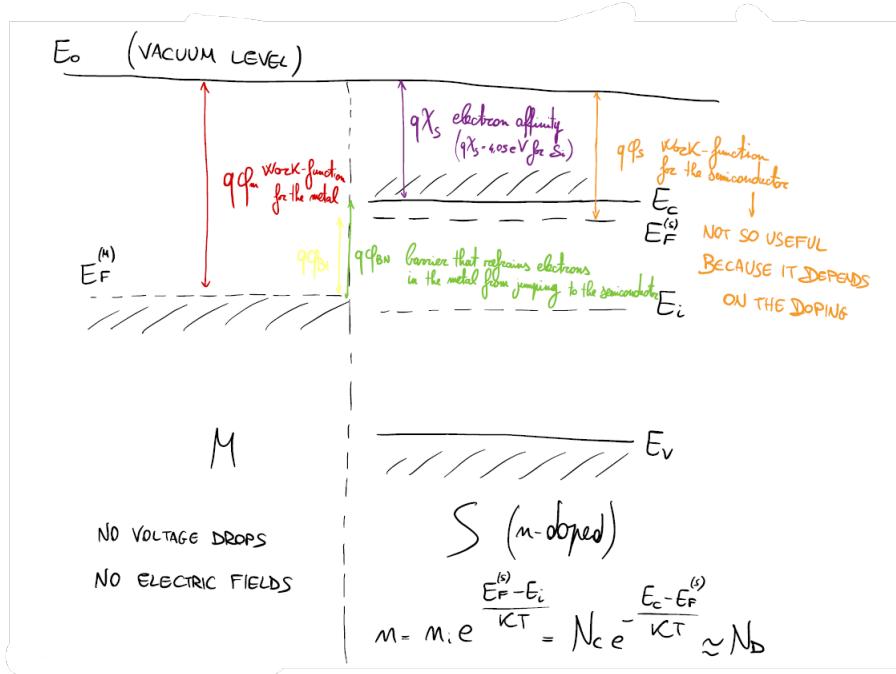
M-S junction

3.1 Basics of the M-S junction

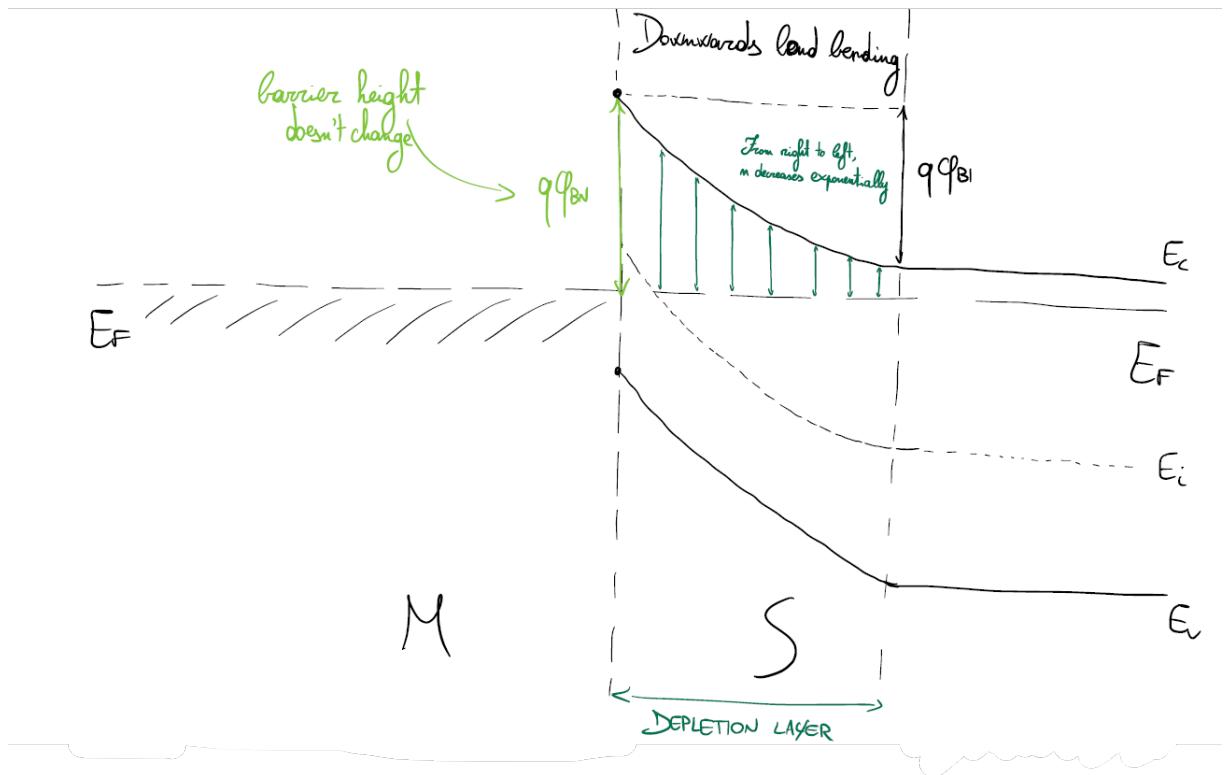
Metal-Semiconductor can be design to be either a rectifying device or an ohmic contact.

Rectifying device (Schottky diode)	Ohmic contact (ideal contact)
<ul style="list-style-type: none"> - Relatively low-doped ($N < 10^{17} \text{ cm}^{-3}$) - Highly resistive - Proper work-function of the metal 	<ul style="list-style-type: none"> - Very high-doped ($N > 10^{19} \text{ cm}^{-3}$) - Lowly resistive (strong current flow with a small voltage drop) - Proper work-function of the metal

Let's investigate the device behavior under thermodynamic equilibrium, separated materials first.



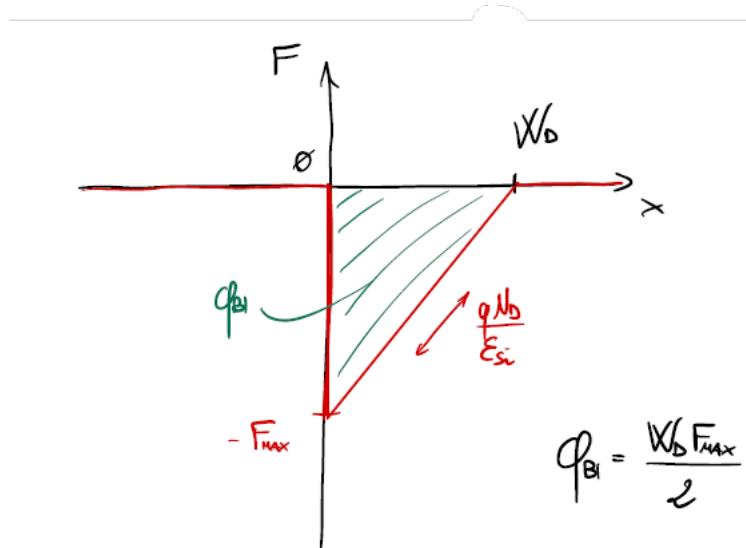
For a Schottky diode, $E_F^{(M)}$ must be below $E_F^{(S)}$ in order to have a rectifying behavior. Now, if we put the materials in contact, we sure notice a band bending which is going to be downwards because $q\phi_{BN}$ does not change.



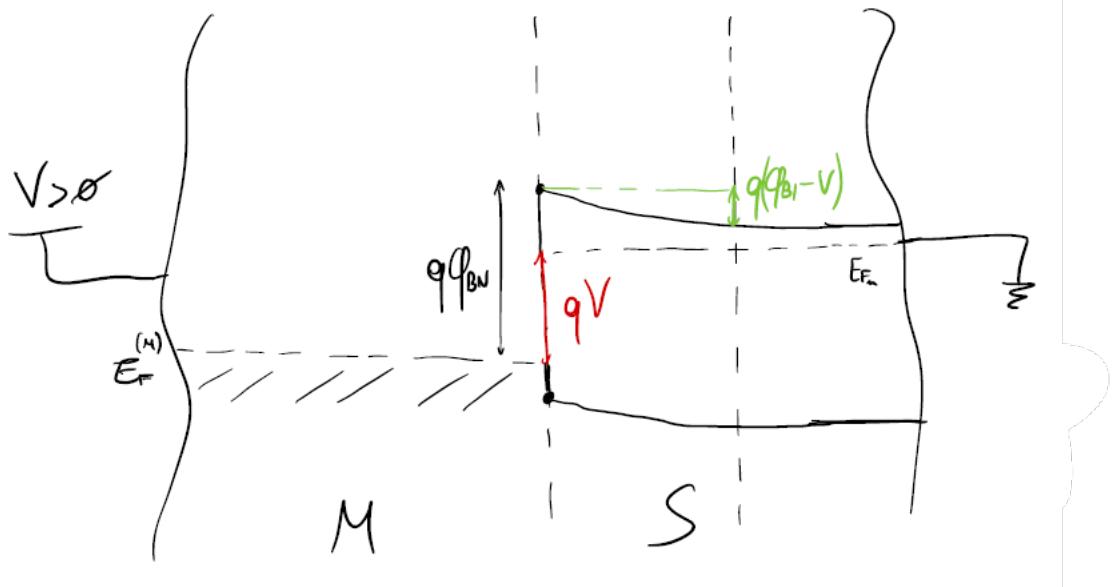
We can write a powerful definition for that barrier:

$$q\phi_{BN} = q(\phi_m - \chi_s) = q\phi_{BI} + KT \log \left(\frac{N_c}{N_D} \right) \quad (3.1)$$

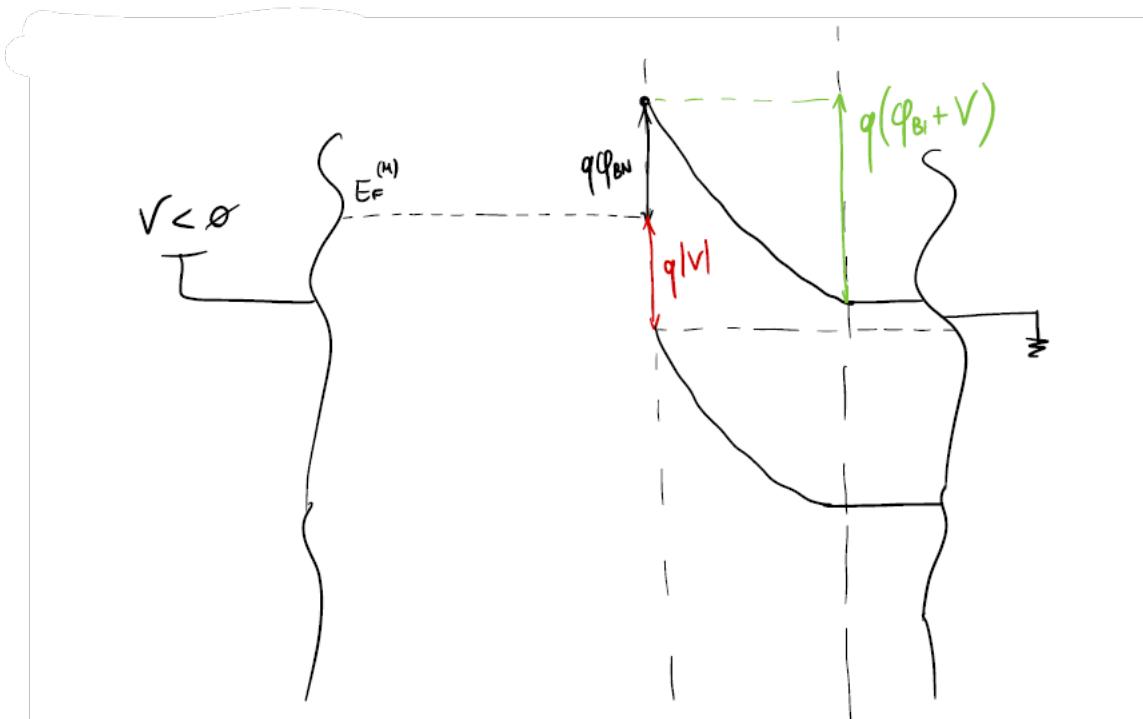
We have been knowing $q\phi_{BI}$ since 2.14, this time considering the Schottky diode as a unilateral p-n junction.



It is time to apply both forward and reverse bias. In the first case we notice that the total band bending decreases, $E_F^{(S)}$ is almost flat in the quasi-neutral region and W_D shrinks.



In the second one, all is the opposite of course.

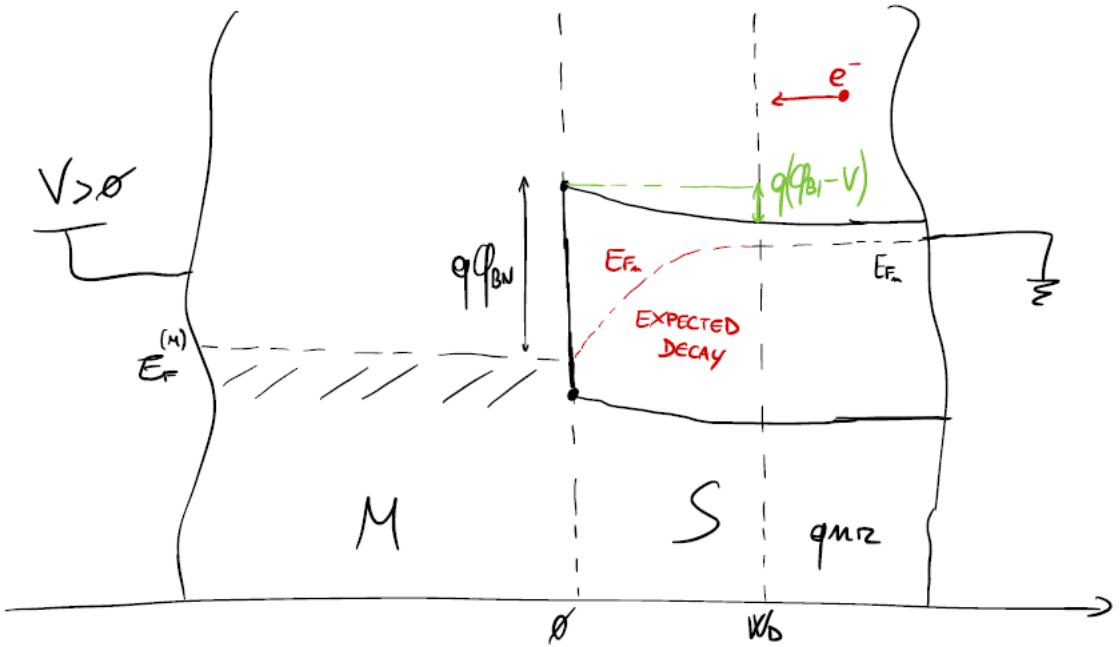


3.2 M-S junction as a Schottky diode

Here are presented a few models on how current flow works inside a Schottky diode.

3.2.1 Current transport: Schottky's model

We want to study how current flows inside the device under forward bias. In this case, electrons flow is not limited by the region where they would be minority carriers as if we had a p-n junction, because *they are always a majority*; this results in the absence of the well-known bottleneck region, therefore we can safely say that **this is a majority carriers device**, where G/R processes have no significant impact (N.B.: p-n junction is a minority carriers device). We have holes as well, but since they are concentrated in the *n*-region, they are negligible.



Flow constraint is localized entirely within the depletion layer because of the scattering, so

$$J_n = qn\mu_n F + qD_n \frac{dn}{dx} = \text{constant} \quad (3.2)$$

Assuming the following boundary conditions

$$\begin{cases} n(W_D) = N_D \\ n(0) = N_c e^{-\frac{E_c - E_F}{KT}} = e^{-\frac{q\phi_{BN}}{KT}} \end{cases}$$

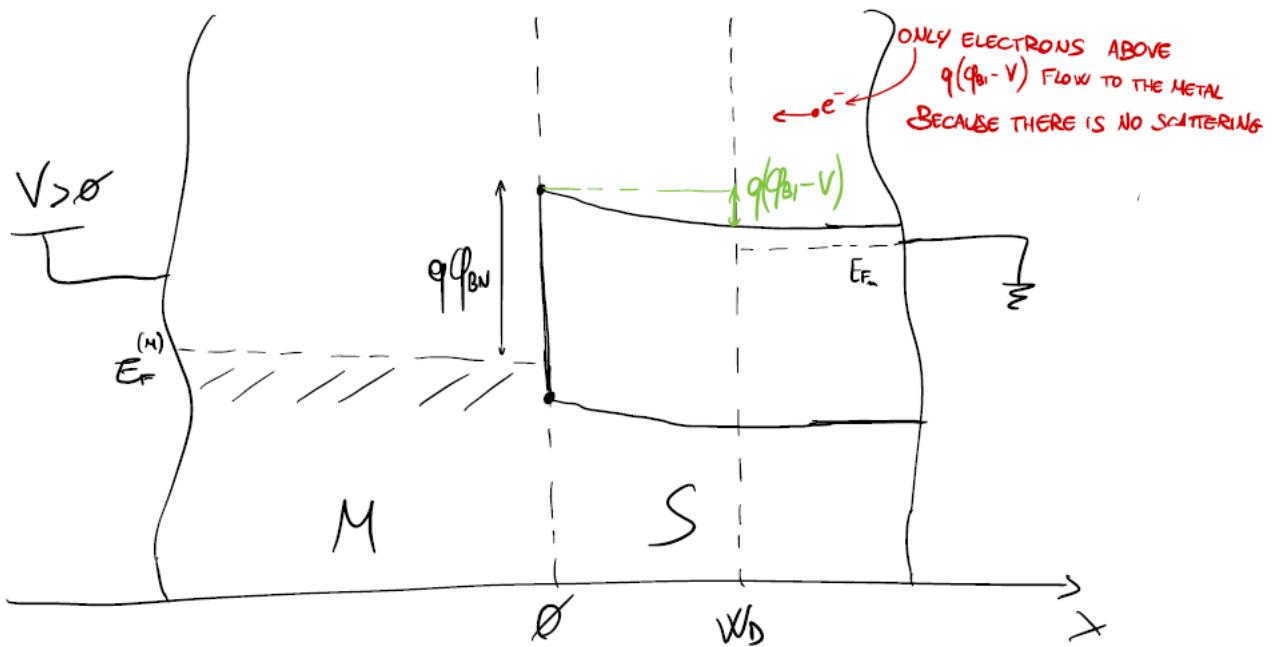
in the end we achieve the expected rectifying behavior

$$J_n(V) = J_0 \left(e^{\frac{qV}{KT}} - 1 \right) = J_S \quad (3.3)$$

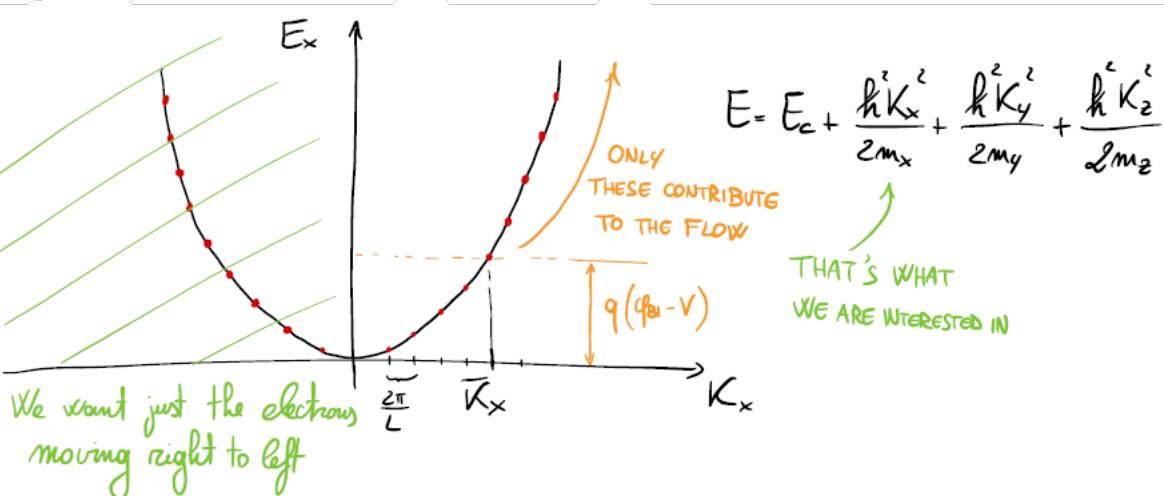
but its prefactor is not the same from the p-n junction! In fact, this equation works fine just for low-mobility semiconductors, while for high-mobility ones (Si, Ge, GaAs), J_0 doesn't show the correct dependencies on temperature and voltage. This is a pure drift/diffusion model, also known as the Schottky's model, and an explanation of why it is not correct for those kind of semiconductors is needed. That happens because at the M-S interface we kept the condition that $E_F^{(M)}$ must coincide with E_{F_n} , but in high-mobility semiconductors the electrons flow towards the metal is enormous, therefore thermodynamic equilibrium is not assured to be restored as electrons concentration may increase as well; electrons flow through the interface, then, cannot be described by drift/diffusion equation because that region is too narrow.

3.2.2 Current transport: Bethe's model

We have seen how in Schottky's model the interface is not handled well, so a new model is required. If we assume that only electrons above the barrier can flow through metal, we are handling correctly the interface behavior while completely neglecting the scattering inside the depletion layer: that is Bethe's model, a pure thermionic one.



Let's recall some Quantum Physics.



Current density from the first available energy level can be found as

$$J_{S \rightarrow M} = \sum_{k_x > \bar{k}_x} \frac{2q}{L^3} v_x(k_x) \mathbf{f}(k_x, k_y, k_z) = \frac{1}{(\frac{2\pi}{L})^3} \int_{\bar{k}_x}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{2q}{L^3} v_x(k_x) \mathbf{f}(k_x, k_y, k_z) dk_x dk_y dk_z$$

We can switch from discrete summation to integration because energy levels are very tight. To simplify calculations we can substitute

$$\begin{cases} \hbar k_x = m_x v_x \\ \hbar k_y = m_y v_y \\ \hbar k_z = m_z v_z \end{cases} \implies \begin{cases} dk_x = \frac{m_x}{\hbar} dv_x \\ dk_y = \frac{m_y}{\hbar} dv_y \\ dk_z = \frac{m_z}{\hbar} dv_z \end{cases}$$

thus

$$\begin{aligned}
J_{S \rightarrow M} &= \frac{2q}{(2\pi)^3} \int_{\bar{v}_x}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} v_x \mathbf{f}(v_x, v_y, v_z) \frac{m_x m_y m_z}{\hbar} dv_x dv_y dv_z = \\
&= \frac{2q}{(2\pi)^3} \frac{m_x m_y m_z}{\hbar} \int_{\bar{v}_x}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} v_x e^{-\frac{E_c - E_{F_n}}{KT}} e^{-\frac{m_x v_x^2}{2KT}} e^{-\frac{m_y v_y^2}{2KT}} e^{-\frac{m_z v_z^2}{2KT}} dv_x dv_y dv_z = \\
&= \frac{2q}{\hbar^3} m_x m_y m_z e^{-\frac{E_c - E_{F_n}}{KT}} \int_{\bar{v}_x}^{+\infty} v_x e^{-\frac{m_x v_x^2}{2KT}} dv_x \int_{-\infty}^{+\infty} e^{-\frac{m_y v_y^2}{2KT}} dv_y \int_{-\infty}^{+\infty} e^{-\frac{m_z v_z^2}{2KT}} dv_z
\end{aligned}$$

The three integrals result in

$$\begin{aligned}
\int_{\bar{v}_x}^{+\infty} v_x e^{-\frac{m_x v_x^2}{2KT}} dv_x &= \frac{KT}{m_x} e^{-\frac{m_x \bar{v}_x^2}{2KT}} \\
\int_{-\infty}^{+\infty} e^{-\frac{m_y v_y^2}{2KT}} dv_y &= \sqrt{\frac{2\pi KT}{m_y}} \\
\int_{-\infty}^{+\infty} e^{-\frac{m_z v_z^2}{2KT}} dv_z &= \sqrt{\frac{2\pi KT}{m_z}}
\end{aligned}$$

and so

$$\begin{aligned}
J_{S \rightarrow M} &= \frac{4\pi q \textcolor{red}{m_0}}{\hbar^3} \frac{\sqrt{m_y m_z}}{\textcolor{red}{m_0}} (KT)^2 e^{-\frac{E_c - E_{F_n}}{KT}} e^{-\frac{m_x \bar{v}_x^2}{2KT}} = A \frac{\sqrt{m_y m_z}}{m_0} T^2 e^{-\frac{E_c - E_{F_n}}{KT}} e^{-\frac{m_x \bar{v}_x^2}{2KT}} = \\
&= A \frac{\sqrt{m_y m_z}}{m_0} T^2 e^{-\frac{q(\phi_{BN} + V)}{KT}}
\end{aligned} \tag{3.4}$$

where A is the Richardson constant. After some Quantum Physics considerations we can rearrange the mass term

$$J_{S \rightarrow M} = A \frac{2m_t + 4\sqrt{m_t m_i}}{m_0} T^2 e^{-\frac{q(\phi_{BN} + V)}{KT}} = A^* T^2 e^{-\frac{q(\phi_{BN} + V)}{KT}} \tag{3.5}$$

and similarly we can consider the electron flow from the metal to the semiconductor

$$J_{M \rightarrow S} = A^* T^2 e^{-\frac{q\phi_{BN}}{KT}} \tag{3.6}$$

Total current density would be

$$J_{TOT} = J_{S \rightarrow M} - J_{M \rightarrow S} = A^* T^2 e^{-\frac{q\phi_{BN}}{KT}} \left(e^{\frac{qV}{KT}} - 1 \right) = J_{0,TH} \left(e^{\frac{qV}{KT}} - 1 \right) \tag{3.7}$$

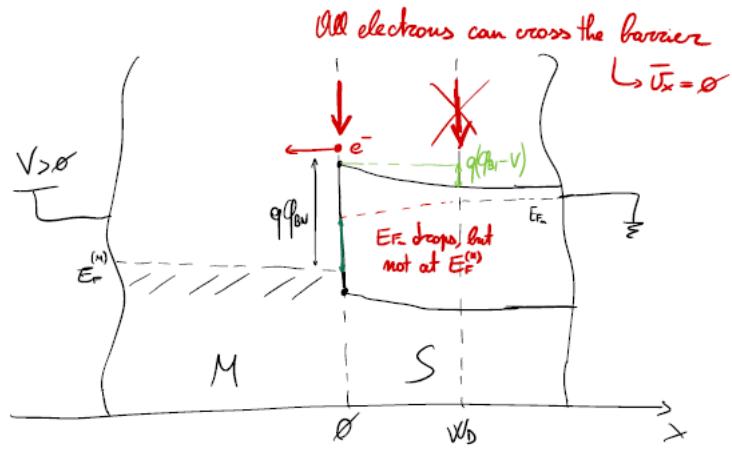
What about the scattering inside the depletion layer?

3.2.3 Current transport: Thermionic/diffusion model

We need to blend together the previous models: let's start with a comparison.

Schottky model	Bethe model
- Pure drift/diffusion	- Pure thermionic
- Transport in the depletion layer	- Transport in the depletion layer
- M-S interface	- M-S interface

We must take the **best aspects** and mix them together to obtain the best suitable model possible. Let's start considering that all the electrons can cross the barrier.



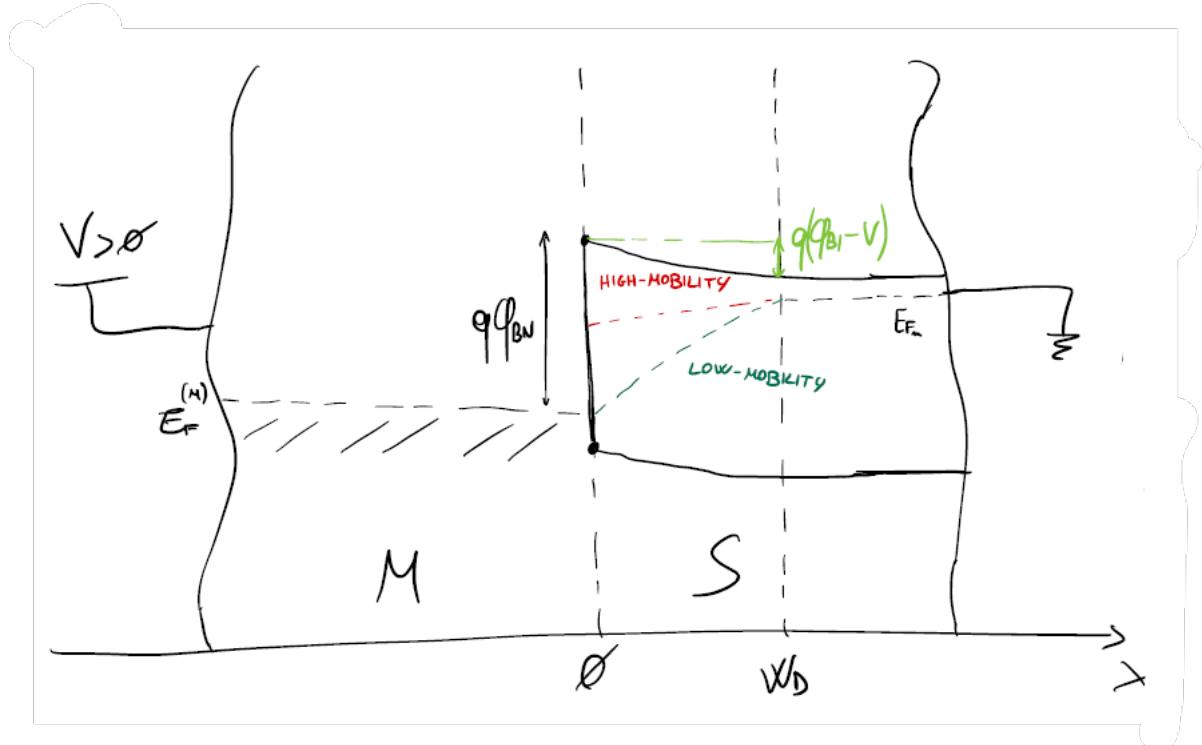
Let's now assume that current density remains constant and that $n(W_D) \approx N_D$; at the interface we get

$$\begin{cases} J_{S \rightarrow M} = A^* T^2 e^{-\frac{E_c - E_{F,n}}{KT}} e^{-\frac{m_e v_x}{2KT}} \frac{N_c}{N_c} = \frac{A^* T^2}{N_c} n(0) \\ J_{M \rightarrow S} = \frac{A^* T^2}{N_c} n_0 \end{cases} \implies J_{th} = \frac{A^* T^2}{N_c} (n(0) - n_0) \quad (3.8)$$

We finally arrive to the condition we were looking for:

$$J_n(0) = J_{th} \quad (3.9)$$

To sum it up, in Schottky's model the bottleneck is represented by the depletion layer, while in Bethe's one it is the interface between the metal and the semiconductor.

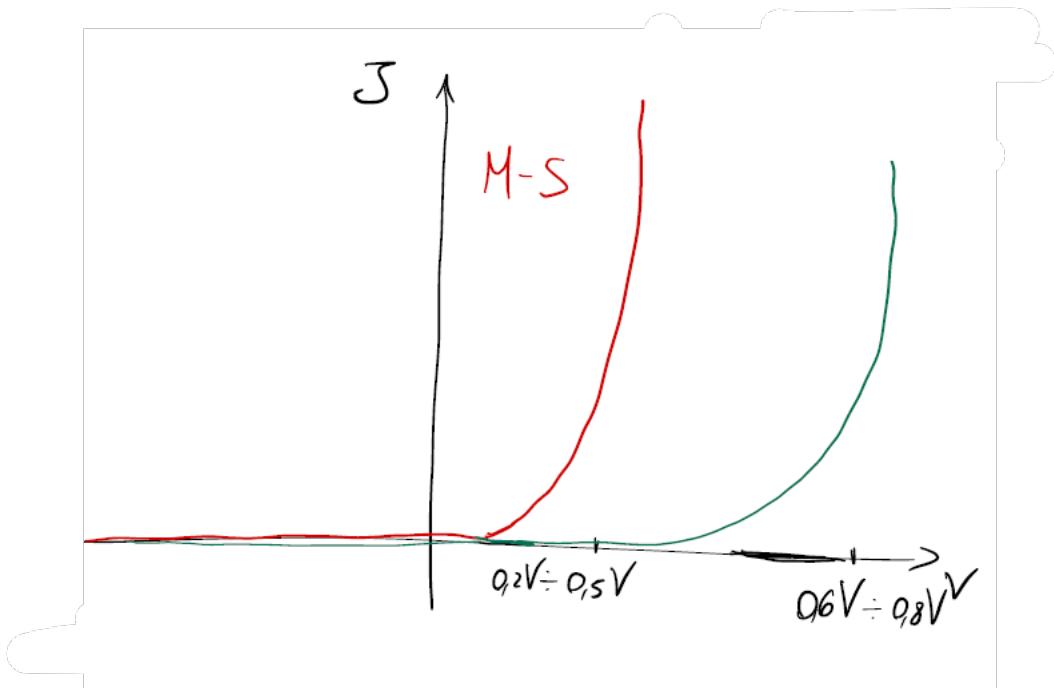


Now we put some numbers for the Si case (we use Bethe's model), so

$$J_n \approx J_{th} = A^* T^2 e^{-\frac{q\phi_{BN}}{KT}} \left(e^{\frac{qV}{KT}} - 1 \right)$$

$$q\phi_{BN} = 0,81 \text{ eV}$$

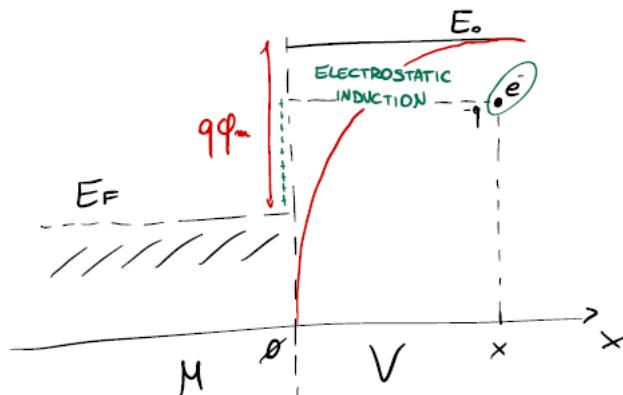
$$J_{0,th} = 5 \cdot 10^{-7} \text{ A/cm}^2 \gg J_{0,pn}$$



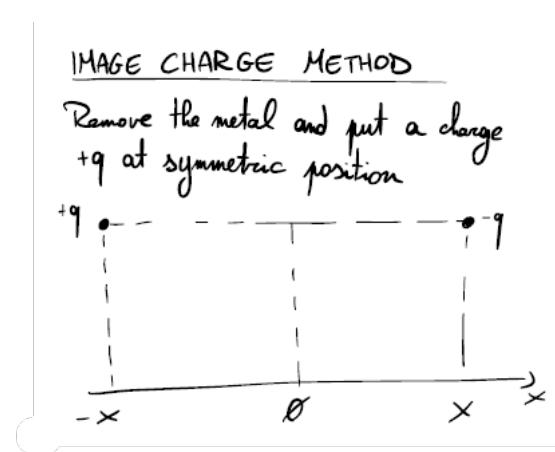
The M-S junction has also a better frequency response.

3.2.4 Schottky effect

Equation 3.7 is good, but it still needs some adjustments due to Schottky effect, a pure electrostatic one. Let's begin considering a metal-vacuum interface, with a negative charge positioned in the vacuum; this charge would attract some positive charges on the metal surface for electrostatic induction, causing a band bending of the vacuum level E_0 .



Analytically, we can come to this conclusion through the **image charge method**.



$$-qF = \frac{1}{4\pi\epsilon_0} \frac{-q^2}{(2x)^2}$$

$$F = \frac{1}{16\pi\epsilon_0} \frac{q}{x^2} = -\frac{d\phi}{dx}$$

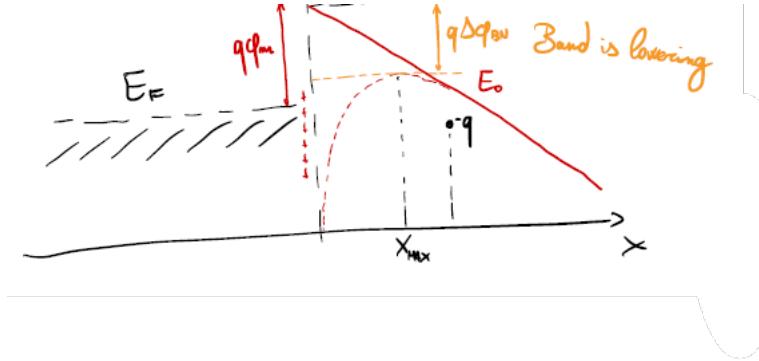
$$\int_{\phi(x)}^{\phi(+\infty)} d\phi = \int_x^{+\infty} -\frac{1}{16\pi\epsilon_0} \frac{q}{x^2} dx$$

$$\phi(+\infty) - \phi(x) = \left. \frac{1}{16\pi\epsilon_0} \frac{q}{x} \right|_x^{+\infty}$$

$$\phi(x) = \phi(+\infty) + \frac{1}{16\pi\epsilon_0} \frac{q}{x}$$

Remember that $\phi = -\frac{E_0}{q}$, so

$$E_0(x) = E_0(+\infty) - \frac{1}{16\pi\epsilon_0} \frac{q^2}{x} \quad (3.10)$$



If we consider $E_0(+\infty) = E_0(0) - qFx$ we can further develop 3.10 as

$$E_0(x) = E_0(0) - qFx - \frac{1}{16\pi\epsilon_0} \frac{q^2}{x} \quad (3.11)$$

As we can see from the previous picture, E_0 bends with a maximum in x_{MAX} ; to find that, we have to compute the derivative

$$\frac{dE_0}{dx} = -qF - \frac{1}{16\pi\epsilon_0} \frac{q^2}{x^2} = 0 \quad \Rightarrow \quad x_{MAX} = \sqrt{\frac{q}{16\pi\epsilon_0 F}} \quad (3.12)$$

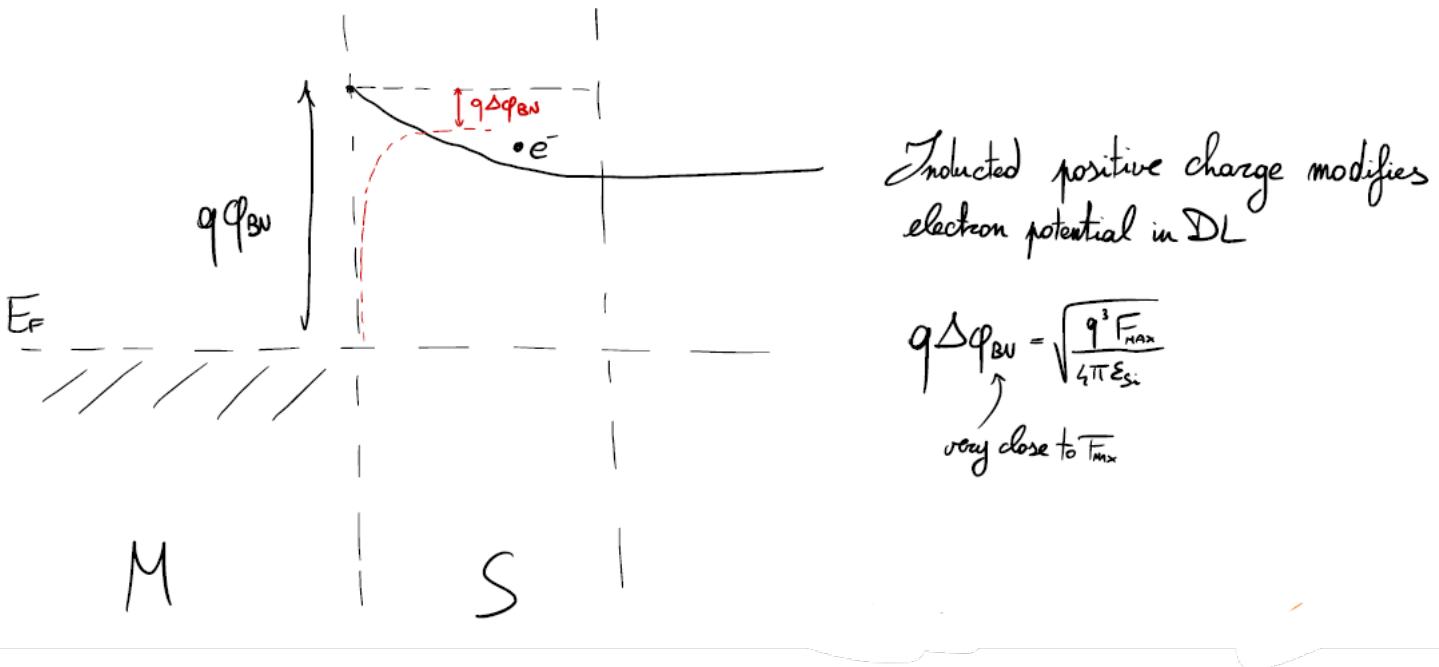
All we are left to do is to find how much the band bending is:

$$E_0(x_{MAX}) = E_0(0) - qF \sqrt{\frac{q}{16\pi\epsilon_0 F}} - \frac{1}{16\pi\epsilon_0} \frac{q^2 \sqrt{16\pi\epsilon_0 F}}{\sqrt{q}} = E_0(0) - \sqrt{\frac{q^3 F}{16\pi\epsilon_0}} - \sqrt{\frac{q^3 F}{16\pi\epsilon_0}} = E_0(0) - \sqrt{\frac{q^3 F}{4\pi\epsilon_0}} \quad (3.13)$$

It is trivial that

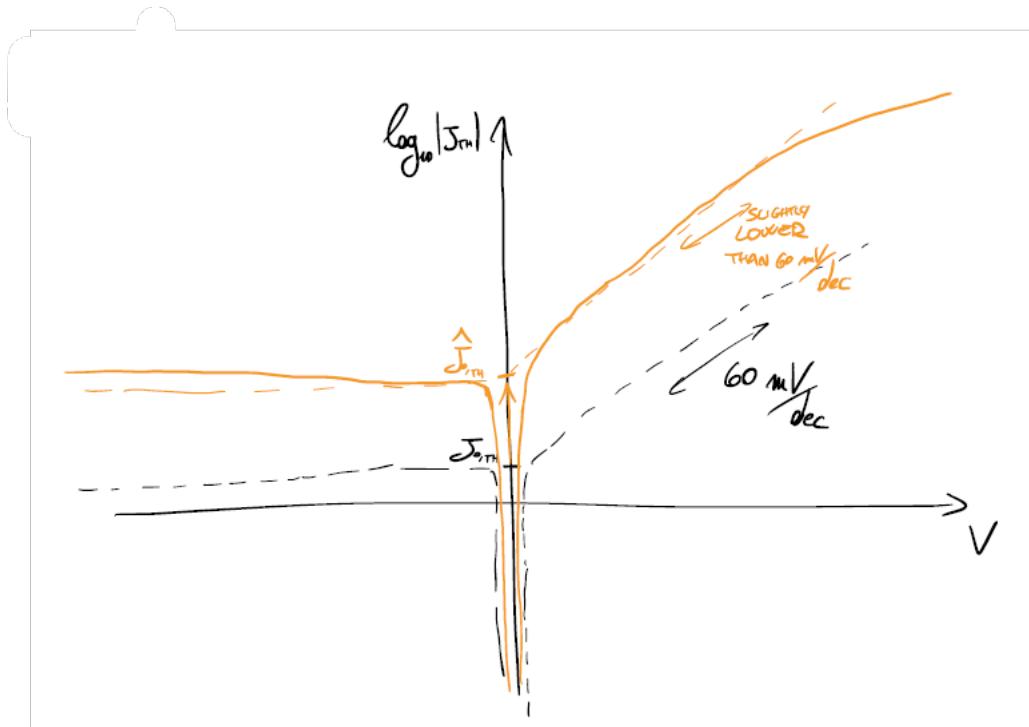
$$q\Delta\phi_{BN} = \sqrt{\frac{q^3 F}{4\pi\epsilon_0}} \quad (3.14)$$

Eventually we apply this result to the actual M-S junction.



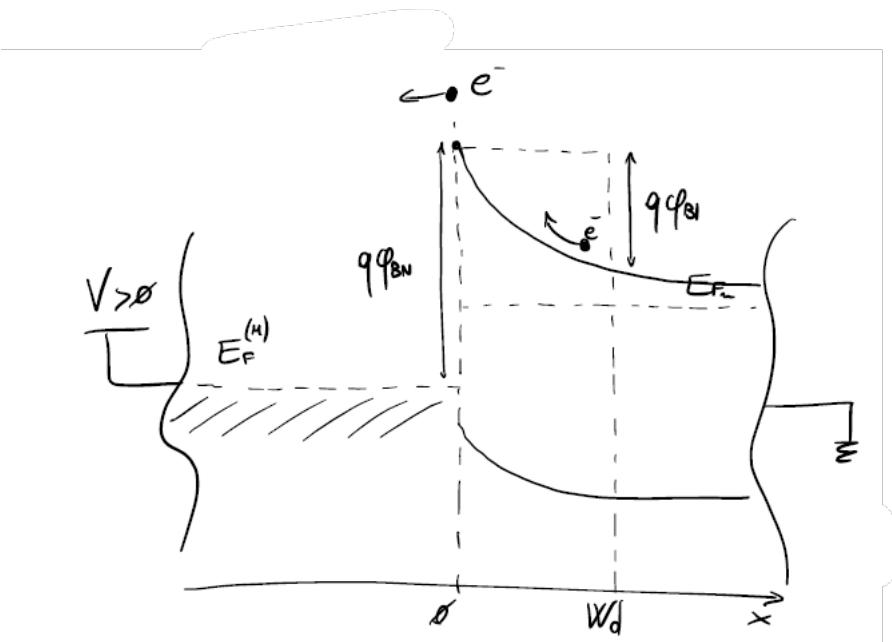
Current density prefactor depends also on the voltage now:

$$J_{\text{TH}} = A^* T^2 e^{-\frac{q(\phi_{\text{BN}} - \Delta\phi_{\text{BN}})}{kT}} \left(e^{\frac{qV}{kT}} - 1 \right) = \hat{J}_{0,\text{TH}}(V) \left(e^{\frac{qV}{kT}} - 1 \right) \quad (3.15)$$

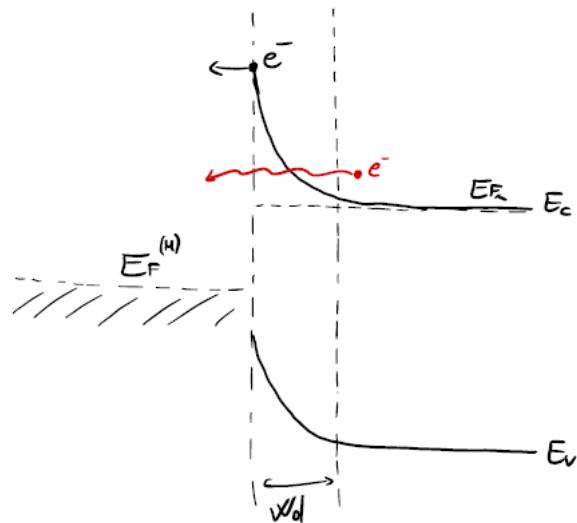


3.3 M-S junction as an ohmic contact

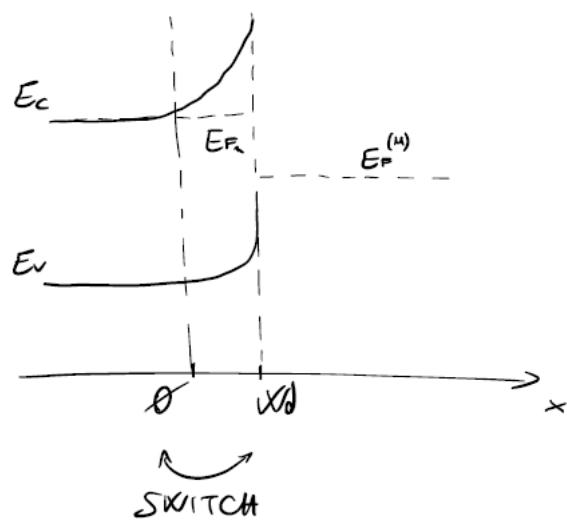
Our junction is now operating as an ohmic contact, which is a crucial aspect for every Integrated Circuit. As we have already seen, current flow is limited by thermionic emission at the interface.



This time, though, W_D is thinner due to high-level doping. Because of that, it is not impossible that **electrons begin to travel through the barrier by tunneling effect**.



We can calculate tunneling probability exploiting WKB approximation and considering $E = E_{F_n} = 0$ as our reference level.



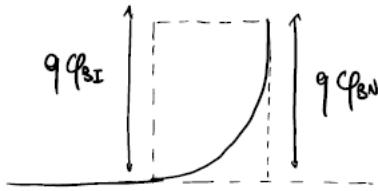
$$\begin{cases} E_c(x) = \frac{q^2 N_D}{2\epsilon_{Si}} x^2 \\ T = \exp \left(-2 \int_0^{W_D} \sqrt{\frac{2m^*(E_c - E)}{\hbar^2}} dx \right) \end{cases} \quad (3.16)$$

Developing the system we obtain

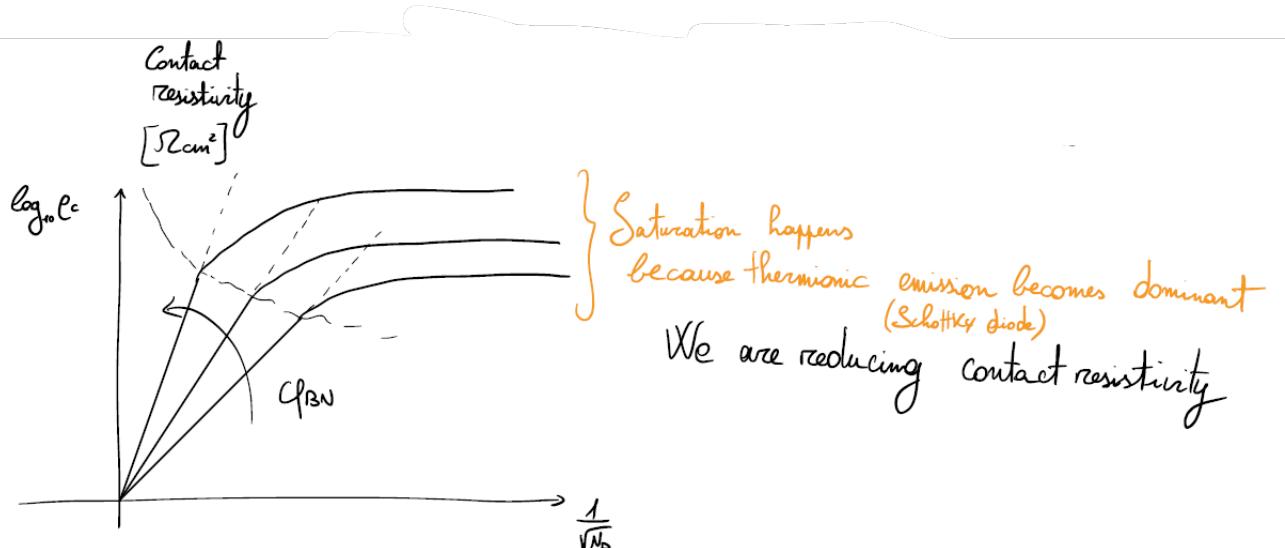
$$\begin{aligned} T &= \exp \left(-2 \int_0^{W_D} \sqrt{\frac{2m^* q^2 N_D}{2\epsilon_{Si}} x^2} dx \right) = \exp \left(-2 \sqrt{\frac{2m^* q^2 N_D}{2\epsilon_{Si}}} \frac{W_D^2}{2} \right) = \\ &= \exp \left(-2 \sqrt{\frac{2m^* q^2 N_D}{2\epsilon_{Si}}} \frac{1}{2} \frac{2\epsilon_{Si}}{q} \frac{1}{N_D} (\phi_{Bi} - V) \right) = \exp \left(-\frac{q(\phi_{Bi} - V)}{E_{00}} \right) \end{aligned} \quad (3.17)$$

where $E_{00} = \frac{qh\sqrt{N_D}}{4\pi\sqrt{m^*\epsilon_{Si}}}$; from Quantum Physics we know that current density is proportional to the transmission probability ($J \propto T$), so

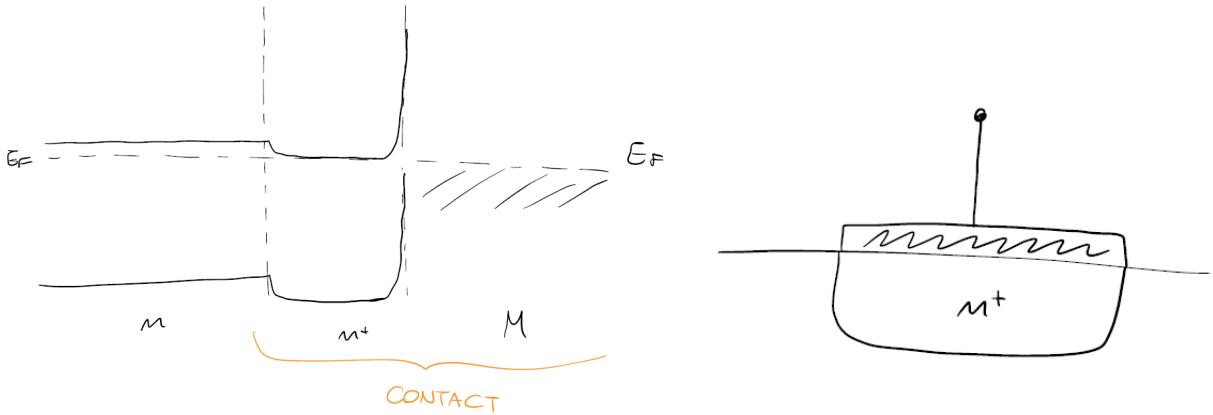
$$\rho_c = \left(\frac{\partial J}{\partial V} \right)^{-1} \Big|_{V=0} \propto \left(e^{-\frac{q(\phi_{Bi} - V)}{E_{00}}} \right)^{-1} \Big|_{V=0} = \frac{E_{00}}{q} e^{\frac{q\phi_{Bi}}{E_{00}}} = \frac{E_{00}}{q} e^{\frac{q\phi_{BN}}{E_{00}}} \quad (3.18)$$



Since $E_{00} \propto N_D$, contact resistivity decreases when the doping is very high.

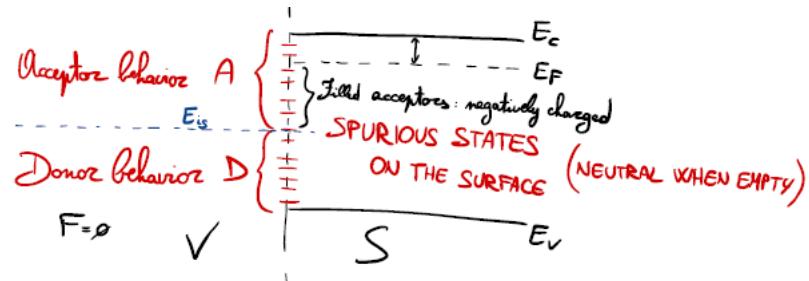


A real contact looks like a n^+ region coupled with a metal.

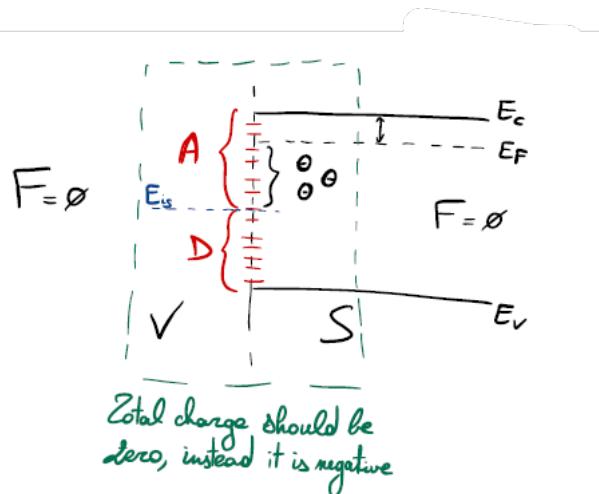


3.4 Interface states

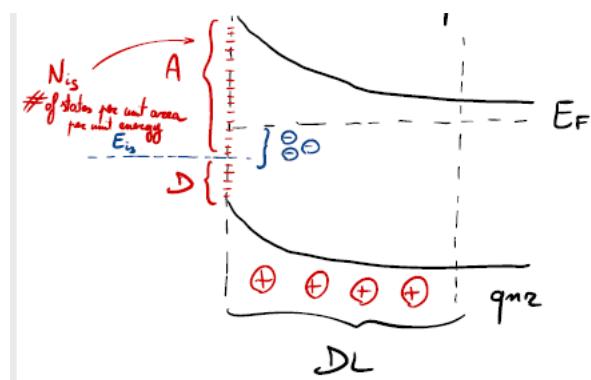
As we know, silicon has a lot of imperfections and impurities, especially on its border to the vacuum: spurious states are introduced inside the energy gap, breaking potential periodicity.



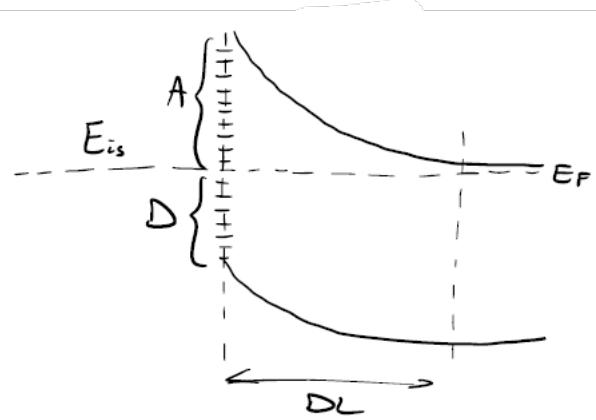
This is a huge problem, because now we have negative charges in the gap nullifying for Gauss law the band diagrams we have considered until here.



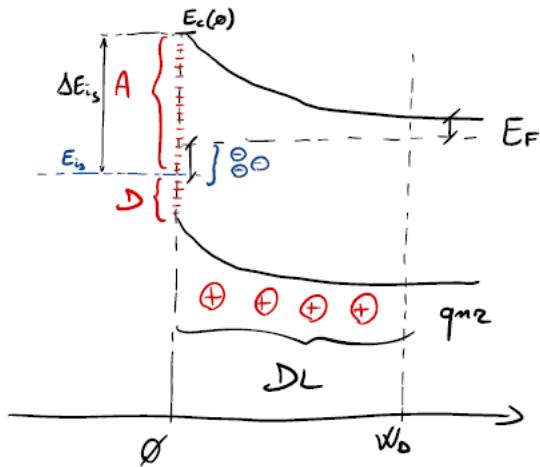
To fix this, bands must inevitably move upwards creating a depletion layer.



It is important to notice that the depletion layer grows when N_{is} is very high.



What is the upper limit for E_{is} ? If we consider $N_{is} \rightarrow +\infty$, it is impossible that $E_{is} > E_F$ as it would result in the formation of positive charges outside of the depletion layer, therefore $E_{is} = E_F$ is our limiting case¹. Assuming then N_{is} constant in the energy gap, it is important to find the charge on the interface.



$$|Q_{is}| = qN_{is}(E_F - E_{is}) = qN_{is}[\Delta E_{is} - (E_c(0) - E_F)] \quad (3.19)$$

$$Q_{DEP} = qN_D W_D = \sqrt{2\epsilon_{Si}qN_D \frac{E_c(0) - E_c(W_D)}{q}} \quad (3.20)$$

In order to have a perfect balance, $|Q_{is}| = Q_{DEP}$ must be imposed, so

$$qN_{is}[\Delta E_{is} - (E_c(0) - E_F)] = \sqrt{2\epsilon_{Si}qN_D \frac{E_c(0) - E_c(W_D)}{q}} \quad (3.21)$$

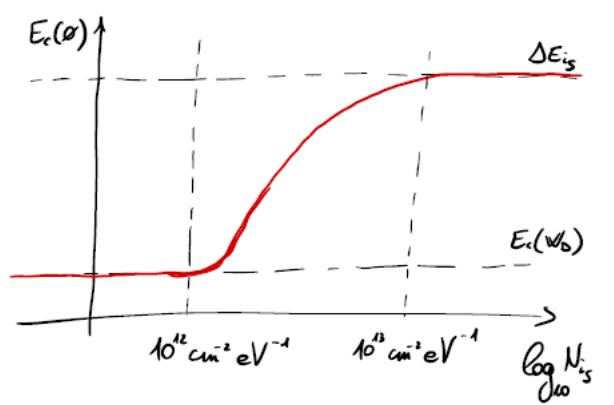
When $N_{is} = 0$, $E_c(0) = E_c(W_D)$

When $N_{is} \rightarrow +\infty$, $E_c(0) \approx \Delta E_{is}$

Everything makes perfect sense. We can even rewrite $E_c(0)$ in a more general form:

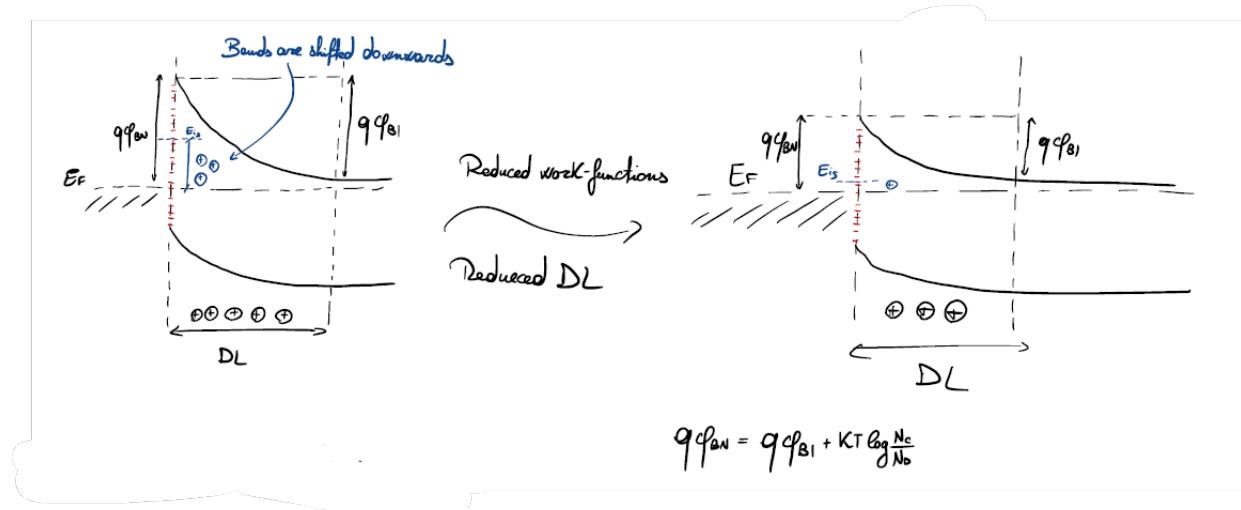
$$E_c(0) = \Delta E_{is} + \frac{\epsilon_{Si}N_D}{q^2N_{is}^2} - \sqrt{\left(\frac{\epsilon_{Si}N_D}{q^2N_{is}^2}\right)^2 + \frac{2\epsilon_{Si}N_D}{q^2N_{is}^2}(\Delta E_{is} - E_c(W_D))} \quad (3.22)$$

¹Notice that flat band condition is compatible only if $N_{is} = 0$



When N_{is} is very large, E_F position is NOT set by the doping concentration, but only by N_{is} itself; this condition is called **Fermi level pinning**.

If we apply these results to the proper M-S junction we would notice that $q\phi_{BN}$ is modified, so bands suffer from a downwards shift.



Here is a tabular reporting some theoretical and actual values of $q\phi_{BN}$; we can clearly see the Fermi level pinning.

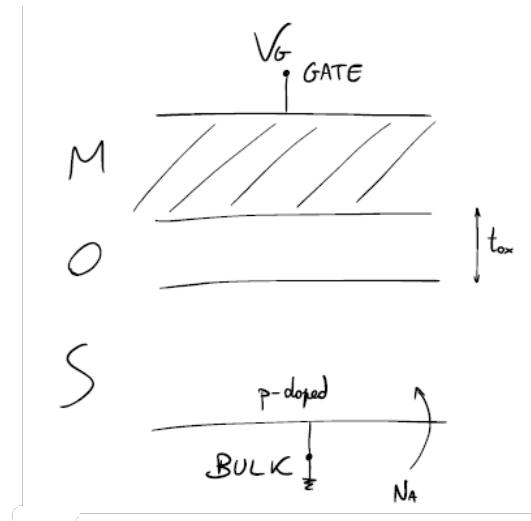
$q\phi_{BN}$	Al-nSi	Au-nSi	Pt-nSi
Theoretically	0,05 eV	0,75 eV	1,25 eV
Empirically	0,81 eV	0,83 eV	0,9 eV

Chapter 4

MOS capacitor

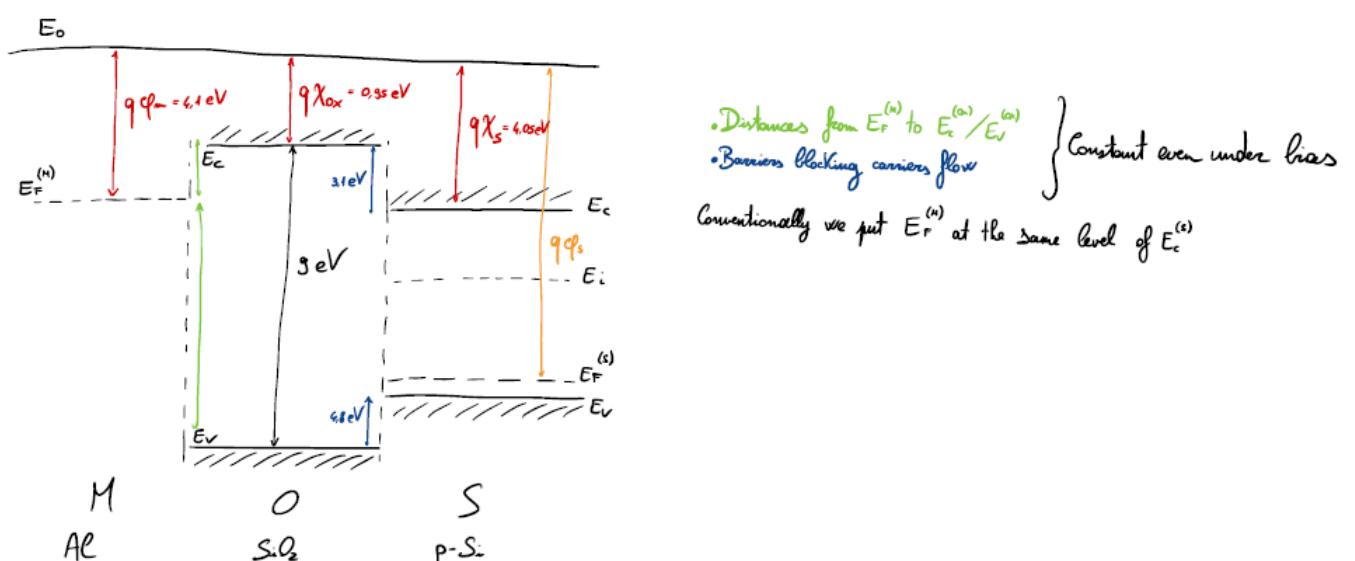
4.1 Basics of the MOS capacitor

The MOS capacitor is the basic element for the CMOS technology; its structure consists in a gate material (typically a metal or a highly doped polycrystalline/amorphous silicon, but not a monocrystalline silicon as *Si* atoms would bond just with the oxide compromising spatial arrangement), which guarantees a high free-carriers density, a thin insulator (typically SiO_2 , which allowed the birth of planar processes) and a base uniformly doped.



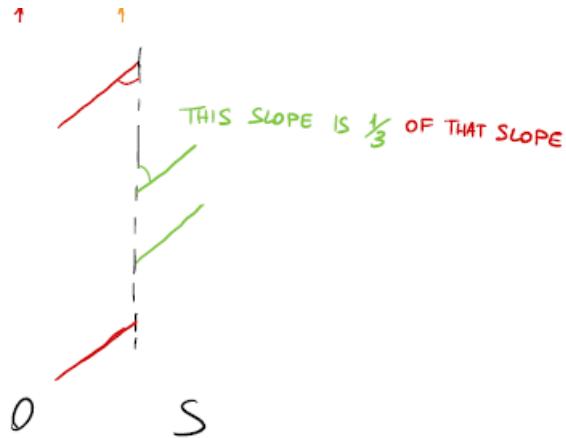
The insulator is introduced to provide a barrier blocking the carriers flow between metal and semiconductor; it is important to underline that high-quality insulators have low spurious states and a high dielectric constant is needed to have a good coupling.

As always, we are interested in the electrostatics: let's see what is like when materials are separated under thermodynamic equilibrium.

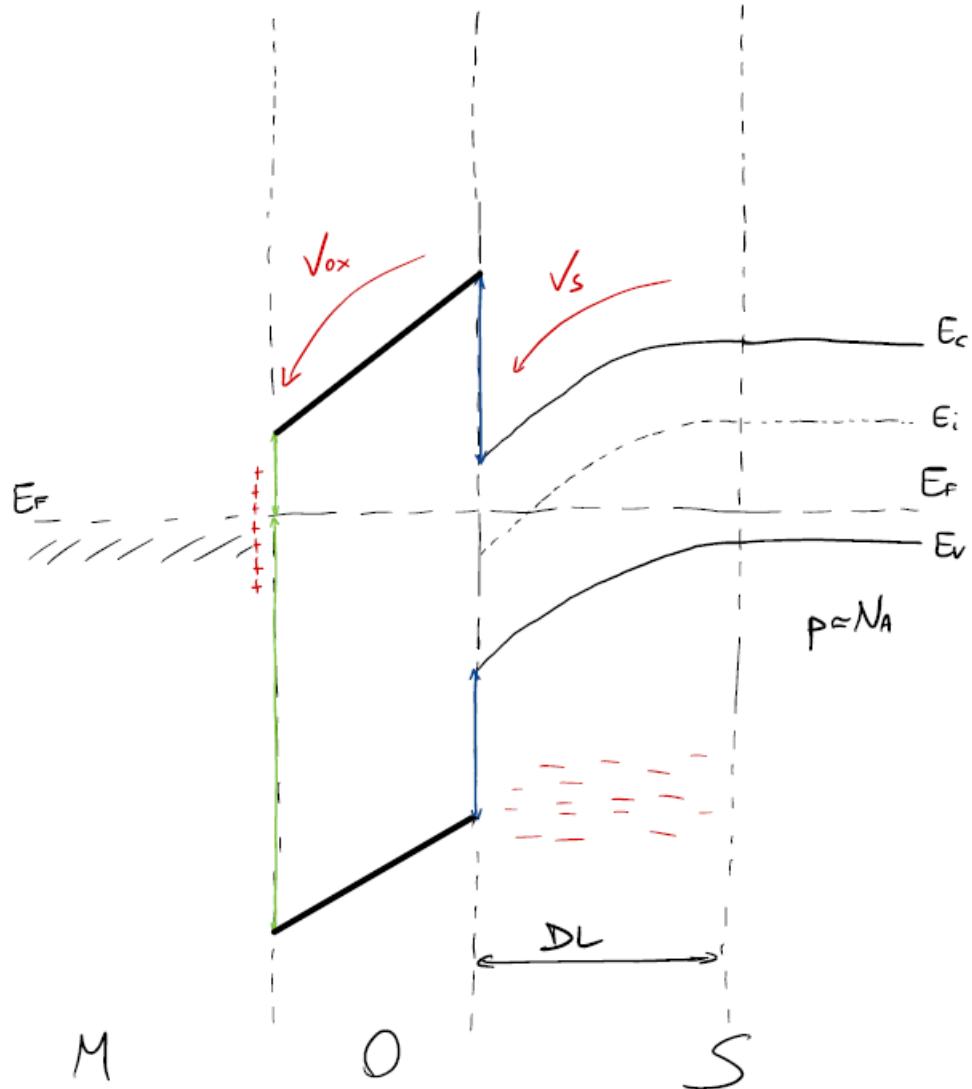


Gauss law is fundamental to understand what happens in the O-S interface, where the relative dielectric constant changes: F is discontinuous, in fact

$$\begin{aligned}\epsilon_{ox} F_{ox} &= \epsilon_{Si} F_S \\ F_{ox} &= \frac{\epsilon_{Si}}{\epsilon_{ox}} F_S \approx 3F_S\end{aligned}\quad (4.1)$$



Since there cannot be any charge inside of the oxide, V_{ox} must be linear.

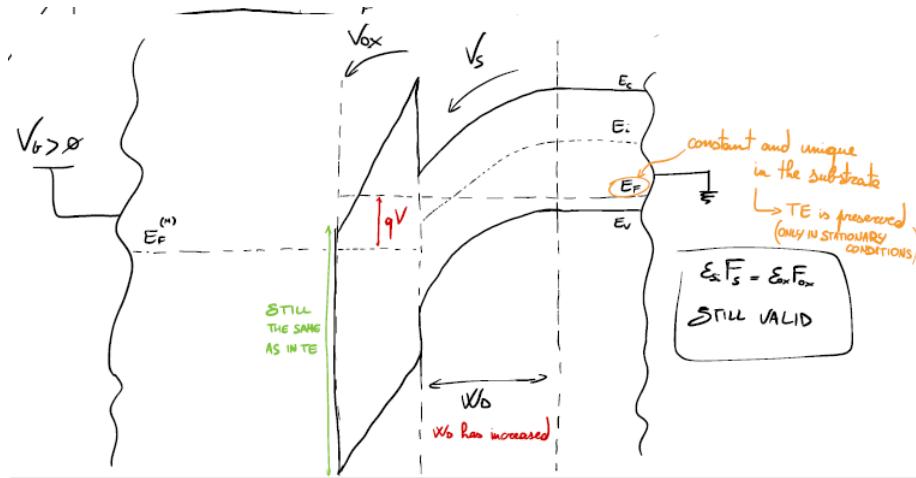


Total voltage drop on the the device is given by $V_s + V_{ox} = \phi_{Bi}$ ¹, which is also the separation between $E_F^{(M)}$ and $E_F^{(S)}$.

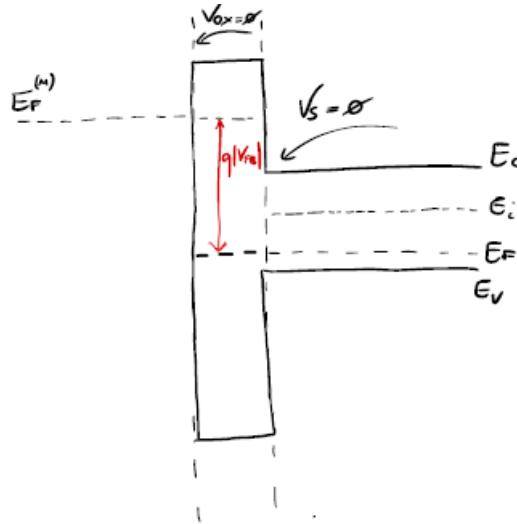
¹Typically $\phi_{Bi} \approx 1 V$

4.2 Gate bias

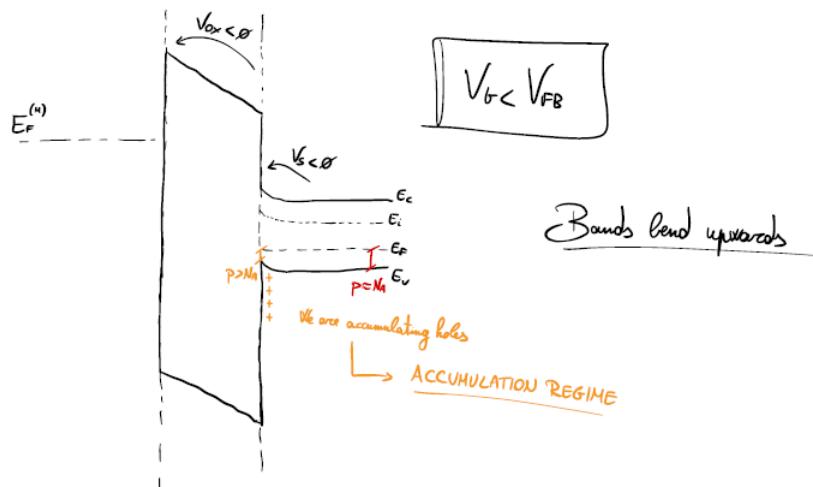
We have seen the device under thermodynamic equilibrium, now we apply a voltage to the gate and see what happens starting from a generic V_G :



In this case $V_s + V_{ox} = \phi_{bi} + V_G$; if we were to put $V_G = -\phi_{bi}$, then we would have $V_s + V_{ox} = 0$, but since those voltages must have the same sign $V_s = V_{ox} = 0$.

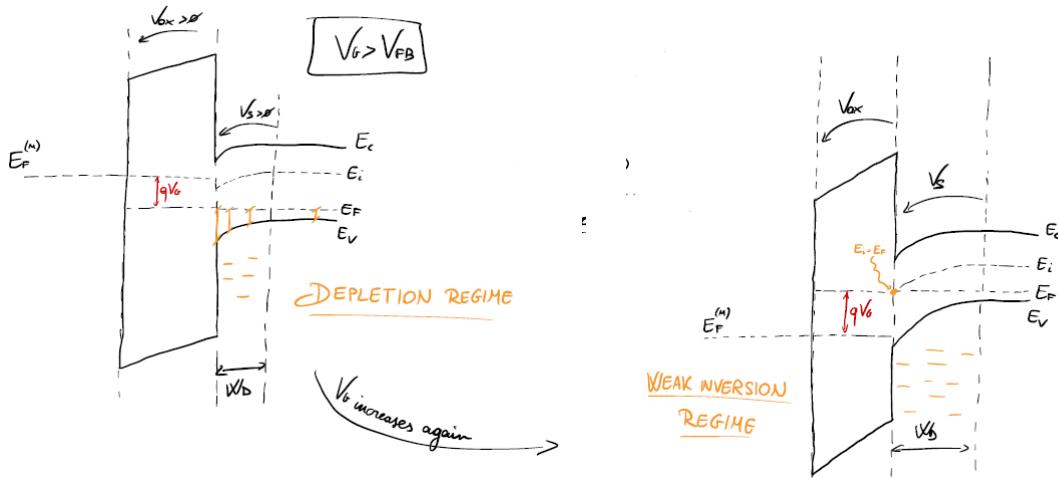


That is known as the **flat band condition** ($V_G = -\phi_{bi} = V_{FB}$), where charge neutrality is total.
Now we go below V_{FB}



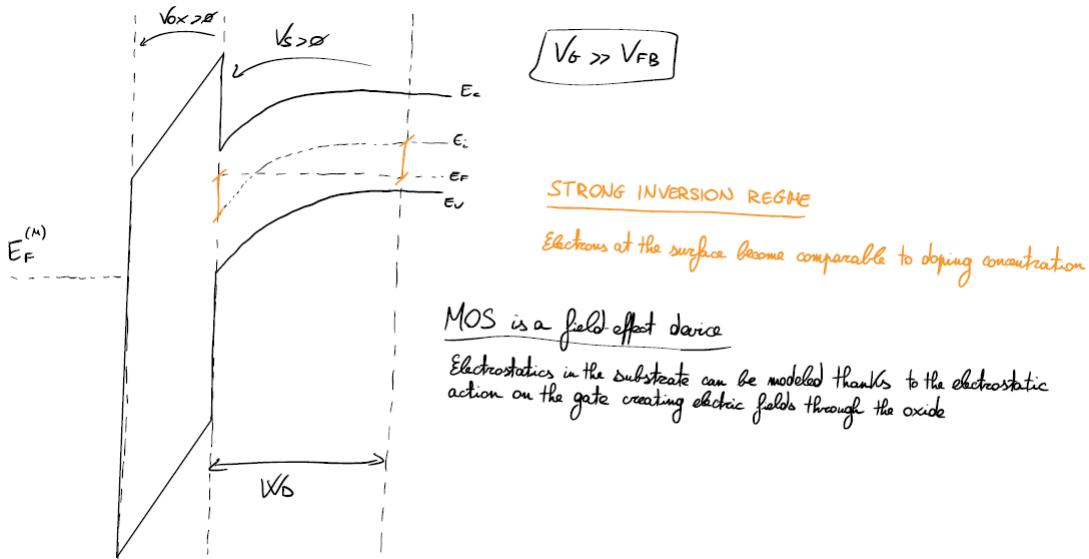
Positive charges are moved to the semiconductor surface: this is the **accumulation regime**.

What if we went above V_{FB} ?



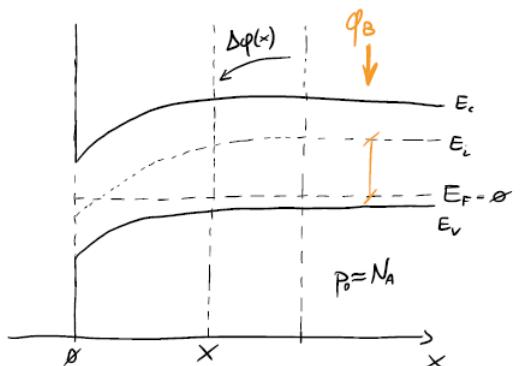
At the silicon surface, electrons and holes are quite the same concentration, still negligible in comparison to the doping concentration; **depletion and weak inversion regimes** can be considered as a single one.

If we keep increasing the gate voltage



4.3 Reformulation of the Poisson equation

We start in strong inversion regime, so we have a situation in the substrate like that:



At the silicon surface $n > N_A$, so the ionized acceptors charge inside the depletion layer becomes less relevant. Due to the

Poisson equation (assuming complete ionization), then

$$\frac{d^2\phi}{dx^2} = -\frac{q}{\epsilon_{\text{Si}}}(p - n + N_D^\leftarrow - N_A^\leftarrow) \quad (4.2)$$

$$\phi(x) = -\frac{E_i(x)}{q} \quad (4.3)$$

We could solve this equation for every section of the MOS, but in the end it is enough to stop at the substrate. Introducing ϕ_B as the potential in the bulk and $\Delta\phi(x)$ as the potential difference between ϕ_B and a random $\phi(x)$ inside the depletion layer, we can write

$$\begin{cases} \phi_B = \phi(x \rightarrow +\infty) = -\frac{E_i(x \rightarrow +\infty)}{q} = -\frac{KT}{q} \log\left(\frac{N_A}{n_i}\right) \\ p_0^2 = n_i e^{\frac{E_i - E_F}{KT}} = N_A \Rightarrow E_i = KT \log\left(\frac{N_A}{n_i}\right) \\ \Delta\phi(x) = \phi(x) - \phi_B \end{cases} \quad (4.4)$$

We know that

$$n = n_i \exp\left(\frac{\overset{\text{Reference level}}{E_F^\leftarrow} - E_i}{KT} \frac{q}{q}\right) = n_i \exp\left(\frac{q\phi}{KT}\right) \quad (4.5)$$

$$n_0 = n_i \exp\left(\frac{q\phi_B}{KT}\right)$$

thus

$$\frac{n}{n_0} = \exp\left(\frac{q(\phi - \phi_B)}{KT}\right) \implies n = n_0 \exp\left(\frac{q\Delta\phi}{KT}\right) \quad (4.6)$$

Same procedure can be applied for the holes, so

$$p = p_0 \exp\left(-\frac{q\Delta\phi}{KT}\right) \quad (4.7)$$

Since the substrate is under thermodynamic equilibrium, 1.21 is still valid; for charge neutrality, then

$$p_0 = n_0 + N_A \Rightarrow N_A = p_0 - n_0$$

without considering approximations. We can finally highlight the dependence of 4.2 on $\Delta\phi$:

$$\frac{d^2\Delta\phi}{dx^2} = -\frac{q}{\epsilon_{\text{Si}}} \left[p_0 \exp\left(-\frac{q\Delta\phi}{KT}\right) - n_0 \exp\left(\frac{q\Delta\phi}{KT}\right) - p_0 + n_0 \right]$$

and then its dependence on p_0 :

$$\frac{d^2\Delta\phi}{dx^2} = -\frac{q}{\epsilon_{\text{Si}}} \left[p_0 \exp\left(-\frac{q\Delta\phi}{KT}\right) - \frac{n_i^2}{p_0} \exp\left(\frac{q\Delta\phi}{KT}\right) - p_0 + \frac{n_i^2}{p_0} \right]$$

Finally, considering that $p_0 \approx N_A$, we get

$$\frac{d^2\Delta\phi}{dx^2} = -\frac{q}{\epsilon_{\text{Si}}} \left[N_A \exp\left(-\frac{q\Delta\phi}{KT}\right) - \frac{n_i^2}{N_A} \exp\left(\frac{q\Delta\phi}{KT}\right) - N_A + \frac{n_i^2}{N_A} \right] \quad (4.8)$$

²Pedices '0' stands for "in the bulk"

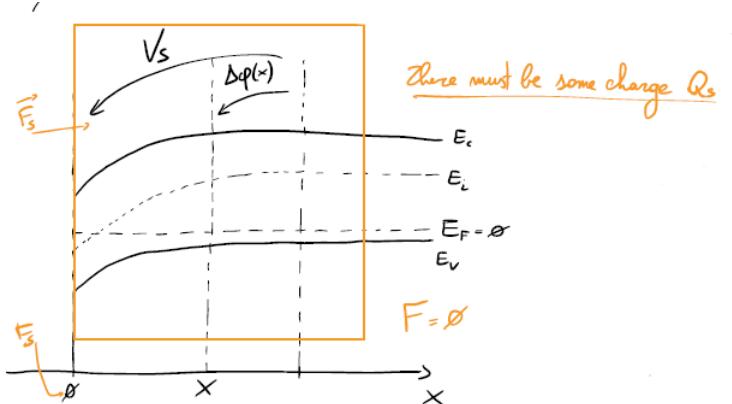
4.4 Calculation of Q_s as a function of V_s

We now want to demonstrate that there is some charge Q_s laying on the semiconductor surface. Starting point is, of course, 4.8; let's integrate

$$\begin{aligned} \int_{\frac{d\Delta\phi}{dx}}^0 \frac{d\Delta\phi}{dx} d\left(\frac{d\Delta\phi}{dx}\right) &= \int_{\Delta\phi}^0 -\frac{q}{\epsilon_{\text{Si}}} \left[N_A \exp\left(-\frac{q\Delta\phi}{KT}\right) - \frac{n_i^2}{N_A} \exp\left(\frac{q\Delta\phi}{KT}\right) - N_A + \frac{n_i^2}{N_A} \right] dx \cancel{\frac{d\Delta\phi}{dx}} \\ -\frac{1}{2} \left(\frac{d\Delta\phi}{dx} \right)^2 &= -\frac{q}{\epsilon_{\text{Si}}} \left[N_A \exp\left(-\frac{q\Delta\phi}{KT}\right) \left(-\frac{KT}{q}\right) - \frac{n_i^2}{N_A} \exp\left(\frac{q\Delta\phi}{KT}\right) \left(\frac{KT}{q}\right) - N_A \Delta\phi + \frac{n_i^2}{N_A} \Delta\phi \right]_{\Delta\phi}^0 \\ \left(\frac{d\Delta\phi}{dx} \right)^2 &= \frac{2q}{\epsilon_{\text{Si}}} \left[-\frac{KT}{q} N_A + \frac{KT}{q} N_A \exp\left(-\frac{q\Delta\phi}{KT}\right) - \frac{KT}{q} \frac{n_i^2}{N_A} + \frac{KT}{q} \frac{n_i^2}{N_A} \exp\left(\frac{q\Delta\phi}{KT}\right) + N_A \Delta\phi - \frac{n_i^2}{N_A} \Delta\phi \right] \\ \left(\frac{d\Delta\phi}{dx} \right)^2 &= \frac{2q}{\epsilon_{\text{Si}}} \frac{KT}{q} N_A \left[-1 + \exp\left(-\frac{q\Delta\phi}{KT}\right) - \frac{n_i^2}{N_A^2} + \frac{n_i^2}{N_A^2} \exp\left(\frac{q\Delta\phi}{KT}\right) + \frac{q\Delta\phi}{KT} - \frac{n_i^2}{N_A^2} \frac{q\Delta\phi}{KT} \right] \end{aligned}$$

Finally, we obtain

$$\frac{d\Delta\phi}{dx} = \pm \sqrt{\frac{2KT N_A}{\epsilon_{\text{Si}}}} \left[\exp\left(-\frac{q\Delta\phi}{KT}\right) + \frac{q\Delta\phi}{KT} - 1 + \frac{n_i^2}{N_A^2} \left(\exp\left(\frac{q\Delta\phi}{KT}\right) - \frac{q\Delta\phi}{KT} - 1 \right) \right]^{1/2} = -F(\Delta\phi) \quad (4.9)$$



Defining $F_s = F(V_s)$ and $Q_s = -\epsilon_{\text{Si}} F_s$, we arrive to the conclusion that

$$Q_s = \pm \sqrt{2\epsilon_{\text{Si}} K T N_A} \left[\exp\left(-\frac{qV_s}{KT}\right) + \frac{qV_s}{KT} - 1 + \frac{n_i^2}{N_A^2} \left(\exp\left(\frac{qV_s}{KT}\right) - \frac{qV_s}{KT} - 1 \right) \right]^{1/2} \quad (4.10)$$

We don't know the electric field as a function of x , but we got the total charge in the substrate given V_s for whatever regime.

4.5 $Q_s - V_s$ curve

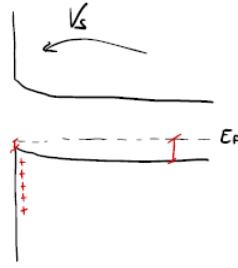
It is time to represent how Q_s changes in the different regimes; first we want to break up 4.10:

$$Q_s = \sqrt{2\epsilon_{\text{Si}} K T N_A} \left[\exp\left(-\frac{qV_s}{KT}\right) + \frac{qV_s}{KT} - 1 + \frac{n_i^2}{N_A^2} \left(\exp\left(\frac{qV_s}{KT}\right) - \frac{qV_s}{KT} - 1 \right) \right]^{1/2}$$

- Holes (accumulation regime)
- Impurities (depletion/weak inversion regime)
- Electrons (strong inversion regime)

Then, for each regime it is possible to approximate Q_s ; let's start from the **accumulation regime ($V_s < 0$)**:

$$Q_s \approx \sqrt{2\epsilon_{\text{Si}}KTN_A} \exp\left(-\frac{qV_s}{KT}\right) \quad (4.11)$$



Continuing with the **flat band ($V_s = 0$)**, it is trivial that

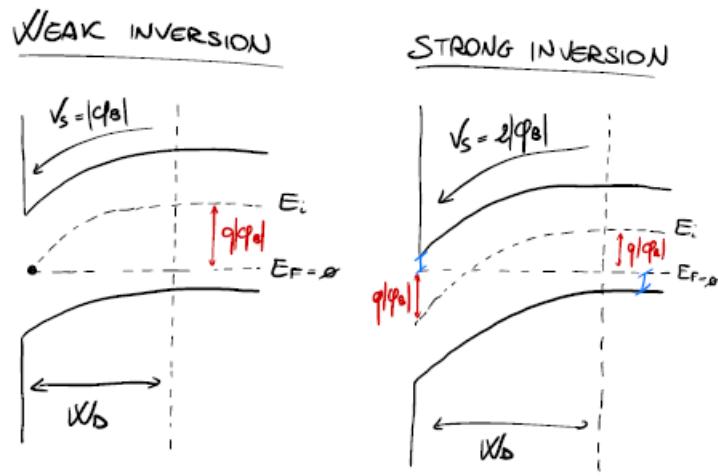
$$Q_s = 0 \quad (4.12)$$

Now we consider **depletion/weak inversion regime ($0 < V_s < 2|\phi_B|$)**:

$$Q_s \approx -\sqrt{2\epsilon_{\text{Si}}KTN_A \frac{qV_s}{KT}} = -qN_A W_D \quad (4.13)$$

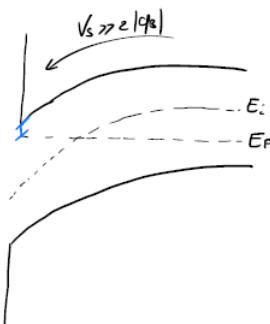
Note that

$$\frac{n_i^2}{N_A^2} \exp\left(\frac{q2|\phi_B|}{KT}\right) = \frac{n_i^2}{N_A^2} \exp\left(\frac{\frac{q2}{q} \frac{KT}{KT} \log\left(\frac{N_A}{n_i}\right)}{KT}\right) = \frac{n_i^2}{N_A^2} \left(\frac{N_A}{n_i}\right)^2 = 1 \quad (4.14)$$

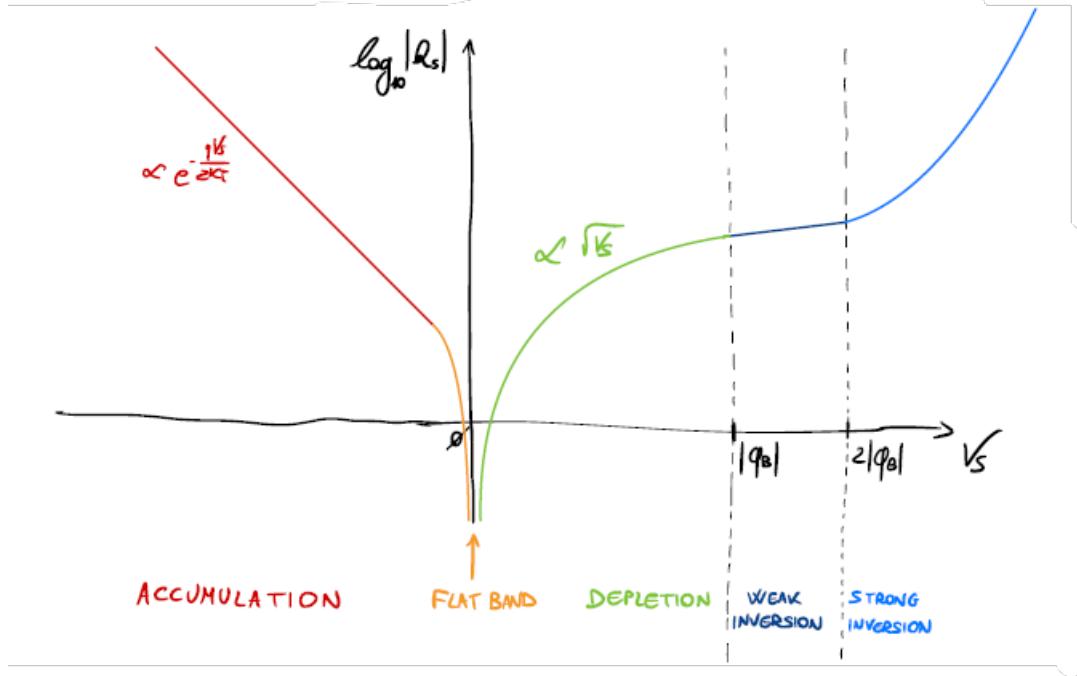


Finally, the **strong inversion regime ($V_s >> 2|\phi_B|$)**:

$$Q_s \approx -\sqrt{2\epsilon_{\text{Si}}KTN_A} \frac{n_i}{N_A} \exp\left(\frac{qV_s}{2KT}\right) \quad (4.15)$$



Eventually we get these results.



4.6 Dependence of V_s and Q_s on V_G

We have seen that $Q_s = Q_s(V_s)$, although we are unable to find $\phi(x)$ for every working regime. Starting from this very important relation

$$\begin{aligned} V_G + \phi_{BI} &= V_s + V_{ox} \\ \downarrow \\ V_G - V_{FB} &= V_s + V_{ox} \end{aligned} \tag{4.16}$$

and rewriting V_{ox} as

$$V_{ox} = F_{ox}t_{ox} = \frac{\epsilon_{Si}}{\epsilon_{ox}}F_st_{ox} = -\frac{Q_s}{\epsilon_{ox}}t_{ox} = -\frac{Q_s}{C_{ox}} \tag{4.17}$$

we arrive at the system

$$\begin{cases} Q_s = Q_s(V_s) \\ V_G - V_{FB} = V_s - \frac{Q_s(V_s)}{C_{ox}} = f(V_s) \end{cases} \tag{4.18}$$

where V_s is the only unknown term; we have to find $V_s(V_G)$ and $Q_s(V_G)$ for each regime.

Accumulation regime We start from 4.11 and put it inside 4.18, so we get

$$V_G - V_{FB} = -\frac{\sqrt{2\epsilon_{Si}KTN_A}}{C_{ox}} \exp\left(-\frac{qV_s}{KT}\right) + V_s^{\text{IRRELEVANT}} \tag{4.19}$$

Doing some math, final result is

$$V_s \approx -\frac{2KT}{q} \log \left[\frac{C_{ox}(V_{FB} - V_G)}{\sqrt{2\epsilon_{Si}KTN_A}} \right] \tag{4.20}$$

$$Q_s \approx C_{ox}(V_{FB} - V_G) \tag{4.21}$$

V_s dependence on V_G is logarithmic, so it is very weak; i.e., if we consider $t_{ox} = 3 \text{ nm}$ and $N_A = 10^{17} \text{ cm}^{-3}$

$$V_{FB} - V_G = 1; V \Rightarrow V_s = -190 \text{ mV}$$

$$V_{FB} - V_G = 2; V \Rightarrow V_s = -225 \text{ mV}$$

Just a small bending is needed, because hole concentration already increases exponentially.

Depletion/weak inversion regime We can refer to Q_s as the charge in the depletion layer, so

$$Q_s = -\sqrt{2\epsilon_{si}qN_A V_s} = Q_{DEP}$$

$$V_G - V_{FB} = V_s + \frac{\sqrt{2\epsilon_{si}qN_A V_s}}{C_{ox}} \quad (4.22)$$

If we were to remove the approximation that the charge in the inversion layer is negligible, though, we would get

$$Q_s = -\sqrt{2\epsilon_{si}KT N_A} \left[\frac{qV_s}{KT} + \frac{n_i^2}{N_A^2} \exp\left(\frac{qV_s}{KT}\right) \right]^{1/2} = -\sqrt{2\epsilon_{si}KT N_A} \sqrt{\frac{qV_s}{KT}} \left[1 + \frac{n_i^2}{N_A^2} \exp\left(\frac{qV_s}{KT}\right) \frac{KT}{qV_s} \right]^{1/2} \stackrel{\text{TAYLOR}}{=} \\ \stackrel{\text{TAYLOR}}{=} -\sqrt{2\epsilon_{si}KT N_A} \sqrt{\frac{qV_s}{KT}} \left[1 + \frac{1}{2} \frac{n_i^2}{N_A^2} \exp\left(\frac{qV_s}{KT}\right) \frac{KT}{qV_s} \right] = -\sqrt{2\epsilon_{si}qN_A V_s} - \sqrt{\frac{2\epsilon_{si}N_A}{qV_s} KT} \frac{1}{2} \frac{n_i^2}{N_A^2} \exp\left(\frac{qV_s}{KT}\right) = \\ = Q_{DEP} + Q_{INV} \quad (4.23)$$

There is an inversion charge growing exponentially with V_s , something that will be useful for the subthreshold currents study. From the previous equations, finally, it is clear that V_s has a linear dependence on V_G , while Q_s has a root dependence.

Threshold voltage We may want to identify another characteristic point, i.e. the threshold trigger to strong inversion V_T . Since we are in-between weak inversion and strong inversion, $V_s = 2|\phi_B|$, therefore we can rewrite 4.18 as

$$V_G = V_{FB} + 2|\phi_B| + \frac{\sqrt{2\epsilon_{si}qN_A 2|\phi_B|}}{C_{ox}} = V_T \quad (4.24)$$

Strong inversion regime For 4.14, it is true that

$$\frac{n_i}{N_A} = \exp\left(-\frac{q2|\phi_B|}{2KT}\right) \quad (4.25)$$

so if we combine this with 4.15 we obtain

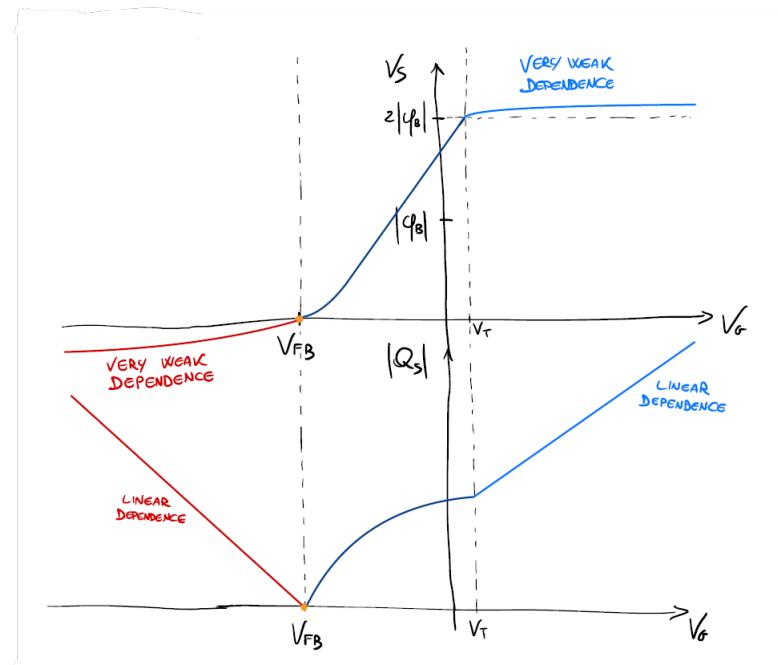
$$Q_s \approx -\sqrt{2\epsilon_{si}KT N_A} \exp\left(\frac{q(V_s - 2|\phi_B|)}{2KT}\right) \quad (4.26)$$

Starting from 4.18 again, it is easy to demonstrate

$$V_s \approx 2|\phi_B| + \frac{2KT}{q} \log\left[\frac{C_{ox}(V_G - V_{FB} - V_s)^3}{\sqrt{2\epsilon_{si}KT N_A}}\right] \approx 2|\phi_B| \quad (4.27)$$

$$Q_s \approx -C_{ox}(V_G - V_{FB} - 2|\phi_B|) \quad (4.28)$$

$$Q_{INV} = Q_s - Q_{DEP} = -C_{ox}(V_G - V_{FB} - 2|\phi_B|) + \sqrt{2\epsilon_{si}qN_A 2|\phi_B|} = -C_{ox}(V_G - V_T) \quad (4.29)$$



³We treat this like it was a given datum

Note that W_D in strong inversion regime does not increase significantly, so

$$W_{D_{max}} = \sqrt{\frac{2\epsilon_{Si}}{q} \frac{1}{N_A} 2|\phi_B|} \quad (4.30)$$

4.7 Small signal model

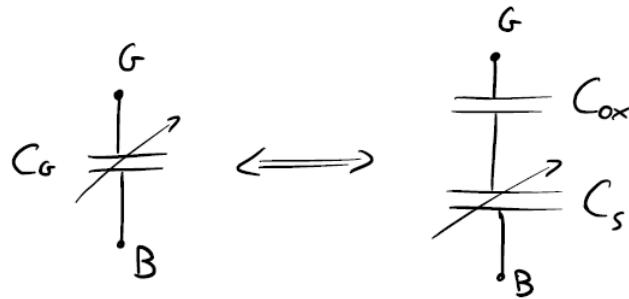
As we have seen previously, Q_s is not linearly dependent on V_G , so we can define just a small signal capacitance

$$C_G = \frac{dQ_s}{dV_G} = C_G(V_G) \quad (4.31)$$

Thus we can rewrite 4.18 as

$$\begin{aligned} \frac{dV_G}{d(-Q_s)} &= \frac{dV_s}{d(-Q_s)} + \frac{1}{C_{ox}} \\ \Downarrow \\ \frac{1}{C_G} &= \frac{1}{C_s} + \frac{1}{C_{ox}} \end{aligned} \quad (4.32)$$

where C_s is the substrate capacitance. Gate capacitance can be represented as the series of substrate and oxide capacitances ($C_G < C_{ox}$).

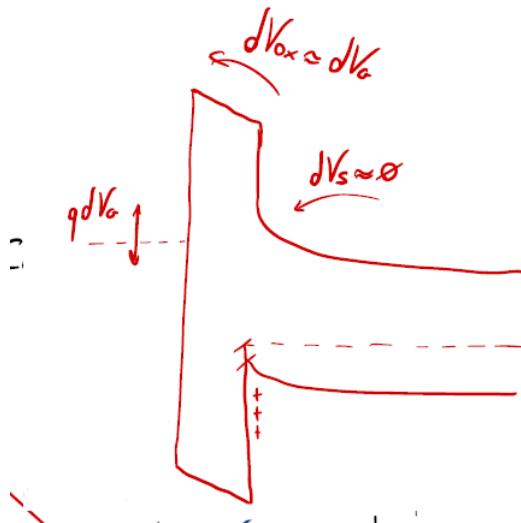


Let's study how C_G changes according to gate bias.

Accumulation regime Q_s is of course 4.11; we can find the substrate capacitance

$$C_s = \frac{d(-Q_s)}{dV_s} = -\sqrt{2\epsilon_{Si}KTN_A} \exp\left(-\frac{qV_s}{2KT}\right) \left(-\frac{q}{2KT}\right) = \frac{Q_s}{2KT/q} = \frac{C_{ox}(V_{FB} - V_G)}{2KT/q} \quad (4.33)$$

If $V_{FB} - V_G \gg \frac{2KT}{q}$, then $C_s \gg C_{ox}$ and $C_G \rightarrow C_{ox}$.



A small voltage modulation is enough to provoke a large charge rearrangement; modulation is negligible in the substrate, so the MOS acts like a metal plate capacitor.

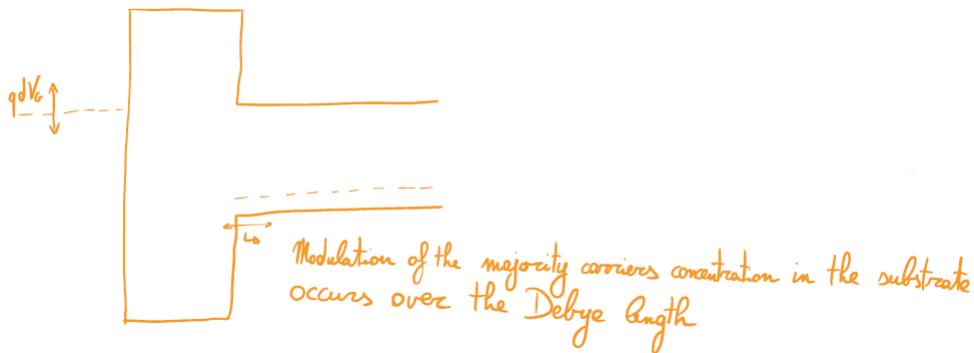
Flat band Approaching the flat band region, $Q_s \neq 0$ yet, therefore

$$Q_s = \sqrt{2\epsilon_{\text{Si}}KTN_A} \left[\exp\left(-\frac{qV_s}{KT}\right) + \frac{qV_s}{KT} - 1 \right]^{1/2} \stackrel{\text{TAYLOR}}{=} \sqrt{2\epsilon_{\text{Si}}KTN_A} \left[1 - \frac{qV_s}{KT} + \frac{1}{2} \left(\frac{qV_s}{KT} \right)^2 + \frac{qV_s}{KT} - 1 \right] = \sqrt{2\epsilon_{\text{Si}}KTN_A} \left[\frac{1}{\sqrt{2}} \frac{qV_s}{KT} \right] = \sqrt{\frac{\epsilon_{\text{Si}}N_A q^2}{KT}} V_s \quad (4.34)$$

and so C_s

$$C_s = \frac{d(-Q_s)}{dV_s} = \sqrt{\frac{\epsilon_{\text{Si}}N_A q^2}{KT}} \sqrt{\frac{\epsilon_{\text{Si}}}{\epsilon_{\text{Si}}}} = \frac{\epsilon_{\text{Si}}}{\sqrt{\epsilon_{\text{Si}}KT/q^2 N_A}} = \frac{\epsilon_{\text{Si}}}{L_D} \quad (4.35)$$

where L_D is the **Debye length**, the rearrangement extension of the carriers over which they screen electric fields perturbations. Given that typically $\epsilon_{\text{Si}} = 11.7\epsilon_0$, $\epsilon_{\text{ox}} = 3.9\epsilon_0$, $L_D \approx 10 \text{ nm}$ and $tox \approx 1 \text{ nm}$, C_s and C_{ox} are comparable.



Depletion/weak inversion regime We are only considering the charge present in the depletion layer, so

$$Q_s \approx Q_{\text{DEP}} = -qN_A W_D = -\sqrt{2\epsilon_{\text{Si}}qN_A V_s} \quad (4.36)$$

$$C_s = \frac{\epsilon_{\text{Si}}}{W_D(V_s)} = C_{\text{DEP}} \ll C_{ox} \quad (4.37)$$

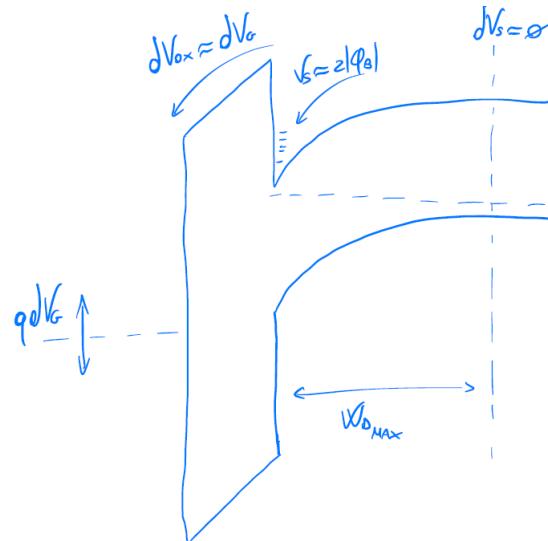
Finally we get

$$C_G = \frac{C_{ox}}{\sqrt{1 + 2C_{ox}^2 \frac{V_G - V_{FB}}{\epsilon_{\text{Si}}qN_A}}} \quad (4.38)$$

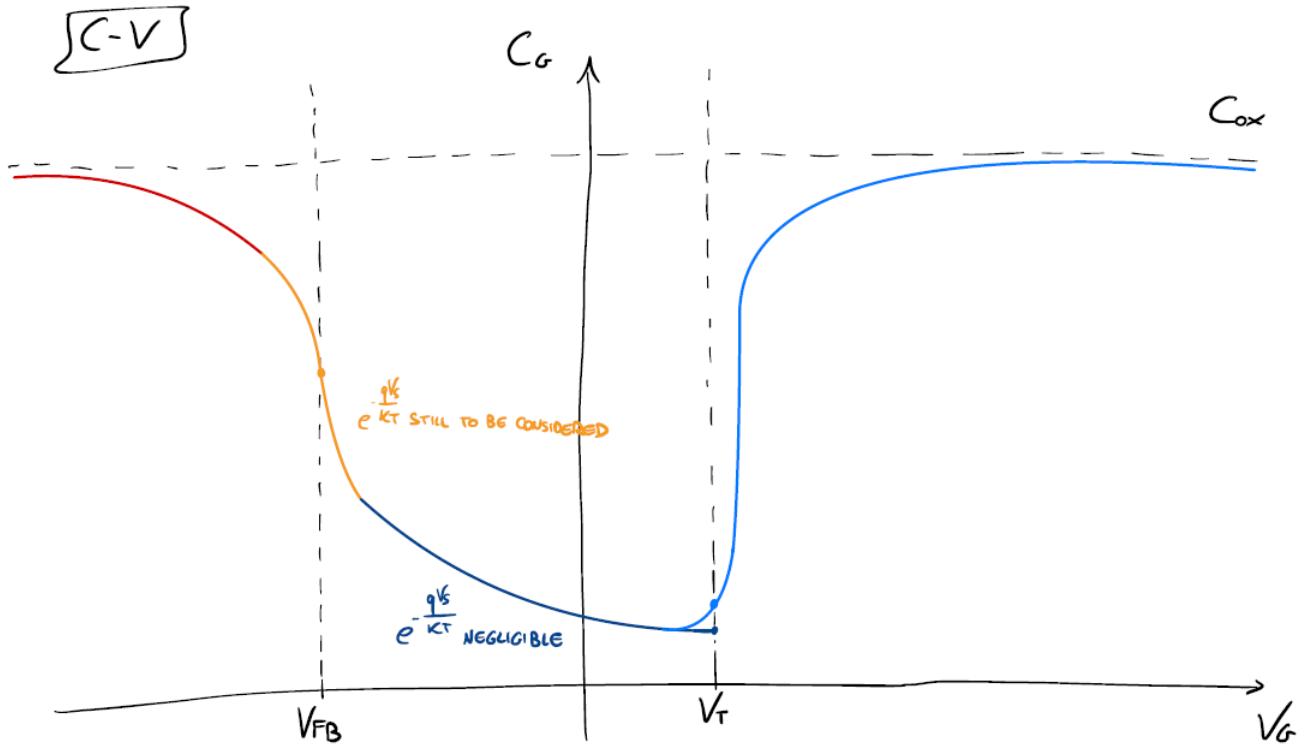
Strong inversion regime Last, we start from 4.15 and immediately find C_s as

$$C_s = \frac{|Q_s|}{2KT/q} = \frac{C_{ox}(V_G - V_{FB} - 2|\phi_B|)}{2KT/q} \quad (4.39)$$

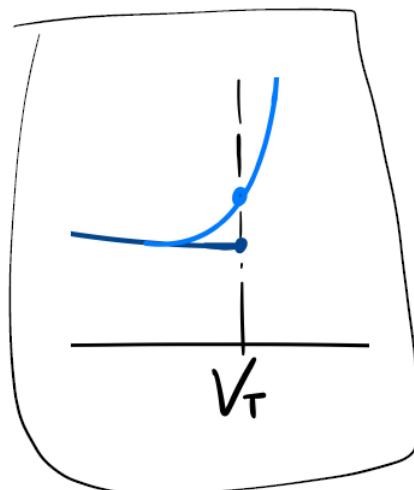
As before, if $V_G - V_{FB} \gg \frac{2KT}{q}$, then $C_s \gg C_{ox}$ and $C_G \rightarrow C_{ox}$.



Now we build the $C - V$ curve:



Note that when $V_G = V_T$, C_G is already rising: small signal voltage can significantly increase the inversion layer charge ($C_s = 2C_{\text{DEP}}$).



Eventually, starting from the curve it is possible to find Q_s just by integrating:

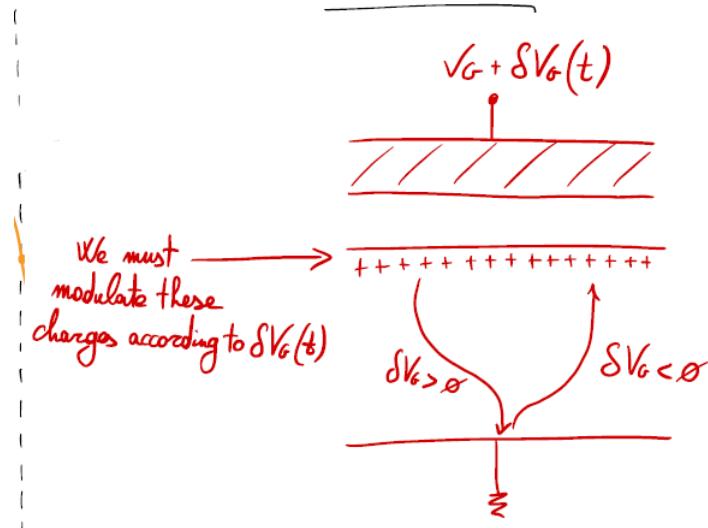
$$Q_s(V_G) = - \int_{V_{FB}}^{V_G} C_G dV_G \quad (4.40)$$

Every analysis we made has been performed under thermodynamic equilibrium in the substrate, even with the small signal; we have never mentioned time either, making this the quasi-static $C - V$ curve.

4.8 $C - V$ curve under low and high frequencies

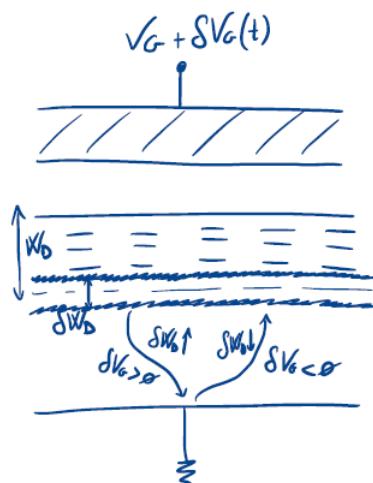
This aforementioned lack of time dependence, though, may be considered quite contradictory as a perturbation is needed to determine a change of charge. We are finally introducing time into the equations.

Accumulation regime Here positive charges are reallocated through the ground contact, as it is shown in the picture.

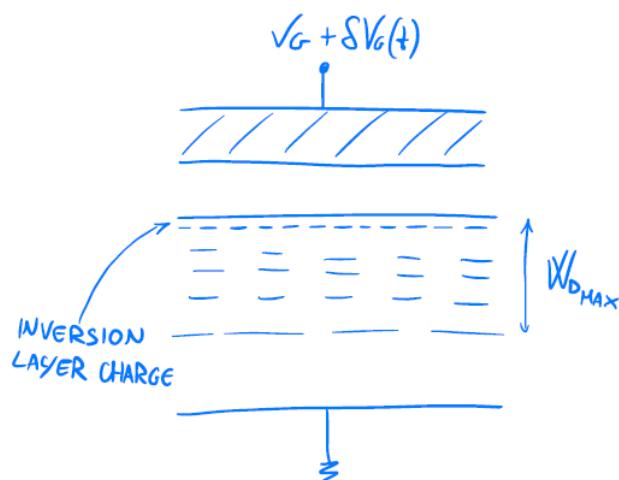


Most common small signal is $\sin(t)$; only charges in the substrate are modulated in a very short time (**Dielectric Relaxation Time**). Throughout our analysis, we are always considering time constants for δV_G larger than DRT: charges are modulated so fast it looks like we are still under thermodynamic equilibrium, making no difference at all.

Depletion/weak inversion regime In this case, we have to modulate the width of the depletion layer to adjust the charges; as before, all is happening so quickly that electrostatics is not perturbed.



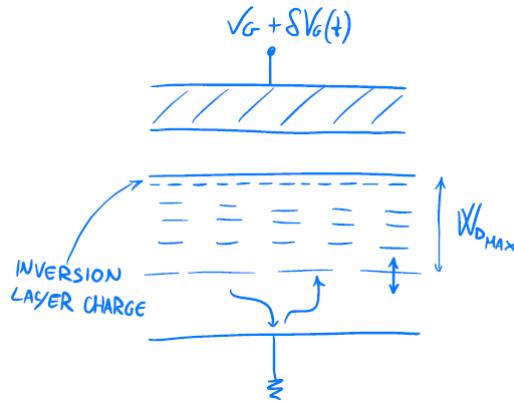
Strong inversion regime Small signal affects only the charges in the inversion layer: this time, electrons cannot be taken by the contact, thus we have to rely only on G/R processes.



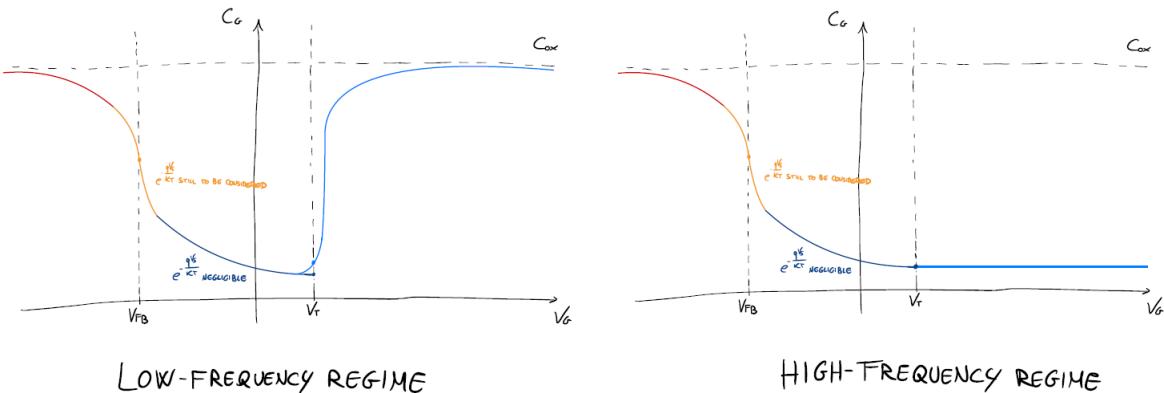
From the $p - n$ junction we know that $J_{DL} \approx \frac{q n_i}{2\tau_0} W_D$; starting from this, we can define our G/R time constant:

$$t_G = \frac{|Q_{DEP}|}{J_{DL}} = \frac{q N_A W_D}{(\frac{q n_i}{2\tau_0}) W_D} = 2\tau_0 \frac{N_A}{n_i} \approx 1 \text{ s} \quad (4.41)$$

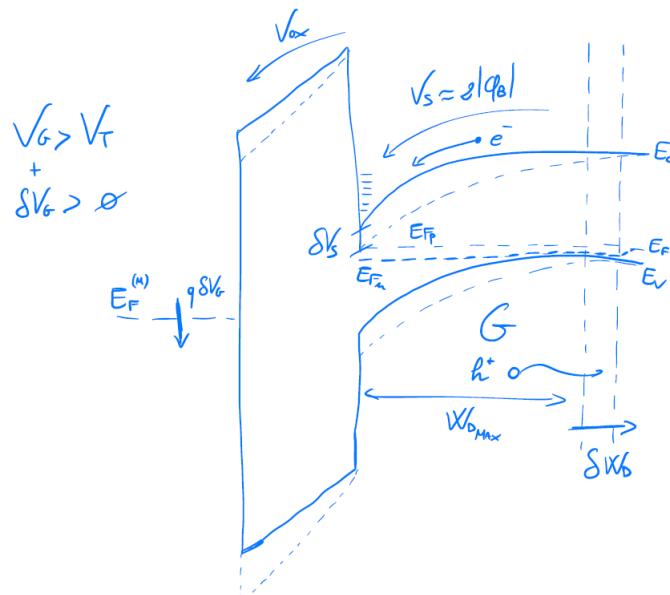
δV_G must change with a time constant larger than t_G ($f_{\text{signal}} < f_{\text{GR}}$), otherwise G/R processes fail their purpose. If that happens, we enter the high-frequency regime as opposed to the low-frequency regime: we are *de facto* modulating the depletion layer ($C_s = \frac{\epsilon_{\text{Si}}}{W_{D_{max}}}$).



Here we can see the main difference between the two regimes.



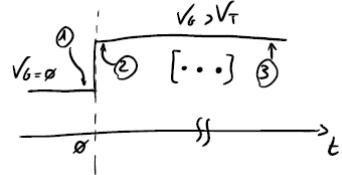
Now focus on the high-frequency regime with $\delta V_G > 0$: since we are increasing the band bending, δV_s is significant. E_{F_n} , then, must move downwards to keep the electron concentration constant.



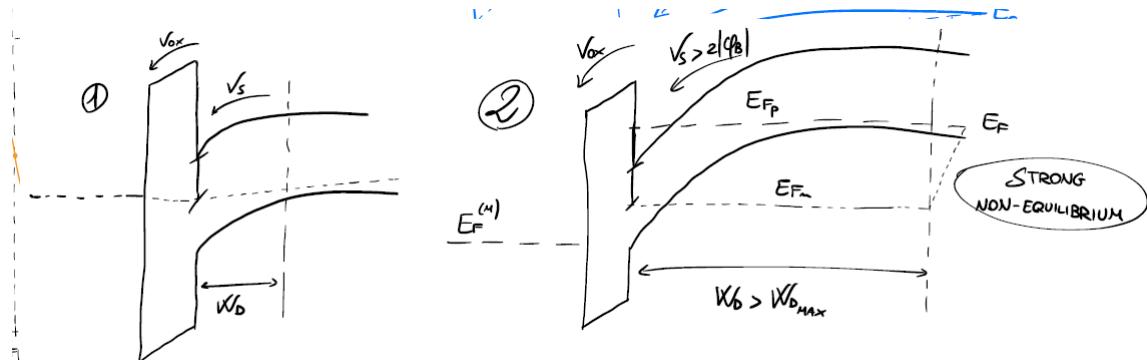
In this case, $E_{F_n} < E_{F_p}$, so we have net generation processes. Non-equilibrium is maintained as E_{F_n} keeps moving according to δV_G : G/R processes keep alternating. On the contrary, in the low-frequency regime, after a lot of time V_G was increased, electrons are stored in the silicon surface moving the bands upwards, thus thermodynamic equilibrium is restored.

4.9 Deep-depletion condition

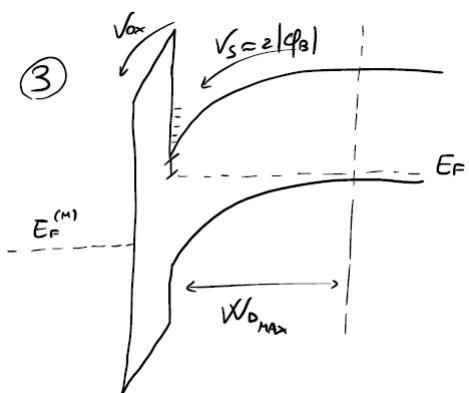
Now we want to know what happens when a large signal is applied. This can be splitted into three segments:



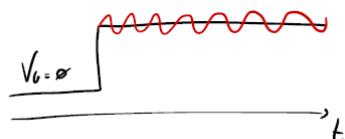
Immediately after the step increase of the voltage ((2)), we don't have a step increase of the electrons, as we have to wait for generation processes; total charge does change, but not because of electrons, that is thanks to depletion layer charges. This brings us to a huge band bending while reaching the **deep-depletion condition**.



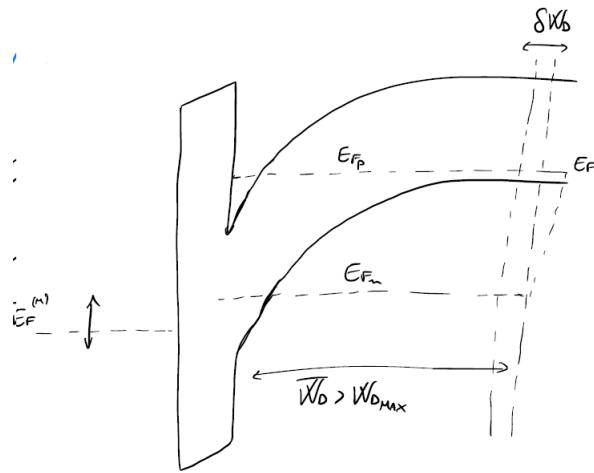
Approaching (3), depletion layer narrows down as charge is relocated in the inversion layer.



Now we want to study what happens if we apply a small signal right after (2).

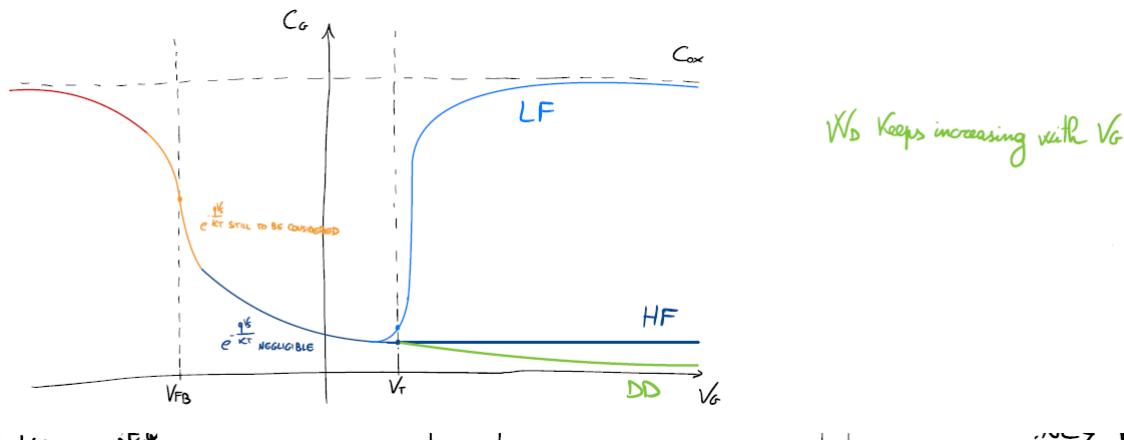


We start from a big disequilibrium; in addition to that we have the small signal keeping the perturbation alive and a depletion layer which is even bigger than $W_{D_{max}}$.



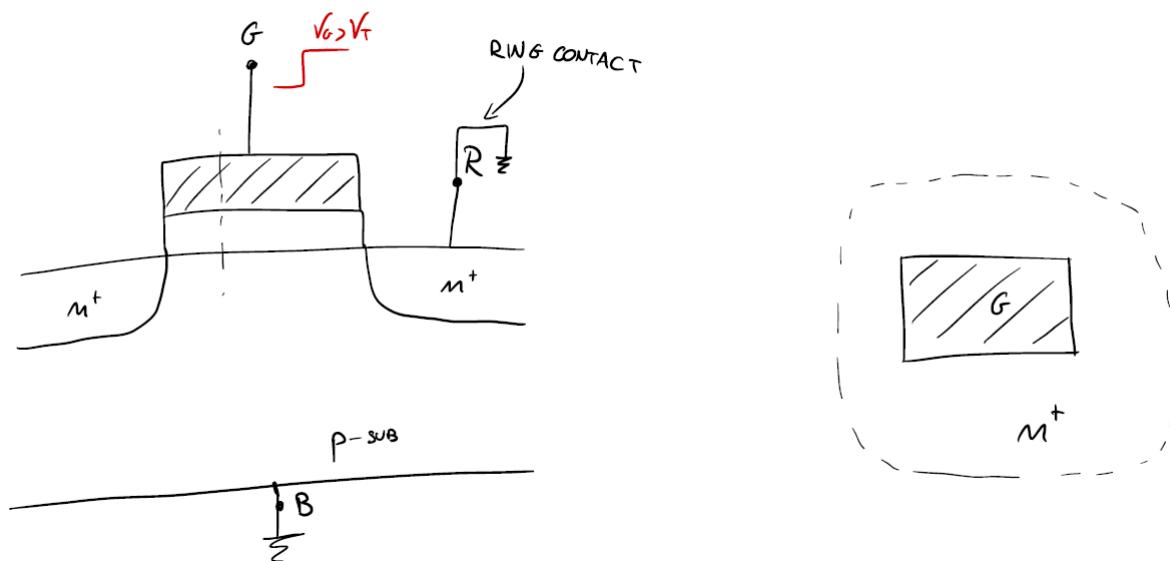
Capacitance value is decreasing too!

$$C_s = \frac{\epsilon_{\text{Si}}}{W_D} \quad (4.42)$$

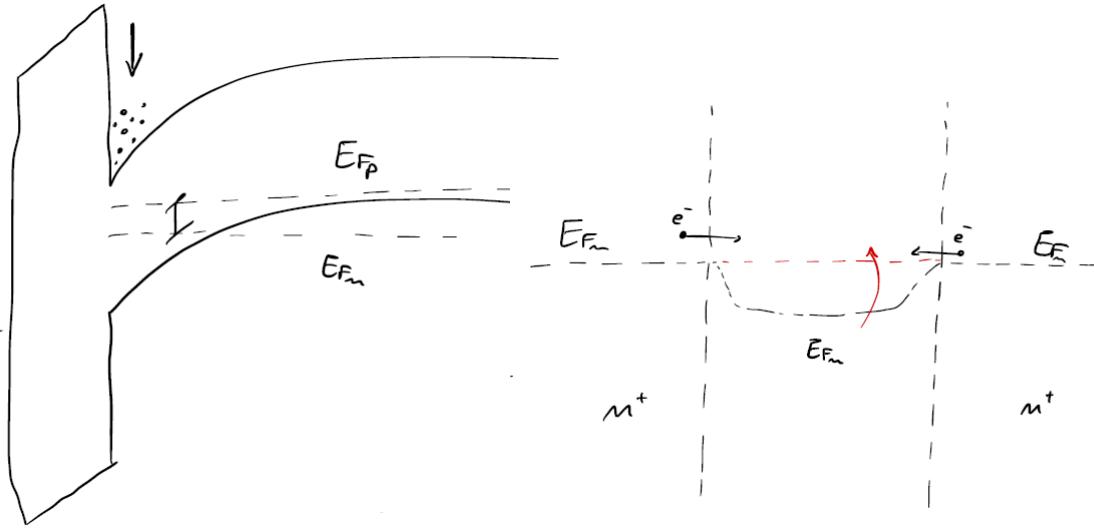


4.10 Basics of the nMOS capacitor with ring

According to the last considerations, the onset of the high-frequency regime is just some tens/hundreds of hertz; obviously, for our purpose, this is not acceptable. We must find a way to reduce electron concentration quicklier than G/R processes: the **nMOS capacitor with ring** is introduced.



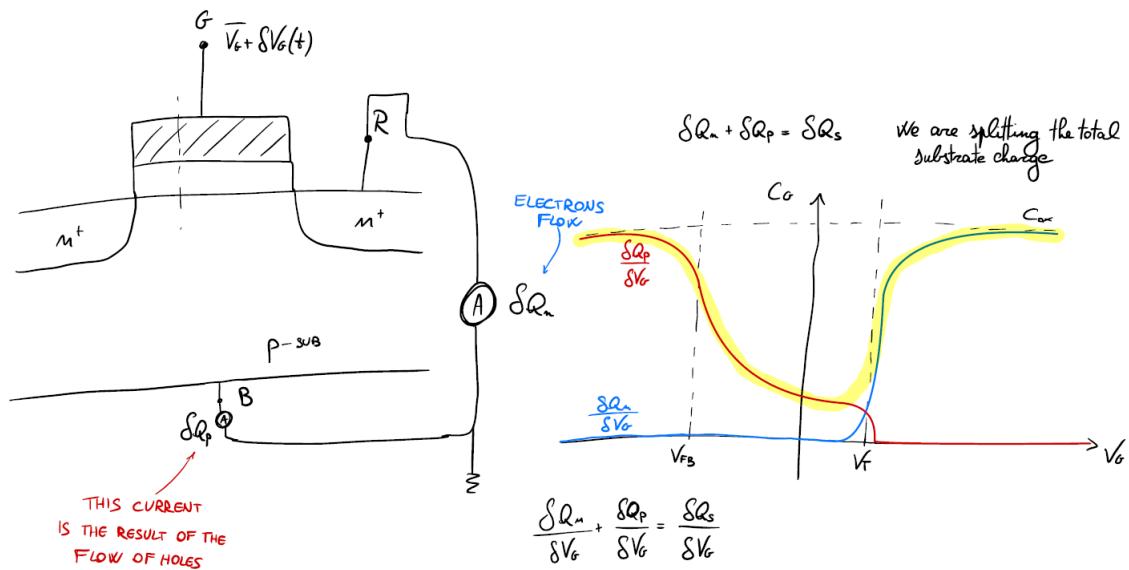
That is a normal MOS capacitor surrounded entirely by a n^+ -region: we now have a 2D device.



We don't have to wait for G/R processes anymore as electrons are taken directly from the n^+ -region. Time constant, in fact, is the transit time to move them from said region to the substrate. This determines a splitting in the total substrate charge

$$\delta Q_s = \delta Q_p + \delta Q_n \quad (4.43)$$

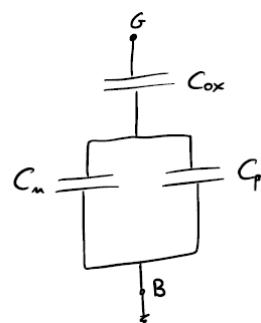
which affects also the $C - V$ curve.



Thus C_G can be considered as the parallel of C_n and C_p :

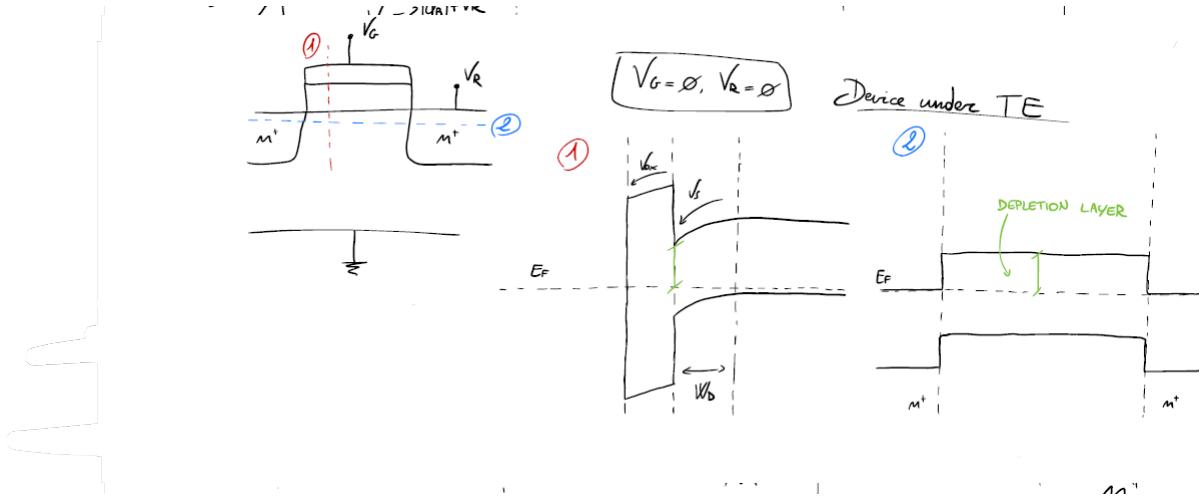
$$\frac{\delta Q_s}{\delta V_s} = \frac{\delta Q_n}{\delta V_s} + \frac{\delta Q_p}{\delta V_s} \quad (4.44)$$

$$C_G = \frac{C_{ox}(C_n + C_p)}{C_{ox} + C_n + C_p} = \frac{C_{ox}C_n}{C_{ox} + C_n + C_p} + \frac{C_{ox}C_p}{C_{ox} + C_n + C_p} = \frac{\delta Q_n}{\delta V_G} + \frac{\delta Q_p}{\delta V_G} \quad (4.45)$$

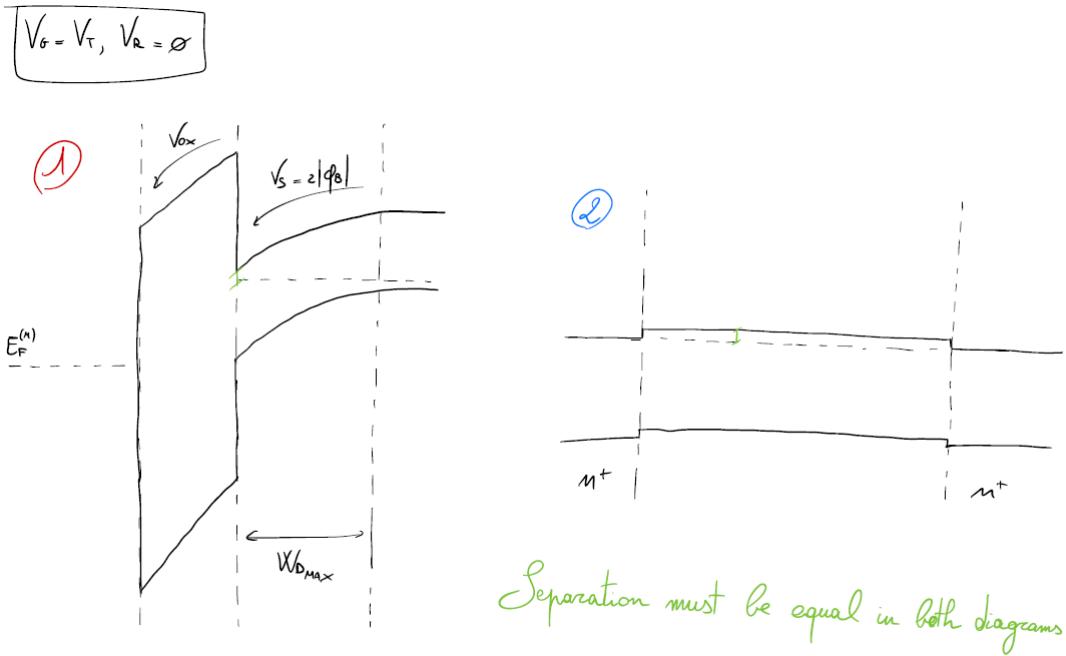


4.11 Impact of the ring bias on the electrostatics

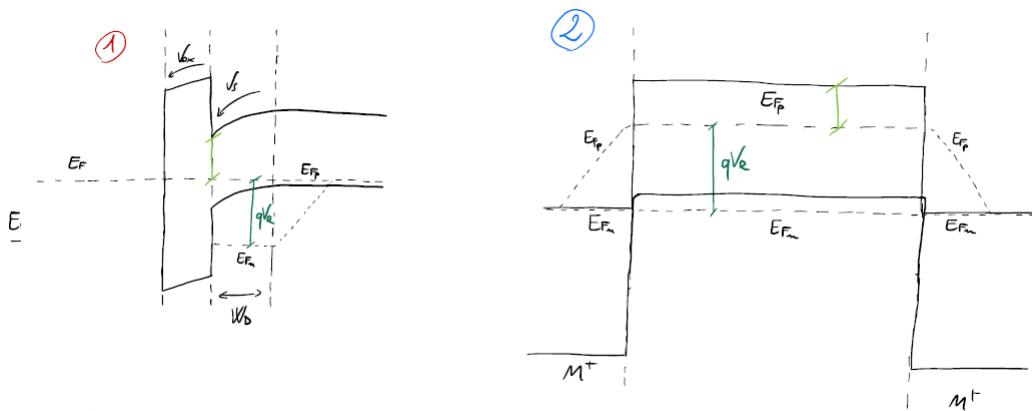
We now discuss the changing in the band diagrams (both vertical and horizontal sections) according to the applied V_G and V_R , starting from thermodynamic equilibrium.



Now we consider strong inversion regime.

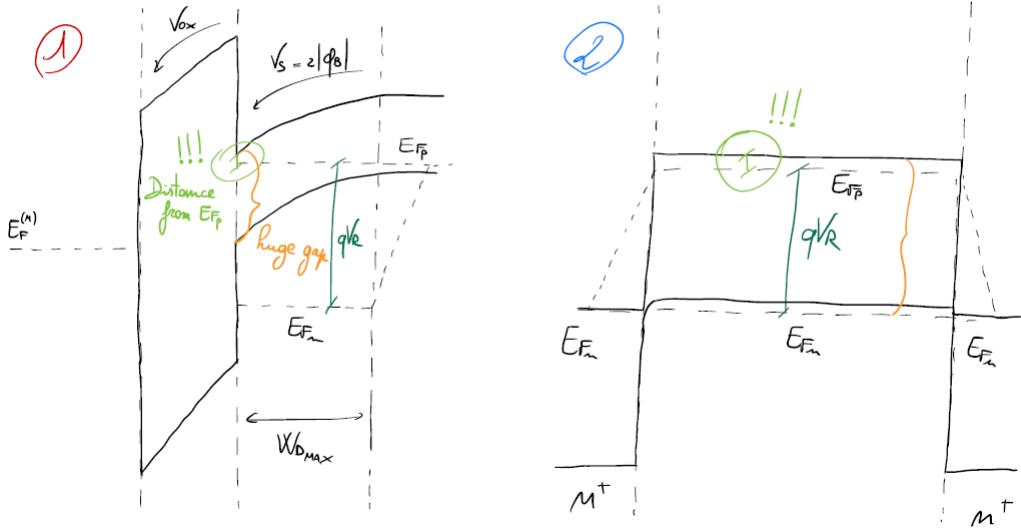


It is time to apply a $V_R > 0$ not to forward-bias the $p - n$ junction ($V_G = 0$).

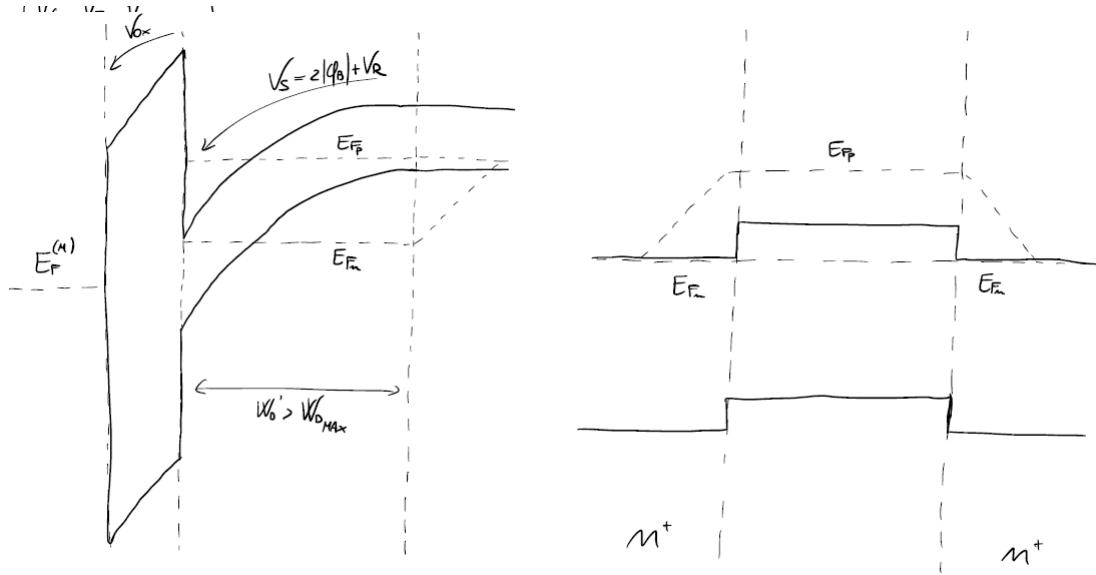


Since electron concentration is negligible, there is no changing in the bands for ①. Last case is the most interesting:

$$V_G = V_T, V_R > 0$$



Pay attention to the fact that the nMOS is NOT at the onset of strong inversion: since $E_c - E_{F_n}$ is very large, electron concentration is still smaller than doping concentration. We then need to find a value V'_T which allows the strong inversion regime to trigger.

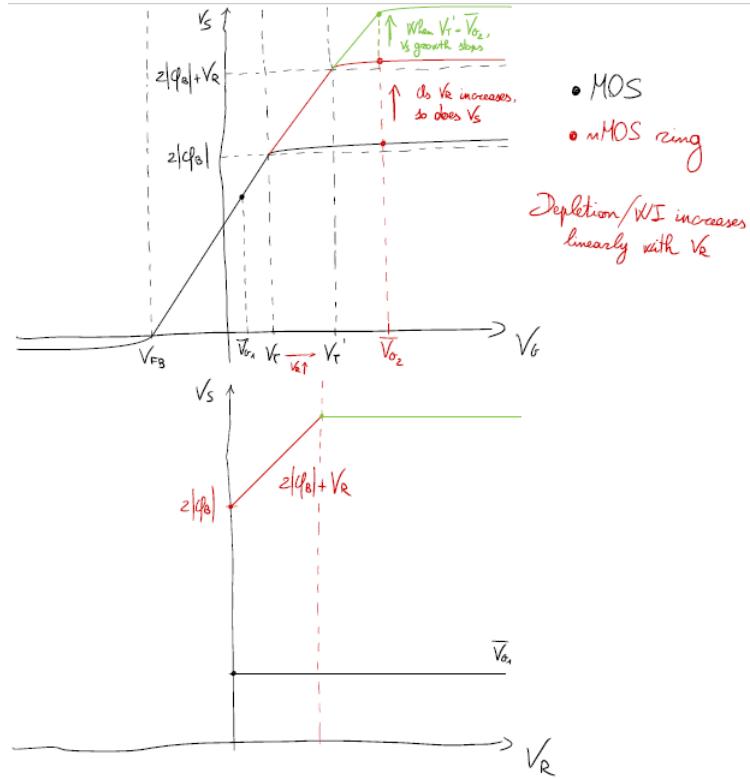


As we can see, $V_s = 2|\phi_B| + V_R$; we have also reached deep-depletion condition, this time is stationary though.

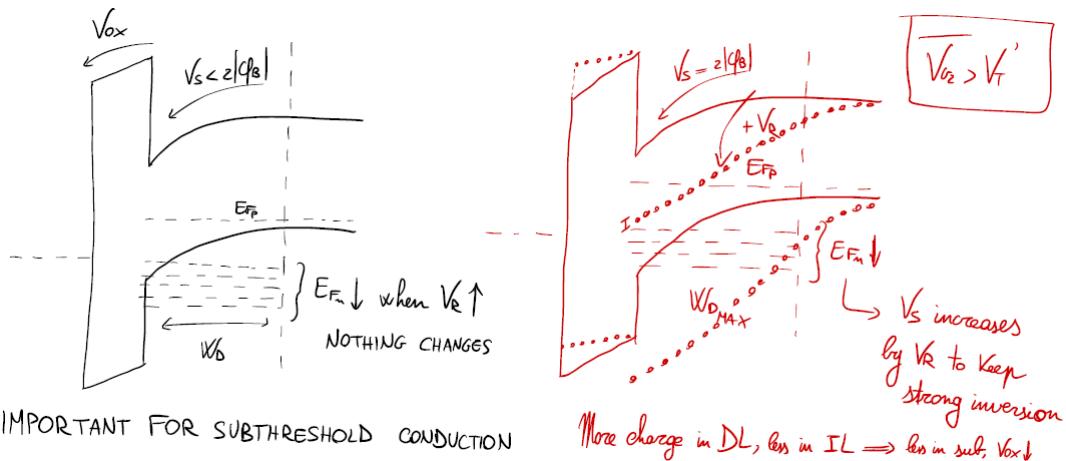
4.12 Dependence of V_s on V_G/V_R

Starting point is 4.24, where we can assume $V_R = 0$; if $V_R > 0$, then

$$\begin{aligned} V'_T &= V_{FB} + 2|\phi_B| + V_R + \frac{\sqrt{2\epsilon_{si}qN_A(2|\phi_B| + V_R)}}{C_{ox}} + \frac{\sqrt{2\epsilon_{si}qN_A2|\phi_B|}}{C_{ox}} - \frac{\sqrt{2\epsilon_{si}qN_A2|\phi_B|}}{C_{ox}} = \\ &= V_T + V_R + \frac{\sqrt{2\epsilon_{si}qN_A}}{C_{ox}} \left[\sqrt{2|\phi_B| + V_R} - \sqrt{2|\phi_B|} \right] \end{aligned} \quad (4.46)$$



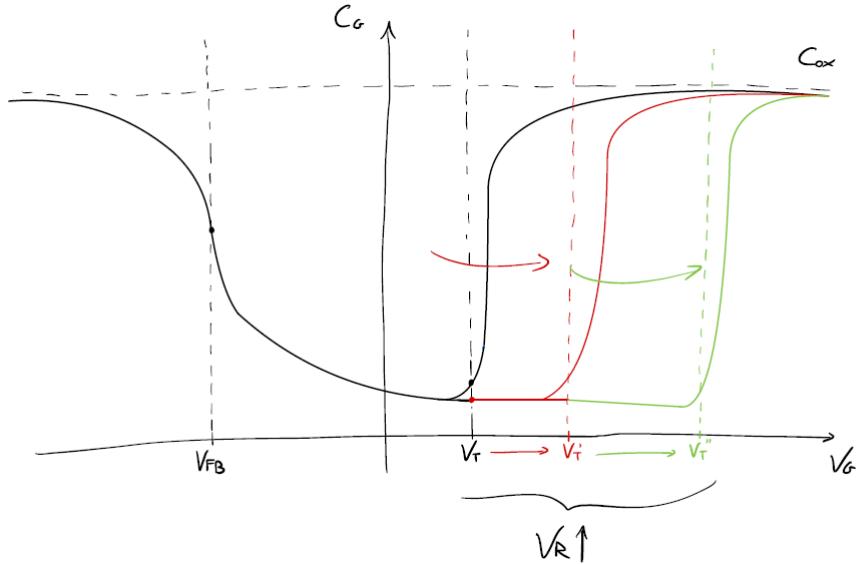
Let's see in detail what's the band bending situation.



If we were to calculate Q_{INV} :

$$Q_{\text{INV}} = Q_s - Q_{\text{DEP}} = -C_{ox}(V_G - V_{FB} - 2|\phi_B| - V_R) - Q_{\text{DEP}} = -C_{ox}(V_G - V'_T) \quad (4.47)$$

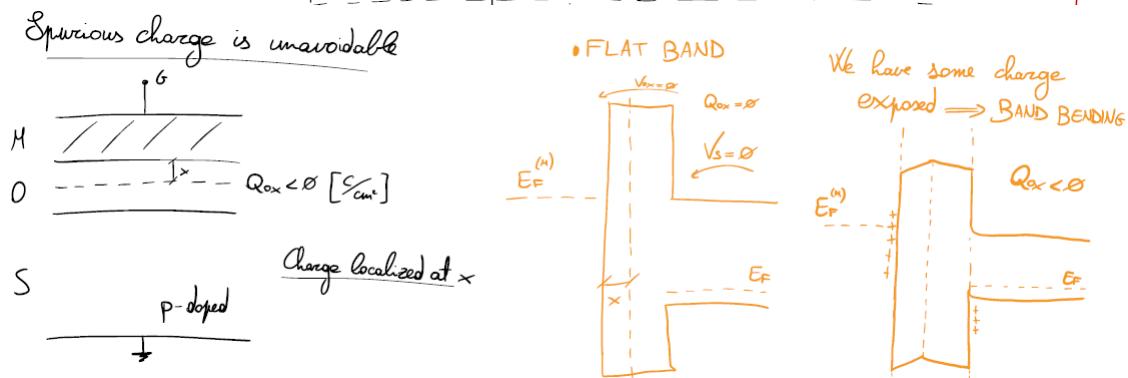
At a certain point $V'_T = \bar{V}_{G_2}$, so V_s will eventually stop increasing because $Q_{\text{INV}} = 0$: we enter depletion/weak inversion regime. C - V curve would look like this:



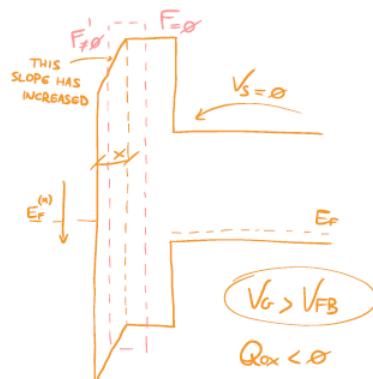
Decreasing trend is prolonged along with V'_T .

4.13 Impact of fixed oxide charge and of interface states on the silicon/oxide interface

Until now we have been considering SiO_2 as a perfect insulator, but now it is time to introduce its defects. First we study the spurious charge Q_{ox} , which can be either positive or negative.



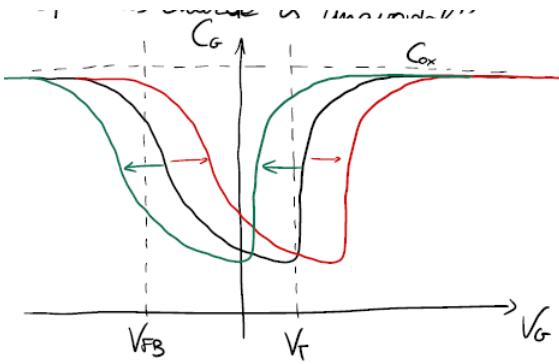
Band slope changes at x , where the charge is located; flat band condition can be partially restored by moving $E_F^{(M)}$ downwards whereas charge doesn't change with gate voltage.



By studying the electrostatics where the slope has increased we can say

$$\frac{Q_{ox}}{\epsilon_{ox}} = -\frac{\Delta V_G}{x} = F' \quad \Rightarrow \quad \Delta V_G = -\frac{Q_{ox}x}{\epsilon_{ox}} = -\frac{Q_{ox}}{C_{xG}} \quad (4.48)$$

where C_{xG} is the capacitance between the gate and x ; ΔV_G is maximum when the charge is exactly at the interface.



- Without Q_{is}
- With $Q_{ox} < 0$
- With $Q_{ox} > 0$

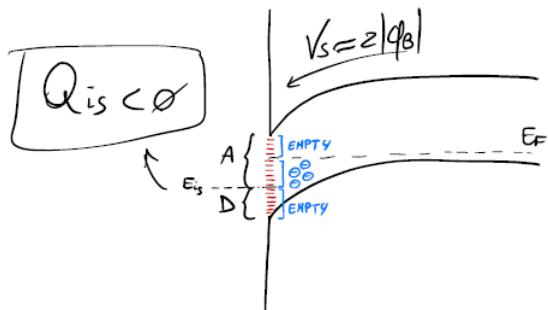
Typically ΔV_G is called ΔV_{FB} or ΔV_T

C-V curve is shifted
To the right when $Q_{ox} < 0$
To the left when $Q_{ox} > 0$

We can also rewrite ΔV_G as a general law:

$$\Delta V_G = \int_0^{t_{ox}} \frac{x \rho_{ox} dx}{\epsilon_{ox}} \quad (4.49)$$

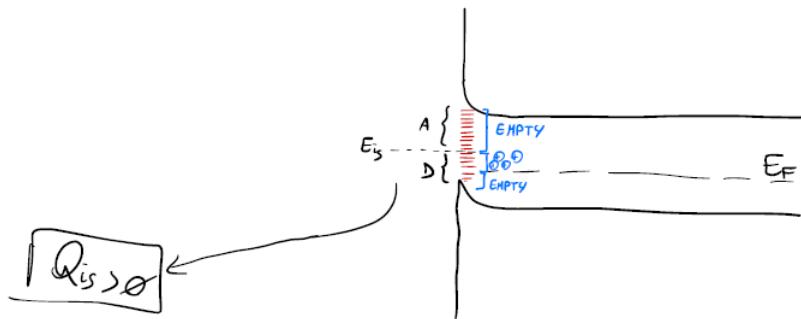
Now let's take a look at the interface states impact, beginning from strong inversion regime.



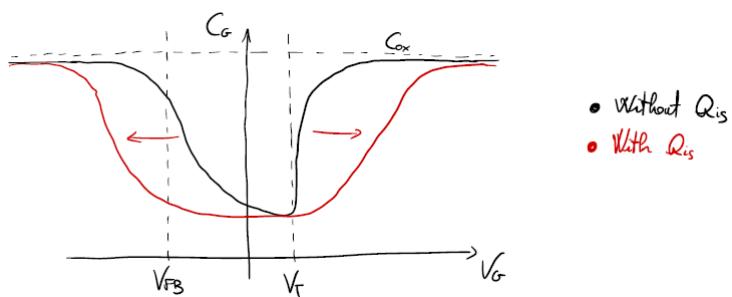
As usual, we have to adjust the band diagram, as there should be a downwards shift.

$$\Delta V_G = -\frac{Q_{is}}{\epsilon_{ox}} t_{ox} = -\frac{Q_{is}}{C_{ox}} \quad (4.50)$$

In accumulation regime we have a different scenario:



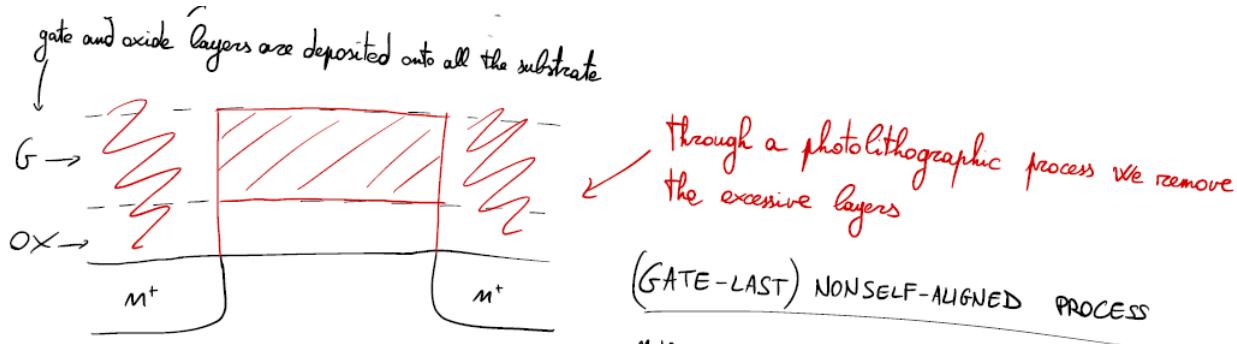
Q_{is} switches sign depending on the working regime.



Interface states give a different contribution to the capacitance according to the frequency of the device; all of this is valid for nMOS transistors too.

4.14 Polysilicon gate

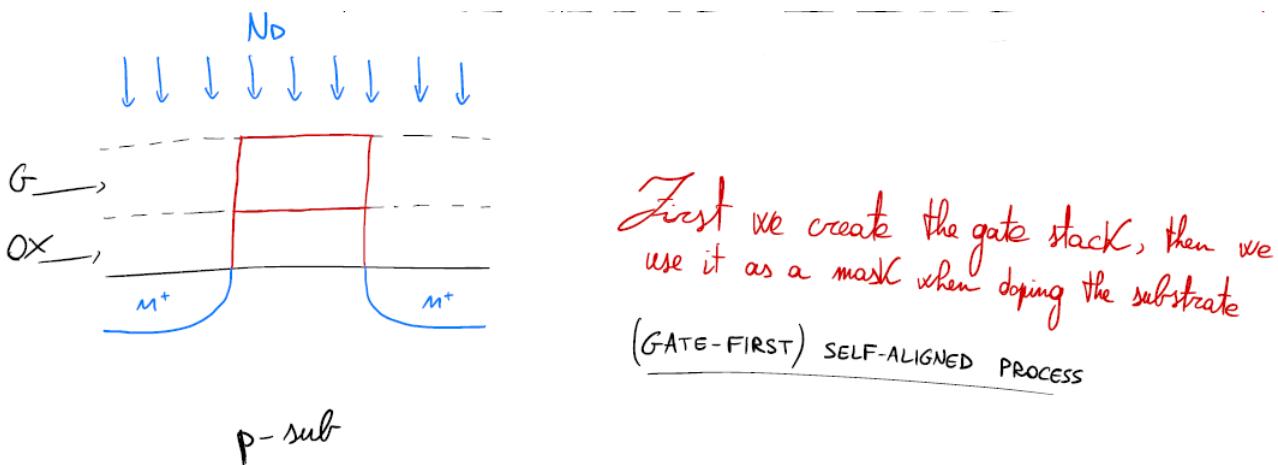
Polysilicon gate played a very important role in the history of the MOS as it introduced the possibility to use self-aligned processes on the silicon wafer; let's start from the beginning, by seeing how we can integrate a MOS capacitor in a ring.



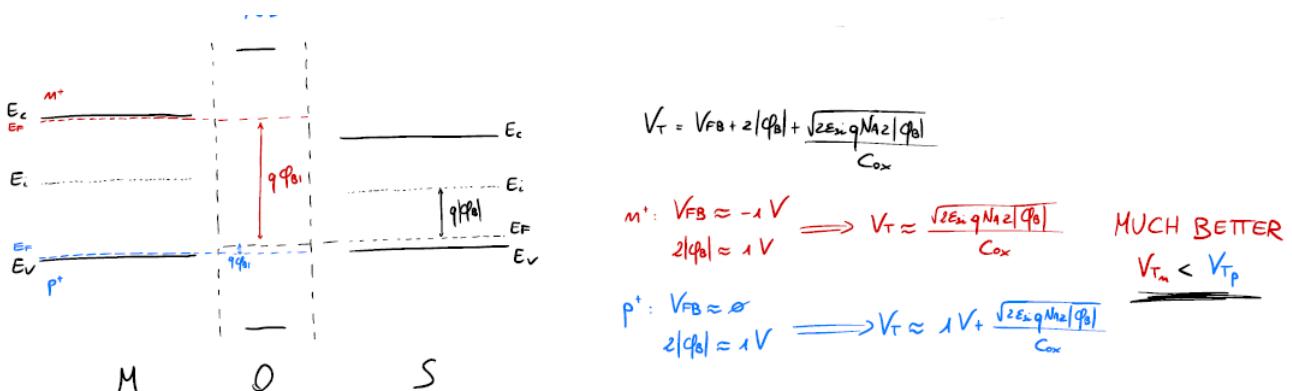
Nothing assures we have a perfect alignment between the n^+ -region and the gate stack: in fact, there are always some misplacements leading to poor performances.



We surely need a better method.



A major issue with this technique is that, since for every process dopants are activated through a high-temperature treatment, the gate (typically Al) actually melts. That is why we need to introduce a **polysilicon gate**, something that also caused the birth of the scaling process for the MOS. First of all, we want the polysilicon to work as a metal, therefore it must have a large doping concentration.



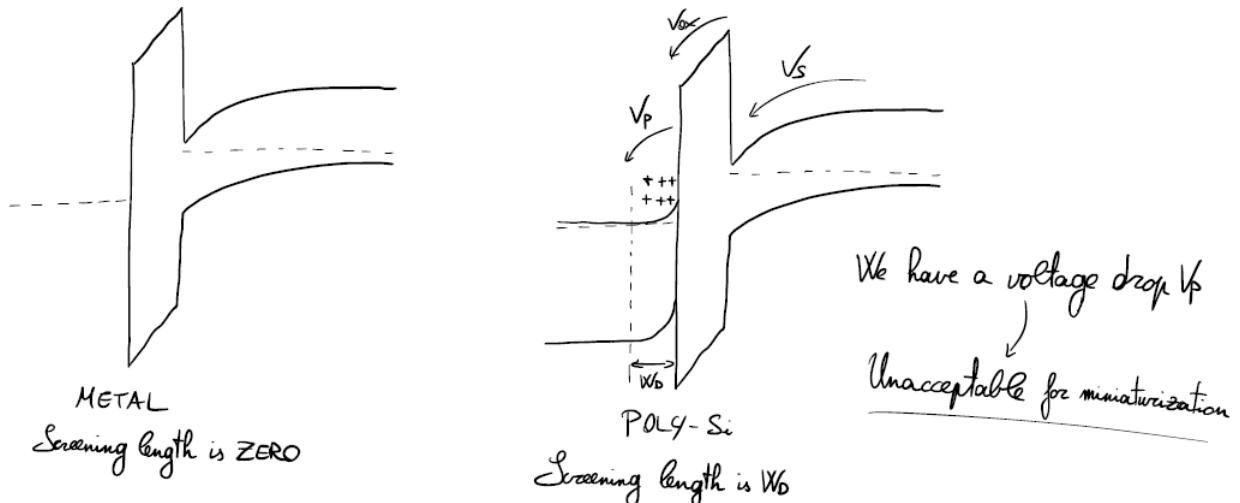
For a p -doped substrate, the best solution is to use a n^+ gate, while for a n -doped substrate it is of course the opposite. There are also two additional advantages in using a polysilicon gate:

1. Had we used a metal, we would have to choose a specific one for both devices, while polysilicon is *universal*;

2. Metal is less stable than polysilicon, because it has a lot of impurities.

Then why in the last 15 years was polysilicon replaced with metals (TiM, TaM)?

Main reason is that, no matter its beneficial aspects, polysilicon is still not a metal, so its free carriers concentration is not comparable to a metal one. Also, it has not the screening capability of a metal: a voltage drop V_p is needed to screen the electric fields coming from the oxide.



This additional voltage drop also introduces a capacitance in series with C_s and C_{ox} :

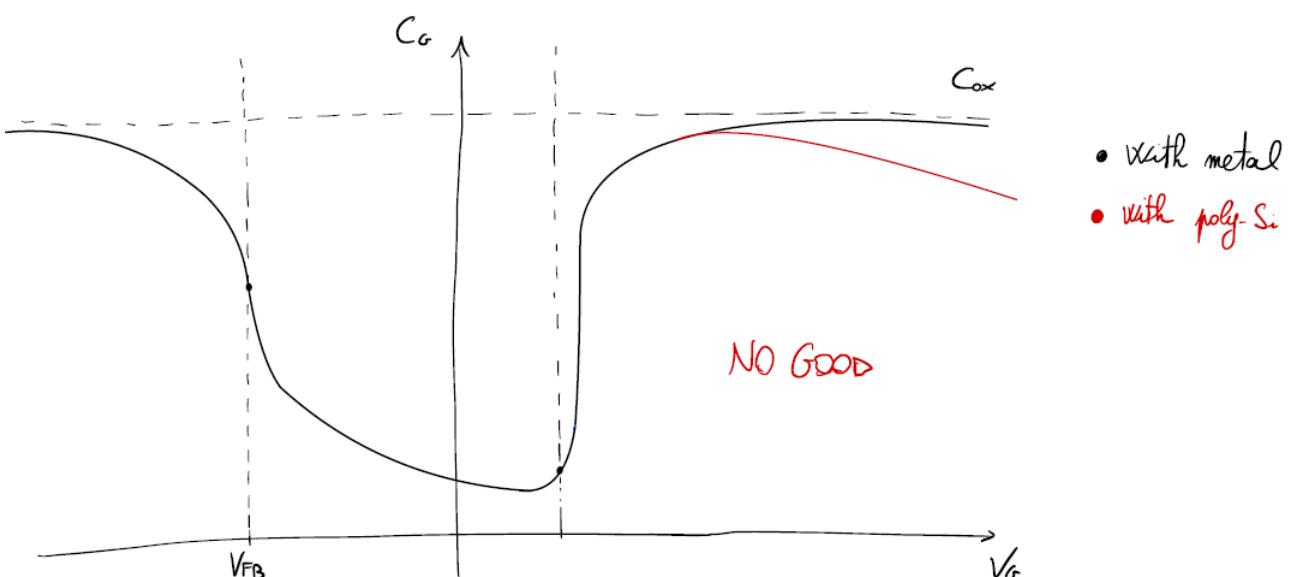
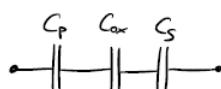
$$V_G - V_{FB} = V_s + V_{ox} + V_p$$

↓

$$\frac{dV_G}{d(-Q_s)} = \frac{dV_s}{d(-Q_s)} + \frac{1}{C_{ox}} + \frac{dV_p}{d(-Q_s)} \quad (4.51)$$

↓

$$\frac{1}{C_G} = \frac{1}{C_s} + \frac{1}{C_{ox}} + \frac{1}{C_p}$$

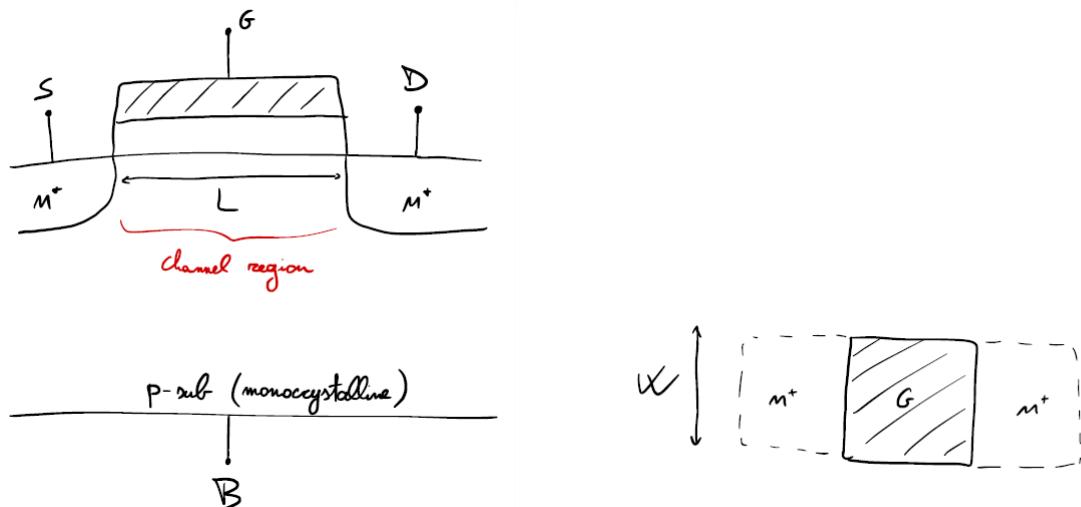


Chapter 5

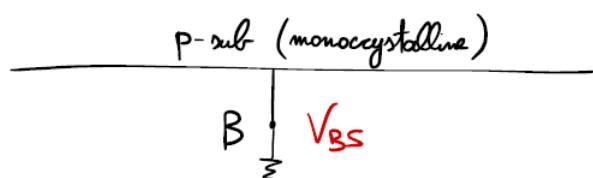
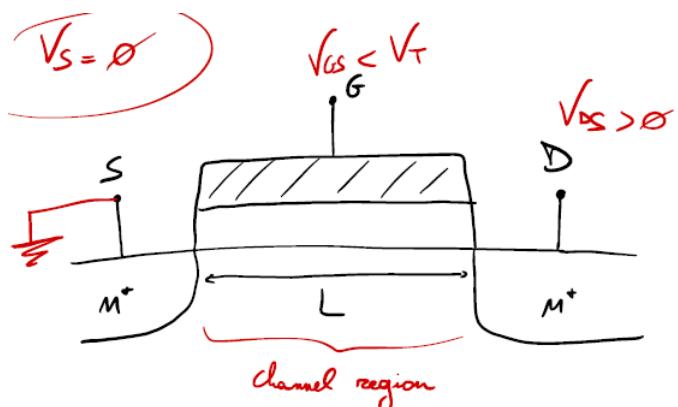
MOS transistor

5.1 Basics of the MOS transistor

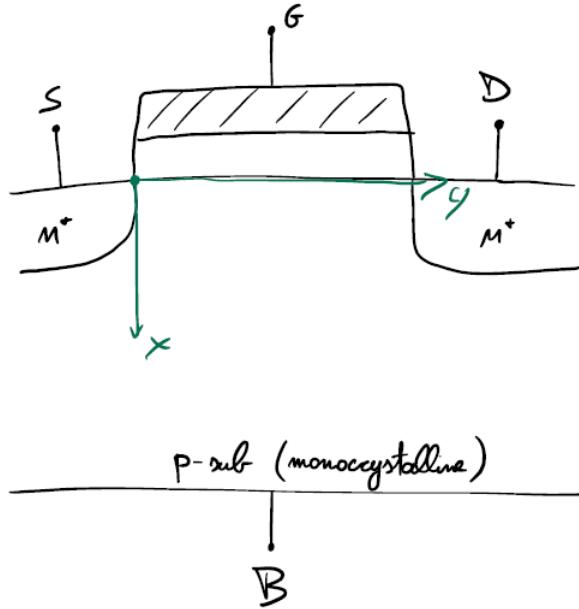
This is probably the most important device; as the MOS capacitor, we are considering a perfect SiO_2 insulator and an ideal metal, but this time the n^+ -regions are independent.



This is of course a bidimensional device with **four** terminals. Let's polarize our transistor by applying these bias voltages: $V_S = 0$, $V_{GS} < V_T$, $V_{DS} > 0$, $V_{BS} = 0$.



We mainly focus on the channel region, the most important one; since $V_{GS} < V_T$, we are working in the depletion/weak inversion regime, so the electron concentration in the channel region is very low. That part of the transistor then works as a bottleneck, determining a small current to flow. In order to get a larger current we may want to increase V_{GS} : current can be in fact controlled by the gate terminal (so can be the electrostatics in the channel region), making the MOS a **Field Effect Transistor**. In this case, electrons are the most important carriers (unipolar transistor). Now let's do some useful approximations.



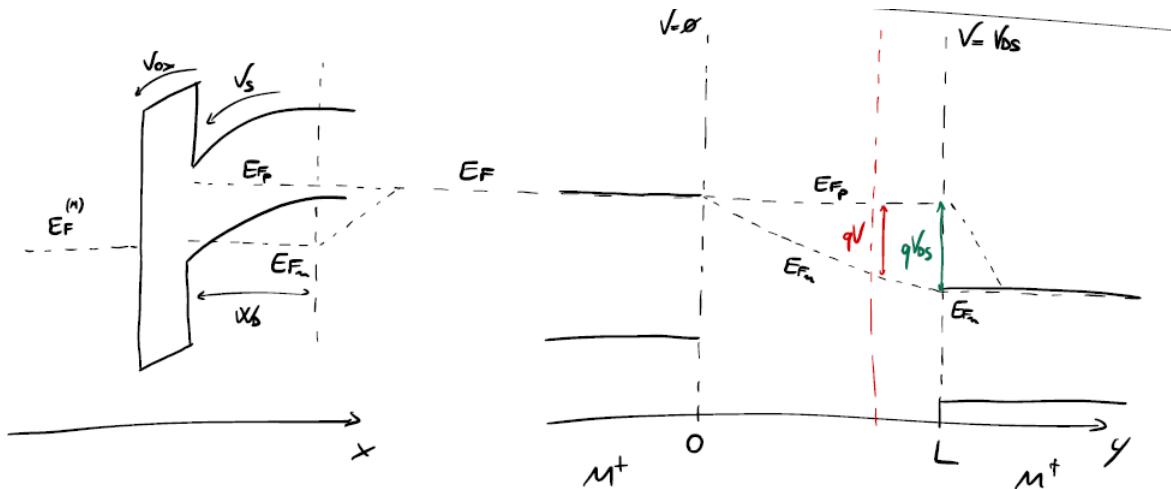
Starting point is of course the 2D Poisson equation:

$$\begin{cases} \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = -\frac{q}{\epsilon_{Si}}(p - n + N_D^+ - N_A^-) \\ \phi = \phi(x, y) \end{cases} \quad (5.1)$$

If we assume to have a *long-channel* MOSFET we can actually introduce the **gradual channel approximation**

$$F_y \ll F_x \quad (5.2)$$

meaning that the horizontal electrostatics contribution is negligible compared to the vertical one (quasi 1D electrostatics).



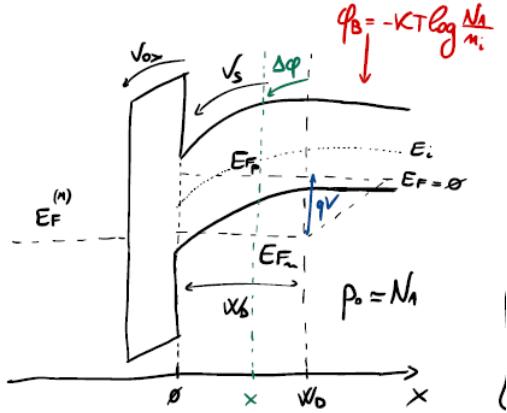
Moving through the cross-section, E_{F_n} changes according to the position, so

$$V(y) = -\frac{E_{F_n}}{q} \quad (5.3)$$

where $V(y)$ is the separation between the two quasi-Fermi levels.

5.2 Electrostatics in the channel region

Electrostatic analysis is the same as in the MOS capacitor, but this time the substrate is not under thermodynamic equilibrium.



4.6 and 4.7 are still valid, but this time

$$p_0 n_0 = n_i^2 \exp\left(\frac{E_{F_n} - E_{F_p}}{KT}\right) = n_i^2 \exp\left(-\frac{qV}{KT}\right) \quad (5.4)$$

Let's consider the onset for strong inversion ($n = N_A$) at $x = 0$:

$$\begin{aligned} n = N_A &= \frac{n_i^2}{N_A} \exp\left(\frac{qV_s}{KT}\right) \exp\left(-\frac{qV}{KT}\right) \\ &\Downarrow \\ \frac{N_A^2}{n_i^2} &= \exp\left(\frac{qV_s}{KT}\right) \exp\left(-\frac{qV}{KT}\right) \\ &\Downarrow \\ \exp\left(\frac{q2|\phi_B|}{KT}\right) &= \exp\left(\frac{qV_s}{KT}\right) \exp\left(-\frac{qV}{KT}\right) \\ &\Downarrow \\ V_s &= 2|\phi_B| + V \end{aligned} \quad (5.5)$$

Following the same steps we did in the last chapter, we eventually get

$$\frac{d^2\Delta\phi}{dx^2} = -\frac{q}{\epsilon_{Si}} \left[N_A \exp\left(-\frac{q\Delta\phi}{KT}\right) - \frac{n_i^2}{N_A} \exp\left(\frac{q\Delta\phi}{KT}\right) \exp\left(-\frac{qV}{KT}\right) - N_A + \frac{n_i^2}{N_A} \right] \quad (5.6)$$

and then

$$Q_s = \pm \sqrt{2\epsilon_{Si} KTN_A} \left[\exp\left(-\frac{qV_s}{KT}\right) + \frac{qV_s}{KT} - 1 + \frac{n_i^2}{N_A^2} \left(e^{-\frac{qV}{KT}} \left(e^{\frac{qV_s}{KT}} - 1 \right) - \frac{qV_s}{KT} \right) \right]^{1/2} \quad (5.7)$$

just to arrive at

$$Q_s \approx -\sqrt{2\epsilon_{Si} KTN_A} \left[\frac{qV_s}{KT} + \frac{n_i^2}{N_A^2} \exp\left(\frac{q(V_s - V)}{KT}\right) \right]^{1/2} = Q_s(V_s, V) \quad (5.8)$$

5.3 Subthreshold and Onstate regimes

Beginning from last statement we can say

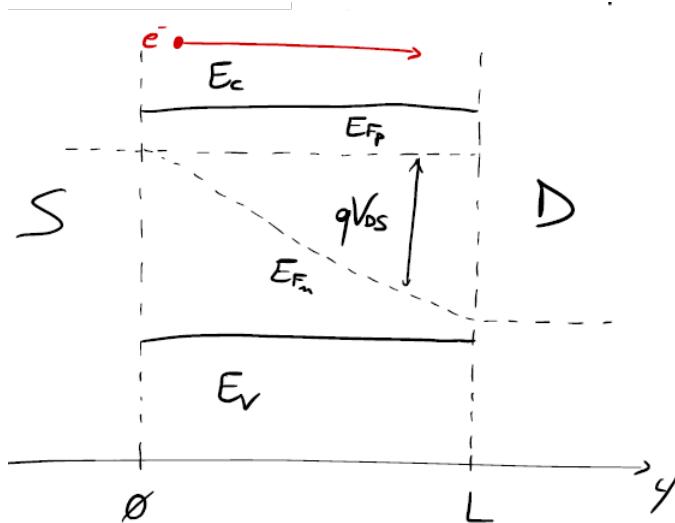
$$Q_s(V_s, V) = Q_{DEP}(V_s, V) + Q_{INV}(V_s, V) \quad (5.9)$$

Assuming we are working in depletion/weak inversion regime it is true that

$$Q_{DEP} = -\sqrt{2\epsilon_{Si} q N_A V_s} \quad (5.10)$$

$$Q_{INV} = -\sqrt{\frac{\epsilon_{Si} q N_A}{2V_s} \frac{KT}{q} \frac{n_i^2}{N_A^2} \exp\left(\frac{q(V_s - V)}{KT}\right)} \quad (5.11)$$

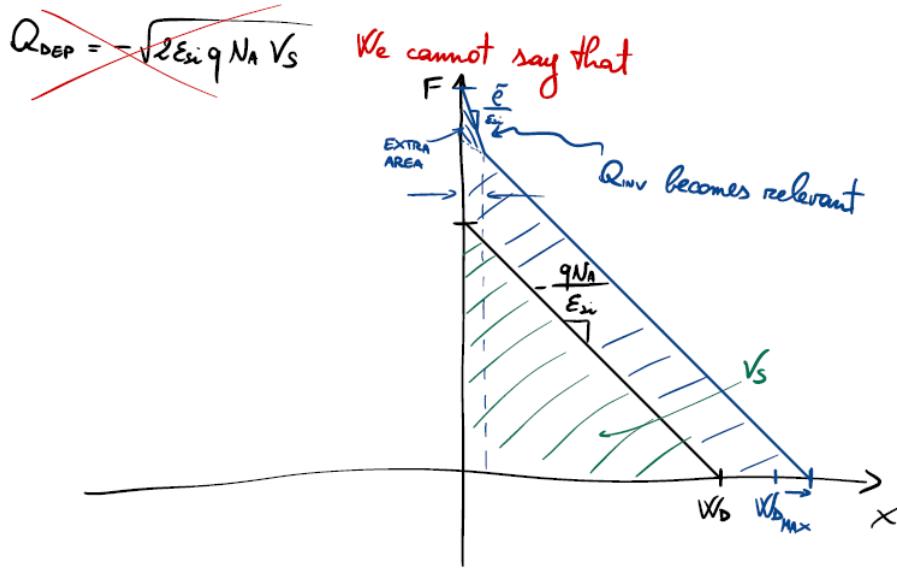
Since electron concentration is negligible for the electrostatics, $V_s \neq V_s(y)$, thus there is no band bending from source to drain.



Electrons move by diffusion

Drift is impossible as bands are almost flat

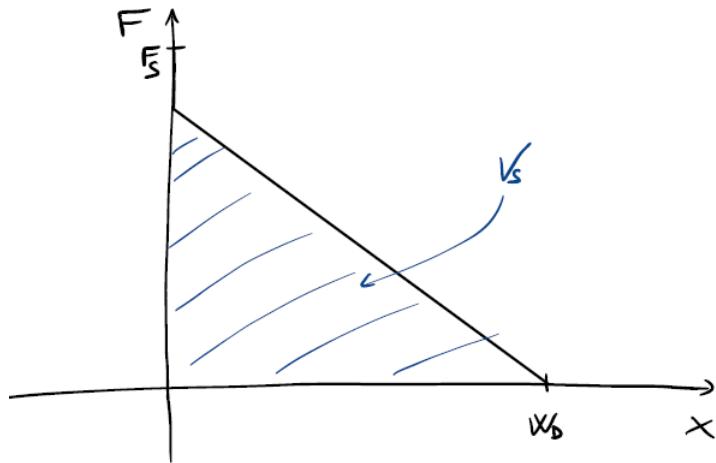
That is the **subthreshold regime**. Now we move to the strong inversion regime, where we actually don't have any expression for Q_{DEP} and Q_{INV} .



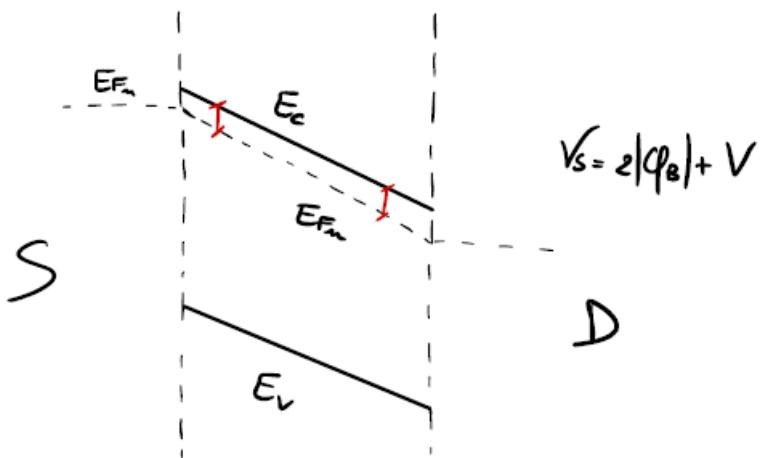
Without any approximation, only in the subthreshold regime we can say that $W_D = \sqrt{\frac{2\epsilon_{Si}}{qN_A} V_s}$ and so

$$Q_{DEP} = -qN_A W_D = -\sqrt{2\epsilon_{Si} q N_A V_s} \quad (5.12)$$

For this purpose we introduce the *charge sheet approximation*, which consists in neglecting the inversion layer width.



Thanks to this, 5.11 is valid, and so we arrive at $Q_{\text{INV}} = Q_s + Q_{\text{DEP}}$. This time $V_s = V_s(y)$, therefore we get a significant band bending along with E_{F_n} .



That is the **onset regime**.

5.4 Continuity equation in the channel

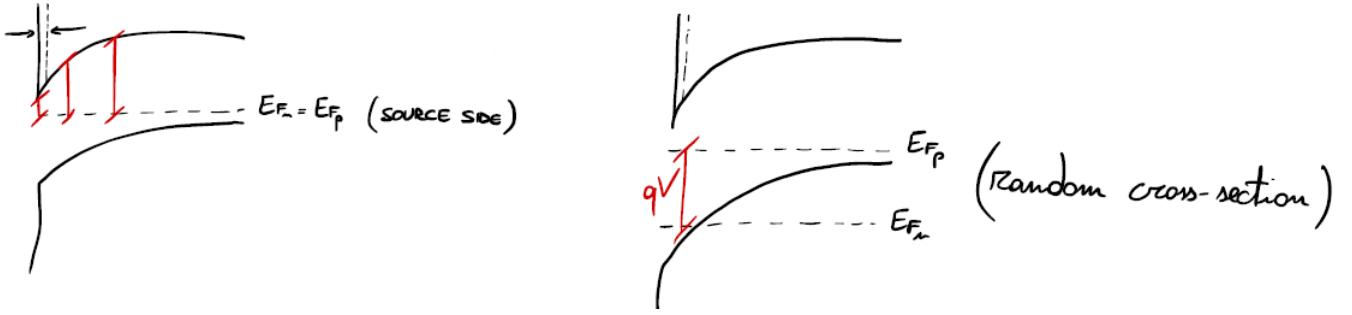
Drain bias shifts E_{F_n} downwards: in this case V dependence is very important, so

$$\begin{cases} V_s = V_s(V_{GS}, V) \\ Q_s = Q_s(V_{GS}, V) \\ Q_{\text{INV}} = Q_{\text{INV}}(V_{GS}, V) \end{cases} \quad (5.13)$$

Poisson equation is not enough because of the non-equilibrium (ϕ and E_{F_n} are unknown): **continuity equation is needed**.

$$J_n = n\mu_n \frac{dE_{F_n}}{dy} = -qn(x, y)\mu_n \frac{dV(y)}{dy} = J_n(x, y) \quad (5.14)$$

We assume stationary conditions: V_{GS} and V_{DS} constant and no G/R processes.



J_n is constant in the transport direction, so $J_n = J_n(x)$:

$$I_{DS} = -W \int_0^{t_{INV}} J_n(x) dx = -W \int_0^{t_{INV}} -qn\mu_n \frac{dV}{dy} dx = -W\mu_n \frac{dV}{dy} \int_0^{t_{INV}} -qndx = -\mu_n W Q_{INV} \frac{dV}{dy} \quad (5.15)$$

That is the differential form for the current; we also want the integral form, thus

$$\begin{aligned} \int_0^L I_{DS} dy &= \int_0^{V_{DS}} -\mu_n W Q_{INV} dV \\ &\Downarrow \\ I_{DS} L &= -\mu_n W \int_0^{V_{DS}} Q_{INV} dV \quad (5.16) \\ &\Downarrow \\ I_{DS} &= -\mu_n \frac{W}{L} \int_0^{V_{DS}} Q_{INV}(V) dV \end{aligned}$$

All we are left to find is Q_{INV} as a function of V .

5.5 Ohmic/parabolic regime

Assuming we are in the onstate regime, let's first regroup the most important equations:

$$\left\{ \begin{array}{l} Q_s = Q_s(V_s, V) \\ V_{GS} - V_{FB} = V_s - \frac{Q_s}{C_{ox}} \\ I_{DS} = -\mu_n \frac{W}{L} \int_0^{V_{DS}} Q_{INV}(V) dV \\ V_s \approx 2|\phi_B| + V \\ Q_s = -C_{ox}(V_{GS} - V_{FB} - V_s) \end{array} \right.$$

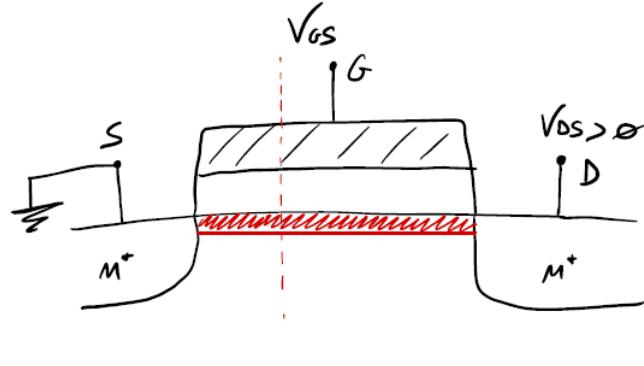
We would get the inversion layer charge to be

$$Q_{INV} = -C_{ox}(V_{GS} - V_{FB} - 2|\phi_B| - V) + \sqrt{2\epsilon_{Si}qN_A(2|\phi_B| + V)} = Q_{INV}(V) \quad (5.17)$$

Two cases are to be considered: $V_{DS} \ll 2|\phi_B|$ and $V_{DS} < 2|\phi_B|$. We start from the first:

$$Q_{INV} = -C_{ox} \left[(V_{GS} - V_{FB} - 2|\phi_B|) - \frac{\sqrt{2\epsilon_{Si}qN_A 2|\phi_B|}}{C_{ox}} \right] = -C_{ox}(V_{GS} - V_T) \quad (5.18)$$

As we can see $Q_{INV} \neq Q_{INV}(V)$, thus it is constant along the channel.

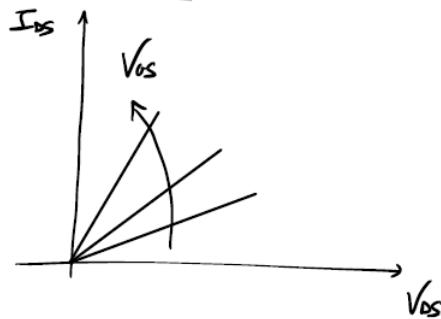


5.16 is then

$$I_{DS} = -\mu_n \frac{W}{L} \int_0^{V_{DS}} -C_{ox}(V_{GS} - V_T) dV = \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_T) V_{DS} \quad (5.19)$$

A linear proportionality exists between I_{DS} and V_{DS} ; the channel shows then a resistive behavior.

$$R_{CH} = \left(\frac{\partial I_{DS}}{\partial V_{DS}} \right)^{-1} = \frac{1}{\mu_n C_{ox} \frac{W}{L} (V_{GS} - V_T)} = \rho_{sh}^{\text{CH}} \frac{L}{W} \quad (5.20)$$



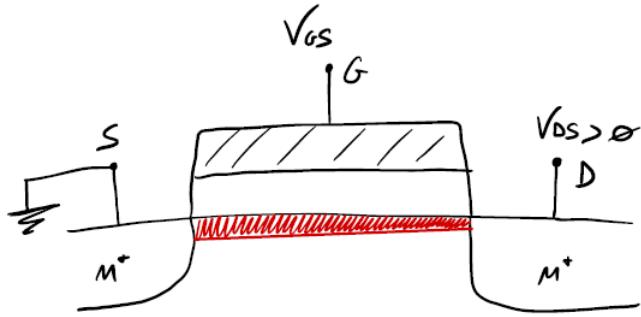
Charge is not modified: this is the **ohmic regime**.

Now we talk about the second case, beginning from rearranging Q_{INV} :

$$\begin{aligned} Q_{\text{INV}} &= -C_{ox}(V_{GS} - V_{FB} - 2|\phi_B| - V) + \sqrt{2\epsilon_{si}qN_A2|\phi_B|} \sqrt{1 + \frac{V}{2|\phi_B|}} \stackrel{\text{TAYLOR}}{=} \\ &\stackrel{\text{TAYLOR}}{=} -C_{ox}(V_{GS} - V_{FB} - 2|\phi_B| - V) + \sqrt{2\epsilon_{si}qN_A2|\phi_B|} \left(1 + \frac{1}{2} \frac{V}{2|\phi_B|} \right) = \\ &= -C_{ox} \left(V_{GS} - V_{FB} - 2|\phi_B| - V - \frac{\sqrt{2\epsilon_{si}qN_A2|\phi_B|}}{C_{ox}} - \sqrt{\frac{2\epsilon_{si}qN_A2|\phi_B|}{4(2|\phi_B|)^2}} \frac{V}{C_{ox}} \right) = \quad (5.21) \\ &= -C_{ox} \left[V_{GS} - V_{FB} - 2|\phi_B| - \frac{\sqrt{2\epsilon_{si}qN_A2|\phi_B|}}{C_{ox}} - V \left(1 + \frac{C_{\text{DEP}}}{C_{ox}} \right) \right] = \\ &= -C_{ox}(V_{GS} - V_T - \textcolor{red}{m}V) = -C_{ox}(V_{GS} - V'_T) \end{aligned}$$

The factor m for now does not retain any physical meaning: it is just a characteristic of the device¹. What is truly interesting, though, is the fact that Q_{INV} decreases when approaching the drain.

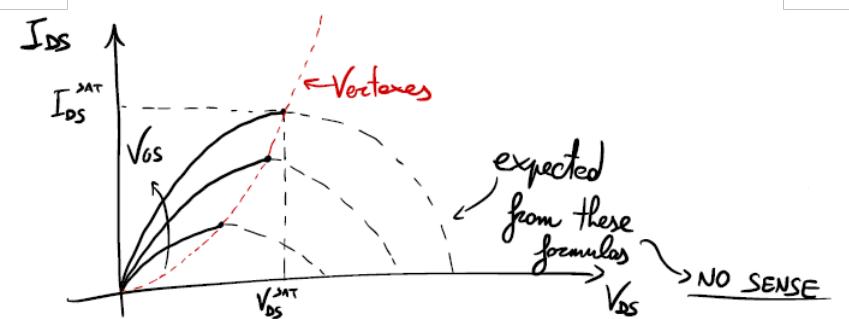
¹It is interesting to note that $V'_T = V_T + mV$ just like for the MOS capacitor with ring it was $V'_T = V_T + V_R$



Eventually 5.16 changes as

$$I_{DS} = -\mu_n \frac{W}{L} \int_0^{V_{DS}} -C_{ox}(V_{GS} - V_T - mV)dV = \mu_n C_{ox} \frac{W}{L} \left[(V_{GS} - V_T)V_{DS} - m \frac{V_{DS}^2}{2} \right] \quad (5.22)$$

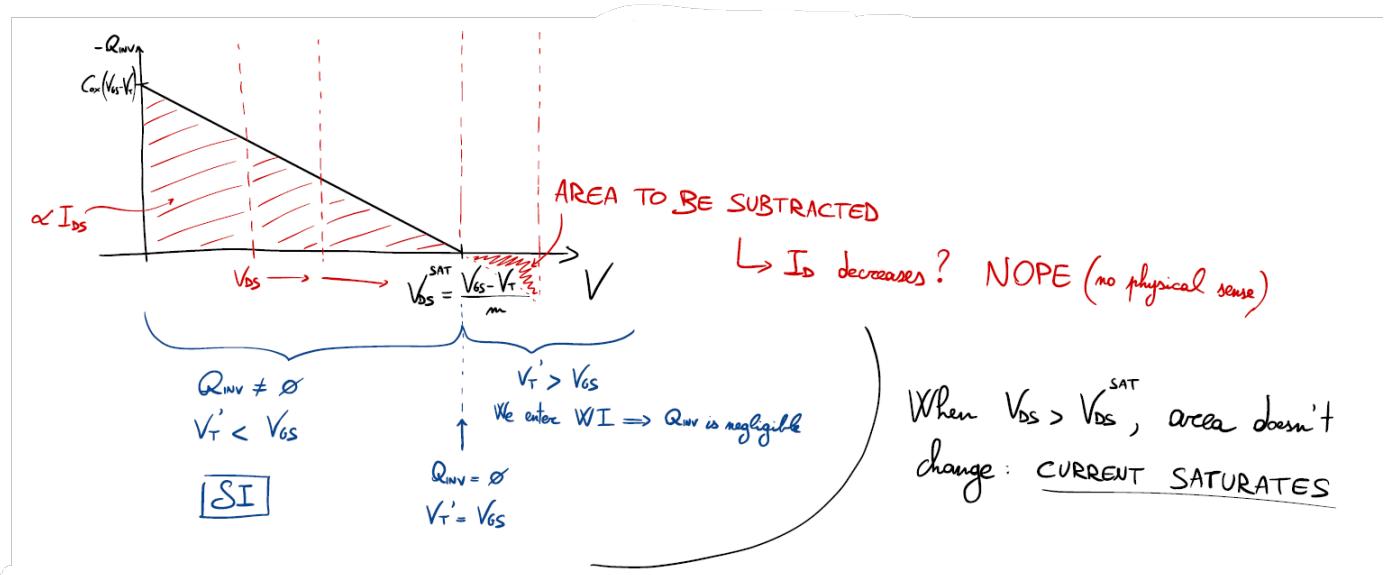
A squared term appears: we get a parabolic behavior with vertexes located in $\left(V_{DS}^{\text{SAT}} = \frac{V_{GS} - V_T}{m}; \quad I_{DS}^{\text{SAT}} = \mu_n C_{ox} \frac{W}{L} \frac{m(V_{DS}^{\text{SAT}})^2}{2} \right)$.

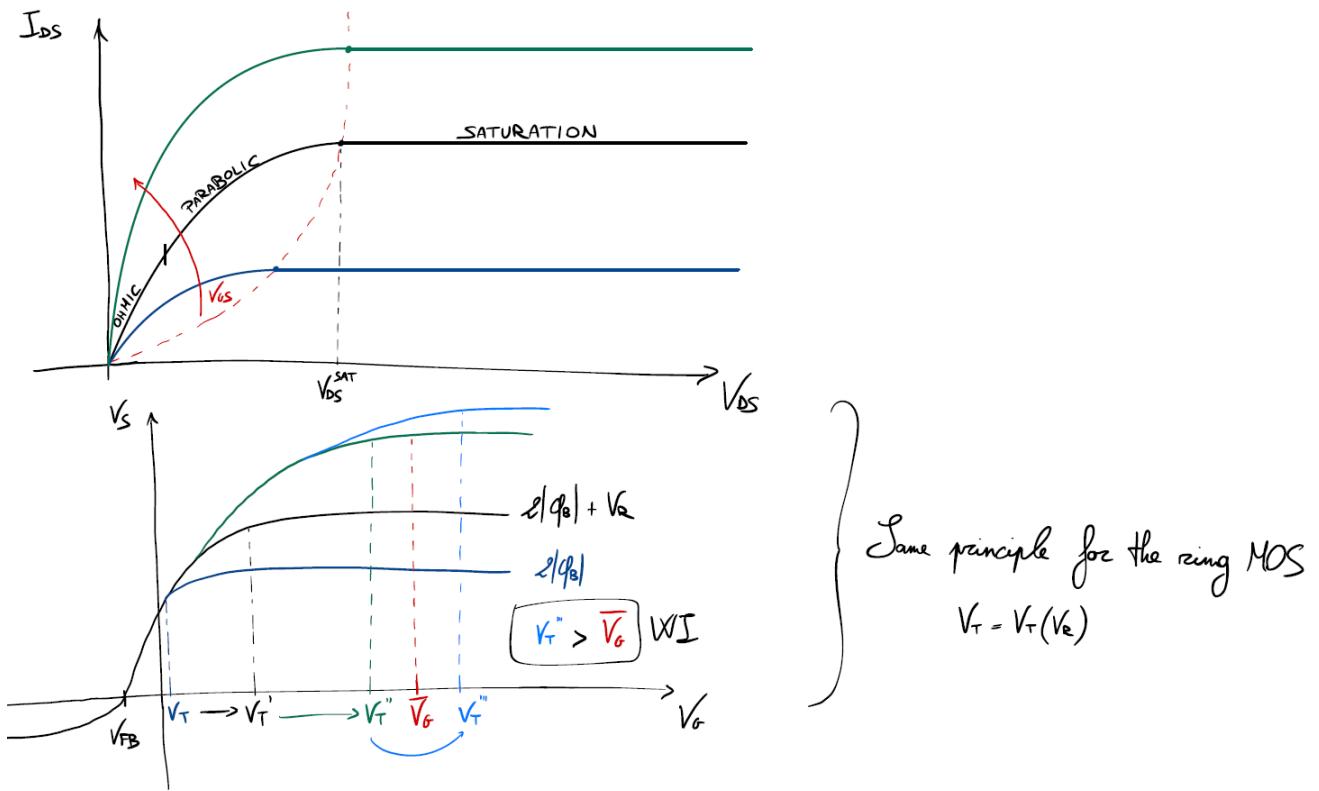


When V_{DS} increases, current increases less because charge in the inversion layer becomes fewer ($R_{ch} \uparrow$); clearly all of this actually makes sense until $V_{DS} \leq V_{DS}^{\text{SAT}}$ is true.

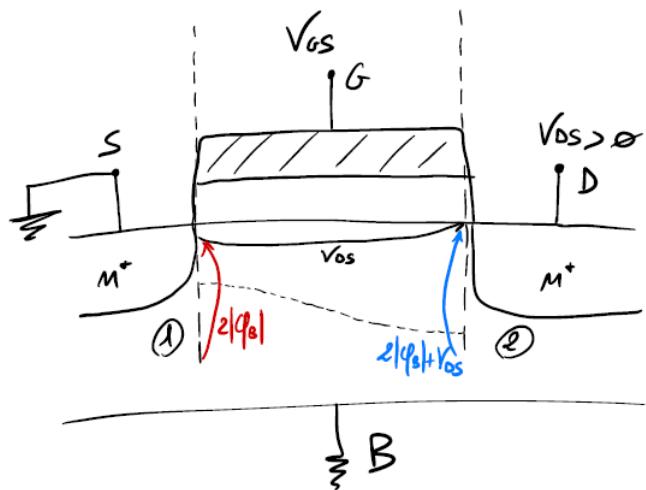
5.6 Saturation regime

We want to know what happens when $V_{DS} > V_{DS}^{\text{SAT}}$; let's start from Q_{inv} :

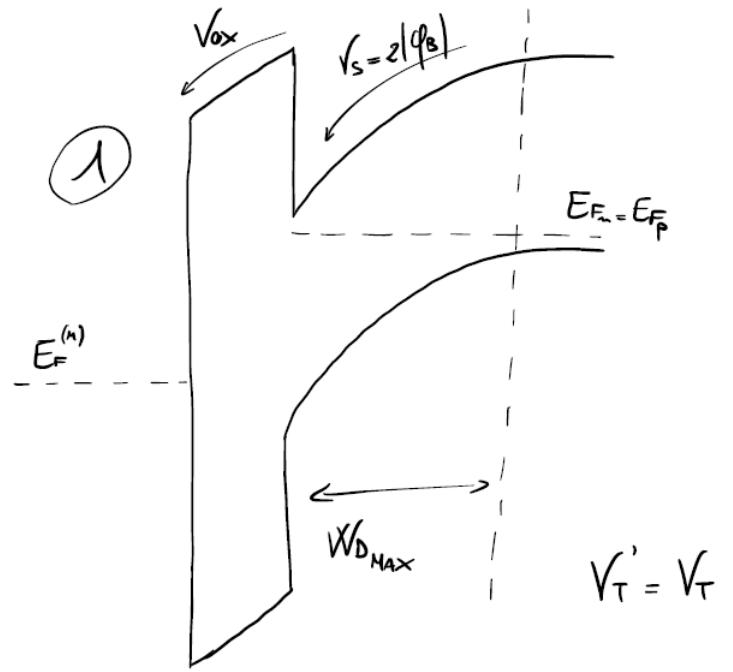




Let's see in detail what happens to the electrostatics along the channel:



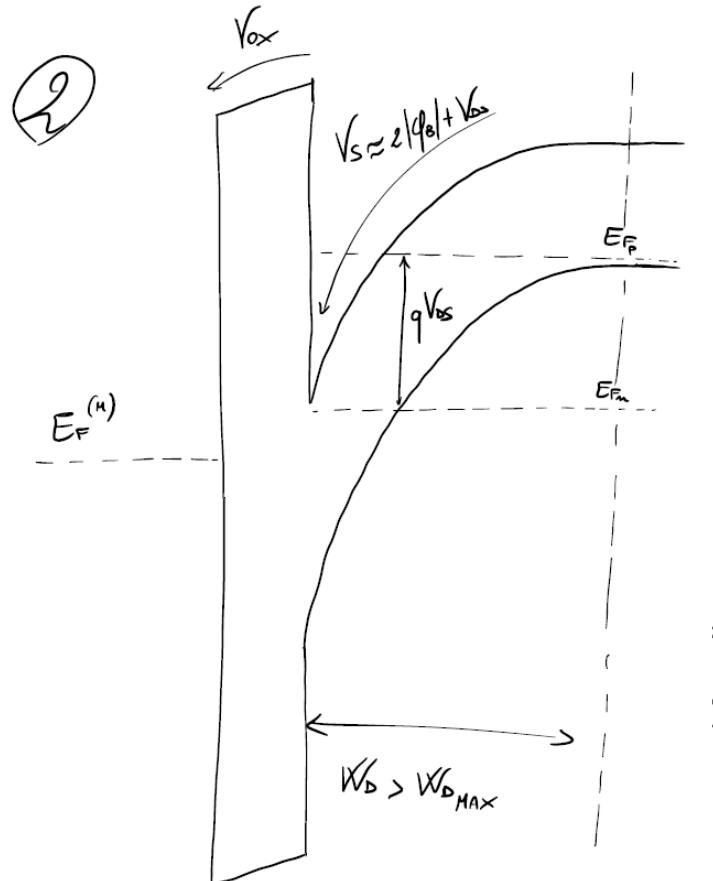
At the source side we have still thermodynamic equilibrium.



At the drain side E_{F_n} is shifted downwards by qV_{DS} and band bending in the oxide decreases. At the same time, depletion layer grows while inversion layer shrinks. Local threshold voltage rises too, like

$$V'_T = V_T + mV_{DS} \quad (5.23)$$

until at a certain point $V'_T = V_{GS}$ and $Q_{INV} = 0$: **drain loses strong inversion condition**. This is the pinch-off condition.



We can conclude that V_{DS}^{SAT} is the maximum voltage drop possible between drain and source; in saturation, drain loses control over channel electrostatics.

5.7 Band diagram along the channel in the onstate regime

We have seen that at the drain we get into weak inversion regime; it is time to refine the analysis, starting from V dependency on y (we integrate the current for a general V).

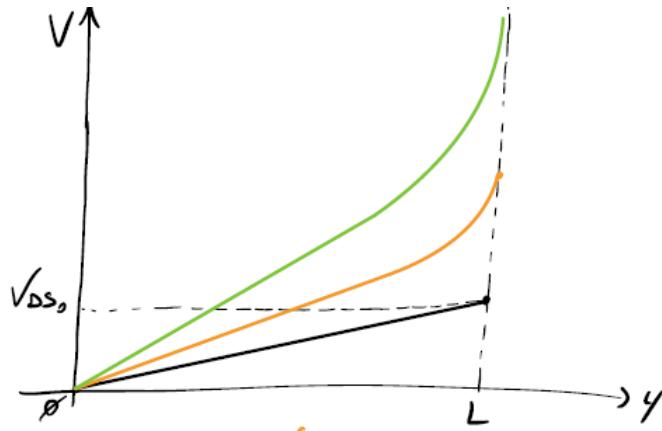
$$I_{DS} = \mu_n C_{ox} \frac{W}{y} \left[(V_{GS} - V_T)V - \frac{mV^2}{2} \right] \quad (5.24)$$

At the drain $y = L$ and that becomes 5.22, so

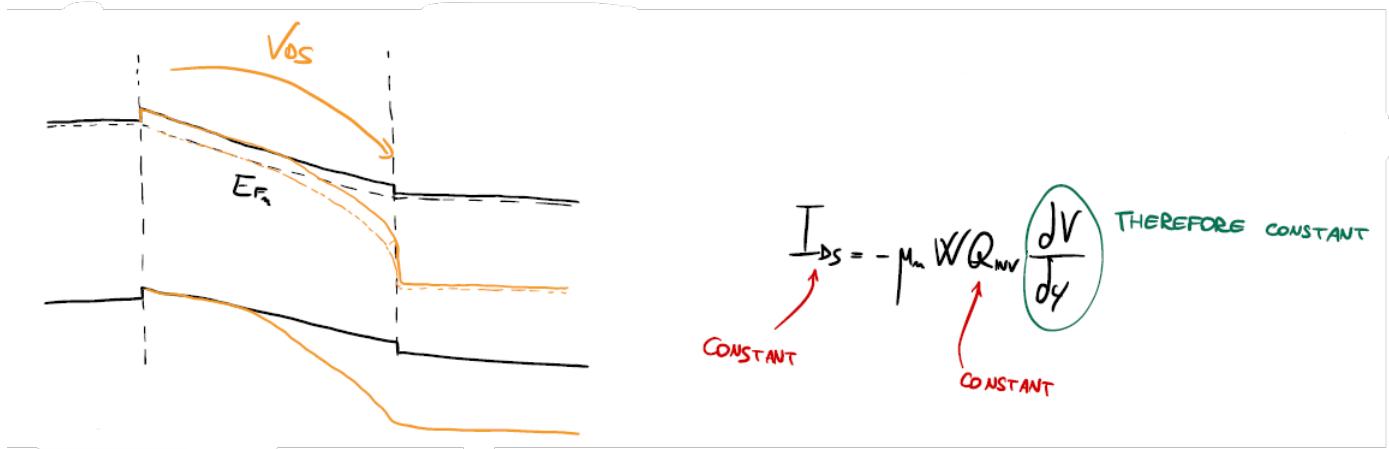
$$V = V(y) = \frac{V_{GS} - V_T}{m} - \sqrt{\left(\frac{V_{GS} - V_T}{m} \right)^2 - \frac{2y}{L} \left(\frac{V_{GS} - V_T}{m} \right) V_{DS} + \frac{y}{L} V_{DS}^2} \quad (5.25)$$

which can be simplified for small V_{DS} as

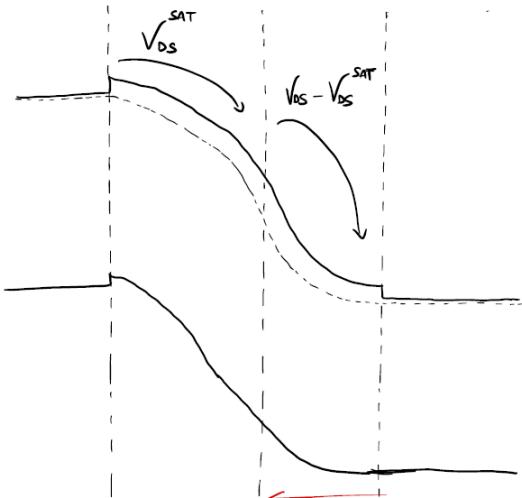
$$V(y) = \frac{V_{DS}}{L} y \quad (5.26)$$



Considering that $V_s = 2|\phi_B| + V$, we can say that bands do shift linearly.

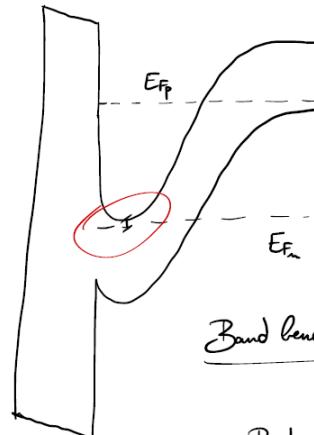
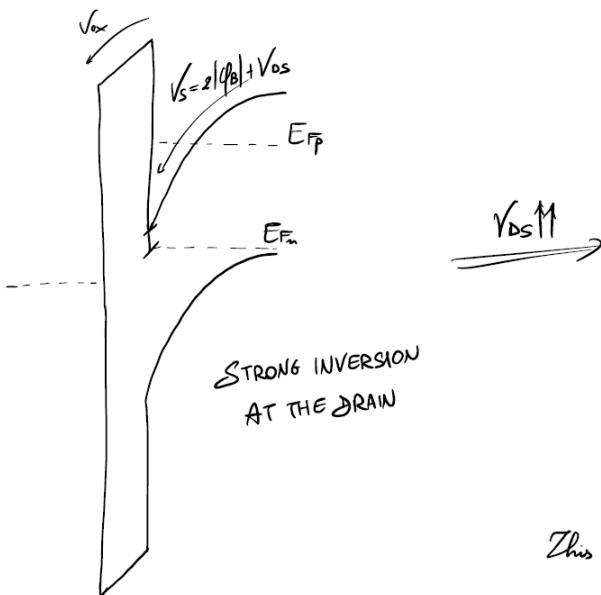


If we consistently increase V_{DS} , since I_{DS} must be kept constant and Q_{INV} is reducing until pinch-off ($Q_{INV} = 0$), E_{F_n} is no longer linearly decreasing when approaching the drain. All of this is valid under gradual channel approximation; in saturation regime under these conditions, though, $F_y \gg F_x$, thus



Close to the drain electrostatics
is strongly bidirectional

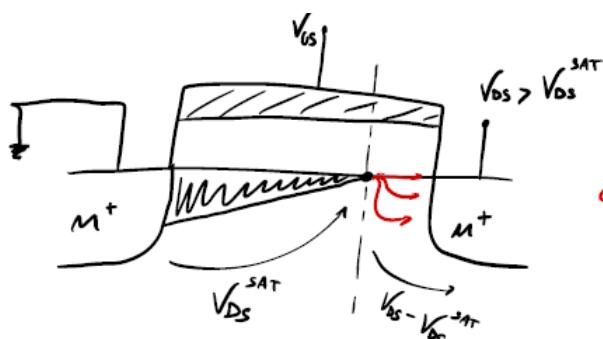
Pinch-off condition moves back to accommodate $V_{DS} - V_{DS}^{SAT}$



This happens because of the bidimensionality

Band bending in the oxide is reversed
↓
Bands at the surface must adapt

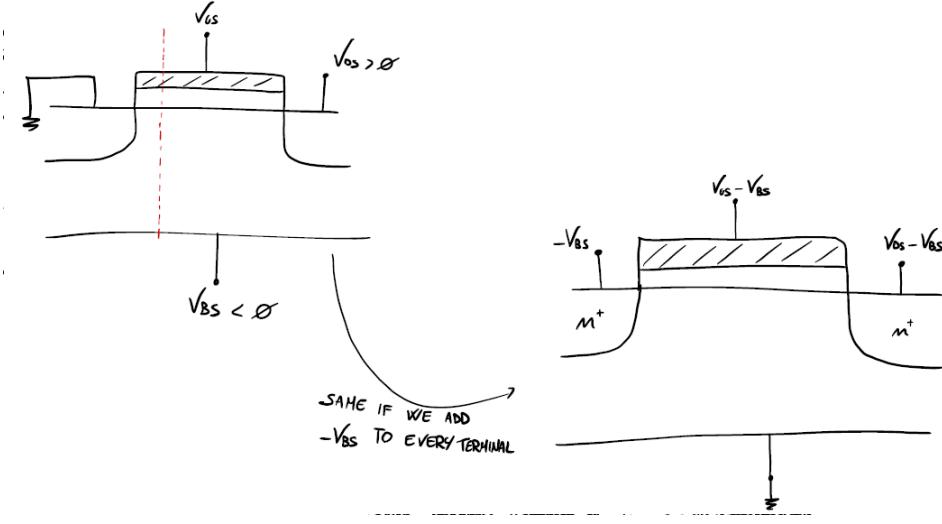
Maximum electron concentration is no longer on the surface.



2D-electrostatics
Current spread in vertical direction
not possible in 1D devices

5.8 Body effect

We now discuss the case where the bulk for some reason is not grounded anymore.



Total charge on the surface is therefore affected as well:

$$Q_s = -C_{ox}(V_{GS} - V_{BS} - V_{FB} - V_s) \quad (5.27)$$

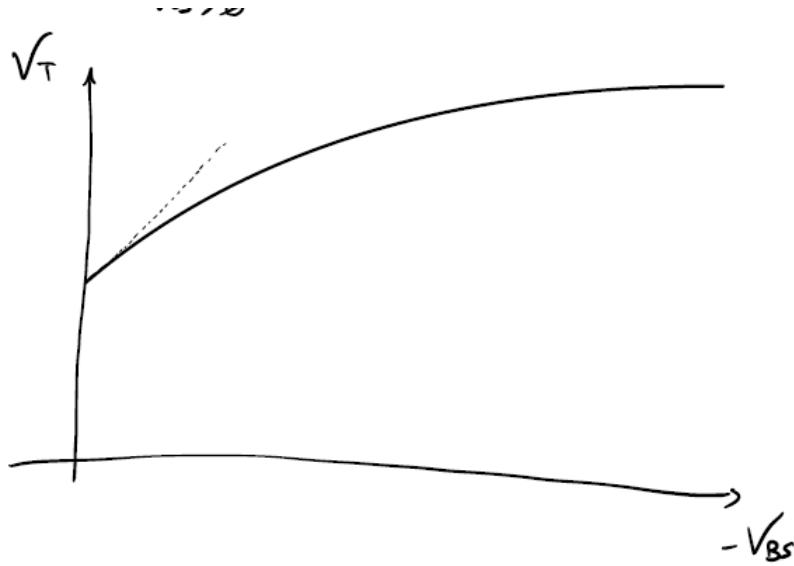
and so is the inversion layer charge:

$$Q_{\text{INV}} = Q_s - Q_{\text{DEP}} = -C_{ox}(V_{GS} - V_{BS} - V_{FB} - 2|\phi_B| - V) + \sqrt{2\epsilon_{\text{Si}}qN_A(2|\phi_B| + V)} \quad (5.28)$$

Considering just the ohmic regime, we can then rewrite 5.28

$$\begin{aligned} Q_{\text{INV}} \approx & -C_{ox}(V_{GS} - V_{BS} - V_{FB} - 2|\phi_B| + V_{BS}) + \sqrt{2\epsilon_{\text{Si}}qN_A(2|\phi_B| - V_{BS})} = \\ & = -C_{ox} \left(V_{GS} - V_{FB} - 2|\phi_B| - \frac{\sqrt{2\epsilon_{\text{Si}}qN_A(2|\phi_B| - V_{BS})}}{C_{ox}} \right) = -C_{ox}(V_{GS} - V_T) \end{aligned} \quad (5.29)$$

Threshold voltage has increased!



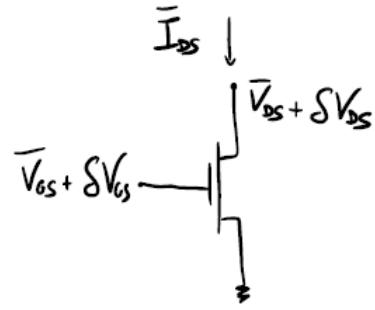
We can introduce a sensitivity parameter:

$$\frac{dV_T}{d(-V_{BS})} \Big|_{V_{BS}=0} = \frac{\sqrt{2\epsilon_{\text{Si}}qN_A}}{C_{ox}} \frac{1}{2} \frac{1}{\sqrt{2|\phi_B| - V_{BS}}} \Big|_{V_{BS}=0} \frac{\sqrt{2\epsilon_{\text{Si}}qN_A}}{C_{ox}} \frac{1}{\sqrt{4 \cdot 2|\phi_B|}} = \frac{C_{\text{DEP}}}{C_{ox}} = m - 1 \quad (5.30)$$

Optimal case is $m = 1$.

5.9 Small signal model

At last we arrived at the small signal model for the MOS transistor; this time two terminals are attached to a signals.

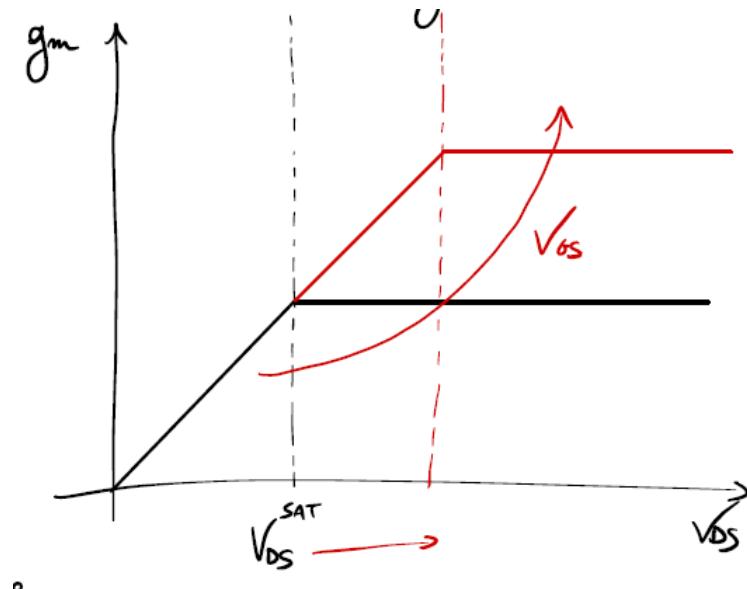


Let's begin by introducing just a gate voltage swing ($\delta V_{DS} = 0$); we need the small signal transconductance

$$g_m = \left(\frac{\partial I_{DS}}{\partial V_{GS}} \right)_{V_{DS}} = \begin{cases} \mu_n C_{ox} \frac{W}{L} V_{DS}, & \text{ohmic regime} \\ \mu_n C_{ox} \frac{W}{L} \frac{V_{GS} - V_T}{m}, & \text{saturation regime} \end{cases} \quad (5.31)$$

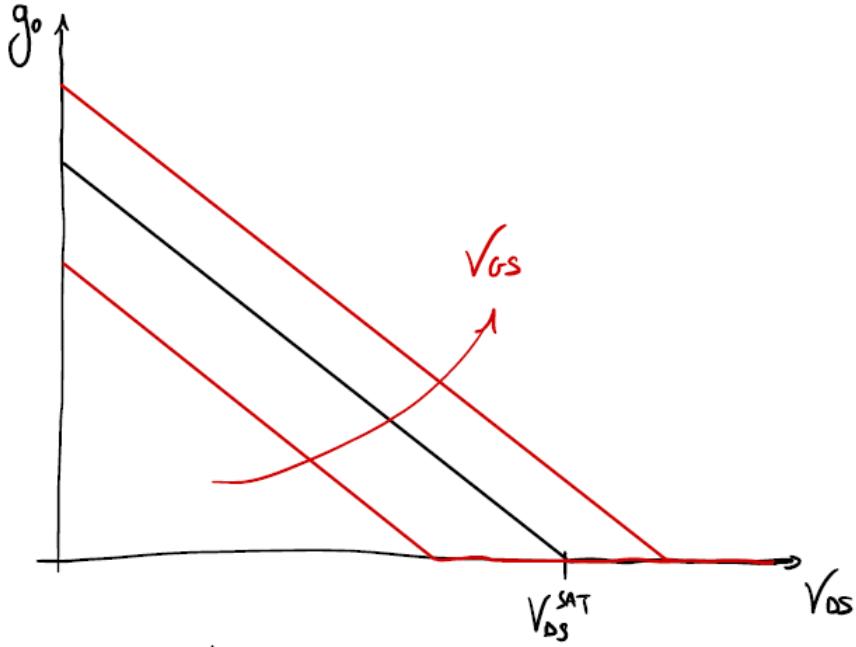
being the current

$$I_{DS} = \begin{cases} \mu_n C_{ox} \frac{W}{L} \left[(V_{GS} - V_T)V_{DS} - \frac{mV_{DS}^2}{2} \right], & \text{ohmic regime} \\ \mu_n C_{ox} \frac{W}{L} \frac{(V_{GS} - V_T)^2}{2m}, & \text{saturation regime} \end{cases} \quad (5.32)$$



Second parameter is the small signal output conductance

$$g_0 = \frac{1}{r_0} = \left(\frac{\partial I_{DS}}{\partial V_{DS}} \right)_{V_{GS}} = \begin{cases} \mu_n C_{ox} \frac{W}{L} [(V_{GS} - V_T) - mV_{DS}], & \text{ohmic regime} \\ 0, & \text{saturation regime} \end{cases} \quad (5.33)$$



Current is dependent on the time-derivative of the voltage applied to the terminals, so we have also some capacitive contributions (intrinsic capacitances come from the inversion layer). First we need the **total** charge in the inversion layer, not just the charge per area, so

$$Q_c = W \int_0^L Q_{\text{INV}}(y) dy = W \int_0^{V_{DS}} Q_{\text{INV}}(y) \frac{dy}{dV} dV$$

This can be known by recalling 5.15:

$$\frac{dy}{dV} = -\frac{\mu_n W Q_{\text{INV}}}{I_{DS}} \quad (5.34)$$

Eventually

$$\begin{aligned} Q_c &= W \int_0^{V_{DS}} -\frac{\mu_n W Q_{\text{INV}}^2}{I_{DS}} dV = -\frac{\mu_n W^2}{I_{DS}} \int_0^{V_{DS}} [-C_{ox}(V_{GS} - V_T - mV)]^2 dV = \\ &= \frac{\mu_n C_{ox}^2 W^2}{I_{DS}} \left[\frac{(V_{GS} - V_T - mV)^3}{3m} \right]_0^{V_{DS}} = \frac{\mu_n C_{ox}^2 W^2}{I_{DS}} \left[\frac{(V_{GS} - V_T)^3}{3m} - \frac{(V_{GS} - V_T - mV_{DS})^3}{3m} \right] \end{aligned}$$

Under strong inversion, we get this final result:

$$Q_c = -WLC_{ox} \frac{2}{3} \frac{m^2 V_{DS}^2 + 3(V_{GS} - V_T)^2 - 3mV_{DS}(V_{GS} - V_T)}{2(V_{GS} - V_T - mV_{DS})} \quad (5.35)$$

In the ohmic regime that becomes

$$Q_c \approx -WLC_{ox}(V_{GS} - V_T) \quad (5.36)$$

while in saturation

$$Q_c \approx -WLC_{ox} \frac{2}{3}(V_{GS} - V_T) \quad (5.37)$$

due to pinch-off. Now we can find the gate capacitance

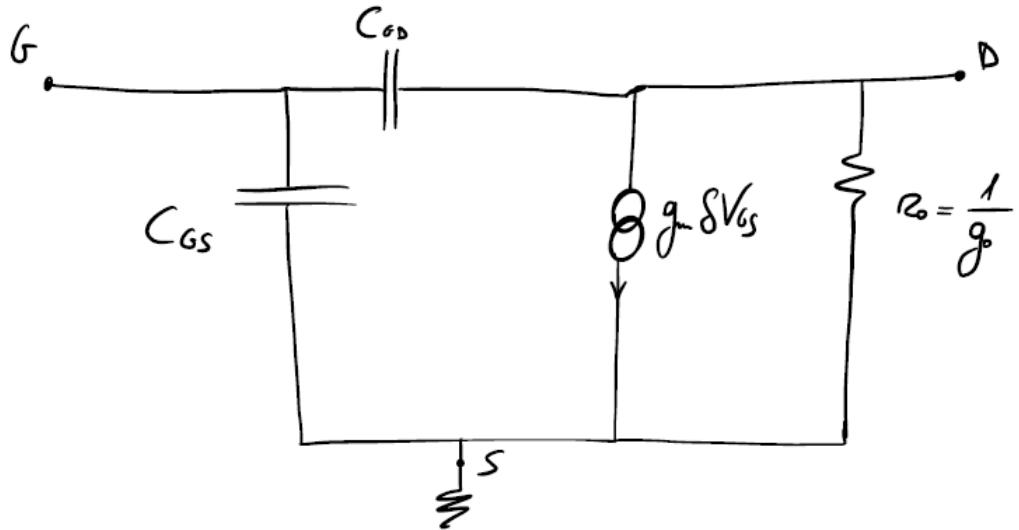
$$C_G = \left(-\frac{\partial Q_c}{\partial V_{GS}} \right)_{V_{DS}} = WLC_{ox} \left[1 - \frac{m^2 V_{DS}^2}{3[2(V_{GS} - V_T) - mV_{DS}]^2} \right] \quad (5.38)$$

and the drain capacitance

$$C_D = \left(-\frac{\partial Q_c}{\partial V_{DS}} \right)_{V_{GS}} = \frac{2}{3} WLC_{ox} \left[1 - \frac{(V_{GS} - V_T)^2}{[2(V_{GS} - V_T) - mV_{DS}]^2} \right] \quad (5.39)$$

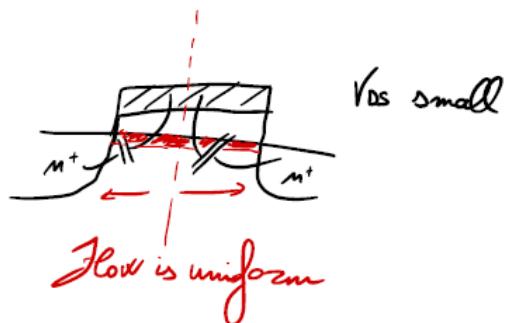
Furthermore

$$\begin{cases} C_G = C_{GS} + C_{GD} \\ C_D = C_{GD} \end{cases} \quad (5.40)$$



Once again, in the ohmic regime

$$\begin{cases} C_G = WLC_{ox} = C_{GS} + C_{GD} \\ C_D = \frac{1}{2}WLC_{ox} = C_{GD} \end{cases} \implies C_{GS} = C_{GD} = \frac{1}{2}WLC_{ox} \quad (5.41)$$



while in saturation

$$\begin{cases} C_G = \frac{2}{3}WLC_{ox} \\ C_D = 0 \end{cases} \implies C_{GD} = 0, C_{GS} = \frac{2}{3}WLC_{ox} \quad (5.42)$$



Drain does not play a role

Last parameter to study is the electron transit time

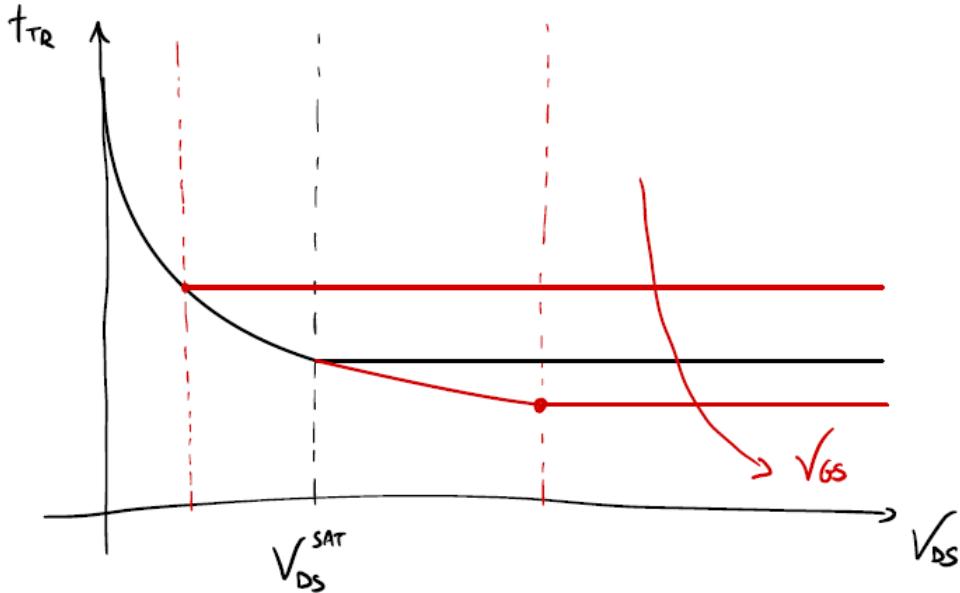
$$t_{TR} = \int_0^L \frac{dy}{v_d} = \frac{\partial Q_c}{\partial I_{DS}} = \frac{\partial Q_c}{\partial V_{GS}} \frac{\partial V_{GS}}{\partial I_{DS}} = \frac{C_G}{g_m} \quad (5.43)$$

We actually don't know what is the drift velocity, so we had to rewrite the equation as the inversion charge disappearing because of the current. As usual, first we see what is like in the ohmic regime

$$t_{\text{TR}} = \frac{WLC_{ox}}{\mu_n C_{ox} \frac{W}{L} V_{DS}} = \frac{L^2}{\mu_n V_{DS}} \quad (5.44)$$

and then in saturation

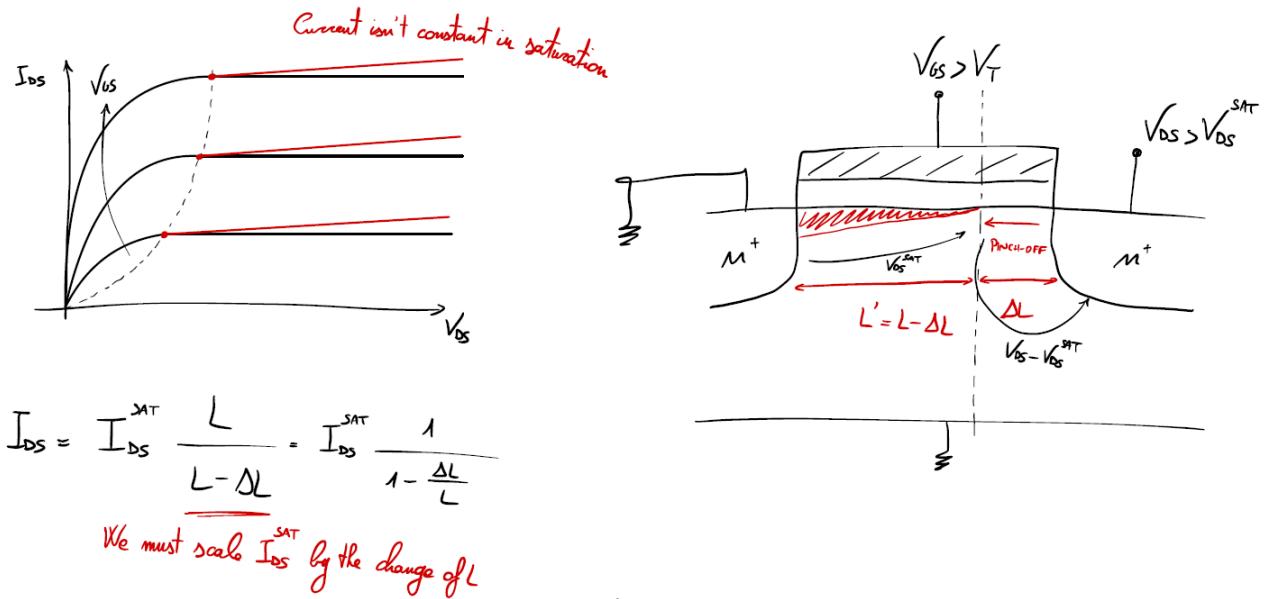
$$t_{\text{TR}} = \frac{2/3 WLC_{ox}}{\mu_n C_{ox} \frac{W}{L} \frac{V_{GS} - V_T}{m}} = \frac{2L^2}{3\mu_n V_{DS}^{\text{SAT}}} \quad (5.45)$$



There is a quadratic dependence on the distance. In order to decrease t_{TR} , finally, we must increase the mobility; this is a very important goal because travel time is strictly related to frequency response.

$$f_T = \frac{g_m}{2\pi(C_{GS} + C_{GD})} = \frac{1}{2\pi t_{\text{TR}}} \quad (5.46)$$

Parasitic elements strongly limit frequency response. Let's now spend a few extra words for the saturation regime.



$$I_{DS} = I_{DS}^{\text{SAT}} \frac{L}{L - \Delta L} - I_{DS}^{\text{SAT}} \frac{1}{1 - \frac{\Delta L}{L}}$$

We must scale I_{DS}^{SAT} by the change of L

Considering that $\Delta L = \frac{V_{DS} - V_{DS}^{\text{SAT}}}{F_p}$, where F_p is the electric field at the drain, and $\Delta L \ll L$, we can write

$$I_{DS} = I_{DS}^{\text{SAT}} \left(1 + \frac{\Delta L}{L} \right) = I_{DS}^{\text{SAT}} \left(1 + \frac{V_{DS} - V_{DS}^{\text{SAT}}}{F_p L} \right) \quad (5.47)$$

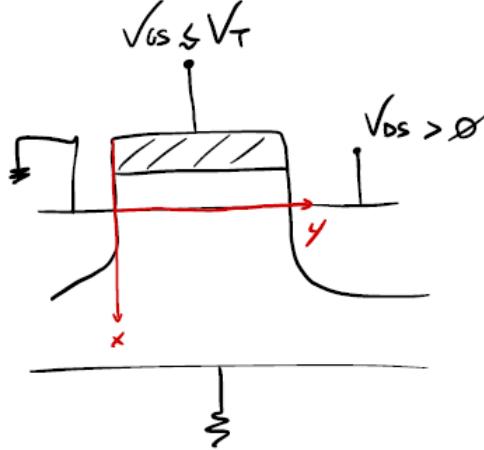
This translates into an actual output resistance

$$r_0 = \left(\frac{\partial I_{DS}}{\partial V_{DS}} \right)^{-1} = \frac{F_p L}{I_{DS}^{\text{SAT}}} = \frac{V_A}{I_{DS}^{\text{SAT}}} \quad (5.48)$$

where V_A is the **Early voltage**.

5.10 Subthreshold operations

In this section we go deeper into the subthreshold regime, so $V_{GS} < V_T$ and all the channel is in weak inversion (electrons are irrelevant). Also, $V_s \neq V_s(y)$ so we get the same band bending in every cross-section ($F_y = 0$).



From 4.23 we know

$$Q_{\text{INV}}(V) = -\sqrt{\frac{\epsilon_{\text{Si}} q N_A}{2V_s}} \frac{KT}{q} \frac{n_i^2}{N_A^2} \exp\left(\frac{q(V_s - V)}{KT}\right) \quad (5.49)$$

thus when we look for the current

$$\begin{aligned} I_{DS} &= -\mu_n \frac{W}{L} \int_0^{V_{DS}} Q_{\text{INV}}(V) dV = \mu_n \frac{W}{L} \sqrt{\frac{\epsilon_{\text{Si}} q N_A}{2V_s}} \frac{KT}{q} \frac{n_i^2}{N_A^2} \exp\left(\frac{qV_s}{KT}\right) \int_0^{V_{DS}} -\exp\left(\frac{qV}{KT}\right) dV = \\ &= \mu_n \frac{W}{L} \sqrt{\frac{\epsilon_{\text{Si}} q N_A}{2V_s}} \left(\frac{KT}{q}\right)^2 \frac{n_i^2}{N_A^2} \exp\left(\frac{qV_s}{KT}\right) \left[1 - \exp\left(-\frac{qV_{DS}}{KT}\right) \right] \end{aligned} \quad (5.50)$$

Nevertheless, we want the dependence to be on V_{GS} , not V_s , so

$$V_{GS} - V_{FB} = V_s - \frac{Q_s}{C_{ox}} \approx V_s - \frac{Q_{\text{DEP}}}{C_{ox}} = V_s + \frac{\sqrt{2\epsilon_{\text{Si}} q N_A} V_s}{C_{ox}}$$

For Taylor $\sqrt{V_s} \approx \sqrt{2|\phi_B|} + \frac{1}{2} \frac{1}{\sqrt{2|\phi_B|}} (V_s - 2|\phi_B|)$, then

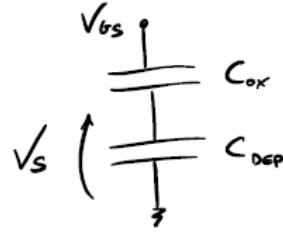
$$\begin{aligned} V_{GS} &= V_{FB} + V_s + \frac{\sqrt{2\epsilon_{\text{Si}} q N_A} 2|\phi_B|}{C_{ox}} + \frac{\sqrt{2\epsilon_{\text{Si}} q N_A}}{C_{ox}} \frac{1}{2} \frac{(V_s - 2|\phi_B|)}{\sqrt{2|\phi_B|}} = \\ &= V_{FB} + 2|\phi_B| + \frac{\sqrt{2\epsilon_{\text{Si}} q N_A} 2|\phi_B|}{C_{ox}} + (V_s - 2|\phi_B|) \left(1 + \frac{C_{\text{DEP}}}{C_{ox}} \right) = V_T + (V_s - 2|\phi_B|)m \end{aligned} \quad (5.51)$$

and finally

$$V_s = 2|\phi_B| + \frac{V_{GS} - V_T}{m} \quad (5.52)$$

Since $V_{GS} \approx V_T$, $V_s \approx 2|\phi_B|$. Alternatively, we could have avoided all calculations just by considering the small signal model:

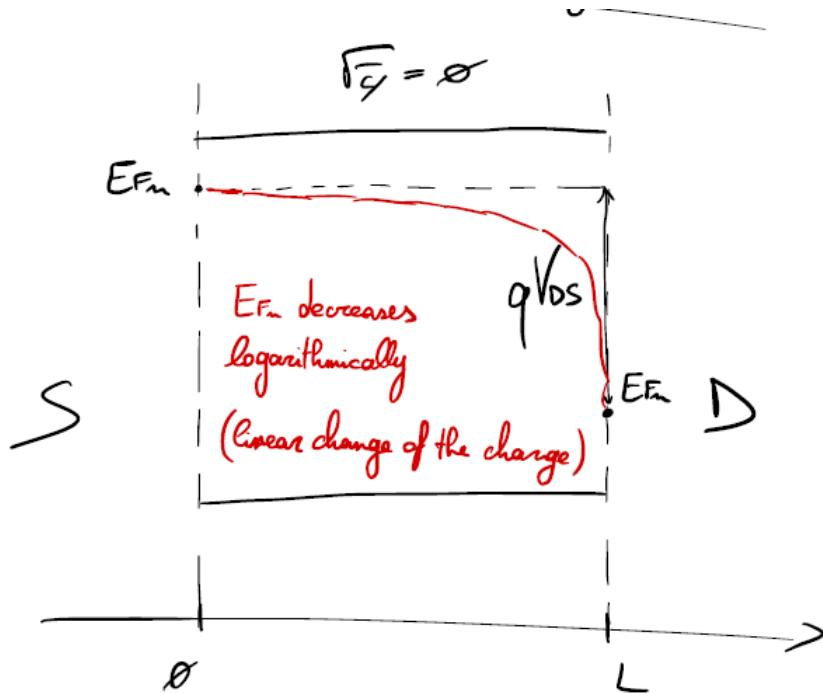
$$V_{GS} + \delta V_{GS} = V_s = 2|\phi_B| + \delta V_{GS} \frac{C_{ox}}{C_{ox} + C_{\text{DEP}}} = 2|\phi_B| + \frac{\delta V_{GS}}{1 + \frac{C_{\text{DEP}}}{C_{ox}}} = 2|\phi_B| + \frac{V_{GS} - V_T}{m} \quad (5.53)$$



This works because physically $1/m$ is the voltage partition over the substrate (if $m = 1$, all the gate voltage drops over the depletion layer; gate has then maximum control over the substrate electrostatics). Now let's get back to 5.50 and substitute 5.52 inside it; final result looks like

$$I_{DS} = \mu_n C_{ox} \frac{W}{L} (m-1) \left(\frac{KT}{q} \right)^2 \exp \left(\frac{q(V_{GS} - V_T)}{mKT} \right) \left[1 - \exp \left(-\frac{qV_{DS}}{KT} \right) \right] \quad (5.54)$$

This is the **subthreshold current**; what is important is the double exponential dependence (even though the second one can be discarded if $V_{DS} \gg KT/q$). Now let's take a look at the electrostatics in the channel region:



Even if E_{F_n} drops it doesn't matter: we have a **pure diffusion current**.

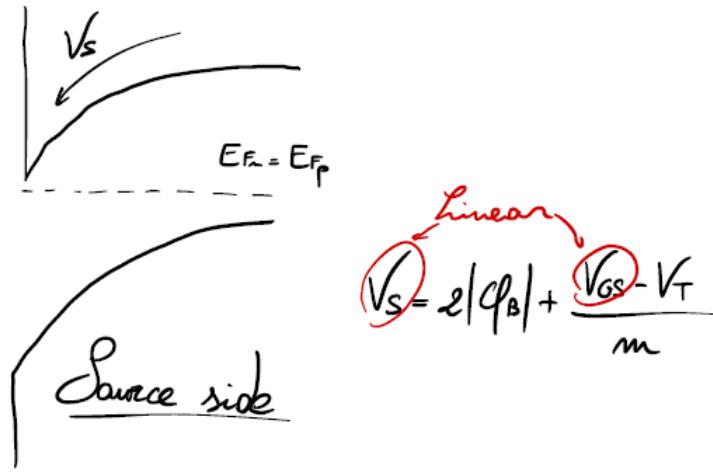
$$I_{DS} = W \int_0^{t_{INV}} J_n(x) dx = -W \int_0^{t_{INV}} q D_n \frac{dn}{dy} dx = WD_n \frac{d}{dy} \int_0^{t_{INV}} -q n dx = WD_n \frac{dQ_{INV}}{dy} \quad (5.55)$$

The derivative must be constant, so Q_{INV} is a linear function of the position.

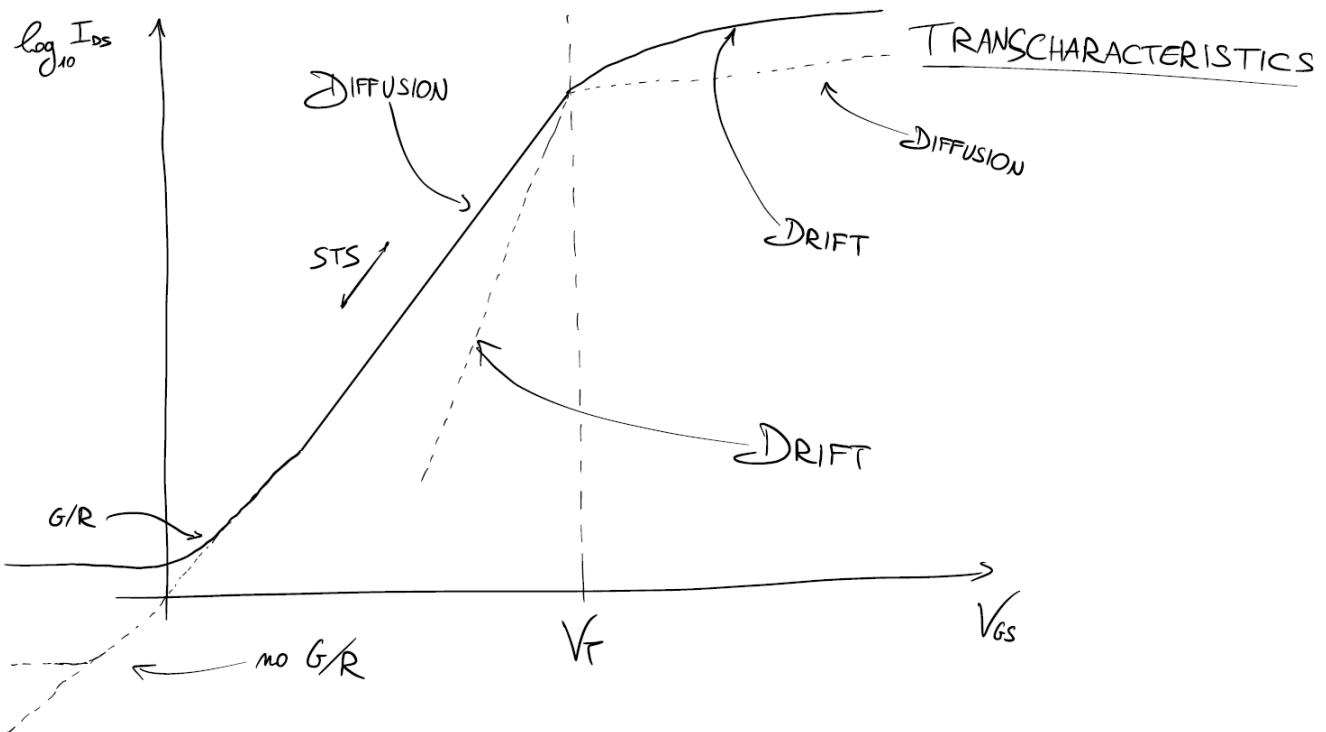
$$I_{DS} = -WD_n \frac{Q_{INV}(y=0) - Q_{INV}(y=L)}{L} \quad (5.56)$$

Taking a closer look to the charge in the two positions:

$$\begin{cases} Q_{INV}(0) = -\sqrt{\frac{\epsilon_{Si} q N_A}{2V_s}} \frac{KT}{q} \frac{n_i^2}{N_A^2} \exp\left(\frac{qV_s}{KT}\right) \\ Q_{INV}(L) = -\sqrt{\frac{\epsilon_{Si} q N_A}{2V_s}} \frac{KT}{q} \frac{n_i^2}{N_A^2} \exp\left(\frac{q(V_s - V_{DS})}{KT}\right) \end{cases} \implies Q_{INV}(0) - Q_{INV}(L) \propto e^{\frac{qV_s}{KT}} \left(1 - e^{\frac{qV_{DS}}{KT}}\right) \quad (5.57)$$



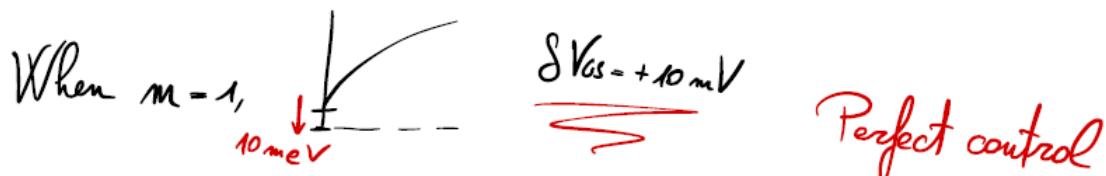
Current increases exponentially with V_{GS} !



We can finally define the **SubThreshold Slope**:

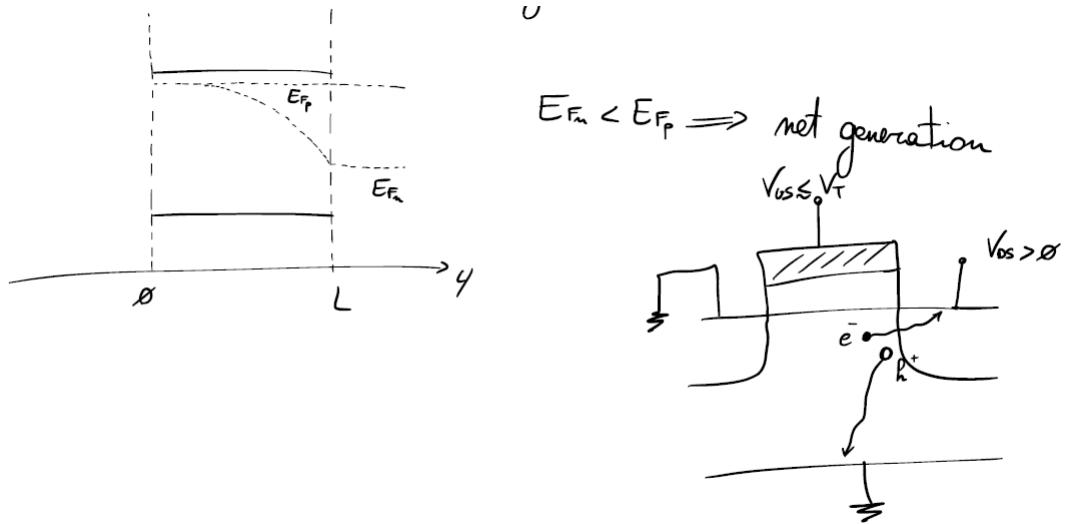
$$\text{STS} = \left[\frac{\partial \log_{10} I_{DS}}{\partial V_{GS}} \right]^{-1} = \frac{KT}{q} \log(10)m \quad (5.58)$$

To have a good device, this must be as steep as possible².



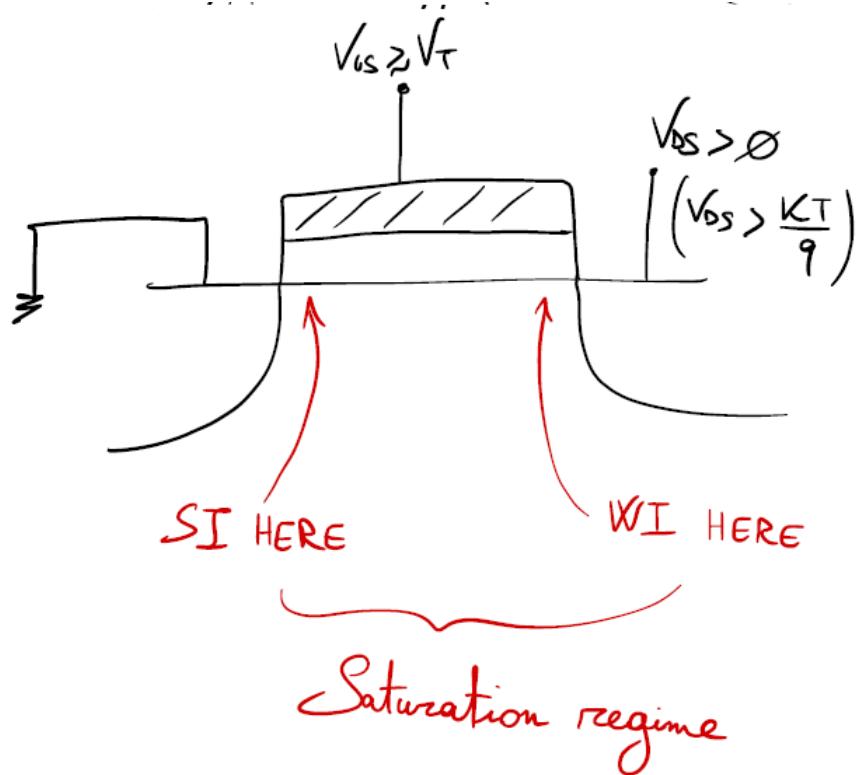
As always, if we enter in low-current regime, G/R processes must be considered.

²At 300K, $\frac{KT}{q} \log(10) = 60 \text{ mV/dec}$

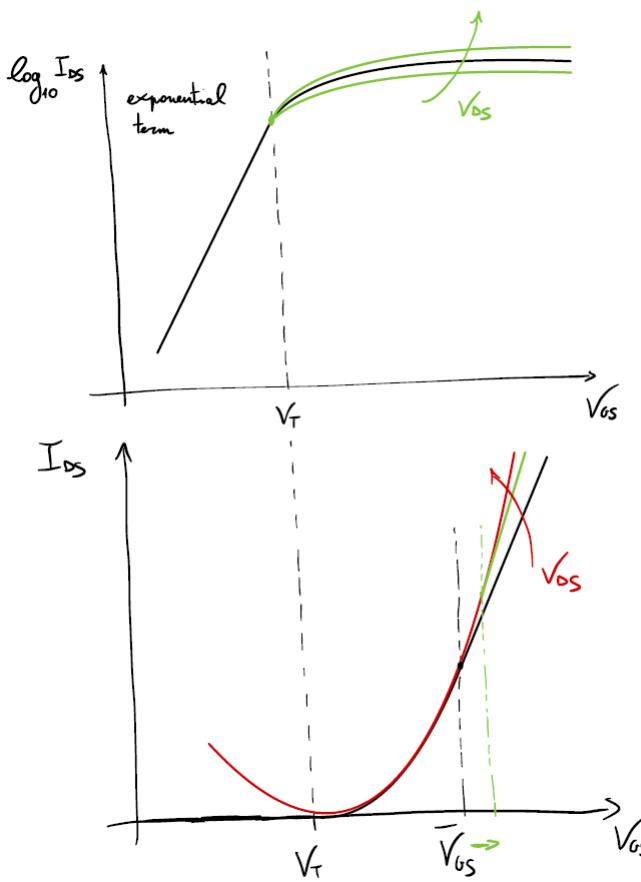


5.11 Impact of different parameters on the transcharacteristics

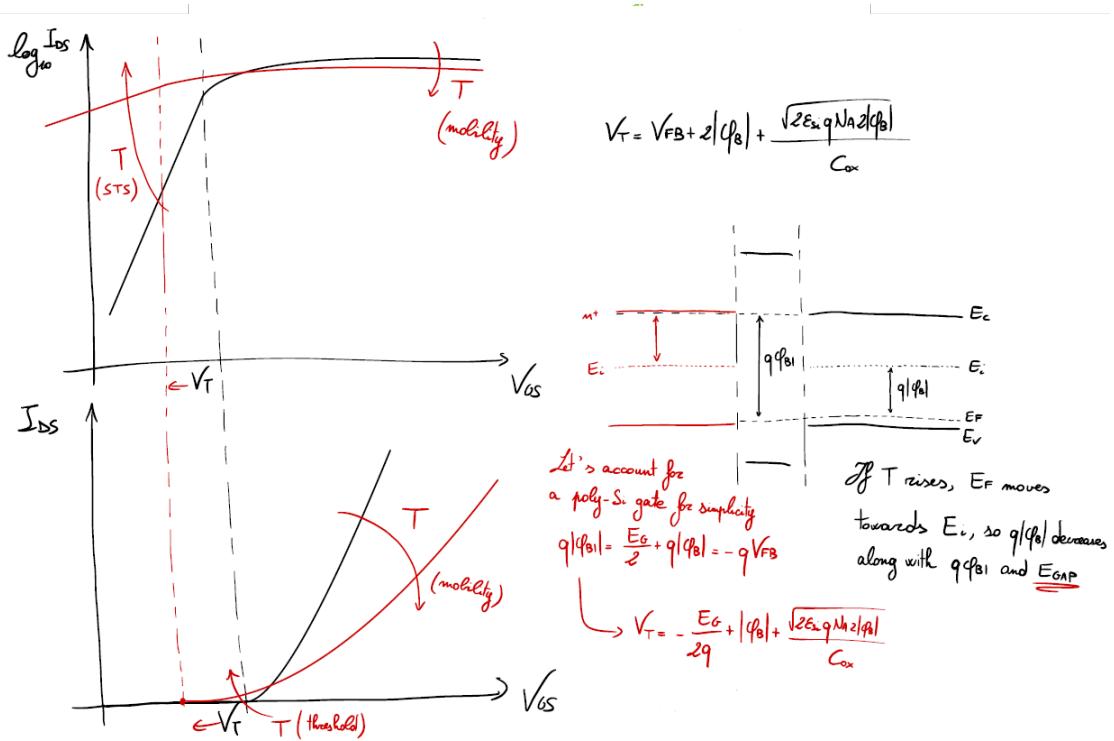
We now see in detail how drain voltage, temperature, residual oxide charge and interface states affect our model. Let's start with V_{DS} :



Here $V'_T = V_T + mV_{DS} = \bar{V}_{GS}$; when $V_{GS} = \bar{V}_{GS}$ we enter ohmic regime.



Subthreshold current on the other hand is independent from V_{DS} (when $V_{DS} > KT/q$), but it is greatly affected by the temperature.



If we were to quantify this impact:

$$\frac{dV_T}{dT} = -\frac{1}{2q} \frac{dE_G}{dT} + \frac{d|\phi_B|}{dT} + \frac{\sqrt{2\epsilon_{si}qN_A}}{C_{ox}} \frac{1}{2|\phi_B|} \frac{1}{dT} \frac{d|\phi_B|}{dT} = -\frac{1}{2q} \frac{dE_G}{dT} + \frac{d|\phi_B|}{dT} \left(1 + \frac{2C_{DEP}}{C_{ox}} \right) = -\frac{1}{2q} \frac{dE_G}{dT} + \frac{d|\phi_B|}{dT} (2m - 1)$$

From 4.4 we can say

$$\frac{d|\phi_B|}{dT} = \frac{K}{q} \log\left(\frac{N_A}{n_i}\right) + \frac{KT}{q} \frac{\mu_i}{N_A} \frac{-\mathcal{N}_A}{n_i^2} \frac{dn_i}{dT} = \frac{K}{q} \log\left(\frac{N_A}{n_i}\right) - \frac{KT}{q} \frac{1}{n_i} \frac{dn_i}{dT} \quad (5.59)$$

then

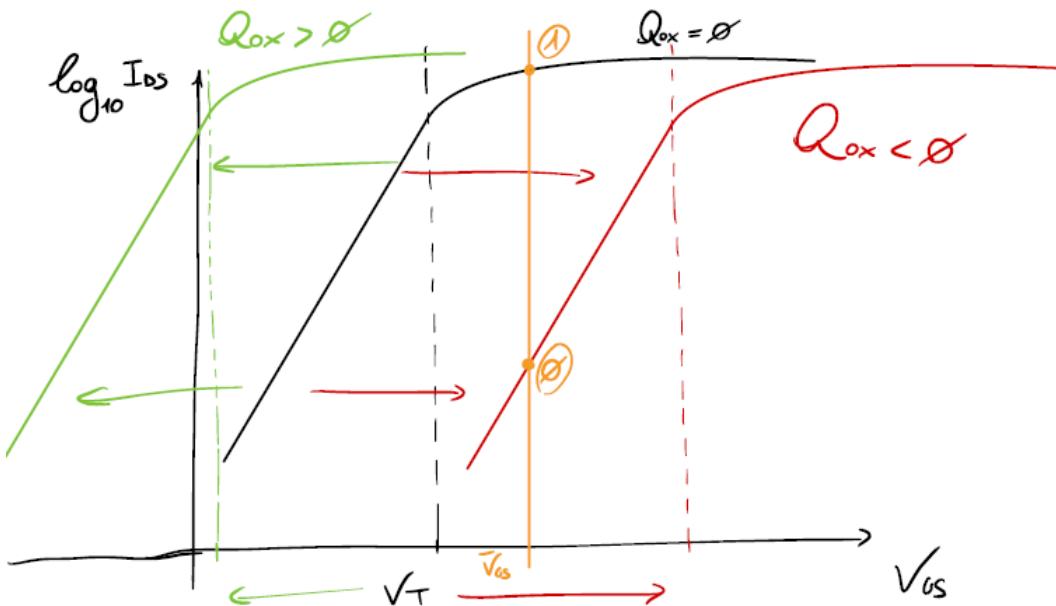
$$\frac{dV_T}{dT} = -(2m-1) \frac{K}{q} \left[\log\left(\frac{\sqrt{N_c N_v}}{N_A}\right) + \frac{3}{2} \right] + \frac{m-1}{q} \frac{dE_G}{dT} < 0 \quad (5.60)$$

Threshold voltage decreases as temperature rises. Also mobility lowers with temperature. Here is an example:

$$\begin{cases} N_A = 10^{16} \text{ cm}^{-3} \\ m = 1.1 \end{cases} \implies \frac{dV_T}{dT} = -1 \text{ mV/K}$$

$$\begin{cases} N_A = 10^{18} \text{ cm}^{-3} \\ m = 1.3 \end{cases} \implies \frac{dV_T}{dT} = -0.7 \text{ mV/K}$$

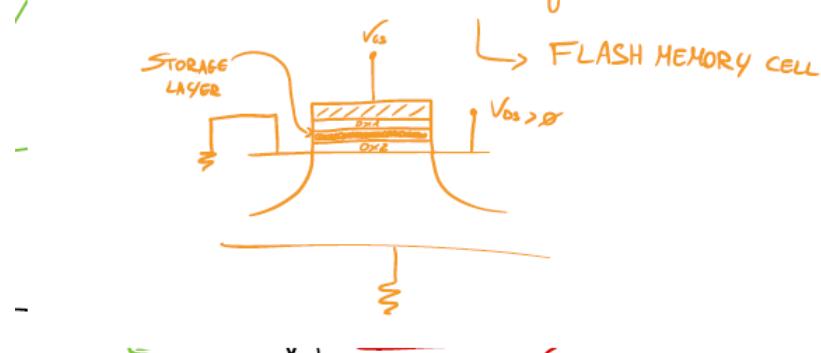
Talking about spurious charges in the oxide, all they do is provoking a rigid shift of the transcharacteristics according to their sign.



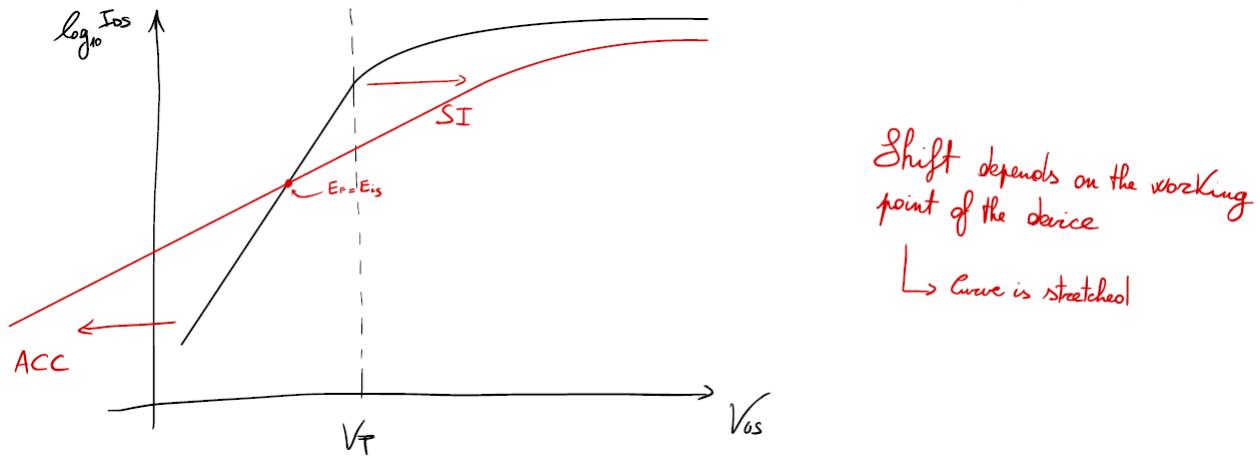
This can actually be engineerable in order to create **flash memory cells**.

If we were able to calibrate Q_{ox} we could shift the transcharacteristics how we like it

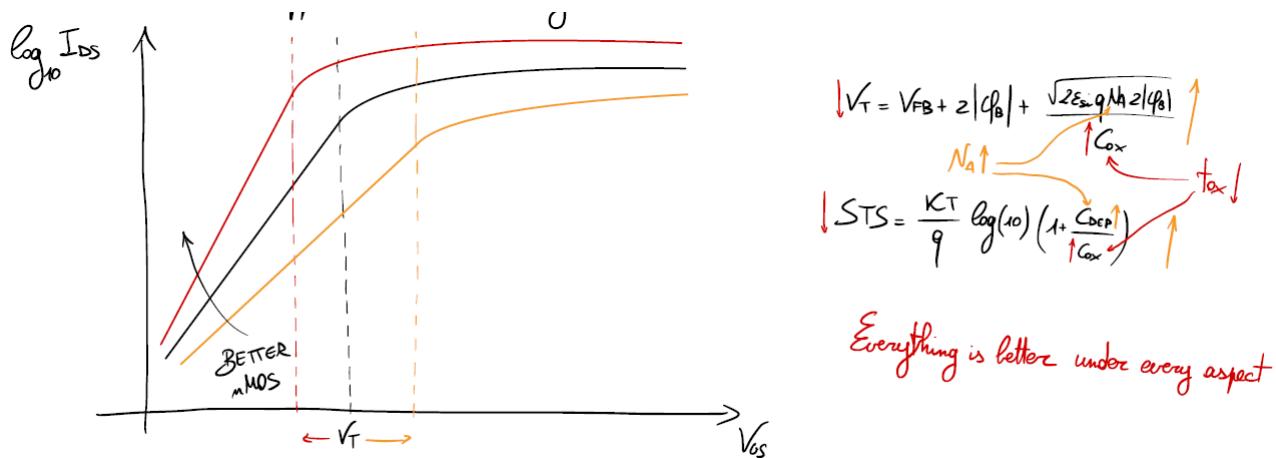
For V_{DS} we would have two possible current flows



Very briefly we now cover the interface states:



Last but not least, we see what happens when we change t_{ox} and N_A :

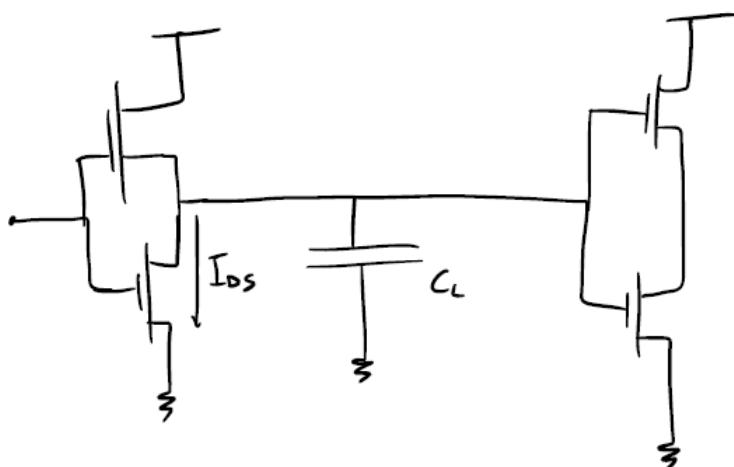


5.12 Short-channel regime

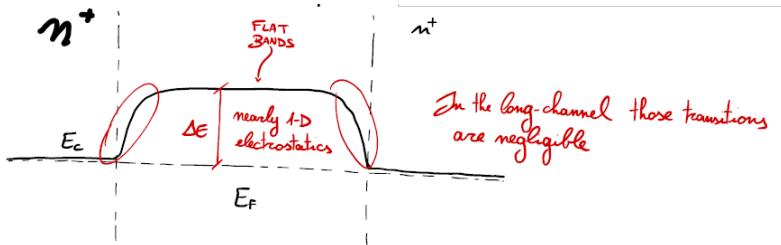
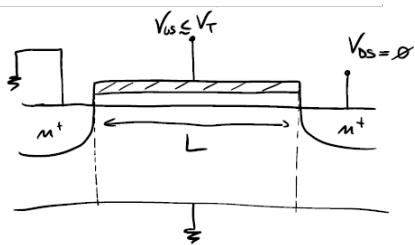
Why would we want to reduce channel length?

$$\begin{cases} t_{tr} \propto L^2 & \text{transit time improves quadratically} \\ I_{DS} \propto \frac{W}{L} & \text{drain current increases} \\ C_G \propto WL & \text{intrinsic capacitance gets smaller} \end{cases} \quad (5.61)$$

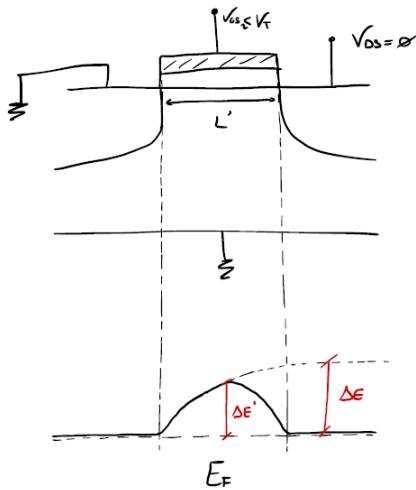
All of this allows us to build way faster logic circuits with higher currents discharging smaller capacitances.



Enhanced performances and cheaper devices are other pros. Nevertheless, if we reduce just L leaving untouched every other parameter we note an actual worsening of the performances; that's because we have entered **short-channel regime**. First, take a step back and consider once again the bidimensional electrostatics in the channel, starting from a long-channel:



Now here is a short-channel:



Here is the problem

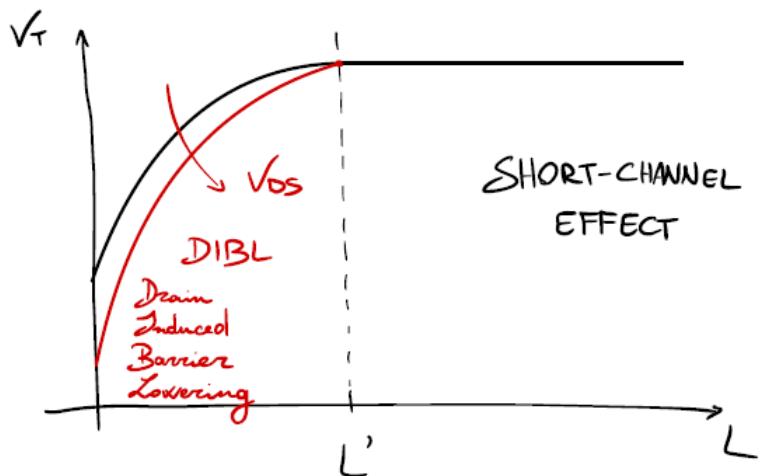
Electrostatics is strongly affected by source and drain

↳ it has become bidimensional

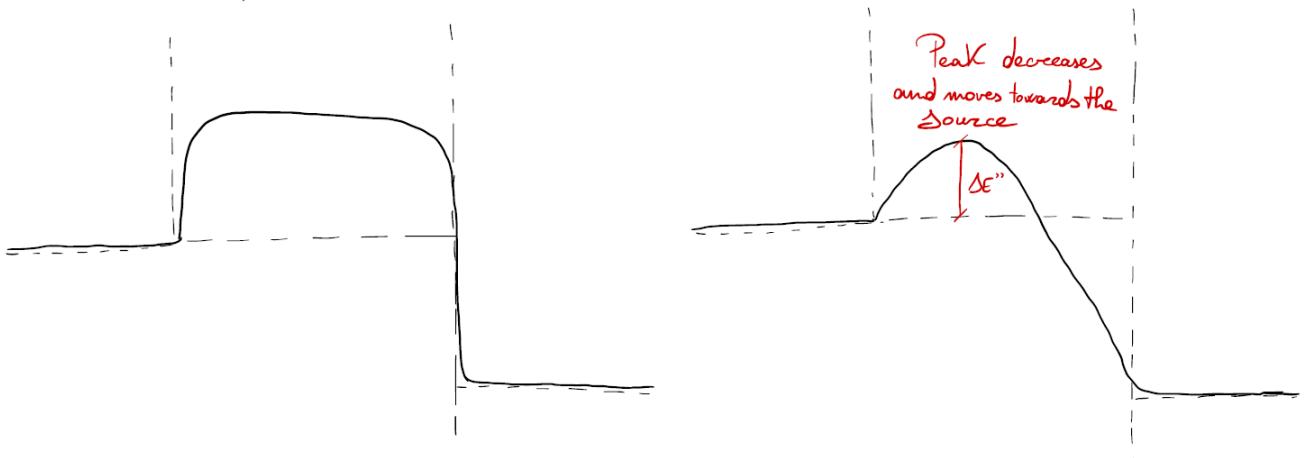
In the long-channel, only in saturation we had bidimensional electrostatics

↳ more gradual approximation is always invalid

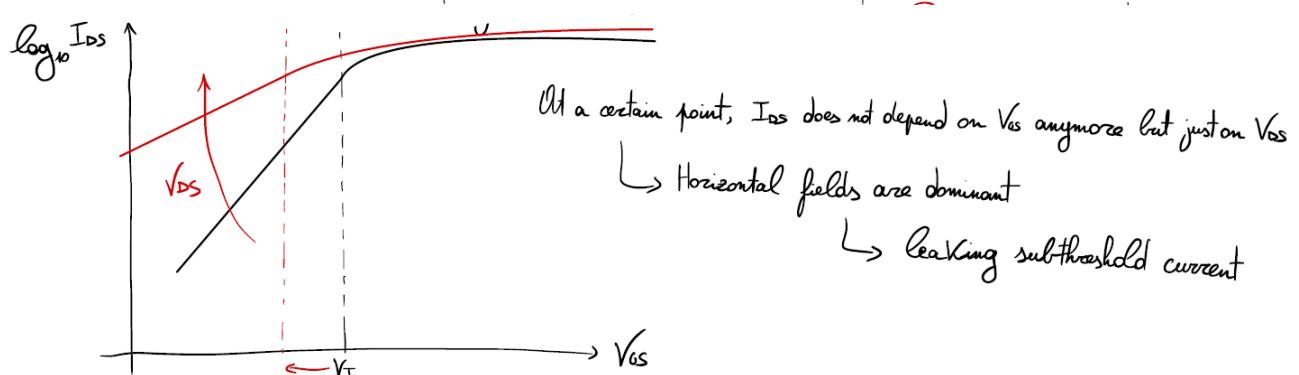
We are forced to solve the 2D Poisson equation; additionally, for the same V_{GS} we have a higher electron concentration than in the long-channel, so V_{GS} must be reduced ($V_T^{\text{SH}} < V_T^{\text{LO}}$).



What if we apply $V_{DS} > 0$?



Output resistance is reduced by DIBL as well; regarding the subthreshold regime, instead



V_T reduction is described as

$$\Delta V_T = \frac{24t_{ox}}{W_{D_{max}}} \left[\sqrt{\phi_{BI}(\phi_{BI} + V_{DS})} - 0.4 \cdot 2|\phi_B| \right] \exp\left(-\frac{\pi L}{2(W_{D_{max}} + 3t_{ox})}\right) \quad (5.62)$$

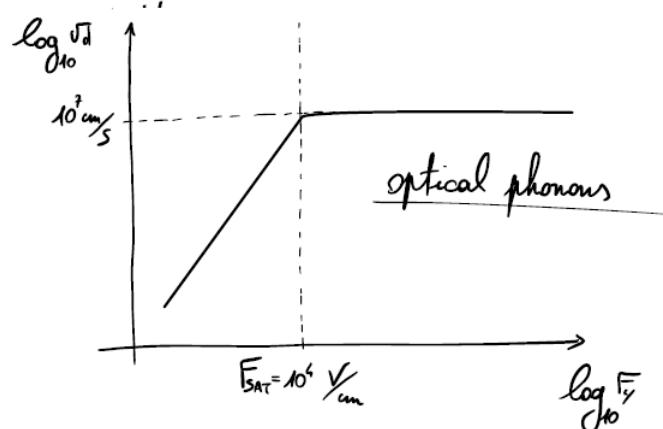
In order to have a good length for the channel we must follow

$$\Delta V_T < 100 \text{ mV} : L > 2(W_{D_{max}} + 3t_{ox}) \quad (5.63)$$

If the gate is closer to the substrate, it has a better control over the electrostatics. With high doping concentration, also, we have steeper transitions at the n^+ -regions, but the gate has a worse control.

5.13 Velocity saturation

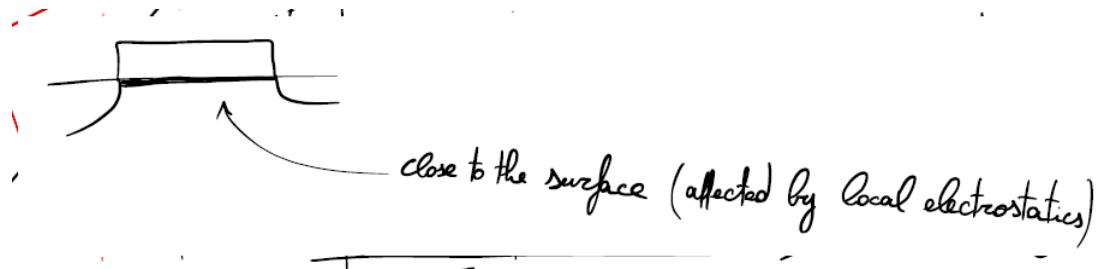
Velocity is proportional to the electric field, but from a certain F onwards it saturates because of the optical phonons: this is the **velocity saturation**.



$$v_d = \frac{\mu_{\text{eff}} F_y}{\left[1 + \left(\frac{F_y}{F_{\text{SAT}}}\right)^n\right]^{1/n}} = \begin{cases} \mu_{\text{eff}} F_y & \text{if } F_y \ll F_{\text{SAT}} \\ \mu_{\text{eff}} F_{\text{SAT}} & \text{if } F_y \gg F_{\text{SAT}} \end{cases} \quad (5.64)$$

where μ_{eff} is the mobility in the channel, not in the bulk, and n is a parameter dependent on the technology, in particular

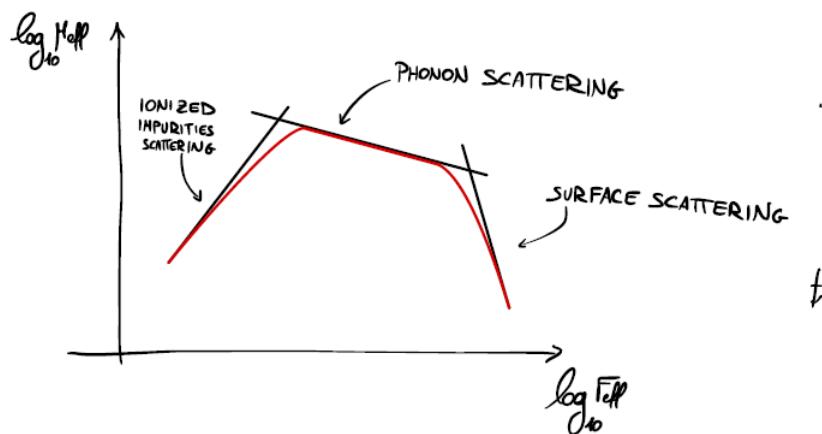
$$\begin{cases} n = 1 & \text{for pMOS} \\ n = 2 & \text{for nMOS} \end{cases} \quad (5.65)$$



We can also define a vertical electric field (carrier/surface interaction) like

$$F_{\text{eff}} = \frac{Q_{\text{DEP}} + \frac{1}{2} Q_{\text{INV}}}{\epsilon_{\text{Si}}} \quad (5.66)$$

which strongly affects the mobility.



We are ready to compute the current.

$$\begin{aligned}
I_{DS} &= -\mu_n W Q_{INV} \frac{dV}{dy} = -\mu_n W Q_{INV} (-F_y) = -v_d W Q_{INV} = -\frac{\mu_{\text{eff}} \frac{dV}{dy}}{1 + \frac{dV}{dy} \frac{1}{F_{\text{SAT}}}} W Q_{INV} \\
I_{DS} \left(1 + \frac{dV}{dy} \frac{1}{F_{\text{SAT}}} \right) &= -\mu_{\text{eff}} \frac{dV}{dy} W Q_{INV} \\
I_{DS} &= \left[-\frac{I_{DS}}{F_{\text{SAT}}} - \mu_{\text{eff}} W Q_{INV} \right] \frac{dV}{dy} \\
\int_0^L I_{DS} dy &= \int_0^{V_{DS}} \left[-\frac{I_{DS}}{F_{\text{SAT}}} - \mu_{\text{eff}} W Q_{INV} \right] dy \tag{5.67} \\
I_{DS} L &= -\frac{I_{DS}}{F_{\text{SAT}}} V_{DS} - \int_0^{V_{DS}} \mu_{\text{eff}} W Q_{INV} dV \\
I_{DS} &= \frac{-\mu_{\text{eff}} \frac{W}{L} \int_0^{V_{DS}} Q_{INV} dV}{1 + \frac{V_{DS}}{L} \frac{1}{F_{\text{SAT}}}}
\end{aligned}$$

Numerator is the same for the long-channel MOSFET; note that when L is very small the denominator becomes way more relevant. Assuming quasi-1D electrostatics, we can say

$$I_{DS} = \frac{-\mu_{\text{eff}} \frac{W}{L} \int_0^{V_{DS}} -C_{ox}(V_{GS} - V_T - mV) dV}{1 + \frac{V_{DS}}{L} \frac{1}{F_{\text{SAT}}}} = \frac{\mu_{\text{eff}} C_{ox} \frac{W}{L} \left[(V_{GS} - V_T)V_{DS} - \frac{mV_{DS}^2}{2} \right]}{1 + \frac{V_{DS}}{L} \frac{1}{F_{\text{SAT}}}} \tag{5.68}$$

$$\frac{\partial I_{DS}}{\partial V_{DS}} = 0 \implies V_{DS}^{\text{SAT}} = \frac{\frac{2}{m}(V_{GS} - V_T)}{1 + \sqrt{1 + 2\frac{V_{GS} - V_T}{m} \frac{1}{L} \frac{1}{F_{\text{SAT}}}}} \tag{5.69}$$

We can distinguish long-channel and short-channel cases:

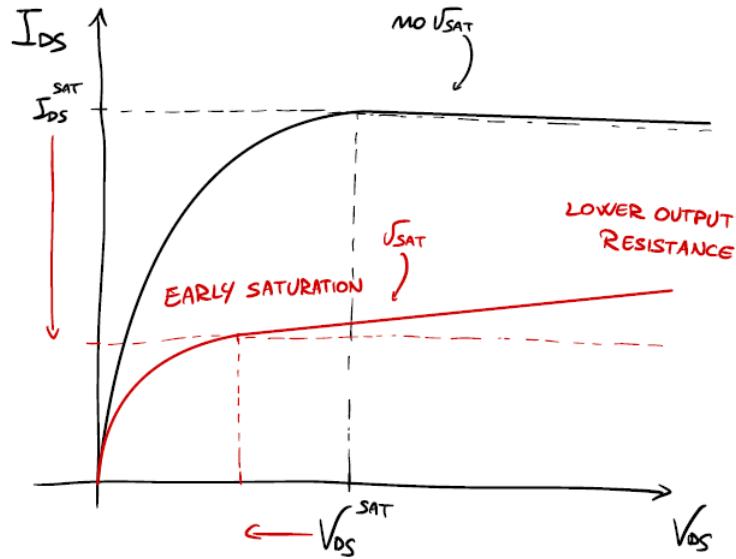
- LONG-CHANNEL:

$$\begin{cases} V_{DS}^{\text{SAT}} = \frac{V_{GS} - V_T}{m} \\ I_{DS}^{\text{SAT}} = \mu_{\text{eff}} C_{ox} \frac{W}{L} \frac{(V_{GS} - V_T)^2}{2m} \end{cases} \implies Q_{INV}(y = L) = 0 \tag{5.70}$$

- SHORT-CHANNEL:

$$\begin{cases} V_{DS}^{\text{SAT}} < \frac{V_{GS} - V_T}{m} \\ I_{DS}^{\text{SAT}} < \mu_{\text{eff}} C_{ox} \frac{W}{L} \frac{(V_{GS} - V_T)^2}{2m} \end{cases} \implies Q_{INV}(y = L) \neq 0 \tag{5.71}$$

Current saturates early because of v_{SAT} , so we still have strong-inversion condition at the drain.



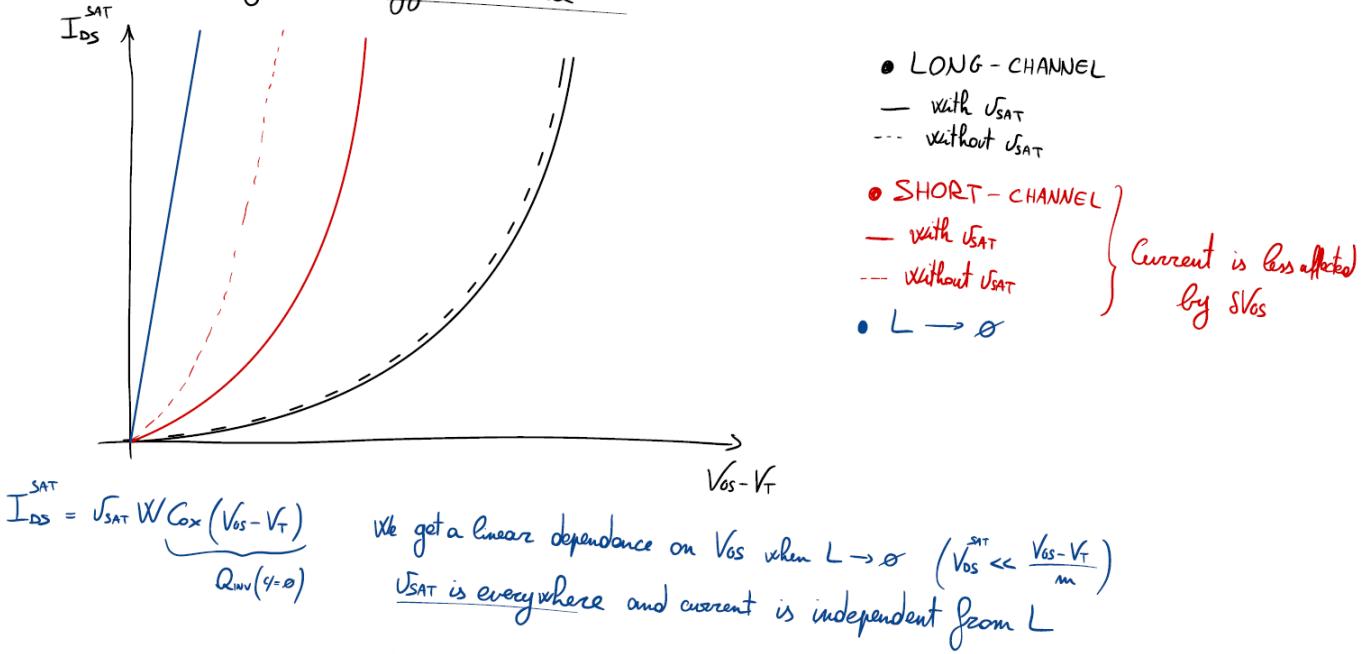
In other terms, current becomes constant because $v_d = v_{\text{SAT}}$ at the drain.

$$I_{DS} = -v_d(y=L)WQ_{\text{INV}}(y=L) \xrightarrow{I_{DS}=I_{DS}^{\text{SAT}}} I_{DS}^{\text{SAT}} = -v_{\text{SAT}}WQ_{\text{INV}}(y=L)$$

$$Q_{\text{INV}} = -C_{ox}(V_{GS} - V_T - mV_{DS}^{\text{SAT}})$$
(5.72)

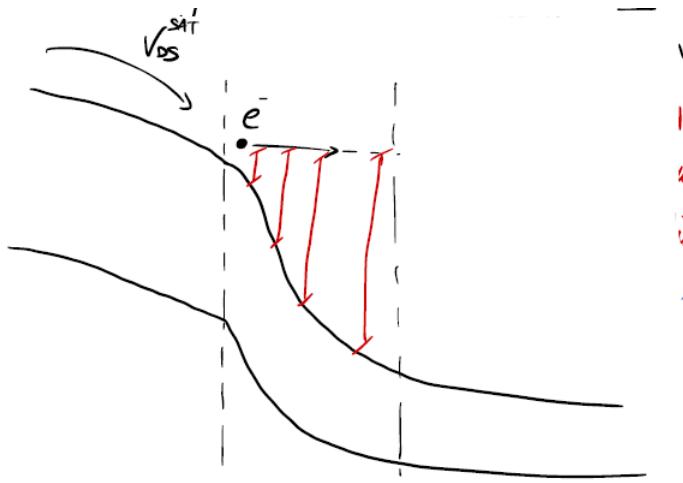
Pinch-off condition and velocity saturation are mutually exclusive: one or the other is reached according to channel length.

Device in saturation regime is strongly bidimensional



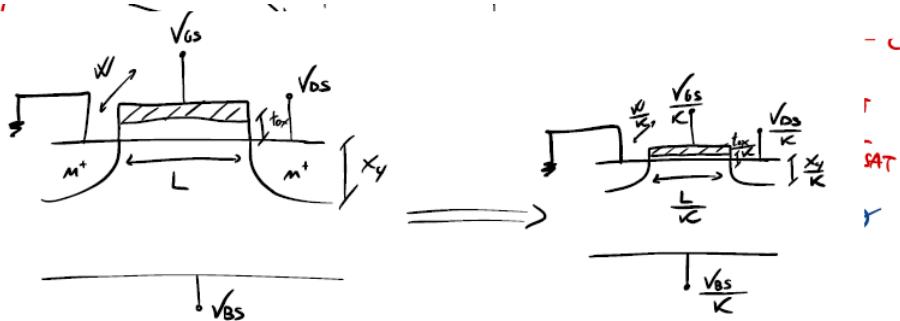
In order to limit F_y , we should intervene on V_{DS} .

Hot electrons As we have seen, v_{SAT} is the maximum drift velocity (scattering), not the maximum in absolute. Electrons, in fact, can reach even $v \gg v_{\text{SAT}}$ thanks to the very high kinetic energies made possible by the band bending; these are called *hot electrons* and constitute a problem for reliability (spurious state creation, oxide charging, etc...).



5.14 Constant-field scaling

Short-channel regime has detrimental effects, unless we correctly rescale every parameter. Here are presented the **constant-field scaling rules** by Dennard: vertical and horizontal electric field ratio should not change in rescaling, while avoiding velocity saturation (do not increase electric fields at all).



All the dimensions and all the voltages have been scaled by $1/k$.

$$-\frac{\partial F_x}{\partial x} - \frac{\partial F_y}{\partial y} = \frac{q}{\epsilon_{si}} N_A \quad \Rightarrow \quad -\frac{\partial F_x}{\partial x/k} - \frac{\partial F_y}{\partial y/k} = \frac{q}{\epsilon_{si}} N'_A \quad (5.73)$$

where $N'_A = kN_A$; doping has increased by k . Let's see how parameters have changed (**RED**: reduced by k ; **BLUE**: increased by k):

- Electric fields

$$F = \frac{V}{L} \rightarrow 1 \quad (5.74)$$

- Carrier velocity

$$v_d \rightarrow 1 \quad (5.75)$$

- Capacitances

$$C_G = C_{ox} W L \rightarrow \frac{1}{k} \quad (5.76)$$

- Current in the onstate regime

$$I_{DS} = \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_T) V_{DS} \rightarrow \frac{1}{k} \quad ?? \quad (5.77)$$

$$V_T \approx \frac{\sqrt{2\epsilon_{si} q N_A 2|\phi_B|}}{C_{ox}} \rightarrow \frac{1}{\sqrt{k}}$$

Everything would be fine if we were able to scale $2|\phi_B|$, but since it is related to the energy gap, we can't; also, position of dopants is relevant, thus we choose a non-uniform doping concentration in the substrate.

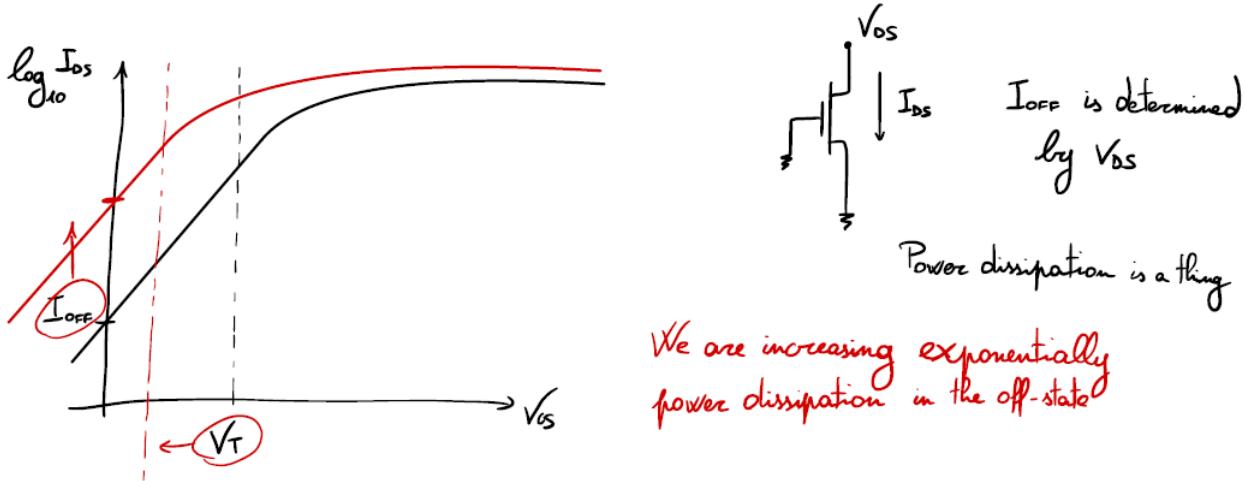
- Transit-time

$$t_{tr} = \frac{L^2}{\mu_n V_{DS}} \rightarrow \frac{1}{k} \quad (5.78)$$

- SubThreshold Slope

$$\text{STS} = \frac{KT}{q} \log(10) \left(1 + \frac{C_{\text{DEP}}}{C_{\text{ox}}} \right) \rightarrow 1 \quad (5.79)$$

This can actually be a problem because we are decreasing V_T without scaling the STS.



- Minimum length condition

$$L = 2(W_{D_{\max}} + 3t_{\text{ox}}) \quad (5.80)$$

$$W_{D_{\max}} = \sqrt{\frac{2\epsilon_{\text{Si}}}{q} \frac{1}{N_A} 2|\phi_B|} \rightarrow \sqrt{\frac{1}{k}}$$

L approaches the second member while the first is smaller just by $1/k$: that's no good at all!

Now let's consider circuit performances:

- Delay of logic gates

$$\tau = \frac{CV}{I_{DS}} \rightarrow \frac{1}{k} \quad (5.81)$$

We get faster circuits.

- Power dissipation

$$P = VI_{DS} \rightarrow \frac{1}{k^2} \quad (5.82)$$

- Density of power dissipation

$$p = \frac{VI}{WL} \rightarrow 1 \quad (5.83)$$

- Integration density (components per unit area)

$$I = \frac{1}{WL} \rightarrow k^2 \quad (5.84)$$

At a first approach, everything looks fine, but we still have some issues for bidimensional electrostatics and offstate power dissipation.

5.15 Generalized scaling

There is always a balance between pros and cons; Dennard's rules are good but present some flaws, so here are introduced the **generalized scaling rules** by Baccaram: we want to avoid moving towards bidimensionality at all costs, so we pay with a little increase of the electric fields. All dimensions are scaled by $1/k$, while all the voltages by α/k ($1 \leq \alpha \leq k$); for $\alpha = 1$ we get back to the constant-field scaling, while for $\alpha = k$ we have constant voltages. Doping concentrations are also increased by αk ; as before let's jump into details.

- Electric fields

$$F = \frac{V}{L} \rightarrow \alpha \quad (5.85)$$

- Carrier velocity

$$v_d \rightarrow \begin{cases} \alpha & \text{long-channel} \\ 1 & \text{velocity saturation} \end{cases} \quad (5.86)$$

- Capacitances

$$C_G = \mathcal{C}_{ox} WL \rightarrow \frac{1}{k} \quad (5.87)$$

- Current in the onstate regime

$$I_{DS} = \mu_n \mathcal{C}_{ox} \frac{W}{L} (V_{GS} - V_T) V_{DS} \rightarrow \begin{cases} \frac{\alpha^2}{k} \\ \frac{\alpha}{k} \end{cases} \quad (5.88)$$

- Transit-time

$$t_{\text{tr}} = \frac{L^2}{\mu_n V_{DS}} \rightarrow \begin{cases} \frac{1}{\alpha k} \\ \frac{1}{k} \end{cases} \quad (5.89)$$

- Delay of logic gates

$$\tau = \frac{CV}{I_{DS}} \rightarrow \begin{cases} \frac{1}{\alpha k} \\ \frac{1}{k} \end{cases} \quad (5.90)$$

- Power dissipation

$$P = VI_{DS} \rightarrow \begin{cases} \frac{\alpha^3}{k^2} \\ \frac{\alpha^2}{k^2} \end{cases} \quad (5.91)$$

- Density of power dissipation

$$p = \frac{VI}{WL} \rightarrow \begin{cases} \alpha^3 \\ \alpha^2 \end{cases} \quad (5.92)$$

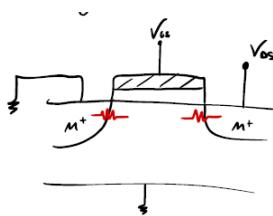
- Integration density (components per unit area)

$$I = \frac{1}{WL} \rightarrow k^2 \quad (5.93)$$

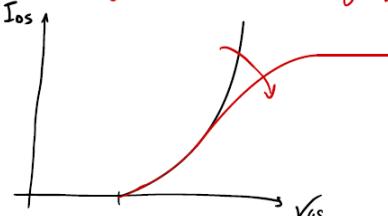
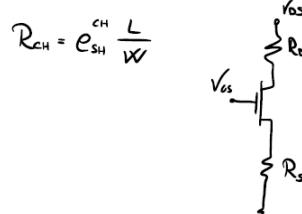
We have many problems concerning power dissipation, even though we fixed the minimum length condition; in fact, for $\alpha = k$

$$W_{D_{\max}} \rightarrow \frac{1}{k} \implies L > 2(W_{D_{\max}} + 3t_{ox}) \quad (5.94)$$

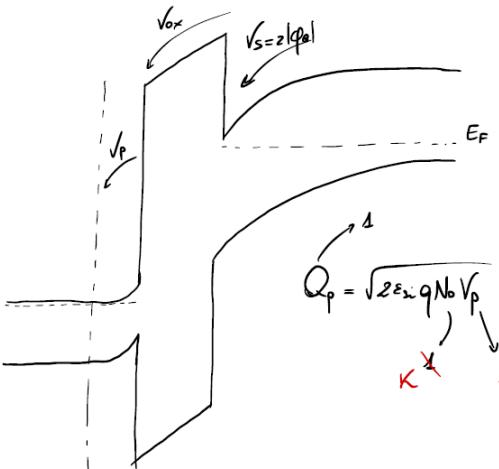
Both sets of rules are valid and offer some good trade-offs. Doping concentration, then, must be increased anyway because of the parasitic resistances.



We have parasitic resistances between channel and n+ region, which do not change if N+ stays the same



Same thing for polysilicon gates:



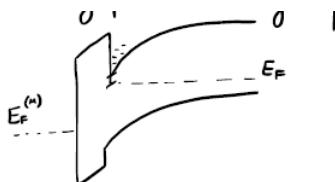
V_p must be reduced, so N_d must be increased as well

Doping concentration arrived to a point where couldn't be further incremented

↳ Back to metal gates

5.16 Dielectric imperfections

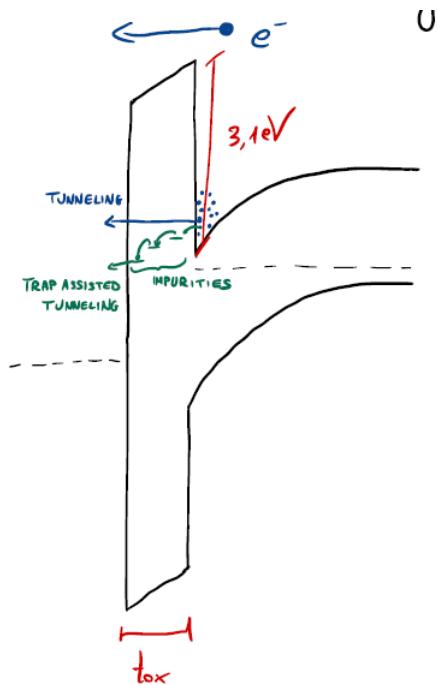
Scaling process constantly requires t_{ox} reduction; this worsens the oxide insulation as electrical conductivity goes up.



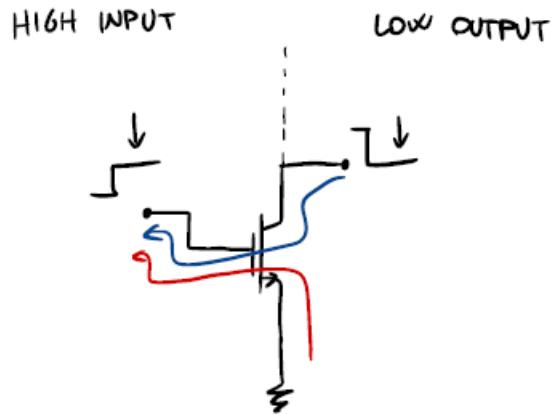
Our entire device is NOT under TE

↳ net processes are allowed

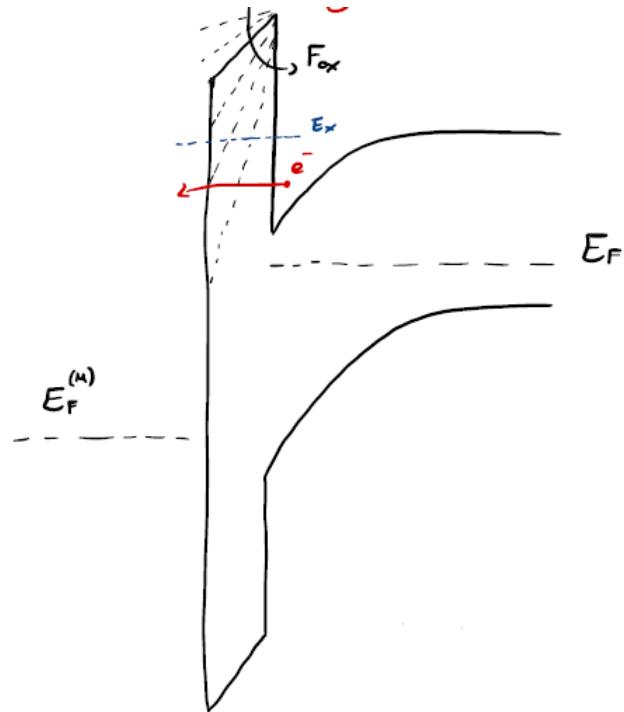
Electrons tend to move from the substrate to the gate, but are normally blocked by the SiO_2 barrier. Said barrier is finite, though, therefore if $E_{e^-} > 3.18 \text{ eV}$ the electron can actually jump to the gate (high V_{GS} , high V_{DS}). At $T = 300 \text{ K}$, these energy values are hardly reached, but still higher temperatures and hot electrons may cause some problems (channel hot electron injection).



We get static power dissipation due to the gate leakage current.

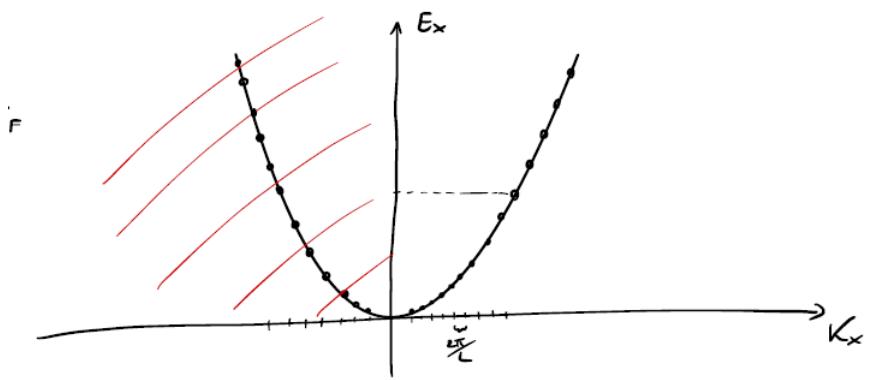


Tunneling We focus our attention just on the tunneling effect.



We have, of course, to recall some Quantum Physics, as we did for module 3.

$$E = E_c + \frac{\hbar^2 K_x^2}{2m_x} + \frac{\hbar^2 K_y^2}{2m_y} + \frac{\hbar^2 K_z^2}{2m_z}$$



As usual, we consider just the energies on the right branch of the parabola. Let's start with the discrete tunneling current

density

$$J_{\text{TUN},1} = \sum_{k_x > 0} \frac{q}{L^3} v_x(k_x) \mathbf{f}(k_x, k_y, k_z) T(E_x) \quad (5.95)$$

and immediately putting it into a continuous space.

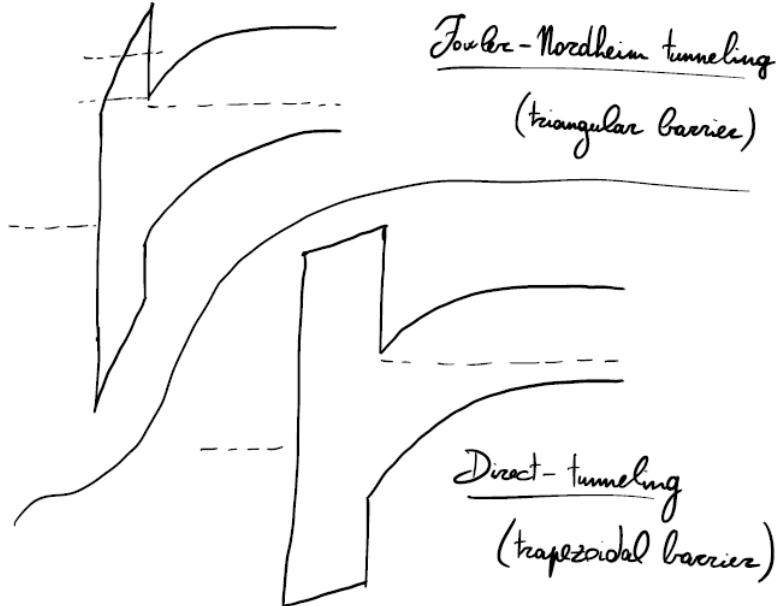
$$J_{\text{TUN},1} = \frac{1}{\left(\frac{2\pi}{L}\right)^3} \int_0^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{q}{L^3} v_x(k_x) \mathbf{f}(k_x, k_y, k_z) T(E_x) dk_x dk_y dk_z \quad (5.96)$$

After a few calculations similar to last time, we eventually get

$$J_{\text{TUN}} = \frac{q(4\sqrt{m_t m_l} + 2m_t)KT}{2\pi^2 \hbar^3} \int_0^{+\infty} T(E_x) \log \left(1 + e^{-\frac{E_x - E_F}{KT}} \right) dE_x \quad (5.97)$$

$$T(E_x) = \exp \left(-2 \int_0^{\bar{x}} \frac{\sqrt{2m^*(E_c - E_x)}}{\hbar} dx \right) \quad (5.98)$$

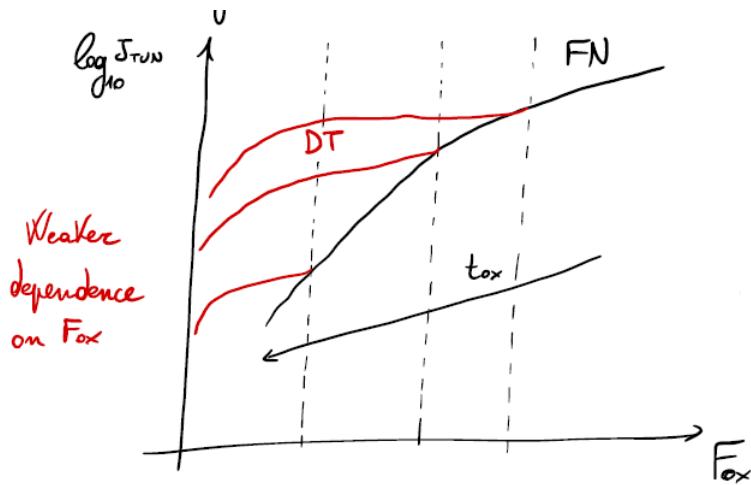
We can exploit some approximations:



For Fowler-Nordheim we get

$$\begin{aligned} J_{\text{TUN}} &= AF_{ox}^2 \exp \left(-\frac{B}{F_{ox}} \right) \\ A &= \frac{q^2(4\sqrt{m_t m_l} + 2m_t)}{16\pi^2 \hbar (q\phi - E_F) m^*} \\ B &= \frac{4\sqrt{2m^*}}{3\hbar q} (q\phi - E_F)^{3/2} \end{aligned} \quad (5.99)$$

where **this is the barrier height**. We have just discovered that there is an exponential dependence on the electric field in the oxide.



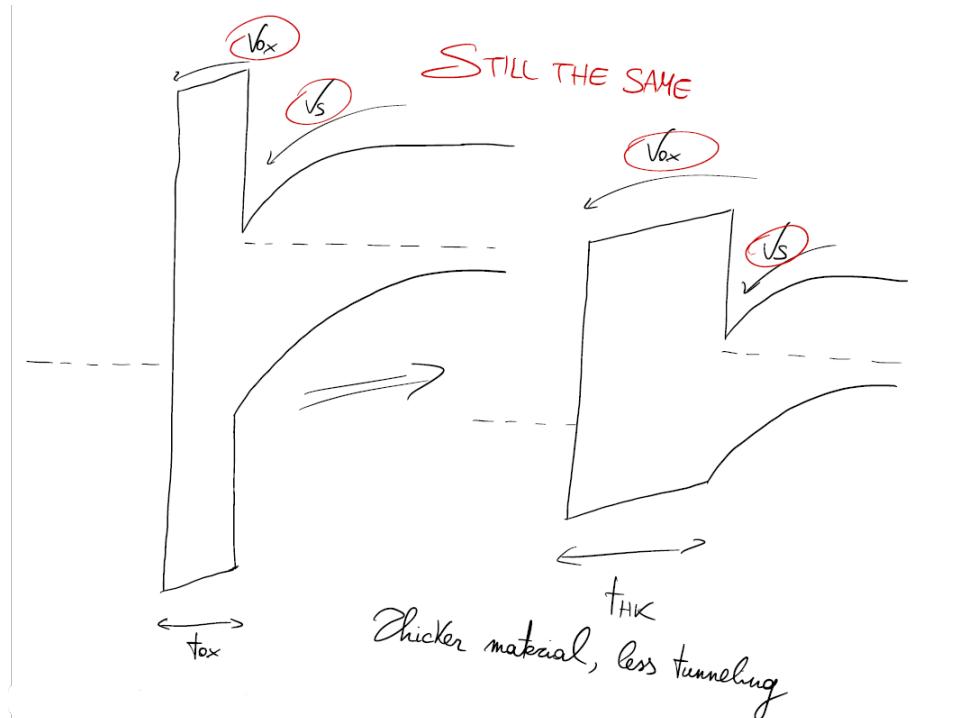
We introduce now the *Fowler-Nordheim condition*:

$$F_{ox} t_{ox} = 3.1 \text{ eV} \quad \Rightarrow \quad F_{ox} = \frac{3.1 \text{ eV}}{t_{ox}} \quad (5.100)$$

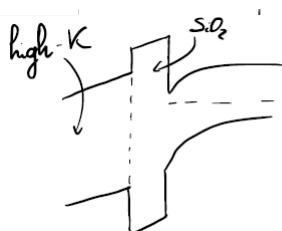
When t_{ox} decreases, there is an exponential increment in J_{TUN} (barrier also changes with F_{ox}).

High-K insulators A possible solution to avoid any leakage current to the gate is the introduction of **high-K insulators**.

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} = \frac{\epsilon_{HK}}{t_{HK}} = C_{HK} \quad \Rightarrow \quad t_{HK} = \frac{\epsilon_{HK}}{\epsilon_{ox}} t_{ox} > t_{ox} \quad (5.101)$$



There is a major drawback though; it is true that we suffer less from tunneling, but high-K materials have a lower barrier and more defect, so process-wise SiO_2 is better.



This is a good combination to avoid spurious states on the interface
A good high-K material is HfO_2

5.17 Design rules

As the final section, a bunch of good design rules are presented. In general, scaling rules alone are not enough, so we need to define and/or quantify a few more parameters:

- Optimal SubThreshold Slope

$$\text{STS} = \frac{KT}{q} \log(10)m \leq 85 \text{ mV/dec} \longrightarrow m \leq 1.4 \quad (5.102)$$

With m so quantified, we can relate oxide thickness and maximum depletion width.

$$m = 1 + \frac{C_{\text{DEP}}}{C_{ox}} = 1 + \frac{\epsilon_{\text{Si}}}{W_{D_{\max}}} \frac{t_{ox}}{\epsilon_{ox}} \approx 1 + \frac{3t_{ox}}{W_{D_{\max}}} \leq 1.4 \implies t_{ox} \leq 0.13 W_{D_{\max}} \quad (5.103)$$

- Maximum threshold voltage shift

$$\Delta V_T^{\text{SCE}} \propto \exp \left(-\frac{\pi L}{2(W_{D_{\max}} + 3t_{ox})} \right) \leq 100 \text{ mV}$$

↓

$$L \geq 2(W_{D_{\max}} + 3t_{ox}) \implies t_{ox} \leq \frac{L}{6} - \frac{W_{D_{\max}}}{3} \quad (5.104)$$

L depends on both t_{ox} and $W_{D_{\max}}$

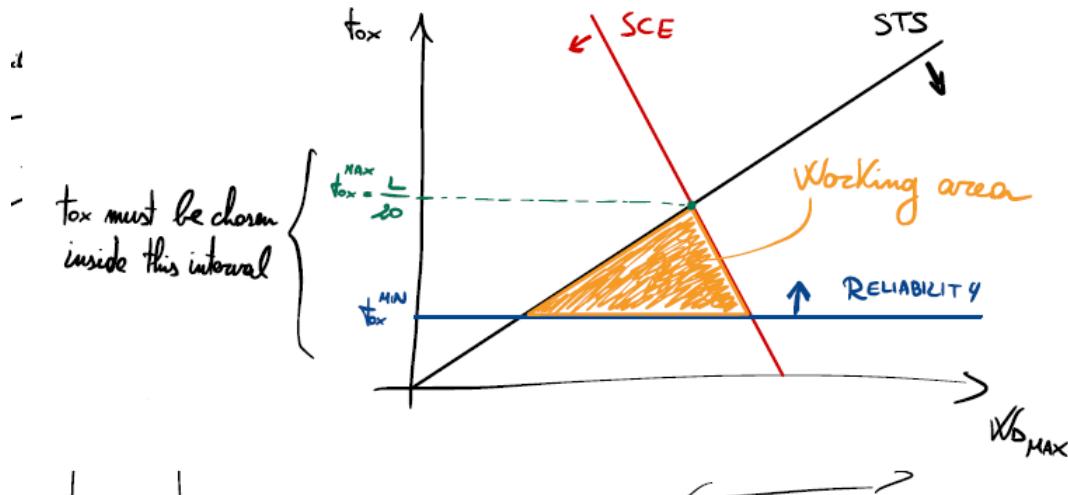
- Maximum electric field in the oxide (reliability requirement)

$$F_{ox} = \frac{V_{DD}}{t_{ox}} \leq F_{ox}^{\max} \implies t_{ox} \geq \frac{V_{DD}}{F_{ox}^{\max}} \quad (5.105)$$

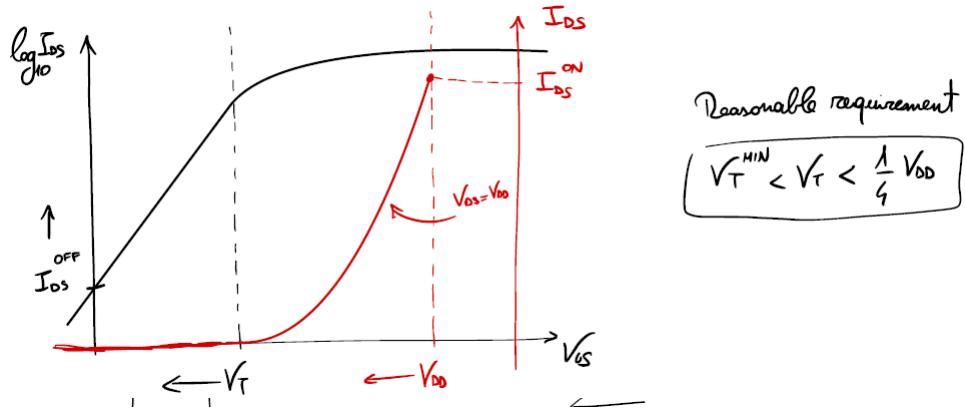
All these requirements must be satisfied simultaneously, so

$$t_{ox}^{\max} > t_{ox}^{\min} \implies \frac{L}{20} > \frac{V_{DD}}{F_{ox}^{\max}} \implies V_{DD} < \frac{L}{20} F_{ox}^{\max} \quad (5.106)$$

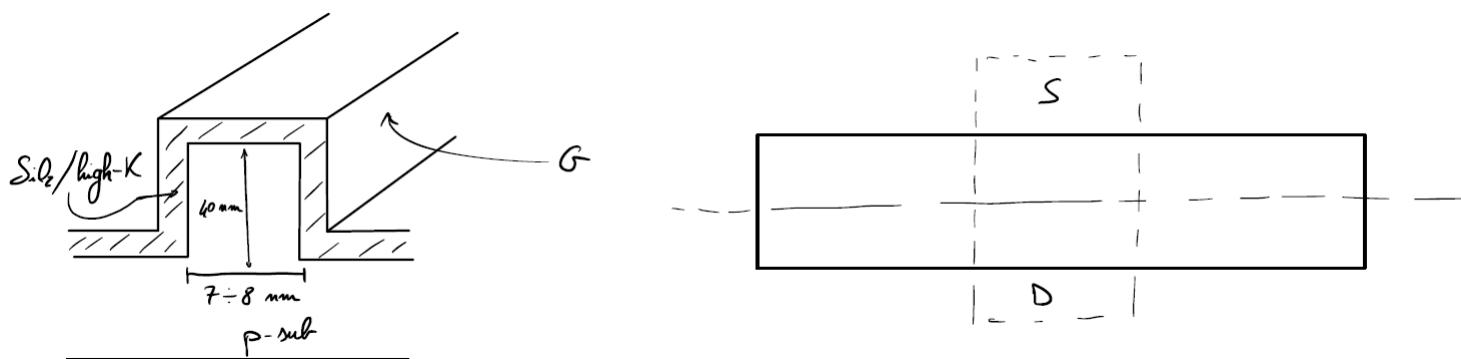
V_{DD} and L may seem independent, but they are not. We can condense all of this in a useful graphic.



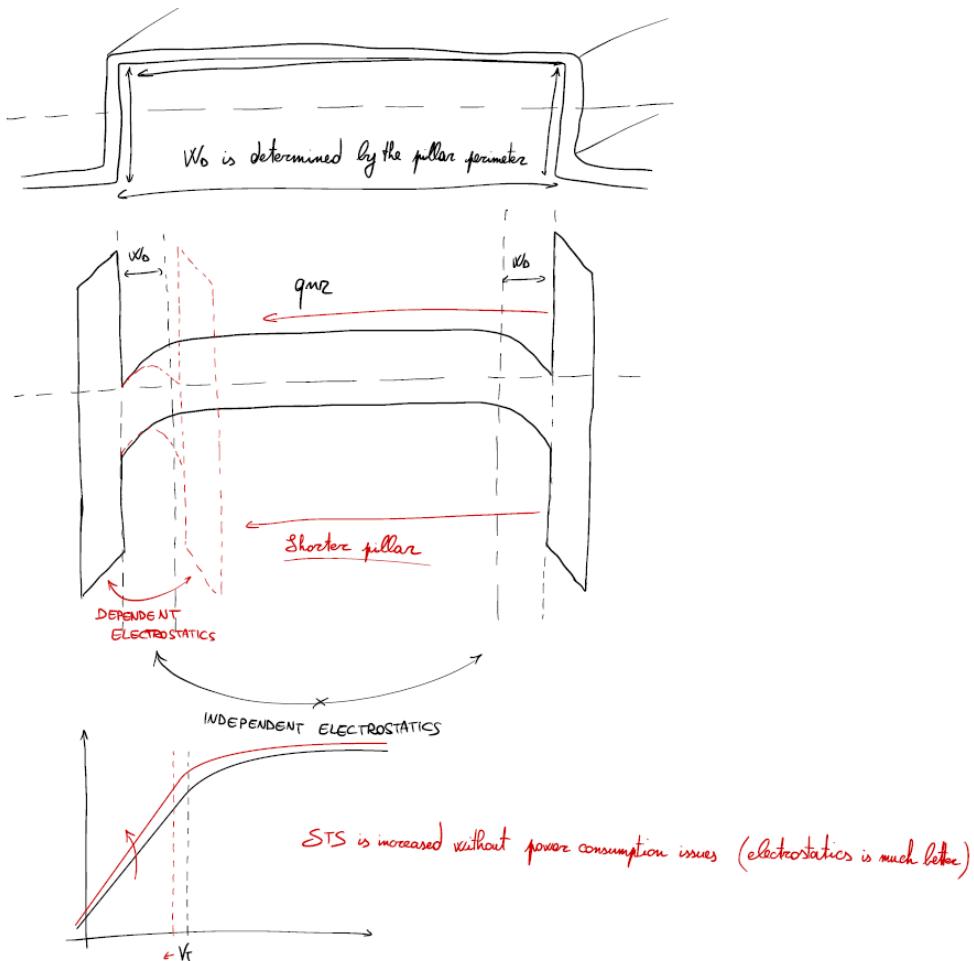
Even V_T requirements must be satisfied to have a good device, even though it is not easy to mediate among all the parameters.



We ran out of design interval for V_T in the MOSFET; it is time to jump to another device, this time tridimensional: the FINFET.

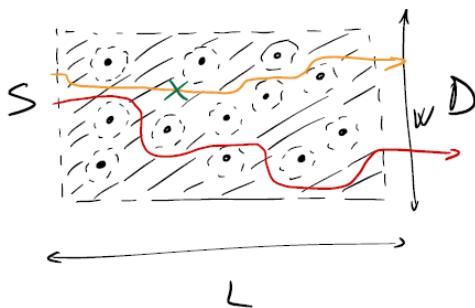


Gate overlaps the pillar on three sides: it has a better electrostatic control over the substrate.



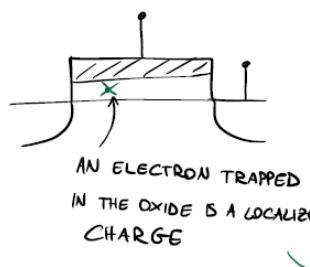
Moore's law This is not a proper physical law, it is more like an empiric statement. Complexity (integration density) of the Integrated Circuits doubles every two years, therefore a device minimum technology-related channel length shrinks with $\sqrt{2}$ proportionality. In the last years, though, this law has been mended as the time to create a new model has become longer than two years.

Finally, it is a concern of ours to show how the atomistic nature of the dopants actually affects our analysis.



Due to atomistic nature of dopants, current flow is not uniform in the channel, especially when L is short

↳ Strong variability in the parameters



AN ELECTRON TRAPPED
IN THE OXIDE IS A LOCALIZED
CHARGE

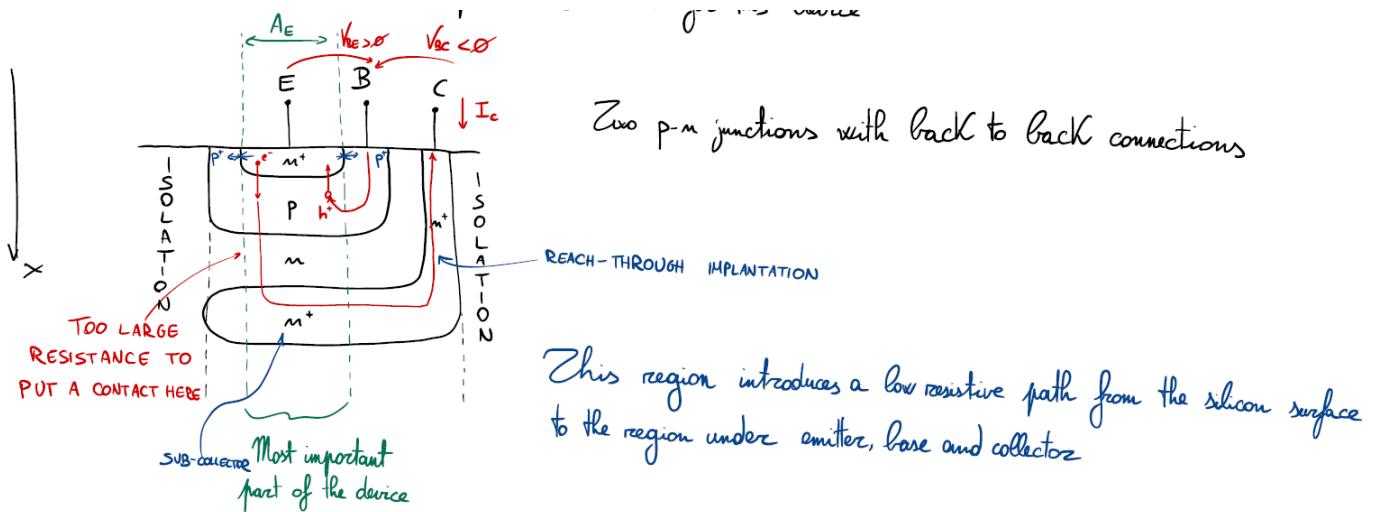
IT CAN STOP THE CURRENT BELOW IT

Chapter 6

Bipolar transistor (BJT)

6.1 Basics of the BJT

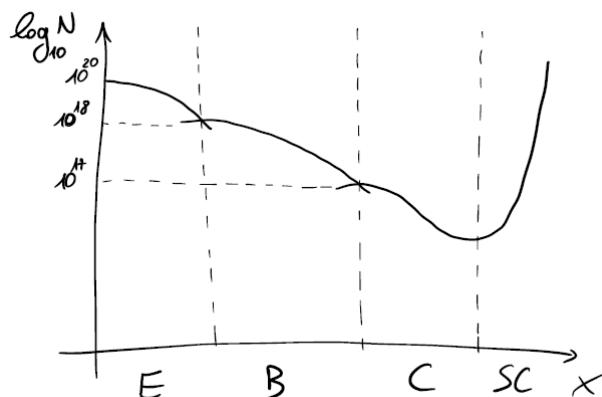
Even though this technology is older than the MOSFET, for some applications (e.g. high-gain circuits) it is still preferred. The most important structure for this device is the vertical *n-p-n* junction.



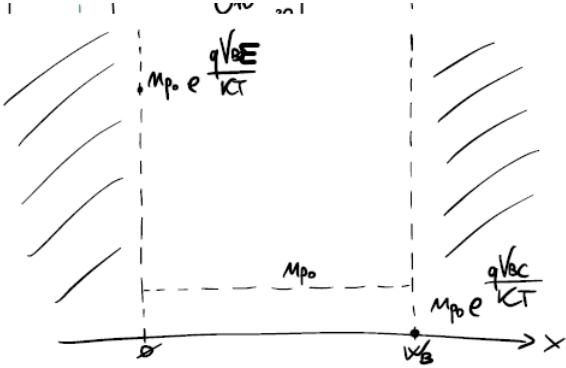
If the *p*-region is narrow there are no recombinations, so we have a net flow of electrons from the emitter to the collector. There is also a hole flow from the base to the emitter, which is not wanted, but unavoidable. Thus we can define

$$\beta = \frac{I_C}{I_B} \approx 100 \quad (6.1)$$

and we want it to be as high as possible; to achieve that, doping concentration in the emitter must be higher than the one in the base (unilateral *p* – *n* junction) and electrons injected in the base must not recombine ($W_B < 100 \text{ nm}$), otherwise we would have just an emitter-base current. Furthermore, only if the two *p* – *n* junctions interact we have a transistor. *Side parts of the base are p⁺-doped to avoid horizontal current*. Finally, doping concentration in the base region must be non-constant.

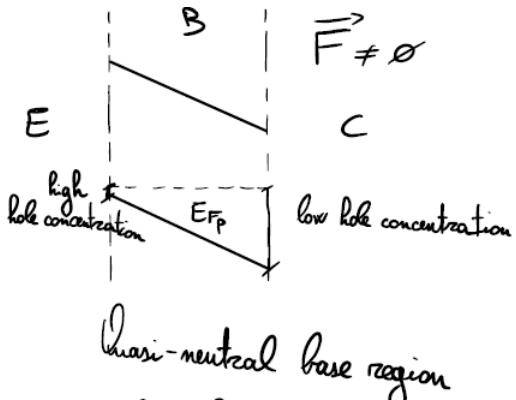


In this device, the bottleneck region for the current is the quasi-neutral base region.



These are the boundary conditions

Let's see in detail the bands in the base region:



If doping were uniform



What about the electric field?

$$p = n_i \exp\left(\frac{E_i - E_{F_p}}{KT}\right) \implies E_i - E_{F_p} = KT \log\left(\frac{p}{n_i}\right)$$

$$\phi = -\frac{E_i}{q} \implies F = -\frac{d\phi}{dx} = \frac{1}{q} \frac{dE_i}{dx} = \cancel{\frac{1}{q} \frac{dE_{F_p}}{dx}} + \frac{KT}{q} \frac{\cancel{p}}{p} \frac{\cancel{\frac{dp}{dx}}}{\cancel{p}} = \frac{KT}{q} \frac{1}{p} \frac{dp}{dx}$$
(6.2)

To be fair $\frac{dE_{F_p}}{dx} \neq 0$ exactly:

$$J_p = p\mu_p \frac{dE_{F_p}}{dx} \implies \frac{dE_{F_p}}{dx} = \frac{J_p}{p\mu_p} = \frac{J_B}{\beta p\mu_p} = \frac{J_C}{(6.3)}$$

But considering these typical numbers

$$\begin{cases} J_C = 10^5 \text{ A/cm}^2 \\ N_A = 10^{18} \text{ cm}^{-3} \\ \beta = 100 \\ \mu_p = 150 \text{ cm}^2/\text{Vs} \end{cases} \implies \frac{1}{q} \frac{dE_{F_p}}{dx} = 40 \text{ V/cm} \implies W_B = 100 \text{ nm} \implies \Delta V_{B,E_{F_p}} = 0.4 \text{ mV}$$
(6.4)

we can say that its contribution is negligible. Now we define the electron current density as

$$J_n = qn\mu_n F + qD_n \frac{dn}{dx} = qn\mu_n \frac{KT}{q} \frac{1}{p} \frac{dp}{dx} + qD_n \frac{dn}{dx}$$
(6.5)

and we consider it for low-injection and high-injection regimes. For the first

$$J_n = qn\mu_n \frac{KT}{q} \frac{1}{N_A^B} \frac{dN_A^B}{dx} + qD_n \frac{dn}{dx} = qn\mu_n F_0 + qD_n \frac{dn}{dx}$$
(6.6)

where F_0 is the built-in electric field; then for the second

$$J_n \approx q\kappa\mu_n \frac{KT}{q} \frac{1}{\kappa} \frac{dn}{dx} + qD_n \frac{dn}{dx} = 2qD_n \frac{dn}{dx}$$
(6.7)

This is a **pure diffusion process** (Webster effect).

6.2 Calculation of I_C

Now we investigate the collector current in both low-level and high-level of injection. Let's start from the first.

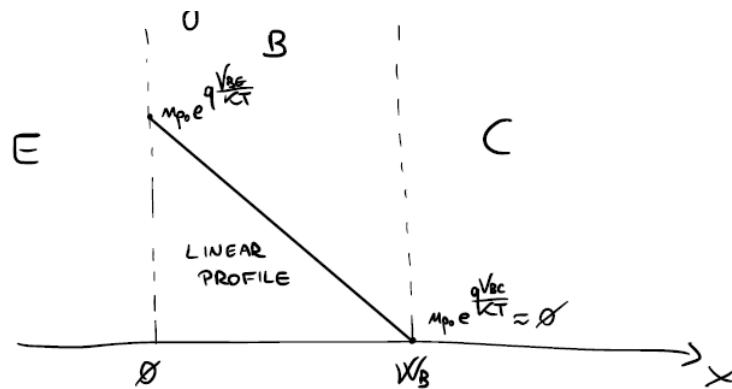
Low-level of injection We already solved the Poisson equation for this:

$$\begin{aligned} J_n &= qD_n \frac{dn}{dx} = \text{constant} \\ \int_0^{W_B} J_n dx &= \int_{n(0)}^{n(W_B)} qD_n dn \\ J_n W_B &= qD_n [n(W_B) - n(0)] \\ J_n &= -qD_n \frac{n(W_B) + n(0)}{W_B} \end{aligned} \quad (6.8)$$

The actual current would be

$$I_C = -A_E J_n = \frac{A_E q n_i^2 e^{\frac{qV_{BE}}{KT}}}{\frac{N_A^B W_B}{D_n}} = \frac{A_E q n_i^2}{G_B} e^{\frac{qV_{BE}}{KT}} \quad (6.9)$$

where A_E is the collector area and G_B is the Gummel number of the base. Keep in mind that W_B is the width of the quasi-neutral base region, not the actual physical width; this is called *prototype transistor*.



High-level of injection This time it is not a pure diffusion current.

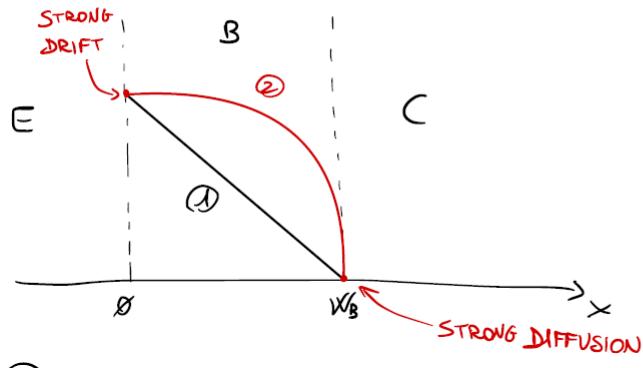
$$\begin{aligned} J_n &= qn\mu_n F + qD_n \frac{dn}{dx} = qn\mu_n \frac{KT}{q} \frac{1}{p} \frac{dp}{dx} + qD_n \frac{dn}{dx} = \frac{qD_n}{p} \left[n \frac{dp}{dx} + p \frac{dn}{dx} \right] = \frac{qD_n}{p} \frac{d(pn)}{dx} = \text{constant} \\ \int_0^{W_B} J_n \frac{p}{qD_n} dx &= \int_{pn(0)}^{pn(W_B)} d(pn) \\ J_n \int_0^{W_B} \frac{p}{qD_n} dx &= px(W_B) - pn(0) = -n_i^2 e^{\frac{qV_{BE}}{KT}} \\ J_n &= -\frac{n_i^2 e^{\frac{qV_{BE}}{KT}}}{\int_0^{W_B} \frac{p}{qD_n} dx} \end{aligned} \quad (6.10)$$

Thus the current is

$$I_C = \frac{A_E q n_i^2 e^{\frac{qV_{BE}}{KT}}}{\int_0^{W_B} \frac{p}{D_n} dx} = A_E \frac{q n_i^2}{G_B} e^{\frac{qV_{BE}}{KT}} \quad (6.11)$$

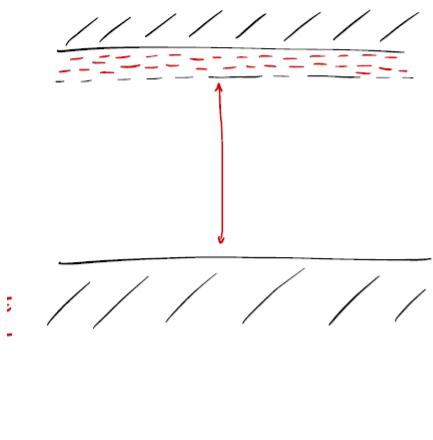
which is the same as before, except for the Gummel number being an integral.

Drift term is proportional to F



Still this isn't the most general case

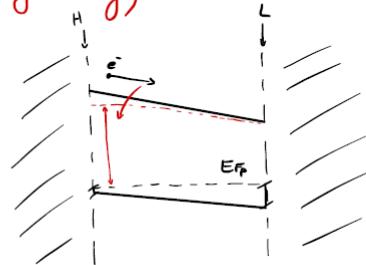
Band-gap narrowing This effect happens when the doping concentration is high.



Wave functions of the atoms start interacting with the defect level, generating sub-energy bands (degeneracy)

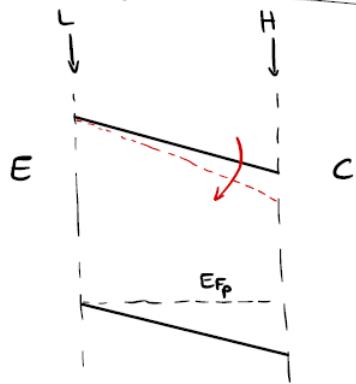
$N \geq 10^{18} \text{ cm}^{-3}$ to cause this

Negative effect



Of course we don't want that to happen, so we add different Ge concentrations to the Si-substrate (SiGe) to voluntarily reduce the band-gap at $x = W_B$.

Engineering of the base material ($\text{Si}_x\text{Ge}_{1-x}$)



We want to change E_{GAP}

$\text{Si}_x\text{Ge}_{1-x}$

Positive effect

$$E_{F_n} = E_{F_p} + kT \log \left(\frac{p_n}{n_e^2} \right)$$

$$p_n = n_e^2 \exp \left(\frac{E_{F_n} - E_{F_p}}{kT} \right)$$

(6.12)

$$n_e^2 = N_c N_v \exp \left(-\frac{E_G - \Delta E_G}{kT} \right) = N_c N_v \exp \left(-\frac{E_G}{kT} \right) \exp \left(\frac{\Delta E_G}{kT} \right) = n_i^2 \exp \left(\frac{\Delta E_G}{kT} \right)$$

Therefore the current density becomes

$$J_n = n\mu_n \frac{dE_{F_n}}{dx} = n\mu_n KT \frac{n_{i_e}^2}{pn} \frac{d\left(\frac{pn}{n_{i_e}^2}\right)}{dx} = qD_n \frac{n_{i_e}^2}{p} \frac{d}{dx} \left(\frac{pn}{n_{i_e}^2}\right) = \text{constant}$$

$$\int_0^{W_B} J_n \frac{p}{qD_n n_{i_e}^2} dx = \int_{\frac{pn}{n_{i_e}^2}(0)}^{\frac{pn}{n_{i_e}^2}(W_B)} d\left(\frac{pn}{n_{i_e}^2}\right) \quad (6.13)$$

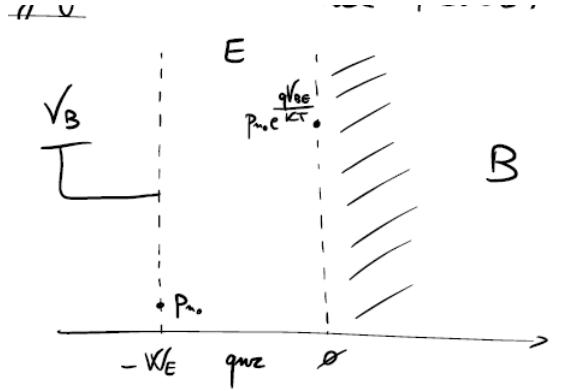
$$J_n = -\frac{e^{\frac{qV_{BE}}{KT}}}{\int_0^{W_B} \frac{p}{qD_n n_{i_e}^2} dx}$$

and so the current

$$I_C = \frac{A_E q n_i^2}{\int_0^{W_B} \frac{n_i^2}{n_{i_e}^2} \frac{p}{D_n} dx} e^{\frac{qV_{BE}}{KT}} = \frac{A_E q n_i^2}{G_B} e^{\frac{qV_{BE}}{KT}} \quad (6.14)$$

6.3 Calculation of I_B

We want the base current to be the smallest possible; in this case, bottleneck region is constituted by the quasi-neutral emitter region.



First we consider a shallow emitter, where the quasi-neutral region is very short. In this case we just repeat the calculations already done in the previous section (this time for J_p) and get

$$I_B = \frac{A_E q n_i^2}{\int_{-W_E}^0 \frac{n_i^2}{n_{i_e}^2} \frac{n}{qD_p} dx} e^{\frac{qV_{BE}}{KT}} = \frac{A_E q n_i^2}{G_E} e^{\frac{qV_{BE}}{KT}} \quad (6.15)$$

Also, under low-level injection

$$G_E \approx \frac{n_i^2}{n_{i_e}^2} \frac{N_D^E}{D_p} W_E \quad (6.16)$$

If instead we have a deep emitter (wide emitter), we should consider also G/R processes; let's take a simple case (low-injection, N_D^E constant and no ΔE_G) as an example:

$$J_p = \frac{qD_p p n_{i_0} e^{\frac{qV_{BE}}{KT}}}{L_p \tanh\left(\frac{W_E}{L_p}\right)} \implies I_B = \frac{A_E q n_i^2}{\frac{N_D^E L_p}{D_p} \tanh\left(\frac{W_E}{L_p}\right)} e^{\frac{qV_{BE}}{KT}} = \frac{A_E q n_i^2}{G_E} e^{\frac{qV_{BE}}{KT}} \quad (6.17)$$

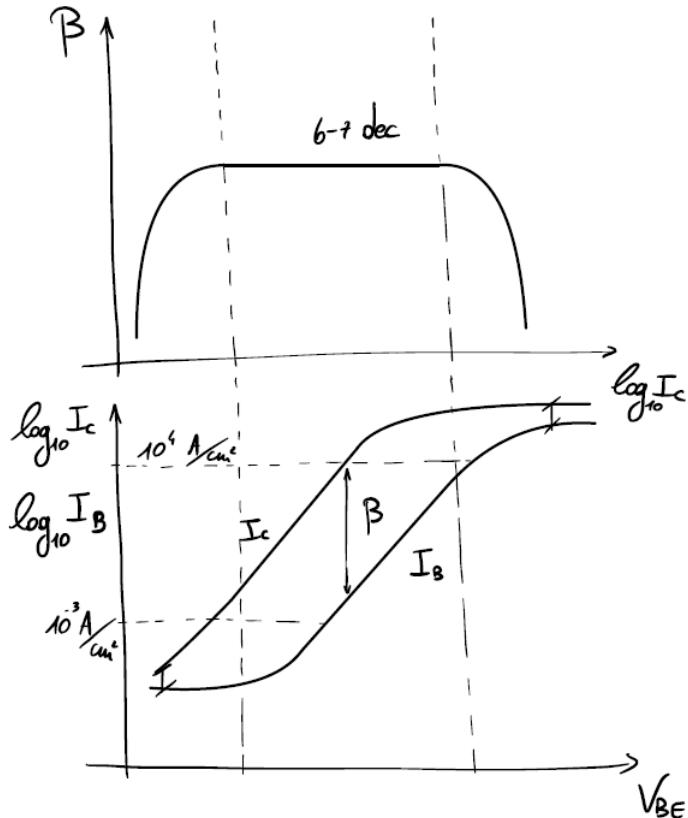
Gummel number is the one that keeps changing. Now let's spend some word for β : as we know, this parameter should be as high as possible; let's rewrite 6.1 as

$$\beta = \frac{A_E q n_i^2}{G_B} e^{\frac{qV_{BE}}{KT}} \frac{G_E}{A_E q n_i^2 e^{\frac{qV_{BE}}{KT}}} = \frac{G_E}{G_B} \quad (6.18)$$

Thus we need $G_E > G_B$; to make it so, doping concentration is the most important parameter to play with. If we then add the band gap narrowing into the equation, it is useful to compare the two Gummel numbers.

$$\begin{cases} G_E = \int_{-W_E}^0 \frac{n_i^2}{(n_{i_e}^2)_E} \frac{n}{qD_p} dx \\ G_B = \int_0^{W_B} \frac{n_i^2}{(n_{i_e}^2)_B} \frac{p}{qD_n} dx \end{cases} \quad (6.19)$$

Reduction of G_E is stronger than G_B when ΔE_G is present, so β decreases with band-gap narrowing.



6.4 High/low-current regimes

It is time to look at the edges of our $I_C - V_{BE}$ curve; first we start with the low-current regime.

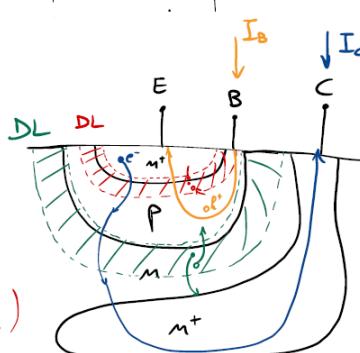
We have to pay attention to G/R processes

What happens in the two DLs?

There is a net recombination in BE DL
(forward bias junction)

I_B decreases less
(120 mV/dec trend instead of 60 mV/dec)

I_C is not affected because it is the current that passes DL (boundary conditions are the edges of quizz)
Base region is not a bottleneck for electrons



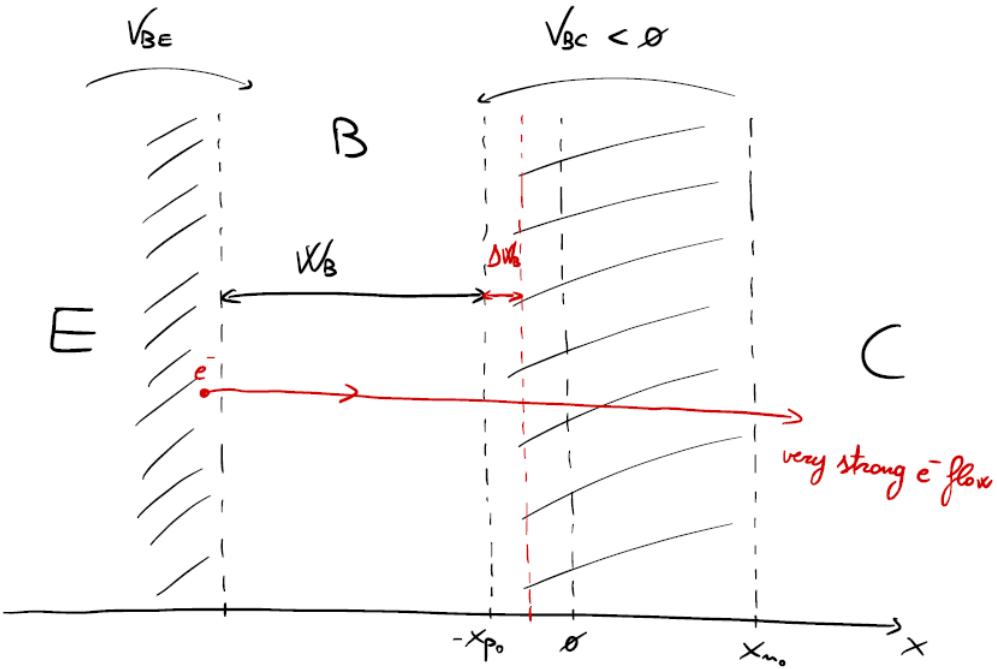
There is a net generation in BC DL
(reverse bias junction)

I_c and I_b contributions
are negligible

In the high-current regime, then, parasitic resistances do have a role as current increases less. Other than this, high-injection in base quasi-neutral region (majority carriers increase) provokes the same effect. For 6.19, it is clear that

$$I_C \propto \frac{1}{G_B}$$

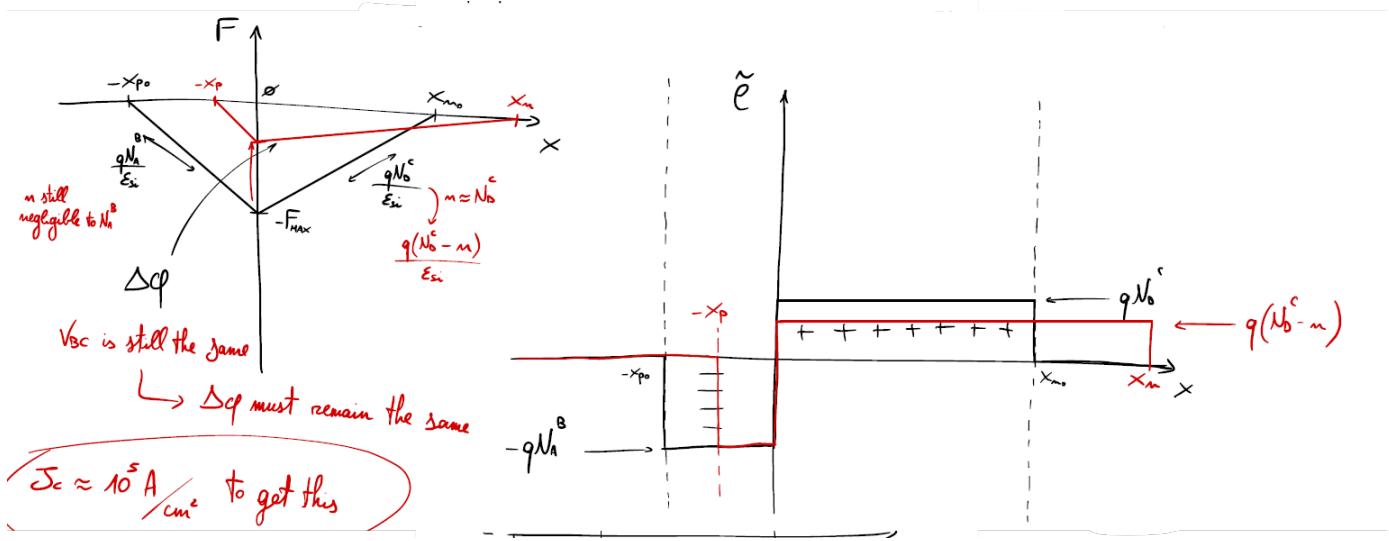
That is known as the modulation of the base conductivity. Furthermore, a third effect also appears and it is typical of the BJT: the **Kirk effect** (base-widening effect).



We assume that electrons move under velocity saturation, so

$$J_n = qnv_{\text{SAT}} = J_c \implies n = \frac{J_c}{qv_{\text{SAT}}} \approx N_D^C \quad (6.20)$$

n increases along with J_c until it becomes comparable with the doping concentration, then there is no depletion layer anymore. This happens because of the two interactive $p-n$ junctions.



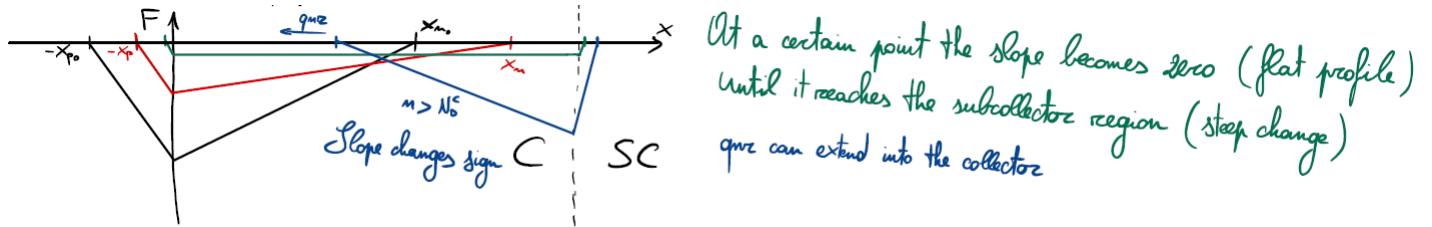
In comparison we have

$$\begin{aligned} N_A^B x_{p0} &= N_D^C x_{n0} \\ \Delta\phi &= \frac{qN_D^C}{2\epsilon_{\text{Si}}} x_{n0}^2 + \frac{qN_A^B}{2\epsilon_{\text{Si}}} x_{p0}^2 \end{aligned} \iff \begin{aligned} (N_A^B + n)x_p &= (N_D^C - n)x_n \\ \Delta\phi &= \frac{q(N_D^C - n)}{2\epsilon_{\text{Si}}} x_n^2 + \frac{q(N_A^B + n)}{2\epsilon_{\text{Si}}} x_p^2 \end{aligned} \quad (6.21)$$

and so

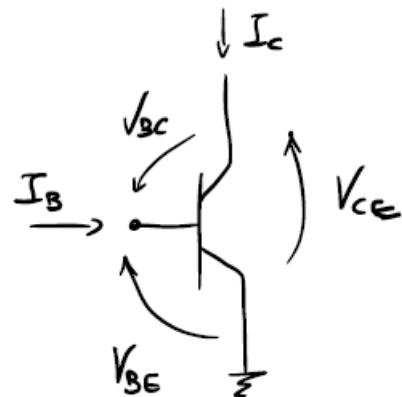
$$\begin{cases} x_n = x_{n0} \sqrt{\frac{1 + n/N_A^B}{1 - n/N_D^C}} \approx \frac{x_{n0}}{\sqrt{1 + n/N_D^C}} > x_{n0} \\ x_p = x_{p0} \sqrt{\frac{1 - n/N_D^C}{1 + n/N_A^B}} \approx x_{p0} \sqrt{1 - \frac{n}{N_D^C}} < x_{p0} \end{cases} \quad (6.22)$$

Which is the limit?



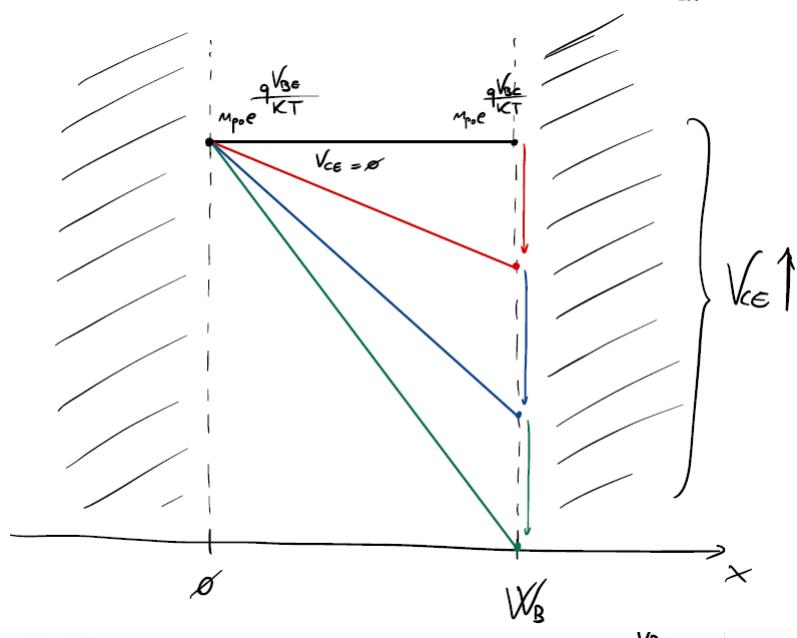
6.5 $I_C - V_{CE}$ curve

Here is presented the most important curve for the BJT.

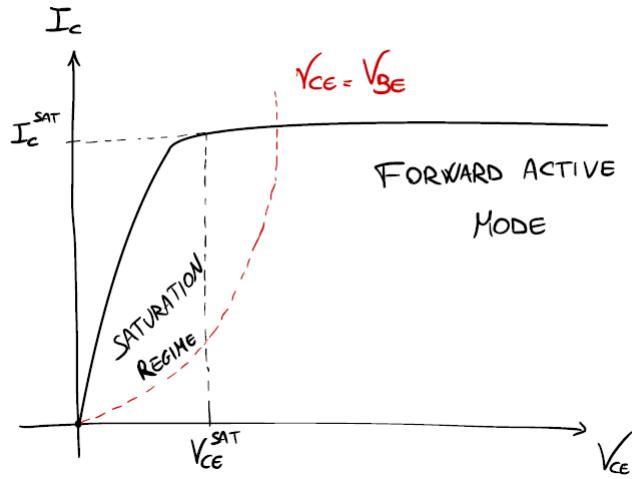


$$V_{CE} = V_{BE} - V_{BC} \quad (6.23)$$

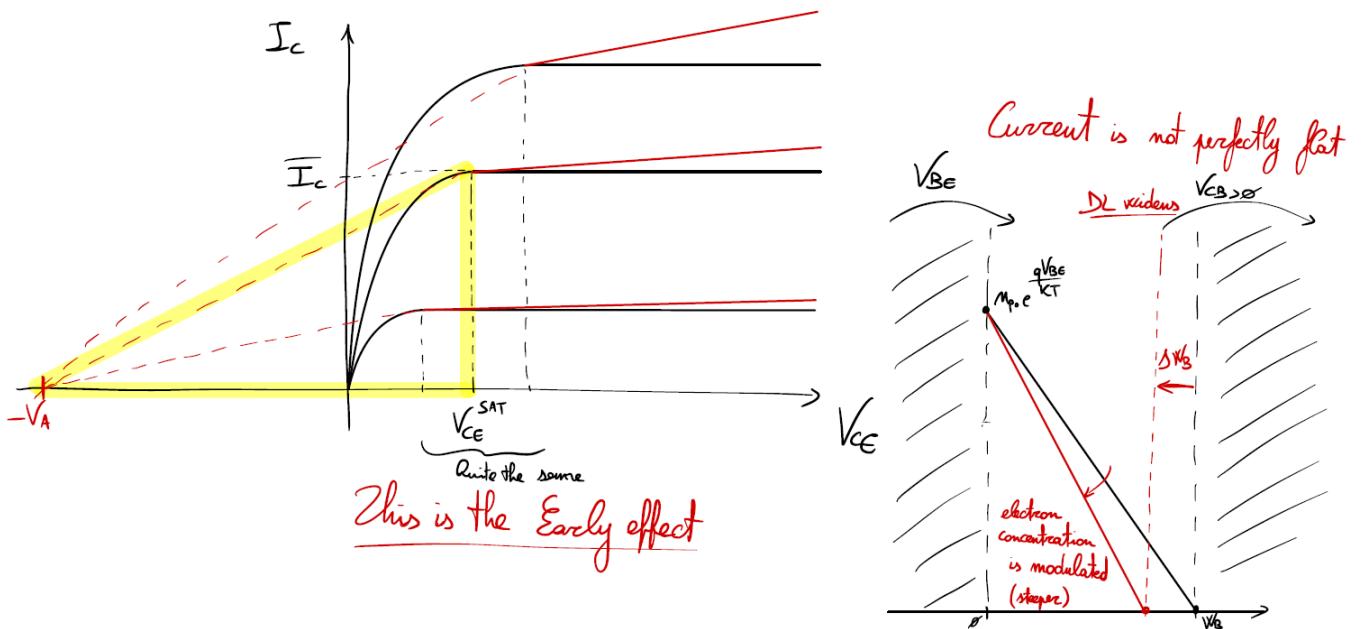
First we consider $V_{CE} = 0$: the two junctions are under forward bias, so holes must be considered. As a consequence of that, we have a small negative I_C which can be easily considered as null.



After that, I_C increases exponentially with V_{CE} until V_{BC} is below V_{BE} by at least KT/q . This is when $V_{CE} = V_{CE}^{\text{SAT}}$ and we enter **forward active regime**.



We do want to work in this regime. Also, Early effect is present as well.



We can easily find V_A with this procedure: first we look at the drawing and see

$$V_A + V_{CE}^{SAT} \xrightarrow{100\text{ mV}} = \frac{\bar{I}_C}{\partial I_C / \partial V_{CE}} \quad (6.24)$$

then we do some calculations

$$I_C = A_E \frac{q n_i^2 D_n}{N_A^B E_B} e^{\frac{q V_{BE}}{kT}}$$

$$\frac{\partial I_C}{\partial V_{CE}} = \frac{A_E q n_i^2 D_n}{N_A^E E_B} e^{\frac{q V_{BE}}{kT}} \left(-\frac{1}{W_B^2} \right) \frac{\partial W_B}{\partial V_{CE}} = -\frac{I_C}{W_B} \frac{\partial W_B}{\partial V_{CE}}$$

$$\frac{\partial W_B}{\partial V_{CE}} = \frac{\partial W_B}{\partial V_{CB}} = -\frac{\partial x_p}{\partial V_{CB}} \frac{q N_A^B}{q N_A^B} = -\frac{C_{DEP}^{BC}}{q N_A^B}$$

Combining all of these produces

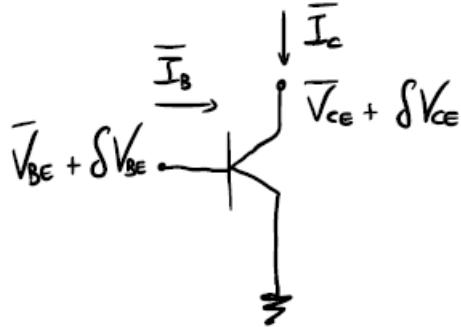
$$V_A = \frac{q N_A^B W_B}{C_{DEP}^{BC}} = \frac{Q_p^V}{C_{DEP}^{BC}} \quad (6.25)$$

If we plug these numbers

$$\begin{cases} W_B = 100 \text{ nm} \\ N_A^B = 10^{18} \text{ cm}^{-3} \\ N_D^C = 2 \cdot 10^{16} \text{ cm}^{-3} \end{cases} \implies V_A = 40 \text{ V} \quad (6.26)$$

6.6 Small signal model

Alas, we have arrived at the end. Of course, what is left to be done is the small signal model for the BJT.



Here are the parameters we need:

- Small signal transconductance

$$g_m = \left(\frac{\partial I_C}{\partial V_{BE}} \right)_{V_{CE}} = \frac{I_C}{KT/q} \quad (6.27)$$

- Input resistance

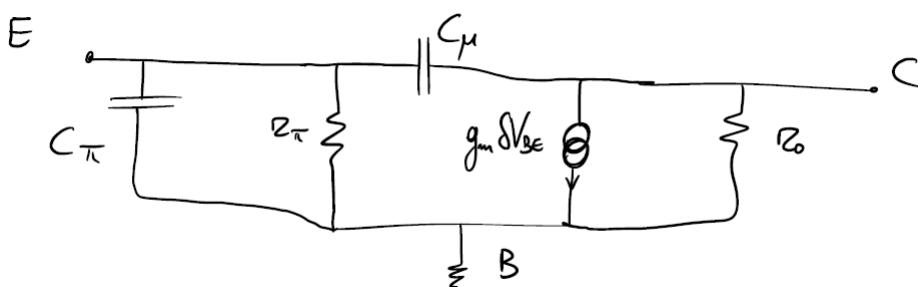
$$r_\pi = \left(\frac{\partial I_B}{\partial V_{BE}} \right)_{V_{CE}}^{-1} = \frac{\beta}{g_m} \quad (6.28)$$

- Output resistance

$$r_0 = \frac{1}{g_0} = \left(\frac{\partial I_C}{\partial V_{CE}} \right)_{V_{BE}} = \frac{V_A}{I_C} \quad (6.29)$$

- Small signal capacitances

$$\begin{cases} C_\pi = C_{DEP}^{BE} + C_{DIFF} \\ C_\mu = C_{DEP}^{BC} \end{cases} \quad (6.30)$$



What is exactly Q_{DIFF} ? It is the diffusion charge stored in the device and can be defined as

$$Q_{\text{DIFF}} = Q_B + Q_E + Q_{BC} + Q_{BE} \quad (6.31)$$

Let's study in details each charge:

- Extra charge in quasi-neutral base region

$$Q_B = I_C t_B \quad (6.32)$$

$$t_B = \frac{W_B^2}{2D_n} \quad (6.33)$$

- Extra charge coming from the minority carriers in quasi-neutral emitter region

$$Q_E = I_B \tau_p = \frac{I_C}{\beta} \tau_p = I_C t_E \quad (6.34)$$

- Free charge from the depletion layer between base and collector regions

$$Q_{BC} = qnW_D^{BC}A_E = qW_D^{BC}A_E \frac{J_c}{qv_{SAT}} = I_C \frac{W_D^{BC}}{v_{SAT}} = I_C t_{BC} \quad (6.35)$$

- Diffusion charge from the depletion layer between base and emitter (no v_{SAT})

$$Q_{BE} = I_C t_{BE} \quad (6.36)$$

Now let's rewrite 6.31

$$Q_{\text{DIFF}} = I_C(t_B + t_E + t_{BC} + t_{BE}) = I_C \tau_F \quad (6.37)$$

where the last time constant is the forward transit time. Trivially, we know the diffusion capacitance because

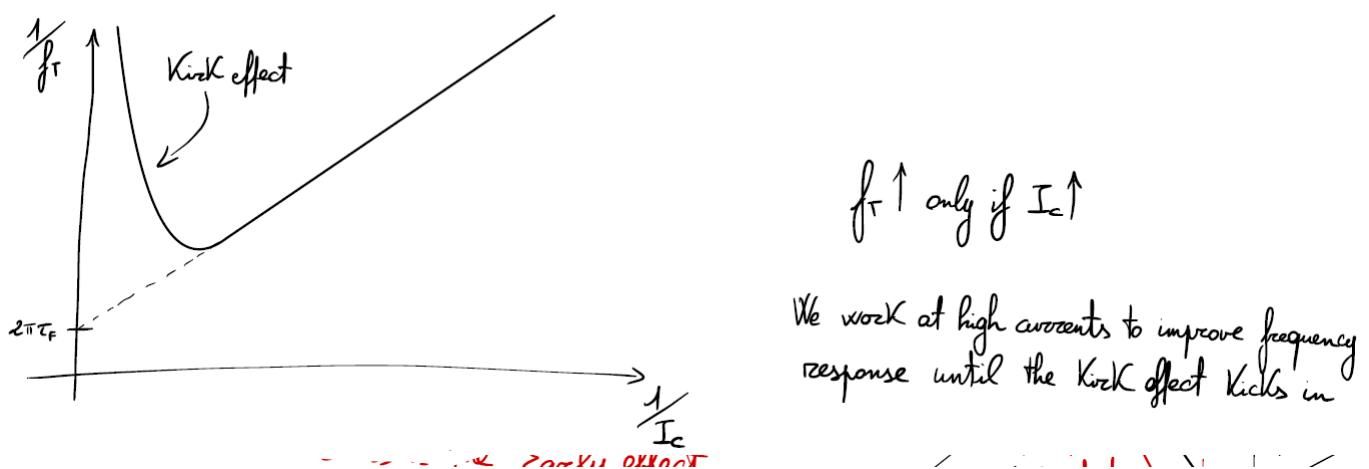
$$C_{\text{DIFF}} = \left(\frac{\partial Q_{\text{DIFF}}}{\partial V_{BE}} \right)_{V_{CE}} = g_m \tau_F \quad (6.38)$$

At this point, we can obtain the cut-off frequency:

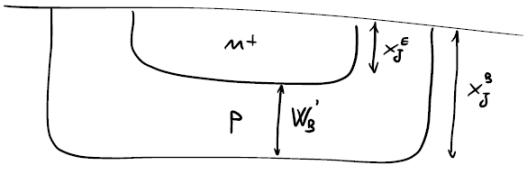
$$f_T = \frac{g_m}{2\pi(C_\pi + C_\mu)} = \frac{g_m}{2\pi(C_{\text{DEP}}^{BE} + C_{\text{DEP}}^{BC} + C_{\text{DIFF}})} = \frac{g_m}{2\pi(C_{\text{DEP}}^{BE} + C_{\text{DEP}}^{BC} + g_m \tau_F)} \quad (6.39)$$

We can also play a little with the last equation and get

$$\frac{1}{f_T} = \frac{2\pi(C_{\text{DEP}}^{BE} + C_{\text{DEP}}^{BC})}{I_C} \frac{KT}{q} + 2\pi \tau_F \quad (6.40)$$



Lastly, here are some considerations about the BJT structure: since it is a vertical device, the width of its most important parts results from counter-implantations of Si.



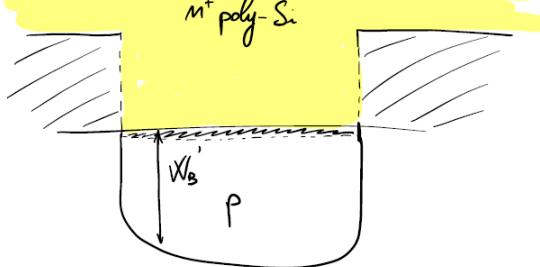
$$W'_B = X_J^s - X_J^e$$

$$W'_B \downarrow \Rightarrow f_T \uparrow$$

If the emitter is deep, I_B is less affected by variability

forced to 100 nm (not very good control on doping)

Scaling is very limited



A poly-Si emitter is introduced

W'_B can be decreased keeping a good emitter behavior

Part II

Exercise classes

Zutorial #1

$$T = 77 \text{ K}$$

$$N_D = 10^{16} \text{ cm}^{-3}$$

$$E_c - E_F = 54 \text{ meV}$$

$$E_c - E_F ?$$

n ?

p ?

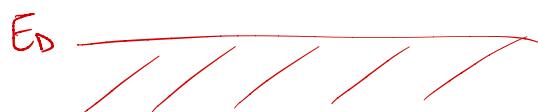
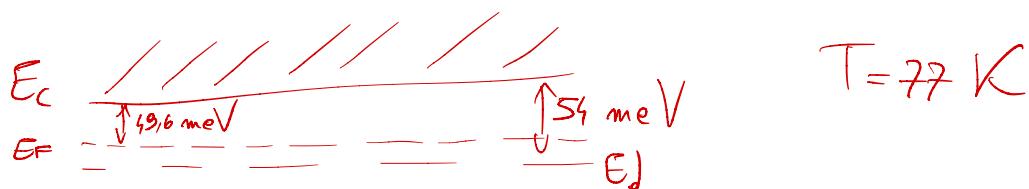
$$M = \frac{N_c'}{2} \left(\sqrt{1 + \frac{4N_D}{N_c'}} - 1 \right)$$

$$N_c' = \frac{N_c}{2} e^{-\frac{E_c - E_F}{KT}}$$

$$M = N_c e^{-\frac{E_c - E_F}{KT_{PO}}}$$

$$E_c - E_F = KT \log \left(\frac{N_c}{N_D} \right) = 46 \text{ meV}$$

$$N_D = N_{D0} + n$$



$M = \frac{n}{p + N_D}$

negligible

$$N_c e^{-\frac{E_c - E_F}{KT}} = N_D \frac{1}{1 + 2e^{-\frac{E_F - E_D}{KT}}} \Rightarrow e^{-\frac{E_c - E_F}{KT}} \left(1 + 2e^{-\frac{E_D - E_F}{KT}} \right) = \frac{N_D}{N_c}$$

$$e^{-\frac{E_c - E_F}{KT}} \left[1 + 2e^{-\frac{E_c - E_F}{KT}} e^{\frac{E_c - E_d}{KT}} \right] = \frac{N_d}{N_c}$$

X X

$$KT = 25,8 \text{ meV} \frac{77 \text{ K}}{300 \text{ K}} = 6,64 \text{ meV}$$

$$2e^{\frac{E_c - E_d}{KT}} \times^2 + \times - \frac{N_d}{N_c} = 0 \Rightarrow E_c - E_F = 49,6 \text{ meV}$$

$$N_c = 2,8 \cdot 10^{19} \text{ cm}^{-3} \left(\frac{77 \text{ K}}{300 \text{ K}} \right)^{\frac{3}{2}} = 3,64 \cdot 10^{18} \text{ cm}^{-3}$$

$$m = N_c e^{-\frac{E_c - E_F}{KT}} = 8 \cdot 10^{15} \text{ cm}^{-3}$$

$$P = \frac{m_i^2}{m} = 1,45 \cdot 10^{-55} \text{ cm}^{-3}$$

$$N_v(77 \text{ K}) = 1,35 \cdot 10^{18} \text{ cm}^{-3}$$

$$m_i^2 = N_c N_v e^{-\frac{E_g}{KT}} = 3 \cdot 10^{40} \text{ cm}^{-3}$$

$$E_g(T) = E_g(0) - \frac{\alpha T^2}{B+T} = 1,165 \text{ eV}$$

$$m = p^+ N_d^+$$

$$\begin{aligned} T &= RT \\ T &> RT \end{aligned} \quad \left\{ \begin{array}{l} \text{handle} \\ \text{together} \end{array} \right. \Rightarrow m \approx p^+ N_d \quad \text{complete ionization}$$

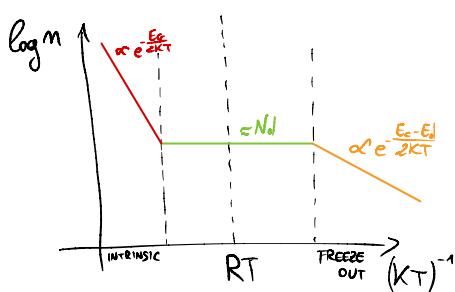
$$m = \frac{m_i^2}{m} + N_d \rightarrow m^2 - N_d m - m_i^2 = 0 \rightarrow m = \frac{N_d + \sqrt{N_d^2 + 4m_i^2}}{2}$$

$$\begin{array}{c} T=RT \\ m \approx N_d \\ T > RT \\ m \approx m_i = \sqrt{N_c N_v e^{-\frac{E_g}{2KT}}} \end{array}$$

$$T \ll RT \Rightarrow m \approx N_d^+ \rightarrow m = N_d \frac{1}{1 + 2e^{-\frac{E_c - E_F}{KT}}}$$

$$m \left(1 + 2e^{-\frac{E_c - E_F}{KT}} \right) \approx N_d \rightarrow m \left(1 + \frac{2}{N_c} N_c e^{-\frac{E_c - E_F}{KT}} e^{\frac{E_c - E_d}{KT}} \right) \approx N_d \rightarrow \frac{2}{N_c} e^{\frac{E_c - E_d}{KT}} m^2 + m - N_d = 0$$

$$m = \frac{-1 + \sqrt{1 + \frac{8N_d}{N_c} e^{\frac{E_c - E_d}{KT}}}}{\frac{4}{N_c} e^{\frac{E_c - E_d}{KT}}} \approx \sqrt{\frac{N_c N_d}{2}} e^{-\frac{E_c - E_d}{2KT}}$$



$$m = \sqrt{\frac{N_c N_d}{2}} e^{-\frac{E_c - E_d}{KT}} = N_c e^{-\frac{E_c - E_F}{KT}}$$

?

$$E_c - E_F = -KT \log \left(\sqrt{\frac{N_d}{2N_c}} e^{-\frac{E_c - E_d}{2KT}} \right)$$

$$E_F = \frac{E_c + E_d}{2} + KT \log \left(\sqrt{\frac{N_d}{2N_c}} \right)$$

Tutorial #2

Si

$$N_D = 10^{16} \text{ cm}^{-3}$$

$$E_T = E_i$$

$$\tau_0 = 10^{-6} \text{ s}$$

$$T = RT$$

Material is perturbed by changing hole concentration

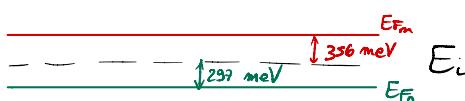
$$R = 10^{14} \text{ cm}^{-3} \text{ s}^{-1} \rightarrow 10^8 \text{ cm}^{-3} \text{ s}^{-1}$$

$$E_{F_n}, E_{F_p} ? \text{ (Q)}$$

$$E_i = E_T \Rightarrow \tau_m \approx \tau_0$$

$$\Delta n = R \tau_0 = 10^{15} \text{ cm}^{-3}$$

$$P_0 = \frac{m_i^2}{N_D} = 10^4 \text{ cm}^{-3}$$



~~$$\Delta n = \tau_0 R = 10^{16} \text{ cm}^{-3} = N_D$$~~

~~$$M = 2 \cdot 10^{16} \text{ cm}^{-3}$$~~

~~$$E_{F_n} - E_i = KT \log\left(\frac{m}{m_i}\right) = 374 \text{ meV}$$~~

~~$$E_{F_p} - E_i = -KT \log\left(\frac{P}{m_i}\right) = 356 \text{ meV}$$~~

$$PM = M_i e^{\frac{E_{F_n} - E_{F_p}}{KT}}$$

$$M = M_i e^{\frac{E_{F_n} - E_i}{KT}}$$

$$\hookrightarrow E_{F_n} - E_i = KT \log\left(\frac{M}{m_i}\right) = 356 \text{ meV}$$

$$P = M_i e^{\frac{E_i - E_{F_p}}{KT}}$$

$$\hookrightarrow E_{F_p} - E_i = KT \log\left(\frac{P}{m_i}\right) = 297 \text{ meV}$$

$$E_{F_n} - E_{F_p} = 653 \text{ meV}$$

~~$$E_{F_n} - E_{F_p} = 730 \text{ meV}$$~~

We are no more under low injection condition

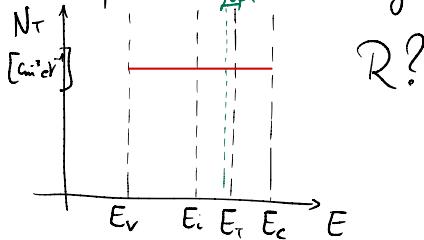
$$R = \frac{MP - M_i^2}{\tau_0 [P + M + 2M_i \cosh\left(\frac{E_T - E_i}{KT}\right)]} = \frac{(M_0 + \Delta P)(P_0 + \Delta p) - M_i^2}{\tau_0 [P_0 + \Delta p + M_0 + \Delta p + 2M_i \cosh\left(\frac{E_T - E_i}{KT}\right)]} = \frac{P_0 M_0 + P_0 \Delta p + M_0 \Delta p + \Delta p^2 - M_i^2}{\tau_0 [P_0 + M_0 + 2\Delta p + 2M_i \cosh\left(\frac{E_T - E_i}{KT}\right)]}$$

$$= \frac{\Delta p (P_0 + M_0 + \Delta p)}{\tau_0 (P_0 + M_0 + 2\Delta p)} = \frac{\Delta p}{\tau_0 (\Delta p)}$$

$$R = \frac{\Delta p(p_0 + m_0 + \Delta p)}{T_0(p_0 + m_0 + 2\Delta p)} \implies R T_0(p_0 + m_0 + 2\Delta p) = \Delta p(p_0 + m_0 + \Delta p)$$

$$\Delta p^2 + \Delta p(p_0 + m_0 - 2RT_0) - RT_0(p_0 + m_0) = 0 \implies \Delta p = 1.6 \cdot 10^{16} \text{ cm}^{-3}$$

Si quasi-neutral layer injected



$$E_{F_n} - E_i \approx 370 \text{ meV}$$

$$E_{F_p} - E_i \approx -360 \text{ meV}$$

Single energy level

$$R = \frac{\int_{E_V}^{E_c} N_T U_{TH} \sigma \Delta p(p_0 + m_0) dE_T}{T_0 [p_0 + m_0 + 2m_i \cos(\frac{E_T - E_i}{KT})]} = \Delta p(p_0 + m_0) U_{TH} \sigma N_T \int_{E_i}^{E_c} \frac{dE_T}{p_0 + m_0 + 2m_i \cosh(\frac{E_T - E_i}{KT})} = 2\Delta p(p_0 + m_0) U_{TH} \sigma N_T \int_{E_i}^{E_c} \frac{dE_T}{p_0 + m_0 + m_i e^{\frac{E_T - E_i}{KT}}} = 2\Delta p(p_0 + m_0) U_{TH} \sigma N_T \int_{E_i}^{E_c} \frac{dE_T}{p_0 + m_0 + m_i e^{\frac{E_T - E_i}{KT}}} \xrightarrow{\text{negligible due to integration interval}}$$

$m_i e^{\frac{E_T - E_i}{KT}} \ll p_0 + m_0 \implies E_T \ll E_i + KT \log\left(\frac{p_0 + m_0}{m_i}\right)$

\cosh is even

$$N_T' = \frac{N_T}{E_G} \quad R = 2\Delta p N_T' U_{TH} \sigma K T \left(\log \frac{p_0 + m_0}{m_i} - \frac{p_0 + m_0}{m_i} e^{\frac{E_T - E_i}{KT}} \right)$$

$$R = \frac{2\Delta p K T}{T_0 E_G} \quad R_o = \frac{R}{2\Delta p K T}$$

Defects further to E_i

Contribute less to recombination

$$\frac{R}{R_o} = \frac{2KT}{E_G} \quad R_o = 0.66$$

$$E_T = E_i$$

Tutorial #3

$$N_A = 10^{17} \text{ cm}^{-3}$$

$$N_D = 5 \cdot 10^{14} \text{ cm}^{-3}$$

$$F = F(x) ?$$

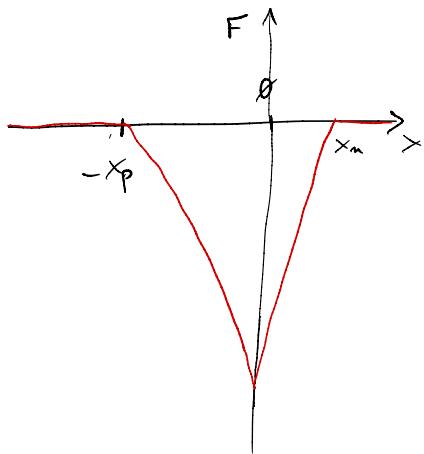
$$TE \text{ and } V = 600 \text{ mV}$$

$$W_d = \sqrt{\frac{2\varepsilon_s}{q} \left(\frac{1}{N_A} + \frac{1}{N_D} \right) (q_{B1} - V)} = 62.8 \text{ nm}$$

$$x_m = W_d \frac{N_A}{N_A + N_D} = 10.5 \text{ nm}$$

$$q_{B1} = \frac{KT}{q} \log\left(\frac{N_A N_D}{m_i^2}\right) = 854 \text{ mV}$$

$$x_p = W_d \frac{N_D}{N_A + N_D} = 52.3 \text{ nm}$$



$$F(\sigma) = F_{\max} = \frac{qN_D}{\epsilon_s} x_m = 81 \text{ KV/cm}$$

$$F_{\max}^{\text{TG}} = 167 \text{ KV/cm}$$

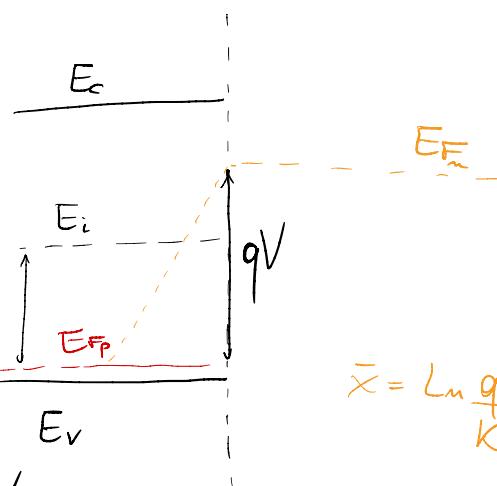
$E_{F_n}, E_{F_p}?$

WIDE-BASE

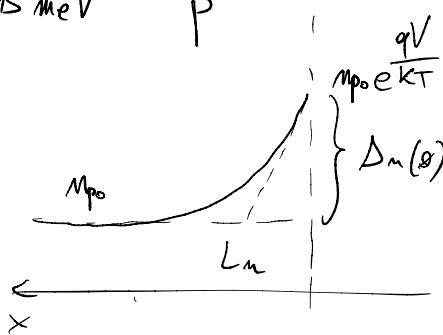
$$M_n = 750 \text{ cm}^2/\text{Vs}$$

$$T_n = 3 \mu\text{s}$$

$$E_i - E_{F_p} = KT \log \frac{N_A}{M_i} = 615 \text{ meV}$$



$$\bar{x} = L_n \frac{qV}{KT} = 1,76 \text{ mm}$$



$$L_n = \sqrt{D_n T_n} = 76,2 \mu\text{m}$$

$$M(x) = M_{p_0} + \Delta_n(x) e^{-\frac{x}{L_n}} = M_i e^{\frac{E_{F_n} - E_i}{KT}}$$

$$M_{p_0} = \frac{M_i^2}{N_A} = 10^3 \text{ cm}^{-3}$$

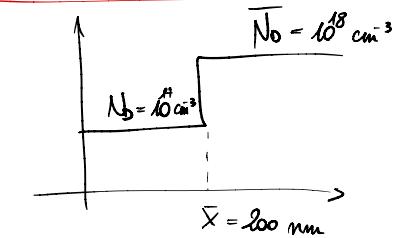
$$M_{p_0} e^{\frac{qV}{KT}} = 1,3 \cdot 10^{13} \text{ cm}^{-3}$$

$$M_{p_0} e^{\frac{qV}{KT}} e^{-\frac{x}{L_n}} = M_i e^{\frac{E_{F_n} - E_i}{KT}} \quad \text{for small } x$$

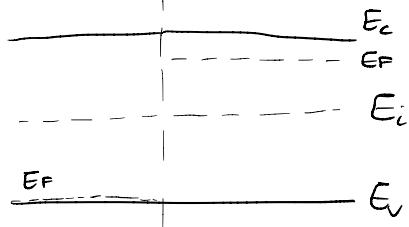
$$E_{F_n} - E_i = KT \log \left(\frac{M_{p_0}}{M_i} e^{\frac{qV}{KT}} e^{-\frac{x}{L_n}} \right) = KT \log \frac{M_{p_0}}{N_A} + KT \frac{qV}{KT} - \frac{KT}{L_n} x$$

$$E_{F_n}(x) = E_{F_p} + qV - \frac{KT}{L_n} x = E_{F_n}(\sigma) - \frac{KT}{L_n} x$$

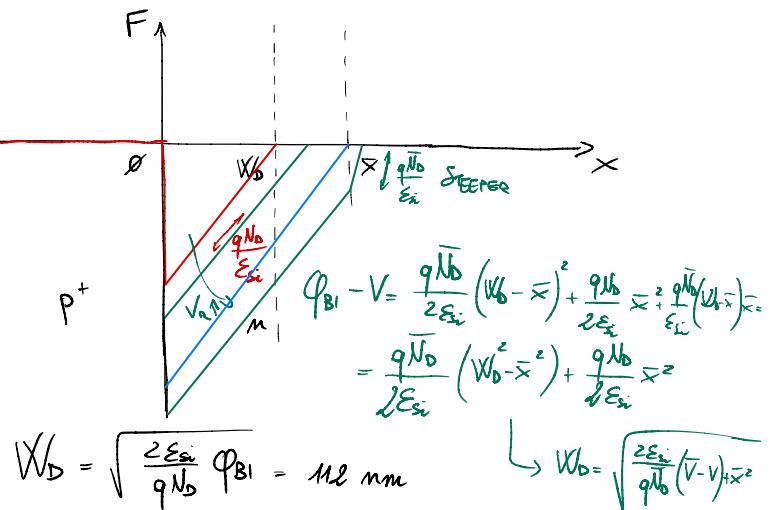
Tutorial #4



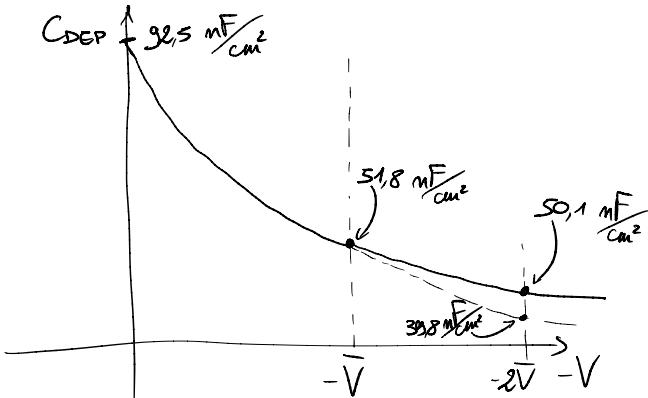
C_{DEP} ?
 W_b ?



$$\varphi_{BI} = \frac{E_g}{2q} + \frac{kT}{q} \log\left(\frac{N_d}{n_i}\right) = 0.97 \text{ V}$$



$$x = \sqrt{\frac{2\varepsilon_s}{qN_d} (\varphi_{BI} - V)} \Rightarrow V = -2.13 \text{ V}$$



$$C_{DEP} = \frac{\varepsilon_s}{W_D} = \frac{\varepsilon_s}{\sqrt{\frac{2\varepsilon_s}{qN_d} (\varphi_{BI} - V)}}$$

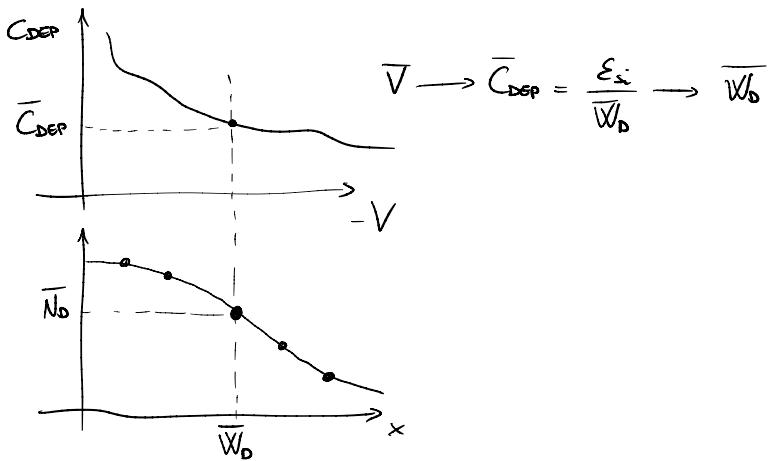
$$\frac{dC_{DEP}}{dV} = \frac{\varepsilon_s}{\sqrt{\frac{2\varepsilon_s}{qN_d}}} \left(-\frac{1}{2} \right) \frac{-1}{(\varphi_{BI} - V)^{3/2}} = \frac{\varepsilon_s}{\sqrt{\frac{2\varepsilon_s}{qN_d} (\varphi_{BI} - V)}} \frac{1}{2(\varphi_{BI} - V)} \frac{\frac{2\varepsilon_s}{qN_d}}{\frac{2\varepsilon_s}{qN_d}} = \frac{\varepsilon_s^2}{W_D^3 qN_d} \frac{\varepsilon_s}{\varepsilon_s} = \frac{C_{DEP}^3}{qN_d \varepsilon_s}$$

$$\therefore N_d = \frac{C_{DEP}^3}{q\varepsilon_s \frac{dC_{DEP}}{dV}}$$

OR

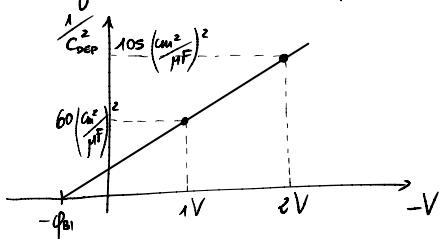
$$\frac{dC_{DEP}}{dV} = -\frac{\varepsilon_s}{W_D^2} \frac{dW_b}{dV} = -\frac{\varepsilon_s}{W_D^2} \left(-\frac{C_{DEP}}{qN_d(W_b)} \right) \frac{\varepsilon_s}{\varepsilon_s} = \frac{C_{DEP}^3}{q\varepsilon_s N_d(W_b)} \Rightarrow N_d(W_b) = \frac{C_{DEP}^3}{q\varepsilon_s \frac{dC_{DEP}}{dV}}$$

$$C_{DEP} = -\frac{d}{dV} \int_0^{W_b} qN_d(x) dx = -q \frac{d}{dV} \left[\tilde{N}_d(W_b) - \tilde{N}_d(0) \right] = -q \tilde{N}_d(W_b) \frac{dW_b}{dV} \Rightarrow \frac{dW_b}{dV} = -\frac{C_{DEP}}{q \tilde{N}_d(W_b)}$$



Tutorial #5

M-S junction with n-doped Si



qφ_{BN}?

$$C_{Dop} = \frac{\epsilon_{Si}}{W_b} = \frac{\epsilon_{Si}}{\sqrt{\frac{2\epsilon_{Si}}{9} \frac{1}{N_d} (q\phi_{Bi} - V)}} \Rightarrow \frac{1}{C_{Dop}^2} = \frac{1}{\epsilon_{Si}} \sqrt{\frac{2\epsilon_{Si}}{9} \frac{1}{N_d} (q\phi_{Bi} - V)}$$

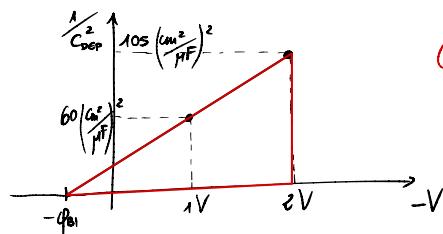
$$W_b = \sqrt{\frac{2\epsilon_{Si}}{9} \frac{1}{N_d} (q\phi_{Bi} - V)}$$

$$\frac{1}{C_{Dop}^2} = \frac{2}{q N_d \epsilon_{Si}} (q\phi_{Bi} - V)$$

We have a linear relation

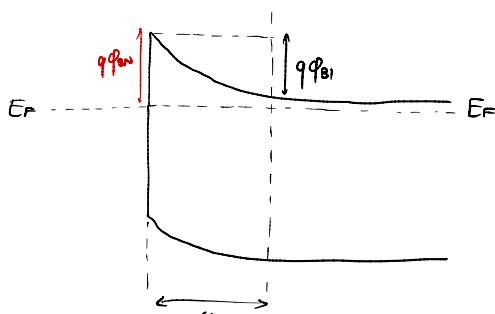
$$\text{Slope: } \frac{2}{q N_d \epsilon_{Si}} = \frac{(150-60) \left[\frac{\text{cm}^2}{\mu\text{F}} \right]^2}{(2-1) [\text{V}]} \Rightarrow N_d = 2,7 \cdot 10^{17} \text{ cm}^{-3}$$

When $\frac{1}{C_{Dop}^2} = 0$, $q\phi_{Bi} = -V$



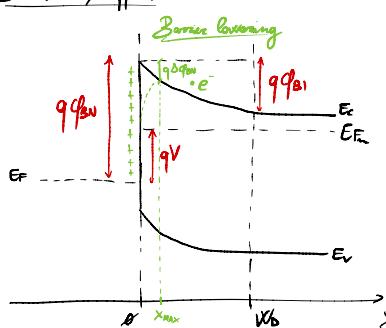
Considering the triangle we have that

$$\frac{\text{HEIGHT}}{\text{BASE}} = \frac{105 \left[\frac{\text{cm}^2}{\mu\text{F}} \right] \cdot 1 [\text{V}]}{(150-60) \left[\frac{\text{cm}^2}{\mu\text{F}} \right]^2} \Rightarrow q\phi_{Bi} = 0,33 V$$



$$q\phi_{BN} = q\phi_{Bi} + kT \log \frac{N_c}{N_d} = 458 \text{ meV}$$

Schottky effect



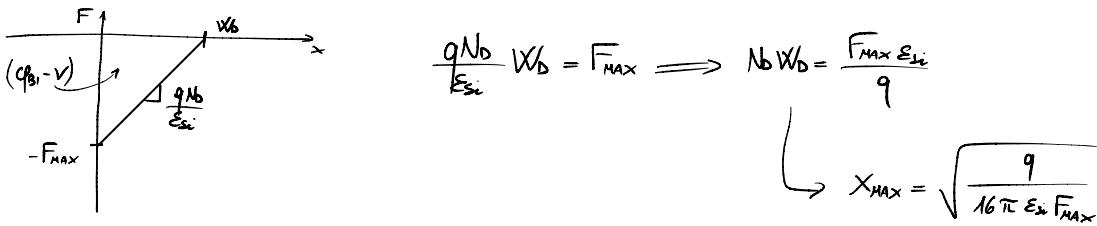
$$E_c(x) = E_c(W_b) + \frac{q^2 N_b}{2 \epsilon_{Si}} (W_b - x)^2 - \frac{1}{16 \pi \epsilon_{Si}} \frac{q^2}{x}$$

$$\frac{dE_c}{dx} = - \frac{q^2 N_b}{2 \epsilon_{Si}} (W_b - x) + \frac{q^2}{16 \pi \epsilon_{Si}} \frac{1}{x^2} = 0 \Rightarrow -N_b (W_b - x_{max}) + \frac{1}{16 \pi x_{max}^2} = 0$$

$$-N_b W_b \left(1 - \frac{x_{max}}{W_b} \right) + \frac{1}{16 \pi x_{max}^2} = 0$$

Schottky effect
happens very close
to the interface

$$x_{max} = \sqrt{\frac{1}{16 \pi N_b W_b}}$$



$$E_c(x_{max}) = E_c(W_b) + \frac{q^2 N_b}{2 \epsilon_{si}} (W_b - x_{max})^2 - \frac{q^2}{16\pi \epsilon_{si}} \frac{1}{x_{max}} = E_c(W_b) + \frac{q^2 N_b}{2 \epsilon_{si}} W_b^2 \left(1 - \frac{x_{max}}{W_b}\right)^2 - \frac{q^2}{16\pi \epsilon_{si}} \frac{1}{x_{max}} =$$

$$= E_c(W_b) + \frac{q^2 N_b}{2 \epsilon_{si}} W_b^2 \left(1 - \frac{2x_{max}}{W_b}\right)^2 - \frac{q^2}{16\pi \epsilon_{si}} \frac{1}{x_{max}} = E_c(W_b) + \frac{q^2 N_b}{2 \epsilon_{si}} W_b^2 - \frac{q^2 N_b W_b x_{max}}{2 \epsilon_{si}} - \frac{q^2}{16\pi \epsilon_{si}} \frac{1}{x_{max}}$$

E_c(0) NO SCHOTTKY

qDelta phi_BN

$$qDelta phi_BN = \frac{q^2 N_b W_b}{2 \epsilon_{si}} 2x_{max} + \frac{q^2}{16\pi \epsilon_{si}} \frac{1}{x_{max}} = \frac{q^2}{\epsilon_{si}} x_{max} \frac{1}{16\pi x_{max}} + \frac{q^2}{16\pi \epsilon_{si}} \frac{1}{x_{max}} = \frac{2q^2}{16\pi \epsilon_{si}} \frac{1}{x_{max}} =$$

$$= \frac{2q^2}{16\pi \epsilon_{si}} \frac{\sqrt{16\pi \epsilon_{si} F_{max}}}{q} = \sqrt{\frac{q^3 F_{max}}{4\pi \epsilon_{si}}}$$

$$N_d = 10^{16} \text{ cm}^{-3}$$

$$\phi_{BI} = 0.4 \text{ V}$$

$$\sqrt{\phi} = \phi$$

$$W_b = 230 \text{ nm}$$

$$x_{max} = 3 \text{ nm}$$

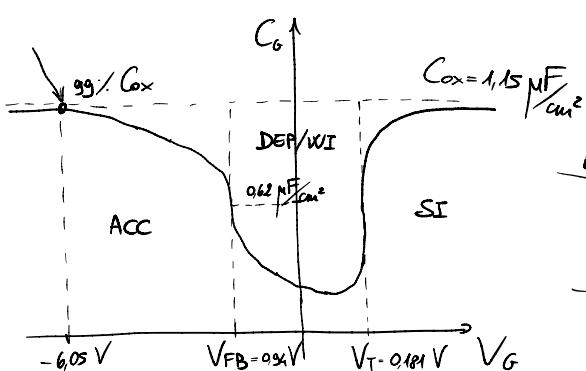
Tutorial #6

MOS capacitor

Si p-doped as substrate, $N_A = 3 \cdot 10^{17} \text{ cm}^{-3}$

Al gate, $q\phi_m = 4.1 \text{ eV}$

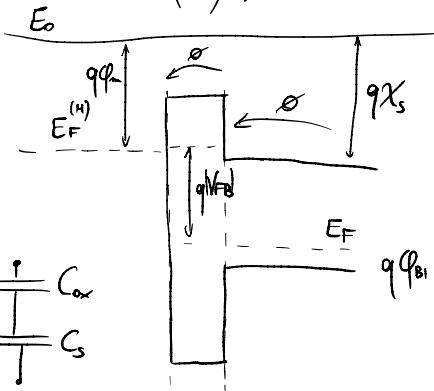
SiO_2 insulator, $t_{ox} = 3 \text{ nm}$



Quasi-static C-V curve?

$V_{FB}, C_G(V_{FB})$?

V_G in accumulation $\rightarrow C_G = 99\% C_{ox}$
 $V_T, C_G(V_T)$?



$$V_S + V_{ox} = \phi_{BI} + V$$

$$V_S = V_{ox} = \phi$$

$$q\phi_{BI} = \left[q\chi_s + \frac{E_g}{2} + kT \ln(\frac{N_A}{n_i}) \right] - q\phi_m = 0.92 \text{ eV}$$

$$V_{FB} = -\phi_{BI} = -0.92 \text{ V}$$

$$\left. \begin{array}{l} C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} = 1,15 \mu F/cm^2 \\ C_s = \frac{\epsilon_{si}}{L_D} = 1,39 \mu F/cm^2 \\ L_D = \sqrt{\frac{\epsilon_{si} kT}{q^2 N_A}} = 7,4 \text{ mm} \end{array} \right\} C_G = \left(\frac{1}{C_s} + \frac{1}{C_{ox}} \right)^{-1} = 0,62 \mu F/cm^2$$

ACCUMULATION : $Q_S \approx \sqrt{2 \epsilon_{si} kT N_A} e^{-\frac{qV_S}{2kT}}$

$$\begin{aligned} C_s &= -\frac{dQ_S}{dV_S} = -\sqrt{2 \epsilon_{si} kT N_A} e^{-\frac{qV_S}{2kT}} \left(-\frac{q}{2kT} \right) = \frac{Q_S}{2kT} = \frac{C_{ox}(V_{FB} - V_G)}{2kT} \\ C_G &= \left(\frac{1}{C_s} + \frac{1}{C_{ox}} \right)^{-1} = 0,99 C_{ox} \implies \frac{C_s}{C_s + C_{ox}} = 0,99 \implies C_s = \frac{0,99 C_{ox}}{0,01} = 99 C_{ox} \end{aligned}$$

$$C_{ox}(V_{FB} - V_G) = 99 C_{ox} \frac{2kT}{q} \implies V_G = V_{FB} - 99 \frac{2kT}{q} = -6,05 \text{ V}$$

$$V_T = V_{FB} + 2|\varphi_B| + \frac{\sqrt{2 \epsilon_{si} q N_A 2 |\varphi_B|}}{C_{ox}} = 121 \text{ mV}$$

$$\varphi_B = -\frac{kT}{q} \log \frac{N_A}{N_i} = -434 \text{ mV}$$

$$2|\varphi_B| = 868 \text{ mV}$$

$$W_{D_{MAX}} = \sqrt{\frac{2 \epsilon_{si}}{q N_A} 2 |\varphi_B|} = 61 \text{ mm}$$

$$Q_{D_{MAX}} = q N_A W_{D_{MAX}} = 292 \text{ nC/cm}^2$$

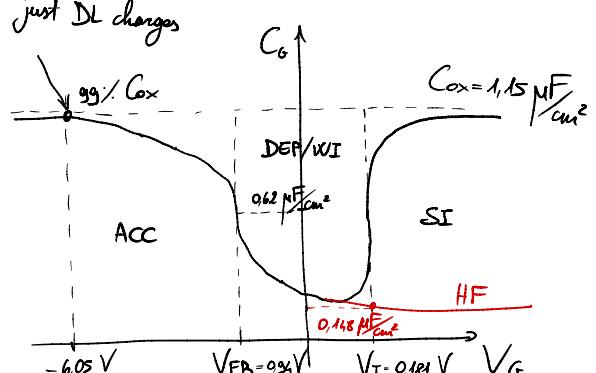
AT THRESHOLD

$$Q_S = -\sqrt{2 \epsilon_{si} kT N_A} \left[\frac{qV_S}{kT} + \frac{N_i^2}{N_A^2} e^{\frac{qV_S}{kT}} \right]^{\frac{1}{2}}$$

DL IL

$$\begin{aligned} C_s &= -\frac{dQ_S}{dV_S} = \sqrt{2 \epsilon_{si} kT N_A} \frac{1}{2} \left[\frac{q}{kT} + \frac{N_i^2}{N_A^2} e^{\frac{qV_S}{kT}} \right] = \\ &\stackrel{V_S = 2|\varphi_B|}{=} \sqrt{2 \epsilon_{si} kT N_A} \frac{1}{2} \frac{q}{kT} \left[\frac{1}{1 + \frac{N_i^2}{N_A^2} e^{\frac{q2|\varphi_B|}{kT}}} \right]^{\frac{1}{2}} \approx \end{aligned}$$

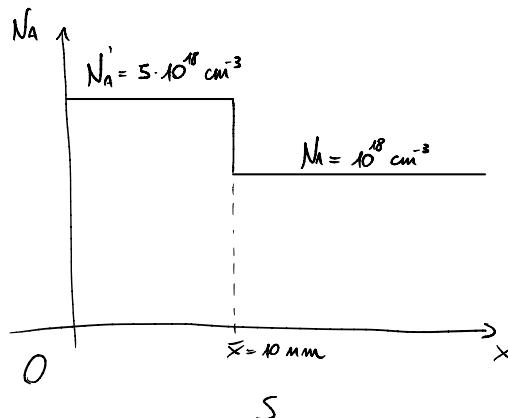
$$\approx \sqrt{2 \epsilon_{si} kT N_A} \frac{1}{2} \frac{q}{kT} \frac{1}{\sqrt{\frac{2q|\varphi_B|}{kT}}} = \sqrt{2 \epsilon_{si} kT N_A} \frac{q}{kT} \frac{\sqrt{kT}}{\sqrt{q2|\varphi_B|}} \frac{\sqrt{\frac{2\epsilon_{si}}{qN_A}}}{\sqrt{\frac{2\epsilon_{si}}{qN_A}}} = \frac{2\epsilon_{si}}{W_{D_{MAX}}} = 2C_{D_{EP}}$$



Inversion and depletion charges are modulated equally

$$C_o = 0,26 \mu F/cm^2$$

Tutorial #7

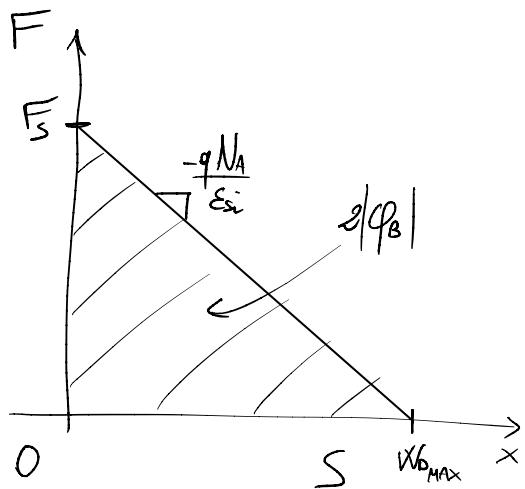


$$V_T = V_{FB} + 2|\phi_B| + \frac{\sqrt{2\epsilon_{Si}qN_Ae|\phi_B|}}{C_{ox}} =$$

$$= V_{FB} + 2|\phi_B| + \frac{qN_AW_{D_{MAX}}}{C_{ox}} =$$

$$= V_{FB} + 2|\phi_B| + \frac{F_{ox}t_{ox}}{C_{ox}} \quad \text{We can use both terms to relate } V_T$$

O: SiO_2 with $t_{ox} = 3 \text{ nm}$
 M: metal with $q\phi_m = 4.145 \text{ eV}$
 V_T ?
 $V_T(N_{A_{uniform}} = 10^{18} \text{ cm}^{-3})$?



$$2|\phi_B| = 2 \frac{K_T}{q} \log \frac{N_A}{n_i} = 931 \text{ mV}$$

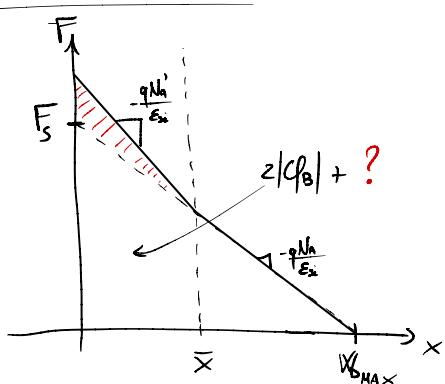
$$W_{D_{MAX}} = \sqrt{\frac{2\epsilon_{Si}}{q} \frac{1}{N_A} 2|\phi_B|} = 34 \text{ nm}$$

$$F_S = \frac{qN_A}{\epsilon_{Si}} W_{D_{MAX}} = 0.53 \text{ MV/cm}$$

$$F_{ox} = \frac{\epsilon_{Si}}{\epsilon_{ox}} F_S = 1.59 \text{ MV/cm}$$

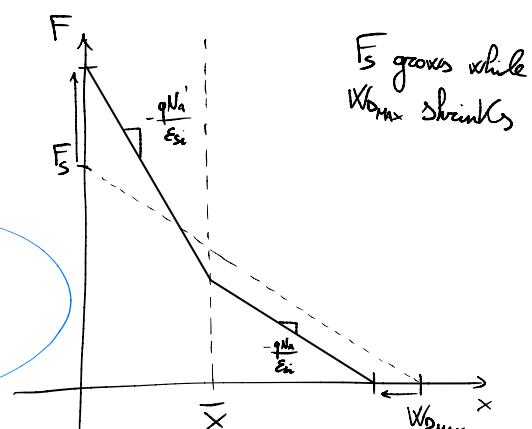
$$V_T = 0.47 \text{ V}$$

NON-UNIFORM NA



Plot is not correct

Total area is preserved



Threshold voltage increases as well

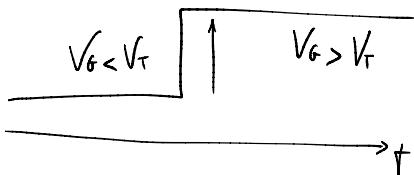
$$2|Q_B| = \frac{qN_A}{2\varepsilon_{Si}} (W_D' - \bar{x})^2 + \frac{qN_A}{\varepsilon_{Si}} (W_D' - \bar{x})\bar{x} + \frac{qN_A}{2\varepsilon_{Si}} \bar{x}^2 = \frac{qN_A}{2\varepsilon_{Si}} \left[W_D'^2 - \bar{x}^2 - 2W_D'\bar{x} + 2W_D'\bar{x} - 2\bar{x}^2 \right] + \frac{qN_A}{2\varepsilon_{Si}} \bar{x}^2 = \\ = \frac{qN_A}{2\varepsilon_{Si}} \left[W_D'^2 - \bar{x}^2 \right] + \frac{qN_A}{2\varepsilon_{Si}} \bar{x}^2$$

$$F_S = \frac{qN_A}{\varepsilon_{Si}} (W_D' - \bar{x}) + \frac{qN_A'}{\varepsilon_{Si}} \bar{x} = 1,04 \text{ MV/cm}$$

$$\underbrace{\frac{2\varepsilon_{Si}}{qN_A} 2|Q_B| = W_D'^2 - \bar{x}^2 + \frac{N_h'}{N_A} \bar{x}^2}_{\rightarrow} \Rightarrow W_D' = \sqrt{W_{D_{MAX}}^2 - \frac{N_h' - N_A}{N_A} \bar{x}^2} = 27,5 \text{ nm}$$

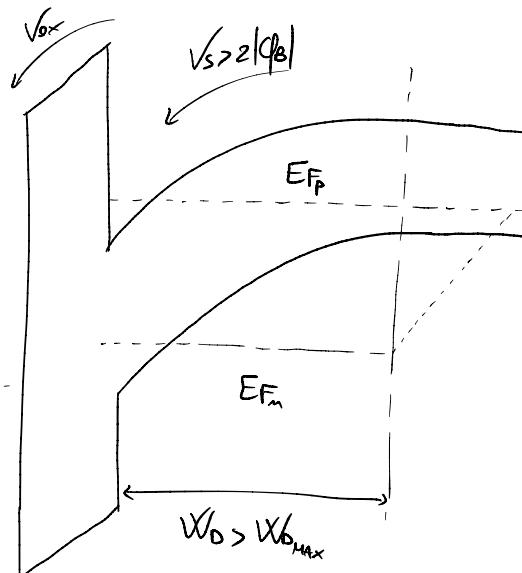
$$V_T = V_{FB} + 2|Q_B| + F_{ox}t_{ox} = 0,94 \text{ V}$$

V_G switch



$$\frac{dQ_{INV}}{dt} = -q \left(\frac{M_i}{2\tau_0} \right) [W_D(t) - W_{D_{MAX}}]$$

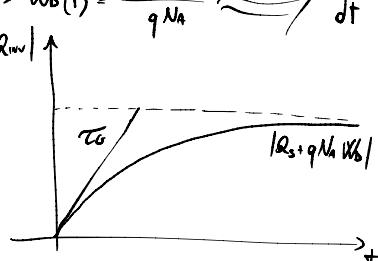
generation rate



$$Q_S = Q_{INV} + Q_{DEP} = Q_{INV} - qN_A W_D(t) \rightarrow W_D(t) = \frac{Q_{INV} - Q_S}{qN_A} \Rightarrow \frac{dQ_{INV}}{dt} = -\frac{M_i}{2\tau_0 N_A} Q_{INV} + \frac{M_i}{2\tau_0 N_A} [Q_S + qN_A W_{D_{MAX}}]$$

$$Q_{INV}(t) = [Q_S + qN_A W_{D_{MAX}}] (1 - e^{-\frac{t}{\tau_0}})$$

$$\tau_0 = 2\tau_0 \frac{N_A}{M_i}$$



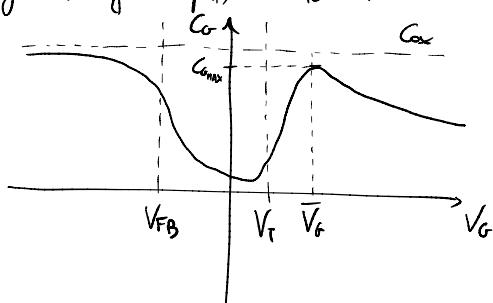
Tutorial #8

MOS capacitor

Si, unif. p-doped, $N_A = 3 \cdot 10^{11} \text{ cm}^{-3}$

Si_3N_4 , $t_{ox} = 3 \text{ nm}$

Poly-Si, unif. n-doped, $N_D = 10^{20} \text{ cm}^{-3}$



Quasi-static C-V curve?

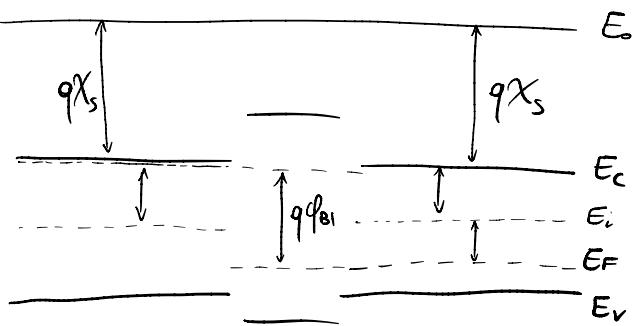
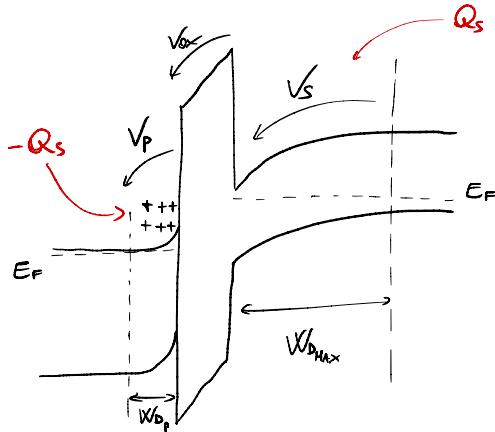
V_G under strong inversion for $C_{G_{MAX}}$?

$$Q_S = -\sqrt{2\varepsilon_{Si} kT N_A} \frac{M_i}{N_A} e^{\frac{q|Q_B|}{kT}}$$

$$C_{ox} = \frac{E_{ox}}{t_{ox}} = 1,15 \mu\text{F/cm}^2$$

$$C_s = \frac{dQ_s}{dV_s} = \frac{C_{ox}(V_g - V_{FB} - 2|Q_B|)}{2KT/q}$$

$$C_p = \frac{dQ_s}{dV_p}$$



$$E_c - E_i = \frac{E_{GAP}}{2}$$

$$E_i - E_F = KT \log \left(\frac{N_A}{m_i} \right)$$

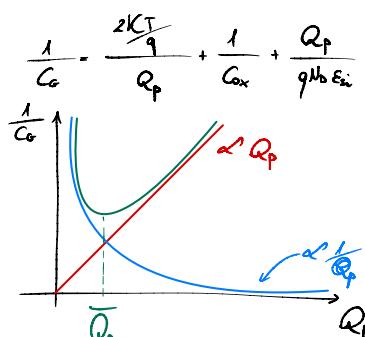
$$qV_{FB} = - \left(\frac{E_{GAP}}{2} + KT \log \left(\frac{N_A}{m_i} \right) \right) = -0.99 \text{ eV}$$

$$C_G \equiv \frac{1}{C_G} \Leftrightarrow \frac{1}{C_G} = C_p + C_{ox}$$

$$\frac{d}{dQ_p} \left(\frac{1}{C_G} \right) = -\frac{2KT}{q} \frac{1}{Q_p^2} + \frac{1}{qN_b \epsilon_{si}} = 0$$

$$C_s = \frac{|Q_B|}{2KT/q} = \frac{Q_p}{2KT/q}$$

$$C_p = \frac{\epsilon_{si}}{X_{Dmax}} \frac{qN_b}{qN_b} = \frac{qN_b \epsilon_{si}}{Q_p}$$



$$\bar{Q}_p = \sqrt{2\epsilon_{si}KT N_b} = \sqrt{2\epsilon_{si}qN_bV_p}$$

$$\frac{1}{C_{G_{MAX}}} = \frac{2KT}{q\sqrt{2\epsilon_{si}KT N_b}} + \frac{1}{C_{ox}} + \frac{\sqrt{2\epsilon_{si}KT N_b}}{\epsilon_{si}qN_b} = \sqrt{\frac{2KT}{q^2\epsilon_{si}N_b}} + \frac{1}{C_{ox}} + \sqrt{\frac{2KT}{q^2\epsilon_{si}N_b}} = \sqrt{\frac{8KT}{q^2\epsilon_{si}N_b}} + \frac{1}{C_{ox}}$$

$C_s = C_p$ at the maximum

$$C_{G_{MAX}} = 1.02 \mu F/cm^2$$

$$2|Q_B| \cdot \frac{eKT}{q} \log \left(\frac{N_A}{m_i} \right) = 868 \text{ nV}$$

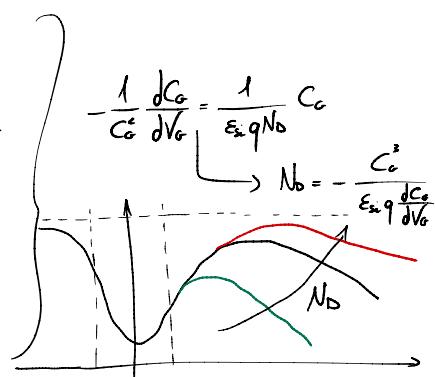
$$V_{FB} = -Q_B = -0.99 \text{ V}$$

Let's focus on the decreasing part of the C-V curve

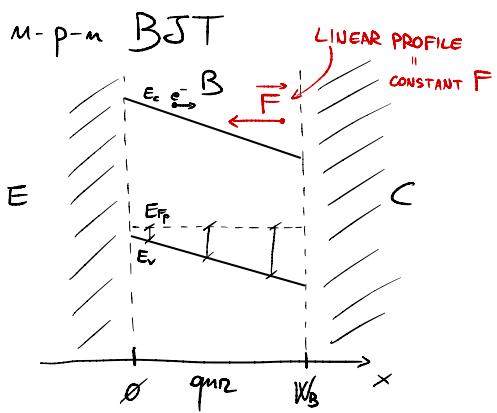
$$\frac{1}{C_G} \approx \frac{1}{C_{ox}} + \frac{1}{C_p}$$

$$\frac{d}{dV_g} \left(\frac{1}{C_G} \right) = -\frac{1}{C_G^2} \frac{dC_G}{dV_g} = \frac{d}{dV_g} \left(\frac{1}{C_p} \right) = \frac{d}{dV_g} \left(\frac{Q_p}{\epsilon_{si}qN_b} \right) = \frac{1}{\epsilon_{si}qN_b} C_G$$

$$C_p = \frac{\epsilon_{si}}{X_{Dmax}} \frac{qN_b}{qN_b} = \frac{\epsilon_{si}qN_b}{Q_p}$$



Tutorial #9



$$N_A^B(x) = N_A^B(0) e^{-\frac{x}{\lambda}}$$

$$N_A^B(0) = 2 \cdot 10^{17} \text{ cm}^{-3}$$

$$\lambda = 200 \text{ nm}$$

$$W_B = 500 \text{ nm}$$

$$V_{BE} > 0, V_{CB} > \frac{kT}{q}$$

low level of injection

$F(x)$? $n(x)$? J_c ?

$$p(x) \approx N_A^B(x)$$

$$F \approx \frac{kT}{q} \frac{1}{p} \frac{dp}{dx} = \frac{kT}{q} \frac{1}{p} \frac{d}{dx} \left[N_A^B(0) e^{-\frac{x}{\lambda}} \right] = \frac{kT}{q} \frac{1}{N_A^B(0) e^{-\frac{x}{\lambda}}} N_A^B(0) e^{-\frac{x}{\lambda}} \left(-\frac{1}{\lambda} \right) = -\frac{kT}{q} \frac{1}{\lambda} = -1.3 \text{ KV/cm}$$

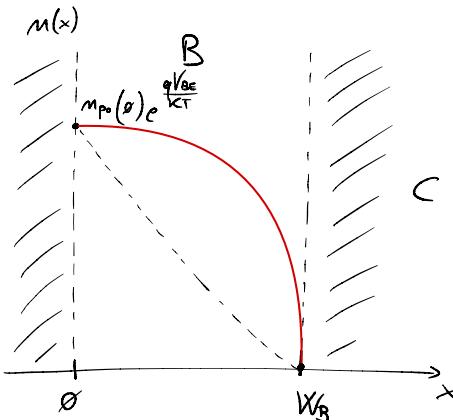
$$J_n = q_m \mu_m F + q D_m \frac{dn}{dx} = q_m \mu_m \left(-\frac{kT}{q} \frac{1}{\lambda} \right) + q D_m \frac{dn}{dx} = -q_m D_m \frac{1}{\lambda} + q D_m \frac{dn}{dx}$$

$$\frac{J_n}{q D_m} = -\frac{1}{\lambda} + \frac{dn}{dx} \implies \frac{dn}{dx} = \frac{1}{\lambda} + \frac{J_n}{q D_m}$$

$$n(x) = \underbrace{A e^{\frac{x}{\lambda}}}_{\text{homogeneous solution}} - \frac{J_n \lambda}{q D_m}$$

$$n(0) = n_p(0) e^{\frac{qV_{BE}}{kT}} = A - \frac{J_n \lambda}{q D_m}$$

$$n(W_B) = 0 = A e^{\frac{W_B}{\lambda}} - \frac{J_n \lambda}{q D_m}$$



$$\begin{cases} A = \frac{J_n \lambda}{q D_m} e^{-\frac{W_B}{\lambda}} \\ n(0) = \frac{J_n \lambda}{q D_m} e^{-\frac{W_B}{\lambda}} - \frac{J_n \lambda}{q D_m} \end{cases}$$

$$\begin{aligned} J_n &= \frac{n(0)}{\frac{\lambda}{q D_m} (e^{-\frac{W_B}{\lambda}} - 1)} \sim J_n = -J_n = \\ &= \frac{n_p(0) e^{\frac{qV_{BE}}{kT}}}{\frac{\lambda}{q D_m} (1 - e^{-\frac{W_B}{\lambda}})} = \\ &= \frac{q m_i^2 e^{\frac{qV_{BE}}{kT}}}{N_A^B(0) \frac{\lambda}{D_m} (1 - e^{-\frac{W_B}{\lambda}})} \end{aligned}$$

$$G_B = \int_0^{W_B} \frac{P}{D_m} dx = \int_0^{W_B} \frac{N_A^B(0) e^{-\frac{x}{\lambda}}}{D_m} dx = \frac{N_A^B(0)}{D_m} e^{-\frac{x}{\lambda}} \Big|_0^{W_B} = \frac{N_A^B(0)}{D_m} \lambda \left(1 - e^{-\frac{W_B}{\lambda}} \right)$$

$$n(x) = \frac{J_n \lambda}{q D_m} e^{-\frac{W_B}{\lambda}} e^{\frac{x}{\lambda}} - \frac{J_n \lambda}{q D_m} = \frac{J_n \lambda}{q D_m} \left[1 - e^{-\frac{W_B-x}{\lambda}} \right]$$

$$\frac{dn}{dx} = \frac{J_n \lambda}{q D_m} e^{-\frac{W_B}{\lambda}} e^{\frac{x}{\lambda}} \frac{1}{\lambda}$$

m-p-n BST

$$N_A^B = 10^{18} \text{ cm}^{-3} \quad W_B = 100 \text{ nm}, D_m = 8,3 \frac{\text{cm}^2}{\text{s}}$$

$$N_D^E = 10^{20} \text{ cm}^{-3} \quad W_E = 100 \text{ nm, shallow}, D_p = 3,1 \frac{\text{cm}^2}{\text{s}}$$

low level of injection

$$V_{BE} > 0 \quad V_{CB} > \frac{kT}{q}$$

$$\beta = \frac{I_c}{I_B} = \frac{G_E}{G_B}$$

β with and without band-gap narrowing effect?

No effect

$$\left. \begin{aligned} G_E &= \frac{N_D^E W_E}{D_p} \\ G_B &= \frac{N_A^B W_B}{D_m} \end{aligned} \right\} \beta = \frac{N_D^E W_E}{D_p} \frac{D_m}{N_A^B W_B} = 535$$

BGN effect

$$\left. \begin{aligned} G_E &= \int_{-W_E}^0 \frac{M_i^2}{(M_{ie})^E} \frac{P}{D_p} dx = \frac{M_i^2}{(M_{ie})^E} \frac{N_D^E}{D_p} W_E \\ G_B &= \int_0^{W_B} \frac{M_i^2}{(M_{ie})^B} \frac{P}{D_m} dx = \frac{M_i^2}{(M_{ie})^B} \frac{N_A^B}{D_m} W_B \end{aligned} \right\} \overline{\beta} = \frac{M_i^2}{(M_{ie})^E} \frac{N_D^E W_E}{D_p} \frac{(M_{ie})^B}{M_i^2} \frac{N_A^B W_B}{D_m} = \frac{(M_{ie})^B}{(M_{ie})^E} \beta = 77$$

$$M_{ie}^2 = M_i^2 e^{\frac{\Delta E_G}{kT}}$$

$$M\text{-doped Si: } (\Delta E_G)^E \approx 18,7 \log \left[\frac{N_D}{7 \cdot 10^{17} \text{ cm}^{-3}} \right] = 92,8 \text{ meV}$$

$$P\text{-doped Si: } (\Delta E_G)^B \approx q \left[\log \left(\frac{N_A}{10^{17} \text{ cm}^{-3}} \right) + \sqrt{\left[\log \left(\frac{N_A}{10^{17} \text{ cm}^{-3}} \right) \right]^2 + 0,5} \right] = 42 \text{ meV}$$

Part III

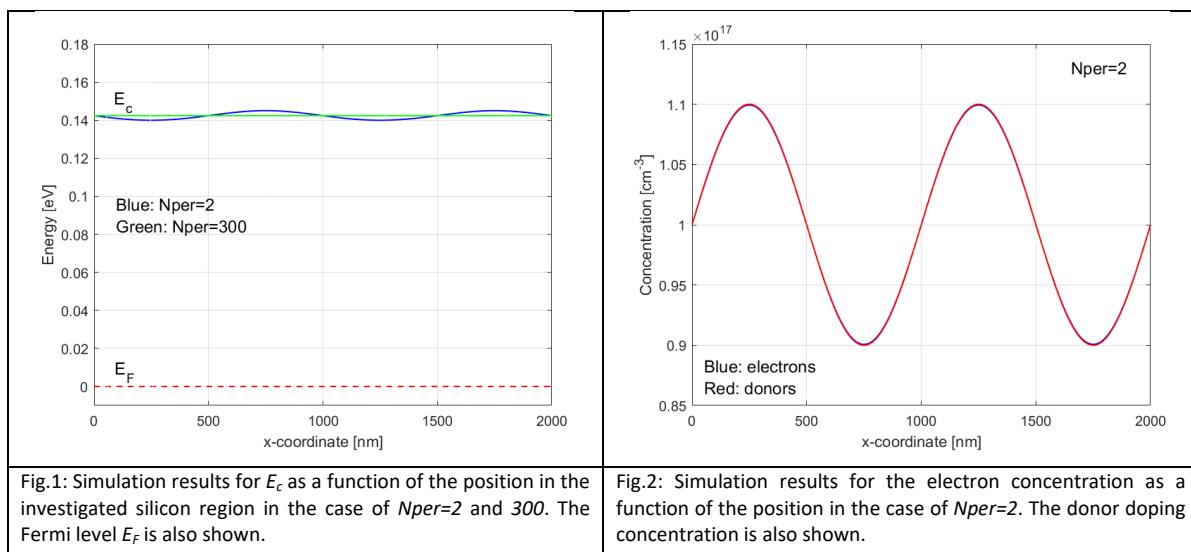
Simulation labs

Lab 1

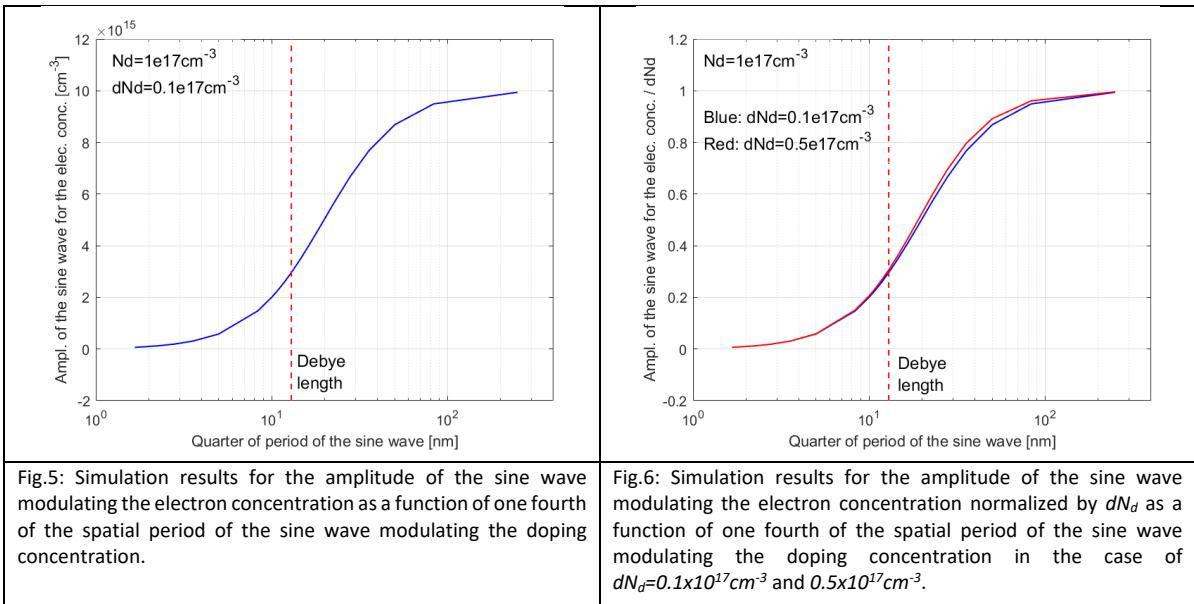
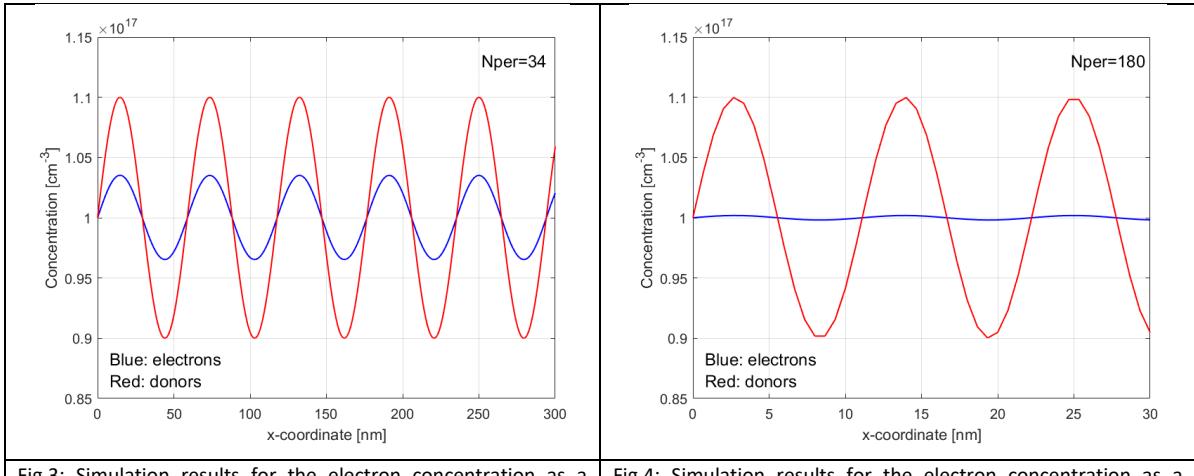
Step 5: Analysis of the simulation results in Matlab

The file with the simulation results generated by Comsol Multiphysics can now be easily loaded into Matlab for a detailed analysis of the dependence of the electrostatics of the investigated silicon region on the periodicity of the sine wave modulating the doping concentration. Students are kindly invited to try to perform this analysis on their own, reproducing the figures reported in this section (anyway, who is still not familiar with Matlab may find the sequence of commands used to generate the reported figures at the end of this section).

Fig.1 shows the profile of the conduction band edge E_c along the investigated silicon region as resulting from the numerical simulations performed in Comsol Multiphysics, in the case of a number of periods of the sine wave modulating the doping concentration $Nper=2$ (blue curve) and $Nper=300$ (green curve). In the former case, the period of the sine wave modulating the doping concentration equals $1\mu m$ and clearly appears from the periodicity of the E_c profile in the material. In this regard, note also that the vertical shift of E_c over the space in the case of $Nper=2$ reveals that the electron concentration follows (at least partially) the change of the donor doping concentration (remember that the sine wave for the doping concentration starts with its positive branch at $x=0$ in the project and that a downward shift of E_c may only result in the growth of the electron concentration in the presence of a constant Fermi level E_F). In the case of $Nper=300$, instead, the period of the sine wave modulating the doping concentration equals about $6.7nm$, but no evidence of that appears from the E_c profile in Fig.1. More in general, the E_c profile in this latter case does not provide any evidence at all of the change of the doping concentration over the space. In fact, E_c is almost flat all over the silicon region, meaning that the electron concentration is almost constant.

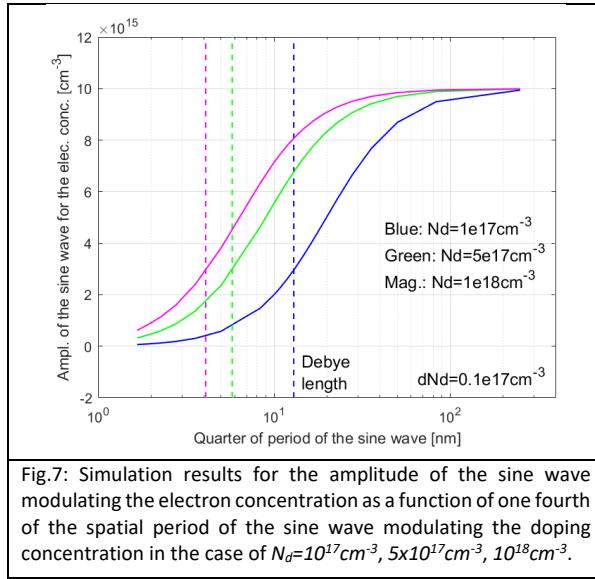


To go into more details with the analysis of the simulation results, Fig.2, Fig.3 and Fig.4 compare the electron concentration profile with the doping concentration profile in the case of $Nper=2, 34$ and 180 , respectively. In the first case (Fig.2), the electron concentration almost equals the doping concentration, following its change over the space. In the second case (Fig.3), instead, a sine wave modulation of the electron concentration in phase with the modulation of the doping concentration clearly appears, but the amplitude of the former is significantly less than that of the latter. In the third case (Fig.4), finally, the modulation of the electron concentration over the space is almost negligible (the same is true also for larger values of $Nper$).



Figs.2-4 revealed that a sine wave modulation of the donor doping concentration gives rise to an in-phase sine wave modulation of the electron concentration, with the amplitude of the latter depending on N_{per} . More specifically, the amplitude of the modulation of the electron concentration approaches the amplitude of the modulation of the doping concentration when N_{per} is low and decreases down to zero when N_{per} is high. For the sake of accuracy, it is now important to point out that the amplitude of the modulation of the electron concentration is actually dependent on the periodicity of the sine wave modulating the doping concentration, with the quantitative trend with N_{per} being just the consequence of that dependence in the presence of a constant length of the silicon region L_{tot} (students can easily verify this point by repeating the simulations in Comsol Multiphysics with different values of L_{tot}). The amplitude of the modulation of the electron concentration is then reported in Fig.5 as a function of the periodicity of the sine wave modulating the doping concentration, with the latter quantified in terms of quarter of period of the wave (the reason for taking the quarter of period of the wave and not the full period will be discussed later). The curve clearly shows that the transition of the amplitude of the modulation of the electron concentration from the value corresponding to the amplitude of the modulation of the doping concentration to zero occurs, roughly, over

the interval from 100nm to a few nm . Almost in the middle of this interval, there is the value of the Debye length corresponding to silicon with average donor doping concentration $N_d=10^{17}\text{cm}^{-3}$, which is about 13nm . This proves that, for the majority carrier concentration and the electrostatics in a semiconductor material to follow a change of the doping concentration, the latter change must occur over a distance much longer than the Debye length. If that is not the case, the majority carrier concentration significantly deviates from the doping concentration.



The reason why the quarter of period and not the full period of the sine wave modulating the doping concentration has been considered along the horizontal axis of Fig.5 can now be easily understood starting from the previous general discussion on the role of the Debye length. First of all, it is worth noting that, irrespective of N_{per} and the amplitude of the modulation of the doping concentration dN_d , the electron concentration equals the average doping concentration N_d at the points where the sine wave modulating the doping concentration is zero (see Figs.2-4). That is due to the fact that i) the sine wave gives rise to a symmetrical modulation of the doping concentration around N_d and ii) the doping concentration is not altered at those points with respect to the case of no modulation of the doping concentration (i.e., $dN_d=0$). Given that the electron concentration equals N_d at the points where the sine wave modulating the doping concentration is zero, the quarter of period of the sine wave represents the distance over which the electron concentration should change to follow the change of the doping concentration. That is, then, roughly the distance to be compared with the Debye length (a more accurate analysis, however, leads to a slightly different definition of the characteristic length to be compared with the Debye length, see *). In particular, if that distance is much longer than the Debye length, the electron concentration can follow the doping concentration over the space (Fig.2) and the amplitude of the modulation of the former concentration almost equals that of the latter (see what happens in Fig.5 above 100nm). If, instead, that distance is comparable or shorter than the Debye length, the electron concentration cannot follow the doping concentration over the space (Figs.3-4). In this latter case, the electron concentration changes more slowly, trying to catch the change of the doping concentration over a distance of a few Debye lengths. Due to the periodic nature of the modulation of the doping concentration over the space considered in the project, however, that slow change of the electron concentration turns into a reduction of the amplitude of the modulation of this latter concentration. This point can be easily understood by considering two adjacent quarter of periods of the sine wave making the doping concentration grow from N_d to N_d+dN_d and then decrease from N_d+dN_d to N_d . Over the first quarter of period, the electron concentration tends to grow from N_d to N_d+dN_d , but to achieve that change a distance of a few Debye lengths would be needed. As a consequence, the increase of the electron concentration at the end of the first quarter of period of the sine wave is smaller than dN_d . During the second

quarter of period, then, the electron concentration tends to decrease back to N_d , because that is what the doping concentration does. As a results, the electron concentration peaks at the same position of the doping concentration, but the amplitude of the peak of the former concentration is smaller than the amplitude of the peak of the latter. In the case the quarter of period of the sine wave of the doping concentration is much shorter than the Debye length, moreover, the peak is not significantly higher than N_d , meaning that the electron concentration remains nearly constant over the space (see what happens in Fig.5 for quarter of periods shorter than a few nm).

To further prove that the capability of the majority carrier concentration to follow the change of the doping concentration depends on the periodicity of the sine wave and not on other parameters, the simulations performed in Comsol Multiphysics have been repeated with $dN_d=0.5 \times 10^{17} \text{ cm}^{-3}$. Fig.6 shows that the increase of the amplitude of the modulation of the doping concentration does not affect at all the previous conclusions. This confirms that it is not the amplitude but the periodicity of the modulation of the doping concentration to be relevant for the results. Finally, Fig.7 shows what obtained by repeating the Comsol Multiphysics simulations with different values of N_d , equal to 10^{17} cm^{-3} , $5 \times 10^{17} \text{ cm}^{-3}$ and 10^{18} cm^{-3} ($dN_d=0.1 \times 10^{17} \text{ cm}^{-3}$). As clearly appearing from the figure, the curve representing the amplitude of the modulation of the electron concentration displays a leftward shift with the growth of N_d , as expected from the corresponding reduction of the Debye length. This represents a further proof of the previous physical picture for the change of the majority carrier concentration over the space.

* Interested students may try to demonstrate that the characteristic length of the sine wave modulation of the doping concentration to be compared with the Debye length is actually the period of the sine wave divided by 2π , i.e., $L_{tot}/N_{per}/2\pi$. Coming to that result is not difficult and requires just to set a sine wave modulation of the donor and electron concentrations in the Poisson equation, so $N_d = \tilde{N}_d + dN_d * \sin(2\pi * N_{per}/L_{tot} * x)$ and $n = \tilde{n}_d + dn * \sin(2\pi * N_{per}/L_{tot} * x)$ and to formulate the electrostatic potential ϕ as a function of n . From the resulting equation, assuming dN_d and dn to be small, the following expression for dn can be easily calculated:

$$dn = \frac{dN_d}{1 + \left(\frac{Ldn * N_{per} * 2\pi}{L_{tot}}\right)^2}$$

From the previous formula, the comparison between the Debye length L_{dn} and $L_{tot}/N_{per}/2\pi$ is clearly evident. Besides, the formula allows to reproduce the trends for the amplitude of the modulation of the electron concentration obtained from the simulations performed in Comsol Multiphysics and shown in Figs.5-7. By plotting dn as a function of $L_{tot}/N_{per}/2\pi$, the Debye length corresponds to the characteristics length at which $dn=dN_d/2$, as also evident from the previous formula for dn .

Lab 2

Step 5: Comparison between the numerical and the analytical results in Matlab

The file with the electrostatic results generated by Comsol Multiphysics can now be easily loaded into Matlab for a comparison with the analytical results obtained during lessons (see the notes on “*The p-n junction*”). Students are kindly invited to try to perform the comparison on their own, reproducing the figures reported in this section (anyway, who is still not familiar with Matlab may find the sequence of commands used to generate the reported figures at the end of the analysis).

Fig.1 shows the band diagram of the investigated *p-n* junction as resulting from the numerical simulations performed in Comsol Multiphysics (blue curves) and from the analytical formulas obtained during lessons (green curves). The edges $-x_p$ and $+x_n$ of the depletion layer obtained from the analytical calculations are also highlighted and the Fermi level E_F is shown as a horizontal red dashed line. As clearly appearing from the figure, the numerical and the analytical results are quite similar, with just a slightly smoother transition of the bands at the edges of the depletion layer in the former case. This confirms the validity of the approximations done in the theoretical analysis of device electrostatics. To be more quantitative in the comparison, Fig.2 shows the discrepancy between the numerical and the analytical results, i.e., the difference between the value of the conduction band edge E_c (or valence band edge E_v) resulting from the former and the latter, as a function of the position along the x-axis. This discrepancy can be considered to be the error on the band diagram (or the electrostatic potential) coming from the approximations involved in the analytical calculations performed during lessons. The figure reveals that the error is always less than the thermal energy kT , i.e., it is small, and peaks close to the edges of the depletion layer. The fact that the error reaches its highest values at the edges of the depletion layer can be explained by considering that there the depletion approximation assumes an abrupt transition of the majority carrier concentration from the doping value to zero. As will be pointed out in the next discussions, that is not possible, since the majority carrier concentration requires a few Debye lengths (L_D) to change.

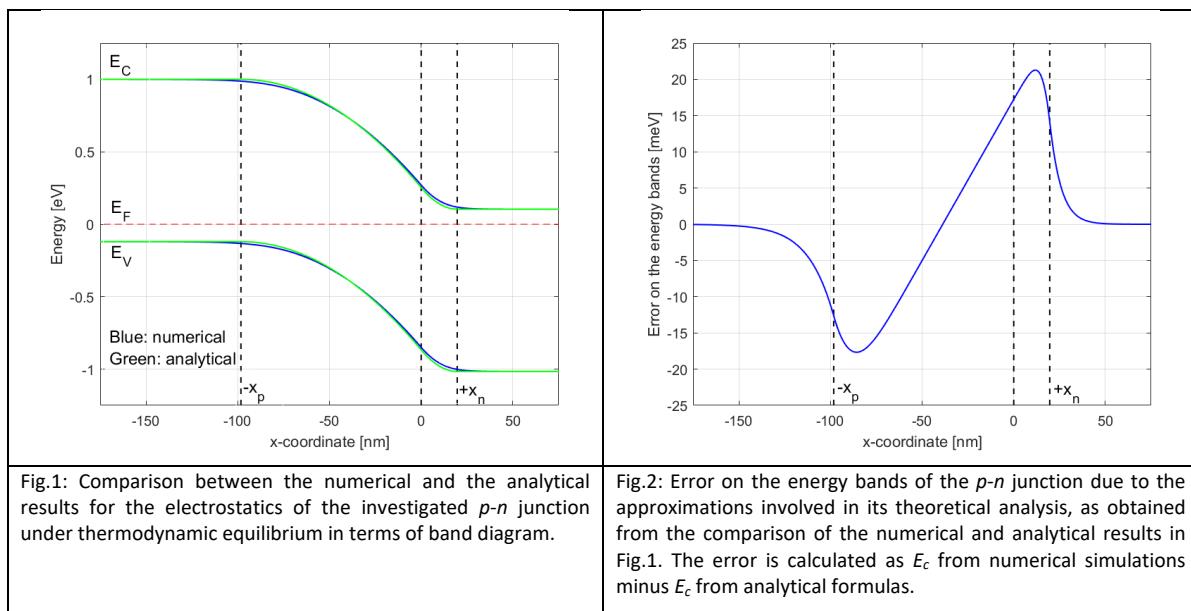


Fig.3 extends the comparison between the numerical and the analytical results to the electric field (F) profile in the device. Even with this metric, the discrepancy between the results is relatively small and consists mainly in a smoother transition of the F at the edges of the depletion layer in the numerical case. Fig.4 shows that well inside the depletion layer the relative error on F is less than 10%, confirming again the validity of the analytical investigations performed during lessons. Only close to the edges of the depletion layer the relative error steeply rises, because the analytical studies assume that F goes to zero at $+x_n$ and $-x_p$. That makes the

relative error reach 100% at those points. Anyway, the low value of F at the edges with respect to the central regions of the depletion layer makes the error not relevant for device operation.

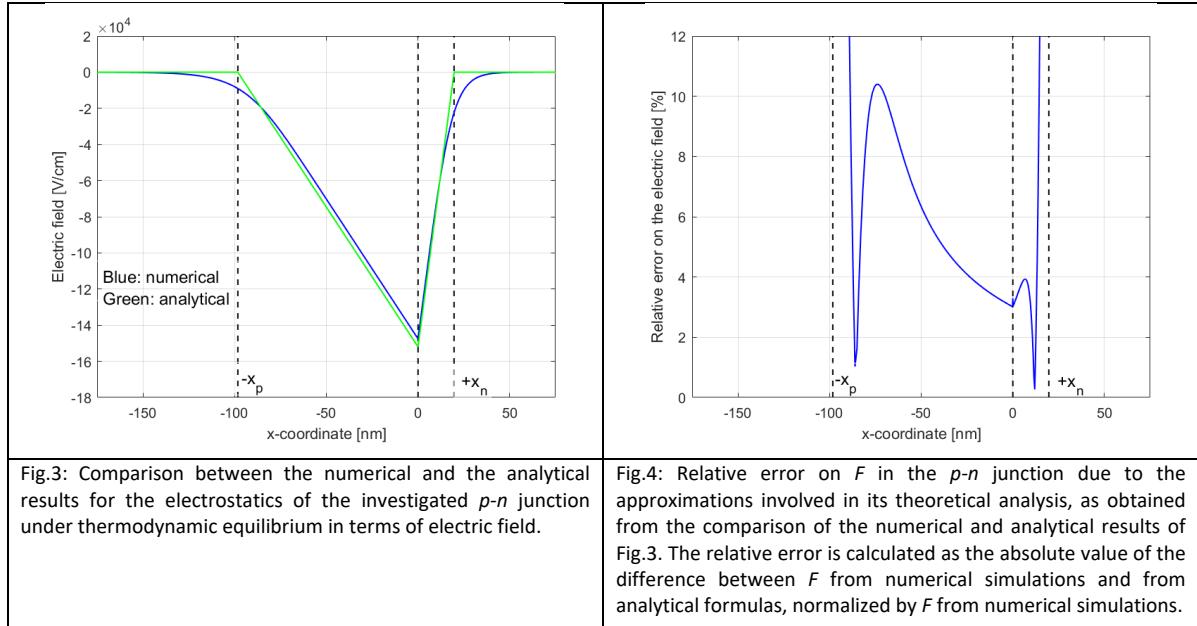
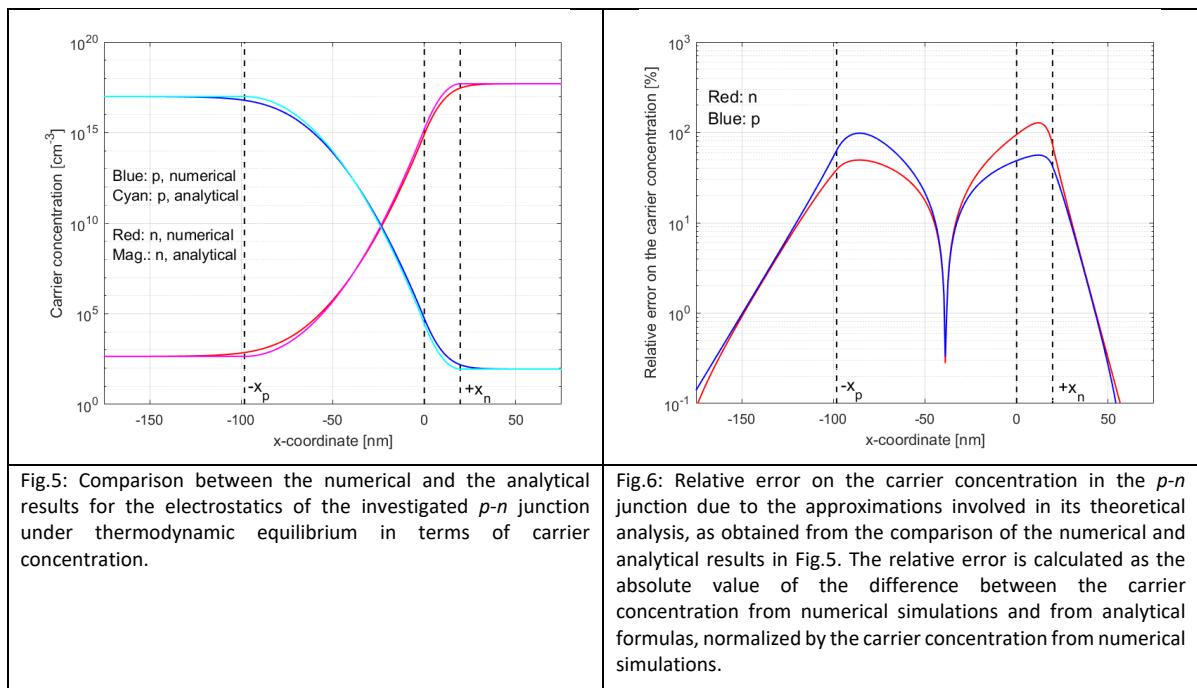
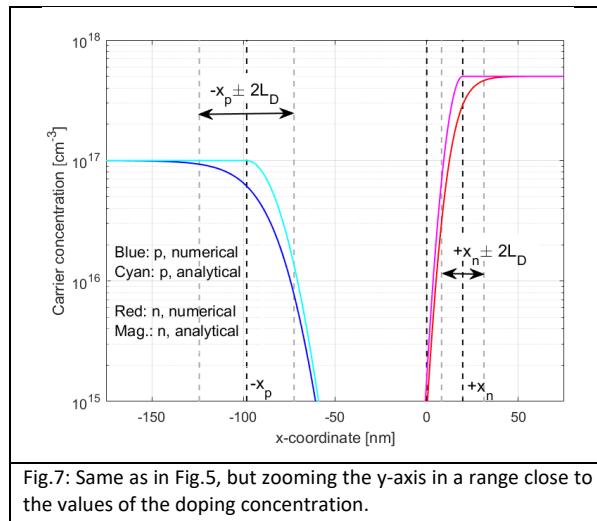


Fig.5 shows that even the profiles for the electron and hole concentrations predicted by the numerical and analytical results are in quite good agreement. This is a direct consequence of the agreement between the results in term of band diagram in Fig.1. Fig.6 shows that the relative error on the concentrations is the highest close to the edges of the depletion layer, where it can reach 100% (this corresponds to an analytical carrier concentration that is twice the carrier concentration coming from numerical simulations). Well inside the depletion layer and in the side regions of the device, instead, the error is significantly smaller.



In order to complete the comparison between the numerical and the analytical results for the electrostatics of the investigated *p-n* junction under thermodynamic equilibrium, it is worth noting that the carrier concentration profiles considered in Figs.5-6 for the analytical case are those coming from the band diagram in Fig.1 and not those corresponding to the depletion approximation. The latter, in fact, would predict a majority carrier concentration equal to the doping concentration in the side regions of the device that are under charge neutrality conditions and an electron and hole concentration equal to zero in the depletion layer. Such a behavior for the carrier concentrations misses that majority carriers require a few Debye lengths L_D to change, as highlighted in Fig.7. In this figure, the hole concentration at the edge of the depletion layer on the *p* side of the junction and the electron concentration at the edge of the depletion layer on the *n* side of the junction are shown to change from the value corresponding to the doping concentration to one tenth of that value in an interval equal to nearly $4L_D$ centered at $-x_p$ and $+x_n$, respectively. That means that the change of the majority carrier concentration at the edges of the depletion layer is smoother than what predicted by the depletion approximation and that is the reason why even the band diagram and the *F* profiles obtained from numerical simulations in Figs.1-3 are slightly smoother than the analytical predictions at those points. The accuracy of the analytical results is, anyway, rather good because the smoother majority carrier profile does not significantly modify the total charge Q_{dep} exposed in the depletion layer. This charge equals $1.57 \times 10^{-7} C/cm^2$ when calculated with the analytical formula and $1.54 \times 10^{-7} C/cm^2$ when calculated from the spatial integration on each side of the device of the majority carrier profile obtained from the numerical simulations. The error on this charge is then just equal to 2.4%. Such a low error reveals that the depletion approximation makes the profile of the charge exposed in the device more abrupt at the depletion layer edges, but the total amount of charge exposed is not affected by it and that is the reason why it allows to reproduce rather well the band diagram and *F* profile in the device.



To further challenge the validity of the analytical results for the electrostatics of a *p-n* junction, students are kindly invited to extend this analysis by going back to Comsol Multiphysics and addressing the case of different doping concentrations (to do that, change the value of the global parameters *Na* and *Nd* in the project). In so doing, it should become clear that the errors coming from the approximations involved in the analytical results are always rather low, unless a unilateral junction with a very large difference (a factor 500 or more) in the doping concentration of the *n* and *p* regions is considered. The reasons why in this latter case the approximation errors grow are left to the students' physical insight. Besides, the impact of the Maxwell-Boltzmann approximation on the results could also be investigated by changing the field **Carrier Statistics** under the tree **Model Properties** in the **Settings** window for Semiconductor.

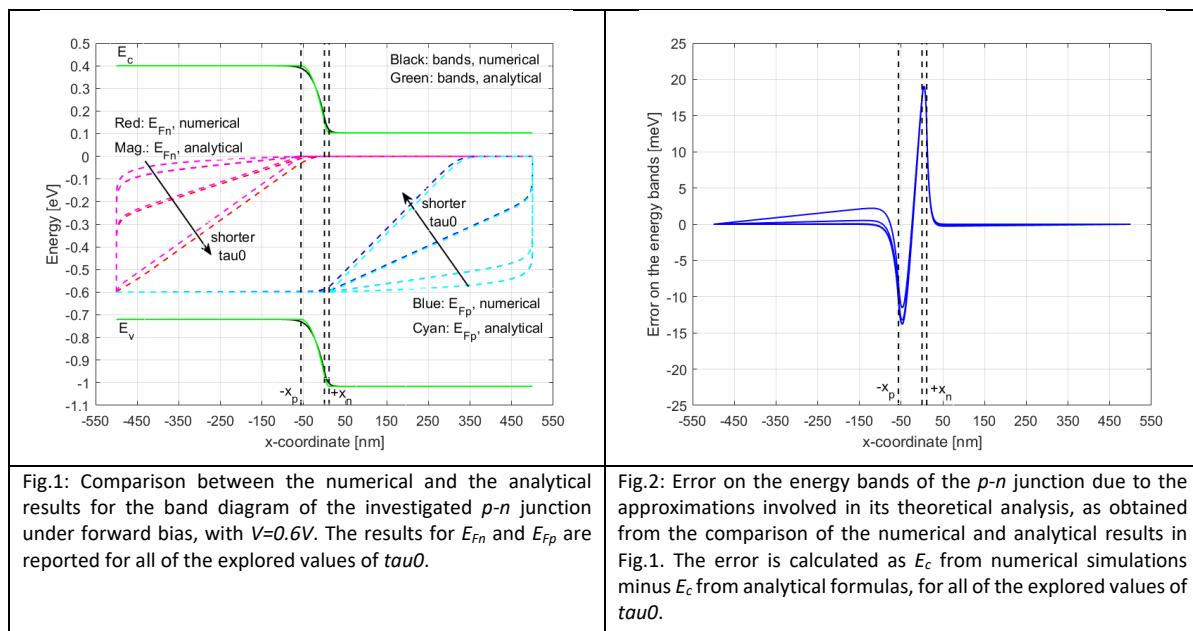
Lab 3

Step 6: Analysis of the results in Matlab and comparison with the analytical results

A. Results from Study 1

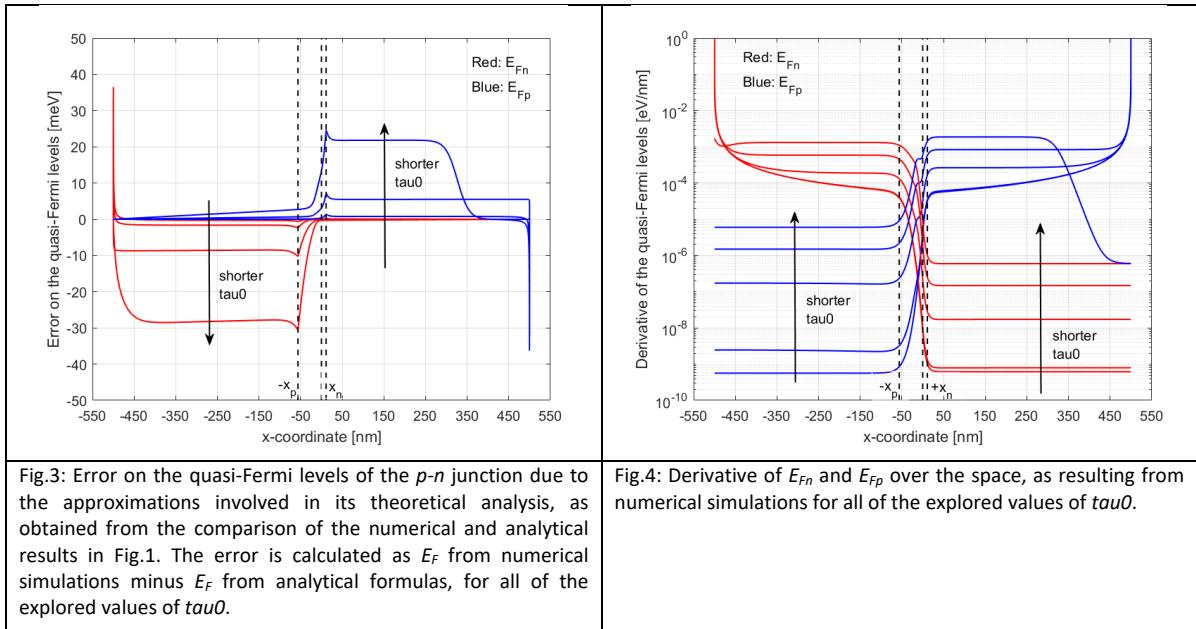
By loading into Matlab the file with the results from Study 1 generated by Comsol Multiphysics, the impact of $\tau_{\text{a}0}$ (carrier lifetime or characteristic time constant of the generation/recombination processes) on the operation of a p - n junction under bias can be explored. Besides, the numerical results obtained from the accurate solution of the Poisson and continuity equations in the device can be compared with the predictions of the ideal diode analysis (see the notes on “The p - n junction”) to check the accuracy of the assumptions and approximations involved in the latter. Students are kindly invited to try to perform some analyses and comparisons on their own, reproducing also the figures reported in this section (the sequence of Matlab commands used to generate the reported figures is, anyway, provided at the end of this section).

Fig.1 shows the band diagram of the investigated p - n junction as resulting from the numerical simulations performed in Comsol Multiphysics and as obtained from the analytical formulas coming from the ideal diode analysis under the same working conditions set in Study 1, i.e., forward bias $V=0.6\text{V}$ and temperature $T=300\text{K}$. In the figure, the quasi-Fermi level for electrons (E_{Fn}) and holes (E_{Fp}) are reported for all of the explored $\tau_{\text{a}0}$ values, while the conduction band edge (E_c) and valence band edge (E_v) are shown just for a single value of the parameter. Note, in fact, that the ideal diode analysis predicts no dependence of the band profile on $\tau_{\text{a}0}$ and numerical results nicely confirm this conclusion. The latter point is proved in Fig.2 by the fact that the discrepancy between the numerical and the analytical results for the band profile does not show any significant dependence on $\tau_{\text{a}0}$. In addition to that, Fig.2 also demonstrates that the analytical calculations provide a rather good description of device electrostatics. Similarly to the case of thermodynamic equilibrium, in fact, the discrepancy between the numerical and the analytical results is always less than the thermal energy kT , i.e., small. This proves the validity of all of the arguments and approximations used at the beginning of the ideal diode analysis to come to the band profile of the p - n junction under bias.



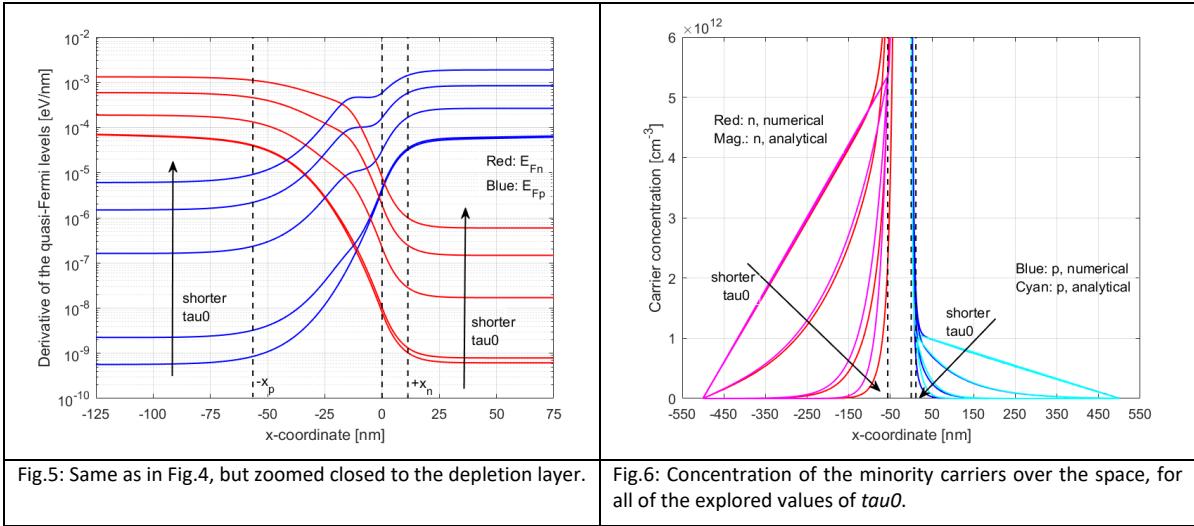
While the impact of $\tau_{\text{a}0}$ on E_c and E_v is negligible, that on E_{Fn} and E_{Fp} is relevant and clear from Fig.1. In particular, the change of $\tau_{\text{a}0}$ in the explored range from 10^{-5}s to $2 \times 10^{-13}\text{s}$ results in a significant change of the behavior of the quasi-Fermi level of minority carriers in the quasi-neutral regions of the device. Since the analytical results for E_{Fn} and E_{Fp} closely match the simulation results, we can get back to the ideal diode analysis to easily explain the observed trends. Focusing on what happens in the quasi-neutral p region, for instance, we can say that with $\tau_{\text{a}0}=10^{-5}\text{s}$ and 10^9s , the diffusion length L_n of electrons (minority carriers in

that region) equals, respectively, $139\mu\text{m}$ and $1.39\mu\text{m}$. Those lengths are significantly longer than the width of the quasi-neutral p region, making the device work as a narrow-base device. That means that the electron concentration is expected to follow a linear profile and E_{Fn} is expected to drop logarithmically from 0eV in the figure down to -0.6eV when moving from the depletion layer edge $-x_p$ to the contact of the p region. When, instead, τ_{00} is reduced to 10^{-12}s and $2 \times 10^{-13}\text{s}$, L_n decreases down to 44nm and 19nm . These lengths are shorter than the width of the quasi-neutral p region, meaning that the device assumes a wide-base behavior for those values of τ_{00} . As a consequence, the electron concentration is expected to decrease exponentially and E_{Fn} is expected to drop linearly from 0eV down to E_{Fp} when moving from $-x_p$ into the quasi-neutral p region. The slope of the linear E_{Fn} trend, moreover, is expected to be equal to kT/L_n , meaning that the drop of E_{Fn} is expected to get steeper for shorter L_n and, in turn, τ_{00} . For τ_{00} equal to 10^{-11}s , finally, an intermediate behavior between the case of narrow-base and wide-base is obtained, since $L_n=139\text{nm}$ is comparable to the width of the quasi-neutral p region. Similar arguments can be used to explain the behavior of E_{Fp} in the quasi-neutral n region.



In order to compare in detail the predictions of the ideal diode analysis with the numerical results obtained through Comsol Multiphysics in terms of quasi-Fermi levels, Fig.3 shows the discrepancy between them extracted from the E_{Fn} and E_{Fp} curves reported in Fig.1. For all of the explored values of τ_{00} but $2 \times 10^{-13}\text{s}$, the discrepancy is relatively small. In this regard, note that the peaks of the curves appearing close to the contacts of the p and n regions are likely due to inaccuracies of the numerical simulations in resolving the steep change of the quasi-Fermi level of minority carriers with the selected mesh for the discretization of the x-axis (this is completely negligible from the standpoint of the overall accuracy of the simulations, since the minority carrier concentration at those points is by orders of magnitude smaller than at the edges of the depletion layer). With this in mind, the peaks of the curves close to the contacts in Fig.3 are properly not a consequence of errors coming from the assumptions and approximations involved in the ideal diode analysis and the latter can be considered quite good even from the standpoint of the resulting quasi-Fermi levels. The only case for which a relevant discrepancy between numerical and analytical results appears (far from the contacts) in Fig.1 is that of $\tau_{00}=2 \times 10^{-13}\text{s}$. In that case, the behavior of E_{Fn} and E_{Fp} predicted by the ideal diode analysis is affected by errors that are comparable or even higher than kT . The reasons for that are related to the quite short diffusion length of minority carriers in the quasi-neutral regions of the device, as it will be better discussed later.

Although the ideal diode analysis provides a careful description of the behavior of E_{Fn} and E_{Fp} in the band diagram of Fig.1 (with the previously mentioned exception of the case with $\tau_{n0}=2 \times 10^{-13}s$), that analysis cannot be used to explore the weak change of the quasi-Fermi levels in the depletion layer of the device. In that analysis, in fact, E_{Fn} is considered to be perfectly flat from the contact of the n region to $-x_p$ and E_{Fp} is considered to be perfectly flat from the contact of the p region to $+x_n$ (unless second-order corrections are added in the analysis). The numerical results, instead, can be used to highlight even the very small changes of the quasi-Fermi levels in those regions. To that aim, Fig.4 shows the spatial derivative of E_{Fn} and E_{Fp} as a function of the position inside the device. Focusing on electrons and E_{Fn} (the discussion can easily be extended to the case of holes and E_{Fp}), an increase by orders of magnitude of the curves appears when moving from the quasi-neutral n region to the quasi-neutral p region. That is required to assure the continuity of the electron current density J_n in the presence of an electron concentration decreasing by orders of magnitude in the same direction. Besides, since the reduction of τ_{n0} gives rise to an increase of J_n (the reasons for that are left to the students' physical insight), an increase of the derivative of E_{Fn} is observed almost all over the device in Fig.4 when τ_{n0} is changed from $10^{-5}s$ to $2 \times 10^{-13}s$.



To better visualize what happens to the derivative of E_{Fn} and E_{Fp} close to the depletion layer of the p - n junction, a zoomed version of Fig.4 is provided in Fig.5. From this latter figure, the derivative of E_{Fn} and E_{Fp} clearly displays a large drop when moving into the depletion layer from, respectively, the quasi-neutral p region and the quasi-neutral n region. That is due to the combined action of the band bending in the depletion layer on the carrier concentration and of the need to keep the continuity of J_n and justifies the approximation made in the ideal diode analysis that inside the depletion layer E_{Fn} and E_{Fp} are constant in the band diagram. Again, the only case in which that approximation is not so good, since the drop of the derivative of E_{Fn} and E_{Fp} entering the depletion layer is not so strong, is that of $\tau_{n0}=2 \times 10^{-13}s$.

To complete the analysis, Fig.6 shows (on a linear y-scale) the behavior of the minority carrier concentration in the quasi-neutral regions of the p - n junction. As previously mentioned, for $\tau_{n0}=10^{-5}s$ and $10^{-9}s$, the device behaves as a narrow-base diode and the minority carrier concentration drops linearly when moving from the edges of the depletion layer to the contacts. When, instead, $\tau_{n0}=10^{-12}s$ and $2 \times 10^{-13}s$, the device reaches a wide-base behavior, with the minority carrier concentration dropping exponentially when entering the quasi-neutral regions from the depletion layer. For $\tau_{n0}=10^{-11}s$, an intermediate behavior is obtained. In the figure, a very good agreement between the numerical and the analytical results clearly appears for all the τ_{n0} values, with the exception, again, of the case with $\tau_{n0}=2 \times 10^{-13}s$. In this latter case, the curve for the minority carrier concentration coming from numerical simulations is significantly shifted towards the depletion layer with respect to that predicted by the ideal diode analysis, meaning that a stronger drop of the carrier concentration occurs inside the depletion layer. That is in agreement with the fact that, in the case of $\tau_{n0}=2 \times 10^{-13}s$, the quasi-Fermi levels display a nonnegligible drop inside the depletion layer in Fig.1. This

effect can be explained by considering that $\tau_{00}=2\times10^{-13}s$ is so short that the minority carrier diffusion length in the quasi-neutral regions is comparable to the Debye length in those regions. Since the bands require about a couple of Debye lengths inside the quasi-neutral regions to become fully flat starting from $-x_p$ and $+x_n$ (see the discussions in the “*Simulation Lab #2 – The p-n junction under thermodynamic equilibrium*”), that means that, when $\tau_{00}=2\times10^{-13}s$, the drop of the minority carrier concentration in the quasi-neutral regions occurs over an interval of the x-axis where a small band bending still exists and affects carrier transport. Due to that, the ideal diode analysis is for $\tau_{00}=2\times10^{-13}s$ not as accurate as for the other τ_{00} values. However, it is worth remarking that values of τ_{00} shorter than $10^{-9}s$, and in particular so short to provide a diffusion length of minority carriers comparable with the Debye length of the quasi-neutral regions, are unphysical. These values were considered in this Lab just to explore the case of a wide-base diode without increasing the width of the p and n regions, because that increase would have made difficult the clear visualization of the band profile and of the behavior of the quasi-Fermi levels in the device given the width of the depletion layer. Besides, in so doing, the ideal diode analysis was challenged with some extreme choice of device parameters and the overall accuracy encountered in these cases represents a general proof of its validity.

For some further investigations, students may try to repeat the analyses for different T or changing FW_p and FW_n (paying attention, in this latter case, to mesh discretization if large changes of device width are introduced). Besides, the case of reverse bias may be explored, trying to understand why some differences appearing in the quasi-Fermi level behavior are in the end irrelevant from the standpoint of device current.

B. Results from Study 2

By loading into Matlab the file with the results from Study 2 generated by Comsol Multiphysics, the J-V characteristics of the investigated p-n junction can be studied as a function of T , comparing the results with the theoretical analyses performed during lessons. Again, students are kindly invited to try to reproduce the analyses and figures reported below on their own (the sequence of Matlab commands used to generate the reported figures is, anyway, provided at the end of this section).

Fig.7 shows the J-V curve of the p-n junction resulting from numerical simulations at $T=200K, 300K, 400K$ and $500K$. Focusing, first, on the curve at $T=300K$, it is easy to recognize all the main features of the J-V dependence discussed during lessons. First of all, the curve displays a rectifying behavior, with J growing exponentially under forward bias and just weakly under reverse bias. Under forward bias, moreover, the growth of J with V displays an ideal slope of $60mV/dec$ in the intermediate J regime (marked as 1 in the figure, see the red dashed line). In the low J regime (marked as 2 in the figure, see the green dashed line), instead, the slope of the curve becomes equal to $\approx 100mV/dec$, which is not far from the expected $120mV/dec$ trend coming from the recombination processes in the depletion layer of the device. The discrepancy can be easily attributed to the fact that the contribution coming from the recombination processes is not fully dominant over the ideal diode behavior in the low J regime at $T=300K$, since the low current bump in the J-V curve arising from it appears only in a narrow voltage interval close to $V=0$. In the high J regime above $\approx 10^4A/cm^2$, finally, the J-V curve flattens with respect to the ideal diode behavior, as expected from high-injection and series resistance in the quasi-neutral regions.

The impact of T on J when the latter is dominated by the ideal diode behavior is clearly the one expected from the analyses performed during lessons, with the height and the slope of the curve in mV/dec growing when T is increased. Fig.8 shows that the value of V needed to have a constant $J=10^2A/cm^2$ decreases almost linearly with the growth of T , with a sensitivity coefficient equal to about $-1.9mV/K$.

From the standpoint of the impact of the generation/recombination processes in the depletion layer on the J-V curves, Fig.7 shows that reducing T from $300K$ to $200K$ makes the low current tail due to these processes far more evident, while increasing T to $400K$ and $500K$ makes it disappear. This is expected due to the different T activation of the contributions to J coming from the ideal diode behavior and from the generation/recombination processes in the depletion layer. In order to explore this activation with close attention, Fig.9 shows the value of J corresponding to $V=-500mV$ as a function of $1/kT$ in a semilogarithmic plot, which is typically referred to as an *Arrhenius plot*. The data points in the plot clearly reveal that in the low T regime, the sensitivity of J to T is weaker than in the high T regime. Quantitatively, the data points in

the two regimes can be fit by assuming an exponential behavior $J \propto \exp(-E_A/kT)$, where E_A is called the activation energy for J , leading to $E_A \approx 1.3\text{eV}$ at high T and $E_A \approx 0.6\text{eV}$ at low T . The E_A values in the high and low T regimes would in principle be equal to, respectively, the energy gap E_G of silicon and $E_G/2$ if J were only affected by the T dependence coming from the exponential term $\exp(-E_G/2kT)$ in the expression of the intrinsic carrier concentration n_i . Due to the additional dependences on T coming from the effective density of states, the diffusion coefficients (note, in this regard, that the T dependence of carrier mobility was not accounted for in the project) and E_G itself, some deviations of E_A from E_G and $E_G/2$ are obtained.

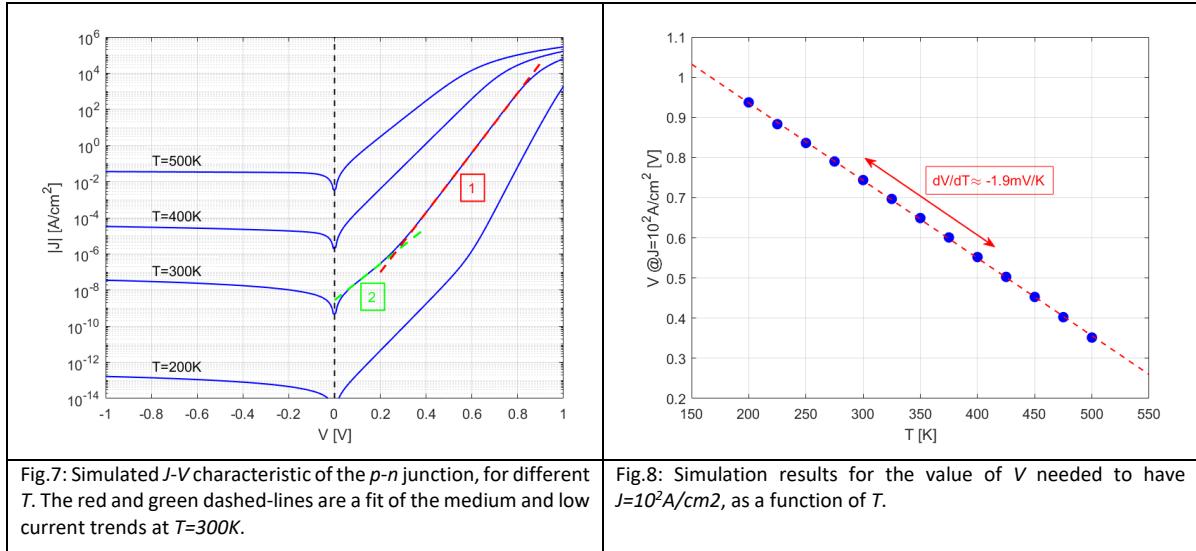
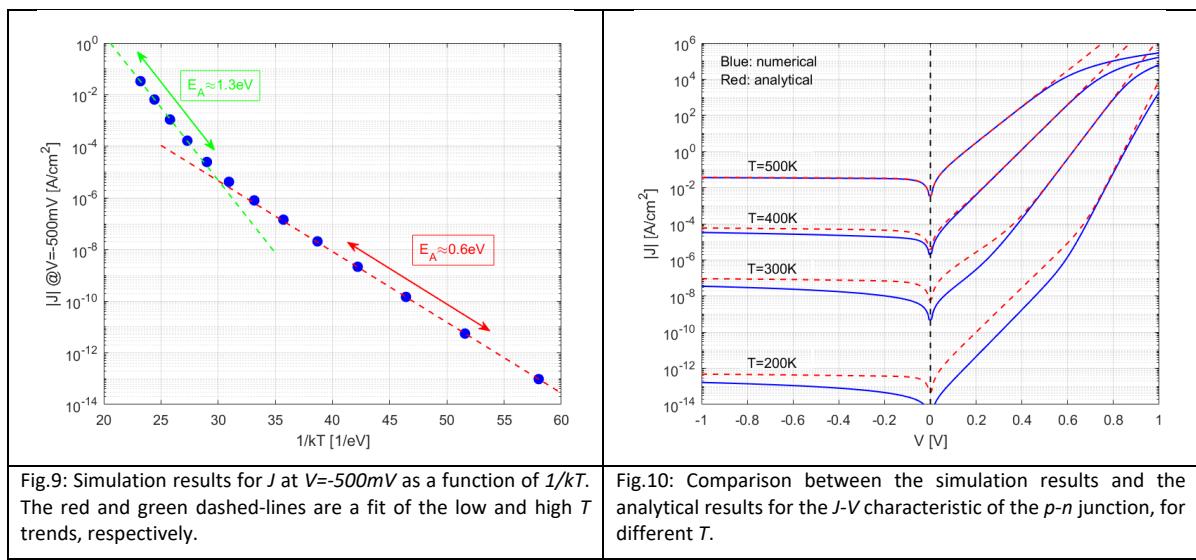


Fig.10 compares, finally, the simulation results with the analytical results obtained during lessons, considering only the low and medium J regimes in the latter case. The analytical results for J coming from the ideal diode analysis display a very good agreement with the simulation results in the intermediate J regime for all T , further confirming the validity of the approximations and assumptions involved in the analysis. The contribution to J coming from the generation/recombination processes in the depletion layer, instead, appears to be overestimated by the analytical results. However, that is not unexpected, since one of the hypotheses of the analysis performed during lessons to assess the contribution to J coming from the generation/recombination processes in the depletion layer is that the rate of these processes equals its maximum value all over the depletion layer.



Lab 4

Step 7: Analysis of the simulation results in Matlab

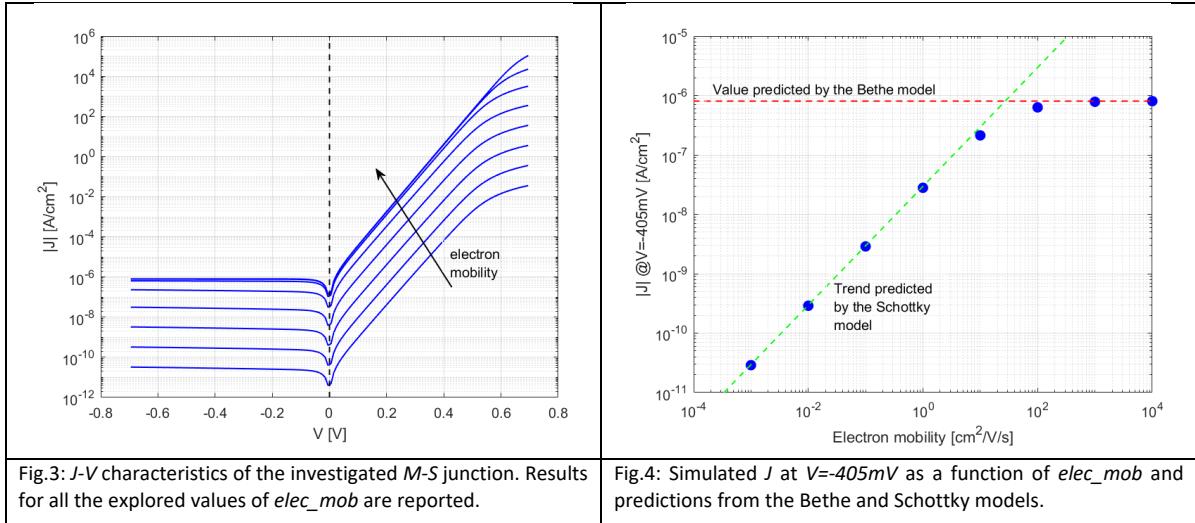
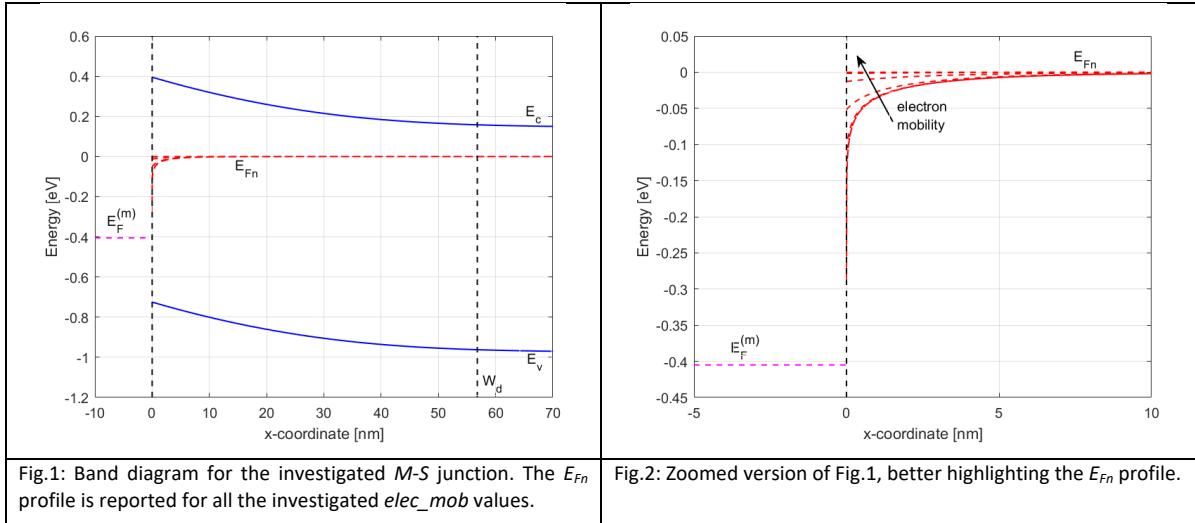
A. Results from Study 1

The files with the simulation results from Study 1 generated by Comsol Multiphysics can now be easily loaded into Matlab to investigate the dependence of device operation on electron mobility in the semiconductor (*elec_mob*). Students are kindly invited to try to perform the investigation on their own, reproducing the figures reported in this section (anyway, who is still not familiar with Matlab may find the sequence of commands used to generate the reported figures at the end of the analysis).

Fig.1 shows the band diagram of the investigated *M-S* junction as resulting from the numerical simulations performed in Comsol Multiphysics, in the case of forward bias with voltage $V=+405mV$ and for all the explored values of *elec_mob*. In agreement with the analyses presented during lessons, the bending of the conduction band edge (E_c) and of the valence band edge (E_v) creates a depletion layer at the surface of the *S* (width $Wd \approx 56nm$). The band bending is not affected by current transport and is, therefore, independent of electron mobility (students can easily verify this point on their own by inspecting the simulation results). The profile of the quasi-Fermi level for electrons (E_{Fn}), instead, significantly changes among the explored values of *elec_mob*. In particular, while the forward bias applied to the device gives rise, in general, to a decrease of E_{Fn} moving from the edge of the quasi-neutral region to the surface of the *S*, the total drop and the value of E_{Fn} at the *S* surface are significantly affected by *elec_mob*. This latter dependence can be better appreciated with the zoomed version of the band diagram reported in Fig.2. In the case of *elec_mob* equal to $10^4cm^2/Vs$ and $10^3cm^2/Vs$, E_{Fn} is almost flat all over the depletion layer. This is due to the fact that, for such high values of *elec_mob*, the bottleneck for electron transport is not the drift/diffusion of carriers through the depletion layer but their thermionic emission at the *M-S* interface. As studied during lessons, when that happens, the continuity of the electron flow requires maximizing the thermionic emission and, in turn, the electron concentration at the *S* surface. To achieve that and to reduce the drift/diffusion of carriers through the depletion layer down to the levels set by thermionic emission, E_{Fn} stays almost flat between the quasi-neutral region and the surface of the *S*. When *elec_mob* is reduced, however, the constraints of drift/diffusion to electron transport become more relevant. As a consequence, the continuity of the electron flow requires to enhance the latter mechanism and to decrease thermionic emission at the *M-S* interface. That is obtained with a more relevant drop of E_{Fn} along the depletion layer, so that the gradient of E_{Fn} (driving the drift/diffusion process) in that region is enhanced and the electron concentration at the *S* surface (driving thermionic emission) is reduced. Fig.2 shows that the E_{Fn} drop is still negligible in the case of $elec_mob=10^2cm^2/Vs$, but becomes extremely relevant in the case *elec_mob* is equal to or smaller than $10^1cm^2/Vs$. In this latter case, the quite low electron mobility makes the drift/diffusion of electrons through the depletion layer the bottleneck for carrier transport, with a negligible role played by thermionic emission at the *M-S* interface. When *elec_mob*= $10^0cm^2/Vs$ (which is a rather unrealistic value of electron mobility for a semiconductor material), finally, E_{Fn} approaches the energy corresponding to the Fermi level in the metal. As a final remark, note that the drop of E_{Fn} occurs quite close to the *S* surface, since the electron concentration reaches there its lowest values in the presence of the parabolic band bending of the depletion layer. As appearing from Fig.2, carefully investigating the drop of E_{Fn} requires then to use a very tight discretization mesh close to the *S* surface in the numerical simulations.

Fig.3 shows the simulated current density vs. voltage (*J-V*) characteristics of the investigated *M-S* junction for all the explored values of *elec_mob*. The curves reveal that, in any case, the device features a rectifying behavior, with the current increasing exponentially by $60mV/dec$ under forward bias and remaining nearly constant under reverse bias. Besides, similarly to the case of E_{Fn} , the dependence of *J* on *elec_mob* highlights that two possible working regimes exist for the device. The first is when electron mobility is high. In that case, *J* assumes its highest value and is almost independent of *elec_mob*. From what studied during lessons, that is expected, since in the case electron mobility is high, carrier transport is limited by thermionic emission at the *M-S* interface and drift/diffusion through the depletion layer does not significantly constrain it. The second regime is when electron mobility is low. In that case, the reduction of *elec_mob* results in a proportional reduction of *J*, which then decreases below the value reached in the high electron mobility regime. Again, that is expected from what studied during lessons, since carrier drift/diffusion through the

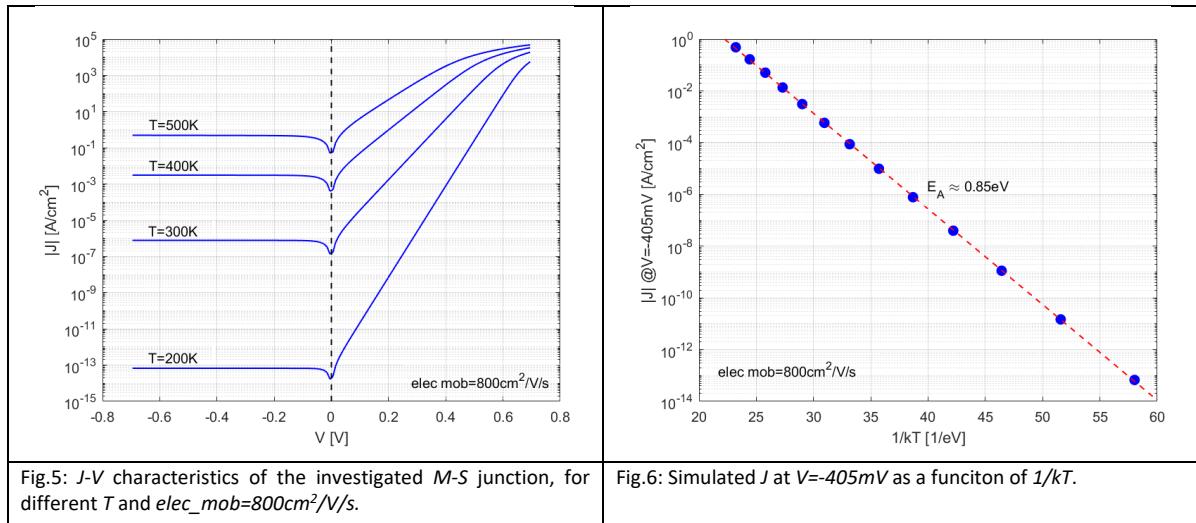
depletion layer represents the bottleneck for current transport in the case electron mobility is low. The two regimes can be better appreciated in Fig.4, where the reverse saturation current density at $V=-405mV$ is reported as a function of $elec_mob$ along with the predictions coming from the Bethe and Schottky models. The figure shows that the transition from the high to the low electron mobility regime occurs when this parameter decreases below a few tens of $cm^2/V/s$ and that the Bethe model and the Schottky model provide a rather accurate description of device behavior in the former and latter regimes, respectively.



B. Results from Study 2

Fig.5 shows the simulated J-V characteristics of the investigated *M-S* junction as a function of temperature T . Results are for $elec_mob=800\text{cm}^2/\text{V/s}$, i.e., the case of silicon with donor doping concentration $N_d=10^{17}\text{cm}^{-3}$. As discussed with the analysis of the results from Study 1, for such a high value of $elec_mob$ electron transport is limited by thermionic emission at the *M-S* interface and the Bethe model provides accurate predictions for J . Similarly to the case of the *p-n* junction, the increase of T results in the growth not only of the slope of the J-V curve under forward bias in mV/dec but also of the reverse saturation current density of the device. To explore this latter dependence, the reverse saturation current density at $V=-405\text{mV}$ is reported in Fig.6 as a function of $1/kT$ (*Arrhenius plot*), with kT being the thermal energy. The nice alignment of the data points along a straight line in the figure allows to say that the dependence of the reverse saturation current density

on T can be reproduced with an exponential relation. That relation is typically written as $J \propto \exp(-E_A/kT)$, with E_A that is called the *activation energy* for the current. From the fit of the data points in Fig.6, $E_A \approx 0.85\text{eV}$ can be extracted. In this regard, there are a few points that are worth highlighting. The first is that, from the Bethe model, the reverse saturation current density is expected to be proportional to $T^2 * \exp(-q\phi_{Bn}/kT)$, where $q\phi_{Bn}$ is the barrier height at the *M-S* interface. As a consequence, E_A should be slightly higher than $q\phi_{Bn}$ (slightly higher due to the T^2 prefactor in the previous formula), as confirmed by the results of Fig.6. The second is that, being E_A close to $q\phi_{Bn}$, the T dependence of the reverse saturation current density of a *M-S* junction is typically weaker than that of a *p-n* junction when the latter is dominated by the ideal diode behavior (E_A slightly higher than E_G is obtained in that case for the *p-n* junction). The last is that the generation/recombination processes in the semiconductor were neglected in the Comsol Multiphysics project and, thus, their possible impact on the results in Figs.5-6 is not accounted for. Due to the high reverse saturation current density of the device at room temperature, however, that impact is not expected to appear but at very very low T .



Lab 5

Step 5: Analysis of the numerical results in Matlab

The files with the simulation results generated by Comsol Multiphysics can now be easily loaded into Matlab for a detailed analysis of the electrostatics of the investigated *MOS* capacitor. Students are kindly invited to try to perform the analysis on their own, reproducing the figures reported in this section (anyway, who is still not familiar with Matlab may find the sequence of commands used to generate the reported figures at the end of the analysis).

Fig.1 shows the band diagram of the investigated *MOS* capacitor as resulting from the numerical simulations performed in Comsol Multiphysics in the presence of a gate voltage $V_G=1.5V$. Although the band profile in the oxide is not directly provided by Comsol Multiphysics, it can be easily determined given that i) there the bands are linear (no oxide charge) and ii) the total voltage drop over the material $V_{ox}=F_{ox} \cdot t_{ox}=F_s \cdot \epsilon_{si}/\epsilon_{ox} \cdot t_{ox}$ can be easily calculated from the simulation results (F_{ox} is the electric field in the oxide, F_s is the electric field at the silicon surface and $\epsilon_{si}, \epsilon_{ox}$ are the dielectric constants of silicon and the oxide, respectively; consider that no interface states are present at the silicon/oxide surface in the simulations). In Fig.1, the energy gap and the electron affinity of SiO_2 were assumed for the oxide layer.

Since $V_G=1.5V$ is much higher than the device threshold-voltage $V_T=0.53V$, the band diagram of Fig.1 is that of a strong-inversion condition. In particular, a depletion layer where the negative charge of ionized acceptors is exposed is present in the substrate, with width $W_d \approx 52nm$. The voltage drop $V_s=1.08V$ over this layer is slightly higher than $2/\phi_B=0.94V$, where ϕ_B is the electrostatic potential in the bulk region. The electron concentration in the inversion layer, moreover, is relevant for device electrostatics, being higher than the doping concentration N_a . This is clearly highlighted in Fig.2, where the spatial profile of the electron concentration resulting from the numerical simulations is shown. Note that this profile cannot be accurately determined via analytical calculations, since calculations do not allow to come to the electrostatic potential and the band diagram as a function of the position in the substrate under strong-inversion. Fig.2 shows that the electron concentration peaks at the silicon surface and rapidly decreases going deeper in the substrate. More specifically, the width over which the electron concentration is higher than the doping concentration is $\approx 3nm$. That width, which is in general less than $5nm$, can be considered as the thickness of the inversion layer. In this regard, it is worth mentioning that, due to the strong confinement of electrons in the potential well created by the oxide and the band bending close to the silicon surface, quantum-mechanical effects typically affect the profile of the electron concentration in the inversion layer of an *MOS* device. As a result of electron confinement, the peak of the electron concentration is not at the silicon/oxide interface but few nm deep into silicon. Even the thickness of the inversion layer changes a bit when quantum-mechanical effects are accounted for, remaining, anyway, less than $5nm$.

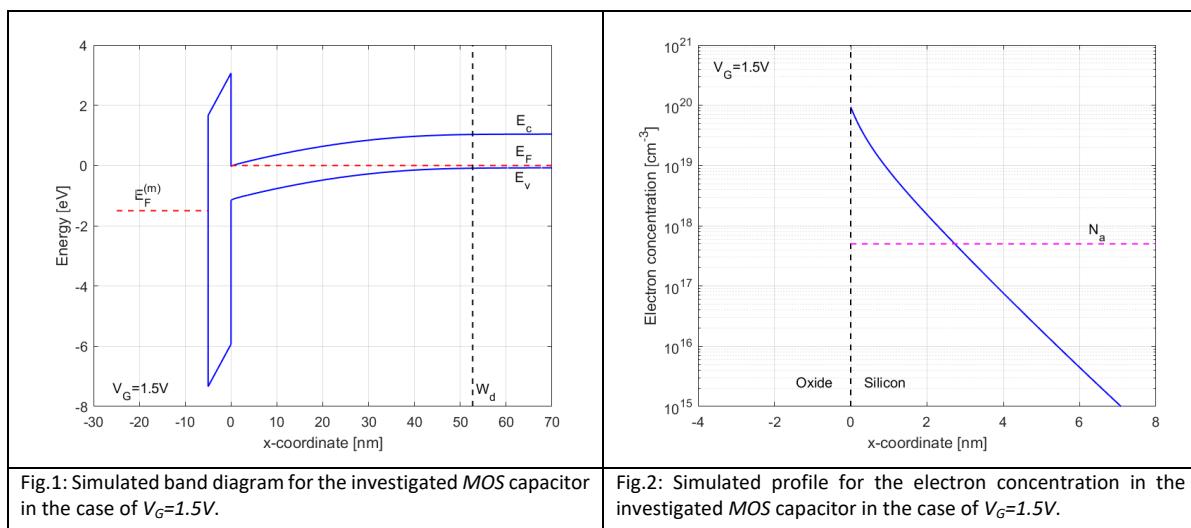


Fig.3 shows the electric field profile in the substrate of the *MOS* capacitor. Due to the strong-inversion condition in the device and the large electron concentration at the silicon surface, the profile is not a pure triangular profile. A pure triangular profile is what would result in the substrate under a depletion or weak-inversion condition (students can verify this point by repeating the analysis for a $V_G < V_T$), due to the fact that the constant charge density of ionized acceptors in the depletion layer would in that case be the only relevant charge density for device electrostatics. Under the strong-inversion condition considered in Figs.1-3, instead, the charge density coming from electrons close to the silicon surface is comparable or even higher than the charge density from ionized acceptors in the depletion layer. As a result, the slope of the electric field profile (which is proportional to the charge density exposed in each point of the substrate) grows from $-qN_a/\epsilon_{si}$ to a much higher value entering the inversion layer, as clearly appearing from Fig.3 (q is the elementary charge). That growth results in a steep increase of the electric field across the thickness of the inversion layer and in an additional contribution to V_s with respect to the case in which only the depletion-layer charge were present over the width W_d (remember that the area under the electric field profile corresponds to V_s). However, as evident from Fig.3, the latter contribution to V_s is very small, since the inversion layer thickness is very narrow. That leads to the so-called *charge-sheet approximation* (which will be better discussed in the analysis of the *MOS* transistor), which consists in neglecting the thickness of the inversion layer and the additional voltage drop over it with respect to the voltage drop over the depletion layer. As a result of this approximation, the electron charge in the inversion layer is considered to be a sheet of charge at the silicon surface (hence the name of the approximation) and the width of the depletion layer is calculated as $W_d = \sqrt{2\epsilon_{si}/q/N_a * V_s}$. Note, in this regard, that the latter formula for W_d would be valid in principle only in the presence of a pure triangular profile for the electric field in the substrate, i.e., in the depletion and weak-inversion regimes. Thanks to the charge-sheet approximation, it can be used to calculate W_d and, in turn, the depletion layer charge Q_{dep} even in the strong-inversion regime (see the red dashed line in Fig.3). From the numerical results, it is easy to realize that the involved approximation error is rather small, since the total depletion layer charge calculated through the formula $Q_{dep} = -qN_a W_d = -4.2 \times 10^{-7} C/cm^2$ is quite close to that resulting from the integration over the substrate of the difference between the acceptor and the hole charge densities, which equals $-4.09 \times 10^{-7} C/cm^2$. The approximation error on Q_{dep} is then just $\approx 3\%$.

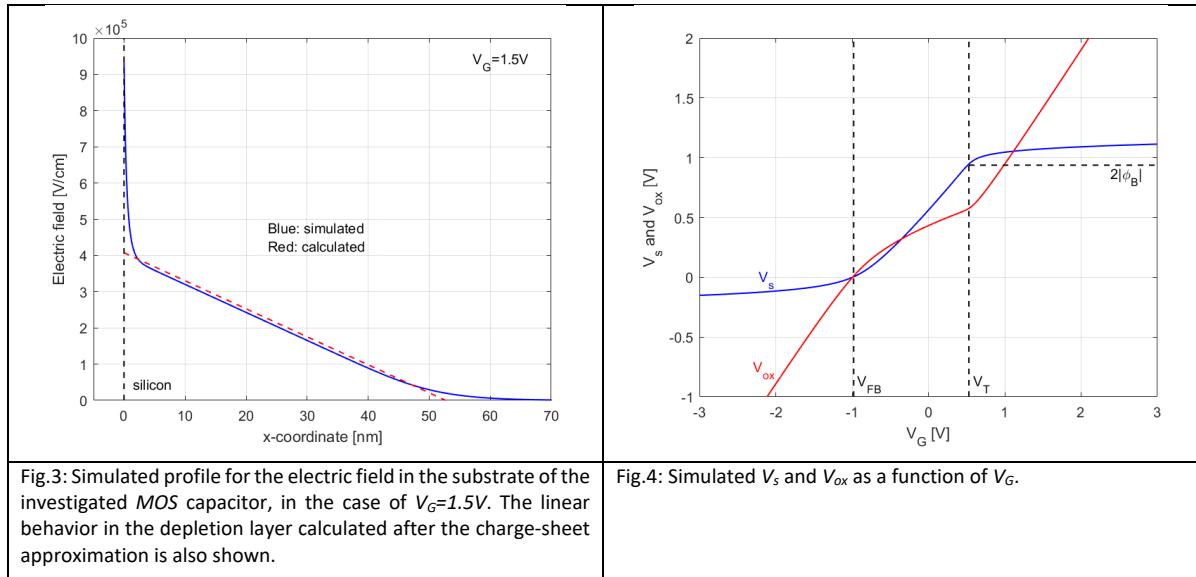
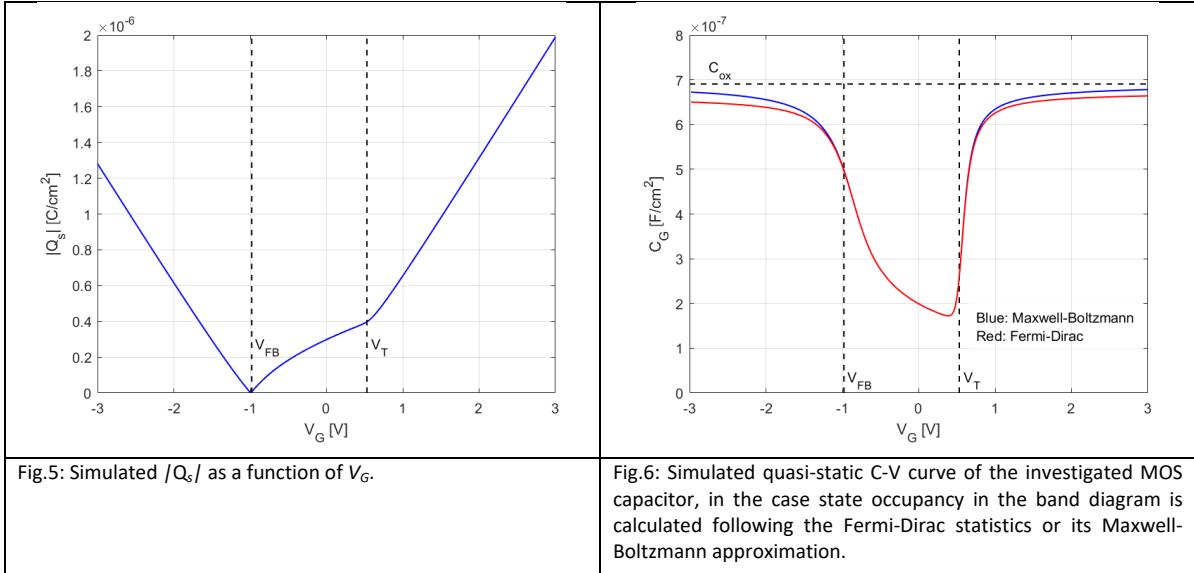


Fig.4 shows the simulation results for the dependence of V_s (blue curve) and V_{ox} (red curve) on V_G . While V_s is directly provided by Comsol Multipysics, V_{ox} has been extracted from the simulation results as $-Q_s/C_{ox}$, with Q_s being the substrate charge and C_{ox} the oxide capacitance. The behavior of V_s as a function of V_G reflects what obtained with simple arguments and analytical calculations during lessons. In particular, V_s becomes weakly negative below the flat-band voltage $V_{FB}=-0.98V$ and weakly increases above $2/\phi_B|=0.94V$ above

$V_T=0.53V$, displaying a trend that can be approximated as parabolic first and as linear later in-between these two voltages. The V_{ox} trend, instead, can be easily related to the Q_s vs. V_G dependence, which is shown in absolute value in Fig.5. As studied during lessons, Q_s (and then V_{ox}) grows almost linearly in the accumulation and strong-inversion regimes, while a square root dependence is reached in-between V_{FB} and V_T (close to V_T , however, a linear approximation of the V_{ox} vs. V_G dependence introduces just a small error, see Fig.4).



To complete the analysis of the stationary electrostatics of the investigated *MOS* capacitor, Fig.6 shows the quasi-static capacitance-voltage (*C-V*) curve of the device. This curve was obtained by calculating the small-signal gate capacitance $C_G=-dQ_s/dV_G$ from the simulated Q_s vs. V_G relation. As studied during lessons, this methodology to extract C_G is based on the assumption that the device follows a path of stationary conditions (or thermodynamic equilibrium conditions in the substrate in the presence of an ideal insulator) when V_G is slightly modified to evaluate C_G . In practical assessments of the *C-V* curve, that is a reasonable approximation only when the small-signal adopted to evaluate C_G is a low-frequency signal and, for that reason, the *C-V* curve reported in Fig.6 is also called a low-frequency *C-V* curve.

The results in Fig.6 provide a careful and comprehensive description of the C_G vs. V_G dependence, without the need for the regional approximations adopted in the calculations performed during lessons. Besides, by repeating the simulations performed in Comsol Multiphysics using the Fermi-Dirac statistics instead of its default Maxwell-Boltzmann approximation (that can be done by modifying the field **Carrier Statistics** under the tree **Model Properties** in the **Settings** window for Semiconductor), the impact of the latter on the quasi-static *C-V* curve can be easily studied. In particular, the curve obtained under the Fermi-Dirac statistics (red curve in Fig.6) displays a slower increase towards C_{ox} in the accumulation and strong-inversion regimes than the curve obtained under the Maxwell-Boltzmann approximation (blue curve in Fig.6). No relevant difference between the results appears, instead, in the depletion and weak-inversion regimes. This can be explained by considering that it is only in the accumulation and strong-inversion regimes that the Fermi level E_F is quite close either to the conduction band edge E_c or to the valence band edge E_v at the silicon surface, which is the condition making the Maxwell-Boltzmann approximation fairly inaccurate. The fact that the Fermi-Dirac statistics predicts a weaker growth of C_G towards C_{ox} in these regimes, then, is due to the weaker dependence of Q_s on E_F-E_c at the silicon surface and, in turn, on V_s resulting from that statistics with respect to the Maxwell-Boltzmann case. From that, a smaller substrate capacitance C_s and C_G result in the Fermi-Dirac case.

To further investigate the stationary electrostatics of an *MOS* capacitor, students are kindly invited to repeat the previous analysis for different doping concentrations N_a and different oxide thicknesses t_{ox} , exploring the change of the shape of the quasi-static *C-V* curve as a function of these latter parameters.

Lab 6

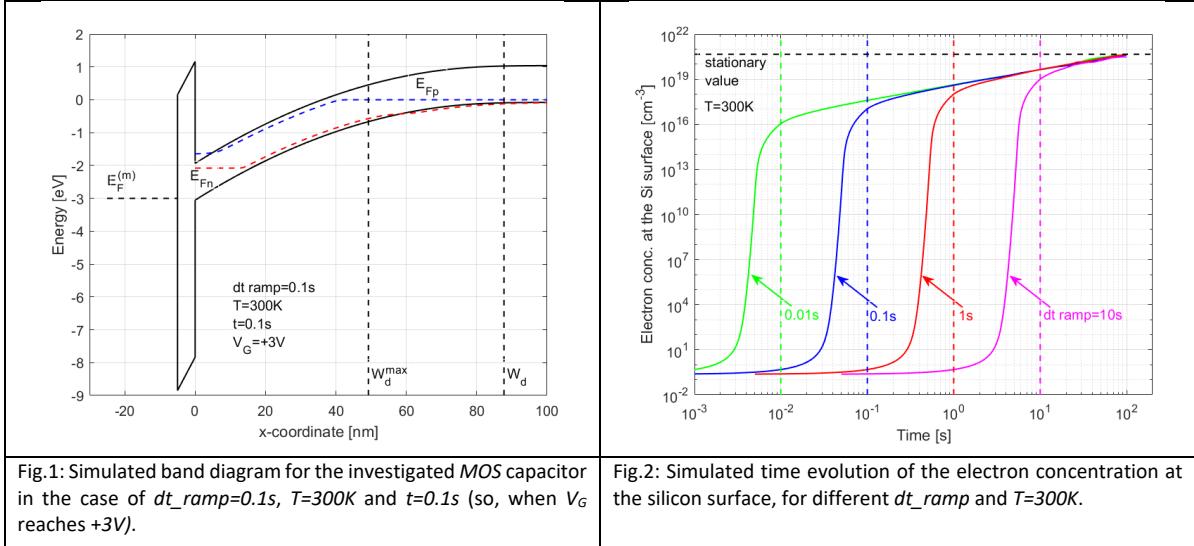
Step 5: Analysis of the numerical results in Matlab

In order to investigate the transient behavior of the MOS capacitor, students are asked to run the Comsol Multiphysics simulations for different values of temperature T and of the duration of the gate voltage ramp dt_ramp . In particular, the following combinations of the two parameters will be considered here: $T=300K$ and dt_ramp equal to $0.01s$, $0.1s$, $1s$ and $10s$; $dt_ramp=0.1s$ and T equal to $300K$, $325K$, $350K$, $375K$, $400K$ and $500K$. Note that the Comsol Multiphysics project was not created to process all these cases together because some of them may take up to 10 minutes on a high-performance PC and may be prohibitively long on low-performance hardware; by running the cases independently, students may skip some of them if they take more than 15 minutes on their PC (another option is to reduce the number of elements of the mesh discretizing the x -axis from 500 to 100, thing that gives rise to negligible changes in most of the results that will be presented here). After exporting the simulation results for each combination of T and dt_ramp , the files can be easily loaded into Matlab for a detailed analysis of the time evolution of the electrostatics of the investigated MOS capacitor. Students are kindly invited to try to perform the analysis on their own, reproducing the figures reported in this section (anyway, who is still not familiar with Matlab may find the sequence of commands used to generate the reported figures at the end of the analysis).

Fig.1 shows the band diagram of the investigated MOS capacitor as resulting from the numerical simulations performed in Comsol Multiphysics in the case of $T=300K$ and $dt_ramp=0.1s$. Results refer to the time instant $t=dt_ramp=0.1s$ and provide, therefore, a picture of the electrostatic condition of the device when the gate voltage ramp reaches its maximum value $V_G=+3V$. The separation between the quasi-Fermi level for electrons E_{Fn} and the quasi-Fermi level for holes E_{Fp} reveals that a relevant nonequilibrium condition is present in the substrate at that stage of the transient. That is, however, expected from what studied during lessons. Note, in fact, that $dt_ramp=0.1s$ is shorter than the typical timescale of seconds over which the generation processes take place in the substrate of an MOS capacitor at room temperature. That means that the generation processes are not able to create all the electrons needed to keep the substrate close to thermodynamic equilibrium as V_G grows from $-3V$ to $+3V$. More specifically, it is when V_G rises above the device threshold voltage V_T that the lack of a sufficient amount of electron generation plays a relevant role for the evolution of device electrostatics. For V_G higher than V_T , in fact, the electron concentration would be relevant for the electrostatics if the substrate were under thermodynamic equilibrium. With a limited amount of carrier generation during the gate voltage ramp, the electron concentration in the inversion layer remains below the values corresponding to thermodynamic equilibrium of the substrate and, due to that, the evolution of device electrostatics must rely more on the depletion layer charge. That means that the charge in the depletion layer, and then the depletion layer width W_d , must increase above their values in the presence of thermodynamic equilibrium to compensate for the lack of electrons close to the silicon surface. As a consequence, W_d grows above the maximum depletion layer width $W_d^{max}=sqrt(2\varepsilon_{si}/q/Na^2/\phi_B)$ and the voltage drop over the substrate V_s grows above $2/\phi_B$ (ϕ_B is the bulk potential, ε_{si} is the dielectric constant of silicon and q is the elementary charge). These effects are clearly evident from Fig.1, where the band bending in the substrate corresponds to $V_s \approx 3V$. The strong band bending in the substrate together with the need to keep the electron concentration in the inversion layer lower than that under thermodynamic equilibrium in the material give rise, finally, to the downward shift of E_{Fn} close to the silicon surface that makes evident substrate nonequilibrium.

From the simulation results of Fig.1 it is possible to see that E_{Fn} and E_{Fp} do not stay flat all over the depletion layer of the MOS capacitor, as typically assumed in the theoretical analyses of device behavior studied during lessons. The correct spatial profiles of E_{Fn} and E_{Fp} are the outcome of the combination of carrier generation and transport throughout the depletion layer and cannot be easily determined with analytical calculations. However, the assumption that these levels remain flat all over the depletion layer does not introduce any meaningful error in the analysis of device electrostatics. Note, in fact, that the downward bending of E_{Fp} moving from the bulk region to the silicon surface does not affect the conclusion that the hole concentration is much less than the doping concentration all over the depletion layer. In a similar way, the upward bending of E_{Fn} from the silicon surface to the edge of the depletion layer is absolutely irrelevant for device electrostatics, since it is only close to the silicon surface that the electron concentration is high enough to

play a role (it is just the distance of E_{Fn} from the conduction band edge E_C at the silicon surface that determines the electron concentration in the inversion layer, not the spatial profile by which E_{Fn} reaches its vertical position at the surface). For these reasons, the analysis of the MOS capacitor (but also of the MOS transistor) can neglect the E_{Fn} and E_{Fp} evolutions over the depletion layer, considering them to be flat.



Since the nonequilibrium condition created in the substrate by the gate voltage ramp triggers carrier generation (see the relative vertical position of E_{Fn} and E_{Fp} in Fig.1), the electrostatics of the MOS capacitor keeps evolving in the stretch of time following the instant at which V_G reaches its maximum value equal to $+3V$, even though during that stretch of time V_G stays constant. More specifically, substrate nonequilibrium is just a transient condition that vanishes over time due to the increase of the electron concentration in the inversion layer resulting from carrier generation in the material. In order to study the transient evolution of substrate electrostatics, then, the electron concentration at the silicon surface was considered. That concentration is shown as a function of time in Fig.2 for $dt_ramp=0.01s, 0.1s, 1s$ and $10s$, with $T=300K$. Results reveal that, irrespective of the value of dt_ramp , the electron concentration reaches its stationary value (which can be easily calculated by means of the Comsol Multiphysics project developed for the “*Simulation Lab #5 – Stationary electrostatics of the MOS capacitor*”) in about $100s$. The reduction of dt_ramp , on the other hand, gives rise to a reduction of the electron concentration at the instant at which V_G reaches $+3V$. With dt_ramp equal to $10s$ and $1s$, that concentration is about, respectively, $10^{19}cm^{-3}$ and $10^{18}cm^{-3}$. That means that, with such long ramp durations, the generation processes are able to create a strong-inversion condition in the MOS capacitor during the ramp itself (remember that the acceptor doping concentration of the substrate is $N_a=5\times10^{17}cm^{-3}$), even though the electron concentration in the inversion layer remains below the stationary value. In the case of $dt_ramp=0.1s$ and $0.01s$, instead, the electron concentration when the ramp reaches $V_G=+3V$ equals about $10^{17}cm^{-3}$ and $10^{16}cm^{-3}$, respectively. In these latter cases, then, the generation processes can give rise just to a weak-inversion condition in the MOS capacitor during the gate voltage ramp and the deep-depletion established in the substrate when V_G reaches $+3V$ is then stronger than in the cases of longer dt_ramp .

By using the simulation results for the substrate charge Q_s and the gate voltage V_G as a function of time obtained for the dt_ramp cases considered in Fig.2 ($T=300K$), the capacitance-voltage ($C-V$) curves of Fig.3 can be extracted. The small-signal capacitance C_G of the MOS capacitor has been calculated by exploiting the variations of Q_s and V_G during the gate voltage ramp, i.e., $C_G=-dQ_s/dV_G$ (in this regard, consider that the discretization of V_G and Q_s during the ramp is tight enough to accurately calculate the previous derivative as a function of V_G by numerical techniques). In the case of $dt_ramp=0.01s$ and $0.1s$, the C_G results are almost overlapped and reproduce the deep-depletion $C-V$ curve of the MOS capacitor. As previously discussed, in fact, for short dt_ramp the electrostatics in the device during the gate voltage ramp is dominated by the

depletion layer charge, since just a weak-inversion condition can be established by the generation processes in the substrate. For $dt_ramp=1s$ and $10s$, instead, a nonnegligible generation of electrons during the gate voltage ramp occurs and, as a result, the electrostatics in the device is not determined just by the depletion-layer charge but also by a partial contribution of the inversion layer charge. As a consequence, the $C-V$ curves corresponding to these cases are a bit higher than the deep-depletion curve obtained for shorter dt_ramp in the interval $V_G > V_T$.

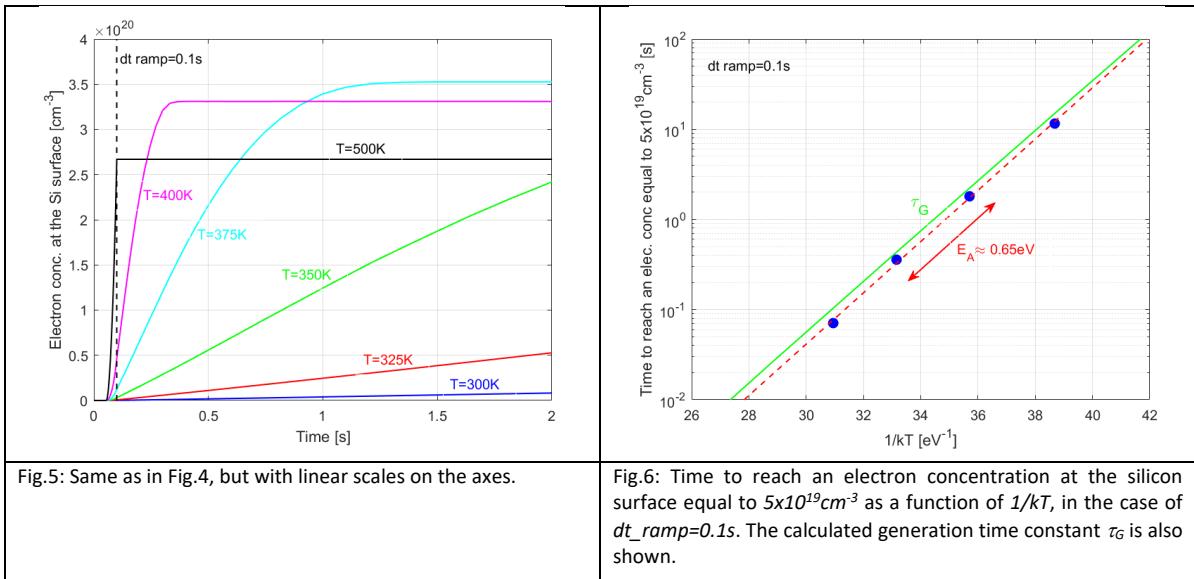
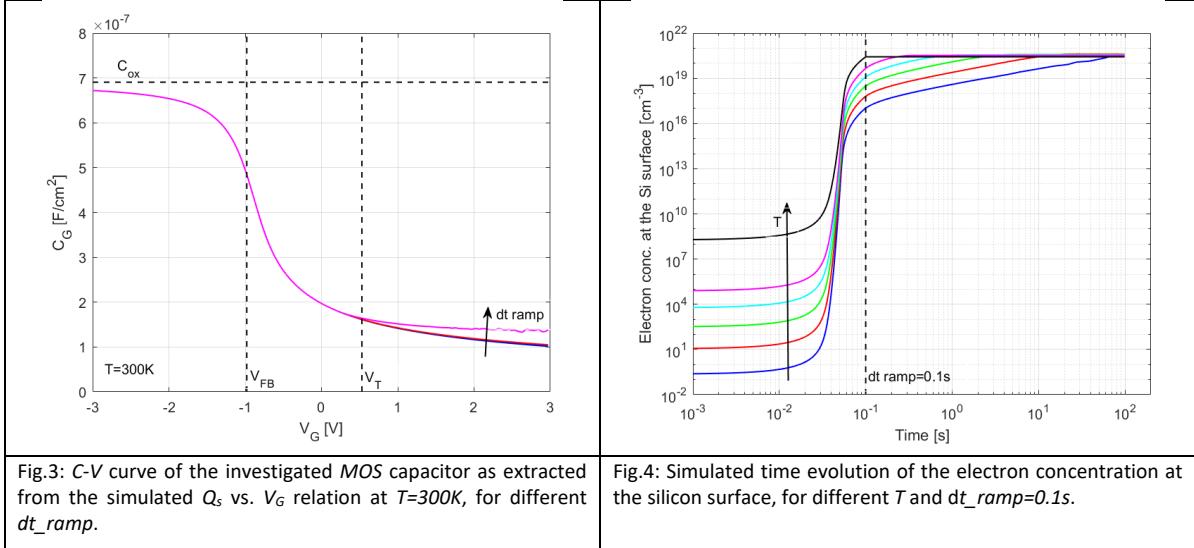
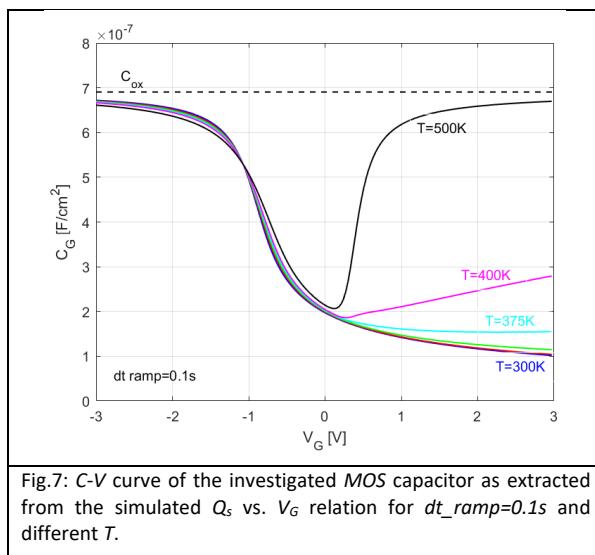


Fig.4 shows the simulated time evolution of the electron concentration at the silicon surface in the case of $dt_ramp=0.1s$ and T ranging from $300K$ to $500K$. Results reveal that the transient condition of the substrate induced by the gate voltage ramp vanishes earlier at higher T . In particular, at $T=500K$ the electron concentration at the silicon surface equals its stationary value already at the instant when V_G reaches $+3V$ (this is clearly evident when looking at the transients on a linear plot, see Fig.5), meaning that in that case a thermodynamic equilibrium condition is preserved in the substrate even during the gate voltage ramp. In order to quantitatively investigate the acceleration of the recovery of stationary electrostatics in the substrate with T , the time needed at $V_G=+3V$ to reach a constant electron concentration at the silicon surface equal to $5 \times 10^{19} \text{ cm}^{-3}$ is reported as a function of $1/kT$ in Fig.6 (Arrhenius plot, kT is the thermal energy). In the

figure, only the data points for $T=300K$, $325K$, $350K$ and $375K$ were considered, since for $T=400K$ and $500K$ a significant growth of the electron concentration already occurs during the gate voltage ramp (see Fig.5) and that results in a significantly different initial condition for the electrostatics in the substrate at the end of the ramp with respect to the other cases. A clear exponential acceleration of the recovery of the stationary condition in the substrate appears from Fig.6 when T increases. That exponential acceleration can be quantified by assuming an exponential dependence $\exp(E_A/kT)$ for the data points in the figure, with E_A representing the so-called *activation energy* of the recovery process. Starting from the slope of the points in the figure, $E_A \approx 0.65eV$ can be extracted.

The exponential acceleration of the recovery of the stationary condition in the substrate with T can be easily explained by considering the T dependence of carrier generation in the substrate. From what studied during lessons, the time constant for the generation processes in the substrate of an *MOS* capacitor can be written as $\tau_G = 2\tau_0 * N_o / n_i$, where τ_0 is the characteristic time of the generation/recombination processes and n_i is the intrinsic carrier concentration. Fig.6 shows that τ_G follows the same T activation of the previously discussed time needed to reach a constant electron concentration at the silicon surface (the quantitative vertical agreement between the two trends in the figure slightly depends on the value of the electron concentration selected to extract the recovery time from the simulation results). This result proves that it is the temperature activation of τ_G (arising from n_i) that determines the temperature activation of the recovery of stationary electrostatics in the device.



To complete the analysis, Fig.7 shows the C - V curves obtained from the simulation results corresponding to the cases considered in Figs.4-6. Due to the faster carrier generation in the substrate at higher T , a relevant change in the behavior of the curves appears for $V_G > V_T$. Over that interval, in fact, C_G decreases with the increase of V_G at $T=300K$, $325K$ and $350K$, following the previously discussed deep-depletion curve of the *MOS* capacitor. At $T=375K$ and $400K$, instead, C_G increases with V_G over that interval, due to a more relevant role played by the growth of the electron concentration in the inversion layer on device electrostatics during the gate voltage ramp. At $T=500K$, finally, the C_G behavior maps the low-frequency C - V curve of the capacitor, since carrier generation is so strong to keep the substrate close to thermodynamic equilibrium even during the gate voltage ramp.

For some further investigations, students may plot the band diagram of the *MOS* capacitor at different times during the gate voltage ramp and during the stretch of time over which V_G stays constant at $+3V$, at different T . In so doing, the dependence of the electrostatics of the device on holes and electrons in the substrate can be better understood, along with the change of the total charge in the substrate over time and the change of the voltage drop over the substrate and over the oxide.

Lab 7

Step 5: Analysis of the numerical results in Matlab

The files with the simulation results generated by Comsol Multiphysics can now be easily loaded into Matlab for a detailed analysis of the electrostatics and of current transport in the investigated *MOS* transistor, addressing the impact of the drain voltage V_{DS} on its ON-state behavior. Students are kindly invited to try to perform the analysis on their own, reproducing the figures reported in this section (anyway, who is still not familiar with Matlab may find the sequence of commands used to generate the reported figures at the end of the analysis).

Fig.1 shows the simulated drain current (I_{DS}) vs. V_{DS} characteristics of the investigated *MOS* transistor corresponding to an applied gate voltage V_{GS} equal to 0.5V, 0.75V, 1V, 1.25V and 1.5V. For all of the selected V_{GS} values, the typical ON-state trend of I_{DS} with V_{DS} featuring the ohmic, parabolic and saturation regimes clearly appears. As studied during lessons, the three regimes hold, respectively, for low, intermediate and high V_{DS} . In the ohmic regime, I_{DS} grows almost linearly with V_{DS} . In the parabolic regime, the I_{DS} increase with V_{DS} becomes markedly less-than-linear, following a parabola with the concavity directed downwards. In the saturation regime, finally, I_{DS} becomes weakly dependent on V_{DS} . From these results, it is possible to come to the conclusion that the threshold-voltage V_T of the investigated device is less than 0.5V. This conclusion is confirmed by the formula $V_T = V_{FB} + 2/\phi_B + \sqrt{2\epsilon_s * qN_a * 2/\phi_B} / C_{ox}$, which leads to $V_T \approx 0.3V$ (V_{FB} is the flat-band voltage, ϕ_B the electrostatic potential in the bulk region, ϵ_s the dielectric constant of silicon, q the elementary charge, N_a the substrate doping concentration and C_{ox} the oxide capacitance per unit area). As a final remark about Fig.1, students are kindly invited to pay attention to the y-axis scale showing the I_{DS} values. First of all, I_{DS} has been reported normalized to the width W of the transistor expressed in μm (pay attention that $W=1\mu m$ was assumed in the Comsol Multiphysics project), which is a quite common way to show the I_{DS} values. After this normalization, I_{DS} spans the typical range from a few hundreds of $\mu A/\mu m$ to few $mA/\mu m$. In this regard, it is worth mentioning that the scaling path followed by the CMOS technology over the years has resulted in the growth of the maximum W -normalized ON-state I_{DS} of *MOS* transistors (for the minimum value of the device channel length L of the technology). That was due to the impossibility to stick to the constant-field scaling rules, which would have maintained the maximum W -normalized I_{DS} constant, and to the need to introduce some generalized-scaling steps along technology evolution.

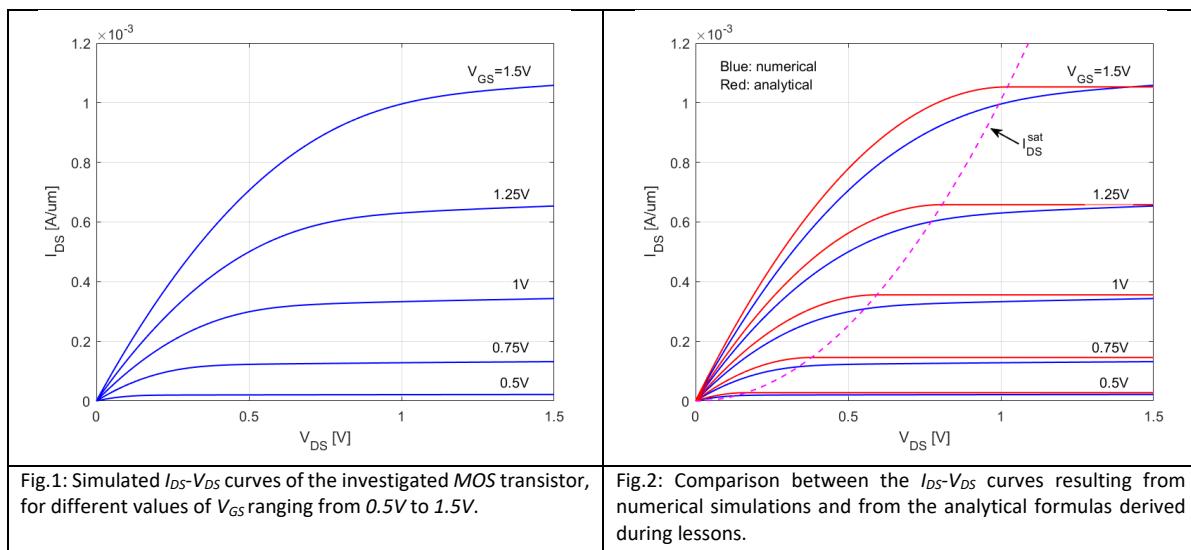


Fig.2 compares the I_{DS} - V_{DS} curves resulting from numerical simulations with the predictions of the analytical formulas derived during lessons, i.e., $I_{DS} = \mu_n C_{ox} W / L * [(V_{GS} - V_T) * V_{DS} - m V_{DS}^2 / 2]$ in the ohmic and parabolic regimes and $I_{DS} = I_{DS}^{sat} = \mu_n C_{ox} W / L * (V_{GS} - V_T)^2 / 2m$ in the saturation regime (μ_n is the electron mobility and $m = 1 + C_{dep}/C_{ox} = 1.17$, with C_{dep} being the minimum depletion-layer capacitance in the substrate corresponding

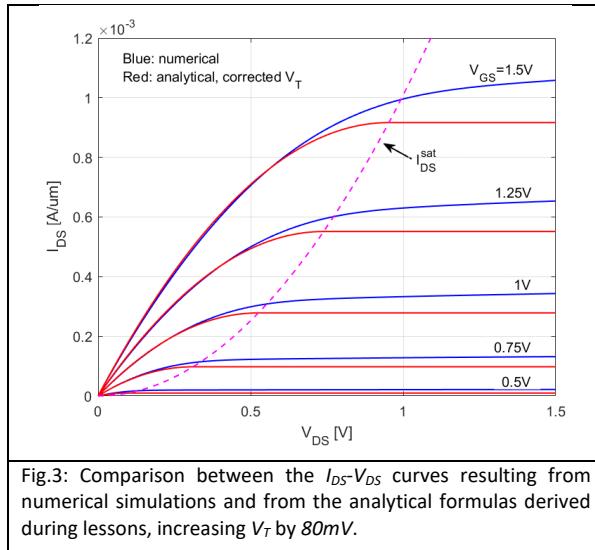
to a value of the surface potential $V_s=2/\phi_B|$). Although a rough general agreement between the two sets of curves appears from the figure, it is clearly evident that the analytical formulas overestimate I_{DS} with respect to what obtained from numerical simulations. In order to understand the reasons for that, it is worth starting from the ohmic and parabolic regimes and considering that the investigated device can be safely assumed to be a long-channel MOS transistor, since $L=300nm$ is significantly longer than the critical length $2(W_d^{max}+3t_{ox})\approx83nm$ separating the long-channel from the short-channel regime (W_d^{max} is the maximum depletion-layer width corresponding to $V_s=2/\phi_B|$ and t_{ox} is the oxide thickness). The long-channel nature of the device allows to safely say that the gradual-channel approximation adopted to come to the analytical expression for I_{DS} in the ohmic and parabolic regimes is valid. The origin of the mismatch between the analytical and the simulation results in these regimes, then, is to be attributed to some other approximation done after it. In this regard, recalling the steps taken during lessons to come to the analytical expression for I_{DS} , it is possible to identify three major approximations following the gradual-channel approximation. The first is the charge-sheet approximation, which allows to calculate the inversion charge Q_{inv} at each point in the channel even under strong-inversion as $Q_{inv}=-C_{ox}*(V_{GS}-V_{FB}-V_s)+sqrt(2\varepsilon_i*qN_a*V_s)$, where the square-root term represents the absolute value of the depletion-layer charge Q_{dep} (remember that Q_{inv} and Q_{dep} are charges per unit area). The second is the regional approximation $V_s=2/\phi_B|+V$, with V representing the quasi-Fermi potential for electrons. The third is that V is smaller than $2/\phi_B|$, even though it is not completely negligible with respect to it. After these three approximations, the expression $Q_{inv}=-C_{ox}*(V_{GS}-V_T-mV)$ was obtained and, from it, I_{DS} was calculated. Among the three, the second approximation is by far the roughest and is mainly responsible for the mismatch between the analytical and the simulation results in the ohmic and parabolic regimes in Fig.2.

In order to understand the role of the approximation $V_s=2/\phi_B|+V$ on device electrostatics, we may focus our attention on the ohmic regime. In that regime, V_{DS} is so small that V is negligible with respect to $2/\phi_B|$, making the previously mentioned third approximation surely appropriate. Setting $V_s=2/\phi_B|$, then, the relations $Q_{dep}=Q_{dep}^{max}=-sqrt(2\varepsilon_i*qN_a*2/\phi_B|)$ and $Q_{inv}=-C_{ox}*(V_{GS}-V_T)$ can be obtained. Although these results provide, of course, the most elementary description of what happens to Q_{inv} and Q_{dep} in an MOS device entering strong-inversion, they do not account for some important details related to device behavior near V_T . In particular, saying that Q_{inv} is zero when $V_{GS}=V_T$ and then linearly grows when V_{GS} increases above V_T misses a careful description of what happens to that charge close to V_T . A more accurate description of device physics should consider that for $V_{GS}<V_T$ both Q_{inv} and the change of Q_{inv} arising from a slight change of V_{GS} are negligible with respect to Q_{dep} and to its change. When $V_{GS}=V_T$, instead, Q_{inv} is still small and negligible with respect to Q_{dep} , but its change arising from a slight change of V_{GS} becomes equal to the change of Q_{dep} (remember, in fact, that when $V_{GS}=V_T$ the total substrate capacitance C_s of the MOS system is twice the value of the depletion-layer capacitance C_{dep} , meaning that the capacitance of the depletion layer and of the inversion layer are equal). For V_{GS} higher than V_T , then, device electrostatics is significantly affected by Q_{inv} and by its change with V_{GS} and that is what makes the device work in the strong-inversion regime (pay attention on that Q_{inv} is not required to be dominant over Q_{dep} for that). Since Q_{inv} is exponentially related to V_s , moreover, the growth of V_{GS} above V_T rapidly makes the inversion-layer capacitance much larger than not only the depletion-layer capacitance but also the oxide capacitance C_{ox} . That means, first of all, that the change of Q_{inv} following a slight change of V_{GS} becomes much stronger than the change of Q_{dep} , which can then be neglected. Then, that also means that the change of Q_{inv} becomes equal to the product between C_{ox} and the amplitude of the V_{GS} variation, since the latter drops mainly over the oxide. From the standpoint of the V_s vs. V_{GS} relation, this phenomenology makes V_s first deviate from the almost linear trend with V_{GS} observed under weak-inversion as soon as V_{GS} rises above V_T and then strongly flatten, reaching almost a saturation, when V_{GS} is well above V_T . It is only when V_s flattens, in particular, that the change of Q_{dep} vanishes and Q_{inv} becomes linearly related to V_{GS} . Therefore, some hints about how much V_{GS} must be increased above V_T to achieve those conditions can be derived by looking at the V_s vs. V_{GS} curve (see the notes on the MOS capacitor to accurately visualize that curve). From the inspection of that curve, V_s must increase by few kT/q above $2/\phi_B|$ before reaching a true saturation (kT/q is the thermal voltage). kT/q is, in fact, the characteristic voltage ruling the exponential dependence of Q_{inv} and, in turn, the inversion-layer capacitance, on V_s . By increasing V_s by few kT/q above $2/\phi_B|$, and then V_{GS} by a similar amount above V_T , the inversion-layer capacitance grows from the value

corresponding to the depletion-layer capacitance to a value much larger than it, giving rise to V_s and Q_{dep} saturation and to a linear dependence of Q_{inv} on V_{GS} .

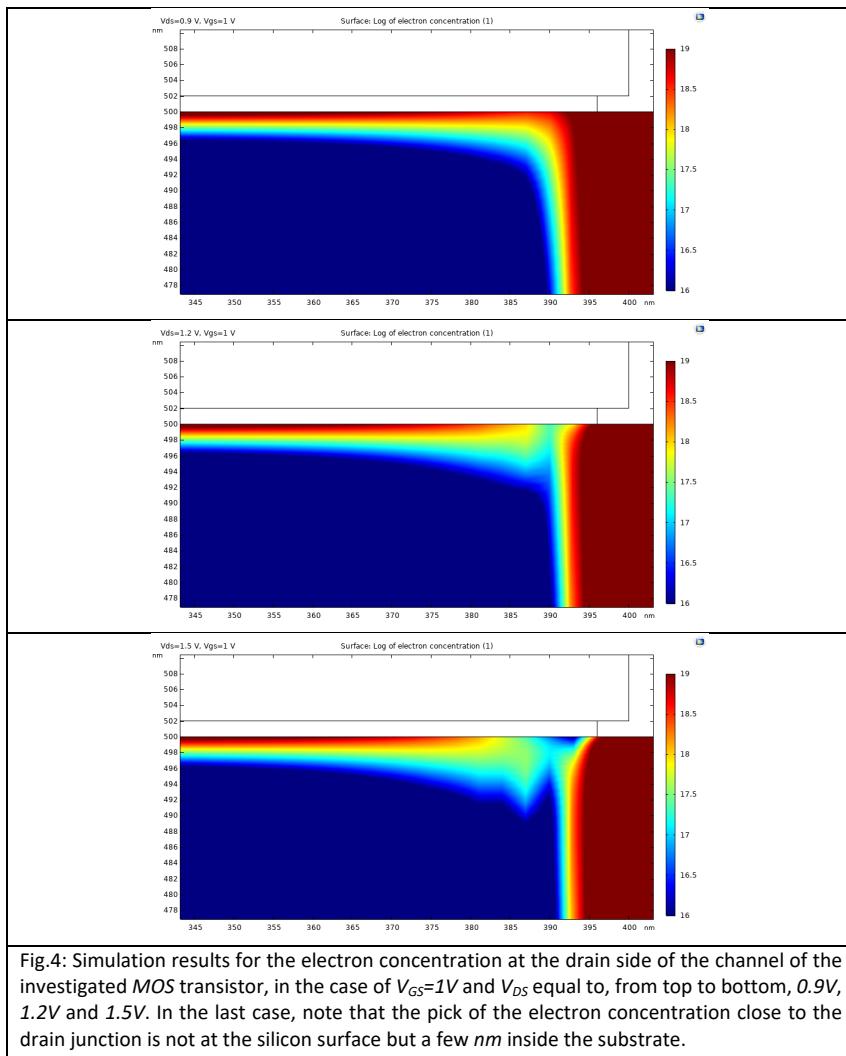
From the previous picture, it should be clear that assuming $V_s=2/\phi_B|$, $Q_{dep}=Q_{dep}^{max}$ and $Q_{inv}=-C_{ox}*(V_{GS}-V_T)$ under strong-inversion represents a rough approximation, since it corresponds to considering the inversion-layer capacitance to abruptly increase from a value much lower than C_{dep} and C_{ox} to a value much higher than them at $V_{GS}=V_T$. That misses that the increase of the inversion-layer capacitance with V_s is fast but, anyway, gradual and that an increase of V_s above $2/\phi_B|$ by few kT/q is needed to make that capacitance much larger than C_{dep} and C_{ox} . In turn, that misses that an increase of V_{GS} above V_T by few kT/q is needed to make V_s saturate, to make the small-signal gate capacitance approach C_{ox} and to make Q_{inv} reach a linear dependence on V_{GS} .

Getting back to the mismatch between the simulation and the analytical results for I_{DS} in the ohmic and parabolic regimes in Fig.2, it is easy to understand that one of the reasons why the formula $I_{DS}=\mu_n C_{ox} W/L * [(V_{GS}-V_T) * V_{DS} - m V_{DS}^2 / 2]$ overestimates the correct I_{DS} coming from simulations is that it underestimates the value of V_{GS} needed to achieve a linear growth of Q_{inv} with V_{GS} . In order to prove this conclusion, the value of V_T used in the formula for I_{DS} was slightly increased by 80mV to take into account the need to increase V_{GS} and V_s by few kT/q to make the inversion-layer capacitance much larger than C_{dep} and C_{ox} and, then, reach V_s saturation and a linear dependence of Q_{inv} on V_{GS} . These V_T -corrected curves are reported in Fig.3 and display a fairly good agreement with the simulation results in the ohmic and parabolic regimes. As a final remark, it is worth pointing out that the increase by 80mV of the V_T value to be adopted in the formulas for I_{DS} to reach such a good agreement with the simulation results is relevant because the V_T of the investigated MOS transistor is just about 300mV. In the case of devices with a much higher V_T and working, in turn, with much higher overdrives ($V_{GS}-V_T$), the correction to be introduced on the V_T value to account for the finite inversion-layer capacitance is typically negligible.



The residual mismatch between the simulation and the analytical results appearing in Fig.3 close to the edge between the parabolic and the saturation regimes can be attributed to both the gradual-channel approximation and the assumption that V and, in turn, V_{DS} are much smaller than $2/\phi_B|$. The latter assumption, in fact, is not so accurate at the edge of the parabolic regime, where V_{DS} is rather large. The gradual-channel approximation, on the other hand, cannot provide a careful description of device electrostatics approaching and entering the saturation regime, as discussed during lessons. That is due to the strongly-bidimensional electrostatics in the channel when V_{DS} approaches the value leading to saturation. In particular, the simplified electrostatic analysis based on the gradual-channel approximation misses the channel-length modulation effect, which is responsible for the weak increase of I_{DS} in the saturation regime (in Fig.3, in fact, the theoretical I_{DS}^{sat} was considered to be perfectly flat under saturation). That effect can be

clearly appreciated from the simulation results for the electron concentration reported in Fig.4. In order to partially correct this approximation, the Early voltage $V_A=L^*F_p$, where F_p is an effective electric field between the position of the pinch-off point in the channel and the drain, was extracted in Fig.5. That was done by extrapolating the linear trend observed in the simulated I_{DS} curves under saturation down to the horizontal axis, obtaining from the intersection $V_A \approx 15V$ (the curve at $V_{GS}=1.5V$ was not considered in the extraction since a too narrow V_{DS} interval is available for it in the saturation regime, leading to a relevant error in the extrapolation). As studied during lessons, V_A can be used to calculate the output resistance of the device in the saturation regime (going from $24k\Omega$ to $740k\Omega$ for the explored V_{GS} values in our case) and to correct the analytical I_{DS} formulas. In order to obtain a gradual increase of I_{DS} due to the channel-length modulation effect when approaching and entering the saturation regime, the correction is typically introduced by multiplying I_{DS} by $(1+V_{DS}/V_A)$ both in the parabolic and in the saturation regime (during lessons, the corrective term $1+(V_{DS}-V_{DS}^{sat})/V_A$ was introduced only for I_{DS} in the saturation regime, leading to a discontinuity of the derivative of the I_{DS} - V_{DS} relation at the edge between the parabolic and the saturation regime). This latter correction partially improves the agreement between the calculated and the simulated I_{DS} - V_{DS} curves at high V_{DS} (students can easily verify this point). To further improve the correctness of the formulas for I_{DS} , these are modified and made more complex in the compact models adopted by circuit simulators. However, the simple expressions studied so far are typically enough for some first order back-of-the-envelope calculations related to circuit operation.



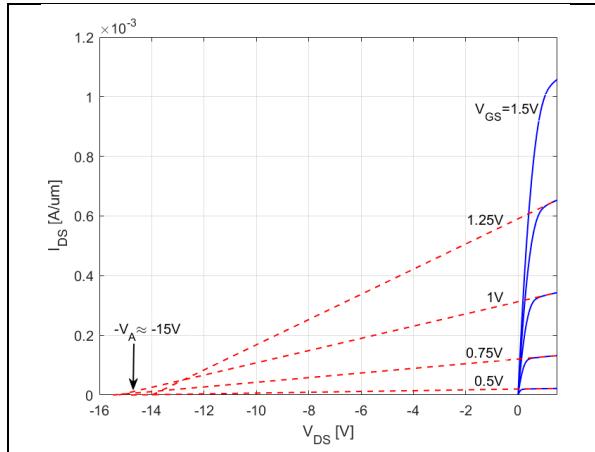


Fig.5: Extraction of the Early voltage V_A of the investigated *MOS* transistor via the linear extrapolation of the simulated I_{DS} trend in saturation down to the horizontal axis.

To complete the analysis of the operation of the investigated *MOS* transistor, Fig.6 shows the simulated conduction band edge (E_c) and E_{Fn} profiles at the channel surface in the source-to-drain direction, in the case of $V_{GS}=1V$ and for increasing V_{DS} from $0V$ to $1.5V$. As discussed during lessons, for low V_{DS} equal to $0.1V$ and $0.3V$ (ohmic regime), the band bending in the horizontal direction is almost linear and so is the bending of the quasi-Fermi level for electrons E_{Fn} . For higher V_{DS} equal to $0.5V$, $0.7V$ and $0.9V$ (parabolic regime), then, the reduction of Q_{inv} moving from the source side to the drain side of the channel (see Fig.7) results in a stronger band bending and E_{Fn} drop close to the drain. For V_{DS} equal to $1.2V$ and $1.5V$ (saturation regime), finally, a very steep band bending and drop of E_{Fn} appear close to the drain, while far from the drain the bands and E_{Fn} keep a profile that is independent of V_{DS} . This latter piece of evidence clearly demonstrates that in the saturation regime the drain loses control of the electrostatics in the channel, but for the channel-length modulation effect. Fig.8 shows the complete band diagram of the device for each V_{DS} condition.

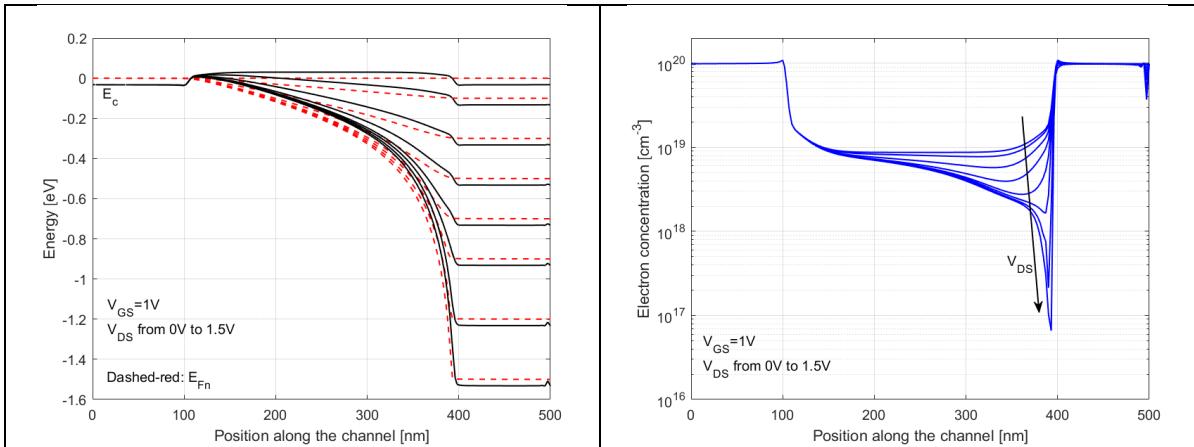


Fig.6: Simulation results for the E_c and E_{Fn} profiles along the channel of the *MOS* transistor, in the case of $V_{GS}=1V$ and for increasing V_{DS} from $0V$ to $1.5V$.

Fig.7: Simulation results for the electron concentration at the surface of the *MOS* transistor, in the case of $V_{GS}=1V$ and for increasing V_{DS} from $0V$ to $1.5V$.

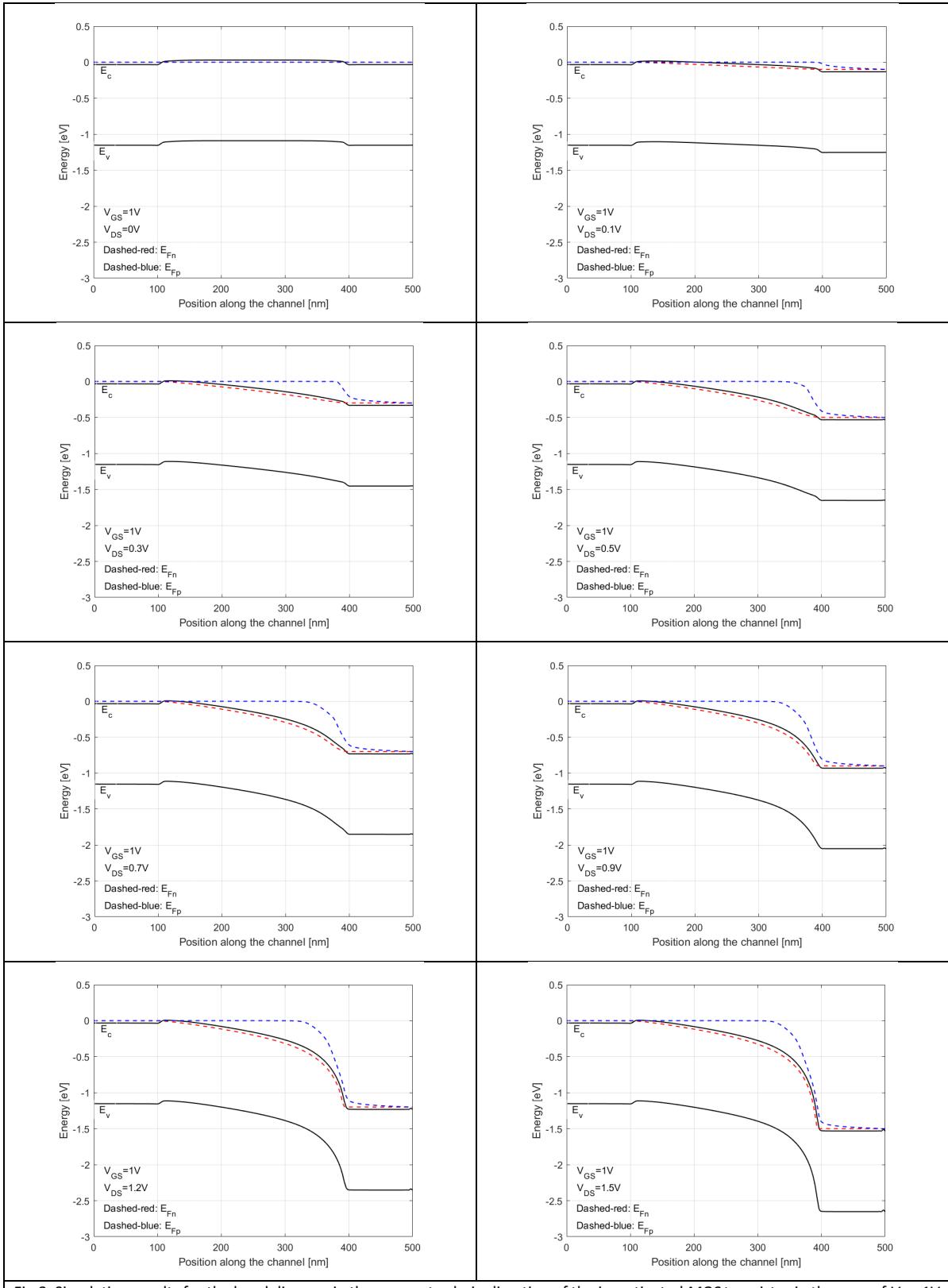
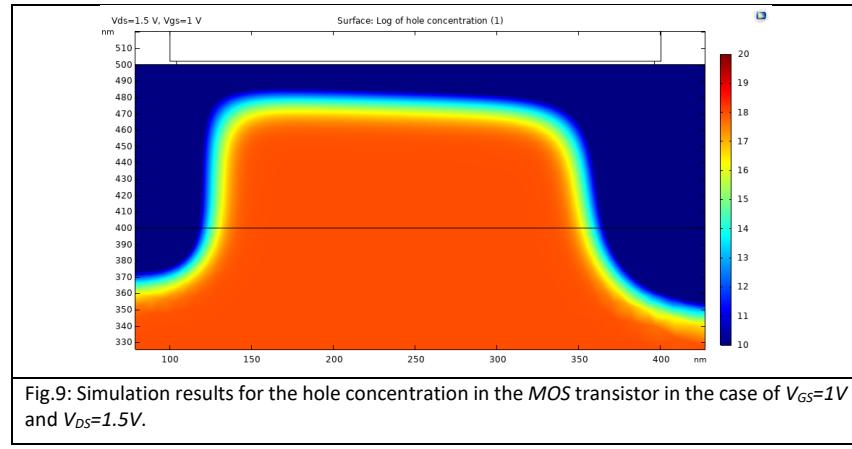


Fig.8: Simulation results for the band diagram in the source-to-drain direction of the investigated MOS transistor in the case of $V_{GS}=1\text{V}$ and V_{DS} equal to 0V (first row, left), 0.1V (first row, right), 0.3V (second row, left), 0.5V (second row, right), 0.7V (third row, left), 0.9V (third row, right), 1.2V (fourth row, left) and 1.5V (fourth row, right).

As a final result related to device electrostatics, Fig.9 shows the hole concentration profile in the substrate at $V_{GS}=1V$ and $V_{DS}=1.5V$. The width of the depletion layer at the source side and its increase moving to the drain side of the channel can be clearly appreciated.



Lab 8

Step 4: Analysis of the numerical results in Matlab

The files with the simulation results generated by Comsol Multiphysics can now be easily loaded into Matlab for a detailed analysis of the electrostatics and of current transport in the investigated *MOS* transistor, addressing the impact of the gate voltage V_{GS} on both its subthreshold and ON-state behavior. Students are kindly invited to try to perform the analysis on their own, reproducing the figures reported in this section (anyway, who is still not familiar with Matlab may find the sequence of commands used to generate the reported figures at the end of the analysis).

Fig.1 shows the simulated drain current I_{DS} , source current I_S and substrate current I_{sub} of the investigated *MOS* transistor as a function of V_{GS} , in the case of drain voltage $V_{DS}=0.1V$ and temperature $T=300K$. The semilogarithmic plot reveals the typical shape of the $I_{DS} - V_{GS}$ transcharacteristics (blue curve) studied during lessons and featuring: i) a very-low current regime where I_{DS} is nearly constant and equal to I_{sub} (magenta curve), due to carrier generation in the channel-region and at the drain-bulk junction of the device producing electrons gathered by the drain contact and holes gathered by the bulk contact; ii) an intermediate current regime where I_{DS} grows exponentially with V_{GS} due to electron diffusion from source to drain at the channel surface, corresponding to the weak-inversion regime of the *MOS* device; iii) a high-current regime above the threshold-voltage $V_T \approx 0.3V$ where the dependence of I_{DS} on V_{GS} turns into a power-law dependence (then flattening in the plot), due to the onset of strong-inversion in the channel. In this regard, it is worth pointing out that the current level corresponding to the generation processes in the substrate is very low at room temperature in the presence of a low $V_{DS}=0.1V$. That means that, under those conditions, it may be difficult to detect it experimentally, being typically overwhelmed by measurement noise (of course, that level may be more easily detected by increasing T or V_{DS}). Besides, it is also worth recalling that the generation current flows between the drain and bulk contacts of the device and does not affect I_S . As a consequence, differently from I_{DS} , I_S preserves the exponential trend corresponding to the subthreshold regime down to V_{GS} values approaching the flat-band voltage $V_{FB} \approx 0.99V$ (see the red curve in Fig.1). For such low V_{GS} values, the dependence of the surface potential V_s in the channel region on V_{GS} weakens, turning from linear to parabolic and then to logarithmic. That weakens the dependence of the inversion charge in the channel on V_{GS} as well, resulting in I_{DS} flattening.

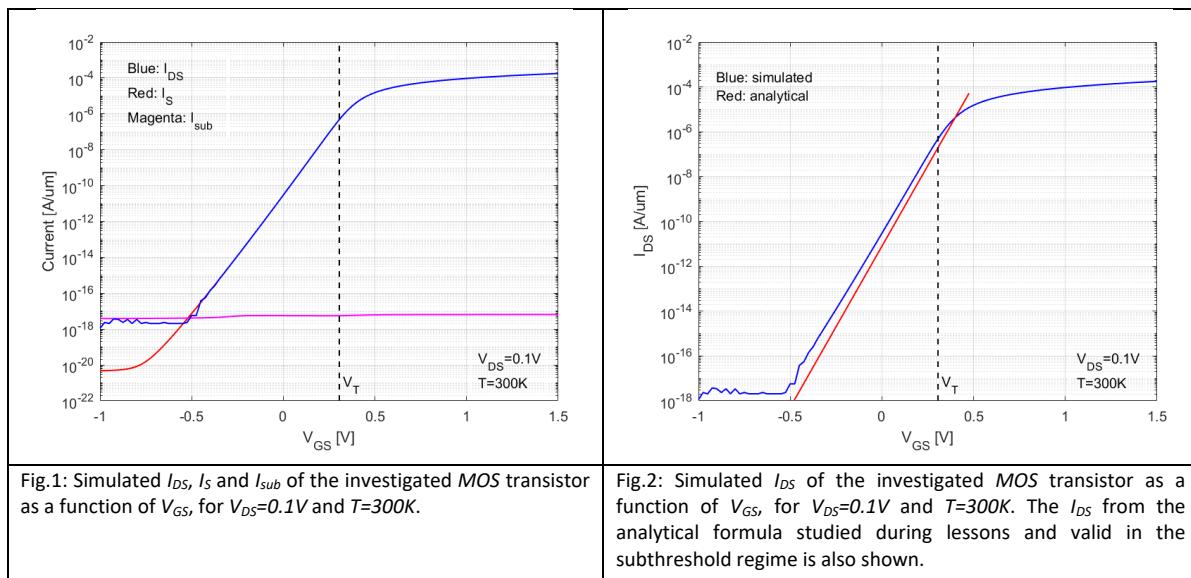
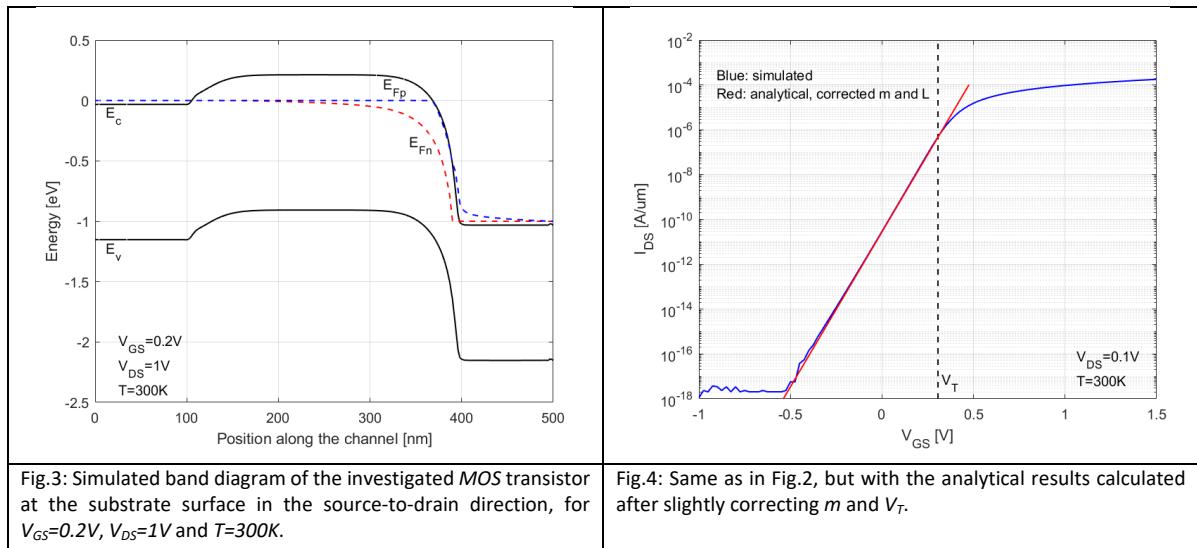


Fig.2 shows a comparison between the predictions of the analytical formula derived during lessons for I_{DS} in the subthreshold regime $I_{DS} = \mu_n C_{ox} W / L * (m-1) * (kT/q)^2 * \exp[q(V_{GS}-V_T)/(mkT)]$ and the simulated $I_{DS} - V_{GS}$ transcharacteristics (μ_n is the electron mobility, C_{ox} is the oxide capacitance, W and L are the channel width

and length, $m=1+C_{dep}/C_{ox}$, C_{dep} is the depletion-layer capacitance evaluated for $V_s=2/\phi_B$, ϕ_B is the bulk potential, kT is the thermal energy, q is the elementary charge). Although a rough agreement between the analytical and the numerical I_{DS} values appears from the figure, some nonnegligible differences between them are also evident. In particular, the analytical I_{DS} is a bit lower than the simulated one and, moreover, displays a stronger growth with V_{GS} , corresponding to a smaller subthreshold slope in mV/dec . Recalling that the subthreshold slope is just $STS=kT/q*\ln(10)*m$, the mismatch between its analytical and numerical value can only be the result of inaccuracies in the assessment of m . These may arise from the depletion approximation and from some nonnegligible bidimensional electrostatic effects coming from the source and drain regions, both slightly affecting the value of C_{dep} . The lower I_{DS} , on the other hand, may be attributed to slight inaccuracies in V_T estimation and to a shorter “effective” channel length of the device. To understand this latter point, Fig.3 shows the band diagram of the device in the source-to-drain direction at the substrate surface in the case of $V_{GS}=0.2V$, i.e., less than V_T , $T=300K$ and $V_{DS}=1V$ ($1V$ and not $0.1V$ was assumed for V_{DS} to better visualize the profile of the quasi-Fermi level for electrons E_{Fn}). Since L is longer than the critical length $2(W_d^{max}+3t_{ox})\approx83nm$ separating the long-channel from the short-channel regime (W_d^{max} is the maximum depletion-layer width corresponding to $V_s=2/\phi_B$ and t_{ox} is the oxide thickness), as expected from what discussed during lessons, the investigated MOS transistor can be safely considered to be a long-channel device and there is a wide region of the channel where the bands are flat and the electrostatics is mainly controlled by the vertical action of the gate. However, Fig.3 also highlights that the length of the transition regions at the sides of the channel, where the electrostatics is significantly affected by the presence of the source and drain n^+ junctions, is not completely negligible with respect to L (in other words, L is not “much” longer than $2(W_d^{max}+3t_{ox})$). This leads to defining a so-called “effective” channel length that is slightly different from L . In this analysis, however, this correction is not introduced and only a slight change of V_T is taken into account to correct the mismatch between the I_{DS} values of Fig.2. Fig.4 shows that, by slightly increasing m from its analytical value equal to 1.17 to 1.21 and by reducing V_T by 20mV with respect to its analytical value (this latter correction will be better validated in the “*Simulation Lab#9 – Scaling the channel length of an MOS transistor*”), the I_{DS} predicted by the analytical formula for the subthreshold regime of the MOS transistor can nicely fit the simulation results. In this regard, it is worth pointing out that circuit simulators rely on compact models for the electrical characteristics of the adopted electron devices that may be based on simple expressions such as those studied for the I_{DS} of an MOS transistor in the subthreshold and in the ON-state regimes, but the parameters involved in those expressions are typically calibrated against simulation results and experimental data for the devices, to have them accurately reproduce device behavior. In addition, further dependences of the parameters in the formulas for the electrical characteristics are introduced to account for additional physical effects that were not considered in a first-order analysis of the devices.



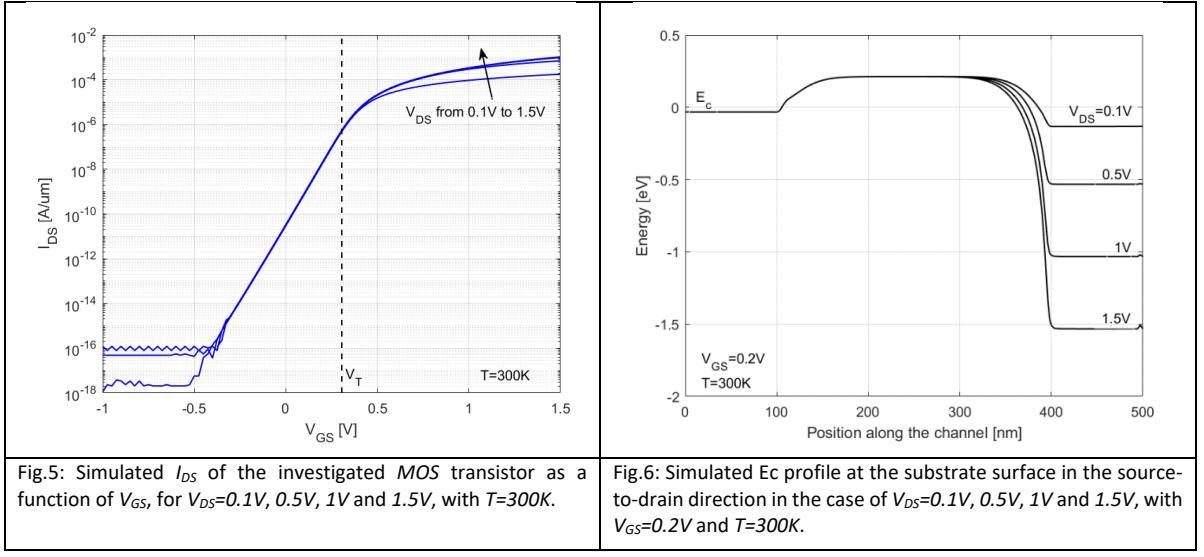
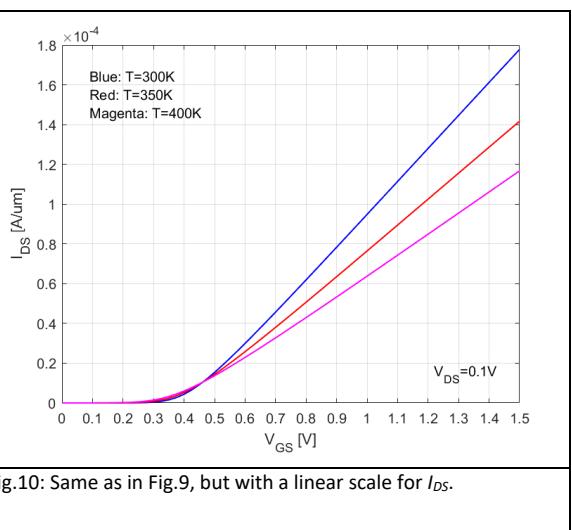
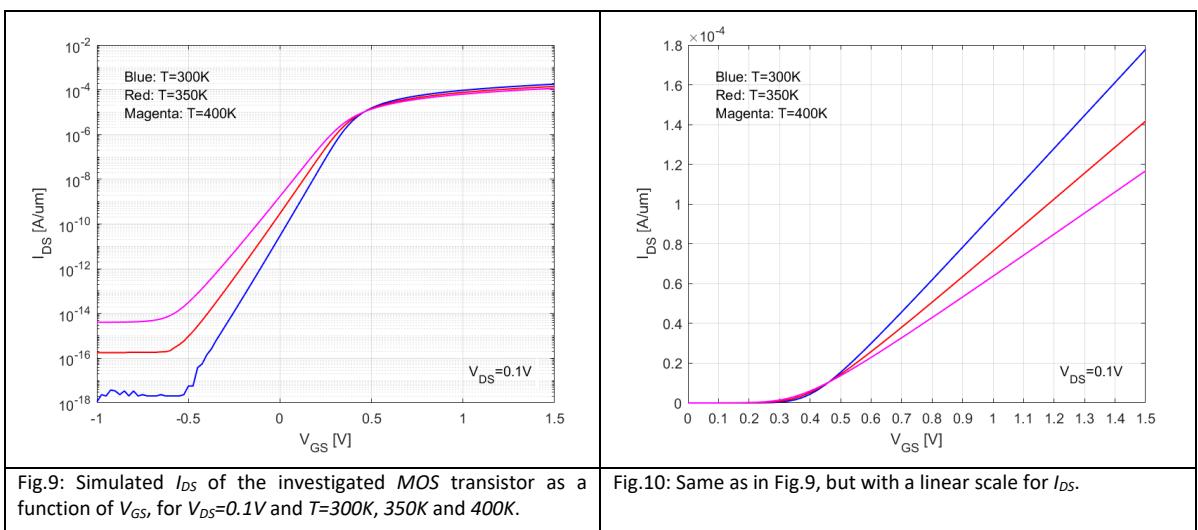
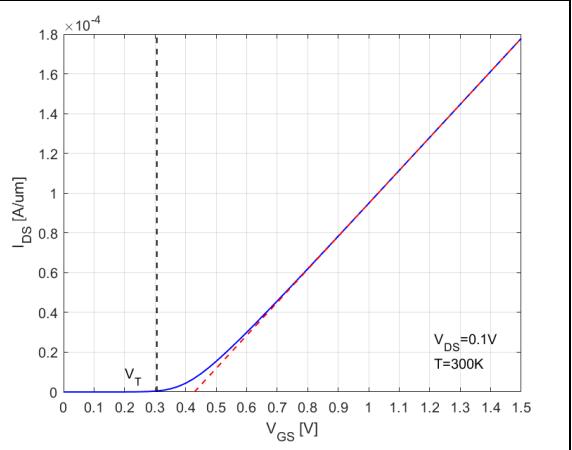
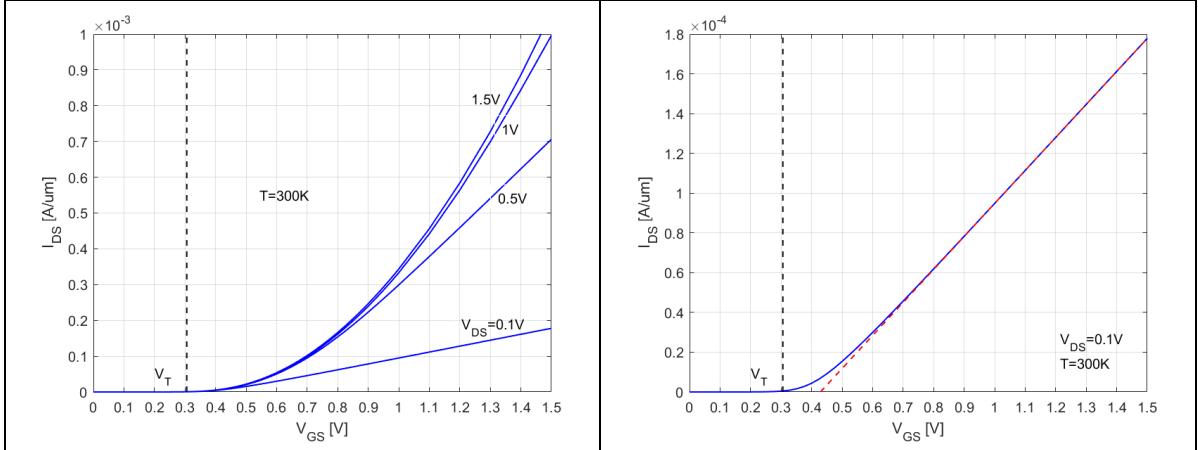


Fig.5 shows the $I_{DS} - V_{GS}$ transcharacteristics of the investigated MOS transistor for different V_{DS} ranging from $0.1V$ to $1.5V$. As expected from the long-channel nature of the device, the increase of V_{DS} does not affect the subthreshold current (pay attention on that the minimum $V_{DS}=0.1V$ is larger than kT/q). That can be easily explained by the fact that the flat part of the band diagram in the source-to-drain direction is not affected by V_{DS} , as shown in Fig.6. The increase of V_{DS} , on the other hand, slightly enlarges the transition region of the bands close to the drain side of the channel, strongly increasing the band bending there. In the ON-state regime, instead, I_{DS} is significantly impacted by the change of V_{DS} . That is yet visible in Fig.5 but can be better appreciated in Fig.7, where a linear scale for I_{DS} is adopted. As studied during lessons, on a linear scale, I_{DS} starts to grow above V_T approximately following, first, the quadratic trend with V_{GS} of the saturation regime $I_{DS}=I_{DS}^{sat}=\mu_n C_{ox} W/L * (V_{GS}-V_T)^2/2m$; when V_{GS} rises above V_T+mV_{DS} , then, the linear trend with V_{GS} of the ohmic/parabolic regimes $I_{DS}=\mu_n C_{ox} W/L * [(V_{GS}-V_T)*V_{DS}-mV_{DS}^2/2]$ appears. From Fig.7, it is clear that, for each V_{DS} , the linear trend detaches from the quadratic curve as its tangent at the V_{GS} value determining the transition from the saturation regime to the ohmic/parabolic regimes. In this regard, it is worth pointing out that considering V_T+mV_{DS} for the latter V_{GS} value and using the analytical expression for I_{DS} in the saturation regime are, of course, approximations. Remember, in fact, that not only the saturation regime but also the onset of saturation cannot be carefully described by the gradual-channel approximation adopted to come to the formulas for I_{DS} . The channel-length modulation effect impacting I_{DS} approaching and entering saturation, for instance, cannot be described following a gradual-channel approximation analysis. Besides, the formulas for I_{DS} were obtained under the hypothesis that V_{DS} is smaller than $2/\phi_B$ and that introduces some errors even in the ohmic/parabolic regimes at high V_{DS} . Another important point worth mentioning is that the value of V_T to be used in the formulas for I_{DS} must be increased a little bit with respect to what predicted by the analytical definition of threshold-voltage of an MOS system. As discussed in the “*Simulation Lab#7 – Long-channel MOS transistor in the ON-state: impact of V_{DS}* ”, this is due to the gradual (and not abrupt) increase of the inversion-layer capacitance entering strong-inversion, which results in a linear growth of the inversion charge Q_{inv} with V_{GS} only a few kT/q above V_T . In order to prove this point, Fig.8 shows the extrapolation down to the horizontal axis of the linear part of the simulated transcharacteristic of the device for $V_{DS}=0.1V$. Since the linear part of the curve corresponds to the ohmic/parabolic regimes and V_{DS} is low, the gradual-channel approximation and the hypothesis that V_{DS} is smaller than $2/\phi_B$ are surely valid. That allows to use the analytical formula for I_{DS} and say that the intercept of the linear extrapolation with the voltage axis should correspond to $V_T+mV_{DS}/2$ and from that the V_T value may be extracted. From the simulation results of Fig.8, $V_T \approx 0.37V$ can be obtained, which is about $70mV$ higher than the analytical value $V_T \approx 0.3V$. This is in reasonable agreement with the V_T correction by $80mV$ adopted in the “*Simulation Lab#7 – Long-channel MOS transistor in the ON-state: impact of V_{DS}* ” to reproduce with the analytical formula for I_{DS} the simulated $I_{DS}-V_{DS}$ curves of

the device in the ohmic/parabolic regimes. To better formalize the need for this correction when addressing the formulas for I_{DS} in the ON-state of an *MOS* transistor, V_T is sometimes called V_{ON} in the formulas.



To complete the analysis of the transcharacteristics of the investigated *MOS* transistor, Figs.9-10 show its dependence on T . The semilogarithmic plot of Fig.9 clearly highlights three major changes induced by the increase of T : i) the growth of the baseline arising from the carrier generation processes in the substrate, since these are thermally activated; ii) the degradation of the *STS* (more mV/dec), due to the kT/q term in its definition; and iii) the reduction of V_T , due to the change of ϕ_B and of the energy gap E_G of silicon with T (in that, the change of the effective density of states for the valence band and the conduction band N_c and N_v , and of the intrinsic carrier concentration n_i are also involved). From the linear plot of Fig.10, on the other hand, the change of device transconductance clearly appears. That arises from the degradation of electron mobility. Note that the previous changes with T make I_{DS} increase in the subthreshold regime and slightly above V_T , while they make I_{DS} decrease deep in the ON-state.

Part IV

Experimental labs

REPORT LAB 1 ELECTRON DEVICES

Andreea Colpos, Sofia Cano Castro, Mattia Marinoni, 30th November 2022

Introduction and aim

For this laboratory three high-voltage transistors (after changing the working table because of electromagnetic noise) were employed, each one with its own dimensions (see next table); our goal was to measure every output resistance of every device at different gate voltages. For this experiment, we used the following values of V_{GS} : [1, 1.25, 1.5, 1.75] [V].

Device name	W [μm]	L [μm]
#2	10	10
#3	0.5	0.6
#4	10	0.6

First thing we did was calculating the flat band voltage V_{FB} , knowing that it is equal to $-\phi_{BI}$, and immediately after the threshold voltage V_T , the capacitance of the oxide C_{ox} and the one of the depletion layer C_{dep} ($t_{ox} = 10.5 \text{ nm}$, $N_a = 2 \cdot 10^{17} \text{ cm}^{-3}$).

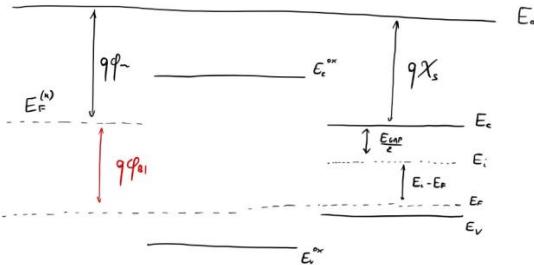


Fig.1: Band diagram and Fermi's levels

$$V_{FB} = -\frac{\left(q\chi_s + \frac{E_{GAP}}{2} + KT \ln\left(\frac{N_a}{n_i}\right) - q\phi_m\right)}{q} = -0.98 \text{ V}$$

$$V_T = V_{FB} + 2|\phi_B| + \frac{\sqrt{2\epsilon_{Si}qN_a2|\phi_B|}}{C_{ox}} = 0.96 \text{ V}$$

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} = 0.328 \frac{\mu\text{F}}{\text{cm}^2}$$

$$C_{dep} = \sqrt{\frac{2\epsilon_{Si}qN_a2|\phi_B|}{4(2|\phi_B|)^2}} = 0.121 \frac{\mu\text{F}}{\text{cm}^2}$$

After performing a total of 12 measurements (4 for each device), we imported the data on Matlab and we've plotted the I-V curves of the different measurements. Then in a single plot we grouped, for each device, the curves obtained with the different gate voltages applied, so it would be easier showing how

they change. Obtained results are supported by the analytical formulas we encountered during theoretical classes, such as the one for the current in saturation regime

$$I_{DS} = \mu_n C_{ox} \left(\frac{W}{L} \right) \left[(V_{GS} - V_T)V_{DS} - \frac{mV_{DS}^2}{2} \right]$$

where $m = 1 + \frac{C_{dep}}{C_{ox}} = 1.37$, a recurrent value.

Knowing that

$$r_0 = \left(\frac{\partial I_{DS}^{sat}}{\partial V_{DS}} \right)^{-1}_{V_{DS} > V_{DS}^{sat}}$$

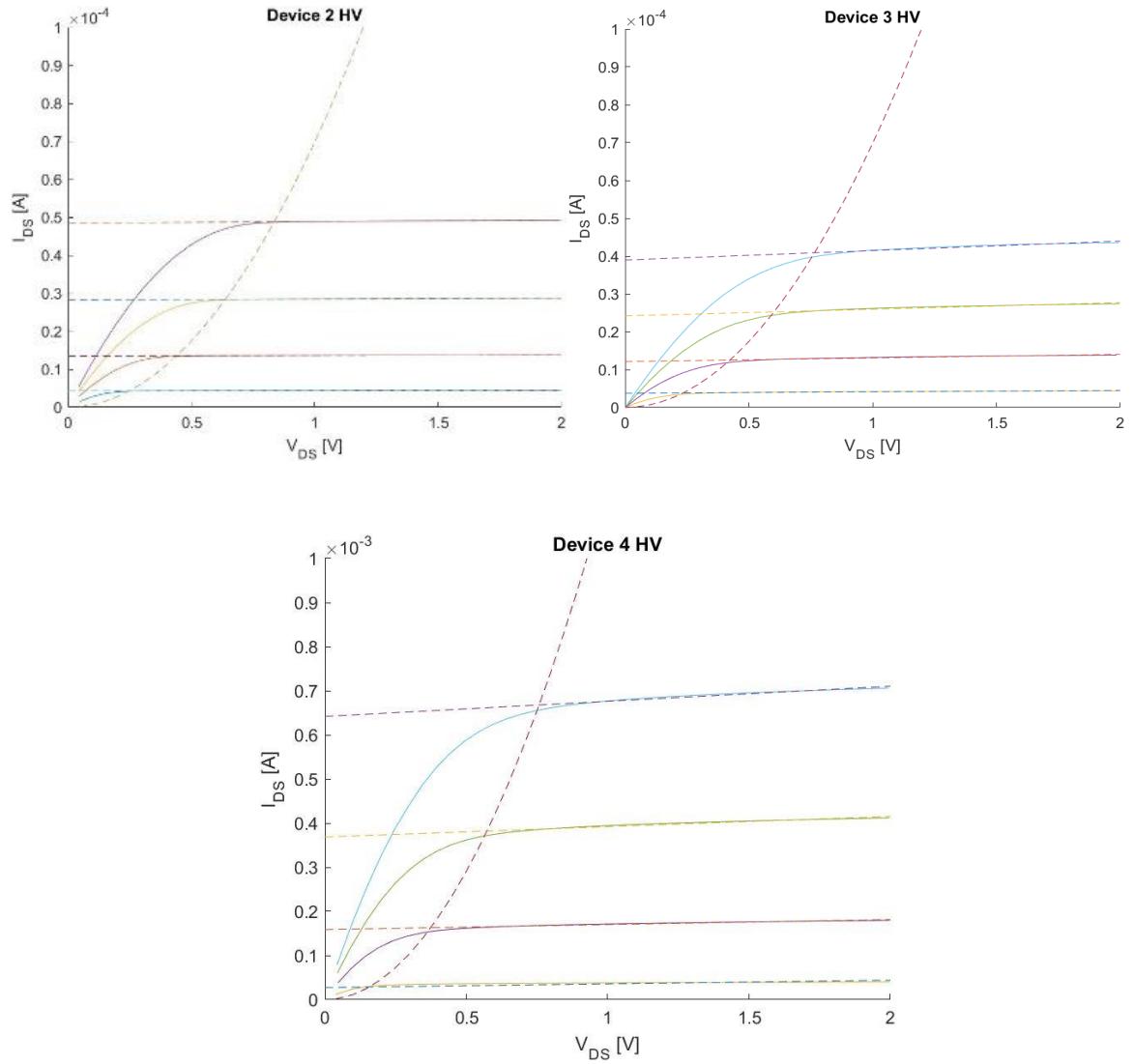
all left to do was just writing a proper Matlab code to elaborate the experimental data.

Plots and results

We managed to obtain the output resistances for every device at the different V_{gs} by exploiting the *polyfit* and *polyval* functions in Matlab; here is a table summarizing what we got:

Device number	V_{gs}	r_0
#2	1 V	14.8 MΩ
#2	1.25 V	7.2 MΩ
#2	1.5 V	4 MΩ
#2	1.75 V	2.6 MΩ
#3	1 V	2.7 MΩ
#3	1.25 V	1 MΩ
#3	1.5 V	568.5 kΩ
#3	1.75 V	399.6 kΩ
#4	1 V	313.7 kΩ
#4	1.25 V	87.6 kΩ
#4	1.5 V	42.8 kΩ
#4	1.75 V	29.2 kΩ

As we can see, the resistance decreases by moving through the saturation regime, this is reasonable and compatible with the plots in fact, what is happening in the device is the following: while V_{gs} increases, the slope of the I-V plot increases, so the resistance that is the reverse of the slope, decreases at higher and higher values of V_{gs} , so when V_{gs} will be at the highest value considered (1.75 [V]), r_0 will be at the lowest value. For all 3 devices we can observe this fact and the V_{gs} in the plots, increase from the bottom to the top.



REPORT LAB 2 ELECTRON DEVICES

Andreea Colpos, Sofia Cano Castro, Mattia Marinoni

Introduction and aim

For this laboratory three low-voltage transistors were employed, each one with its own dimensions (next table); our goal was to measure the STS (SubThreshold Slope) and the DIBL (Drain Induced Barrier Lowering) of each device. For this experiment we used 4 different voltages for V_{DS} : [0.25, 0.5, 0.75, 1] [V].

Device number	W[μm]	L[μm]
#3	1	0.085
#4	10	0.085
#5	1	0.095

The devices have all the $N_a = 1.5 \cdot 10^{18} [\text{cm}^{-3}]$ and same $t_{ox} = 2.2 [\text{nm}]$. We proceeded to calculate the threshold voltage, C_{ox} and C_{dep} .

$$V_T = V_{FB} + 2|\phi_B| + \frac{\sqrt{2\epsilon_{Si}qN_a2|\phi_B|}}{C_{ox}} = 0.43 [\text{V}]$$

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} = 0.157 [\frac{\mu\text{F}}{\text{cm}^2}]$$

$$C_{dep} = \sqrt{\frac{2\epsilon_{Si}qN_a2|\phi_B|}{4(2|\phi_B|)^2}} = 0.354 [\frac{\mu\text{F}}{\text{cm}^2}]$$

After performing a total of 12 measurements, we imported the data on Matlab and we've plotted the I_{DS} - V_{GS} curves of the different measurements. Then in a single plot we grouped, for each device, the curves obtained with the different drain voltages applied, so it would be easier showing how they change. Obtained results are supported by the analytical formulas we encountered during theoretical classes, such as the one for the current in the subthreshold regime:

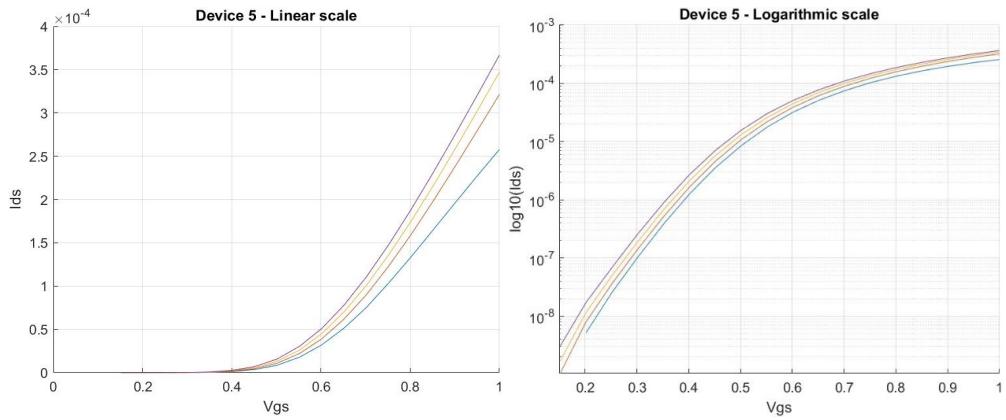
$$I_{DS} = \mu_n C_{ox} \left(\frac{W}{L}\right) (m - 1) \left(\frac{kT}{q}\right)^2 e^{q(\frac{V_{GS}-V_T}{m kT})} \left[1 - e^{-q(\frac{V_{DS}}{kT})} \right]$$

where $m = 1 + \frac{C_{dep}}{C_{ox}} = 1.22$, a recurrent value. The term in blue could be neglected in our analysis since we have a V_{DS} large enough in every case. Theoretically, the expected STS can be calculated as

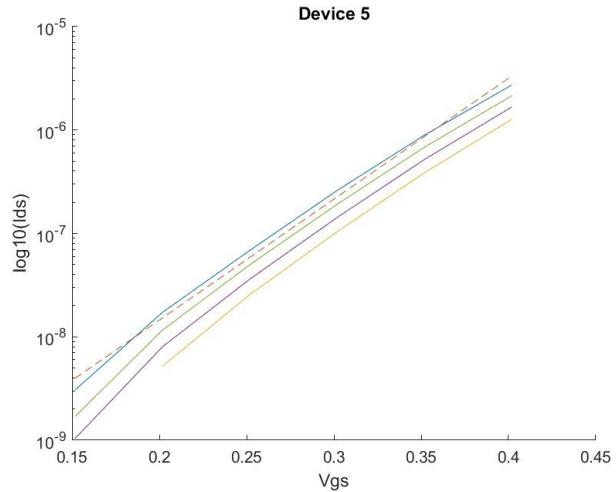
$$STS = \left(\frac{\partial \log_{10} I_{DS}}{\partial V_{GS}} \right)^{-1} = [...] = \frac{kT}{q} \ln(10) m = 73.2 \frac{\text{mV}}{\text{dec}}$$

Plots and results

When we plotted the line in the logarithmic scale, by using the function polyfit, Matlab gave as a result the slope of the line (that is the STS). These are the diagrams for device #5; in the graphs, from the bottom to the top, V_{DS} increases.

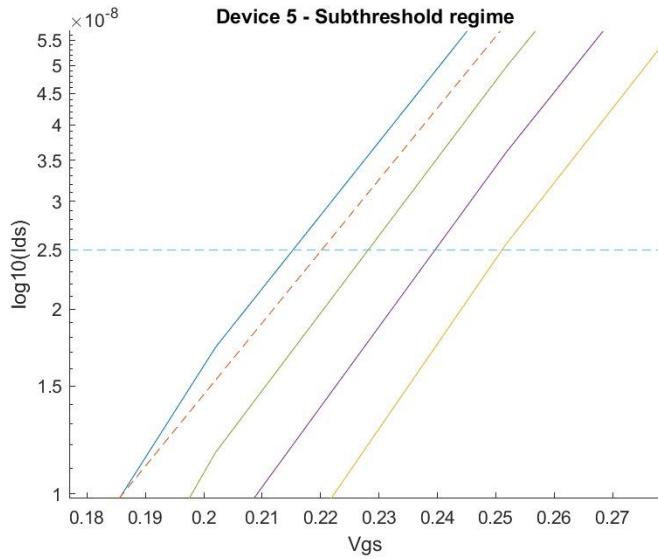


Considering values only in the subthreshold regime we see the obtained fitted line (dotted) which showed a value $STS = 85.4 \frac{mV}{dec}$, comparable to the theoretical one.



The following table shows the intersections between a constant current, arbitrarily chosen as $I_{DS} = 25 nA$, and the plots presented before.

V_{DS} [V]	V_{GS} [V]
1	0.216
0.75	0.227
0.50	0.24
0.25	0.251



Here we can observe how the curves (and consequentially the threshold voltage) shift from right to left, by changing V_{DS} . That is due to the fact that, when the channel is too short like in this case ($L = 95 \text{ nm}$), electrostatics in the horizontal direction becomes relevant (if not overwhelming over the vertical one), thus the gradual channel approximation is no longer valid. If V_{DS} keeps growing, then, the gate-drain band transition becomes so tight that we start to see a peak moving closer to the source and getting progressively lower, incrementing the electrons in the channel; at a certain point, electron concentration becomes so high that the nMOS device is turned on even in subthreshold regime (subthreshold leakage current). Finally, in order to find the shift for the threshold voltage, we have to calculate the difference between the V_{GS} values listed previously; results are reported in the following table.

$\Delta V_{DS} [\text{V}]$	$ \Delta V_T [\text{V}]$
0.5 – 0.25	0.011
0.75 – 0.5	0.013
1 – 0.75	0.011

In this report, we considered only one device (#5) because for the other devices, unfortunately the data from the experiments result to be seriously distorted (the difference between the maximum and the minimum current values was of very few nA) so also the data imported for the elaboration in Matlab were wrong. We elaborated the graphs equally, in the following we present a graph showing the described situation for device #4.

