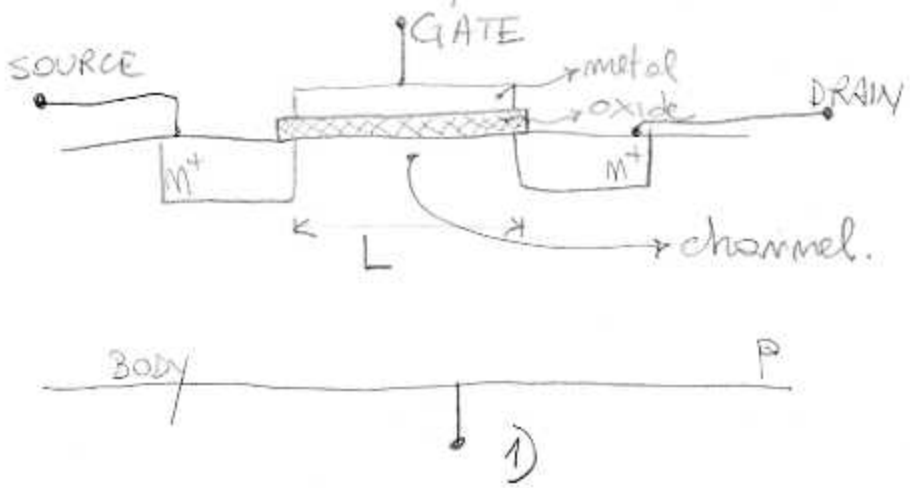
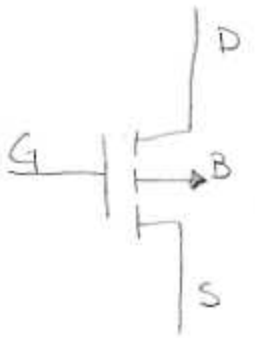
MOSFET: Metal-Oxide-Semiconductor Field-Effect Transistor

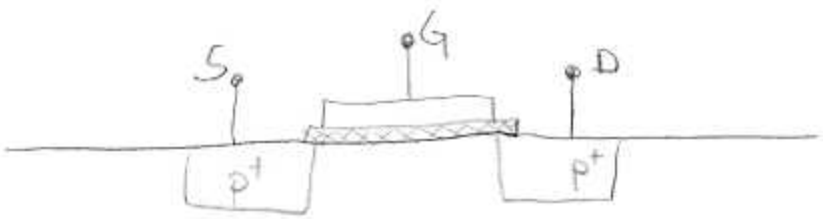
The structure of this transistor is like the following:

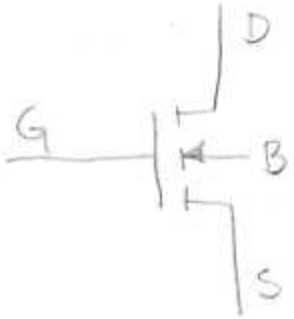


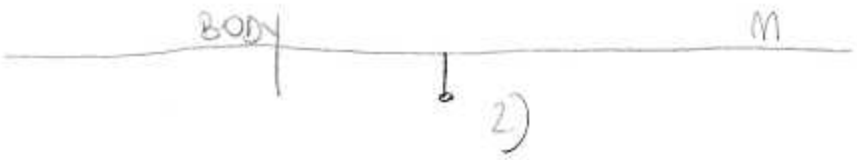


(symbol

or

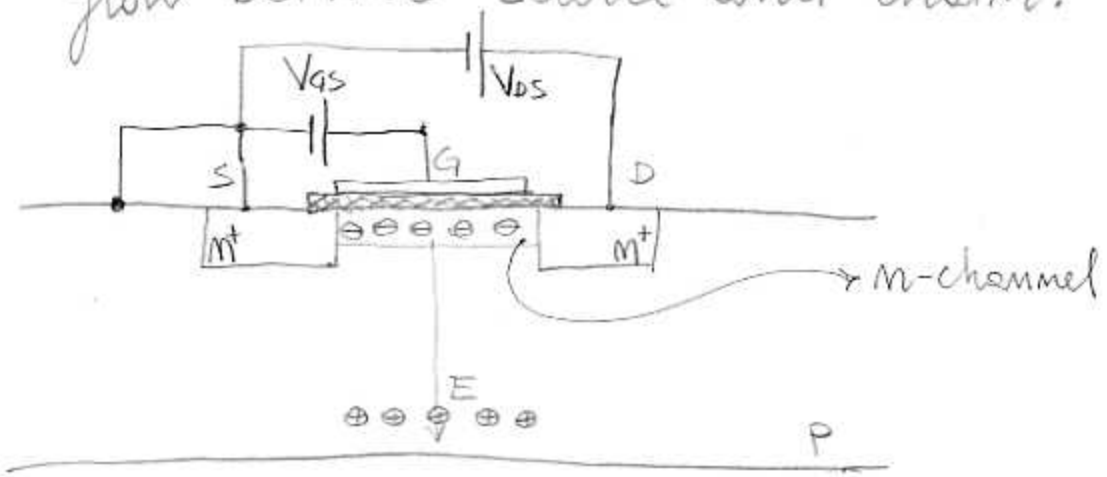






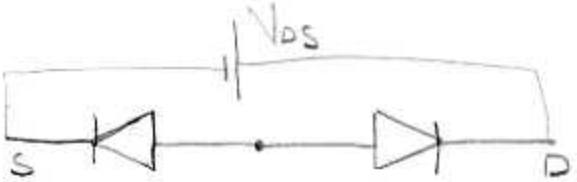
The device in 1) is called n-channel MOSFET and is built on a p substrate. The device in 2) is called p-channel MOSFET and is built on an n substrate.
Usually source and body are connected together and this will be the configuration that we will consider.

MOS is of course referred to the transistor structure while FET is referred to the way it works. We use the electric field effect generated by a voltage across gate and body (or gate and source when body and source are connected together) to allow current to flow between source and drain.



Consider an n-channel MOS where a voltage source is connected between gate and source and drain an source.

If $V_{GS}=0$ then the electric field in the "capacitor" gate-body is zero and there is no effect on the chargers that are present in the substrate. The device can be modeled as two diodes:



This circuit cannot conduct current no matter what is V_{DS} because one of the two diodes will always be OFF.

If $V_{GS} > 0$ then $E \neq 0$ and it's directed from the gate to the body.

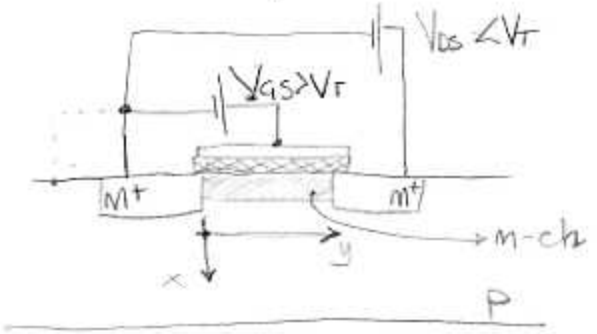
Positive charges then move down in the substrate and negative charges move up towards the channel. This is because a charge is subject to a force $f = qE$.

Notice that in the p substrate there are much more positive charges (holes) than electrons but the electron density is not zero.

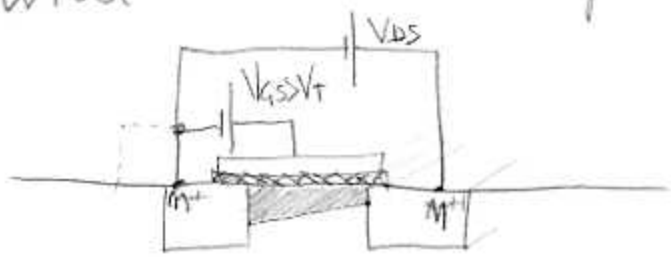
If E is too small, meaning that V_{GS} is too small, the number of electrons in the channel, or equivalently the negative charge density, is very small. The density then would not be sufficient to allow a significant current between source and drain: <u>CUT OFF REGION.</u>

For a certain value of V_{GS}, the number of electrons in the channel becomes sufficient to allow conduction of current between S and D.

The value of V_{GS} that allow conduction is called threshold voltage and denoted by V_T.



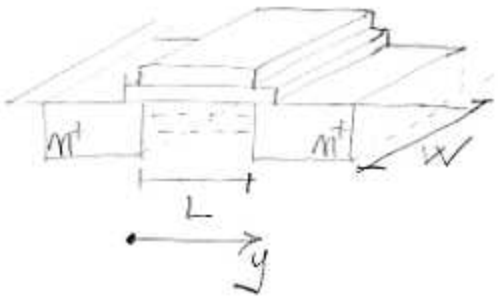
Consider the case where $V_{GS} > V_T$ (hence the channel is formed) and V_{DS} small. If we establish a x,y coordinate system than the potential in the channel is a function of (x,y) and we shall denote it $V(x,y)$. We are interested in the y axes so we will consider only $V(y)$. If V_{DS} is very small, then we can consider $V(y)$ almost constant otherwise $V(y)$ will change with y and the situation will be like the following:



The reason is that the total potential difference across the channel is V_{DS} so the electric field changes depending on y and hence the charge density changes.

For small V_{DS} we can analyze the MOSFET as follows.

The current depends on the charge density and on their speed (because it is the number of charges in the unit time).



$$I_D(y) = W Q_n(y) v(y)$$

This is the linear density of the current at position y.

$Q_n(y)$ is the electrons density at a position y while $v(y)$ is their velocity. Velocity of charges is proportional to the electric field:

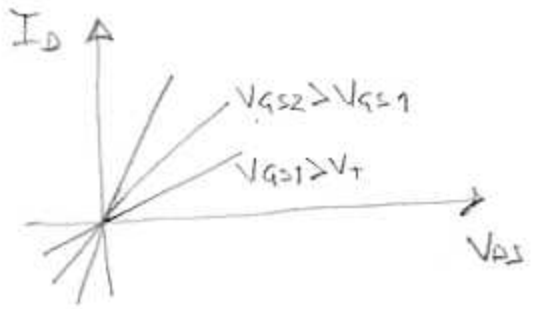
$$v(y) = -\mu_n E(y) = \mu_n \frac{\partial V(y)}{\partial y}$$

So

$$I_D(y) = W Q_n(y) \mu_n \partial V(y)/\partial y$$

We can express $Q_n(y)$ as a function of Voltage by following a simple reasoning. The structure M-O-Si is a capacitor whose capacity can be denoted by C_{ox}. The voltage across this capacitor is the excess gate voltage with respect to V_T

In triode region, the MOSFET behaves like a variable resistor whose resistance depends on the value $(V_{GS}-V_T)$ which is the excess voltage w.r.t. V_T. The resistance of course depends on the phisical properties of the transistor.



We said that in order to be able to conduct current, the voltage difference between the gate and the substrate must be greater than V_T.
at the drain, this voltage difference is

$$V_{GD} = V_{GS} - V_{DS} > V_T \Rightarrow V_{DS} < V_{GS} - V_T .$$

What happens if $V_{DS} \geq V_{GS} - V_T$?
Consider the case $V_{DS} = V_{GS} - V_T \Rightarrow V_{GD} = V_T$
So il like the voltage is just enough to guarantee continuity of the channel. If V_{DS} is a little bit greater than $V_{GS} - V_T$ then there will be a point in the channel where $V < V_T$

on one side and $V(y)$ on the other side. So:

$$Q_n(y) = -C_{ox}[V_{GS} - V_T - V(y)]$$

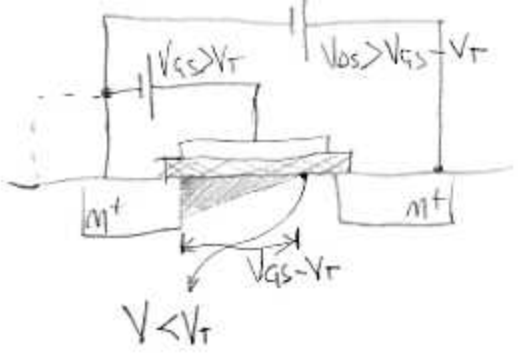
the total current is then the integral of the linear density current along the entire channel:

$$I_D = \int_0^L I_D(y)\,dy = \mu_n W C_{ox} \int_0^{V_D} [V_{GS} - V_T - V(y)]\,dV =$$

$$= \mu_n C_{ox} \frac{W}{L} \left[\left(V_{GS} - V_T - \frac{1}{2} V_{DS}\right) V_{DS} \right] =$$

$$= K \left[(V_{GS} - V_T) V_{DS} - \frac{1}{2} V_{DS}^2 \right]$$

If V_{DS} is very small $\Rightarrow I_D \simeq K[(V_{GS} - V_T)V_{DS}]$. The MOSFET then looks like a resistor between source and drain, whose resistance is equal to $\frac{1}{K(V_{GS} - V_T)}$. When the MOSFET operates in this conditions, we shall say that it operates in TRIODE REGION



P

This phenomenon is called pinch-off. The current cannot increase anymore even if V_{DS} increases, because the voltage between the source and the pinch-off point is constant and equal to $V_{GS}-V_T$. This region of operation is called

SATURATION

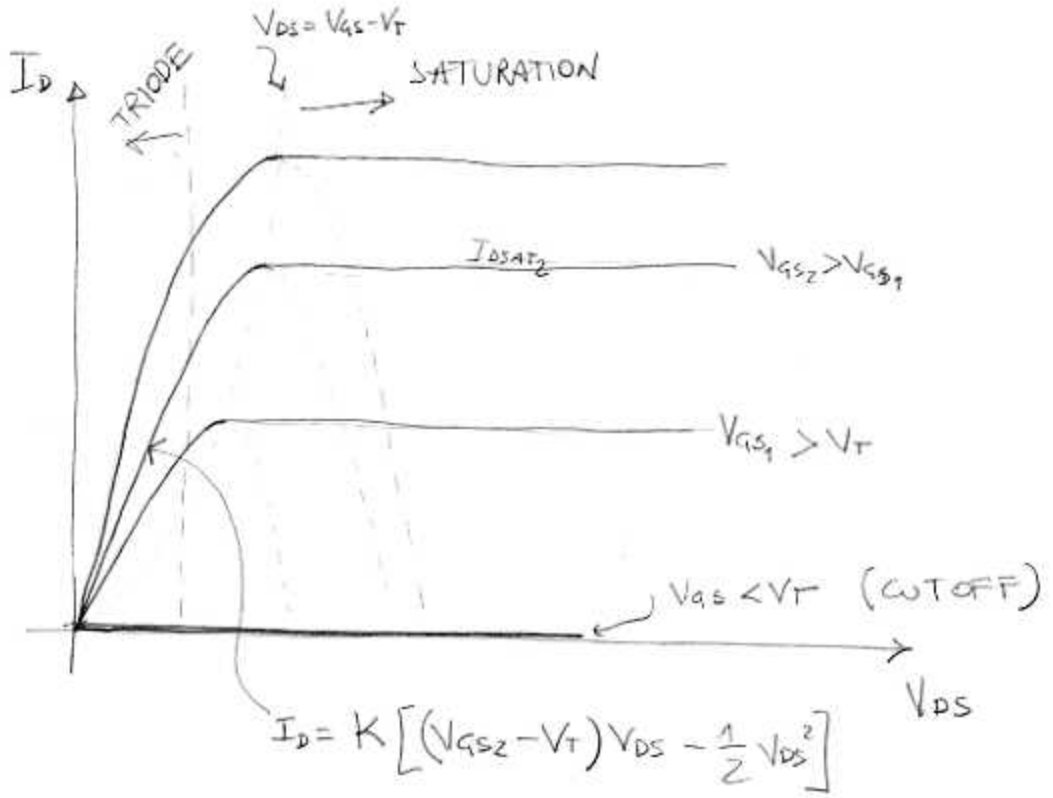
Saturation happens for $V_{DS}=V_{GS}-V_T$ after which the current remains constant to

$$I_{DSAT} = k\left[(V_{GS}-V_T)(V_G-V_T)-\frac{1}{2}(V_{GS}-V_T)^2\right] =$$

Substituting $V_{DS}=V_{GS}-V_T$

$$= \frac{k}{2}(V_{GS}-V_T)^2$$

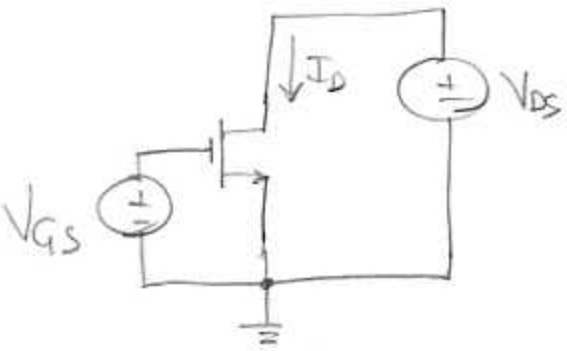
Putting everything together we can plot the I-V characteristic of the mosfet:

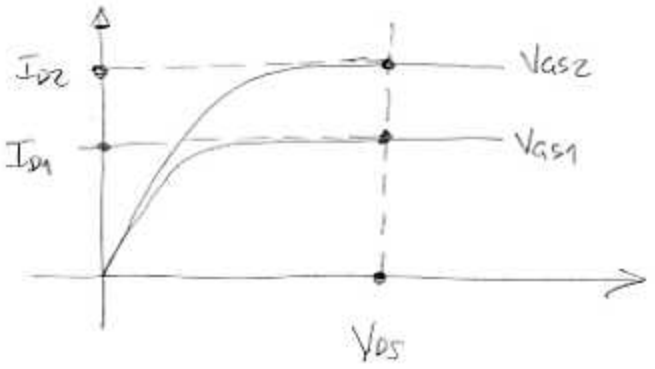


There few things that are important to note. Befor saturation, the current equation is quadratic and if V_{DS} is small then it can be approximated by a line (which is a resistor between drain an source).

In saturation the current doesn't depend on V_{DS}. The saturation current is a function of V_{GS}.

If the transistor is in saturation and we change V_{GS} a little bit, then the the drain current is going to change:





V_{DS} is constat in this example (it is a voltage source). If we change V_{GS} then I_D is going to change and in particular:

$$I_{D1} = \frac{K}{2}\left(V_{GS1} - V_T\right)^2$$

$$I_{D2} = \frac{K}{2}\left(V_{GS2} - V_T\right)^2$$

What we can do is to use a resistor in order to transform a change in current in a change in voltage.

LOAD LINE

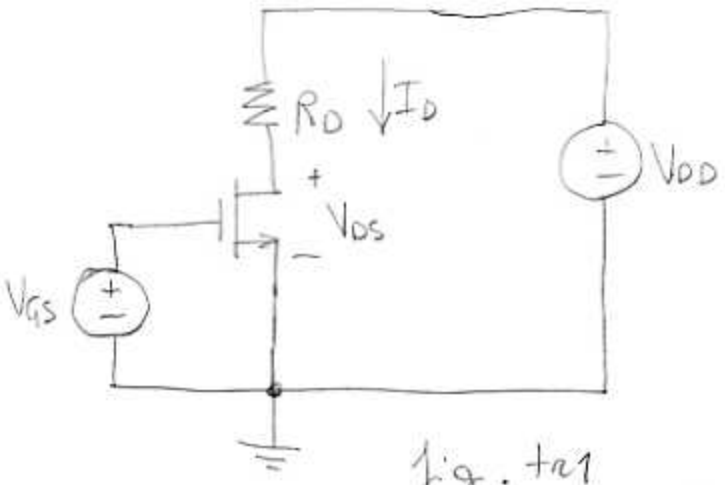


fig. ta1

Now using kVL

$$V_{DD} - V_{DS} - R_D I_D = 0$$

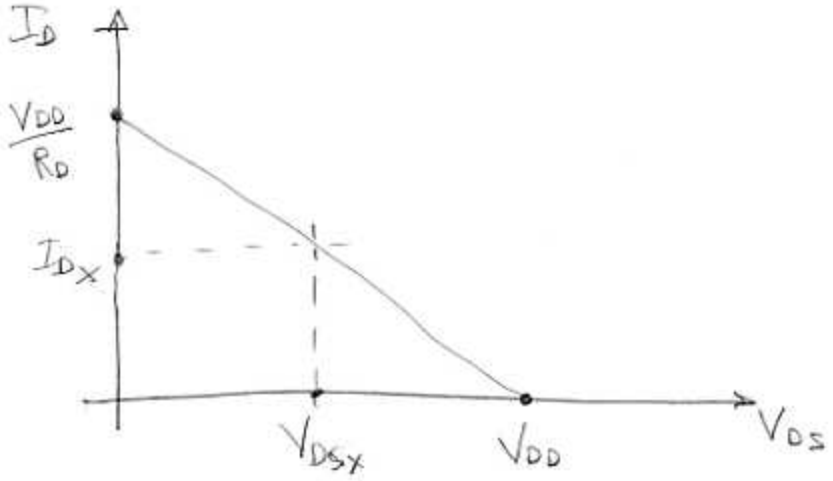
$$\Rightarrow V_{DS} = V_{DD} - R_D I_D \quad (*)$$

This means that now if V_{GS} changes, I_D is going to change and also V_{DS} is going to change (it is not constant anymore because it's equal to V_{DD} minus the voltage drop on R_D).

Equation (*) relate V_{DS} and I_D in the following way:

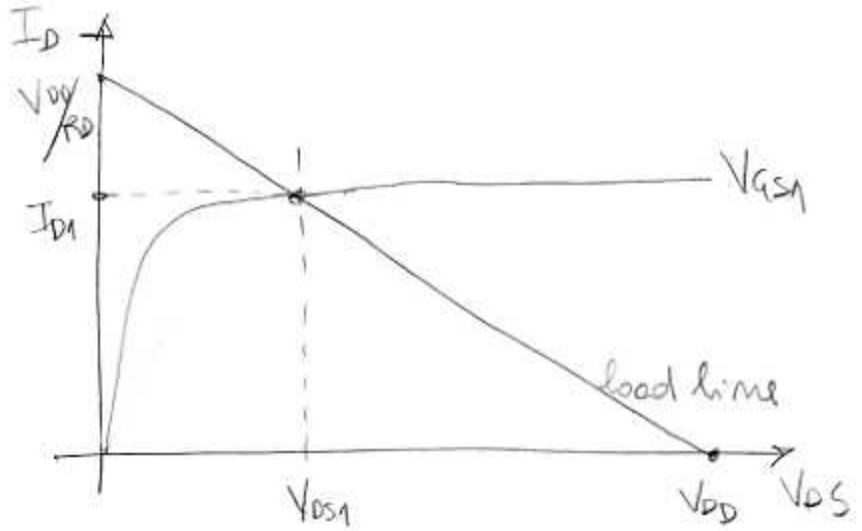
$$I_D = \frac{V_{DD} - V_{DS}}{R_D} = -\frac{V_{DS}}{R_D} + \frac{V_{DD}}{R_D}$$

this equation is a line in the V_{DS}-I_D plane. This line has a negative angular coefficient an intersect the V_{DS} axis for $V_{DS} = V_{DD}$ an the I_D axis for $I_D = V_{DD}/R_D$



This is called the "load line". It gives V_{DS} once we fix I_D.

The transistor V_{DS}, I_D characteristic is going to provide another equation. The intersection of the load line with transistor characteristic gives us the current I_D and the voltage V_{DS}. So basically if we fix a certain V_{GS1} then we select a curve for the transistor in figure tr1.



It say

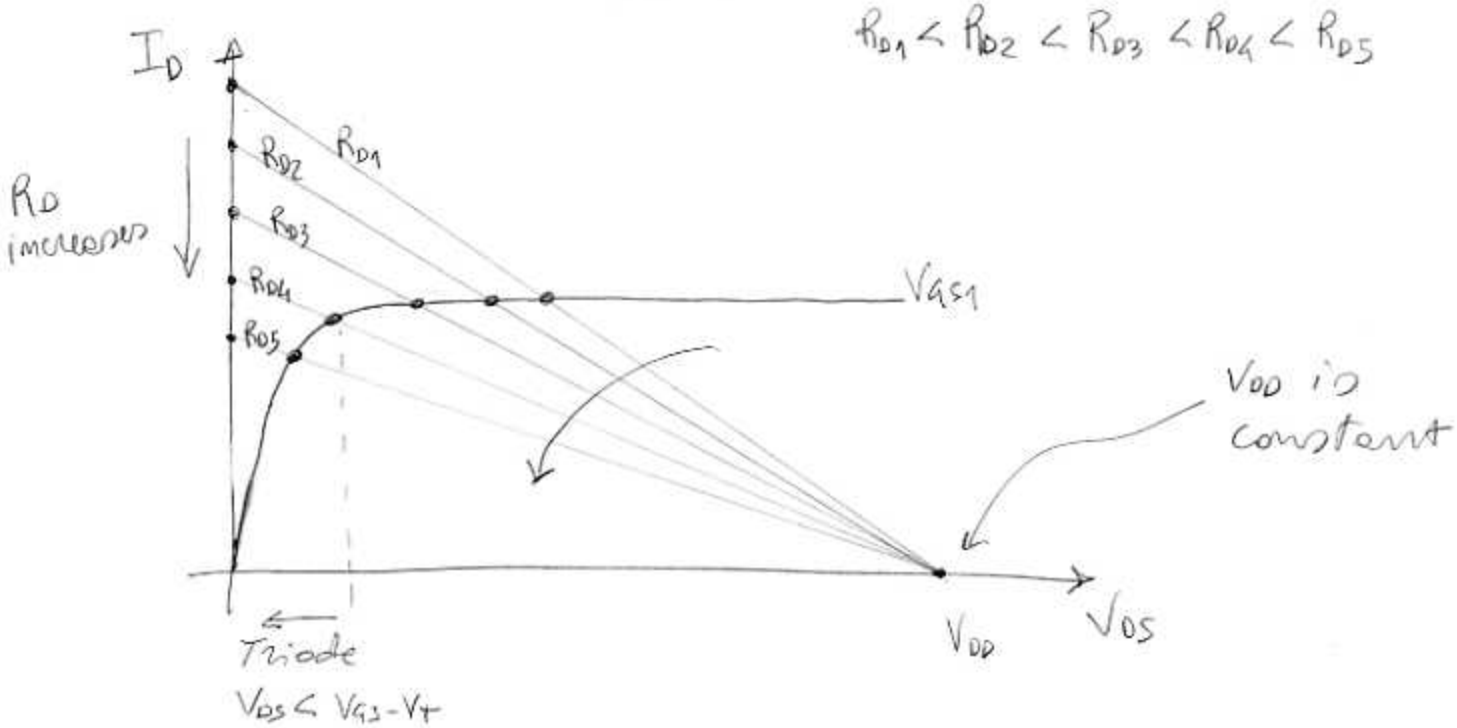
$I_{D1} = \frac{K}{2}(V_{GS1} - V_T)^2$

in saturation.

Now using the intersection with the load line we can know V_{DS1}

What happens if we increase R_D and mantain $V_{GS} = V_{GS1}$?

The transistor characteristic is going to stay the same. Since the transistor was originally in saturation, I_D doesn't change. So the voltage drop $R_D I_D$ will increase and V_{DS} will decrease. When V_{DS} becomes less than $V_{GS1} - V_T$ the transistor switches from saturation to triode. Using the load line, this is easy to verify:



The load line analys is a graphical method to compute V_{DS} and I_D given V_{DD}, R_D and V_{GS}. It is also useful to understand what happens when these three quantities change.

SMALL SIGNAL APPROXIMATION

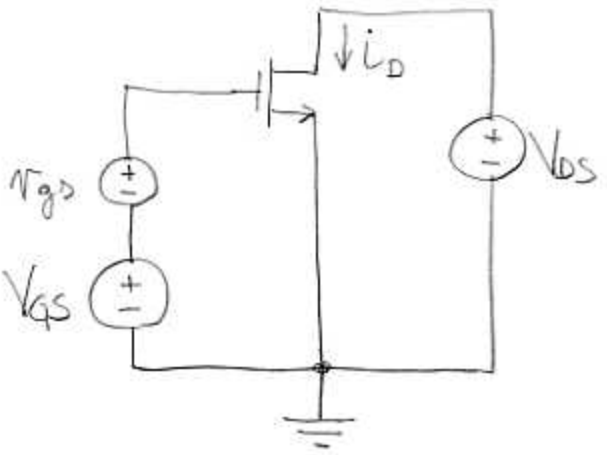
We want to study the behaviour of a transistor in saturation in the case that the voltage between gate and source is the superposition of two voltage: a DC voltage which is constant with time and an AC voltage which is not constant.

I will use the following notation:

V_{GS} upper case letters with upper case subscript for DC quantities

v_{gs} lower case letters with lower case subscript for AC quantities

$v_{GS} = V_{GS} + v_{gs}$ lower case letters with upper case subscript for the sum of both



the gate to source voltage is:

$$v_{GS} = V_{GS} + v_{gs}$$

Consider $v_{gs}=0$, so only the DC voltage sources are present. Also assume that $V_{GS} > V_T$ and $V_{DS} \geq V_{GS} - V_T$ so the transistor is in saturation

Then:

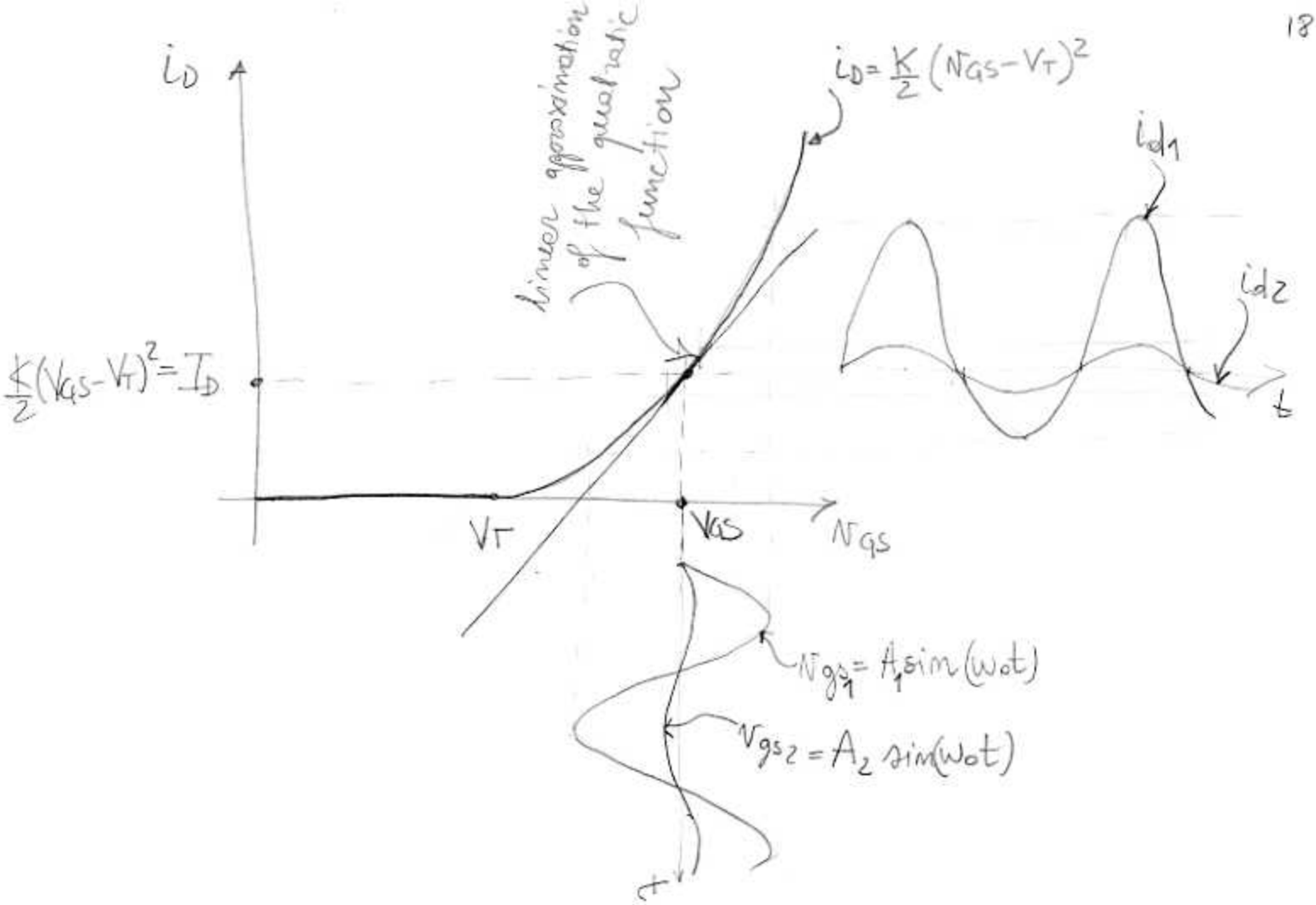
$$I_D = \frac{K}{2}\left(V_{GS} - V_T\right)^2$$

Now let $v_{gs} \neq 0$ then we have

$$i_D = \frac{K}{2}\left(v_{GS} - V_T\right)^2 = \frac{K}{2}\left(V_{GS} - V_T + v_{gs}\right)^2 =$$

$$= \frac{K}{2}\left[\left(V_{GS} - V_T\right)^2 + 2\left(V_{GS} - V_T\right)v_{gs} + v_{gs}^2\right] =$$

$$= I_D + K\left(V_{GS} - V_T\right)v_{gs} + \frac{K}{2}v_{gs}^2 = I_D + i_d$$



The figure represents i_D as function of V_{GS} when the transistor is in saturation

$V_{GS} = V_{GS} + V_{gs}$ and the two components (V_{GS} and V_{gs}) are represented in figure. V_{gs1} is a big signal while V_{gs2} is small. You can see the effect on id. i_{d1} is the drain current when the input is big (V_{gs1}) whereas i_{d2} is the drain currentent when V_{gs} is small (V_{gs2}) you can see how i_{d1} is not a sinusoid but is "distorted" by the quadratic term. i_{d2} is basically a sinusoid and is proportional to V_{gs2}.

for small signal we can neglect v_{gs}^2 meaning

$$K v_{gs}(V_{GS}-V_T) \gg \frac{K}{2} v_{gs}^2 \Rightarrow v_{gs} \ll 2(V_{GS}-V_T)$$, then

$$i_d \simeq K(V_{GS}-V_T) v_{gs} = g_m v_{gs}$$

$g_m \doteq K(V_{GS}-V_T)$ is called transconductance

So basically we have neglected the square term and we obtained a linear model:

$$v_{GS} = V_{GS} + v_{gs}$$

$$i_D = I_D + g_m v_{gs}$$

The behavior of the circuit can the be described as superposition of a DC circuit where all AC sources are set to zero, and an AC circuit where all DC sources are set to zero.

The general analysis then can be divided into two steps:

- DC (consider AC sources equal zero)
- AC (consider DC sources equal zero)

This approximation is called small signal approx. because it holds for

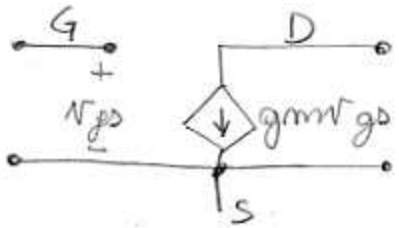
$$v_{gs} \ll K(V_{GS} - V_T)$$

When all DC sources are zero we have:

$$i_d = g_m v_{gs}$$

then the transistor is behaving like a voltage controlled current source.

We can built then the equivalent small signal model for a transistors:



This model is of course linear (remember that we have neglected v_{gs}^2 on purpose) and is valid for AC signals only.
When dealing with transistor then we always distinguish two phases:

- DC analysis (load line)
- AC analysis (use the equivalent circuit and then Node-Voltage analysis)

In the first phase we set all AC sources to 0 and using the load line analysis (or alternatively working out the non linear equations) we compute V_{GS}, V_{DS} and I_D. We call this point the quiescent operating point and denote the quantis with a subscript q: V_{GSQ}, V_{DSQ}, I_{DQ}.

In the second phase we set all DC sources to $\emptyset$, we substitute the transistor with the linear equivalent circuit, and we apply standard techniques for solving linear circuits, namely node-voltage analysis or mesh-current analysis.

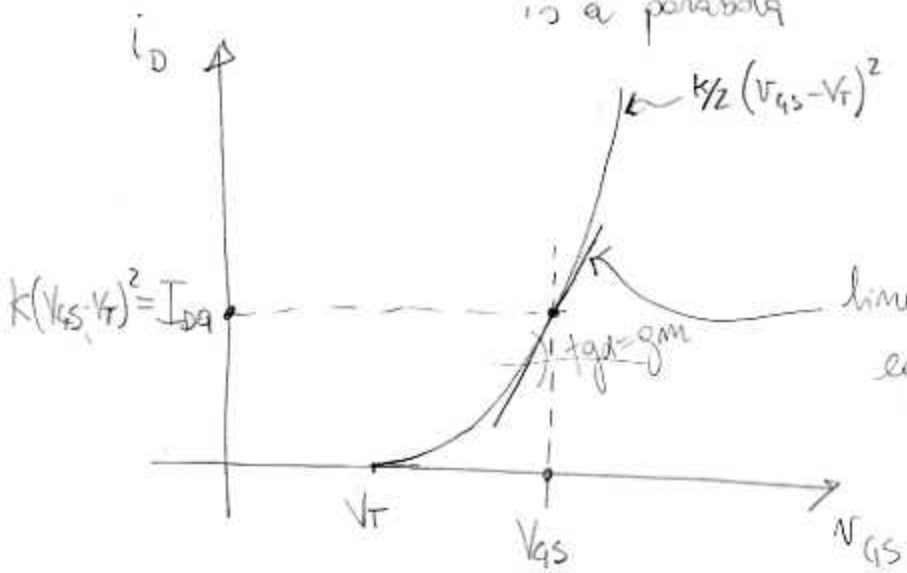
Note that parameters of the small signal equivalent circuit depend on the quiescent point.

This procedure corresponds to the following.

The graph $i_D = f(v_{GS})$ is shown in

figure 1 for the saturation region)

$$i_D = \underbrace{\frac{K}{2}\left(v_{GS} - V_T\right)^2}_{\text{is a parabola}} = \frac{K}{2}\left(V_{GS} - V_T\right)^2 + K\left(V_{GS} - V_T\right)v_{gs} + \frac{K}{2}v_{gs}^2$$



linear approximation of the equation for v_{gs} small

$$g_m = \left.\frac{\partial i_D}{\partial v_{GS}}\right|_{V_{GS}}$$