The Field-Effect Transistor

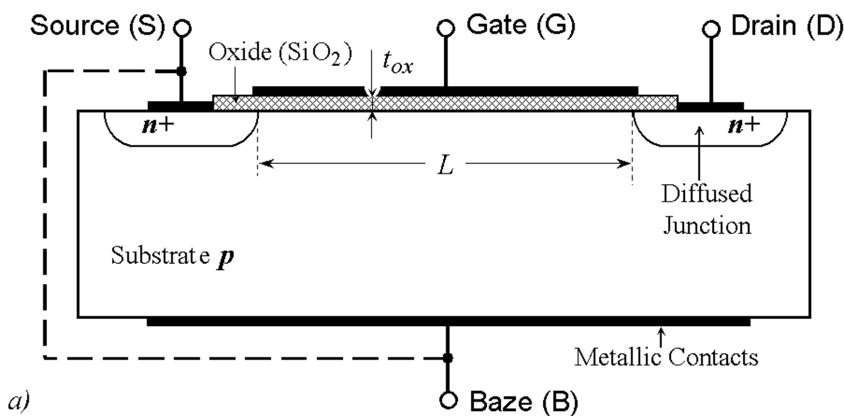
The basic principle involved is the use of the voltage between two terminals called gate and source to control the current flowing through a pair of terminals, drain and source, connected to both ends of a channel. The field effect transistor (FET) acts basically as a voltage-controlled current source.
FET are also called unipolar transistors because current is conducted by majority carriers (electrons or holes) flowing through one type of semiconductor (n-type for n-channel FETs and p-type for p-channel FETs) in contrast to bipolar transistors (where current passes through n-type and p-type semiconductor materials in series).
There are two main types of FETs: the junction FET (JFET) and the metal-oxide semiconductor FET (MOS-FET).

Metal Oxide Semiconductor Field Effect Transistors

The gate of the MOSFET is insulated from the channel by a silicon dioxide layer.
MOS transistors are either of the p-channel or the n-channel type.
MOSFET can be either depletion or enhancement type. The enhancement type MOSFET is the most widely used electronic device (in integrated circuits or as a discrete device).

Structure and Physical Operation

An n-channel enhancement type MOSFET is presented in the next figure.



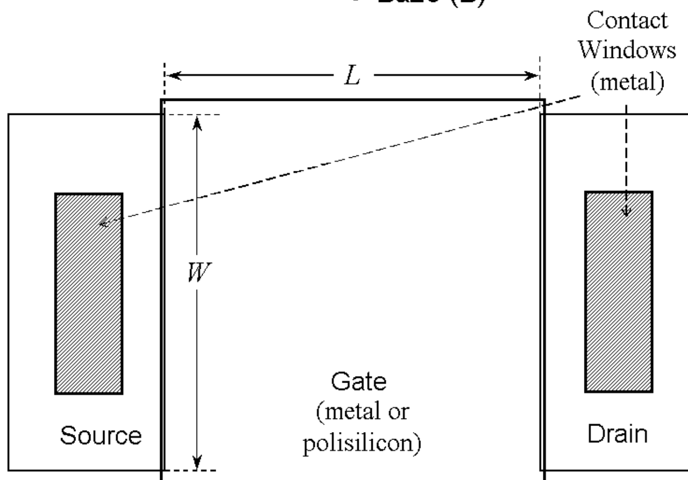


Figure – The physical structure of an enhancement type, *n*-channel MOSFET:

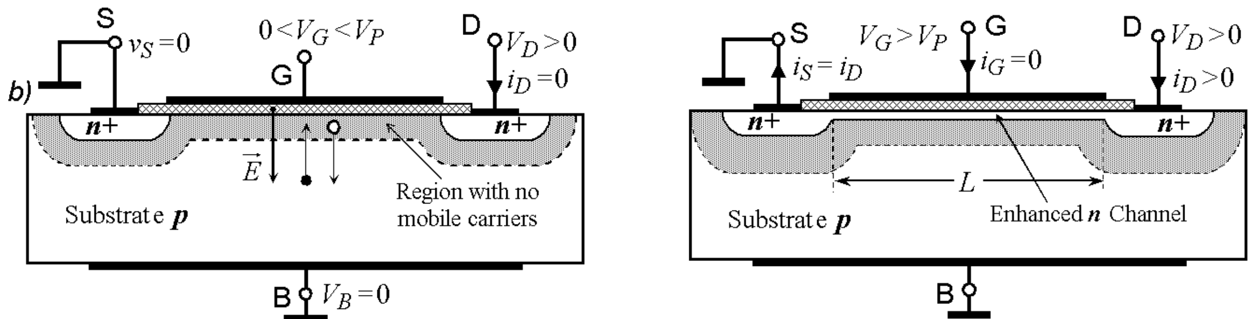
a) Section,

b) Upper view.

The typical TEC-MOS parameters are:

- Oxide thickness, t_{ox}=0,02…0,1 μm;
- Channel length, L=1…10μm;
- Channel width, W=2…500μm.

The n-channel transistor is formed by two heavily doped n^+ regions diffused into a lightly doped p material called the substrate (or bulk). The two n^+ regions (drain and source) are separated by a distance L referred as **the device length**. At the surface (between drain and source) there is a gate electrode that is separated from the silicon by a thin dielectric material (silicon dioxide). The oxide thickness is t_{ox}; W is the MOSFET width.

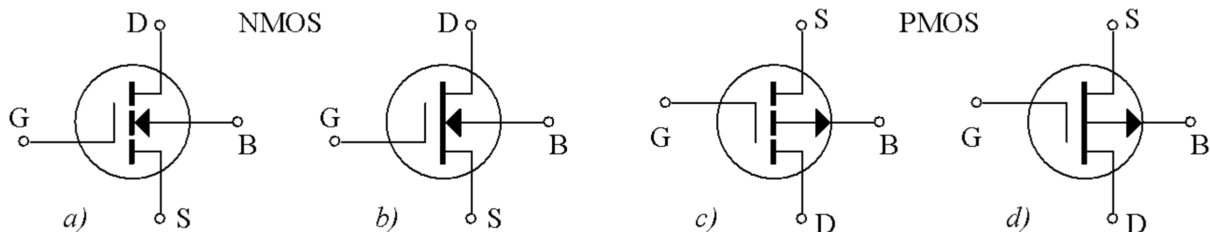


If $v_{GS} = 0$, the path from drain to source includes two diodes back-to-back, which means that no drain current can flow, $i_D = 0$. To cause current to flow from drain to source we first have to create an "n-channel". This can be done by applying a positive voltage at the gate: $v_{GS} > 0$. This voltage produces an electrical field that will attract electrons from substrate and cause them to accumulate at the surface under the oxide layer.
The minimum gate-source voltage that creates the n-type inversion layer under the gate is called the threshold voltage V_T.
The conductivity of the channel is enhanced by increasing the gate-to-source voltage (and thus pulling more electrons into the channel).

Circuit symbol for MOS transistors

Here are some hints to analyze the MOS symbols:

- The space between the line representing the gate electrode and the channel denotes the insulating oxide layer;
- The channel is represented by a thick line either as a broken line (for enhancement devices, where no channel exists –device is normally off) or as a normal (filled) line for depletion devices, channel exists without bias –device is normally on);
- The gate terminal is drawn closer to the source;
- The arrow at the substrate (body) line points in the direction of the substrate-to-channel *pn* junction (and indicate the type of the device).
- The circle indicates the case (for discrete devices).



In many applications the substrate is electrically connected to the source (this is the usual case for the discrete MOS transistors).

Static Characteristics

On the left part of figure there are represented the output characteristics (there are few characteristics for few input voltages: V_{GS}). The saturation voltage $V_{DSsat}=V_{GS}-V_T$, separates the two regions of each output characteristic in a resistive region (or triode region, where the output current, i_D depends on the output voltage, v_{DS}) and a saturation region (or pinch-off region, where the drain current does not depend on the drain-to-source voltage).
On the right side the transfer characteristics in saturation (for $V_{DS}>V_{GS}-V_T$).

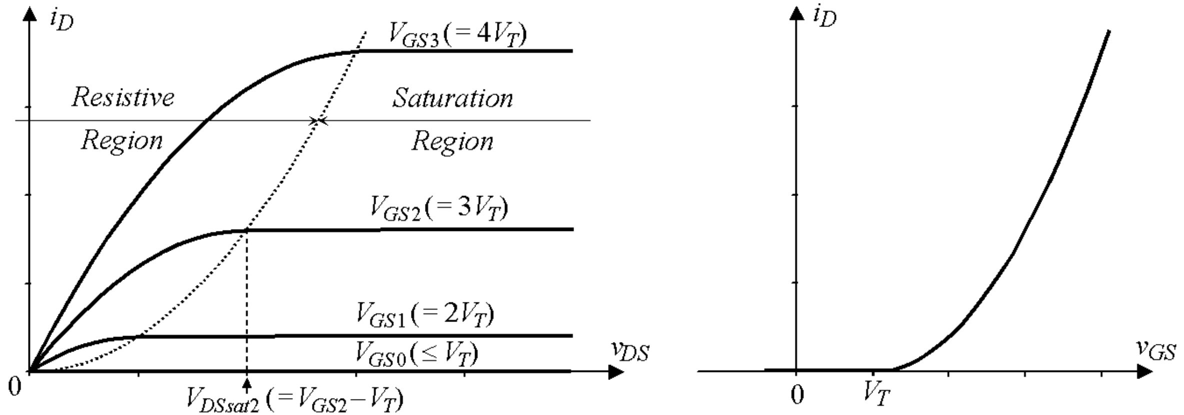


Figure The static characteristics of an enhancement type, n-channel, MOSFET:
a) Drain (output) characteristic, *b)* Transfer characteristics in saturation ($V_{DS}>v_{GS}-V_T$).

The Linear Region Equations

The linear region of the output characteristics is characterized by: $v_{GS}>V_T$, $v_{DS}>0$, $v_{DS}\cong 0$ (more precise: $v_{DS}\ll v_{GS}-V_T$).

The gate capacitance is: $C_g=\frac{\varepsilon_{ox}A}{t_{ox}}=\frac{\varepsilon_{ox}WL}{t_{ox}}=C_{ox}WL$, where $C_{ox}=\frac{\varepsilon_{ox}}{t_{ox}}$ is the oxide capacitance per unit area, ε_{ox} is the permittivity of the silicon oxide and A is the gate area.

The charge of the mobile electrons in channel is: $Q=C_g(V_{GS}-V_T)=C_{ox}\cdot W\cdot L\cdot(V_{GS}-V_T)$.

The transit time for mobile elctrons is: $\tau=\frac{L}{v}=\frac{L}{\mu_n E_l}=\frac{L}{\mu_n\frac{v_{DS}}{L}}=\frac{L^2}{\mu_n\cdot v_{DS}}$, where μ_n is the electron mobility and E_l is the longitudinal electric field.

The electron and hole mobility are related: $\mu_n\cong 2{,}5\mu_p$.

The drain current is: $i_D=\frac{Q}{\tau}=\frac{C_{ox}WL(V_{GS}-V_T)}{L^2}\mu_n v_{DS}=\mu_n C_{ox}\frac{W}{L}\cdot(V_{GS}-V_T)\cdot v_{DS}$, or:

$$i_D=k_n'\frac{W}{L}\cdot(V_{GS}-V_T)\cdot v_{DS}=k_n\cdot(V_{GS}-V_T)\cdot v_{DS};$$

$k_n'=\mu_n C_{ox}$ is the device transconductance parameter, $k_n=k_n'(W/L)$ is the transconductance parameter and W/L is the geometrical factor of MOSFET.

The mobile charge Q connects drain and source. It acts as a resistance connecting drain and source. The actual value of the resistance (given by the shunt resistance under the gate) is: $r_{DS} = \frac{v_{DS}}{i_D} = \frac{1}{k_n' \frac{W}{L}(V_{GS} - V_T)} = \frac{1}{k_n(V_{GS} - V_T)}$.

The non-linear and saturation regions

The model of MOS-FET in the non-linear region can be found by integrating the source to drain current by considering a variable gate to channel voltage through the channel length (for a significant v_{DS}, the voltage decreases from source to drain). The average channel height is reduced, the channel conductance is decreasing and the output characteristics slope is decreasing also. The integration result is: $i_D = k_n \cdot \left(V_{GS} - V_T - \frac{v_{DS}}{2} \right) \cdot v_{DS}$.

The saturation region equation can be found from the previous equation by replacing $v_{DS} = V_{DS\text{sat}} = V_{GS} - V_T$; starting from this point the drain current remains almost constant to:

$$i_D = \frac{k_n}{2} \cdot (V_{GS} - V_T)^2 .$$

That represents the FET transfer characteristic in saturation; it indicates that the FET in saturation is basically a voltage controlled current source.
The practical MOS-FET has a finite output resistance and:

$$i_D = \frac{k_n}{2} \cdot (v_{GS} - V_T)^2 (1 + \lambda \cdot v_{DS})$$

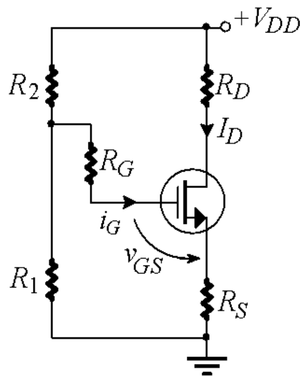
λ is the channel length modulation parameter; it can be computed by $\lambda = 1/V_A$ (the inverse of the Early voltage as it is used by the BJTs). λ depends on the MOS length L; for short channels λ is larger and it determines a low drain resistance: $r_d = 1/(\lambda I_D) = V_A/I_D$.
The channel length modulation effect is usually neglected for the first approximation (paper and pencil) computing.

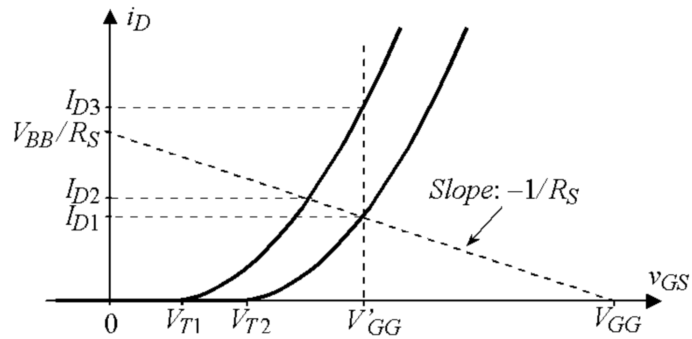
Breakdown and Input Protection

As the voltage of the drain is increased, a value is reached at which the substrate to channel junction breaks down (avalanche breakdown) and cause the current to increase rapidly with v_{DS}.
In the MOSFET another kind of breakdown occurs when the gate-to-source voltage exceeds a certain value (e.g. 50 V); this is the breakdown of the gate oxide and it results a permanent damage of the device results.
Because of the MOSFET very high input impedance, small amount of charge on the C_g can cause the breakdown voltage to be exceeded. To prevent this, gate protection circuits are usually included at the MOS inputs (e.g. clamping diodes) and certain handling precautions should be taken: MOS devices are shipped in conductive foam, the handler’s wrist is connected to earth with a high value resistance, never remove a device from a power-on circuit, do not apply signal to a power-off circuit.

Biasing the Enhancement MOS-FET

The typical bias circuit for a discrete NMOS, is presented in the figure.





The voltage at the gate is given by the voltage divider rule: $V_{GG} = \frac{R_1}{R_1 + R_2} V_{DD}$.

In the absence of R_S (R_S=0), $V_{GS} = V_{GG}$, the drain bias current I_D will be highly dependent on the device parameters (k and V_T).

Including the resistance R_S, the drain bias current is: $I_D = \frac{V_{GG}}{R_S} - \frac{1}{R_S} V_{GS}$ (the oblique line).

On the figure there are presented two possible transfer characteristics (with same k but different V_T – V_{T1} and V_{T2}). In the absence of R_S, the load line is a vertical one (at $V_{GS} = V'_{GG}$) and the current variation is high: from I_{D1} to I_{D3}.
With R_S, the load line is represented by a dashed line with the slope "$-1/R_S$" and the current variation is much less: from I_{D1} to I_{D2}.

Small Signal Operation of the FETs

In the quiescent point (Q point) of the transistor the drain current depends on the drain-source and gate-source voltage: $i_D = f(v_{GS}, v_{GDS})$. For small signall (and low frequency) retaining the first two terms of the Taylor series around Q point:

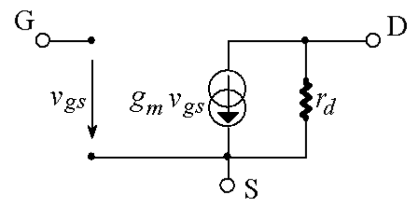
$$i_D = i_D\big|_Q + \left.\frac{\partial i_D}{\partial v_{GS}}\right|_Q \mathrm{d}v_{GS} + \left.\frac{\partial i_D}{\partial v_{DS}}\right|_Q \mathrm{d}v_{DS}$$

The bias current is: $i_D\big|_Q = I_D$ and for small signal $\mathrm{d}v_{GS} = v_{gs}$, $\mathrm{d}v_{GS} = v_{gs}$. Since $i_D = I_D + i_d$ (*dc*+*ac* or bias+signal), the *ac* part of current is:

$$i_d = \left.\frac{\partial i_D}{\partial u_{GS}}\right|_P u_{gs} + \left.\frac{\partial i_D}{\partial u_{DS}}\right|_P u_{ds} = g_m u_{gs} + g_d u_{ds}.$$

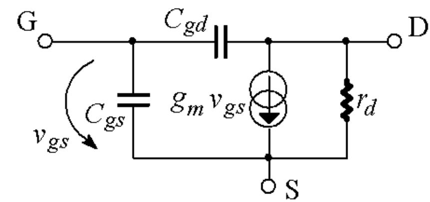
The small-signal low-frequency model is presented in the next figure (it is based on the previous equation and on the fact that the gate current is practically zero).

The two parameters of the FET are:



- The transconductance: $g_m = \left.\frac{\partial i_D}{\partial v_{GS}}\right|_{Q,\, v_{ds}=0}$ and
- The drain (output) conductance: $g_d = \left.\frac{\partial i_D}{\partial v_{DS}}\right|_{Q,\, v_{gs}=0}$.

The small-signal parameters depend on the Q point. For MOSFET in saturation, the parameters are:



$g_m = k'\frac{W}{L}(V_{GS} - V_P) = k \cdot (V_{GS} - V_P)$ and

$g_d = \lambda I_D$, that gives $r_d = \frac{1}{g_d} = \frac{1}{\lambda I_D} = \frac{V_A}{I_D}$.

The previous figure presents the high frequencies model; it results from the low-frequency model completed by two capacitors usually in the fraction of pF to pF range:

- the gate capacitance: C_{gs} and
- a parasitic drain to gate capacitance C_{gs}.

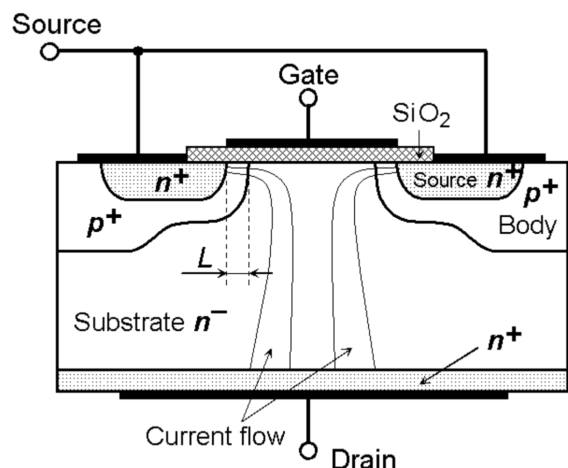
MOS Power Transistors, D-MOS

For an n-channel MOSFET operating in pinch-off, the drain current is:

$$i_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}\cdot(V_{GS} - V_T)^2.$$

To increase the current capability of the MOSFET its width W should be made large and its channel length L should be made as short as possible. The short length produces a reduction of the breakdown voltage. Such a device can not controll high voltages that aare used by power applications; special structures are used.

The most popular structure of a short-channel (1 to 2 micrometers) MOSFET with high breakdown voltage is the double-diffusion or DMOS transistor. The physical structure of a DMOS is presented in figure. The device is realized on a lightly doped substrate (on the top of a heavily doped region used to connect the drain terminal). Two successive diffusion regions are realized: one for the heavily doped body (p-type) and the next one (inside the body) for the heavily doped region for the source (n-type).



A gate voltage $v_{GS} > V_T$ induces a lateral n-channel in the heavily doped p-type body region underneath the gate oxide. Current is then conducted by electrons from the source through the short channel (L) and then vertically down the substrate to the drain.

The breakdown voltage can be very high (e.g. 600 V) because the depletion region between the substrate and body extends mostly in the lightly doped substrate. The MOS transistor has a high current capability (e.g. 50 A) also.

Characteristics of Power MOSFET

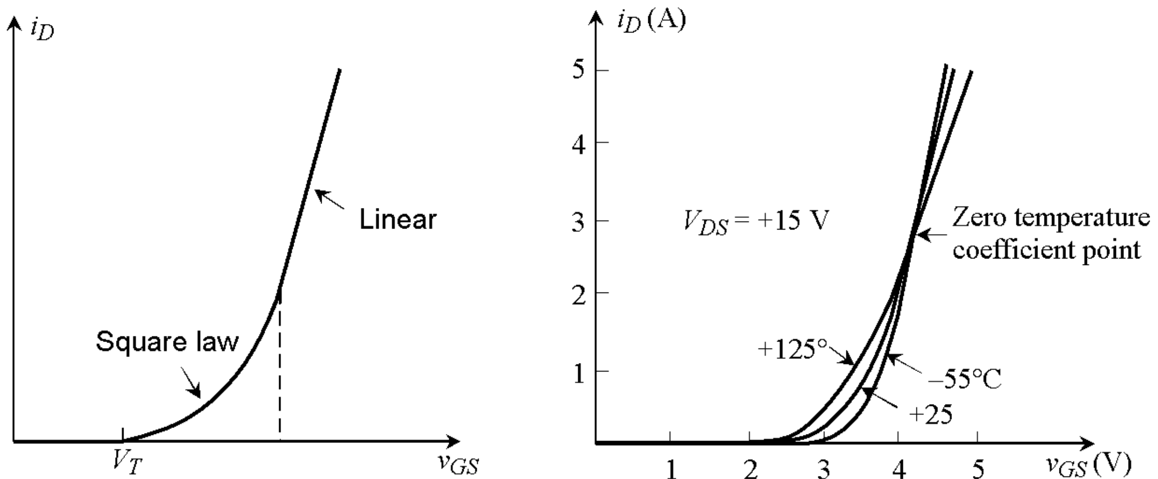
The usual value of threshold voltage is $V_T = 2\ldots4\,\mathrm{V}$ and in pinch-off at relatively small drain currents there is a square law (as for a normal MOS). For high currents, the transfer characteristics becomes linear, because the carrier velocity value, U_{sat}, get saturated to a value of about $5 \cdot 10^6$ cm/s (for electrons in silicon):

$$i_D = \frac{1}{2} C_{ox} W \cdot U_{sat} (v_{GS} - V_P).$$

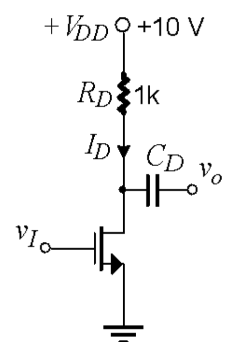
In the linear region the transconductance would be constant: $g_m = \frac{\partial i_D}{\partial v_{GS}} = \frac{1}{2} C_{ox} W \cdot U_{sat}$.

Its value is given in the transistor data-sheet as a parameter.

In the next figures it is indicated a generic transfer characteristics for a MOS power transistor and in the second figure it is indicated the transfer characteristic family for an actual MOS (type IRF 630 – a 9 A, 200 V NMOS). In this second figure it can be seen the dependence of the transfer characteristics on the temperature. For the given example the thermal coeficient of the threshold voltage V_T is from –3 to –6 mV/°C.



Applications



S12 – P1. *a)* Compute the small-signal, low-frequency voltage gain of the NMOS amplifier in figure for $V_{GS} = 2$ V.

The MOS transistor parameters are: $V_T = 0.7$ V, $W/L = 100$, $k' = 60\ \mu A/V^2$, $\lambda = 1/(L \cdot V_E)$ with $V_E = 4$ V/µm (Early voltage per unit area) and $L = 5$ µm.

b) What is the value of voltage gain if the drain resistance is replaced by a current generator that generates the statical drain current with an infinite internal resistance (ideal current generator).

c) Design a maximum gain amplifier for an active load (circuit from *b*). You can modify: V_{GS}, L and W. (V_T, k' and V_E are constants that depends on technology).

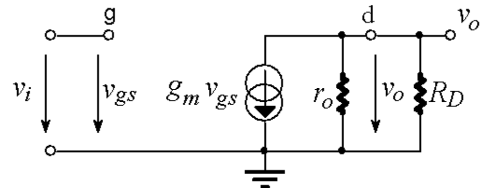
a) In saturation (for $V_{DS} > V_{GS} - V_T$), the drain current is: $i_D = \frac{k'}{2} \cdot \frac{W}{L} \cdot (V_{GS} - V_T)^2$.

The bias point is: $I_D = 30\mu \cdot 100 \cdot (2 - 0.7)^2 = 5.07\,\text{mA}$ and
$V_{DS} = V_{CC} - R_D I_D = 10 - 1k \cdot 5.07m = 4.93\,\text{V}$; $V_{DS} > V_{DSsat} (= 2 - 0.7 = 1.3\,\text{V})$.

The transconductance is: $g_m = k' \frac{W}{L}(V_{GS} - V_T) = 60\mu \cdot 100 \cdot 1.3 = 7.8\,\text{mA/V}$. That value is low, compared to a BJT operated at the same current ($g_{mBJT} = I_C / V_T \cong 40 \cdot 5m = 200\,\text{mA/V}$).

The output resistance is:
$r_o = \frac{1}{\lambda \cdot I_D} = \frac{V_E \cdot L}{I_D} = \frac{4 \cdot 5}{5.07m} \cong 4\,\text{k}\Omega$.

The voltage gain can be computed on the small signal equivalent circuit presented in the next figure:



$A_v = \frac{v_o}{v_i} = -g_m \cdot (R_D \parallel r_o) = -7.8m \cdot 0.8k = -6.2$ (much lower than the voltage gain of a BJT).

b) With an ideal current source instead of the drain resistance, the voltage gain is greater: $A_v = -g_m \cdot r_o = -7.8m \cdot 4k = -30.8$.

c) We recomputed the gain literally (for the ideal current generator case):

$$A_v = -g_m \cdot r_o = -k' \frac{W}{L}(V_{GS} - V_T) \cdot \frac{V_E \cdot L}{\frac{k'}{2}\frac{W}{L}(V_{GS} - V_T)^2} = -\frac{2 \cdot V_E \cdot L}{V_{GS} - V_P}.$$

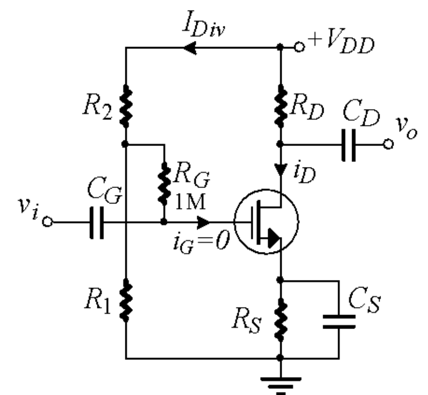
One can see that the voltage gain increases with L and decreases with $V_{GS} - V_T$. As an example for $L = 20\,\mu\text{m}$ and $V_{GS} - V_T = 0.3\,\text{V}$:

$A_v = -\frac{V_E \cdot L}{V_{GS} - V_P} = -\frac{2 \cdot 4 \cdot 20\mu}{0.3} = -533$. In the mean time the current decreases dramatically:

$I_D = 30\mu \cdot 25 \cdot (0.3)^2 = 67.5\,\mu\text{A}$ $g_m = 60\mu \cdot 25 \cdot 0.3 = 0.45\,\text{mA/V}$

and $r_o = \frac{4 \cdot 20}{67.5\mu} \cong 1.185\,\text{M}\Omega$.

S13 – P1. Consider an enhancement MOSFET with $V_T = 2\,\text{V}$ and $I_D = 0.25\,\text{mA}$ at $V_{GS} = 2\,\text{V}$.



a) Design the network that bias the device at using the biasing arangement of figure with $V_{DD} = 20\,\text{V}$. Consider $V_S = 4\,\text{V}$ and $I_{Div} = 0.1\,\text{mA}$. Assume the maximum swing of $\pm 4\,\text{V}$ is required at the drain.

b) What is the voltage gain A_v for $\lambda = 0.025\,\text{V}^{-1}$.

c) Find the percentage change in I_D if the FET is replaced by another having the same k value but $V_T = 3\,\text{V}$.

a) In saturation (for $V_{DS} > V_{GS} - V_T$), the drain current is: $i_D = \frac{k}{2} \cdot (V_{GS} - V_T)^2$.

The transconductance parameter can be computed from the FET data:

$$k = \frac{2I_D}{(V_{GS} - V_T)^2} = \frac{2 \cdot 0.25m}{1} = 0.5\,\text{mA/V}^2.$$

In the required bias point: $V_{GS} = V_T + \sqrt{\frac{2I_D}{k}} = 2 + \sqrt{\frac{2 \cdot 4m}{0.5m}} = 4\ \text{V}$ and $R_S = \frac{V_S}{I_D} = \frac{4}{1m} = 4\,\text{k}\Omega$.

The gate current is zero, $I_{Div} = i_G + I_1 = I_1$ and $R_1 + R_2 = \frac{V_{DD}}{I_{Div}} = \frac{20}{0.1m} = 200\,\text{k}\Omega$.

$V_G = V_{GS} + V_S = 8\ \text{V}$ and $V_G = \frac{R_1}{R_1 + R_2} V_{DD}$ gives: $R_1 = (R_1 + R_2) \cdot \frac{V_G}{V_{DD}} = 200k \cdot \frac{8}{20} = 80\,\text{k}\Omega$.

$v_{DS} > V_{DSsat} = V_{GS} - V_T = 2\,\text{V}$, $V_{D\min} = V_S + V_{DSsat} + \hat{V}_o = 4 + 2 + 4 = 10\,\text{V}$,

$V_{D\max} = V_{DD} - \hat{V}_o = 20 - 4 = 16\,\text{V}$. The corresponding drain resistances (limits) are:

$R_D = \frac{V_{DD} - V_D}{I_D} = 4...10\,\text{k}\Omega$. For the maximum R_D we get the maximum voltage gain.

b) The ac equivalent circuit is similar with the one from the previous problem (except the resistance in the gate – that are not relevant for this problem). Based on this schematic the voltage gain can be computed as: $A_v = -g_m (R_D \parallel r_o)$. The MOS parameters are:

$g_m = \left.\frac{di_D}{dv_{GS}}\right|_Q = k(V_{GS} - V_T) = 0.5m \cdot 2 = 1\,\text{mA/V}$; $g_d = \lambda \cdot I_D = 0.025\,\text{mA/V}$; $r_d = \frac{1}{g_d} = 40\,\text{k}\Omega$ and the voltage gain is: $A_v = -1m \cdot 8k = -8$ (the minus sign indicates an inverting amplifier).

c) From the transfer characteristic and KVL on gate loop, we get:

$V_G - V_{GS} = \frac{k}{2} R_S (V_{GS} - V_T)^2$, $V_{GS}^2 - 5 \cdot V_{GS} + 1 = 0$, $V_{GS} = \frac{5 \pm \sqrt{25 - 4}}{2} = 4.8\,\text{V}$; The 2nd solution is $0.21\ \text{V}$ is not possible. The current is: $I_D = \frac{0.5m}{2} \cdot (4.8 - 3)^2 = 0.81\,\text{mA}$ ($\varepsilon = -19\,\%$).