MOSFET Characteristics- Theory and Practice

Debapratim Ghosh
deba21pratim@gmail.com

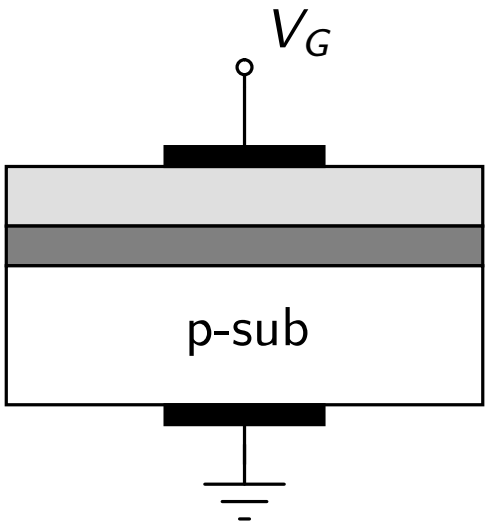
Electronic Systems Group
Department of Electrical Engineering
Indian Institute of Technology Bombay
20 September 2012

Introduction- the MOSFET

- **M**etal **O**xide **S**emiconductor **F**ield **E**ffect **T**ransistor
- The name describes nearly everything about the device itself.
- The first three words **M**etal **O**xide **S**emiconductor describes the layer-wise structure of the device.
- The last three words **F**ield **E**ffect **T**ransistor describes the principle of operation.

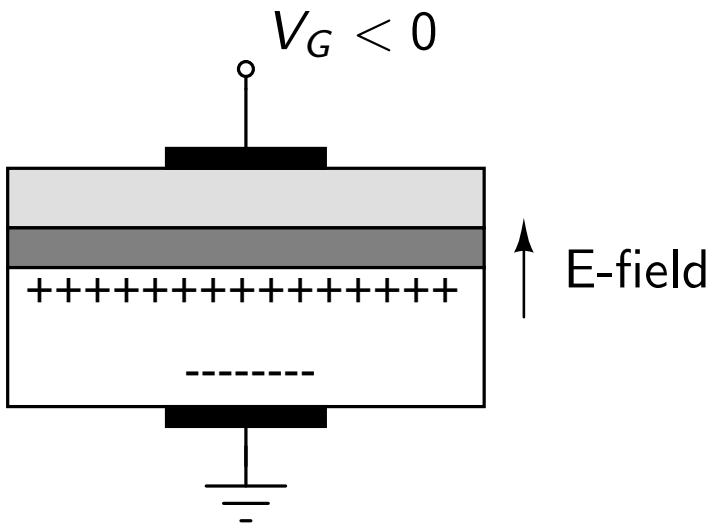
MOS Structure Physics

- MOS transistors can be of two types- NMOS and PMOS.
- An NMOS has a lightly doped p-substrate (where there is scarcity of electrons).
- The metal terminal is called the Gate.
- The oxide layer (usually SiO_2) is an insulator.
- The p-type substrate is grounded while the gate voltage V_G is varied.
- We will see how the MOS structure behaves as V_G is varied.



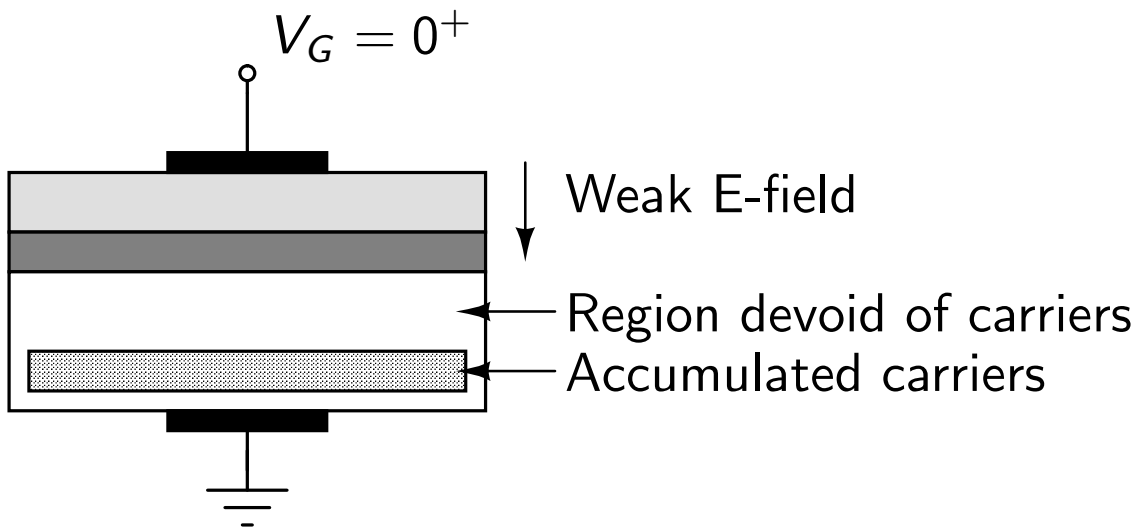
MOS Structure Physics- Accumulation

- Let us apply a negative gate voltage i.e. $V_G < 0$.
- This negative V_G sets up an electric field through the oxide.
- The electrons (minority carriers) are pushed away towards ground, and the holes (majority carriers) are pushed towards the oxide.
- The region below the oxide is now devoid of n-type charge carriers.
- This region of operation is called **accumulation region**.



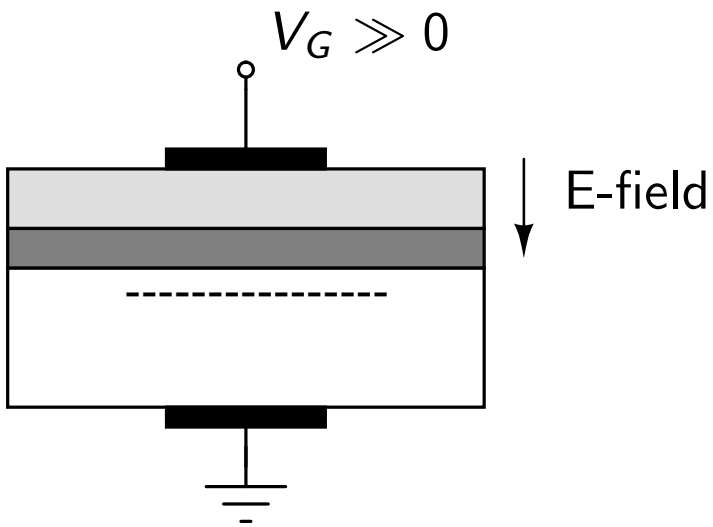
MOS Structure Physics- Depletion

- Let us apply a small positive gate voltage.
- This small V_G sets up a weak electric field though the oxide.
- The holes are now pushed away from the oxide, deep into the substrate.
- However, the electric field is too weak to pull all the minority electrons towards the oxide.
- At this time, the immediate region below the oxide is devoid of any mobile charges (electrons or holes).
- This region of operation is called **depletion region**.



MOS Structure Physics- Inversion

- Let us now increase the gate voltage V_G .
- As V_G increases, the electric field becomes stronger, and the minority electrons accumulate below the oxide.
- At a certain value of V_G , the concentration of moblie electrons becomes so high that the region just below the oxide becomes as n-type as the rest of the substrate is p-type.
- This region of operation is called **inversion region**.
- These accumulated electrons can now be used to generate a current.



MOS Threshold Voltage

To accumulate mobile electrons below the oxide, the gate voltage VG has to be sufficiently high to cross the "threshold". This is governed by a number of factors.

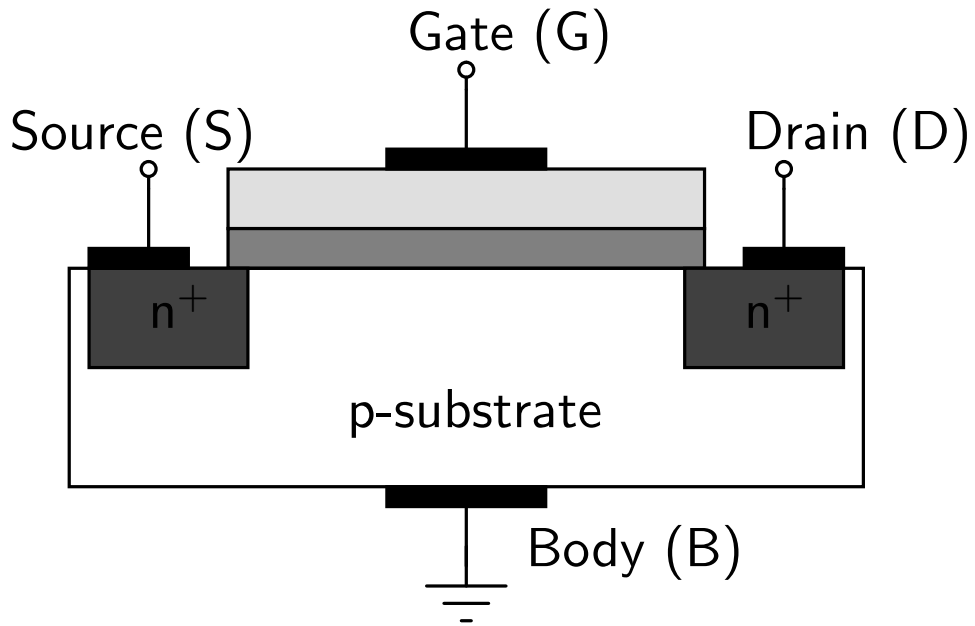
- The work-function difference between the gate and the silicon substrate (this leads to the "flatband" voltage).
- The gate voltage component required to bring about surface inversion (surface just below the oxide).
- The concentration of acceptor ions in the substrate.
- The concentration of trapped charges inside the oxide.
- The substrate voltage (so far weve assumed it to be ground).

Note:

- The threshold voltage V_{TN} for NMOS is positive.
- The threshold voltage V_{TP} for PMOS is negative.

The MOS Transistor

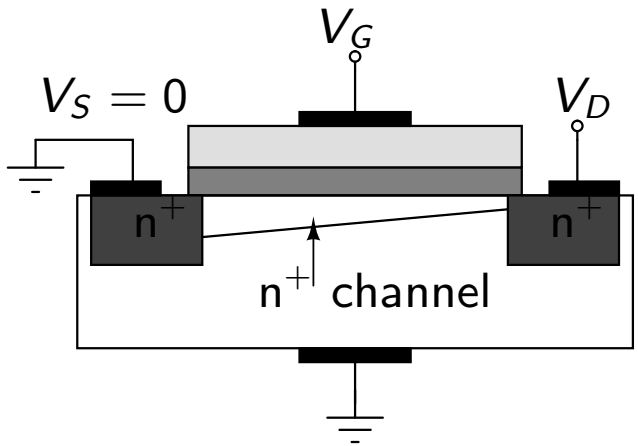
- Once the threshold has been crossed, we need to make the electrons move, i.e. set up a current.
- For this, we need two more terminals- Source (S) and Drain (D), and a potential across them to control the flow of electrons.
- The drain and source are heavily-doped n-type regions.
- We now have a 4-terminal device- drain, source, gate and body.
- The drain and source can be interchanged!



MOS Transistor Characteristics- Linear Region

- Assume that $V_G > V_{TN}$ and $V_{GS} - V_{TN} > V_{DS}$.
- The device is on as the threshold has been crossed. The inversion layer (full of electrons) is now a connecting path between the two n^+-type source and drain regions.
- Due to a nonzero V_{DS} , electrons flow from the drain to the source via the inversion layer. The inversion layer is now called a channel.
- The current flowing in the channel is called the drain current (I_D). For this bias condition, I_D is given by

$$I_D = \frac{k_n}{2}(2(V_{GS} - V_{TN})V_{DS} - V_{DS}^2) \tag{1}$$



MOS Transistor Characteristics- Linear Region (cont'd...)

Based on our discussion so far, try to do the following exercises.

- For the above biasing, plot a graph of I_D v/s V_{GS} as you increase V_{GS}, starting from 0V. You may assume that V_{DS} is small (though not necessary). Now you know why this is called the **linear** region!

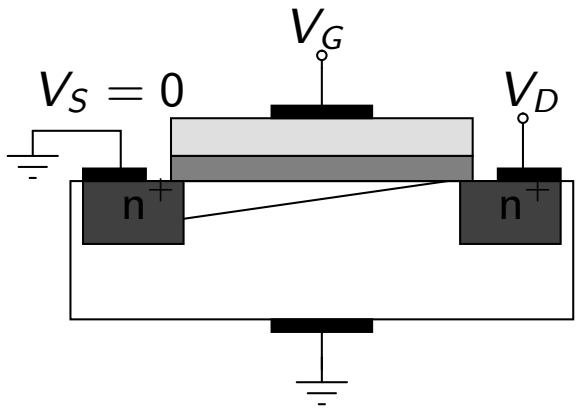
- Now for a given V_{GS}, plot a graph of I_D v/s V_{DS} as you increase V_{DS}, starting from 0V. At what value of V_{DS} is the I_D maximum?

MOS Transistor Characteristics- Saturation Region

- We have seen that the I_D reaches a maxima when $V_{DS} = V_{GS} - V_{TN}$.
- At this time, we see that the $V_{GD} = V_{TN}$. At this time, the channel depth at the drain-substrate interface is zero. This is called pinch-off.
- When V_{DS} is increased further, $V_{GD} < V_{TN}$ and the pinchoff point shifts towards the source.
- The I_D is now very weakly dependent on V_{DS}. The channel voltage is equal to $V_{DS,sat} = V_{GS} - V_{TN}$. The rest of the drain bias voltage is across the pinched-off region.

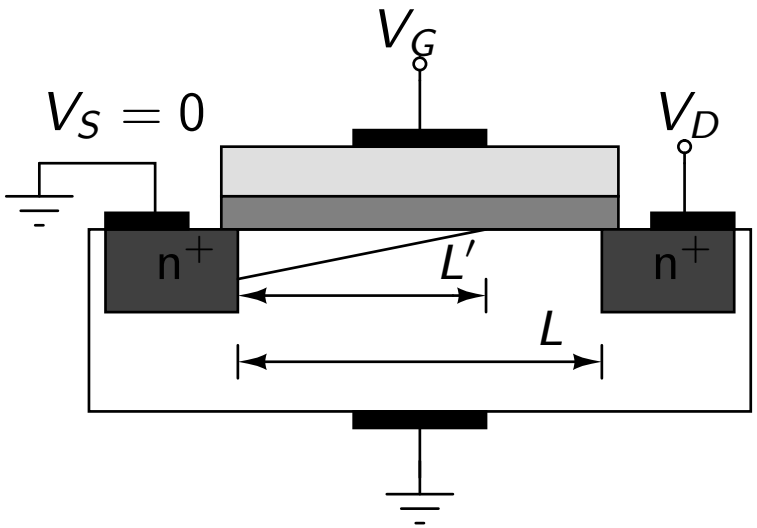
Substituting $V_{DS} = V_{GS} - V_{TN}$ in equation (1), we get

$$I_D = \frac{k_n}{2}(V_{GS} - V_{TN})^2 \tag{2}$$



The Saturation Region (cont'd...)

- From equation (2) we see that I_D is now independent of V_{DS}.
- The plot of I_D v/s V_{DS} in the saturation region is a straight line parallel to the VDS axis.
- That does not happen practically. If the effective length after pinch-off, i.e. L' , is significantly less, I_D does change with V_{DS}!
- We know that, $k_n = \mu_n C_{ox}(\frac{W}{L})$. After pinch-off, we have $k_n = \mu_n C_{ox}(\frac{W}{L'})$.
- Assume $L' = L - \Delta L$. Substitute this in (2).



The Saturation Region (cont'd...)

We now have

$$I_D = \left(\frac{1}{1-\frac{\Delta L}{L}}\right)\frac{\mu_n C_{ox}}{2}\left(\frac{W}{L}\right)(V_{GS}-V_{TN})^2 \tag{3}$$

It can also be shown that, $\Delta L \propto \sqrt{V_{DS}-V_{DS,sat}}$. Using power series, we get

$$1-\frac{\Delta L}{L} = 1-\lambda V_{DS}$$

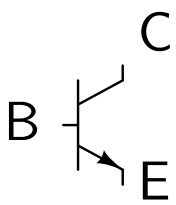
Assuming $\lambda V_{DS} \ll 1$, equation (3) now becomes

$$I_D = \frac{\mu_n C_{ox}}{2}\left(\frac{W}{L}\right)(V_{GS}-V_{TN})^2(1+\lambda V_{DS}) \tag{4}$$

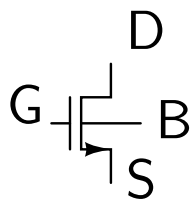
Clearly, the decrease in channel length causes I_D to be linearly varying with V_{DS}! This is called **channel length modulation**, and is a critical issue in IC design.

BJT and MOSFET- A Comparison

Bipolar Junction Transistor	MOS Transistor
1. Current-controlled current source	1. Voltage-controlled current source
2. Current flows due to both electrons and holes (bipolar)	2. Current flows due to one type of carrier (unipolar)
3. No two terminals are interchangeable.	3. Source and Drain can be interchanged.
4. No two terminals are strictly isolated.	4. Gate is isolated by means of an insulator.



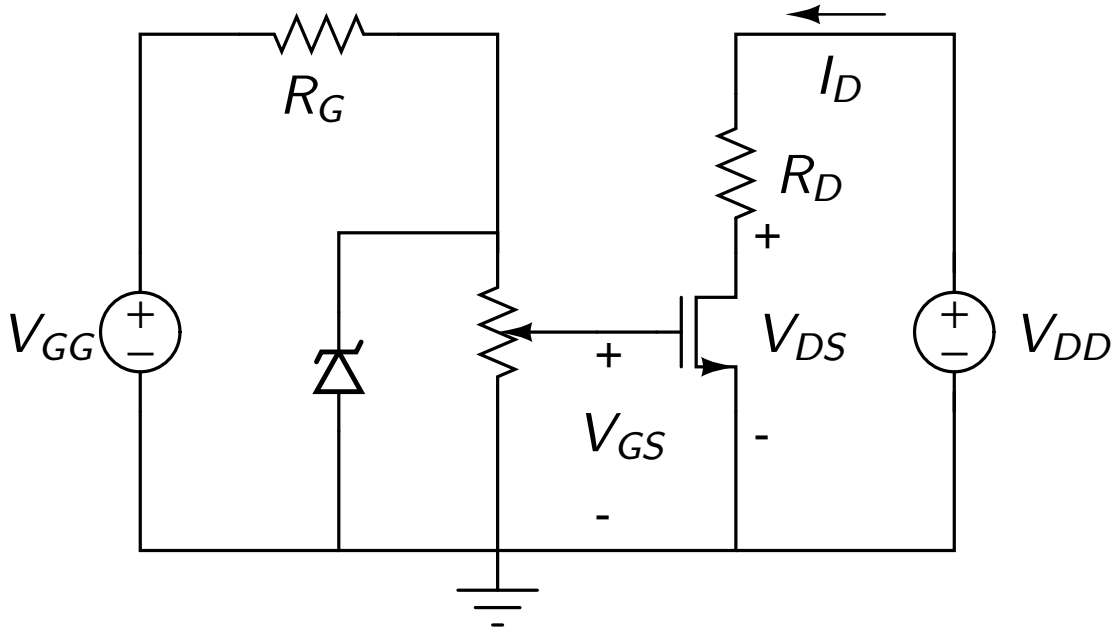
An n-p-n BJT



An n-channel MOSFET

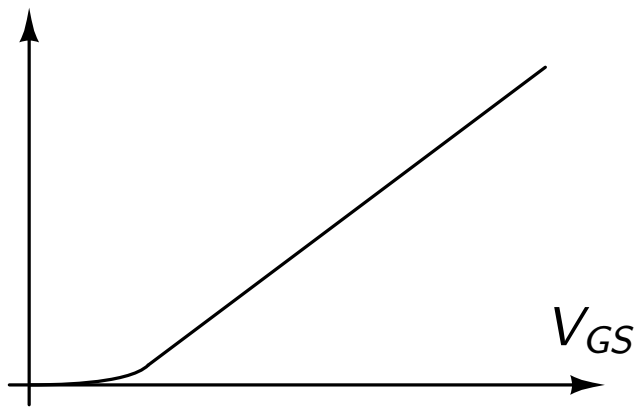
Experiment- Part 1

- In this part, we will measure the NMOS threshold voltage. We will use the IC CD4007.
- Connect the NMOS substrate to ground, and the PMOS substrate to V_{DD}.
- We will operate the NMOS in the linear region. Apply a small V_{DS} of around 0.25 V and keep it constant for a set of I_D v/s V_{GS} readings.
- Vary V_{GS} from 0 to V_{DD} and note I_D.



Experiment- Part 1 (cont'd...)

We expect to see an I_D v/s V_{GS} plot like this.

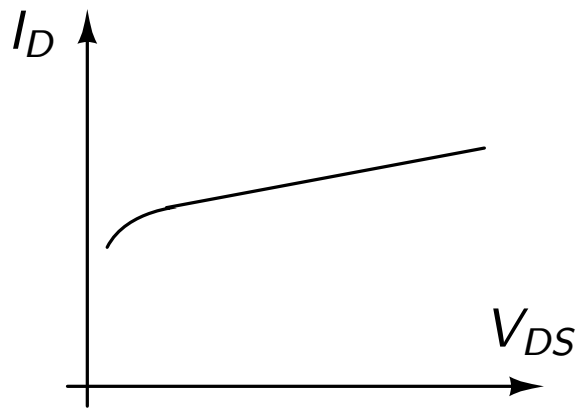


Extrapolating the linear portion of the plot to find the intercept on the V_{GS} axis gives us V_{TN}.

Q: Why do we need to extrapolate?

Experiment- Part 2

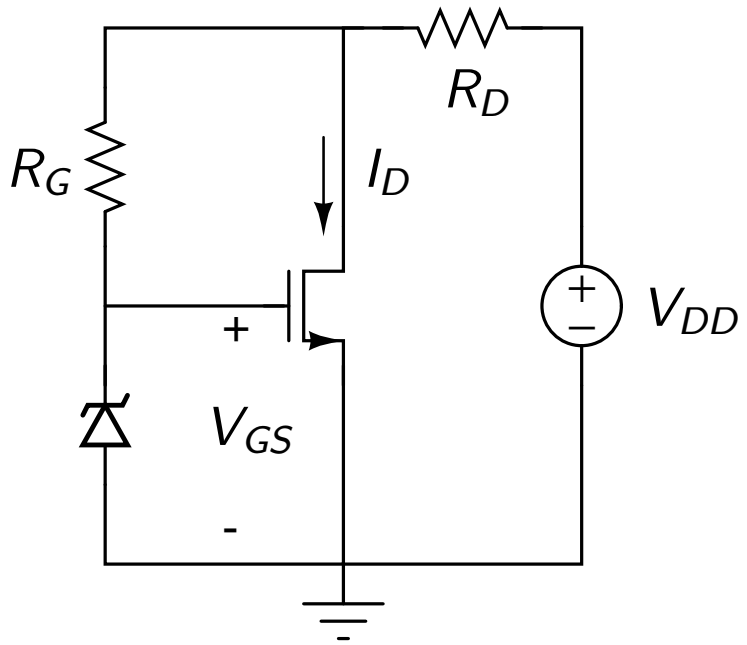
- In this part, we investigate the $I_D - V_{DS}$ characteristics.
- The circuit to be used is the same as in Part 1.
- For a fixed value of V_{GS} , vary V_{DS} to get different values of I_D.
- The expected I_D v/s V_{GS} plot is as shown.



- This plot will help you find the Early Voltage V_A. **How**?

Experiment- Part 3

- In this part, we look at the $I_D - V_{GS}$ relationship for an NMOS in the saturation region.
- We make $V_{GS} < V_{DS}$. This ensures that $V_{DS} > V_{GS} - V_{TN}$.
- We know that in the saturation region, $I_D = \frac{k_n}{2}(V_{GS} - V_{TN})^2$.
- What kind of an I_D v/s V_{GS} plot do you expect?

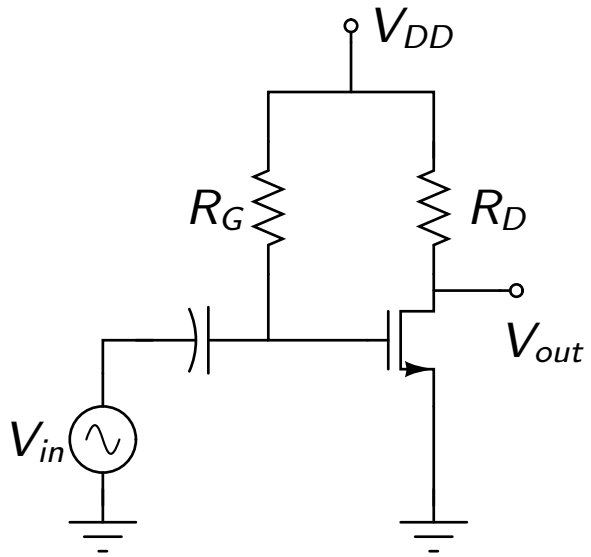


Experiment- Part 4

- In this part, we will measure the small signal transconductance g_m of the NMOS, defined as

$$g_m = \left.\frac{\partial i_d}{\partial v_{gs}}\right|_{V_{DS}} \tag{5}$$

- Here, i_d and v_{gs} are small-signal quantities. Measure g_m using the circuit shown.
- Bias the NMOS in saturation with $V_{GS} = 2\text{V}$ and $V_{DS} = 5\text{V}$.
- Now apply a sine wave and find out the voltage gain $A_v = V_{out}/V_{in}$. Also, $A_v = g_m R_D$. From this you can find the unknown g_m.



Food For Thought

* Bring out some differences between a BJT and a MOSFET, **other than** the ones mentioned in this document.
* Rather than using a metal for the gate, the contemporary VLSI industry uses a material called **polysilicon**. What is it and what advantages does it offer over using a metal for the gate?
* Suppose we take a different approach to measuring g_m than the one explained. Partially differentiating equation (2) w.r.t V_{GS}, we find

$$g_m = k_n(V_{GS} - V_{TN})$$

Eliminating k_n, we get

$$g_m = \frac{2I_D}{(V_{GS} - V_{TN})}$$

Is this procedure correct? What differences do you expect to find between the values of g_m calculated by these 2 methods?