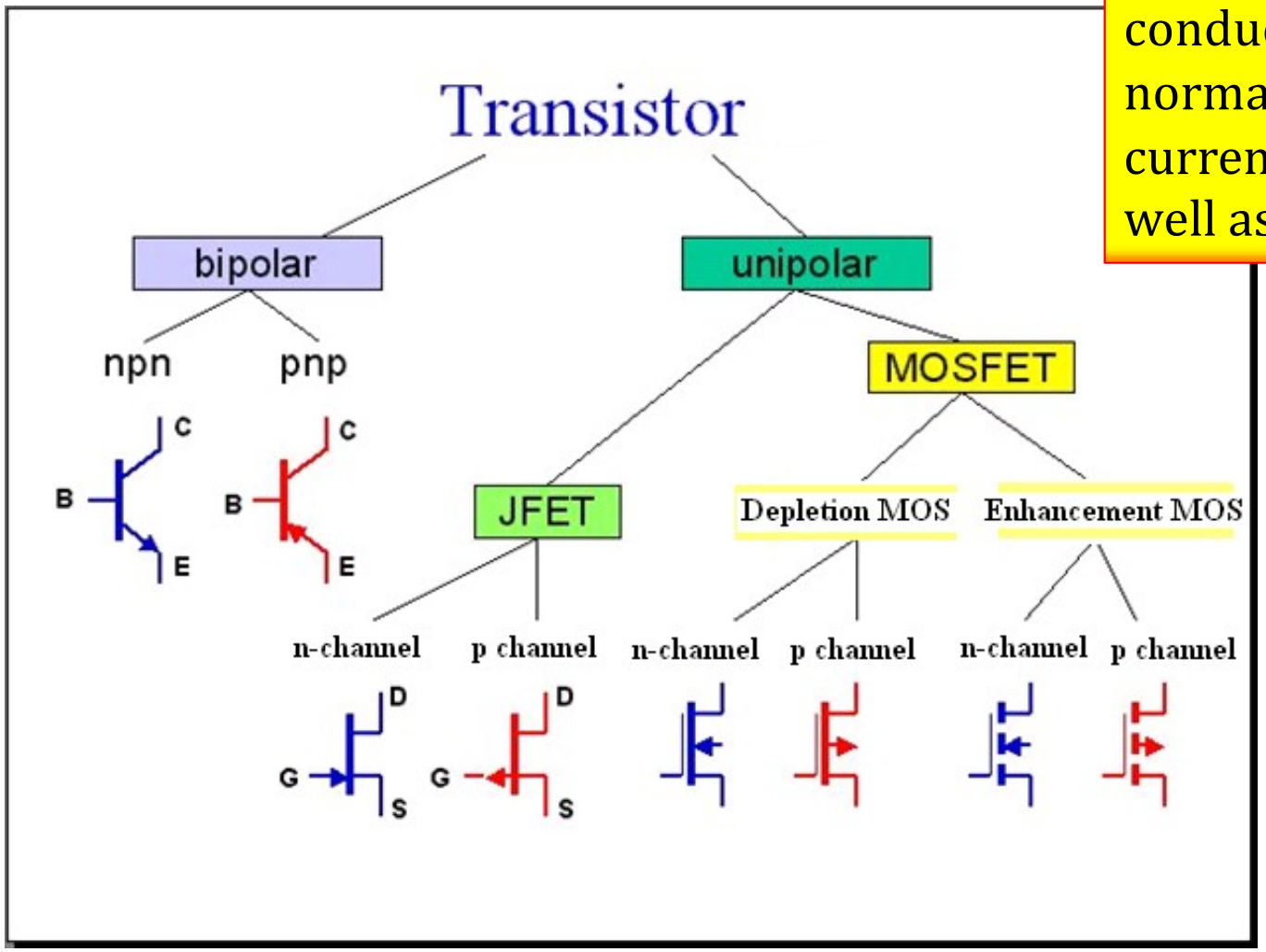
Transistors - MOSFET

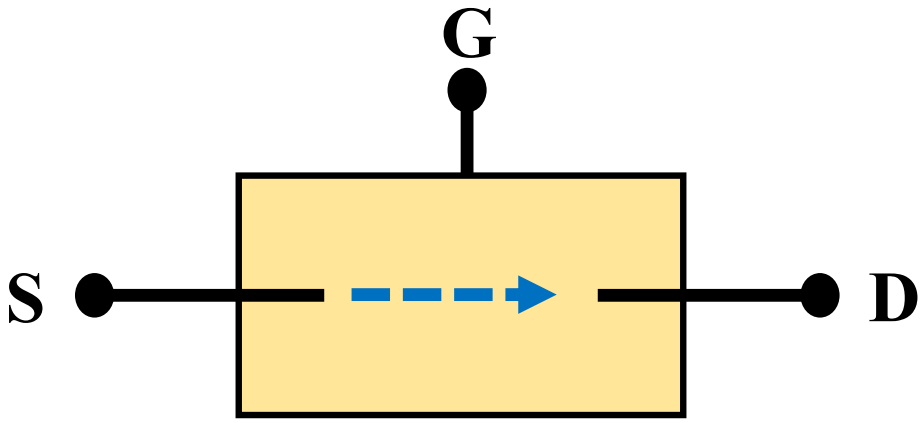
1) Concept of Field Effect Transistors (FETs)
2) Metal Oxide Semiconductor FETs (MOSFETs)
   - » Three terminal device
3) MOSFET working principle
   - » Threshold voltage
4) Operating modes
   - » Cut off, Triode, Saturation
5) Amplifiers and biasing
6) Small signal analysis

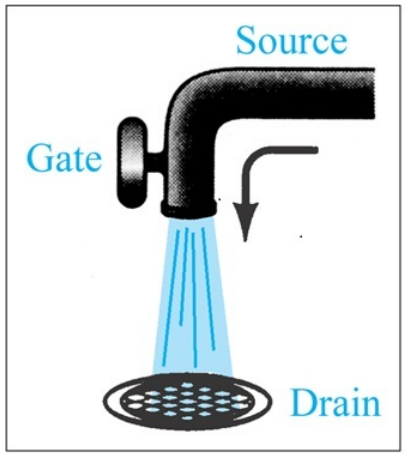
What is a transistor?

It is a three terminal semi-conductor device, which normally offers voltage or current **amplification**, as well as switching.



Field Effect Transistors (FETs)

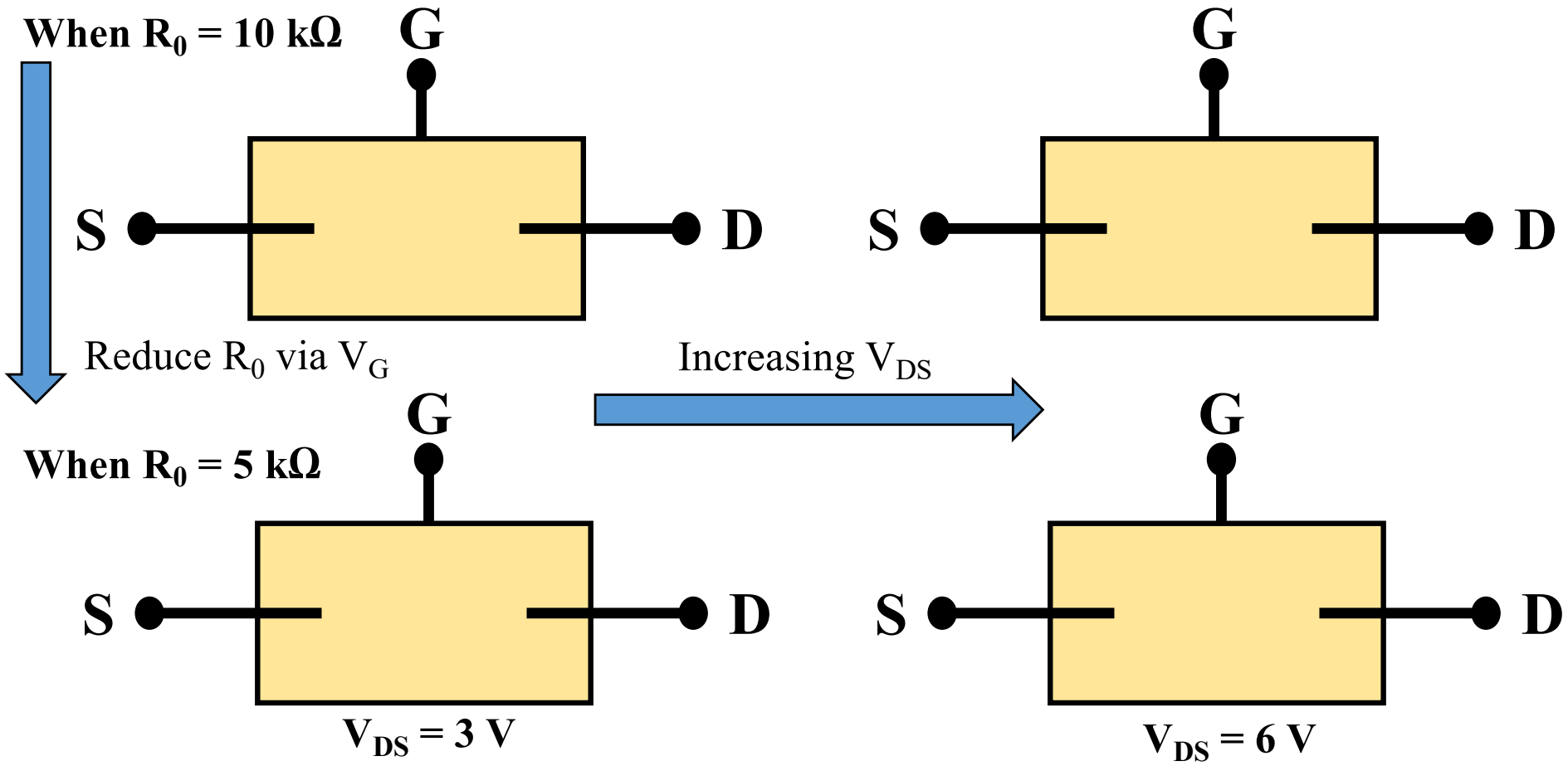




1) An FET device has three terminals: Gate (G), source (S) and drain (D)
   - » G terminal controls current flow between S and D
2) Could think of a FET device as a resistor between S and D
   - » Voltage applied to G controls resistance between S and D
3) So you could think of an FET device as a tap, where G controls the water flow between S and D

The field effect

Field effect refers to the mechanism by which the gate controls the current between S and D. As we will see later, the voltage applied at the gate controls the resistance of the channel between S and D. In the following example, V_G controls R_0.



Turning on and off through the gate

The previous slide illustrates what happens to the current between S and D as we change the resistance as controlled by the voltage at G.

It is also possible shut off the path between S and D so that an open circuit appears between S and D.

Under such conditions, we can see that no current can run between S and D for whatever voltage is applied across S and D.

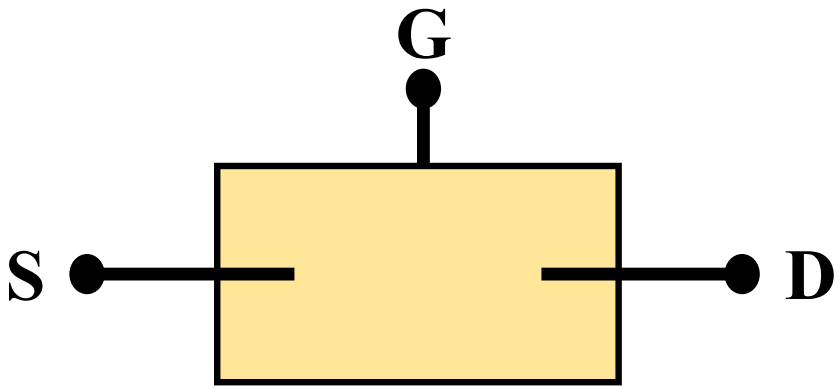
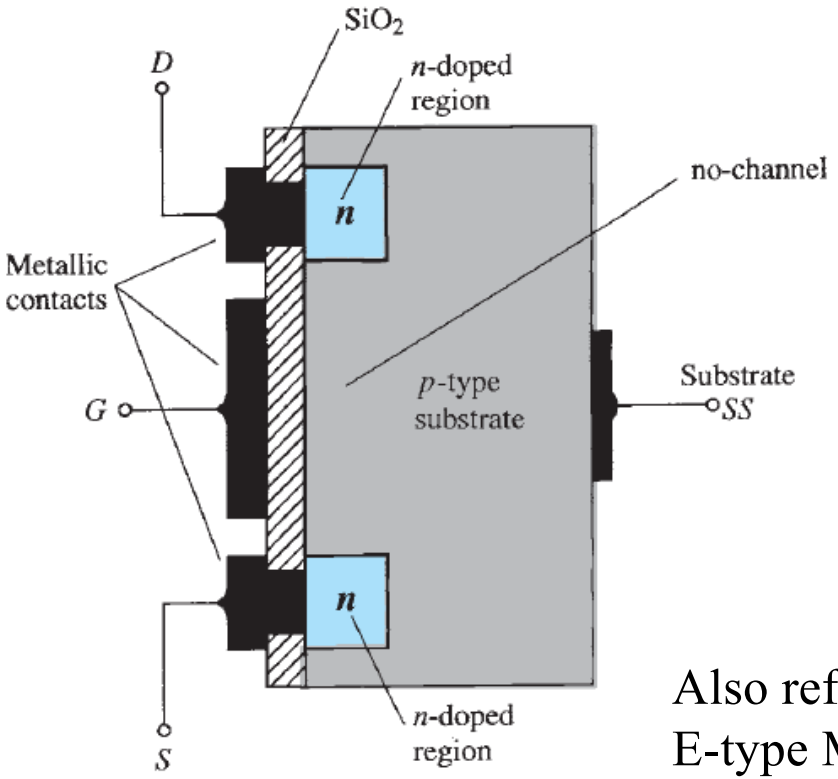


Illustration when the device is turned off

Enhancement-type MOSFET

MOSFET: Metal Oxide Semiconductor FET



Gate is electrically isolated from the substrate by the thin layer of silicon oxide which is insulating.

Metal: Refers to the gate that is made of metallic contacts

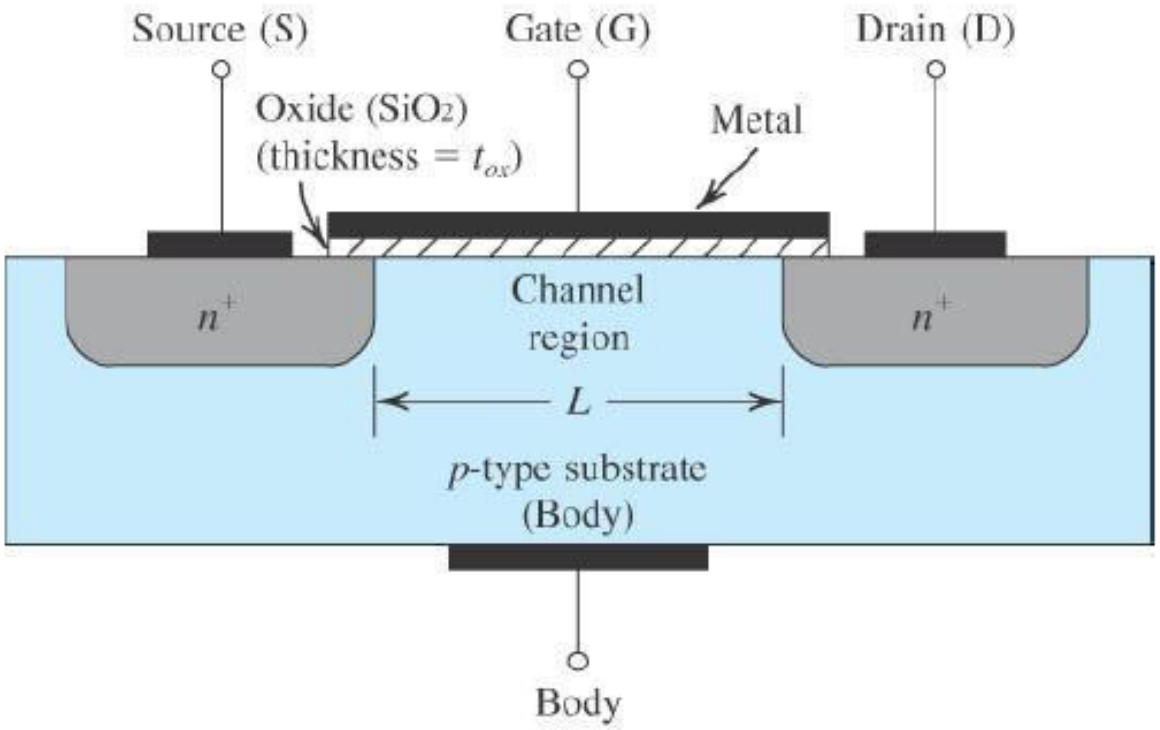
Oxide: Refers to the insulating silicon oxide

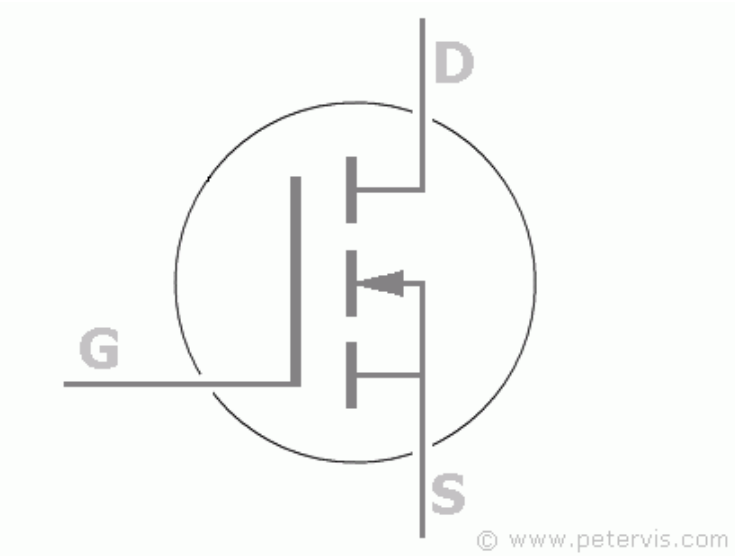
Semiconductor: Refers to the semiconductor substrate

Also referred to as an E-type MOSFET
E-type MOSFET has no channel built into the device

This example shows an E-type n-channel MOSFET (or NMOS for short)

Enhancement NMOS Schematic and Symbol



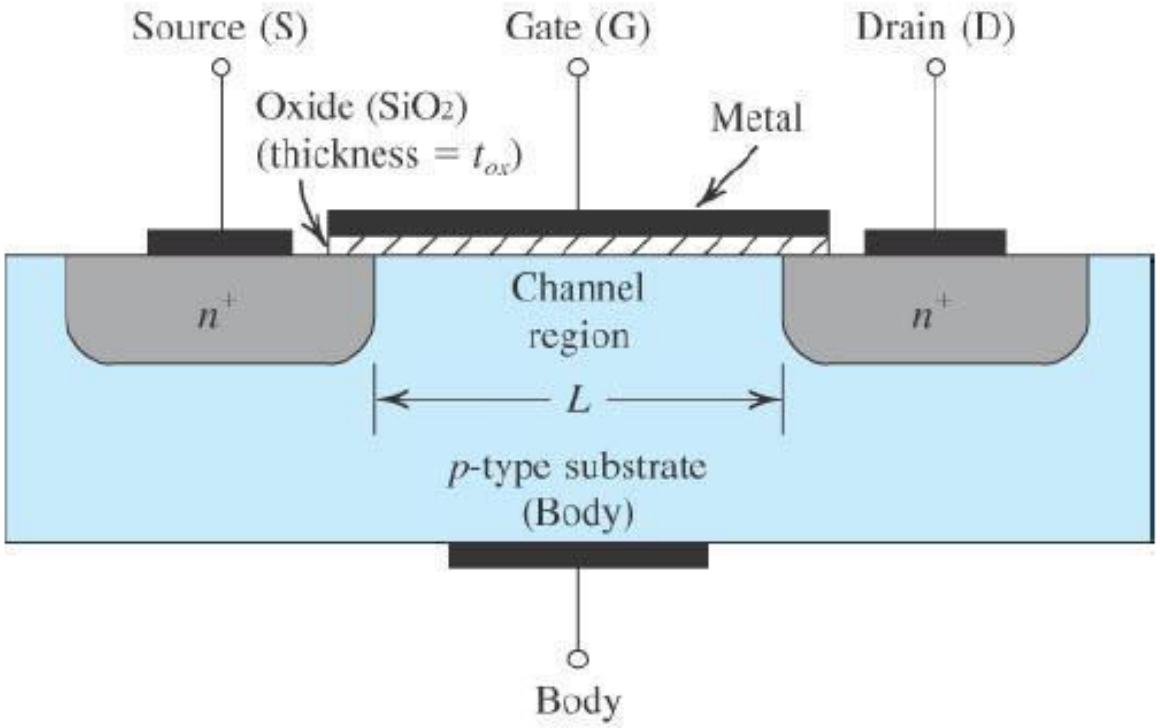


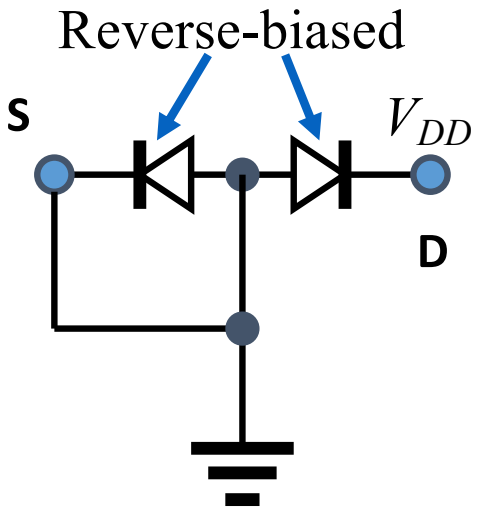
The "+" indicates heavier doping (i.e. higher carrier concentration); the source and drain are doped more heavily.

Current is supposed to flow only through the channel region and commonly referred to as the drain current and takes the symbol I_D.

But we see that the channel regions is originally p-type. The channel separates the source and drain regions that are n-type.

Channel: pathway for current to flow



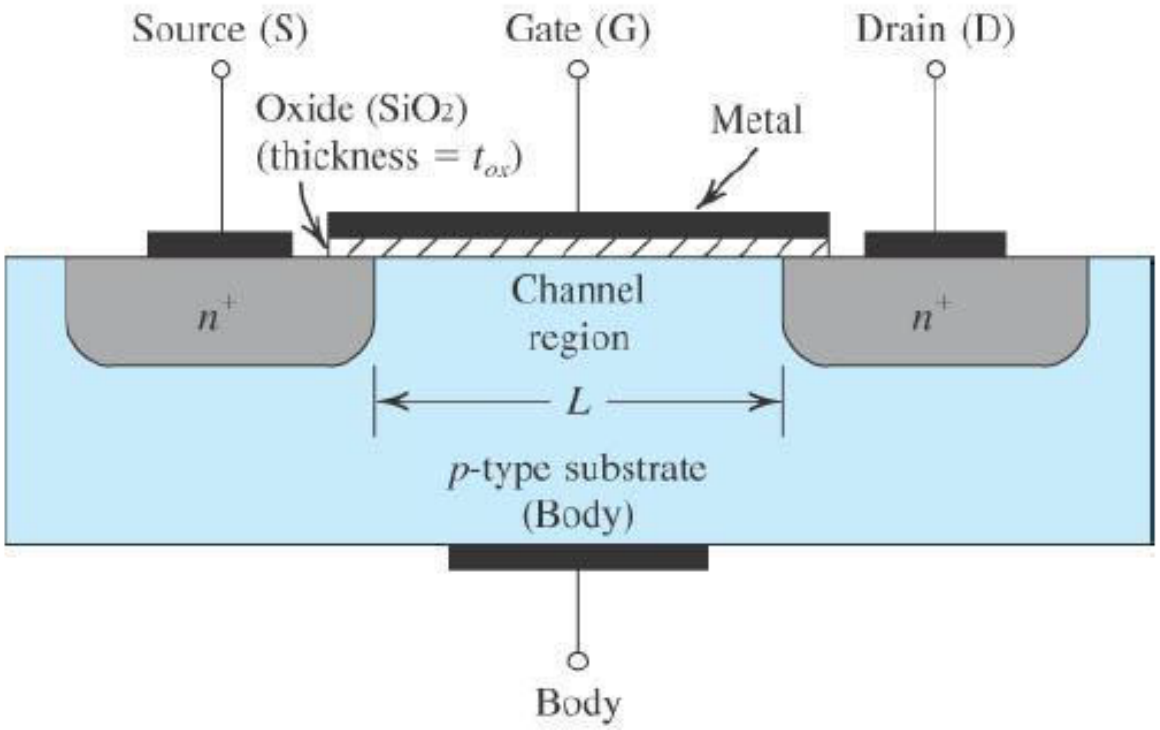


When $V_{GS} = 0V$, $I_D = 0A$

The figure above shows what the MOSFET looks like by default where there is no channel. At the interface of the source and body as well as drain and body, you get PN junctions. In fact, as shown in the figure on the right, these PN junctions are located back to back.

Imagine what happens when you apply a voltage across S and D: One diode conducts while the other is reverse biased. The net result is that no current flows between S and D in the default state (remember that currents run in loops and must thus complete a full cycle).

Formation of channel



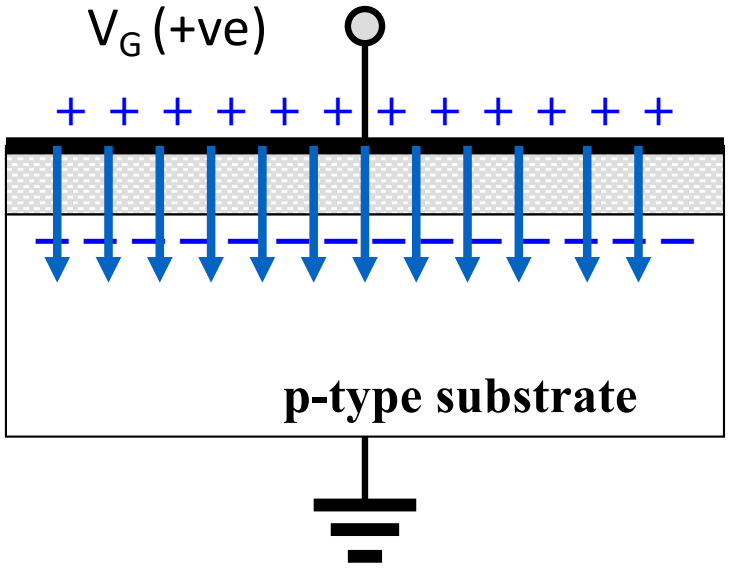
V_T is referred to as the Threshold Voltage. It is the minimum voltage required to form a channel between source and drain.

To establish a path for current to flow between D and S, a n-type channel that connects S and D must first be formed.

Once a n-type channel is formed between S and D, electrons have a continuous conducting path between S and D.

In order to form a channel, we have to apply a minimum amount of voltage between the gate and the body. This minimum voltage is called the Threshold Voltage (V_T).

Formation of channel in NMOS



V_T is a constant that depends on the model of the MOSFET.

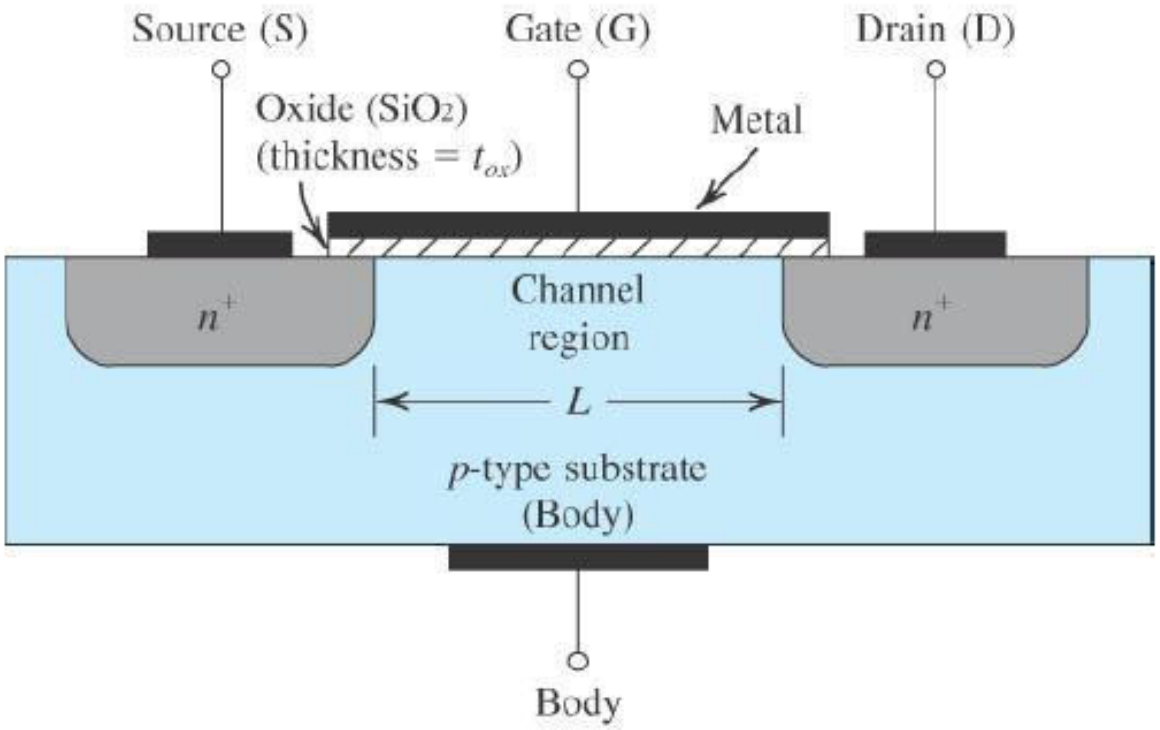
When $V_G > V_T$:

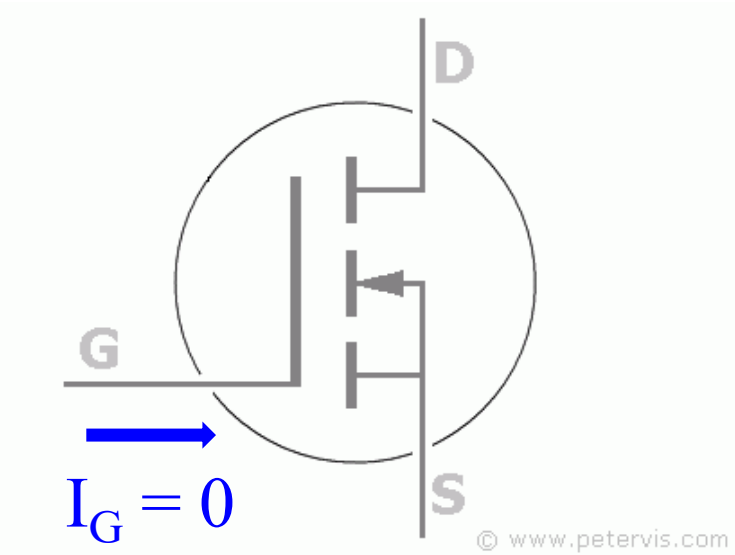
Electrons which are the minority carriers in p-type are attracted towards the surface at the silicon-oxide interface.

The high concentration of electrons creates a shallow n-type channel near this surface.

Increasing V_G pulls more electrons to the surface thereby increasing the concentration of carriers which reduces the resistivity of the channel. So for the same voltage across drain and source, drain current increases as V_G is increased.

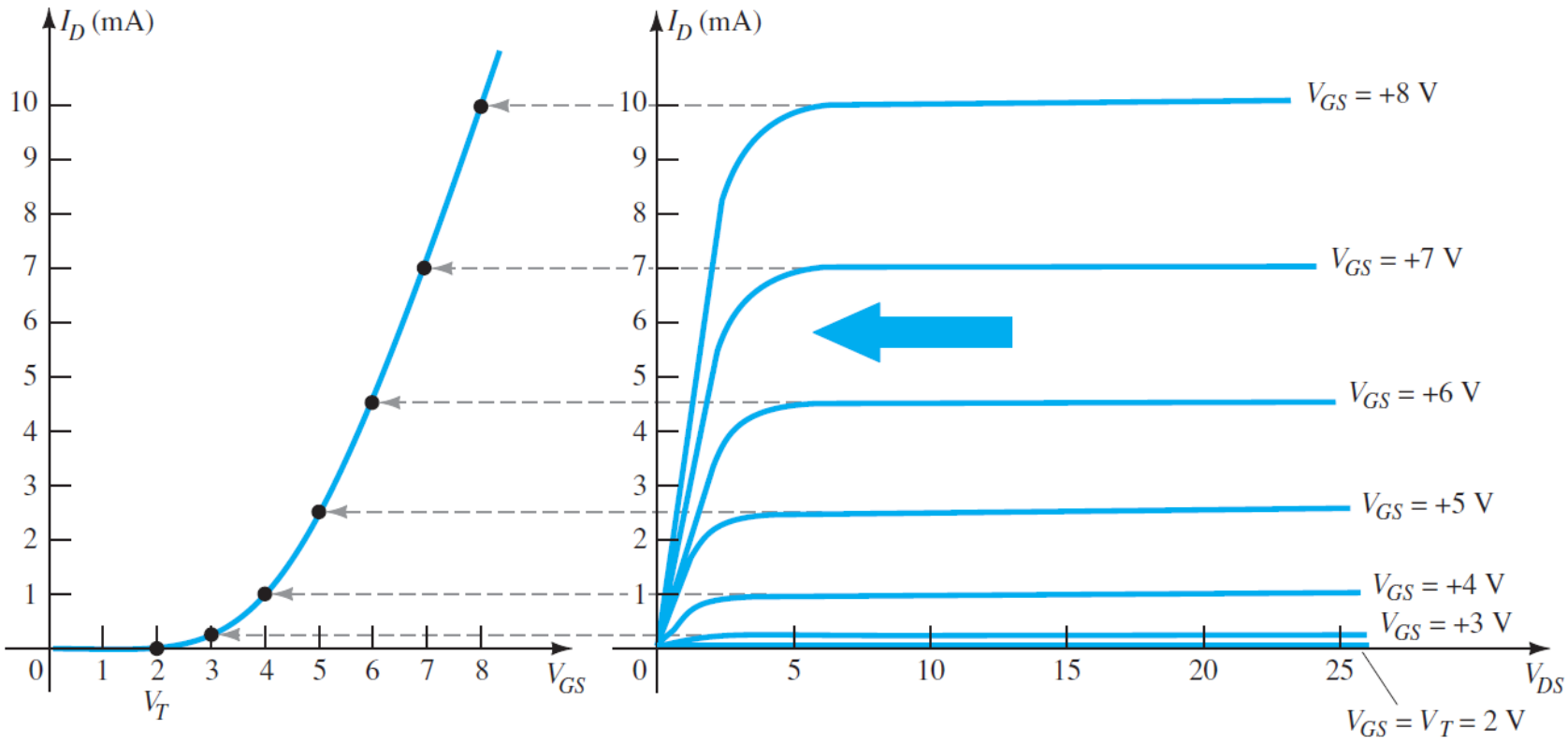
Gate current





The gate current in a MOSFET is ideally zero as the oxide blocks current flowing between the metal gate and the semiconductor substrate.

Transfer characteristics of E-type NMOS



Note that current (I_D) only starts to flow when V_{GS} exceeds V_T. In this example, V_T is 2V. V_T is thus the voltage when the MOSFET starts to turn on.

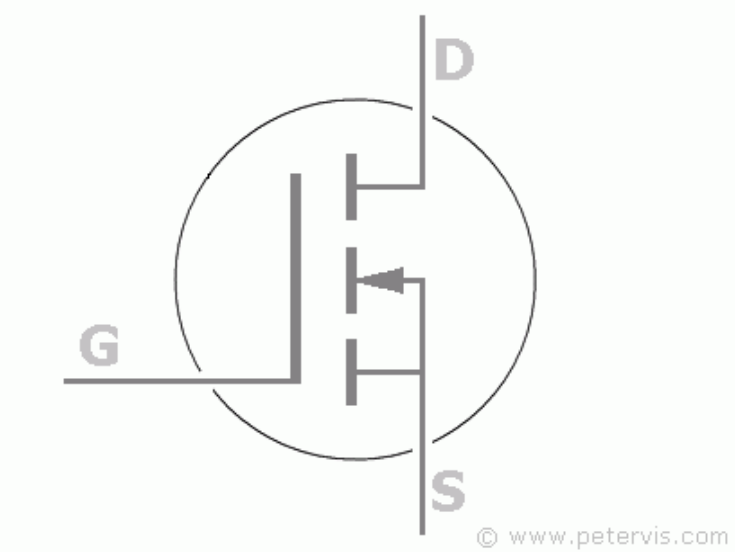
Modes of operation

There are three modes of operation in a MOSFET

1) Cutoff: When there is no drain current because no channel has been formed.

2) Triode: When the level of drain current is determined by the voltages between G-S (V_{GS}), and D-S (V_{DS}).

3) Saturation: When the level of drain current is determined mostly by the voltage between G-S (V_{GS}).



The choice of which of these modes the MOSFET will work in depends on the values of V_{GS} and V_{GD} relative to V_T.

Cutoff mode

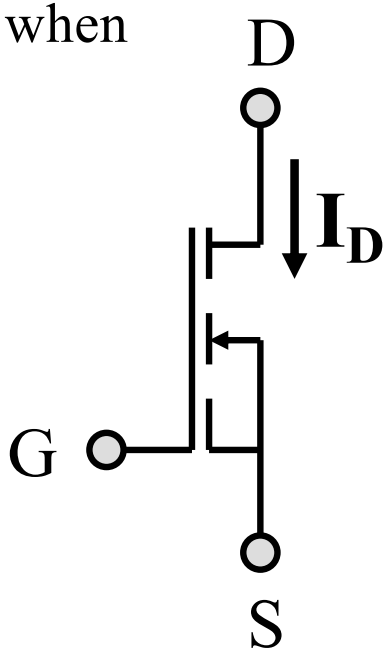
The first mode we shall consider is when the device is off. This is known as Cutoff. As previously pointed out, this occurs when the threshold voltage has not been reached.

Cutoff Region

$V_{GS} < V_T$, $V_{GD} < V_T$

As a result of this,

$I_D = 0A$



Cutoff Region

- **V_{GS} and V_{GD} both less than V_T**
- Channel is off at both source and drain
- No conduction and no current flow between S & D

Triode mode

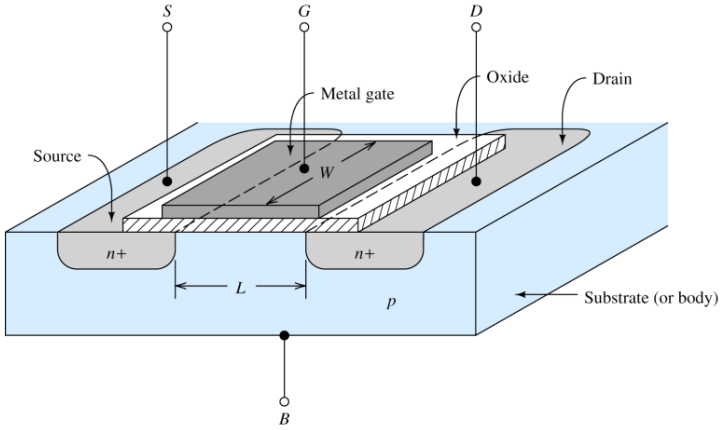
As mentioned previously, when the gate voltage exceeds the threshold the MOS can conduct and therefore turns on. When it turns on, it can operate in one of two possible regions. One of these 2 modes is known as the triode or ohmic region.

Triode/Ohmic Region

$V_{GS} > V_T$, $V_{GD} > V_T$

The drain current is given by:

$$I_D = K\,[2(V_{GS} - V_T)V_{DS} - V_{DS}^2]$$



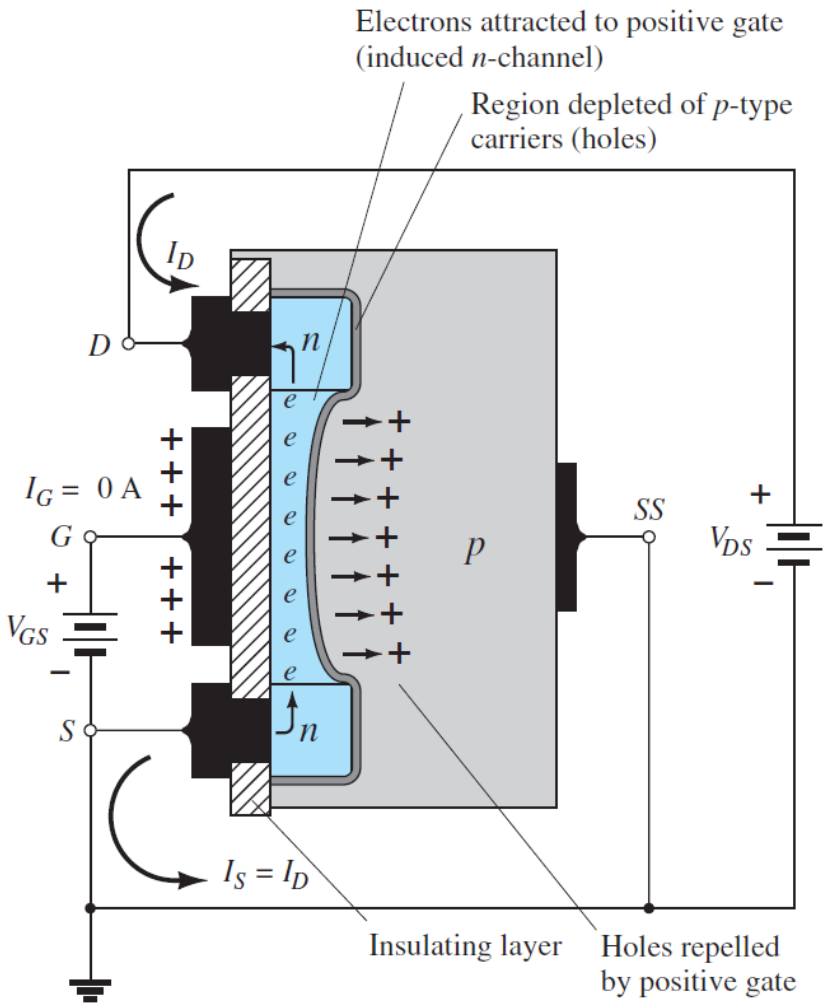
K is known as the conductance parameter, and is defined as: $K = \left(\frac{W}{L}\right)\left(\frac{\mu C_{ox}}{2}\right)$

Triode/Ohmic Region

- **Both V_{GS} and V_{GD} are greater than V_T**
- Channel is on at source and also on at drain
- Current is dependent of both V_{DS} and V_{GS}

W: channel width
L: channel length
μ: mobility of carriers
C_{ox}: Oxide capacitance

MOS in Triode mode



Near both the drain and source:

⇨Voltage difference > V_T

⇨Conducting channel is formed on both sides

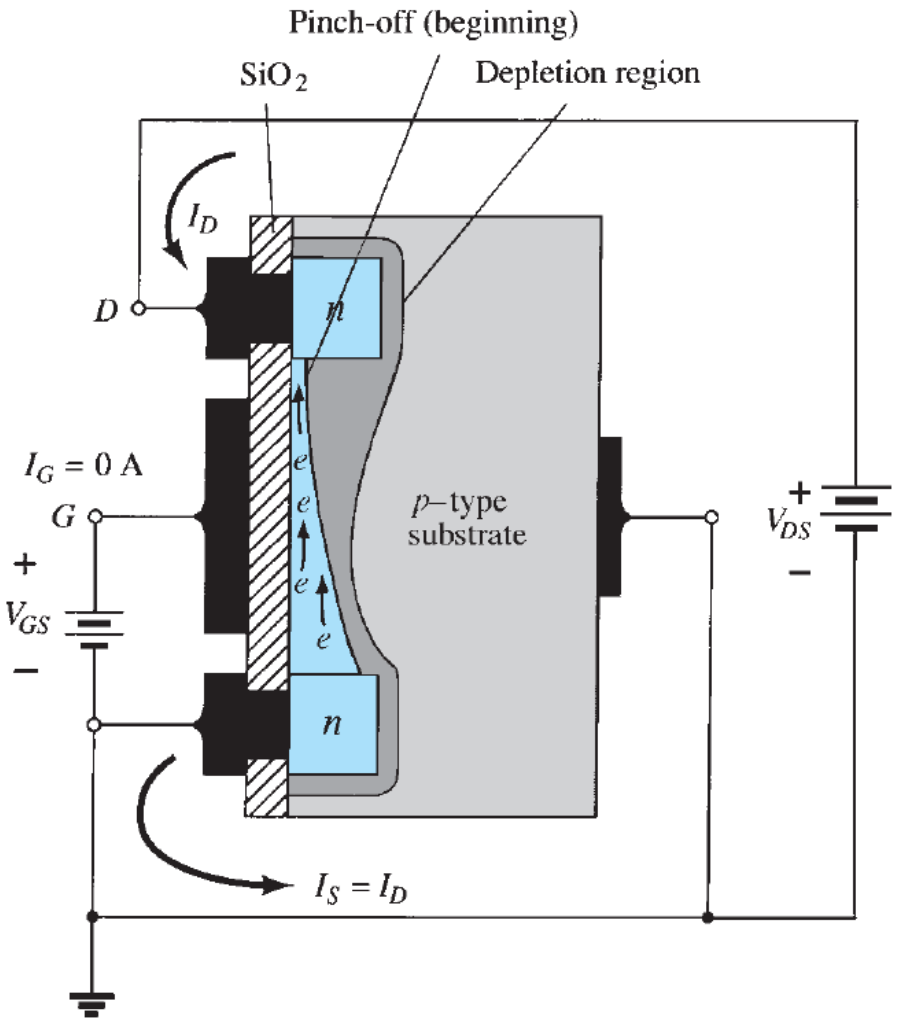
In fact everywhere under the gate oxide is larger than V_T so there is a continuous channel between the drain and source.

⇨Increasing V_{GS} thus increases the number of carriers, resulting in a larger drain current

⇨Increase V_{DS} increases the voltage drop across the channel, resulting in a larger drain current

I_D is dependent therefore on both V_{DS} and V_{GS} (Refer to equation on previous slide)

Increasing V_{DS} while fixing V_{GS}



Given that $V_{DS} > 0V$ then $V_{GD} < V_{GS}$

The field applied by the gate close to S will have to be stronger than close to D.

Channel close to D will be more narrow compared to closed to S, as illustrated in the diagram.

When V_{GD} finally reduces to be equal to V_T (just meets the condition to form a channel), the channel close to D will pinch off, as shown in the diagram.

Example: V_{GS} fixed at 4V and $V_T = 2V$

When $V_{DS} = 1.0V$, $V_{GD} = 3V$_____

When $V_{DS} = 1.5V$, $V_{GD} = 2.5V$_____

When $V_{DS} = 2.0V$, $V_{GD} = 2V$_____

Saturation mode

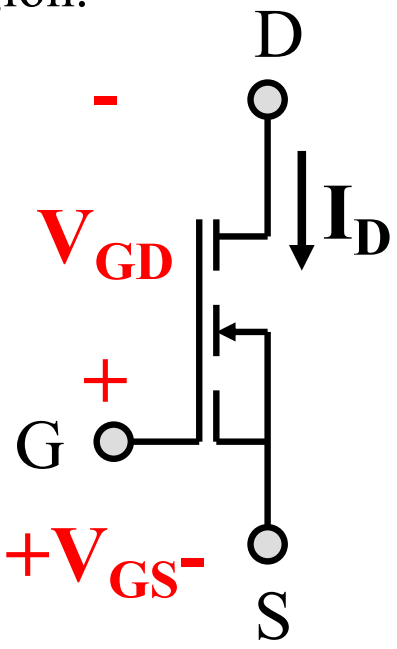
While the MOS is still on, if we continue to increase V_{DS} without increasing V_{GS}, there will come a point when the voltage near the drain becomes less than the threshold ($V_{GD} < V_T$). This is known as the saturation region.

Saturation Region

$V_{GS} > V_T$, $V_{GD} < V_T$

The drain current is given by:

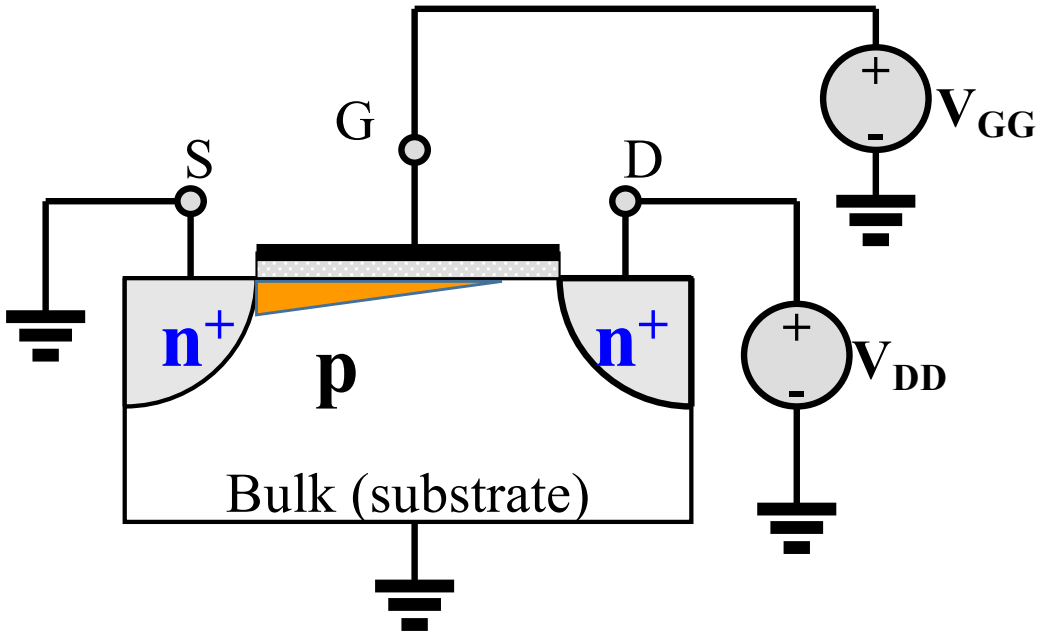
$$i_D = K (V_{GS} - V_T)^2$$



Saturation Region

- **$V_{GS} > V_T$ and $V_{GD} < V_T$**
- Channel is on at source but off at drain
- Current is nearly independent of V_{DS}, depending only on V_{GS}
- One might expect the MOS to operate like in cutoff under these bias conditions but it does not.

NMOS in Saturation



The channel in saturation mode is typically represented as tapered or wedged (as shown in the figure). Large currents still conduct via the channel due to the large voltage drop across drain and source, but changes little with increasing V_{DS}.

Pinch-off (going from triode to saturation)

The channel near the drain begins for fall below the threshold condition when

$V_{GD} = V_T$

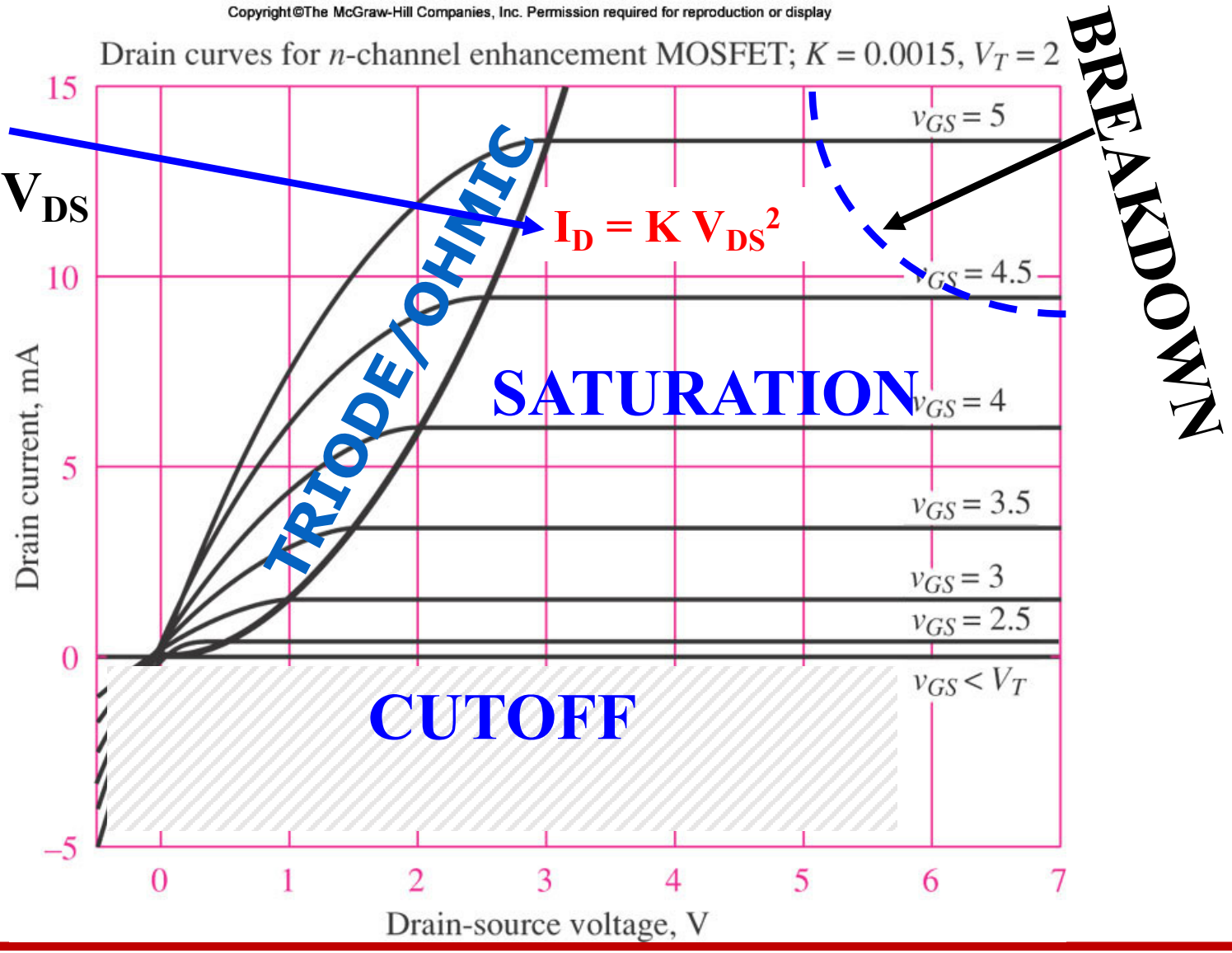
In terms of the drain-source voltage, this corresponds to:

$V_{DS}(\text{pinch-off}) = V_{GS} - V_T$

Operating Regions

When

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



Summary of operating modes

- **Cut-Off:**

  $V_{GS} < V_T, V_{GD} < V_T$

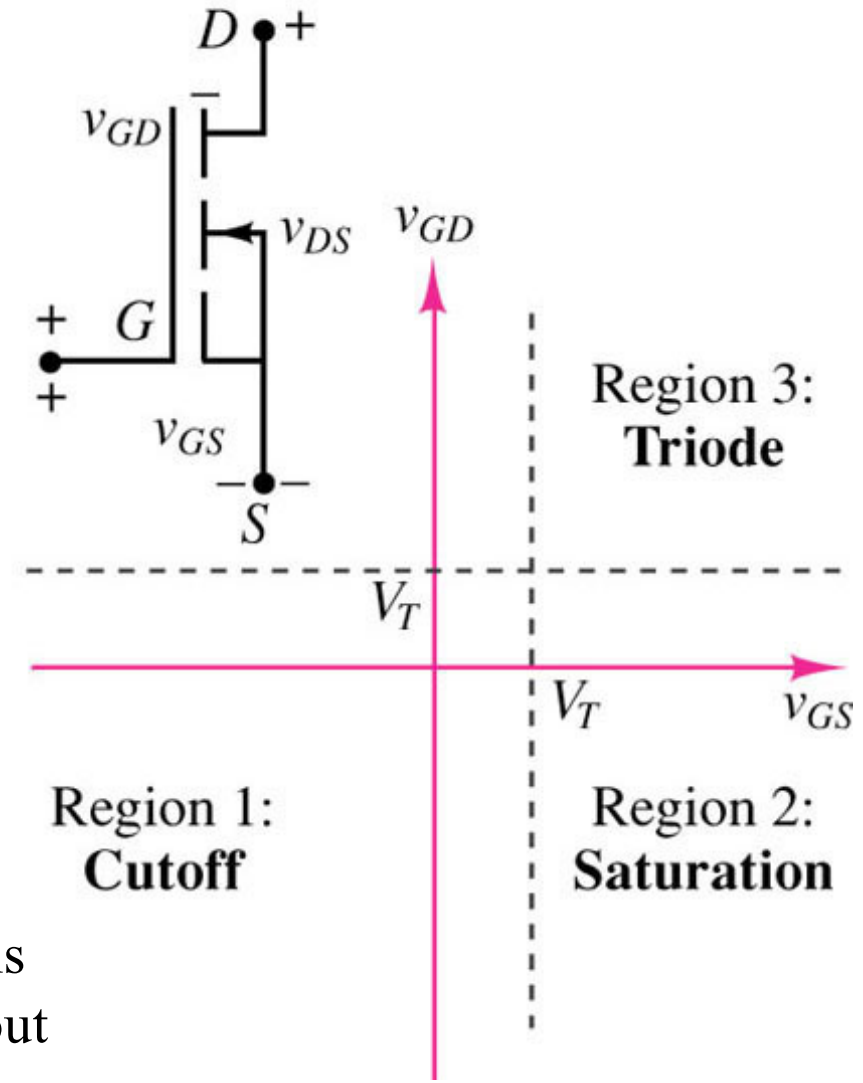
- **Linear/Triode/Ohmic:**

  $V_{GS} > V_T, V_{GD} > V_T$

- **Saturation:**

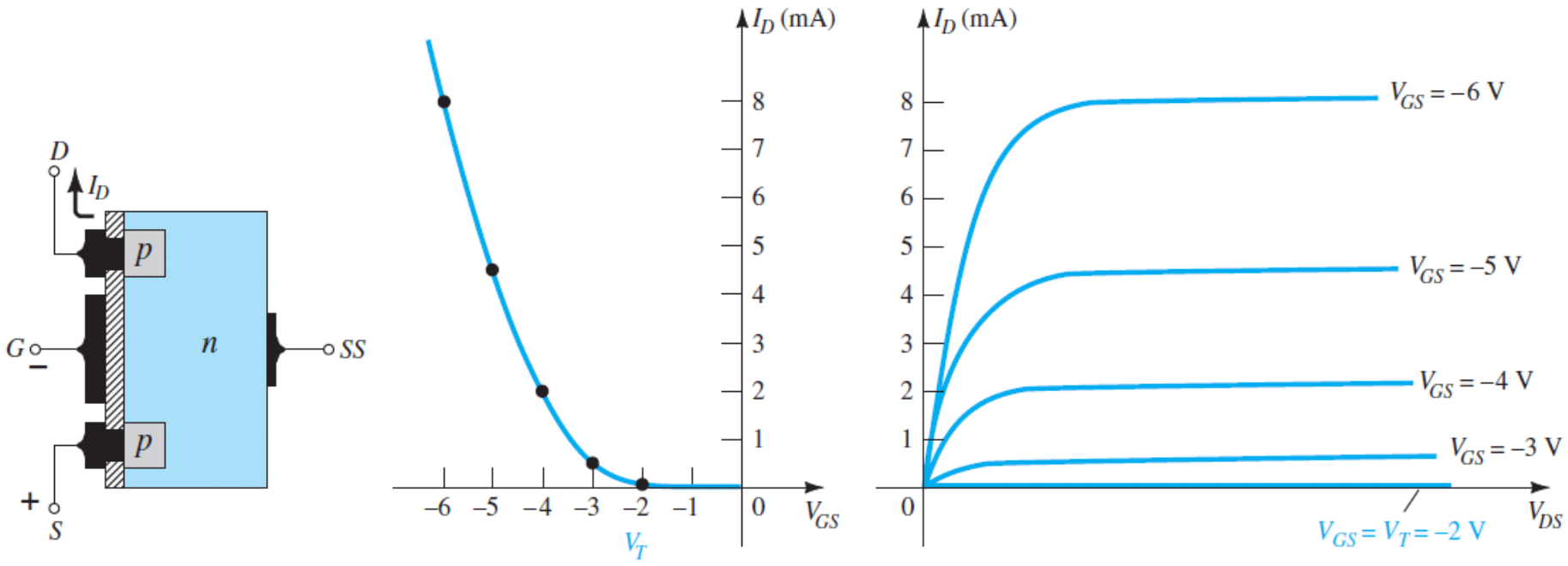
  $V_{GS} > V_T, V_{GD} < V_T$

For n-channel enhancement MOS



Power consumed by MOSFET = $I_D V_{DS}$
This sets a limit on the operating conditions
Too much power and the transistor burns out

E-type PMOS



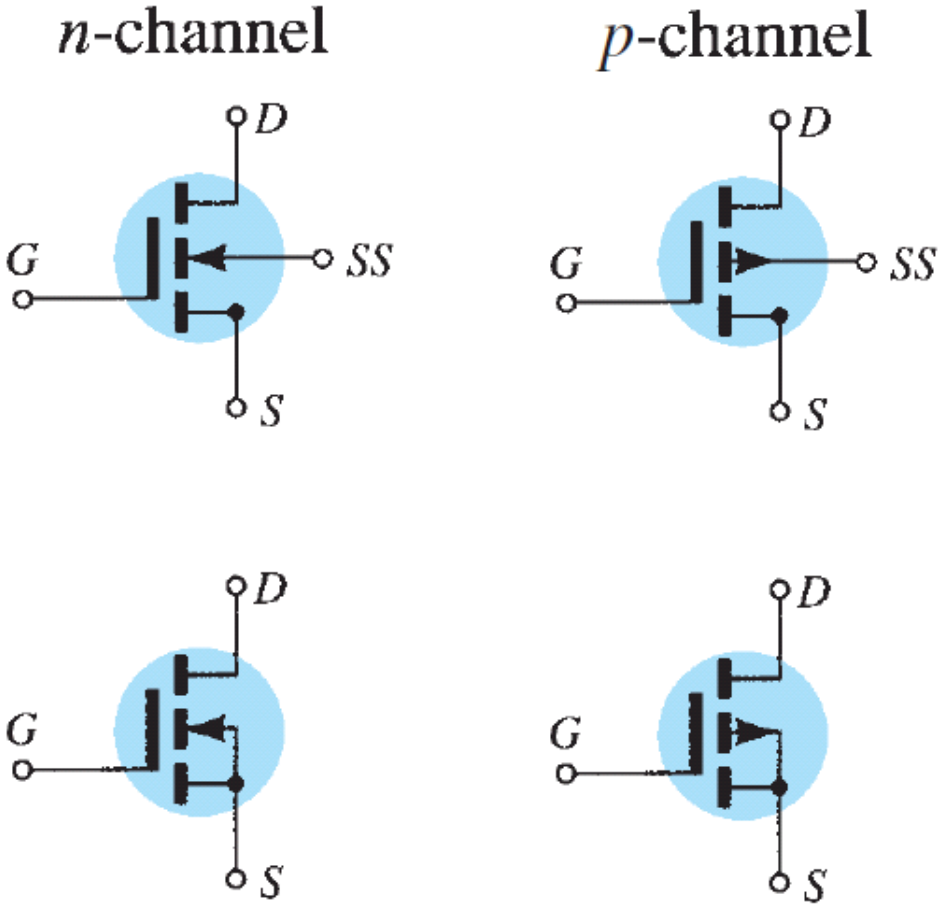
PMOS: p-channel MOSFET

Structure of the PMOS is the same as the NMOS except that the carriers are swapped:

Source and drain are p-type (instead of n-type), body is n-type (instead of p-type), and the channel to be formed is p-typed (thus called p-channel MOS).

The p-channel enhancement-type MOSFET is similar to the n-channel MOSFET (NMOS), except that the voltage polarities and current directions are reversed.

E-type PMOS and NMOS symbols



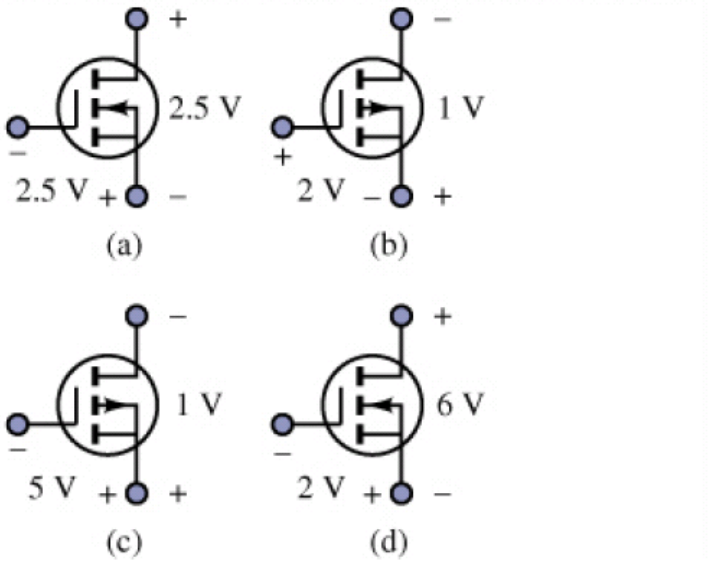
When the body or substrate has an independent connection

When the body or substrate is connected to the source

Arrow points from p-type to n-type

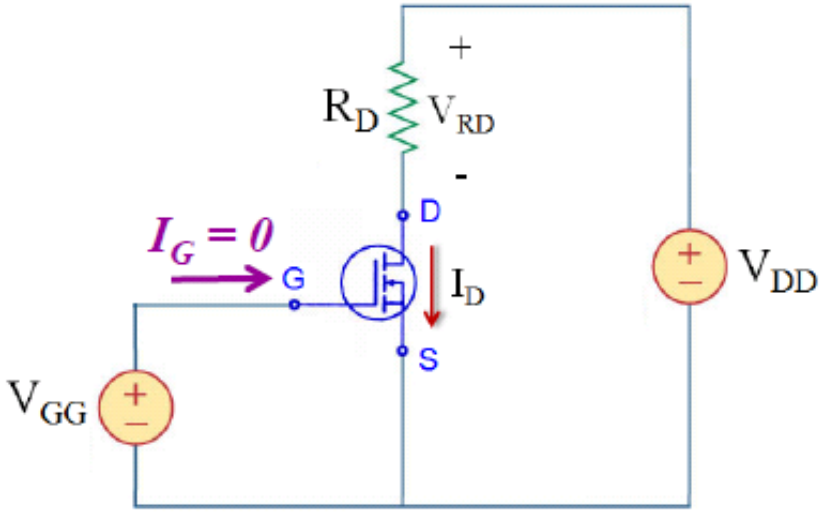
Practice example on operating modes

The transistors have $|V_T| = 3V$. Determine the operating region for each.
For NMOS: $V_T = 3V$; V_{GS} needs to exceed this voltage for NMOS to turn on.
For PMOS: $V_T = -3V$_____; V_{GS} needs to be more negative than this voltage for the PMOS to turn on.



FET Amplifiers: Operating point

One of the applications of FETs is to realize amplifiers. To do this, we need to determine the point among the family curves where the FET going to work at. This point is defined by the value of I_D, V_{GS} and V_{DS}. This referred to as the operating point.



V_{in} is an AC voltage (we will consider this in the graph on the next slide but not here)

$I_G = 0 \rightarrow V_{GS} = V_{GG}$

So we know the value of V_{GS}

Sum voltages across R_D and D-S:

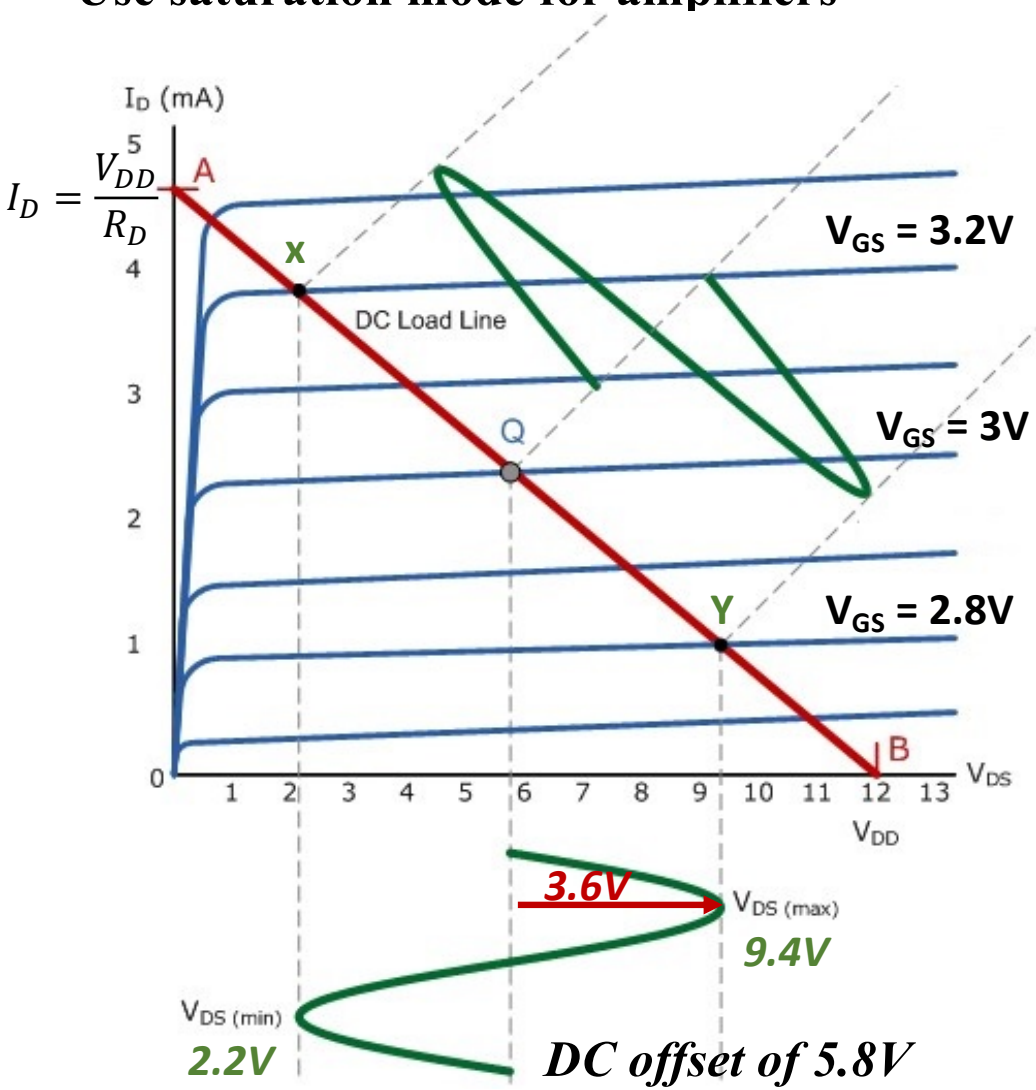
$$V_{DS} + I_D R_D = V_{DD}$$

$$I_D = -\frac{1}{R_D} V_{DS} + \frac{V_{DD}}{R_D}$$

To see how amplification is achieved graphically, we draw the above load line on the I_D vs. V_{DS} curve of the FET and find the intersection point

FET Amplifiers: Operating point

Use saturation mode for amplifiers



$V_{GS} = 3V$

$$I_D = -\frac{1}{R_D}V_{DS} + \frac{V_{DD}}{R_D}$$

Intersection point: I_D = 2.3 mA, V_{DS} = 5.8 V

If V_{in} has an amplitude of 0.2V, V_G will vary with an amplitude of 0.2V at a rest value of V_{GS} = 3V

V_{DS} will also vary but with a larger amplitude of ~3.6V

Phase of sinusoid across V_{DS} is inverted relative to V_{IN} (inverting)

Gain = -3.6/0.2 = -18 (INVERTING)

FET Small signal equivalent circuit

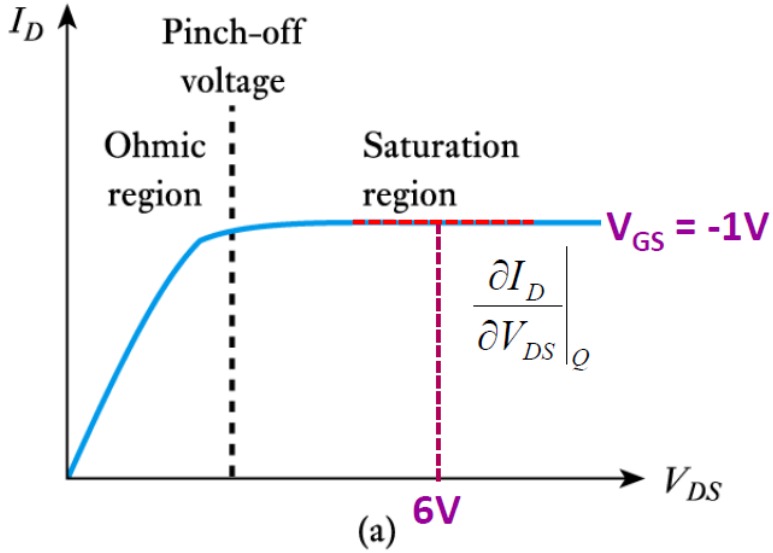
1) Within a small range, we see from the variation of I_D vs V_{GS} and I_D vs V_{DS} is approximate linear
2) Instead of using graphs determine gain, we can instead define the linear characteristics of the FET for a given operating point
3) We can do so using a small signal equivalent circuit (SSEC)
4) This will then allow us to analyze the FET using linear circuit theory that we have learnt so far

Small signal = Small changes

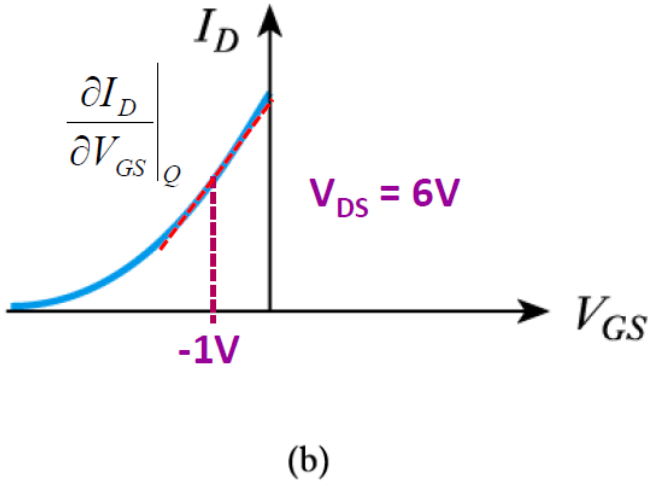
What does I_D depend on to control it?

I_D depends on _____ and _____

$I_D = f(\_\_\_\_\_, \_\_\_\_\_)$



For example, let us say that the operating point (Q) is at :
I_D = 2.5mA, V_{GS} = -1V, V_{DS} = 6V



Express the above using chain rule: $\Delta I_D = \left.\frac{\partial I_D}{\partial V_{GS}}\right|_{V_{DS}} \Delta V_{GS} + \left.\frac{\partial I_D}{\partial V_{DS}}\right|_{V_{GS}} \Delta V_{DS}$

We then replace the partial differentials with new symbols:

$$r_o = \left.\frac{\partial V_{DS}}{\partial I_D}\right|_{V_{GS}}, \quad g_m = \left.\frac{\partial I_D}{\partial V_{GS}}\right|_{V_{DS}}$$

We can see that r_o is usually very large

Small signal = Small changes

Next, we also replace the variables with new symbols. Note that we use small letters to represent small changes/signals and their related parameters. This is so that we do not confuse them with the DC values defined at point Q.

This gives us: $i_d = \Delta I_D$, $v_{gs} = \Delta V_{GS}$, $v_{ds} = \Delta V_{DS}$

Remember: $I_D \neq i_d$, $V_{GS} \neq v_{gs}$, $V_{DS} \neq v_{ds}$

Recall: I_D, V_{GS},V_{DS} correspond to the Q point and are DC values. There are also constants for a given circuit

Note: i_d, v_{gs}, v_{ds} are small AC signals that result from the interaction of V_{in} and the circuit. They cannot be derived from DC analysis. Remember they are variables whose values are not meant to be determined.

Constants: g_m and r_o (defined at the operating point Q)

Variables: i_d, v_{gs}, v_{ds}

$$\boldsymbol{i_d = g_m v_{gs} + v_{ds}/r_o}$$

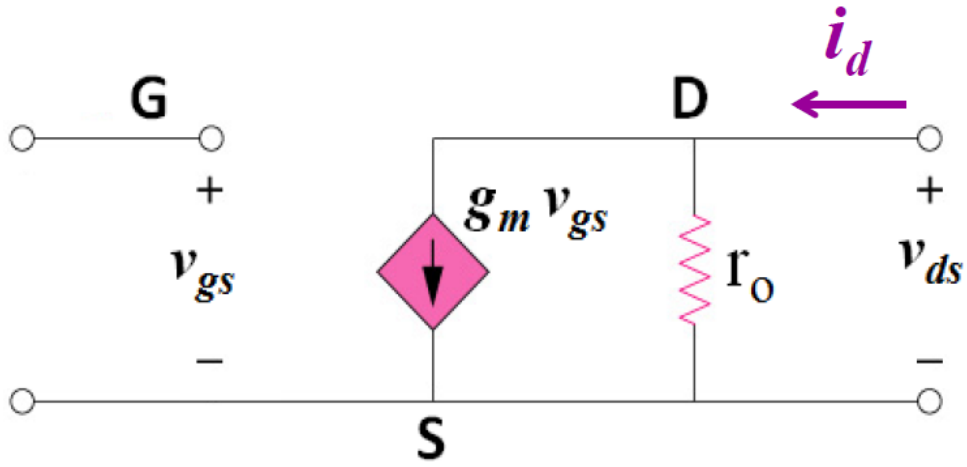
$$\Delta I_D = \left.\frac{\partial I_D}{\partial V_{GS}}\right|_{V_{DS}} \Delta V_{GS} + \left.\frac{\partial I_D}{\partial V_{DS}}\right|_{V_{GS}} \Delta V_{DS}$$

Small signal equivalent circuit (SSEC)

The chain rule equation is now a simple KCL equation:

$$i_d = g_m v_{gs} + v_{ds}/r_o$$

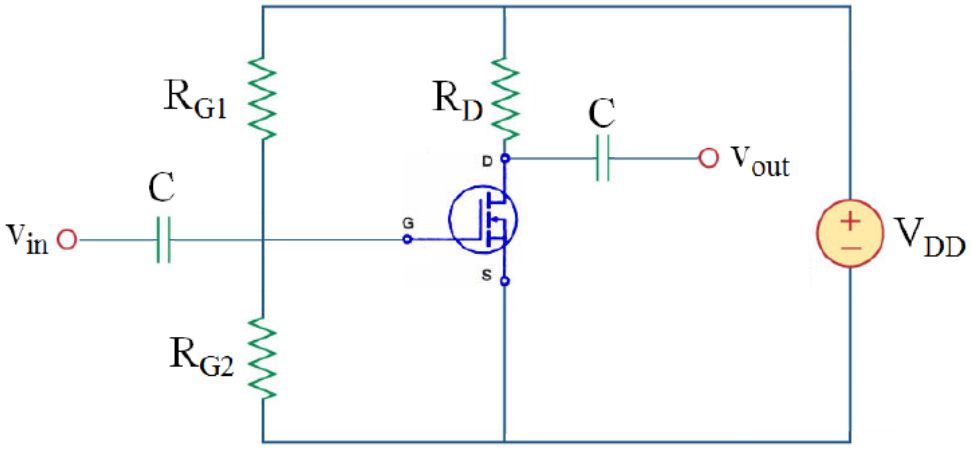
$$r_o = \left.\frac{\partial V_{DS}}{\partial I_D}\right|_{V_{GS}}, \quad g_m = \left.\frac{\partial I_D}{\partial V_{GS}}\right|_{V_{DS}}$$



NOTE: The SSEC only shows variables that are small signals.
DC values have no small signals and thus assigned to be ZERO.

The result is the small signal equivalent circuit that describes how small changes in the voltages results in changes in the drain current

Common source amplifier: Biasing



Given V_{DD} = 8V, I_D = 4.5 mA, R_D = 1 kΩ

First we consider the operation point (Q) where we apply DC analysis:

At the input side:

$V_G = [R_{G2}/(R_{G1}+R_{G2})]*V_{DD}$

At the output side:

$V_{DD} = V_{DS} + I_D*R_D$

So V_{DS} =

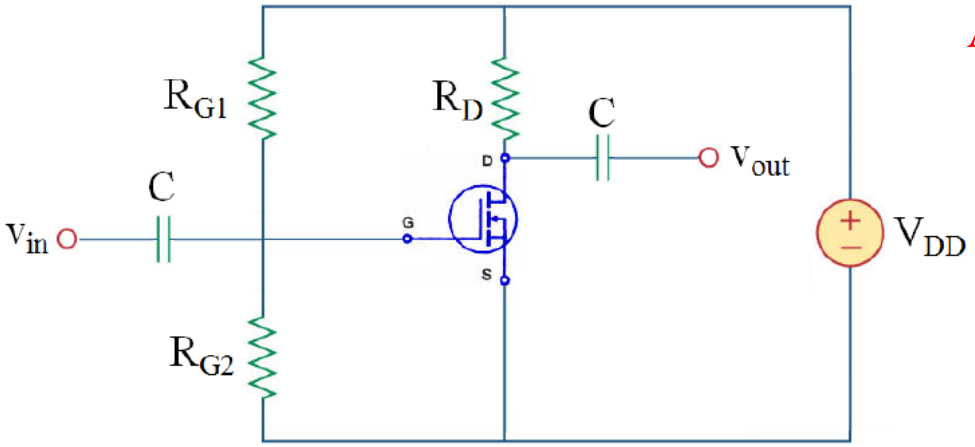
The capacitors is known as decoupling capacitors, and their job is to allow AC signals to go and in out of the amplifier circuit without interfering with the biasing point.

At the input side: AC voltage can be applied on the resistors R_{G1} and R_{G2} but the DC current that flows through R_{G1} and R_{G2} cannot flow out towards V_{in}.

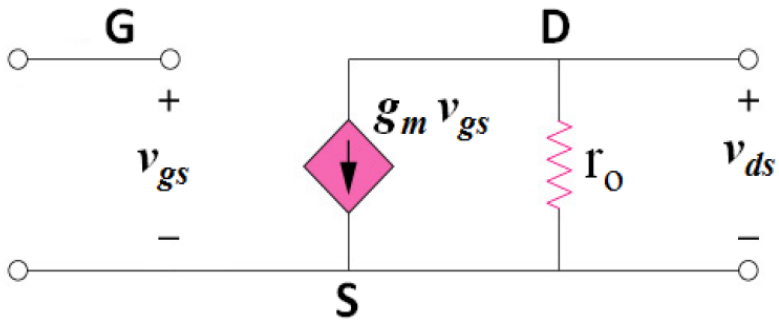
At the output side: The DC voltage across V_{DS} is filtered out through the capacitor, thus V_{out} is simply an AC voltage. The current through R_D cannot flow out into V_{out} through C.

Common source amplifier: Small signals

Next we will derive the SSEC for the circuit that we previously analyzed using a graph



Always start by drawing the FET SSEC



V_{DD}, as a DC source, is shorted to ground

At output: $v_{out} = -g_m v_{gs}(r_o \parallel R_D)$

At input: $v_{in} = v_{gs}$

Combined: $v_{out}/v_{in} = -g_m(r_o \parallel R_D)$

Note: We never attempted to find the value of the variables (v_{gs}, v_{ds})

Gain depends on the size of g_m (FET) and R_D (circuit) where $r_o >> R_D$