# 4.1 Device Structure and Physical Operation

**Reading Assignment:** *pp. 235-248*

**Chapter 4** covers **Field Effect Transistors** (          )

Specifically,  Metal Oxide Semiconductor Field Effect Transistors (              ).

We will study 2 types of MOSFETs:

    **1.**

    **2.**

Each of these types can likewise be an *n*-channel MOSFET (          ), or a *p*-channel MOSFET (          ).

**A.**   NMOS Enhancement Structure

**Q:**

**A:** HO: The Structure of an NMOS Enhancement FET

## B.  The Induced Channel

_____

**HO: Creating a Channel for Current Flow**

## C. The Modes of a MOSFET Transistor

Like a junction diode, we find that the behavior of a MOSFET is best described in terms of three operating modes:
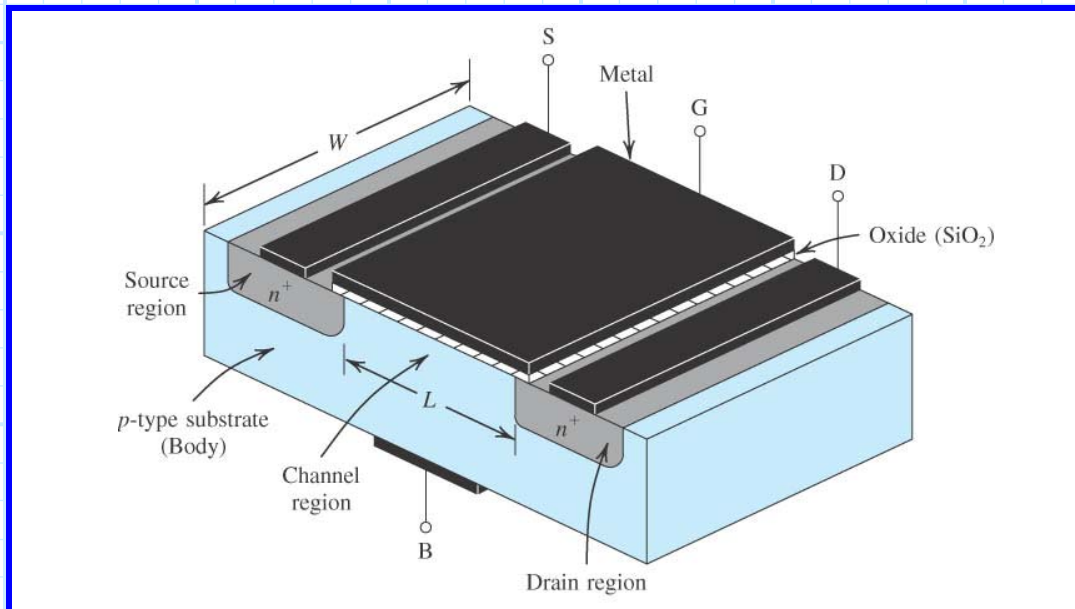
1.
2.
3.

**HO: Applying a Drain Voltage to an NMOS Device**

## D. The p-Channel Enhancement MOSFET

Another type of MOSFET is the p-channel (PMOS) device.

**HO: PMOS and CMOS**

# The Structure of an NMOS Enhancement FET



An **NMOS Enhancement FET** is a **FOUR** terminal device!

Moreover, each terminal has a specific **name**:

1. Source (**S**)
2. Drain (**D**)
3. Gate (**G**)
4. Body (**B**)

Each terminal is associated with a **metal electrode** that is attached to the semiconductor device.

\* The **Body** electrode is connected directly to the $p$-type substrate.

**\*** Two **heavily** doped $n$-type "wells" are implanted into the $p$-type substrate. The **Source** and **Drain** electrodes are each connected to one of these $n^+$ wells.

**\*** The region between these $n^+$ wells is called the **channel**. The channel has **two** important geometries—**channel width** $W$, and **channel length** $L$.
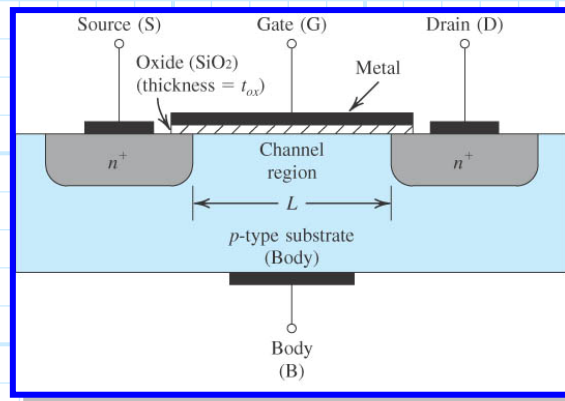


**\*** **Typical** values for channel **length** $L$ are 0.1 to 3 μm (1 μm is 0.001 millimeter!), while channel **width** $W$ is typically 0.2 to 100 μm.

**\*** The **Gate** electrode rests on top of the channel, but is **not** connected directly to it.  Instead, the channel and gate electrode are **separated** by a thin (e.g., 2-5 nm) layer of Silicon Dioxide (SiO₂).

**\*** Silicon Dioxide is essentially glass!  Glass is a very good **insulator**—thus, no current can flow from the gate into the MOSFET device!

**\*** Thus, the Silicon Dioxide layer is **sandwiched** between the metal Gate electrode and the p-type channel.  It is these **three** materials that give the **MOS**FET its name—**M**etal (Gate electrode) **O**xide (SiO₂) **S**emiconductor ($p$-type channel) FET.

# Creating a Channel for Current Flow

When we first look at an **NMOS** device, it appears that **no** current can flow from the Drain electrode to the Source electrode (or vice versa) as we must contend with two **p-n junctions**!
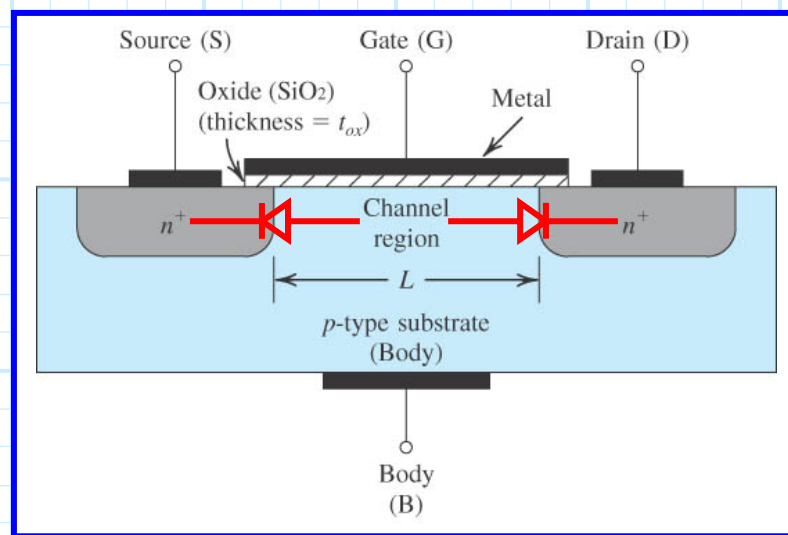


**\*** Current seemingly cannot flow **into** channel from the Drain, as this would require current flowing from an $n$-type (cathode) region into a $p$-type (anode) region.

**\*** Likewise, current cannot flow **into** channel from the Source, as this would require current flowing from an $n$-type (cathode) region into a $p$-type (anode) region.

**\*** Recall that we have previously determined that current **cannot** flow into (or out of) the channel from (into) the **gate**, as the SiO2 layer is a very good **insulator**!

**Q:** *Pardon me, but this NMOS device does **not** appear to be particularly **useful**. I mean, what good is a device if **no** current can flow into it?*

**A:** An NMOS device would indeed be useless **if** no current could flow from drain to source. However, we can **modify** the channel so that this current **can** indeed flow!

We must **induce a channel**—that is, create a thin layer of *n*-type Si connecting the source and drain!

To do this, we place a **positive voltage** at the **gate** electrode. This creates an **electric field** within the *p*-type substrate, which pushes the positively charged holes in the *p*-type substrate away from the gate electrode—a **depletion region** is formed in the Silicon under the gate!
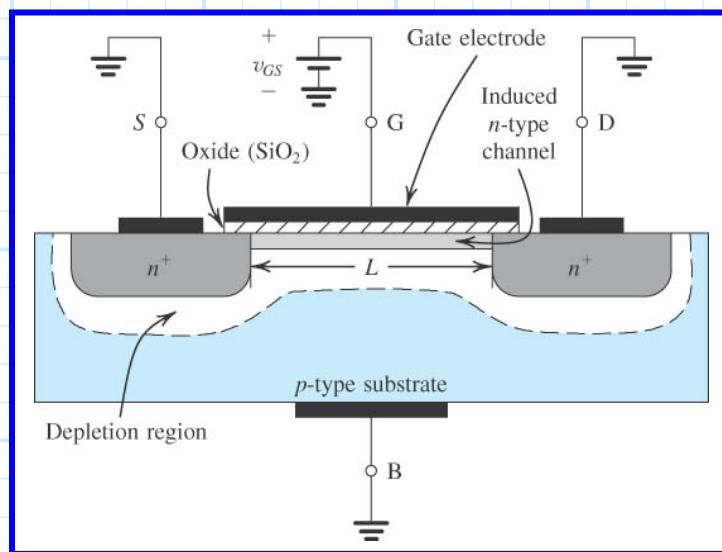
The electric field under the gate electrode will **repel** positively charged holes, but will **attract** negatively charged free electrons!

> **Q:** *I see! The minority carriers in the p-type substrate (i.e., **free electrons**) are attracted to the **gate** electrode!*
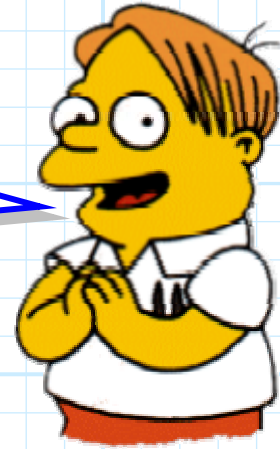
**A:** True! But we also find that many of the free electrons attracted to the gate come from the **heavily doped** $n^+$ wells under the source and drain electrodes.

\*   Of course, there is a Silicon Dioxide **insulator** separating the gate electrode and the Silicon substrate, so the free-electrons attracted by the gate electrode simply "**pile up**" at the **top** of the Silicon substrate, just **under** the $SiO_2$ layer.

\*   The result is an "**inversion layer**"—A **thin** layer in the *p*-type silicon where the majority carriers are actually **free electrons**!

\*   This inversion layer forms a ***n*-type conducting channel** connecting the *n+* Silicon well under the **drain** to the *n+* Silicon well under the **source**.  By applying a positive voltage to the gate, we have **induced a conducting channel**!

In other words, current flowing from drain to source **no longer** encounters any ***p-n* junctions**!

**Q:** *So, will **any** positive gate voltage suffice for inducing a channel, **or** must this gate voltage be somehow sufficiently **large**?*

**A:** The later. The gate voltage must be **sufficiently large** to create an inversion layer—it must be sufficiently large to **induce** a conducting channel.

In fact, the voltage value must exceed some **threshold.**

First some **definitions**:

$v_G$ =  The gate electrode potential with respect to **ground**.

$v_S$ = The source electrode potential with respect to **ground**.

$v_{GS}$ = $v_G$ - $v_S$ = The gate electrode potential with respect to the **source**.

We find that for a channel to be induced with in an **NMOS** device, the voltage $v_{GS}$ must exceed a **threshold voltage**:

$$v_{GS} > V_t \qquad \text{to induce an NMOS channel}$$

Moreover, we find that the amount by which $v_{GS}$ **exceeds** the threshold voltage is a **very important** parameter for determining NMOS behavior.  We call this value the **excess gate voltage**—this value is **very important**!

$$v_{GS} - V_t \doteq \text{excess gate voltage}$$

Thus, we can say:

$$v_{GS} - V_t > 0 \quad \text{to induce an NMOS channel}$$

# Applying a Drain Voltage to an NMOS Device

Say we apply a voltage at the **gate** of an NMOS device that is sufficiently large to **induce** a conducting channel (i.e., $v_{GS} - V_t > 0$).

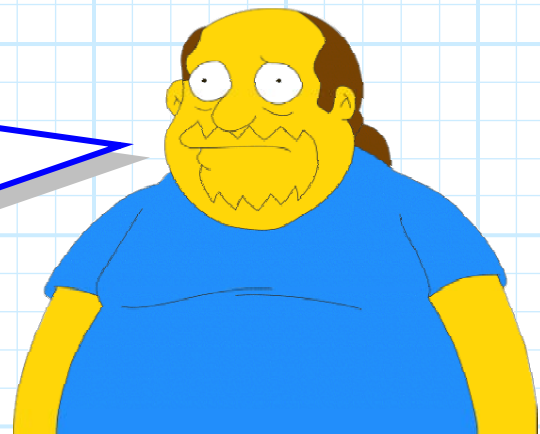Now, say that we additionally place a voltage at the NMOS **drain** electrode, such that:

$$v_{DS} > 0$$

where:

$$v_{DS} = v_D - v_S \doteq \text{Drain-to-Source Voltage}$$

Now guess what happens—**current** begins to flow through the **induced channel**!

**Q:**  *Current! I thought current could **not** flow because of the two p-n junctions in the NMOS device!*

**A:**  Remember, that was **before** we applied a sufficient **gate voltage**.  With this voltage applied, an *n*-type channel is **induced,** forming a **conducting channel** from drain to source!

Recall that because of the SiO$_2$ layer, the gate current is **zero** (i.e., $i_G = 0$).

Thus, all current **entering the drain** will **exit the source**. We therefore conclude that:

$$i_S = i_D$$

As a result, we refer to the channel current for NMOS devices as simply the **drain current** $i_D$.
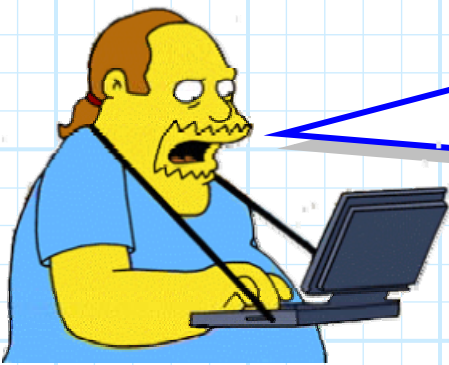


**Q:** *So, I see that you have now defined current $i_D$ and voltages $v_{GS}$ and $v_{DS}$. Just how are these parameters **related**?*

**A:** First, we find that an **increasing** $v_{GS}$ or, more specifically, an increasing **excess gate voltage** $v_{GS}-V_t$ will result in a **higher channel conductivity** (in other words, a lower channel resistivity).
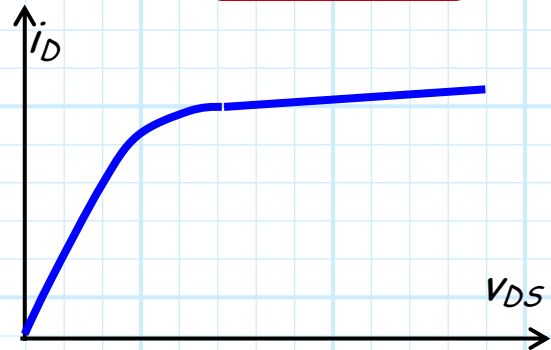
Thus, we find that the **drain current** $i_D$ will **increase** as a positive **excess gate voltage** $v_{GS}-V_t$ **increases** (assuming that $v_{DS} > 0$).

This process, of increasing the induced channel conductivity by increasing the excess gate voltage, is otherwise known as **channel enhancement**. This is where the **enhancement MOSFET** gets its name!

**Q:** *OK, but what about the relationship between drain current $i_D$ and voltage $v_{DS}$?*
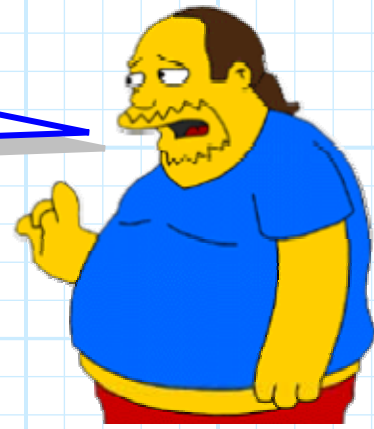
**A:** This relationship is a little **complicated**! Generally speaking, however, a **positive** $v_{DS}$ results in a **positive** $i_D$, and the **larger** the $v_{DS}$, the **larger** the drain current $i_D$.

More specifically, we find that when $v_{DS}$ is **small** (we'll see how small later), the drain current will be **directly proportional** to the voltage drain to source $v_{DS}$.

$$i_D \propto v_{DS} \quad \textbf{if } v_{DS} \textbf{ small}$$

In other words, if $v_{DS}$ is **zero**, the drain current $i_D$ is **zero**. Or, if the voltage $v_{DS}$ **increases** by 10%, the drain current will likewise **increase** by 10%. Note this is **just like a resistor**!

$$i = v\left(\frac{1}{R}\right) \qquad \therefore \ i \propto v$$

Thus, **if** (and **only** if!) $v_{DS}$ is small, the induced channel behaves **like a resistor**—the current through the channel ($i_D$) is **directly proportional** to the voltage across it ($v_{DS}$).
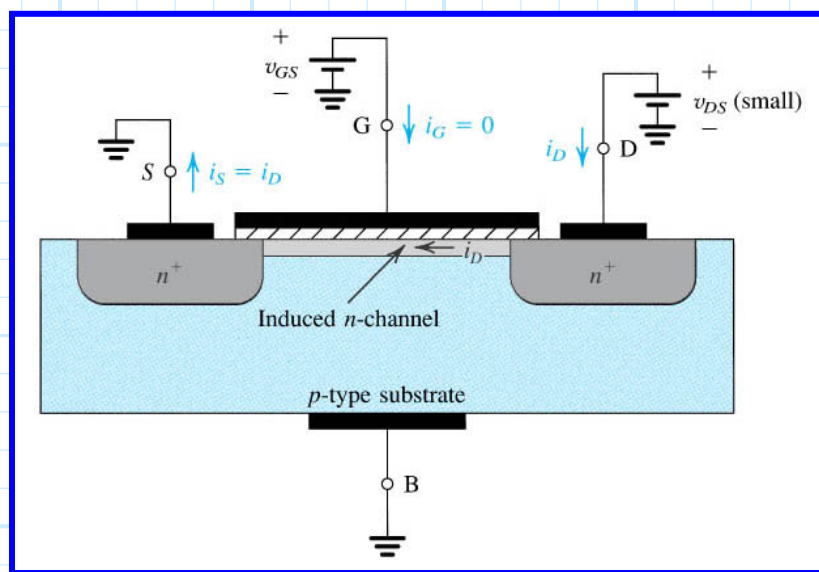
In other words, we can (for small values of $v_{DS}$), define a **channel resistance** $r_{DS}$:

$$\text{Since} \quad i_D \propto v_{DS}, \quad \frac{v_{DS}}{i_D} \doteq r_{DS} \quad \textbf{(if } v_{DS} \textbf{ small)}$$

Note that this resistance value depends on the **conductivity** of the induced channel—which in turn is dependent on the **excess gate voltage**!

Thus, if we were to **plot** drain current $i_D$ versus $v_{DS}$ for various excess gate voltages, we would see something like this:
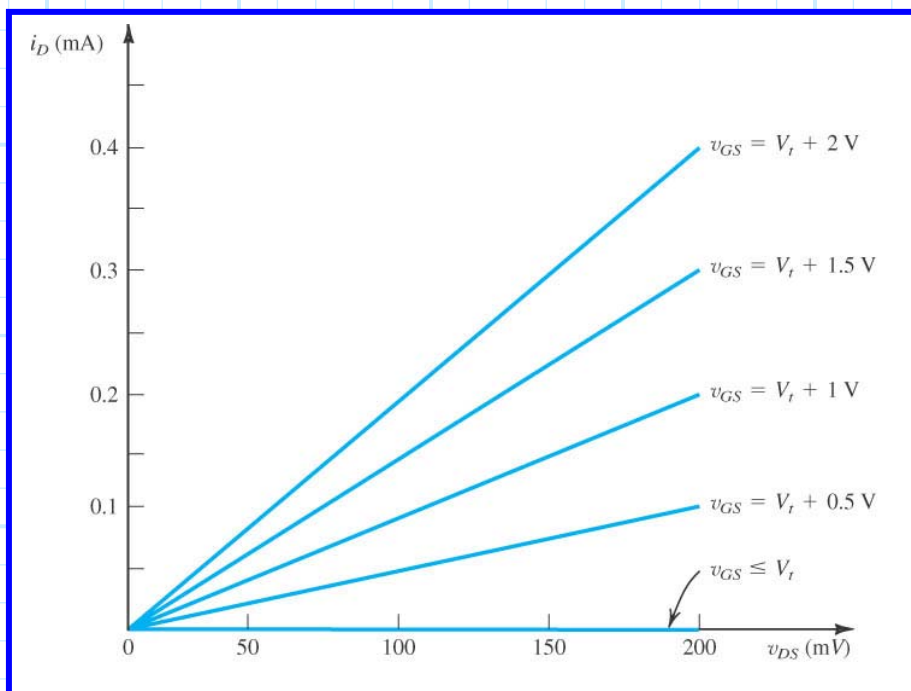
**Q:** *Yawn! It is apparent that an NMOS transistor is so* **simple** *that virtually any intergalactic traveler should be able to understand it. It's just a* **voltage controlled resistor**—*right?*
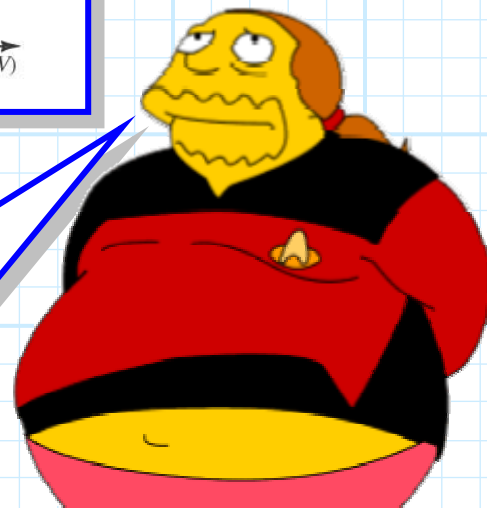
**A:** **WRONG!** Remember, channel resistance $r_{DS}$ **only** has meaning if $v_{DS}$ is **small**—and most often $v_{DS}$ will **not** be small!

As $v_{DS}$ **increases** from our presumably small value, we find that **strange things** start to happen in our **channel**!

Recall that primarily, the **free-electrons** in our inversion layer (the induced channel) were attracted to the **gate** from the heavily doped **n+ Silicon** regions under the **drain** and source.

But the **gate** now has **competition** in attracting these free electrons!

It was "easy" to attract free electrons to the gate when the **gate electrode voltage** was much **larger** than both the drain and source voltage (i.e., when $v_{GS} \gg v_{DS}$). But as the **drain** voltage increases, it begins to attract **free electrons** of its **own**!

Recall that **positive current entering** the drain will actually consist mainly of **free electrons exiting** the drain! As a result, the **concentration** of free-electrons in our inversion layer will begin to **decrease** in the vicinity of the **drain**.

In other words, **increasing $v_{DS}$** will result in **decreasing** channel **conductivity**!
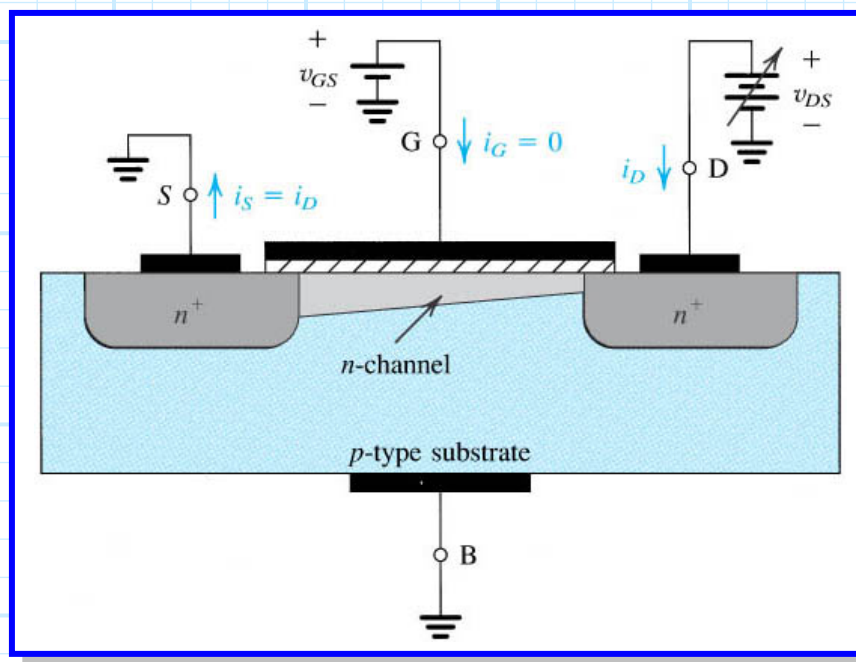
Thus, increasing the $v_{DS}$ will have **two effects** on the NMOS device:

> 1. Increasing $v_{DS}$ will **increase the potential difference** (voltage) across the conducting channel, an effect that works to **increase** the drain current $i_D$
>
> 2. Increasing $v_{DS}$ will **decrease the conductivity** of the induced channel, and effect that works to **decrease** the drain current $i_D$.

For **small** values of $v_{DS}$, the second effect is **tiny**, so that the increase in drain current is directly **proportional** to the increase in voltage $v_{DS}$ (hence, we can define channel **resistance** $r_{DS}$). For example, a **10% increase** in $v_{DS}$ will result in a **10% increase** in drain current.
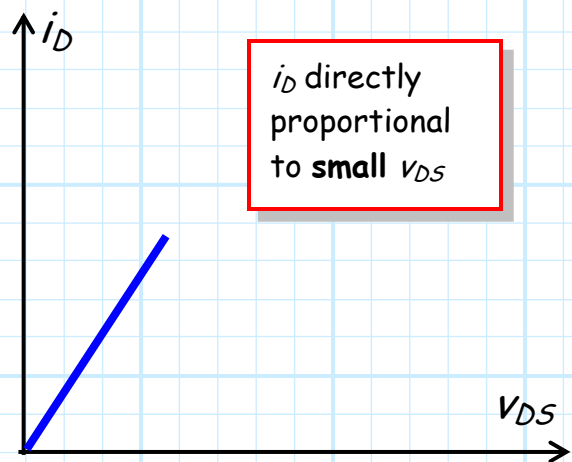


However, as $v_{DS}$ increases, the **second effect** will become more and **more pronounced**. We find then that the drain current will **no longer** be directly **proportional** to the voltage $v_{DS}$. The reduction in channel conductivity will begin to "**counteract**" the increase in potential across the channel.

For example, a **10% increase** in $v_{DS}$ may result in only a **9% increase** in $i_D$. Likewise, if we increase $v_{DS}$ another 10%, the drain current may then increase **only 8%** (and so on).

Eventually, we find that the an increase in $v_{DS}$ will result in **no** further increase drain current $i_D$!! Effect 2 will **completely** "counteract" effect 1, so that there is **no more** increase in drain current as $v_{DS}$ increases.
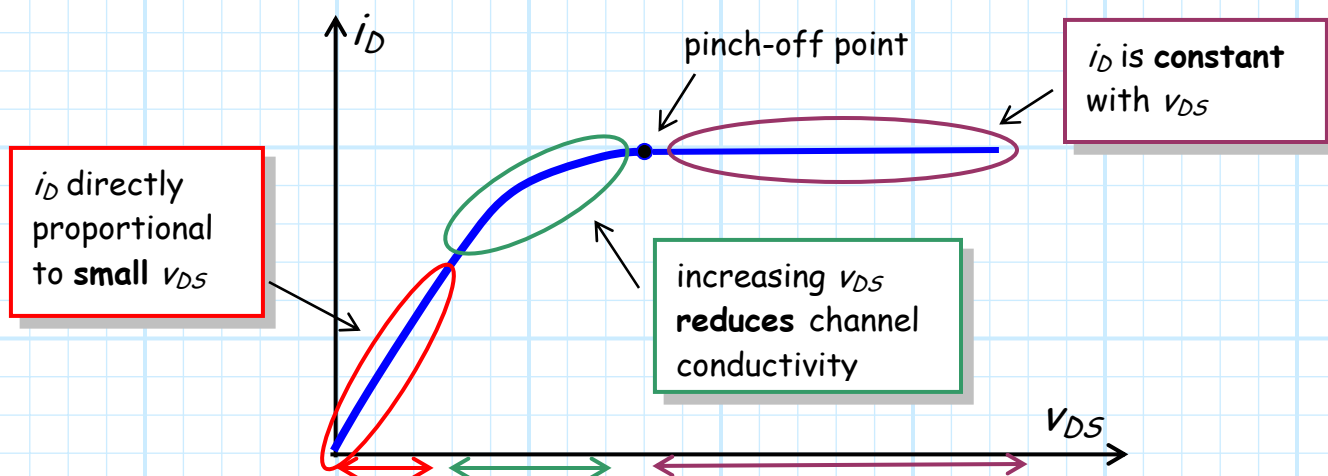
increasing $v_{DS}$ **reduces** channel conductivity

$i_D$ directly proportional to **small** $v_{DS}$
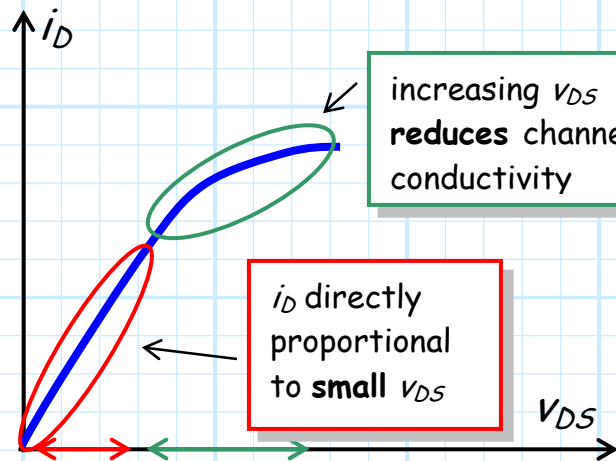
When this occurs, we say that we have "**pinched-off**" the induced channel—in other words the **channel is in pinch off**.
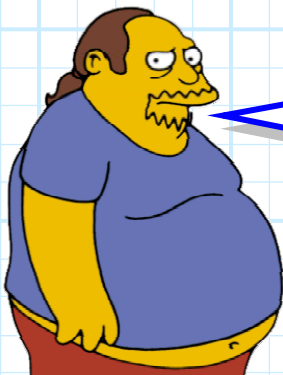
**Q:** *So, if we **continue** to increase $v_{DS}$ after the channel is "**pinched off**", does the drain current actually begin to **decrease**?*

**A:** **NO!** A **interesting** thing happens when the channel is in pinch off. As we further increase $v_{DS}$, the drain current $i_D$ will remain **unchanged** (approximately)! That is, the drain current will be a **constant** (approximately) with respect to $v_{DS}$.

pinch-off point

$i_D$ is **constant** with $v_{DS}$

$i_D$ directly proportional to **small** $v_{DS}$

increasing $v_{DS}$ **reduces** channel conductivity

Note that there are **three distinct channel conditions** in for NMOS operation.

* Depending on the value of $v_{GS}$, we can have an **induced channel,** or **no** conducting channel at all!

* Then **if** we have an **induced channel** (i.e., $v_{GS} - V_t > 0$), (depending on the value of $v_{DS}$) the channel can be either be **pinched-off** or **not**!

Each of these **three** possibilities has a **name**—they are the names of our **NMOS transistor modes**!

1. **Cutoff** - When $v_{GS} - V_t < 0$, **no** channel is induced (no inversion layer is created), and so $i_D$=0. We call this mode **CUTOFF**.

2. **Triode** - When an induced channel **is** present (i.e., $v_{GS} - V_t > 0$), but the value of $v_{DS}$ is **not** large enough to pinch-off this channel, the NMOS is said to be in **TRIODE** mode.

3. **Saturation** - When an induced channel **is** present (i.e., $v_{GS} - V_t > 0$), and the value of $v_{DS}$ **is** large enough to pinch-off this channel, the NMOS is said to be in **SATURATION** mode.

We can **summarize** these modes in a table:

| MODE | INDUCED CHANNEL? | CHANNEL PINCH-OFF? |
|---|---|---|
| **CUTOFF** | NO | N/A |
| **TRIODE** | YES | NO |
| **SATURATION** | YES | YES |
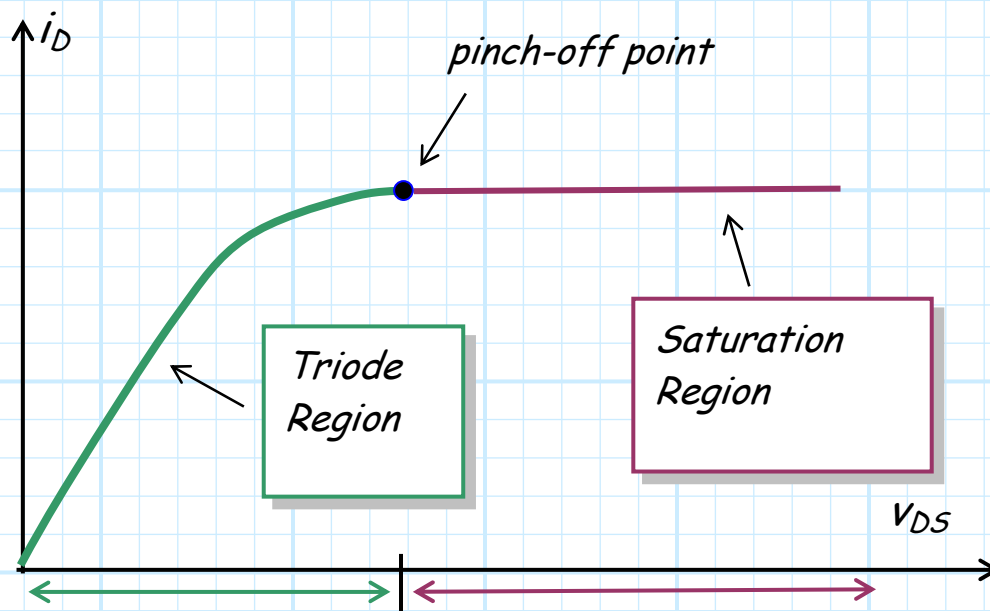
# PMOS and CMOS

In addition to an *n*-channel MOSFET device (i.e., NMOS), we can build *p*-channel MOSFET (i.e., **PMOS**) device.

The structure of a PMOS device is essentially the same as an NMOS transistor, except that wherever there **was *n*-type** Silicon there is **now *p*-type** Silicon—and wherever there **was *p*-type** Silicon there is **now *n*-type** Silicon!

Specifically, the PMOS channel is part of a ***n*-type substrate** lying between two heavily doped ***p+* wells** beneath the source and drain electrodes.

Generally speaking, a PMOS transistor is only constructed in **consort** with an NMOS transistor.  This "pair" of NMOS and PMOS transistors is known as **C**omplementary **MOS**FETs— **CMOS** for short!
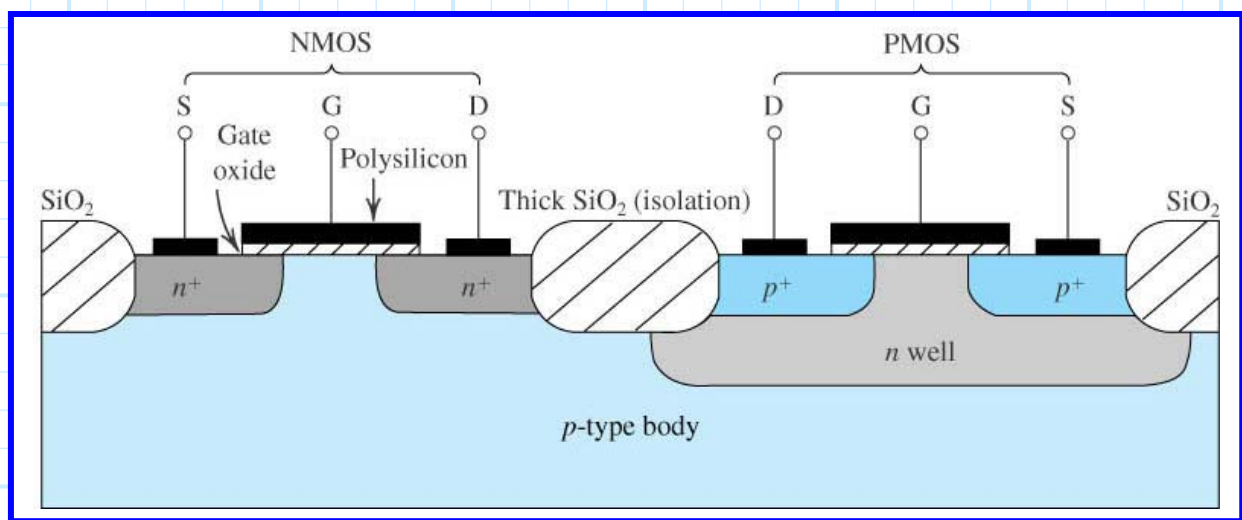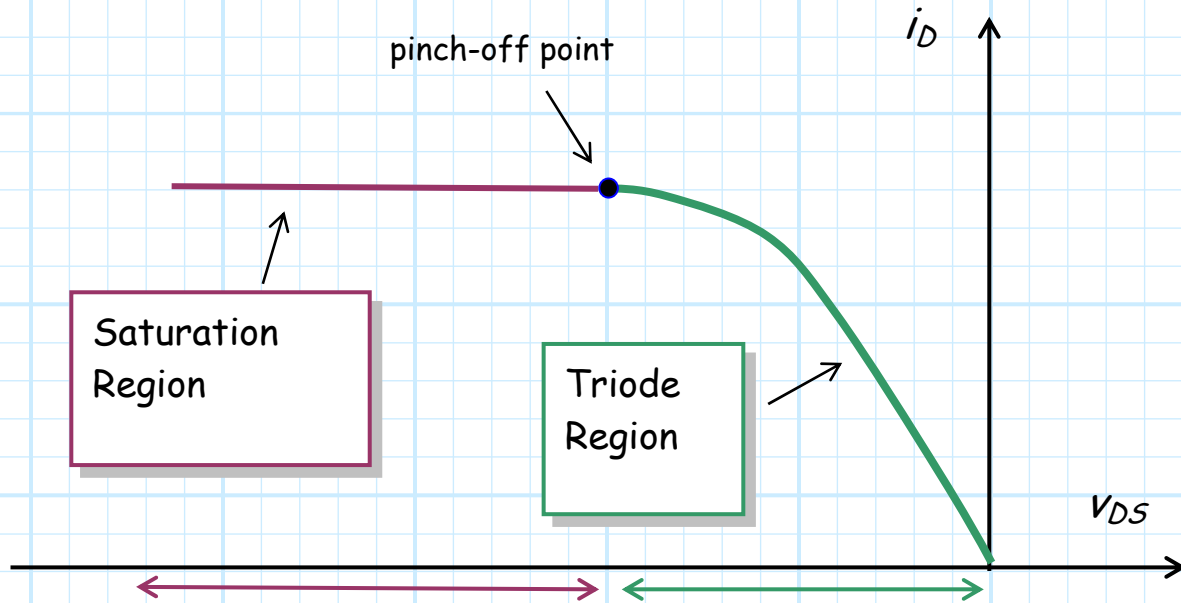
The operation of a PMOS transistor is in many ways **similar** to that of the NMOS device, but in many ways they are also **quite different**!

For example, for a **PMOS** device we find:

* To create an inversion layer in the **n-type** substrate, we must attract **holes** to the gate electrode.

* As a result, a **p-type** channel will be induced, connecting the p+ wells at the drain and the source.

* However, to attract **holes** toward the gate, the voltage $v_{GS}$ must be sufficiently **negative**! The threshold voltage $V_t$ is thus a **negative** value, so that a channel is induced only if $v_{GS} < V_t$ (i.e., $v_{GS}$ is more **negative** than $V_t^I$).

* As a result, a channel is induced in a **PMOS** device **only** if the excess gate voltage $v_{GS} - V_t$ is **negative** (i.e., $v_{GS} - V_t < 0$).

* Likewise, we find that we typically get current to flow through this channel by making the voltage $v_{DS}$ **negative**. If we make the voltage $v_{DS}$ sufficiently **negative**, the **p-type** induced channel will **pinch off**.

* Note that when $v_{DS}$ is **negative**, the drain current will flow **from** the PMOS **source, to** the PMOS **drain** (i.e., exactly **opposite** that of the NMOS device with a positive $v_{DS}$).

* Thus, for a **PMOS** device, we **define** current flowing from source to drain as **positive current**((i.e., exactly **opposite** that of the NMOS device).

pinch-off point

$i_D$

Saturation Region

Triode Region

$v_{DS}$

The **PMOS** $i_D$ vs. $v_{DS}$ Curve