JOINT INSTITUTE
交大密西根学院

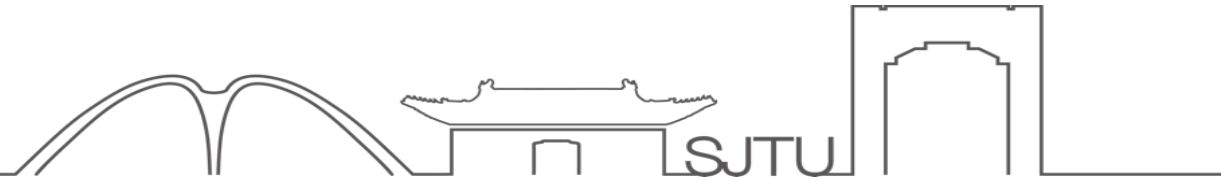
ECE3110J/VE311 Electronic Circuits

MOSFET

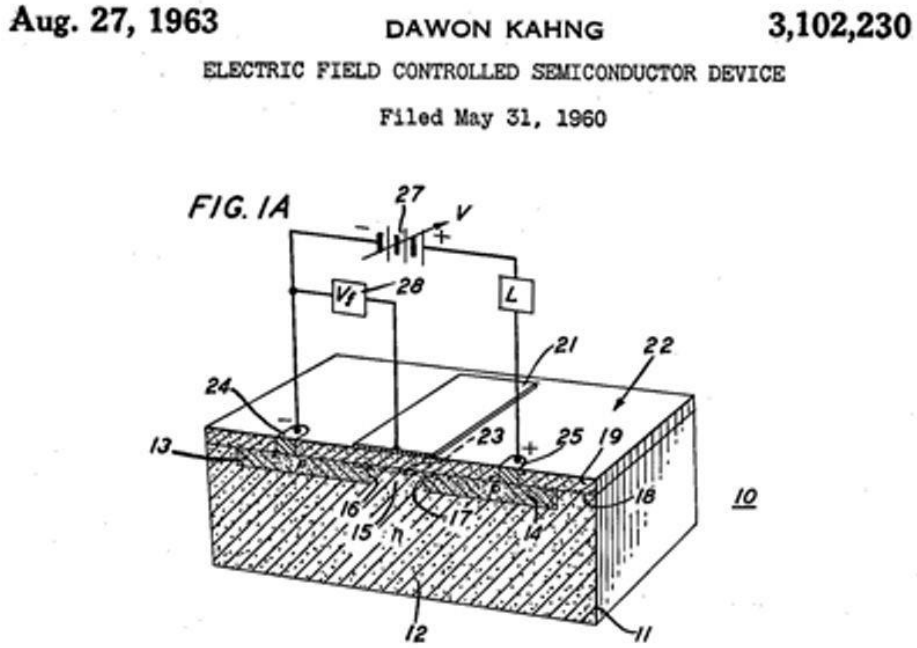
Basic MOS Device Physics

Design of Analog CMOS Integrated Circuits, Chapter 2
Fundamentals of Microelectronics, Chapter 6

Yuljae Cho, *PhD*
Associate Professor
Joint Institute, SJTU



*General Introduction



Aug. 27, 1963 DAWON KAHNG 3,102,230

ELECTRIC FIELD CONTROLLED SEMICONDUCTOR DEVICE

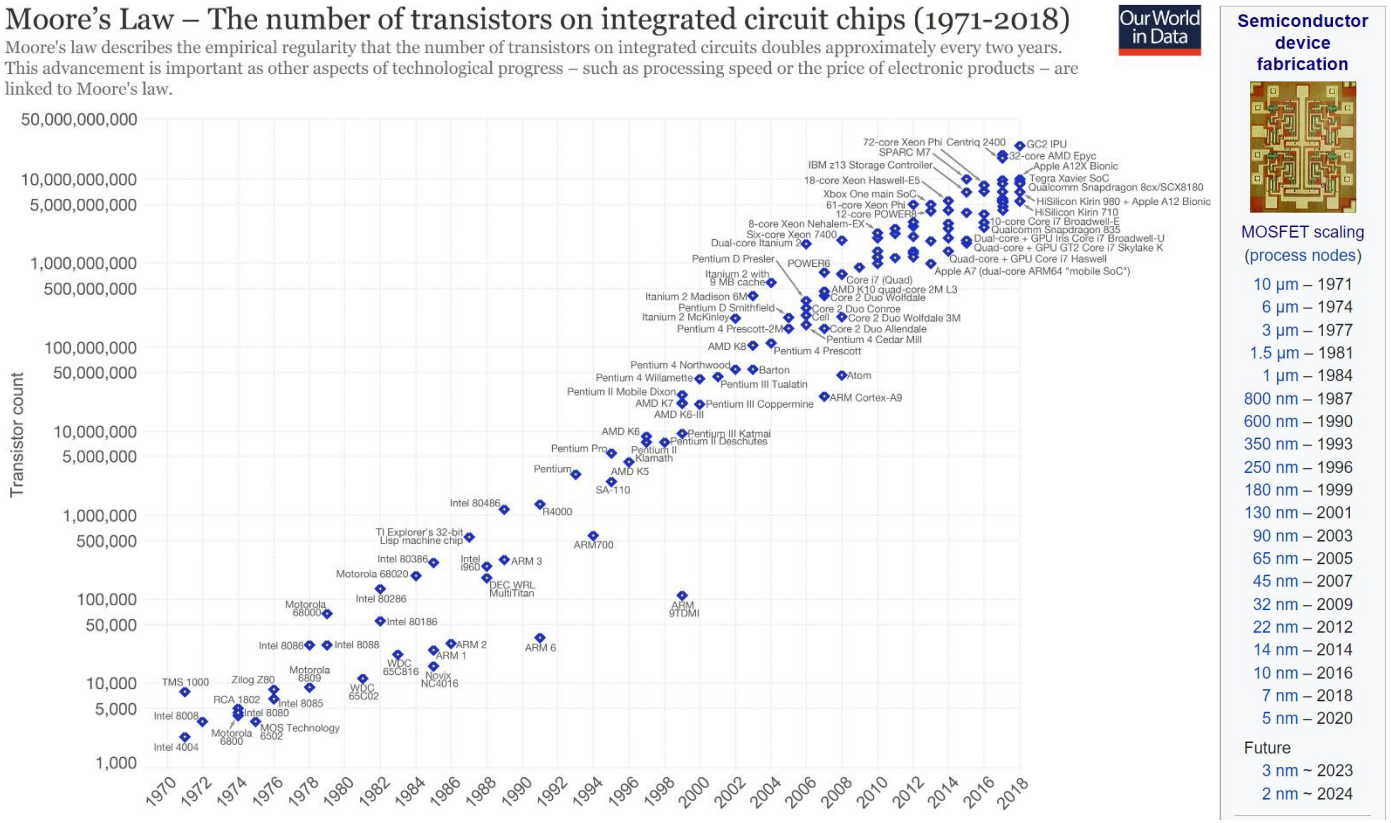
Filed May 31, 1960

FIG. 1A

US Patent Office

In 1959 M. M. (John) Atalla and Dawon Kahng at Bell Labs achieved the first successful insulated-gate field-effect transistor (FET).

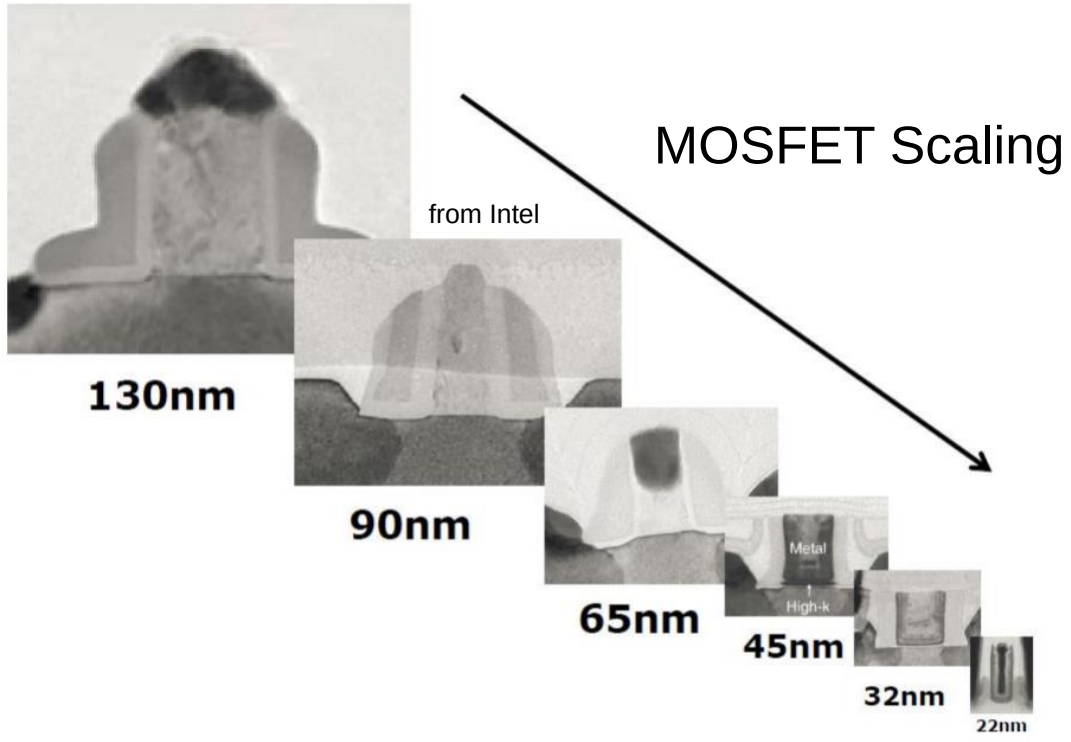
History of Electronics



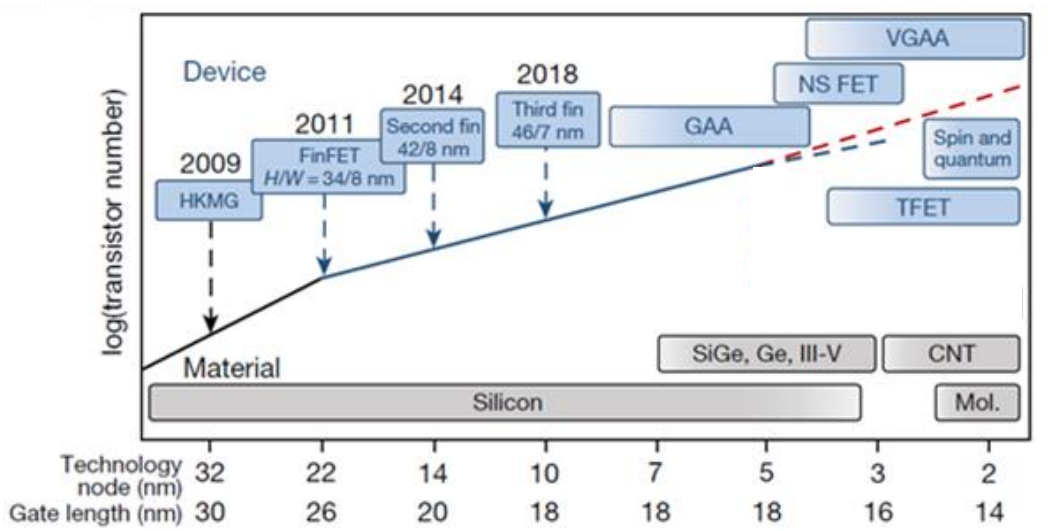
Since the commercial introduction of the integrated circuit, increases in transistors' density have been achieved through a continued reduction in **minimum feature size.** Many companies in a semiconductor industry around the world are actively working on below 3 nm processes.

MOSFET Scaling

Although we have enjoyed significant performance improvement through scaling so far, we **have reached practical limitations**, and it is hard to keep on track with ideal scaling with respect to the “constant-field” perspective.

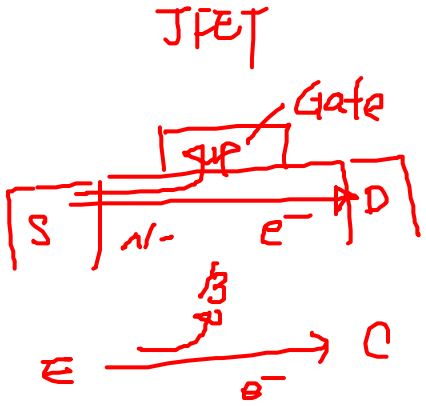


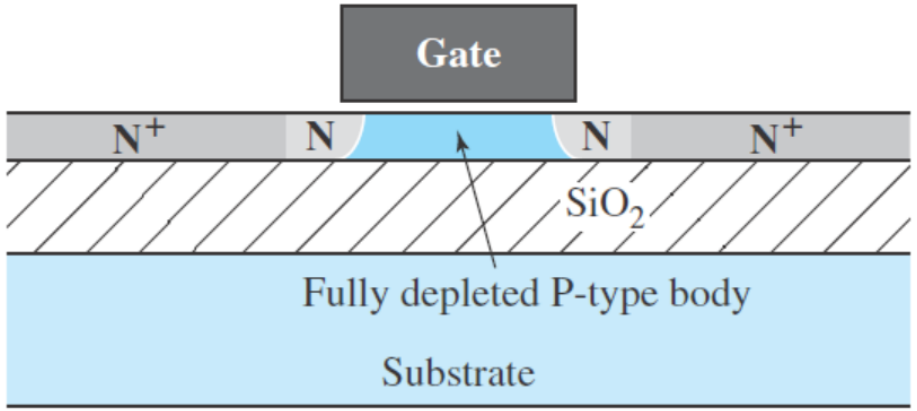
Nature, Vol 591, **43 (2021)**

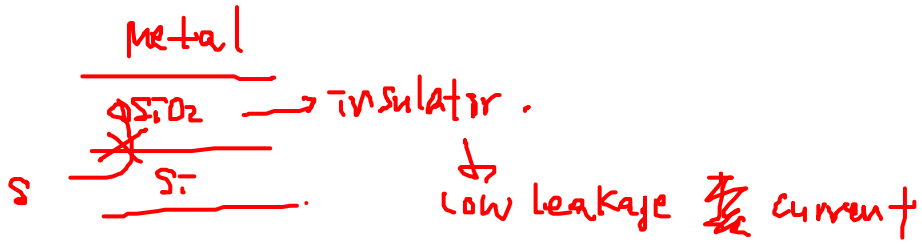


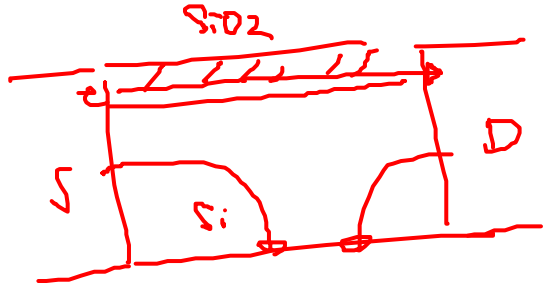
Advanced MOSFETs

- **Silicon-on-Insulator (SOI) MOSFETs**



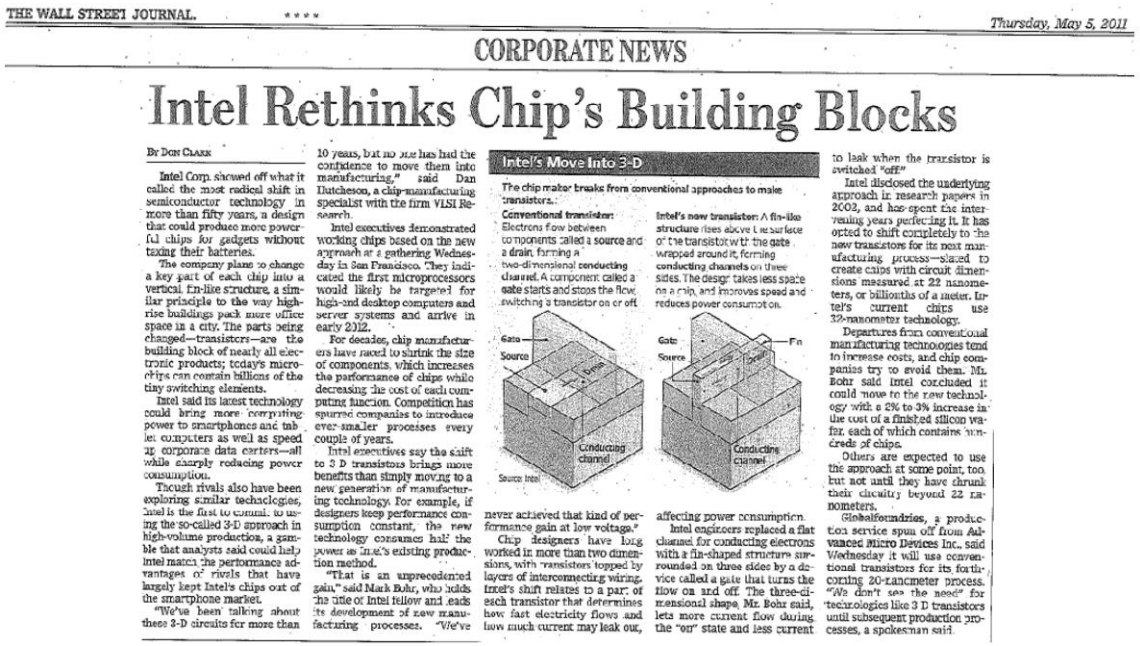






The SOI technology provides a thin monocrystalline film of silicon on an insulating layer (typically, SiO_2). Because no current can flow through the buried SiO_2 layer, much better control of the leakage current is possible in SOI MOSFETs. The top silicon film has to be very thin so that it is fully depleted. These MOSFETs are also called **ultra-thin-body (UTB) MOSFETs**.

- **Double-Gate MOSFETs: FinFET**



THE WALL STREET JOURNAL. Thursday, May 5, 2011

CORPORATE NEWS

Intel Rethinks Chip's Building Blocks

By Don Clark

Intel Corp. showed off what it called the most radical shift in semiconductor technology in more than fifty years, a design that could produce more powerful chips for gadgets without taxing their batteries.

The company plans to change a key part of each chip into a vertical, fin-like structure, a similar principle to the way high-rise buildings pack more office space in a city. The parts being changed—transistors—are the building block of nearly all electronic products; today's microchips can contain billions of the tiny switching elements.

Intel said its latest technology could bring more computing power to smartphones and tablet computers as well as speed up corporate data centers—all while sharply reducing power consumption.

Though rivals also have been exploring similar technologies, Intel is the first to commit to using the so-called 3-D approach in high-volume production, a gamble that analysts said could help Intel match the performance advantages of rivals that have largely kept Intel's chips out of the smartphone market.

"We've been talking about these 3-D circuits for more than 10 years, but no one has had the confidence to move them into manufacturing," said Dan Hutcheson, a chip-manufacturing specialist with the firm VLSI Research.

Intel executives demonstrated working chips based on the new approach at a gathering Wednesday in San Francisco. They indicated the first microprocessors would likely be targeted for high-end desktop computers and server systems and arrive in early 2012.

For decades, chip manufacturers have raced to shrink the size of components, which increases the performance of chips while decreasing the cost of each computing function. Competition has spurred companies to introduce ever-smaller processes every couple of years.

Intel executives say the shift to 3-D transistors brings more benefits than simply moving to a new generation of manufacturing technology. For example, if designers keep performance consumption constant, the new technology consumes half the power as Intel's existing production method.

"That is an unprecedented gain," said Mark Bohr, who holds the title of Intel fellow and leads its development of new manufacturing processes. "We've never achieved that kind of performance gain at low voltage."

Chip designers have long worked in more than two dimensions, with transistors topped by layers of interconnecting wiring. Intel's shift relates to a part of each transistor that determines how fast electricity flows and how much current may leak out, affecting power consumption.

Intel engineers replaced a flat channel for conducting electrons with a fin-shaped structure surrounded on three sides by a device called a gate that turns the flow on and off. The three-dimensional shape, Mr. Bohr said, lets more current flow during the "on" state and less current to leak when the transistor is switched "off."

Intel disclosed the underlying approach in research papers in 2002, and has spent the intervening years perfecting it. It has opted to shift completely to the new transistors for its next manufacturing process—sliced to create chips with circuit dimensions measured at 22 nanometers, or billionths of a meter. Intel's current chips use 32-nanometer technology.

Departures from conventional manufacturing technologies tend to increase costs, and chip companies try to avoid them. Mr. Bohr said Intel concluded it could move to the new technology with a 2% to 3% increase in the cost of a finished silicon wafer, each of which contains hundreds of chips.

Others are expected to use the approach at some point, too, but not until they have shrunk their circuitry beyond 22 nanometers.

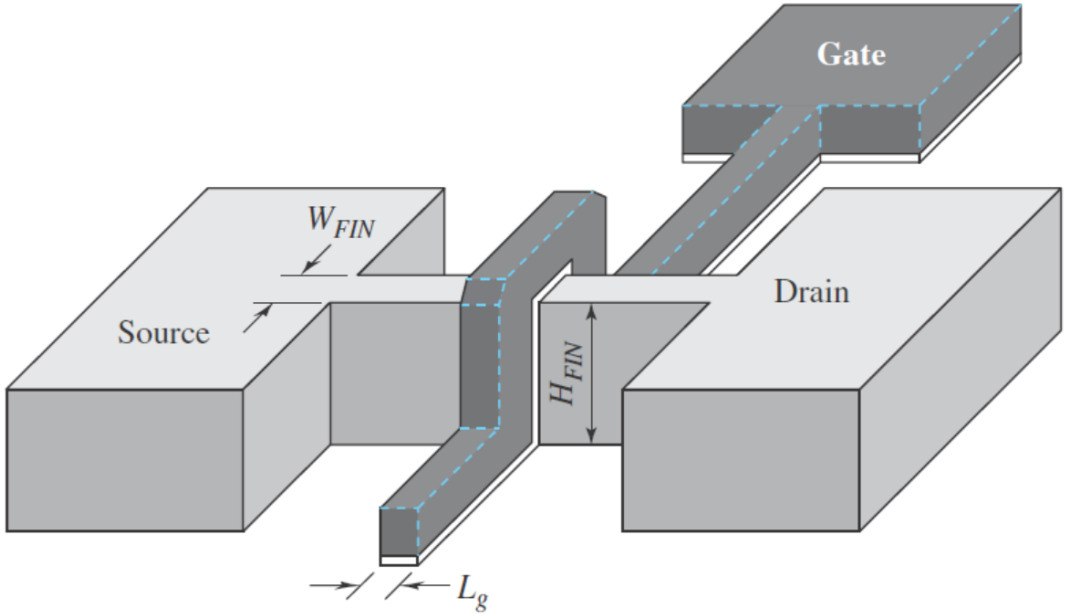
GlobalFoundries, a production service spun off from Advanced Micro Devices Inc., said Wednesday it will use conventional transistors for its forthcoming 20-nanometer process. "We don't see the need" for technologies like 3-D transistors until subsequent production processes, a spokesman said.

Intel's Move Into 3-D

The chip maker breaks from conventional approaches to make transistors.

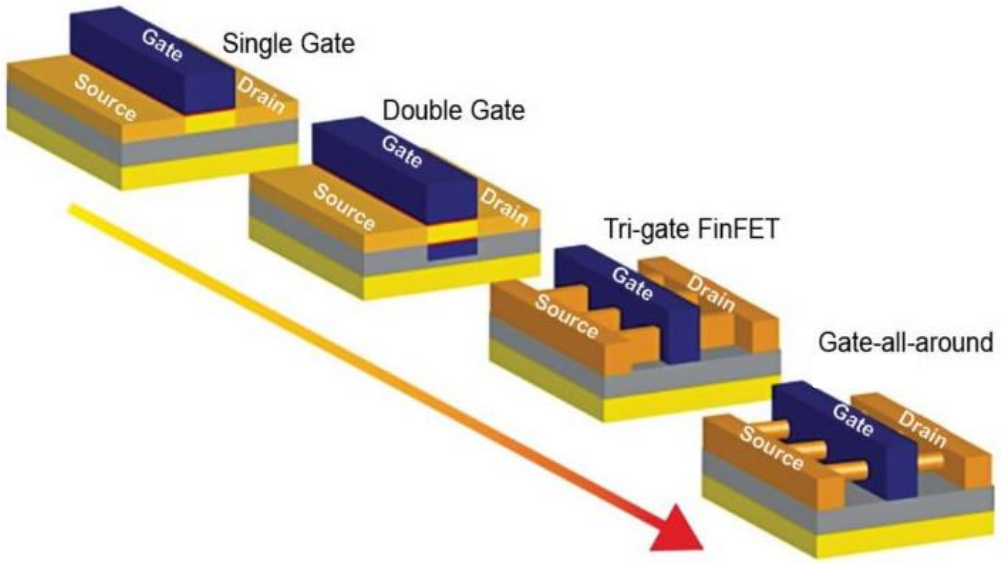
Conventional transistor: Electrons flow between components called a source and a drain, forming a two-dimensional conducting channel. A component called a gate starts and stops the flow, switching a transistor on or off.

Intel's new transistor: A fin-like structure rises above the surface of the transistor with the gate wrapped around it, forming conducting channels on three sides. The design takes less space on a chip, and improves speed and reduces power consumption.



On May 4, 2011, Intel Corporation announced the most radical shift in semiconductor technology in 50 years. **A new 3D transistor design** (double-gate MOSFETs or **FinFET**) will enable the production of integrated-circuit chips that operate faster with less power. Much better control of the MOSFET body in the gate area can be achieved.

Beyond MOSFET and Multi-gate FET?



https://www.electronicsforu.com/

CMOS technology now faces **two problems** that together result in high power consumption.

(1) stopping the rising leakage currents that degrade the switching ratio
(2) increasing difficulty in further reducing the supply voltage.


www.BANDICAM.com
5/4nm
0:00 / 1:46

Boltzmann tyranny

FETs in today's integrated circuits require **at least 60 mV** of gate voltage to increase the current by **one order** of magnitude at room temperature.

→ **The Boltzmann limit hinders** the use of the conventional MOSFETs as a switch for **ultralow supply voltages.**

NANO LETTERS

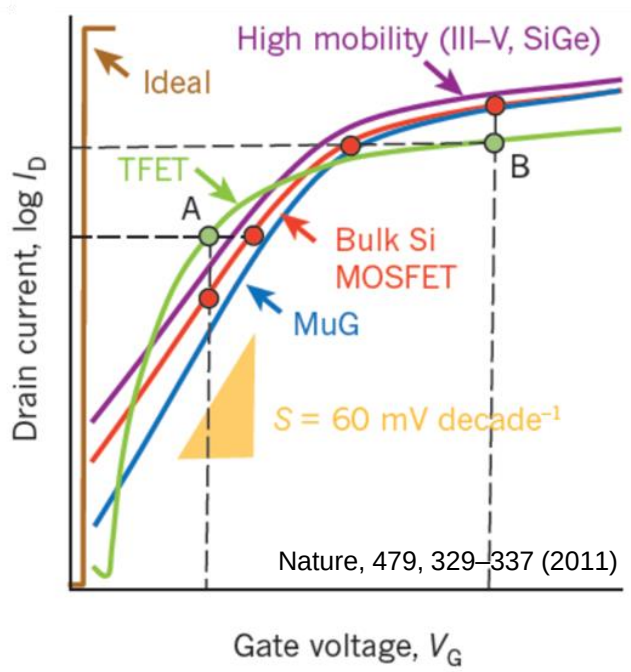
pubs.acs.org/NanoLett Letter

60 mV/dec.

Overcoming Boltzmann's Tyranny in a Transistor via the Topological Quantum Field Effect

Muhammad Nadeem,* Iolanda Di Bernardo, Xiaolin Wang, Michael S. Fuhrer,* and Dimitrie Culcer*

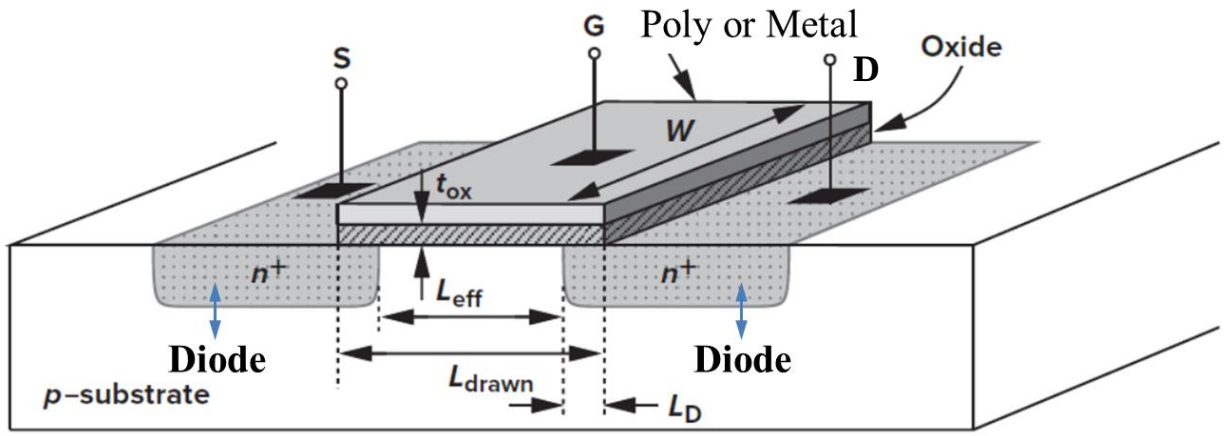
Cite This: *Nano Lett.* 2021, 21, 3155–3161 Read Online



Nature, 479, 329–337 (2011)

General Considerations

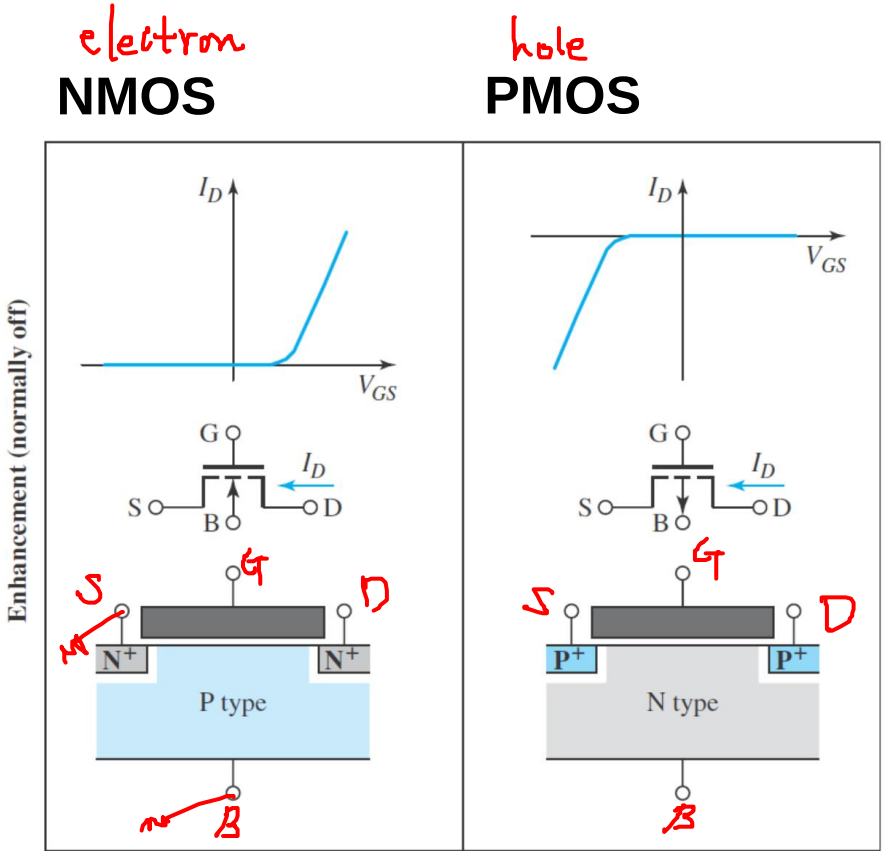
Recall BJT: any voltage-controlled current source can provide signal amplification. **MOSFETs also behave as such controlled sources** but their characteristics are different from those of BJTs.



- MOS = Metal-Oxide-Semiconductor
- FET = Field effect Transistor

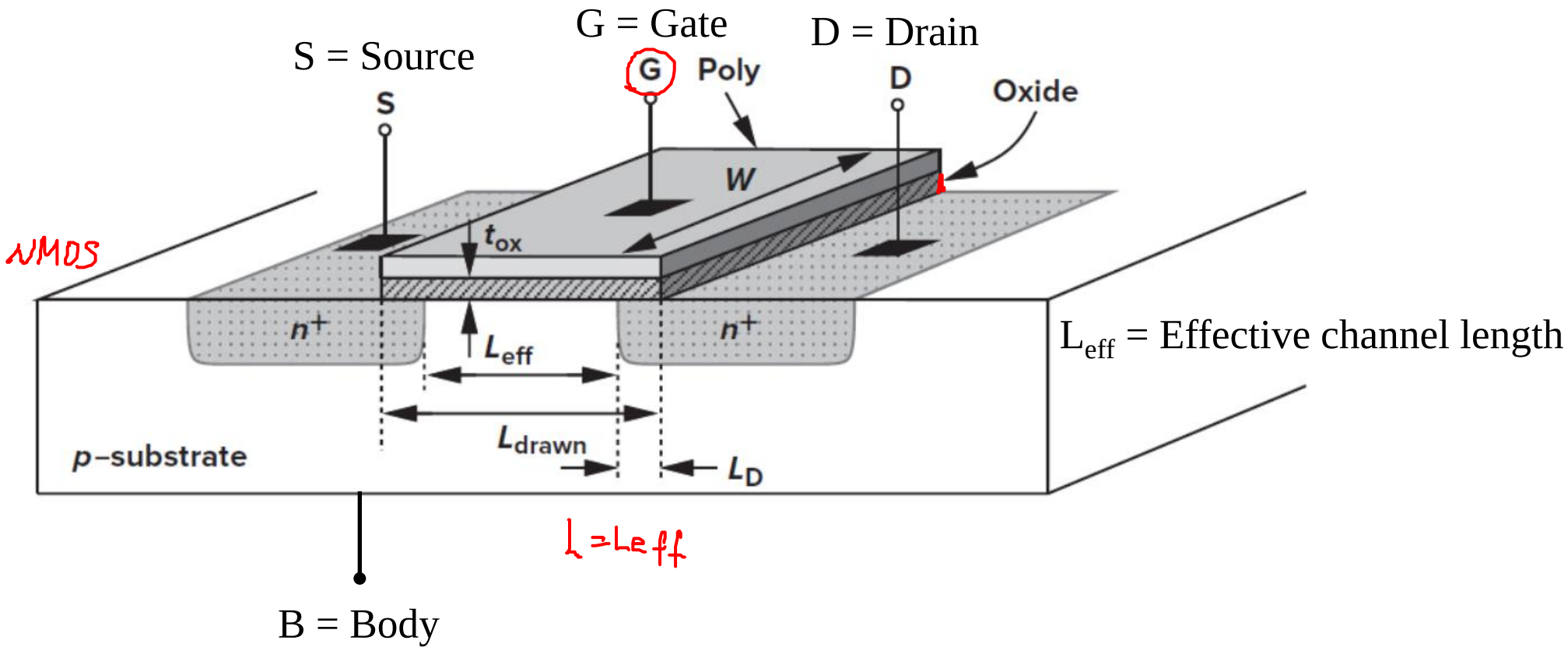
Field Effect Transistor: Charge carriers (current) in a channel are controlled by V_{GS} (or V_G) and V_{DS} (or V_D). FET will show I-V curves in relation with I_{DS}-V_{GS} and I_{DS}-V_{DS}.

MOSFET structure



The MOSFET is a four-terminal device:

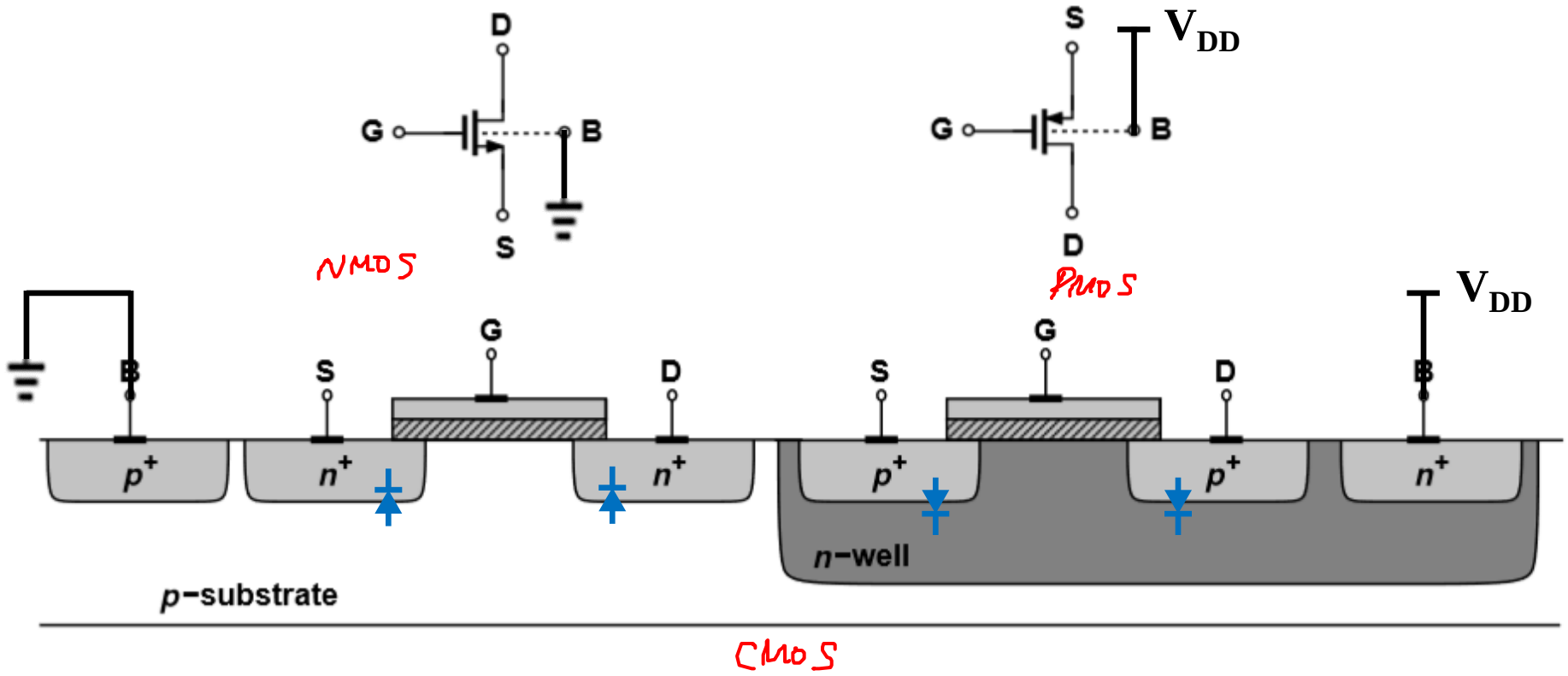
- Silicon substrate body (B), gate (G), source (S), and drain (D).
- NMOS is built on a p-type Si whereas PMOS is built on an n-type Si.
- Generally, source and body are grounded. (NMOS)



L_{eff} = Effective channel length

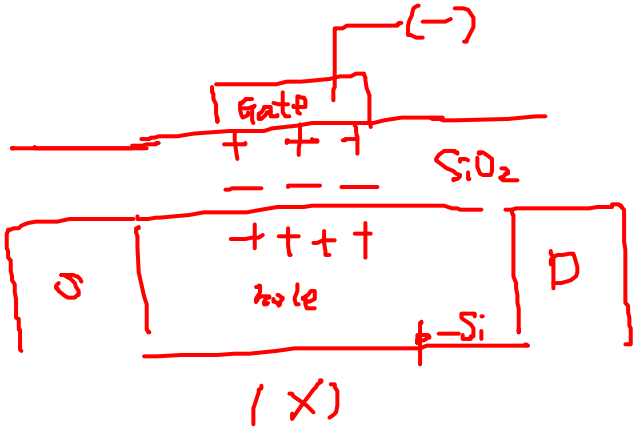
The lateral dimension of the gate along the source-drain path is called the **length, L**, and that **perpendicular to the length** is called the **width, W**. The actual distance between S and D is slightly less than L. To avoid confusion, we write, $\mathbf{L_{eff} = L_{drawn} - 2L_D}$, L_{eff} and the gate oxide thickness, t_{ox}, play an important role in the performance of MOS circuits.

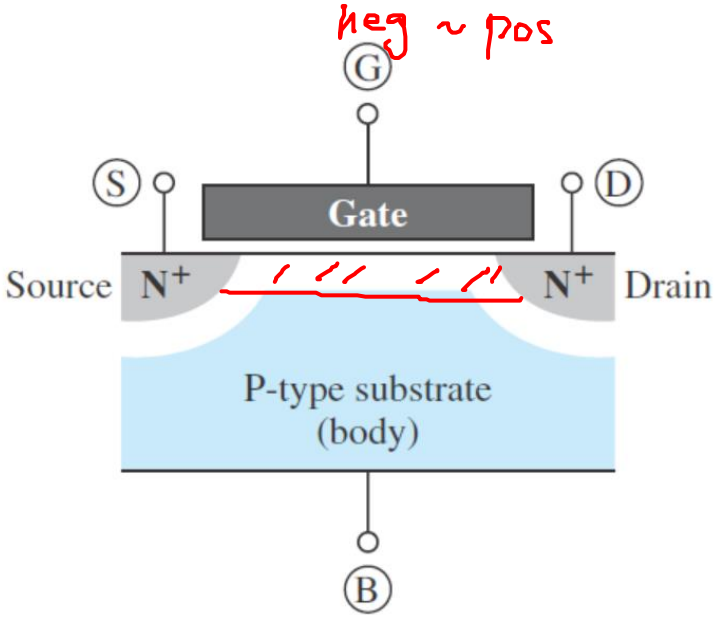
Complementary MOS (CMOS) technologies

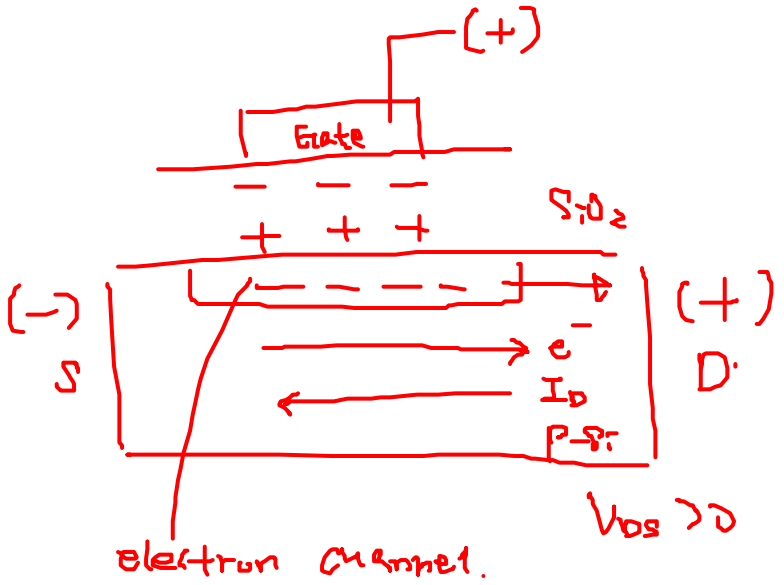


- CMOS = Complementary MOS
- Substrate (Body) of NMOS is generally connected to ground.
- N-well (Body) of PMOS is generally connected to V_{DD}.

N-channel enhancement MOSFET working mechanism

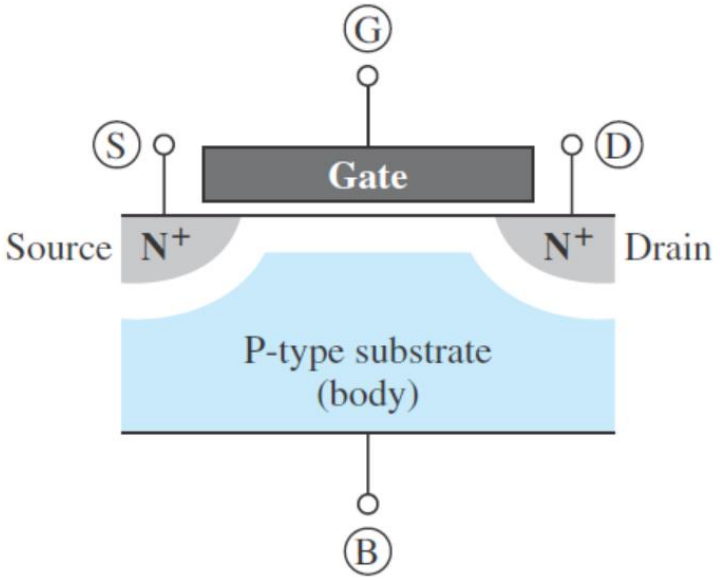






Gate voltage (V_{GS}) controls the state of the silicon surface underneath.

- **Negative V_{GS}** attracts the **holes** from the p-type Si to the surface: **Accumulation.**
- **Positive $V_{GS} > V_T$** creates **a layer of electrons** at the surface: **Inversion.**
- These two states are used to make a voltage-controlled switch.



(i) On state: The layer of electrons at the surface is contacted at the source and drain. **The existence of the electron layer (channel*)* corresponds to the on state** – the electron channel virtually short circuits the source and the drain regions.

(ii) Off state: When the gate voltage is **below the threshold voltage**, **the electron layer (the channel) disappears from the surface**, and the source and drain N^+ regions are isolated by the P-type substrate.

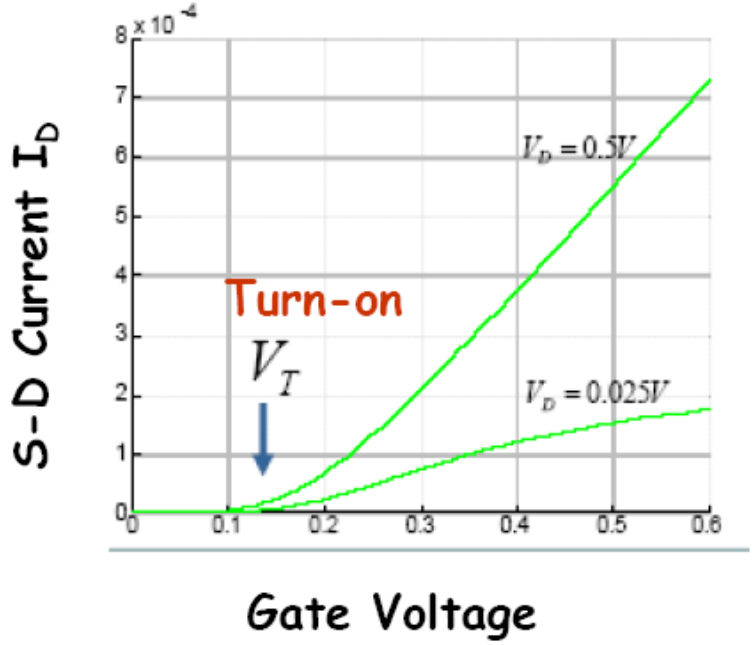
MOS I/V Characteristics

We will analyze the **generation and transport of charge in MOSFETs** as a function of the terminal voltages. Our objective is to **derive equations for the I/V characteristics** such that we can elevate our abstraction from device physics level to circuit level.

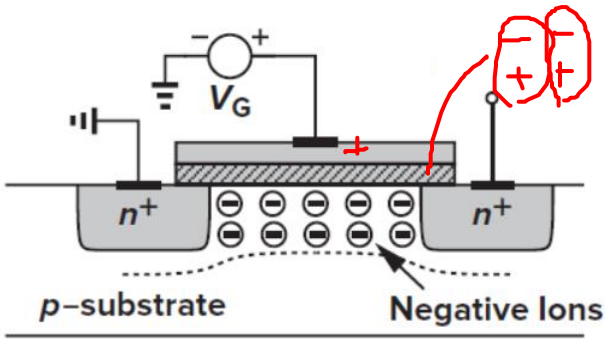
Threshold Voltage (NMOS)

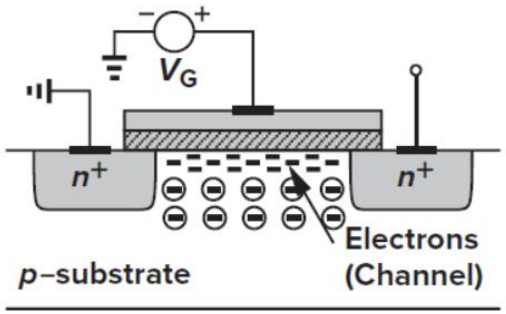
The threshold voltage V_{TH} of a MOSFET is the minimum V_{GS} that is need to create a conduction path between the source and drain terminals.

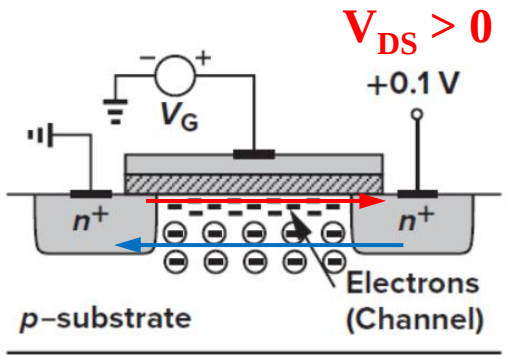
$$V_{TH} = \Phi_{MS} + 2\Phi_F + \frac{Q_{dep}}{C_{ox}}$$



Assume we change V_G from negative to positive





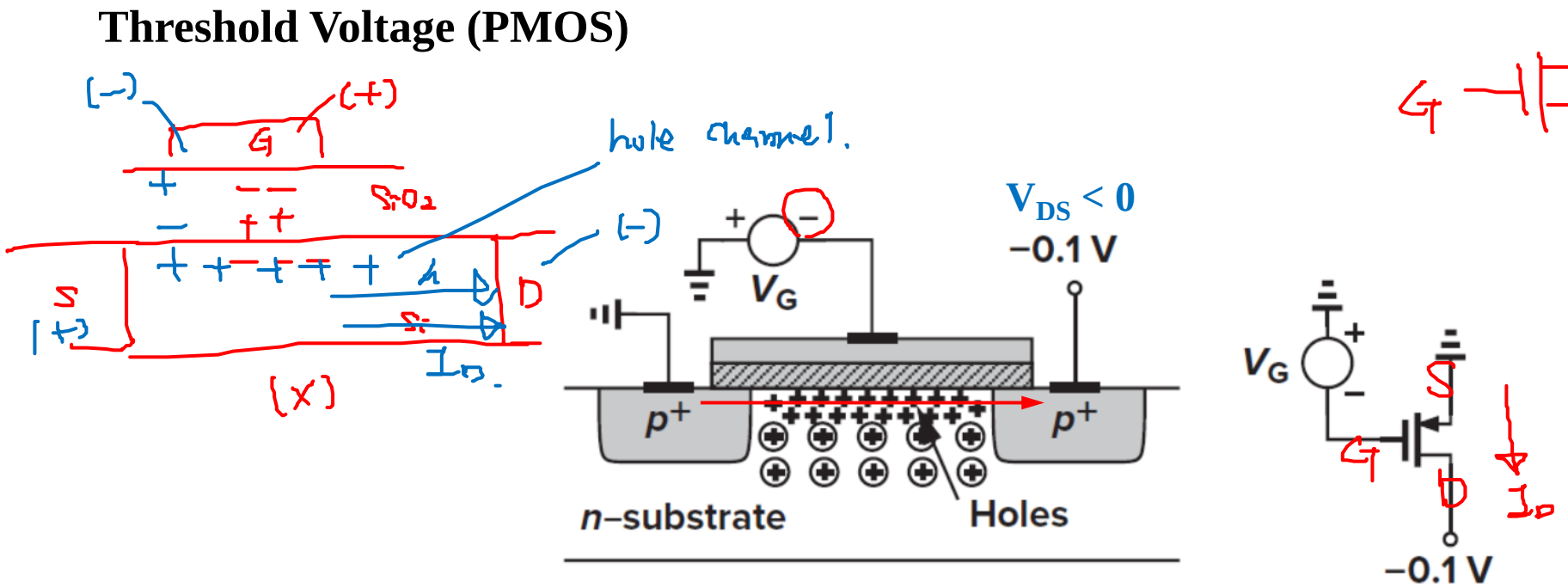


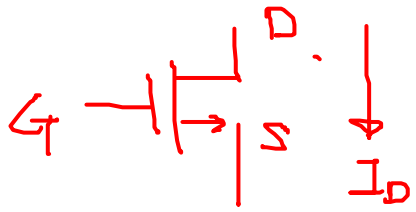
Electron direction

Drain current (I_D) direction

- By applying **positive V_{GS}**, electrons are induced in the channel, or holes are repelled from the channel, forming a **depletion region** at the interface between oxide and Si.
- When V_{GS} (> 0) is large enough, **a channel of electrons (inversion layer)** is formed at the interface.
- The value of V_{GS} at which the inversion layer forms is the **threshold voltage (V_{TH})**. The device turns on abruptly for $V_{GS} \geq V_{TH}$. NMOS
- **With the bias of V_{DS}**, electrons flow **from source to drain.**

Threshold Voltage (PMOS)

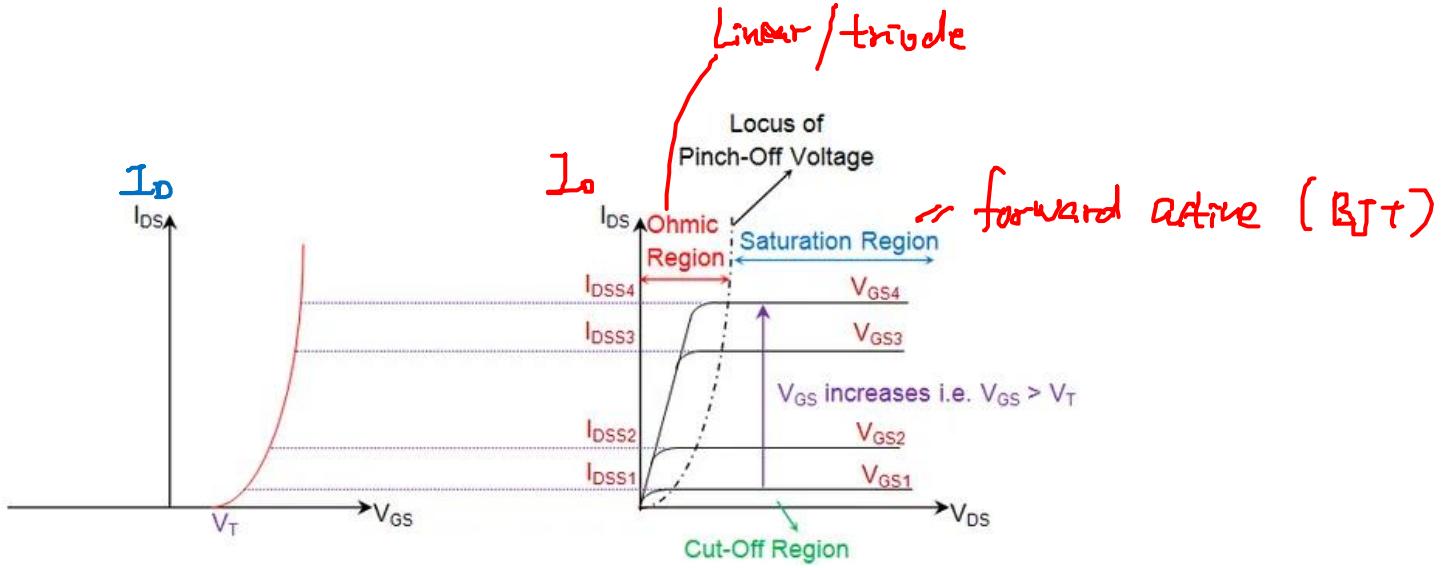




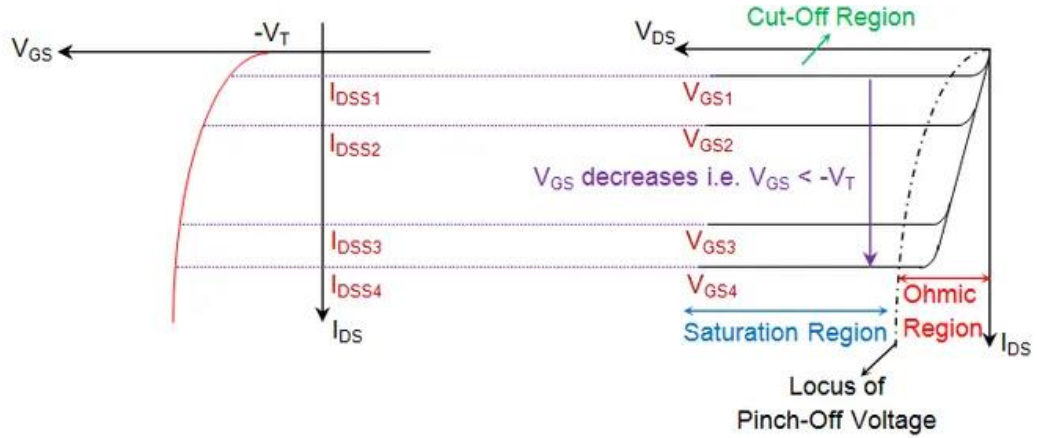
Hole and drain current (I_D) direction

The turn-on phenomenon in a PMOS device is similar to that of NFETs, but with all the polarities reversed. If the V_{GS} **becomes sufficiently *negative***, an inversion layer consisting of holes is formed at the oxide-silicon interface, providing a conduction path between the source and the drain. V_{TH} **of PMOS < 0.**

I-V characteristics



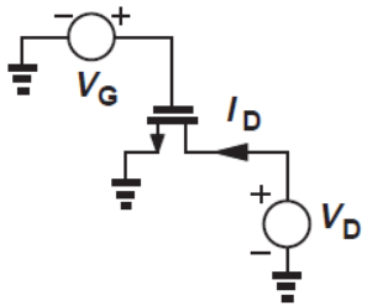
N-channel enhancement mode



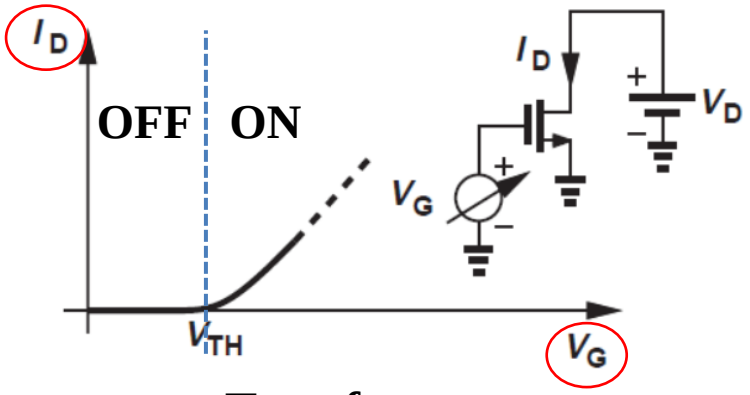
P-channel enhancement mode

Modes of MOSFET

(1) $V_{GS} < V_{TH}$: No channel exists, the **device is off**, and $I_D = 0$ regardless of the value of V_{DS}.
(2) $V_{GS} > V_{TH}$: The channel is formed. With $V_{DS} > 0$, the **device is on** and there is $I_D > 0$.

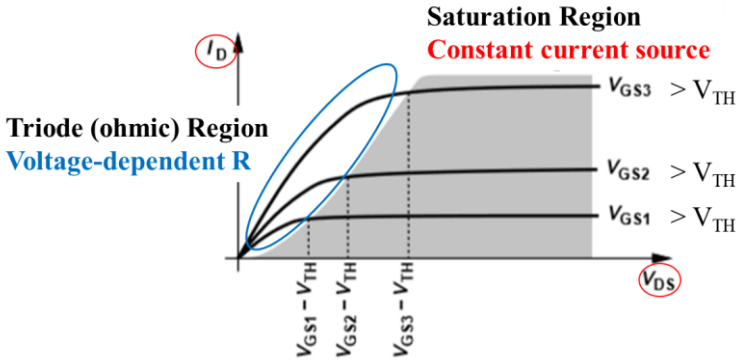


MOSFET Bias: V_{GS} and V_{DS}



Transfer curve

By varying V_{GS} at fixed V_{DS}



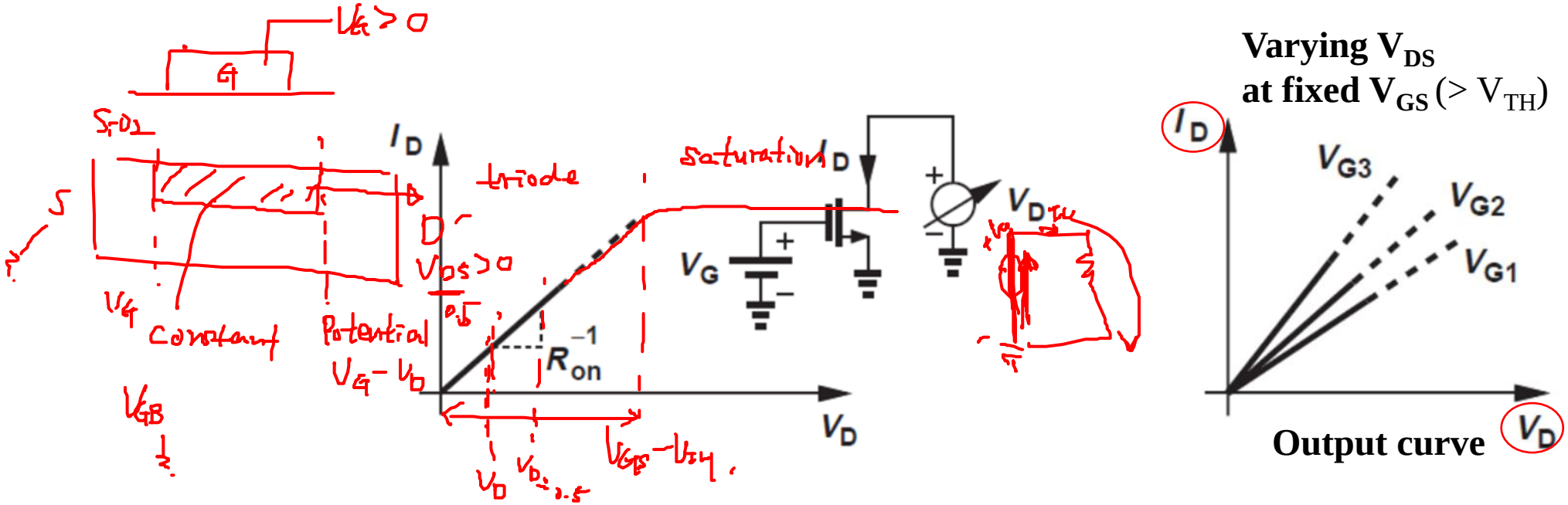
Output curve

By varying V_{DS} at fixed V_{GS}

In an output curve of MOSFET, we can determine two modes (regions): **(1) Triode or ohmic region and (2) saturation region.**

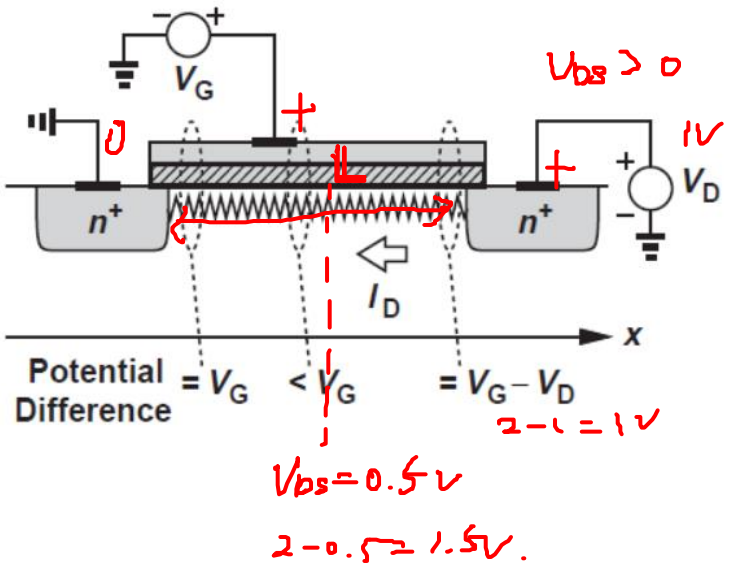
MOSFET as a Variable Resistor (Triode Region)

The value of MOSFET resistor changes with the gate voltage. **Such a voltage-dependent resistor** proves extremely useful in analog and digital circuits.

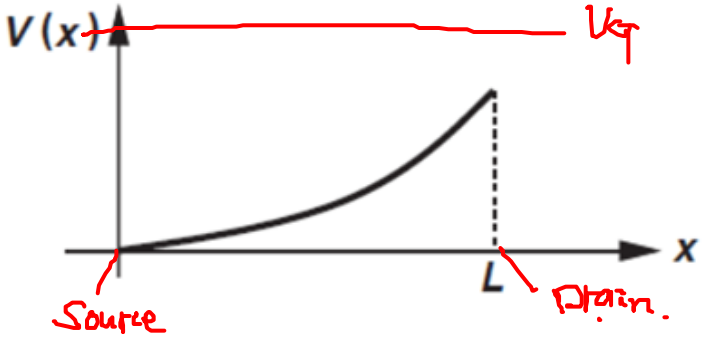


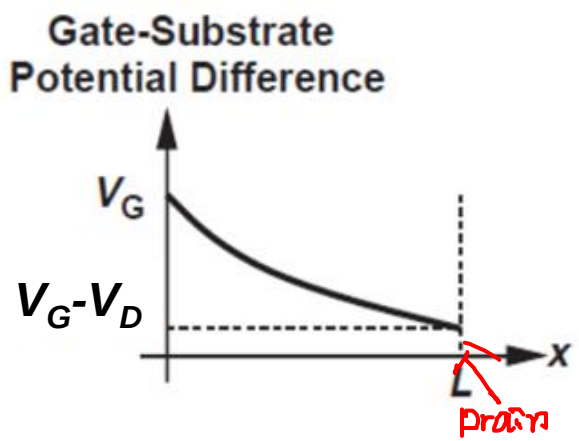
As V_{GS} controls the channel conductivity (before saturation of MOSFET), the **source-drain path** may act as a simple resistor, yielding the I_D-V_{DS} characteristic. The **slope of the characteristic is equal to $1/R_{on}$,** where R_{on} denotes the "on-resistance" of the transistor.

Channel Pinch-Off and Saturation



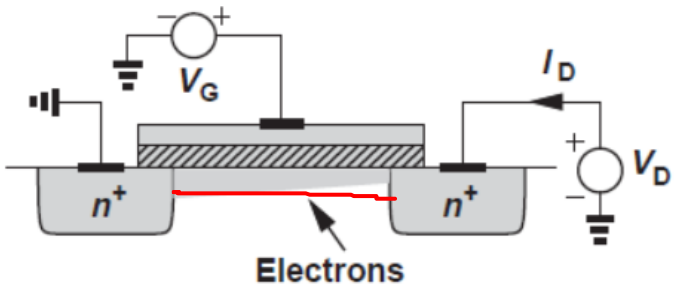
Potential profile btw S-D

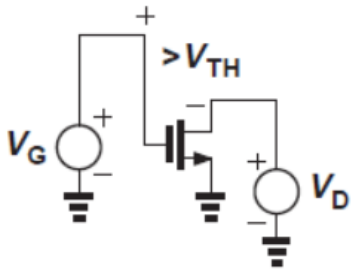




*As the source is grounded generally, $V_G = V_{GS}$ and $V_D = V_{DS}$

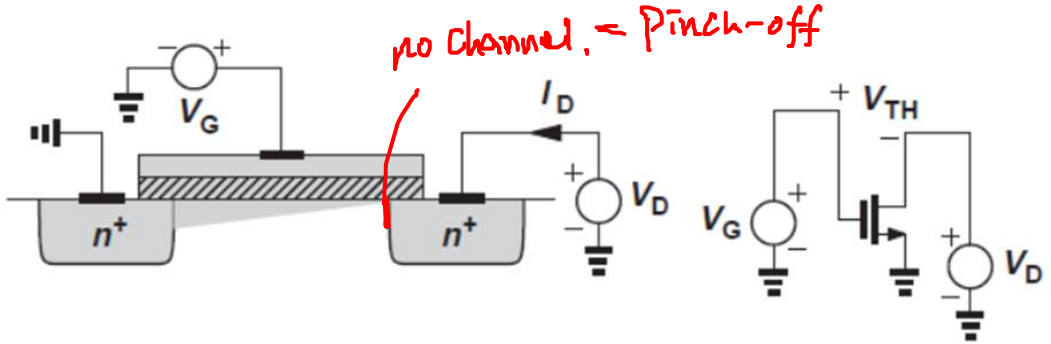
To understand **Pinch-Off**, we make two observations: (1) to form a channel, the potential difference between the gate and the oxide-silicon interface must **exceed V_{TH}**; (2) if $V_{DS} > 0$ with the grounded source, the **V_{DS} at each point along the channel with respect to ground increases** from the source towards the drain.



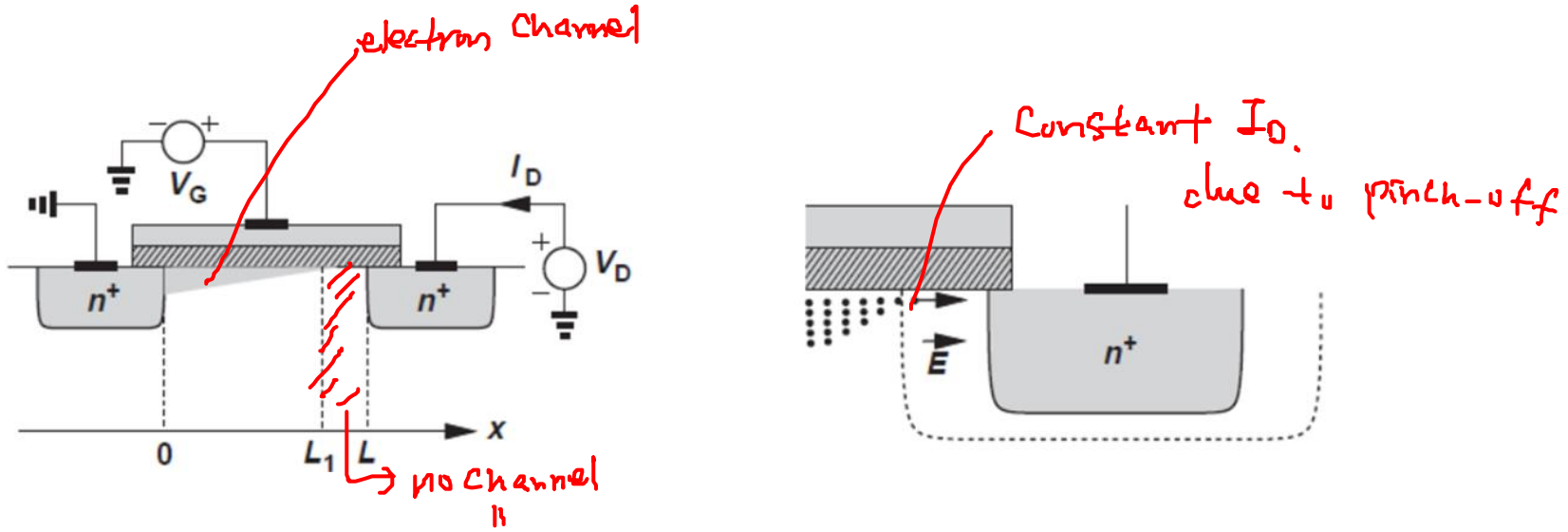


$\mathbf{V_{GS} - V_{DS} > V_{TH}}$, No pinch-off

$\mathbf{V_{GS} - V_{DS} = V_{TH}}$,
Pinch-off at the drain side



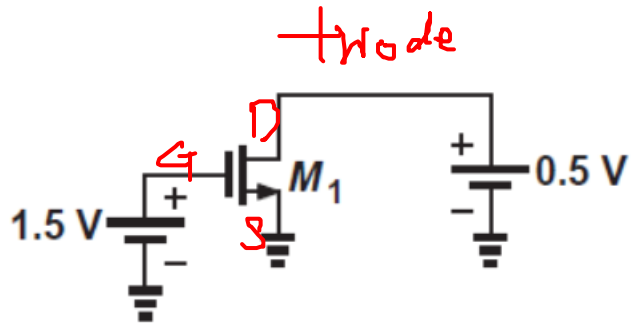
If the V_{DS} is high enough to produce $\mathbf{V_{GS} - V_{DS} \leq V_{TH}}$, then the channel ceases to exist near the drain. At x = L, the channel is **pinched off**.



$\mathbf{V_{GS} - V_{DS} < V_{TH}}$, Pinch-off expands further towards source side

As V_{DS} further increases ($V_{DS} > V_{GS} - V_{TH}$), the voltage difference between the gate and the substrate falls to V_{TH} at some point $L_1 < L$. The drain voltage no longer affects the current significantly, and the **MOSFET acts as a constant current source**.

Example 1. Determine the region of operation of M_1 in each of the circuits. $V_{TH} = 0.4$ V.



Triode

$V_G = 1.5V$ } $V_{GS} = 1.5V$

$V_S = 0V$

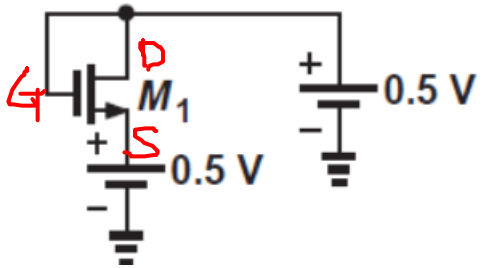
$V_D = 0.5V$ $V_{DS} = 0.5V$

$V_{GS} - V_{TH} > V_{DS}$

$1.5 - 0.4 \quad 0.5$

"triode"

$V_{GS} > V_{TH} \Rightarrow$ on.



off

$V_G = 0.5$

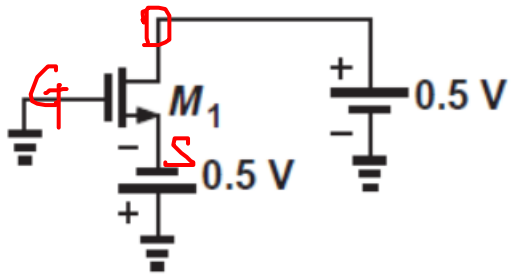
$V_D = 0.5$

$V_S = 0.5$

$V_{GS} = 0 < V_{TH}$

off.

Saturation



$V_G = 0V$

$V_S = -0.5$

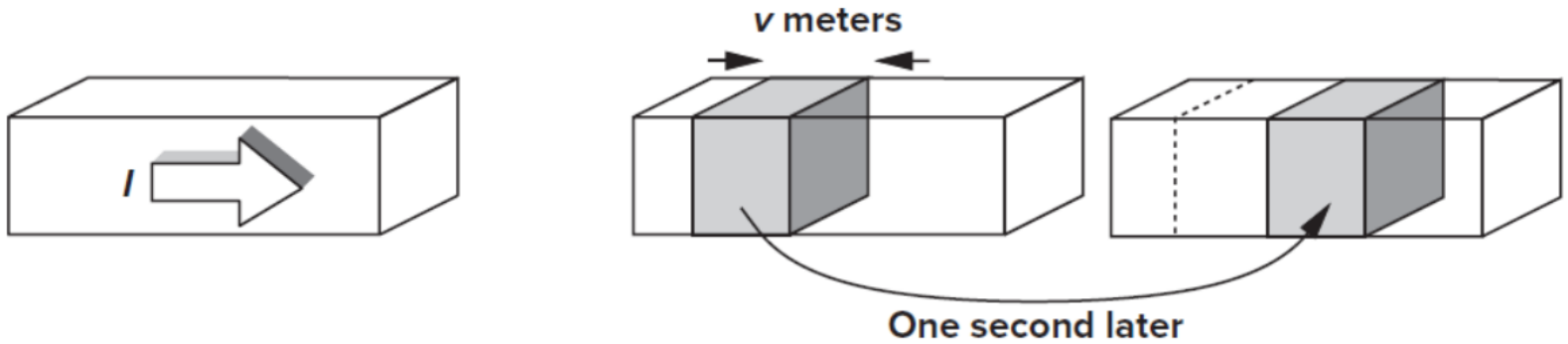
$V_D = 0.5$

$V_{GS} = 0.5 > V_{TH} \Rightarrow$ on

$V_{DS} = 1V$

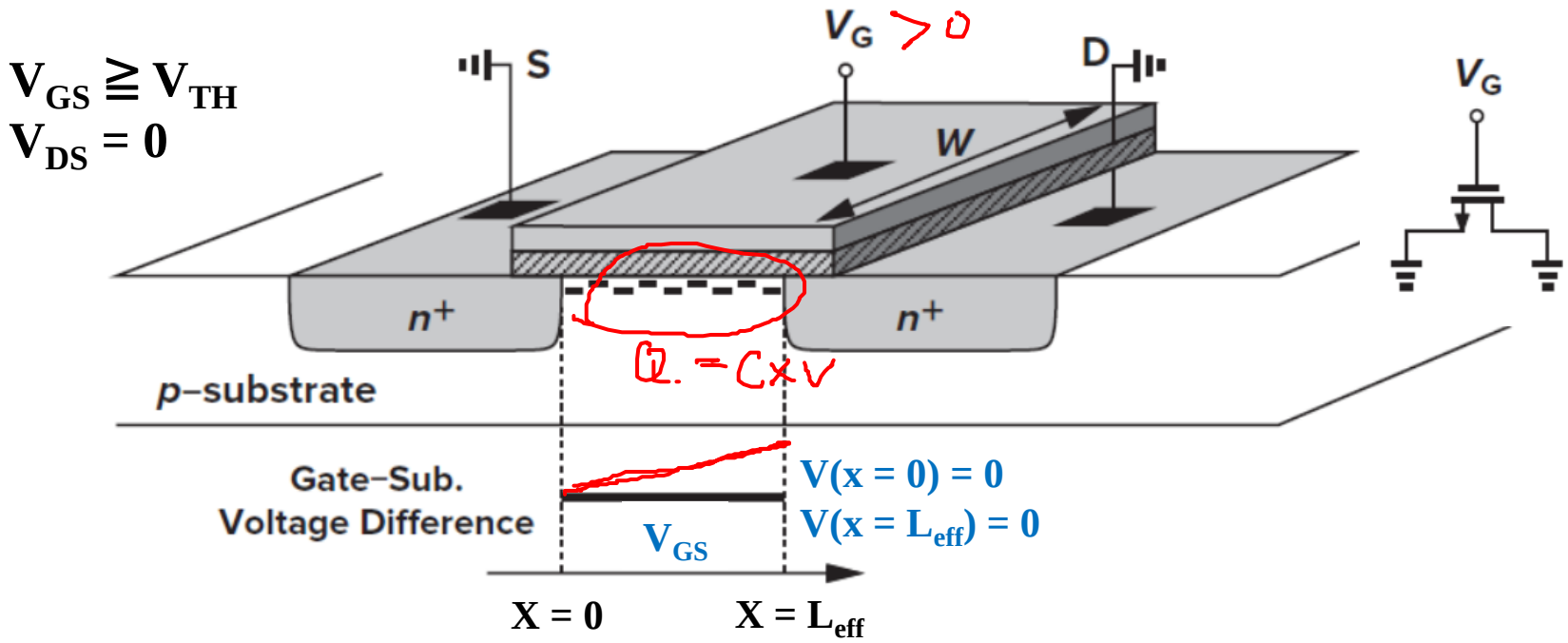
$V_{GS} - V_{TH} < V_{DS}$

Derivation of I-V Characteristics (NMOS)



$I = Q/t$

Let's consider a semiconductor bar carrying a current *I*. If the mobile charge density along the direction of current is Q_d coulombs per meter and the velocity of the charge is v meters per second, then ***I* [C/s] = Q_d [C/m]· v [m/s]**



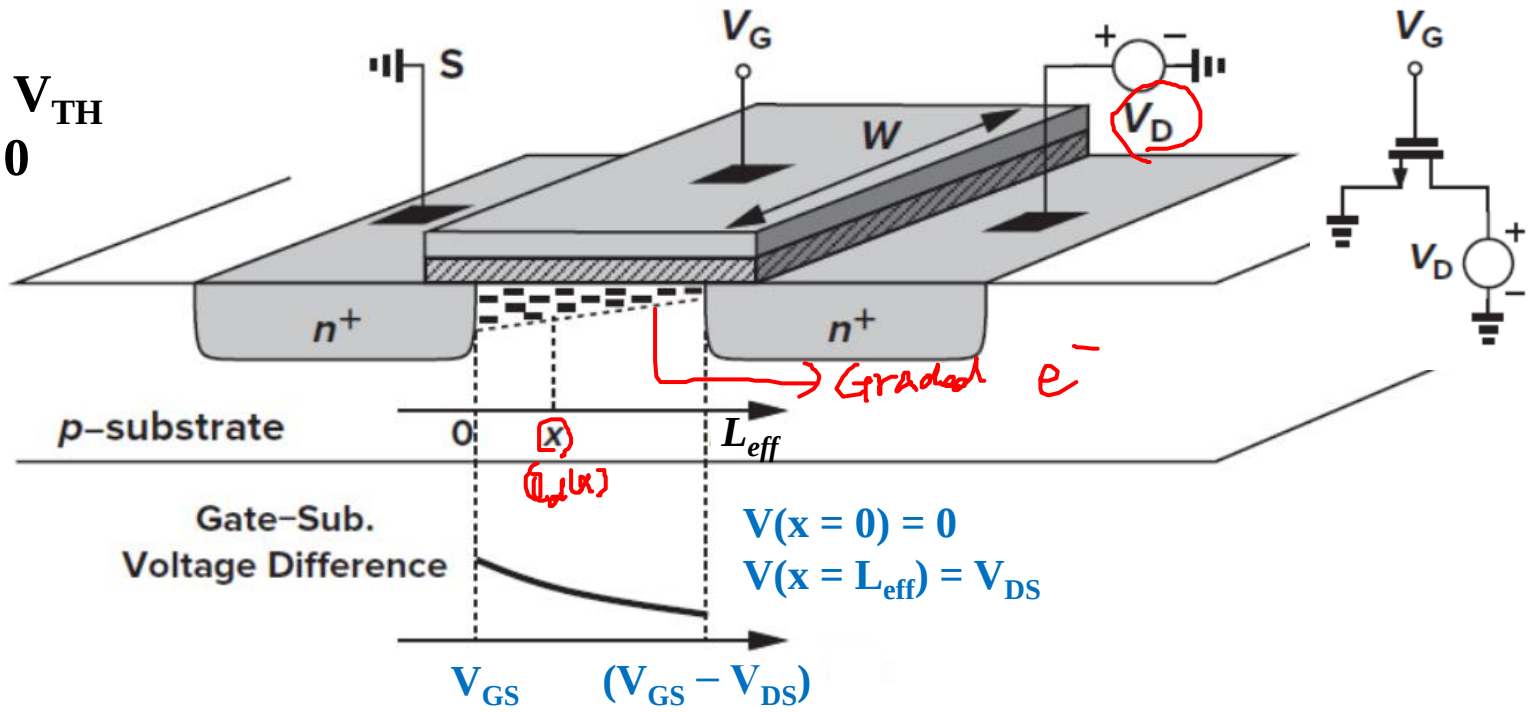
The onset of **inversion occurs at $V_{GS} = V_{TH}$**, the inversion charge density produced by the gate-oxide capacitance is proportional to $V_{GS} - V_{TH}$. For $V_{GS} \geq V_{TH}$, any charge placed on the gate must be **mirrored by the charge in the channel**, yielding a uniform channel charge density [charge per unit length along L_{eff}]

$$Q = WL_{eff}C_{ox}(V_{GS} - V_{TH}) \quad \text{(unit: coulomb)}$$

$$Q_d = WC_{ox}(V_{GS} - V_{TH}) \quad \text{(unit: coulomb·m}^{-1}\text{)}$$

C_{ox} (gate oxide capacitance per unit area)
$= \epsilon_{\text{silicon oxide}} / t_{ox}$
$= [8.85 \times 10^{-12}\ \text{(F/m)} \times 3.9] / t_{ox}$

$V_{GS} \geqq V_{TH}$
$V_{DS} > 0$



Since the channel potential varies from the source to the drain, the local voltage difference between the gate and the channel **varies from V_{GS} near the source to $V_{GS} - V_{DS}$ near the drain**. Thus, the **charge density at a point x along the channel** can be written as

$$I_D = Q_d \times v = Q_d \times (\mu_n \mathcal{E}) = -WC_{ox}[V_{GS} - V_{TH} - V(x)] \times (\mu_n \mathcal{E})$$

$$= WC_{ox}[V_{GS} - V_{TH} - V(x)] \times \mu_n \times \frac{dV(x)}{dx} \qquad \mathcal{E} = -dV(x)/dx$$

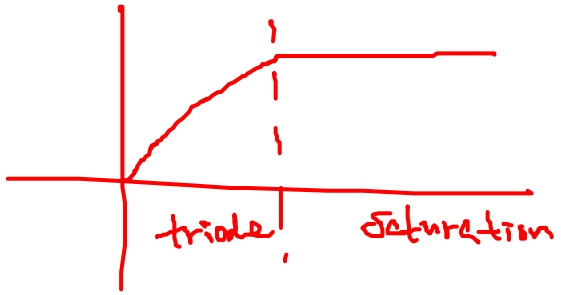
$$I_D = Q_d \times v = Q_d \times (\mu_n \mathcal{E}) = -WC_{ox}[V_{GS} - V_{TH} - V(x)] \times (\mu_n \mathcal{E})$$

$$= WC_{ox}[V_{GS} - V_{TH} - V(x)] \times \mu_n \times \frac{dV(x)}{dx} \qquad \mathcal{E} = -dV(x)/dx$$

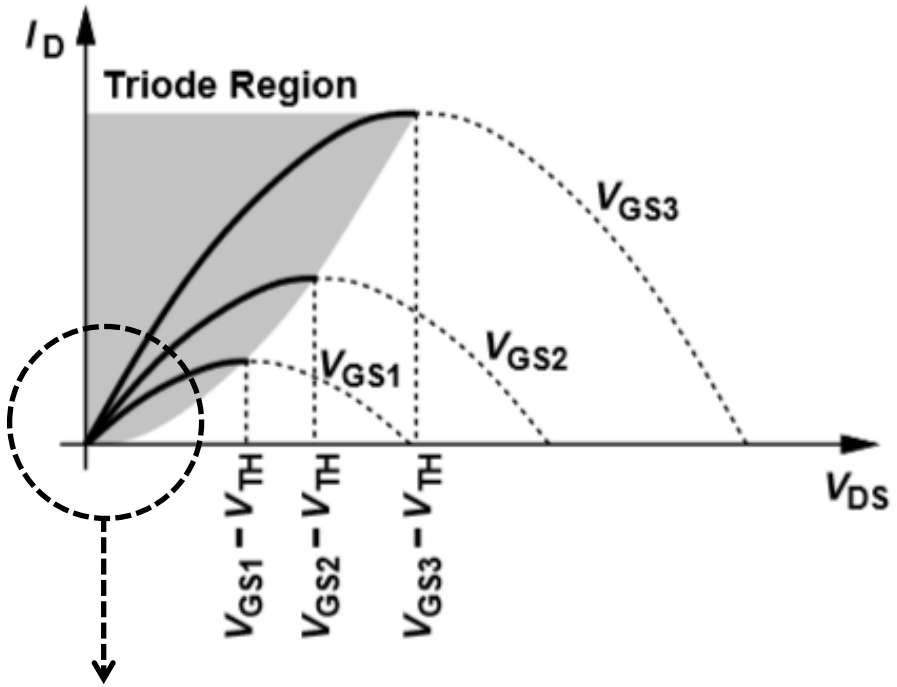
Integrate along the channel $\int_{x=0}^{x=L_{eff}} I_D \cdot dx = \int_{V(0)=0}^{V(L)=V_{DS}} \mu_n C_{ox} W[V_{GS} - V_{TH} - V(x)] \cdot dV(x)$

Drain current

$$\mathbf{I_D = \mu_n C_{ox} \frac{W}{L_{eff}} [(V_{GS} - V_{TH})V_{DS} - \frac{1}{2} V_{DS}^2]}$$



Triode Region



In the **Triode region**, $\mathbf{V_{DS} < V_{GS} - V_{TH}}$

$$I_D = \mu_n C_{ox} \frac{W}{L_{eff}} \left[(V_{GS} - V_{TH}) V_{DS} - \frac{1}{2} V_{DS}^2 \right]$$

When $\mathbf{V_{DS} = V_{GS} - V_{TH}}$ (overdrive voltage)

$$I_{D,max} = \frac{1}{2} \mu_n C_{ox} \frac{W}{L_{eff}} (V_{GS} - V_{TH})^2$$

Deep triode region: $\mathbf{V_{DS} \ll 2(V_{GS} - V_{TH})}$**, and thus** $\mathbf{V^2_{DS}}$ **can be ignored in the equation**

$$I_D = \mu_n C_{ox} \frac{W}{L_{eff}} \left[(V_{GS} - V_{TH}) V_{DS} - \frac{1}{2} V_{DS}^2 \right] \rightarrow \boldsymbol{I_D = \mu_n C_{ox} \frac{W}{L_{eff}} (V_{GS} - V_{TH}) V_{DS}}$$

→ Linear relationship of I_D-V_{DS}

Triode Region

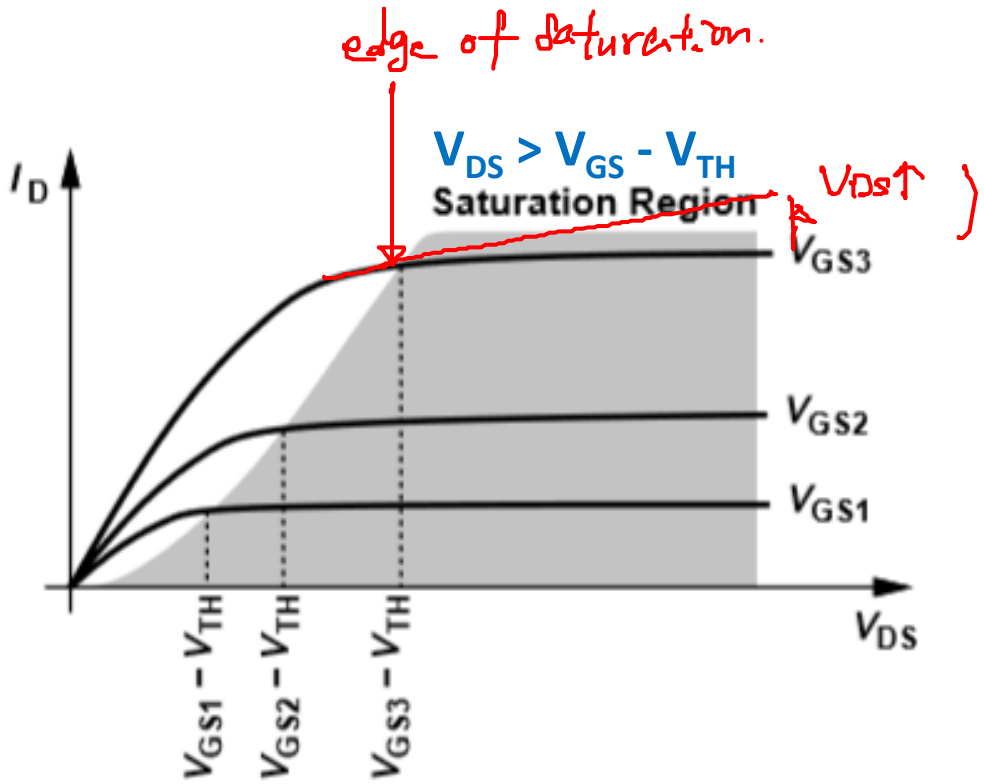
Deep triode region: $I_D = \mu_n C_{ox} \frac{W}{L_{eff}} \left[(V_{GS} - V_{TH}) V_{DS} - \frac{1}{2} {V_{DS}}^2 \right] \rightarrow \boldsymbol{I_D = \mu_n C_{ox} \frac{W}{L_{eff}} (V_{GS} - V_{TH}) V_{DS}}$

The linear relationship of I_D-V_{DS} implies that the path from S to D can be represented **by a linear R**

$$R_{on} = \frac{1}{\mu_n C_{ox} \left(\frac{W}{L_{eff}} \right) (V_{GS} - V_{TH})}.$$

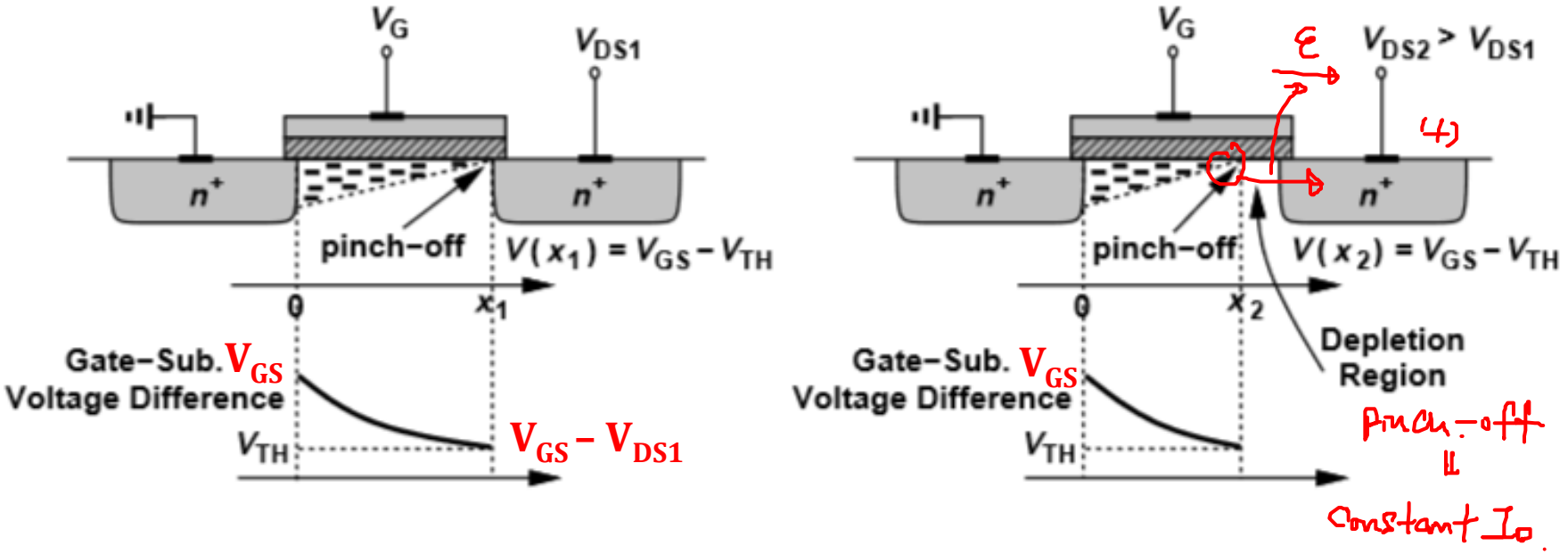
- For digital circuit, MOSFET operates in **deep triode region for switching**.
- Reducing t_{ox} and L_{eff} can improve speed.

Saturation Region

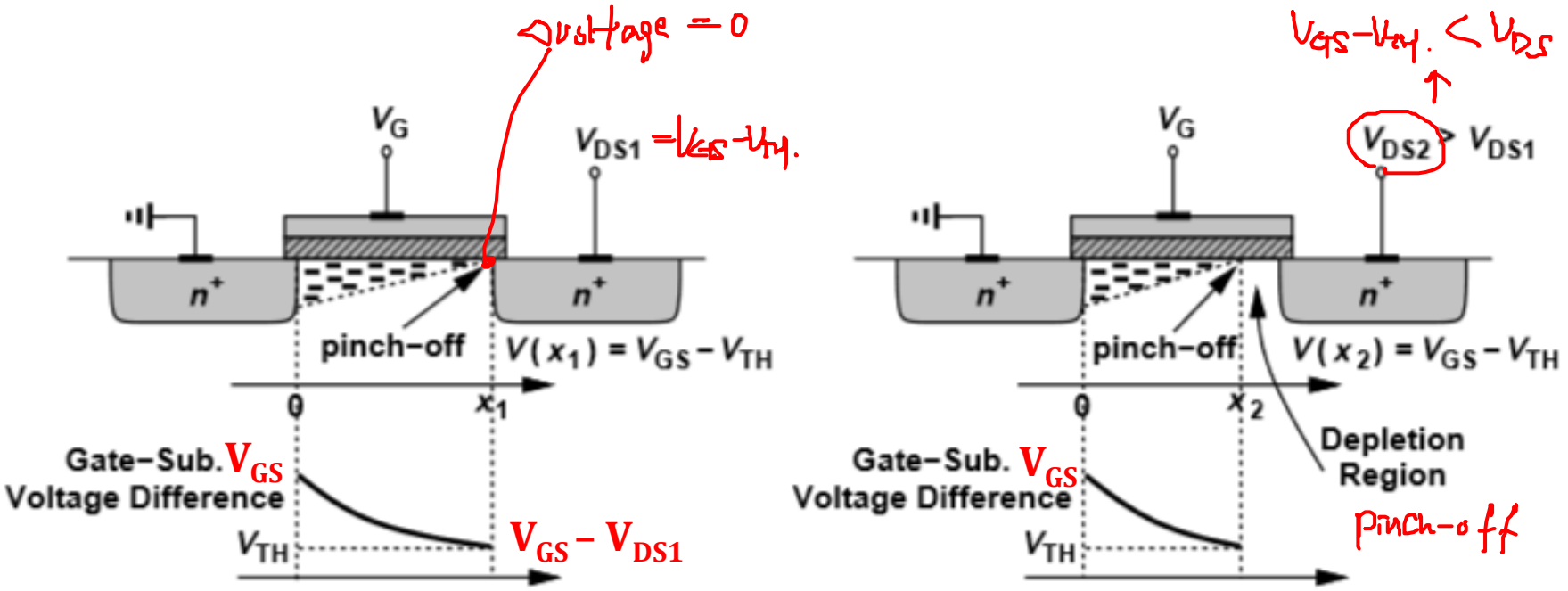


When the $V_{DS} > V_{GS} - V_{TH}$, the drain current does not follow the parabolic behavior. In fact, I_D becomes relatively constant and the device operates in the **saturation region**.

$V_{DS} = V_{GS} - V_{TH}$ is the minimum value for the NMOS to operate in saturation region.
For **$V_{DS} > V_{GS} - V_{TH}$**, I_D becomes relatively constant.



The local density of the inversion-layer charge is proportional to $V_{GS} - V(x) - V_{TH}$. Thus, if V(x) approaches $V_{GS} - V_{TH}$, then $Q_d(x)$ drops to zero. In other words, if V_{DS} is slightly greater than $V_{GS} - V_{TH}$, then the inversion layer stops at $x \leq L$, and the channel is **pinched off**. As V_{DS} increases further, **the point at which Q_d equals zero gradually moves toward the source**.



When $V_{DS} = V_{GS} - V_{TH}$, the saturation occurs. Thus, substitute V_{DS} with $V_{GS} - V_{TH}$ then

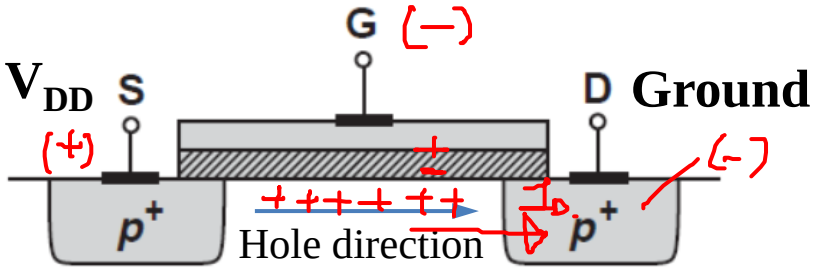
$$I_D = \mu_n C_{ox} \frac{W}{L_{eff}} \left[(V_{GS} - V_{TH}) V_{DS} - \frac{1}{2} V_{DS}^2 \right] \rightarrow I_D = \frac{1}{2} \mu_n C_{ox} \frac{W}{L'} (V_{GS} - V_{TH})^2$$

L': the point at which Q_d drops to zero

PMOS → hole channel.

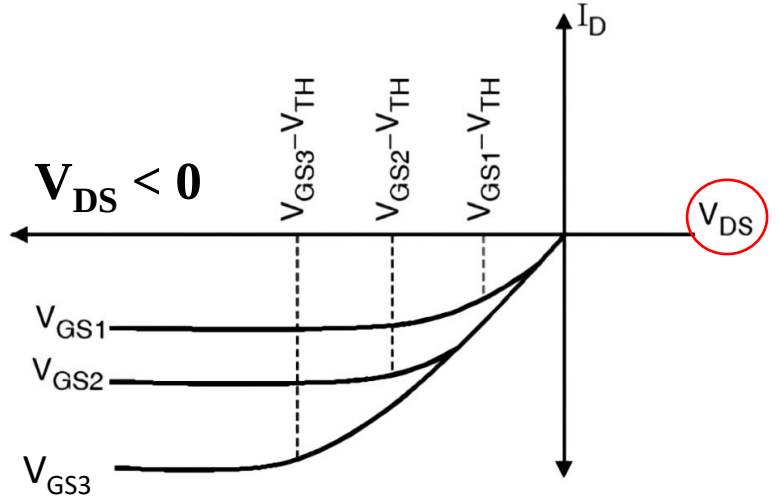
h^+

$V_{SD} > 0$ or $V_{DS} < 0$



NMOS : S → e → D

S ← I_D ← D

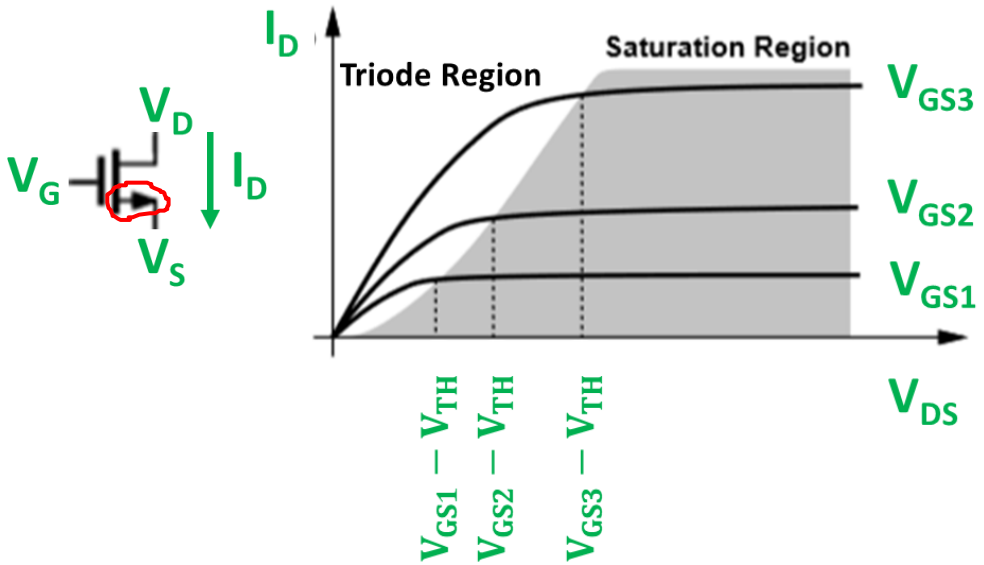


Triode region: $I_D = \mu_p C_{ox} \frac{W}{L} \left[(V_{GS} - V_{TH}) V_{DS} - \frac{1}{2} V_{DS}^2 \right]$

Saturation region: $I_D = \frac{1}{2} \mu_p C_{ox} \frac{W}{L'} (V_{GS} - V_{TH})^2$

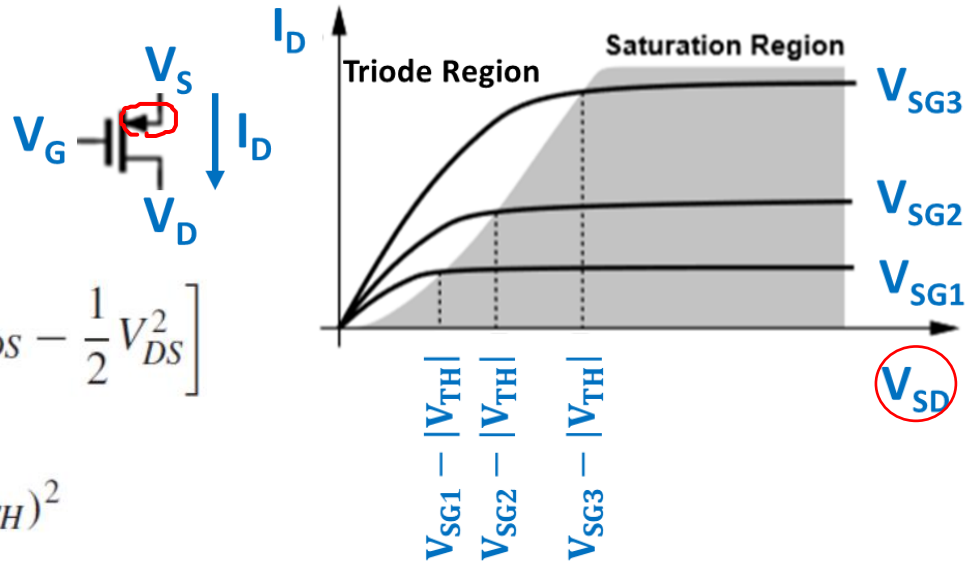
NMOS: Electrons from **Source to Drain**, and thus, the drain current from **Drain to Source**.
PMOS: Holes from **Source to Drain**, and thus, the drain current from **Source to Drain**,
i.e. current direction is opposite.

NMOS *vs* PMOS



Triode region: $I_D = \mu_n C_{ox} \frac{W}{L_{eff}} \left[(V_{GS} - V_{TH}) V_{DS} - \frac{1}{2} {V_{DS}}^2 \right]$

Saturation region: $I_D = \frac{1}{2} \mu_n C_{ox} \frac{W}{L'} (V_{GS} - V_{TH})^2$



Triode region: $I_D = -\mu_p C_{ox} \frac{W}{L} \left[(V_{GS} - V_{TH}) V_{DS} - \frac{1}{2} V_{DS}^2 \right]$

Saturation region: $I_D = -\frac{1}{2} \mu_p C_{ox} \frac{W}{L'} (V_{GS} - V_{TH})^2$

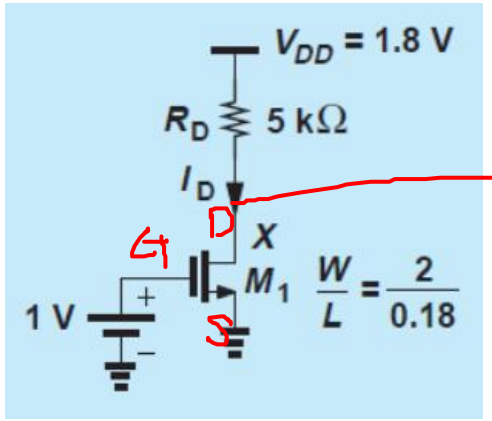
Example 2. Calculate the bias current ($\mathbf{I_D}$) and V_{DS} of M_1. Assume M_1 is in a saturation mode, $\mu_n C_{ox}$ = 100 μA/V^2 and V_{TH} = 0.4 V. If the gate voltage increases by 10 mV, what is the change in the drain voltage?

$$\mathbf{I_D}(sat) = \frac{\mathbf{1}}{\mathbf{2}}\mu_n \mathbf{C_{ox}} \frac{\mathbf{W}}{\mathbf{L'}}(\mathbf{V_{GS}} - \mathbf{V_{TH}})^2$$

$(1-0.4)^2$

$I_D(sat) = \frac{1}{2} \times 100\mu \times \frac{2}{0.18} \times 0.6^2$

$= 200\mu A.$



$V_{GS} = 1V$

V_D

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

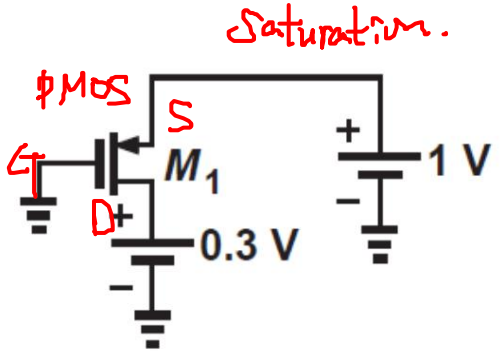
$1.8 - 200\mu A \times 5k = V_{DS} = 0.8V$

$\Rightarrow V_{GS} - V_{TH}$.

$V_{GS} = 1.01V$

$= A.$

Example 3. Determine the region of operation of M_1 in each circuit below. V_{TH} = -0.4 V



Saturation.

$V_G = 0$

$V_S = 1V$

$V_D = 0.3V$

$V_{GS} = -1V$

$V_{DS} = -0.7V$

$V_{TH} = -0.4\ V.$

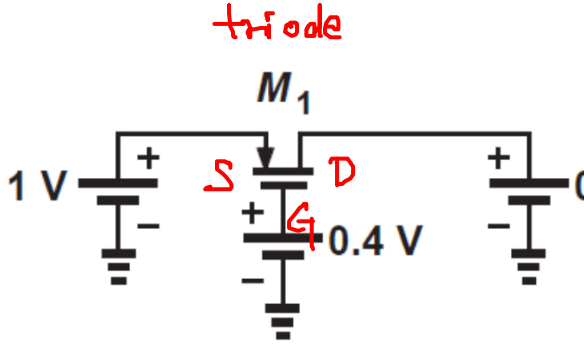
on.

$|V_{GS}| - |V_{TH}| < |V_{DS}|$

1V 0.4 0.7

PMOS $V_{SG} - |V_{TH}| \gtrless V_{SD}$.

NMOS $V_{GS} - V_{TH} \gtrless V_{DS}$



triode

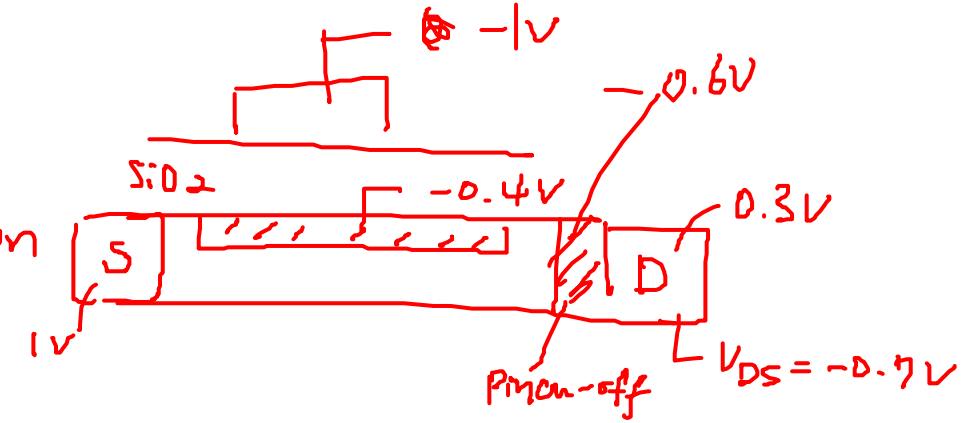
$V_G = 0.4$

$V_S = 1V$

$V_D = 0.9$

$V_{SG} = 0.6\ V$

$V_{SD} = 0.1\ V$

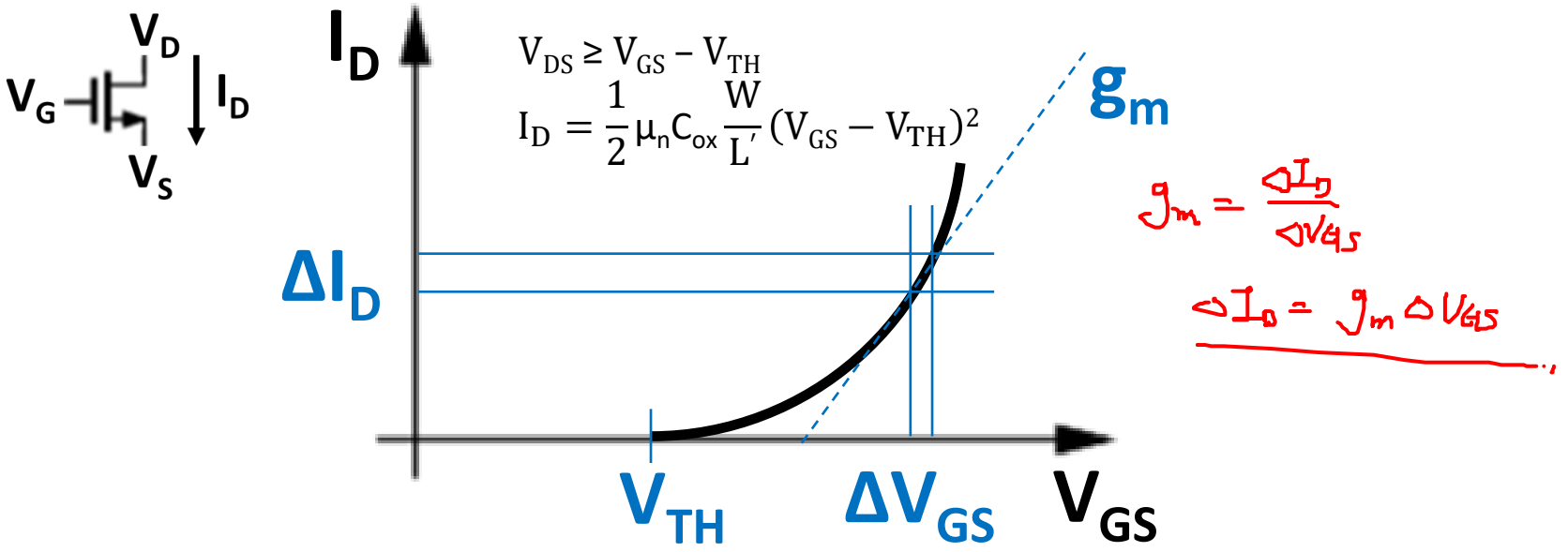


Transconductance

Since a MOSFET operating in saturation produces a current in response to its gate-source overdrive voltage, we may define a figure of merit that indicates **how well a device converts a voltage to a current, called the transconductance (g_m)**: **Change in I_D divided by the change in V_{GS}**. Transconductance is usually defined in the saturation region.

$$g_m = \frac{\partial I_D}{\partial V_{GS}} = \mu_n C_{ox} \frac{W}{L'} (V_{GS} - V_{TH}) = \sqrt{2\mu_n C_{ox} \frac{W}{L'} I_D} = \frac{2I_D}{V_{GS} - V_{TH}}$$

g_m represents the **sensitivity of the device**: for a high g_m, a small change in V_{GS} results in a large change in I_D.

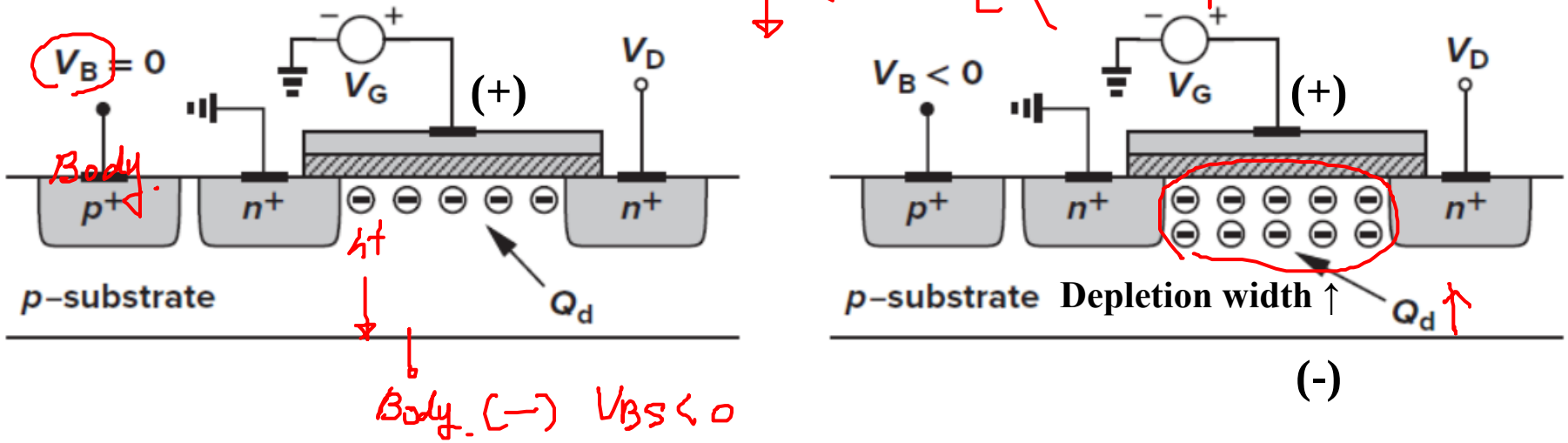


- For a given NMOS, **g_m changes according to the DC biasing** condition.
- If a **small signal** is applied to a NMOS with defined biasing values, we assume the signal amplitude is small enough that the variation in **g_m is negligible**.

Second-Order Effects

Body Effect

Till now, we assumed that the bulk and the source of the MOSFET were tied to ground. We will apply **a bias to the bulk (body)** to see changes in MOSFET operation.



As $\mathbf{V_B}$ **or** $\mathbf{V_{BS}}$ **becomes more negative (**$V_{SB} > 0$**)**, more holes are attracted to the substrate connection, **leaving a larger negative charge behind**. The depletion region becomes wider.

More depletion charge (bound charge) $Q_d \uparrow \rightarrow$ $\mathbf{V_{TH}}$ **increases**

$$V_{TH} = \Phi_{MS} + 2\Phi_F + \frac{Q_{dep}}{C_{ox}}$$

Modified threshold voltage with body effect

$$V_{TH} = V_{TH0} + \gamma(\sqrt{|2\Phi_F + V_{SB}|} - \sqrt{|2\Phi_F|})$$

$$\Phi_F = \frac{kT}{q}\ln\frac{N_{sub}}{n_i} \qquad \gamma = \frac{\sqrt{2q\epsilon_{Si}N_{sub}}}{C_{ox}}$$

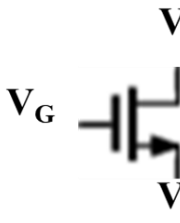
For the transconductance,

$$g_{mb} = \frac{\partial I_D}{\partial V_{BS}} \left(= -\frac{\partial I_D}{\partial V_{SB}}\right) = \frac{\partial I_D}{\partial V_{TH}} \cdot \frac{\partial V_{TH}}{\partial V_{BS}} \qquad I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L'}(V_{GS} - V_{TH})^2$$

$$= -\mu_n C_{ox}\frac{W}{L'}(V_{GS} - V_{TH}) \cdot \frac{\partial V_{TH}}{\partial V_{BS}} = \mu_n C_{ox}\frac{W}{L'}(V_{GS} - V_{TH}) \cdot \frac{\gamma}{2}\frac{1}{\sqrt{|2\Phi_F + V_{SB}|}}$$

$= g_m$ (circled) $= \eta$

$$= g_m \cdot \eta$$



$\Delta I_D = \Delta V_{GS} \times g_m$ (>0)

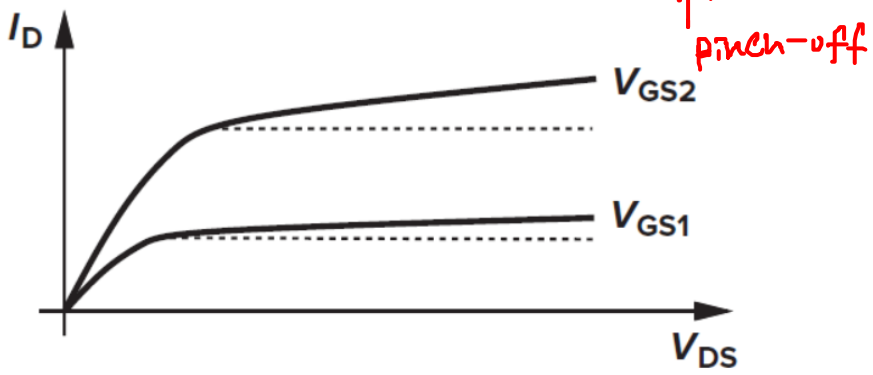
$\Delta I_D = \Delta V_{BS}\ (>0) \times g_{mb}$ ($\Delta V_{SB} \times g_{mb} \uparrow I_D$)

$\gamma = 0$ $g_{mb} = 0$

- V_{GS} positively increases, I_D increases.
- V_{BS} positively increases, i.e. V_{SB} negatively increases, V_{TH} decreases and thus I_D increases.
- Or, V_{SB} leads to changes in V_{TH} and thus I_D.

Channel-Length Modulation

The actual length of the channel gradually decreases as the potential difference between the gate and the drain decreases. L is in fact a function of V_{DS}. This effect is called **Channel-Length Modulation**.



$$L' = L_{eff} - \Delta L$$

$$\frac{1}{L'} = \frac{1}{L_{eff} - \Delta L} = \frac{1}{L_{eff}} \cdot \frac{1}{1 - \frac{\Delta L}{L_{eff}}} \approx \frac{1}{L_{eff}} \cdot \left(1 + \frac{\Delta L}{L_{eff}}\right)$$

$$I_D = \frac{1}{2}\mu_n C_{ox} \frac{W}{L'}(V_{GS} - V_{TH})^2$$

$$I_D = \frac{1}{2}\mu_n C_{ox} \frac{W}{L_{eff}}(V_{GS} - V_{TH})^2 \left(1 + \frac{\Delta L}{L_{eff}}\right)$$

$$= \frac{1}{2}\mu_n C_{ox} \frac{W}{L_{eff}}(V_{GS} - V_{TH})^2(1 + \lambda V_{DS})$$

The channel-length modulation results in a **nonzero slope in the I_D/V_{DS}** characteristic and hence a nonideal current source between D and S in saturation. The parameter **λ represents the relative variation in length for a given increment in V_{DS}**.

With the channel-length modulation,

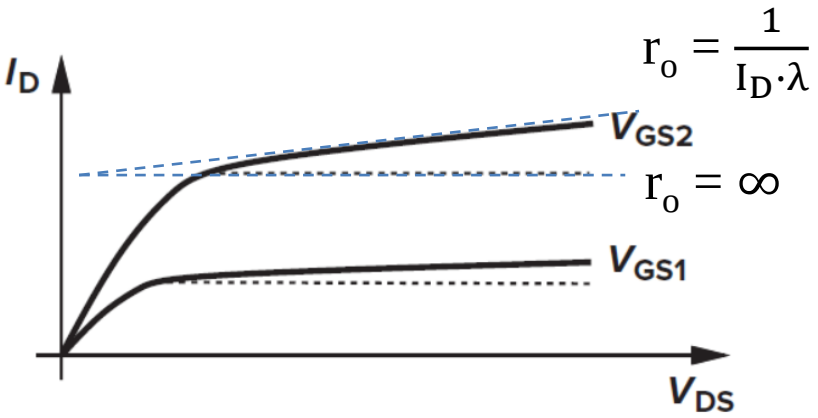
$$I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS}-V_{TH})^2(1+\lambda V_{DS})$$

$$\mathbf{g_m = \frac{\partial I_D}{\partial V_{GS}} = \mu_n C_{ox}\frac{W}{L_{eff}}(V_{GS}-V_{TH})(1+\lambda V_{DS})}$$

$$\mathbf{r_o = \frac{\partial V_{DS}}{\partial I_D} = 1/\frac{\partial I_D}{\partial V_{DS}} = \frac{1}{\frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS}-V_{TH})^2\cdot\lambda}}$$

$$\mathbf{\approx \frac{1}{I_D\cdot\lambda}}$$

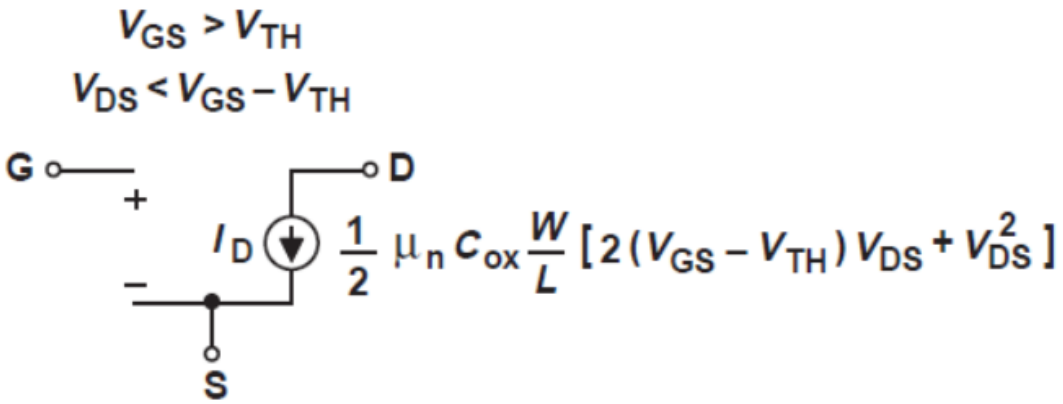
$\lambda = 0 \quad r_o = \infty$

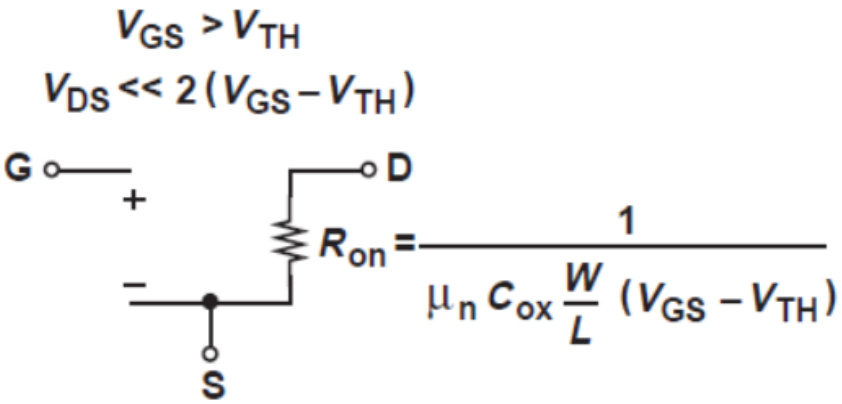


Channel-length modulation is reflected in g_m and r_o by λ

MOS Device Model

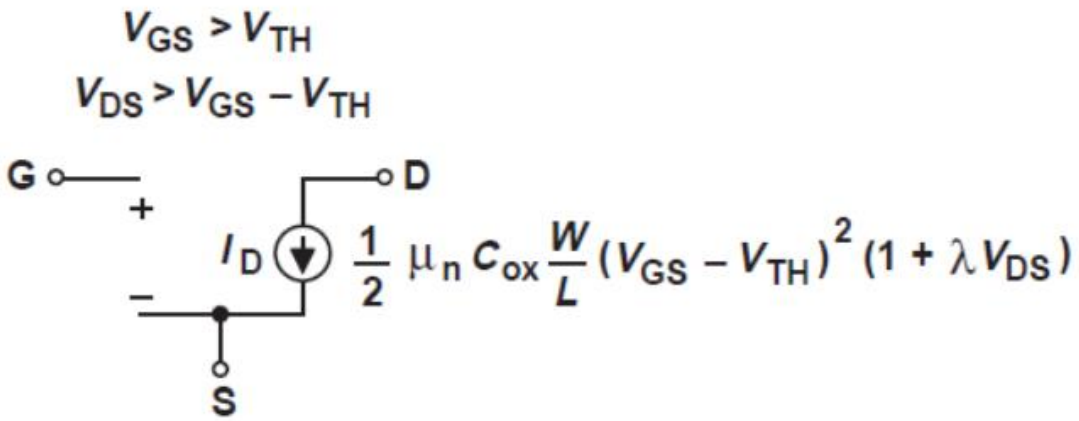
Large-Signal Model





For $\mathbf{V_{DS} < V_{GS} - V_{TH}}$, the model must reflect the triode region, but it can still incorporate a voltage-controlled current source. If $\mathbf{V_{DS} \ll 2(V_{GS} - V_{TH})}$, the transistor can be viewed as a voltage-controlled resistor.

Large-Signal Model



In the saturation region, the transistor acts as **a voltage-controlled current source**. I_D does depend on V_{DS} and is therefore not an ideal current source. In all three cases, the gate remains an open circuit to represent the **zero gate current**.

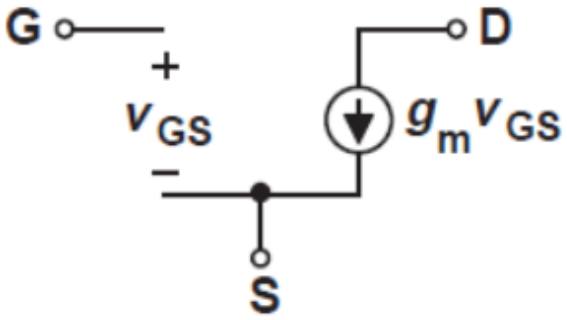
Small-Signal Model (NMOS)

If the bias currents and voltages of a MOSFET are only **slightly disturbed by small signals**, **the nonlinear large-signal models** can be reduced to **linear small signal representations**.

Since in many analog circuits, MOSFETs are biased in **the saturation region**, we derive the corresponding small-signal model.

$$I_D(sat) = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS}-V_{tn})^2$$

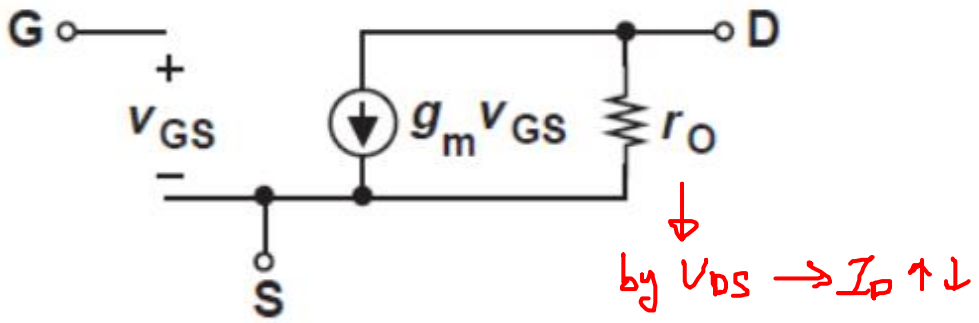
(1) Viewing the MOSFET as a voltage-controlled current source



$\mathbf{i_D = g_m v_{GS}}$ and the gate remains open

$\Delta v_{GS} \uparrow \rightarrow \Delta I_D \uparrow$ (from Drain to Source)

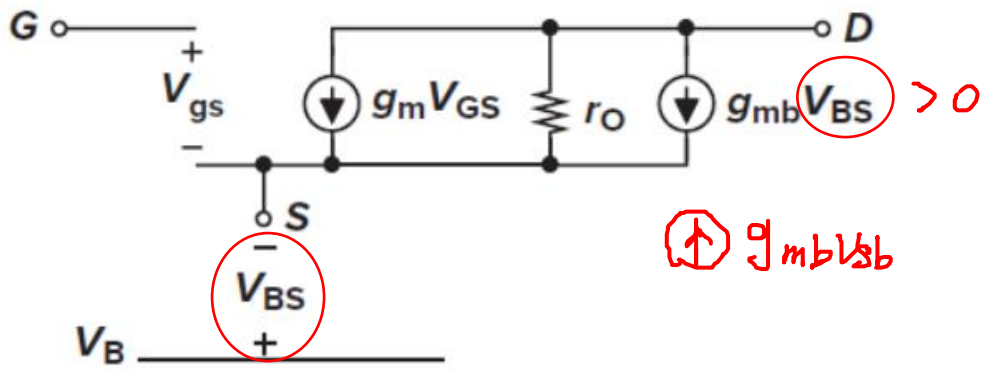
(2) To represent **channel-length modulation**, i.e., variation of i_D with v_{DS}, we add a resistor as below.



$$\mathbf{r_o} = \frac{\partial V_{DS}}{\partial I_D} = 1/\frac{\partial I_D}{\partial V_{DS}} = \frac{1}{\frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS}-V_{TH})^2 \cdot \lambda}$$

$$\approx \frac{\mathbf{1}}{\mathbf{I_D \cdot \lambda}}$$

(3) With all other terminals held at a constant voltage, **I_D is a function of the body voltage ($V_{BS} > 0$ or $V_{SB} < 0$)**. Modeling this dependence by a current source connected between D and S, we add the value as $g_{mb}V_{BS}$ (or $g_{mb}V_{SB}$). At saturation, $\mathbf{g_{mb} = \eta g_m}$



$$g_{mb} = \frac{\partial I_D}{\partial V_{BS}} = \frac{\partial I_D}{\partial V_{TH}} \cdot \frac{\partial V_{TH}}{\partial V_{BS}}$$

$$= \mathbf{g_m \cdot \eta}$$

$\Delta V_{BS}\uparrow$, $V_{TH}\downarrow$ I_D increases

Small-Signal Model for NMOS

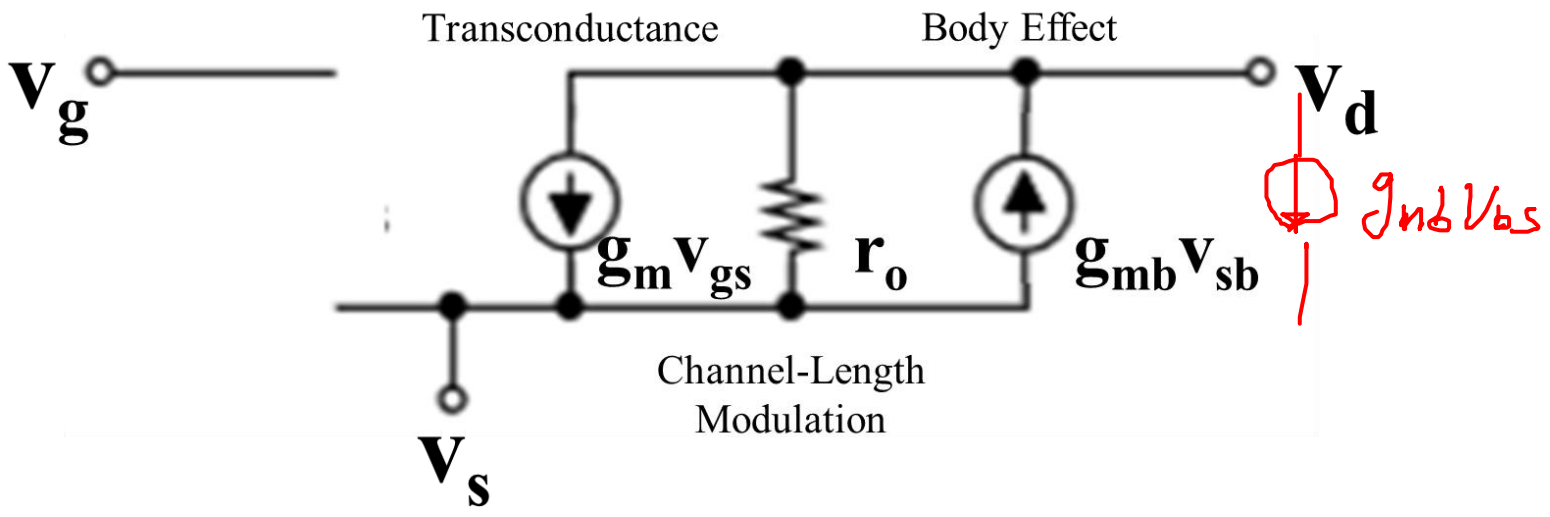
$$V_d = V_D + v_d$$

$$V_g = V_G + v_g$$

$$V_s = V_S + v_s$$

$I_D \downarrow$ $i_d = g_m v_{gs}$ $\uparrow$ $i_d = g_{mb} V_{sb}$

$\downarrow$ $i_d = g_{mb} v_{bs}$



Small-Signal Model for PMOS

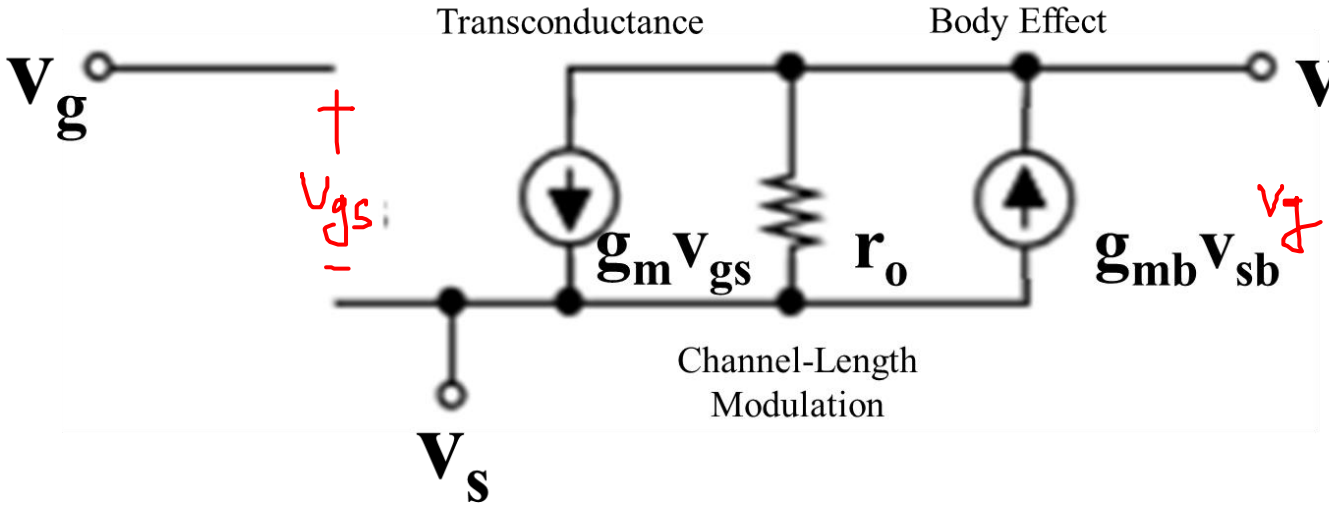
$$V_s = V_S + v_s$$

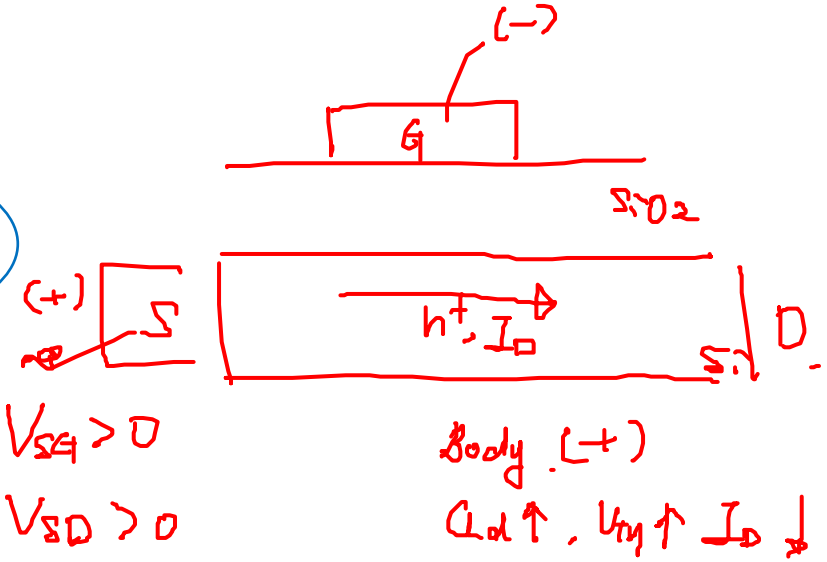
$$V_g = V_G + v_g$$

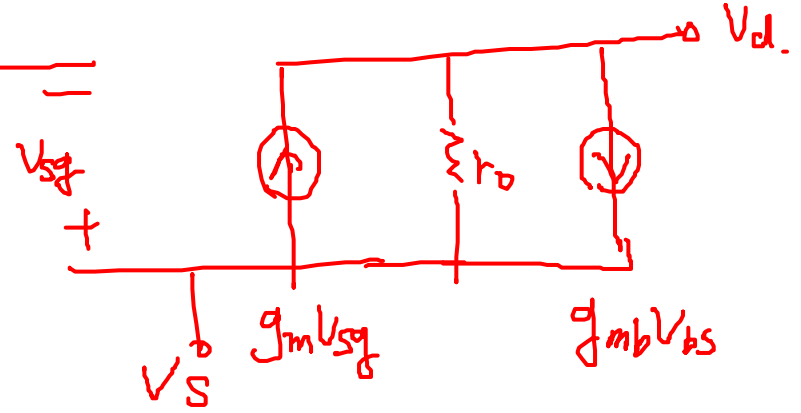
$$V_d = V_D + v_d$$

$I_D \uparrow \; i_d = g_m v_{gs}$

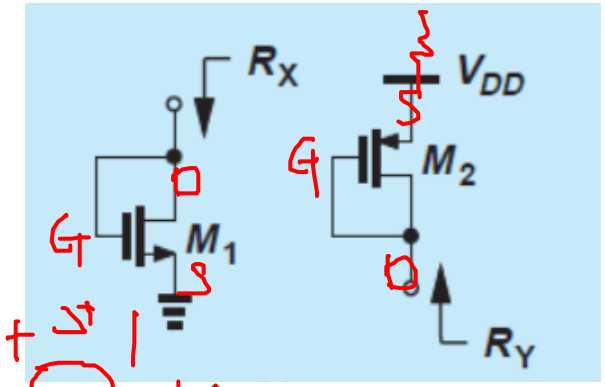
$\downarrow \; i_d = g_{mb} v_{sb}$







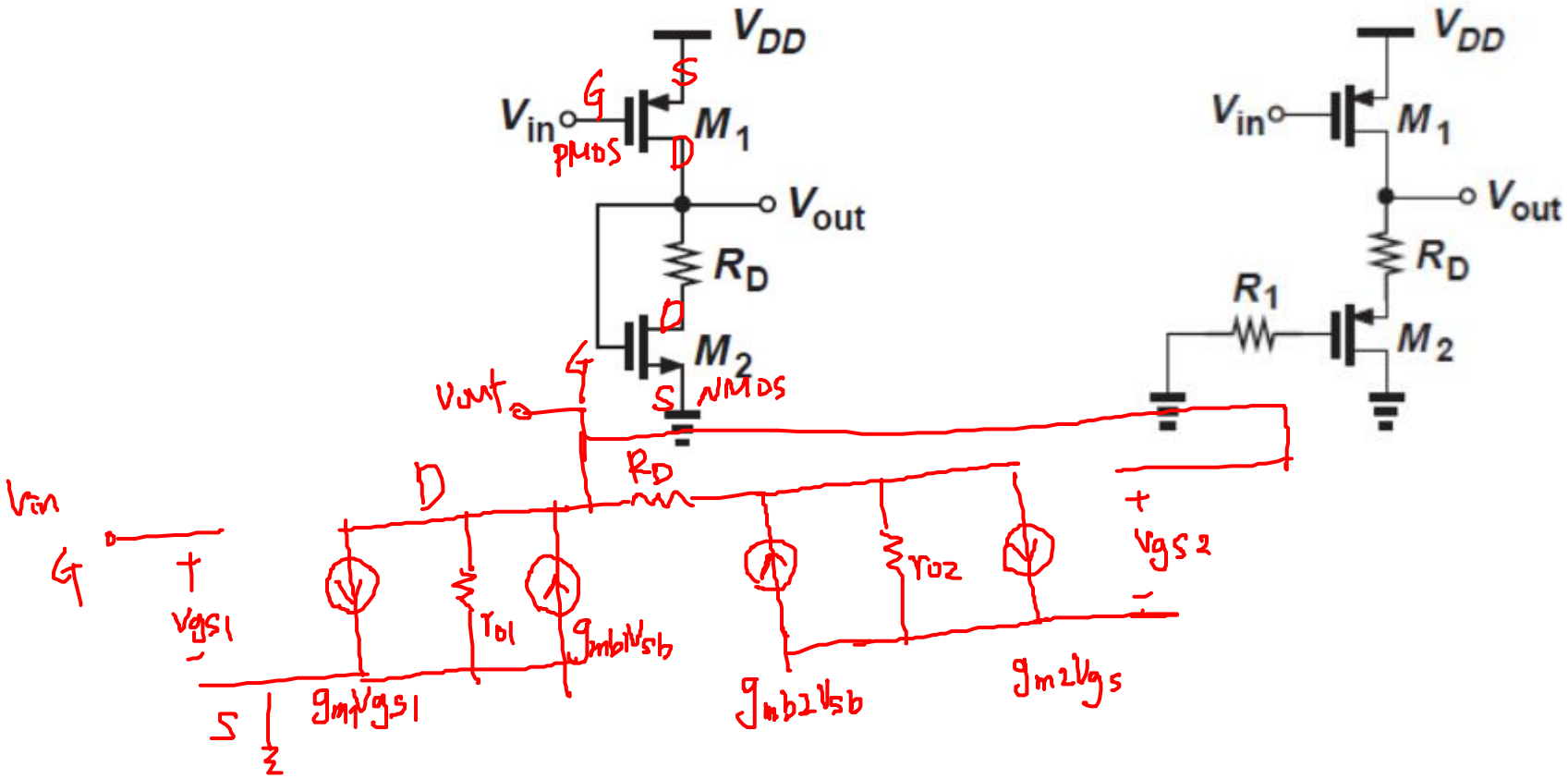
Example 4. For the configurations shown below, determine the **small-signal resistances** R_X and R_Y. Assume $\lambda \neq 0$. **No body effect.**



$$R_X = \frac{1}{g_{m1}} \| r_{o1}$$

$$R_Y = \frac{1}{g_{m2}} \| r_{o2}$$

Example 5. Construct the small-signal model of each circuit if all of the transistors operate in saturation and $\lambda \neq 0$.



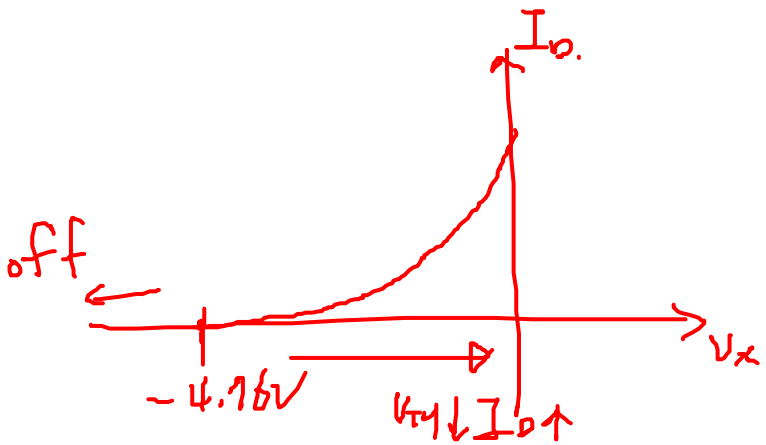
Example 6. Sketch I_D as a function of V_X increasing from $-\infty$ to 0. Assume $V_{TH0} = 0.6$ V, $\gamma = 0.4\ V^{1/2}$ and $2\Phi_F = 0.7$ V.

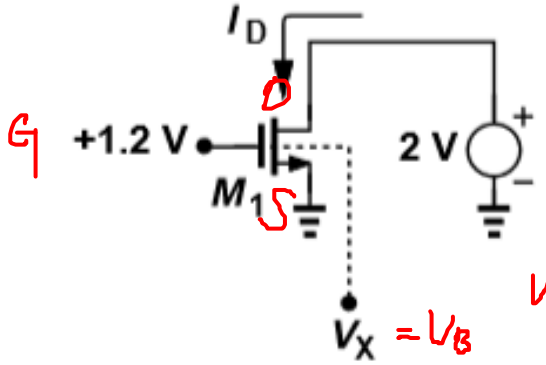
To turn off NMOS

$V_{GS} \geq V_{TH}$

$1.2 \geq a$

$V_X \leq -4.76\,V$





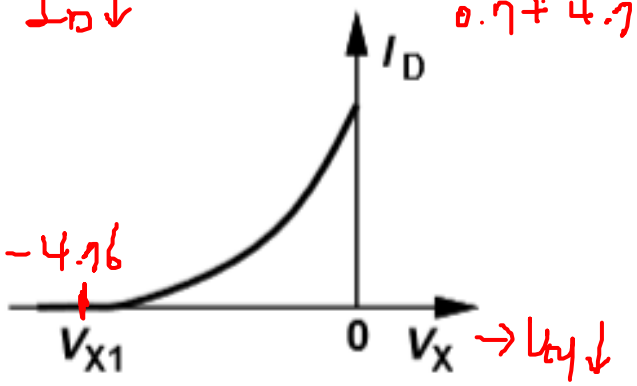
$V_{SB} = -V_X$

$= a$

$= 0.6 + 0.4\left(\sqrt{0.7 - V_X} - \sqrt{0.7}\right)$

$$V_{TH} = V_{TH0} + \gamma\left(\sqrt{|2\Phi_F + V_{SB}|} - \sqrt{|2\Phi_F|}\right)$$

$I_D\downarrow$ $\quad 0.7 + 4.76.$



NMOS vs PMOS in Performance

- PMOS devices are quite inferior to NMOS in most CMOS technology.
- Lower mobility of holes ($\mu_p C_{ox} \approx 0.5\mu_n C_{ox}$) yield lower current drive and conductance.
- NMOS exhibit higher output resistance, providing more ideal current sources and higher voltage gain.
- It is preferable to use NMOS rather than PMOS wherever possible.

$\mu_e < \mu_h$.

$$\frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS}-V_{TH})^2$$

- Please bring a calculator = Sorry, the calculation is dirty
- I have uploaded sample exams = I will show you how things change fast
- I heard that exam was hard last year = I am going to hear it one more time this year
- The exam is easy this year = Other instructors said it is easy
- Please write the detailed processes = I want to know how you got it completely wrong
- There are partial scores = I cannot give you zero score..