MOSFET Fundamentals

for Analog/RF Circuit Design

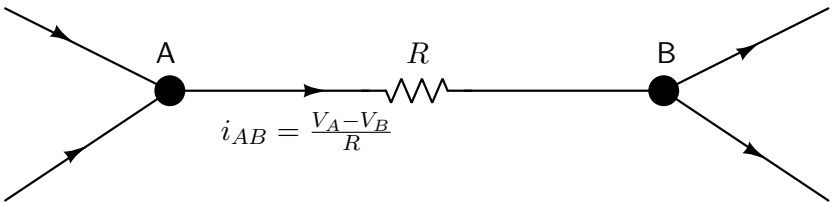
Nagarjuna Nallam

Department of Electrical Engineering
IIT Kanpur, India

Contents

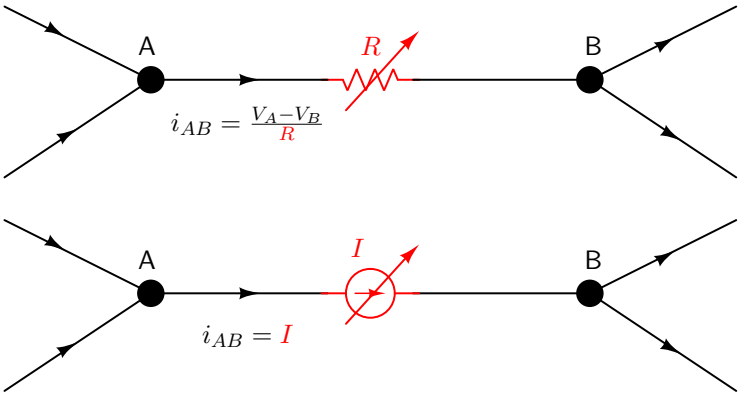
MOSFET Introduction and Operation

Transistor



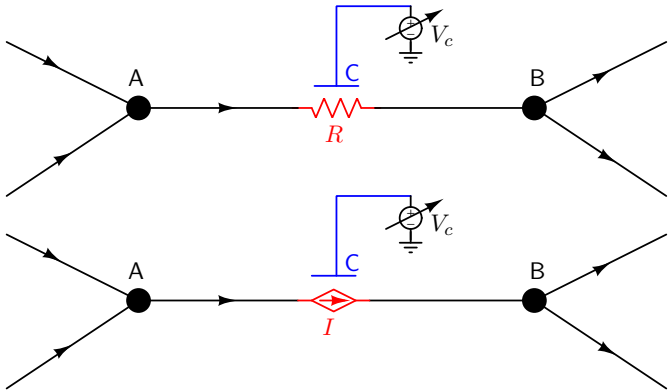
- How can we control the current flowing between the nodes A and B in the figure?

Transistor



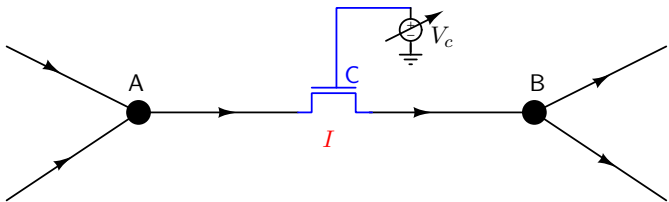
- We need a third node to (electronically) tune the resistance or current-flow without disturbing other circuit connections.

Transistor



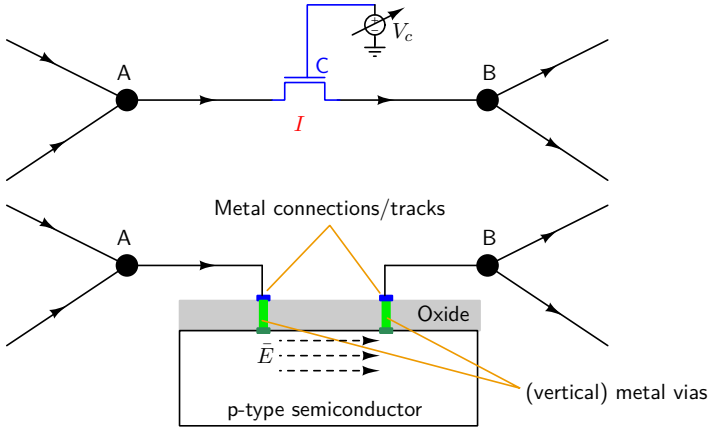
- Transistor - Controlling either the resistance or current flow between two terminals by varying the voltage or current-through the other terminal!

Why semiconductor?



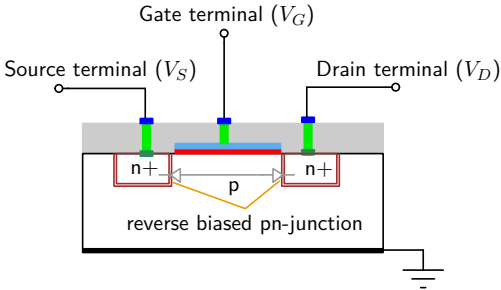
- Metals and Dielectrics are not useful as variable resistors. (why?)
- Let's use a doped semiconductor and build a transistor.

Why not just semiconductor?



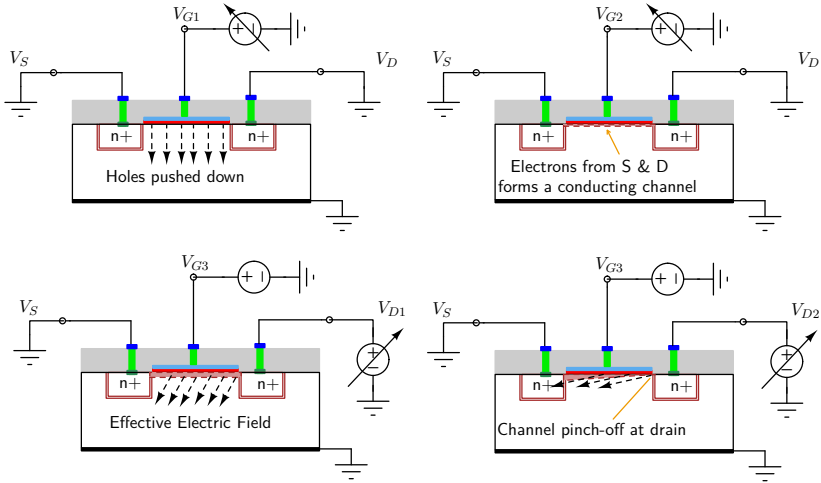
- How to control the current or resistance? How to add the third terminal?
- Can we get all values of resistances from few Ohms to Mega Ohms?

Metal-Oxide-semiconductor Field-Effect Transistor

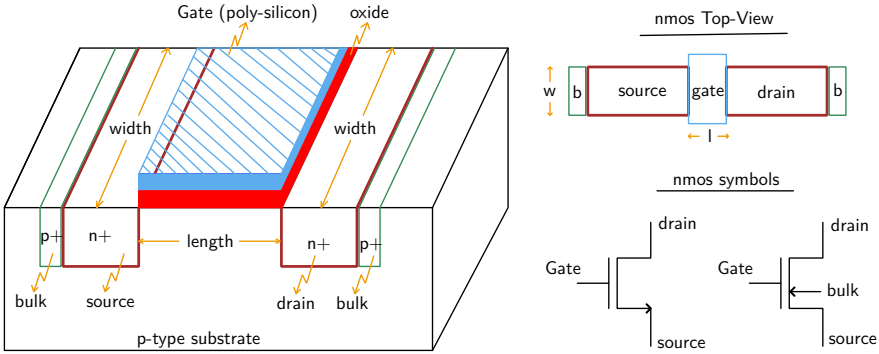


- We have added a thin oxide layer and a metal (poly-Si in modern technologies) above the current conducting region to control the electric field capacitively.
- We have also added n+ regions at the A and B terminals. These n+ regions form two back to back diodes with p-type substrate. If we connect the substrate to the lowest potential in the circuit, the diodes are reverse biased and no current flows between source and drain under normal conditions (for zero gate voltage).

nMOSFET operation

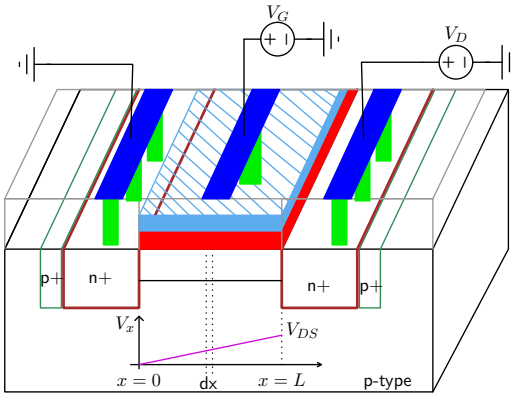


nMOSFET: Cross-sectional view



- For fabrication, we draw the top-view of all (physical) layers and send it to fab. We need not to draw some layers - oxide/dielectric layers like in a PCB design as they are only used for isolation between conducting layers.

nMOSFET: I-V Characteristics



Assumptions:

Gate is an equvi-potential surface

V_{DS} increases linearly from S to D

Overdrive Voltage:

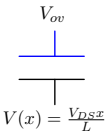
Out of total gate-to-source voltage V_{GS} V_{Th} is used to keep the channel intact and remaining $V_{GS} - V_{Th}$ will be used to control the channel current.

Overdrive Voltage $V_{ov} = V_{GS} - V_{Th}$

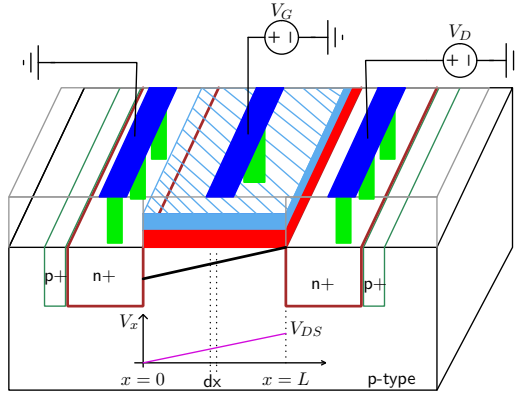
Charge per unit length in the channel at a distance $x = C_{ox}WV_{ov}$ when $V_{DS} = 0$

$$q_d(x) = C_{ox}W(V_{ov} - V(x)) \text{ when } V_{DS} \neq 0$$

where C_{ox} is the gate oxide capacitance per unit area



nMOSFET: I-V Characteristics



$$q_d(x) = C_{ox}W(V_{ov} - V(x))$$

Current $I = q_d(x).v_{drift}$

Drift velocity, $v_{drift} = \mu_n|E_H|$

Current, $I_D = C_{ox}W(V_{ov} - V(x))\mu_n\frac{dV}{dx}$

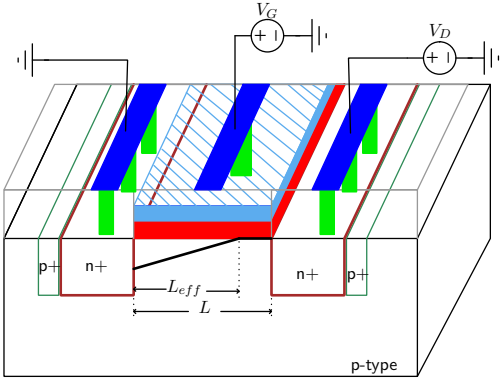
$$\int_{x=0}^{L} I_D dx = \int_{V(x)=0}^{V_{DS}} \mu_n C_{ox} W(V_{ov} - V(x))dV$$

$$\boxed{I_D = \mu_n C_{ox}\frac{W}{L}(V_{ov}V_{DS} - \frac{V_{DS}^2}{2})}$$

Once the pinch-off occurs at the drain (when $V_{DS} \geq V_{ov}$), the current does not increase further with V_{DS}.

Saturation Current, $\boxed{I_{Dsat} = \mu_n C_{ox}\frac{W}{L}(\frac{V_{ov}^2}{2})}$

nMOSFET: Channel Length Modulation



~~The current does not change with V_{DS} in saturation.~~

$$\frac{1}{L_{eff}} = \frac{1}{L-\Delta L} \approx \frac{1+\Delta L/L}{L}$$

Let $\Delta L/L = \lambda V_{DS}$

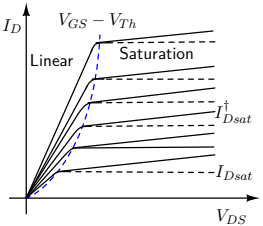
$$I_{Dsat} = \mu_n C_{ox} \frac{W}{L}(\frac{V_{ov}^2}{2})(1+\lambda V_{DS})$$

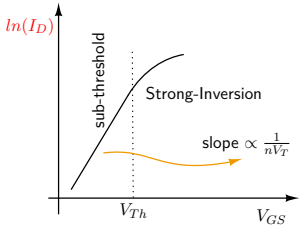
λ is a model parameter; $\lambda = \frac{1}{V_E L}$

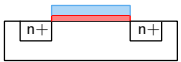
V_E is similar to Early voltage in BJTs and is constant for a given technology

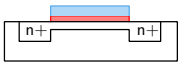
- Note: In a MOSFET source and drain are interchangeable. Terminal with highest potential acts as a drain and the other one acts as a source.

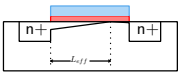
nMOSFET: I-V Characteristics











$$I_{Dsub} = I_{D0}\frac{W}{L}e^{\frac{V_{GS}}{nV_T}}$$

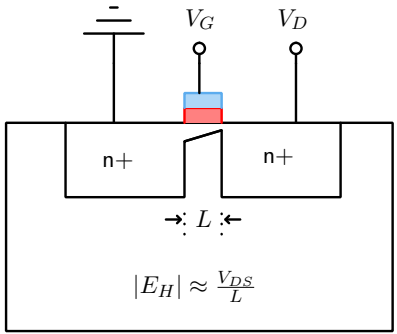
Saturation in subthreshold: $V_{GS} < V_{Th}$ and $V_{DS} > 3V_T$; $V_T = \frac{kT}{q}$

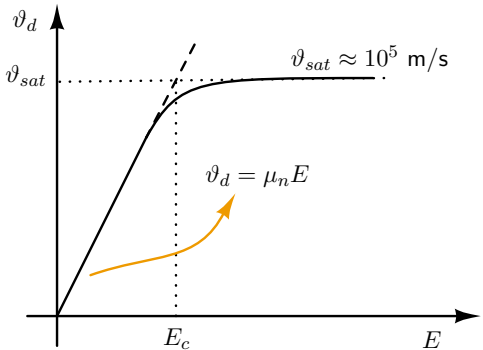
$$I_D = \mu_n C_{ox}\frac{W}{L}((V_{GS} - V_{Th})V_{DS} - \frac{V_{DS}^2}{2})$$

$$I_{Dsat} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{Th})^2$$

$$I^{\dagger}_{Dsat} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{Th})^2(1 + \lambda V_{DS})$$

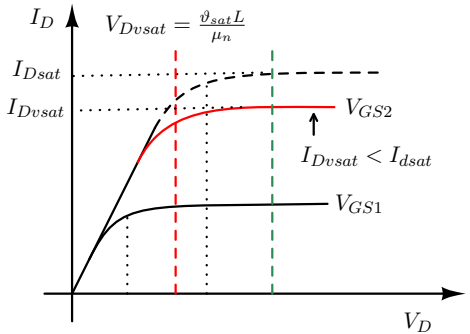
nMOSFET: Velocity Saturation

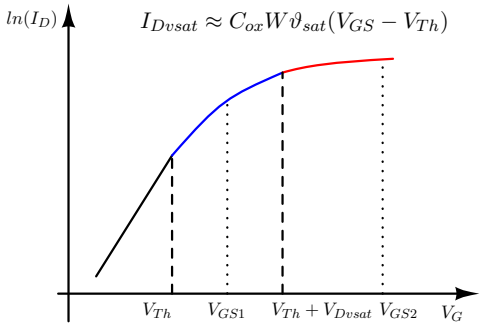




- As the channel length decreases, the electric field in the channel increases. The drift velocity of electrons saturates after a critical field (E_c).
- If $V_{DS} > \frac{\vartheta_{sat} L}{\mu_n}$, the transistor current enters the velocity saturation. Saturated drift velocity of electrons is $\approx 10^5$ m/s.

nMOSFET: Velocity Saturation





- In velocity saturation, current $I_{Dvsat} = q_d.\vartheta_{sat} = C_{ox} W (V_{GS} - V_{Th}) \vartheta_{sat}$.
- Note that I_{Dvsat} is independent of channel length, and proportional to V_{ov}.
- Also, $I_{Dvsat} < I_{Dsat}$ and $V_{Dvsat} = \frac{\vartheta_{sat} L}{\mu_n}$ is a constant for a given technology and channel length.

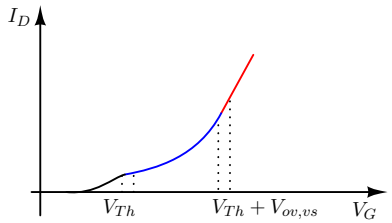
nMOSFET: A note on V_{ov} for velocity saturation voltage

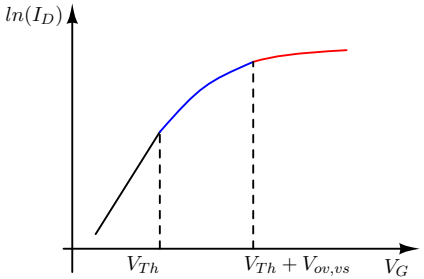
- What we have seen so far, is the velocity saturation due to high V_{DS} (i.e, high horizontal electric field) at low-to-moderate V_{GS}.
- High V_{GS} (or high vertical electric field), also causes the electron mobility degradation even for low-to-moderate V_{DS}.
- The effect of high V_{GS} on current is usually modelled by a parameter θ.

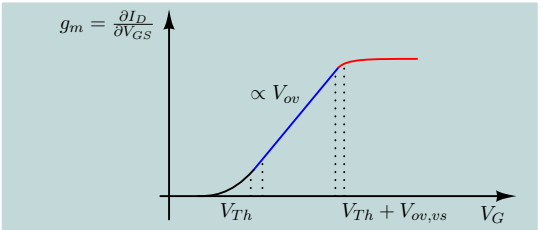
$$I_D = \mu_n C_{ox} \frac{W}{L} \frac{V_{ov}^2}{1 + \theta V_{ov}}$$

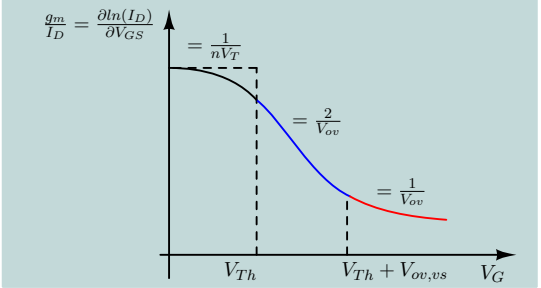
- We define velocity saturation at which $g_m = \frac{\partial I_D}{\partial V_{GS}}$ becomes constant. If we equate the g_m expressions for saturation (use the above expression involving θ) and velocity saturation (from prev. slide), we can show that $\boxed{V_{ov,vs} = \frac{1}{\theta} = 2n \frac{\vartheta_{sat}}{\mu_n} L}$.
- A rough estimate for velocity saturation is $\boxed{V_{ov,vs} \approx 5L}$; where L is the channel length in μm.

nMOSFET: Transition between regions

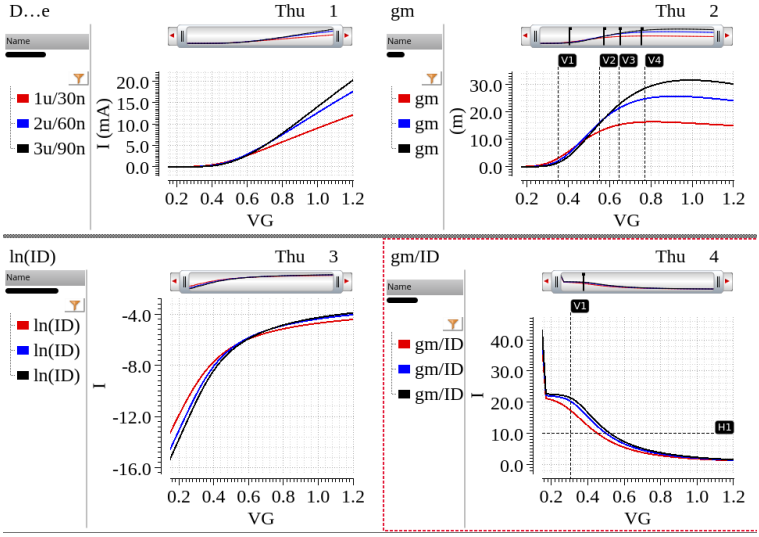




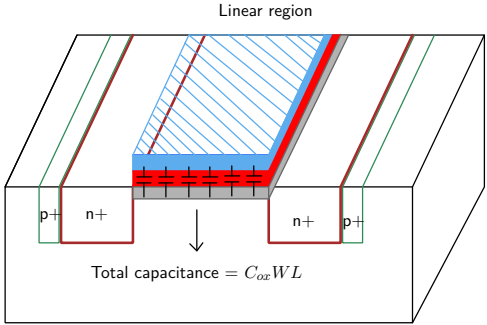


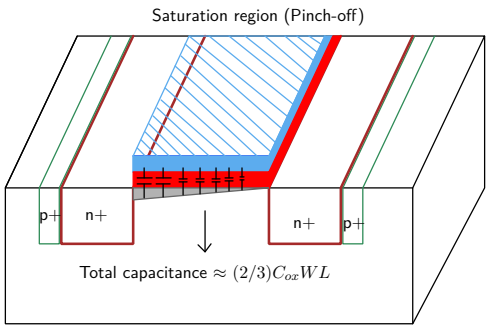


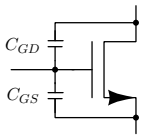
nMOSFET: Simulated IV Characteristics in 28nm CMOS



MOSFET: Channel Capacitance



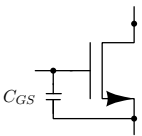




No-terminal called channel

We assign half of channel capacitance to source and other half to drain

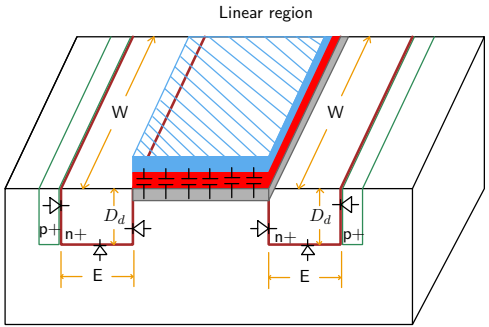
$C_{GS} = C_{GD} = \frac{1}{2}C_{ox}WL$

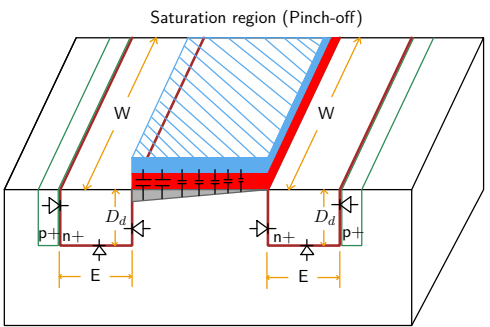


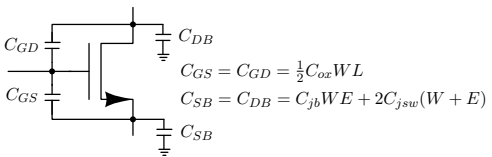
$C_{GS} \approx \frac{2}{3}C_{ox}WL$

$C_{GD} \approx 0$ due to channel

MOSFET: Junction Capacitances

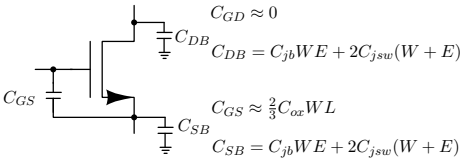






$$C_{GS} = C_{GD} = \tfrac{1}{2} C_{ox} W L$$

$$C_{SB} = C_{DB} = C_{jb} W E + 2 C_{jsw} (W + E)$$



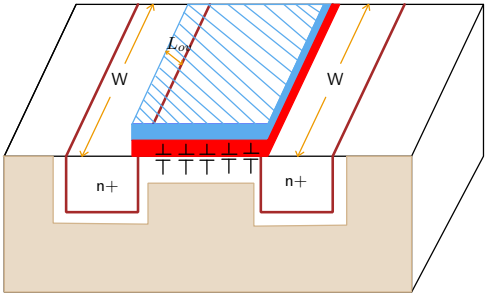
$$C_{DB} = C_{jb} W E + 2 C_{jsw} (W + E)$$

$$C_{GS} \approx \tfrac{2}{3} C_{ox} W L$$

$$C_{SB} = C_{jb} W E + 2 C_{jsw} (W + E)$$

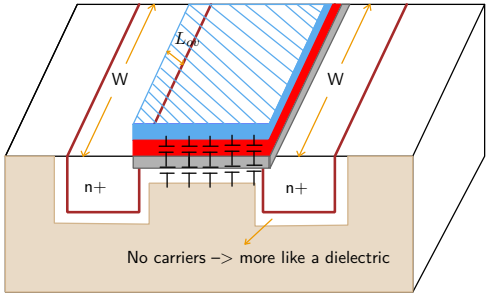
C_{jb} is per unit area and C_{jsw} is per unit length.

MOSFET: Overlap and Depletion Capacitances



When $V_{GS} < V_{Th}$, bottom plate is p-sub (C_D).

Overlap capacitance $= C_{ov} W L_{ov}$ is always present.

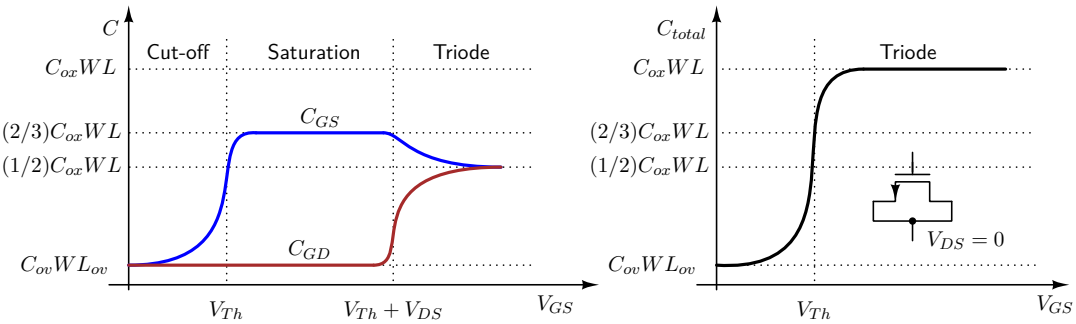


When $V_{GS} > V_{Th}$, two capacitors are in series.

One is gate-to-channel (C_{ox}) and the other is channel-to-substrate ($C_D = \frac{\epsilon_{si}}{t_{si}}$).

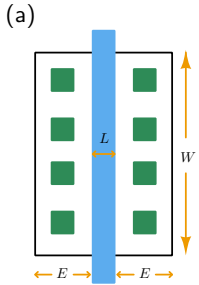
$C_D/C_{ox} = n - 1$, where n is the subthreshold gate coupling coefficient. Typically n ranges from 1.2 to 1.5

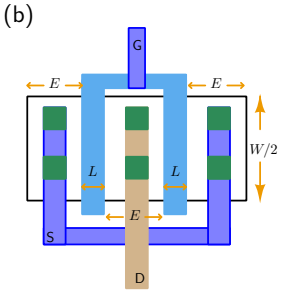
MOSFET: CV Characteristics

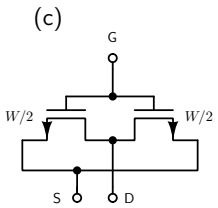


- Overlap capacitance is always present, but shown only when oxide capacitance is not there in the above figure. Also, junction capacitances are not included.
- In many applications, MOSFET is also used as a capacitor by shorting source and drain terminals as shown above.

MOSFET: Folding or Gate-Fingers







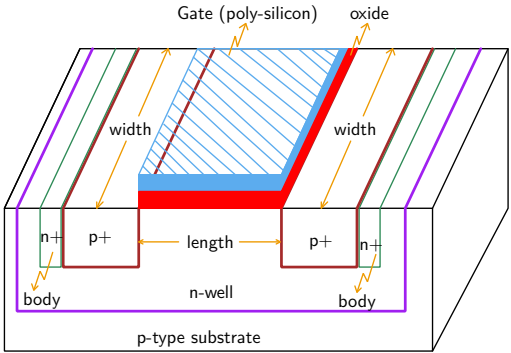
$$C_{GS} = C_{GD} = C_{jb}WE + 2C_{jsw}(W + E)$$

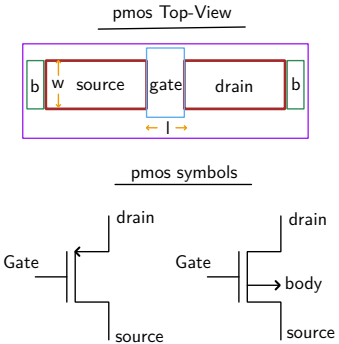
$$C_{GD} = C_{jb}\frac{WE}{2} + 2C_{jsw}(\frac{W}{2} + E)$$

$$C_{GS} = C_{jb}WE + 2C_{jsw}(W + E)$$

- We can reduce the junction capacitance at drain by using two fingers of $W/2$ width instead of a single finger with width W.
- The real benefit is coming due to shared drain (diffusion) between two half-transistors.

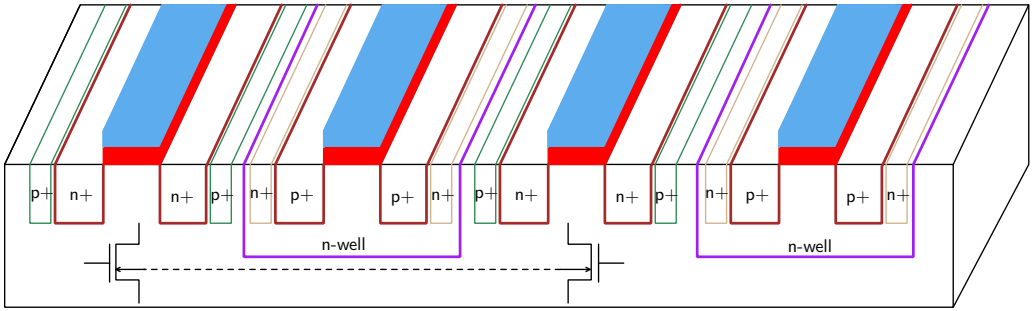
pMOSFET: Cross-sectional view





- pmos is realized in an n-well on the same p-type substrate as nmos.

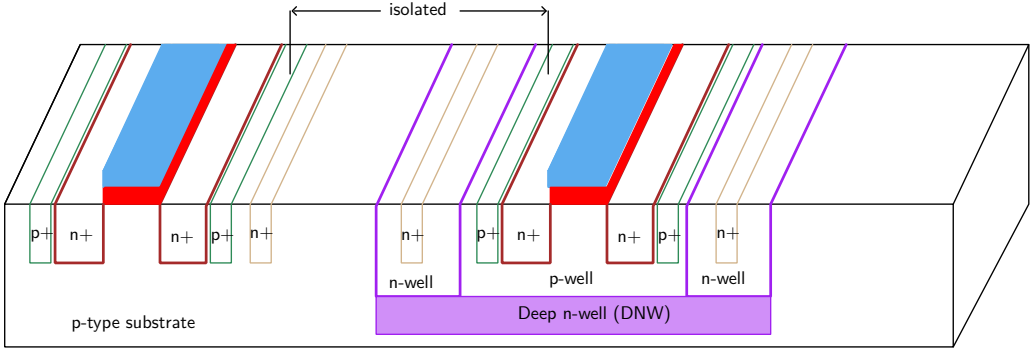
Body/bulk contact



All nmos share the same p-type substrate

- In majority of designs (definitely in digital design), nmos devices are realized on the same substrate, and hence their body/bulk terminals are shorted (via substrate).
- Unlike nmos, each pmos is realized in it's own n-well in majority of designs. Therefore the body terminal of each pmos device can be connected individually to any node in a circuit.

Tripple-well process



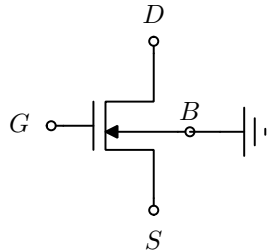
- nmos transistor can also be realized in a p-well (isolated by n-wells on both-sides + at the bottom) if the process supports (Tripple-well process). The bottom layer is called deep N-well layer. Mainly used for critical analog/RF circuits only.
- Draw the top-view of this transistor showing all layers.

Effect of source-to-body potential

$$V_{Th} = V_{Th0} + \gamma[\sqrt{2\Phi_F + V_{SB}} - \sqrt{2\Phi_F}]$$

where γ is the body-effect parameter, and

$\Phi_F = V_T ln(\frac{N_A}{n_i})$ is the bulk Fermi potential



As the source voltage increases

↓

Width of deplection region (reverse biased pn-junction) increases

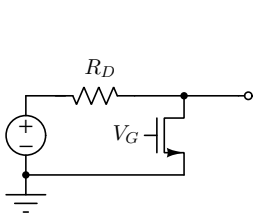
↓

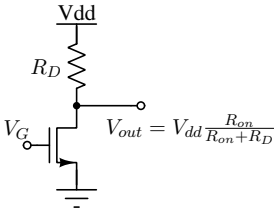
Threshold voltage increases as the amount of charge in depletion region increses

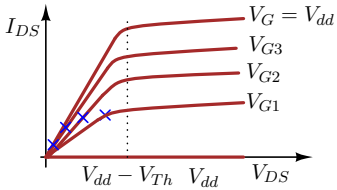
- Most common scenarios: series switches, common-gate and common-drain stages, common-source with a resistor in the source.

MOSFET Models for Circuit Analysis and Design

MOSFET as a tunable resistor



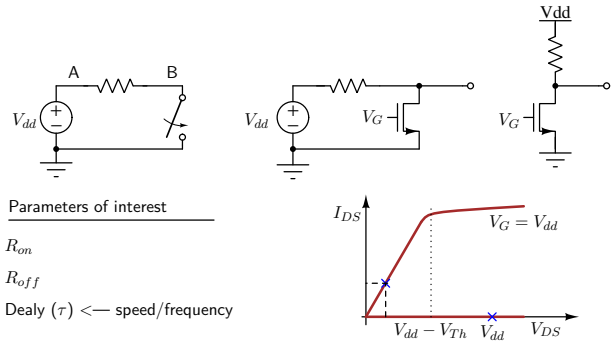




For small values of V_{DS}, $I_D \approx \mu_n C_{ox} \frac{W}{L}(V_{GS} - V_{Th})V_{DS}$

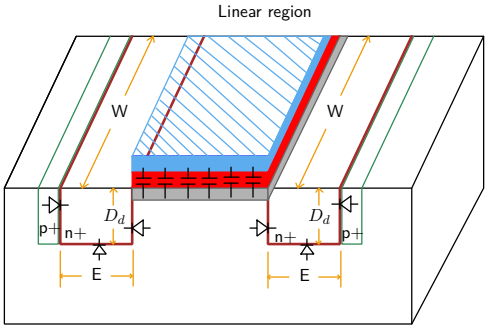
$$\Rightarrow \boxed{R_{on} = \frac{V_{DS}}{I_D} = \frac{1}{\mu_n C_{ox} \frac{W}{L}(V_{GS} - V_{Th})}}$$

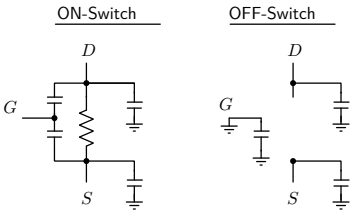
MOSFET as a switch

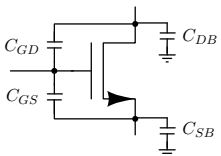


- MOSFET behaves like a resistor in the linear region –> an ON-switch.
- MOSFET behaves like an open circuit in the cut-off region –> Off-switch.

MOSFET: Switch model







$$C_{GS} = C_{GD} = \tfrac{1}{2} C_{ox} W L$$

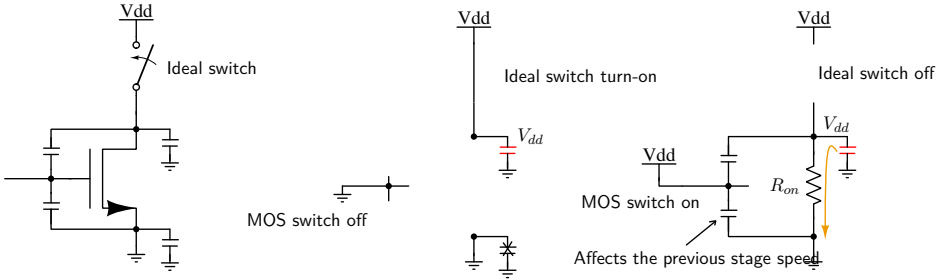
$$C_{SB} = C_{DB} = C_{jb} W E + 2 C_{jsw} (W + E) D_d$$

Note: In cut-off, no channel is prsent.

$C_{GS} = C_{GD} = 0$ for off-switch

but C_{DB}, and C_{SB} are present

Switch Size Vs. Performance



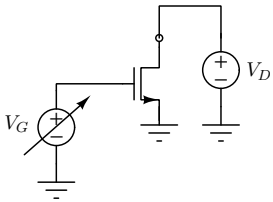
Larger the C_{off}, larger is the charge it stores during off-period.

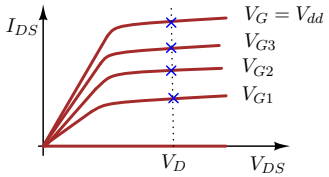
When the switch turns-on, $C_{off}V_{dd}$ needs to be dicharged through R_{on}

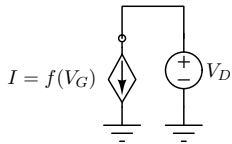
$R_{on}C_{off}$ is a good measure of switch intrinsic speed. ← Smaller is faster.

$$R_{on}C_{off} \approx \frac{C_j(E+2D_d)L_{min}}{\mu_n C_{ox}(V_{dd}-V_{Th})}$$ is independent of switch width.

MOSFET as a Variable Current Source

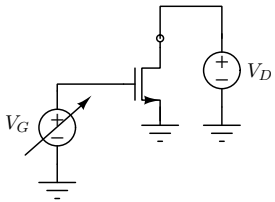


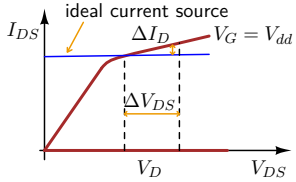


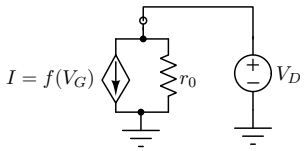


- For sufficiently high fixed V_{DS}, $I_D \approx \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{Th})^2$; Neglecting channel-length modulation
- MOSFET in saturation acts as a variable current source whose value is controlled by the gate voltage or overdrive voltage. A voltage-controlled-current-source (VCCS) model is shown above.
- Note: We are discussing about DC current here.

MOSFET: r_0





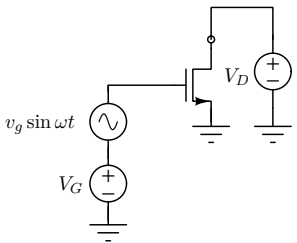


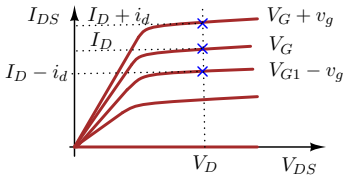
$$r_0 = \frac{\Delta V_{DS}}{\Delta I_D} = \frac{1}{\partial I_D / \partial V_{DS}} \approx \frac{1}{\lambda I_D}$$

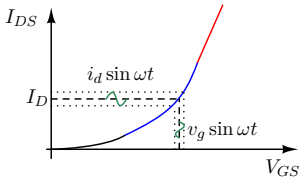
- MOSFET output current depends on V_{DS} due to channel-length modulation. This can be modeled by a resistor between drain and source terminals. Ideal current source should have ∞ output impedance.
- The output impedance of the MOSFET current source is an incremental resistance.

$$\boxed{r_0 \approx \frac{1}{\lambda I_D}}$$

MOSFET as a Transconducting Amplifier



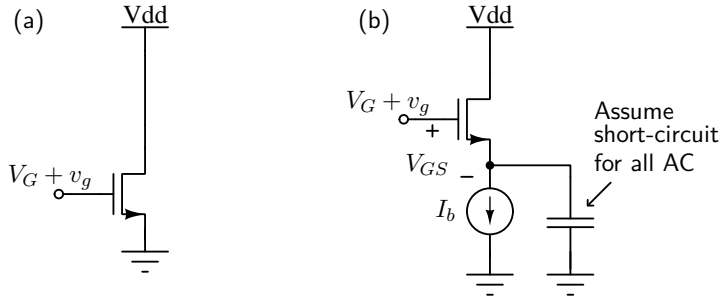




Transconductance, $g_m = \frac{\partial I_D}{\partial V_{GS}} = \frac{i_d}{v_{gs}}$

- If we add a small sinusoidal signal ($v_g \sin \omega t$) to the gate voltage V_G, the drain current contains multiple harmonics of ω depending on the order of nonlinearity. The amplitude of the fundamental current depends on the slope of $I_D - V_{GS}$ curve.
- Different expressions for g_m in S.I. : $g_m = \mu_n C_{ox} \frac{W}{L} V_{ov} = \sqrt{2\mu_n C_{ox} \frac{W}{L} I_D} = \frac{2I_D}{V_{ov}}$

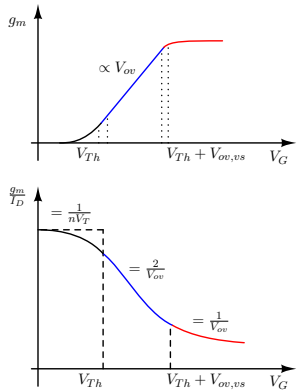
Transconductance in Strong Inversion



- What is the DC current flowing through the transistor in (a) and (b)?
- How does the g_m vary with V_G and W/L in (a)?
- How does the g_m vary with I_b and W/L in (b)?

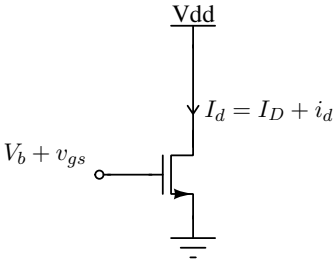
Transconductance Efficiency - g_m/I_D

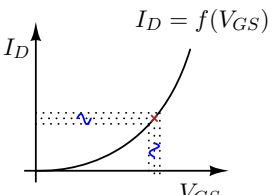
Region	I_D	g_m	g_m/I_D
Weak Inv.	$I_{D0}(W/L)e^{\frac{V_{GS}}{nV_T}}$	$\frac{I_D}{nV_T}$	$\frac{1}{nV_T}$
Strong Inv.	$(1/2)\mu_n c_{ox}(W/L)V_{ov}^2$	$\frac{2I_D}{V_{ov}}$	$\frac{2}{V_{ov}}$
Velocity Sat.	$c_{ox}W\vartheta_{sat}(V_{ov})$	$\frac{I_D}{V_{ov}}$	$\frac{1}{V_{ov}}$

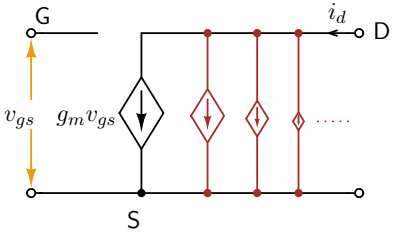


- g_m/I_D represents transconductance (dc-to-ac conversion) efficiency.
- Except for n, g_m/I_D is almost independent of technology parameters!
- $1/V_T = 40/\text{v}$ at room temperature. For MOSFET, $g_m/I_D \leq 25$.
- $g_m/I_D = 10$ is a good compromise between speed and efficiency, and is most commonly used at moderate frequencies.

Transconductance Nonlinearity







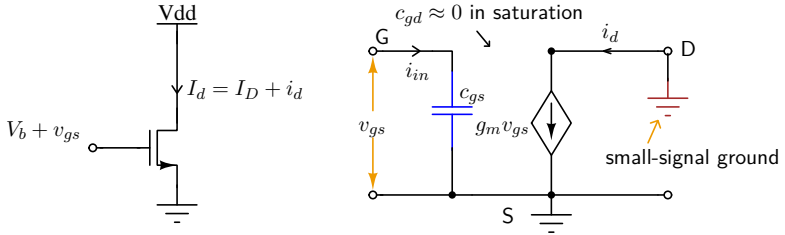
$$I_d = I_D + f'(V_b)v_{gs} + \frac{f''(V_b)}{2!}v_{gs}^2 + \frac{f'''}{3!}v_{gs}^3 + \dots$$

$$i_d = c_1 v_{gs} + c_2 v_{gs}^2 + c_3 v_{gs}^3 + \dots$$

- For a sinusoidal input, $v_{gs} = A\cos\omega t$, considering upto third order non-linearity

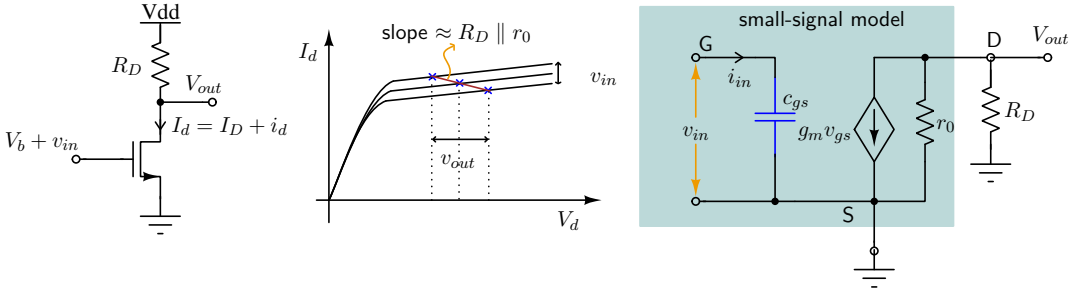
$$i_d = \frac{c_2 A^2}{2} + (c_1 A + \frac{3c_3 A^3}{4})\cos\omega t + \frac{c_2 A^2}{2}\cos 2\omega t + \frac{c_3 A^3}{4}\cos 3\omega t$$

Unity current gain frequency (f_t)



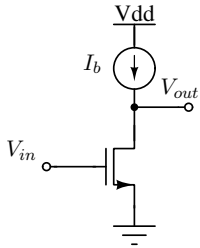
- At low frequencies, the input impedance at the gate is very high and can be considered as an open circuit.
- At high frequencies, the input impedance starts reducing due to the gate capacitance (c_{gs}). The input current into the gate is $i_{in} = v_{gs}(sc_{gs})$, and current gain is $i_{out}/i_{in} = g_m/(sc_{gs})$.
- We can use the MOSFET as a transconductor upto a frequency at which the (short-circuited) current gain is unity, $\boxed{f_t = g_m/(2\pi c_{gs}) \propto 1/L^2}$ (for saturation). This is often used to compare different technologies.

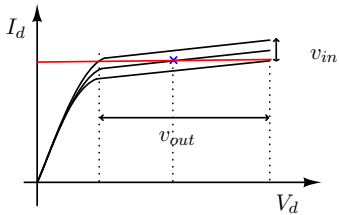
Voltage Amplification

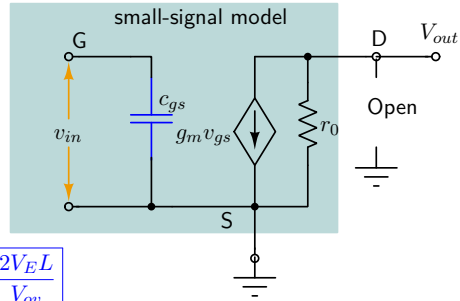


- Add a resistor at the drain so that the current generated by MOSFET (by its transconductance) flows through the resistor. The small signal output voltage $v_{out} \approx -g_m R_D$.
- If we consider the output impedance of the MOSFET, $r_0 = 1/(\lambda I_D)$, the voltage gain is $\approx g_m(R_D \parallel r_0)$. As seen previously, at high frequencies input capacitance should also be considered.

Intrinsic Voltage Gain





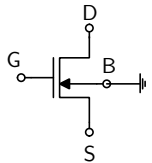


$$A_{oc} = g_m r_0 \approx \frac{2}{\lambda V_{ov}} = \frac{2V_E L}{V_{ov}}$$

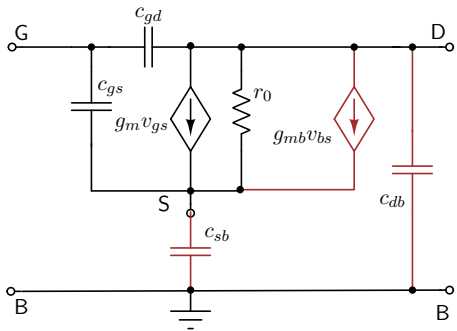
- We can achieve maximum voltage gain, called intrinsic gain, if $R_D \to \infty$. One possibility is having an ideal current source at the drain.
- $A_{oc} = g_m r_0 = \frac{V_E L}{V_{ov}}$. Intrinsic gain is proportional to length!!

Small-Signal Model Including Body Effect

Valid if the transistor is in Saturation.

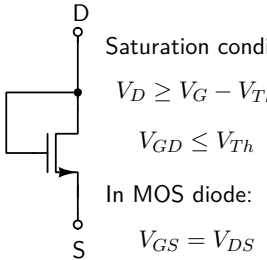


$$g_{mb} = \frac{\partial I_D}{\partial V_{BS}} = \frac{\partial I_D}{\partial V_{Th}} \cdot \frac{-\partial V_{Th}}{\partial V_{SB}} = g_m \frac{\gamma}{2\sqrt{2\Phi_F + V_{SB}}} = g_m \eta$$



- Since the threshold voltage depends on the source-to-bulk potential, drain current also depends on V_{BS}. In other words, V_{BS} has similar effect as V_{GS} and hence bulk is called back-gate. v_{bs} also gives a small-signal current $g_{mb}v_{bs}$ which is in parallel with $g_m v_{gs}$ in the model.
- All capacitances are shown in the small signal model above. c_{gd} in saturation is due to overlap and layout-parasitic capacitances. C_{gb} is negligible in saturation.

MOS Diode



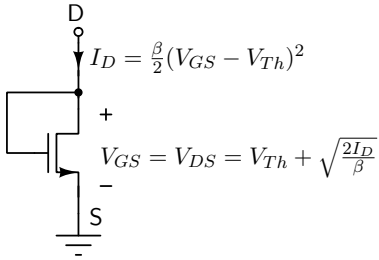
Saturation condition:

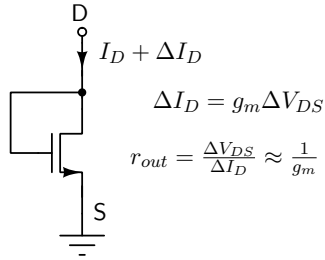
$V_D \geq V_G - V_{Th}$

$V_{GD} \leq V_{Th}$

In MOS diode:

$V_{GS} = V_{DS}$





- When the gate is shorted to the drain, the transistor is always in saturation (can be either in strong inversion or weak inversion depending on the current flowing).
- MOSFET, in general, is a unidirectional device $\Rightarrow$ applying a current at drain does not generate a voltage at gate! But in a MOS diode, we can generate the V_{GS} $(= V_{DS})$ by applying a current I_D at the drain. $\leftarrow$ Inverse function!
- The incremental or small-signal impedance seen between the two terminals of the MOS diode is $\approx 1/g_m$. If $V_{SB} > 0$, the $r_{out} \approx 1/(g_m + g_{mb})$.

References

Text book

- Razavi B., "Design of analog CMOS integrated circuits," 2nd ed., McGraw-Hill Education, 2017.

For step-by-step design examples

- Phillip E. Allen, Douglas R. Holberg, "CMOS Analog Circuit Design," Oxford University Press, 2002.

For a collection of circuit topologies/architectures

- Willy M. C. Sansen, "Analog Design Essentials," Springer-Verlag, Berlin, Heidelberg, 2006.

Thank you!

nallam@iitk.ac.in

https://home.iitk.ac.in/∼nallam/