Session 10: Solid State Physics

MOSFET

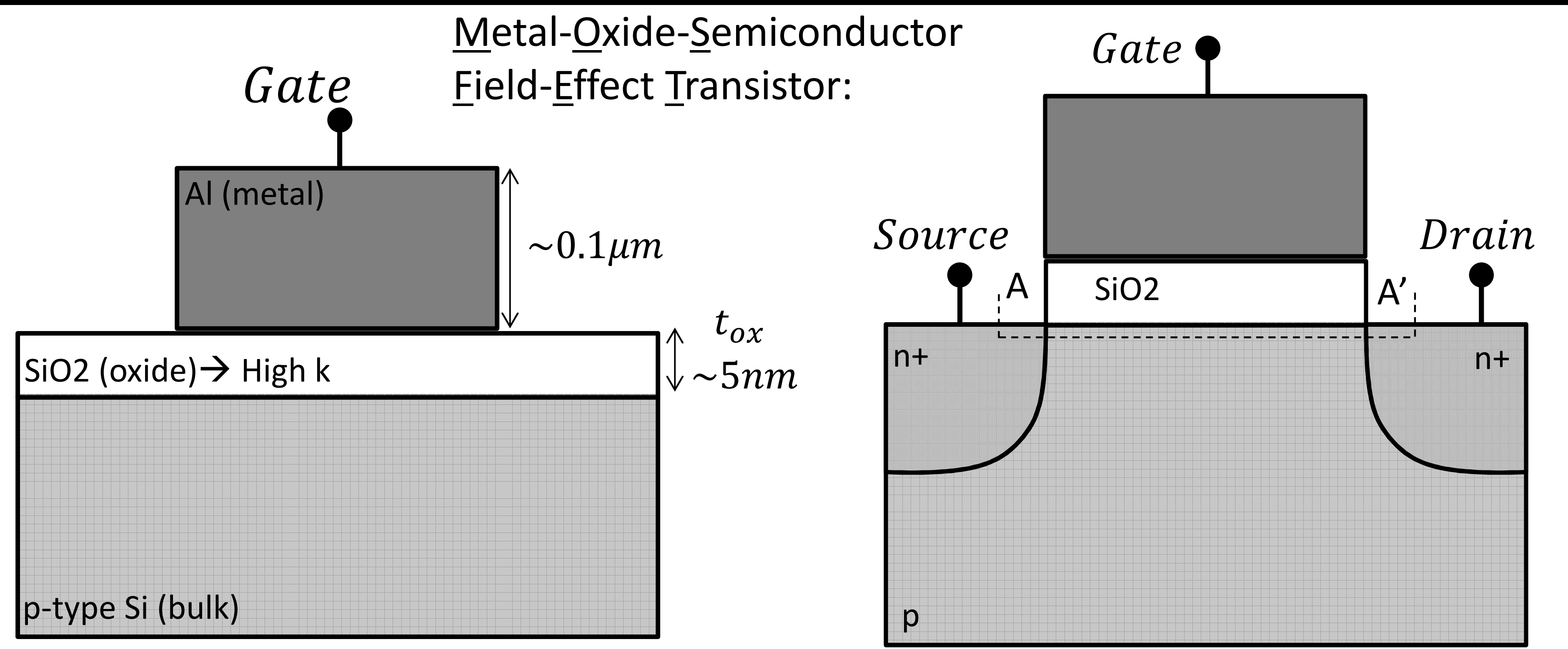
Outline

- A
  - B
  - C
  - D
  - E
- F
  - G
- H
- I
- J

MOSCap → MOSFET

1. I
2.
3.
4.
5.

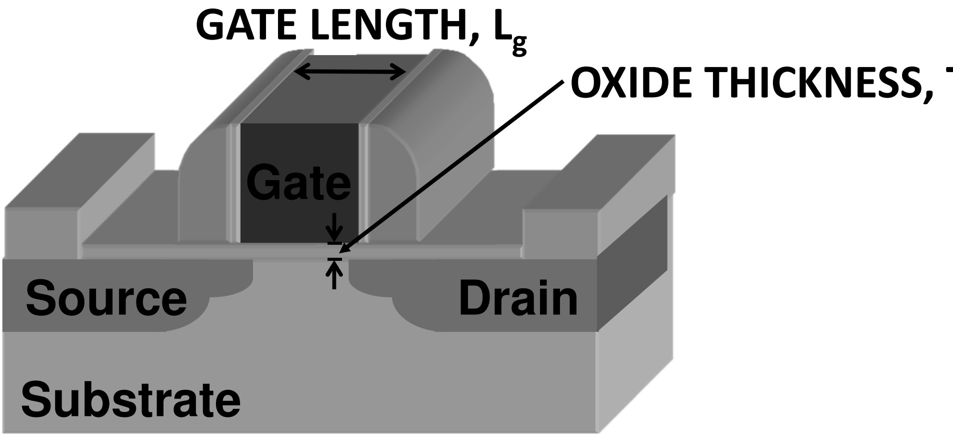
Metal-Oxide-Semiconductor
Field-Effect Transistor:



Drift current flowing between 2 doped regions ("source" & "drain")
is modulated by varying the voltage on the "gate" electrode.

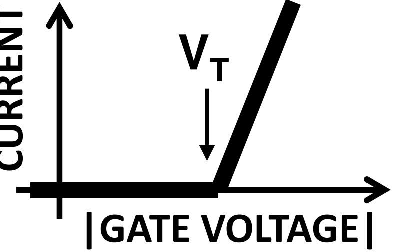
MOSFET

1. I
2.
3.
4.
5.

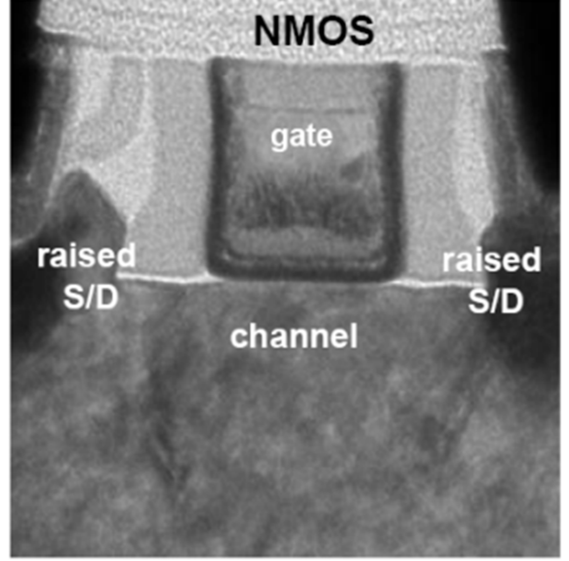


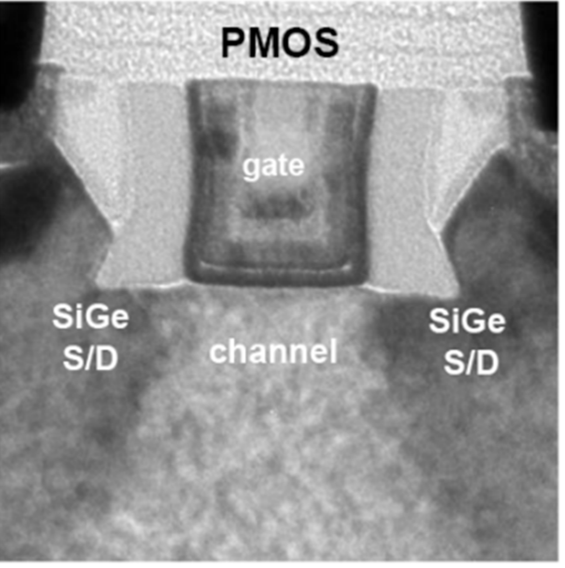
Desired characteristics:

- High ON current
- Low OFF current



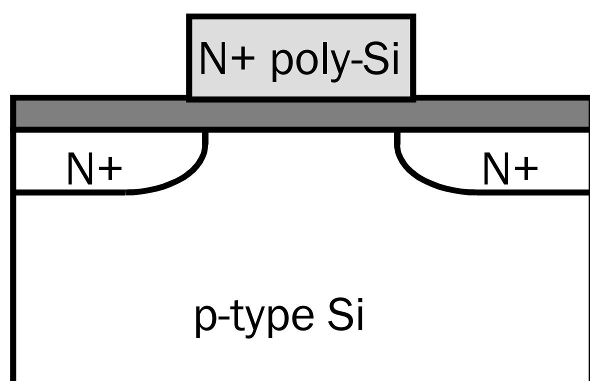
Intel's 32nm CMOSFETs





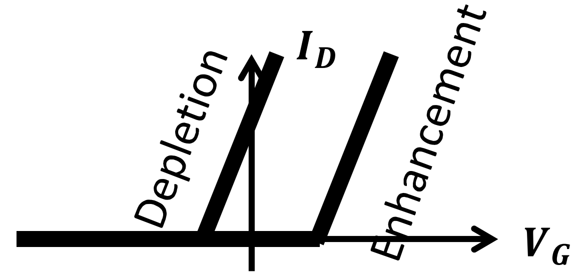
N-channel vs. P-channel

NMOS

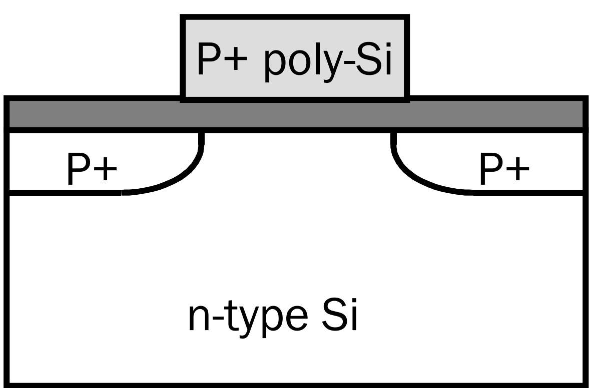


- For current to flow, $V_{GS} > V_T$
- Enhancement mode: $V_T > 0$
- Depletion mode: $V_T < 0$

Transistor is ON when V_G=0V

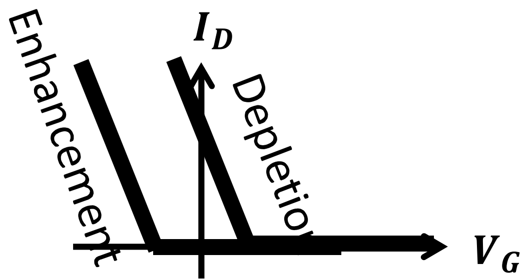


PMOS



- For current to flow, $V_{GS} < V_T$
- Enhancement mode: $V_T < 0$
- Depletion mode: $V_T > 0$

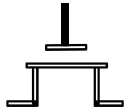
Transistor is ON when V_G=0V



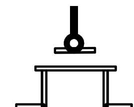
CMOS Devices and Circuits

CIRCUIT SYMBOLS

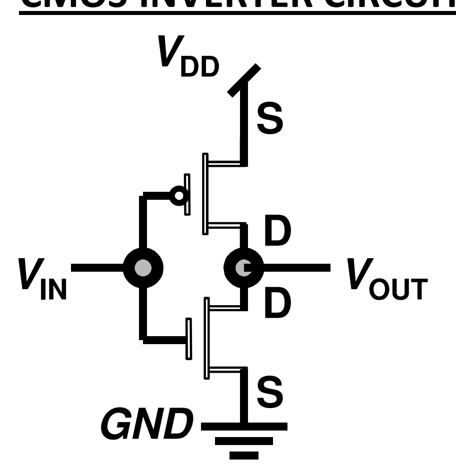
N-channel MOSFET

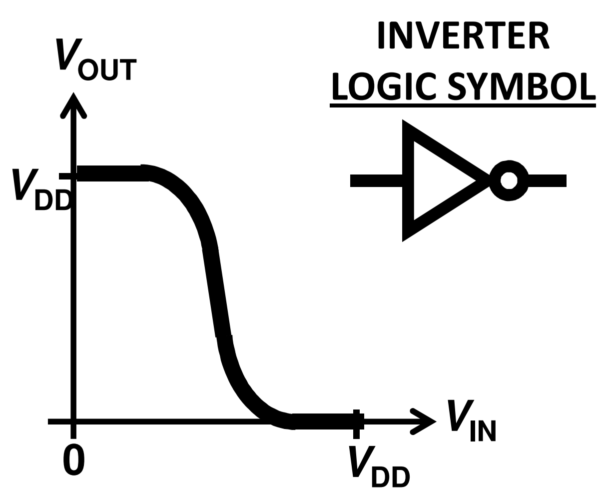


P-channel MOSFET



CMOS INVERTER CIRCUIT



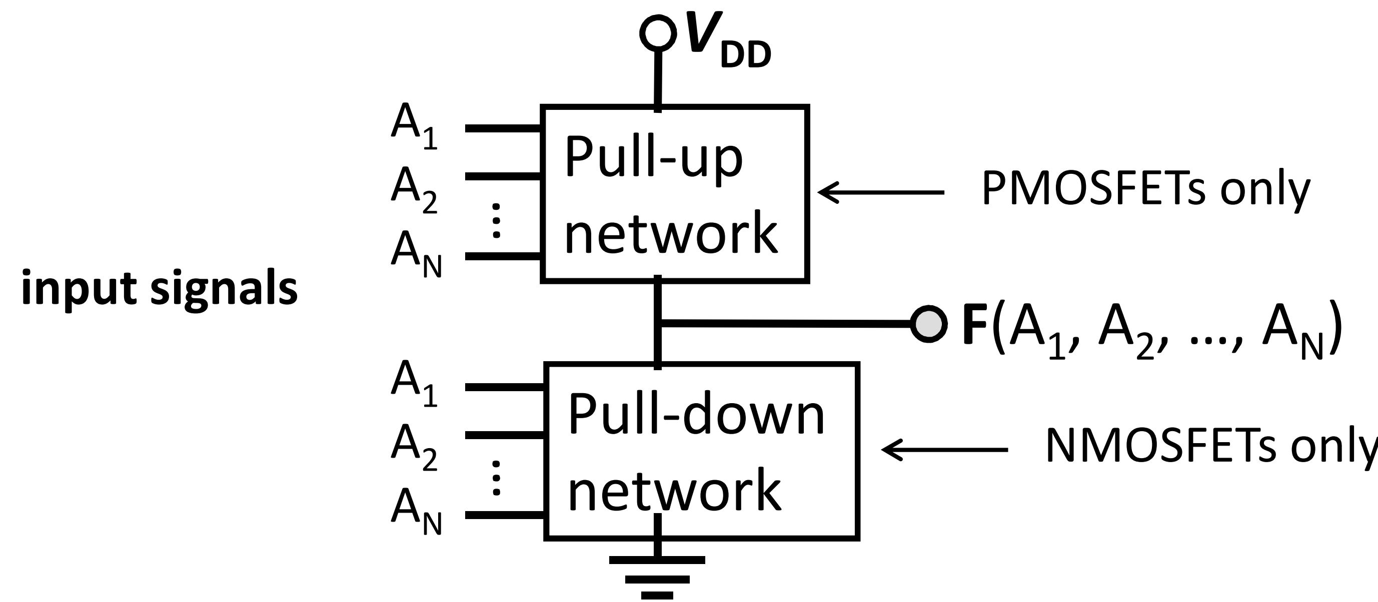


- When $V_G = V_{DD}$, the NMOSFET is on and the PMOSFET is off.

  When $V_G = 0$, the PMOSFET is on and the NMOSFET is off.

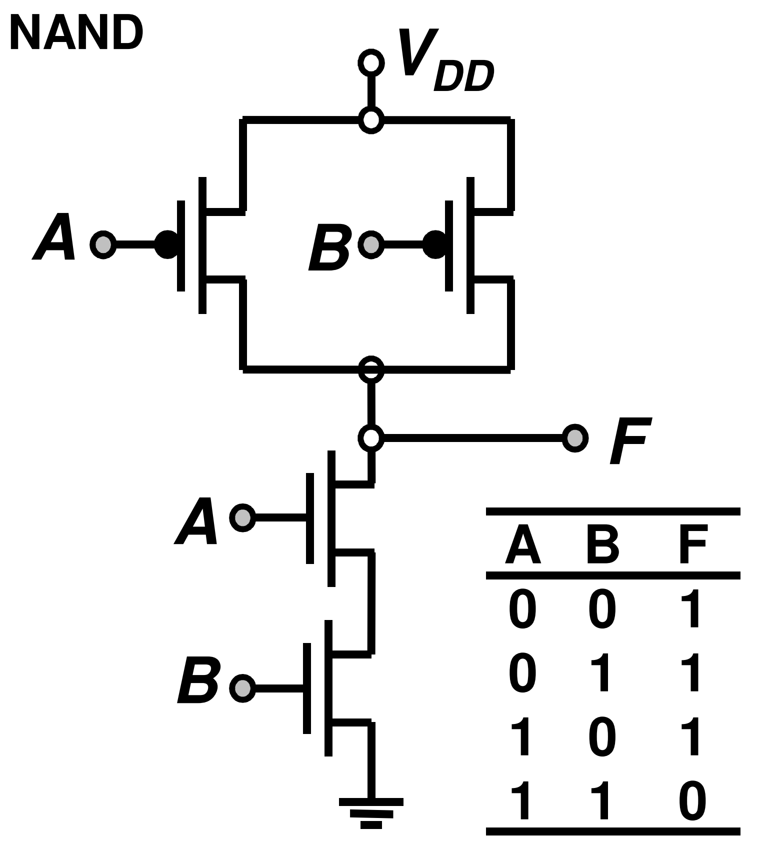
"Pull-Down" and "Pull-Up" Devices

- In CMOS logic gates, NMOSFETs are used to connect the output to GND, whereas PMOSFETs are used to connect the output to VDD.
  - An NMOSFET functions as a ***pull-down device*** when it is turned on (gate voltage = V_{DD})
  - A PMOSFET functions as a ***pull-up device*** when it is turned on (gate voltage = ***GND***)

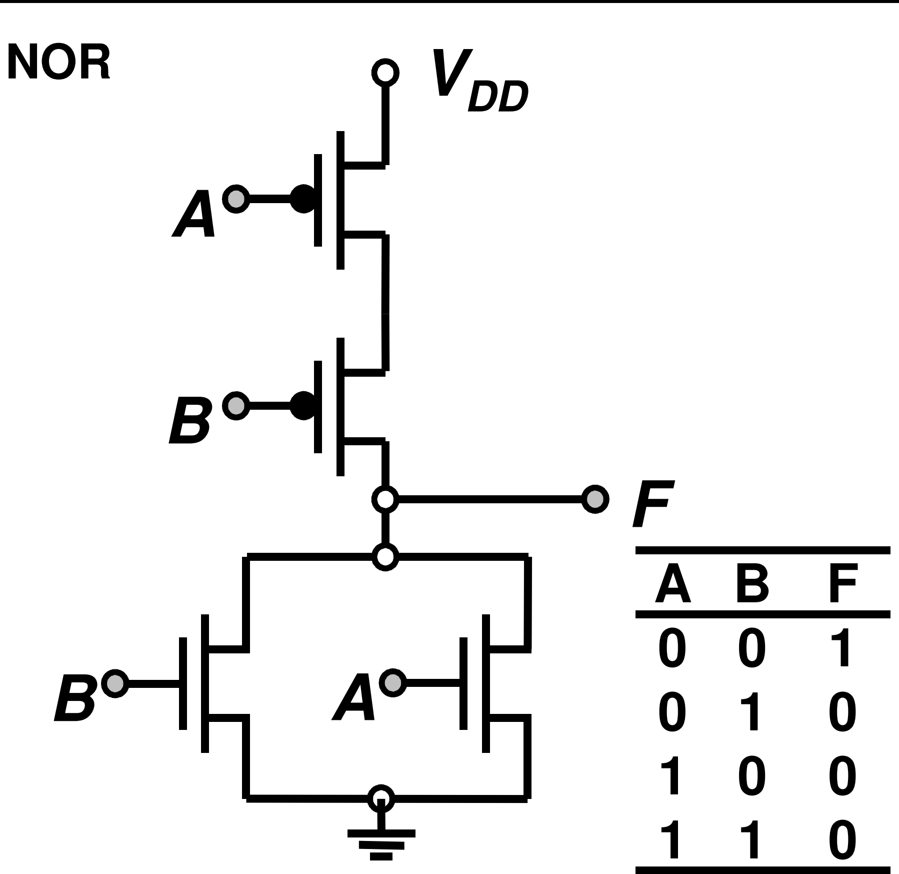


CMOS NAND / NOR Gate

1. I
2.
3.
4.
5.

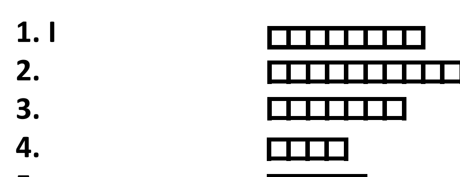


A	B	F
0	0	1
0	1	1
1	0	1
1	1	0



A	B	F
0	0	1
0	1	0
1	0	0
1	1	0

Qualitative Theory of the NMOSFET



Gate

Source *Drain*

A SiO2 A’

n+ n+

p

$V_G = 0$

$V_G > 0$

$V_G = V_T$

E_C

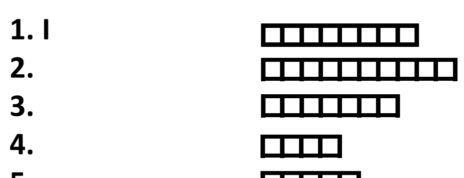
E_F

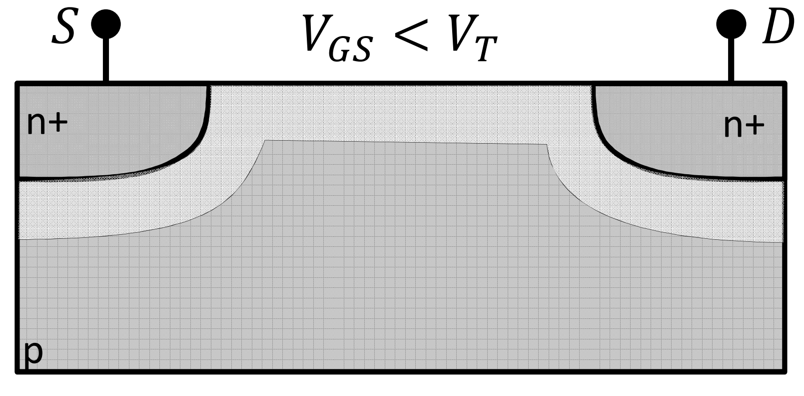
Source *Drain*

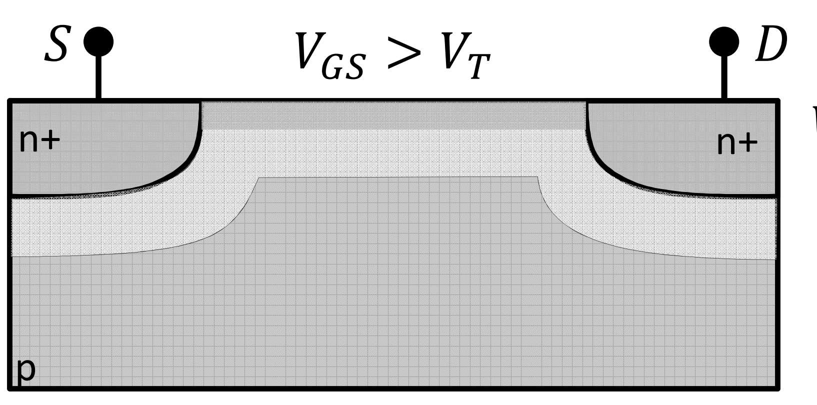
E_V

The potential barrier to electron flow from the source into the channel region is lowered by applying $V_{GS} > V_T$

Qualitative Theory of the NMOSFET

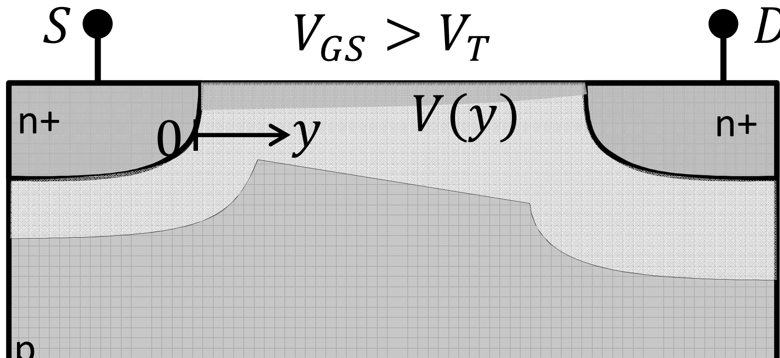




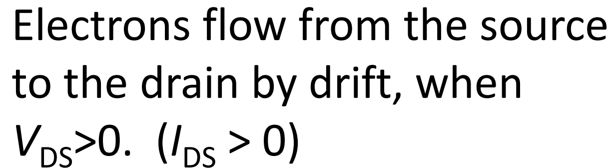


$V_{DS} \approx 0$

$V_{GS} > V_T$ → Inversion-layer "channel" is formed



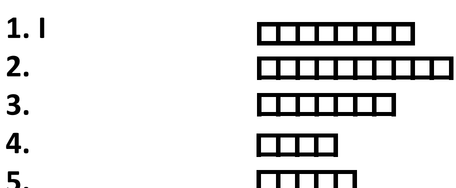
$V_{DS} > 0$



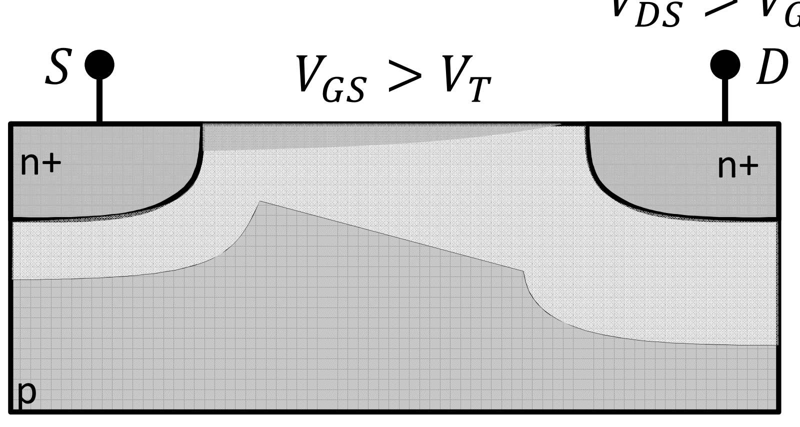
Electrons flow from the source to the drain by drift, when V_{DS}>0. ($I_{DS} > 0$)

The channel potential ($V_c(y)$) varies from V_S at the source end to V_D at the drain end.

Qualitative Theory of the NMOSFET



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



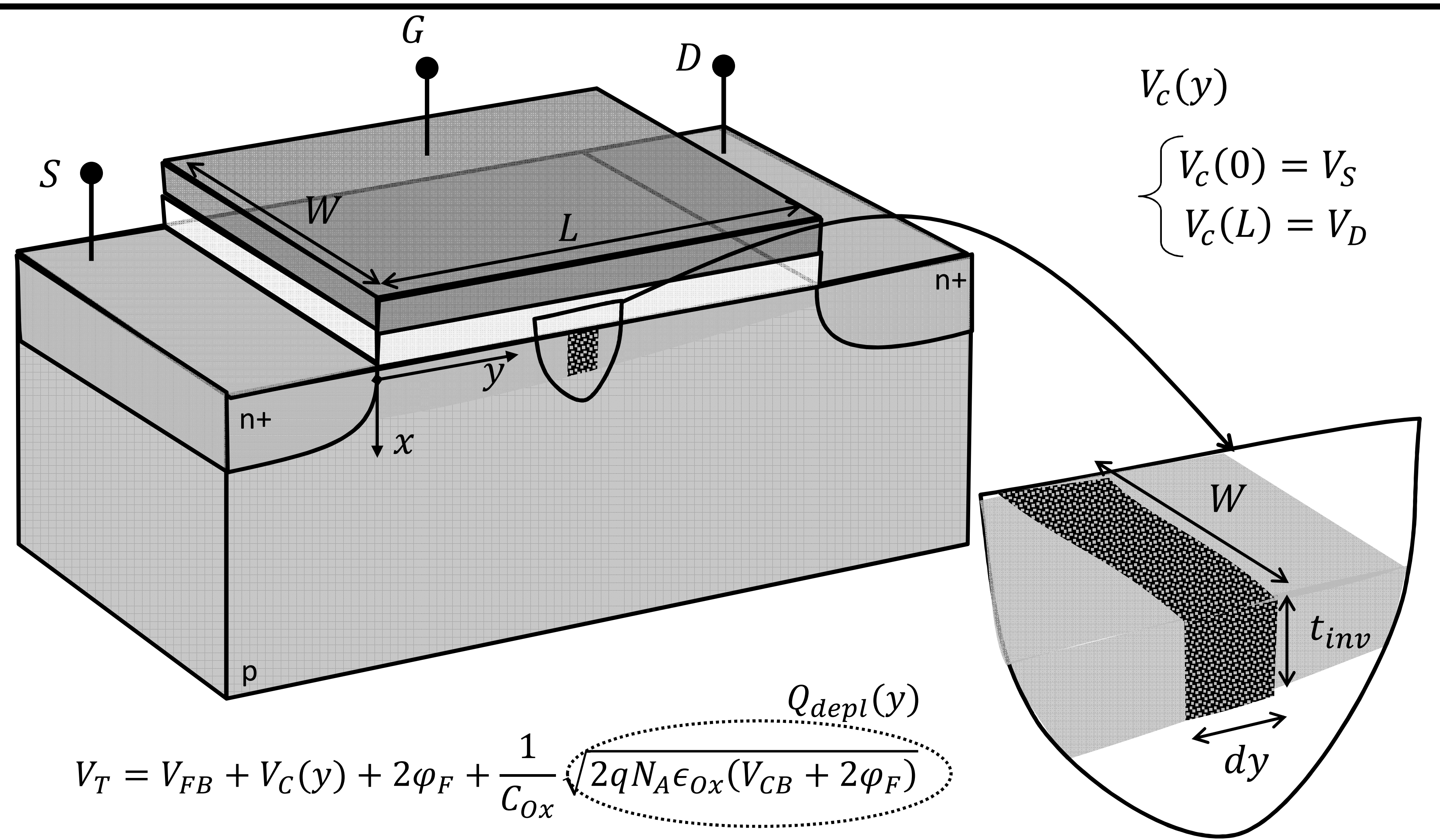
$V_{DS_{sat}} = V_{GS} - V_T$

$V_{GS} > V_T$ → Inversion-layer "channel" is formed

Electrons flow from the source to the drain by drift, when V_{DS}>0. ($I_{DS} > 0$)

The channel potential ($V_c(y)$) varies from V_S at the source end to V_D at the drain end.

MOSFET I-V Curve



$$V_T = V_{FB} + V_C(y) + 2\varphi_F + \frac{1}{C_{Ox}}\sqrt{2qN_A\epsilon_{Ox}(V_{CB} + 2\varphi_F)}$$

MOSFET I-V Curve

$$Q_{depl}(y)$$

$$V_T(y) = V_{FB} + V_C(y) + 2\varphi_F + \frac{1}{C_{Ox}}\sqrt{2qN_A\epsilon_{Ox}(V_{CB} + 2\varphi_F)}$$

$$V_c(y)$$

$$\begin{cases} V_c(0) = V_S \\ V_c(L) = V_D \end{cases}$$

$$Q_{inv} = -C_{Ox}(V_G - V_T(y))$$

$$Q_{inv} = -C_{Ox}\left(V_G - V_{FB} - V_C(y) - 2\varphi_F - \frac{Q_{depl}(y)}{C_{Ox}}\right)$$

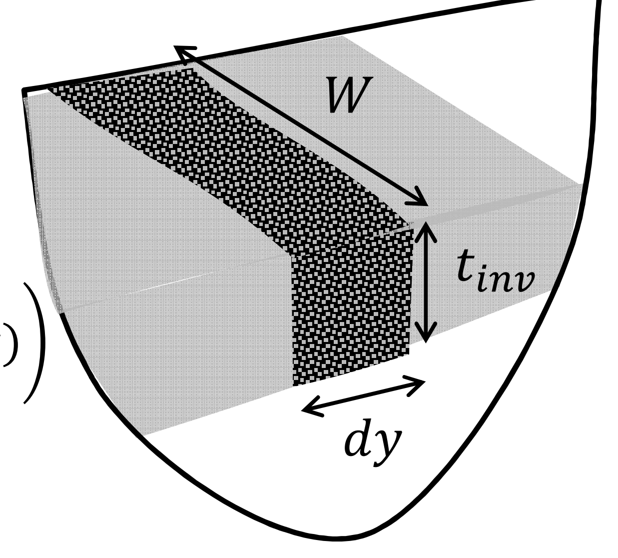
Depletion Region Approximation:

$Q_{inv}(y)$ but $Q_{depl}(y) \approx Q_{depl}(0)$

$$Q_{depl}(y) \approx \sqrt{2qN_A\epsilon_{Ox}(V_{SB} + 2\varphi_F)}$$

$$Q_{inv} = -C_{Ox}\left(V_G \underbrace{- V_{FB} - V_S - 2\varphi_F - \frac{Q_{depl}(0)}{C_{Ox}}}_{V_T(0)} + V_s - V_C(y)\right)$$

$$Q_{inv}(y) = -C_{Ox}(V_G - V_T(0) + V_s - V_C(y))$$



MOSFET I-V Curve

1. I
2.
3.
4.
5.

$$Q_{inv}(y) = -C_{Ox}(V_G - V_T(0) + V_s - V_C(y))$$

$V_c(y)$

$$\begin{cases} V_c(0) = V_S \\ V_c(L) = V_D \end{cases}$$

Simply call $V_T(0)$ as V_T $\quad Q_{inv}(y) = -C_{Ox}(V_G - V_T + V_s - V_C(y))$

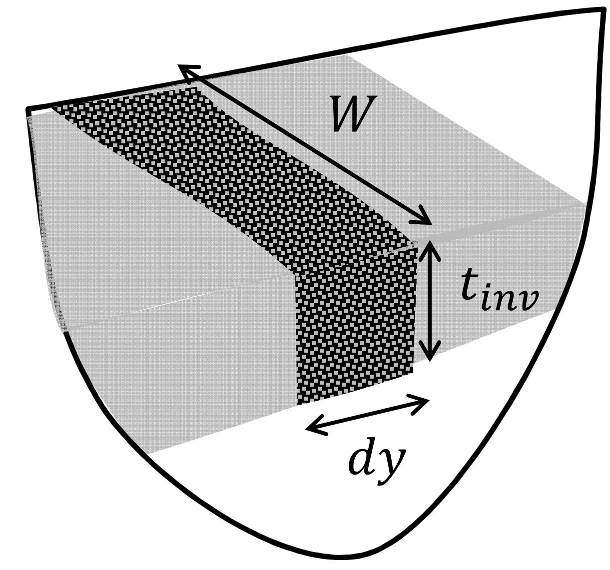
$$dV_c(y) = I_{DS} \cdot dR = I_{DS} \frac{dy}{\sigma W t_{inv}} = \frac{I_{DS} dy}{(q\mu_{eff} n) W t_{inv}} = \frac{I_{DS} dy}{\underbrace{(qnt_{inv})}_{-Q_{inv}(y)\ [C/cm^2]} \mu_{eff} W}$$

$$\int_0^L I_{DS} dy = \int_{V_S}^{V_D} -\mu_{eff} W Q_{inv}(y) dV_c$$

$$I_{DS} L = \mu_{eff} W \int_{V_S}^{V_D} [C_{Ox}(V_G - V_T + V_s - V_C(y))] dV_c$$

$$I_{DS} = \frac{W}{L} \mu_{eff} C_{Ox}\left(V_{GS} - V_T - \tfrac{1}{2} V_{DS}\right) V_{DS}$$

$$\left.\frac{\partial I_{DS}}{\partial V_{DS}}\right|_{V_{DS_{sat}}} = 0$$

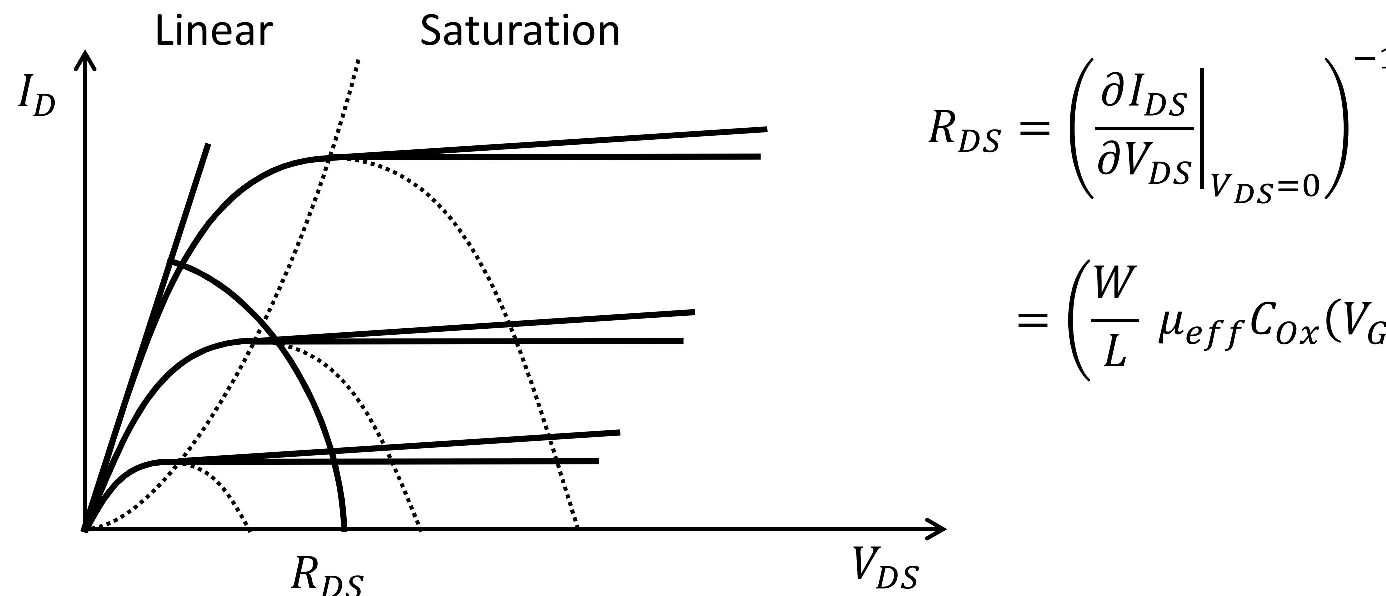


MOSFET I-V Curve

Linear

$$I_{DS} = \begin{cases} \frac{W}{L}\,\mu_{eff}C_{Ox}\left(V_{GS} - V_T - \frac{1}{2}V_{DS}\right)V_{DS} & V_{DS} < V_{DS_{sat}} \\ \frac{W}{2L}\,\mu_{eff}C_{Ox}(V_{GS} - V_T)^2 & V_{DS} < V_{DS_{sat}} = V_{GS} - V_T \end{cases}$$

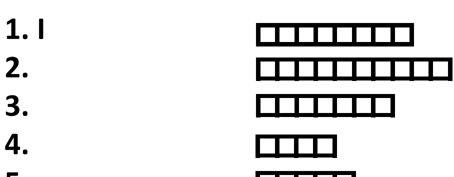
Saturation



$$R_{DS} = \left(\frac{\partial I_{DS}}{\partial V_{DS}}\bigg|_{V_{DS}=0}\right)^{-1}$$

$$= \left(\frac{W}{L}\,\mu_{eff}C_{Ox}(V_{GS} - V_T)\right)^{-1}$$

Field-Effect Mobility, μ_{eff}



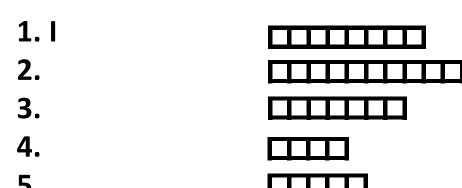
Effective vertical electric filed in the inversion layer for NMOS

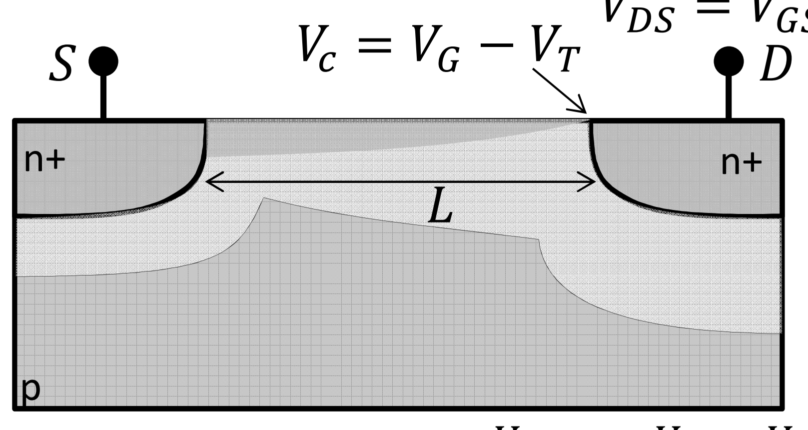
Scattering mechanisms:

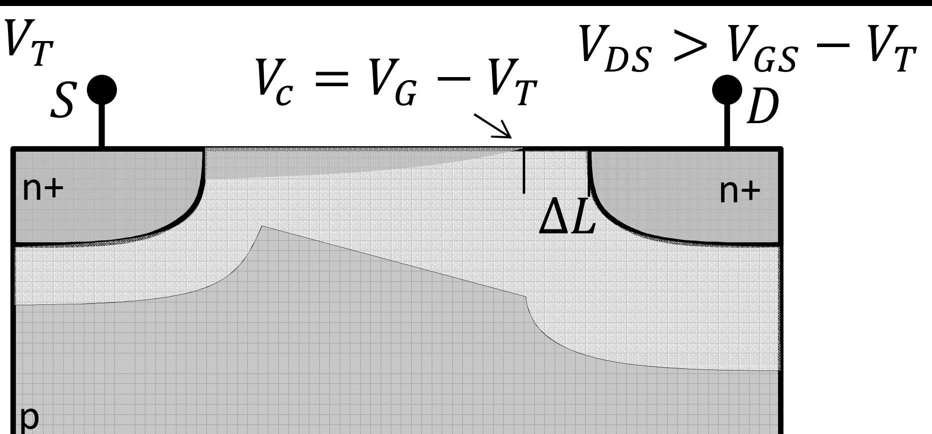
- Coulombic scattering
- phonon scattering
- surface roughness scattering

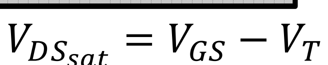
Effective vertical electric filed in the inversion layer for PMOS

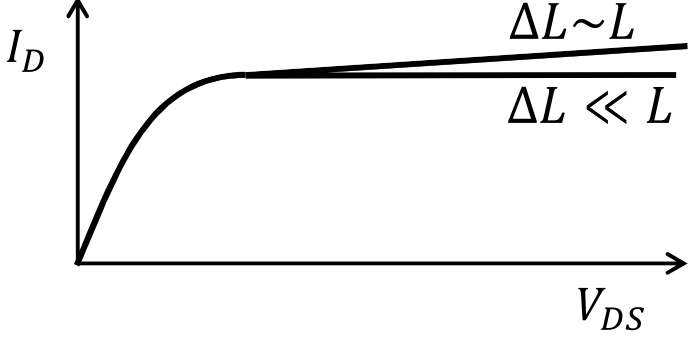
MOSFET Saturation Region of Operation











As V_D is increased above V_G-V_T, the length ΔL of the “pinch-off” region increases. The voltage applied across the inversion layer is always V_{Dsat}=V_{GS}-V_T, and so the current saturates.

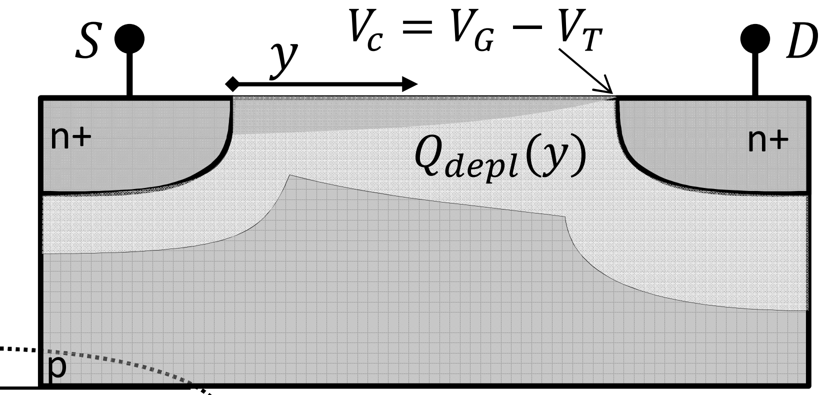
If ΔL is significant compared to L, then I_{DS} will increase slightly with increasing V_{DS}>V_{Dsat}, due to “channel-length modulation”

“Square Law Theory” ?

1. I
2.
3.
4.
5.

$$I_{DS} = \begin{cases} \frac{W}{L}\,\mu_{eff}C_{Ox}\left(V_{GS} - V_T - \frac{1}{2}V_{DS}\right)V_{DS} & V_{DS} < V_{DS_{sat}} \quad \text{Linear} \\ \frac{W}{2L}\,\mu_{eff}C_{Ox}(V_{GS} - V_T)^2 & V_{DS} < V_{DS_{sat}} = V_{GS} - V_T \quad \text{Saturation} \end{cases}$$

Depletion Region Approximation:



$Q_{inv}(y)$ but $Q_{depl}(y) \approx Q_{depl}(0)$

$$Q_{depl}(y) \approx \sqrt{2qN_A\epsilon_{Ox}(V_{SB} + 2\varphi_F)}$$

$$V_T(y) = V_{FB} + V_C(y) + 2\varphi_F + \frac{1}{C_{Ox}}\underbrace{\sqrt{2qN_A\epsilon_{Ox}(V_{CB} + 2\varphi_F)}}_{Q_{depl}(y)}$$

$$Q_{inv}(y) = -C_{Ox}(V_G - V_T(0) + V_s - V_C(y))$$

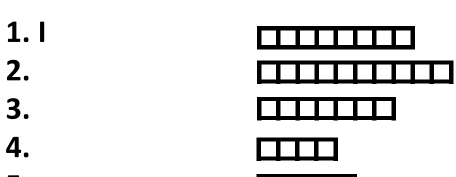
$Q_{depl}(y) > Q_{depl}(0)$

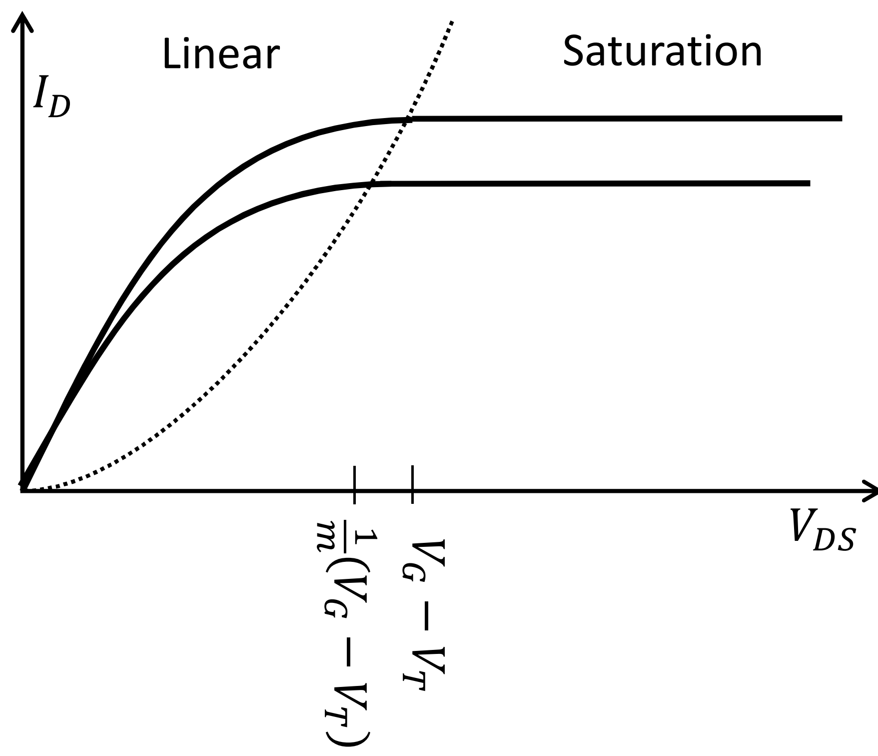
$V_T(y) > V_T(0)$

$Q'_{inv}(y) < Q_{inv}(y)$

$I'_{DS} < I_{DS}$

Modified (Bulk-Charge) I-V Model





Linear $V_{DS} < V_G - V_T$

$$I_{DS} = \frac{W}{L}\,\mu_{eff}C_{Ox}\left(V_{GS} - V_T - \tfrac{1}{2}V_{DS}\right)V_{DS}$$

Linear $V_{DS} < \frac{1}{m}(V_G - V_T)$

$$I_{DS} = \frac{W}{L}\,\mu_{eff}C_{Ox}\left(V_{GS} - V_T - \tfrac{m}{2}V_{DS}\right)V_{DS}$$

Saturation $V_{DS} > V_G - V_T$

$$I_{DS} = \frac{W}{2L}\,\mu_{eff}C_{Ox}(V_{GS} - V_T)^2$$

Saturation $V_{DS} > \frac{1}{m}(V_G - V_T)$

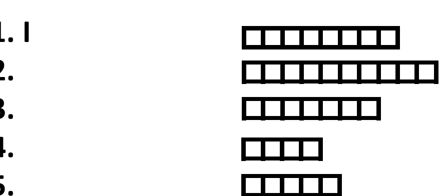
$$I_{DS} = \frac{W}{2mL}\,\mu_{eff}C_{Ox}(V_{GS} - V_T)^2$$

Bulk charge factor

$$m = 1 + \frac{C_{dep}}{C_{Ox}} = 1 + \frac{3t_{Ox}}{W_T}$$

Typically $1.1 < m < 1.4$

The Body Effect



Note that V_T is a function of V_{SB}:

$$V_T = V_{FB} + 2\varphi_F + \frac{1}{C_{Ox}}\sqrt{2qN_A\epsilon_{Ox}(V_{SB} + 2\varphi_F)}$$

$$V_T = V_{T0} + \frac{1}{C_{Ox}}\sqrt{2qN_A\epsilon_{Ox}}\left(\sqrt{(V_{SB} + 2\varphi_F)} - \sqrt{2\varphi_F}\right)$$

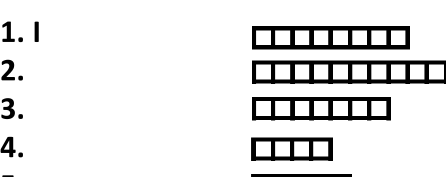
$$= V_{T0} + \gamma\left(\sqrt{(V_{SB} + 2\varphi_F)} - \sqrt{2\varphi_F}\right)$$

where γ is the *body effect parameter*

$$\gamma = \frac{1}{C_{Ox}}\sqrt{2qN_A\epsilon_{Ox}}$$

When the source-body pn junction is reverse-biased, $|V_T|$ is increased. Usually, we want to minimize g so that I_{Dsat} will be the same for all transistors in a circuit.

The Body Effect



NAND

V_{DD}, A, B, F, (a) A, (b) B, (1)

	A	B	F
	0	0	1
	0	1	1
(*)	1	0	1
	1	1	0

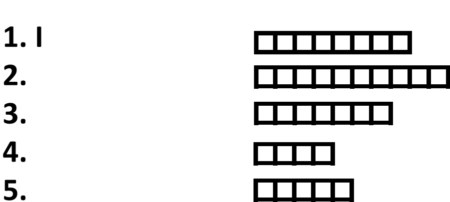
$V_A = V_{DD}$

$V_B = 0$

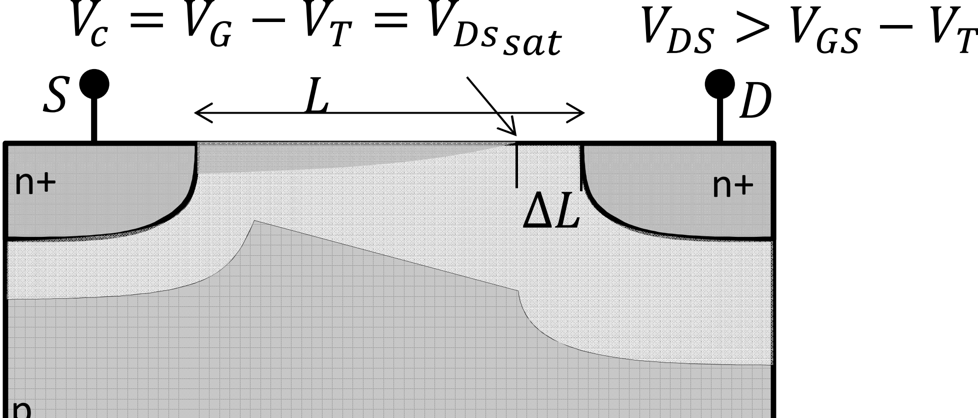
$V_{(1)} = V_{DD}$

$V_{T_a} > V_{T_b}$

λ: Channel Length Modulation Parameter



$$I_{D_{sat}} \propto \frac{1}{L-\Delta L} = \frac{1}{L}\left(1+\frac{\Delta L}{L}\right)$$



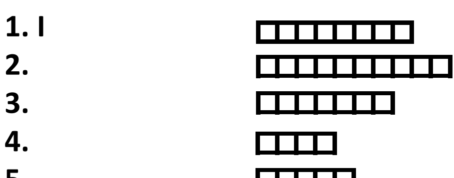
$$\Delta L \propto V_{DS} - V_{DS_{sat}}$$

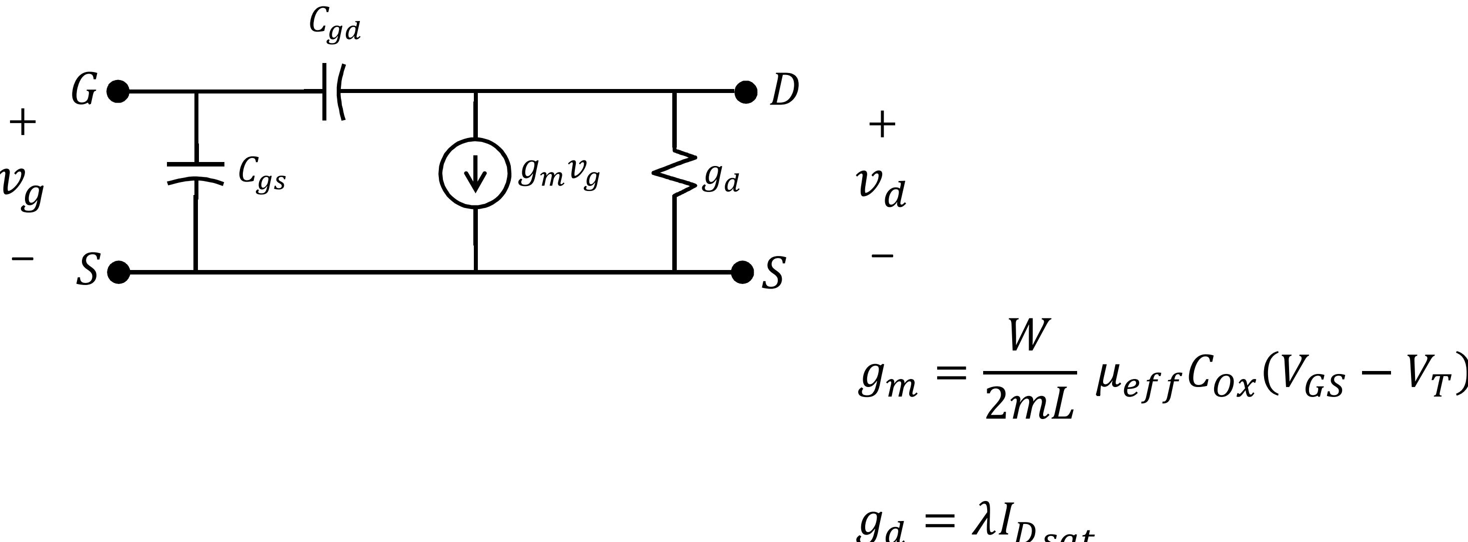
$$\frac{\Delta L}{L} \propto \lambda\left(V_{DS} - V_{DS_{sat}}\right)$$

$$\lambda \sim \frac{1}{L}$$

$$I_{DS} = \frac{W}{2mL}\,\mu_{eff} C_{Ox} (V_{GS} - V_T)^2 \left(1 + \lambda\left(V_{DS} - V_{DS_{sat}}\right)\right)$$

MOSFET: Small Signal Model

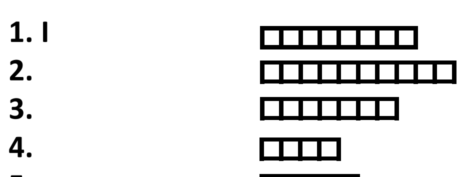


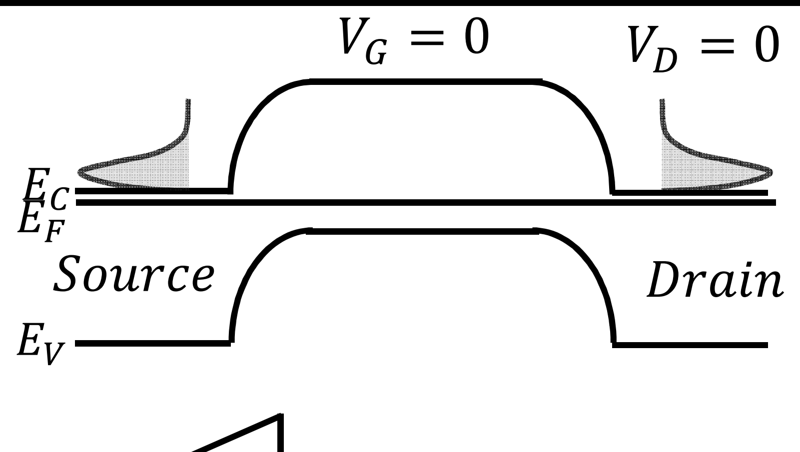


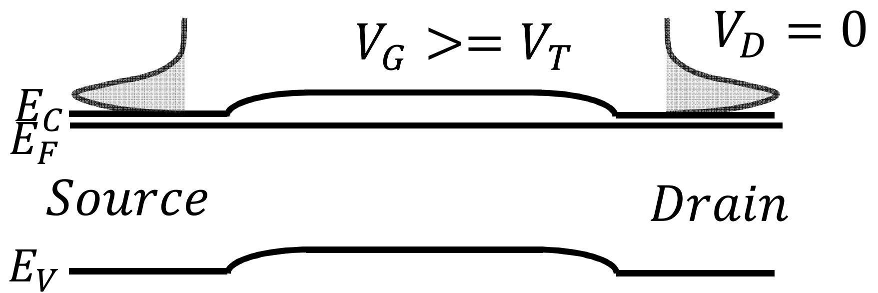
cut-off frequency:

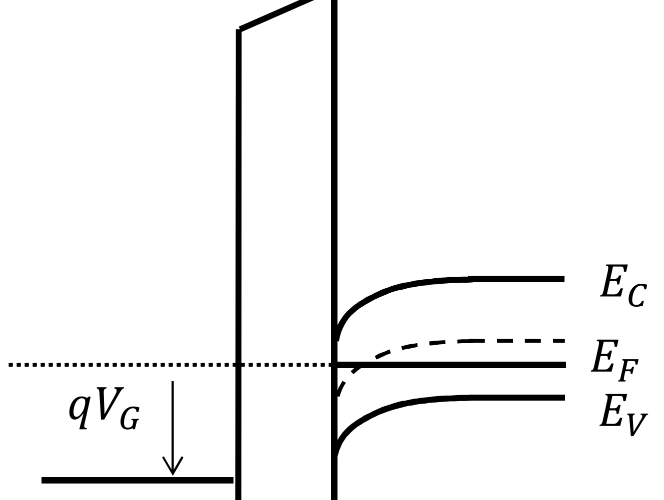
$$f_{max} \nearrow \quad \longrightarrow \quad \frac{g_m}{2\pi C_{Ox}} \quad \propto \quad \frac{1}{L} \searrow$$

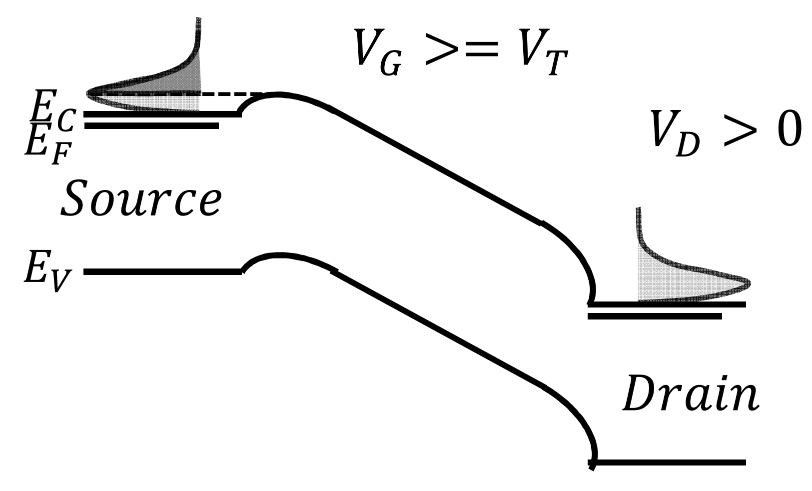
Sub-Threshold Current



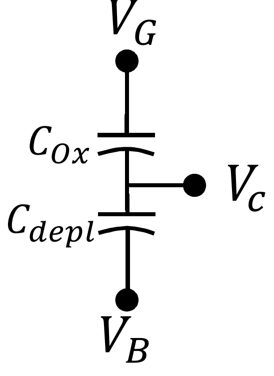








Sub-Threshold Current



Similarly:

$$m = 1 + \frac{C_{depl}}{C_{Ox}} = 1 + \frac{3t_{Ox}}{W_T}$$

$$V_c = \frac{C_{Ox}}{C_{Ox} + C_{depl}} V_{GB}$$

$V_G < V_T$

$V_D > 0$

E_C

E_F

$Source$

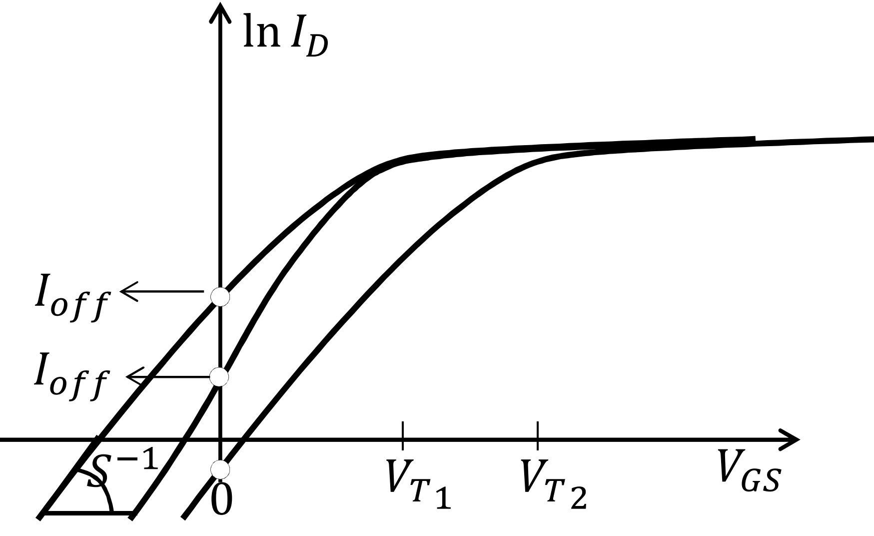
E_V

$Drain$

$$I_{DS} = \frac{W}{L}\,\mu_{eff} C_{Ox}(m-1)\left(\frac{kT}{q}\right)^2 e^{q(V_G - V_T)/mkT}\left(1 - e^{-qV_{DS}/kT}\right)$$

$$S = \left(\frac{d\log_{10} I_{DS}}{dV_{GS}}\right)^{-1} = \underbrace{\frac{kT}{q}\ln 10}_{60\ mV}\left(1 + \frac{C_{depl}}{C_{Ox}}\right)$$

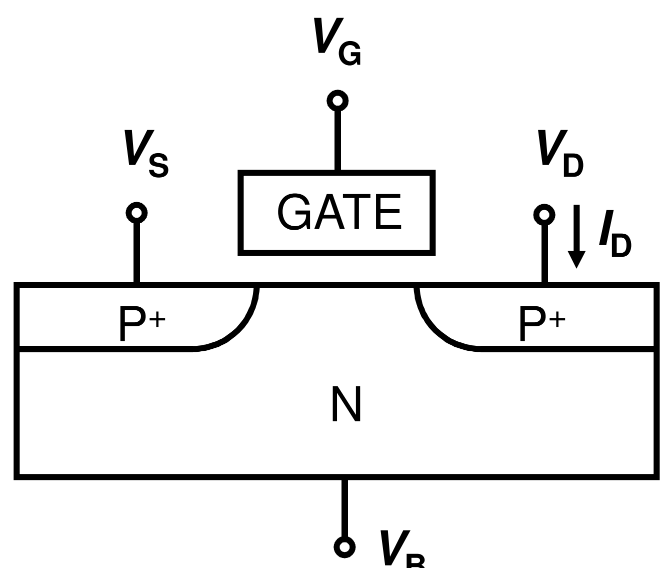
$$m \searrow \begin{cases} N_A \searrow \implies C_{depl} \searrow \implies \text{retrograde} \\ t_{Ox} \searrow \implies C_{Ox} \nearrow \\ T \searrow \quad \text{(low-temperature)} \end{cases}$$



P-Channel MOSFET

- The PMOSFET turns on when $V_{GS} < V_T$
  - Holes flow from SOURCE to DRAIN

    $\Rightarrow$ DRAIN is biased at a ***lower*** potential than the SOURCE

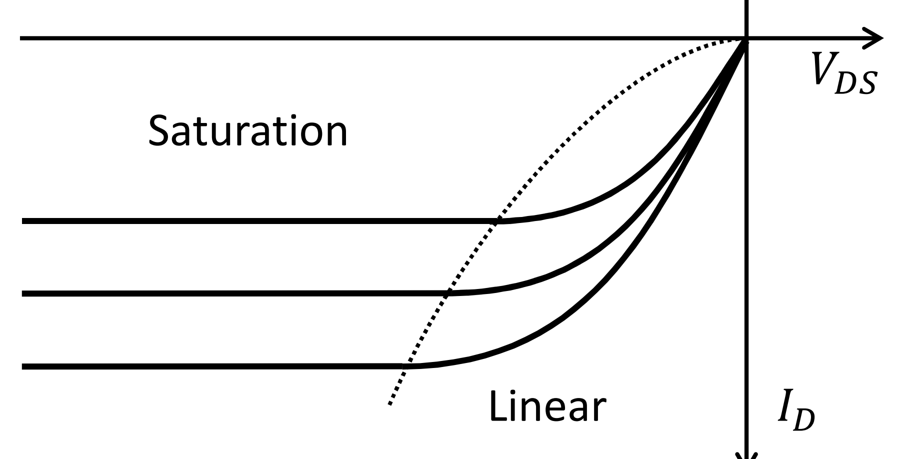


- $V_{DS} < 0$
- $I_{DS} < 0$
- $|I_{DS}|$ increases with
  - $|V_{GS} - V_T|$
  - $|V_{DS}|$ (linear region)

- In a CMOS technology, the PMOS & NMOS threshold voltages are usually symmetric about 0, *i.e.* $V_{Tp} = -V_{Tn}$

PMOSFET I-V

1. I
2.
3.
4.
5.



Linear $\quad 0 < |V_{DS}| < \dfrac{|V_G - V_T|}{m}$

$$I_{DS} = -\frac{W}{L}\ \mu_{eff} C_{Ox}\left(V_{GS} - V_{Tp} - \tfrac{m}{2}V_{DS}\right)V_{DS}$$

bulk-charge factor

$$m = 1 + \frac{C_{depl}}{C_{Ox}} = 1 + \frac{3t_{Ox}}{W_T}$$

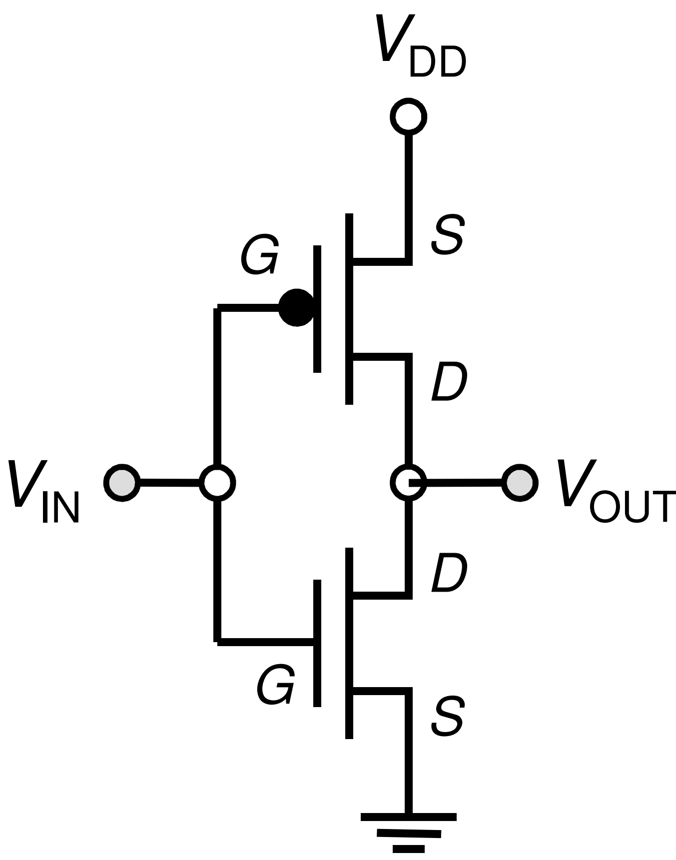
Saturation $\quad |V_{DS}| > \dfrac{|V_G - V_T|}{m}$

$$I_{DS} = I_{D_{sat}} = -\frac{W}{2mL}\ \mu_{eff} C_{Ox}(V_{GS} - V_T)^2$$

CMOS Inverter: Intuitive Perspective

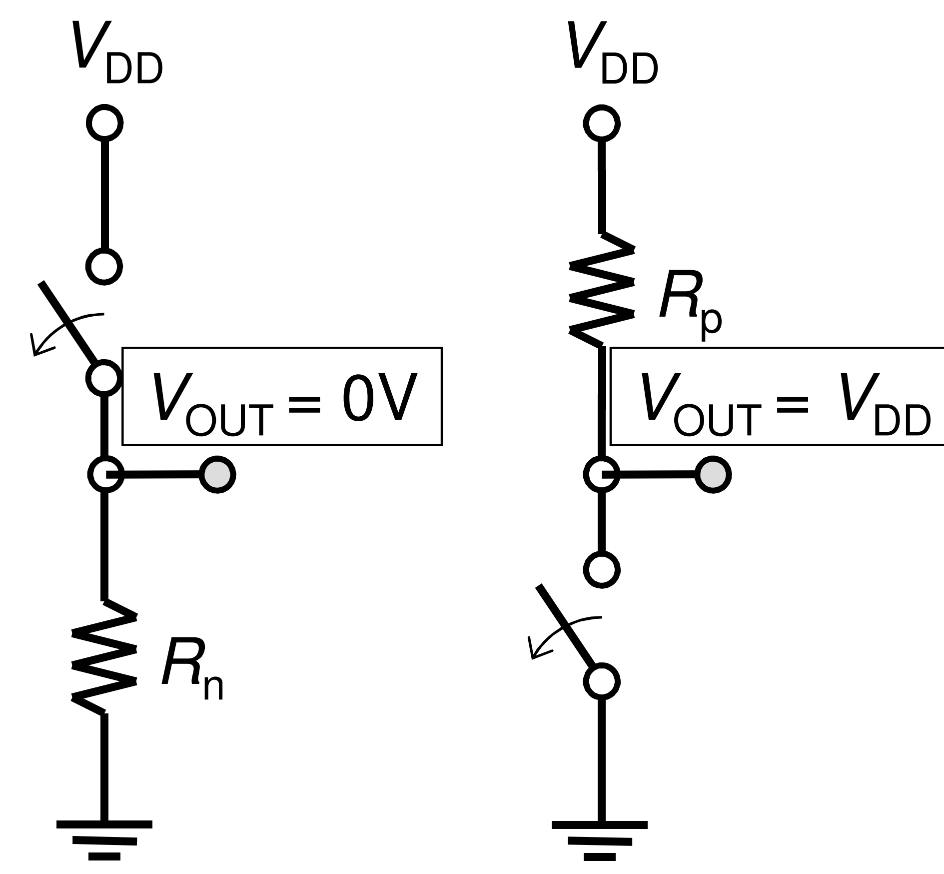
1. I
2.
3.
4.
5.

CIRCUIT



Low static power consumption, since one MOSFET is always off in steady state

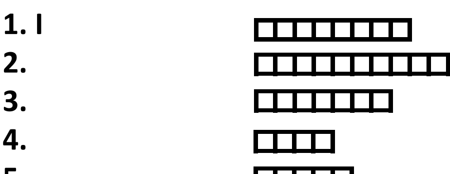
SWITCH MODELS



$V_{IN} = V_{DD}$

$V_{IN} = 0V$

Voltage Transfer Characteristic



V_{OUT}

N: sat
P: sat

N: off
P: lin

C

V_{DD}

N: sat
P: lin

A

B

D

E

N: lin
P: sat

N: lin
P: off

0

0

V_{DD}

V_{IN}

V_{DD}

S

G

D

V_{IN}

V_{OUT}

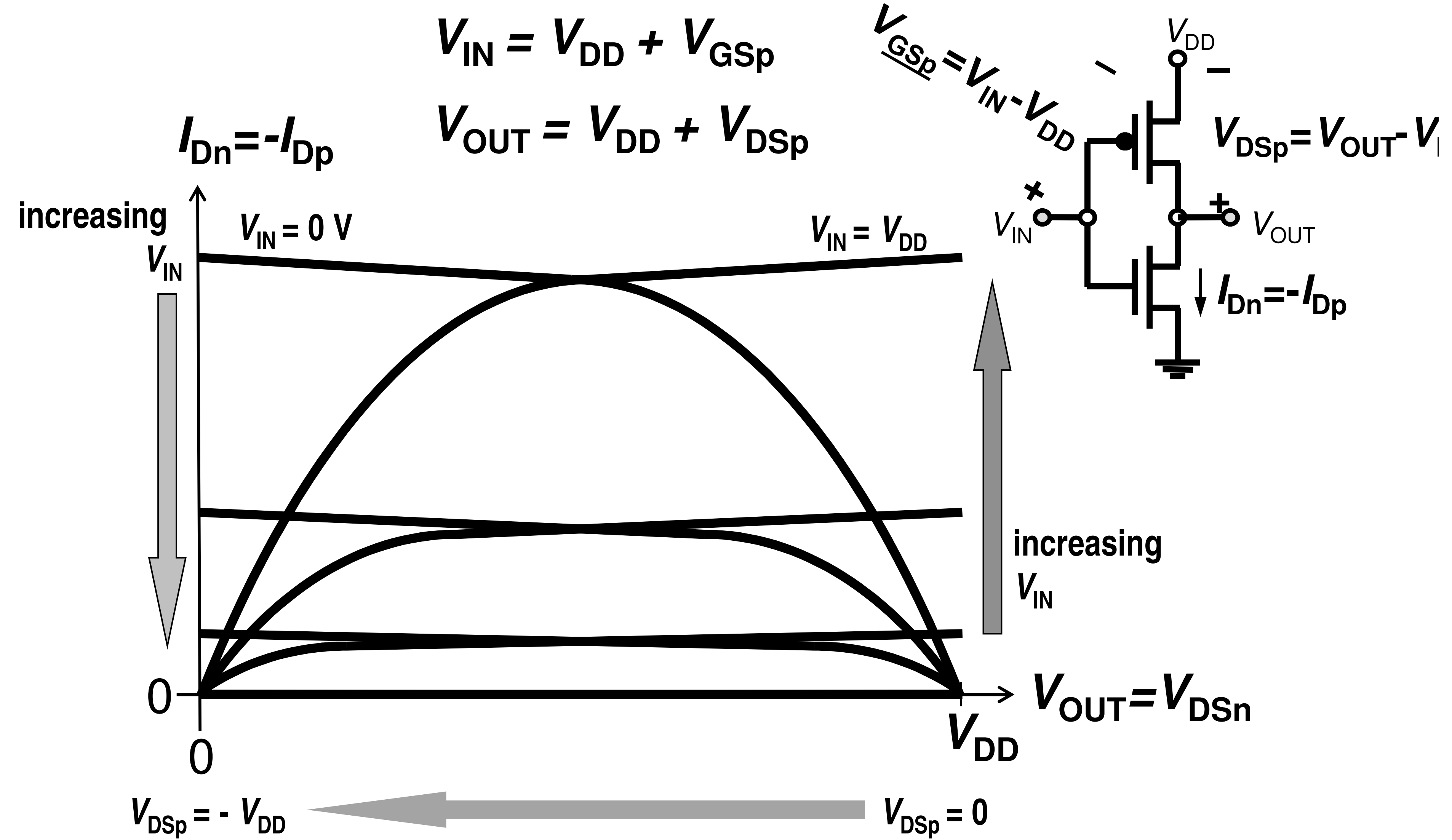
D

G

S

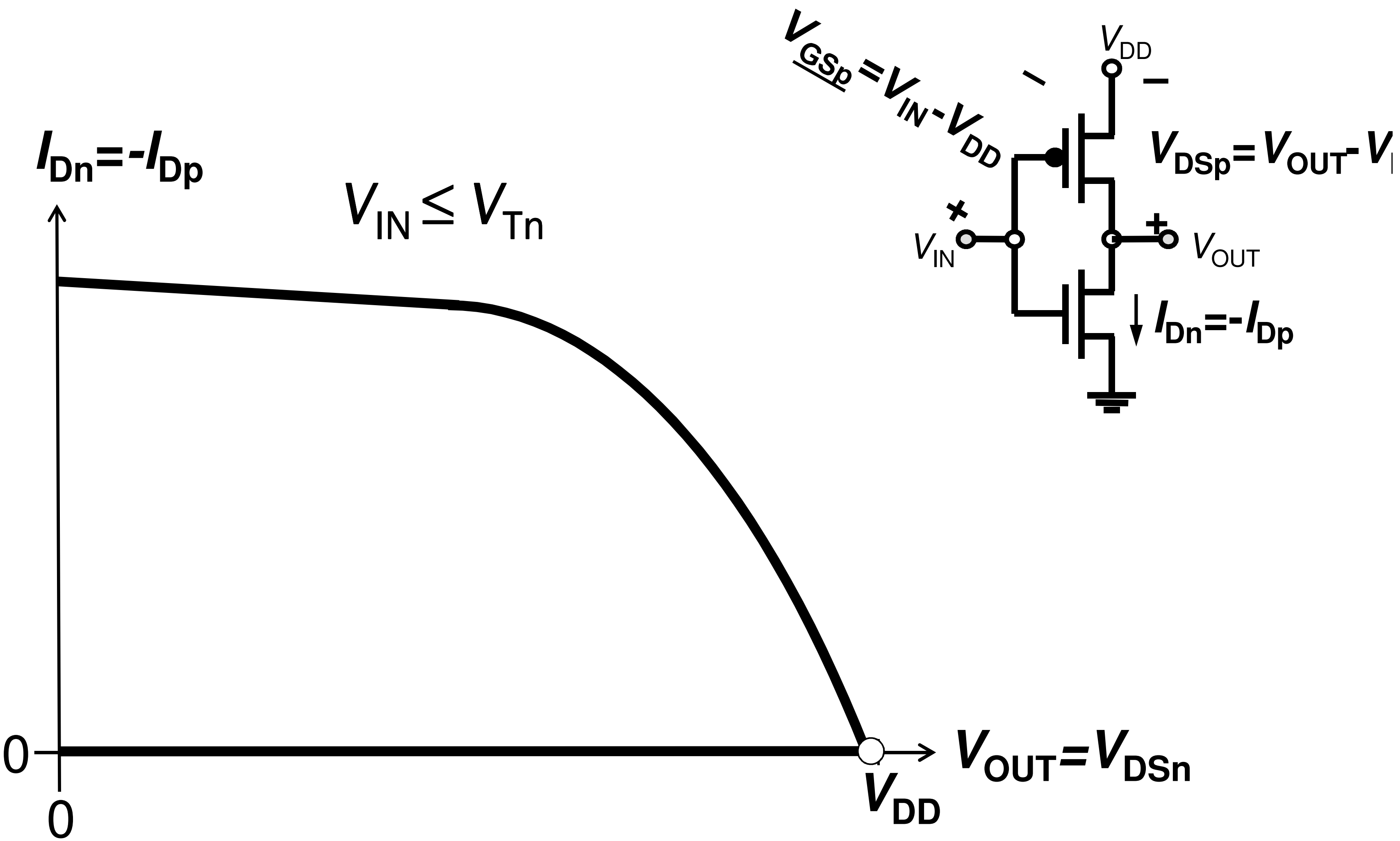
CMOS Inverter Load-Line Analysis

1. I
2.
3.
4.
5.

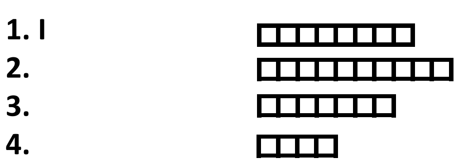


Load-Line Analysis: Region A

1. I
2.
3.
4.
5.



Load-Line Analysis: Region B



$V_{GSp}=V_{IN}-V_{DD}$

V_{DD}

$V_{DSp}=V_{OUT}-V_{DD}$

V_{IN}

V_{OUT}

$I_{Dn}=-I_{Dp}$

$I_{Dn}=-I_{Dp}$

$V_{DD}/2 > V_{IN} > V_{Tn}$

0

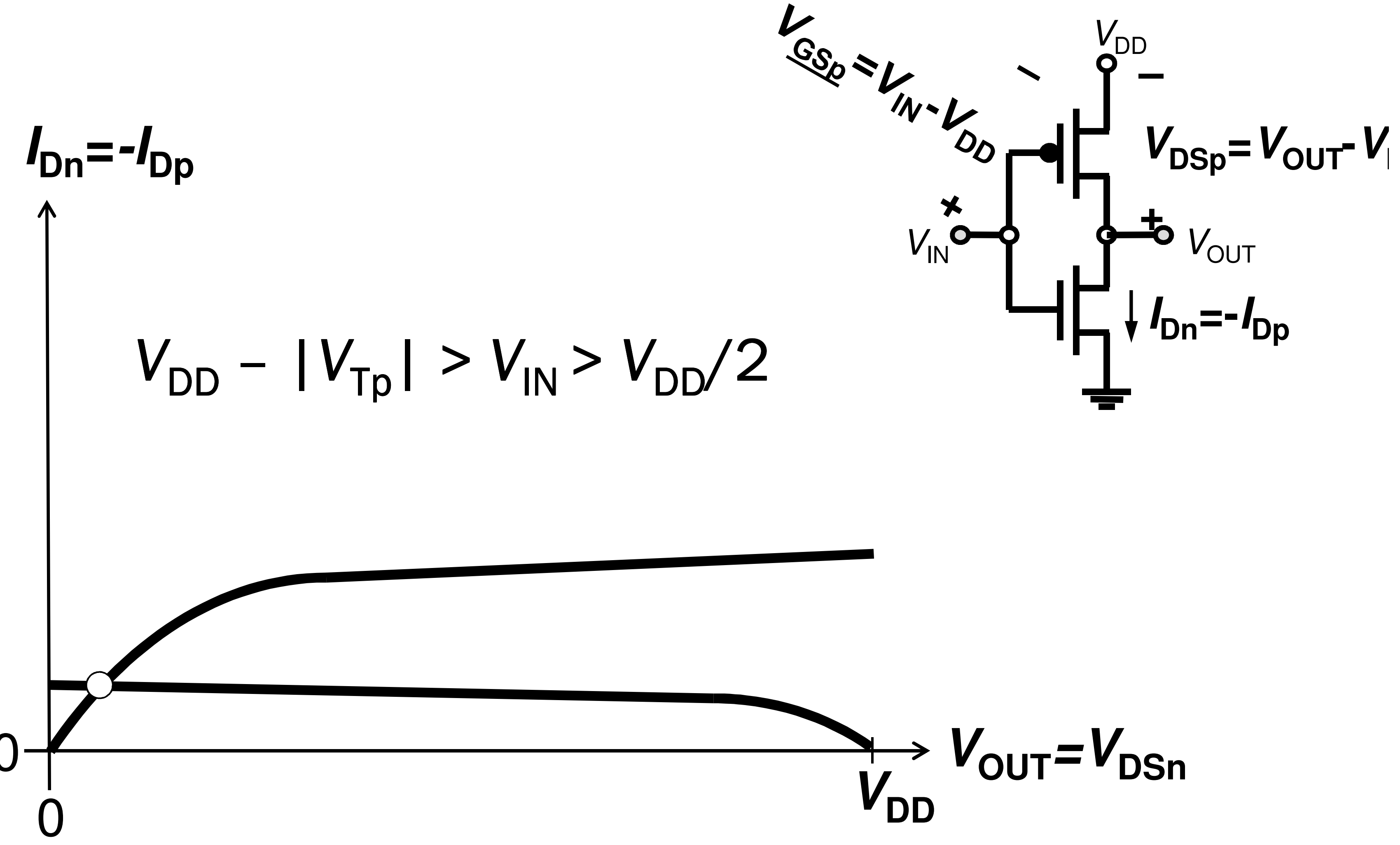
0

V_{DD}

$V_{OUT}=V_{DSn}$

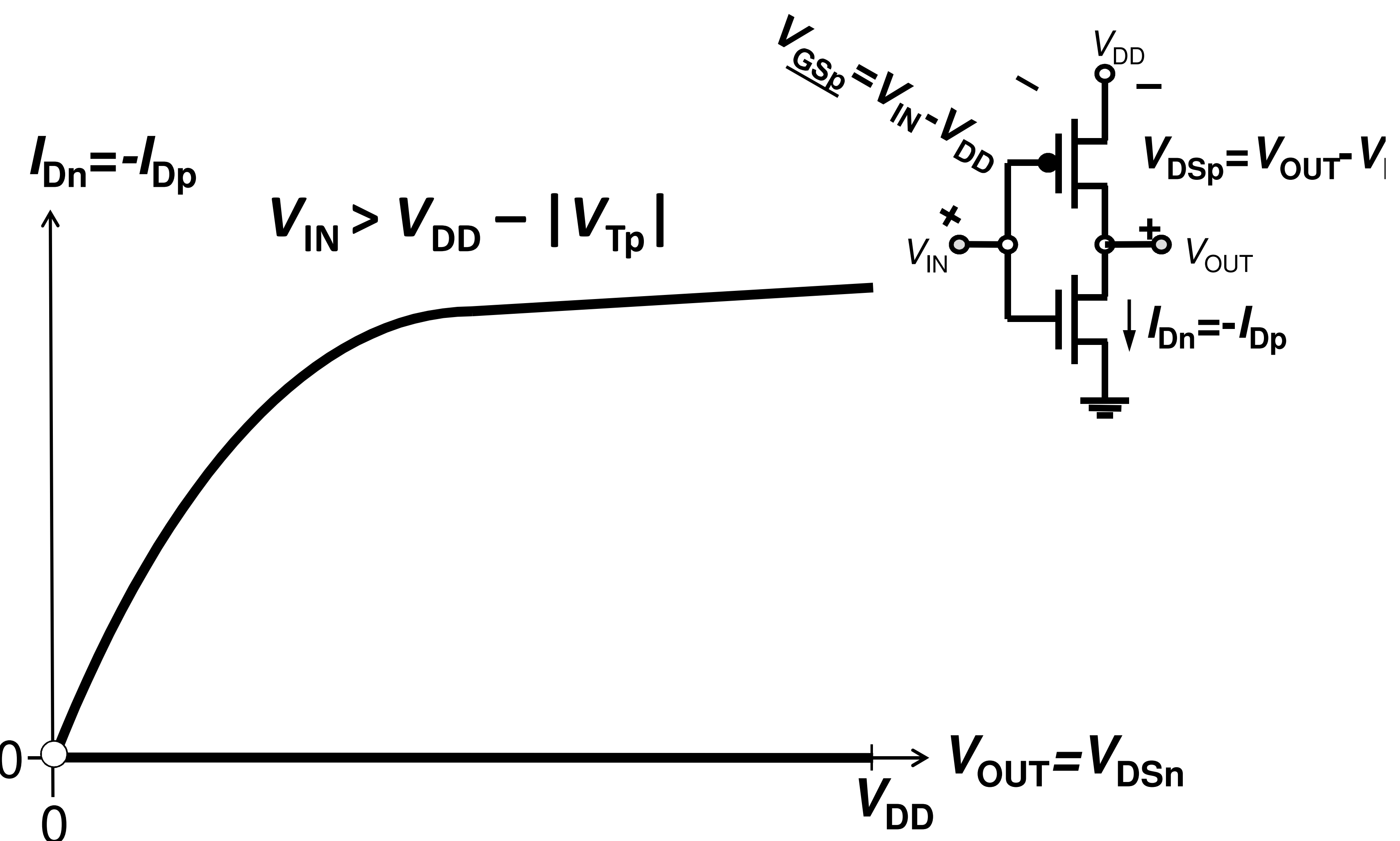
Load-Line Analysis: Region D

1. I
2.
3.
4.
5.



Load-Line Analysis: Region E

1.
2.
3.
4.
5.



MOSFET Scaling

1. I
2.
3.
4.
5.

MOSFETs have been steadily miniaturized over time

1970s: ~ 10 mm

Today: ~30 nm

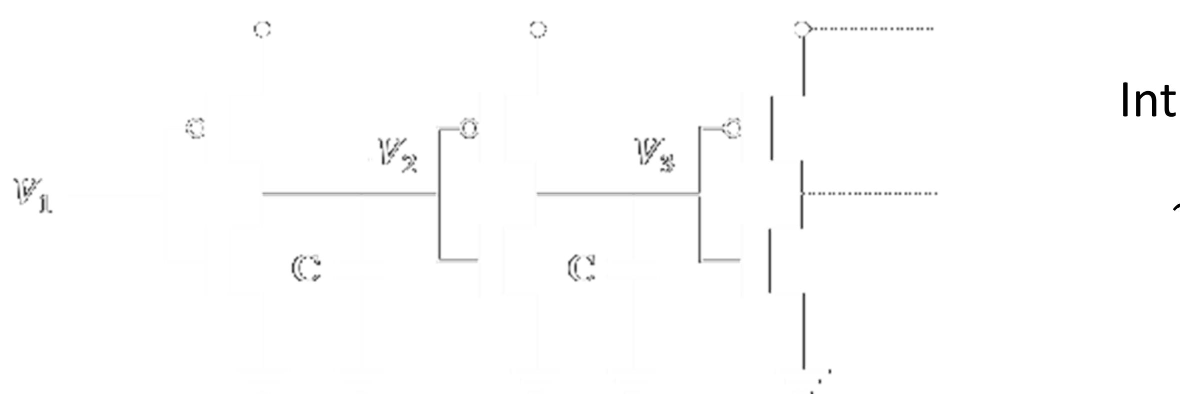
Reasons:

Improved circuit operating speed

Increased device density --> lower cost per function

As MOSFET lateral dimensions (*e.g.* channel length L) are reduced:

- I_{Dsat} increases → decreased effective "R"
- gate and junction areas decrease → decreased load "C"
  - → faster charging/discharging (*i.e.* t_d is decreased)



Intrinsic Delay

$$\tau = \frac{C_{Ox}\, V_{DD}}{I_{ON}}$$

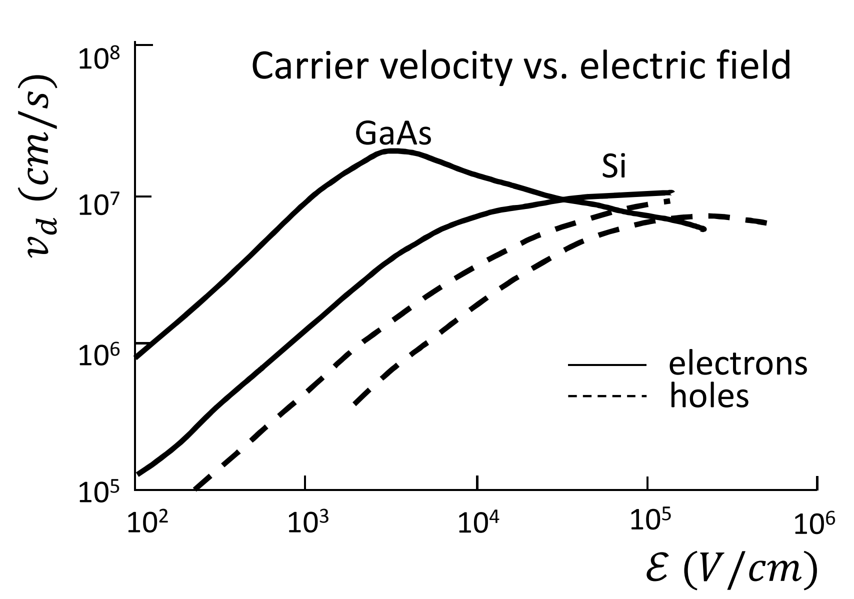
Velocity Saturation

Velocity saturation limits I_{Dsat} in sub-micron MOSFETS

Simple model:

$$v = \begin{cases} \dfrac{\mu \mathcal{E}}{1 + \dfrac{\mathcal{E}}{\mathcal{E}_{sat}}} & for\ \ \mathcal{E} < \mathcal{E}_{sat} \\ v_{sat} & for\ \ \mathcal{E} \geq \mathcal{E}_{sat} \end{cases}$$

$$\mu\, \mathcal{E}_{sat} = 2 v_{sat}$$



$$v_{sat} = \begin{cases} 8 \times 10^6\ cm/s \ \text{ for } e^{-} \text{in Si} \\ 6 \times 10^6\ cm/s \ \text{ for } h^{+} \text{in Si} \end{cases}$$

If $\mathcal{E} < \mathcal{E}_{sat}$: $\quad \mu \mapsto \dfrac{\mu}{1 + \dfrac{\mathcal{E}}{\mathcal{E}_{sat}}}$

1. I
2.
3.
4.
5.

MOSFET I-V with Velocity Saturation

In linear region:

$$\mu \mapsto \frac{\mu}{1+\frac{\mathcal{E}}{\mathcal{E}_{sat}}}$$

$$I_D = \frac{W}{L}\,\frac{\mu_{eff}C_{Ox}}{1+\frac{V_{DS}}{L\mathcal{E}_{sat}}}\left(V_{GS}-V_T-\frac{m}{2}V_{DS}\right)V_{DS} = \frac{I_{D\,LongChannel}}{1+\frac{V_{DS}}{L\mathcal{E}_{sat}}}$$

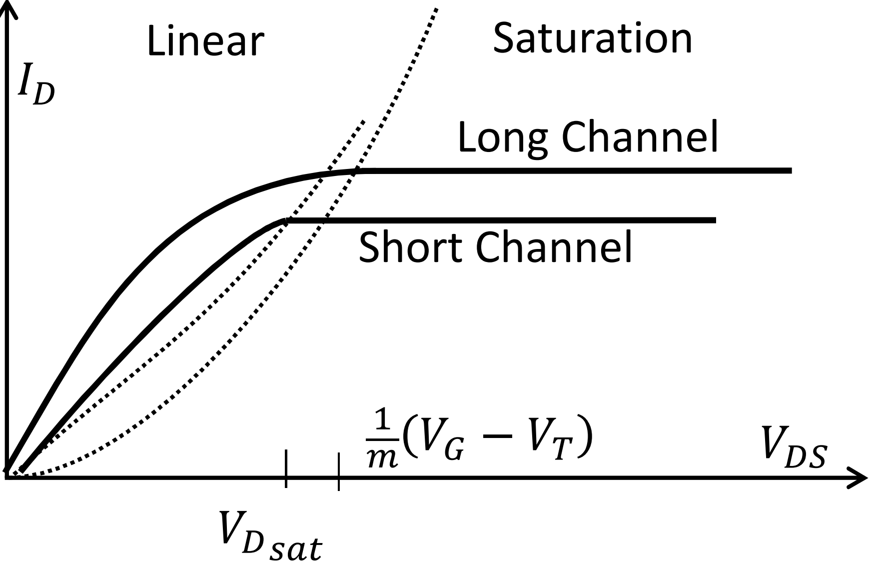
MOSFET is Long channel if $L\mathcal{E}_{sat} \gg V_{GS}-V_T$

$$\frac{1}{V_{D\,sat}} = \frac{m}{V_{GS}-V_T} + \frac{1}{L\mathcal{E}_{sat}}$$

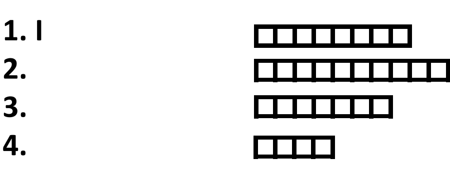
$V_{GS} = 1.8\,V, t_{Ox} = 3\,nm,$
$V_T = 0.25\,V, W_T = 45\,nm$

μ_{eff} =200 cm^{2}V^{-1}s^{-1} , $m = 1 + 3T_{oxe}/W_T = 1.2$
$\mathcal{E}_{sat}$= $2v_{sat}$ / μ_{eff} = 8×10^4 V/cm

$$L = 10\mu m \rightarrow V_{D\,sat} = 1.3V$$
$$L = 1\mu m \rightarrow V_{D\,sat} = 1.1V$$
$$L = 100nm \rightarrow V_{D\,sat} = 0.5V$$
$$L = 30nm \rightarrow V_{D\,sat} = 0.2V$$



I_{Dsat} with Velocity Saturation



In saturation region:

$$V_{DS} \mapsto V_{GS} - V_T$$

$$I_{D_{sat}} = \frac{W}{2mL} \frac{\mu_{eff} C_{Ox}}{1 + \frac{V_{GS} - V_T}{L\mathcal{E}_{sat}}} (V_{GS} - V_T)^2 = \frac{I_{Dsat}{}_{LongChannel}}{1 + \frac{V_{GS} - V_T}{L\mathcal{E}_{sat}}}$$

Very short channel length: $L\mathcal{E}_{sat} \ll V_{GS} - V_T$

$$I_{D_{sat}} = \frac{W}{2m} \mu_{eff} \mathcal{E}_{sat} C_{Ox} (V_{GS} - V_T) = \frac{W}{m} v_{sat} C_{Ox} (V_{GS} - V_T)$$

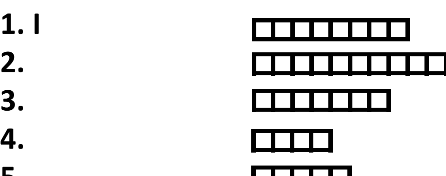
- $I_{D_{sat}}$ is proportional to $V_{GS} - V_T$ rather than $(V_{GS} - V_T)^2$
- $I_{D_{sat}}$ is not dependent on L ☹

To improve modern MOSFETs: $I_{ON} \nearrow$

$C_{Ox} \nearrow$ high-k dielectric

$v_{sat} \nearrow$ strained Si

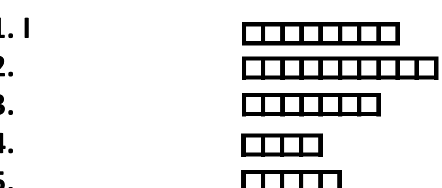
Short- vs. Long-Channel NMOSFET



Short-channel NMOSFET:

- I_{Dsat} is proportional to V_{GS}-V_{Tn} rather than $(V_{GS}$-$V_{Tn})^2$
- V_{Dsat} is lower than for long-channel MOSFET
- Channel-length modulation is apparent

Velocity Overshoot



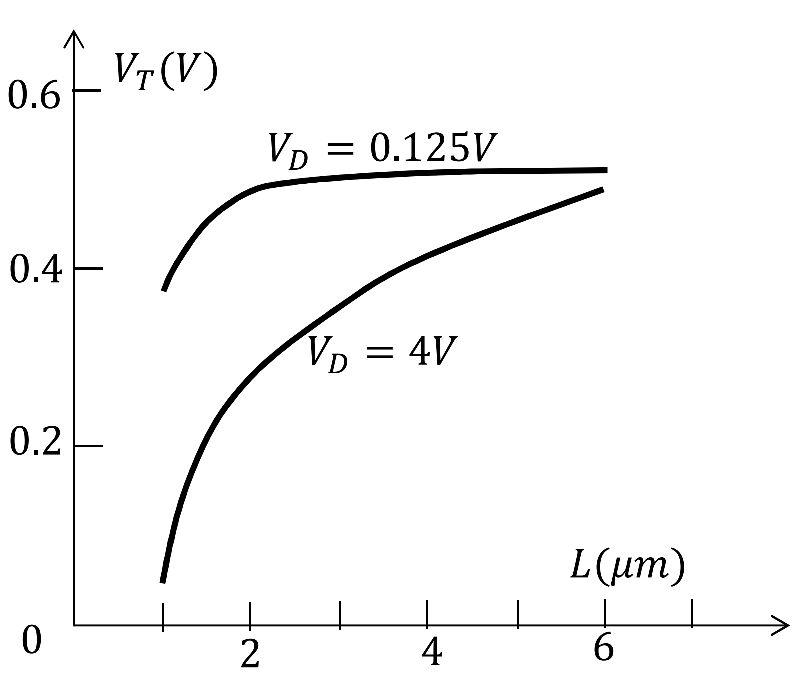
When L is comparable to or less than the mean free path, some of the electrons travel through the channel without experiencing a single scattering event
→ projectile-like motion (“ballistic transport”)

The Short Channel Effect (SCE)

1. I
2.
3.
4.
5.

"V_T roll-off"

- $|V_T|$ decreases with L
- Effect is exacerbated by high values of $|V_{DS}|$



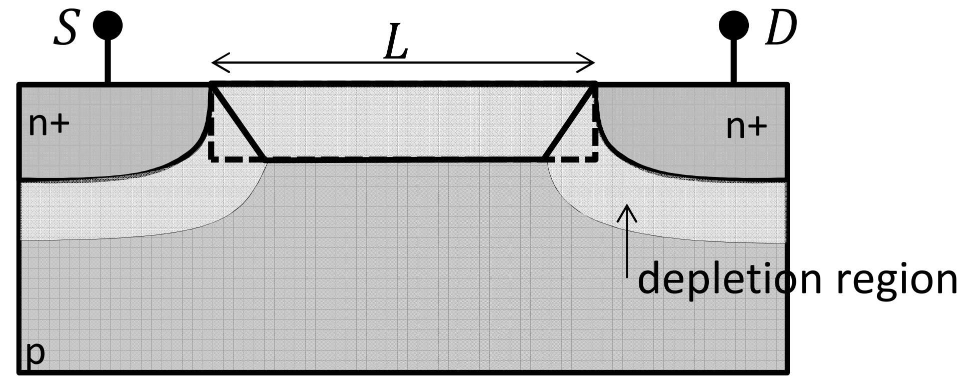
This effect is undesirable (i.e. we want to minimize it!) because circuit designers would like VT to be invariant with transistor dimensions and bias condition

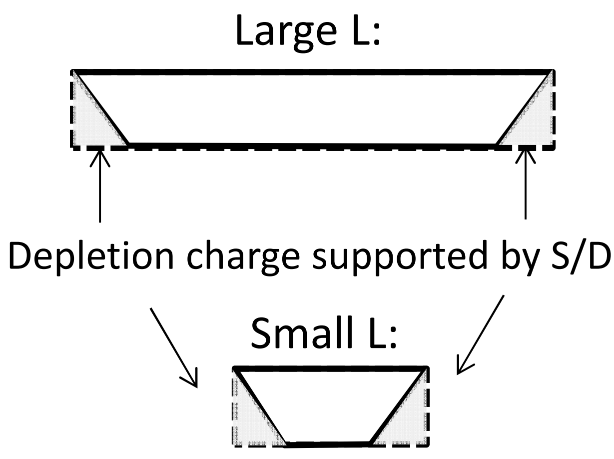
Qualitative Explanation of SCE

Before an inversion layer forms beneath the gate, the surface of the Si underneath the gate must be depleted (to a depth W_T)

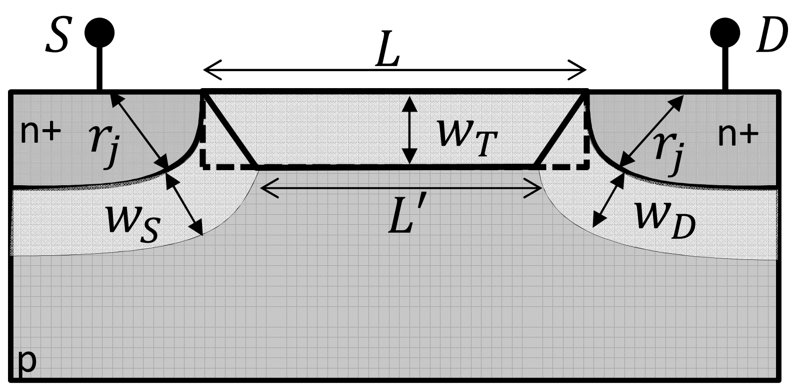
The source & drain pn junctions assist in depleting the Si underneath the gate. Portions of the depletion charge in the channel region are balanced by charge in S/D regions, rather than by charge on the gate. Less gate charge is required to invert the semiconductor surface (i.e. $|V_T|$ decreases)

$$\searrow V_T = V_{FB} + 2\varphi_F + \frac{Q_{depl}\searrow}{C_{Ox}}$$





V_T Roll-Off: First-Order Model



$$V_T = V_{FB} + 2\varphi_F + \frac{Q_{depl}}{C_{Ox}}$$

$$\Delta V_T \equiv |V_T| - \left|V_{T\,Long\ Channel}\right|$$

$$\Delta V_T \propto \frac{qN_A}{C_{Ox}} W_T \left(1 - \frac{L + L'}{2L}\right)$$

$$L' = L - 2r_j\left[\sqrt{1 + \frac{2W_T}{r_j}} - 1\right] \quad \rightarrow \quad \Delta V_T = \frac{-qN_A W_T}{C_{Ox}} \frac{r_j}{L}\left[\sqrt{1 + \frac{2W_T}{r_j}} - 1\right]$$

Minimize ΔV_T by

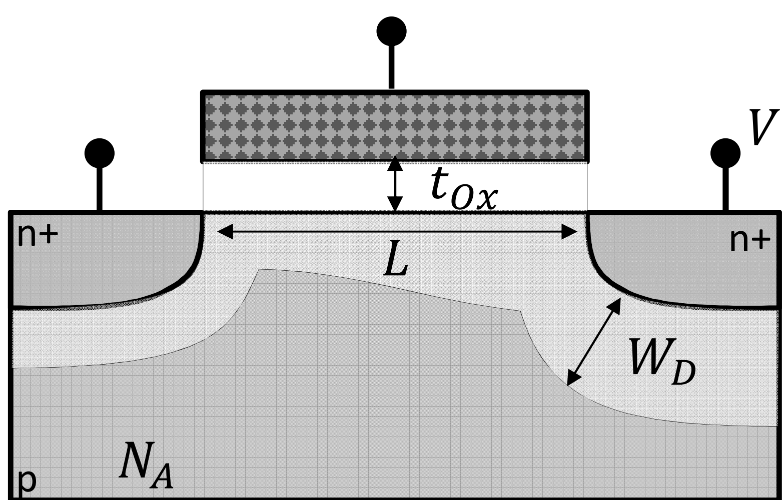
- reducing t_{Ox}
- reducing r_j
- increasing N_A (trade-offs: degraded μ_{eff}, m)

MOSFET vertical dimensions should be scaled along with horizontal dimensions!

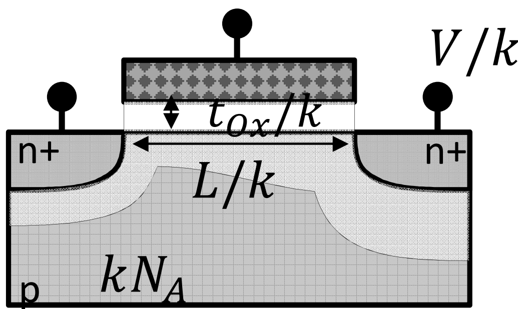
MOSFET Scaling: Constant-Field Approach

MOSFET dimensions and the operating voltage (V_{DD}) each are scaled by the same factor $k > 1$, so that the electric field remains unchanged.

Original device



Scaled device



Constant-Field Scaling Benefits

1. I
2.
3.
4.
5.

	Parameter	Multiplication factor (k>1)
Scaling assumptions	Device dimensions (t_{Ox}, L, W, r_j)	$1/k$
	Doping concentration (N_A, N_D)	k
	Voltage (V)	$1/k$
Derived scaling behavior of device parameters	Electric field ($\mathcal{E}$)	1
	Carrier velocity (v)	1
	Depletion-layer width (W_D)	$1/k$
	Capacitance ($C = \epsilon A/t$)	$1/k$
	Inversion charge density (Q_{inv})	1
	Current drift (I)	$1/k$
	Channel resistance (R_{ch})	1
Derived scaling behavior of circuit parameters	Circuit delay time ($\tau \sim CV/I$)	$1/k$
	Power diss. per circuit ($P \sim VI$)	$1/k^2$
	Power-delay product per circuit ($P\tau$)	$1/k^3$
	Circuit density ($\propto 1/A$)	k^2
	Power density (P/A)	1

Failure of Constant-Field Scaling

Since V_T cannot be scaled down aggressively, the operating voltage (V_{DD}) has not been scaled down in proportion to the MOSFET channel length:

Feature Size (μm)	**Power-Supply Voltage (V)**	**Gate Oxide Thickness (Å)**	**Oxide Field (MV/cm)**
2	5	350	104
1.2	5	250	2.0
0.8	5	180	2.8
0.5	3.3	120	2.8
0.35	3.3	100	3.3
0.25	2.5	70	3.6

MOSFET Scaling: Generalized Approach

1. I
2.
3.
4.
5.

	Parameter	Multiplication factor (k>1)	
Scaling assumptions	Device dimensions (t_{Ox}, L, W, r_j)	$1/k$	
	Doping concentration (N_A, N_D)	αk	
	Voltage (V)	α / k	
Derived scaling behavior of device parameters	Electric field ($\mathcal{E}$)	α	
	Depletion-layer width (W_D)	$1/k$	
	Capacitance ($C = \epsilon A/t$)	$1/k$	
	Inversion charge density (Q_{inv})	α	
		Long ch.	Vel Sat.
	Carrier velocity (v)	α	1
	Current drift (I)	α^2/k	α/k
Derived scaling behavior of circuit parameters	Circuit delay time ($\tau \sim CV/I$)	$1/\alpha k$	$1/k$
	Power diss. per circuit ($P \sim VI$)	α^3/k^2	α^2/k^2
	Power-delay product per circuit ($P\tau$)	α^2/k^3	
	Circuit density ($\propto 1/A$)	k^2	
	Power density (P/A)	α^3	α^2

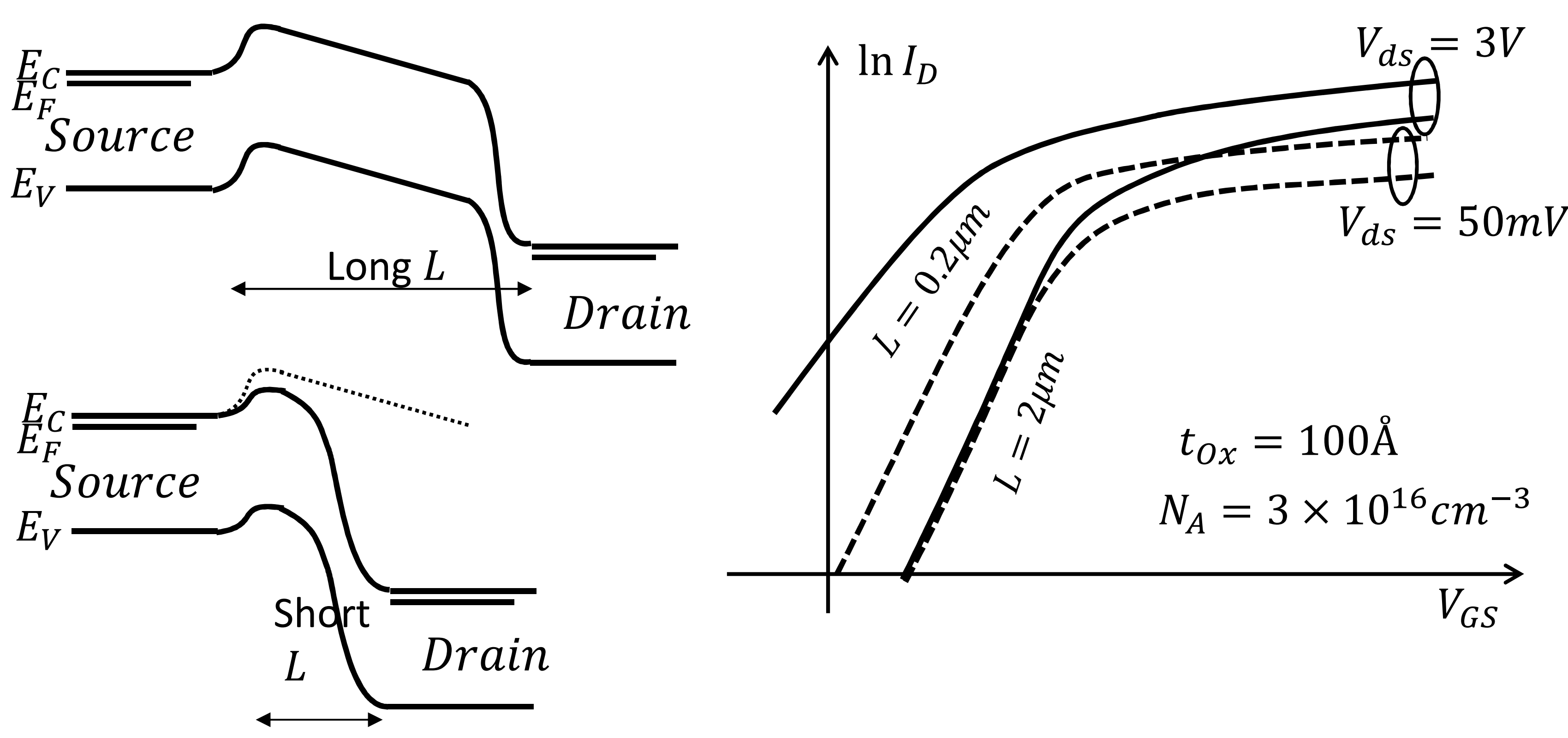
Electric field intensity increases by a factor $\alpha > 1$

N_{body} must be scaled up by α to suppress short-channel effects

Drain Induced Barrier Lowering (DIBL)

As the source and drain get closer, they become electrostatically coupled, so that the drain bias can affect the potential barrier to carrier diffusion at the source junction

→ V_T decreases (*i.e.* OFF state leakage current increases)



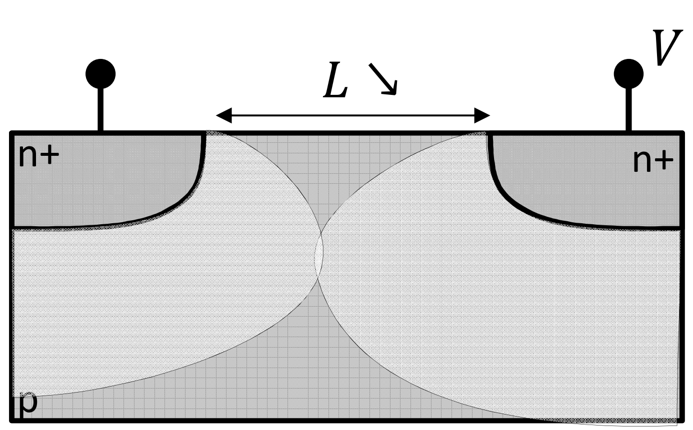
Punchthrough

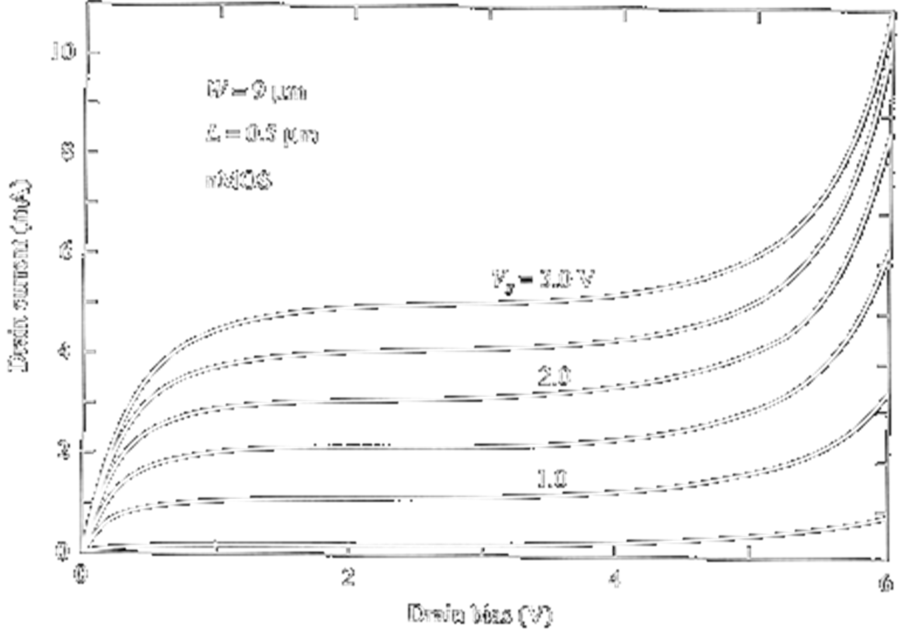
1. I
2.
3.
4.
5.

A large drain bias can cause the drain-junction depletion region to merge with the source-junction depletion region, forming a sub-surface path for current conduction.

→ I_{Dsat} increases rapidly with V_{DS}

This can be mitigated by doping the semiconductor more heavily in the sub-surface region, i.e. using a "retrograde" doping profile.





Source and Drain (S/D) Structure

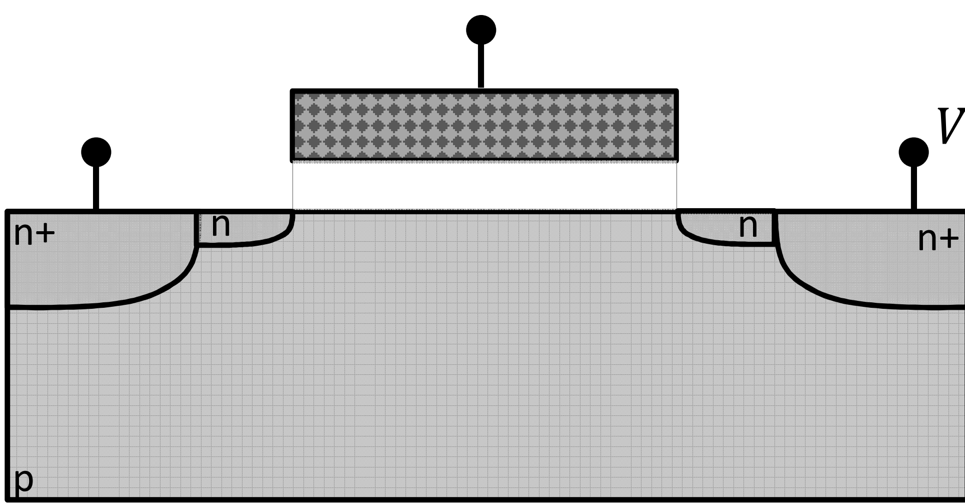
1. I
2.
3.
4.
5.

To minimize the short channel effect and DIBL, we want shallow (small r_j) S/D regions – but the parasitic resistance of these regions increases when r_j is reduced.

$$R_{source}, R_{drain} \propto \rho/W r_j$$

where ρ = resistivity of the S/D regions

Shallow S/D "extensions" may be used to effectively reduce r_j with a relatively small increase in parasitic resistance



$\mathcal{E}$ - Field Distribution Along the Channel

1. I
2.
3.
4.
5.

The lateral electric field peaks at the drain end of the channel.

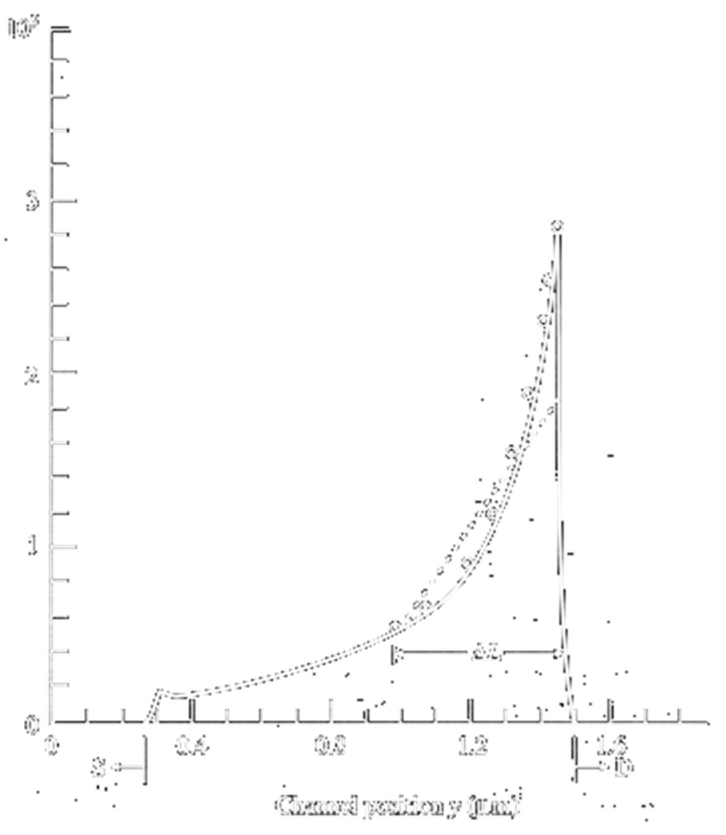
$\mathcal{E}_{peak}$ can be as high as 10^6 V/cm

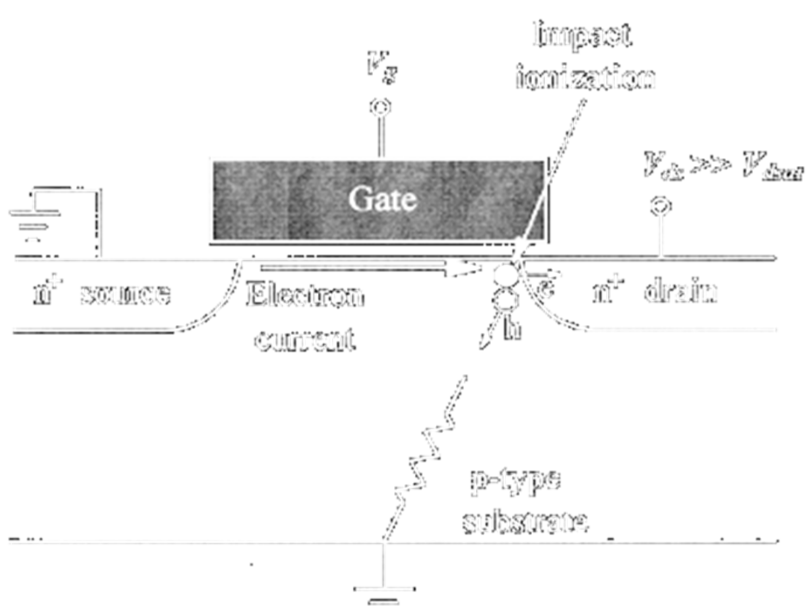
High E-field causes problems:

Damage to oxide interface & bulk

(trapped oxide charge → V_T shift)

substrate current due to impact ionization:

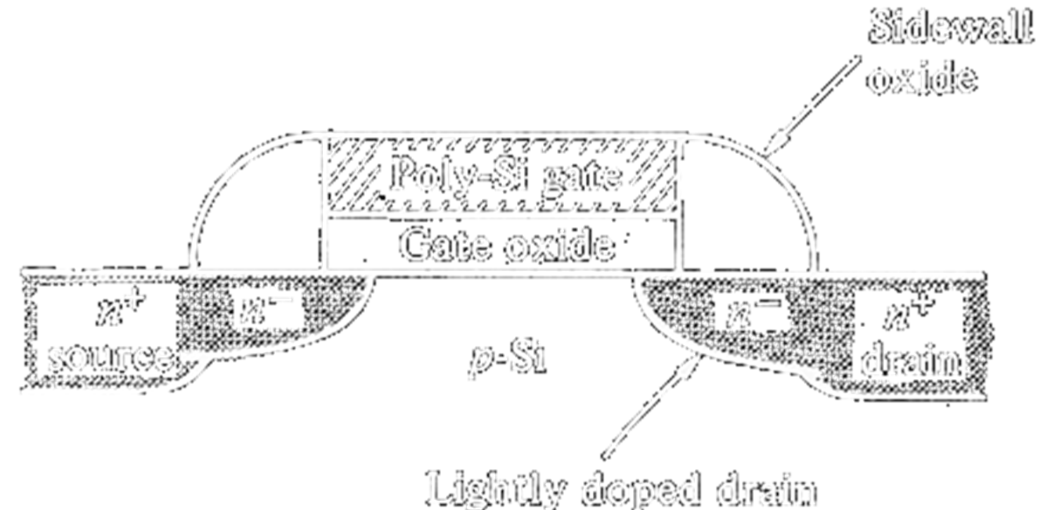




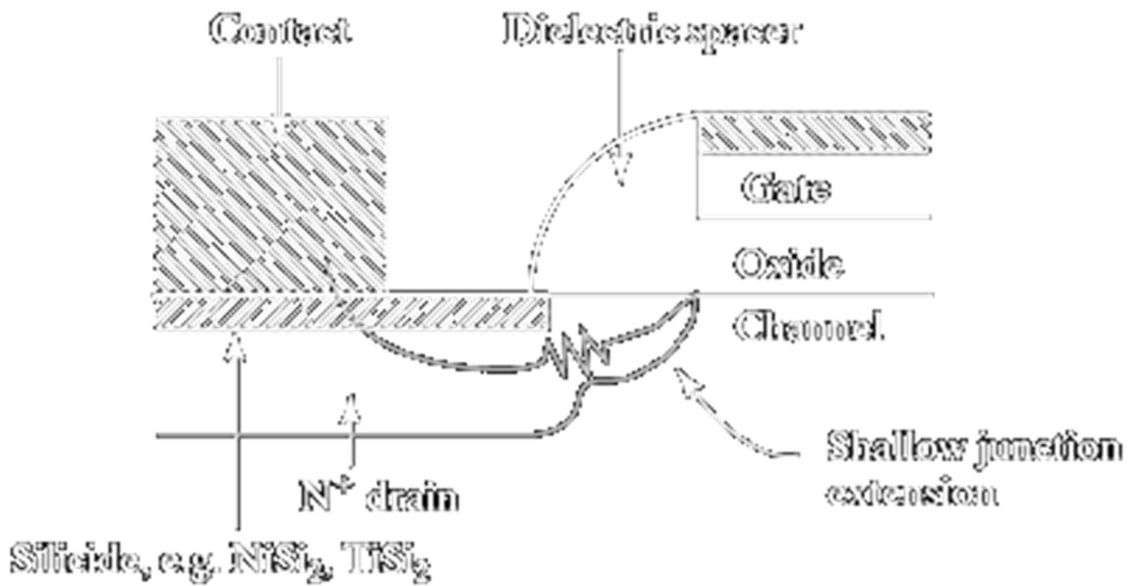
Lightly Doped Drain (LDD) Structure

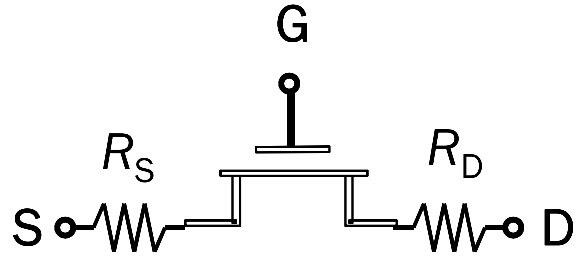
Lower pn junction doping results in lower peak E-field

- ✓ "Hot-carrier" effects are reduced
- ✗ Parasitic resistance is increased



Parasitic Source-Drain Resistance





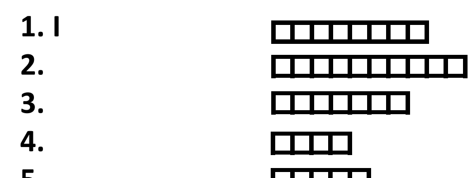
For short-channel MOSFET, $I_{Dsat0} \propto V_{GS} - V_T$, so that

$$I_{Dsat} = \frac{I_{Dsat0}}{1 + \frac{I_{Dsat} R_s}{V_{GS} - V_T}}$$

→ I_{Dsat} is reduced by ~15% in a 0.1 mm MOSFET.

V_{Dsat} is increased to $V_{Dsat0} + I_{Dsat}(R_S + R_D)$

Summary: MOSFET OFF State vs. ON State



OFF state ($V_{GS} < V_T$):

- I_{DS} is limited by the rate at which carriers diffuse across the source pn junction
- Minimum subthreshold swing S, and DIBL are issues

ON state ($V_{GS} > V_T$):

- I_{DS} is limited by the rate at which carriers drift across the channel
- Punchthrough is of concern at high drain bias
    - I_{DSat} increases rapidly with V_{DS}
- Parasitic resistances reduce drive current
    - source resistance R_S reduces effective V_{GS}
    - source & drain resistances R_S & R_D reduce effective V_{DS}