The Devices:
MOS Transistors

References:

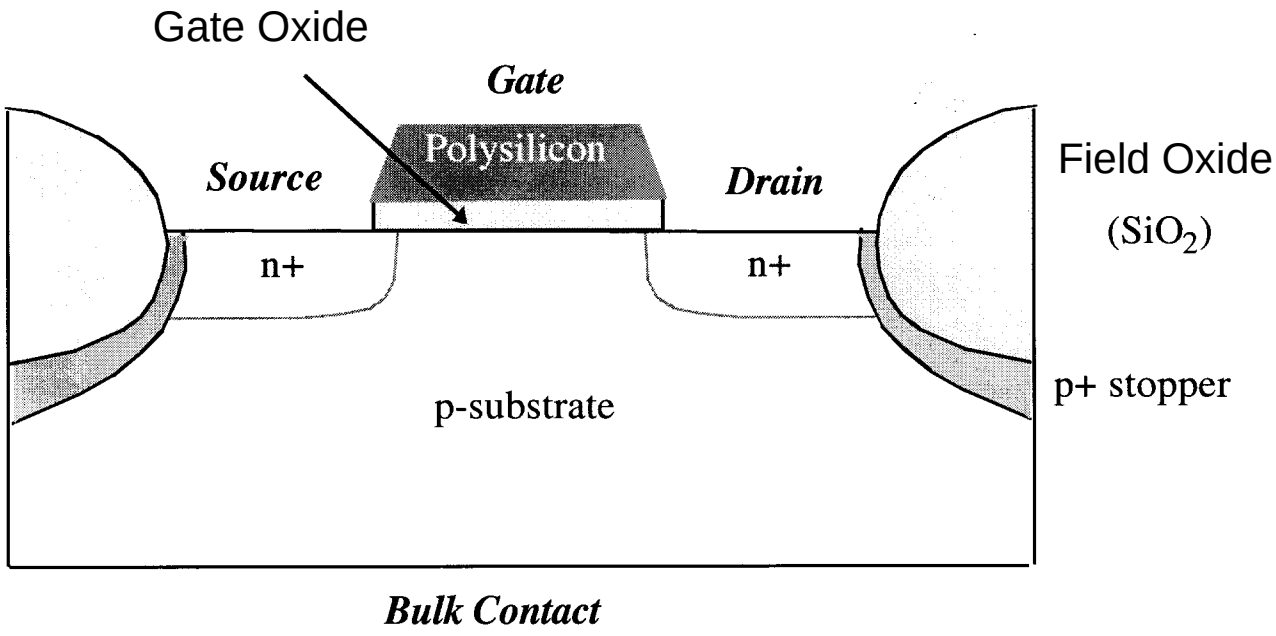
Semiconductor Device Fundamentals,

R. F. Pierret, Addison-Wesley

Adapted from: *Digital Integrated Circuits: A Design Perspective*, J. Rabaey, Prentice Hall © UCB

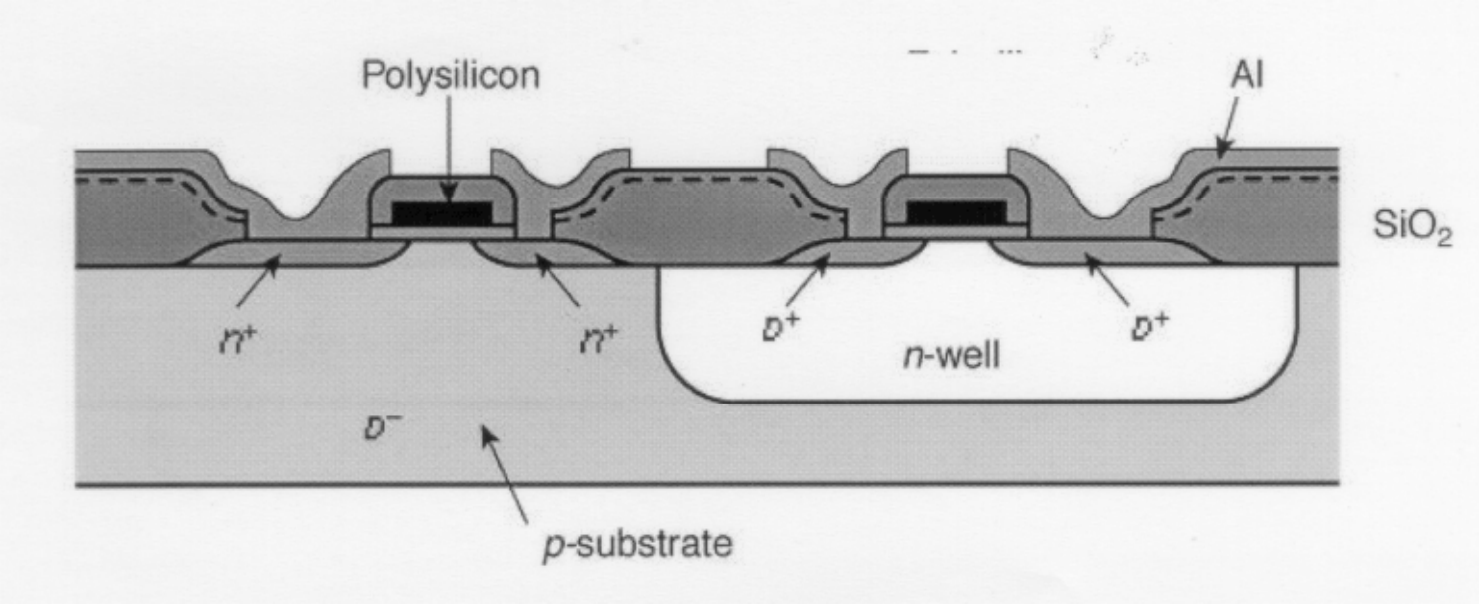
MOS Transistor
(Metal-Oxide-Semiconductor)

NMOS Transistor

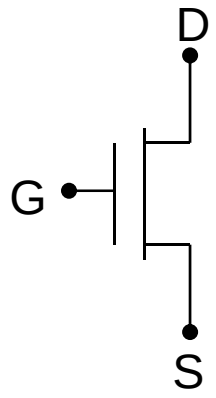


CROSS-SECTION of NMOS Transistor

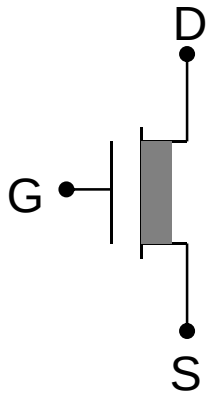
Cross-Section of CMOS Technology



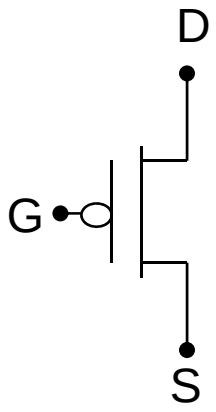
MOS transistors - types and symbols



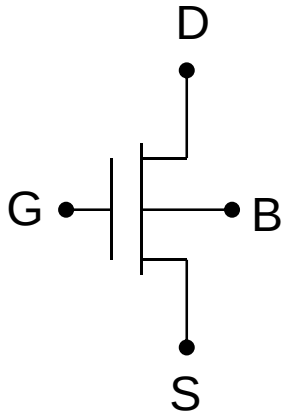
NMOS Enhancement



NMOS Depletion

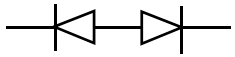


PMOS Enhancement



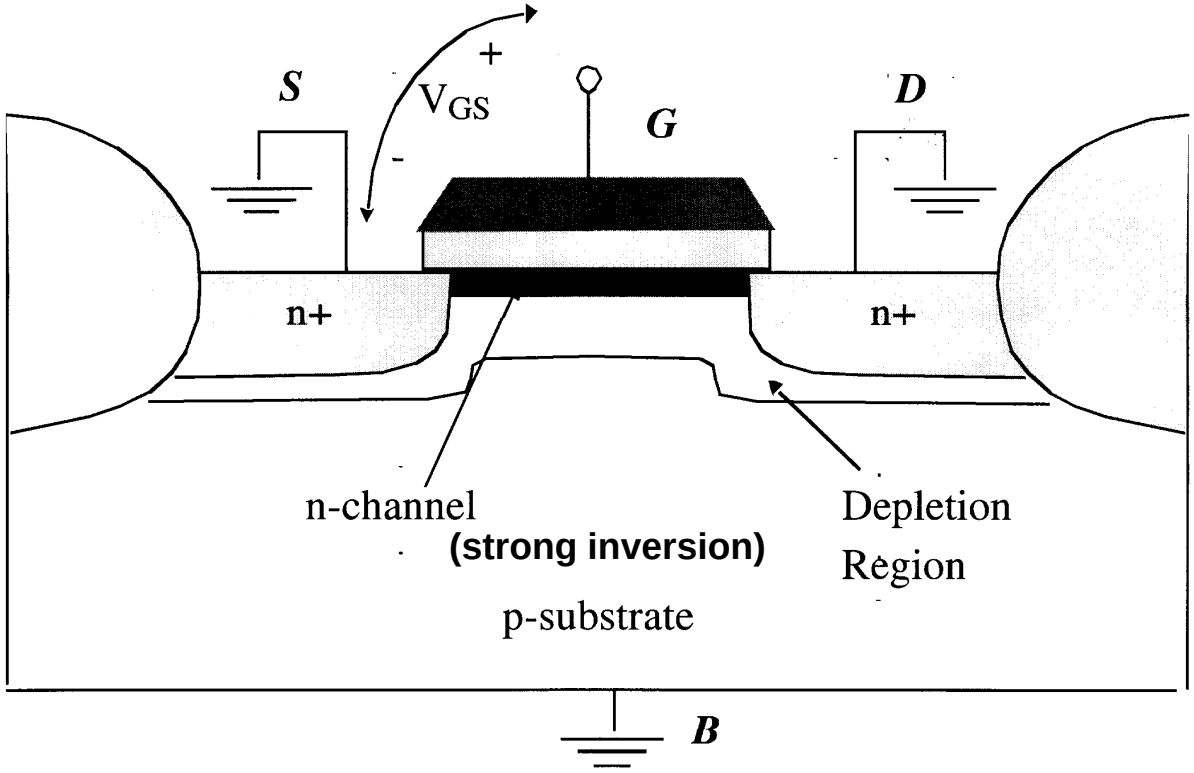
NMOS with Bulk Contact

Threshold Voltage: Concept



$V_T = V_{FB} + V_B + V_{ox}$

$V_B = 2\phi_F$ Fermi potential



Threshold Voltage: Concept

- Threshold voltage due to ideal MOS structure
  - Voltage to invert the character of the surface region from n-type to p-type and vice versa
  - Voltage drop due to gate oxide
- Threshold voltage due to non-ideal MOS structure
  - Difference in the work functions of metal and semiconductor
  - Charges in the gate oxide
  - Ion-implantation
  - Body effect
  - ...

Depletion Width and Electric Field

- Poisson's equation $\frac{d\mathrm{E}}{dx} = \frac{\rho}{K_S \varepsilon_0} \cong -\frac{qN_A}{K_S \varepsilon_0} \quad (0 \le x \le W)$

  K_S : dielectric constant

  ε_0 : permitivity of free space

- Electric Field $\mathrm{E}(x) = \frac{qN_A}{K_S \varepsilon_0}(W - x) \quad (0 \le x \le W)$

- Depletion width $\phi(x) = \frac{qN_A}{2K_S \varepsilon_0}(W - x)^2 \quad (0 \le x \le W)$

$$\phi_S = \frac{qN_A}{2K_S \varepsilon_0} W^2 \Leftrightarrow W = \left[\frac{2K_S \varepsilon_0}{qN_A} \phi_S \right]^{1/2}$$

$$W_{\max} = \left[\frac{4K_S \varepsilon_0}{qN_A} \phi_F \right]^{1/2} \qquad \mathrm{E}_{S,\max} = \left[\frac{4qN_A}{K_S \varepsilon_0} \phi_F \right]^{1/2}$$

Threshold Adjustment by Ion Implantation

- Implant a relatively small, precisely controlled number of either boron or phosphorus ions into the near-surface region of semiconductor
- Implantation of boron causes a positive shift in threshold voltage
- Implantation of phosphorus causes a negative shift
- Like placing additional “fixed” charges

$$\Delta V = -\frac{Q_I}{C_{ox}} \qquad Q_I = \pm qN_I$$

$$(+): \text{donor} \quad (-): \text{acceptor}$$

Back Biasing or Body Effect

- V_{SB} is normally positive for n-channel devices, negative for p-channel devices
- Always increases the magnitude of the ideal device threshold voltage
- Inversion occurs at $\phi_S = (2\phi_F + V_{SB})$
- Increases the charges stored in depletion region

$$Q_B = \sqrt{2qN_A\varepsilon_{si}(2\phi_F + V_{SB})}$$

Threshold voltage

$$V_T = V_{FB} + V_B + V_{ox}$$

$$V_T = \left(\Phi_{ms} - \frac{Q_{ox}}{C_{ox}} - \frac{Q_I}{C_{ox}} \right) - 2\Phi_F - \frac{Q_B}{C_{ox}}$$

The Threshold Voltage

$$V_T = \phi_{MS} + 2\phi_F + \frac{Q_B}{C_{ox}} - \frac{Q_I}{C_{ox}} - \frac{Q_M \gamma_M}{C_{ox}} - \frac{Q_F}{C_{ox}} - \frac{Q_{IT}(2\phi_F)}{C_{ox}}$$

- In general $V_{FB} = \phi_{MS} - \frac{Q_I}{C_{ox}} - \frac{Q_M \gamma_M}{C_{ox}} - \frac{Q_F}{C_{ox}} - \frac{Q_{IT}(0)}{C_{ox}}$

$$V_T = V_{FB} + V_B + V_{ox}$$

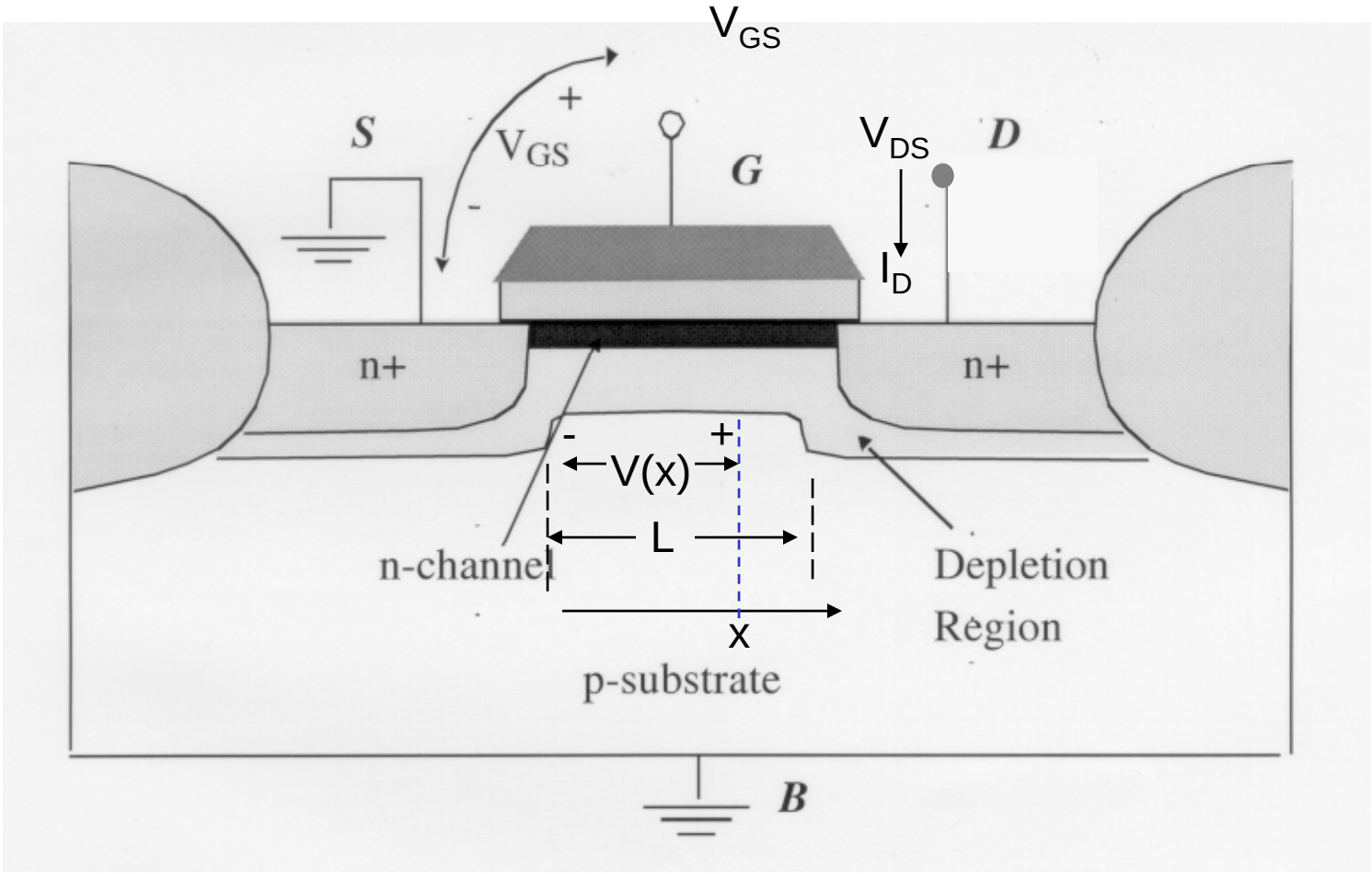
$$V_B = 2\phi_F$$

- NMOS: $V_{SB} > 0$, PMOS: $V_{SB} < 0$

$$V_{ox} = \frac{K_s}{K_0} x_0 \sqrt{\frac{2qN_A}{K_S \varepsilon_0}(2\phi_F + V_{SB})} \quad \text{for NMOS}$$

$$V_{ox} = -\frac{K_s}{K_0} x_0 \sqrt{\frac{2qN_D}{K_S \varepsilon_0}(-2\phi_F - V_{SB})} \quad \text{for PMOS}$$

Current-Voltage Relations



At x, the gate to channel voltage equals V_{GS} - V(x)

Transistor in Linear Region

- Assume that the voltage exceeds V_T all along the channel
- Induced charge/area at point x

$$Q_i(x) = -C_{ox}[V_{GS} - V(x) - V_T]$$

- Current $I_D = -v_n(x).Q_i(x).W$

$v_n(x)$: drift velocity $v_n = -\mu_n E(x) = \mu_n \frac{dV}{dx}$

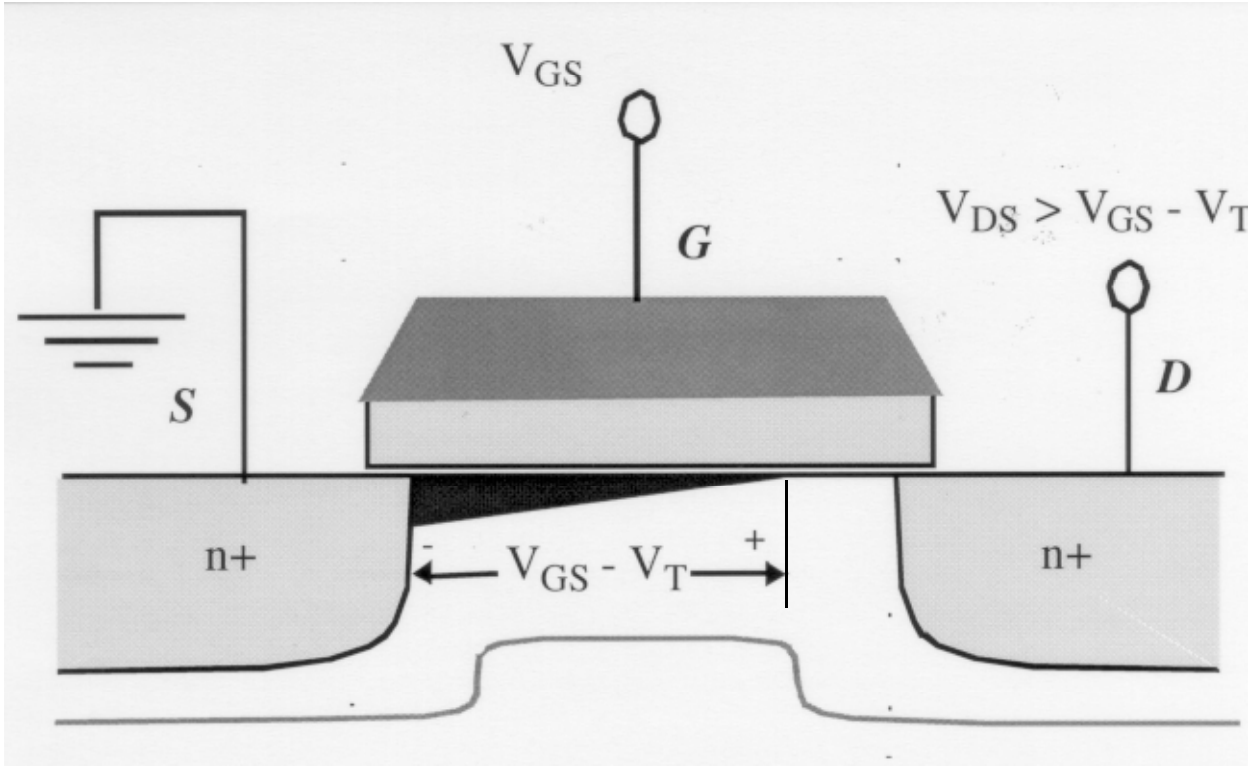
$$\therefore I_d dx = \mu_n.C_{ox}.W(V_{GS} - V - V_T)dV$$

- Integrating over the length of the channel L

$$I_D = K'_n \frac{W}{L}((V_{GS} - V_T).V_{DS} - \frac{V_{DS}^2}{2})$$

$$K'_n = \mu_n C_{ox} = \mu_n \frac{C_{ox}}{T_{ox}}$$

Transistor In Saturation



Transistor in Saturation

- If drain-source voltage increases, the assumption that the channel voltage is larger than V_T all along the channel ceases to hold.

- When V_{GS} - V(x) < V_T pinch-off occurs

- Pinch-off condition

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

Saturation Current

- The voltage difference over the induced channel (from pinch-off to the source) remains fixed at V_{GS} - V_T and hence, the current remains constant.

- Replacing V_{DS} by V_{GS}-V_T in equation for I_D yields

$$I_D = \frac{K'_n}{2}\frac{W}{L}(V_{GS} - V_T)^2$$

- Effective length of the conductive channel is modulated by applied V_{DS} - Channel Length Modulation

Current-Voltage Relations

Cut-off: $V_{GS} \leq V_T$, $I_{DS} \approx 0$

Linear Region: $V_{DS} < V_{GS} - V_T$

$$I_D = k_n' \frac{W}{L}\left((V_{GS} - V_T)V_{DS} - \frac{V_{DS}^2}{2} \right)$$

$$k'_n = \mu_n C_{ox} = \frac{\mu_n \varepsilon_{ox}}{t_{ox}}$$ Process Transconductance Parameter

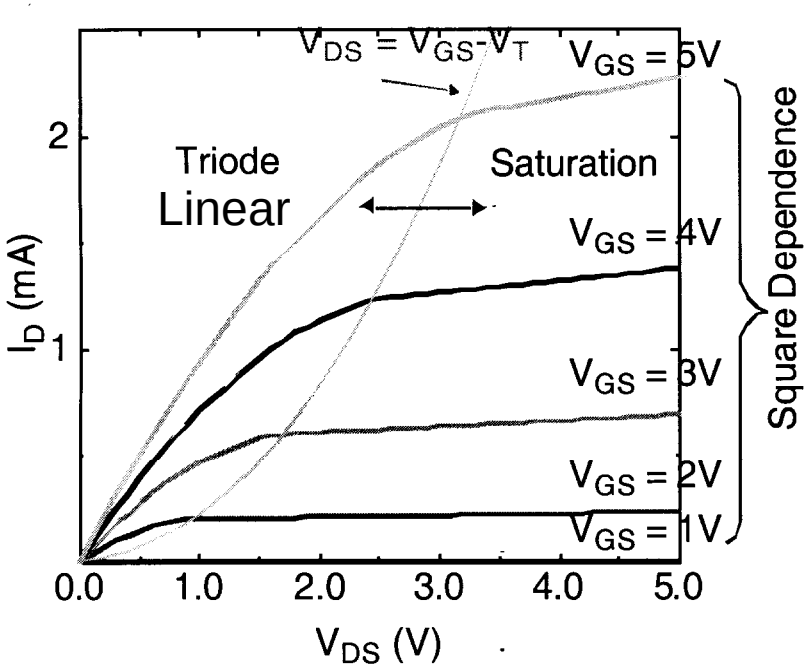
Saturation Mode: $V_{DS} \geq V_{GS} - V_T$

Channel Length Modulation

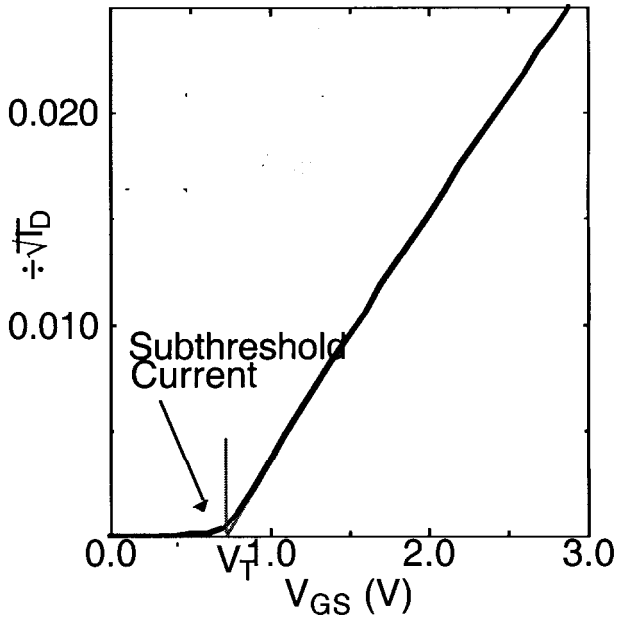
$$I_D = \frac{k'_n}{2}\frac{W}{L}(V_{GS} - V_T)^2(1 + \lambda V_{DS})$$

I-V Relations

Linear: $V_{DS} < V_{GS} - V_T$



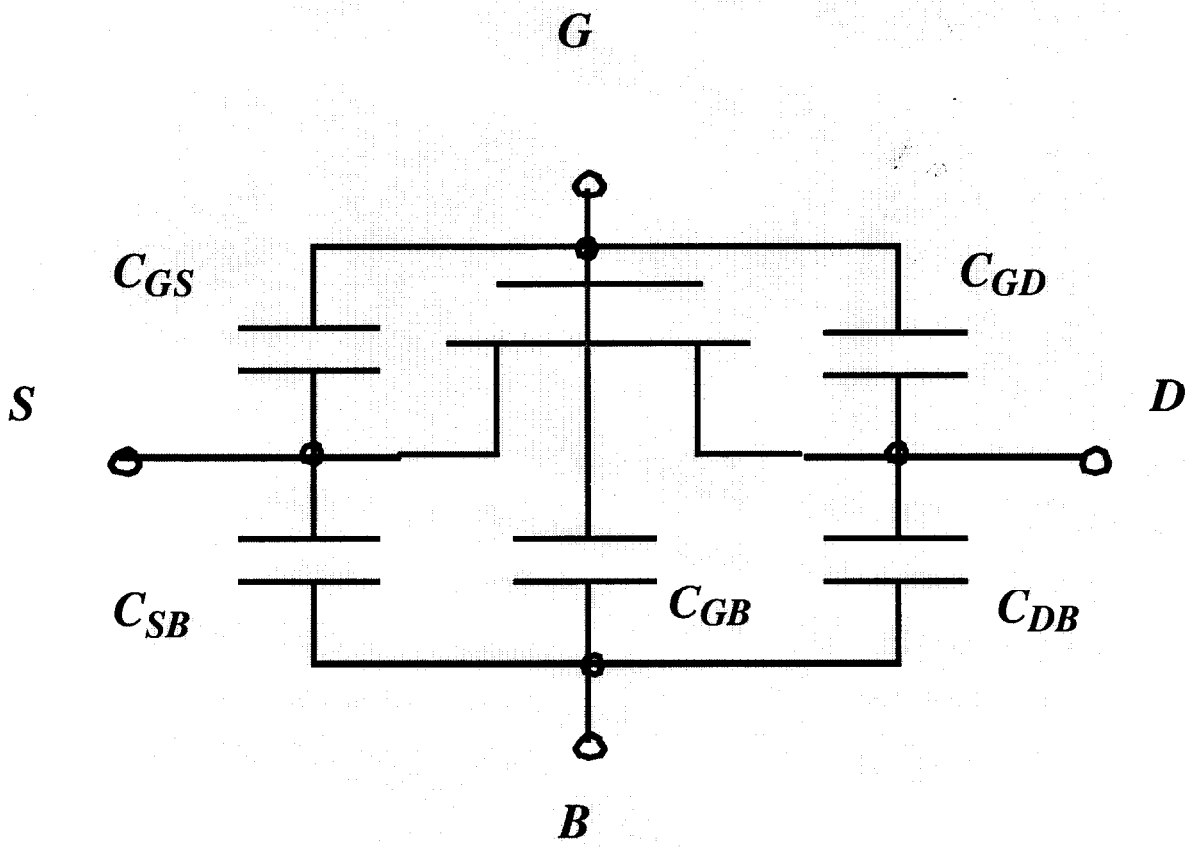
(a) ID as a function of VDS



(b) $\sqrt{I_D}$ as a function of V_{GS} (for V_{DS} = 5V)

NMOS Enhancement Transistor: W = 100 μm ,L = 20 μm

Dynamic Behavior of MOS Transistor



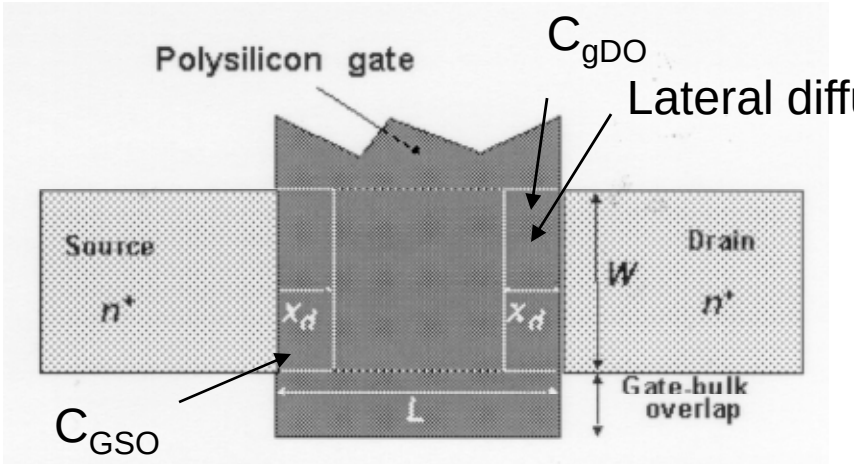
Source of Cap. - Basic MOS structure

- channel charge

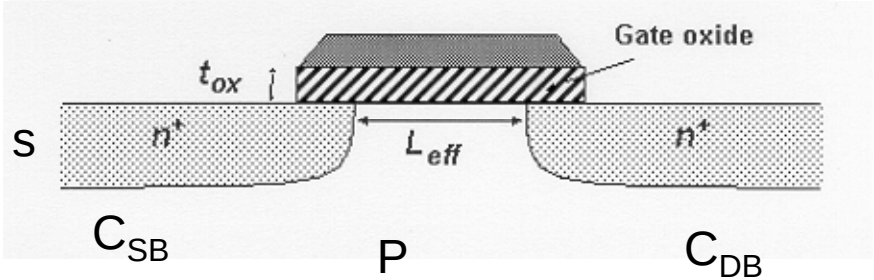
- depletion region of resource bias p-n junctions

The Gate Capacitance

(a) Top view



(b) Cross-section



$$C_{gate} = \frac{\varepsilon_{ox}}{t_{ox}} WL$$

Can be decomposed into a number of elements each with a different behavior

Parasitic capacitance between gate and source (drain) called **Overlap Capacitance** (linear)

$$C_{gsO} = C_{gdO} = C_{ox}.x_d.W = C_o.W$$

Channel Capacitance: C_{gs}, C_{gd}, and C_{gb}

Cut-Off: no channel, total capacitance = $C_{ox}WL_{eff}$ appears between gate and bulk

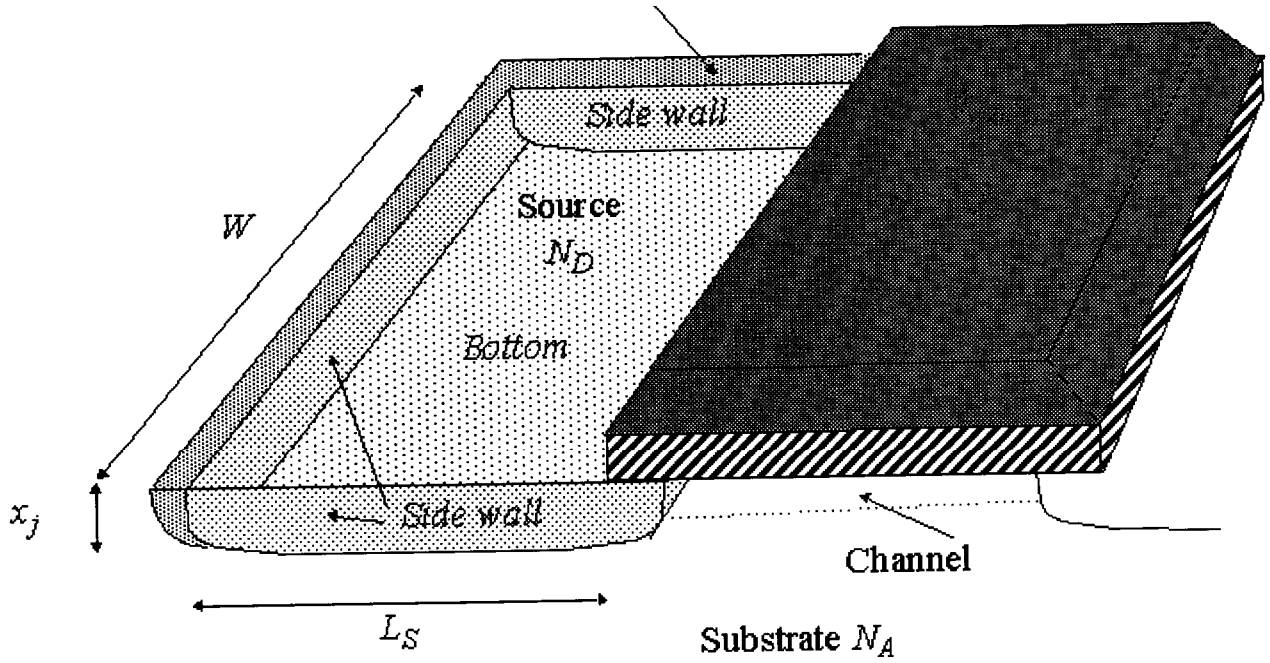
Triode Region: Inversion layer **-** acts as conductor $\therefore C_{gb} = 0$

Symmetry dictates $C_{gs} \approx C_{gd} \approx \dfrac{C_{ox}WL_{eff}}{2}$

Saturation: Pinch off, $\therefore C_{gd} \approx 0, C_{gb} = 0$

C_{gs} averages $(2/3)C_{ox}WL_{eff}$

Diffusion Capacitance (Junction Capacitance)



Reverse biased source-bulk and drain-bulk pn junctions

- **Bottom plate**

$$C_{bottom} = C_j W L_s,$$

- **Side-wall junctions** - formed by source (N_D) and P^+ channel stop (N_A^+)

- graded junction (m=1/3)

$$C_{sw} = C'_{jsw} x_j (w + 2L_s)$$

$$= C_{jsw}(W + 2L_s)$$

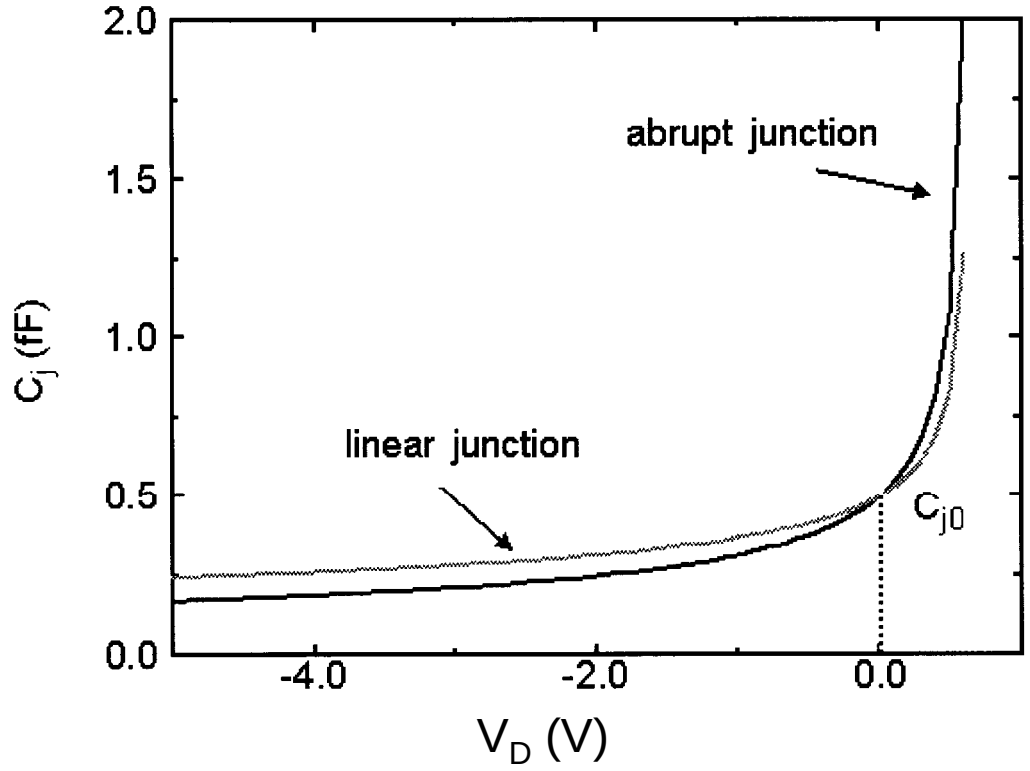
$$C_{jsw} = C'_{jsw} x_j \quad , \qquad x_j = \text{junction depth}$$

- $C_{diff} = C_{bottom} + C_{sw}$

$$= C_j * \text{Area} + C_{jsw} \times \text{Perimeter}$$

$$= C_i L_s W + C_{isw} (2L_s + W)$$

Junction Capacitance



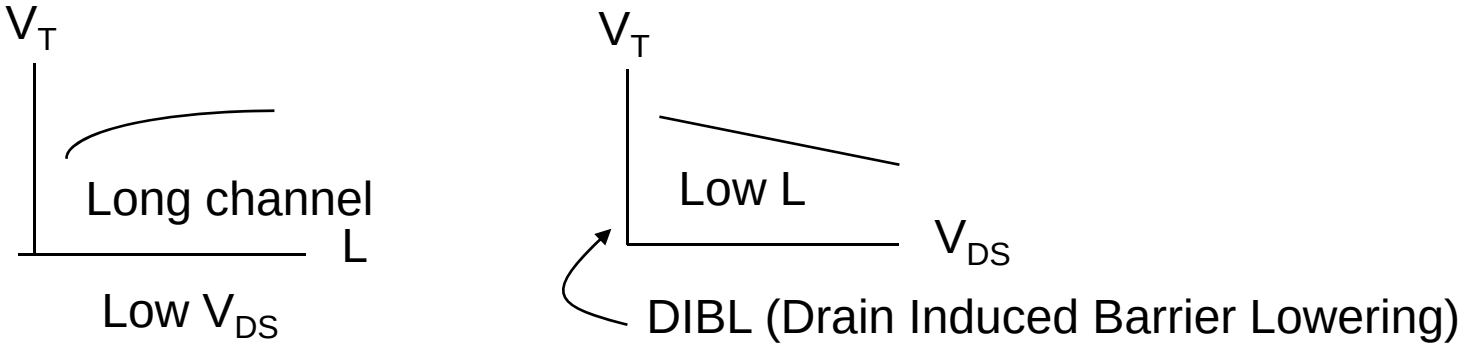
$$C_j = \frac{C_{j0}}{(1 - V_D / \phi_0)^m}$$

The Sub-Micron MOS Transistor

- Threshold Variations (Manufacturing tech., V_{SB})
- Parasitic Resistances
- Velocity Saturation and Mobility Degradation
- Subthreshold Conduction
- Latchup

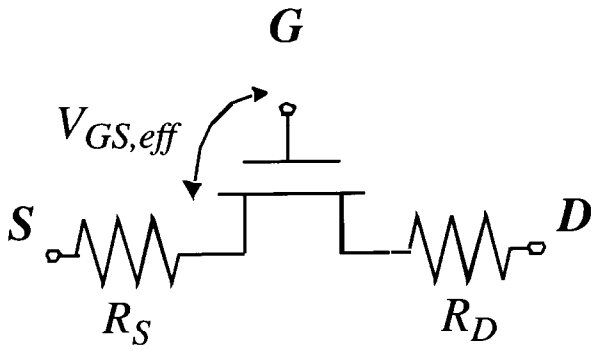
Threshold Variations

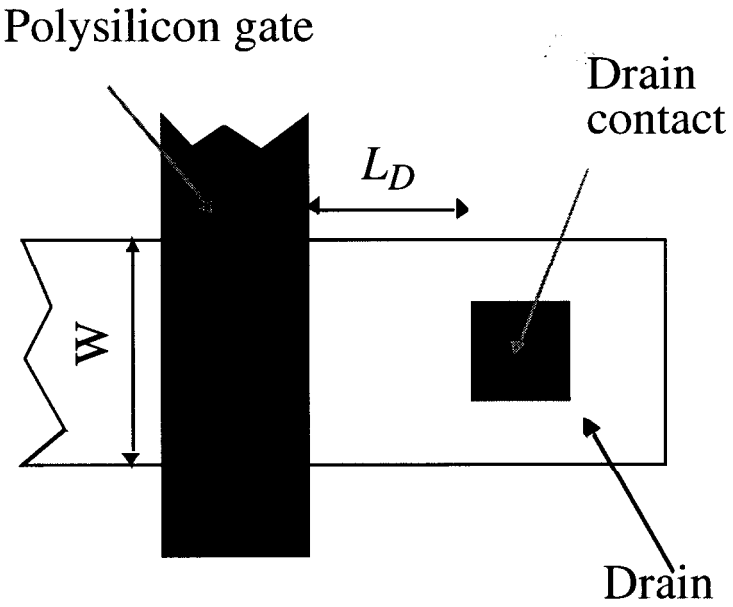
- In derivation of V_T the following assumption were made:
  - charge beneath gate originates from MOS field effects
  - ignores depletion region the source and drain junctions (reverse biased)
- A part of the region below the gate is already depleted (by source & drain fields), a smaller V_T suffices to cause strong inversion
- V_T decreases with L
- Similar effect can be obtained by increasing V_{DS} or V_{DB} as it increases drain-junction depletion region



- V_T can also drift over time (Hot-carrier effect)
  - Decreased device dimensions
  - Increase in electrical field
  - Increasing velocity of electrons, can leave Si surface and enter gate oxide
  - Electrons trapped in gate oxide change V_T (increases in NMOS, decreases in PMOS)
- For a electron to be hot, electric field of 10^4 V/cm is necessary
  - Condition easily met for sub-micron devices

Parasitic Resistances

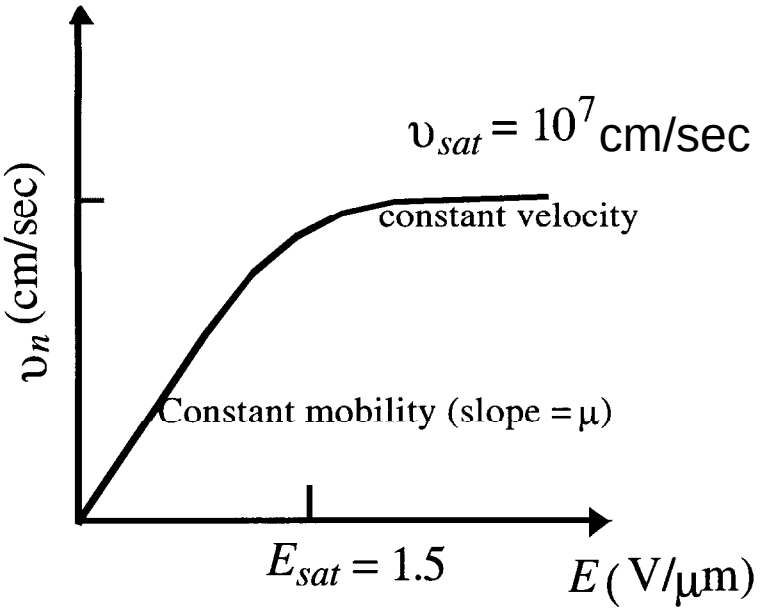




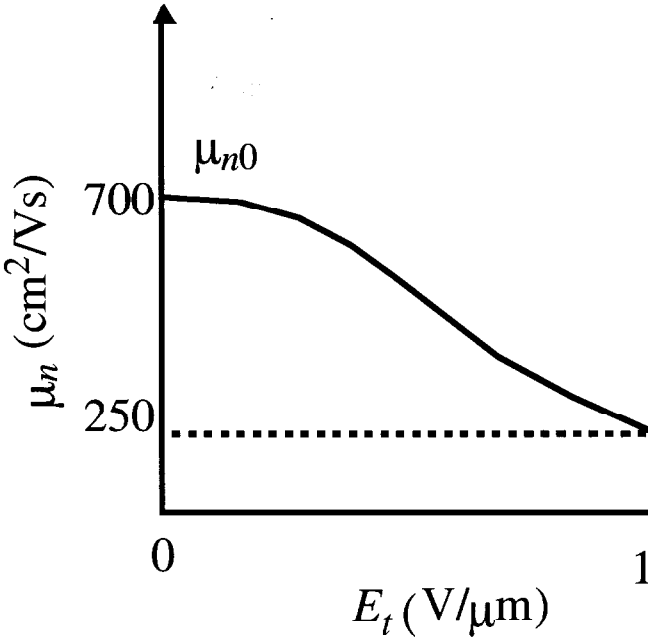
$$R_S = \frac{L}{W} R_{[]} + R_C$$

Solutions: cover the diffusion regions with low-resistivity material such as titanium or tungsten, or make the transistor wider

Velocity Saturation (1) short channel devices

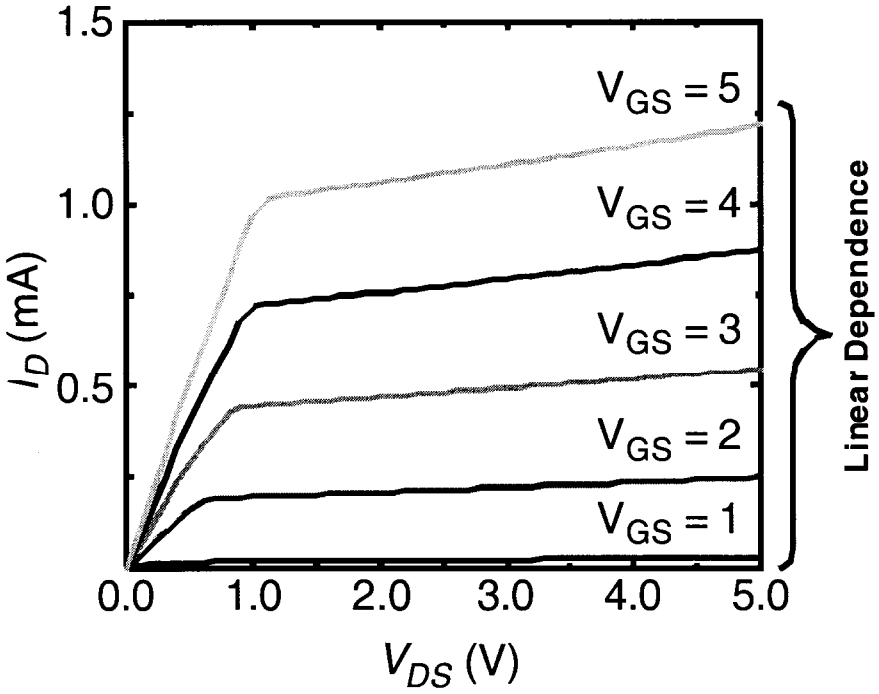


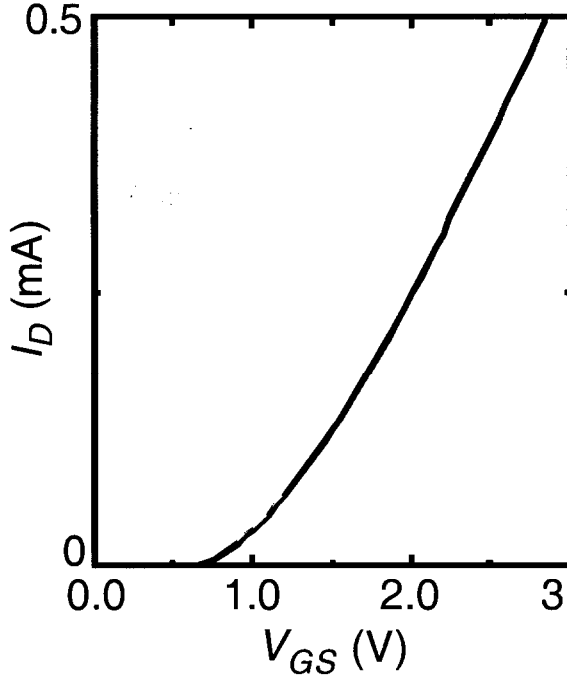
(a) Velocity saturation



(b) Mobility degradation

Velocity Saturation (2)





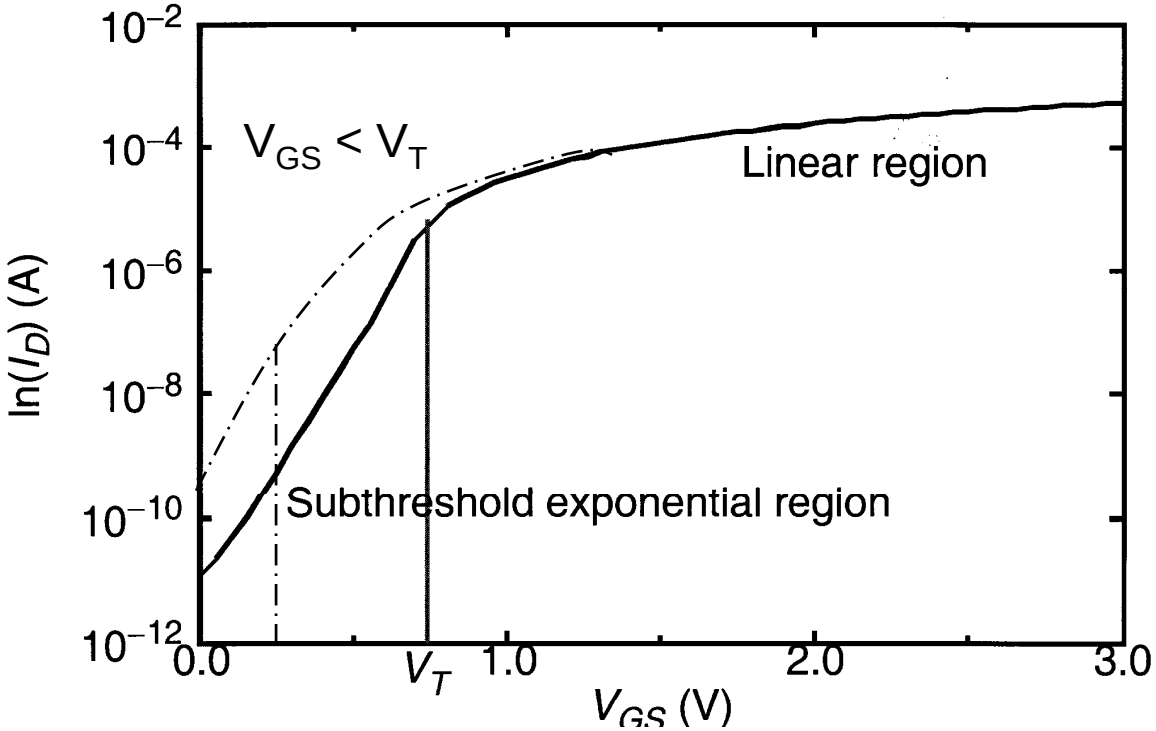
$$I_{DSAT} = v_{SAT} C_{ox} W (V_{GS} - V_{DSAT} - V_T)$$

Linear Dependence on V_{GS}

independent on L ⟵ current drive cannot be improved by decreasing L

Sub-threshold Conduction

$$I_D = K \cdot e^{(V_{gs} - V_t) q / nkT} \left(1 - e^{V_{ds} q / kT}\right)$$

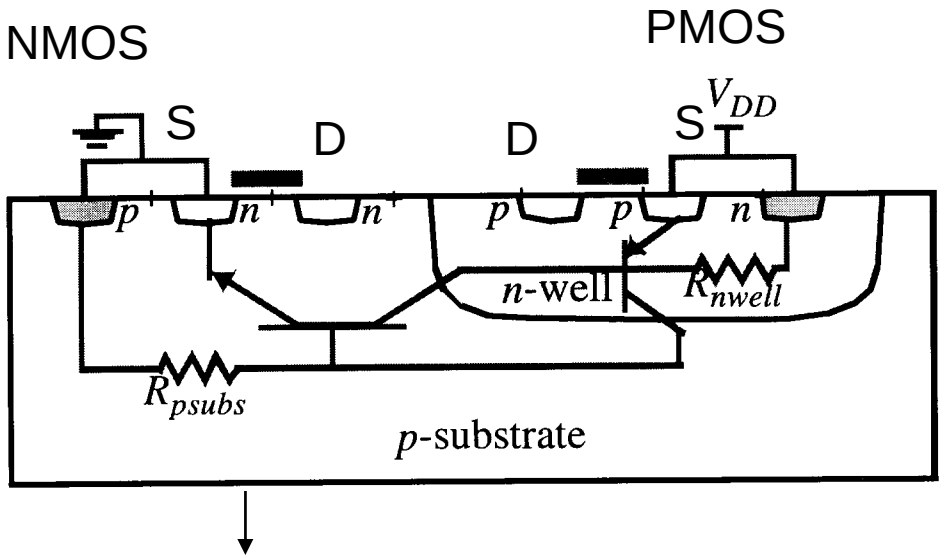


SOI has better sub-threshold leakage

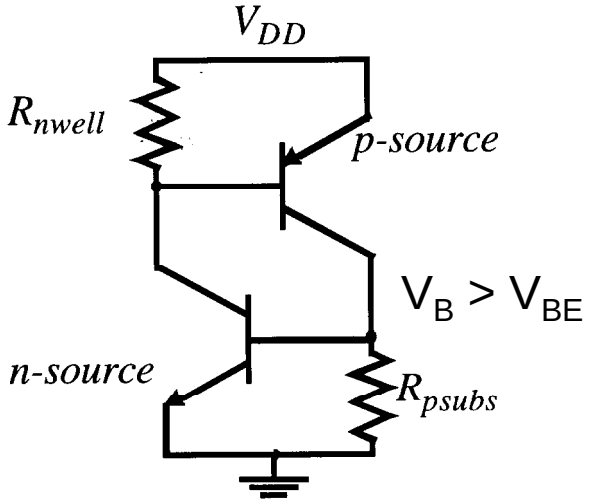
(Inverse) Rate of decrease of current : $\left(\frac{d}{dV_{GS}} \ln(I_D)\right)^{-1} = \underbrace{\frac{KT}{q} \ln 10 (1+\alpha)}$

60mV/decade At T= 300^{o}K

Latchup



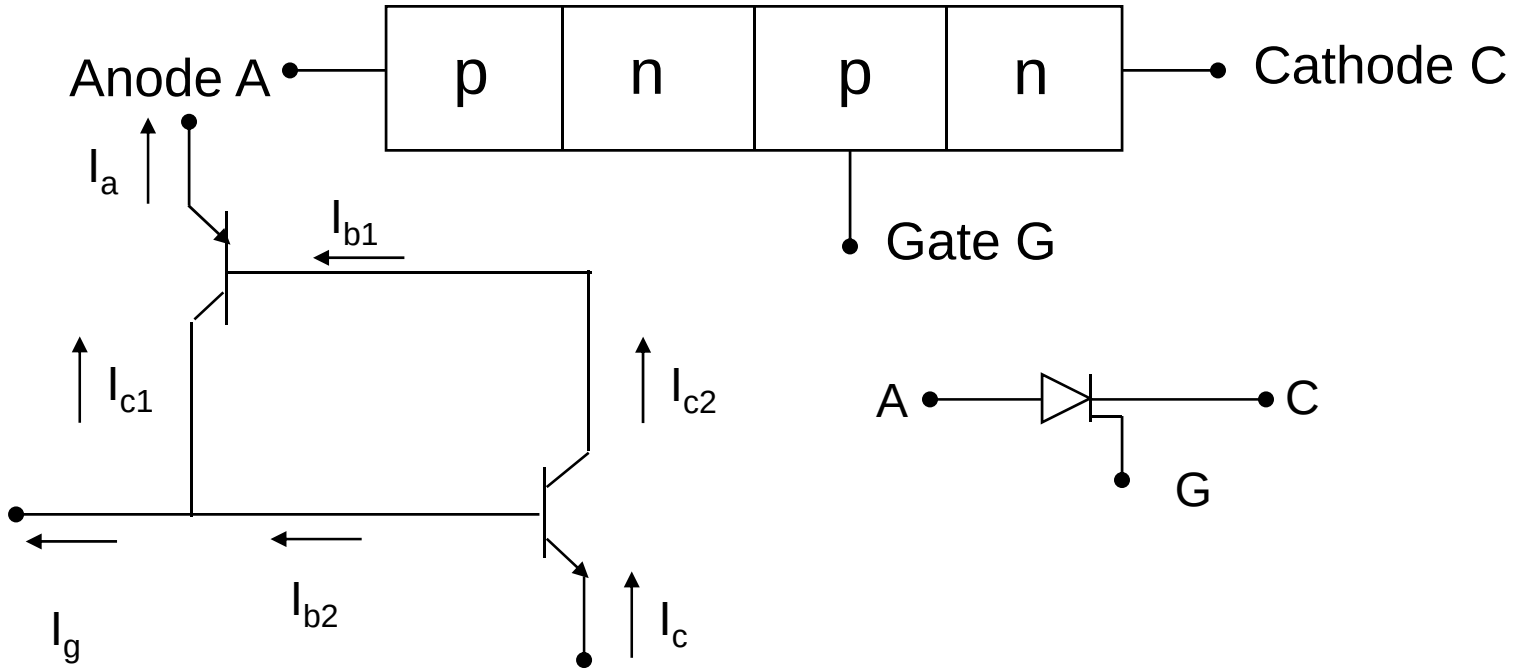
(a) Origin of latchup



(b) Equivalent circuit

Latchup

- Parasitic circuit effect
- Shorting of V_{DD} and V_{SS} lines resulting in chip self-destruction or system failure with requirements to power down
- To understand latchup consider: Silicon Controlled Rectifiers (SCRs)



Latchup - cont.

If $I_g \uparrow \Rightarrow I_{c2} \uparrow$

I_{c2} is the base current I_{b1} of the p-n-p transistor

$\because \quad I_g \uparrow \Rightarrow I_{b1} \uparrow \Rightarrow I_{c1} \uparrow \Rightarrow I_{b2} \uparrow$

(magnitude of current increases)

If the gain of the transistor are β_1 and β_2

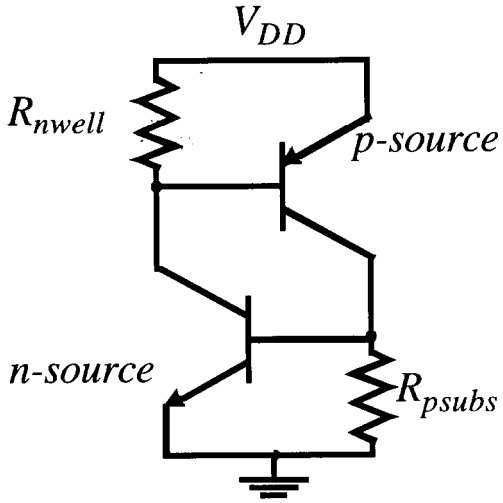
Then if $\beta_1 \beta_2 \geq 1$, the feedback action will turn device ON permanently and current will self destruct device.

Latchup Triggering

- Parasitic n-p-n & pin-p has to be triggered and holding state to be maintained
- Can be triggered by transient currents
  - Voltages during power-up
  - Radiation pulses
  - Voltages or current beyond operating range

$$I_{ntrigger} \approx \frac{V_{pnp-on}}{\alpha_{npn}.R_{well}}$$

Lateral triggering



α_{npn} : Common base gain of n-p-n transistor

Similarly, vertical triggering → due to the voltage drop across $R_{substrate}$ as current is injected into the emitter

Latchup Triggering - cont.

- Triggering occurs due to (mainly) I/O circuits where internal voltages meet external world and large currents can flow
  - When NMOS experiences undershoot by more than 0.7V, the drain is forward biased, which initiates latchup
  - When PMOS experiences overshoot by more than 0.7V, the drain is forward biased, which initiates latchup

Latchup Prevention

Analysis of the circuit shows that for latchup to occur the following inequality has to be true

$$\beta_{npn}\beta_{pnp} > 1 + \frac{(\beta_{npn}+1)(I_{Rsub} + I_{Rwell}.\beta_{pnp})}{I_{DD} - I_{Rsub}}$$

where $I_{Rsub} = \frac{V_{benpn}}{R_{sub}}$

$I_{Rwell} = \frac{V_{bepnp}}{R_{well}}$

$I_{DD} =$ total supply current

The feedback current flowing into n-p-n base is collector current offset by I_{Rsub}. To cause the feedback, this current must be greater than initial n-p-n base current, I_b.

Prevention of latchup

- Reduce the resistor values (substrate & well) and reduce the gain of parasitic transistors
- Latchup resistant CMOS process
- Layout techniques

Process option
- that reduces gain of parasitic transistors

- Si starting material with a thin epitaxial layer on highly doped Substrate
  - decreases substrate resistance
  - provide a sink for collector current of vertical p-n-p transistor
- as epi layer is thinned latch-up improves
- retrograde well structure
  - highly doped area at the bottom of the well
  - top lightly doped
  - reduces well-resistance deep in the well without deteriorating performance of transistors

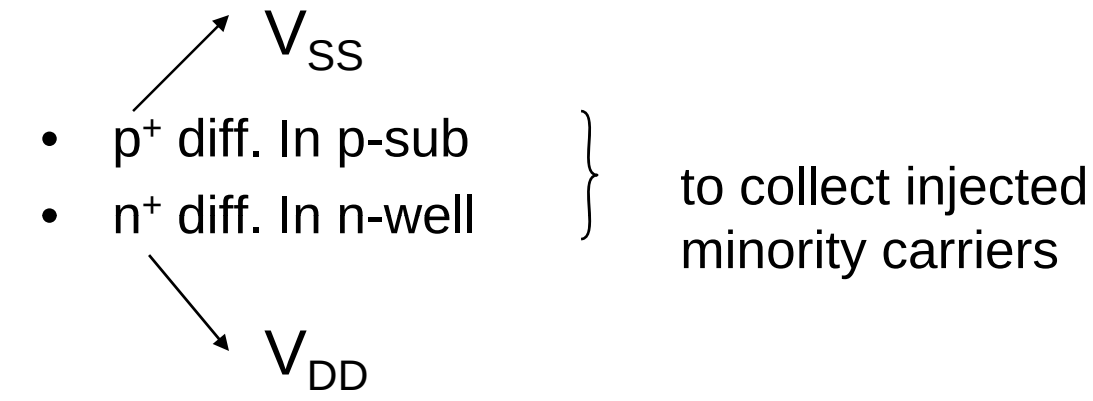
How about β_{npn} or β_{pnp}?

- Hard to reduce
- For 1 μ n-well process

$$\beta_{pnp} \sim 10 - 100$$

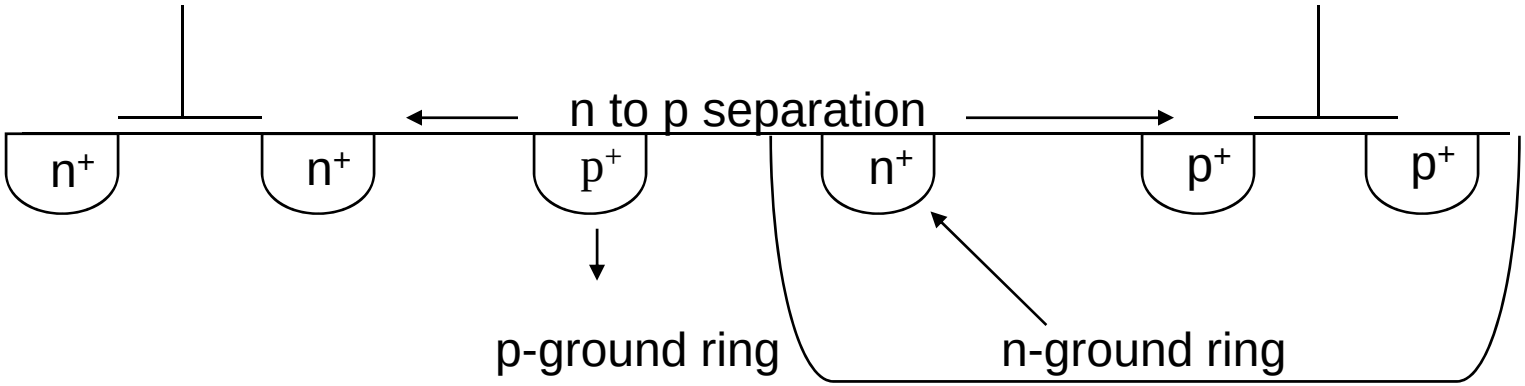
$$\beta_{npn} \sim 2 - 5$$

Guard Ring


V_{SS}
• p^+ diff. In p-sub
• n^+ diff. In n-well
V_{DD}
to collect injected
minority carriers

I/O Latchup Prevention

- Reduce β
  - use guard rings → act as dummy → collect minority currents and prevent minority carriers from being injected into respected bases
  - area expensive
  - only used in special space-borne applications where radiation is important
  - mainly used in I/O circuits only
- I/O Rules
  - separate (physically) n and p transistors
  - p+ guard rings connected to V_{ss} around n-transistors
  - n+ guard rings connected to V_{DD} around p-transistors



Latchup Prevention Techniques

- Every well must have a substrate contact of the appropriate type
- Substrate contact directly to metal to Supply pad (no diffusion or poly underpasses in the supply rails)
- Substrate contact as close to Source reduces R_{well} and R_{sub}
  - Conservative rule: one supply contact for every supply connection
  - Less conservative: a substrate contact for every 5-10 transistors or every 25 to 100
- Layout n-transistors with packing of n-devices towards V_{ss} & similarly for p-devices (V_{DD})
  - avoid convoluted structures that intertwine n- and p-devices

Spice Models

- Level 1: Long Channel Equations - Very Simple
- Level 2: Physical Model - Includes Velocity Saturation and Threshold Variations
- Level 3: Semi-Emperical - Based on curve fitting to measured devices
- Level 4 (BSIM): Emperical-Simple and Popular

Main MOS Spice Parameters

Parameter Name	Symbol	SPICE Name	Units	Default Value
SPICE Model Index		LEVEL	-	1
Zero-Bias Threshold Voltage	VT0	VT0	V	0
Process Transconductance	k'	KP	A/V2	2.E-5
Body-Bias Parameter	g	GAMMA	V0.5	0
Channel Modulation	l	LAMBDA	1/V	0
Oxide Thickness	tox	TOX	m	1.0E-7
Lateral Diffusion	xd	LD	m	0
Metallurgical Junction Depth	xj	XJ	m	0
Surface Inversion Potential	2|fF|	PHI	V	0.6
Substrate Doping	NA,ND	NSUB	cm-3	0
Surface State Density	Qss/q	NSS	cm-3	0
Fast Surface State Density		NFS	cm-3	0
Total Channel Charge Coefficient		NEFF	-	1
Type of Gate Material		TPG	-	1
Surface Mobility	m0	U0	cm2/V-sec	600
Maximum Drift Velocity	umax	VMAX	m/s	0
Mobility Critical Field	xcrit	UCRIT	V/cm	1.0E4
Critical Field Exponent in Mobility Degradation		UEXP	-	0
Transverse Field Exponent (mobility)		UTRA	-	0

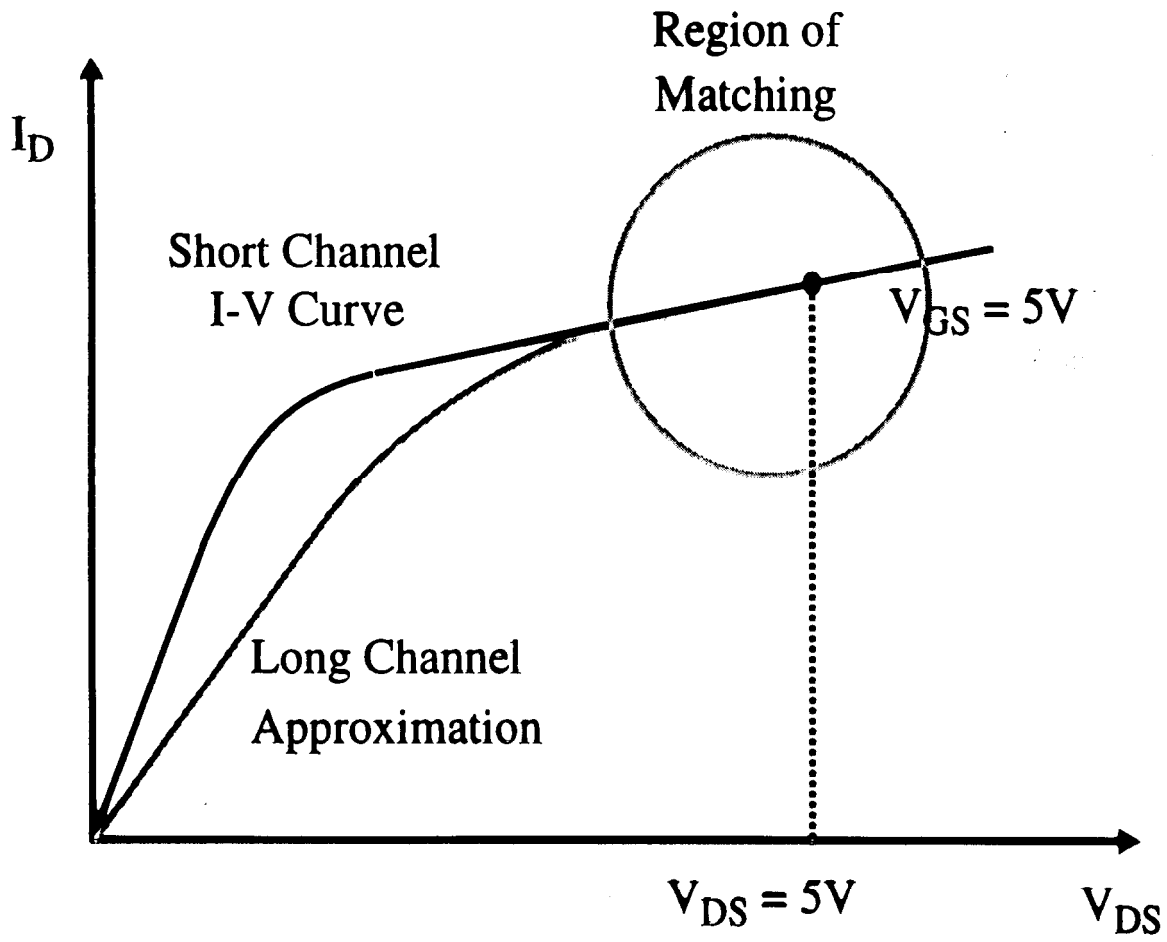
SPICE Parameters for Parasitics

Parameter Name	Symbol	SPICE Name	Units	Default Value
Source resistance	R_S	RS	Ω	0
Drain resistance	R_D	RD	Ω	0
Sheet resistance (Source/Drain)	R_o	RSH	$\Omega/_o$	0
Zero Bias Bulk Junction Cap	C_{j0}	CJ	F/m^2	0
Bulk Junction Grading Coeff.	m	MJ	-	0.5
Zero Bias Side Wall Junction Cap	C_{jsw0}	CJSW	F/m	0
Side Wall Grading Coeff.	m_{sw}	MJSW	-	0.3
Gate-Bulk Overlap Capacitance	C_{gbO}	CGBO	F/m	0
Gate-Source Overlap Capacitance	C_{gsO}	CGSO	F/m	0
Gate-Drain Overlap Capacitance	C_{gdO}	CGDO	F/m	0
Bulk Junction Leakage Current	I_S	IS	A	0
Bulk Junction Leakage Current Density	J_S	JS	A/m^2	1E-8
Bulk Junction Potential	ϕ_0	PB	V	0.8

SPICE Transistor Parameters

Parameter Name	Symbol	SPICE Name	Units	Default Value
Drawn Length	L	L	m	-
Effective Width	W	W	m	-
Source Area	AREA	AS	m2	0
Drain Area	AREA	AD	m2	0
Source Perimeter	PERIM	PS	m	0
Drain Perimeter	PERIM	PD	m	0
Squares of Source Diffusion		NRS	-	1
Squares of Drain Diffusion		NRD	-	1

Matching Manual and SPICE Models



Technology Evolution

Year of Introduction	1994	1997	2000	2003	2006	2009
Channel length (μm)	0.4	0.3	0.25	0.18	0.13	0.1
Gate oxide (nm)	12	7	6	4.5	4	4
V_{DD} (V)	3.3	2.2	2.2	1.5	1.5	1.5
V_T (V)	0.7	0.7	0.7	0.6	0.6	0.6
NMOS I_{Dsat} (mA/μm) (@ $V_{GS} = V_{DD}$)	0.35	0.27	0.31	0.21	0.29	0.33
PMOS I_{Dsat} (mA/μm) (@ $V_{GS} = V_{DD}$)	0.16	0.11	0.14	0.09	0.13	0.16