CPE/EE 427, CPE 527
VLSI Design I
MOS Transistor Theory

Department of Electrical and Computer Engineering
University of Alabama in Huntsville

Aleksandar Milenkovic

Outline

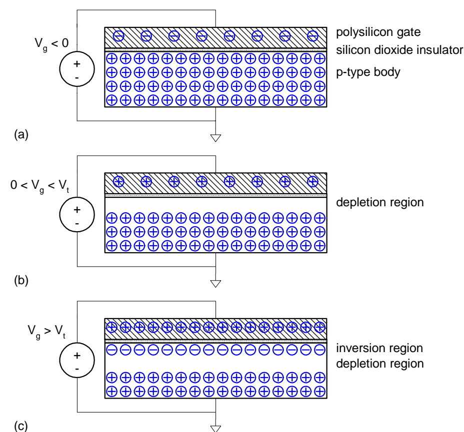
- Introduction
- MOS Capacitor
- nMOS I-V Characteristics
- pMOS I-V Characteristics
- Gate and Diffusion Capacitance
- Non-Ideal IV Effects

Introduction

- So far, we have treated transistors as ideal switches
- An ON transistor passes a finite amount of current
  - Depends on terminal voltages
  - Derive current-voltage (I-V) relationships
- Transistor gate, source, drain all have capacitance
  - $I = C (\Delta V/\Delta t) \rightarrow \Delta t = (C/I) \Delta V$
  - Capacitance and current determine speed
- Also explore what a "degraded level" really means

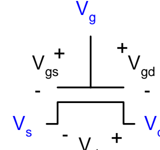
MOS Capacitor

- Gate and body form MOS capacitor
- Operating modes
  - Accumulation
  - Depletion
  - Inversion



Terminal Voltages

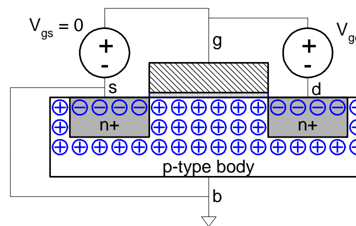
- Mode of operation depends on V_g, V_d, V_s
  - $V_{gs} = V_g - V_s$
  - $V_{gd} = V_g - V_d$
  - $V_{ds} = V_d - V_s = V_{gs} - V_{gd}$



- Source and drain are symmetric diffusion terminals
  - By convention, source is terminal at lower voltage
  - Hence $V_{ds} \geq 0$
- nMOS body is grounded. First assume source is 0 too.
- Three regions of operation
  - *Cutoff*
  - *Linear*
  - *Saturation*

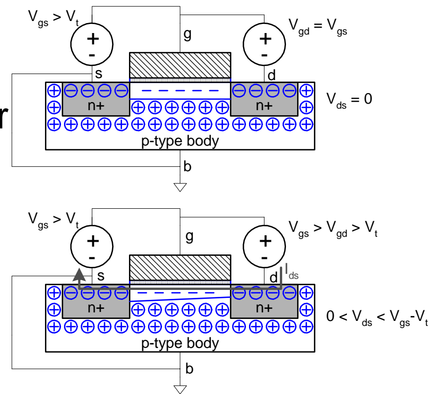
nMOS Cutoff

- No channel
- $I_{ds} = 0$



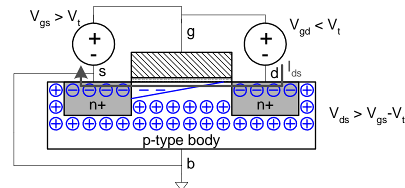
nMOS Linear

- Channel forms
- Current flows from d to s
  - e^- from s to d
- I_{ds} increases with V_{ds}
- Similar to linear resistor



nMOS Saturation

- Channel pinches off
- I_{ds} independent of V_{ds}
- We say current saturates
- Similar to current source

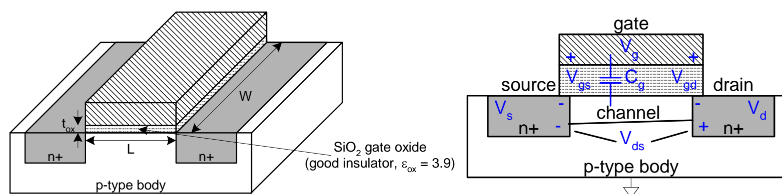


I-V Characteristics

- In Linear region, I_{ds} depends on
  - How much charge is in the channel?
  - How fast is the charge moving?

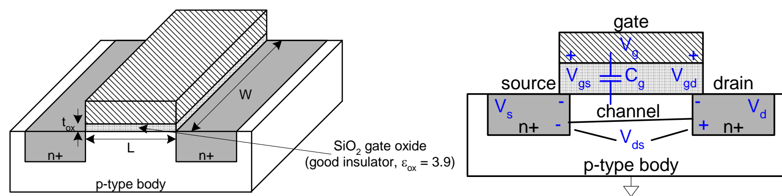
Channel Charge

- MOS structure looks like parallel plate capacitor while operating in inversion
  - Gate – oxide – channel
- $Q_{channel}$ =



Channel Charge

- MOS structure looks like parallel plate capacitor while operating in inversion
  - Gate – oxide – channel
- $Q_{channel} = CV$
- $C =$

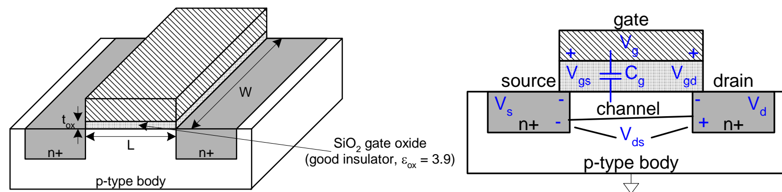


Channel Charge

- MOS structure looks like parallel plate capacitor while operating in inversion
  - Gate – oxide – channel
- $Q_{channel} = CV$
- $C = C_g = \varepsilon_{ox}WL/t_{ox} = C_{ox}WL$

$C_{ox} = \varepsilon_{ox} / t_{ox}$

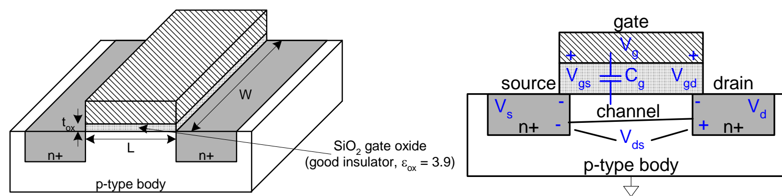
- $V =$



Channel Charge

- MOS structure looks like parallel plate capacitor while operating in inversion
  - Gate – oxide – channel
- $Q_{channel} = CV$
- $C = C_g = \varepsilon_{ox}WL/t_{ox} = C_{ox}WL$
- $V = V_{gc} – V_t = (V_{gs} – V_{ds}/2) – V_t$

$C_{ox} = \varepsilon_{ox} / t_{ox}$



Carrier velocity

- Charge is carried by e-
- Carrier velocity *v* proportional to lateral E-field between source and drain
- *v* =

Carrier velocity

- Charge is carried by e-
- Carrier velocity *v* proportional to lateral E-field between source and drain
- $v = \mu E$ μ called mobility
- E =

Carrier velocity

- Charge is carried by e-
- Carrier velocity *v* proportional to lateral E-field between source and drain
- $v = \mu E$ μ called mobility
- $E = V_{ds}/L$
- Time for carrier to cross channel:
  - *t* =

Carrier velocity

- Charge is carried by e-
- Carrier velocity v proportional to lateral E-field between source and drain
- $v = \mu E$ μ called mobility
- $E = V_{ds}/L$
- Time for carrier to cross channel:
  - $t = L / v$

nMOS Linear I-V

- Now we know
  - How much charge $Q_{channel}$ is in the channel
  - How much time t each carrier takes to cross

$$I_{ds} =$$

nMOS Linear I-V

- Now we know
  - How much charge $Q_{channel}$ is in the channel
  - How much time *t* each carrier takes to cross

$$I_{ds} = \frac{Q_{\text{channel}}}{t}$$

$$=$$

nMOS Linear I-V

- Now we know
  - How much charge $Q_{channel}$ is in the channel
  - How much time *t* each carrier takes to cross

$$I_{ds} = \frac{Q_{\text{channel}}}{t}$$

$$= \mu C_{\text{ox}} \frac{W}{L}\left(V_{gs} - V_t - {V_{ds}}/{2}\right) V_{ds}$$

$$= \beta\left(V_{gs} - V_t - {V_{ds}}/{2}\right) V_{ds}$$

$$\beta = \mu C_{\text{ox}} \frac{W}{L}$$

nMOS Saturation I-V

- If $V_{gd} < V_t$, channel pinches off near drain
  - When $V_{ds} > V_{dsat} = V_{gs} - V_t$
- Now drain voltage no longer increases current

$$I_{ds} =$$

nMOS Saturation I-V

- If $V_{gd} < V_t$, channel pinches off near drain
  - When $V_{ds} > V_{dsat} = V_{gs} - V_t$
- Now drain voltage no longer increases current

$$I_{ds} = \beta\left(V_{gs} - V_t - \frac{V_{dsat}}{2}\right)V_{dsat}$$

nMOS Saturation I-V

- If $V_{gd} < V_t$, channel pinches off near drain
  - When $V_{ds} > V_{dsat} = V_{gs} - V_t$
- Now drain voltage no longer increases current

$$I_{ds} = \beta\left(V_{gs} - V_t - \frac{V_{dsat}}{2}\right)V_{dsat}$$

$$= \frac{\beta}{2}\left(V_{gs} - V_t\right)^2$$

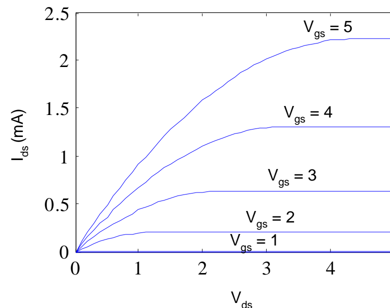
nMOS I-V Summary

- *Shockley* 1st order transistor models

$$I_{ds} = \begin{cases} 0 & V_{gs} < V_t & \text{cutoff} \\ \beta\left(V_{gs} - V_t - \frac{V_{ds}}{2}\right)V_{ds} & V_{ds} < V_{dsat} & \text{linear} \\ \frac{\beta}{2}\left(V_{gs} - V_t\right)^2 & V_{ds} > V_{dsat} & \text{saturation} \end{cases}$$

Example

- For a 0.6 μm process
  - From AMI Semiconductor
  - t_{ox} = 100 Å
  - μ = 350 $cm^2/V*s$
  - V_t = 0.7 V
- Plot I_{ds} vs. V_{ds}
  - V_{gs} = 0, 1, 2, 3, 4, 5
  - Use W/L = 4/2 λ



$$\beta = \mu C_{ox} \frac{W}{L} = (350)\left(\frac{3.9 \bullet 8.85 \cdot 10^{-14}}{100 \cdot 10^{-8}}\right)\left(\frac{W}{L}\right) = 120 \frac{W}{L} \mu A / V^2$$

9/11/2006 VLSI Design I; A. Milenkovic 25

pMOS I-V

- All dopings and voltages are inverted for pMOS
- Mobility μ_p is determined by holes
  - Typically 2-3x lower than that of electrons μ_n
  - 120 $cm^2/V*s$ in AMI 0.6 μm process
- Thus pMOS must be wider to provide same current
  - In this class, assume $\mu_n / \mu_p = 2$
  - *** plot I-V here

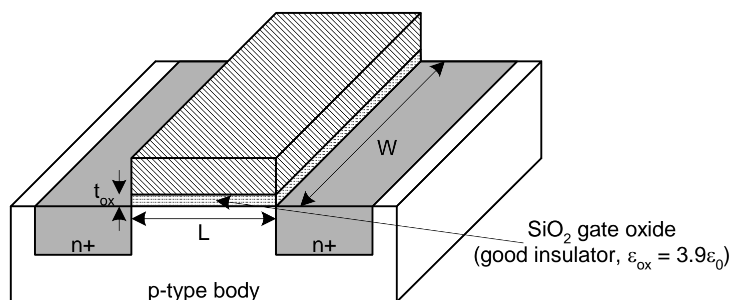
9/11/2006 VLSI Design I; A. Milenkovic 26

Capacitance

- Any two conductors separated by an insulator have capacitance
- Gate to channel capacitor is very important
  - Creates channel charge necessary for operation
- Source and drain have capacitance to body
  - Across reverse-biased diodes
  - Called diffusion capacitance because it is associated with source/drain diffusion

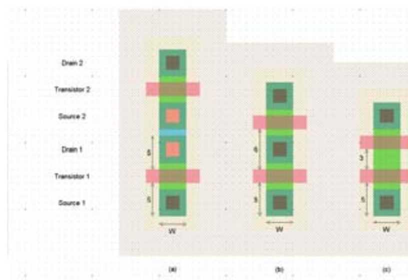
Gate Capacitance (simple model)

- Approximate channel as connected to source
- $C_{gs} = \varepsilon_{ox}WL/t_{ox} = C_{ox}WL = C_{permicron}W$
- $C_{permicron}$ is typically about 2 fF/μm

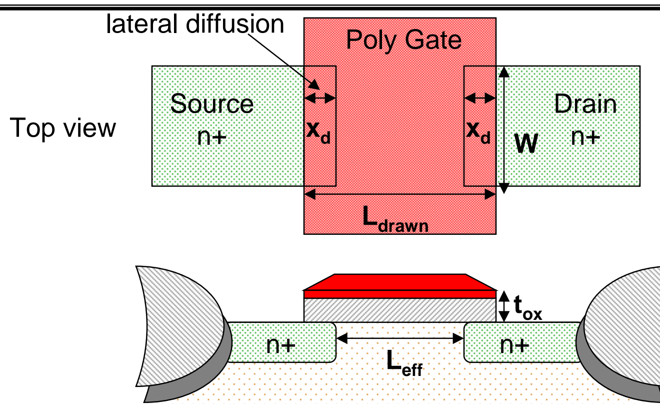


Diffusion Capacitance (simple model)

- C_{sb}, C_{db}
- Undesirable, called *parasitic* capacitance
- Capacitance depends on area and perimeter
  - Use small diffusion nodes
  - Comparable to C_g for contacted diff
  - ½ C_g for uncontacted
  - Varies with process



MOS Structure Capacitances

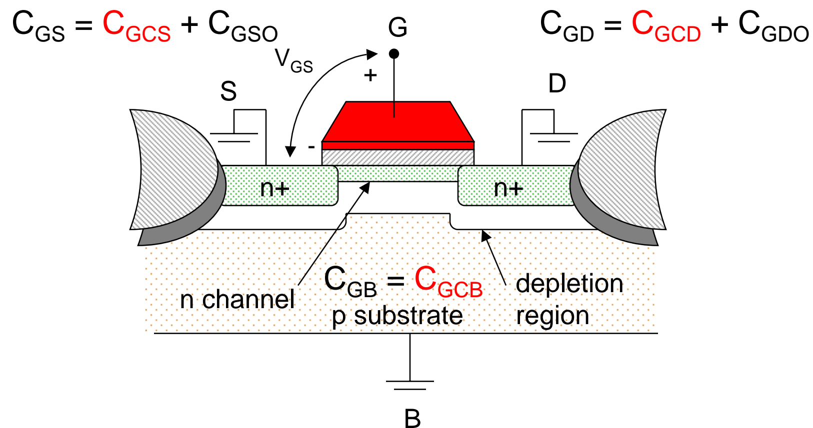


Overlap capacitance (linear)

$$C_{GSO} = C_{GDO} = C_{ox}\, x_d\, W = C_o\, W$$

MOS Channel Capacitances

- The gate-to-channel capacitance depends upon the operating region and the terminal voltages



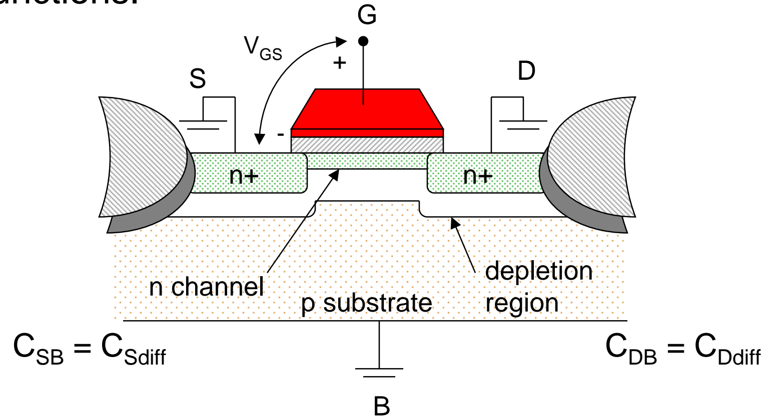
Average Distribution of Channel Capacitance

Operation Region	C_{GCB}	C_{GCS}	C_{GCD}	C_{GC}	C_G
Cutoff	$C_{ox}WL$	0	0	$C_{ox}WL$	$C_{ox}WL + 2C_oW$
Resistive	0	$C_{ox}WL/2$	$C_{ox}WL/2$	$C_{ox}WL$	$C_{ox}WL + 2C_oW$
Saturation	0	$(2/3)C_{ox}WL$	0	$(2/3)C_{ox}WL$	$(2/3)C_{ox}WL + 2C_oW$

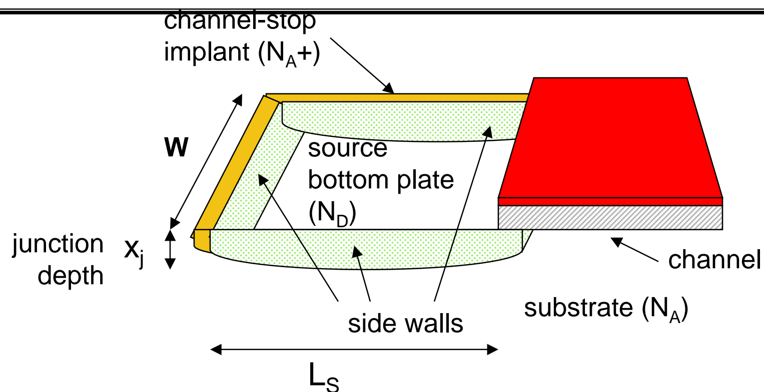
- Channel capacitance components are nonlinear and vary with operating voltage
- Most important regions are cutoff and saturation since that is where the device spends most of its time

MOS Diffusion Capacitances

- The junction (or diffusion) capacitance is from the reverse-biased source-body and drain-body pn-junctions.



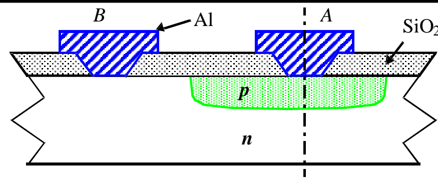
Source Junction View



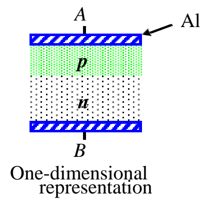
$$C_{diff} = C_{bp} + C_{sw} = C_j \, AREA + C_{jsw} \, PERIMETER$$

$$= C_j L_S W + C_{jsw} (2L_S + W)$$

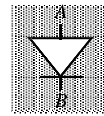
Recap: The Diode



Cross-section of *pn*-junction in an IC process



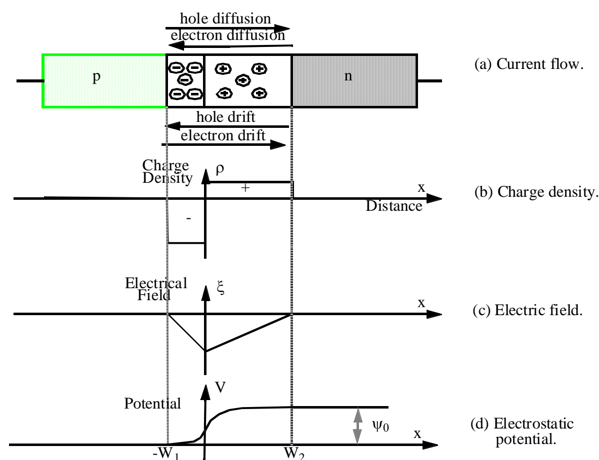
One-dimensional representation



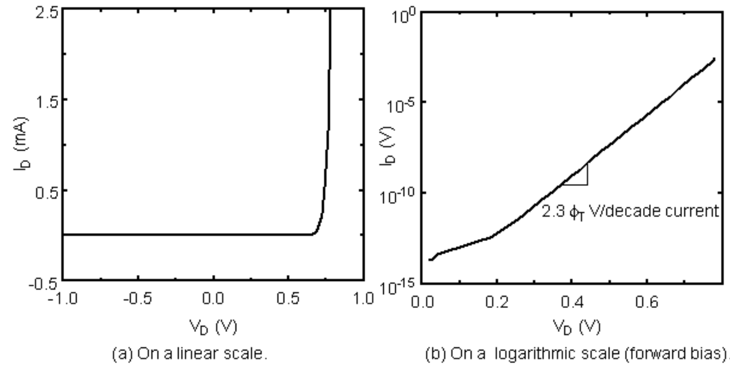
diode symbol

Mostly occurring as parasitic element in Digital ICs

Recap: Depletion Region



Recap: Diode Current



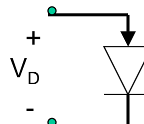
(a) On a linear scale. (b) On a logarithmic scale (forward bias).

$$I_D = I_S\left(e^{V_D/\phi_T} - 1\right)$$

Recap: Ideal Diode Equation

- The *ideal diode equation* (for both forward and reverse-bias conditions) is

$$I_D = I_S(e^{V_D/\phi_T} - 1)$$

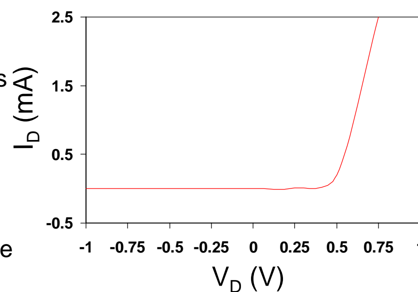


where V_D is the voltage applied to the junction

- a forward-bias lowers the potential barrier allowing carriers to flow across the diode junction
- a reverse-bias raises the potential barrier and the diode becomes nonconducting

$\phi_T = kT/q = 26\text{mV}$ at 300K

I_S is the saturation current of the diode



MOS Diffusion Capacitances

$C_{diff} = C_{bp} + C_{sw} = C_{jbp}\ AREA + C_{jsw}\ PERIMETER$
$= C_{jbp} L_S W + C_{jsw} (2L_S + W)$

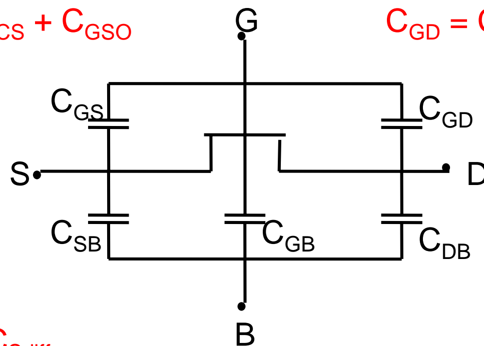
Source parasitic:
$C_{jbs} = C_J(1+V_{sb}/\psi_0)^{-Mj}$
- C_J – junction capacitance at zero bias
- Mj – junction grading coefficinet (0.5 – 0.33)
- ψ_0 - – built-in potential $\{\psi_0 = v_T \ln(N_A N_D/n_i^2)\}$

$C_{jbssw} = C_{JSW}(1+V_{sb}/\psi_0)^{-Mjsw}$
- C_{JSW} – junction capacitance at zero bias
- M_{jsw} – junction grading coefficinet (0.5 – 0.33)

MOS Capacitance Model

$C_{GS} = C_{GCS} + C_{GSO}$

$C_{GD} = C_{GCD} + C_{GDO}$



$C_{SB} = C_{Sdiff}$

$C_{DB} = C_{Ddiff}$

$C_{GB} = C_{GCB}$

Transistor Capacitance Values for 0.25μ

Example: For an NMOS with L = 0.24 μm, W = 0.36 μm, $L_D = L_S =$ 0.625 μm

$C_{GSO} = C_{GDO} = C_{ox}\, x_d\, W = C_o\, W =$

$C_{GC} = C_{ox}\, WL =$

so $C_{gate_cap} = C_{ox}WL + 2C_oW =$

$C_{bp} = C_j\, L_S\, W =$

$C_{sw} = C_{jsw}\, (2L_S + W) =$

so $C_{diffusion_cap} =$

	C_{ox} (fF/μm²)	C_o (fF/μm)	C_j (fF/μm²)	m_j	ϕ_b (V)	C_{jsw} (fF/μm)	m_{jsw}	ϕ_{bsw} (V)
NMOS	6	0.31	2	0.5	0.9	0.28	0.44	0.9
PMOS	6	0.27	1.9	0.48	0.9	0.22	0.32	0.9

Transistor Capacitance Values for 0.25μ

Example: For an NMOS with L = 0.24 μm, W = 0.36 μm, $L_D = L_S = 0.625$ μm

$C_{GSO} = C_{GDO} = C_{ox}\, x_d\, W = C_o\, W = 0.11$ fF

$C_{GC} = C_{ox}\, WL = 0.52$ fF

so $C_{gate_cap} = C_{ox}WL + 2C_oW = 0.74$ fF

$C_{bp} = C_j\, L_S\, W = 0.45$ fF

$C_{sw} = C_{jsw}\, (2L_S + W) = 0.45$ fF

so $C_{diffusion_cap} = 0.90$ fF

	C_{ox} (fF/μm²)	C_o (fF/μm)	C_j (fF/μm²)	m_j	ϕ_b (V)	C_{jsw} (fF/μm)	m_{jsw}	ϕ_{bsw} (V)
NMOS	6	0.31	2	0.5	0.9	0.28	0.44	0.9
PMOS	6	0.27	1.9	0.48	0.9	0.22	0.32	0.9

Non-Ideal I-V Effects

- I_{ds}(sat) increases less than quadratically with increasing V_{gs}
  - Velocity saturation & mobility degradation
- I_{ds}(sat) increases slightly with V_{ds}
  - Channel length modulation
- V_T is influenced by the V_{sb}
  - Body effect
- There is current flow in nominally OFF transistors
  - Subthreshold conduction (junction leakage, tunnel)

$$I_{ds} = \begin{cases} 0 & V_{gs} < V_t \quad \text{cutoff} \\ \beta\left(V_{gs} - V_t - \frac{V_{ds}}{2}\right)V_{ds} & V_{ds} < V_{dsat} \quad \text{linear} \\ \frac{\beta}{2}\left(V_{gs} - V_t\right)^2 & V_{ds} > V_{dsat} \quad \text{saturation} \end{cases}$$

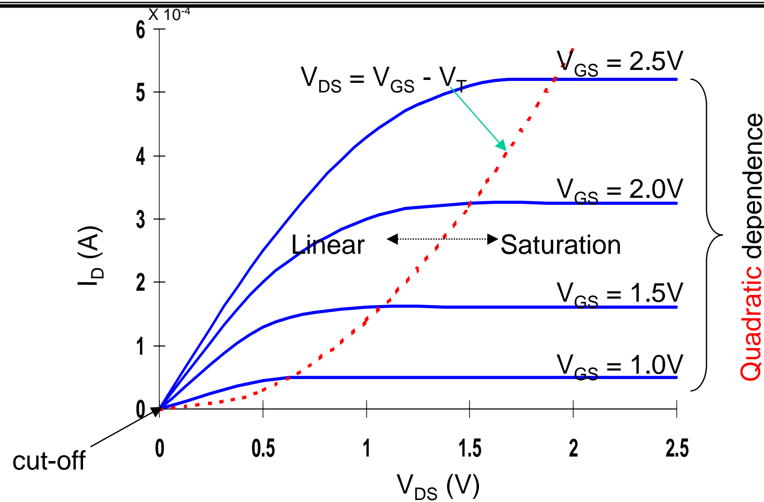
9/11/2006 VLSI Design I; A. Milenkovic 43

Current Determinates

- For a fixed V_{DS} and V_{GS} (> V_T), I_{DS} is a function of
  - the distance between the source and drain – L
  - the channel width – W
  - the threshold voltage – V_T
  - the thickness of the SiO_2 – t_{ox}
  - the dielectric of the gate insulator (SiO_2) – ε_{ox}
  - the carrier mobility
    - for nfets: $\mu_n = 500$ cm^2/V-sec
    - for pfets: $\mu_p = 180$ cm^2/V-sec

9/11/2006 VLSI Design I; A. Milenkovic 44

Long Channel I-V Plot (NMOS)

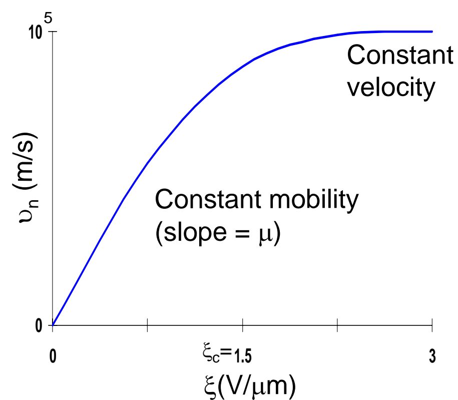


NMOS transistor, 0.25um, L_d = 10um, W/L = 1.5, V_{DD} = 2.5V, V_T = 0.4V

9/11/2006 VLSI Design I; A. Milenkovic 45

Short Channel Effects

- Behavior of short channel device mainly due to



- Velocity saturation – the velocity of the carriers saturates due to scattering (collisions suffered by the carriers)

- For an NMOS device with L of .25µm, only a couple of volts difference between D and S are needed to reach velocity saturation

9/11/2006 VLSI Design I; A. Milenkovic 46

Voltage-Current Relation: Velocity Saturation

For short channel devices

- Linear: When $V_{DS} \leq V_{GS} - V_T$

  $I_D = \kappa(V_{DS})\, k'_n\, W/L\, [(V_{GS} - V_T)V_{DS} - V_{DS}^2/2]$

  where

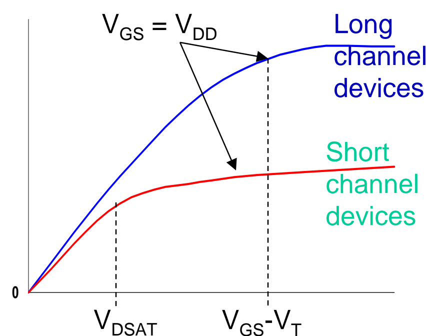
  $\kappa(V) = 1/(1 + (V/\xi_c L))$ is a measure of the degree of velocity saturation

- Saturation: When $V_{DS} = V_{DSAT} \geq V_{GS} - V_T$

  $I_{DSat} = \kappa(V_{DSAT})\, k'_n\, W/L\, [(V_{GS} - V_T)V_{DSAT} - V_{DSAT}^2/2]$

9/11/2006 VLSI Design I; A. Milenkovic 47

Velocity Saturation Effects



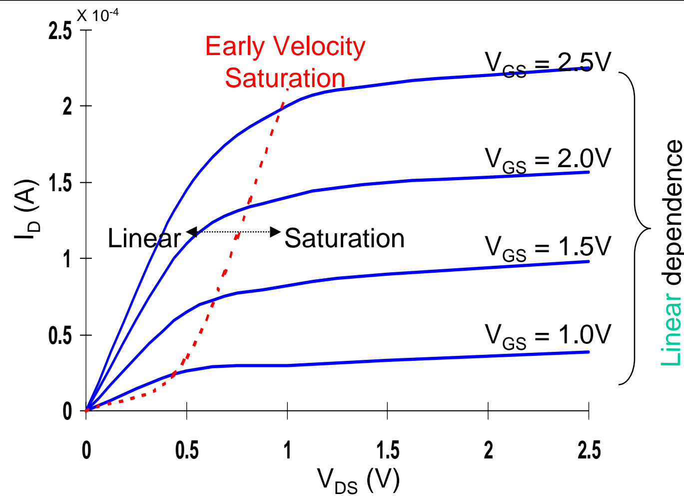
For short channel devices and large enough $V_{GS} - V_T$

- $V_{DSAT} < V_{GS} - V_T$ so the device enters saturation before V_{DS} reaches $V_{GS} - V_T$ and operates more often in saturation

- I_{DSAT} has a linear dependence wrt V_{GS} so a reduced amount of current is delivered for a given control voltage

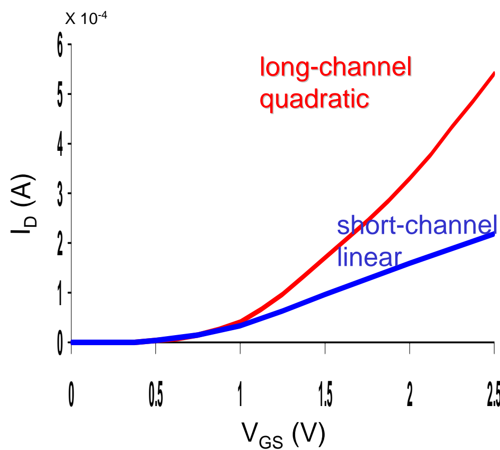
9/11/2006 VLSI Design I; A. Milenkovic 48

Short Channel I-V Plot (NMOS)



NMOS transistor, 0.25um, L_d = 0.25um, W/L = 1.5, V_{DD} = 2.5V, V_T = 0.4V

MOS I_D-V_{GS} Characteristics

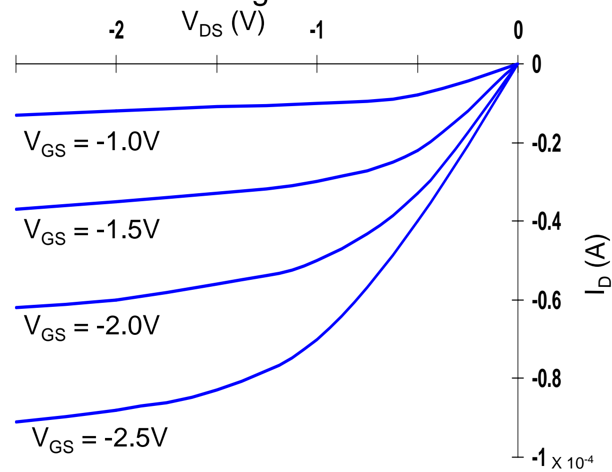


(for V_{DS} = 2.5V, W/L = 1.5)

- Linear (short-channel) versus quadratic (long-channel) dependence of I_D on V_{GS} in saturation
- Velocity-saturation causes the short-channel device to saturate at substantially smaller values of V_{DS} resulting in a substantial drop in current drive

Short Channel I-V Plot (PMOS)

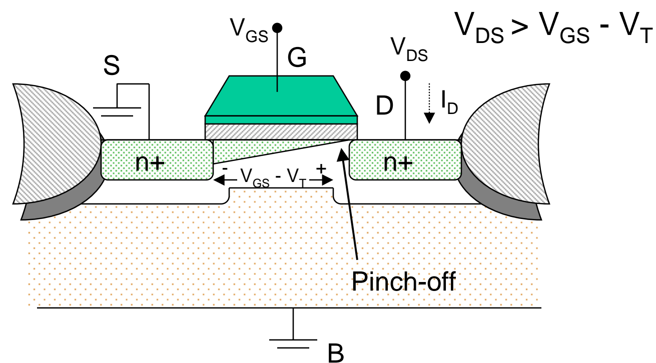
- All polarities of all voltages and currents are reversed



PMOS transistor, 0.25um, L_d = 0.25um, W/L = 1.5, V_{DD} = 2.5V, V_T = -0.4V

Transistor in Saturation Mode

Assuming $V_{GS} > V_T$



The current remains constant (saturates).

Voltage-Current Relation: Saturation Mode

For long channel devices

- When $V_{DS} \geq V_{GS} - V_T$

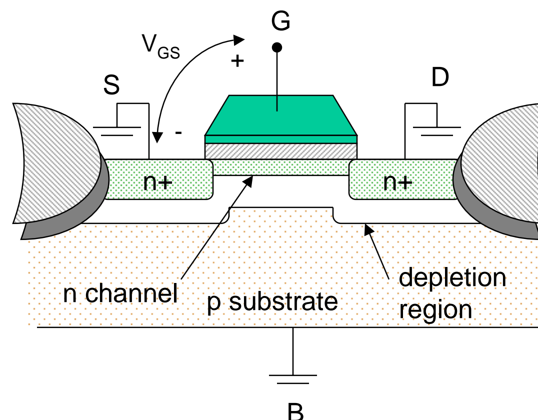
  $$I_D' = k'_n/2 \; W/L \; [(V_{GS} - V_T)^{2}]$$

  since the voltage difference over the induced channel (from the pinch-off point to the source) remains fixed at $V_{GS} - V_T$
- However, the effective length of the conductive channel is modulated by the applied V_{DS}, so

  $$I_D = I_D' (1 + \lambda V_{DS})$$

  where λ is the channel-length modulation (varies with the inverse of the channel length)

Threshold Voltage Concept



The value of V_{GS} where strong inversion occurs is called the threshold voltage, V_T

The Threshold Voltage

$$V_T = V_{T0} + \gamma(\sqrt{|-2\phi_F + V_{SB}|} - \sqrt{|-2\phi_F|})$$

where

V_{T0} is the threshold voltage at $V_{SB} = 0$ and is mostly a function of the manufacturing process

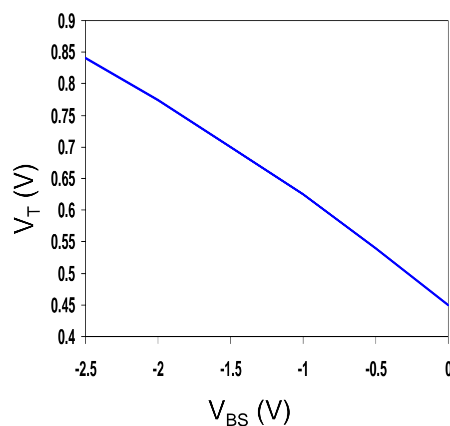
- Difference in work-function between gate and substrate material, oxide thickness, Fermi voltage, charge of impurities trapped at the surface, dosage of implanted ions, etc.

V_{SB} is the source-bulk voltage

$\phi_F = -\phi_T \ln(N_A/n_i)$ is the Fermi potential ($\phi_T = kT/q = 26mV$ at 300K is the thermal voltage; N_A is the acceptor ion concentration; $n_i \approx 1.5 \times 10^{10}$ cm^{-3} at 300K is the intrinsic carrier concentration in pure silicon)

$\gamma = \sqrt{(2q\varepsilon_{si}N_A)}/C_{ox}$ is the body-effect coefficient (impact of changes in V_{SB}) ($\varepsilon_{si}=1.053 \times 10^{-10}$F/m is the permittivity of silicon; $C_{ox} = \varepsilon_{ox}/t_{ox}$ is the gate oxide capacitance with $\varepsilon_{ox}=3.5 \times 10^{-11}$F/m)

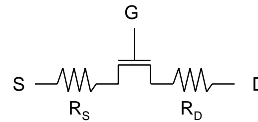
The Body Effect



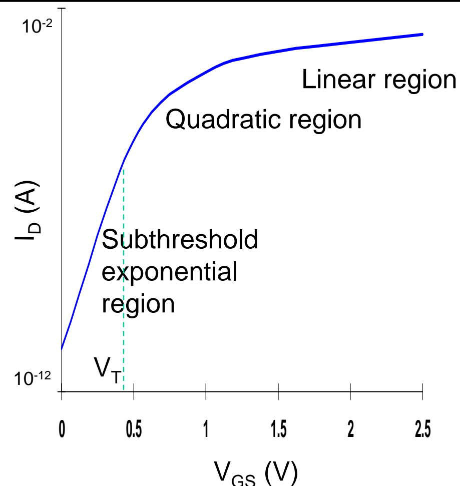
- V_{SB} is the substrate bias voltage (normally positive for n-channel devices with the body tied to ground)
- A negative bias causes V_T to increase from 0.45V to 0.85V

Other (Submicon) MOS Transistor Concerns

- Velocity saturation
- Subthreshold conduction
  - Transistor is already partially conducting for voltages below VT
- Threshold variations
  - In long-channel devices, the threshold is a function of the length (for low VDS)
  - In short-channel devices, there is a drain-induced threshold barrier lowering at the upper end of the VDS range (for low L)
- Parasitic resistances
  - resistances associated with the source and drain contacts
- Latch-up



Subthreshold Conductance



$I_D \sim I_S\ e^{(qV_{GS}/nkT)}$ where $n \geq 1$

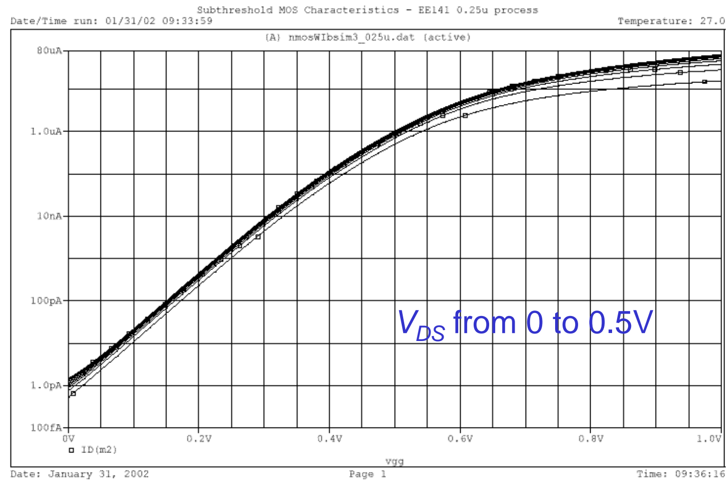
- Transition from ON to OFF is gradual (decays exponentially)
- Current roll-off (slope factor) is also affected by increase in temperature

$S = n\ (kT/q)\ \ln(10)$
(typical values 60 to 100 mV/decade)

- Has repercussions in dynamic circuits and for power consumption

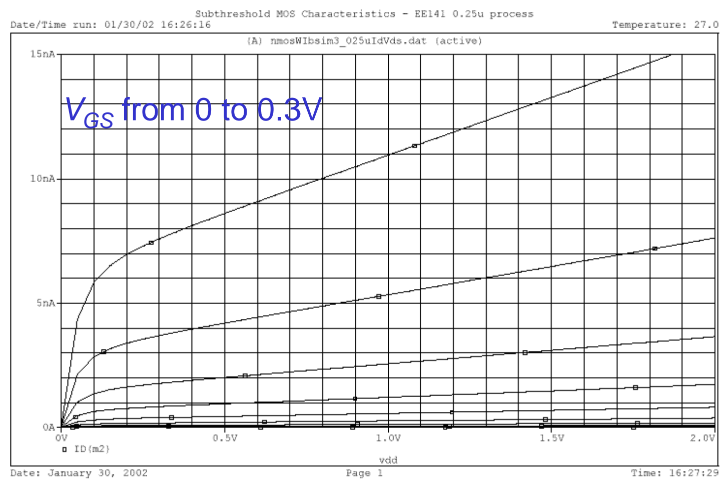
Subthreshold I_D vs V_{GS}

$$I_D = I_S\, e^{(qV_{GS}/nkT)}\,(1 - e^{-(qV_{DS}/kT)})(1 + \lambda V_{DS})$$

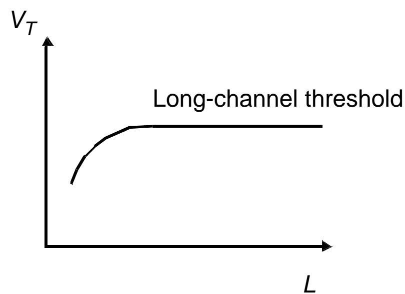


Subthreshold I_D vs V_{DS}

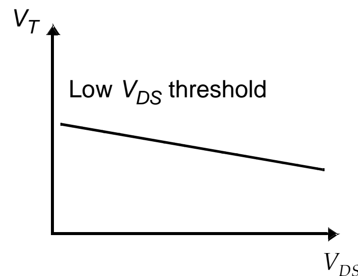
$$I_D = I_S\, e^{(qV_{GS}/nkT)}\,(1 - e^{-(qV_{DS}/kT)})(1 + \lambda V_{DS})$$



Threshold Variations



Threshold as a function of the length (for low V_{DS})



Drain-induced barrier lowering (for low L)

Voltage-Current Relation: Linear Mode

For long-channel devices (L > 0.25 micron)

- When $V_{DS} \le V_{GS} - V_T$

$$I_D = k'_n \, W/L \, [(V_{GS} - V_T)V_{DS} - V_{DS}^2/2]$$

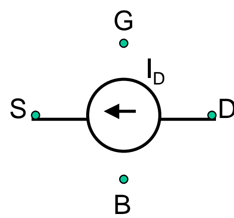
  where

  $k'_n = \mu_n C_{ox} = \mu_n \varepsilon_{ox}/t_{ox}$ = is the process transconductance parameter (μ_n is the carrier mobility (m^2/Vsec))

  $k_n = k'_n$ W/L is the gain factor of the device

- For small V_{DS}, there is a linear dependence between V_{DS} and I_D, hence the name resistive or linear region

The MOS Current-Source Model



$I_D = 0$ for $V_{GS} - V_T \leq 0$

$I_D = k' W/L [(V_{GS} - V_T)V_{min} - V_{min}^2/2](1+\lambda V_{DS})$
for $V_{GS} - V_T \geq 0$

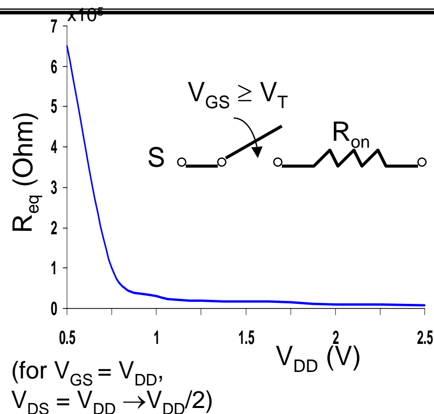
with $V_{min} = \min(V_{GS} - V_T, V_{DS}, V_{DSAT})$
and $V_{GT} = V_{GS} - V_T$

- Determined by the voltages at the four terminals and a set of five device parameters

	V_{T0}(V)	$\gamma(V^{0.5})$	V_{DSAT}(V)	k'(A/V²)	$\lambda(V^{-1})$
NMOS	0.43	0.4	0.63	115 x 10^{-6}	0.06
PMOS	-0.4	-0.4	-1	-30 x 10^{-6}	-0.1

9/11/2006 VLSI Design I; A. Milenkovic 63

The Transistor Modeled as a Switch



(for $V_{GS} = V_{DD}$, $V_{DS} = V_{DD} \rightarrow V_{DD}/2$)

Modeled as a switch with infinite off resistance and a finite on resistance, R_{on}

- Resistance inversely proportional to W/L (doubling W halves R_{on})
- For $V_{DD} >> V_T + V_{DSAT}/2$, R_{on} independent of V_{DD}
- Once V_{DD} approaches V_T, R_{on} increases dramatically

V_{DD}(V)	1	1.5	2	2.5
NMOS(kΩ)	35	19	15	13
PMOS (kΩ)	115	55	38	31

R_{on} (for W/L = 1)
For larger devices divide R_{eq} by W/L

9/11/2006 VLSI Design I; A. Milenkovic 64

Next Time: The CMOS Inverter