EASTERN WASHINGTON UNIVERSITY

MOS Transistor Theory

Slides adapted from:

N. Weste, D. Harris, CMOS VLSI Design,
© Addison-Wesley, 3/e, 2004

EASTERN WASHINGTON UNIVERSITY

Outline

- The Big Picture
- MOS Structure
- Ideal I-V Charcteristics
- MOS Capacitance Models
- Non ideal I-V Effects
- Pass transistor circuits
- Tristate Inverter
- Switch level RC Delay Models

EASTERN WASHINGTON UNIVERSITY

The Big Picture

- 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

EASTERN WASHINGTON UNIVERSITY

MOS Transistor Symbol

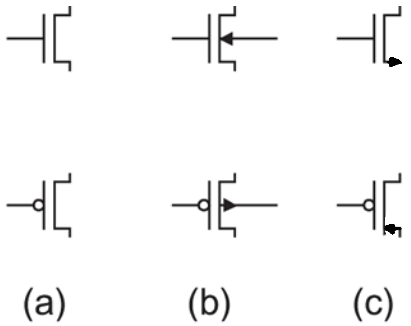


FIG 2.1 MOS transistor symbols

EASTERN WASHINGTON UNIVERSITY

MOS Structure

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

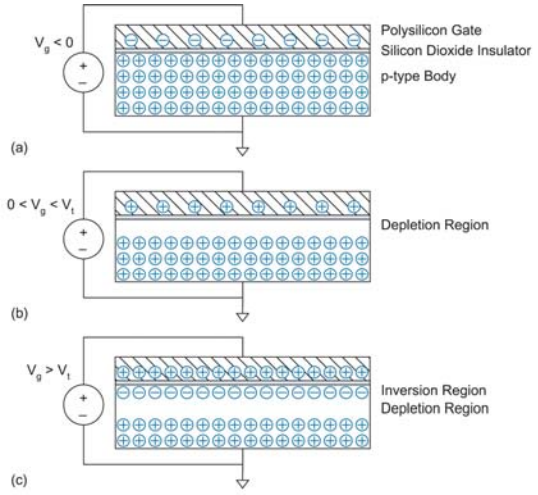
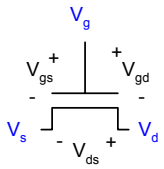


FIG 2.2 MOS structure demonstrating (a) accumulation, (b) depletion, and (c) inversion

EASTERN WASHINGTON UNIVERSITY

nMOS Transistor Terminal Voltages

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

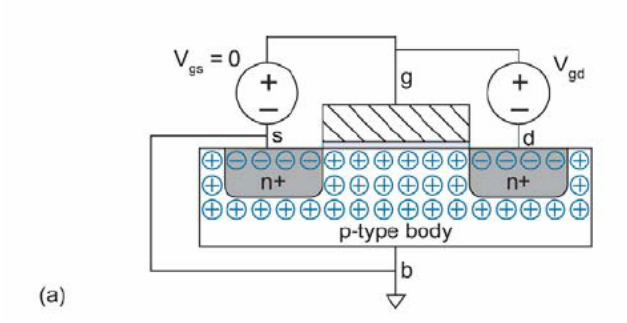


- Source and drain are symmetric diffusion terminals
  - By convention, source is terminal at lower voltage
  - Hence $V_{ds} \geq 0$
- nMOS body is grounded. First assume source is 0 too.
- Three regions of operation
  - Cutoff
  - Linear
  - Saturation

EASTERN WASHINGTON UNIVERSITY

nMOS in cutoff operation mode

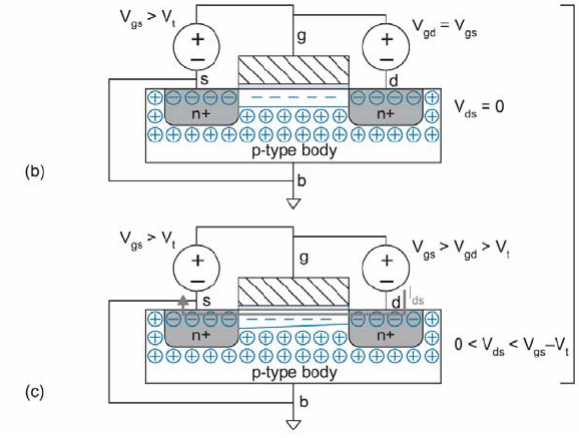
- No channel
- $I_{ds} = 0$



EASTERN WASHINGTON UNIVERSITY

nMOS in linear operation mode

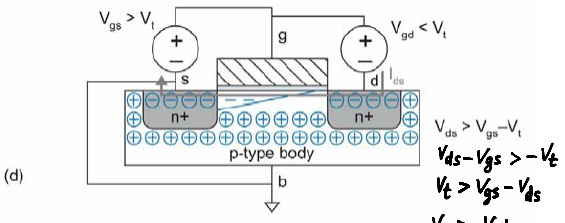
- Channel forms
- Current flows from D
  - e^- from S to D
- I_{ds} increases with V_{ds}
- Similar to linear resistor



EASTERN WASHINGTON UNIVERSITY

nMOS in Saturation operation mode

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



EASTERN WASHINGTON UNIVERSITY

pMOS Transistor

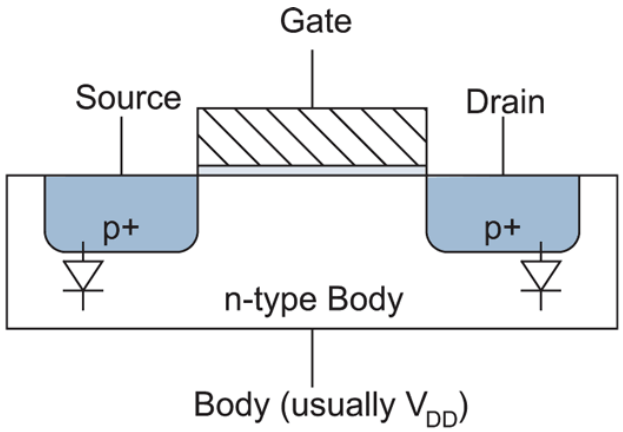


FIG 2.4 pMOS transistor

EASTERN WASHINGTON UNIVERSITY

I-V Characteristics (nMOS)

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

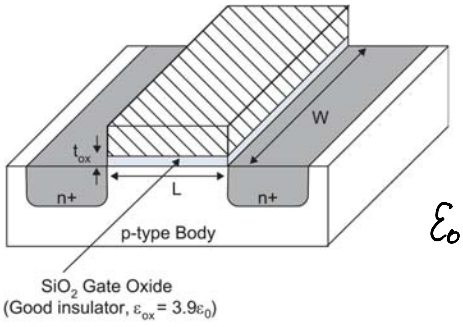


FIG 2.6 Transistor dimensions

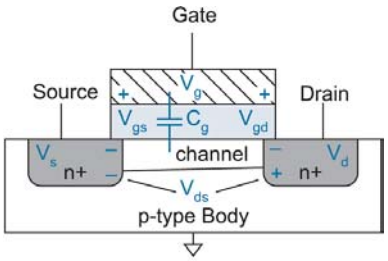
$\varepsilon_0 = 8.85 \times 10^{-12}$ F/m

EASTERN WASHINGTON UNIVERSITY

Channel Charge

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

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



Average gate to channel potential:

$V_{gc} = (V_{gs} + V_{gd})/2 = V_{gs} - V_{ds}/2$

FIG 2.5 Average gate to channel voltage

EASTERN WASHINGTON UNIVERSITY

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:
  - $t = L / v$

EASTERN WASHINGTON UNIVERSITY

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

Many textbooks define:
$k' = \mu C_{ox}$ ← technology dependent parameters

but be careful !!!
MOSIS:
$k' = \mu \frac{C_{ox}}{2}$

technology and geometric parameters

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

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

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

EASTERN WASHINGTON UNIVERSITY

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

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

EASTERN WASHINGTON UNIVERSITY

nMOS I-V Summary

- first order transistor models

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

EASTERN WASHINGTON UNIVERSITY

I-V characteristics of nMOS Transistor

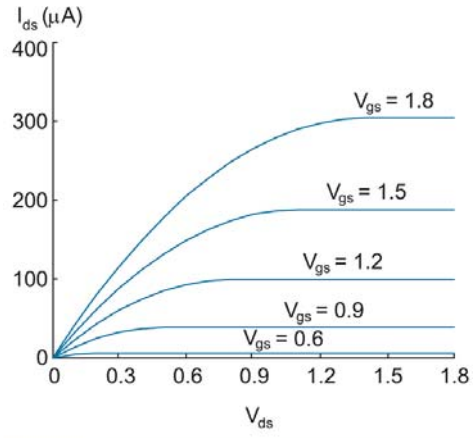
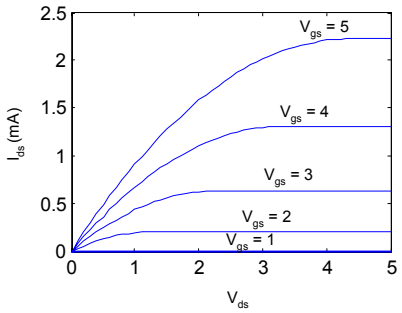


FIG 2.7 I-V characteristics of ideal nMOS transistor

EASTERN WASHINGTON UNIVERSITY

Example

- 0.6 μm process from AMI Semiconductor
  - t_{ox} = 100 Å
  - m = 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



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

EASTERN WASHINGTON UNIVERSITY

pMOS I-V Characteritics

- 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 mm process
- Thus pMOS must be wider to provide same current
  - In this class, assume $\mu_n / \mu_p = 2$

EASTERN WASHINGTON UNIVERSITY

pMOS I-V Summary

- first order transistor models

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

EASTERN WASHINGTON UNIVERSITY

I-V characteristics of pMOS Transistor

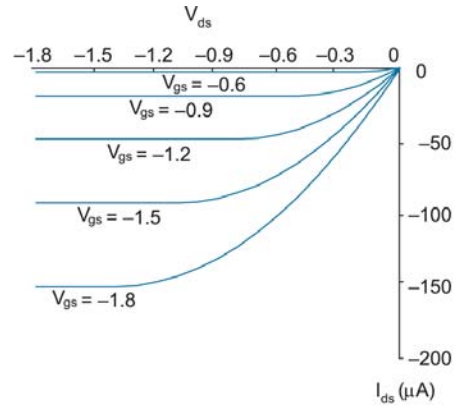


FIG 2.8 I-V characteristics of ideal pMOS transistor

EASTERN WASHINGTON UNIVERSITY

Capacitances of a MOS Transistor

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

EASTERN WASHINGTON UNIVERSITY

Gate Capacitance

- When the transistor is off, the channel is not inverted
  $C_g = C_{gb} = \varepsilon_{ox} WL/t_{ox} = C_{ox} WL$
- Let's call $C_{ox} WL = C_0$
- When the transistor is on, the channel extends from the source to the drain (if the transistor is unsaturated, or to the pinchoff point otherwise)
  $C_g = C_{gb} + C_{gs} + C_{gd}$

EASTERN WASHINGTON UNIVERSITY

Gate Capacitance

Table 2.1 Approximation of intrinsic MOS gate capacitance

Parameter	Cutoff	Linear	Saturation
C_{gb}	C_0	0	0
C_{gs}	0	$C_0/2$	$2/3\ C_0$
C_{gd}	0	$C_0/2$	0
$C_g = C_{gs} + C_{gd} + C_{gb}$	C_0	C_0	$2/3\ C_0$

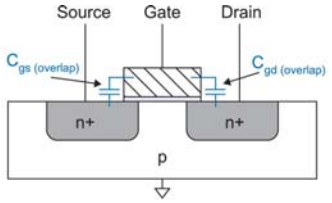


FIG 2.10 Overlap capacitance

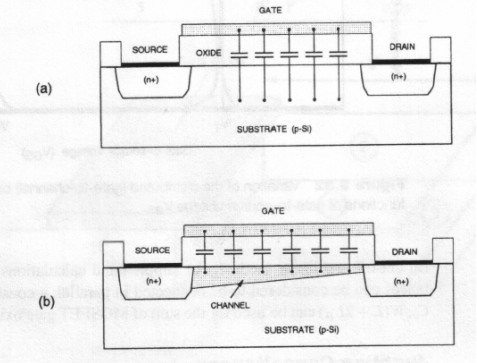
In reality the gate overlaps source and drain. Thus, the gate capacitance should include not only the intrinsic capacitance but also parasitic overlap capacitances:

$C_{gs}(\text{overlap}) = C_{ox}\ W\ L_D$
$C_{gs}(\text{overlap}) = C_{ox}\ W\ L_D$

EASTERN WASHINGTON UNIVERSITY

Detailed Gate Capacitance

Capacitance	Cutoff	Linear	Saturation
C_{gb} (total)	C_0	0	0
C_{gd} (total)	$C_{ox}WL_D$	$C_0/2 + C_{ox}WL_D$	$C_{ox}WL_D$
C_{gs} (total)	$C_{ox}WL_D$	$C_0/2 + C_{ox}WL_D$	$2/3\ C_0 + C_{ox}WL_D$



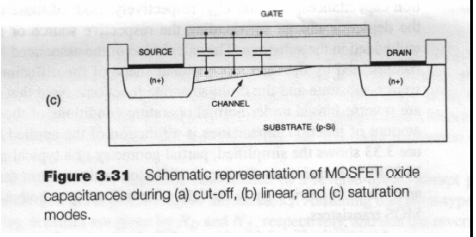


Figure 3.31 Schematic representation of MOSFET oxide capacitances during (a) cut-off, (b) linear, and (c) saturation modes.

Source: M-S Kang, Y. Leblebici, *CMOS Digital ICs*, 3/e, 2003, McGraw-Hill

EASTERN WASHINGTON UNIVERSITY

Diffusion Capacitance

- C_{sb}, C_{db}
- Undesired capacitance (parasitic)
- Due to the reverse biased p-n junctions between source diffusion and body and drain diffusion and body
- Capacitance depends on area and perimeter
  - Use small diffusion nodes
  - Comparable to C_g for contacted diffusion
  - ½ C_g for uncontacted
  - Varies with process

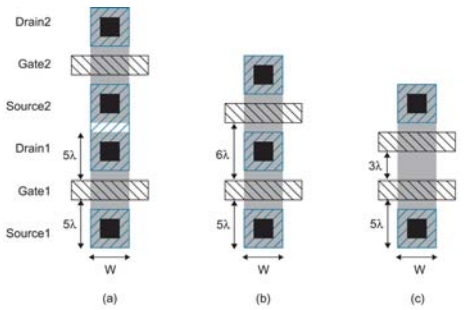


FIG 2.9 Diffusion region geometries

EASTERN WASHINGTON UNIVERSITY

Lumped representation of the MOSFET capacitances

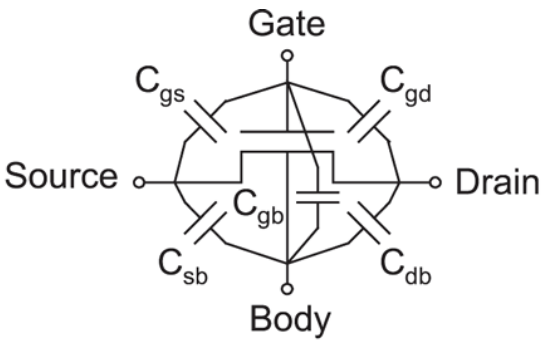


FIG 2.14 Capacitances of an MOS transistor

EASTERN WASHINGTON UNIVERSITY

Non-ideal I-V effects

- The saturation current increases less than quadratically with increasing V_{gs}
  - Velocity saturation
  - Mobility degradation
- Channel length modulation
- Body Effect
- Leakage currents
  - Sub-threshold conduction
  - Junction leakage
  - Tunneling
- Temperature Dependence
- Geometry Dependence

EASTERN WASHINGTON UNIVERSITY

Velocity saturation and mobility degradation

- At strong lateral fields resulting from high V_{ds}, drift velocity rolls off due to carrier scattering and eventually saturates
- Strong vertical fields resulting from large V_{gs} cause the carriers to scatter against the surface and also reduce the carrier mobility. This effect is called mobility degradation

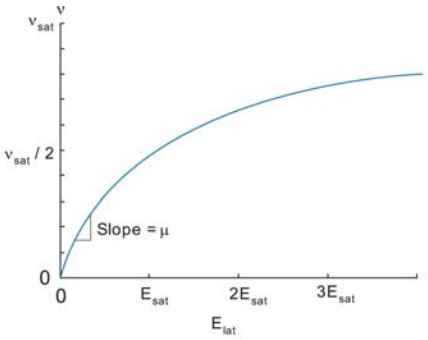


FIG 2.16 Carrier velocity vs. electric field

EASTERN WASHINGTON UNIVERSITY

Channel length modulation

- The reverse biased p-n junction between the drain and the body forms a depletion region with length L' that increases with V_{db}. The depletion region effectively shorten the channel length to: $L_{eff} = L - L'$
- Assuming the source voltage is close to the body votage $V_{db} \sim V_{sb}$. Hence, increasing V_{ds} decrease the effective channel length.
- Shorter channel length results in higher current

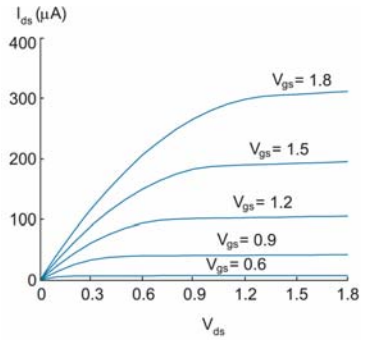


FIG 2.18 I-V characteristics of nMOS transistor with channel length modulation

EASTERN WASHINGTON UNIVERSITY

Body Effect

- The potential difference between source and body V_{sb} affects (increases) the threshold voltage
- Threshold voltage depends on:
  - V_{sb}
  - Process
  - Doping
  - Temperature

EASTERN WASHINGTON UNIVERSITY

Subthreshold Conduction

- The ideal transistor I-V model assumes current only flows from source to drain when $V_{gs} > V_t$.
- In real transistors, current doesn't abruptly cut off below threshold, but rather drop off exponentially
- This leakage current when the transistor is nominally OFF depends on:
  - process (ε_{ox}, t_{ox})
  - doping levels (N_A, or N_D)
  - device geometry (W, L)
  - temperature (T)
  - (Subthreshold voltage (V_t))

EASTERN WASHINGTON UNIVERSITY

Junction Leakage

- The p-n junctions between diffusion and the substrate or well for diodes.
- The well-to-substrate is another diode
- Substrate and well are tied to GND and VDD to ensure these diodes remain reverse biased
- But, reverse biased diodes still conduct a small amount of current that depends on:
  - Doping levels
  - Area and perimeter of the diffusion region
  - The diode voltage

FIG 2.19 Reverse-biased diodes in CMOS circuits

EASTERN WASHINGTON UNIVERSITY

Tunneling

- There is a finite probability that carriers will tunnel though the gate oxide. This result in gate leakage current flowing into the gate
- The probability drops off exponentially with t_{ox}
- For oxides thinner than 15-20 Å, tunneling becomes a factor

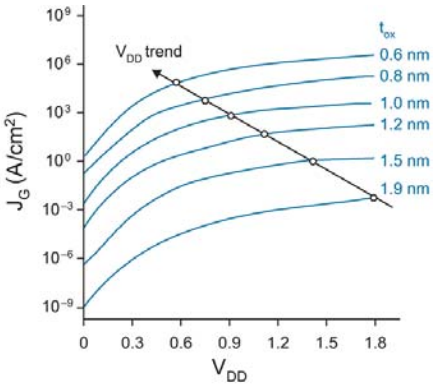


FIG 2.20 Gate leakage current from [Song01]

EASTERN WASHINGTON UNIVERSITY

Temperature dependence

- Transistor characteristics are influenced by temperature
  - μ decreases with T
  - V_t decreases linearly with T
  - $I_{leakage}$ increases with T
- ON current decreases with T
  OFF current increases with T
  - Thus, circuit performances are worst at high temperature

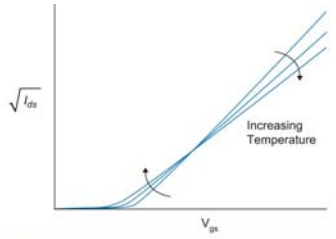


FIG 2.21 I–V characteristics of nMOS transistor in saturation at various temperatures

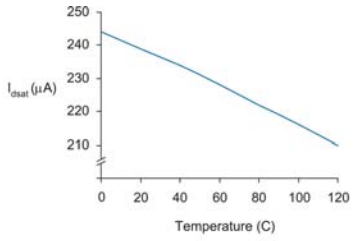


FIG 2.22 I_{dsat} vs. temperature

EASTERN WASHINGTON UNIVERSITY

Geometry Dependence

- Layout designers draw transistors with W_{drawn}, L_{drawn}
- Actual dimensions may differ from some factor X_W and X_L
- The source and drain tend to diffuse laterally under the gate by L_D, producing a shorter effective channel
- Similarly, diffusion of the bulk by W_D decreases the effective channel width
- In process below 0.25 μm the effective length of the transistor also depends significantly on the orientation of the transistor

$$L_{eff} = L_{drawn} + X_L - 2L_D$$

$$W_{eff} = W_{drawn} + X_W - 2W_D$$

EASTERN WASHINGTON UNIVERSITY

Impact of non-ideal I-V effects

- Threshold is a significant fraction of the supply voltage
- Leakage is increased causing gates to
  - consume power when idle
  - limits the amount of time that data is retained
- Leakage increases with temperature
- Velocity saturation and mobility degradation result in less current than expected at high voltage
  - No point in trying to use high VDD to achieve fast transistors
  - Transistors in series partition the voltage across each transistor thus experience less velocity saturation
    - Tend to be a little faster than a single transistor
    - Two nMOS in series deliver more than half the current of a single nMOS transistor of the same width
- Matching: same dimension and orientation

high V_t (better performance)
low V_t (lower power)

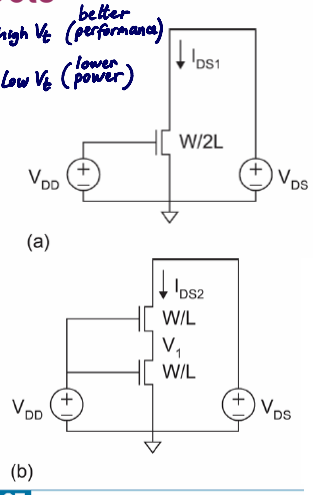


FIG 2.37 Current in series transistors

EASTERN WASHINGTON UNIVERSITY

Pass Transistors

- nMOS pass transistors pull no higher than V_{DD}-V_{tn}
  - Called a degraded “1”
  - Approach degraded value slowly (low I_{ds})
- pMOS pass transistors pull no lower than $|V_{tp}|$
  - Called a degraded “0”
  - Approach degraded value slowly (low I_{ds})

EASTERN WASHINGTON UNIVERSITY

Pass transistor Circuits

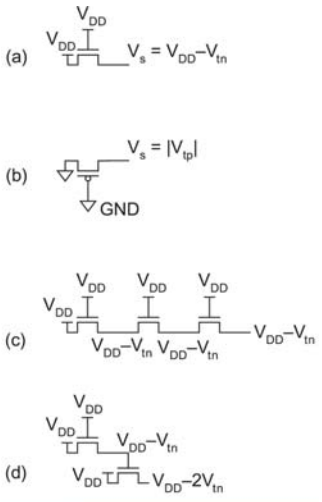
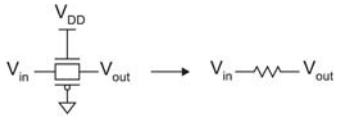


FIG 2.31 Pass transistor threshold drops

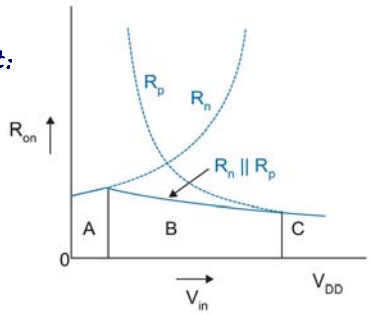
EASTERN WASHINGTON UNIVERSITY

Transmission gate ON resistance



At a given operating point:

$$R = \left(\frac{\partial I_{ds}}{\partial V_{ds}}\right)^{-1}$$



Input voltage Vin is swept from GND to VDD

FIG 2.32 Resistance of a transmission gate as a function of input voltage

EASTERN WASHINGTON UNIVERSITY

Tri-state Inverter

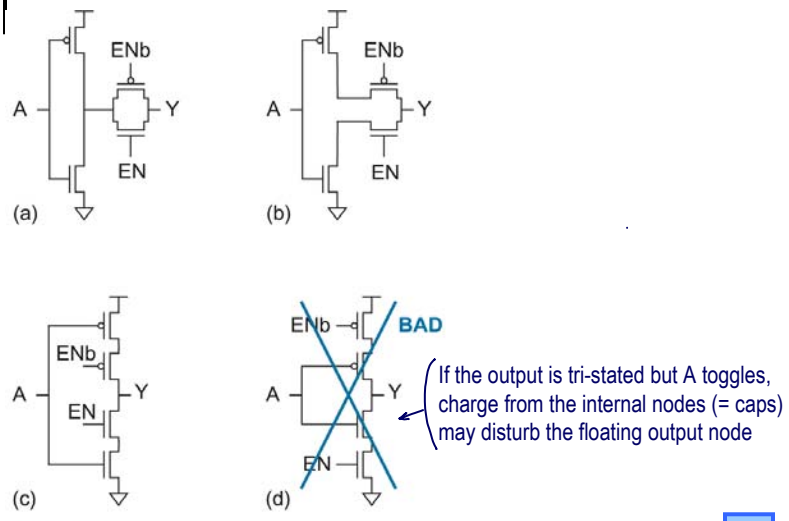


FIG 2.33 Tristate inverter

EASTERN WASHINGTON UNIVERSITY

Effective resistance of a transistor

- First-order transistor models have limited value
  - Not accurate enough for modern transistors
  - Too complicated for hand analysis
- Simplification: treat transistor as resistor
  - Replace $I_{ds}(V_{ds}, V_{gs})$ with effective resistance R
  - $I_{ds} = V_{ds}/R$
  - R averaged across switching range of digital gate
- Too inaccurate to predict current at any given time
  - But good enough to predict RC delay (propagation delay of a logic gate)

EASTERN WASHINGTON UNIVERSITY

RC Values

- Capacitance
  - $C = C_g = C_s = C_d = 2$ fF/μm of gate width
  - Values similar across many processes
- Resistance
  - $R \approx 6$ KΩ*μm in 0.6um process
  - Improves with shorter channel lengths
- Unit transistors
  - May refer to minimum contacted device (4/2 λ)
  - or maybe 1 μm wide device
  - Doesn't matter as long as you are consistent

EASTERN WASHINGTON UNIVERSITY

RC Delay Models

- Use equivalent circuits for MOS transistors
  - ideal switch + capacitance and ON resistance
  - unit nMOS has resistance R, capacitance C
  - unit pMOS has resistance 2R, capacitance C
- Capacitance proportional to width
- Resistance inversely proportional to width

EASTERN WASHINGTON UNIVERSITY

Switch level RC models

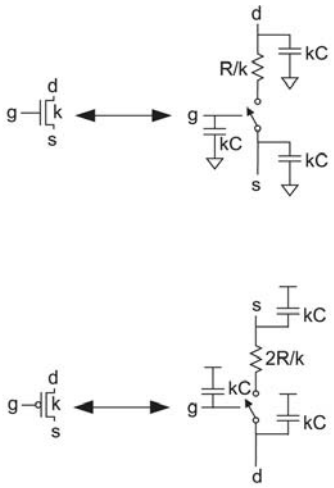


FIG 2.34 Equivalent RC circuit models

EASTERN WASHINGTON UNIVERSITY

Inverter Delay Estimate

- Estimate the delay of a fanout-of-1 inverter

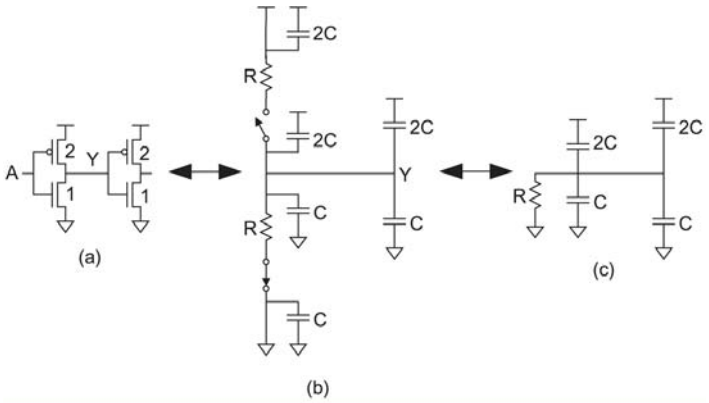


FIG 2.35 Inverter propagation delay

delay = 6RC

EASTERN WASHINGTON UNIVERSITY

Resistance of a unit transmission gate

- The effective resistance of a transmission gate is the parallel of the resistance of the two transistor
- Approximately R in both directions
- Transmission gates are commonly built using equal-sized transistors
- Boosting the size of the pMOS only slightly improve the effective resistance while significantly increasing the capacitance

Effective resistance of a transistor passing a value in its poor direction ≈ double

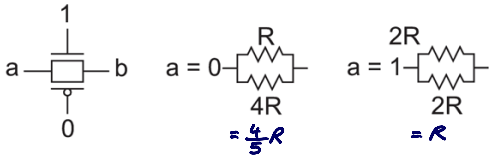


FIG 2.36 Effective resistance of a unit transmission gate

EASTERN WASHINGTON UNIVERSITY

Summary

- Models are only approximations to reality, not reality itself
- Models cannot be perfectly accurate
  - Little value in using excessively complicated models, particularly for hand calculations
- To first order current is proportional to W/L
  - But, in modern transistors L_{eff} is shorter than L_{drawn}
    - Doubling the L_{drawn} reduces current more than a factor of two
    - Two series transistors in a modern process deliver more than half the current of a single transistor
- Use Transmission gates in place of pass transistors
- Transistor speed depends on the ratio of current to capacitance
  - Sources of capacitance (voltage dependents)
    - Gate capacitance
    - Diffusion capacitance