

CS/EE 6710

MOS Transistor Models
Electrical Effects
Propagation Delay

V_{gs} Forms a Channel

$V_{gs} < V_t$ $V_{gs} = V_t$ $V_{gs} > V_t$

V_t : The threshold voltage to turn on the transistor

N-type Transistor

+V_{gs} V_{ds} i electrons

MOS Capacitor

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

Another Cutaway View

Thanks to National Central University for Some images

Transistor Characteristics

- ▶ Three conduction characteristics
 - ▶ **Cutoff Region**
 - ▶ No inversion layer in channel
 - ▶ $I_{ds} = 0$
 - ▶ **Nonsaturated, or linear region**
 - ▶ Weak inversion of the channel
 - ▶ I_{ds} depends on V_{gs} and V_{ds}
 - ▶ **Saturated region**
 - ▶ Strong inversion of channel
 - ▶ I_{ds} is independent of V_{ds}
- ▶ As an aside, at very high drain voltages:
 - ▶ "avalanche breakdown" or "punch through"
 - ▶ Gate has no control of I_{ds} ...

nMOS Cutoff: $V_{gs} < V_t$

- ▶ No channel
- ▶ $I_{ds} = 0$

Basic N-Type MOS Transistor

- ▶ Conditions for the regions of operation
 - ▶ **Cutoff:** If $V_{gs} < V_t$, then I_{ds} is essentially 0
 - ▶ V_t is the "Threshold Voltage"
 - ▶ **Linear:** If $V_{gs} > V_t$ and $V_{ds} < (V_{gs} - V_t)$ then I_{ds} depends on both V_{gs} and V_{ds}
 - ▶ Channel becomes deeper as V_{gs} goes up
 - ▶ **Saturated:** If $V_{gs} > V_t$ and $V_{ds} > (V_{gs} - V_t)$ then I_{ds} is essentially constant (Saturated)

nMOS Linear: $V_{gs} > V_t$, small V_{ds}

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

Transistor Gain

- ▶ β is the MOS transistor gain factor
- ▶ $\beta = (\mu\epsilon/t_{ox})(W/L)$
 - Process-dependent
 - Layout dependent
- ▶ μ = mobility of carriers
 - ▶ Note that N-type is twice as good as P-type
- ▶ ϵ = permittivity of gate insulator
 - ▶ $\epsilon = 3.9 \epsilon_0$ for SiO_2 ($\epsilon_0 = 8.85 \times 10^{-14}$ F/cm)
- ▶ T_{ox} = thickness of gate oxide
- ▶ Also, $\epsilon/t_{ox} = C_{ox}$ The oxide capacitance
 - ▶ $\beta = (\mu C_{ox})(W/L) = k'(W/L) = KP(W/L)$
- ▶ Increase W/L to increase gain

nMOS Saturation: $V_{ds} > V_{gs} - V_t$

- ▶ Channel pinches off
 - ▶ Conduction by drift because of positive drain voltage
 - ▶ Electrons are injected into depletion region
- ▶ I_{ds} independent of V_{ds}
- ▶ We say that the current saturates
- ▶ Similar to current source

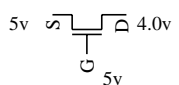
Example

- ▶ We will be using a $0.6 \mu\text{m}$ process for your project
 - ▶ From AMI Semiconductor
 - ▶ $t_{ox} = 100 \text{ \AA}$
 - ▶ $\mu = 350 \text{ cm}^2/\text{V}^*\text{s}$
 - ▶ $V_t = 0.7 \text{ V}$
- ▶ Plot I_{ds} vs. V_{ds}
 - ▶ $V_{gs} = 0, 1, 2, 3, 4, 5$
 - ▶ Use $W/L = 4/2 \lambda$

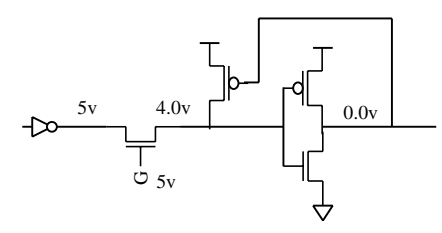
$$\beta = \mu C_{ox} \frac{W}{L} = (350) \left(\frac{3.9 \cdot 8.85 \cdot 10^{-14}}{100 \cdot 10^{-8}} \right) \left(\frac{W}{L} \right) = 120 \frac{W}{L} \mu\text{A}/\text{V}^2$$

"Saturated" Transistor

- ▶ In the $0 < (V_{gs} - V_t) < V_{ds}$ case
 - ▶ I_{ds} Current is effectively constant
 - ▶ Channel is "pinched off" and conduction is accomplished by drift of carriers
 - ▶ Voltage across pinched off channel (I.e. V_{ds}) is fixed at $V_{gs} - V_t$
 - ▶ This is why you don't use an N-type to pass 1's!
 - ▶ High voltage is degraded by V_t
 - ▶ If V_t is 1.0v, 5v in one side, 4.0v out the other

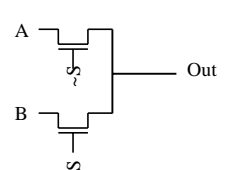


N-type Pass Transistors



- ▶ Another option is a "keeper" transistor fed back from the output
 - ▶ This pulls the internal node high when the output is 0
 - ▶ But is disconnected when output is high
- ▶ Make sure the size is right...

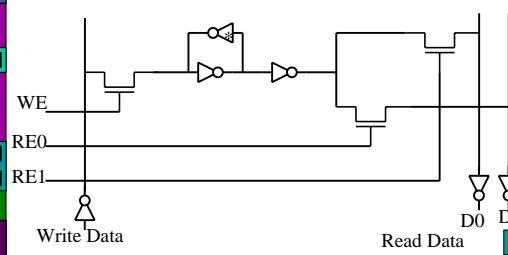
Aside: N-type Pass Transistors



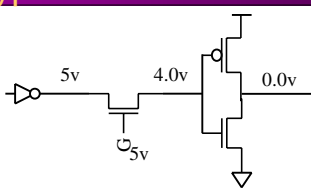
- ▶ If it weren't for the threshold drop, N-type pass transistors (without the P-type transmission gate) would be nice
 - ▶ 2-way Mux Example...

N-type Pass Transistors

- ▶ In practice, they are used fairly often, but be aware of what you're doing
 - ▶ For example, read/write circuits in a [Register File](#)



N-type Pass Transistors



- ▶ On one hand, the degraded high voltage from the pass transistor will be restored by the inverter
- ▶ On the other hand, the P-device may not turn off completely resulting in extra power being used

Back to the Saturated Transistor

- ▶ What influences the constant I_{ds} in the saturated case?
 - ▶ Channel length
 - ▶ Channel width
 - ▶ Threshold voltage V_t
 - ▶ Thickness of gate oxide
 - ▶ Dielectric constant of gate oxide
 - ▶ Carrier mobility μ
 - ▶ Velocity Saturation

Threshold Voltage: V_t

- ▶ The V_{gs} voltage at which I_{ds} is essentially 0
 - ▶ $V_t = .67v$ for nmos and $-.92v$ for pmos in our process
 - ▶ Tiny I_{ds} is exponentially related to V_{gs} , V_{ds}
 - ▶ Take 5720/6720 for "subthreshold" circuit ideas
- ▶ V_t is affected by
 - ▶ Gate conductor material
 - ▶ Gate insulator material
 - ▶ Gate insulator thickness
 - ▶ Channel doping
 - ▶ Impurities at Si/insulator interface
 - ▶ Voltage between source and substrate (V_{sb})

I_{ds} Curves

$\beta = \frac{\mu e}{t_{ox}} \left(\frac{W}{L} \right) = \mu C_{ox} \left(\frac{W}{L} \right)$

Cutoff Region
 $V_{gs} < V_t$
 $I_{ds} = 0$

Triode (Linear) Region
 $V_{gs} - V_t > V_{ds} > 0$
 $I_{ds} = \beta \left[(V_{gs} - V_t)V_{ds} - \frac{V_{ds}^2}{2} \right]$

Saturation Region
 $V_{gs} - V_t > V_{ds} > 0$
 $I_{ds} = \beta \frac{(V_{gs} - V_t)^2}{2}$

2nd Order Effect: Body Effect

- ▶ A second order effect that raises V_t
- ▶ Recall that V_t is affected by V_{sb} (voltage between source and substrate)
 - ▶ Normally this is constant because of common substrate
 - ▶ But, when transistors are in series, V_{sb} ($V_s - V_{substrate}$) may be changed

$V_{t2} > V_{t1}$

P-type Transistor

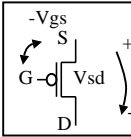
Basic DC Equations for I_{ds}

- ▶ **Cutoff Region**
 - ▶ $V_{gs} < V_t$, $I_{ds} = 0$
- ▶ **Linear Region**
 - ▶ $0 < V_{ds} < (V_{gs} - V_t)$
 - ▶ $I_{ds} = \beta[(V_{gs} - V_t)V_{ds} - V_{ds}^2/2]$
 - ▶ Note that this is only "linear" if $V_{ds}^2/2$ is very small, i.e. $V_{ds} \ll V_{gs} - V_t$
- ▶ **Saturated Region**
 - ▶ $0 < (V_{gs} - V_t) < V_{ds}$
 - ▶ $I_{ds} = \beta[(V_{gs} - V_t)^2/2]$

P-type Transistor

P-type Transistors

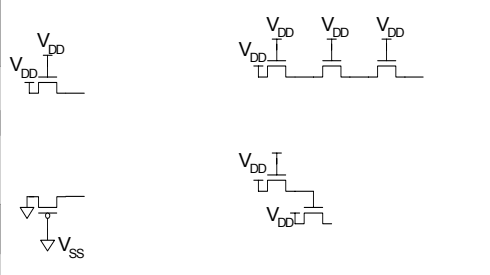
- ▶ Source is Vdd instead of GND
 - ▶ $V_{sg} = (V_{dd} - V_{in})$, $V_{sd} = (V_{dd} - V_{out})$, V_t is negative
- ▶ **Cutoff:** $(V_{dd} - V_{in}) < -V_t$, $I_{ds} = 0$
- ▶ **Linear Region**
 - ▶ $(V_{dd} - V_{out}) < (V_{dd} - V_{in} + V_t)$
 - $I_{ds} = \beta[(V_{dd} - V_{in} + V_t)(V_{dd} - V_{out}) - (V_{dd} - V_{out})^2/2]$
- ▶ **Saturated Region**
 - ▶ $((V_{dd} - V_{in}) + V_t) < (V_{dd} - V_{out})$
 - $I_{ds} = \beta[(V_{dd} - V_{in} + V_t)^2/2]$



2nd Order Effect: Velocity Saturation

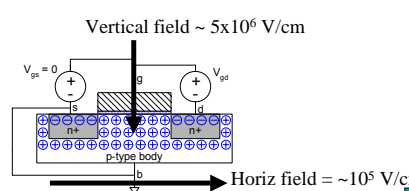
- ▶ With weak fields, current increases linearly with lateral electric field
- ▶ At higher fields, carrier drift velocity rolls off and saturates
 - ▶ Due to carrier scattering
 - ▶ Result is less current than you think!
 - ▶ For a 2μ channel length, effects start around $4v V_{dd}$
 - ▶ For 180nm, effects start at $0.36v V_{dd}$!

Pass Transistor Ckts



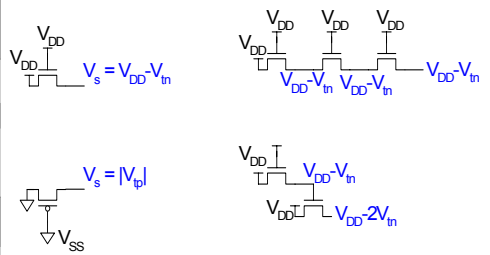
2nd Order Effect: Velocity Saturation

- ▶ When the carriers reach their speed limit in silicon...
- ▶ Channel lengths have been scaled so that vertical and horizontal EM fields are large and interact with each other



Vertical field $\sim 5 \times 10^5$ V/cm
 Horiz field $= \sim 10^5$ V/cm

Pass Transistor Ckts



$V_s = V_{DD} - V_{in}$
 $V_s = |V_{tp}|$
 $V_{DD} - V_{in}$
 $V_{DD} - 2V_{in}$

2nd Order Effect: Velocity Saturation

- ▶ When the carriers reach their speed limit in silicon...
- ▶ Means that relationship between I_{ds} and V_{gs} is closer to linear than quadratic
- ▶ Also the saturation point is smaller than predicted
- ▶ For example, 180nm process
 - ▶ 1st order model = 1.3v
 - ▶ Really is 0.6v

2nd Order Effect: Velocity Saturation

- ▶ This is a basic difference between long- and short-channel devices
 - ▶ The strength of the horizontal EM field in a short channel device causes the carriers to reach their velocity limit early
 - ▶ Devices saturate faster and deliver less current than the quadratic model predicts

2nd Order Effect: Body Effect

- **Body Effect** - V_t is a function of voltage between source and substrate

$$V_t = V_{t0} + \gamma \sqrt{(2\phi_b + V_{sb}) + 2\sqrt{\phi_b}}$$

$$\phi_b = \frac{kT}{q} \ln\left(\frac{N_A}{N_i}\right)$$

$$\gamma = \frac{t_{ox}}{\epsilon_{ox}} \sqrt{2q\epsilon_{si}N_A} = \frac{1}{C_{ox}} \sqrt{2q\epsilon_{si}N_A}$$

National Central University 2. CMOS Transistor Theory

2nd Order Effect: Velocity Saturation

- ▶ Consider two devices with the same W/L ratio in our process ($V_{gs}=5v$, $V_{dd}=5v$)
 - ▶ 100/20 vs 4.6/1.2
 - ▶ They should have the same current...
 - ▶ Because of velocity saturation in the short-channel device, it has 47% less current!

2nd Order Effect: Body Effect

- ▶ Consider an nmos transistor in a 180nm process
 - ▶ Nominal V_t of 0.4v
 - ▶ Body is tied to ground
 - ▶ How much does the V_t increase if the source is at 1.1v instead of 0v?
 - ▶ Because of the body effect, V_t increases by 0.28v to be 0.68v!

2nd Order Effect: Body Effect

- ▶ A second order effect that raises V_t
- ▶ Recall that V_t is affected by V_{sb} (voltage between source and substrate)
 - ▶ Normally this is constant because of common substrate
 - ▶ But, when transistors are in series, V_{sb} ($V_s - V_{substrate}$) may be changed

2nd Order Effect

- **Channel Length Modulation** - Channel length is a function of V_{ds} . When V_{ds} increase, the depletion region of the pinch off at drain shorten the channel length.

$$L_{eff} = L - L_{short}$$

$$L_{short} = \sqrt{2 \frac{\epsilon_{si}}{qN_A} (V_{ds} - (V_{gs} - V_t))}$$

$$I_{ds} = \frac{kW}{2L} (V_{gs} - V_t)^2 (1 + \lambda V_{ds})$$

2nd Order Effect

- Mobility Variation -**
The mobility of the carrier decreases when the carrier density increases. Therefore, when V_{gs} is large. The density of the carrier in the channel increases. As a result, the mobility decreases.

$$\mu = \frac{\text{Average_carrier_drift_velocity}(V)}{\text{Electrical_Field}(E)}$$

$$\mu_n = 600 \text{ cm}^2 / V \cdot \text{sec}$$

$$\mu_p = 250 \text{ cm}^2 / V \cdot \text{sec}$$

Inverter Switching Point

- Inverter switching point is determined by ratio of β_n/β_p
 - If $\beta_n/\beta_p = 1$, then switching point is $V_{dd}/2$
 - If W/L of both N and P transistors are equal
 - Then $\beta_n/\beta_p = \mu_n/\mu_p =$ electron mobility / hole mobility
 - This ratio is usually between 2 and 3
 - Means ratio of $W_{p\text{tree}}/W_{n\text{tree}}$ needs to be between 2 and 3 for $\beta_n/\beta_p = 1$
 - For this class, we'll use $W_{p\text{tree}}/W_{n\text{tree}} = 2$

2nd Order Effect

- Fowler-Nordheim Tunneling**
When the gate oxide is very thin, a current can flow from gate to source by electron tunneling through the gate oxide.

$$I_{FN} = C_1 W L E_{ox}^2 e^{-\frac{-E_o}{E_{ox}}}$$

$$E_{ox} = \frac{V_{gs}}{t_{ox}}$$

- Drain Punchthrough**
When the drain voltage is high enough, the depletion region around the drain may extend to the source. Thus, causing current to flow irrespective of the gate voltage.

Inverter Switching Point

2nd Order Effect

- Impact Ionization - Hot Electrons**
When the source-drain electric field is too large, the electron speed will be high enough to break the electron-hole pair. Moreover, the electrons will generate the gate oxide, causing a gate current.
- Subthreshold Region**
The cutoff region is also referred to as the subthreshold region, where I_{ds} increase exponentially with V_{ds} and V_{gs} .

Inverter Operating Regions

Region	NMOS	PMOS
A	Off	Linear
B	Sat	Linear
C	Sat	Sat
D	Linear	Sat
E	Linear	Off

Gate Sizes

- ▶ Assume minimum inverter is $W_p/W_n = 2/1$ ($L = L_{min}$, $W_n = W_{min}$, $W_p = 2W_n$)
 - ▶ This becomes a 1x inverter
- ▶ To drive larger capacitive loads, you need more gain, more I_{ds}
 - ▶ Increase widths to get 2x inverter
 - ▶ W_p/W_n is still 2/1, but W_p and W_n are double the size
 - ▶ For most gates, diminishing returns after about 4x size

Inverter Noise Margin

- ▶ To maximize noise margins, select logic levels at:
 - ▶ unity gain point of DC transfer characteristic

Inverter β Ratios

Performance Estimation

- ▶ First we need to have a model for resistance and capacitance
 - ▶ Delays are caused (to first order) by RC delays charging and discharging capacitors
- ▶ All these layers on the chip have R and C associated with them
- ▶ Mostly this is handled in the Spectre simulator
 - ▶ But it's good to have an idea what's going on

Inverter Noise Margin

How much noise can a gate see before it doesn't work right?

Resistance

- ▶ $R = (\rho/t)(L/W) = R_s(L/W)$
 - ▶ ρ = resistivity of the material
 - ▶ t = thickness
 - ▶ R_s = sheet resistance in Ω/square
- ▶ Typical values of R_s

	Min	Typ	Max
M3	0.03	0.04	0.05
M1, M2	0.05	0.07	0.1
Poly	15	20	30
Silicide	2	3	6
Diffusion	10	25	100
Nwell	1k	2k	5k

Capacitance

- ▶ Three main forms:
 - ▶ Gate capacitance (gate of transistor)
 - ▶ Diffusion capacitance (drain regions)
 - ▶ Routing capacitance (metal, etc.)

$C_g = C_{gb} + C_{gs} + C_{gd}$
 Approximated by
 $C = C_{ox}A$
 C_{ox} = thin oxide cap
 A = area of gate

RC Wire Delay Example

- ▶ $R = 20\Omega/\text{sq}$
- ▶ $C = 4 \times 10^{-4} \text{ pF}/\mu\text{m}$
- ▶ $L = 2\text{mm}$
- ▶ $K = 0.7$
- ▶ $T = kRC L^2/2$
- ▶ $T = (0.7)(20)(4 \times 10^{-15})(2000)^2 / 2 \text{ s}$
- ▶ delay = 11.2 ns

Routing Capacitance

- ▶ First order effect is layer->substrate
 - ▶ Approximate using parallel plate model
 - ▶ $C = (\epsilon/t)A$
 - ▶ ϵ = permittivity of insulator
 - ▶ t = thickness of insulator
 - ▶ A = area
 - ▶ Fringing fields increase effective area
- ▶ Capacitance between layers becomes very complex!
 - ▶ Crosstalk issues...

RC Wire/Buffer Delay Example

- ▶ Now split into 2 1mm segments with a buffer
- ▶ $T = 2 \times (0.7)(20)(4 \times 10^{-15})(1000)^2/2 + T_{\text{buf}}$
 $= 5.6\text{ns} + T_{\text{buf}}$
- ▶ Assuming T_{buf} is less than 5.6ns (which it will be), the split wire is a win

Distributed RC on Wires

- ▶ Wires look like distributed RC delays
 - ▶ Long resistive wires can look like transmission lines
 - ▶ Inserting buffers can really help delay

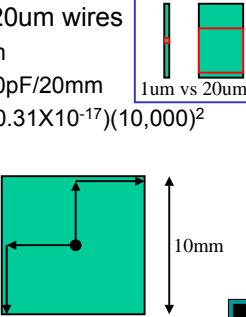
- ▶ $T_n = RCn(n+1)/2$
- ▶ $T = kRC L^2/2$ as the number of segments becomes large
 - ▶ K = constant (i.e. 0.7)
 - ▶ R = resistance per unit length
 - ▶ C = capacitance per unit length
 - ▶ L = length of wire

Another Example: Clock

- ▶ 50pF clock load distributed across 10mm chip in 1 μm metal
 - ▶ Clock length = 20mm
 - ▶ $R = 0.05\Omega/\text{sq}$, $C = 50\text{pF}/20\text{mm}$
 - ▶ $T = (0.7)(RC/2)L^2 = (6.25 \times 10^{-17})(20,000)^2 = 17.5\text{ns}$

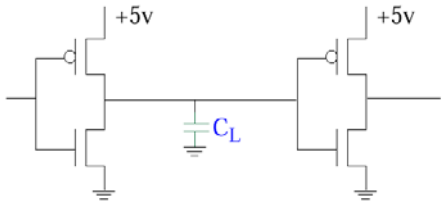
Different Distribution Scheme

- ▶ Put clock driver in the middle of the chip
- ▶ Widen clock line to 20um wires
 - ▶ Clock length = 10mm
 - ▶ $R = 0.05\Omega/\text{sq}$, $C = 50\text{pF}/20\text{mm}$
 - ▶ $T = (0.7)(RC/2)L^2 = (0.31 \times 10^{-17})(10,000)^2 = 0.22\text{ns}$
 - ▶ Reduces R by a factor of 20, L by 2
 - ▶ Increases C a tiny bit



Propagation Delay

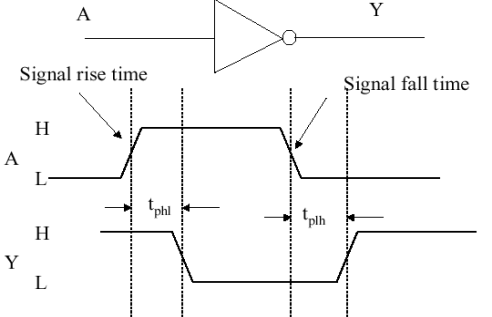
- Recall that it takes time to charge capacitors
- Recall that the gate of a transistor looks like a capacitor
- Wires have resistance and capacitance also!



Capacitance Design Guide

- ▶ Get a table of typical capacitances per unit square for each layer
 - ▶ Capacitance to ground
 - ▶ Capacitance to another layer
- ▶ Add them up...
- ▶ See, for example, Tables 4.8, 4.9 in your book

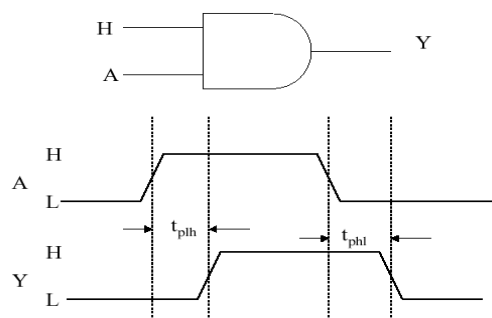
Inverting Propagation Delay



Wire Length Design Guide

- ▶ How much wire can you use in a conducting layer before the RC delay approaches that of a unit inverter?
 - ▶ Metal3 = 2,500u
 - ▶ Metal2 = 2,000u
 - ▶ Metal1 = 1,250u
 - ▶ Silicide = 150u
 - ▶ Poly = 50u
 - ▶ Diffusion = 15u

Non-Inverting Delay



Where to Measure Delay?

If use 50% point (input) to 50% point (output), can produce negative delays (slow input slope, fast output slope).
 A better way is to use the 30% and 70% points on the signals.

Process Corners

- ▶ When parts are specified, under what operating conditions?
- ▶ **Temp:** three ranges
 - ▶ Commercial: 0 C to 70 C
 - ▶ Industrial: -40 C to 85 C
 - ▶ Military: -55 C to 125 C
- ▶ **Vdd:** Should vary $\pm 10\%$
 - ▶ 4.5 to 5.5v for example
- ▶ **Process variation:**
 - ▶ Each transistor type can be slow or fast

Example Non-Inverting Gate

What Else Affects Gate Delay?

Input slew and output load both effect timing. For a FIXED input slope, FIXED environment, a simple timing model is:

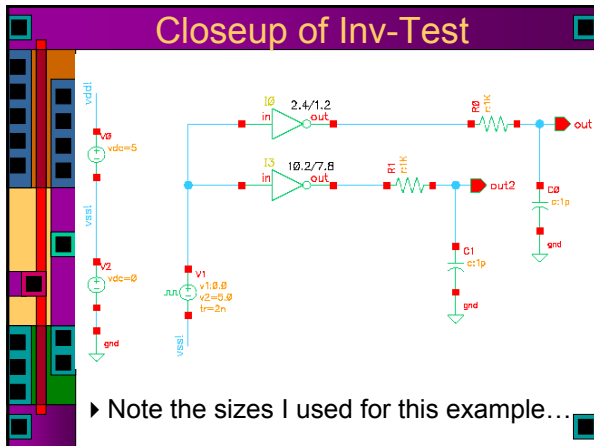
$$\text{delay} = T_{\text{load}} + K * \text{Cload}$$

T_{load} is the delay of the gate with no external load.
 K is different for TPLH, TPHL since it represents the channel resistance. Same equation is used for Slew values.

What Affects Gate Delay?

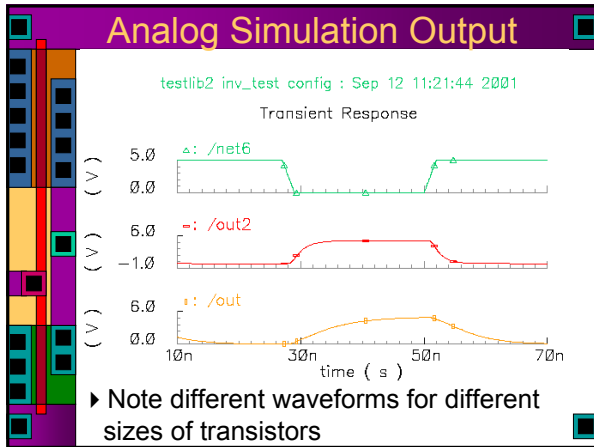
- ▶ **Environment**
 - ▶ Increasing Vdd decreases delay
 - ▶ Decreasing temperature decreases delay
 - ▶ Fabrication effects, fast/slow devices
- ▶ Usually measure delay for at least three cases:
 - ▶ Best - high Vdd, low temp, fast N, Fast P
 - ▶ Worst - low Vdd, high temp, slow N, Slow P
 - ▶ Typical - typ Vdd, room temp (25C), typ N, typ P

Inv_Test Schematic



RC Delay Model

- ▶ 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



RC Values

- ▶ Capacitance
 - ▶ $C = C_g = C_s = C_d = 2 \text{ fF}/\mu\text{m}$ of gate width
 - ▶ Values similar across many processes
- ▶ Resistance
 - ▶ $R \approx 6 \text{ K}\Omega \cdot \mu\text{m}$ in 0.6 μm process
 - ▶ Improves with shorter channel lengths
- ▶ Unit transistors
 - ▶ May refer to minimum contacted device ($4/2 \lambda$)
 - ▶ Or maybe $1 \mu\text{m}$ wide device
 - ▶ Doesn't matter as long as you are consistent

Effective Resistance

- ▶ Shockley models have limited value
 - ▶ Not accurate enough for modern transistors
 - ▶ Too complicated for much 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 of digital gate
- ▶ Too inaccurate to predict current at any given time
 - ▶ But good enough to predict RC delay

Inverter Delay Estimate

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

Inverter Delay Estimate

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

What's a Standard Load?

step or high slew realistic input slope

Must include realistic load, realistic input slew

Fanout 4 (FO4) load

Inverter Delay Estimate

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

What About Gates in Series

▶ Basically we want every gate to have the delay of a “standard inverter”

- ▶ Standard inverter starts with 2/1 P/N ratio
- ▶ Gates in series? Sum the conductance to get the series conductance
- ▶ $\beta_{n\text{-eff}} = 1 / (1/\beta_1 + 1/\beta_2 + 1/\beta_3)$
- ▶ $\beta_{n\text{-eff}} = \beta_n / 3$
- ▶ Effect is like increasing L by 3
- ▶ Compensate by increasing W by 3

Inverter Delay Estimate

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

= 6RC

Power Dissipation

▶ Three main contributors:

1. Static leakage current (P_s)
2. Dynamic short-circuit current during switching (P_{sc})
3. Dynamic switching current from charging and discharging capacitors (P_d)

▶ Becoming a HUGE problem as chips get bigger, clocks get faster, transistors get leakier!

- ▶ Power typically gets dissipated as heat...

Static Leakage Power

- ▶ Small static leakage current due to:
 - ▶ Reverse bias diode leakage between diffusion and substrate (PN junctions)
 - ▶ Subthreshold conduction in the transistors
- ▶ Leakage current can be described by the diode current equation
 - ▶ $I_o = I_s(e^{qV/kT} - 1)$
 - ▶ Estimate at 0.1nA – 0.5nA per device at room temperature

Short-Circuit Dissipation

- ▶ So, with short-circuit current on every transition of the output, integrate under that current curve to get the total current
 - ▶ It works out to be:
 - ▶ $P_{sc} = B/12(V_{dd} - 2V_t)^3 (T_r / T_p)$
 - ▶ Assume that $T_r = T_f$, $V_{tn} = -V_{tp}$, and $B_n = B_p$
 - ▶ Note that P_{sc} depends on B, and on input waveform rise and fall times

Static Leakage Power

- ▶ That's the leakage current
- ▶ For static power dissipation:
 - ▶ $P_s = \text{SUM of } (I \times V_{dd})$ for all n devices
 - ▶ For example, inverter at 5v leaks about 1-2 nW in a .5u technology
 - ▶ Not much...
 - ▶ ...but, it gets MUCH worse as feature size shrinks!

Dynamic Dissipation

- ▶ Charging and discharging all those capacitors!
 - ▶ By far the largest component of power dissipation
 - ▶ $P_d = C_L V_{dd}^2 f$
- ▶ Watch out for large capacitive nodes that switch at high frequency
 - ▶ Like clocks...

Short-Circuit Dissipation

- ▶ When a static gate switches, both N and P devices are on for a short amount of time
 - ▶ Thus, current flows during that switching time

Total Power

- ▶ These are pretty rough estimates
- ▶ It's hard to be more precise without CAD tool support
 - ▶ It all depends on frequency, average switching activity, number of devices, etc.
 - ▶ There are programs out there that can help
- ▶ But, even a rough estimate can be a valuable design guide
 - ▶ $P_{total} = P_s + P_{sc} + P_d$