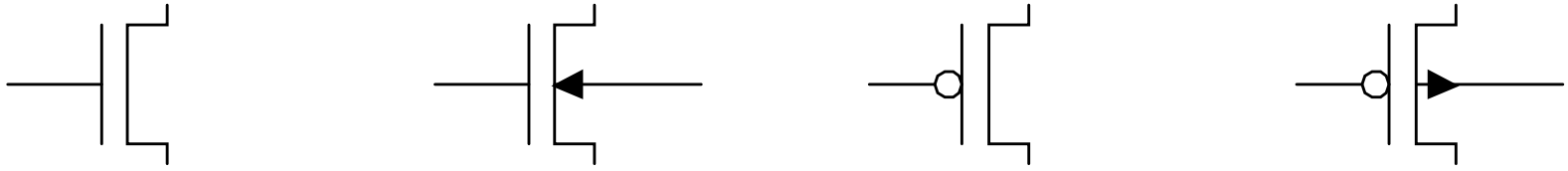
Lecture 1: Ideal/nonideal MOS transistor theory and reliability

Outline

- Introduction
- MOS Capacitor
- nMOS I-V Characteristics
- pMOS I-V Characteristics
- Gate and Diffusion Capacitance
- Pass Transistors
- RC Delay Models
- Nonideal MOS Transistor
- Reliability

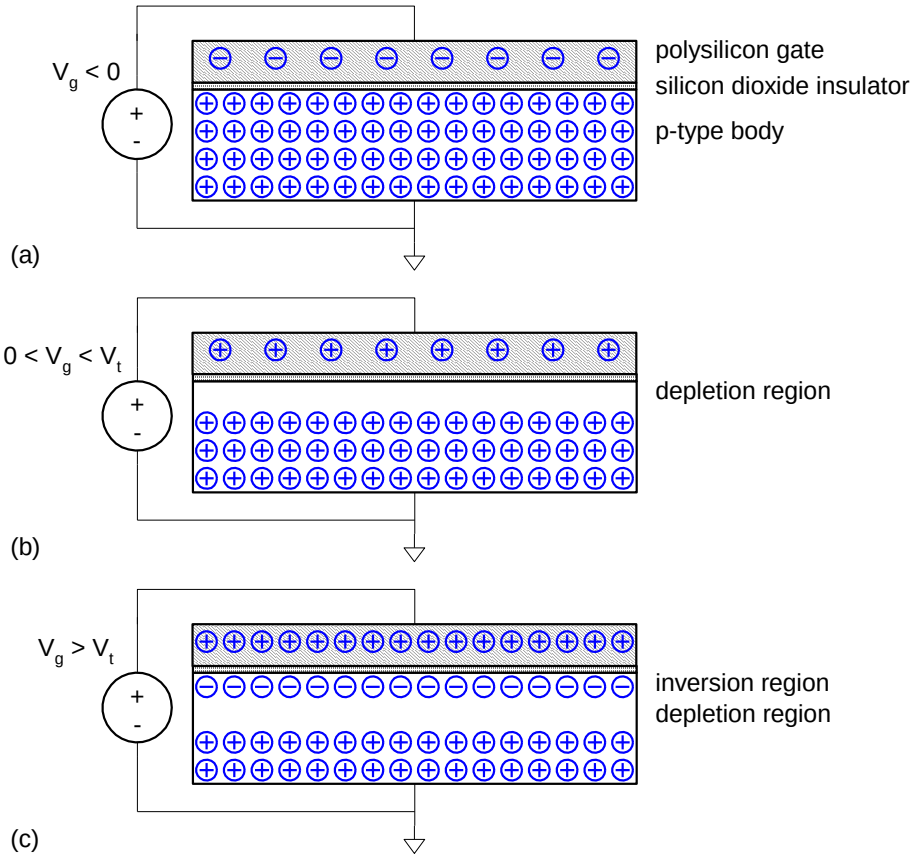
Introduction

- 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



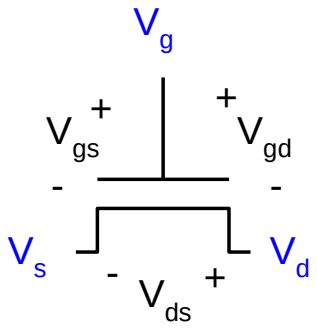
MOS Capacitor

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



Terminal Voltages

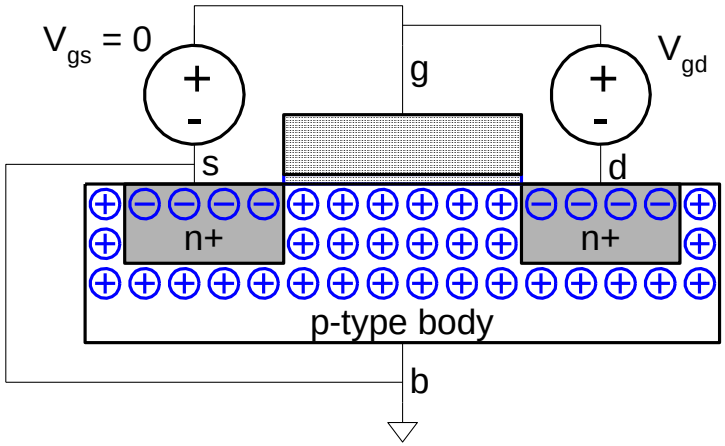
- Mode of operation depends on V_g, V_d, V_s
  - $V_{gs} = V_g - V_s$
  - $V_{gd} = V_g - V_d$
  - $V_{ds} = V_d - V_s = V_{gs} - V_{gd}$



- 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*

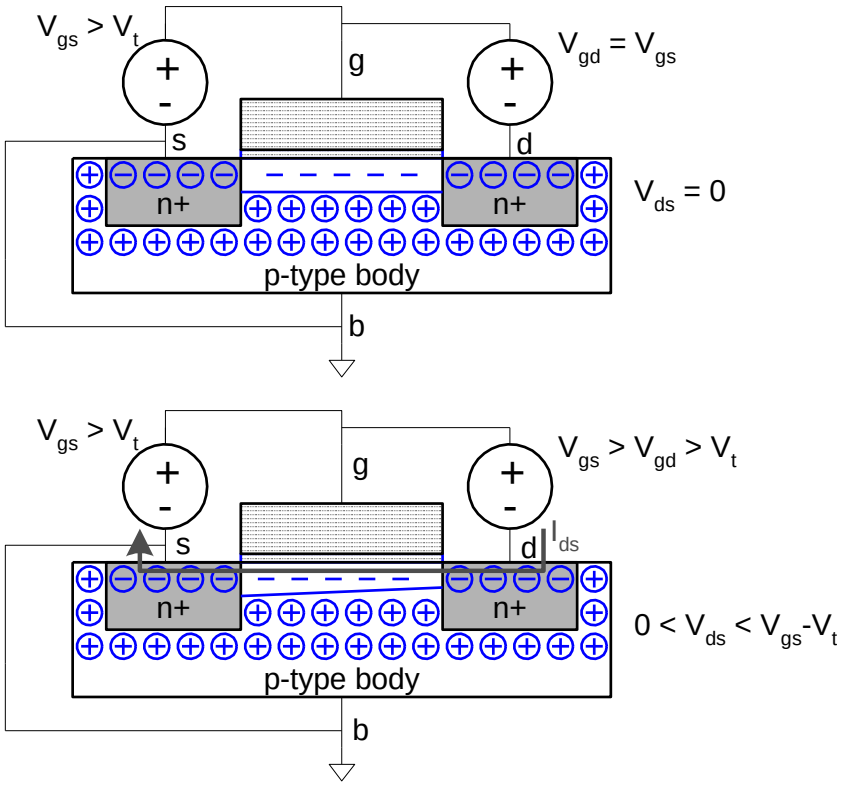
nMOS Cutoff

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



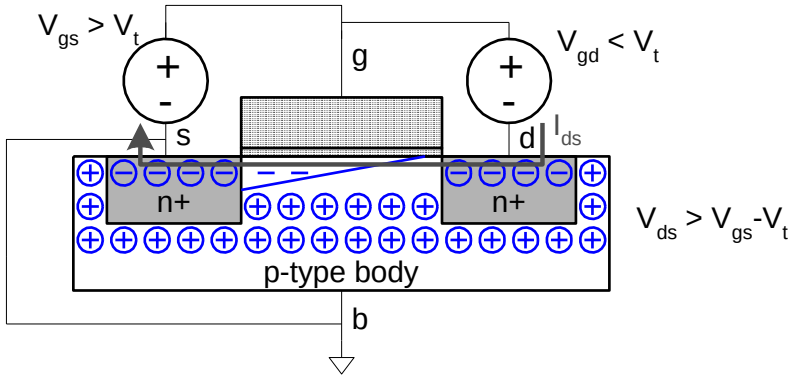
nMOS Linear

- 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

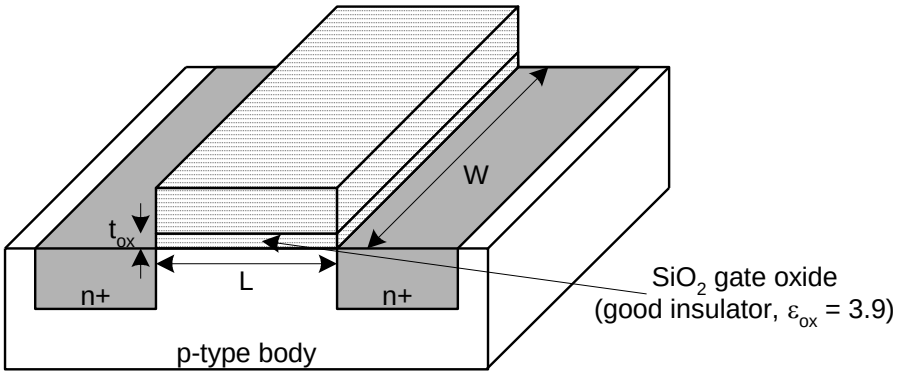


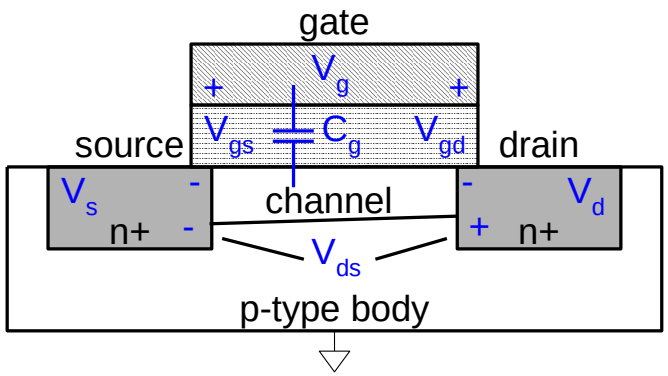
I-V Characteristics

- 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
- $Q_{channel} = CV$
- $C = C_g = \varepsilon_{ox} WL/t_{ox} = C_{ox} WL$ $\quad C_{ox} = \varepsilon_{ox} / t_{ox}$
- $V = V_{gc} - V_t = (V_{gs} - V_{ds}/2) - V_t$





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

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

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

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

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

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

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

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

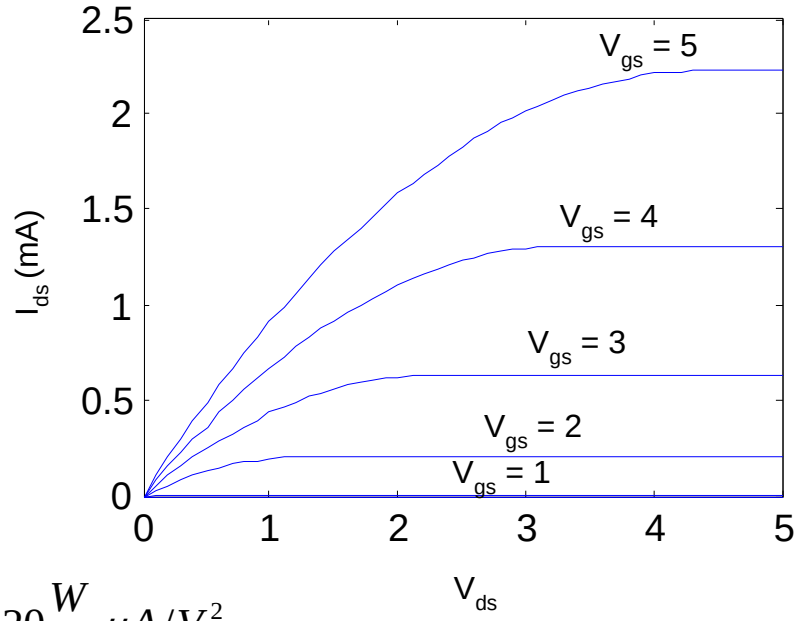
nMOS I-V Summary

- *Shockley* 1st order transistor models

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

Example

- We will be using a 0.6 μm process for your project
  - 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



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

pMOS I-V

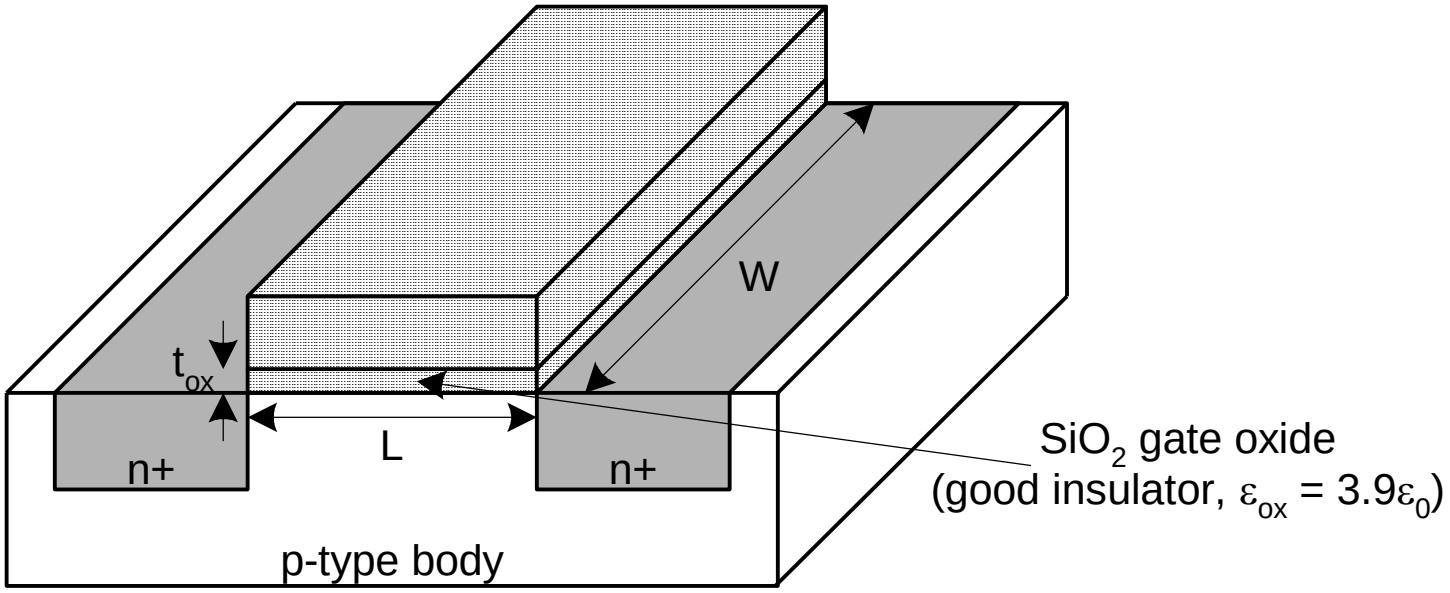
- All dopings and voltages are inverted for pMOS
- Mobility μ_p is determined by holes
  - Typically 2-3x lower than that of electrons μ_n
  - 120 $cm^2/V*s$ in AMI 0.6 μm process
- Thus pMOS must be wider to provide same current
  - In this class, assume $\mu_n / \mu_p = 2$
  - *** plot I-V here

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

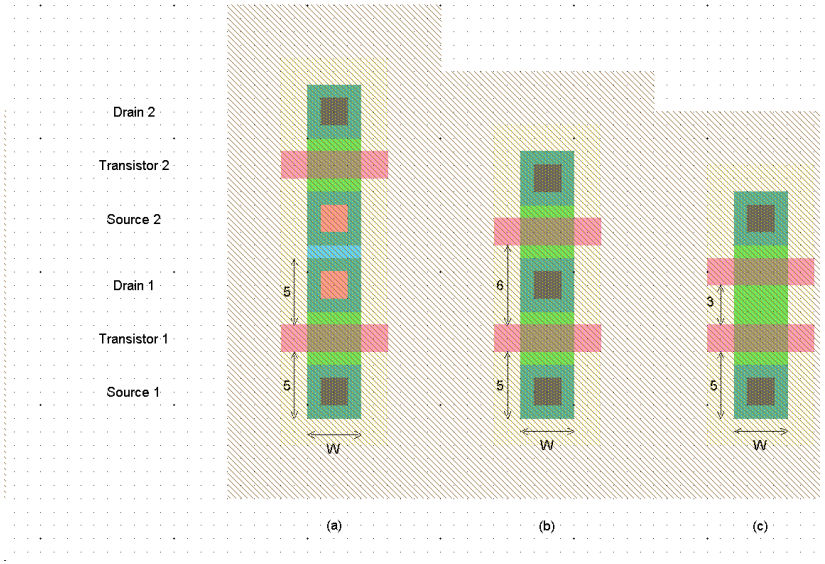
Gate Capacitance

- 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



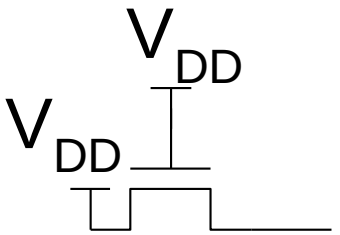
Diffusion Capacitance

- 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

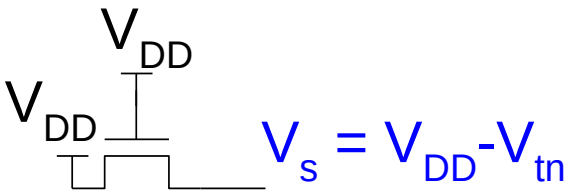


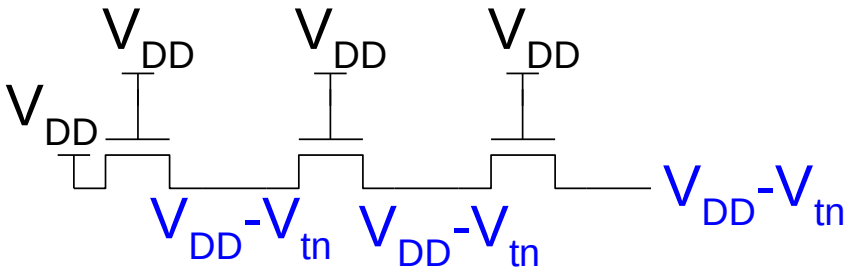
Pass Transistors

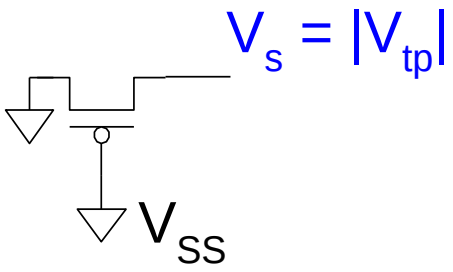
- We have assumed source is grounded
- What if source > 0?
  - e.g. pass transistor passing V_{DD}
- $V_g = V_{DD}$
  - If $V_s > V_{DD}-V_t$, $V_{gs} < V_t$
  - Hence transistor would turn itself off
- nMOS pass transistors pull no higher than $V_{DD}-V_{tn}$
  - Called a degraded "1"
  - Approach degraded value slowly (low I_{ds})
- pMOS pass transistors pull no lower than V_{tp}

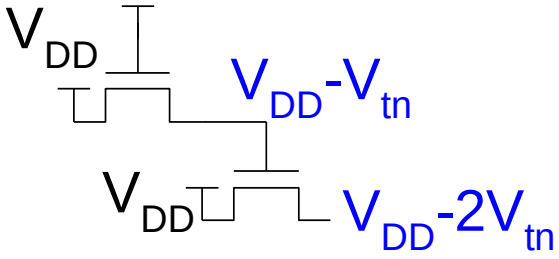


Pass Transistor Ckts







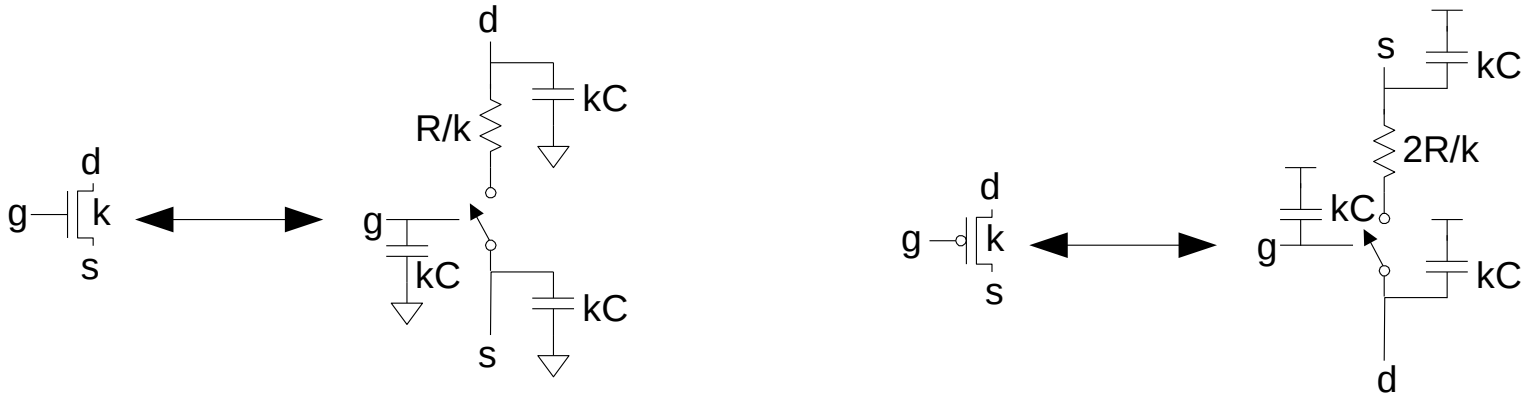


Effective Resistance

- 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

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

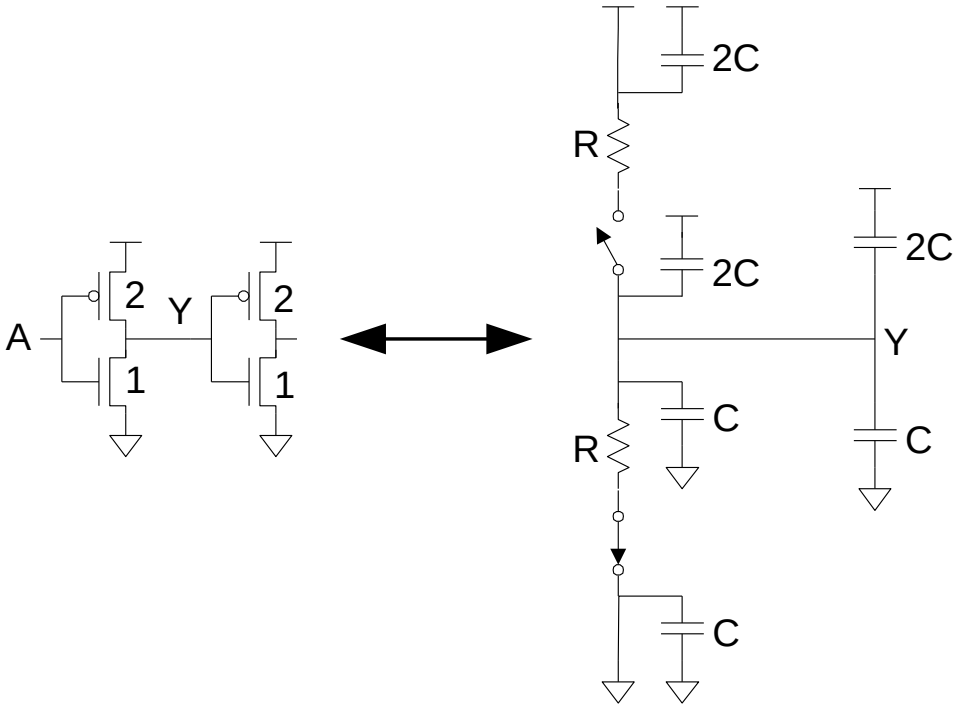


RC Values

- Capacitance
  - $C = C_g = C_s = C_d = 2$ 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

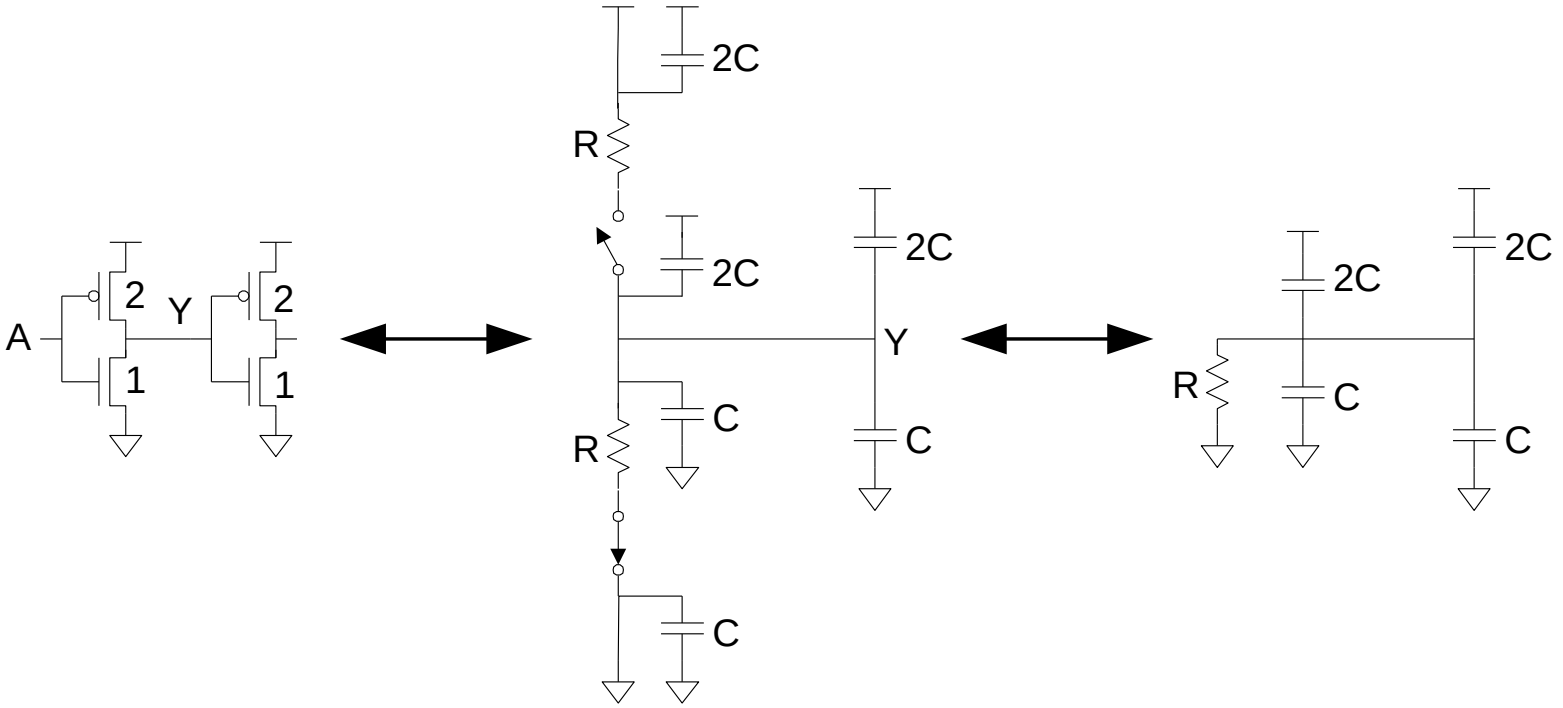
Inverter Delay Estimate

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



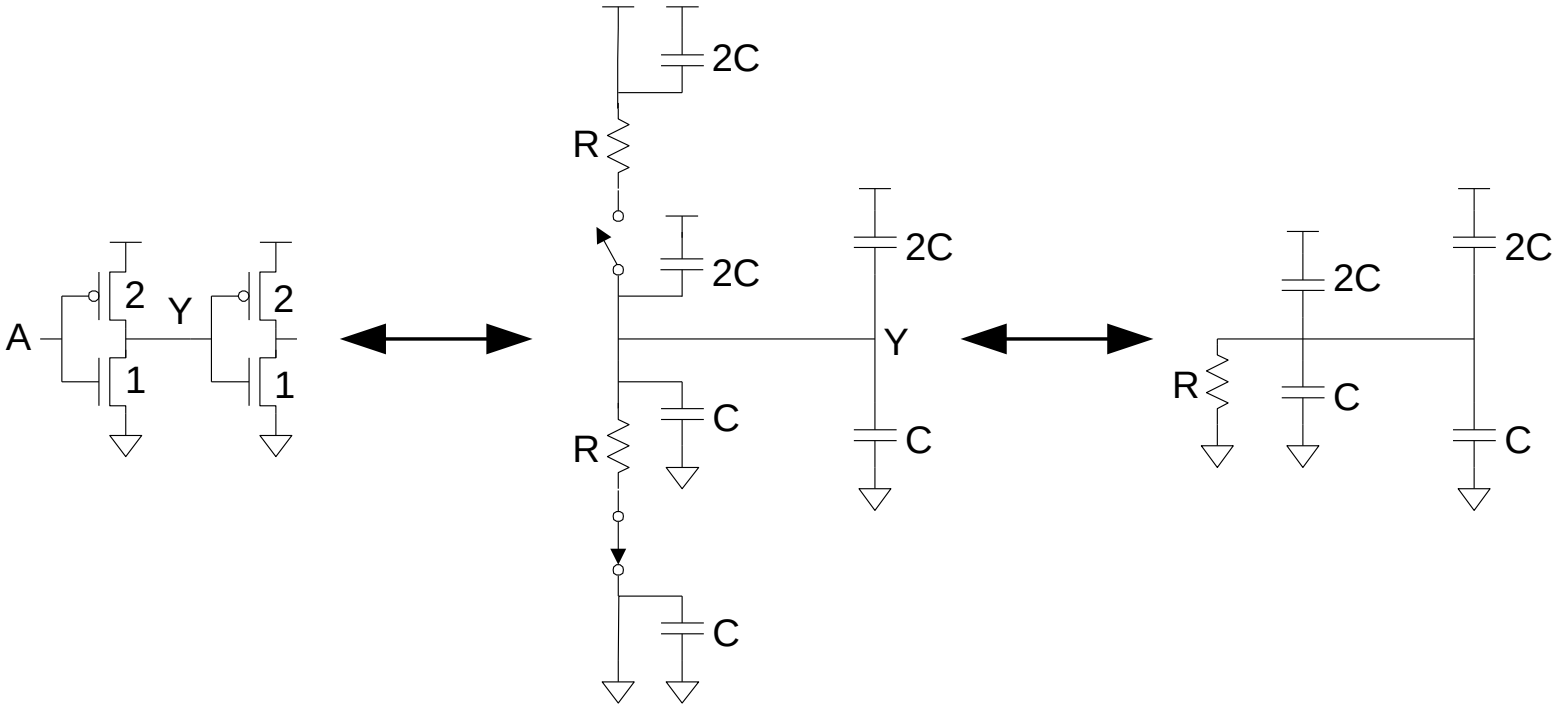
Inverter Delay Estimate

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



Inverter Delay Estimate

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



d = 6RC

Nonideal Transistors

Outline

- Transistor I-V Review
- Nonideal Transistor Behavior
  - Velocity Saturation
  - Channel Length Modulation
  - Body Effect
  - Leakage
  - Temperature Sensitivity
- Process and Environmental Variations
  - Process Corners

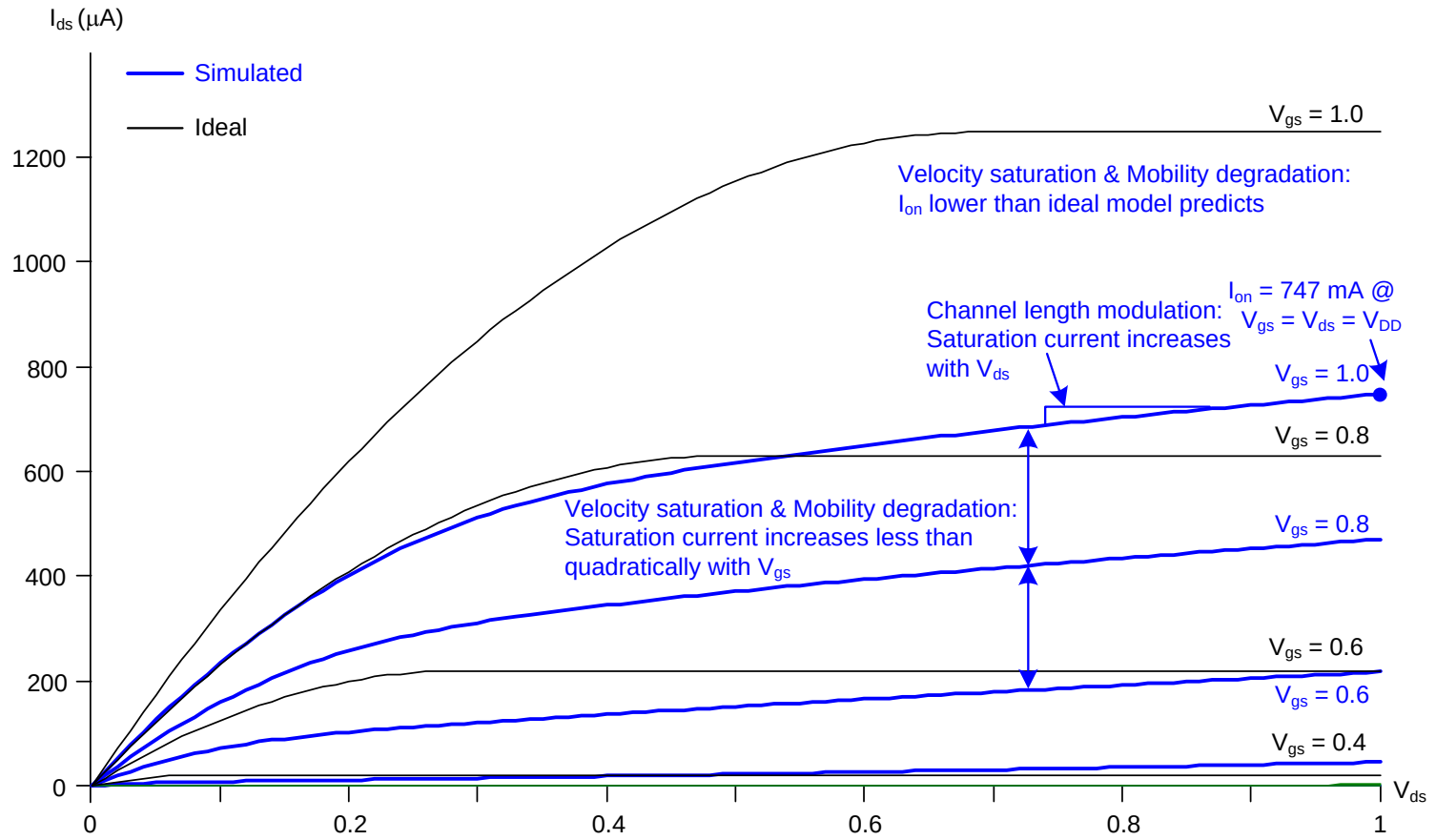
Ideal Transistor I-V

- *Shockley* 1st order transistor models
- accurate for only old technology (long-channel model)

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

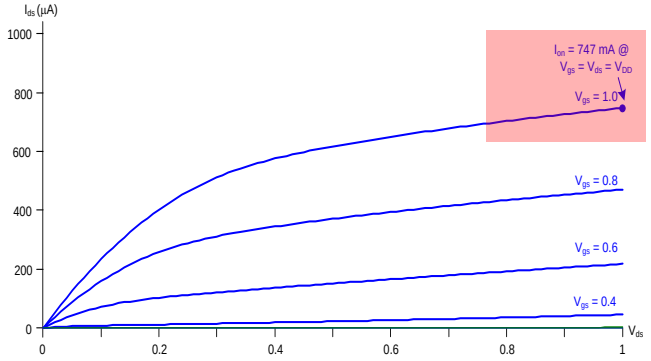
Ideal vs. Simulated nMOS I-V Plot

- 65 nm IBM process, V_{DD} = 1.0 V

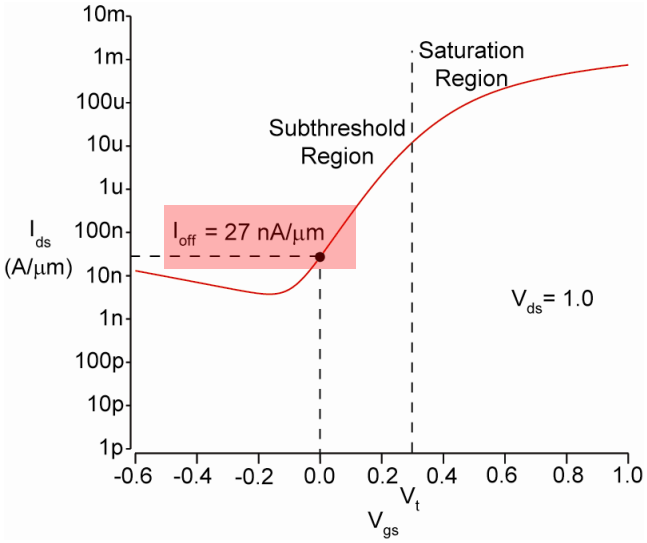


ON and OFF Current

- $I_{on} = I_{ds}$ @ $V_{gs} = V_{ds} = V_{DD}$
  - Saturation



- $I_{off} = I_{ds}$ @ $V_{gs} = 0$, $V_{ds} = V_{DD}$
  - Cutoff



Electric Fields Effects

- Vertical electric field: $E_{vert} = V_{gs} / t_{ox}$
  - Attracts carriers into channel
  - Long channel: $Q_{channel} \propto E_{vert}$
- Lateral electric field: $E_{lat} = V_{ds} / L$
  - Accelerates carriers from drain to source
  - Long channel: $v = \mu E_{lat}$

Coffee Cart Analogy

- Tired student runs from VLSI lab to coffee cart
- Freshmen are pouring out of the physics lecture hall
- V_{ds} is how long you have been up
  - Your velocity = fatigue × mobility
- V_{gs} is a wind blowing you against the glass (SiO_2) wall
- At high V_{gs}, you are buffeted against the wall
  - *Mobility degradation*
- At high V_{ds}, you scatter off freshmen, fall down, get up
  - *Velocity saturation*
    - Don't confuse this with the saturation region

Mobility Degradation

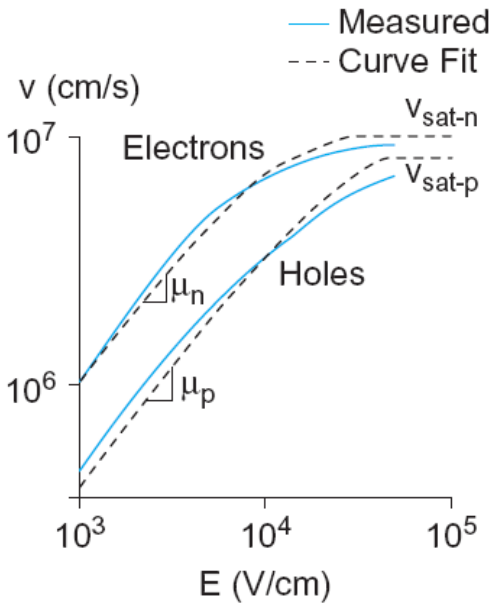
- High E_{vert} effectively reduces mobility
  - Collisions with oxide interface

$$\mu_{\text{eff}-n} = \frac{540\frac{\text{cm}^2}{\text{V}\cdot\text{s}}}{1+\left(\frac{V_{gs}+V_t}{0.54\frac{\text{V}}{\text{nm}}t_{\text{ox}}}\right)^{1.85}} \qquad \mu_{\text{eff}-p} = \frac{185\frac{\text{cm}^2}{\text{V}\cdot\text{s}}}{1+\frac{\left|V_{gs}+1.5V_t\right|}{0.338\frac{\text{V}}{\text{nm}}t_{\text{ox}}}}$$

Velocity Saturation

- At high E_{lat}, carrier velocity rolls off
  - Carriers scatter off atoms in silicon lattice
  - Velocity reaches v_{sat}
    - Electrons: 10^7 cm/s
    - Holes: 8 x 10^6 cm/s
  - Better model

$$v = \begin{cases} \dfrac{\mu_{\text{eff}} E}{1 + \dfrac{E}{E_c}} & E < E_c \\ v_{\text{sat}} & E \geq E_c \end{cases} \qquad E_c = \frac{2 v_{\text{sat}}}{\mu_{\text{eff}}}$$



Vel Sat I-V Effects

- Ideal transistor ON current increases with V_{DD}^2

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

- Velocity-saturated ON current increases with V_{DD}

$$I_{ds} = C_{\text{ox}} W \left(V_{gs} - V_t\right) v_{\max}$$

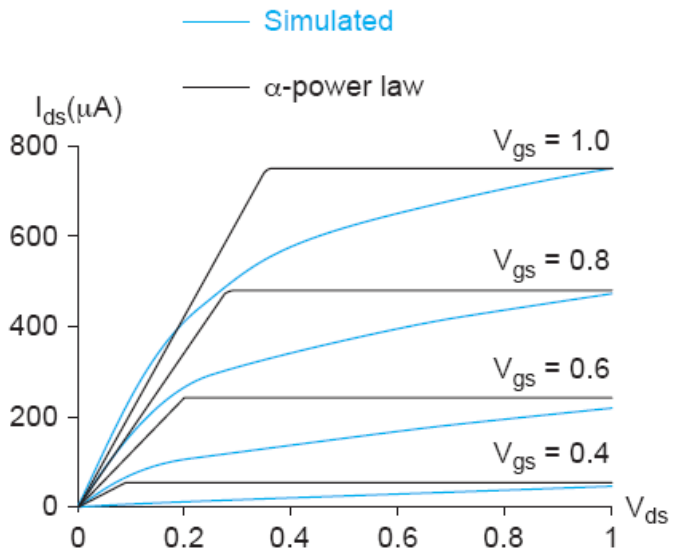
- Real transistors are partially velocity saturated
  - Approximate with α-power law model
  - $I_{ds} \propto V_{DD}^{\alpha}$
  - $1 < \alpha < 2$ determined empirically (≈ 1.3 for 65 nm)

α-Power Model

$$I_{ds} = \begin{cases} 0 & V_{gs} < V_t \quad \text{cutoff} \\ I_{dsat}\dfrac{V_{ds}}{V_{dsat}} & V_{ds} < V_{dsat} \quad \text{linear} \\ I_{dsat} & V_{ds} > V_{dsat} \quad \text{saturation} \end{cases}$$

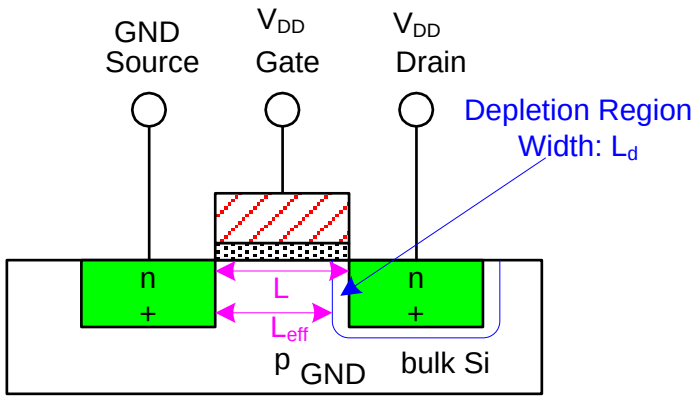
$$I_{dsat} = P_c \frac{\beta}{2}\left(V_{gs} - V_t\right)^{\alpha}$$

$$V_{dsat} = P_v \left(V_{gs} - V_t\right)^{\alpha/2}$$



Channel Length Modulation

- Reverse-biased p-n junctions form a *depletion region*
  - Region between n and p with no carriers
  - Width of depletion L_d region grows with reverse bias
  - $L_{eff} = L - L_d$
- Shorter L_{eff} gives more current
  - I_{ds} increases with V_{ds}
  - Even in saturation



Chan Length Mod I-V

$$I_{ds} = \frac{\beta}{2}\left(V_{gs} - V_t\right)^2 \left(1 + \lambda V_{ds}\right)$$

- λ = *channel length modulation coefficient*
  - not feature size
  - Empirically fit to I-V characteristics

Threshold Voltage Effects

- V_t is V_{gs} for which the channel starts to invert
- Ideal models assumed V_t is constant
- Really depends (weakly) on almost everything else:
  - Body voltage: *Body Effect*
  - Drain voltage: *Drain-Induced Barrier Lowering*
  - Channel length: *Short Channel Effect*

Body Effect

- Body is a fourth transistor terminal
- V_{sb} affects the charge required to invert the channel
  - Increasing V_s or decreasing V_b increases V_t

$$V_t = V_{t0} + \gamma\left(\sqrt{\phi_s + V_{sb}} - \sqrt{\phi_s}\right)$$

- ϕ_s = *surface potential* at threshold

$$\phi_s = 2v_T \ln\frac{N_A}{n_i}$$

  - Depends on doping level N_A
  - And intrinsic carrier concentration n_i
- γ = *body effect coefficient*

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

Body Effect Cont.

- For small source-to-body voltage, treat as linear

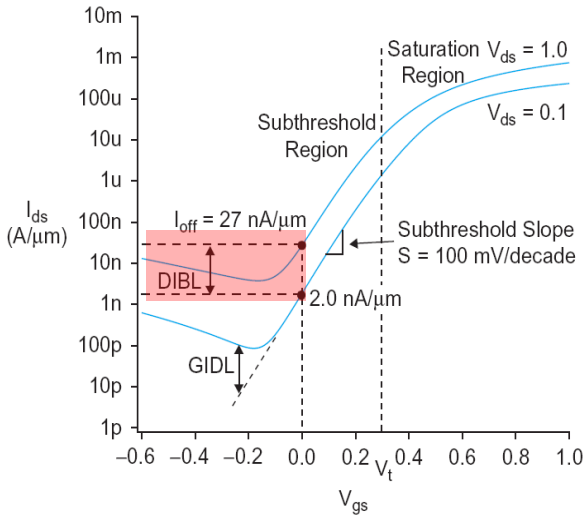
$$V_t = V_{t0} + k_\gamma V_{sb}$$

$$k_\gamma = \frac{\gamma}{2\sqrt{\phi_s}} = \frac{\sqrt{\frac{q\varepsilon_{si} N_A}{v_T \ln \frac{N_A}{n_i}}}}{2C_{ox}}$$

DIBL

- Electric field from drain affects channel
- More pronounced in small transistors where the drain is closer to the channel
- Drain-Induced Barrier Lowering
  - Drain voltage also affect V_t

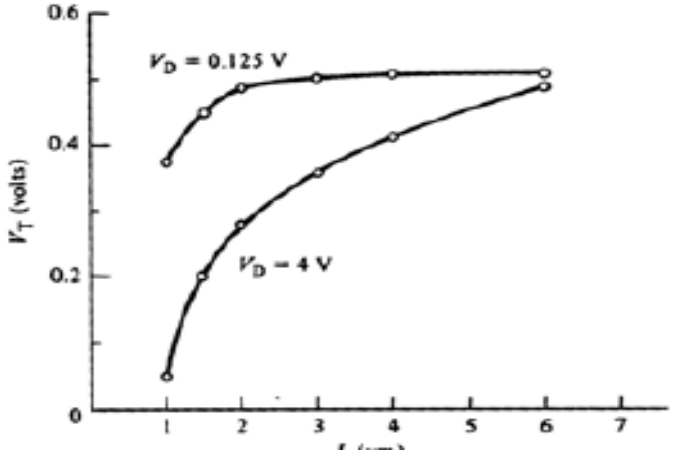
$$V_t' = V_t - \eta V_{ds}$$

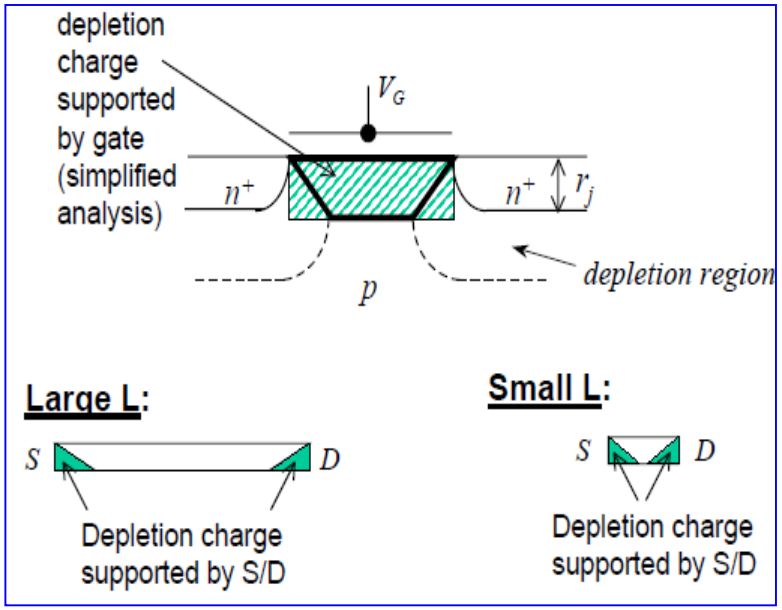


- High drain voltage causes current to increase.

Short Channel Effect

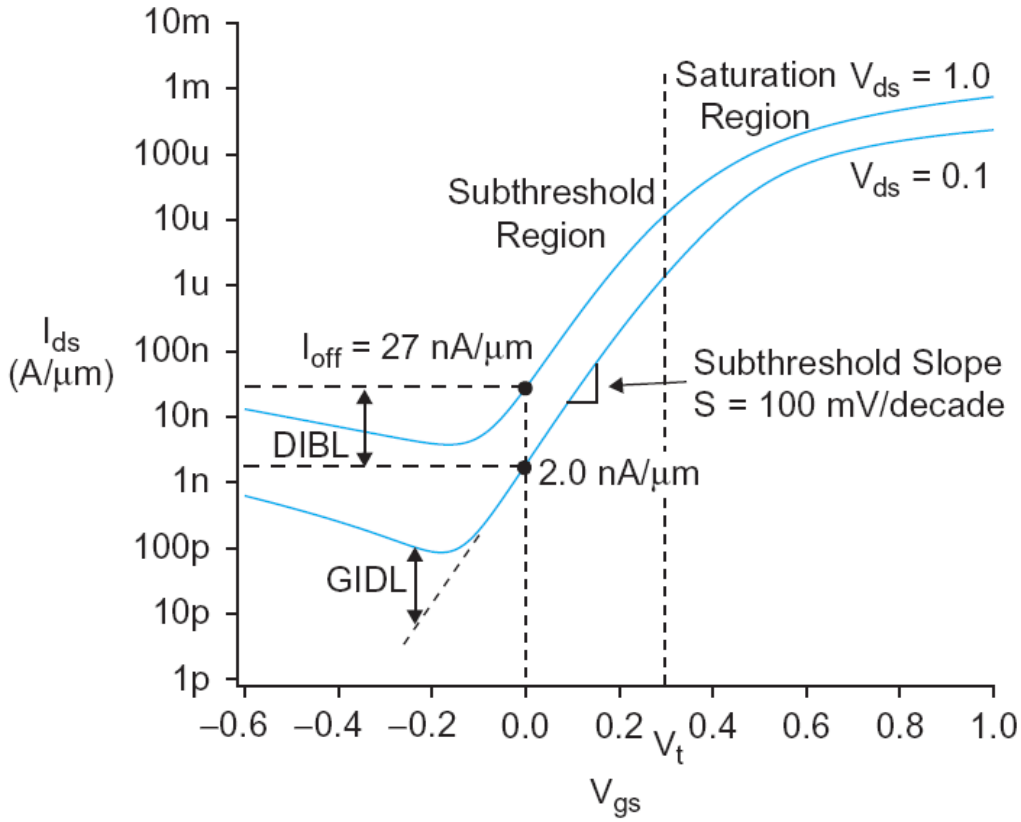
- V_t roll-off: decreases with L
  - Exacerbated by high Vds
  - Undesirable (constant V_t is better for circuit designers!!)
- In small transistors, source/drain depletion regions extend into the channel
  - Impacts the amount of charge required to invert the channel
  - And thus makes V_t a function of channel length
  - The smaller the L, the greater % of charge balanced by the S/D pn junctions
  - Some processes exhibit a reverse short channel effect in which V_t increases with L (due to halo doping)





Leakage

- What about current in cutoff?
- Simulated results
- What differs?
  - Current doesn't go to 0 in cutoff



Leakage Sources

- Subthreshold conduction
  - Transistors can't abruptly turn ON or OFF
  - Dominant source in contemporary transistors
- Gate leakage
  - Tunneling through ultrathin gate dielectric
- Junction leakage
  - Reverse-biased PN junction diode current

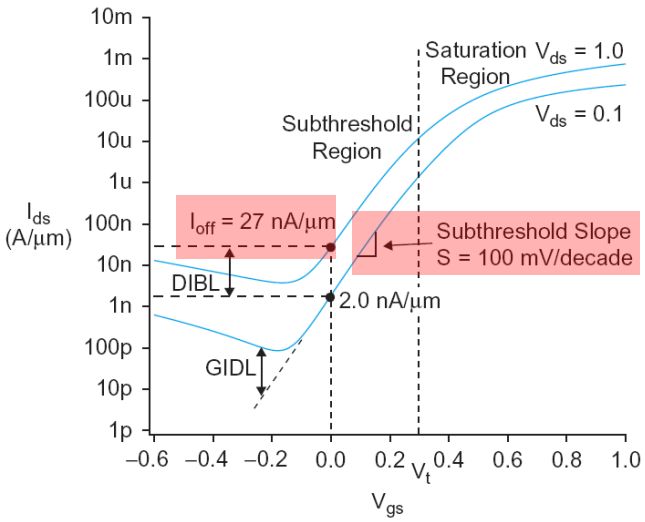
Subthreshold Leakage

- Subthreshold leakage exponential with V_{gs}

$$I_{ds} = I_{ds0} e^{\frac{V_{gs} - V_{t0} + \eta V_{ds} - k_\gamma V_{sb}}{n v_T}} \left(1 - e^{\frac{-V_{ds}}{v_T}}\right)$$

- I_{ds0} current at threshold V_{t0}
- n is process dependent
  - typically 1.3-1.7
- Rewrite relative to I_{off} on log scale

$$I_{ds} = I_{off} 10^{\frac{V_{gs} + \eta\left(V_{ds} - V_{dd}\right) - k\gamma V_{sb}}{S}} \left(1 - e^{\frac{-V_{ds}}{v_t}}\right)$$



- S ≈ 100 mV/decade @ room temperature

$$S = \left[\frac{d\left(\log_{10} I_{ds}\right)}{dV_{gs}}\right]^{-1} = n v_T \ln 10$$

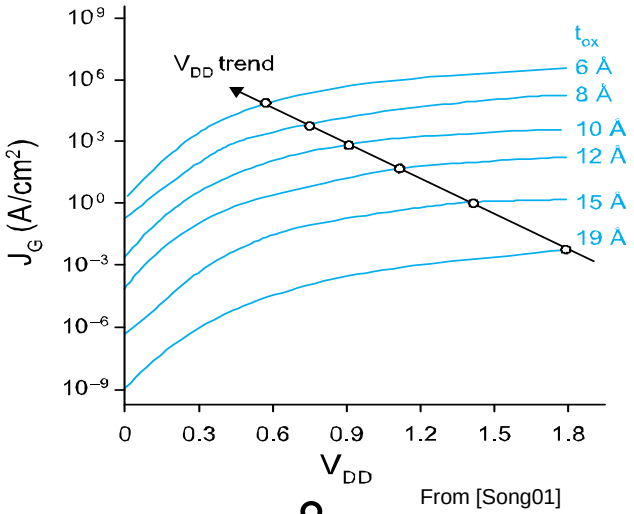
- Why?: The smaller L, the smaller voltage swing between ON & OFF → smaller barrier →By thermal emission of carriers over the potential barrier by V_t → OFF to ON (weak inversion)!!

Gate Leakage

- Carriers tunnel through very thin gate oxides
- Exponentially sensitive to t_{ox} and V_{DD}

$$I_{gate} = WA\left(\frac{V_{DD}}{t_{ox}}\right)^2 e^{-B\frac{t_{ox}}{V_{DD}}}$$

  - A and B are tech constants
  - Greater for electrons
    - So nMOS gates leak more

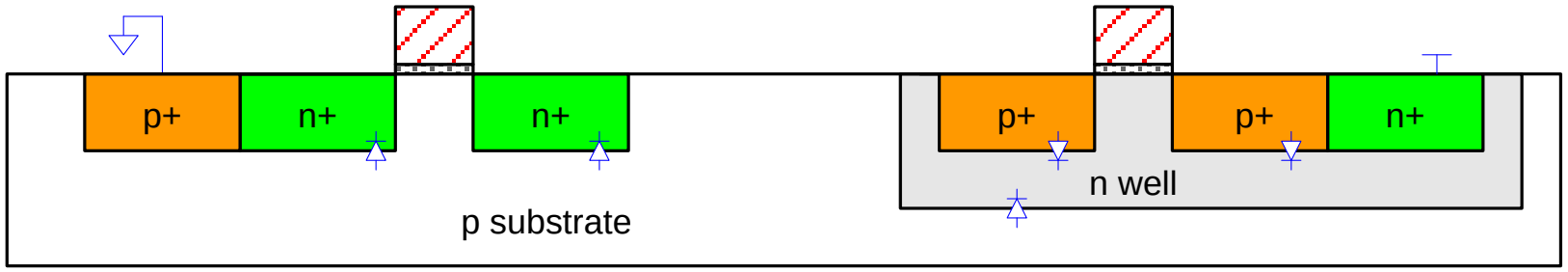


From [Song01]

- Negligible for older processes (t_{ox} > 20 Å)
- Critically important at 65 nm and below (t_{ox} ≈ 10.5 Å)

Junction Leakage

- Reverse-biased p-n junctions have some leakage
  - Ordinary diode leakage
  - Band-to-band tunneling (BTBT)
  - Gate-induced drain leakage (GIDL)



Diode Leakage

- Reverse-biased p-n junctions have some leakage

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

- At any significant negative diode voltage, $I_D = -I_s$
- I_s depends on doping levels
  - And area and perimeter of diffusion regions
  - Typically < 1 fA/μm^2 (negligible)

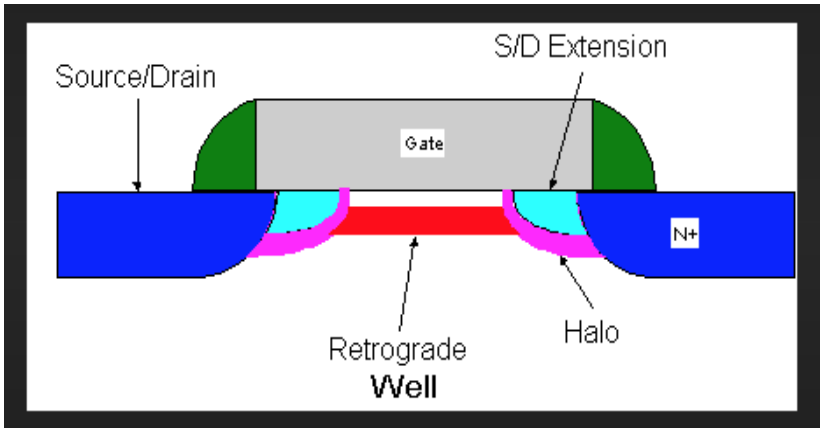
Band-to-Band Tunneling

- Tunneling across heavily doped p-n junctions
  - Especially sidewall between drain & channel when *halo doping* is used to increase V_t
- Increases junction leakage to significant levels

$$I_{BTBT} = WX_j A \frac{E_j}{E_g^{0.5}} V_{dd} \, e^{-B\frac{E_g^{1.5}}{E_j}}$$

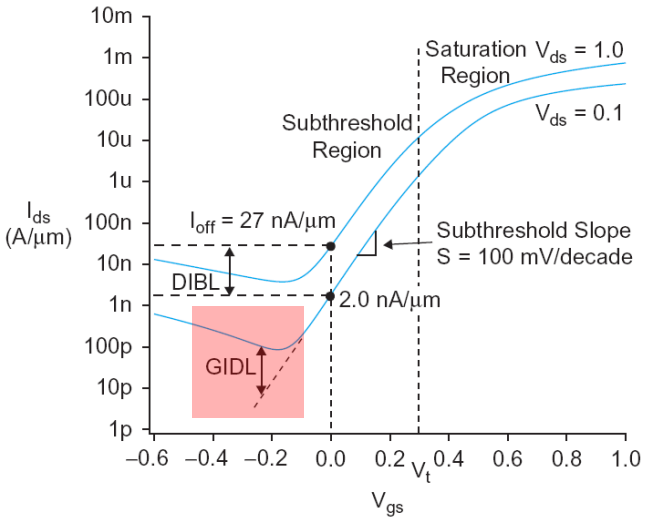
$$E_j = \sqrt{\frac{2qN_{halo}N_{sd}}{\varepsilon\left(N_{halo}+N_{sd}\right)}}\left(V_{DD} + v_T \ln\frac{N_{halo}N_{sd}}{n_i^2}\right)$$

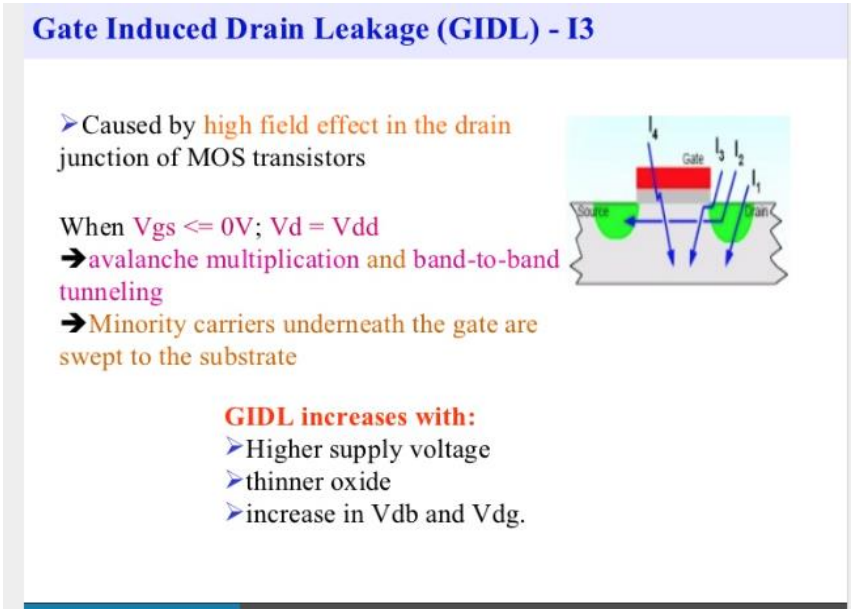
  - X_j: sidewall junction depth
  - E_g: bandgap voltage
  - A, B: tech constants



Gate-Induced Drain Leakage

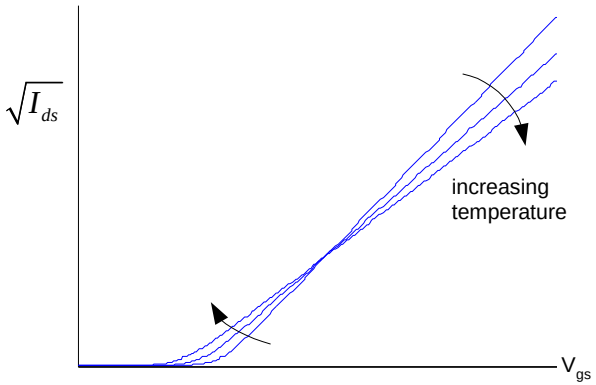
- Occurs at overlap between gate and drain
  - Most pronounced when drain is at V_{DD}, gate is at a negative voltage
  - Thwarts efforts to reduce subthreshold leakage using a negative gate voltage





Temperature Sensitivity

- Increasing temperature
  - Reduces mobility
  - Reduces V_t (see eq. 2.52 p. 85)
- I_{ON} decreases with temperature
- I_{OFF} increases with temperature

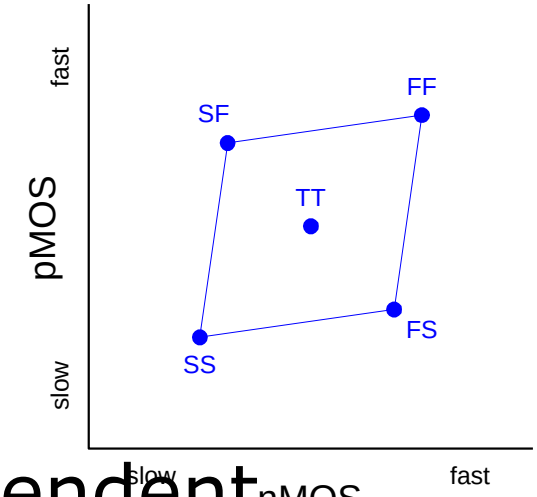


So What?

- So what if transistors are not ideal?
  - They still behave like switches.
- But these effects matter for…
  - Supply voltage choice
  - Logical effort
  - Quiescent power consumption
  - Pass transistors
  - Temperature of operation

Parameter Variation

- Transistors have uncertainty in parameters
  - Process: L_{eff}, V_t, t_{ox} of nMOS and pMOS
  - Vary around typical (T) values
- Fast (F)
  - L_{eff}: short
  - V_t: low
  - t_{ox}: thin
- Slow (S): opposite
- Not all parameters are independent for nMOS and pMOS



Environmental Variation

- V_{DD} and T also vary in time and space
- Fast:
  - V_{DD}: high
  - T: low

Corner	Voltage	Temperature
F		
T	1.8	70 C
S		

Process Corners

- Process corners describe worst case variations
  - If a design works in all corners, it will probably work for any variation.
- Describe corner with four letters (T, F, S)
  - nMOS speed
  - pMOS speed
  - Voltage
  - Temperature

Important Corners

- Some critical simulation corners include

Purpose	**nMOS**	**pMOS**	**V_{DD}**	**Temp**
Cycle time				
Power				
Subthreshold leakage				

- As temperature increases, Tr is slower, Vt is lower(see eq. 2.52 p. 85) and leakage is increased!!

Variation, Noise, and Reliability

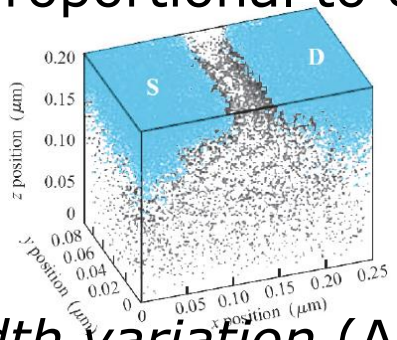
Variation

- Process
  - Threshold
  - Channel length
  - Interconnect dimensions
- Environment
  - Voltage
  - Temperature
- Aging / Wearout

Process Variation

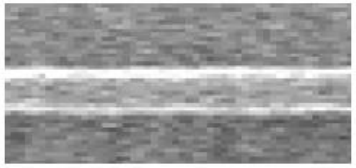
- Threshold Voltage
  - Depends on placement of dopants in channel
  - Standard deviation inversely proportional to channel area

$$\sigma_{V_t} = \frac{t_{ox}}{\varepsilon_{ox}} \frac{\sqrt[4]{q^3 \varepsilon_{si} \phi_b N_a}}{\sqrt{2LW}} = \frac{A_{V_t}}{\sqrt{LW}}$$



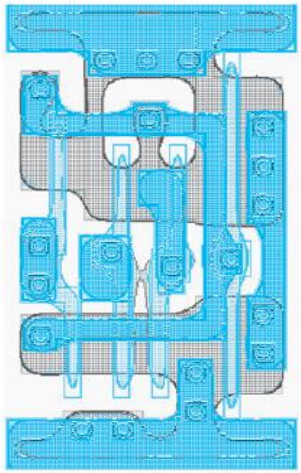
[Bernstein06]

- Channel Length
  - Systematic *across-chip linewidth variation* (ACLV)
  - Random line edge roughness (LER)



Courtesy Texas Instruments

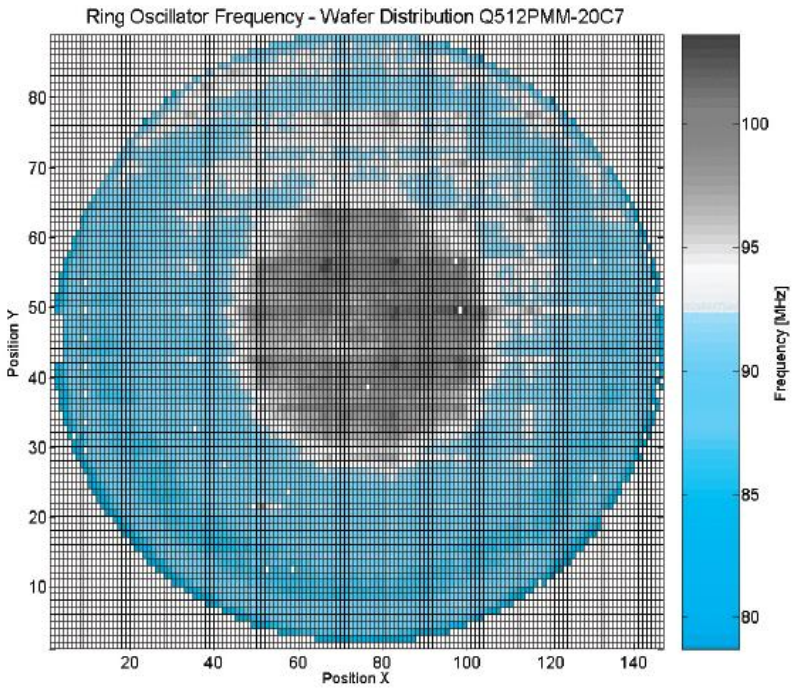
- Interconnect
  - Etching variations affect w, s, h



Courtesy Larry Pileggi

Spatial Distribution

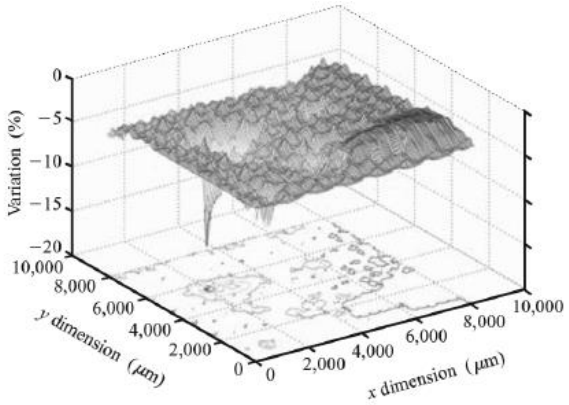
- Variations show spatial correlation
  - *Lot-to-lot* (L2L)
  - *Wafer-to-wafer* (W2W)
  - *Die-to-die* (D2D) / *inter-die*
  - *Within-die* (WID) / *intradie*
- Closer transistors match better



Courtesy M. Pelgrom

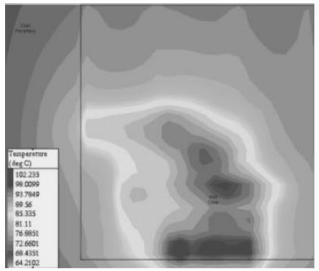
Environmental Variation

- Voltage
  - V_{DD} is usually designed +/- 10%
  - Regulator error
  - On-chip droop from switching activity
- Temperature
  - Ambient temperature ranges
  - On-die temperature elevated by chip power consumption



Courtesy IBM

Standard	Minimum	Maximum
Commercial	0 °C	70 °C
Industrial	−40 °C	85 °C
Military	−55 °C	125 °C



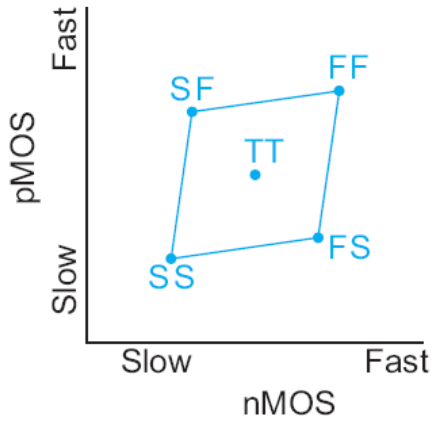
[Harris01b]

Aging

- Transistors change over time as they wear out
  - Hot carriers
  - Negative bias temperature instability
  - Time-dependent dielectric breakdown
- Causes threshold voltage changes
- More on this later…

Process Corners

- Model extremes of process variations in simulation
- Corners
  - Typical (T)
  - Fast (F)
  - Slow (S)
- Factors
  - nMOS speed
  - pMOS speed
  - Wire
  - Voltage
  - Temperature



Corner	Voltage	Temperature
F	1.98	0 °C
T	1.8	70 °C
S	1.62	125 °C

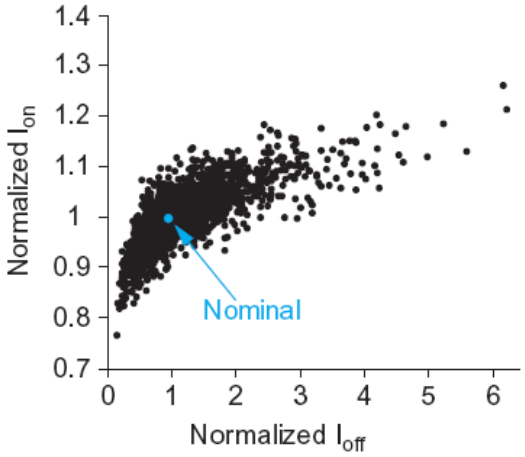
Corner Checks

- Circuits are simulated in different corners to verify different performance and correctness specifications

Corner					Purpose
nMOS	pMOS	Wire	V_{DD}	Temp	
T	T	T	S	S	Timing specifications (binned parts)
S	S	S	S	S	Timing specifications (conservative)
F	F	F	F	F	Race conditions, hold time constraints, pulse collapse, noise
S	S	?	F	S	Dynamic power
F	F	F	F	S	Subthreshold leakage noise and power, overall noise analysis
S	S	F	S	S	Races of gates against wires
F	F	S	F	F	Races of wires against gates
S	F	T	F	F	Pseudo-nMOS and ratioed circuits noise margins, memory read/write, race of pMOS against nMOS
F	S	T	F	F	Ratioed circuits, memory read/write, race of nMOS against pMOS

Monte Carlo Simulation

- As process variation increases, the worst-case corners become too pessimistic for practical design
- Monte Carlo: repeated simulations with parameters randomly varied each time
- Look at scatter plot of results to predict yield
- Ex: impact of V_t variation
  - ON-current
  - leakage

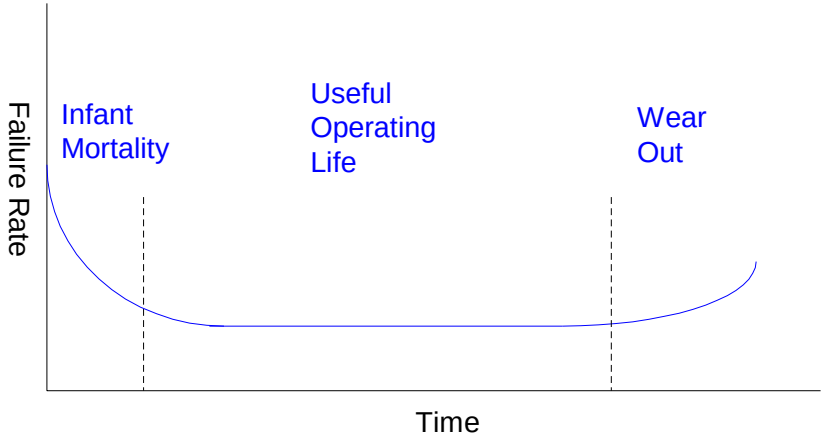


Noise

- Sources
  - Power supply noise / ground bounce
  - Capacitive coupling
  - Charge sharing
  - Leakage
  - Noise feedthrough
- Consequences
  - Increased delay (for noise to settle out)
  - Or incorrect computations

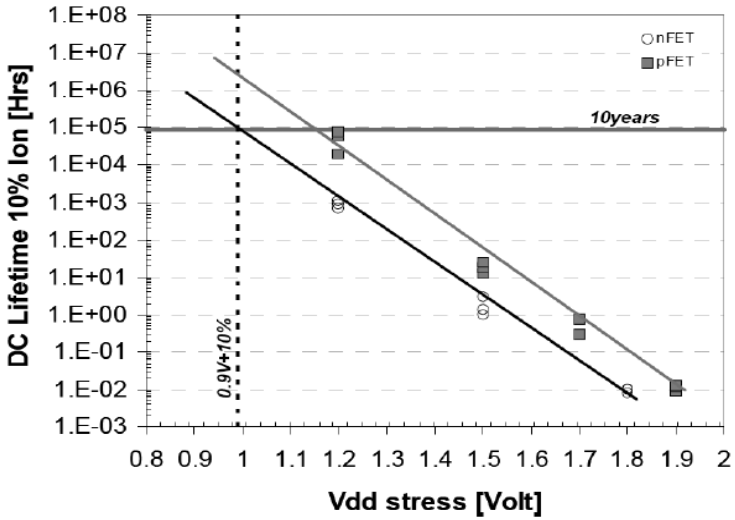
Reliability

- Hard Errors
  - Oxide wearout
  - Interconnect wearout
  - Overvoltage failure
  - Latchup
- Soft Errors
- Characterizing reliability
  - Mean time between failures (MTBF)
    - # of devices x hours of operation / number of failures
  - Failures in time (FIT)
    - # of failures / thousand hours / million devices



Accelerated Lifetime Testing

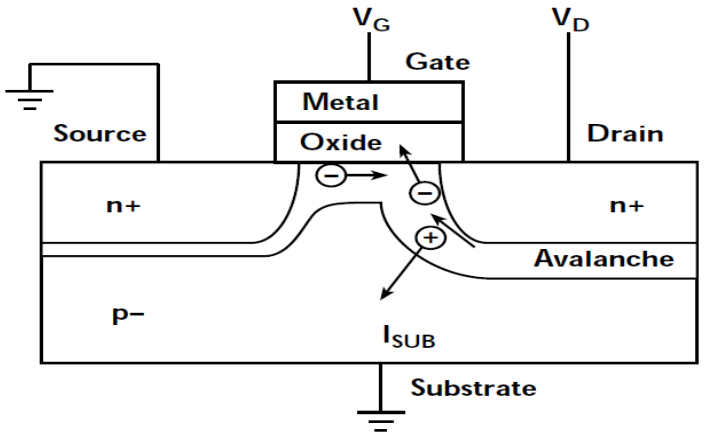
- Expected reliability typically exceeds 10 years
- But products come to market in 1-2 years
- Accelerated lifetime testing required to predict adequate long-term reliability (gate oxide of IBM 32nm process)



[Arnaud08]

Hot Carriers (Electron)

- As Trs switch, electric fields across channel impart high energies to some carriers
  - These "hot" carriers may be blasted into the gate oxide where they become trapped
  - Accumulation of charge in oxide causes shift in V_t over time
  - Eventually V_t shifts too far for devices to operate correctly
  - Current: nMOS→ decreased, pMOS → increased
  - Cause "slow device"
- Choose V_{DD} to achieve reasonable product lifetime
  - Worst problems for inverters and NORs with fast rising inputs and long propagation delays
    - Max when Isub is large, which typically occurs when nMOS TRs are in saturation while the input rises.

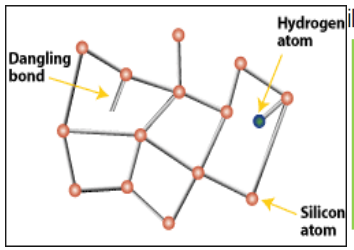


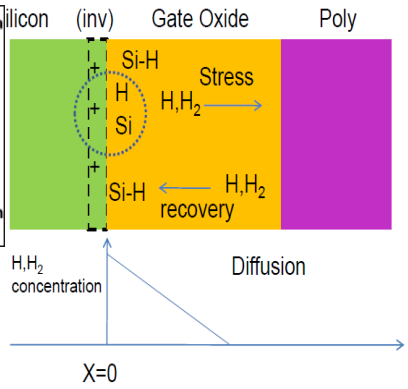
Today's ULSI MOSFET devices feature extremely short channel lengths and high electric fields. In these high electric fields, carriers are accelerated to high velocities, reaching a maximum kinetic energy (hot) near the device drain. If the carrier energy is high enough, impact ionization can occur (Drain Avalanche Hot Carrier Effect), creating electron-hole pairs ***(Figure 1)***. These holes and electrons can attain enough energy to surmount the Si–SiO2 barrier and become trapped in the gate oxide. Trapped charges cause device degradation and enhanced substrate current (ISUB).

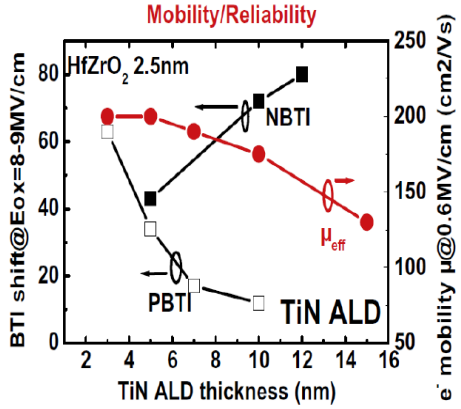
NBTI

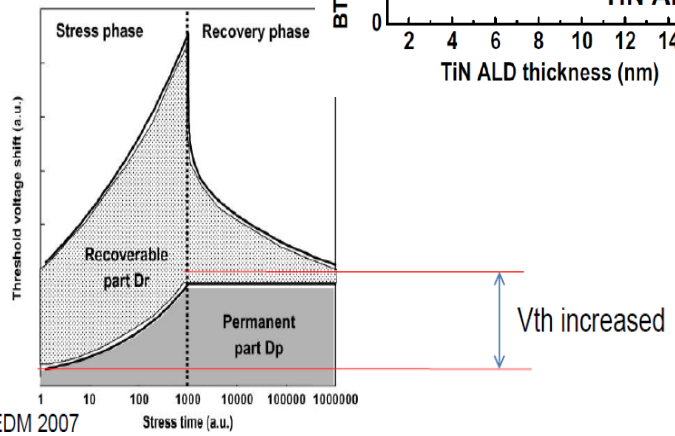
- *Negative bias temperature instability*
- Electric field applied across oxide forms dangling bonds called traps at Si-SiO_2 interface
- Accumulation of traps causes V_t shift
- Most pronounced for pMOS transistors with strong negative bias (V_g = 0, V_s = V_{DD}) at high temperature
  - Vth increase with negative bias, Vgs=-Vdd
  - Recover with zero bias, Vgs=0
  - 7% to 10% frequency degradation

$$\Delta V_t = k e^{\frac{E_{ox}}{E_0}} t^{0.25} \qquad E_{ox} = V_{DD}/t_{ox}$$









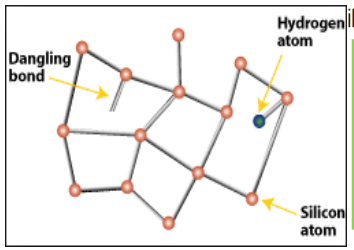
NBTI(*Negative bias temperature instability*)

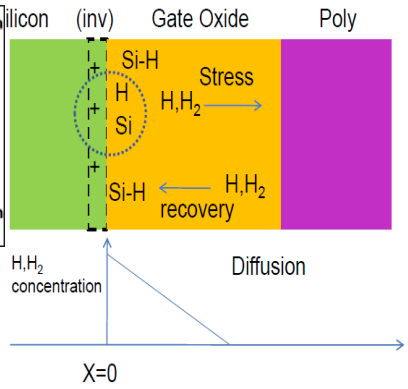
- Most pronounced for pMOS transistors with strong negative bias (Vg = 0, Vs = VDD) at high temperature
  - Vth increase with negative bias, Vgs=-Vdd
  - Recover with zero bias, Vgs=0
  - 7% to 10% frequency degradation 0

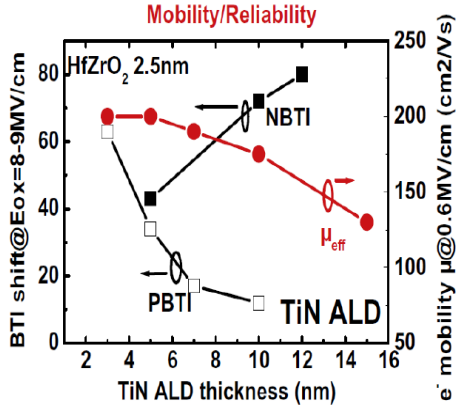
$$\Delta V_t = k e^{\frac{E_{ox}}{E_0}} t^{0.25} \qquad E_{ox} = V_{DD}/t_{ox}$$

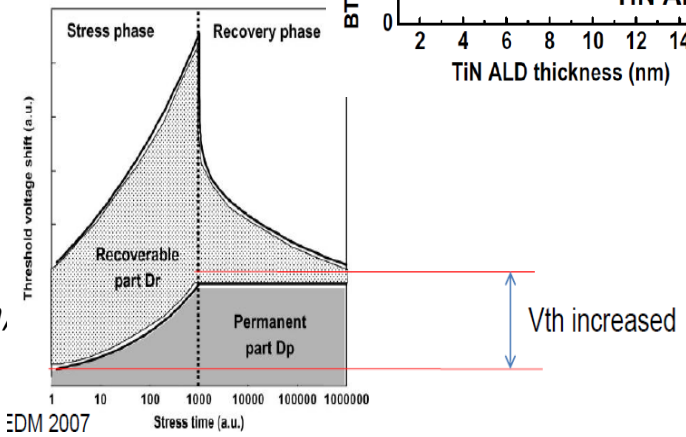
- At Si-SiO_2 interface, Si is trapped at the top, passivated by O and Hydron gas, Si-H bond are formed
- At strong negative bias, Si-H bonds are broken, H is moving toward Si or SiO2 and holes are trapped at the interface
- Accumulation of traps causes V_t to shift

- *Refer to "modeling and simulation of NBTI", Phd dissertation, Robert Entner.*







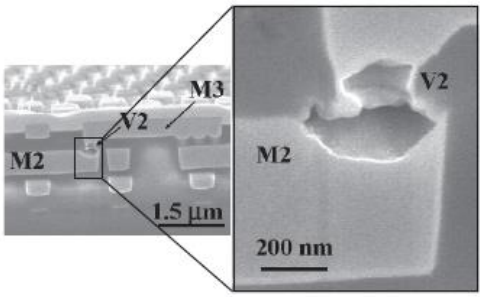


TDDB

- *Time-dependent dielectric breakdown*
  - Gradual increase in gate leakage when an electric field is applied across an oxide
  - a.k.a *stress-induced leakage current*
- For 10-year life at 125 C, keep E_{ox} below ~0.7 V/nm

Electromigration

- "Electron wind" causes movement of metal atoms along wires
- Excessive electromigration leads to open circuits
- Most significant for unidirectional (DC) current
  - Depends on current density J_{dc} (current / area)
  - Exponential dependence on temperature
  - Black's Equation: $$MTTF \propto \frac{e^{\frac{E_a}{kT}}}{J_{dc}^{\ n}}$$
  - Typical limits: $J_{dc} < 1 - 2$ mA / μm^2



[Christiansen06]

See the videos at http://www.cmosvlsi.com

Self-Heating

- Current through wire resistance generates heat
  - Oxide surrounding wires is a thermal insulator
  - Heat tends to build up in wires
  - Hotter wires are more resistive, slower
- Self-heating limits AC current densities for reliability

$$I_{rms} = \sqrt{\frac{\int_0^T I(t)^2 dt}{T}}$$

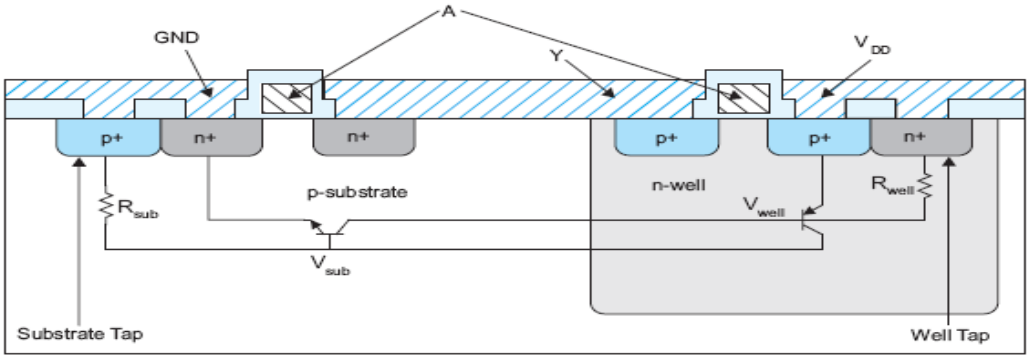
  - Typical limits: $J_{rms} < 15$ mA / μm^2

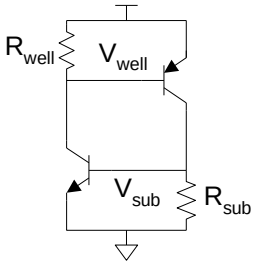
Overvoltage Failure

- High voltages can blow out tiny transistors
- *Electrostatic discharge* (ESD)
  - kilovolts from static electricity when the package pins are handled
- *Oxide breakdown*
  - In a 65 nm process, $V_g \approx 3$ V causes *arcing* through thin gate oxides
- *Punchthrough*
  - High V_{ds} causes depletion region between source and drain to touch, leading to high current flow and destructive overheating

Latchup

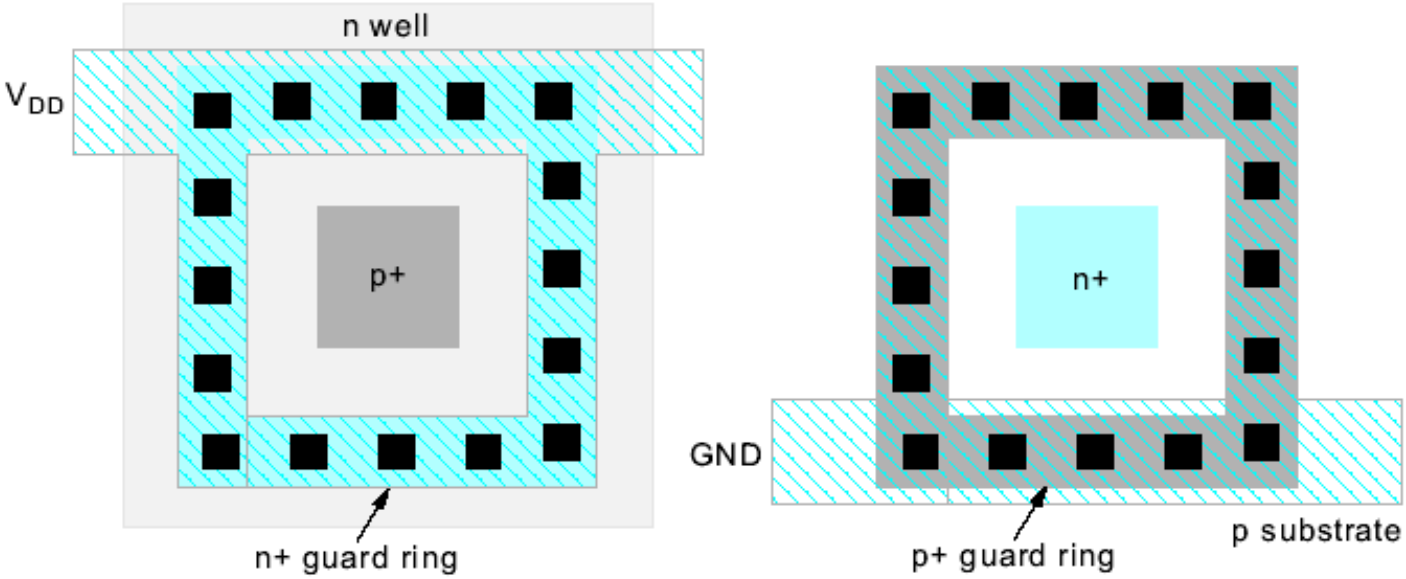
- Latchup: positive feedback leading to V_{DD} – GND short
  - Major problem for 1970's CMOS processes before it was well understood
- Avoid by minimizing resistance of body to GND / V_{DD}
  - Use plenty of substrate and well taps





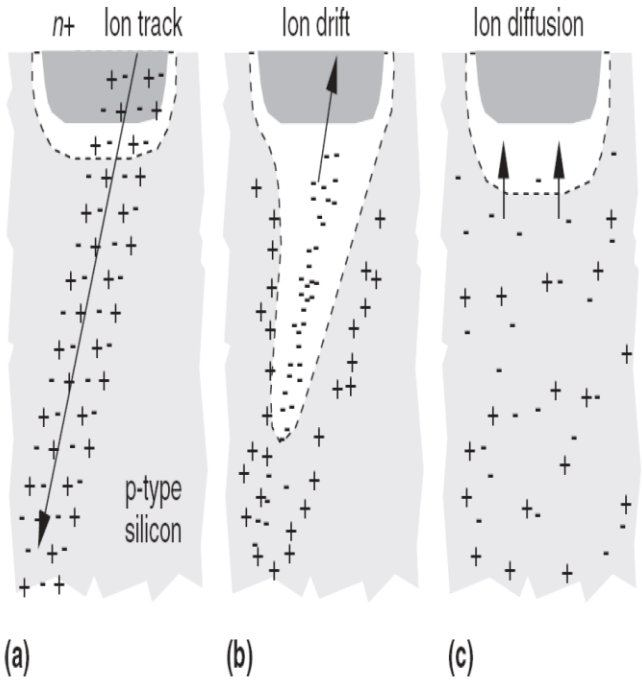
Guard Rings

- Latchup risk greatest when diffusion-to-substrate diodes could become forward-biased
- Surround sensitive region with guard ring to collect injected charge



Soft Errors

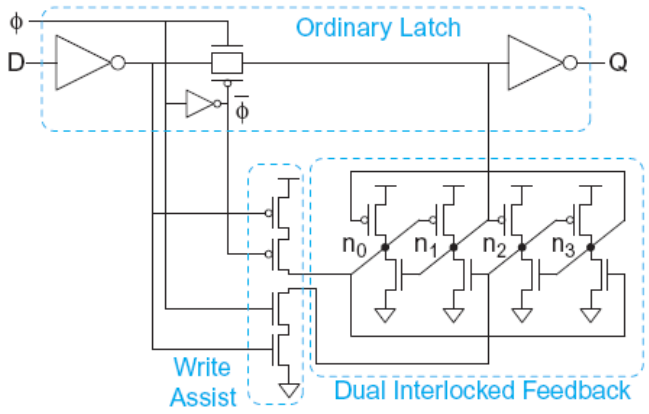
- In 1970's, DRAMs were observed to occasionally flip bits for no apparent reason
  - Ultimately linked to alpha particles and cosmic rays
- Collisions with particles create electron-hole pairs in substrate
  - These carriers are collected on dynamic nodes, disturbing the voltage
- Minimize soft errors by having plenty of charge on dynamic nodes
- Tolerate errors through ECC, redundancy



[Baumann05]

Radiation Hardening

- Radiation hardening reduces soft errors
  - Increase node capacitance to minimize impact of collected charge
  - Or use redundancy
  - E.g. *dual-interlocked cell*



- Error-correcting codes
  - Correct for soft errors that do occur

Summary

- Static CMOS gates are very robust
  - Will settle to correct value if you wait long enough
- Other circuits suffer from a variety of pitfalls
  - Tradeoff between performance & robustness
- Very important to check circuits for pitfalls
  - For large chips, you need an automatic checker.
  - Design rules aren't worth the paper they are printed on unless you back them up with a tool.