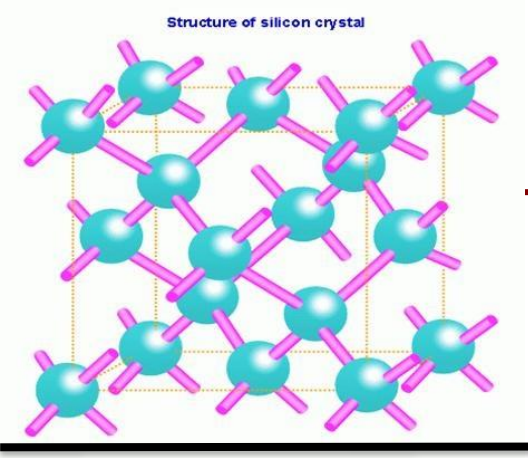
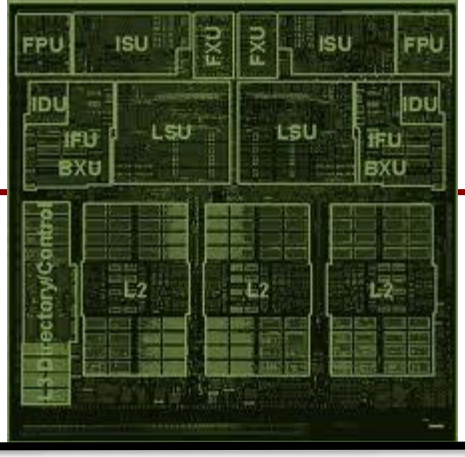
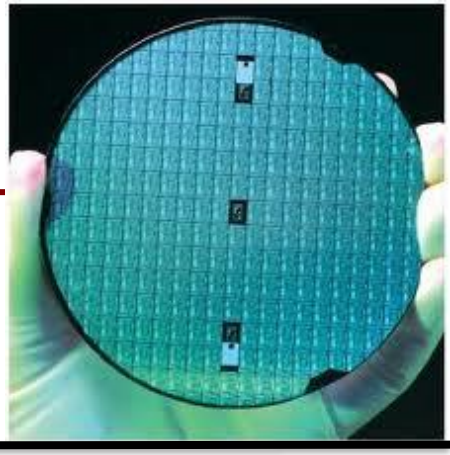






ECE 225

Lecture 11

Interconnect Design under Thermal & Reliability Constraints / Power Dissipation & Minimum Operation Voltage

Prof. Kaustav Banerjee

Electrical and Computer Engineering

University of California, Santa Barbara

Interconnect Thermal Effects and Reliability

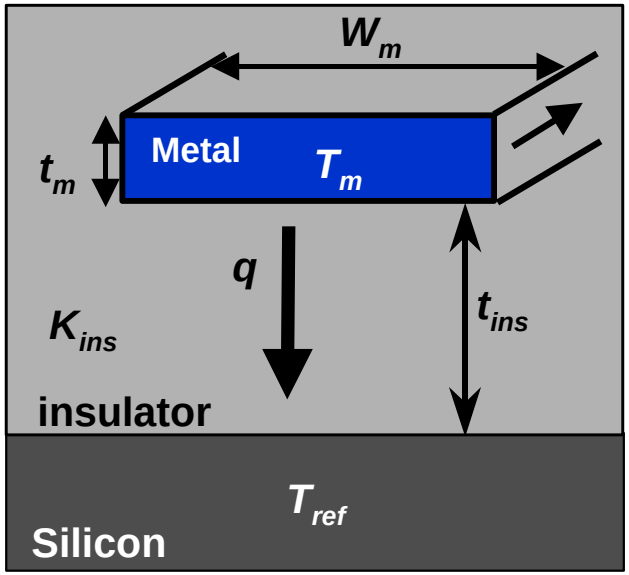
K. Banerjee and A. Mehrotra, IEEE Circuits and Devices Magazine, pp. 16-32, Sept. 2001.

Self-Heating or Joule Heating

$$\Delta T_{self-heating} = \left(T_m - T_{ref}\right) = I_{rms}^2 R \theta_j$$

θ_j is the effective thermal impedance:

$$\theta_j = \frac{t_{ins}}{K_{ins} L W_{eff}}$$

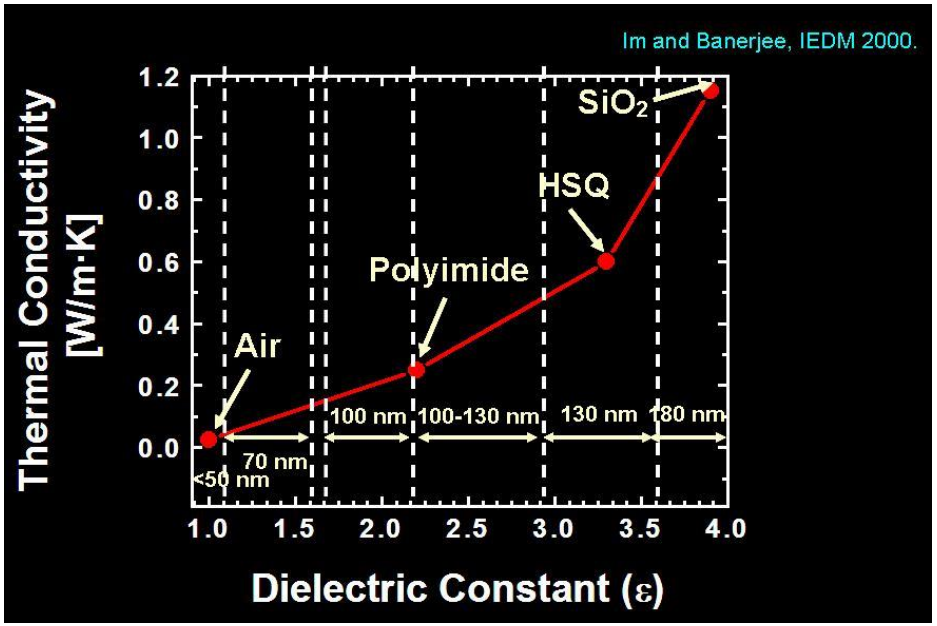


W_{eff} is the effective metal width to account for quasi-2D heat conduction.

Interconnect Scaling Trends

Implications of Technology Scaling on Thermal Effects

- Low-k materials have lower thermal conductivity than silicon dioxide.
- Scaling trends that cause increasing thermal effects are:
  - increasing current density
  - increasing interconnect levels
  - low-k dielectrics
  - increasing thermal coupling

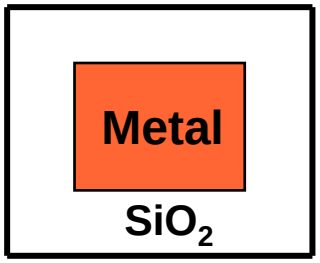


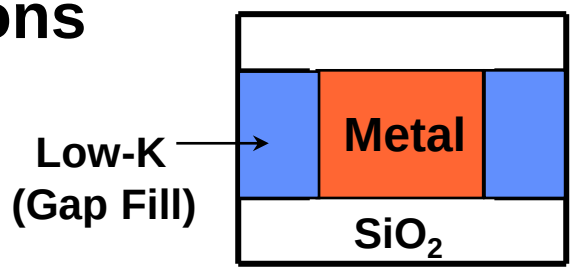
Interconnect Scaling Trends

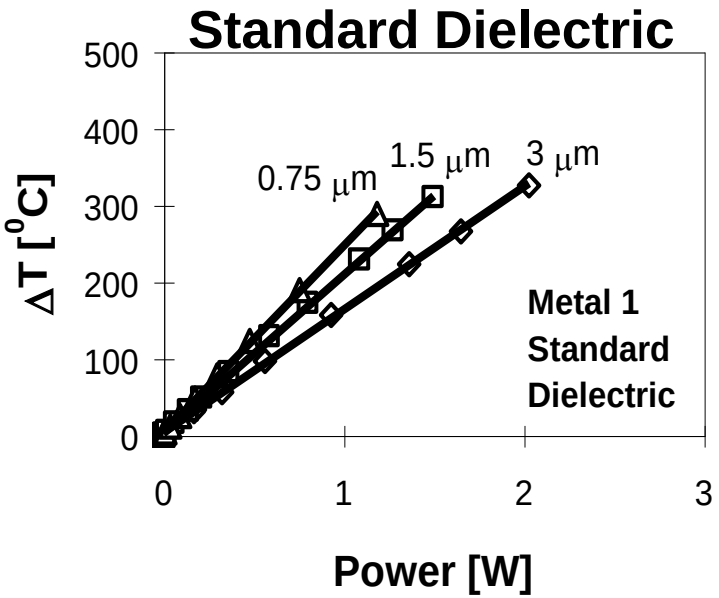
Impact of Line Width Scaling Using Low-k

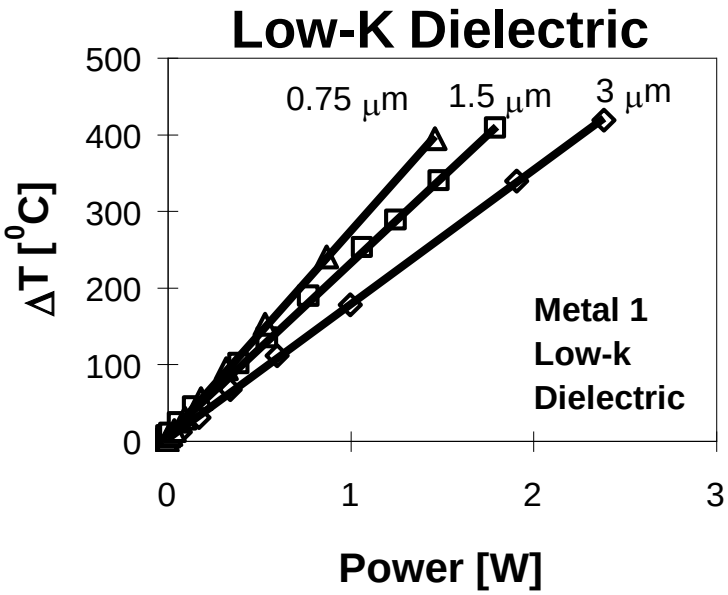
Banerjee et al., IEDM, 1996

DC Conditions









- As W decreases SH increases.
- Low-k increases SH by 10-15%

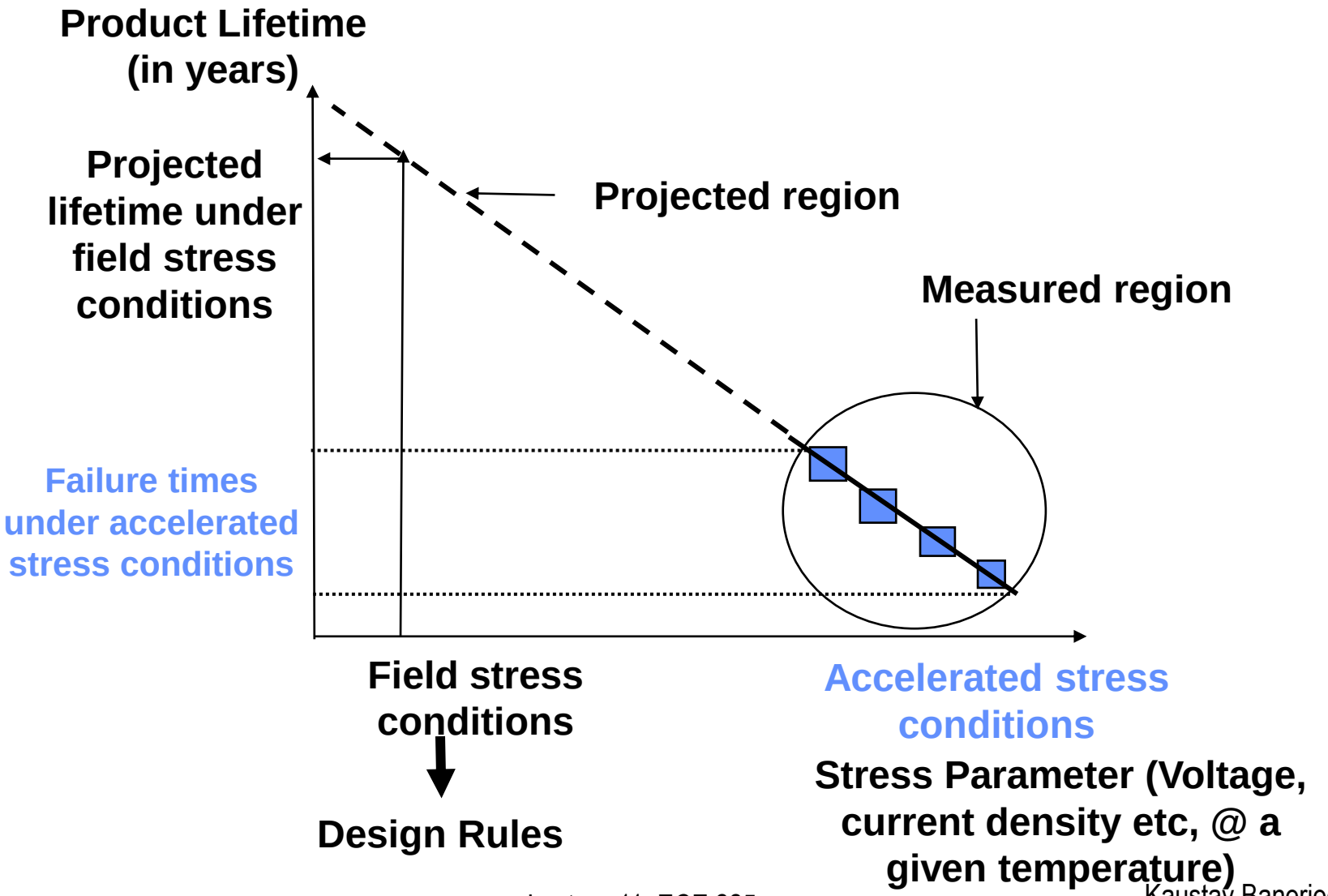
Electromigration Reliability

- Transport of mass in metal interconnects under an applied current density
- EM lifetime reliability modeled using **Black's equation** given by,

$$TTF = A\, j_{avg}^{-2}\, exp\left(\frac{Q}{k_B T_m}\right)$$

- EM stress data gives a technology limit of the current density (j_{avg}) - - *for a required failure rate and a desired lifetime at a reference temperature* T_{ref} *(~ 100* ^{0}*C)*
- The j_{avg} limit does not comprehend **self-heating**

VLSI Reliability: Lifetime Estimation and Design Rule Generation



Self-Consistent Design

Typical EM Analysis

- Accelerated EM stress data yields *A, Q*, and *n* in Black's equation, and a value of log-normal σ_{LN}.
- EM stress data + Black's equation gives a technology limit to the maximum allowed current density (j_{avg}) for the required failure rate and a desired lifetime at a reference temperature T_{ref} (~ 100 ^{0}C).
- Typical goal: achieve a 10 year lifetime.
- The j_{avg} limit does not comprehend self heating, and is therefore not self-consistent.

Self-Consistent Design

Current Density Definitions

- **Peak, Average, and RMS current densities:**

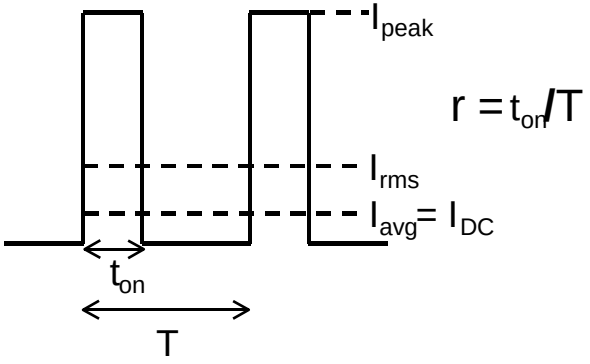
$$j_{peak} = \frac{I_{peak}}{A} \qquad j_{avg} = \frac{1}{T}\int_0^T j(t)dt \qquad j_{rms} = \sqrt{\frac{1}{T}\int_0^T j^2(t)dt}$$

A **is the cross sectional area of interconnect,** ***T*** **is the time period of the current waveform.**

- **For an unipolar waveform:**

$$j_{avg} = r\, j_{peak} \qquad j_{rms} = \sqrt{r}\, j_{peak}$$

r **is the duty factor**



- **EM is determined by j_{avg}, and self-heating by j_{rms}.**

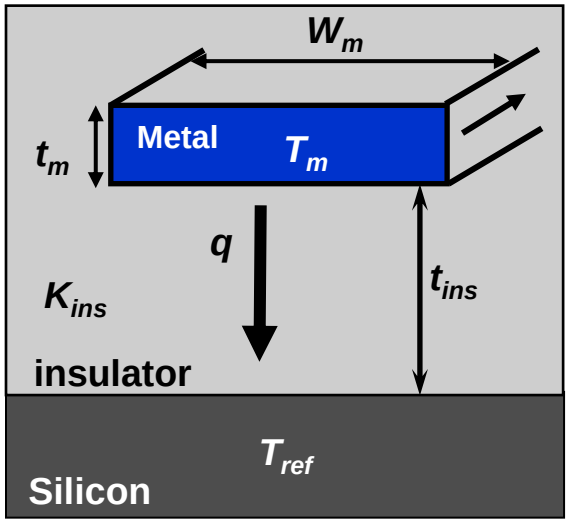
Self-Consistent Design

Impact of Self-Heating on EM

- EM lifetime given by: $TTF = A\, j^{-n} \exp\left(\frac{Q}{k_B T_m}\right)$
- Due to self-heating: $T_m = T_{ref} + \Delta T_{self\text{-}heating}$

$$\Delta T_{self-heating} = (T_m - T_{ref}) = I_{rms}^2 R R_\theta$$

$$j_{rms}^2 = \frac{(T_m - T_{ref}) K_{ins} W_{eff}}{t_{ins} t_m W_m \rho_m (T_m)} \quad (1)$$



R_θ is the effective thermal impedance given by,

$$R_\theta = \frac{t_{ins}}{K_{ins} L W_{eff}}$$

W_{eff} is the effective metal width to account for quasi-2D heat conduction.

Self-Consistent Design

What is Self-Consistent Design?

- Typically, design rules specify j_{avg} from EM and j_{rms} from self-heating separately.
- Self-consistent approach: comprehends EM and self-heating simultaneously.
- In order to achieve an EM reliability lifetime goal mentioned earlier, the lifetime at any j_{avg} and metal temperature T_m, should be equal to or greater than the lifetime value (e.g., 10 year) under the design rule current density (j_0).

$$\frac{exp\left(\frac{Q}{k_B T_m}\right)}{j_{avg}^2} \geq \frac{exp\left(\frac{Q}{k_B T_{ref}}\right)}{j_0^2} \qquad (2)$$

Self-Consistent Design

What is Self-Consistent Design?

- Using the relationship between j_{avg}, j_{rms}, j_{peak} and r for an unipolar waveform described earlier, it can be shown that,

$$\frac{j_{avg}^2}{j_{rms}^2} = r \qquad (3)$$

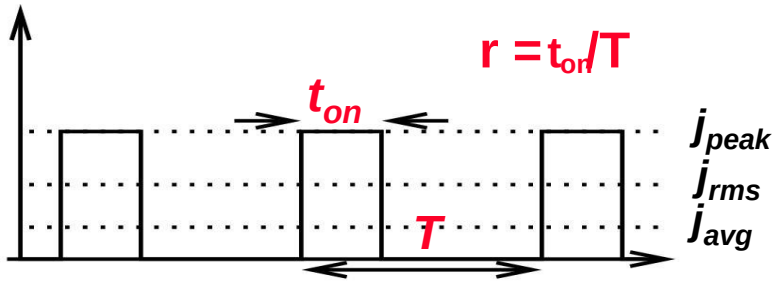
- Incorporating the j^2_{rms} and j^2_{avg} values from (1) and (2) in (3) yields the self-consistent equation,

$$r = j_0^2 \frac{\exp\left(\frac{Q}{k_B T_m}\right)}{\exp\left(\frac{Q}{k_B T_{ref}}\right)} \frac{t_{ins}\, t_m\, W_m\, \rho_m(T_m)}{(T_m - T_{ref}) K_{ins}\, W_{eff}}$$

This is a single equation in the single unknown temperature T_m. Once the self-consistent T_m is obtained, the corresponding j_{rms} and j_{peak} and j_{avg} values can be calculated.

Coupled Analysis

Unipolar Case:



- **Coupled equation:**

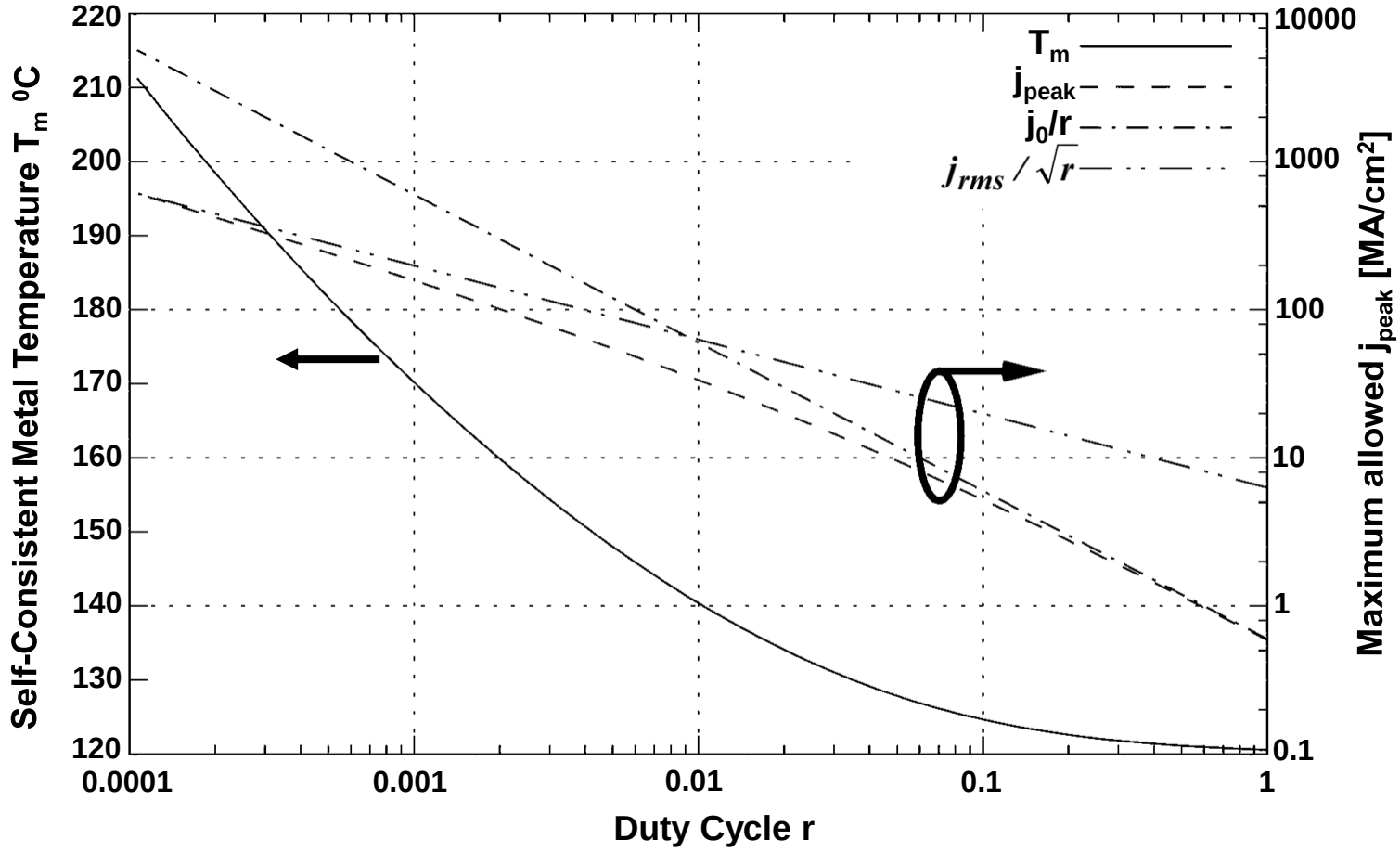
$$\frac{j_{avg}^2}{j_{rms}^2} = r$$

$$r = j_0^2 \frac{\exp\left(\frac{Q}{k_B T_m}\right)}{\exp\left(\frac{Q}{k_B T_{ref}}\right)} \frac{t_{ins} t_m W_m \rho_m (T_m)}{(T_m - T_{ref}) K_{ins} W_{eff}}$$

Once T_m is calculated, j_{rms} and j_{peak} can be obtained

Coupled Analysis

Unipolar Case: Metal 6, 180 nm node, j_0 = 0.6 MA/cm^2



Forms of Power Dissipation

- **Dynamic Power**

➢ **Due to charging/discharging of capacitors….**

- **Leakage Power**

➢ **Subthreshold leakage**

➢ **Gate leakage**

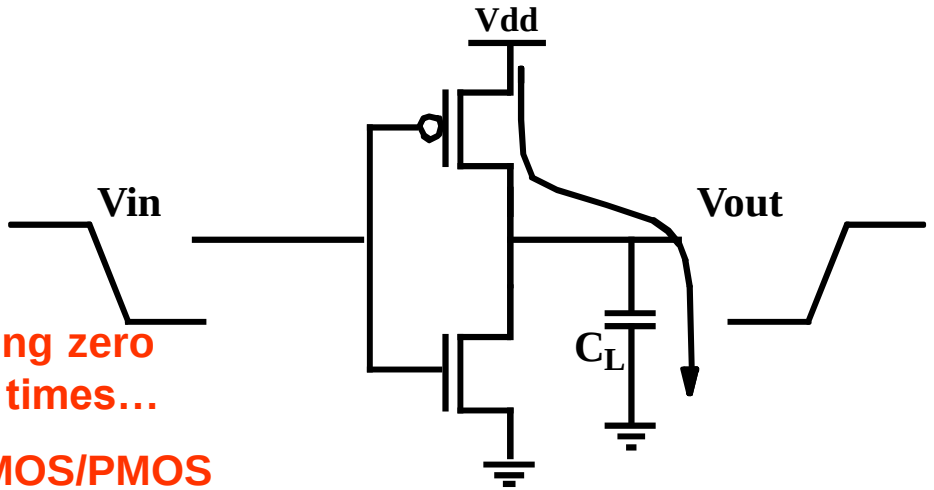
➢ **Junction leakage**

➢ **other mechanisms…DIBL, GIDL**

- **Short-Circuit Power**

➢ **When NMOS and PMOS are both turned ON….**

Dynamic Power Dissipation



Assuming zero rise/fall times…

So that NMOS/PMOS are never ON simultaneously…

Energy/transition = $C_L * V_{dd}^2$

Low-to-High transition:

- C_L is charged through PMOS and discharged through NMOS
- During charging, voltage of C_L rises from 0 to V_{dd}… energy is drawn from the power supply
- Part of this energy is stored in C_L while rest gets dissipated in the PMOS

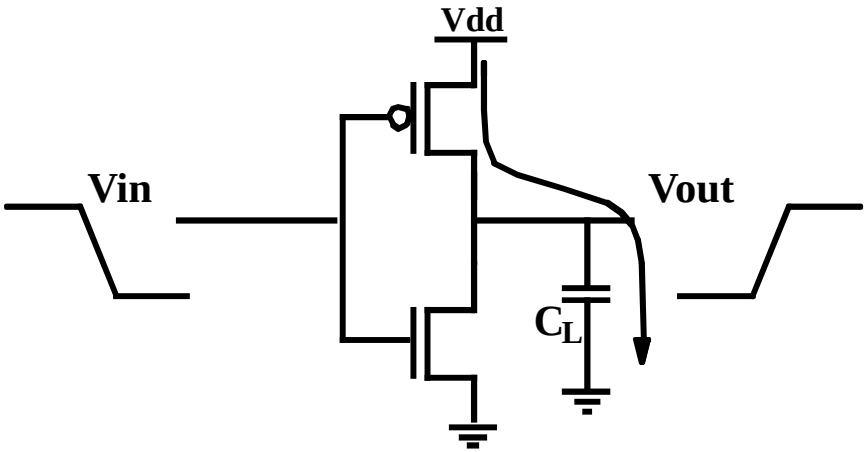
Note: Here C_L is an external capacitor….

- Energy/transition is not a function of transistor sizes!
- Need to reduce C_L, V_{dd}, and f to reduce power.

Energy taken from supply during transition:

$$E_{V_{DD}} = \int_0^\infty i_{V_{DD}}(t) V_{DD}\, dt = V_{DD} \int_0^\infty C_L \frac{dv_{out}}{dt}\, dt = C_L V_{DD} \int_0^{V_{DD}} dv_{out} = C_L V_{DD}^2$$

Dynamic Power Dissipation



High-to-Low transition:

- **During the discharge phase, charge is removed from C_L and its energy is dissipated in the NMOS**
- **Each switching cycle consists of a L-H and H-L transition and takes a fixed amount of energy = $C_L V_{dd}^2$**

Energy taken from supply during transition:

$$E_{V_{DD}} = \int_0^\infty i_{V_{DD}}(t) V_{DD}\, dt = V_{DD} \int_0^\infty C_L \frac{dv_{out}}{dt}\, dt = C_L V_{DD} \int_0^{V_{DD}} dv_{out} = C_L V_{DD}^2$$

Energy stored in the capacitor:

$$E_C = \int_0^\infty i_{V_{DD}}(t) v_{out}\, dt = \int_0^\infty C_L \frac{dv_{out}}{dt} v_{out}\, dt = C_L \int_0^{V_{DD}} v_{out}\, dv_{out} = \frac{C_L V_{DD}^2}{2}$$

Where is the other half of the energy? ….Dissipated in the PMOS

Dynamic Power Dissipation

$$\text{Power} = \text{Energy/transition} * f = C_L * V_{dd}^2 * f$$

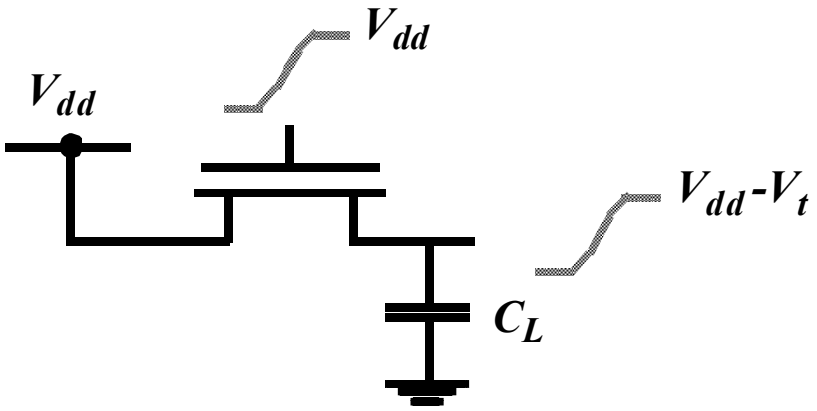
- "f" accounts for how often the gate is switched ON and OFF…
- f increases with scaling…since gate delay decreases

Consider a 0.25 um CMOS technology with a clock-rate of 500 MHz and average load capacitance of 15 fF/gate assuming a fan-out of 4: the power consumption per gate = 50 μW (for V_{dd} = 2.5 V)

One-million such gates would result in a power consumption of 50W! (fortunately, not all gates are switching at the clock frequency…)

Power dissipation in a complex gate is more complex to estimate…since the switching frequency depends on the nature and statistics of the input signals. This is accommodated by replacing "f" with "p.F", where p is the probability that a clock event results in a 0 to 1 transition and F is the maximum possible event-rate (or clock rate) of inputs

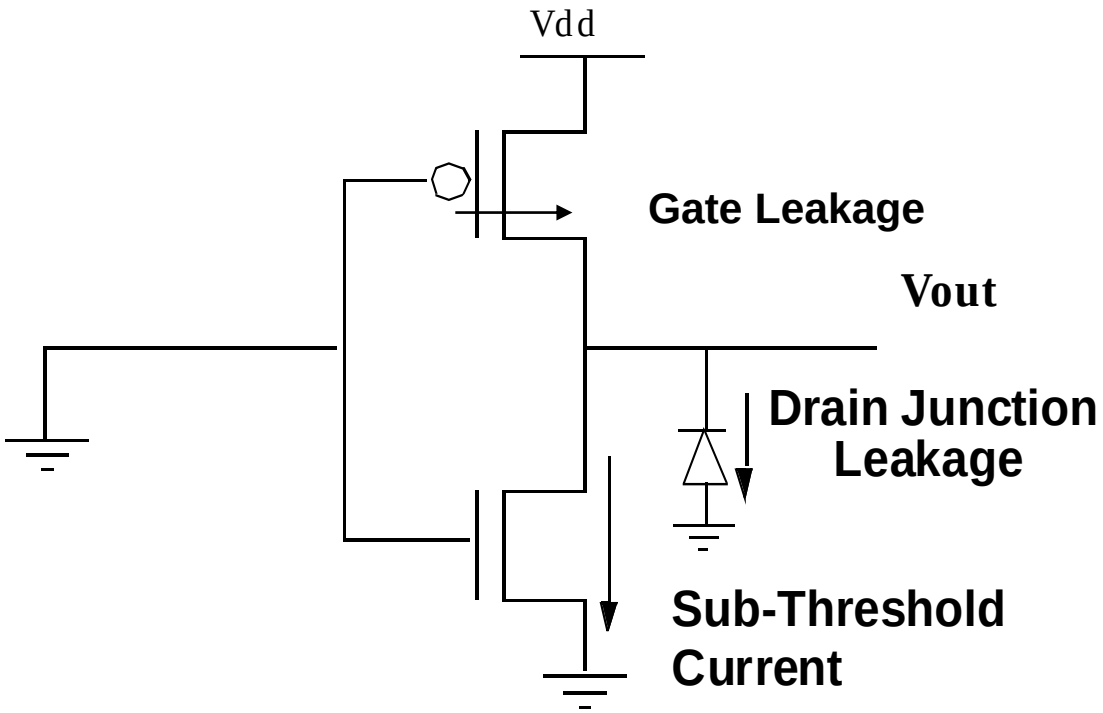
Modification for Circuits with Reduced Swing



$$E_{0 \rightarrow 1} = C_L \bullet V_{dd} \bullet (V_{dd} - V_t)$$

- **Can exploit reduced swing to lower power (e.g., reduced bit-line swing in memory)**

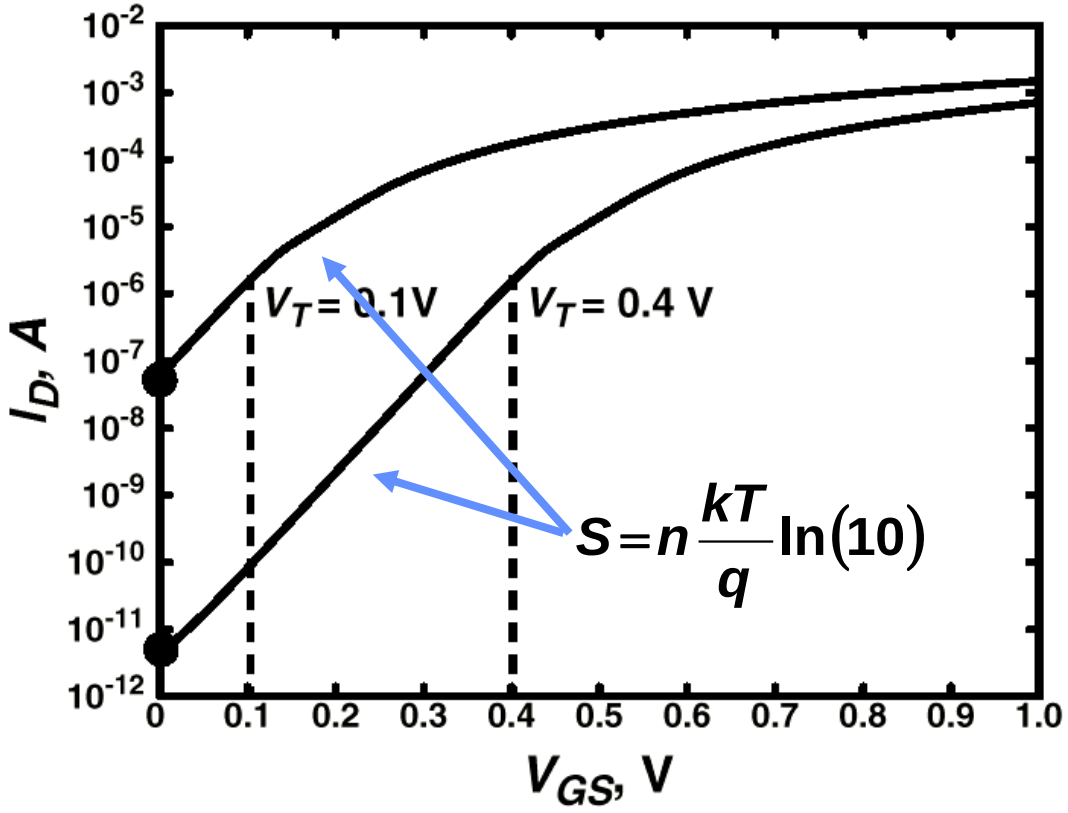
Primary Leakage Mechanisms

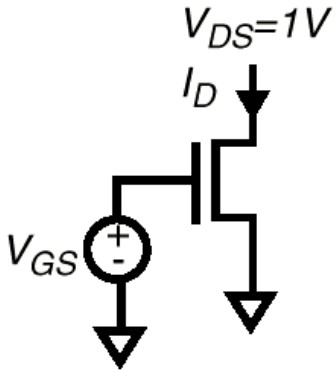


Sub-threshold current one of most compelling issues in low-power and energy-efficient circuit design!

Subthreshold Leakage Component

Subthreshold Slope = S = 60 mv/decade (for ideal transistor with n=1)



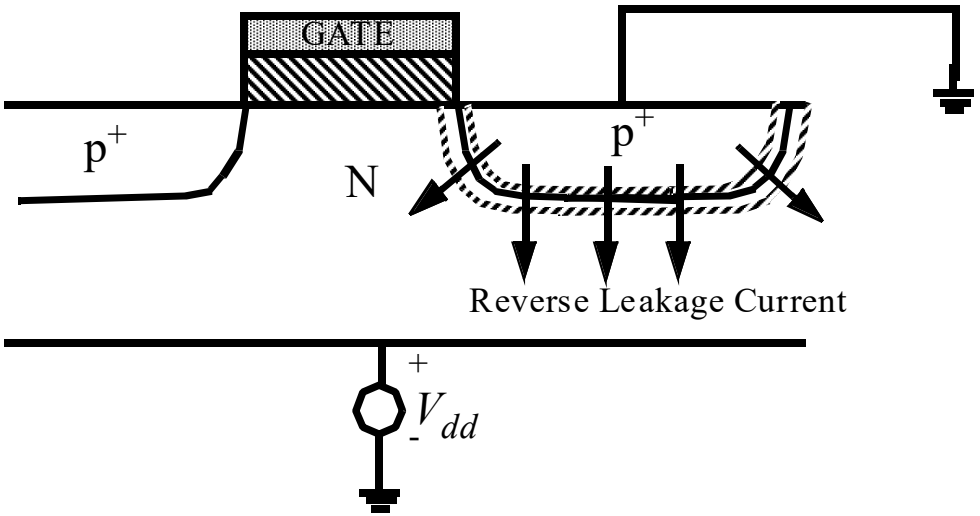


- **Leakage control is critical for low-voltage operation**

Note: DIBL will further increase subthreshold leakage!

Reverse-Biased Diode Leakage

Much smaller than subthreshold leakage….

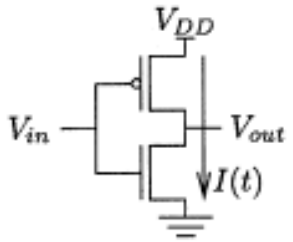


$$I_{DL} = J_S \times A$$

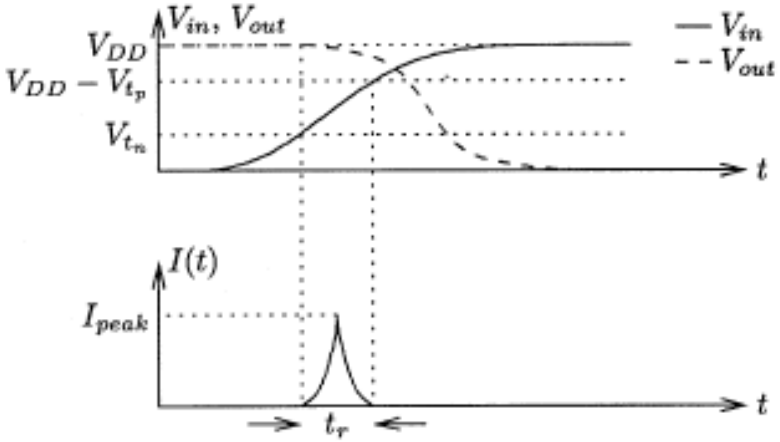
J_S = 10-100 pA/μm^2 at 25 deg C for 0.25μm CMOS

J_S doubles for every 9 deg C!

Short Circuit Currents



(a)
CMOS
inverter.



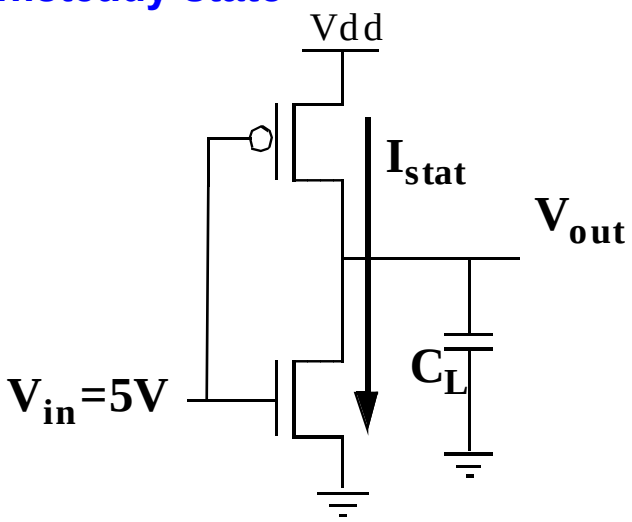
(b) Voltage and current waveforms.

Fig. 5. Voltage and current waveforms of a CMOS inverter.

K. Banerjee and A. Mehrotra, IEEE Transactions on Electron Devices, Vol. 49, No. 11, 2002.

Static Power Consumption

In the absence of switching….steady-state operation



Sources:

Due to all previously mentioned leakage currents……

$$P_{stat} = P_{(In=1)}.V_{dd} \cdot I_{stat}$$

Wasted energy …should be avoided in almost all cases…

Leakage Currents

- **Effect of leakage current**
  - "Wasted" power: power consumed even when circuit is inactive
  - Leakage power raises temperature of chip – positive feedback effect
  - Can cause functionality problem in some circuits: memory, dynamic logic, etc.
- **Reducing transistor leakage**
  - Long-channel devices
  - Small drain voltage
  - Large threshold voltage V_T

Leakage Power Reduction

- Process scaling
  - V_T reduces with each new process (historically)
  - Leakage increases ~10X!
- Leakage vs. performance tradeoff:
  - For high-speed, need small V_T and L
  - For low leakage, need high V_T and large L
- One solution: dual-V_T process
  - Low-V_T transistors: use in critical paths for high speed
  - High-V_T transistors: use to reduce power

Principles for Power Reduction

- Prime choice: Reduce voltage!
  - Dynamic power has quadratic dependence on Vdd
  - Hard to reduce Vdd without reducing Vt
  - Design at very low voltages (< 0.4 V) still open question
  - Improve Subthreshold Swing to ~60 mV/decade
  - Steep-slope transistors: such as TFET
- Reduce physical capacitance
  - Device Sizing
- Reduce switching activity
  - Exploit new architectures: Non Von-Neumann

What is the Lowest Possible Power Supply?

V_{dd} must be greater than a few times kT/q…..

Why is that?

For static CMOS, if minimum gain (near switching threshold) =1, then $V_{dd,min}$ = (2-4) kT/q

Refer to the derivation discussed in the class….

Consider effect of variability…