
ECE 225

High-Speed Digital IC Design

Lecture 7

Interconnect Technology and Scaling Issues-3

(Thermal, Reliability and Power)

Prof. Kaustav Banerjee

Electrical and Computer Engineering

E-mail: kaustav@ece.ucsb.edu

Interconnect Thermal Effects and Reliability

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

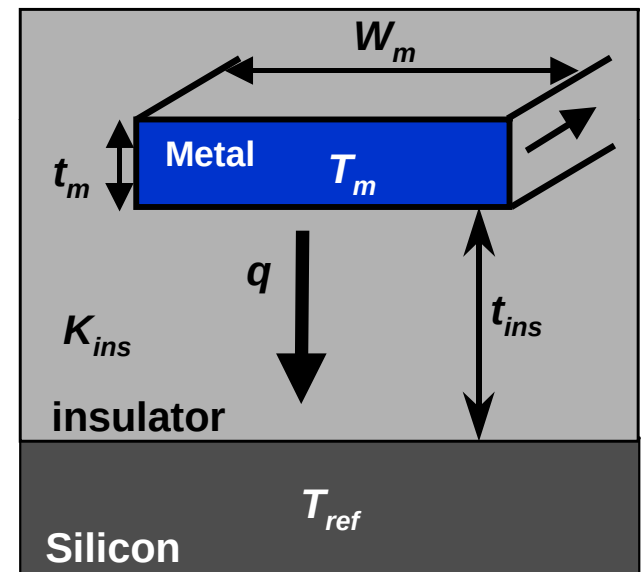
Self-Heating or Joule Heating

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

θ_j is the effective thermal impedance:

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

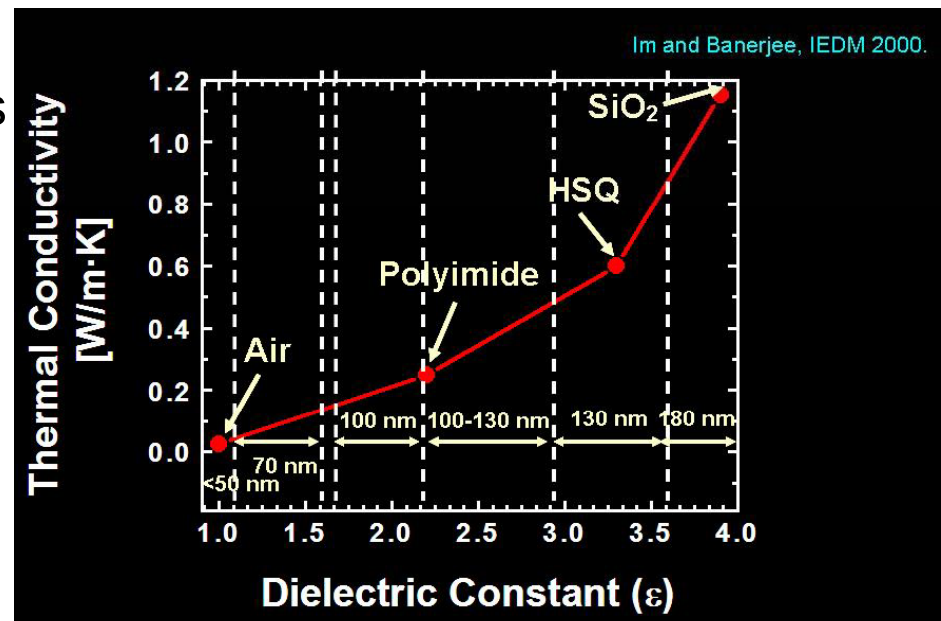
W_{eff} is the effective metal width to account for quasi-2D heat conduction.



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

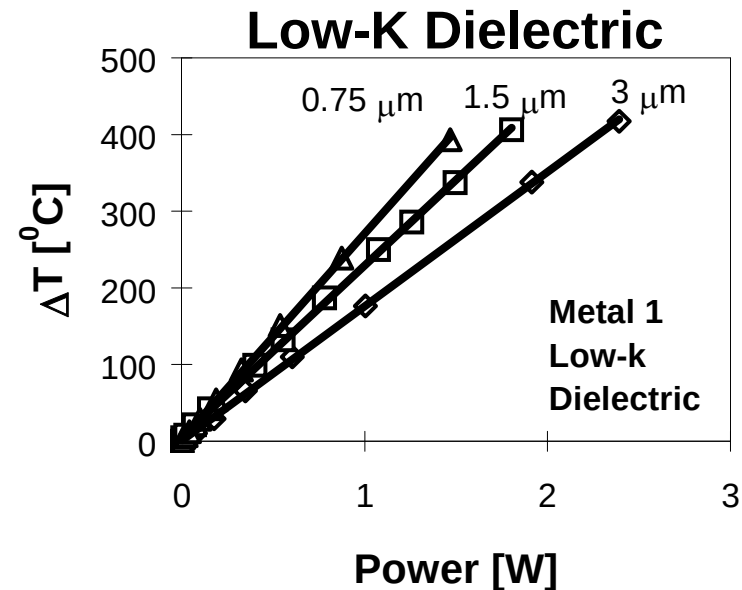
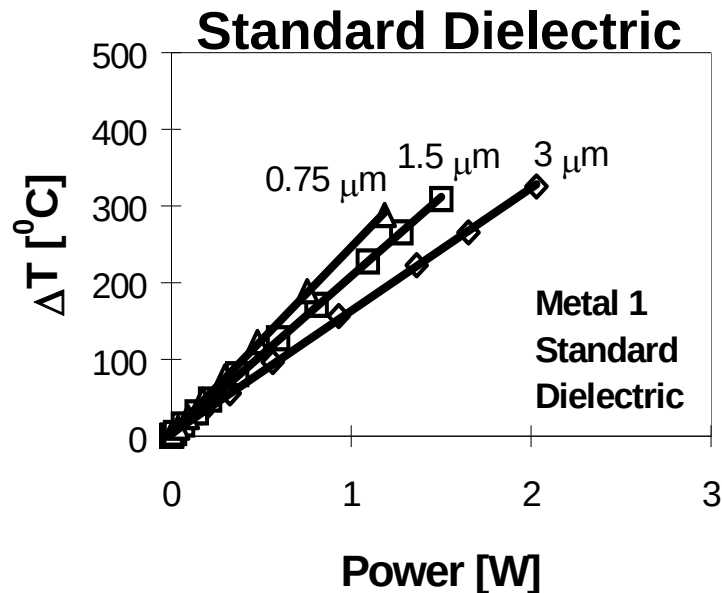
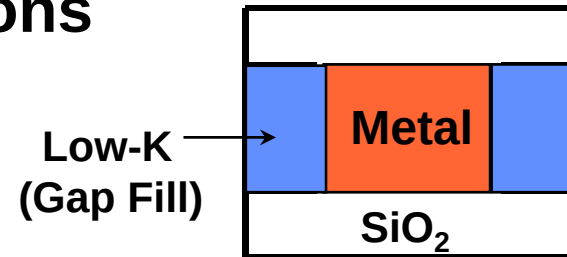
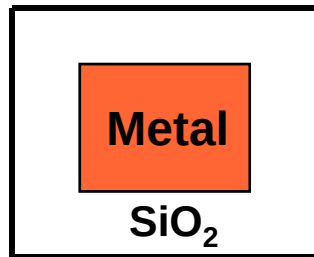


Interconnect Scaling Trends

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%

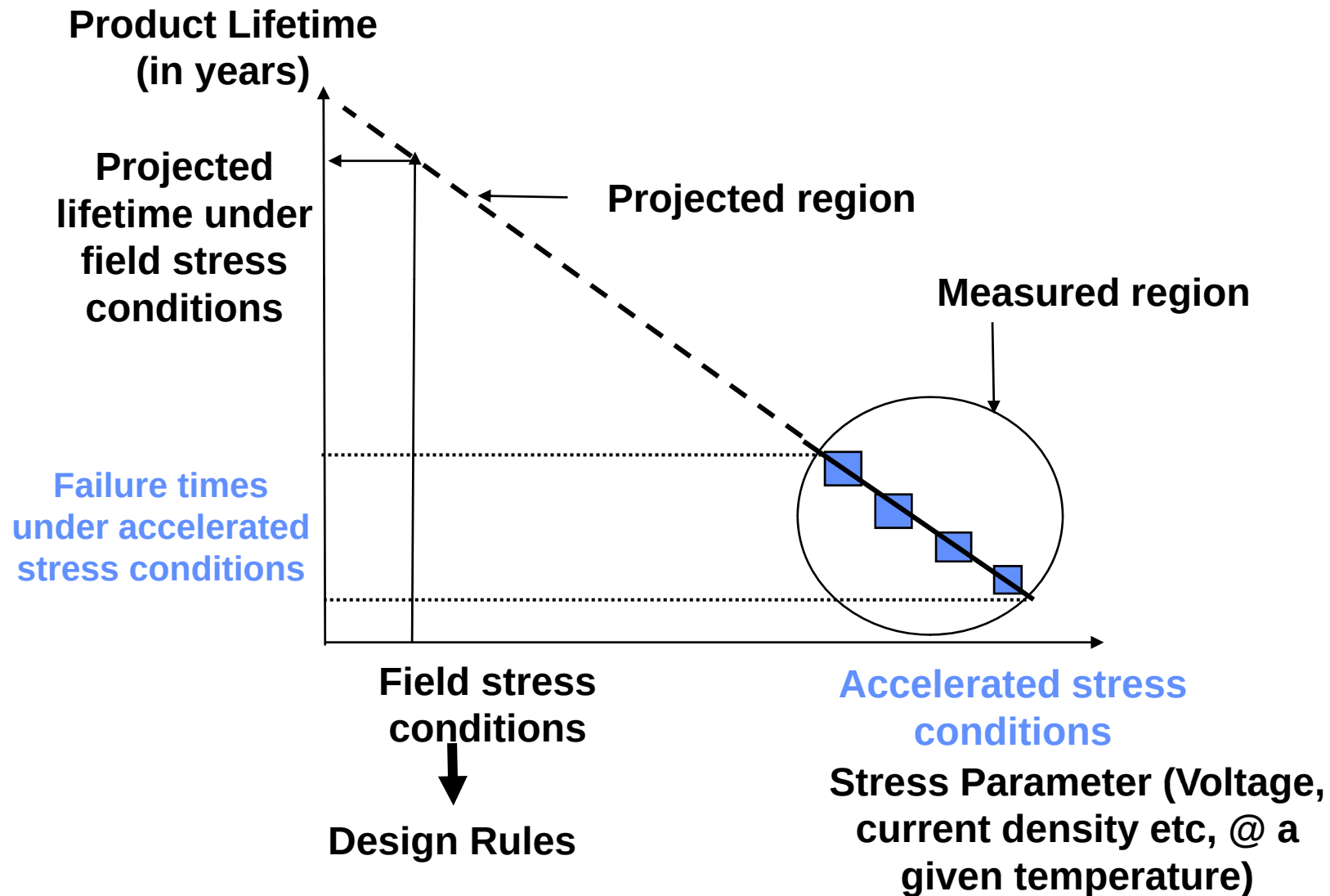
Electromigration Reliability

- 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} ($\sim 100^\circ\text{C}$)*
- 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} ($\sim 100^\circ\text{C}$).
- Typical goal: achieve a 10 year lifetime.
- The j_{avg} limit does not comprehend self heating, and is therefore not self-consistent (Hunter et al., TED 1996).

Self-Consistent Design

Current Density Definitions

■ Peak, Average, and RMS current densities:

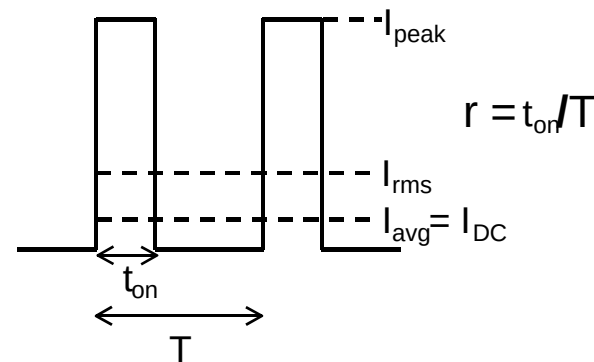
$$j_{peak} = \frac{I_{peak}}{A} \quad j_{avg} = \frac{1}{T} \int_0^T j(t) dt \quad j_{rms} = \sqrt{\frac{1}{T} \int_0^T j^2(t) dt}$$

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

■ For an unipolar waveform:

$$j_{avg} = r j_{peak} \quad j_{rms} = \sqrt{r} j_{peak}$$

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: $T_m = T_{ref} + \Delta T_{self-heating}$

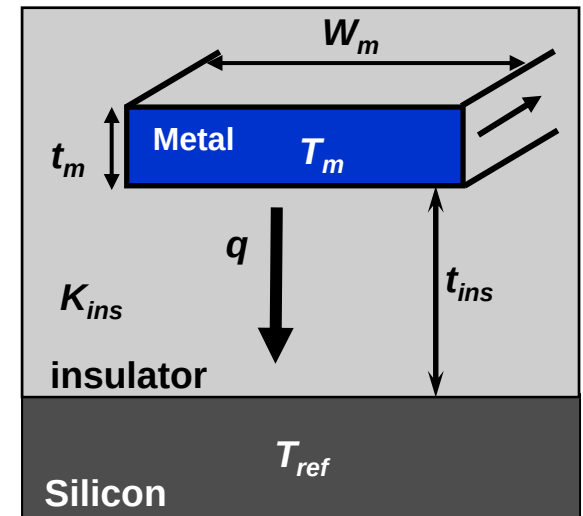
$$\Delta T_{self-heating} = (T_m - T_{ref}) = I_{rms}^2 R R_\theta$$

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

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).

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

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,

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

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

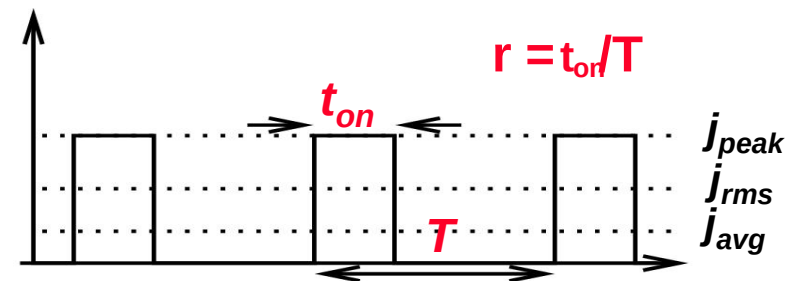
$$r = j_0^2 \frac{\exp\left(\frac{Q}{k_B T_m}\right)}{\exp\left(\frac{Q}{k_B T_{ref}}\right)} \frac{t_{ins} t_m W_m \rho_m (T_m)}{(T_m - T_{ref}) K_{ins} W_{eff}}$$

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:



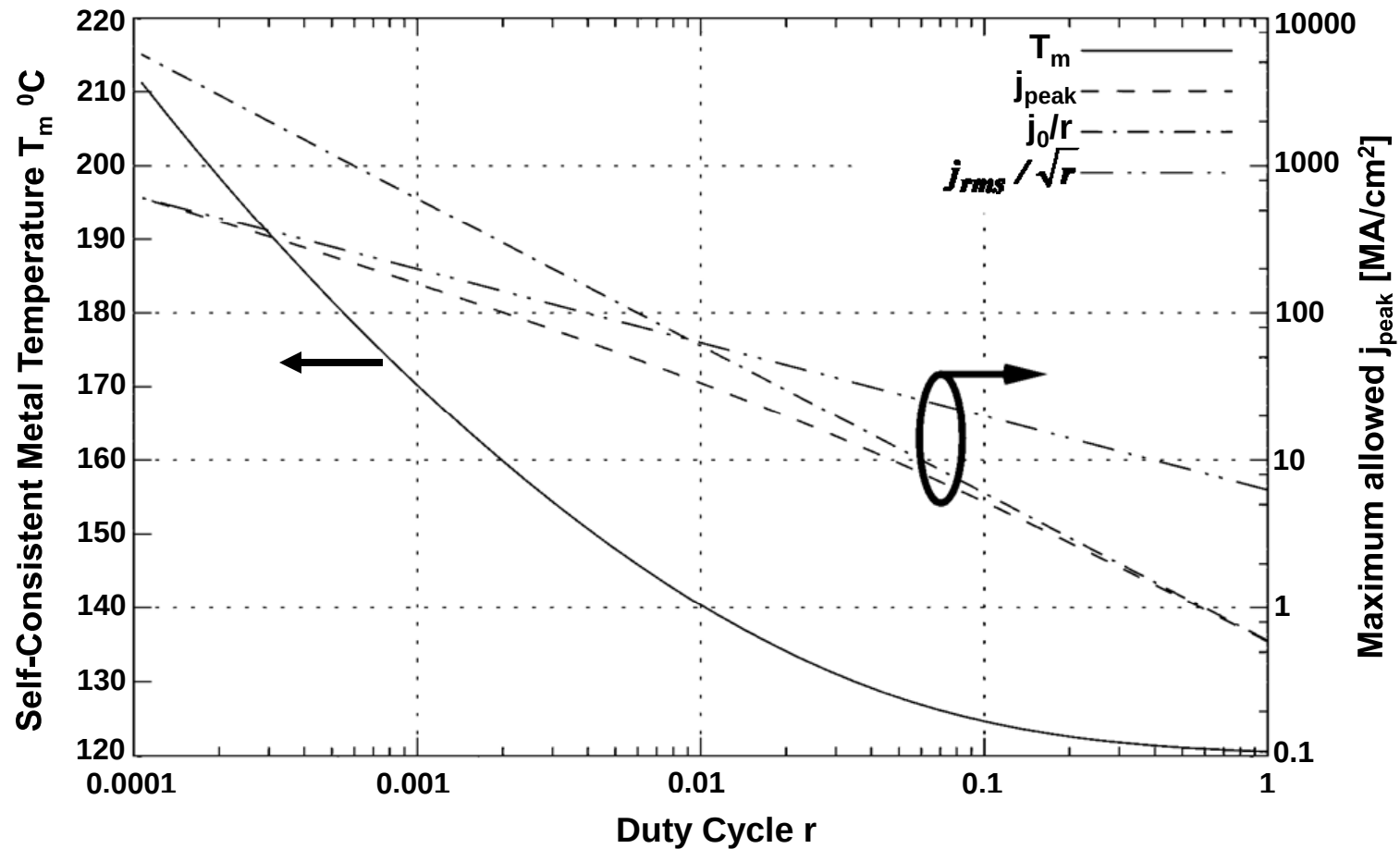
$$\frac{j_{avg}^2}{j_{rms}^2} = r$$

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

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

Coupled Analysis

Unipolar Case: Metal 6, 180 nm node, $j_0 = 0.6 \text{ MA/cm}^2$



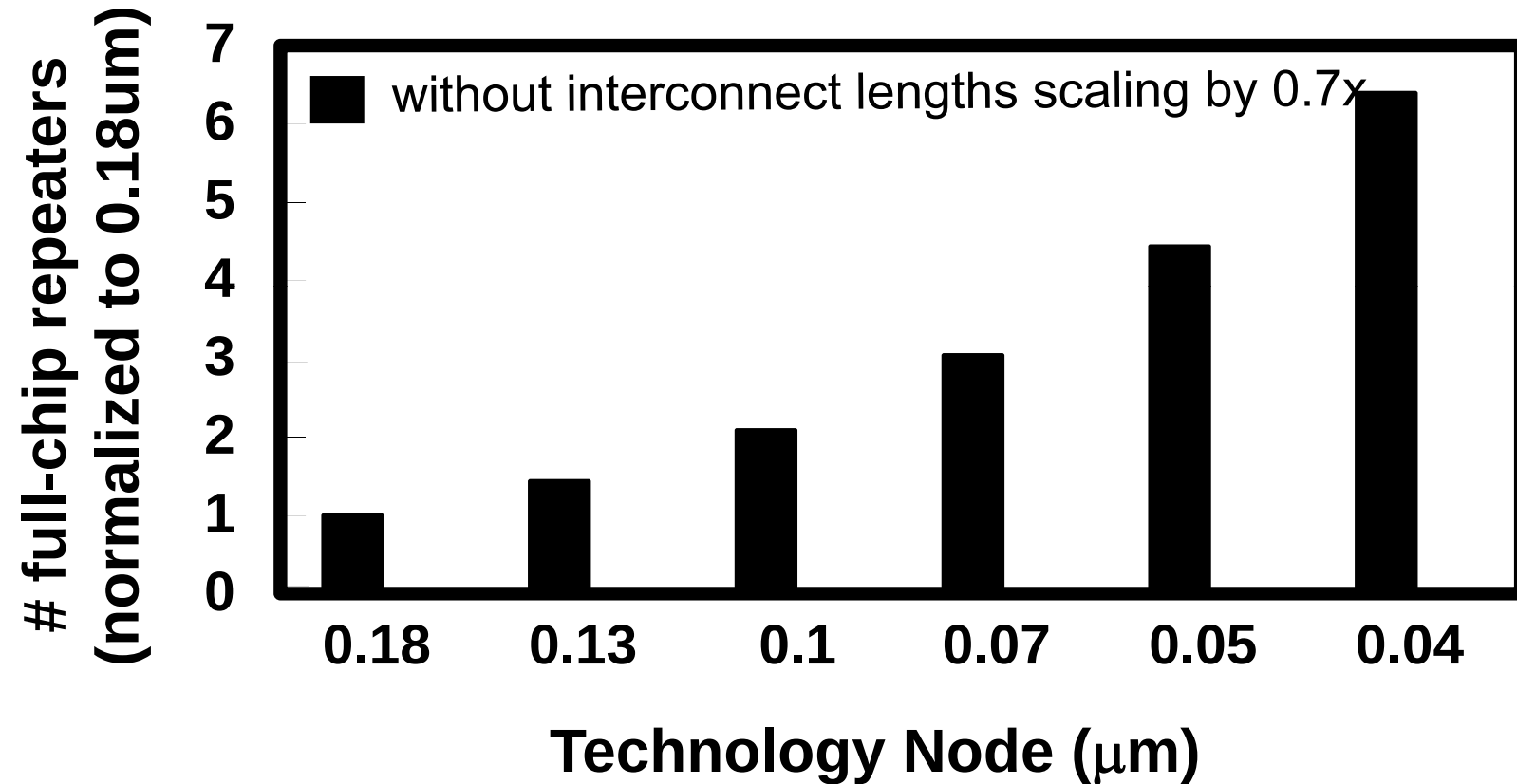
Interconnect Delay and Power Dissipation

K. Banerjee and A. Mehrotra, IEEE Transactions on Electron Devices, Vol. 49, No. 11, pp. 2001-2007, November 2002.

Performance Optimization: Global Wires

- Increasing chip complexity and scaling
 - number and (electrical) length of global lines & latency increase
- Repeater insertion used for lowering the signal delays
- Delay of optimally repeatered line is linear in its length
- High-performance designs: **large number of repeaters ($> 10^6$ for sub-100 nm designs)**

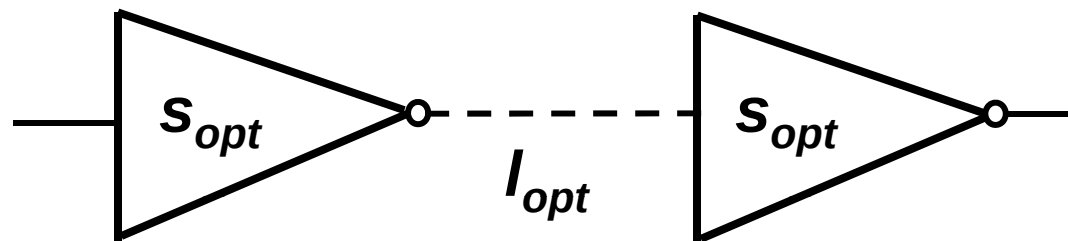
Increasing Number of Repeaters



Source: Intel Corporation

Optimal Repeater Insertion

- Long interconnects can be optimally buffered



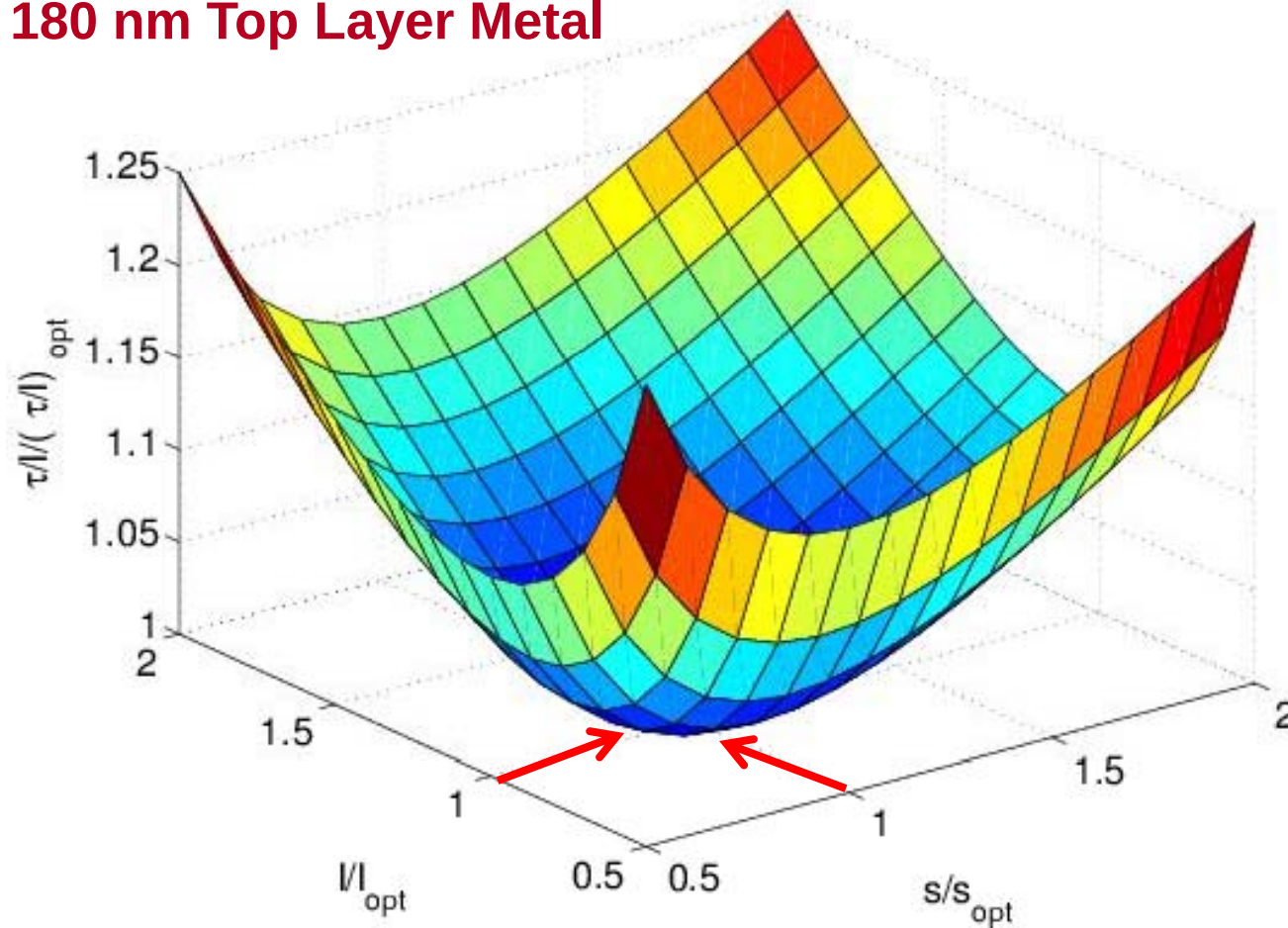
- Large size: could be ~ 450 times the minimum sized inverters available in a relevant technology (K. Banerjee et al., DAC 1999)
- Can contribute to significant on-chip **power dissipation!**

Delay Vs Power Tradeoff for Global Wires

- All global wires are not on the critical path
- Optimal buffering is not always necessary
- Some delay penalty can be tolerated on non-critical wires
- **Potential for large power savings** by using smaller repeaters and larger inter-repeater wire length

Interconnect Delay Optimization

180 nm Top Layer Metal



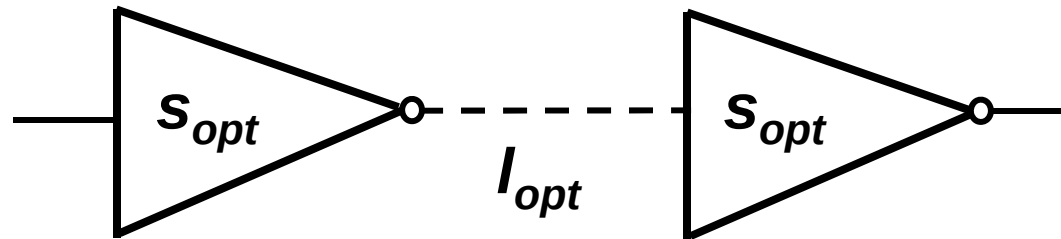
For $s = s_{opt}/2$ and $l = 2 l_{opt}$; Delay Penalty is **only 25%**

Power-Optimal Repeater Insertion

Banerjee and Mehrotra, TED 2002.

- A methodology to estimate repeater size and inter-repeater wire length which minimizes the total interconnect power dissipation for a given delay penalty
- Estimate power-optimal buffering schemes for various ITRS technology nodes
- Analyze relative importance of various components of power dissipation as a function of technology scaling

Methodology



delay per unit length:

$$\frac{\tau}{\ell} = \frac{1}{\ell} r_s (c_0 + c_p) + \frac{r_s}{s} c + r_s c_0 + \frac{1}{2} r c \ell$$

For $\ell = \ell_{opt}$, $s = S_{opt}$:

$$(\tau / \ell)_{opt} = 2 \sqrt{r_s c_0 r c} \left\{ 1 + \sqrt{0.5 \left(1 + \frac{c_p}{c_0} \right)} \right\}$$

Methodology

$$P_{repeater} = P_{switching} + P_{leakage} + P_{short\ circuit}$$

$$P_{switching} = \alpha (s(c_p + c_o) + I_c) V_{DD}^2 f_{clk}$$

$$P_{leakage} = V_{DD} I_{leakage} = V_{DD} I_{off} W_{n_{min}} s$$

$$P_{short\ circuit} = \alpha t_r V_{DD} I_{peak} f_{clk} = \alpha t_r V_{DD} W_{n_{min}} s I_{short\ circuit} f_{clk}$$

- α is the activity factor (=0.15)
- t_r = rise time of input voltage to change from V_{tn} to $V_{DD} - V_{tp}$

Methodology

If the delay penalty is f , then: $\frac{\tau}{I} = (1 + f)(\tau / I)_{opt}$

$$\frac{P_{repeater}}{I} = k_1 \left(s / I (c_p + c_0) + c \right) + k_2 \frac{s}{I} + k_3 s$$

 switching  leakage  short-circuit

$$k_1 = \alpha V_{DD}^2 f_{clk}$$

$$k_2 = \frac{3}{2} V_{DD} I_{off_n} W_{n_{min}}$$

$$k_3 = \alpha V_{DD} W_{n_{min}} I_{short\ circuit} f_{clk} \log_e 3 (1 + f)(\tau / I)_{opt}$$

Methodology

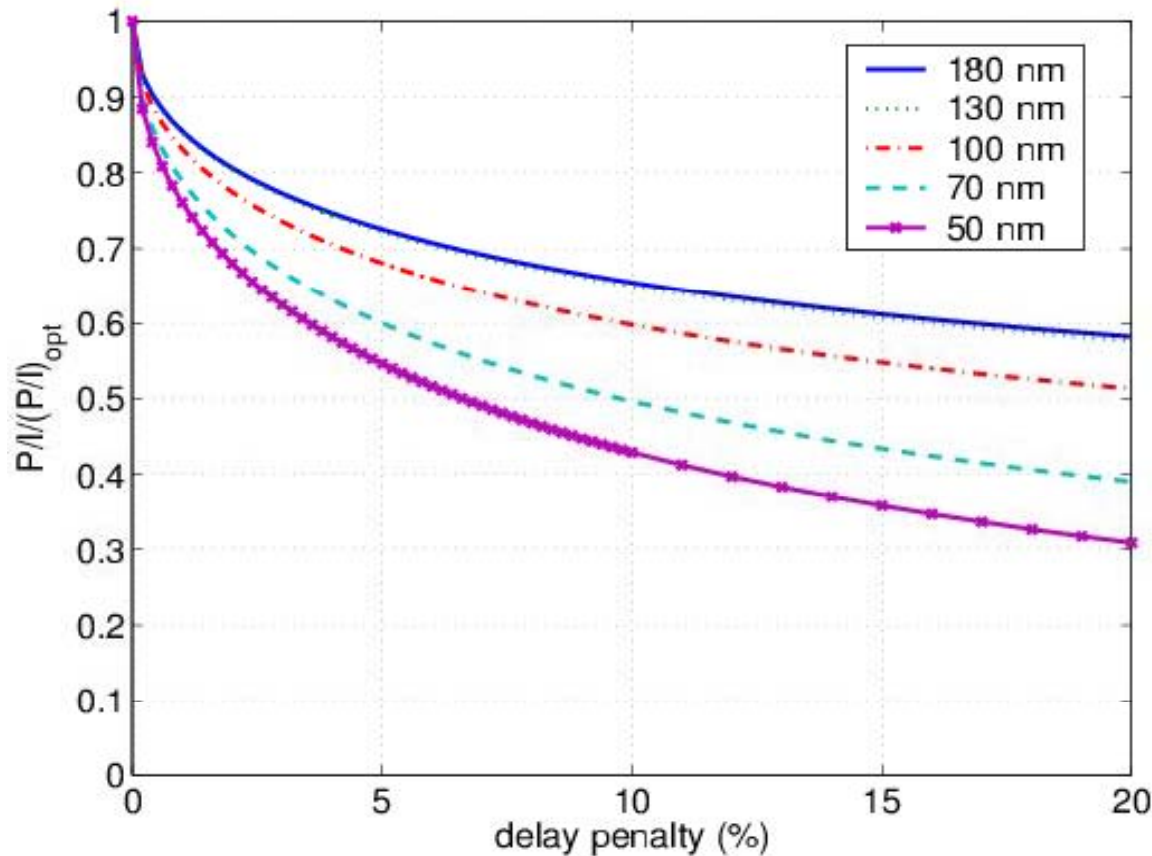
Minimize P_{repeater} w.r. t. s :

$$\frac{P_{\text{repeater}}}{I} = k_1 \left(s / I (c_p + c_0) + c \right) + k_2 \frac{s}{I} + k_3 s$$


switching leakage short-circuit

- Involves three non-linear equations with three unknowns: I , s , and dI/ds
- Used Newton-Raphson to solve numerically

Results: Power Vs Delay Penalty



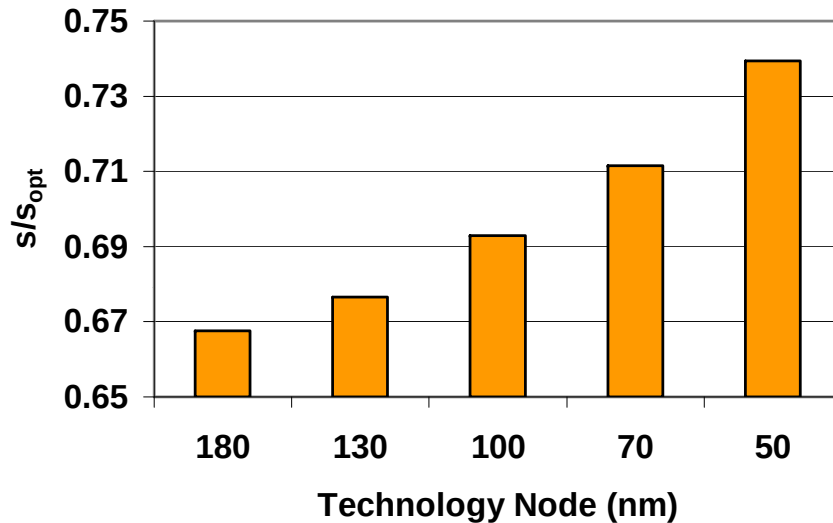
■ Normalized power per unit length reduces as delay penalty increases

■ Incremental reduction in P/I is high for small values of the delay penalty

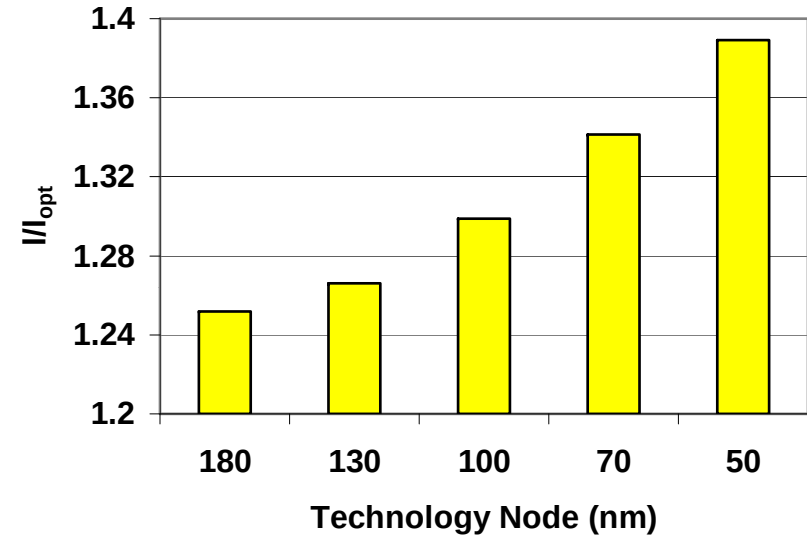
■ 180 nm and 130 nm results are almost identical

■ However, normalized P/I decreases with technology scaling:
leakage power

Optimization Results: 5% delay penalty



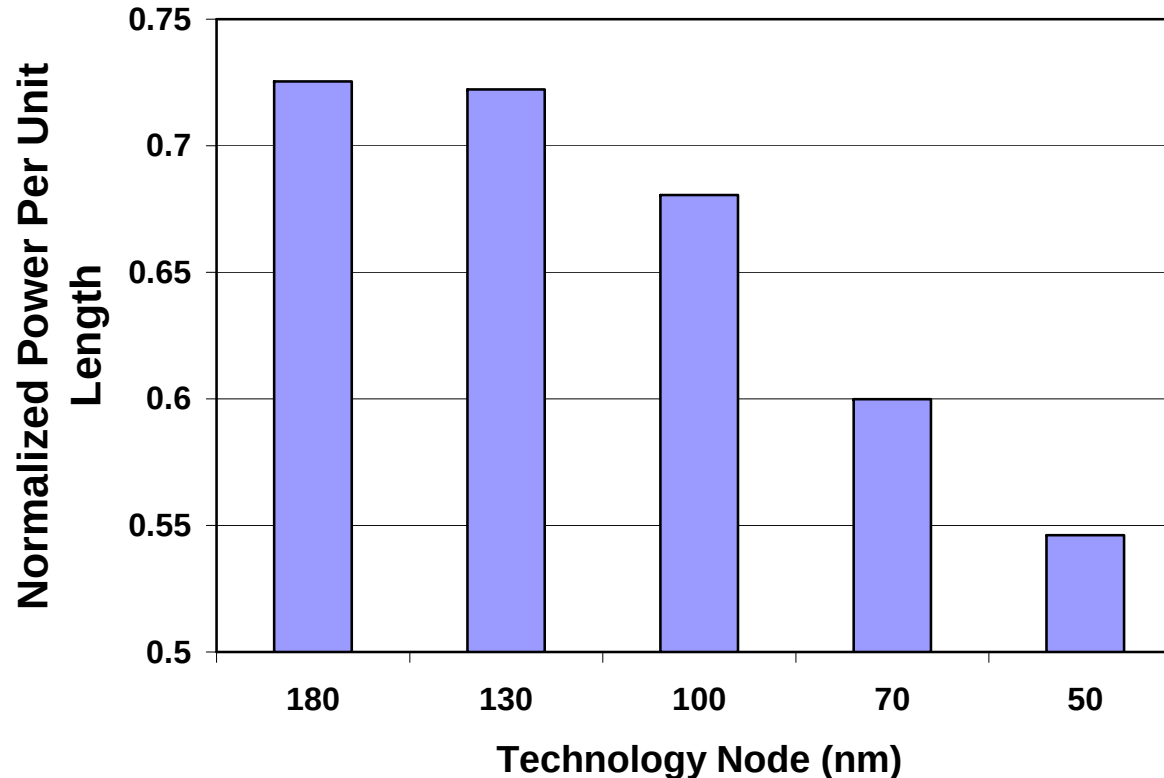
$$S < S_{opt}$$



$$I > I_{opt}$$

For optimal power dissipation: **repeater size** needs to be reduced and **inter-buffer wire length** needs to be increased

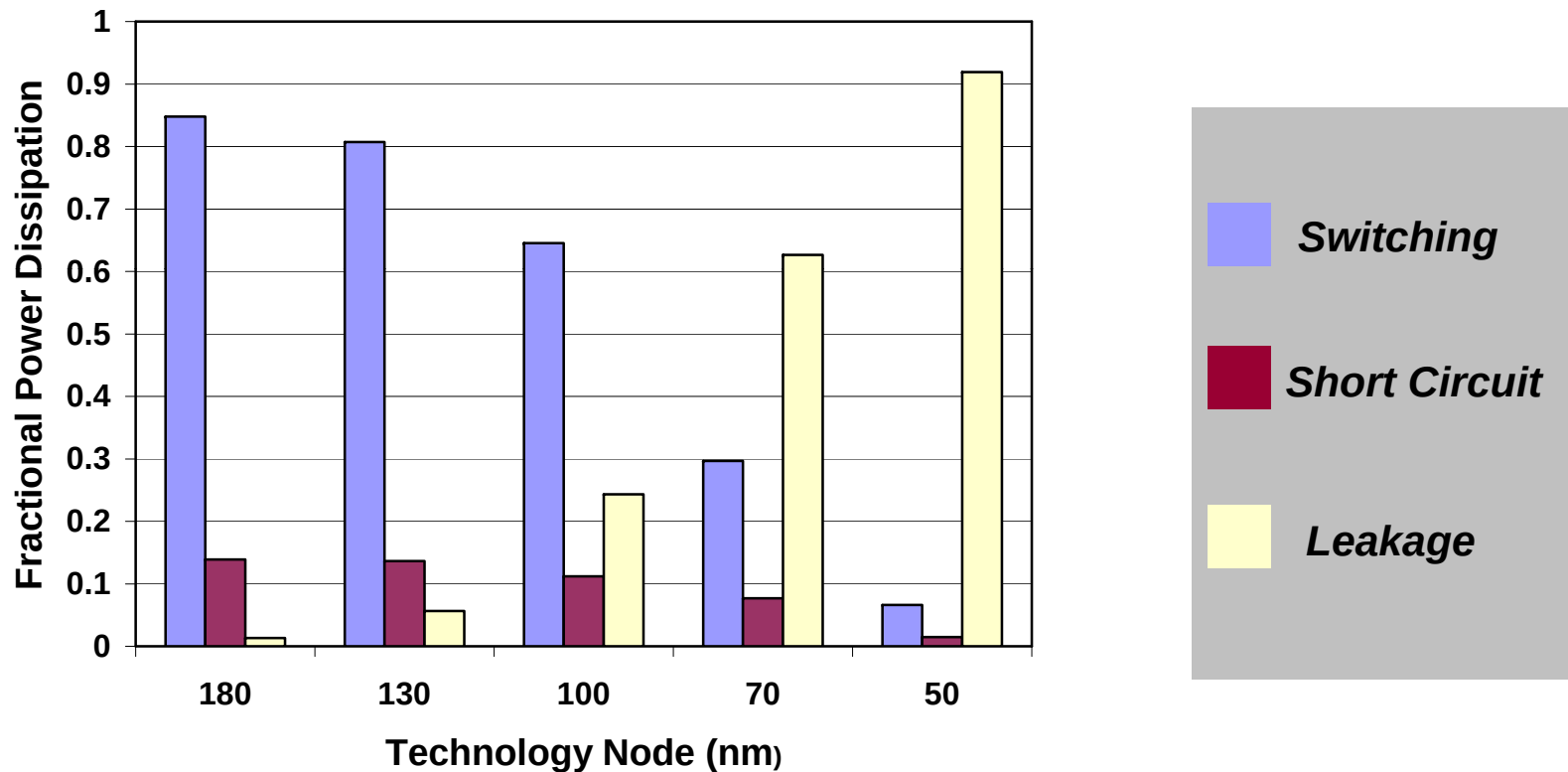
Optimization Results: 5% delay penalty



Normalized power per unit length decreases rapidly with scaling: **due to substantial increase in leakage current**

Optimization Results: 5% delay penalty

Relative contributions of the three components of PD

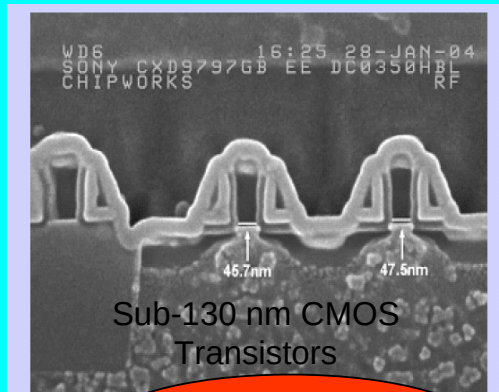


- Leakage power is the dominant component for advanced nodes
- Short circuit power is also non-trivial across all nodes

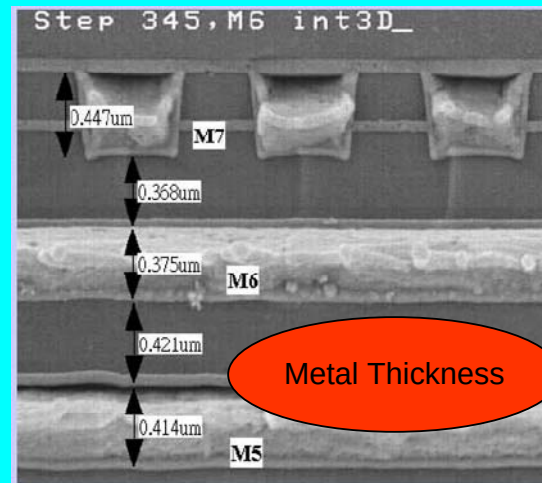
Implications of Parameter Variations

V. Wason and K. Banerjee, IEEE International Symposium on Low Power Electronic Design (ISLPED), 2005, pp. 131-136.

Parameter (P-V-T) Variations

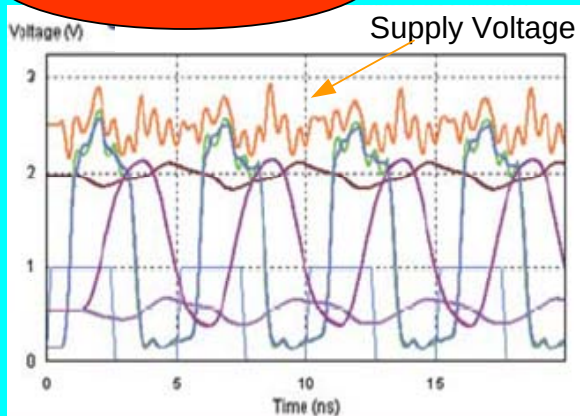


Transistor channel length

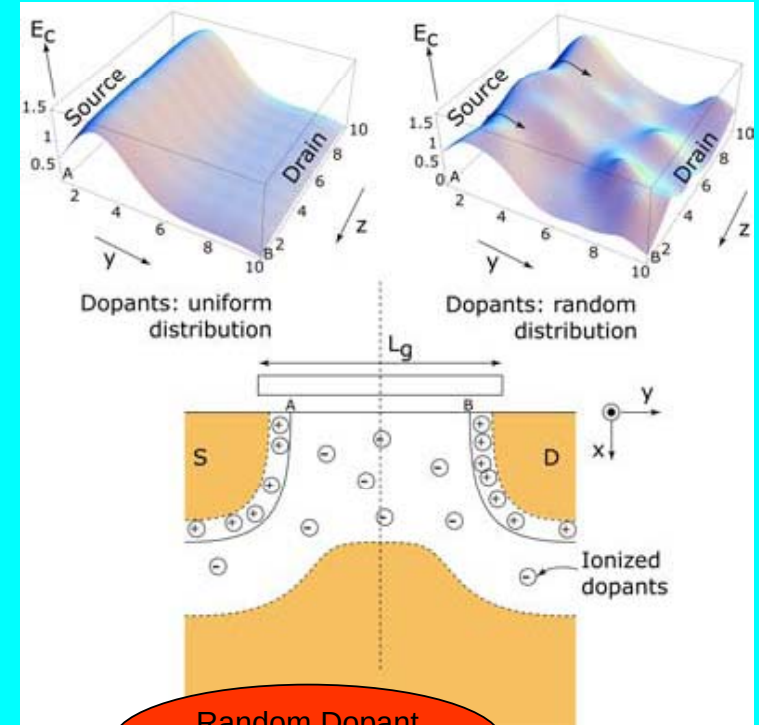
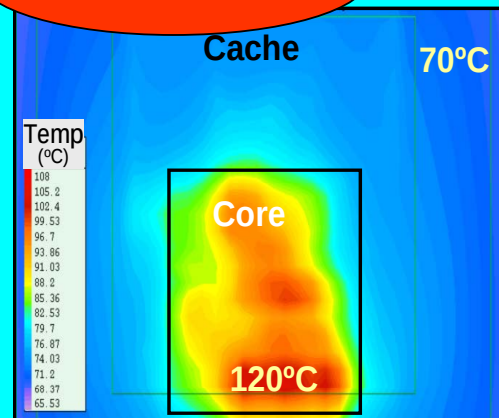


Metal Thickness

Power Supply Voltage



Chip Temperature

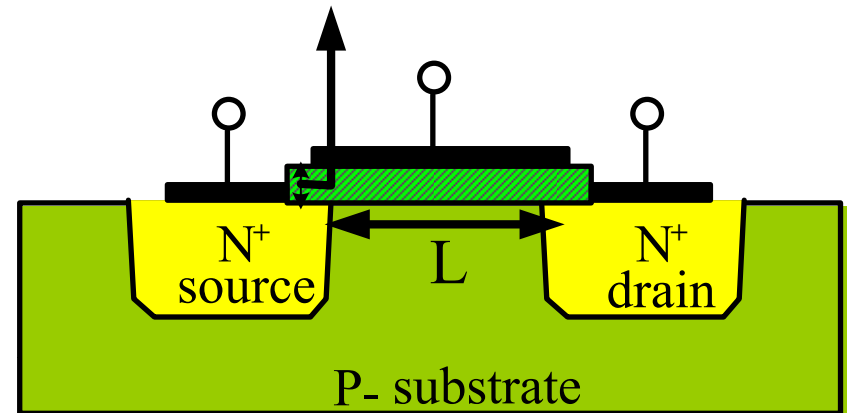


Random Dopant Fluctuations

Process Variations

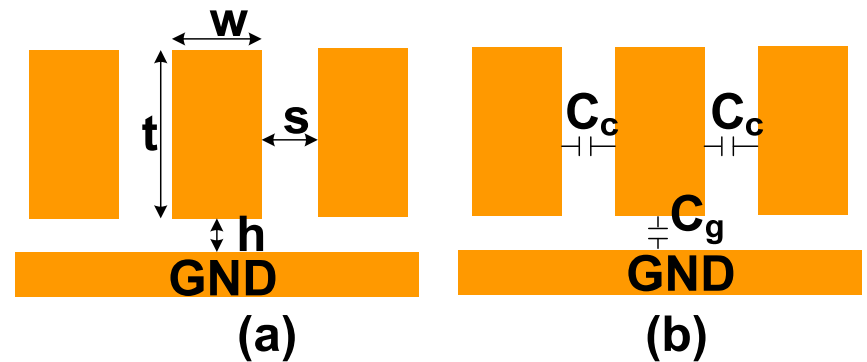
❑ Device Variations

- Channel Length (L)
- Oxide Thickness (t_{ox})
- Dopant Density (N_a)

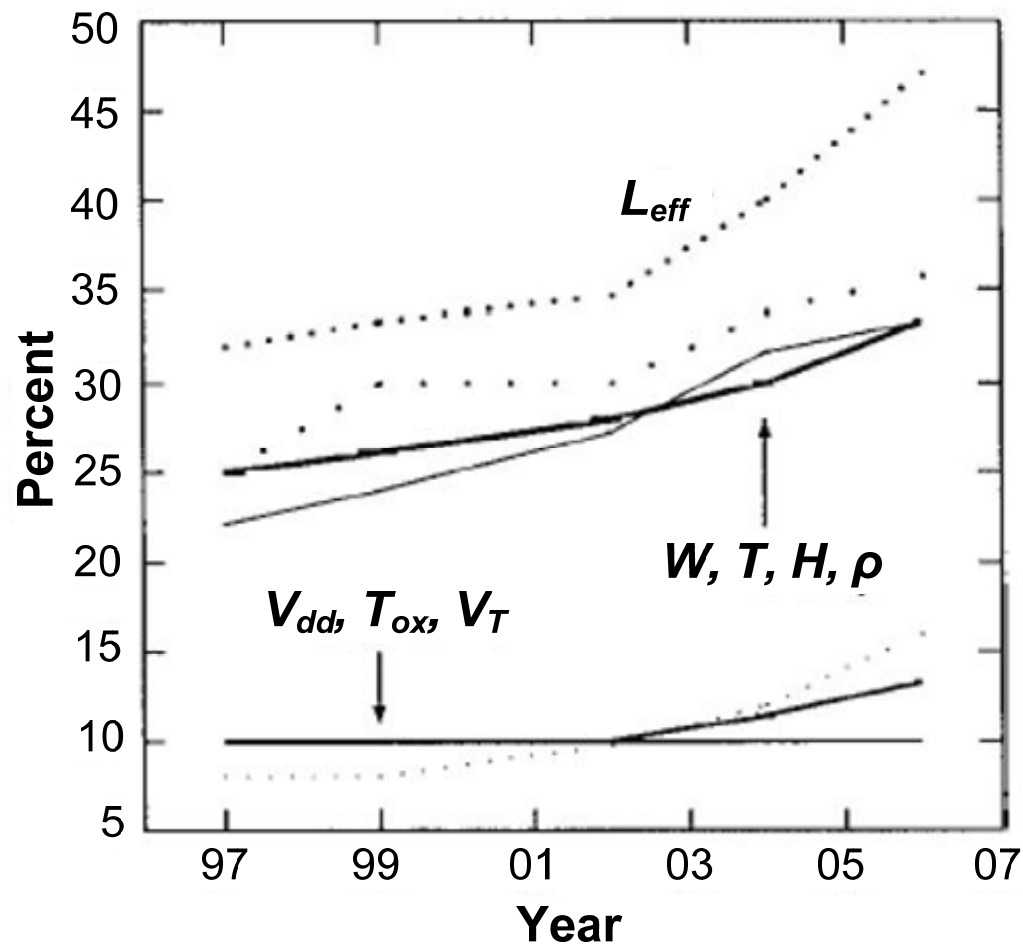


❑ Interconnect Variations

- Line Width (w)
- Spacing (s)
- Metal Thickness (t)
- Dielectric Thickness (h)

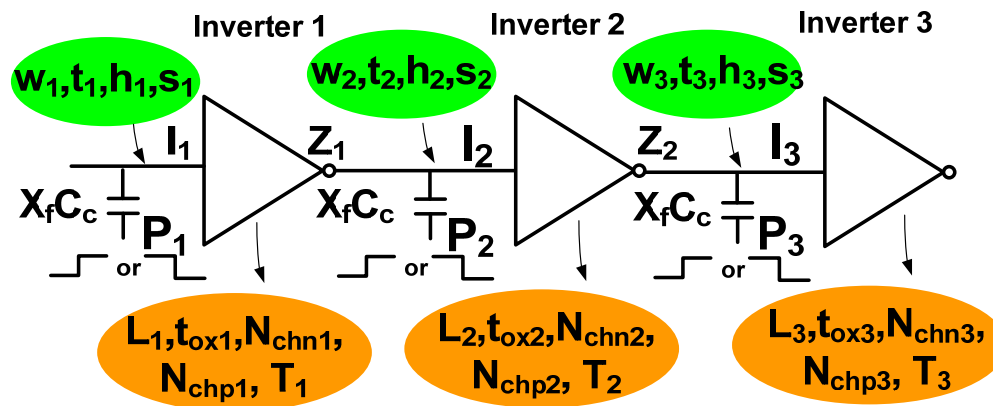


Device vs. Interconnect Variations



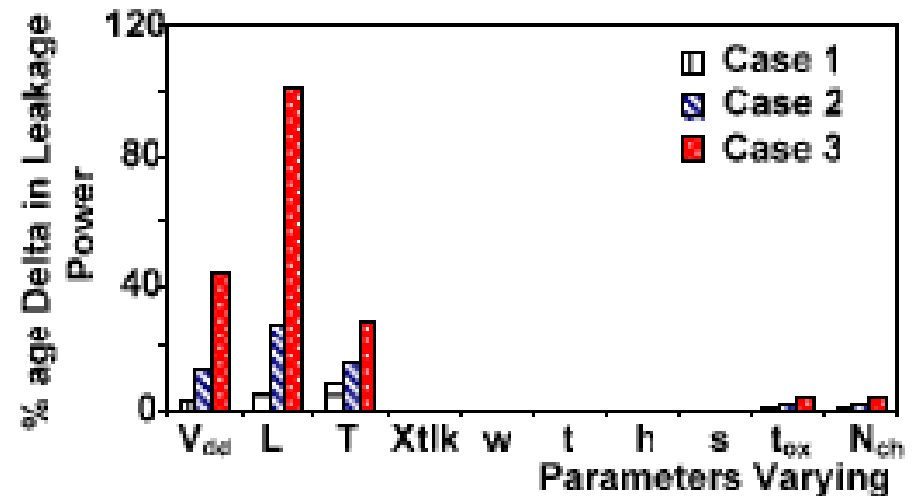
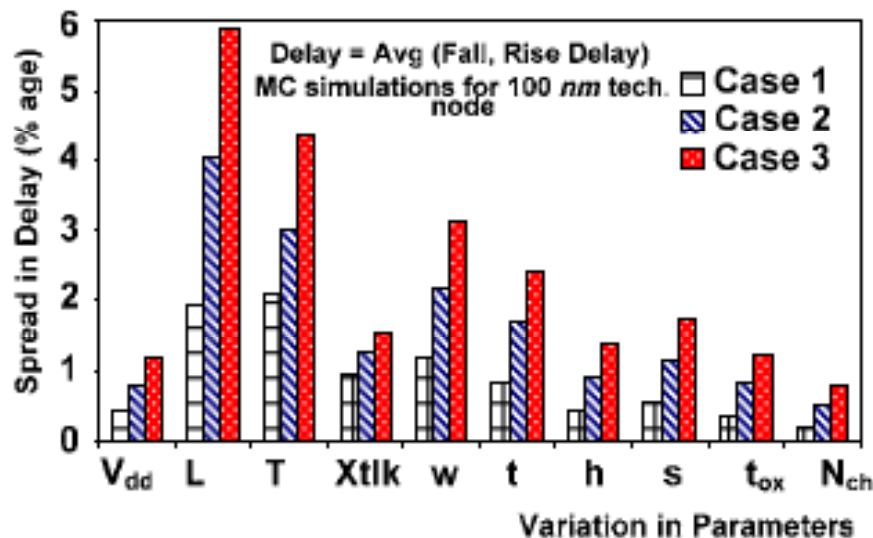
S. Nassif (IBM)

Sensitivity Analysis of Delay and Power



3 cases considered;
% variations increase
from case 1 to case 3

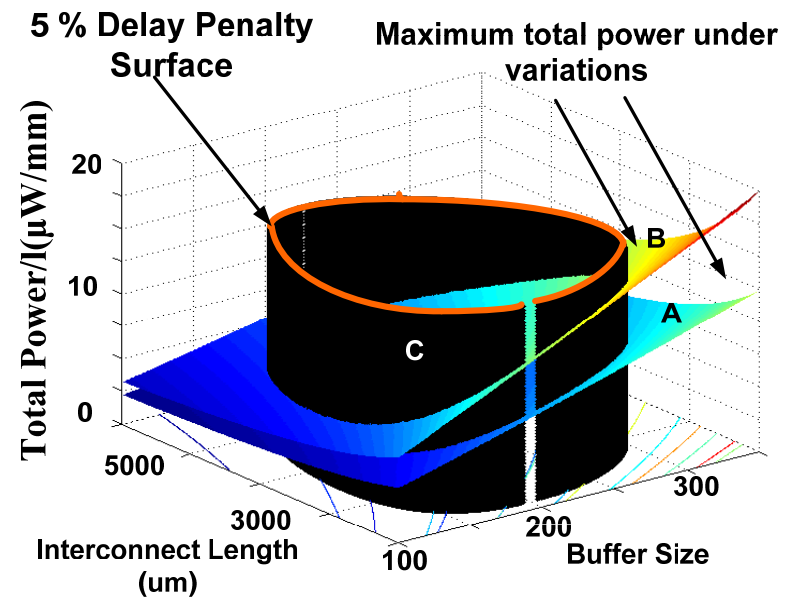
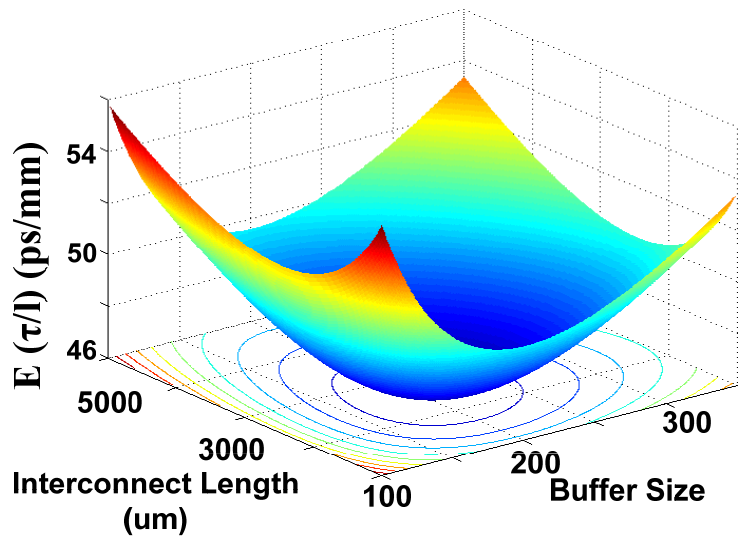
Wason and Banerjee, ISLPED 2005



■ Variations have significant impact on leakage power and delay

Interconnect Design Considering Variations

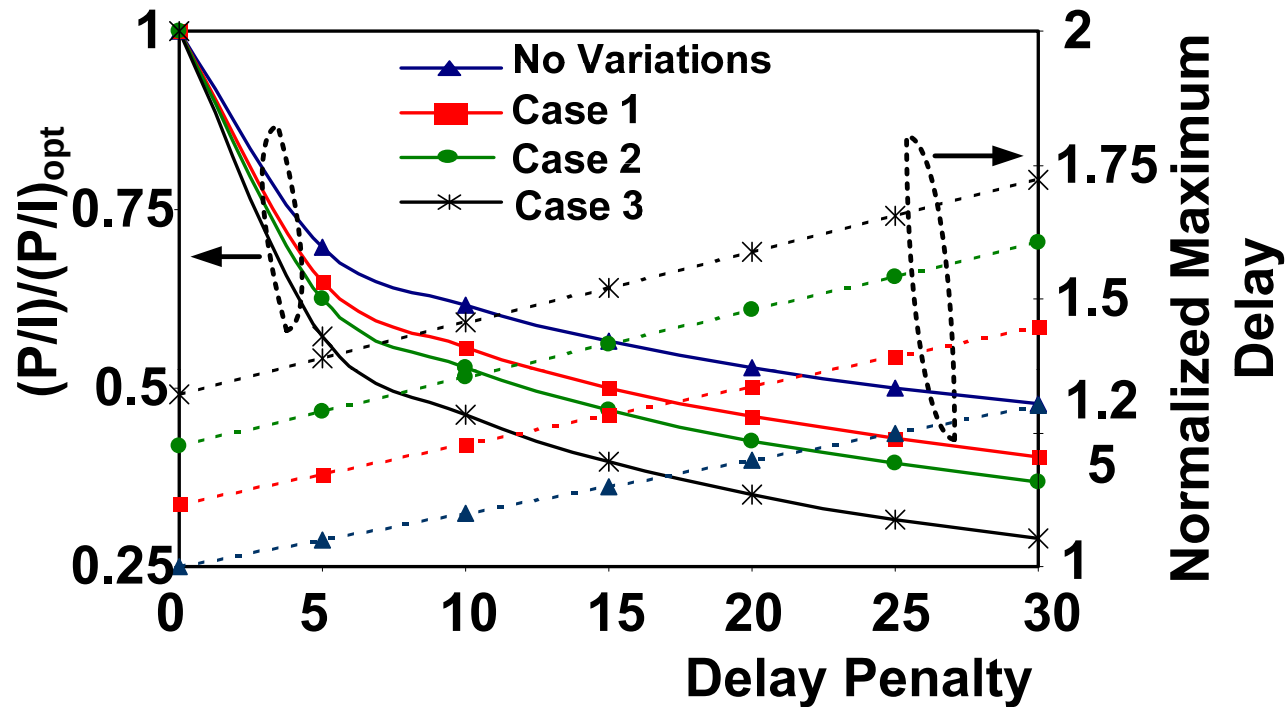
- Significant Implications for Interconnect Design
- Power-Optimal Repeater Insertion- tradeoff delay with power



Wason and Banerjee, ISLPED 2005

Variations can significantly impact power-optimal repeater insertion

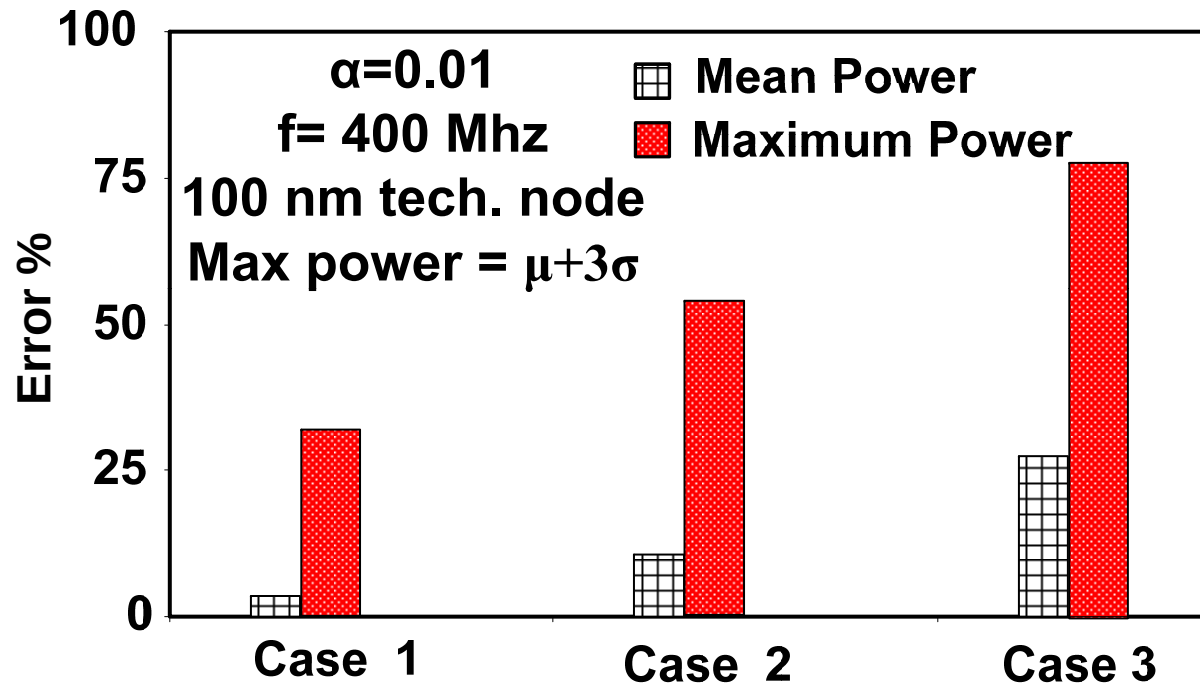
Optimization under Parameter Variations



Wason and Banerjee, ISLPED 2005

- The normalized delay increases under variations
- Power savings are higher for the same penalty in delay, as variations increase
- Importance of power-optimal repeater insertion increases

Implications for Power Estimations



■ Error in power estimation can be significant if variations are neglected