Area Optimization for Leakage Reduction and Thermal Stability in Nanometer Scale Technologies

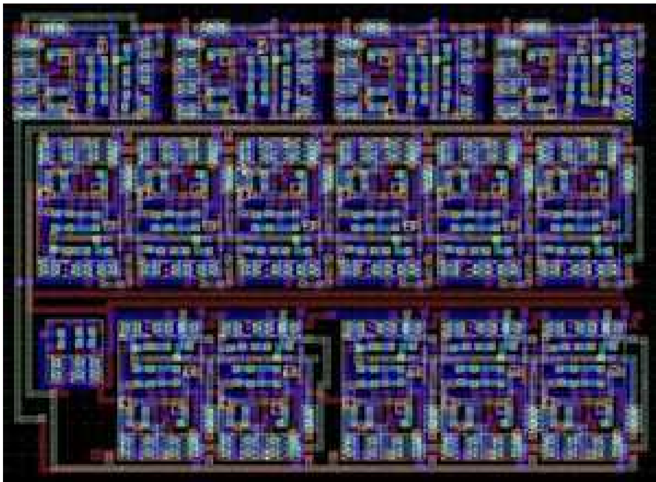
Ja Chun Ku and Yehea Ismail
Northwestern University, Evanston, IL

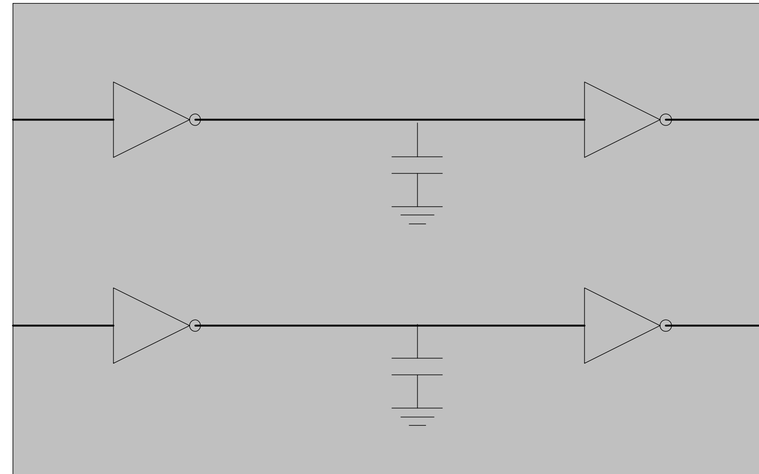
ASP-DAC 2006

Outline

- Introduction
- Analytical Models
  - Power
  - Delay
  - Thermal
- Area Optimization
  - Leakage Reduction
  - Thermal Stability
- Conclusion

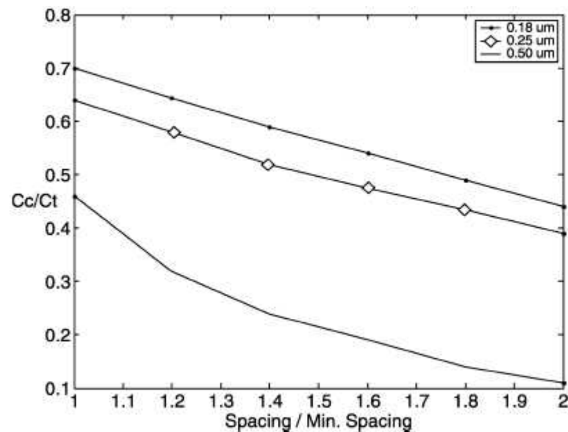
Traditional View



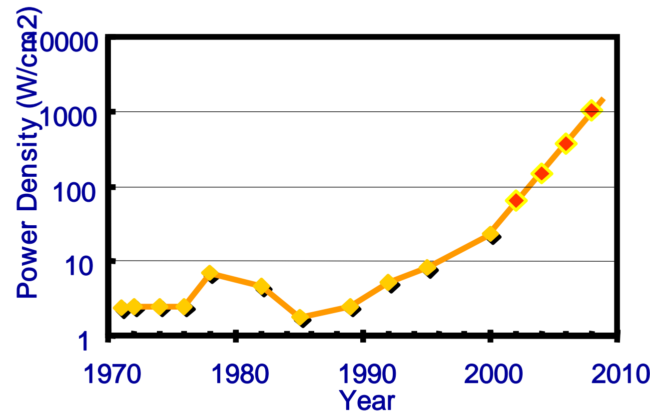


- Minimum Area
  - Minimized interconnect capacitance
  - Considered best for both power and delay

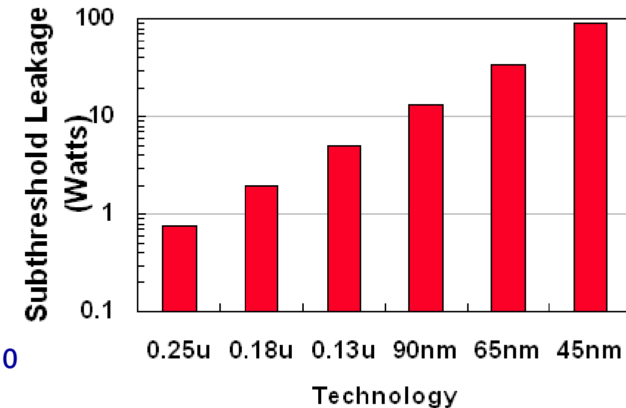
Trends in Nanoscale Technology



[R. Kumar]



[Intel]

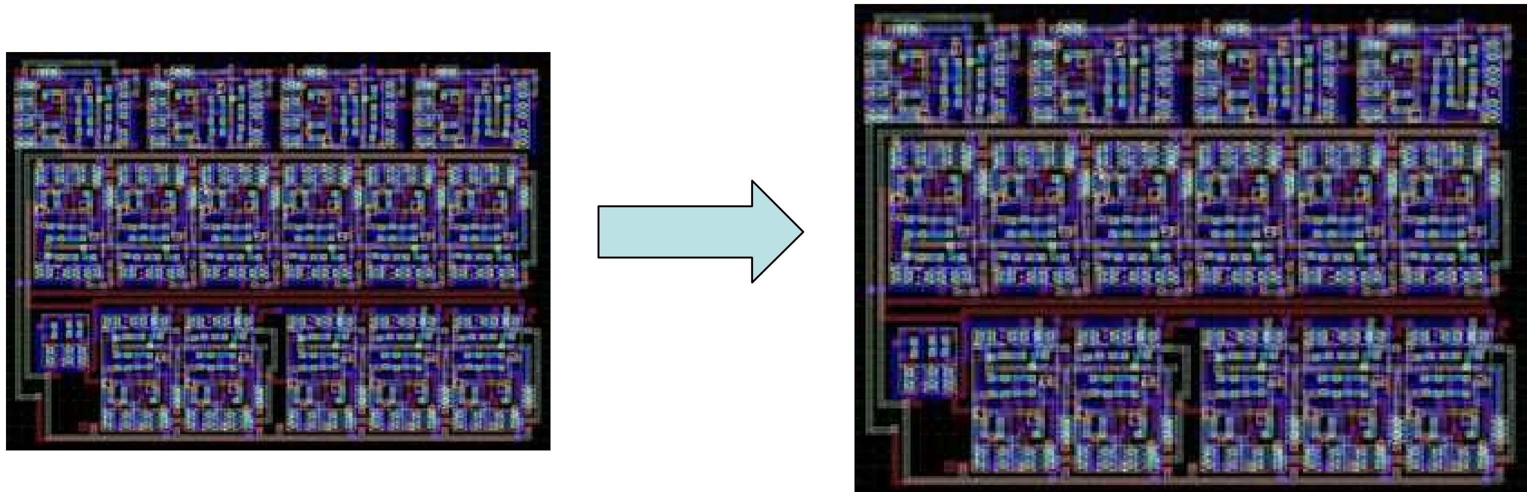


[S. Borkar]

1. Increasing coupling capacitance
2. Increasing power density (temperature)
3. Increasing leakage power

Minimum area: no longer optimum for power and/or delay!

Effects of Increasing Area



- Decrease in power density → Drop in junction temperature → Reduction in leakage power → …
- Increase in interconnect ground capacitance → Increase in dynamic power and delay
- Optimum point exists for power & delay tradeoff
- Prevents thermal runaway

Outline

- Introduction
- Analytical Models
  - Power
  - Delay
  - Thermal
- Area Optimization
  - Leakage Reduction
  - Thermal Stability
- Conclusion

Dynamic Power Model

$$P = P_{dynamic} + P_{leakage}$$

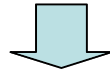
$$P_{dynamic} = Na\left(C_g + MCF_p C_c\right)V_{dd}^2 f$$

Dynamic power is temperature-independent as long as the clock frequency is not affected

Leakage Power Model

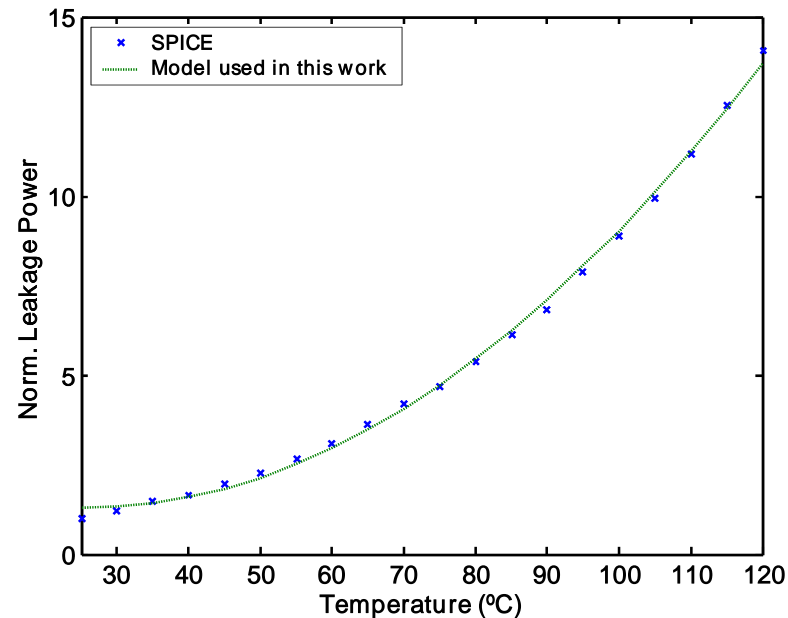
$$P_{leakage} = N I_{sub}(T) V_{dd}$$

$$I_{sub}(T) = \mu(T) C_{ox} \left(\frac{W}{L}\right)(m-1)\left(\frac{kT}{q}\right)^2 e^{q(V_g - V_t(T))/mkT}\left(1 - e^{-qV_{ds}/kT}\right)$$



$$I_{sub}(T) = W I_{sub0}\left\{c_1 (T_j - T_0)^2 + c_2 (T_j - T_0) + c_3\right\}$$

[Similar to Su et al. ISLPED '03]



Outline

- Introduction
- Analytical Models
  - Power
  - Delay
  - Thermal
- Area Optimization
  - Leakage Reduction
  - Thermal Stability
- Conclusion

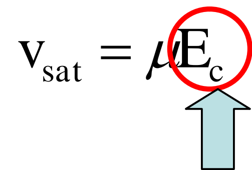
Delay Model

$$D = N_c \frac{\left(C_g + MCF_d C_c\right)V_{dd}}{2I_D(T)}$$

Drain current mostly in the saturation region during transition

$$I_D(T) = KWv_{sat}(T)\left(V_{dd} - V_t(T)\right)^{\alpha}$$

$$v_{sat} = \mu E_c$$



Not a constant!

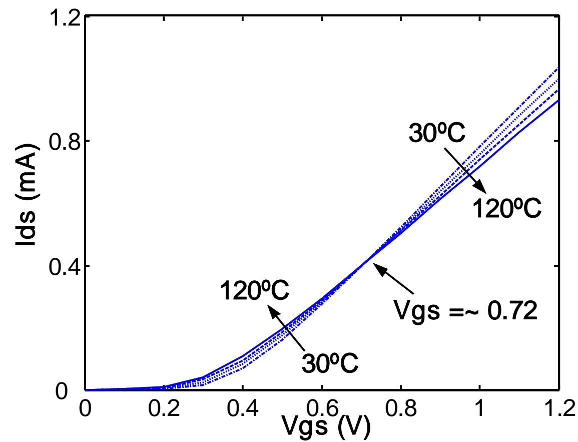
$$v_{sat}(T) = v_{sat}(T_0) - \eta(T - T_0)$$

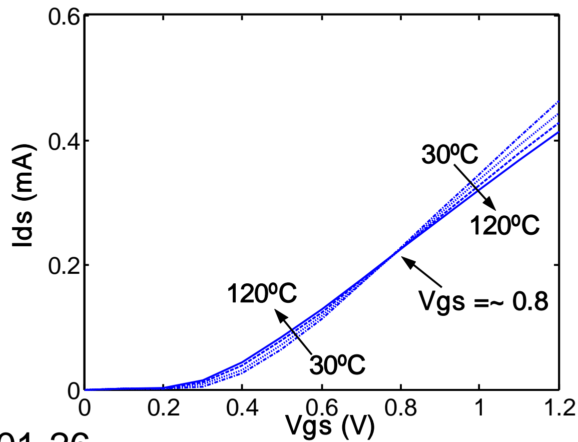
$$V_t(T) = V_t(T_0) - \kappa(T - T_0)$$

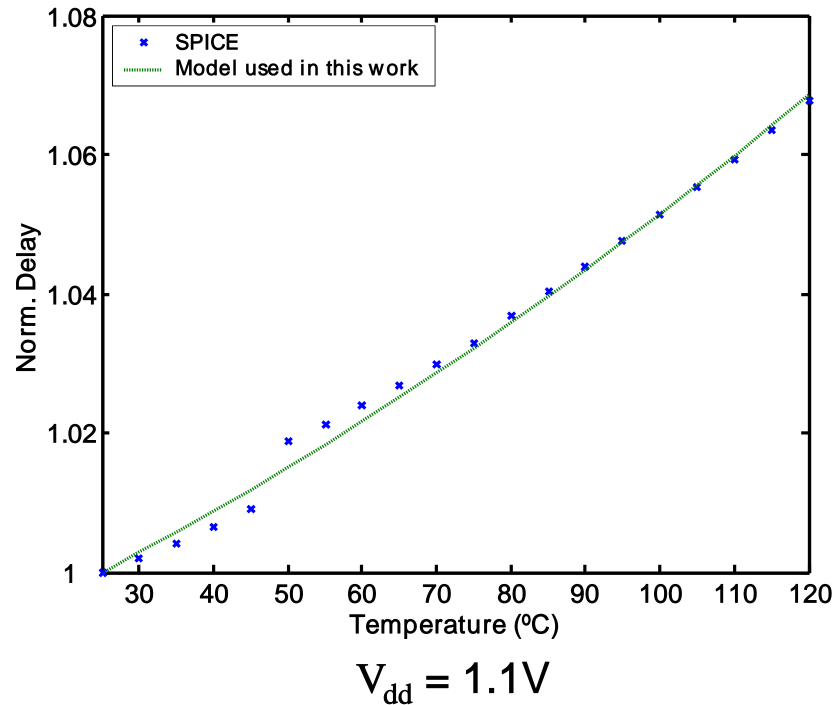
Delay Model (cont.)

$$I_D(T) = KWv_{sat}(T)(V_{dd} - V_t(T))^{\alpha}$$

Temperature dependence of I_D is dependent on the value of V_{dd}





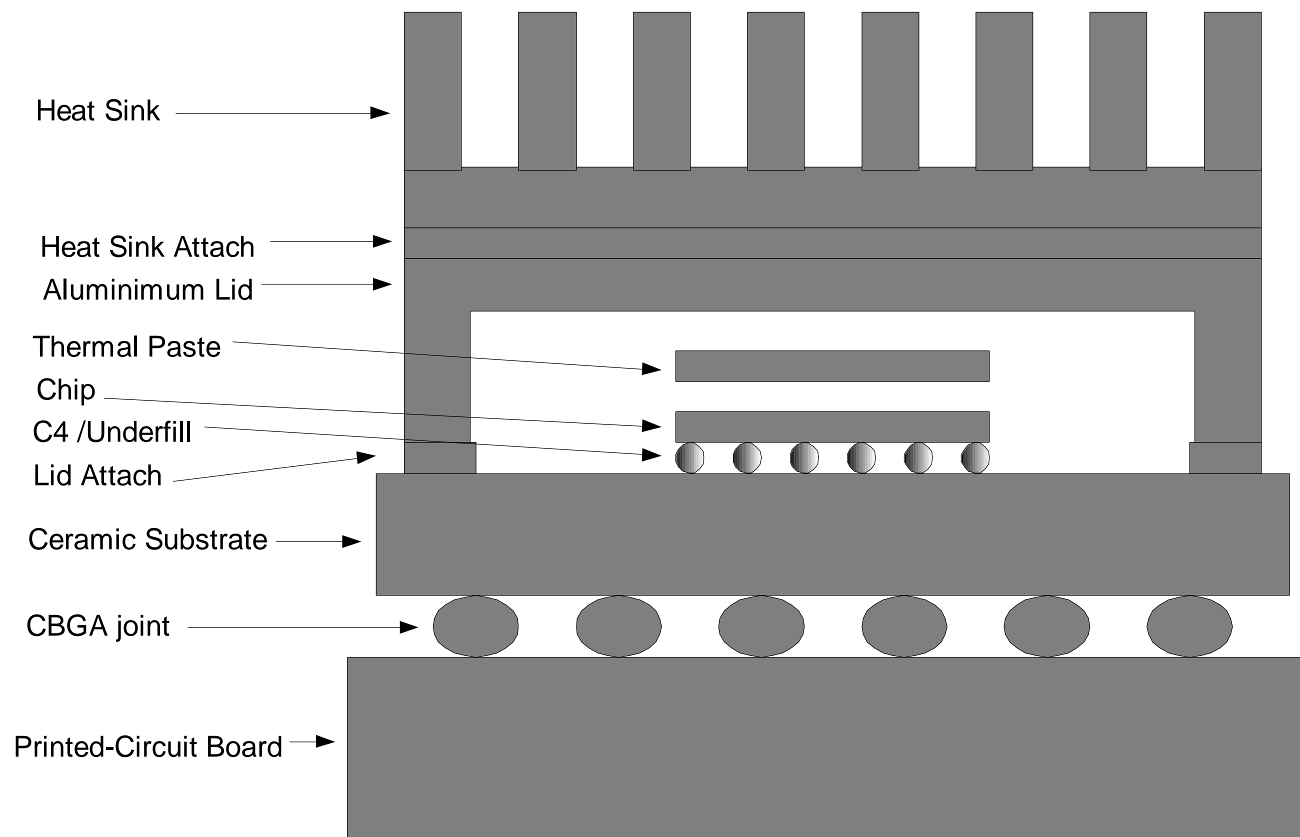


V_{dd} = 1.1V

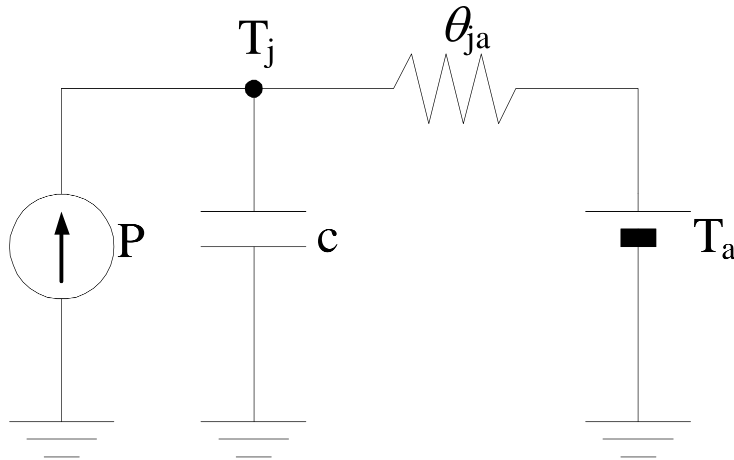
Outline

- Introduction
- Analytical Models
  - Power
  - Delay
  - Thermal
- Area Optimization
  - Leakage Reduction
  - Thermal Stability
- Conclusion

Typical Flip-Chip C4 Package

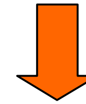


Thermal Model



$$\theta_{ja} c \frac{dT_j}{dt} + T_j = P(T_j)\theta_{ja} + T_a$$

$$\theta_{ja} = \theta_{thermalpaste} + \theta_{heatsink}$$



$$\theta = R_{th} / A_c$$

$$T_j = \left(P(T_j) / A_{chip} \right) R_{thermalpaste} + P(T_j)\theta_{heat\,sin\,k} + T_a$$

Thermal Model (cont.)

$$\theta_{ja} c \frac{dT_j}{dt} = AT_j^{\ 2} + BT_j + C$$

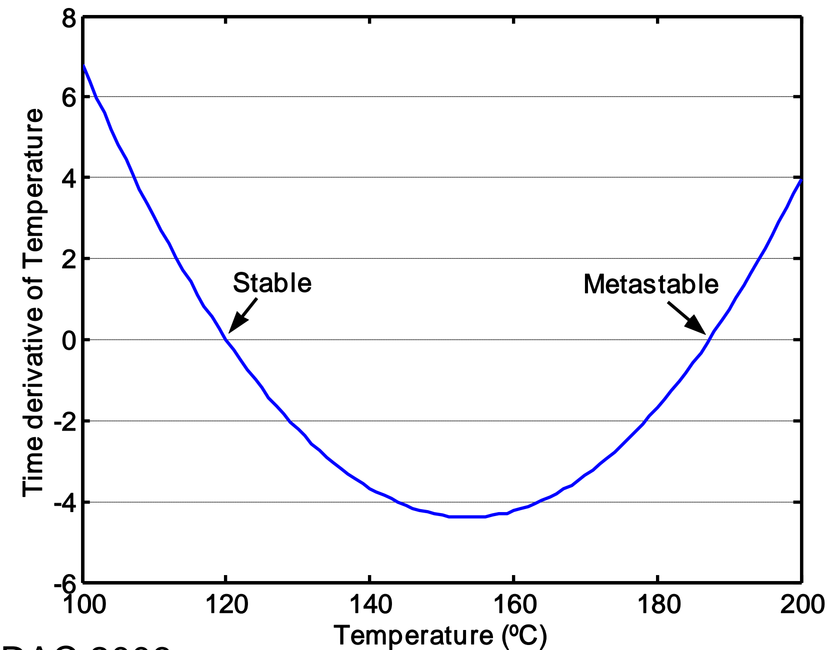
where

$$A = c_1 \theta_{ja} NWI_{sub0} V_{dd}$$

$$B = \theta_{ja} NWI_{sub0} V_{dd} \left(c_2 - 2c_1 T_0\right) - 1$$

$$C = T_a + \theta_{ja} NV_{dd} \left\{ a\left(C_g + MCF_p C_c\right) V_{dd} f + WI_{sub0} \left(c\, T_0^{\ 2} - c_2 T_0 + c_3\right) \right\}$$

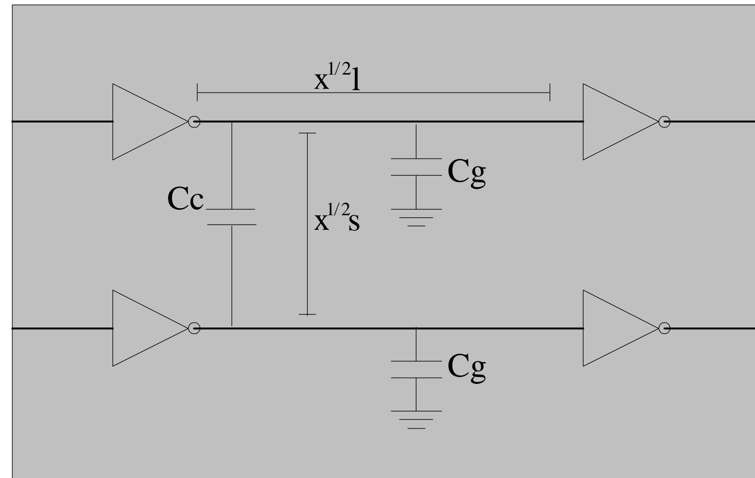
$$T_j = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}$$



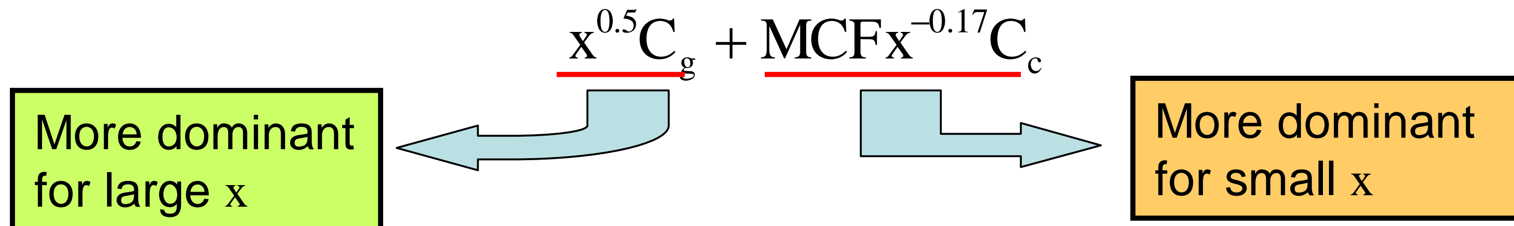
Outline

- Introduction
- Analytical Models
  - Power
  - Delay
  - Thermal
- Area Optimization
  - Leakage Reduction
  - Thermal Stability
- Conclusion

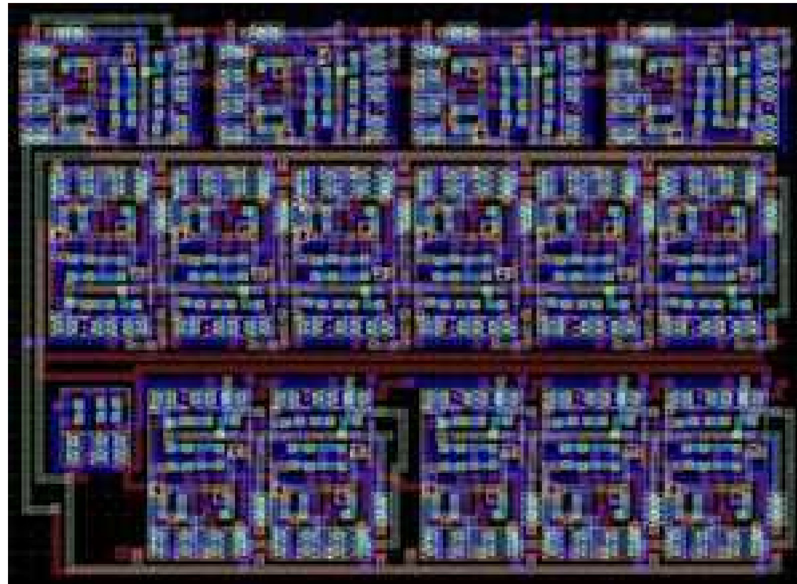
Impact of Area Scaling



- Ground capacitance:
  increases linearly as l increases
- Coupling capacitance:
  increases linearly as l increases
  decreases superlinearly as s increases



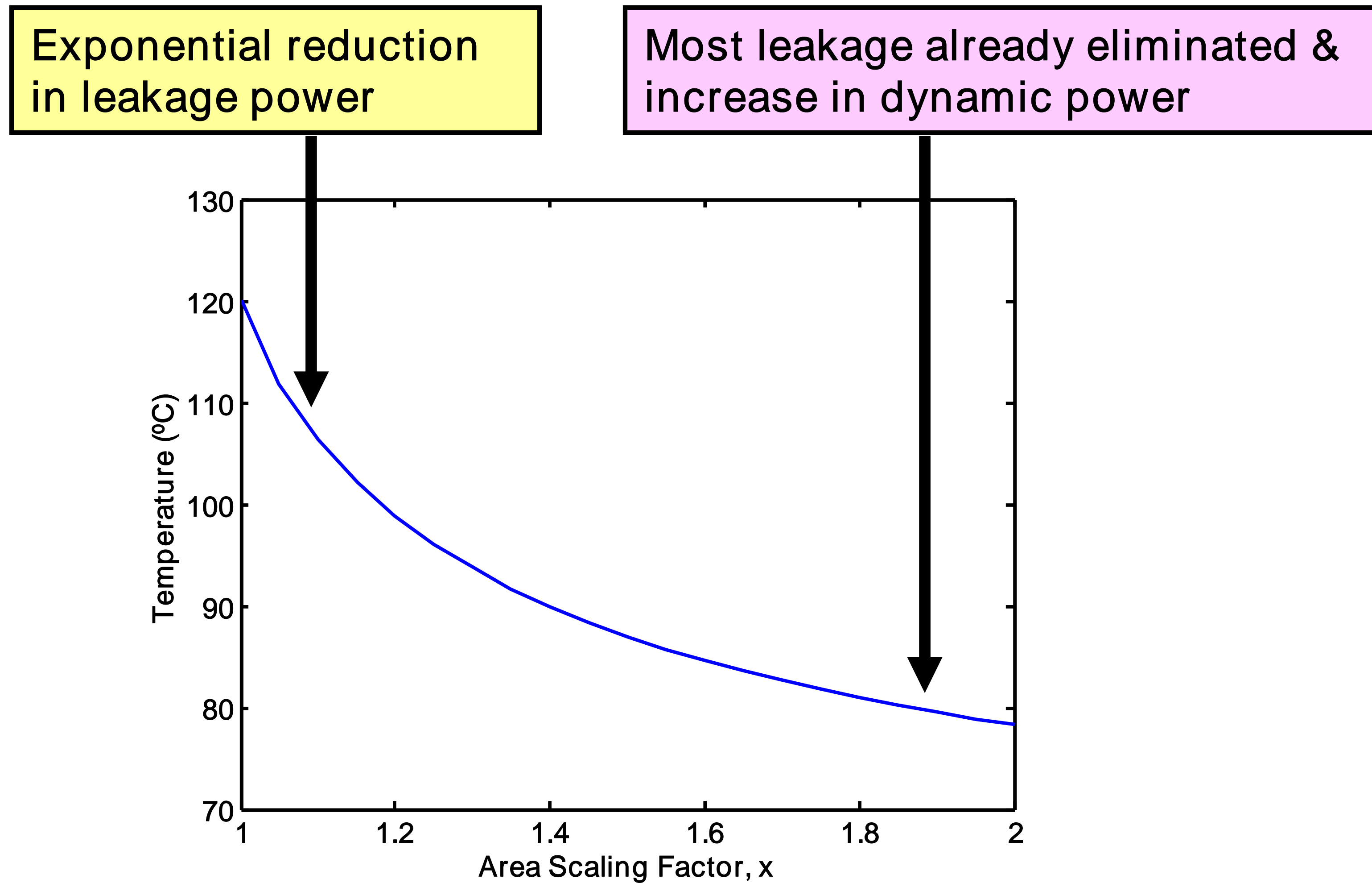
Simulation



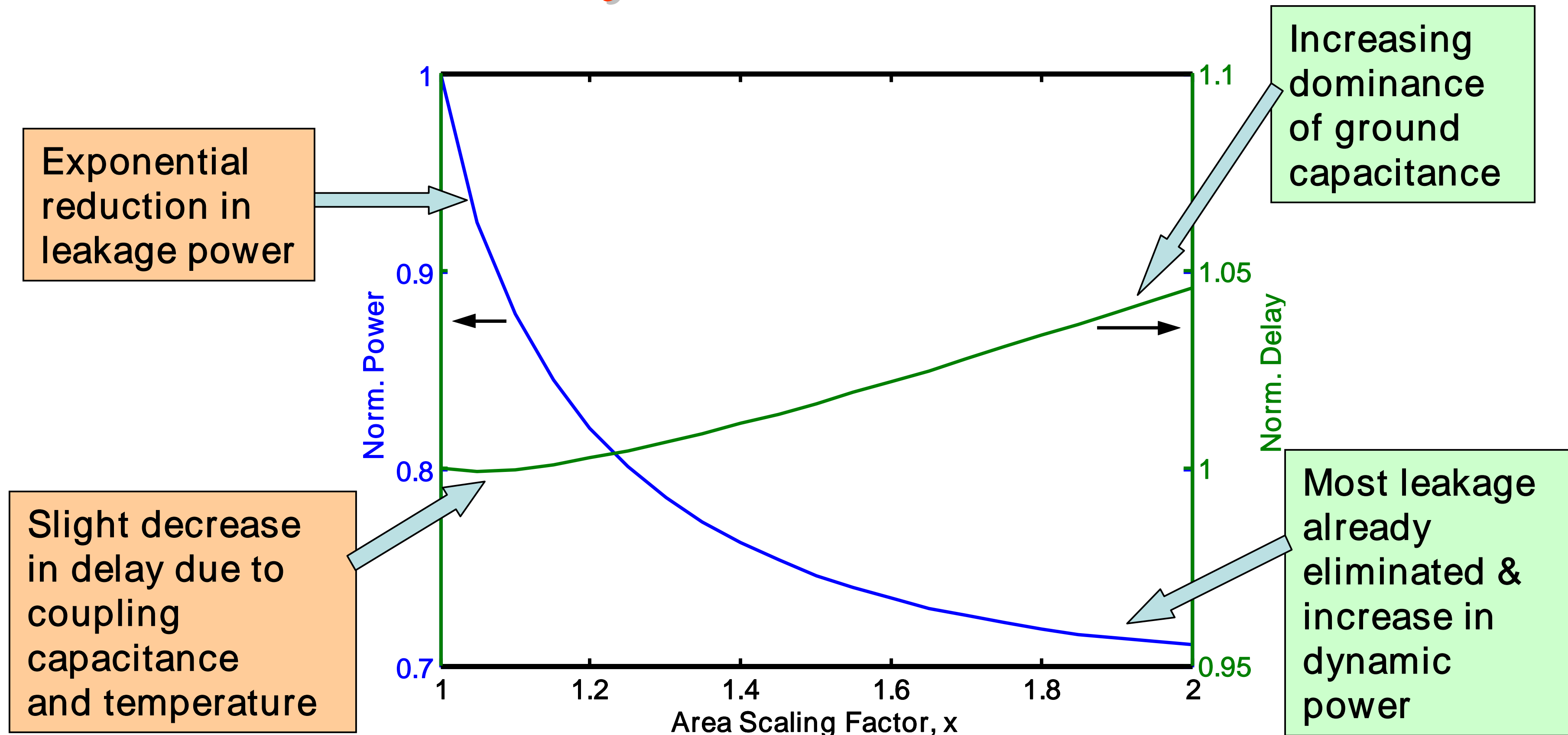
16-bit adder

- Parameters extracted from layout
- Scaled for 70nm BPTM technology

Temperature vs. Area



Power & Delay vs. Area

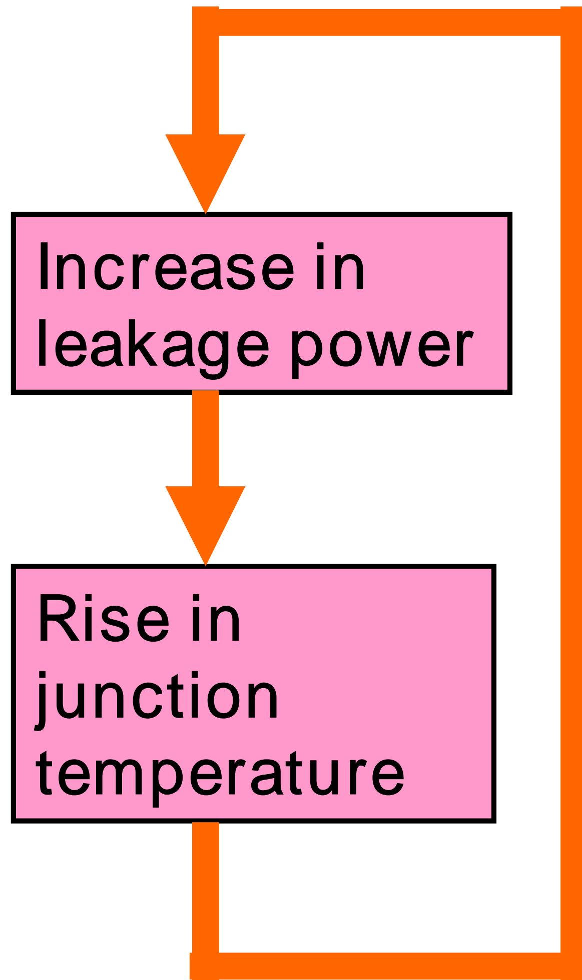


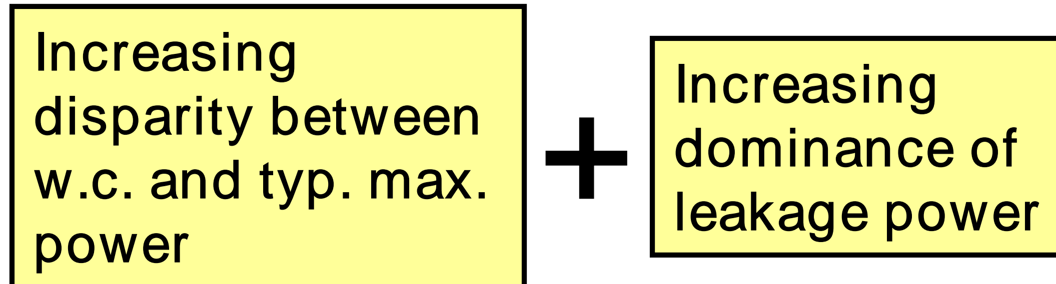
e.g. 15% increase in area
→ 17% reduction in power & no change in delay

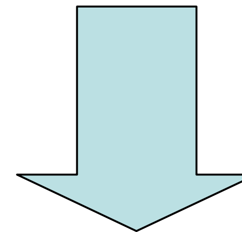
Outline

- Introduction
- Analytical Models
  - Power
  - Delay
  - Thermal
- Area Optimization
  - Leakage Reduction
  - Thermal Stability
- Conclusion

Possibility of Thermal Runaway







- Possibility of thermal runaway increases
- Dynamic techniques to prevent thermal runaway

Thermal Stability Condition

$$\theta_{ja} c \frac{dT_j}{dt} = AT_j^{\,2} + BT_j + C$$

where

$$A = c_1 \theta_{ja} NWI_{sub0} V_{dd}$$

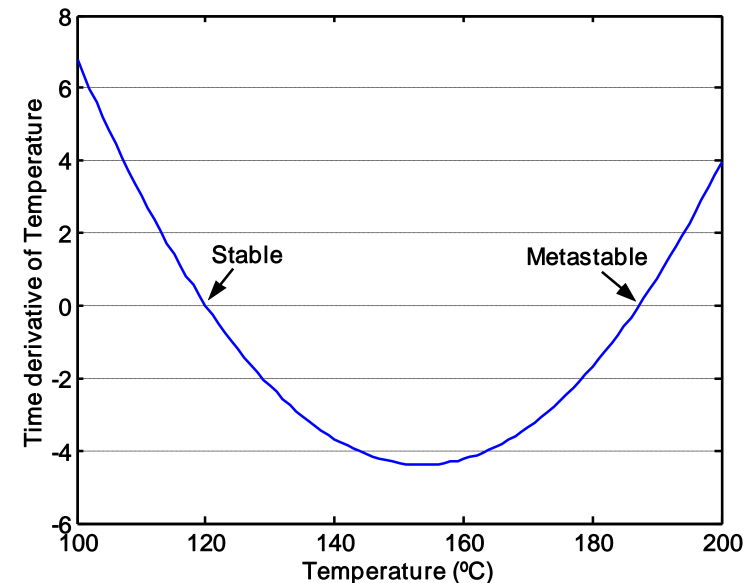
$$B = \theta_{ja} NWI_{sub0} V_{dd} \left(c_2 - 2c_1 T_0\right) - 1$$

$$C = T_a + \theta_{ja} NV_{dd} \left\{ a\left(C_g + MCF_p C_c\right) V_{dd}\, f + WI_{sub0}\left(c\, T_0^{\,2} - c_2 T_0 + c_3\right)\right\}$$

$$T_j = \frac{-B - \sqrt{B^2 - 4AC}}{2A}$$

$$B^2 - 4AC \geq 0$$

➡ Condition for thermal stability



Existing Techniques

- Existing dynamic techniques to prevent thermal runaway
  - Dynamic frequency scaling
  - Dynamic voltage scaling
  - Global clock gating

- Common drawbacks
  - Whole chip has to be slowed down
  - Require sensors on-chip: sensitive to lithographic & supply-current variations

Critical Area for Thermal Stability

$B^2 - 4AC = 0$ ⟹ Critical condition for thermal stability

$$X\theta_c^{\,2} - Y\theta_c + Z = 0$$

where

$$X = N^2 W I_{sub0} V_{dd}^{\,2}\left\{W I_{sub0}\left(c_2^{\,2} - 3c_1^{\,2} T_0^{\,2} - 4c_1 c_3\right) - 4c_1 a\left(C_g + MCF_p C_c\right)V_{dd}\, f\right\}$$

$$Y = NWI_{sub0}V_{dd}\left(2c_2 - 4c_1T_0 + 4c_1T_a\right)$$

$$Z = 1$$

$$\theta_c = \frac{Y - \sqrt{Y^2 - 4X}}{2X}$$

$$A_{min} = \frac{R_{thermalpaste}}{\theta_c - \theta_{heat\,sin\,k}}$$

Conclusion

- Area used as a degree of freedom to
  - Reduce leakage
  - Prevent thermal runaway
- Increasing area:
  - Reduces thermal resistance → lower temperature → leakage power significantly reduced
  - Increases interconnect ground capacitance
- Analytical model for steady-state temperature derived
- Increasing area of a hot spot can significantly reduce its power consumption
- Possibility and condition for thermal runaway discussed
- Use of area to maintain thermal stability w/o slowing down the whole chip

Thank you!