Implications on the Design

Ramon Canal

NCD – Master MIRI

Departament d'Arquitectura de Computadors
UNIVERSITAT POLITÈCNICA DE CATALUNYA

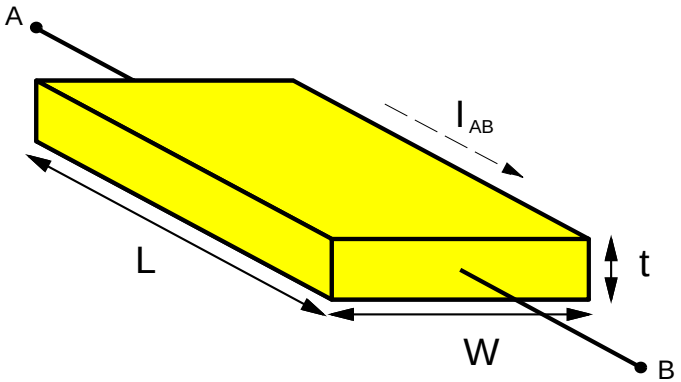
Agenda

- VLSI Basics
  - Resistance:
  - Capacity:
- Energy Consumption
  - Static
  - Dynamic
  - Thermal maps
- Voltage Scaling
- Metrics

Resistance

- In general:

$$R_{AB} = \rho \frac{L}{tW} \qquad I_{AB} = \frac{V_{AB}}{R_{AB}}$$



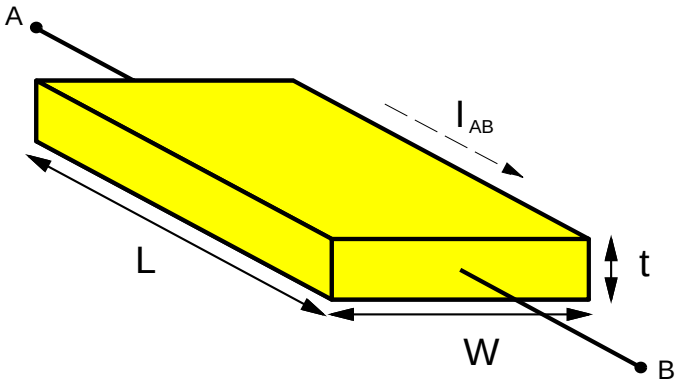
Material	ρ(Ω-m)
Silver (Ag)	1.6 x 10^{-8}
Cupper (Cu)	1.7 x 10^{-8}
Gold (Au)	2.2 x 10^{-8}
Aluminium (Al)	2.7 x 10^{-8}
Tugnsten (W)	5.5 x 10^{-8}

- For a given technology
  - *t* is constant
  - Basic unit of measurement: Square resistance (L=W=2λ): ❑, R_s

Resistance

- In general:

$$R_{AB} = \rho \frac{L}{tW} \qquad I_{AB} = \frac{V_{AB}}{R_{AB}}$$



Material	Sheet Res. (Ω/❐)
n, p well diffusion	1000 to 1500
n+, p+ diffusion	50 to 150
n+, p+ diffusion with silicide	3 to 5
polysilicon	150 to 200
polysilicon with silicide	4 to 5
Aluminum	0.05 to 0.1

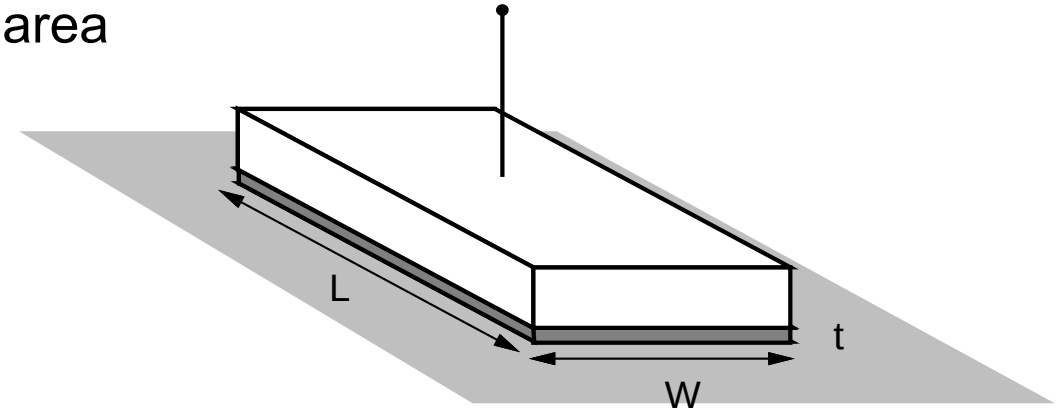
- For a given technology

$$R_s = \frac{\rho}{t}$$

  - *t* is constant
  - Basic unit of measurement: Square resistance (L=W=2λ): ❑, R_s

Capacity

- Directly depending on the area
- In general:



$$C_{LW} = \frac{\varepsilon_0 \varepsilon_{ins}}{t} LW$$

- For a given technology:
  - The permitivity ε_{ins} is constant
  - *t* is constant
  - Basic unit of measurement: (L=W= 2λ): Cg

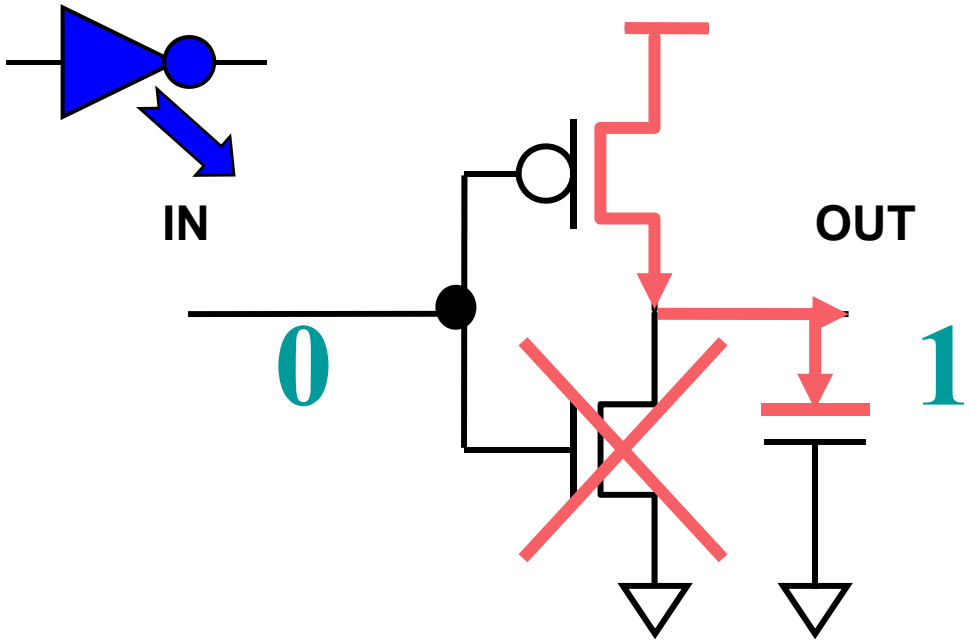
$$C_g = \frac{\varepsilon_0 \varepsilon_{ins}}{t}(2\lambda \times 2\lambda) \quad \rightarrow \quad C_{LW} = \frac{LW}{4\lambda^2} C_g$$

Power

- Two types:
  - Dynamic: proportional to the activity
  - Static: “just for being there”

Dynamic Power

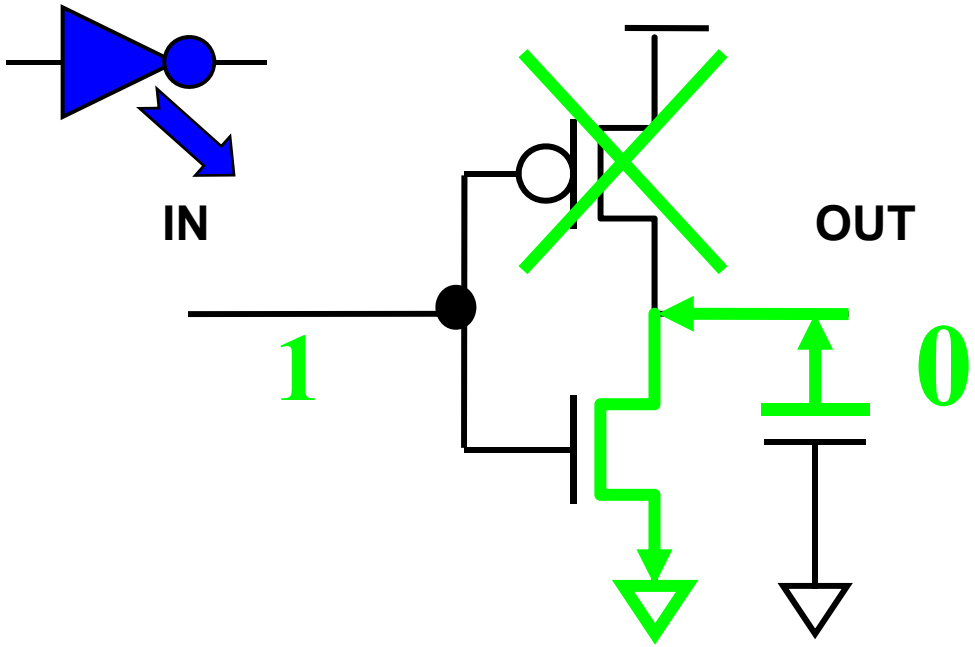
- Charging and discharging capacitors



- The capacitors are the transistors/buses connected to the output

Dynamic Power

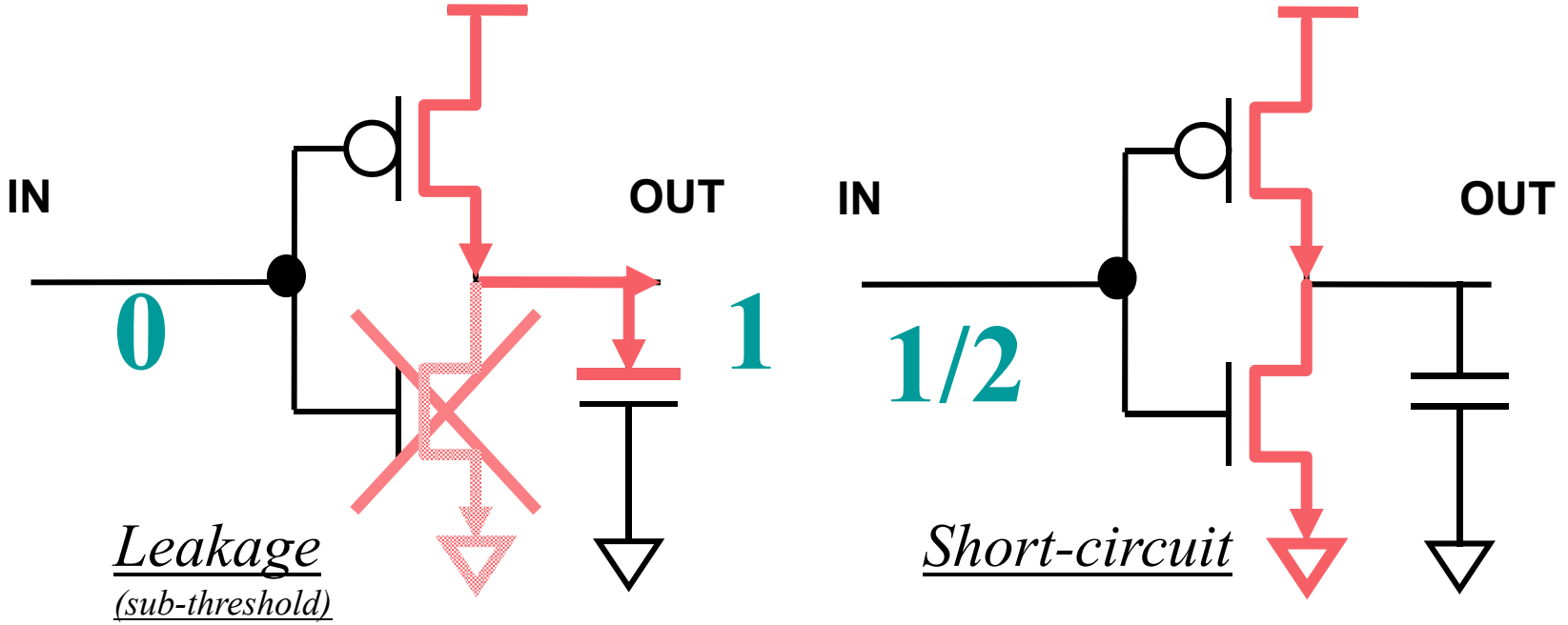
- Charging and discharging capacitors



- The capacitors are the transistors/buses connected to the output

Static Power

- Short-circuit and leakage currents of the transistor



- Leakage is growing dramatically
  - ~7% (0'25µm), ~20% in (130nm) process technology, ~50% in 65nm

Power and Energy

Energy

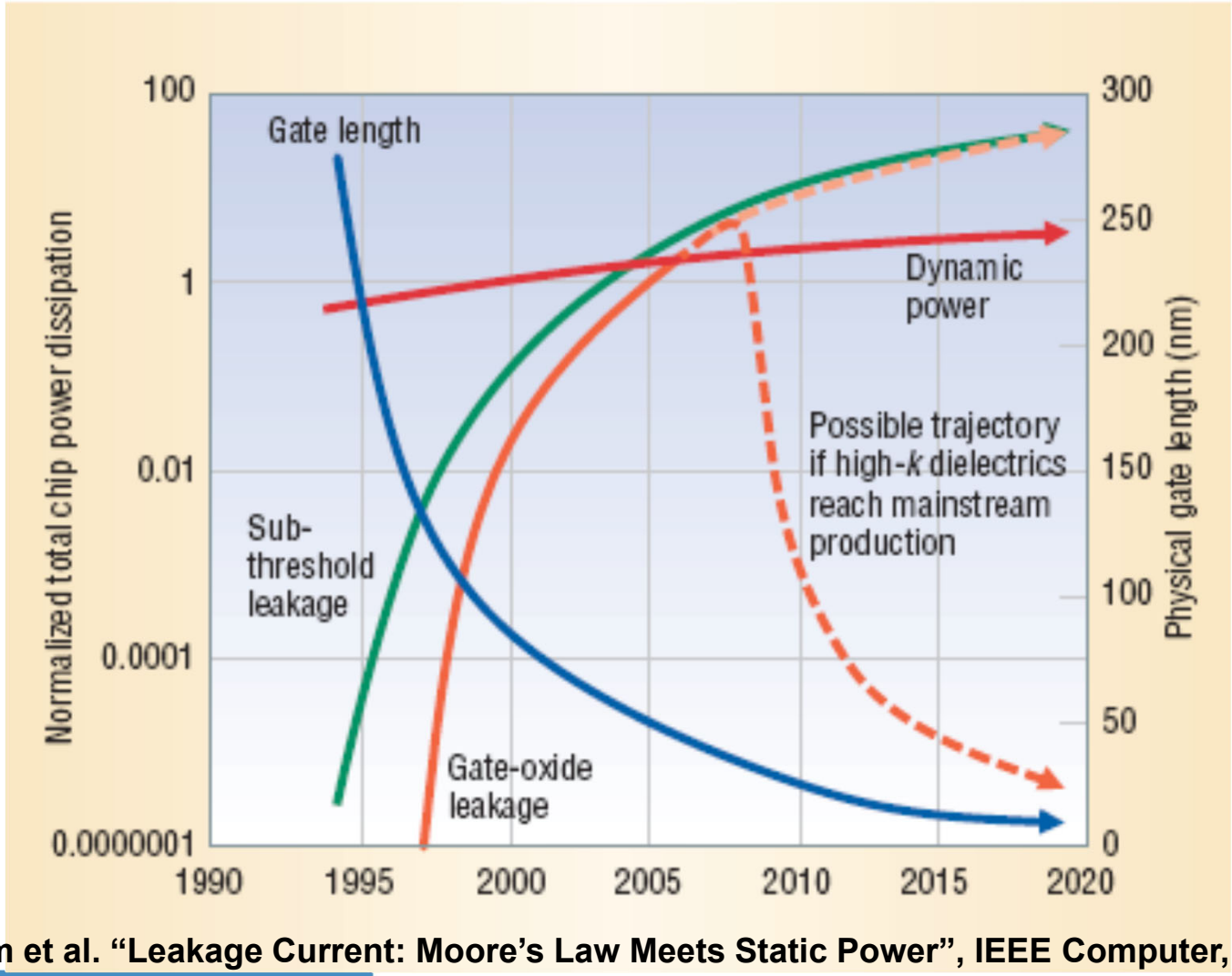
- *"the potential for causing changes"*
  - Mesured in Joules
- Important for
  - Battery duration – less need of energy ➔ longer life
  - Electricity bill – less need of energy ➔ saving $$

Power and Energy

Energy

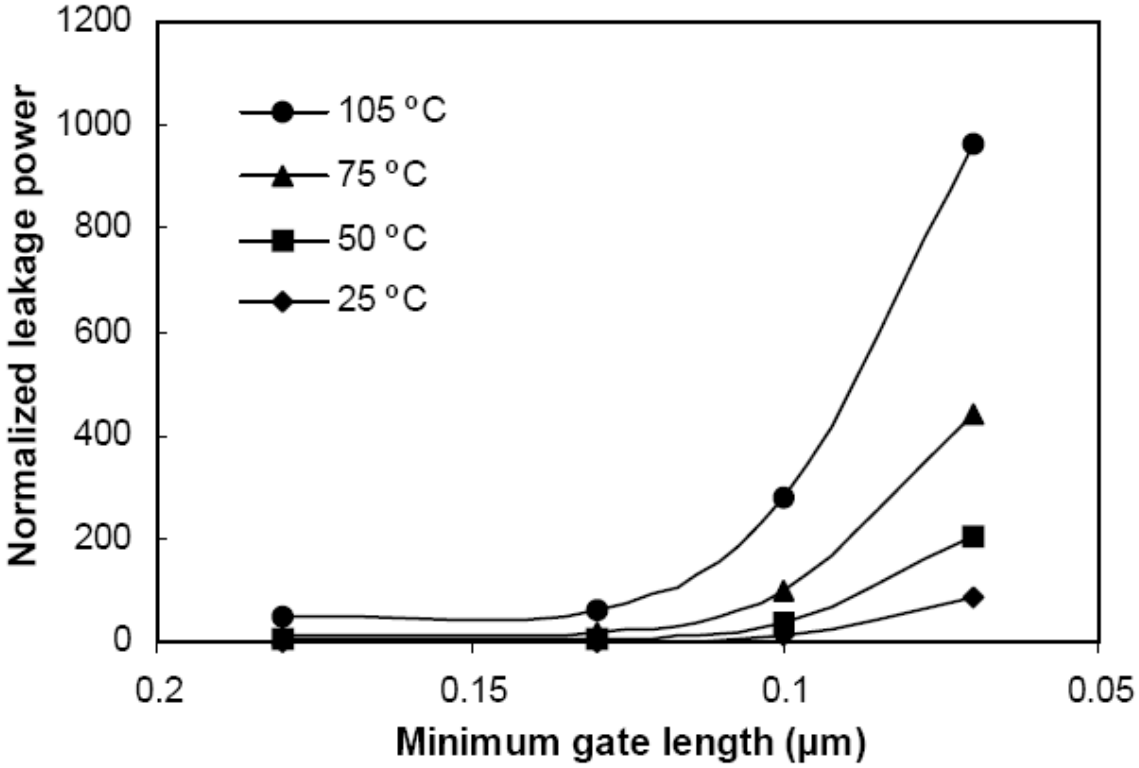
- Dynamic energy (CMOS):
  - Proportional to the activity, the capacitance and the square of the supply voltage ($E = \alpha C V^2$)
- Static energy (CMOS):
  - Short-circuit: $E_{sc} = t_{sc}\ V_{DD}\ I_{peak}\ P_{0\rightarrow1}$
  - Leakage: $E_{leakage} = V_{DD}\ I_{leakage}$

Importance of leakage



Kim et al. "Leakage Current: Moore's Law Meets Static Power", IEEE Computer, 2003.

Thermal effects on leakage

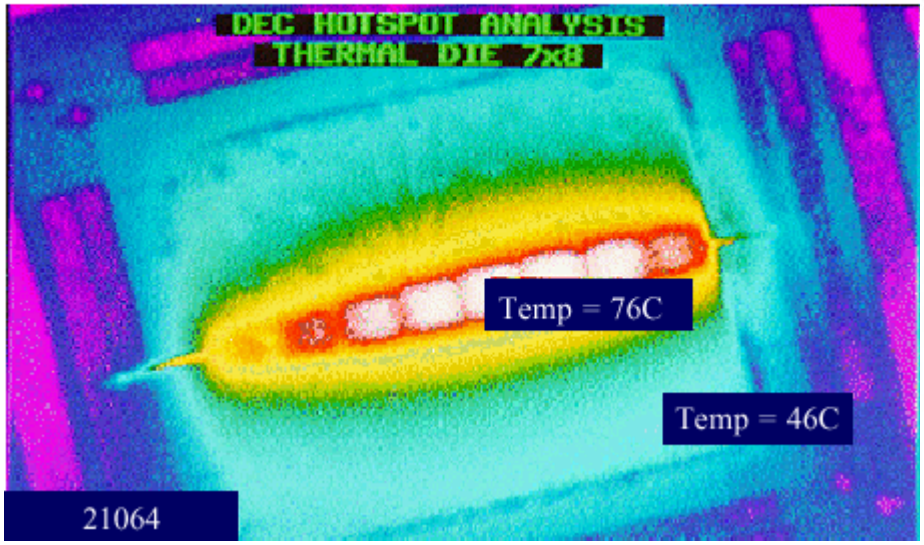


The circuit simulation parameters including threshold voltage were obtained from the Berkeley Predictive Spice Models. The leakage power numbers were obtained by HSPICE simulations.

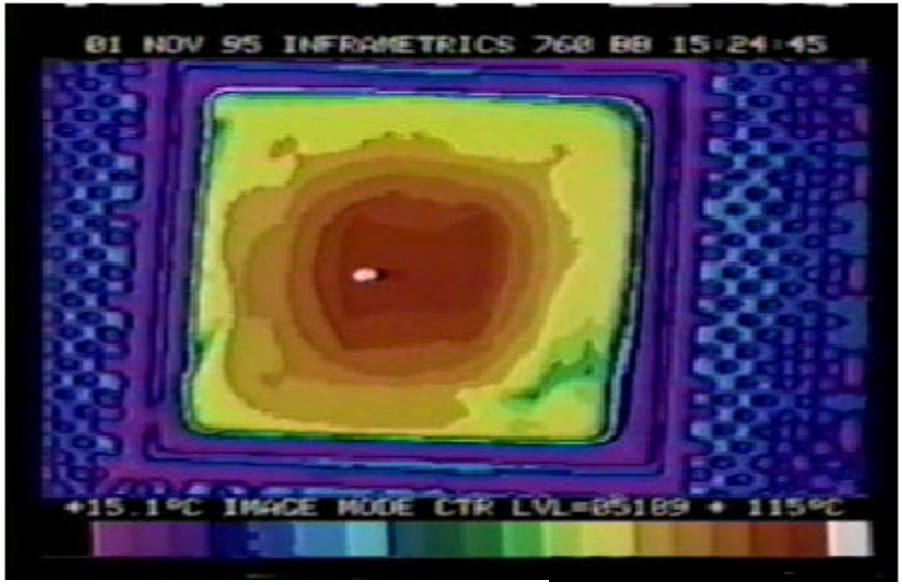
Flautner et al. "Drowsy Caches: Simple Techniques for Reducing Leakage Power", ISCA 2002

Alpha hot spots

21064 Thermal Plot



21164 Thermal Plot



	Power (Watts)	Freq. (MHz.)	Die Size (mm^2)	Vdd
Alpha 21064	30	200	234	3.3
Alpha 21164	50	300	299	3.3

Area 30%

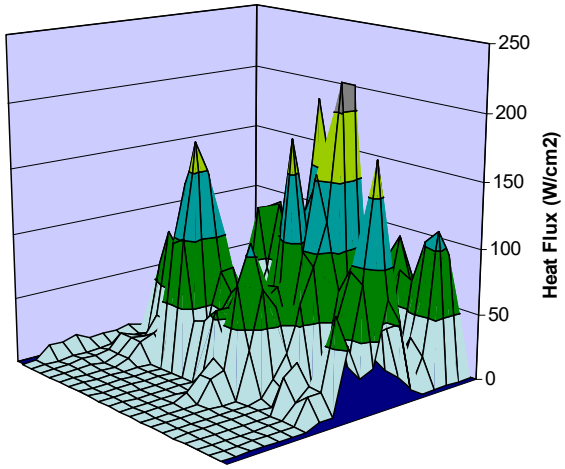
Freq. 50%

Power 67%

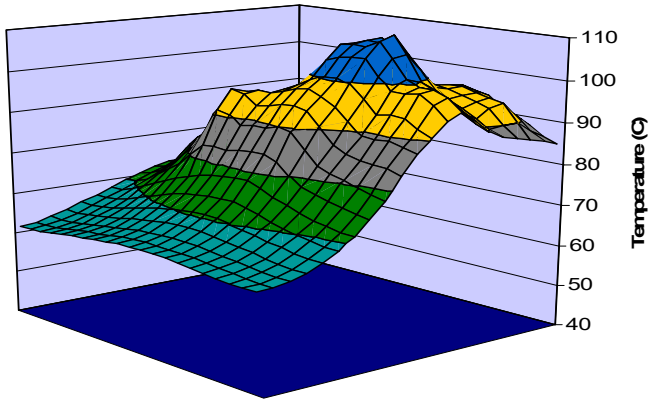
Source - CoolChips-99

Hot Spots and thermal problems

Power Map Pentium IV



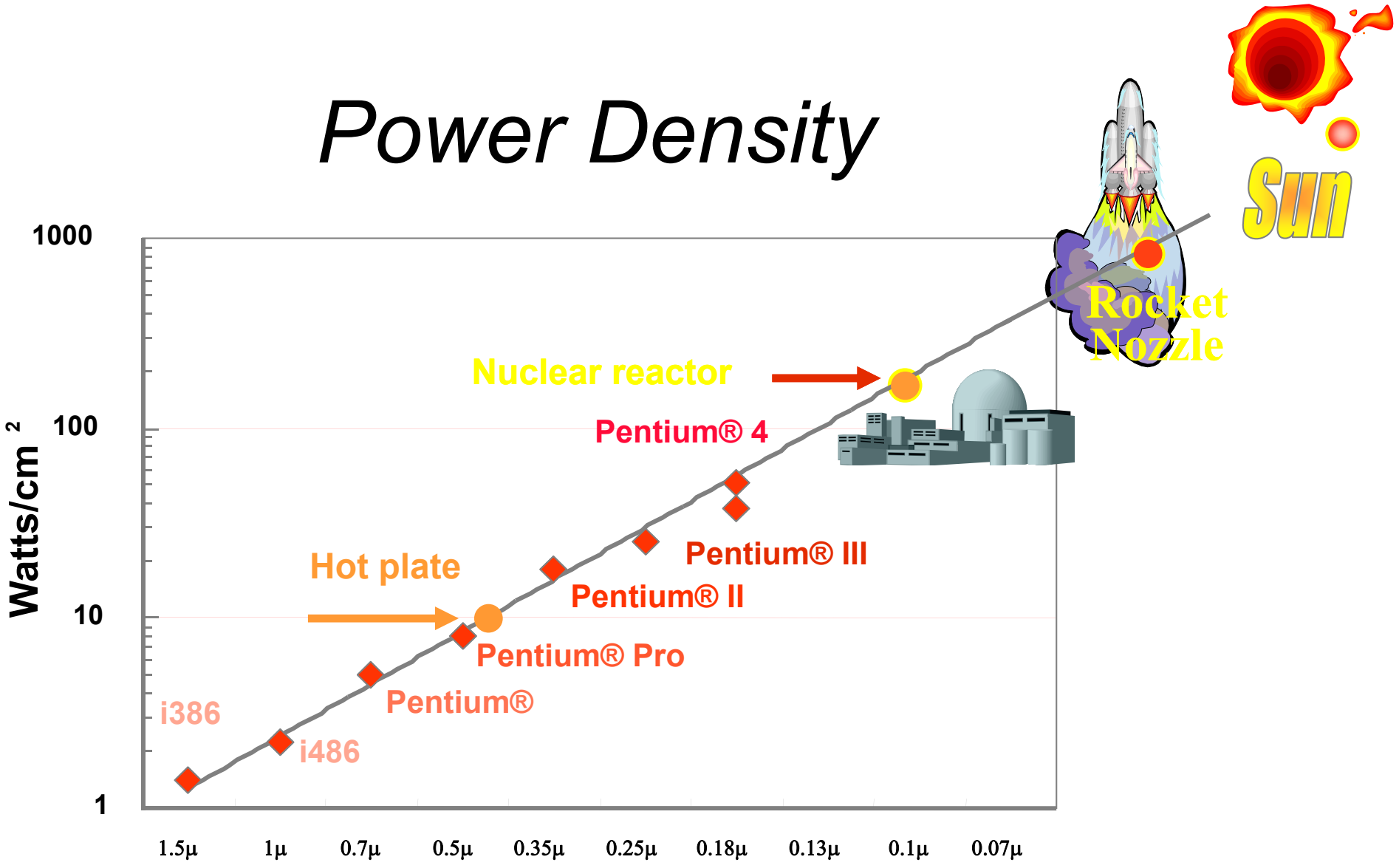
On-Die Temperature Pentium IV



- Silicon is not a good heat conductor
- Temperature ↑ ➔ *leakage* ↑ ➔ power ↑➔ temperature ↑
- A big circuit does not help. *Hot* spots must be reduced.
- With a good layout, hotspots can be put apart and thus reducing the heat dissipation needs –*power envelope*.

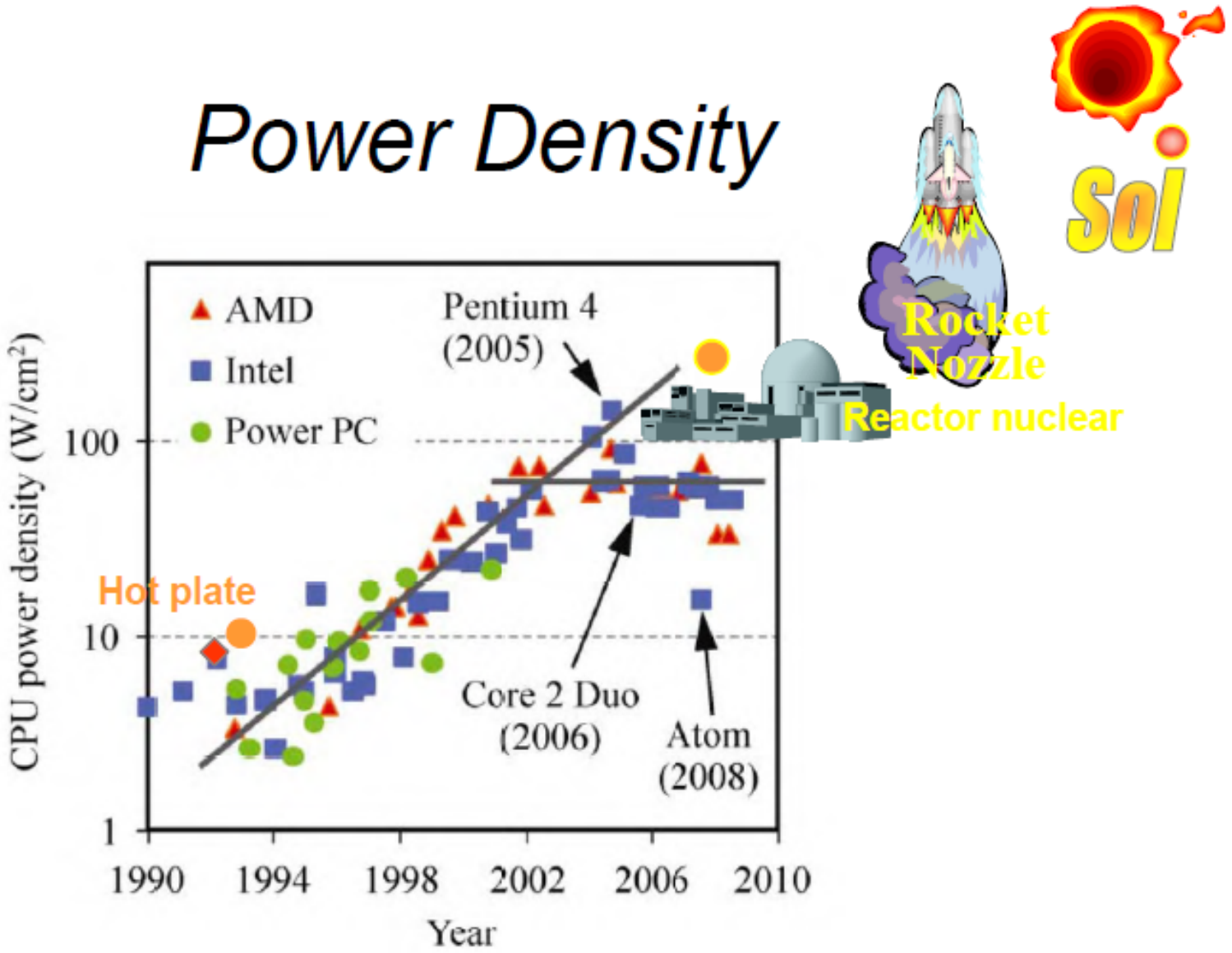
* "New Microarchitecture Challenges in the Coming Generations of CMOS Process Technologies" – Fred Pollack, Intel Corp. Micro32 conference key note - 1999.

Power Density



Power density is too high to keep the contacts cool enough

Power Density



Power and Energy

Power

- *"amount of work done per unit of time"*
  - Measured in Watts
- Important for:
  - Higher power ➔ Higher currents (I)
    - Processors are limited (power delivery constraints)
  - Higher power ➔ Higher temperature
    - Processors are limited again (power envelop)

Power and Energy

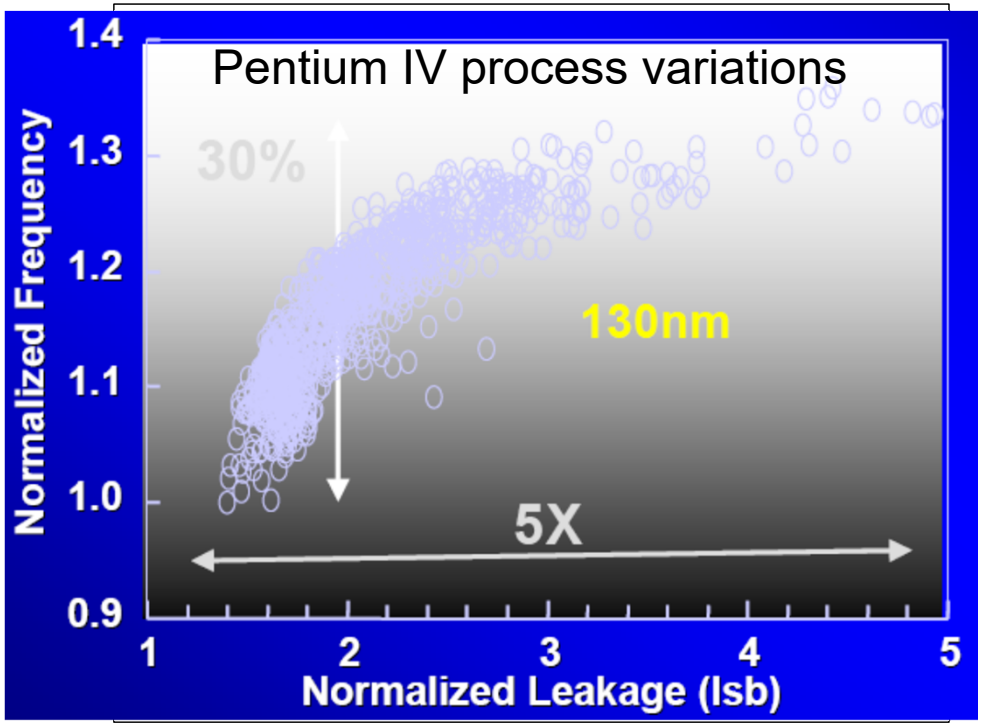
- CMOS
  - $P_{dynamic} = \alpha C V^2 f$
    (α: activity factor, C: capacitance, V: voltatge, f: frequency)
  - $P_{static} = nIV$
    (n: number of transistors *off*, I: leakage current, V: voltatge)

Power Envelop / Thermal Design Power

- **CPU "power envelop"**
  - Maximum power the cooling mechanisms can dissipate.
- **Limitted by**
  - System power budget
  - Processor power budget
  - Usually it is around:
    Server <140W, Desktop 50-100W, Laptop 20-35W, PDA <10W
- **Larger systems can afford large cooling mechanisms such as**
  - *Heat sinks*
  - *Heat pipes*
  - Better TIM (*Thermal Interface Materials*)
- **Power distribution is important:**
  - The most uniform the distribution is the better the dissipation capabilities

Voltage Scaling

- **Given a voltage range, the higher the voltage the higher the frequency the circuit can opperate**
- **Good approach to trade-off power and frequency.**
  - Statically, when manufacturing the circuit
  - Dynamically, while the circuit is opperating
    - Intel's SpeedStep® Technology
    - Transmeta LongRun
    - AMD PowerNow, Cool'n quiet
- **The voltage/frequency range depends on the technology used.**



Voltage Scaling (cont.)

- **Impact on power consumption:**

  Frequency 20% ↓➔ ↓ 20% voltage

  ➔ 35% energy reduction ($\alpha CV^2 = \alpha C(0.8V)^2 = \alpha CV^2 \cdot 0.64$

  ➔ 50% power reduction ($\alpha CV^2f = \alpha CV^2f0.8^3 = \alpha CV^2f \cdot 0.51$)

- **Minimal impact on performance:**

  Frequency 20% ↓ ➔ performance 10%-15% ↓*

- ***Voltage scaling* leverages energy consumption and performance**

* Depending on the ratio between the core and the external busses' frequency and the cache size.

Leveraging power and performance

***Voltage scaling* may be not enough to reduce power**

We need better:

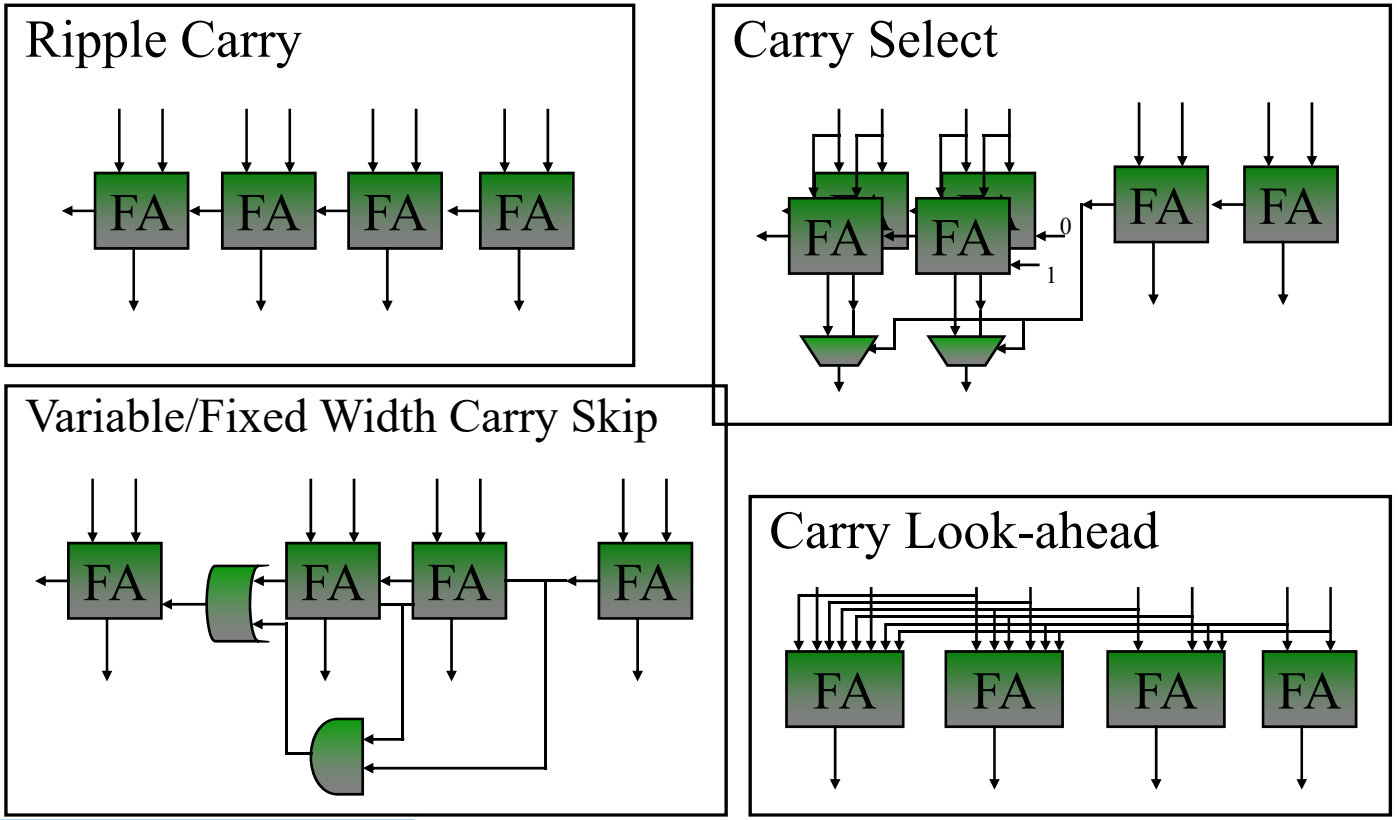
1. **Designs (Computer Architecture)**
    - **Power-Aware designs**
        - **Energy*Delay (EDP) metrics**
2. **Technology (physics)**

Power - Performance Metrics

- Power α C V^2 f
- Metric: suppose we introduce a technique in the design that it has a cost in energy consumption and an effect on performance:
  1. Power/Perf (➔ Energy), assuming the same technology and voltage.
     - Good for:
       - Battery life, electricity bill.
       - Savings in a fixed *power envelope* – without voltage scaling.
  2. Power/$Perf^2$ (➔ Energy*Delay)
     - Balance between performance and energy consumption.
  3. Power/$Perf^3$ (➔ Energy*$Delay^2$)
     - Independent of voltage scaling
     - Good to evaluate architectural techniques that can go over the benefits of voltage scaling

Example: Adder

- There exist several adder designs:
  - Ripple, select, skip (x2), Look-ahead, conditional-sum.
  - Each on has a different trade-off in delay and power requirements



Power and Performance Figures

- Callaway i Swartzlander*:

	Energy (pJ)	Delay (nSec)
Ripple Carry	117	54.27
Constant Width Carry Skip	109	28.38
Variable Width Carry Skip	126	21.84
Carry Lookahead	171	17.13
Carry Select	216	19.56
Conditional Sum	304	20.05

  - Better alternative:
    - Best power – "*constant width carry skip*"
    - Best delay – "*carry look-ahead*"

* "Estimating the power consumption of CMOS adders" - Callaway, T.K.; Swartzlander, E.E., Jr. 11th Symposium on Computer Arithmetic, 1993. Proceedings.

Conclusions

- VLSI Basics
  - Resistance & Capacity are the basis.
- Energy Consumption
  - Important design factor (at the moment more important than area or delay)
- Voltage Scaling: good but not enough
- New metrics for power-aware designs

Reminders

- 1st assignment (*Historical Perspective*) due in 1 week (February 24th)
- 2nd/3rd assignments (*Logic Gates…*) due in 2/3 weeks (March 3rd, 10th)
  - Lectures next Tuesday
- 1st Lab session Wed. March 10th (3 weeks from now).