Topics

- Power Dissipation
- Technology Scaling

Where Does Power Go in CMOS?

- Dynamic Power Consumption
  - Charging and discharging capacitors
  - Short circuit currents: short circuit path between supply rails during switching
- Static Power Consumption
  - Leakage: leaking diodes and transistors

Power dissipation

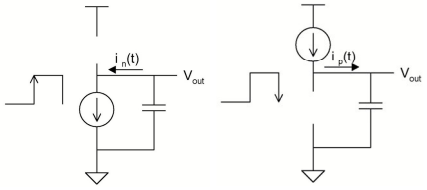
- Static power dissipation
  - In theory, CMOS has no static power dissipation
  - There is a slight current (sub-threshold leakage current and gate leakage current) on the order of 0.1-0.5nA per device
  - At 5V supply voltage, 0.5-2.5 nW static power dissipation per device
  - Million-gate chip will have 0.5-2.5 mW static power dissipation
  - Billion-gate chip ???

Power dissipation

- Dynamic power dissipation
  - Proportional to load capacitance and frequency
  - Proportional to square of the supply voltage
    - Reducing supply voltages can effectively reduce dynamic power
    - However, reduced supply voltage will increase delays
    - There is a limit on reduction of supply voltage

Power dissipation

- Dynamic power dissipation
  - Switching causes charging/discharging current flow which will cause power dissipation



Power dissipation

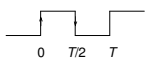
- Dynamic power dissipation

$$P = \frac{1}{T}\left[\int_0^{\frac{T}{2}} i_n(t)V_{out}dt + \int_{\frac{T}{2}}^{T} i_p(t)(V_{DD} - V_{out})dt\right]$$

$$= \frac{1}{T}\left[-C_L\int_{V_{DD}}^{0} V_{out}dV_{out} + C_L\int_0^{V_{DD}} (V_{DD} - V_{out})dV_{out}\right]$$

$$= \frac{C_L}{T}\left[\frac{V_{DD}^2}{2} + \frac{V_{DD}^2}{2}\right]$$

$$= C_L V_{DD}^2 f$$

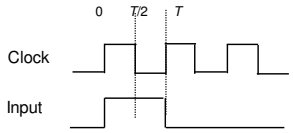


Power dissipation

- Example

$$V_{DD} = 5V$$
$$f = 1GHz$$
$$C_L = 1pF$$
$$P = C_L V_{DD}^2 f$$
$$= 1pf \cdot 5^2 \cdot 1GHz$$
$$= 25mW$$

Power dissipation



$$P = \alpha C_L V_{DD}^2 f$$

- α is the activity factor - i.e. the probability that a clock event results in a transition

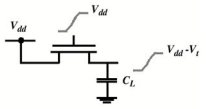
Power dissipation

- Example: NAND gate with independent and uniformly distributed inputs.
  - Probability of NAND gate outputs: P0=0.25 and P1=0.75.
  - Therefore, probability of 0→1 transition is $\alpha_{0\to1}$ = P0P1 =0.1875

$$V_{DD} = 5V$$
$$f = 1GHz$$
$$C_L = 1pF$$

$$P = \alpha C_L V_{DD}^2 f$$
$$= 0.1875 \cdot 1pf \cdot 5^2 \cdot 1GHz$$
$$= 4.7mW$$

Modification for Circuits with Reduced Swing

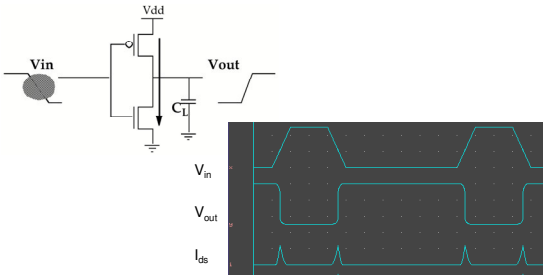


- Can exploit reduced swing to lower power (e.g., reduced bit-line swing in memory)

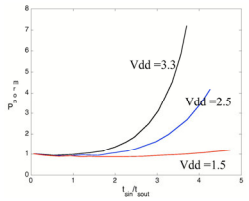
Power dissipation

- Short circuit current dissipation
  - Short circuit current occurs when both transistors are "on" temporarily
  - Proportional to the ratio of rise time
  - Since the rise time is usually much less than period *T*, it can be usually ignored

Short Circuit Currents

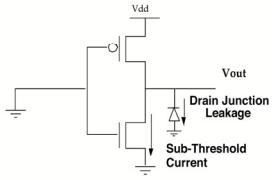


Minimizing Short-Circuit Power



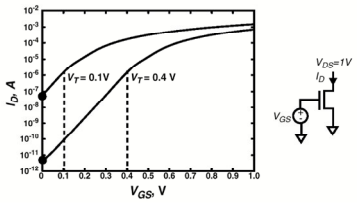
- Keep the input and output rise/fall times the same
- If $V_{dd} < V_{tn} + |V_{tp}|$ then short-circuit power can be *eliminated*!

Leakage



- Sub-threshold current is one of most difficult issues in low-power circuit design!

Sub-threshold Leakage Component



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

Principles for Power Reduction

- The first choice: Reduce voltage!
  - Recent years have seen an acceleration in supply voltage reduction
  - Design at very low voltages still open question
- Reduce switching activity
- Reduce physical capacitance

Goals of Technology Scaling

- Make things cheaper:
  - Want to sell more functions (transistors) per chip for the same money
  - Build same products cheaper, sell the same part for less money
  - Price of a transistor has to be reduced
  - But also want to be faster, smaller, lower power

Technology Scaling

- Goals of scaling the dimensions by 30%:
  - Reduce gate delay by 30% (increase operating frequency by 43%)
  - Double transistor density
  - Reduce energy per transition by 30%
- Die size used to increase by 14% per generation
- Technology generation spans 2-3 years

Technology Scaling Models

- Full Scaling (Constant Electrical Field)
  - ideal model — dimensions and voltage scale together by the same factor *S*
- Fixed Voltage Scaling
  - most common model until recently — only dimensions scale, voltages remain constant
- General Scaling
  - most realistic for today's situation — voltages and dimensions scale with different factors

Scaling Relationships for Long Channel Devices

Parameter	Relation	Full Scaling	General Scaling	Fixed Voltage Scaling
W, L, t_{ox}		1/S	1/S	1/S
V_{DD}, V_T		1/S	1/U	1
N_{SUB}	V/W_{depl}^2	S	S^3/U	S^2
Area/Device	WL	$1/S^2$	$1/S^2$	$1/S^2$
C_{ox}	$1/t_{ox}$	S	S	S
C_L	$C_{ox}WL$	1/S	1/S	1/S
k_n, k_p	$C_{ox}W/L$	S	S	S
I_{av}	$k_{n,p}V^2$	1/S	S/U^2	S
t_p (intrinsic)	C_LV / I_{av}	1/S	U/S^2	$1/S^2$
P_{av}	C_LV^2/t_p	$1/S^2$	S/U^3	S
PDP	C_LV^2	$1/S^3$	$1/SU^2$	1/S

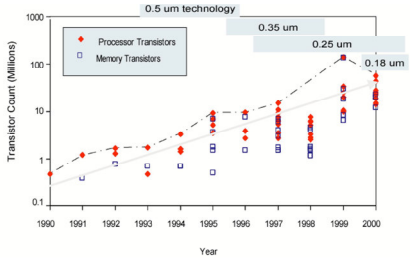
Technology Evolution (2000 data)

- International Technology Roadmap for Semiconductors (ITRS)

Year of Introduction	1999	2000	2001	2004	2008	2011	2014
Technology node [nm]	180		130	90	60	40	*30*
Supply [V]	1.5-1.8	1.5-1.8	1.2-1.5	0.9-1.2	0.6-0.9	0.5-0.6	*0.3-0.6*
Wiring levels	6-7	6-7	7	8	9	9-10	*10*
Max frequency [GHz],Local-Global	1.2	1.6-1.4	2.1-1.6	3.5-2	7.1-2.5	11-3	*14.9 -3.6*
Max µP power [W]	90	106	130	160	171	177	*186*
Bat. power [W]	1.4	1.7	2.0	2.4	2.1	2.3	*2.5*

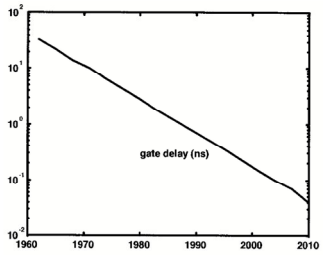
Node years: 2007/65nm, 2010/45nm, 2013/32nm, 2016/23nm

Technology Scaling (1)



Number of components per chip

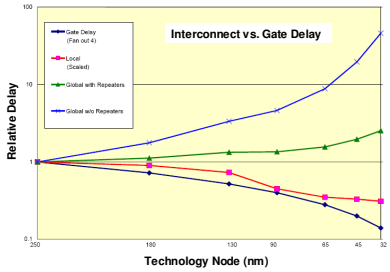
Technology Scaling (2)



t_p decreases by 13%/year
50% every 5 years!

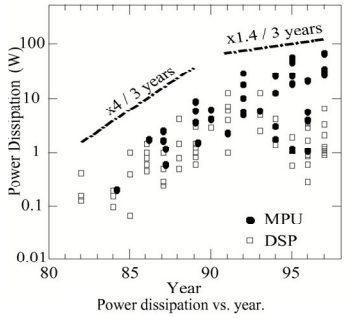
Propagation Delay

Technology Scaling (3)



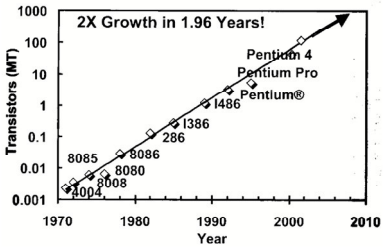
ITRS2005

Technology Scaling (4)



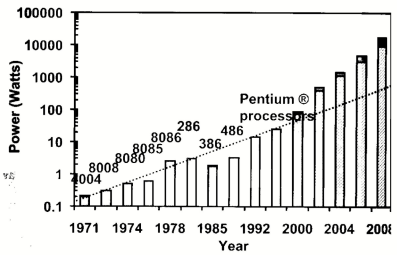
Power dissipation vs. year.

µProcessor Scaling



P.Gelsinger: µProcessors for the New Millennium, ISSCC 2001

µProcessor Power



P.Gelsinger: µProcessors for the New Millennium, ISSCC 2001

Processor Technology Challenges

- Power dissipation
  - Pentium 4W/cm^2
  - Hot plate 10W/cm^2
  - Pentium II 20W/cm^2,
  - Pentium III 50W/cm^2
  - Nuclear reactor 200W/cm^2
    (Source: S. Borkar, IEEE Micro 1999)
- Design variations
  - PTV variations (process parameters, temperature and voltages including threshold voltages)
- Wire delays
  - Clock network, memory bandwidth

2010 Outlook

- Performance 2X/16 months
  - 1 TIP (terra instructions/s)
  - 30 GHz clock
- Size
  - No of transistors: 2 Billion
  - Die: 40*40 mm
- Power
  - 10kW!!
  - Leakage: 1/3 active Power