Introduction to Low-Power Design

Introduction to Low-Power Design

- Motivation
- Sources of Power Dissipation
- Power Reduction Through Voltage Scaling
- Low-Power Design Methodology

Why low-power devices?

- Practical reasons:
  - Reduce power requirements of high-throughput, portable applications
- Financial reasons:
  - Reduce packaging costs
- Technological reasons:
  - Heating prevents the implementation of higher-density chips and limits functionalities
- Environmental reasons:
  - Green computers...

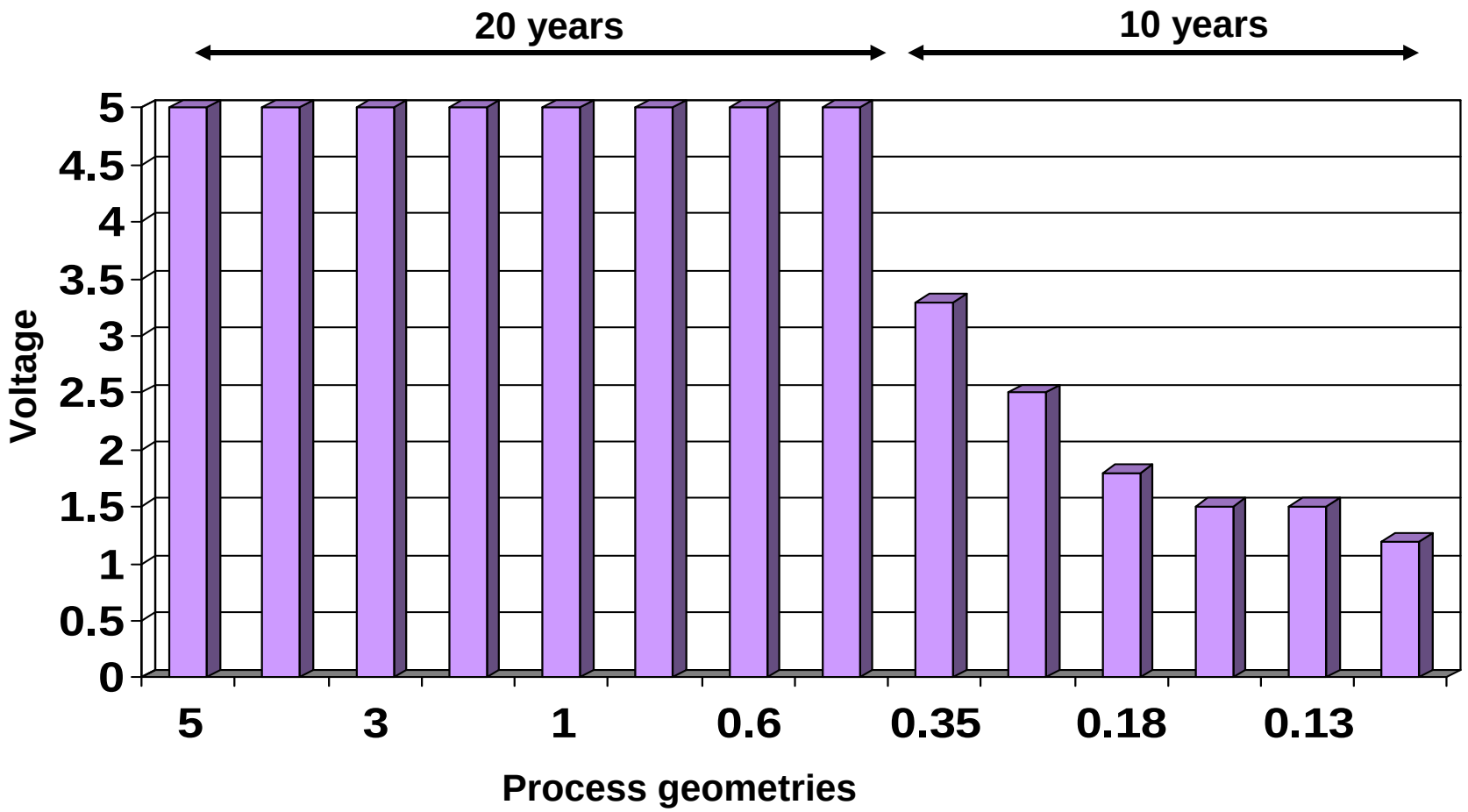
Driving Forces

- Advent of deep sub-micron technologies.
- Increasing market share of mobile applications.
- Limitations of battery technology.

Deep Sub-Micron Technologies

- Smaller geometries:
  - Higher device densities.
  - Higher clock frequencies.
- Consequence:
  - Greater power consumption in spite of lower supply voltages:
    - Technology scales faster than supply voltage.

Voltage Trends



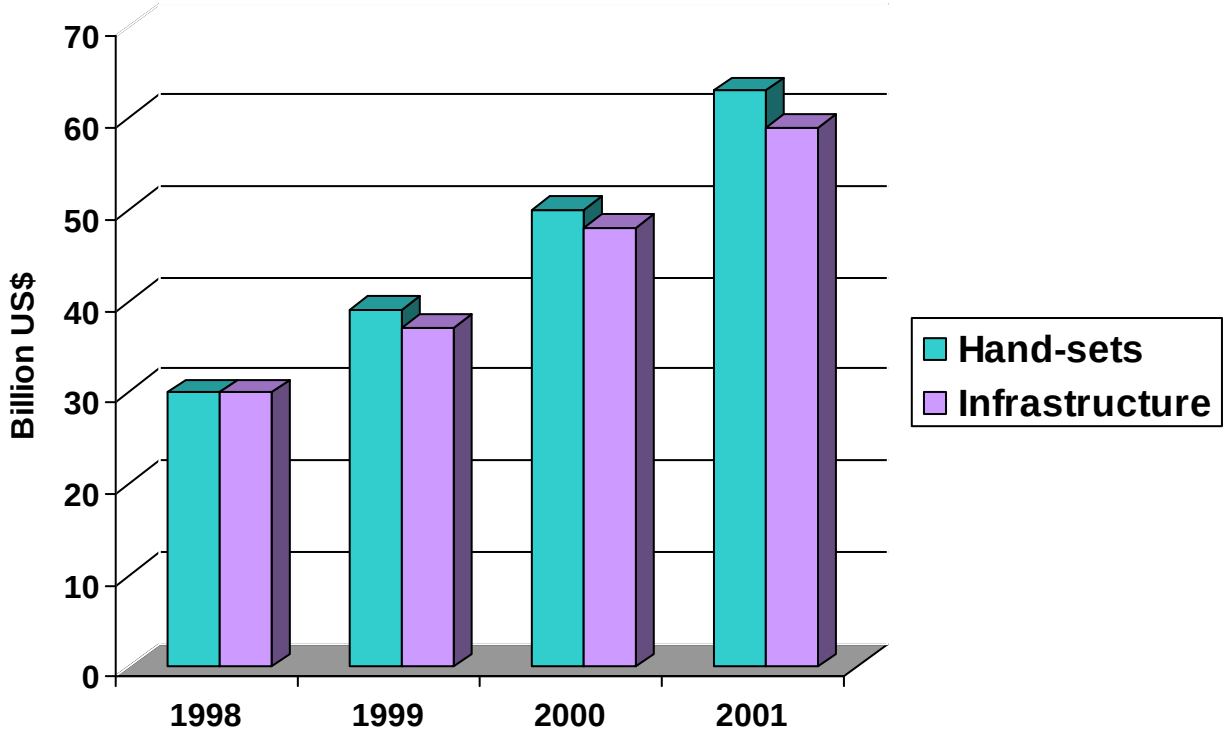
Power Trends

Example: processor

	1992	1994	1998	1999	2001	2005
Technonoly (μ)	0.75	0.5	0.35	0.25	0.18	0.13
Clock speed (MHz)	200	300	667	750	1000	3000
Transistors (millions)	1.68	9.3	15.2	15.2	100	500
Voltage	3.3	3.3	2.3	2.1	1.5	1.3
Power (W)	30	50	72	90	100	~100

Mobile Electronics (I)

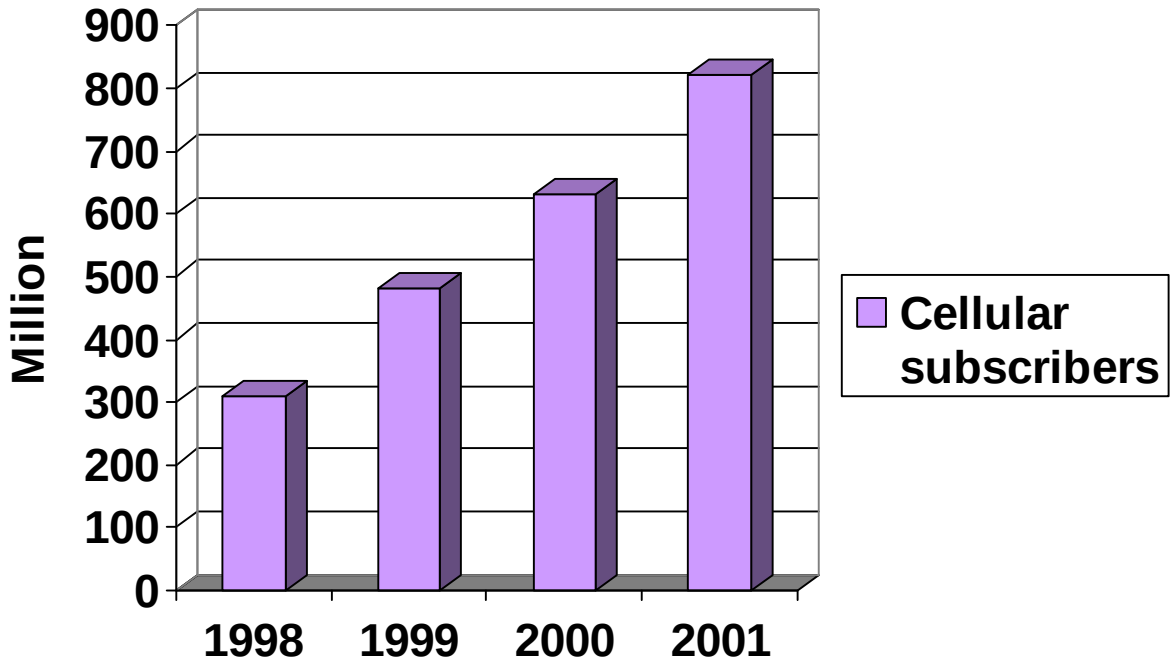
- Wireless communication:
  (appliances and infrastructure)



600 million mobile phones produced in 2001.

Mobile Electronics (II)

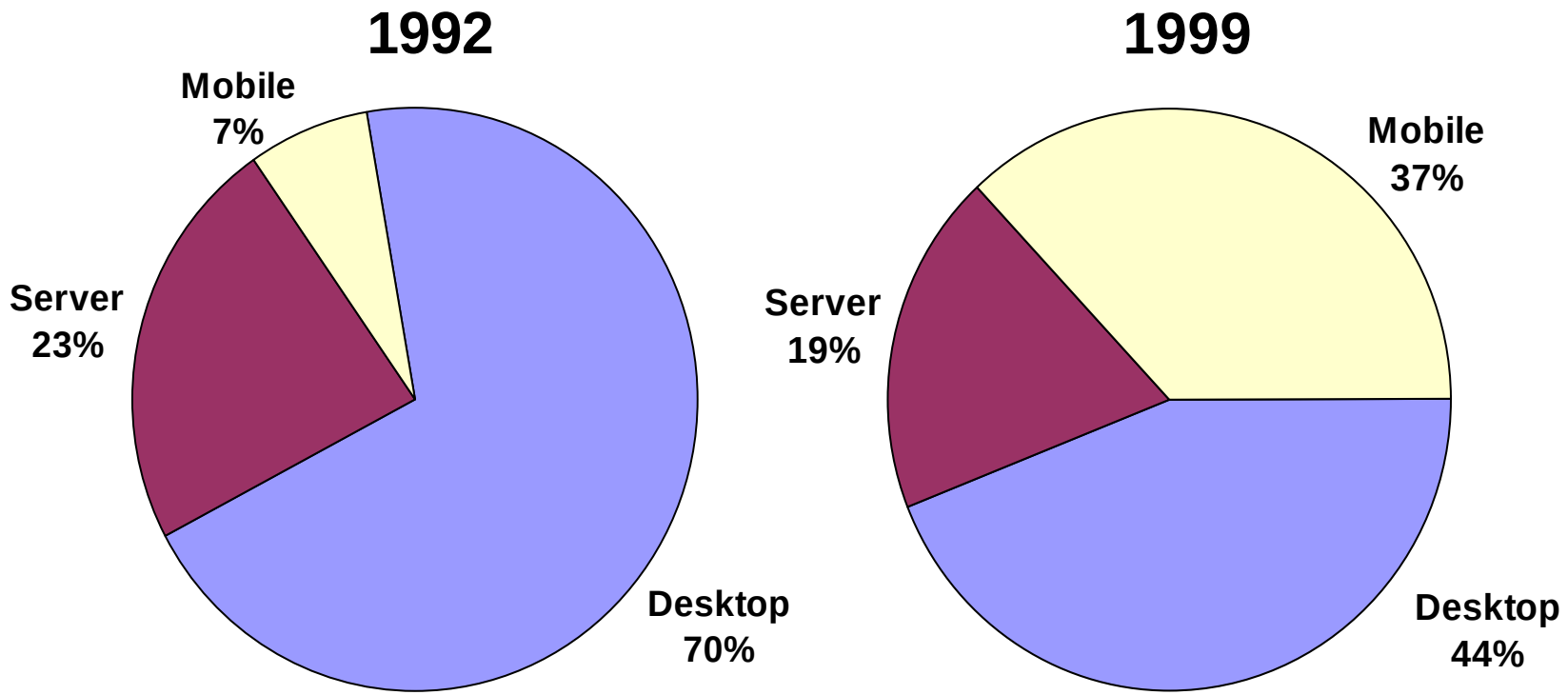
- Cellular network subscribers:



1.9 billion subscribers predicted for year 2004.

Mobile Electronics (III)

- PC market:

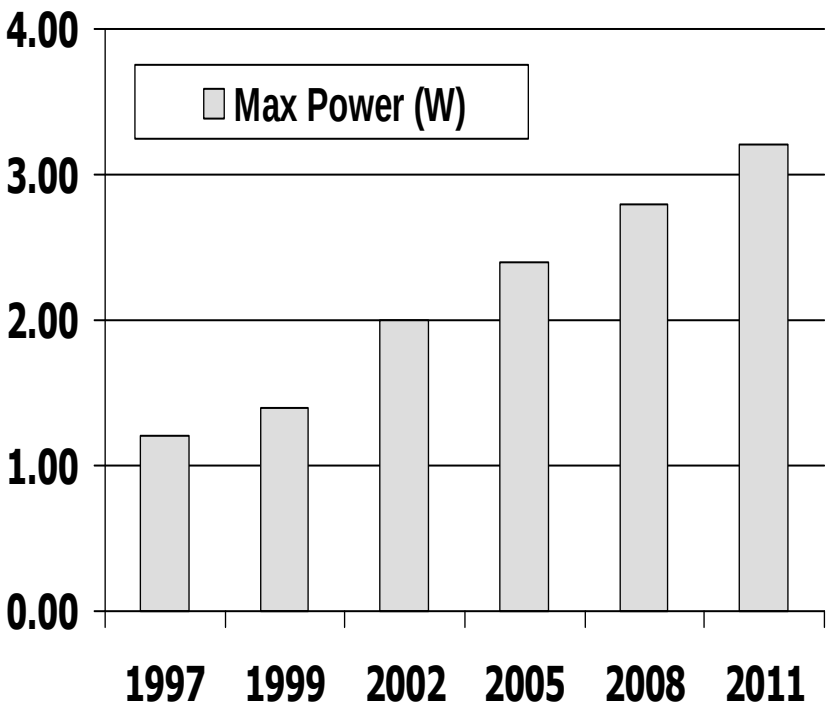


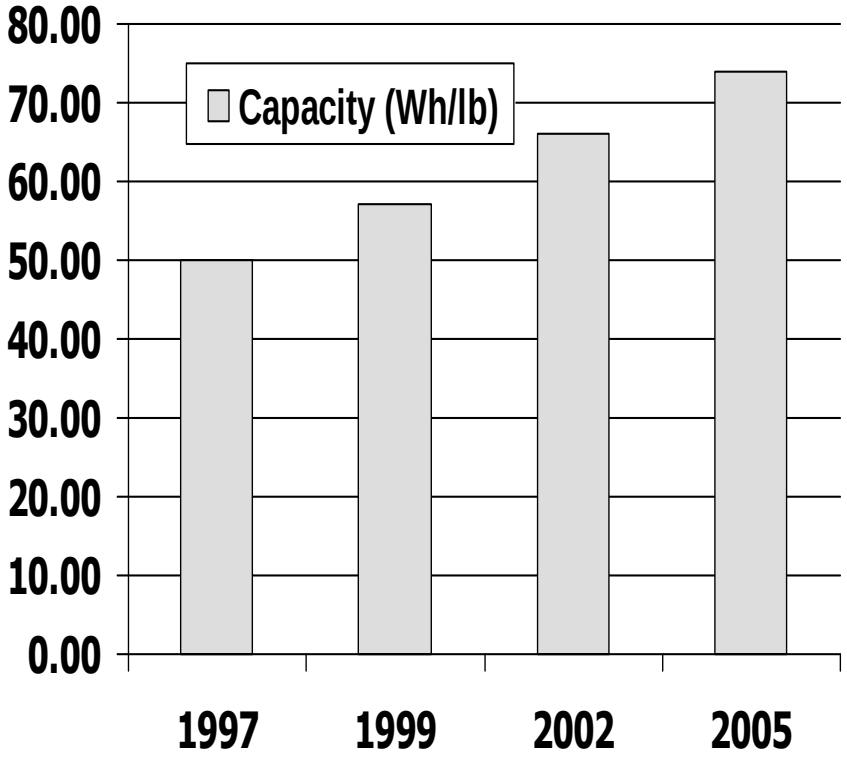
Mobile Electronics (IV)

- Semiconductor market for portable applications:
  [source: R. Chesson, ST, Dec. 2000]
  - 1999: 20%.
  - 2004: 40%.

Low-Power Devices: Battery limitations

- Battery max power and capacity increases 10-15% per year
- Increasing gap w.r.t. power demand





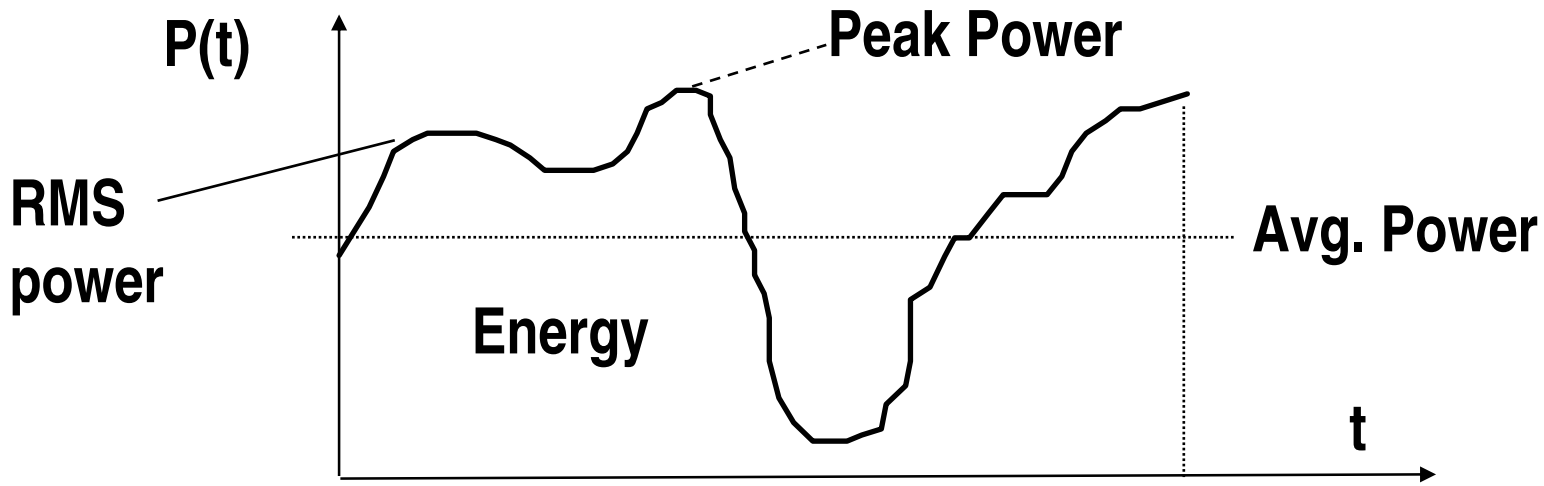
Not Only Mobile

- 20% of electrical energy consumed in Amsterdam is used for telecom.
- In the US, Internet is responsible for 9% of the electrical energy consumed nation-wide.
  - This grows to 13% if all computer applications are considered.
- The transfer of 4 MBytes of data through the net consumes the energy of 1 Kg of coal.

[source: 2000 CO_2 conference, Amsterdam, NL]

Power Metrics

- *Average power*: related to battery life-time
- *Peak power*: related to reliability
- *RMS power*: related to cycle-by-cycle power
- *Energy* = power X time: related to power*time product



Technology issues

CMOS Circuits

- CMOS technology is predominant in the realization of today's ICs.
- CMOS devices are intrinsically low-power consuming.
- CMOS has become the reference technology.

Power Dissipation in CMOS Circuits

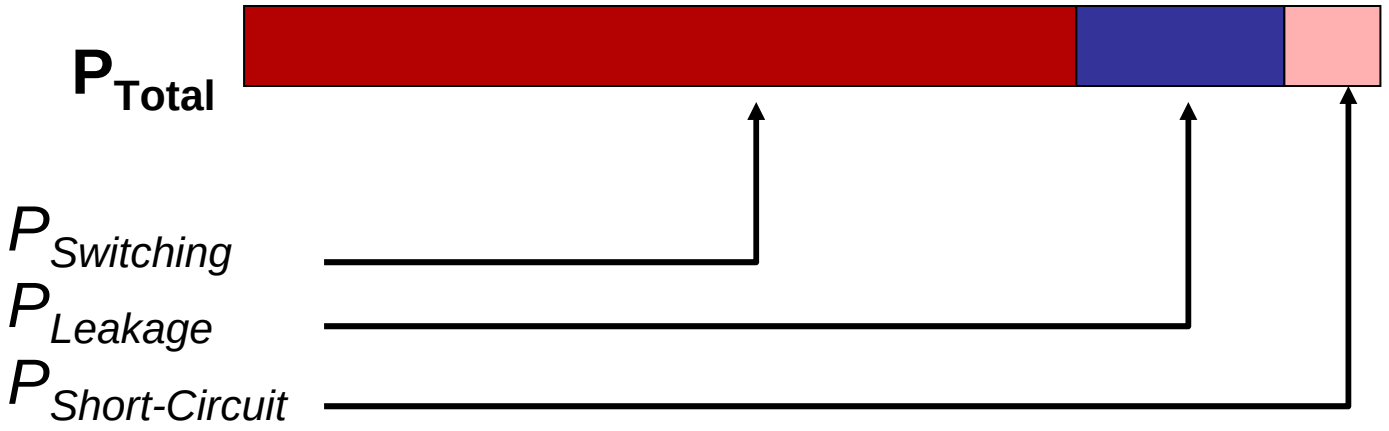
[CSB92]

- There are 3 major sources of power dissipation in a CMOS circuit:

$$P_{Total} = P_{Switching} + P_{Leakage} + P_{Short\text{-}Circuit}$$

- $P_{Switching}$ is due to charging and discharging capacitors.
- $P_{Leakage}$ is due to leaking diodes and transistors.
- $P_{Short\text{-}Circuit}$ is due to direct-paths short circuit currents.

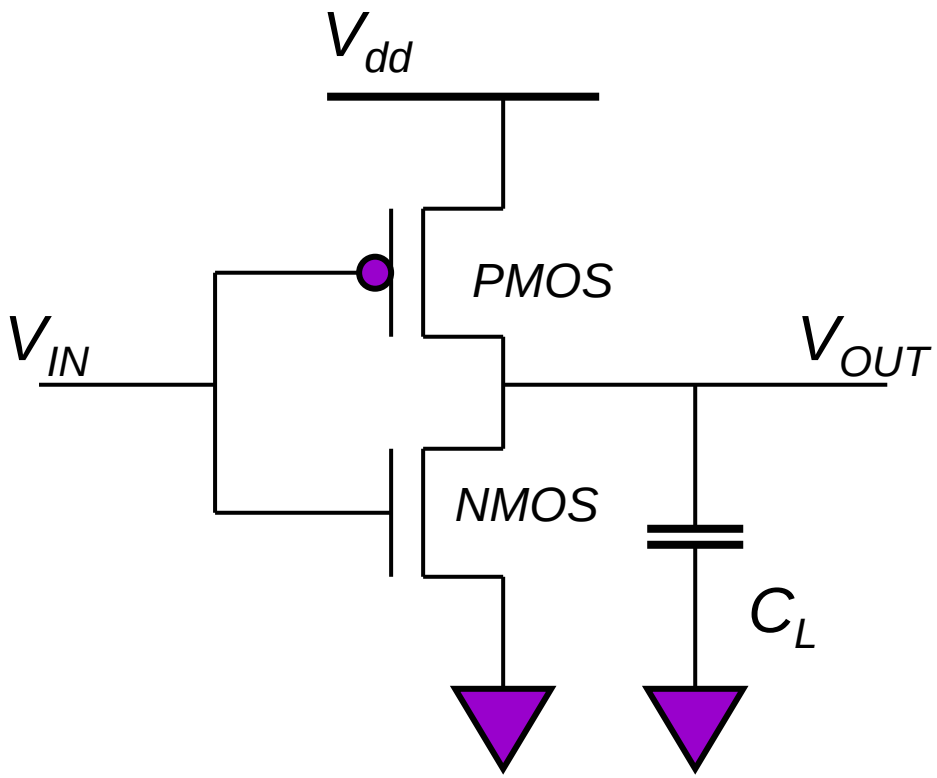
Power Dissipation in CMOS Circuits (Cont.)



- For old technologies (0.7μm and above), $P_{Switching}$ was predominant.
- For new technologies, $P_{leakage}$ becomes critical.
  - Minimization of $P_{Leakage}$ *:* (Mostly) technology problem.
  - Minimization of $P_{Switching}$ *:* Design problem.

Switching Power (1)

- CMOS inverter:



$V_{IN} = 0 \Rightarrow V_{OUT} = 1$

$V_{IN} = 1 \Rightarrow V_{OUT} = 0$

Switching Power (2)

- Transition 0 $\rightarrow$ V_{dd} at the output node:
  Energy dissipated by the PMOS: $1/2\ C_L\ V^2_{dd}$
  Energy stored in C_L: $1/2\ C_L\ V^2_{dd}$.

- Transition $V_{dd} \rightarrow 0$ at the output node:
  Energy dissipated by the NMOS: $1/2\ C_L\ V^2_{dd}$

- If transitions occur at the clock frequency f_{Clock} the average switching power is:

$$P_{Switching} = 1/2\ C_L\ V^2_{dd}\ f_{Clock}$$

Switching Power (3)

- In general, switching does not occur at the clock frequency.
- Let π_{SW} be the probability that the output node makes a transition at each clock cycle. π_{SW} is called the switching activity of the gate.
- The average $P_{Switching}$ for a CMOS gate is given by:

$$\boldsymbol{P_{Switching} = 1/2\ \pi_{SW}\ C_L\ V^2_{dd}\ f_{Clock}}$$

- We define:

$$C_{Eff} = \pi_{SW}\ C_L$$

- To minimize $P_{Switching}$:
  - **Reduce V_{dd}**

Leakage Power

- Due to leakage currents
  - Two types of currents
  - Most significant one due to leakage through the channel of an off transistor: I_{ds}
- Important because related to idle components!!!

Supply voltage reduction

Supply Voltage Reduction

- Model of switching power

$$P_{Switching} = \frac{1}{2} C_{Eff} V_{dd}^2 f_{Clock}$$

- Why not reduce V_{dd} ?
  - Example:
    - *Use a* $V_{dd}' = 0.5\ V_{dd}$
    - $P_{switching}' = 0.25\ P_{switching}$!!

Supply Voltage Reduction

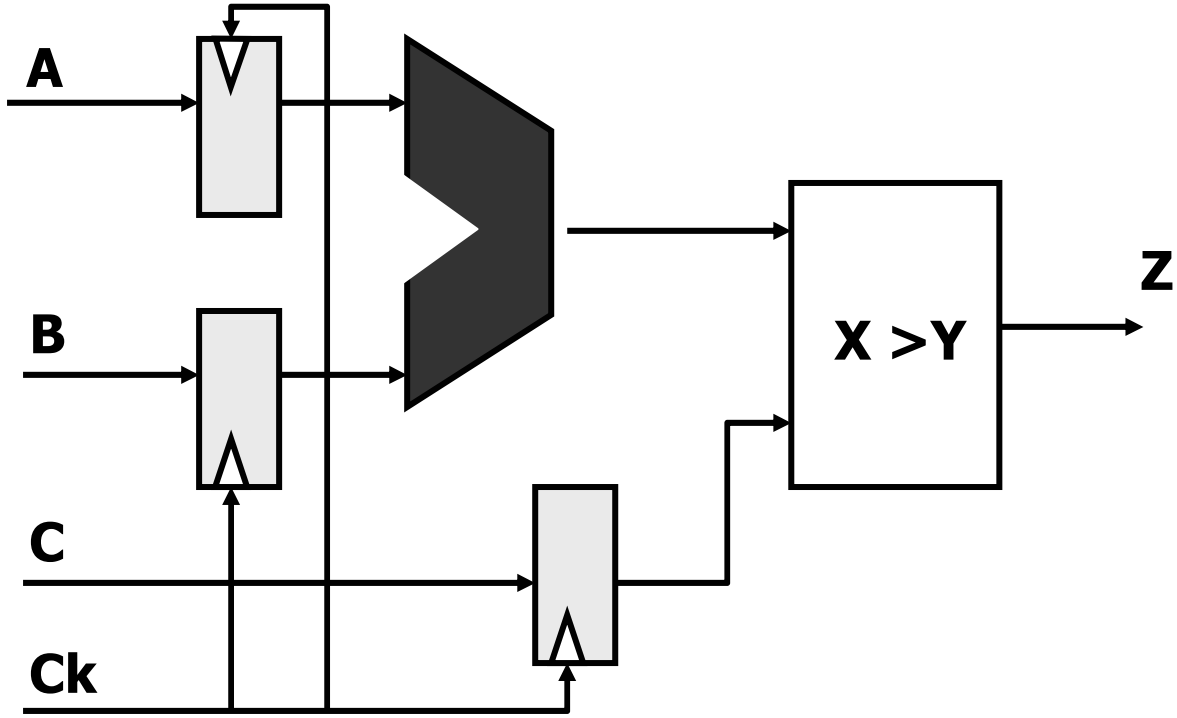
- Drawback:
  - Supply voltage reduction affects circuit speed!
  - Noise à robustness to erros
- Solutions:
  - Technology: improve manifacturing
  - Architecture:
    - Parallelization
    - Pipelining
  - Software:
    - Operating System
    - Applications

Architecture-Driven Supply Voltage Reduction

- Strategy:
  - Modify the architecture of the system so as to make it faster
  - Reduce V_{dd} so as to restore the original speed ➡ Power consumption has decreased
- Typical architectural solutions:
  - **Parallelization**
  - **Pipelining**
- Drawback:
  - Overhead of additional circuitry required to compensate the speed degradation

Example: Reference Architecture

- Adder-Comparator Data-Path
  - Z = (A+B)>C ?



Example: Reference Architecture (2)

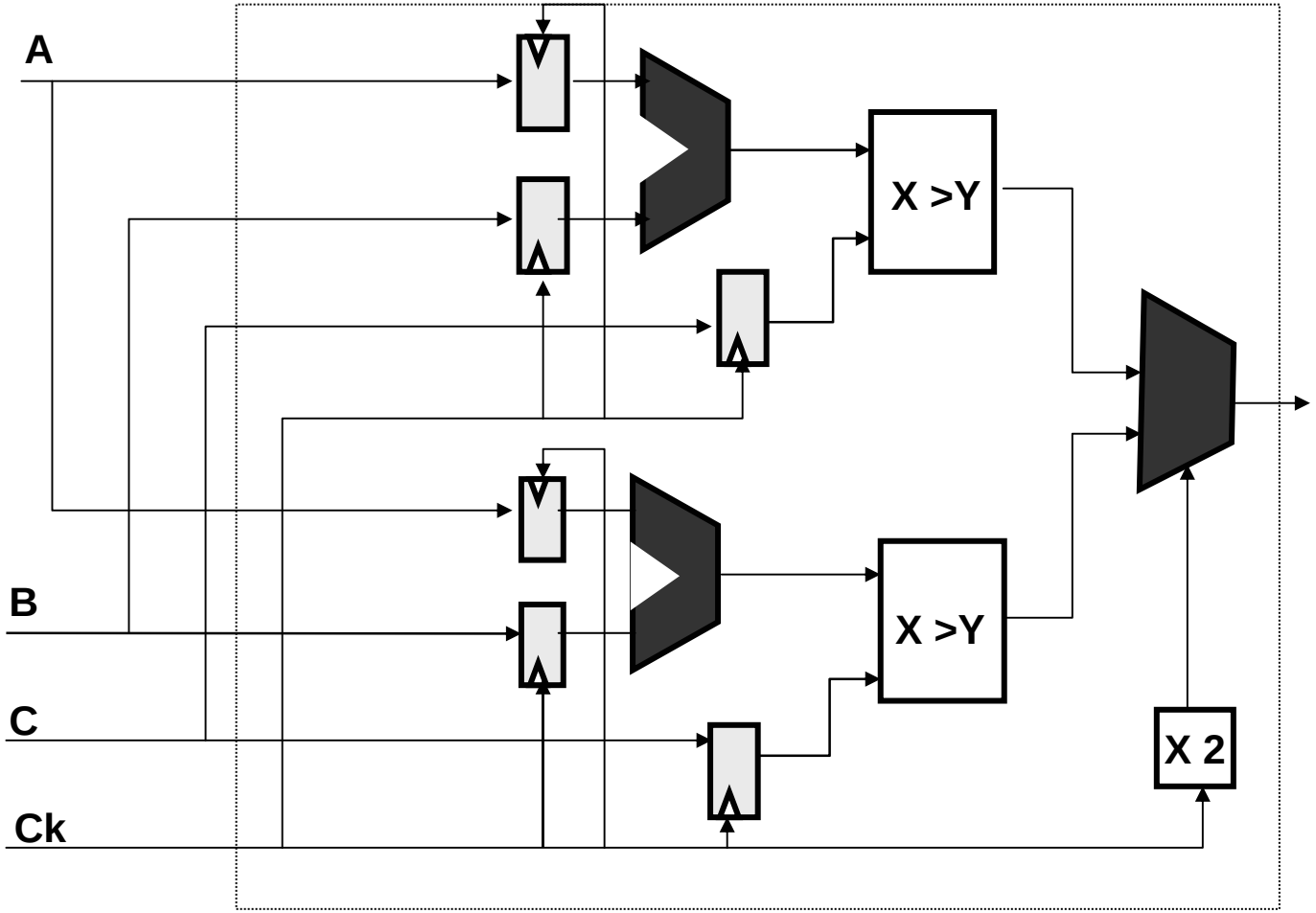
- Supply voltage: $V_{Ref} = 5V.$
- Delay (adder + comparator): *25ns*
  - *12.5ns* in each stage
  - Clock period: $T_{Ref} = 25ns.$
- Total **effective** capacitance: C_{Ref}.

- Power consumption:

$$P_{Ref} = 1/2 C_{Ref} V^2_{Ref} f_{Ref}$$

where $f_{Ref} = 1/T_{Ref}$

Example: Parallel Architecture



Example: Parallel Architecture (2)

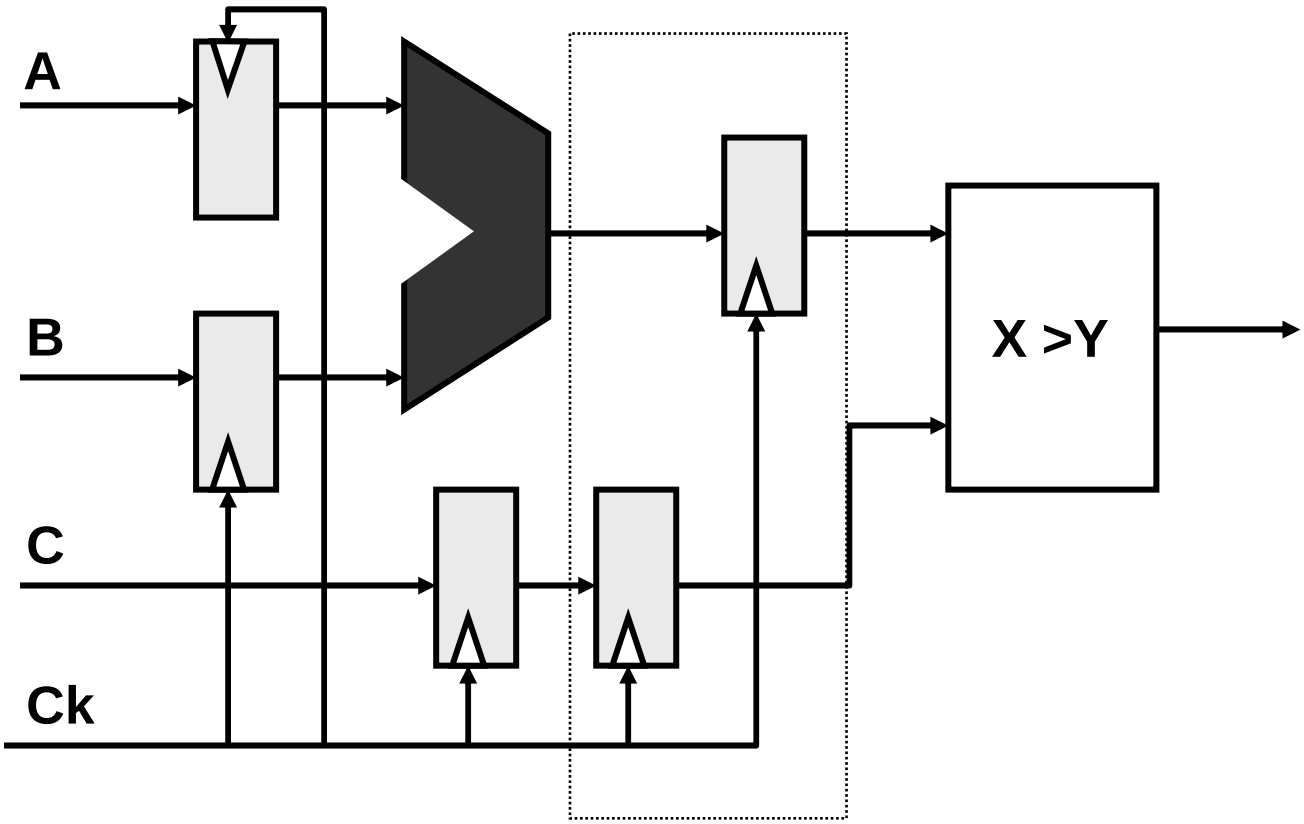
- Supply voltage: $V_{Par} = 2.9V = 0.58\ V_{Ref}$
  - Speed of adder and comparator must be reduced
- The same throughput is still guaranteed if the clock period is doubled: $T_{Par} = 50ns$
- Since $T_{Par} = 2T_{Ref}$ the clock frequency becomes:

$$f_{Par} = f_{Ref} / 2$$

- Total effective capacitance: $C_{Par} = 2.15\ C_{Ref}$ (factor 2.15 instead of 2 is due to overhead)
- Power consumption:

$$P_{Par} = 1/2 C_{Par} V^2_{Par} f_{Par}$$
$$= 1/2\ (2.15\ C_{Ref})(0.58\ V_{Ref})^2 (f_{Ref}/2) \approx \mathbf{0.36 P_{Ref}}$$

Example: Pipelined Architecture



Example: Pipelined Architecture (2)

- Supply voltage: $V_{Pipe} = 2.9V = 0.58\ V_{Ref}$
  - Speed of adder and comparator must be reduced
- The same throughput is still guaranteed even if adder and comparator work at half speed (*25ns* instead of *12.5ns*)
- $T_{Pipe} = T_{Ref} \Rightarrow f_{Pipe} = f_{Ref}$
- Total effective capacitance: $C_{Pipe} = 1.15\ C_{Ref}$ (smaller area than parallel realization).
- Power consumption:

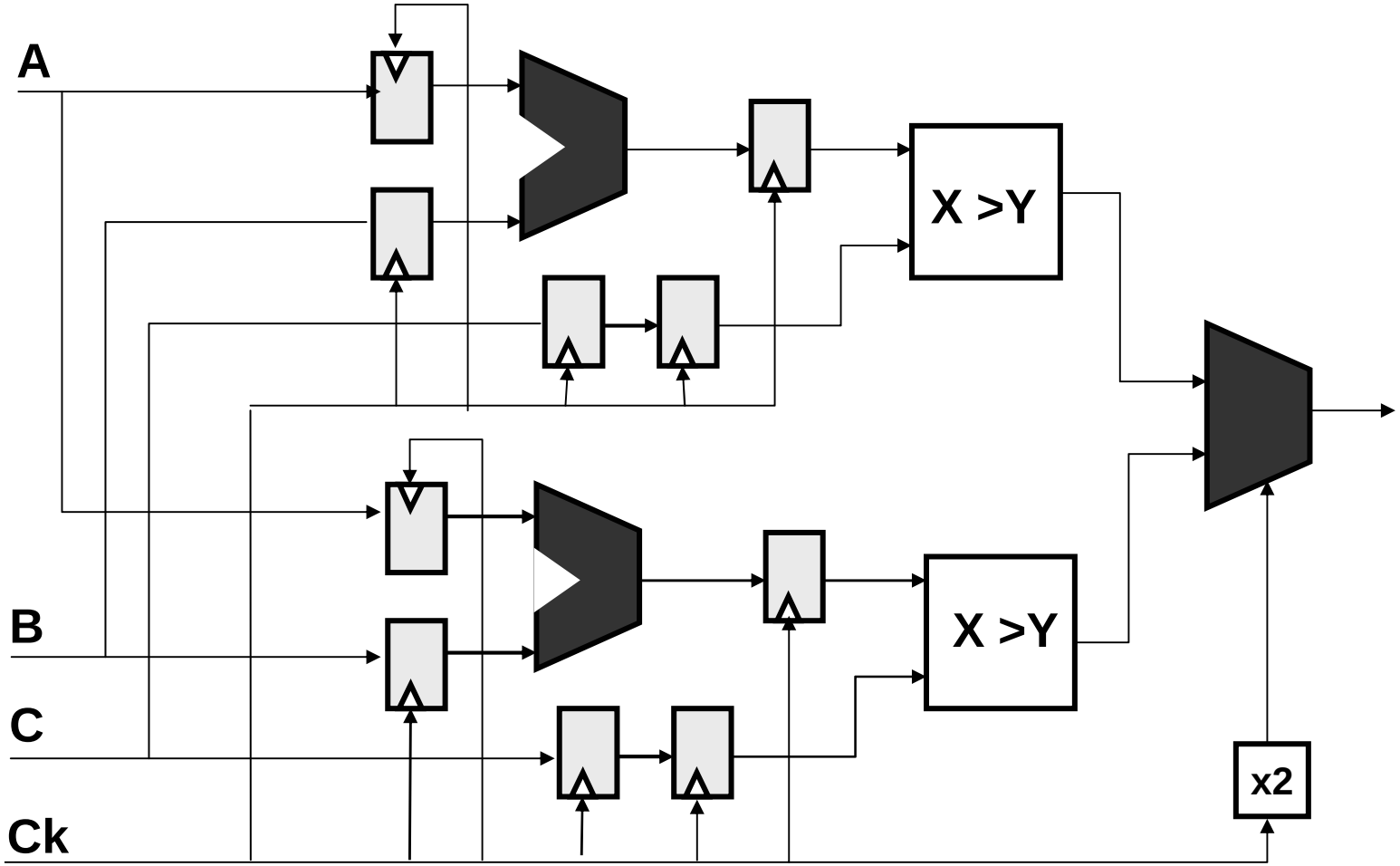
$$P_{Pipe} = 1/2\ C_{Pipe} V^2_{Pipe} f_{Pipe}$$

$$= 1/2\ (1.15\ C_{Ref})(0.58\ V_{Ref})^2 f_{Ref} \approx \mathbf{0.39 P_{Ref}}$$

Example: Pipelined Architecture (3)

- Power consumption reduced by a factor of 2.5, approximately as in the case of parallel realization.
- Area penalty is much smaller than in the parallel case (~50%)

Example: Parallel and Pipelined Architecture



Example: Parallel and Pipelined Architecture (2)

- The critical path is reduced by a factor of 4.
- Speed requirements for adder and comparator can be reduced by the same factor
- Supply voltage: $V_{pp} = 2.0V = 0.4\ V_{Ref}$
- Total effective capacitance: $C_{pp} = 2.5\ C_{Ref}$
- Due to the parallelism in the architecture, the throughput is preserved if the system is clocked at: $T_{pp} = 50ns = 2T_{Ref}$
- The clock frequency is then: $f_{pp} = f_{Ref} / 2$
- Power consumption:

$$P_{pp} = 1/2 C_{pp} V_{pp}^2 f_{pp} = 1/2\ (2.5\ C_{Ref})(0.4\ V_{Ref})^2 (f_{Ref}/2) \approx \mathbf{0.2P_{Ref}}$$

Example: Summary

- Comparison of the four architectures:

Architecture	V_{dd}	C_{Eff}	Area	Power
Reference	5V	1	1	1
Parallel	2.9V	2.15	3.4	0.36
Pipelined	2.9V	1.15	1.3	0.39
Par+pipe	2.0V	2.50	4.0	0.2

Reducing C_{eff}

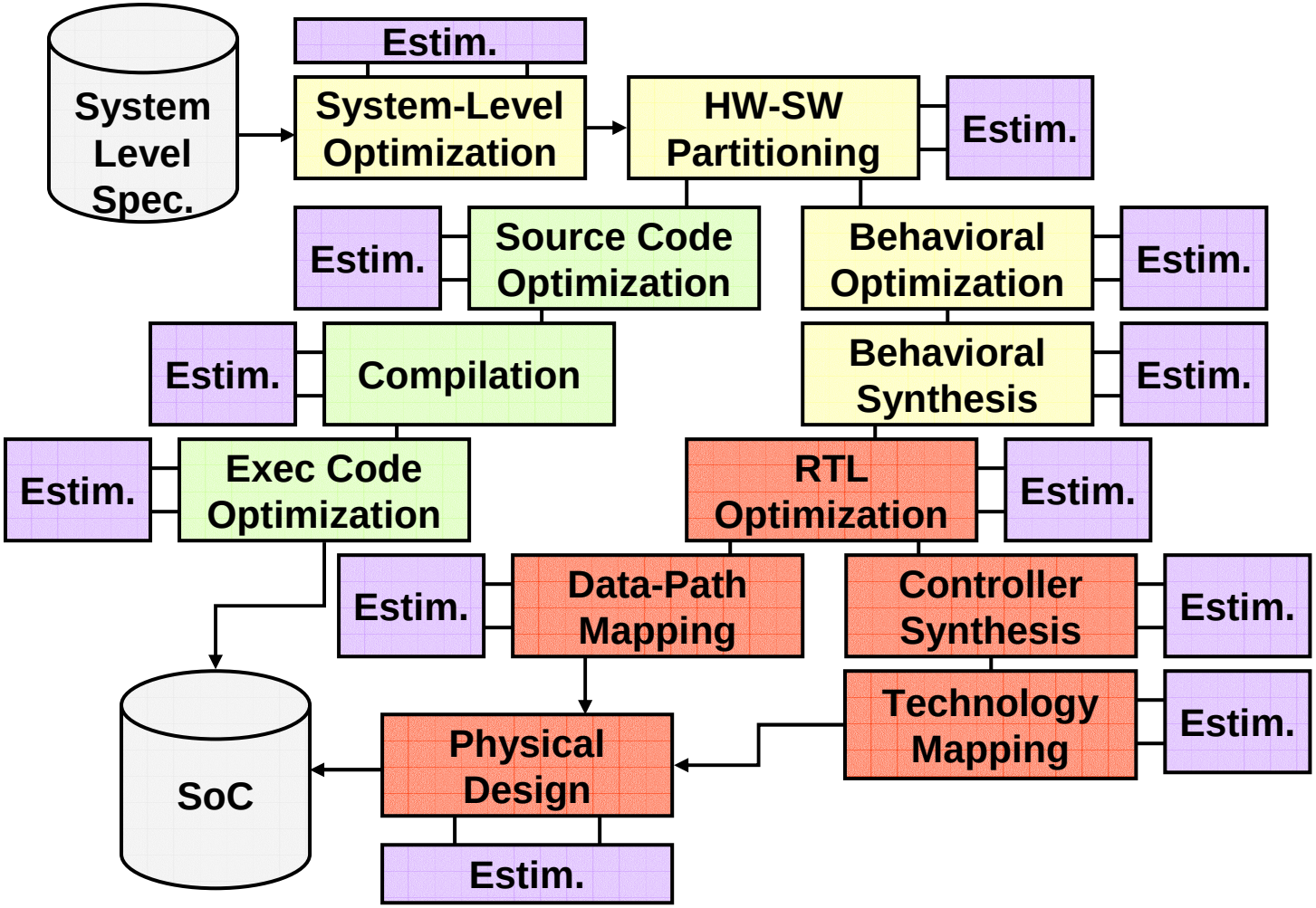
Reducing C_{eff}

- Factors influencing $C_{eff} = \pi_{SW} C_L$
  - Circuit function
  - Input probabilities
  - Circuit topology
  - Circuit technology
- **Design** and **synthesis** techniques have been developed to reduce both C_L and π_{SW} at all stages of the design process.

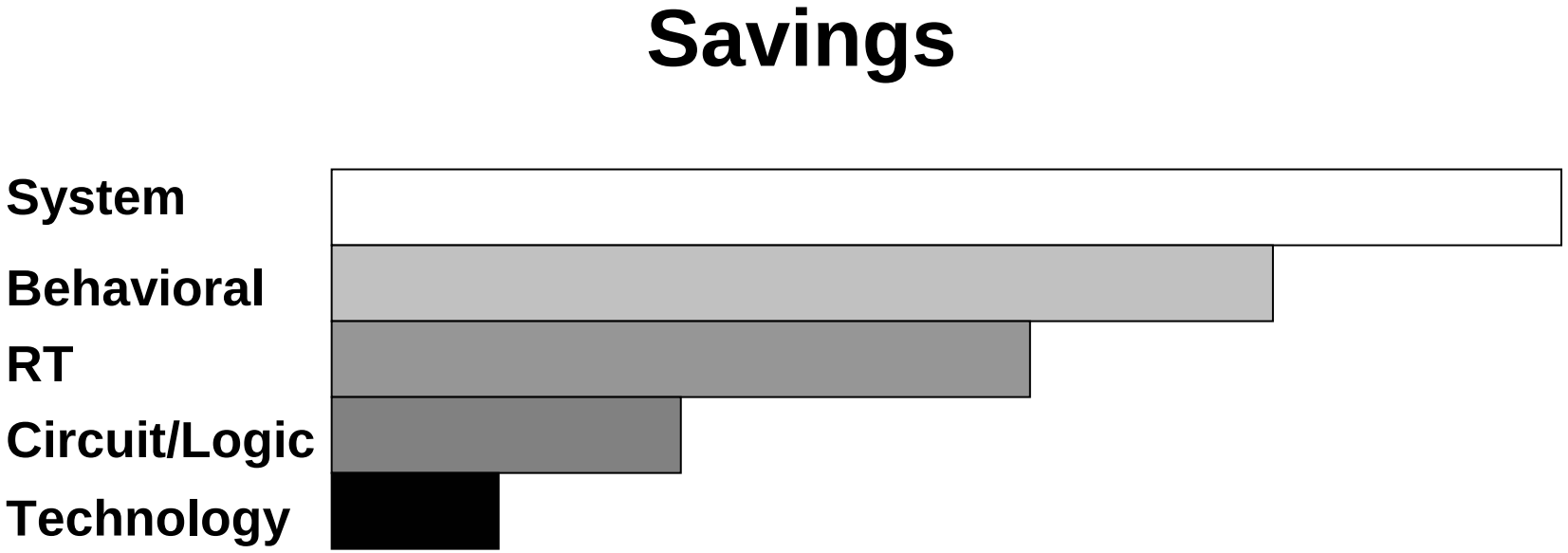
Low-Power Design Flow

- To avoid costly re-design steps, it is mandatory to be able to optimize the power dissipation during the early stages of the design process.
- At each level of abstraction, various design alternatives have to be explored.
- Accurate power estimation tools must provide feedback on the quality of each design choice.
- The availability of a power estimator for each level of abstraction is then fundamental in a low-power design flow.

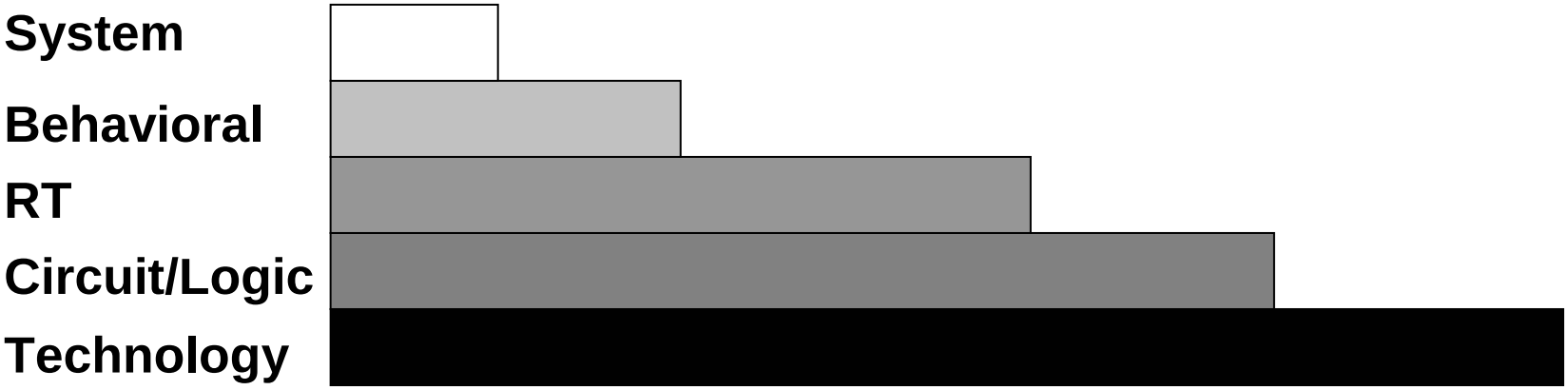
Low-Power Design Flow (Cont.)



Power Optimization



Power Estimation



Others methods to reduce power

- Different clock rates in the chip or on the board
- Suspend idle components of the chip/board
  - In particular I/O: USB, disks, wireless
- Operating System: optimize CPU utilization
  - Process scheduling
  - Reduce clock rate of the CPU during periods of low activity
- Application software

Summary

- The electronics market calls for low-power circuits and systems.
- We have analyzed the the major sources of power dissipation in CMOS circuits
  (CMOS technology is predominant in modern chips).
- Power optimization is required at all levels of the design process.
- We have presented an overview of one of the most common approaches to the design of low-power circuits: Supply voltage reduction.
- We have also introduced a possible flow for designing low-power circuits.