Τμήμα Μηχανικών Η/Υ και Πληροφορικής,
Πανεπιστήμιο Ιωαννίνων

Introduction to low-power microprocessor design - Dynamic power and dynamic power reduction design techniques

Vasileios Tenentes

University of Ioannina

E-mail: tenentes@cse.uoi.gr

ARM® University
Worldwide Education Program

Outline

1: Overview of microprocessors manufacturing process

2: Motivation: Why to care about power efficient microrprocessors?

3: Performance of microprocessors

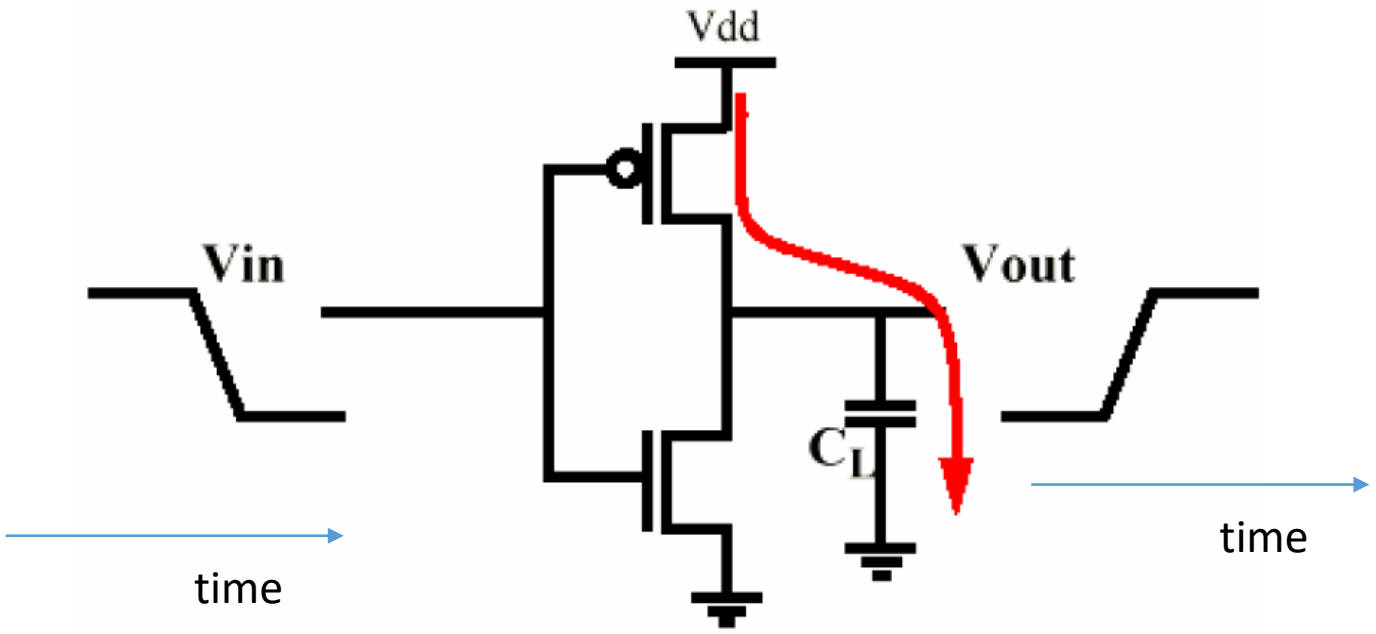
4: Dynamic power and dynamic power reduction design techniques

5: Static power and static power reduction techniques

6: The future, Introduction to IoT applications!!!

Dynamic Power

So what about energy and power?



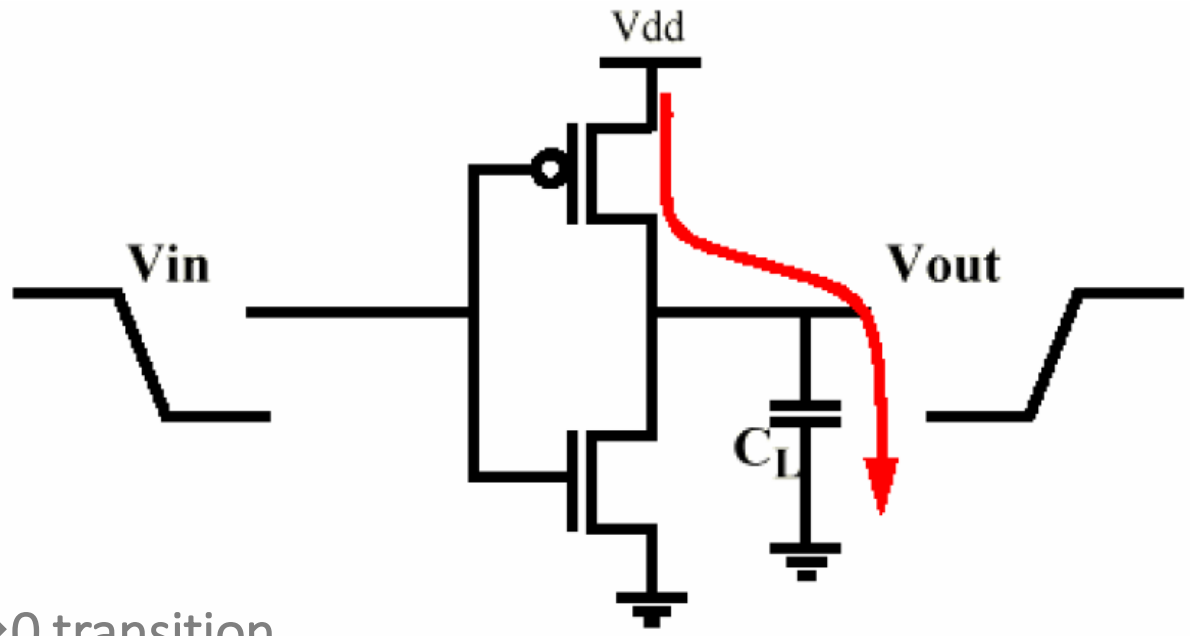
Energy for a 0→1→0 transition

$$Energy \propto Capacitive\ load \times Voltage^2$$

Only half energy is lost on a single transition 0→1 (or 1→0):

$$Energy \propto 1/2 \times Capacitive\ load \times Voltage^2$$

What about energy and power?



Energy for a 0→1→0 transition

$$Energy \propto Capacitive\ load \times Voltage^2$$

Only half energy is lost on a single transition 0→1 (or 1→0):

$$Energy \propto 1/2 \times Capacitive\ load \times Voltage^2$$

Power (it is called Dynamic Power)

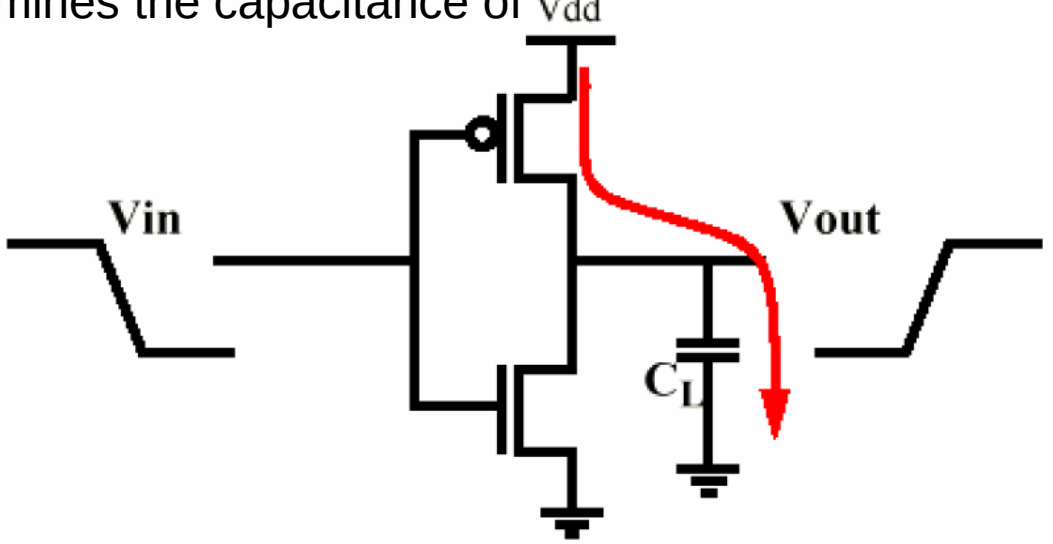
$$Power \propto 1/2 \times Capacitive\ load \times Voltage^2 \times Frequency\ switched$$

Dynamic Power

$$Power \propto 1/2 \times Capacitive\ load \times Voltage^2 \times Frequency\ switched$$

Frequency switched is a function of the clock rate.

Capacitive load per transistor is a function of both the number of transistors connected to an output (called the *fanout*) and the technology, which determines the capacitance of both wires and transistors



Models are a way to describe what we perceive and why not use it for analysis and prediction

Dynamic Power P_d model for a transistor:

$$P_d = V^2_{dd} * F_{clk} * C_L * E_{SW} * 0.5$$

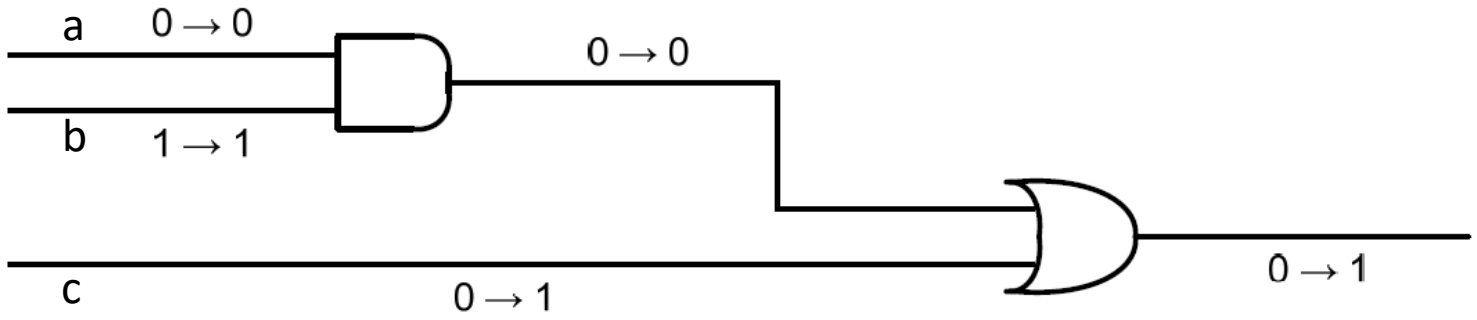
F_{clk} = clock frequency, C_L = capacitance seen by gate,

E_{SW} = gate switching activity (0-1 toggles)

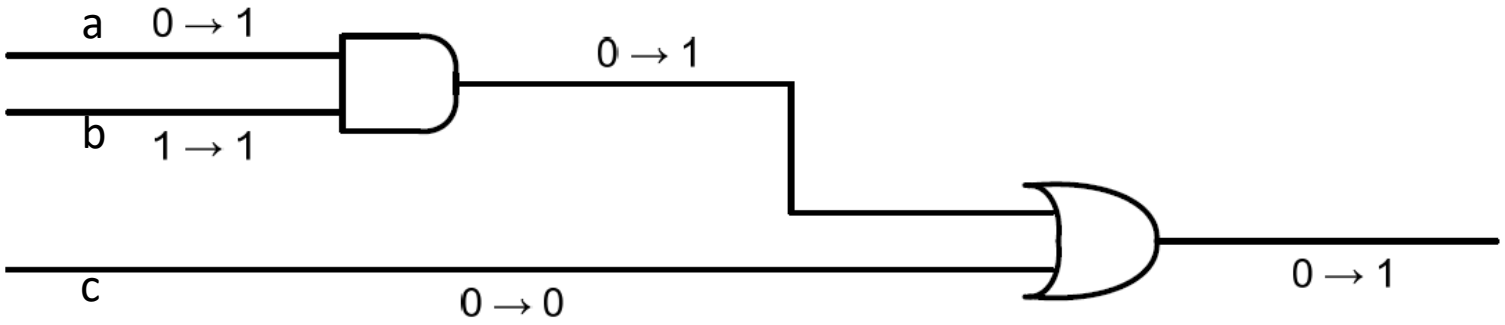
****there is another component of dynamic, which we haven't talked yet…

Workload dependence of dynamic power

- Switching activity depends on input patterns



OR gate consumes power if input 2 of OR gate changes form 0 to 1



AND and OR gate consume power if input 1 of AND gate changes form 0 to 1

Dynamic Power

Dynamic Power P_d model for a transistor:

$$P_d = V^2_{dd} * F_{clk} * C_L * E_{SW} * 0.5$$

Lumped model for a whole RTL netlist:

it is called switching activity coefficient and is a [0-1] variable and is workload dependent

$$P_{dynamic} = 0.5 * V^2_{dd} * F_{clk} * \eta * C_L * \alpha$$

operating frequency and voltage

Constant that depends on technology

Average load switching, it is usually constant for groups of workloads

This model (after fitted) has a 15% error compared to actual measurements

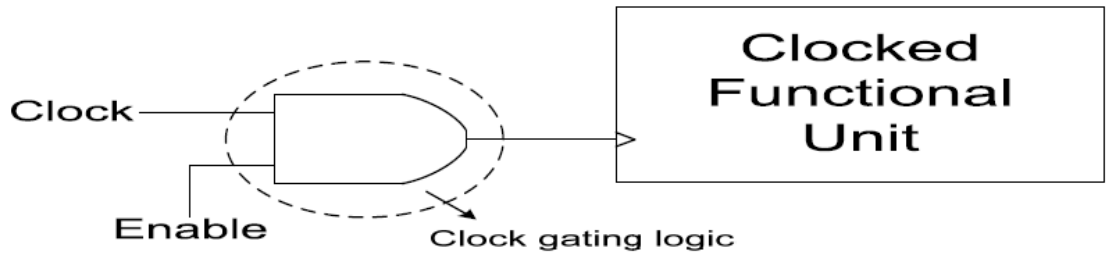
Dynamic power reduction design techniques

Dynamic power reduction happens at various design levels

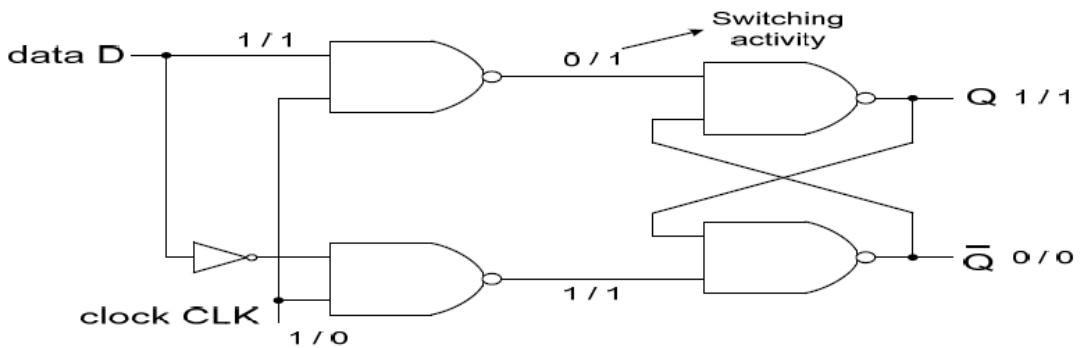
- Physical layout level = shorter interconnect → smaller C_L
- Transistor level = transistor with smaller W → smaller C_L
- Architectural level
  - Switching activity reduction: Clock gating
  - V_{dd} reduction: Multiple V_{dd}

Clock gating: architectural level dynamic power reduction

•When clocked modules (memory, register files, …) are inactive (stored values not charging), use clock gating to disable the clock feeding module



D Flip-Flop



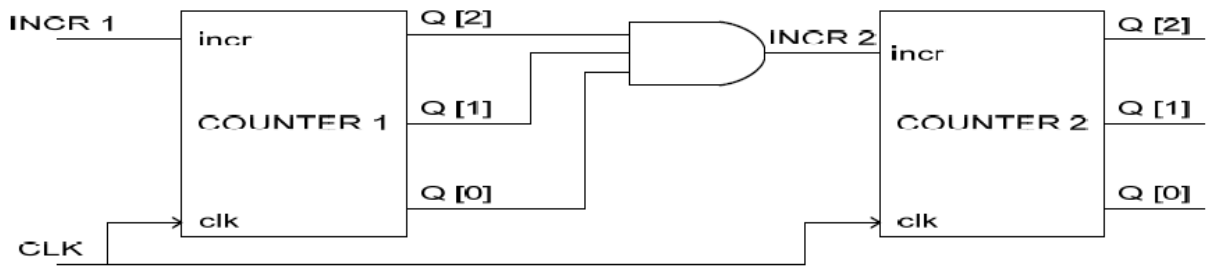
•FF consumes power due to switching activity even when input is not changing

•Clock gating reduces switching activity/dynamic power, up to 50% reduction

Example 1: Clock gating two, 3-bit counters

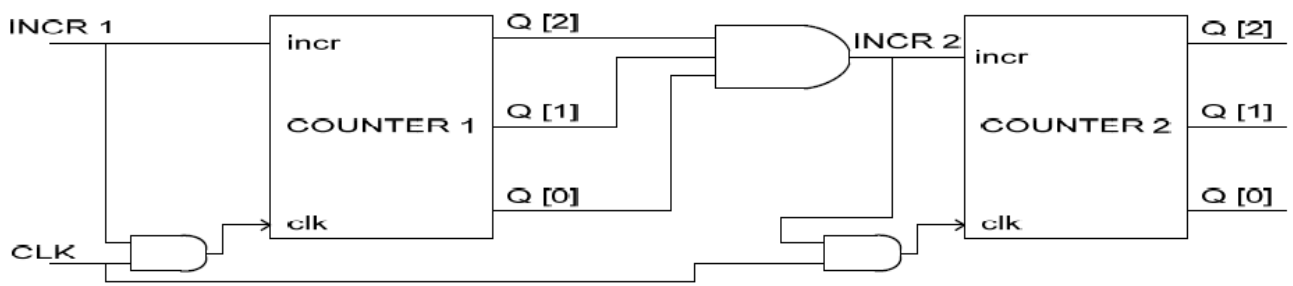
•Each counter increment on every clock cycle when INCR1 and INCR2 high

• INCR1 always high, COUNTER 1 increment every clock; COUNTER 2 increment only once every 8 clocks, but consume power during each clock cycle

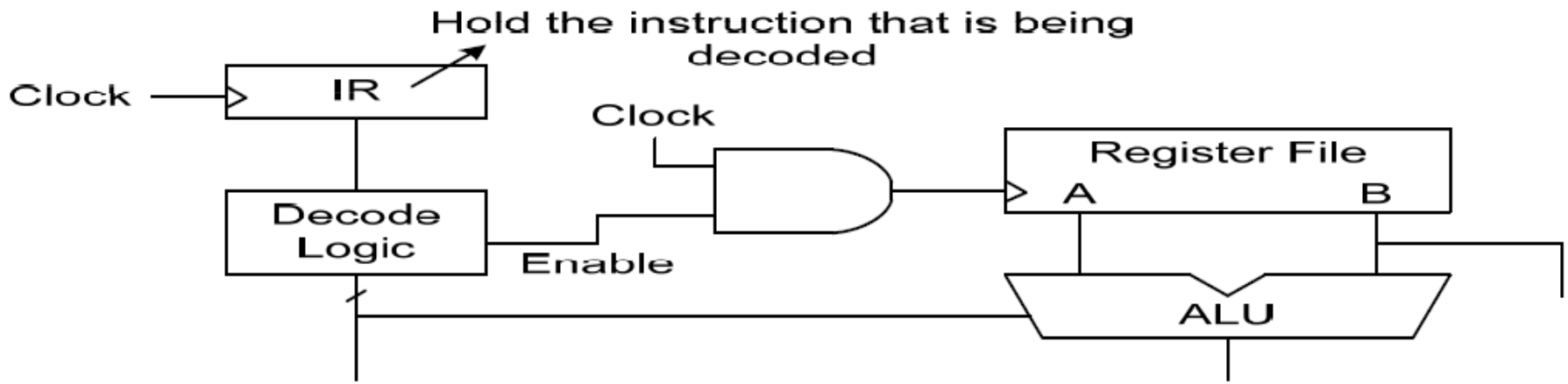


•Clock gating logic (2 AND) cause counters to be clocked only when their increment signals high

•COUNTER 2 not clocked when INCR2 is low



Example 2: Clock gating processor register file

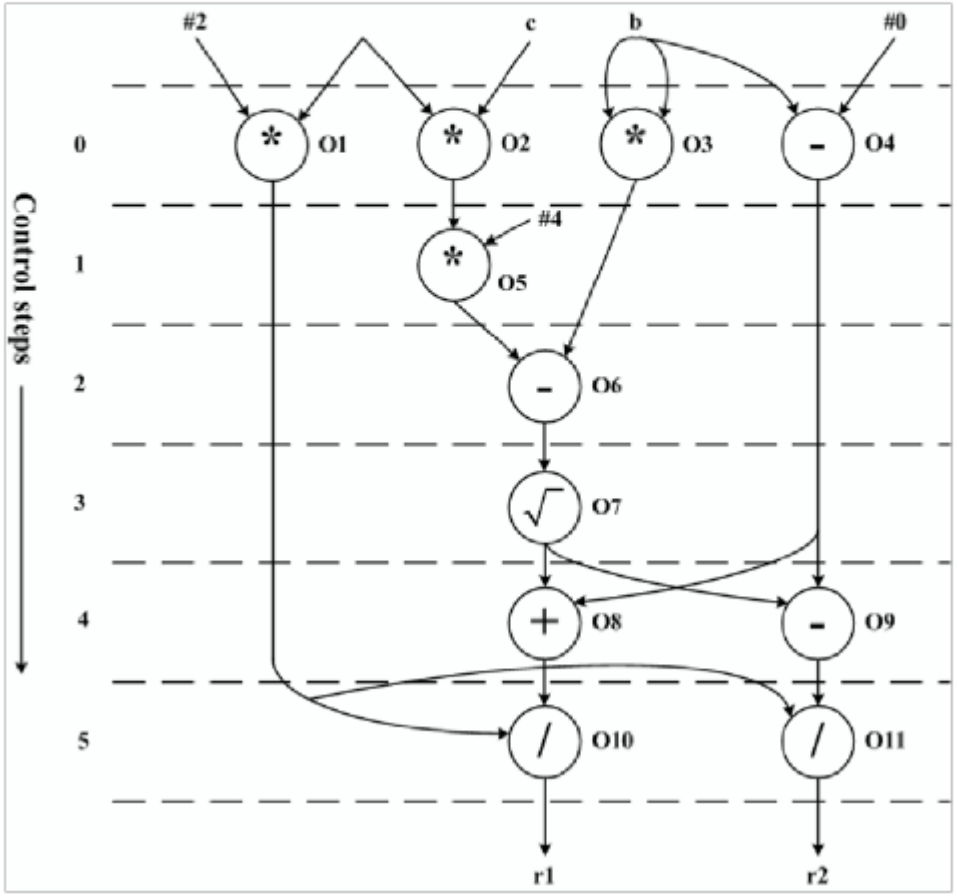


- Only clock register file when instruction decoding require the use of register file

Multi-Vdd (multiple power domains) by example

- Modules on critical path use highest Vdd (to meet required timing constraints)
- Modules on non-critical paths use lower Vdd (reducing power)

Example



Assume Vdd= 2.2V, Vt=0.5v, FUs delay=10ns

Critical path operations: o2,05,o6,07,o8,o9, o10, o11
Operations: o1,o3,o4 are candidates for multi-cycled FUs

Critical path length= 10+10+10+10+10+10= 60ns
Delay of multi-cycled o1= 50ns, o3=20ns, o4=40ns

Delay of digital circuit, $D= Kd*Vdd/(Vdd-Vt)^2$

Where D functional unit delay, Vdd supply voltage, Vt threshold voltage of the functional unit, and Kd a constant

Consider multi-cycled FU o1

$50= (Kd*Vdd)/(Vdd-0.5)^2$ ……. (1)
$10=(Kd*2.2)/(2.2-0.5)^2$ … …. (2)

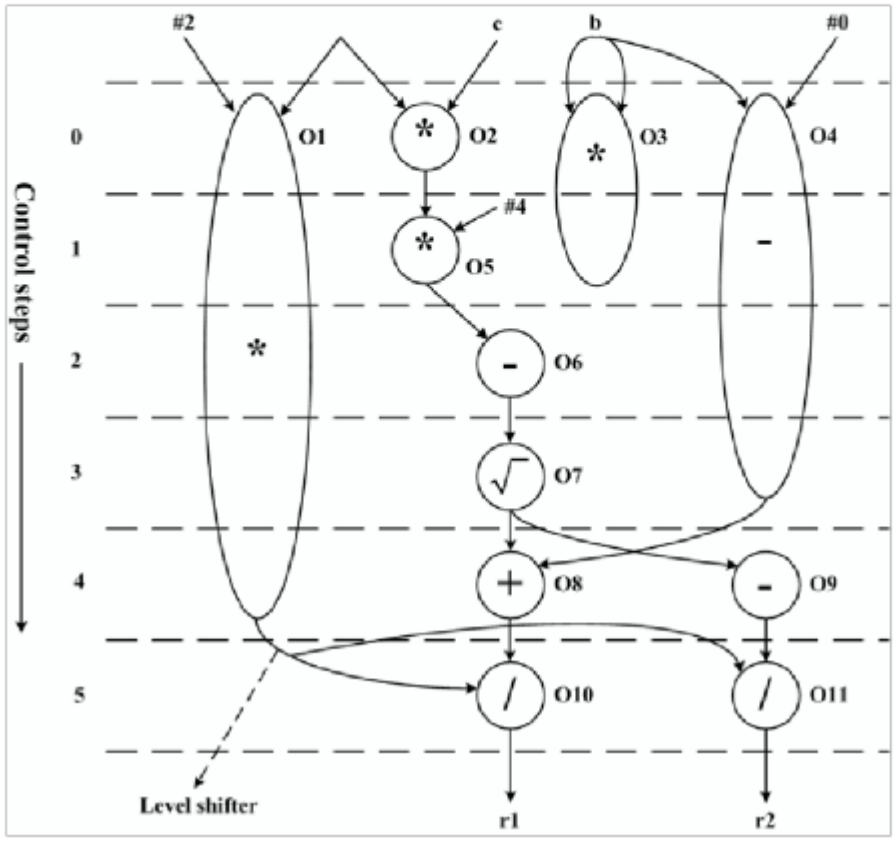
…

Multi-Vdd (multiple power domains)

From Eq.(2), Kd=13.1. Substitute Kd in Eq.(1) and solve for Vdd gives Vdd=1V.

The multi-cycled * (o1) will operate with Vdd=1V.

Similarly, multi-cycled FU o3, Vdd=1.48V.



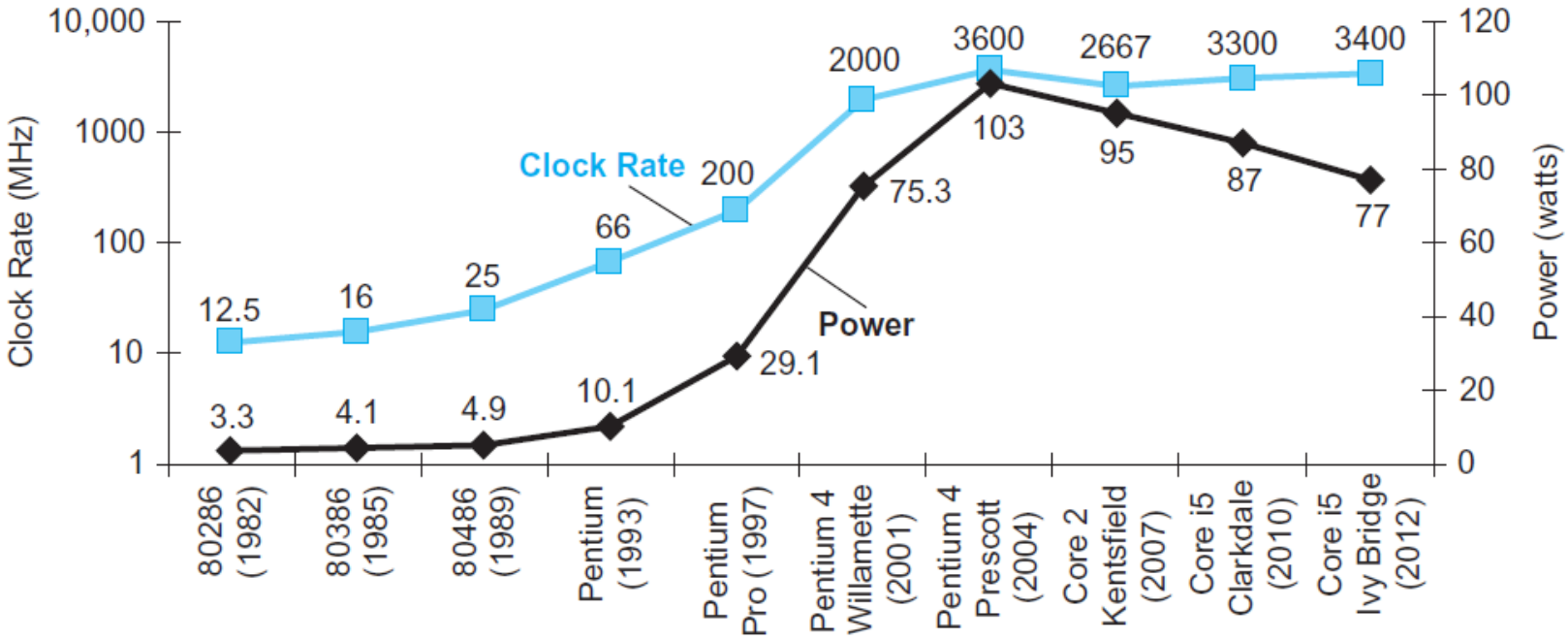
Multi-Vdd Design Requirements

- Use of multiple supply voltages
- Level shifters needed after reduced voltage module

Multi-Vdd design flow will be presented at leakage reduction techniques

Let's see what was happening few years ago again

$$P_{dynamic} = V^2_{dd} * F_{clk} * \eta * C_L * \alpha$$



How could clock rates grow by a factor of 1000 while power grew by only a factor of 30?

On each new generation energy and power are reduced by lowering Vdd (power is a function of the voltage squared). Vdd was reduced about 15% per generation. In 20 years, it has gone form 5 V to 1 V, which is why the increase in power is only 30 times.

Further Vdd reduction is difficult because it has dangerously approached threshold voltage Vth of transistors and reliability issues arise. This is what **near-threshold computing** topic is all about.

Relative power…

$$Power \propto 1/2 \times Capacitive\ load \times Voltage^2 \times Frequency\ switched$$

Relative Power

EXAMPLE

Suppose we developed a new, simpler processor that has 85% of the capacitive load of the more complex older processor. Further, assume that it has adjustable voltage so that it can reduce voltage 15% compared to processor B, which results in a 15% shrink in frequency. What is the impact on dynamic power?

ANSWER

$$\frac{\text{Power}_{\text{new}}}{\text{Power}_{\text{old}}} = \frac{\langle \text{Capacitive load} \times 0.85 \rangle \times \langle \text{Voltage} \times 0.85 \rangle^2 \times \langle \text{Frequency switched} \times 0.85 \rangle}{\text{Capacitive load} \times \text{Voltage}^2 \times \text{Frequency switched}}$$

Thus the power ratio is

$$0.85^4 = 0.52$$

Hence, the new processor uses about half the power of the old processor.