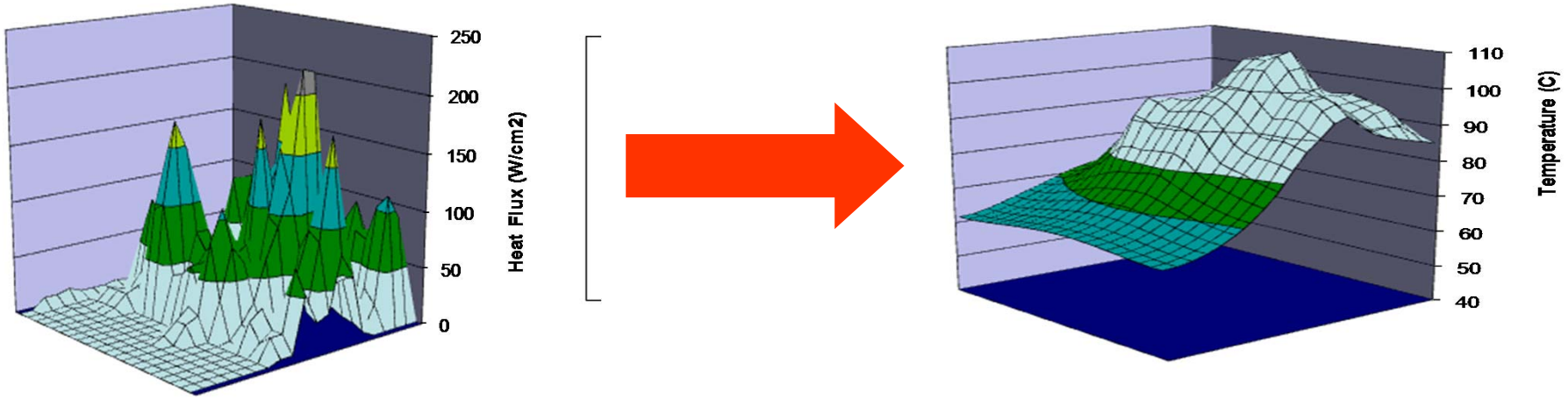
Recent interesting work on power/leakage reduction

Ramon Canal
NCD – Master MIRI

Hot Spots and Thermal Issues

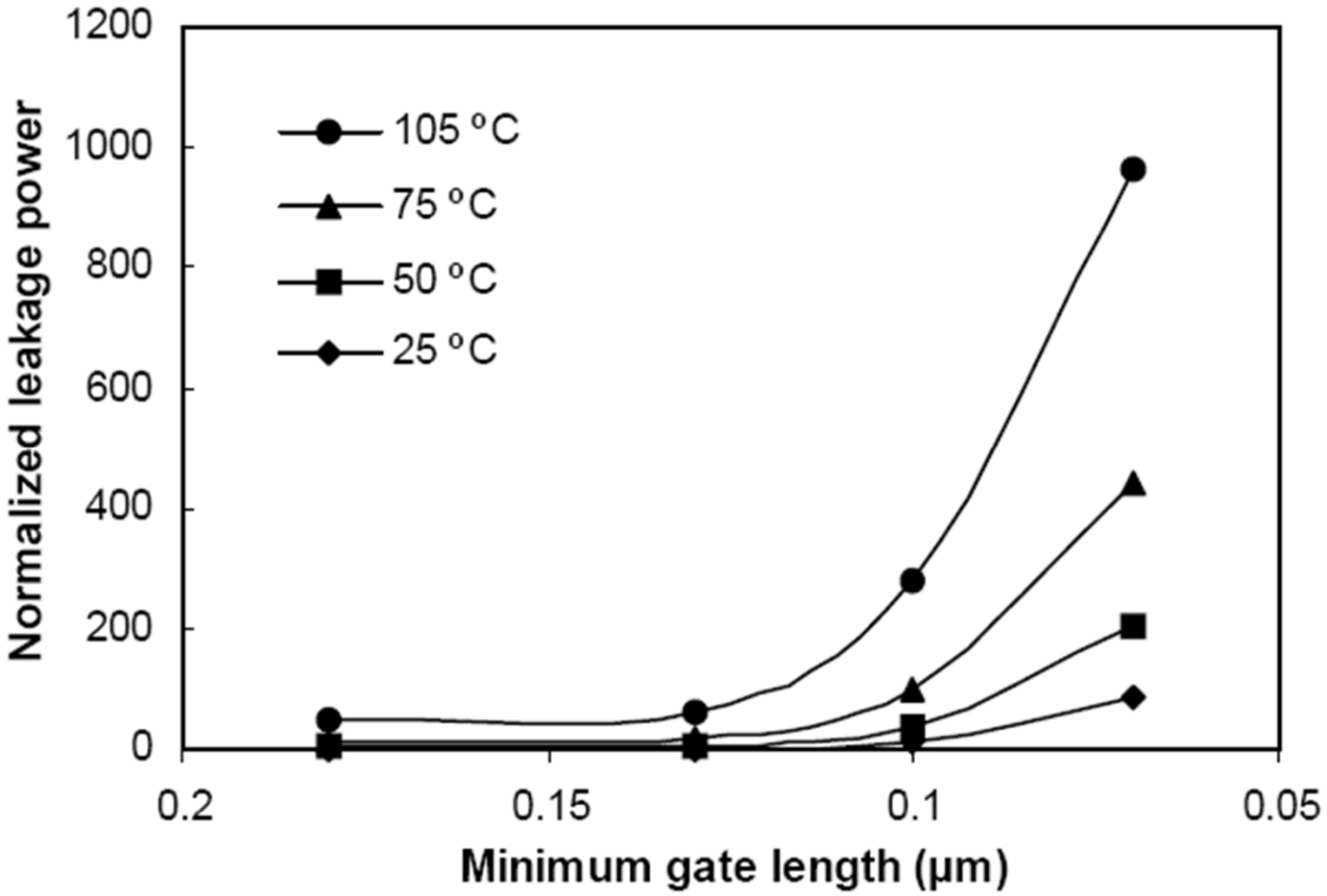
Power Map Pentium IV

On-Die Temperature Pentium IV



- Silicon is not a good heat conductor
- Temperature ↑ → *leakage* ↑ → power ↑ → temperature ↑
- Having a larger circuit does not help. The *hot* spots must be reduced.
- A good layout can separate the *hot spots* and thus reduce the temperature peaks –*power envelope*.

Thermal effects on leakage



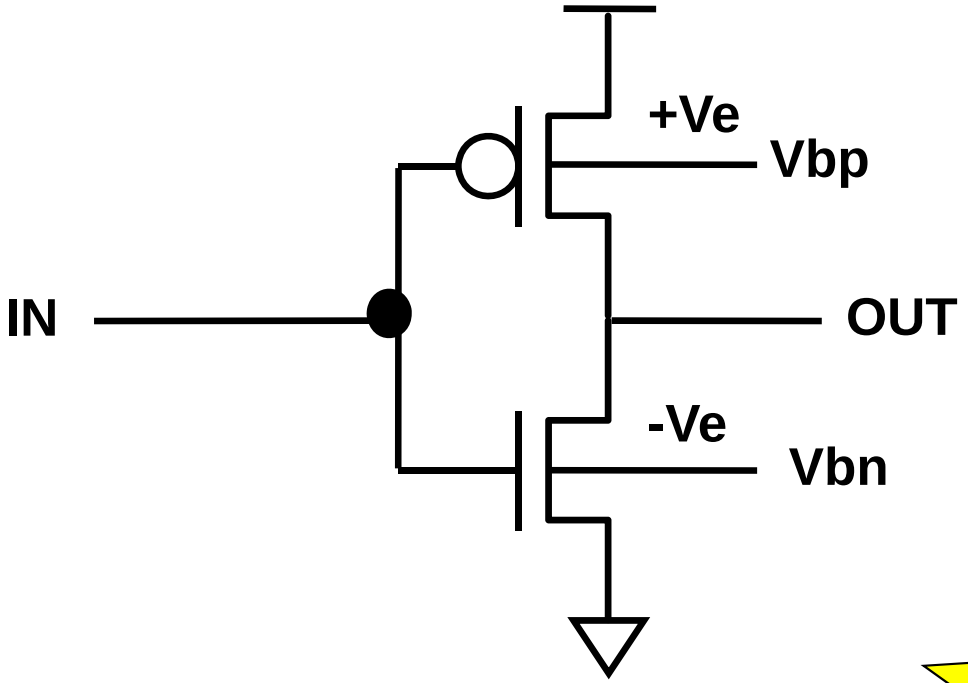
Flautner et al. "Drowsy Caches: Simple Techniques for Reducing Leakage Power", ISCA 2002

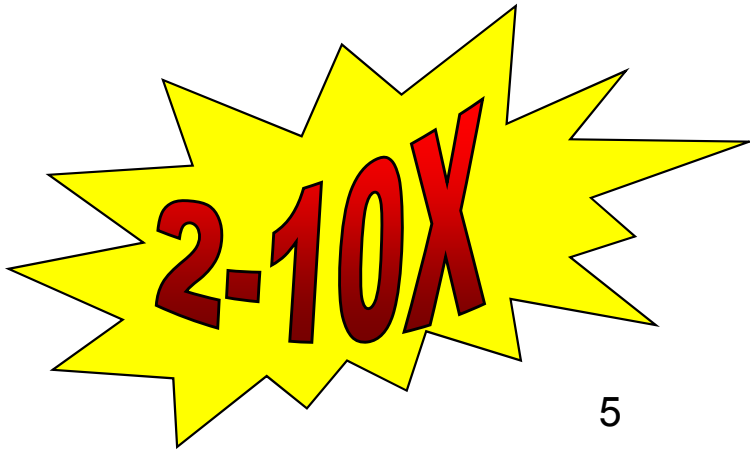
Low leakage techniques

- Circuit level
  - Body bias
  - Stack effect
  - Sleep transistors (Vdd-gating)
- Architecture level
  - Drowsy caches
  - Cache Decay
- System level
  - Compiler-directed
  - Voltage-Frequency Scaling

Circuit level techniques

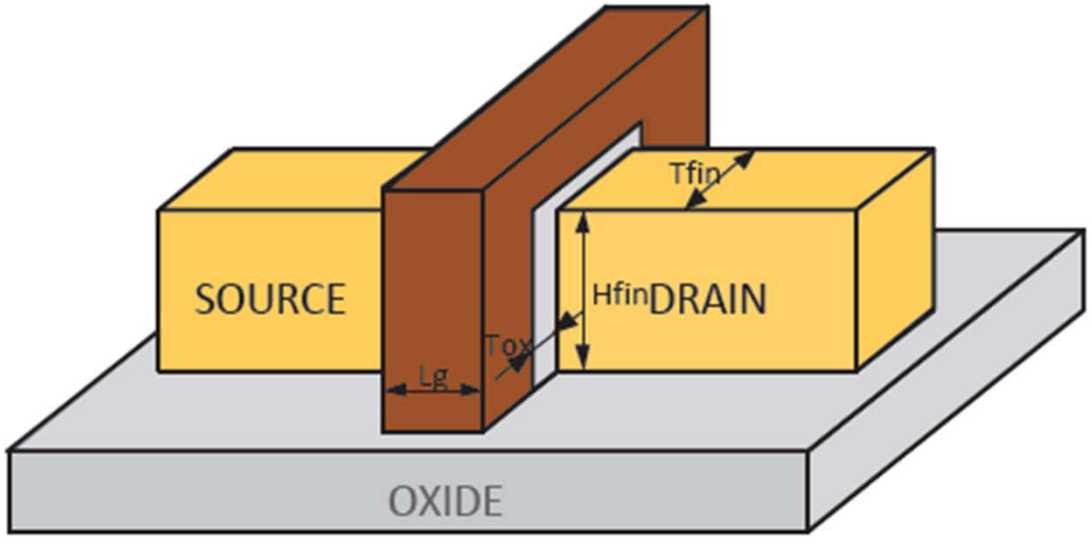
Body bias (Planar CMOS)

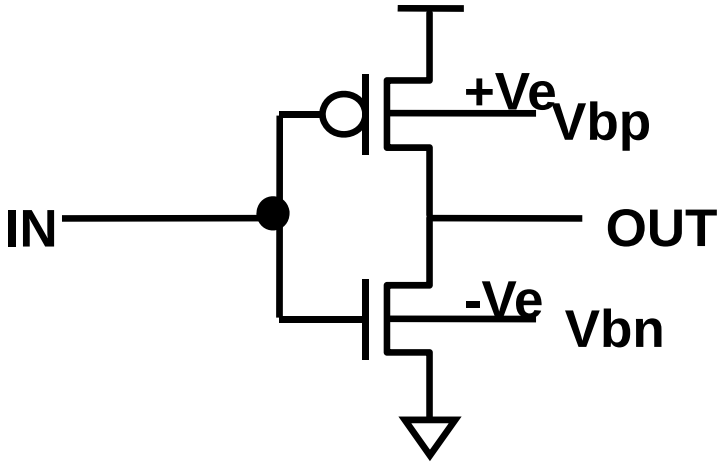


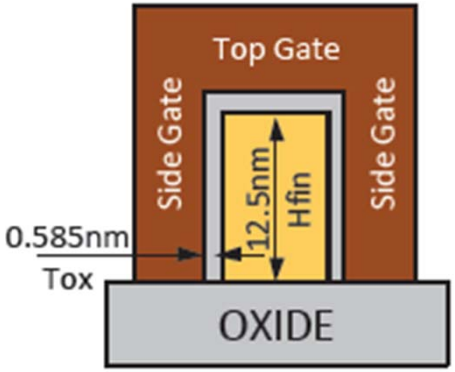


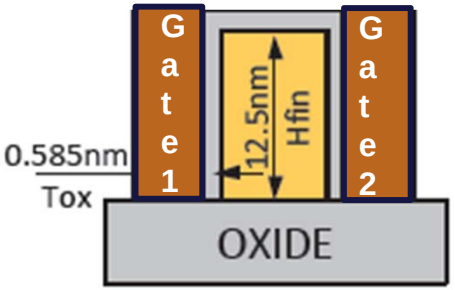
Circuit level techniques

Backgate bias (FinFET)



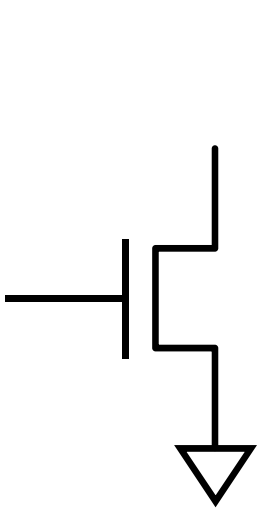


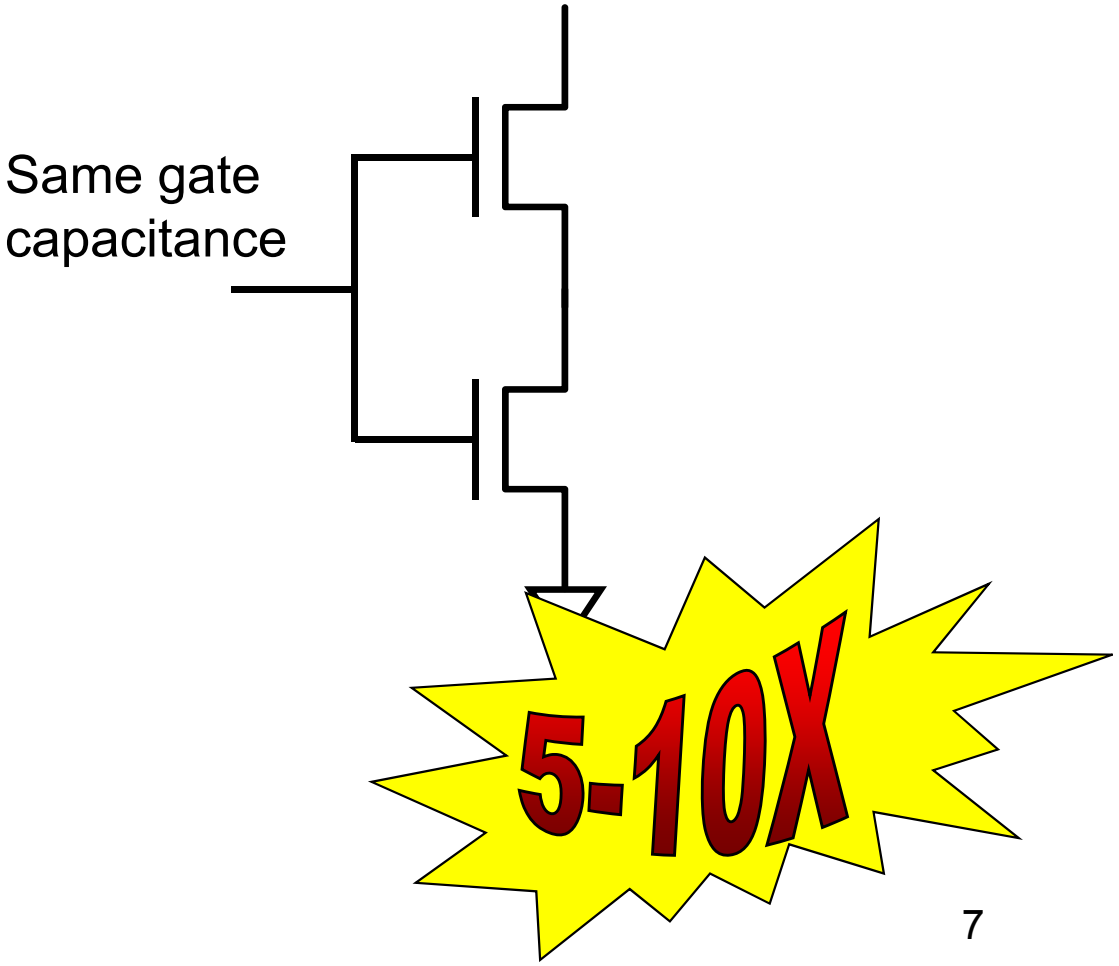




Circuit level techniques

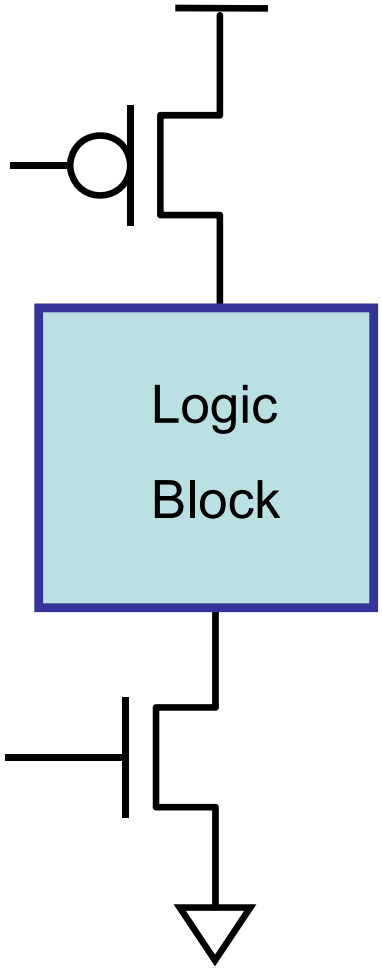
Stack effect (Chen et al. –Hong Kong U. ISLPED 1998)

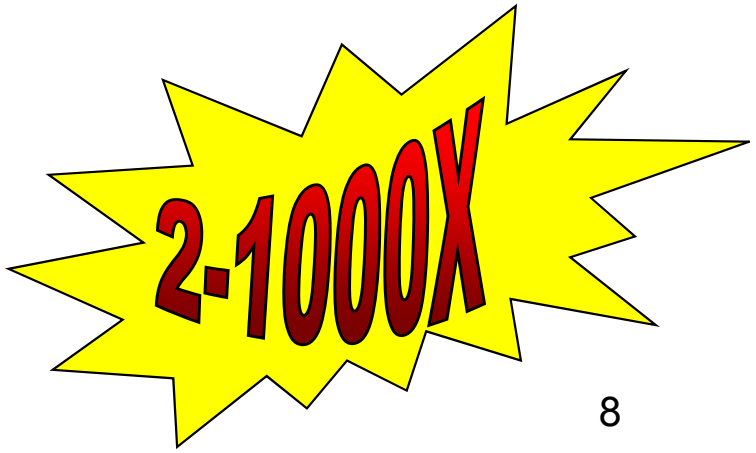




Circuit level techniques

Sleep transistors (Powell et al. –Purdue U.- ILSPED 2000)





Circuit level techniques

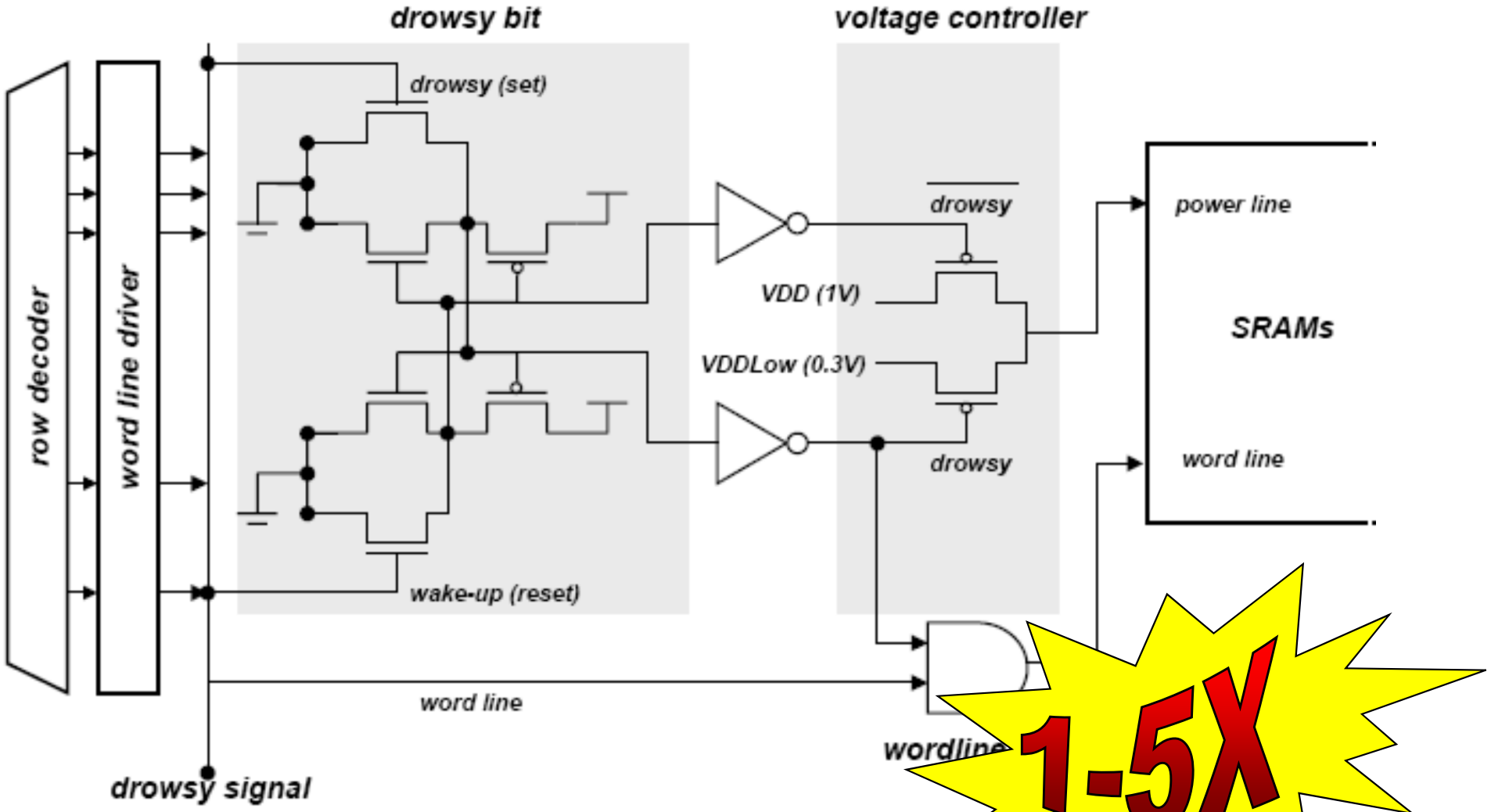
- Circuit level
  - Body bias
  - Stack effect
  - Sleep transistors (Vdd-gating)
- Aims:
  - Reduce leakage
- Consequences
  - Extra transistors
  - Multiple speed transistors
    - Maybe leading to multiple speed units, datapaths…

Architecture level techniques

- Architecture level
  - Drowsy caches
  - Cache Decay
- Aims:
  - Reduce leakage by turning-off/"low powering" unused (or possibly unused) cache lines
- Consequences
  - Extra circuitry
  - Access delays
  - Execution penalties

Architecture level techniques

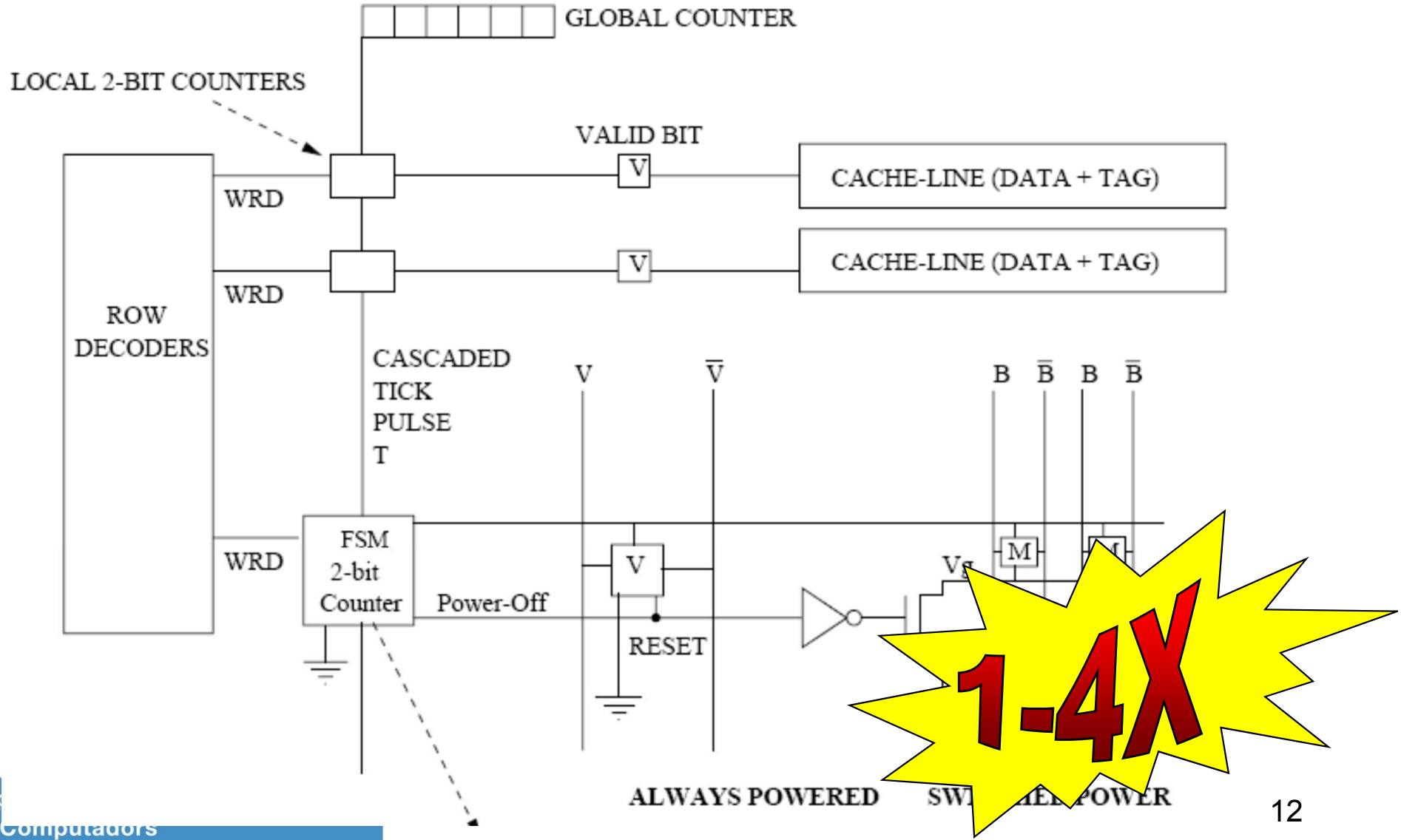
Drowsy Caches (K. Flautner et al. –ARM, U. Michigan- ISCA 2002)



Note that, for simplicity, the word line, bit lines, and two pass transistors of the drowsy bit are not shown in this picture.

Architecture level techniques

Cache Decay (S. Kaxiras et al. –Agere, Princeton U.- ISCA 2001)



Architecture level techniques

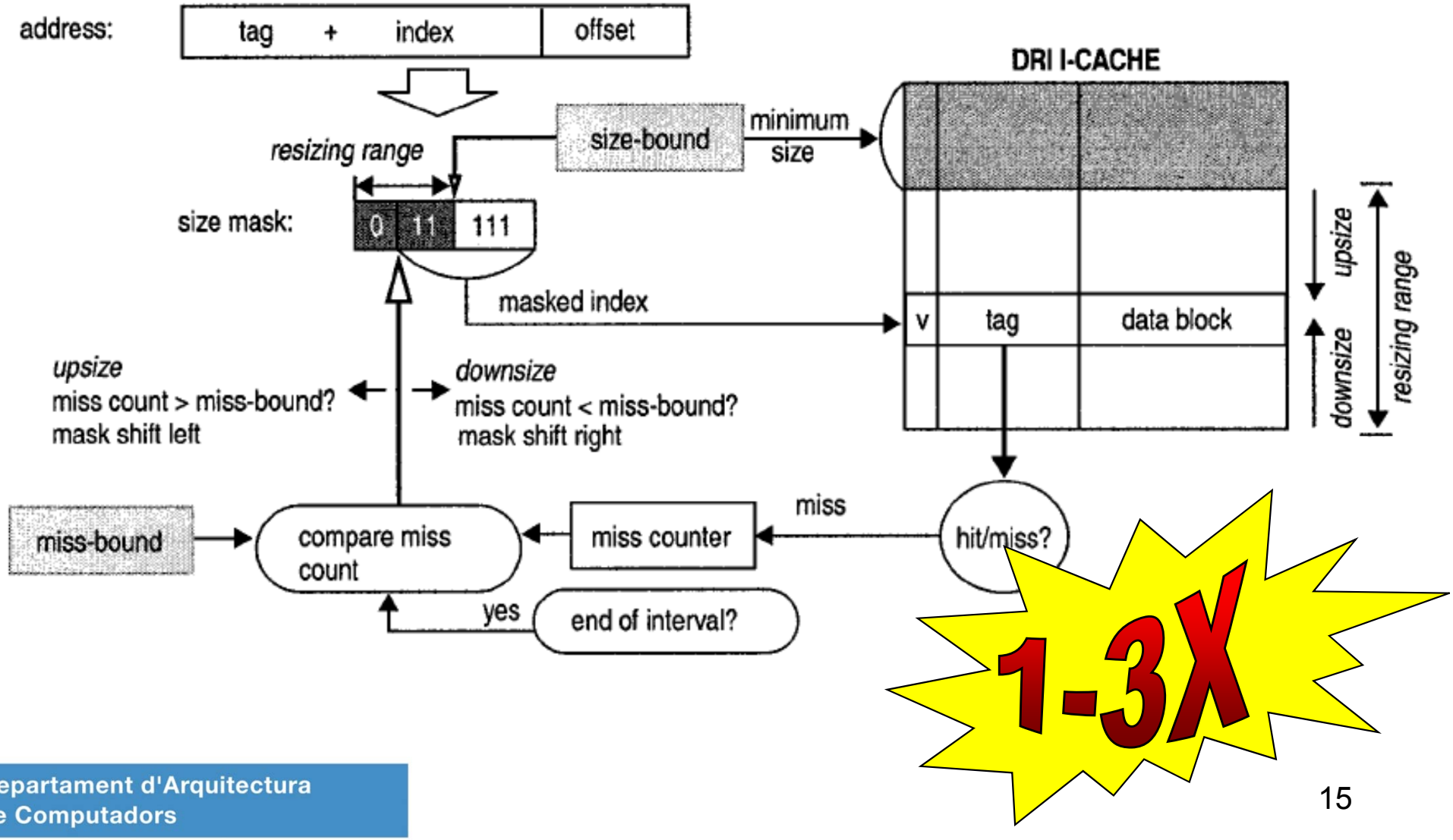
- Architecture level
  - Drowsy caches
  - Cache Decay
- Aims:
  - Reduce leakage by turning-off/"low powering" unused (or possibly unused) cache lines
- Consequences
  - Extra circuitry
  - Access delays
  - Execution penalties

System level techniques

- System level
  - Compiler-directed cache-decay/drowsy caches
  - Cache bypassing
  - Voltage-Frequency Scaling
- Aims:
  - Reduce leakage by turning-off/"low powering" unused (or possibly unused) cache lines
- Consequences
  - Difficult compile-time decisions
  - Extra circuitry
  - Access delays
  - Execution penalties

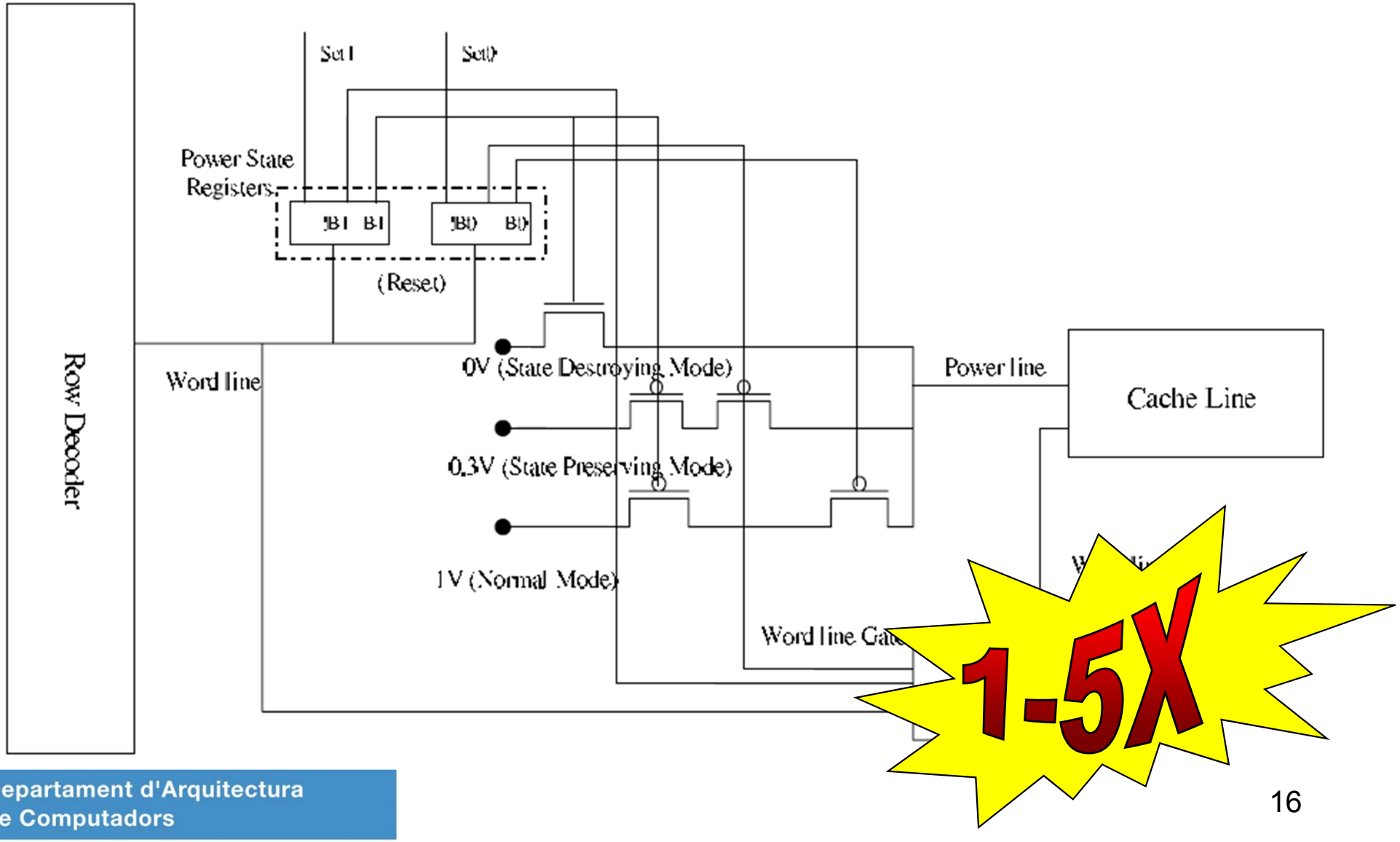
System level techniques

Adaptative Cache (Powell et al. –Purdue U.- IEEE Trans. VLSI 2/2001)



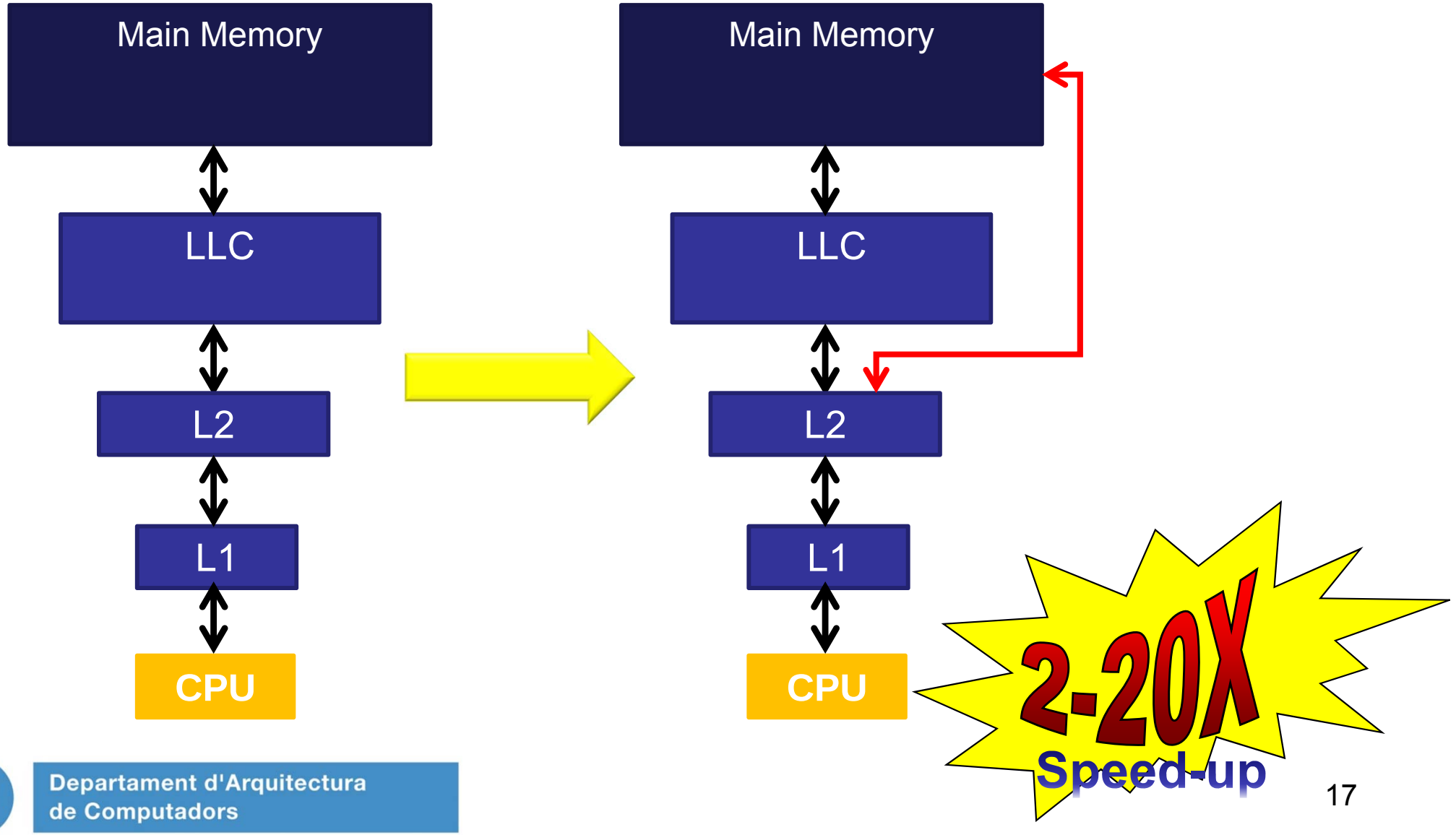
System level techniques

Compiler-directed (Zhang et al. –Penn. State U.- MICRO 2002)



System level techniques

Cache bypassing (Chi et al. –Phillips.- HICSS 1989). Revisited!



System level techniques

- System level
  - Compiler-directed cache-decay/drowsy caches
  - Cache bypassing (hardware and/or software)
  - Voltage-Frequency Scaling
- Aims:
  - Reduce leakage by turning-off/"low powering" unused (or possibly unused) cache lines
- Consequences
  - Difficult compile-time decisions
  - Extra circuitry
  - Access delays
  - Execution penalties

Conclusions

- Need for power, energy and thermal awareness.
  - (TDP) Thermal Design Power
  - Even commercially advertised power efficiency (Cool'n quiet –AMD)
- Extensive use of power/performance metrics such as Energy Delay
  - MIPS/W (Intel)