

# Multicore: Why is it happening now?

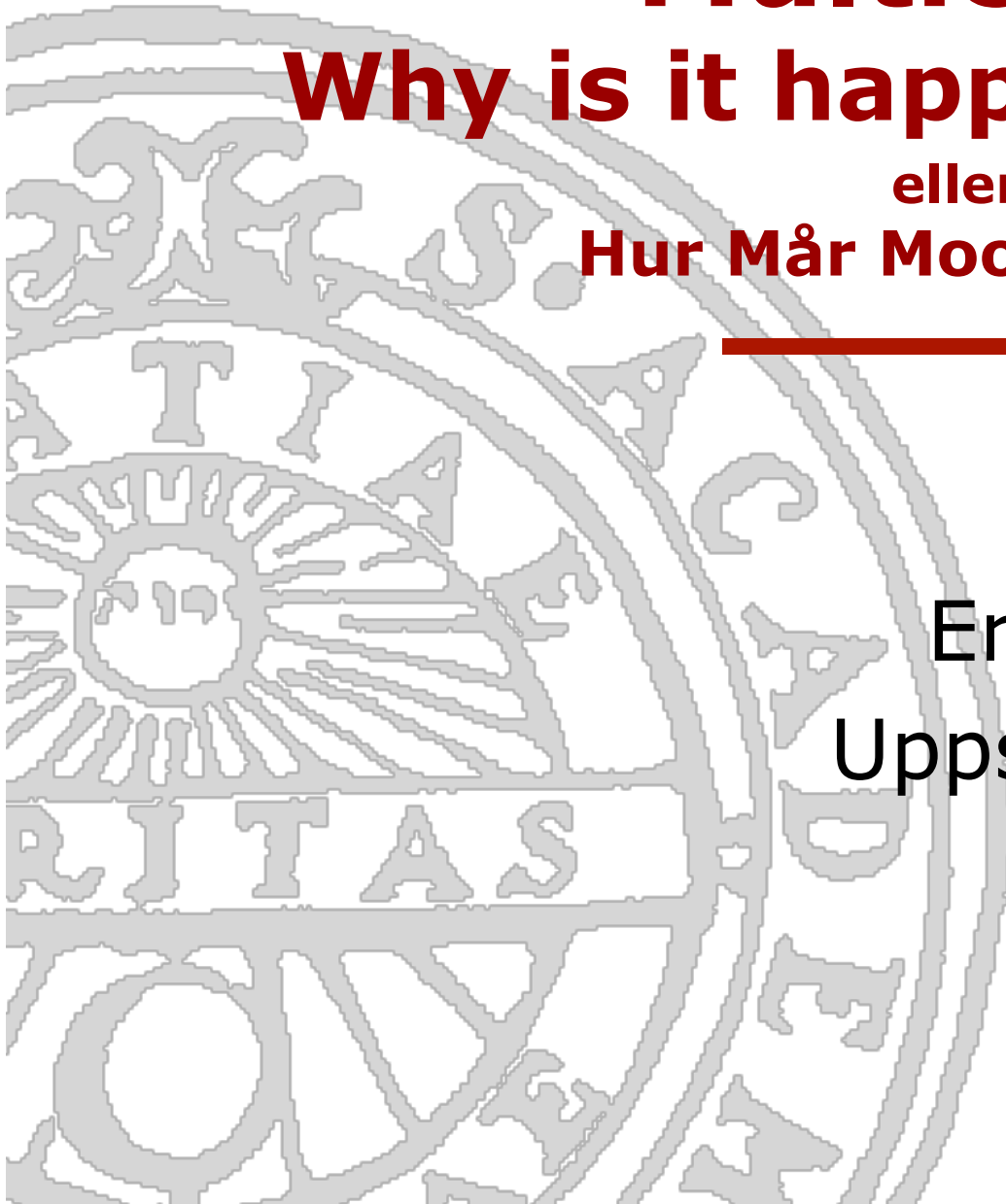
eller

## Hur Mår Moore's Lag?

---

Erik Hagersten

Uppsala Universitet





# Outline of these lectures

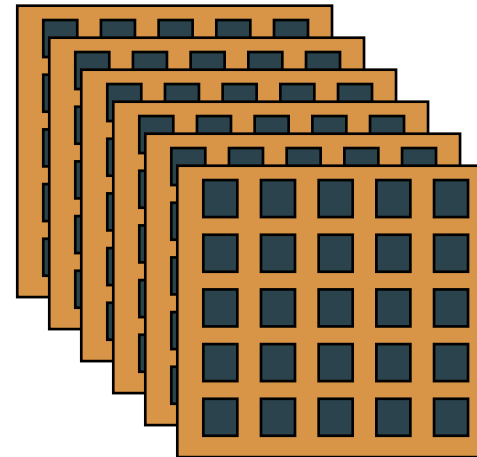
1. Processor implementations
2. Caches and memory system
3. Multiprocessors
4. HW optimizations
- 5. Multicore processors**
6. SW optimizations

# Darling, I shrunk the computer

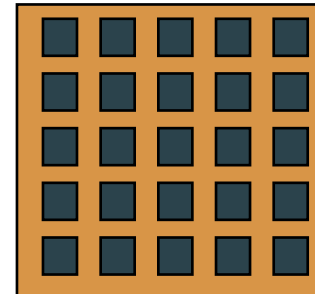


UPPSALA  
UNIVERSITET

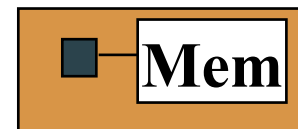
**Mainframes**



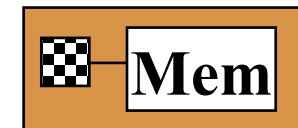
**Super Minis:**



**Microprocessor:**



**Multicore: Many CPUs on a chip!**



PDC  
Summer  
School  
2017

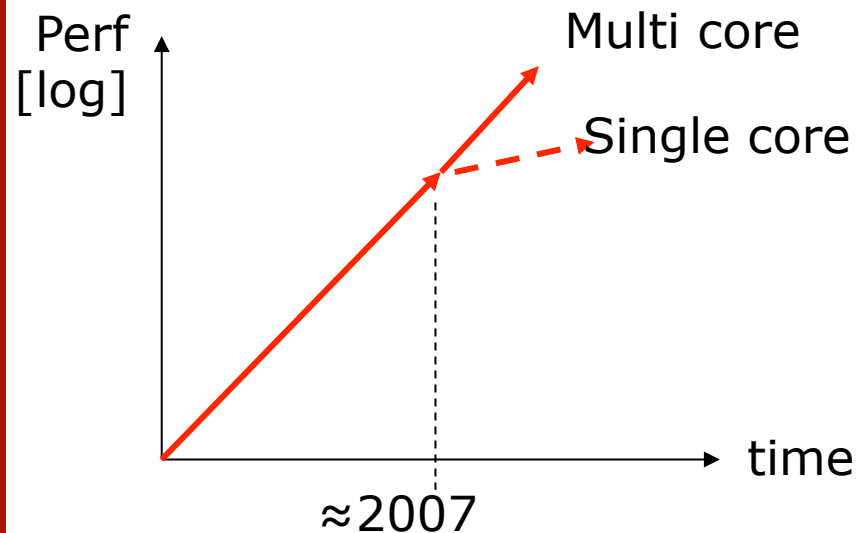


# Outline

- Why multicore now?
- Performance bottlenecks in MCs
- Commercial offerings
- Reflection for the future



# Everybody is doing it! But, why now?

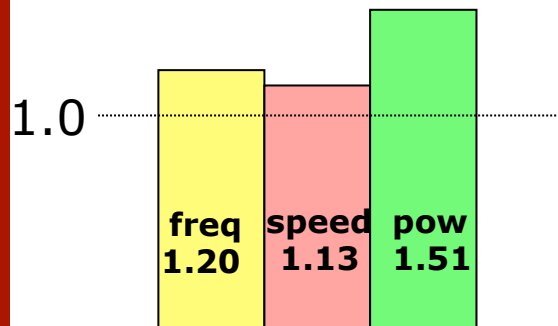


1. Not enough ILP to get payoff from using more transistors on a single core
2. Signal propagation delay  $\gg$  transistor delay
3. Power consumption  $P_{\text{dyn}} \sim C \cdot f \cdot V^2$

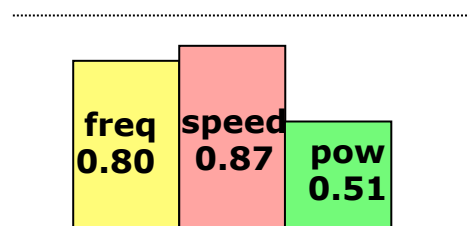


# Example: Freq. Scaling

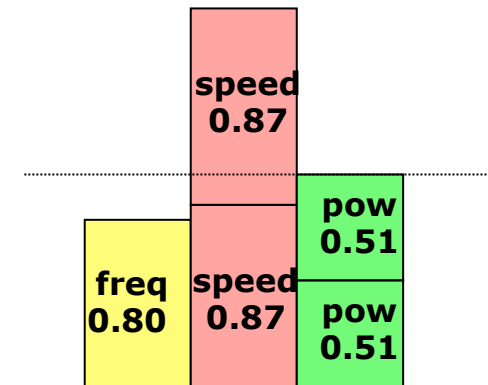
$$P_{\text{dyn}} = C * f * V^2 \approx \text{area} * \text{freq} * \text{voltage}^2$$



20% higher freq.

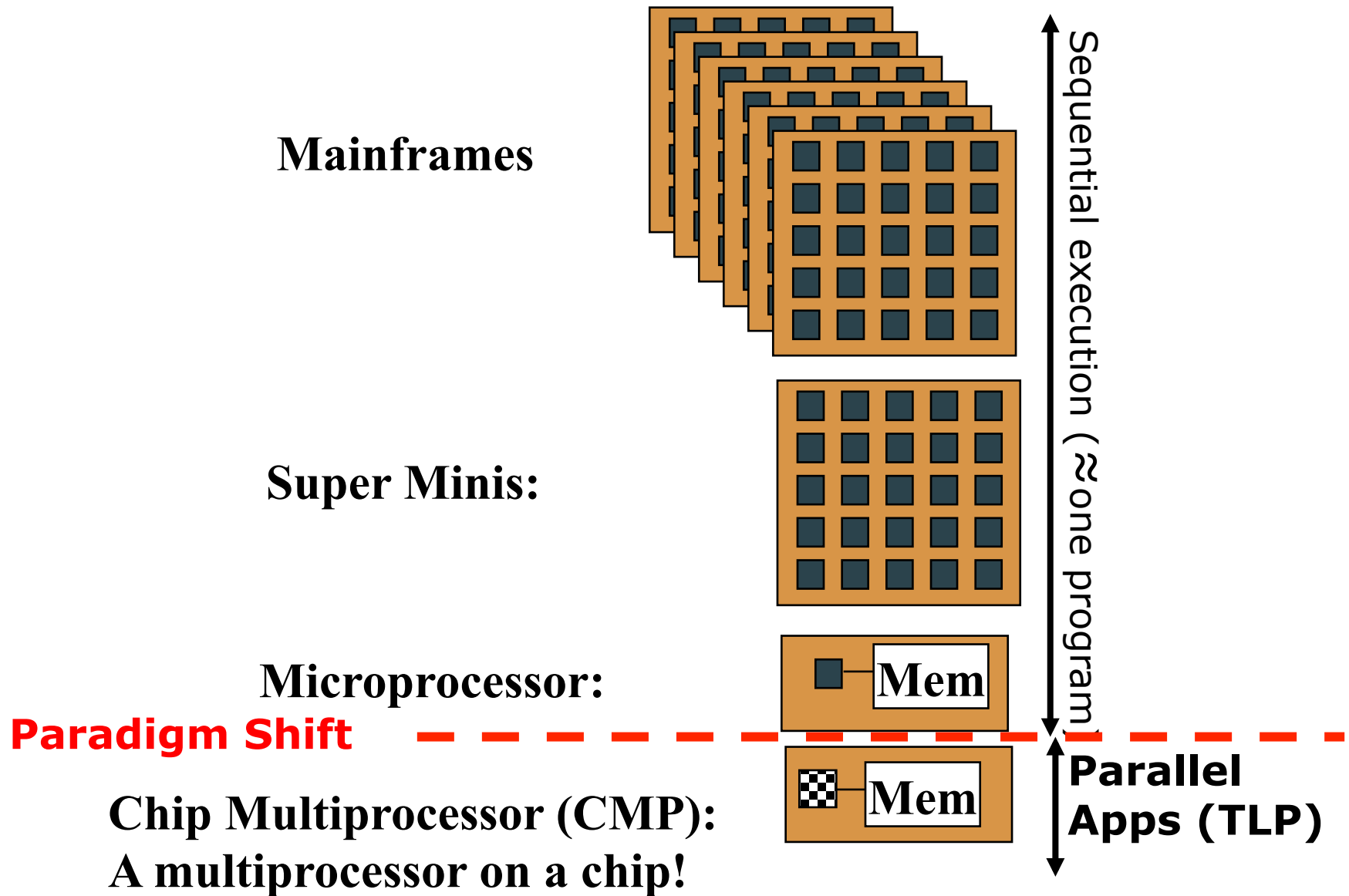


20% lower freq.

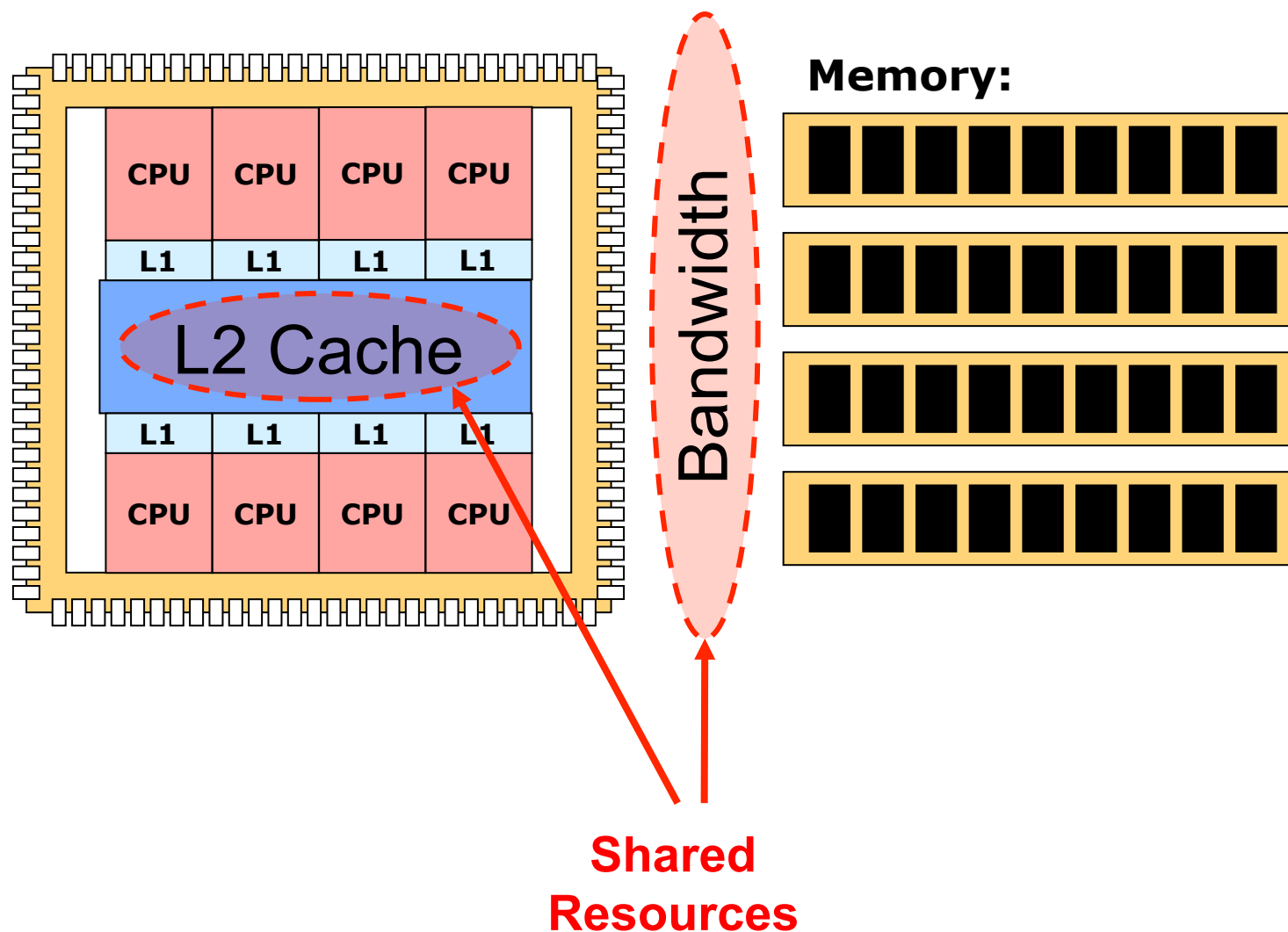


20% lower freq.  
Two cores

# Darling, I shrunk the computer



# Shared Resources







# **Commercial snapshot eller "Kärnornas krig"**

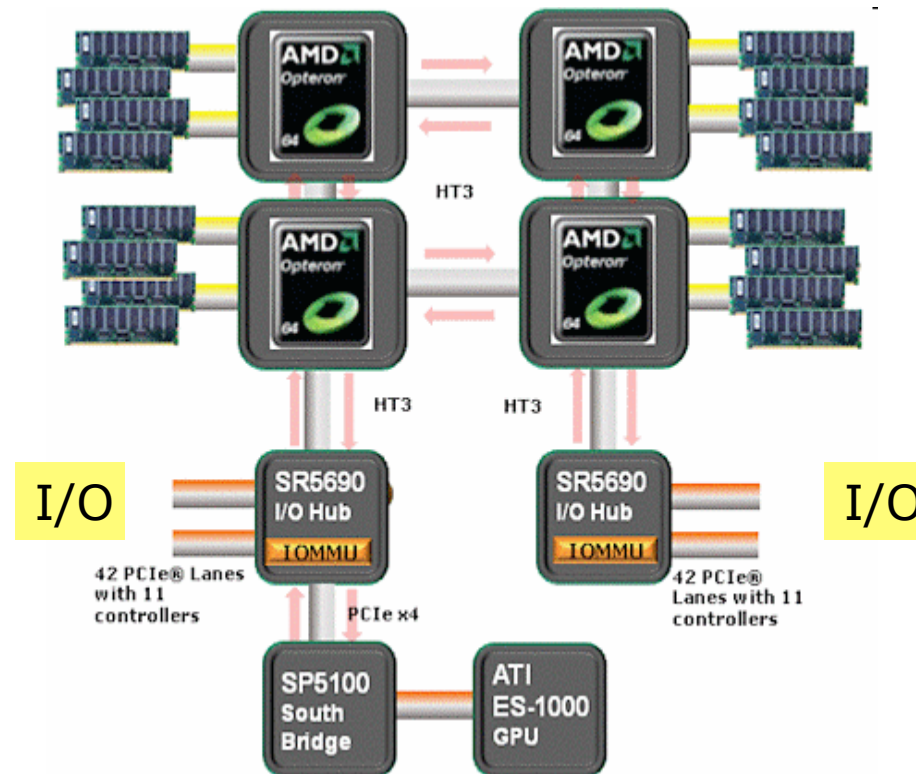
---

**(Examples of early multicores)**

Erik Hagersten  
Uppsala Universitet



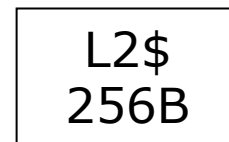
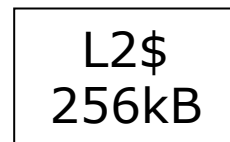
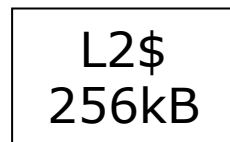
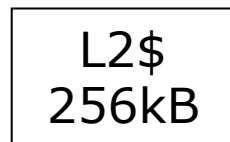
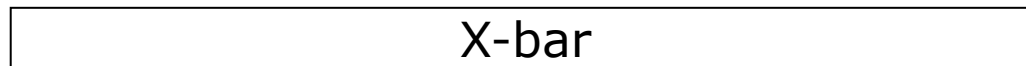
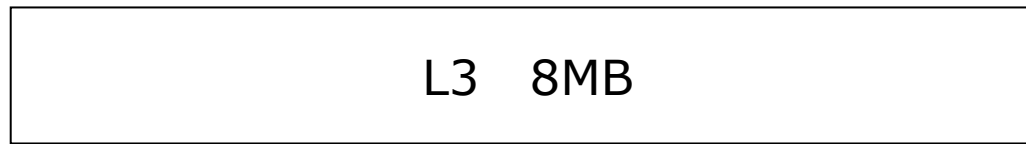
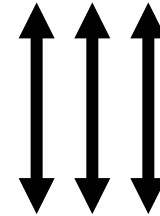
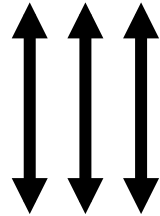
# AMD MC System Architecture



# Intel: Nehalem, Core i7 Q1 2009 (4 cores)

QuickPath Interconnect

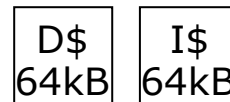
3x DDR-3 DRAM



CPU, 2 thr.



CPU, 2 thr.



CPU, 2 thr.

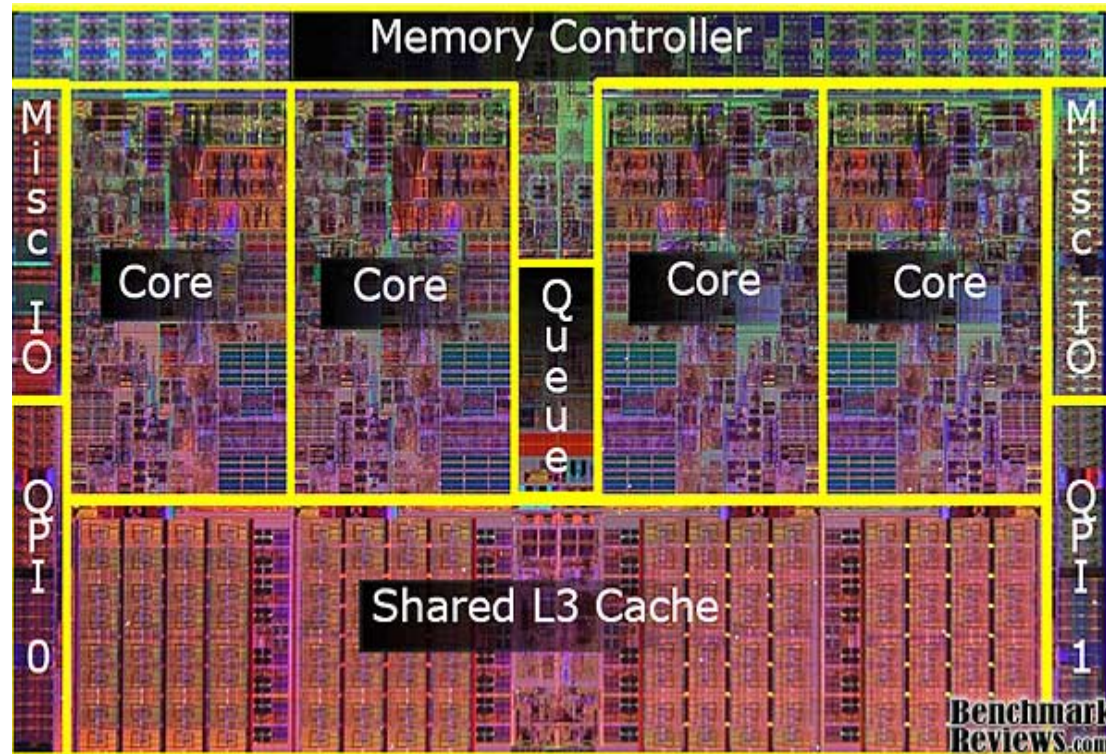


CPU, 2 thr.

Up to 4 cores x 2 threads

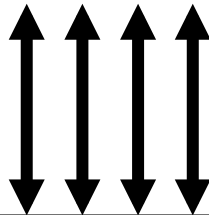


# Nehalem "Core i7"

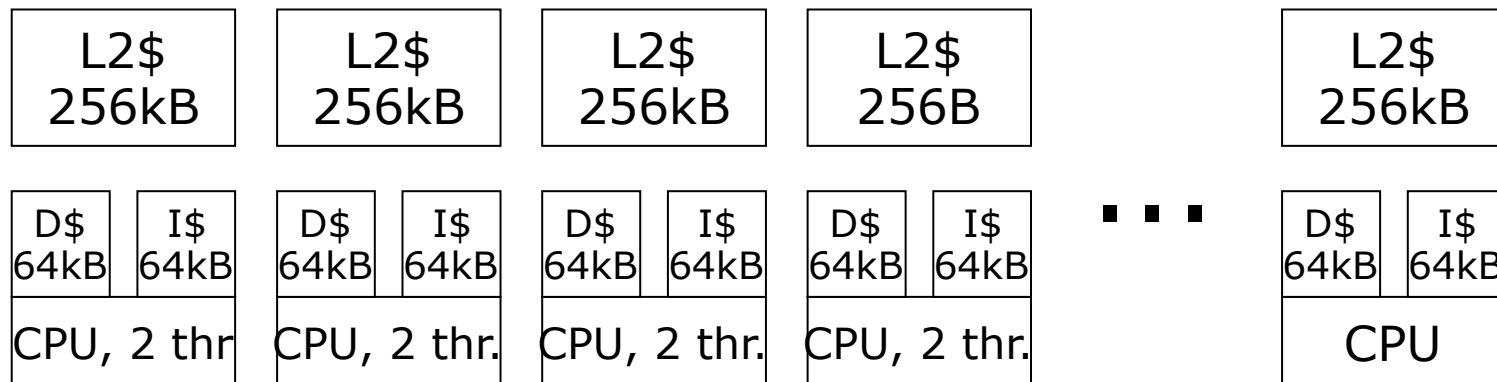
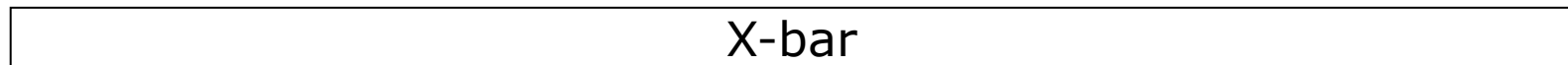
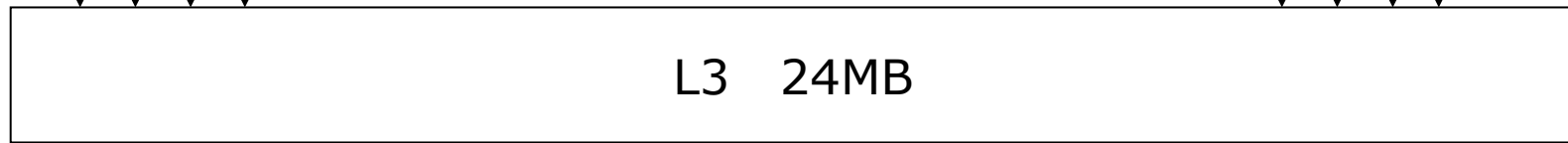
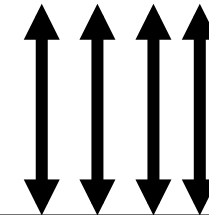


# Intel: "Nehalem-Ex" (i7)

QuickPath Interconnect



4 x DDR-3



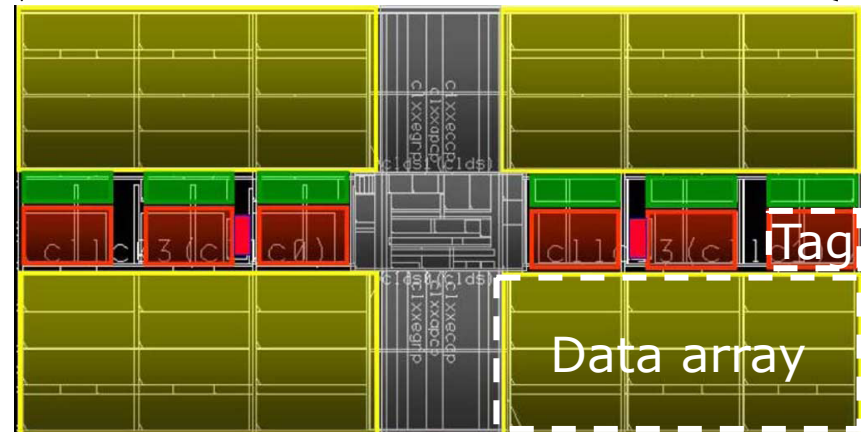
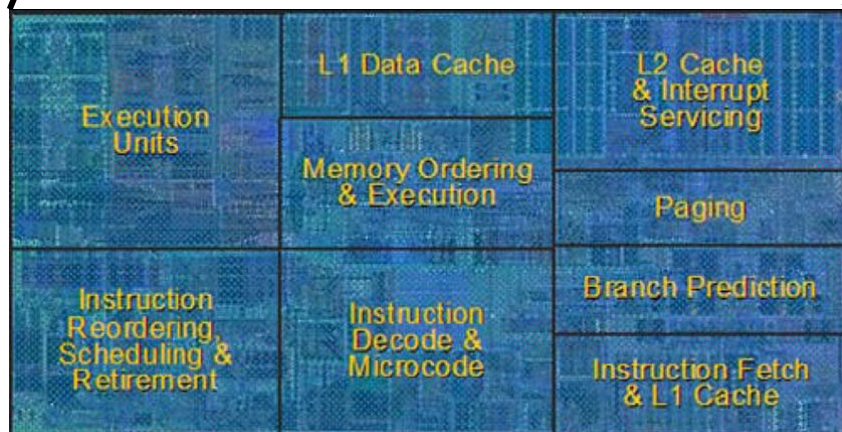
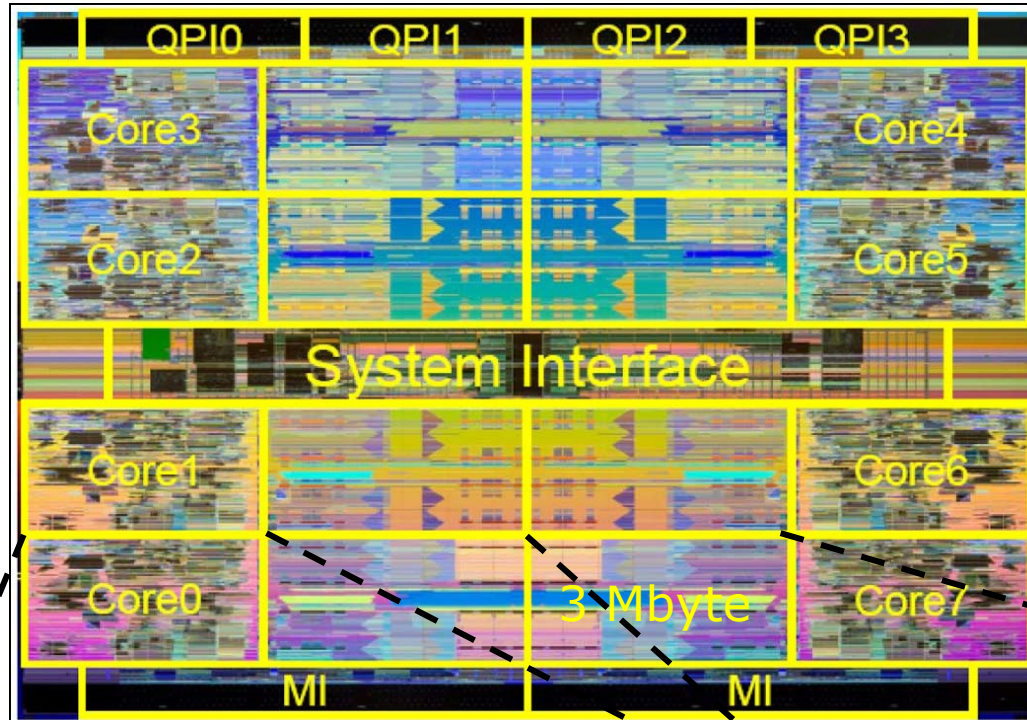
8 cores x 2 threads

Multicores 13



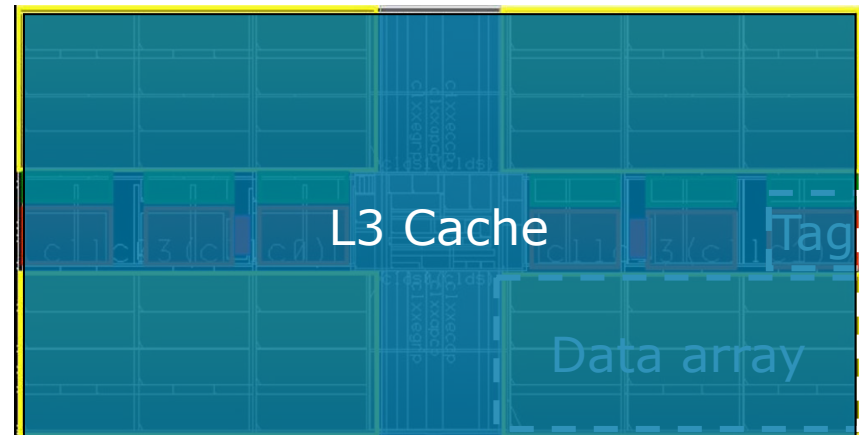
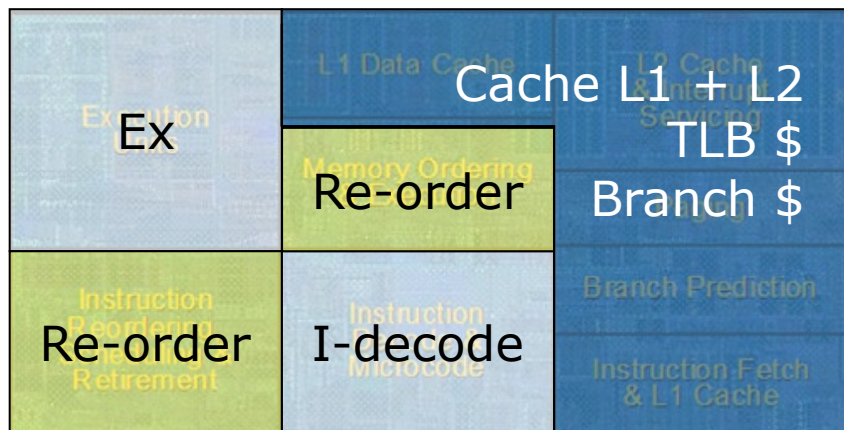
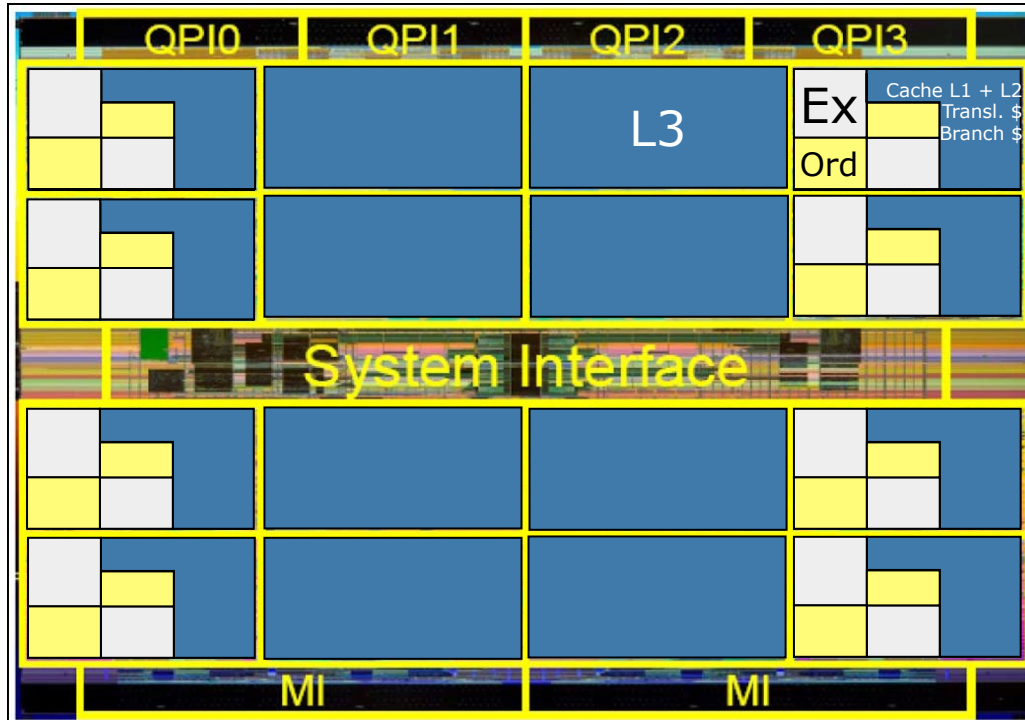


# How is the silicon used (i7-Ex)?



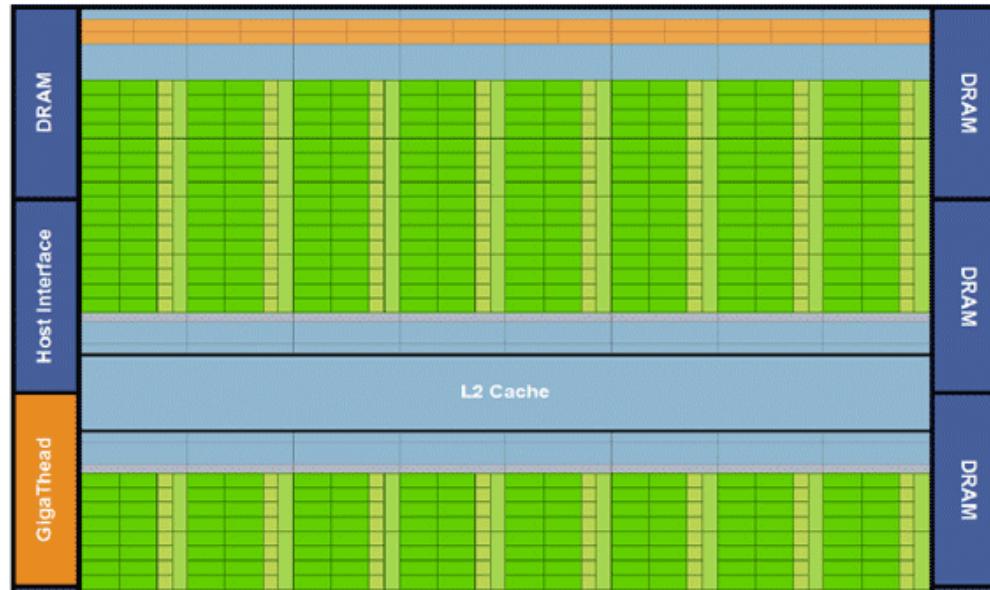


# How is the silicon used?





# GPUs and Accelerators: Lots of hype in HPC!



- 512 "cores" (**P**)
- 16 **P**/StreamProcessor (**SP**)
- Full DP-FP IEEE support
- 64kB L1 cache /**SP**
- 768kB global shared cache
- Atomic instructions
- ECC correction
- Debugging support
- ...

- **CUDA language. Requires "heroic programming" [M.Bull, EPCC]**
- **How much is an nVIDIA core worth vs. an x86 core?**
- **"GPU core"  $\approx$  one "vector slot" in SIMD vector instructions**
- **nVIDIA used to say 100x performance [over some old x86]**
- **Their claims today are more modest**
- **What about programmer productivity?**
- **Still: For some application areas, GPUs works really well**



# What matters for multicore performance?

- Are we buying...
  - ✱ CPU frequency?
  - ✱ Number of cores?
  - ✱ MIPS and FLOPS?
  - ✱ Memory bandwidth?
  - ✱ Cache capacity?
  - ✱ Memory capacity?
  - ✱ Performance/Watt?