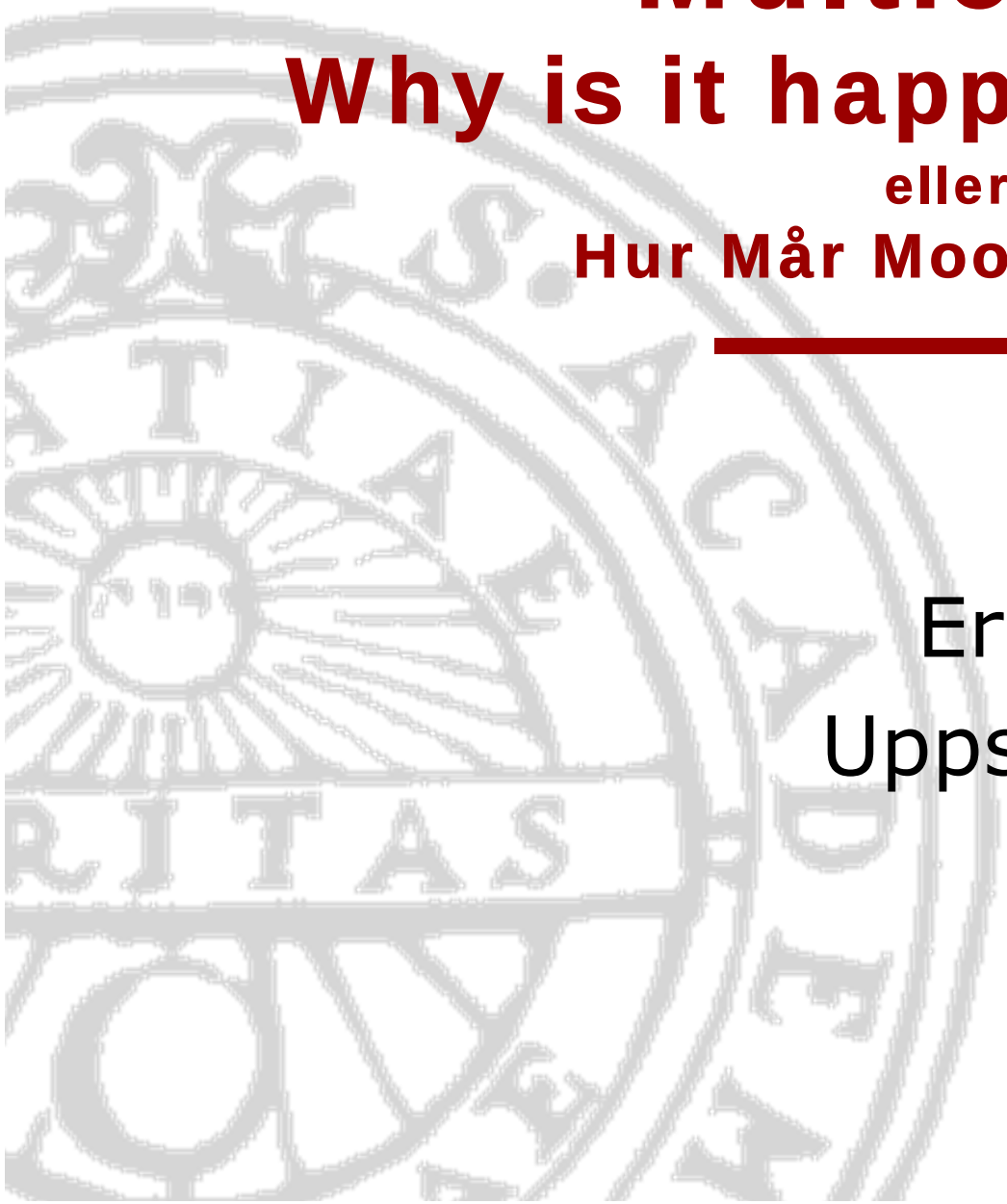


Multicore:
Why is it happening now?
eller
Hur Mår Moore's Lag?

Erik Hagersten
Uppsala Universitet

UPPSALA UNIVERSITET

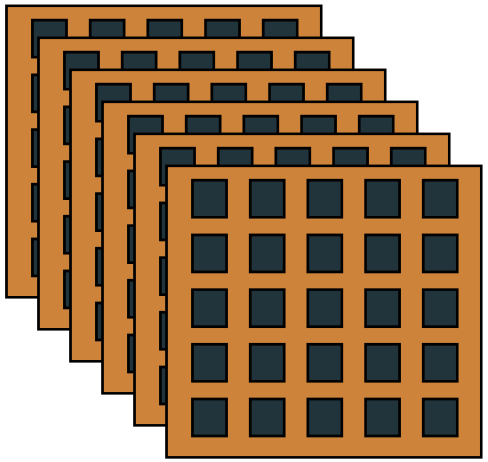
Outline of these lectures

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

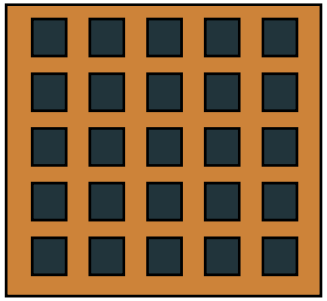
PDC Summer School 2014

Darling, I shrunk the computer

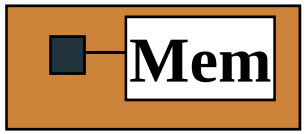
Mainframes



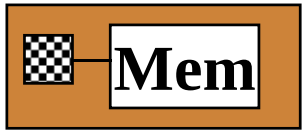
Super Minis:



Microprocessor:



Multicore: Many CPUs on a chip!



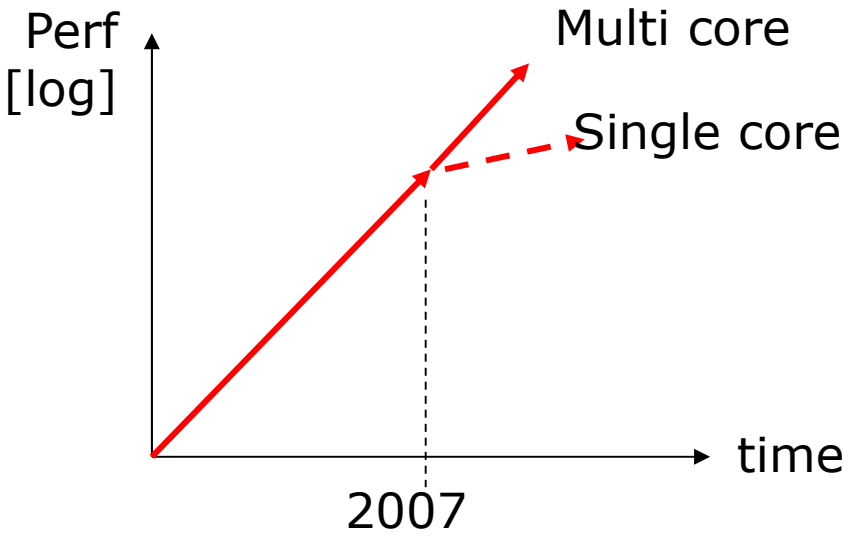
UPPSALA UNIVERSITET

Outline

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

PDC Summer School 2014

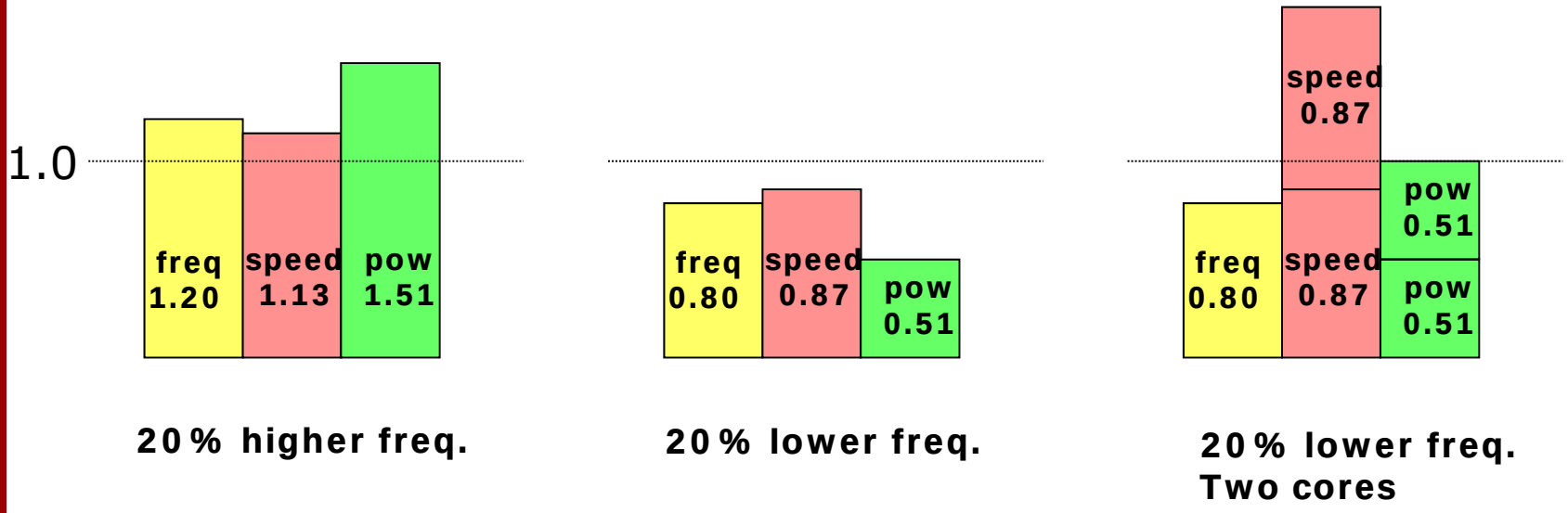
Everybody is doing it! But, why now?



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

Example: Freq. Scaling

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

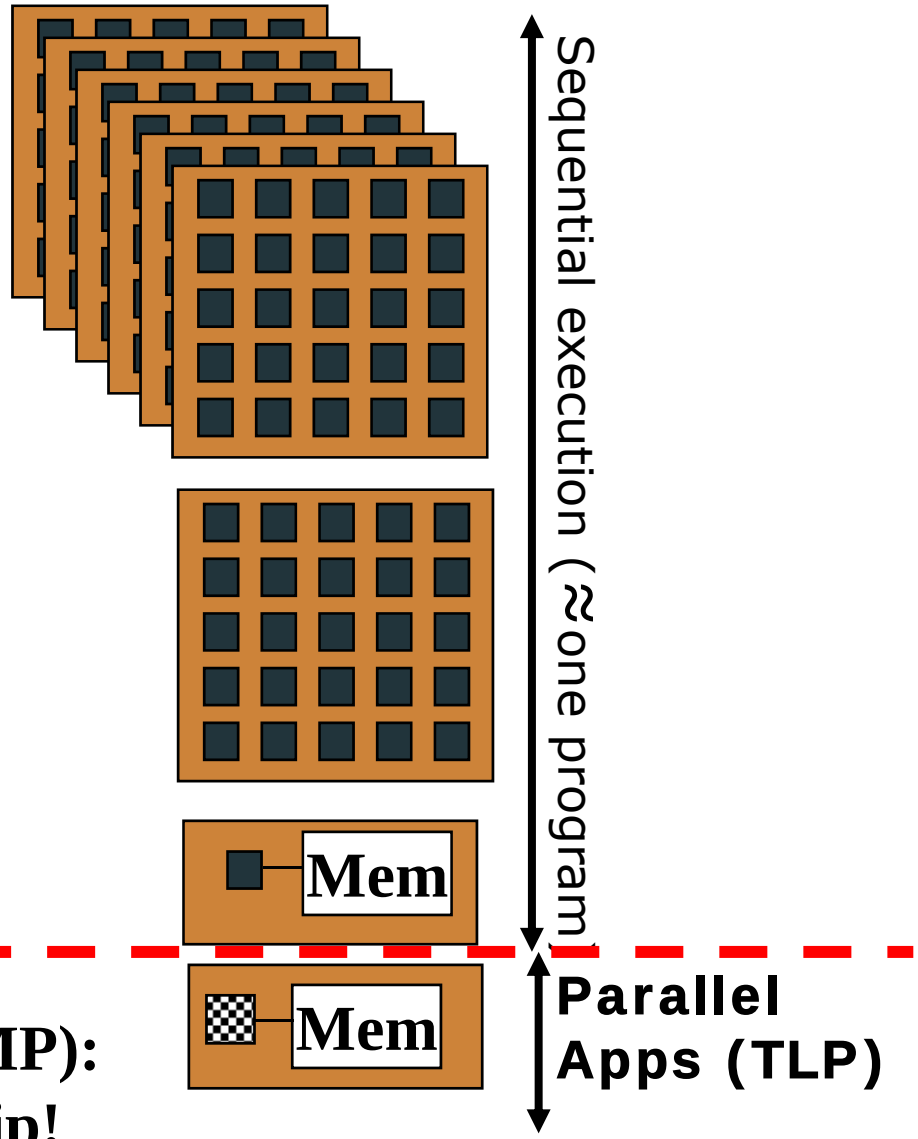


20% higher freq.

20% lower freq.

20% lower freq. Two cores

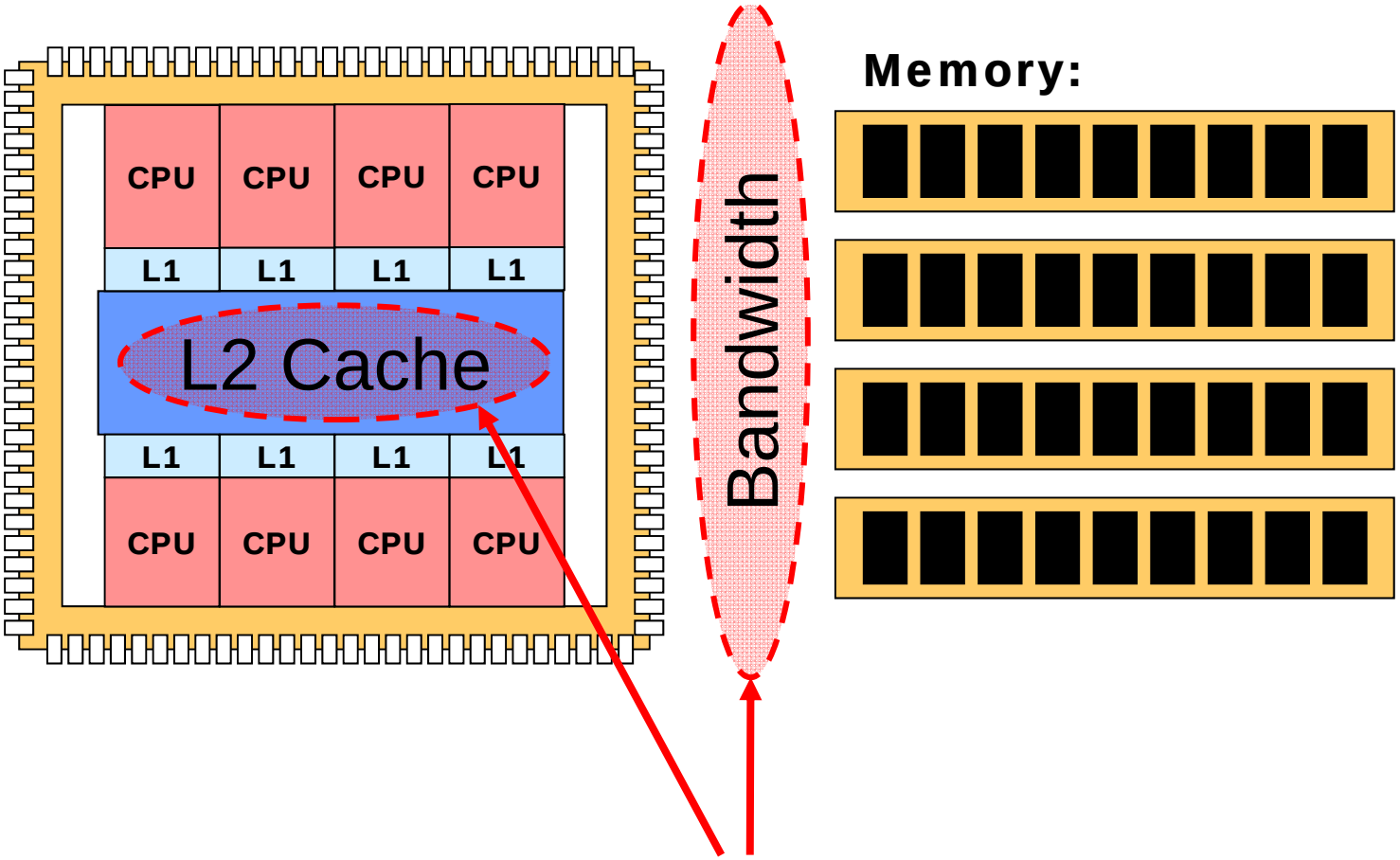
Darling, I shrunk the computer

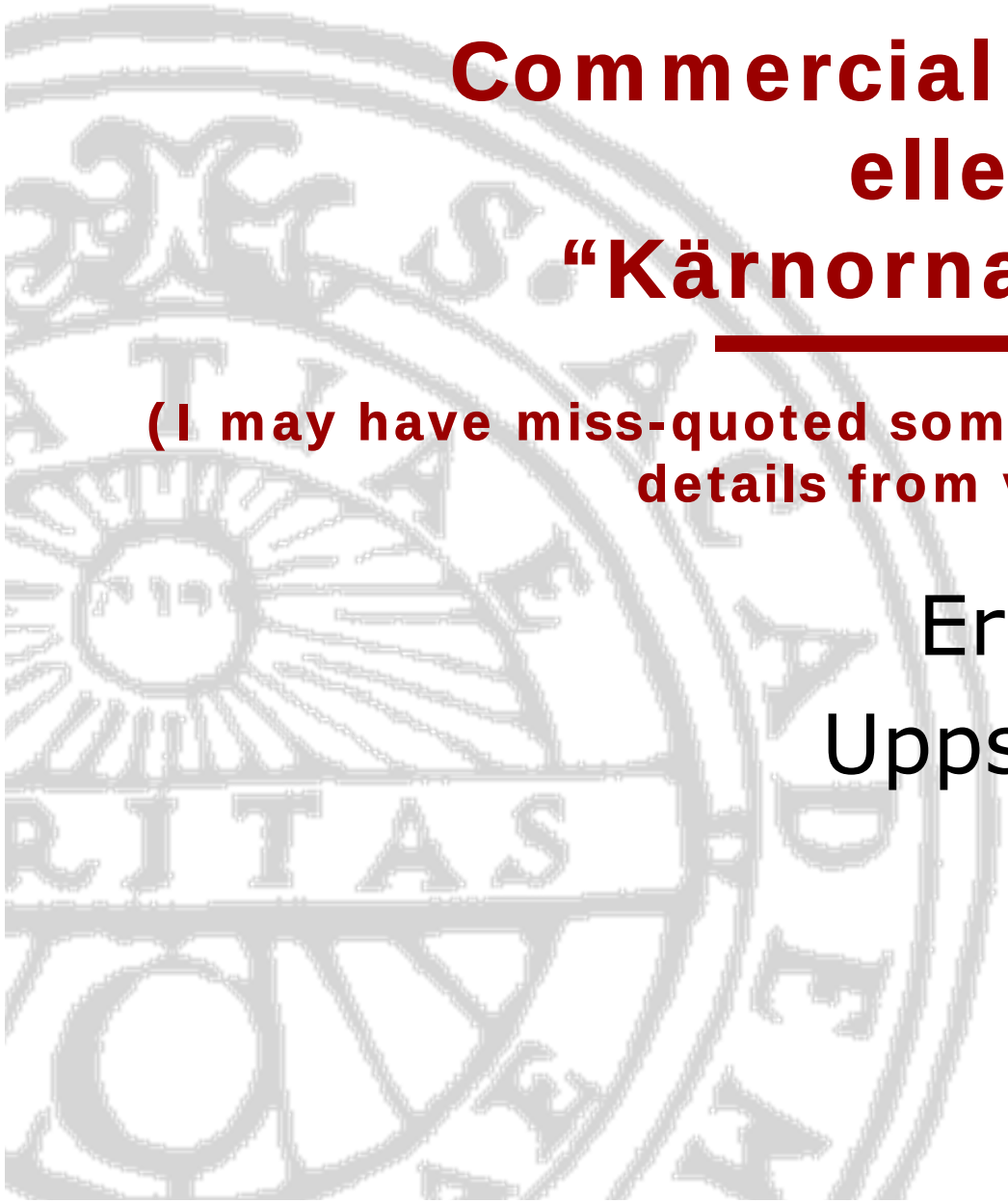


Paradigm Shift

Chip Multiprocessor (CMP): A multiprocessor on a chip!

Shared Resources





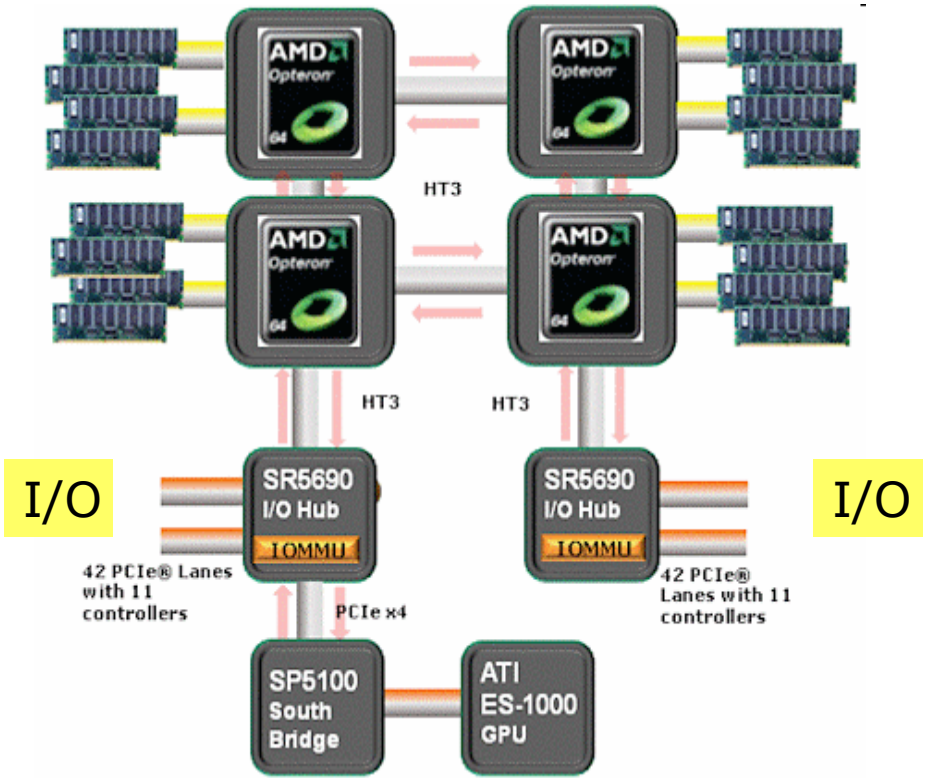
Commercial snapshot eller "Kärnornas krig"

(I may have miss-quoted some details, get architecture details from vendors)

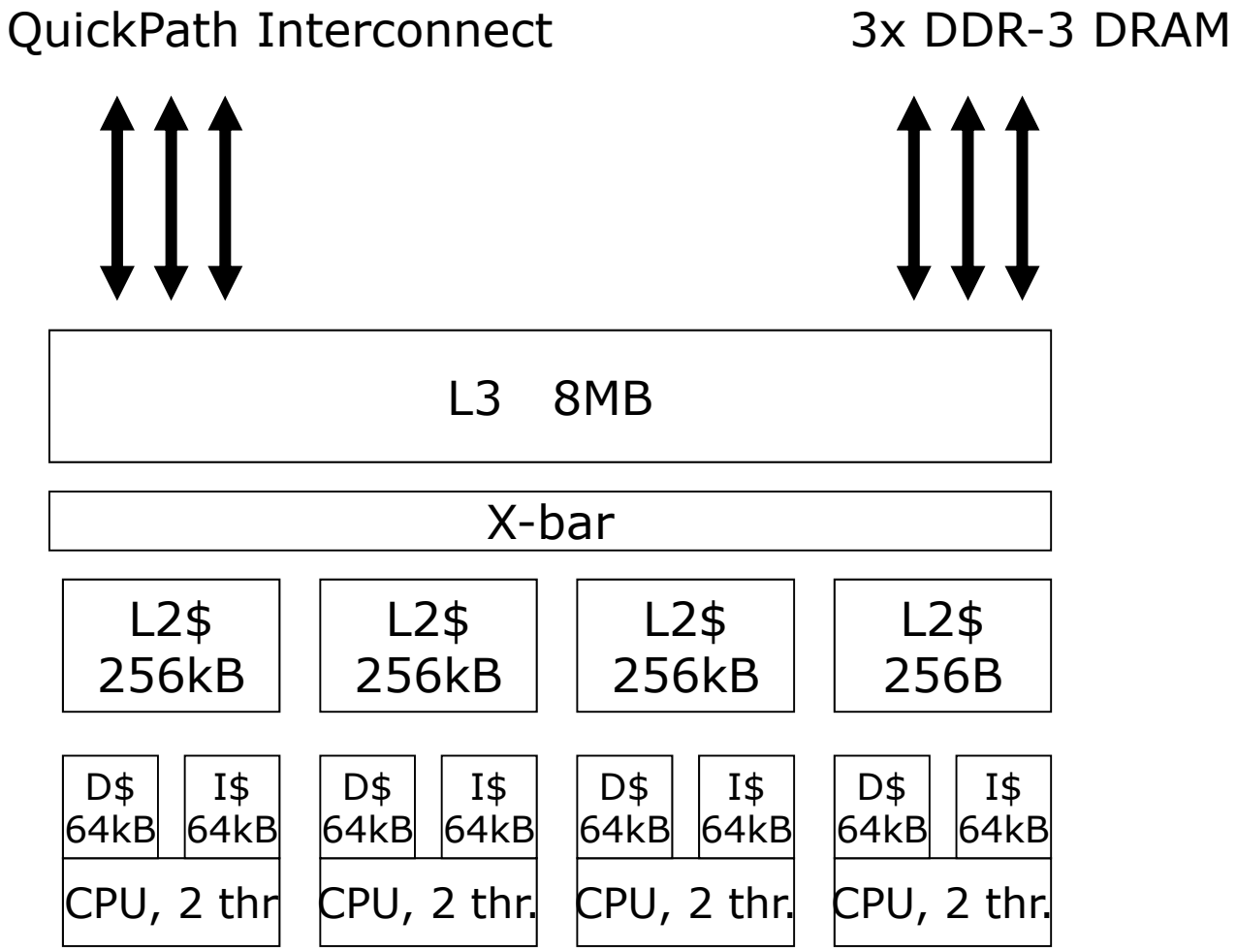
Erik Hagersten

Uppsala Universitet

AMD MC System Architecture



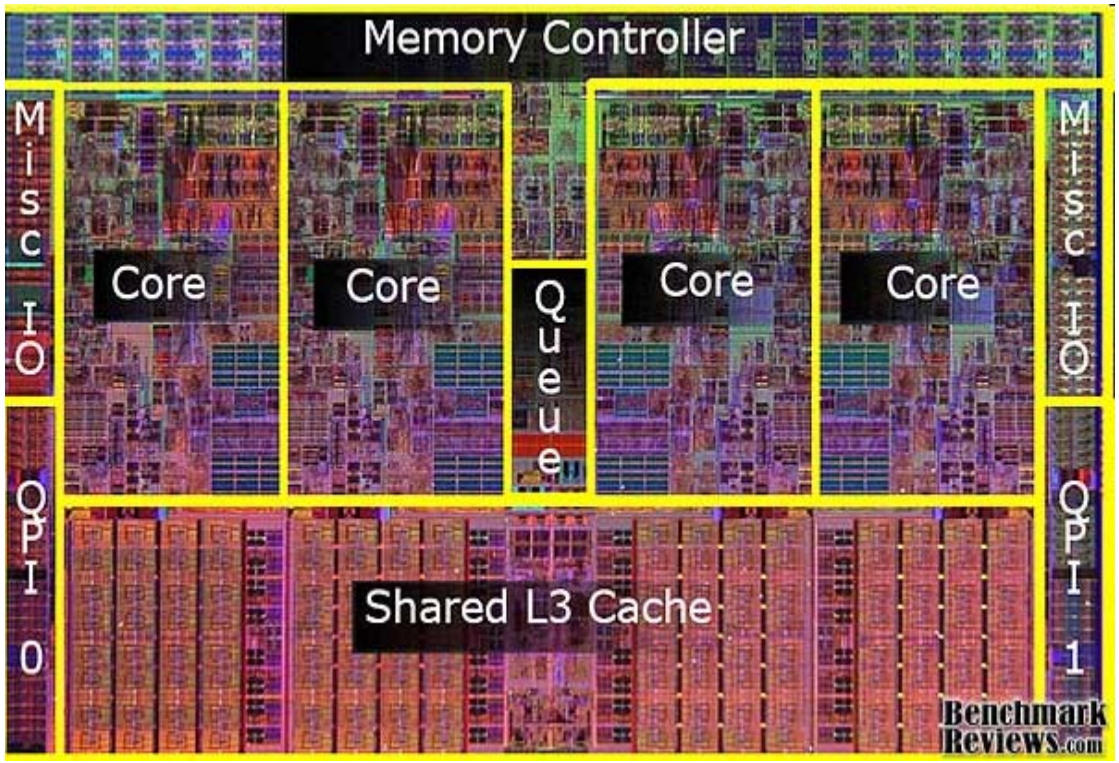
Intel: Nehalem, Core i7 Q1 2009 (4 cores)



Up to 4 cores x 2 threads

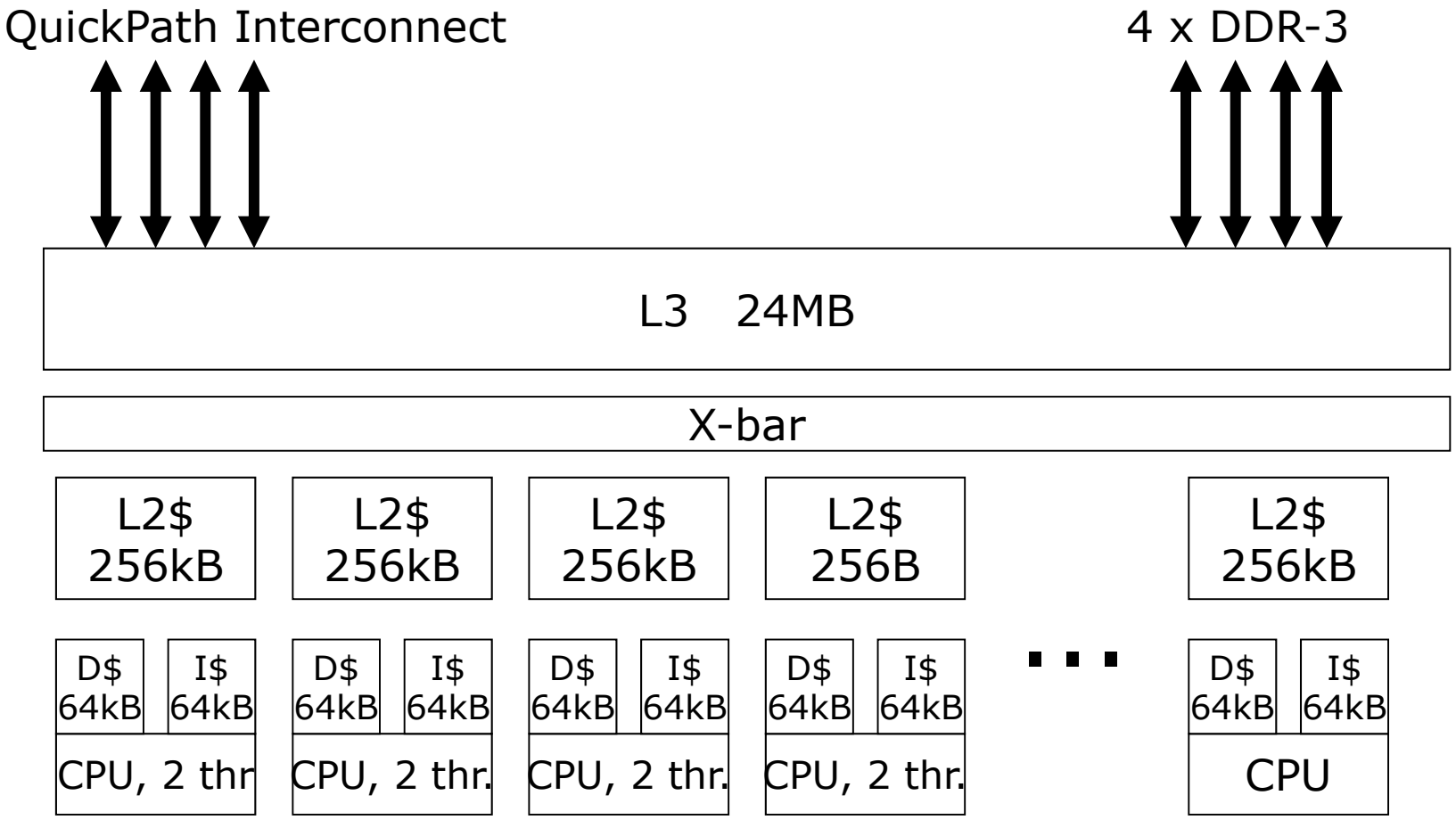
UPPSALA UNIVERSITET

Nehalem ”Core i7”



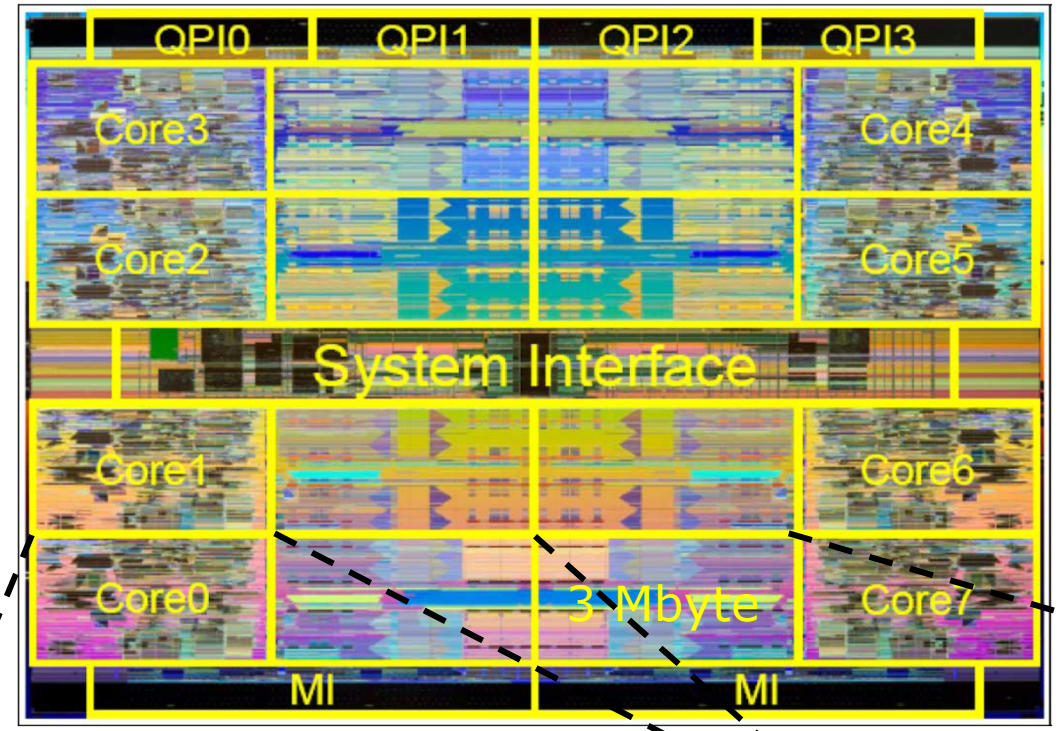
UPPSALA UNIVERSITET

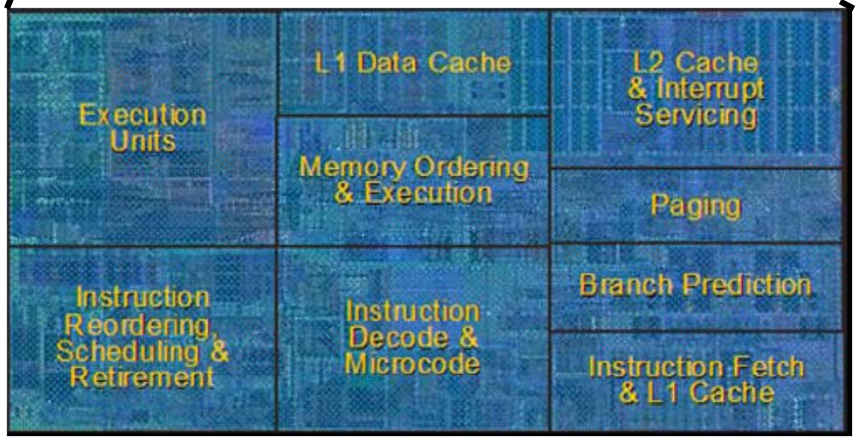
Intel: ”Nehalem-Ex” (i7)

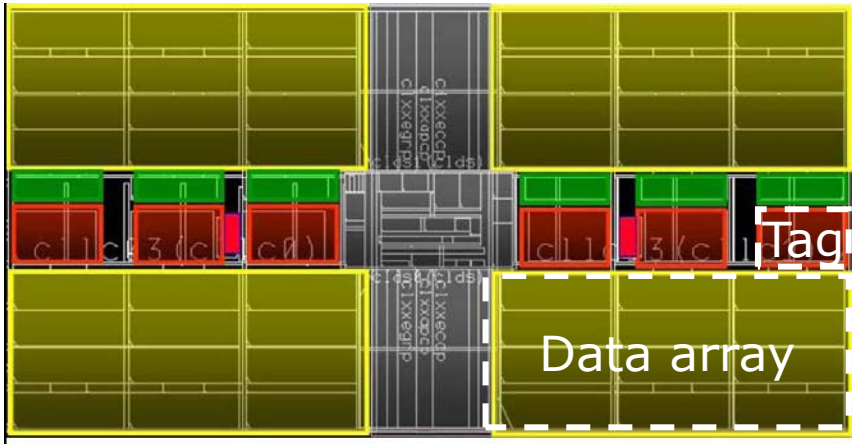


8 cores x 2 threads

How is the silicon used (i7-Ex)?





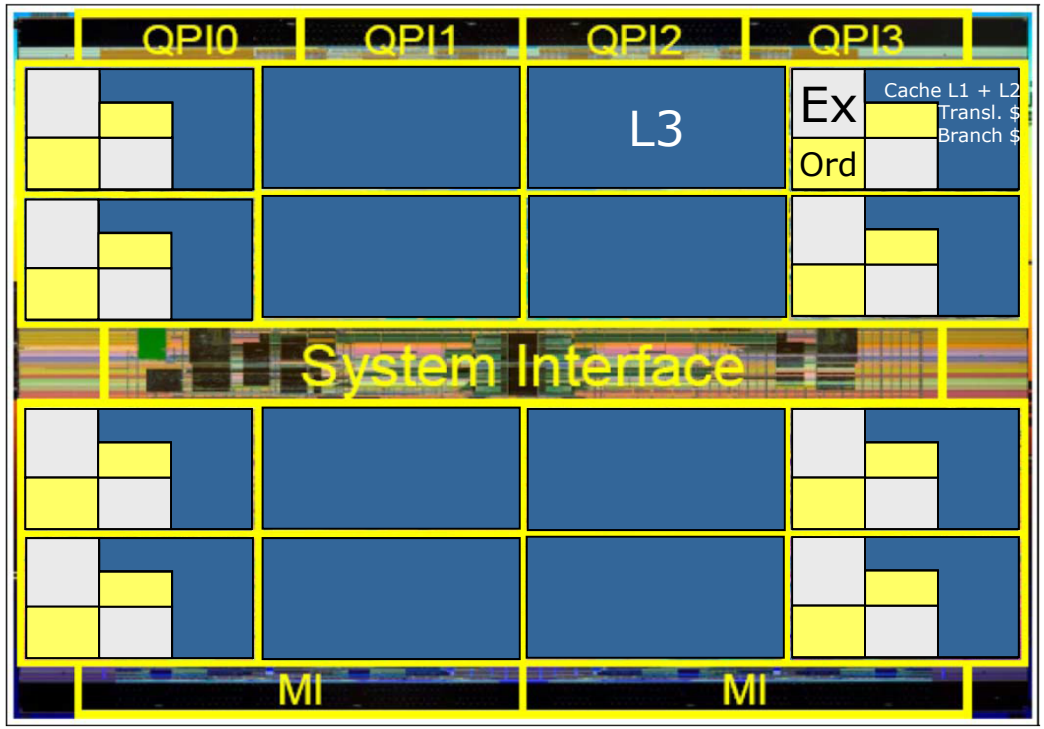


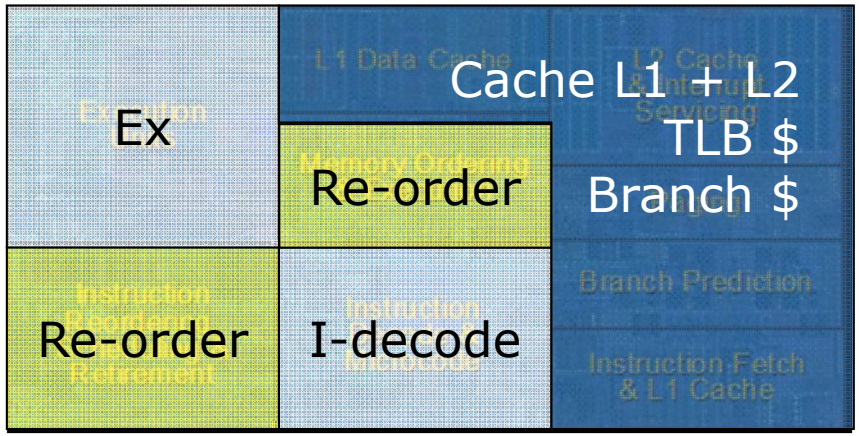
Source: JSSC Jan 2010, Rusu et. al

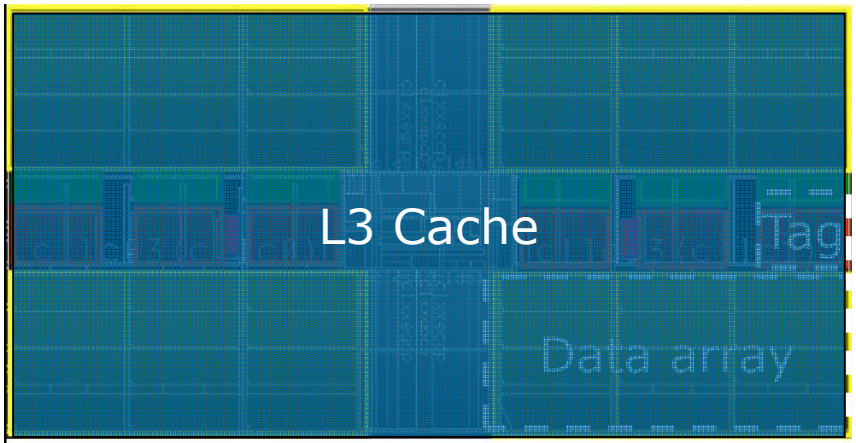
Multicore

UPPSALA UNIVERSITET

How is the silicon used?



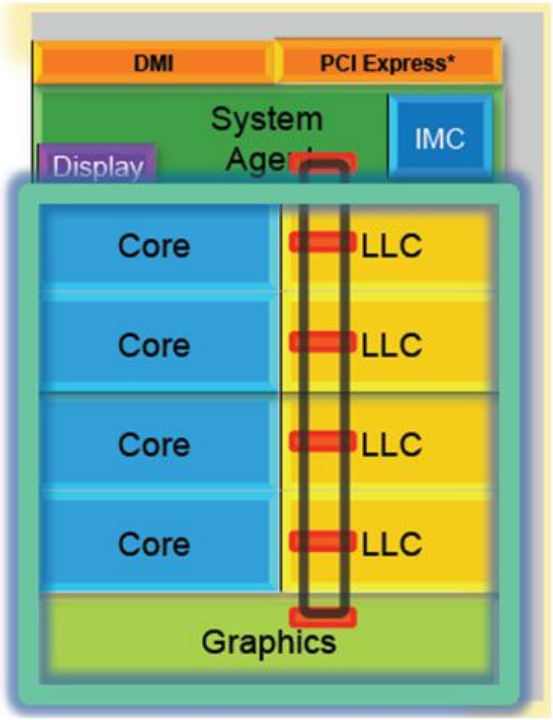




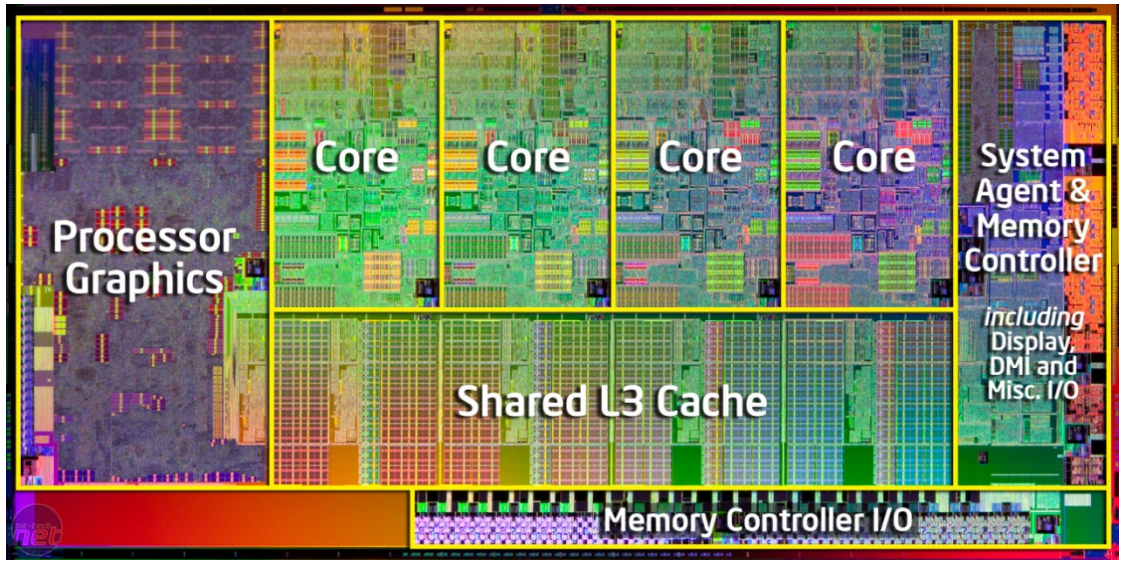
Multicore

Source: JSSC Jan 2010, Rusu et. al

More recent:
Sandy Bridge, 32 nm & Ivy Bridge, 22nm



- Integrated graphics!
- Ring bus for coherent LLC
- Up to 12 cores (24 threads)
- ~3GHz-ish



UPPSALA UNIVERSITET

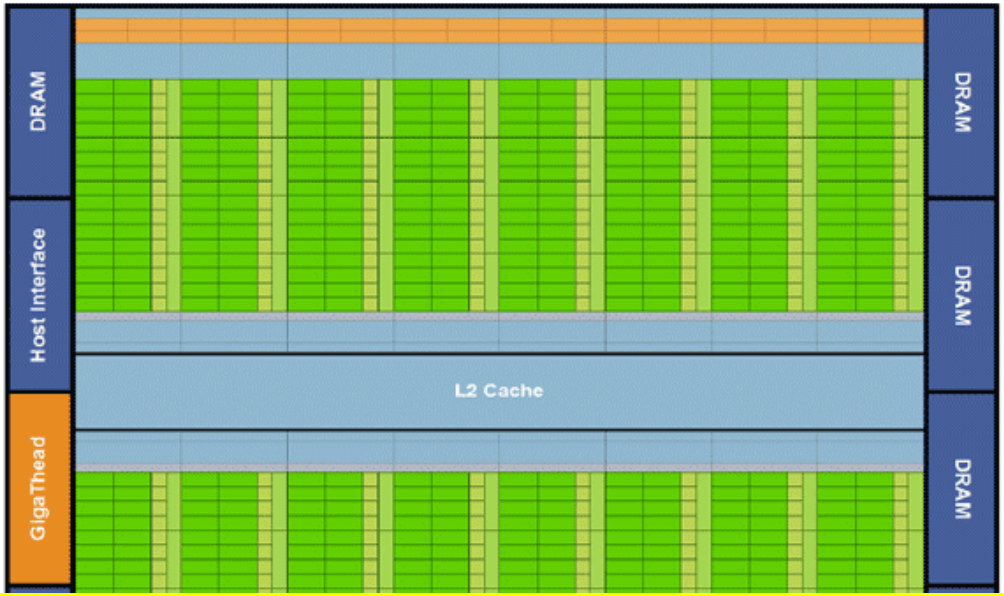
General Purpose:
Is there a demand for more cores?

- Intel released their quad core at Supercomputing 2006.
- 8 years later we are only(!!) at 12 cores and 24 threads.
- Core complexity and caches are favored over number of cores

Still, certain market segments and special applications need cores...

GPUs and Accelerators: Lots of hype in HPC!

Fermi from nVIDIA



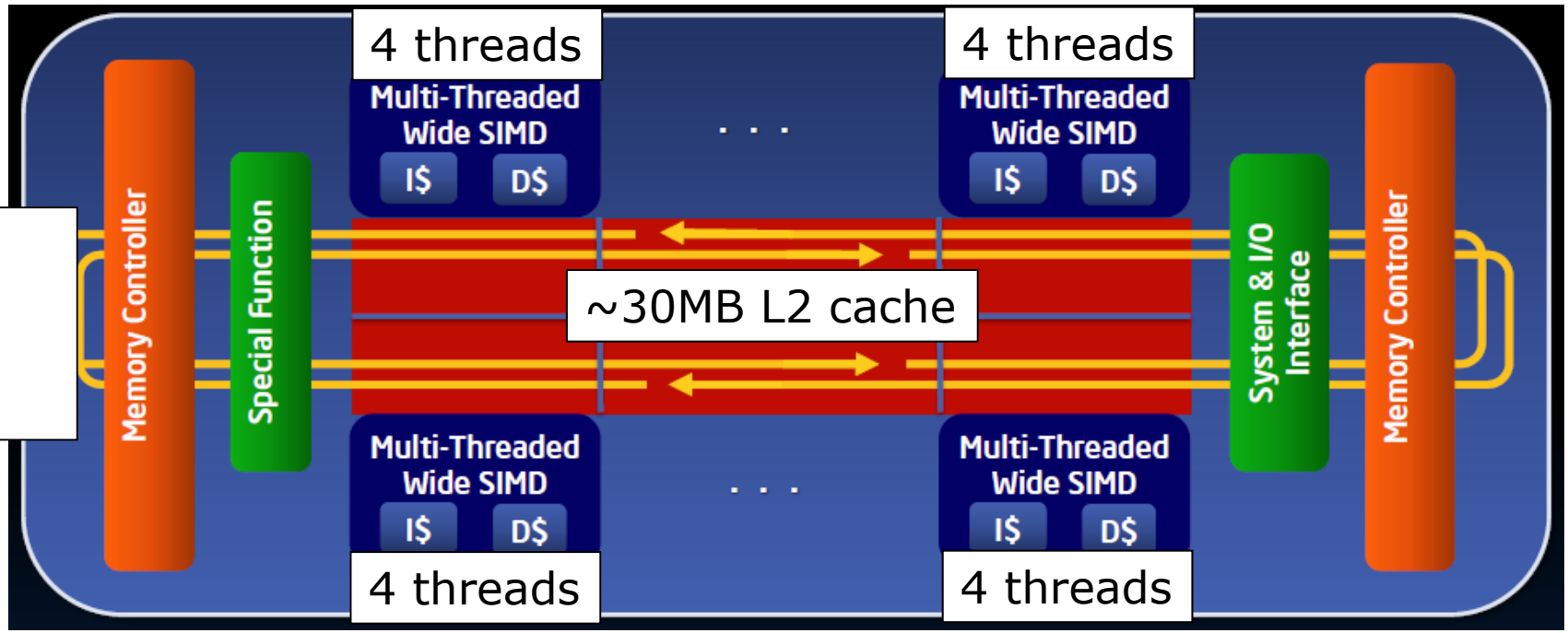
- 512 ”cores” (***P***)
- 16 ***P***/StreamProcessor (***SP***)
- Full DP-FP IEEE support
- 64kB L1 cache /***SP***
- 768kB global shared cache
- Atomic instructions
- ECC correction
- Debugging support
- …

- **CUDA language. Requires ”heroic programming” [M.Bull, EPCC]**
- **How much is an nVIDIA core worth vs. an x86 core?**
- **nVIDIA used to say 100x performance [over some old x86]**
- **Their claims today are more modest**
- **What about programmer productivity?**

David BS to tell you more…

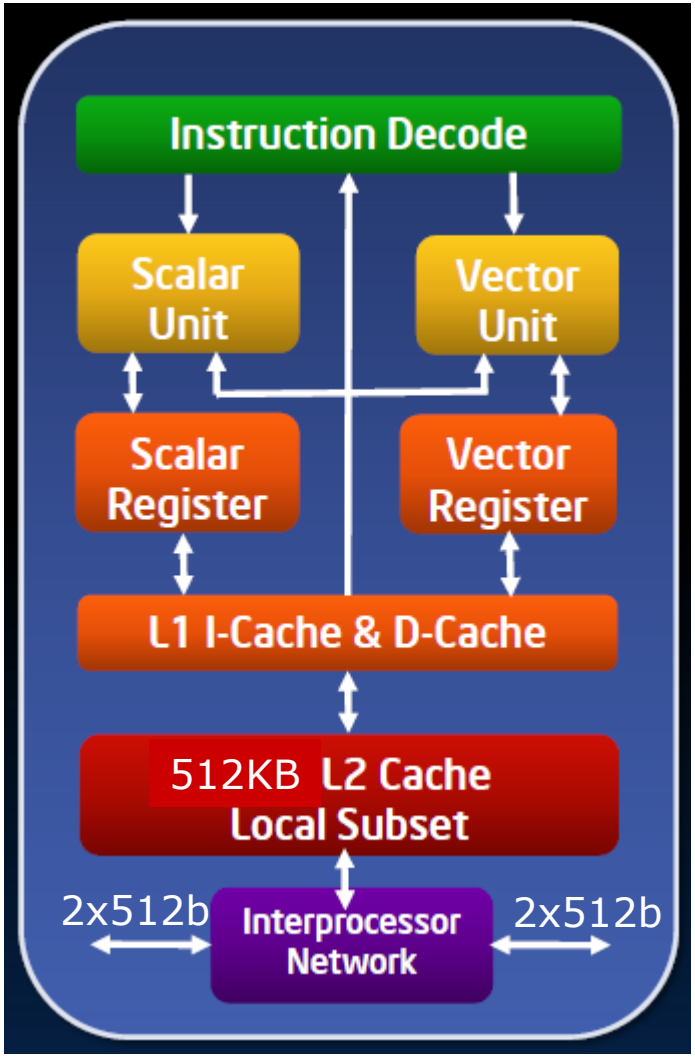
PD
Su
Sc
2014

Intel's MIC HPC architecture



- Knights Corner (Intel Xeon Phi), 22nm
- ~60 cores, X86 instructions, 4 threads
- Coherent caches. Runs Linux. "Just recompile and run"
- > 1TFLOPS (DP).
- 2% of the transistors devoted to x86 instructions

The MIC core "tile"



- 2x superscalar in-order
- 4 Threads SMT/core
- 512b SIMD instr. (16x SP)
- Scatter/gather operations
- 32k L1 I and D cache
- 512kB coherent L2
- 16 simple HW prefetch streams
- 2x512bit ringbus

- Next Gen: Out-of-order

UPPSALA UNIVERSITET

MIC Tool Stack

- It is ”just an x86 Linux box”
- Use your favorit language
- Intel compiler, tuning tool etc.
- 3rd party tools
- #1 on Top500 list (Tianhe-2, China)

- Question:
  - Can we afford 200+ threads/chip?
  - Is single-threaded performance important?

UPPSALA UNIVERSITET

What matters for multicore performance?

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