CERN openlab

A short introduction into

Hardware

... in general

... and what do we have at CERN

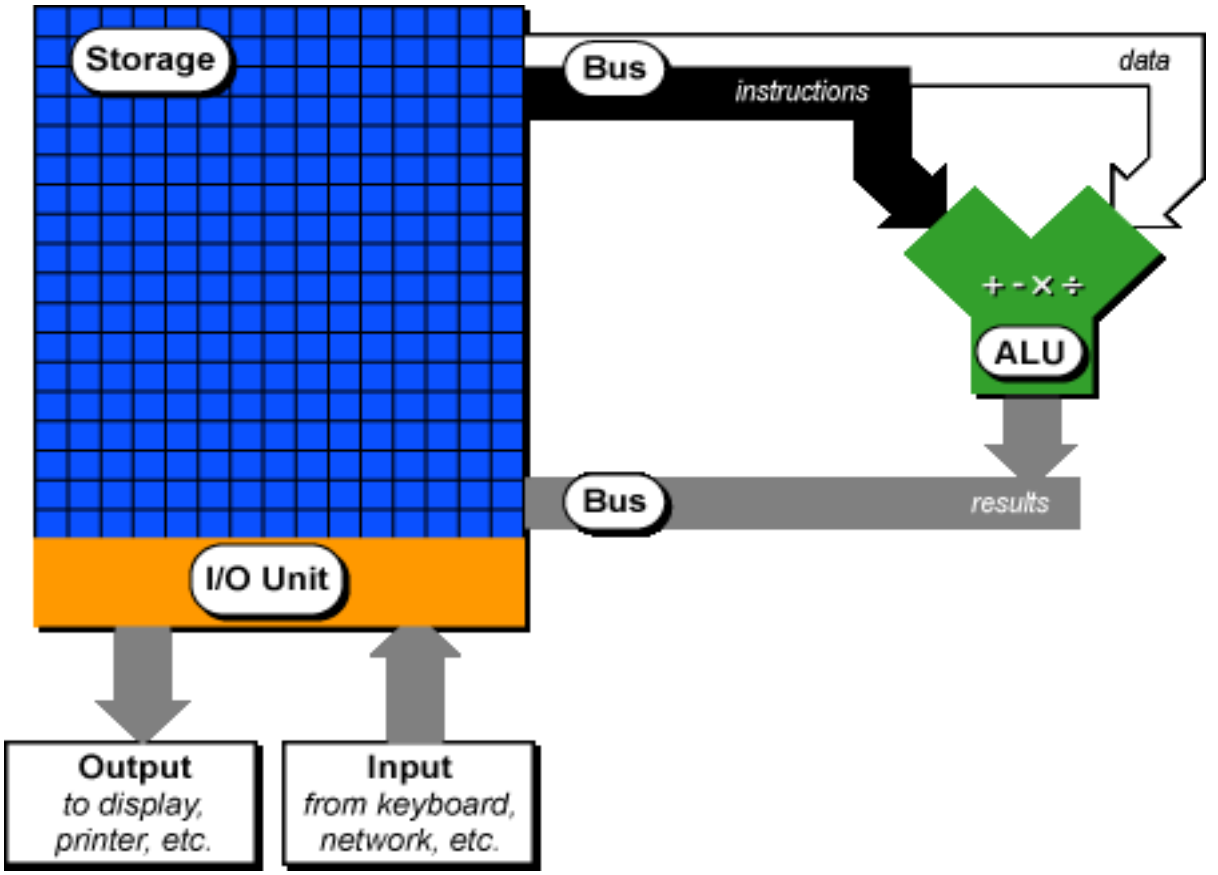
What you're about to hear

- **In general**
  - CPUs
  - Storage and I/O (i.e. network)
- **CERN Computer Center**
  - Batch systems
  - Disk server
  - Tape systems
  - Network
  - The rest

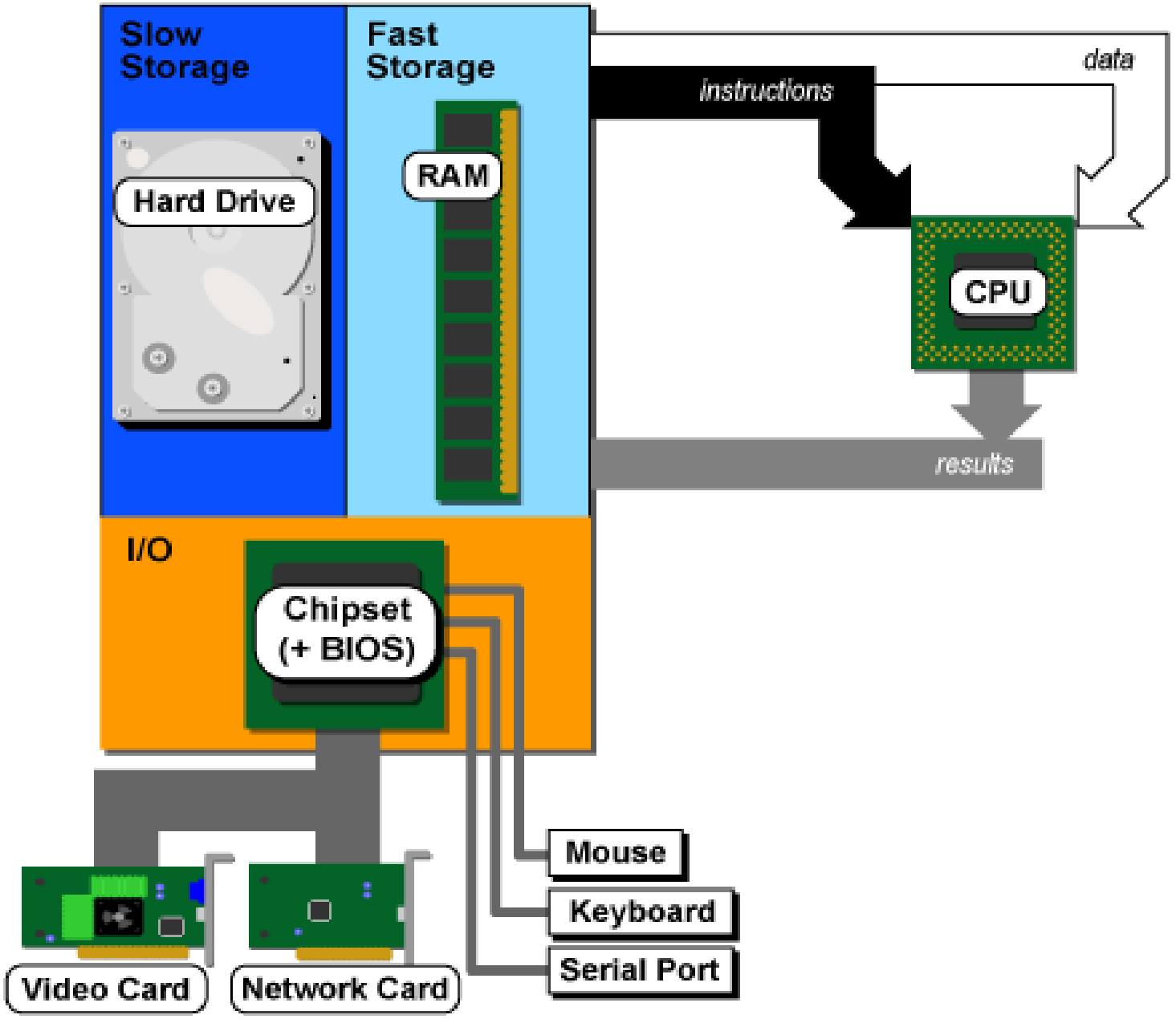
What's *Hardware* anyway ??

“If it hurts when it falls on your feet ... then it's hardware !”

How does a computer look?



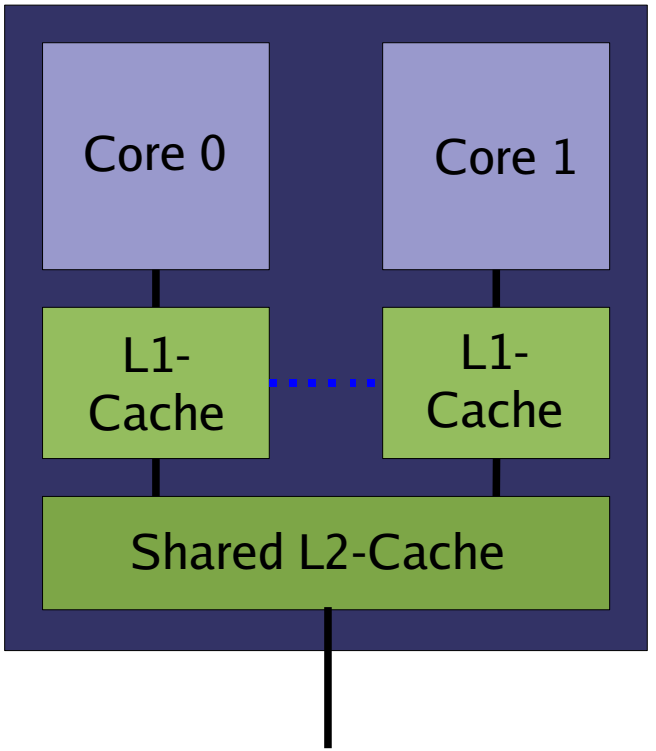
... or in a little more detail ...



The CPU – well Intel *"Core (2)"*

- Intels latest microarchitecture
- based on the "P6" and the Pentium-M (mobile) architectures
- Dual-Core design with shared L2- Cache
  - First incarnation: "Yonah/Sossaman"
    - 32-bit only
  - Just arrived: "Woodcrest/Conroe/Merom"
    - EM64T
    + other refinements of the microarchitecture

Intel "*Core*" ...

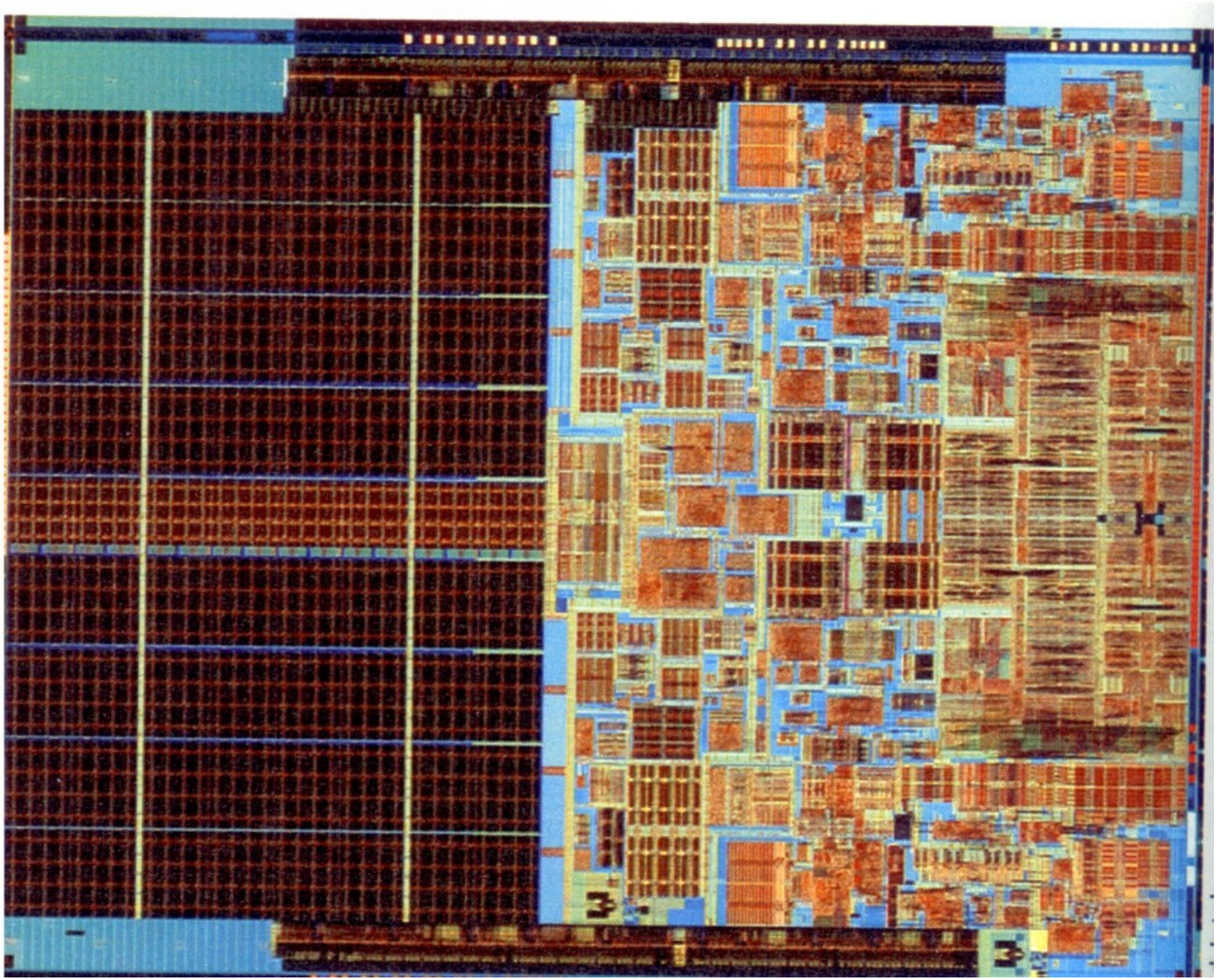


The general layout of a Woodcrest/Conroe/Merom "processor"

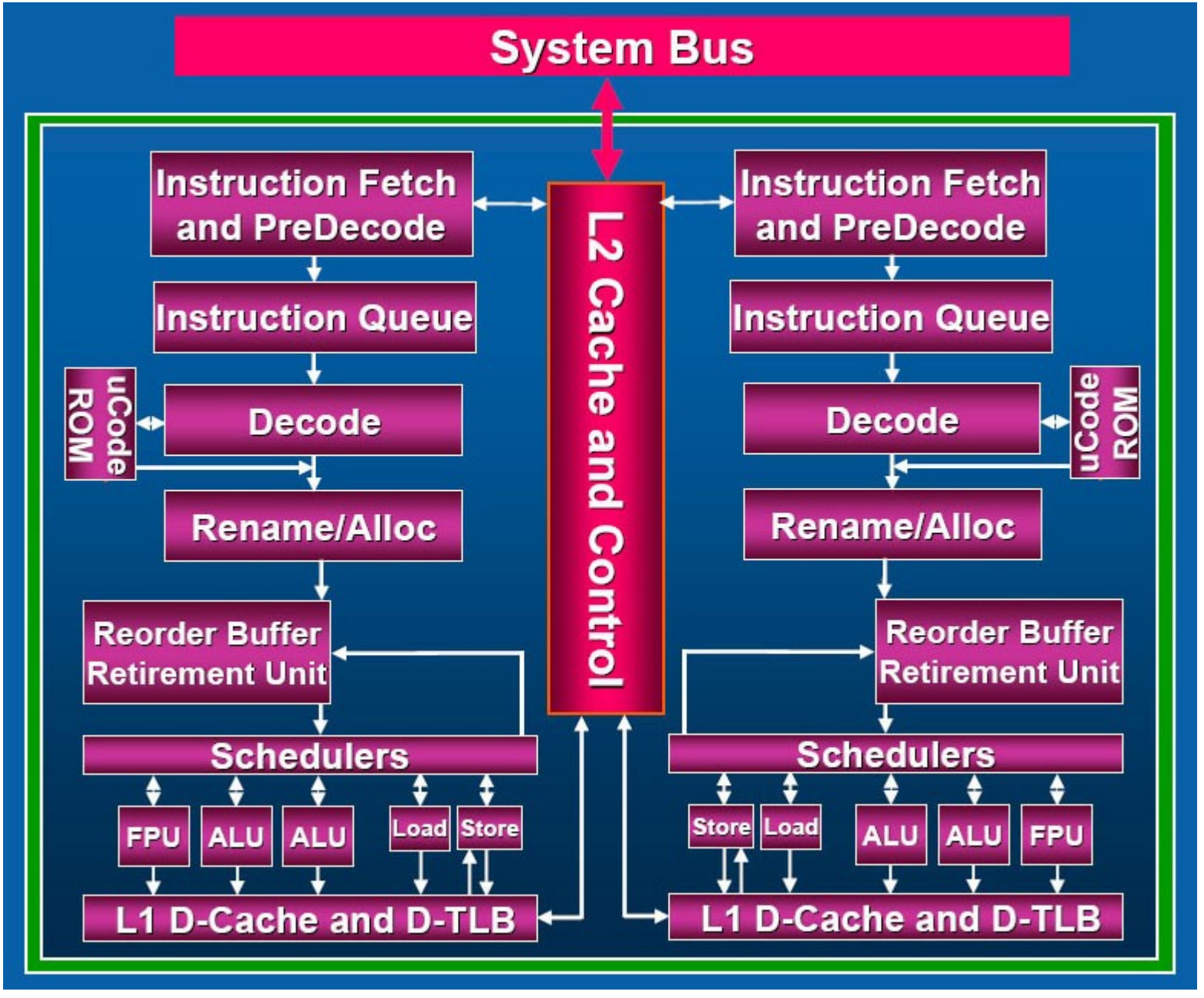
- The L2-Cache is smart enough to hold information which is used by both cores only once!

CERN openlab

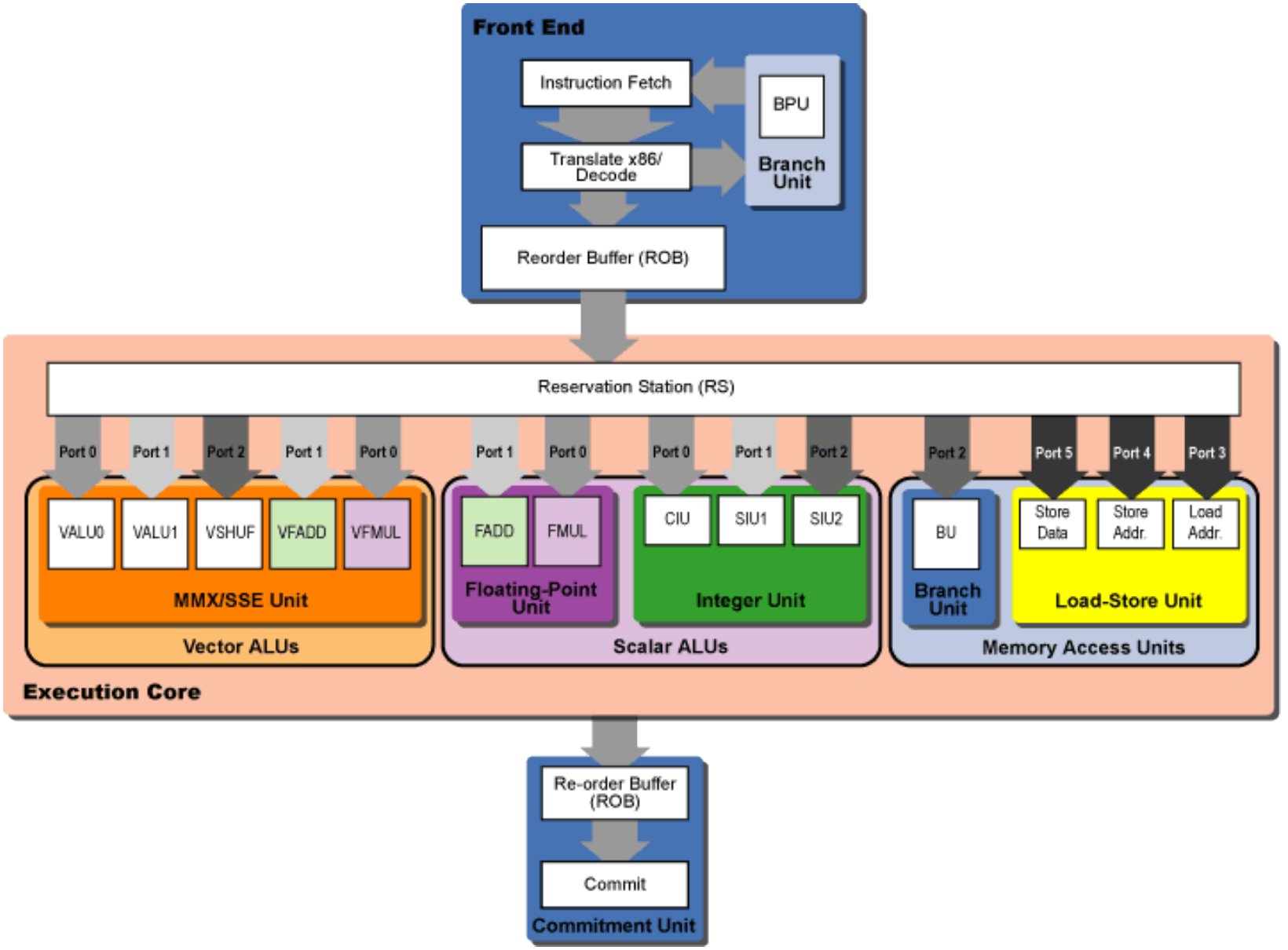
... with a picture ...



... a schematic view ...

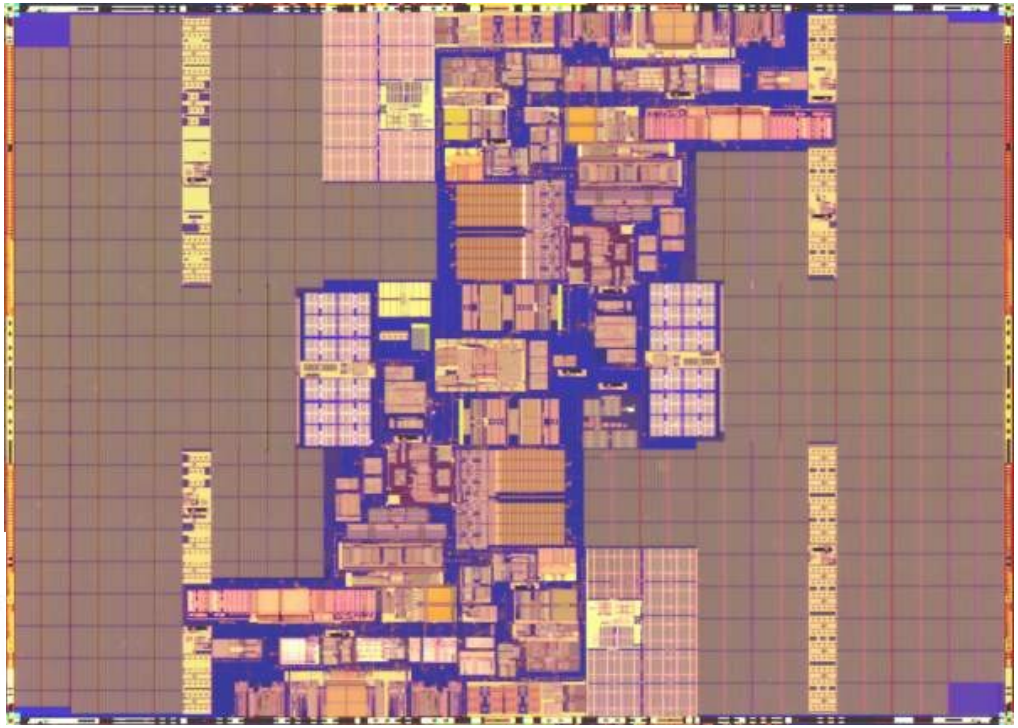


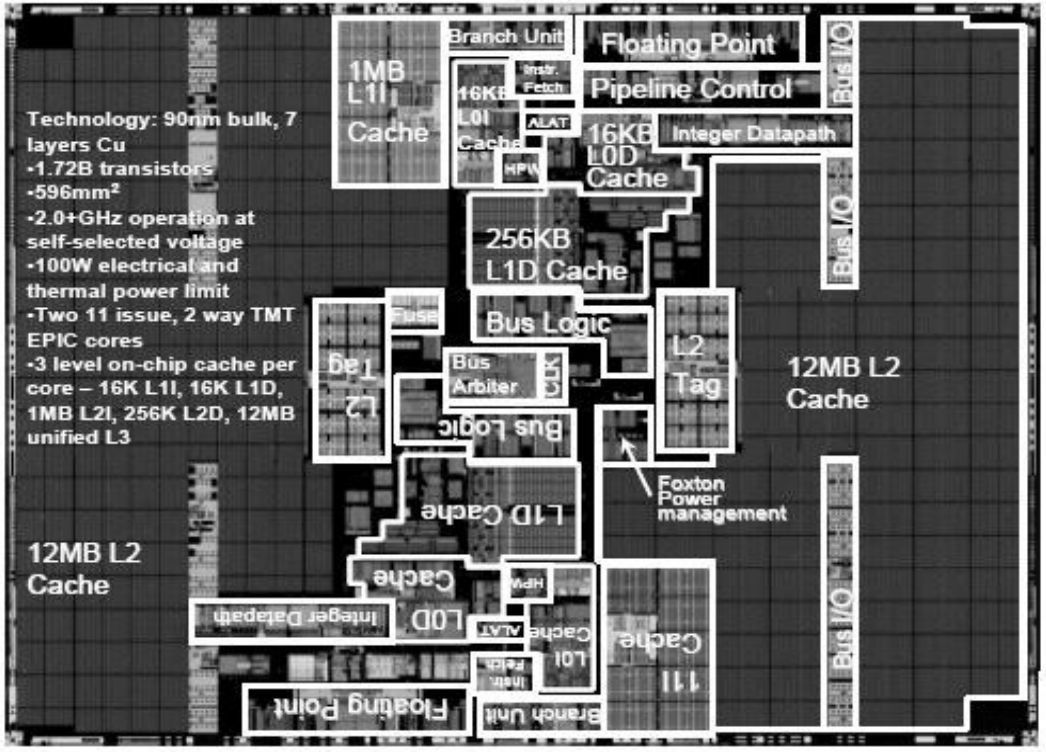
... and a deep look inside ...



Detour to Montecito ...

- Next Gen. Itanium processor
- Dual-core design
  - 1.72 billion transistors
    - ~57M for core logic
    - ~107M L1/L2 Caches
    - ~1550M L3 Cache
    - ~7M bus and I/O logic





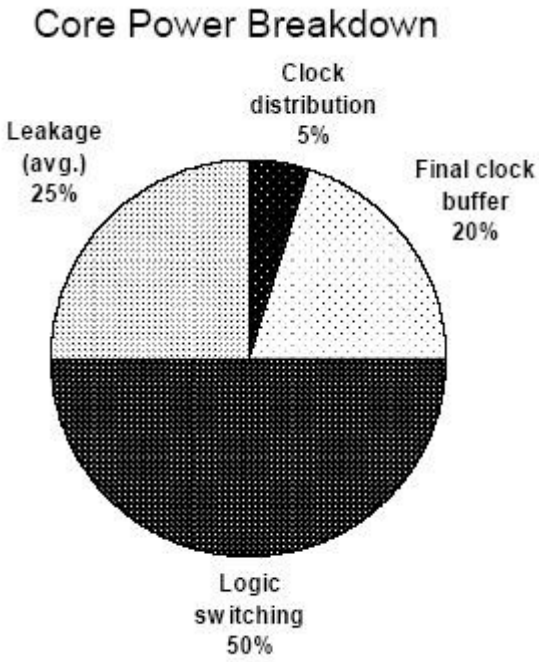
Remarks on multi-core CPUs

- The only way to have the per-socket performance keep increasing in the future (“Moore's Law”)
- The performance of a single core will not increase as much as it used to in the past :-(

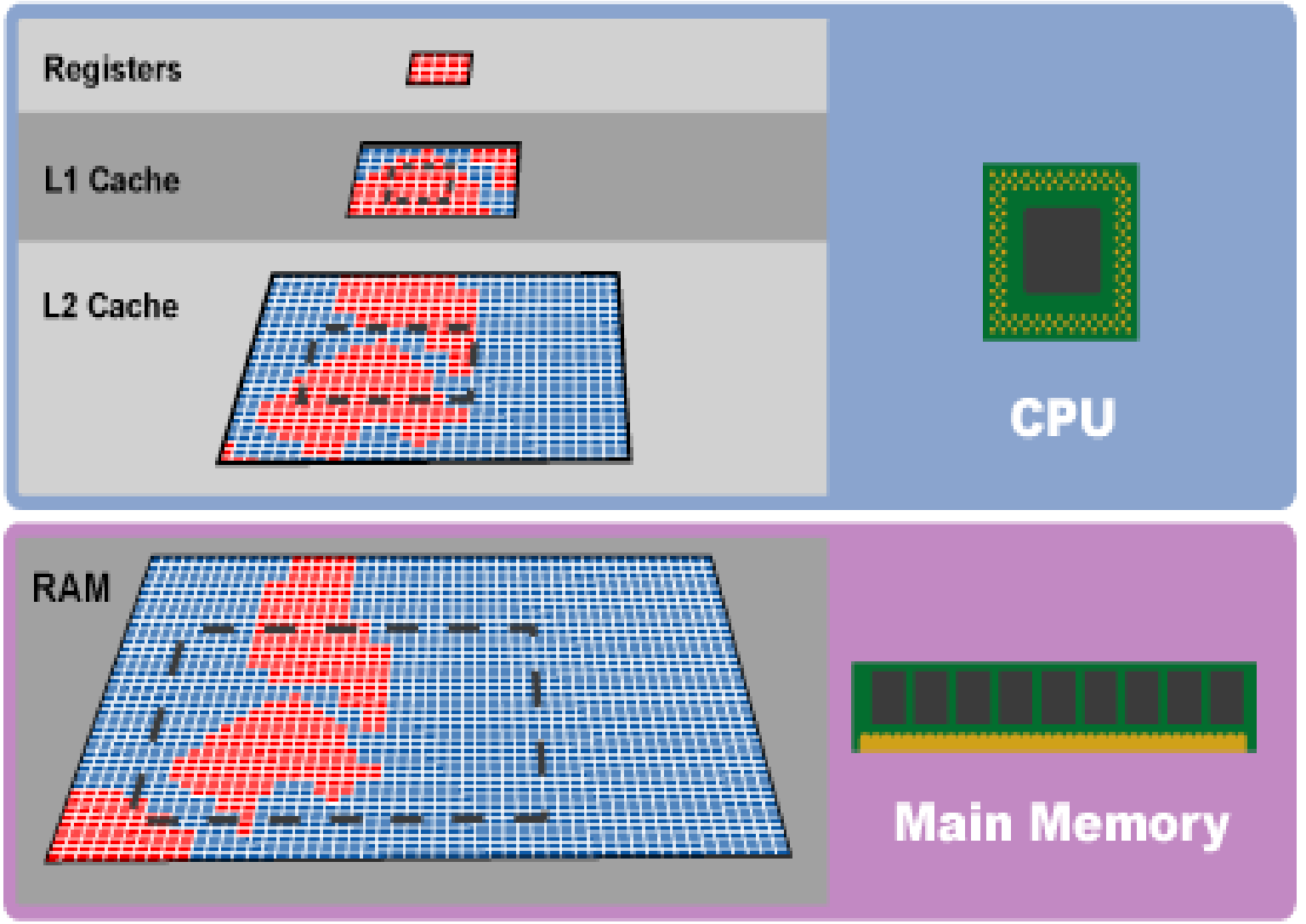
  - ➔ Performance gain mainly through multi-core
  - ➔ Serious implications on software design
    - ➔ keyword: Multi-threading
  - ➔ Very fast connection to main memory is crucial

Power Consumption

- under full load a CPU consumes between 65W and 130W
- The biggest issue with the now obsolete “Netburst” microarchitecture
- ~ 25% of the consumption is caused by leakage currents!!
- Power consumption of memory becomes important
  - ~10W per 1GB under load
  - in a Woodcrest system with 8GB RAM the memory consumes almost as much as the CPUs...



The memory hierarchy



Memory – some numbers

Memory Type	Access Time	Typical Size	Technology	Managed by
Registers	1 cycle	1kB	same as CPU	compiler
Level 1 Cache	2 – 4 cycles	8kB – 64kB	SRAM	hardware/compiler
Level 2 Cache	5 – 20 cycles	256kB – 4MB	SRAM	hardware/compiler
Level 3 Cache	15 – 50 cycles	0 – 24MB	SRAM	hardware/compiler
Main Memory	130 – 500 cycles	1GB – 64GB	DRAM	OS/user
Hard Disk	$10–30*10^6$ cycles	160 – 750GB	Magnetic	OS/user

- getting for data from main memory takes very long
  - ... and the CPU is sitting around just converting power to heat ...
  - try to “prefetch” data into the cache (usually L2-Cache)

If you start from a “worst case” scenario – always go to main memory – prefetching alone could speed up your application by a factor of 20 ... but then, this scenario never occurs these days

Storage and I/O

- Storage
  - Disk : up to several PB (PetaByte)
  - Tape : much more than disks (factor 10 - ...)
- I/O – concentrate on networking
  - Ethernet
    - LAN: 1Gb, 10Gb
    - WAN: 10Gb

Disk Storage

- Most common storage type (each PC has a disk...)
- basically two technologies
  - SATA (I/II) used in PCs and low end servers
    - 1 – 24 disks
    - usually 4-8 ports on motherboards
    - up to 24 ports on special add-on cards
  - SCSI/SAS used in high end servers
    - 1 – “take-your-favourite-number” disks

SATA: Serial Advanced Technology Attachment; SCSI: Small Computer System Interface; SAS: Serial Attached SCSI

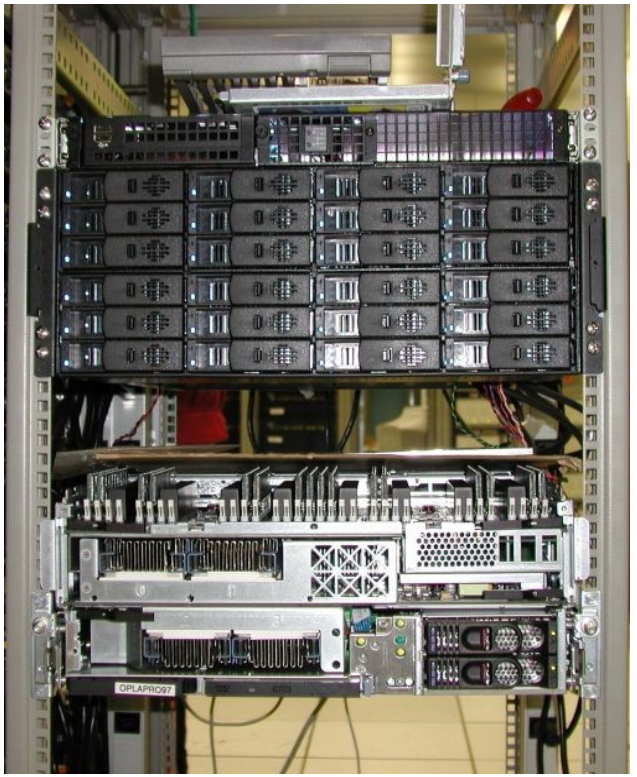
Disk Storage – II

- Disks are inherently “unsafe”
  - Failure rates are relatively high
  - Data recovery after a crash difficult
- ... so disks are organised in RAID systems
  - different RAID levels provides different levels of redundancy and performance.
  - RAID Level 0, 1, 5 or 6 are most commonly used
  - Combinations possibles, e.g. RAID 50
  - have a look at http://en.wikipedia.org/wiki/RAID for details

RAID: Redundant Array of Independent/Inexpensive Disks

Disk Storage – III

Typical Performance		
	single disk	**RAID**
single stream	~50MB/s	up to several GB/s
multi stream	down to a few 10kB/s	up to several GB/s



For example our “custom made” server

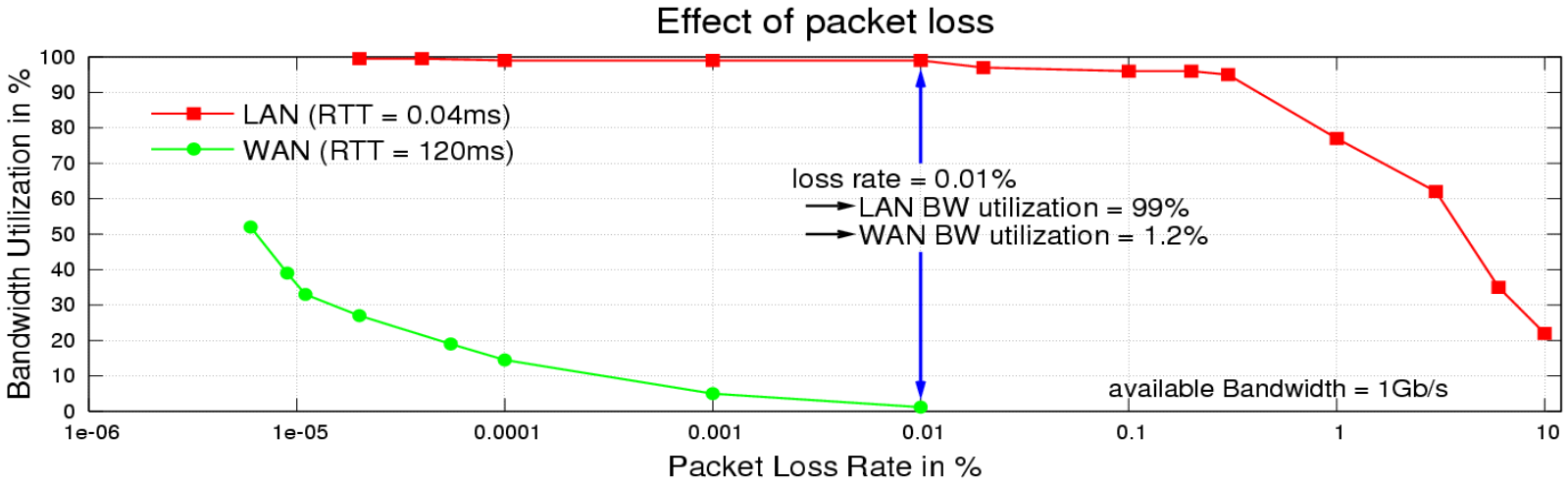
- single disk: ~50MB/s
- 24 disks in RAID0: ~1GB/s

Networking

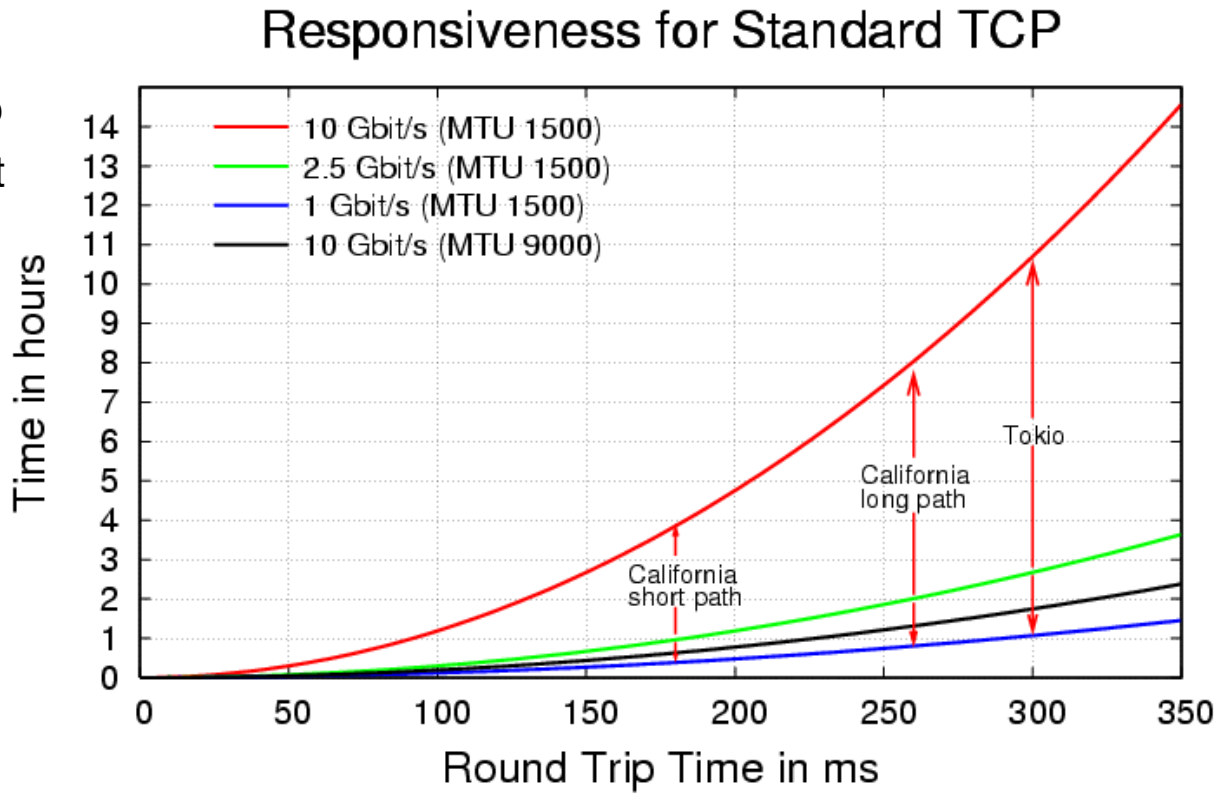
- Ethernet – based on TCP/IP
  - LAN – Local Area Network
    - 1Gb/s links to the hosts
    - 10Gb/s backbone infrastructure (+ some hosts)
  - WAN – Wide Area Network (≅ Internet)
    - 2.5Gb/s widely used
    - 10Gb/s going into production

For a more detailed look: my summer student lecture 2004

LAN vs. WAN



Responsiveness:
essentially the time to recover from a packet loss...



Further reading

- Quite a lot of in depth information can be found at arstechnica.com (in fact, I did “borrow” some of my graphics there)

  http://arstechnica.com/articles/paedia/cpu/core.ars

  http://arstechnica.com/articles/paedia/cpu/caching.ars

- A comparison of Intel and AMD processors

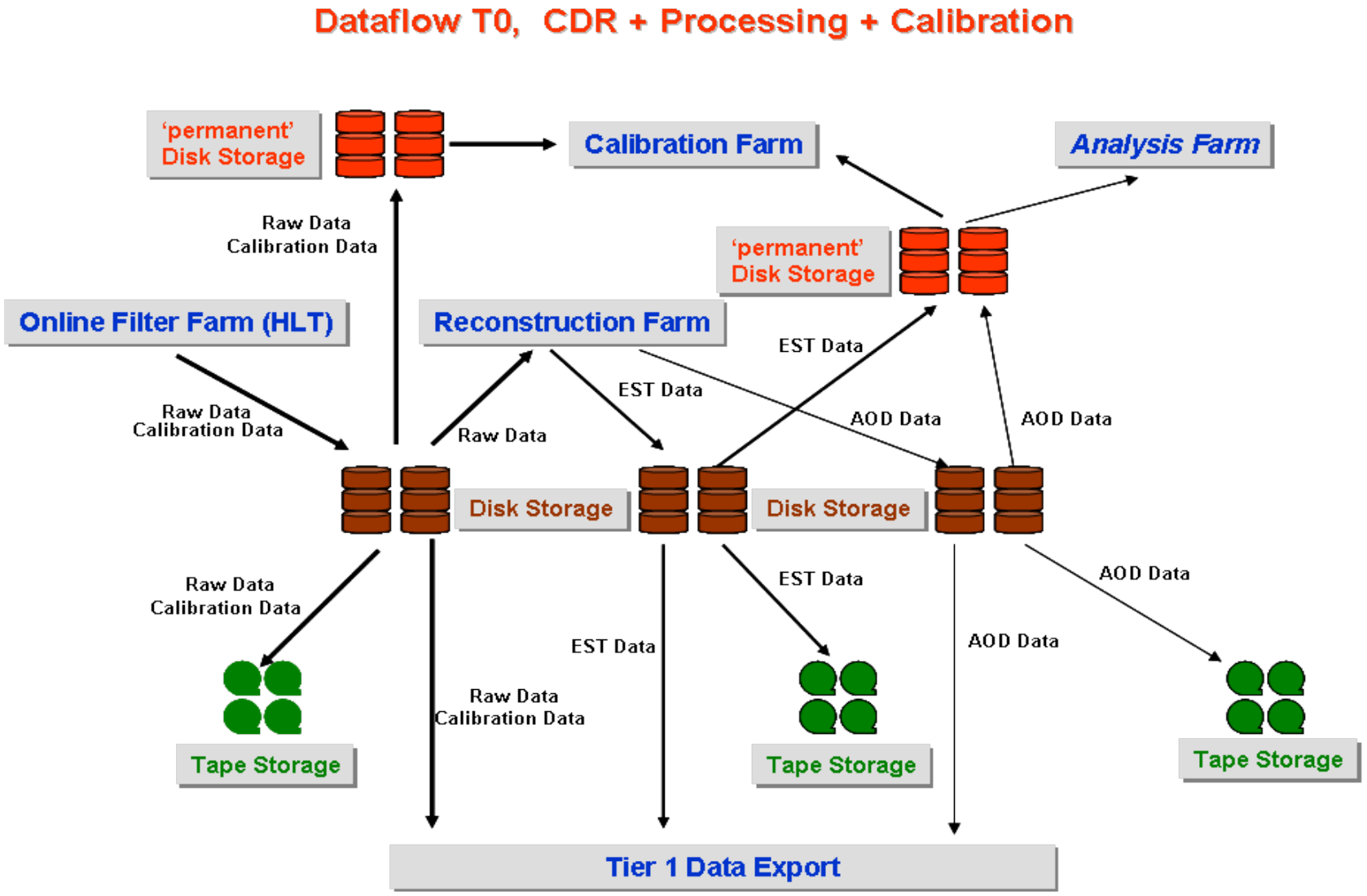
  http://www.anandtech.com/cpuchipsets/showdoc.aspx?i=2748&p=2

  If you want to know “Core (2)” in almost every detail:

  http://www.behardware.com/articles/623-1/intel-core-2-duo-test.html

What we have at CERN

... a lot of machines :-))



CERN CC in numbers

- Current Physics Computing
  - ~3000 Dual-CPU compute-nodes
  - ~1.6PB usable diskspace
  - 10 “old” tape-robots á 5000 tapes
    - new robots under test
- ... in 2007 – 2008
  - ~10000 Dual-Socket (?) compute-nodes
  - ~10PB usable diskspace
  - ~20 – 40PB tapespace
  - CPU and disk limited by the 2.5MW available for the CC

The CPU nodes

The diskserver

The “old” tape silos

The IBM tape silo

The new STK tape silo

The network setup