CERN, T0 and CAF resources

GDB, 10th October 2007

T0 == Central Data Recording, first pass processing

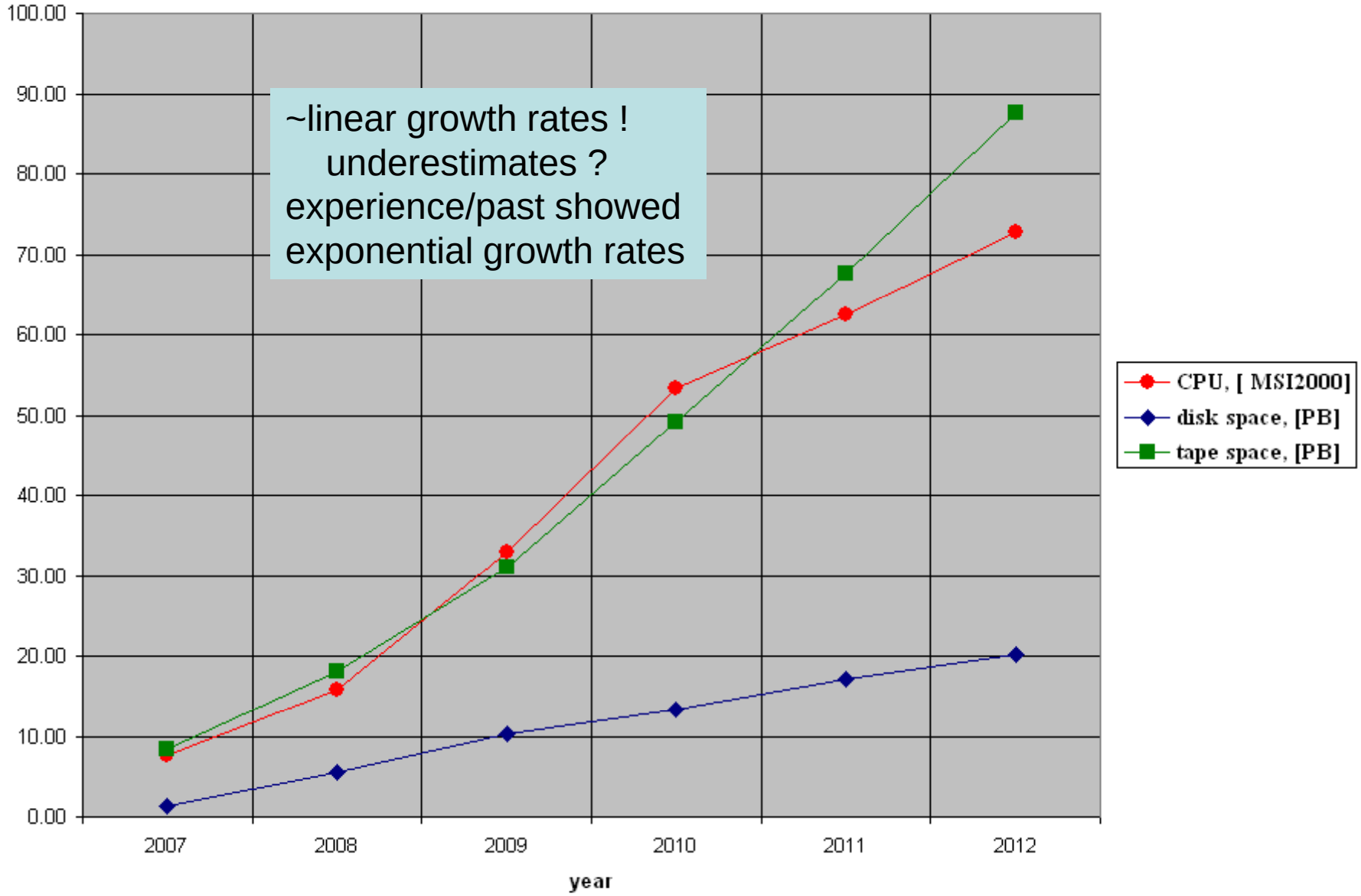
CAF == calibration and alignment, analysis, T1, T2, T3
means something different for each experiment

the distinction between he two is done in a logical manner, based on shares and priorities, sometimes via dedication

the hardware is the same for both

- CPU nodes for processing
- Disk server for storage
- Tape libraries, tape drives and tape server
- Service nodes (Grid services, Castor, Data bases, VO-boxes, experiment

Growth rates based on the latest experiment requirements at CERN

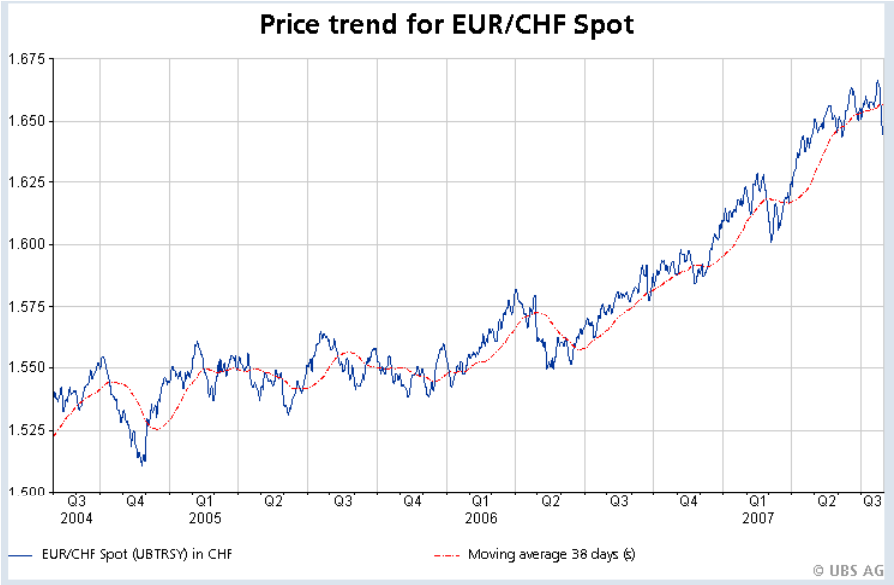


2008 resources

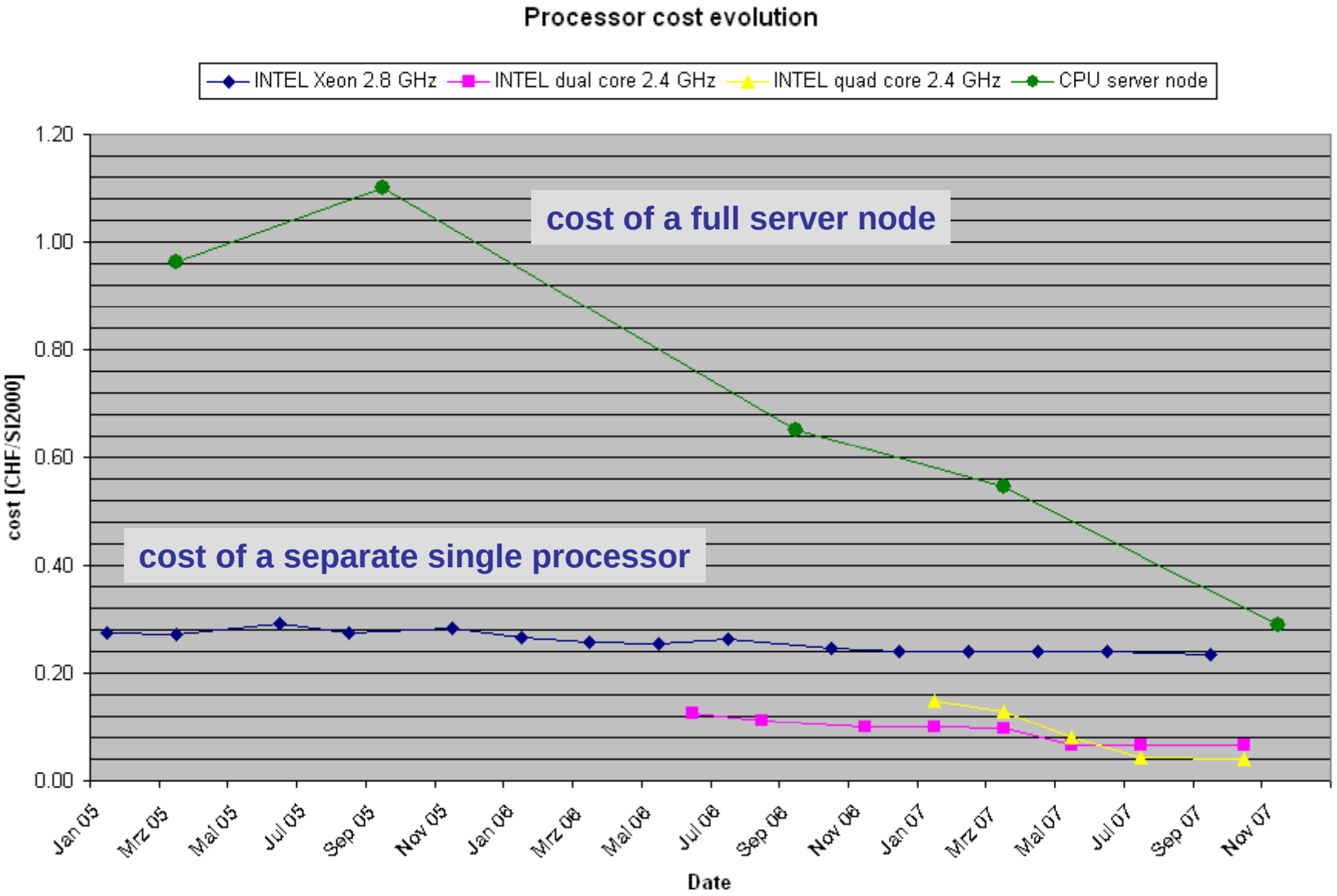
- 2008 purchases and tenders are already on the way
- we will receive in November 2007 and February 2008
  in total :
  ~ 1200 CPU nodes (~4 MCHF)
  ~ 700 Disk servers (~6 MCHF)
  ~ 350 service nodes (~3 MCHF)
  (don't underestimate the service investment!)
- the tape system is already sized for 2008 (~8 PB capacity)
- also planning for an upgrade of the internal network infrastructure

Input parameters for the following pages

- the basic cost of equipment is taken from a European price comparison web side
- the full server costs are taken from our tenders
- our budget is in CHF, but most costs are normally quoted in EURO
- the transformation into CHF takes into account a reduction of 20% (VAT) and an exchange rate factor of 1.55 – 1.65



Processors I



Processors II

- less than 40% of a node costs are the processors
  33% CPU, 36% memory, 31% chassis,motherboard,power-supply,2 disks
- 2007 was a special year , heavy price war between INTEL and AMD,
  INTEL pushing their quad-cores (even competing with their own products)
- new trend → dual motherboard per 1u unit, very good power supply
  power efficiencies as good as for blades
- our purchases will consist out of these nodes with a possibility of getting also blades

Processors III

Technology trends :

- INTEL aims to have a two year cycle
  architecture improvements and structure reduction
  45nm products already announced by INTEL, 32nm in 2009

- multi-core 2 3 4 6 8……100

BUT

what to do with the expected billion transistors and multi-cores ?

the market is not clear → wide spread activities of INTEL e.g.

-- initiatives to get multithreading into the software, quite some time away, complicated
especially debugging (we have a hard time to get our 'simple' programs to work)

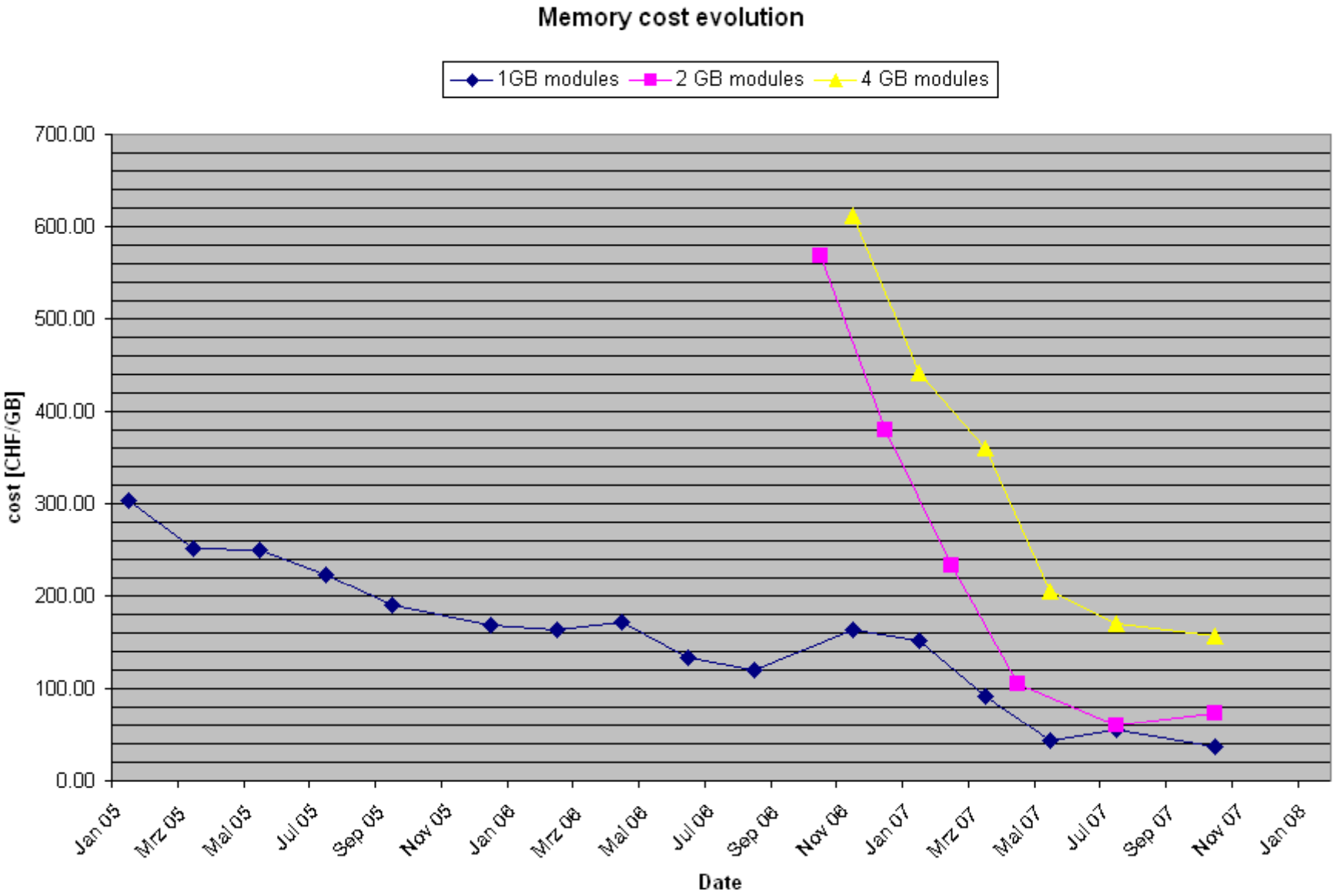
-- co-processors (audio, video, physics engine for games, etc.)

-- merge of CPU and GPU (graphics) , AMD + ATI→ combined processors,
NVIDIA → use GPU as processor, INTEL → move graphics to the cores

-- on the fly re-programmable cores (FPGA like)

not clear where we are going
specialized hardware in the consumer area → change of price structure for us

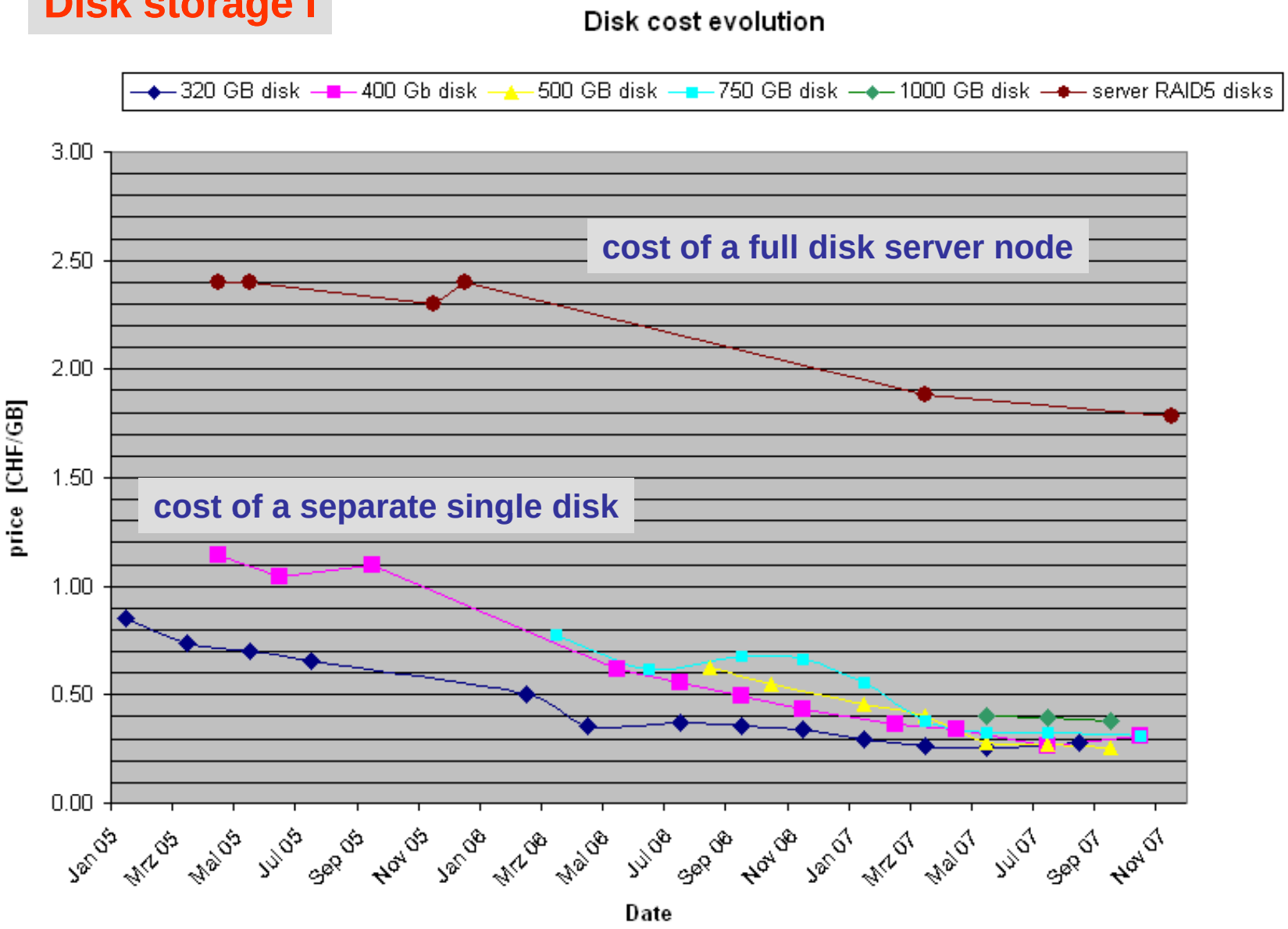
Memory I


Memory cost evolution
1GB modules
2 GB modules
4 GB modules
700.00
600.00
500.00
400.00
300.00
200.00
100.00
0.00
cost [CHF/GB]
Jan 05
Mrz 05
Mai 05
Jul 05
Sep 05
Nov 05
Jan 06
Mrz 06
Mai 06
Jul 06
Sep 06
Nov 06
Jan 07
Mrz 07
Mai 07
Jul 07
Sep 07
Nov 07
Jan 08
Date

Memory II

- still monthly fluctuations in costs, up and down
- large variety of memory modules
  different frequency and latency
  533 and 667 MHz about 10% cost difference, factor 2 for 1 Ghz
  higher frequency goes along with higher latency CAS
- how does HEP code depend on memory speed ?
- DDR3 upcoming , more expensive in the beginning
- is 2 GB per core really enough ?

Disk storage I



Disk storage II

Trends

- cost evolution of single disks is still good (~ factor 2 per year, model dependent)

- lots of infrastructure needed
  upgrade of CPU and memory
  → footprint of applications : RFIO, Gridftp, buffers, new functions,
  checksums, RAID5 consistency checks, data integrity probes

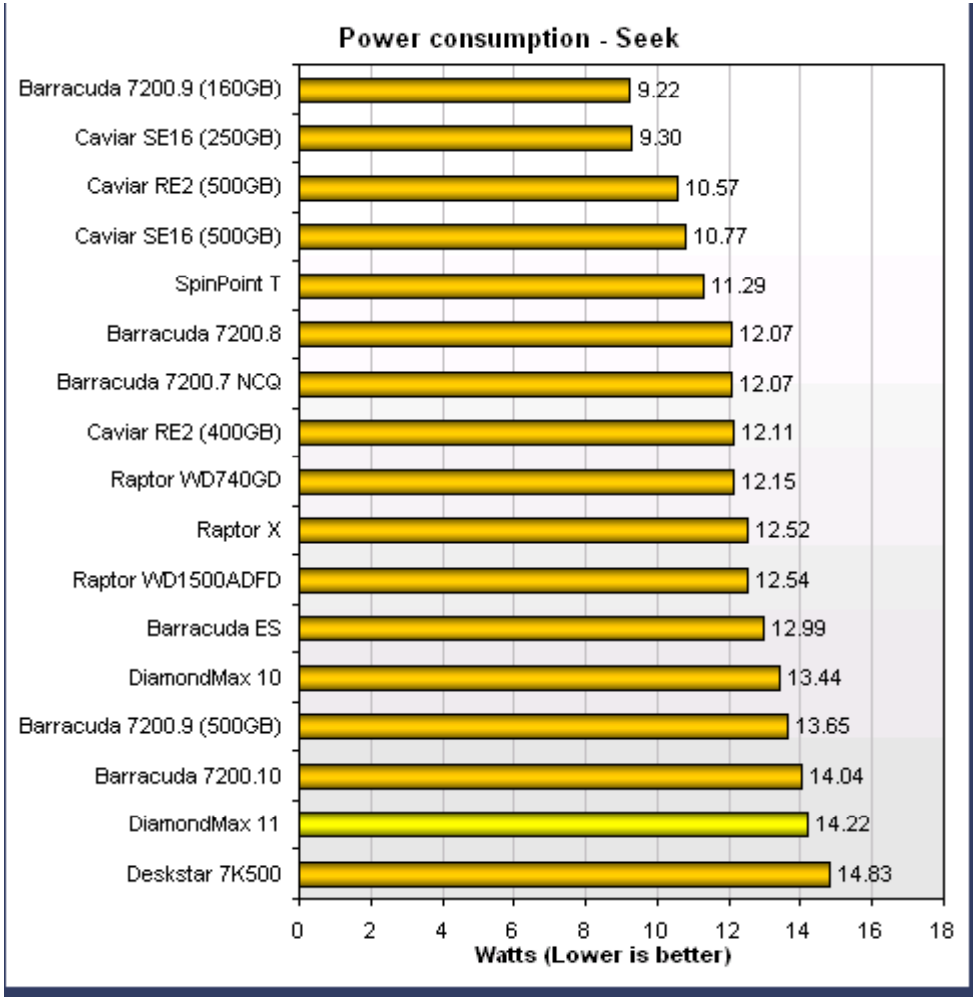
- need disk space AND spindles → use smaller disks or buy more → increase overall costs
  one 'effective' disk per core !? → higher growth rate than space requirements

- solid-state-disks, much more expensive (factor ~70) → data base area

- hybrid disks
  good for VISTA (at least in the future, does not work yet…)
  but higher price e.g. new Seagate disks + 256 MB flash == + 25% costs
  general trend for notebooks
  can't profit in our environment → seldom cache reuse

Disk storage III

power consumption for different disks (full write/read load)



power usage per disk is
only very little size dependent

assuming a disk server is a moderate CPU server plus ~20 disks
→
the power consumption per disk server will be about constant, not depending on the size (TB)