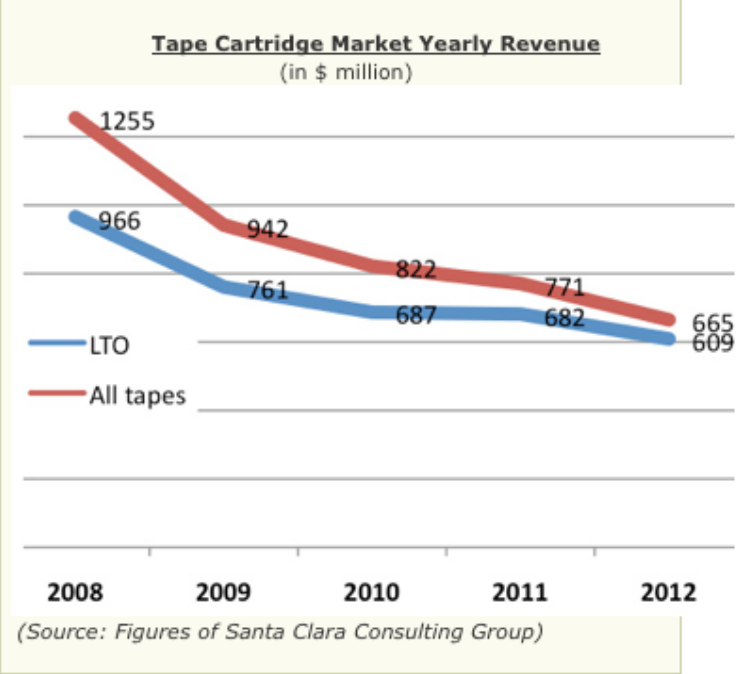
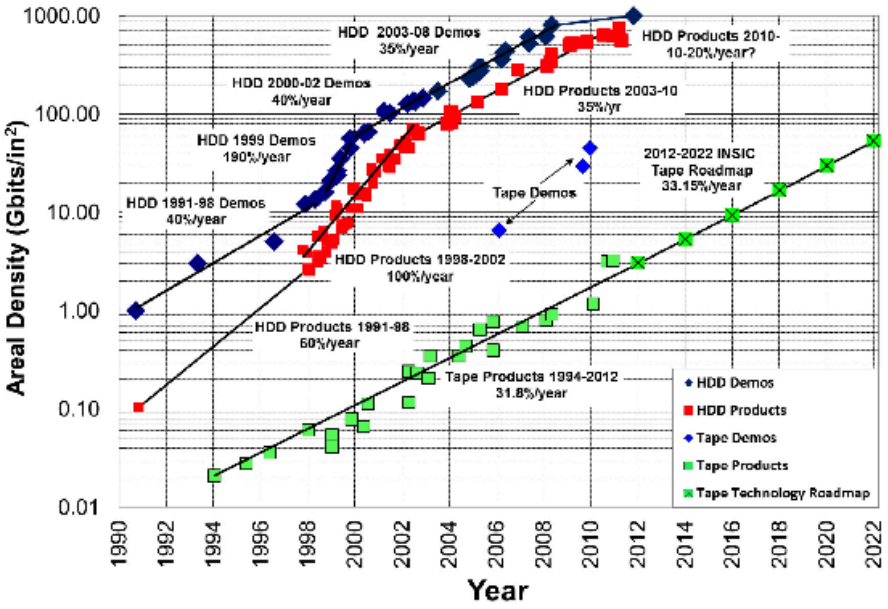
Computing Technology and Markets 2013
A brief overview

Tapes I



Tape cartridge market:
-10% growth rate year by year



Technology still improving
Step function every 2-3 years
Current cartridge density is
Oracle: 5 TB , IBM: 4 TB, LTO-6: 2.5 TB

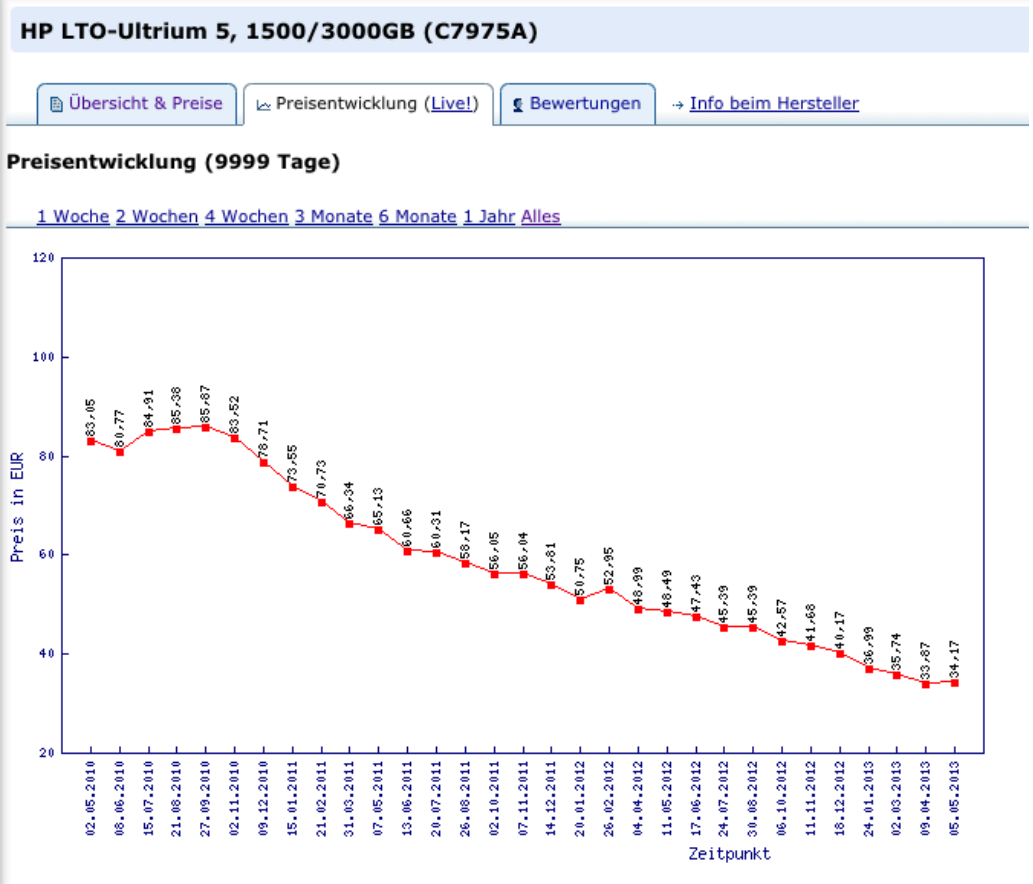
Next cycle will lead to 6-8 TB cartridges

Tapes II

LTO has 93% market share
Enterprise tapes ~2%

Factor 2.5 cost decrease over 3 years for cartridges

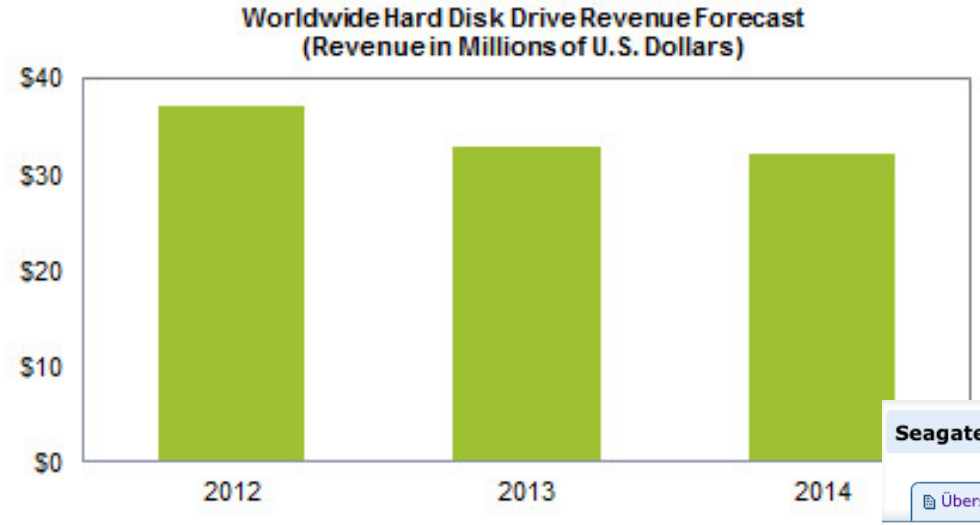
Today: 0.03 – 0.04 euro/GB



~23 Exabyte of tapes were sold in 2012

To be compared with > 1000 exabyte of worldwide data produced per year

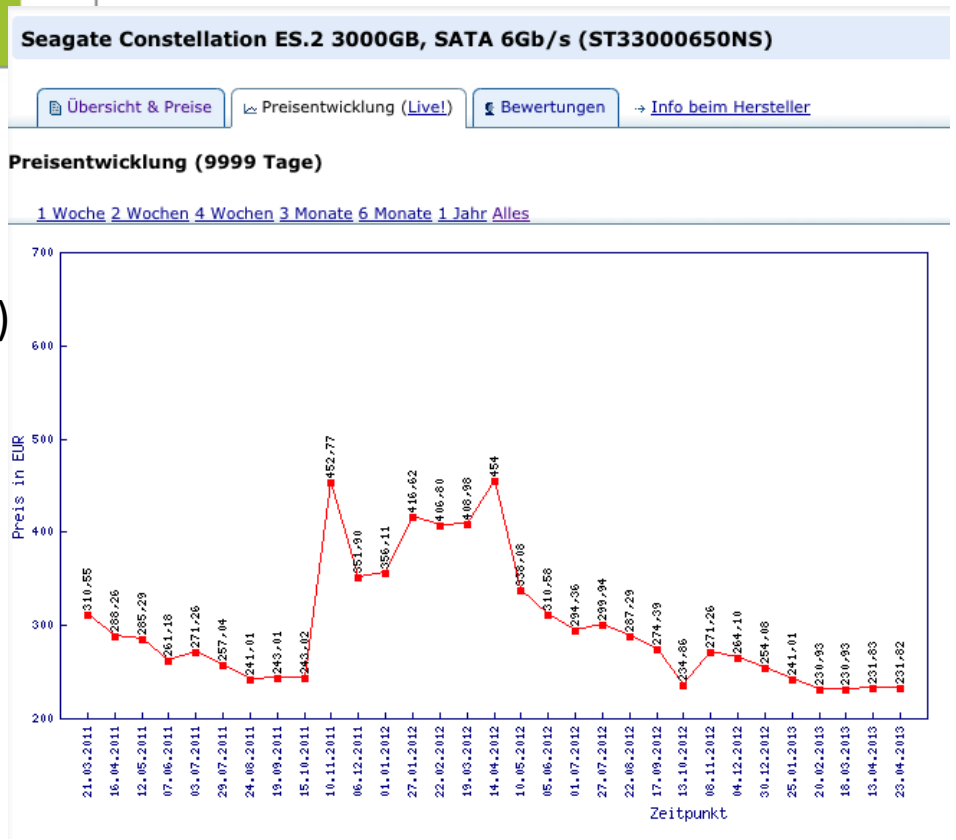
Disks I



Source: IHS iSuppli Research, February 2013

Fluctuating disk prices,
Difference between consumer disks
And enterprise SATA disks is up to a factor 2

0.04 – 0.09 Euro/GB (raw disks)



Decline of HDD market:
-Strong decline of desktop PCs (Q1 2013 -14%)
-Notebooks, Tablets and smartphones sales increase demands for flash memory (SSDs)
-consolidation of cloud storage

Disks II

475 million HDDs were sold in 2012 (630 million in 2011)
a decline of 8% is expected for 2013

69 million SSDs were sold in 2012
a growth rate of 48% is expected for 2013

- The perpendicular recording technology reached its density limit of 1 TB/inch2
- For the next ~3 years the recording technology will be based on intermediate schemes
  - SMR (shingled magnetic recording) from Seagate
  - Helium filled disks from Western Digital (Hitachi)

  4 TB disks in 2013/2014 moving to 5-6 TB,
- Disks based on the HAMR (Heat Assisted Magnetic Recording) technology will start to appear in ~2016, expensive technology, slowdown of price/space improvements

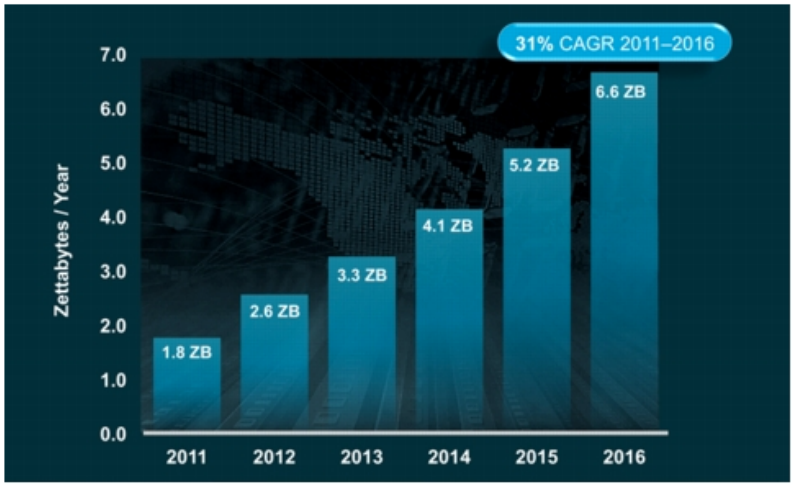
Problem:

Large space per spindle, little increase in sequential and random access speed
→ Overall computing architecture: job IO streams per spindle causing disk congestion

Quick calculation → HEP 2015 resources: 100-140 K disks (RAID6 or mirrored) versus 350 K processing cores

Network I

Figure 1. Global Data Center IP Traffic Growth



The Internet is forecast to reach the zettabyte era in 2016, but the data center has already en

Figure 2. Global Data Center Traffic by Destination

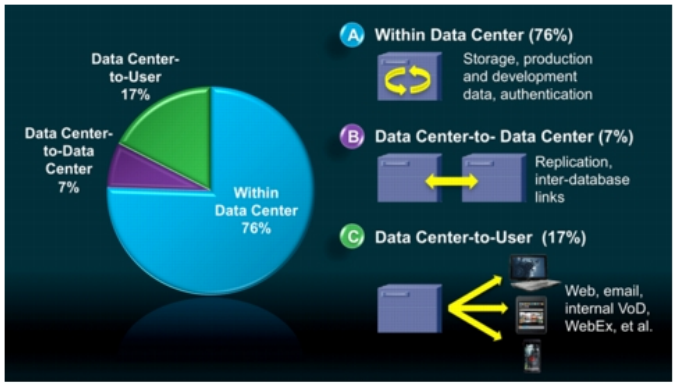
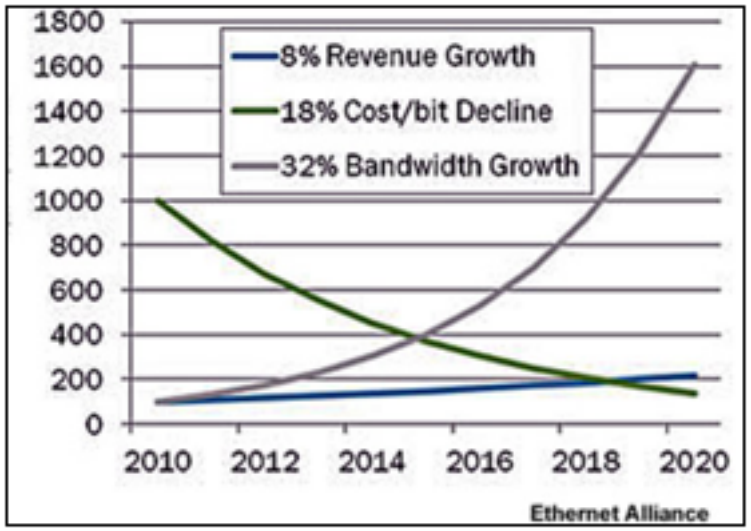


Table 1. Global Data Center Traffic, 2011-2016

Data Center IP Traffic, 2011-2016							
	2011	2012	2013	2014	2015	2016	CAGR 2011-2016
By Type (EB per Year)							
Data center to user	299	438	561	714	912	1,160	31%
Data center to data center	118	173	222	284	365	468	32%
Within data center	1,338	1,940	2,468	3,126	3,969	5,021	30%

Internal traffic in the CERN computer centre in 2012: 500 PB/year
→ Maybe ~2 EB/year for the sum of the HEP centres
Compared to 2 ZB/year worldwide

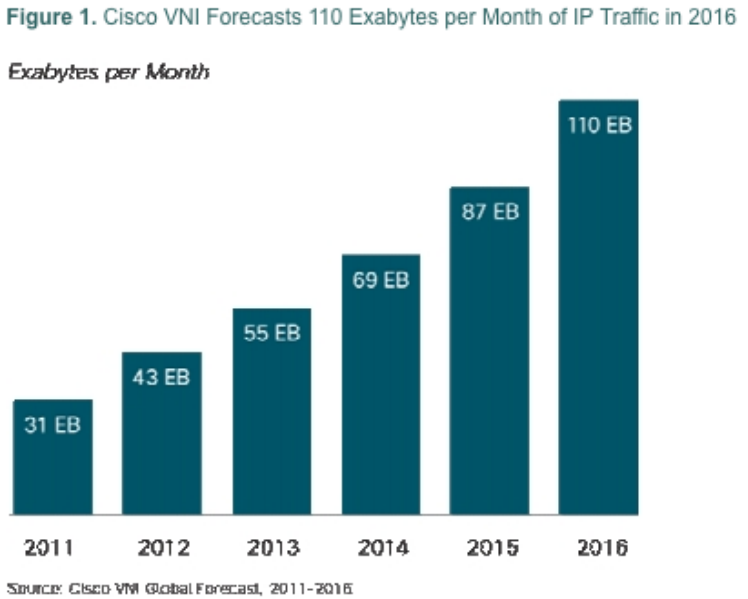
Network II



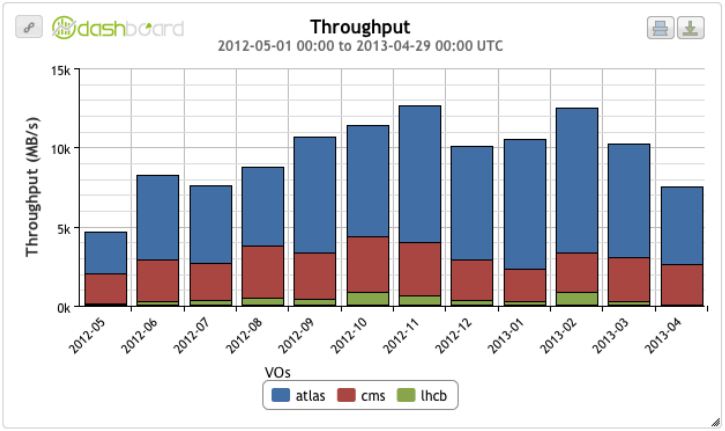
Cost reduction of factor 3 in 3 years for 10 Gbit WAN connectivity

HEP IP traffic is negligible

Do we still special network arrangements ? (LHCONE, LHCOPN)



Worldwide IP traffic 2012: 516 EB/year



HEP WAN traffic 2012: 300 PB per year

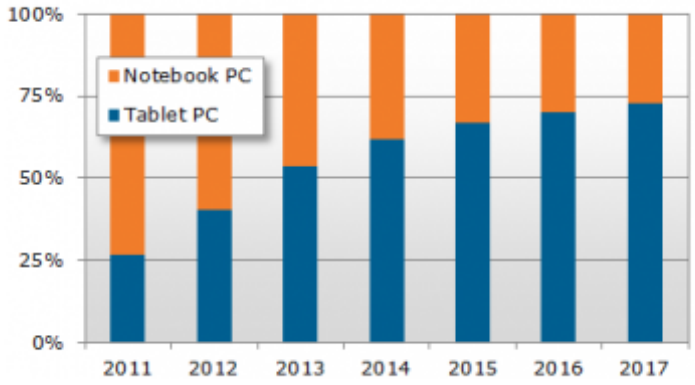
Processors and Servers I

Units shipped [M]
2012 → 2013

Simple Phones	750 → 900
Smartphones	722 → 1000
Tablets	128 → 200
Notebooks	202 → 200
Desktops	148 → 142
Server	10 → 9.5
HPC	0.1 → 0.1

HEP is here

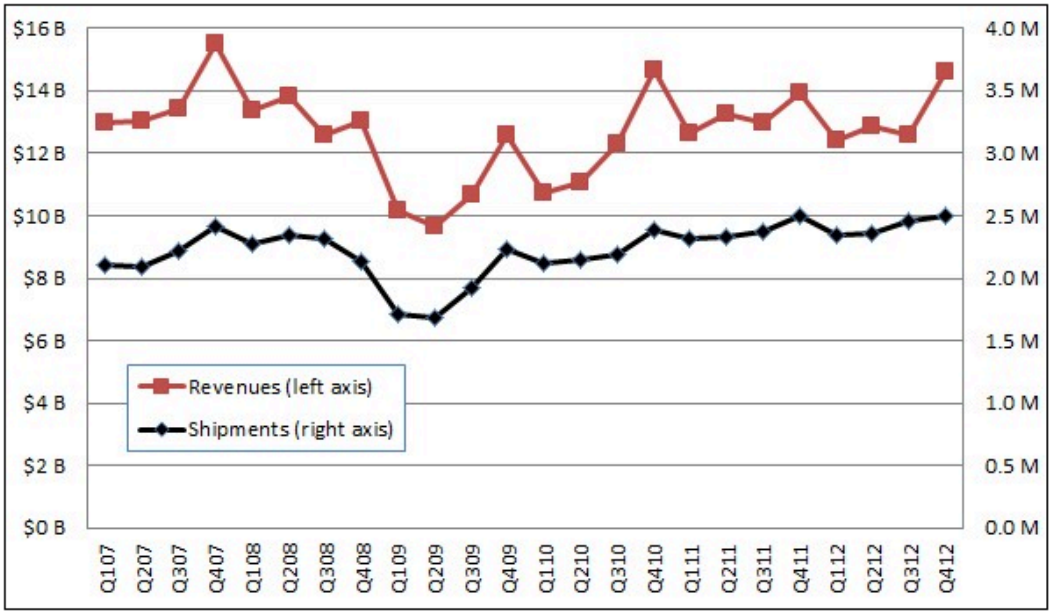
Heavy PC decline in Q1 2013 (-14%)



In 2013:

- #smartphones >= #simple phones
- #tablets >= #notebooks

Processors and Servers II



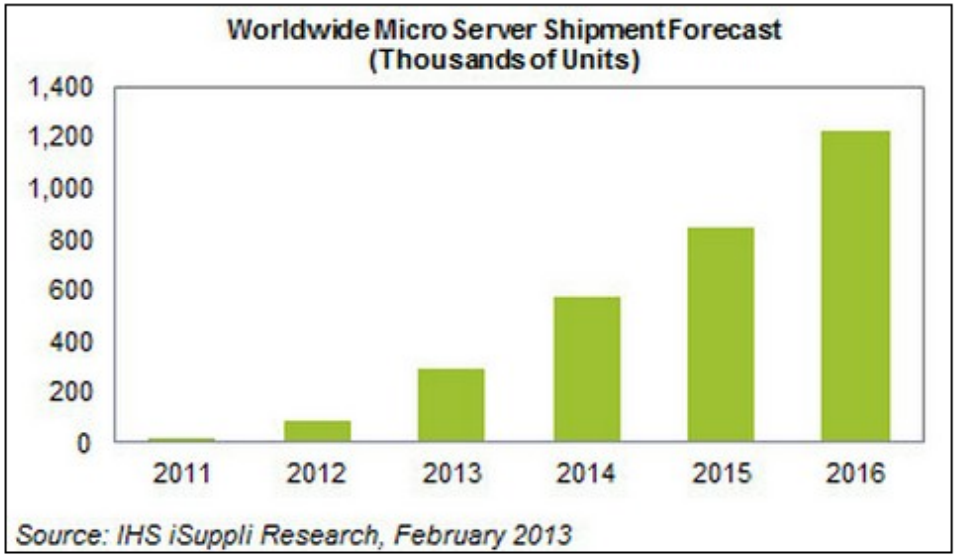
The Great Recession is not hard to spot in Gartner's quarterly server revenue and shipment numbers

HEP is buying from the server market (dual CPU, ECC memory, server-grade Disks)

- X86 represents 98% of the shipments and 74% of the revenues
- Market share INTEL 96% AMD 4% (steady decline over the last years)
- Shipments are more or less flat over the last 6 years (market saturation !?)
- Improvement only via price/performance increase for new processor generations

Processors and Servers III

A new development in the server market is the introduction of Micro Server. Last year about 80000 were sold and the predictions claim a market share of 10% by 2016.



Microserver ships are expected to grow by a factor of 50 between 2011 and 2016

The design is based on ARM or ATOM Processors and is aimed at special Web/cloud task with high energy efficiency. (Facebook, Google, Yahoo, etc.)

Supplier: Seamicro/AMD, Calxeda, HP (Moonshot), E4 (AKA, Boston Virids, etc.

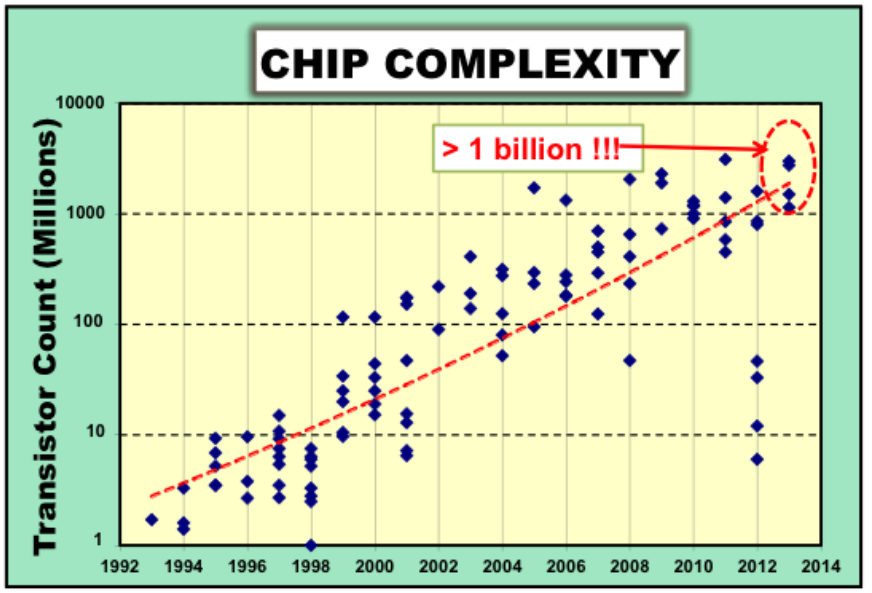
During the last 18 month this was seen as the domain of the ARM designs. The latest Intel Atom architectures (Silvermont) lead to analyst prediction of Intel taking the majority of the Micro Server market during the next years. The ARMv8 architecture might change the picture again at the end of the year.

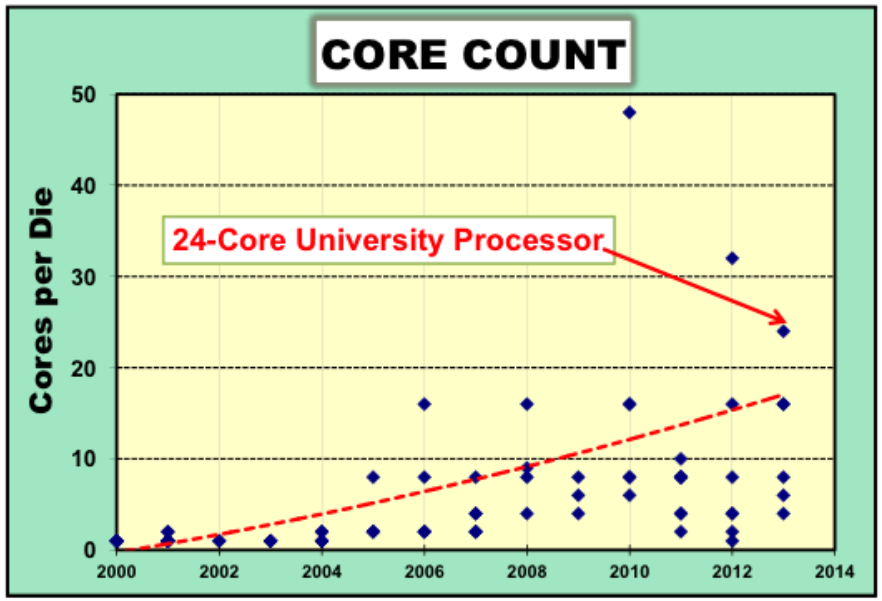
Processors and Servers IV

Moore's original Law still holds:
The amount of transistors per chip is about doubling every 21 month
(factor 1.5 increase per year)

But the really important variant of Moore's Law is:
The price/performance value for CPU servers is decreasing about a factor 1.25 per year
(factor 2 improvement every 3 years)

The increase of cores for the mainstream processors is still following a linear law:
plus 1-2 per year





Connections I

Problem: **Flat budgets and increasing computing requests**

Solution: **Increase efficiency**

- Better utilization of processors (Cycles-per-Instruction, memory access, floating point units, vector units, etc.)
  - → 'traditional' server improvements, Intel Xeon EP
    more than the 20-25% price/performance improvement per year
  - → utilization of co-processors: GPUs (Nvidia, AMD), Xeon PHI (Intel)
  - → lower end processors in micro-servers (ARM and Atom)
  - → integrated CPU+GPU processors (APU design from AMD, TegraX from Nvidia, Samsung Exynos)
  - → more 'exotic': Adapteva 99$ supercomputer, Altera Cyclone ARM+FPGA

- Improve general computing architecture:
  Network versus **Disk** versus **Tape** versus **CPU**

Carefully considering the 'side-effects', taking a holistic view

Connections II

Good : potential of increasing processor usage by factor >10

Bad : potential to overall decrease efficiency by factors

e.g.

Site spending profile and correlation of the different activities is important
CERN estimation (strongly site dependent)
CPU:20% disk:35% tape:12% network:18% electricity:10%
Cut the disk space in half and double the CPU capacity !?? or does this lead to a doubling of Network costs ??

Side-effects: faster programs means higher IO to disk (local/remote), changes in the network connectivity of CPU servers, change of network blocking factors

The markets will have the last word about the technology.
High end GPUs for computing are wide-spread in the HPC market.
Issue: high end cards are a niche market, subsidised internally with the revenues from the commodity gamer cards.
2012 sales: 550M graphics chips, 57M discrete cards, ~1m professional cards
Desktop pc decline has a large impact on the discrete graphics card market!

Multi-threading of HEP code is needed to make use of co-processors,
But it does not improve costs yet: cutting memory by 50% on CPU servers saves ~5%; correlates cores and thus 'concentrates' IO on network, disk and memory; requires dedicated nodes → efficiency of usage