

The TechLab Project

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
20-25% price/performance improvement per year (vendors)
 - 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

Technology

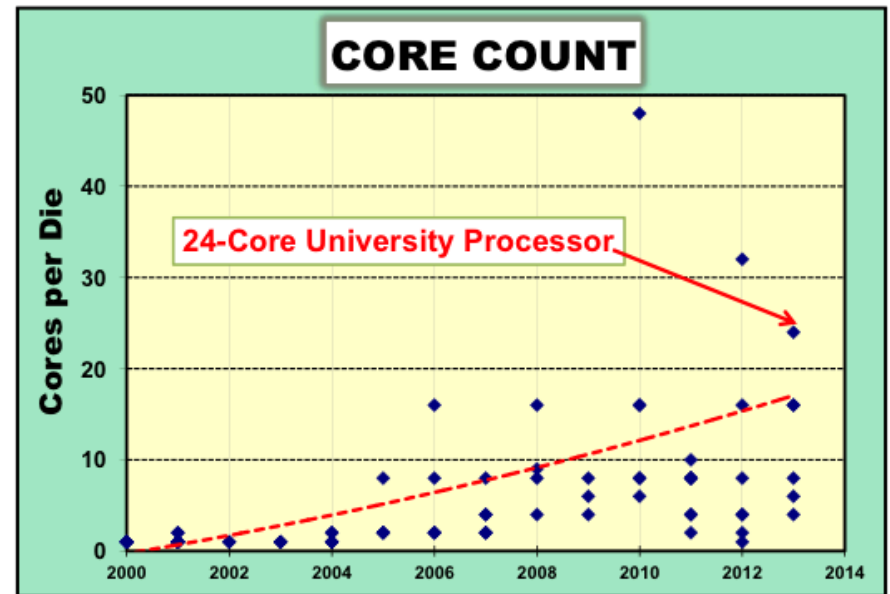
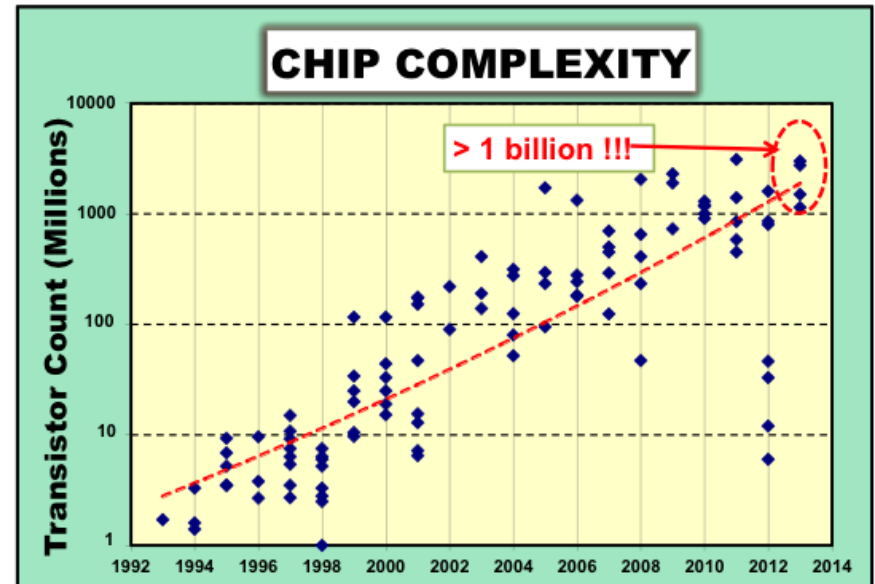
Moore's original Law still holds:

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



Markets

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  HEP is here

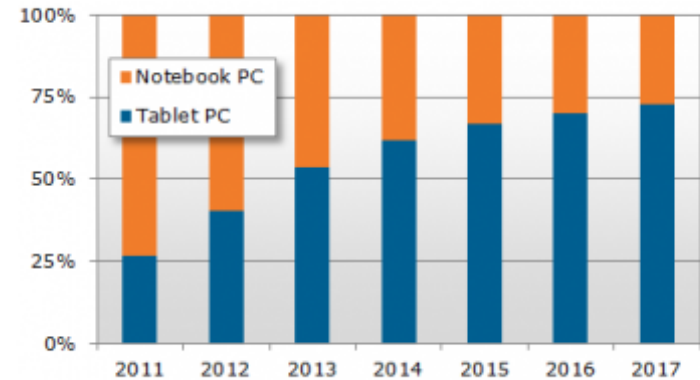
HPC 0.1 → 0.1

In 2013:

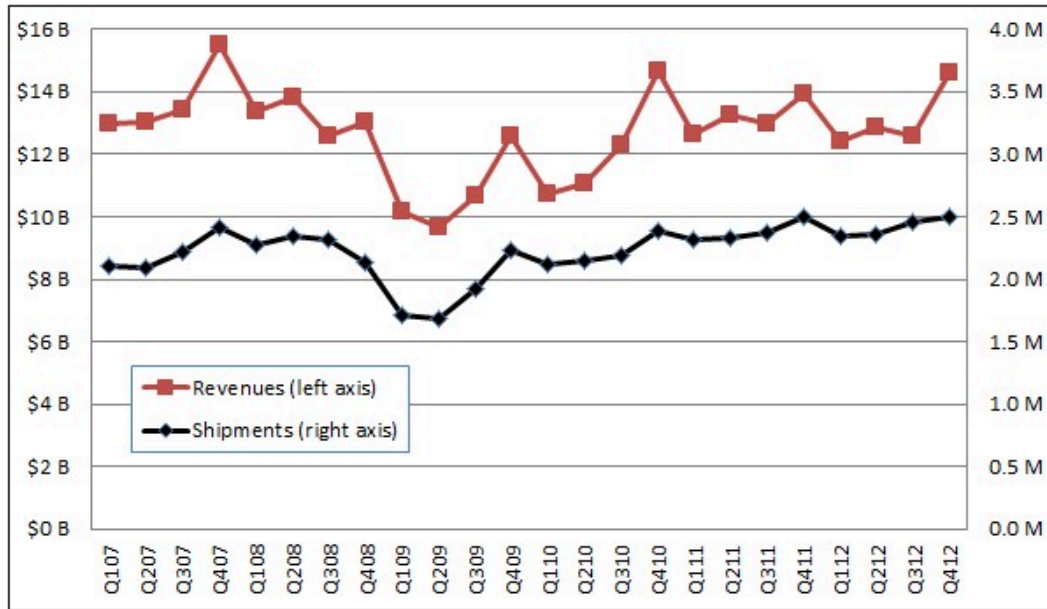
- #smartphones >= #simple phones
- #tablets >= #notebooks



Heavy PC decline in Q1 2013 (-14%)



Servers I



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

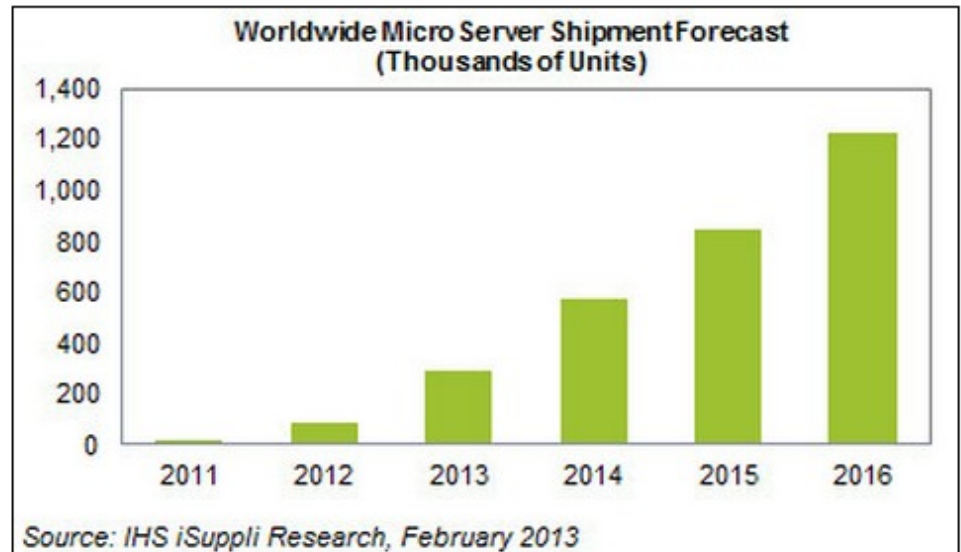
Servers I

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.

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 Viridis, 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.



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

TechLab Project I

Collaboration between IT and PH (initiated by SFT)

Provide a technology setup with hardware and software for advanced code development (efficiency improvements, new architectures, etc.)

Large phase-space !

Combinations of processor architectures and compilers

Processors : Intel, AMD, ARM server, desktop and notebook type

Accelerators : Nvidia, AMD, Intel, Xilinx

New designs : Tiler, Adapteva, Kalray,.....

Compiler : gcc, Intel, CAPS, Portland,

Languages : C++, CUDA, OpenCL,.....

What else and how to focus without too many restrictions ??!

TechLab Project II

**IT will make available a set of new processor architectures
and the corresponding infrastructure**

Starting point → (choices are based on current market and technology developments)

End of June : 16 and 32 core servers with Nvidia Tesla and Xeon Phi

**September : 4 Intel server with Nvidia Tesla
4 Intel server with Xeon Phi
HP Moonshot micro server with 10 blades (ATOM based)
E4 with 8 blades based on Tegra3 (ARM plus GPU)**

Availability of compiler, debugging and monitoring tools (Intel, CAPS, perfmon, etc.)

**Some man-power devoted to gather expertise in IT and support the setup.
Best-effort 'service'**

Regular upgrades of hardware and software (suggestions !!??)

**Close collaboration needed to ensure a balance between
→ bleeding edge 'wild' technology evaluation
→ 'stability' to ensure reproducible results and benchmarks**

Caveats I

Good : potential of increasing processor usage by factor >10

Bad : potential to overall decrease efficiency by factors

Need to take into account the Site-Computing-Architectures and the Computing Market developments

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

Caveats II

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: 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

Today we have based our computing model on standard servers (Intel and AMD processors).

Today all new approaches (GPU, Xeon Phi, ARM, micro-server) have a worse price/performance
>= factor 2

But technologies and market are changing fast..... need to react pro-active and be prepared

Summary

- New TechLab project will be hosted by IT
Hardware, software and some man-power support (best effort)
- Needs close collaboration with PH and experiments
- Regular checkpoints for a holistic view:
Software technology + hardware technology + markets
+ site-architecture + computing model

→ decide on the 'right' approach for hardware and software 'upgrades'