Computing Technology and Markets

Bernd Panzer-Steindel / CTO / CERN

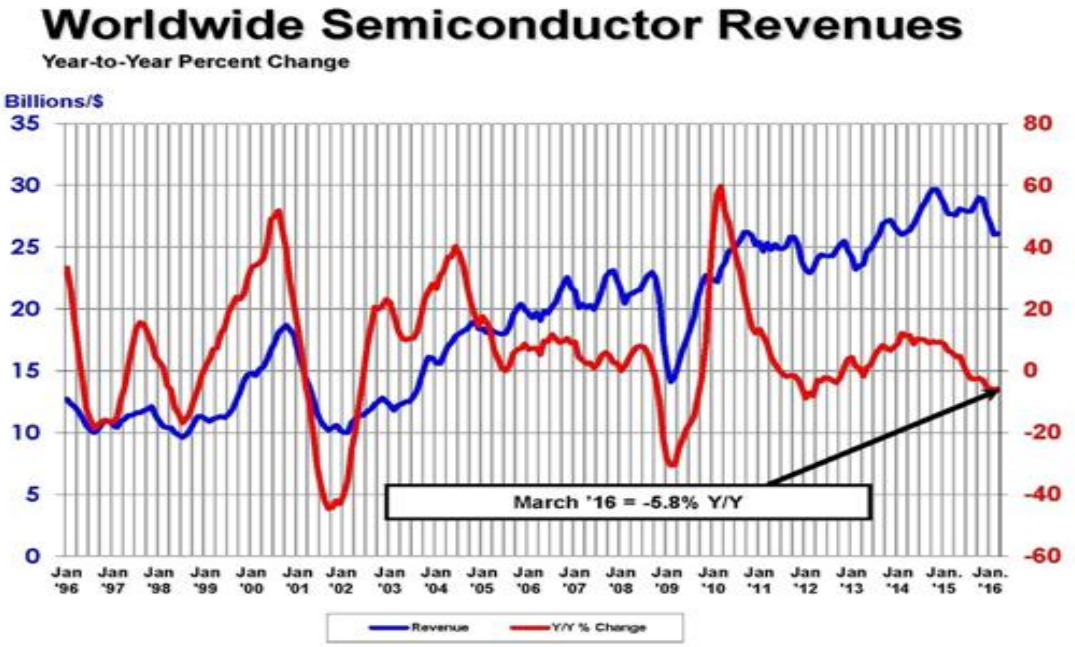
Outline

- Semiconductor market
- Device market
- Processors
- Hard Disk
- Solid-State Disks
- Memory
- Tapes
- Server

- Summary
- References

General Market

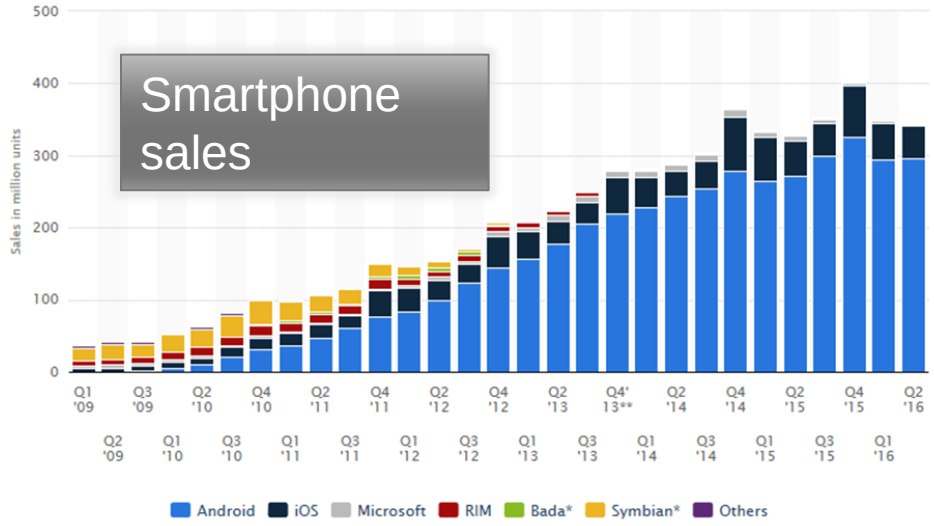
Few companies dominating the markets

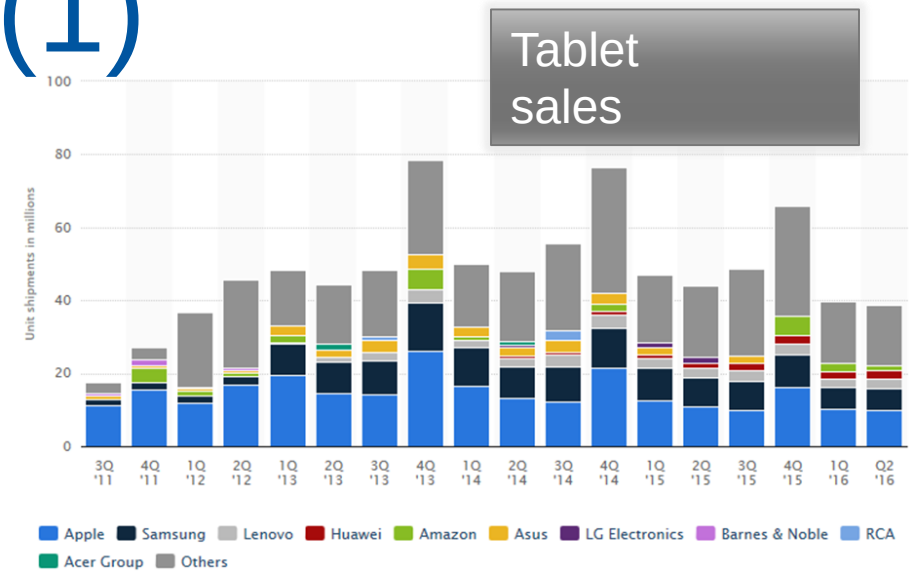


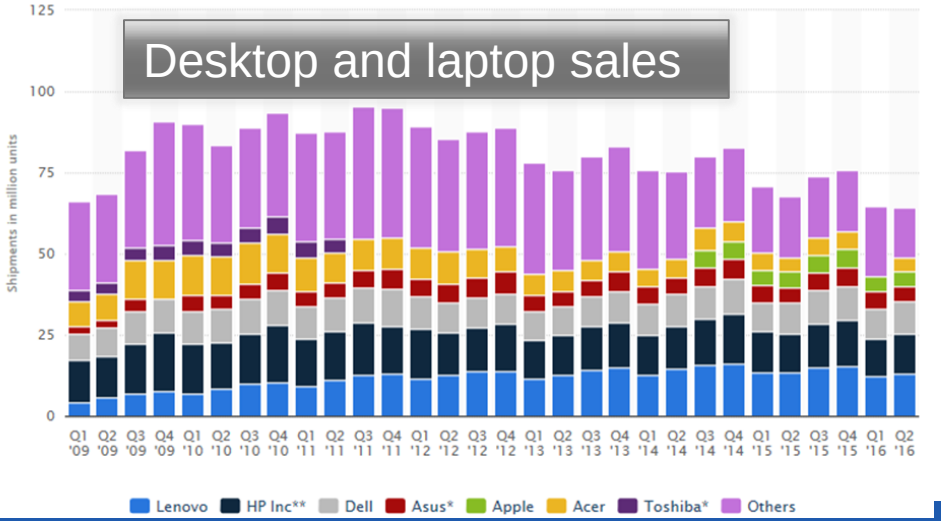
Server CPUs	Intel (99%)
FPGA	Xilinx (49%), Intel (38%)
GPU	Intel (72%), Nvidia (14%), AMD (14%)
Hard disks	Western Digital (44%), Seagate (40%), Toshiba
Tape drives	HP, IBM, Oracle
Tape media	Fujifilm, Sony
NAND	Samsung (45%), Toshiba, Western Digital, Intel
DRAM	Samsung (47%), Hynix, Micron/Intel

Forecast for 2016: -1%
Total: ~330 B$/y

Device Markets (1)



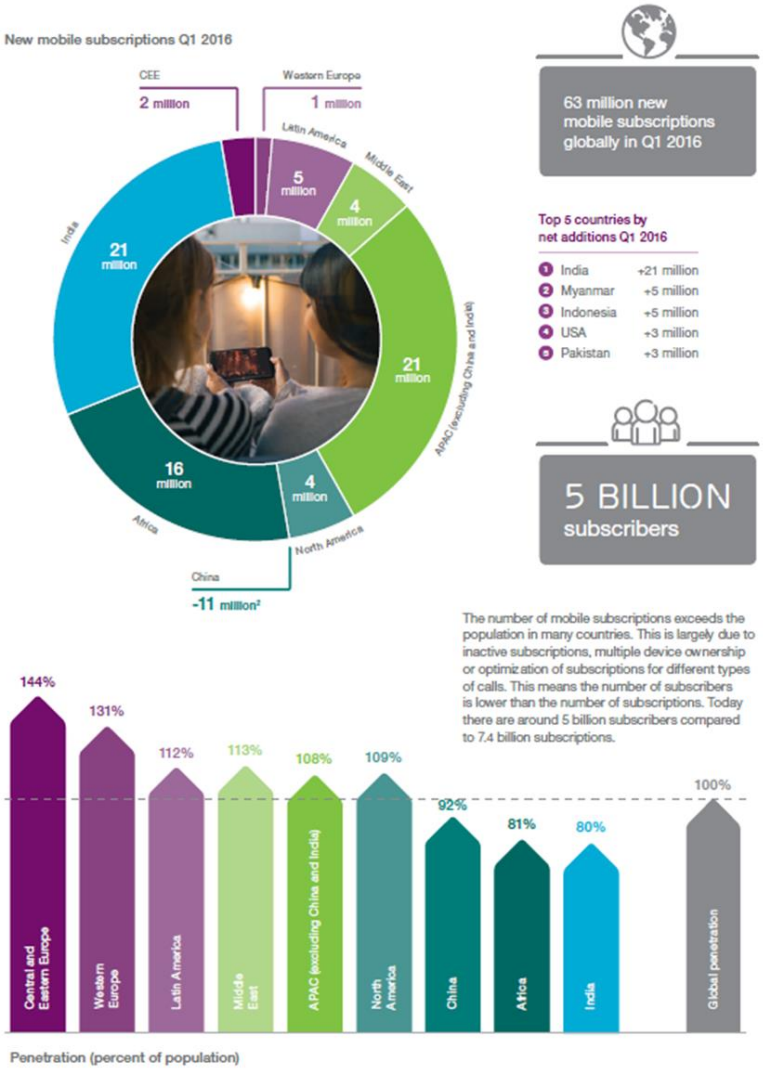


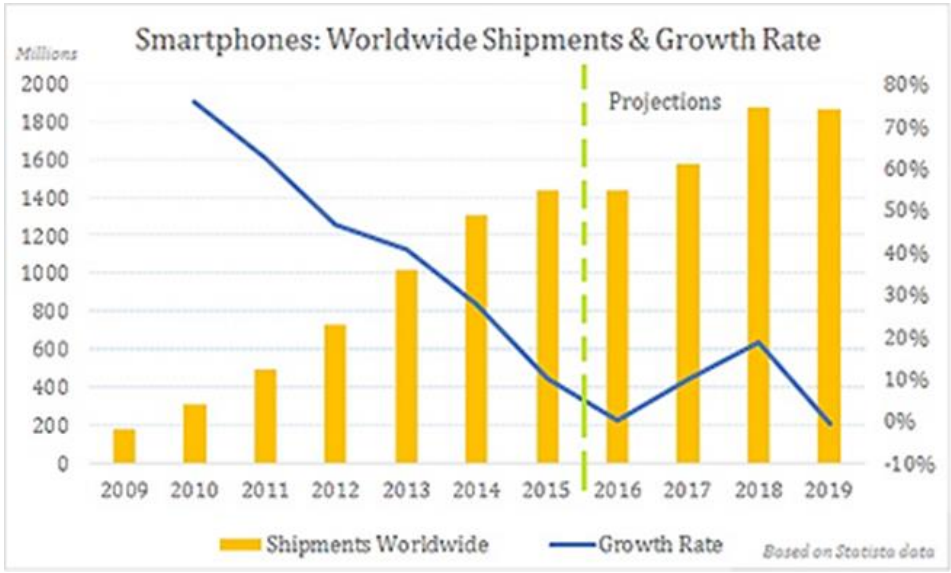


Market saturation:
minimal or negative growth rates
Longer product lifetimes

Smartphones	0-2 %
Tablets	-12%
Desktops and laptops	-7%
Servers	-3%

Device Markets (2)

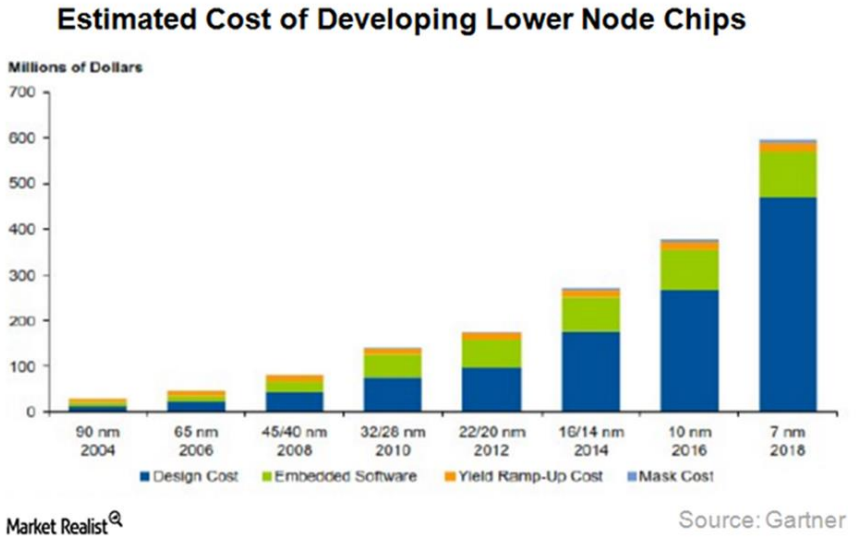




Saturation:
7.3 B phone subscriptions world-wide – more than the population

Replacement bump expected in 2018

Processors (1)



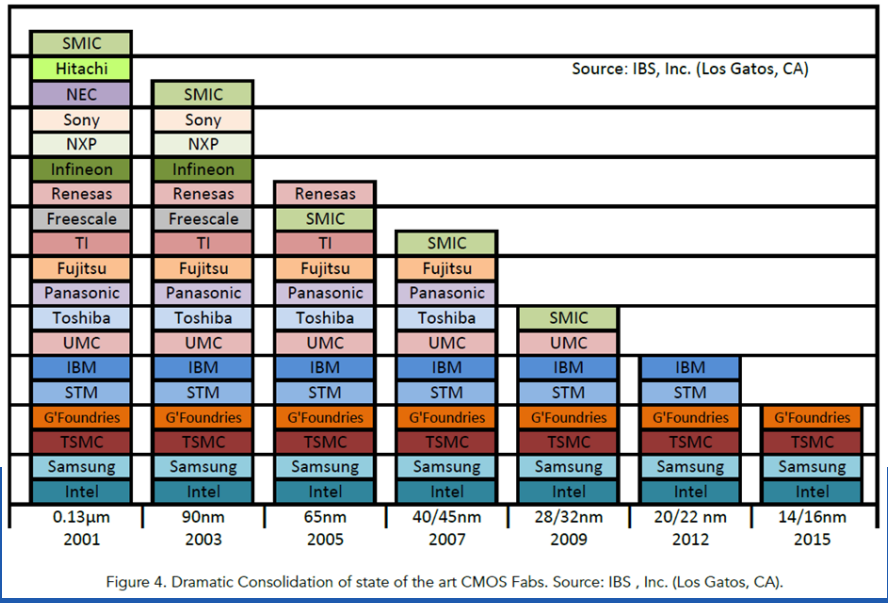
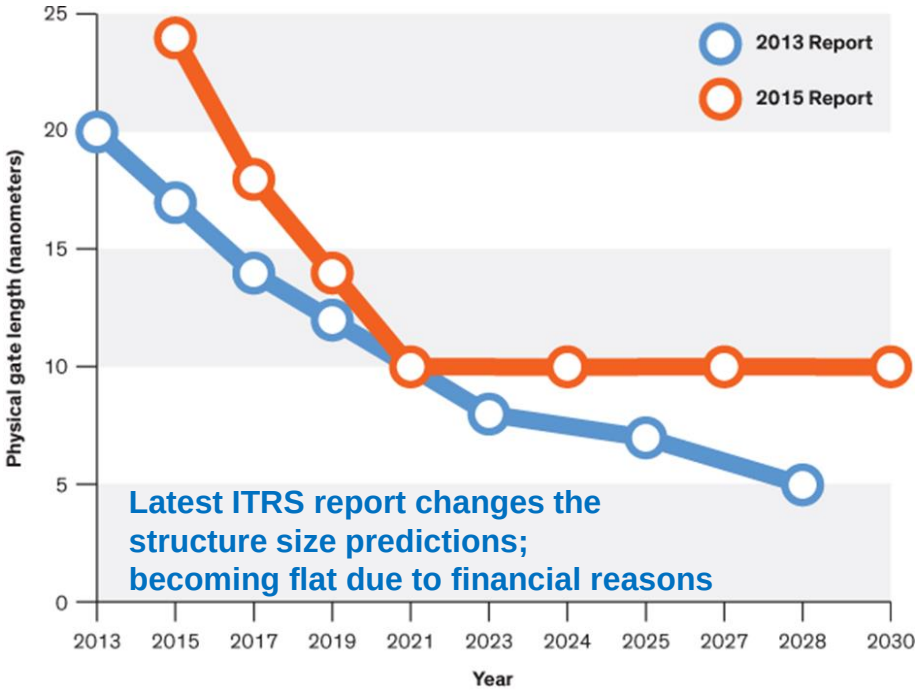


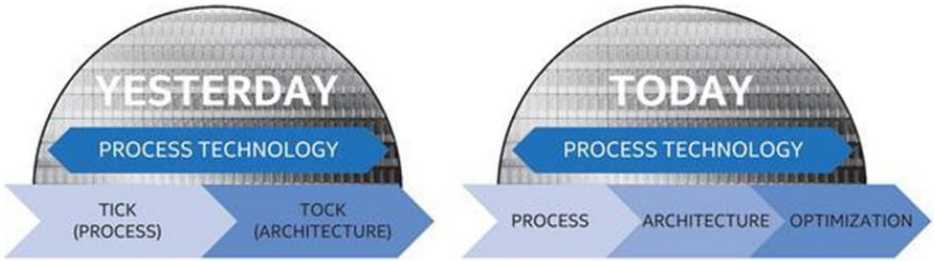
Figure 4. Dramatic Consolidation of state of the art CMOS Fabs. Source: IBS , Inc. (Los Gatos, CA).

Non-linear costs for development

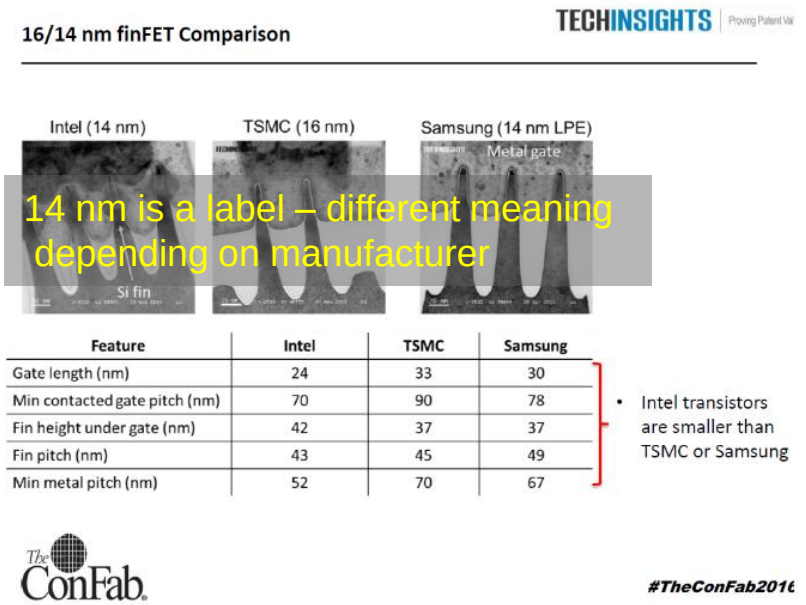
- Only four companies able to fabricate 14 nm chips
- 10 nm Samsung fab costs $14 B

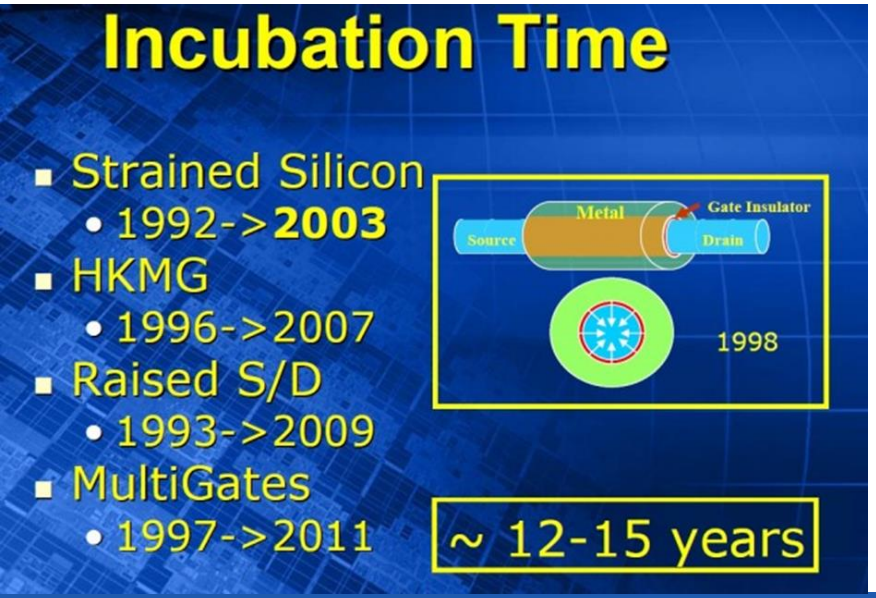


Processors (2)



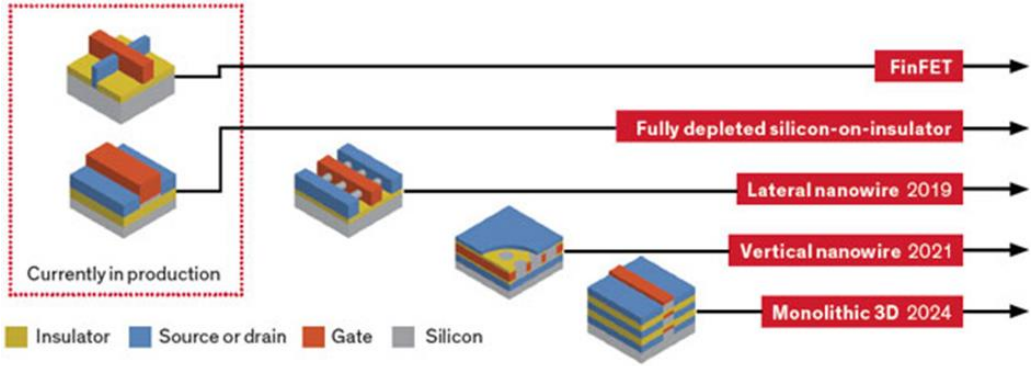
Intel moved from 2-year cycle to 3 years or more





Decrease of feature size goes along with new material technologies

R&D → production needs 12-15 years



7nm structures need new technologies: nanowires and non-silicon material

Accelerators: GPU (1)

Embedded market shares (CPU+GPU): Intel 72%,
Nvidia 16%, AMD 12%

Discrete GPU cards: Nvidia 77%, AMD 23%

Desktop and notebook shipments declining

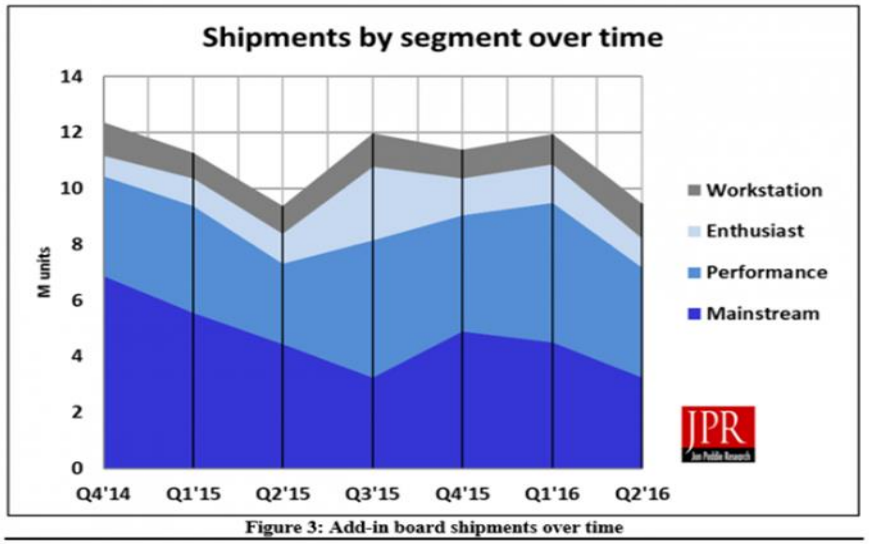


Figure 3: Add-in board shipments over time

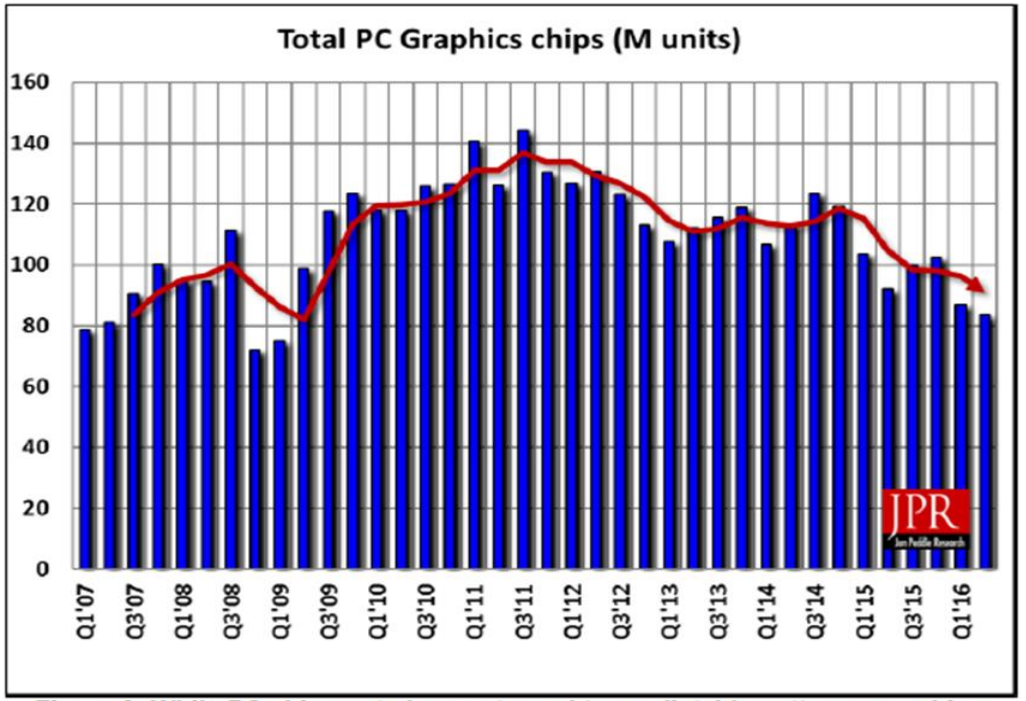


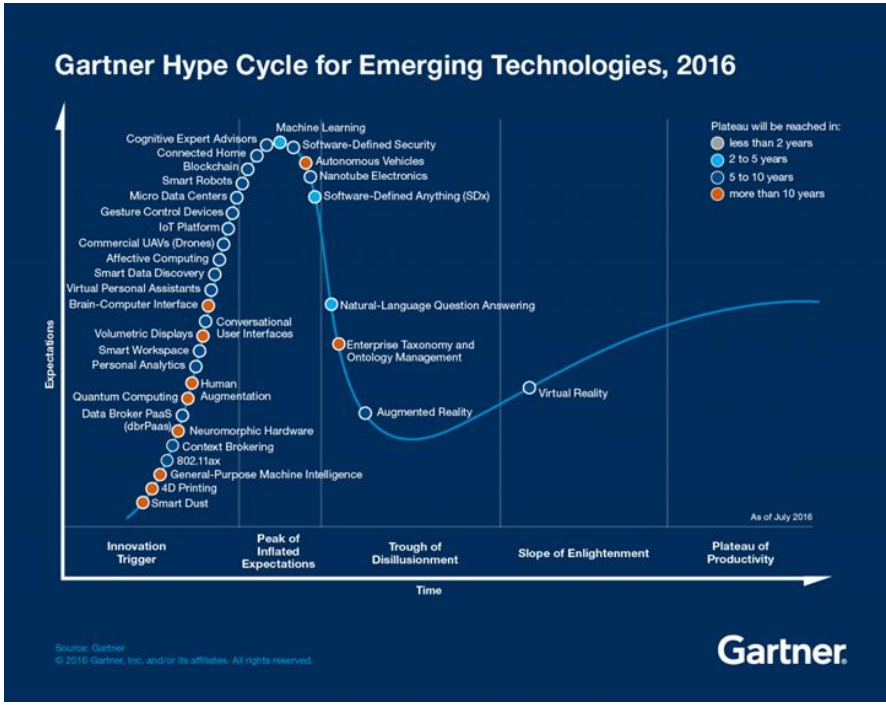
Figure 3: While PC shipments have returned to predictable patterns, graphics shipments have been erratic and defy any seasonal attributes

Focus: high-end Gamer (DP and FP16 artificially reduced)

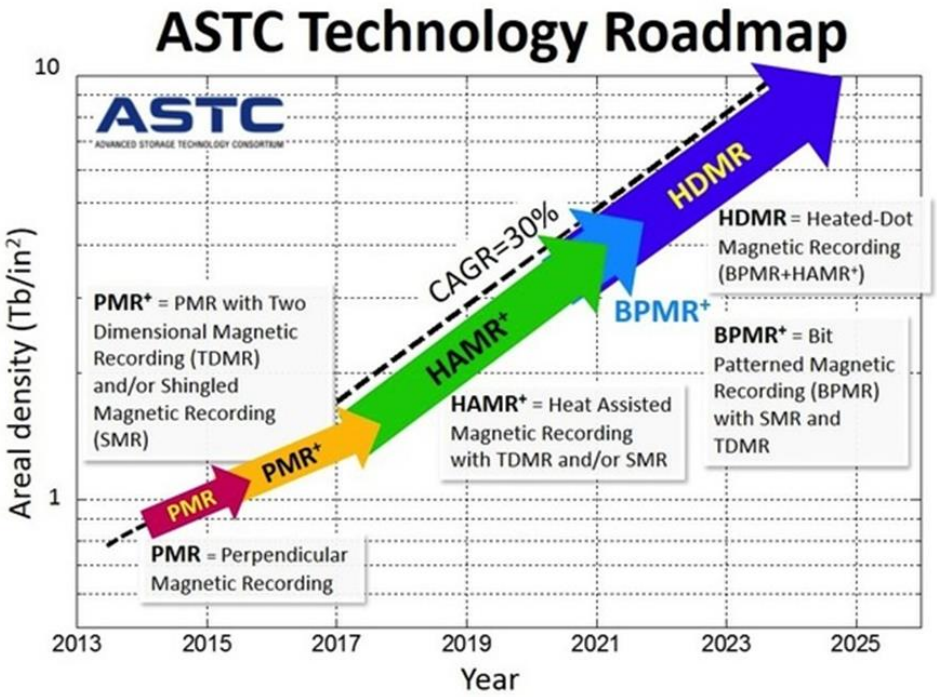
Professional workstation cards and HPC: small niche, ~2 million cards per year (compared to 350 million total GPUs)

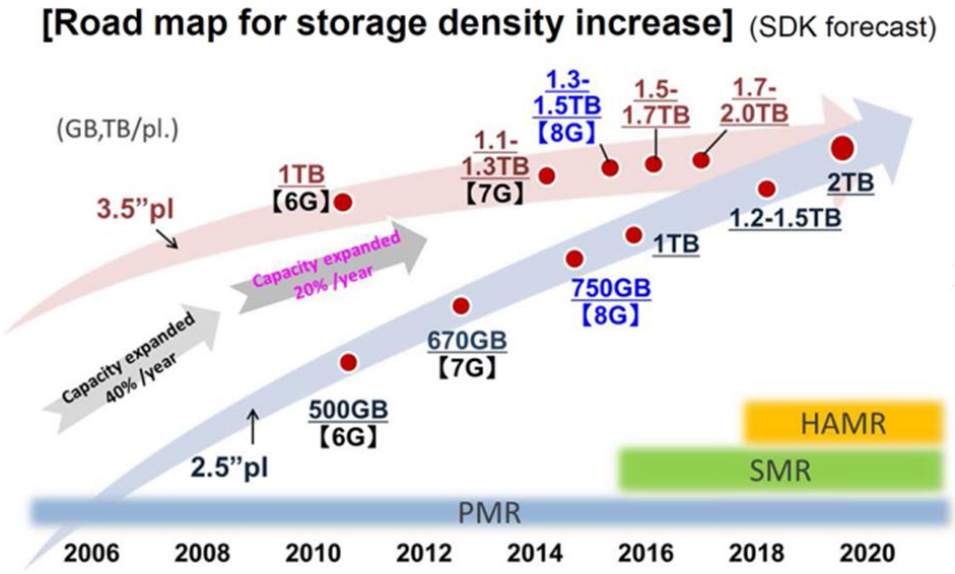
Accelerators: GPU (2)

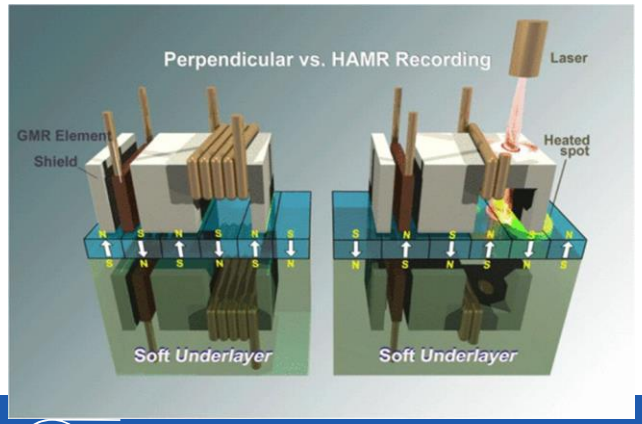
- New focus for graphic cards : machine learning
- Move to FP16 and even INT8 architectures, less precision → 8 bit processing !
- Google TPU Tensor Processing Unit
- New start-ups with special processor designs:
  e.g. KnuEdge, Nervana (just bought by Intel), krtkl, Eyeriss
- Essentially not usable as general purpose processors (online?!)
- Intel changing strategies also for their KnightsXX processors, 'forking' models (increase FP16 and decrease DP)
  ~100k units per year, very small market
- Qualcomm plans to add neuromorphic chips into the smartphone



Hard Disks (1)







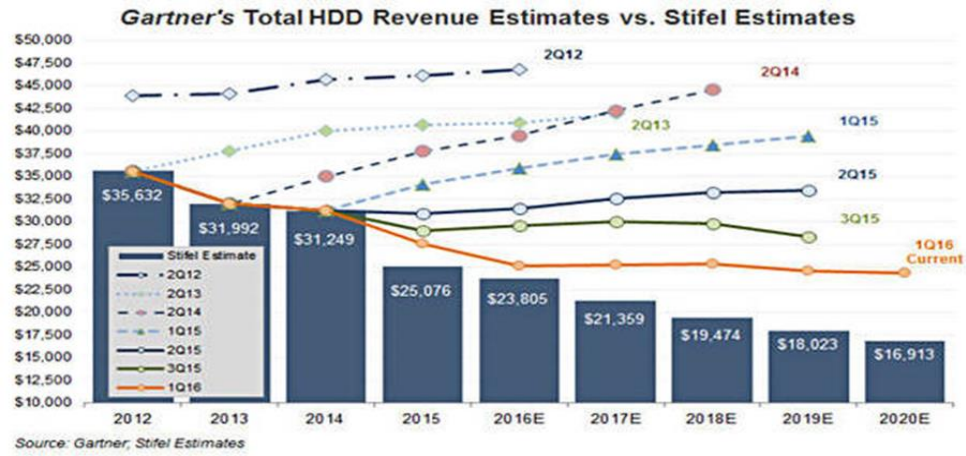
PMR limit at 1 TbPSI
SMR adds ~25%, market small
HAMR should provide 5 TbPSI

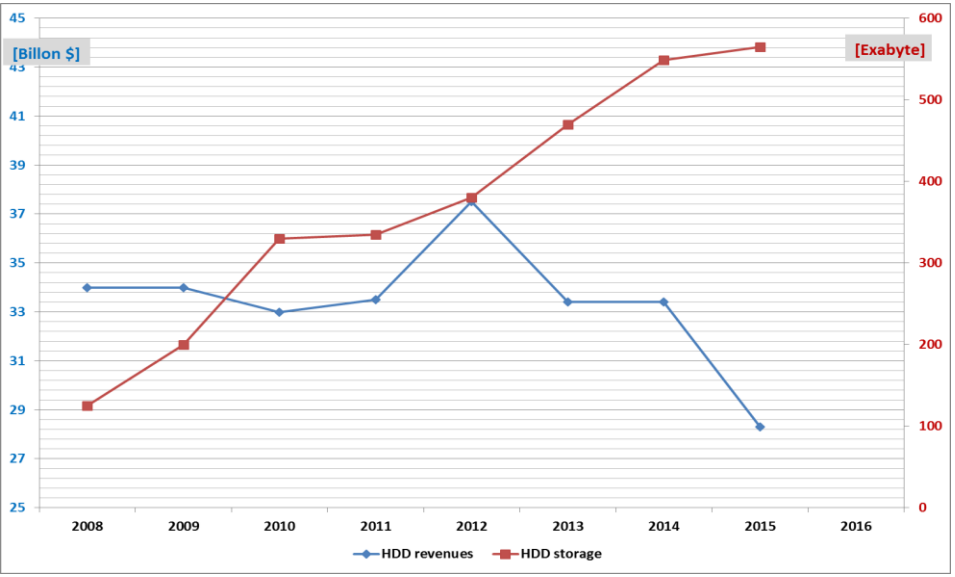
HAMR delayed, production in 2018

Combining bit density (30% annual growth rate) and volume density (number of platters, helium) → 100 TB in 2025 conceivable

Hard Disks (2)

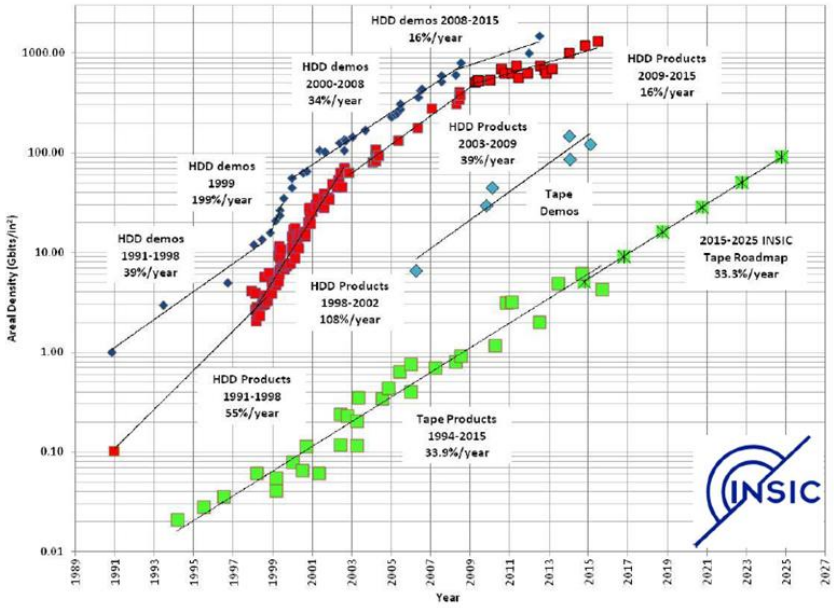
Continuous decrease in revenues
Forecast changes every year





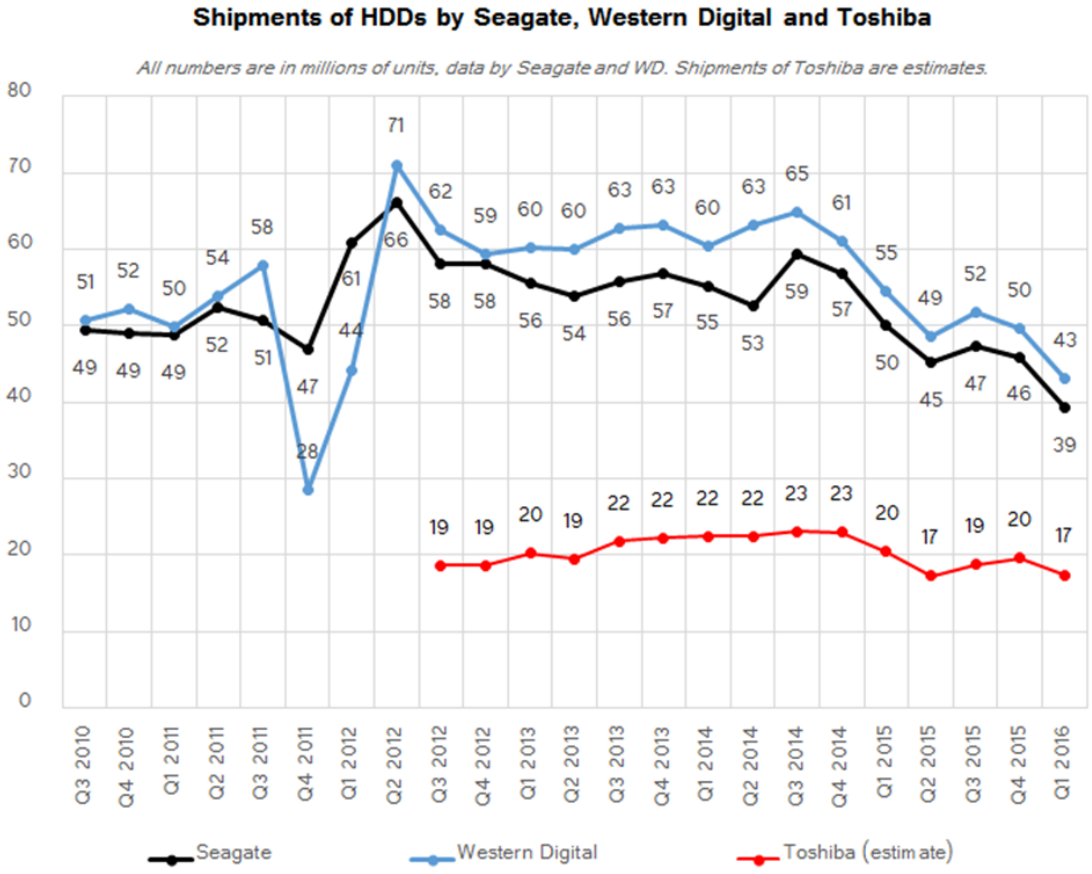
Areal Density Trends

Chart provided courtesy of the Information Storage Industry Consortium (INSIC)



Areal density improvement dropped from ~40% to 16% per year

Hard Disks (3)

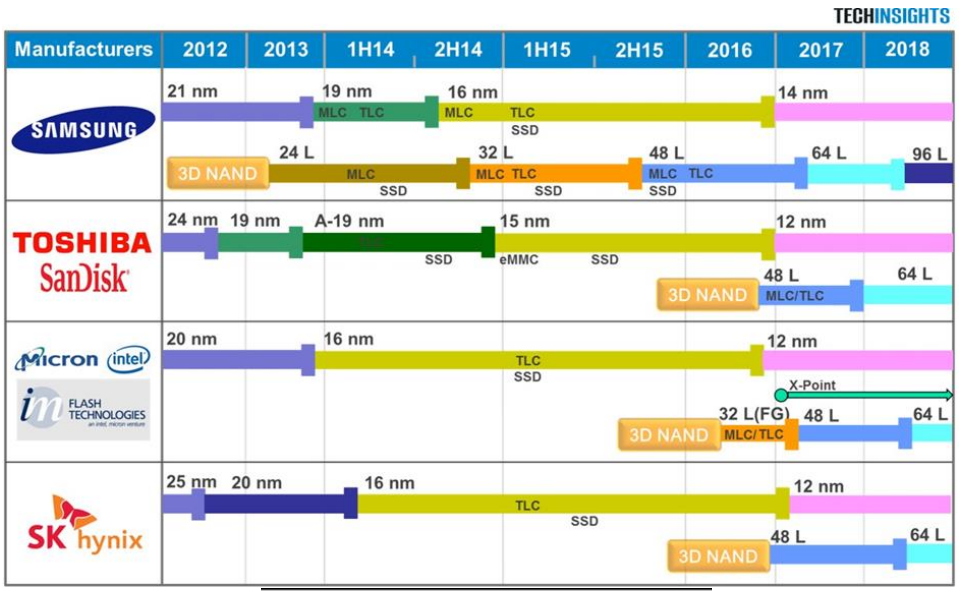


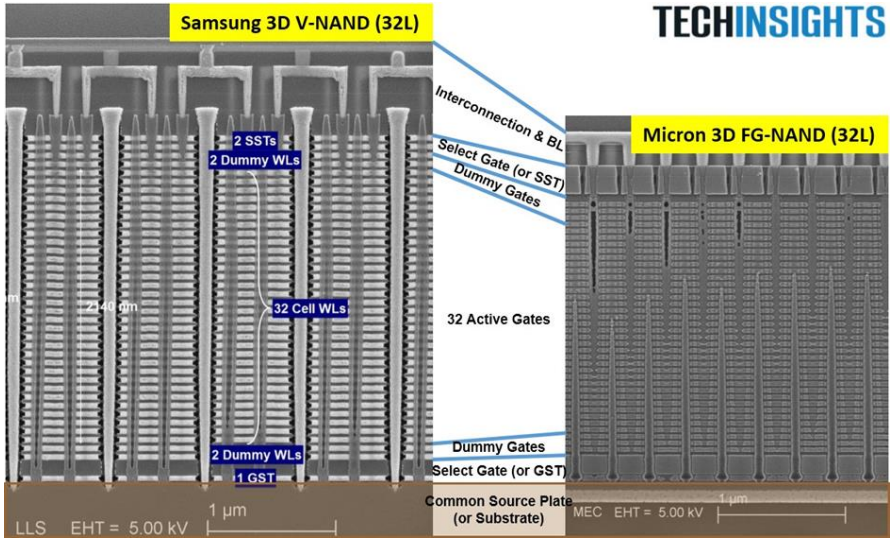
- HDD sales decreasing, related to PC sales decline
- Pressure from SSDs in the notebook area and in the enterprise performance drives (FC, 15krpm)
- Stable sales for capacity cloud drives
- HDD/SSD mergers
  e.g. WesternDigital bought SanDisk

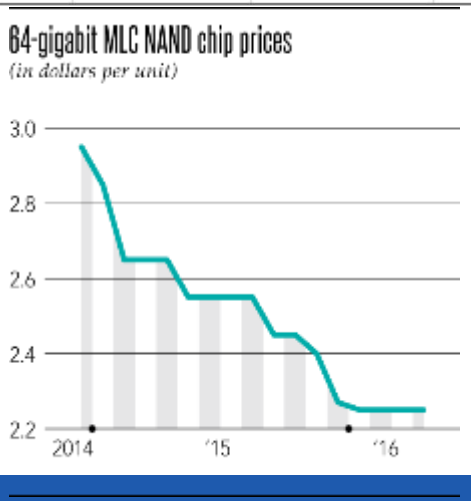
Solid-State Disks (1)

NAND:

- 2D scaling came to an end 2 years ago
- 3D: Samsung 48 layer products in the market; announced 4th generation (64 layers) for next year

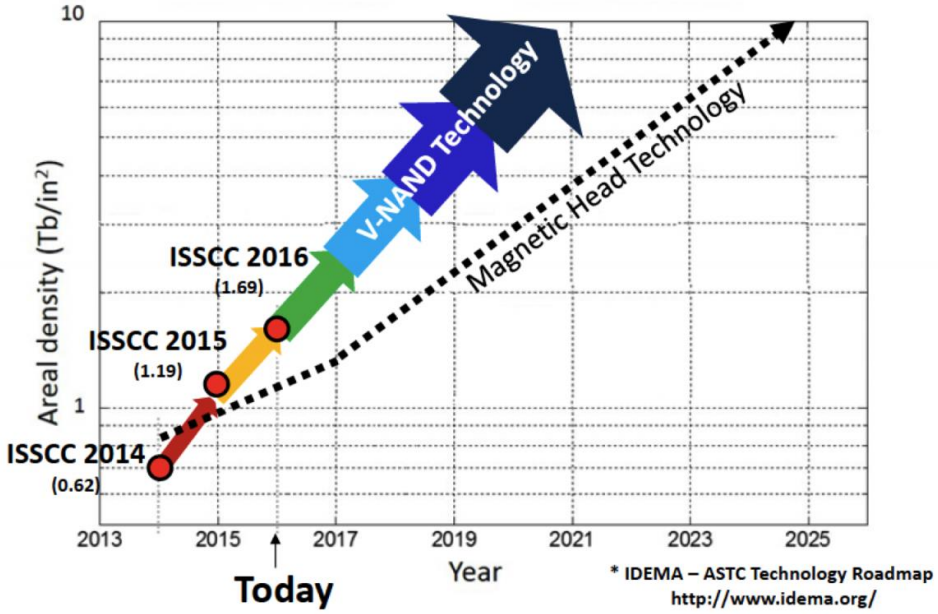






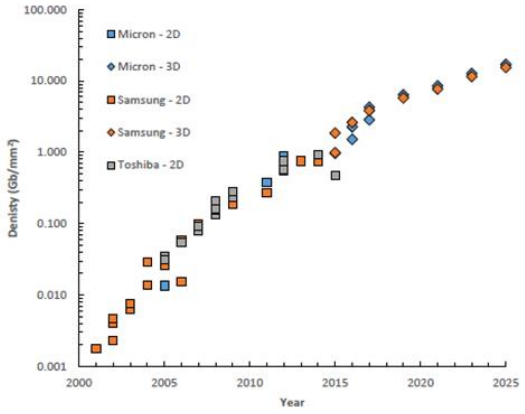
Same name, different technologies and sizes

Solid-State Disks (2)

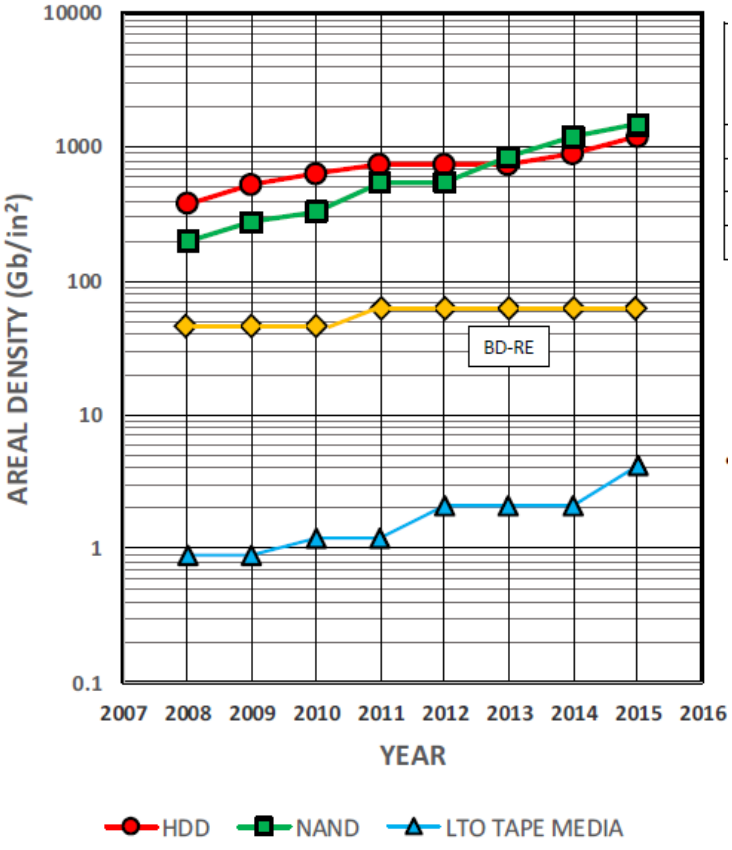


3D NAND – scaling in the third dimension

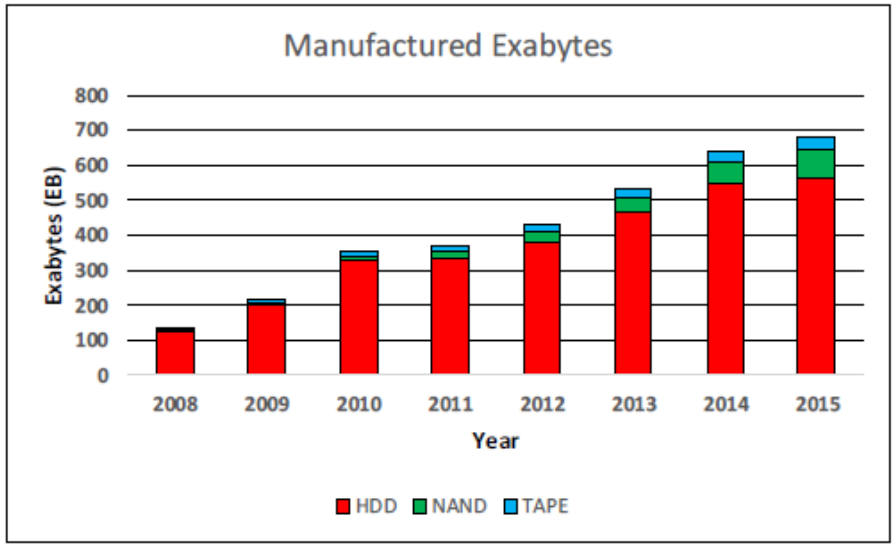
- 2D NAND scaling beyond 16nm/15nm is uneconomical.
- 3D NAND adds additional layers for scaling in place of 2D lithographic scaling.
- Bit density is continuing to scale with the potential for terabit NAND die.

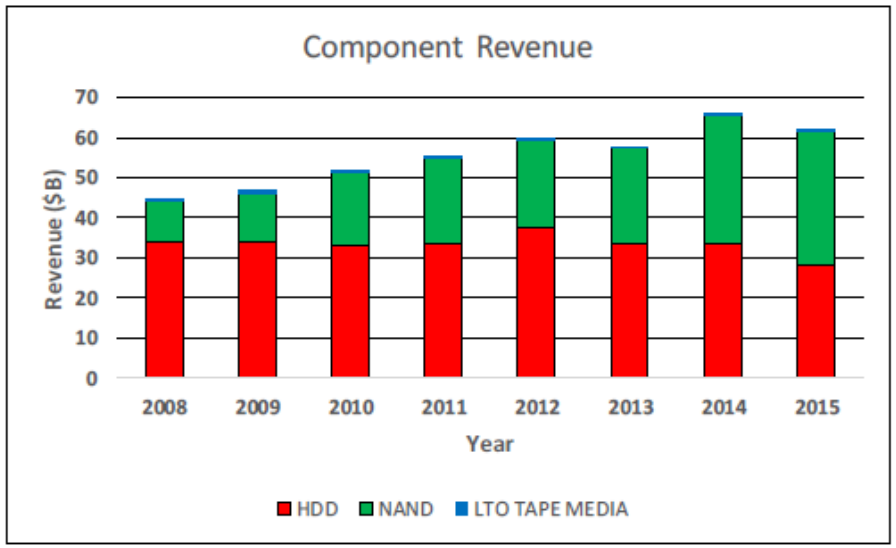


NAND density has surpassed HDD density



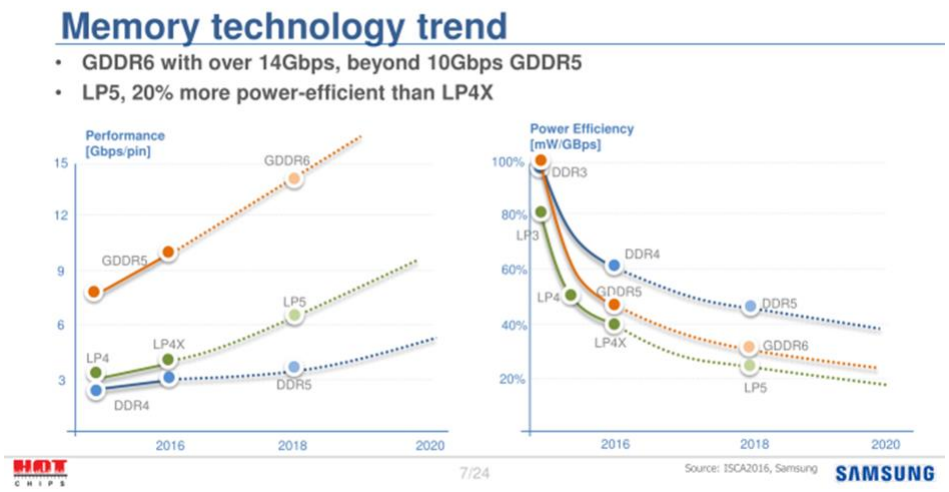
Solid-State Disks vs. Hard Disks

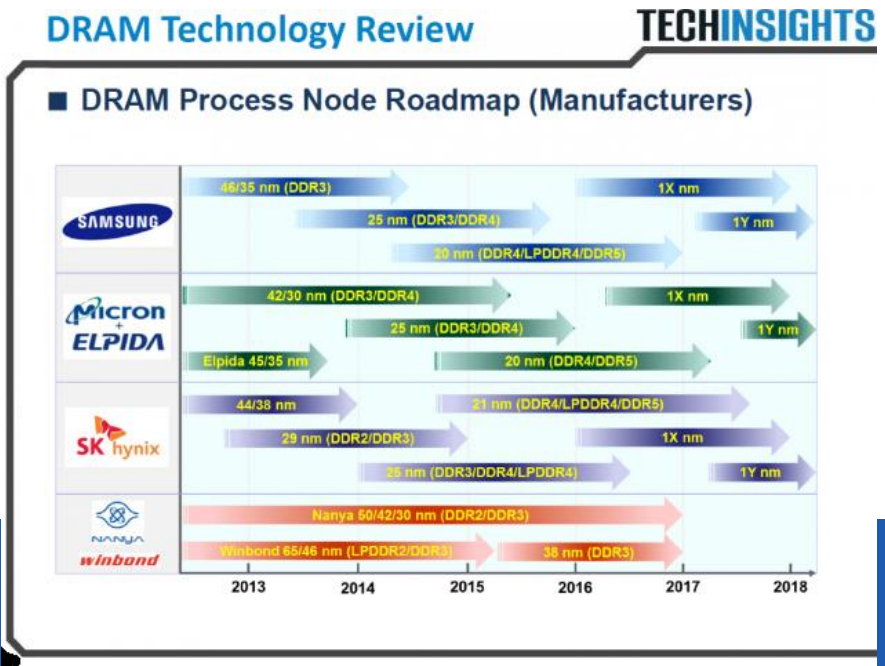




- 14 times more HDD capacity than SSD
- Price per TB decreasing about the same way
- Difference SSD/HDD costs per TB ~5-10 will slowly decrease
- Fab investment of 100-200 B$ necessary to achieve HDD ExaByte deliveries

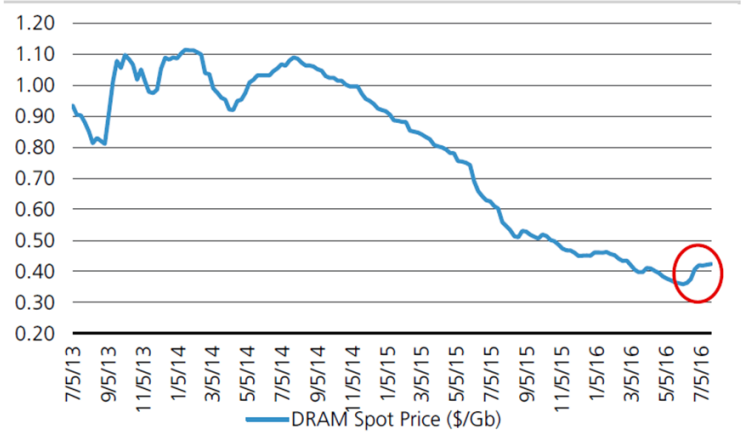
Memory: DRAM





Limited future improvements on performance and energy efficiency

Figure 1: DRAM Spot Price Trend



Source: DRAMeXchange

2 Chinese companies will enter the DRAM market in 2017

Further price decay likely

New Memory Technologies

- 3d xpoint: new technology from Intel and Micron, presumably a variant of Phase Change Memory

  Specs are changing:

  Announcement 2015: 1000x faster, 1000x endurance, 10x denser than NAND

  IDF 2016: 10x faster, 3x endurance, 4x denser than NAND

  Will enter the high end server market in Q1 2017

- Memristors: developed since 2008; HPE now collaborating with SanDisk (ReRAM)

- Spin torque MRAM in larger production units available (Everquest + Globalfoundries)

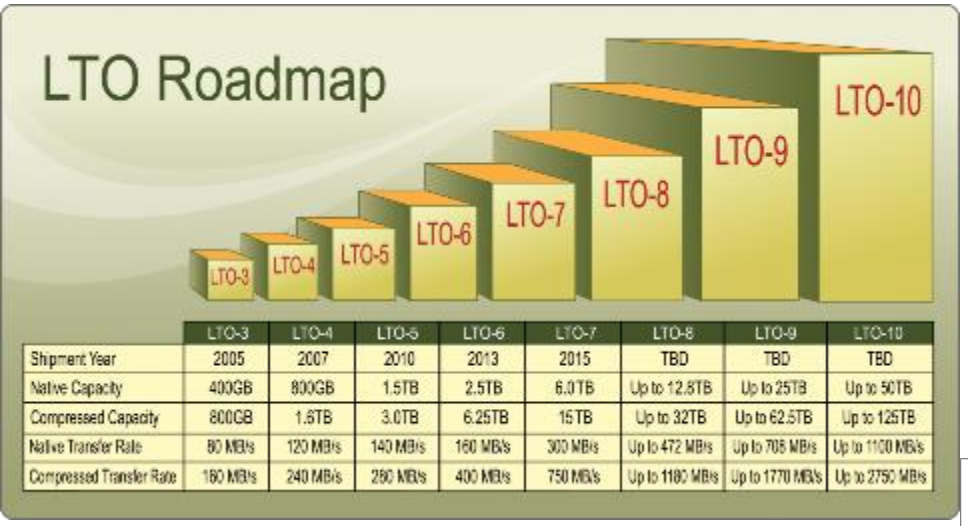
  Low density and high price

- Tantalum memory, Rice University

- RRAM or ReRAM , various new categories being developed: Oxide RAM (OxRAM), Conductive-Bridge RAM (CBRAM) or Self-Rectifying Cells (SRC)

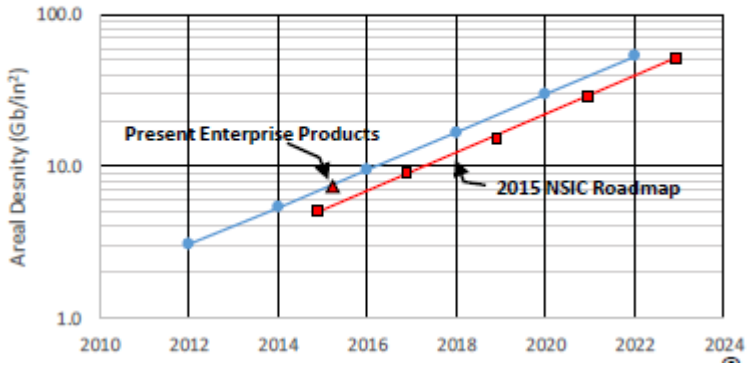
→ But… NAND fab investments are high, extended technology lifetime with 3D, hard to replace in the short term

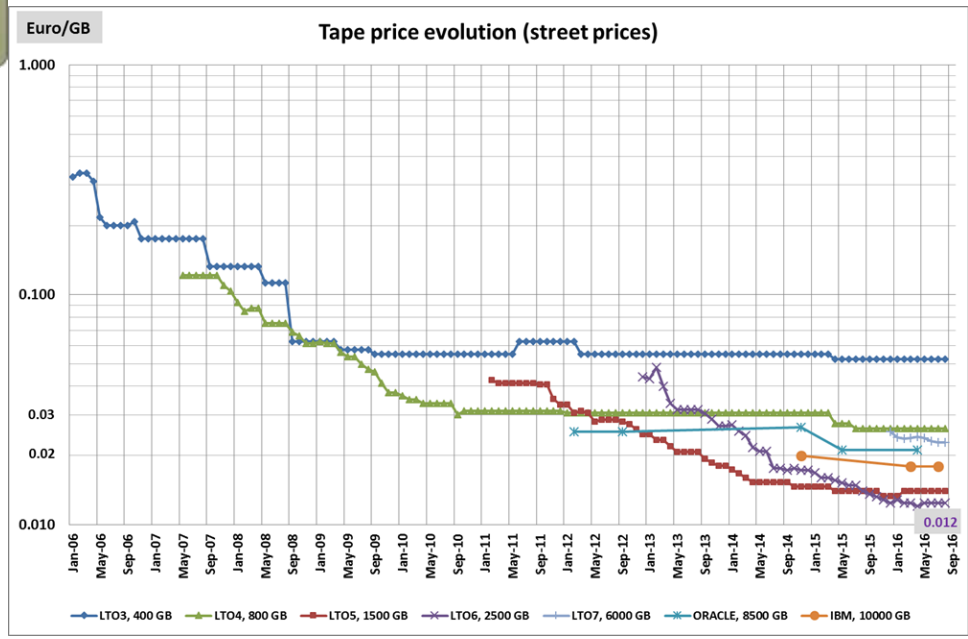
Magnetic Tapes (1)



	LTO-3	LTO-4	LTO-5	LTO-6	LTO-7	LTO-8	LTO-9	LTO-10
Shipment Year	2005	2007	2010	2013	2015	TBD	TBD	TBD
Native Capacity	400GB	800GB	1.5TB	2.5TB	6.0TB	Up to 12.8TB	Up to 25TB	Up to 50TB
Compressed Capacity	800GB	1.6TB	3.0TB	6.25TB	15TB	Up to 32TB	Up to 62.5TB	Up to 125TB
Native Transfer Rate	80 MB/s	120 MB/s	140 MB/s	160 MB/s	300 MB/s	Up to 472 MB/s	Up to 708 MB/s	Up to 1100 MB/s
Compressed Transfer Rate	160 MB/s	240 MB/s	280 MB/s	400 MB/s	750 MB/s	Up to 1180 MB/s	Up to 1770 MB/s	Up to 2750 MB/s

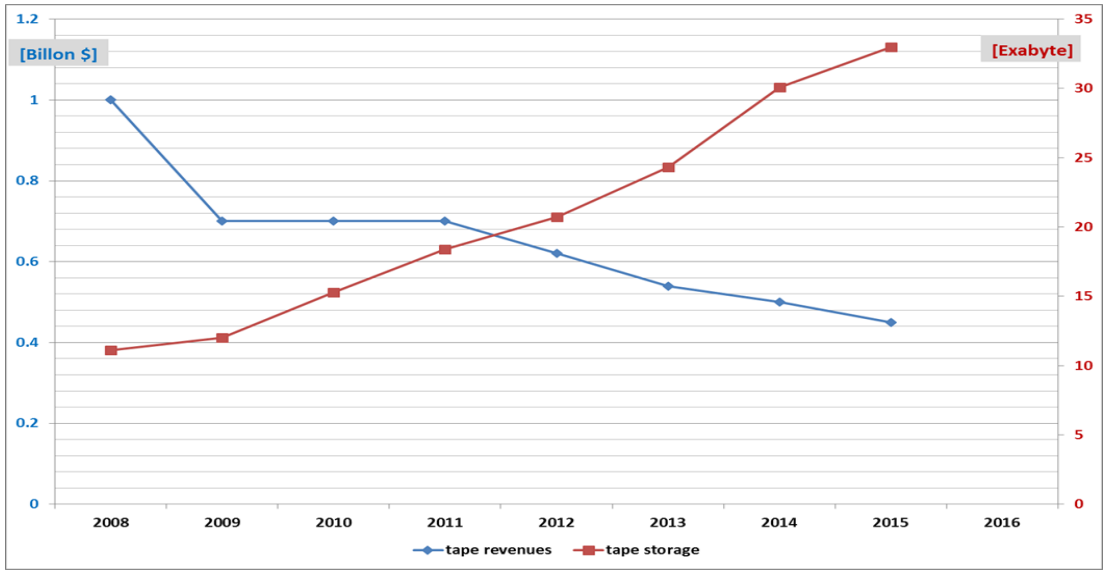
- TAPE: source NSIC 2013

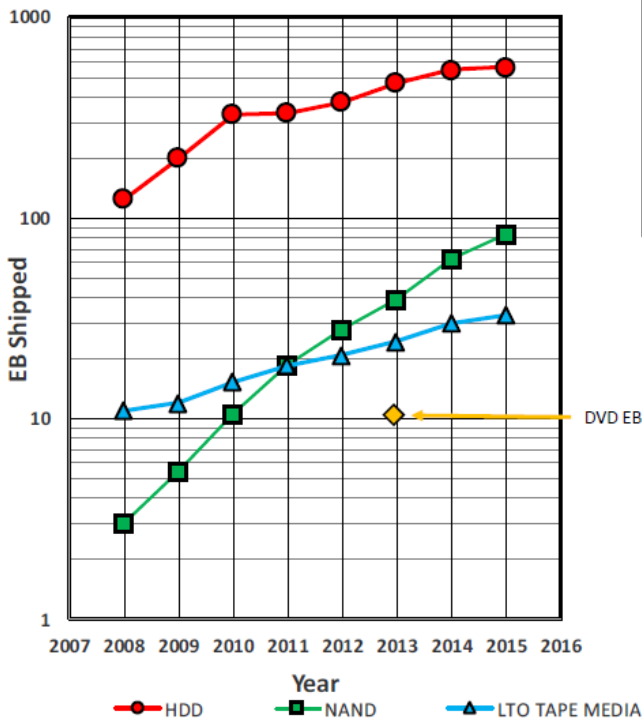




- Enterprise drives:
  Oracle 2017: 8.5 TB → 12 TB
  IBM 2018: 10 TB → 16 TB
- Technology in the lab:
  Fujifilm 154 TB, Sony 185 TB,
  IBM 220 TB
- Good improvements of price/capacity

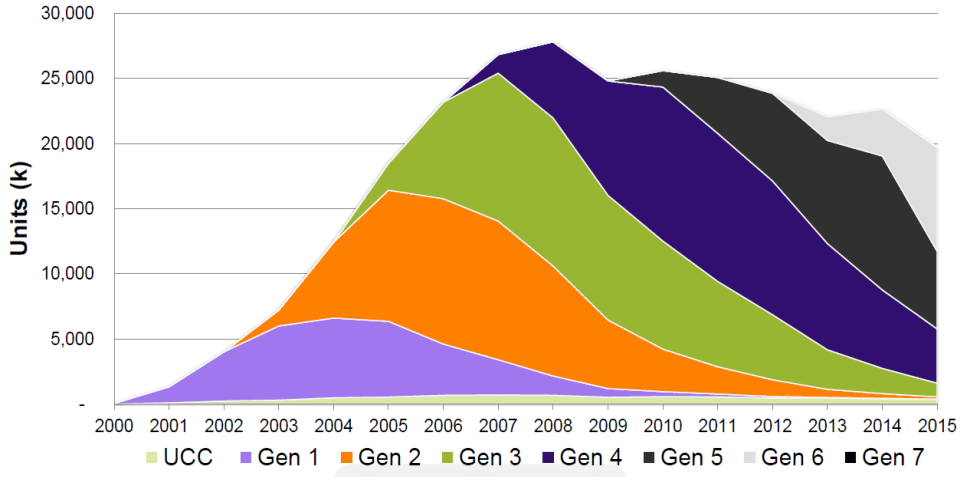
Magnetic Tapes (2)





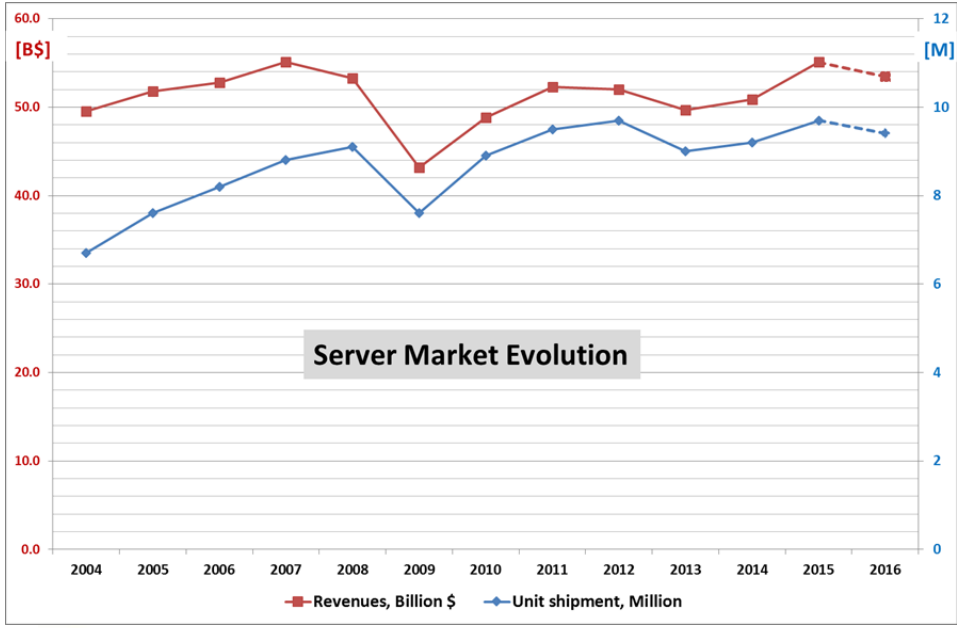
- More NAND than LTO shipped
- Steady decrease of tapes shipped and revenues
- Will Oracle and/or IBM sell or drop these products?

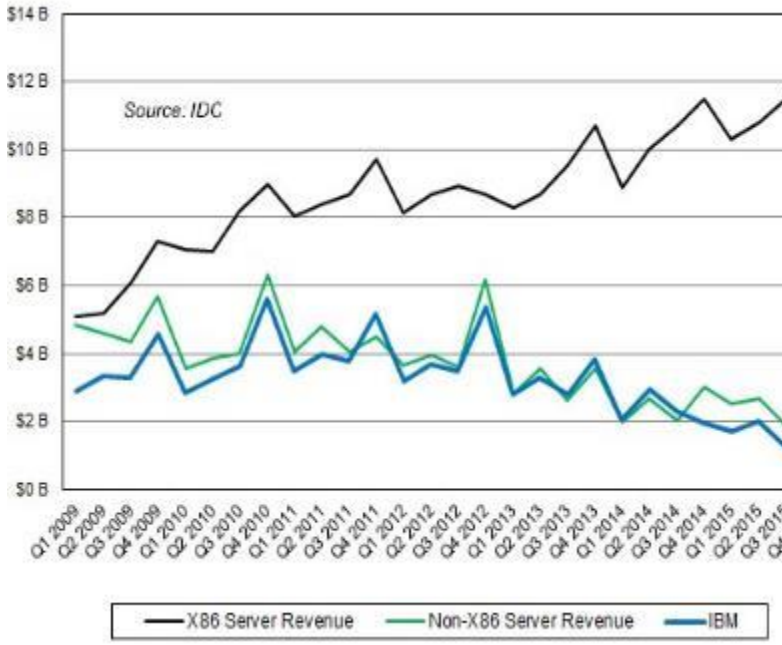
Unit Shipments: Calendar Year



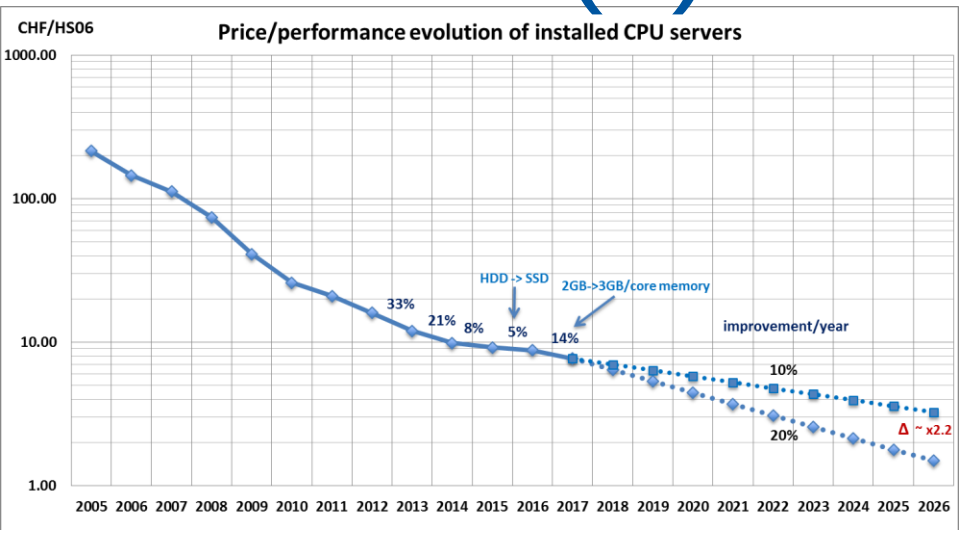
Servers (1)

- Server market is saturated: flat revenues and unit shipments
- High profit market
- Single vendor: Intel, 99% market share
- Several initiatives to change that:
- OpenPower (IBM): consortium with many members
  - But revenues still going down, little impact so far
  - Announcement of POWER9 might help
- ARM server:
  - AppliedMicro , Qualcomm, Cavium: new high end products Announcements for 2H2017 (third ARMv8 Wave 2017-2018), First two waves had little impact
  - Phytium (China), “Mars” processor
- AMD with new processor design (Zen) in 2017
- Fujitsu ARM-powered supercomputer
  - Add large vector instructions to the ARM design
  - Aimed for 2020, now ~2022





Servers (2)



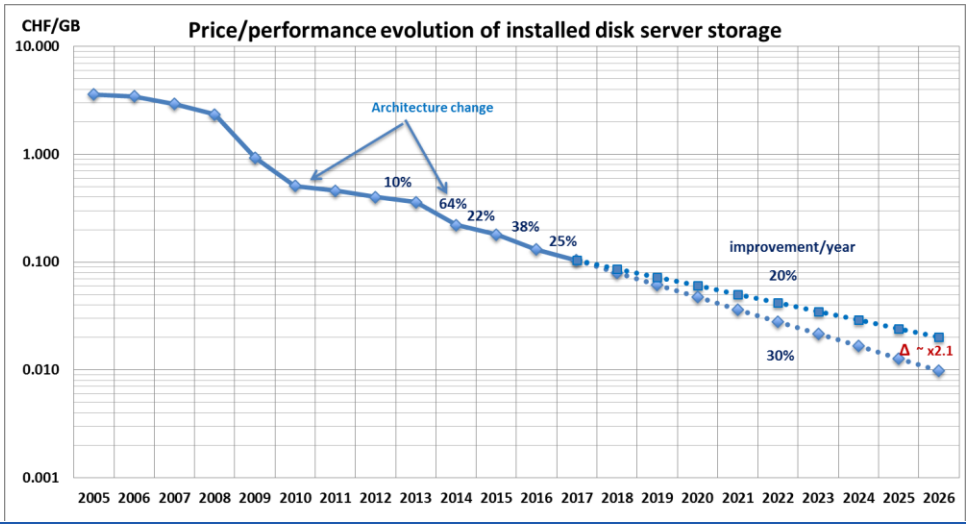
Preliminary extrapolation of CPU and disk server costs (based on CERN procurements)

Pessimistic and reasonable improvement extrapolations

Influence of changing software and hardware architecture requirements to be taken into account (programs, data model, data centre, …)

e.g. CERN moves from 2 to 3 GB/core (+8% cost), Driven by experiment usage AND technology boundary conditions

- Moore's Law and Kryder's Law are slowing down
  - 18 months → >= 3 years
- Real cost/performance evolution driven by financial and market aspects rather than technology



Summary (1)

- Device markets (smartphones, tablets, PCs, notebooks, servers, HPC) saturated – negative growth
  - Replacement market
- Moore's Law in trouble, financial issues
  - Not clear how this effects price/performance evolution
  - So far okay for CPU and disk servers
- Technology improvements still continuing, but requires high CAPEX

  End-product price tag evolution more complicated
- Market dominance of few companies increases, competition diminishing

References

http://electroiq.com/blog/2016/05/global-semiconductor-sales-increase-slightly-in-march/
http://www.statista.com/statistics/266219/global-smartphone-sales-since-1st-quarter-2009-by-operating-system/
http://www.statista.com/statistics/263393/global-pc-shipments-since-1st-quarter-2009-by-vendor/
http://www.statista.com/statistics/276651/global-media-tablet-shipments-since-3rd-quarter-2011-by-vendor/
https://www.ericsson.com/res/docs/2016/ericsson-mobility-report-2016.pdf
http://www.nasdaq.com/article/the-evolution-of-smartphone-markets-where-growth-is-going-cm619105
http://www.potomacinstitute.org/steps/images/PDF/Articles/FritzeSTEPS_2016Issue3.pdf
http://www.pcper.com/news/Processors/Intel-officially-ends-era-tick-tock-processor-production
http://semimd.com/chipworks/
http://www.extremetech.com/extreme/223022-the-myths-of-moores-law
http://www.forbes.com/sites/gartnergroup/2016/08/29/track-three-trends-in-the-2016-gartner-hype-cycle-for-emerging-technologies/#59fc4d787286
http://jonpeddie.com/publications/market_watch
http://www.anandtech.com/show/10613/discrete-desktop-gpu-market-trends-q2-2016-amd-grabs-market-share-but-nvidia-remains-on-top
http://www.computerworld.com/article/3041947/data-storage/how-these-technologies-will-blow-the-lid-off-data-storage.html
http://www.computerworld.com/article/2852233/want-a-100tb-disk-drive-youll-have-to-wait-til-2025.html
http://www.anandtech.com/show/9866/hard-disk-drives-with-hamr-technology-set-to-arrive-in-2018
http://www.theregister.co.uk/2016/05/31/hdd_revenues_to_plummet_as_ssd_penetration_rises/
http://www.anandtech.com/show/10315/market-views-hdd-shipments-down-q1-2016
http://www.trendfocus.com/ssd-cq116_update/
http://www.forbes.com/sites/tomcoughlin/2016/02/03/flash-memory-areal-densities-exceed-those-of-hard-drives/#5dc24d2b4026
http://www.anandtech.com/show/10589/hot-chips-2016-memory-vendors-discuss-ideas-for-future-memory-tech-ddr5-cheap-hbm-more
http://asia.nikkei.com/Business/Trends/NAND-flash-memory-prices-likely-to-climb-again
http://www.techinsights.com/techinsights/about-techinsights/articles/deep-dive-into-the-intel-micron-3D-32L-FG-NAND/
http://amigobulls.com/articles/micron-technology-inc-stock-is-the-next-big-idea-for-2016
http://wccftech.com/micron-compete-samsung-16-nm-dram/
http://storageconference.us/2016/Slides/BobFontana.pdf
http://searchsolidstatestorage.techtarget.com/feature/New-memory-technologies-generate-attention-as-successors-to-NAND-flash
http://www.itjungle.com/tfh/tfh061316-story05.html
http://www.forbes.com/sites/tomcoughlin/2016/01/15/digital-storage-projections-for-2016-part-2/2/#35b1916a3aa8
http://www.lto.org/wp-content/uploads/2016/03/LTO_Media-Shipment-Report_3.22.16.pdf
http://semiaccurate.com/2016/09/12/intels-xpoint-pretty-much-broken/