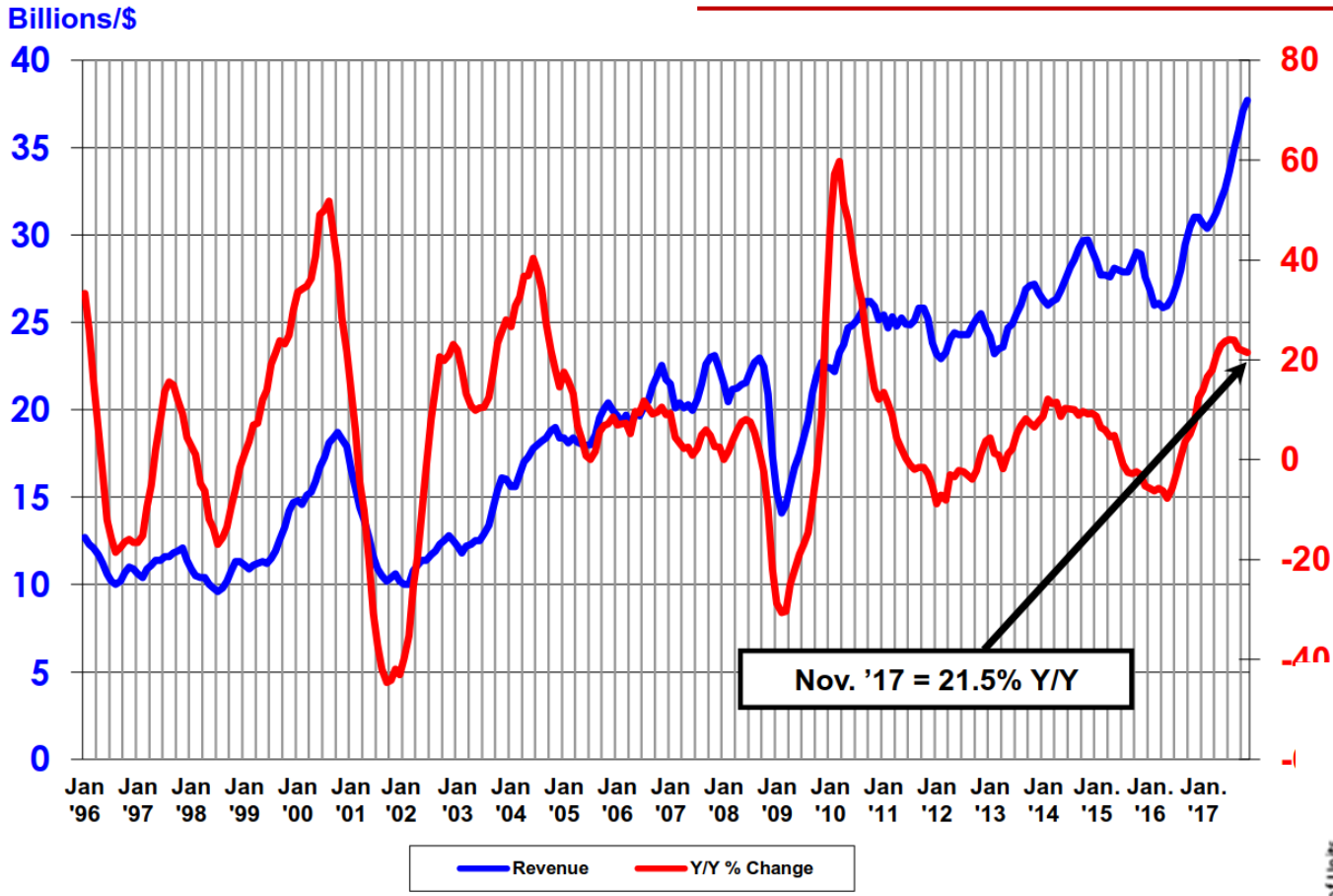
IT Technology and Markets, Status and Evolution

General Semiconductor Market



Source: WSTS

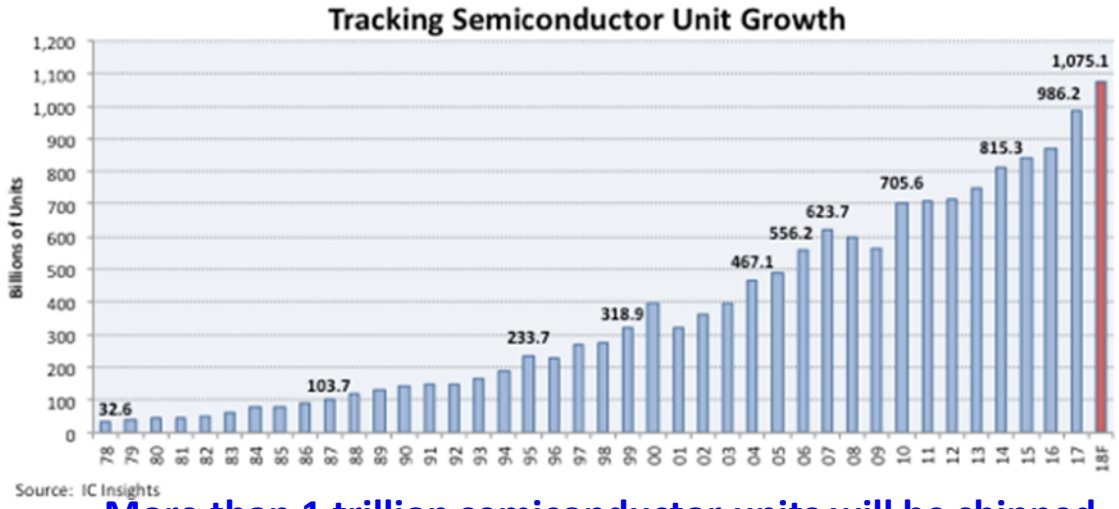
Large jump (21.5%) in revenue and growth rate in 2017 which will partly continue in 2018
Major reason: RAM price increase by 120%

Total revenue per year has now crossed 400 B

Biggest semiconductor company is now Samsung (61 B$), followed by Intel (58 B$)
10 companies make >50% of revenues

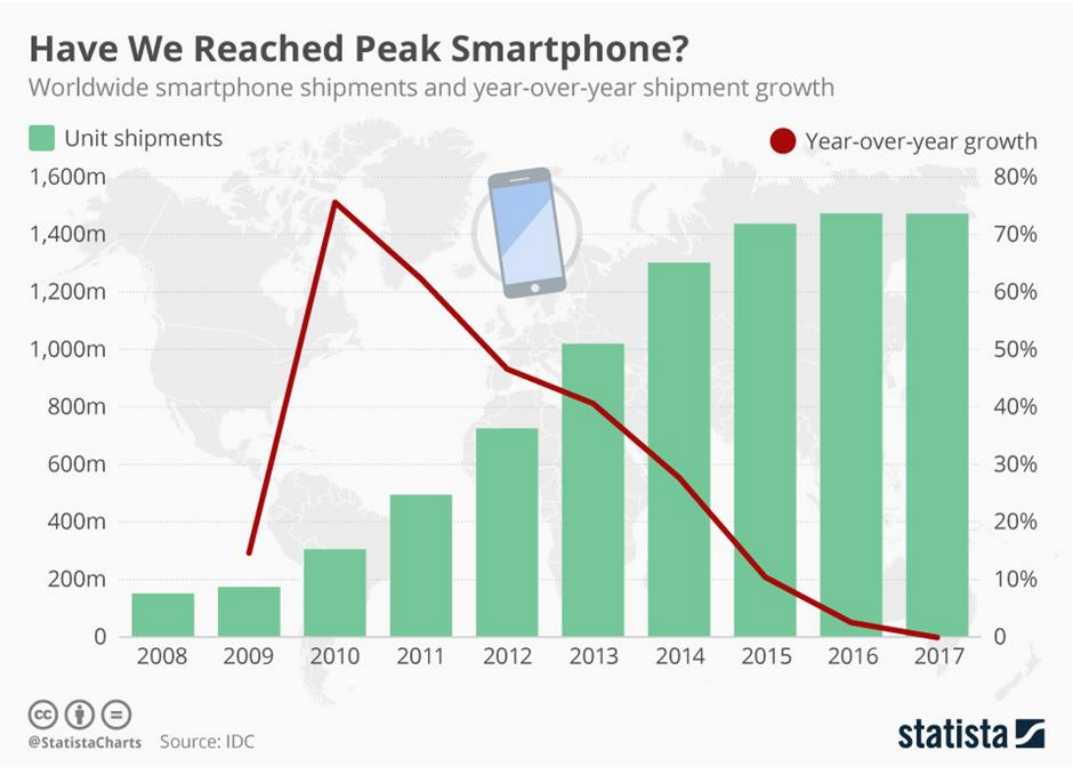
Total IT spending expected in 2018: 3.7 trillion $, 4.5% growth rate

Wafer supply companies have increased prices in 2017 and will increase by 20% in 2018 and further in 2019

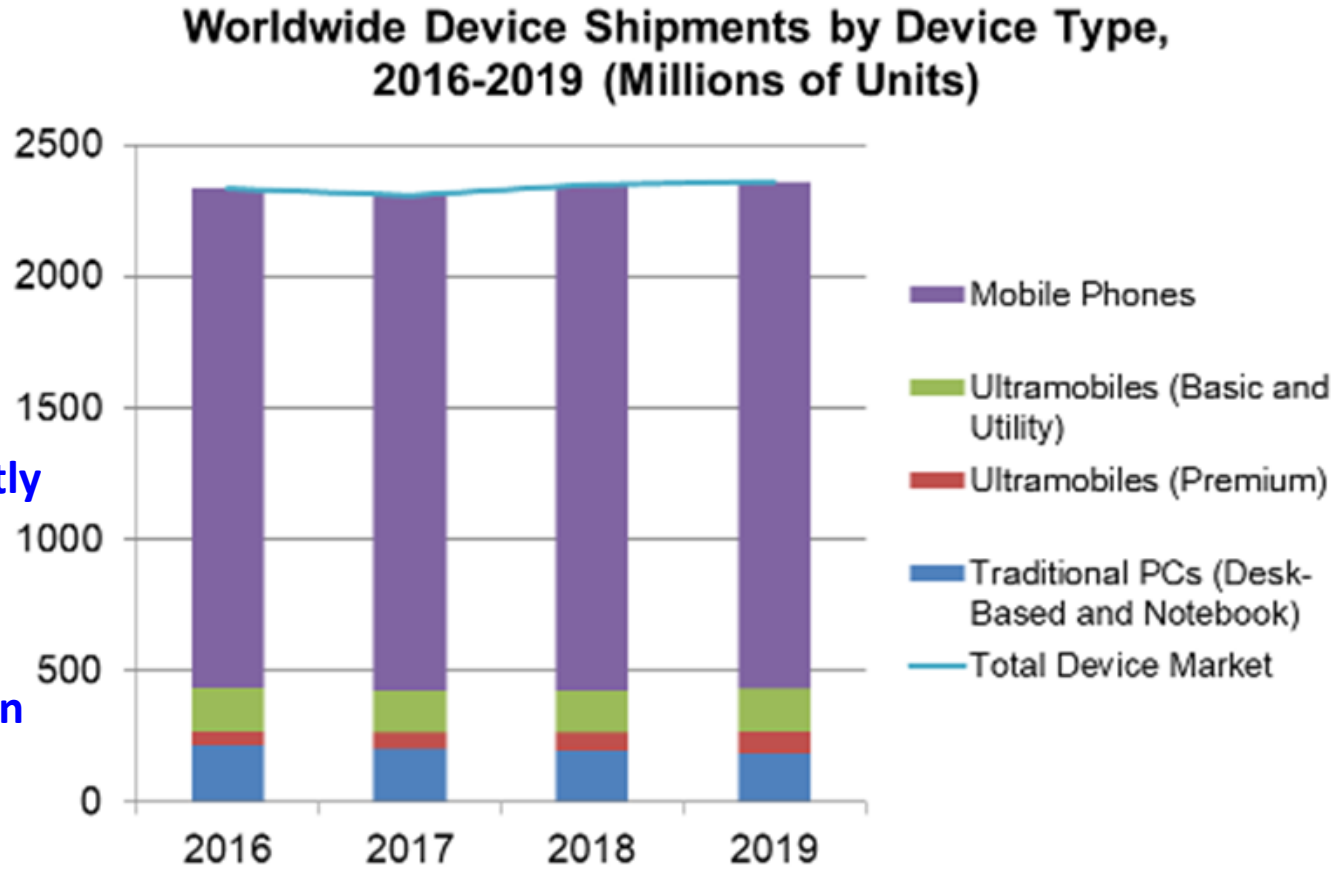


More than 1 trillion semiconductor units will be shipped in 2018 (12% are microprocessors, DRAM, NAND, etc)

Device Markets

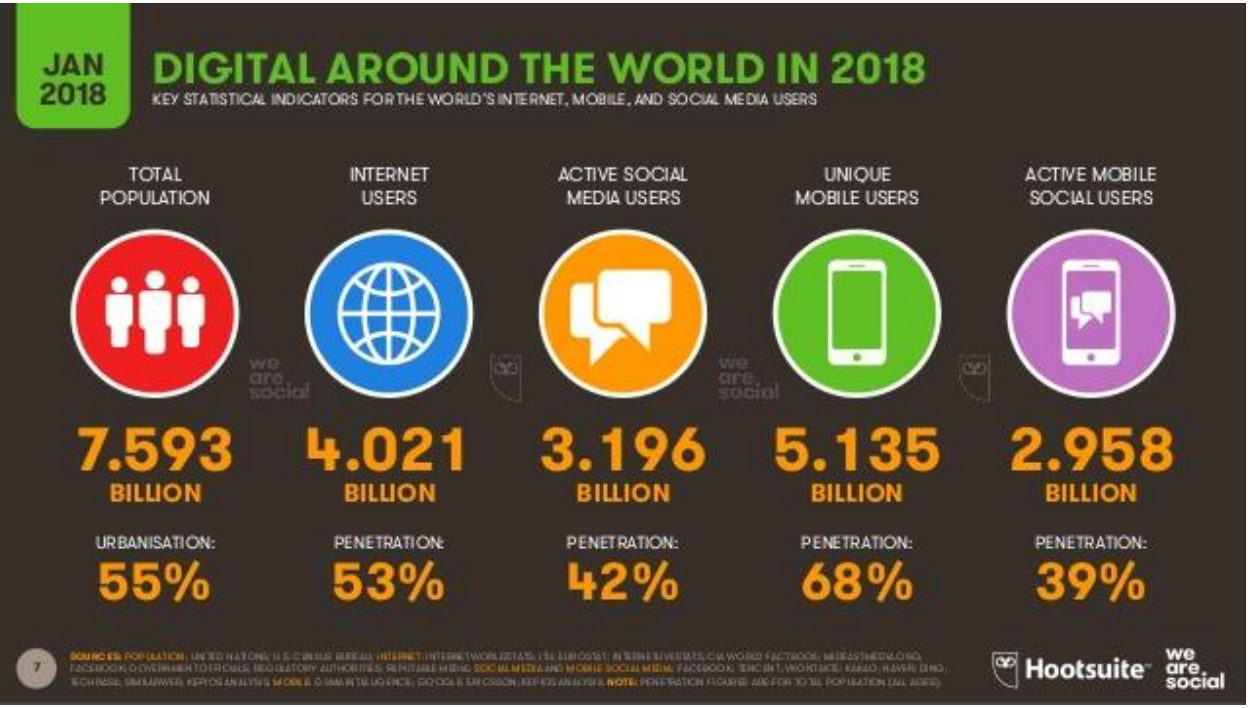


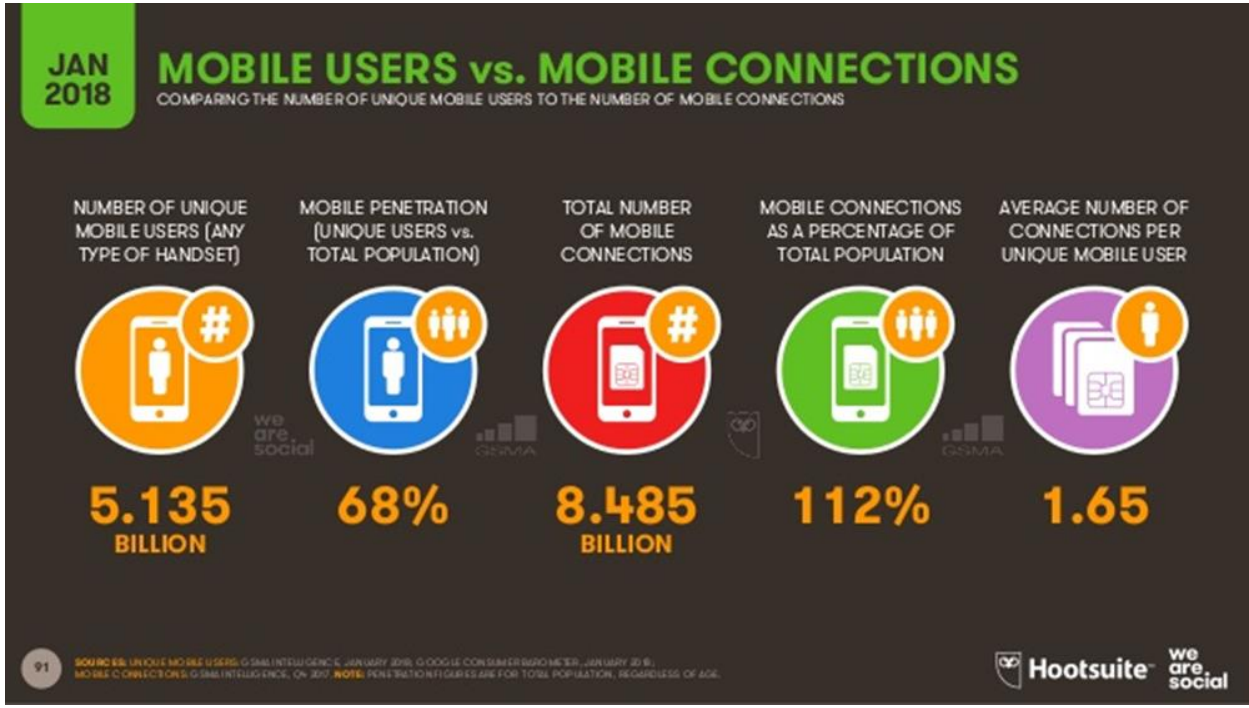
Overall Computing Device market is flat, becomes replacement market



- **PCs, notebooks and tablets sales declining constantly**
- **Smartphones sales are flat**
- **Attractiveness of replacement is decreasing**
  Only marginal differences between smartphone models and generations , small and little innovation
  → Lifetime increase

World Internet Population





Saturation effects

Limited growth rates for internet devices due to limited growth rate of internet users

Already high market penetration in the population

68% of the world population have a phone

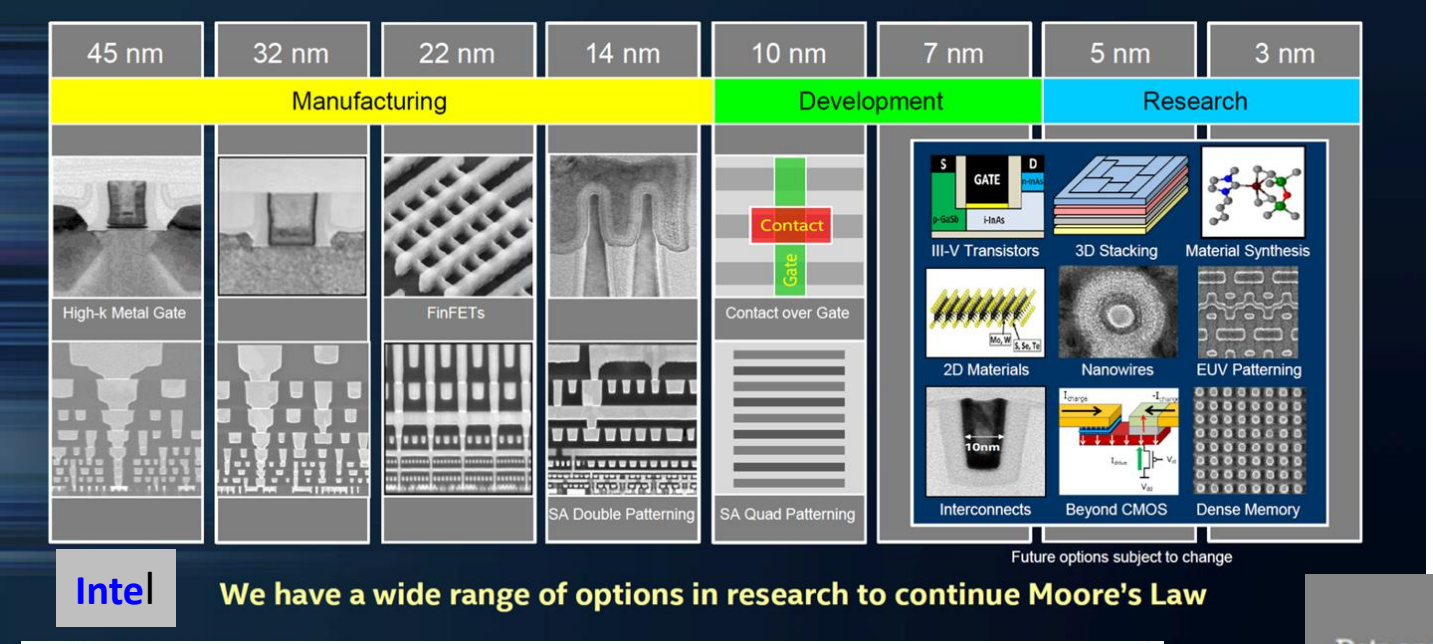
Processor Technology

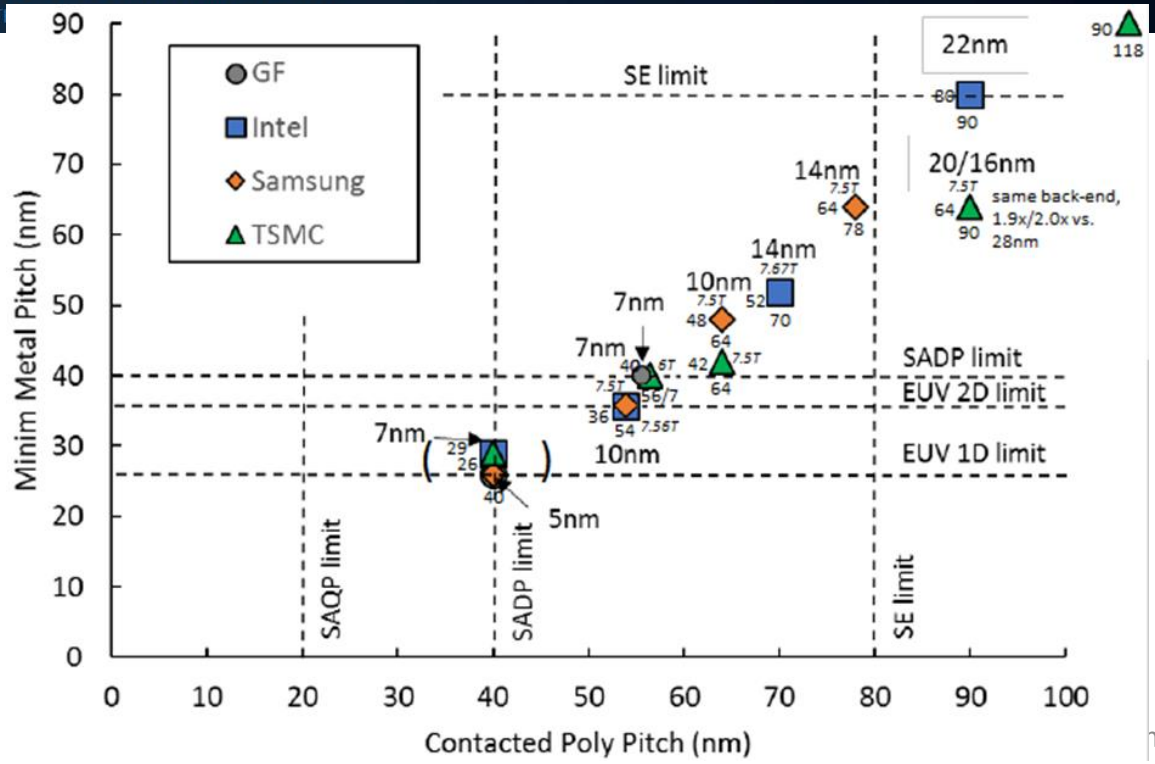
Intel has problems with their 10nm process

TSMC building fab 18 for their 5nm process,
Will be finished in 2020; 950000 m2 for 17 B$

There is no norm for the process names:
10 nm Intel compares to a 7 nm Samsung/TSMC process

Below 7nm new technologies are needed
(nanowires, non-silicon materials), very expensive





Industry FinFET Lithography Roadmap, HVM Start

Data announced by companies during conference calls, press briefings and in press releases

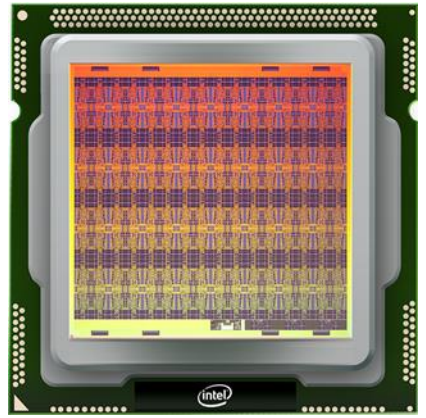
	2016	2017		2018		2019		2020		2021
		1H	2H	1H	2H	1H	2H	1H	2H	
GlobalFoundries	14LPP				7nm DUV	7nm with EUV*				
Intel	14 nm 14 nm+		14 nm++ 10 nm			10 nm+ 10 nm++				
Samsung	14LPP 14LPC	10LPE		10LPP		8LPP 10LPU	7LPP		6 nm* (?)	
SMIC	28 nm**	14 nm in development								
TSMC	CLN16FF+ CLN16FFC	CLN10FF CLN16FFC		CLN7FF CLN12FFC		CLN12FFC/ CLN12ULP	CLN7FF+		5 nm* (?)	
UMC	28 nm**		14nm	no data						

*Exact timing not announced
**Planar

New Processor Architectures

There is a plethora of new processor designs, all with a focus on Machine Learning:

- **Intel: Mobileye EyeQ5 (vision processing, autonomous cars), Nervana Neural Network Processor, Movidius MyriadX VPU**
- **ARM: Project Trillium, Machine Learning processor, Object Detection processor**
- **Graphcore IPU (Intelligent Processing Unit)**
- **Google second generation of Tensor Processing Unit TPU**
- **NeuPro AI processor from CEVA**
- **Neuromorphic chips from IBM (TrueNorth, 64 M neurons + 16 B synapsis) and Intel (Loihi, 130 K neurons + 130 M synapsis)**
- **Nvidia is enhancing their graphics cards, Titan V (110 Tflops Deep Learning), Xavier (SoC, 20 TOPS, vision accelerator)**



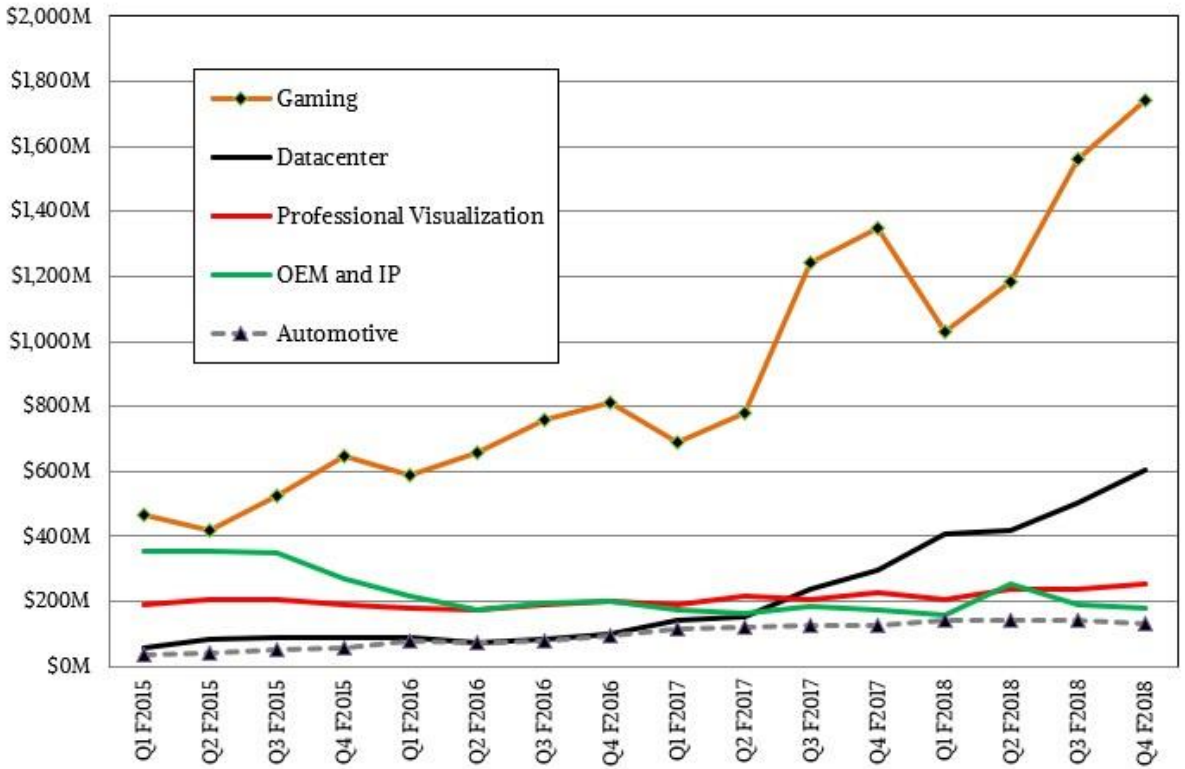
All high-end smartphones are integrating AI chip enhancements (Qualcomm-neural processing engine, Apple- A11 Bionic chip, etc.)
The market for these special chips will reach 5-10 B$ in 2022

The keyword is LOCAL data processing also major impact on IoT
→ much less network, cloud storage and cloud processing needed

Accelerators

GPUs

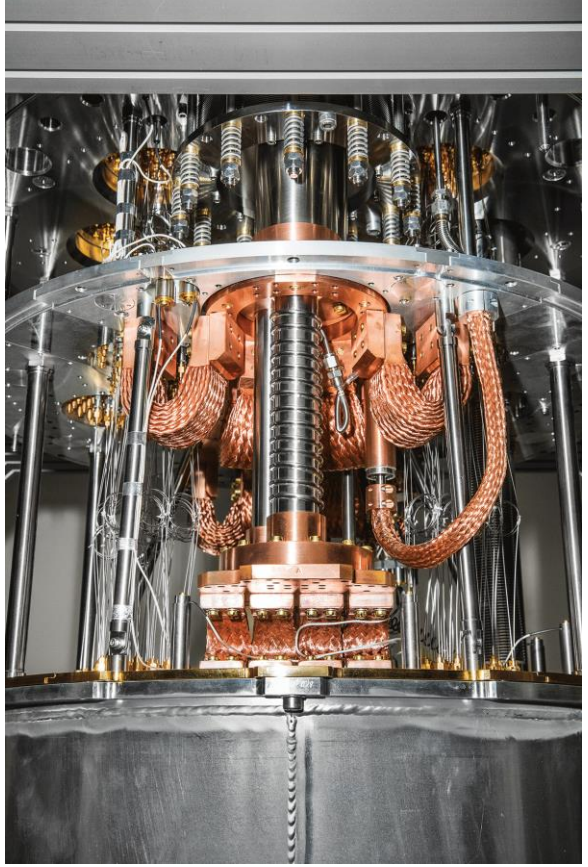
- Dedicated Graphics Cards market leader in Nvidia
- High end card Tesla V100 (14 TFlops SP, 110 Tflops ML, 12nm process)
- Gaming key driver for the market (plus AI and crypto mining).
- Large price increases (up to x2), crypto mining + high memory prices
- Change license policy → no gamer cards in the center
  (GTX 1080 TI = 700$ → Tesla V100 9000$, but also DP performance 20-30 higher)



1 B PC gamer worldwide
PC gaming hardware market 35 B$, total gaming market 165 B$

other accelerators:

- Intel stopped the Xeon Phi line (Knights Mill last product)
  Not clear what the replacement will be
- Microsoft Project brainwave, based on Intel (Altera) Stratix FPGA
- Xilinx ACAP , Project Everest many-core SoC, programmable DSPs
  50 B transistors, TSMC 7nm process
- Chinese Matrix-2000 DSP accelerator for Exascale HPC
  (current No 1 supercomputer)



Quantum Computing

Considerable progress during the last 2 years. Number of qubits sharply rising
→ Intel 49-qubit, IBM 50 , Google 72 for a Quantum gate computer
→ D-wave 2000 qubits, but not a general quantum computer (e.g. no shor's algorithm, no factorization)

Various implementations from Ion traps to silicon, focus is on silicon to re-use the fabrication process of standard chips
Coherence time is still well below 1 ms, limits the time for quantum calculations

Key problem is the Error handling:
Mitigate by combining qubits

- N physical qubits == one logical qubit , where N varies between 10 and 10000
- Use error correction in software, deal with approximative results
  Machine learning algorithms

The Programming model is completely new; not clear how many algorithm can be 'converted' for a quantum computer;
Very, very high cost structure

→ prognosis: Irrelevant for HL-LHC

Optical Computing

Renaissance of optical computing, this time focused on neural networks

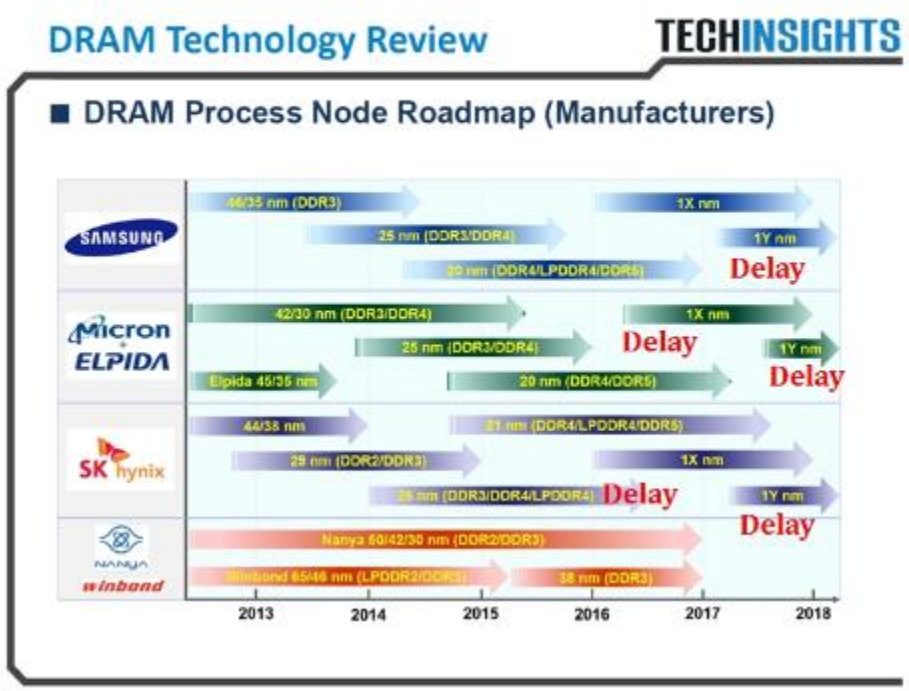
- Optalysis First implementation of a Convolutional Neural Network with Optical Processing Technology
- Lightelligence Deep learning with coherent nanophotonics circuits
- Lightmatter Photonics for AI

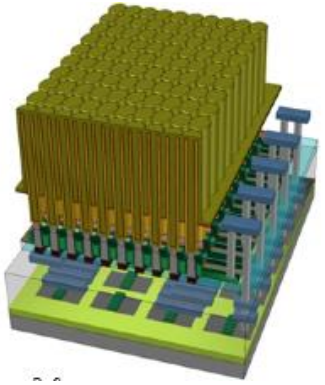
DRAM Memory

~70 B$ market

DRAM price increase during the last 18 month: ~120%
Trend to increase a further 5-10% in 2018

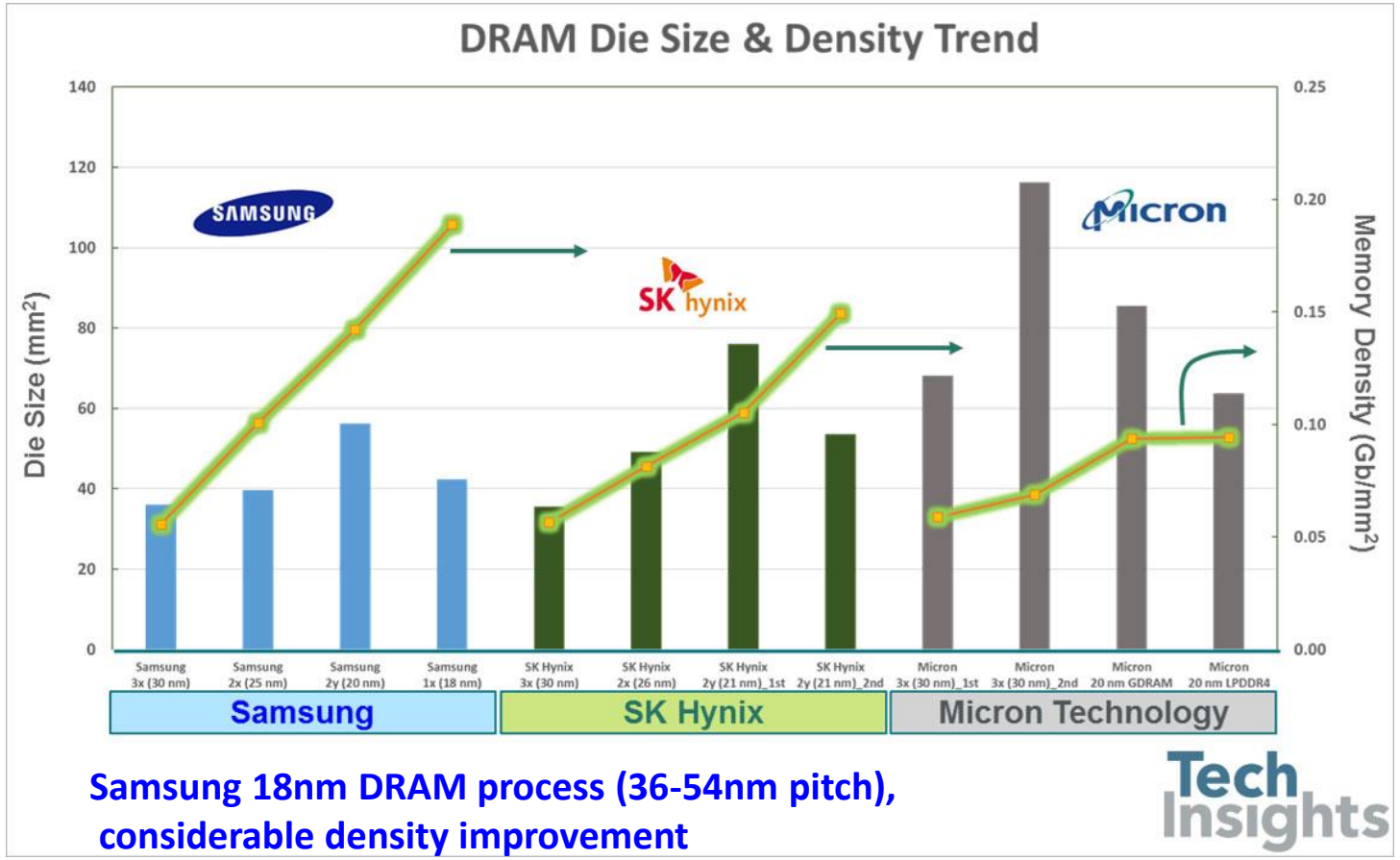
DRAM Roadmap Plan vs. Reality





DRAM scaling slowed down
Capacitor aspect ratio increases exponentially with smaller cell size
→ much higher Fabrication costs

3D DRAM not yet available



Samsung 18nm DRAM process (36-54nm pitch),
considerable density improvement
But it is linear not exponential !

New Memory Technologies

Technology Comparison

SNIA PERSISTENT MEMORY PM SUMMIT
JANUARY 24, 2018 | SAN JOSE, CA

Technology	FeRAM	MRAM	ReRAM	PCM	DRAM	NAND Flash
Nonvolatile	Yes	Yes	Yes	Yes	No	Yes
Endurance	10^{12}	10^{12}	10^{6}	10^{8}	10^{15}	10^{3}
Write Time	100ns	~10ns	~50ns	~75ns	10ns	10µs
Read Time	70ns	10ns	10ns	20ns	10ns	25µs
Power Consumption	Low	Medium/Low	Low	Medium	Very High	Very High
Cell Size (f^2)	15-20	6-12	6-12	1-4	6-10	4
Cost ($/Gb)	$10/Gb	$30-70/Gb	Currently High	$0.16/Gb	$0.6/Gb	$0.03/Gb

14

Several contenders for a new memory technology
Ideally replacing DRAM and NAND at the same time.

No cost effective solution yet.

Resistive RAM, 40nm process, Fujitsu/Panasonic
Aimed at Neuromorphic Computing

Magnetic RAM, 80nm process , Everspin, first 1-2GB SSDs

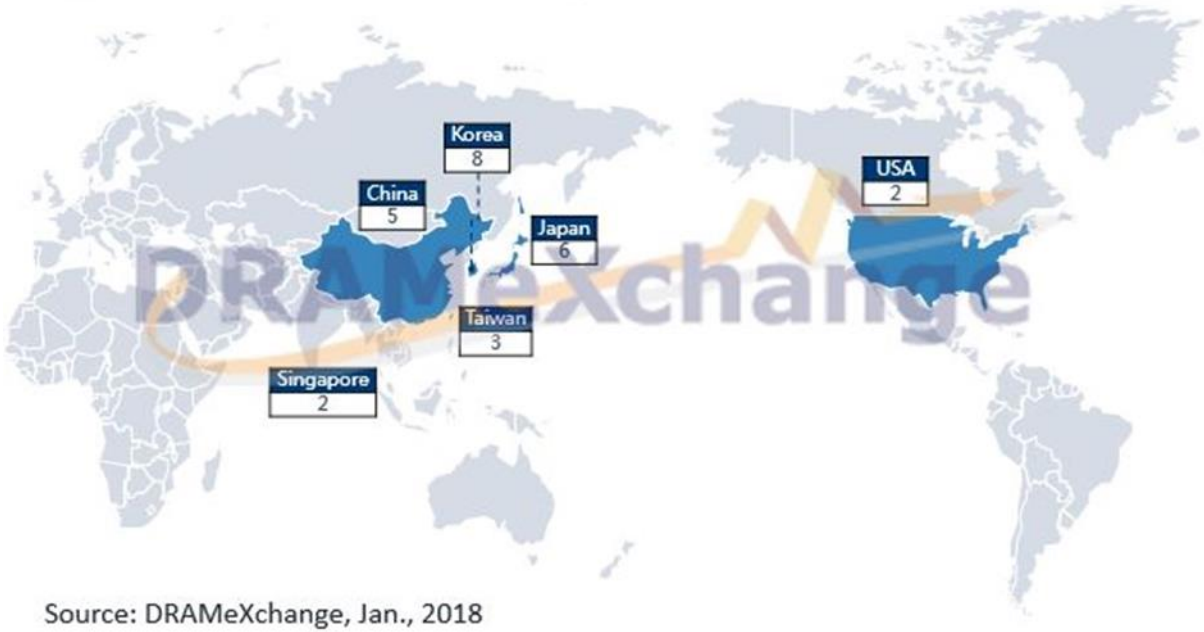
PCM Intel Optane, in production, but focus not clear

Ferroelectric RAM, very small scale products, difficult to scale

NAND Storage

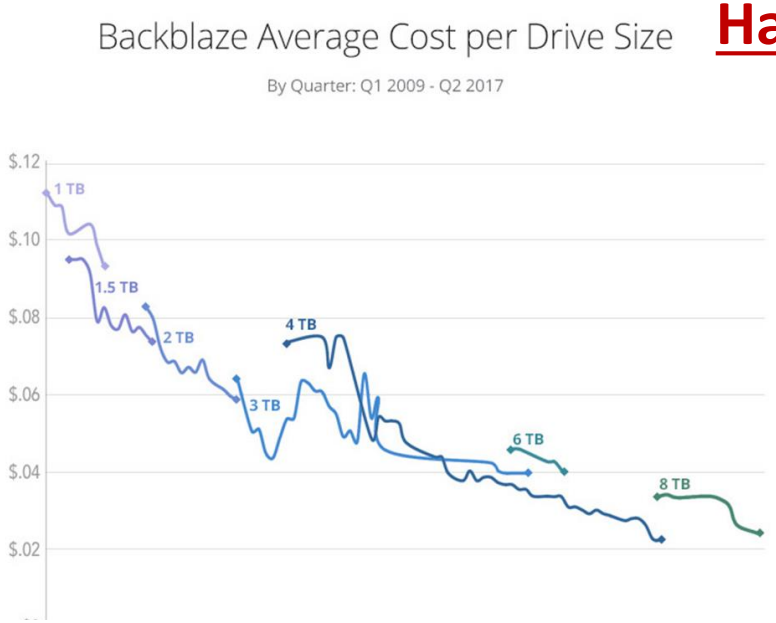
- 60 B$ market
- Fabrication moved from 2D to 3D, 64 layers in the market, 96 layer production started, 128 layers expected for 2020
- NAND prices increased over the last 18 month, high request for smartphones and SSDs (Apple busy 20% of the world-wide NANDs), now 50% of all sold NAND chip are 3D, in 2017 the largest consumer of NAND chips were SSDs (surpassing smartphones)
- 2018 price trend not clear +- 5-10% price in- or decrease → 4 new Chinese fabs will start production this year
- 4-bit cells are now feasible with 3D : ECC code easier with 2D cell size increased ; lab demos exist with hundreds of layers
- investment 3D fabrication process is up to 5x higher than 2 D, ~10B$ for fabrication facility
- Technical challenges: > 64 layers show exponential scaling problems (current density, cell uniformity)
  a wafer stays up to 3 month in the fab before the 100 defect-free layers are done
- Density improvements are now linear, adding 8/16/32 layers

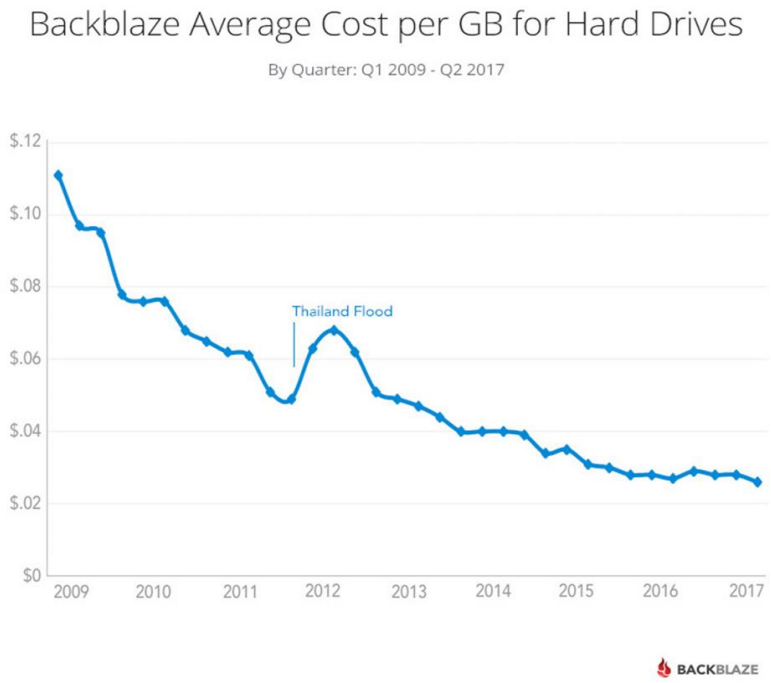
Figure: NAND Flash Factories Map in 2020

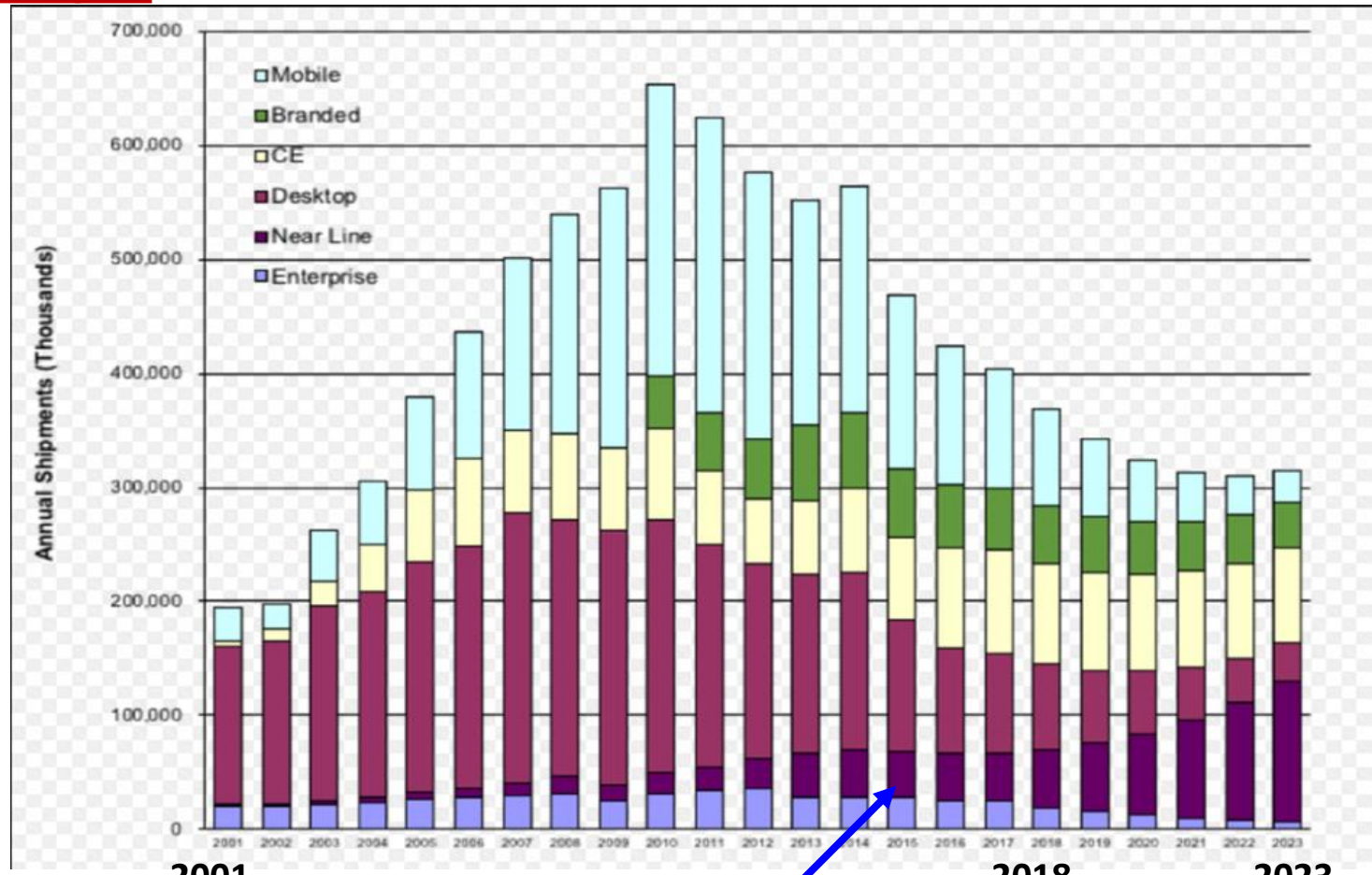


Source: DRAMeXchange, Jan., 2018

Hard Disk Storage I







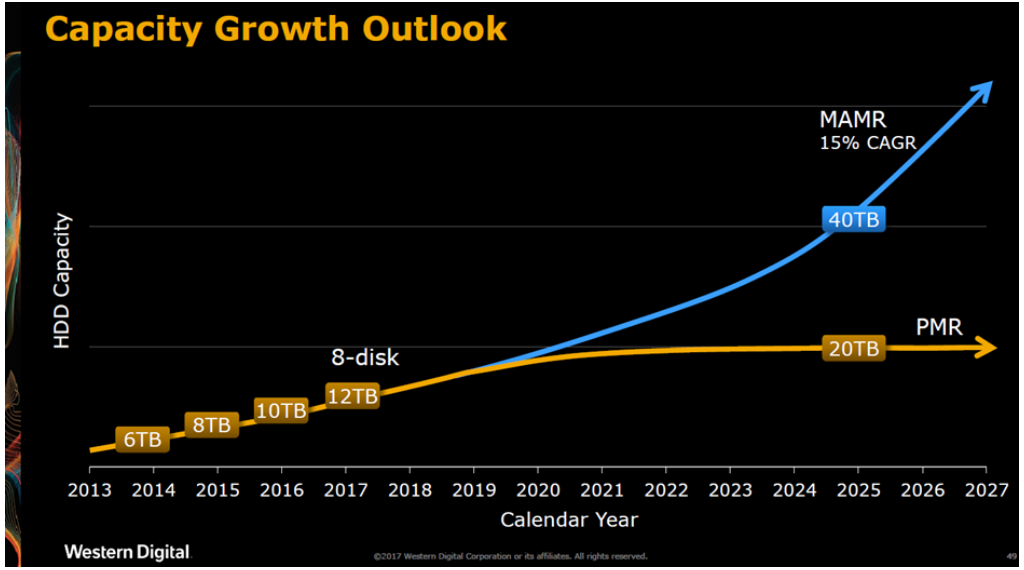
2001 2018 2023

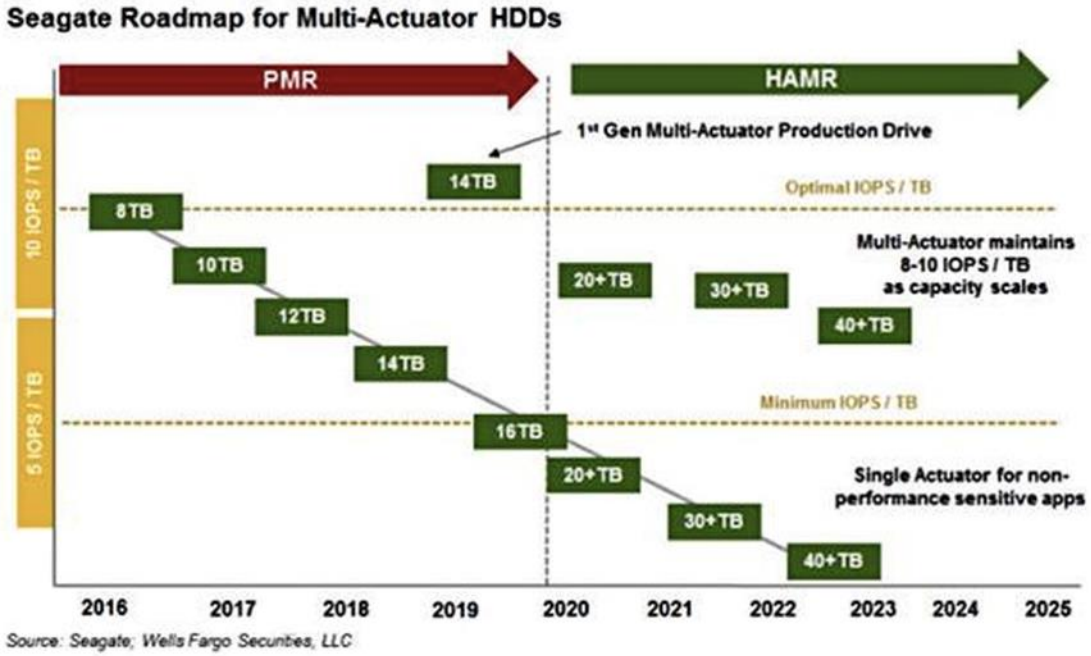
Only growth rate in Near Line disks (high capacity), HEP and Cloud Storage area

Desktop, Mobile, Enterprise replaced by SSDs

Price/space evolution flattening

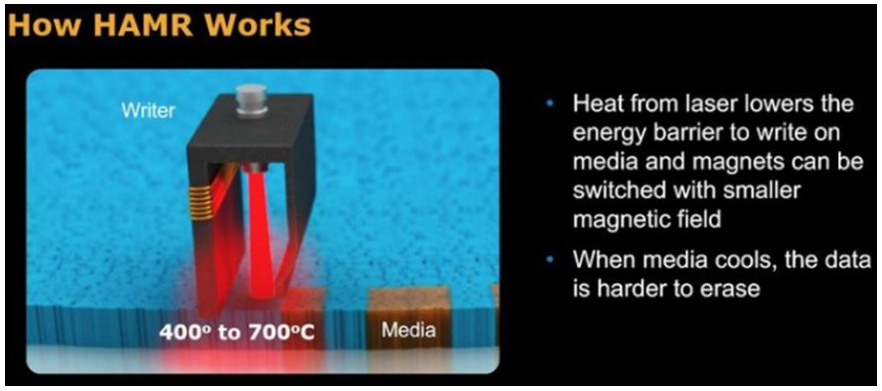
Hard Disk Storage II



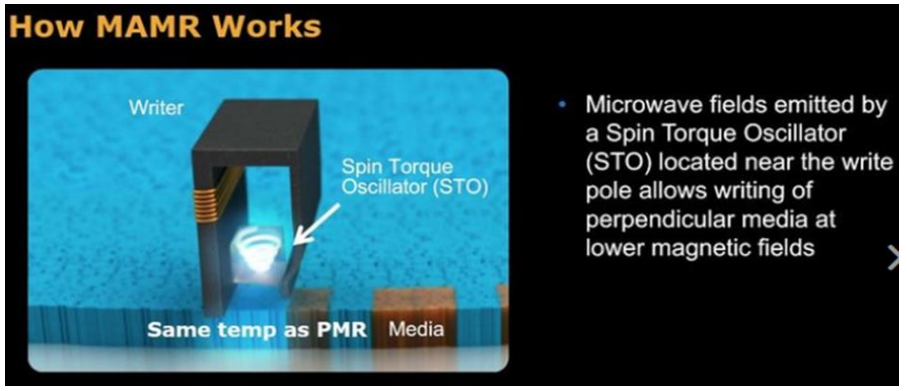


9 platter in one disk 14 TB capacity today He filled
Max with SMR is probably around 20 TB per HDD

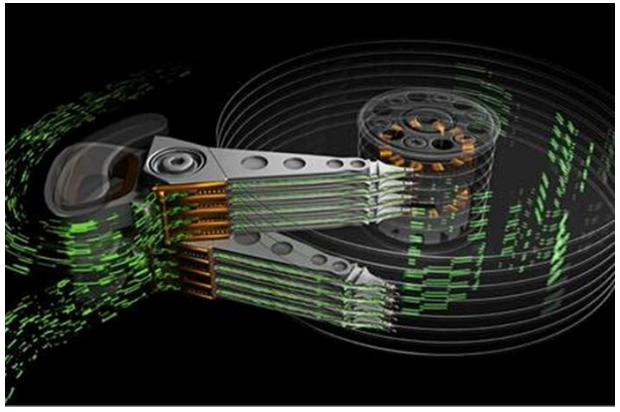
Seagate HAMR first products now in 2020



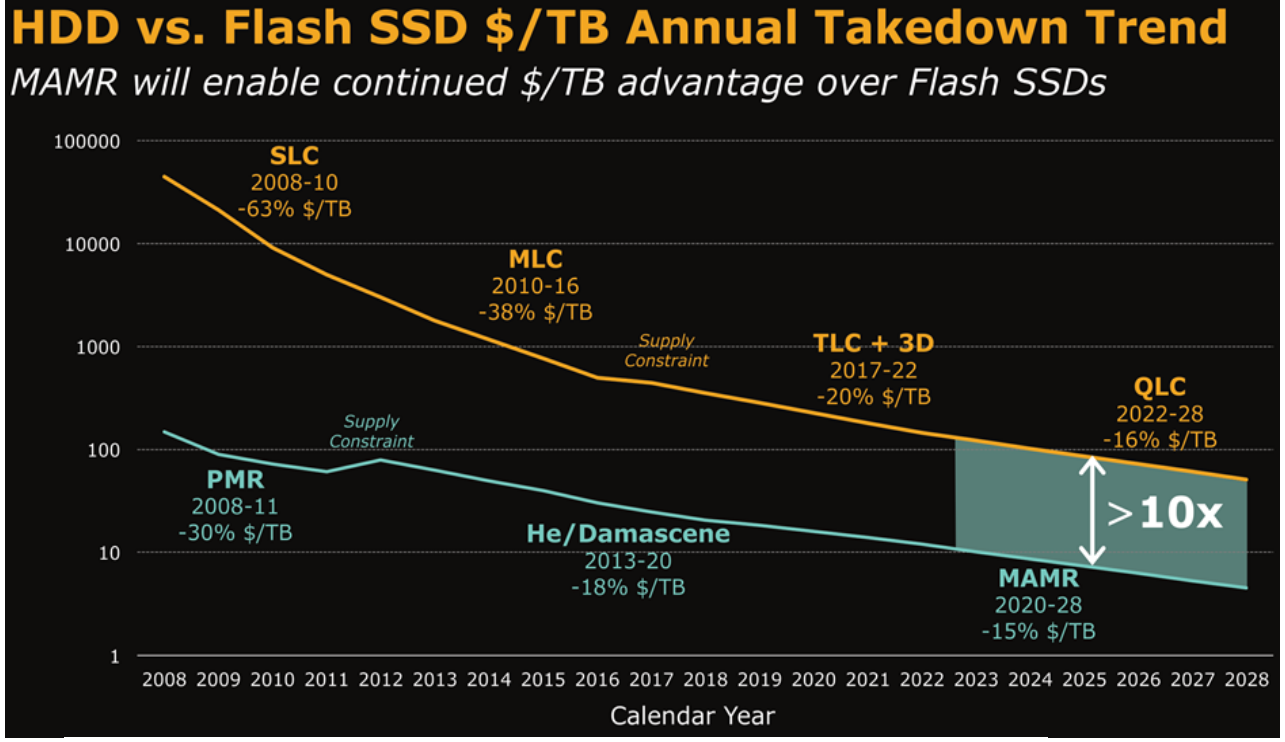
Western Digital new density approach: MAMR production in 2019

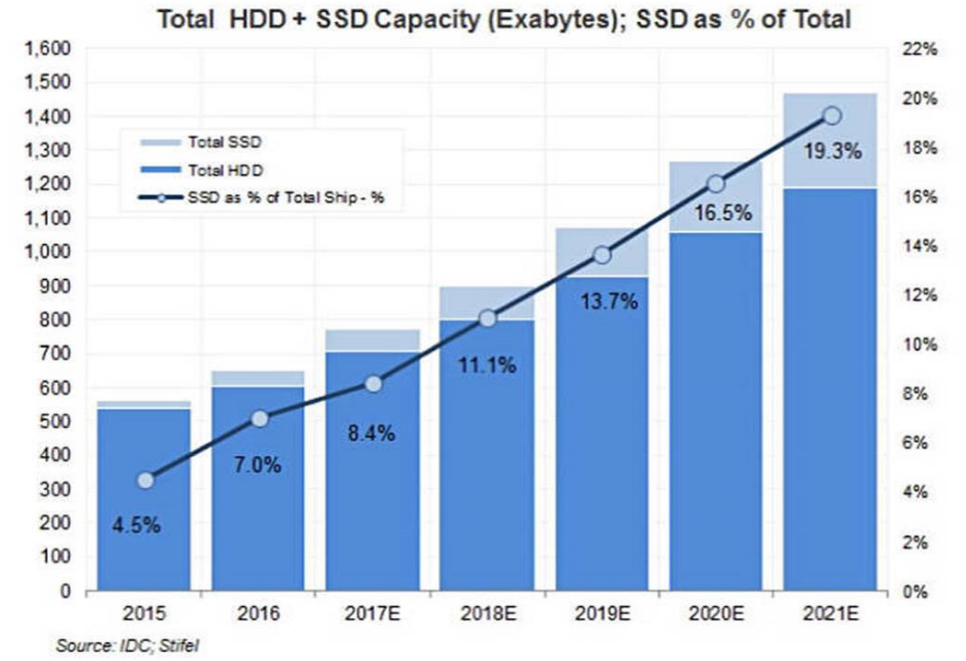


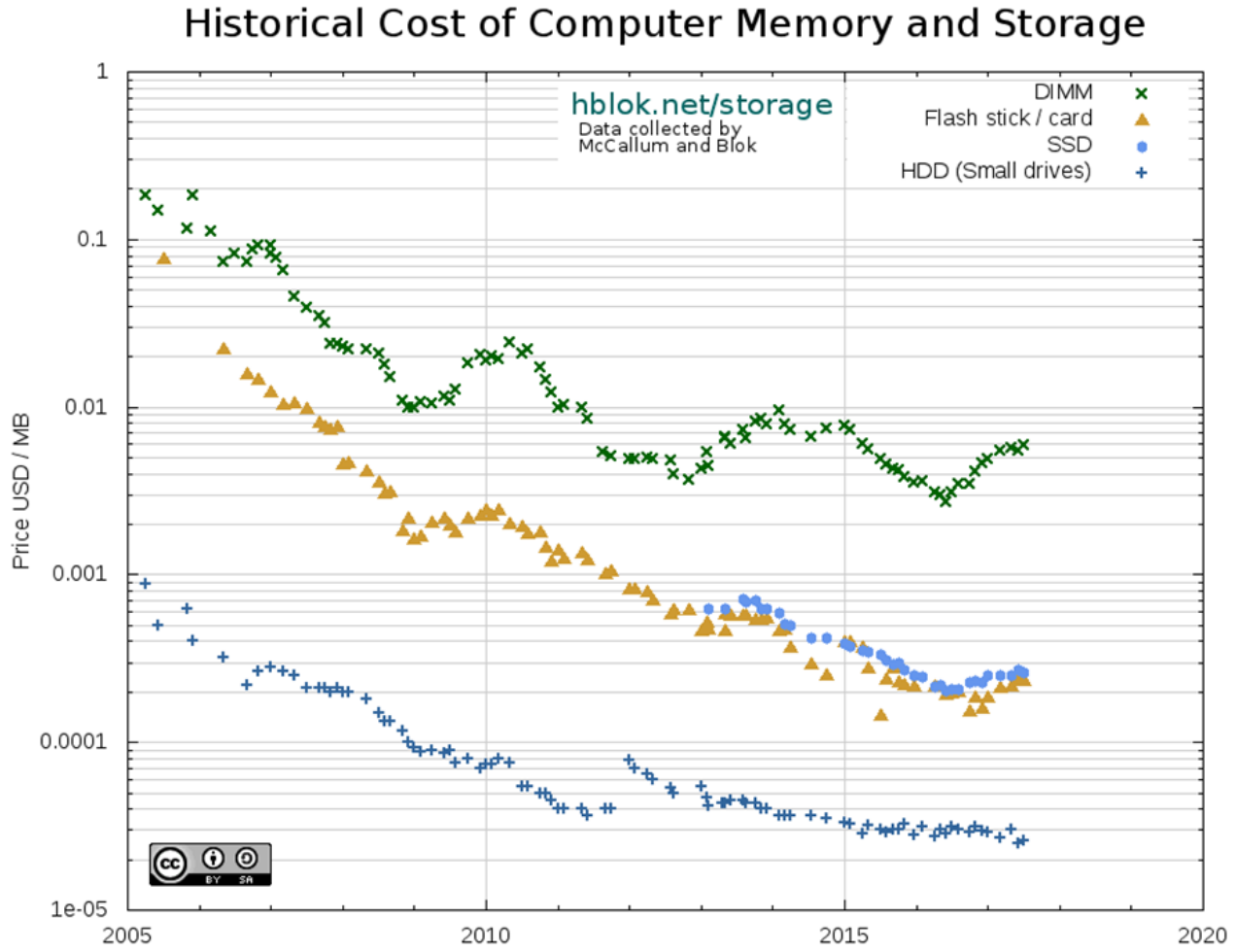
Seagate: multiple actuators per HDD to keep IOPS/TB constant



Solid State Disk Storage



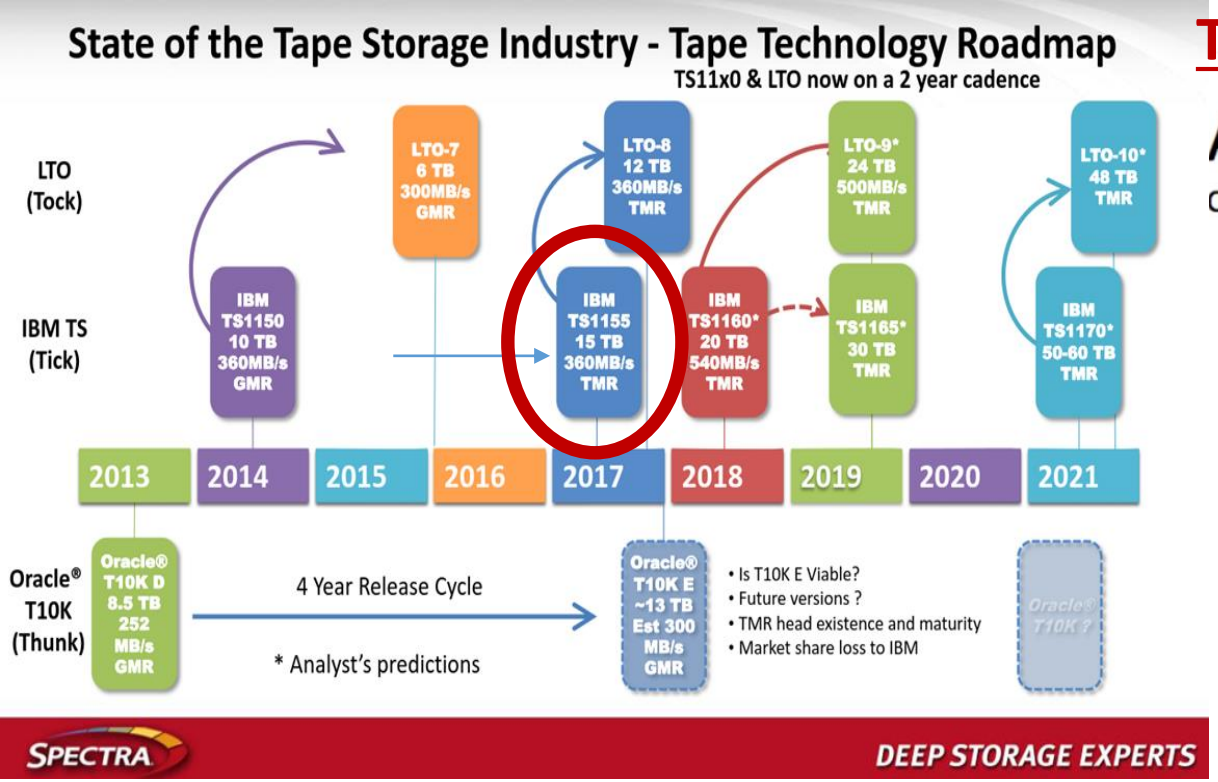


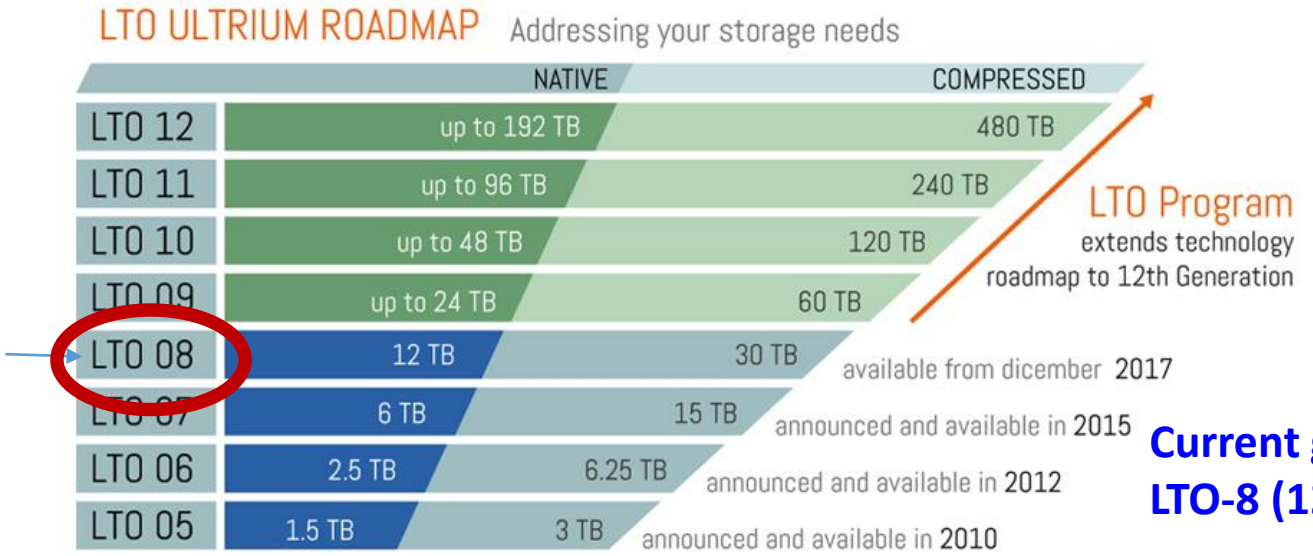


SSD versus HDD, Price difference in capacity drives will stay high For the foreseeable future

Slowdown of yearly price improvements in all areas

Tape Storage I

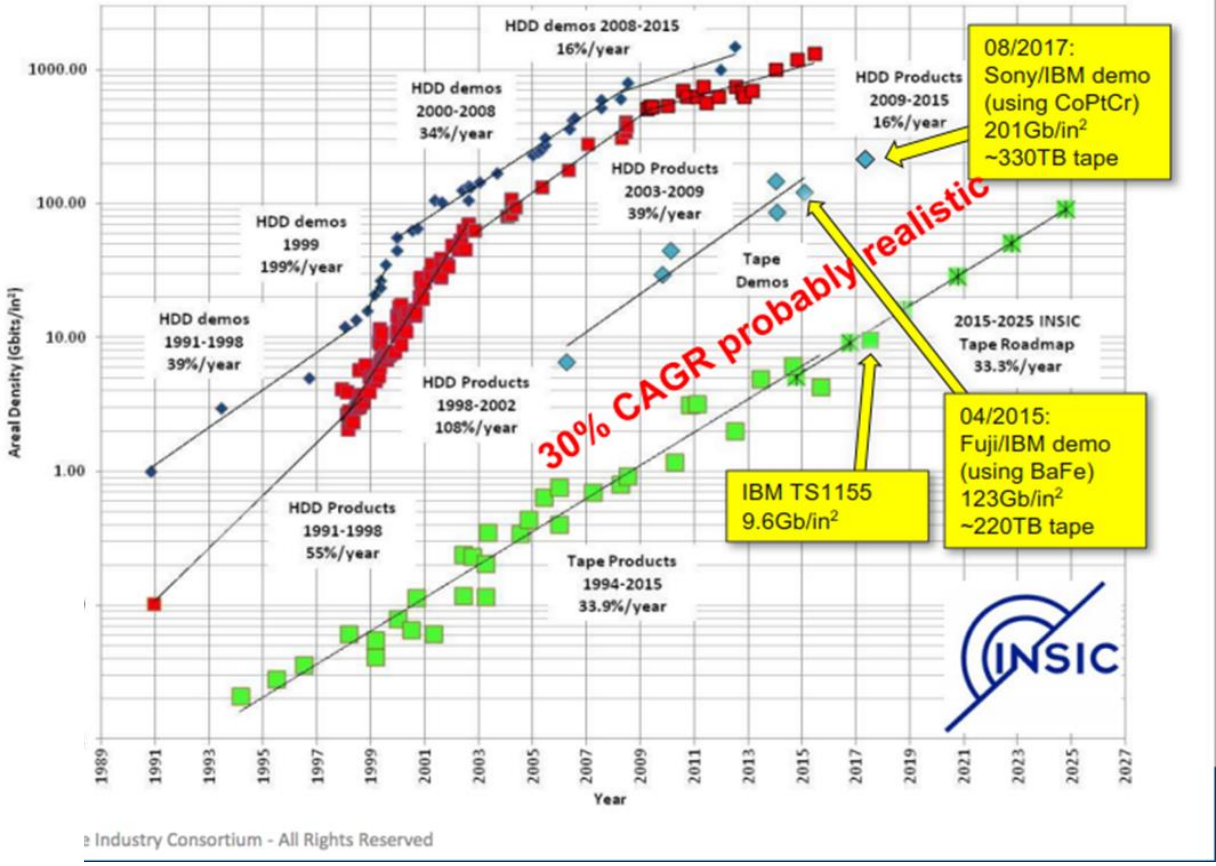




Quite some headroom for density improvements, x10 compared to HDD

Areal Density Trends

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

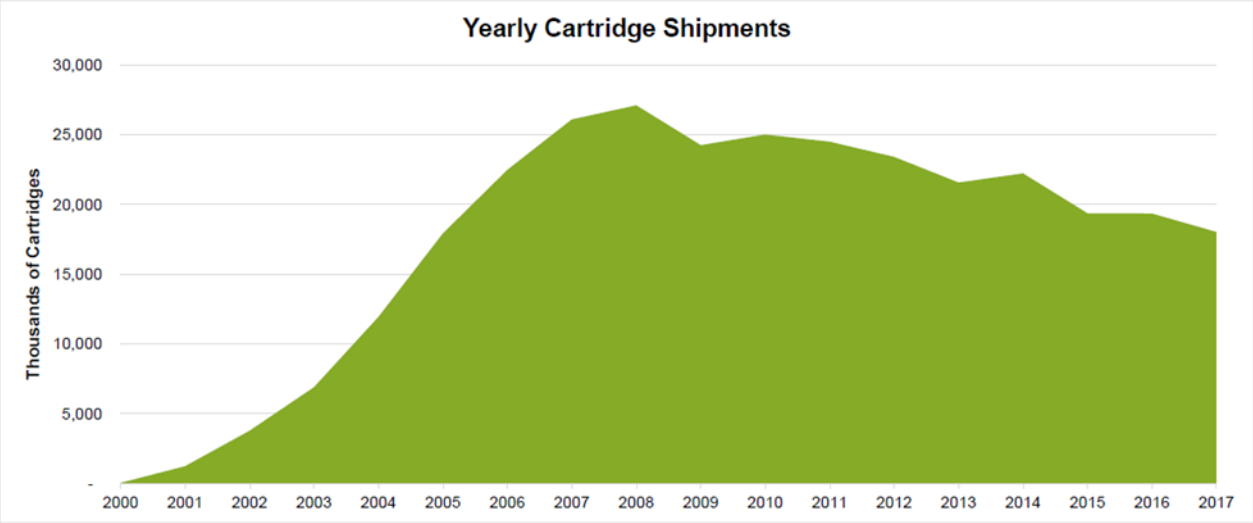


Current generation
LTO-8 (12 TB) , TS1155 (15 TB)

Technology change to
Tunnel Magnetoresistive heads
(used already in HDDs)
For IBM TS1155 and LTO-8

Tape Storage II

Unit Shipments: Calendar Year



LTO tape market domination >95%
Enterprise tapes 4%

44 EB of tape media in 2017 compared to 750 EB HDD
Linear increase in EB sold per year

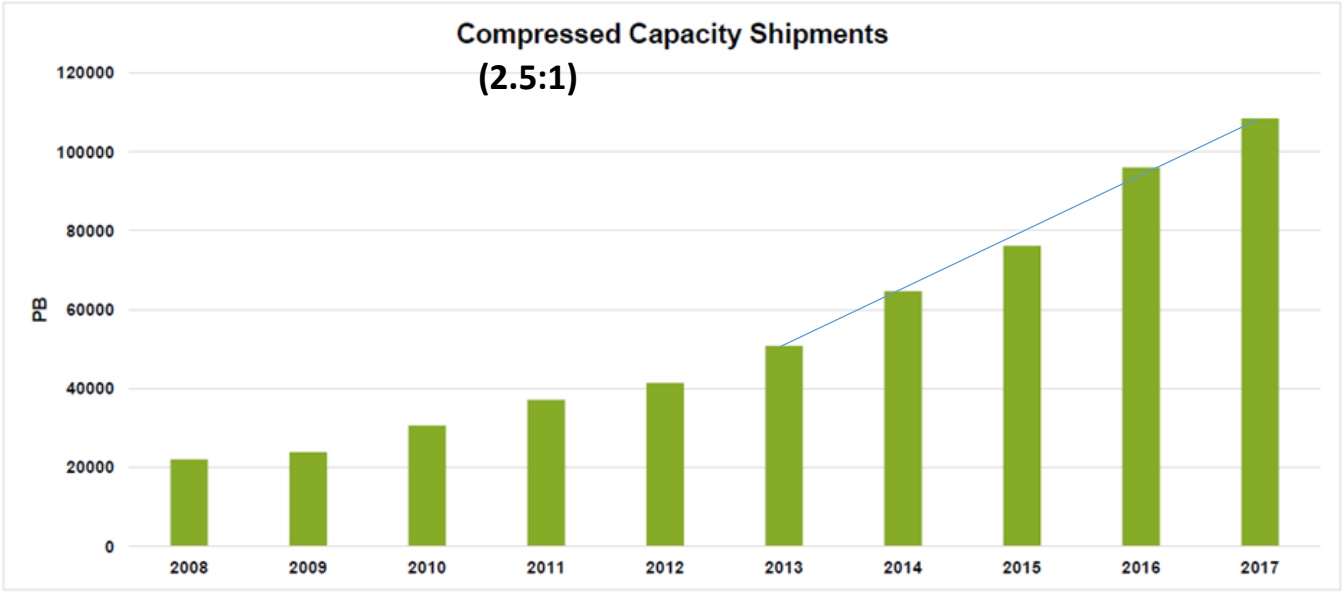
Declining media shipment since 10 years

factor 2 decrease in #drives sold over the last 4 years

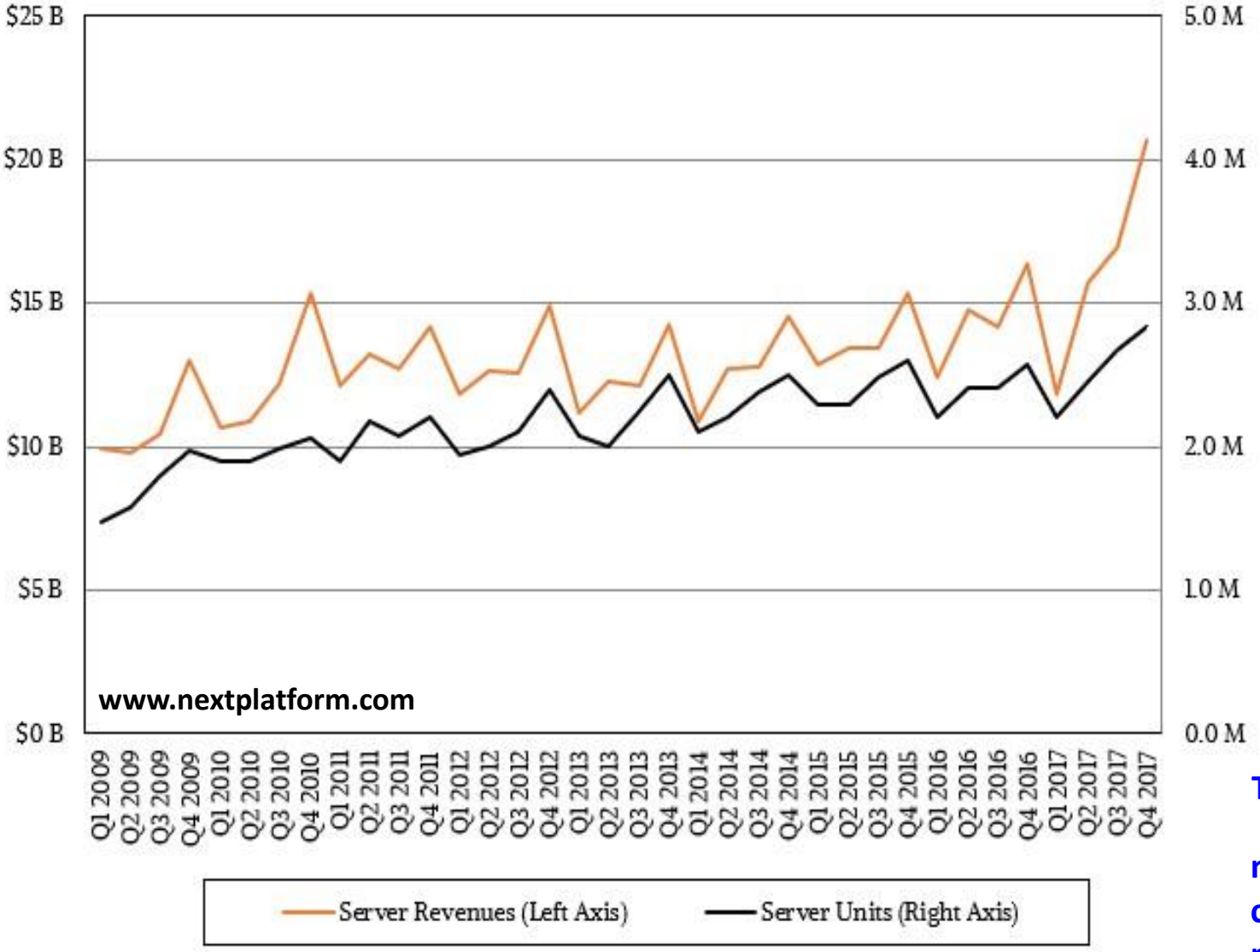
Only two suppliers of media: Fujifilm and Sony
Fujifilm only supplier in the US (patent 'war')

Only IBM left for LTO and Enterprise drives

Total Capacity Shipped: Calendar Year



Server Market I



The total server market revenues reached 20.7 B$
In Q4 2017 with 2.8 M servers shipped.
Large revenue jump → general price increases, memory price explosion, big Iron sales (z14 IBM), HPC/AI investment

The market is split into three parts based on the cost per server:

1. **< 25000$ 15.8B$ (HEP buys < 5000 $/server)**
2. **25k-250k 1.9 B$**
3. **> 250k 2.9B$**

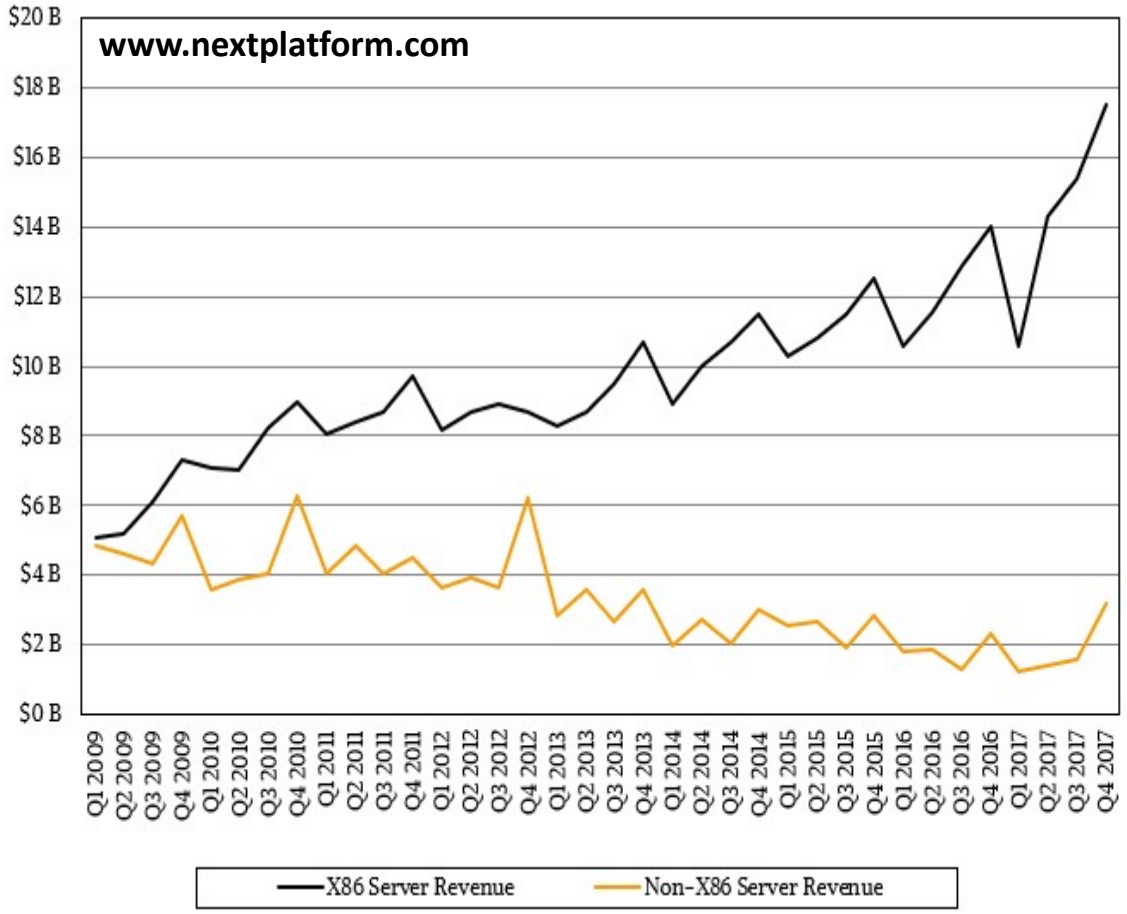
Intel takes 85% of the revenues and ships 99.3 % of the x86 servers processors

The increase in units sold is due to several factors:

new Skylake architecture, shift from DELL/HP/etc to cheaper ODM sellers and high demand for high end machines with GPUs (> 25000$ a piece)

Hyperscale data centers (>100000 server) grew in 2017 from 300 to 390
Amazon, Google, Microsoft, IBM have at least 45 centers each

Server Market II



Intel x86 dominating server market, 99.3% of server units

Possible contender:

- **IBM Power9**
  Aimed at high-end server and HPC/AI market
  (combined with Nvidia GPUs), not power efficient,
  14nm process, no plans for 10/7nm

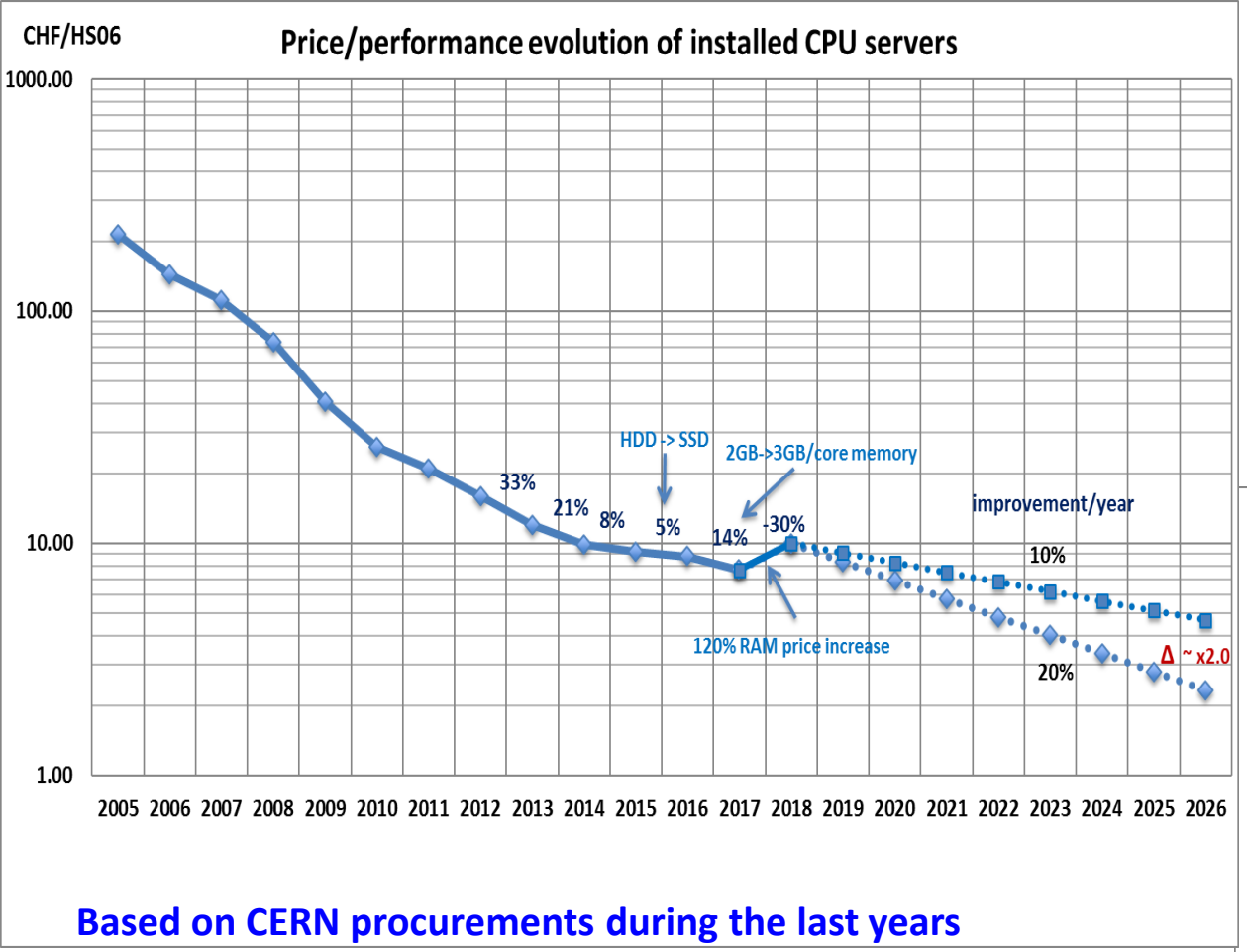
- **AMD EPYC, low market penetration so far, 2018 will be key year**

- **ShenWei 260-core processor (based on alpha, 6 Tflops SP)**
  China only, TaihuLight supercomputer, public market?

ARM

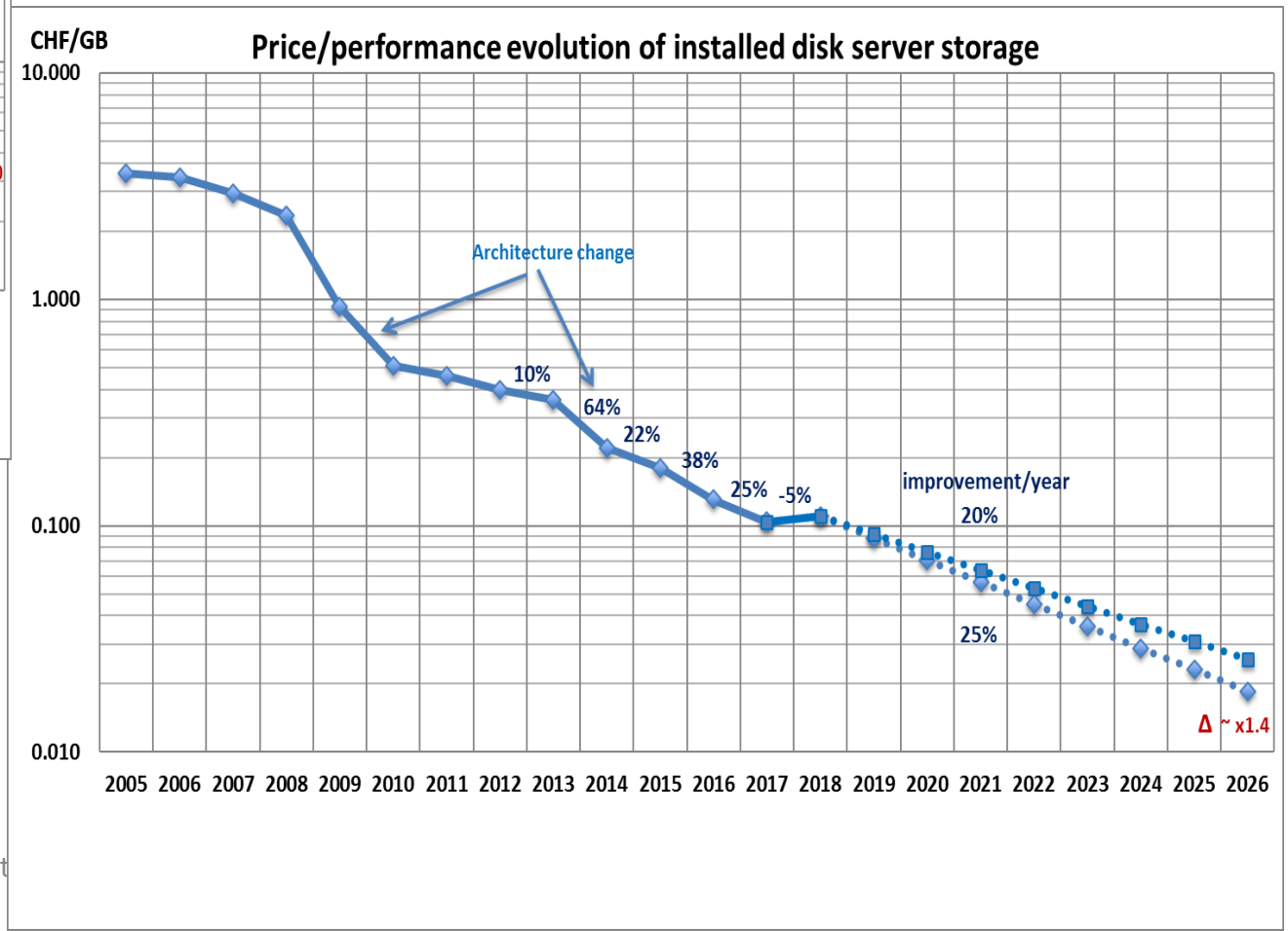
Applied Micro → Ampere : new design 32 core 3.3 GHz end 2018,
Qualcomm Centriq : 48-core, 2.6 GHz, 10nm process, first contract with cloud gaming company
(but strong rumor that Qualcomm is cancelling the project…..)
Cavium → Marvell : ThunderX2 (available in techlab, power/HS06 similar to Intel Broadwell, price/performance x2 off)

Server Cost Evolution



Based on CERN procurements during the last years

Disk server: very hard to estimate the REAL costs of HDDs
e.g. there are 70 different 6TB HDD models in the market with a price difference of a factor 2.5. At CERN we saw price difference of a factor >2 between low street prices and purchase prices. more variations between 6 TB and 8TB disks



Current assumption:

Future CPU server price/performance improvement: 15%

Future Disk server price/space improvement: 20%

Summary

- Technology progress per se is good, but obstacles ahead (CPU, RAM, NAND)
- Key computing markets in the hand of very few companies
- Price/performance advances are slowing down
- Memory prices will still increase, but major price reductions expected end 2018
- New processors and architectures are mainly focused on Machine Learning, highly profitable markets expected
- HDD still key storage for the foreseeable future, SSDs not cost effective
- Have to closely watch the tape development
- Server market still 99% Intel some competition is advancing, but not yet convincing

References

http://www.digitaltvnews.net/?p=30009
https://www.statista.com/chart/12798/global-smartphone-shipments/
http://www.transformingnetworkinfrastructure.com/topics/virtualization/articles/437082-almost-13-billion-cloud-waste-predicted-2018.htm
https://www.gartner.com/newsroom/id/3845563
https://www.semiconductors.org/news/2018/01/02/global_sales_report_2017/global_semiconductor_sales_increase_21.5_percent_year_to_year_in_november/
https://www.qstar.com/index.php/ltfs-linear-tape-file-system/
http://techinsights.com/about-techinsights/overview/blog/samsung-18-nm-dram-cell-integration-qpt-and-higher-uniformed-capacitor-high-k-dielectrics/
https://www.extremetech.com/computing/249075-foundry-futures-tsmc-samsung-globalfoundries-intel-gear-7nm-beyond
https://www.quora.com/When-is-Samsung-going-to-produce-7nm-and-is-it-comparable-to-Intel-10nm-in-terms-of-performance
https://www.neogaf.com/threads/glofo-7nm-details-revealed-the-most-likely-process-node-for-the-next-gen-consoles.1402083/
http://www.icinsights.com/news/bulletins/Semiconductor-Shipments-Forecast-To-Exceed-1-Trillion-Devices-In-2018/
https://www.forbes.com/sites/tomcoughlin/2018/02/05/hdd-growth-in-nearline-markets/2/#66cf02aa3e39
http://www.theregister.co.uk/2018/03/21/seagate_to_drop_multiactuator_hamr_in_2020/
https://www.theregister.co.uk/2017/12/19/seagate_disk_drive_multi_actuator/
http://www.tomshardware.com/news/seagate-wd-hamr-mamr-20tb,35821.html
https://www.extremetech.com/computing/266031-ibms-power9-dent-x86-server-market-oems-prep-new-systems-emphasize-gpu-compute
https://www.nextplatform.com/2018/03/01/server-market-booms-last/
https://www.backblaze.com/blog/hard-drive-cost-per-gigabyte
https://www.theregister.co.uk/2017/12/22/bit_price_decline_and_computestorage_closeness_to_send_enterprise_flash_use_sky_high/
https://hblok.net/blog/posts/2017/12/
https://www.businesswire.com/news/home/20180314005070/en/Record-Breaking-Amount-Total-Tape-Capacity-Shipments
https://www.dramexchange.com/WeeklyResearch/Post/5/4871.html
https://www.snia.org/sites/default/files/PM-Summit/2018/presentations/14_PM_Summit_18_Analysts_Session_Oros_Final_Post_UPDATED_R2.pdf
https://marketrealist.com/2018/01/micron-optimistic-falling-nand-prices
https://www.nextplatform.com/2018/02/09/just-large-can-nvidias-datacenter-business-grow/
https://www.eetimes.com/author.asp?section_id=36&doc_id=1331415