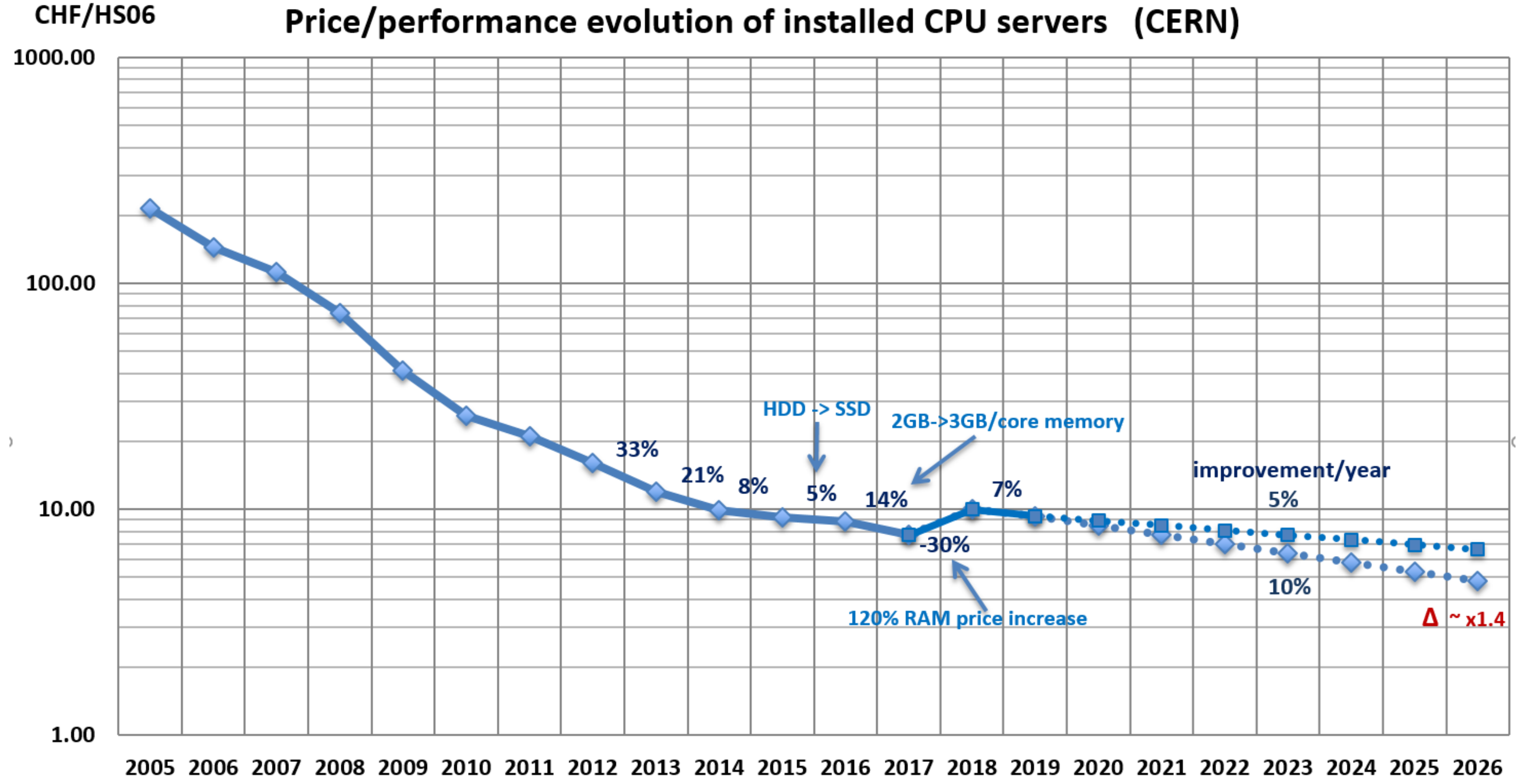
During the last years HEP used the assumption of a ~20% improvement rate for CPU and disk resources ($/HS06, $/GB) to extrapolate the future costing of computing equipment under the boundary condition of flat budgets.

CERN and several T1 sites now report deviations from the ~20% improvement rates.

It looks like the ~20% number is too optimistic and needs to be revised.



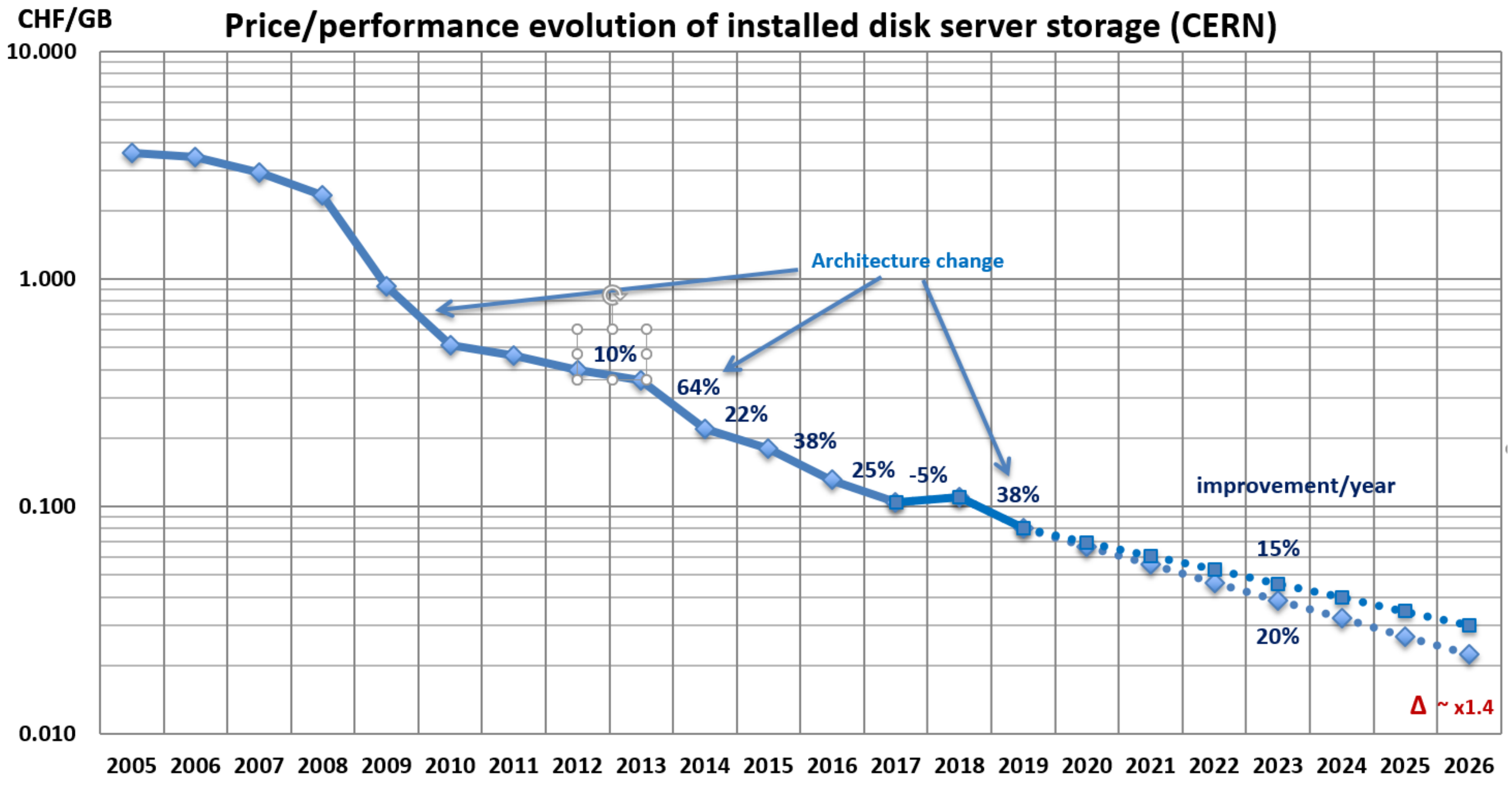
Fluctuating memory price evolution

- Slowdown in feature size reduction in processors (14nm → 10nm → 7nm)
  Especially the Intel problems with 10nm in 2018/2019

- Also increase in SSD prices last year, 20-30% decrease in 2019

- Memory and NAND prices are improving in 2019 (30-40% ?!),
  But industry (Samsung, Micron, Hynix,…..) have started to throttle production to stop the price decrease

Samsung DIMM 16GB, DDR4-2133, CL15-15-15, reg ECC



How expensive will the new 16 TB HAMR/MAMR drives be ?

When do we need more spindles then the experiment space requests will require ?
Space ←→ Performance

The total number of disks sold per year is continuously decreasing.

Strongly site depended ! Economy of scale; different server architectures; special one-off pricing; different data management systems

Cost improvement of basic components ← → infrastructure overhead reduction (varying contribution to the yearly improvement rate)
Changes of the disk server architecture at CERN have made a large impact on the cost decrease.
One-front end node (CPU server) with disk trays attached: went from 1 to 2 to 4 to 8 trays.
Overhead < 10% and much less sensitive to CPU/memory/SSD price changes.
Possible next steps: mirrored → erasure-code; replace only disks and keep trays/front-ends for >8 years.

Hardware details are also defined by dependencies on:

- **Computing models (e.g. copy data to local disks on the worker nodes)**
- **Infrastructure software (e.g. memory/disk increase due to virtualization)**
- **Data management software (e.g. memory/disk/cpu requirements on disk servers)**
- **General service infrastructure (e.g. extra computing needs for metadata)**
- **Experiment software (e.g. different memory/core requirements : 4 experiments (MC, reconstruction, analysis))**
- **Data access models (e.g. caching levels, IO versus space, redundancy levels → performance implications)**
- **……………….**

Changing improvement levels have a dual impact: general resource increase AND replacement of old resources.
The replacement cost is lifetime dependent. To minimize the cost at CERN we went to an 'open-ended' lifetime, i.e. at least 4 years, but equipment is only replaced when there is a need for space/power or the failure rate is higher than a dedicated repair-team can handle (cost of the team).
Strongly depends on the site infrastructure and boundary conditions.

The WLCG Cost and Performance Modeling working group (Markus) received numbers from several Tier 1 sites and the they show a similar picture: 20% is too optimistic
The large variance of these numbers also point to strong site dependencies.

Proposal for the future assumptions of cost improvements (just a starting point for discussions) :

- **~10 % for CPUs**
- **~15 % for disk space**
- **~20 % for tape space (stays the same as before), BUT the future of tape per se is problematic !**
  There are strong tendencies in the computing models to 'replace' disk space with tape
  needs very careful attention !

Need more input and discussions from the T1/T2 sites
Yearly adjustment of the figures !?
Weighted average !?