Ian Bird
CERN IT & WLCG
Washington DC, 23rd March 2015

Evolution of HEP Computing towards FCC

Introduction

- What are the prospects for computing in the FCC era?
  - No easy answer
- The question will really be: what can we afford?
  - What physics can be done with the computing we can afford?
  - Iterative – evolves as technology and costs evolve
- Extrapolating computing technology 20 years into the future is not obvious
  - Although historically the trends are optimistic

Topics

- What can we say/assume about the costs of computing?

- Technology trends
  - What could we expect in the next 20 years?

- What can the HEP community do to evolve and prepare?

Computing costs?

Computing costs

- For the LEP era (Tevatron, BaBar, etc) the costs of computing became commodity
  - Significant computing power available
  - Creativity allowed us to expand our needs to make use of all that was available
  - Computing "just got done" – there were more than enough resources available
  - This period may have been an anomaly
- Prior to that computing had been more expensive
  - And mostly done by large centres with large machines

Costs …

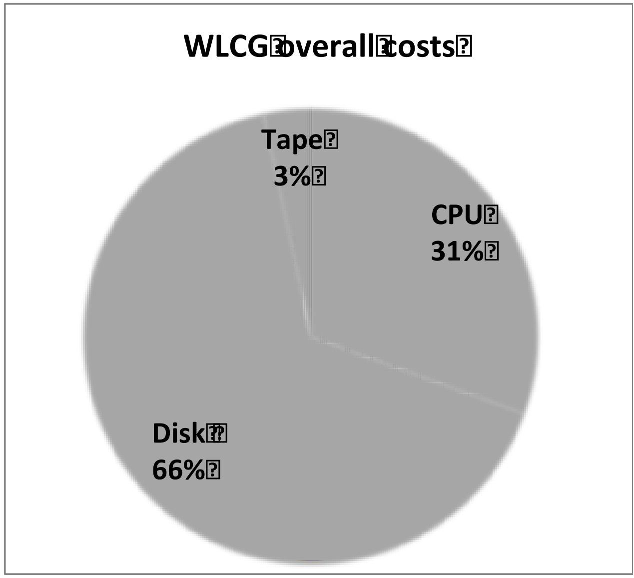
- For LHC the computing requirements led to costs estimates that seemed very high, and for some time the costs were not really discussed …

- A back-of-the-envelope calculation shows that the global yearly cost of WLCG *hardware* is approx 100M CHF/$/€
  - We do not look at the real cost – contributions are given in terms of capacity
  - 5-year cost is ~same as the construction cost of ATLAS or CMS

Cost outlook

- Will really depend on technology
  - Today this is driven by costs of commodity computing
    - Not always optimised for our use – e.g. driven by phones, tablets, etc.; ultra-low power considerations
  - Also driven by HPC requirements – large machines
    - Again, not necessarily optimal for us in the way that PC's were
  - Networking is the exception – we benefit no matter the driver
- To understand the costs of computing in FCC era we can assume that what is acceptable is
  - Computing budgets remain at the levels of today, or
  - Computing budgets (5yr) equivalent to the construction cost of a detector
  - And is a recurring cost – continual yearly replacement – equipment has 3-5 year life

Components of cost



Obviously:

- CPU and computing itself
- Storage – disk, and tape
    - Very different costs – not just hardware, but also power
- Networks

But not to forget:

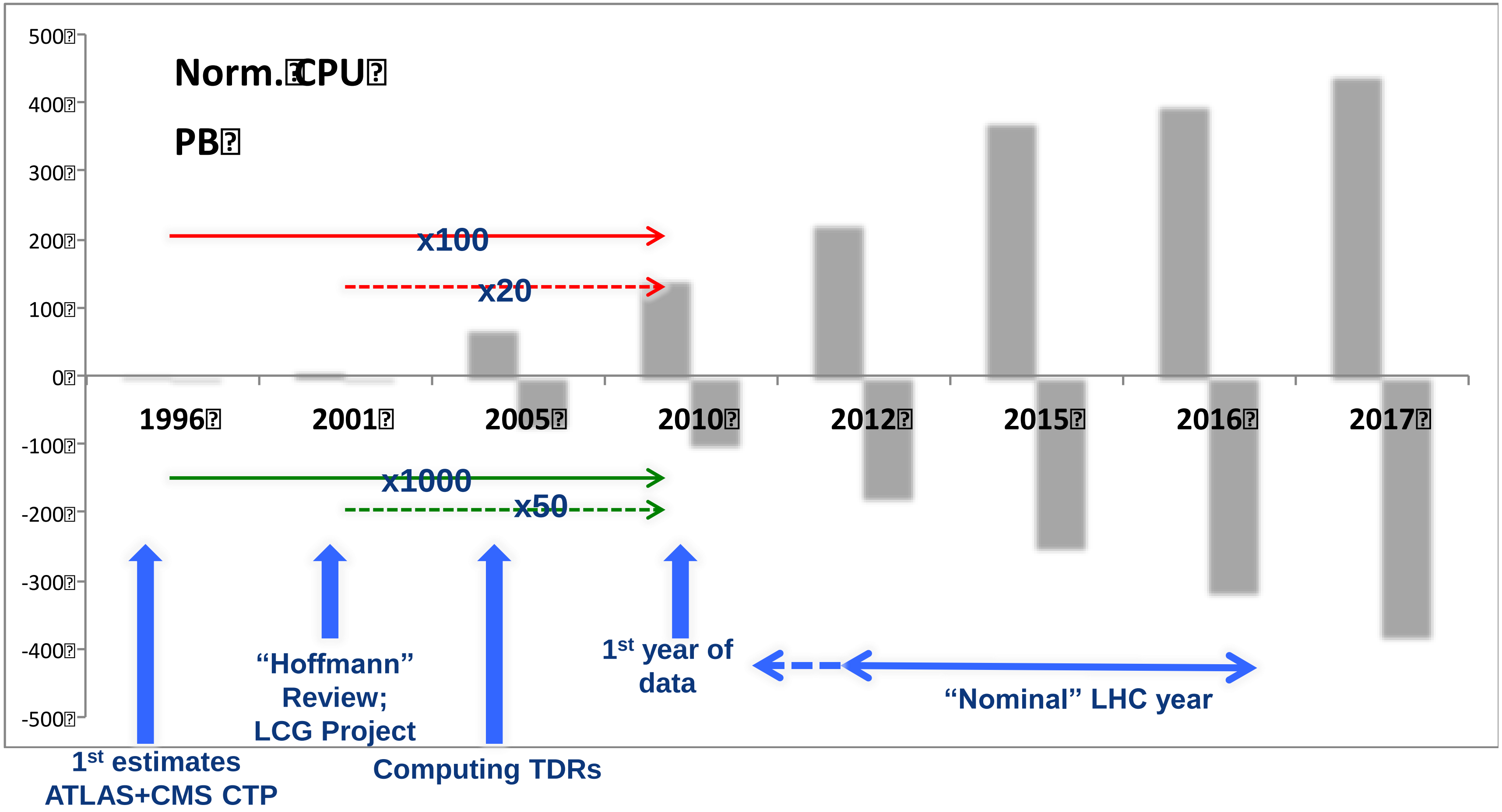
- ☐ Compute facilities
    - These are expensive and its not always obvious that building new facilities ourselves is still cost-effective
        - Associated operational cost
- ☐ Electricity
    - Becoming more expensive, and, more (Tier 2) sites are having to pay these costs now

The costs of facilities and power leads us to think that commercially provisioned compute may soon be more cost effective for HEP:

- They can benefit from huge scale of facility and operation, and locate DC's in regions of cheap power and cooling

How well do we estimate?

What was/is needed for a “nominal” LHC year



1st estimates ATLAS+CMS CTP

“Hoffmann” Review; LCG Project

Computing TDRs

1st year of data

“Nominal” LHC year

Technology outlook

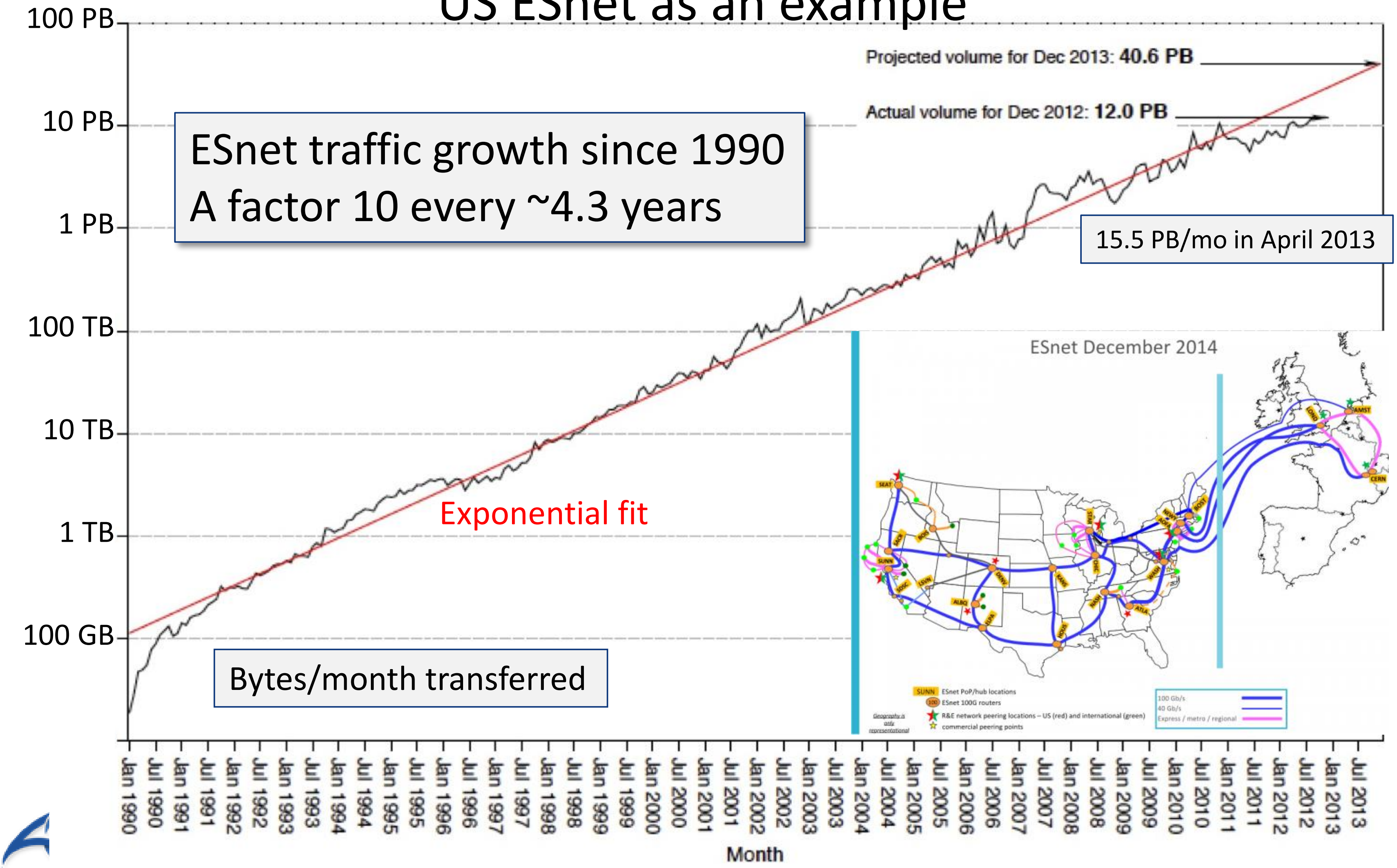
Disclaimer

- Technology companies will not give roadmaps more than 2-3 years in advance
  - We have seen many times real products very different from what we may have seen in NDA roadmaps

- Can use experience, history, and guesswork

Networking growth has been dramatic

US ESnet as an example



ESnet traffic growth since 1990

A factor 10 every ~4.3 years

15.5 PB/mo in April 2013

Exponential fit

Bytes/month transferred

Networks

- ❑ Growth has been exponential
- ❑ For WLCG this has been a key to success
  - Enables us to move away from strict hierarchy to a more peer-peer structure
  - Introducing the ability to federate data infrastructure allows us to reduce disk costs
- ❑ This is driven by consumer services
  - Video streaming, sports, etc.
  - Growth is likely to continue exponentially
  - Today 100 Gbps is ~commodity
  - 1-10 Tbps by HL-LHC
- ❑ The networking concern for HEP is connectivity to all of our collaborators
  - Again, network access to large data repositories and compute facilities is simpler than moving data to physicists

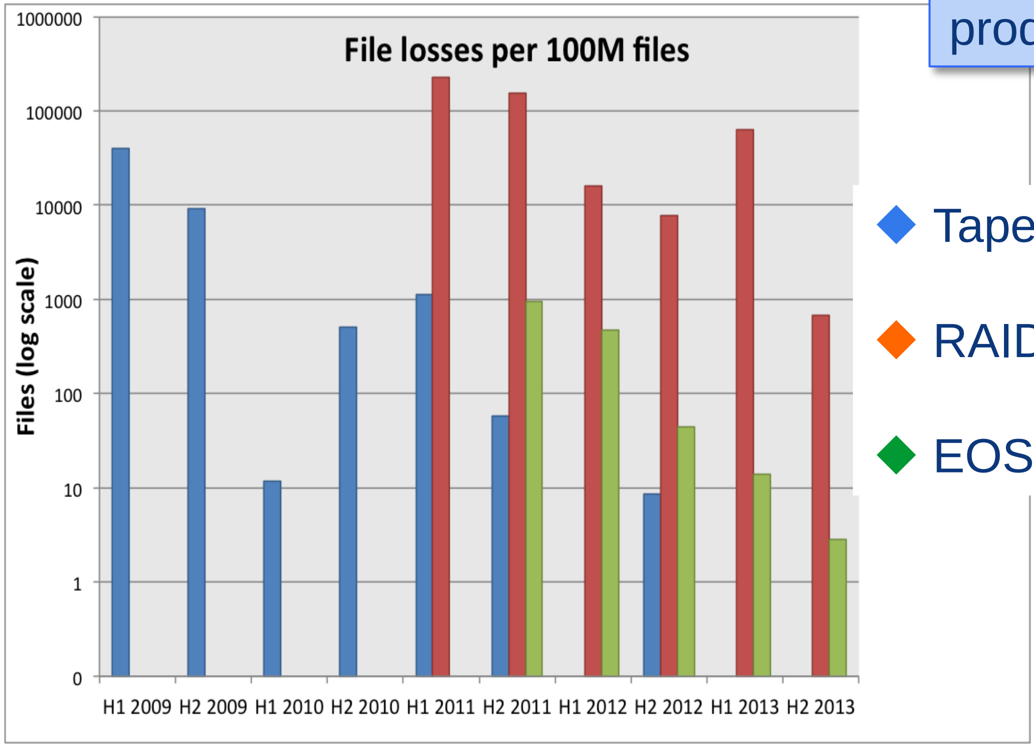
Archive storage

Tape is a long way from being dead …

Reliability and “bit” preservation

- Data reliability significantly improved over last 5 years
  - From annual bit loss rates of O(10^{-12}) (2009) to O(10^{-16}) (2012)
  - New drive generations + less strain (HSM mounts, TM “hitchback”) + verification

CERN measurements on production systems

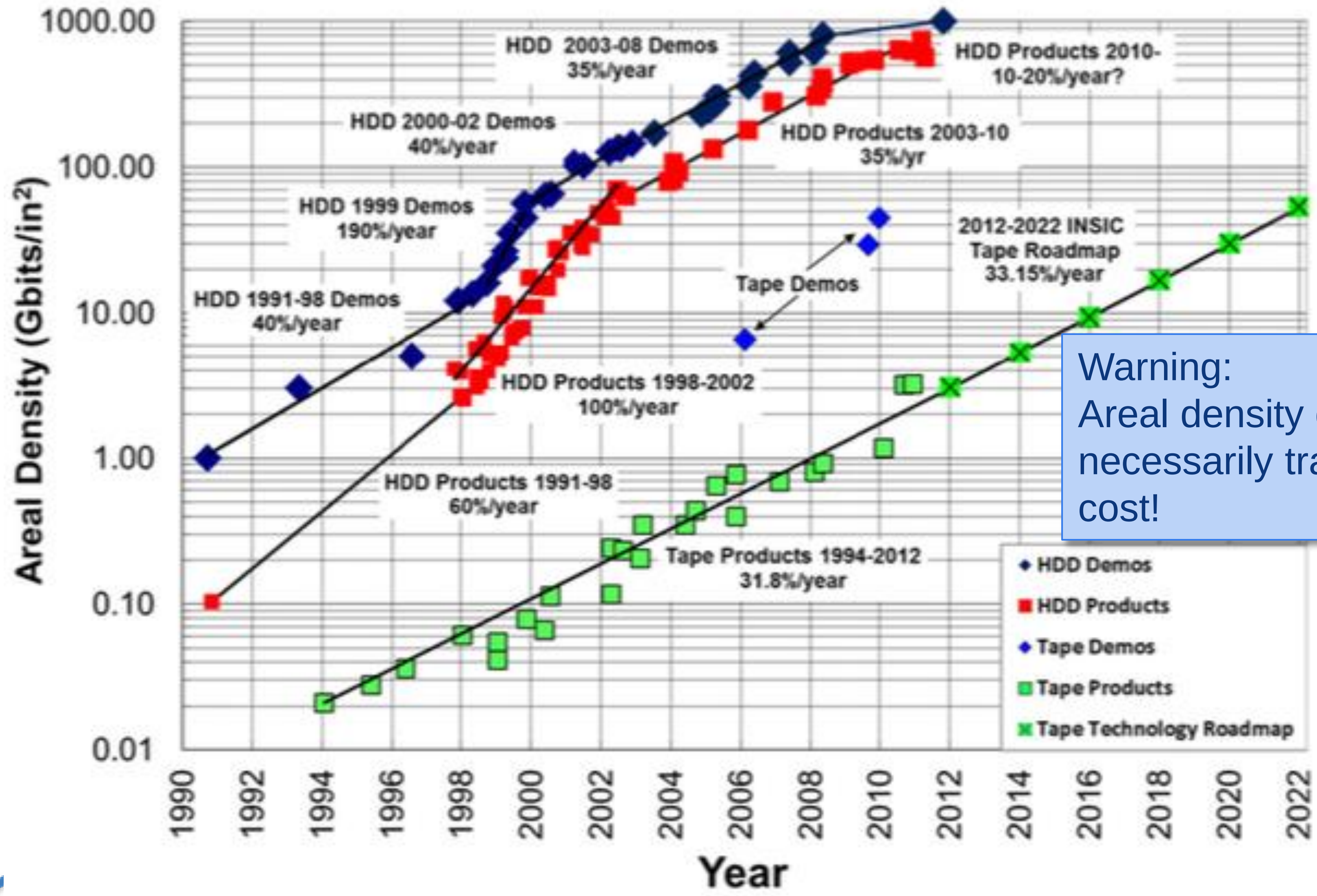


◆ Tape

◆ RAID disk

◆ EOS disk

Tape roadmap

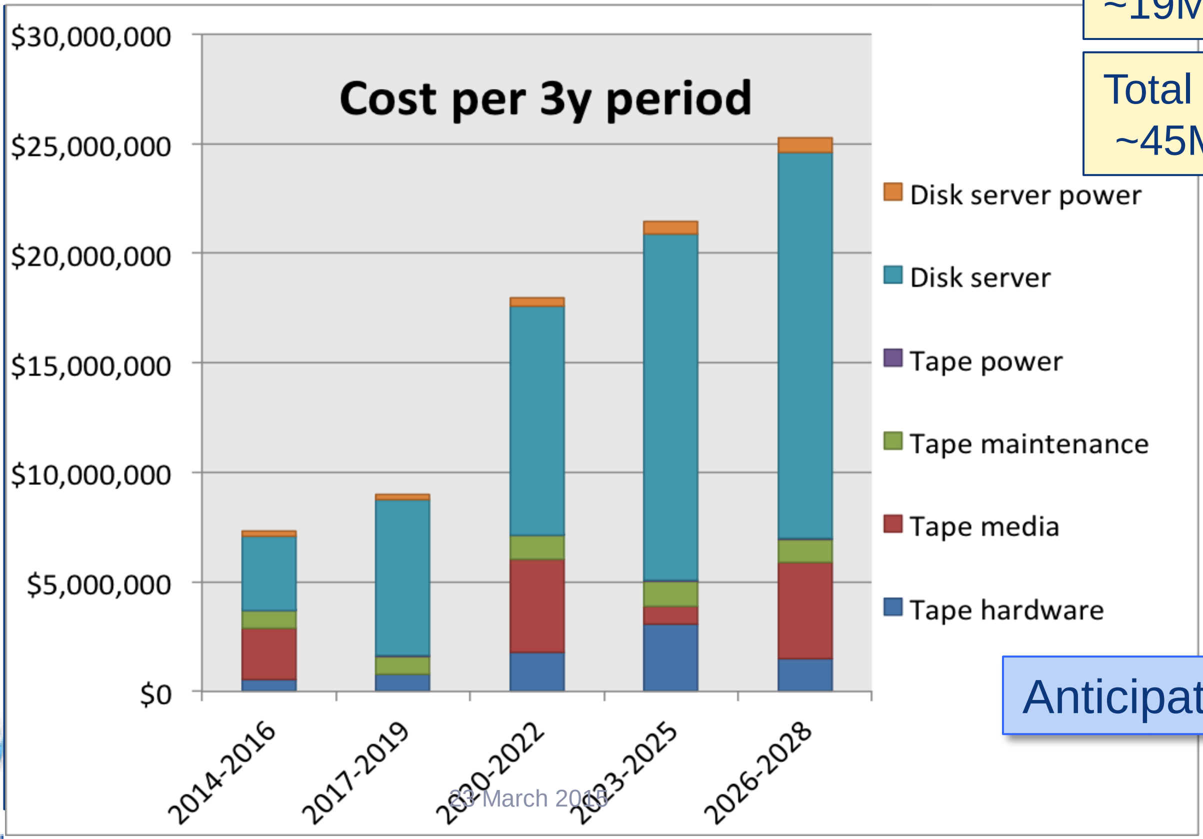


Warning:
Areal density does not necessarily translate to cost!

(Source: INSIC 2012-2022 International Magnetic Tape Storage Roadmap)

Estimates for HL-LHC

- Cost prediction - with many assumptions:
  - No paradigm change…!
  - 10% disk cache (with 20% redundancy overhead)
  - 3y cycle for disks and tape drives, and 6 years for reusable enterprise tape media (repack every 3y)
  - Tape libraries upgraded/replaced around 2020-2025



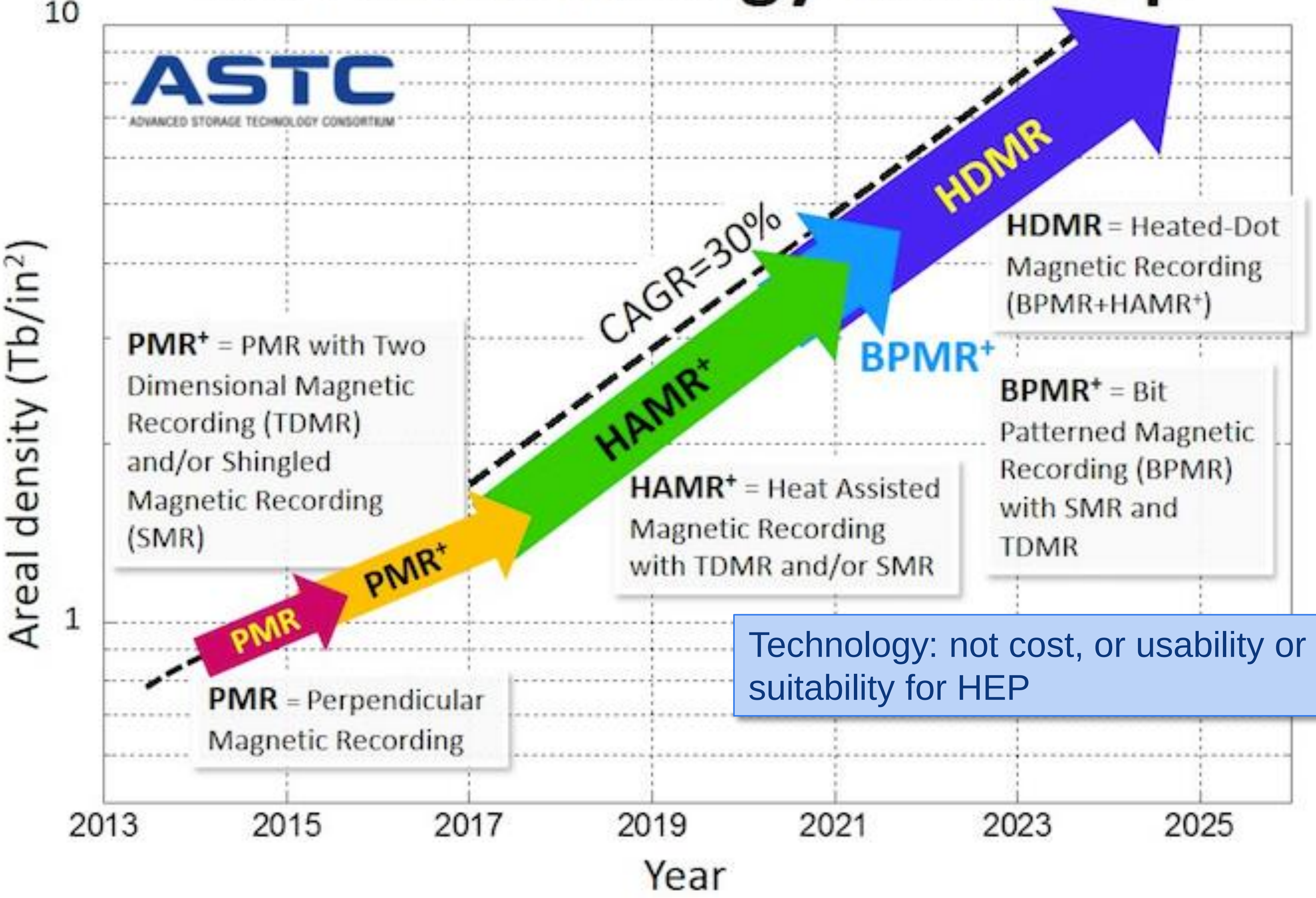
Total 2020-2028 tape:
~19M CHF (2.1M CHF / year)

Total 2020-2028 10% disk:
~45M CHF (5M CHF / year)

Anticipate continued evolution

Disk storage

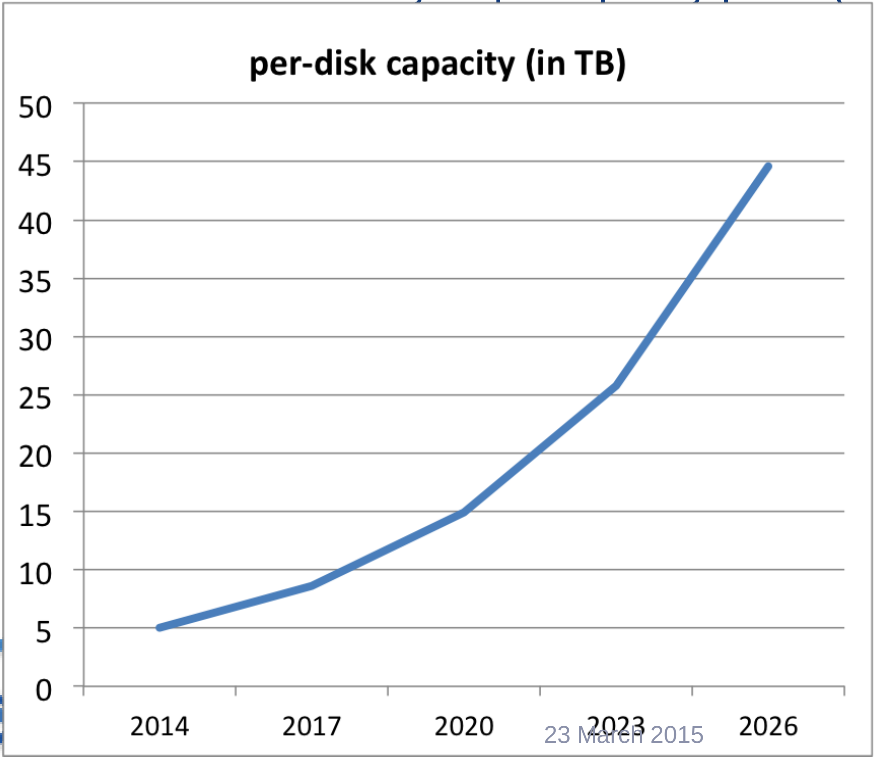
ASTC Technology Roadmap

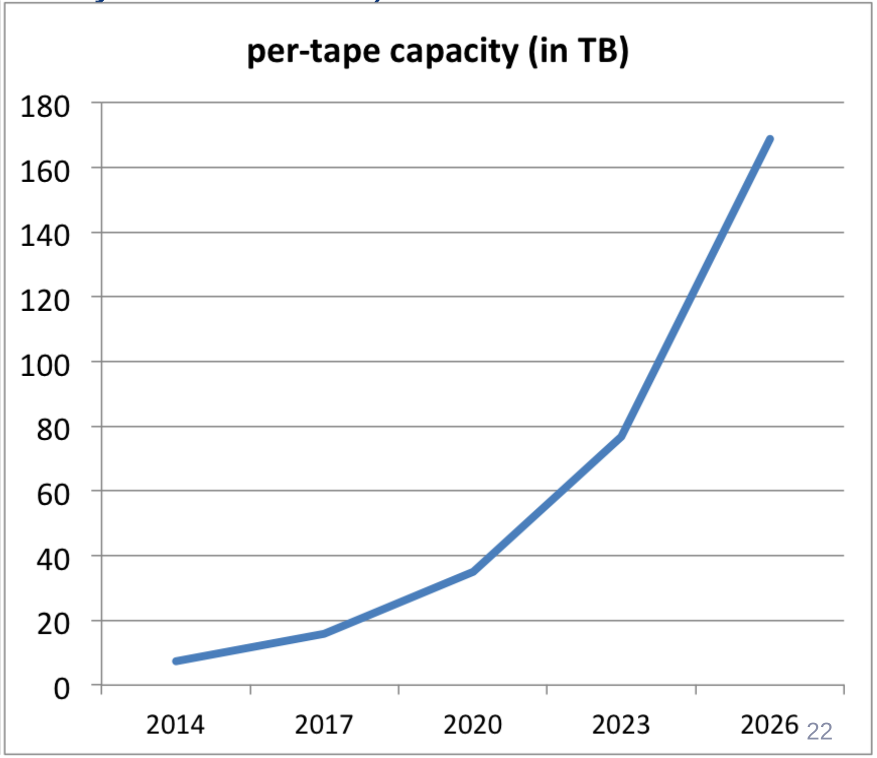


Technology: not cost, or usability or suitability for HEP

Longer term?

- Disk growth
  - New techniques anticipated – continue to grow capacity
  - May not be so easy to use (e.g. shingled disks)
- Technology/market forecast (…risky for 15 years!)
  - INSIC Roadmap:
    - +30% / yr tape capacity per $ (+20%/yr I/O increase)

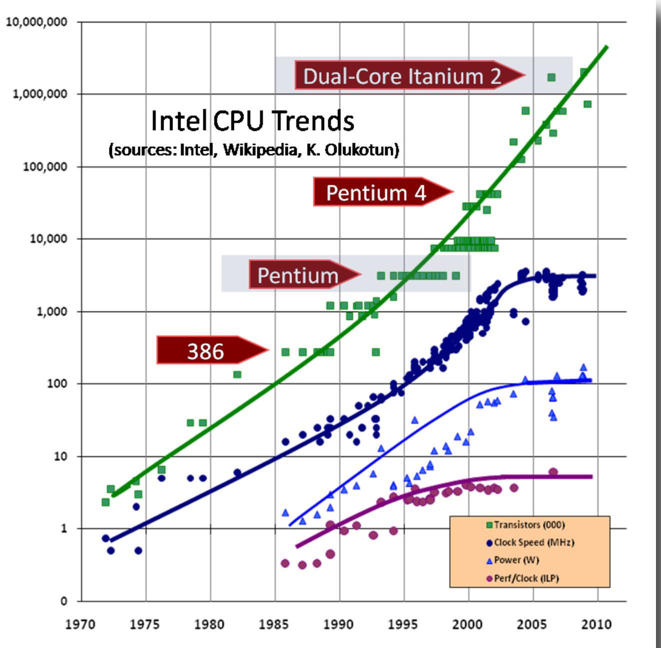




Compute growth

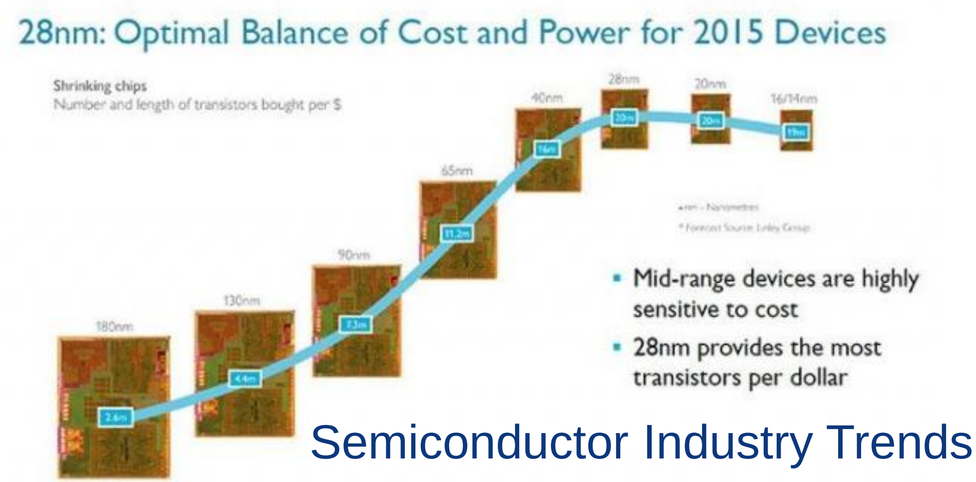
Roadmaps for computing

- ❑ "Moore's law is dead" …
  - Not quite yet …
  - Depends who, and what question, you ask
- ❑ Close to physical limits for feature size
- ❑ But:
  - Can still pursue bringing down the costs at a given feature size
  - Reducing the power requirements
  - Etc.



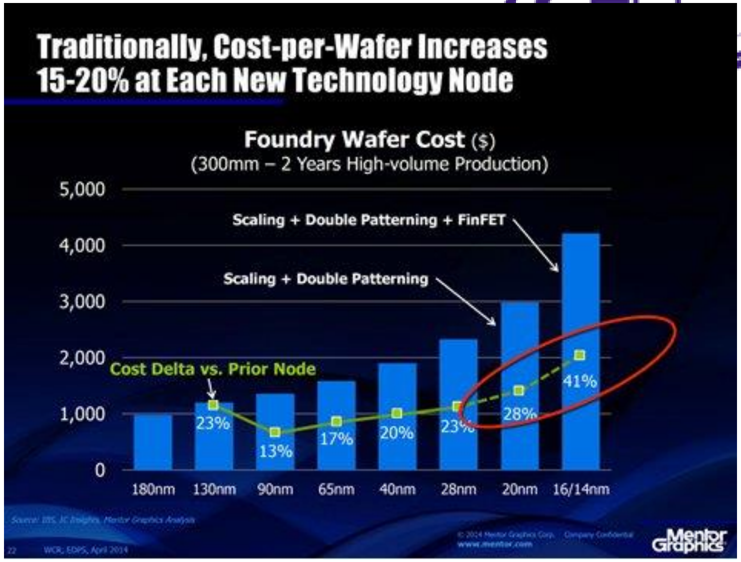
© 2009 Herb Sutter

Semiconductor Industry Trends



28nm: Optimal Balance of Cost and Power for 2015 Devices

- Mid-range devices are highly sensitive to cost
- 28nm provides the most transistors per dollar



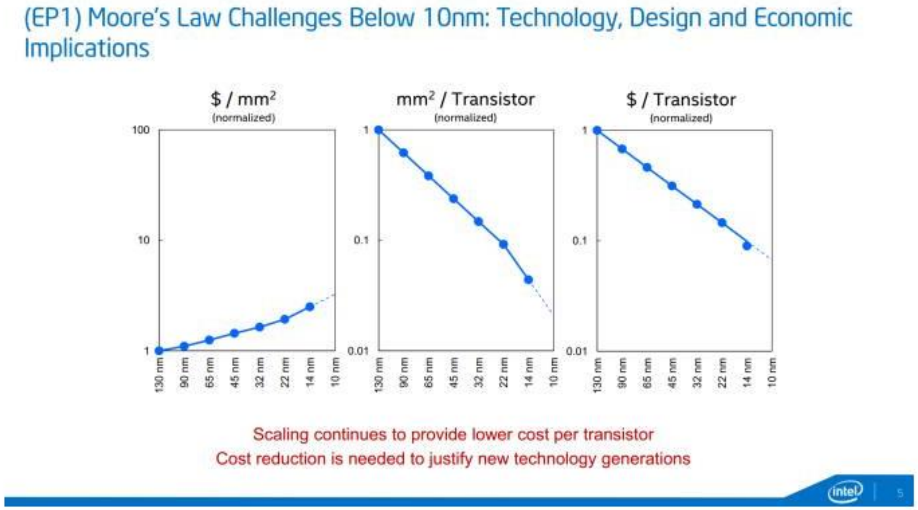
Traditionally, Cost-per-Wafer Increases 15-20% at Each New Technology Node

Source: IBS, IC Insights, Mentor Graphics Analysis

22 WCR, EDPS, April 2014

© 2014 Mentor Graphics Corp. Company Confidential
www.mentor.com

INTEL claims to overcome this up to the 10nm node scale



(EP1) Moore's Law Challenges Below 10nm: Technology, Design and Economic Implications

Scaling continues to provide lower cost per transistor
Cost reduction is needed to justify new technology generations

Hints that we have reached the end of Moore's law simple scaling

Fabrication units have now price-tags of > 10B$ (latest Samsung fab = 14.7 B$)
<5 companies worldwide are capable of financing this

Trends in HEP computing

- Distributed computing is here to stay
    - Actually we had it 30 years ago, and seriously 15-20 years ago
- Ideal general purpose computing (x86 + Linux may be close to the end
    - May be more effective to specialise
        - GPU and other specialised farms
        - HPC machines
        - Commodity processors (“x86”, ARM, etc)
    - Used for different purposes – lose flexibility but may gain significantly in cost

Trends – Data centres

- Moving data around the world to 100's of sites is unnecessarily expensive
  - Much better to have large scale DC's (still distributed but O(10) not O(100)) – connected via v high bandwidth networks
  - Bulk processing capability should be located close or adjacent to these
  - Data access via the network – but in a truly "cloud-like" way – don't move data out except the small data end-products

Data centres

- Our Data Centres may become exactly that – dedicated to data
- Compute resources are quite likely to be commercially available much cheaper
  - Don’t know how they will be presented (hosted, cloud, xxx, …)
  - Already see today commercial compute costs are comparable to our costs
- Not likely, or desirable, that we will give up ownership of our data
  - Will still need our large data facilities and support

“Tier 2”-like resources

- Today these are crucial
  - >50% of CPU provisioned here
  - More importantly today these give access to the experiment data
    - And get us synergistic use of spare resources
- And, engagement of skilled people
- Don’t want to lose this
  - But there are many workloads that are still suited to this type of resource

Opportunistic resources

- Today this has become more important
  - Opportunistic use of:
    - HPC’s
    - Large cloud providers
    - Other offers for “off-peak” or short periods
    - Etc.
    - All at very low or no cost (for hardware)
  - But scale and cost are unpredictable
- Also growing in importance:
  - Volunteer computing (citizen science)
    - BOINC-like (LHC@home, ATLAS/CMS/LHCb@home, etc)
    - Now can be used for many workloads – as well as the outreach opportunities

Trends – Architectures

- Will need to be able to make use of specialised CPU architectures
  - Different problems (event generation, simulation, reconstruction, analysis) may all be better suited to different architecture types
  - We need flexibility in software and in our ability to use existing and new architectures

Trends – software

- Recognizing the need to re-engineer HEP software
  - New architectures, parallelism everywhere, vectorisation, data structures, etc.
- Set up HEP Software Foundation (HSF)
  - Community wide – buy in from major labs, experiments, projects
  - Goals:
    - Address rapidly growing needs for simulation, reconstruction and analysis of current and future HEP experiments,
    - Promote the maintenance and development of common software projects and components for use in current and future HEP experiments,
    - Enable the emergence of new projects that aim to adapt to new technologies, improve the performance, provide innovative capabilities or reduce the maintenance effort,
    - Enable potential new collaborators to become involved,
    - Identify priorities and roadmaps,
    - Promote collaboration with other scientific and software domains.

What should HEP do?

Evolution?

- ☐ Today we have WLCG –
  - Scope is LHC
- ☐ - and international e-infrastructures
  - Which support other HEP and other sciences

- ☐ We see requests from other HEP experiments (Belle-II, ILC, AMS, etc) to be able to make use of the WLCG structures
  - Not really the compute/storage resources
    - Most experiments have their own funded allocations
  - But want to benefit from the structure
    - Support, networks, policies, operations, security, etc
  - And of course many of the sites are common
- ☐ And its not just HEP now – sites will be common with LSST, CTA, SKA, etc.,etc.
  - Really need the infrastructures to be as common as possible
  - Otherwise the support load and cost is unsupportable

Evolution of facilities

- Today we have LHC (WLCG as the computing facility)
- Recognise that between now and FCC, we have potentially many international facilities/collaborations involving global HEP community
  - HL-LHC, Belle-II, Neutrino facilities, ILC/linear collider
  - Etc.
- Thus, we should build on our working infrastructure to evolve towards FCC, serving the needs of these facilities and learning from them

Evolution of structure

- Distinguish between infrastructure and high level tools
- We need to continue to build and evolve the basic global HEP (+others) computing infrastructure
  - Networks, AAA, security, policies, basic compute and data infrastructure and services, operational support, training, etc.
  - This part MUST be common across HEP and co-existing science
  - This part must also be continually evolving and adapting with technology advances
- Need a common repository/library of proven and used middleware and tools
  - A way to help re-use of high and low level tools that help an experiment build a computing system to make use of the infrastructure
  - The proto-HSF today could be a seed of this
- We must try and make this a real common effort and remove a lot of today's duplication of solutions
  - While retaining the ability and agility to innovate
  - The cost of continuing to support unnecessary duplication is too high

Skills

- Difficult to find and retain people with appropriate skills
  - Lack of a career path outside of Labs is a major concern
  - This seems to become a more and more significant problem
- Effort on Computing and Software needs to be treated by the community at the same level as detector building and other key tasks

Conclusions

- 20-year technology extrapolations are unrealistic
  - And miss game-changing events such as mainframe→PC transition
- Computing technology (networks, compute, storage) is being driven by consumer markets
  - Good: much more influential than science
  - Bad: directions may not be easy to adopt
- We must be flexible and adaptable to technology and commercial trends
- Make use of our existing working system to operate and evolve towards FCC, meanwhile serving the intermediate needs of the HEP (and broader science) community