Le sfide del calcolo scientifico

Stefano Piano – INFN sez. Trieste

Introduction

- The scientific computing is a broad field, and I cannot cover all its challenges with my presentation and my knowledge

- I will focus on LHC exp's software and computing with a time horizon of future challenges in the next 15 years
  - The LHC computing resources are provided within the WLCG infrastructure
  - Since the beginning of LHC computing (~2001), INFN has been deeply involved in the development of the Grid middleware and in the management of the Grid infrastructure for the LHC exp's

- Special thanks to those who provided material and suggestions for these slides:
  - Simone Campana (WLCG project leader)
  - Concezio Bozzi (LHCb CC), Alessandro Di Girolamo and Zach Marshall (ATLAS CC's), James Letts and Danilo Piparo (CMS CC's)
  - Roberto Baccomi, Benigno Gobbo and Alessandro Tirel (INFN-TS)

Computing evolution @ INFN-TS

2002 ACID (COMPASS) + ALITS (ALICE)

CPU	120 HS06	44 server (dual PIII) 88 CPUs (cores) 10.240 TB RAM
DISK	2.5 TBn (ACID) + 1.6 TBn (ALITS)	NFS + ODBMS Objectivity
TAPE	3 TBn (ACID)	Ultrium STK L40 (2xLTO)
NET	Storage 1000 Mbps Ethernet	Ethernet 100 Mbps + 1000 Mbps

2022 FARM-TS

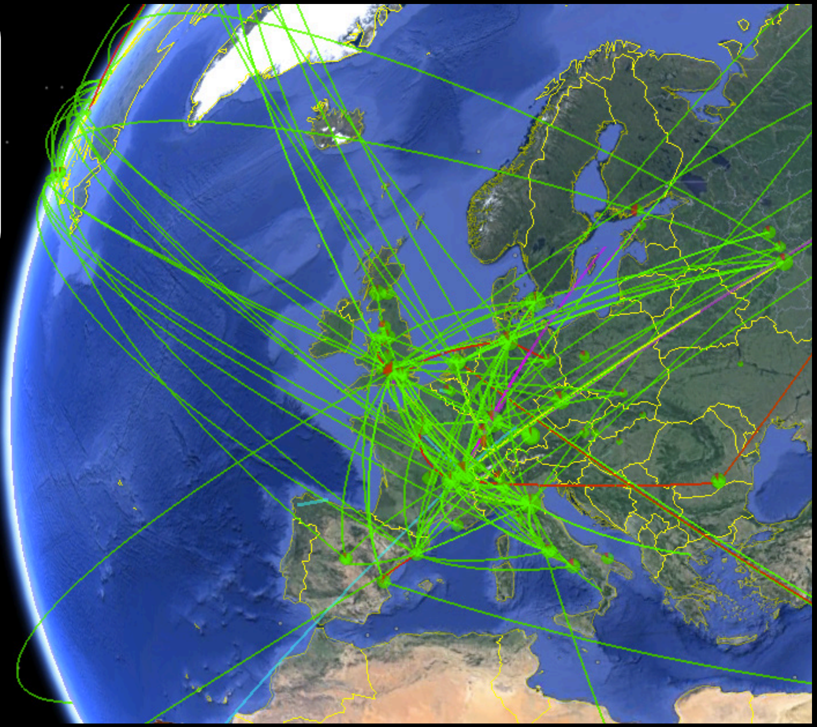
CPU	20000 HS06	100 server (87 WN's) 1376 cores 10.240 TB RAM
DISK	300 TBn + 950 TBn + 500 TBn	GPFS + EOS FS + GlusterFS
TAPE	80/320 TBn (Backup)	Quantum Scalar i2000 (2xLTO4/2xLTO3)
NET	Storage 10 Gb/s Ethernet 8 Gb/s Fibre Channel	Ethernet 1 + 10 Gb/s GARR 10 Gb/s

Scale of LHC computing needs today

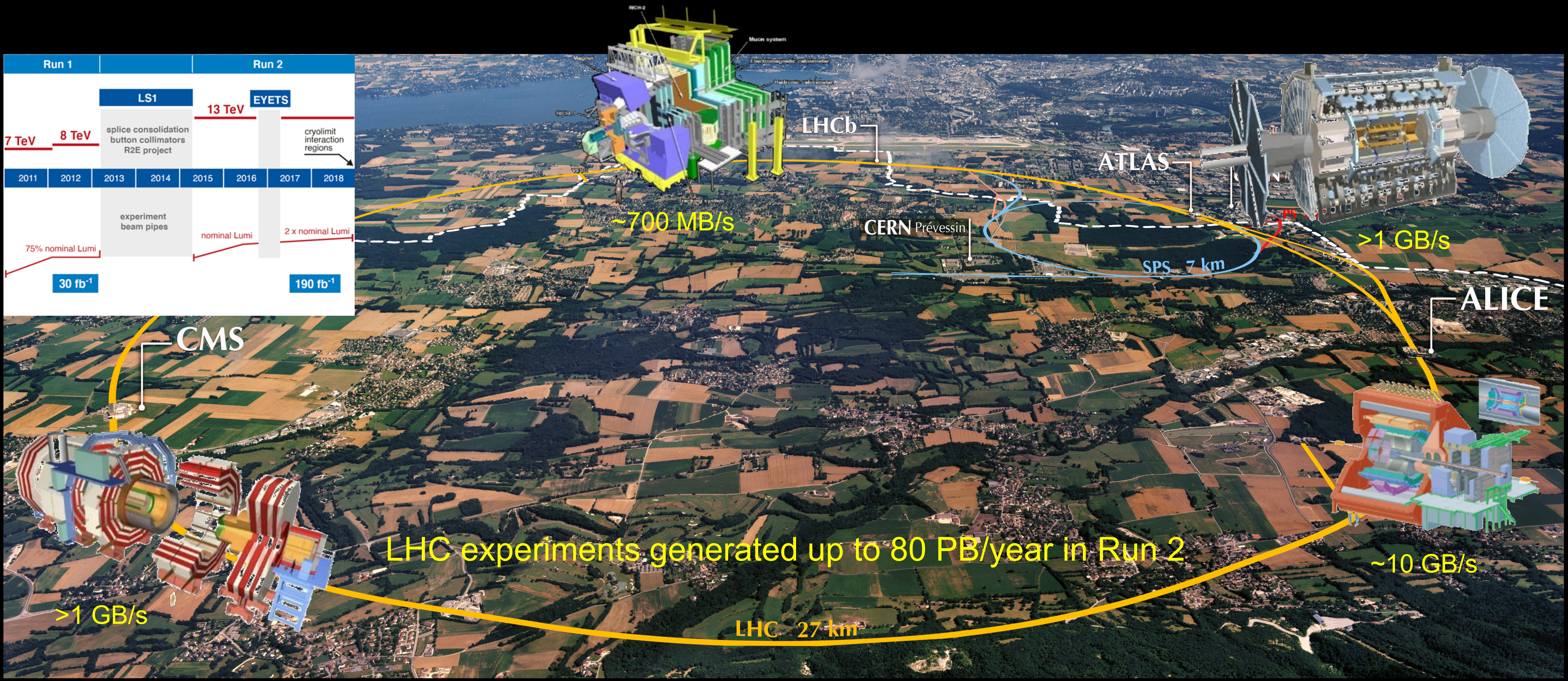
- CPU: >1 million cores fully occupied (~8.6 10^6 HS06)
- Entering the Exabyte-scale data (2022 pledges for all LHC exp's):
  - Disk: 750 PB
  - Tape: 1.2 EB

INFN: crucial contribution to LHC computing:
CPU 850 10^3 HS06
DISK 73 PB
TAPE 100 PB

- Networking:
  - LHCOPN 1.3 Tbps from Tier0 to 14 Tier1's
  - LHCOne overlay of 10-100 Gbps networks
    - LHC: Tier0 + 14 Tier1's + 155 Tier2's + Tier3's + AF's
  - Other HEP experiments share a part of such an infrastructure
    - BelleII, Pierre Auger, NovA, XENON, JUNO
  - Other sciences will use many of the same facilities

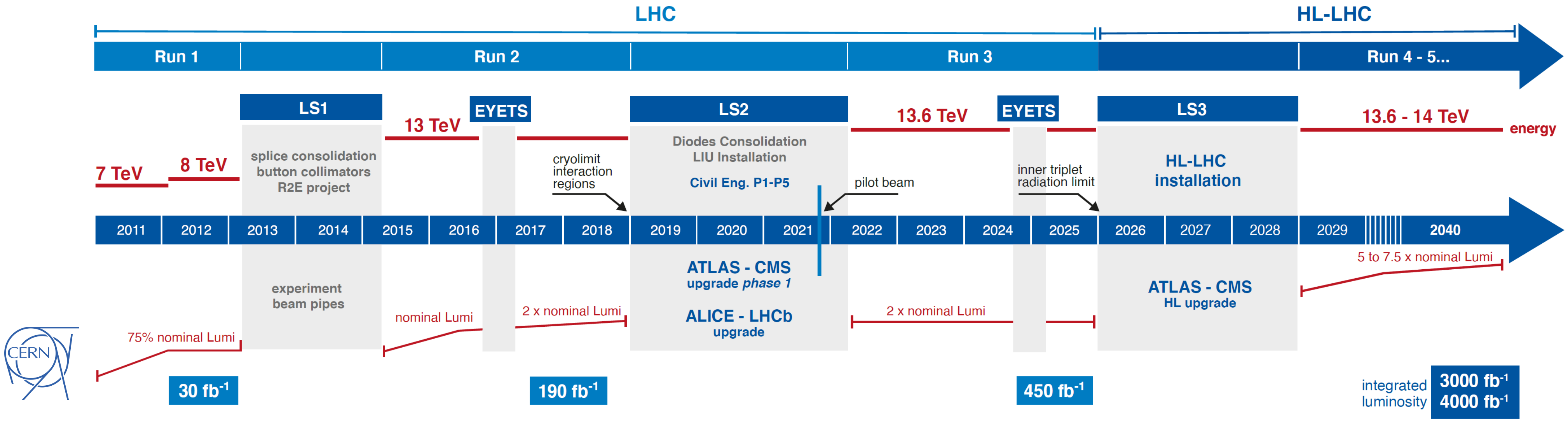


Run 2 data volume



I.Bird, WLCG: new challanges and collaboration with other science projects (LHCOPN/LHCONE 2020 workshop)

LHC/HL-LHC plan: LS2 upgrades



ALICE upgrade

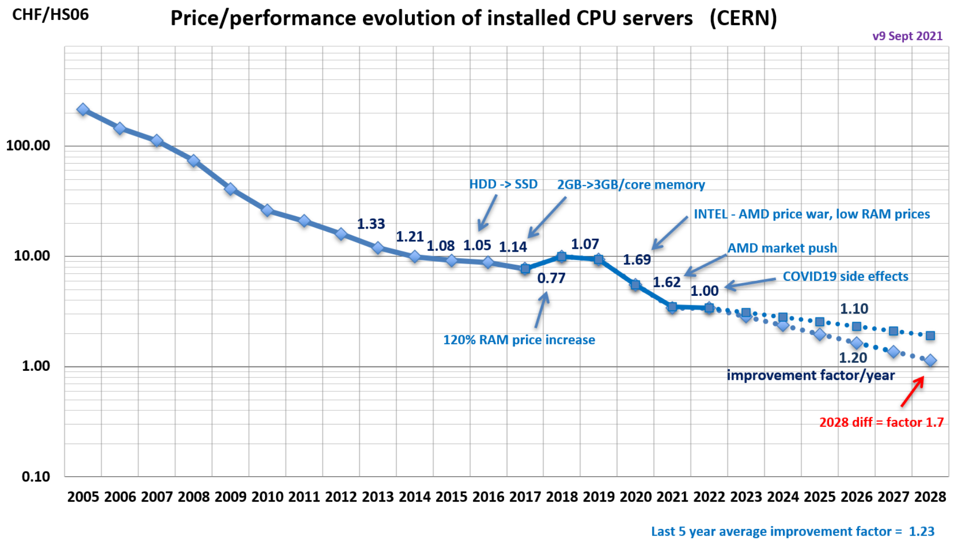
- From 1 kHz to 50 kHz Pb-Pb interaction rate
- Collect 13 nb^{-1} of Pb-Pb collisions at 5 TeV (inspected ~1 nb^{-1} during Run 1 and Run 2)
- 100x more recorded MB events wrt Run1 and Run2

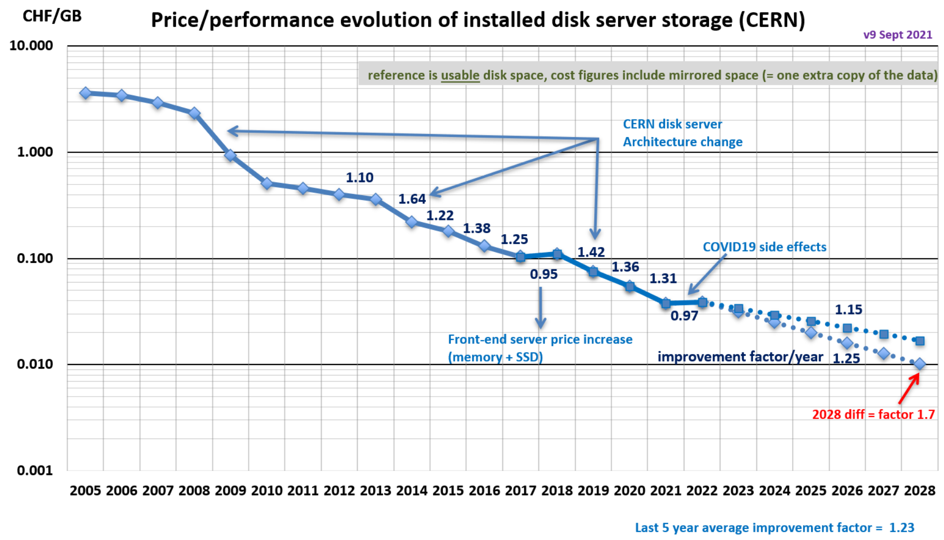
LHCb upgrade

- Raising the instantaneous luminosity by a factor five to 2 x $10^{33}cm^{-2}s^{-1}$
- Implementing a full software trigger to overcome limitation of L0 hardware trigger

Hardware cost evolution

- Hardware cost is more and more dominated by market trends rather than technology and science has no influence on these markets
- The assumption of +20%/year in storage and CPU for the same budget is still holding, but with large fluctuations and our budget outlooks are constrained (flat)



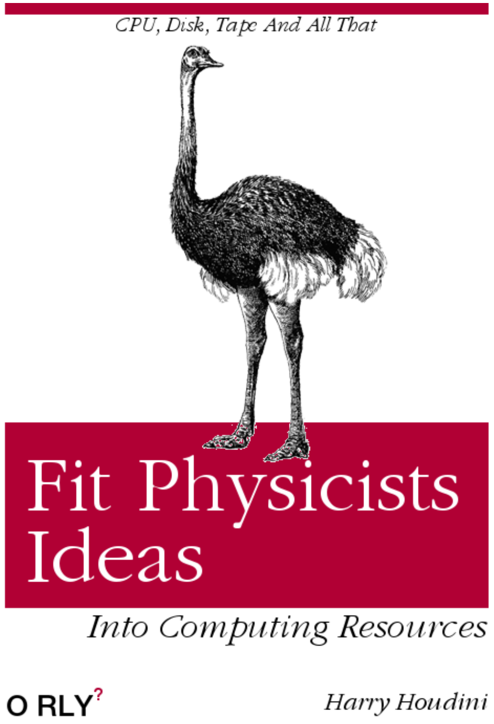


S.Campana, Computing - challenges and future directions (ECFA 2021)

A big challenge in data handling

- ALICE @ Run 3 and 4: 100 times more recorded collisions
- But ... projections assume constant funding every year for LHC computing
- Technology improvements will bring ~20% more resources every year
- Computing increases by factor 5 in 10 years (“flat budget” scenario)
- Need to gain factor 20 (disk and CPU) through smarter strategy and algorithms maintaining the physics performance

- Similar challenge for LHCb: 30x increase in throughput from the upgraded detector (10x physics event rate x factor 3 increase in average event size due to larger pile-up)

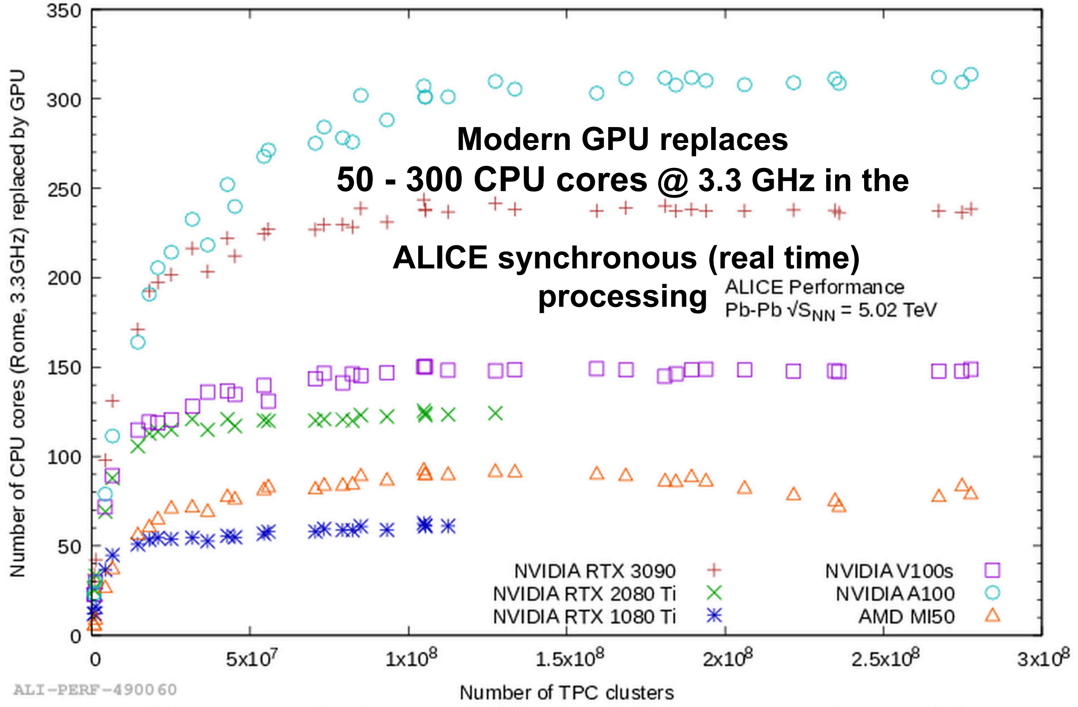


Courtesy of LHCb collaboration

C.Bozzi, Software e computing in LHCb: la sfida di Run3 (e oltre) (seminar @ CNAF 2021)
S.Piano, The ALICE O² computing model for Run 3 and 4 (seminar @ CNAF 2021)

Heterogeneous architecture

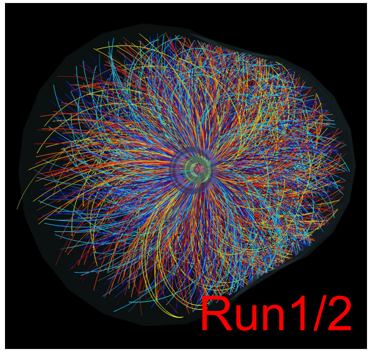
- Heterogeneous architectures: complementing CPU capacity with accelerators (e.g. GPUs)
  - GPUs offer more theoretical FLOPS in a compact package
  - Lower cost than CPUs per theoretical FLOPS
- Playing a fundamental role in Run-3 already, in most online systems. Non exhaustive examples:
- **ALICE:** Speed up from GPU usage + from algorithmic improvements + tuning on CPUs
  - porting of asynchronous (offline) reconstruction code to GPUs well advanced thanks to common online – offline O2 framework
- **LHCb:** exploitation of heterogeneous architectures, thanks to Allen framework:
  - for partial reconstruction in Run 3 (HLT1)
  - full reconstruction in Run 4 and beyond ?



D.Rohr, Usage of GPUs in ALICE Online and Offline processing during LHC Run 3 (vCHEP 2021)

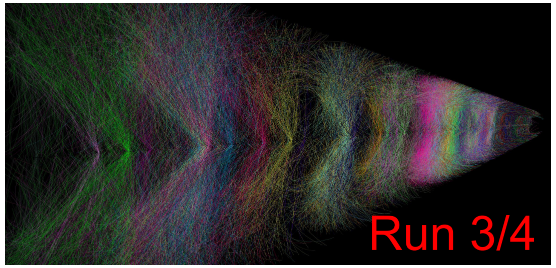
Run 3 ALICE data flow

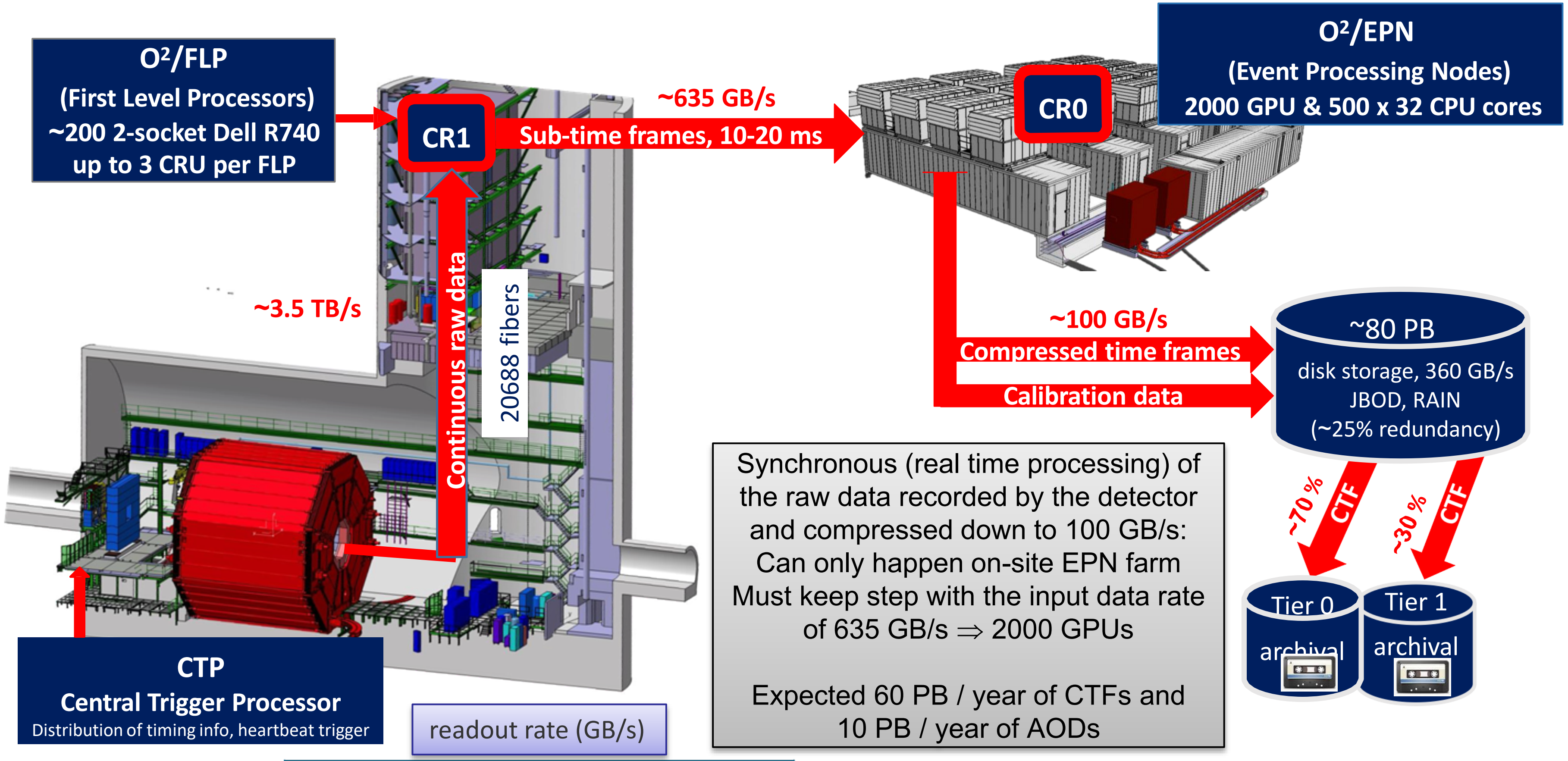
From triggered Run 1/2 to un-triggered Run 3/4



MB Pb-Pb event

Multiple events in one TimeFrame





Synchronous (real time processing) of the raw data recorded by the detector and compressed down to 100 GB/s:
Can only happen on-site EPN farm
Must keep step with the input data rate of 635 GB/s ⇒ 2000 GPUs

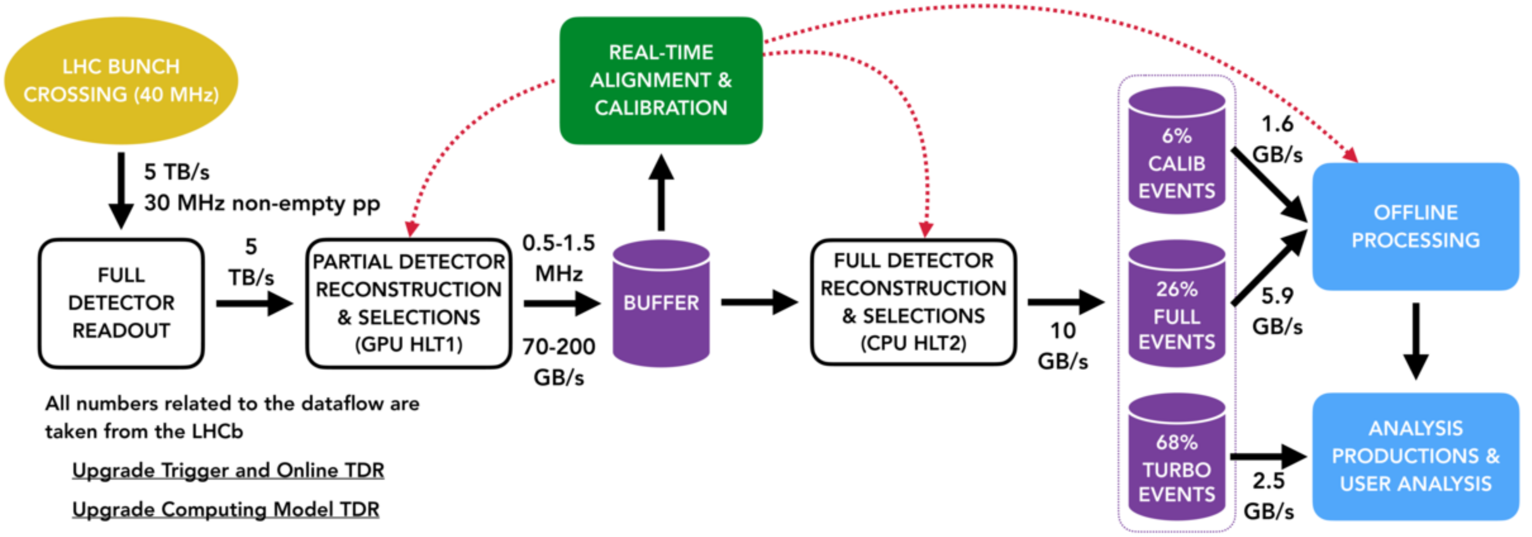
Expected 60 PB / year of CTFs and 10 PB / year of AODs

readout rate (GB/s)

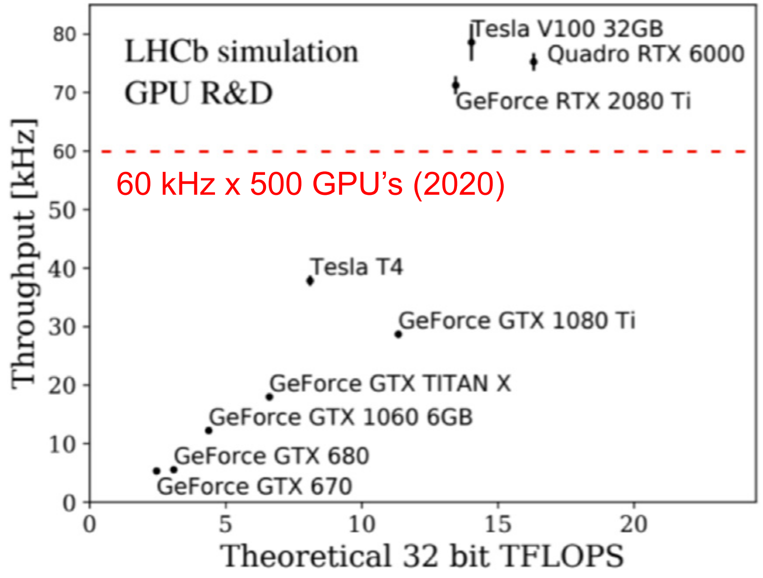
raw data: 3465			
TPC 3400	ITS 40	TRD 4	Others 21

R.Shahoyan, ALICE continuous readout and data reduction strategy for Run3 and 4 (CHEP 2019)

Run 3 LHCb data flow



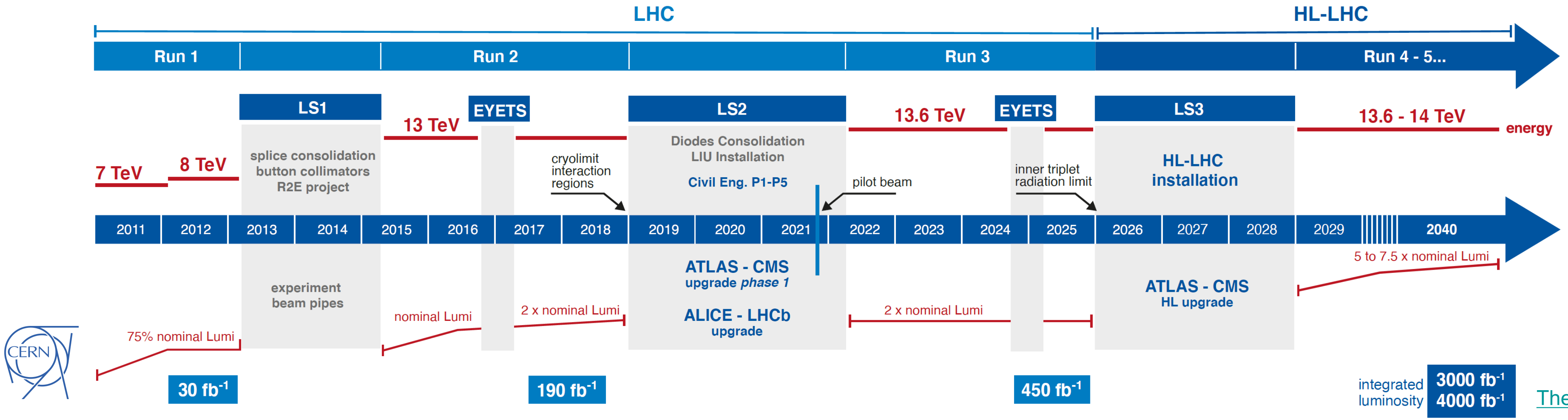
1 year later: software optimizations (Allen) + new consumer NVIDIA cards GeForce RTX 3090 reached ~233 kHz (~4x the red line shown on the plot)



- Remove Hardware trigger in favour of a fully software based one
- Event reconstruction at collision rate
- Full detector read-out at 40 MHz (visible collision rate: 30MHz)
- HLT1 173 EB servers with GPU and HLT2 up to 4000 servers

C.Bozzi, Software e computing in LHCb: la sfida di Run3 (e oltre) (seminar @ CNAF 2021)

LHC/HL-LHC plan: Run 4 and beyond



The HL-LHC project

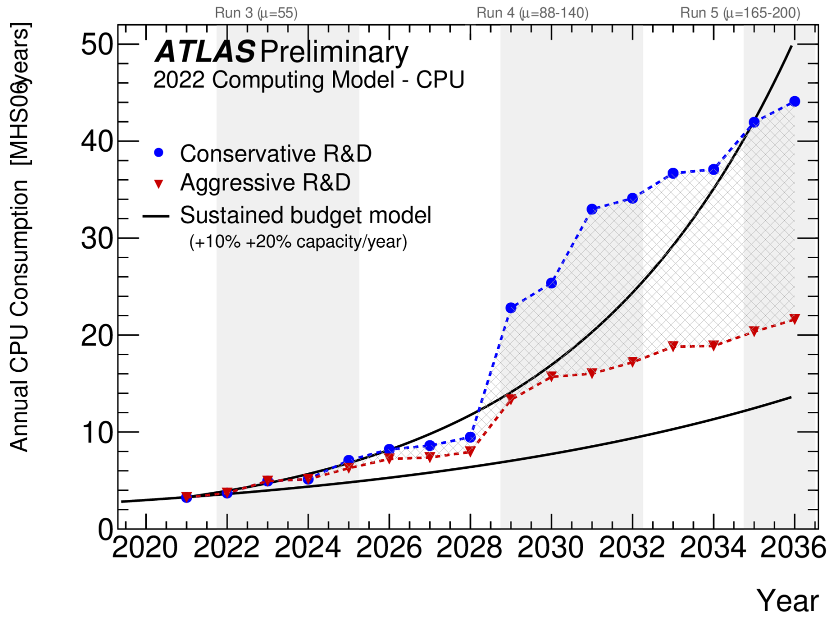
HL-LHC project

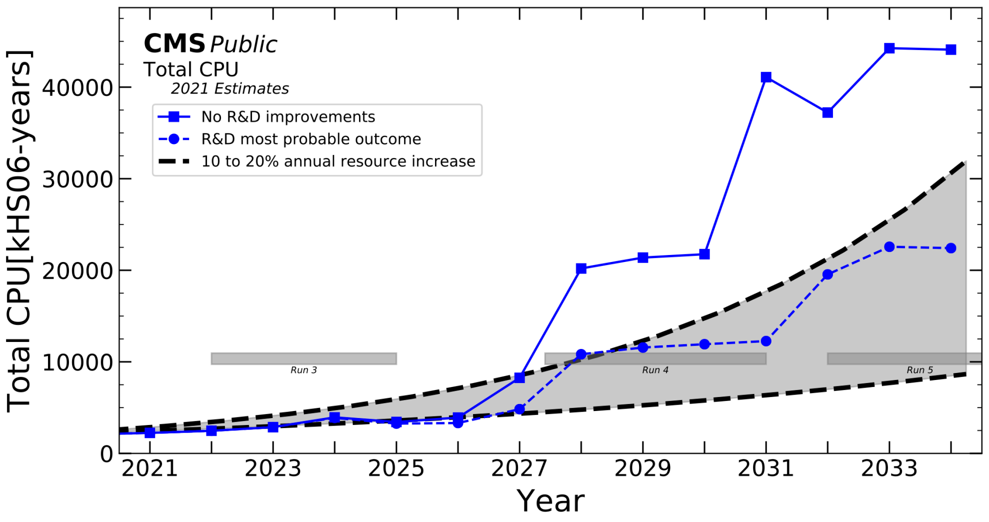
- Expected to be operational from 2029
- Objective is to increase the integrated luminosity by a factor of 10 beyond the LHC's design value
- During Run1/2 189 fb^{-1} delivered to both ATLAS and CMS, expected to provide 4000 fb^{-1} by 2040 (> 20x)

ATLAS and CMS HL upgrades

- ATLAS and CMS are facing the challenge of an instantaneous luminosity increase from $2x10^{34}$ $cm^{-2}s^{-1}$ (end of Run 2) to $7.5x10^{34}$ $cm^{-2}s^{-1}$ (Run 5)
- While for ALICE and LHCb Run-4 will be at a similar scale than Run-3. New plans for Run-5 and beyond.

Expected CPU needs for HL-LHC



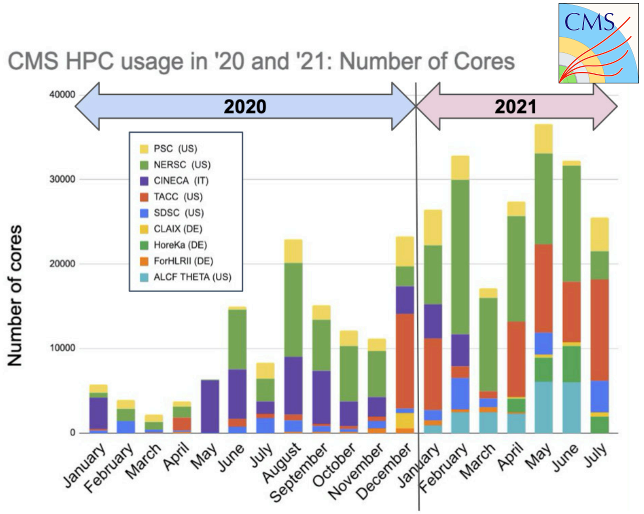


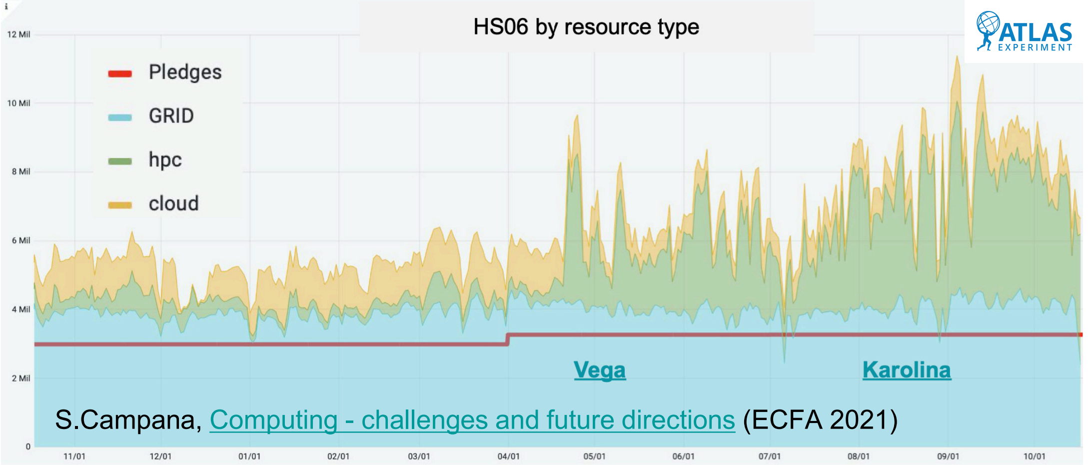
- The gap between available and needed resources can be filled up, assuming the main R&D activities are successful. There are still large uncertainties.
- Investing in further (identified) R&D activities would fill this gap further. Need more effort
- Development and retainment of software and computing expertise in the scientific community as well as with specialists becomes more and more crucial

ATLAS Collaboration, Computing and Software - Public Results
CMS Collaboration, Offline and Computing Public Results

High-performance computing

- A mitigation for the gap in resources comes from opportunistic CPUs.
  - HPC centers offer a unique opportunity: substantial investment of national and supranational entities (exascale)
  - Need to incorporate HPCs keeping integration costs in check, transparently for the operations
- Accessing and using resources at HPC centers comes with different challenges:
  - Diversity in access and usage policies, edge services, system architectures
  - Heterogeneous computing architectures: non x86 CPUs and different GPUs
  - CPU remains central; offload compute-intensive/specialized algorithms to compute accelerators
  - Opportunistic CPU but no opportunistic disk
- Software portability and the success in integrating accelerators will play an important role:
  - Requires significant investment in training and development
  - Much harder for smaller collaboration
  - In 2021: CMS used 10x more capacity at HPC sites wrt 2019,
  - and ATLAS exploited 2.2 MHS06 HPC CPU capacity





S.Campana, Computing - challenges and future directions (ECFA 2021)

Software R&D

Physics improvements vs cost of CPU is very difficult, **not** just save CPU, but better physics

- Event Generators
  - Not always optimized performance codes but pushing for precision physics
  - Reengineering for vector CPUs and GPUs
  - Reducing the fraction of events with negative weights

V.Pascuzzi, HL-LHC Computing Challenges - how can the physics performance be maintained? (LHCP 2021)

- Reconstruction
  - Experiment dependent
  - Exploit capabilities of modern systems and language features
    - SIMD, MIMD, SIMT, compute accelerators
  - R&D to offload part of the reconstruction code to accelerators (CMS Hcal)
  - ML techniques, e.g., GNNs, proving very effective

- Simulation
  - Most CPU-intensive offline software component
  - Workhorse of LHC simulation is Geant4
  - Complemented with fast simulation, including Machine Learning techniques Parameterization- and ML-based approximations to account for >50% of MC events during HL-LHC
    - GANs for particle shower (ATLAS) and particle identification (LHCb)
    - Other novel techniques, e.g., ReDecay, shower libraries
  - AdePT and Celeritas (Geant4 Task Force for R&Ds) aim for fully GPU-based detector simulation frameworks/toolkits
    - CUDA initial target, bearing in mind portability 'solutions'
    - Focus first on EM physics

Software R&D

Physics improvements vs cost of CPU is very difficult, **not** just save CPU, but better physics

- Event Generators
  - Not always optimized performance codes but pushing for precision physics
  - Reengineering for vector CPUs and GPUs
  - Reducing the fraction of events with negative weights

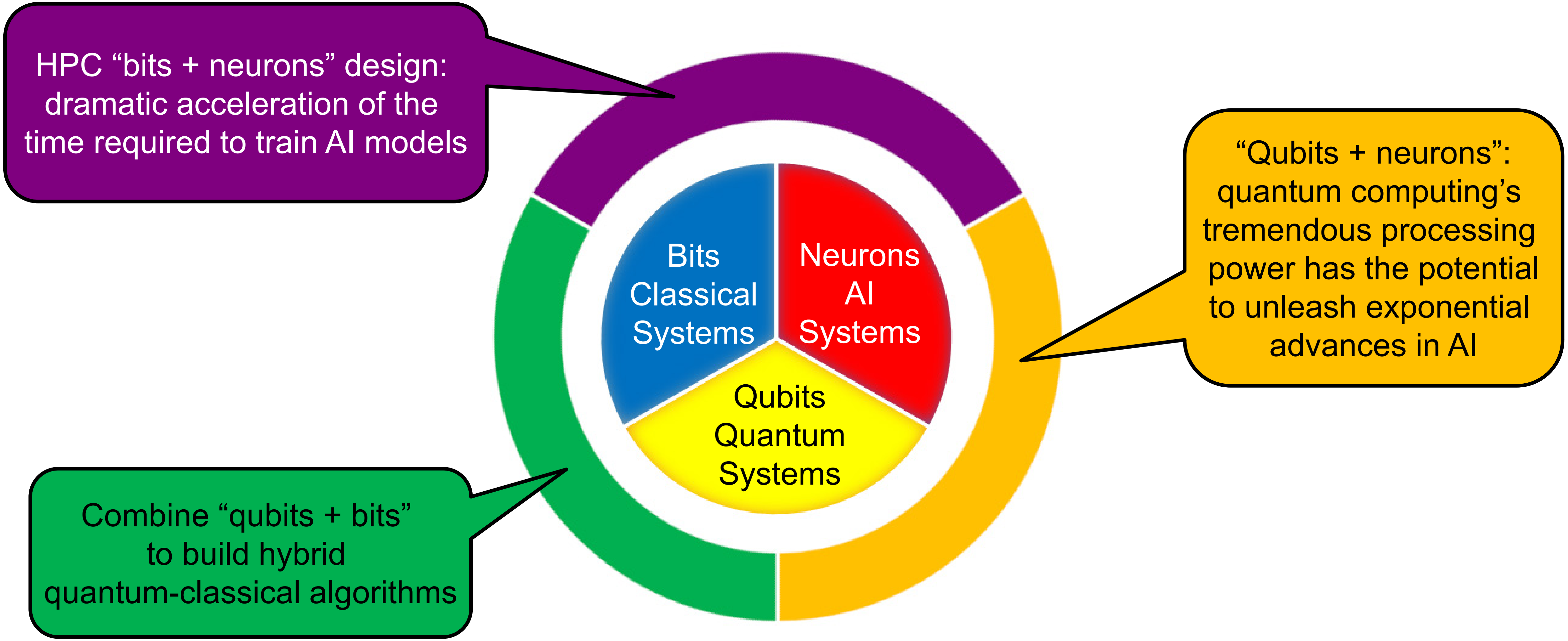
V.Pascuzzi, HL-LHC Computing Challenges - how can the physics performance be maintained? (LHCP 2021)

- Reconstruction
  - Experiment dependent
  - Exploit capabilities of modern systems and language features
    - SIMD, MIMD, SIMT, compute accelerators
  - R&D to offload part of the reconstruction code to

- Simulation
  - Most CPU-intensive offline software component
  - Workhorse of LHC simulation is Geant4
  - Complemented with fast simulation, including Machine Learning techniques Parameterization- and ML-based approximations to account for >50% of MC events during HL-LHC
    - GANs for particle shower (ATLAS) and particle identification (LHCb)
    - Other novel techniques, e.g., ReDecay, shower libraries
  - AdePT and Celeritas (Geant4 Task Force for R&Ds) aim for fully GPU-based detector simulation frameworks/toolkits
    - CUDA initial target, bearing in mind portability

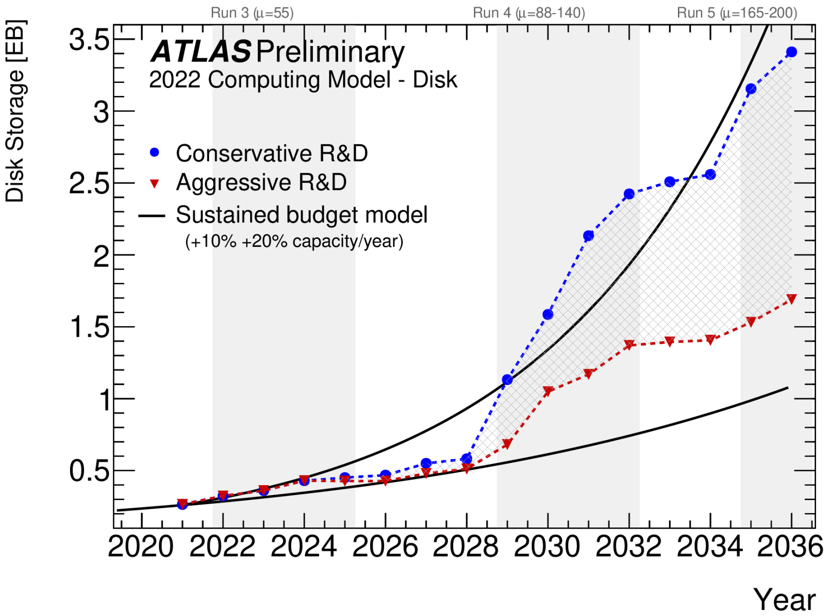
Needed talented Physicists skillful in SW development

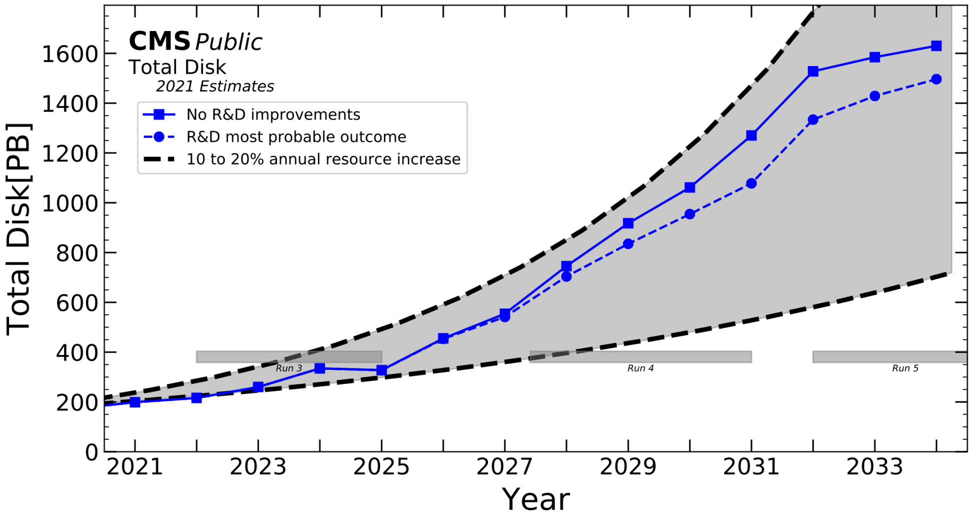
Key technology components of future computing systems



D.Gil and W.M.J. Green, The Future of Computing: Bits + Neurons + Qubits (ISSCC 2020)

Expected disk needs for HL-LHC

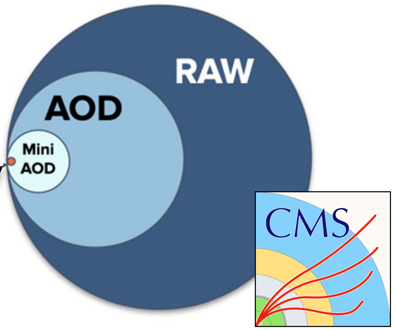




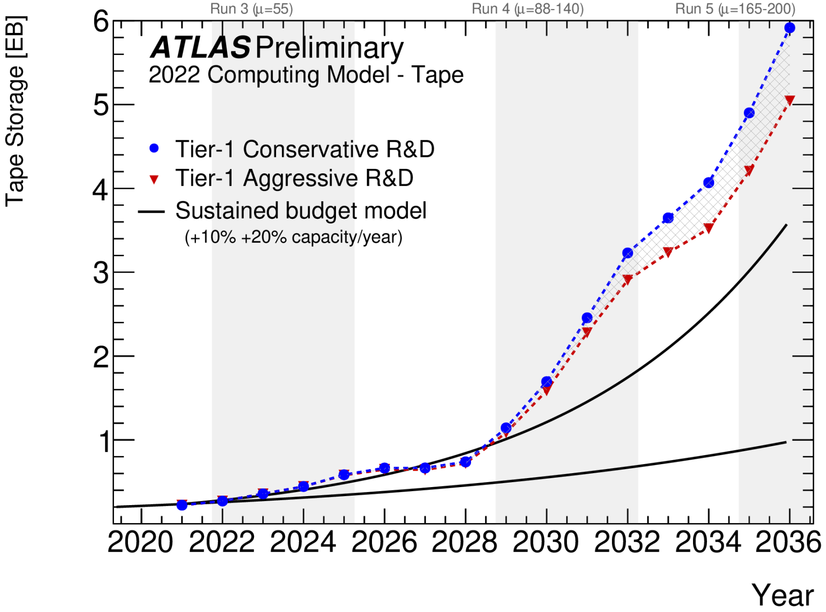
- Disk storage needs are dominated by the amount of reconstructed data, in different formats and versions. The strategy focuses on
  - Reduced analysis formats
    - Larger formats will not be generally needed on disk
  - Data carousels

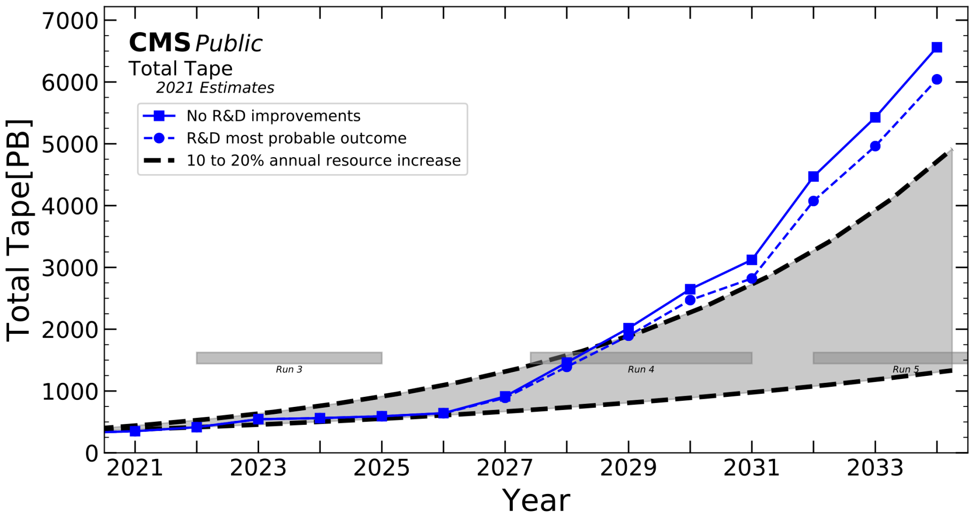
ATLAS Collaboration, Computing and Software - Public Results
CMS Collaboration, Offline and Computing Public Results

NanoAOD:
- 1kB in Run-3
- <5kB in HL-LHC



Expected archive storage needs for HL-LHC



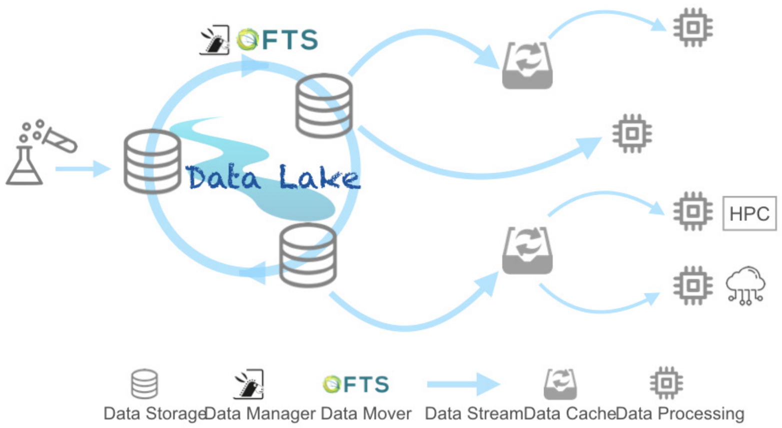


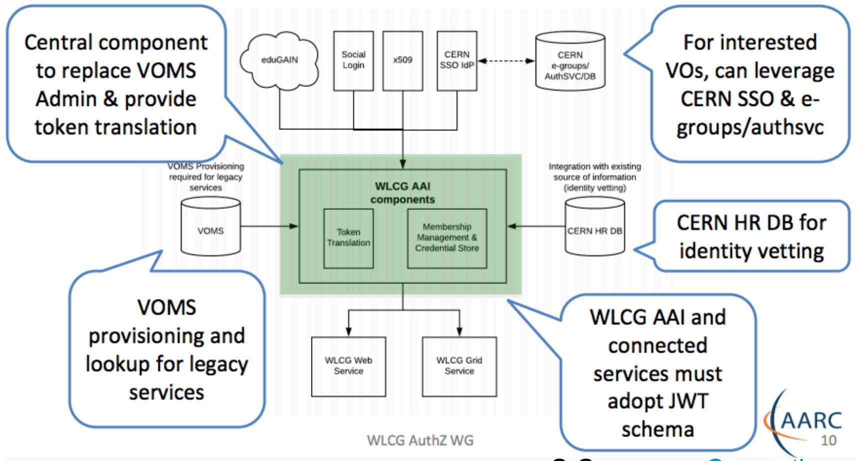
- Archive storage needs (tape) for RAW, AOD and MC data
  - Possible gains with compression/suppression, but moderate
  - Be prepared to invest in the tape volume we need and optimize the rest
- Management of exabyte-scale data
  - Infrastructure and tools

S.Campana, Computing - challenges and future directions (ECFA 2021)
ATLAS Collaboration, Computing and Software - Public Results
CMS Collaboration, Offline and Computing Public Results

Infrastructure and services sustainability

- The HL-LHC challenge is not just about resources.
- The sustainability of the infrastructure over the next 15 years is also challenging
  - Progress towards a more flexible model integrating a heterogeneous landscape of facilities, Data Lake, HPC, Cloud resources
  - Authentication and Authorization Infrastructure challenge
  - Network R&D activities, Data Transfer Optimization, Third Party Copy





S.Campana, Computing - challenges and future directions (ECFA 2021)

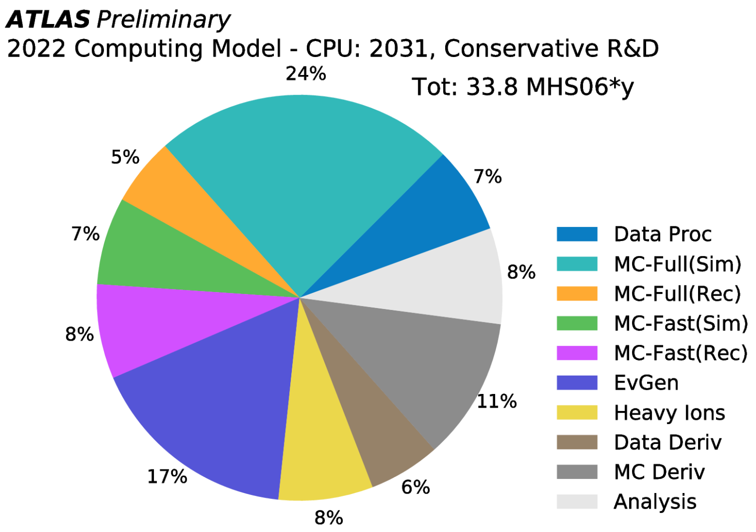
Summary

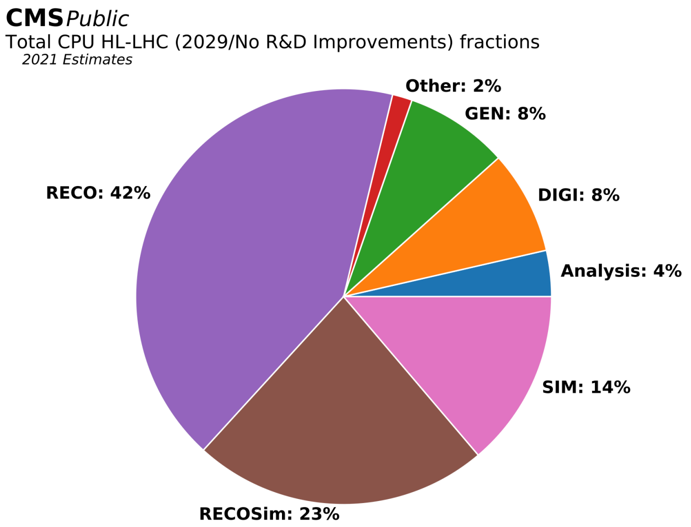
- LHC computing was very successful in providing the global environment for HEP physics
- HL-LHC presents major challenges for LHC computing
  - Management and analysis of exabyte-scale data
  - Keeping the computing needs within the fixed flat investment

- How can these challenges be overcome?
  - Increase software performance by adopting modern coding techniques and tools
  - Fully exploit the features offered by modern HW architectures
    - Execution of codes and validation of outputs across various compute resources
  - Novel developments exploiting parallel computing, ML and AI
  - Synergies and collaborations across scientific disciplines and with Industry partners
- Getting performant software and computing infrastructure requires significant investment in programming and computing skills
  - Training, sustained support and career paths for computing experts


Thank you!!!
INFN
TRIESTE
Istituto Nazionale di Fisica Nucleare
Sezione di Trieste

Expected CPU breakdown for HL-LHC





- ATLAS CPU needs: > 20% reconstruction, ~30% simulation (Geant4 + Fast Sim), 18% event generators
- CMS CPU needs are dominated by reconstruction of data and MC (>80%)

ATLAS Collaboration, Computing and Software - Public Results
CMS Collaboration, Offline and Computing Public Results

Event generators, GEANT4 and ROOT are common software libraries used by both ATLAS and CMS. All the other HEP experiments can profit from their improvements.