ATLAS S&C toward the HL-LHC challenges

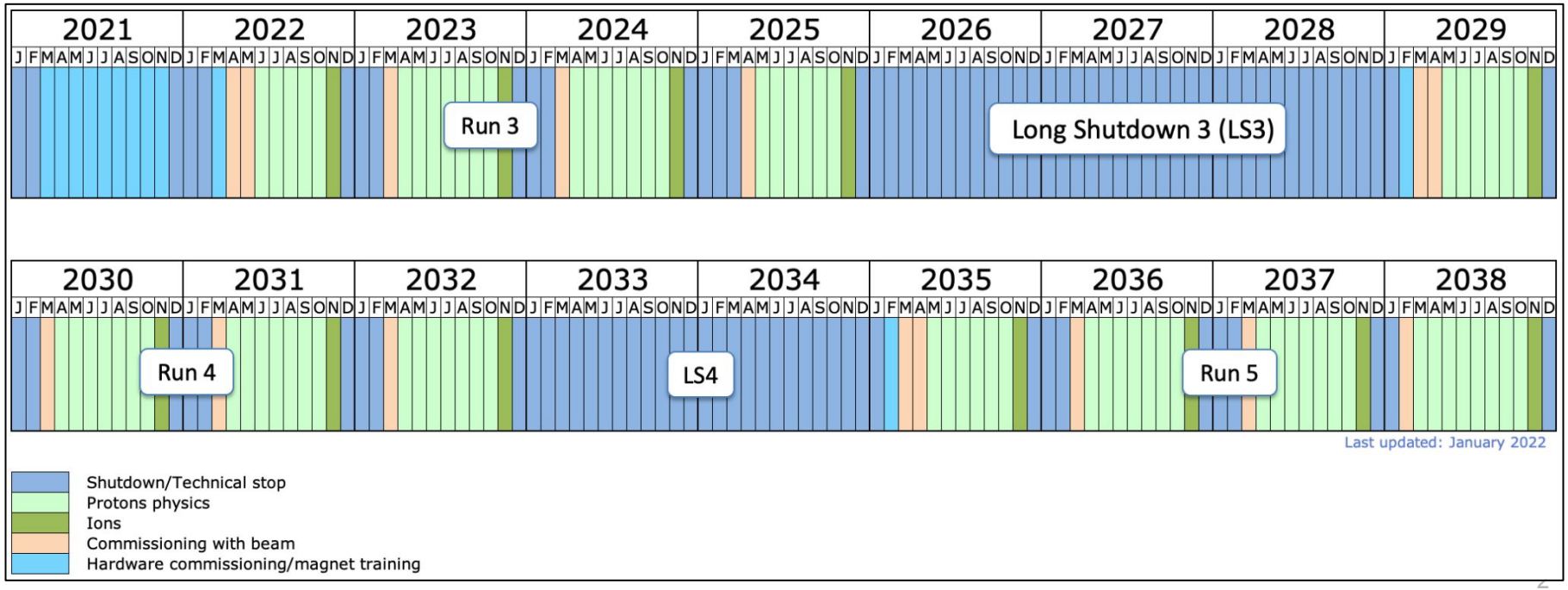
HEP-CCE All Hands Meeting

12 April 2023

Alessandro Di Girolamo (CERN) and Zach Marshall (LBNL)

LHC schedule

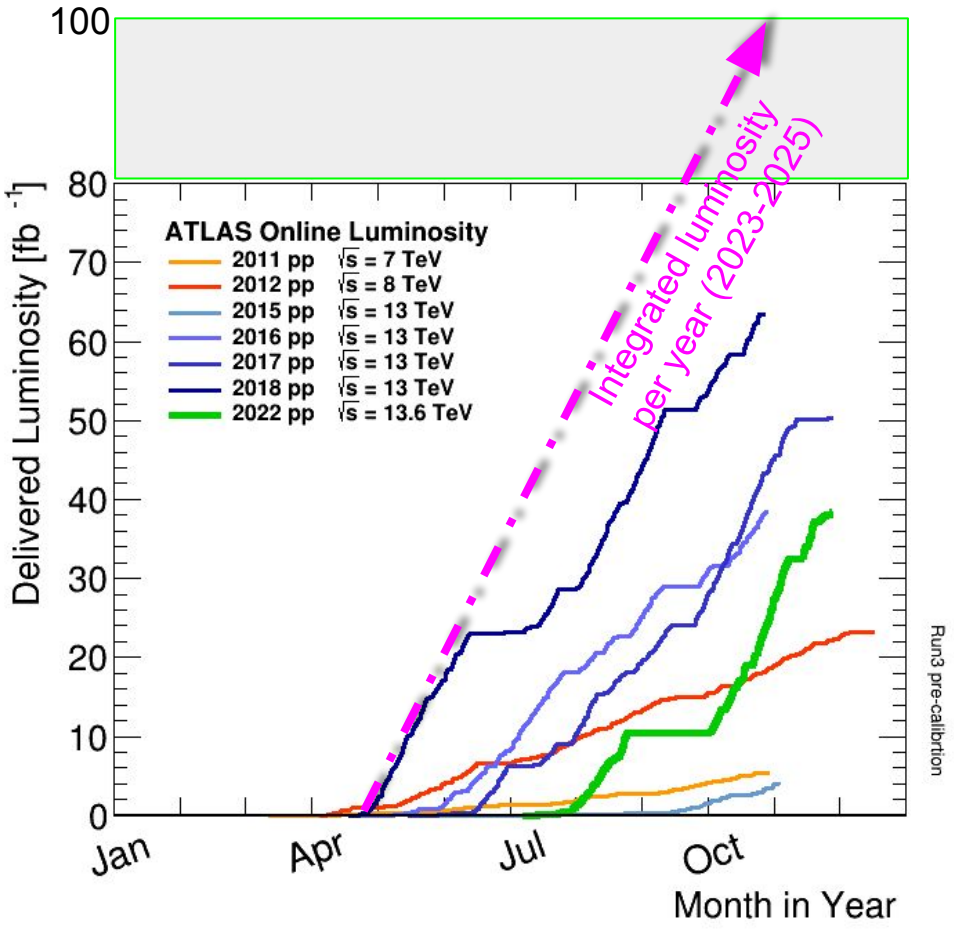
- We are just entering the second year of Run 3
  - Run 3 will end in 2025
- HL-LHC will start in 2029
  - Run 4: 2029–2032, Run 5: 2035–2038. Formal plan extends beyond 2040.



The Physics challenge: Run 3

Our “interesting” data set size is measured in fb^{-1}

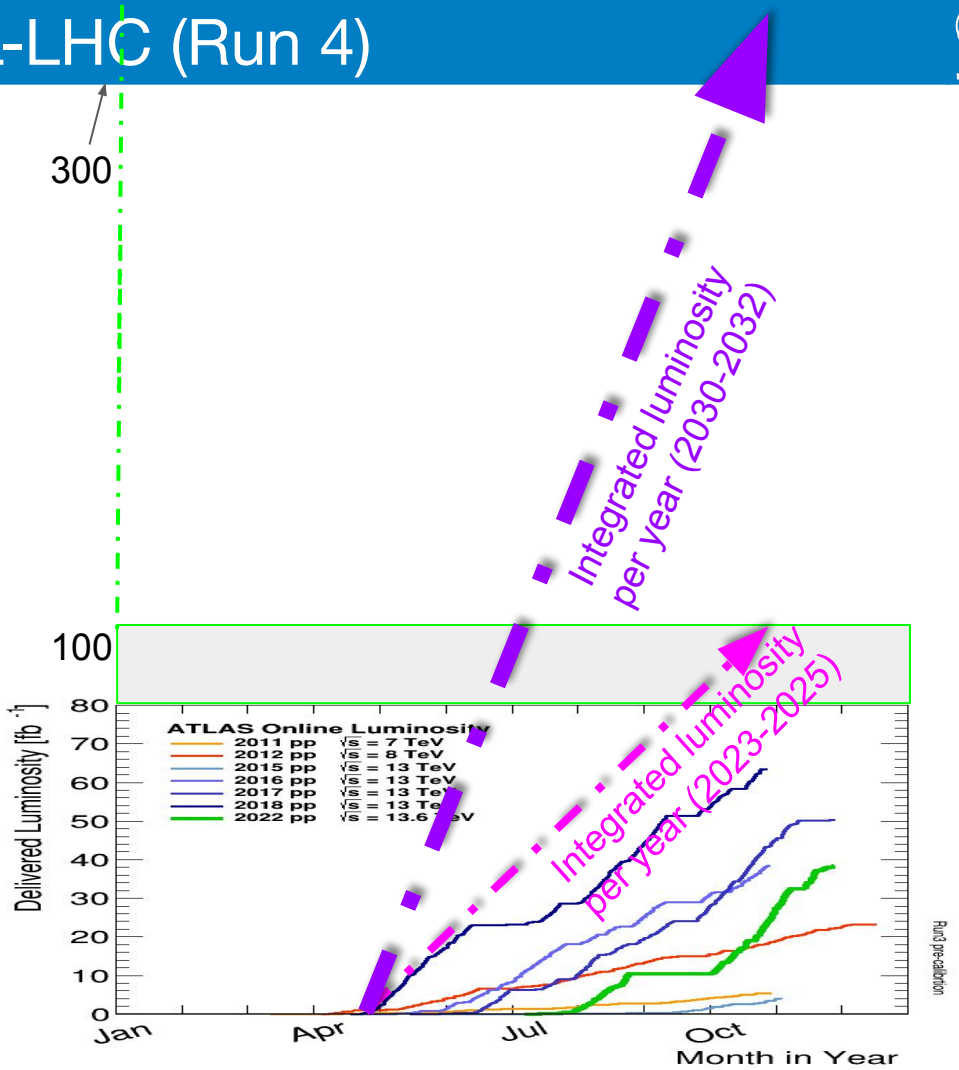
- Trigger rate (incl. delayed) ~ 3.5 kHz
- RAW size ~ 1.3 MB/ev
- Pile up ~ 60
- ~ 30B MC events/year



The Physics challenge: HL-LHC (Run 4)

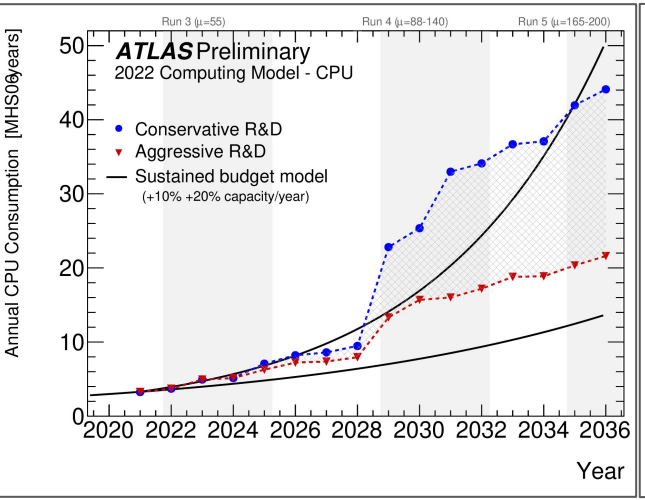
Our “interesting” data set size is measured in fb^{-1}

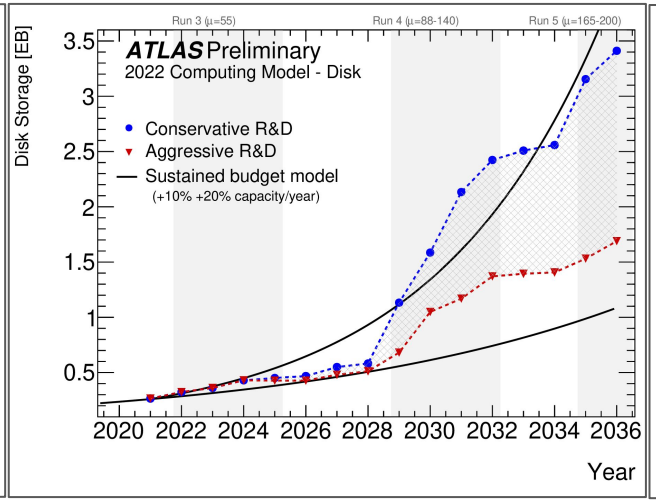
- Trigger rate ~ 10 kHz
- RAW size ~ 4 MB/ev
- Pile up ~ 140+
- ~ 150B MC events/year

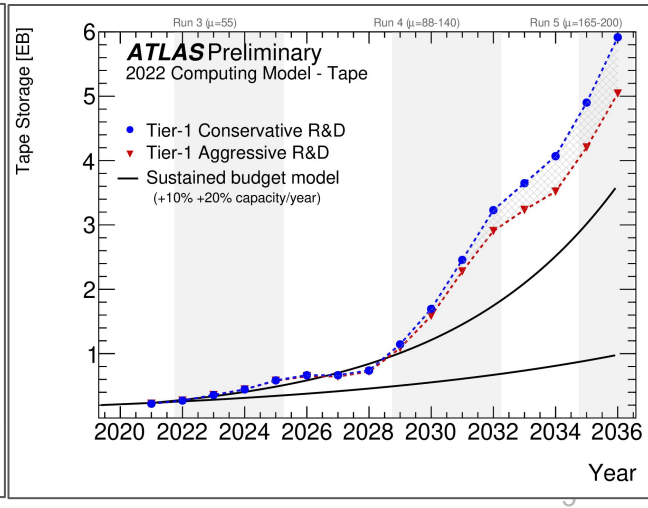


The Computing Challenges

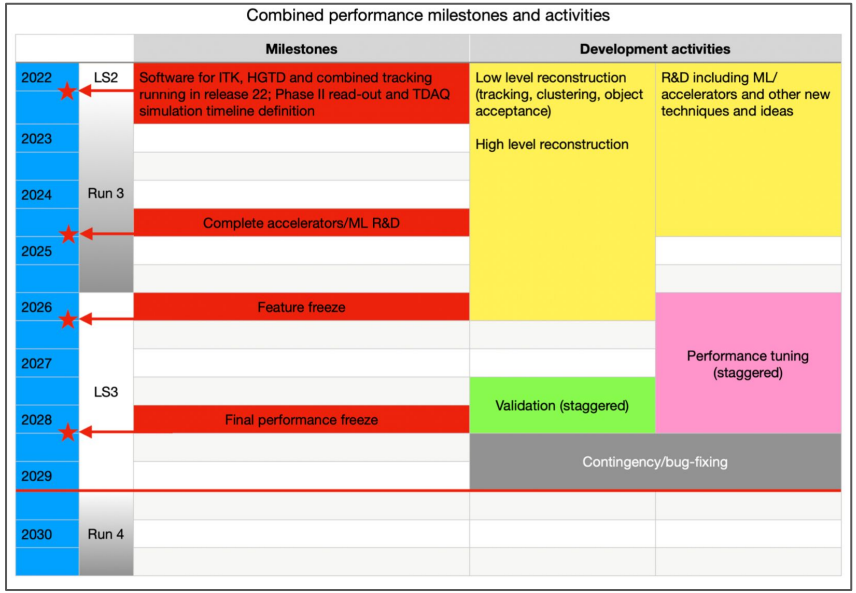
- We hope you're all familiar with these plots by now!
  - Model updated in Spring 2023 — no significant change to these projections
- Need for major R&D (or budgetary) effort to achieve HL-LHC physics potential
  - We have defined *Conservative R&D* and *Aggressive R&D* scenarios
    - N.b. some projects for which we were/are not able to estimate the concrete impact are not (yet) included (e.g. GPU usage, FastChain simulation)
  - The black lines indicate the "flat budget" of 10% (lower line) and 20% (upper line)







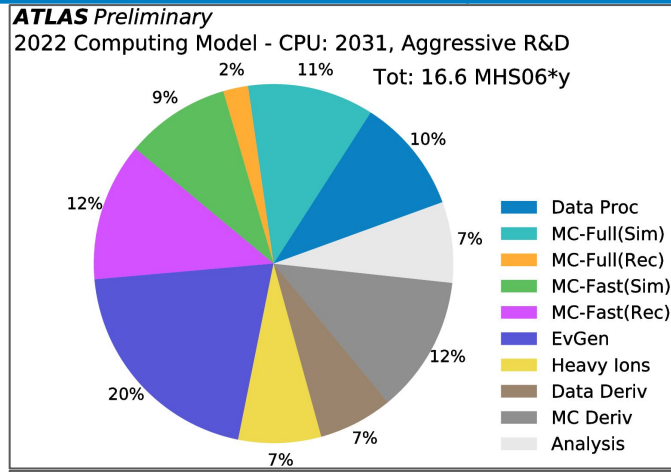
Getting There from Here

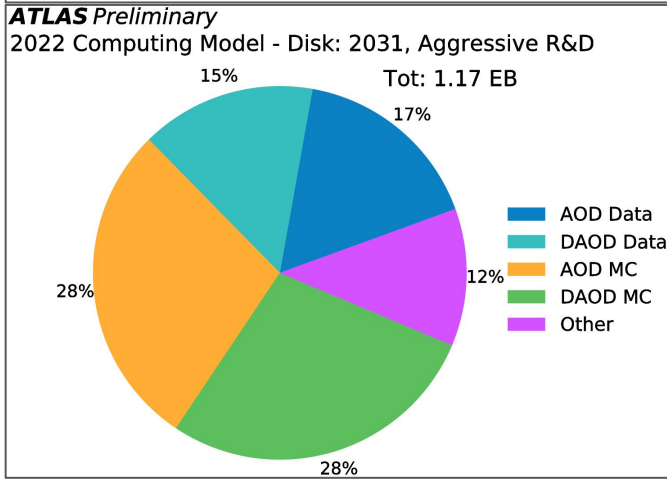


- ATLAS S&C HL-LHC Roadmap
- Defines milestones and deliverables to get to the HL-LHC successfully
- R&D is ongoing
  - We've just spent some time discussing and reviewing "demonstrator" prototype projects for the HL-LHC
  - We see already lot of engagement!
- Integration and validation will require **time**
  - Late arriving R&D is risky
- In 2025 we expect to have in the TDR a detailed path to HL-LHC data taking
  - For example, accelerators: Yes or No
  - This is also the timeline for a decision from our trigger group on accelerators

Pain Points (Processing / CPU)

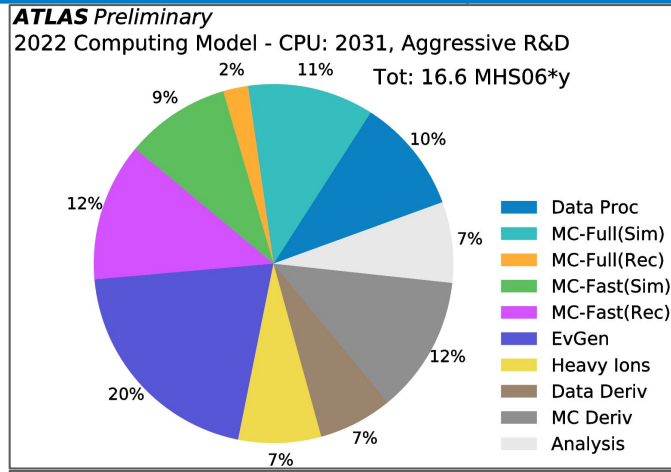
- No single application dominates CPU in 2031
- That's good news and bad news for us
  - No silver bullet to "solving" our resource crunch
  - Also not fatal if one workflow isn't improved
  - Can diversify our R&D — lots of interesting projects!
- Lots of effort and ideas around these problems
  - Very happy with the diverse portfolio HEP-CCE is investing in; many of the key problems are covered
- Biggest (by some metric) "CPU" efforts currently in:
  - Faster simulation (Geant4 optimization / on GPU, better fast simulation, FastChain…)
  - ML/accelerator-based charged particle tracking
  - New approaches to analysis
- Effort spread around reasonably well
  - Other ML/accelerator approaches to reconstruction, event generation, etc
  - Cleaning up "waste" (e.g. unused / failed production)

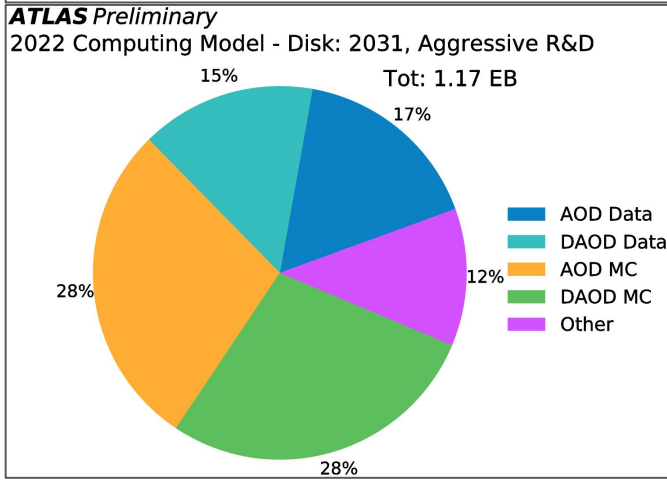




Pain Points (Storage)

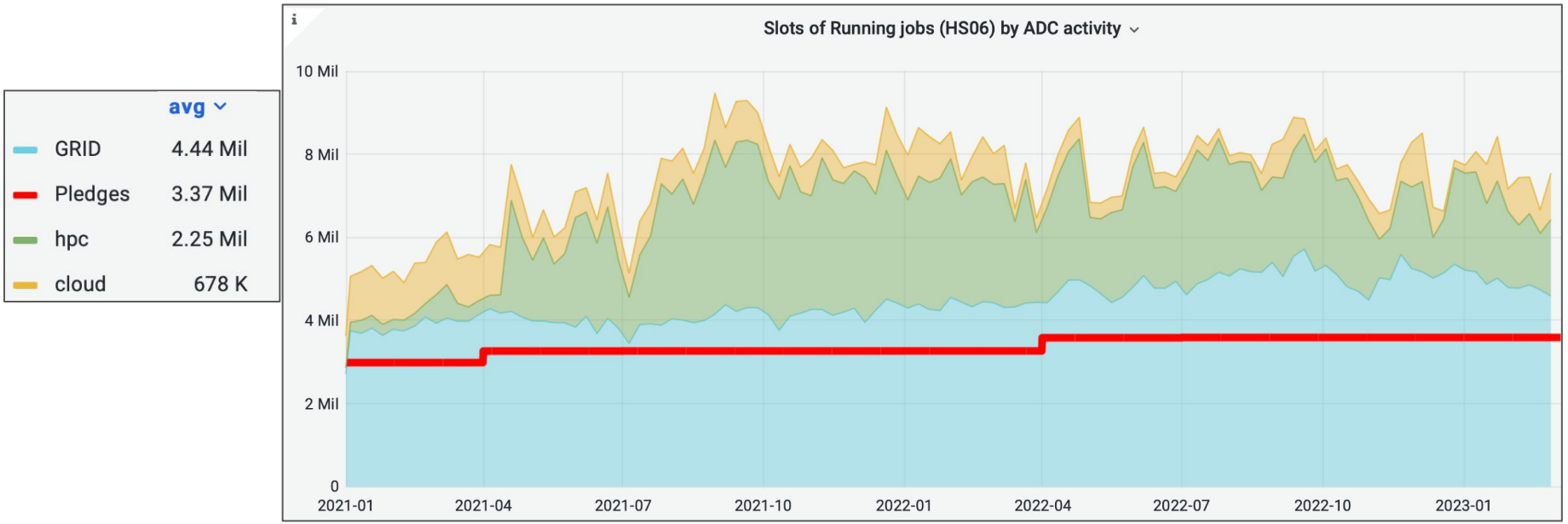
- Several ongoing disk efforts as well
  - RNTuple (of course)
  - Lossy compression: difficulty is not the infrastructure but the physics validation
  - Augmentation to support sparse additional data
  - Constant revision / review of file contents
  - Alternative compression/settings — delicate balance
  - More aggressive deletion / recreation
- Anticipating our PHYSLITE format will serve a wide variety of analyses in Run 4
  - The disk model is driven by “remnants”: how many analyses *don’t* use PHYSLITE, and what they use instead. This is where the hard work goes!
- Very successful model for data distribution
  - Our “data carousel” uses tape effectively as a warm storage medium, reducing disk needs
  - Already have a mechanism for replication of popular datasets, and expecting to continue this way





High-Performance Computing

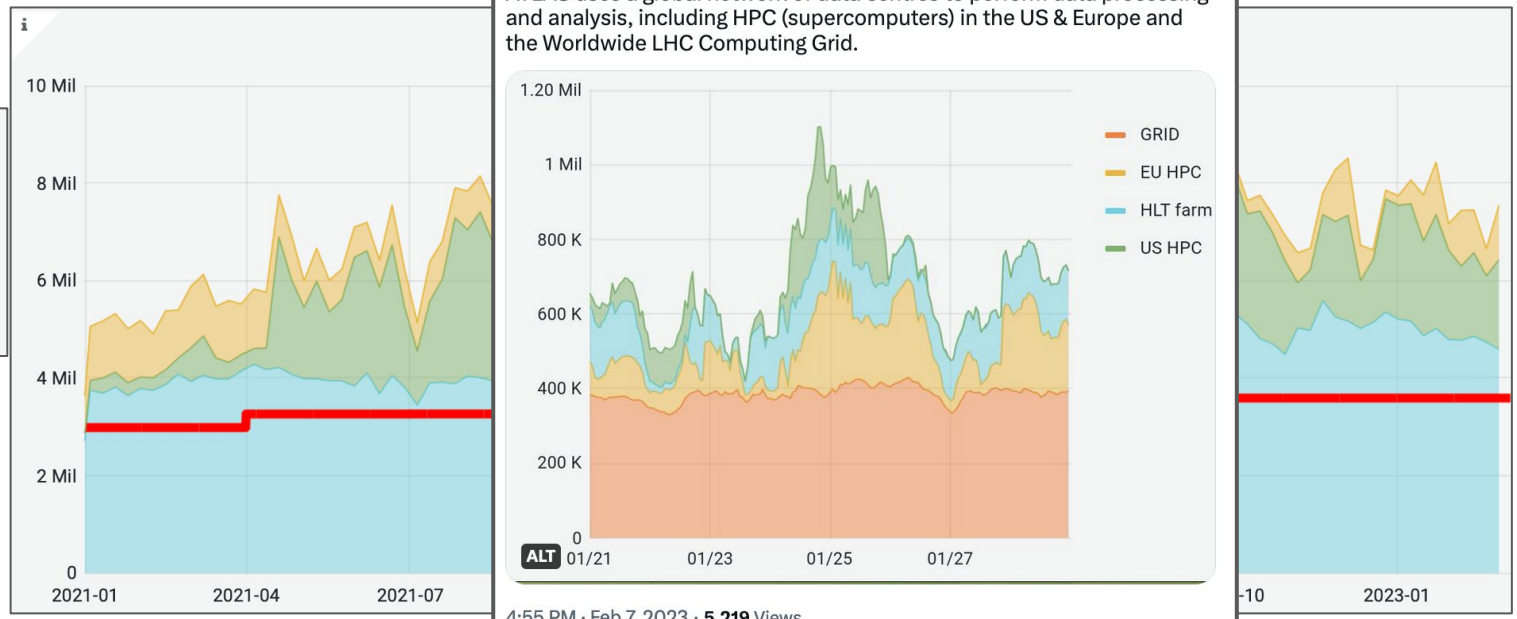
- We have benefited enormously from HPC systems (especially since 2021)



High-Performance Computing

- We have benefited enormously from HPC systems (especially since 2021)
- Recently allowed us to cross 1M simultaneous cores for the first time

	avg
GRID	4.44 Mil
Pledges	3.37 Mil
hpc	2.25 Mil
cloud	678 K



ATLAS Experiment ✓
@ATLASexperiment

New record! 💻 For the first time, over 1 million CPU cores simultaneously contributed to ATLAS computing.

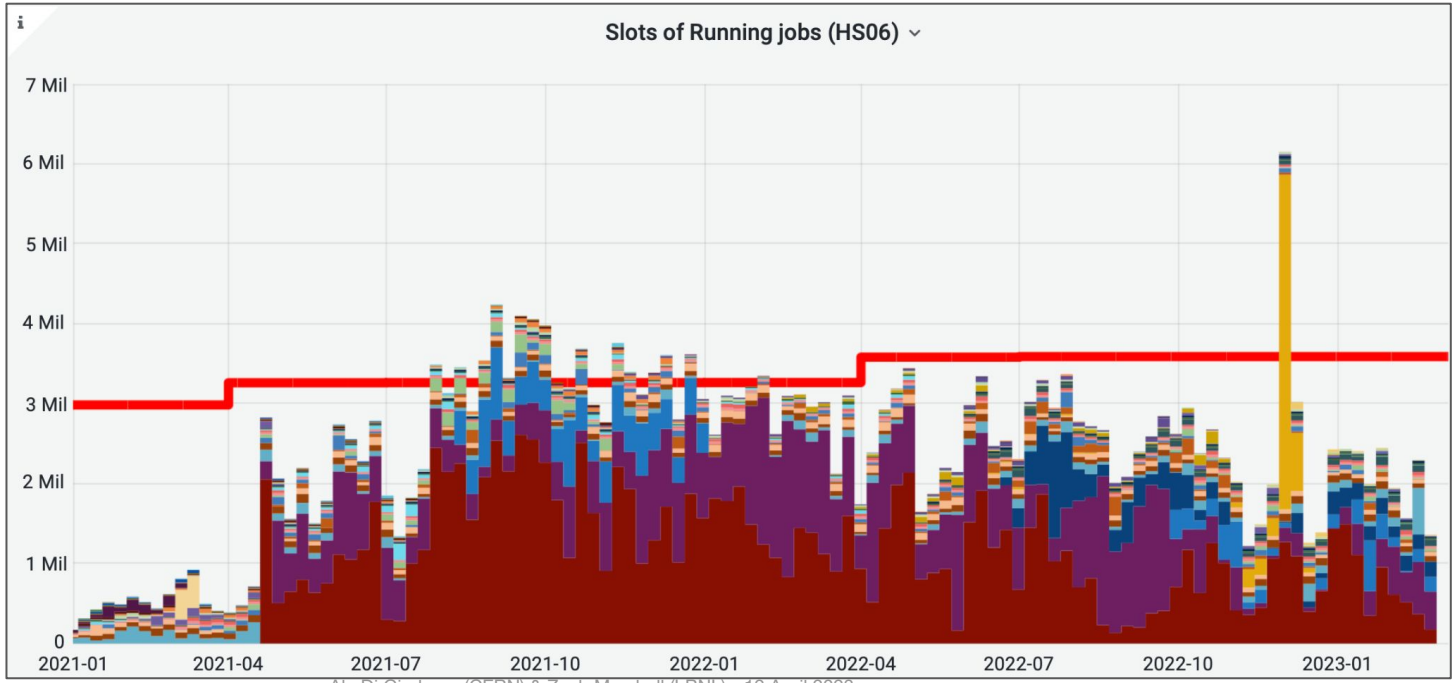
ATLAS uses a global network of data centres to perform data processing and analysis, including HPC (supercomputers) in the US & Europe and the Worldwide LHC Computing Grid.

4:55 PM · Feb 7, 2023 · **5,219** Views

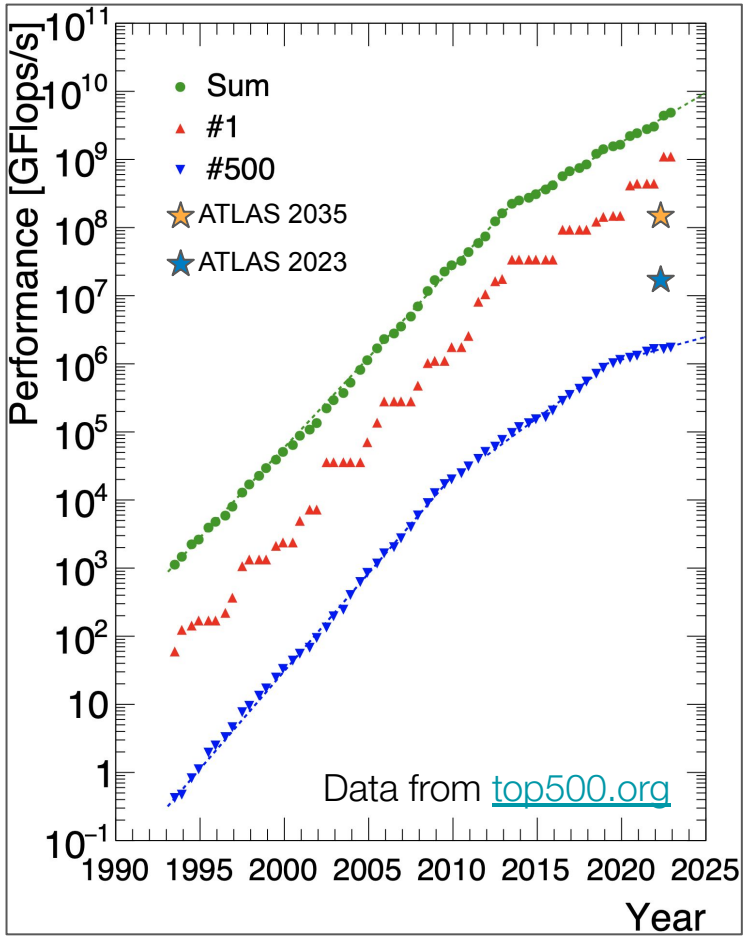
High-Performance Computing

- We have benefited enormously from HPC systems (especially since 2021)
- Recently allowed us to cross 1M simultaneous cores for the first time
- We have a nice mixture of “transparent” and “complex” HPCs today
  - Transparent: Mare Nostrum, Norway, Vega, Karolina, CSCS, (Leonardo), …
  - Complex: Cori/Perlmutter, Toubkal, …

	max	avg
Pledges	3.59 Mil	3.37 Mil
Vega_MCORE	2.61 Mil	1.02 Mil
Vega	1.96 Mil	600 K
praguelcg2_Karolina_MCORE	899 K	134 K
Vega_FULLNODE	1.21 Mil	86.8 K
NERSC_Cori_p2_mcore	582 K	70.8 K
CSCS-LCG2	109 K	67.3 K
RIVR.UM	127 K	58.4 K
NERSC_Perlmutter	4.19 Mil	51.0 K
Vega_largemem	238 K	36.6 K
pic_MareNostrum4	197 K	35.3 K
LRZ-LMU_MUC	226 K	28.7 K
NSC	36.5 K	25.9 K
IFIC_MareNostrum4	55.3 K	21.6 K
CSCS-LCG2-ALPS	108 K	18.1 K
TACC-FRONTERA-UCORE	203 K	16.4 K
HPC2N	43.5 K	16.3 K
CERN-HPC	37.3 K	13.1 K
praguelcg2_Barbora_MCORE	114 K	12.2 K
pic_MareNostrum4_Test	158 K	12.0 K
TOKYO_HPC	394 K	10.4 K
UAM_MareNostrum4	50.0 K	10.1 K
praguelcg2_Salomon_MCORE	79.8 K	9.95 K
LRZ-LMU_MUC_ES	113 K	9.07 K
CERN-HPC_ES	35.8 K	6.83 K
ALCF_Theta	138 K	6.37 K
UM6P	34.7 K	3.69 K
DESY-HH_HPC	14.3 K	3.60 K
MPPMU-DRACO_MCORE	43.2 K	3.19 K



HPC and Heterogeneity Philosophy



- We need more compute for the HL-LHC
- We need to keep our resources diverse
  - Expect HPCs to be an important component
- We do **not** need to run everywhere!!
  - Even a small HPC today could deliver a huge fraction of our required cycles
  - Vega is barely top 100 and is easily our #1 site — and we only use the CPU partition
  - This means we can find *the most friendly* HPC machines to use
- Corollary: we don't need to run on all hardware
  - We are validating ARM now and have some GPU developments in the pipeline
  - It appears likely that ARM+CPU *might* already be sufficient for us
  - Portability languages will be key for us to port to other hardware

HEP-CCE Project Questions / Comments (I)

- We haven't seen a lot of visible impact (yet?) from some HEP-CCE projects
  - Maybe that's natural — HL-LHC isn't here yet, and these things are hard to deploy
  - OTOH, several ATLAS members deeply involved in HEP-CCE and doing great things
- Our Roadmap includes a decision about (portability) languages in 2024
  - This is very much because we expect to learn from HEP-CCE's recommendations in 2023
  - ATLAS members are deeply involved in this effort; we don't think you're missing anything
- We are aware of the IOS activities (and ATLAS members are involved)
  - We haven't seen a killer use-case for e.g. the HDF5 writing yet
  - When we talk about "I/O limiting" in our applications, it is *rare that we really mean the I/O layer itself*. Usually we mean I/O+decompression+ROOT overheads+access overheads, and recent measurements indicate that these last two are the most important to us.
- Offloading (e.g. to GPU) data efficiently is an ongoing project in our Roadmap
  - Perhaps looking for further overlap/collaboration here would be useful
- Intelligent data compression is certainly interesting to us
- Without thinking deeply about it:
  - Metadata is an ongoing fight for ATLAS in a variety of ways. Perhaps there's some opportunity here for collaboration / harmonization / common tooling?

HEP-CCE Project Questions / Comments (II)

- It isn't clear to us what benefit object store R&D can bring
  - We're happy to learn more, of course
- We expect to run significant LO, NLO, and (hopefully) NNLO generation in HL-LHC
  - Accelerating any of these is welcome; we will be happy to test prototypes when ready
  - Our biggest (in CPU) sample is NLO Sherpa V+jets, followed by Powheg NLO ttbar
  - We generate many different samples with MadGraph (more configurations than Sherpa)
  - It would not be shocking (to me at least) if this were to remain the case in the HL-LHC
- GPU-based Geant4 would be a wonderful development
  - Recall that GeantV encountered *serious* difficulties when it got to hadronic physics, and neither Celeritas nor Adept has gotten there yet. Attacking this issue is important.
  - We are working on various steps to prepare ourselves (e.g. geometry offloading)
- We are aware of some of the FuncX/Parsl work
  - We don't see a big potential impact for this yet but are happy to learn more and discuss it

Bonus Point(s)

- Significant interest in power / carbon issues worldwide
  - Is there interest in a HEP-CCE project in that space?
  - Still some time to recommend best practices before big purchases for HL-LHC

Summary

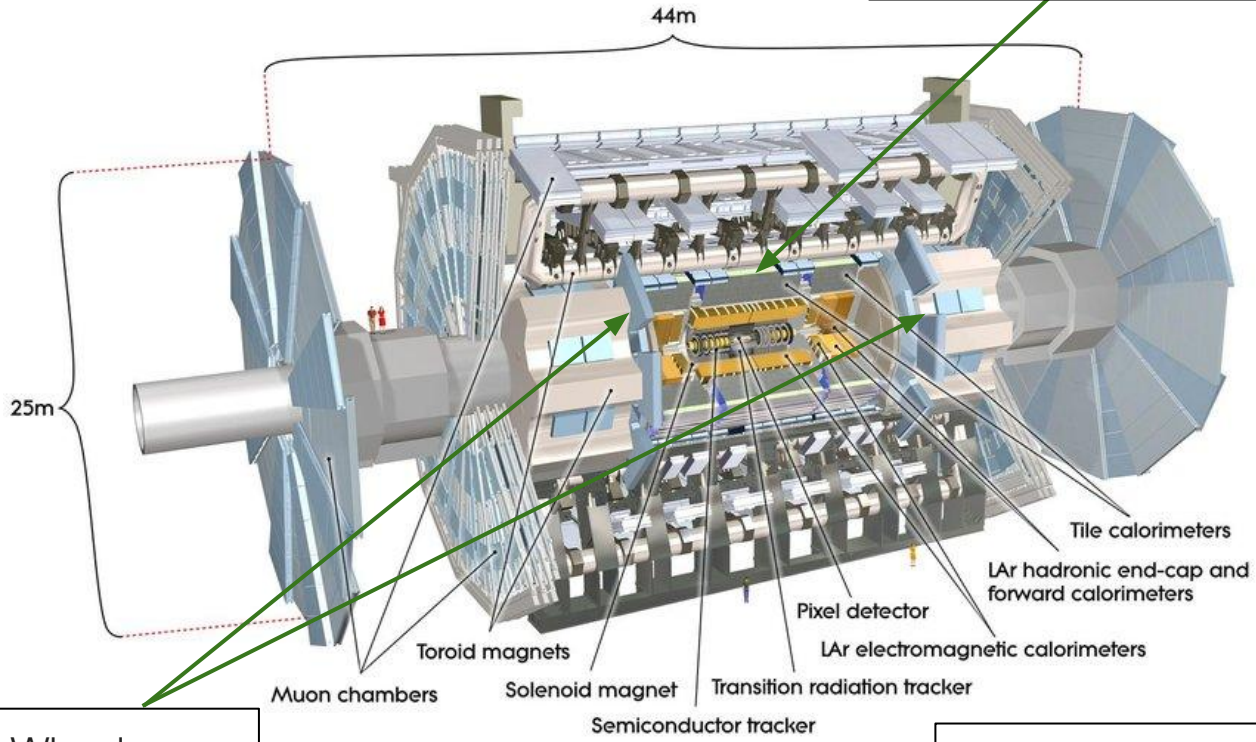
- ATLAS is facing interesting, difficult, but solvable software and computing challenges for the HL-LHC
  - One of the biggest challenges not mentioned here is supporting and retaining skilled developers — your help is always welcome!
- Now is a great time for R&D, demonstrators, prototypes, and pilot projects!
  - From experience we know how long and painful integration in our frameworks and full physics validation are: we should take this into consideration to manage our expectations!
- Focusing our efforts on common, shared objectives is paramount
  - The *way* in which we work can make the difference between success and failure!
  - Fragmented efforts are lethal — and ineffective
- This was a “quick” overview of some of the challenges
  - Much more in our HL-LHC Roadmap, and we are happy to discuss further with anyone interested in contributing!

Extras

The ATLAS detector now

Phase-I upgrade

New LAr L1 trigger



New Small Wheels

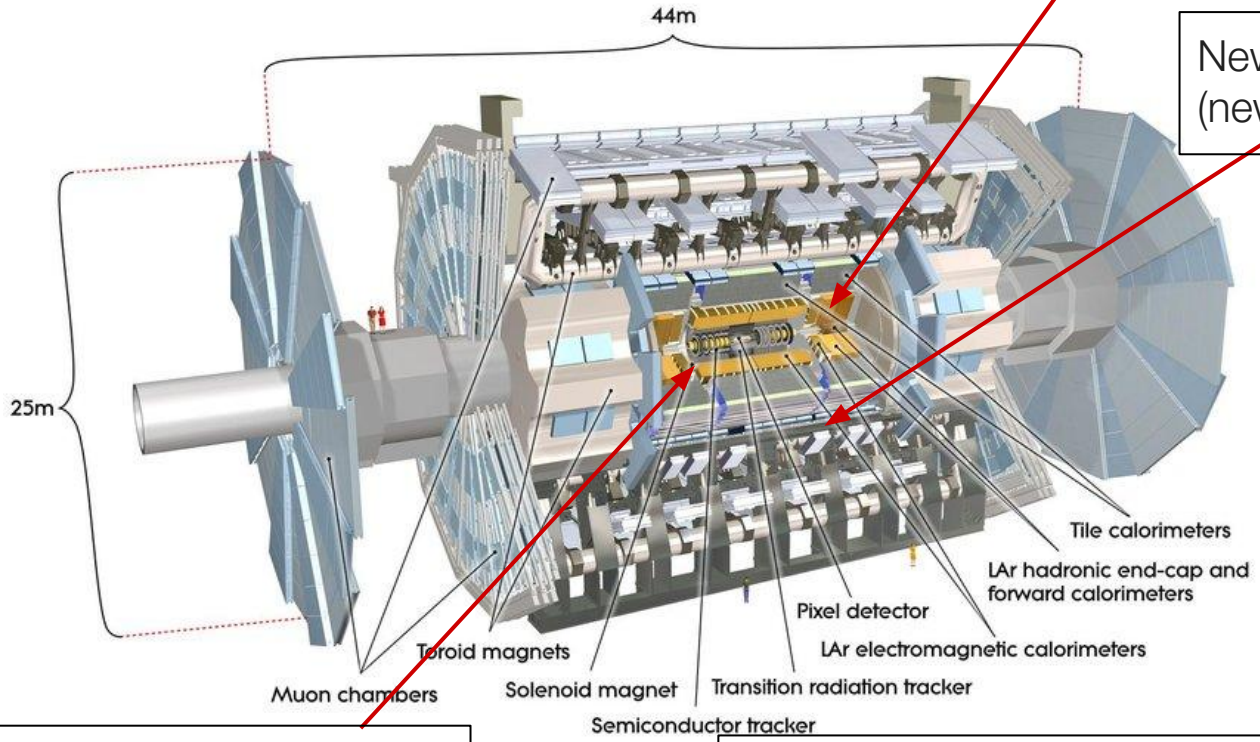
New SW: Athena Multi-Threaded

The ATLAS detector in HL-LHC

Phase-II upgrade

High Granularity Timing Detector (HGTD)

New muon chambers (new RPC and sMDT)



New Inner Tracking detector (ITk)

Upgraded Trigger and Data Acquisition system

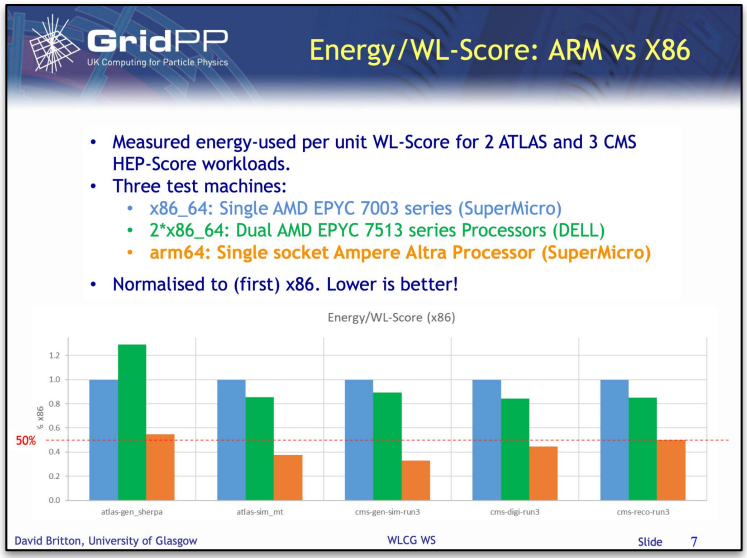
The ATLAS Collaboration

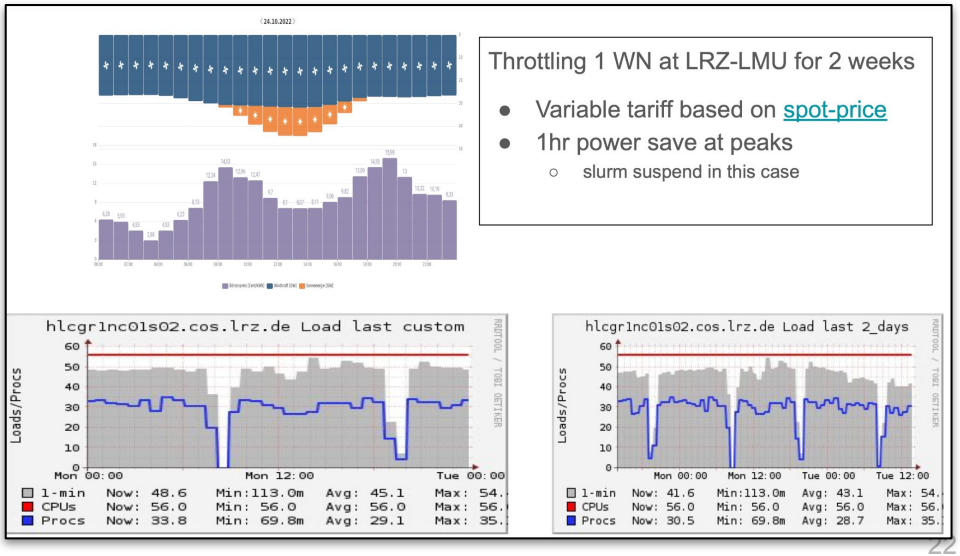
Some caveats: lots of “business as usual”

- We have to keep on running the experiment while we are planning for major upgrades
  - As Karl Jakobs said, “We are building a new ATLAS while we are running ATLAS”
  - Failing is not an option
- Lot of efforts that need to go into non-R&D work (or at the boundaries)
  - Maintaining our current software
  - Updates to database infrastructure
  - Improvements in metadata handling
  - Upgrade geometry and digitization
  - SW performance improvements
  - Re-tuning of Fast and G4 simulation for Run 4 new detector, and re-tuning reco
  - Distributed computing: lots of fundamental stuff, building blocks (tokens, OS, network, storage technologies)
- R&D projects are on top of this:
  - Balance (between R&D and “business as usual”) is key
  - And we need to have a strong focus on “impact”

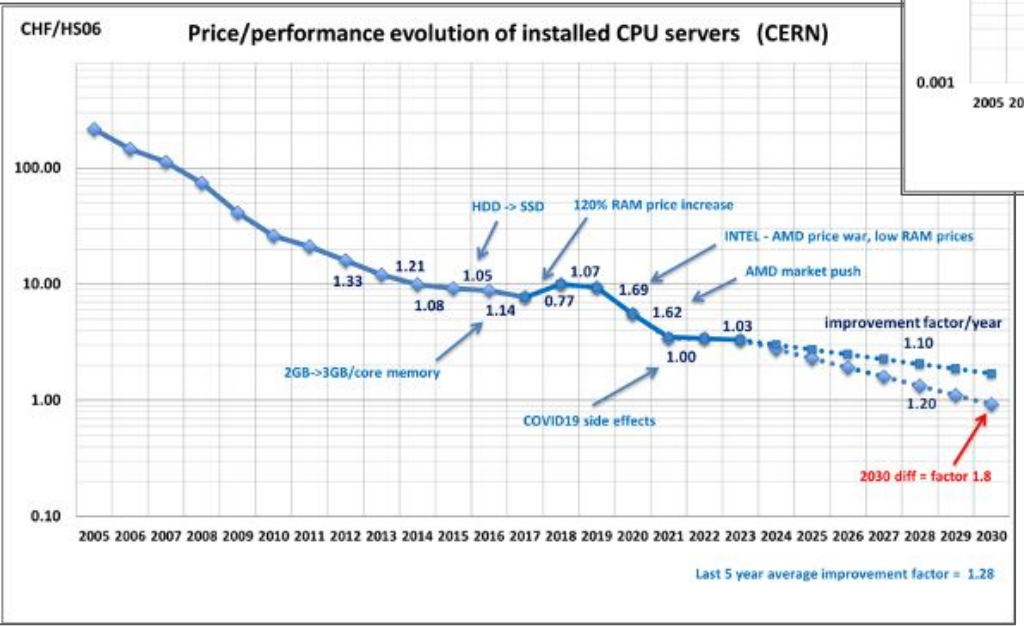
Power and Energy

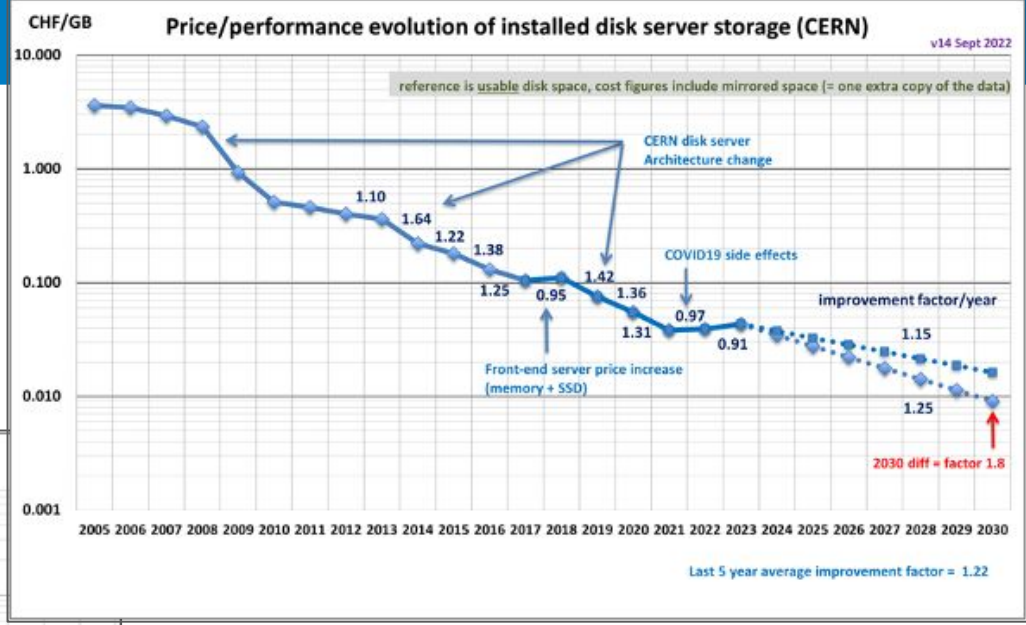
- Interesting & useful discussion of energy consumption during WLCG Workshop
  - Discussions dominated up to now by ATLAS members + sites
- Happy to engage further on energy consumption, power, C02, etc
- Various positive steps in terms of energy reduction
  - ATLAS full (Geant4) simulation fully validated on ARM (and now working on evgen+reco)
  - Clearly defined list of priorities for sites in case of power-shedding needs (switching off disk should be the last resort)





CERN Hardware cost





Challenges: Tracking

- Charged particle tracking is a great example of (constantly) hot R&D
  - It is slow in the HL-LHC – it's always been one of the heavier parts of the reconstruction
- Lots of work was done in preparation for Run 3
  - The vast majority of the speed up came from physics-driven optimizations of the algorithms
  - We should be investing in some of these optimizations now to avoid wasted code optimizations
  - N.B. physicists will find ways to use the CPU again!!