Computing for ATLAS

Torre Wenaus

Nuclear and Particle Physics Software (NPPS) Group Leader, BNL Physics Department

U.S. ATLAS HL-LHC Computing Co-Manager

July 28, 2023

Computational HEP Traineeship Summer School, Princeton University

Brookhaven National Laboratory

U.S. DEPARTMENT OF ENERGY

Computing for ATLAS - This talk

"Computing for ATLAS" is a broad topic but the "Big Picture R&D" session topic focuses things nicely, with plenty (too much) to talk about

I'll give you a look forward: how is ATLAS preparing its computing for HL-LHC

ATLAS has recently organized its HL-LHC R&D activities in preparation for setting its plan in the next two years (a coming Technical Design Report)

I will give you a selective tour of our R&D projects

Drawing heavily on talks from our ATLAS teams, and CHEP23

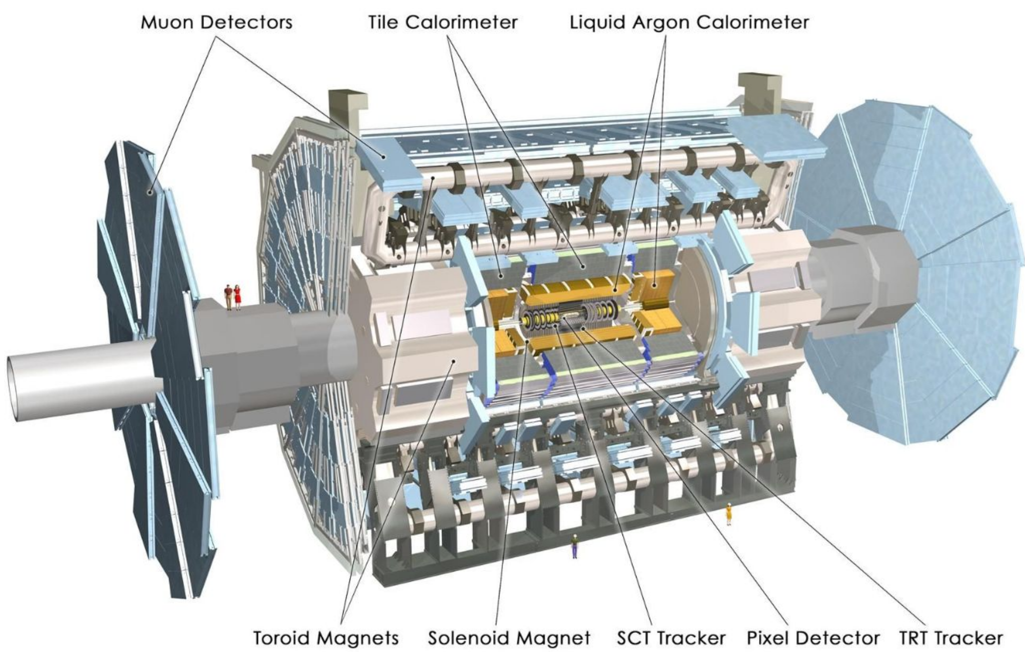
Join ATLAS in the next few years and help us with these!

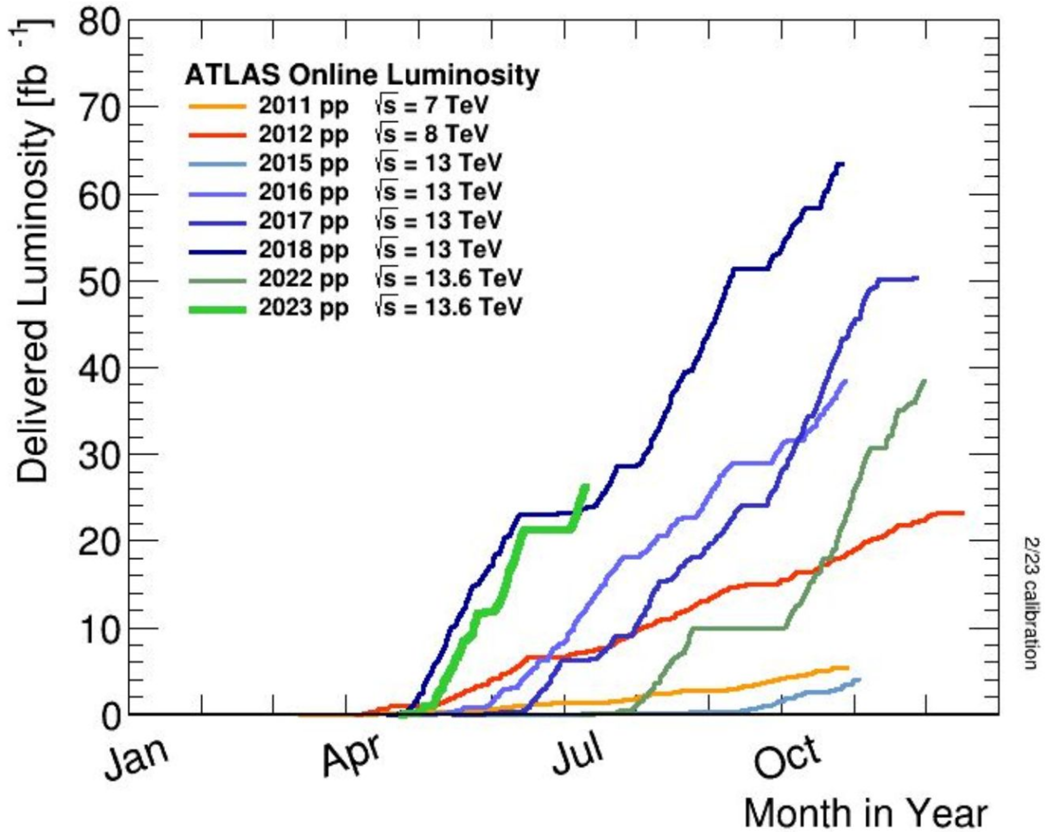
This is an informal & interactive event please ask questions!!!

ATLAS to date

2023 ATLAS Detector

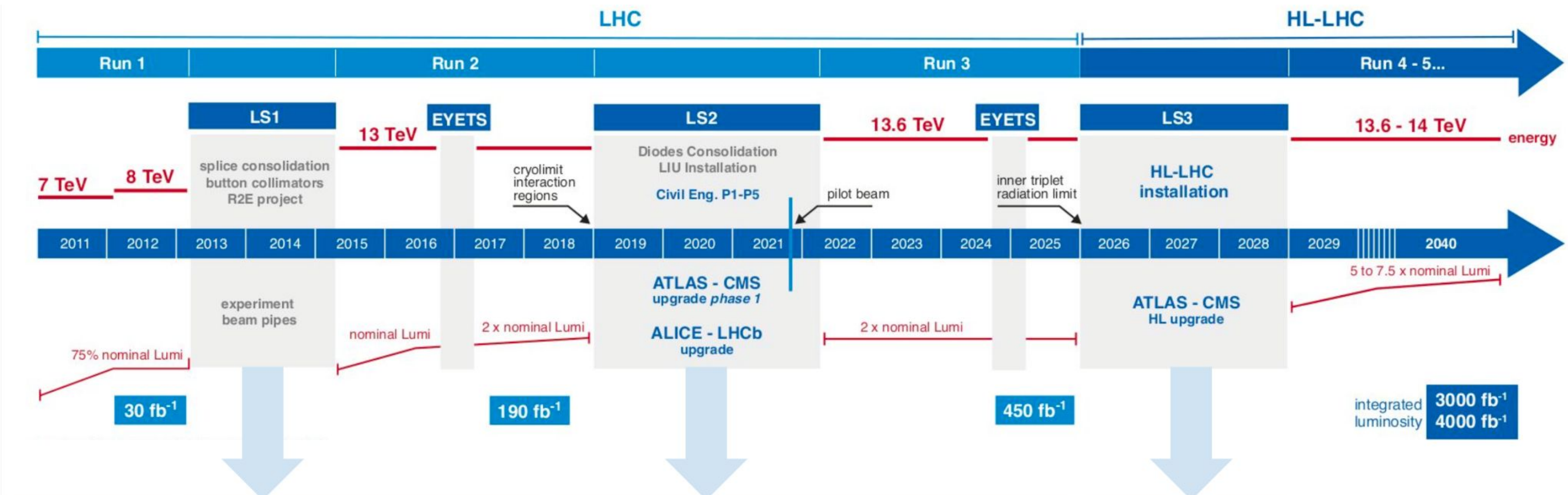
- Inner Detector: Si and TRT tracking
- Calorimeters: LAr, Tile
- Muons: precision tracking & trigger
- Forward: luminosity & diffractive physics
- Magnets: 2T solenoid (tracking), toroid (muon)





Run	Dates	CM Energy	Recorded Lumi
1	2011-12	7,8 TeV	26 fb^{-1}
2	2015-18	13 TeV	149 fb^{-1}
3	2022-...	13.6 TeV	55 fb^{-1} (so far)

LHC and ATLAS evolution



ATLAS Upgrades

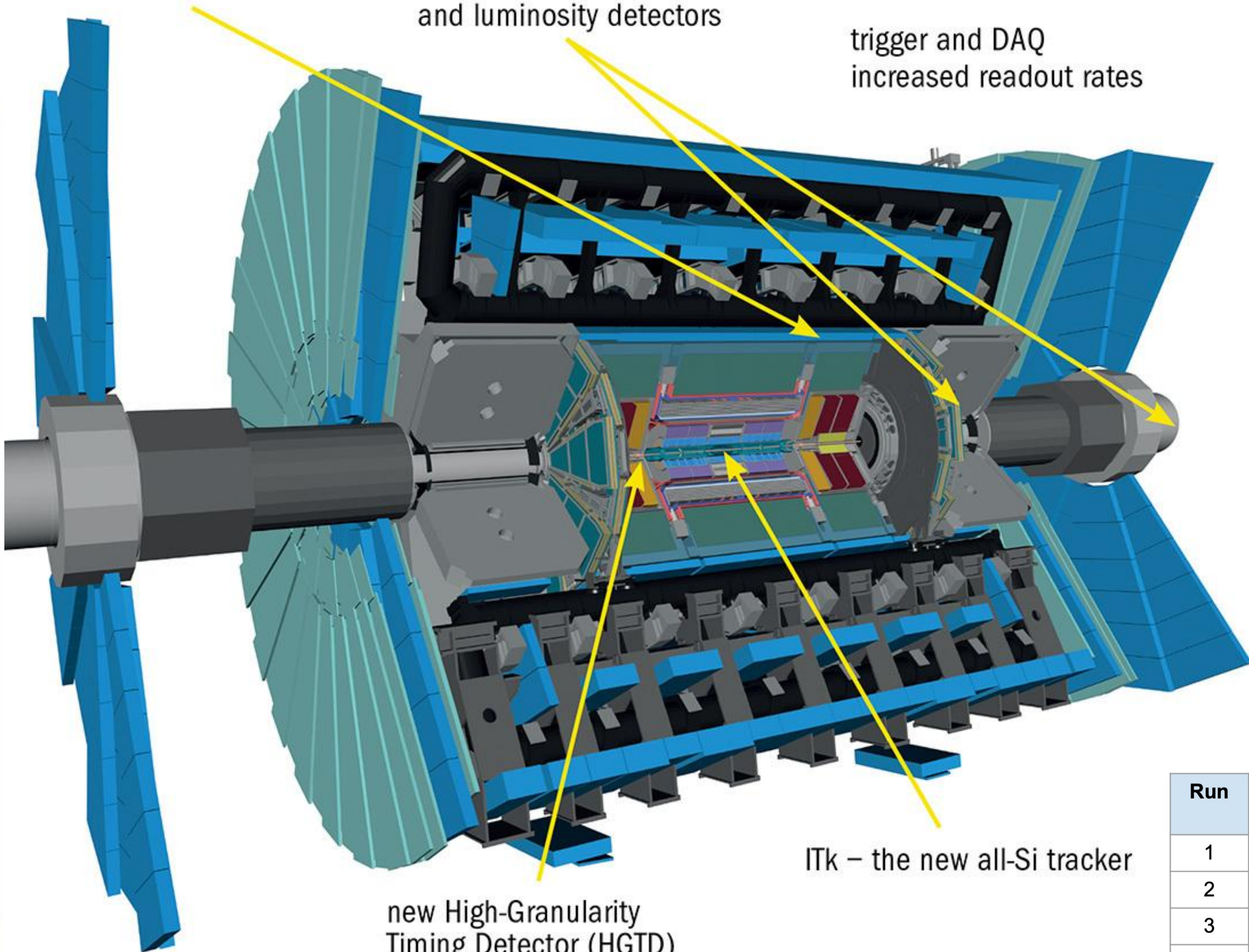
- New inner tracking layer (IBL)
- Complete Forward muon system: detectors + readout
- Trigger: topology at L1 + streamlined dataflow

- Trigger: new feature extractors at L1
- Calorimeter electronics for trigger
- New forward muon system (NSW)
- More performant/uniform readout system (FELIX)

HL-LHC upgrades

Article: A new ATLAS for the high-luminosity era

Major ATLAS upgrades for HL-LHC



An order of magnitude more data starting from 2029

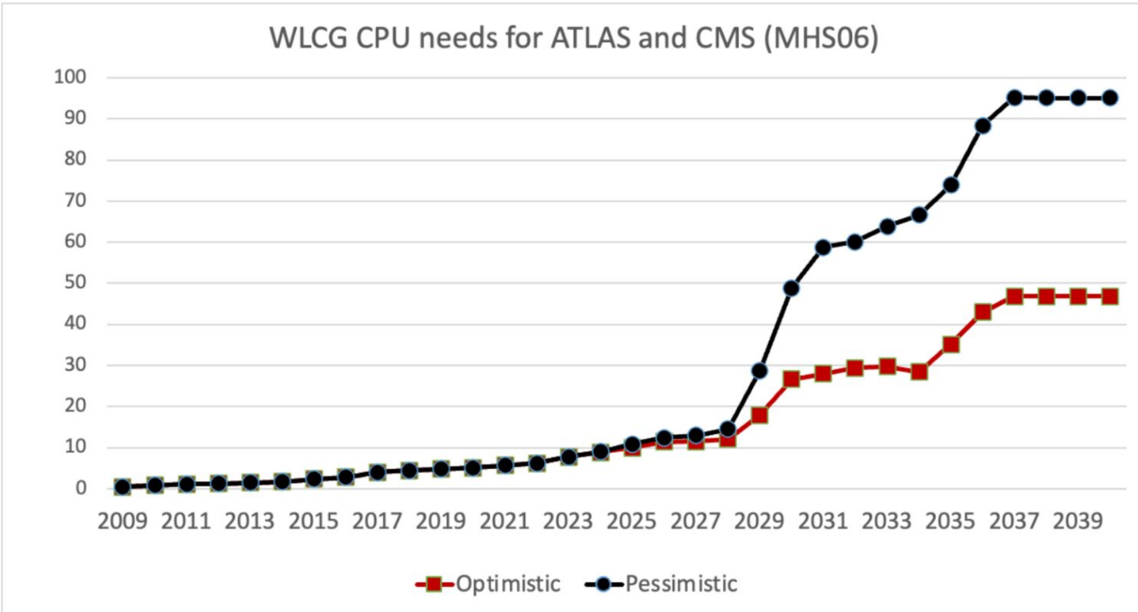
Precision tests of the properties of the Higgs boson

About 170 million Higgs bosons and 120,000 Higgs-boson pairs delivered to ATLAS over a period of about 10 years

Improved sensitivity to many new-physics scenarios

Run	Years	CM Energy (TeV)	Bunch Spacing (ns)	Peak Lumi ($x10^{34}$ cm^{-2} s^{-1})	Pileup (pp / BX)	Total Int. Lumi (fb^{-1})
1	2010-12	7-8	50	0.8	43	23
2	2015-18	13	25	1.9	55	150
3	2022-25	13.6	25	2-3	55-80	450
4…	2026…	13.6-14	25	5-7.5	140-200	3-4,000

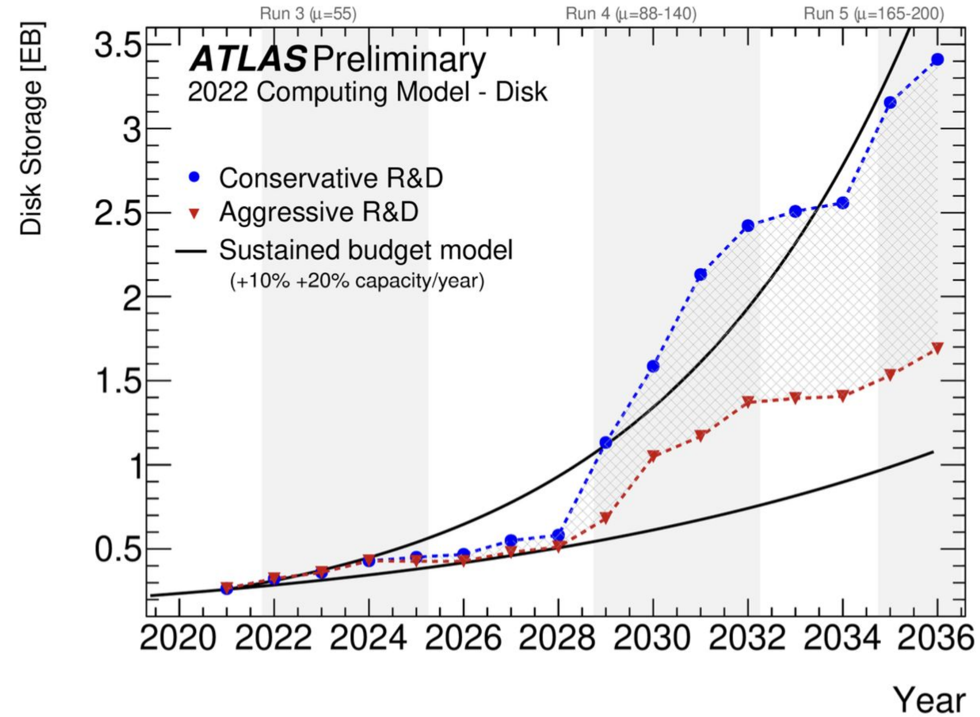
Computing needs for HL-LHC



ATLAS (conservative R&D) + CMS (no R&D improvement) = **PESSIMISTIC** scenario

ATLAS (aggressive R&D) + CMS (weighted probable scenario) = **OPTIMISTIC** scenario

Simone Campana, CHEP 2023



- If we do minimal R&D (blue curve), we cannot afford HL-LHC computing with 10x more events and higher event complexity
- Much the same for storage (most expensive component)
- To afford the computing we need, substantial R&D is required (red curves)
- ATLAS's R&D program (a subset) is the topic of this talk

Selected ATLAS R&D topics

The **processing** challenge

- Fast simulation
- Parallelism
- The ARM platform

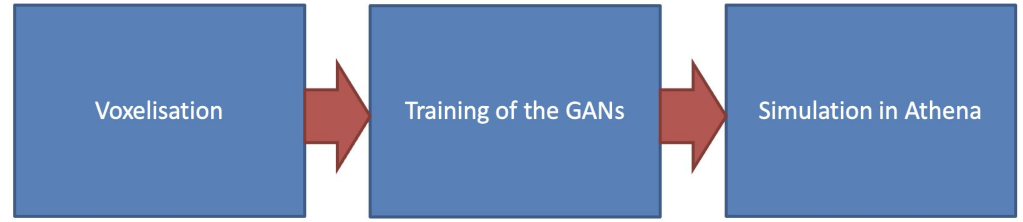
The **data storage and management** challenge

- Data carousel & data on demand
- PHYSLITE: compact efficient data

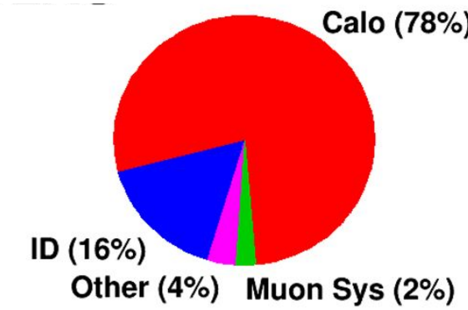
The **complexity and scaling** challenge

- Commercial clouds for analysis
- Large scale complex workflows for AI/ML and analysis
- CREST: next generation conditions database

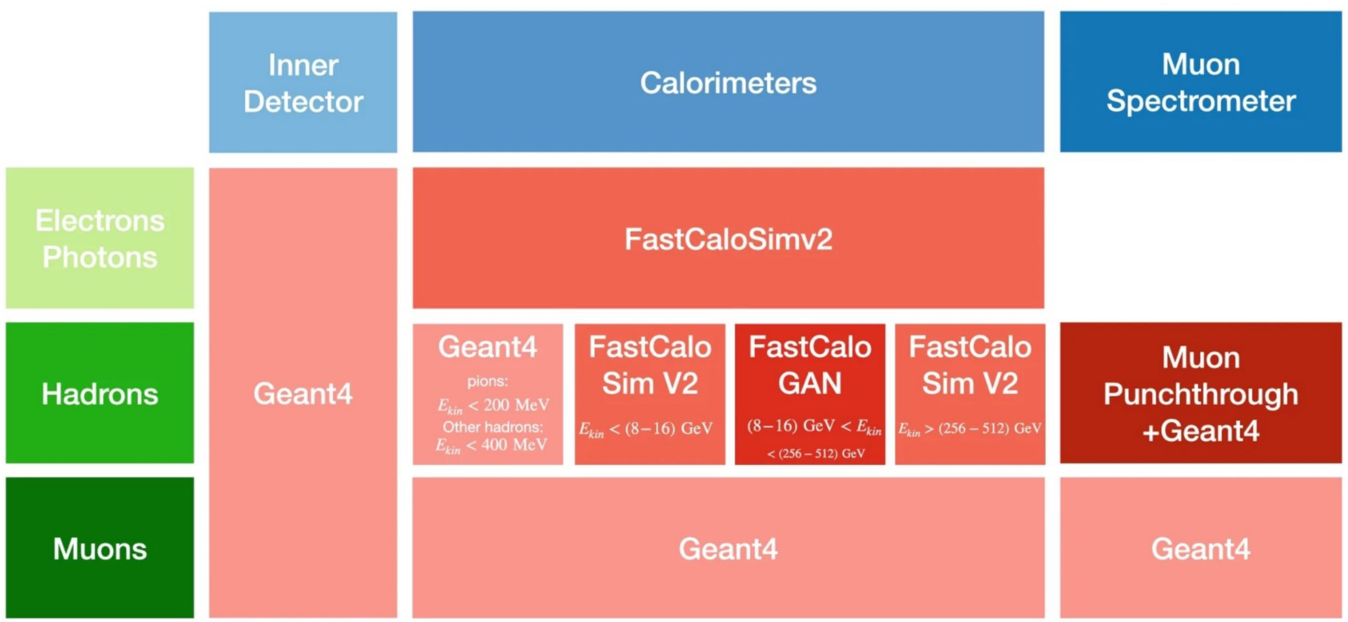
Fast Simulation: AtlFast3

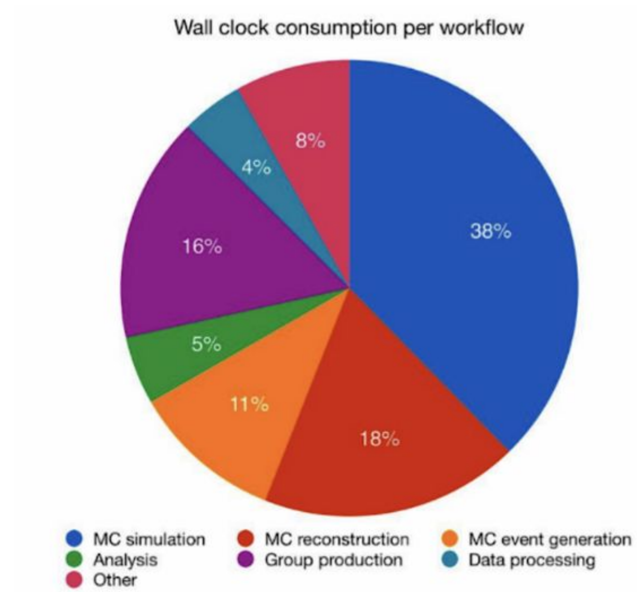


- Simulation is the biggest CPU resource consumer
- Fully detailed Geant4 detector simulation very costly, particularly calorimetry
- The ATLAS fast simulation in its recently released new & improved version AtlFast3 (AF3) is an important answer to this
- ATLAS' plan is to run 90% of the simulation using fast simu, reaching 75% in Run 3
- AF3 includes an ML component FastCaloGAN for select particle types/energies, improving on jet modeling over V2
  - electrons, photons, charged pions, 100 eta slices, 15 energy points
- Ongoing: better voxelization, new GAN settings, more use cases



Subdetector CPU fraction for 50 ttbar events
MC16 Candidate Release



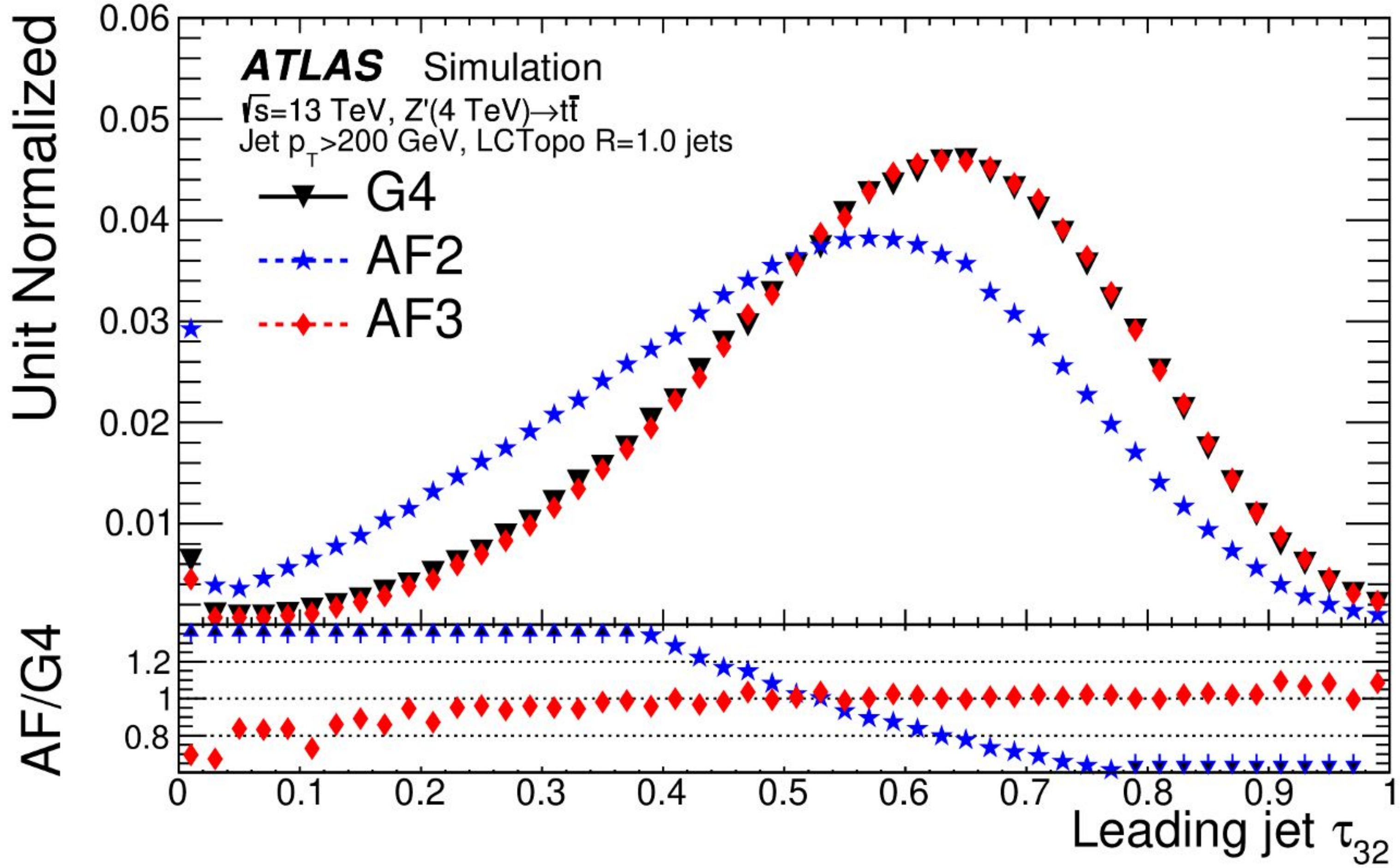


Brookhaven National Laboratory

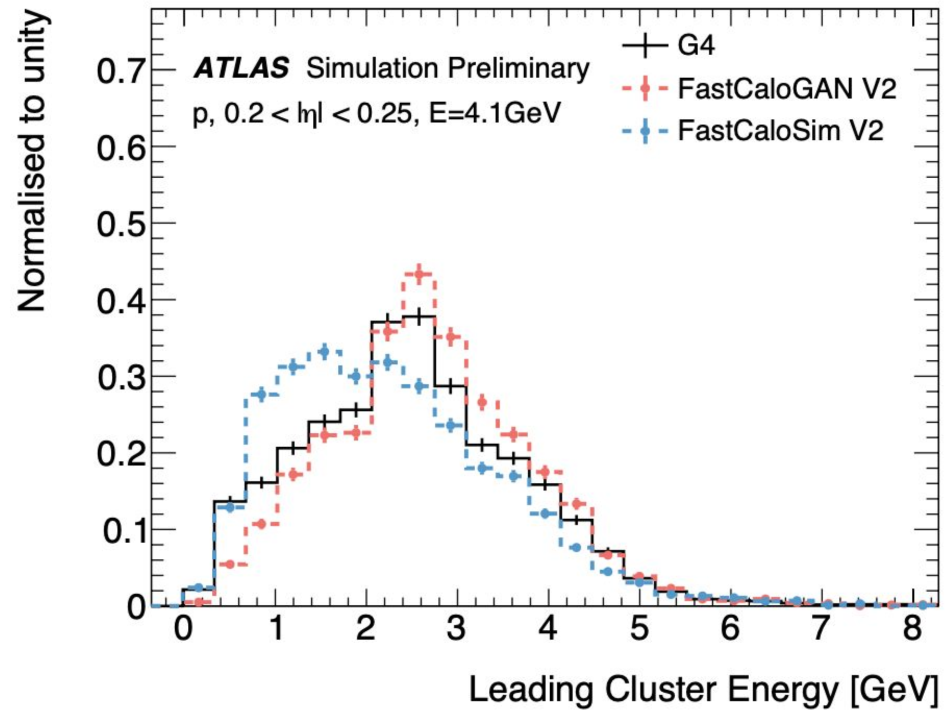
Marilena Bandieramonte et al, FastCaloGAN: A FAST SIMULATION OF THE ATLAS CALORIMETER WITH GANs, CHEP 2023

AtlFast3 performance

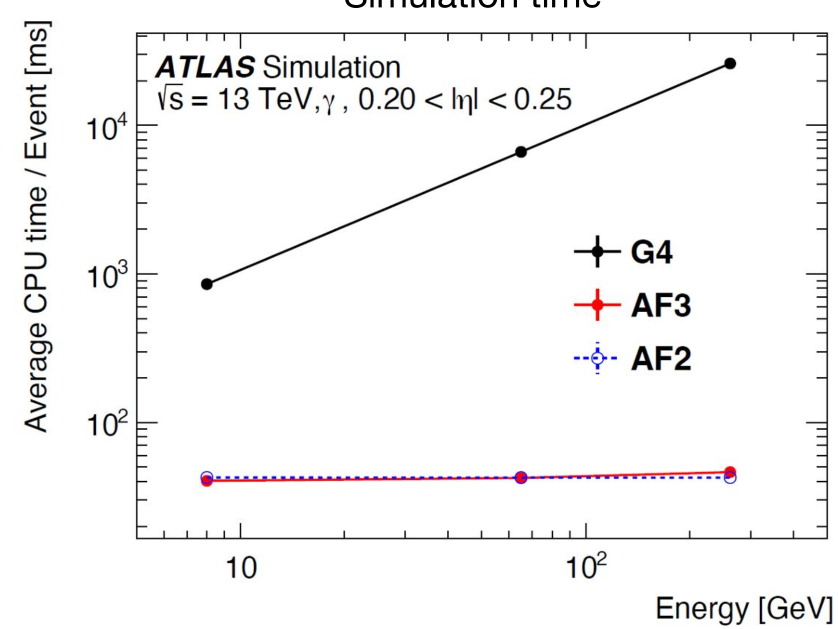
AF3 performance - boosted jets



GAN based calo simu in AF3 improves on V2



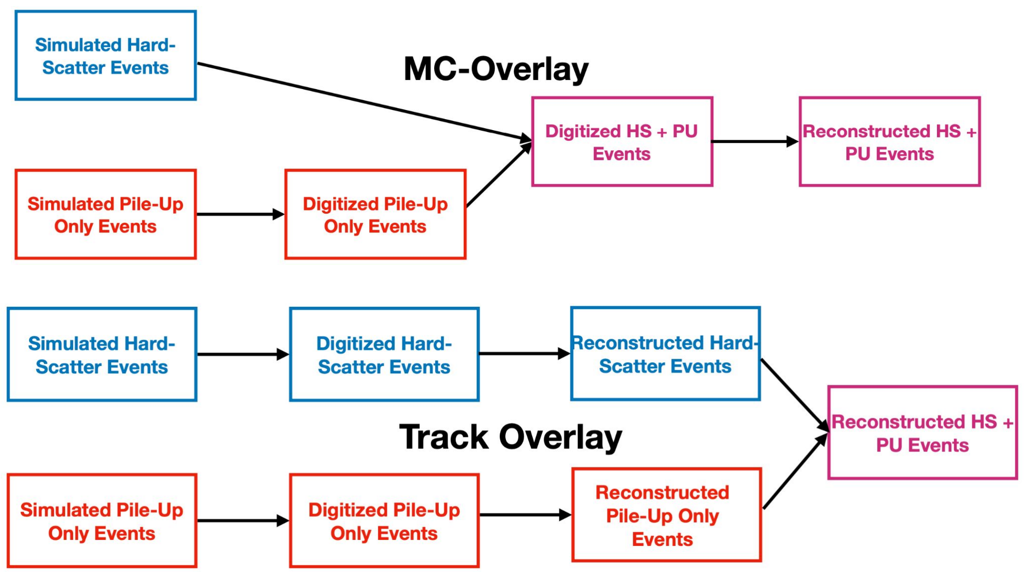
Simulation time

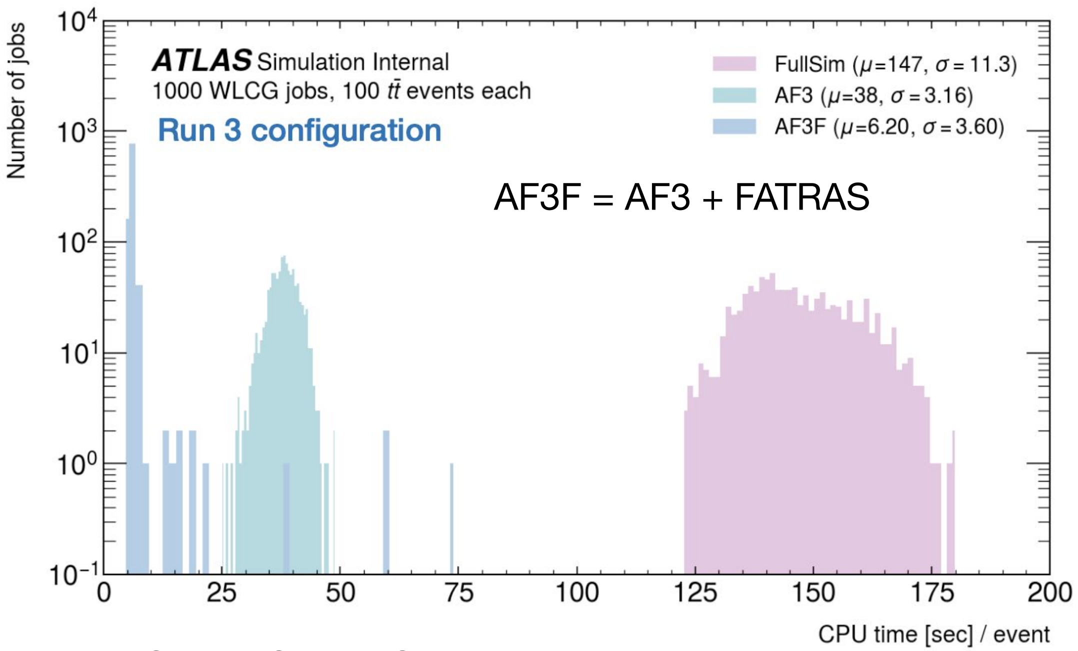


Michele Faucci Gianelli, AltFast3 - The new fast simulation in ATLAS, CHEP 2023

FastChain

- AtlFast3 brings us part way to the HL-LHC fast simulation we want
- FastChain brings us further: replace more CPU-intensive parts of the production chain
- FATRAS: reduce simulation time in the inner tracker
  - Simplified tracker geometry, migrating to new ACTS tracking software
  - Not used in production to date. Goal: Run 4 ATLAS implemented and physics validated in mid 2026
- Track overlay: reduce pileup reco time by overlaying reconstructed pileup tracks
  - More than half of event reco time spent on inner detector tracking, track overlay can reduce significantly
  - ML algorithm chooses when to use track overlay vs unreconstructed; negligible degradation observed
- Single workflow from event generator to analysis data products without intermediate files: substantial tape/disk savings



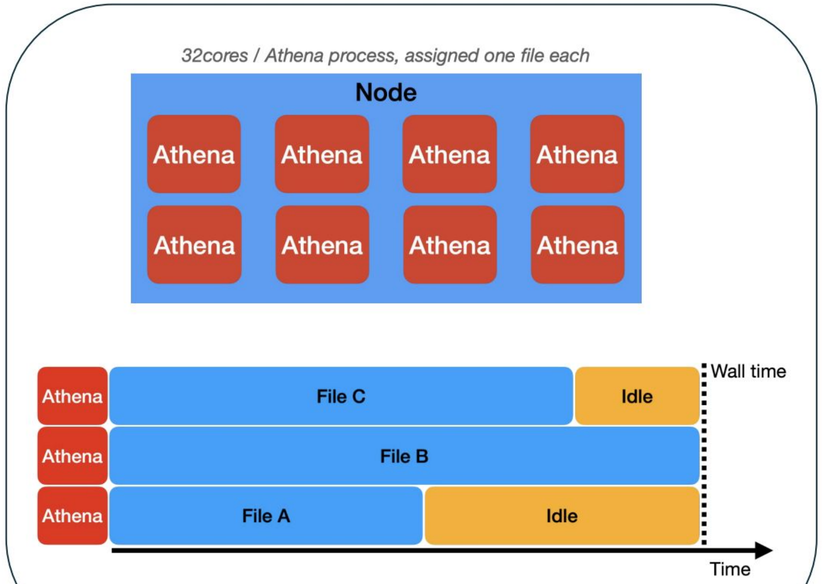


Will Leight, Faster simulated track reconstruction in the ATLAS Fast Chain, CHEP 2023
Martina Javurkova, Fast Chain Workflows in ATLAS, US ATLAS technical meeting

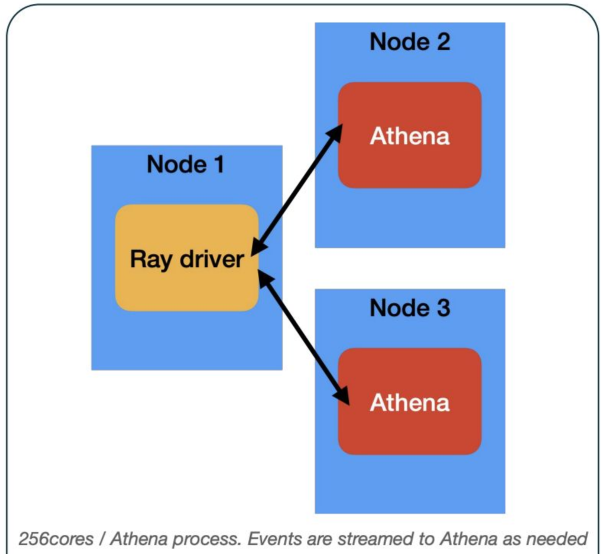
Brookhaven National Laboratory

Parallelism

- ATLAS has a long history of leveraging opportunistic resources including HPCs
- Developed the Event Service for efficient fine grained processing on highly parallel machines (such as DOE supercomputers)
- Next-gen HPC-specific implementation, Raythena, has just been developed
- However, ATLAS offline software is not GPU-ready
- Achieving parallelized accelerator-capable offline software performing efficiently across the GPU resources available to ATLAS has been and continues to be an enormous, costly R&D effort
- And once we have GPU-capable software, achieving high GPU utilization is a further major challenge and development effort
- Instead of getting into all that, I'm going to talk about a promising new platform…

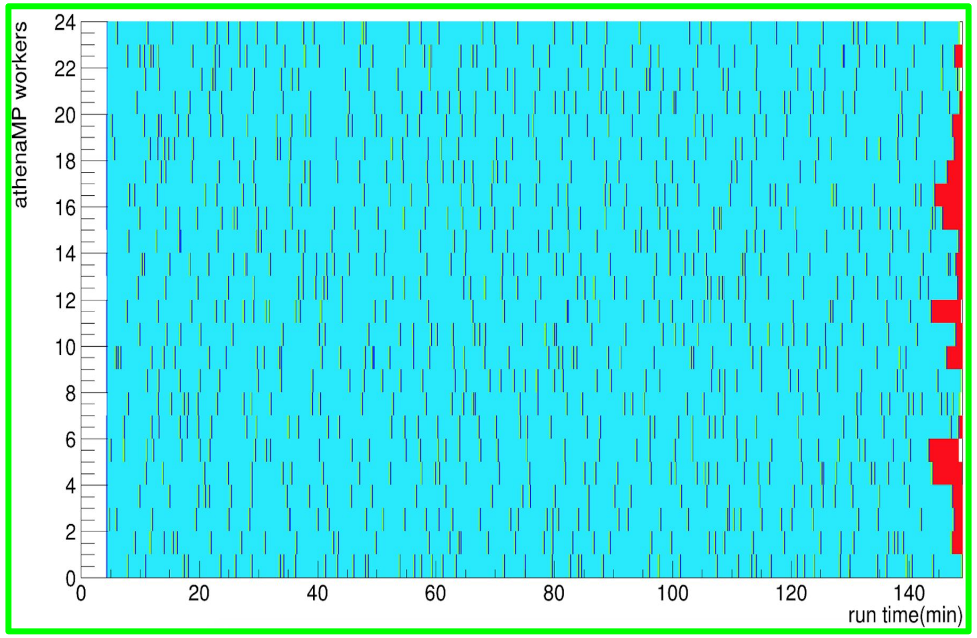


In Production: File-level parallelism



Raythena: Event-level parallelism

No wasted core time



Brookhaven National Laboratory

Miha Muskinja, Raythena: a vertically integrated scheduler for ATLAS applications on heterogeneous distributed resources, CHEP 2023

The ARM Platform

What is ARM?

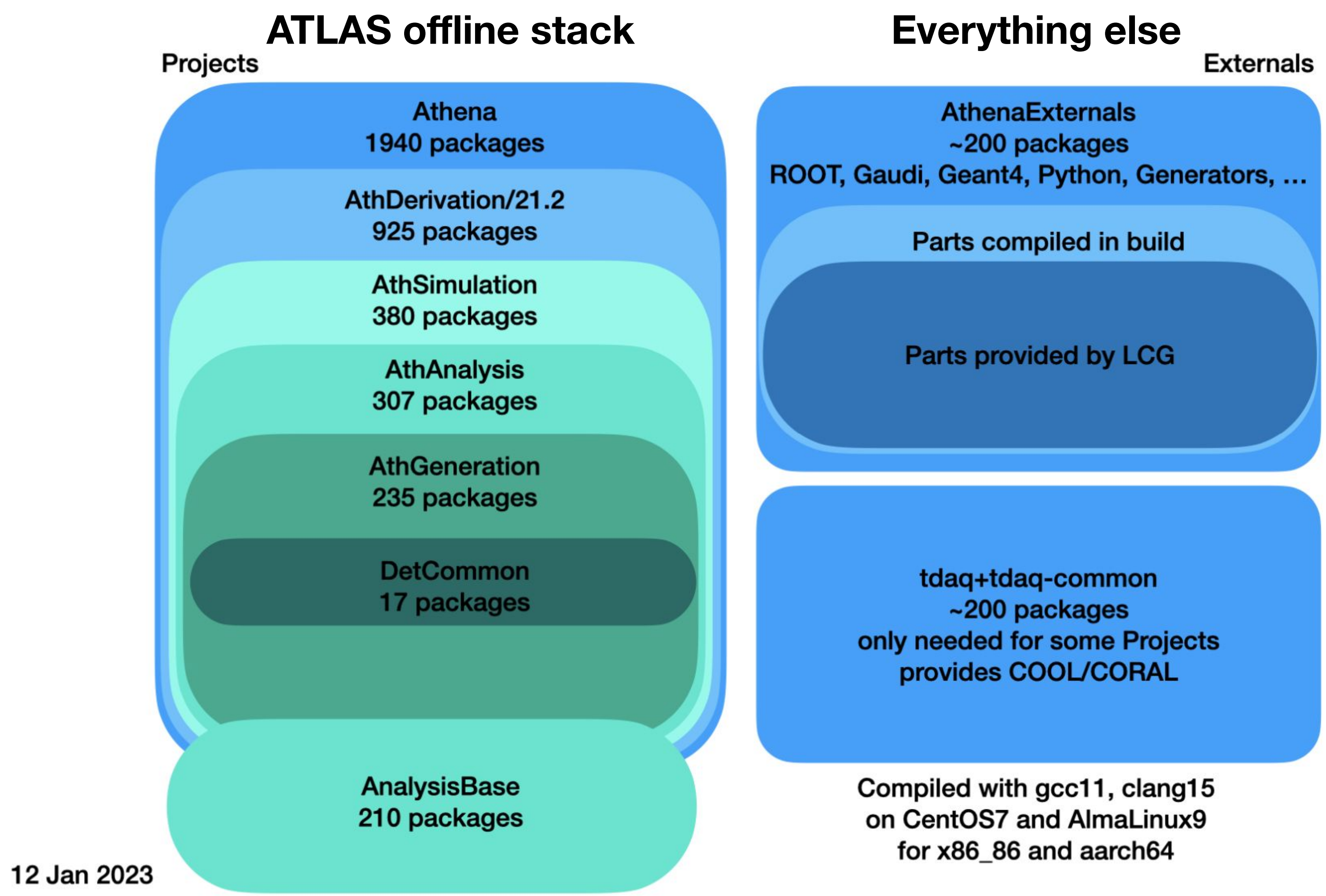
- Family of reduced instruction set computer (RISC) instruction set architectures
    - Arm Ltd. develops the architectures and licenses them to other companies
    - Cellphones, Macs, supercomputers (Fugaku in Japan, #2 on top 500 as of Nov 2022)
    - Amazon AWS Graviton processors
    - Google Cloud (Ampere Altra/Neoverse-N1)
- **Low costs, minimal power consumption, lower heat generation than competitors**
- **Much simpler to port our software than reengineering for GPUs**
- **More HPCs are coming with ARM architecture (Europe for sure, also US?)**
- **Grid sites starting to be interested in ARM when experiment software is validated**

ATLAS established an ARM R&D program drawing on our commercial cloud R&D

- Full simu and reco software stacks, and their dependencies, ported
    - With the help of external dependency maintainers
- ARM fully integrated in ATLAS PanDA/Rucio based production system
- Geant4 and reconstruction physics validation passed using AWS resources

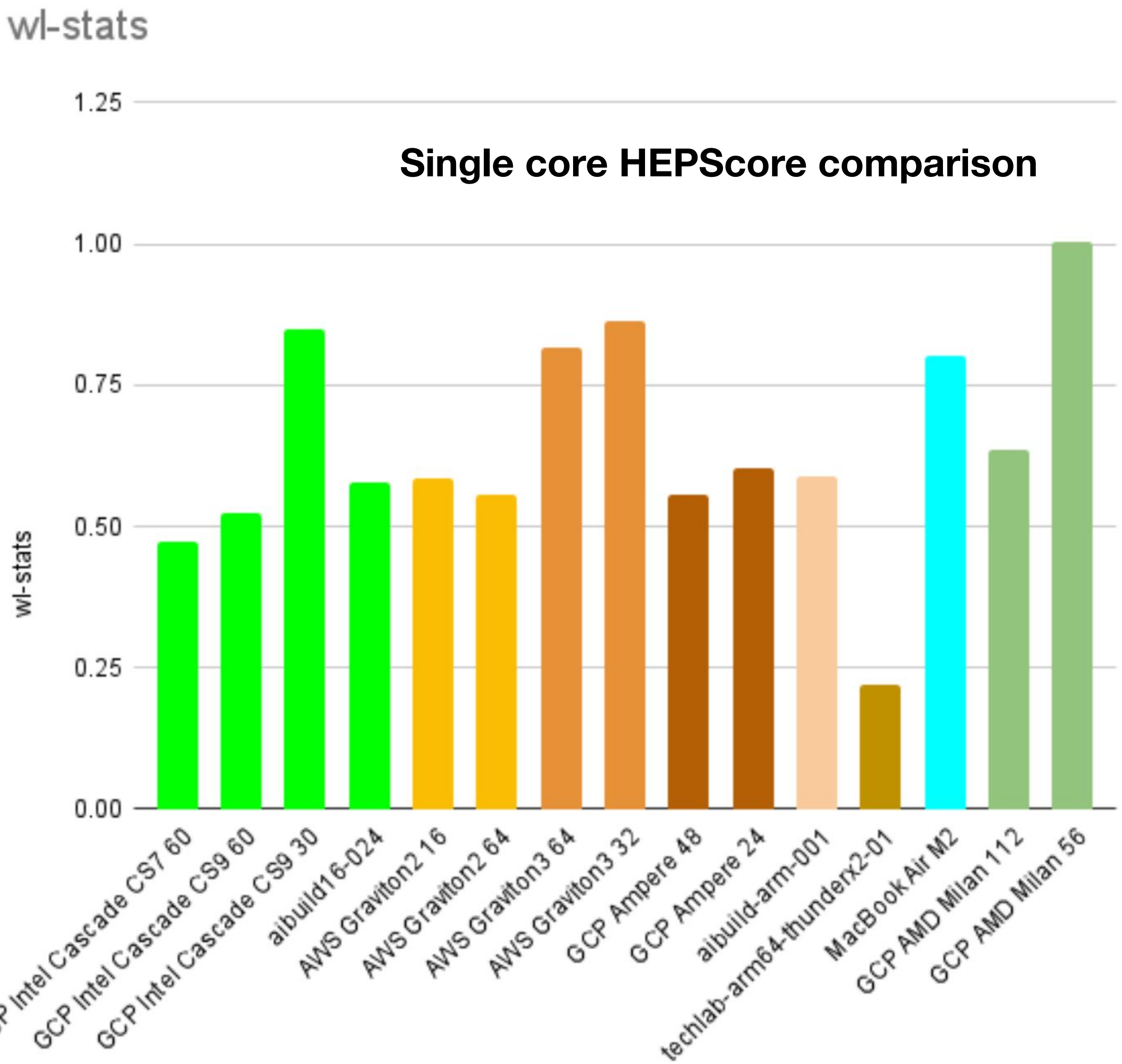
Johannes Elmsheuser, THE ATLAS EXPERIMENT SOFTWARE ON ARM, CHEP 2023
Dave Britton, ARM Update, CHEP 2023 (experience with ARM in UK)

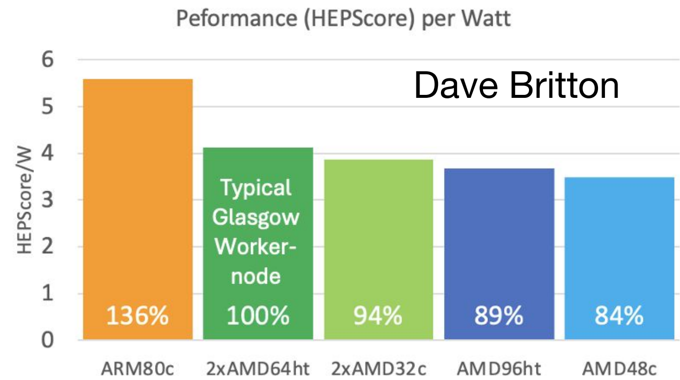
The ARM Port

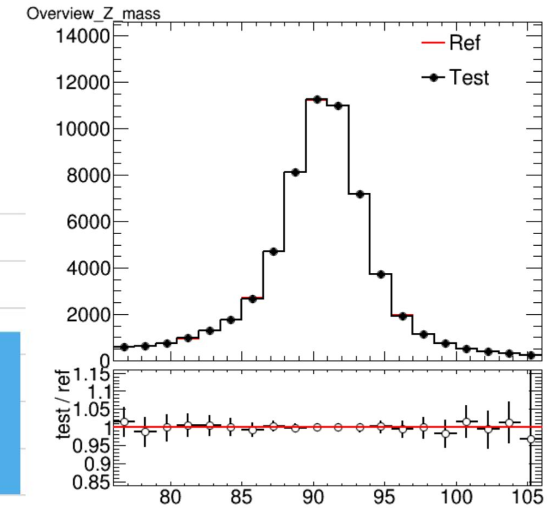


- 3 major blocks build by different teams:
  - Projects: ATLAS offline software
  - Externals: CERN EP/SFT trough LCG layers and ATLAS offline software
  - TDAQ: ATLAS trigger/DAQ

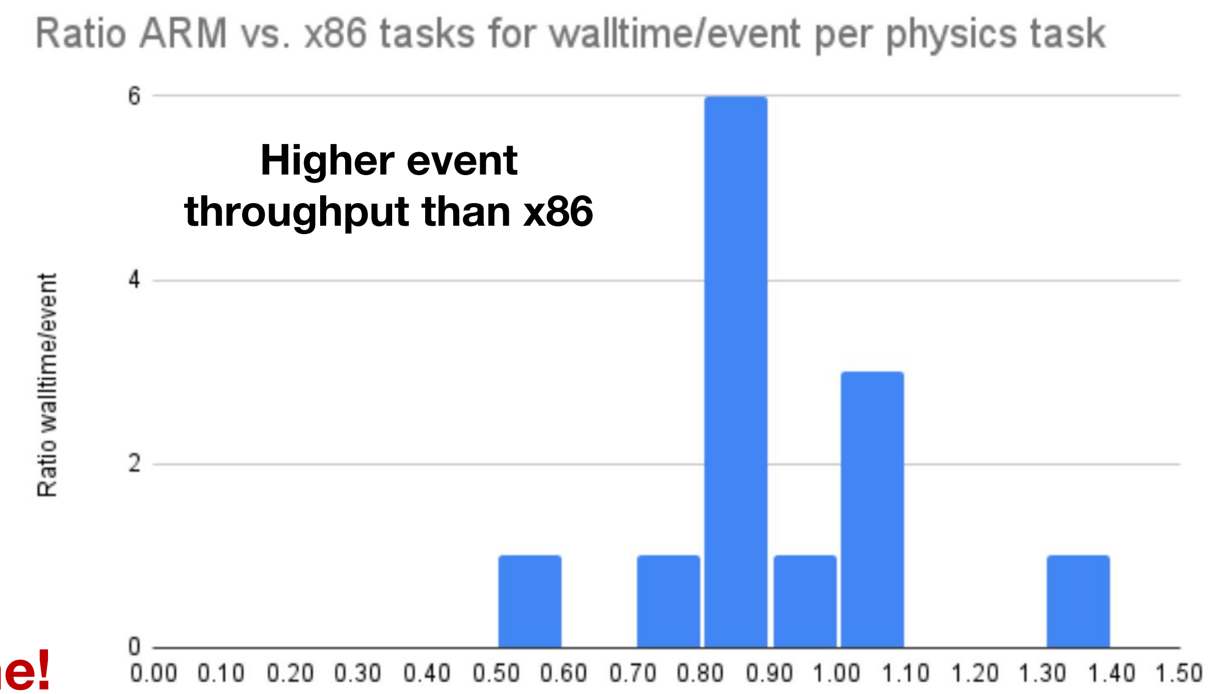
ARM validation and performance







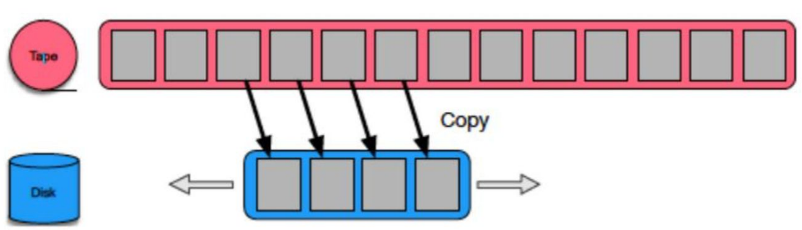
Reconstructed MC $Z \rightarrow \mu\mu$ invariant mass



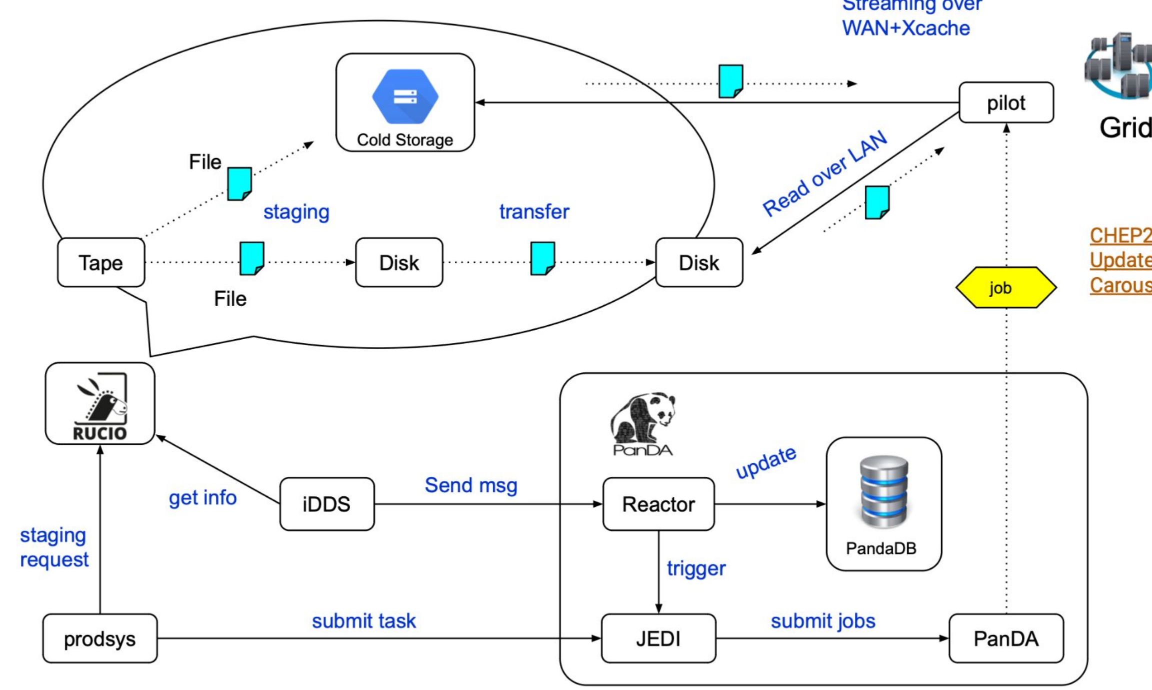
ARM on HPCs and grid: Build 'em and we will come!

Data Carousel

- The most expensive component of present and HL-LHC computing is disk storage - use tape when we can!
  - (Yes our tape costs go up but less fast)
- Data Carousel was one of our first HL-LHC directed R&D programs
- Fine grained disk-efficient tape staging
  - Process it as soon as it appears, delete
- In production for several years, runs all workflows, it cut our analysis data disk footprint in half
- Fruitful collaboration with Tier 1 computing centers to optimize tape systems to perform well



Data Carousel concept



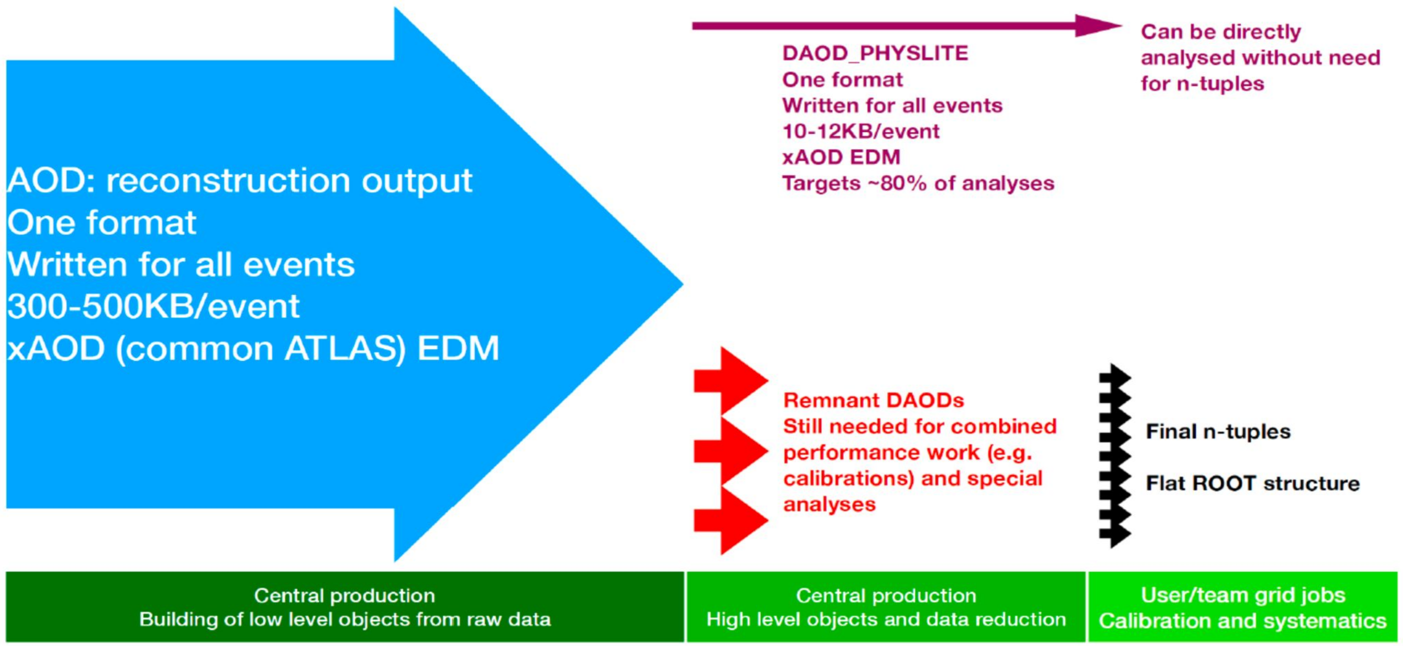
[CHEP2023 Talk: X. Zhao, et al. Updates to the ATLAS Data Carousel Project]

Now entering a Phase 2 round of R&D

- Smart writing: place data on tape intelligently to optimize retrieval
- Data on demand: delete disk replicas more quickly than presently
  - Only ~20% of explicitly requested disk resident data is accessed in the year after the request
  - Explore both regeneration on demand, and tape retrieval on demand
  - Complex regeneration workflow vs. extra load on tape system

PHYSLITE: Compact Efficient Analysis Data

- PHYSLITE is a new compact, common data format for ATLAS for Run 4
- Calibrated objects ready for fast analysis
- One version serving 80% of physics analysis in Run 4
- Run 4 goal: PHYSLITE directly analyzable without NTuple step
- Run 3 prototype in place to use and optimize
- But, other formats take up >50% of the DAOD storage. Reducing this a further project
- RNTuple will replace TTree
  - Implementation in place
  - Improved file size and read throughput
  - Performance optimization continues
- More R&D underway…
  - Event augmentation to ameliorate "one format for all"
  - Lossy compression
  - Columnar analysis



Jana Schaarschmidt, PHYSLITE - A new reduced common data format for ATLAS, CHEP 2023
Florine de Geus, Integration of RNTuple in ATLAS Athena, CHEP 2023

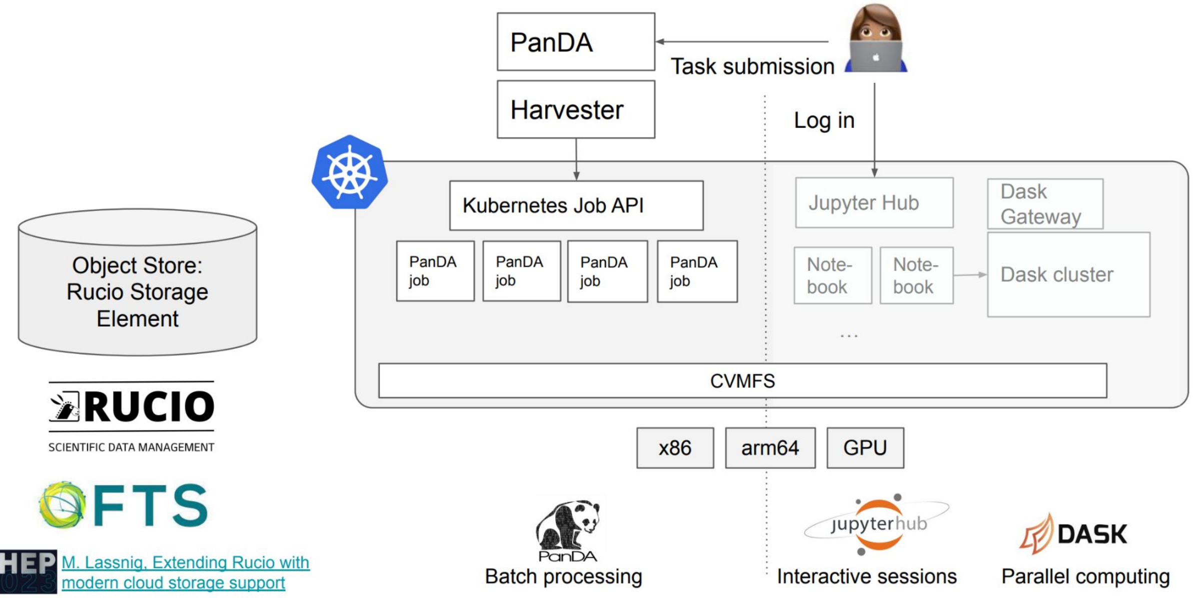
Commercial Clouds for Analysis

ATLAS has R&D collaborations with Google and Amazon AWS

- AWS valuable particularly for the ARM port
- Google collaboration and progressed and broadened through 2 phases over 3 years, phase 3 in discussion
- Testbeds for analysis tools and approaches (K8s, Dask, Jupyter interactivity)
- Investigate special resources and workflows (GPUs, ML, fitting)
- Particularly suited for bursty analysis usage
- Evaluating total cost of ownership as a Tier site
- Continues a long ATLAS tradition of using opportunistic and heterogeneous resources

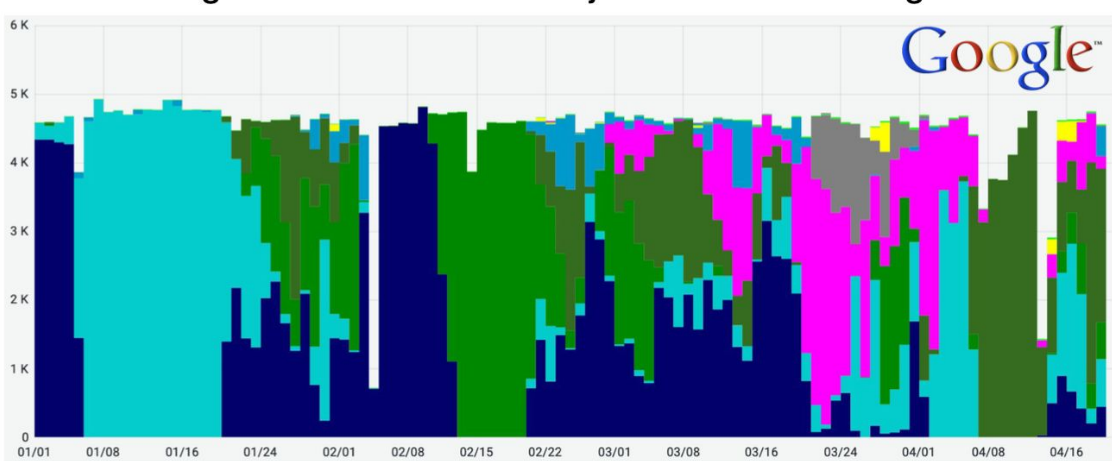
Cloud native integration

CHEP 2021 [Seamless integration of commercial Clouds with ATLAS Distributed Computing]

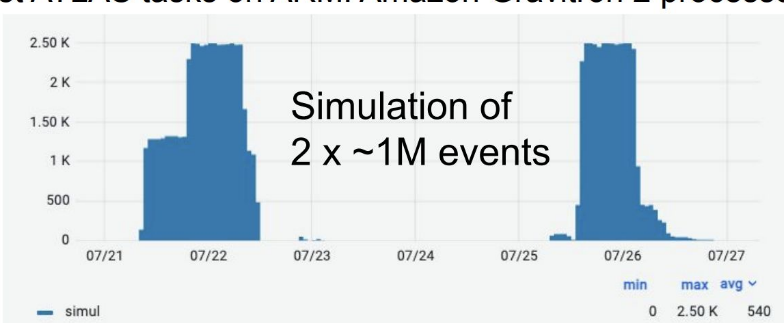


CHEP 2023 [M. Lassnig, Extending Rucio with modern cloud storage support]

Running various w/flows on 5k job slots at the Google site



First ATLAS tasks on ARM: Amazon Gravitron 2 processors



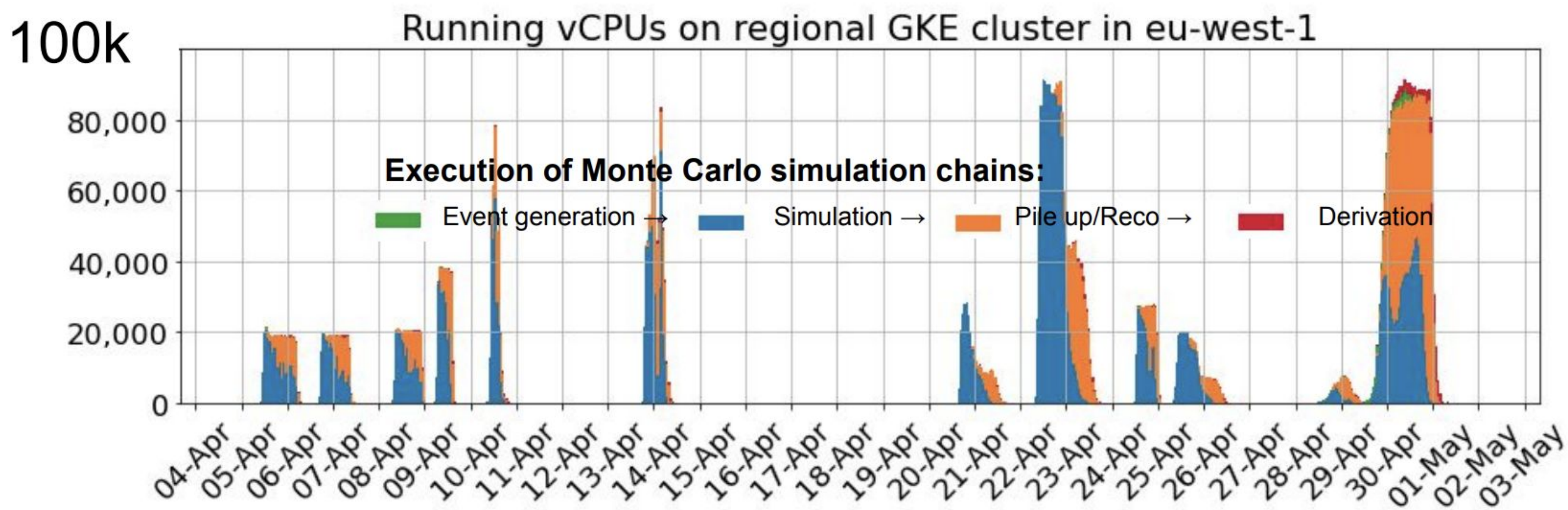
Fernando Barreiro Megino, Accelerating science: the usage of commercial clouds in ATLAS Distributed Computing, CHEP 2023

Brookhaven National Laboratory

Google cloud as a “bursty” resource

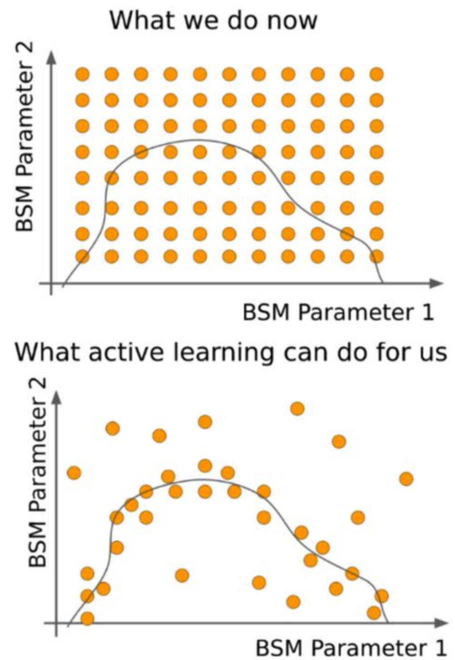
Google Cloud

- A cloud advantage: elasticity
- Enable users to spike quickly into large resources to minimize turnaround time
- If commercial clouds have a cost/benefit lowest hanging fruit, this will be it
- Cloud resources enabled R&D (GPU, mega-memory nodes, ARM, Dask) that would have been very difficult to carry out on the Grid or on-prem clusters



Large scale complex workflows for AI/ML and analysis

- ➢ Active Learning (AL) using iterative regression on limit setting to increase analysis efficiency
  - Traditional: A brute-force grid/random search
  - New: A search session with multiple iterations
    - The search space for the next iteration is defined based on the results of old iterations
- ➢ A single workflow with multiple loops
  - Various tasks on different resource/service providers
  - No human intervention in transitions from one provider to another
- ➢ AL-based analysis for mono-Hbb exclusion limit, dark-Z search, and heavy higgs search in ATLAS



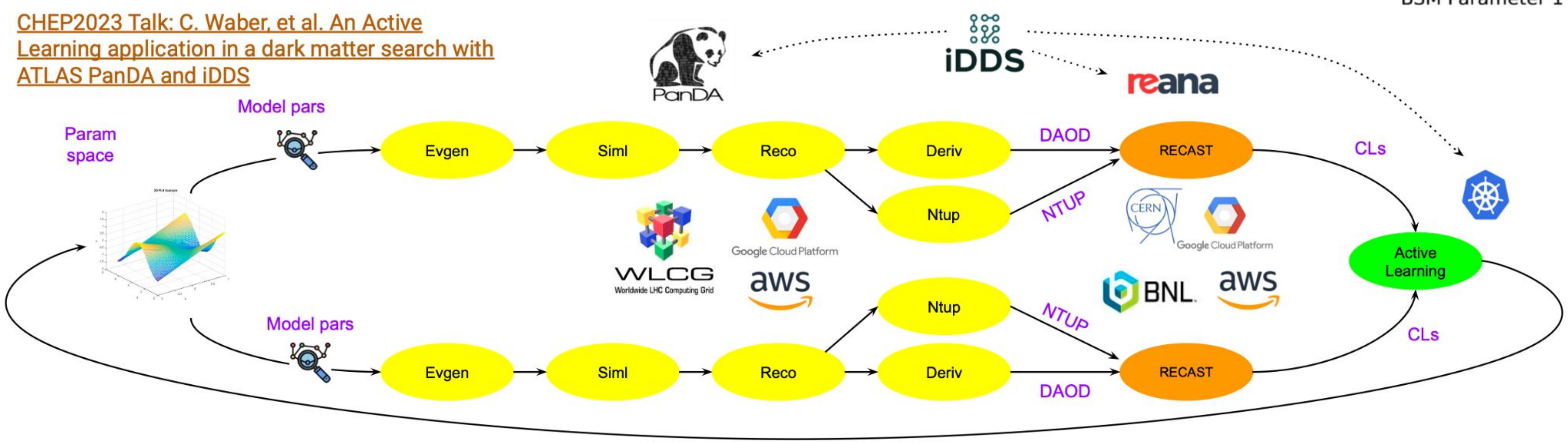
cf. The demo Christian presented yesterday morning

This builds on the HPO service

Workflows described using standard tools (CWL, snakemake python)

Workflows integrated across multiple resources: grid, cloud, HPC, REANA, GitLab CI

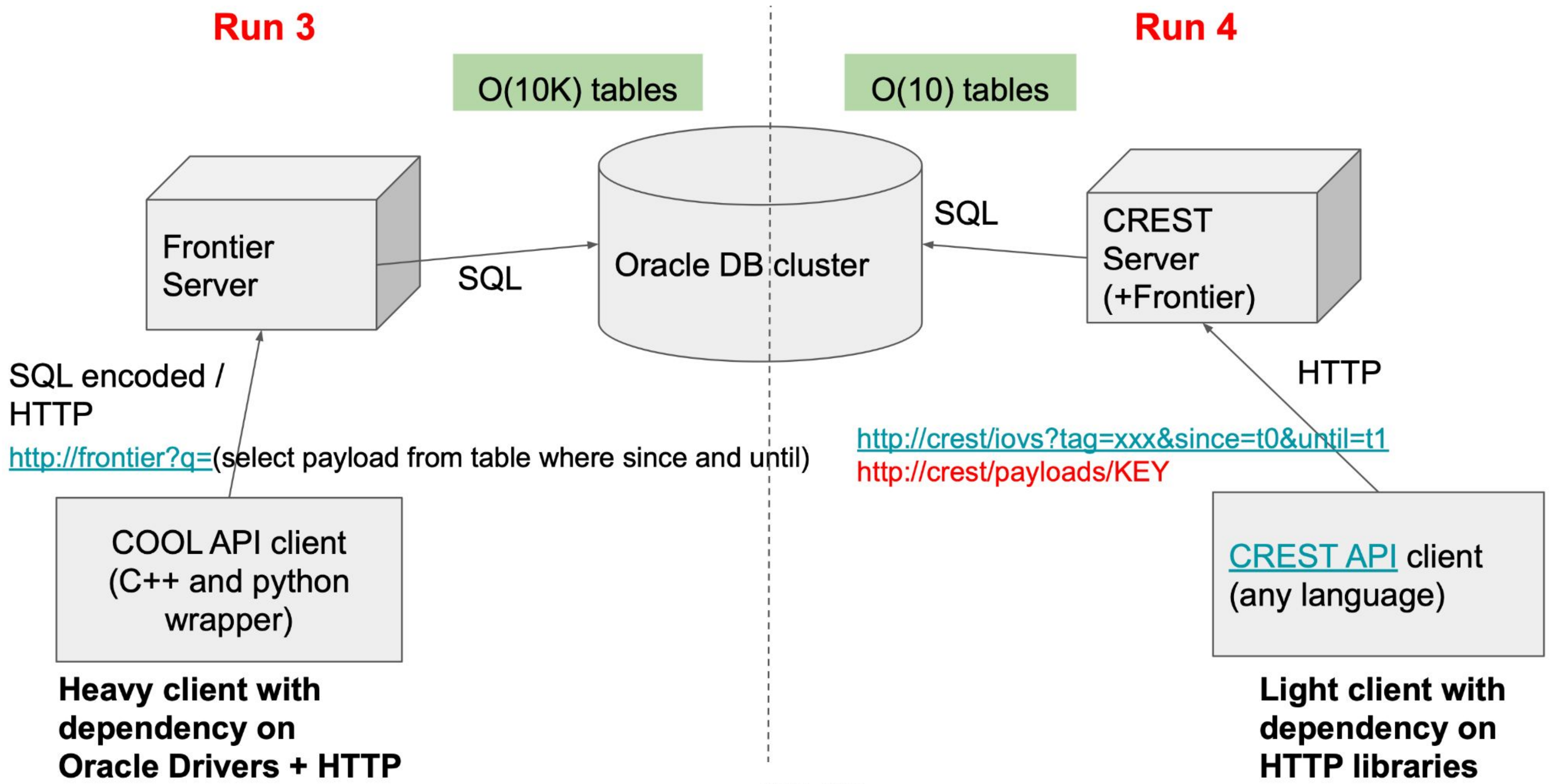
CHEP2023 Talk: C. Waber, et al. An Active Learning application in a dark matter search with ATLAS PanDA and iDDS



Tadashi Maeno, Utilizing Distributed Heterogeneous Computing with PanDA in ATLAS, CHEP 2023

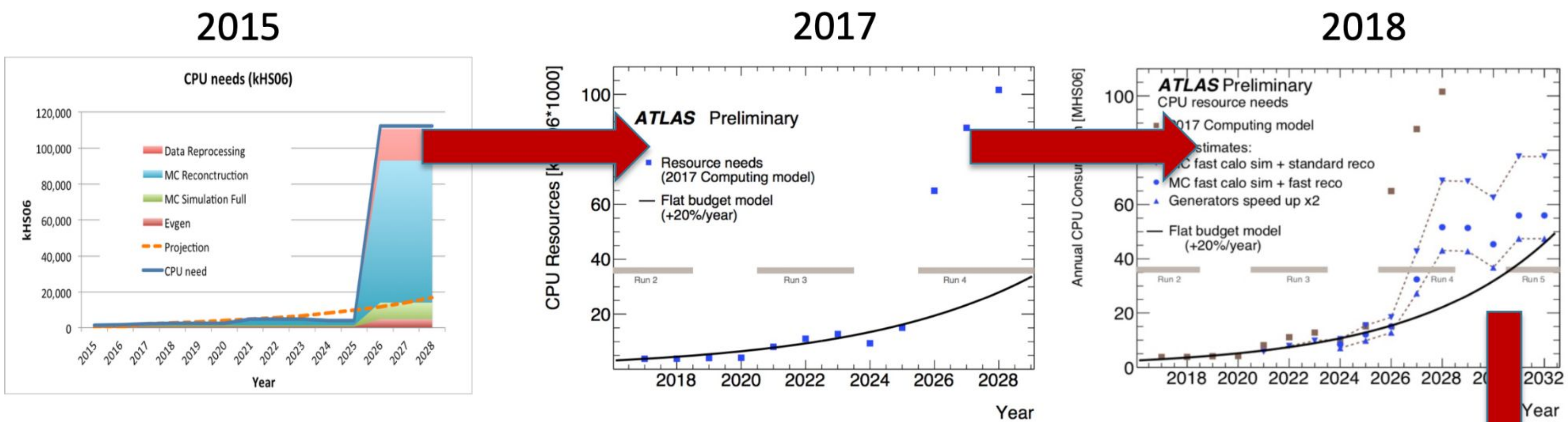
CREST: Next Gen Conditions Database

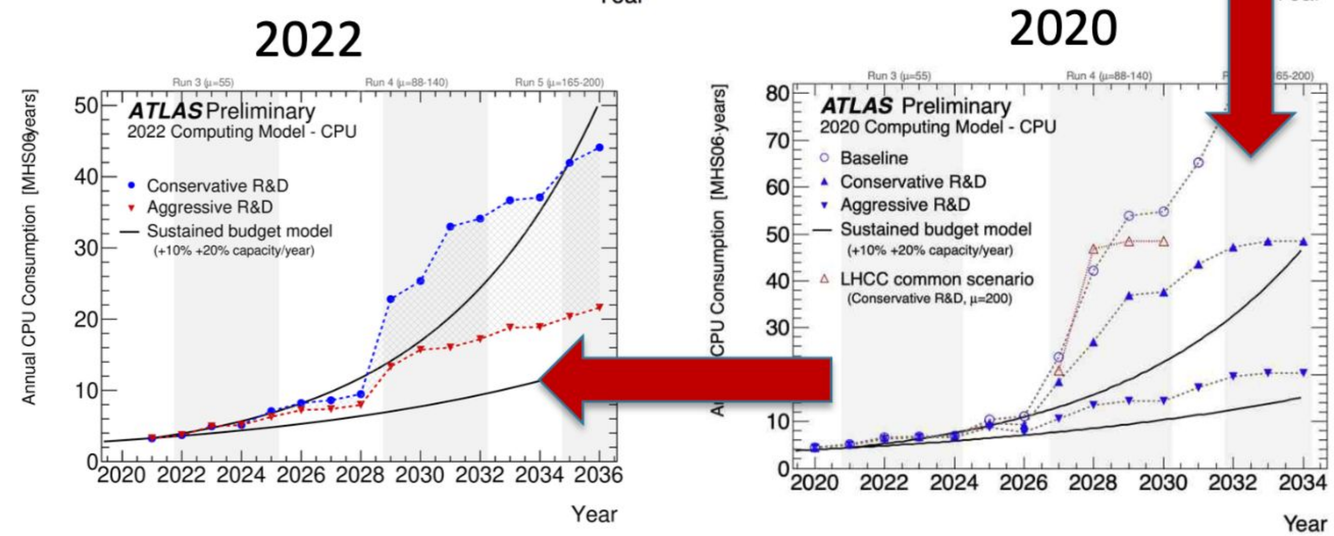
- Current ATLAS distributed conditions DB (COOL + Frontier) has scaling limitations
  - Arising particularly from coupling between request metadata (interval of validity) and the data payload
  - Prevents effective caching, when an http based service should excel at caching
- The new CREST system addresses this, following (broadly) the HSF conditions database reference design
  - HSF design arose from discussions on common architecture (not software) among ATLAS, CMS, LHCb
  - Design now the basis for other experiments also (Belle II, sPHENIX (RHIC), ~~ProtoDUNE~~)
- Not bound to Oracle
  - CERN may move away
  - Many RDBMSs supported
- Simple data model
- REST API
- Optimal caching
- C++ client
  - python, typescript, … via OpenApi
- Demonstrator by end 2023



Finally

- R&D efforts advancing on many fronts to bring HL-LHC processing and storage into affordability, and facilitate analysis at HL-LHC scale and with modern tool set
- Improvement have been steady over the years
- Challenge over the next months: quantify the improvements the R&D efforts have brought
- All will inform the ATLAS HL-LHC Computing Technical Design Report coming in a few years





2015 projections: resource needs = 10x more than budget allows

2022 projections: resource needs compatible with budget (optimistic scenario)

Simone Campana
CHEP 2023

Brookhaven National Laboratory

Thank you all contributors

Much thanks to the many contributors to CHEP 2023, ATLAS S&C meetings and the US ATLAS technical meetings for the excellent materials I've pillaged for this talk