INFN
Istituto Nazionale di Fisica Nucleare

UNIVERSITÀ
DEGLI STUDI
DI MILANO

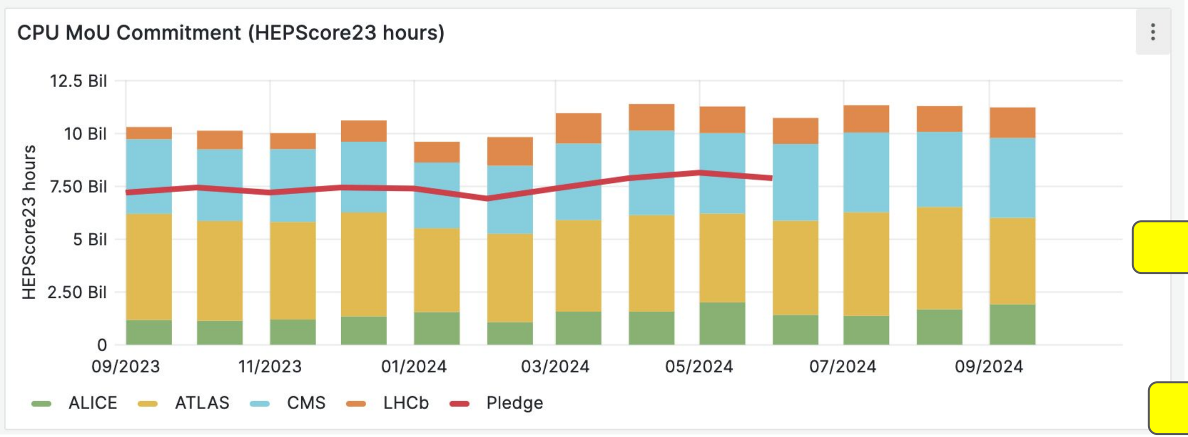
Perspectives for HEP computing: from LHC to FCC

L. Carminati

Milano, 29 ottobre 2024

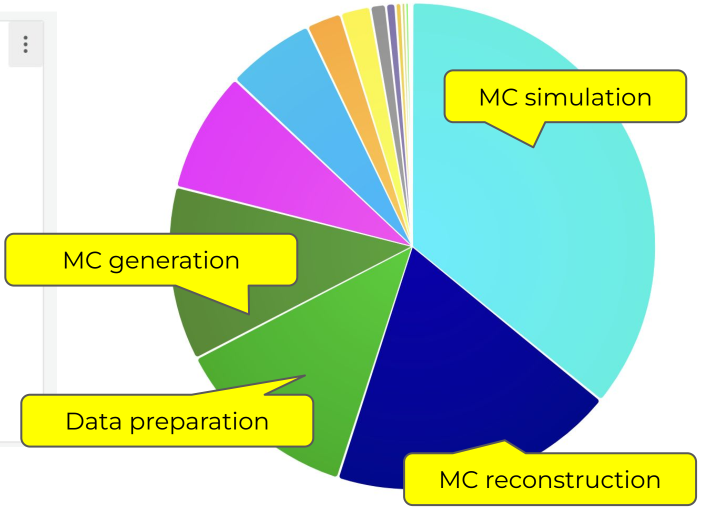
WLCG (Worldwide LHC computing grid)

- LHC computing based on the grid paradigm :
  - 170 computing centers, 40 counties
  - hierarchical structure T0/T1/T2 (T3).
- 10 Bil HS03 hours -> 1.3 M standard cores distributed all over the world and running 24/7
- Data (re)processing, MC event generation/simulation/reconstruction

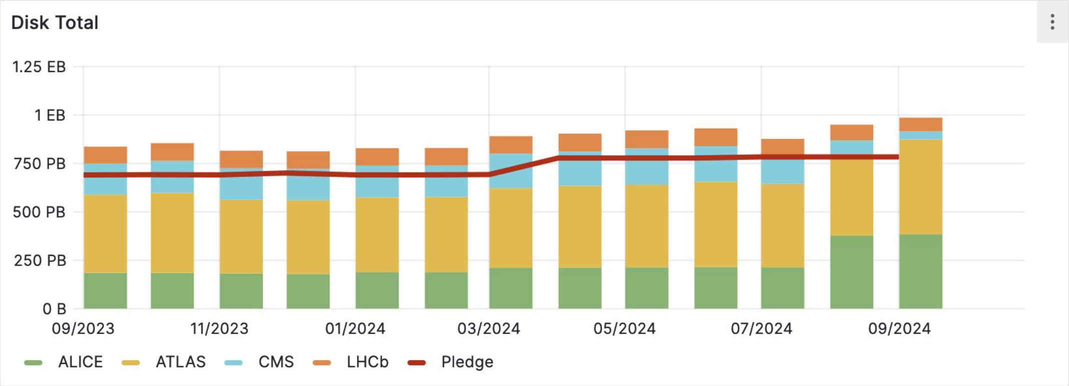


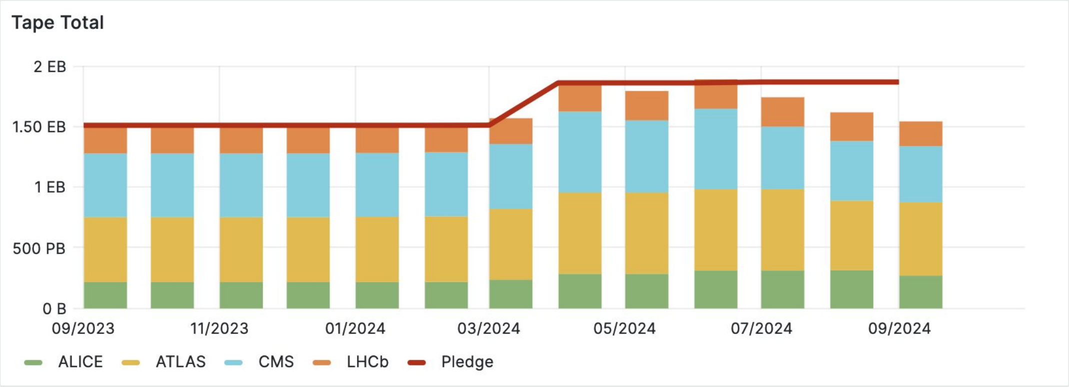
CPU MoU Commitment (HEPScore23 hours)

ATLAS



WLCG (Worldwide LHC computing grid)





ATLAS : ~ 350 PB of disk storage

❑ Analysis object data or derived analysis object data

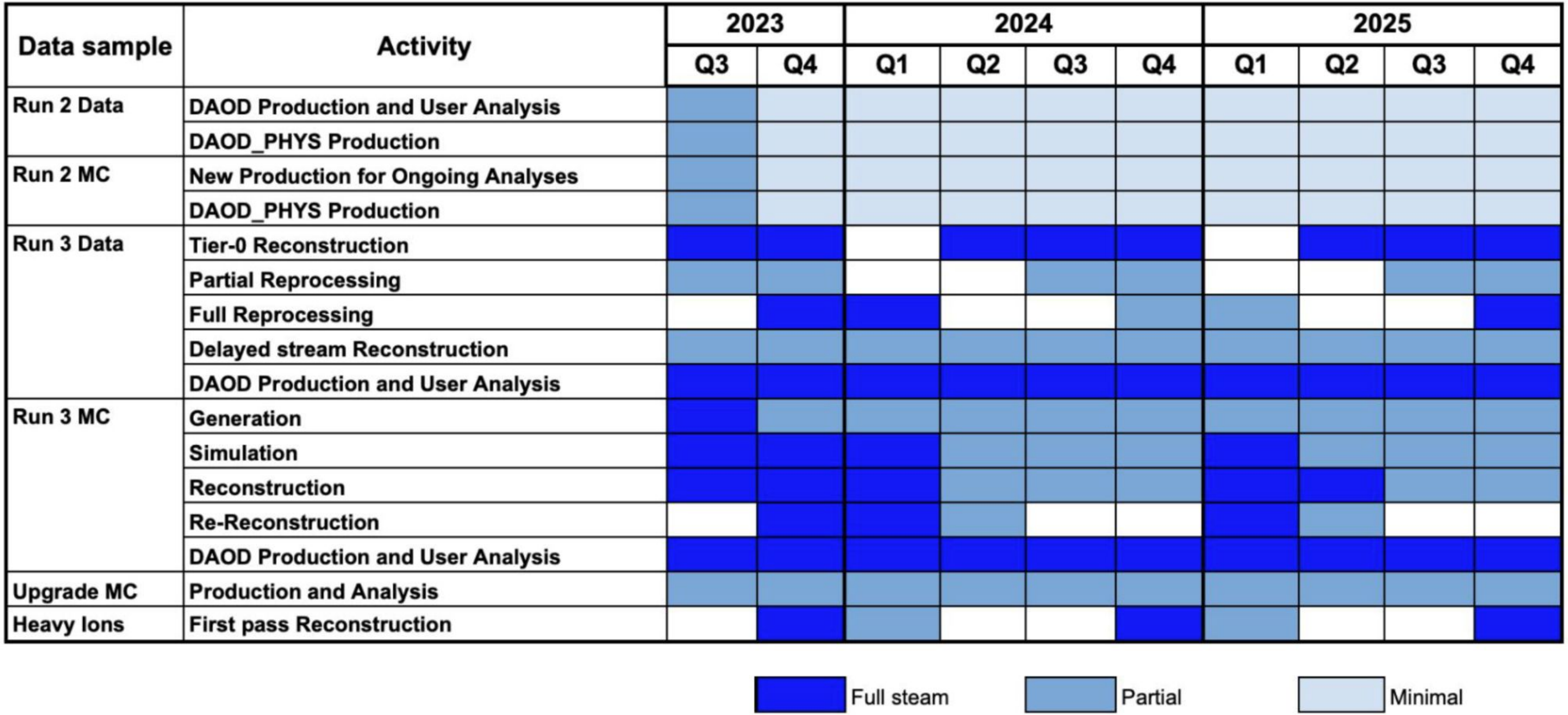
ATLAS : ~ 400 PB of tape

❑ 280 PB of RAW data

❑ 75 PB of analysis object data

Milano is hosting an ATLAS T2 :
~ 4000 cores running 24/7
~ 2 PB of disk space

Typical ATLAS computing workplan

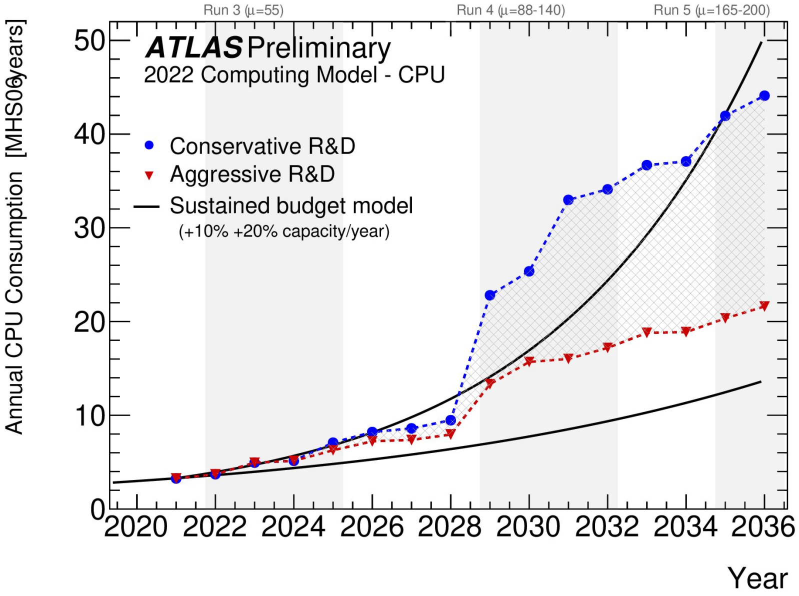


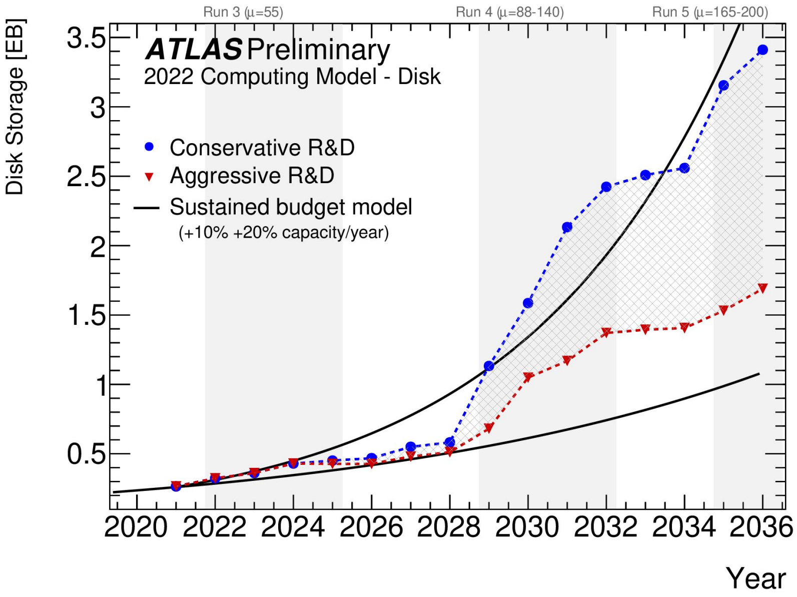
Data sample	Activity	2023		2024				2025			
		Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Run 2 Data	DAOD Production and User Analysis	▨	░	░	░	░	░	░	░	░	░
	DAOD_PHYS Production	▨	░	░	░	░	░	░	░	░	░
Run 2 MC	New Production for Ongoing Analyses	▨	░	░	░	░	░	░	░	░	░
	DAOD_PHYS Production	▨	░	░	░	░	░	░	░	░	░
Run 3 Data	Tier-0 Reconstruction	■	■		■	■	■		■	■	■
	Partial Reprocessing	▨	▨			▨	▨			▨	▨
	Full Reprocessing		■	■			▨	▨			■
	Delayed stream Reconstruction	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
	DAOD Production and User Analysis	■	■	■	■	■	■	■	■	■	■
Run 3 MC	Generation	■	▨	▨	▨	▨	▨	▨	▨	▨	▨
	Simulation	■	■	■	▨	▨	▨	■	▨	▨	▨
	Reconstruction	■	■	■	▨	▨	▨	■	■	▨	▨
	Re-Reconstruction		■	■	▨			■	▨		
	DAOD Production and User Analysis	■	■	■	■	■	■	■	■	■	■
Upgrade MC	Production and Analysis	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
Heavy Ions	First pass Reconstruction		■	▨			■	▨			■

■ Full steam ▨ Partial ░ Minimal

Hot topics : ATLAS Computing requirements for HL-LHC

ATLAS HL-LHC Computing Conceptual Design Report : projections of ATLAS computing requirements for Run3 and HL-LHC to fully exploit the machine physics potential is quite scaring !



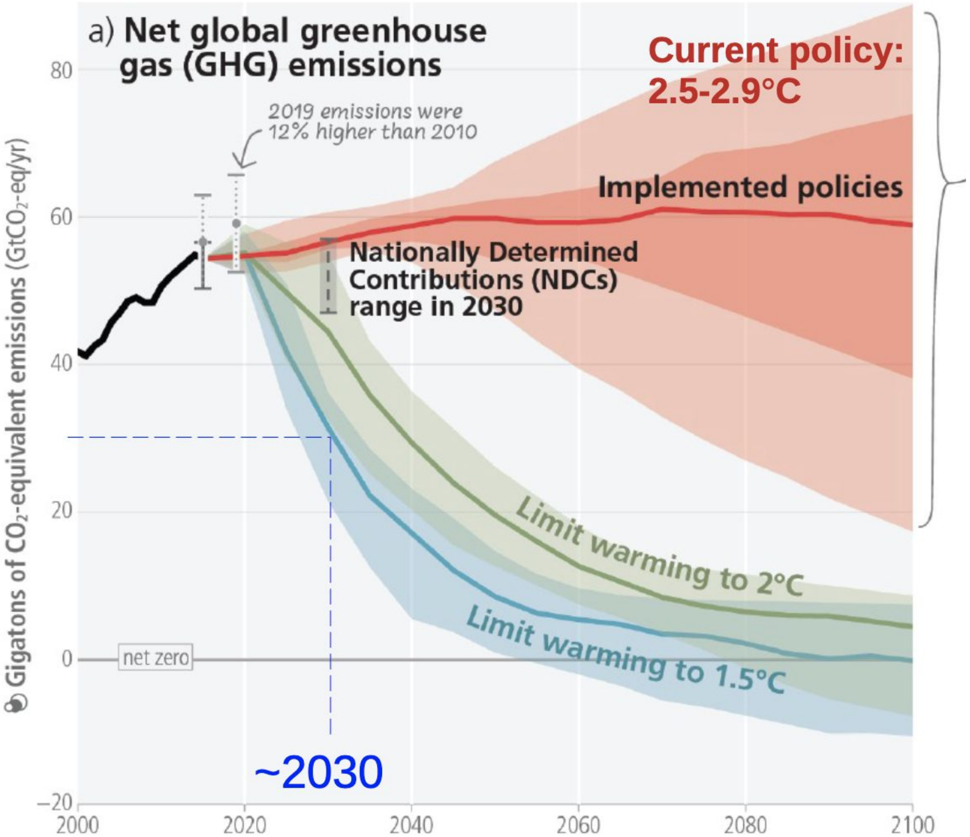


Discussion started on possible strategies to meet the demanding requirements of HL-LHC computing

- optimisation (both speed and flexibility) of the experiment (e.g. reconstruction, simulation) and non-experiment (e.g. generation) software
- optimisation of the available hardware infrastructure usage

Sustainability in HEP computing

https://agenda.infn.it/event/34738/contributions/191871/attachments/106878/150865/SustainComputingLohwasser_v2.pdf



❑ Reduction to zero emissions around 2100 : 50% of the reduction should be achieved by ~2030 → in 7 years

❑ WLCG requirements : from ~ 100% increase in energy needs (pessimistic scenario) to 10% (optimistic scenario)

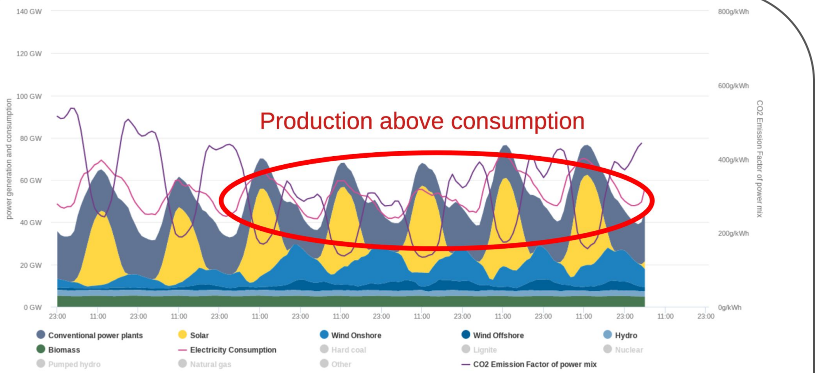
How to reconcile these ?

Sustainability in HEP computing

https://agenda.infn.it/event/34738/contributions/191871/attachments/106878/150865/SustainComputingLohwasser_v2.pdf

https://indico.cern.ch/event/1211083/contributions/5225666/attachments/2580240/4450193/2023-CERN-ATLAS-Geneva-final.pdf

- Expand CO_2-free energies (factor 12)
  - Renewable power for computing: processors and cooling;
  - Consider district heating and site selection;
  - Job scheduling according to energy availability; ...
- Increase energy efficiency (factor 2)
  - Optimised processors (clocks, GPUs),
  - architecture, cooling system,
  - software, quantum computing?, ...
- Save energy (factor 2)
  - Prioritise research questions
  - Optimise debugging, statistics and precision;
  - Modular and reusable software;
  - Modular and repairable hardware, reduce purchases;



Rather than having the grid adapt to the data center, we need to have the data center adapt to the grid .. The data center must “dance with the grid.”

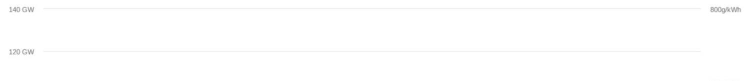
Reminder: <u>Paris agreement is in principle legally binding</u>

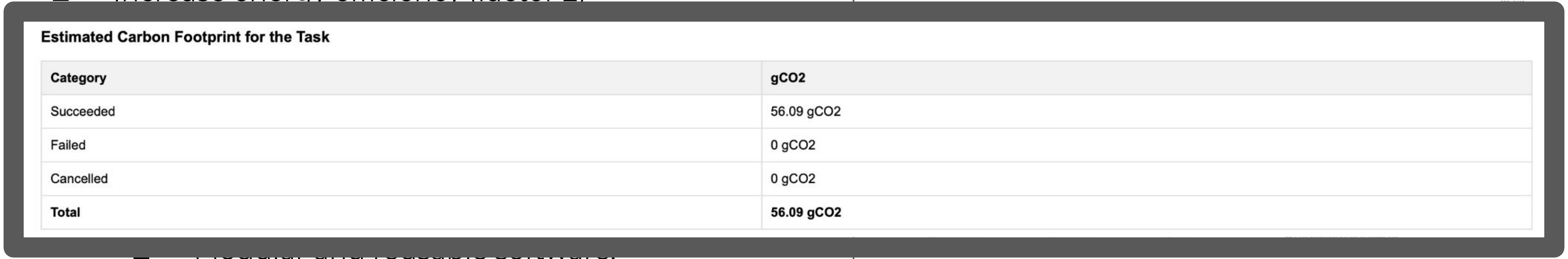
- pressure on us / our savings might need to be increased
- gives us negotiating power if we have a clear plan and strategy with demonstrable impacts and realistically achievable objectives in line with 1.5°C

Sustainability in HEP computing

https://agenda.infn.it/event/34738/contributions/191871/attachments/106878/150865/SustainComputingLohwasser_v2.pdf

- ❑ Expand CO2-free energies (factor 12)
  - ❑ Renewable power for computing: processors and cooling;
  - ❑ Consider district heating and site selection;
  - ❑ Job scheduling according to energy availability; ...
- ❑ Increase energy efficiency (factor 2)





Estimated Carbon Footprint for the Task

Category	gCO2
Succeeded	56.09 gCO2
Failed	0 gCO2
Cancelled	0 gCO2
Total	**56.09 gCO2**

  - ❑ Modular and reusable software;
  - ❑ Modular and repairable hardware, reduce purchases;

Rather than having the grid adapt to the data center, we need to have the data center adapt to the grid .. The data center must “dance with the grid.”

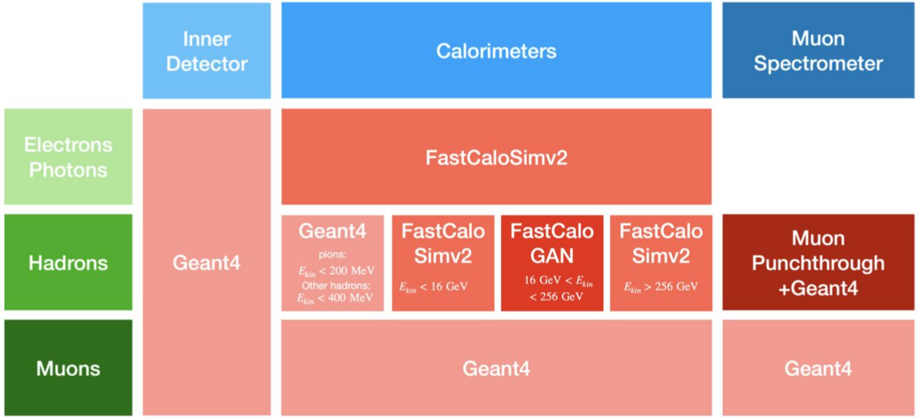
Reminder: Paris agreement is in principle legally binding

- ❑ pressure on us / our savings might need to be increased
- ❑ gives us negotiating power if we have a clear plan and strategy with demonstrable impacts and realistically achievable objectives in line with 1.5°C

Software R&D

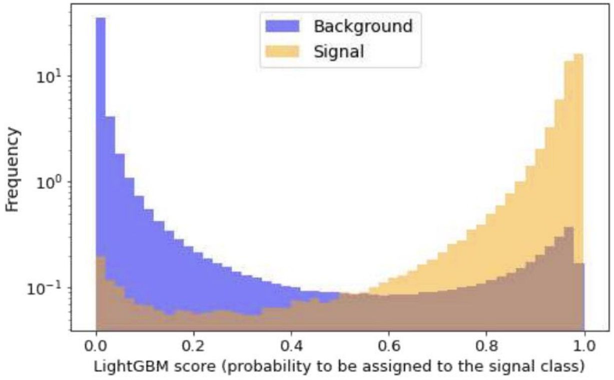
Intensive R&D program to improve the software performance

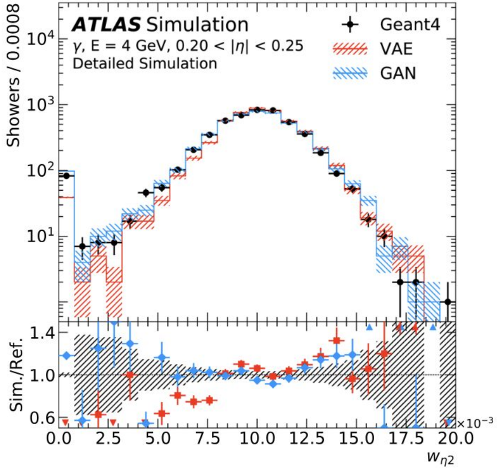
- Several improvements in the CPU performance of full simulation : full simulation is about 2 times faster than it was for run 2.
- Improve fast simulation performance : project to reduce by a factor up to 50 the required computing time. Trade off between accuracy and computing performance
- Keep communication channels open to improve event generation
- Improved analysis models
- Better use of existing infrastructure (eg. Tape Carousel, central role of Tapes)
- Heterogeneous computing :
    - low power architectures
    - use of accelerators (eg GPU)
    - adapt to easily integrate external resource (eg cloud)



AI/QC intermezzo

- Multivariate analysis commonplace in in HEP since 30 years
- ”Modern” ML now making paradigm-shifting contributions
  - Driven by industry, dedicated solutions needed for HEP, e.g.
    - Classification, tagging, calibration
    - Generative models in simulation, generation, lattice gauge theory
    - Unsupervised classification for anomaly detection (BSM searches)
  - Software and hardware needs might not be aligned with industry
    - standard – partnership / direct contributions beneficial
- Quantum Computing is also a paradigm shift
  - Rapid development of (open-source) software and hardware on several platforms
  - HEP use cases: lattice gauge theory, event generation, data analysi





New computing actors : EuroHPC

VEGA : 10 petaflops (peak)

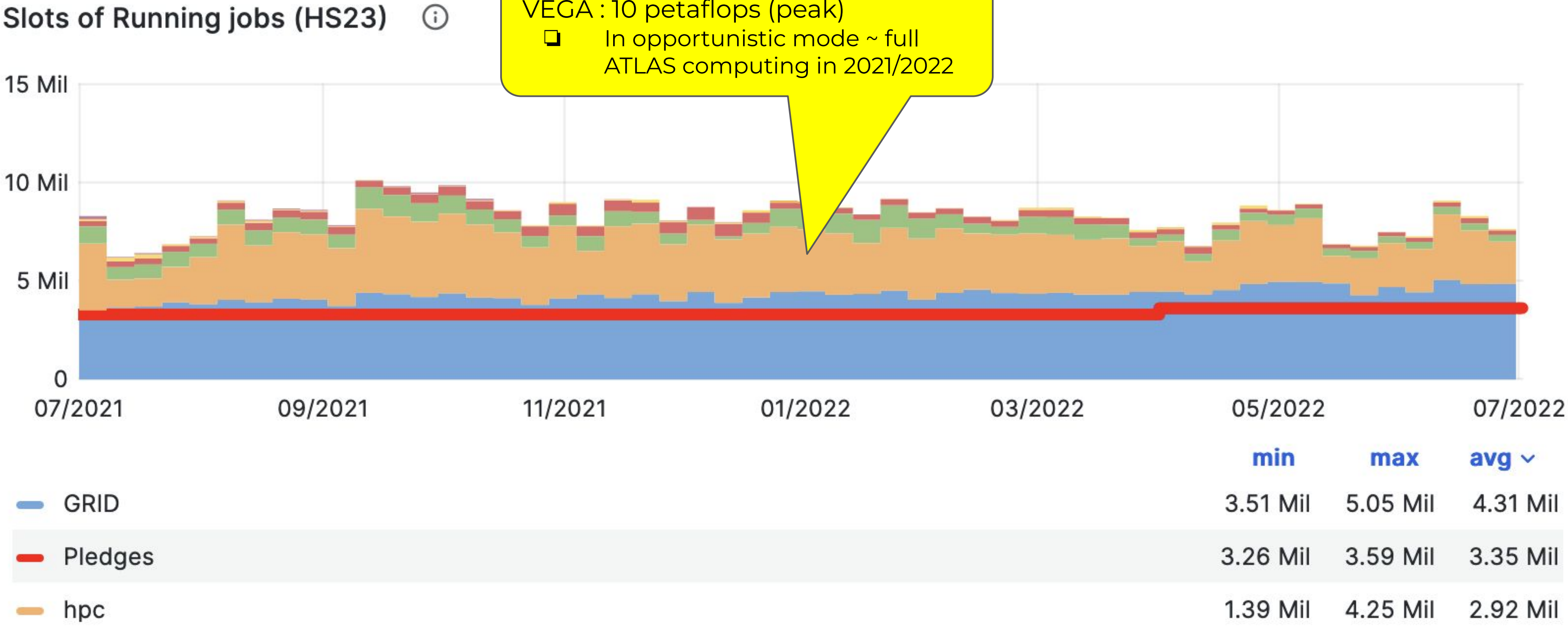
- In opportunistic mode ~ full ATLAS computing in 2021/2022

EuroHPC
Joint Undertaking

LEONARDO : 315 petaflops (peak)

JUPITER : 1 exaflops (peak) !

New computing actors : EuroHPC

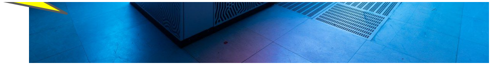


Slots of Running jobs (HS23)

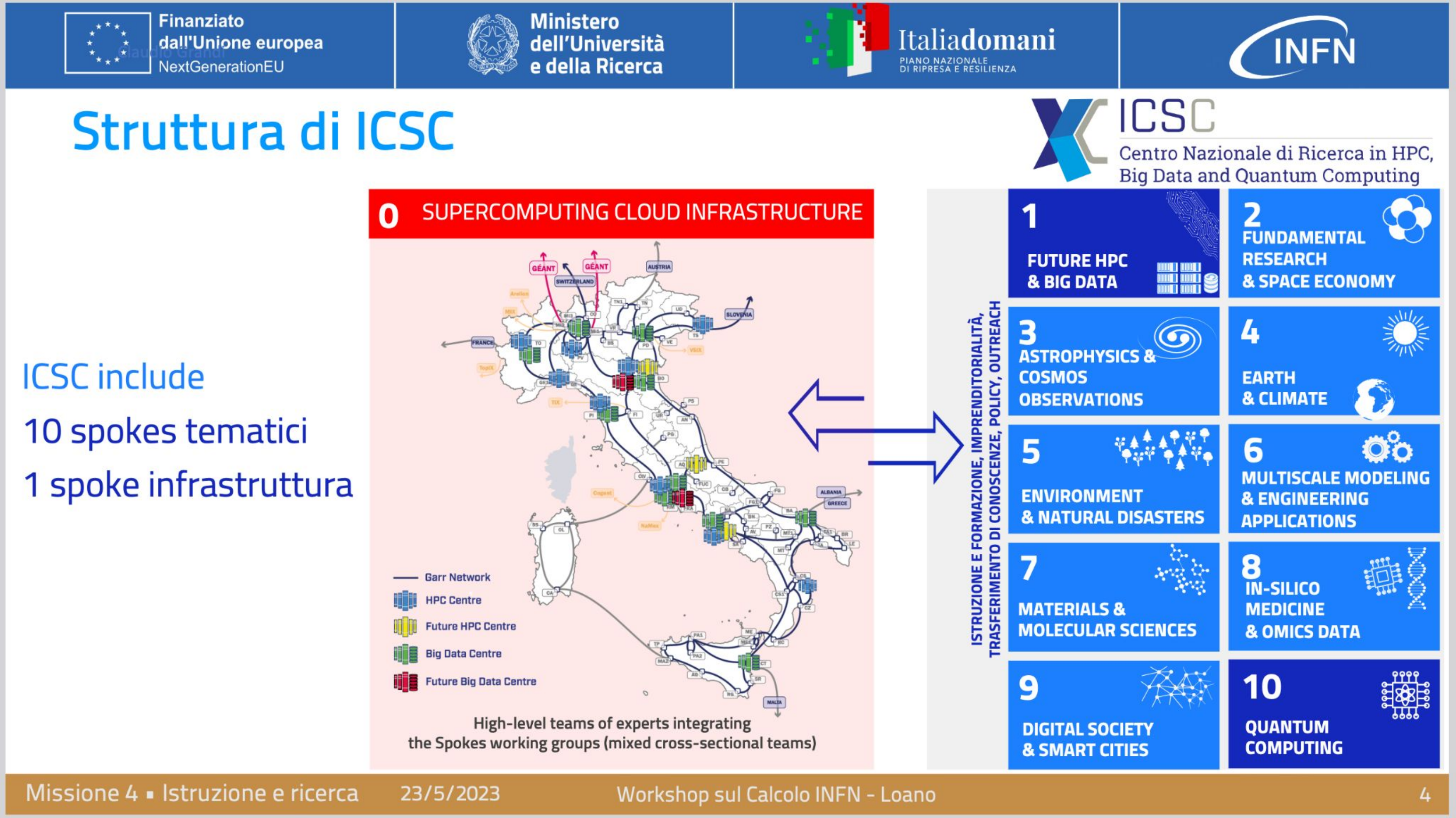
VEGA : 10 petaflops (peak)

- In opportunistic mode ~ full ATLAS computing in 2021/2022

	min	max	avg
GRID	3.51 Mil	5.05 Mil	4.31 Mil
Pledges	3.26 Mil	3.59 Mil	3.35 Mil
hpc	1.39 Mil	4.25 Mil	2.92 Mil



New computing actors : ICSC



Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INFN

Struttura di ICSC

ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

ICSC include

10 spokes tematici

1 spoke infrastruttura

0 SUPERCOMPUTING CLOUD INFRASTRUCTURE

— Garr Network

▥ HPC Centre

▥ Future HPC Centre

▥ Big Data Centre

▥ Future Big Data Centre

High-level teams of experts integrating the Spokes working groups (mixed cross-sectional teams)

ISTRUZIONE E FORMAZIONE, IMPRENDITORIALITÀ, TRASFERIMENTO DI CONOSCENZE, POLICY, OUTREACH

1 FUTURE HPC & BIG DATA

2 FUNDAMENTAL RESEARCH & SPACE ECONOMY

3 ASTROPHYSICS & COSMOS OBSERVATIONS

4 EARTH & CLIMATE

5 ENVIRONMENT & NATURAL DISASTERS

6 MULTISCALE MODELING & ENGINEERING APPLICATIONS

7 MATERIALS & MOLECULAR SCIENCES

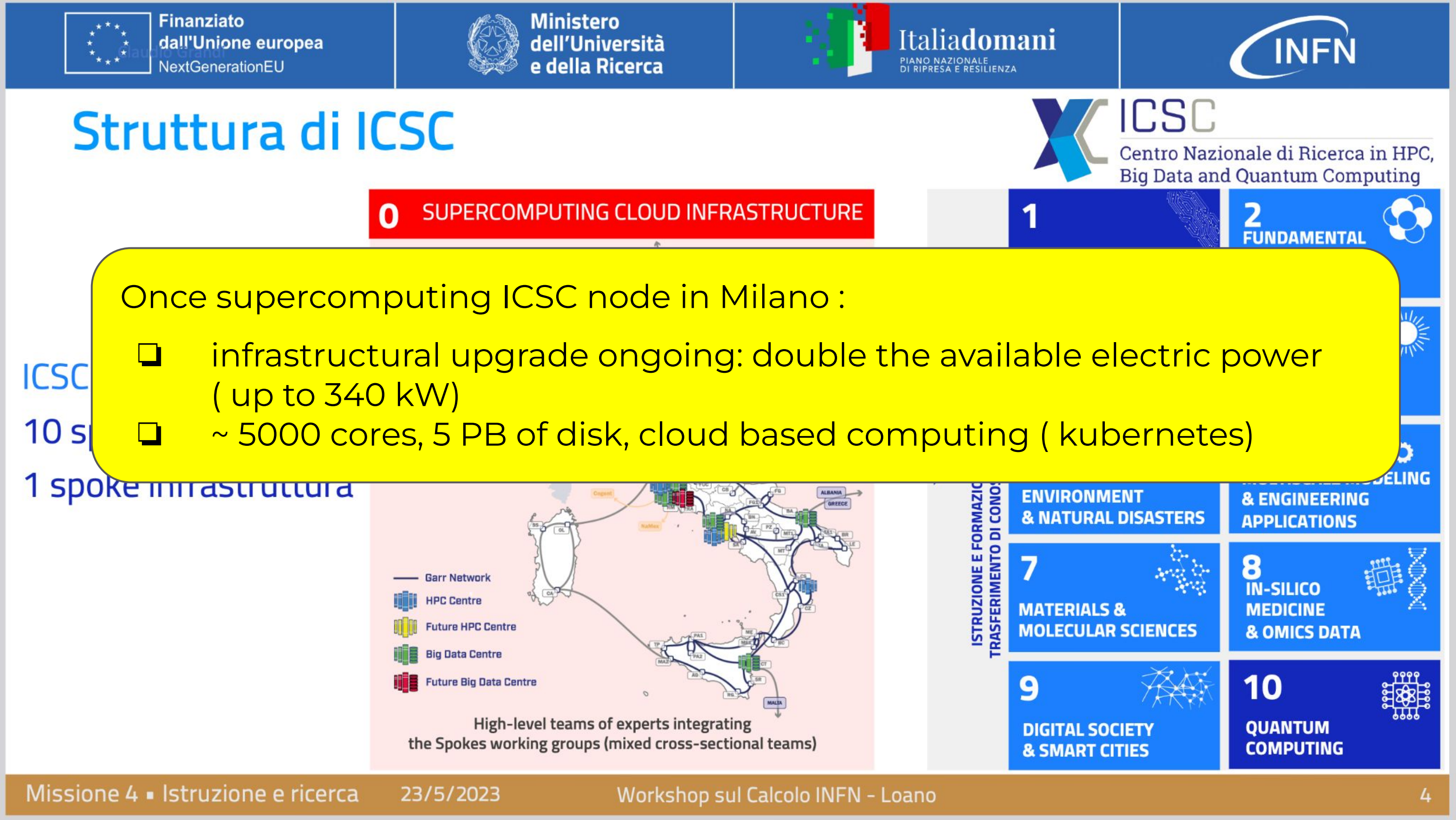
8 IN-SILICO MEDICINE & OMICS DATA

9 DIGITAL SOCIETY & SMART CITIES

10 QUANTUM COMPUTING

Missione 4 ▪ Istruzione e ricerca 23/5/2023 Workshop sul Calcolo INFN - Loano 4

New computing actors : ICSC



Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

INFN

Struttura di ICSC

ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

0 SUPERCOMPUTING CLOUD INFRASTRUCTURE

ICSC

10 s

1 spoke infrastruttura

Garr Network

HPC Centre

Future HPC Centre

Big Data Centre

Future Big Data Centre

High-level teams of experts integrating the Spokes working groups (mixed cross-sectional teams)

ISTRUZIONE E FORMAZIO

TRASFERIMENTO DI CONOS

1

2 FUNDAMENTAL

ENVIRONMENT & NATURAL DISASTERS

ELING & ENGINEERING APPLICATIONS

7 MATERIALS & MOLECULAR SCIENCES

8 IN-SILICO MEDICINE & OMICS DATA

9 DIGITAL SOCIETY & SMART CITIES

10 QUANTUM COMPUTING

Missione 4 ▪ Istruzione e ricerca 23/5/2023 Workshop sul Calcolo INFN - Loano 4

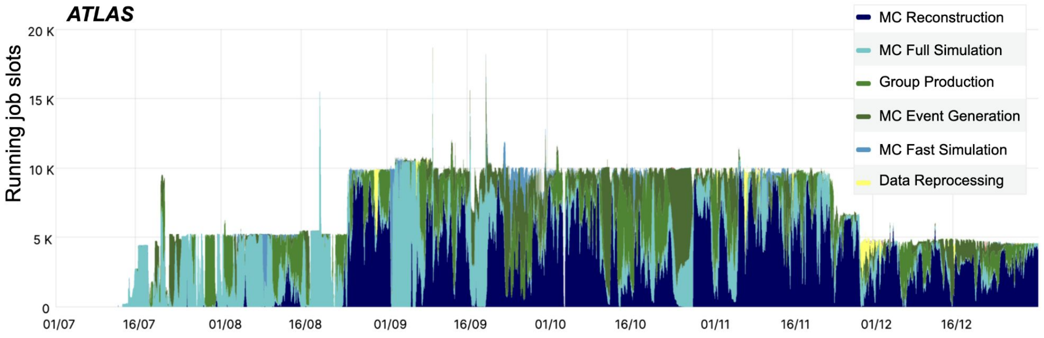
Once supercomputing ICSC node in Milano :

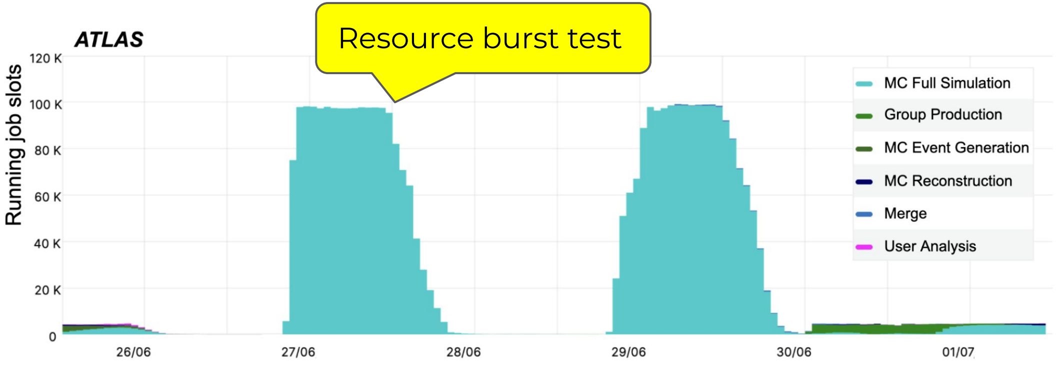
- ❏ infrastructural upgrade ongoing: double the available electric power (up to 340 kW)
- ❏ ~ 5000 cores, 5 PB of disk, cloud based computing (kubernetes)

New computing actors : Google based computing

https://arxiv.org/pdf/2405.13695

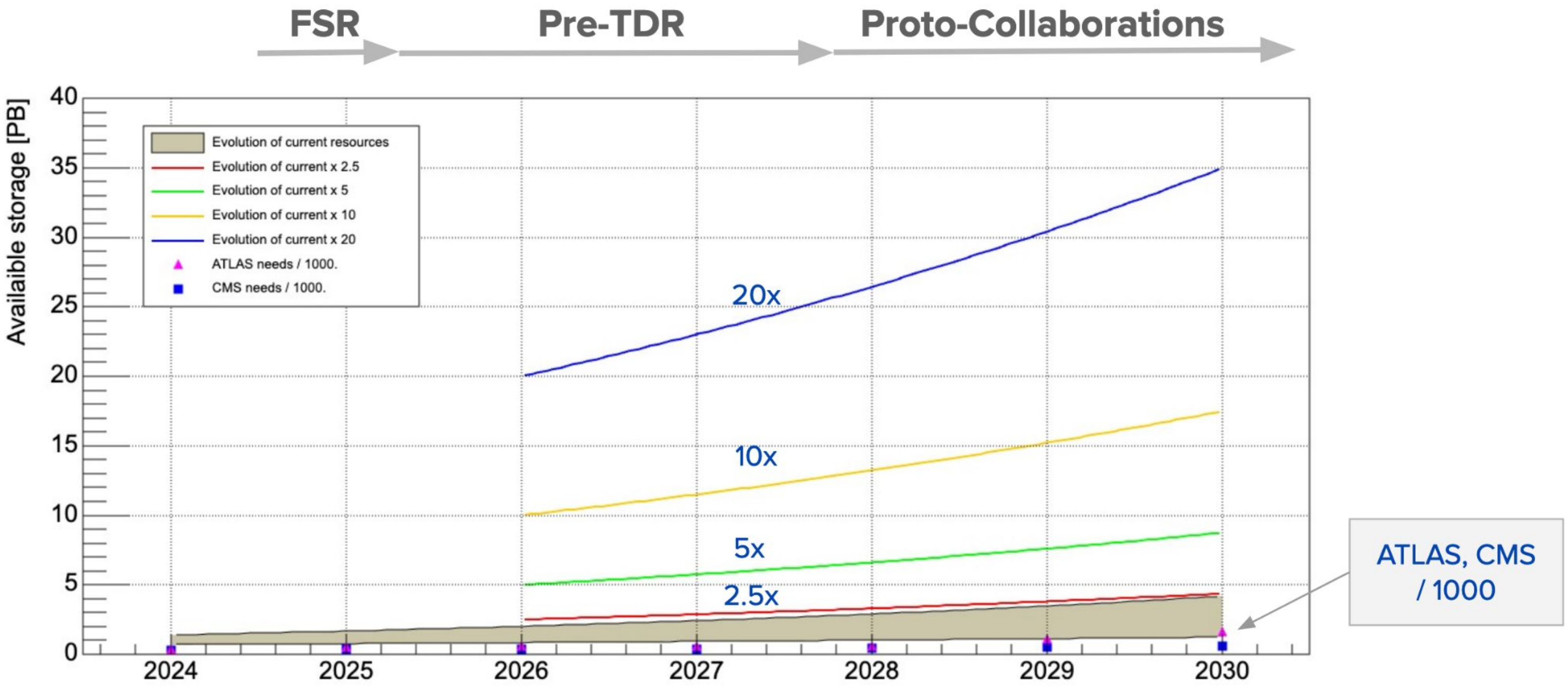
- Create a medium size ATLAS computing center on Google cloud resources
- 7,000 compute cores together with up to 7 PB of storage, and an estimated egress of no more than 0.7 PB per month
- From July 2022 to October 2023, project cost a total of $3.162M at list–price compared to the $849,458 paid via the subscription agreement
- ATLAS tools adapted to cloud based infrastructure in a cloud–agnostic way
- Establishing a viable and cost effective subscription agreement between experiment and commercial cloud provider is clearly a critical consideration of the TCO,




Resource burst test

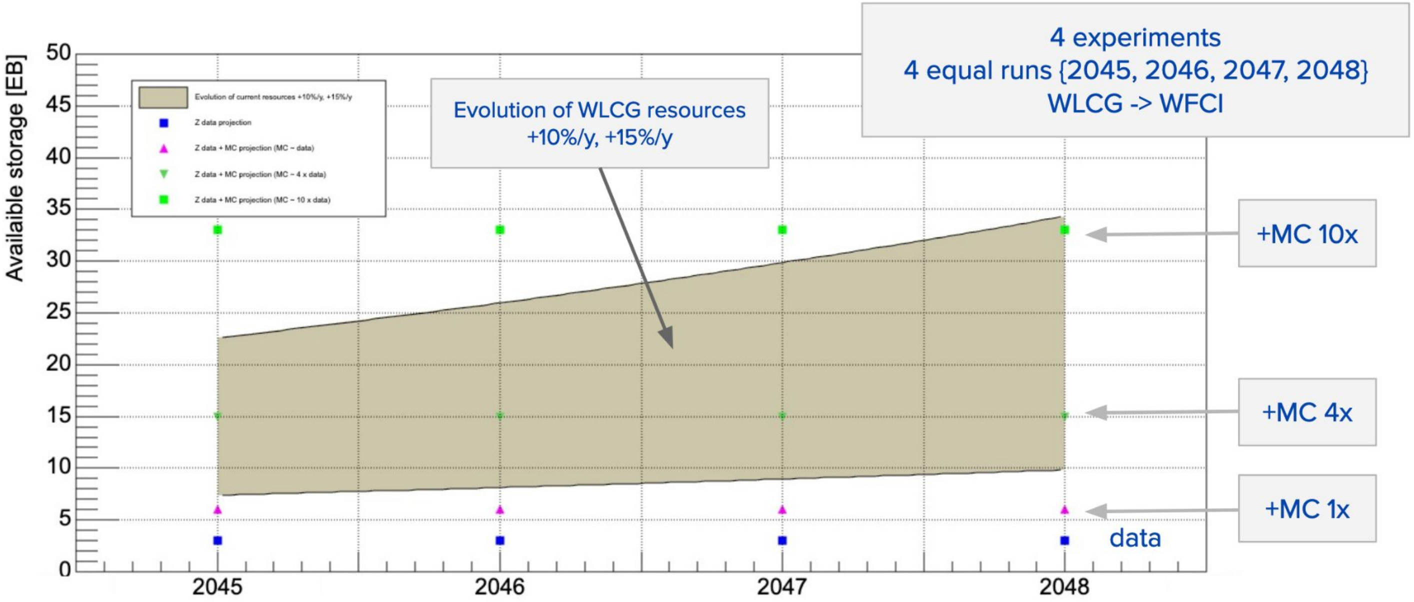
Computing perspective for FCC

FCC Week



Computing perspective for FCC

- Integrated luminosities
  - Nominal: {90, 12, 5, 0.2, 1.5} ab^{-1} at $\sqrt{s}$ = {91.2, 160, 240, 350, 365} GeV : # of evts: 3x10^{12} visible Z decays, 10^{8} WW events, 10^{6} ZH events, 10^{6} tt events
- Baseline event sizes / processing time for hadronic evts at Z
  - DELPHES: full stat sample size 30 PB, ≈ 10^{10} s/core ≈ 0.5 MHS06[2]
  - Full sim (CLD) full stat sample sizes: 3 EB, ≈ 3·10^{13} s/core ≈ 10-15 MHS06[2] / detector [3]



Evolution of WLCG resources
+10%/y, +15%/y

4 experiments
4 equal runs {2045, 2046, 2047, 2048}
WLCG -> WFCI

+MC 10x

+MC 4x

+MC 1x

data

(tentative) Conclusions

❑ Software and computing are fundamental parts of HEP initiatives, sometimes overlooked

❑ The panorama is evolving very quickly, new actors and technologies are appearing
  ❑ HEP community needs to be involved in all different R&Ds

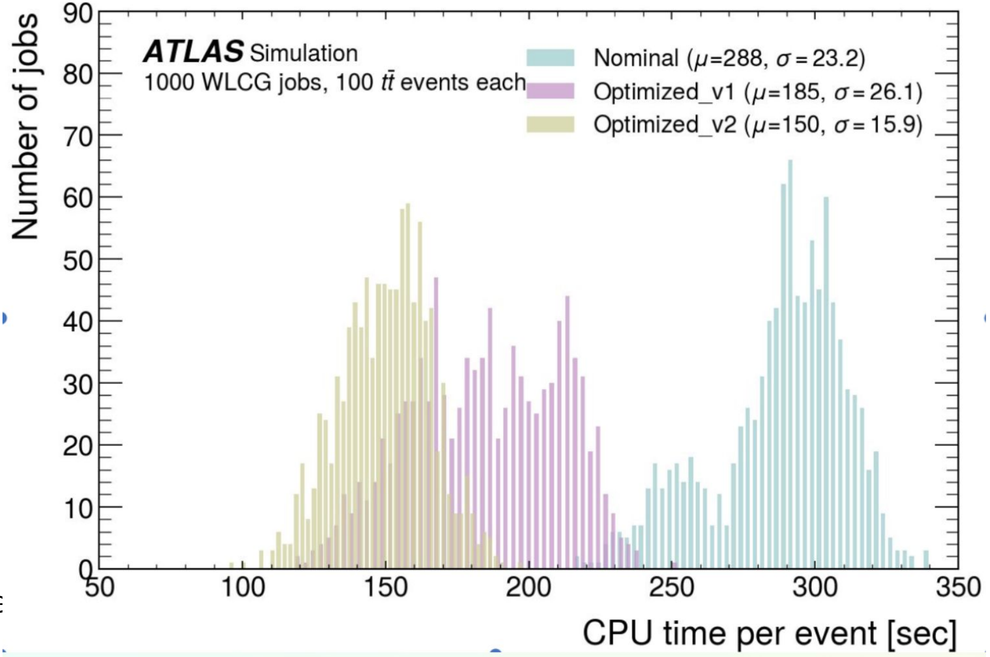
❑ Difficult to make predictions on a long time scale, nevertheless need to be prepared to take advantage of the upcoming opportunities (and experiments are intensively working on this)

Backup

More efficient full simulation

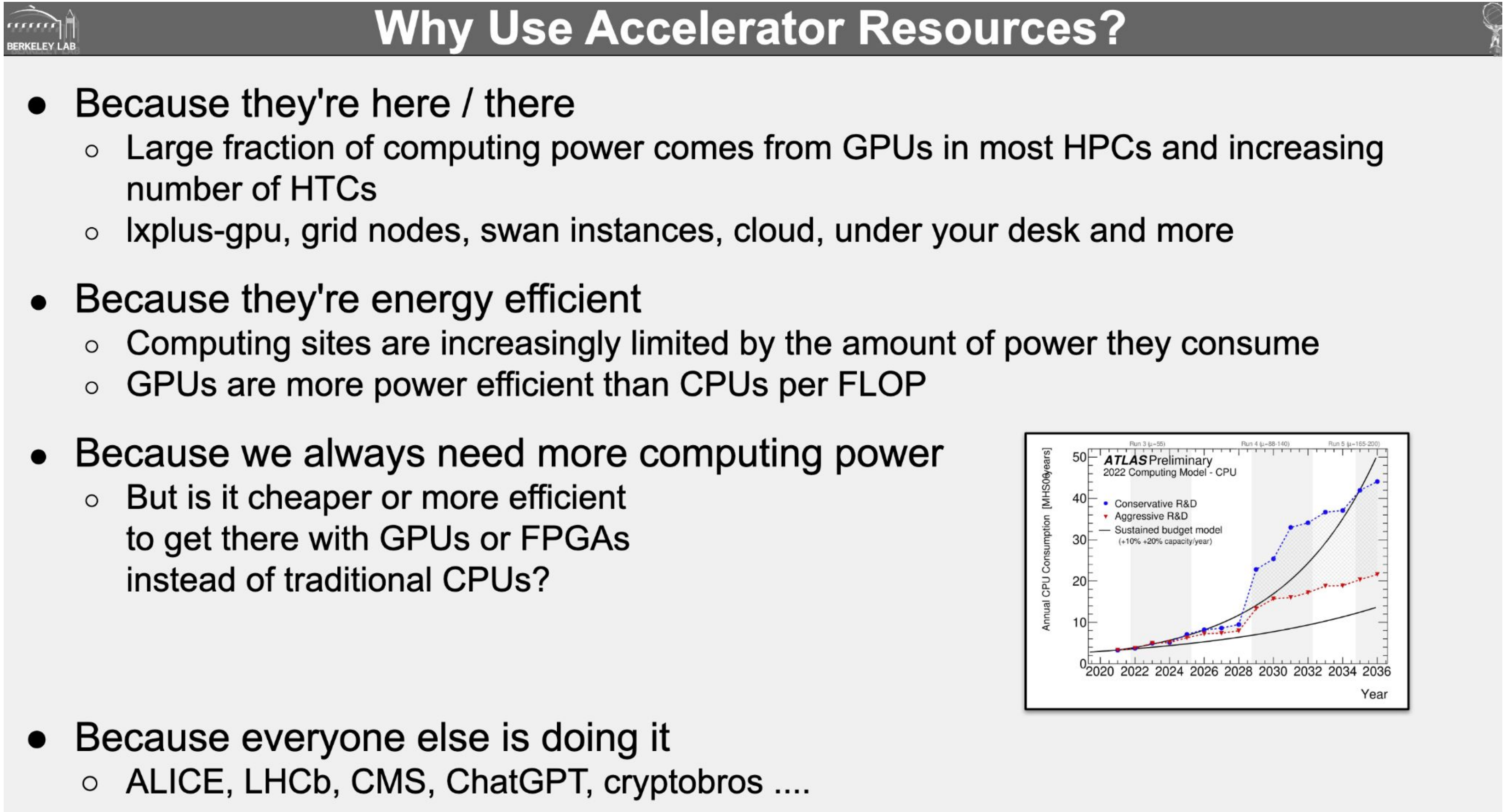
https://indico.cern.ch/event/1180455/contributions/5318122/attachments/2672908/4633983/AtlasWeek_june23.pdf

- ❏ Several improvements in the CPU performances of full simulation (with no loss of accuracy) have been implemented in the last two years : full simulation is about 2 times faster than it was for run 2.
- ❏ Work to further speed up full simulation for future MC campaigns and HL-LHC continue
- ❏ A new EMEC geometry (using G4 supported shapes) makes the simulation of EMEC 2.5 times faster and GPU compatible
- ❏ In parallel, prototypes of GPU simulation are being developed in collaboration with Adept and Celeritas (two groups developing GPU simulation across experiments).
- ❏ Expect to be able to test a GPU workflow for the HL LHC Computing TDR next year



Heterogeneous computing

https://indico.cern.ch/event/1180455/contributions/4958179/attachments/2675554/4639712/Heterogeneous%20Computing%20in%20ATLAS.pdf

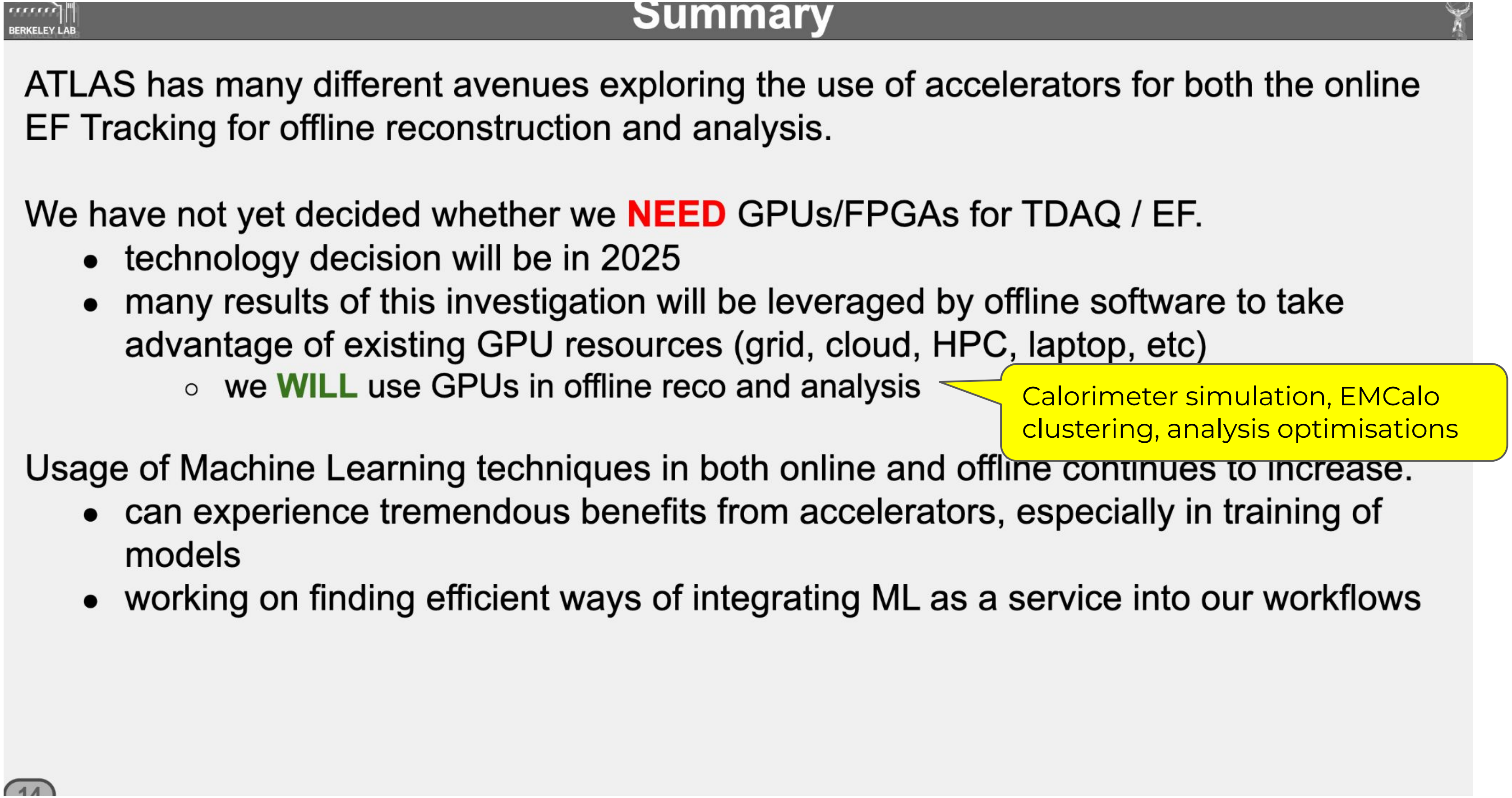


Why Use Accelerator Resources?

- Because they're here / there
  - Large fraction of computing power comes from GPUs in most HPCs and increasing number of HTCs
  - lxplus-gpu, grid nodes, swan instances, cloud, under your desk and more
- Because they're energy efficient
  - Computing sites are increasingly limited by the amount of power they consume
  - GPUs are more power efficient than CPUs per FLOP
- Because we always need more computing power
  - But is it cheaper or more efficient to get there with GPUs or FPGAs instead of traditional CPUs?
- Because everyone else is doing it
  - ALICE, LHCb, CMS, ChatGPT, cryptobros ....

Heterogeneous computing

https://indico.cern.ch/event/1180455/contributions/4958179/attachments/2675554/4639712/Heterogeneous%20Computing%20in%20ATLAS.pdf



Summary

ATLAS has many different avenues exploring the use of accelerators for both the online EF Tracking for offline reconstruction and analysis.

We have not yet decided whether we **NEED** GPUs/FPGAs for TDAQ / EF.

- technology decision will be in 2025
- many results of this investigation will be leveraged by offline software to take advantage of existing GPU resources (grid, cloud, HPC, laptop, etc)
  - we **WILL** use GPUs in offline reco and analysis

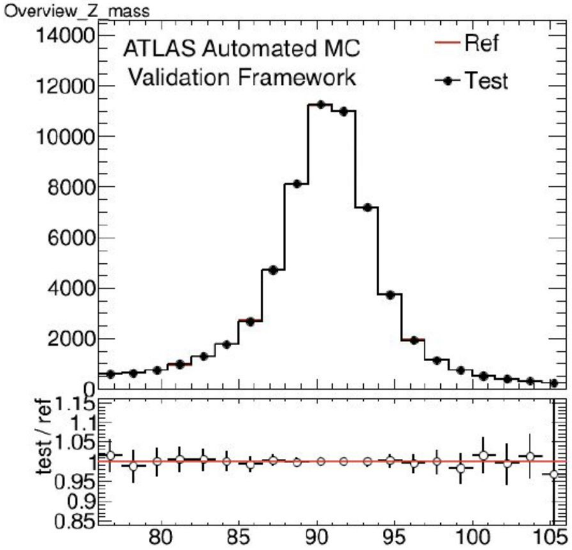
Calorimeter simulation, EMCalo clustering, analysis optimisations

Usage of Machine Learning techniques in both online and offline continues to increase.

- can experience tremendous benefits from accelerators, especially in training of models
- working on finding efficient ways of integrating ML as a service into our workflows

Heterogeneous computing

- ❑ ARM CPUs have low cost and minimal power consumption.
- ❑ Some future HPCs plan to have ARM CPU partitions
- ❑ Athena, AthSimulation, AnalysisBase ported to the aarch64-centos7-gcc11-opt platform last year
- ❑ Technical physics validation successfully passed using Amazon
- ❑ Cloud Gravition2 ARM CPU fully integrated into PanDA/Rucio:
  - ❑ Full simulation Athena,23.0.3 (ATLPHYSVAL-872)
  - ❑ Reconstruction Athena,23.0.14 (ATLPHYSVAL-919)
- ❑ 3 ATLAS workflows (Sherpa, Geant4, Reco) integrated into the new WLCG HEPscore benchmark available for x86_64 and ARM
- ❑ WLCG sites start to get interested in ARM
  - ❑ Glasgow team shows competitive result for events/kWh and throughput

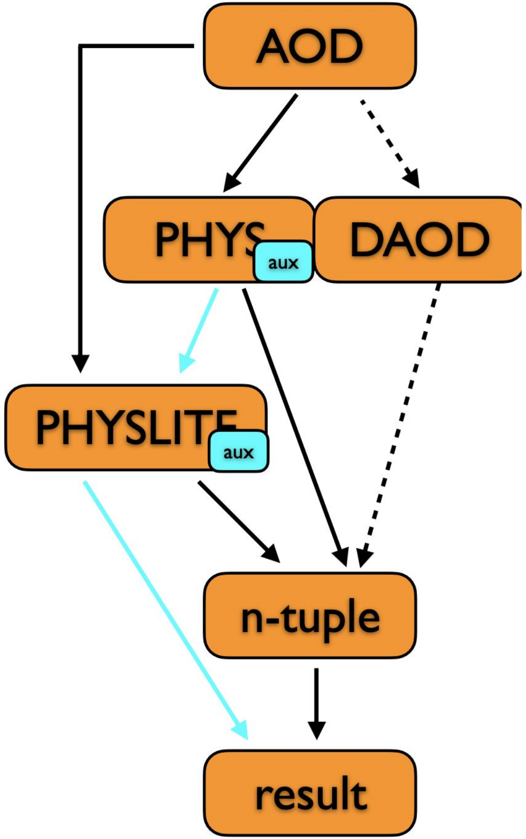


Reconstructed MC $Z\rightarrow\mu\mu$ invariant mass

Analysis model(s)

https://indico.cern.ch/event/1180455/contributions/5359601/attachments/2673977/4637700/AMG-2023-06-27.pdf

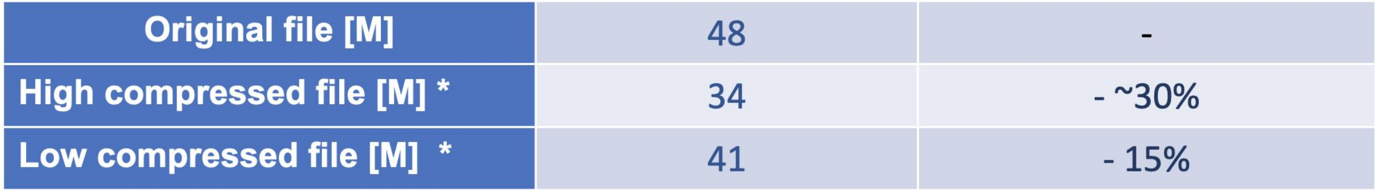
- ❑ The output of the data and MC reconstruction is stored in Analysis Object Data (AOD) files
- ❑ In run 2 these datasets were processed in the derivation framework which produces about 80 different derived AOD (DAOD) formats .
- ❑ In run 3 DAOD_PHYS now the default format for analysis
- ❑ One more DAOD format (especially in view of run4/5) : PHYSLITE
  - ❑ object CP corrections already applied
  - ❑ extra good object thinning
  - ❑ not all content from PHYS
- ❑ future: use PHYSLITE like an n-tuple push towards columnar approach + CP tools, closer to python ecosystem (lot of work ongoing on this subject)
- ❑ future:
  - ❑ event augmentation : save extra data for some events, further reduce need for custom DAODs
  - ❑ lossy compression



Lossy compression

https://indico.cern.ch/event/1295808/contributions/5446232/attachments/2665301/4618464/DAODPHYS_LossyCompression.pdf

- ❑ In order to strengthen the specific characteristics of the PHYSLITE format, studies are ongong to :
  - ❑ study the minimum necessary variable precision to reduce the size after compression;
  - ❑ review and remove unnecessary content.
- ❑ Lossy float compression on DAOD_PHYSLITE can result in a significant reduction of the storage footprint by removing unnecessary bytes;



Original file [M]	48	-
High compressed file [M] *	34	- ~30%
Low compressed file [M] *	41	- 15%

- ❑ All these steps must be implemented very carefully to prevent loss of data information -> they must not have an impact on the final physics results and on the uncertainties;
- ❑ This is beneficial as long as the additional source of error introduced by the truncation/compression is not dominant over the observables' instrumental precision;
- ❑ Interactions with performance/analysis groups ongoing in order to identify this optimal limit.
- ❑ (more info here https://cds.cern.ch/record/2865166?ln=it for non lossy compression studies)

Eventindex/Eventpicking

https://cds.cern.ch/record/2870346/files/ATL-SOFT-PROC-2023-027.pdf

- The ATLAS experiment collects several billion of collisions from the LHC every year. In addition, simulated events are generated to compare the results of the analysis of real data with different physics models.
- All these data need to be catalogued in a large, high performance and high reliability system that can provide ATLAS members the possibility to search for and retrieve one or more individual events from the tens of millions of data files, in order to perform detailed checks, more refined analyses or produce event displays – the so-called “event picking” operations
- The EventIndex design started in 2013 and the first implementation was fully functional for the start of LHC Run 2 in 2015 : necessary to upgrade it in order to stand the expected higher data rates for Run 3 (2022 onwards) and beyond.
- The new EventIndex for LHC Run 3 have demonstrated improved efficacy in handling large amounts of data and accessibility for users.
- In march 2023 the data store contained 540 billion event records belonging to 280’000 datasets, occupying 47 TB in HBase.

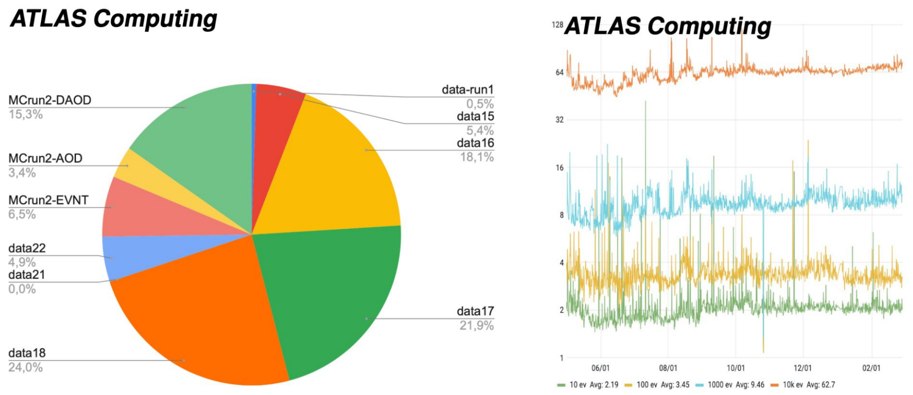


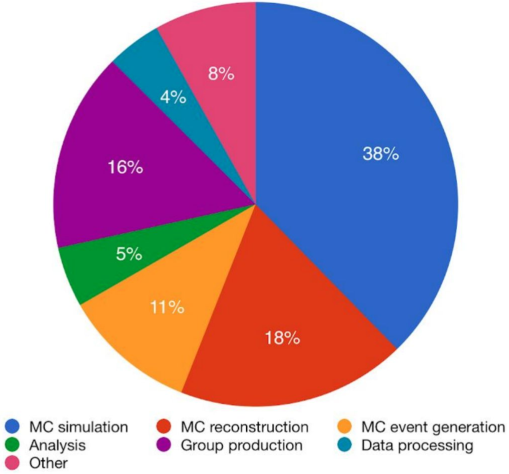
Figure 3. Left: fraction of event records stored by data type at the end of March 2023. Right: lookup times in seconds for queries for 10, 100, 1000 and 10’000 events from performance tests executed between May 2022 and February 2023.

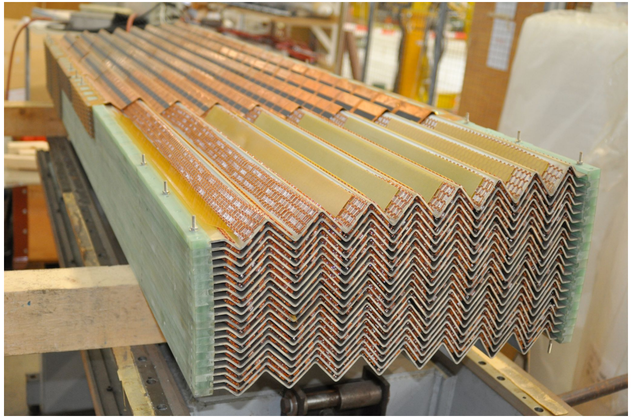
MC events production

ATL-SOFT-PROC-2021-017

Multipurpose experiments cover a wide ranging physics program from precision measurements to searches for new physics

- Monte Carlo events (both hard scatter and pile up) are functional to this process
- Typically the number of simulated MC events is ~ 2.5 the number of data events !
- Most of ATLAS CPU time used for MC detector simulation and ~80-90 of detector simulation time spent on calorimeters (complex geometries)

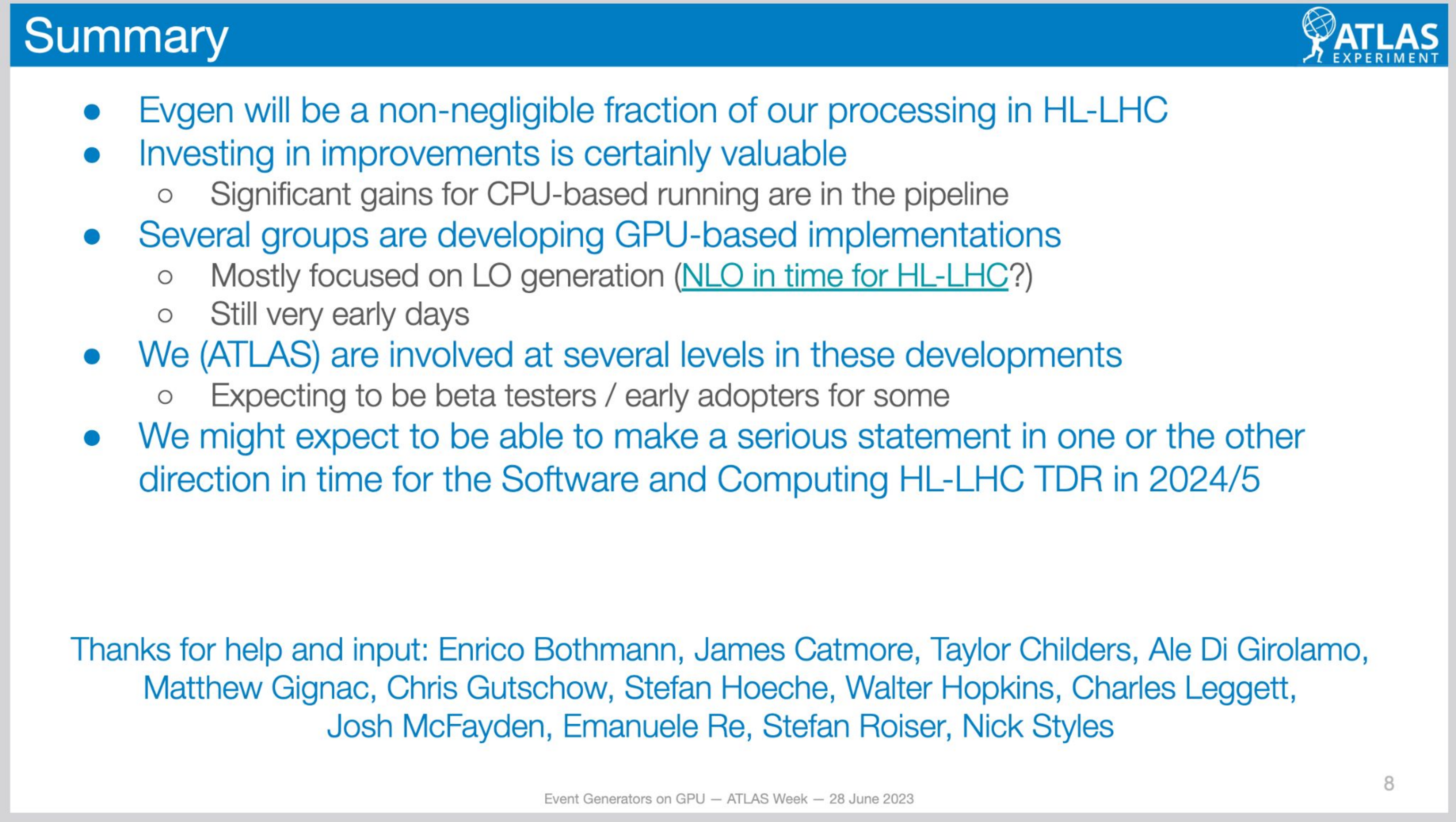




To reduce the impact of the preparation of MC events :

- Optimise G4 full simulation
- Pushing on fast (calo) simulation
  - Reduce simulation time keeping as much accuracy as possible + memory efficiency
  - Increase the number of analyses using FastSim : Run 3: >50% events with fast simulation, Run 4: >75% events with fast simulation
- Part of the full-simulation on accelerators (e.g. GPUs)

Event generation on GPU



Summary

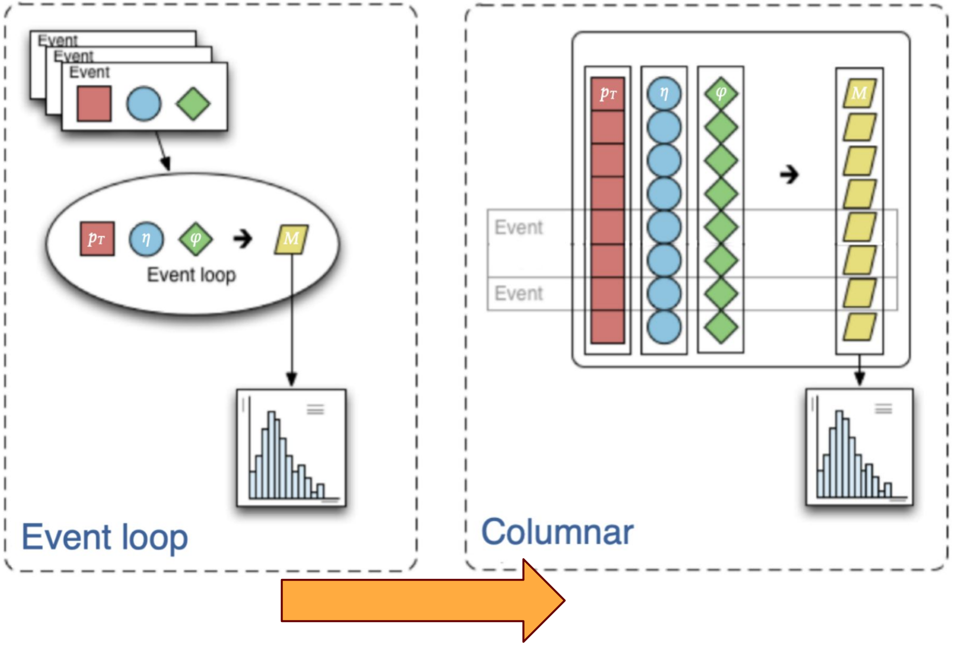
- Evgen will be a non-negligible fraction of our processing in HL-LHC
- Investing in improvements is certainly valuable
  - Significant gains for CPU-based running are in the pipeline
- Several groups are developing GPU-based implementations
  - Mostly focused on LO generation (NLO in time for HL-LHC?)
  - Still very early days
- We (ATLAS) are involved at several levels in these developments
  - Expecting to be beta testers / early adopters for some
- We might expect to be able to make a serious statement in one or the other direction in time for the Software and Computing HL-LHC TDR in 2024/5

Thanks for help and input: Enrico Bothmann, James Catmore, Taylor Childers, Ale Di Girolamo, Matthew Gignac, Chris Gutschow, Stefan Hoeche, Walter Hopkins, Charles Leggett, Josh McFayden, Emanuele Re, Stefan Roiser, Nick Styles

8

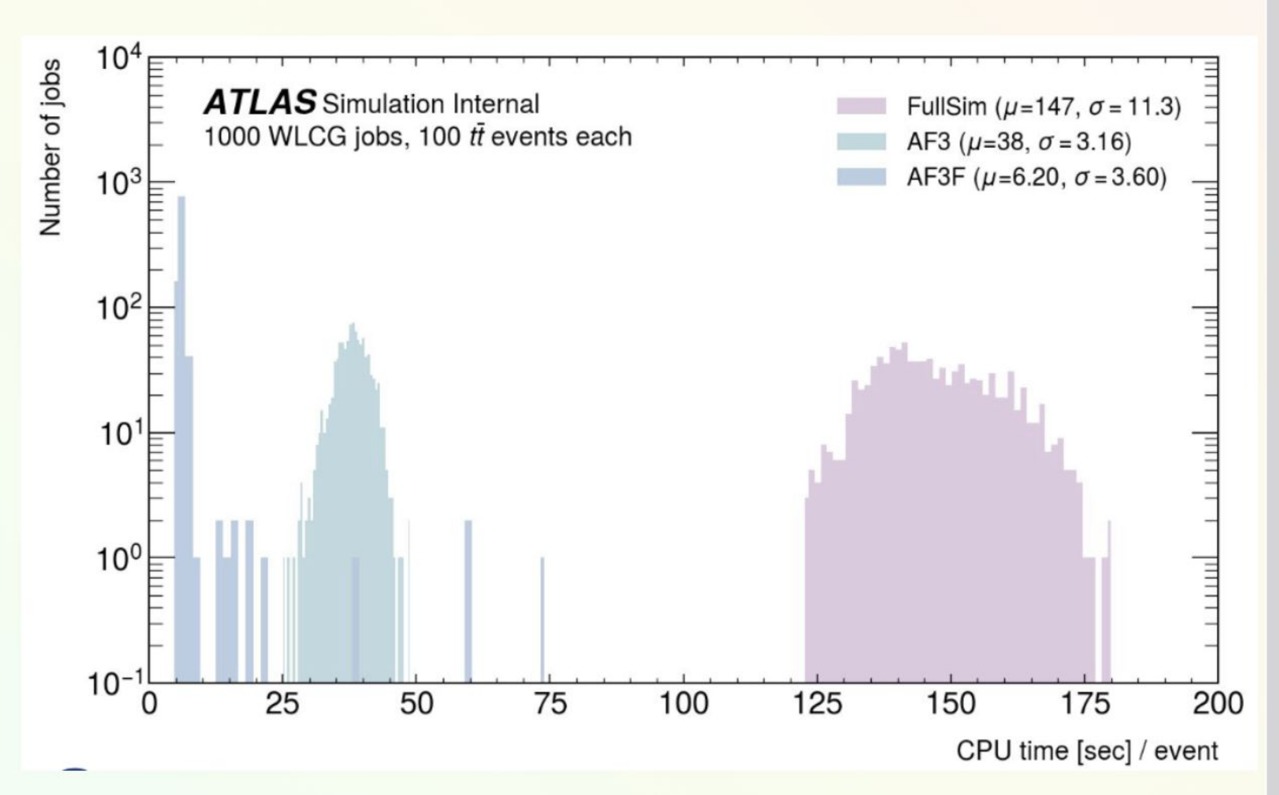
Event Generators on GPU — ATLAS Week — 28 June 2023

Analysis model(s) :



- event-wise analysis:
  - look at one event at a time
  - do all operations on single event
  - move on to next event
- columnar analysis:
  - perform one operation at a time
  - apply operation to many events
  - move on to next operation
- benefits of columnar analysis:
  - better code performance
  - integrate with industry standards and ML ecosystem/tools
- glue ATLAS CP tool to user-friendly python interface: demo exists using egamma cp tools
- incorporate CP tools for on-the-fly systematics directly from PHYSLITE reduce intermediate formats - exchange CPU for disc model

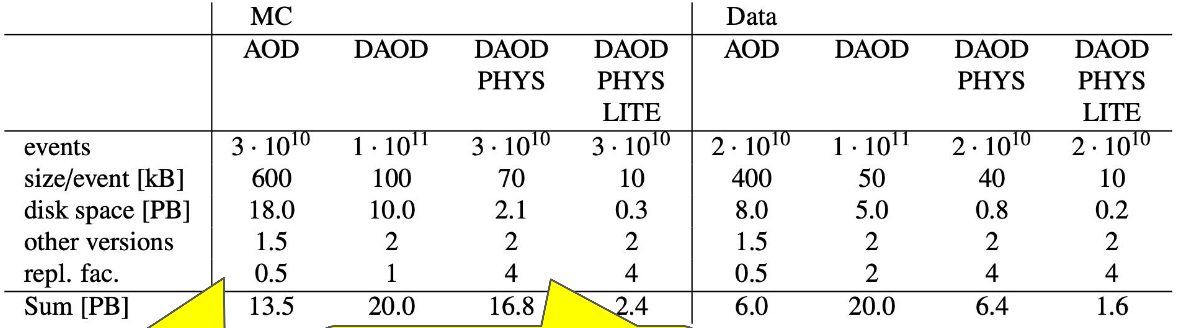
Simulation(s)



A new analysis model

Dalla presentazione 2022 a Pisa

- ❑ Introduce instead a new single DAOD_PHYS targeted for all (>~80%) physics analysis (~50 kB/event).
- ❑ For the HL-LHC a new smaller DAOD_PHYSLITE format (10 kB/event) that contains already calibrated physics objects and will be centrally produced with frequent updates, typically every few months.
  - ❑ Computing management pushing to consider DOAD_PHYSLITE as THE analysis format (ie avoid conversion to ntuples or other). QT on columnar analysis on DAOD_PHYLITE
- ❑ Allow exceptions for performance groups, B-physics (separate stream), long lived particle searches....



	MC				Data			
	AOD	DAOD	DAOD	DAOD	AOD	DAOD	DAOD	DAOD
			PHYS	PHYS			PHYS	PHYS
				LITE				LITE
events	$3 \cdot 10^{10}$	$1 \cdot 10^{11}$	$3 \cdot 10^{10}$	$3 \cdot 10^{10}$	$2 \cdot 10^{10}$	$1 \cdot 10^{11}$	$2 \cdot 10^{10}$	$2 \cdot 10^{10}$
size/event [kB]	600	100	70	10	400	50	40	10
disk space [PB]	18.0	10.0	2.1	0.3	8.0	5.0	0.8	0.2
other versions	1.5	2	2	2	1.5	2	2	2
repl. fac.	0.5	1	4	4	0.5	2	4	4
Sum [PB]	13.5	20.0	16.8	2.4	6.0	20.0	6.4	1.6

50% of AOD on tape

4 replicas of derived data formats, 2 versions kept

- ❑ Run2 AM requires 132 PB
- ❑ Run3 AM would require ~85 PB

- ❑ The new model opens to the possibility of the creation of Analysis Facilities (few PB of disk space):
  - ❑ <u>common resources and storage for full DAOD_PHYSLITE in Italy ?</u>