東京大学
素粒子物理国際研究センター
International Center for Elementary Particle Physics
The University of Tokyo

Challenges in Collider physics

Future of Artificial Intelligence for Science in Japan (FAIRS Japan 2024)

3rd Dec 2024

ICEPP/UTokyo

Masahiko Saito

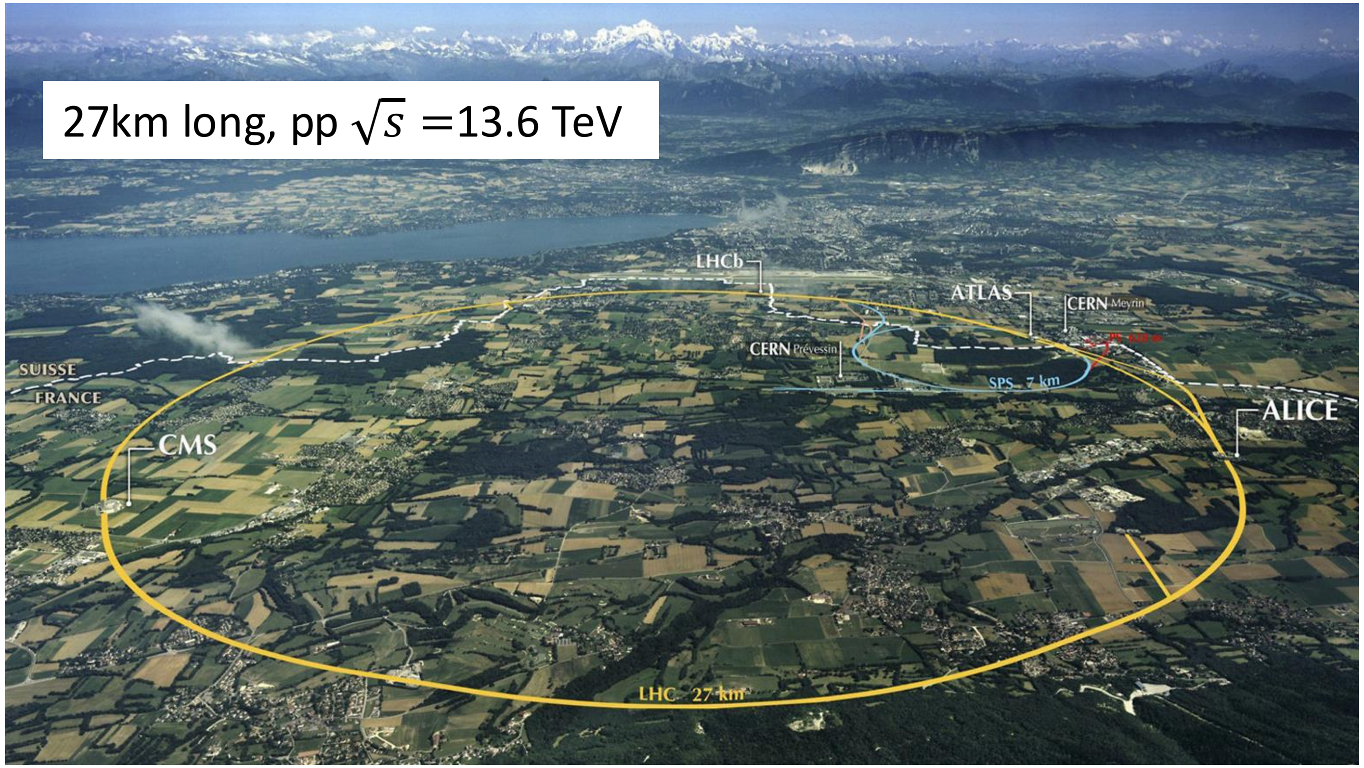
About me and this talk

Computing Cluster in ICEPP/UTokyo

- **Background**
  - PhD on the search for **supersymmetric particles** at the ATLAS experiment in 2019.
  - Now working on operating of a **distributed computing cluster** for the ATLAS experiment and developing **machine learning** for collider physics.

- **This talk**
  - Talk about the **challenges** in collider physics.
    - But focusing only on **ATLAS/LHC**. And highly biased by my interests!
    - Many ML/AI applications in ATLAS/LHC! But they are out of the scope of this talk.

Large Hadron Collider (LHC)

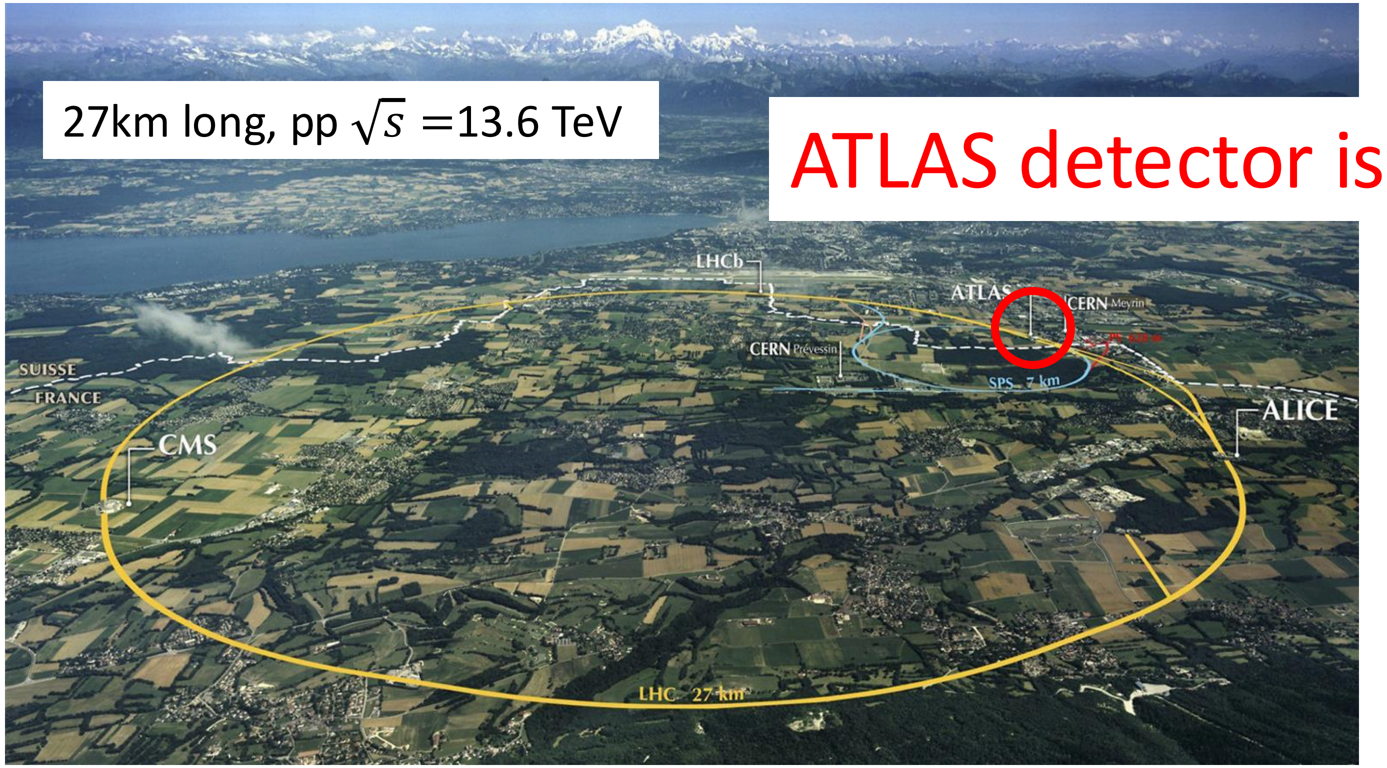


- **<u>Collider physics</u>** : Study the behavior of elementary particles by artificially creating a high-energy state by the collider.
- The most powerful collider in the world is the **LHC (Large Hadron Collider)**
  - In operation since 2008. Discovered Higgs boson in 2012.

Large Hadron Collider (LHC)

27km long, pp $\sqrt{s}$ =13.6 TeV

ATLAS detector is located here

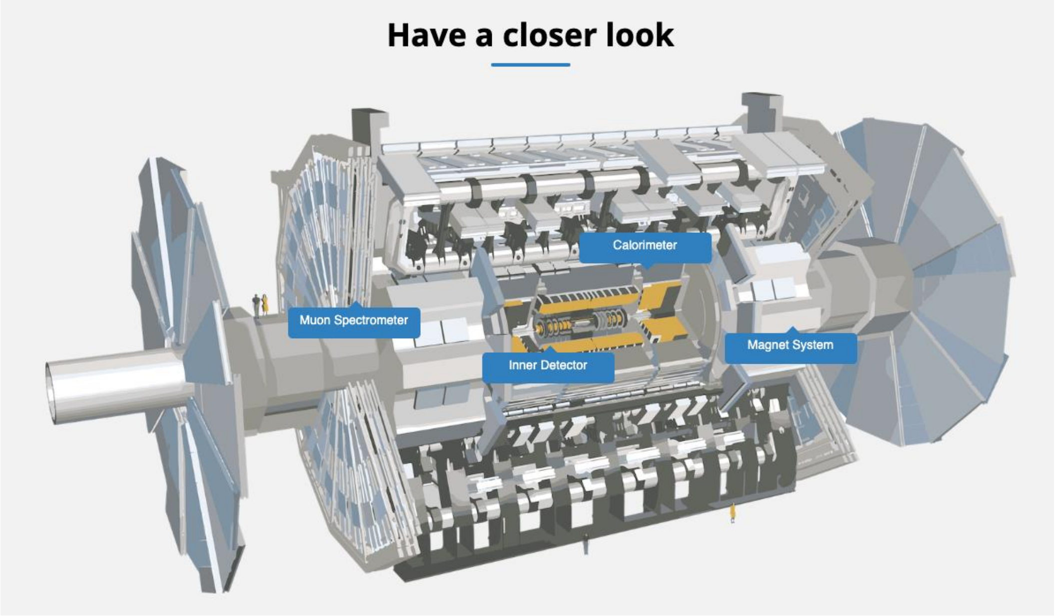


- **<u>Collider physics</u>** : Study the behavior of elementary particles by artificially creating a high-energy state by the collider.
- The most powerful collider in the world is the **LHC (Large Hadron Collider)**
  - In operation since 2008. Discovered Higgs boson in 2012.

ATLAS experiment

- ATLAS collaboration: **6000** members from **42** countries. (~150 members from Japan)
- ATLAS detector: **multipurpose** detector

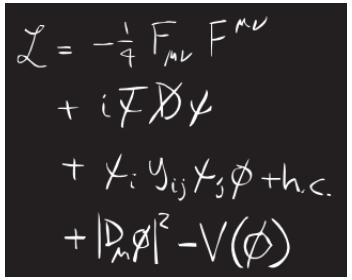
- <u>Major scientific milestone</u>
  - Discovery of Higgs boson ← **Achieved!**
  - **Precise measurements** of the Standard Model: Top quark, EW boson, **Higgs boson,** …
  - **Discovery** of beyond the Standard Model: Dark matter, Supersymmetry, Heavy Higgs, …

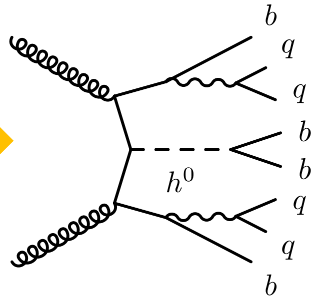


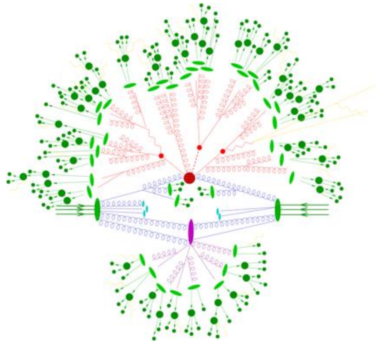
ATLAS website

Production pipeline

Nature



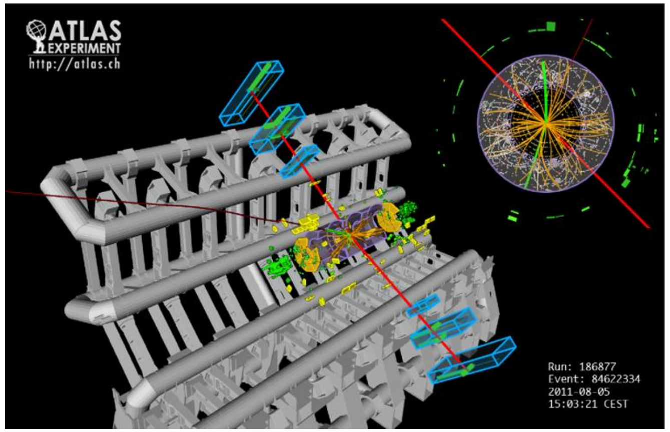




Theory

Parton-level O(10)

Stable particle O(100)



Sensor output O(10^8)

Trigger

40 MHz

Hardware Trigger

100 kHz

Software Trigger

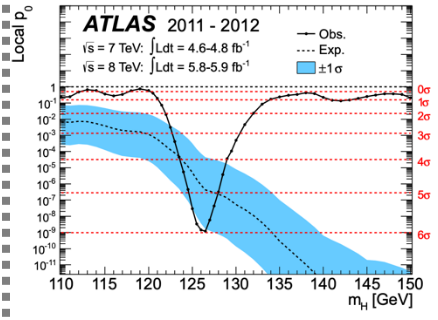
3 kHz

8 GB/s

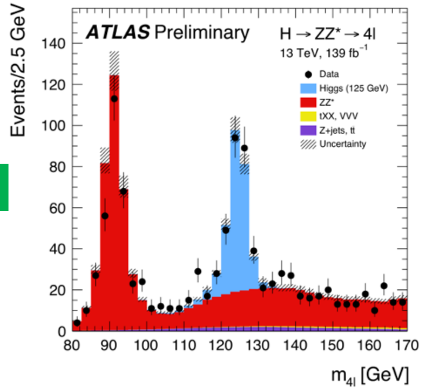
Distribute data to data centers
ATLAS has 1 EB data
(and 600k CPU cores)

Offline analysis

Hypothesis testing

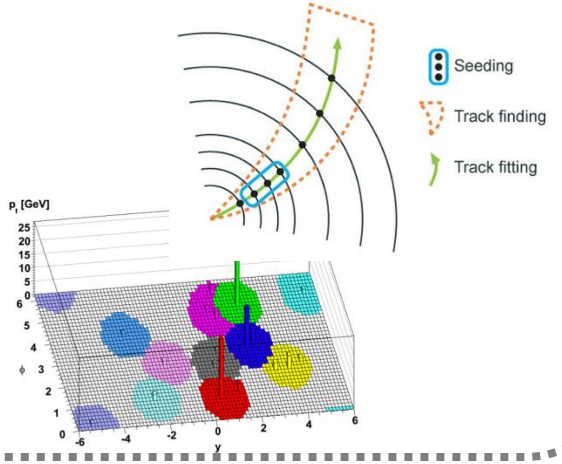


Event reconstruction



Combine four lepton 4-vectors

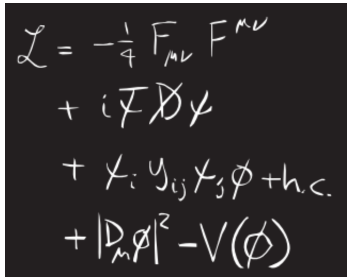
Particle reconstruction
Particle identification



Offline computing

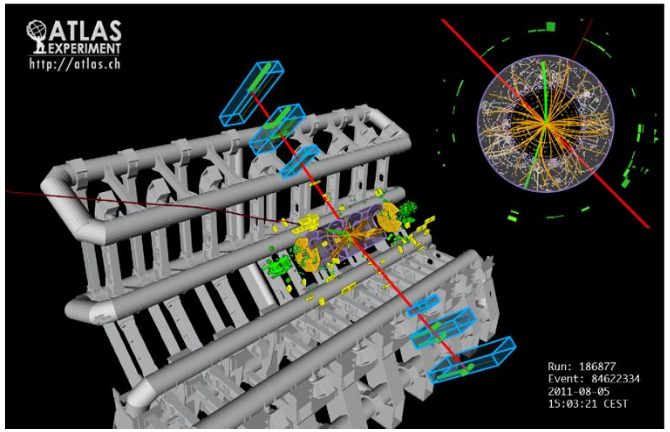
Production pipeline

Simulation



Theory

Event generation
Detector simulation



Sensor output $O(10^8)$

Trigger

40 MHz

Hardware Trigger

100 kHz

Software Trigger

3 kHz

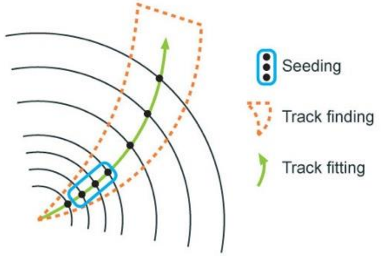
8 GB/s

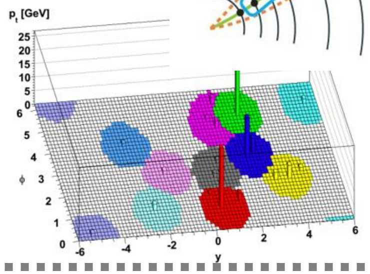
Distribute data to data centers
ATLAS has 1 EB data (and 600k CPU cores)

Offline analysis

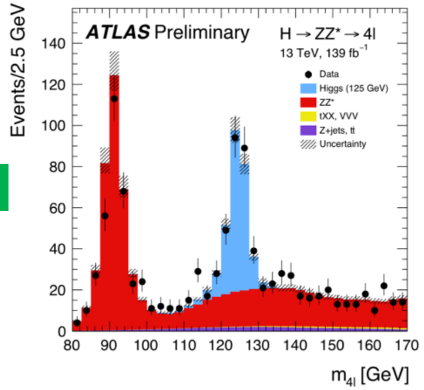
Offline computing

Particle reconstruction
Particle identification



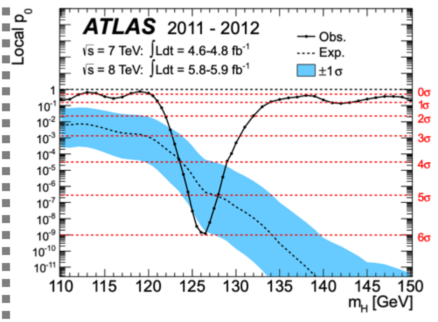


Event reconstruction



Combine four lepton 4-vectors

Hypothesis testing

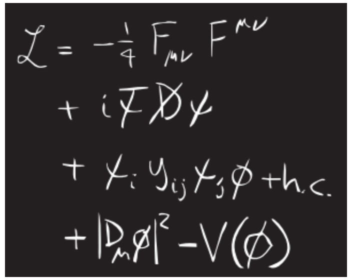


Production pipeline

Detector simulation:
a few min / event

Luminosity: $2 \times 10^{34}\text{cm}^{-2}s^{-1}, \langle\mu\rangle \sim 40$

Simulation

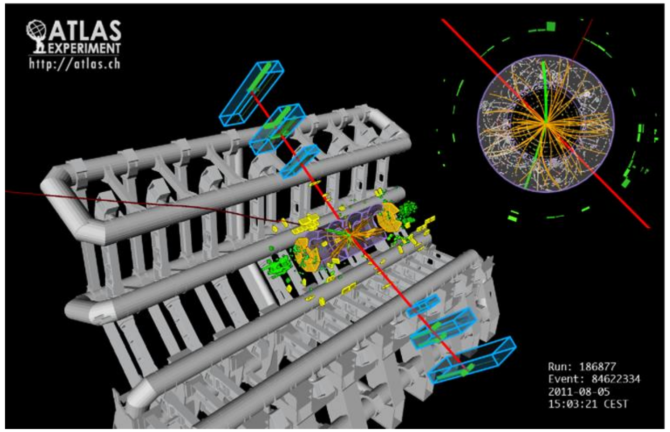


SHERPA

GEANT4
A SIMULATION TOOLKIT

Theory

Event generation
Detector simulation



Sensor output **O(10^8)**

Trigger

40 MHz

Hardware Trigger

100 kHz

Software Trigger

3 kHz

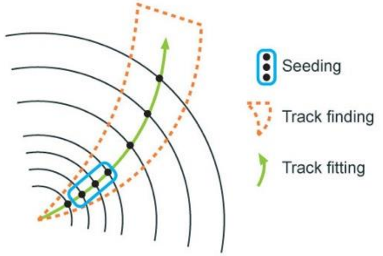
8 GB/s

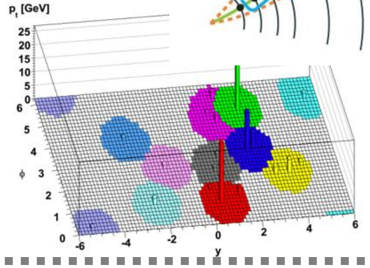
Distribute data to data centers
ATLAS has **1 EB** data
(and 600k CPU cores)

Offline analysis

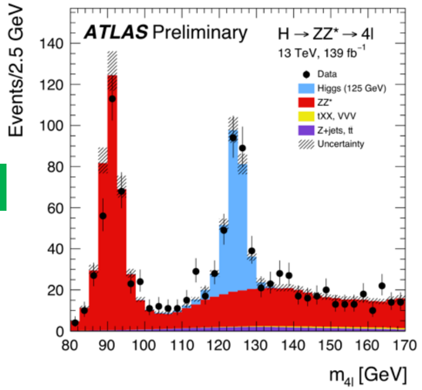
Offline computing

Particle reconstruction
Particle identification



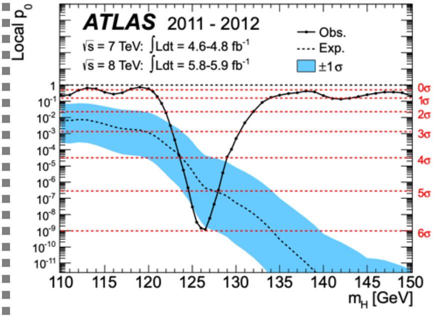


Event reconstruction

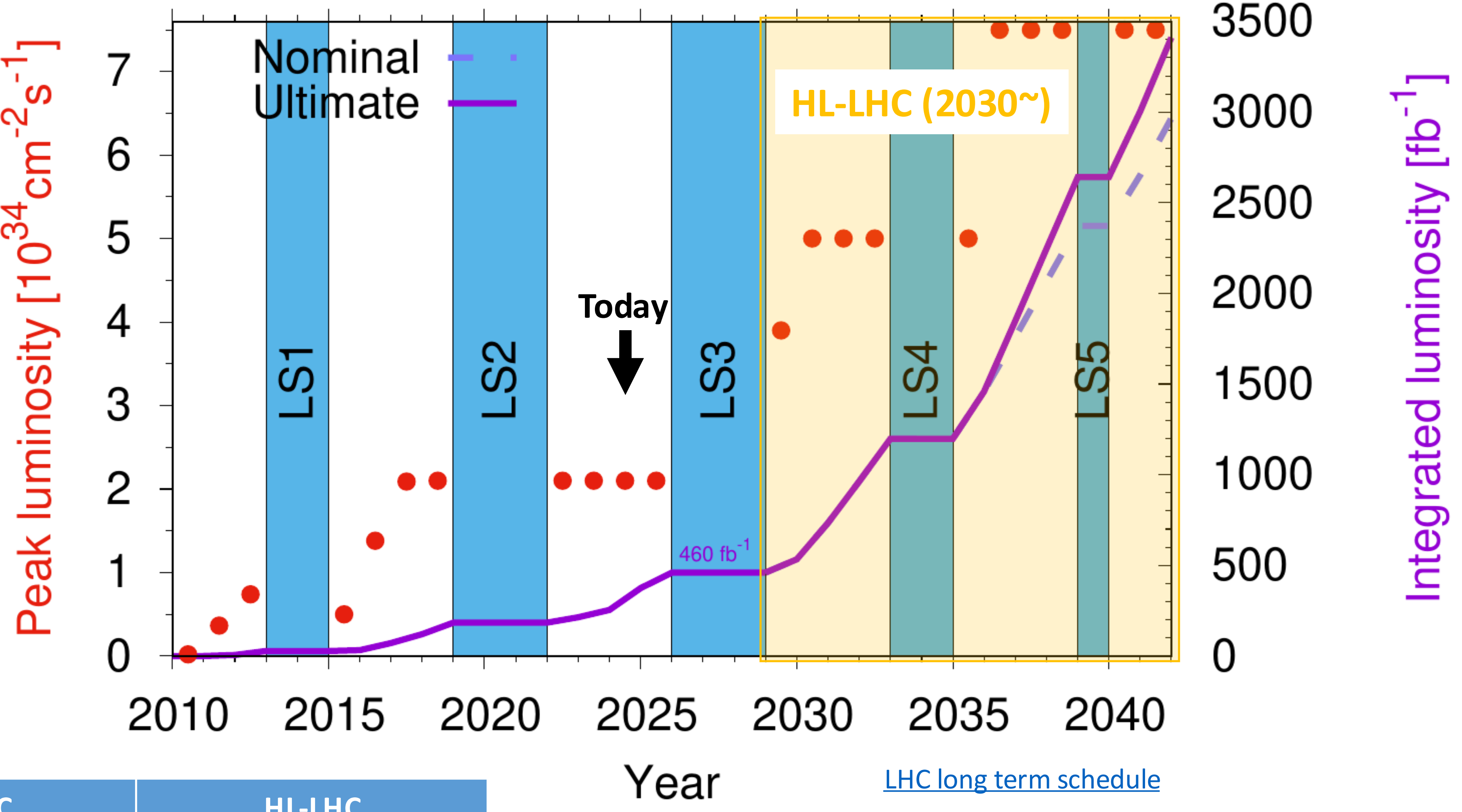


Combine four lepton 4-vectors

Hypothesis testing



Upgrade plan: High Luminosity LHC (HL-LHC)

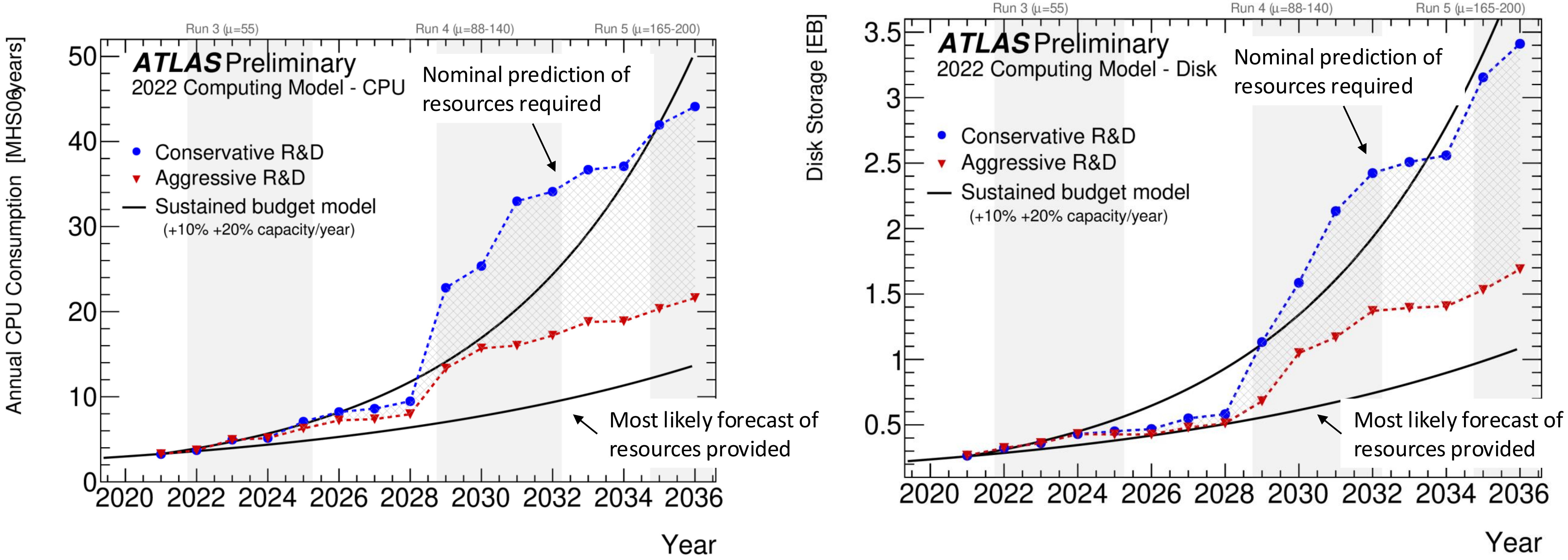


LHC long term schedule

	LHC	HL-LHC
Peak lumi	$2 \times 10^{34} \mathrm{cm}^{-2} s^{-1}$	$\mathbf{7.5 \times 10^{34} cm^{-2} s^{-1}}$
$\langle \mu \rangle$	~40	**~200**
Int. lumi	~450 fb^{-1}	**~3500 fb^{-1}**

Challenges for the HL-LHC: Computing resources

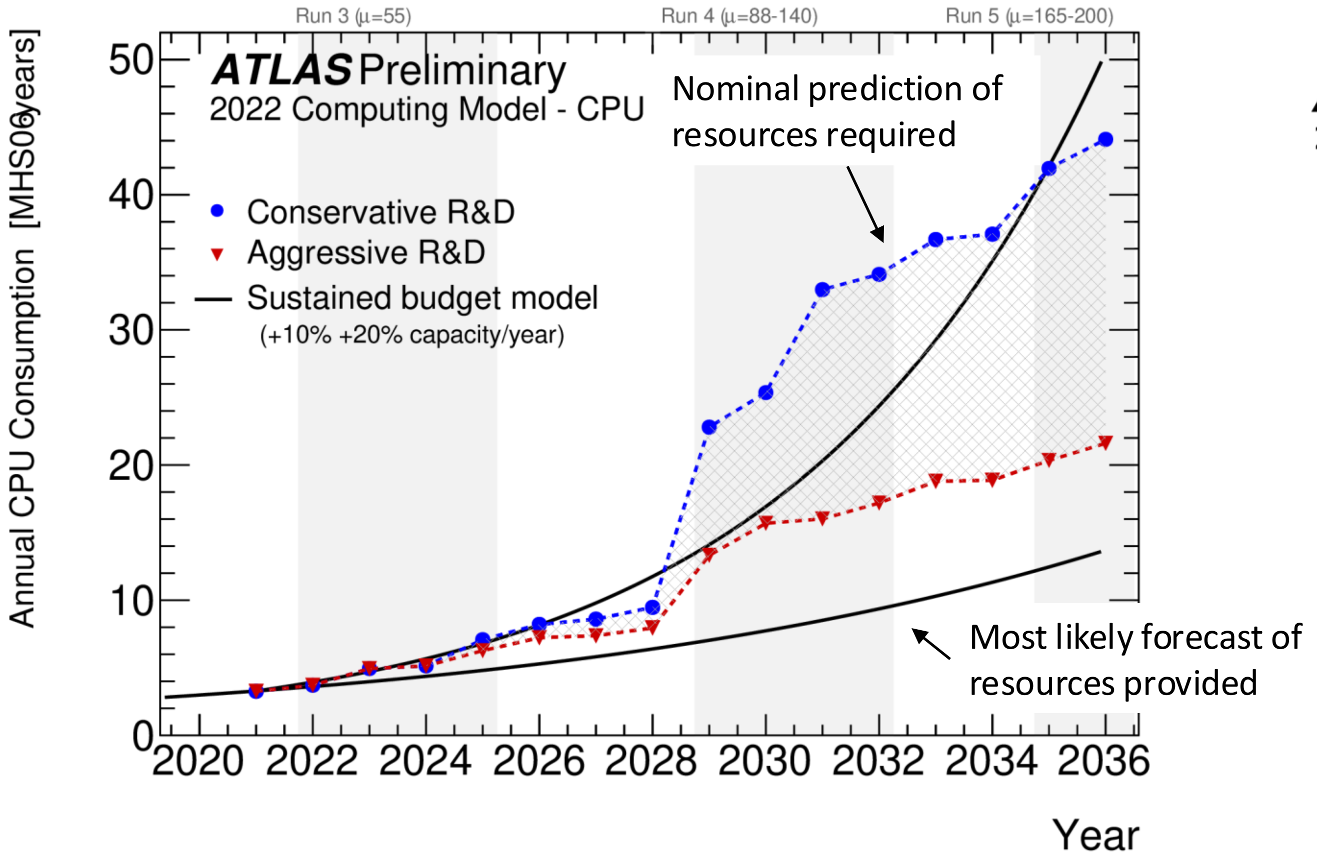
ATLAS S&C HL-LHC Roadmap

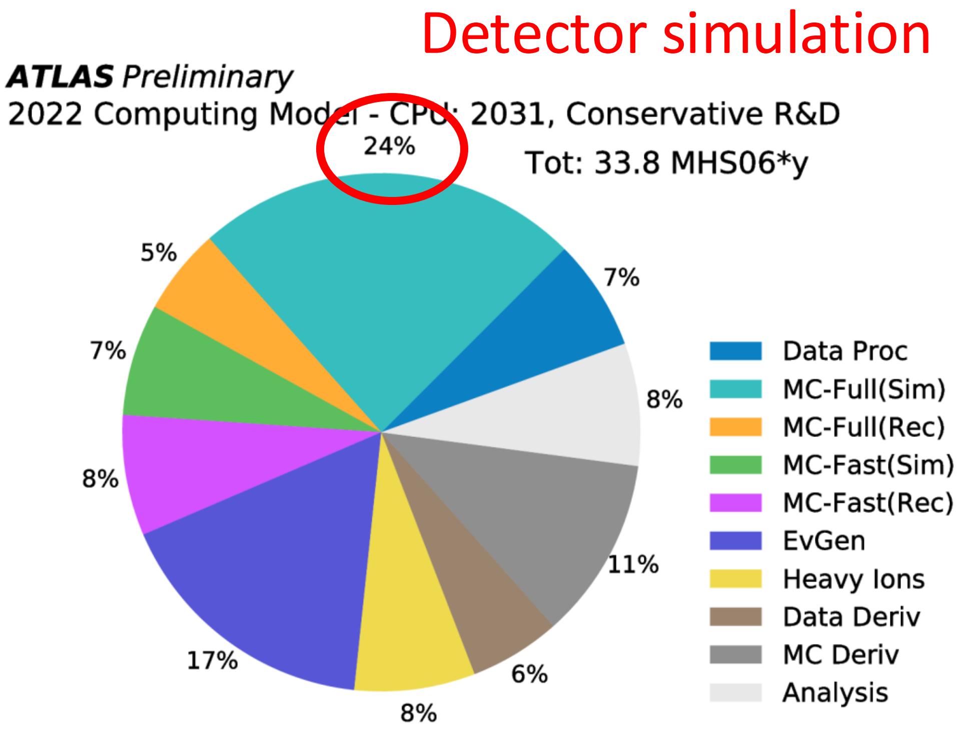


- Due to the increase in **experimental data** and the **difficulty of data processing**, **computing resources will be short.**
  → could lead to **a deterioration in scientific output**

Challenges for the HL-LHC: Computing resources

ATLAS S&C HL-LHC Roadmap

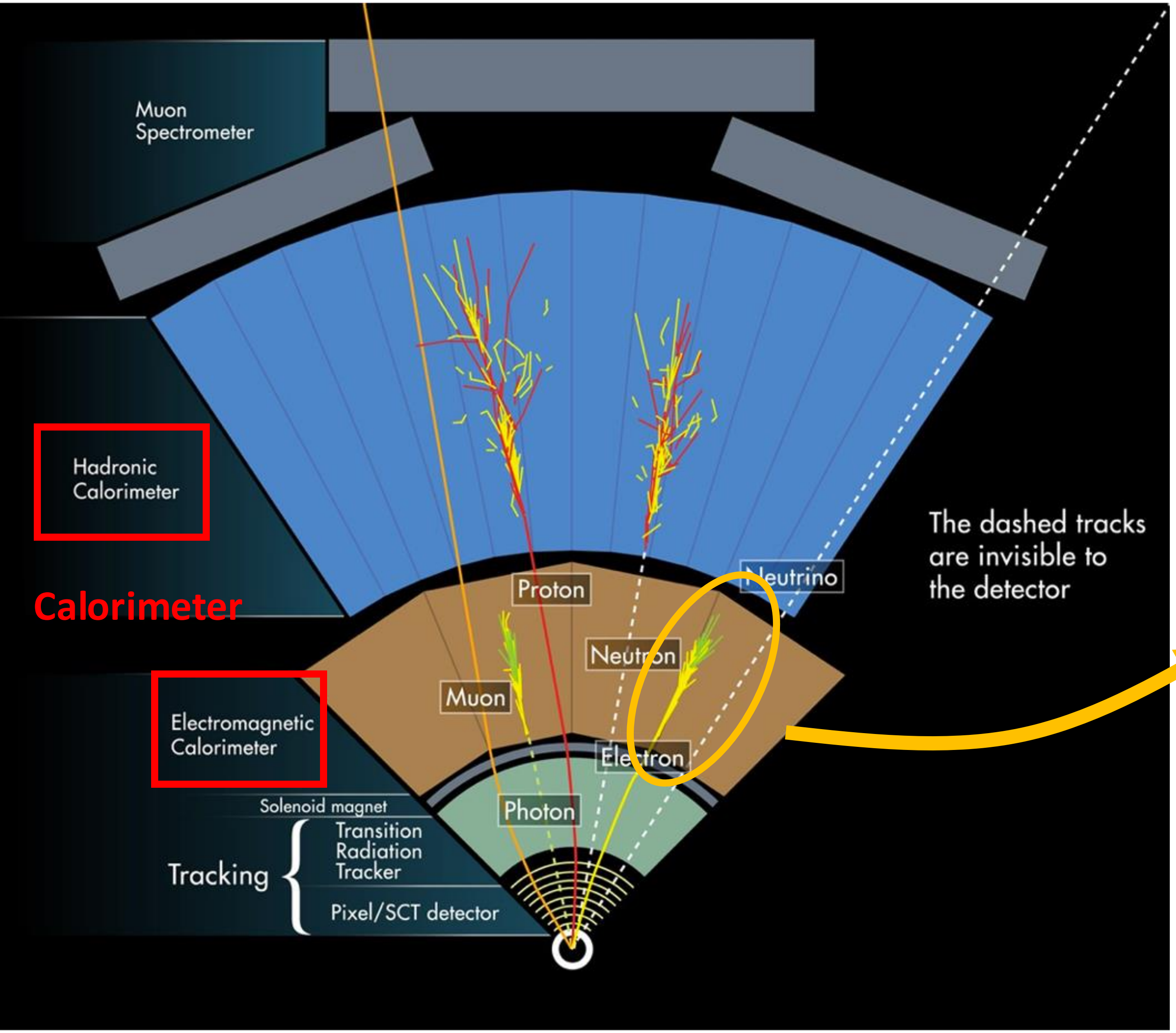




- Due to the increase in **experimental data** and the **difficulty of data processing**, **computing resources will be short.**
  → could lead to **a deterioration in scientific output**

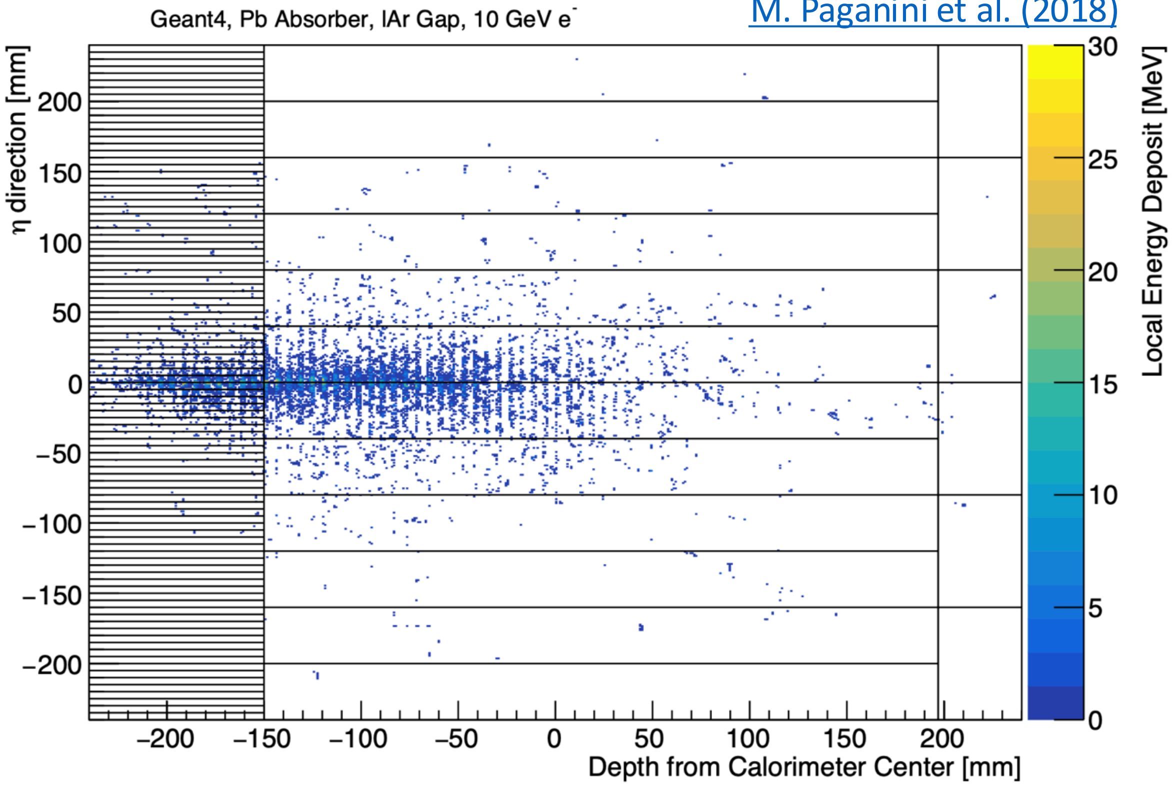
Challenges for the HL-LHC: Detector simulation

Particle interaction in detector



Shower in calorimeter

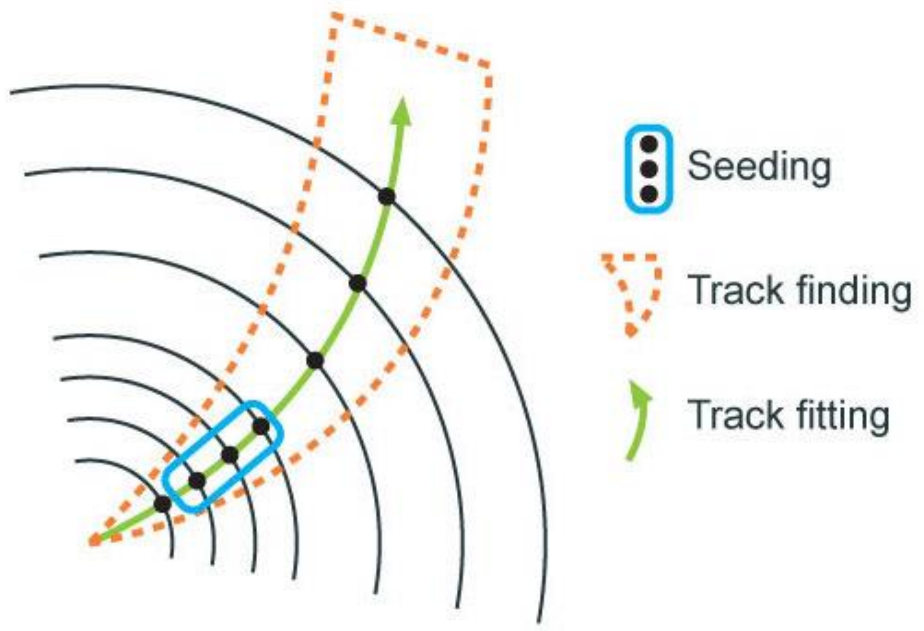
M. Paganini et al. (2018)



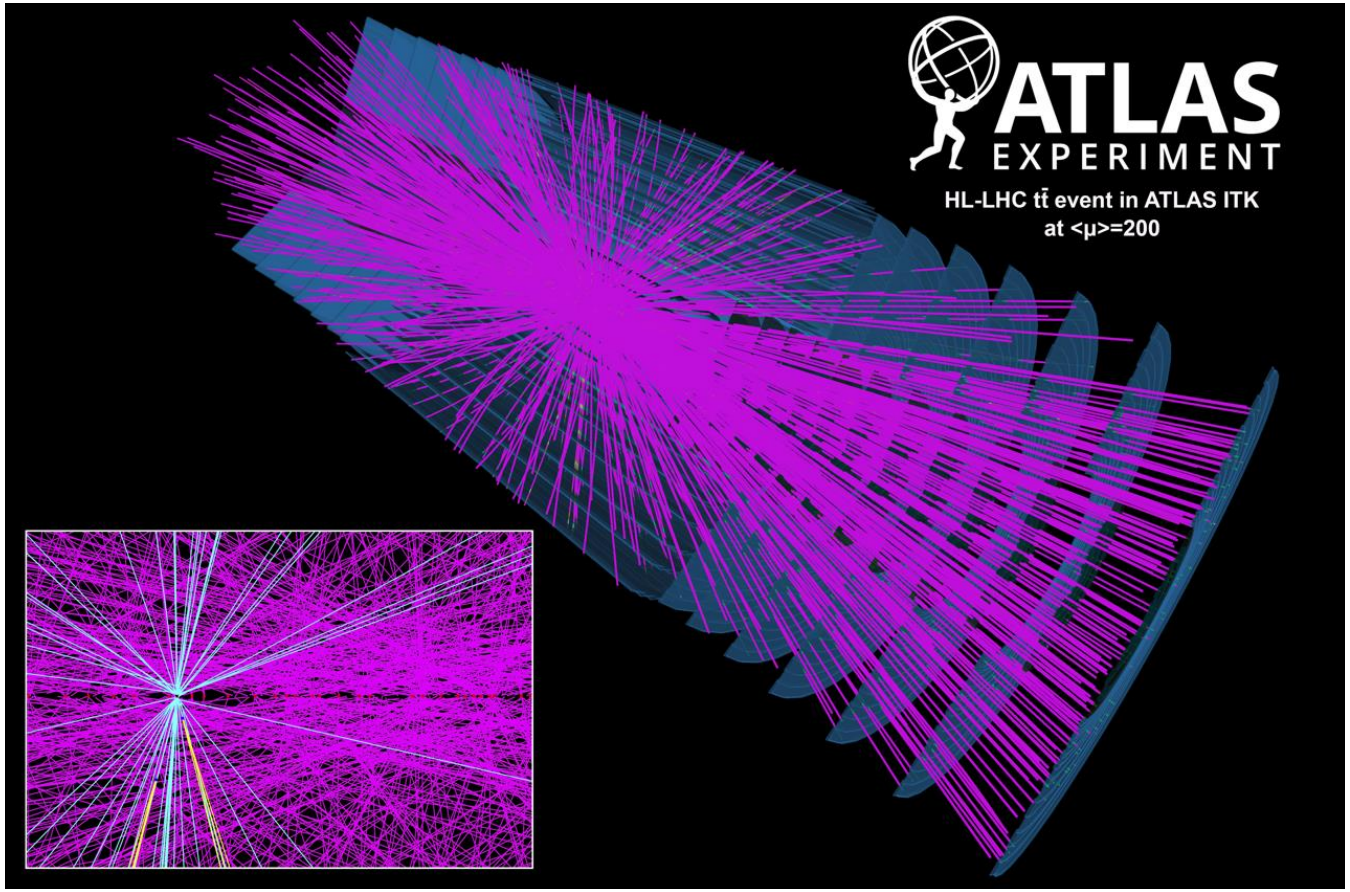
- Particle energy is measured in a calorimeter where it produces a **particle shower**
- Particle shower is simulated by Geant4. But it's computationally expensive.
- **ML-based fast simulation is an active topic**

Challenges for the HL-LHC: Track reconstruction (tracking)

ATLAS-PHOTO-2019-020

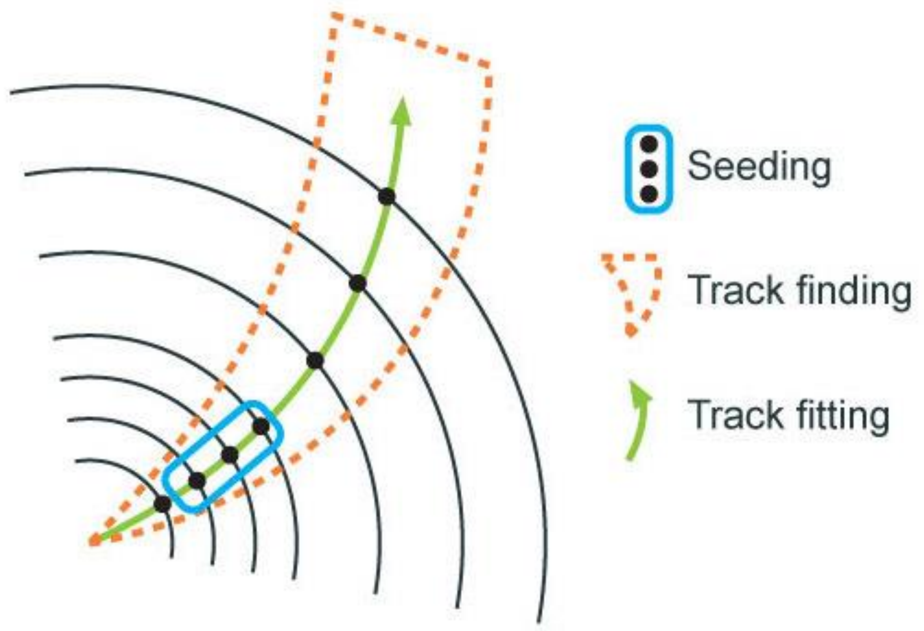


ATLAS SW doc



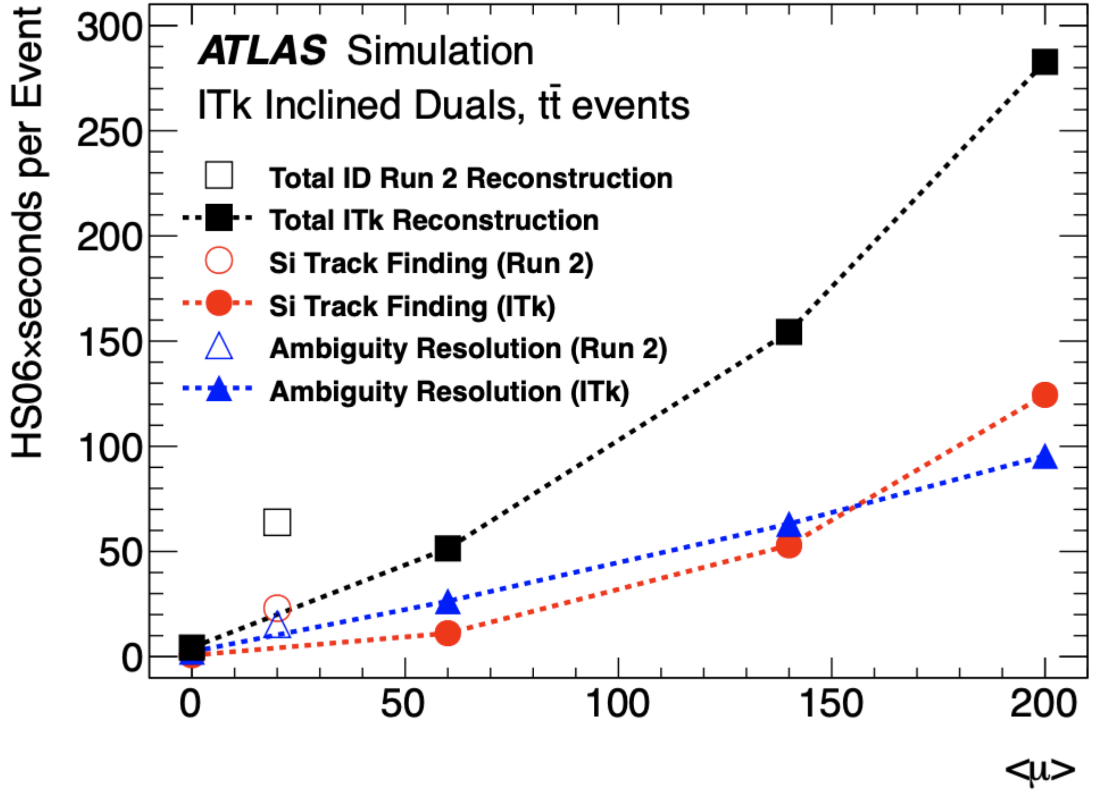
- Track reconstruction is a process of reconstructing the trajectories left by the charged particles by connecting the activated sensor space points.

Challenges for the HL-LHC: Track reconstruction (tracking)



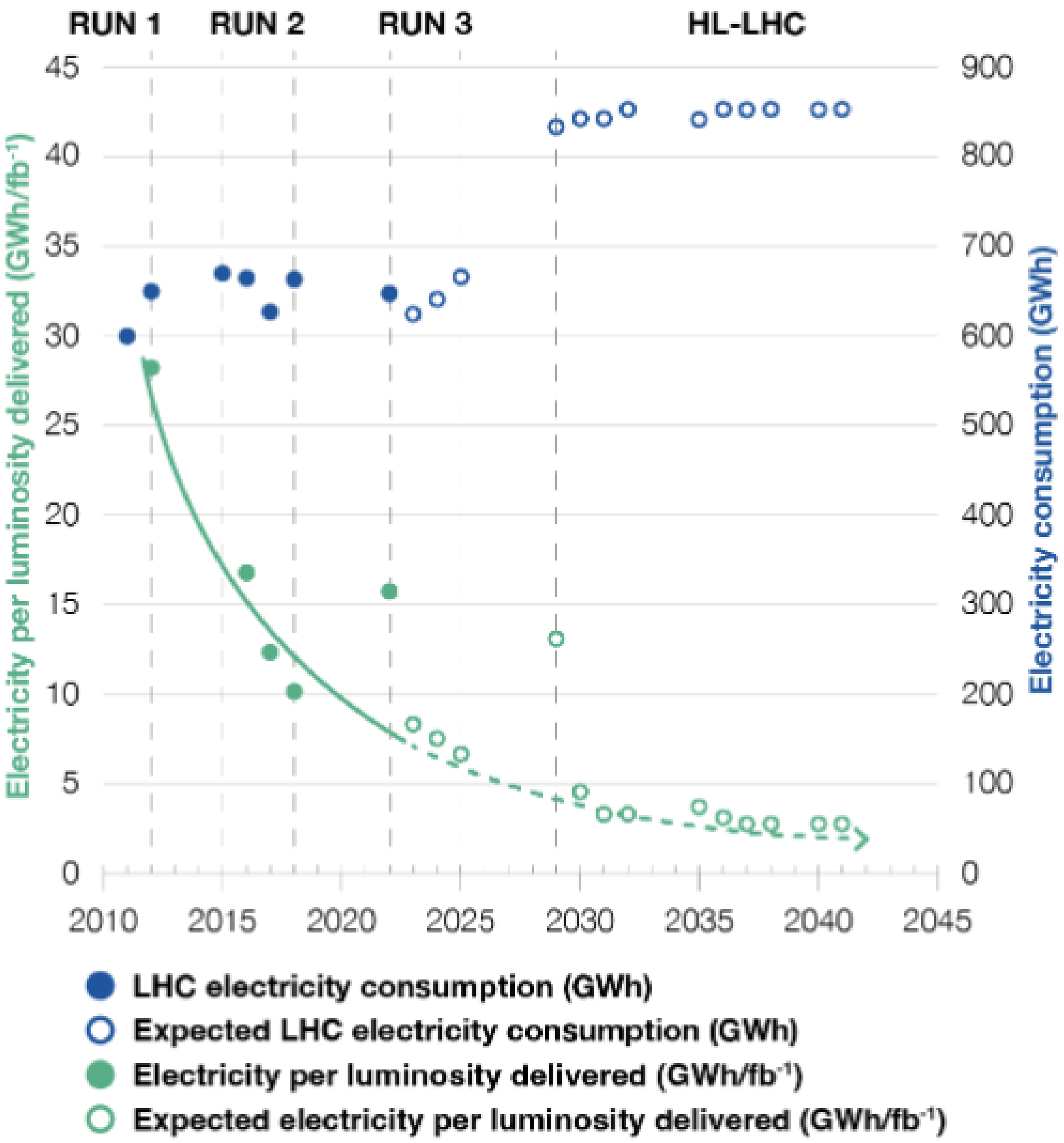
ATLAS SW doc

CERN-LHCC-2017-021



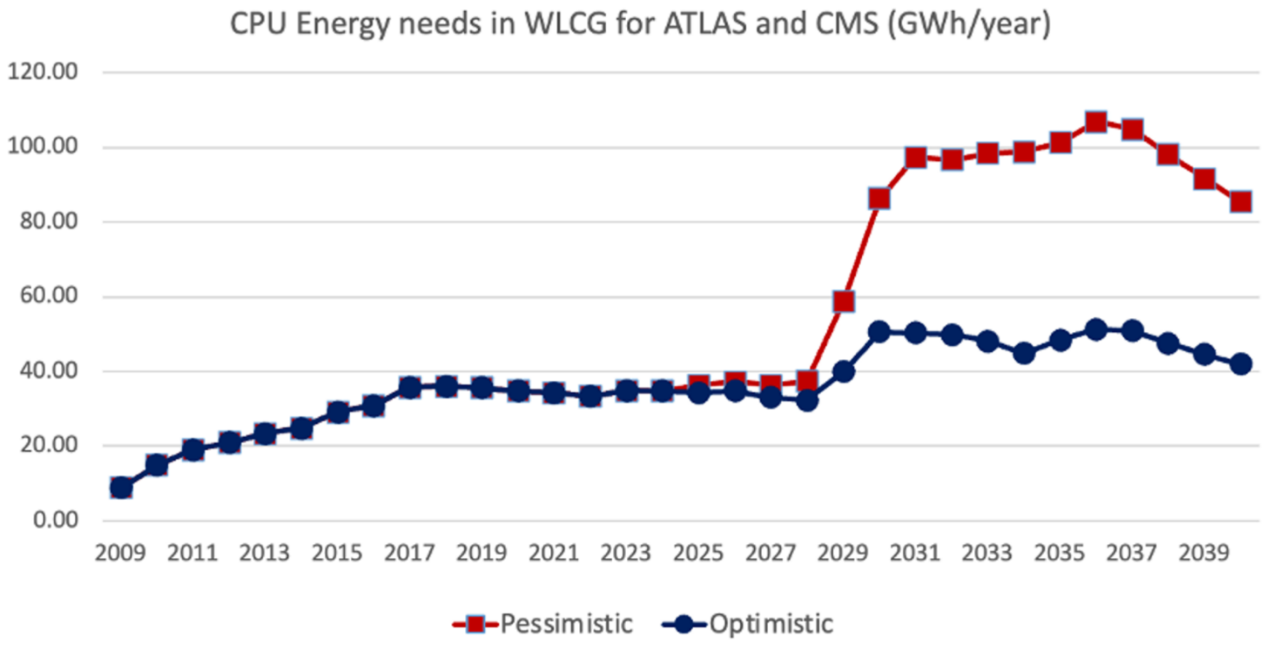
- Track reconstruction is a process of reconstructing the trajectories left by the charged particles by connecting the activated sensor space points.
- Track reconstruction is strongly affected by **pileup** → **increases of the computation time.**
  - because track seeds consist of three combinatorial triplets.

Challenges for the HL-LHC: Electricity consumption



CERN Environment Report

D. Britton, et.al., CHEP2023



- **Power consumption** needs to be considered.
- Very large deep learning models (e.g. LLM) consume a lot of power.

Challenges for application/expansion of ML/AI (from my interests)

Domain adaptation

- Simulations are **not** perfect.
- The more powerful model tends to exploit the incompatibility.
- It is necessary to bridge the gap between simulations and data.

Explainability

- Demand to know why the AI predicts/generates them. It is necessary to verify the process.
- **Uncertainty**

Model agnostic search / Anomaly detection

- New physics might be beyond the scope of human thinking.
- We need a search that does not rely on the specific signal models.

Integrate multiple tasks into a single model

- The pipeline is made up of several tasks. The overall prediction could be improved if they were used as a differentiable chain.
- How to balance multiple losses is one of the issues.

Summary

- This talk focuses on the **challenges of LHC/ATLAS**.
- ATLAS handles **huge** and **complex** experimental data with **fast** and **sophisticated** processing.
- The upgraded collider, **HL-LHC**, will start in 2030. It will bring more difficult challenges. There are many areas that can be solved or mitigated by **AI/ML**.
  - detector simulation, tracking, power consumption, …
  - trigger, DAQ, particle reconstruction/identification, event generation, …