Brookhaven National Laboratory

U.S. DEPARTMENT OF ENERGY

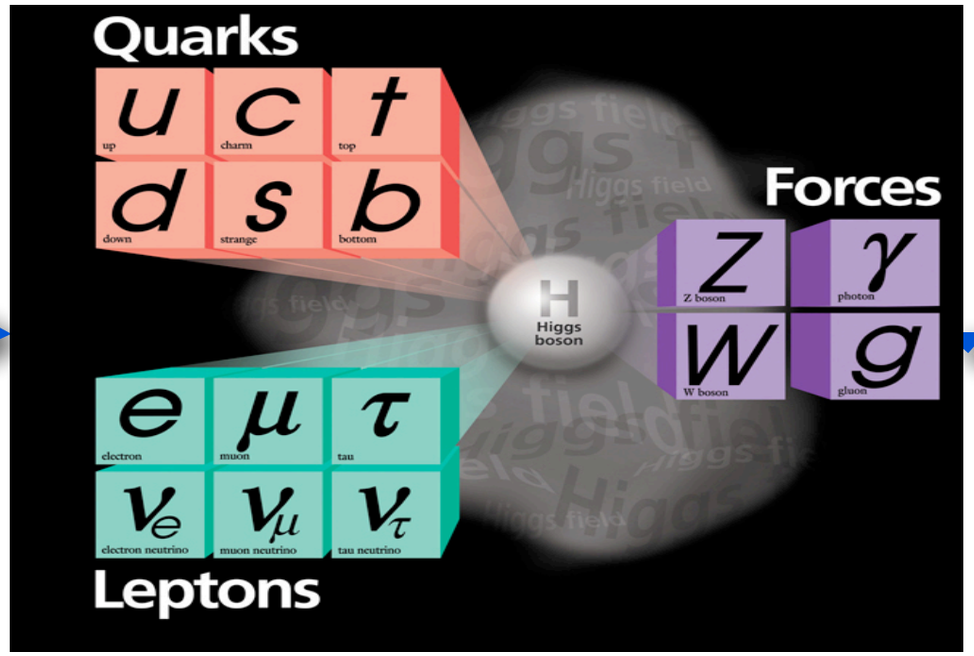
From Raw data to Physics Results (1/3)

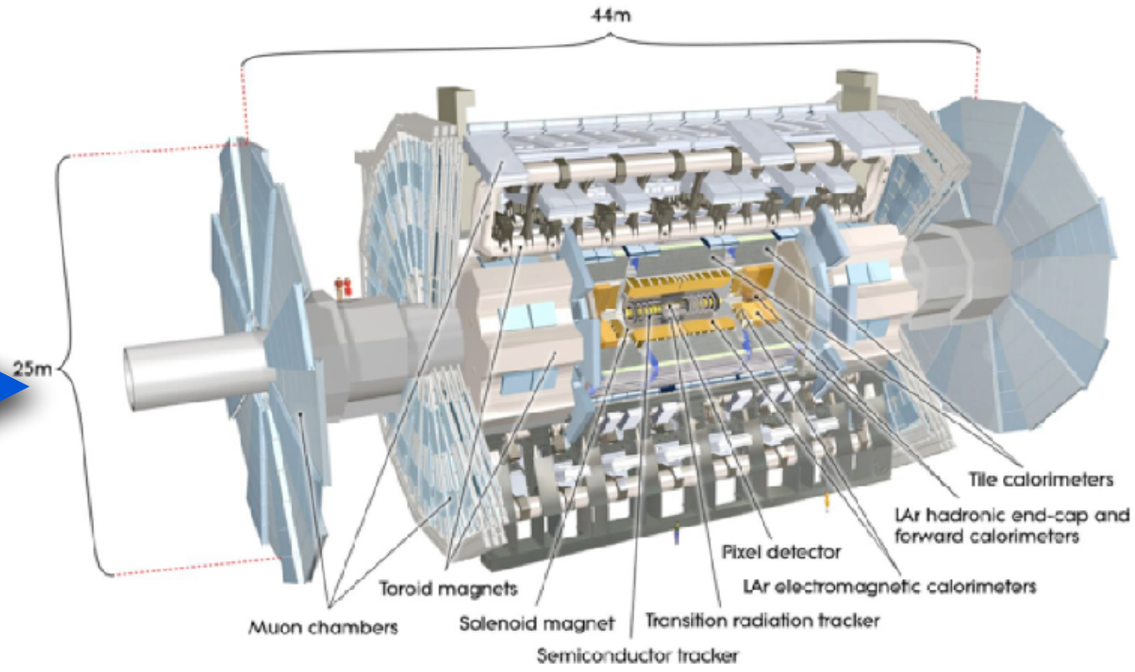
Paul Laycock

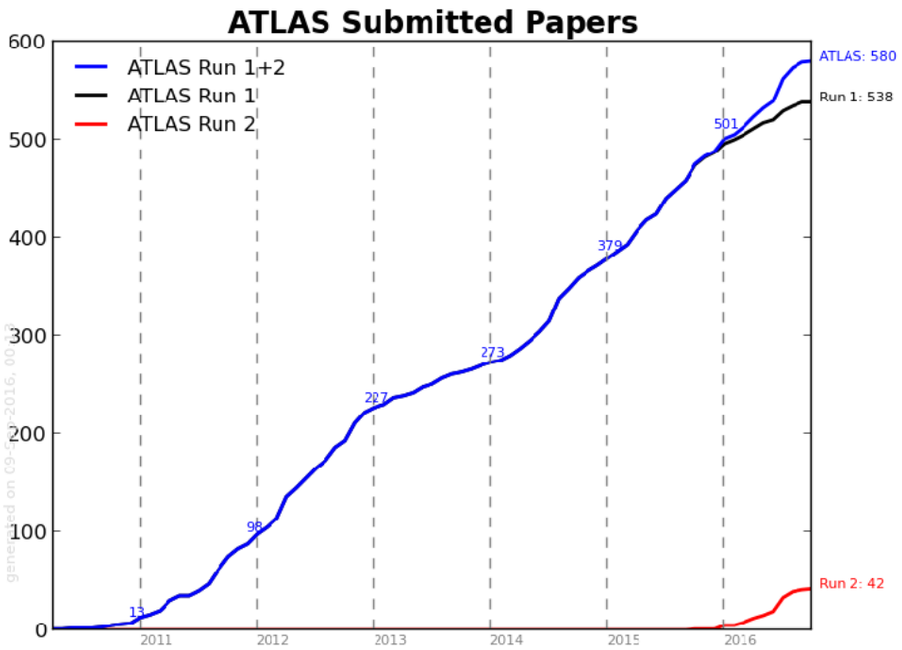
July 4th 2022

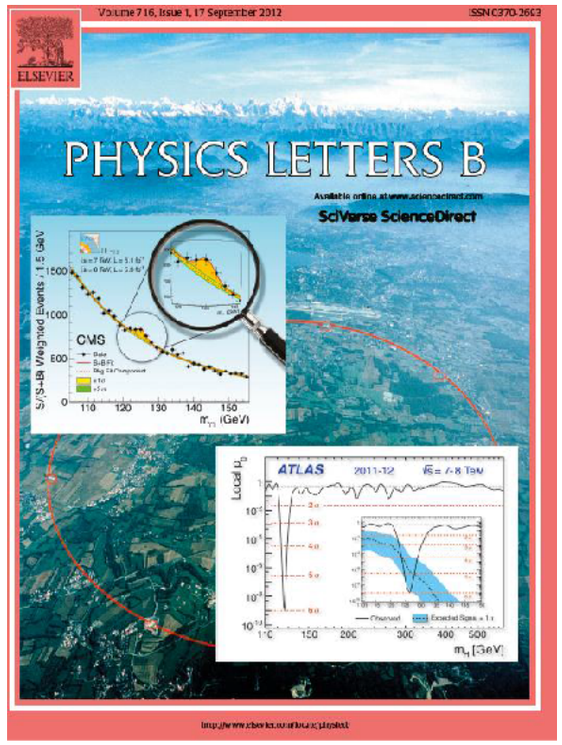
@BrookhavenLab

The particle physics cycle





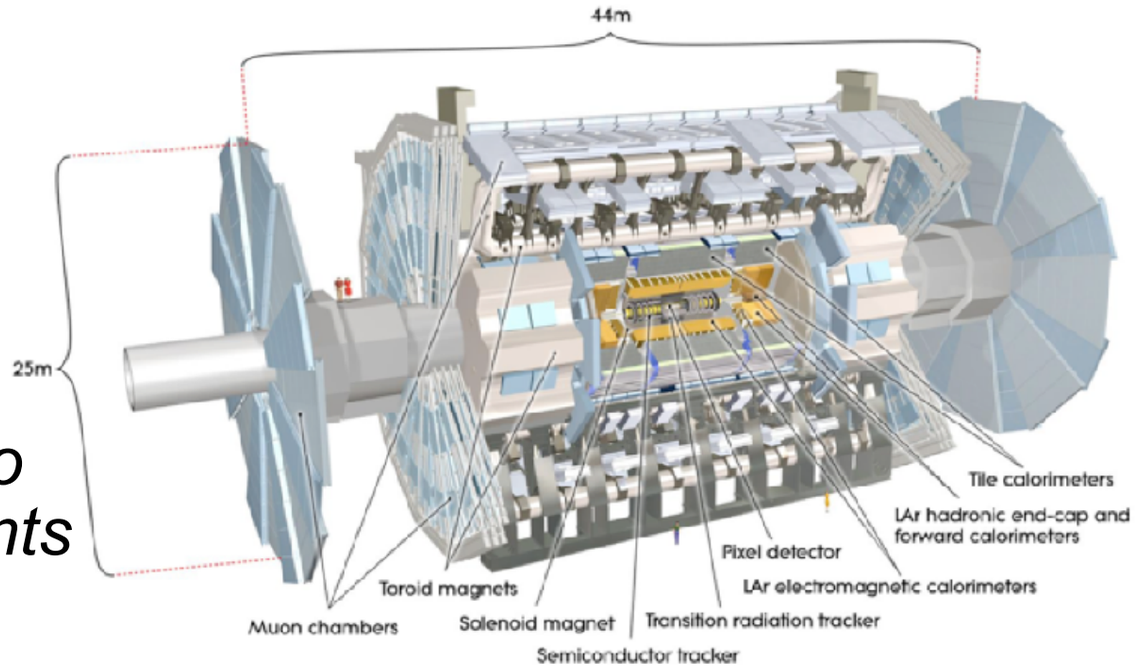




Experimental physics

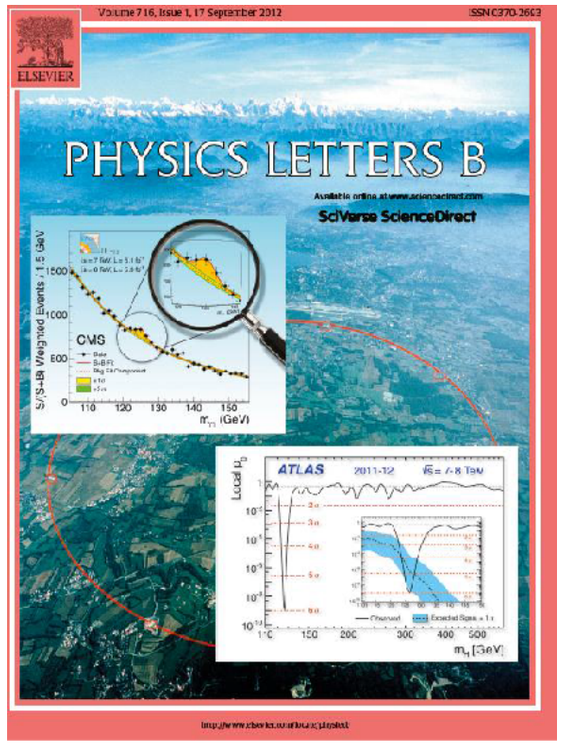
Much of the work of the experimental physicist is running experiments and extracting measurements from them

Note *- Experimental physicists also need to propose, design and build new experiments (see previous slide)*



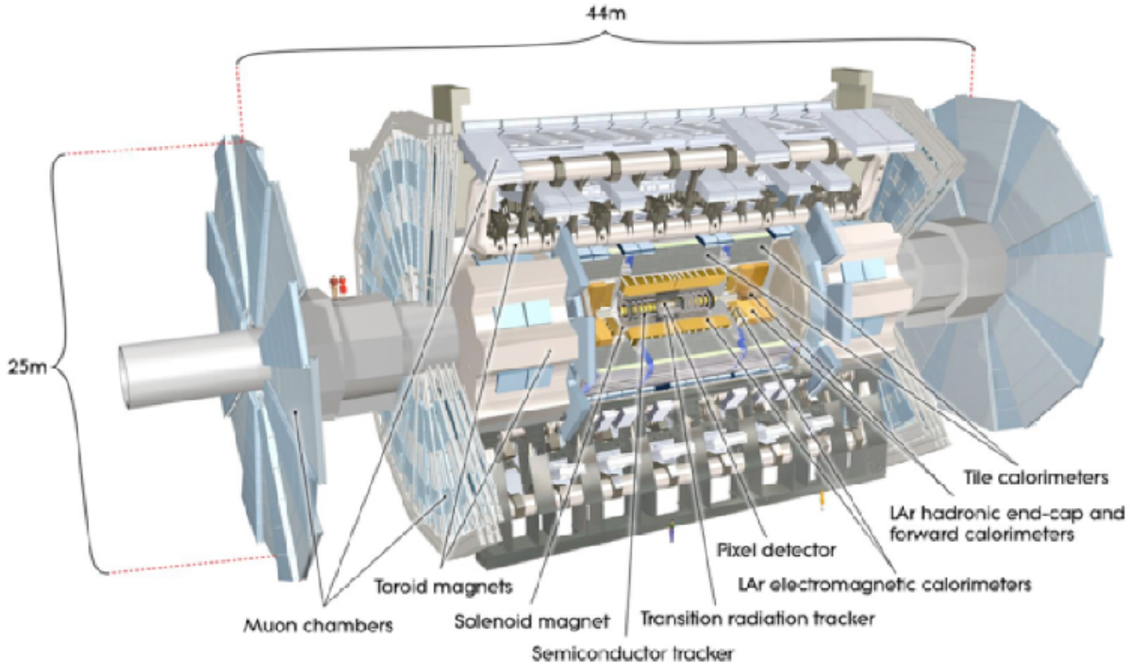
These lectures are focused on understanding how we turn raw experimental detector data into physics results that we can publish

- Results must be ***accurate***
- with well understood ***precision***
- It's important to understand the difference between these two words, we often confuse them



Brookhaven National Laboratory

Course outline



Lecture 1

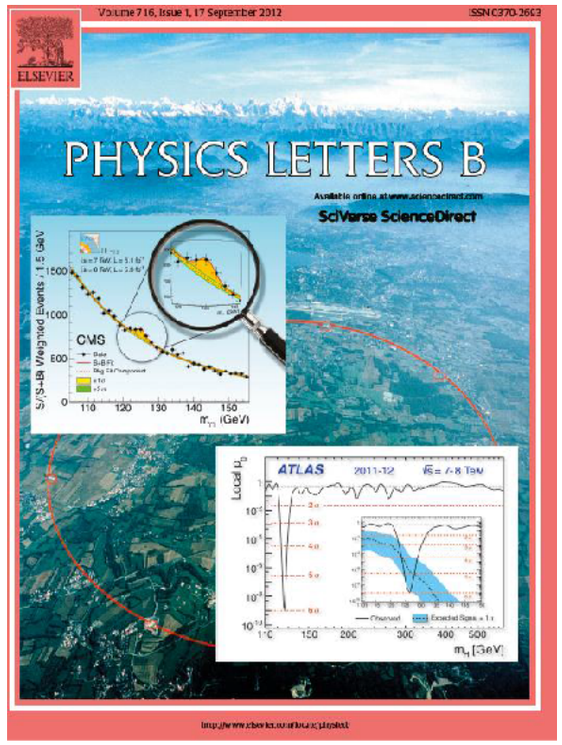
- The journey of raw data from the detector to a publication

Lecture 2

- How we reconstruct fundamental physics processes from raw detector data

Lecture 3

- How we extract our signals from the mountain of data, finding needles in the haystack



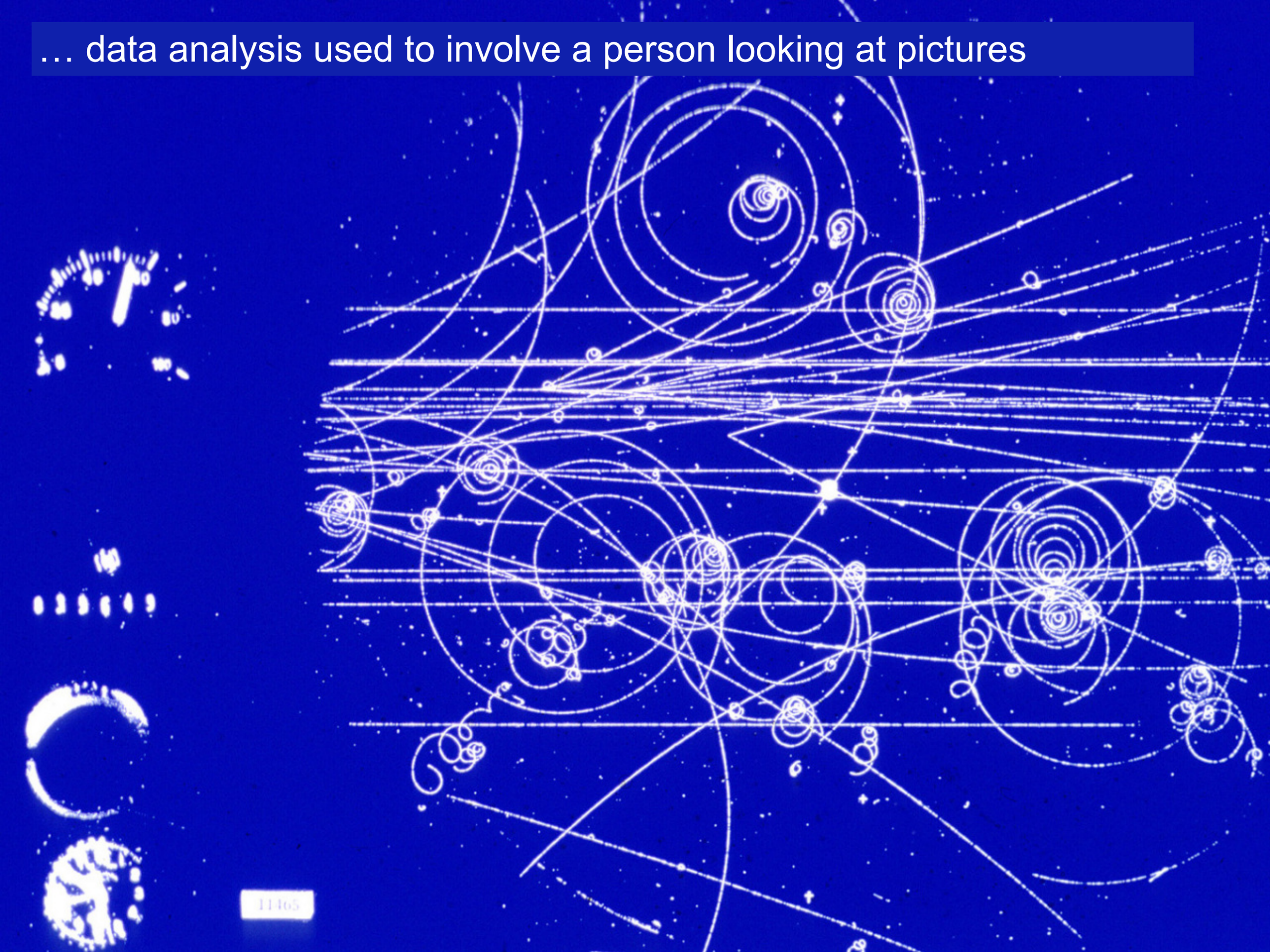
Brookhaven National Laboratory

Experiments at CERN

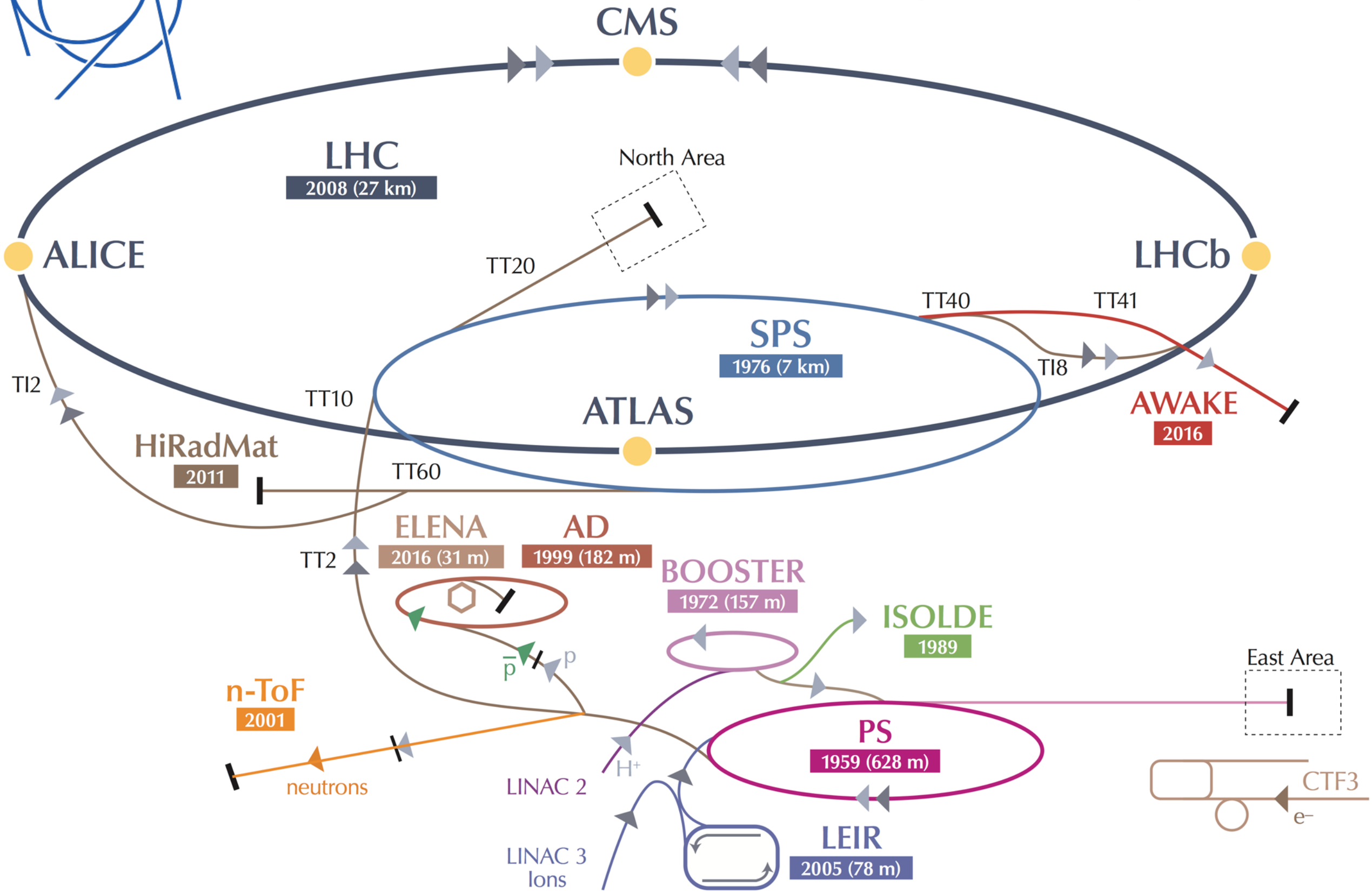
In the 1960s we used Bubble chambers, the one that you can see in the Microcosm was used…

Brookhaven National Laboratory

… data analysis used to involve a person looking at pictures



Today @ CERN we have huge rates of collisions so that we can produce very rare events



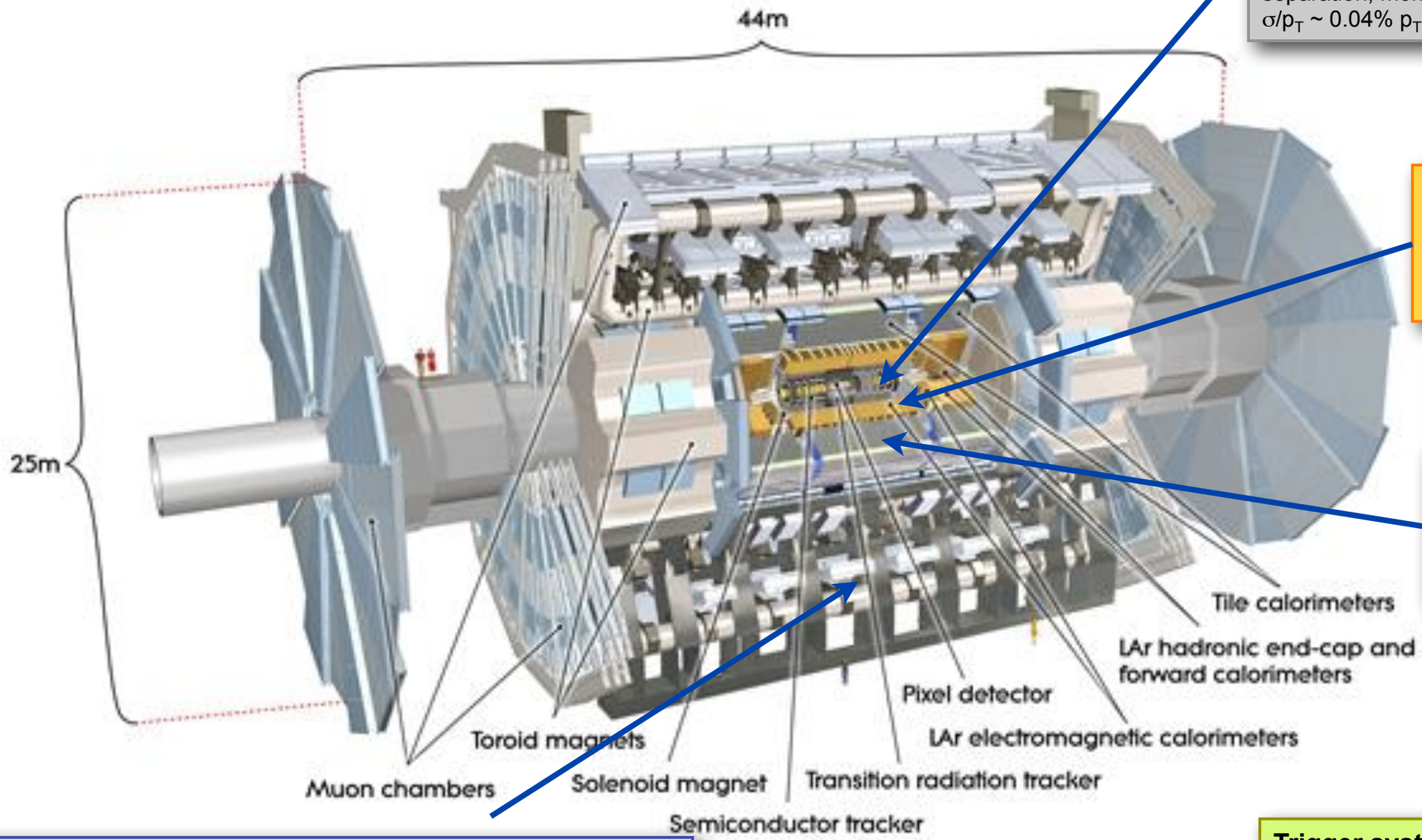
The ATLAS Detector @ LHC

L ~ 46 m, ∅ ~ 22 m, 7000 tons
~10^8 electronic channels

Inner Tracker (|η|<2.5, B=2T): Si Pixels, Si strips, Trans. Rad. Det. Precise tracking and vertexing, e/π separation, momentum resolution: σ/p_T ~ 0.04% p_T (GeV) ⊕ 1.5%

EM calorimeter: Pb-LAr Accordion, e/γ trigger, id. and meas., energy res.: σ/E ~ 10%/√E ⊕ 0.7%

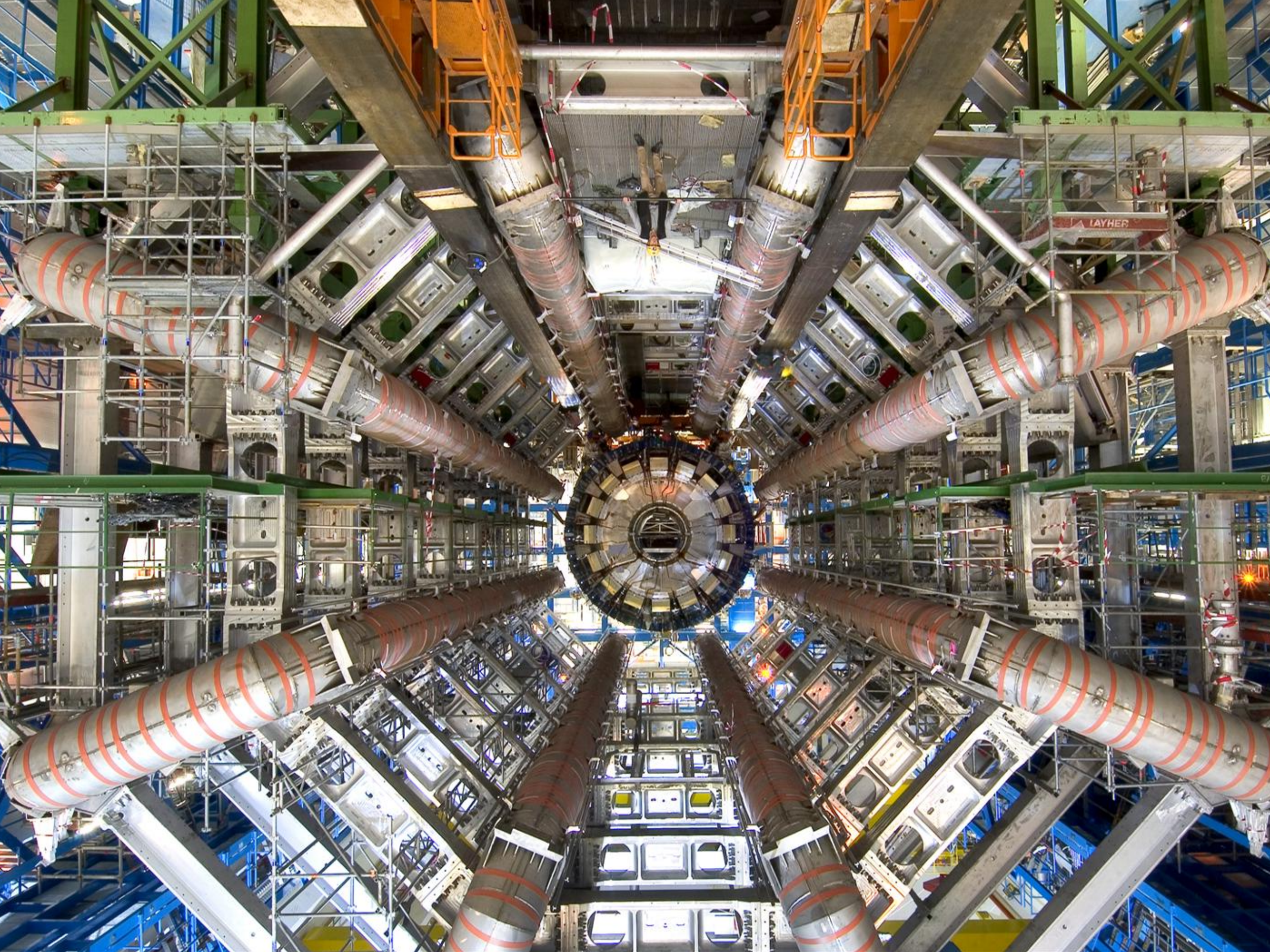
HAD calorimetry (|η|<5): Fe/ scintillator Tiles (cen), Cu/W-LAr (fwd). trigger and meas. of jets and $E_{T,miss}$, energy res.: σ/E ~ 50%/√E ⊕ 3%

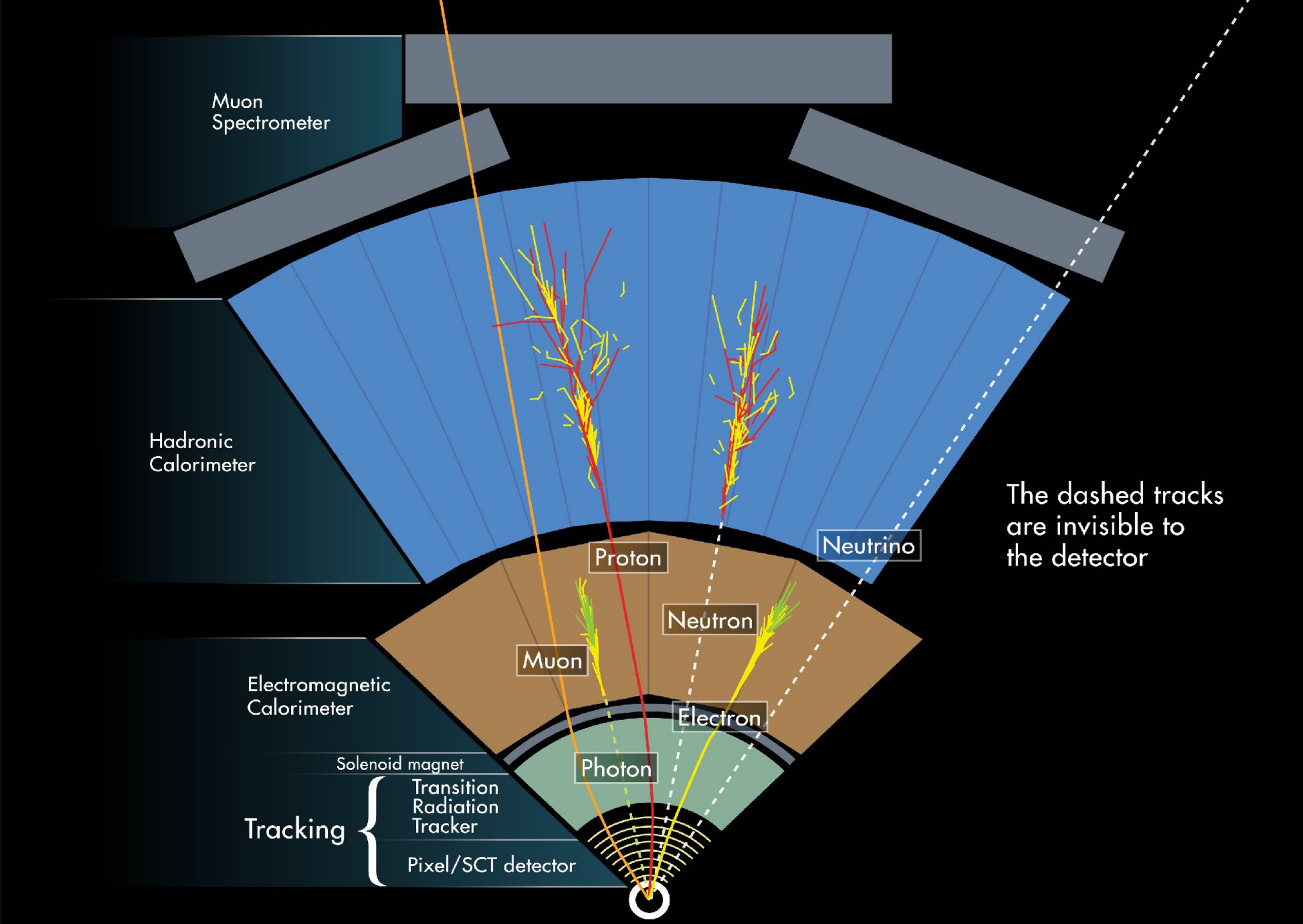


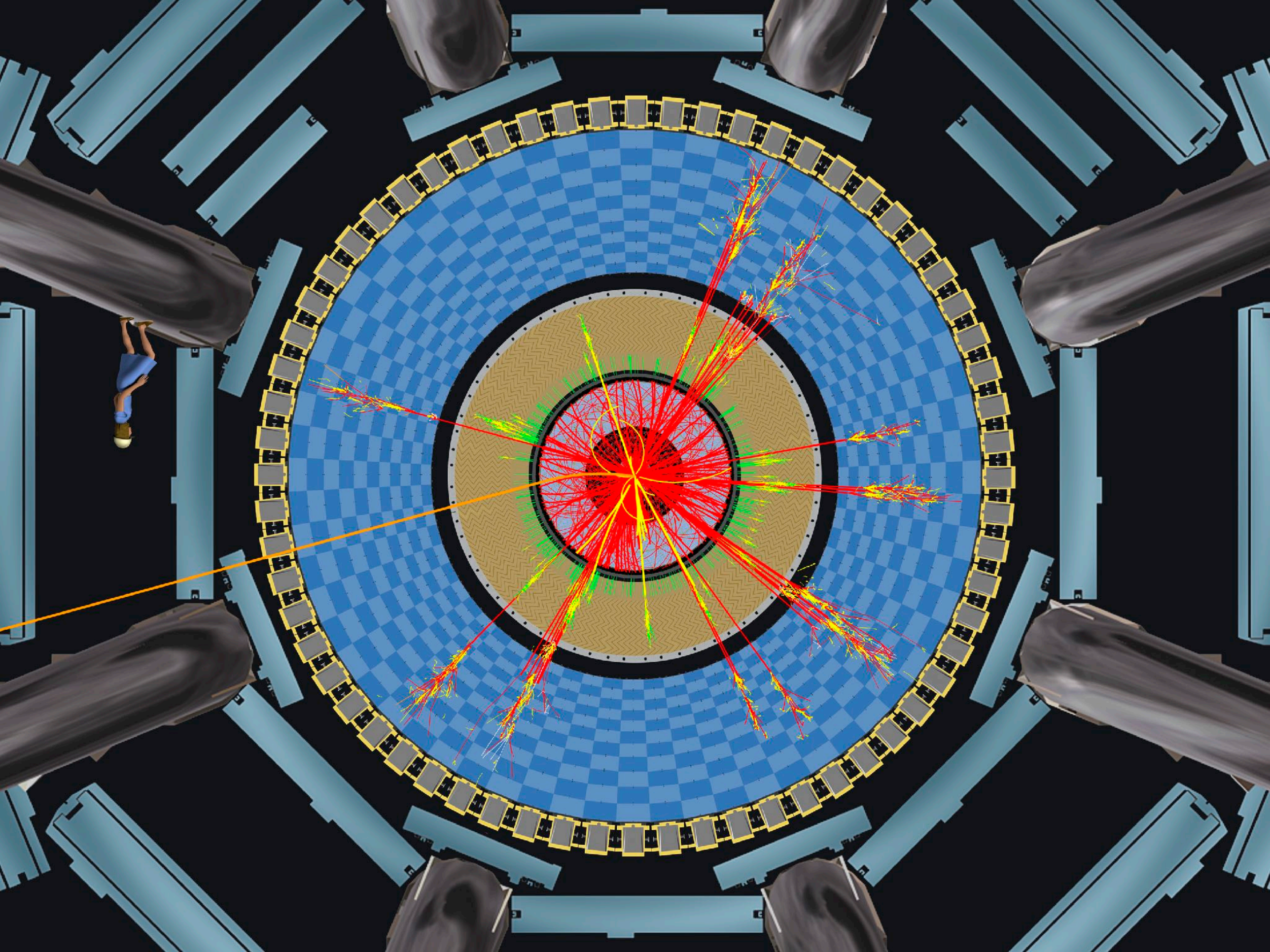
Muon Spectrometer: air-core toroids with gas-based muon chambers. trigger and meas. with momentum resolution < 10% up to E_μ ~ 1 TeV

Trigger system: 3-levels reducing the IA rate from 40 MHz to ~200 Hz

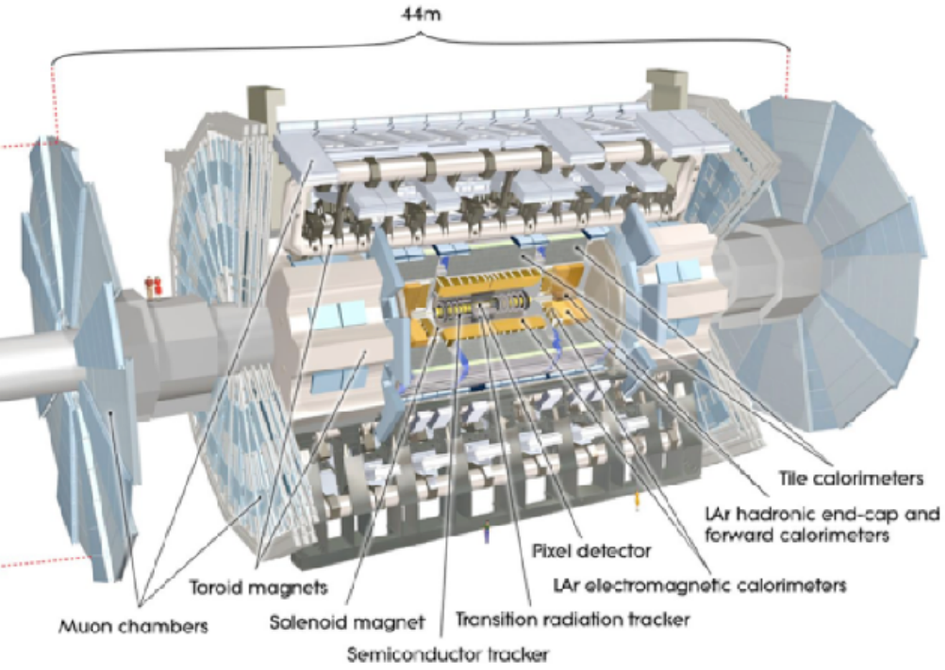
Millions of detector readout channels read out to reconstruct one "event"

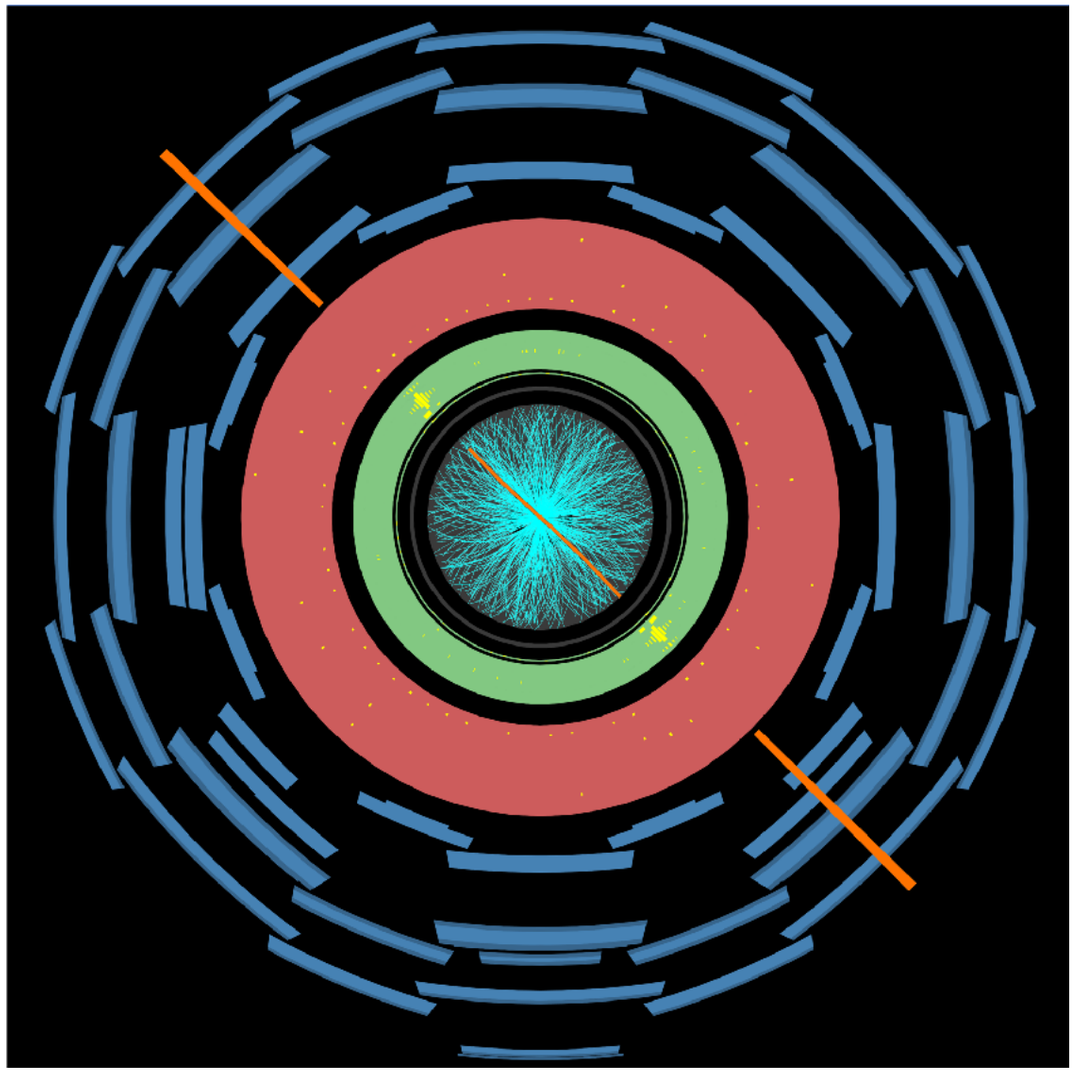

LAYHER


Muon
Spectrometer
Hadronic
Calorimeter
Electromagnetic
Calorimeter
Solenoid magnet
Tracking
Transition
Radiation
Tracker
Pixel/SCT detector
Proton
Neutrino
Neutron
Muon
Electron
Photon
The dashed tracks
are invisible to
the detector



Event displays





Event displays are great ways for us to visualise what happened in a particle collision

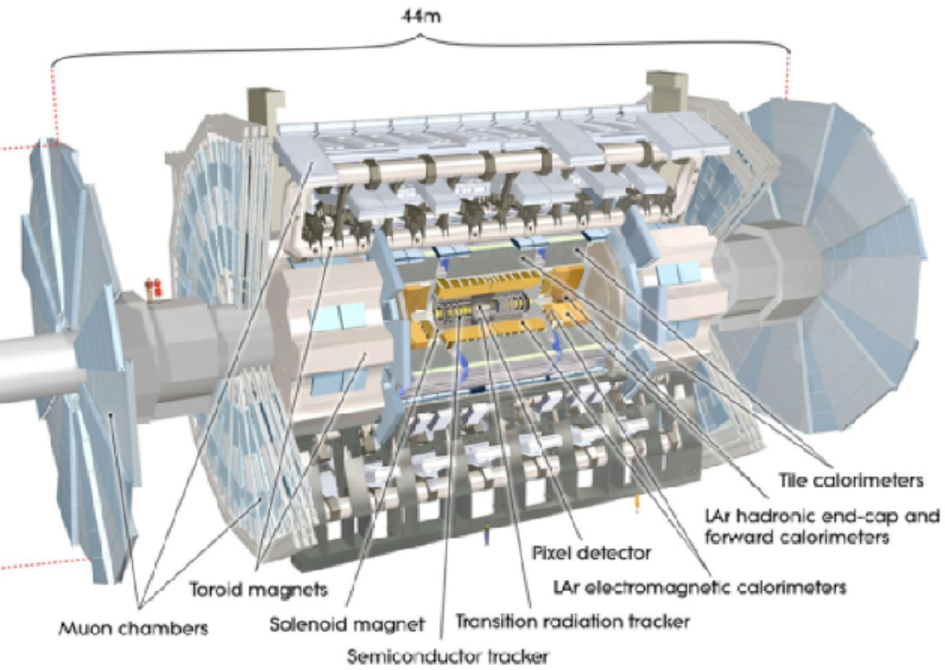
In this **ATLAS event display** (*right*) of a real proton collision, we are looking down the beam pipe, so the plane of the display is transverse to the proton beam direction

***Question**: Can you quantify the momentum in this plane **before** the proton collision*

- *What does that tell you about the distribution of momentum **after** the collision?*
- *Can you say which fundamental particle(s) is (are) observed in the event?*

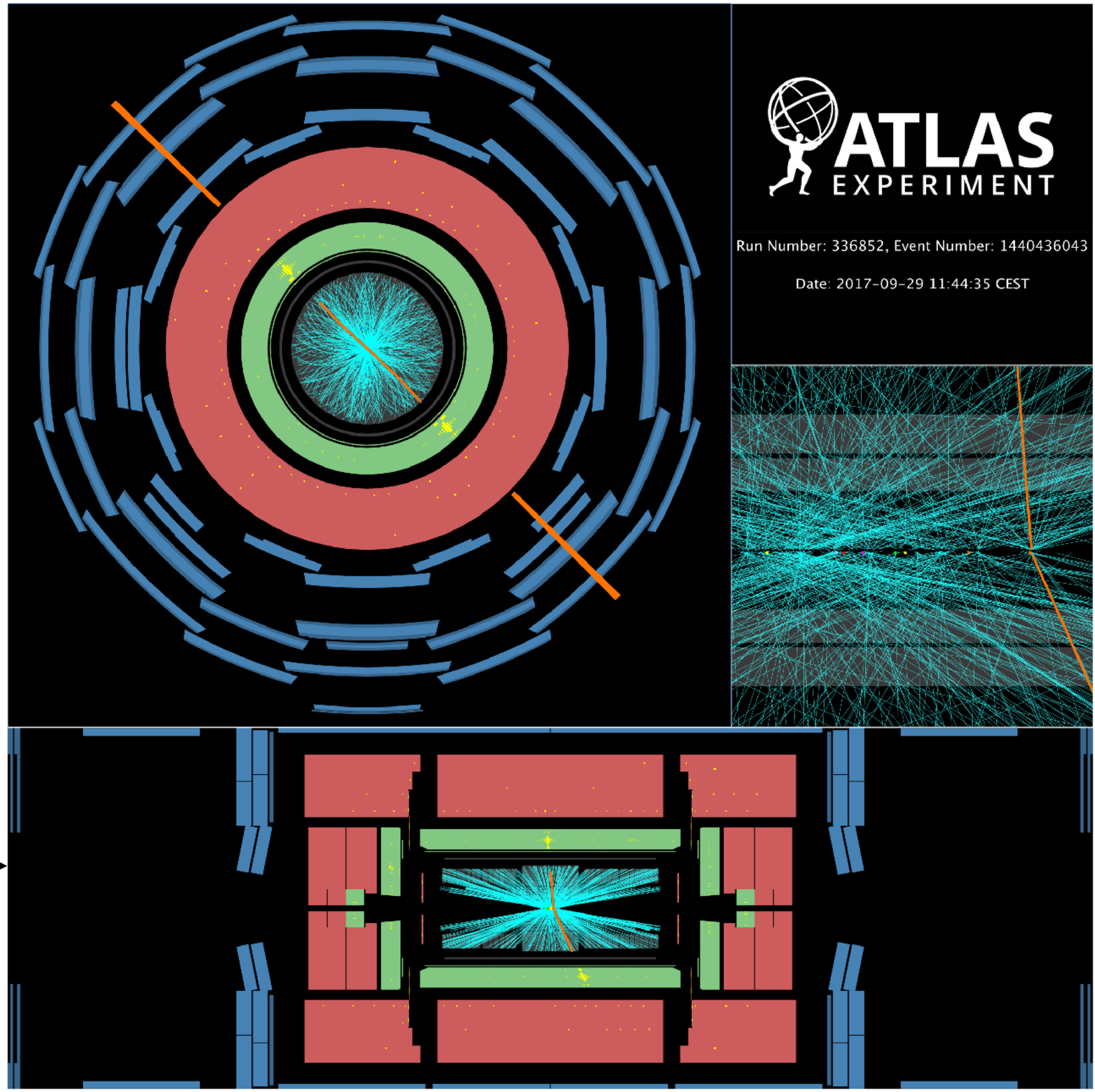
Brookhaven National Laboratory

Event displays



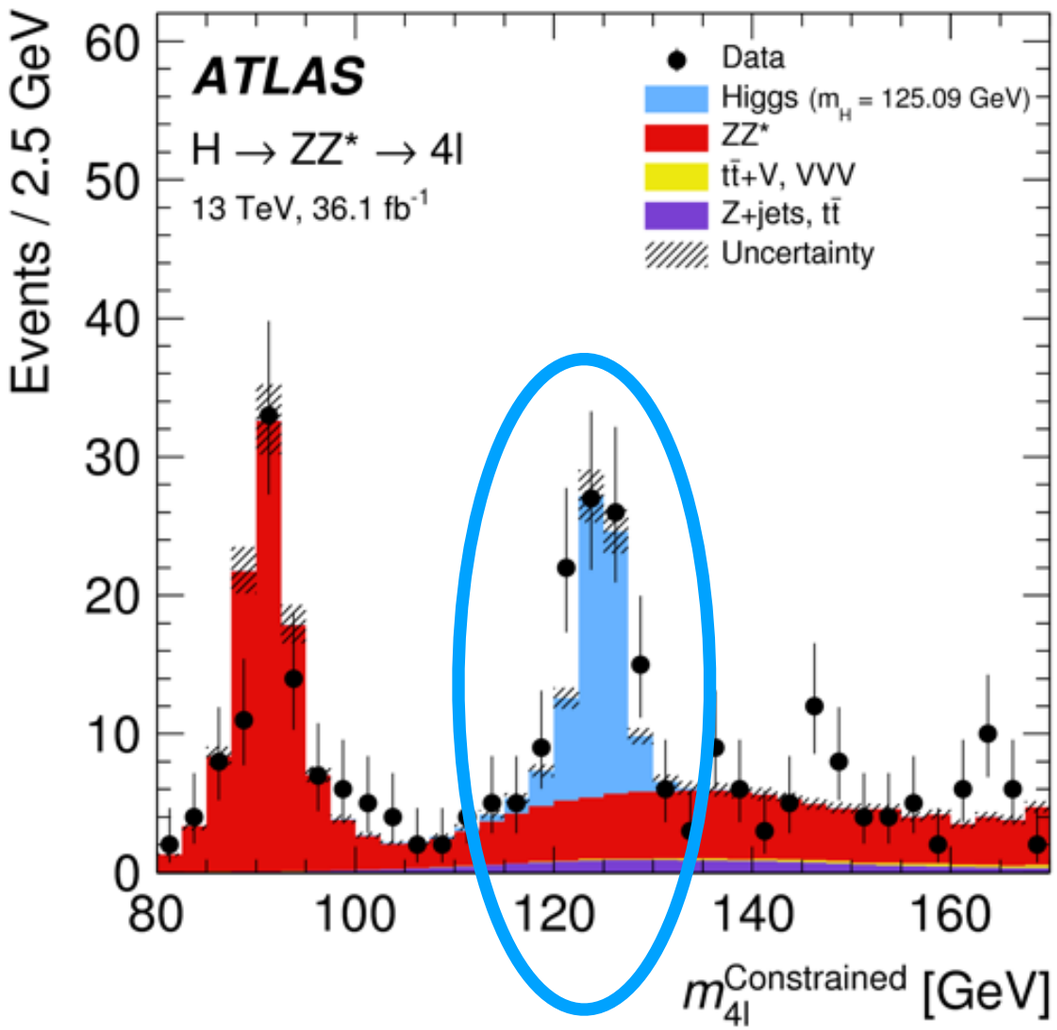
This view shows the plane in the proton beam direction

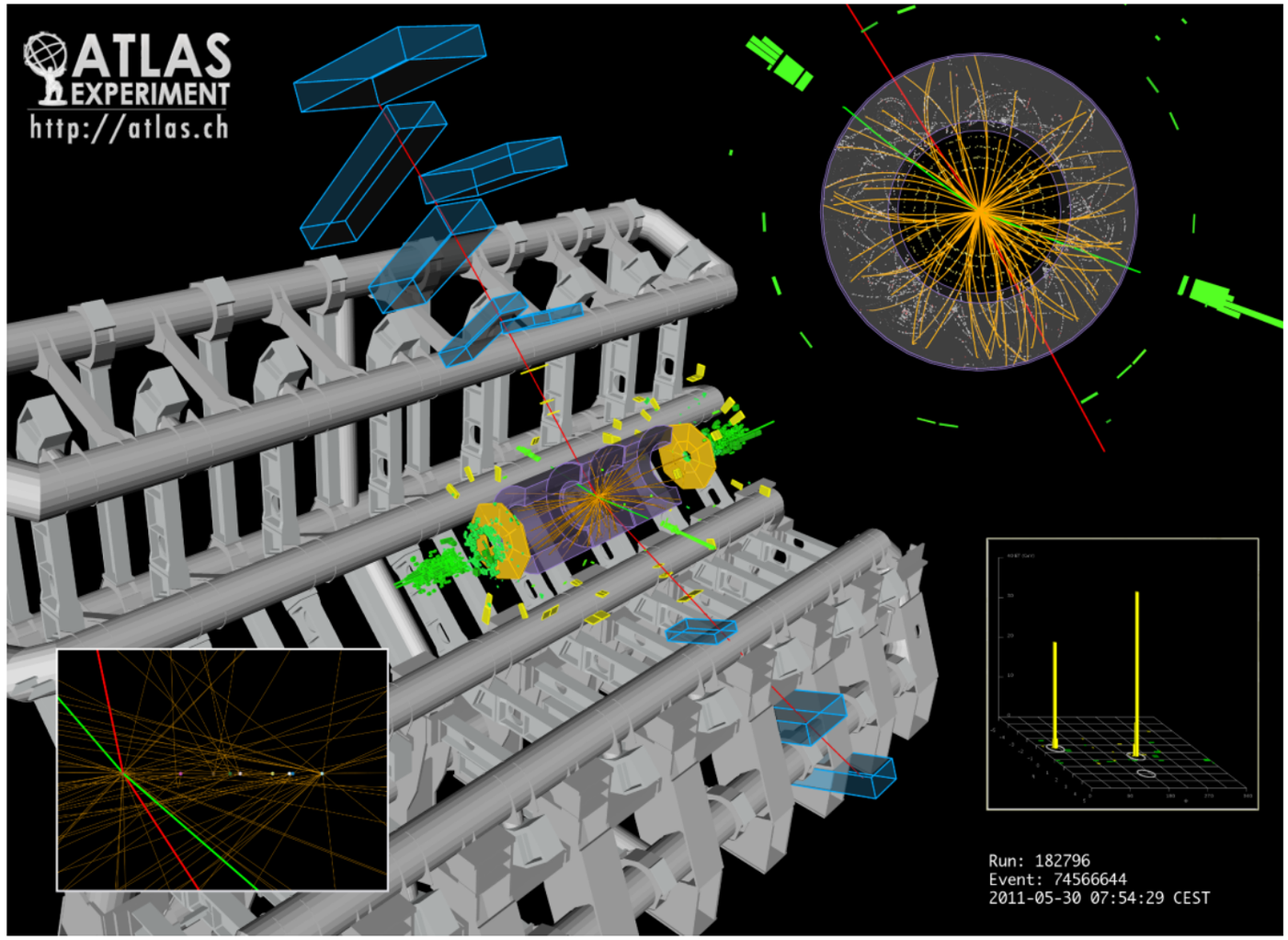
Both **2D** views are often used to provide complementary information



Brookhaven National Laboratory

Discovering the Higgs Boson





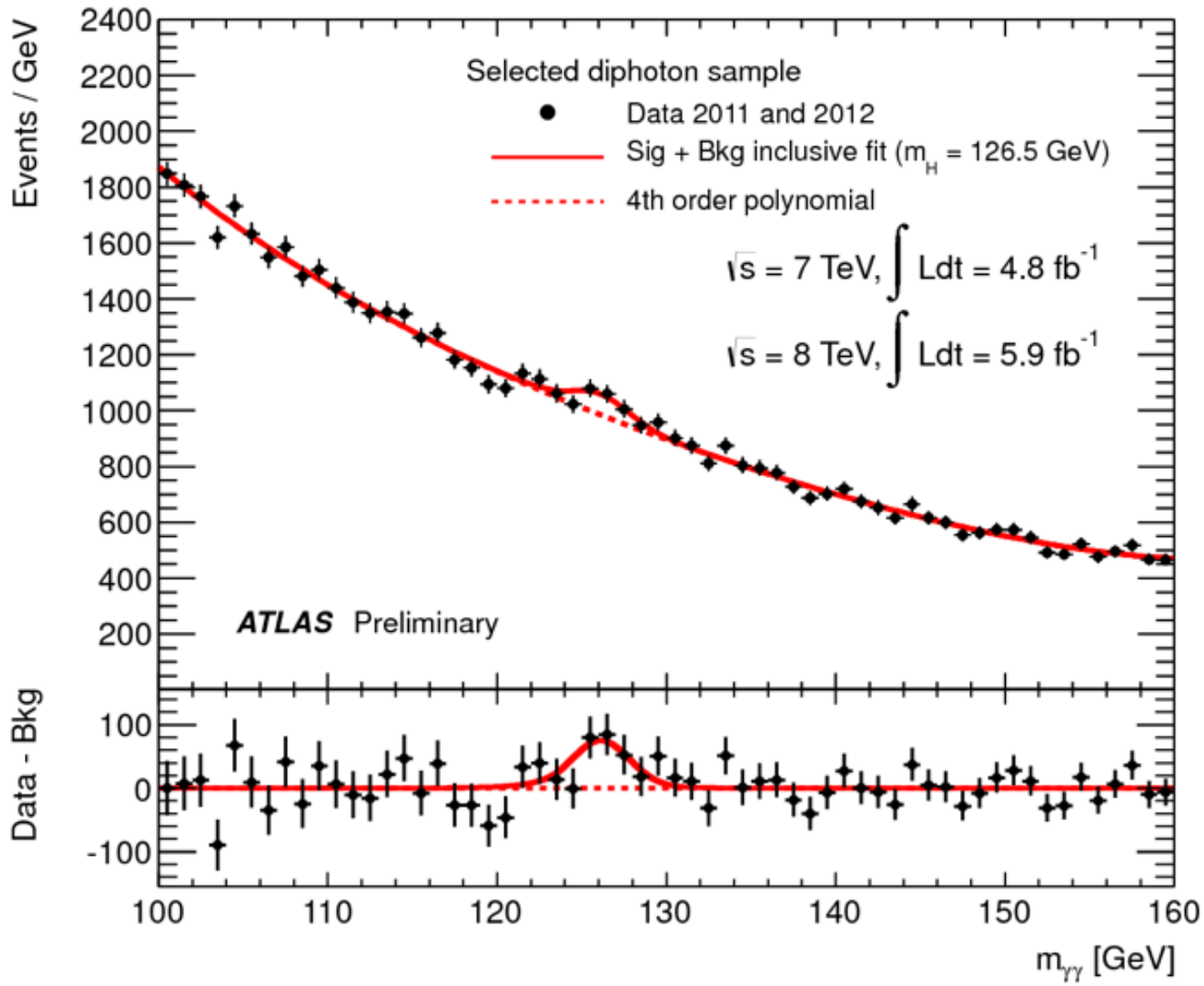
Strategy for this channel (there are several) - we look for events with **two Z bosons** that have decayed to **four leptons**, e.g. two electrons and two muons in the event on the right

If the **two Z bosons** were produced by the **decay of a Higgs boson**, when we reconstruct the invariant mass of the system we should see a **peak** at the **Higgs boson mass**

Brookhaven National Laboratory

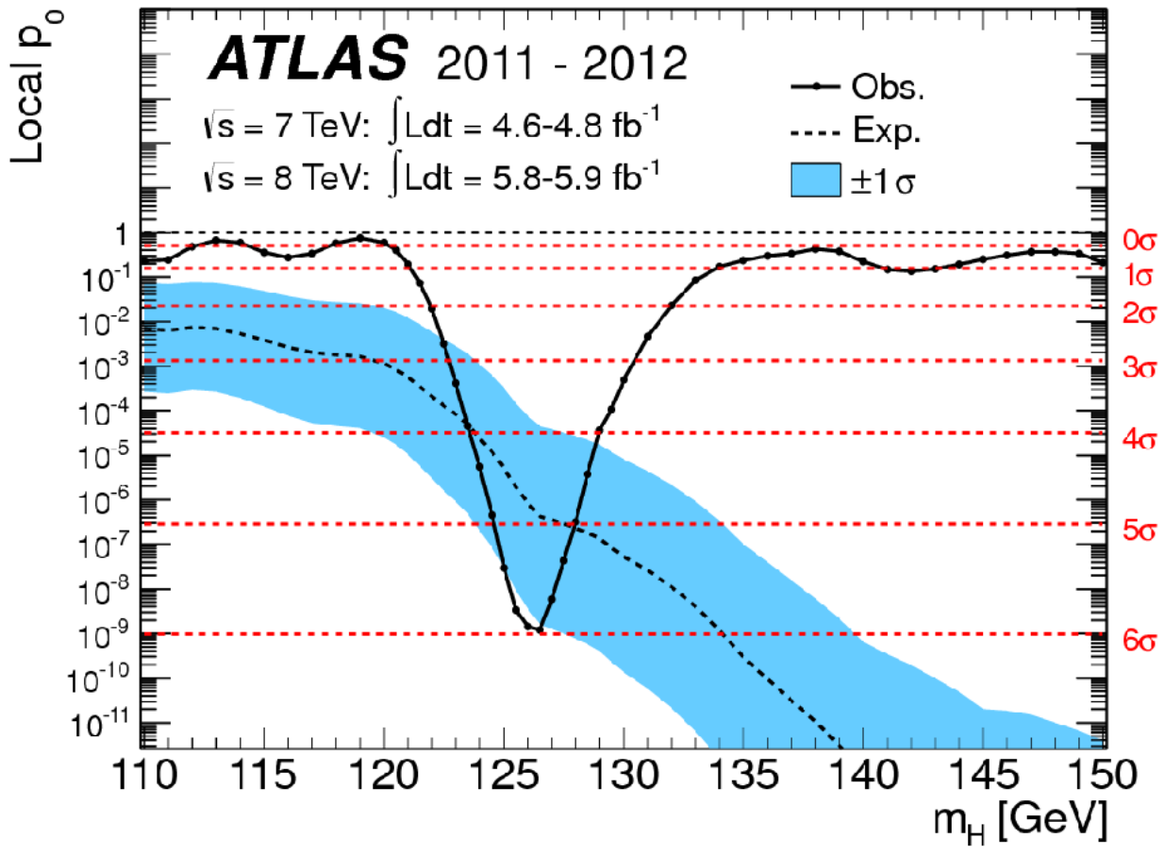
Needles in haystacks

- There are billions of events and the ones we are really interested in are ***very rare***
- Often the interesting events are also ***very difficult to distinguish*** from background
  - Requires ***high precision detectors***, which means ***lots of data*** for each event
- The data are structured but each event is different - ***unique data science challenge***



Brookhaven National Laboratory

Higgs discovery on July 4th 2012

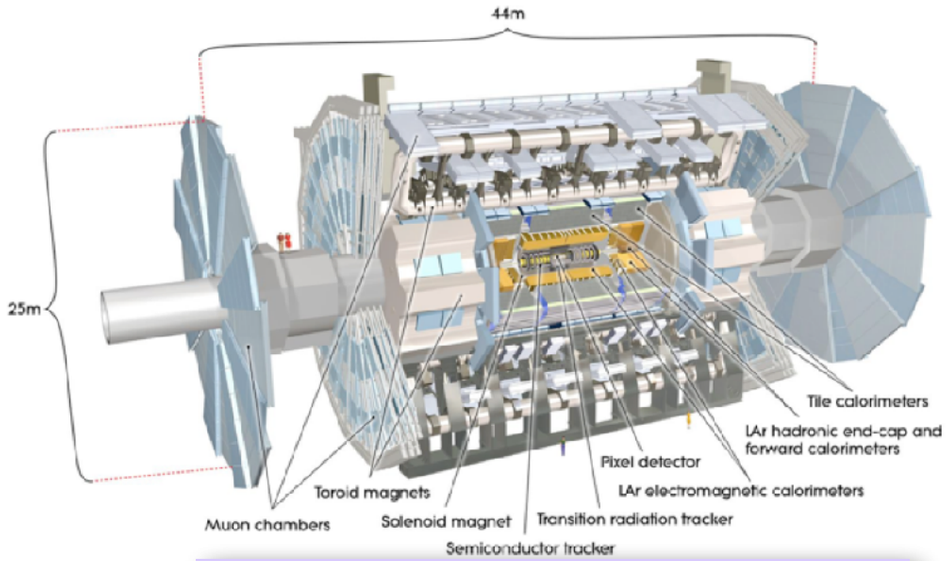


In 2012 the number of observed events (**6σ**) was consistent with, and in excess of the number of events expected for a standard model Higgs (**5σ**)

Question *- Imagine we had several more Large Hadron Colliders, with a total of 9 independent measurements possible. Roughly how many measurements would you expect to lie* ***outside*** *the ±1σ blue band?*

Brookhaven National Laboratory

Exabyte-scale physics analysis



Exabytes of Data

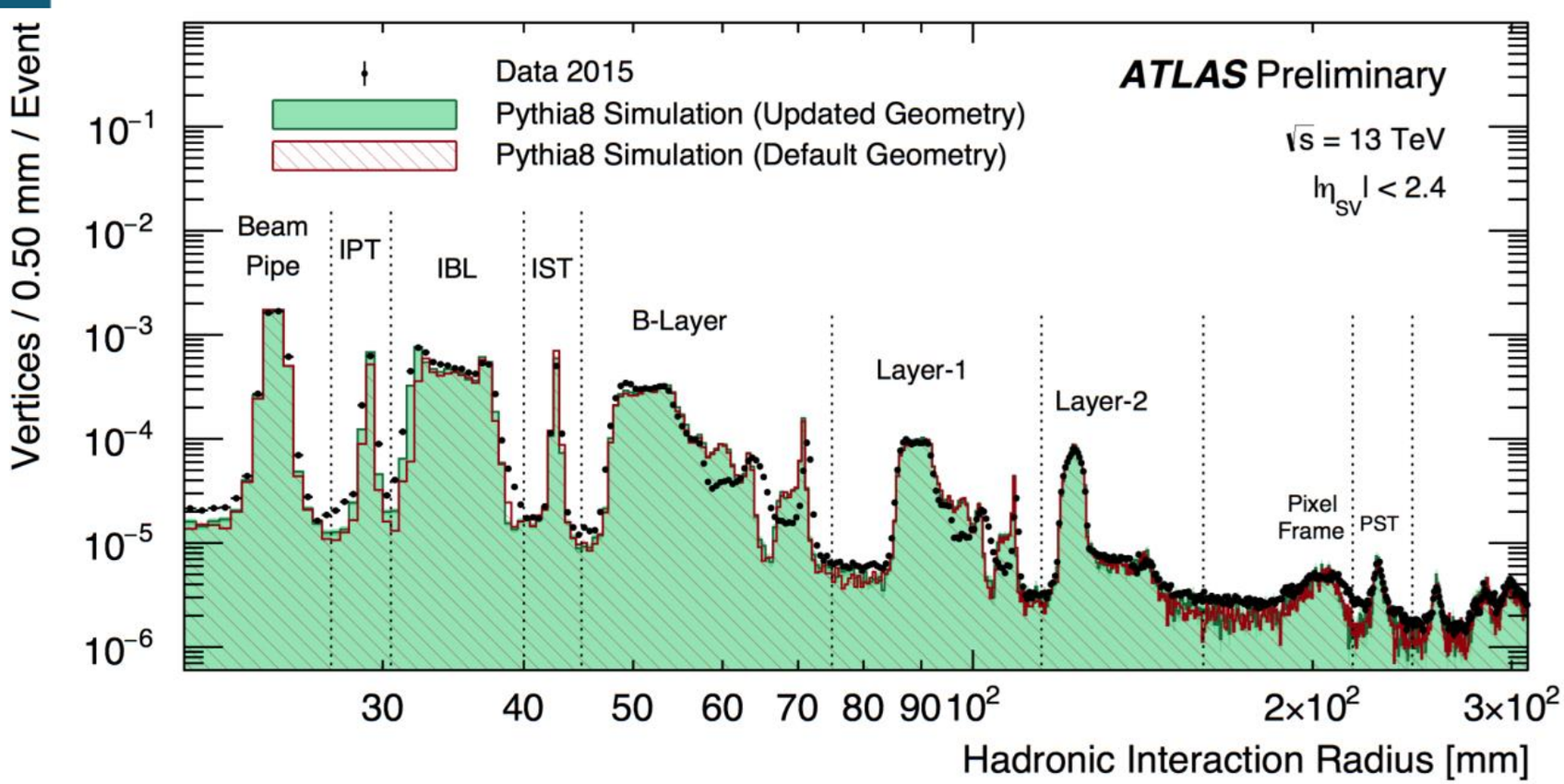
Simulation and understanding detectors

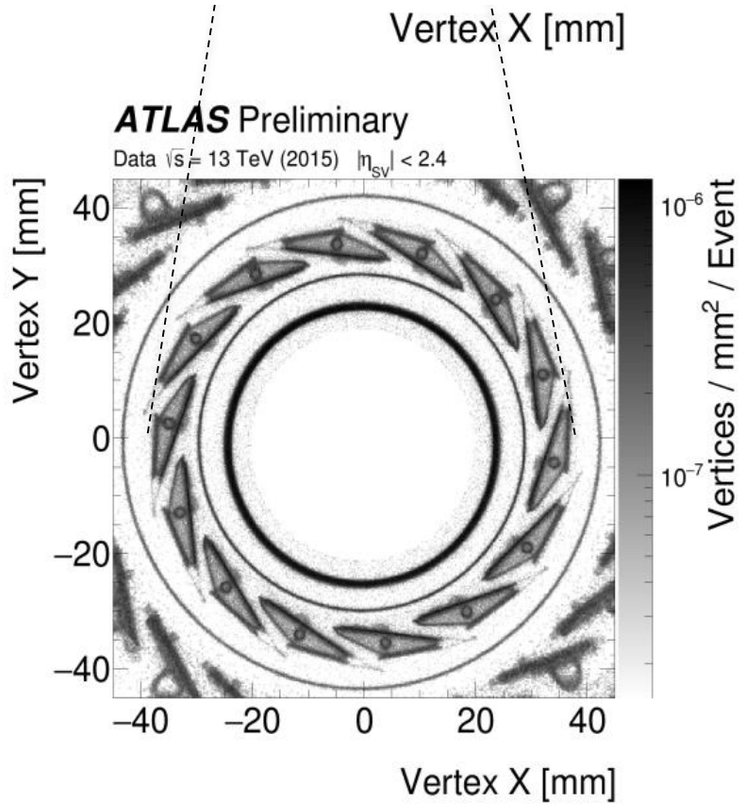
We use **simulations** to model the detector as **accurately** and **precisely** as possible

We then ***test*** that our simulations are accurate ***using real data***

We correct our simulations if necessary

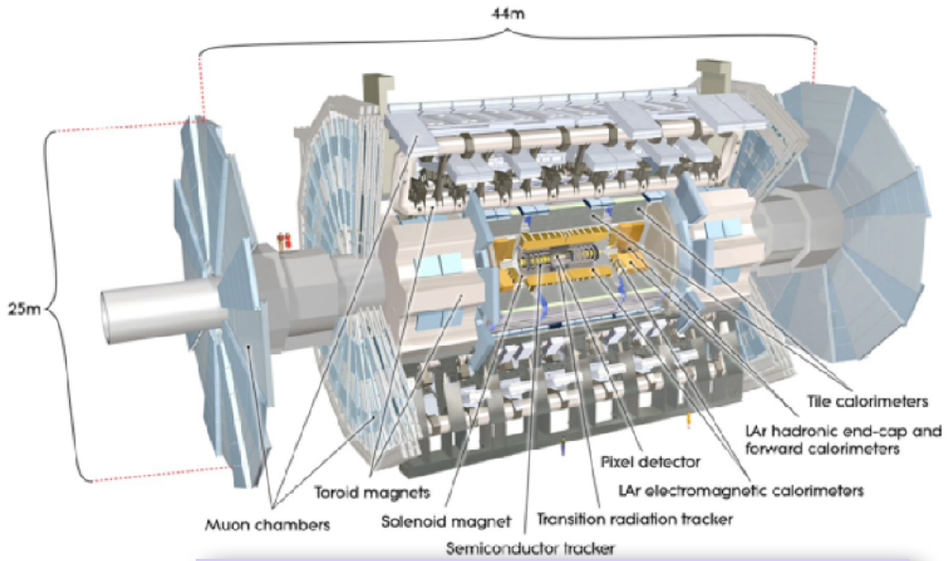
Once our simulation is an ***accurate model*** of our detector, we can use it to ***correct the data for detector response***





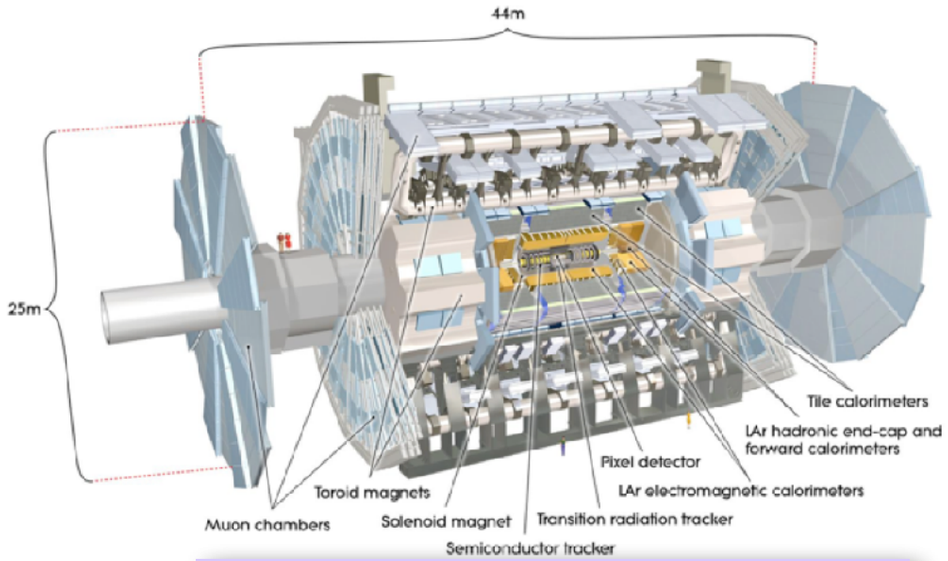
Brookhaven National Laboratory

Exabyte-scale physics analysis

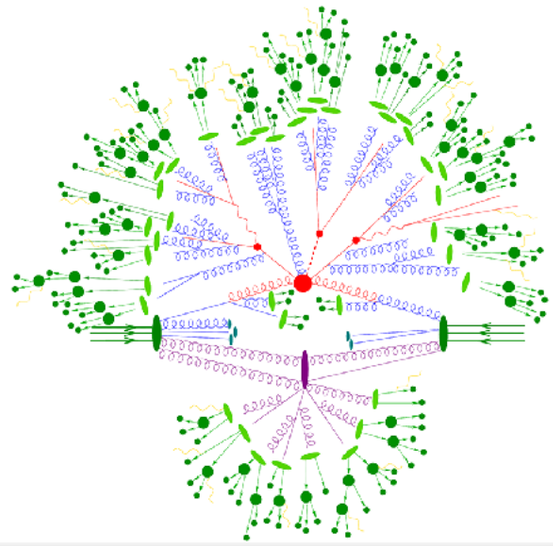


Exabytes of Data

Exabyte-scale physics analysis

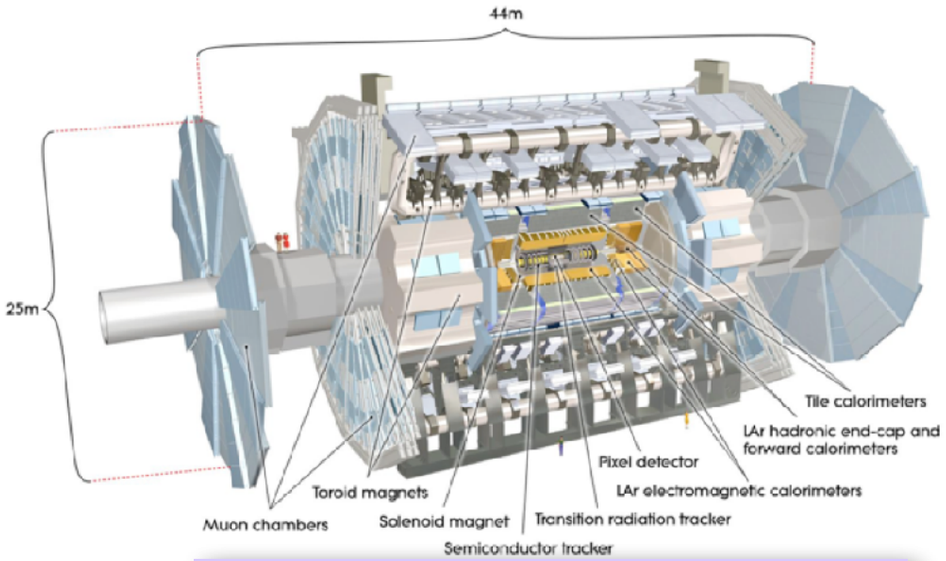


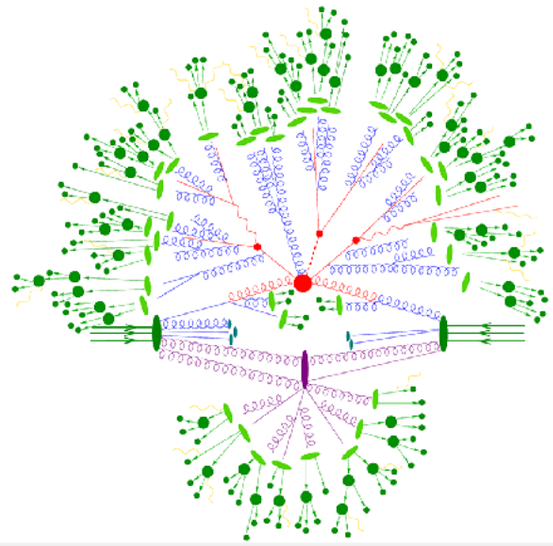
Exabytes of Data



Exabytes of Simulation

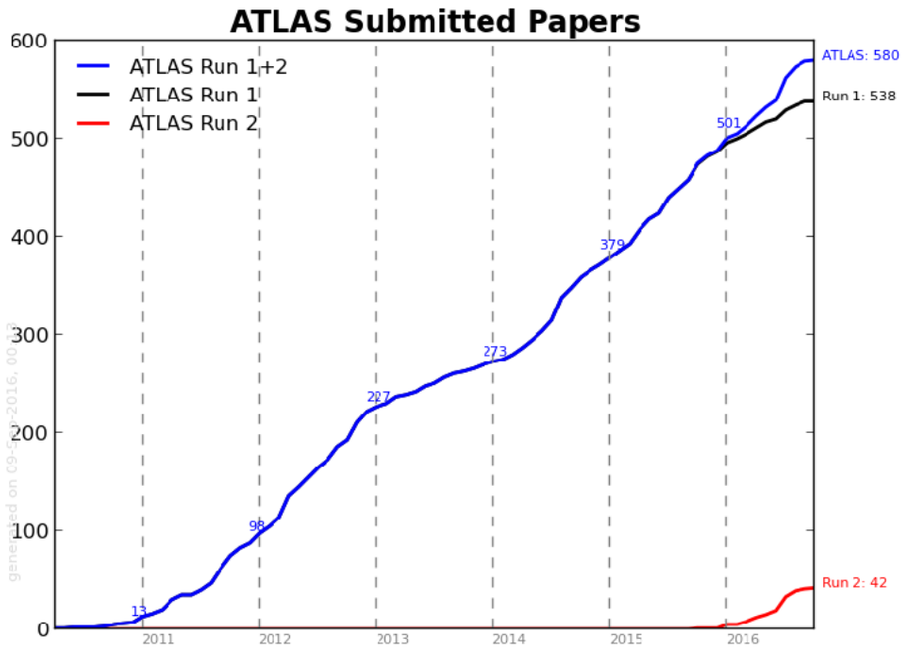
Exabyte-scale physics analysis





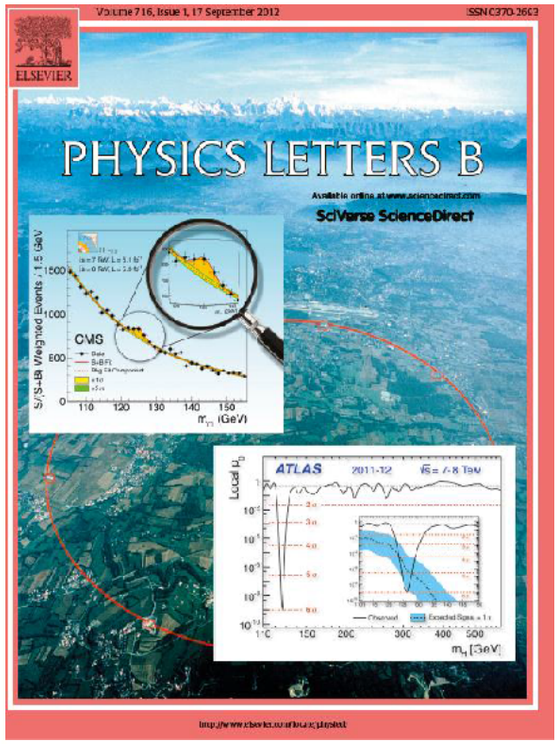
Exabytes of Data

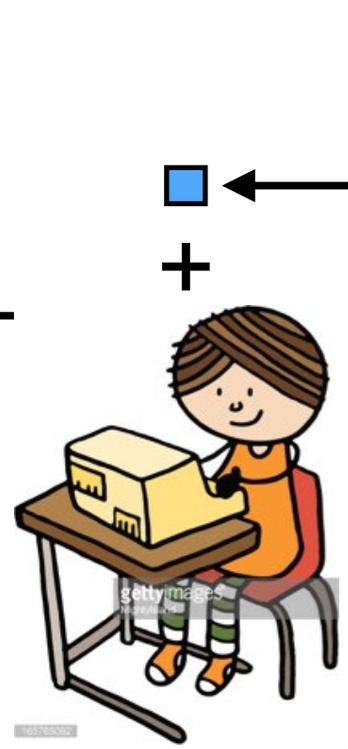
Exabytes of Simulation

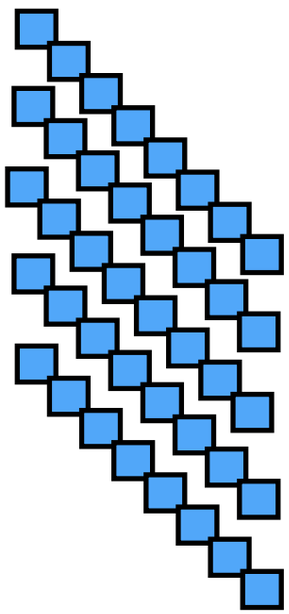


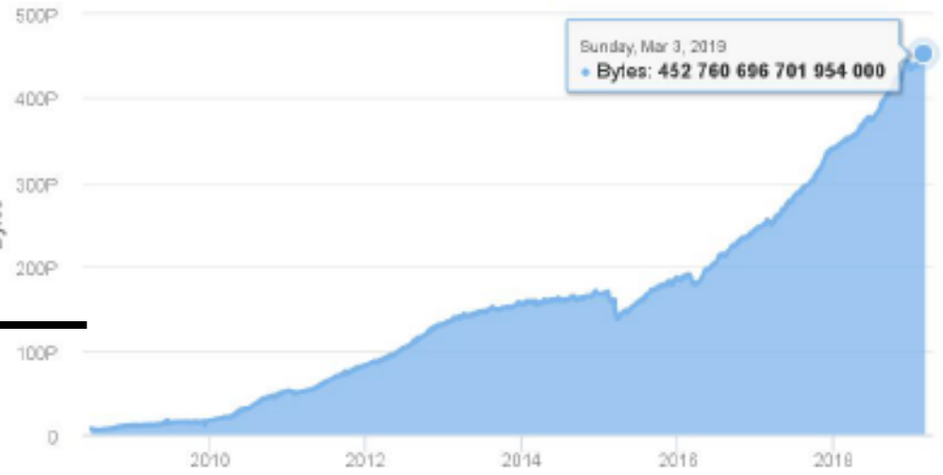
Publish!

Data analysis







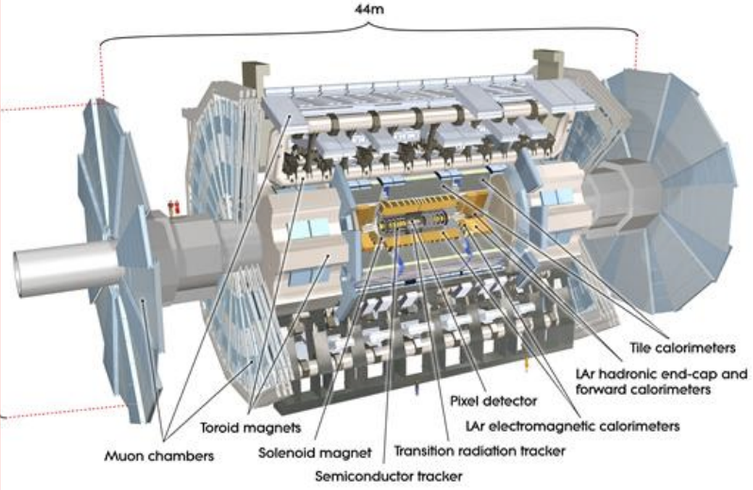


Analysis is performed on only a fraction of the data

How?

Brookhaven National Laboratory

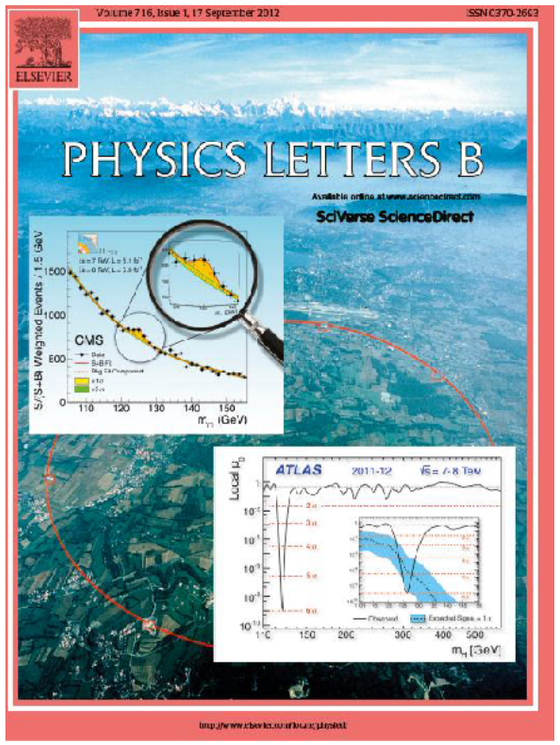
Data's journey

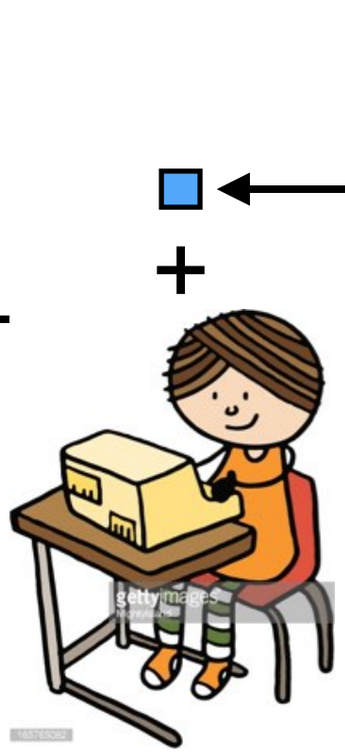


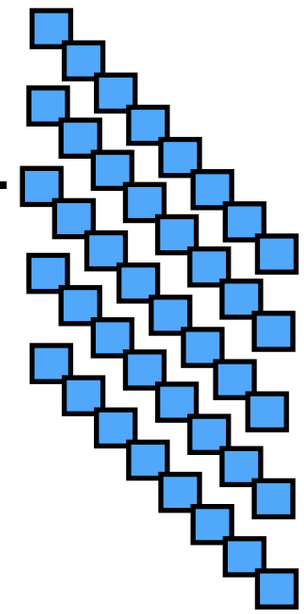
Trigger/DAQ

Data Preparation

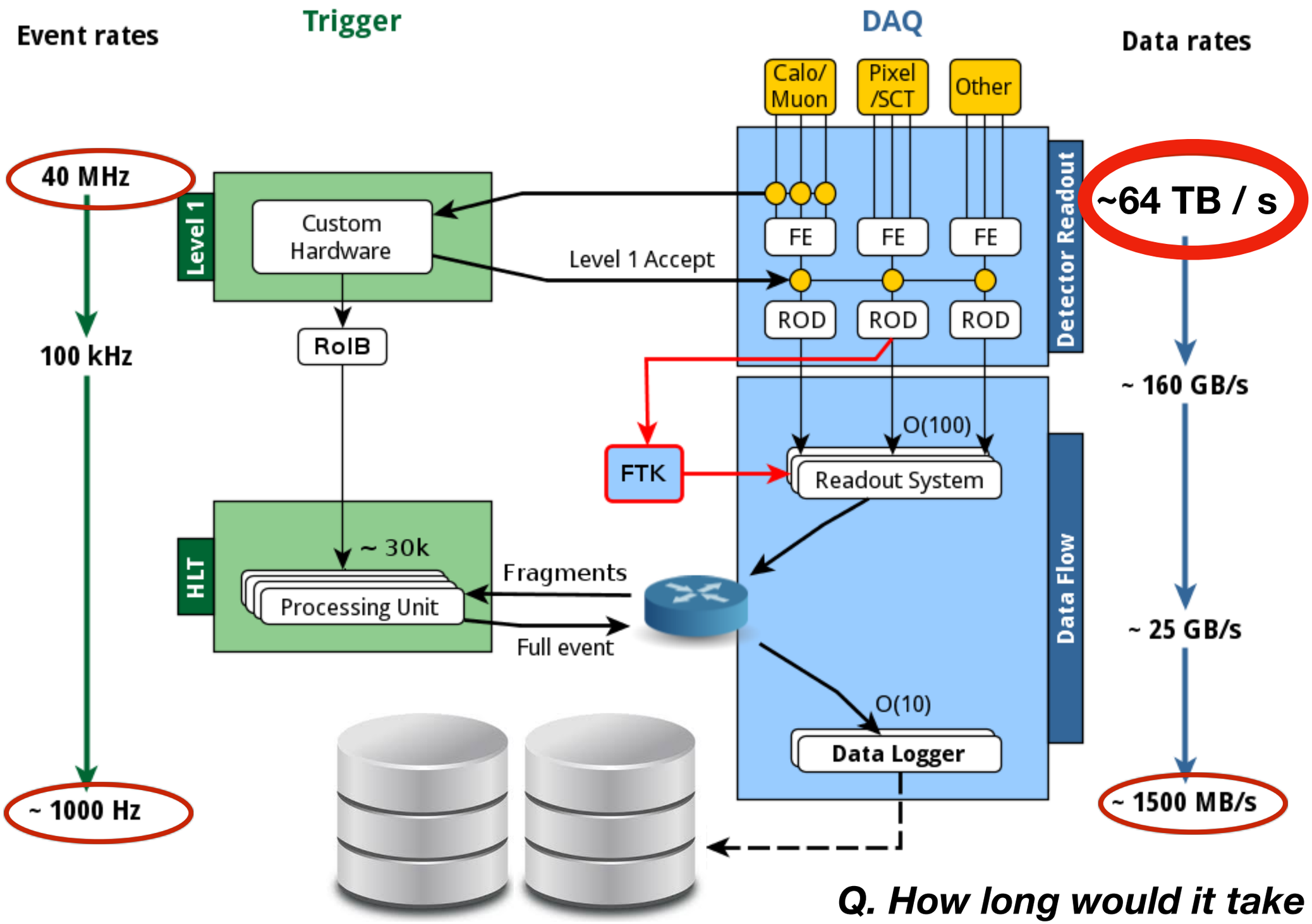
Distributed computing





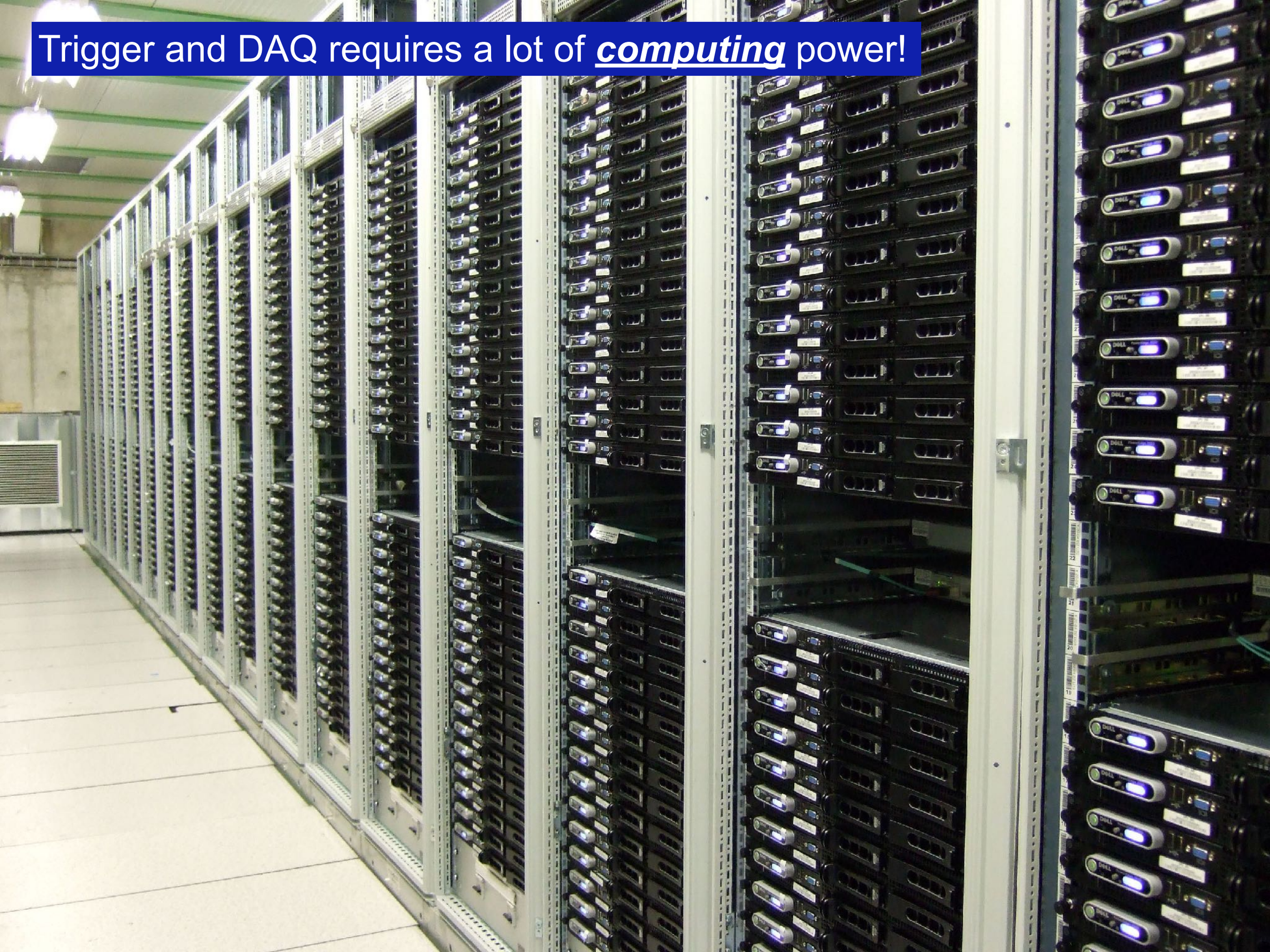


The Atlas Trigger and DAQ

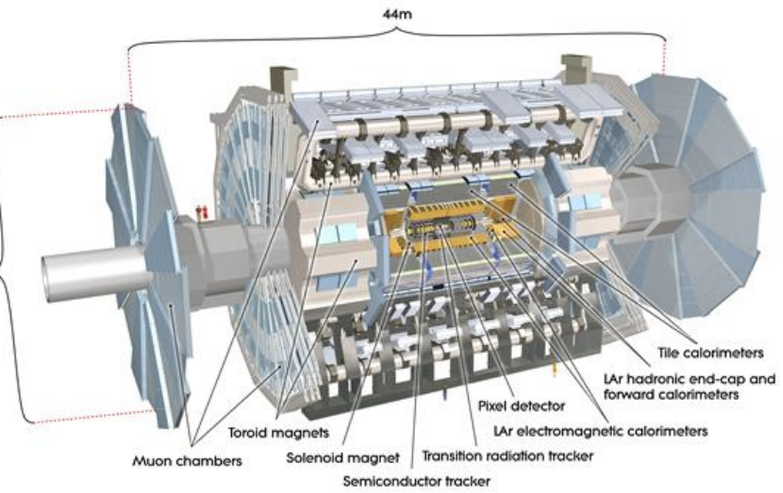


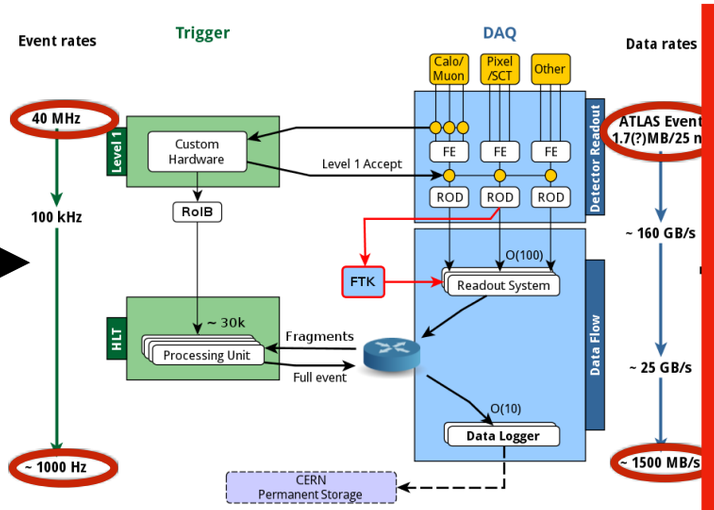
Q. How long would it take to generate 500PB of data if there were no Trigger?

Trigger and DAQ requires a lot of ***computing*** power!

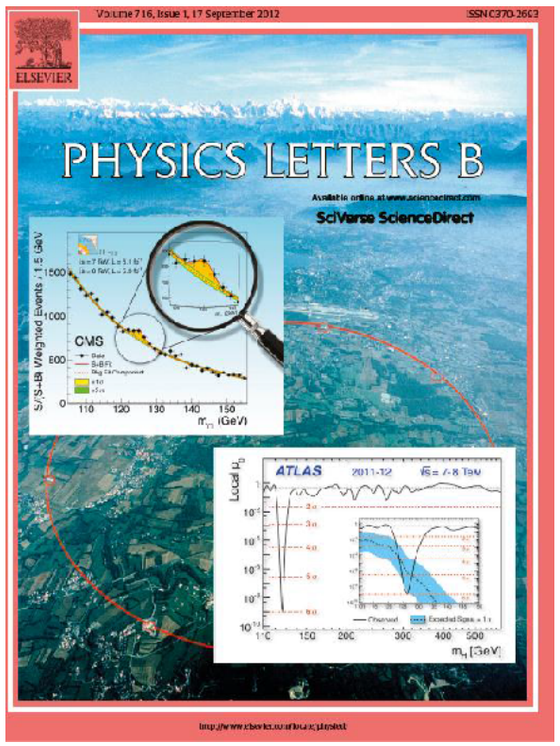


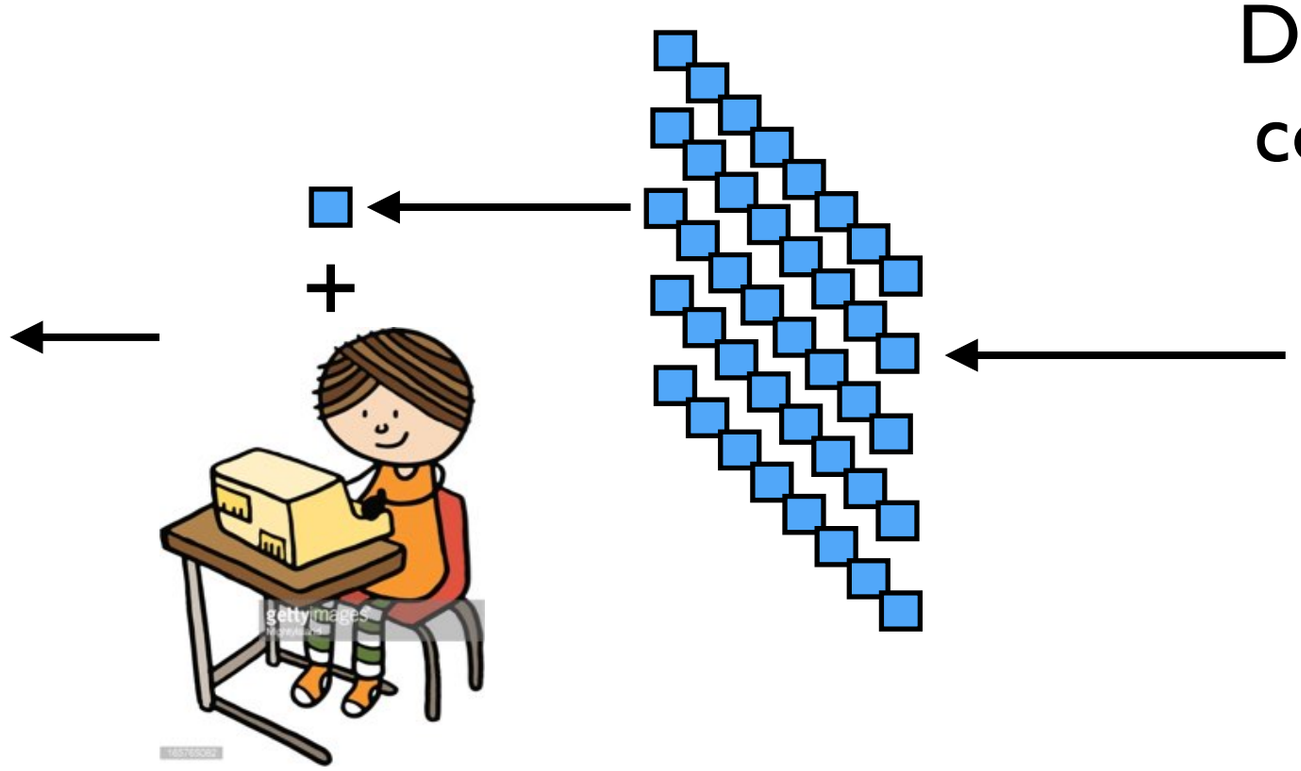
Data's journey





Data Preparation





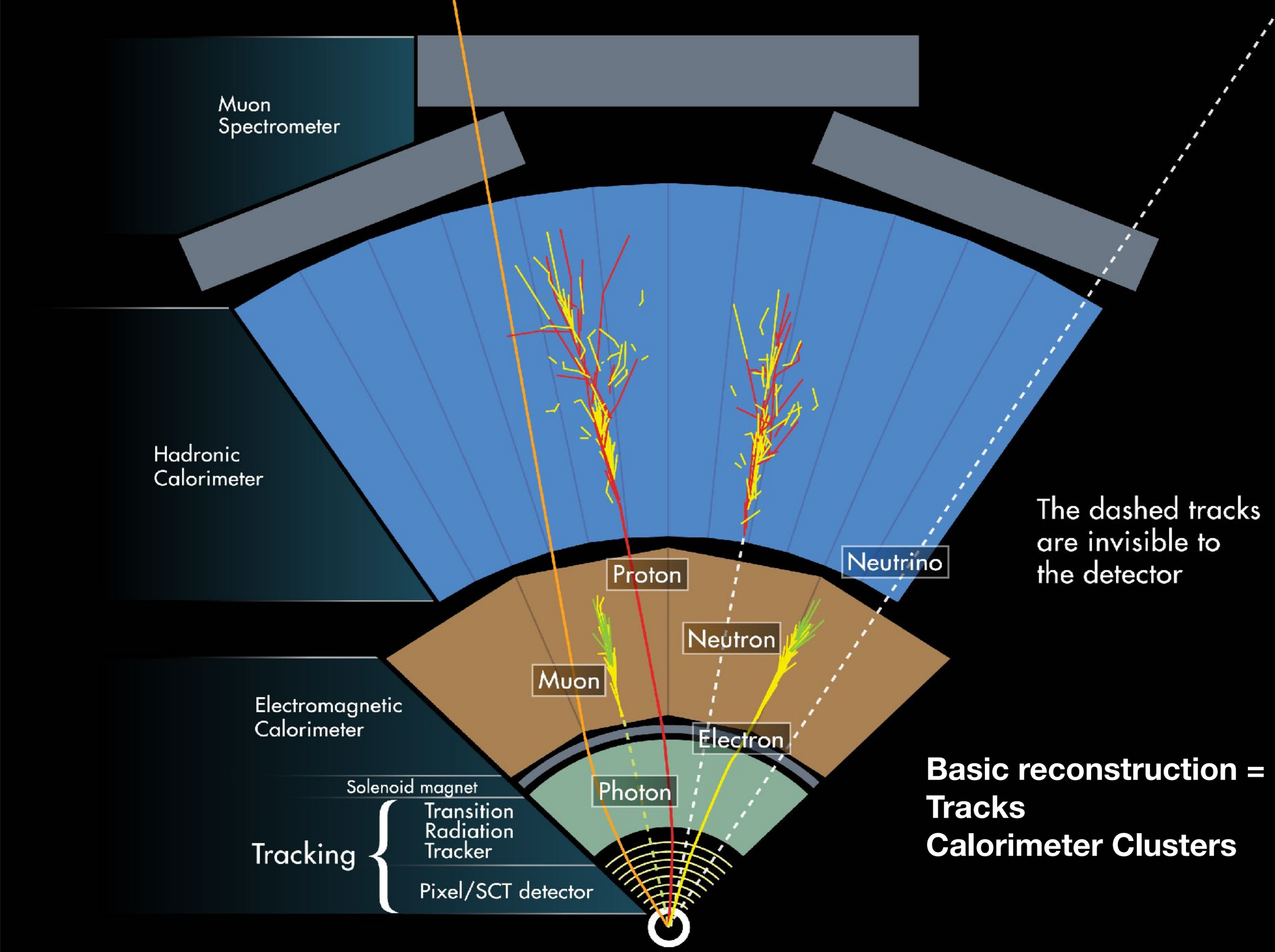
Distributed computing

Data Preparation

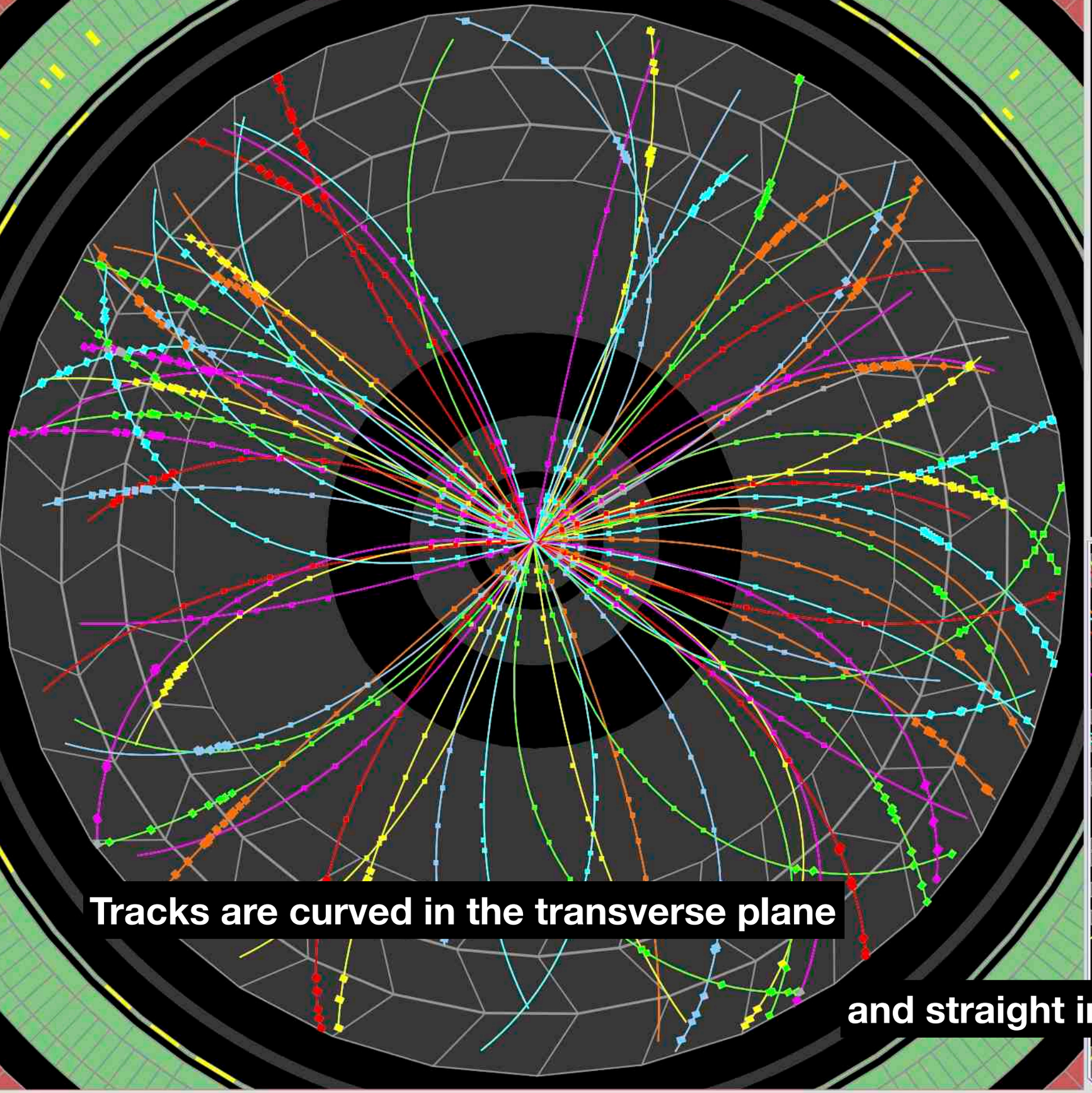
Three major steps to **prepare data for physics analysis** and achieve

- reliable, high quality data (yes, we ***reject*** low quality data)
- the ***best performance*** from our detectors
- readiness for ***physics analysis***

1. **Reconstruct physics signals** from the data
   - Produce information like how many muons does the event have?



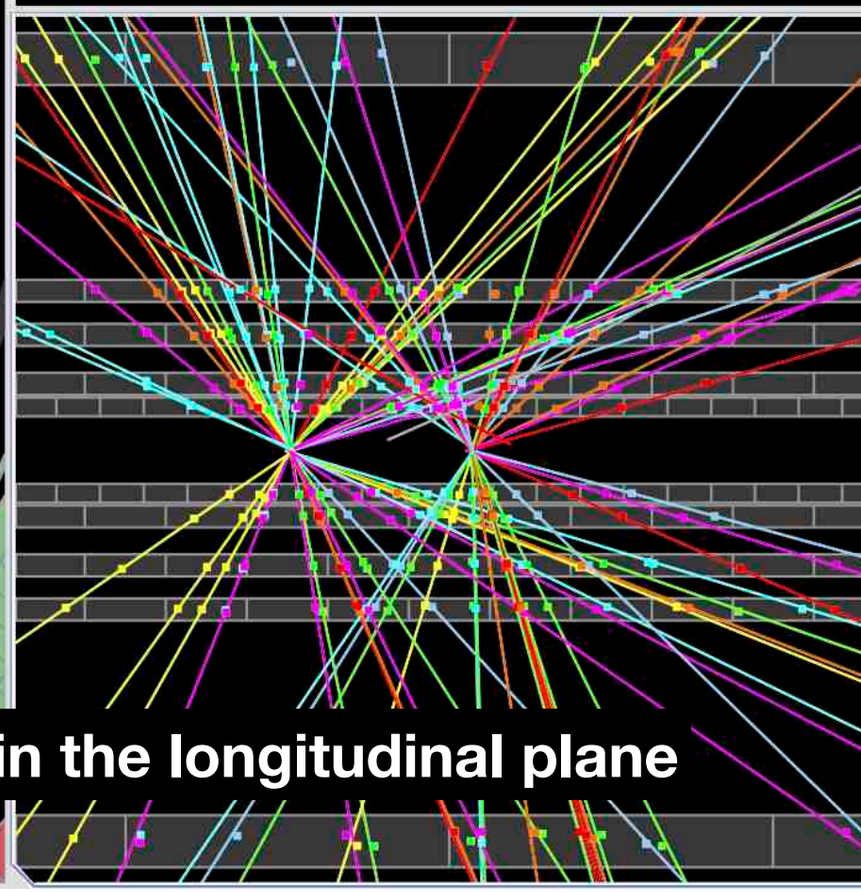
Basic reconstruction =
Tracks
Calorimeter Clusters



ATLAS EXPERIMENT

Run Number: 265545, Event Number: 5720351

Date: 2015-05-21 10:39:54 CEST

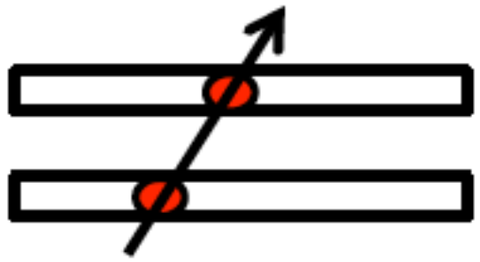


Tracks are curved in the transverse plane

and straight in the longitudinal plane

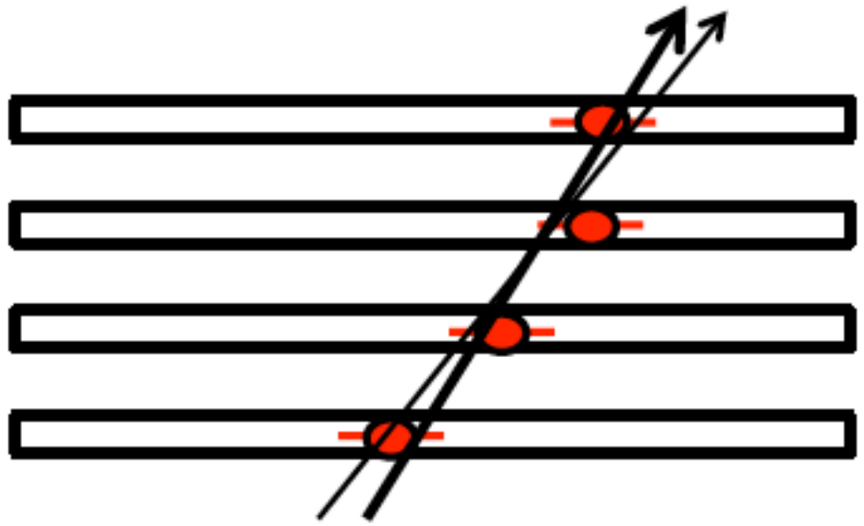
Track fitting

- Perfect measurement – ideal



- Imperfect measurement – reality

- Small errors and more points help to constrain the possibilities



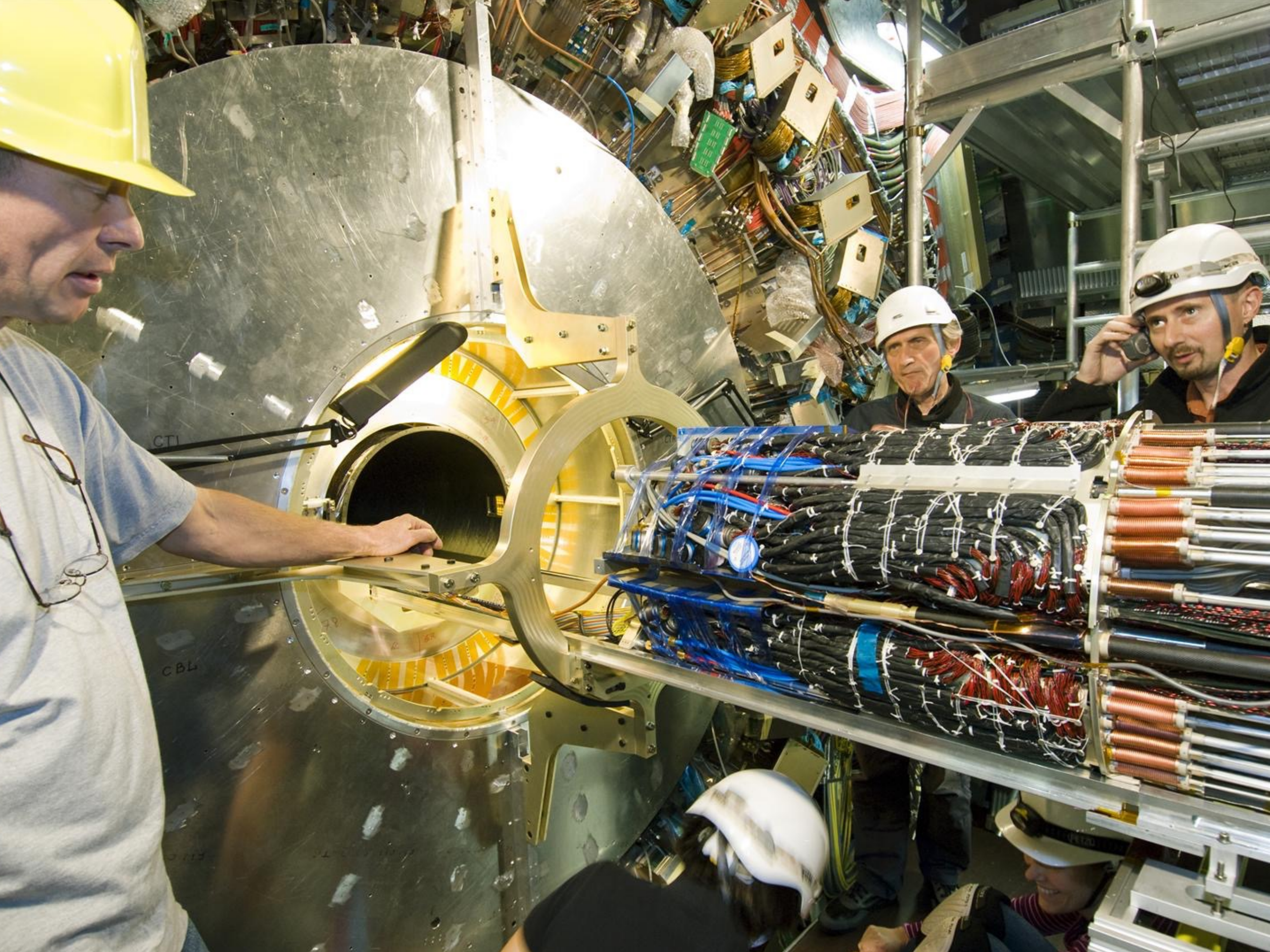
- **Quantitatively:**
  - Parameterize the track;
  - Find parameters by Least-Squares-Minimization;
  - Obtain also uncertainties on the track parameters.

Data Preparation

Three major steps to **prepare data for physics analysis** and achieve

- reliable, high quality data (yes, we ***reject*** low quality data)
- the ***best performance*** from our detectors
- readiness for ***physics analysis***

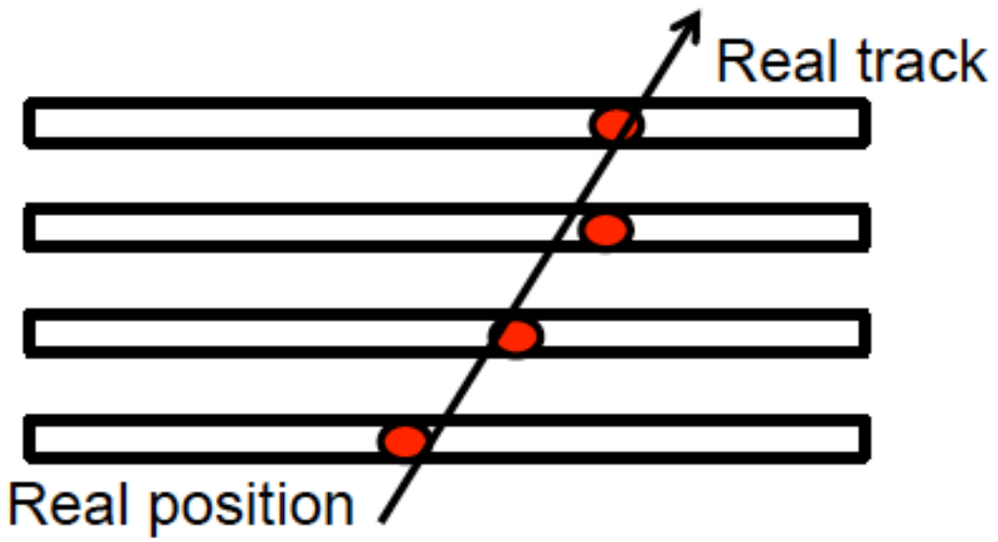
1. **Reconstruct physics signals** from the data 
   - Produce information like how many muons does the event have?
2. **Calibrate** the detectors
   - Correct imperfections, account for changes over time…



Real detector effects

- **Presence of Material**
  - Coulomb scattering off the core of atoms
  - Energy loss due to ionization
  - Bremsstrahlung
  - Hadronic interaction

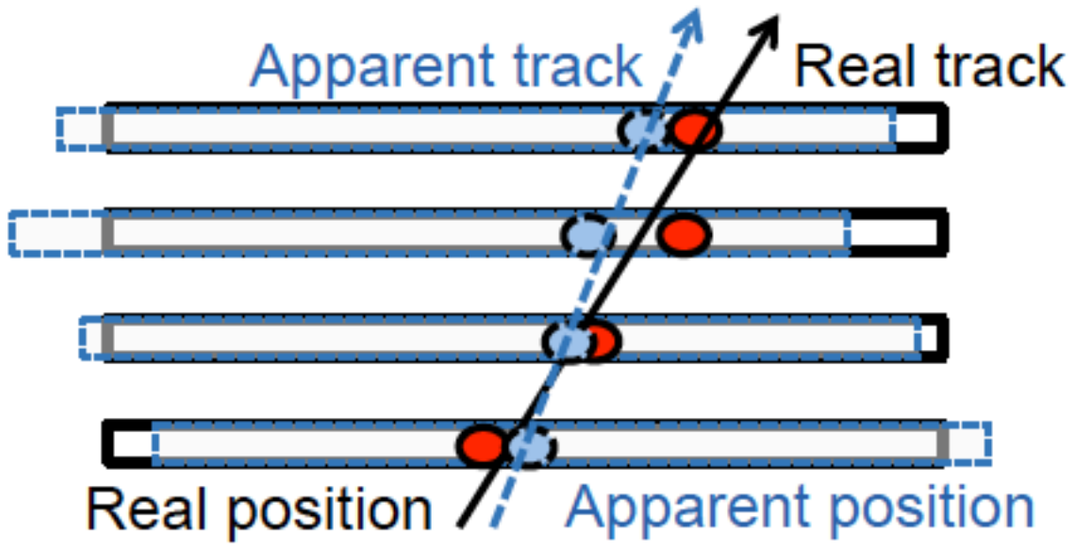
- **Misalignment**
  - Detector elements not positioned in space with perfect accuracy.
  - Alignment corrections derived from data and applied in track reconstruction.



Correcting detector effects - calibration

- **Presence of Material**
  - Coulomb scattering off the core of atoms
  - Energy loss due to ionization
  - Bremsstrahlung
  - Hadronic interaction

- **Misalignment**
  - Detector elements not positioned in space with perfect accuracy.
  - Alignment corrections derived from data and applied in track reconstruction.



Data Preparation

Three major steps to **prepare data for physics analysis** and achieve

- reliable, high quality data (yes, we ***reject*** low quality data)
- the ***best performance*** from our detectors
- readiness for ***physics analysis***

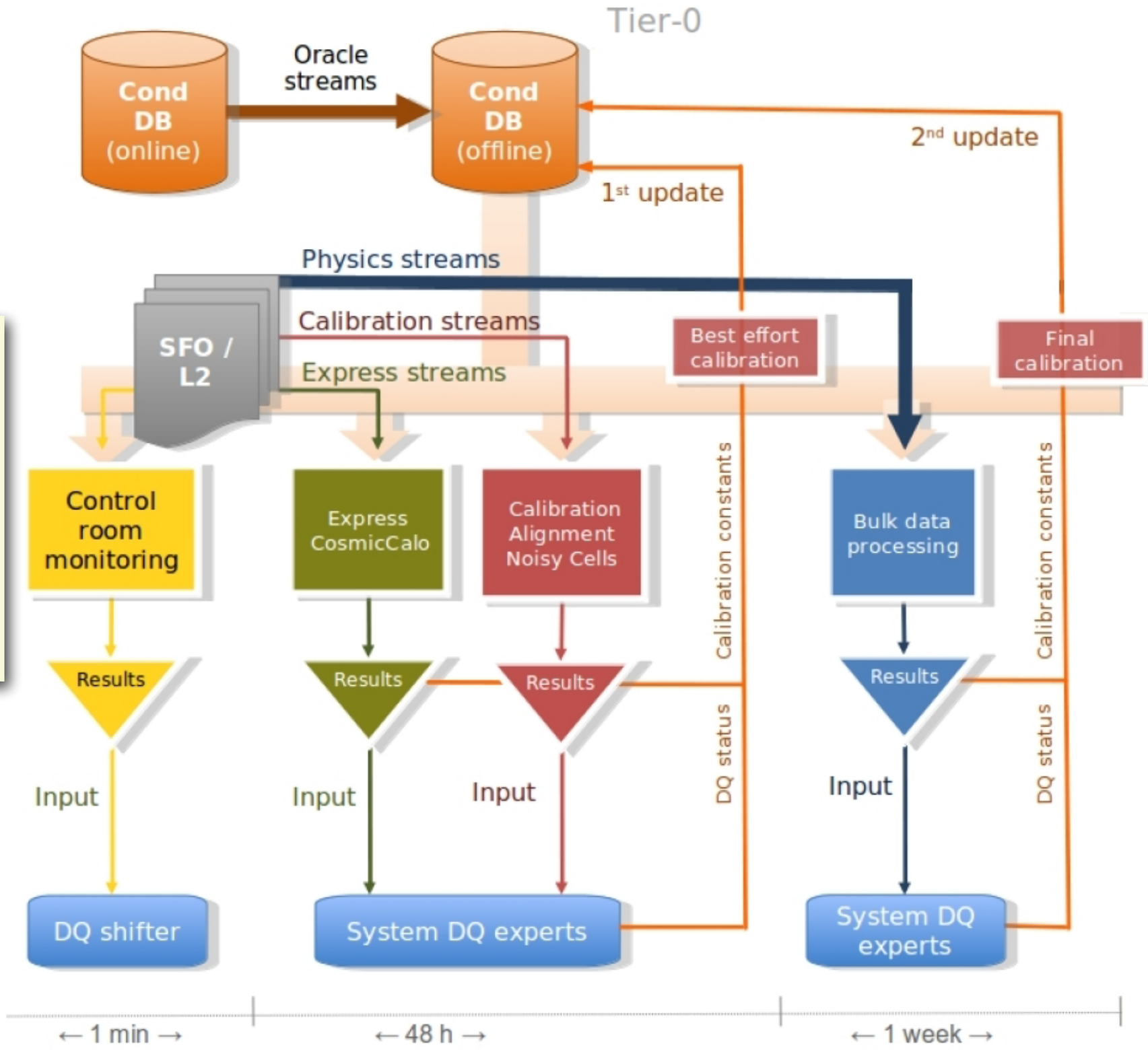
1. **Reconstruct physics signals** from the data ✓
   - Produce information like how many muons does the event have?
2. **Calibrate** the detectors ✓
   - Correct imperfections, account for changes over time…
3. Make sure that the **data quality** is excellent, also in real time
   - Maximise the amount of useful data

Data Quality

Check during data taking

Check a fraction of the data with a quick calibration

Check all of the data with the best calibration - publish this data !!



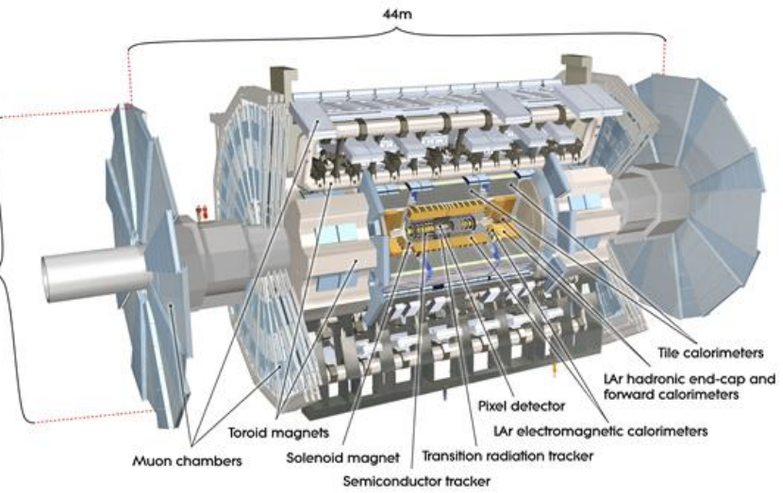
Data Preparation

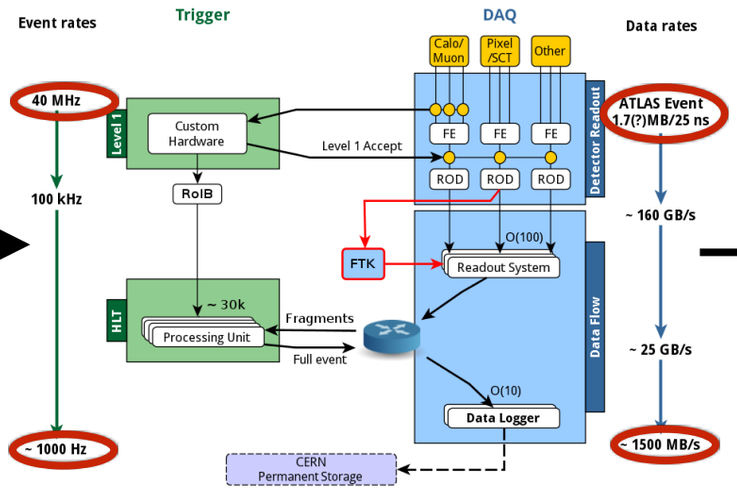
Three major steps to **prepare data for physics analysis** and achieve

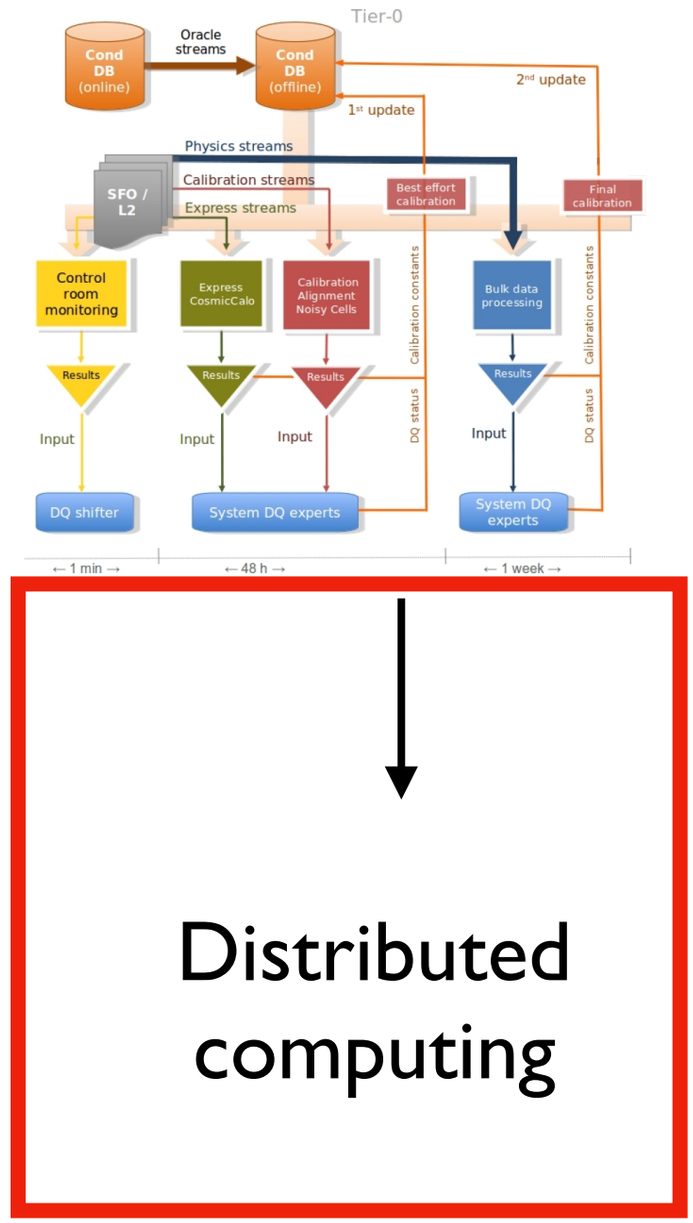
- reliable, high quality data (yes, we ***reject*** low quality data)
- the ***best performance*** from our detectors
- readiness for ***physics analysis***

1. **Reconstruct physics signals** from the data ✓
   - Produce information like how many muons does the event have?
2. **Calibrate** the detectors ✓
   - Correct imperfections, account for changes over time…
3. Make sure that the **data quality** is excellent, also in real time
   - Maximise the amount of useful data ✓

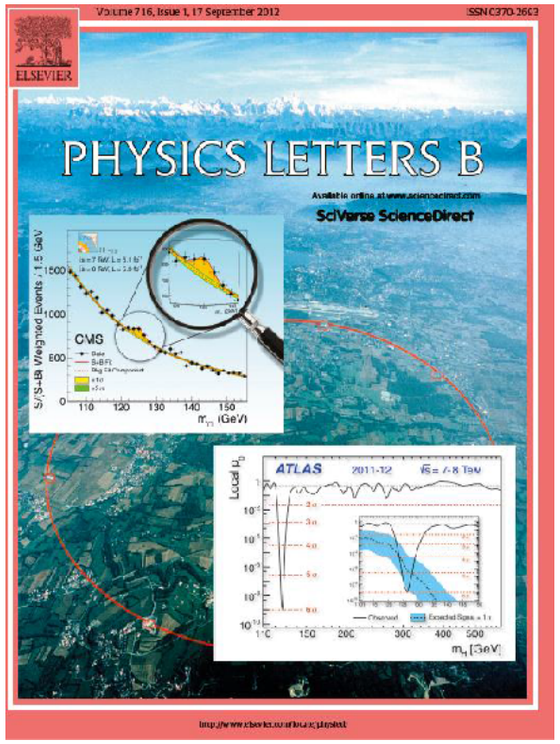
Data's journey

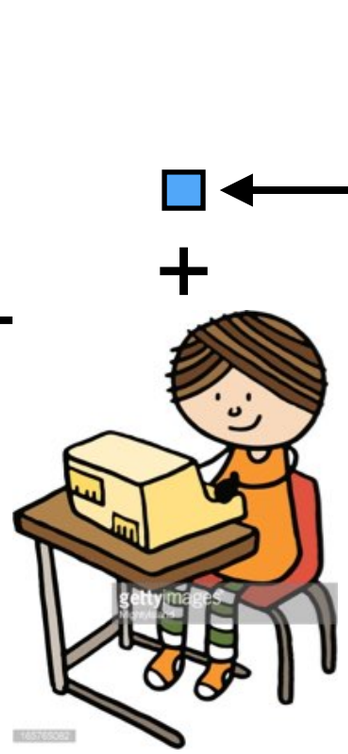


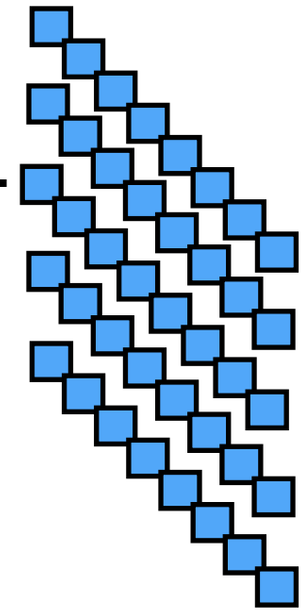




Distributed computing







The Worldwide LHC Computing Grid

- Now the data has been ***prepared for physics analysis***, it's time to extract our favourite physics signal!

- Many experiments, particularly those at the **LHC**, use computing sites all over the world via **the grid** to
  - harness all of that ***computing power***
  - enable collaborators ***worldwide*** to access the data

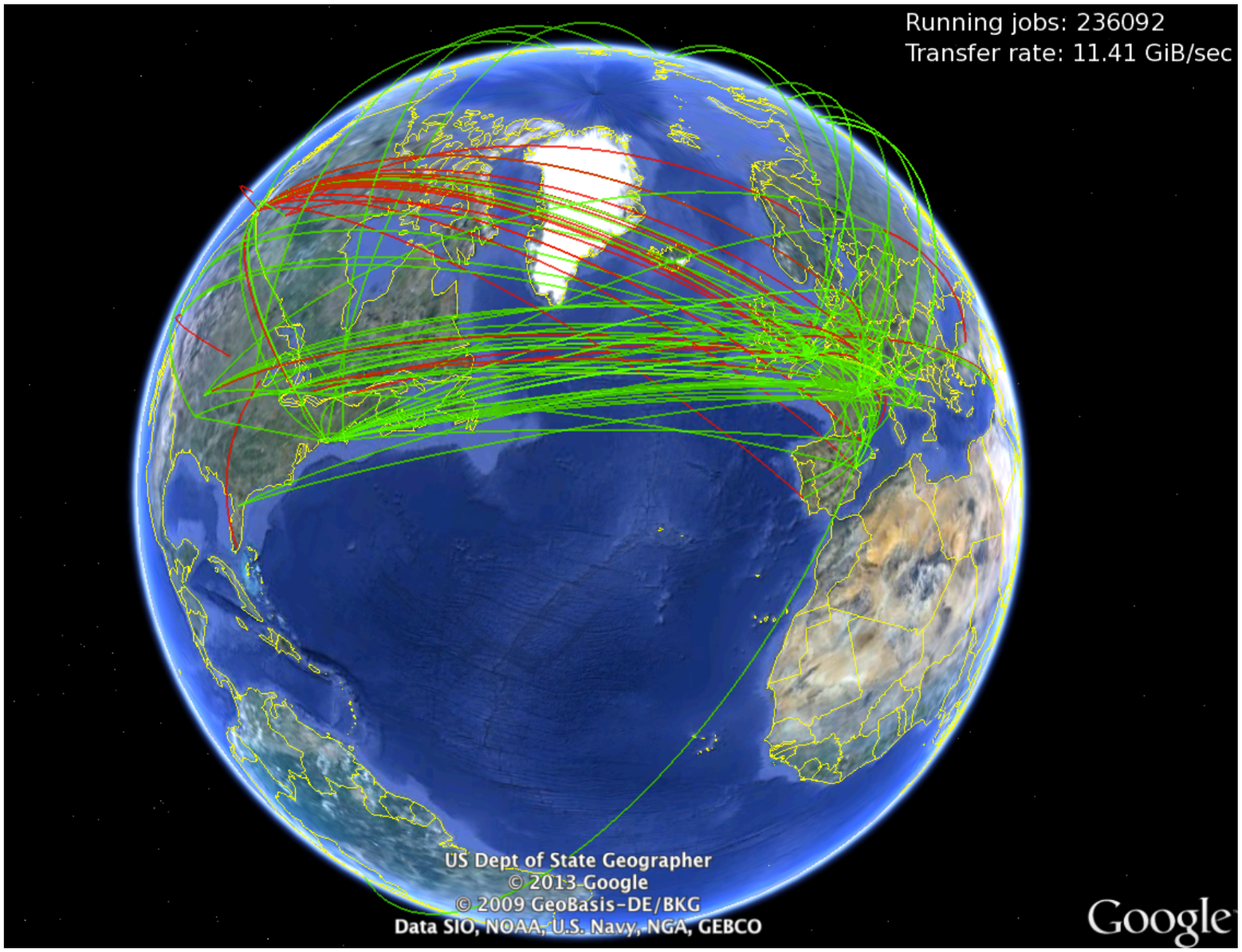
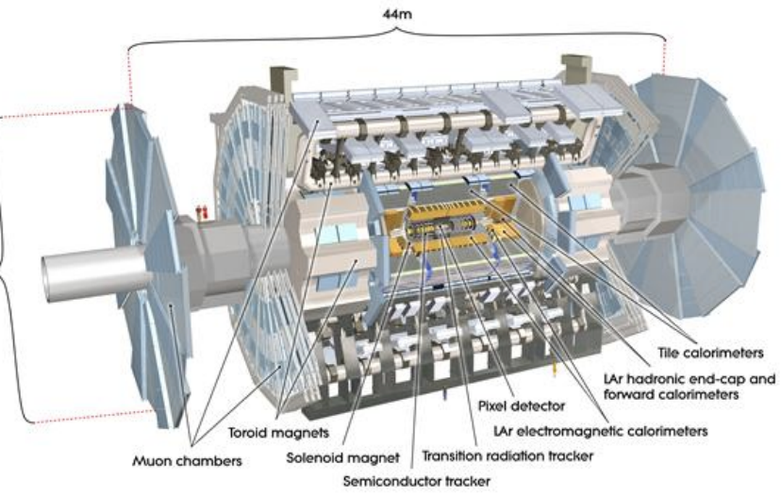


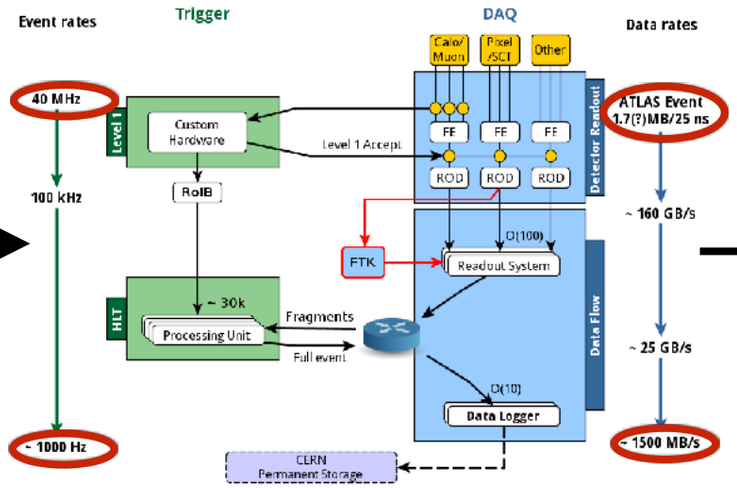
Image taken from the **WLCG** GoogleEarth Dashboard

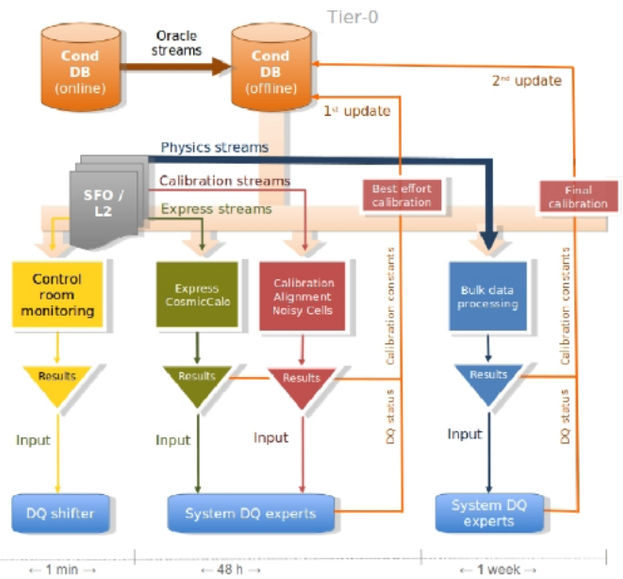
- http://wlcg.web.cern.ch/wlcg-google-earth-dashboard

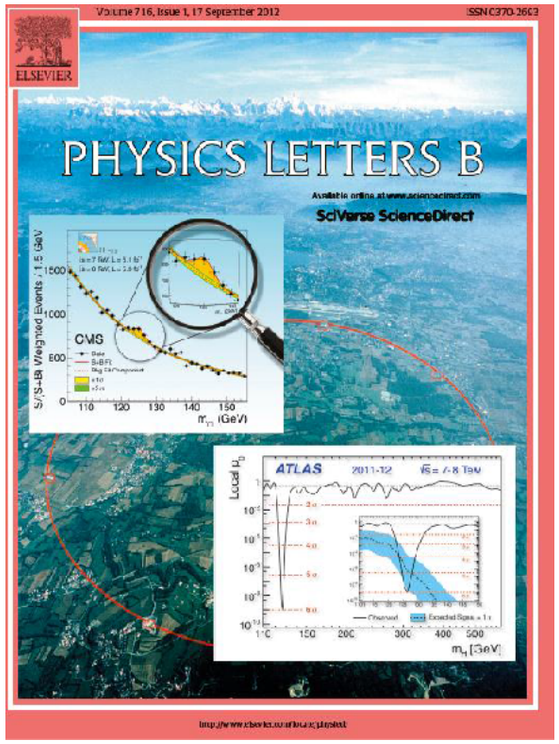
Brookhaven National Laboratory

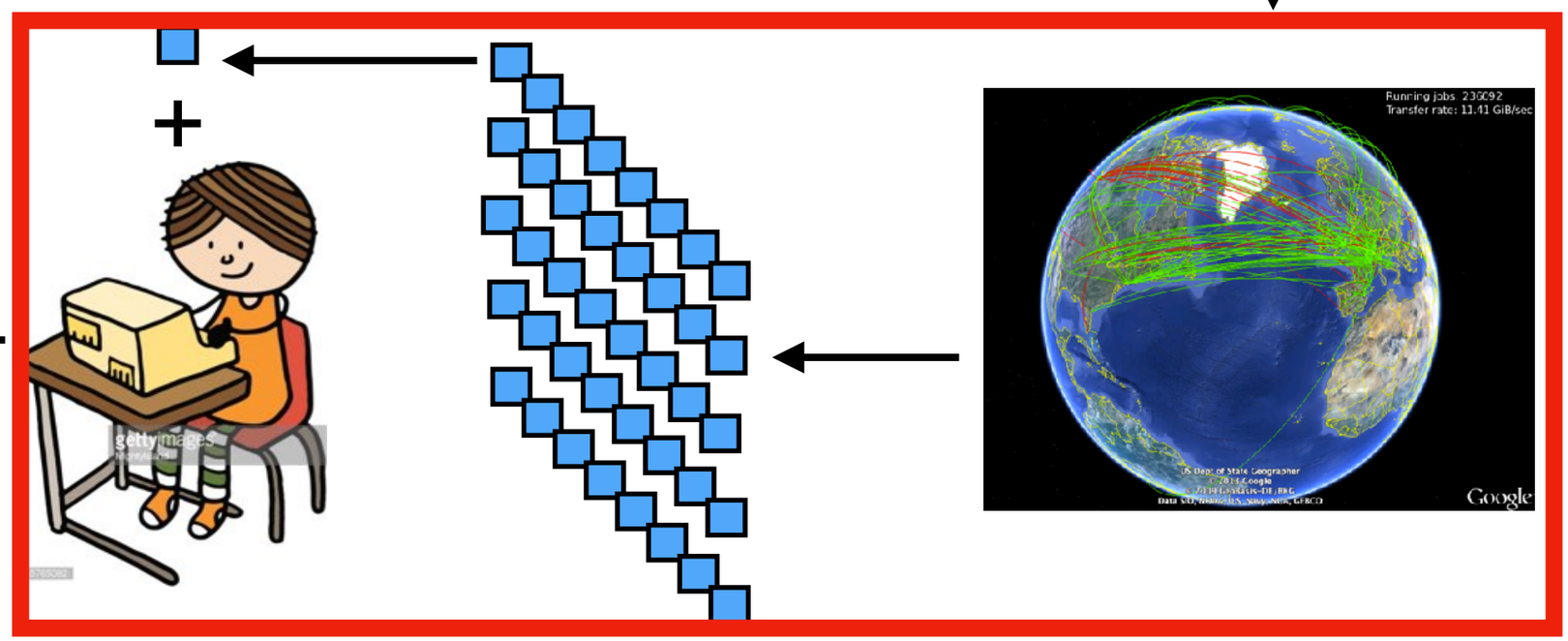
Data's journey











We did it !

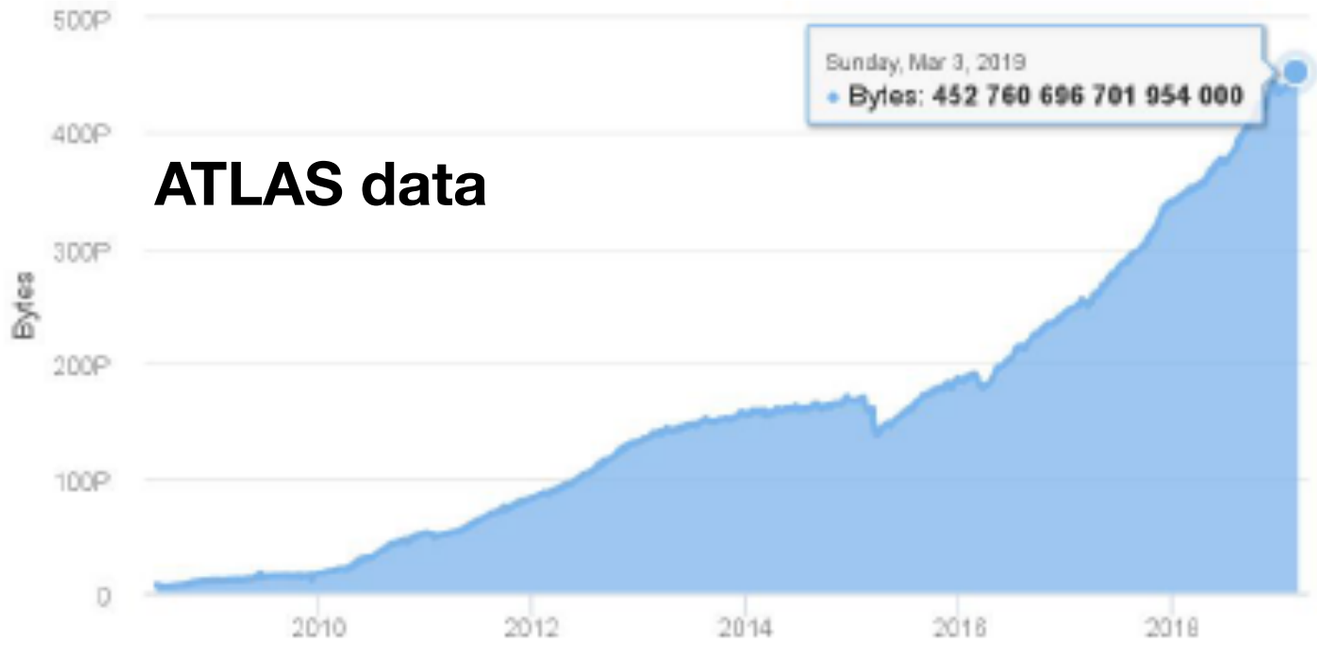
Our data is calibrated and with good data quality
and we've reconstructed the physics objects in the data

- ***it is reliable, accurate and ready for physics analysis***

More detail on these topics in Lecture 2

Now we can extract our measurements in Lecture 3

***Question**: How long would it take to read all of the ATLAS data? (Assume for simplicity you have off-the-shelf SSDs with read speed ~500MB/s)*



Brookhaven National Laboratory

Contact details

Dr Paul Laycock, based at CERN
Office: B40 4-C-16

email: laycock@bnl.gov or paul.james.laycock@cern.ch