Brookhaven National Laboratory

U.S. DEPARTMENT OF ENERGY

Computing Challenges in ~~Nuclear and~~ Particle Physics

Paul Laycock

August 3rd 2021

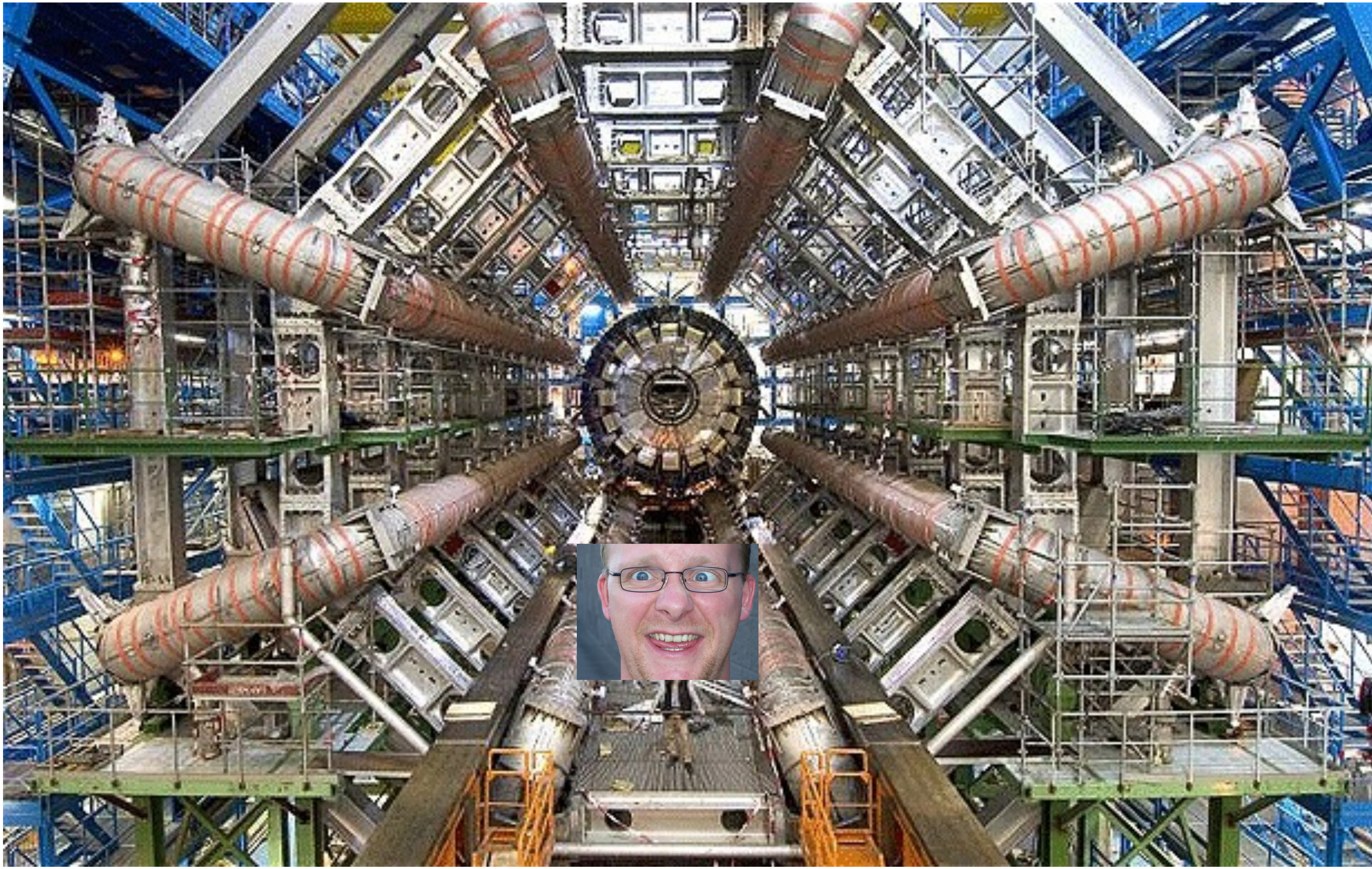
Particle physics and data

	BaBar	H1	ZEUS	HERMES	Belle	BESIII	CDF	DØ
End of data taking	07.04.08	30.06.07	30.06.07	30.06.07	30.06.10	2017	30.09.11	30.09.11
Type of data to be preserved	RAW data Sim/rec level Data skims in ROOT	RAW data Sim/rec level Analysis level ROOT data	Flat ROOT based ntuples	RAW data Sim/rec level Analysis level ROOT data	RAW data Sim/rec level	RAW data Sim/rec level ROOT data	RAW data Rec. level ROOT files (data+MC)	Raw data Rec. level ROOT files (data+MC)
Data Volume	2 PB	0.5 PB	0.2 PB	0.5 PB	4 PB	6 PB	9 PB	8.5 PB
Desired longevity of long term analysis	Unlimited	At least 10 years	At least 20 years	5-10 years	5 years	15 years	Unlimited	10 years

The ATLAS and CMS experiments already have 100s of petabytes of data

Brookhaven National Laboratory

Particle Physics and me



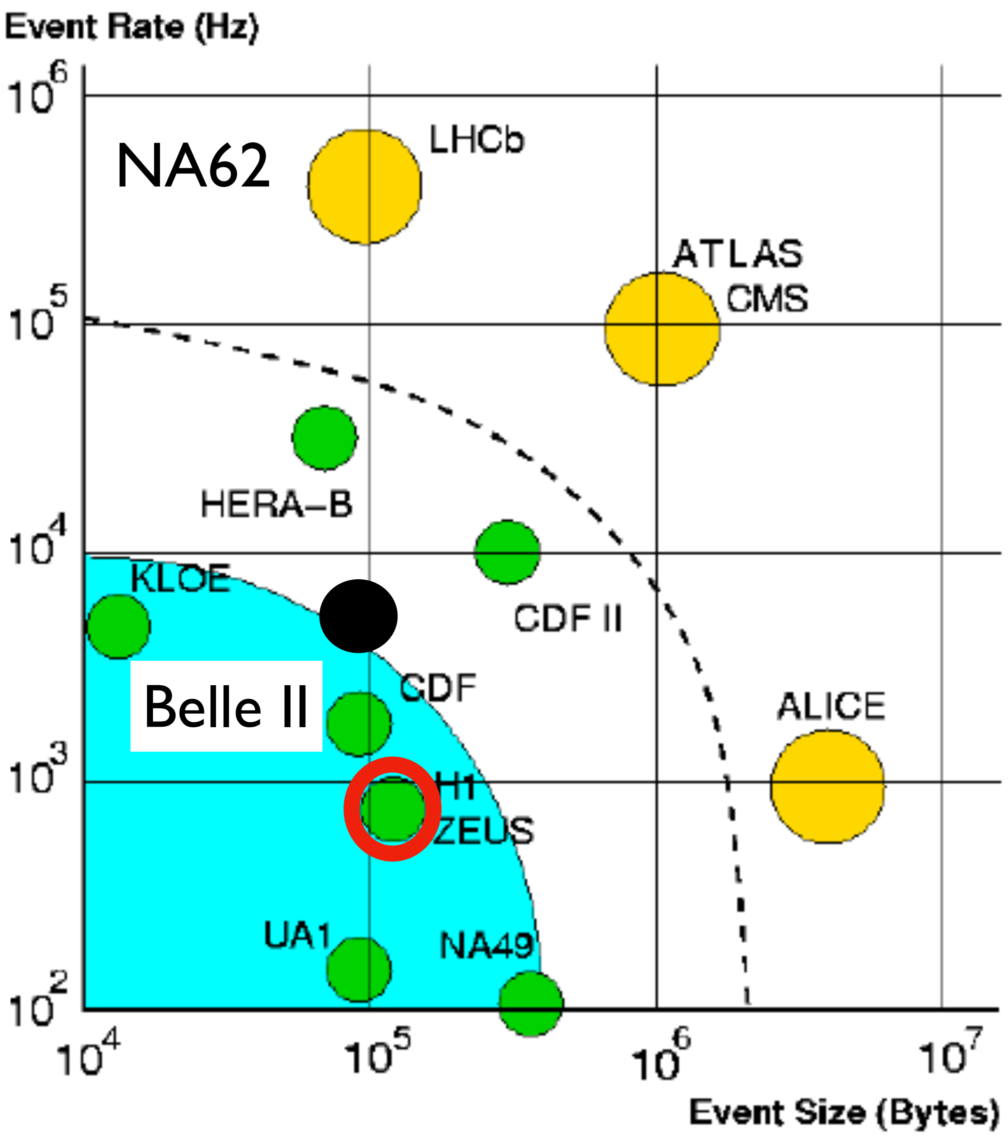
Physics Data

Plot modified from:
"GridPP: development of the UK computing Grid for particle physics."

H1

- Proton structure and QCD
- small event sizes and rates

DAQ throughput =
event rate * event size



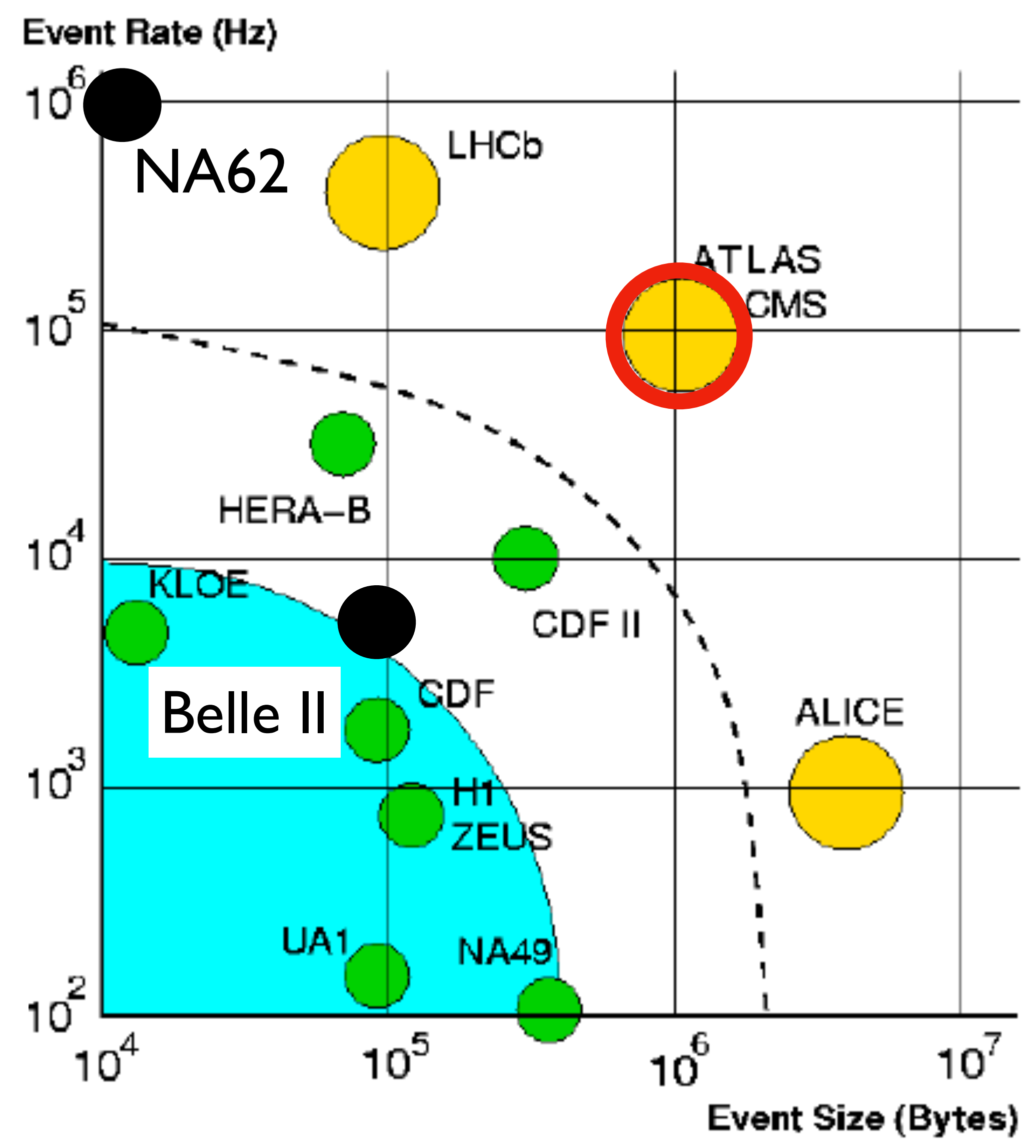
Physics Data

Plot modified from:
"GridPP: development of the UK computing Grid for particle physics."

H1
- Proton structure and QCD
- small event sizes and rates

ATLAS
- Higgs, searches for new physics
- big event sizes and rates

DAQ throughput =
event rate * event size



Physics Data

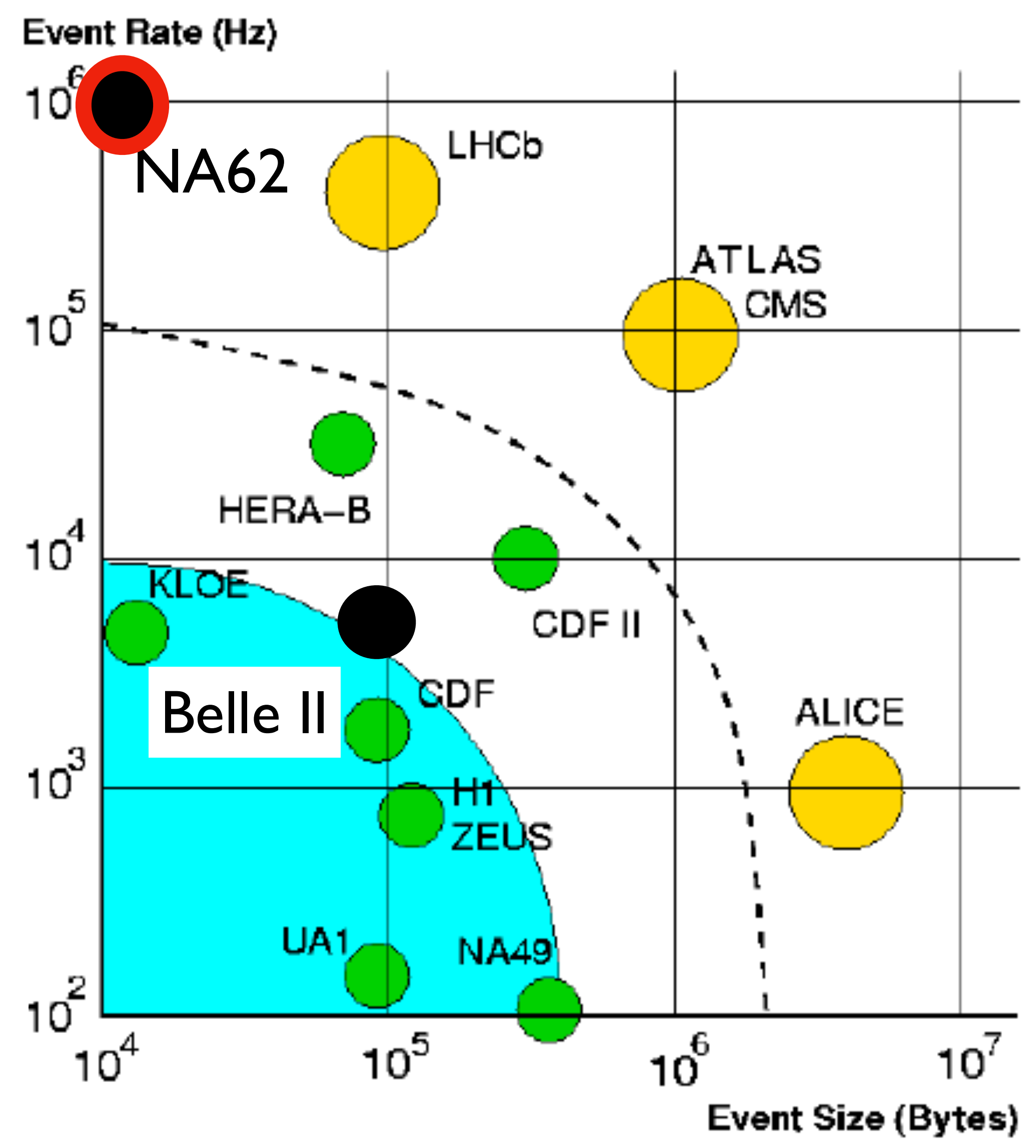
Plot modified from:
"GridPP: development of the UK computing Grid for particle physics."

H1
- Proton structure and QCD
- small event sizes and rates

ATLAS
- Higgs, searches for new physics
- big event sizes and rates

NA62
- Ultra-rare kaon decays
- huge rates of small size events

DAQ throughput =
event rate * event size



Physics Data

Plot modified from:
"GridPP: development of the UK computing Grid for particle physics."

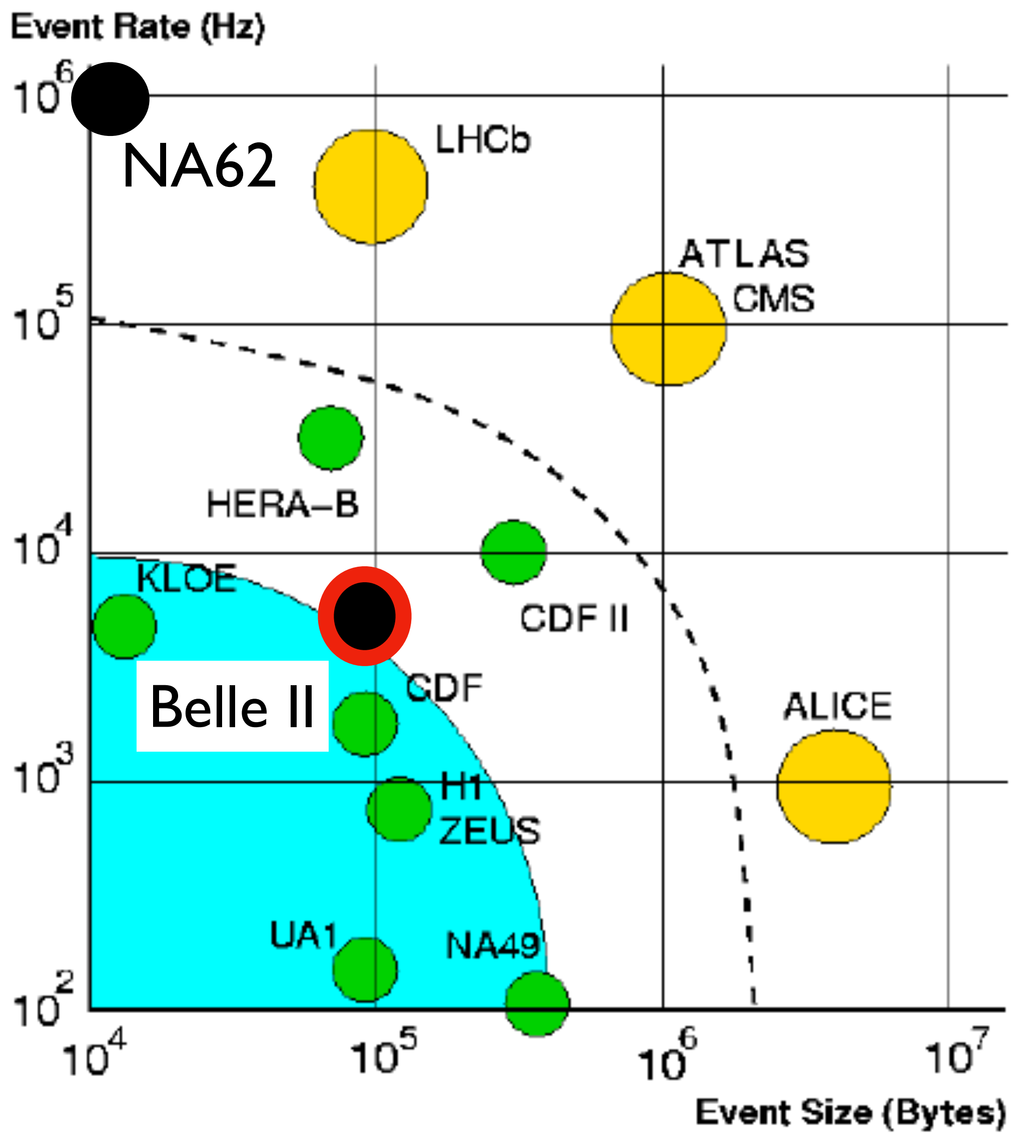
H1
- Proton structure and QCD
- small event sizes and rates

ATLAS
- Higgs, searches for new physics
- big event sizes and rates

NA62
- Ultra-rare kaon decays
- huge rates of small size events

Belle II
- Ultra-rare B decays
- modest event sizes and rates

DAQ throughput = event rate * event size



Physics Data

Plot modified from:
"GridPP: development of the UK computing Grid for particle physics."

H1
- Proton structure and QCD
- small event sizes and rates

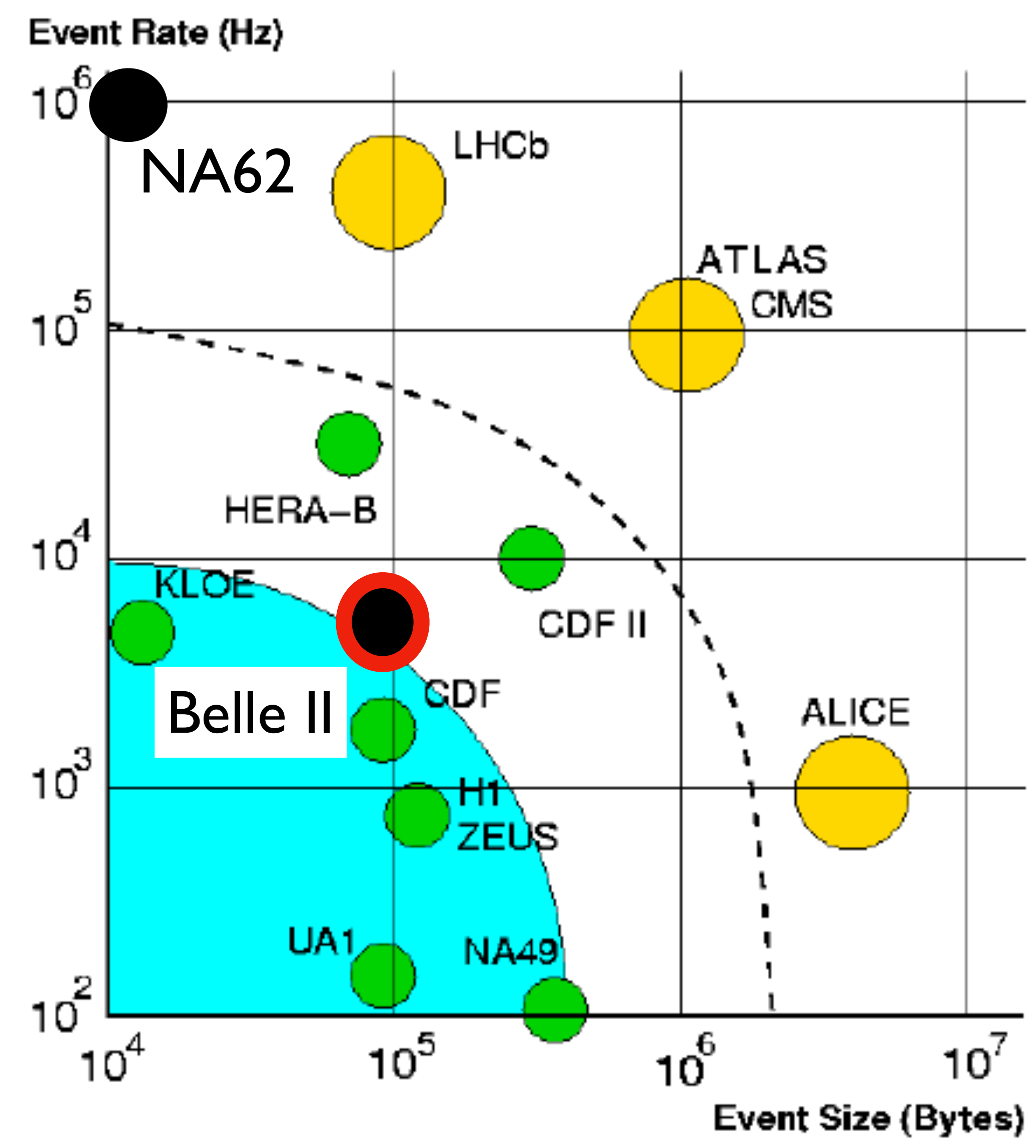
ATLAS
- Higgs, searches for new physics
- big event sizes and rates

NA62
- Ultra-rare kaon decays
- huge rates of small size events

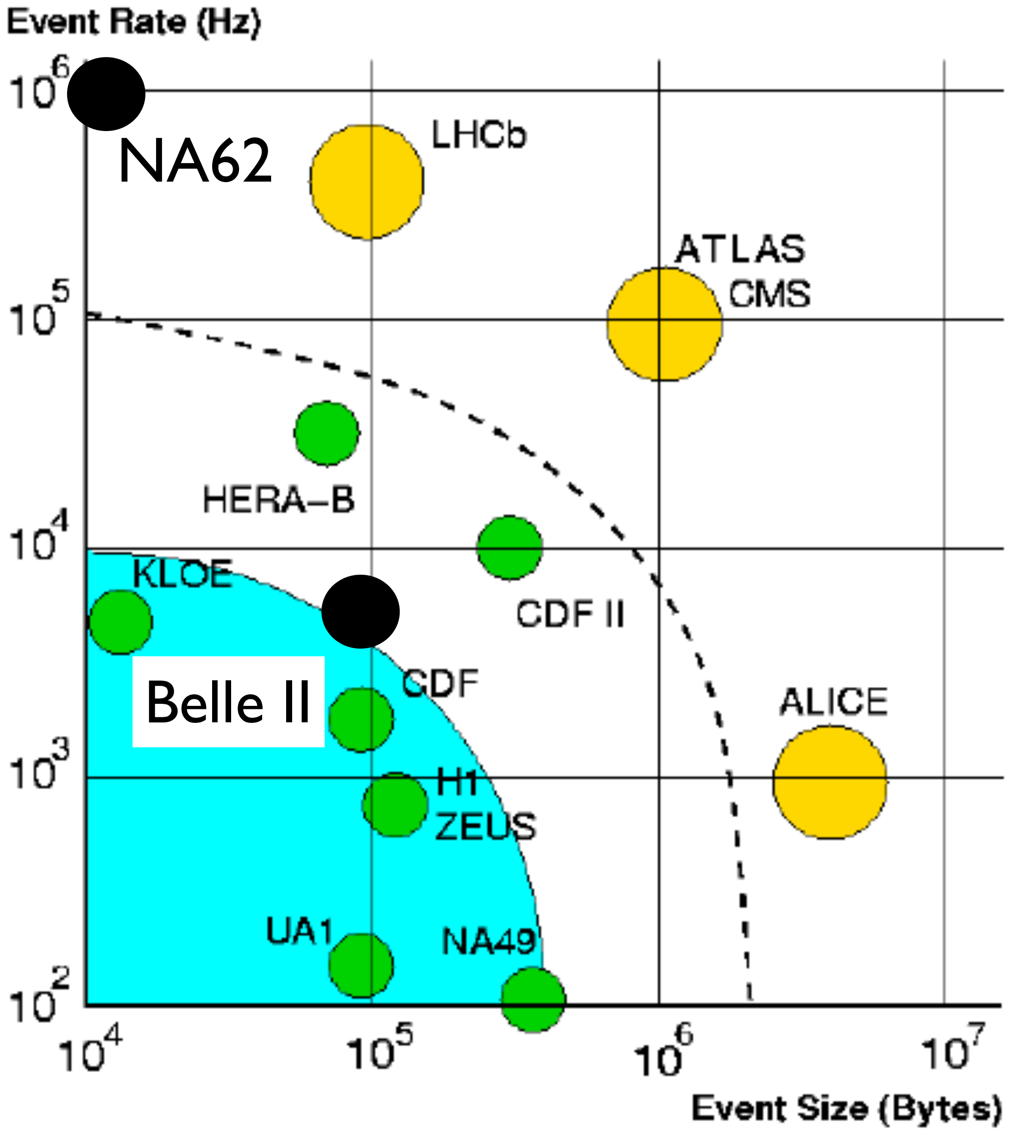
Belle II
- Ultra-rare B decays
- modest event sizes and rates

<u>Question: What about DUNE??</u>

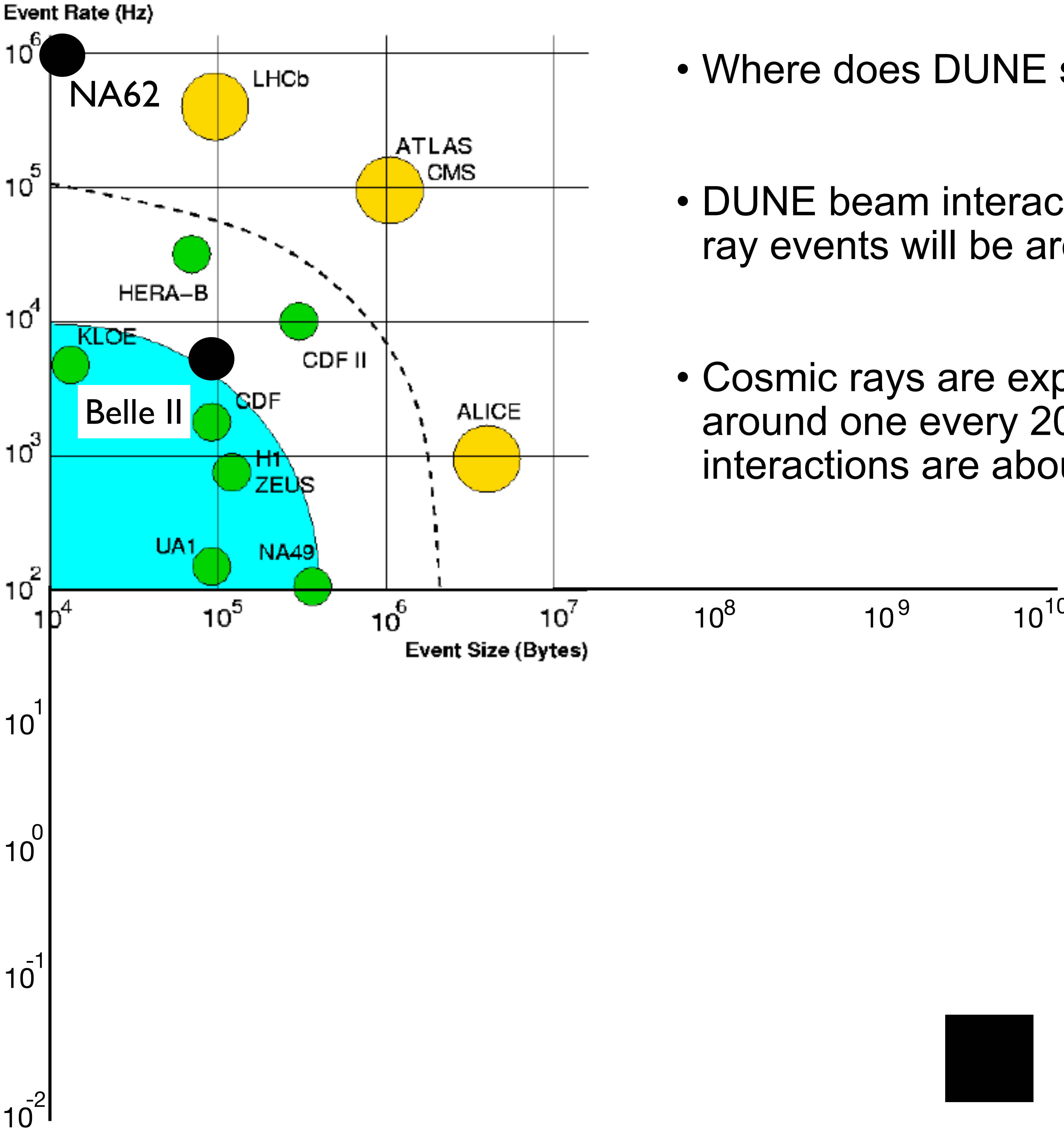
DAQ throughput = event rate * event size



Brookhaven National Laboratory

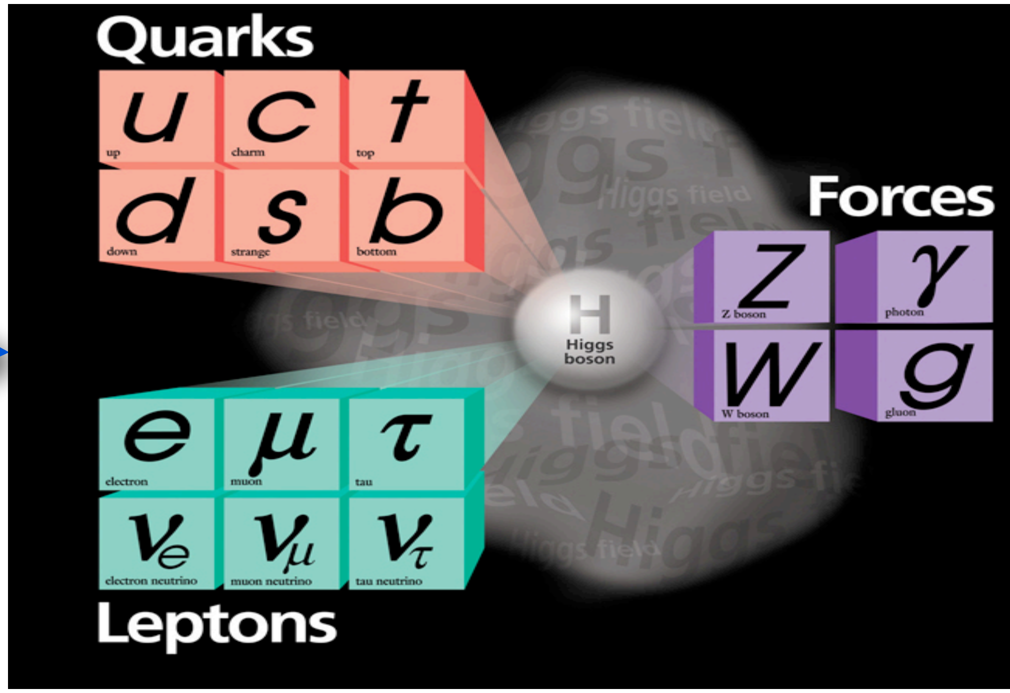


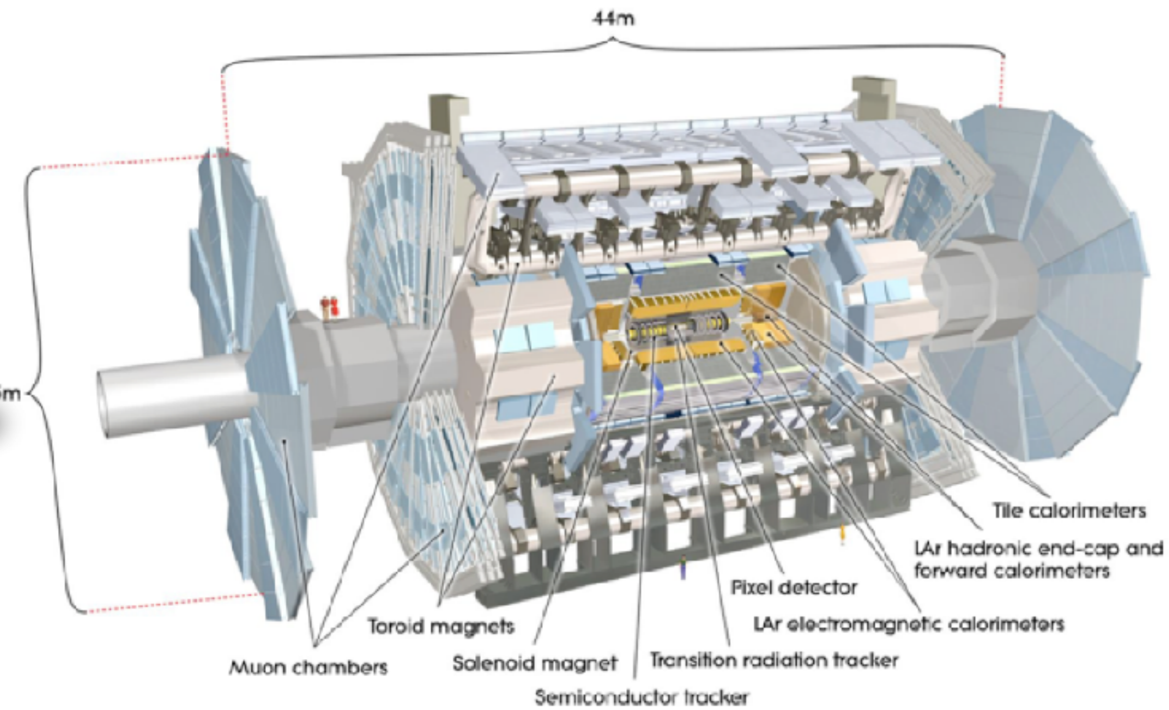
- Where does DUNE sit on this figure?
- DUNE beam interactions and cosmic ray events will be ??? Bytes in size
- Cosmic rays are expected at a rate of ????

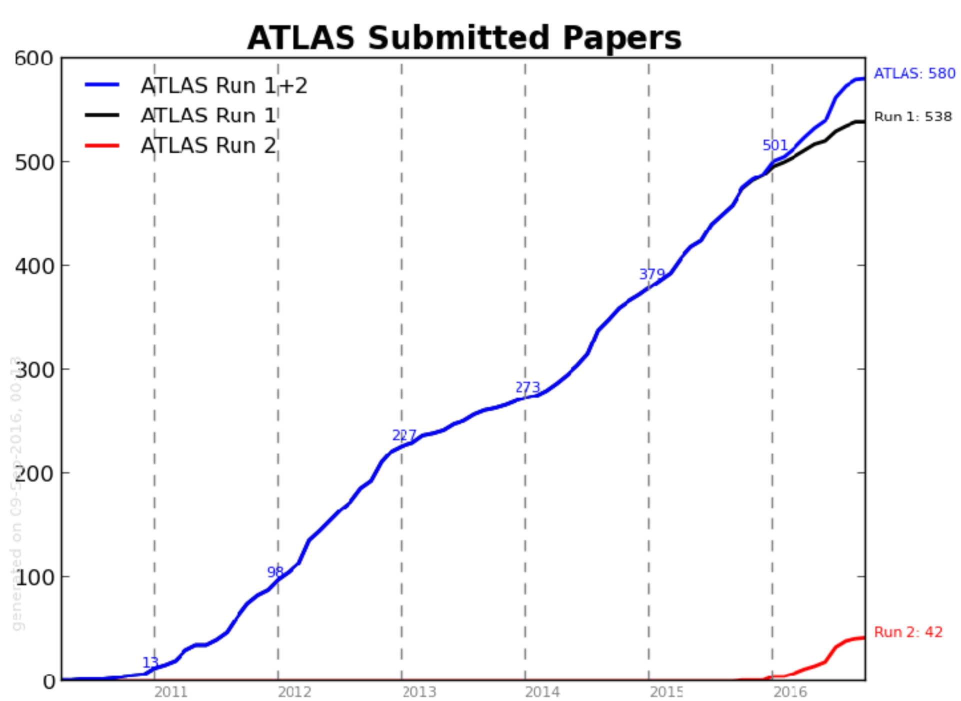


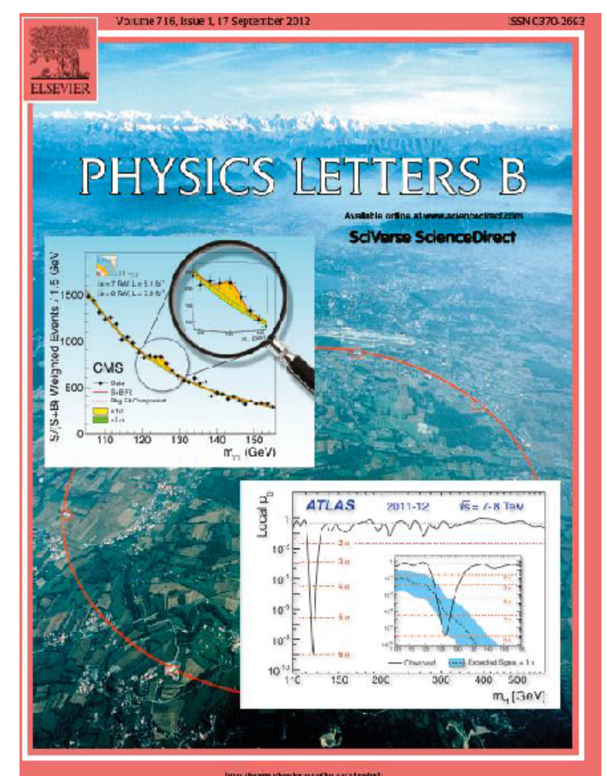
- Where does DUNE sit on this figure?
- DUNE beam interactions and cosmic ray events will be around 6GB in size
- Cosmic rays are expected at a rate of around one every 20s (beam interactions are about 40 / day)

The particle physics cycle

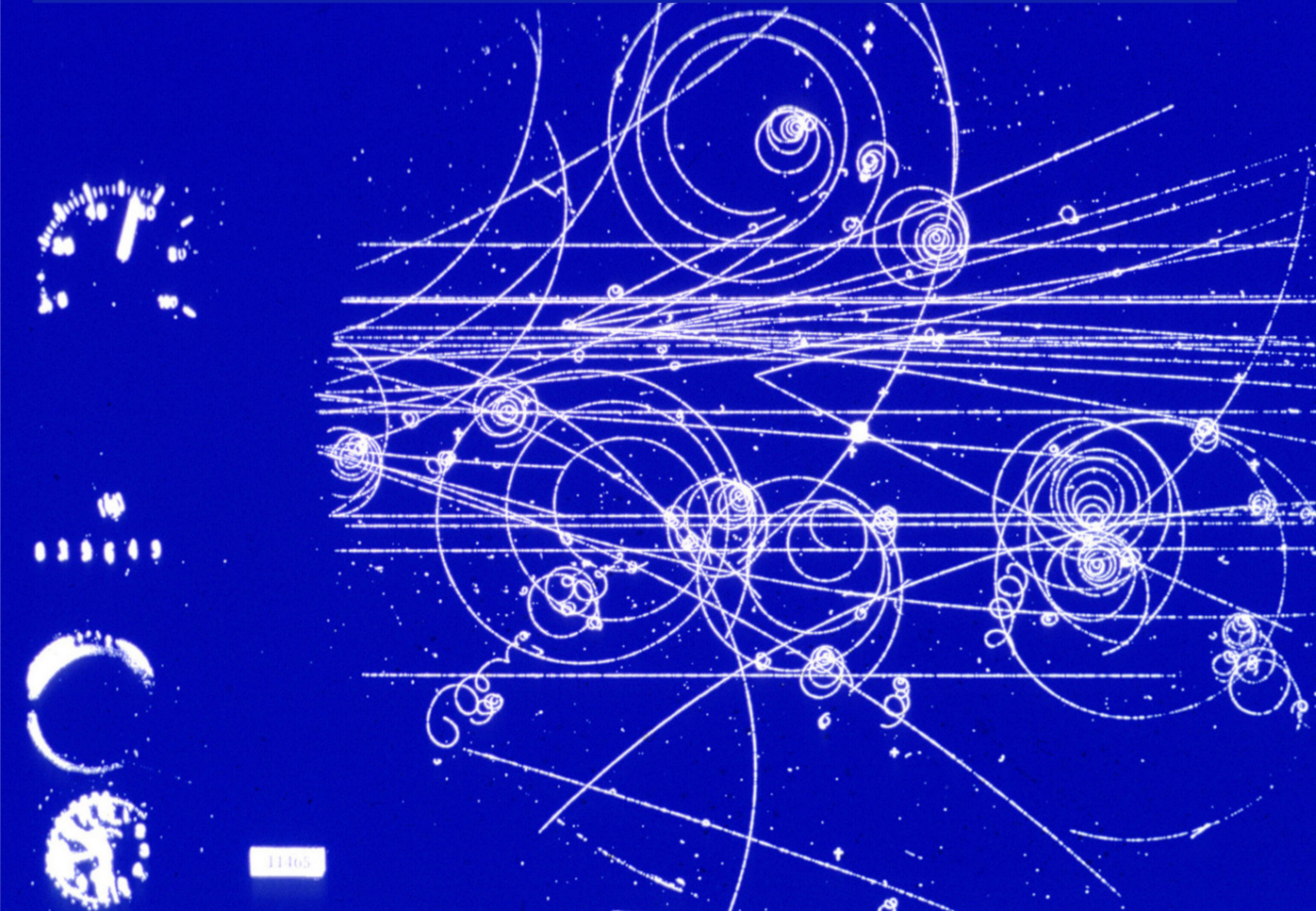




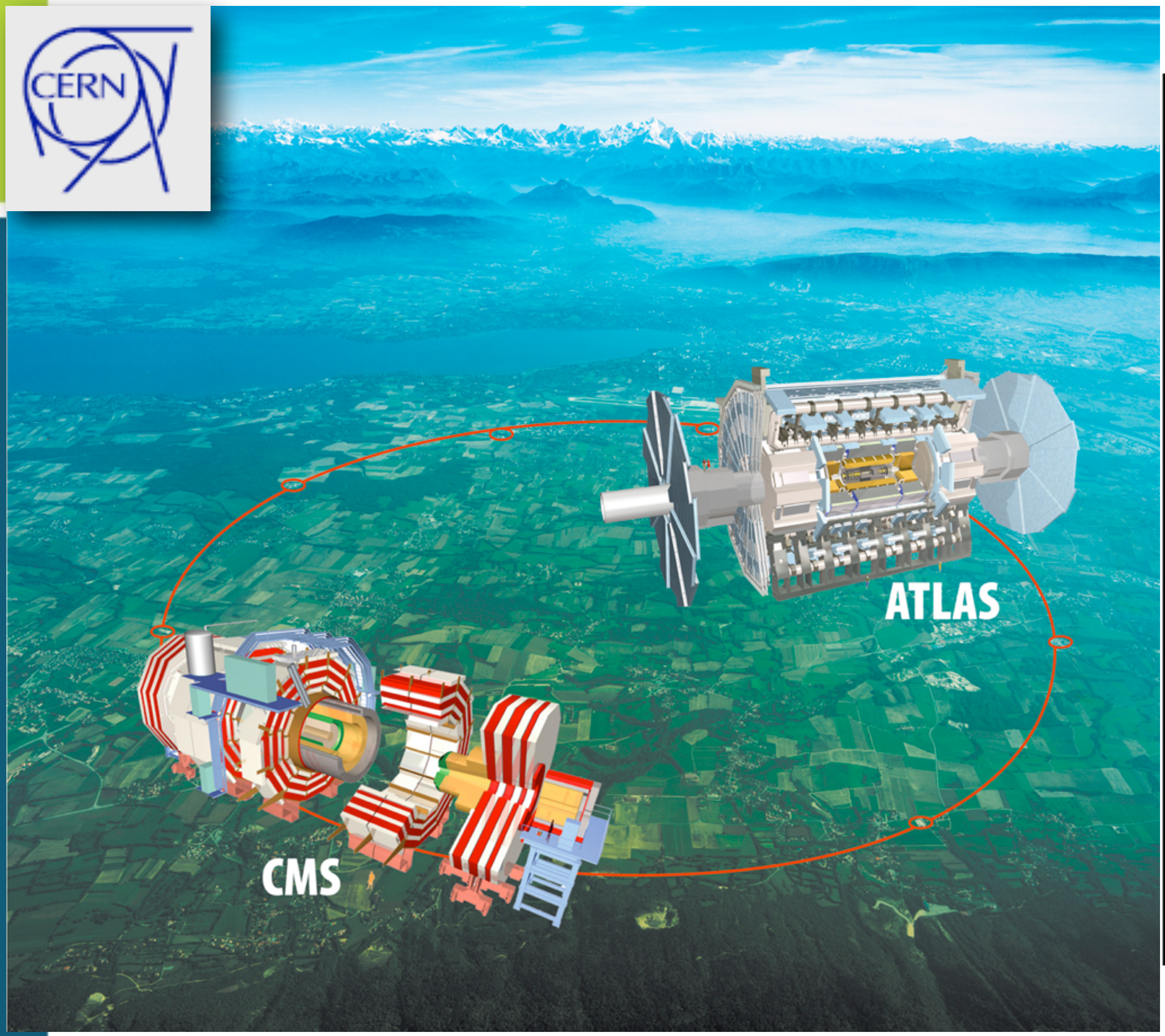




… data analysis used to involve a person looking at pictures



ATLAS and the Large Hadron Collider



I'll use ATLAS as an example of a modern experiment
similar things are done on other experiments

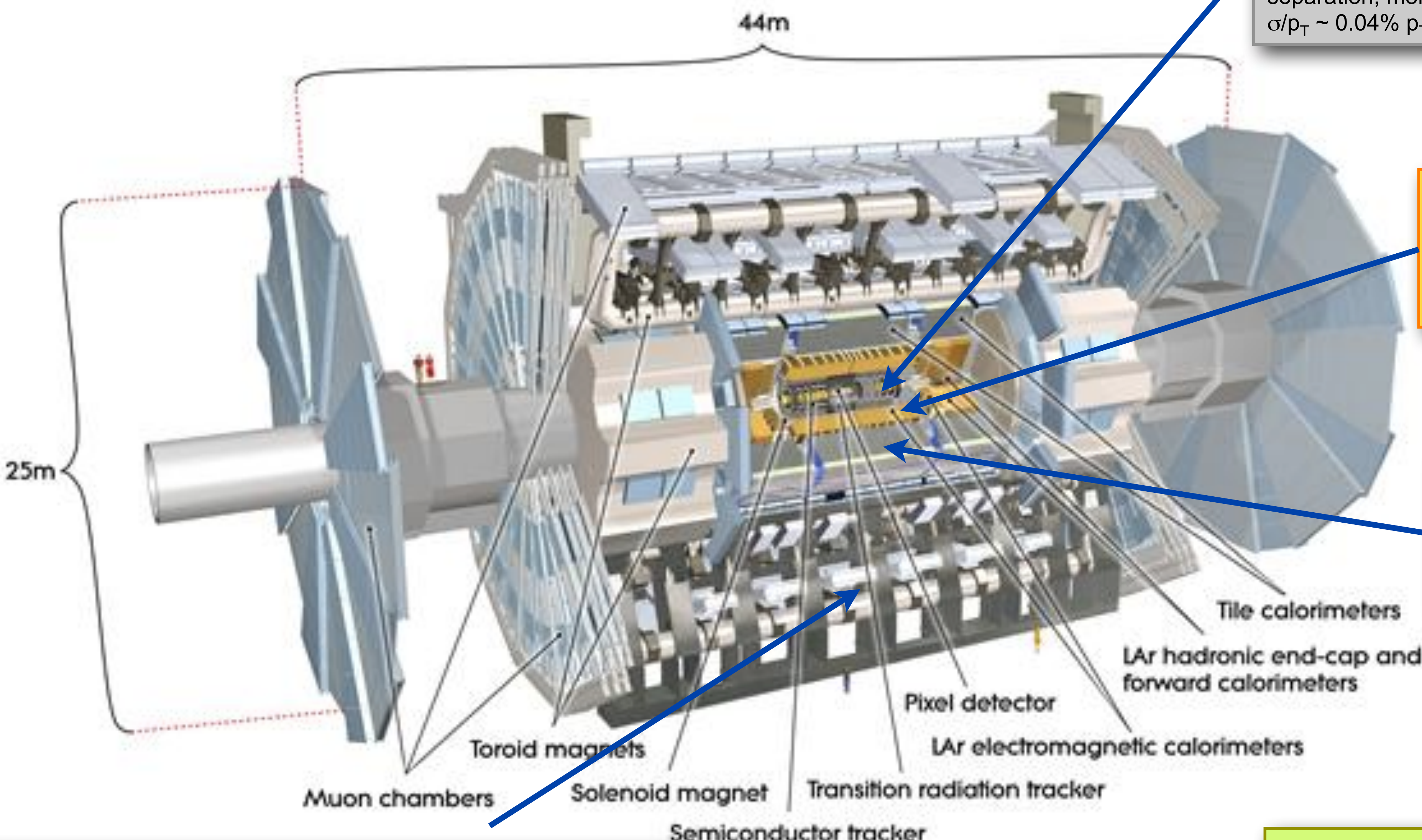
The ATLAS Detector @ LHC

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

Inner Tracker ($|\eta|<2.5$, B=2T):
Si Pixels, Si strips, Trans. Rad. Det.
Precise tracking and vertexing, e/π separation, momentum resolution:
$\sigma/p_T \sim 0.04\%\ p_T\,(\text{GeV}) \oplus 1.5\%$

EM calorimeter:
Pb-LAr Accordion, e/γ trigger, id. and meas., energy res.: $\sigma/E \sim 10\%/\sqrt{E} \oplus 0.7\%$

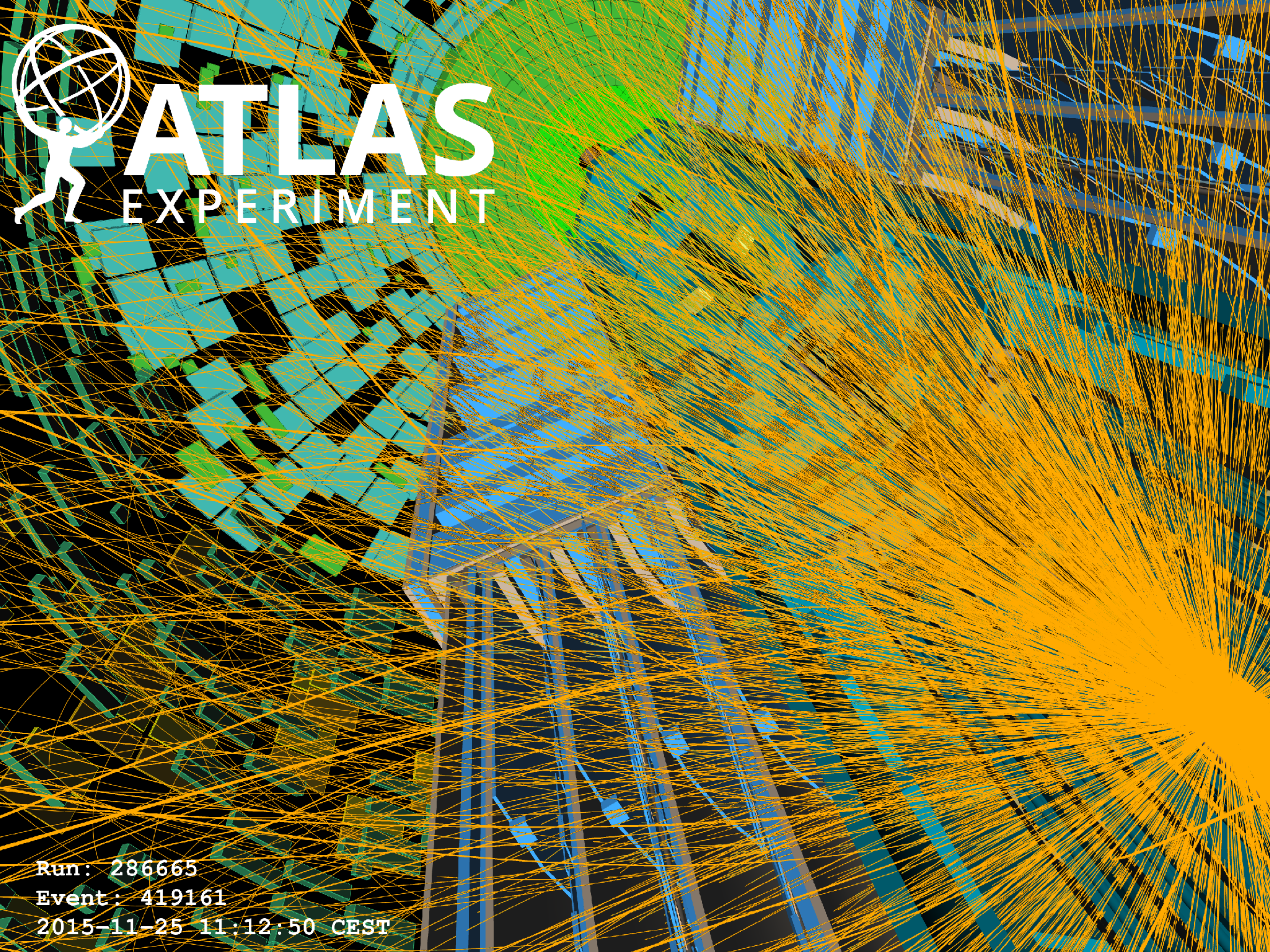
HAD calorimetry ($|\eta|<5$): Fe/scintillator Tiles (cen), Cu/W-LAr (fwd). trigger and meas. of jets and $E_{T,miss}$, energy res.: $\sigma/E \sim 50\%/\sqrt{E} \oplus 3\%$



Muon Spectrometer: air-core toroids with gas-based muon chambers. trigger and meas. with momentum resolution < 10% up to $E_\mu \sim 1$ 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”


ATLAS
EXPERIMENT
Run: 286665
Event: 419161
2015-11-25 11:12:50 CEST

Analysing a lot of data

ATLAS has about 500 PBs of data

ATLAS data

Now analyse it !

***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
Hint: there are about 30 million seconds / year)

Brookhaven National Laboratory

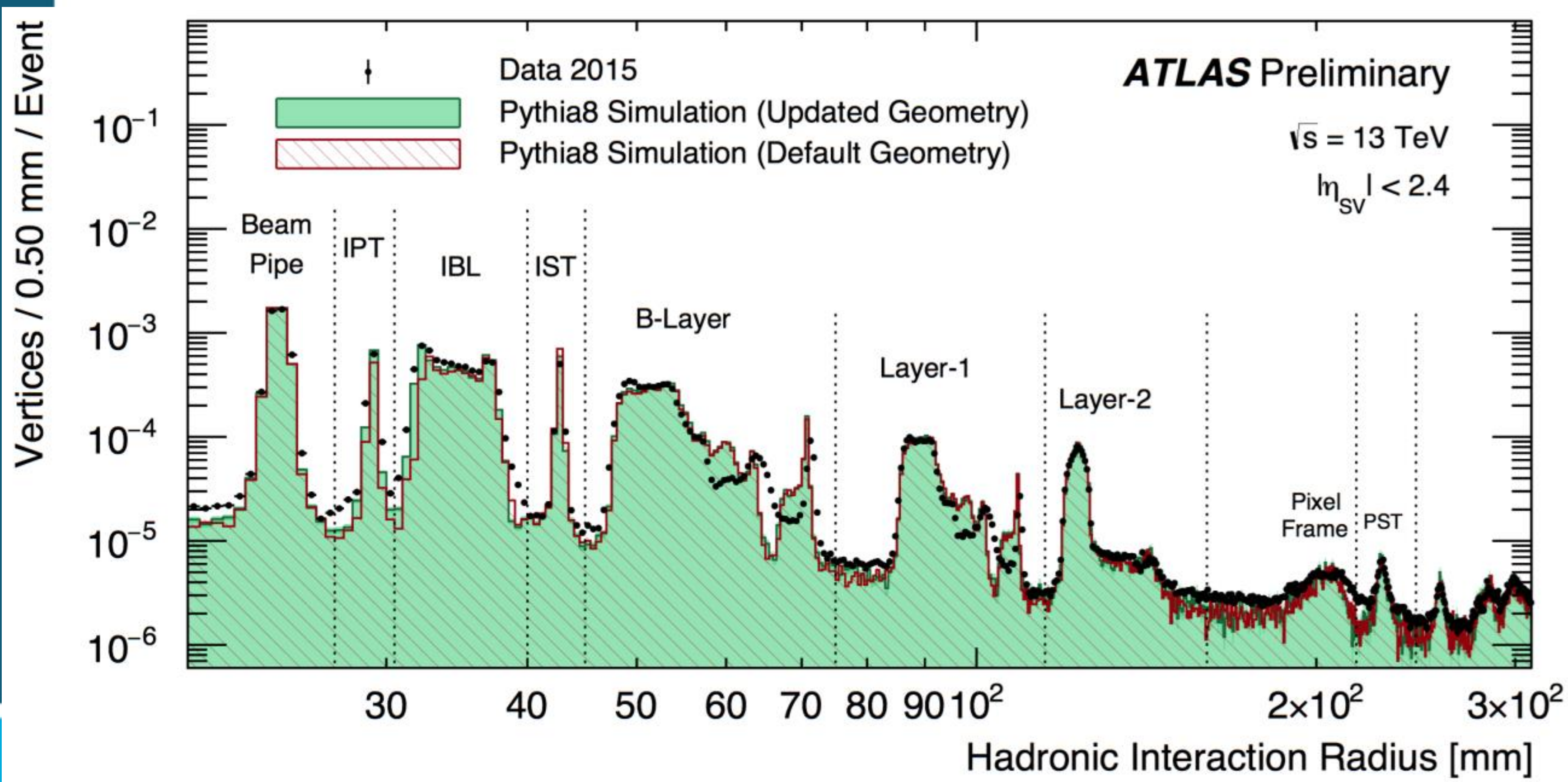
Simulation and understanding detectors

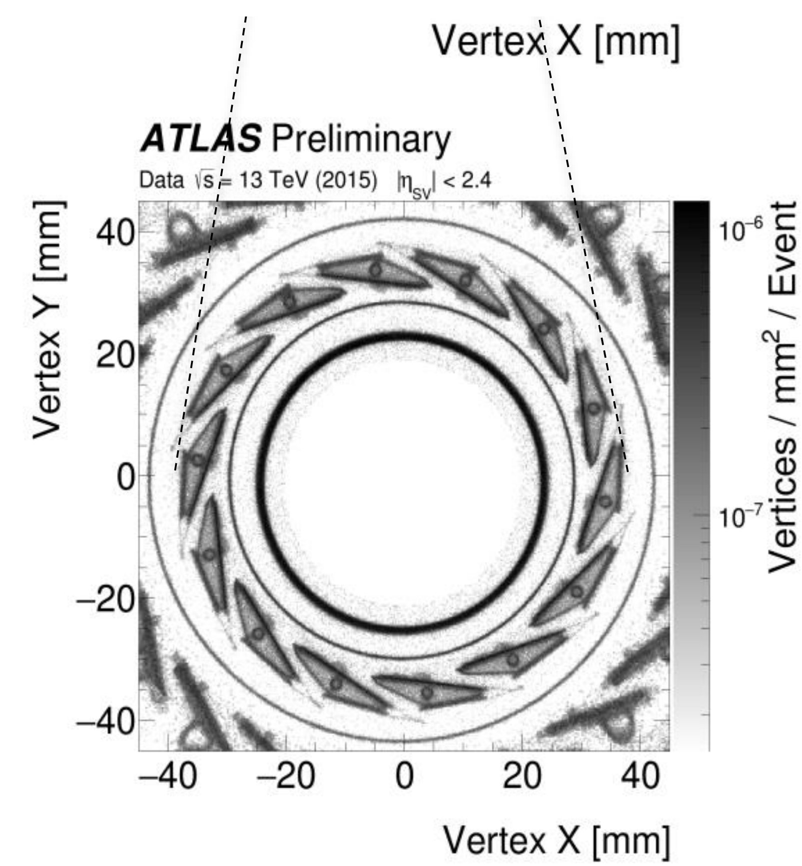
We use **simulations** to model the detector as **accurately** and **precisely** as possible

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

We correct our simulations if necessary

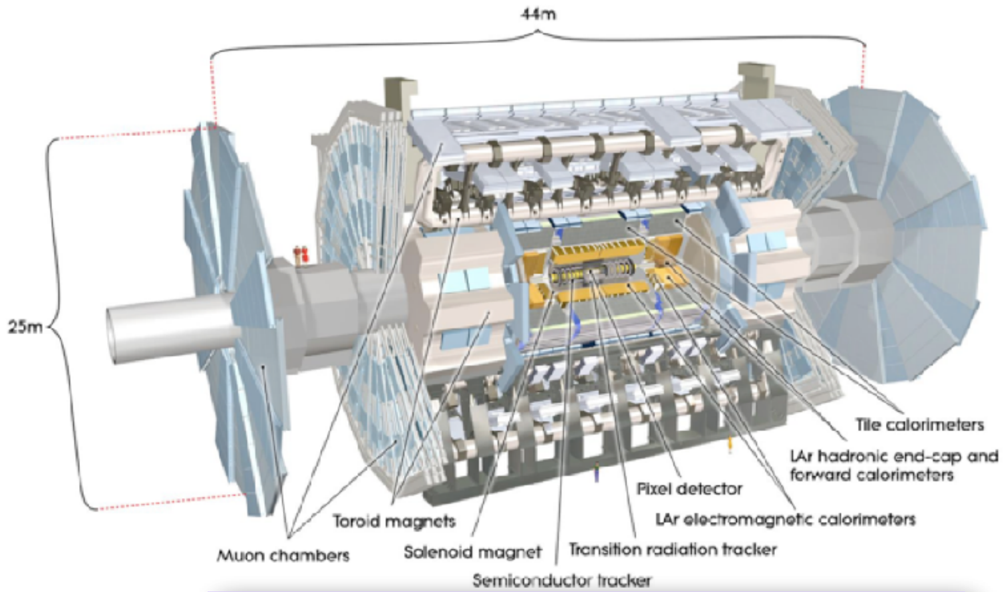
- Once our simulation is an ***accurate model*** of our detector, we can use it to ***correct the data for detector response***





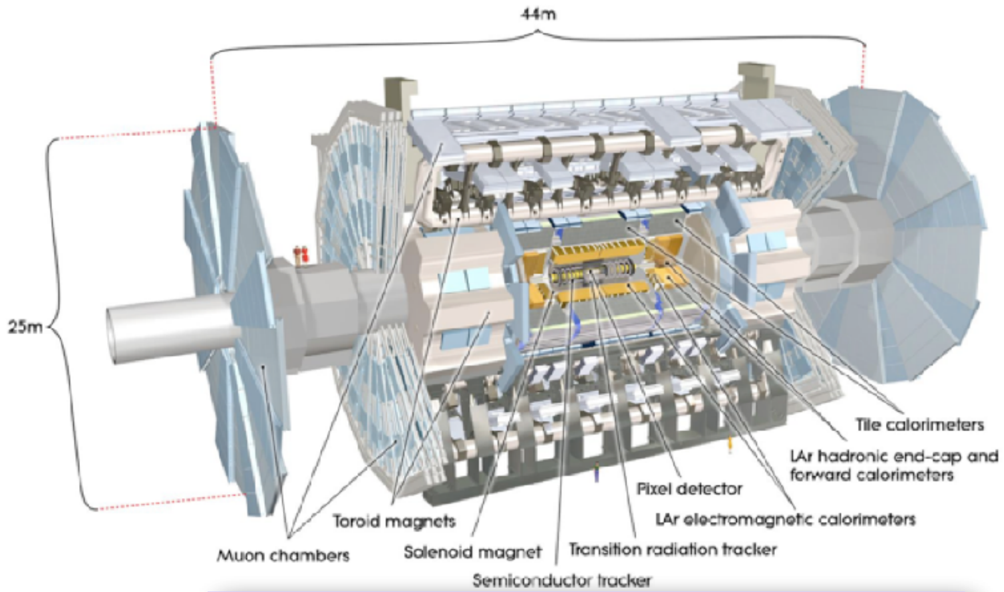
Brookhaven National Laboratory

Exabyte-scale physics analysis

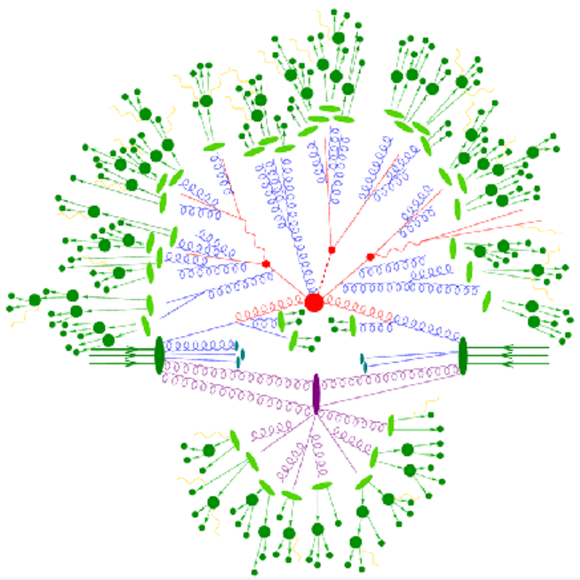


Exabytes of Data

Exabyte-scale physics analysis

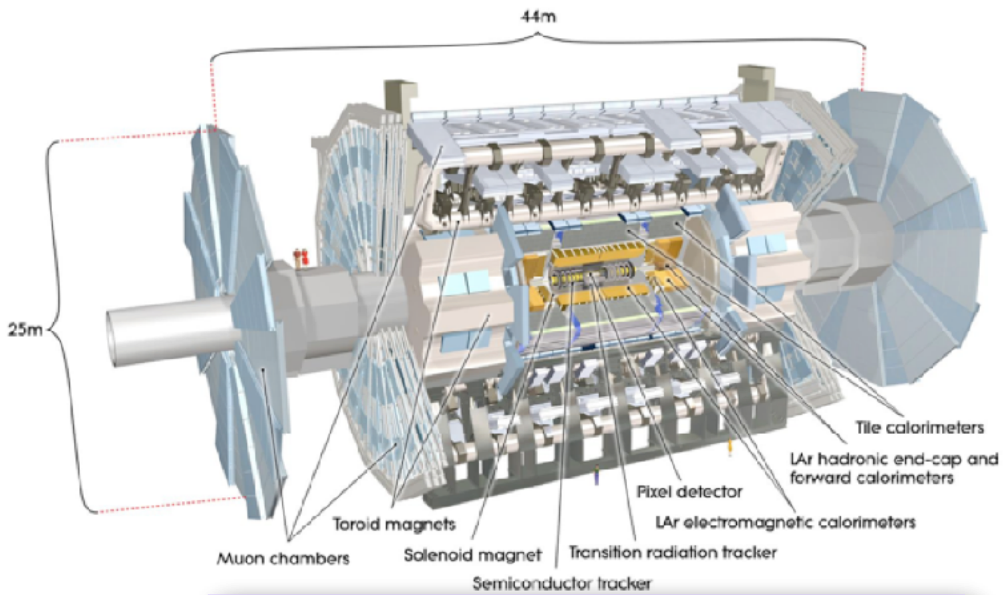


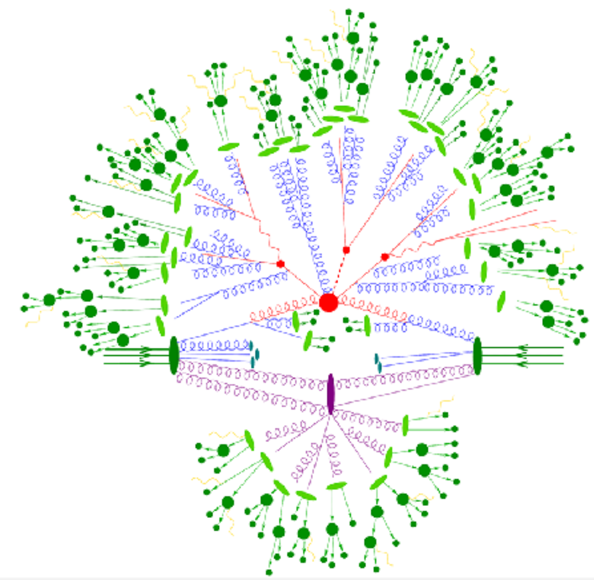
Exabytes of Data



Exabytes of Simulation

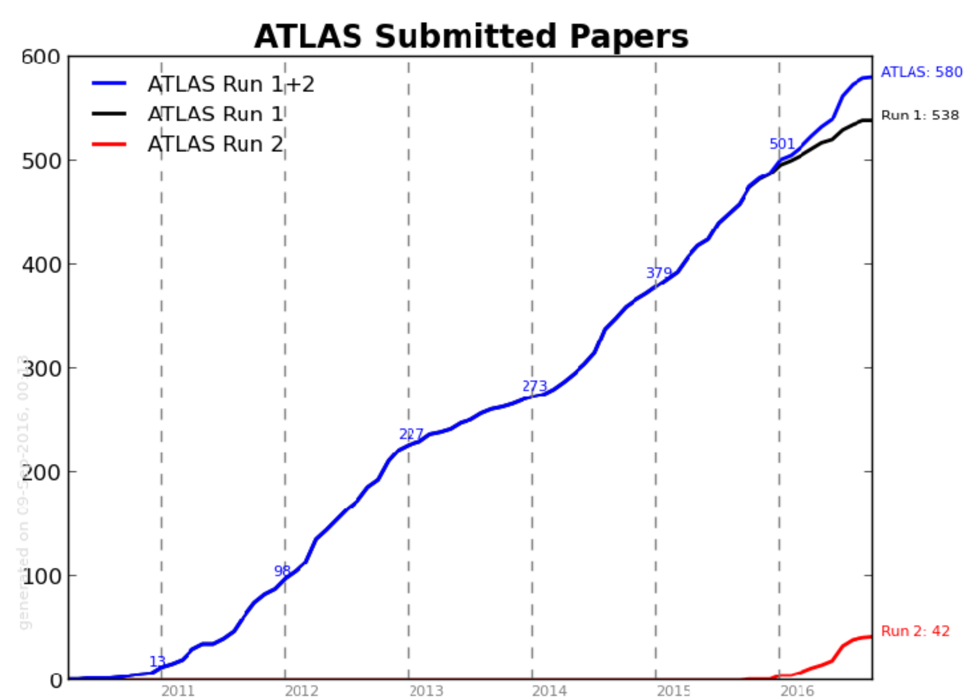
Exabyte-scale physics analysis





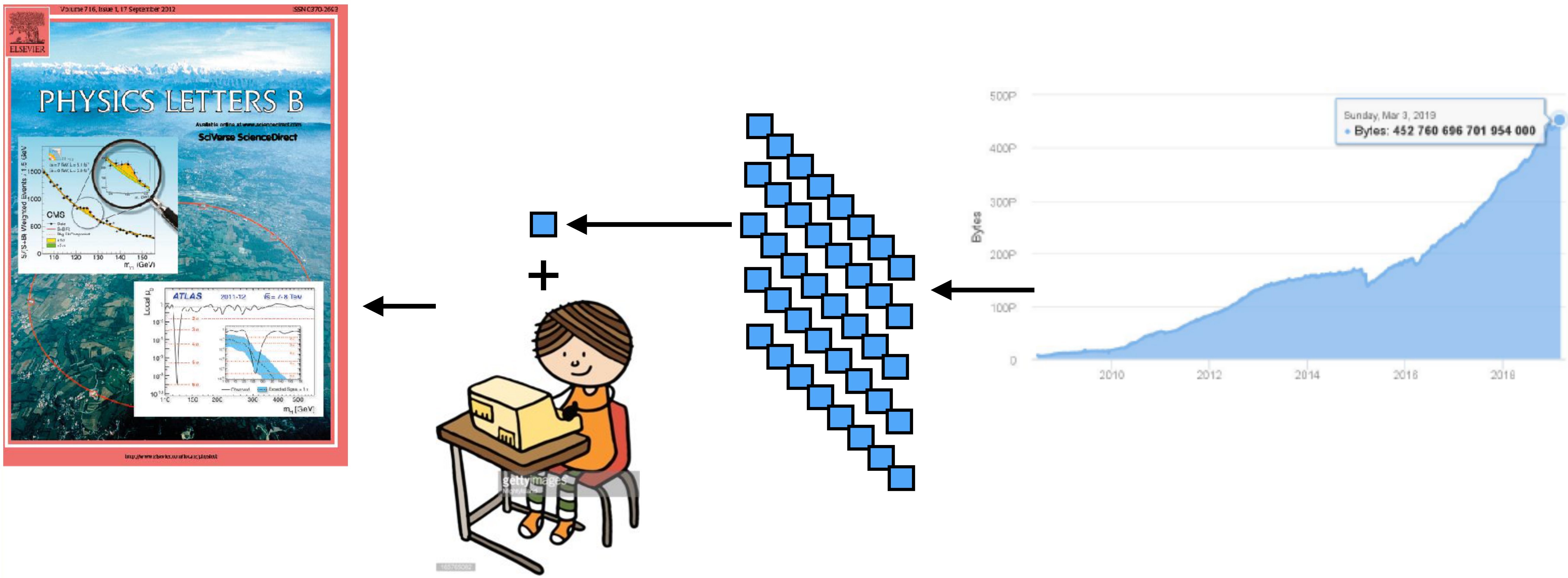
Exabytes of Data

Exabytes of Simulation



Publish!

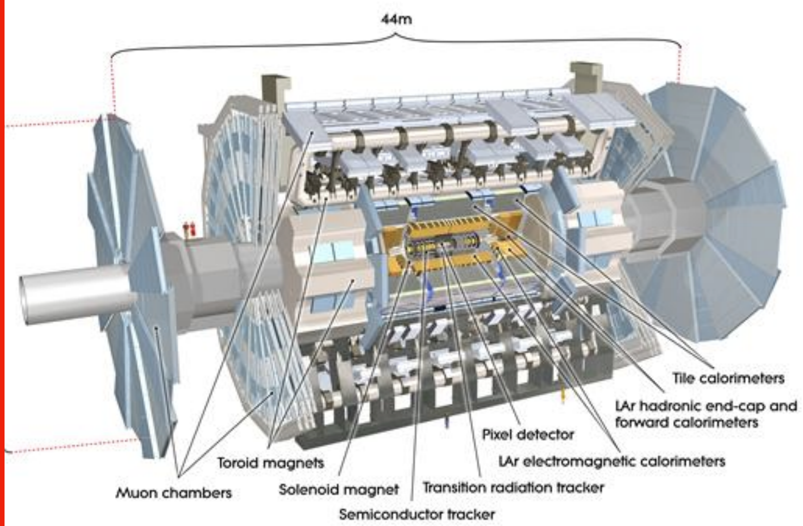
Data analysis



Analysis is performed on a fraction of the data (more on that later!)

Brookhaven National Laboratory

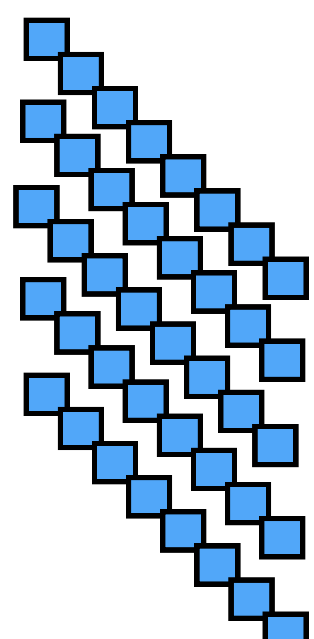
Data’s journey

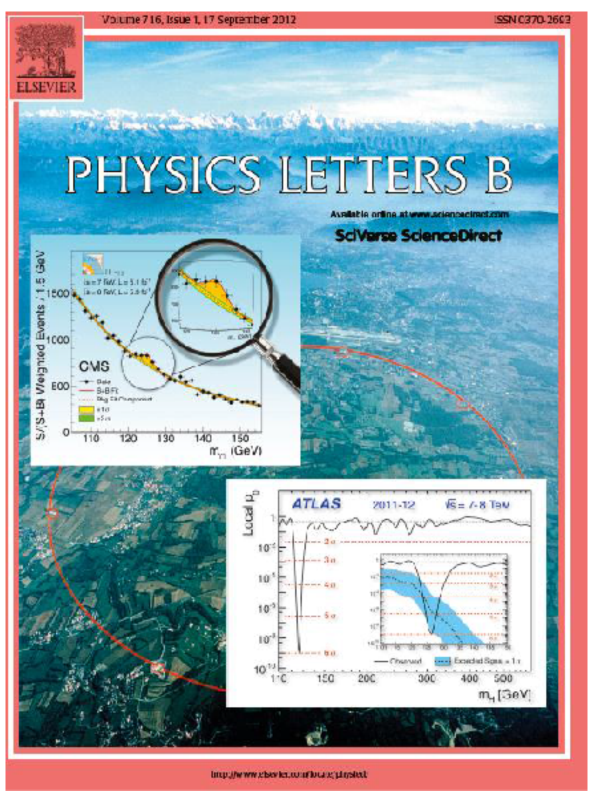


Trigger/DAQ → Data Preparation

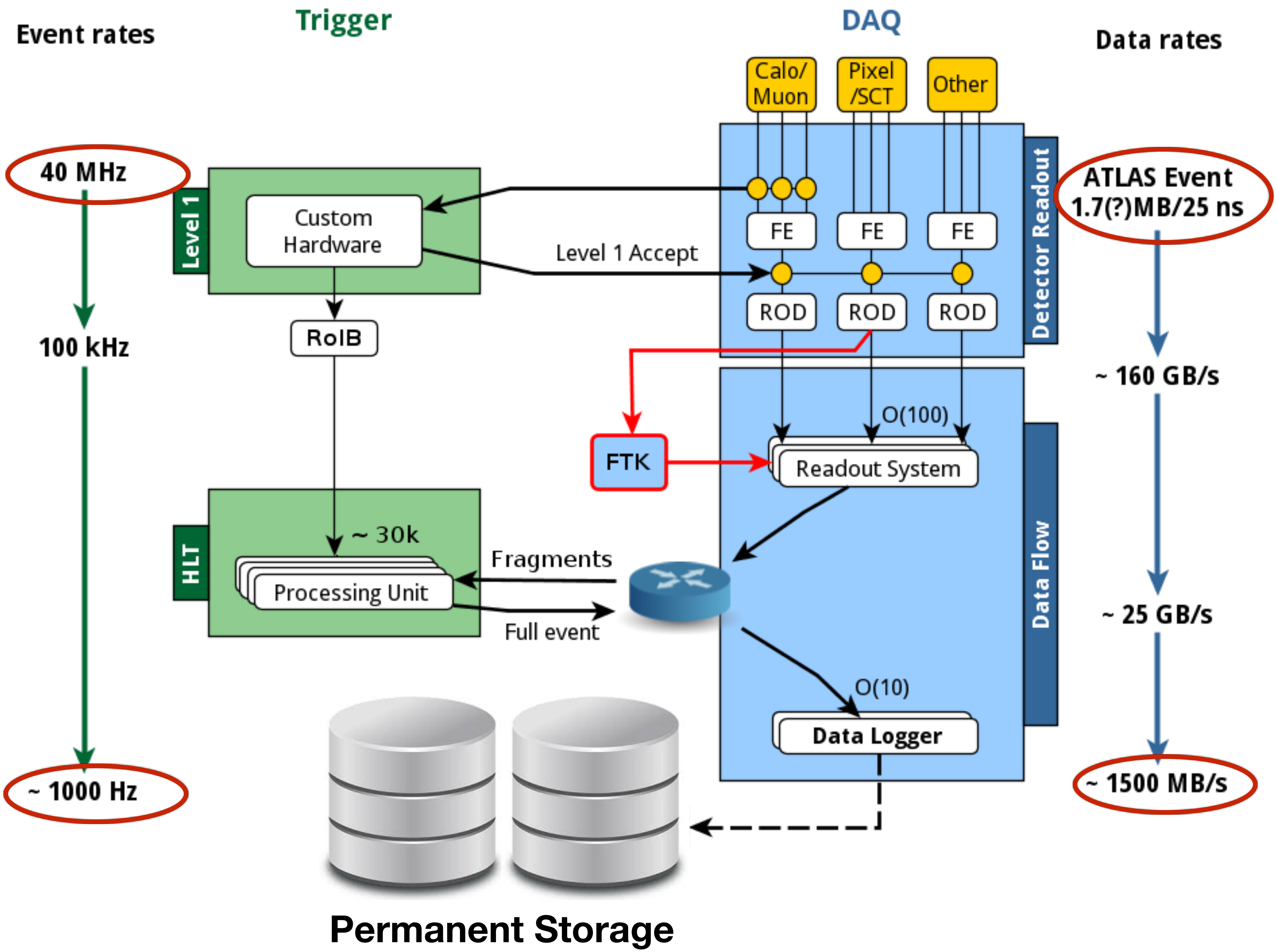
↓

Distributed computing

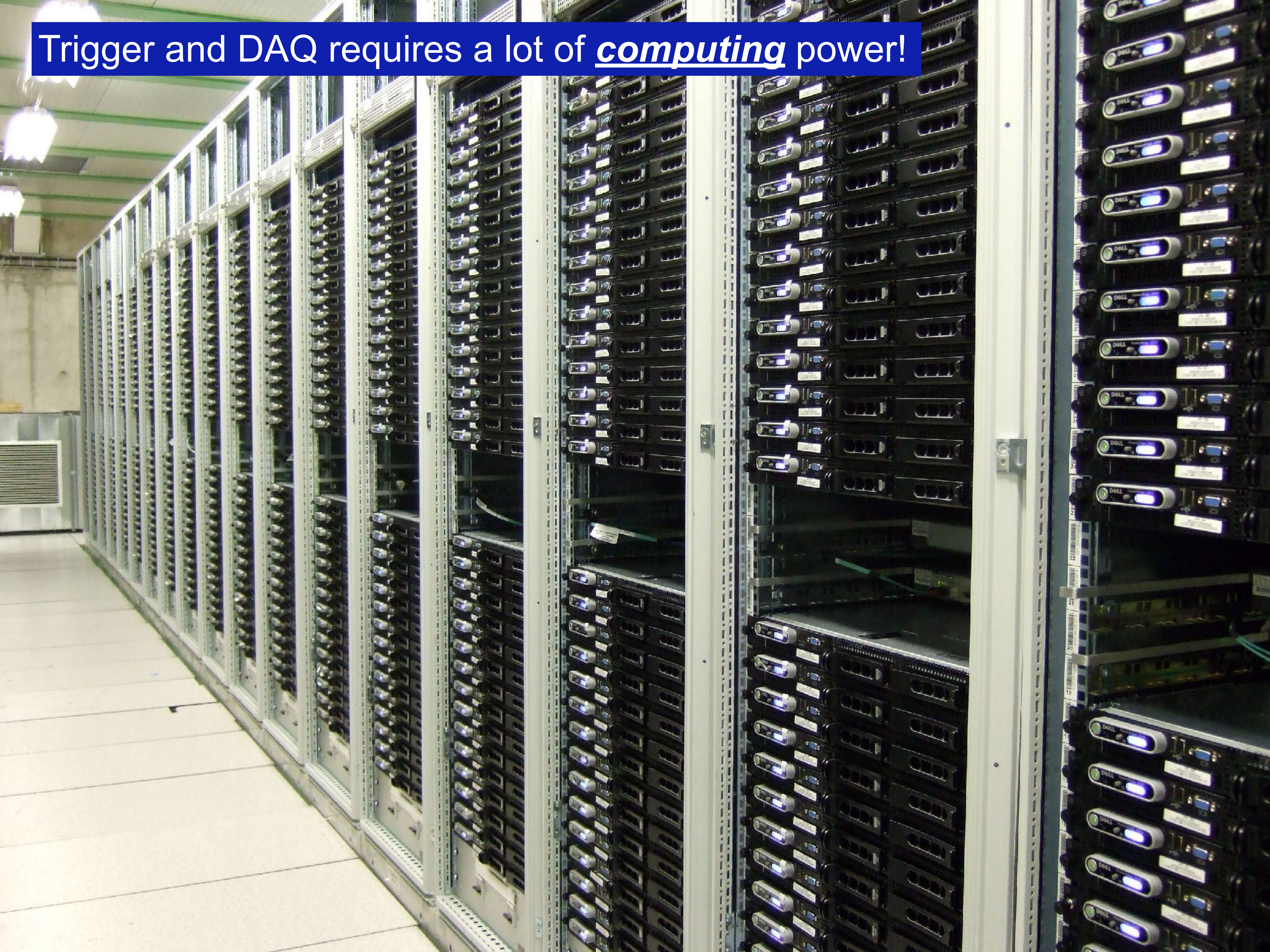




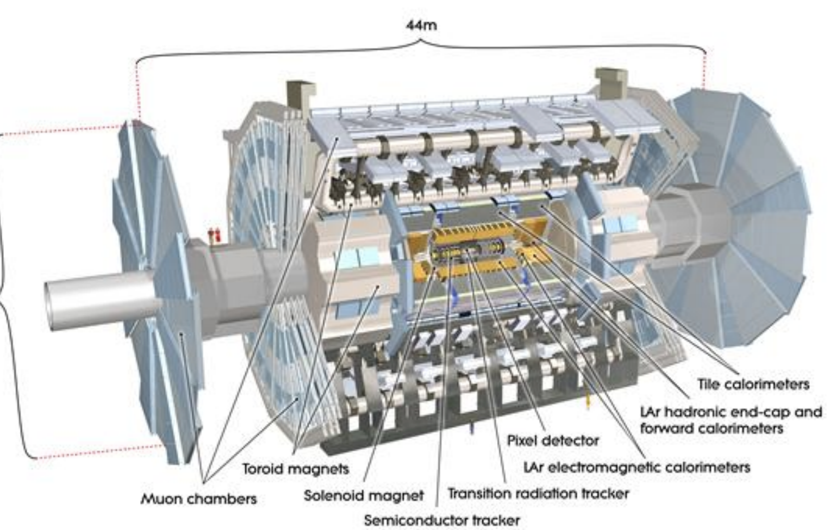
The Atlas Trigger and DAQ

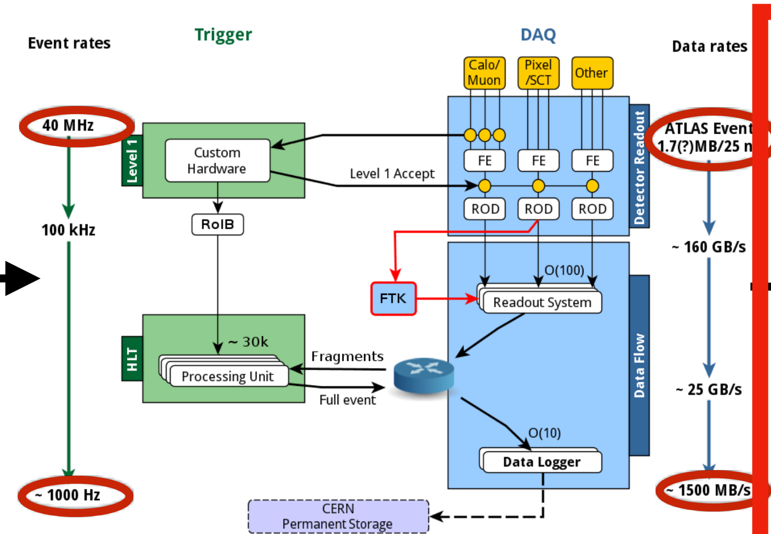


Trigger and DAQ requires a lot of ***computing*** power!



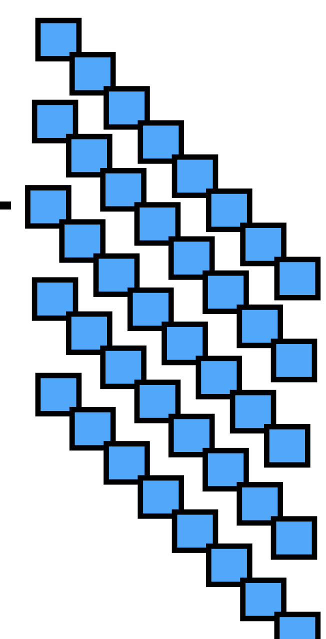
Data’s journey

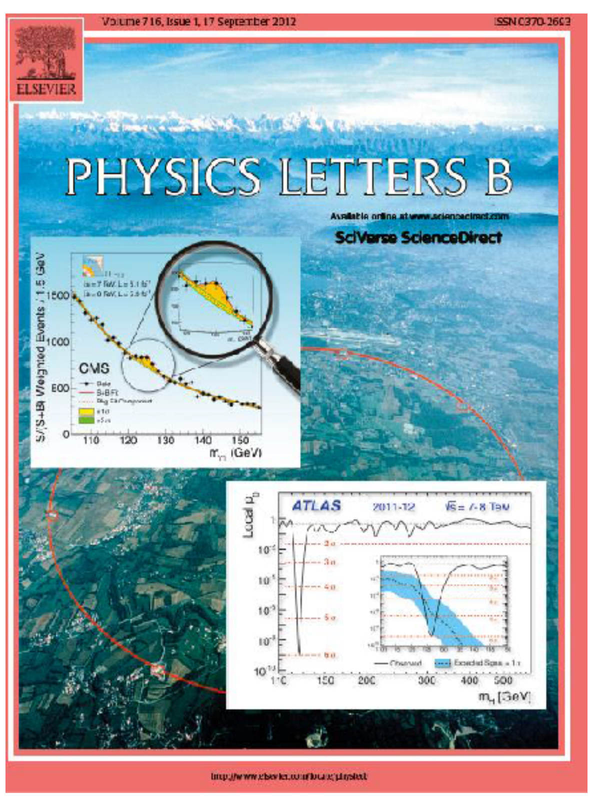




Data Preparation

Distributed computing





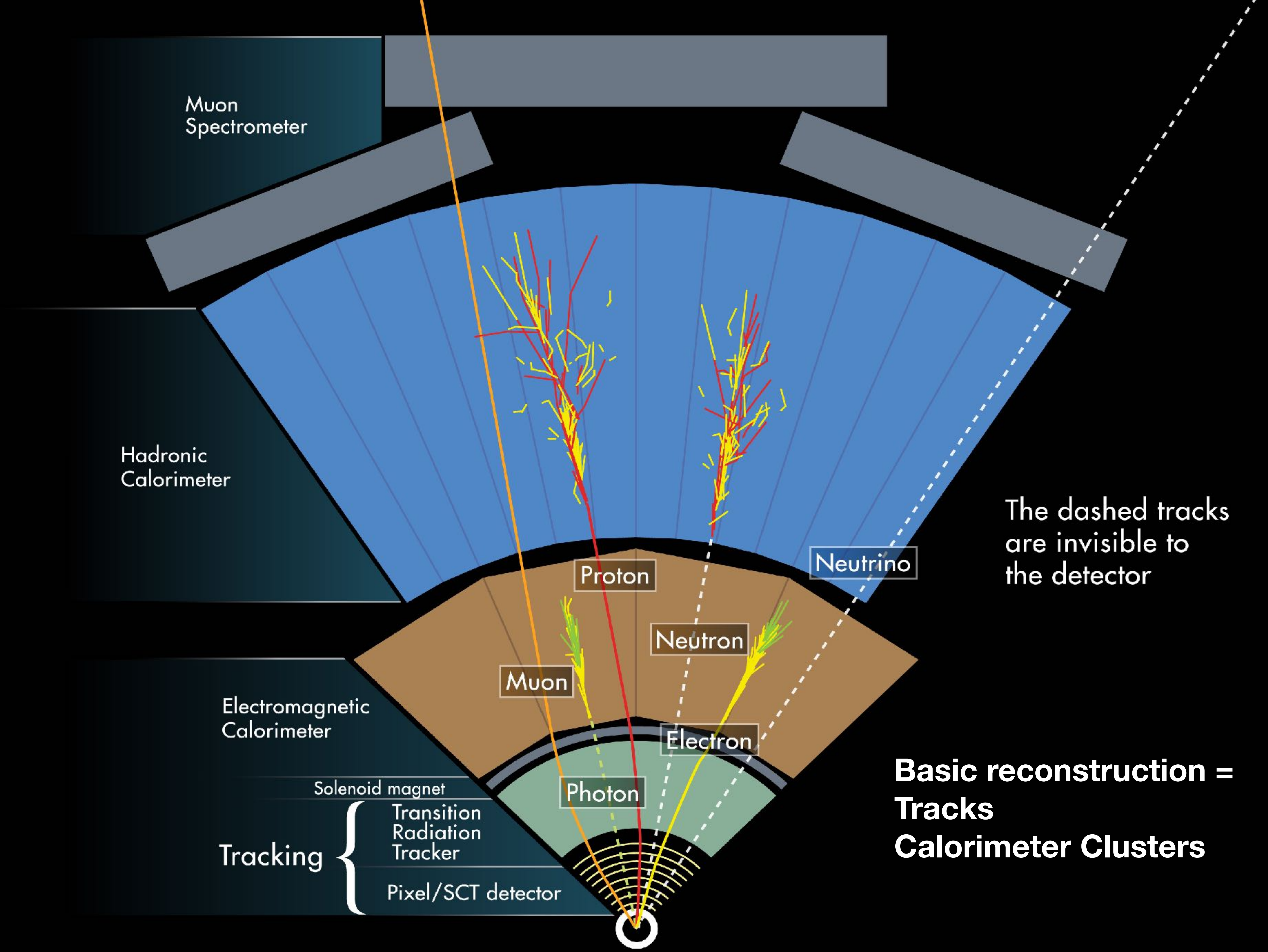
Data Preparation

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***

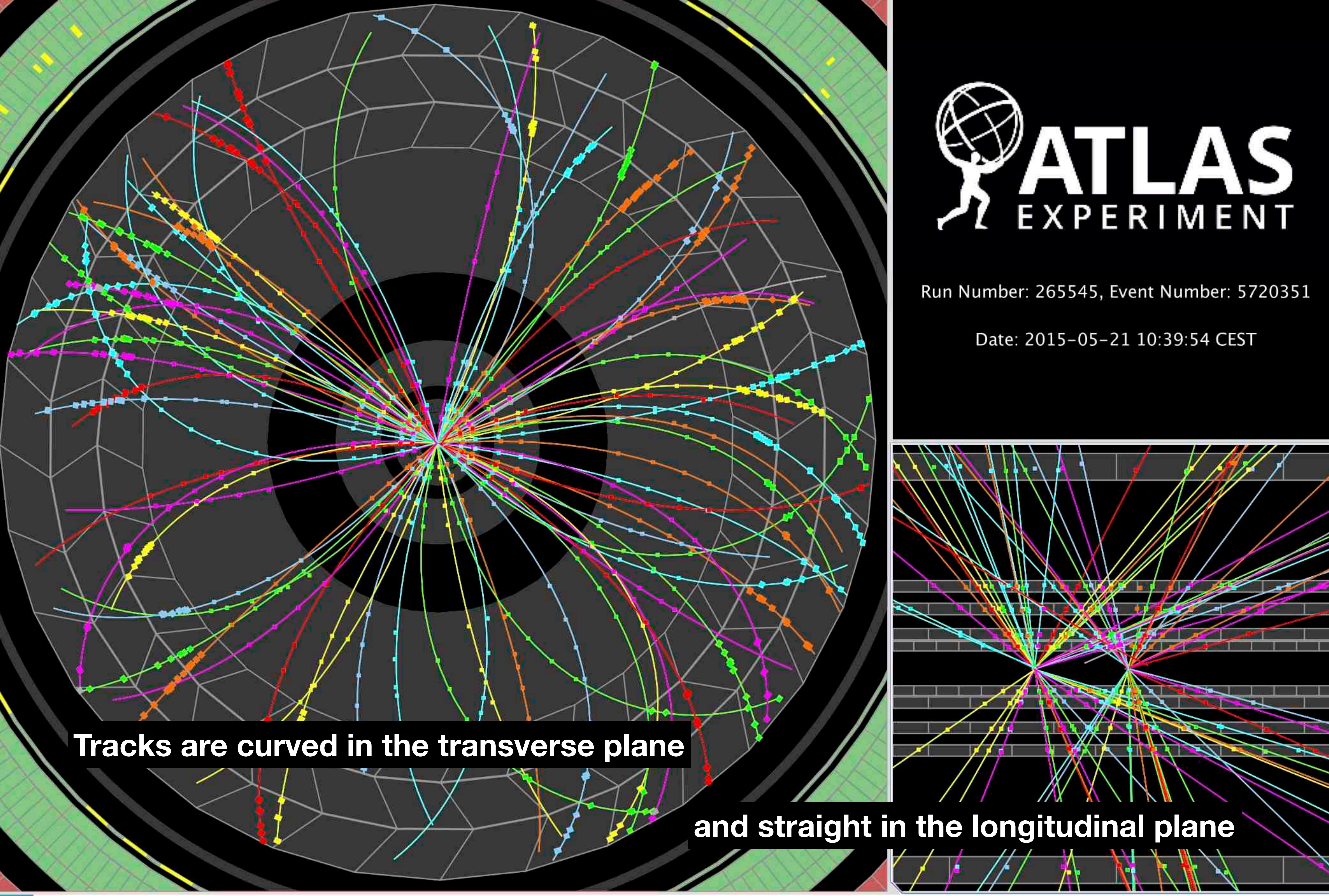
1. **Reconstruct physics signals** from the data
   - Produce information like how many muons does the event have?

Brookhaven National Laboratory



The dashed tracks are invisible to the detector

Basic reconstruction =
Tracks
Calorimeter Clusters

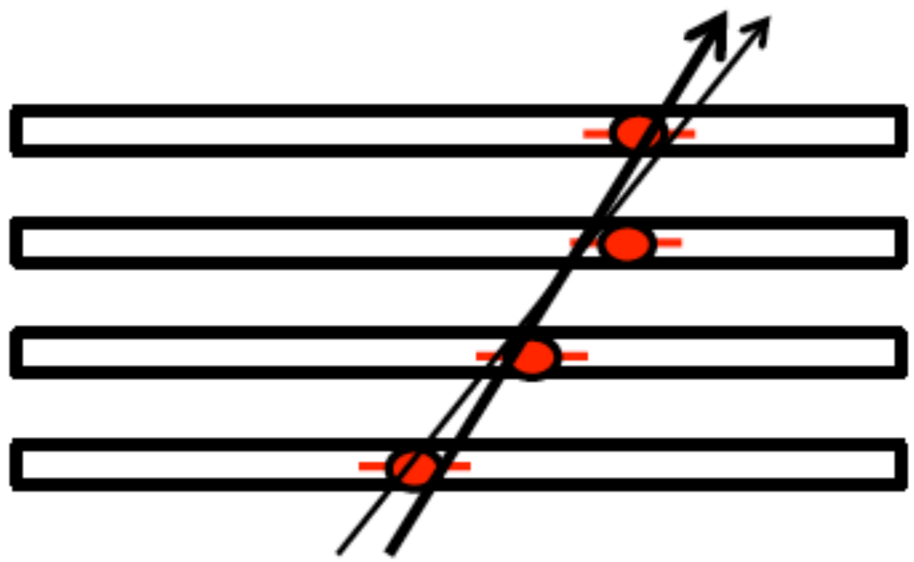

ATLAS
EXPERIMENT
Run Number: 265545, Event Number: 5720351
Date: 2015-05-21 10:39:54 CEST
Tracks are curved in the transverse plane
and straight in the longitudinal plane

Track fitting

- Perfect measurement – ideal

- Imperfect measurement – reality

- Small errors and more points help to constrain the possibilities



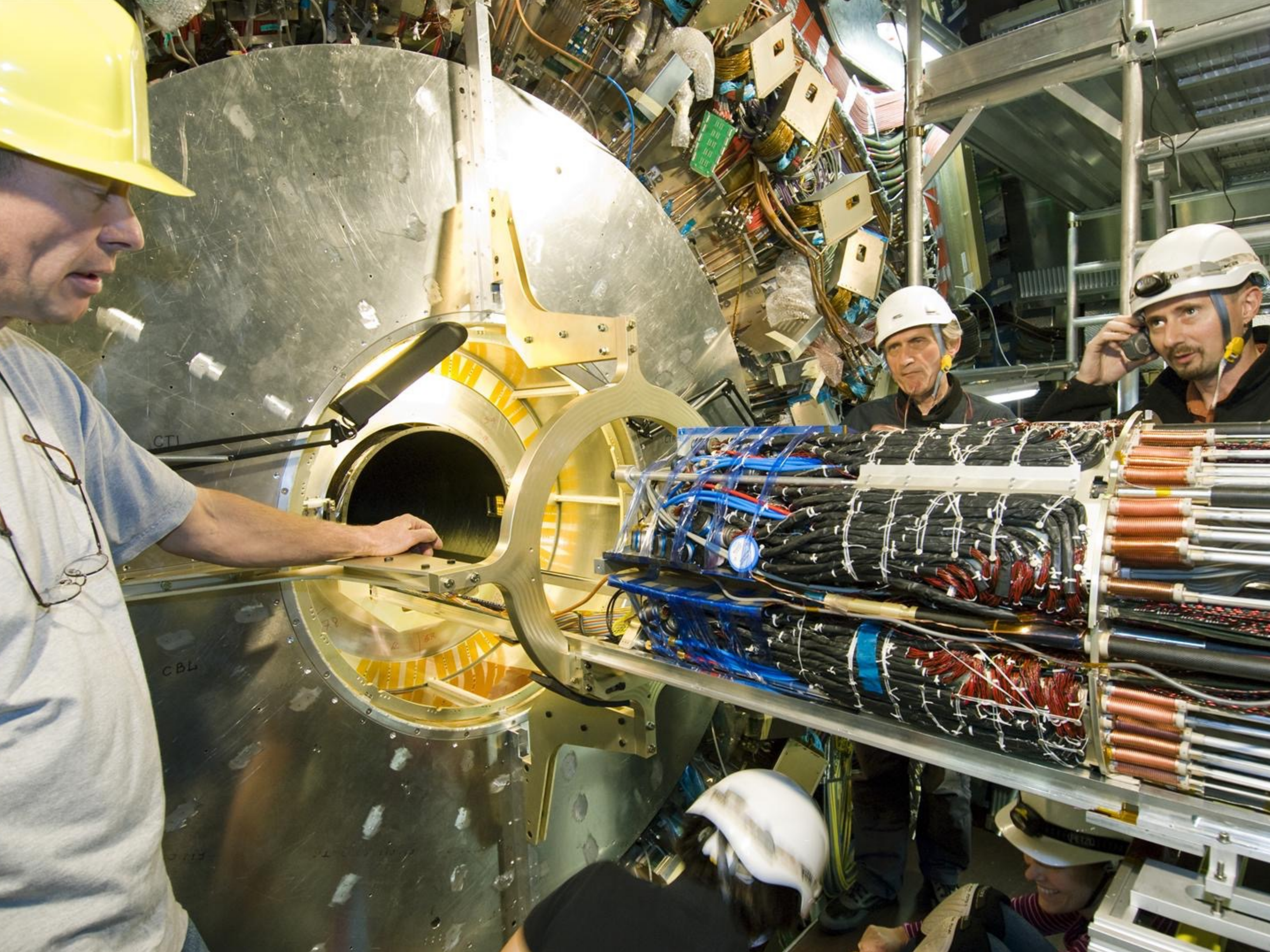
- Quantitatively:
  - Parameterize the track;
  - Find parameters by Least-Squares-Minimization;
  - Obtain also uncertainties on the track parameters.

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***

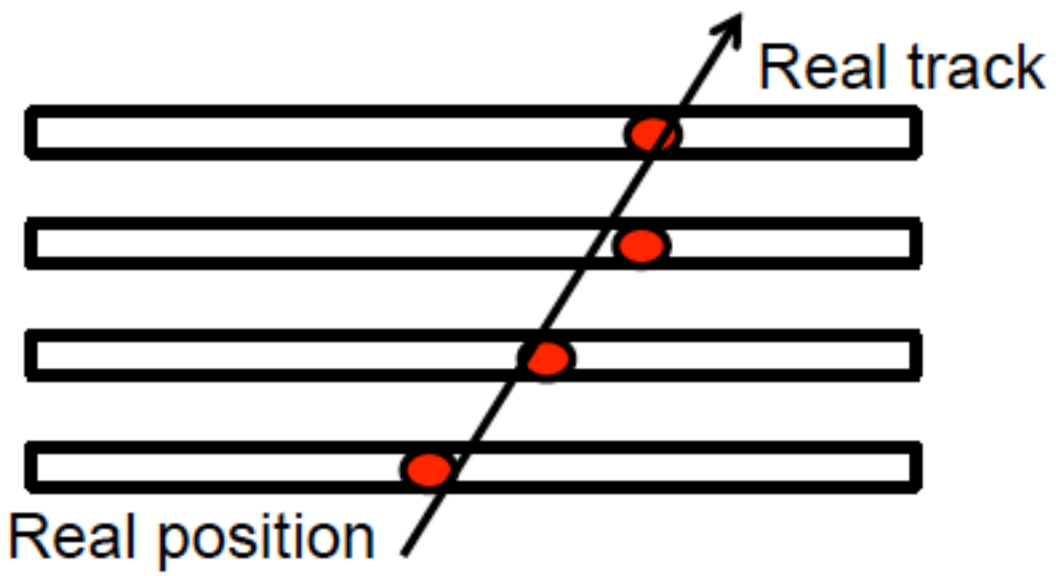
1. **Reconstruct physics signals** from the data ✔
   - Produce information like how many muons does the event have?
2. **Calibrate** the detectors
   - Correct imperfections, account for changes over time…


CTI
CBL

Real detector effects

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

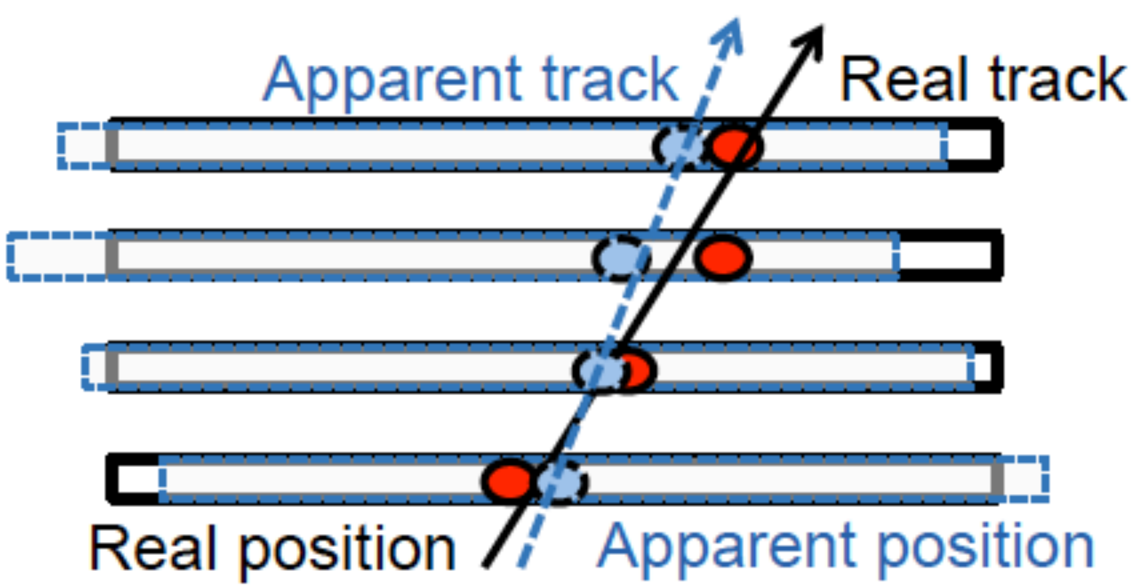
- **Misalignment**
  - Detector elements not positioned in space with perfect accuracy.
  - Alignment corrections derived from data and applied in track reconstruction.



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

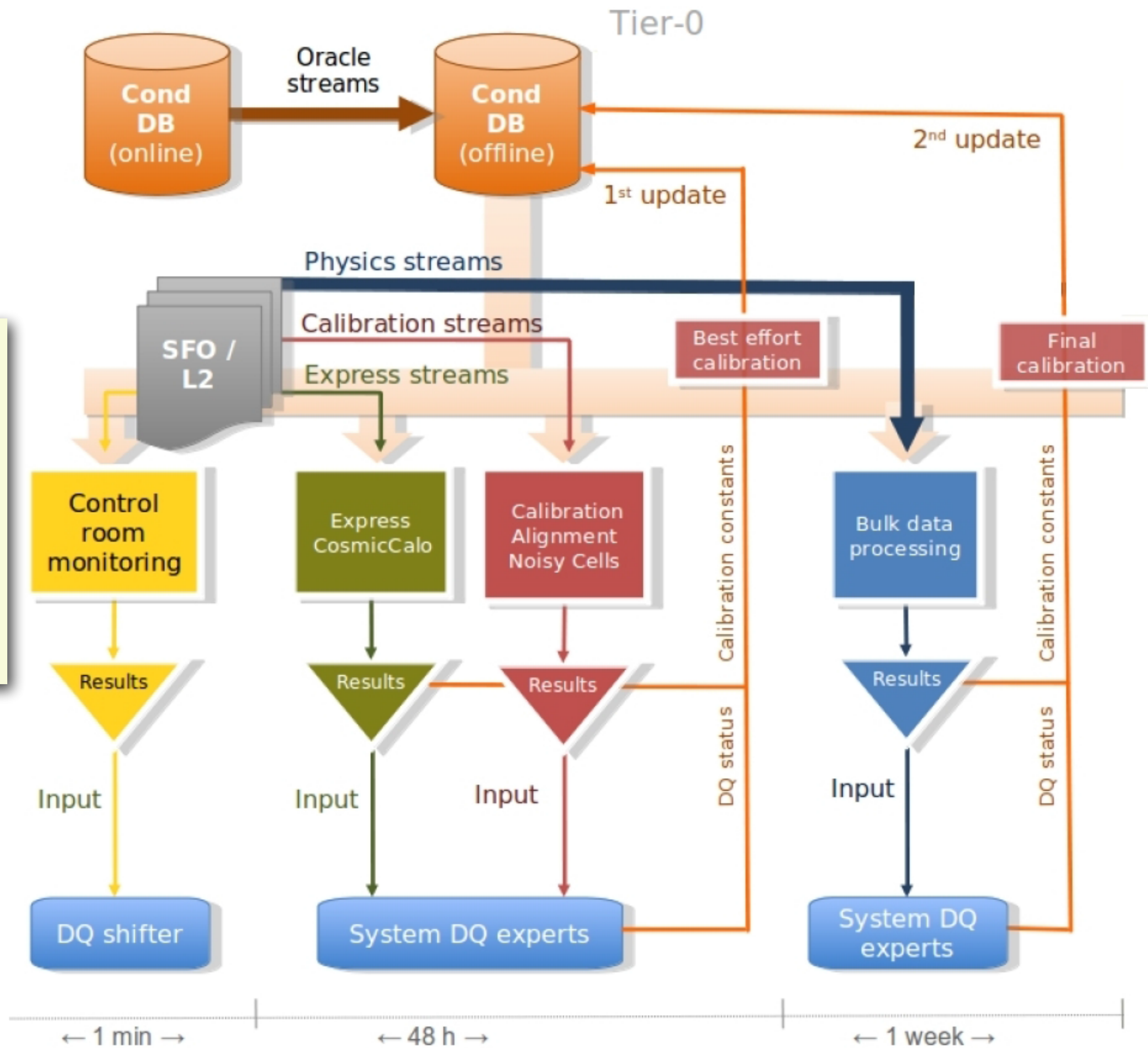
Brookhaven National Laboratory

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 !!



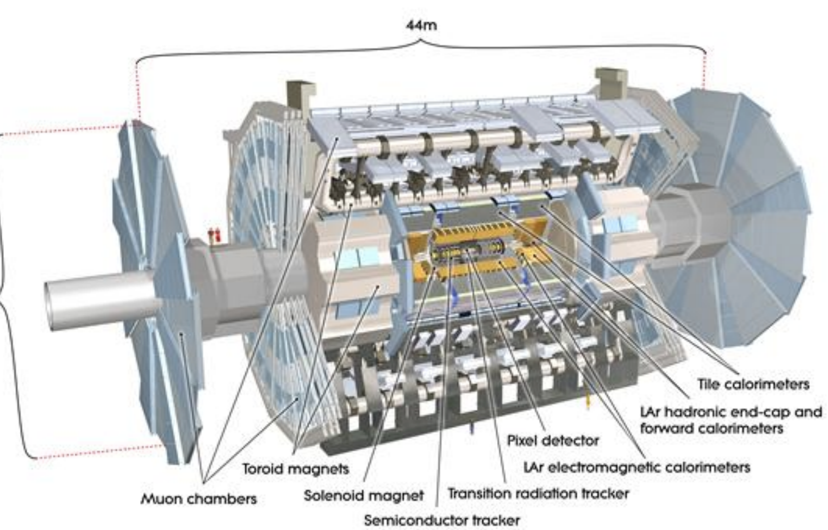
Data Preparation

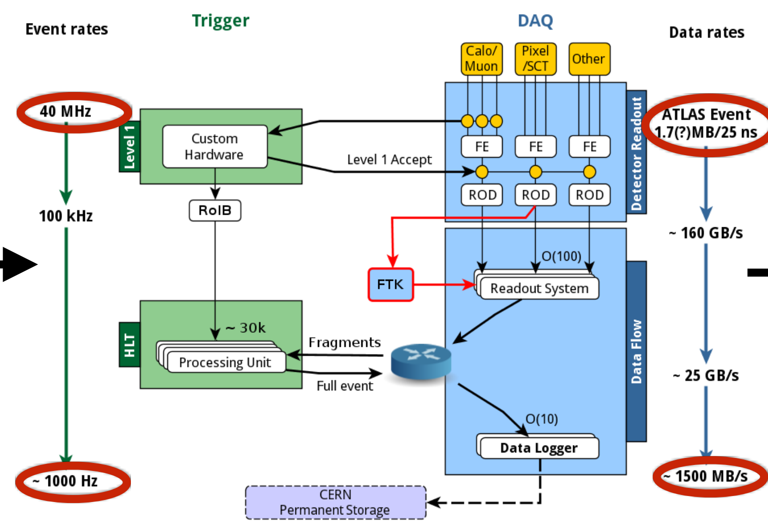
Three major steps to **prepare data for physics analysis** and achieve

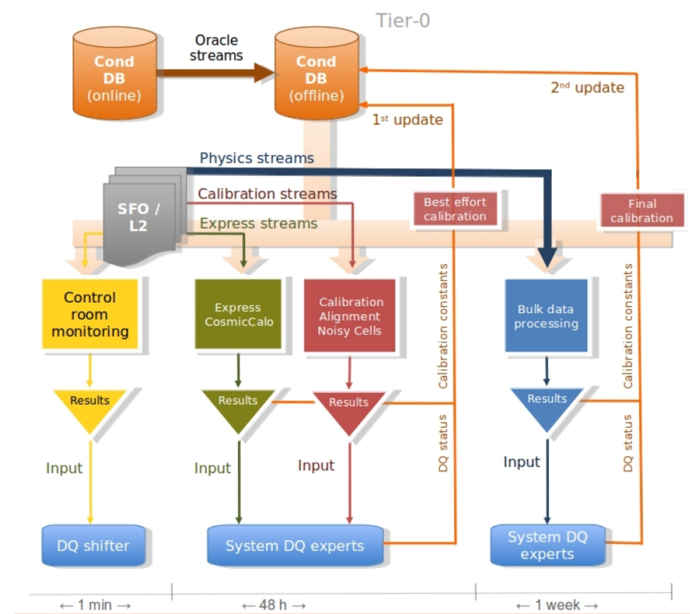
- reliable, high quality data (yes, we ***reject*** low quality data)
- the ***best performance*** from our detectors
- readiness for ***physics analysis***

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 ✓

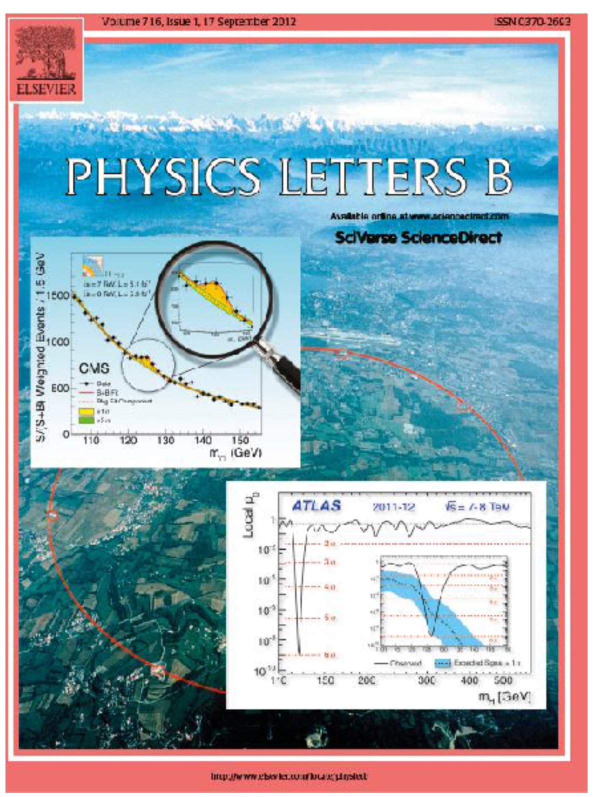
Data's journey







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

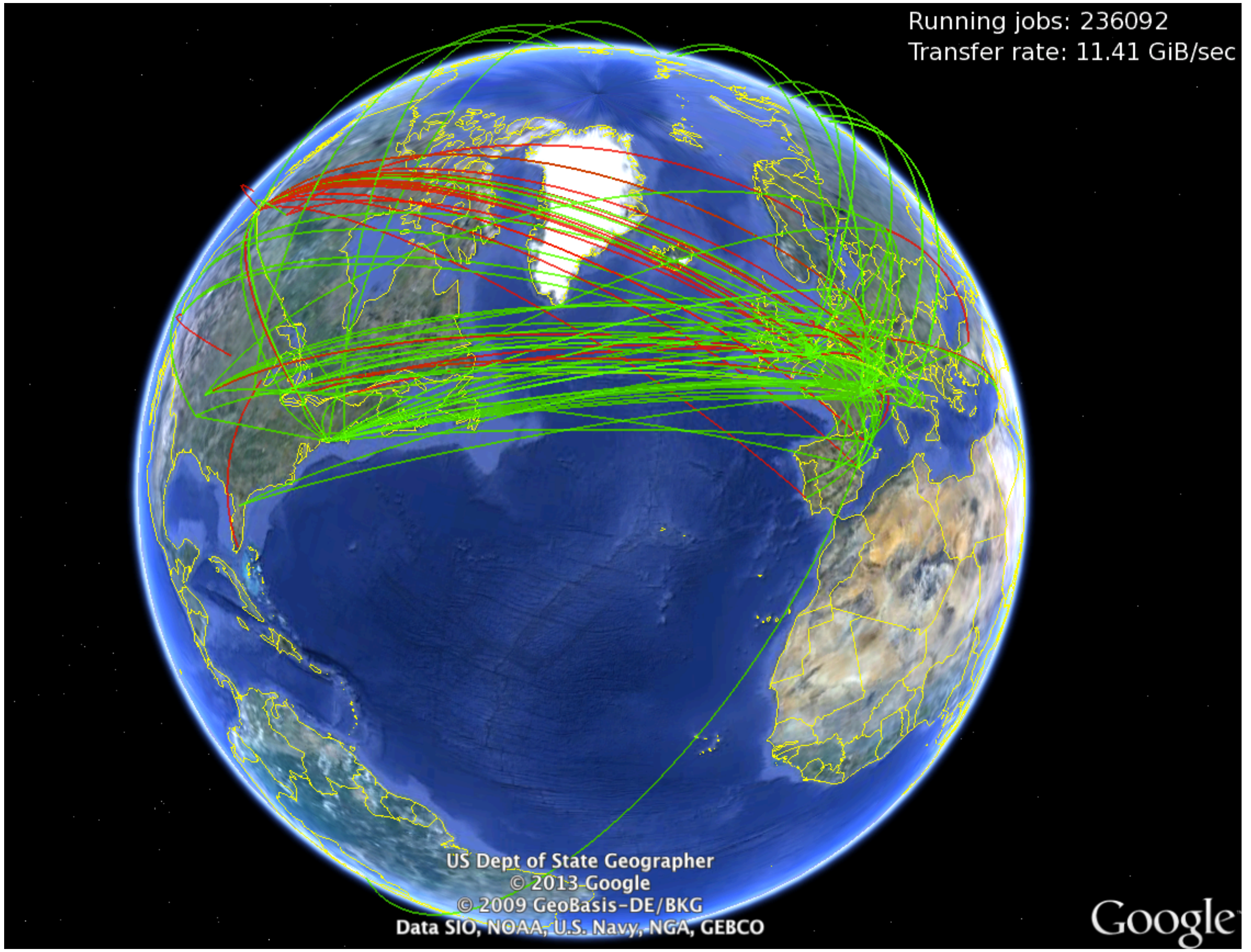
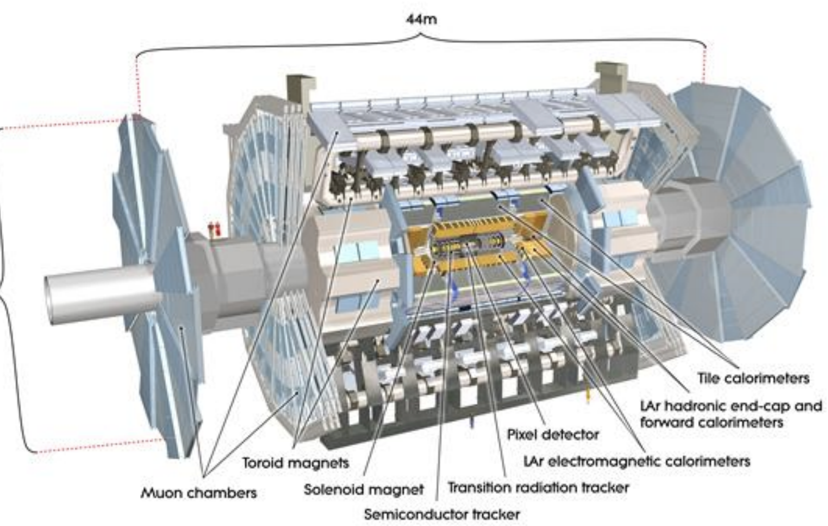
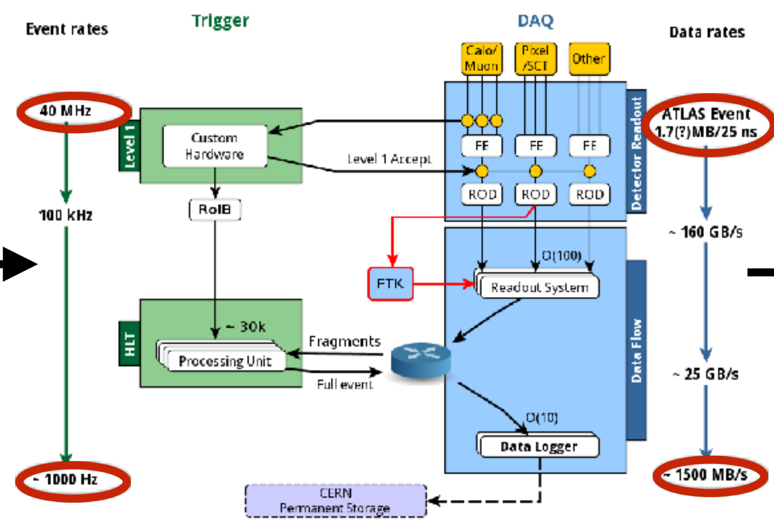


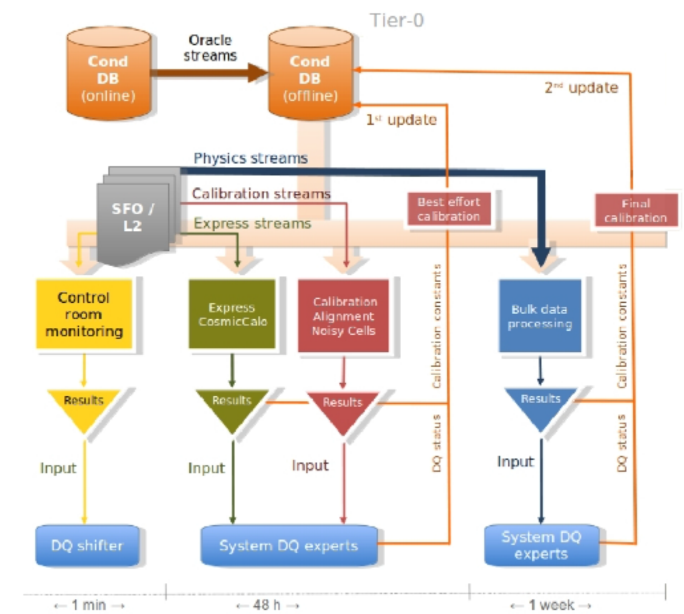
Image taken from the **WLCG** GoogleEarth Dashboard

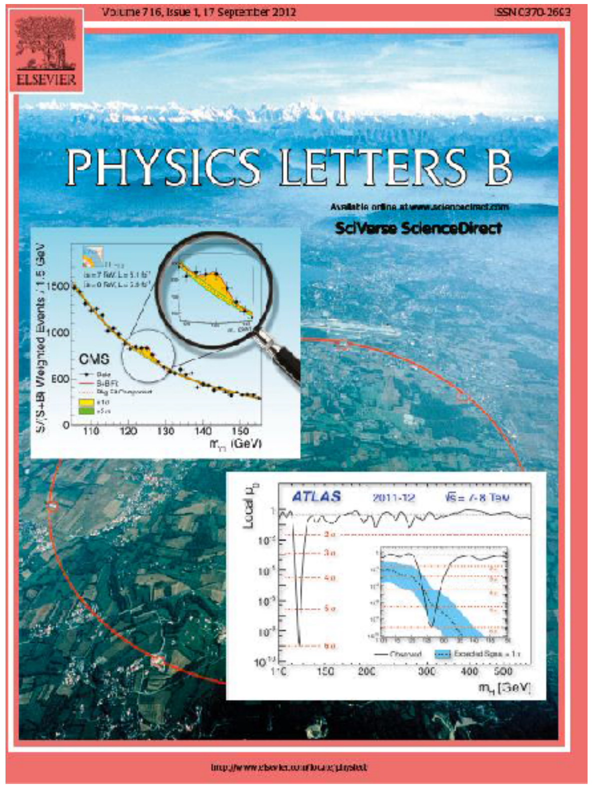
- http://wlcg.web.cern.ch/wlcg-google-earth-dashboard

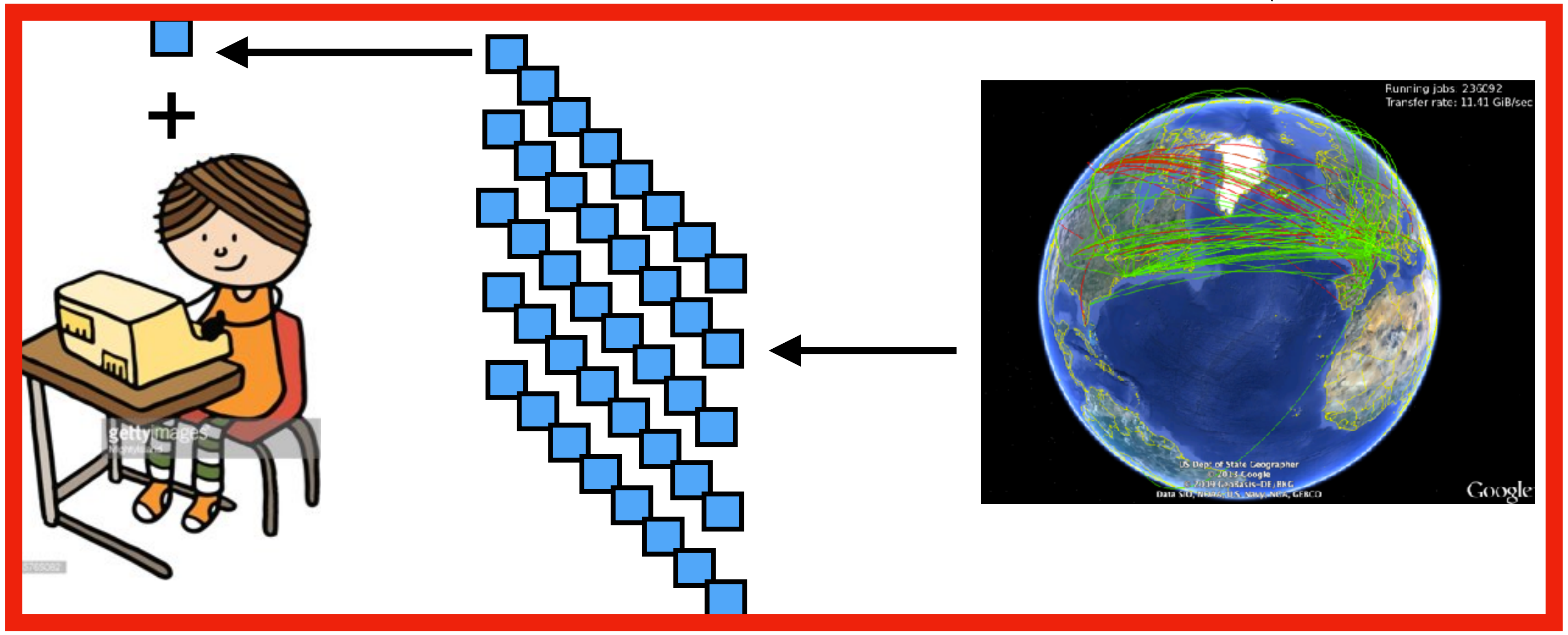
Data’s journey











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***

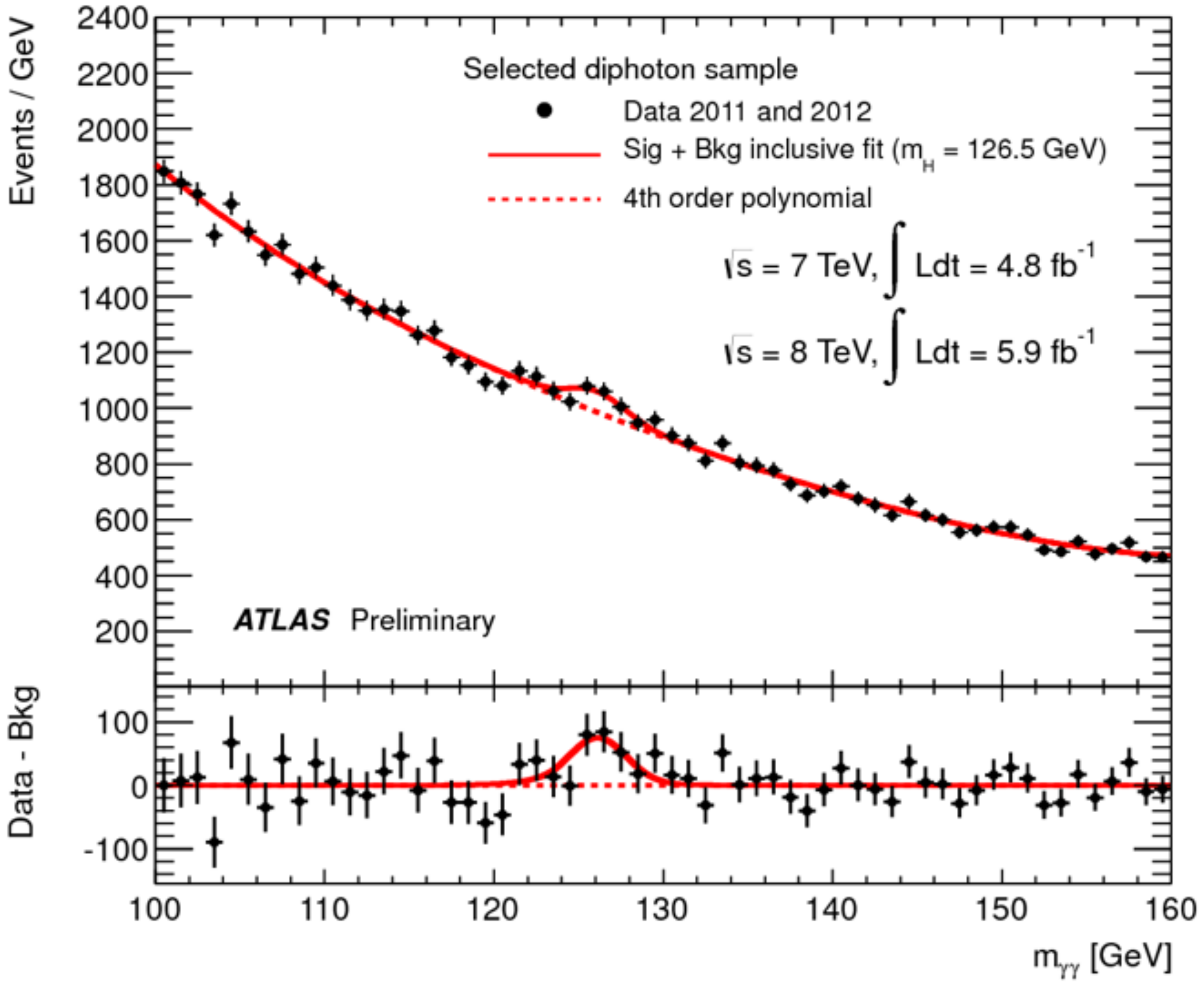
Data reduction proceeds via a two-pronged approach:

Select only the events that you are interested in

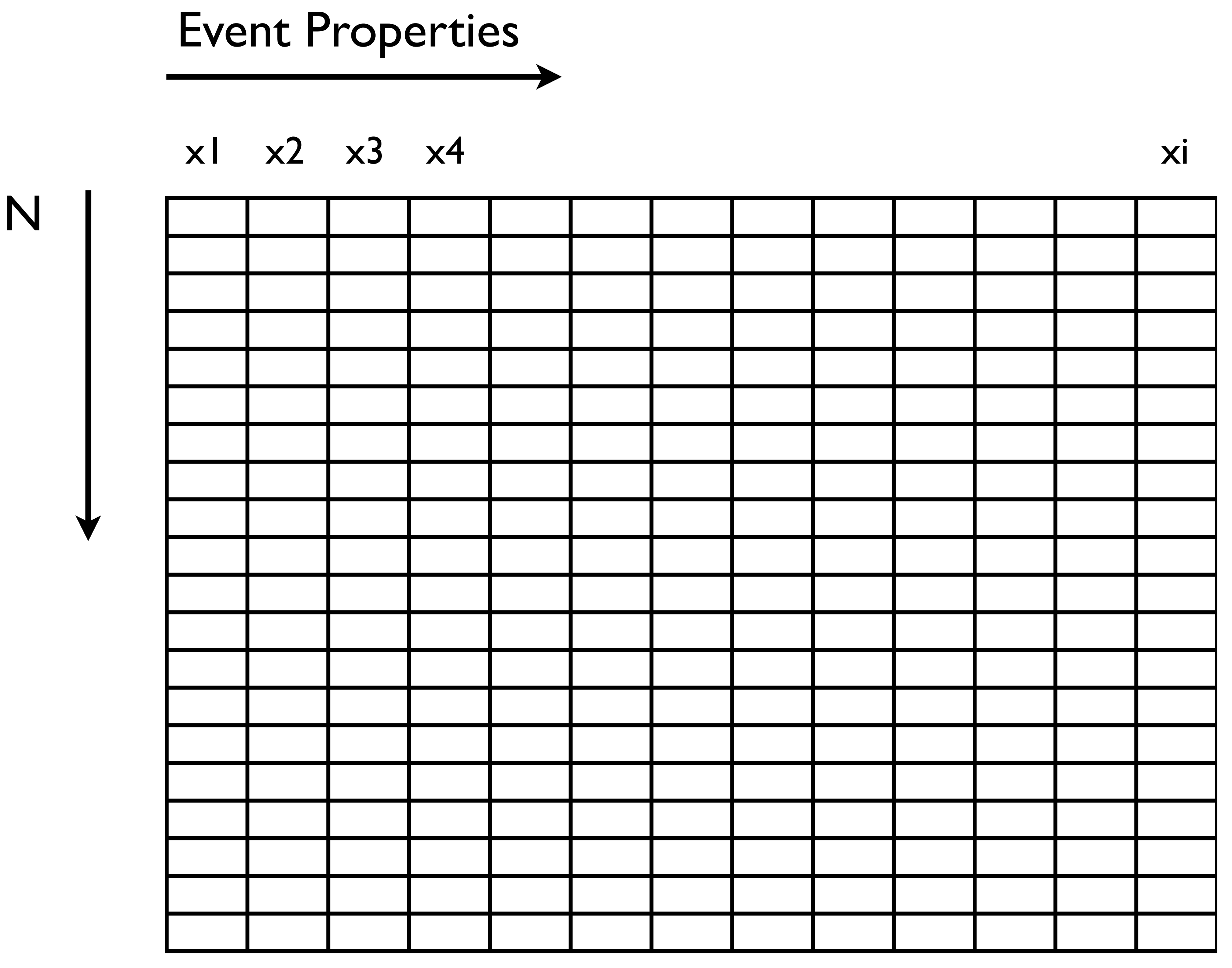
- e.g. events with two photons

Keep only the information you need

Final statistical inference is only performed on the reduced data

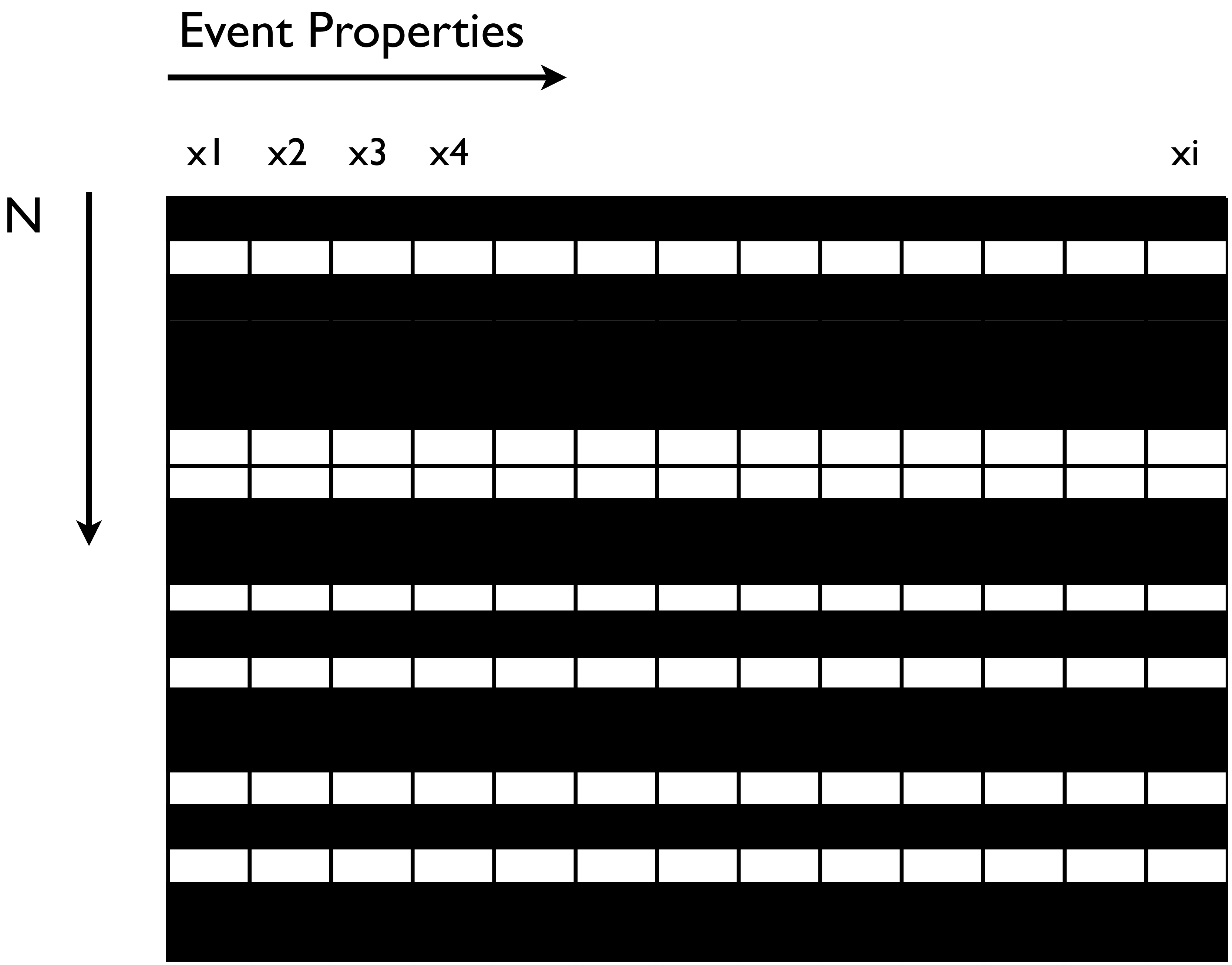


Slice and dice - data reduction



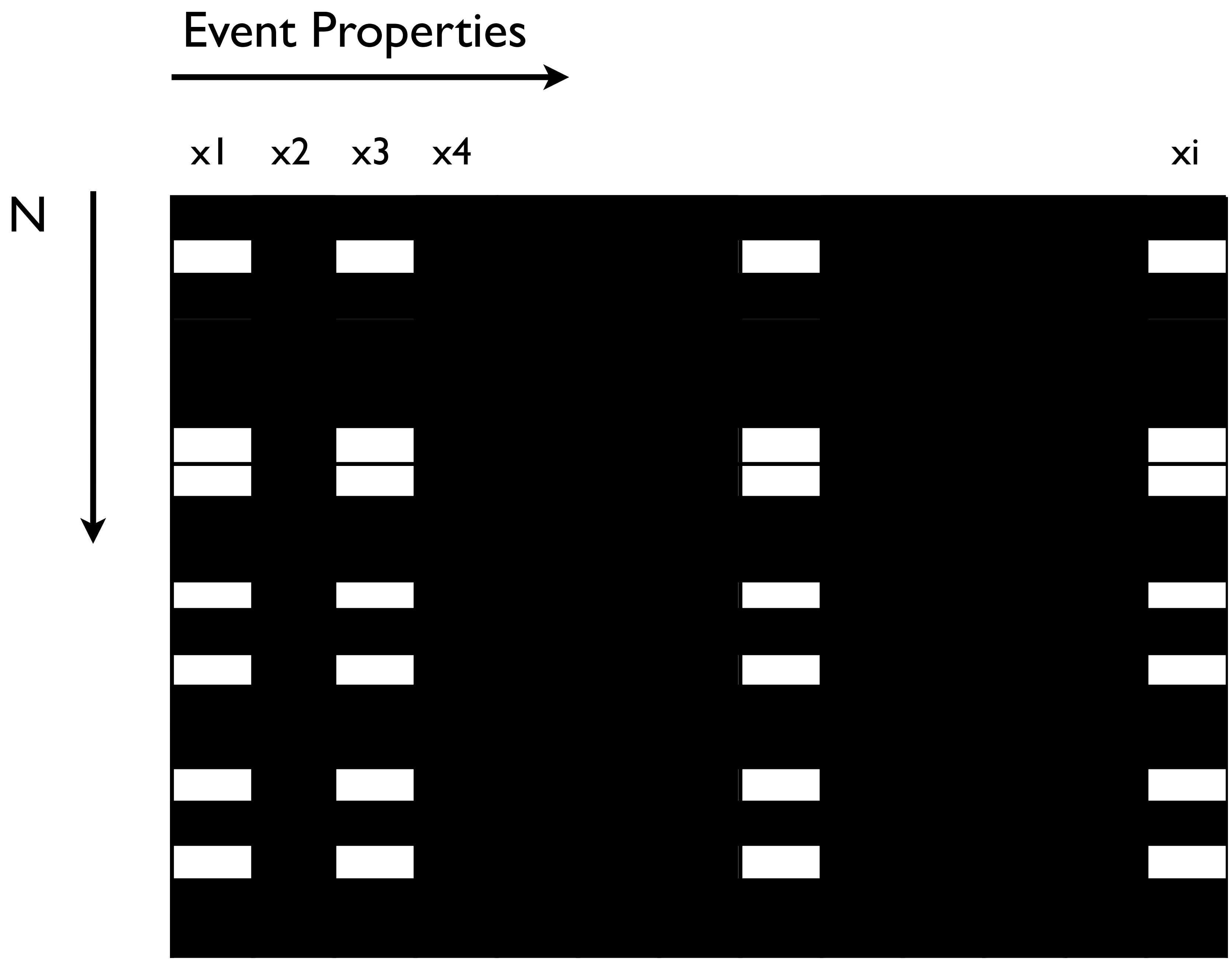
- ***N*** events * ***i*** event properties makes for a LOT of data !

Slice and dice - data reduction



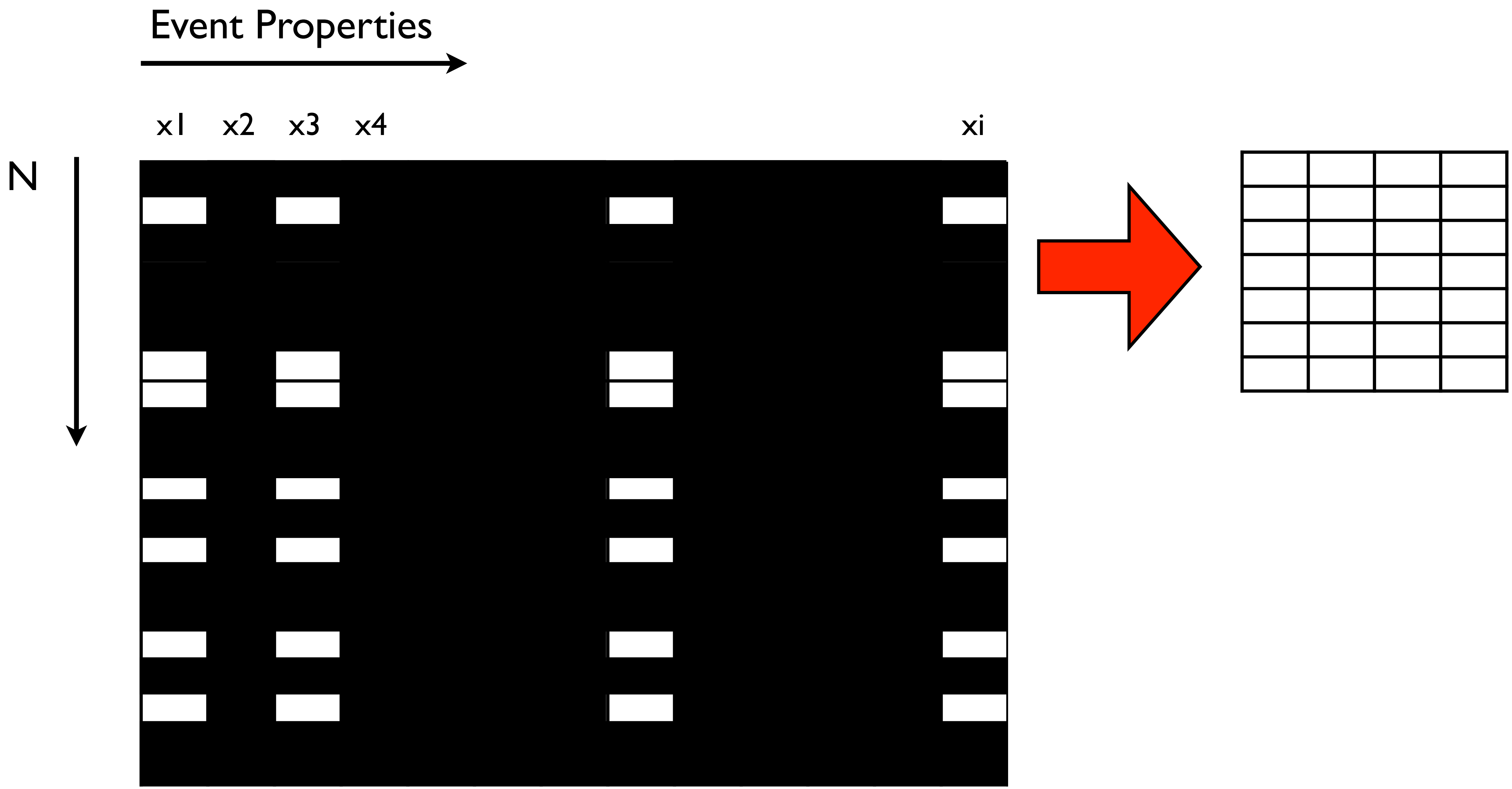
- Select ***only the interesting*** events

Slice and dice - data reduction



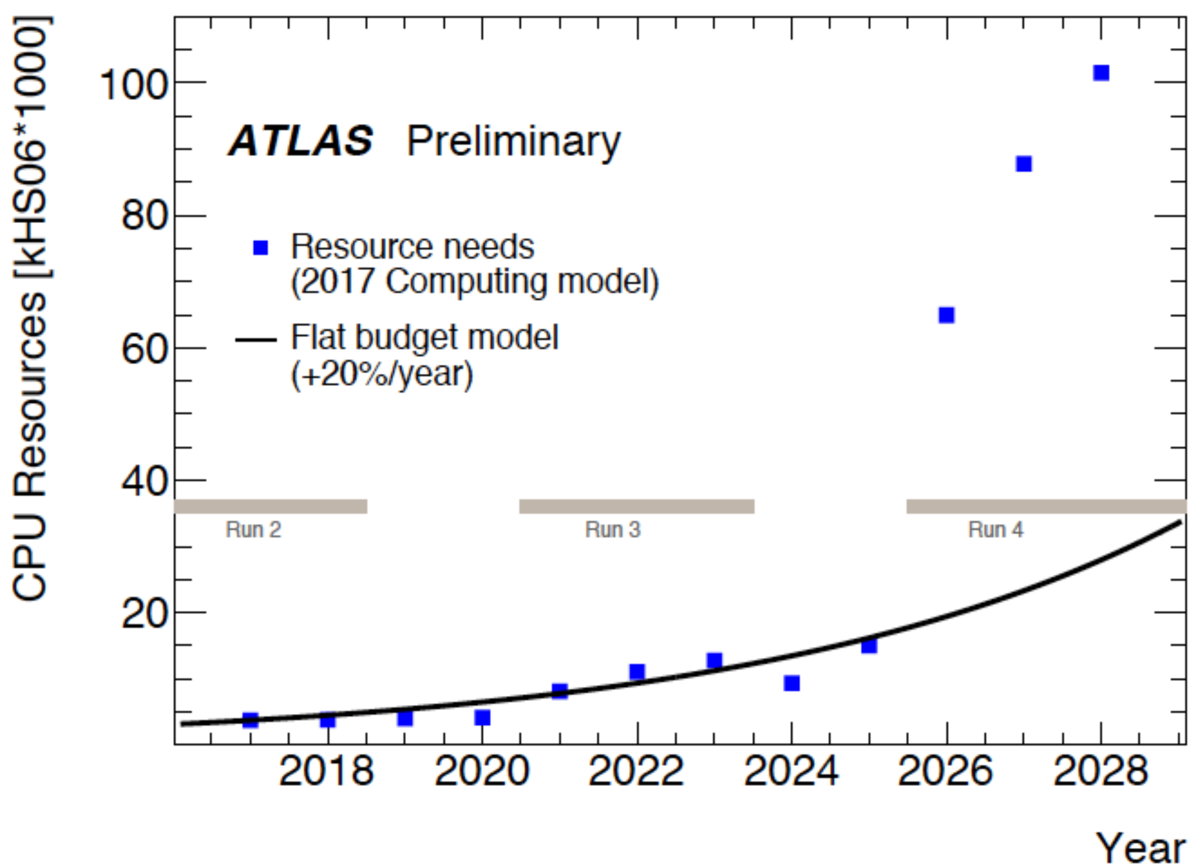
- Select ***only the interesting*** event properties

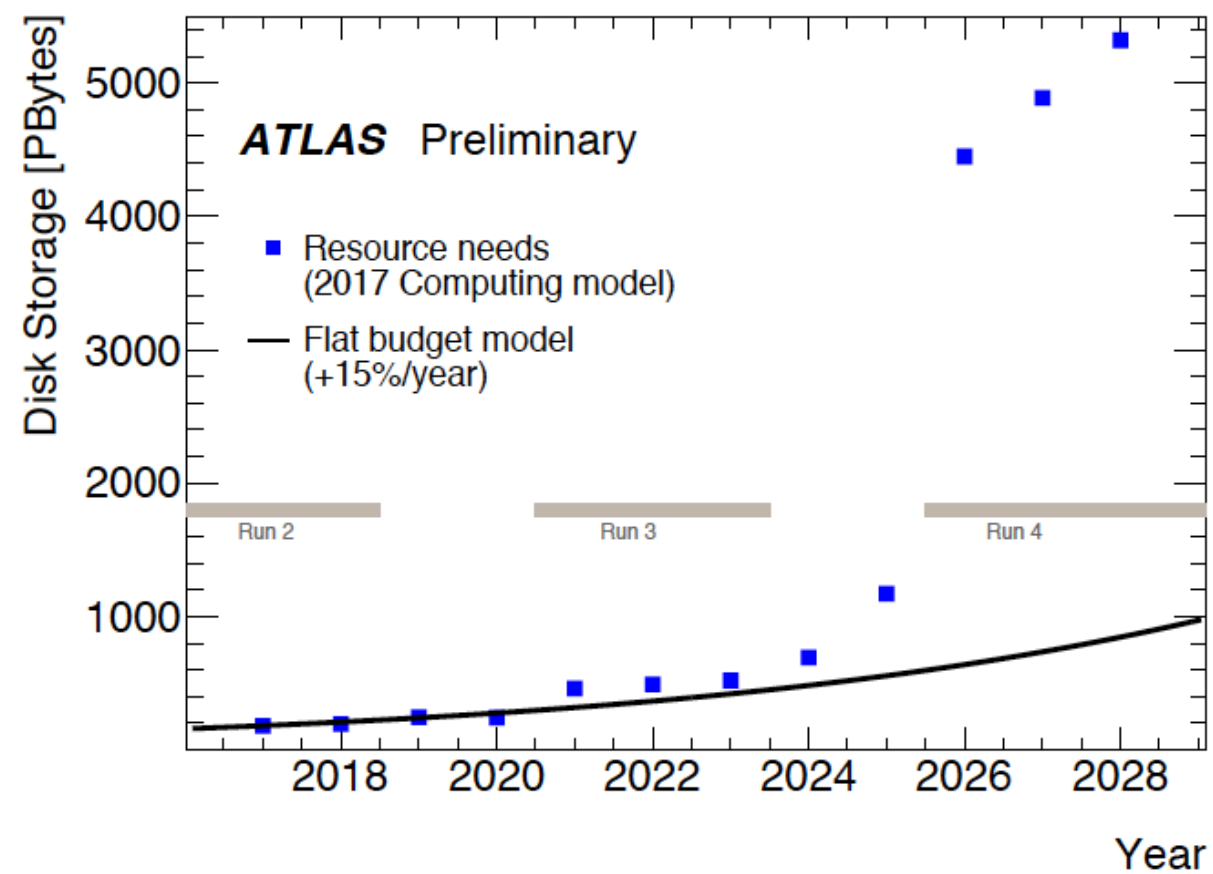
Slice and dice - data reduction



- From now on you only need to process a smaller amount of data

Future software and computing





The data challenge is going to get a lot harder than standard progress can handle, c.f. HEP Software Foundation paper on software and computing R&D for the 2020s:

http://arxiv.org/pdf/1712.06982.pdf

Conclusions

Physics experiments produce a lot of data!

- ***Triggers** reduce the raw event rate*
- ***Data Preparation** is crucial to get the best quality reconstructed and calibrated data*
- ***Data reduction** is crucial to enable analysis, otherwise there's simply too much data to process*

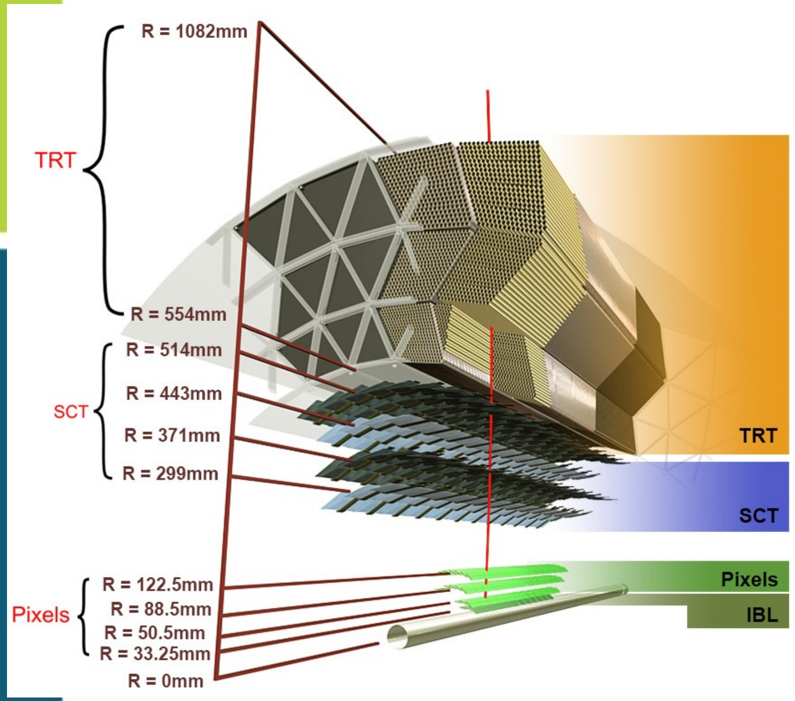
All of this requires *computing, computing, computing!*

- and not only hardware, we need great software too

The future will bring even more computing challenges that will need to be solved

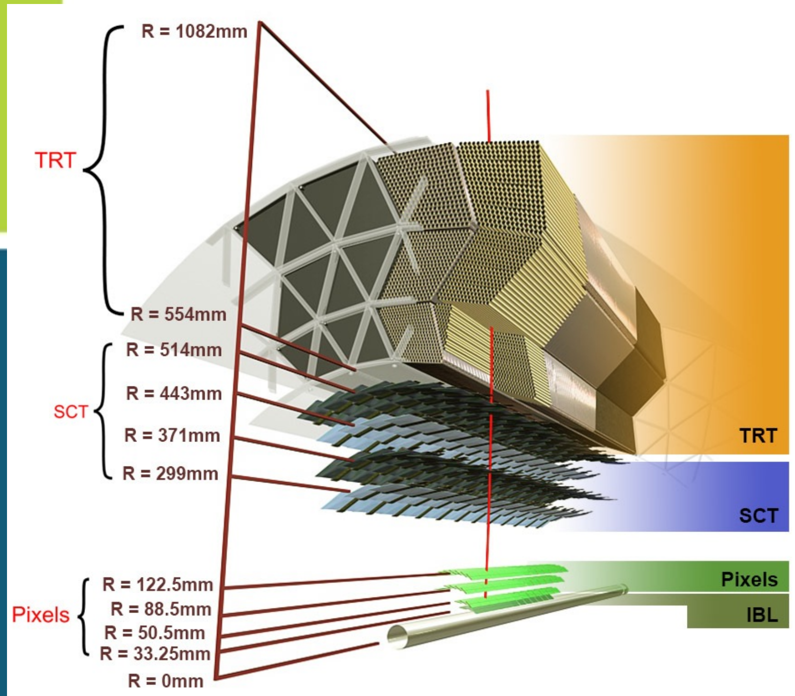
Backup

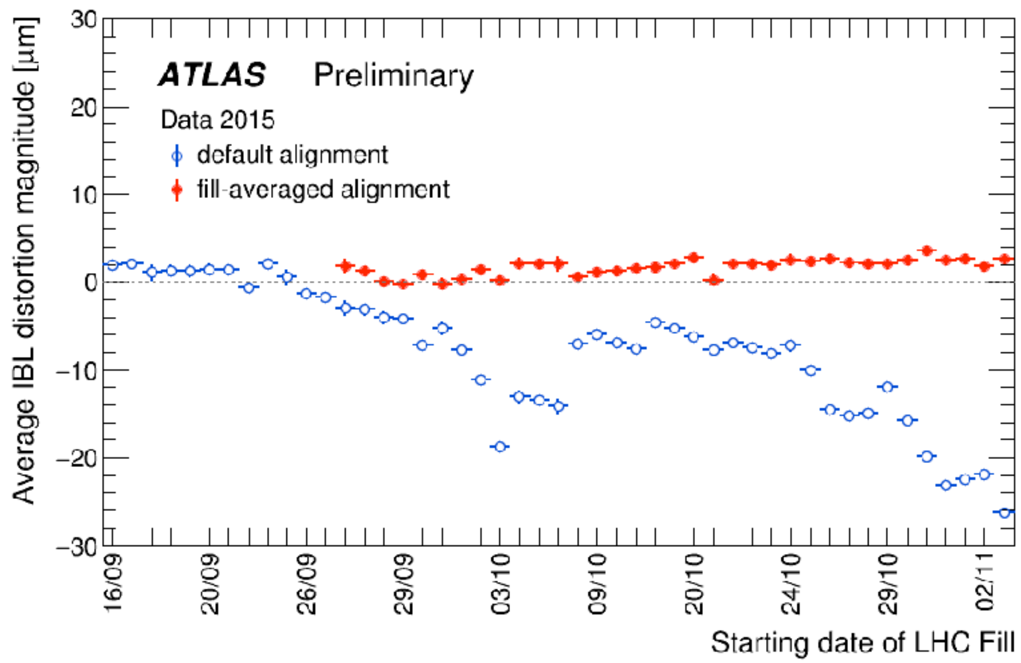
Calibration



During the break between Run 1 and Run 2, ATLAS inserted the IBL, an extra layer of silicon tracker close to the beam pipe

Calibration



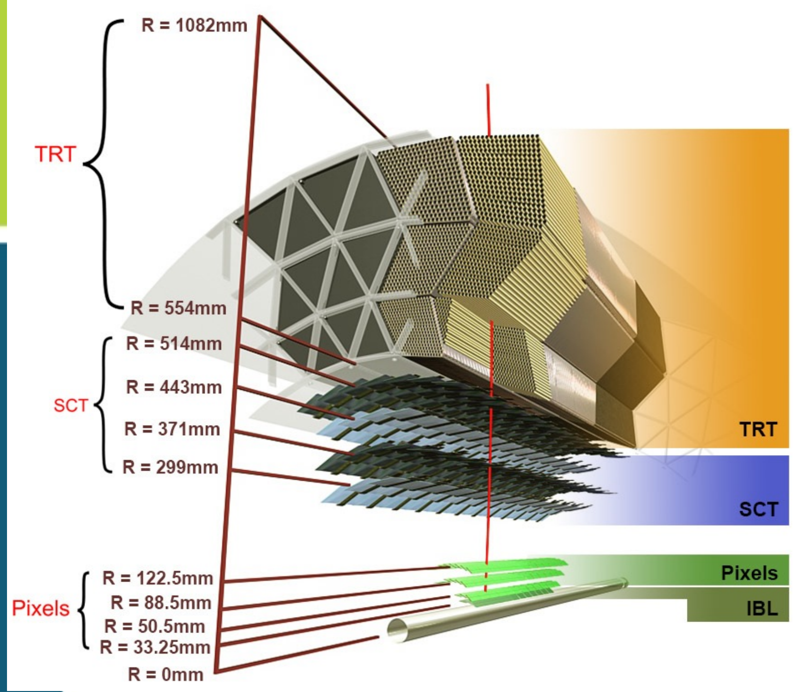


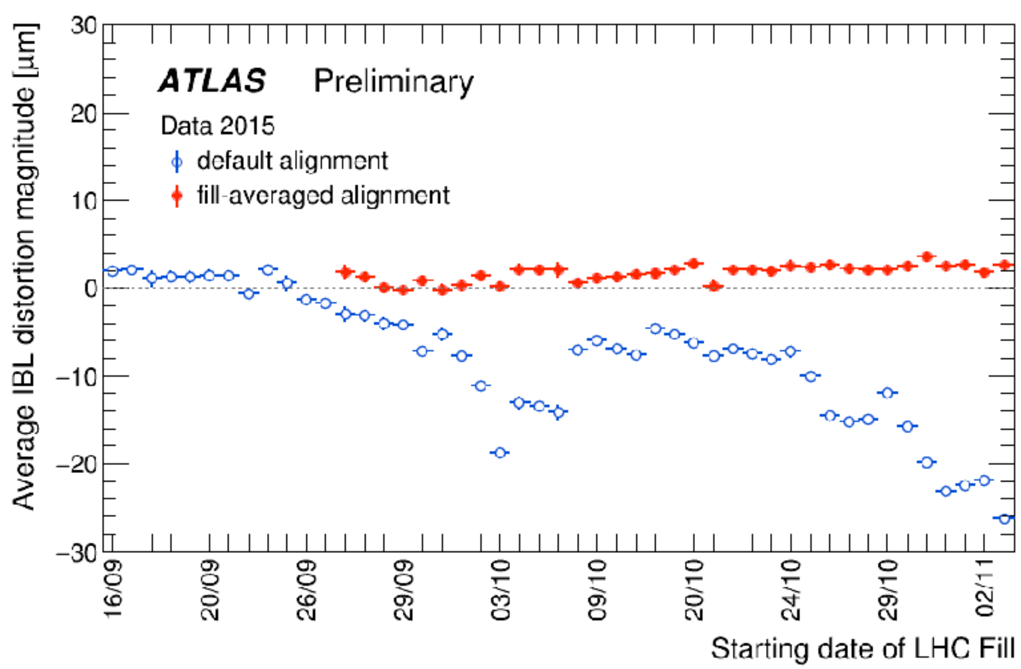
During the break between Run 1 and Run 2, ATLAS inserted the IBL, an extra layer of silicon tracker close to the beam pipe

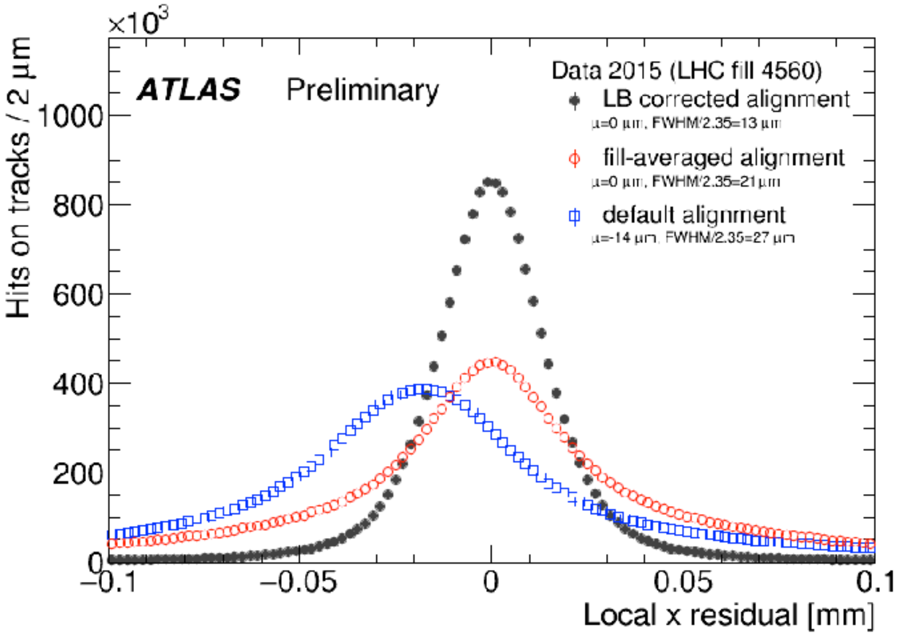
At the start of data taking in Run 2, it started to move

As time went on, the movement was very significant, much more than the detector precision so the movement could really be seen in physics distributions and data quality

Calibration







During the break between Run 1 and Run 2, ATLAS inserted the IBL, an extra layer of silicon tracker close to the beam pipe

At the start of data taking in Run 2, it started to move

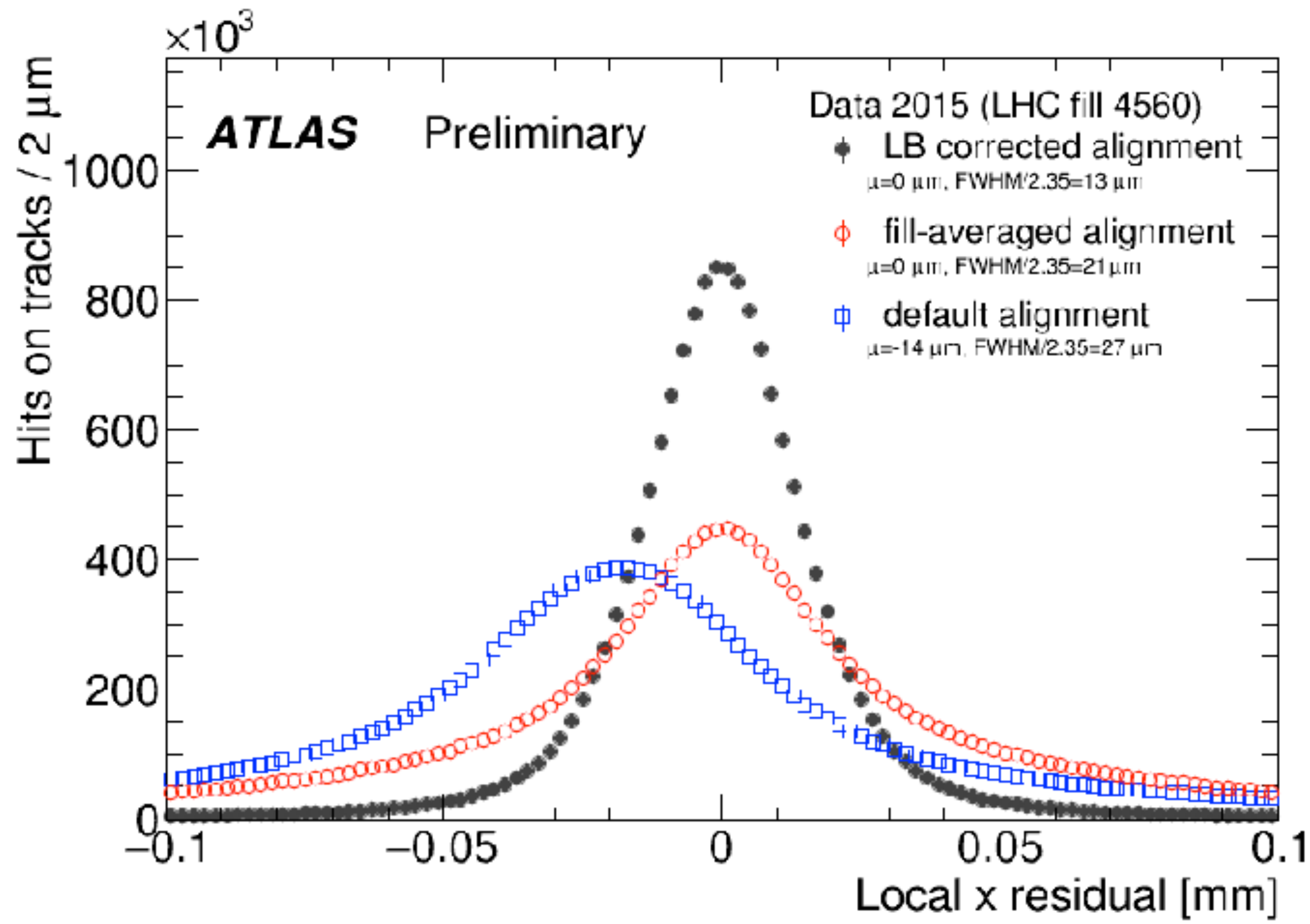
As time went on, the movement was very significant, much more than the detector precision so the movement could really be seen in physics distributions and data quality

ATLAS quickly implemented and commissioned a correction procedure as part of its calibration process

Following the correction the performance of the detector was back to nominal

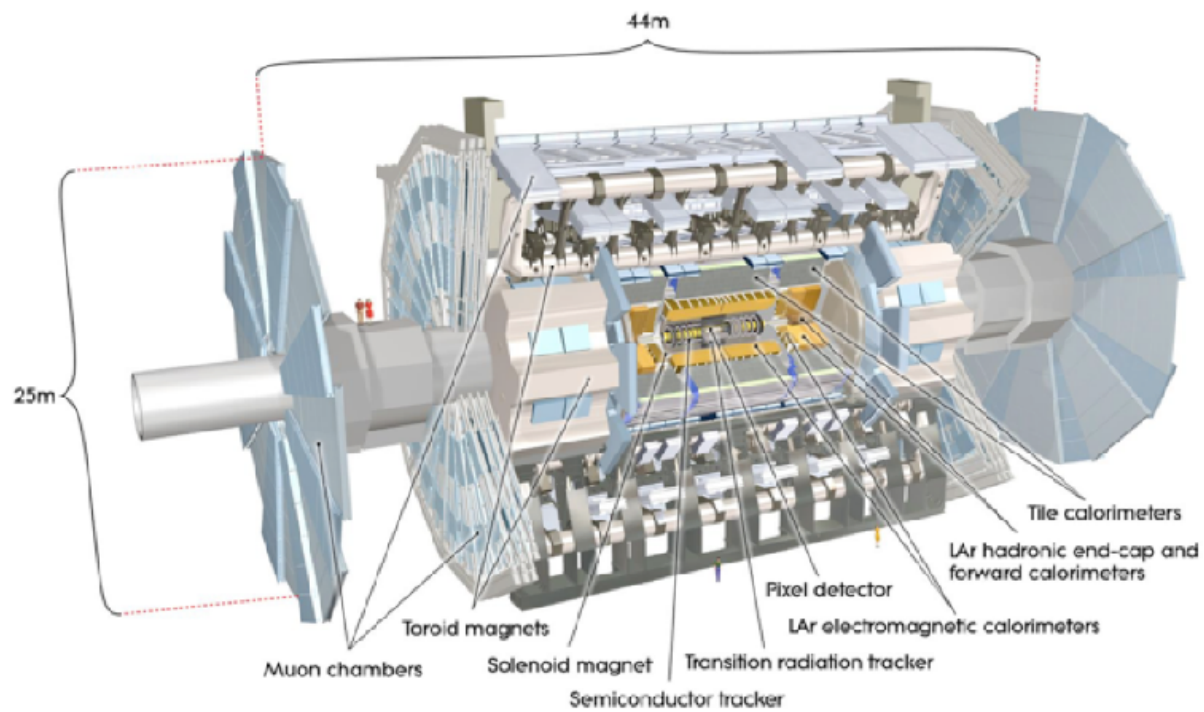
Brookhaven National Laboratory

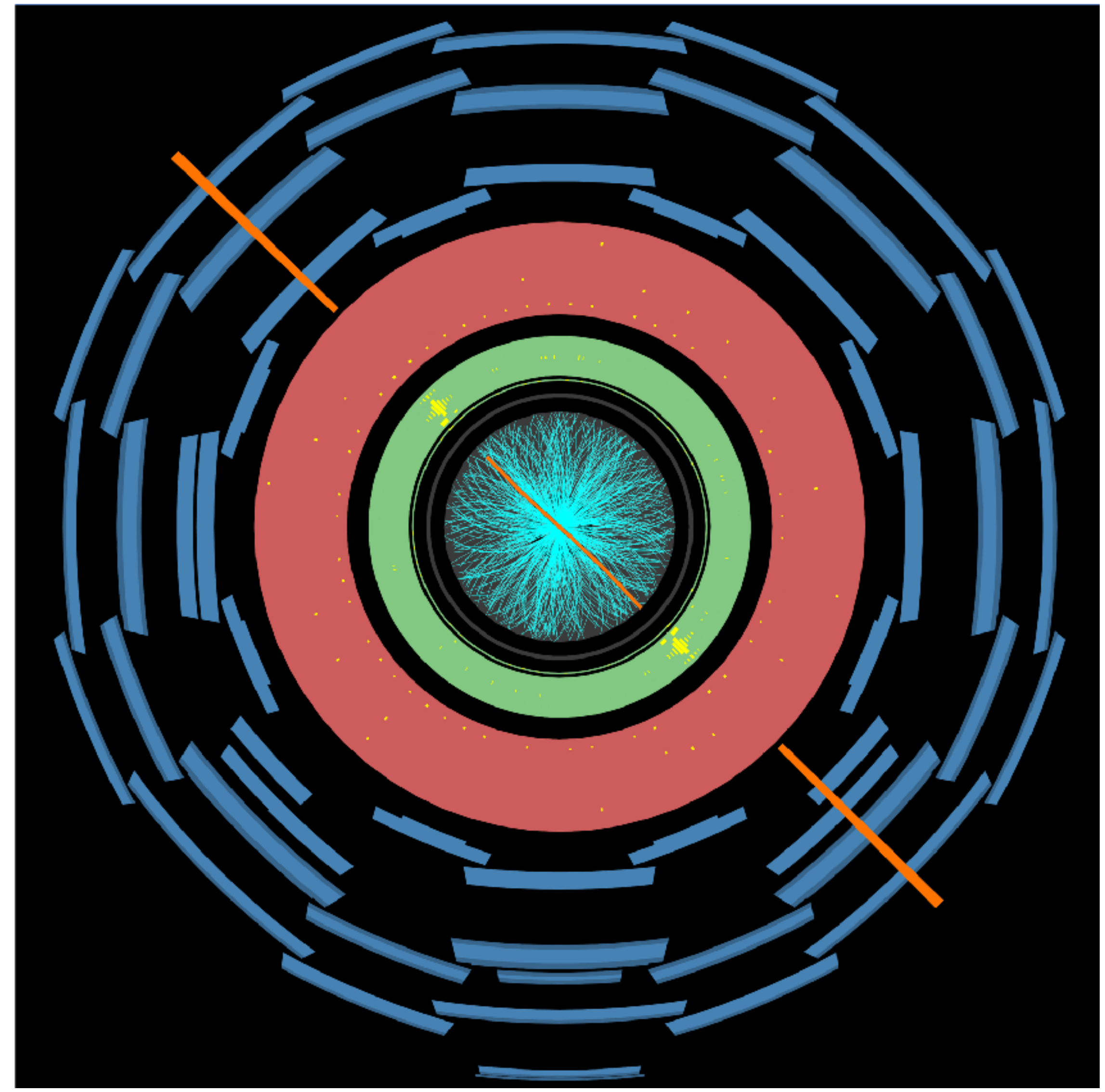
What makes good data quality?



- The **ATLAS IBL** is a good example of a data quality plot
- We need a reference, here that would be the **black** histogram, how we expect the data to look
- If the data quality shifter sees the blue or red histogram, they will raise the alarm!

What am I?





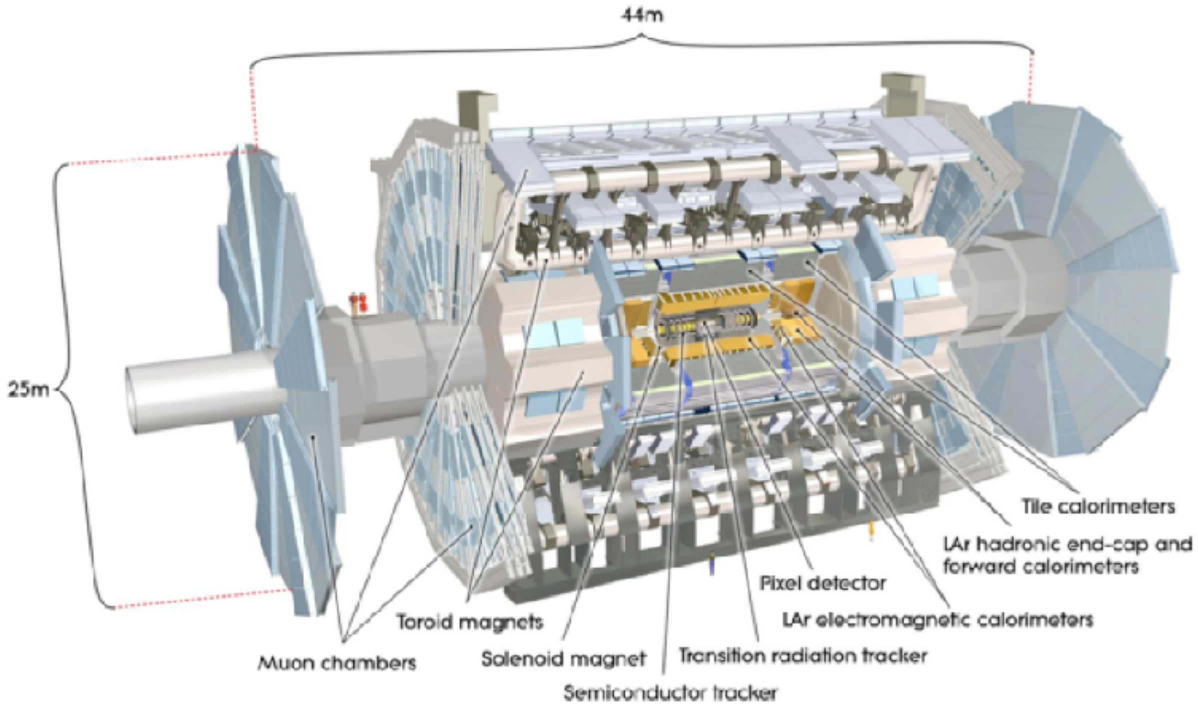
In the event display, right, we are looking down the beam pipe, which we call ***the transverse plane*** here

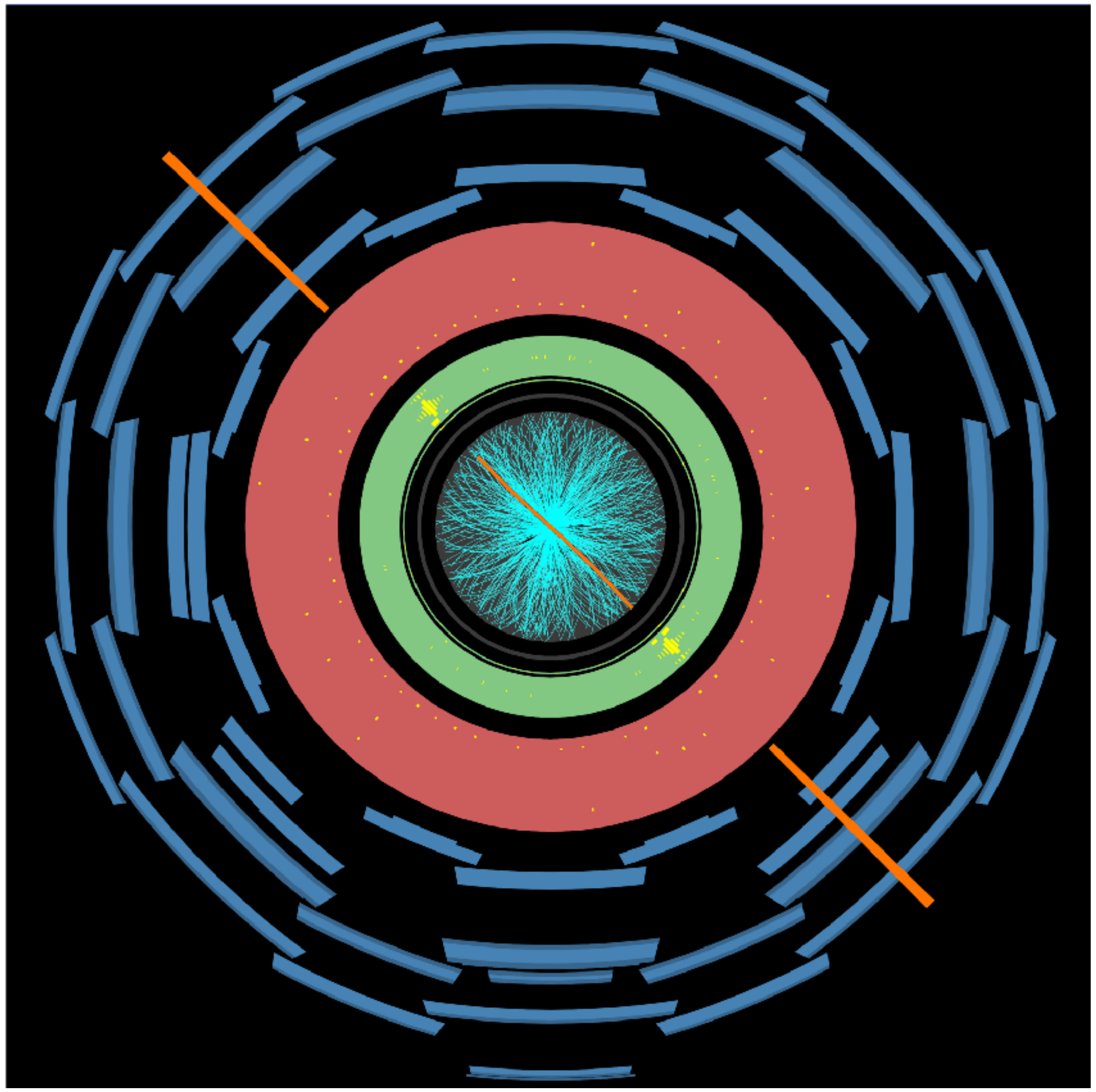
***Question**: What is the sum of the momentum in this plane **before** and **after** the particle collision*

Question: What physics process are we observing in this event?

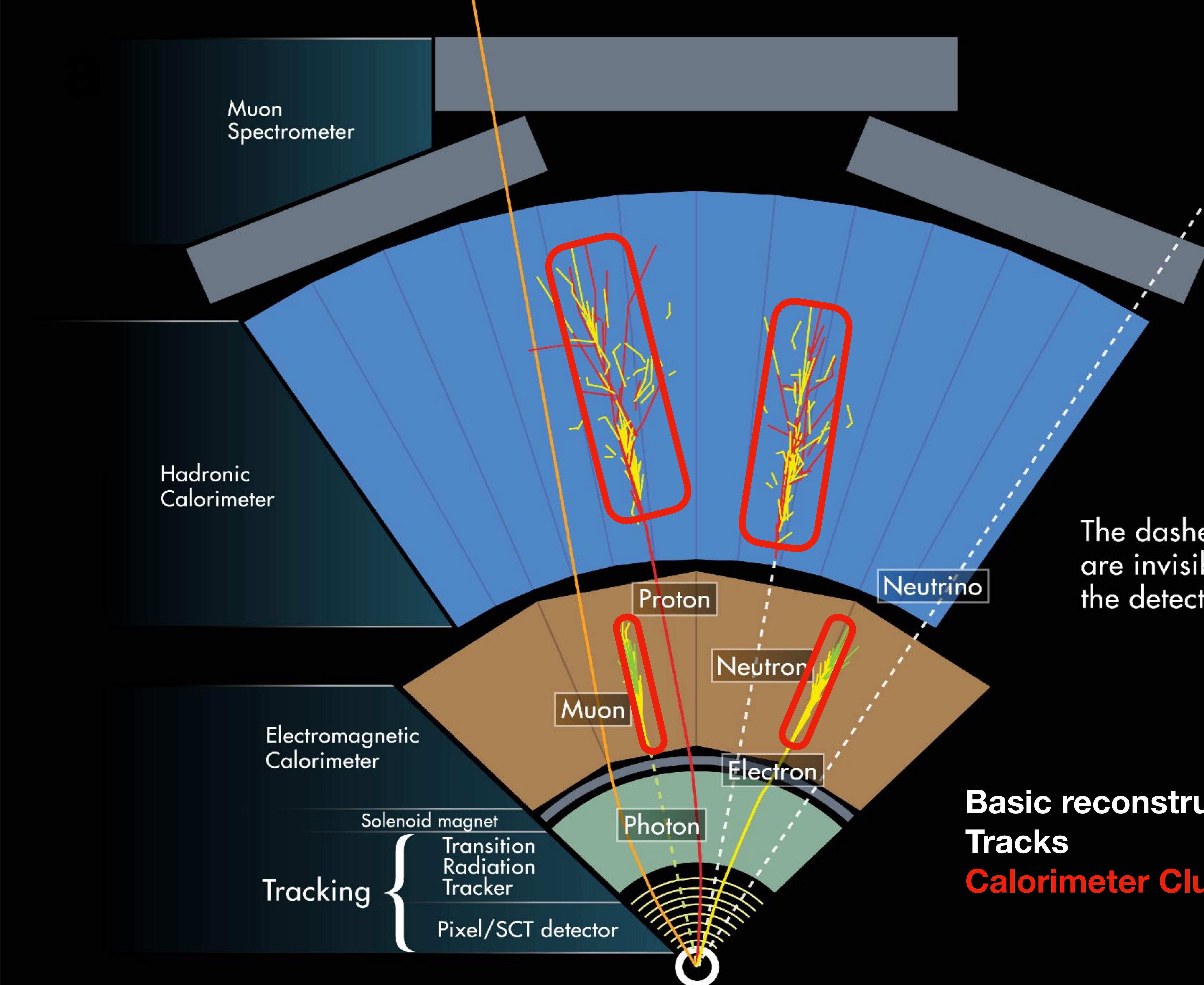
Brookhaven National Laboratory

Neutrinos



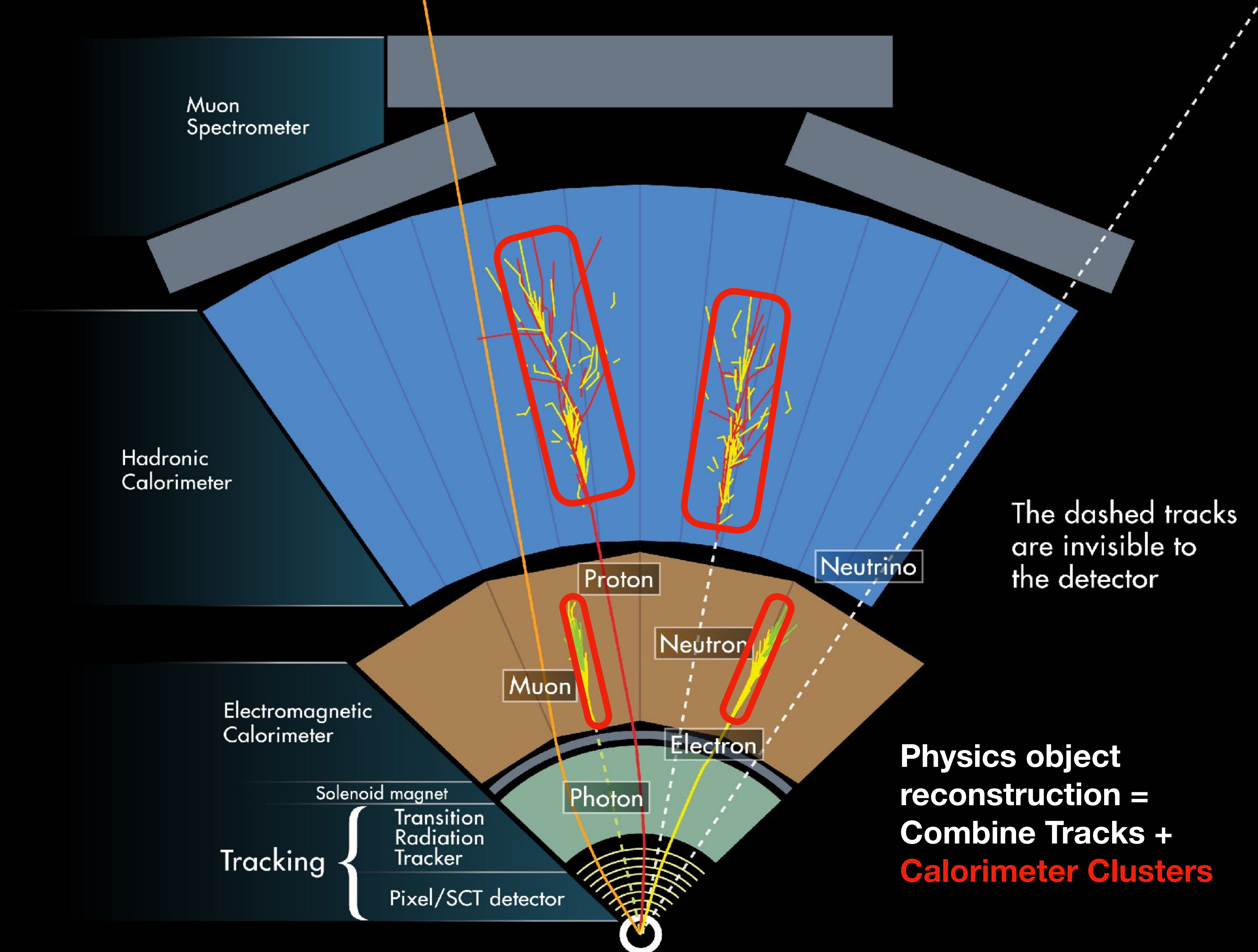


- ***Question**: How would this event display look if we had produced a **W boson** instead of a **Z boson** ?*



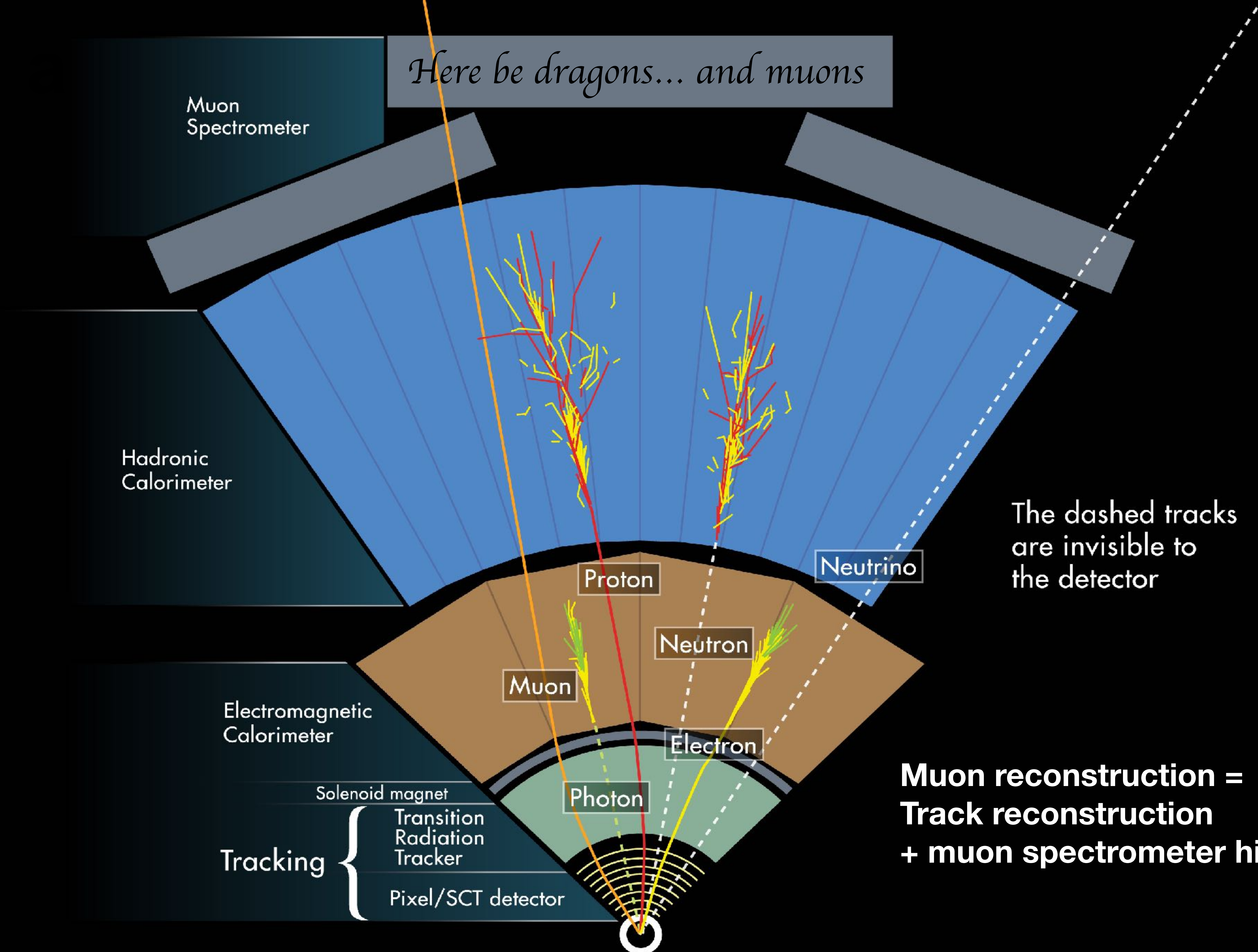
The dashed tracks are invisible to the detector

Basic reconstruction =
Tracks
Calorimeter Clusters



Physics object reconstruction = Combine Tracks + Calorimeter Clusters

Here be dragons… and muons



Muon reconstruction =
Track reconstruction
+ muon spectrometer hits