From Raw data to Physics Results (2/3)

Paul Laycock

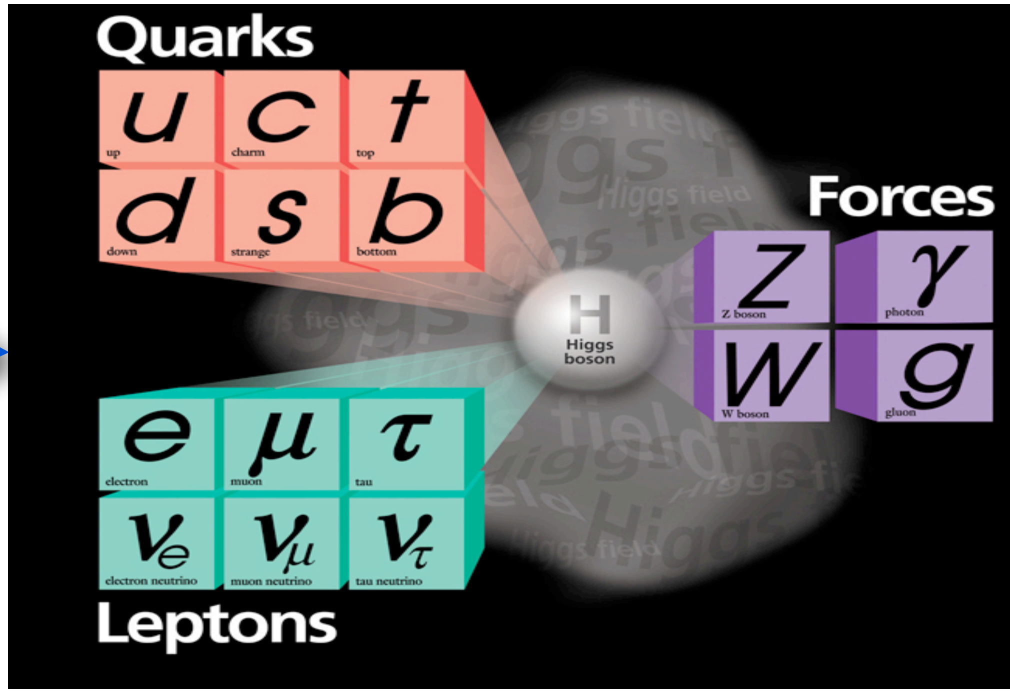
July 4th 2022

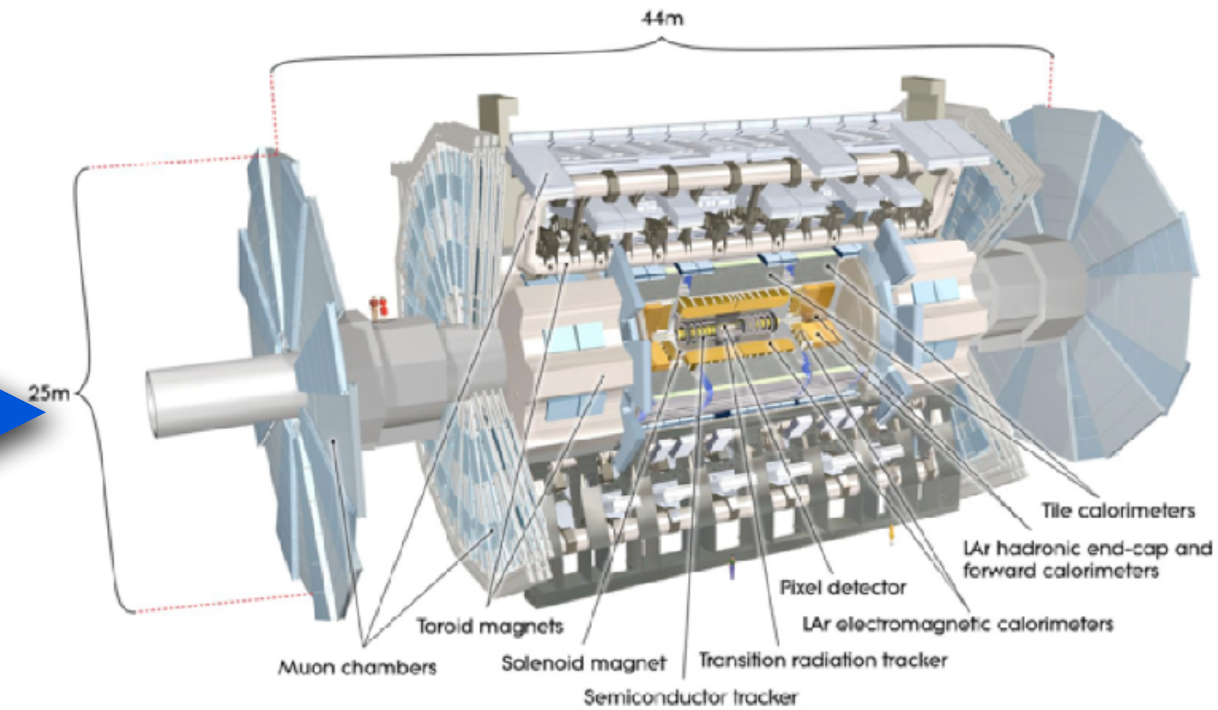
Brookhaven National Laboratory

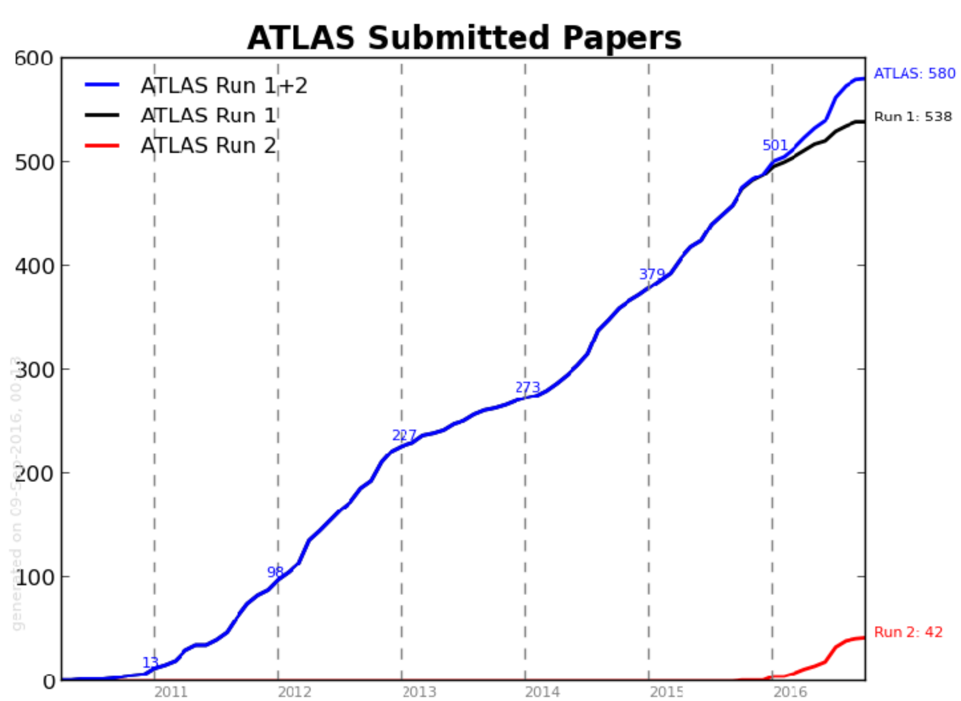
U.S. DEPARTMENT OF ENERGY

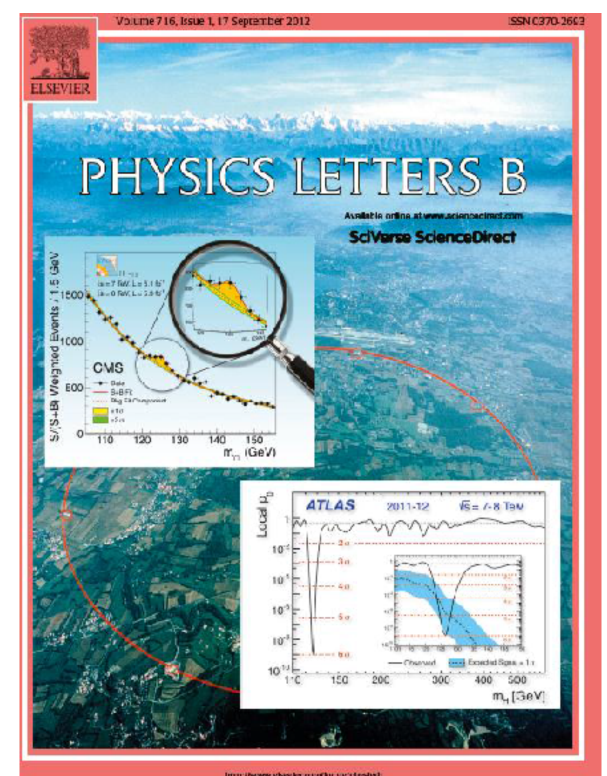
@BrookhavenLab

The particle physics cycle









Course outline

Lecture 1

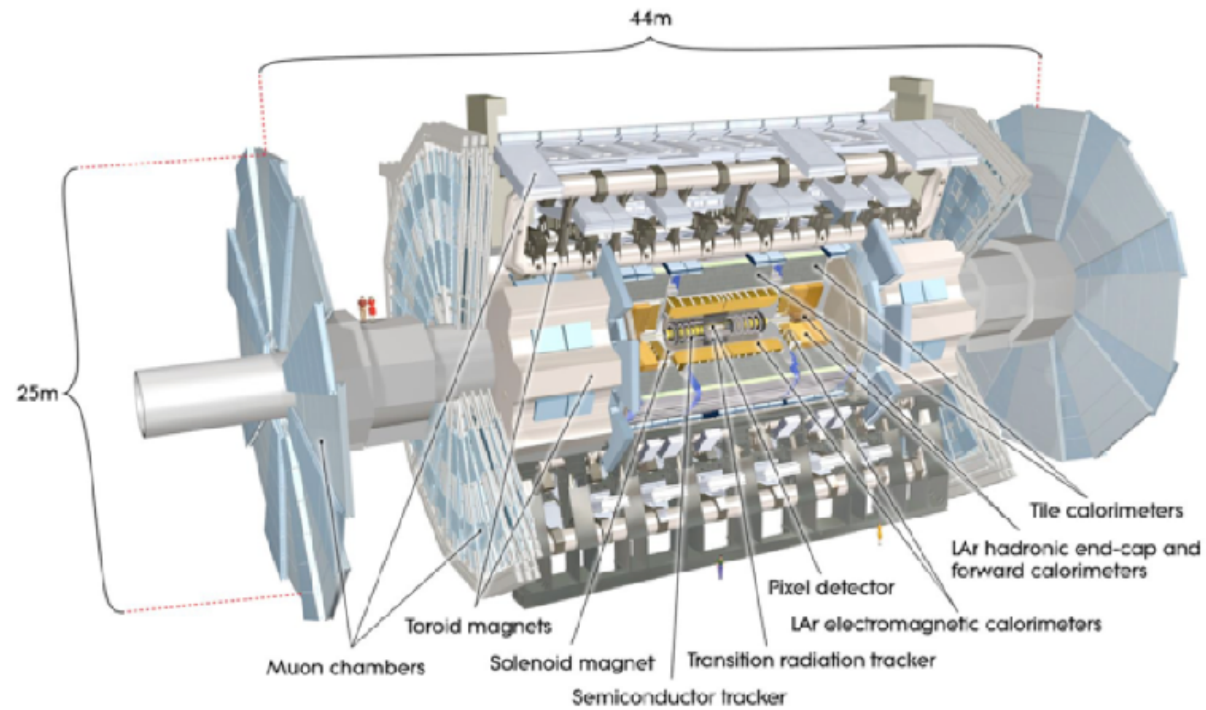
- The journey of raw data from the detector to a publication

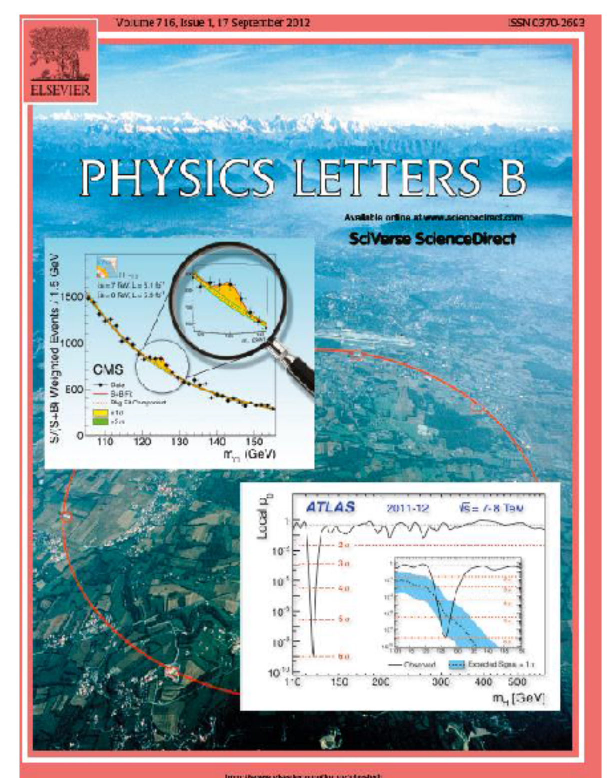
Lecture 2

- How we reconstruct fundamental physics processes from raw detector data

Lecture 3

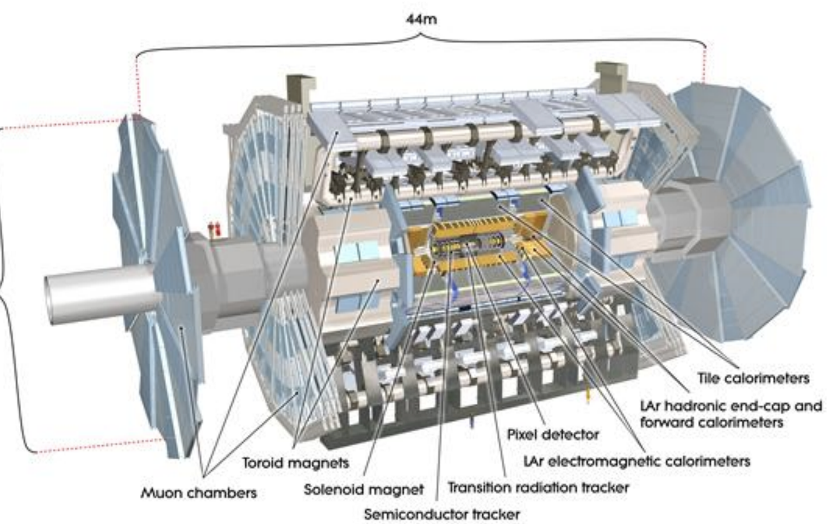
- How we extract our signals from the mountain of data, finding needles in the haystack

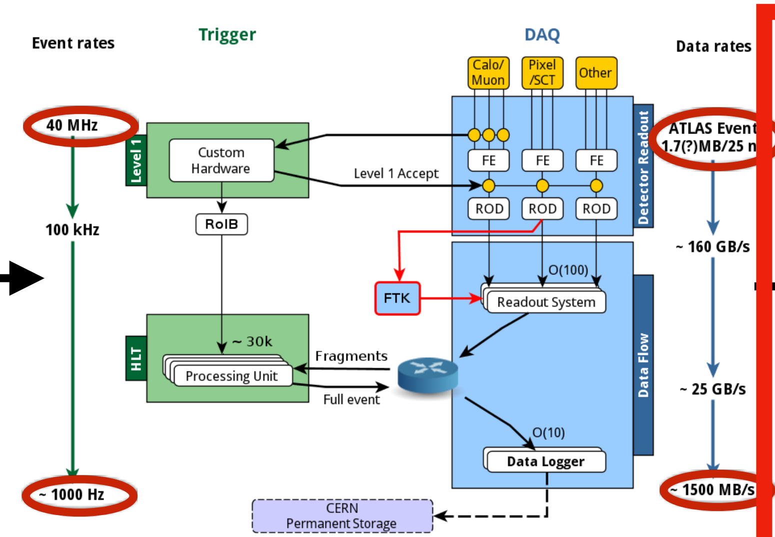




Brookhaven National Laboratory

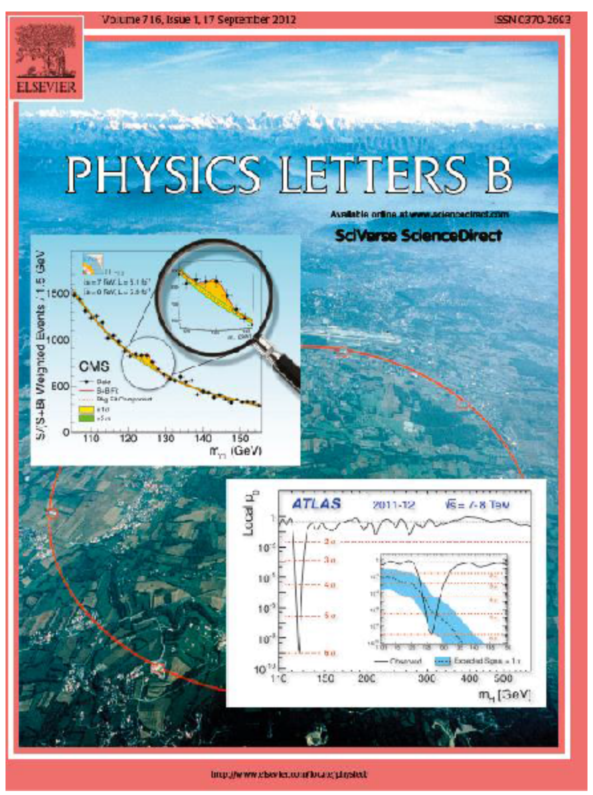
Data’s journey

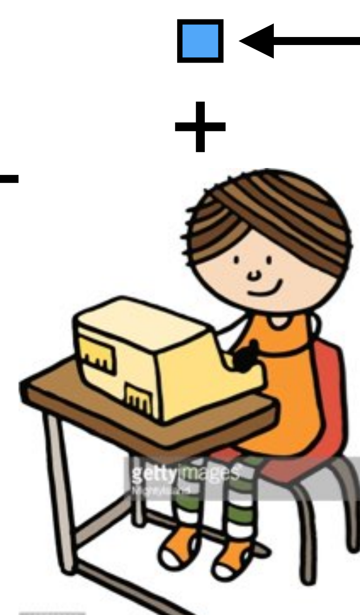


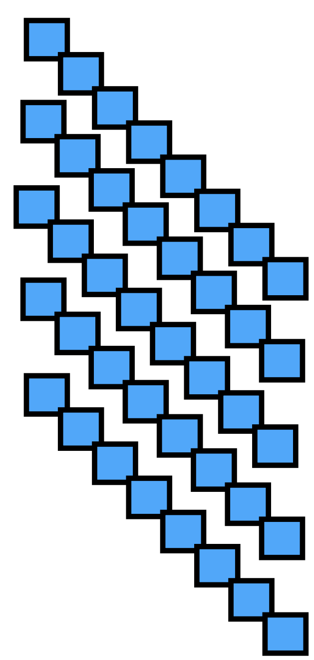


Data Preparation

Distributed computing



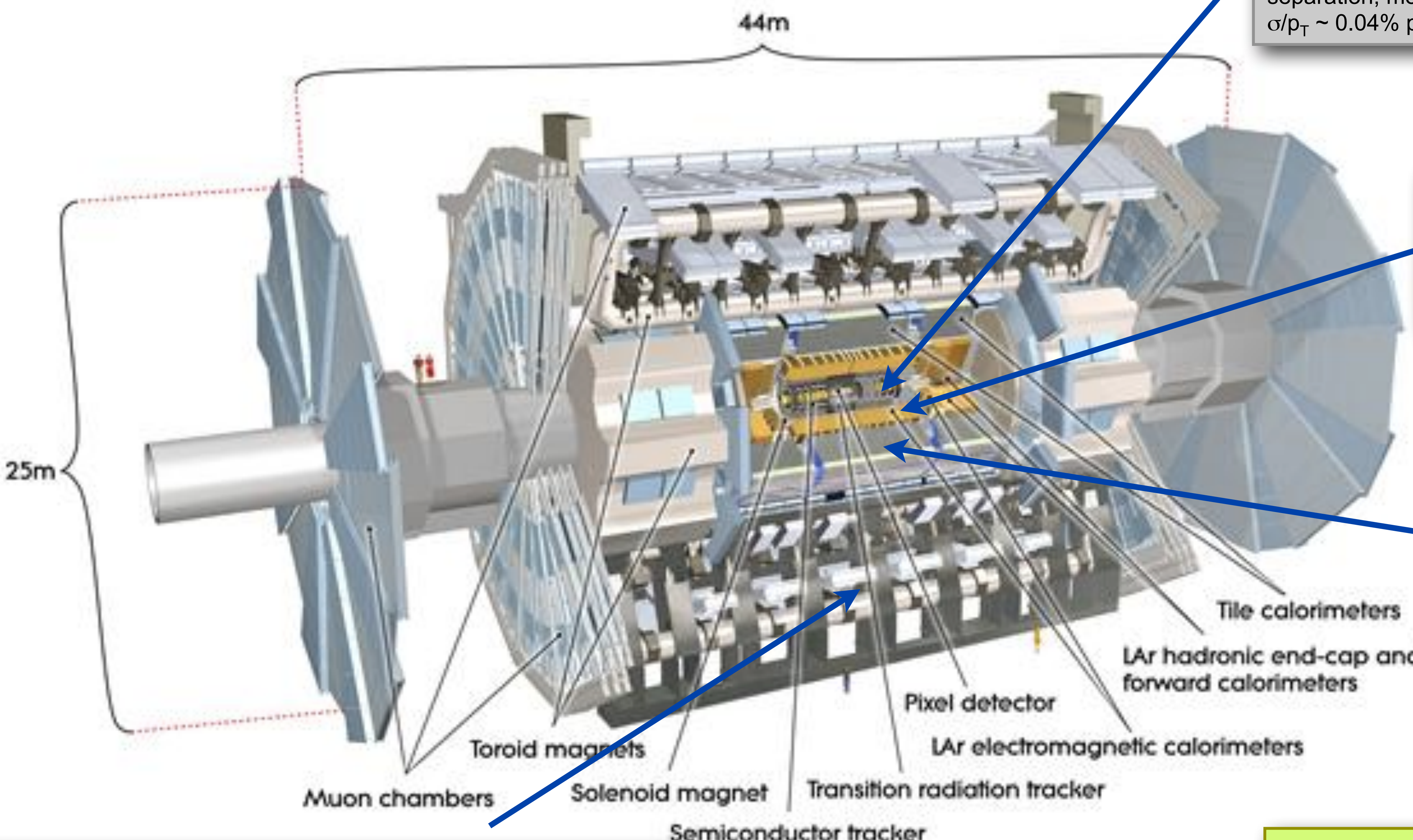




The ATLAS Detector @ LHC

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

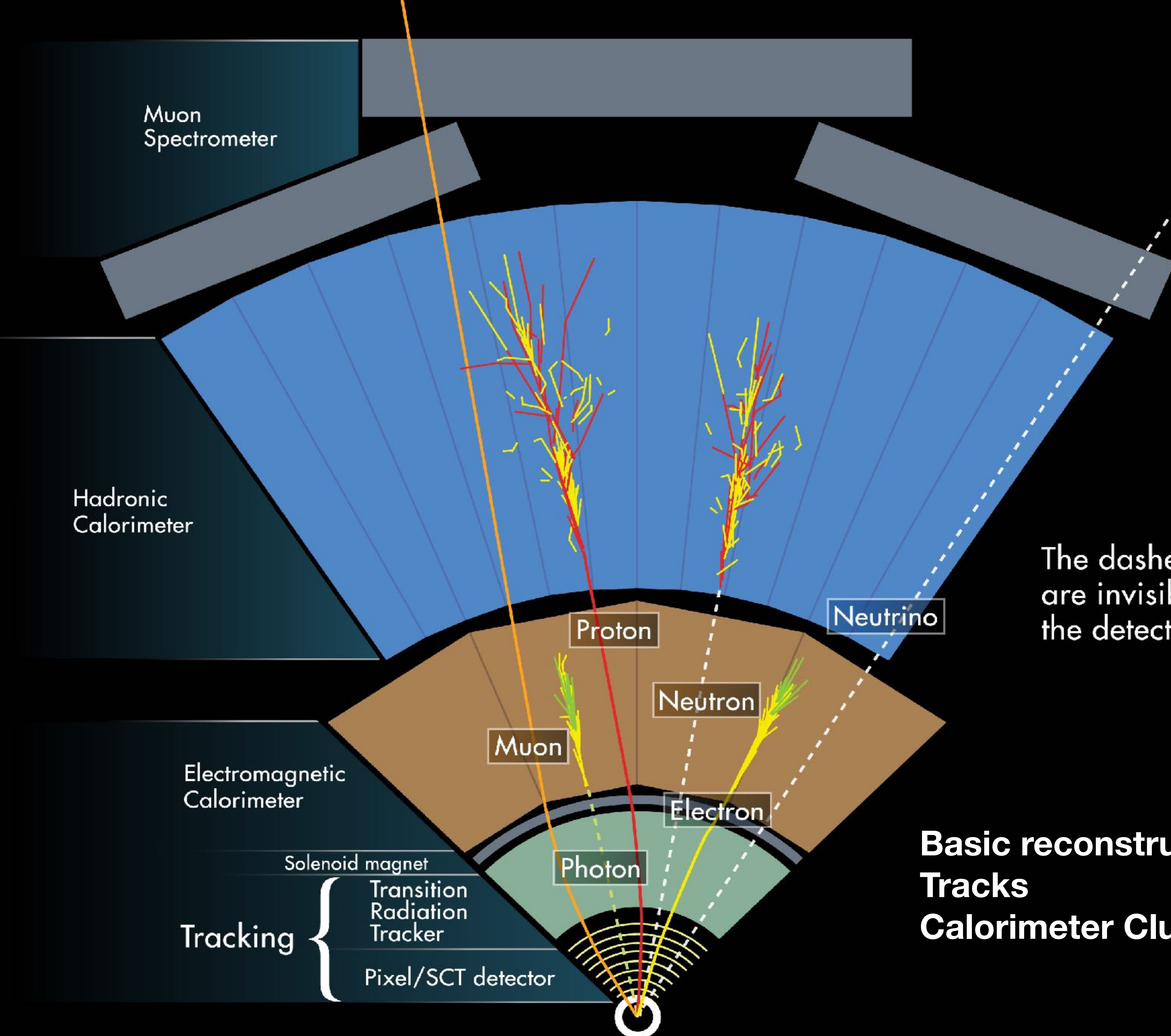
Millions of detector readout channels read out to reconstruct one "event"

Data Preparation

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

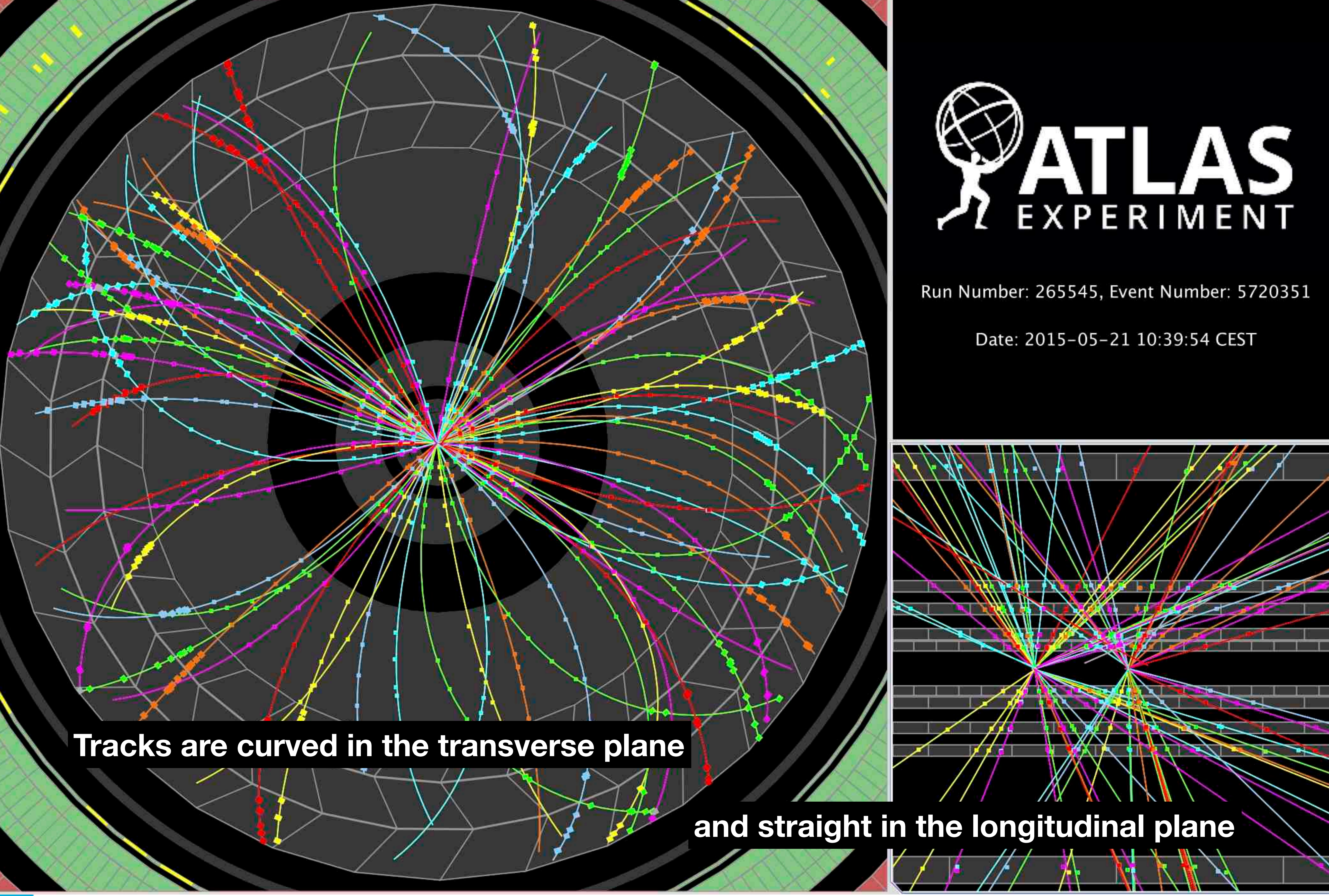
- reliable, high quality data (yes, we ***reject*** low quality data)
- the ***best performance*** from our detectors
- readiness for ***physics analysis***

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



The dashed tracks are invisible to the detector

Basic reconstruction =
Tracks
Calorimeter Clusters



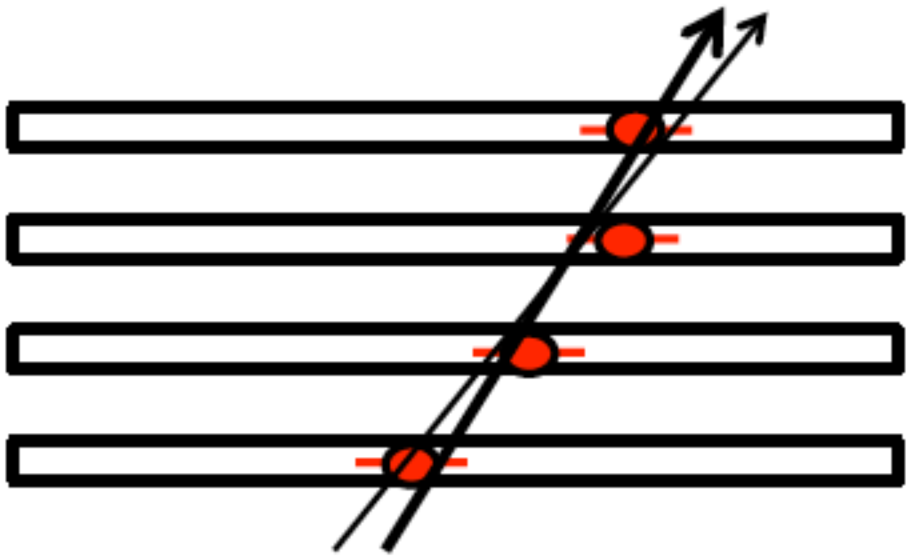
Modern track reconstruction uses Machine Learning: <u>Connect the Dots</u>

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.

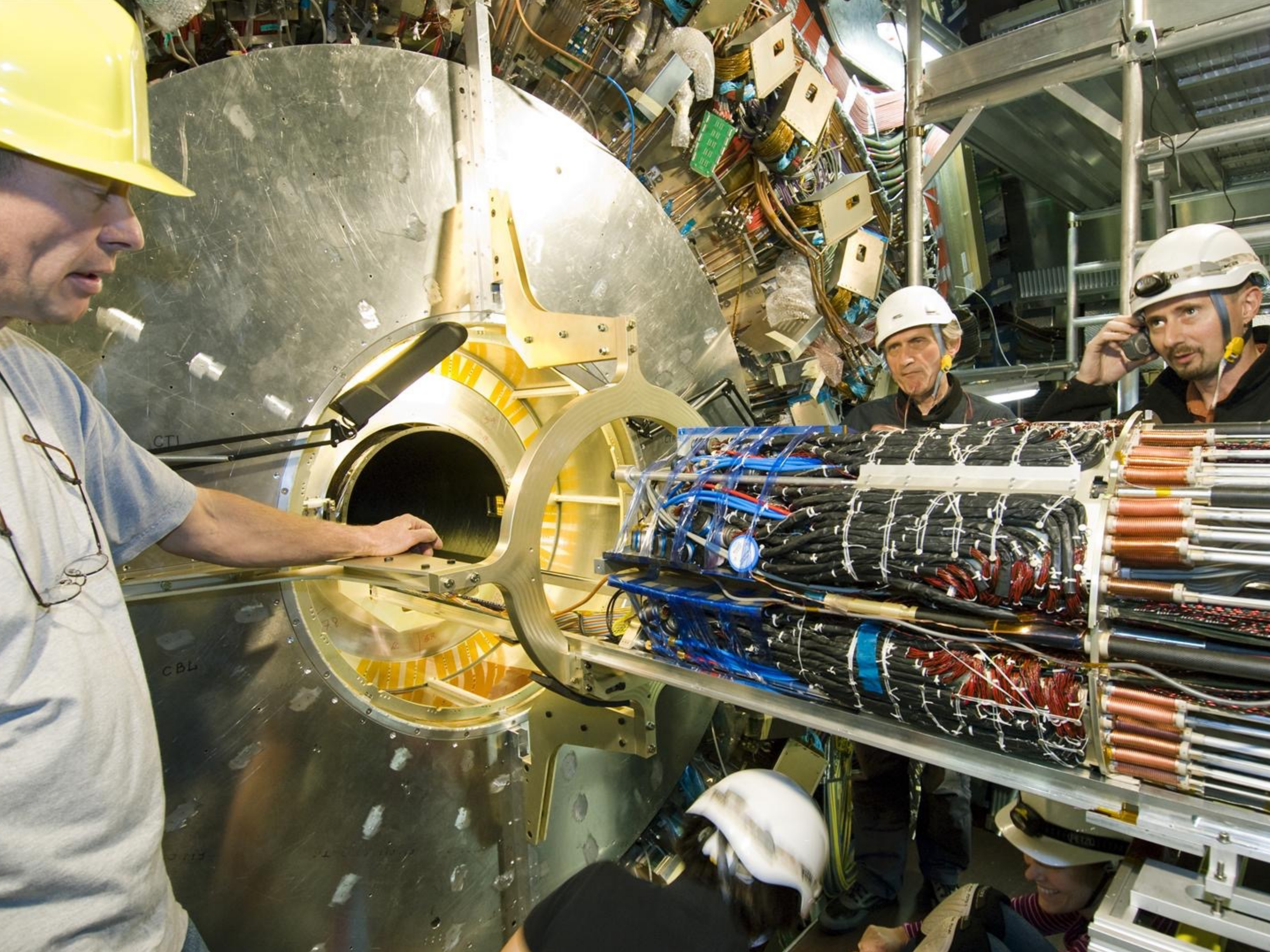
Data Preparation

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…

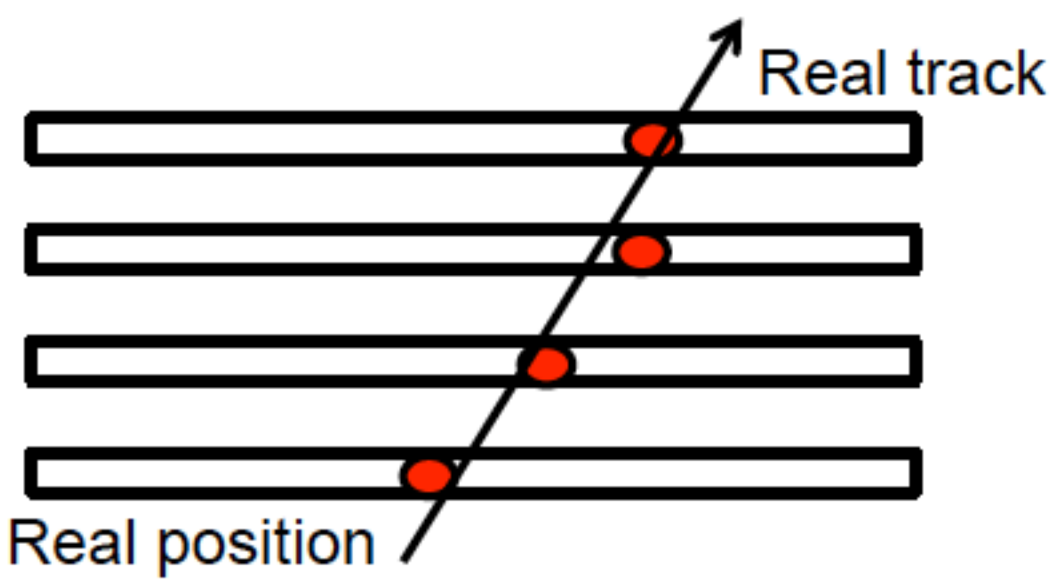
Brookhaven National Laboratory


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.

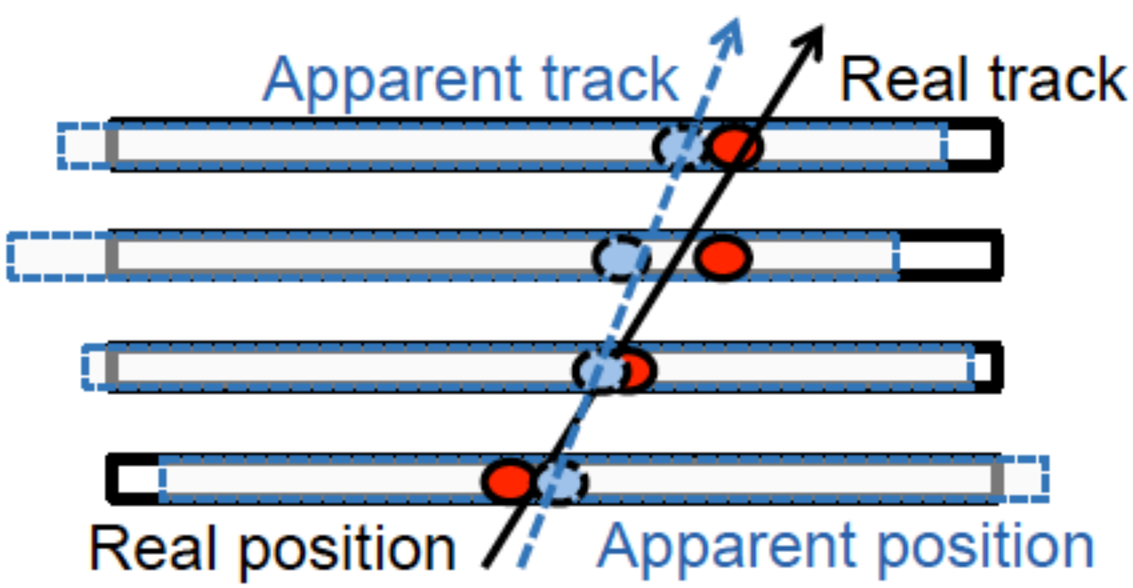


Correcting detector effects - calibration

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

Q. What effects would we see due to the presence of material?

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



Real vs perfect tracking detectors

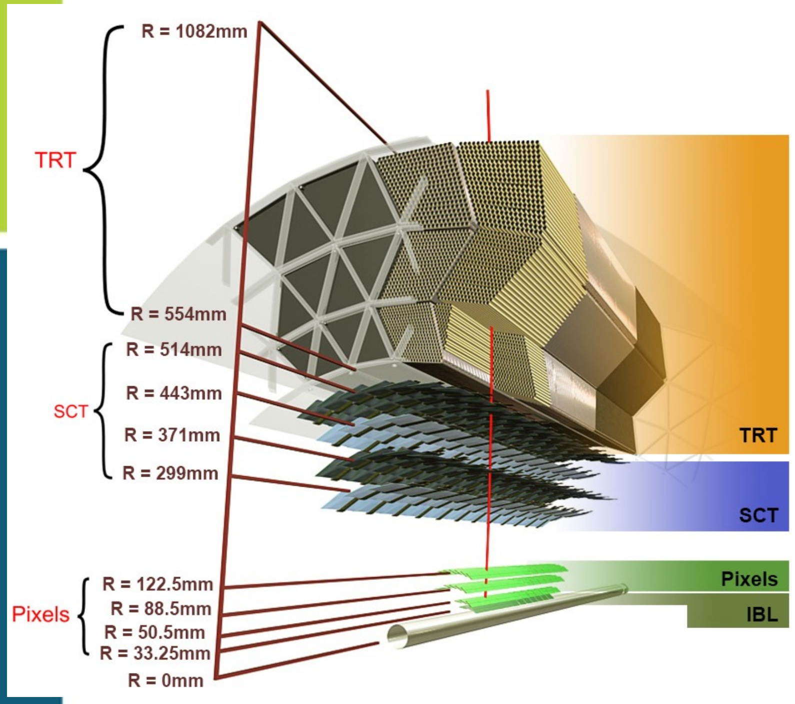
The perfect tracking detector

- **is constructed from zero mass material**
- **has no noise**
- **is 100% efficient**
- **and has infinite resolution**

A real tracking detector

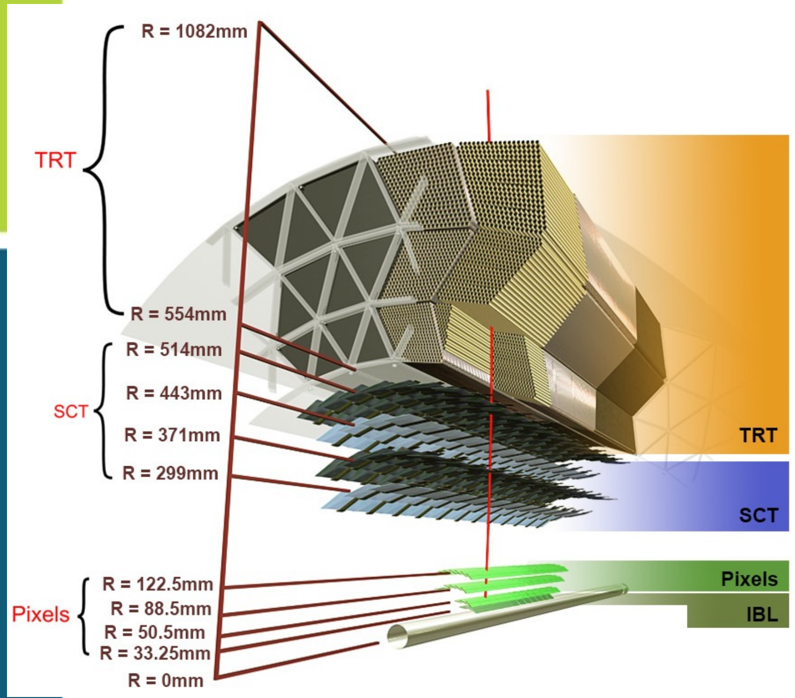
- **is constructed from real material**
  - **particles interact with the detector and scatter, altering the particle trajectory**
- **suffers from noise**
  - **noise can be confused with particle tracks**
- **has less than 100% efficiency**
  - **particles are not always detected and there can even be dead regions**
- **has finite resolution**
  - **it may not always be possible to resolve two particle trajectories**

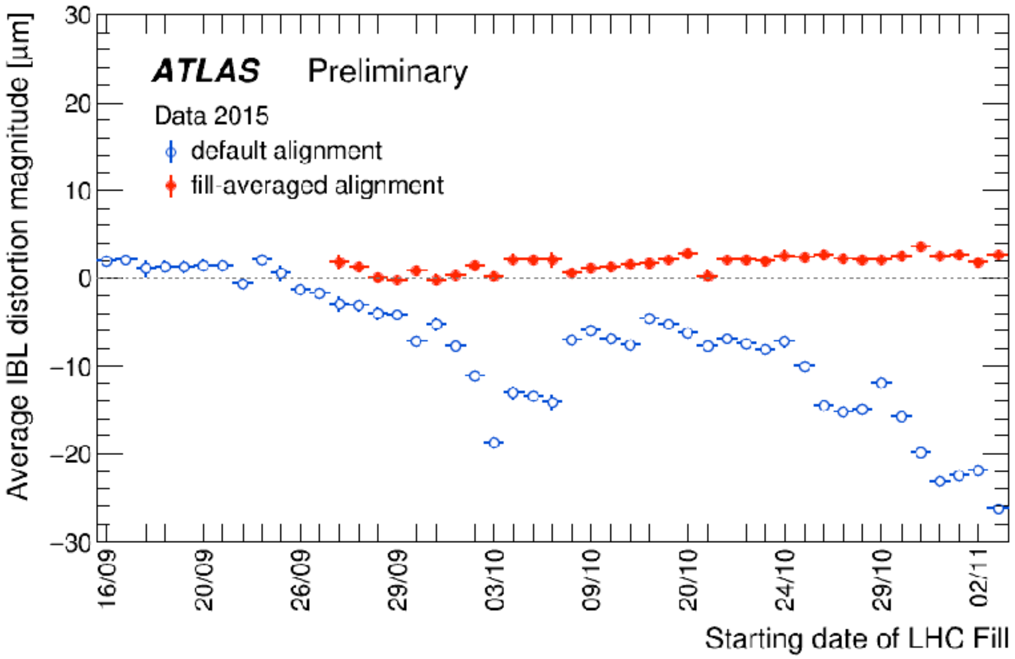
Calibration



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



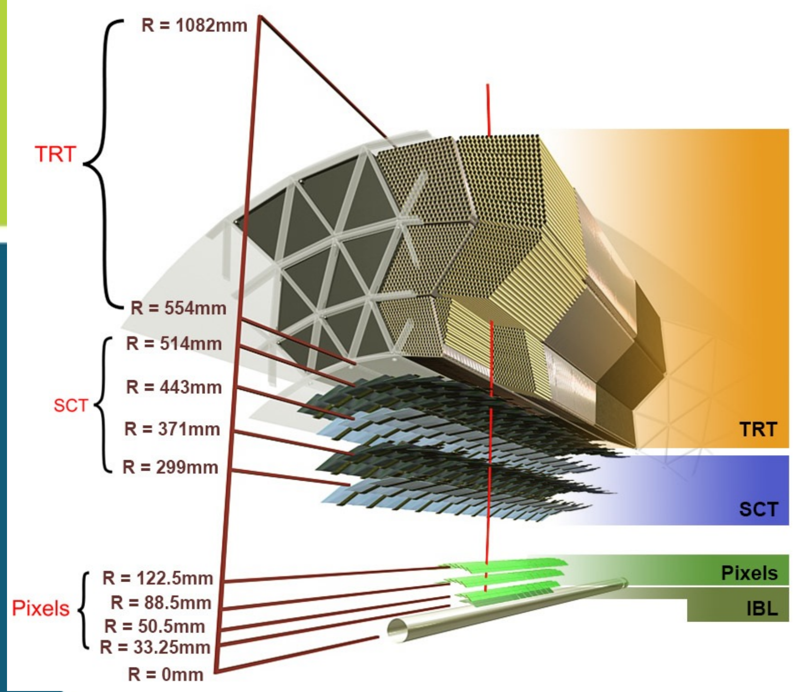


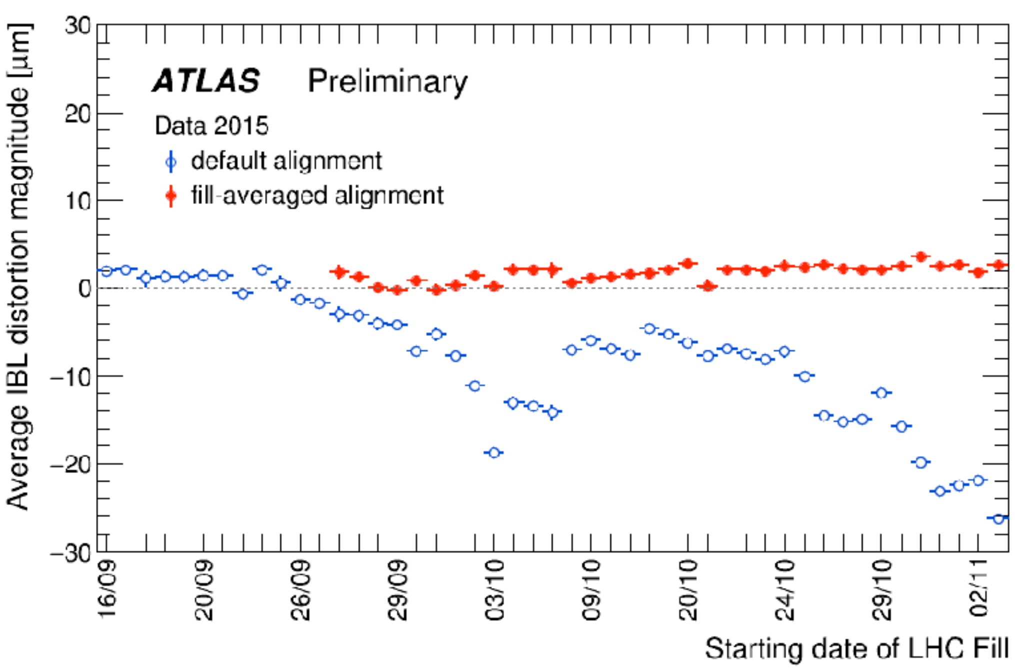
During the break between Run 1 and Run 2, ATLAS inserted the IBL, an extra layer of silicon tracker close to the beam pipe

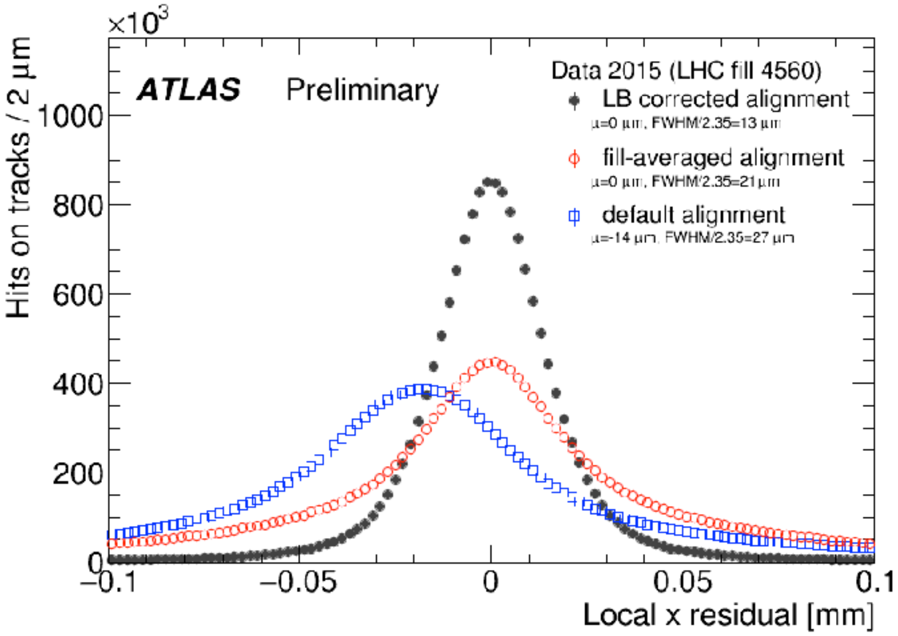
At the start of data taking in Run 2, it started to move

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

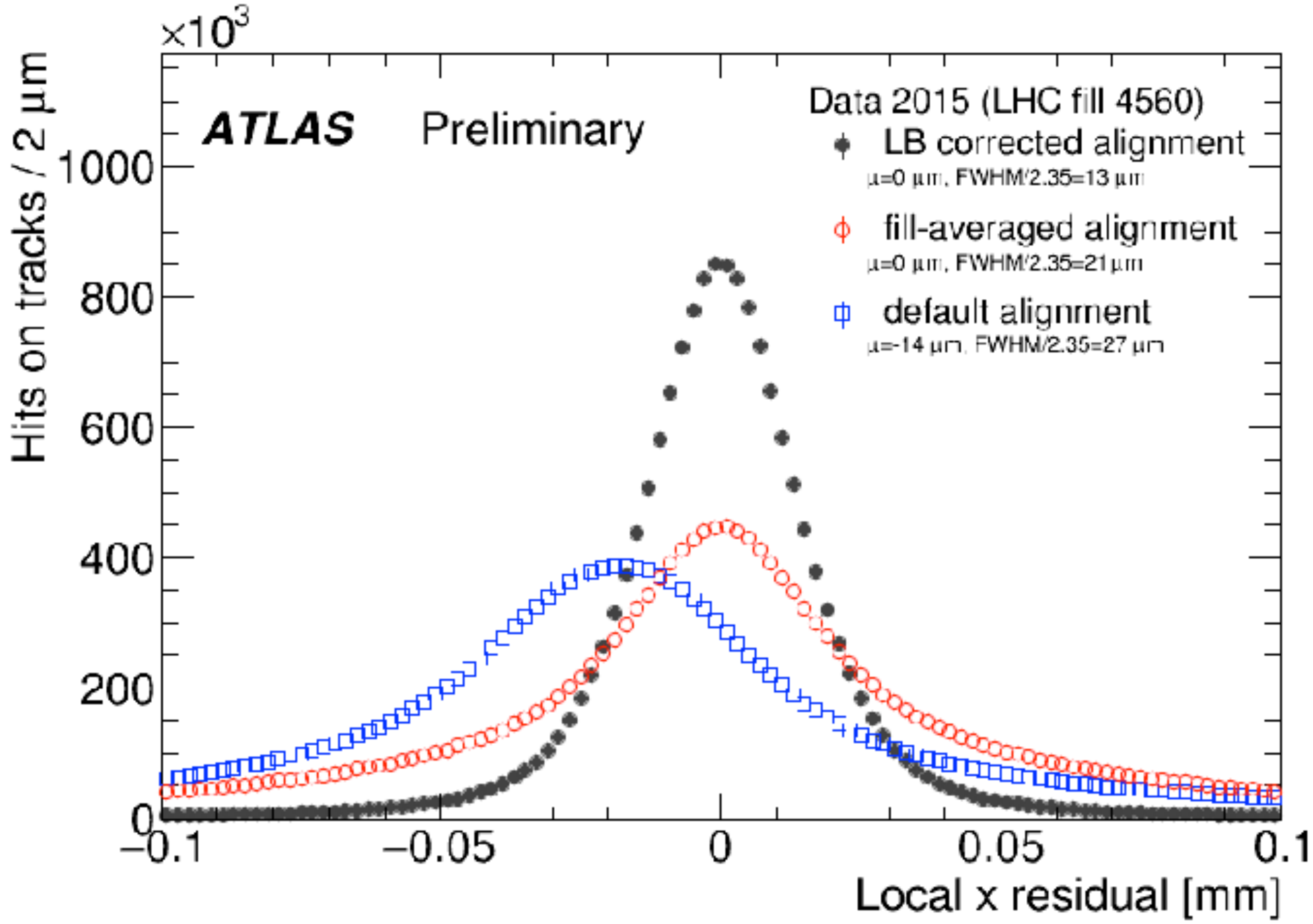
At the start of data taking in Run 2, it started to move

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

Calibration quality



Thinking back to the difference between ***accuracy*** and ***precision***, which versions of the data are ***accurate***, and which are ***precise***? Which are both?

Brookhaven National Laboratory

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***

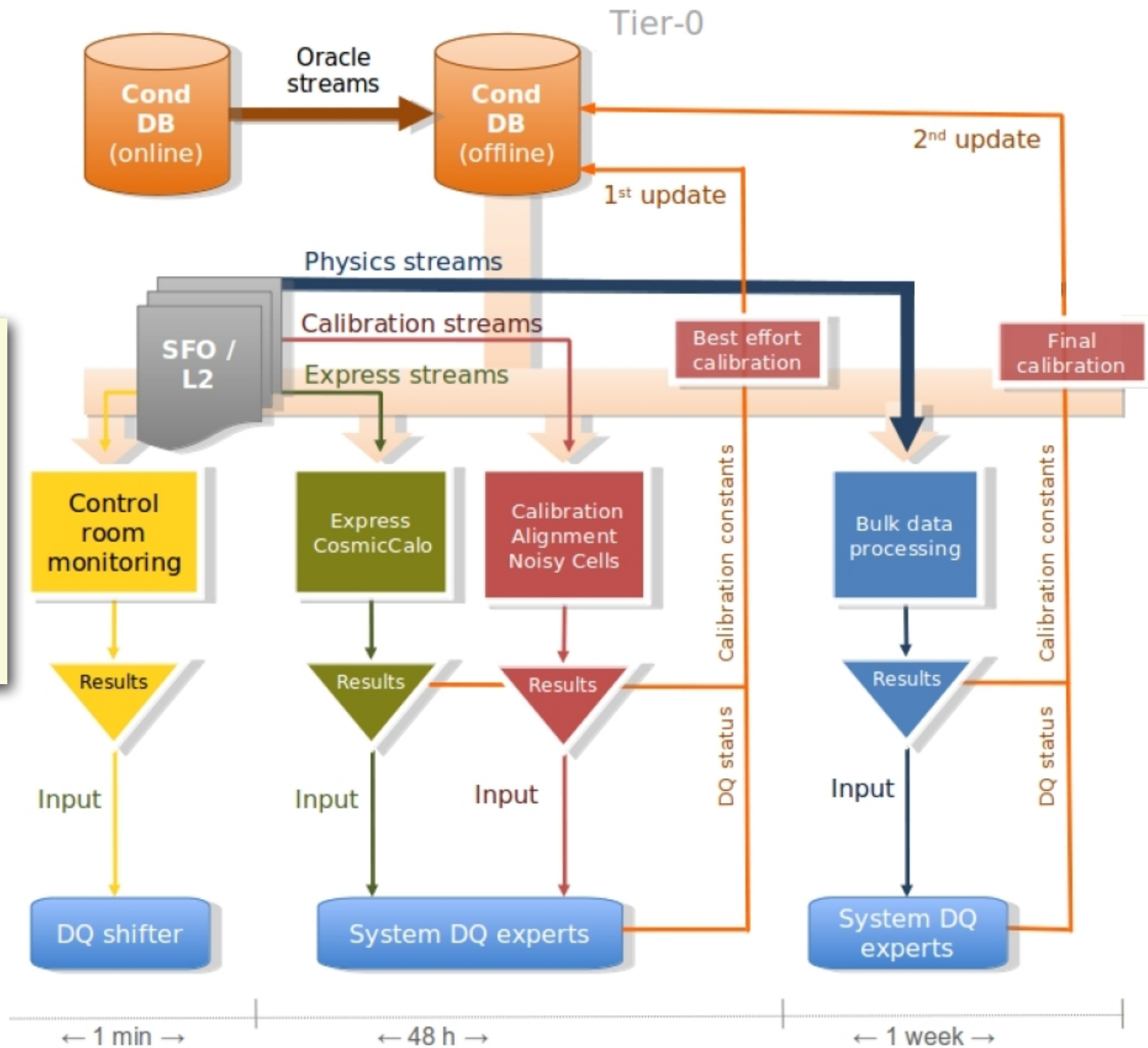
1. **Reconstruct physics signals** from the data ✔
   - Produce information like how many muons does the event have?
2. **Calibrate** the detectors ✔
   - Correct imperfections, account for changes over time…
3. Make sure that the **data quality** is excellent, also in real time
   - Maximise the amount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

Data Quality

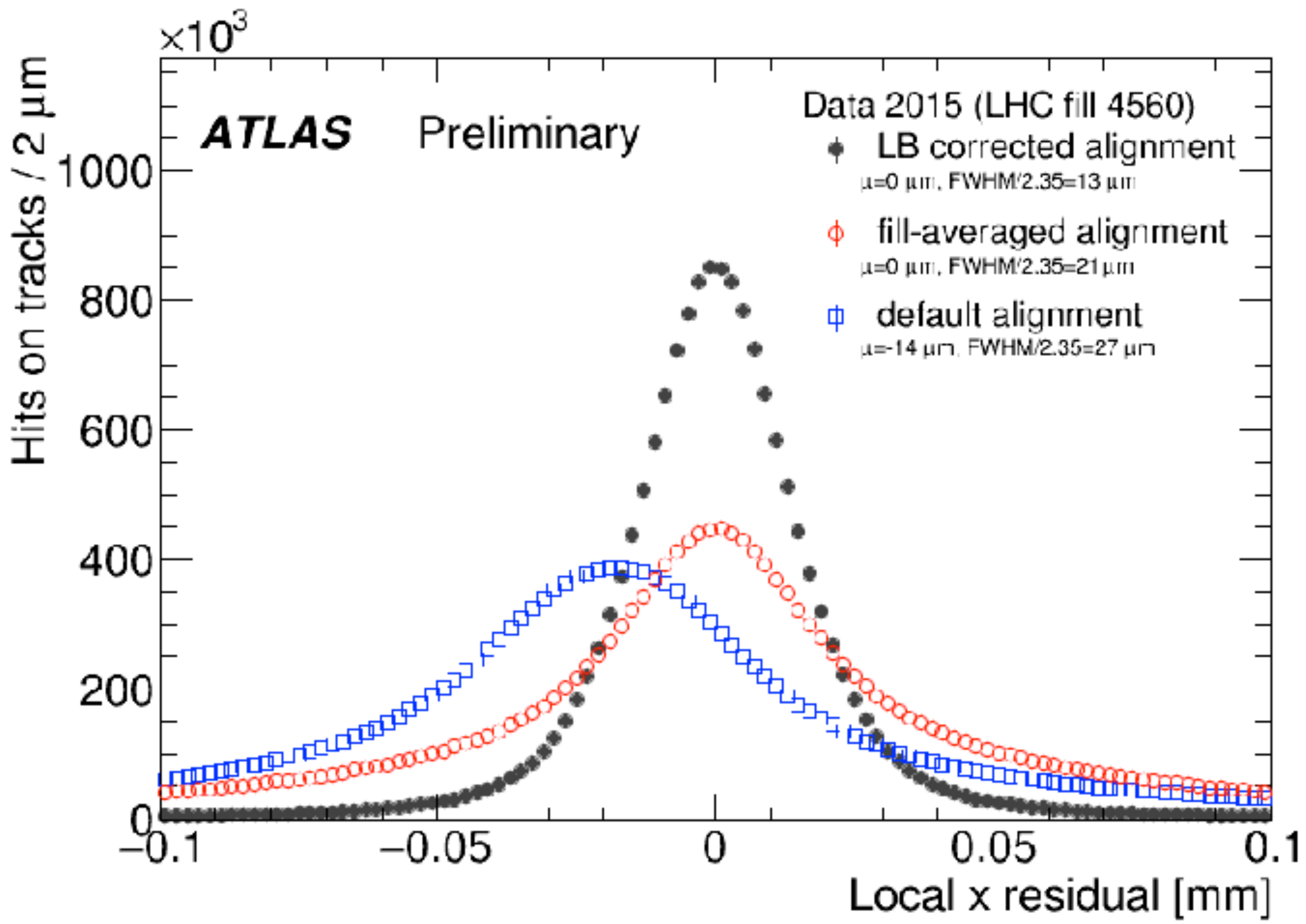
Check during data taking

Check a fraction of the data with a quick calibration

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



What makes good data quality?



The **ATLAS IBL** is a good example of a ***data quality*** problem

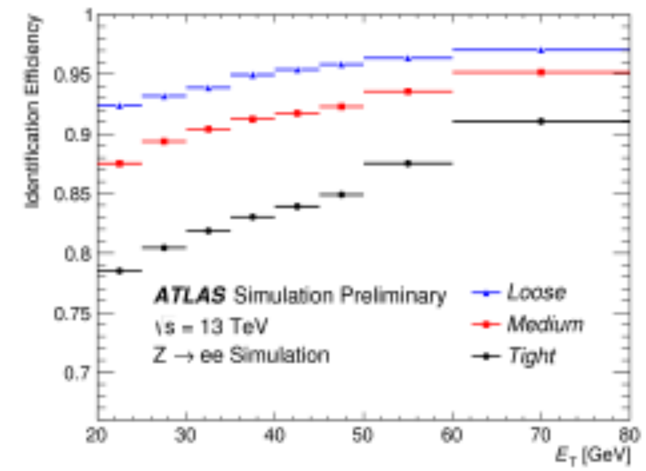
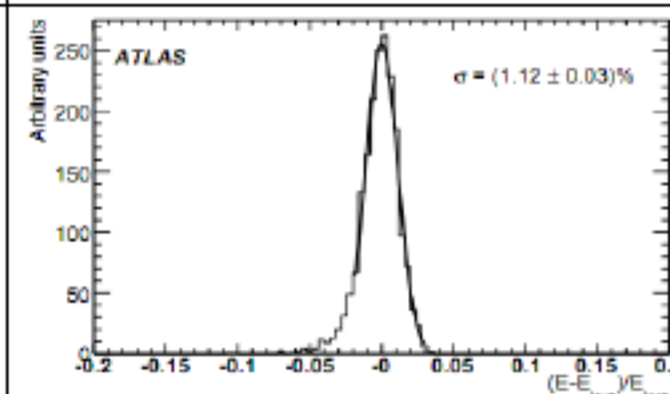
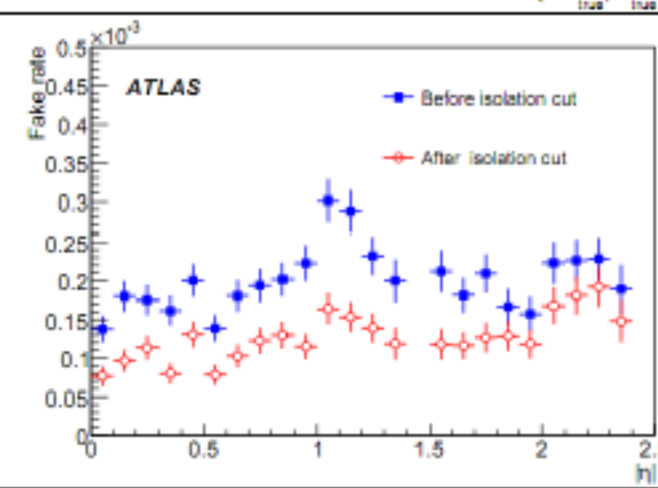
- Potential data quality issues need to be monitored

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!

Brookhaven National Laboratory

Reconstruction figures of merit and data quality

	Definition	Example		Needs be:
Efficiency	how often do we reconstruct the object	electron identification efficiency = (number of reconstructed electrons) / (number of true electrons) in bins of transverse momentum		High
Resolution	how accurately do we reconstruct the quantity	energy resolution = (measured energy – true energy) / (true energy)		Good
Fake rate	how often we reconstruct a different object as the object we are interested in	a jet faking an electron, fake rate = (Number of jets reconstructed as an electron) / (Number of jets) in bins of pseudorapidity		Low

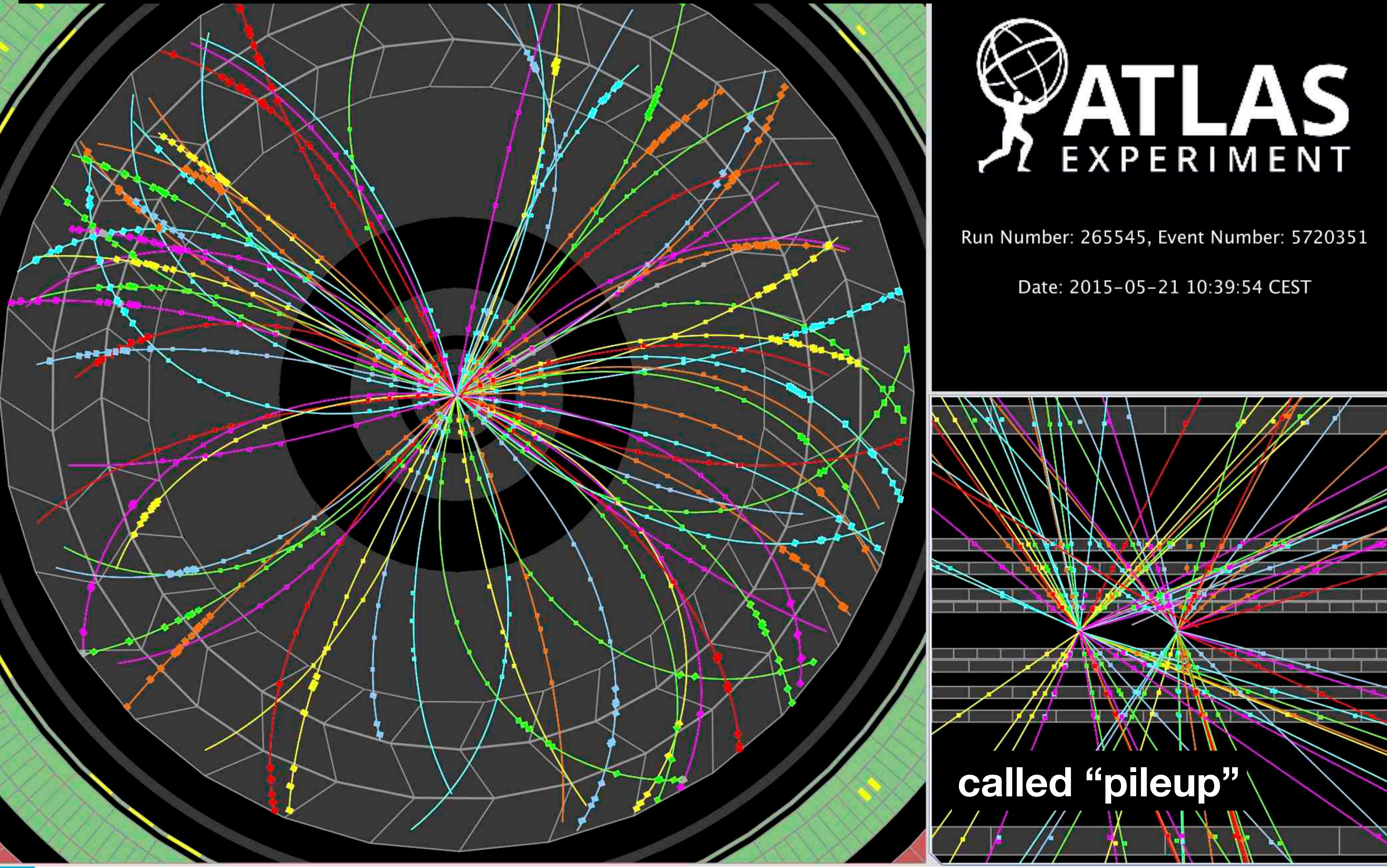
Data Preparation

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***

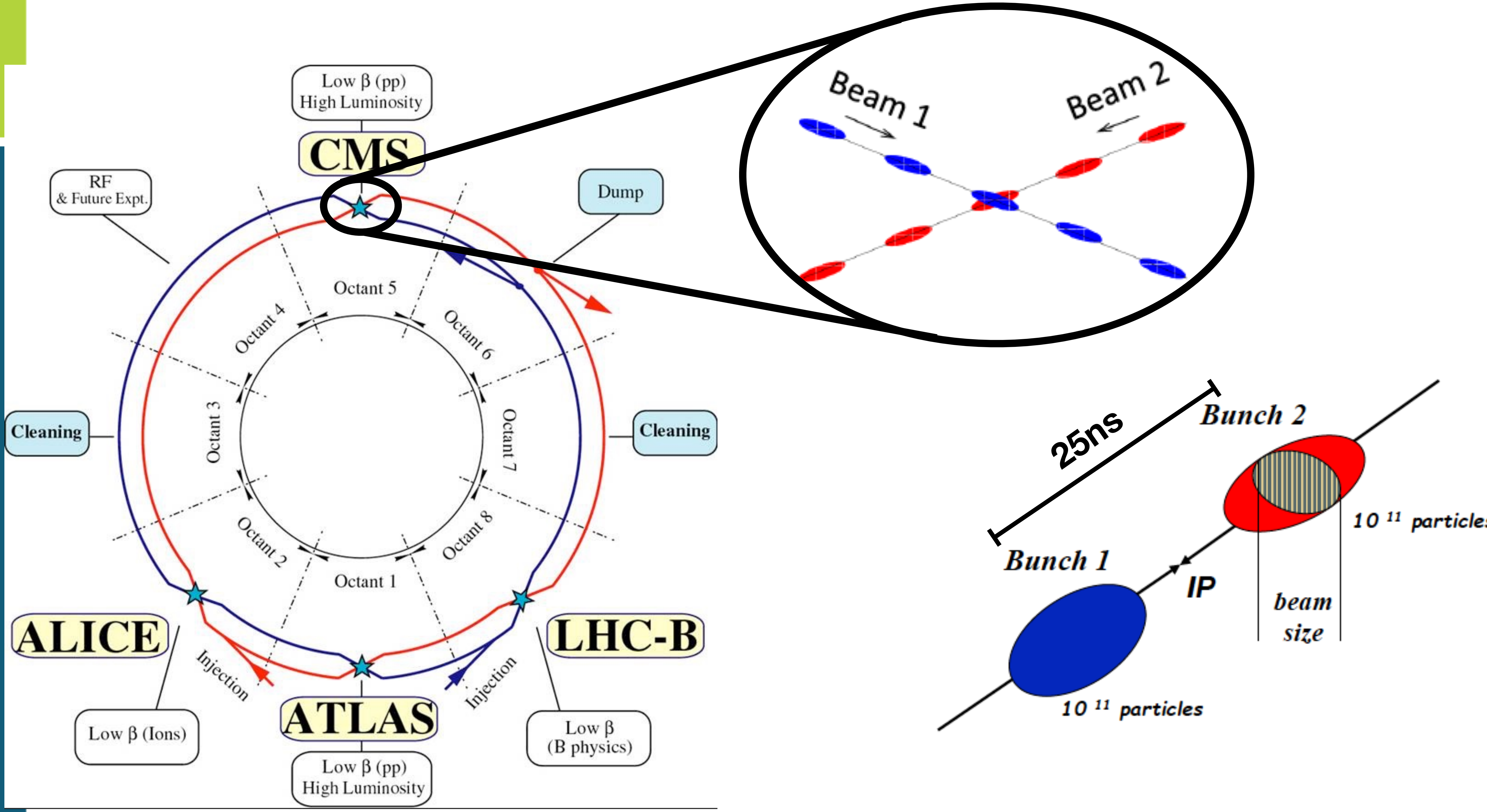
1. **Reconstruct physics signals** from the data ✔
   - Produce information like how many muons does the event have?
2. **Calibrate** the detectors ✔
   - Correct imperfections, account for changes over time…
3. Make sure that the **data quality** is excellent, also in real time
   - Maximise the amount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

At the LHC: more than one proton collision - more than one vertex

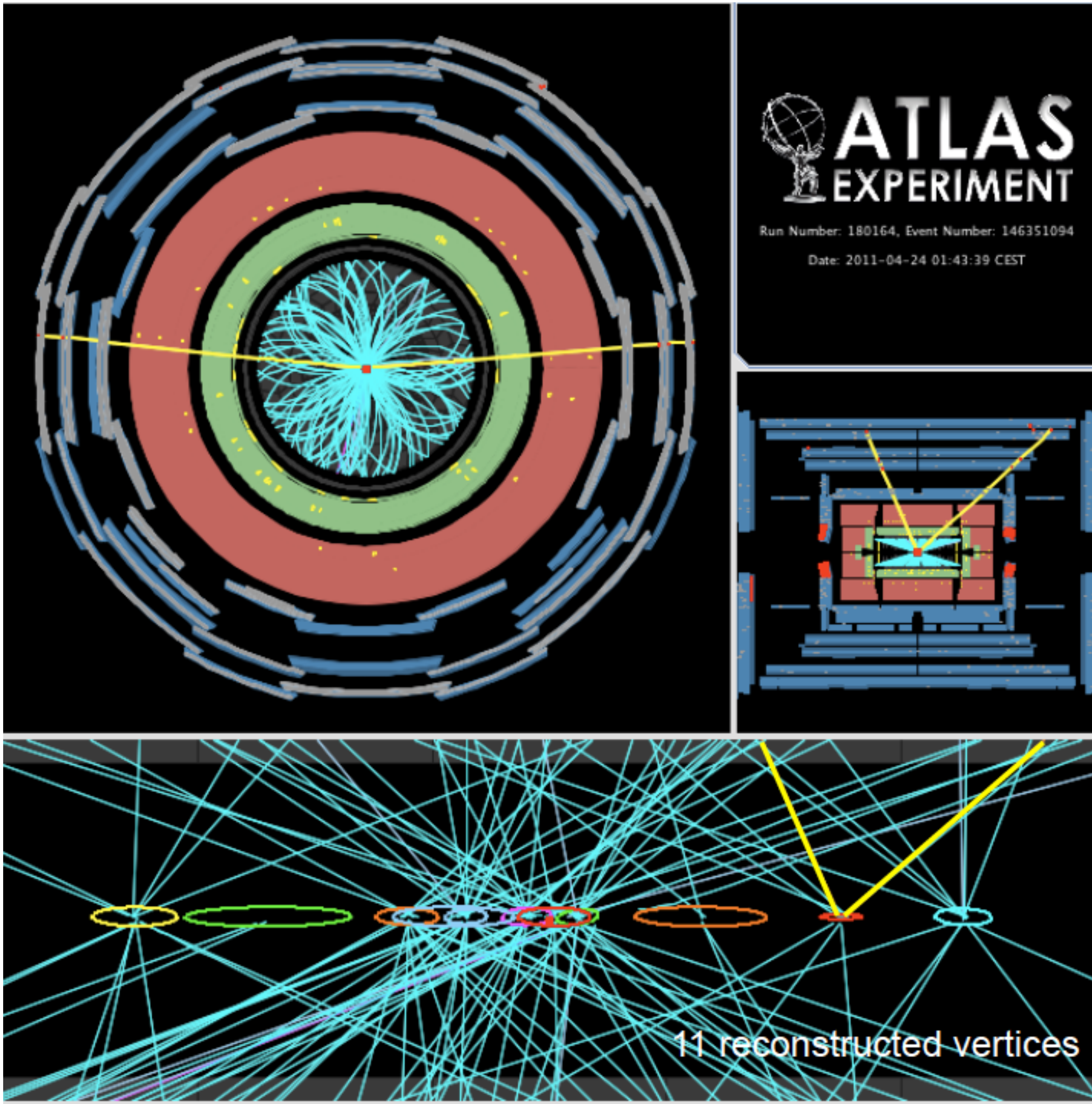


LHC collisions

Figures adapted from Michaela Schaumann's third lecture (11/07/19) on "Particle Accelerators and Beam Dynamics"



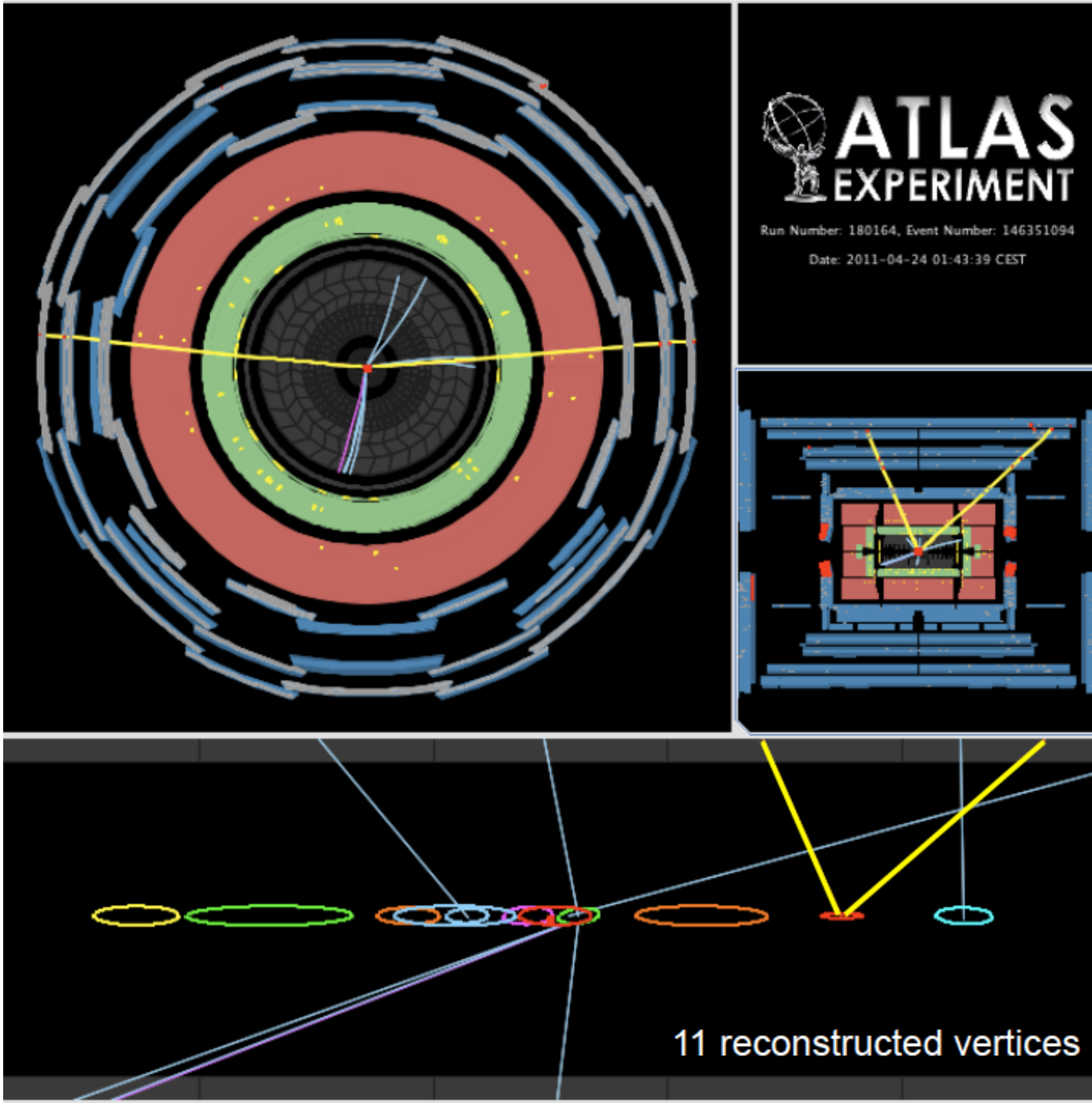
- The LHC accelerates ***bunches of 10^{11} protons*** separated by 25ns gaps



Z->μμ event;
2011 data.

The more bunches are squeezed, the higher the luminosity, the larger the number of simultaneous proton collisions in one recorded event

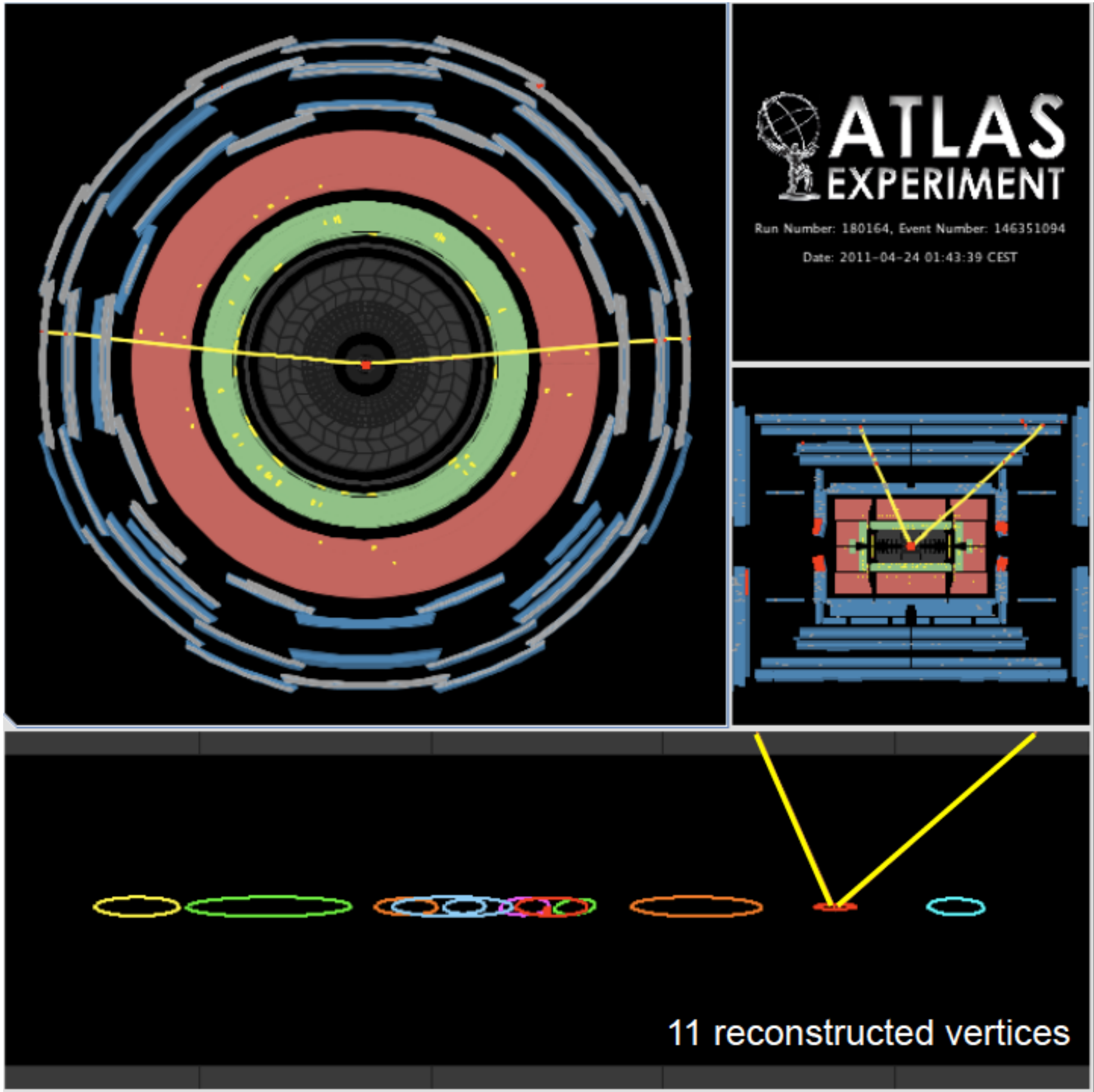
Track $p_T > 0.5$ GeV



Z->μμ event;
2011 data.

Most proton collisions are low momentum and uninteresting. We can remove them simply by making a cut on the transverse momentum.

Track pT > 2 GeV

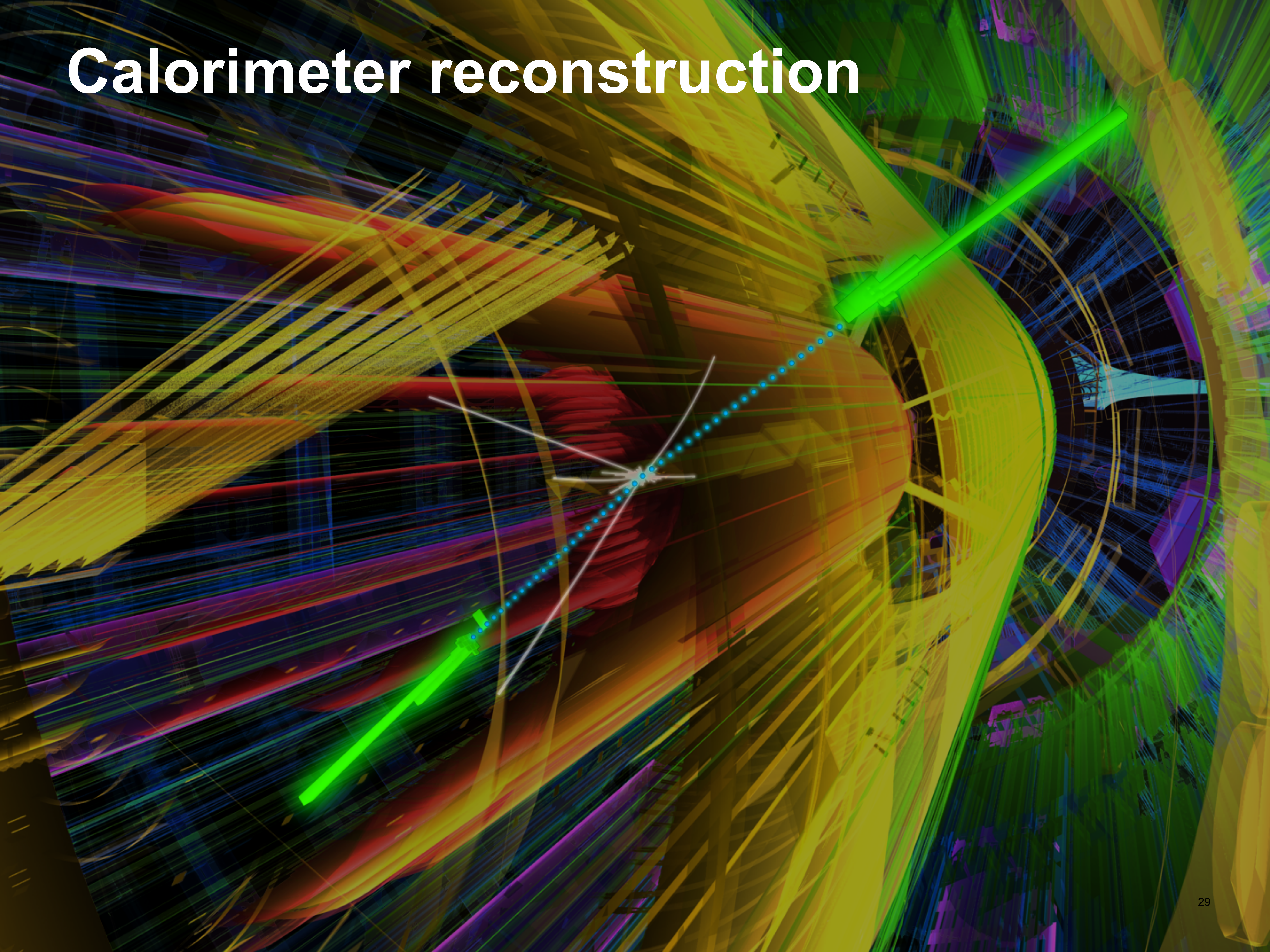


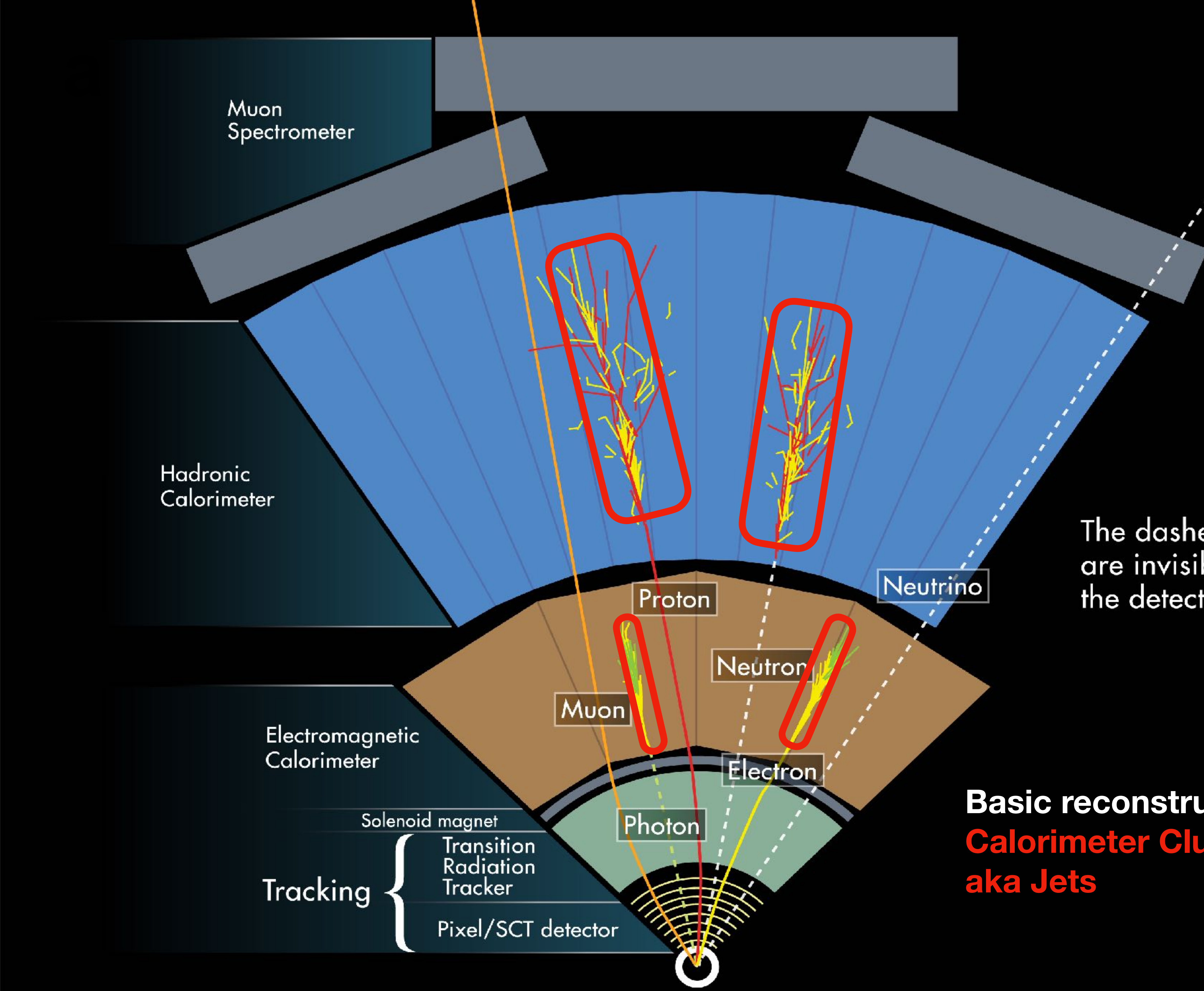
Z->μμ event;
2011 data.

Once we increase the transverse momentum cut sufficiently, we are left with only the interesting proton collision.

Track pT > 10 GeV

Calorimeter reconstruction

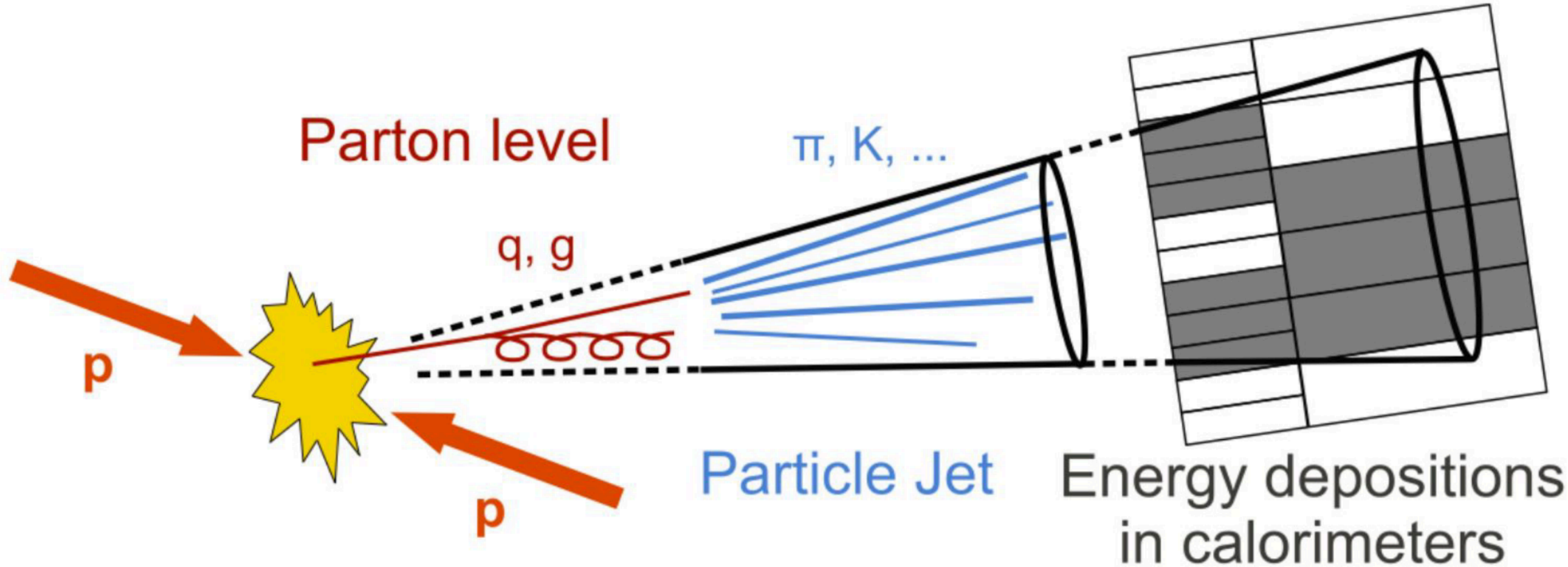




The dashed tracks are invisible to the detector

Basic reconstruction =
Calorimeter Clusters, aka Jets

Jet reconstruction



Quarks and gluons ***hadronize*** quickly and we detect ***sprays of hadronic particles*** in our detectors - we call these ***jets***

Jets are used as proxies for the initial particle(s), we reconstruct them using ***jet algorithms***

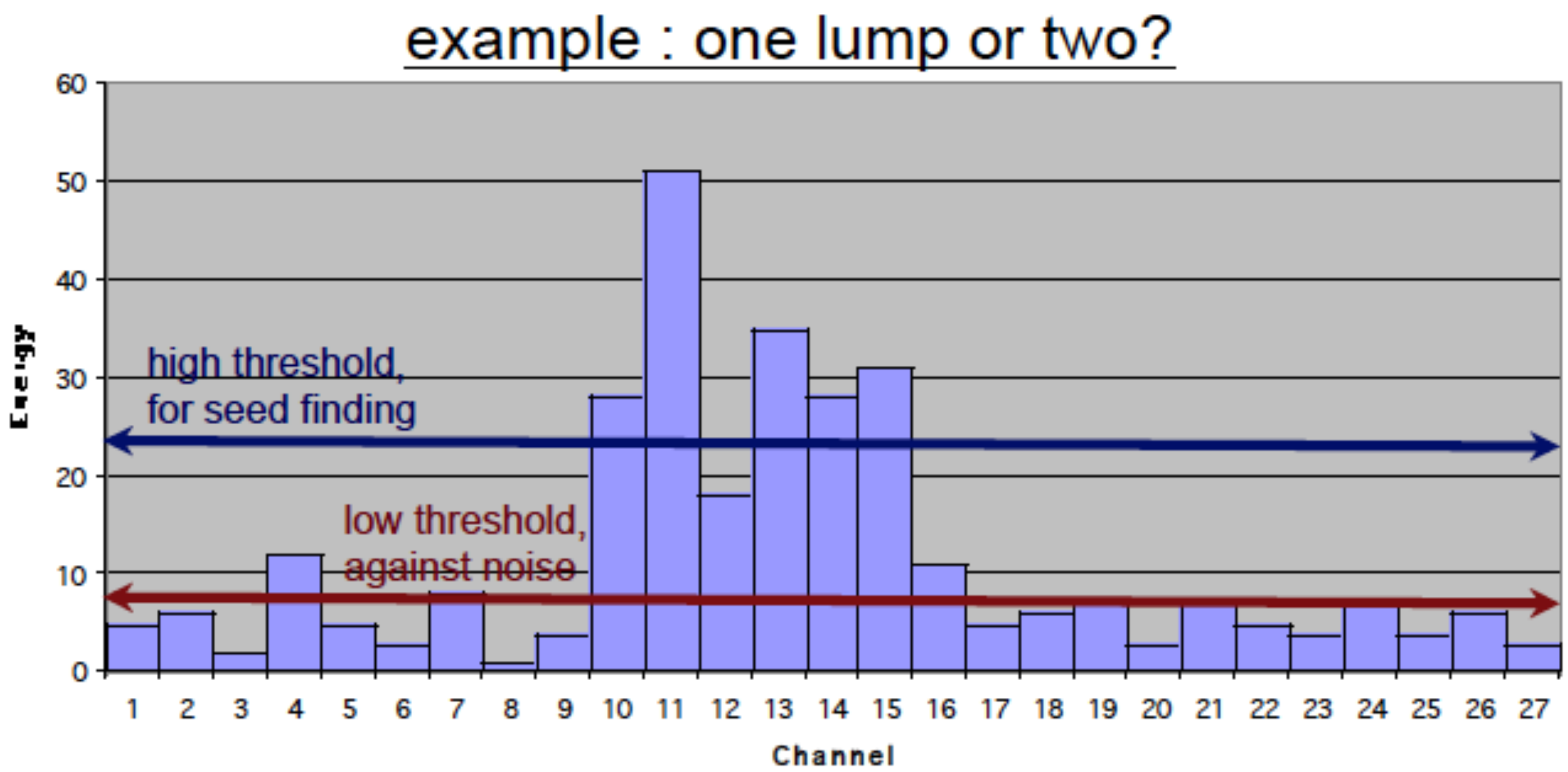
Jet algorithms are effectively ***clustering algorithms*** with special properties (*infra-red safe*)

- Perfecting these algorithms is an active field of research for several decades
- The original idea has been extended to e.g. reconstruction of top quarks

We run ***several*** jet algorithms in reconstruction and then choose the best one depending on the ***physics case***

Modern jet reconstruction uses Machine Learning!

Cluster-finding in practice



Algorithms and ***thresholds*** need careful study and tuning, especially early on in the experiment's lifetime

- Particularly need to study the sensitivity to noise (depends on the detector and the environment)

If ***thresholds*** are ***too low*** - pick up ***noise***

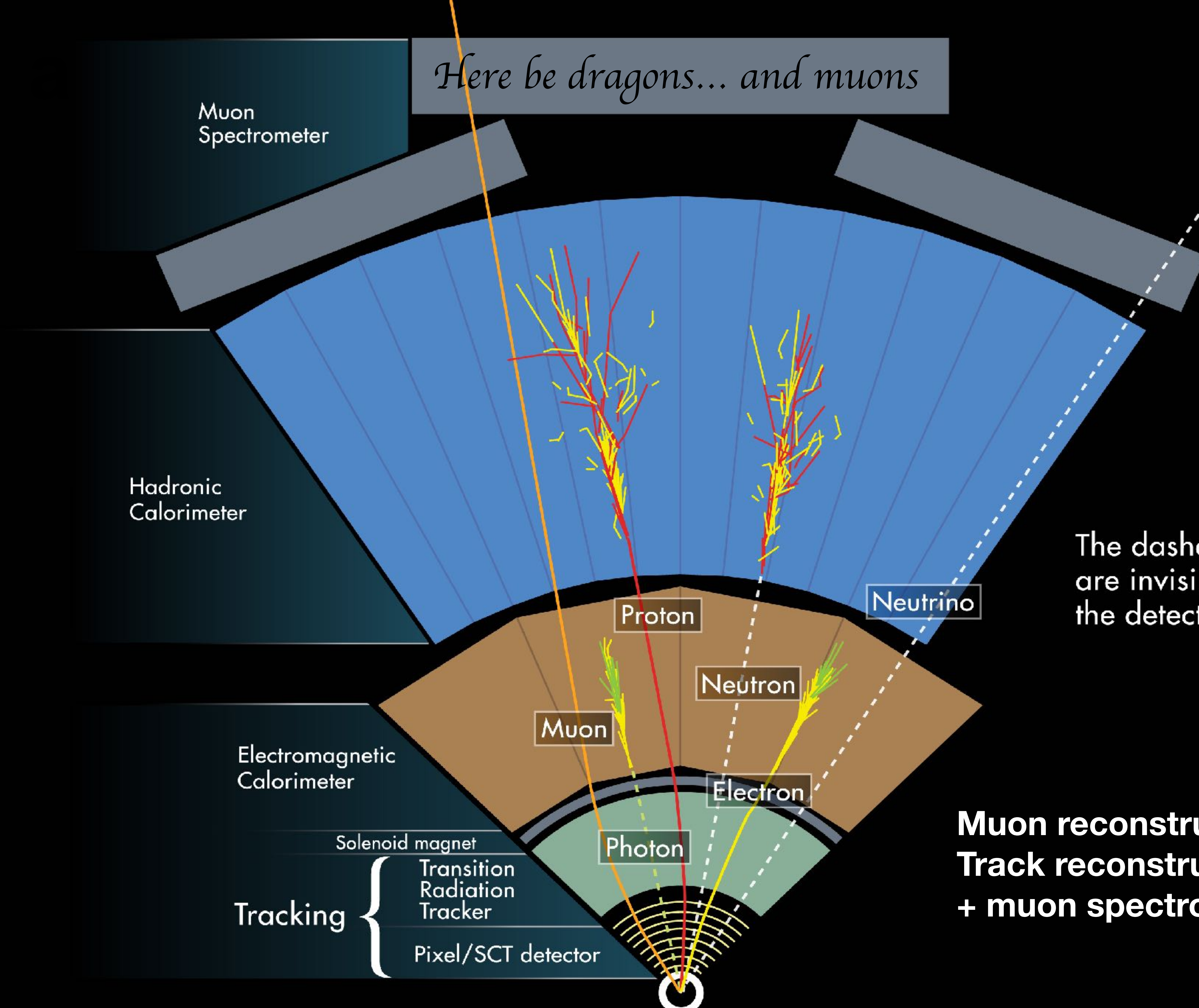
If ***thresholds*** are ***too high*** - ***lose too much of the energy actually deposited by the particle***

- The loss of some of the energy needs to be corrected for (calibrated) and ***large calibration factors*** lead to ***large measurement uncertainties***

Calorimeter calibration takes a lot of work!

Brookhaven National Laboratory

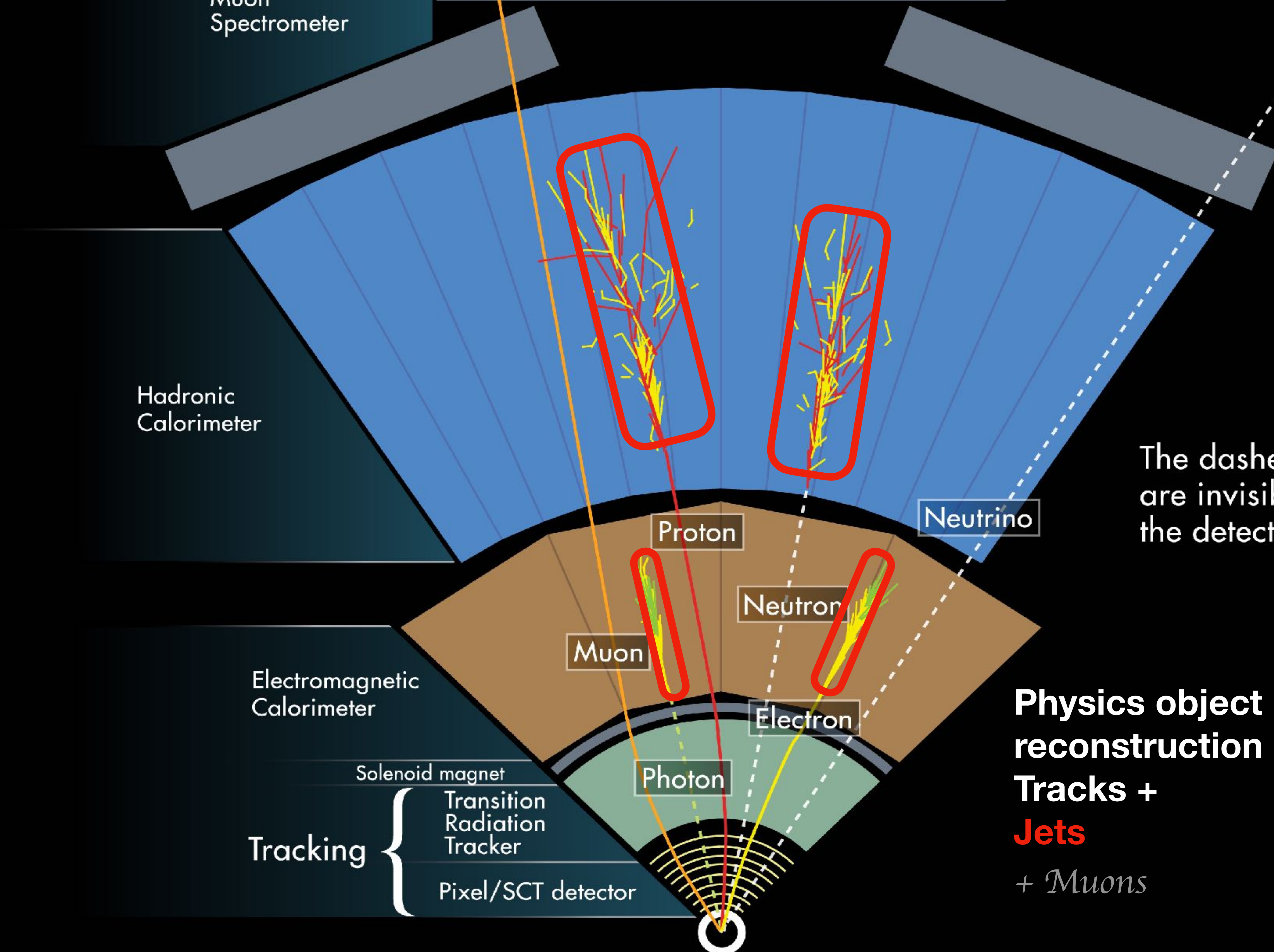
Here be dragons... and muons



The dashed tracks are invisible to the detector

Muon reconstruction = Track reconstruction + muon spectrometer hits

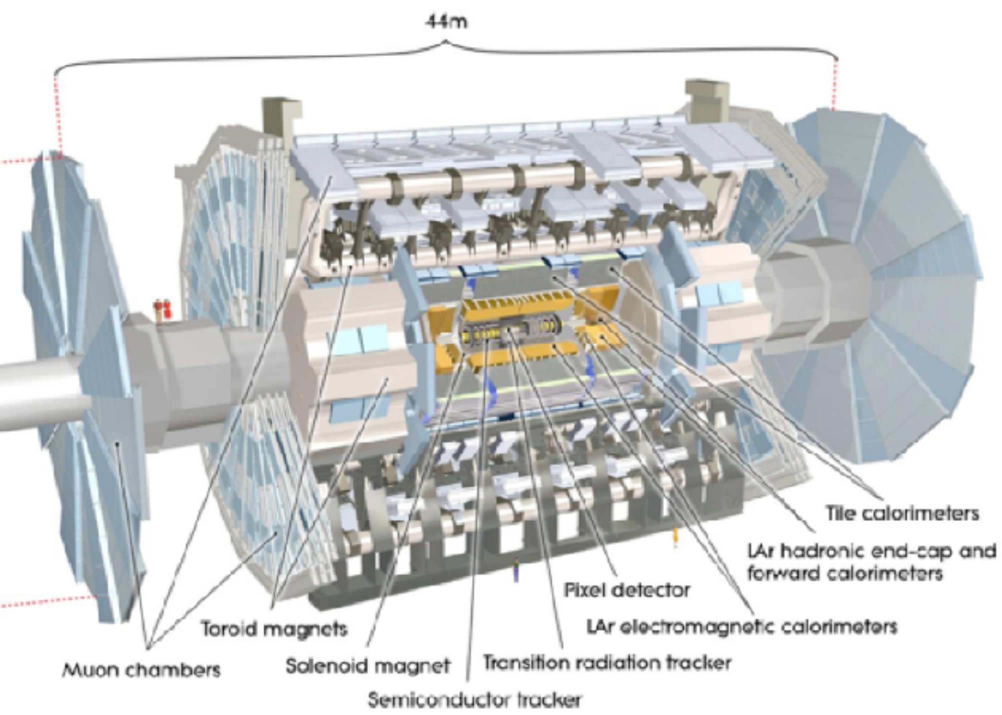
Here be dragons… and muons

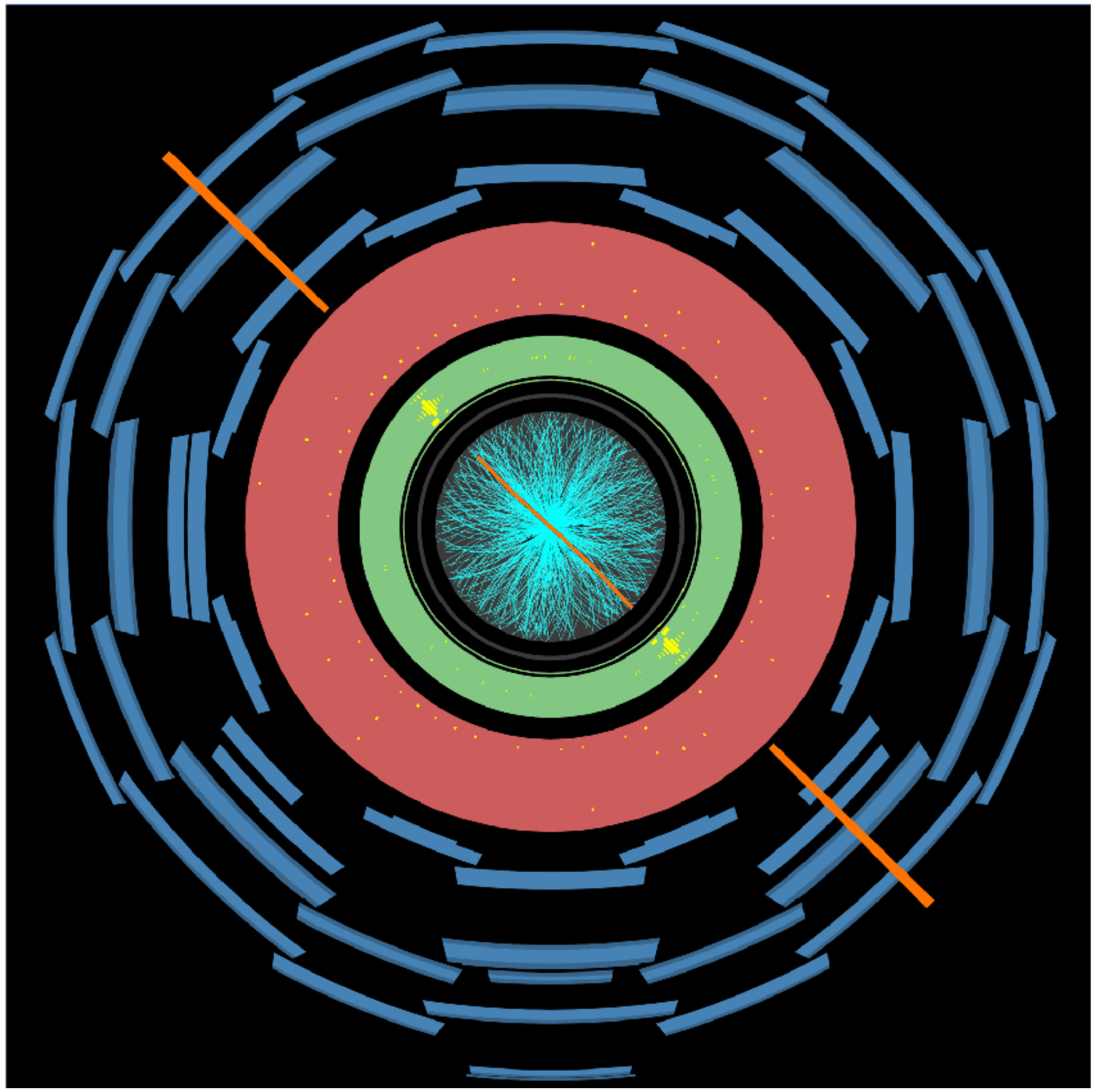


The dashed tracks are invisible to the detector

Physics object reconstruction = Tracks +
Jets
+ Muons

Neutrinos





Let's look at the simplest case for reconstructing neutrinos

Remember, we are looking down the beam pipe, so the plane of the display is transverse to the proton beam direction

***Recall**: 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?*

***Q**. How would this look if we had a **W boson** instead of a **Z boson** ?*

Brookhaven National Laboratory

Today - Machine Learning

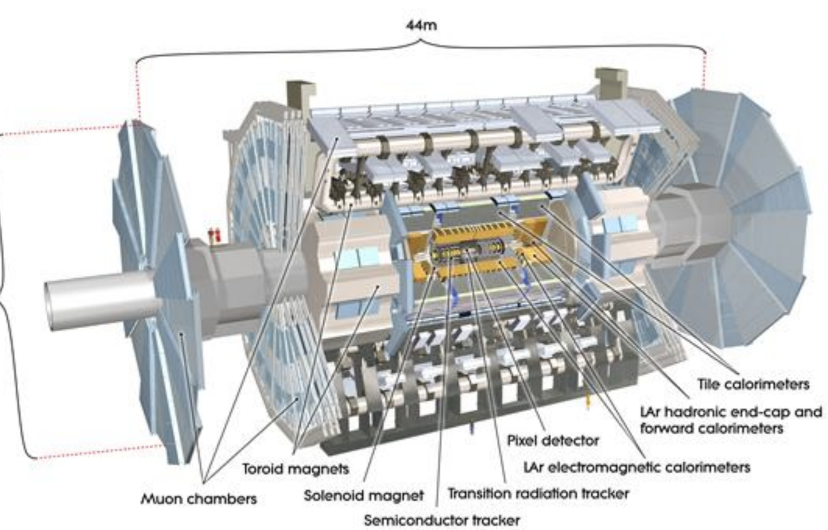
Modern simulation, reconstruction and analysis employ heavy use of Machine Learning techniques. See Foundations of Statistics for an introduction to the key concepts. There are also some excellent resources online, e.g.:

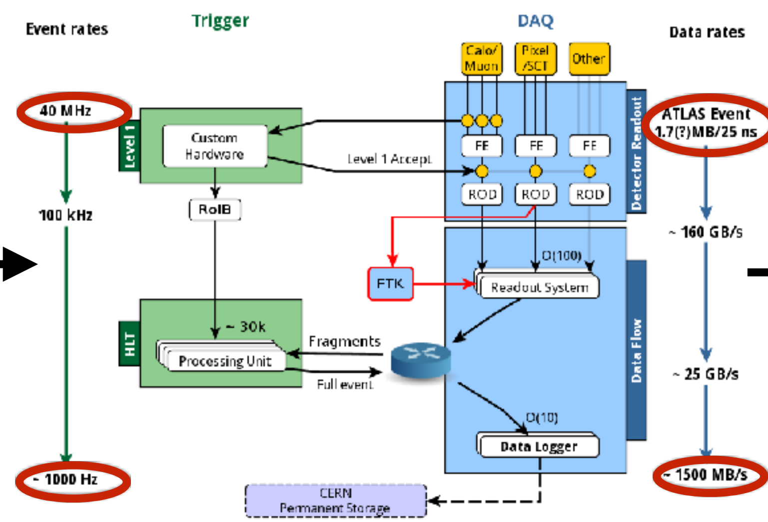
Google Machine Learning Crash Course

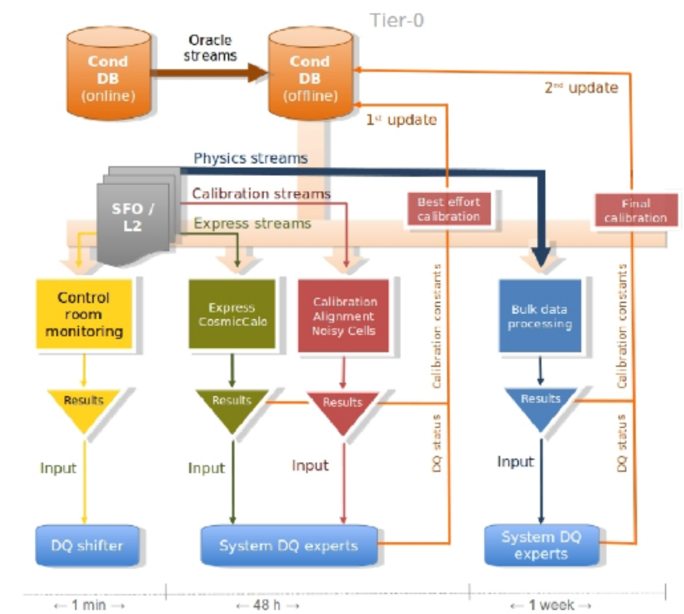
Machine Learning and AI for Scientists

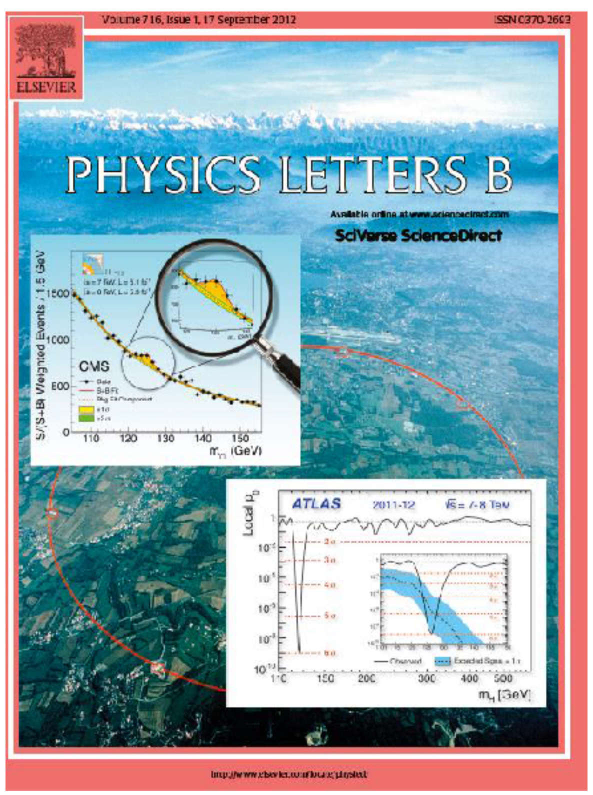
Track Reconstruction @ Connect the Dots

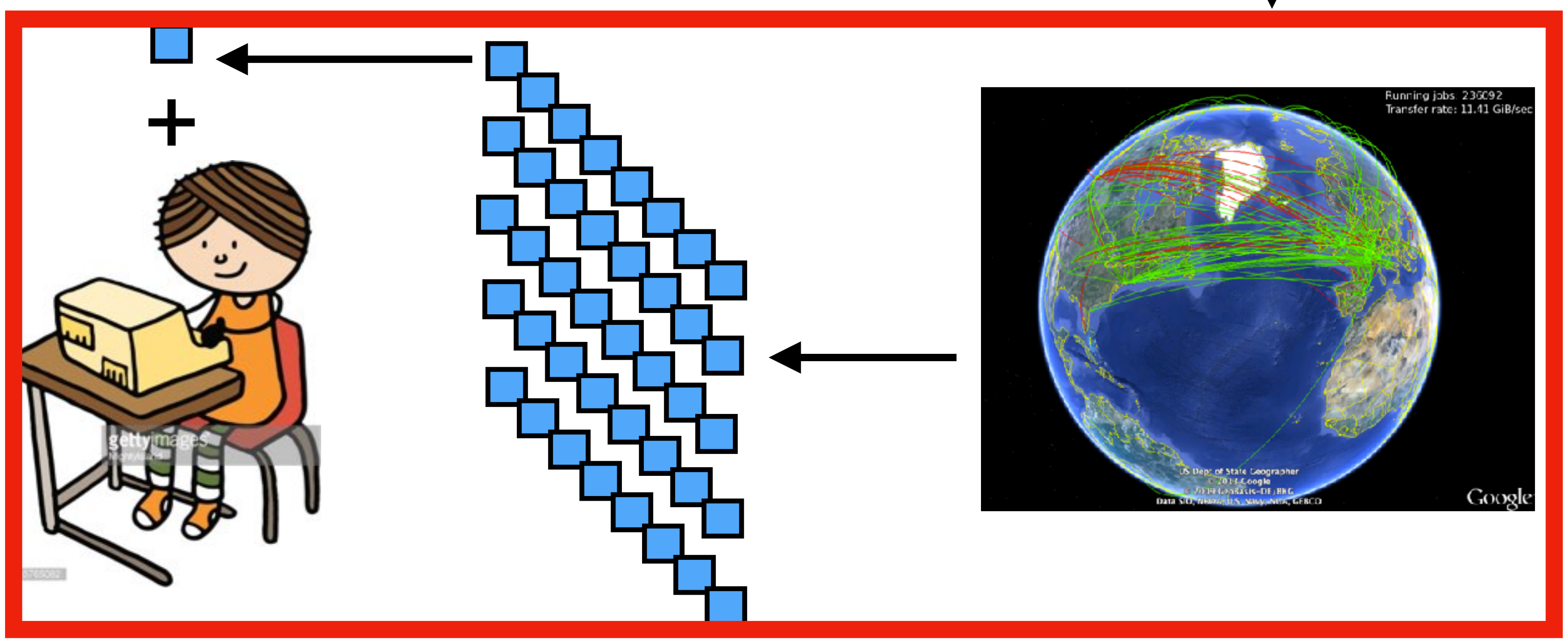
Data’s journey - next time, analysis!











Contact details

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

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

In-person Q&A:
Monday 13:30-15:00 in Salle Anderson, B40 S2-A01

Brookhaven National Laboratory

Q&A time