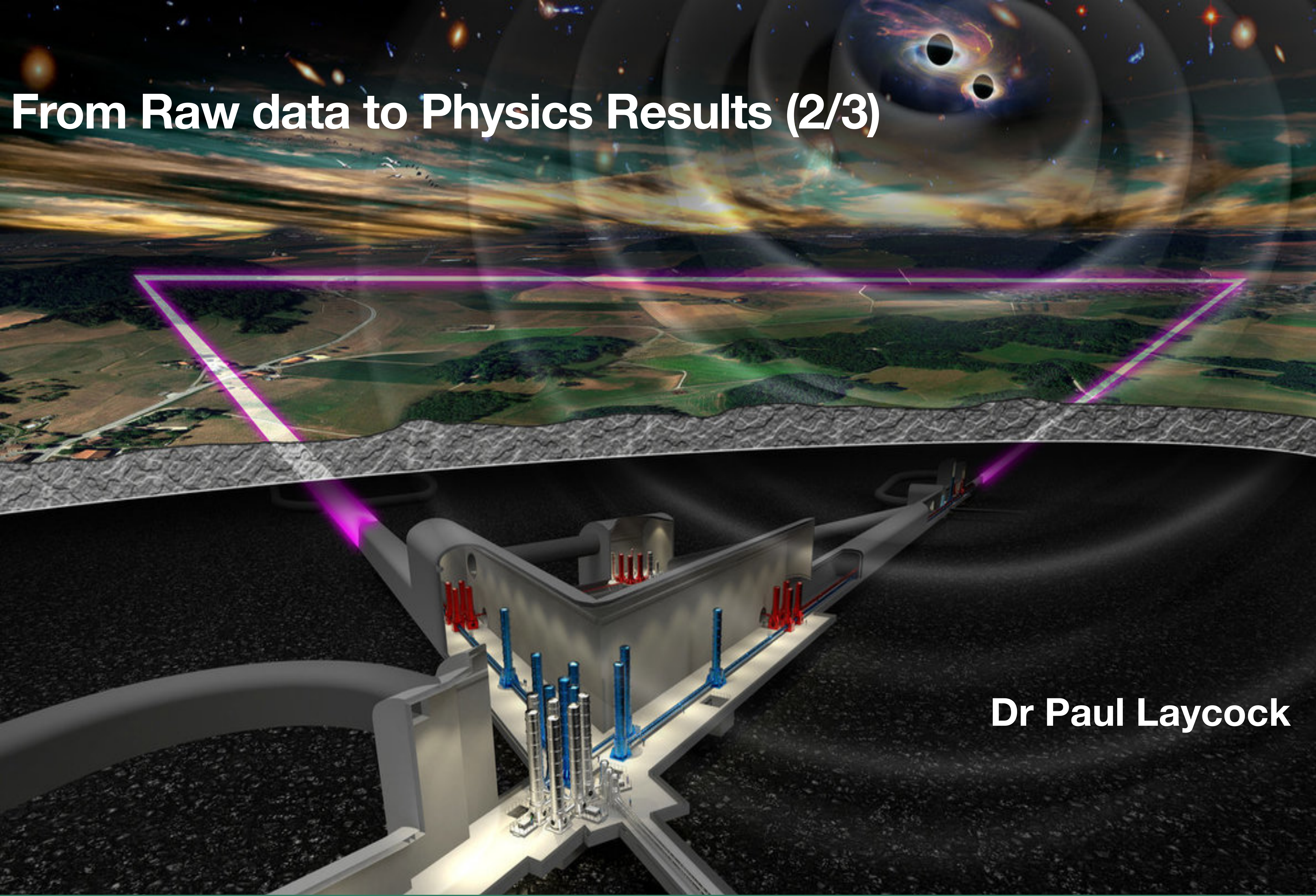
From Raw data to Physics Results (2/3)

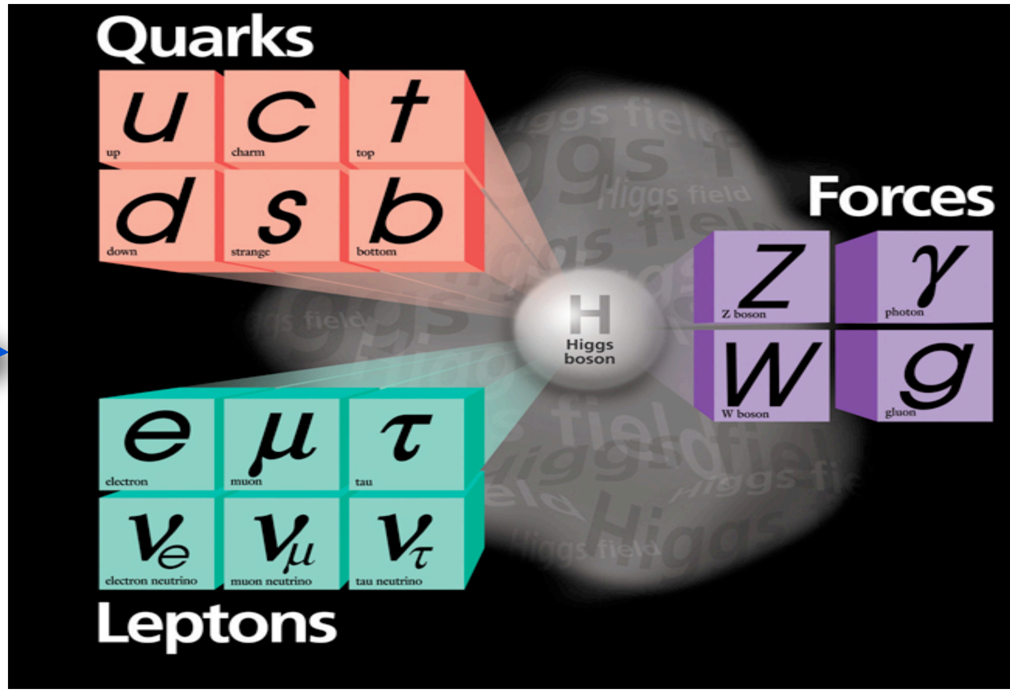


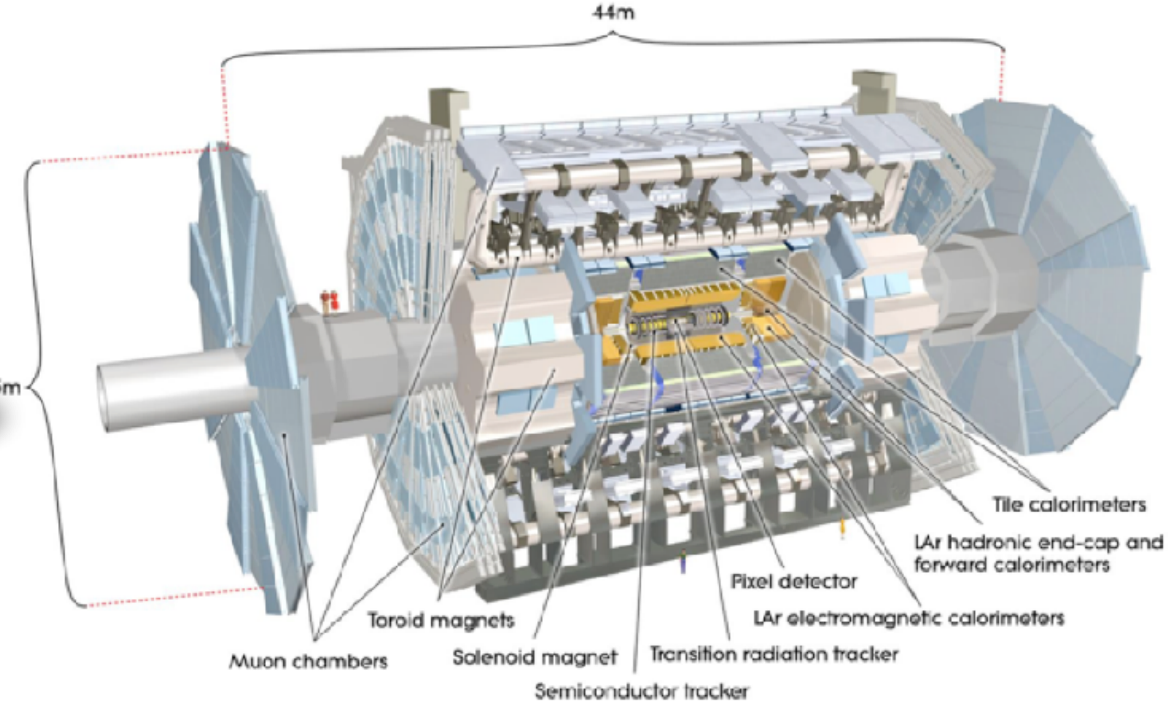
Dr Paul Laycock

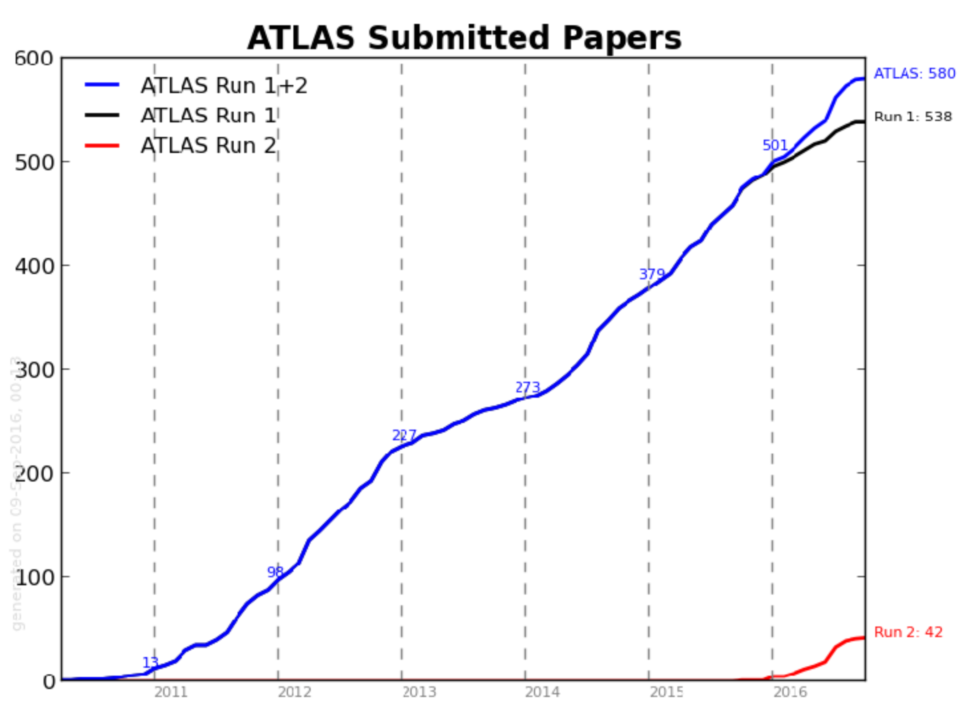
UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES
Département d'astronomie

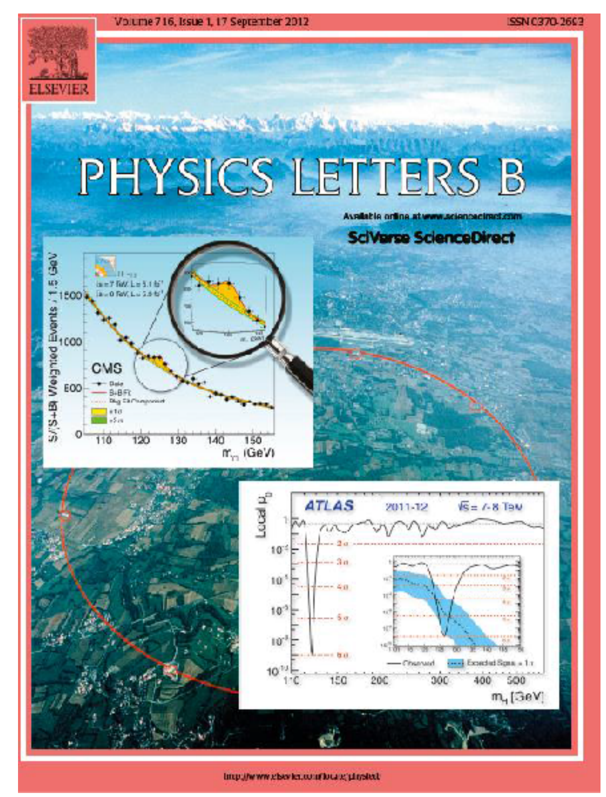
GWSC GRAVITATIONAL WAVE SCIENCE CENTER

The particle physics cycle



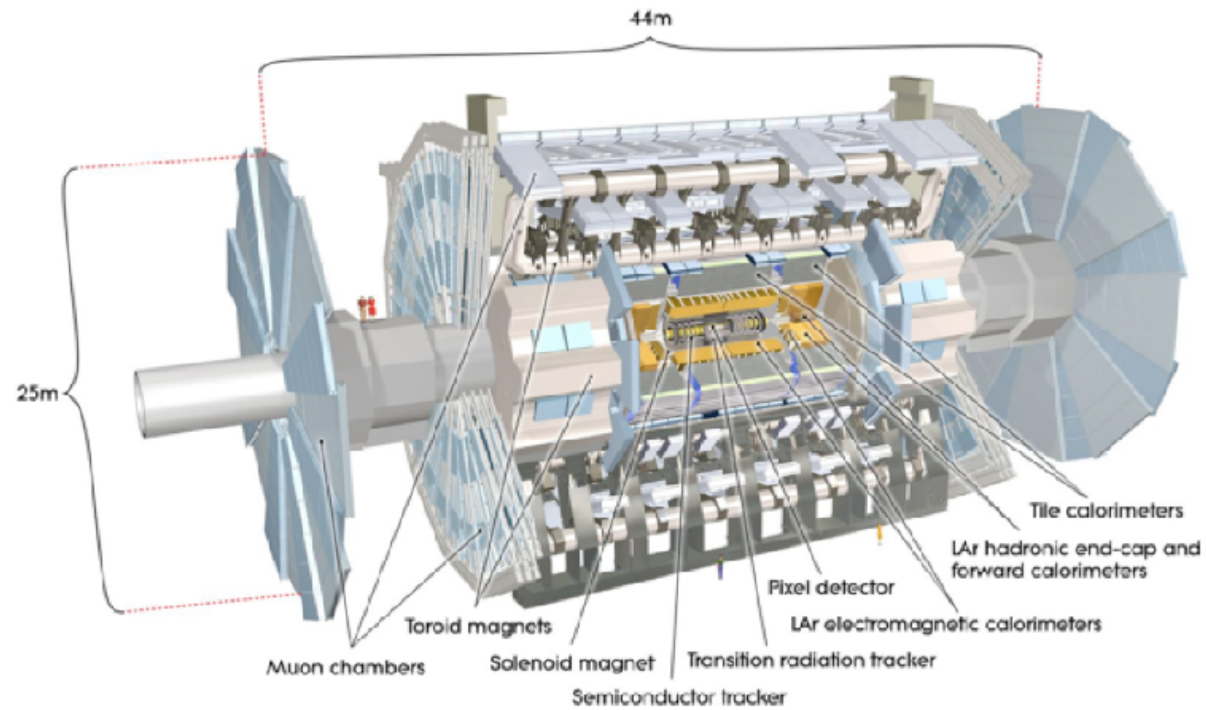


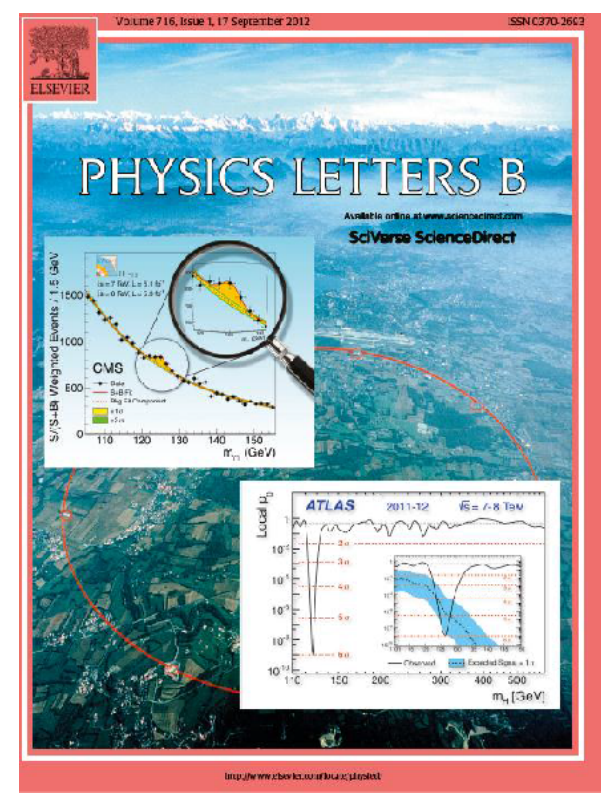




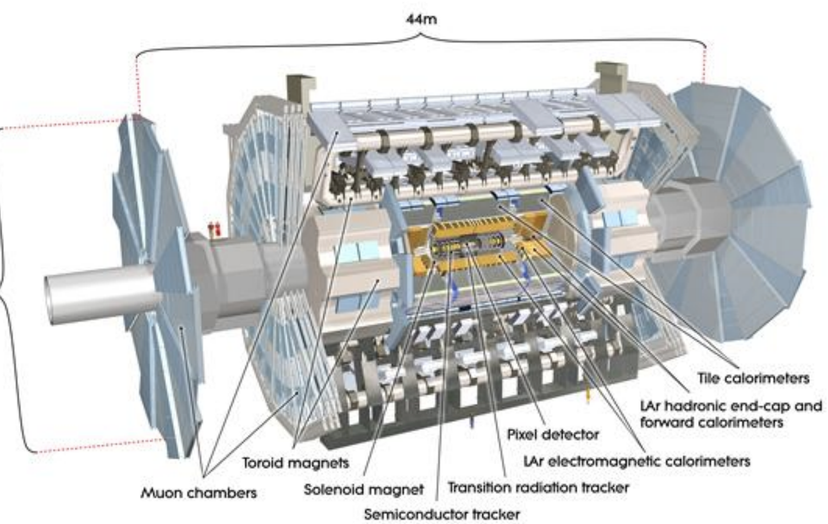
Course outline

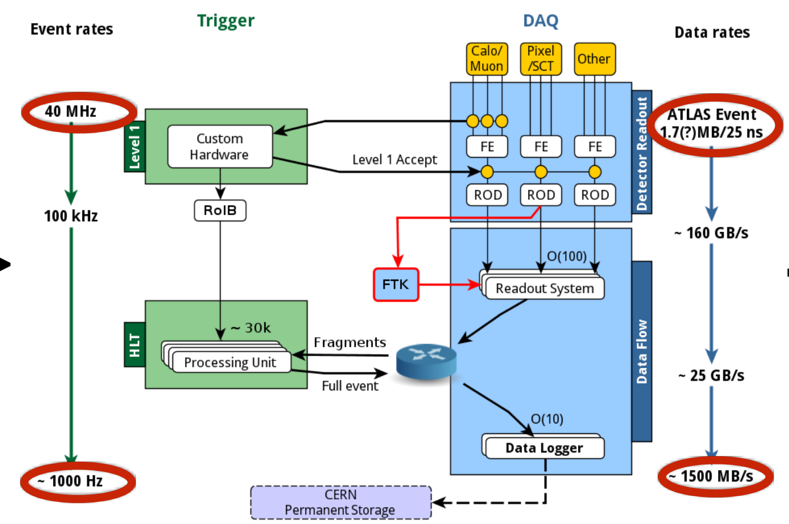
- **Lecture 1**
  - The journey of raw data from the detector to a publication
- **Lecture 2**
  - How we reconstruct fundamental physics processes from raw detector data
- **Lecture 3**
  - How we extract our signals from the mountain of data, finding needles in the haystack





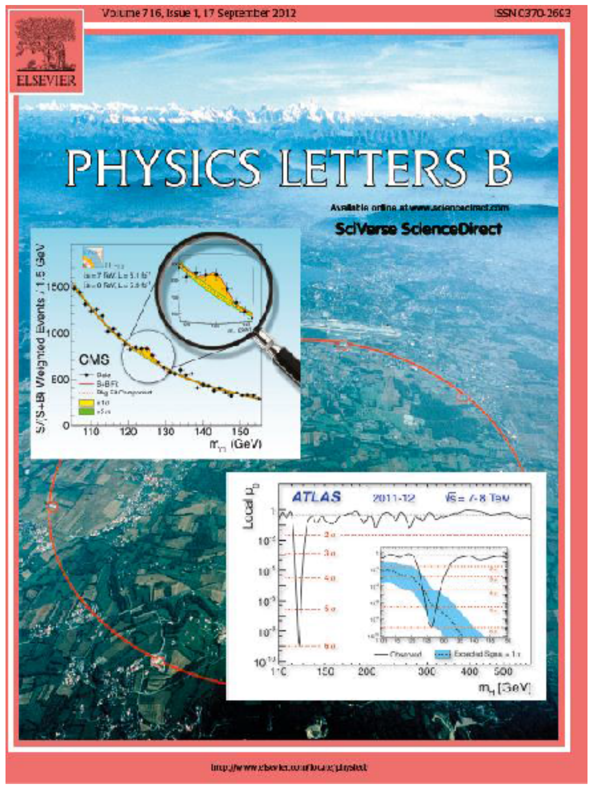
Data's journey

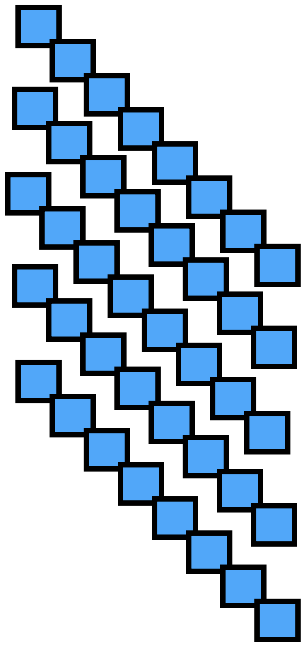




Data Preparation

Distributed computing



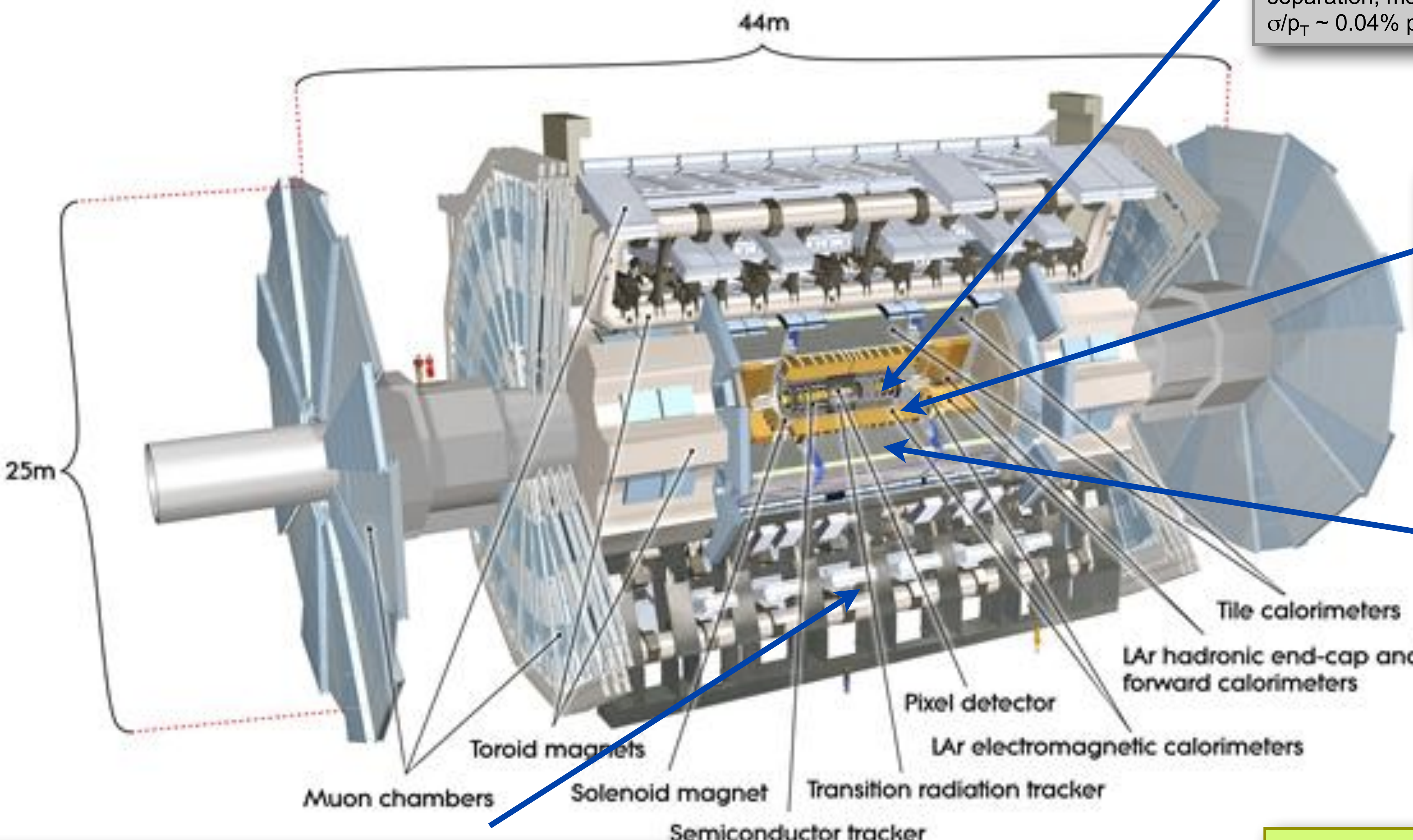


The ATLAS Detector @ LHC

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:
$\sigma/p_T \sim 0.04\%\ p_T\,(\mathrm{GeV}) \oplus 1.5\%$

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

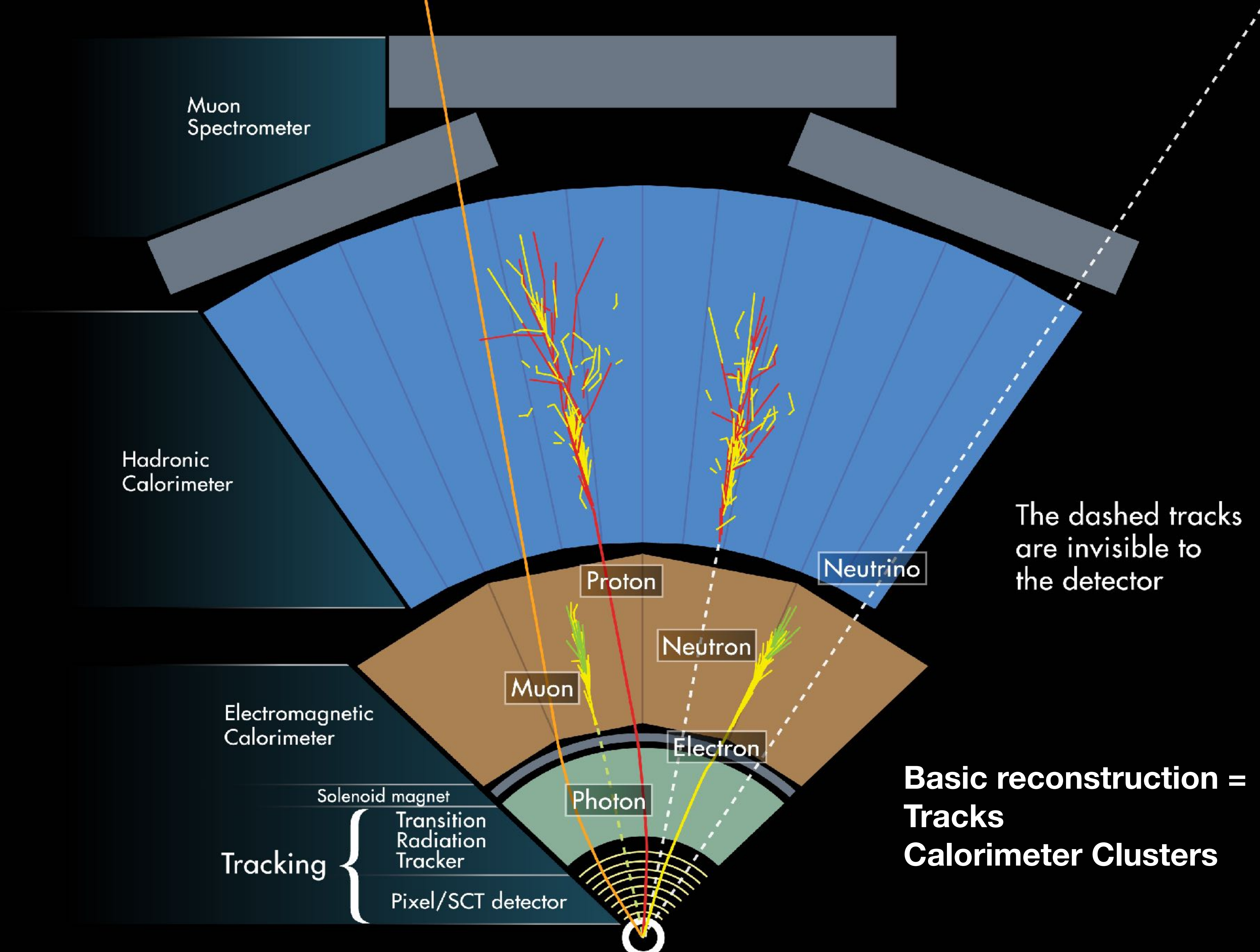
Trigger system: 3-levels reducing the IA rate from 40 MHz to ~200 Hz

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

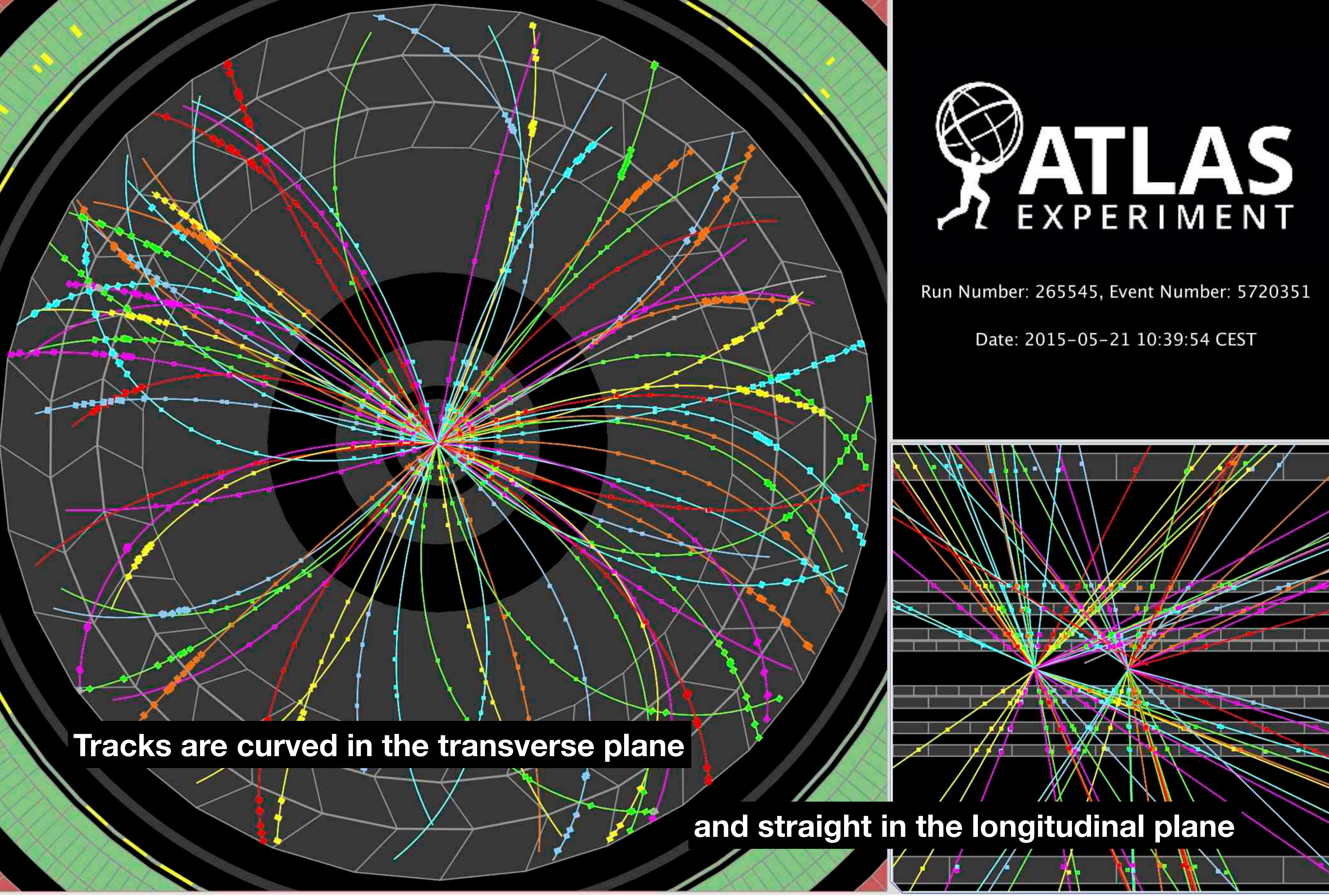
Data Preparation

- 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***

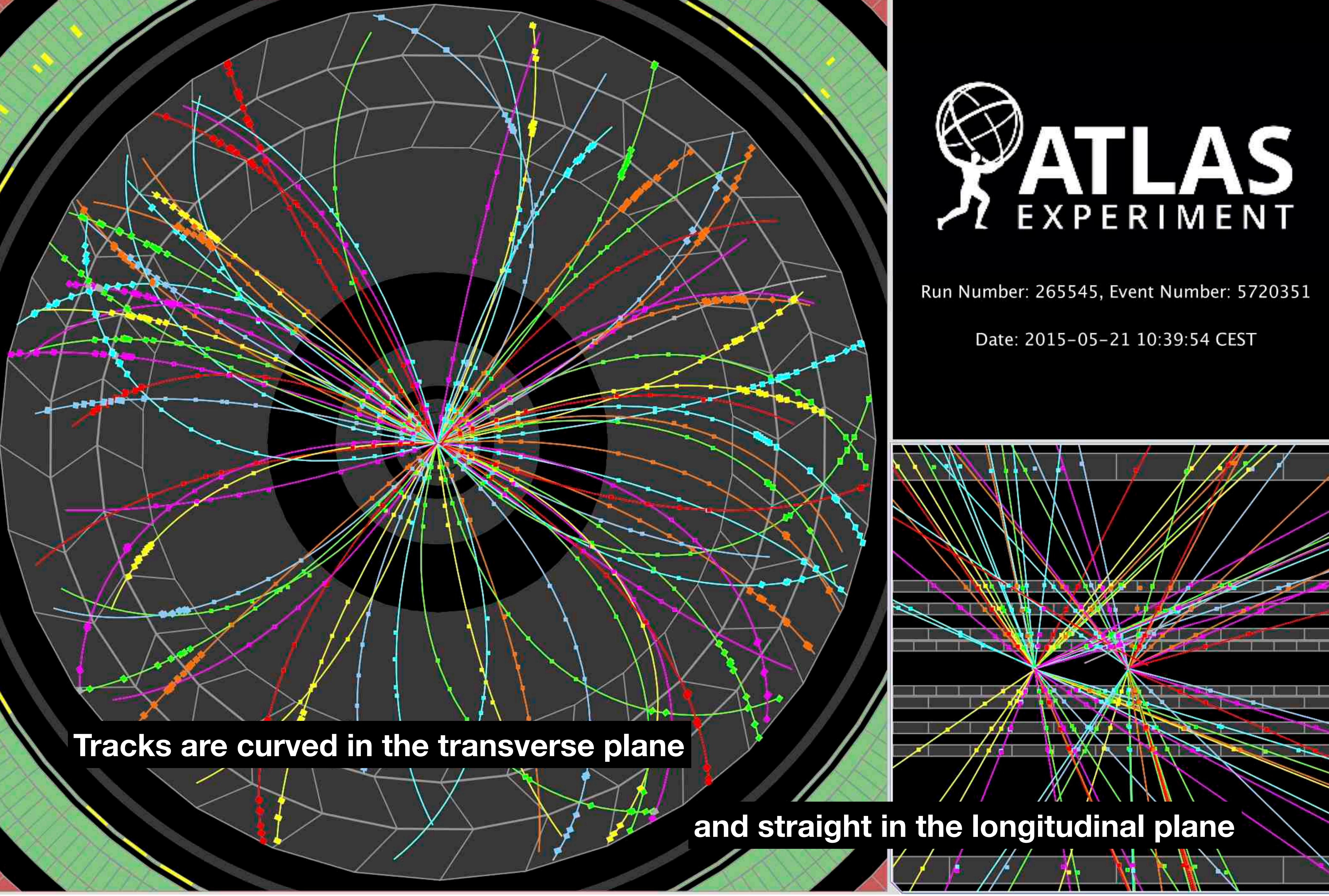
1. **Reconstruct physics signals** from the data
    - Produce information like how many muons does the event have?



Basic reconstruction =
Tracks
Calorimeter Clusters



This is a pattern recognition problem, which technique might be used to solve it?



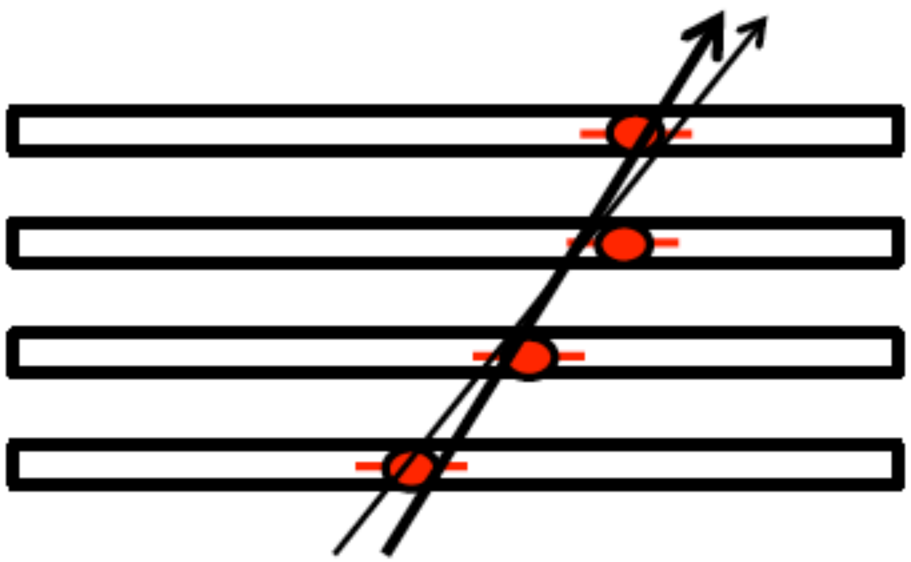
Modern track pattern recognition uses Machine Learning: Connect the Dots

Track fitting

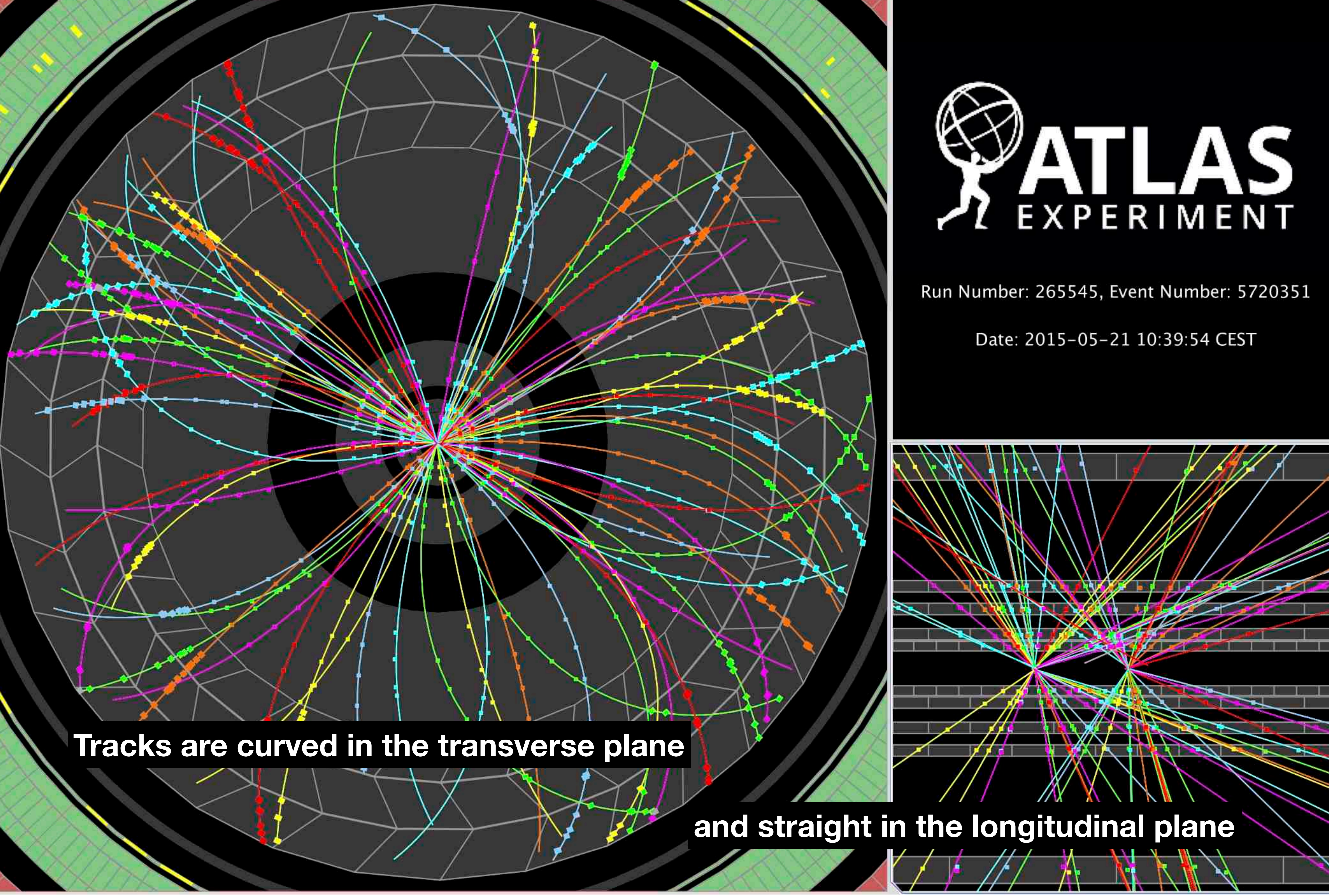
- Perfect measurement – ideal

- Imperfect measurement – reality

- Small errors and more points help to constrain the possibilities

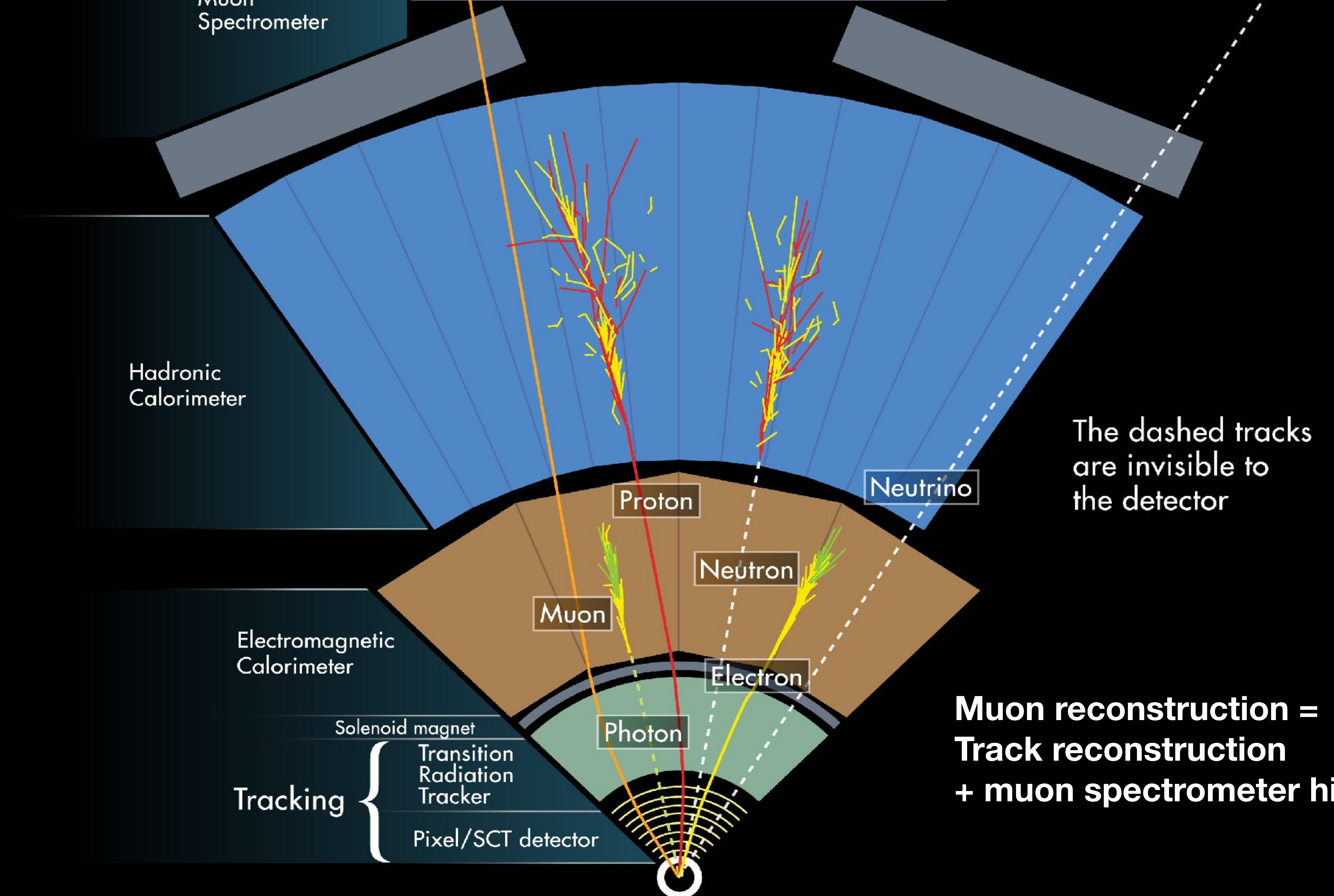


- Quantitatively:
  - Parameterize the track;
  - Find parameters by Least-Squares-Minimization;
  - Obtain also uncertainties on the track parameters.



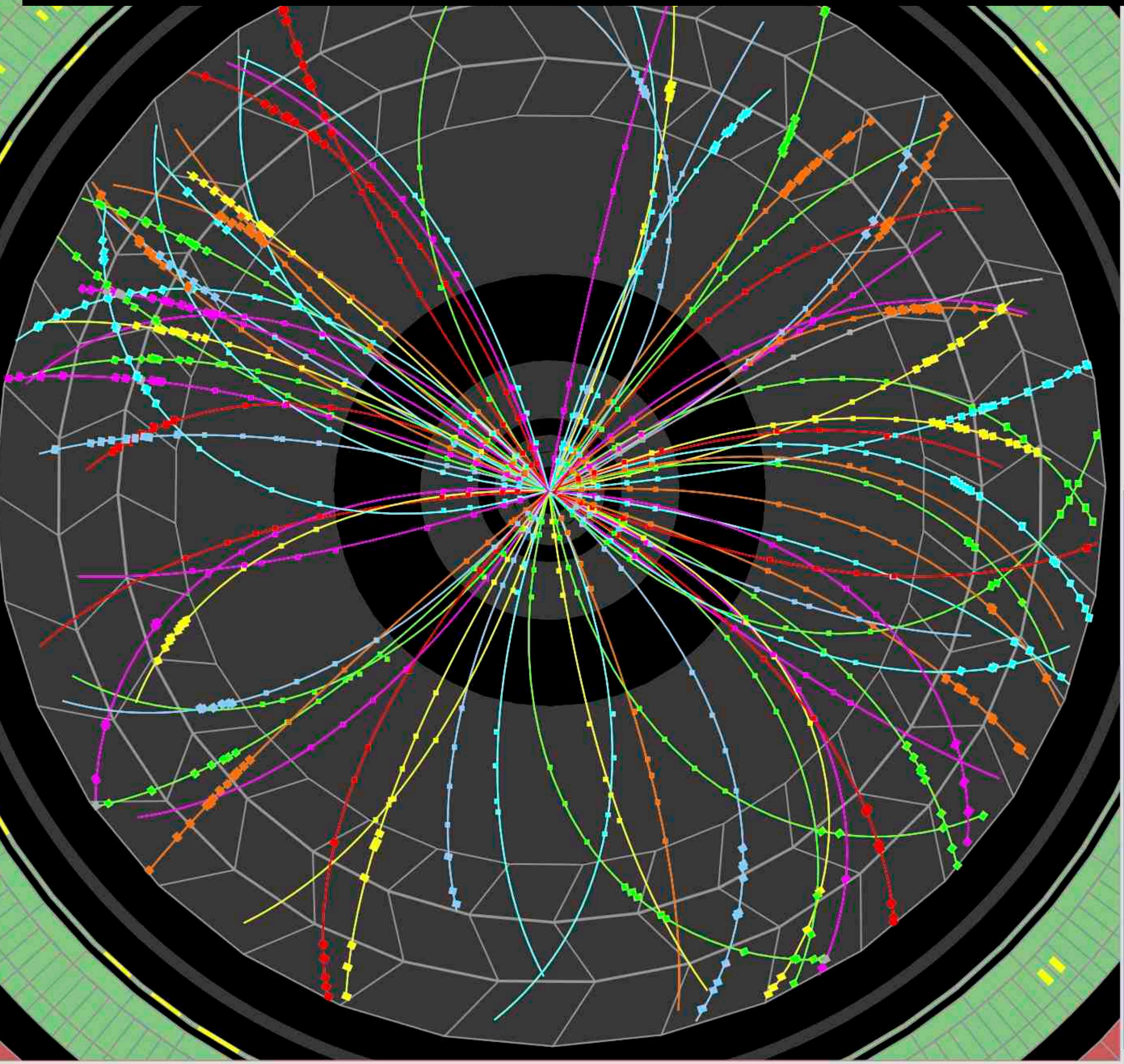
Q. What would be a good track model ?

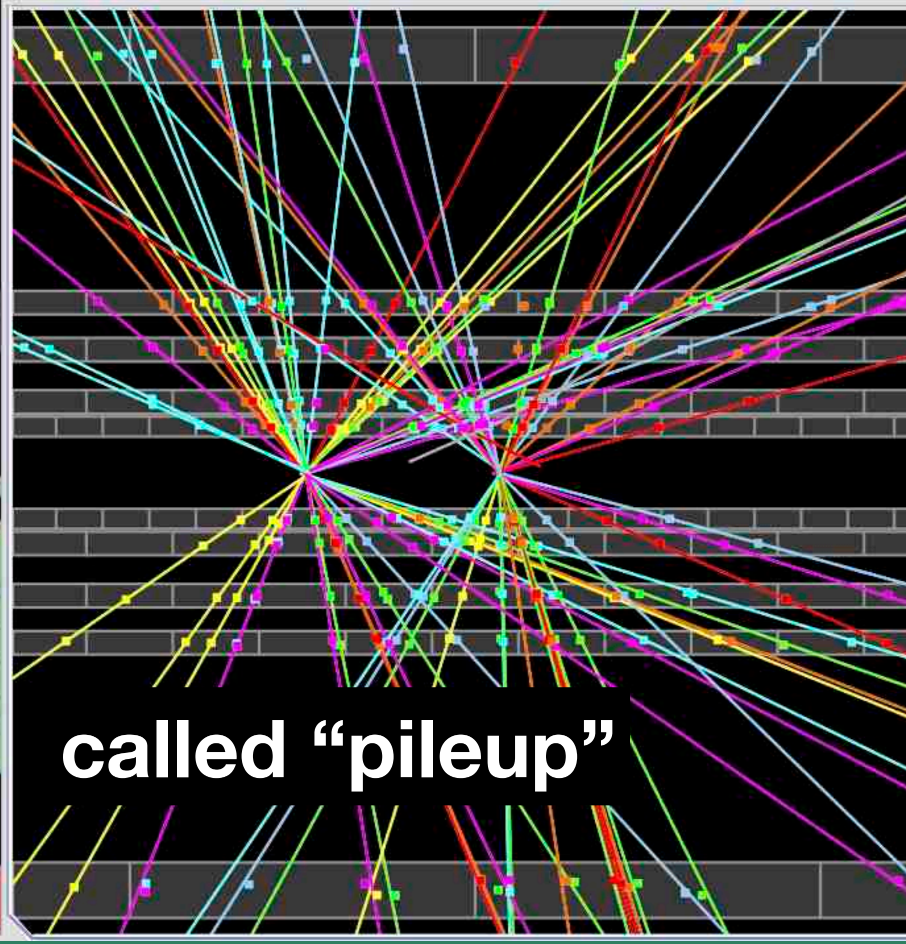
Here be dragons… and muons



Muon reconstruction =
Track reconstruction
+ muon spectrometer hits

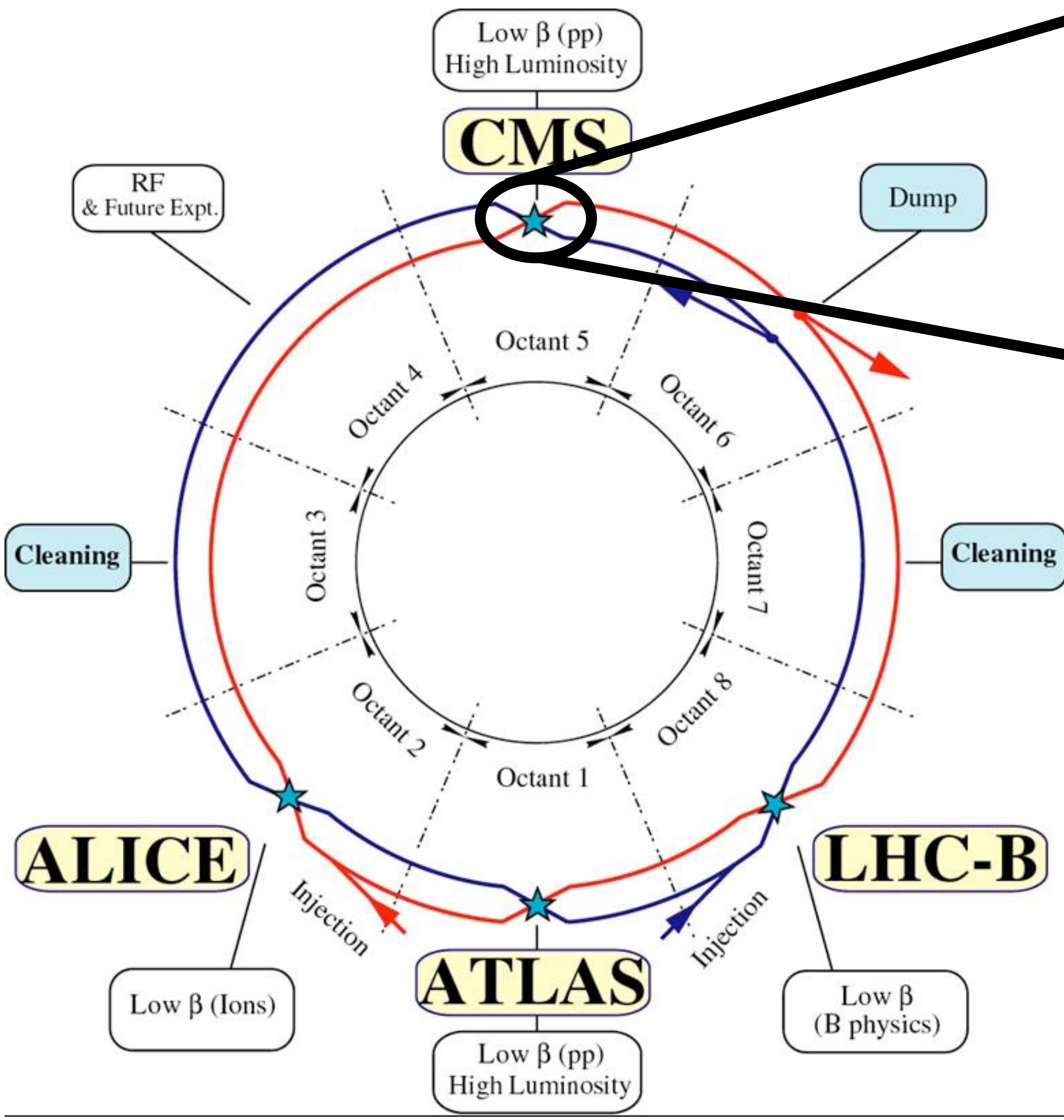
At the LHC: more than one proton collision - more than one vertex

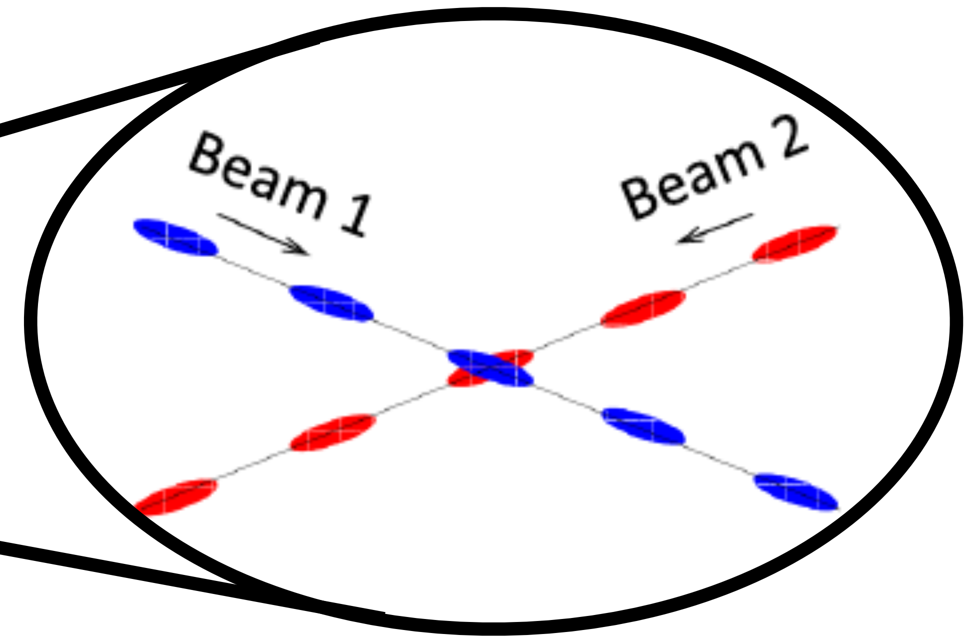


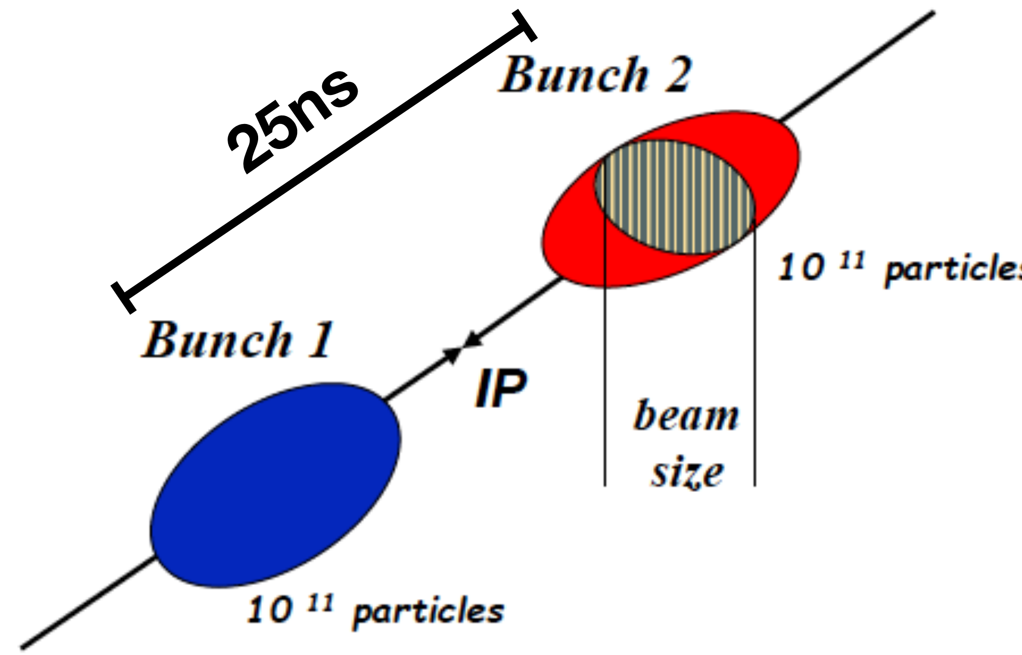


LHC collisions

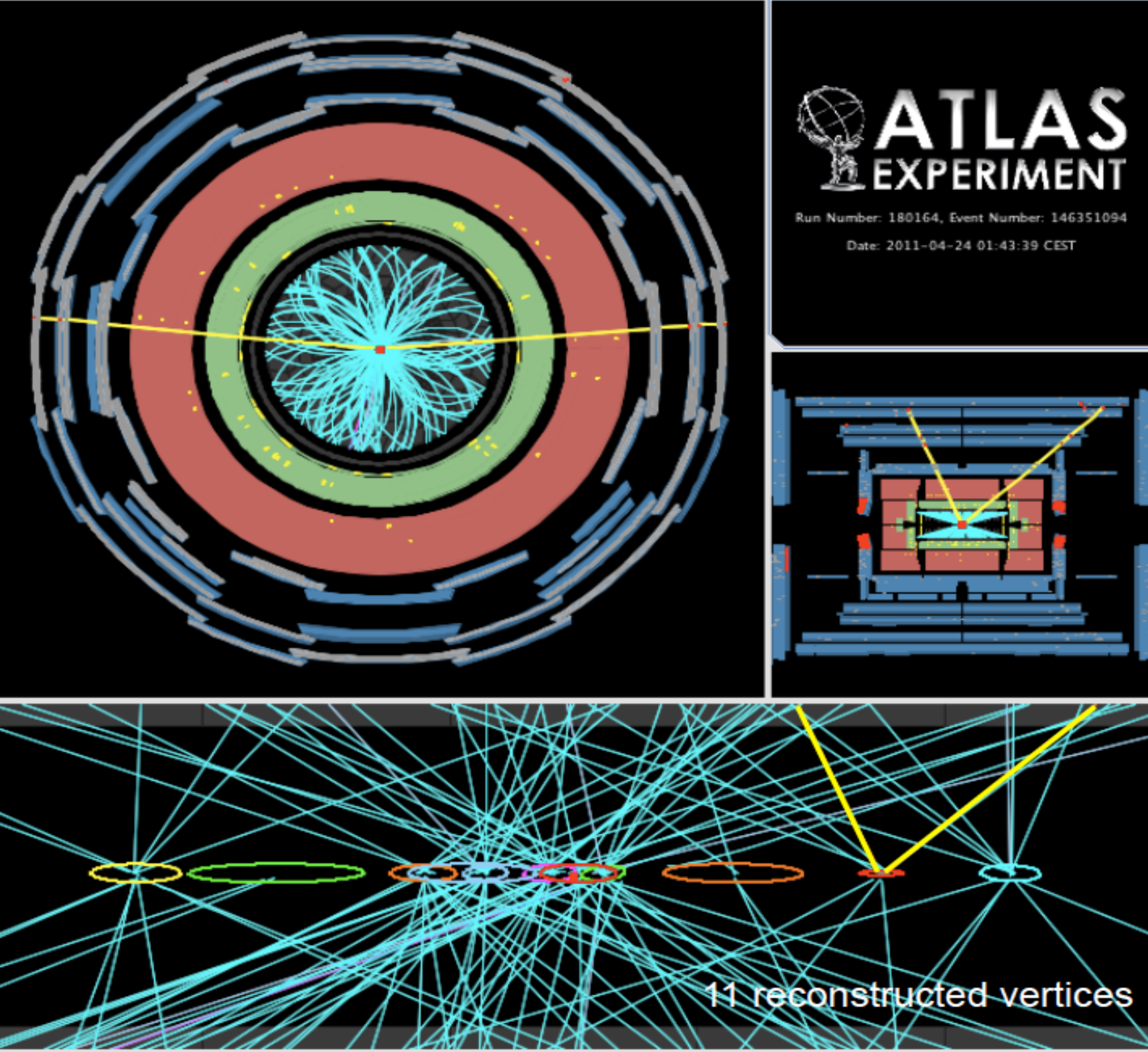
Figures adapted from Michaela Schaumann's third lecture (11/07/19) on "Particle Accelerators and Beam Dynamics"







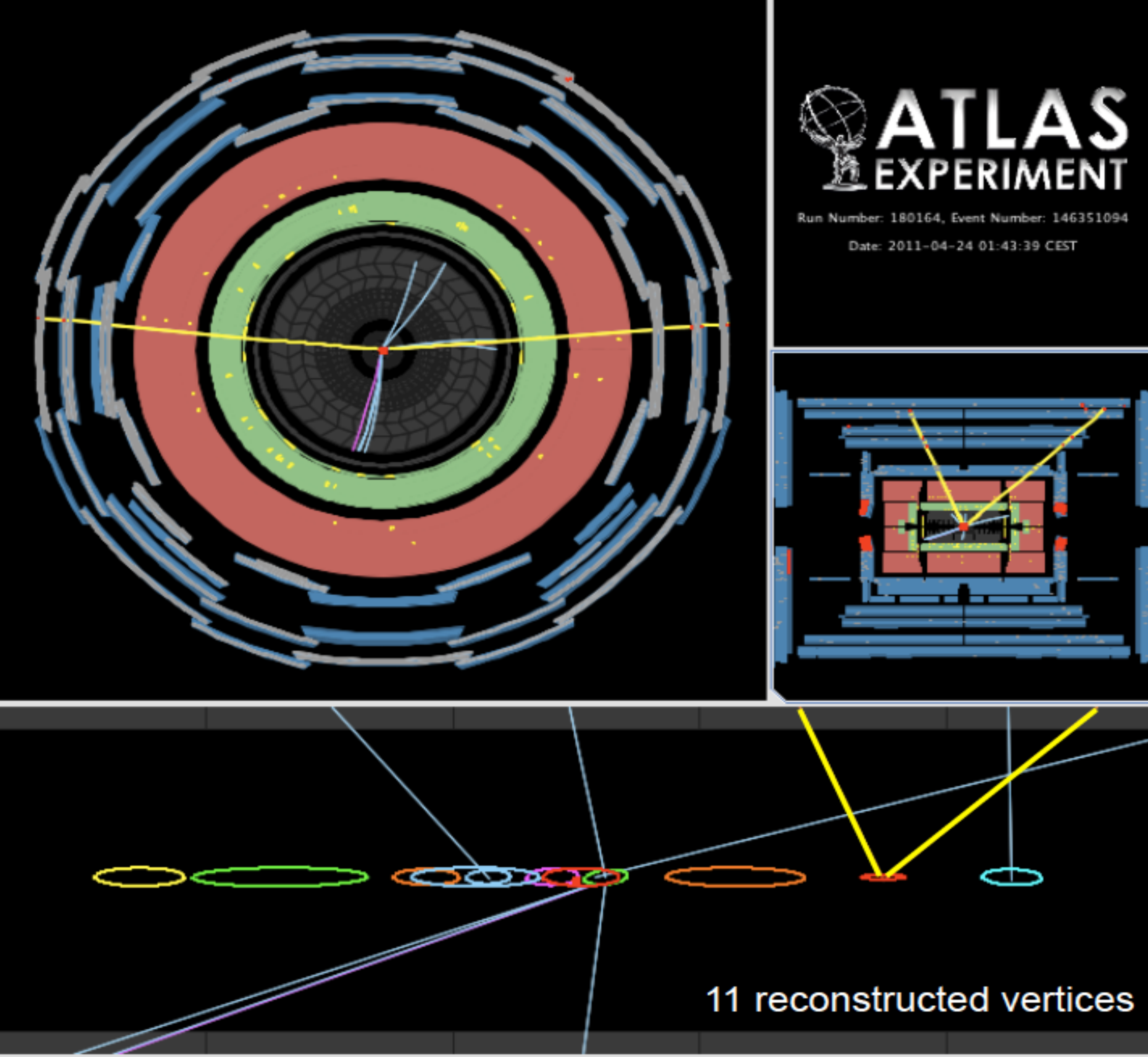
- The LHC accelerates ***bunches of 10^{11} protons*** separated by 25ns gaps



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

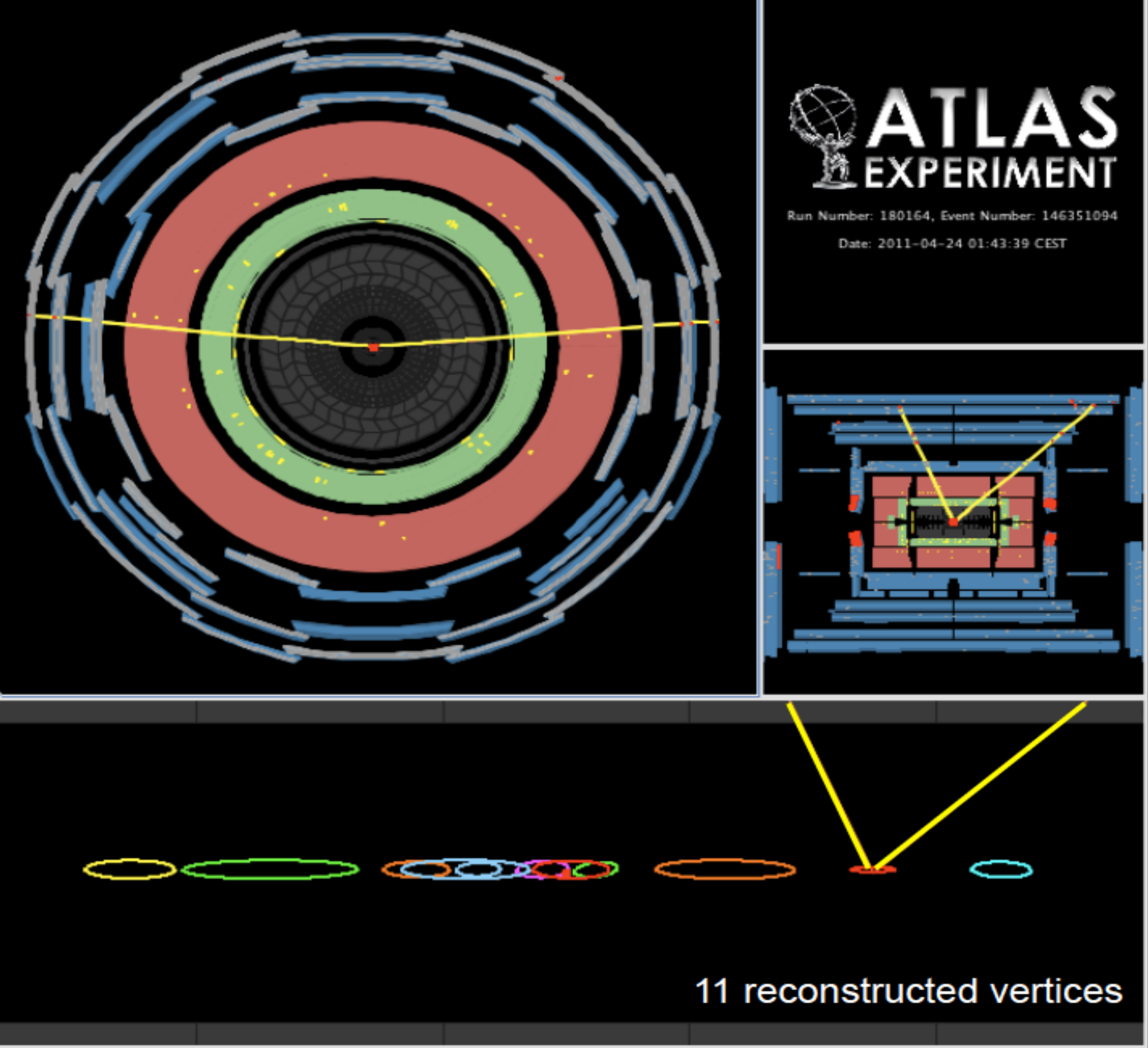
Track pT > 0.5 GeV



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

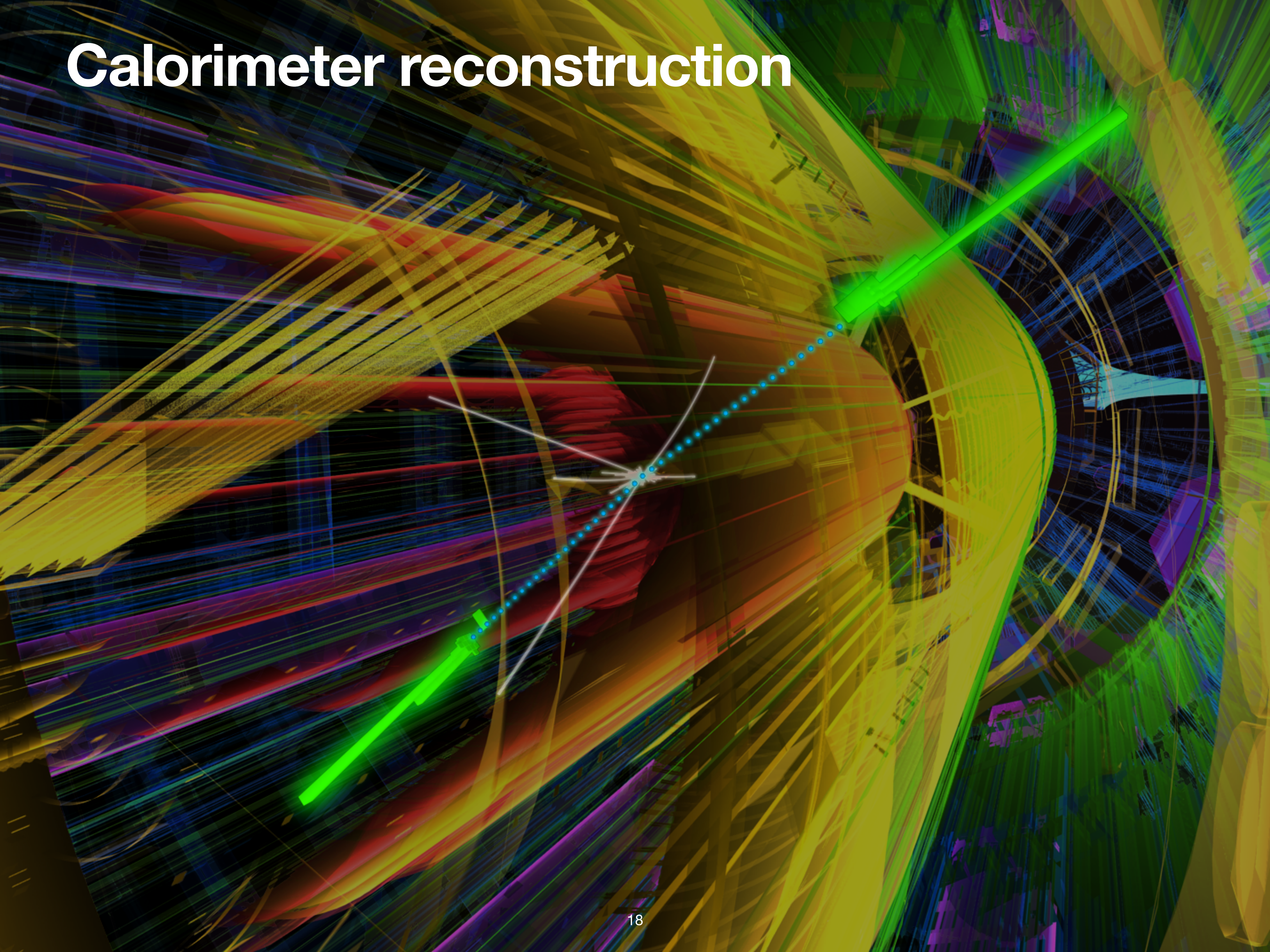


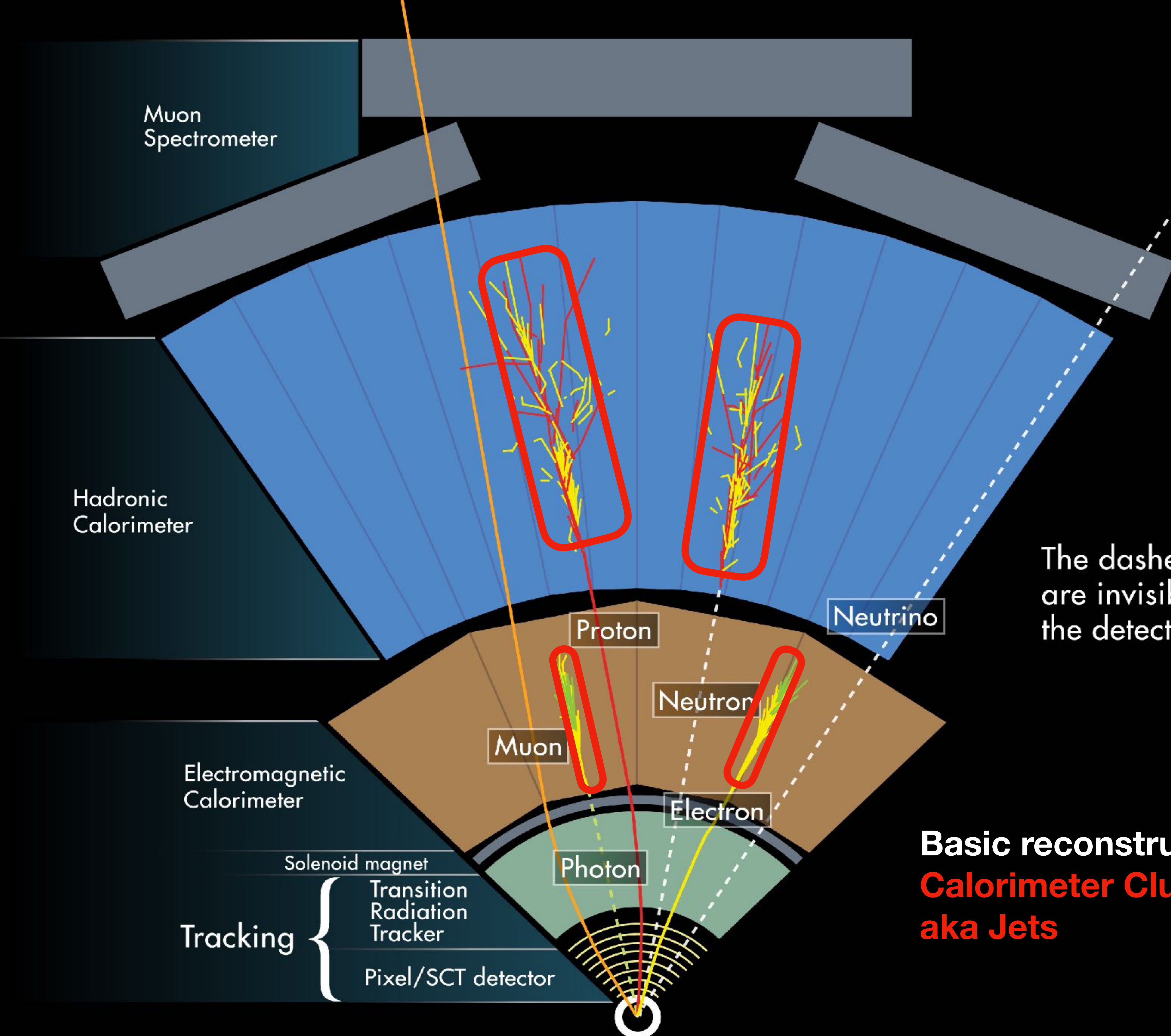
Z->μμ event;
2011 data.

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

Track pT > 10 GeV

Calorimeter reconstruction

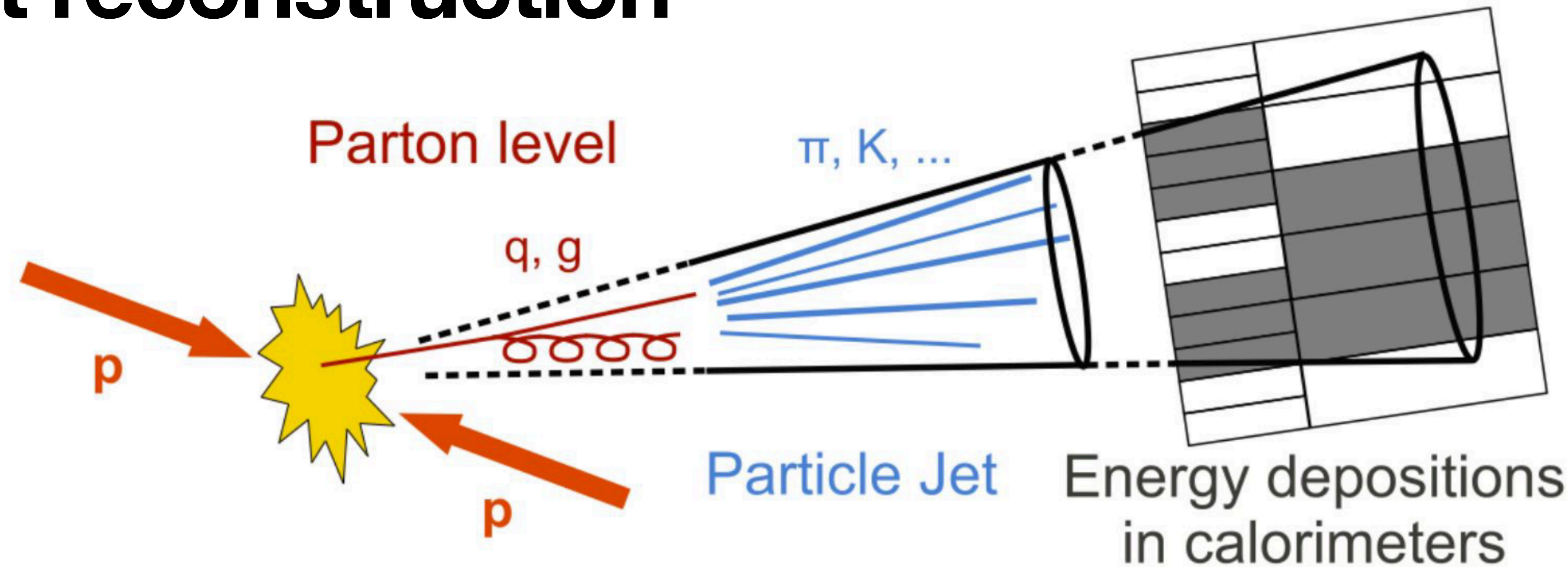




The dashed tracks are invisible to the detector

Basic reconstruction =
Calorimeter Clusters, aka Jets

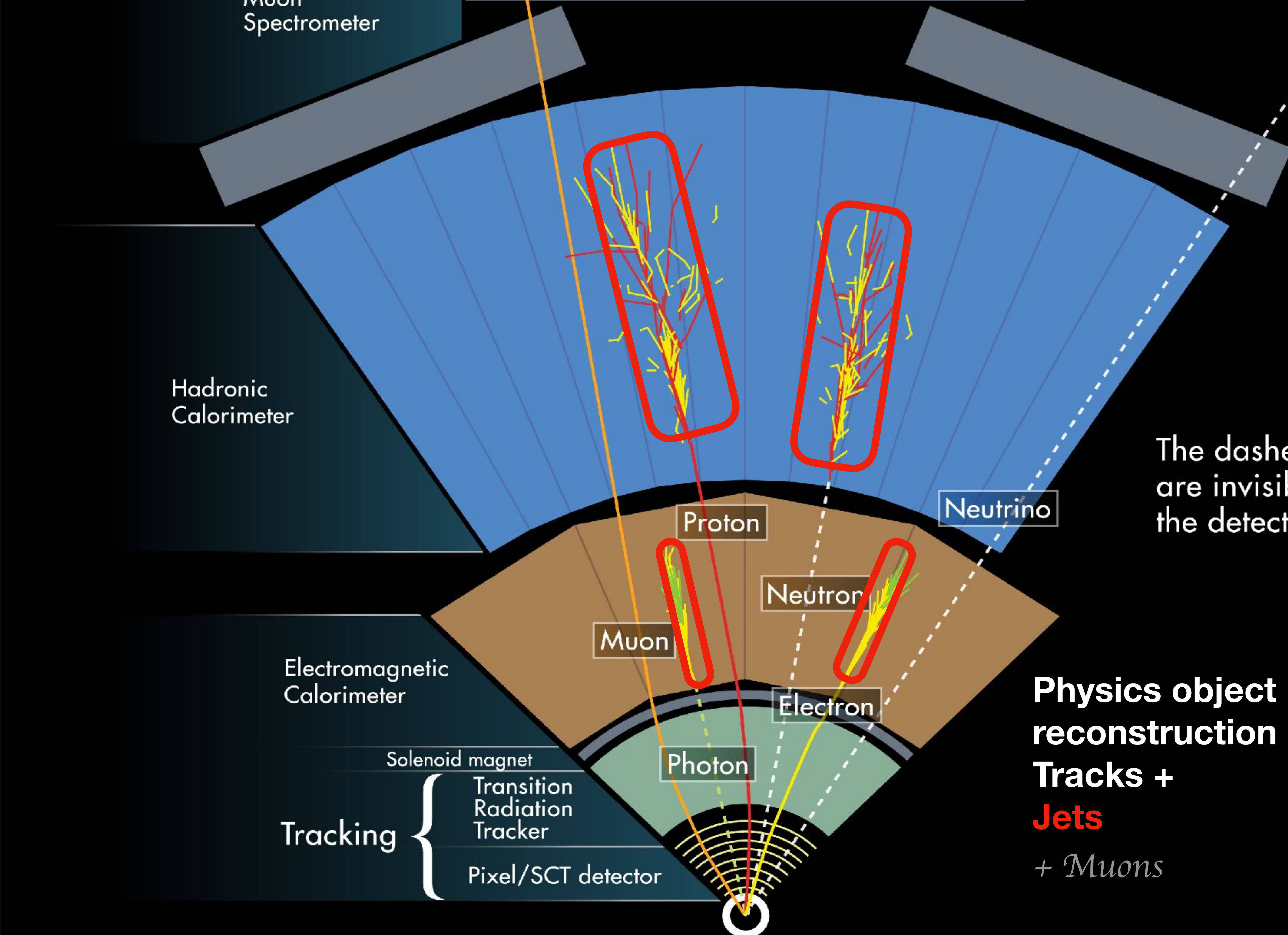
Jet reconstruction



- Quarks and gluons ***hadronize*** quickly and we detect ***sprays of hadronic particles*** in our detectors - we call these ***jets,*** proxies for the initial particle(s), we reconstruct them using ***jet algorithms***
- Hadronic particles leave energy deposits in the ***cells*** of the calorimeter, to reconstruct the energy of the hadronic particle, e.g. a pion, we need to sum the energy of the ***cluster*** of cells in which the pion deposited energy
- Deciding which cells belong to which cluster is a pattern recognition problem

Modern jet reconstruction uses Machine Learning!

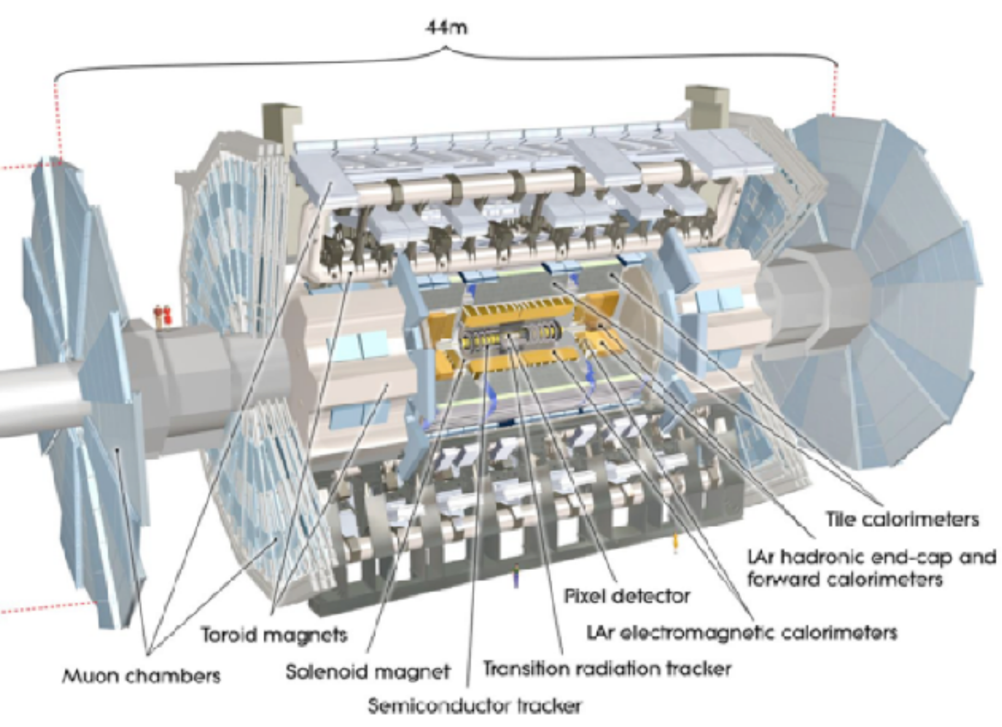
Here be dragons… and muons

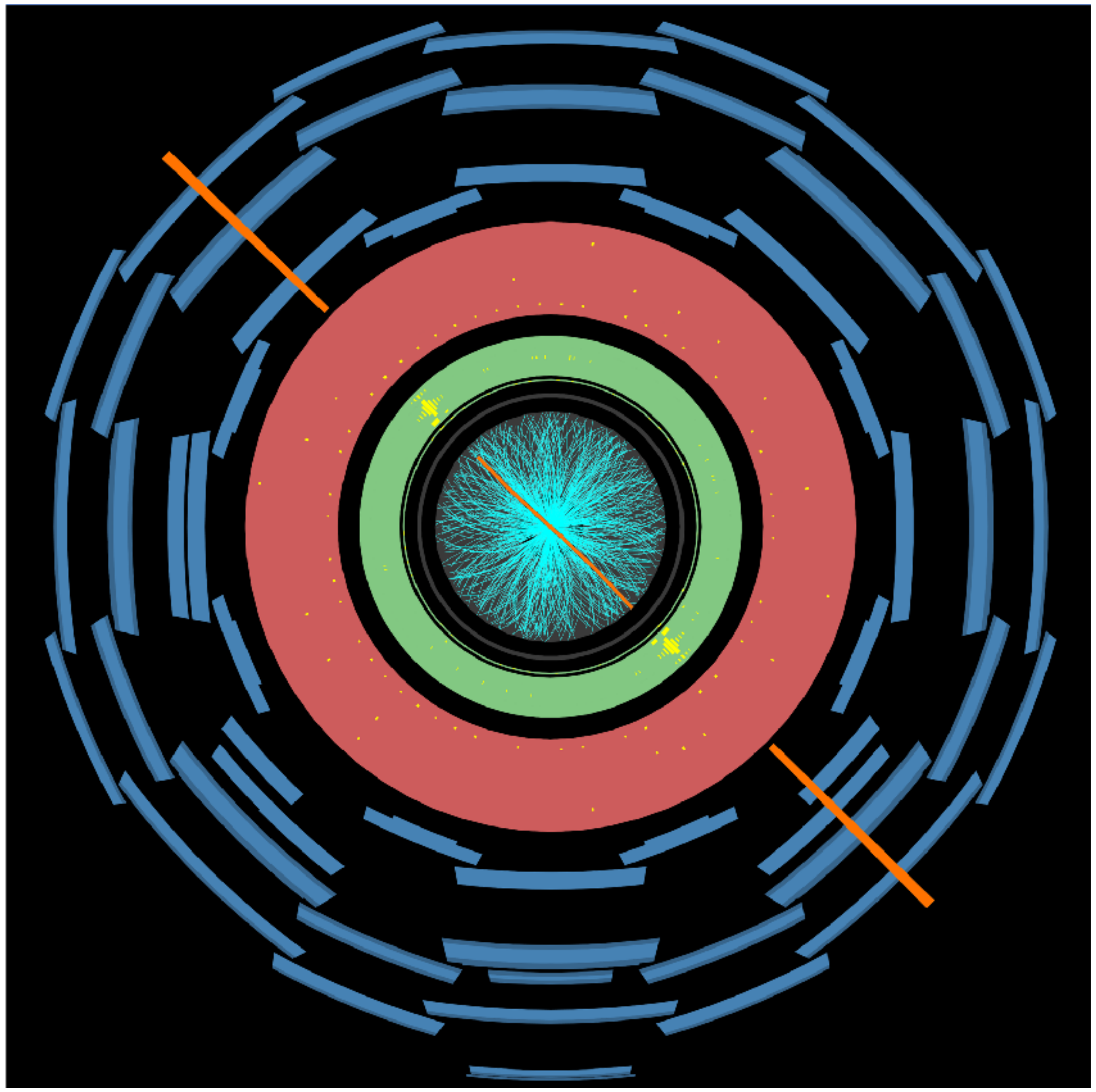


The dashed tracks are invisible to the detector

Physics object reconstruction = Tracks +
Jets
+ Muons

Neutrinos

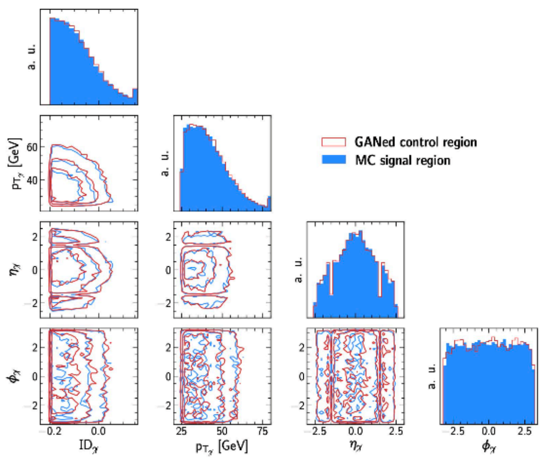


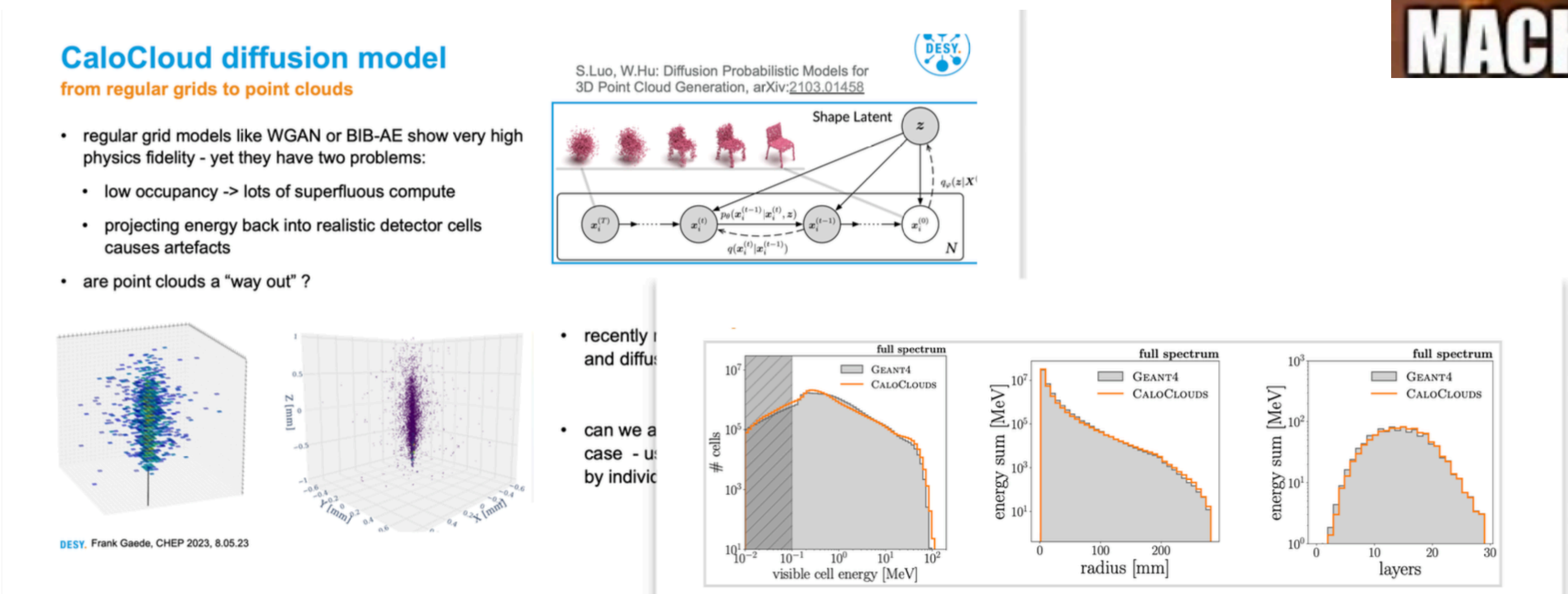


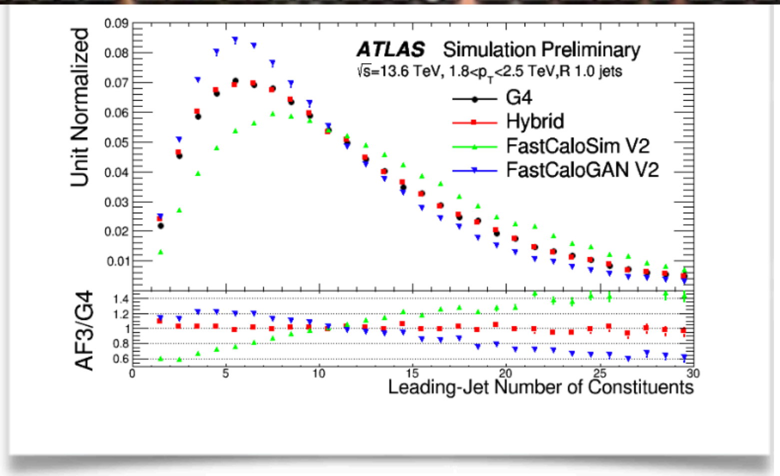
- Let's look at the simplest case for reconstructing neutrinos
- Remember, we are looking down the beam pipe, so the plane of the display is transverse to the proton beam direction
- ***Recall***: *Can you quantify the momentum in this plane* ***before*** *the proton collision*
    - *What does that 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*** *?*

Reconstruction today







- 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.:

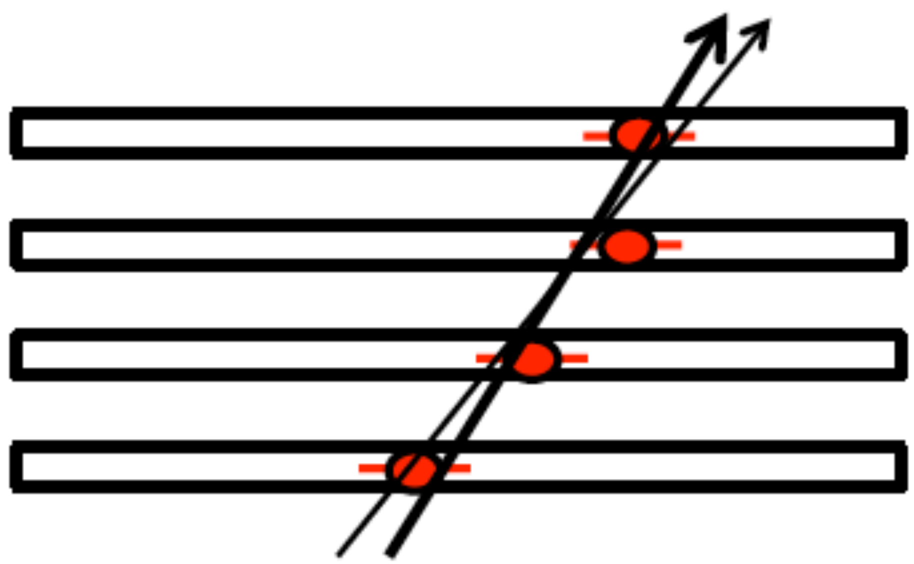
- Google Machine Learning Crash Course

Track fitting

- Perfect measurement – ideal

- Imperfect measurement – reality

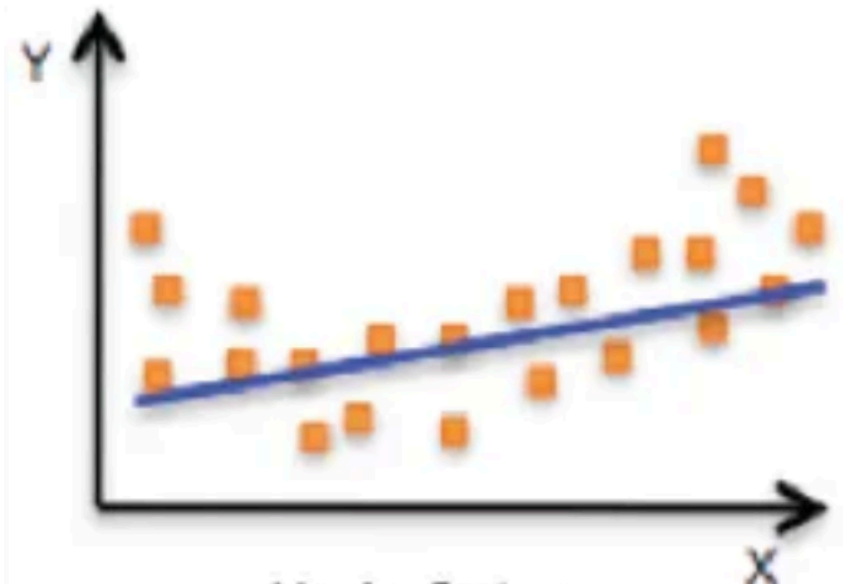
- Small errors and more points help to constrain the possibilities



What is the connection between least-squares minimisation and machine learning?

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

Machine learning (regression)

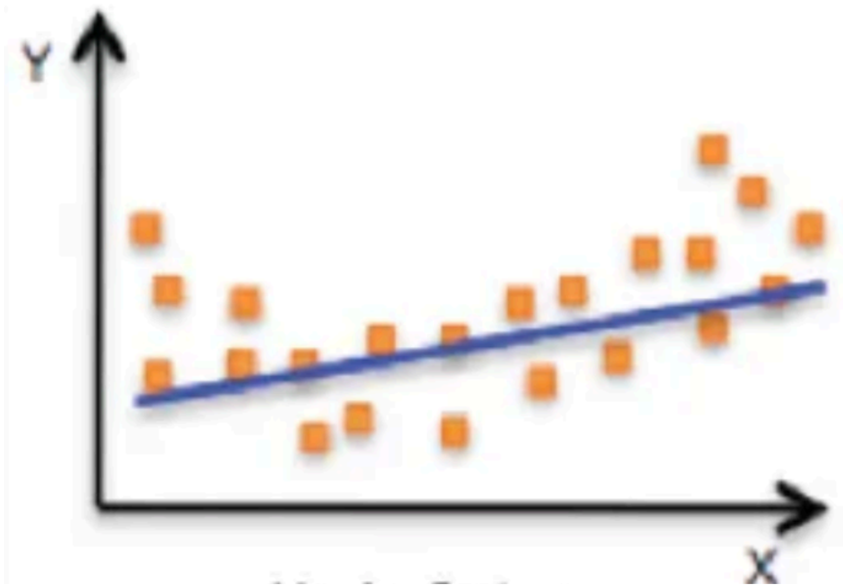


$$y_{model} = (x * a_{model}) + b_{model}$$

$$L = \sum_{N} (y_{model} - y_{data})^2$$

- Linear least squares minimisation compares a model to data
- *L* is the sum of the (squared) differences between the model prediction and the data
- Minimising *L* gives us the best parameters of the model

Machine learning (regression)



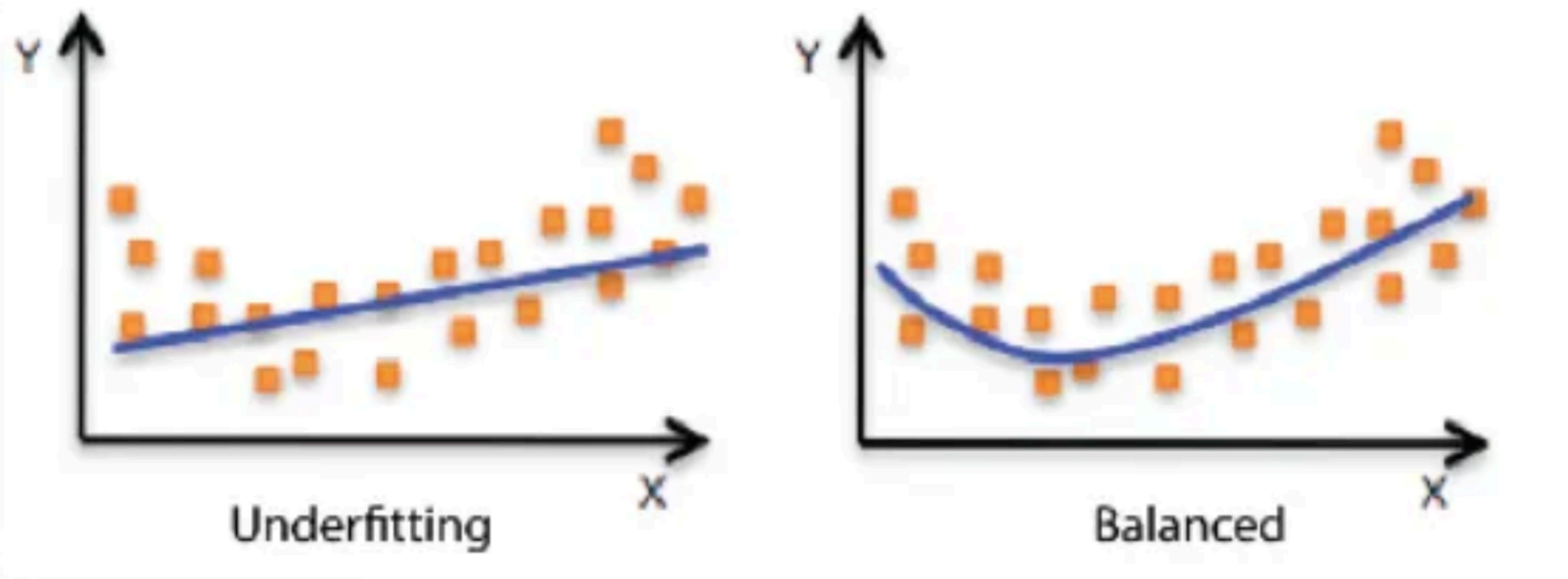
$$y_{model} = (x * a_{model}) + b_{model}$$

$$L = \sum_{N} (y_{model} - y_{data})^2$$

- Linear least squares minimisation compares a model to data
- *L* is the sum of the (squared) differences between the model prediction and the data
- Minimising *L* gives us the best parameters of the model
- We are often in a situation where we need to guess the model

Machine learning (regression)

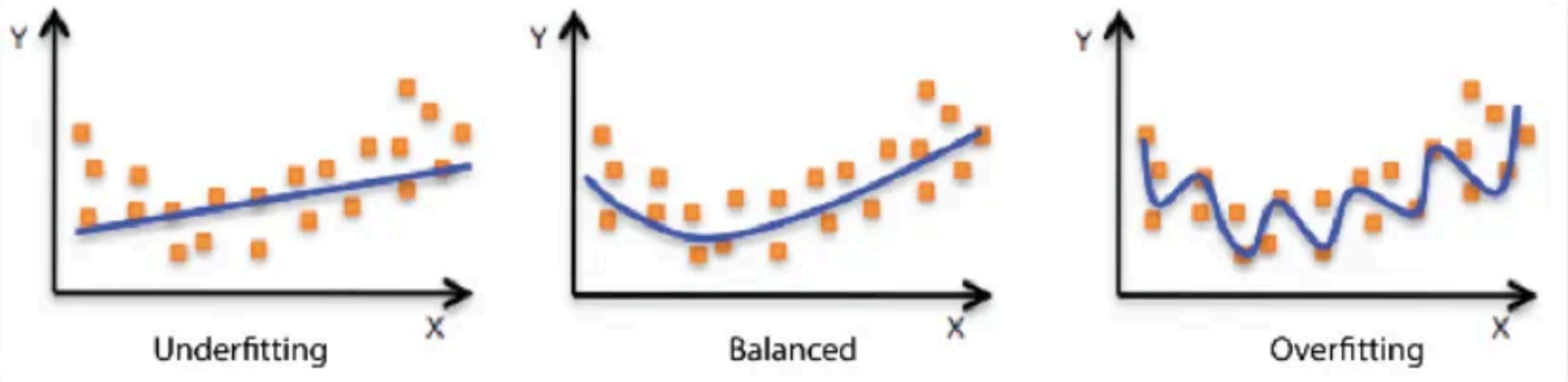
$$L = \sum_N (y_{model} - y_{data})^2$$



- Increasing the #parameters of the model will often achieve a better description of the data (reduce *L*)

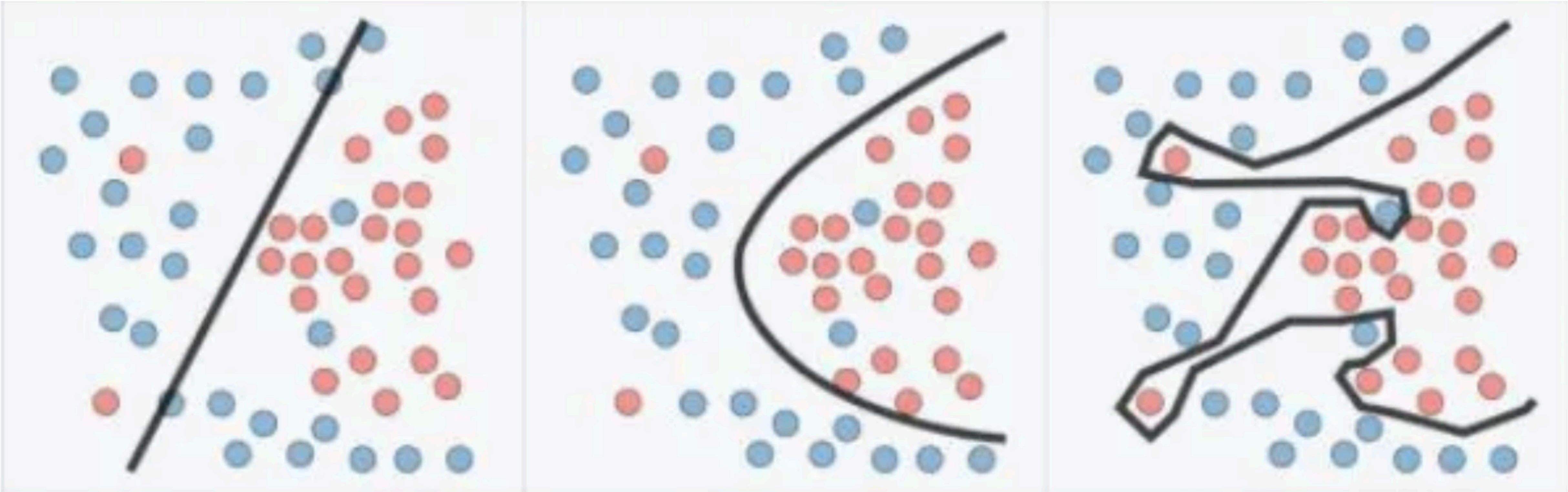
Machine learning (regression)

$$L = \sum_{N} (y_{model} - y_{data})^2$$



- Increasing the #parameters of the model will often achieve a better description of the data (reduce *L*)
- But this has drawbacks, we want a model that can describe ALL data, including data we haven't seen yet / data not included in the fit
- Important to strike a balance between #parameters and quality of fit (*L*)

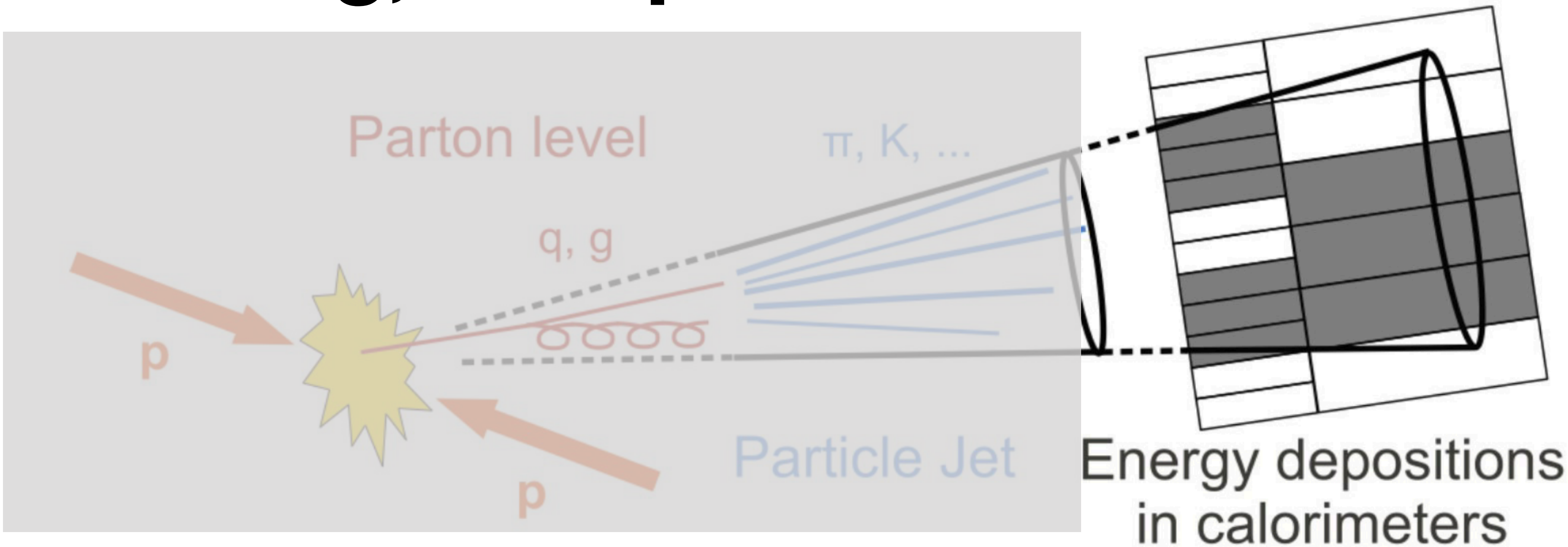
ML classification (supervised)



Underfitting Balanced Overfitting

- Classification works in a similar way, here we model the separation between two populations, red and blue
    - ***We fit a model that describes the shape that separates the data points***
- Again we need to be careful not to overfit our training data or our model will not be general enough to describe new data that were not included in the fit

Clustering, unsupervised ML classification



- Sometimes there is no model, for example reconstructing clusters of energy deposits in calorimeters

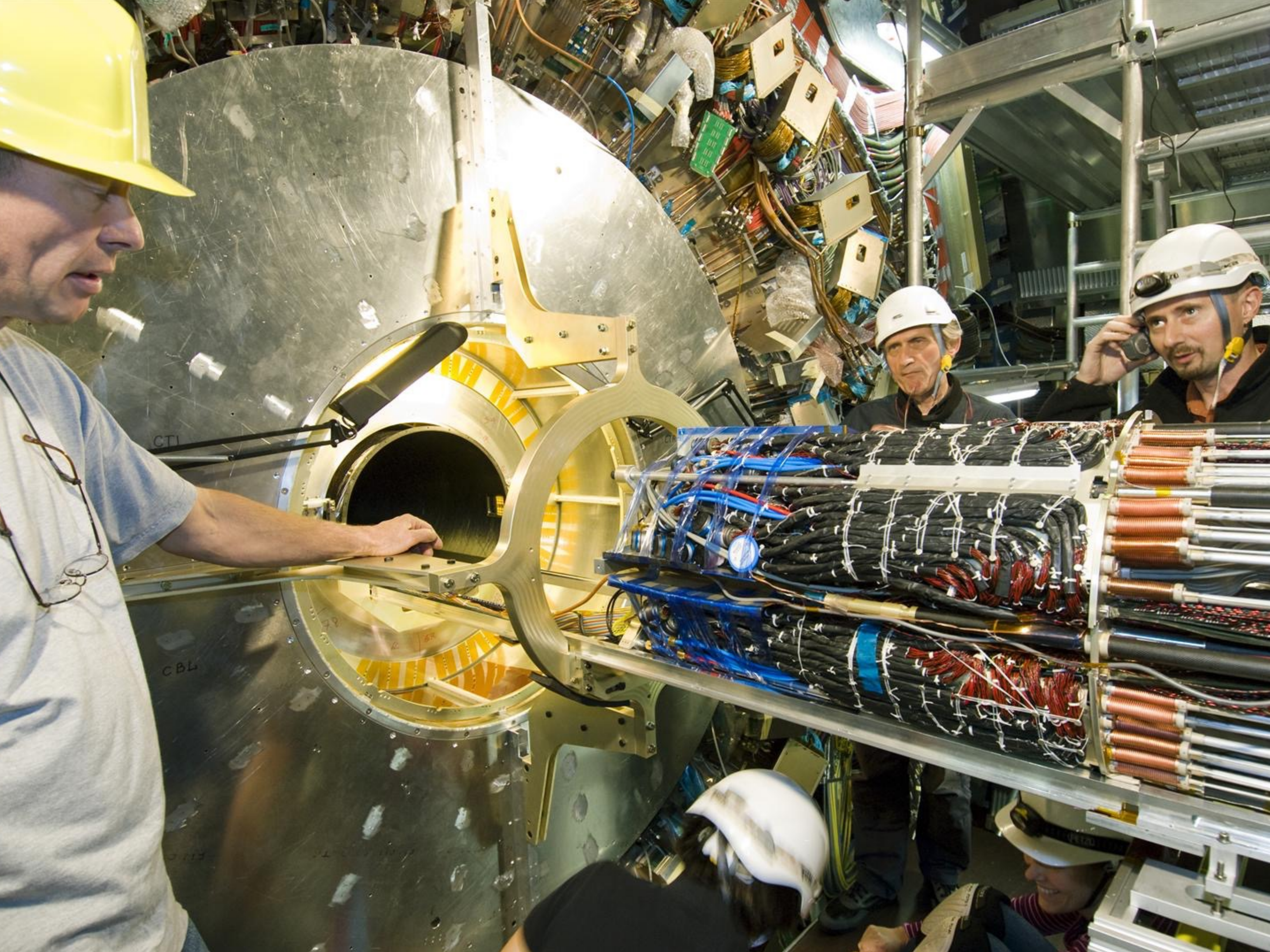
- Instead of defining a number of clusters to reconstruct and tuning that model, we cluster energy deposits (cells) around a varying number of centres ($N_{clusters} = 1,2,3...$)

- We need a metric to choose the best solution ($N_{clusters}$), e.g. increasing the number of clusters by 1 did not improve the total cluster quality by >10%

Data Preparation

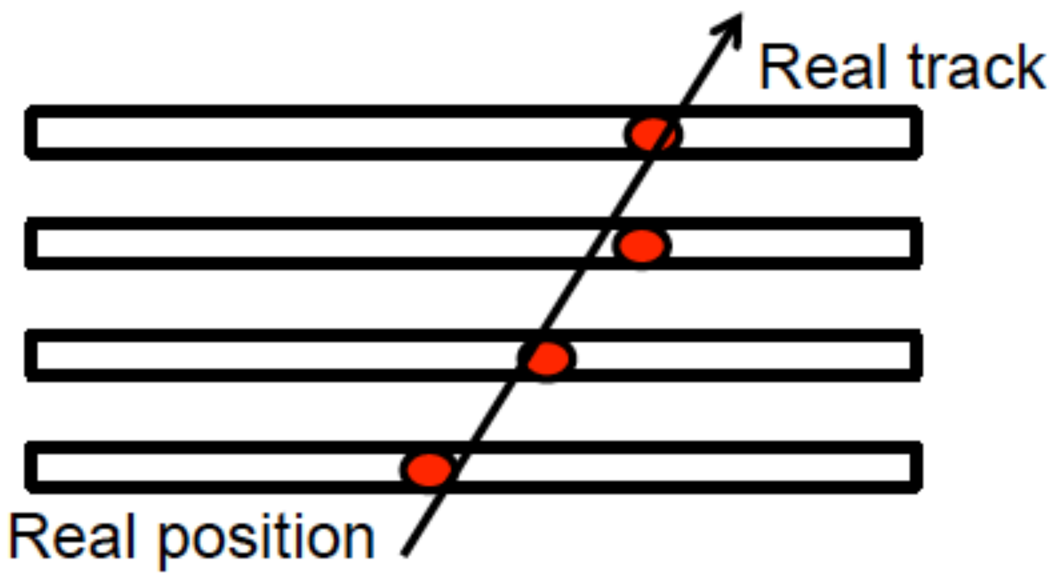
- Three major steps to **prepare data for physics analysis** and achieve
  - reliable, high quality data (yes, we ***reject*** low quality data)
  - the ***best performance*** from our detectors
  - readiness for ***physics analysis***

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.

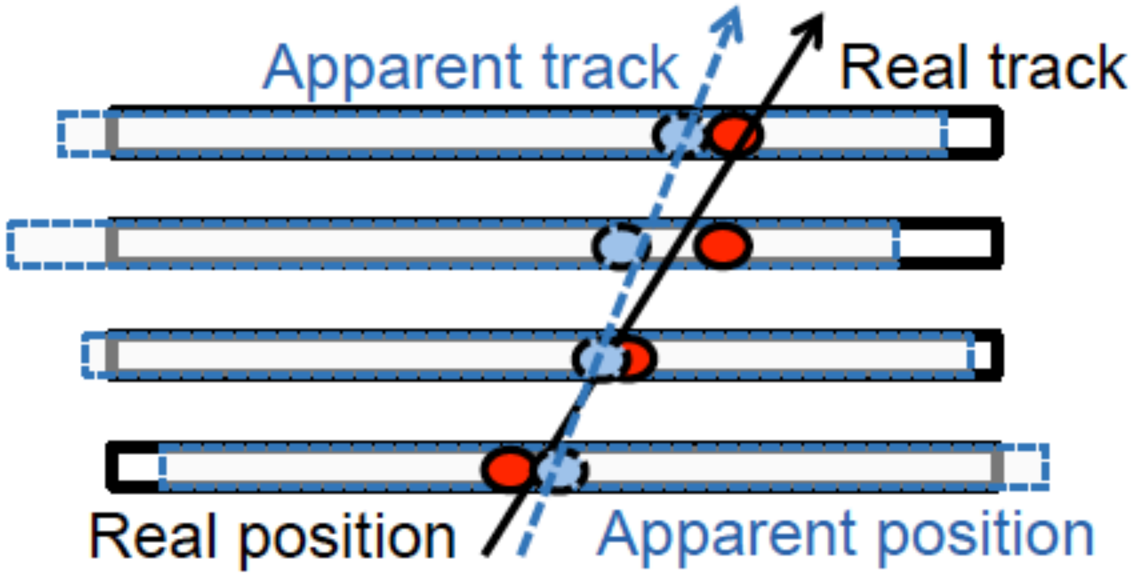


Correcting detector effects - calibration

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

Q. How could we measure detector misalignment ?

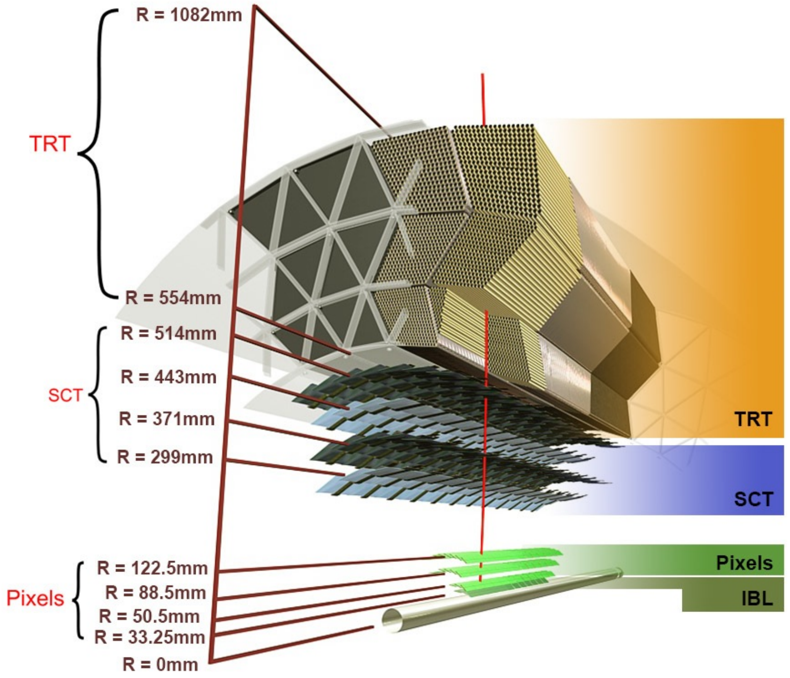
- **Misalignment**
  - Detector elements not positioned in space with perfect accuracy.
  - Alignment corrections derived from data and applied in track reconstruction.



Real vs perfect tracking detectors

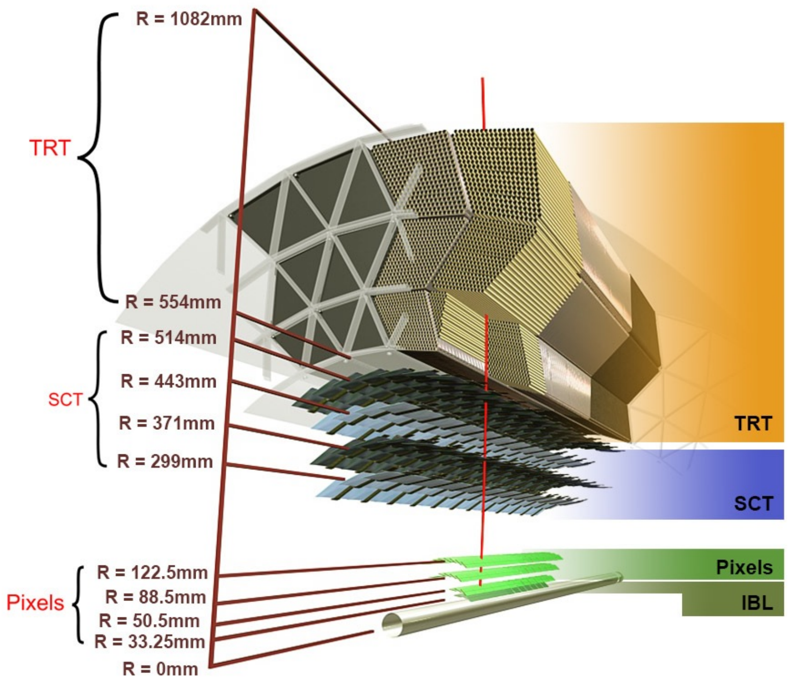
- The perfect tracking detector
    - is constructed from zero mass material
    - has no noise
    - is 100% efficient
    - has perfect resolution
- A real tracking detector
    - is constructed from real material
        - particles interact with the detector and scatter, altering the particle trajectory
    - suffers from noise
        - noise can be confused with particle tracks
    - has less than 100% efficiency
        - particles are not always detected, there can be dead regions
    - has finite resolution
        - it may not always be possible to resolve two particle trajectories

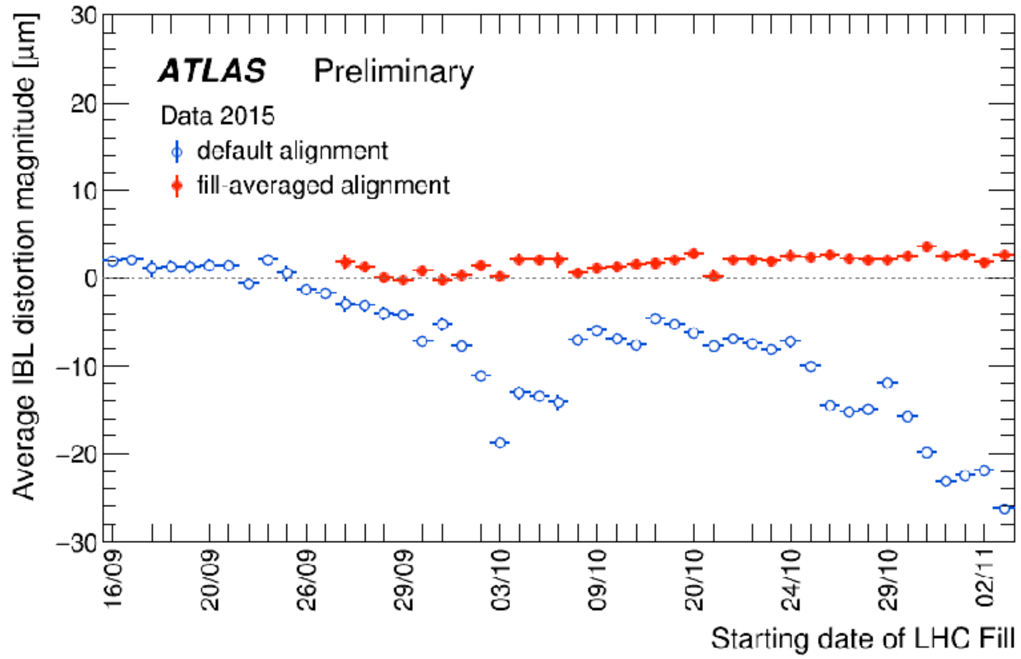
Calibration



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

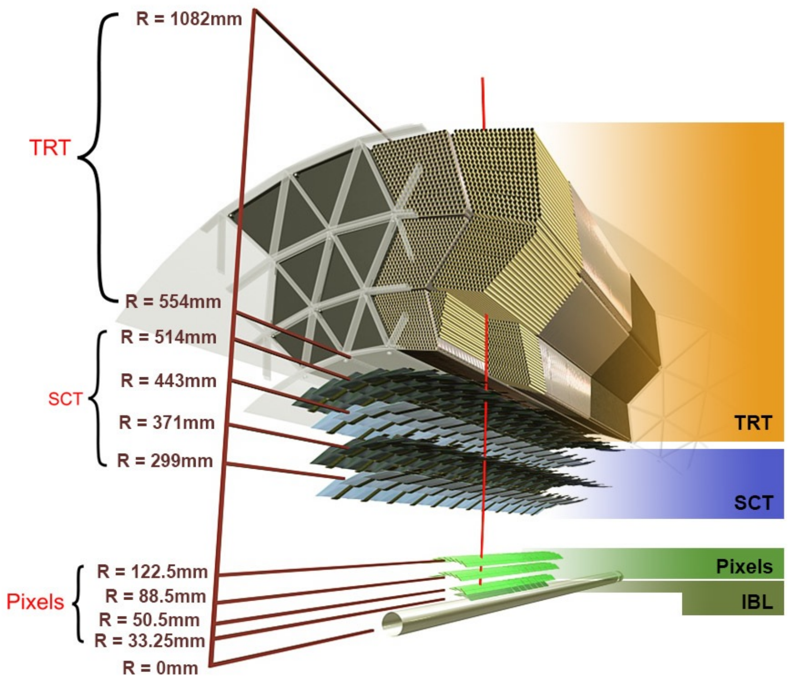
Calibration

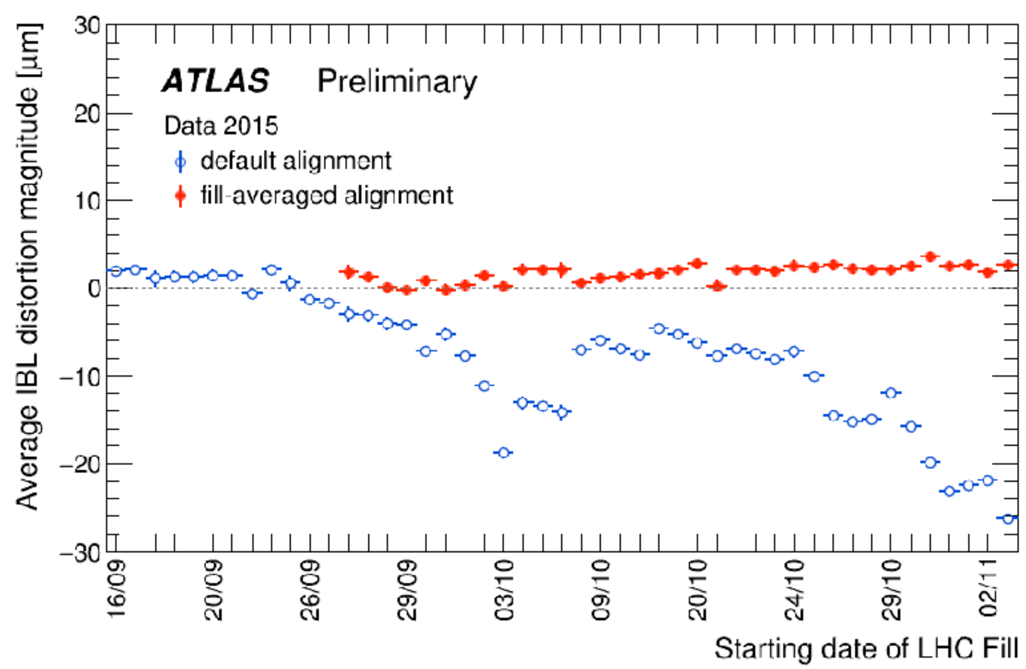


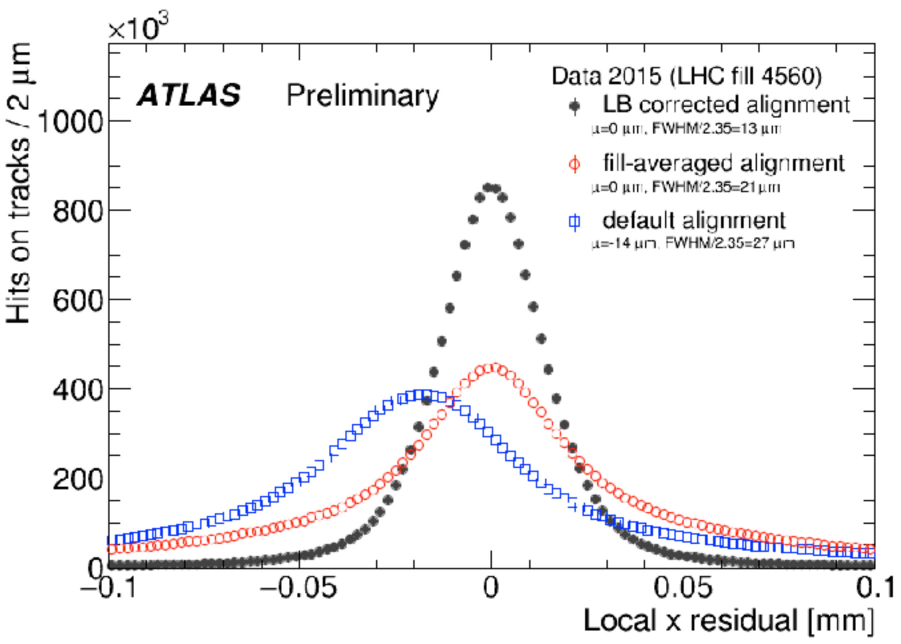


- During the break between Run 1 and Run 2, ATLAS inserted the IBL, an extra layer of silicon tracker close to the beam pipe
- At the start of data taking in Run 2, it started to move during data taking runs

Calibration

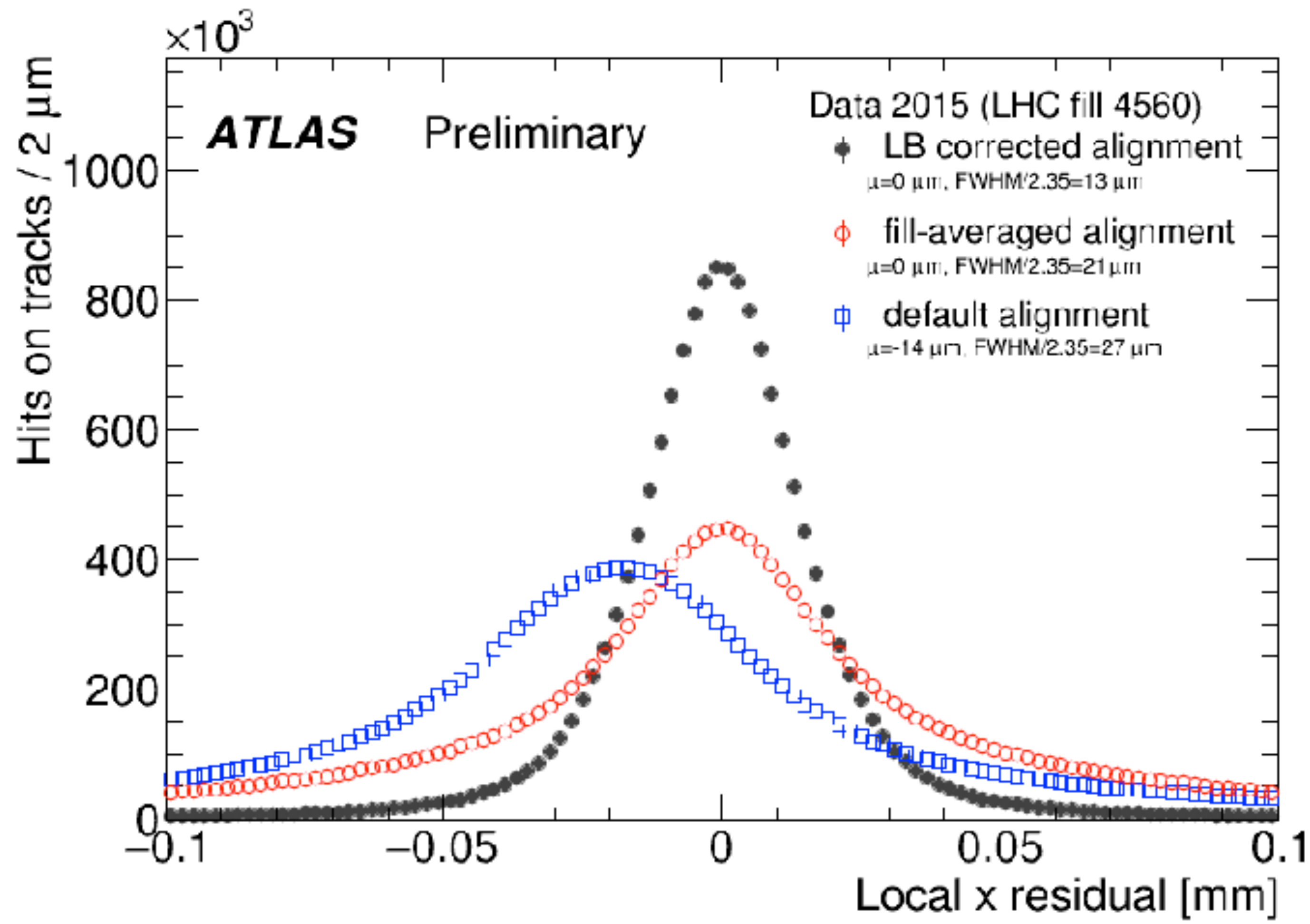






- During the break between Run 1 and Run 2, ATLAS inserted the IBL, an extra layer of silicon tracker close to the beam pipe
- At the start of data taking in Run 2, it started to move during data taking runs
- As time went on, the movement was very significant, much more than the detector precision, so the movement could be seen in physics distributions and data quality
- ATLAS implemented and commissioned a correction procedure per run (red) and then per **sub-run (black)** as part of its calibration process
- Following the correction the performance of the detector was back to nominal

Calibration quality



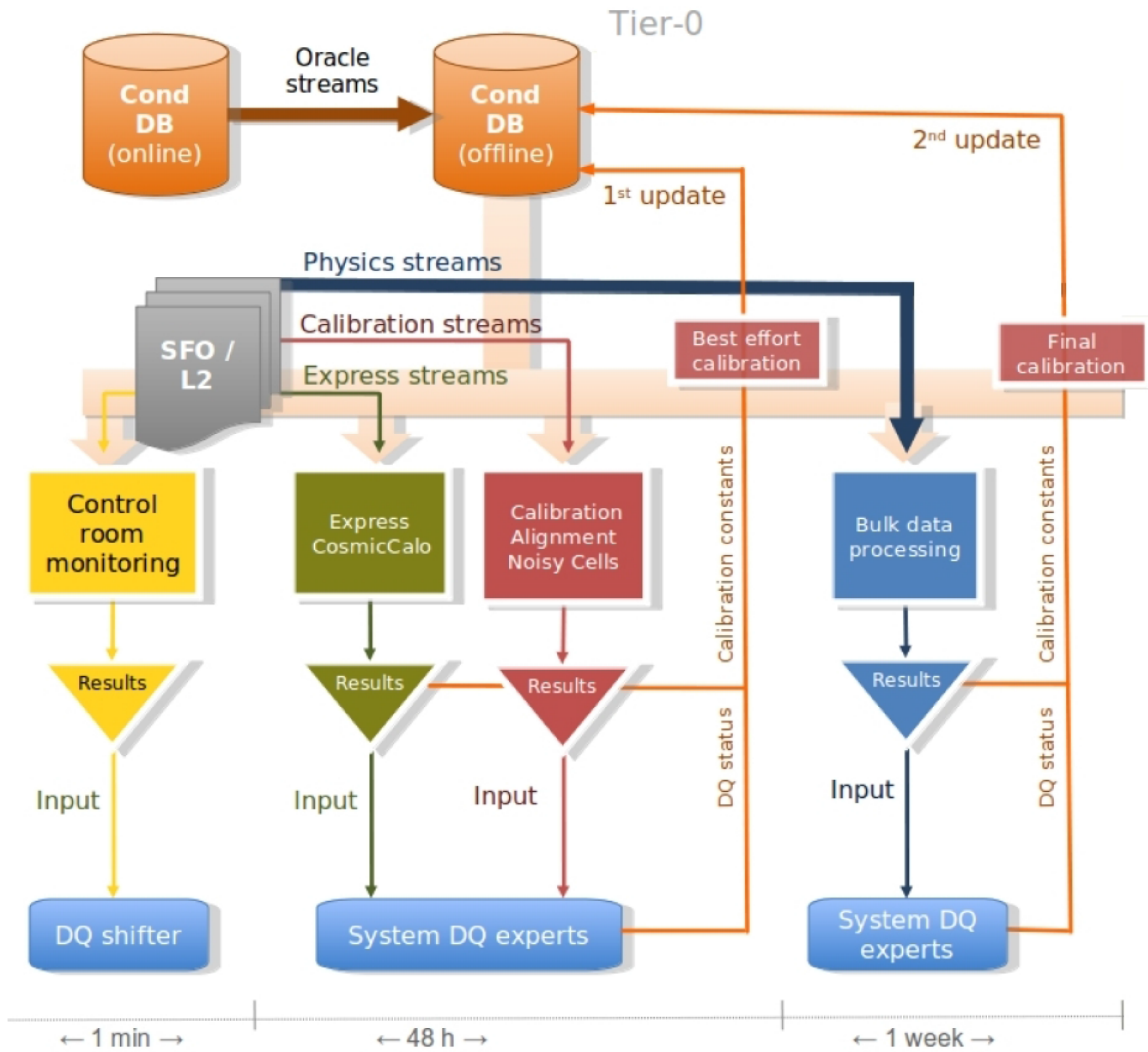
- Explain why the correction procedure per run (red) still resulted in a loss of precision and the per **sub-run (black)** procedure was necessary to retrieve the nominal detector performance

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



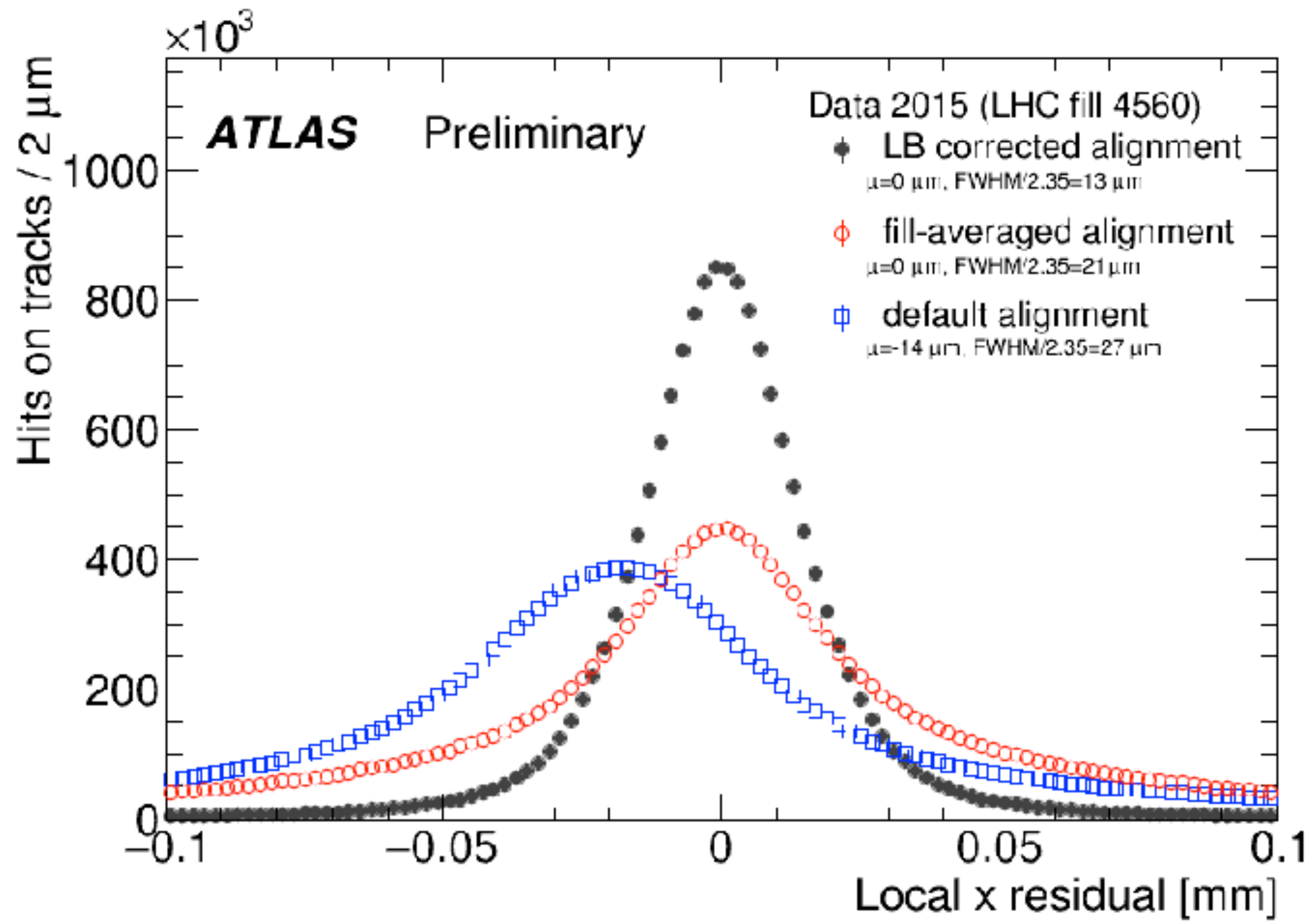
Check during data taking

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?



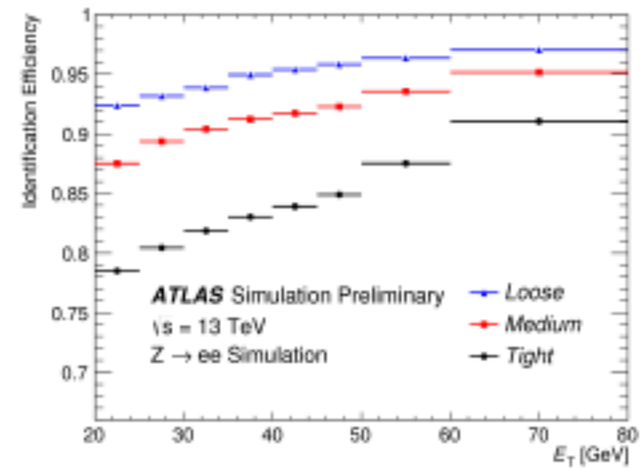
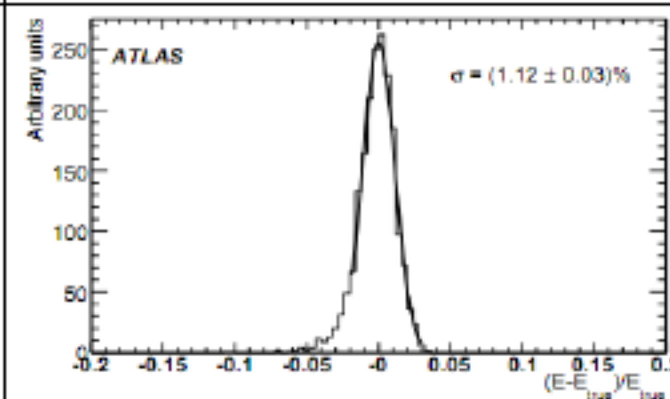
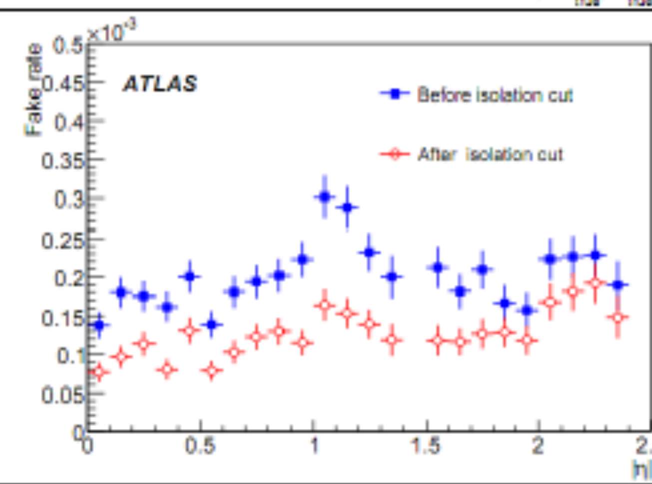
- The **ATLAS IBL** is a good example of a ***data quality*** problem

    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!

Reconstruction figures of merit and data quality

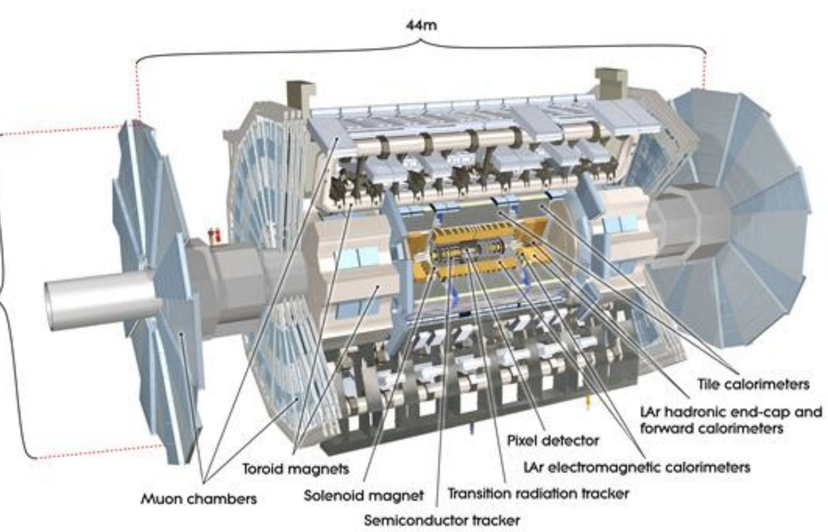
	Definition	Example		Needs be:
Efficiency	how often do we reconstruct the object	electron identification efficiency = (number of reconstructed electrons) / (number of true electrons) in bins of transverse momentum		High
Resolution	how accurately do we reconstruct the quantity	energy resolution = (measured 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

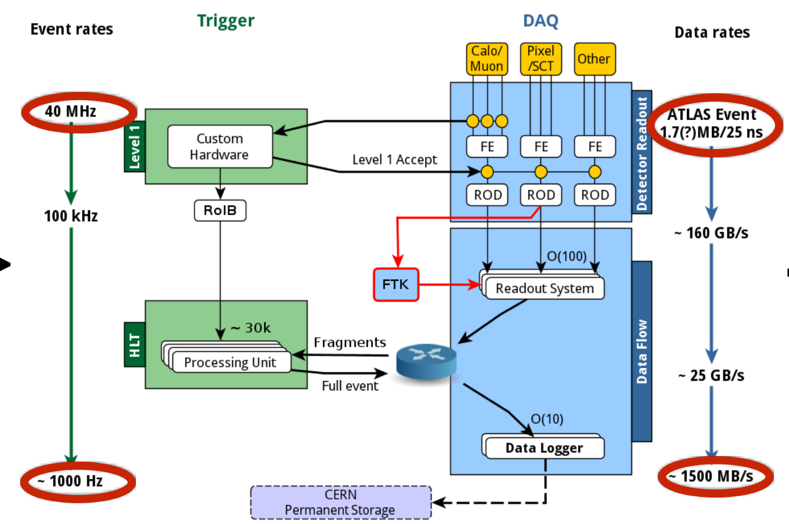
Data Preparation

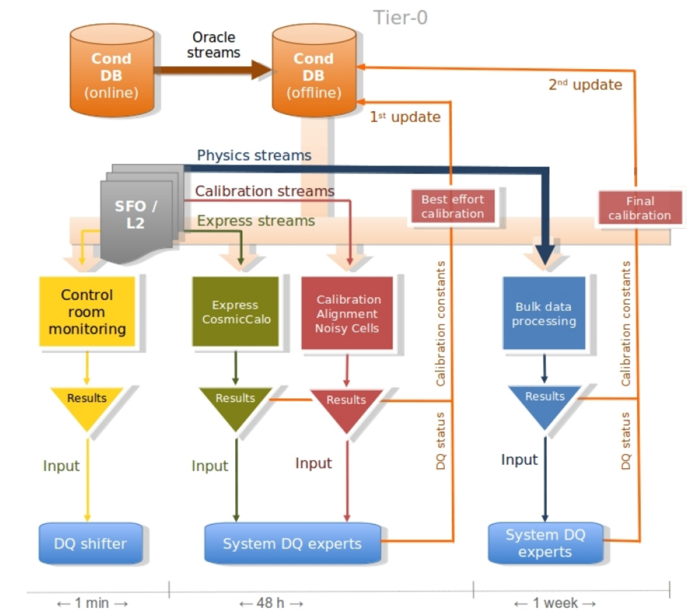
- Three major steps to **prepare data for physics analysis** and achieve
  - reliable, high quality data (yes, we ***reject*** low quality data)
  - the ***best performance*** from our detectors
  - readiness for ***physics analysis***

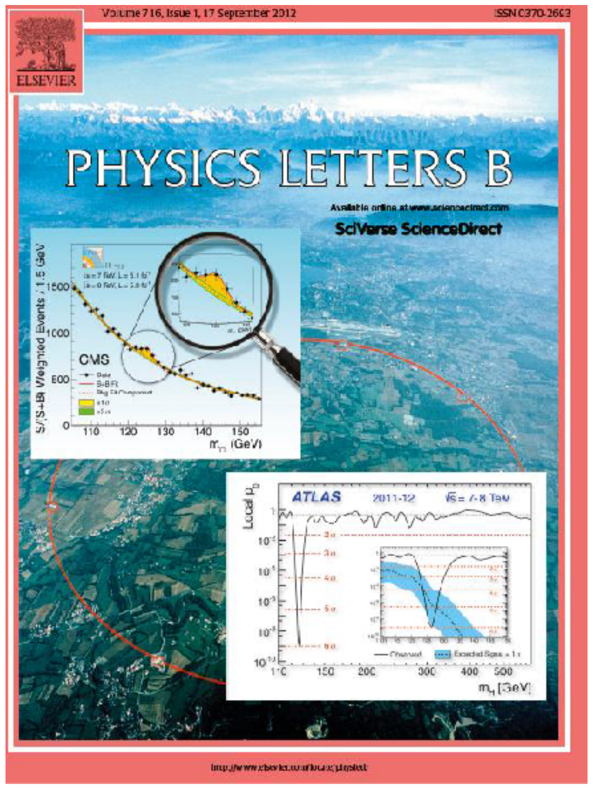
1. **Reconstruct physics signals** from the data ✓
   - Produce information like how many muons does the event have?
2. **Calibrate** the detectors ✓
   - Correct imperfections, account for changes over time…
3. Make sure that the **data quality** is excellent, also in real time ✓
   - Maximise the amount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

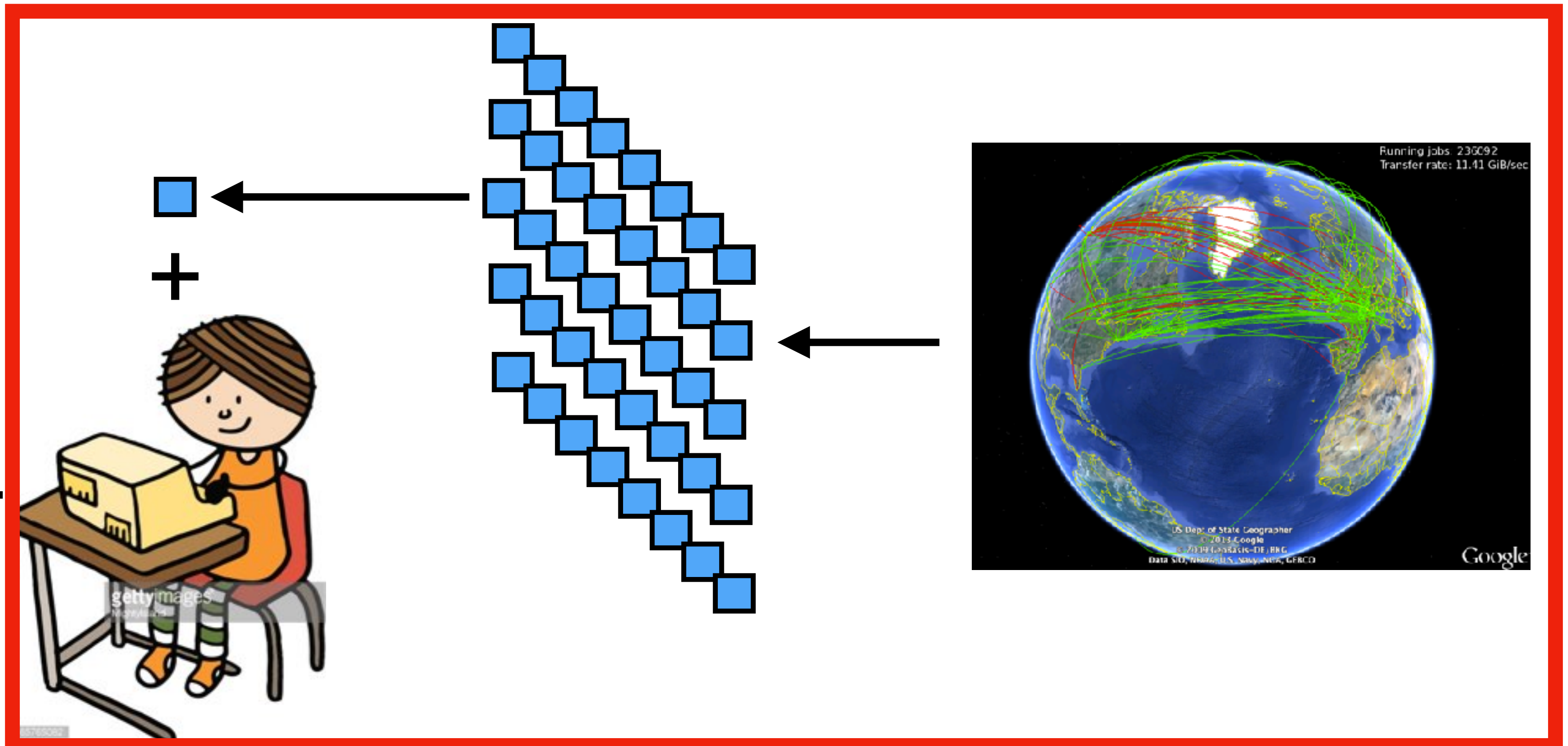
Data's journey - next time, analysis!











Contact details

- I am usually based at Geneva Observatory in Versoix, but will be here at CERN Wednesday 3rd through Friday 5th July
    - Tonight you will find me here

- email: paul.laycock@unige.ch