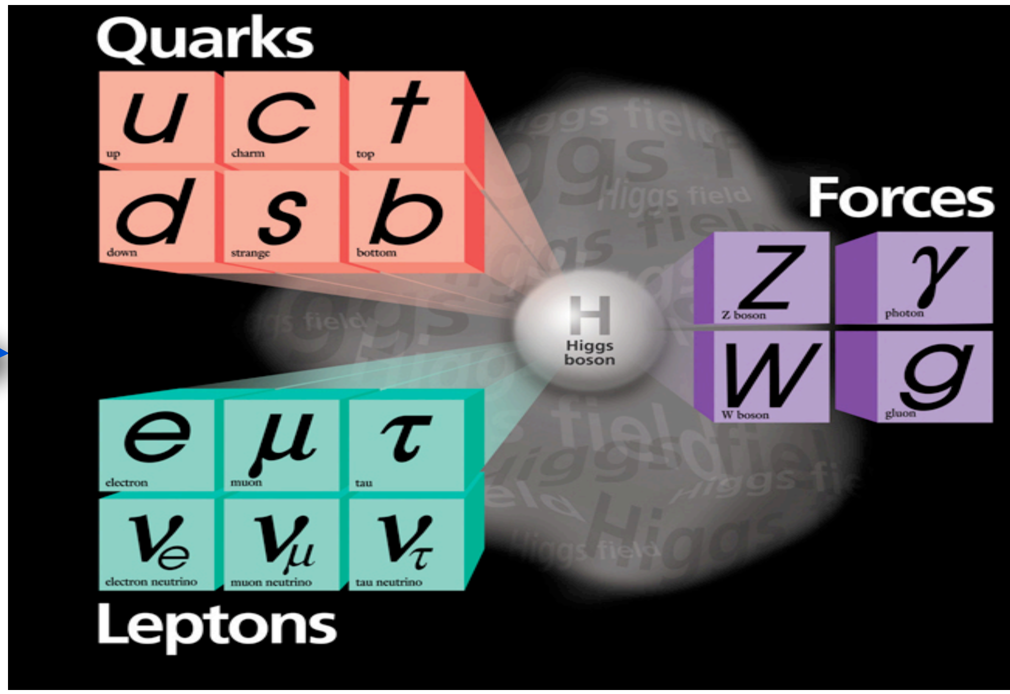
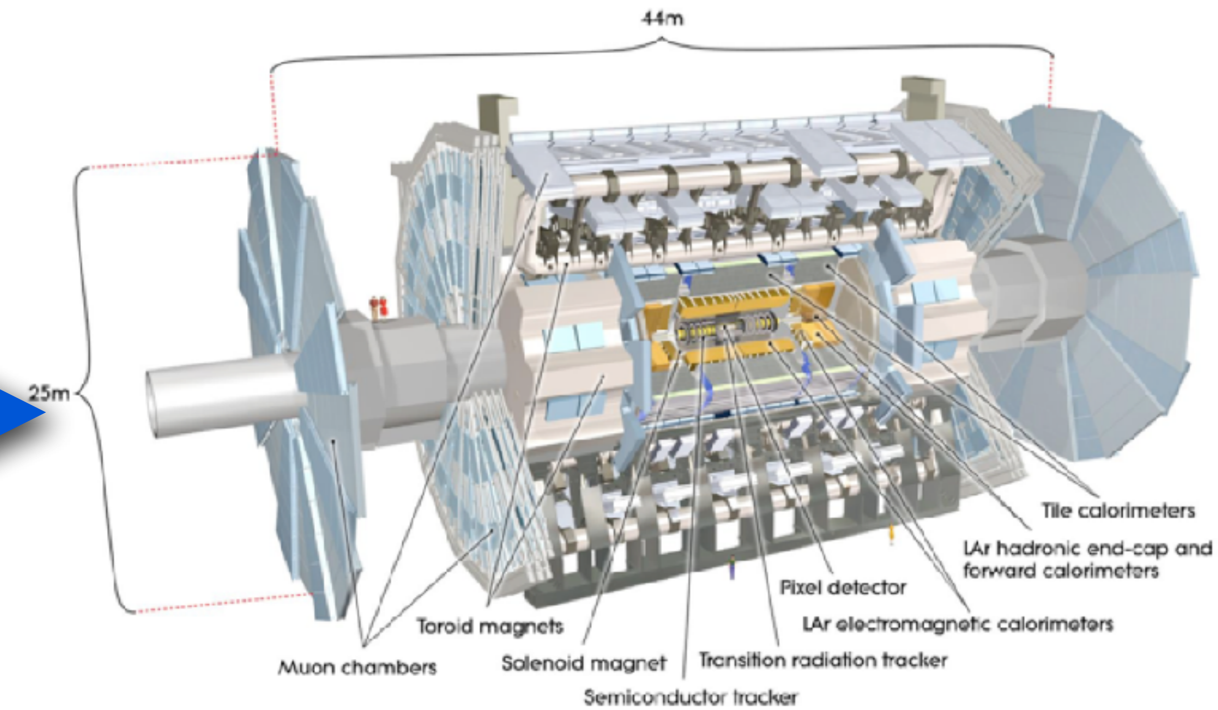
From Raw data to Physics Results (3/3)

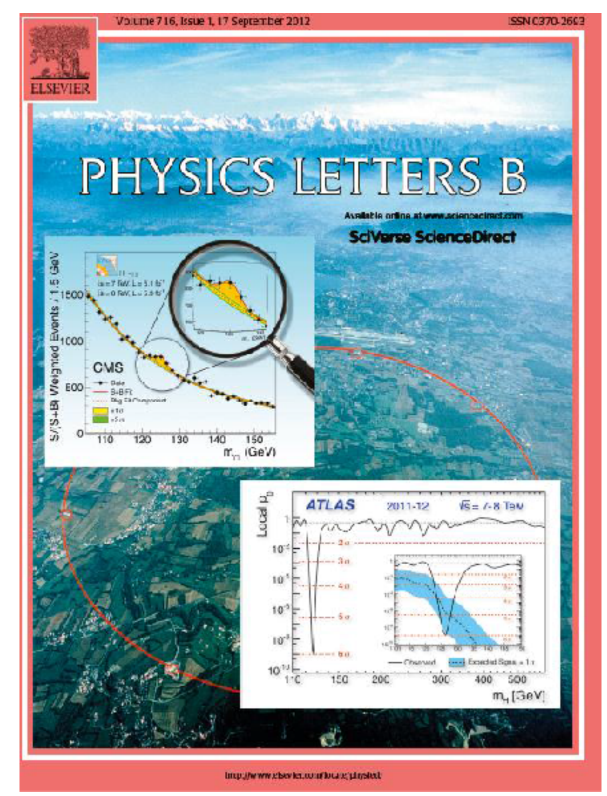
Paul Laycock

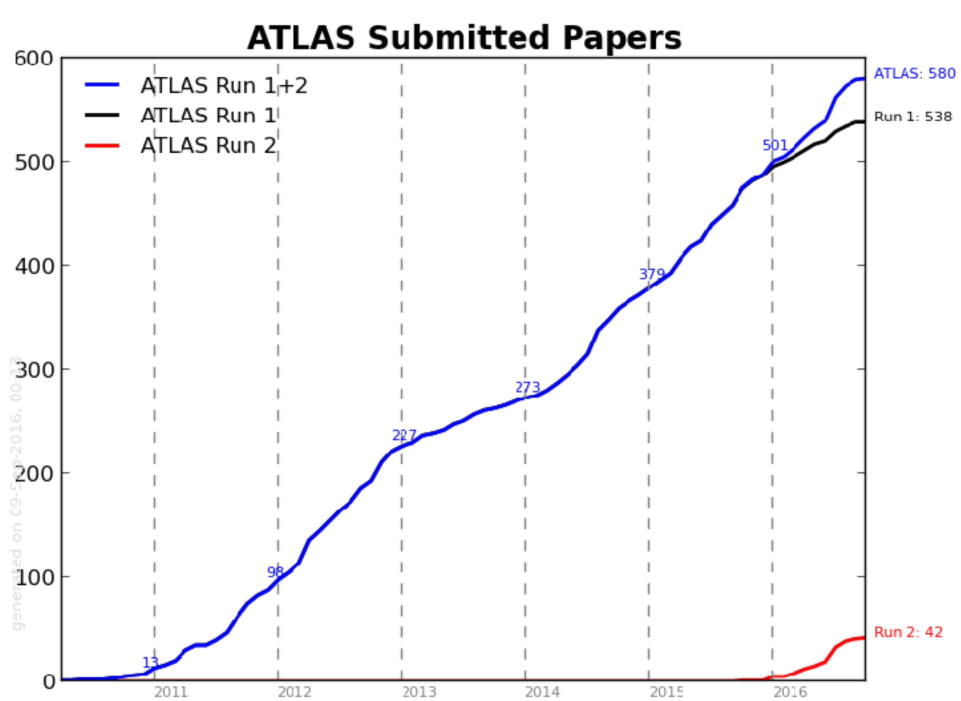
July 5th 2022

The particle physics cycle

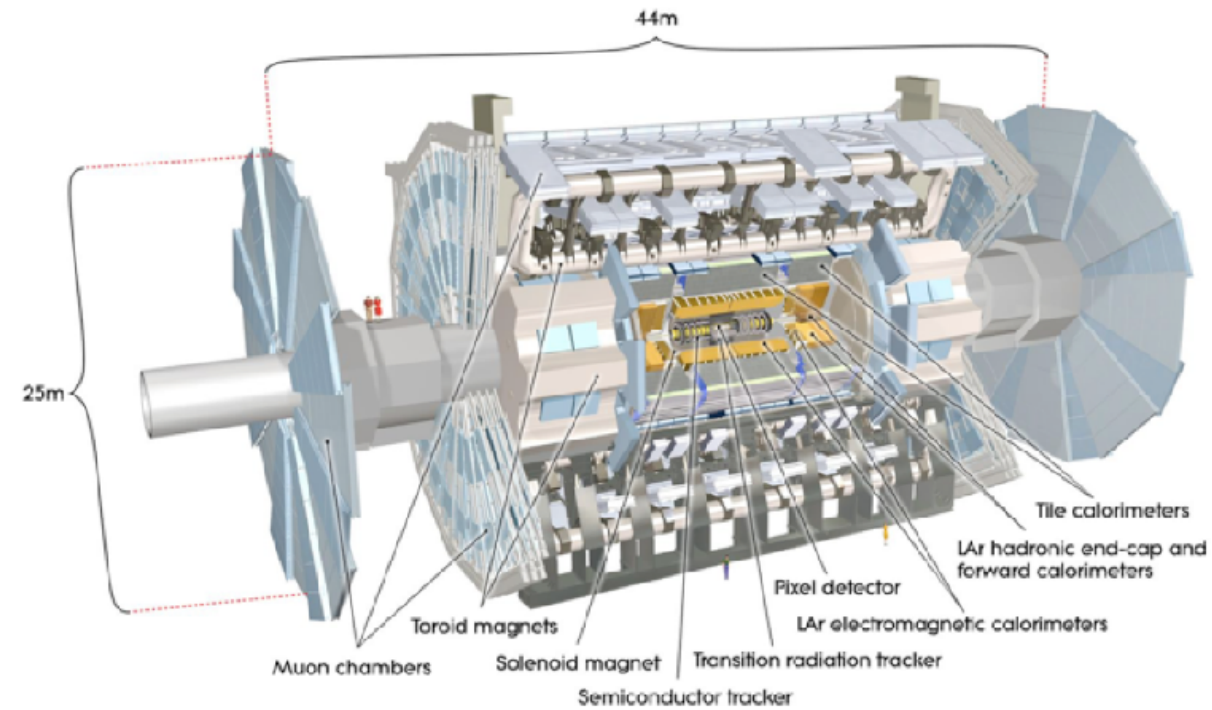








Course outline

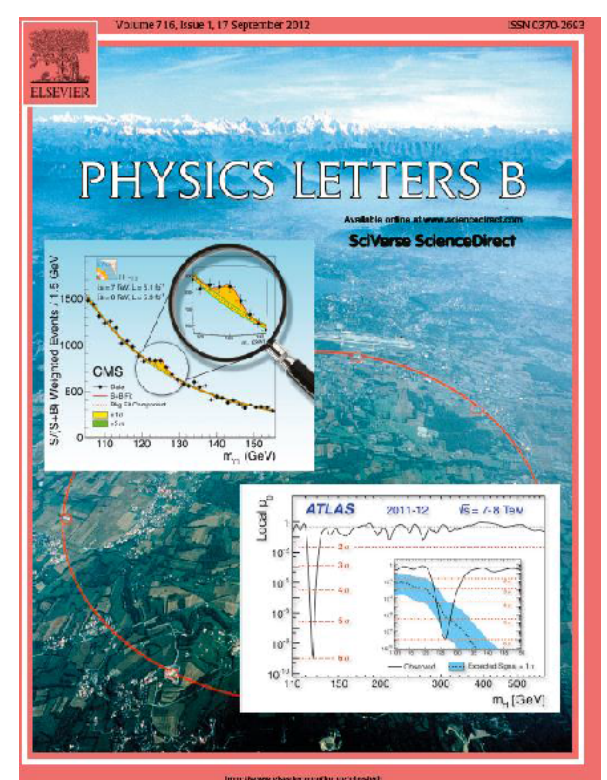


Lecture 1

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

Lecture 2

- How we reconstruct fundamental physics processes from raw detector data

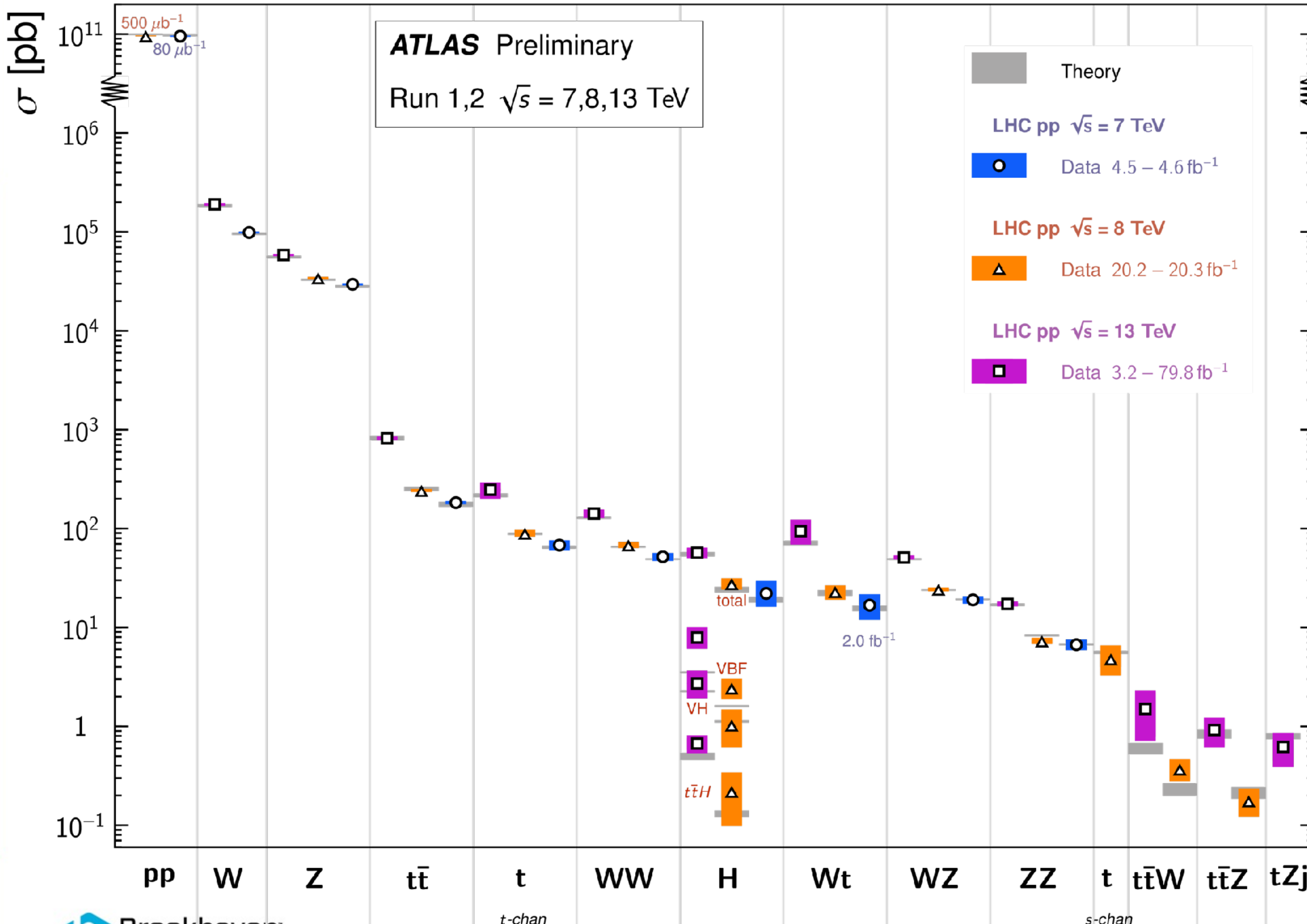


Lecture 3

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

Standard Model Total Production Cross Section Measurements

Status: July 2018



Brookhaven National Laboratory

Measuring cross sections

$$\sigma = \frac{N}{L}$$

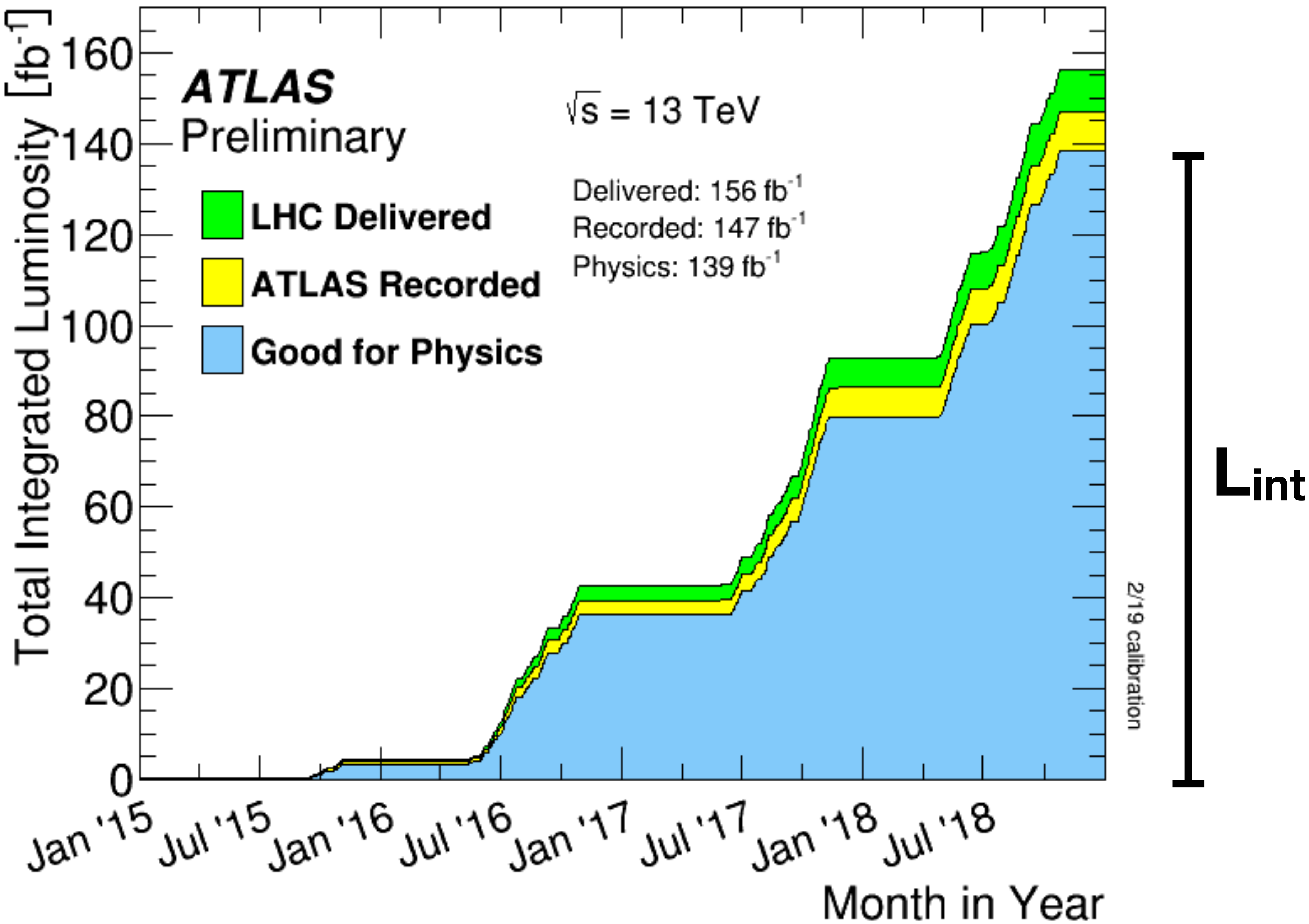
The cross section for a process is defined as the number of events divided by luminosity

Measuring cross sections

$$\sigma = \frac{N}{L_{int}}$$

The cross section for a process is defined as the number of events divided by the integrated luminosity, $\boldsymbol{L_{int}}$, which measures how much data we have collected

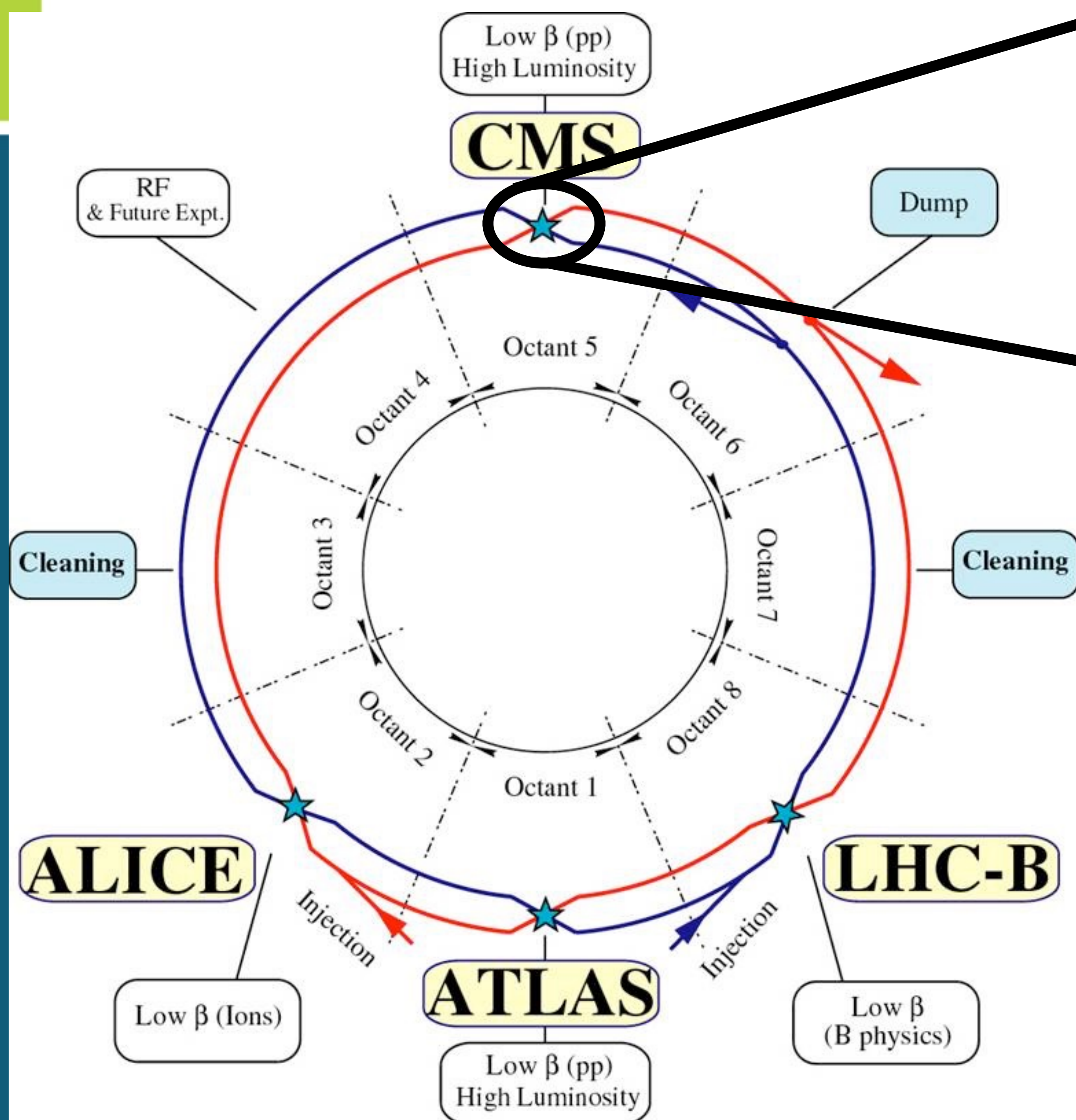
ATLAS Luminosity

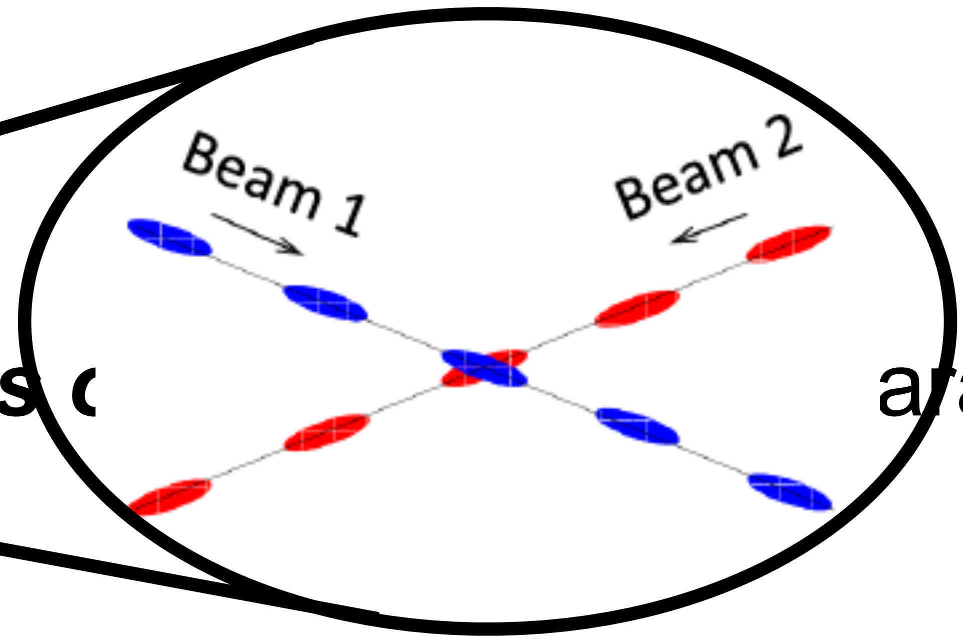


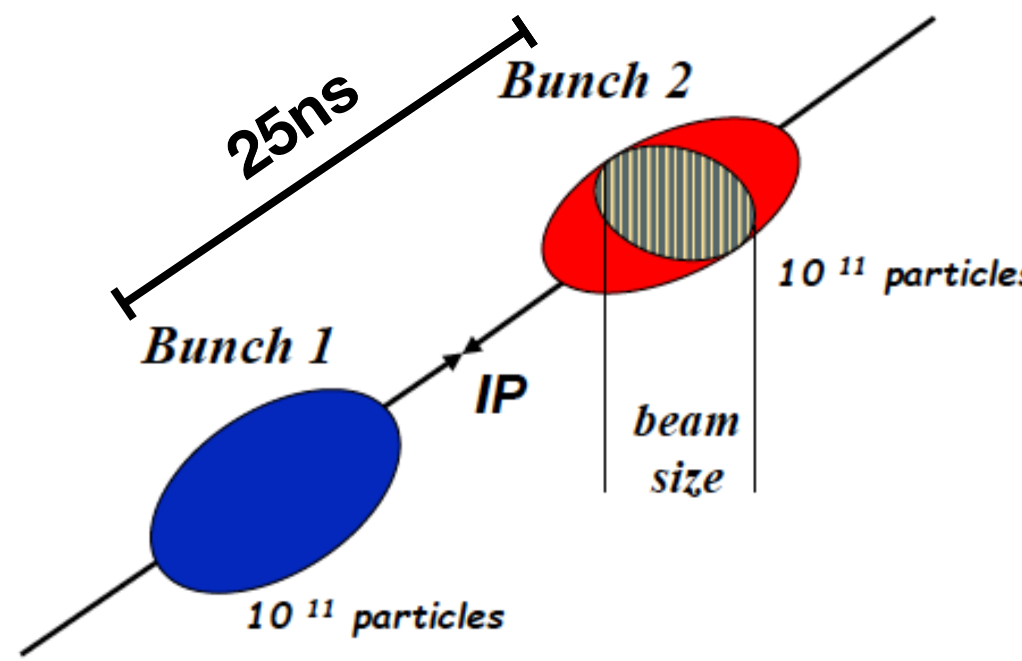
Question: Why does ATLAS record less data than the LHC delivers?
How do we know the integrated luminosity delivered?

LHC collisions

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



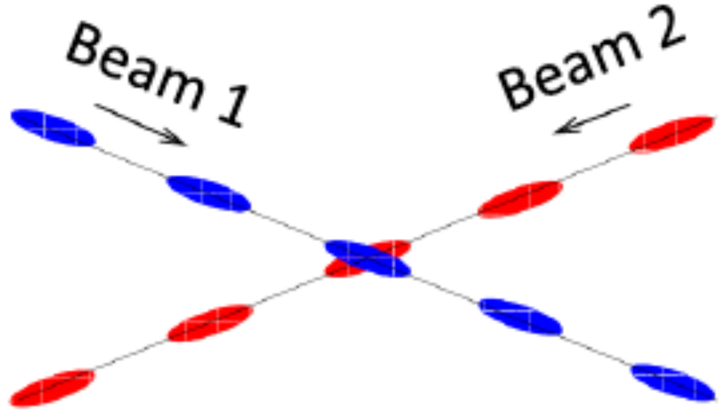


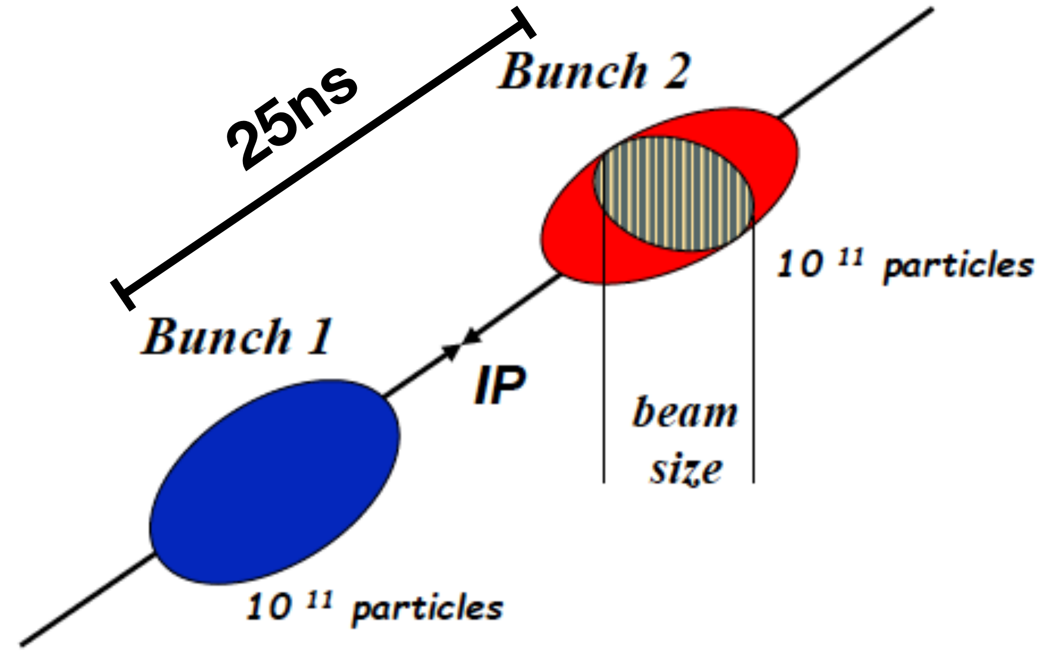


- The LHC accelerates ***bunches of 10^{11} protons*** separated by 25ns gaps

Brookhaven National Laboratory

Measuring Luminosity at the LHC



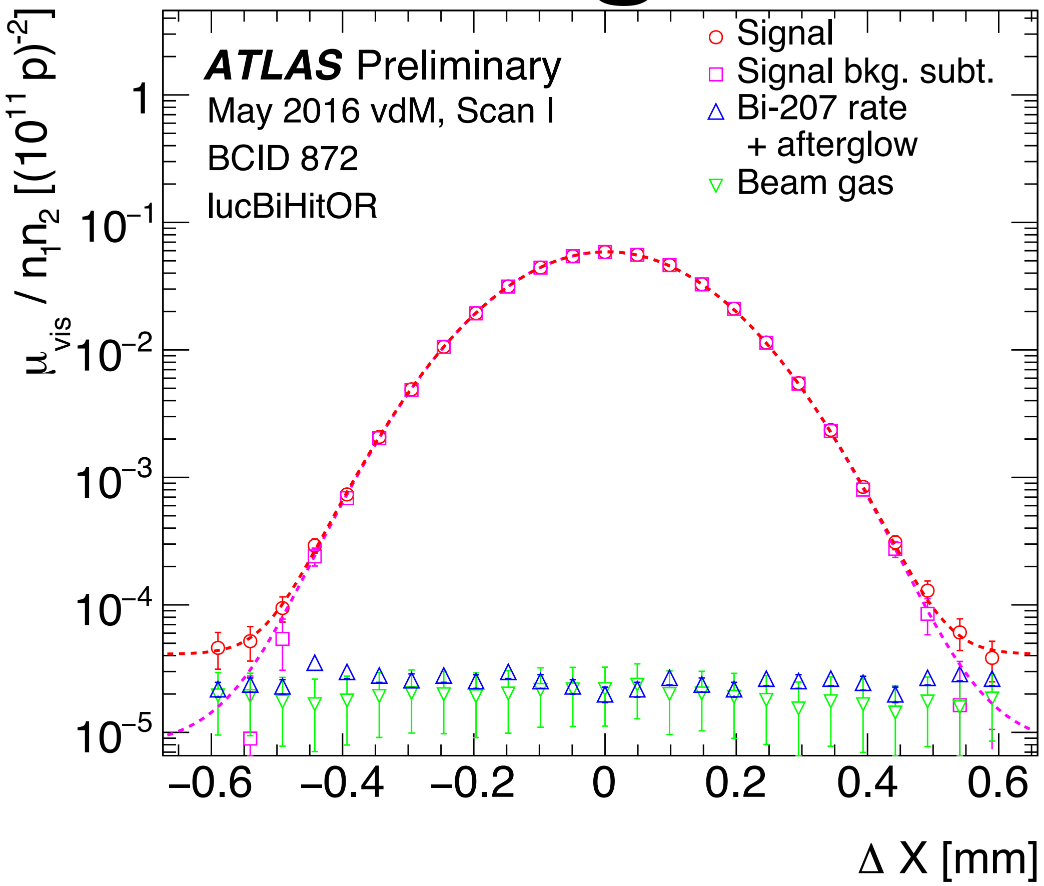


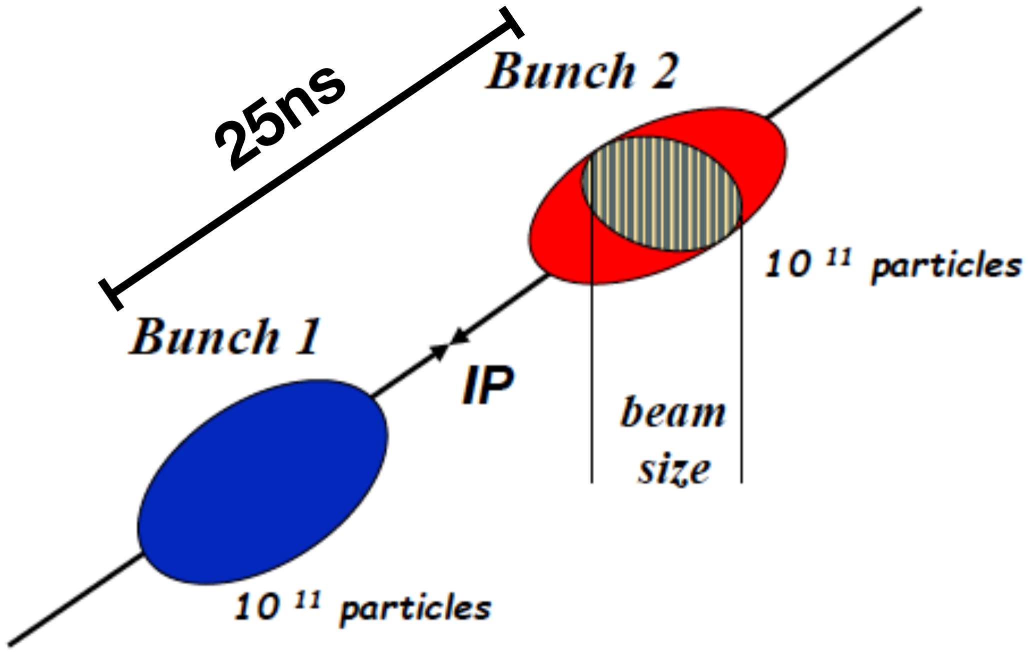
Ingredients for a measurement of the luminosity

- Measuring the **size** of the beams (for a certain LHC configuration)
  - This requires a dedicated measurement where we scan the beams across each other in the **horizontal** and **vertical** directions - a ***van der Meer scan***
- Measuring the beam **currents** in each bunch
  - This is done during collisions, integrating all of the bunch currents and knowing their size, we can calculate the luminosity
- Make **many cross checks** because this is such a crucial measurement

Brookhaven National Laboratory

Measuring Luminosity at the LHC





Ingredients for a measurement of the luminosity

- Measuring the **size** of the beams (for a certain LHC configuration)
  - This requires a dedicated measurement where we scan the beams across each other in the **horizontal** and **vertical** directions - a ***van der Meer scan***
- Measuring the beam **currents** in each bunch
  - This is done during collisions, integrating all of the bunch currents and knowing their size, we can calculate the luminosity
- Make **many cross checks** because this is such a crucial measurement

Measuring cross sections

$$\sigma = \frac{N}{L_{int}}$$

The cross section for a process is defined as the number of events divided by the integrated luminosity, $\boldsymbol{L_{int}}$, which measures how much data we have collected

$$\sigma = \frac{N_{obs}}{A\,.\,\epsilon\,.\,L_{int}}$$

$\boldsymbol{N_{obs}}$ in data needs to be corrected for the detector acceptance, ***A***, for selecting those events. The reconstruction efficiency, ***ε***, is a product of all of the efficiencies that we need to measure and ensure that they are the same in our data and simulation ***- how?***

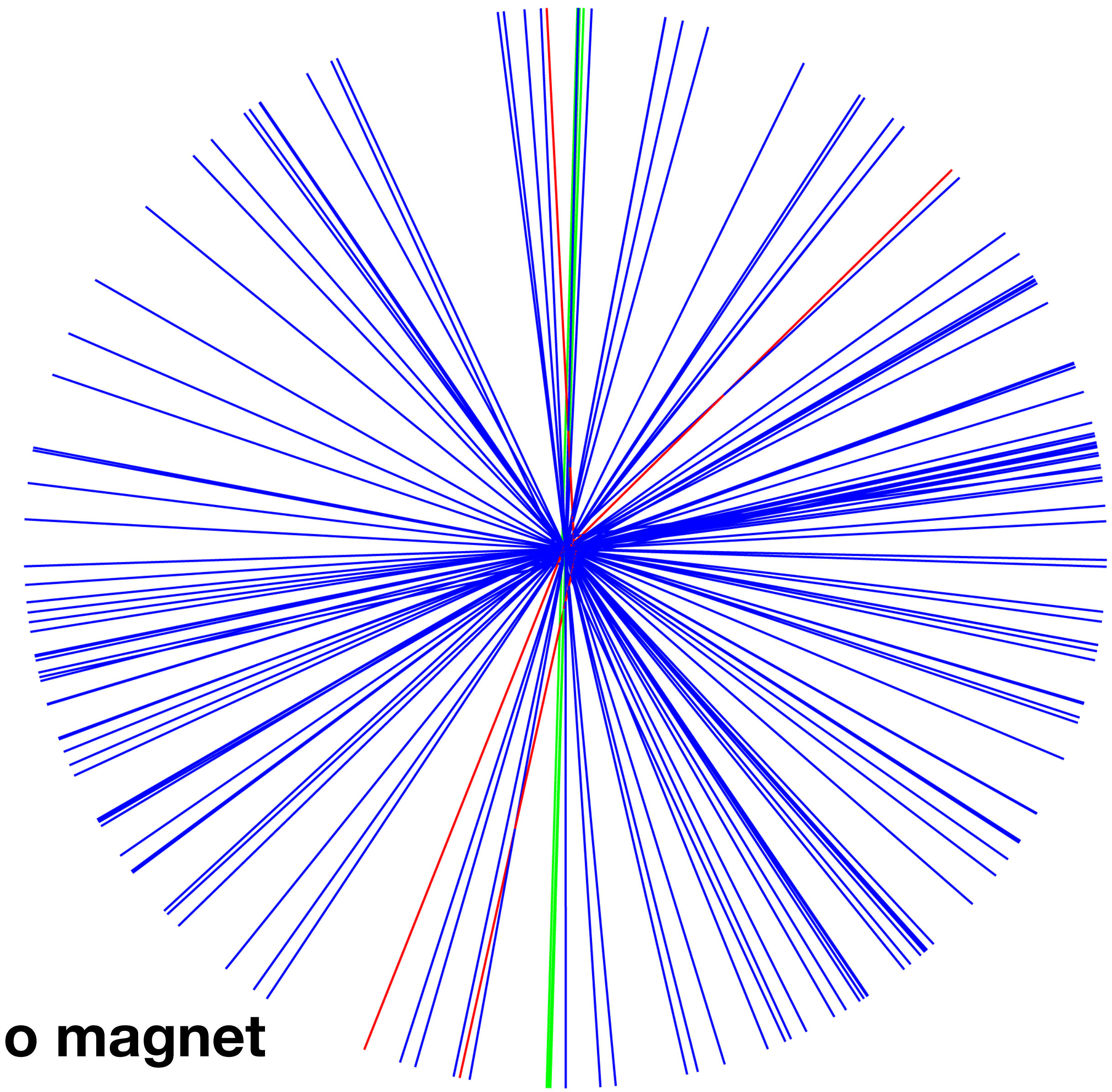
Brookhaven National Laboratory

Before the detector, came the simulation

When designing detectors, we ***simulate detector response*** to physics of interest

Adding a ***solenoid magnet*** makes it possible to measure momentum (and charge) in our tracker by measuring curvature in the transverse plane

Interesting physics is often at ***high momentum***, e.g. four high momentum muon tracks here



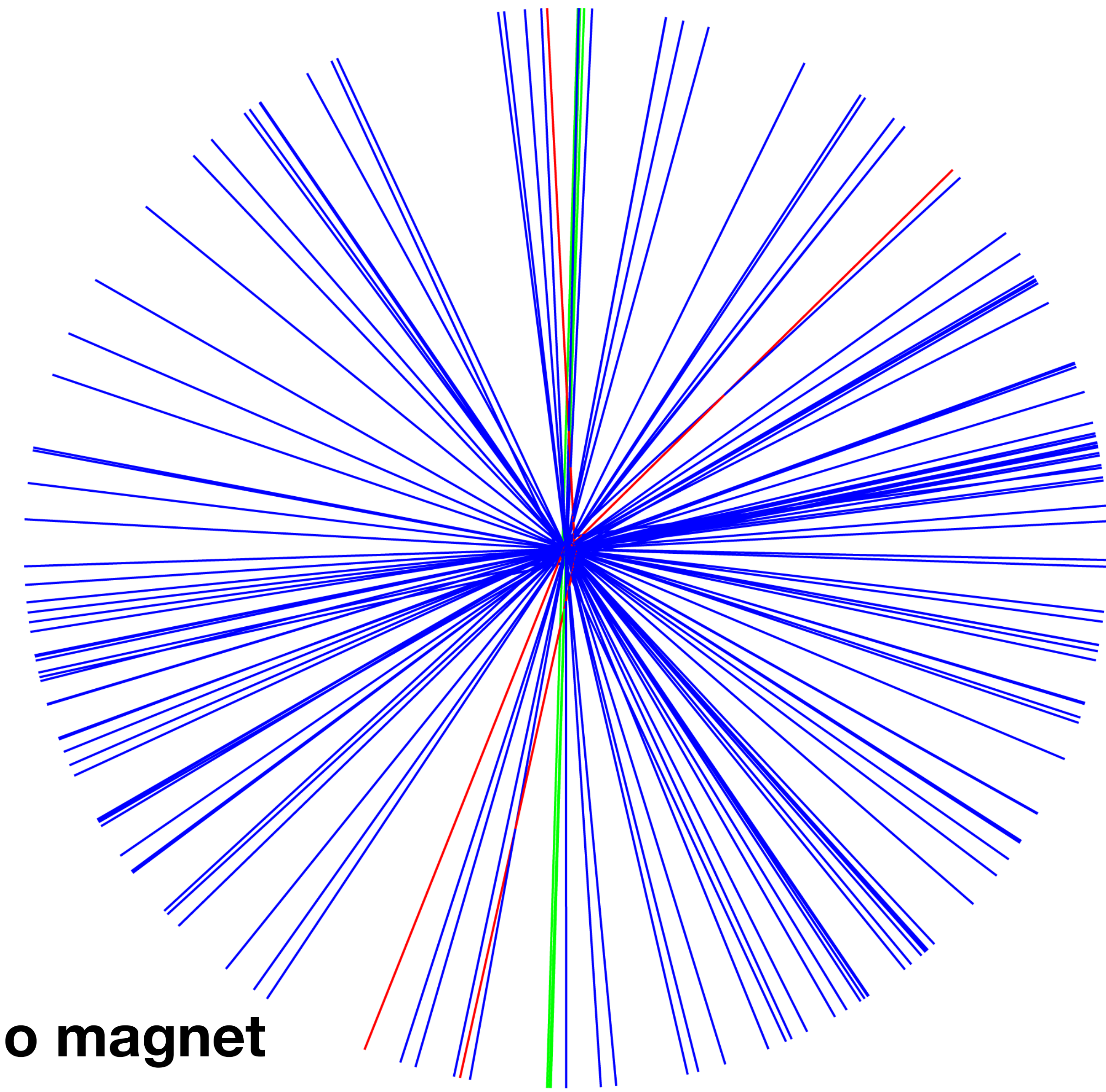
No magnet

Before the detector, came the simulation

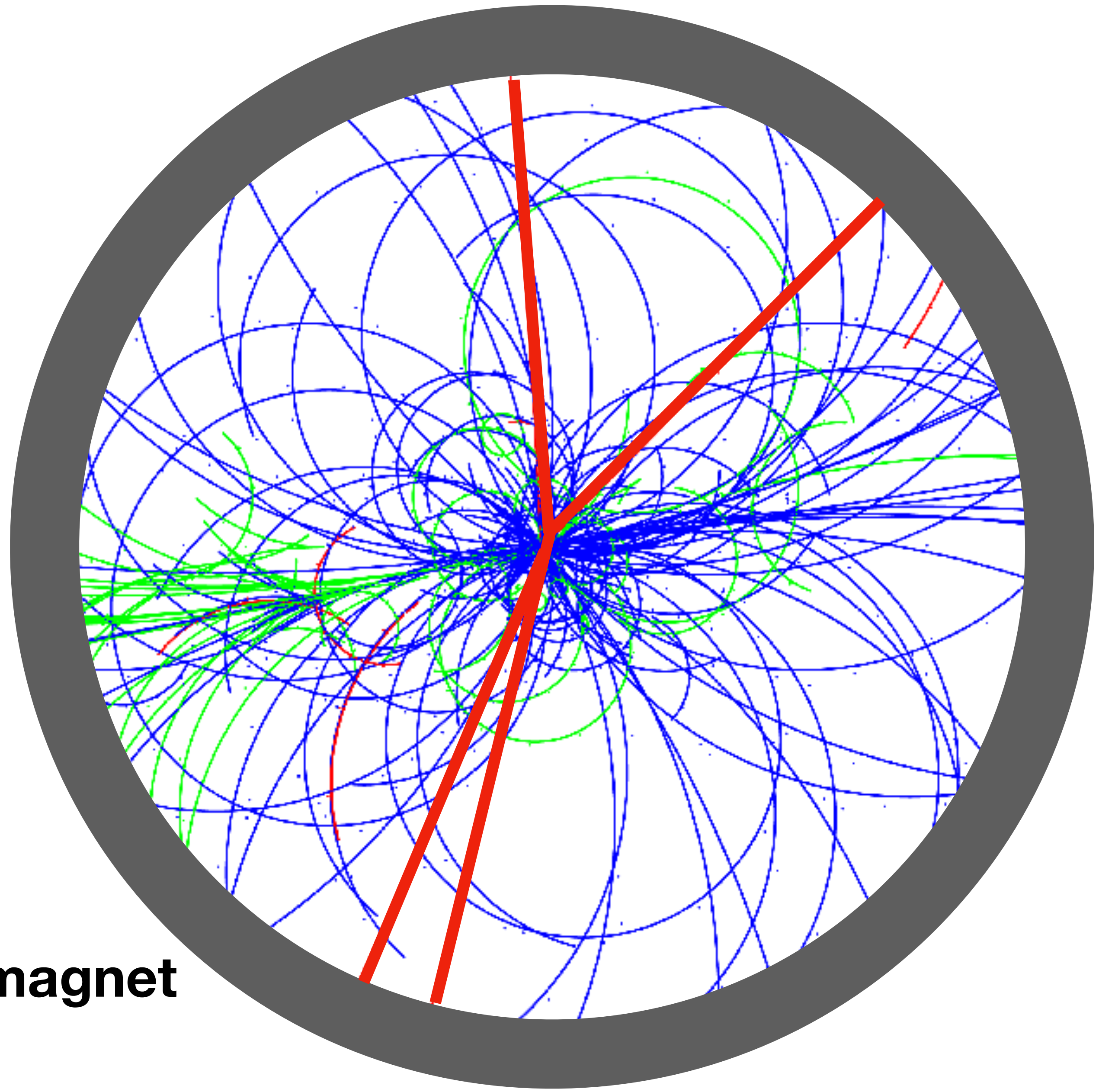
When designing detectors, we ***simulate detector response*** to physics of interest

Adding a ***solenoid magnet*** makes it possible to measure momentum (and charge) in our tracker by measuring curvature in the transverse plane

Interesting physics is often at ***high momentum***, e.g. four high momentum muon tracks here



No magnet



+ magnet

Brookhaven National Laboratory

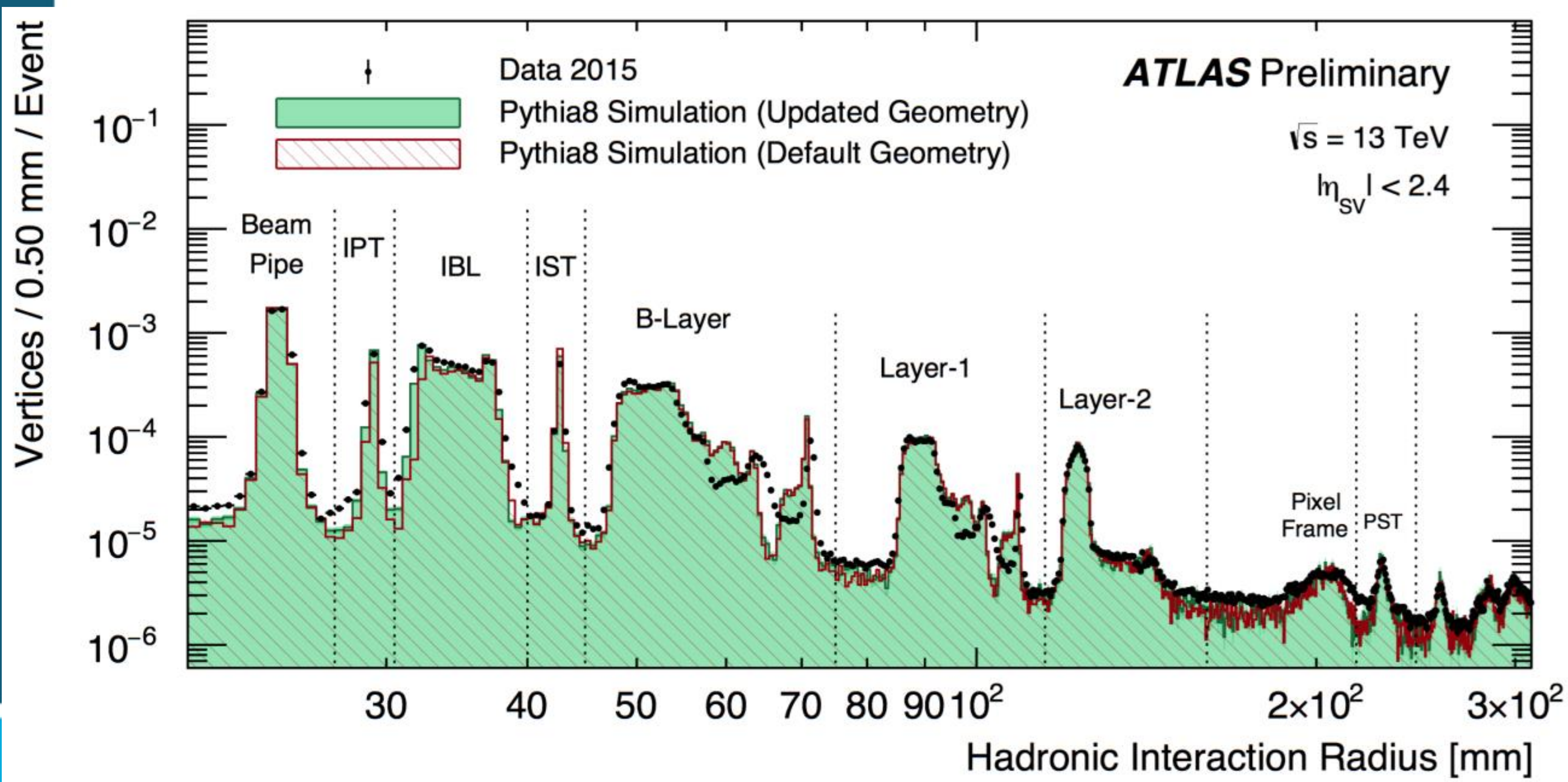
Simulation and understanding detectors

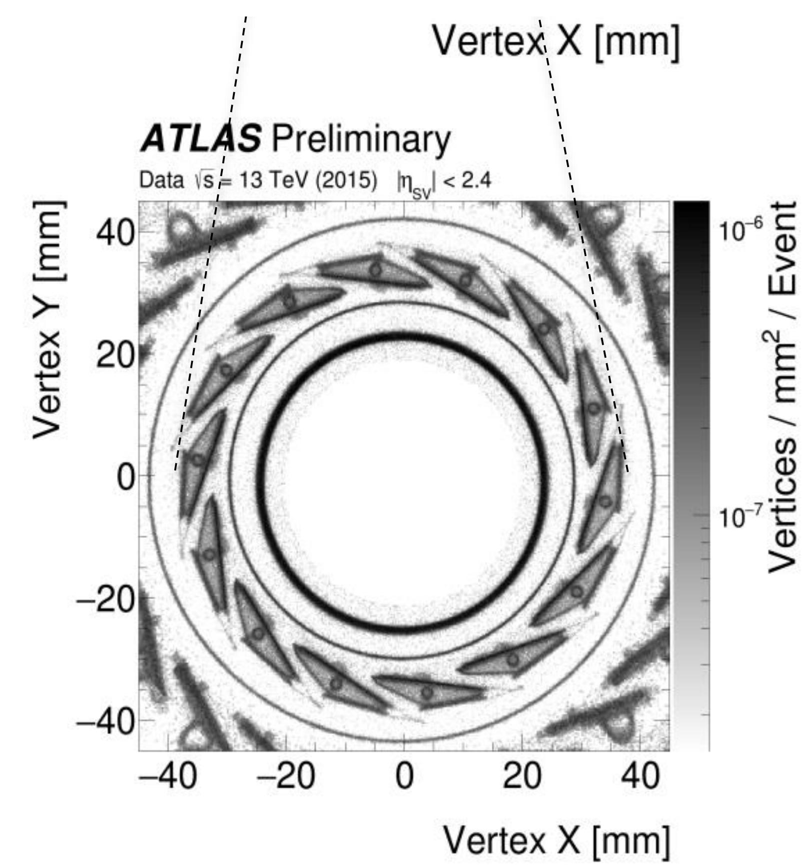
We use **simulations** to model the detector as **accurately** and **precisely** as possible

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

Real vs perfect tracking detectors

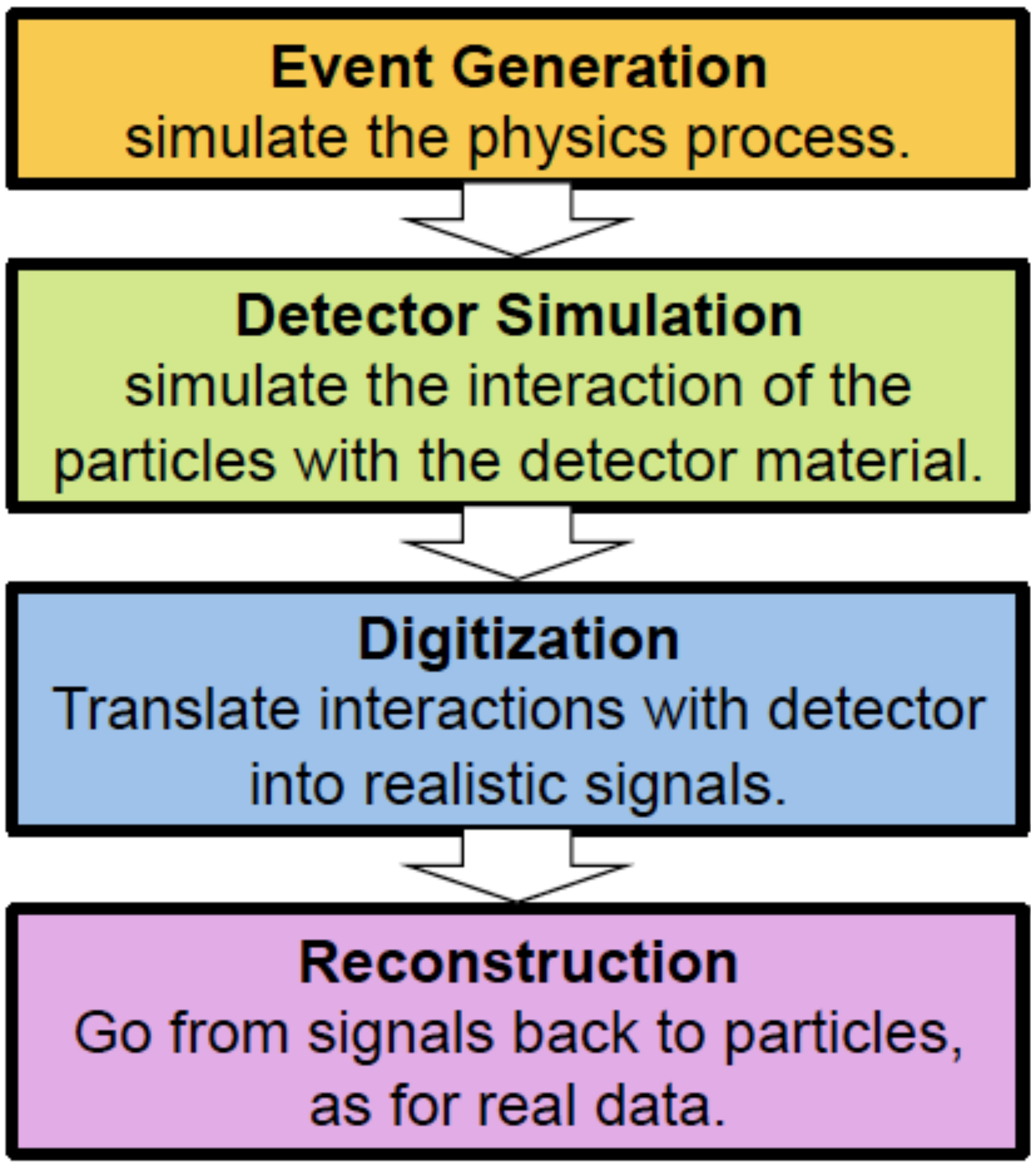
The perfect tracking detector

- is constructed from zero mass material
- has no noise
- is 100% efficiency
- 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

Simulation chain



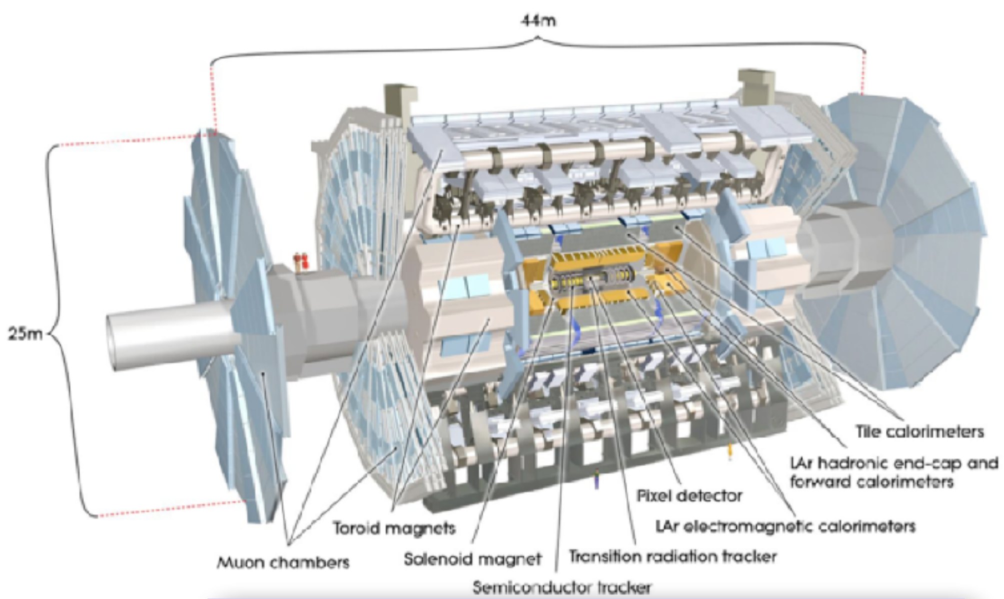
→ **Course on Detectors**

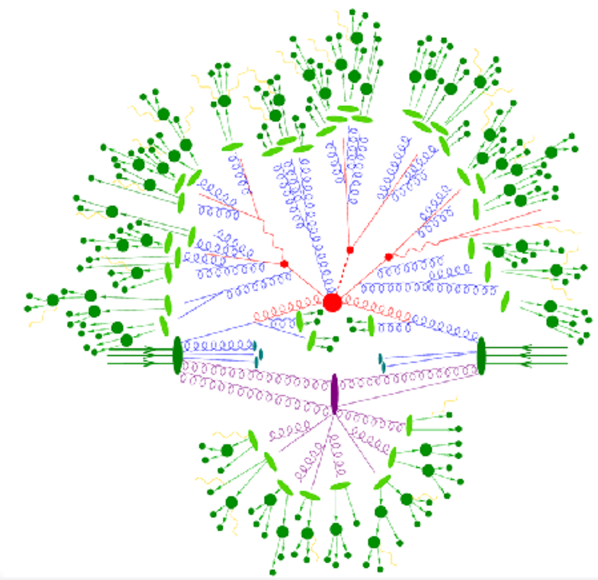
→ **Course on Electronics /Trigger/DAQ**

→ **This Course**

Brookhaven National Laboratory

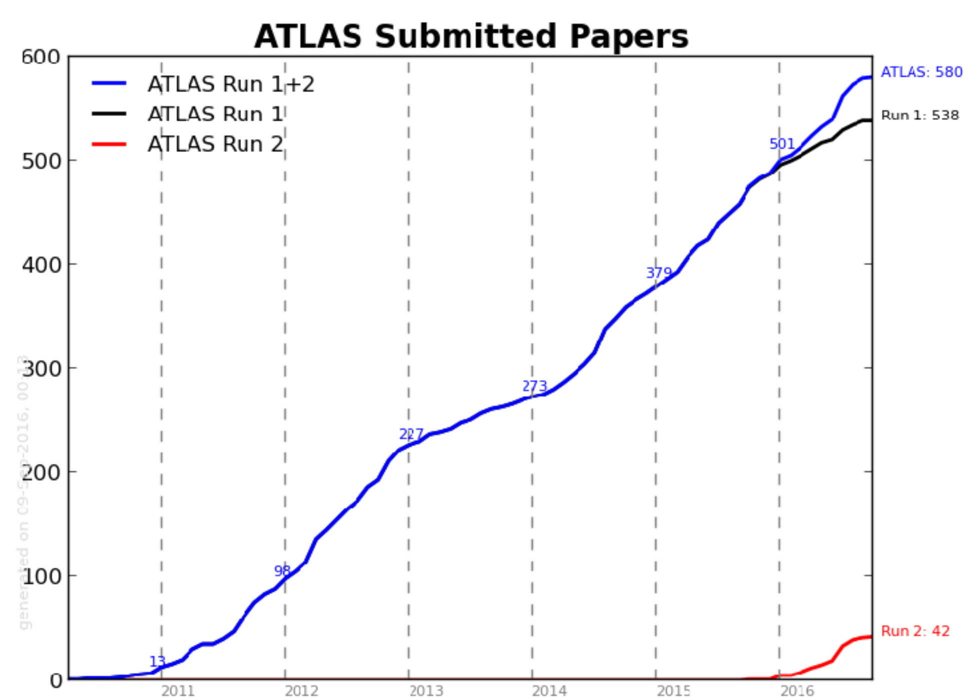
Exabyte-scale physics analysis





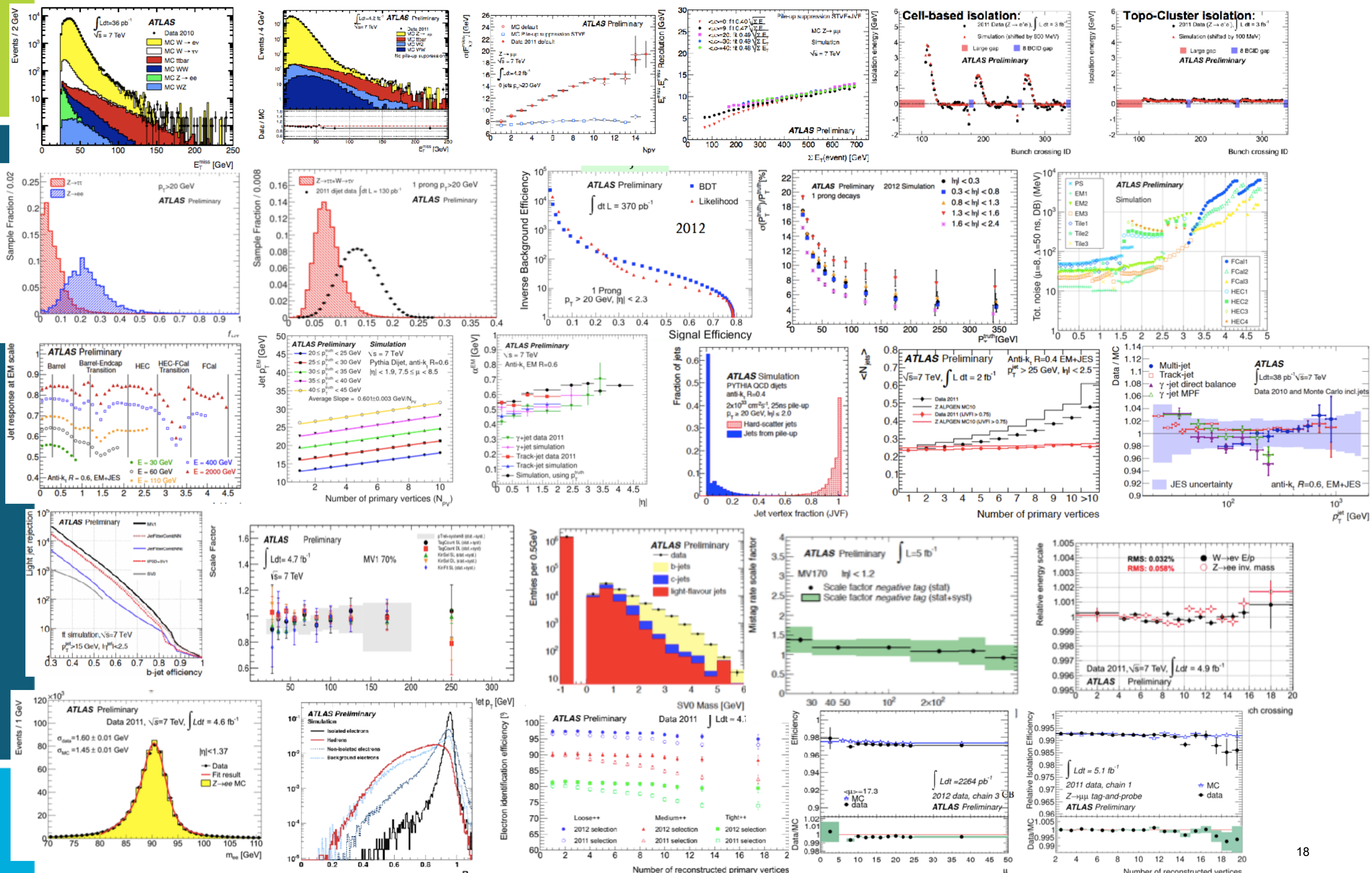
Exabytes of Data

Exabytes of Simulation

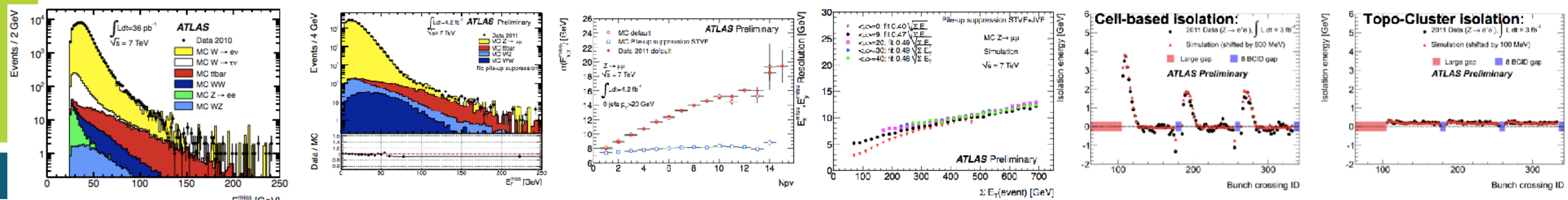


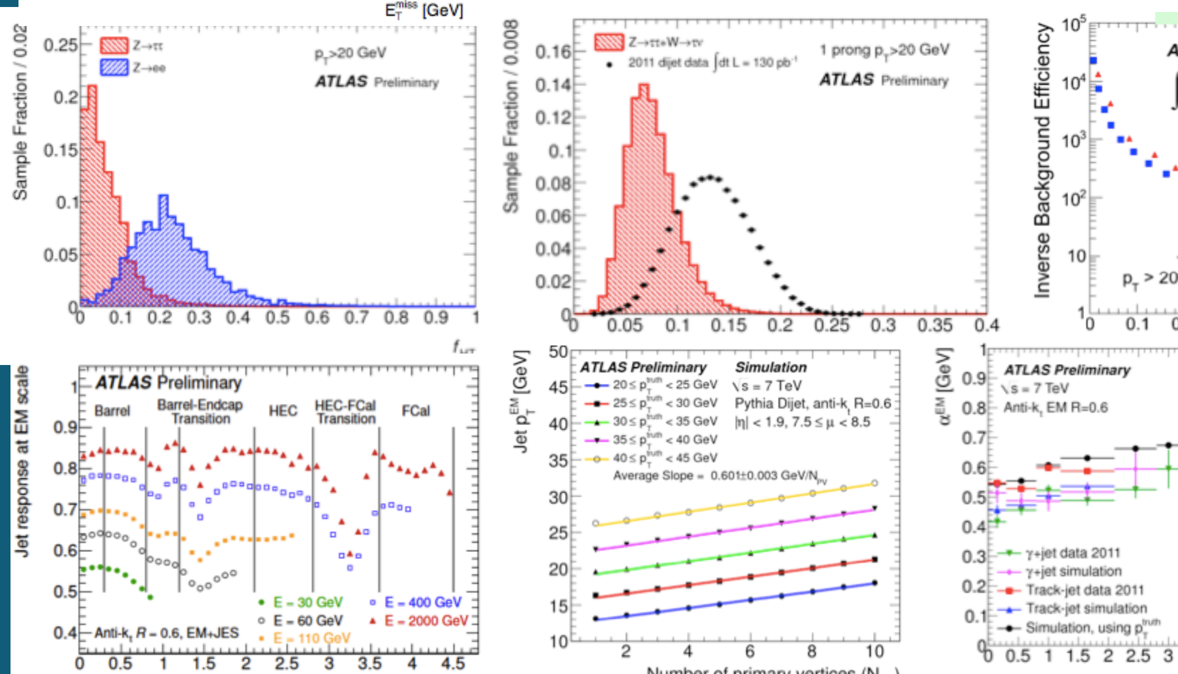
Publish!

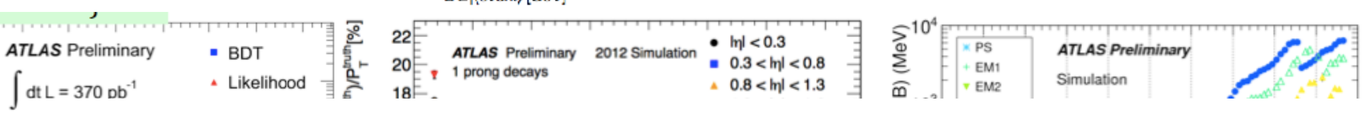
Ingredients to the ATLAS physics program



Ingredients to the ATLAS physics program





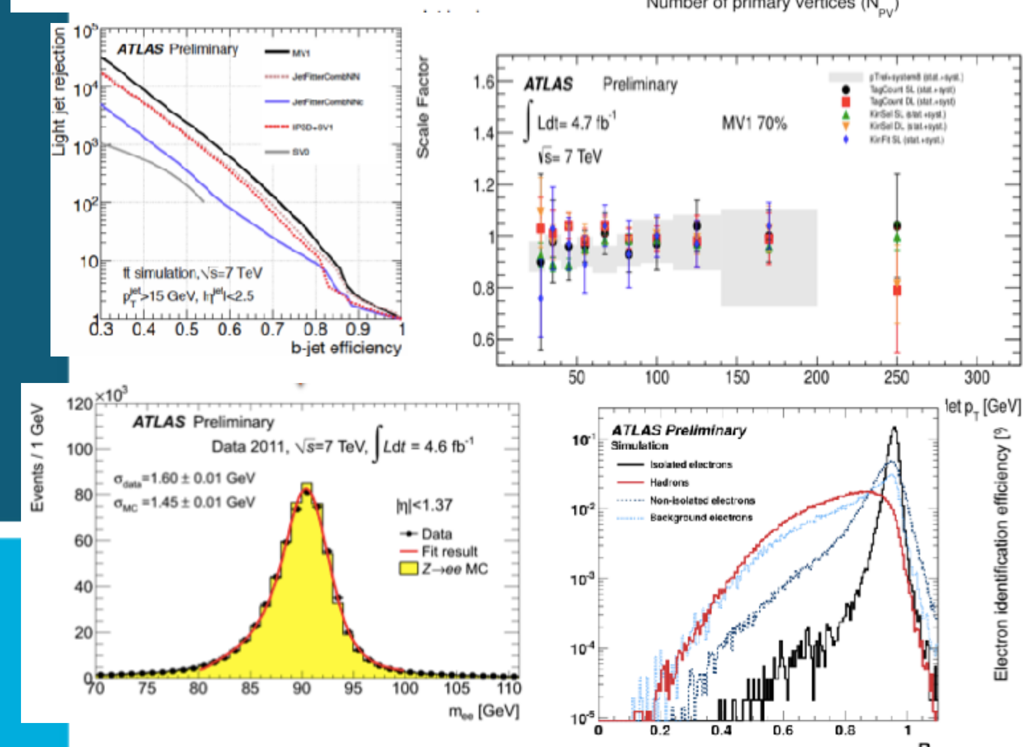


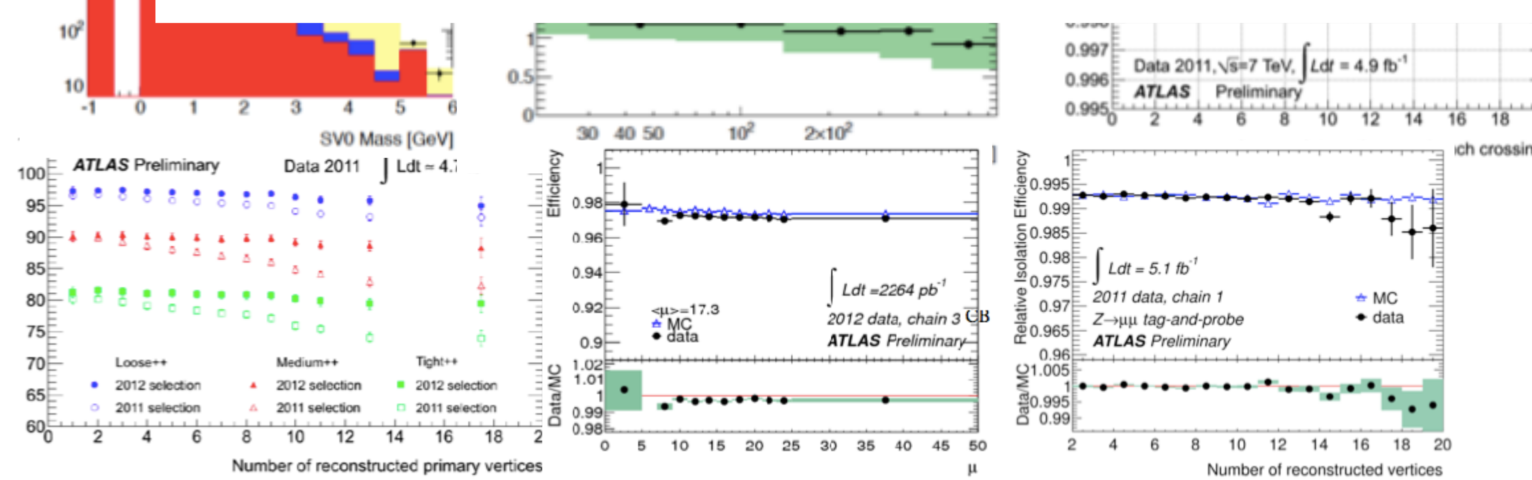
Compare data and simulation in every conceivable way !

We need to check that the simulation does a good job of describing the data

It's more important to understand how particles interact with our detector than to understand the underlying physics model

WHY?





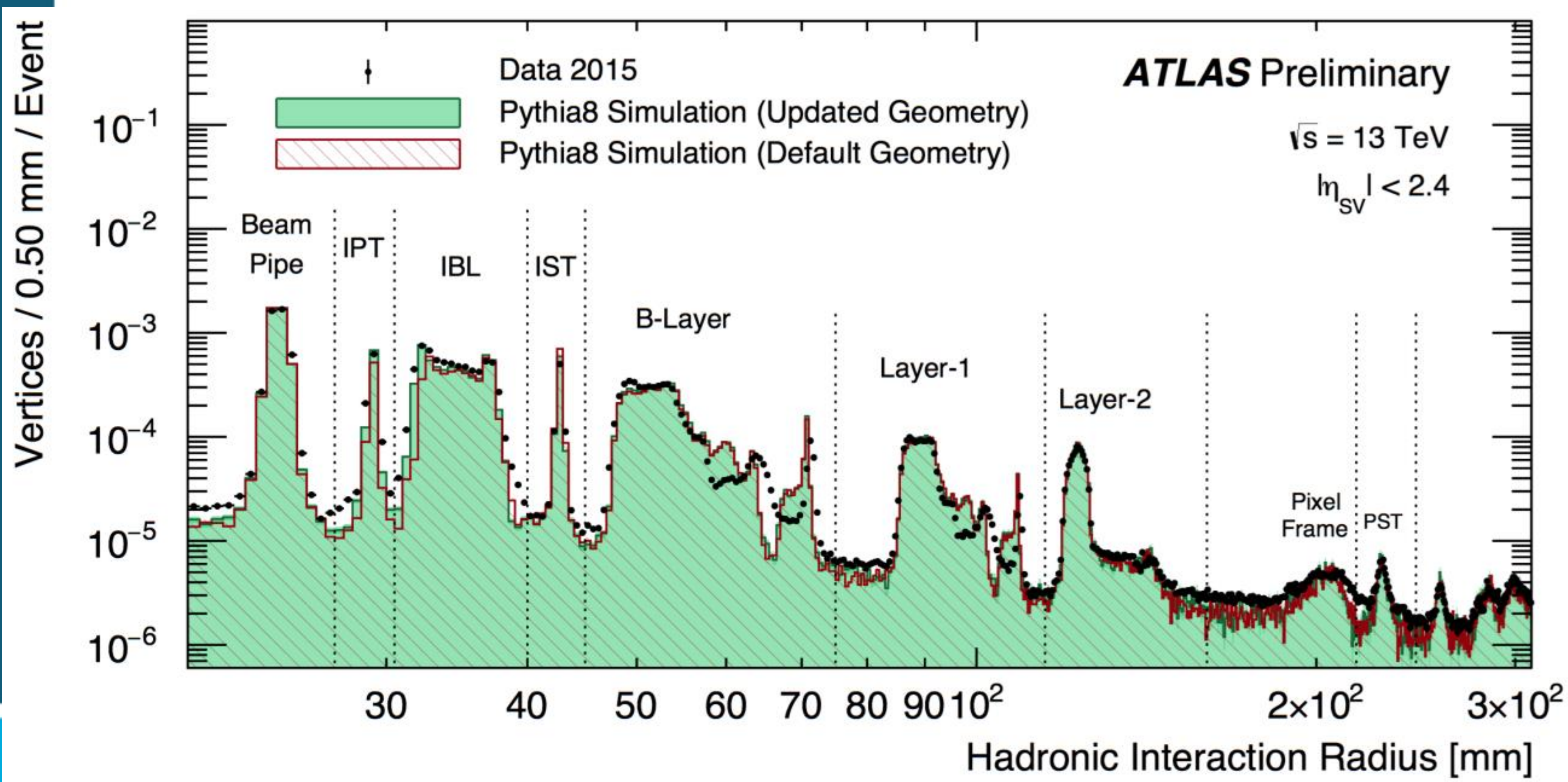
Simulation and understanding detectors

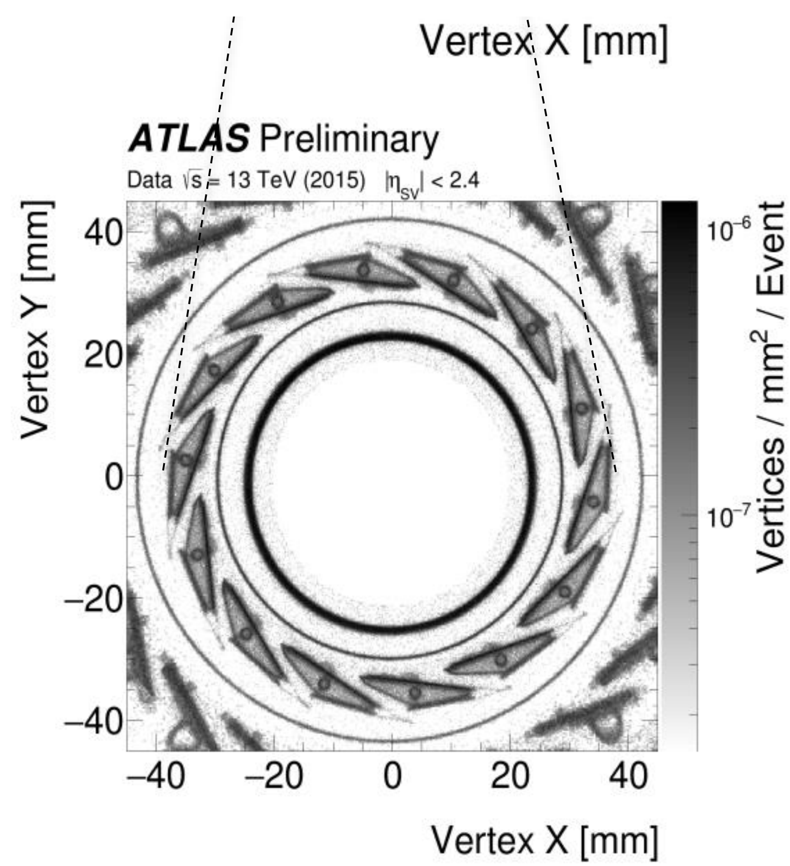
We use **simulations** to model the detector as **accurately** and **precisely** as possible

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***





Measuring cross sections

$$\sigma = \frac{N}{L_{int}}$$

The cross section for a process is defined as the number of events divided by the integrated luminosity, ***L_{int}***, which measures how much data we have collected

$$\sigma = \frac{N_{obs}}{A\,.\,\epsilon\,.\,L_{int}}$$

N_{obs} in data needs to be corrected for the detector acceptance, ***A***, for selecting those events. The reconstruction efficiency, ***ε***, is a product of all of the efficiencies that we need to measure and ensure that they are the same in our data and simulation

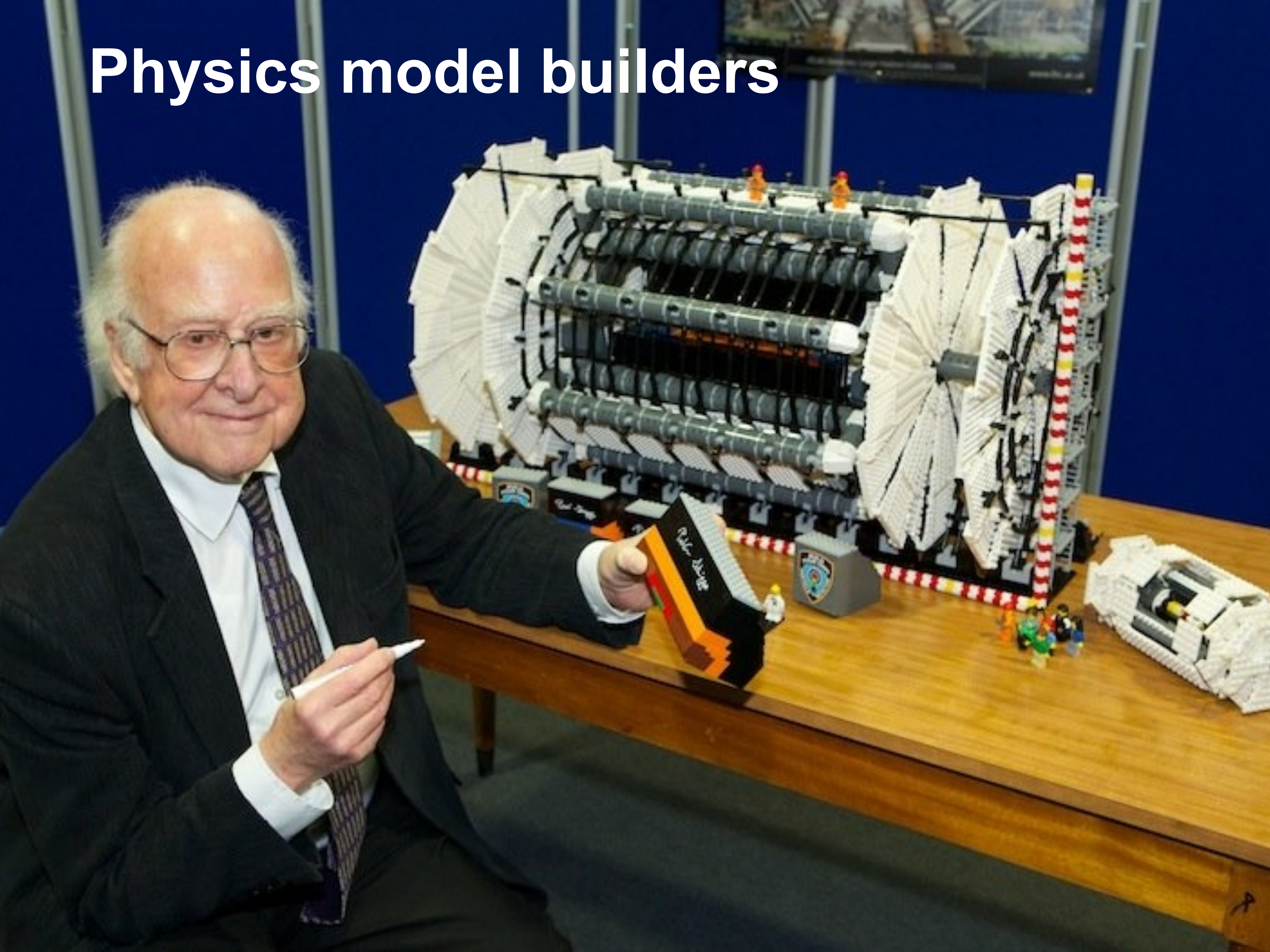
$$\sigma = \frac{N_{obs} - N_{bkg}}{A\,.\,\epsilon\,.\,L_{int}}$$

Finally, in reality we will have some background to our signal and we need to subtract those events that are not part of our signal process

Now we can compare this to the theoretical cross section!

Brookhaven National Laboratory

Physics model builders



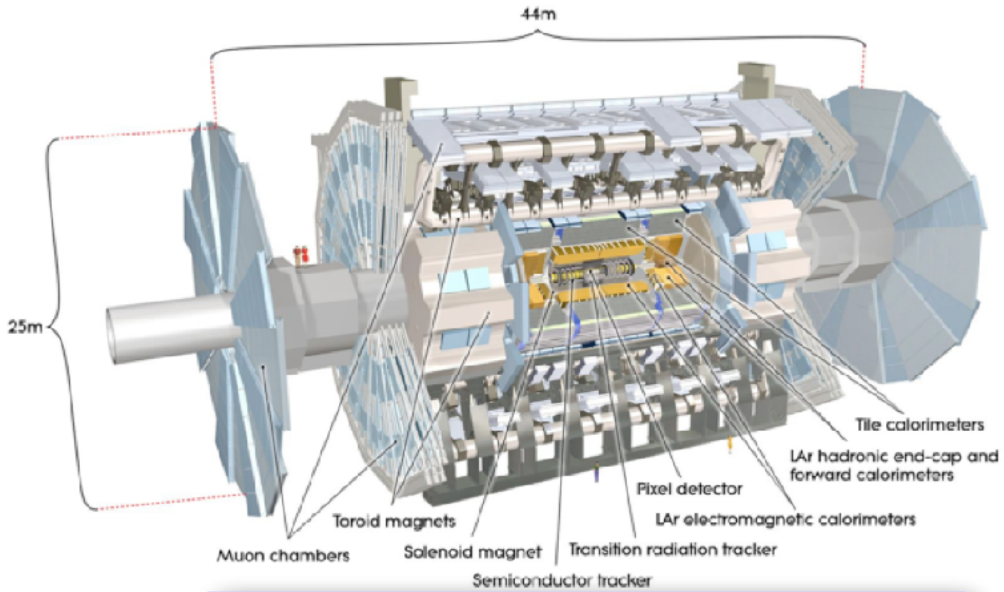
Physics event generators

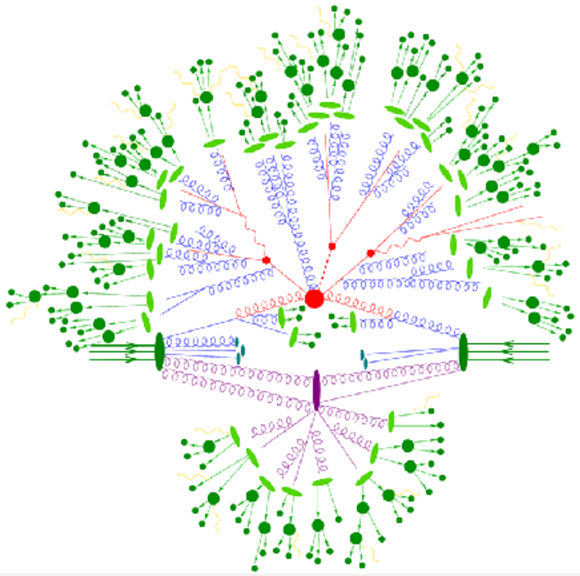
There are lots of different physics models implemented in physics event generators, depending on the type of physics that you're interested in

We want to see if reality looks like theory (and which one !)

Brookhaven National Laboratory

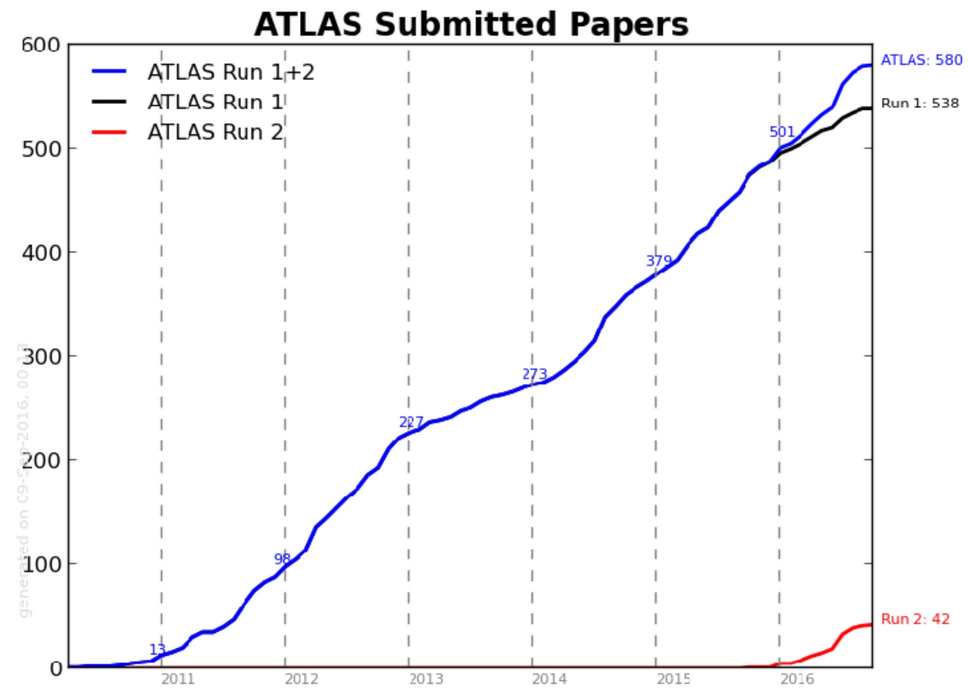
Exabyte-scale physics analysis





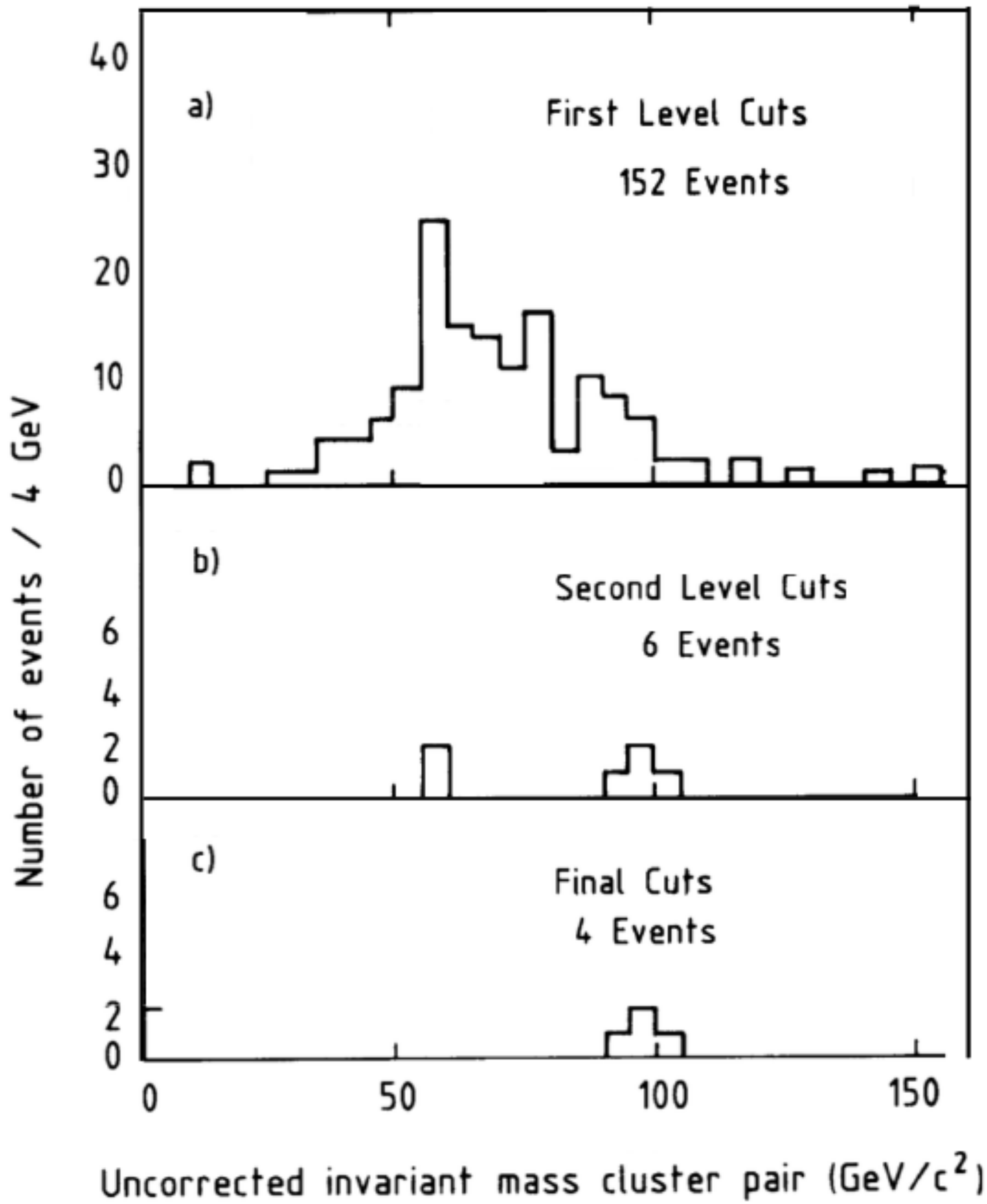
Exabytes of Data

Exabytes of Simulation



Publish!

First - measuring the Z boson



Z->ee in UA1

Two EM clusters with E_T>25GeV.

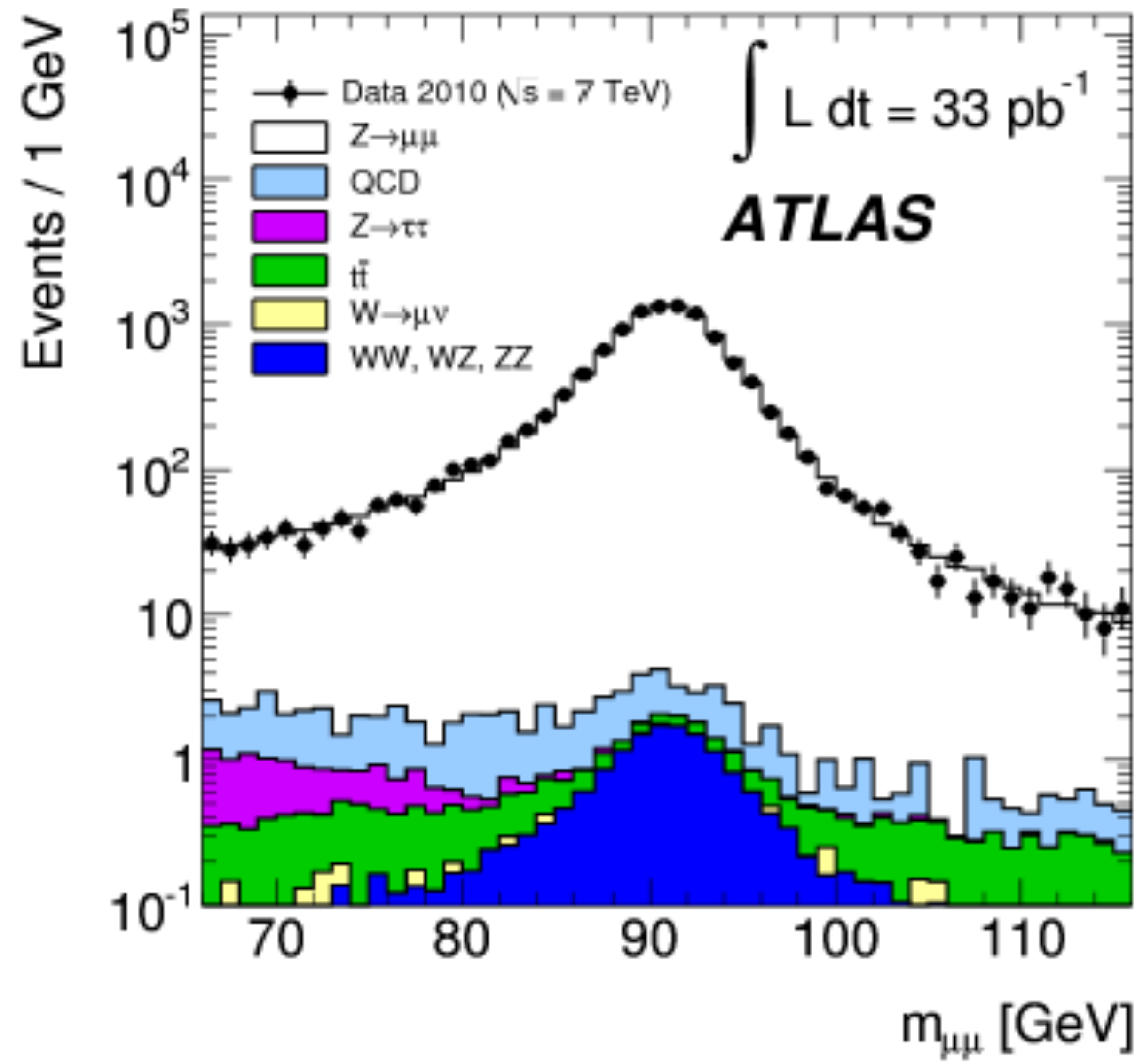
As above plus a track with p_T>7GeV pointing to the cluster. Hadronic and track isolation requirements applied.

A second custer has also an isolated track.

Brookhaven National Laboratory

Measuring the Z boson at ATLAS

$$\sigma = \frac{N_{obs} - N_{bkg}}{A\,.\,\epsilon\,.\,L_{int}}$$



Select events with (here) two muons

Question: what other selections can we apply to the muons?

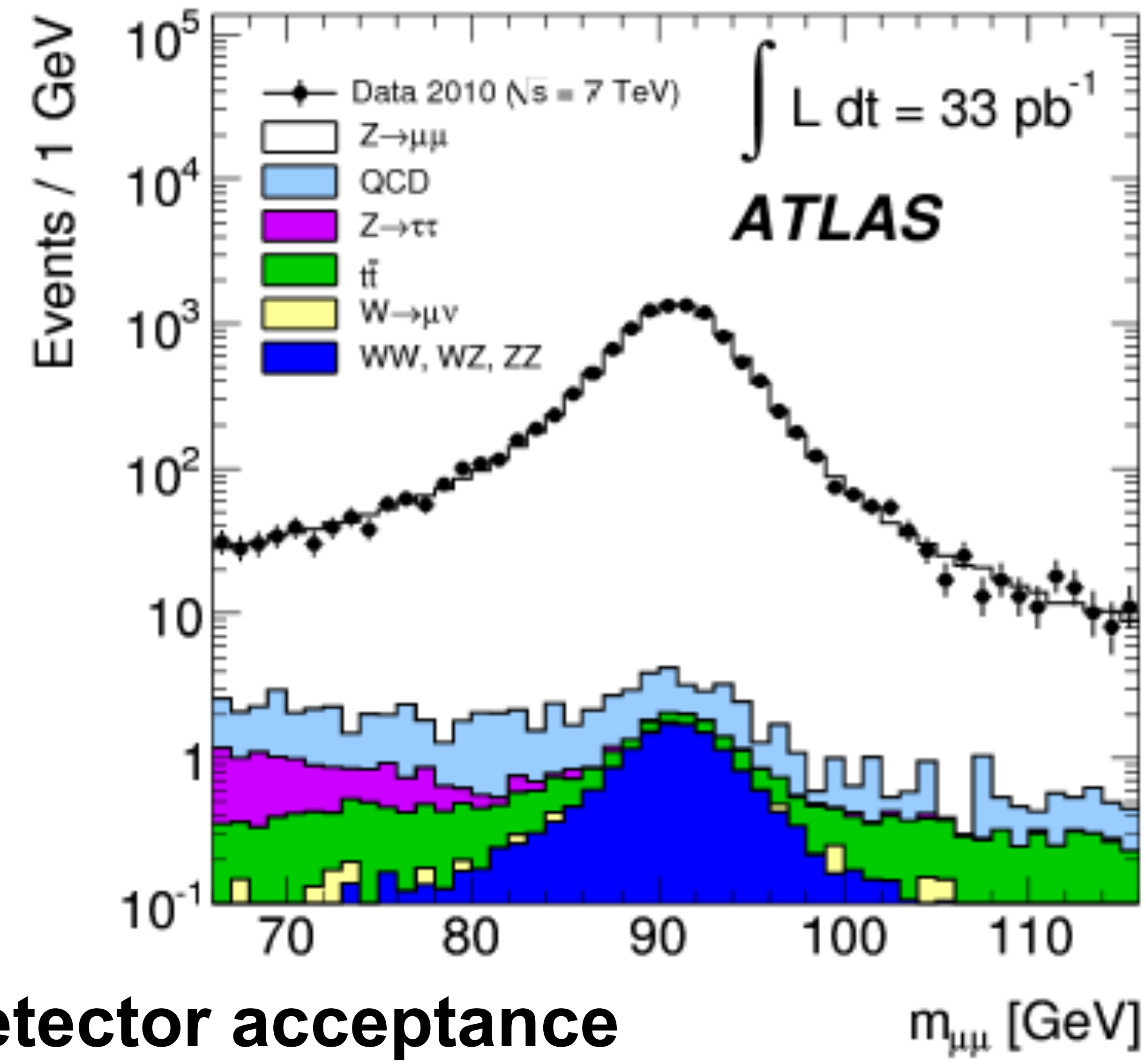
Here I have only considered events with two muons

Question: is this the cross section for Z boson production?

Brookhaven National Laboratory

Measuring the Z boson at ATLAS

$$\sigma = \frac{N_{obs} - N_{bkg}}{A \cdot \epsilon \cdot L_{int}}$$



Backgrounds are small but still need to be measured and subtracted

We will quote a fiducial cross section corresponding to good detector acceptance

After making the event selection, applying the same selection to all of the simulations of background processes, and measuring my acceptance and efficiencies (and knowing the luminosity) - am I done?

Measuring the Z boson at ATLAS

$$\sigma = \frac{N_{obs} - N_{bkg}}{A \, . \, \epsilon \, . \, L_{int}}$$

Table 5: Measured fiducial $Z \rightarrow \ell^+\ell^-$ differential and integrated cross sections for electron and muon channels.

		$Z \rightarrow e^+e^-$				$Z \rightarrow \mu^+\mu^-$			
$\lvert y_{\ell\ell}\rvert^{min}$	$\lvert y_{\ell\ell}\rvert^{max}$	$d\sigma/d\lvert y_{\ell\ell}\rvert$ [pb]	$\delta\sigma_{stat}$ [pb]	$\delta\sigma_{syst}$ [pb]	$\delta\sigma_{lumi}$ [pb]	$d\sigma/d\lvert y_{\ell\ell}\rvert$ [pb]	$\delta\sigma_{stat}$ [pb]	$\delta\sigma_{syst}$ [pb]	$\delta\sigma_{lumi}$ [pb]
0.0	0.5	99.9	2.5	1.6	1.9	105.2	2.4	1.1	2.0
0.5	1.0	100.3	2.7	1.6	1.9	101.9	2.3	1.0	1.9
1.0	1.5	89.2	2.7	1.4	1.7	89.8	2.1	0.8	1.7
1.5	2.0	59.6	2.4	1.2	1.1	61.0	1.8	0.6	1.1
2.0	2.5	19.6	1.3	0.7	0.4	20.3	1.2	0.2	0.4
0.0	2.5	369.0	5.3	4.7	6.9	377.9	4.4	3.4	7.1

No ! You would like to publish with the smallest ***uncertainties*** possible

Every ingredient to the analysis comes with an uncertainty

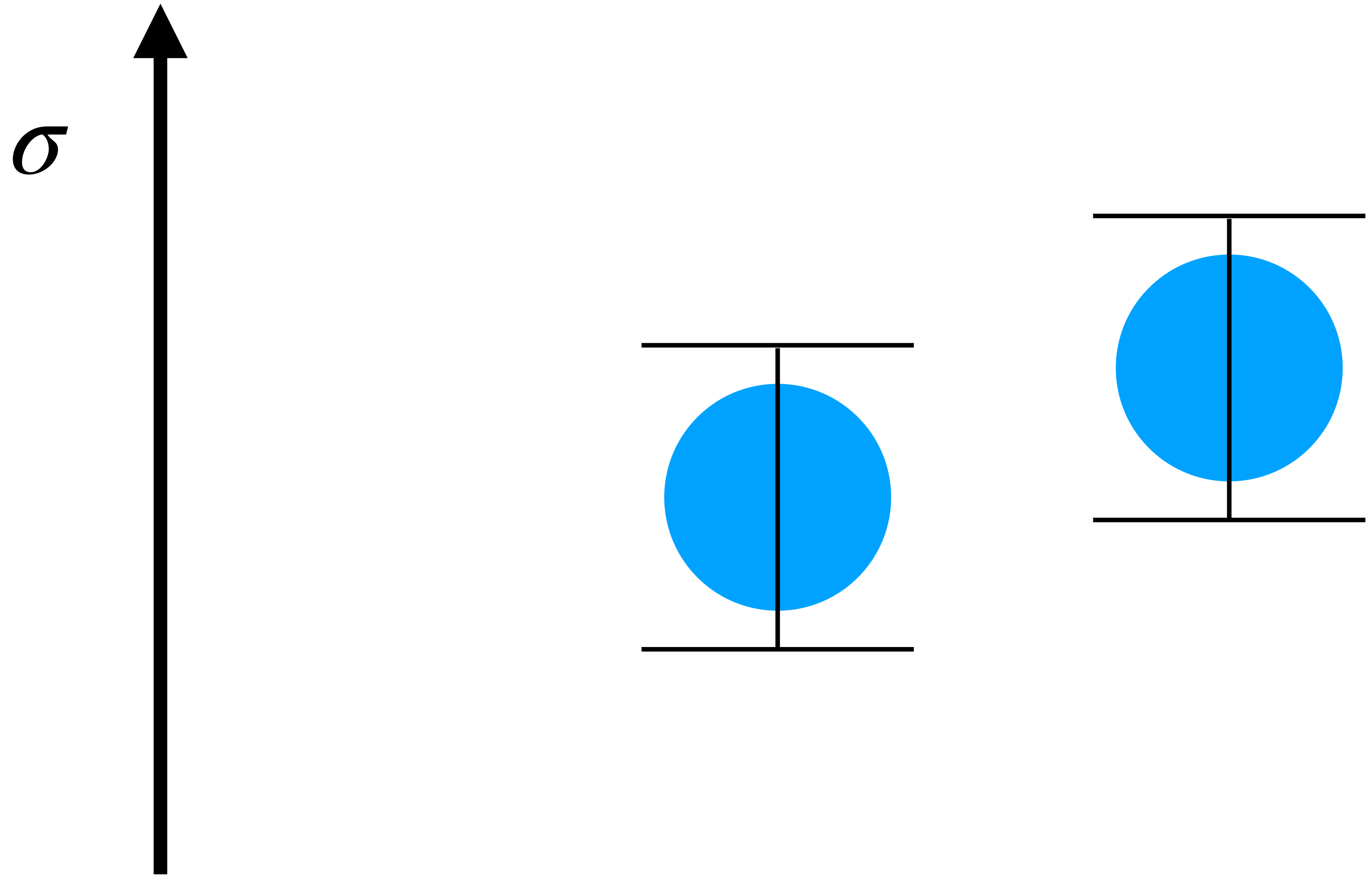
N_{obs} has a ***statistical*** uncertainty

N_{bkg} is typically composed of several sources (different physics processes) with corresponding ***statistical*** and ***systematic*** contributions to the final uncertainty

A and particularly **ε** have many ***systematic*** components stemming from each reconstruction algorithm that we employed

Finally, L_{int} also has an uncertainty that dictates how well we know the absolute scale of the measurement - a ***normalisation*** uncertainty

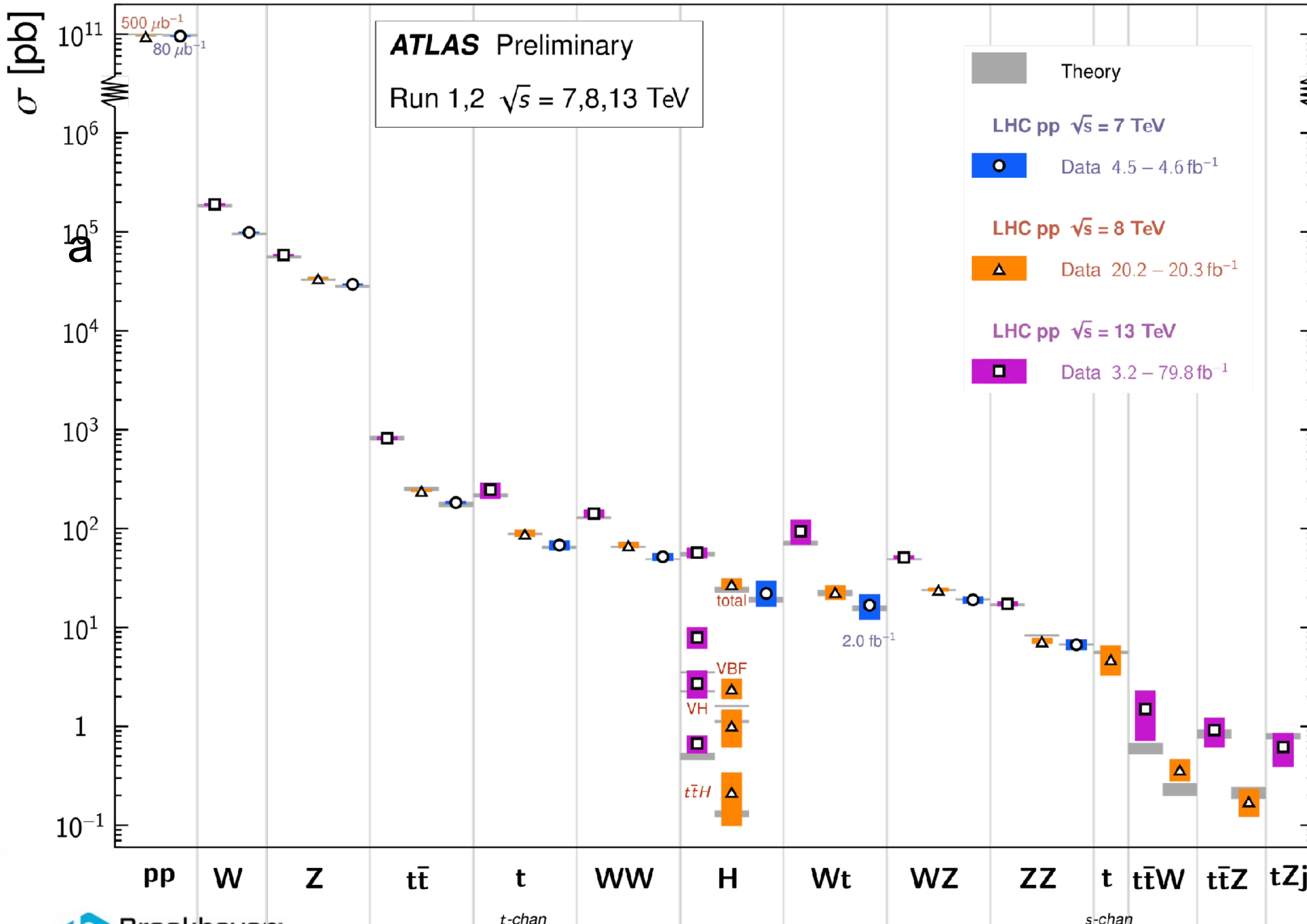
Uncertain about uncertainties



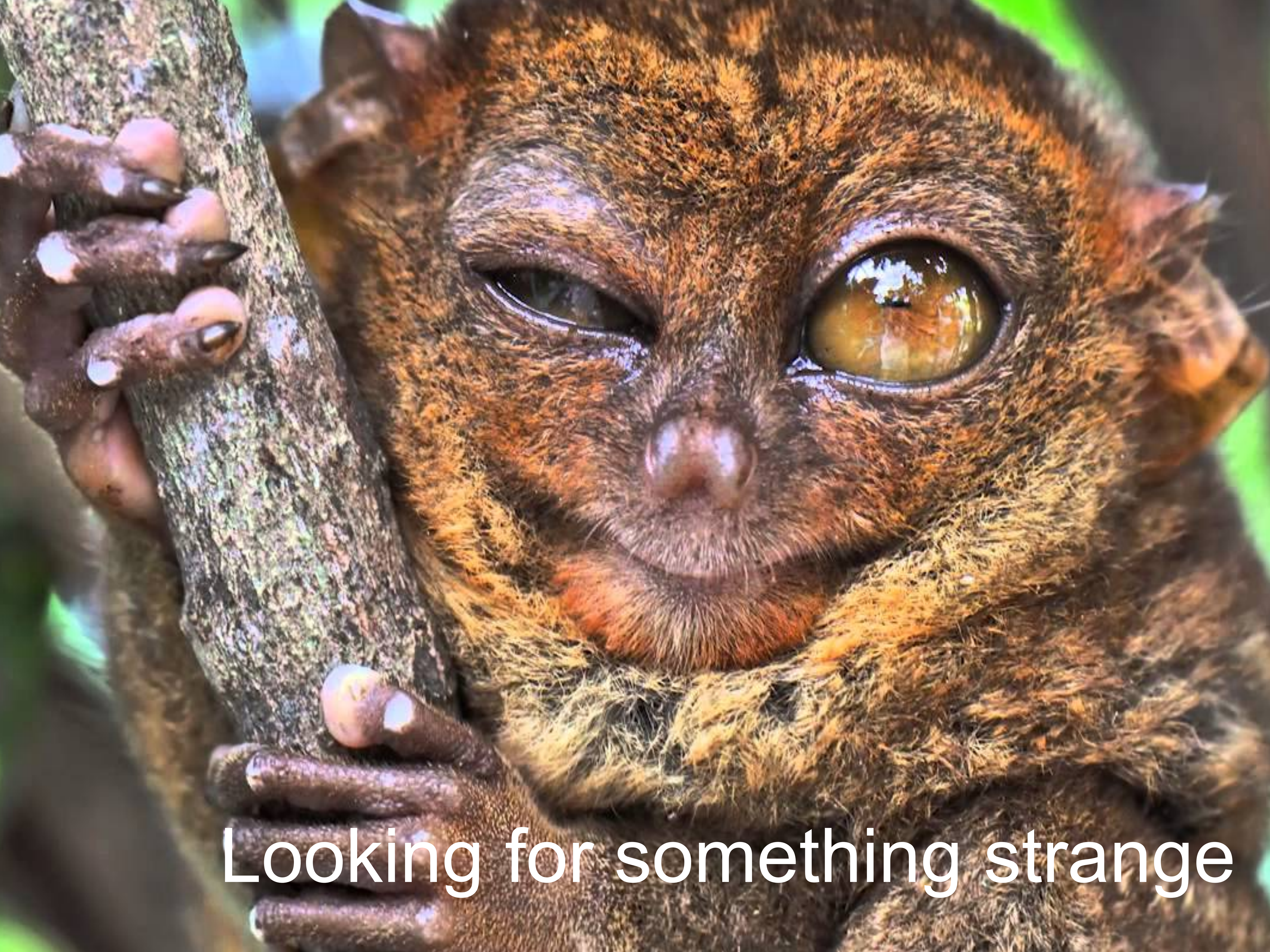
How are ***statistical***, ***systematic*** and ***normalisation*** uncertainties correlated across the individual measurements?

Standard Model Total Production Cross Section Measurements

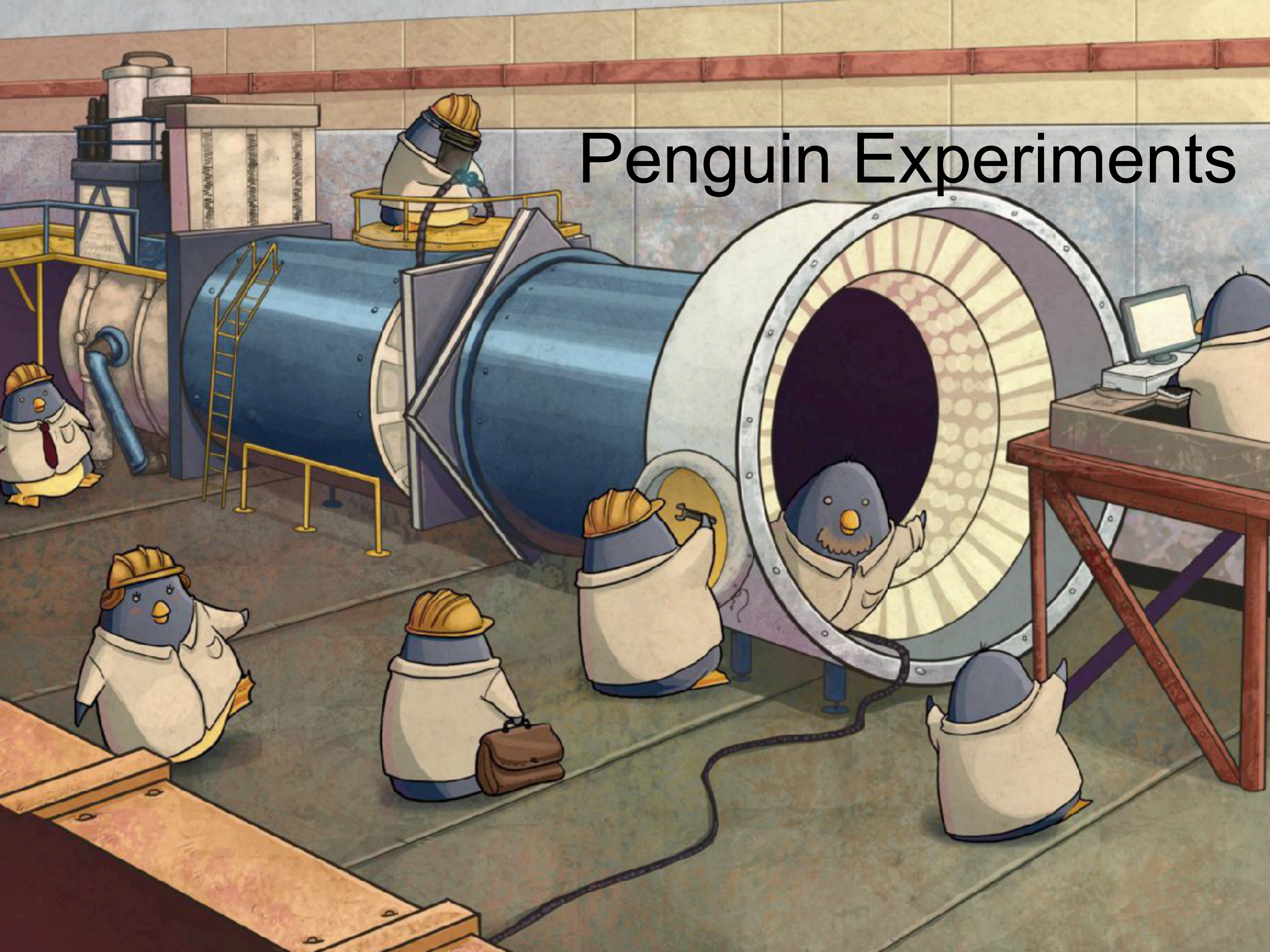
Status: July 2018

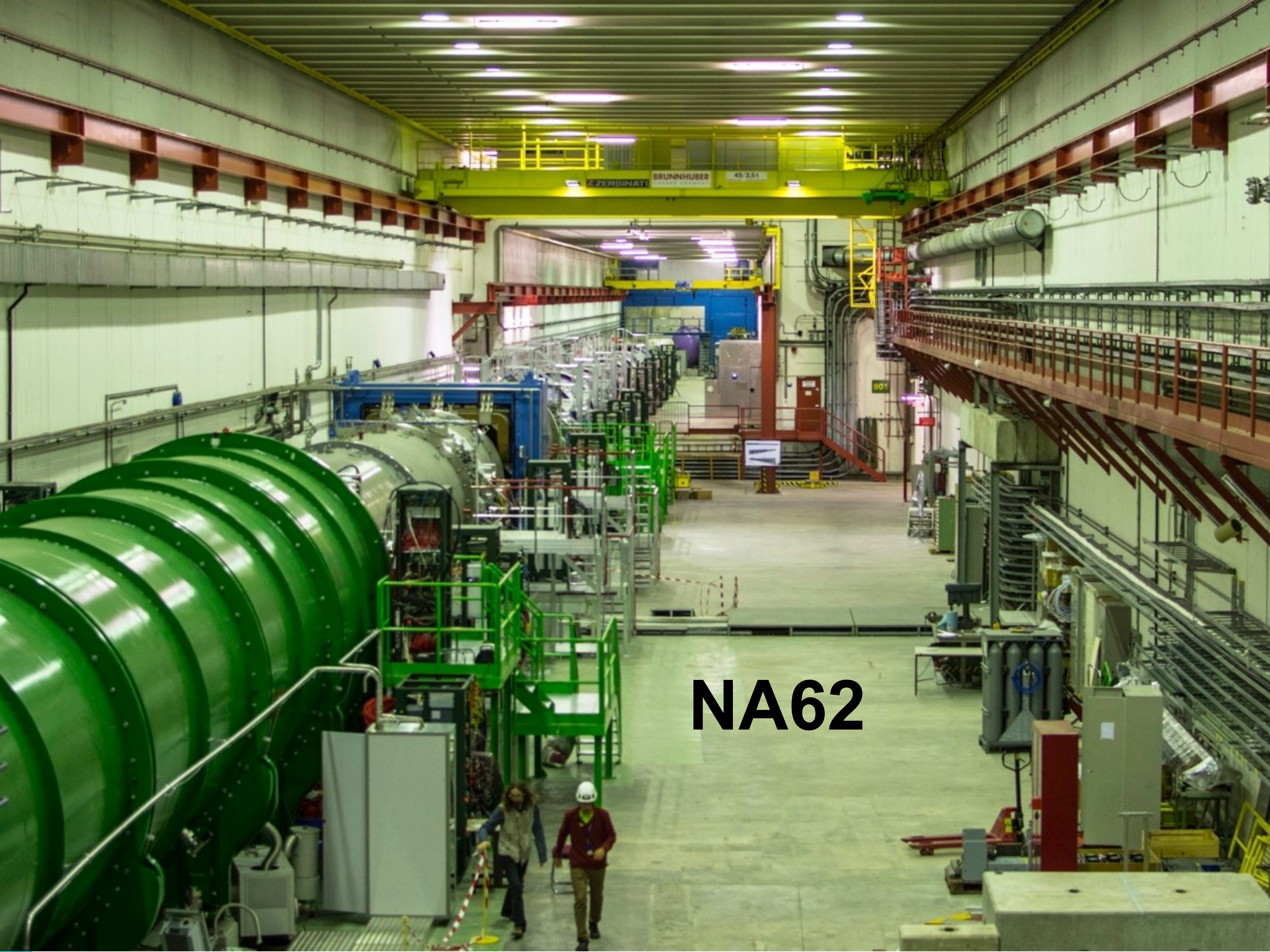


Brookhaven National Laboratory

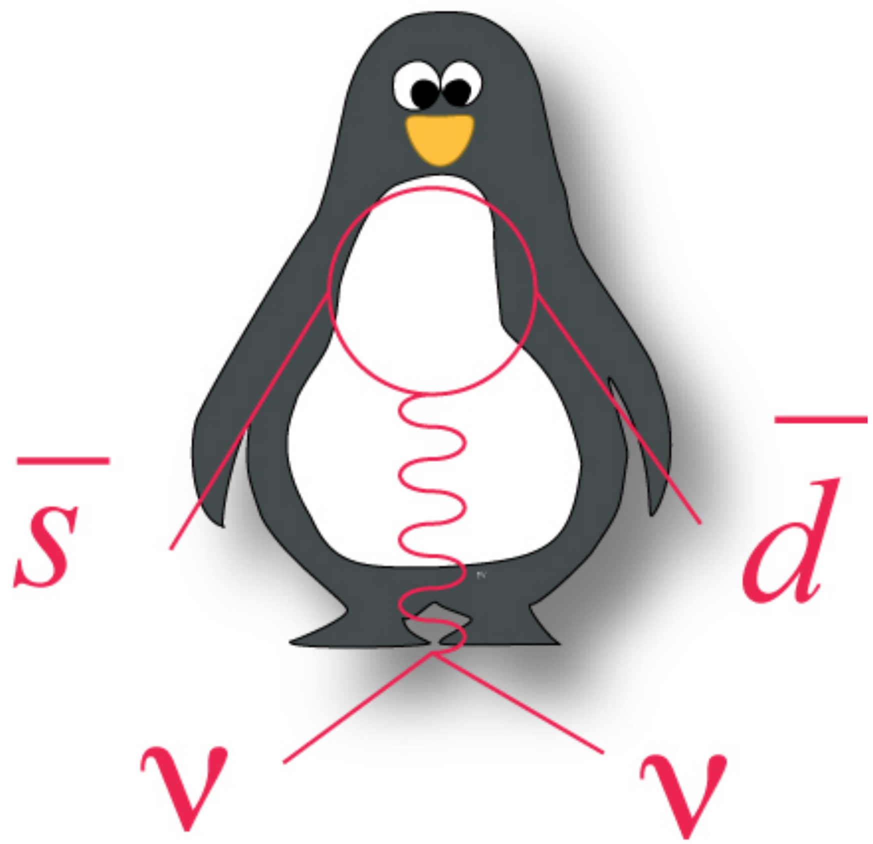

Looking for something strange

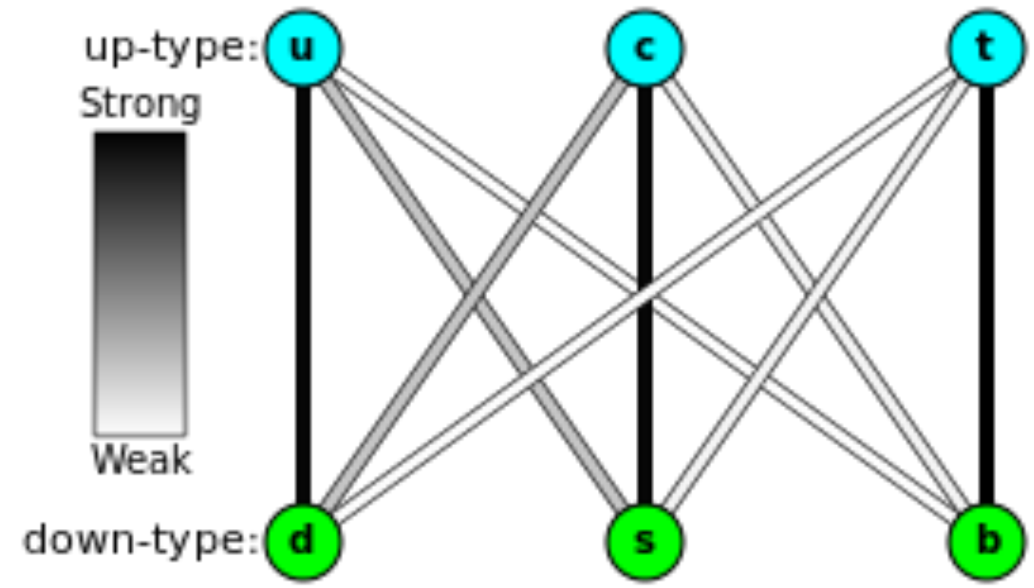
Penguin Experiments




ZERBINATI
BRUNNHUBER
NA62

How strange quarks turn into other quarks





$$BR(K^+ \rightarrow \pi^+ \nu\bar{\nu}) = (8.4 \pm 1.0) \times 10^{-11}$$ [Buras et al. JHEP 1511 (2015) 33]

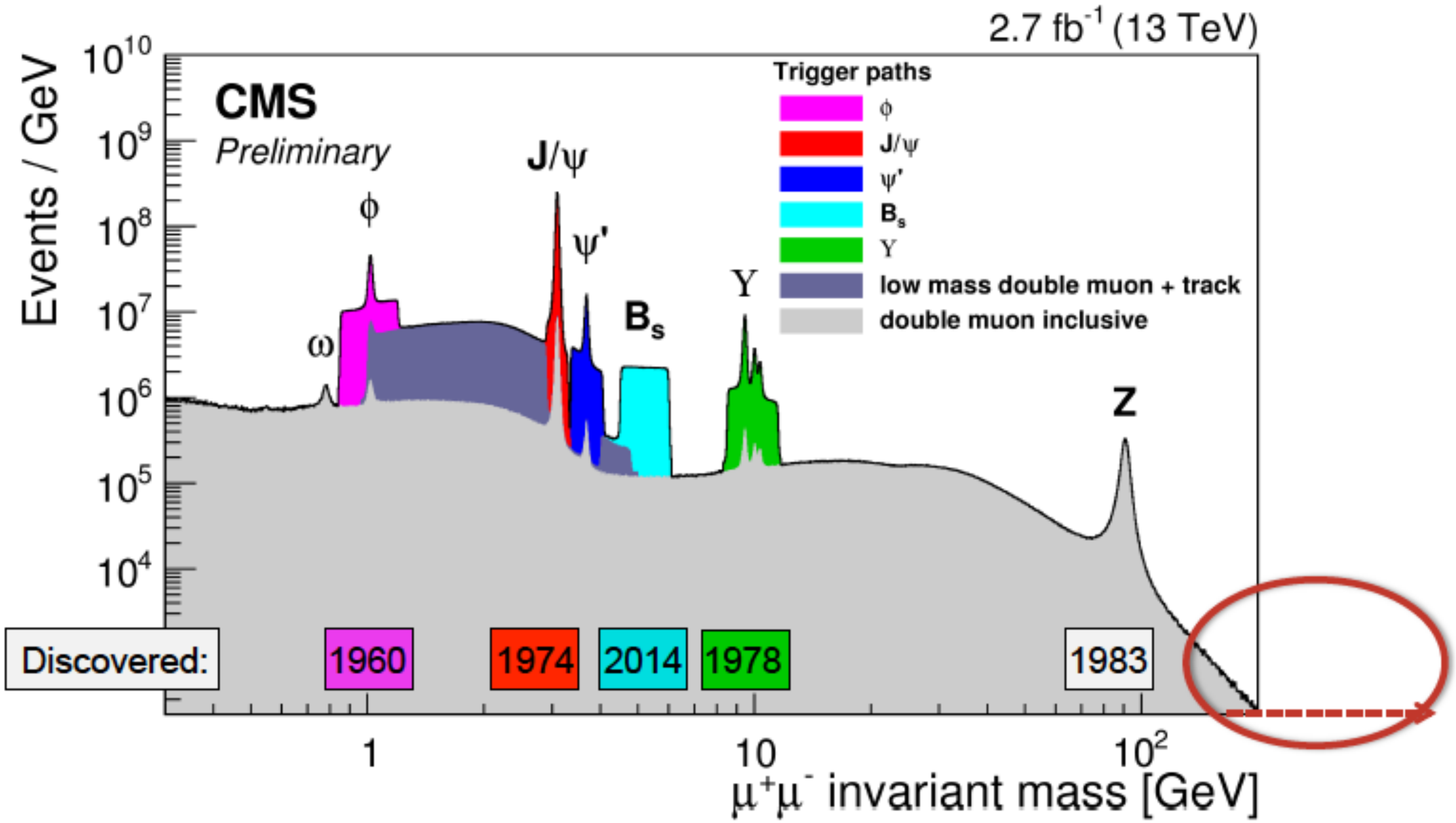
Theory predicts this happens ~100 times in a million million kaon decays:

1) Record a million million kaon decays
2) Analysis - do we count 100 signal events?
   Yes - Congratulations, that's very impressive!
   No - *Congratulations, you've discovered new physics!*

1 in 10 billion is the same probability as finding you in the Earth's population

Brookhaven National Laboratory

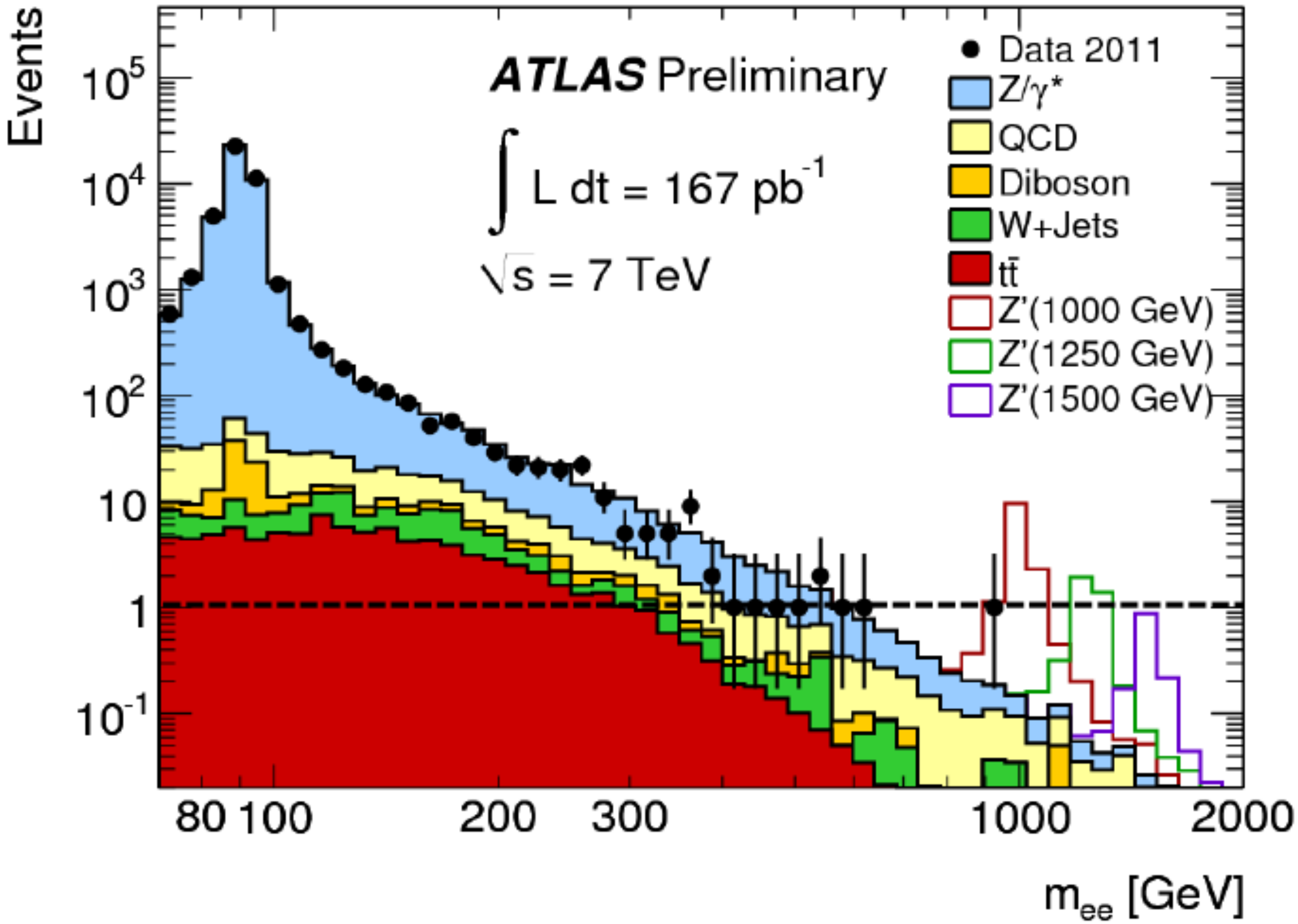
Elements of a search



What's out there?

Elements of a search

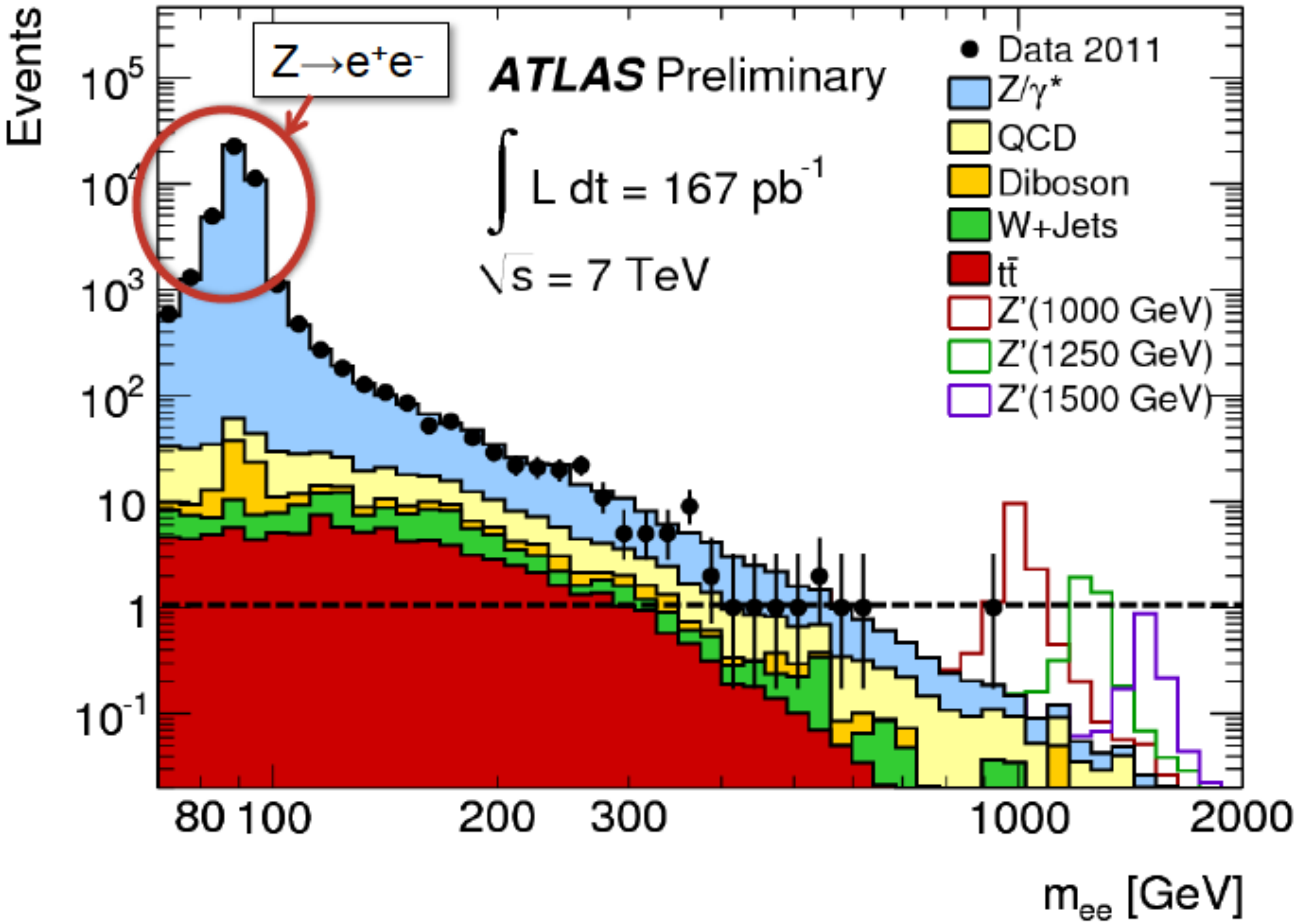
- Like Z->ee but at higher mass.



Brookhaven National Laboratory

Elements of a search

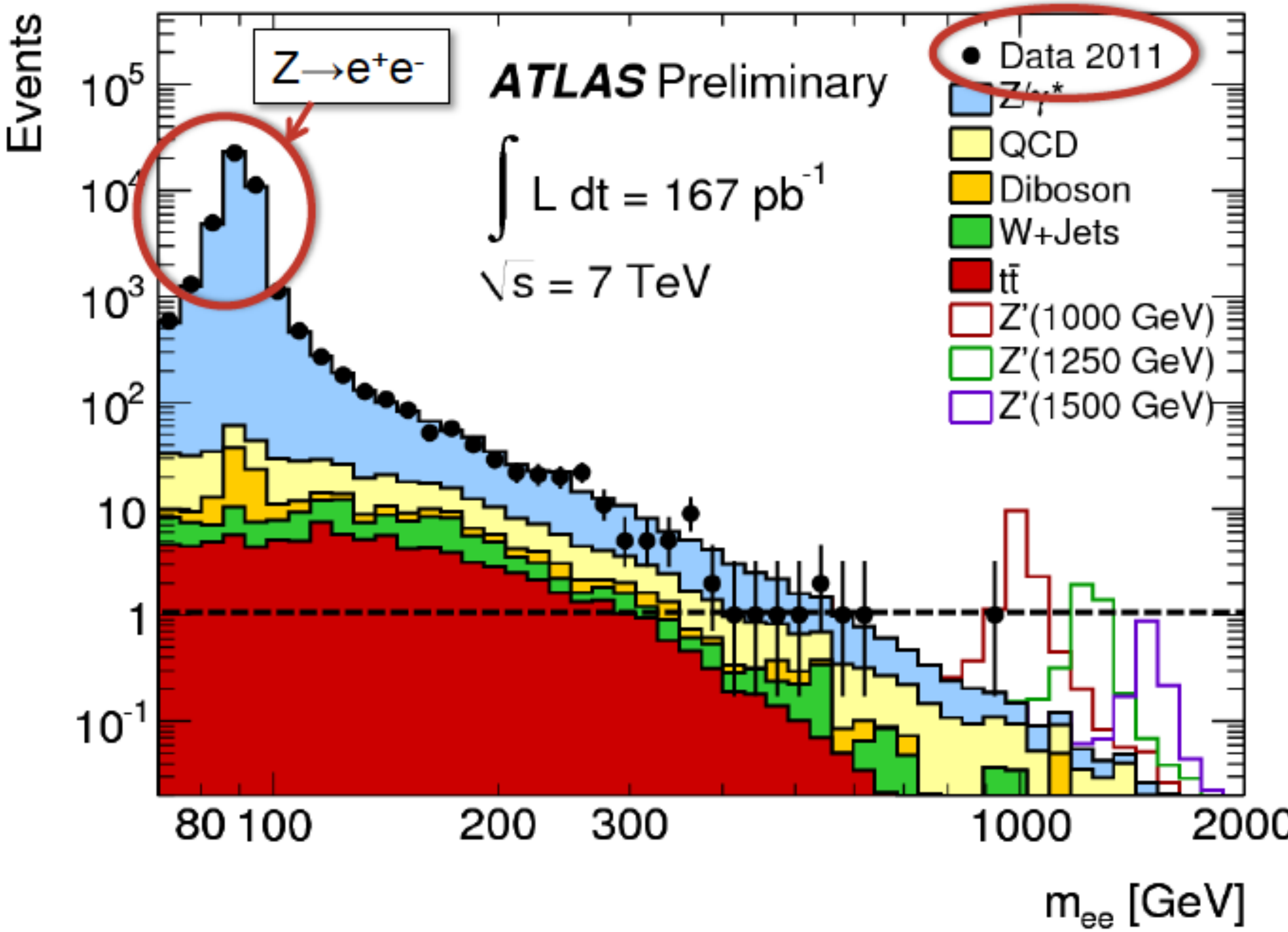
- Like Z->ee but at higher mass.



Brookhaven National Laboratory

Elements of a search

- **Like Z->ee but at higher mass.**

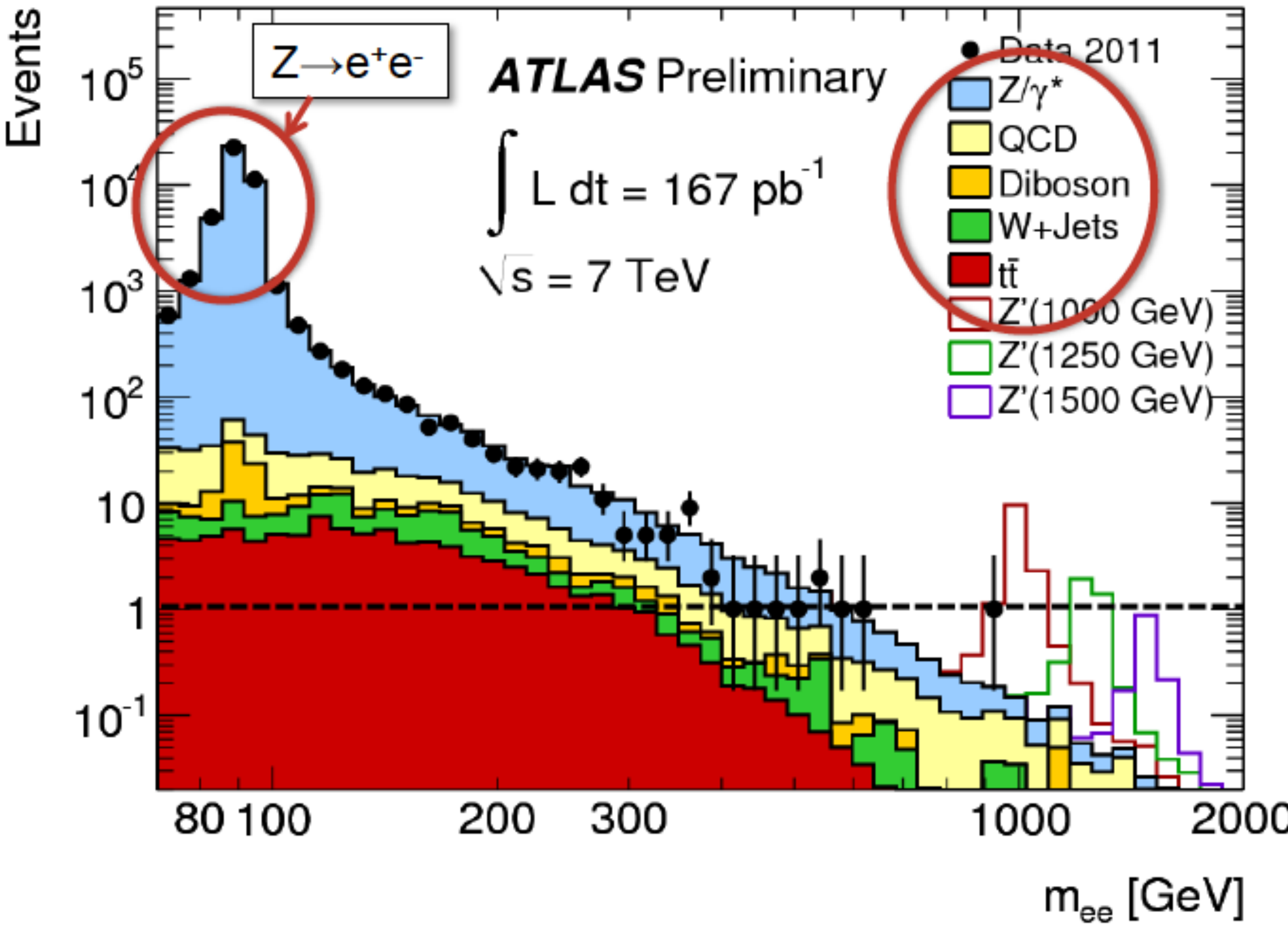


Select 2 electron candidates and plot their invariant mass for:

1. **Data**

Elements of a search

- Like Z->ee but at higher mass.

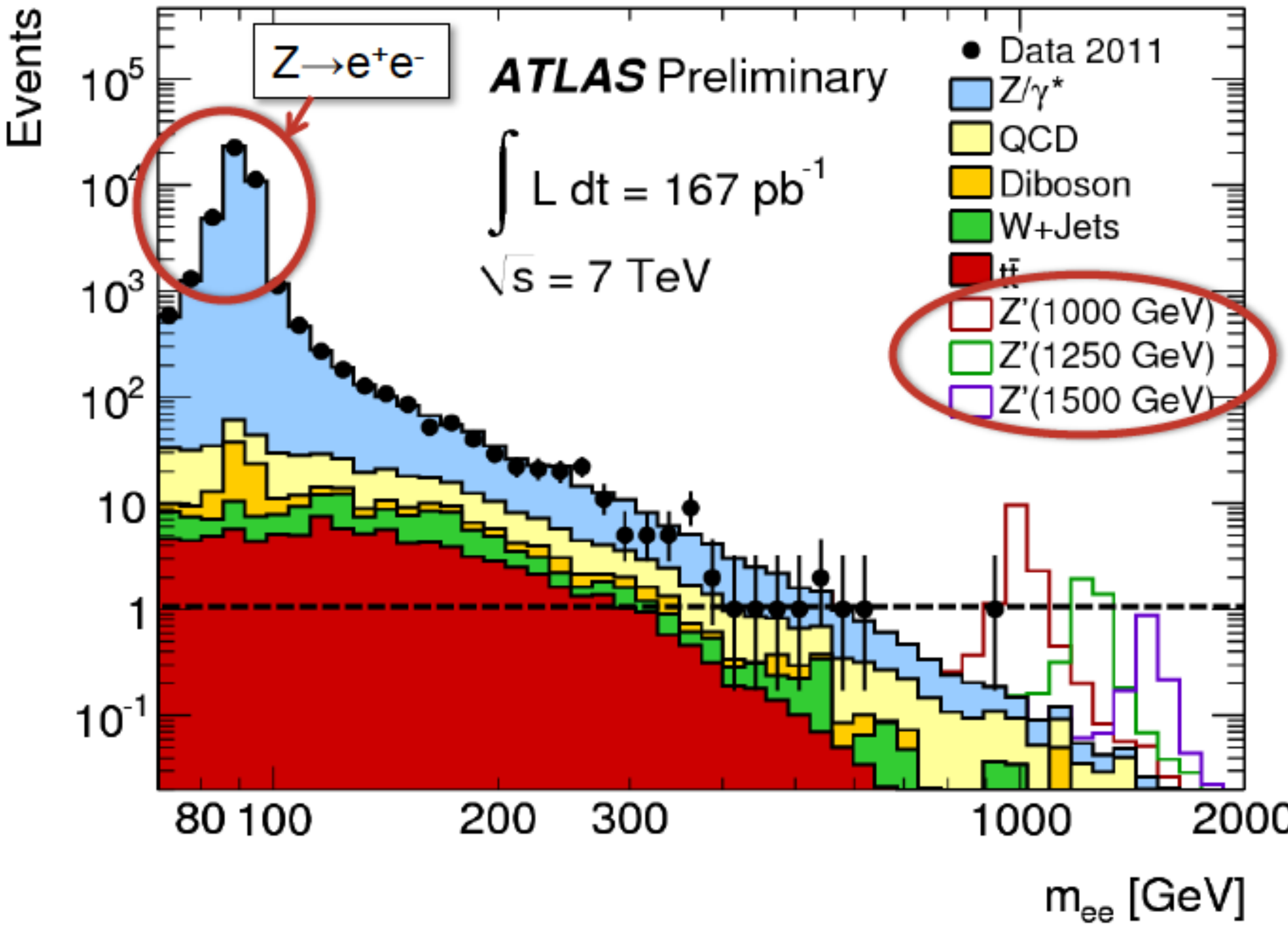


Select 2 electron candidates and plot their invariant mass for:

1. **Data**
2. **Simulated background events**

Elements of a search

- Like Z->ee but at higher mass.

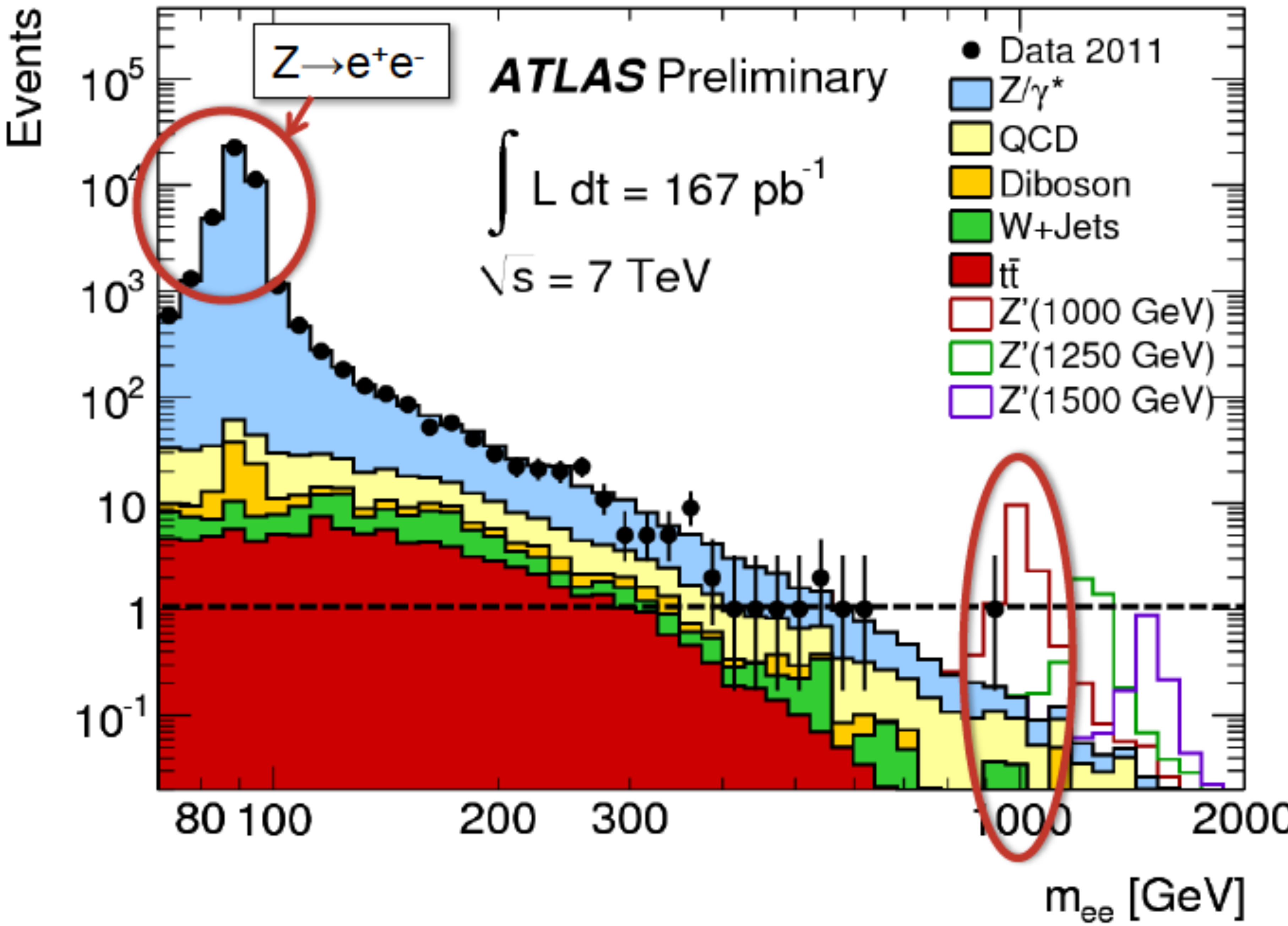


Select 2 electron candidates and plot their invariant mass for:

1. **Data**
2. **Simulated background events**
3. **Simulated signal with different masses**

Elements of a search

◎ **Like Z->ee but at higher mass.**



Select 2 electron candidates and plot their invariant mass for:

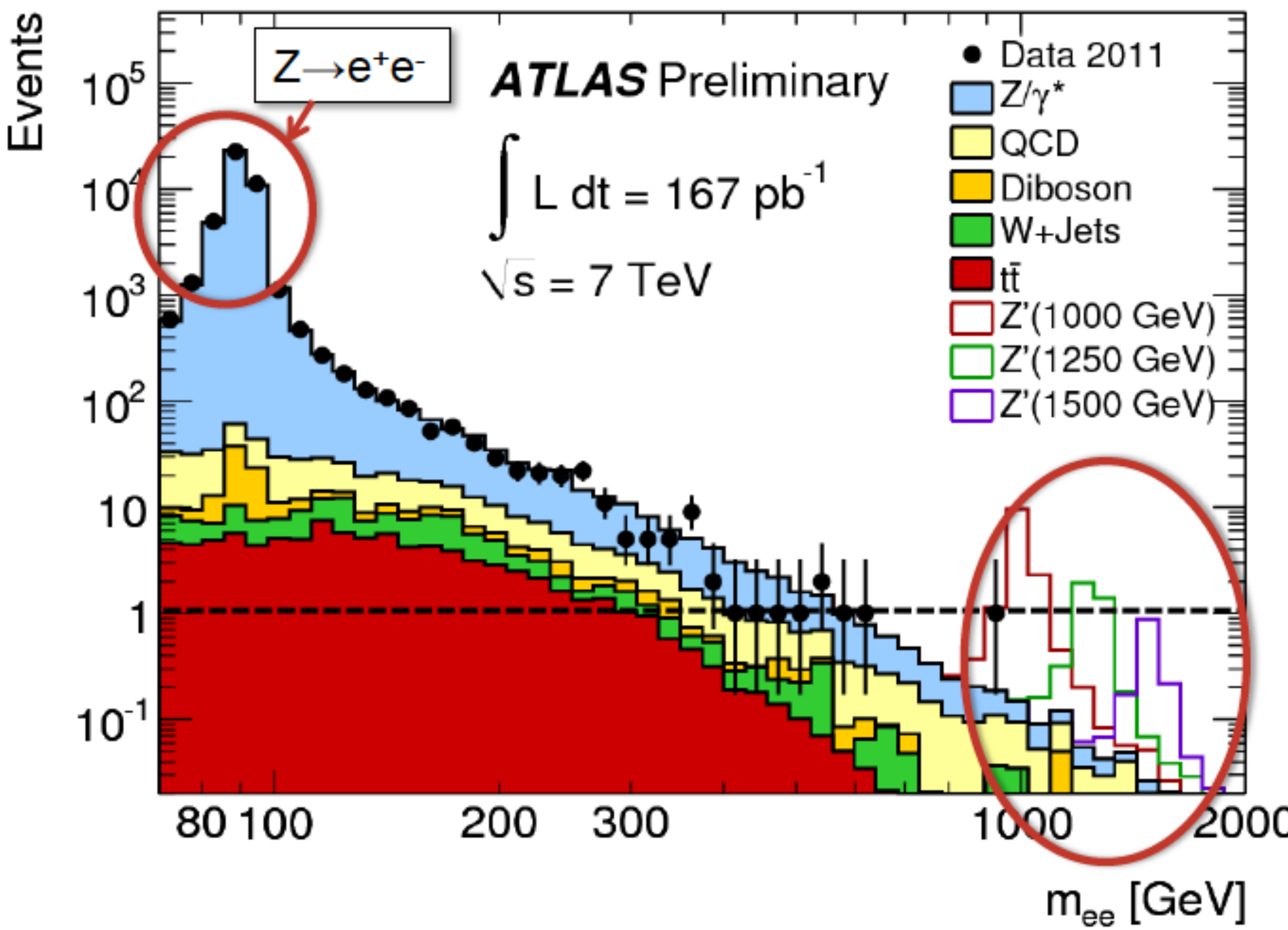
1. **Data**
2. **Simulated background events**
3. **Simulated signal with different masses**

Data inconsistent with a 1TeV Z'

Brookhaven National Laboratory

Elements of a search

- Like Z->ee but at higher mass.



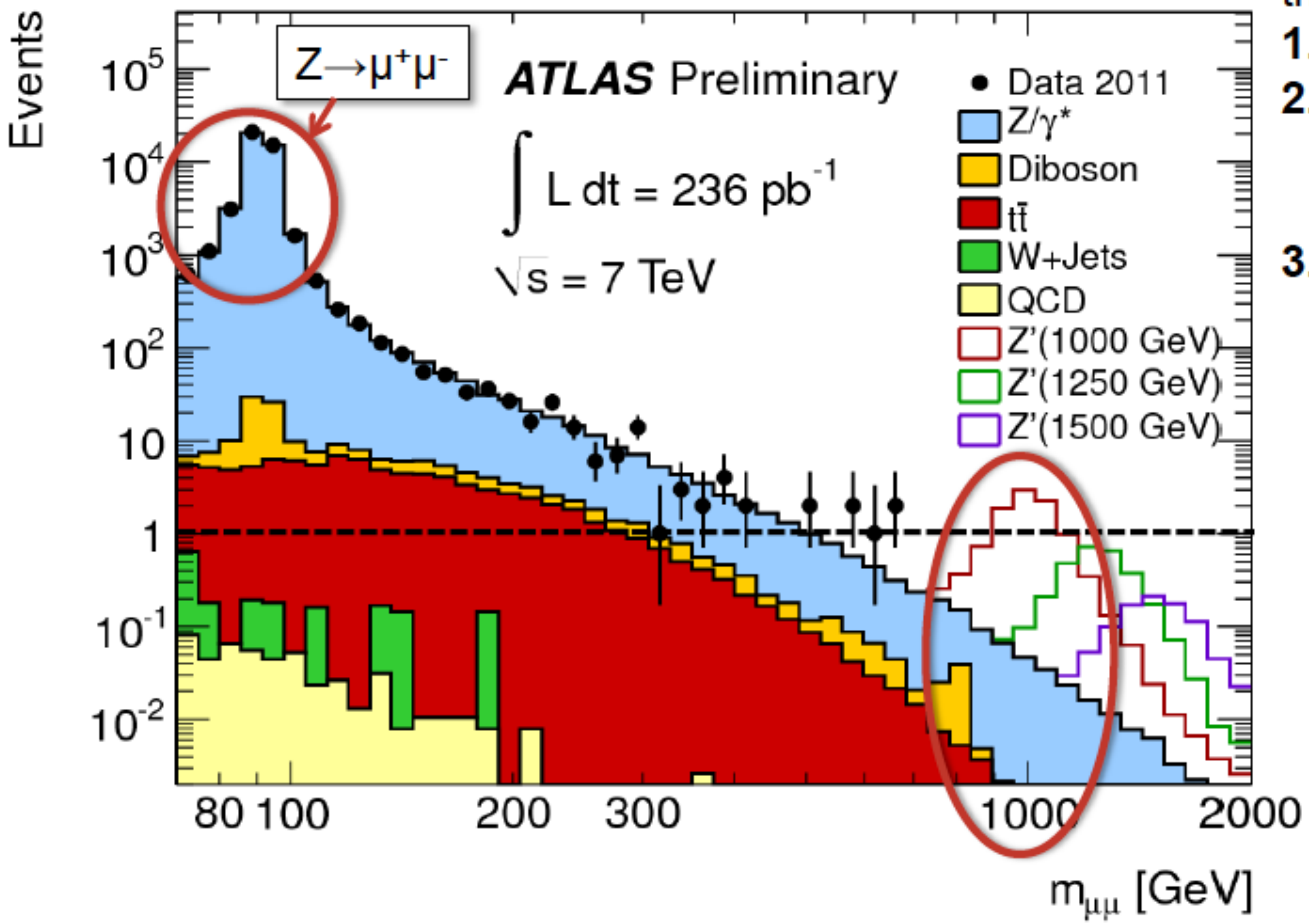
Select 2 electron candidates and plot their invariant mass for:

1. **Data**
2. **Simulated background events**
3. **Simulated signal with different masses**

Cross-section decreases with mass (higher the mass of the Z', the more data needed to discover it)

Elements of a search

- **And similar for muons**

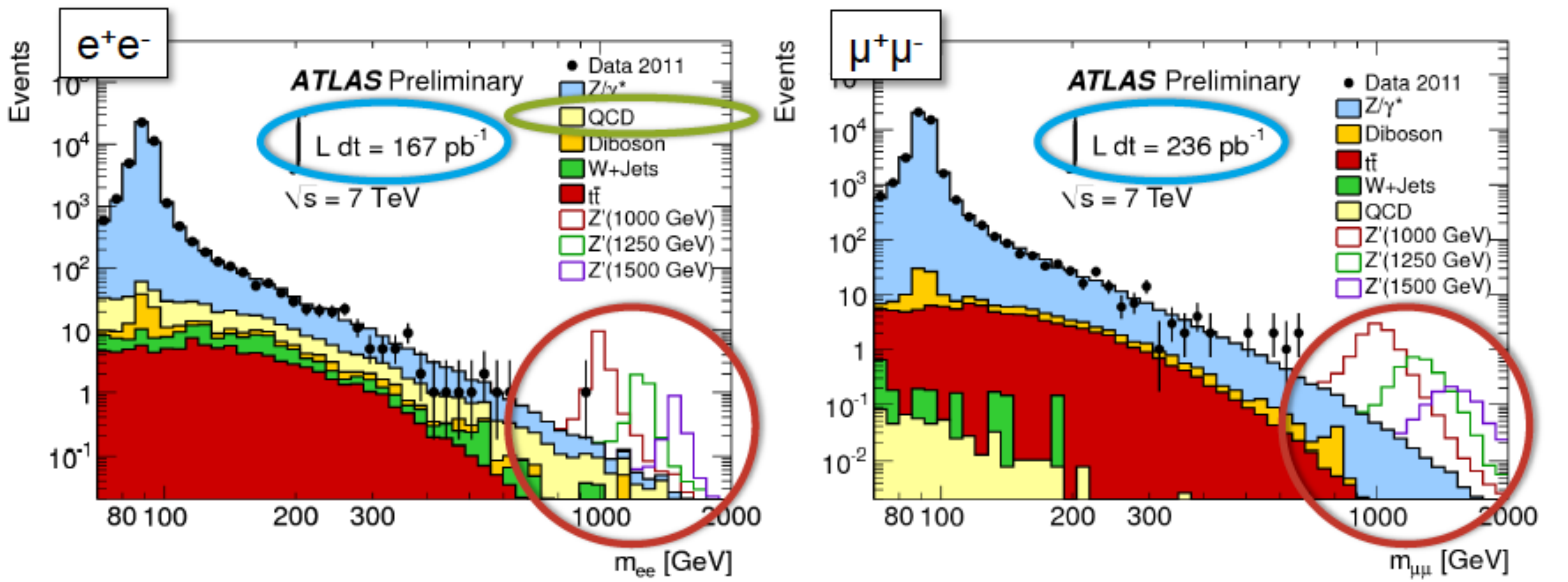


Select 2 muon candidates and plot their invariant mass for:

1. **Data**
2. **Simulated background events**
3. **Simulated signal with different masses**

Data inconsistent with a 1TeV Z'

Elements of a search

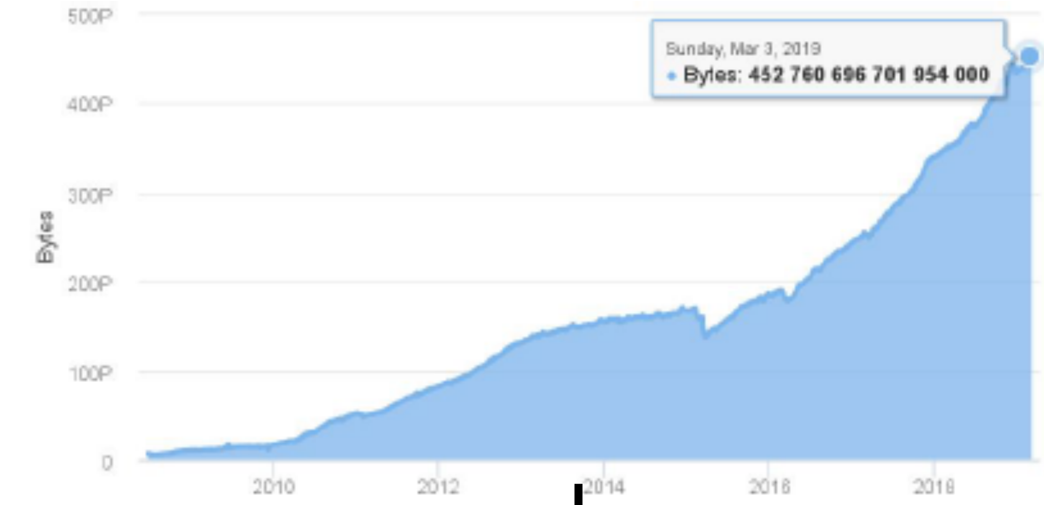


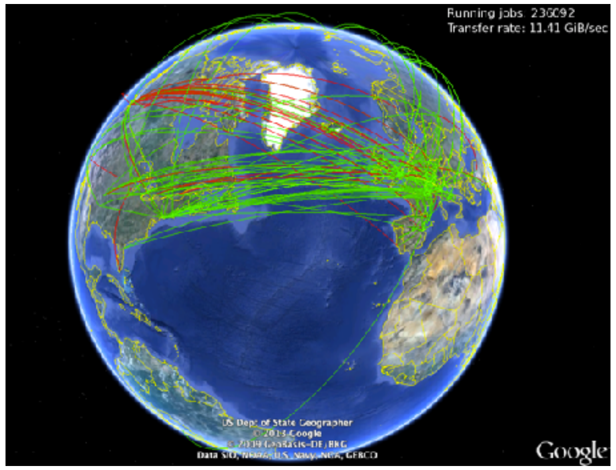
Why is the resolution worse in the muon channel?

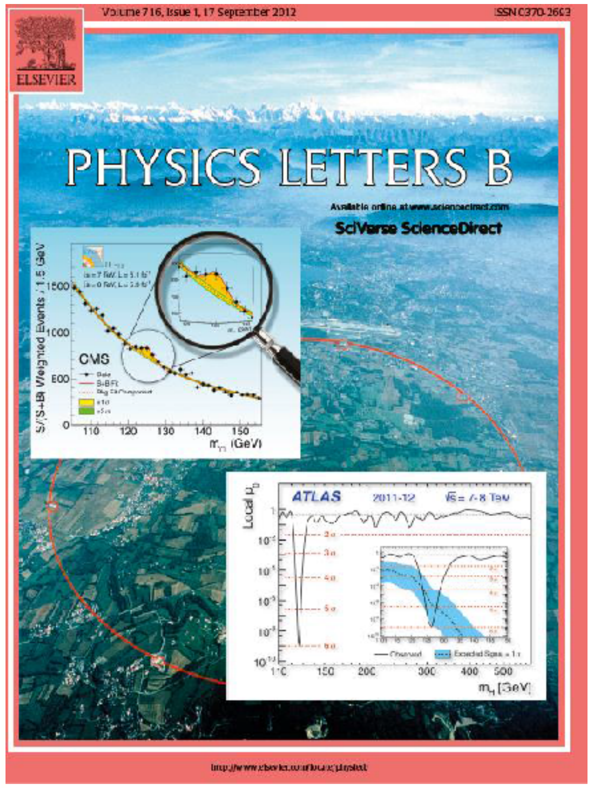
Differences in:
- **Resolution**
- **Background composition**
- **Dataset**

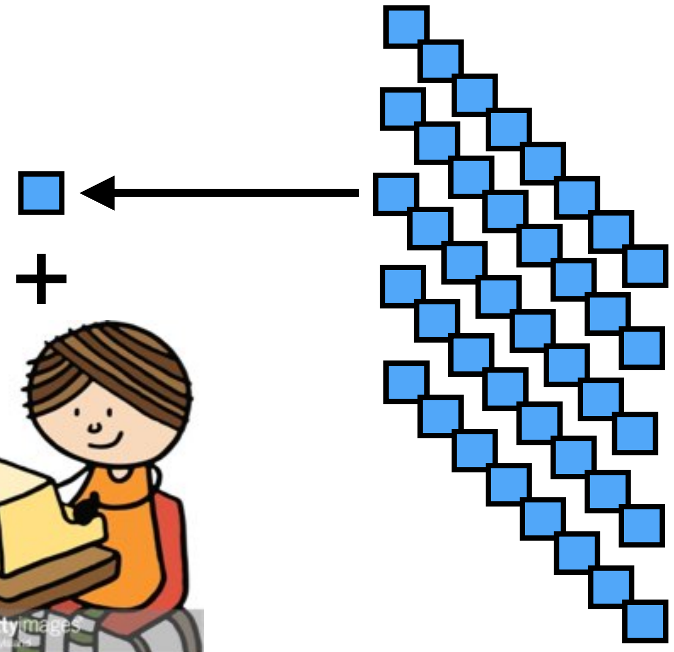
Brookhaven National Laboratory

Data analysis









Needles in haystacks

- We record billions of events

- The data are structured but each event is different - ***unique data science challenge***

Data reduction proceeds via a two-pronged approach:

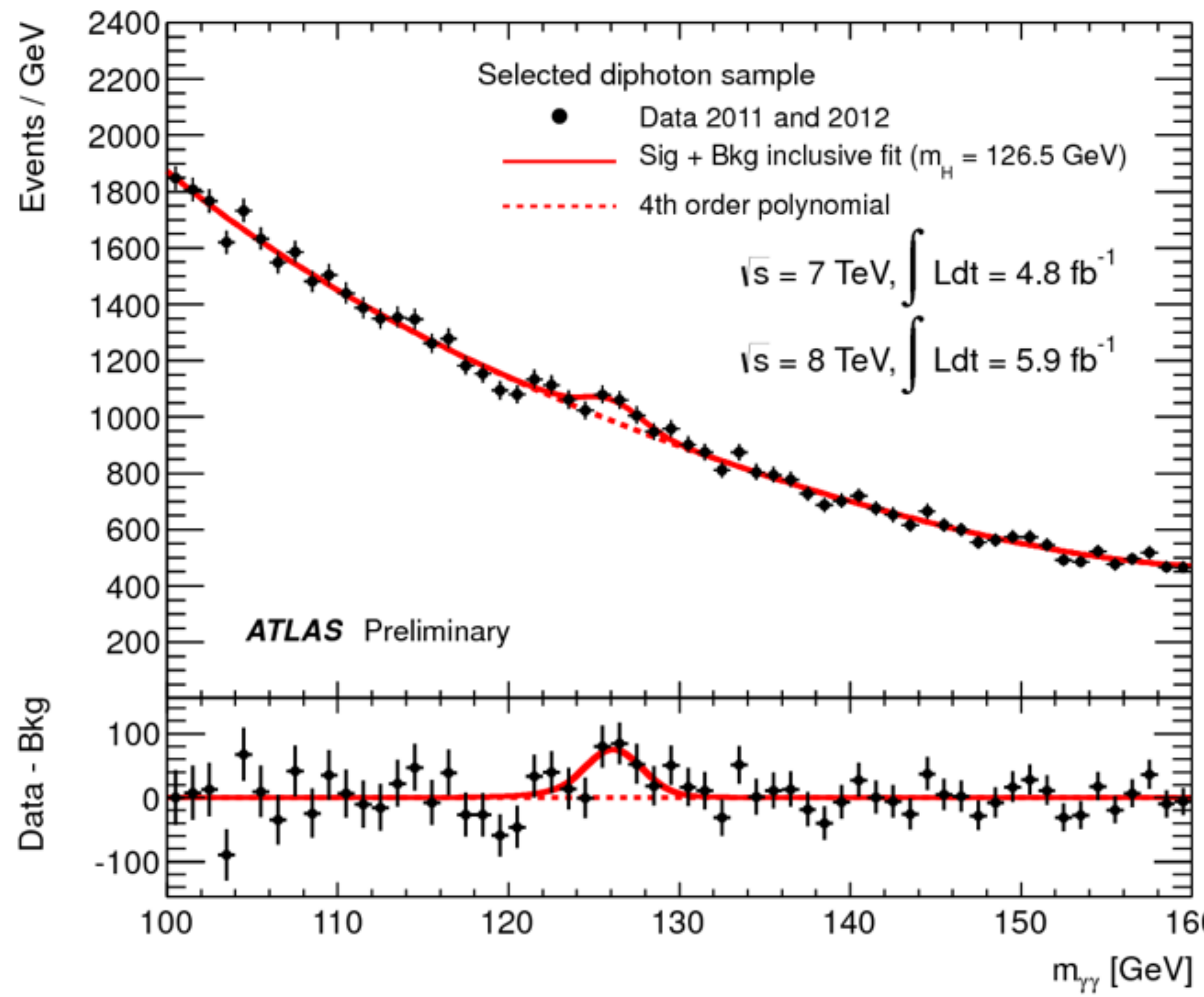
Select only the events that you are interested in

- ***e.g. events with two photons***

Keep only the information you need

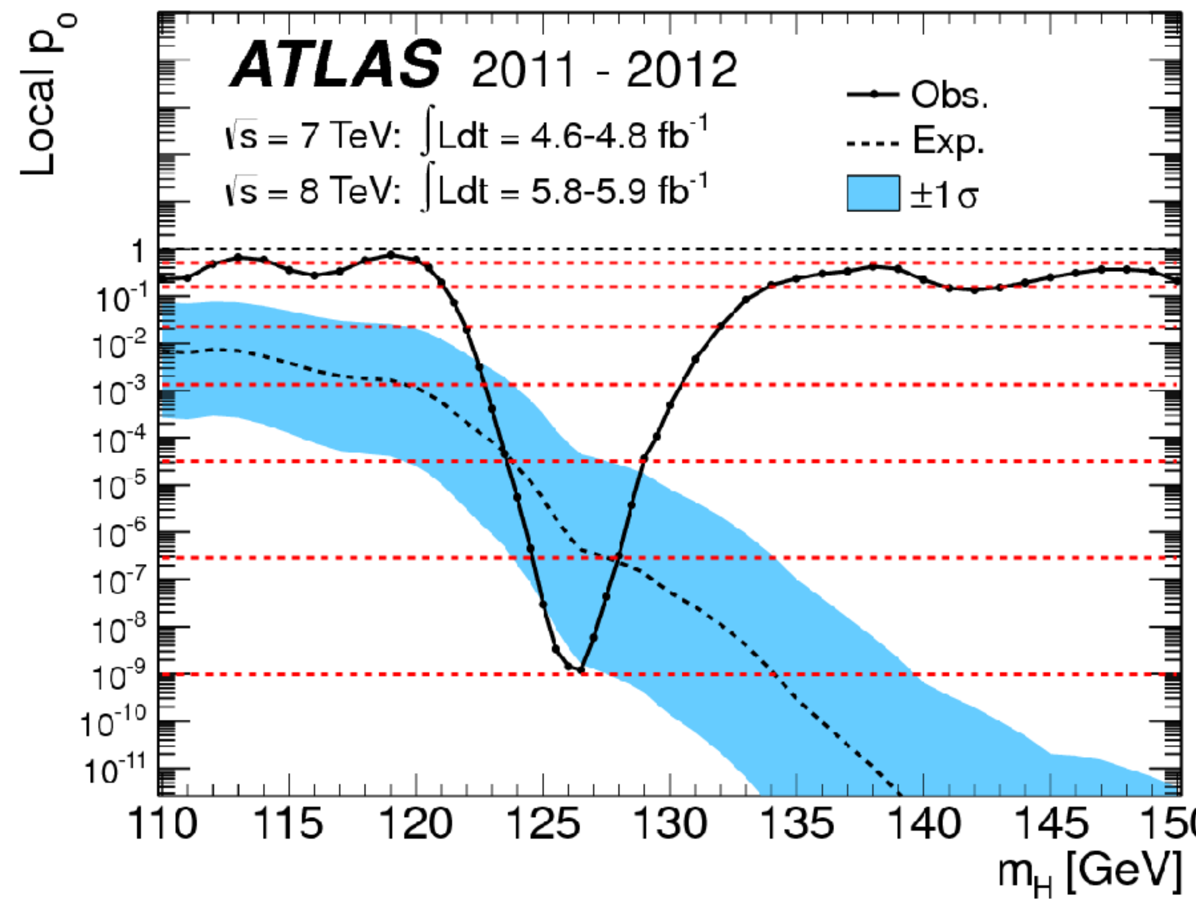
Throw away the rest !

Final statistical inference is only performed on the reduced data



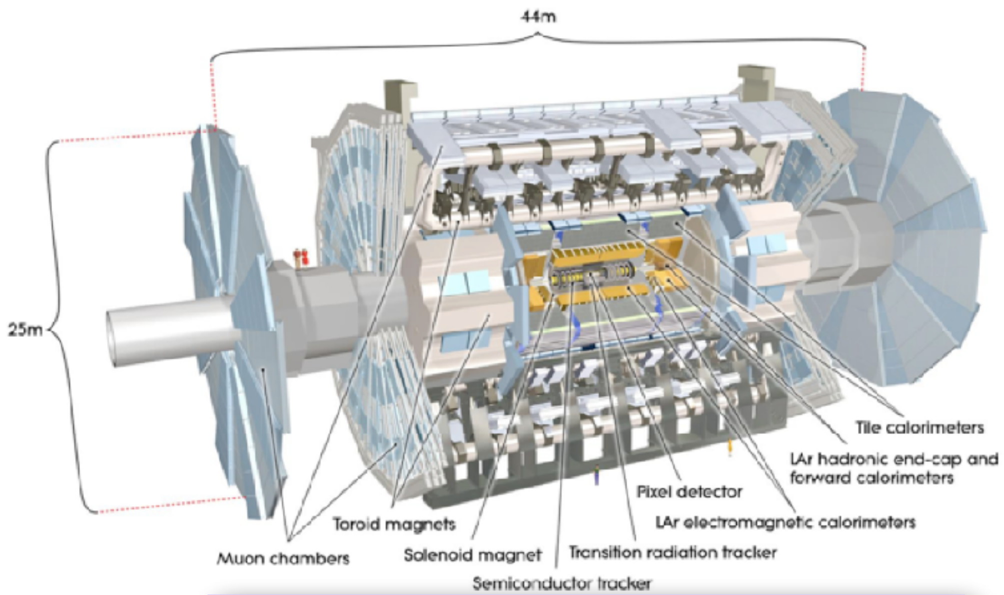
Brookhaven National Laboratory

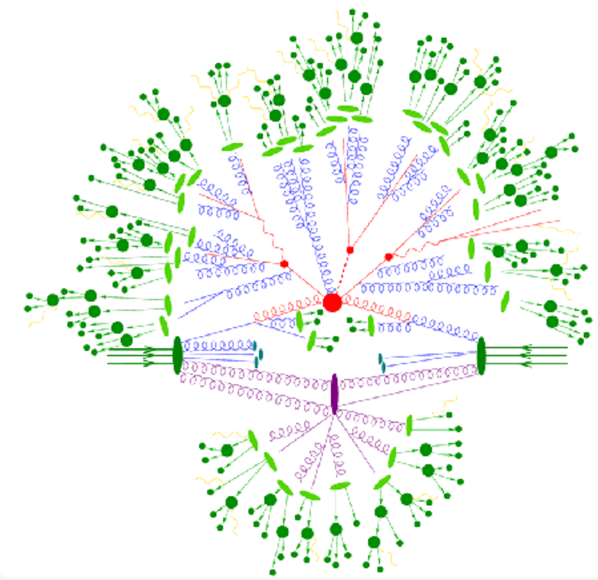
Higgs discovery in 2012



Now you can appreciate a little better the effort that went into this plot - and why we were very happy !

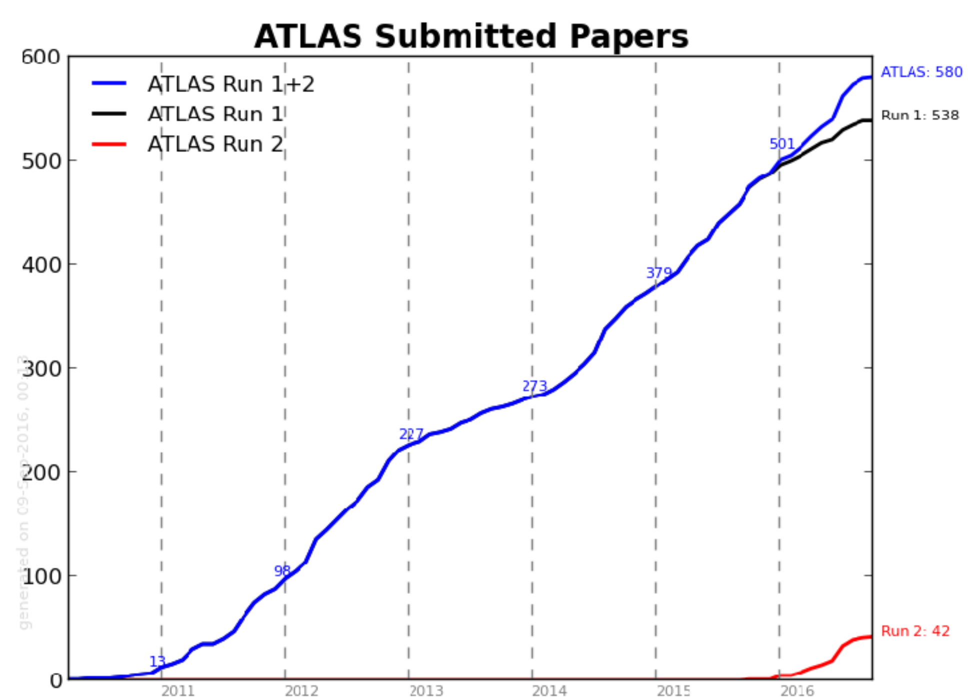
Now you know how to do exabyte-scale physics analysis!





Exabytes of Data

Exabytes of Simulation



Publish!

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 again!:
Tuesday 13:30-15:00 in Salle Anderson, B40 S2-A01