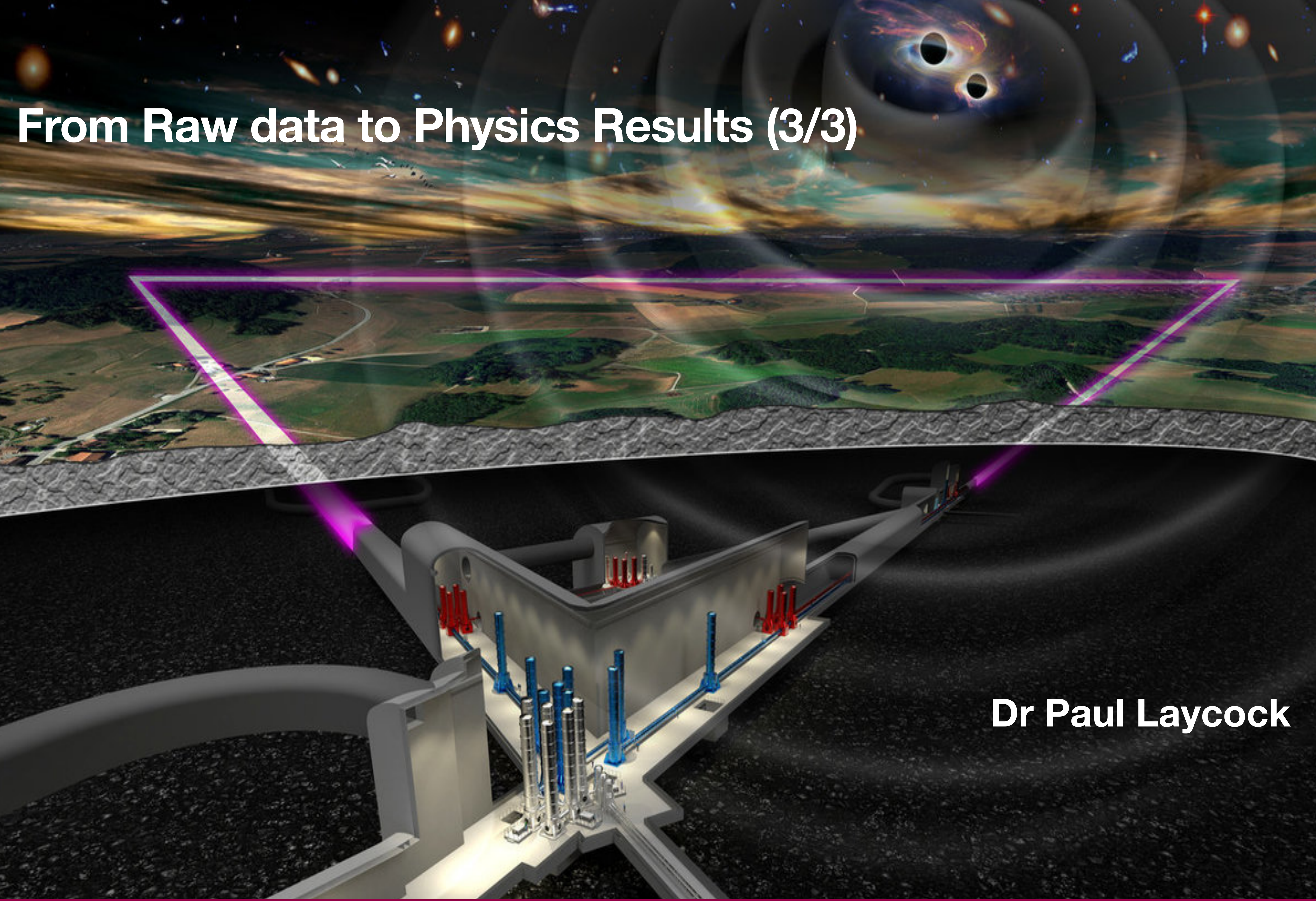
From Raw data to Physics Results (3/3)

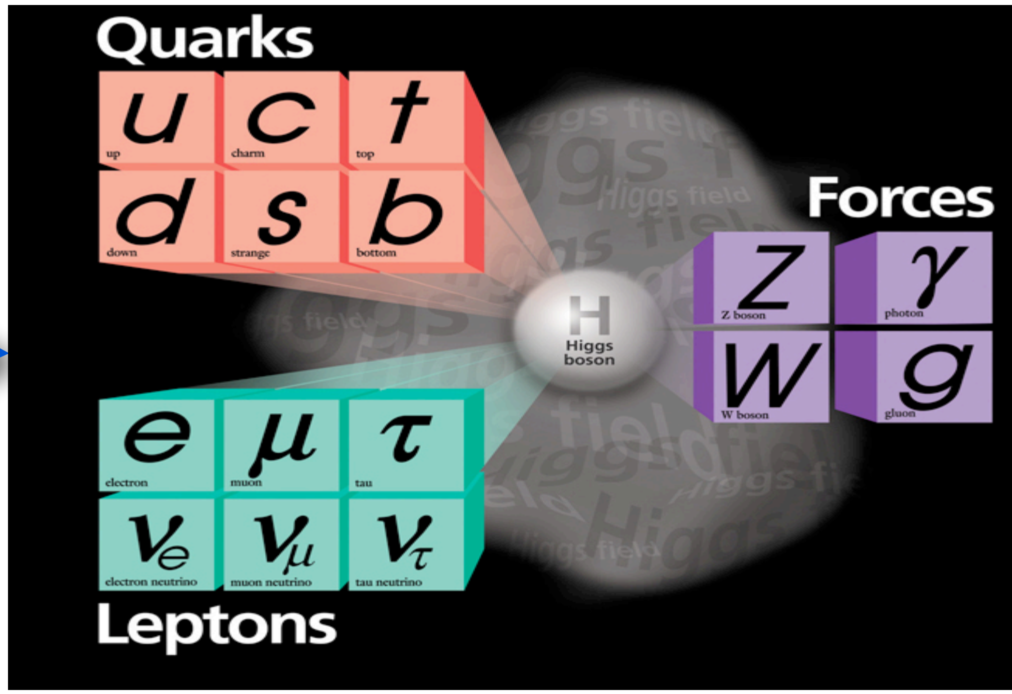


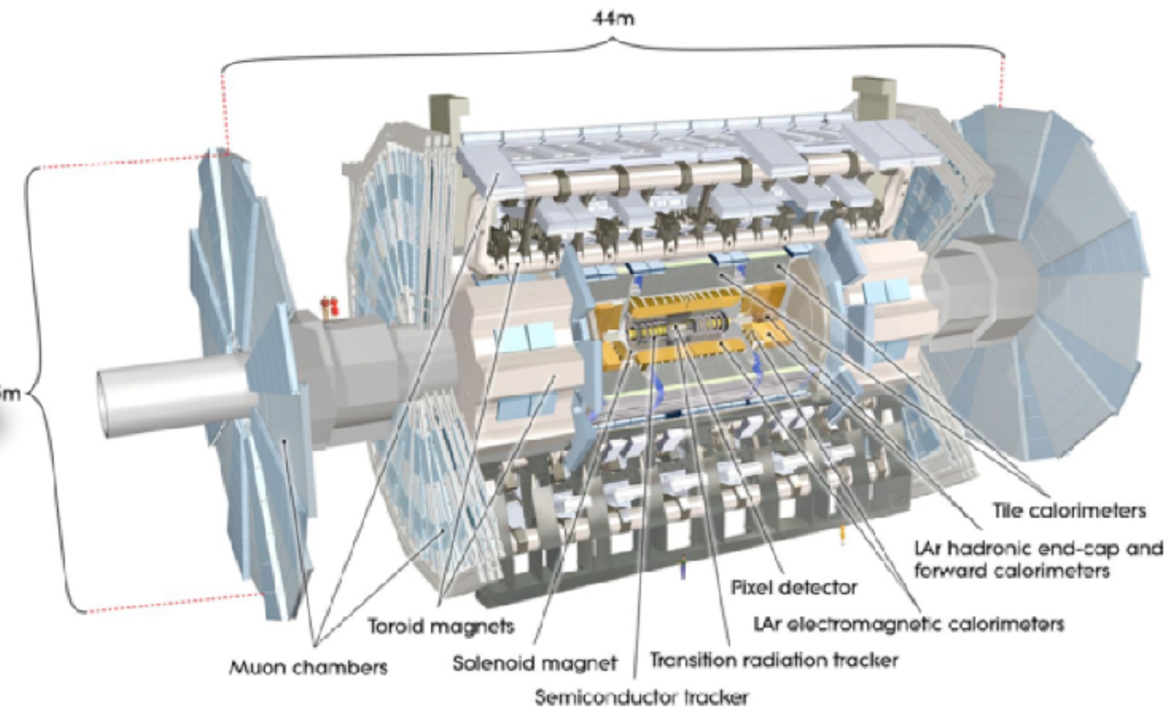
Dr Paul Laycock

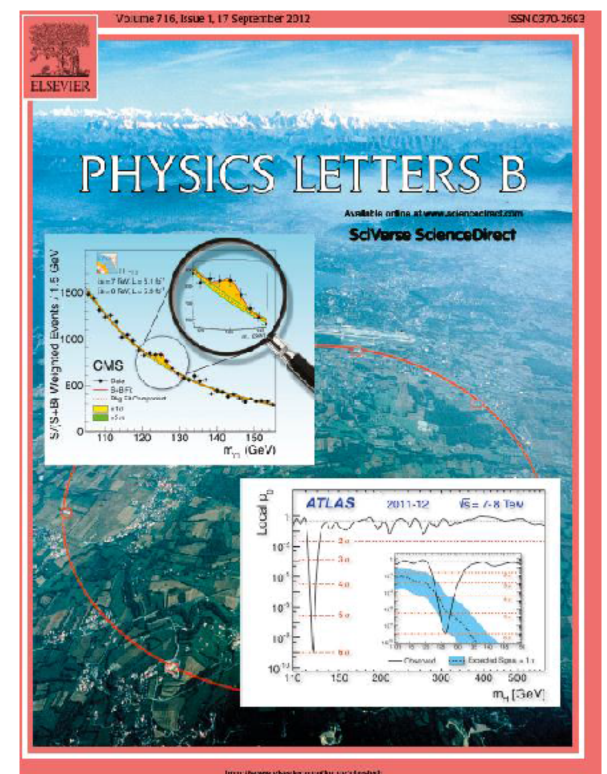
GWSC
GRAVITATIONAL WAVE SCIENCE CENTER

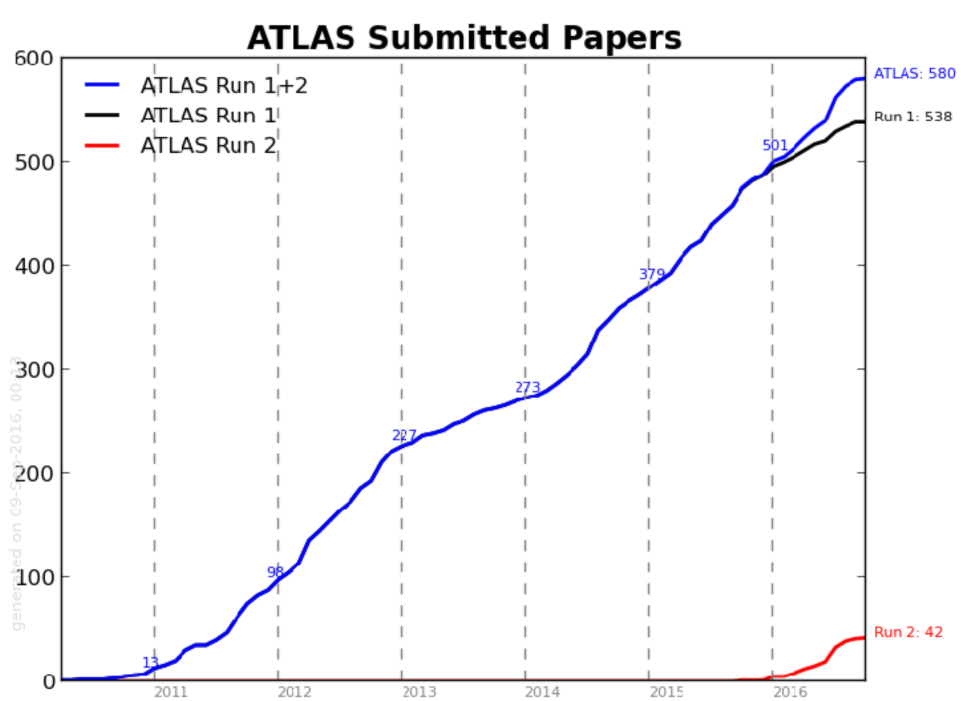
UNIVERSITÉ DE GENÈVE
Faculté des Sciences
Département d'astronomie

The particle physics cycle



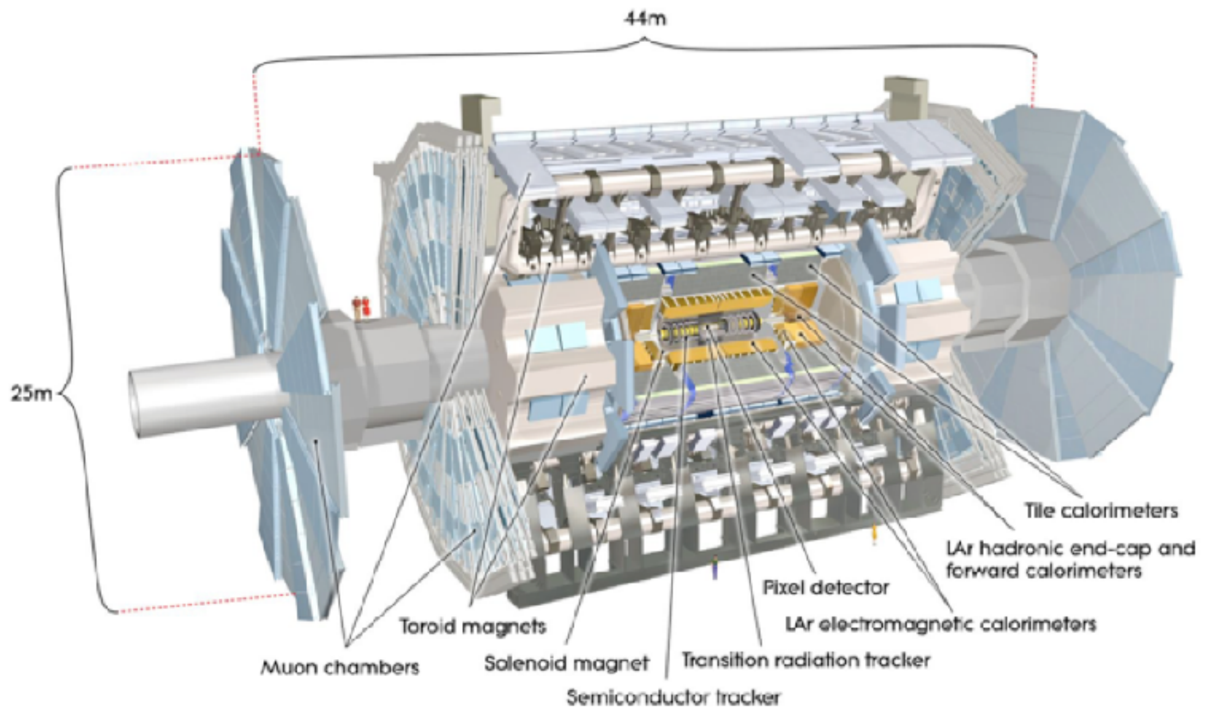


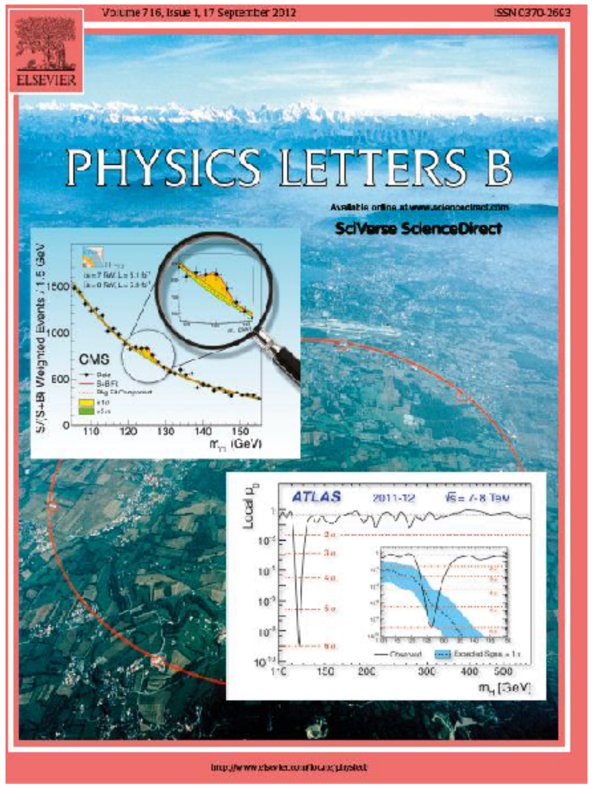


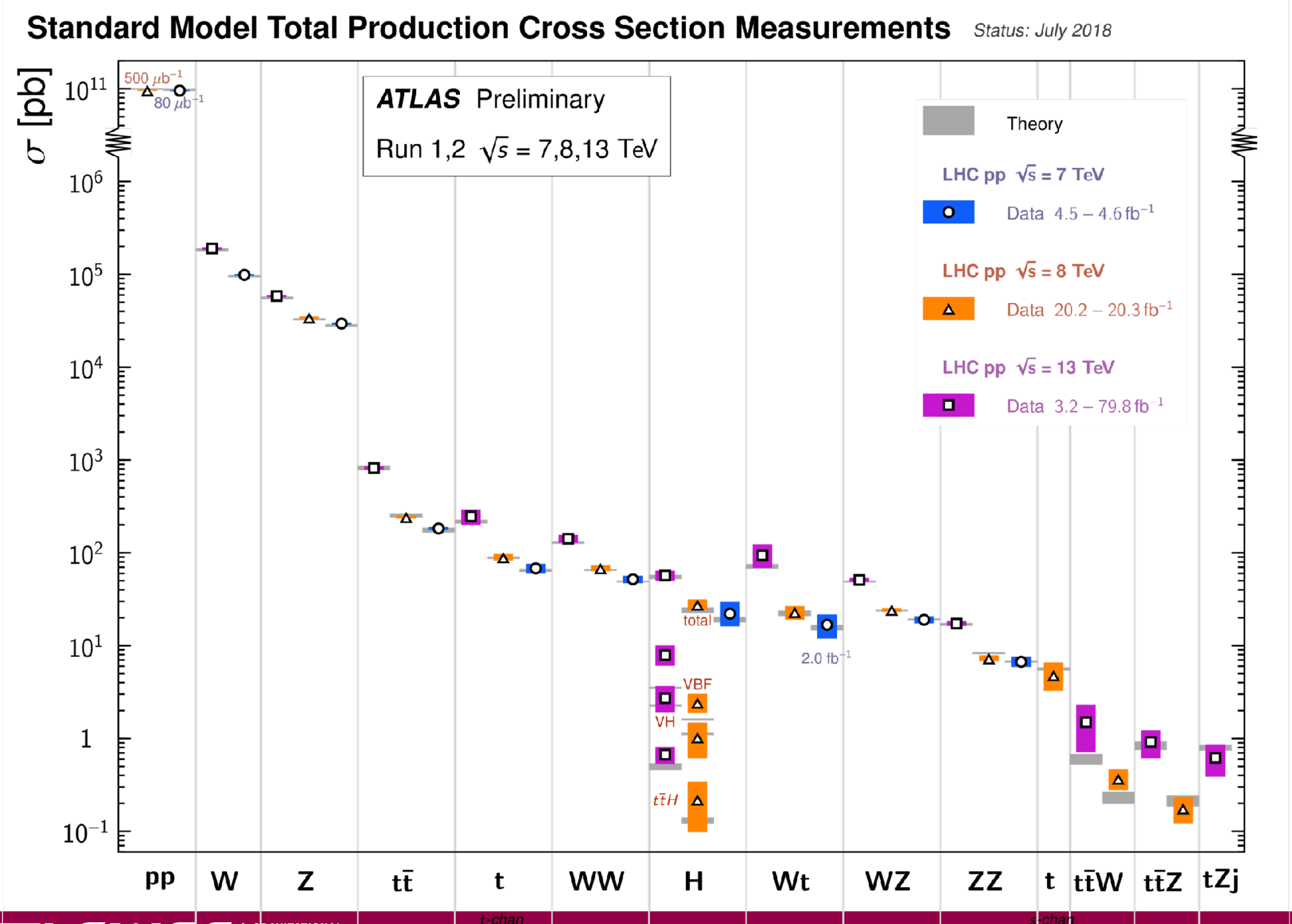


Course outline

- **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
ATLAS Preliminary
Run 1,2 √s = 7,8,13 TeV
Theory
LHC pp √s = 7 TeV
Data 4.5 – 4.6 fb⁻¹
LHC pp √s = 8 TeV
Data 20.2 – 20.3 fb⁻¹
LHC pp √s = 13 TeV
Data 3.2 – 79.8 fb⁻¹
σ [pb]
500 μb⁻¹
80 μb⁻¹
total
VBF
VH
tt̄H
2.0 fb⁻¹
pp
W
Z
tt̄
t
t-chan
WW
H
Wt
WZ
ZZ
t
s-chan
tt̄W
tt̄Z
tZj

Measuring cross sections

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

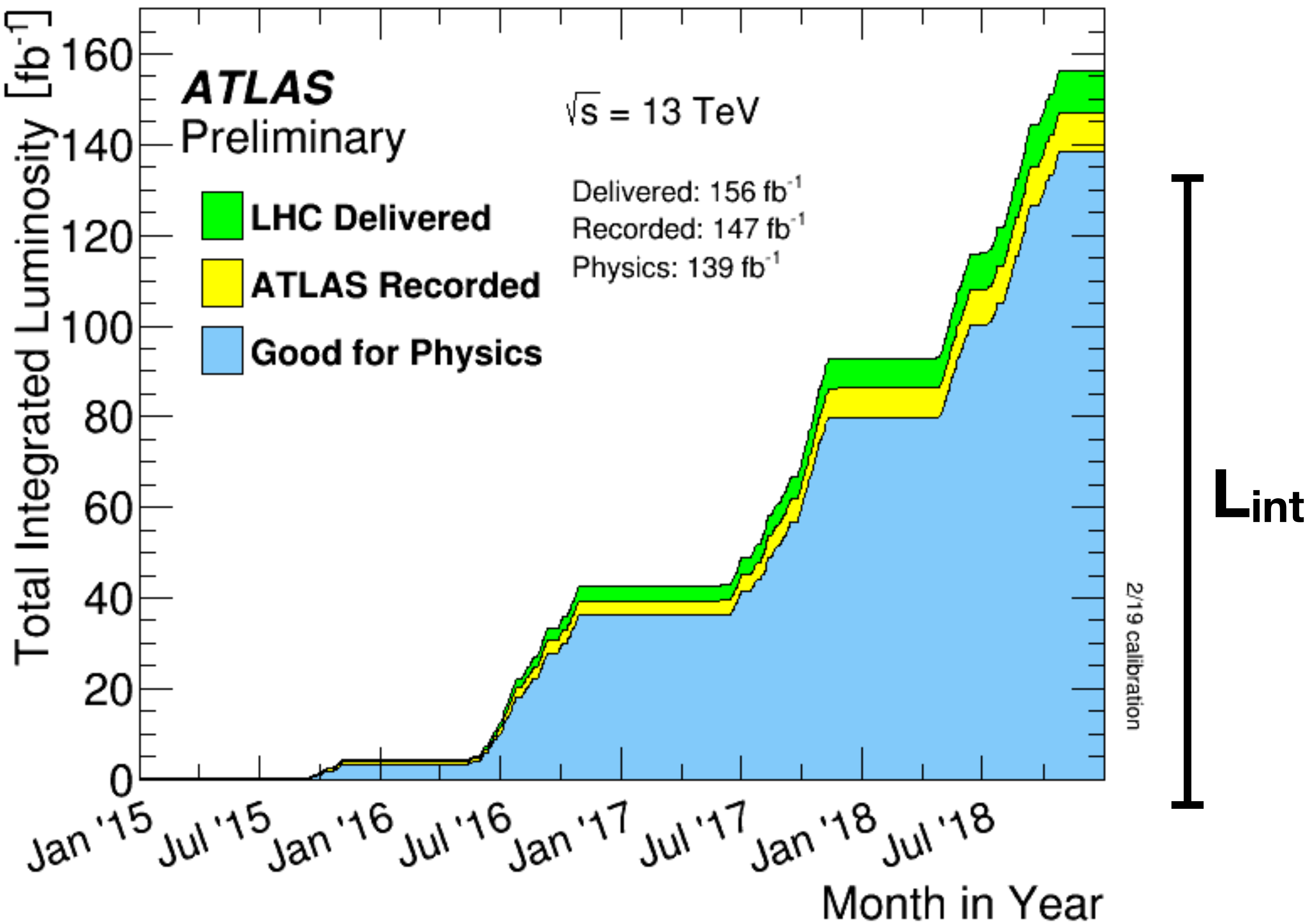
- The cross section for a process is defined as the number of events divided by luminosity

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, ***L_{int}***, which measures how much data we have collected

ATLAS Luminosity

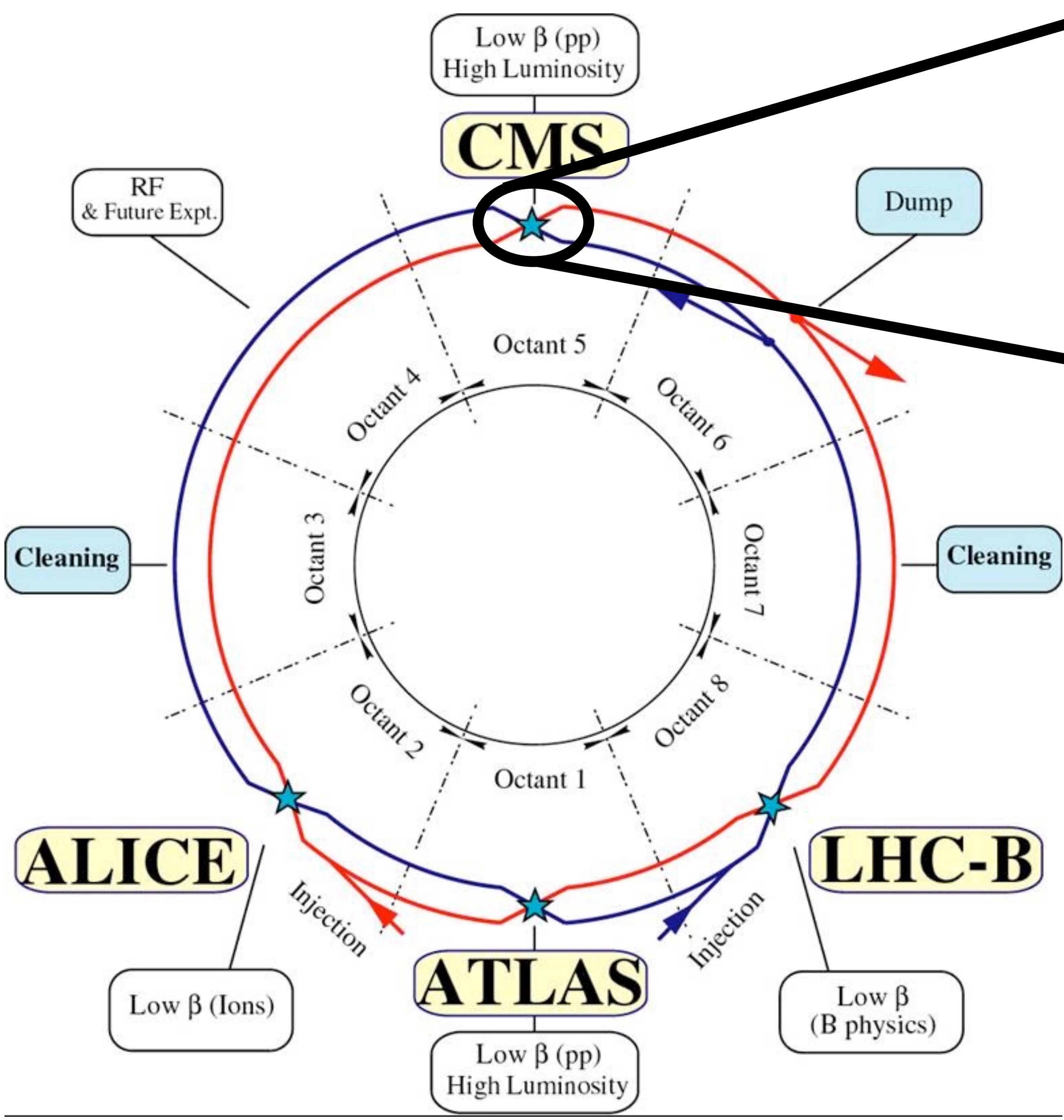


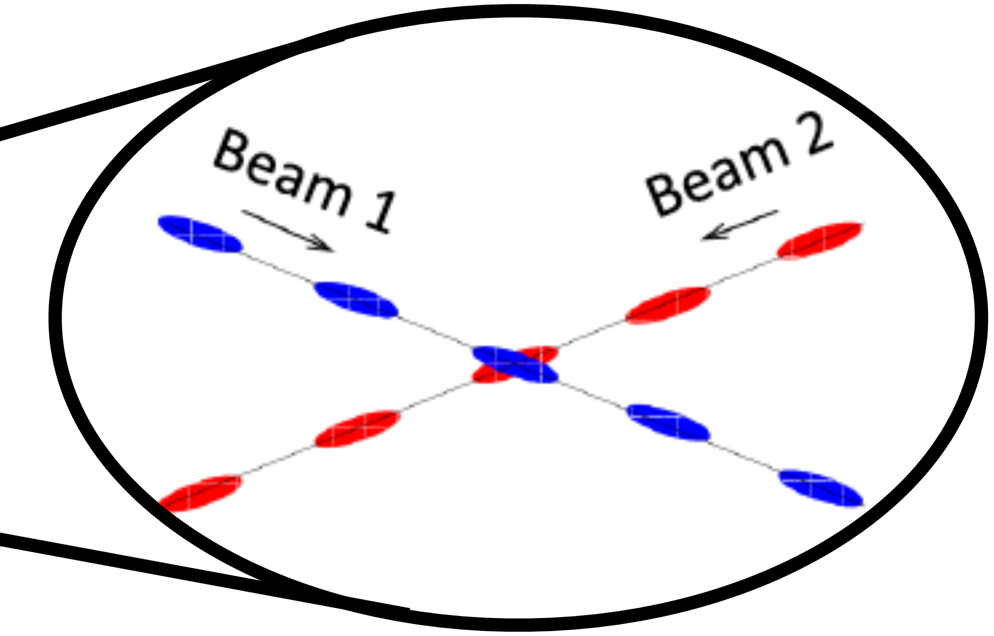
- ***Question: Why does ATLAS record less data than the LHC delivers?***

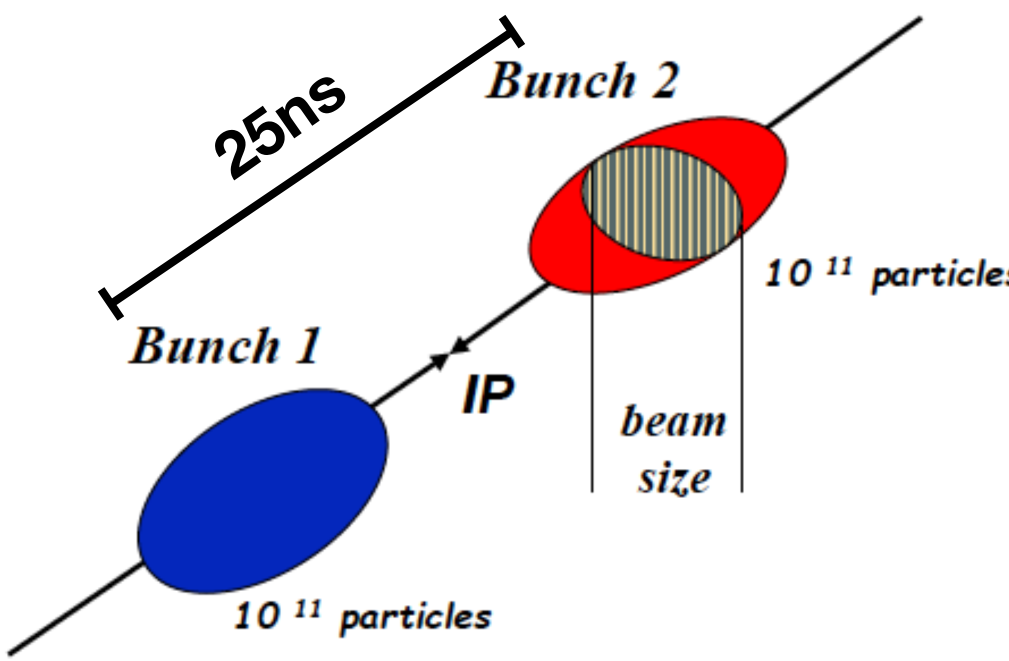
- ***How do we know the integrated luminosity delivered?***

LHC collisions

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

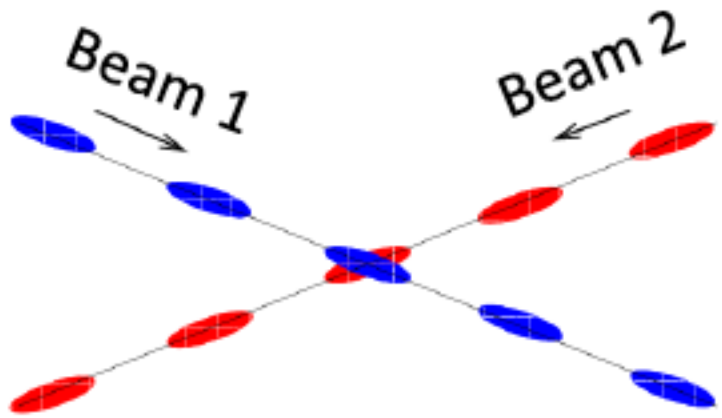


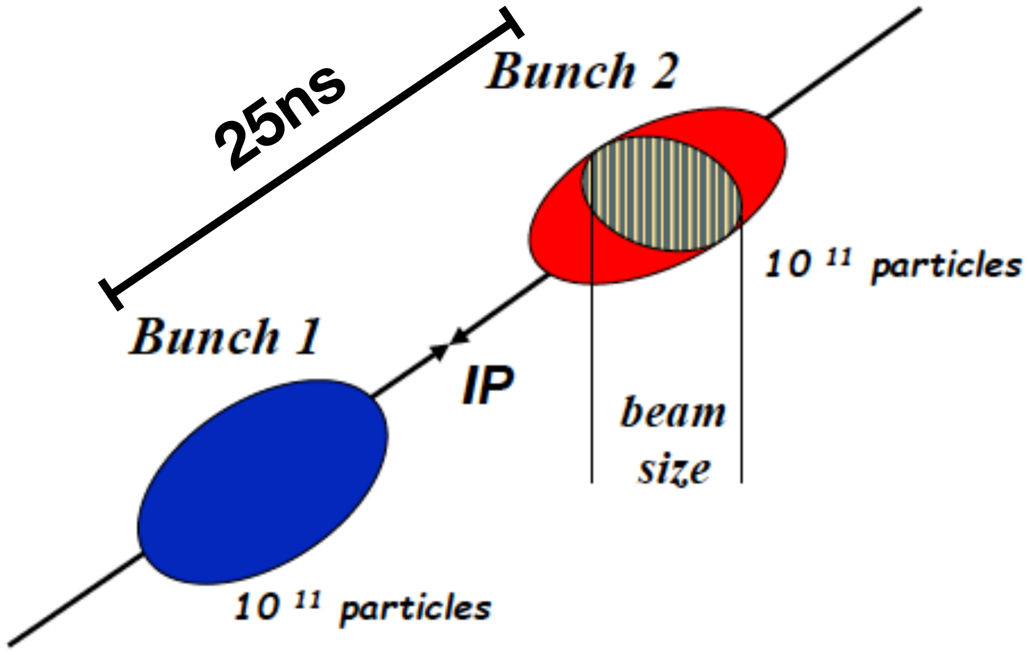




- The LHC accelerates ***bunches of 10^{11} protons*** separated by 25ns gaps

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

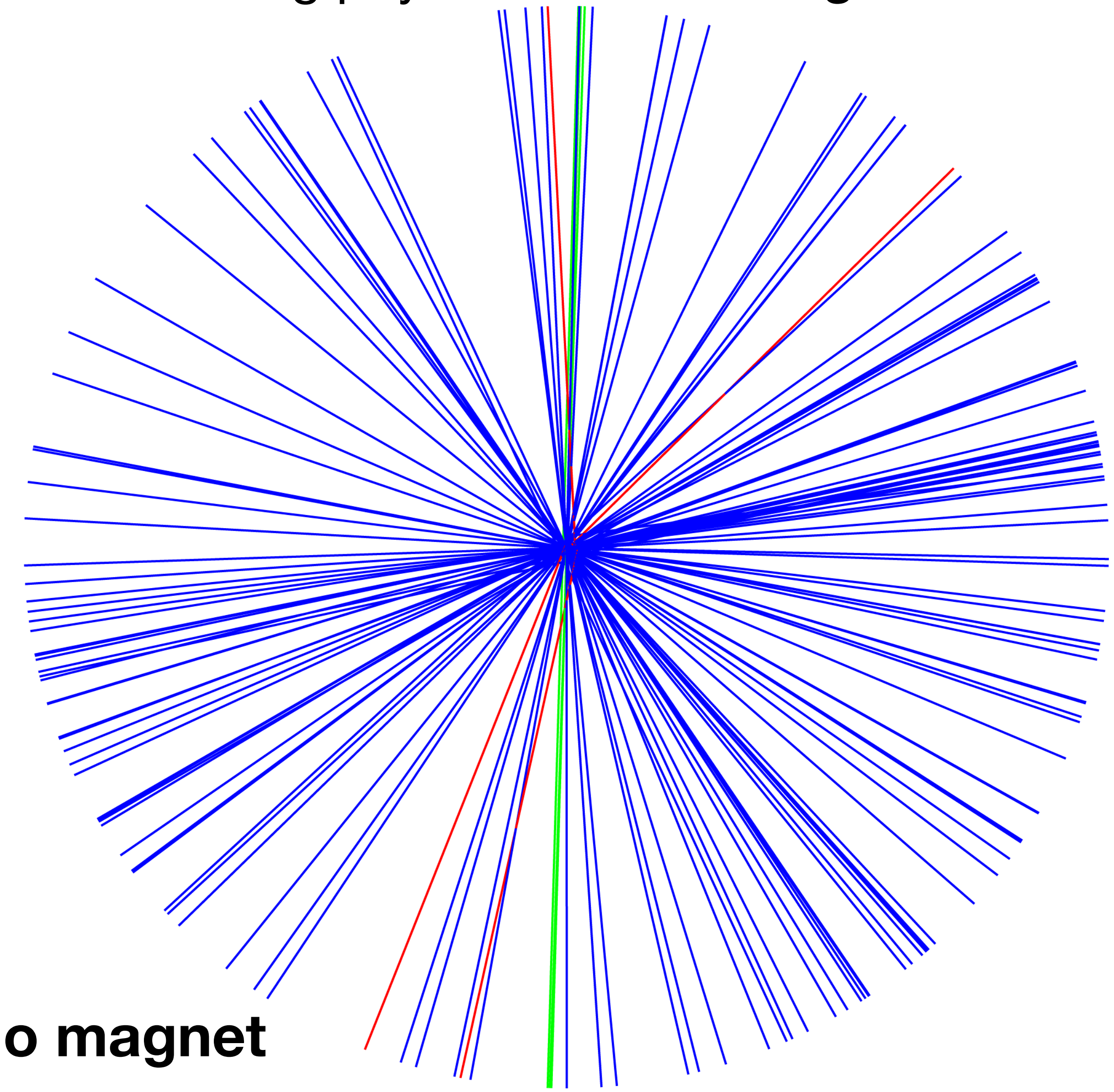
$$\sigma = \frac{N_{obs}}{A\,.\,\epsilon\,.\,L_{int}}$$

- $\boldsymbol{N_{obs}}$ in data needs to be corrected for the detector acceptance, ***A***, for selecting those events. The reconstruction efficiency, $\boldsymbol{\varepsilon}$, 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

Did I mention that simulation is important ?

Before the detector, came the simulation

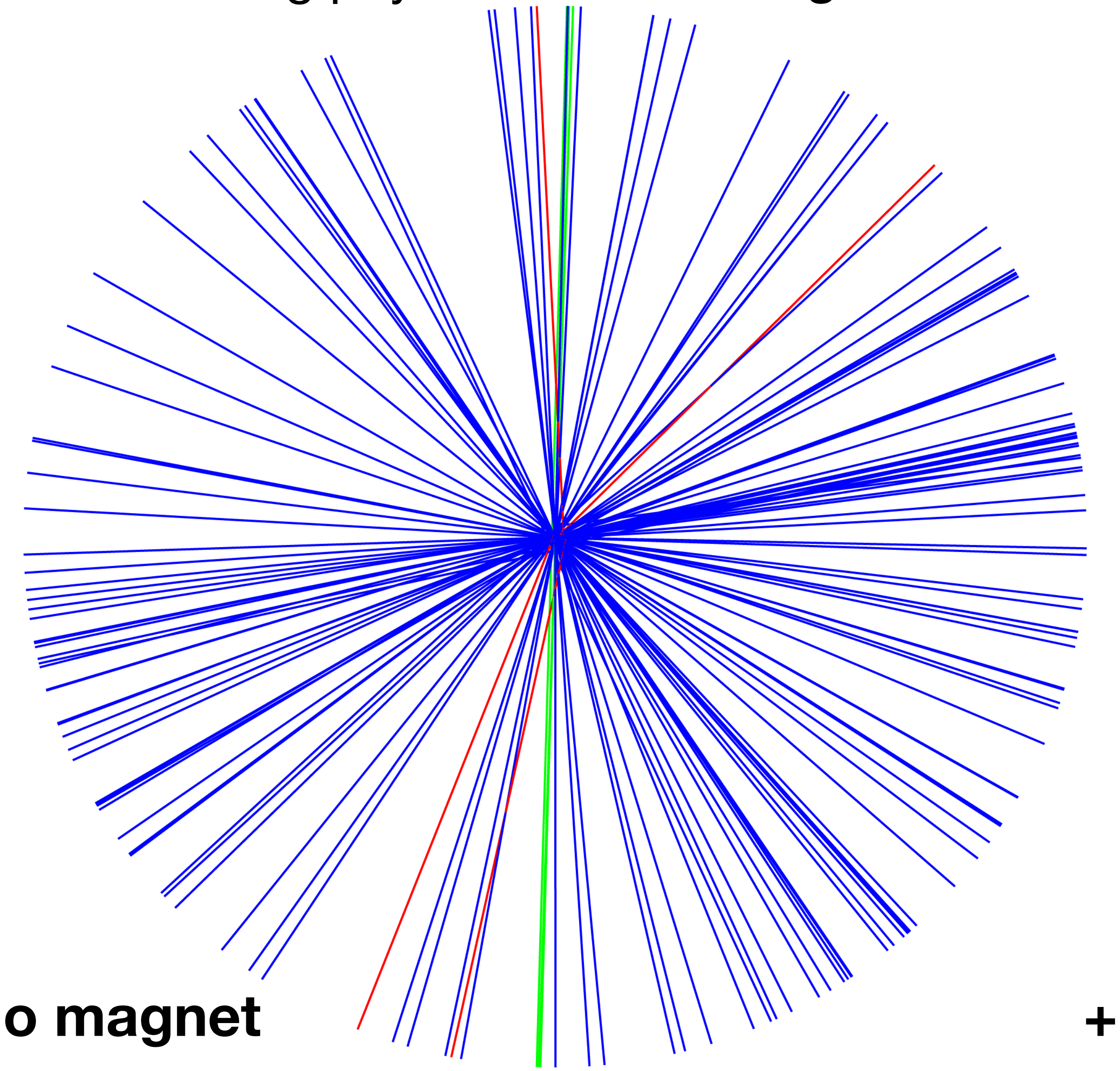
- When designing detectors, we ***simulate detector response*** to physics of interest
- Adding a ***solenoid magnet*** makes it possible to measure momentum (and charge) in our tracker by measuring curvature in the transverse plane
- Interesting physics is often at ***high momentum***, e.g. four high momentum muon tracks here



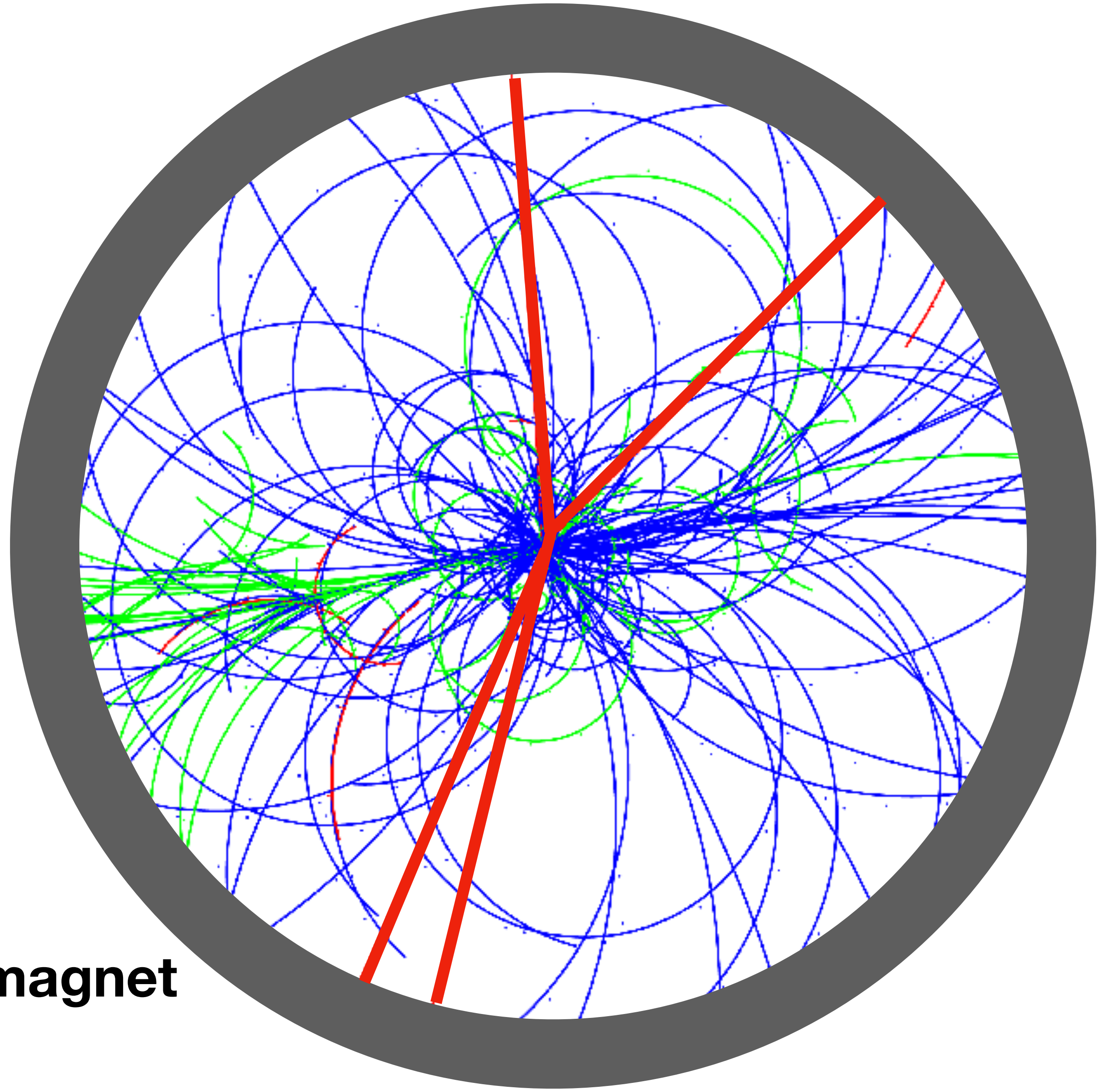
No magnet

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



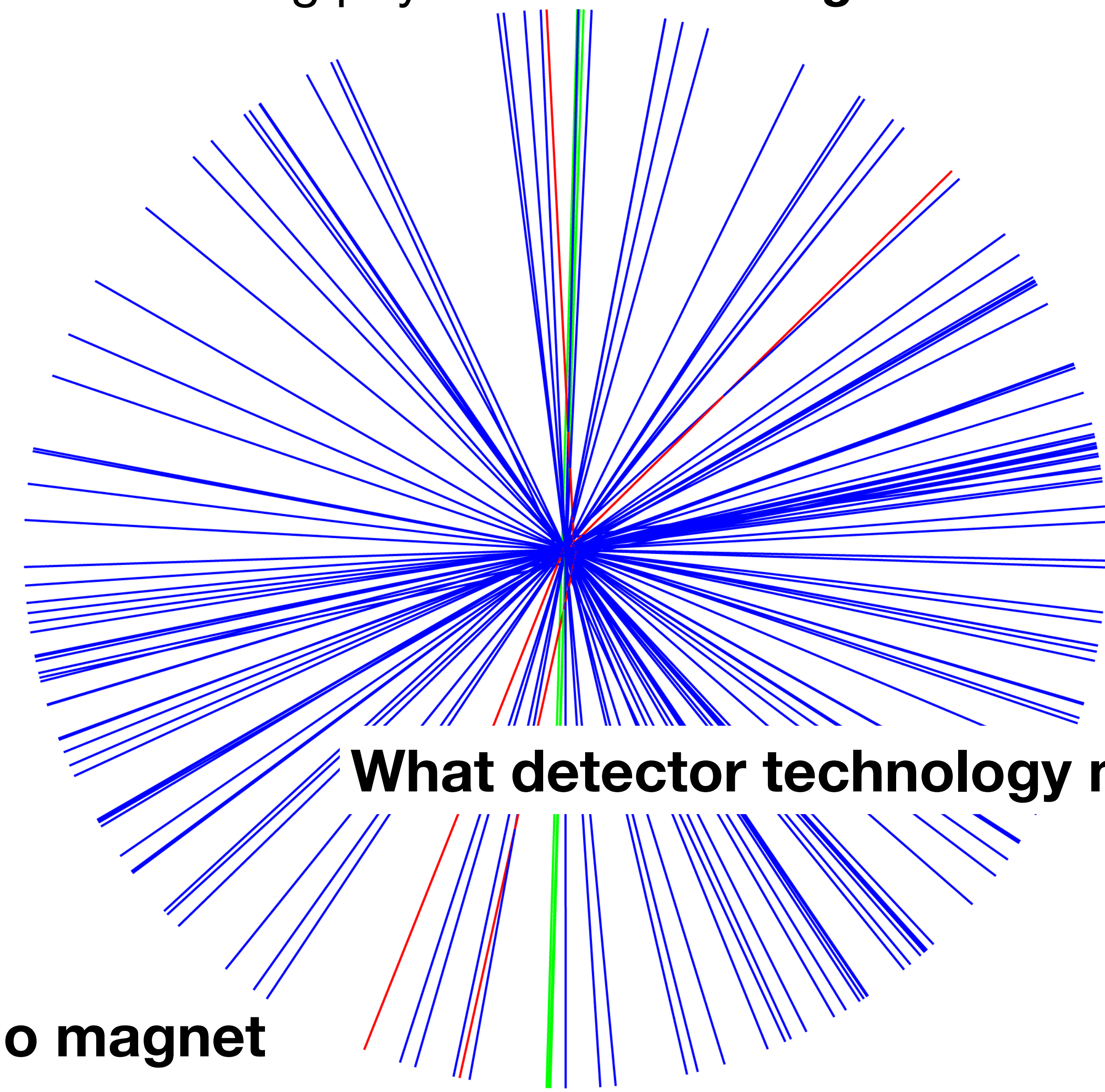
No magnet



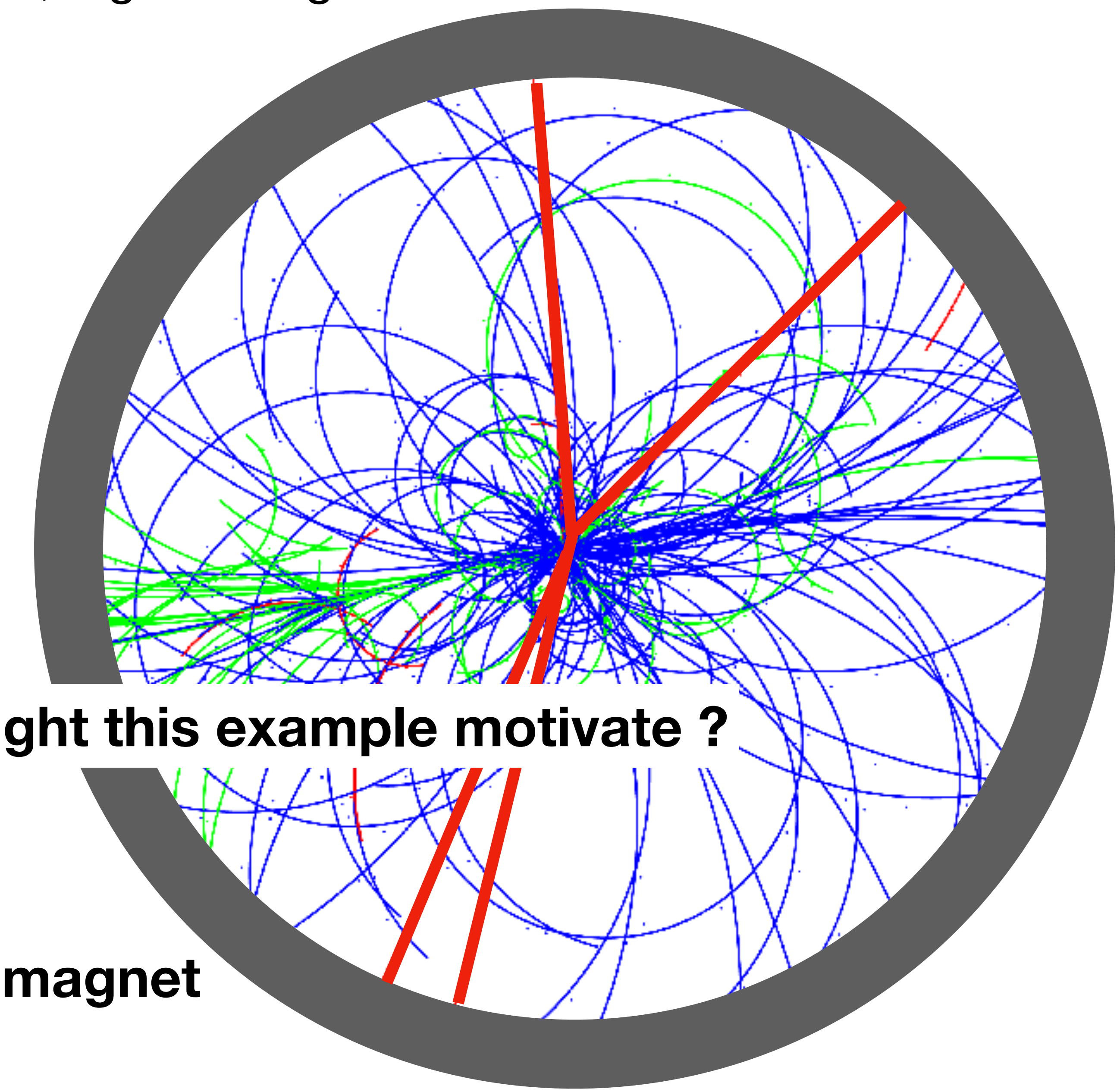
+ magnet

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

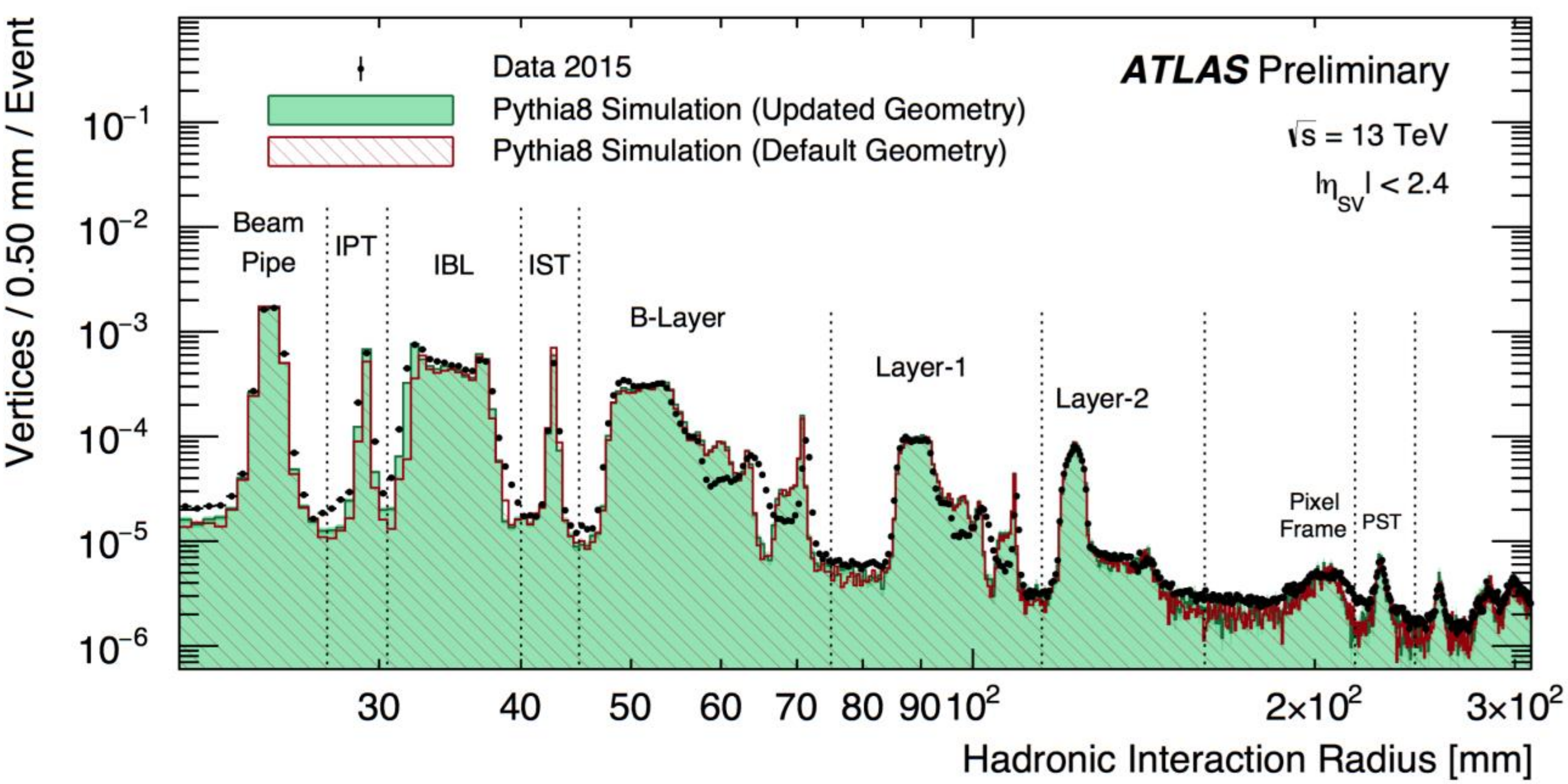


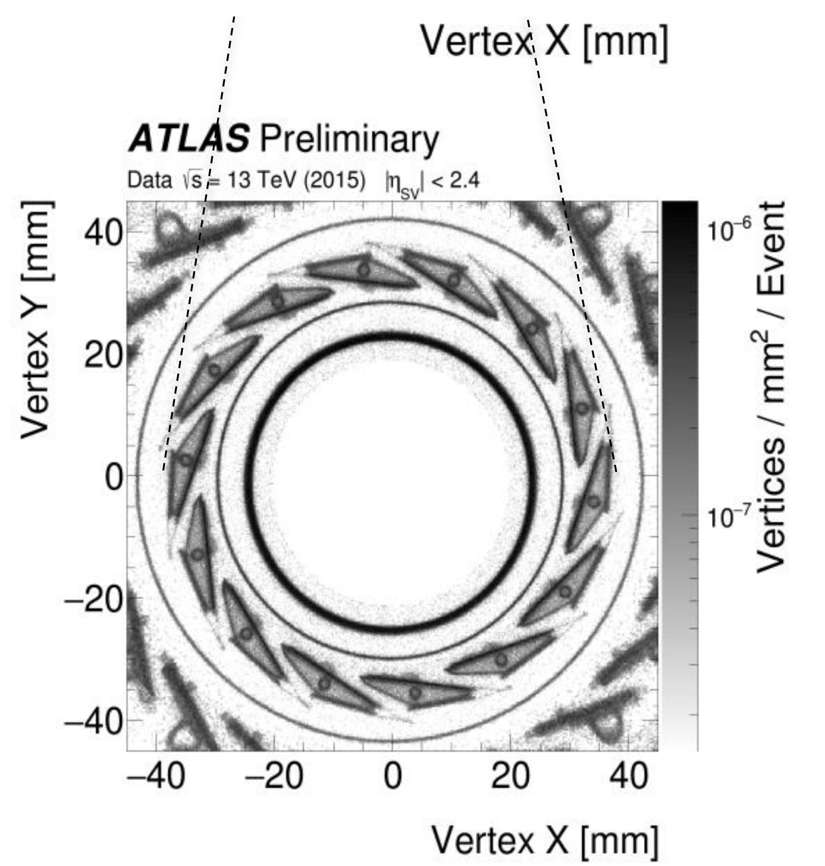
+ magnet

What detector technology might this example motivate ?

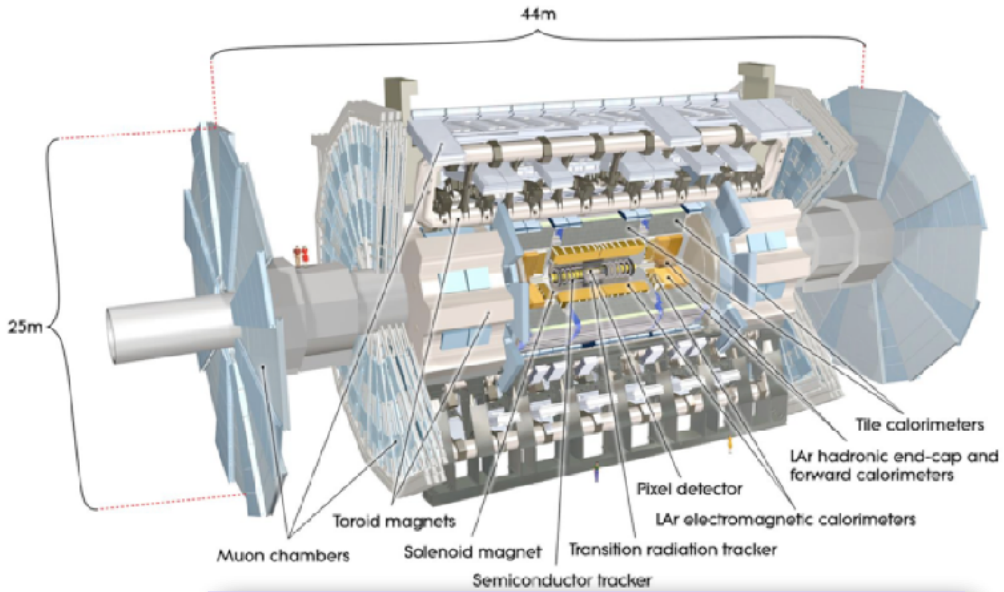
Simulation and understanding detectors

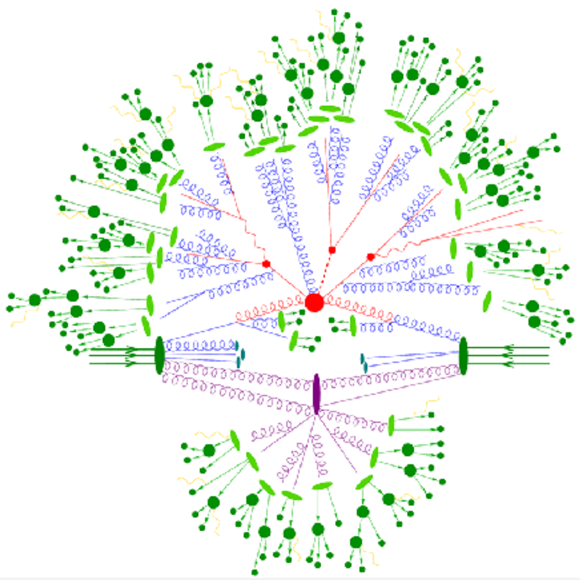
- We use **simulations** to model the detector as **accurately** and **precisely** as possible
- We then ***test*** that our simulations are accurate ***using real data***
- We correct our simulations if necessary
- Once our simulation is an ***accurate model*** of our detector, we can use it to ***correct the data for detector response***





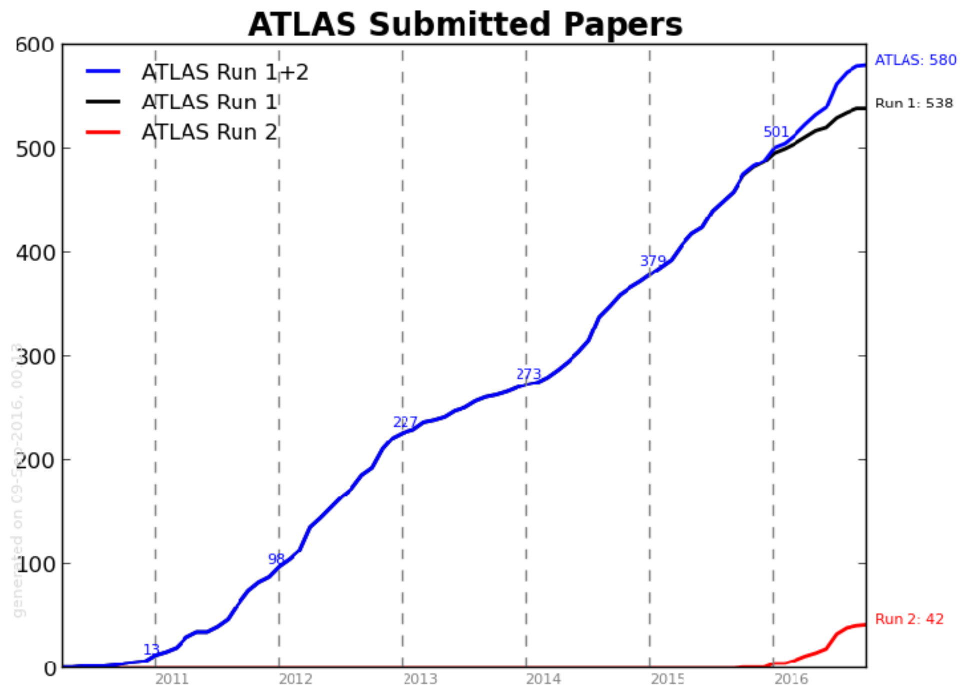
Exabyte-scale physics analysis





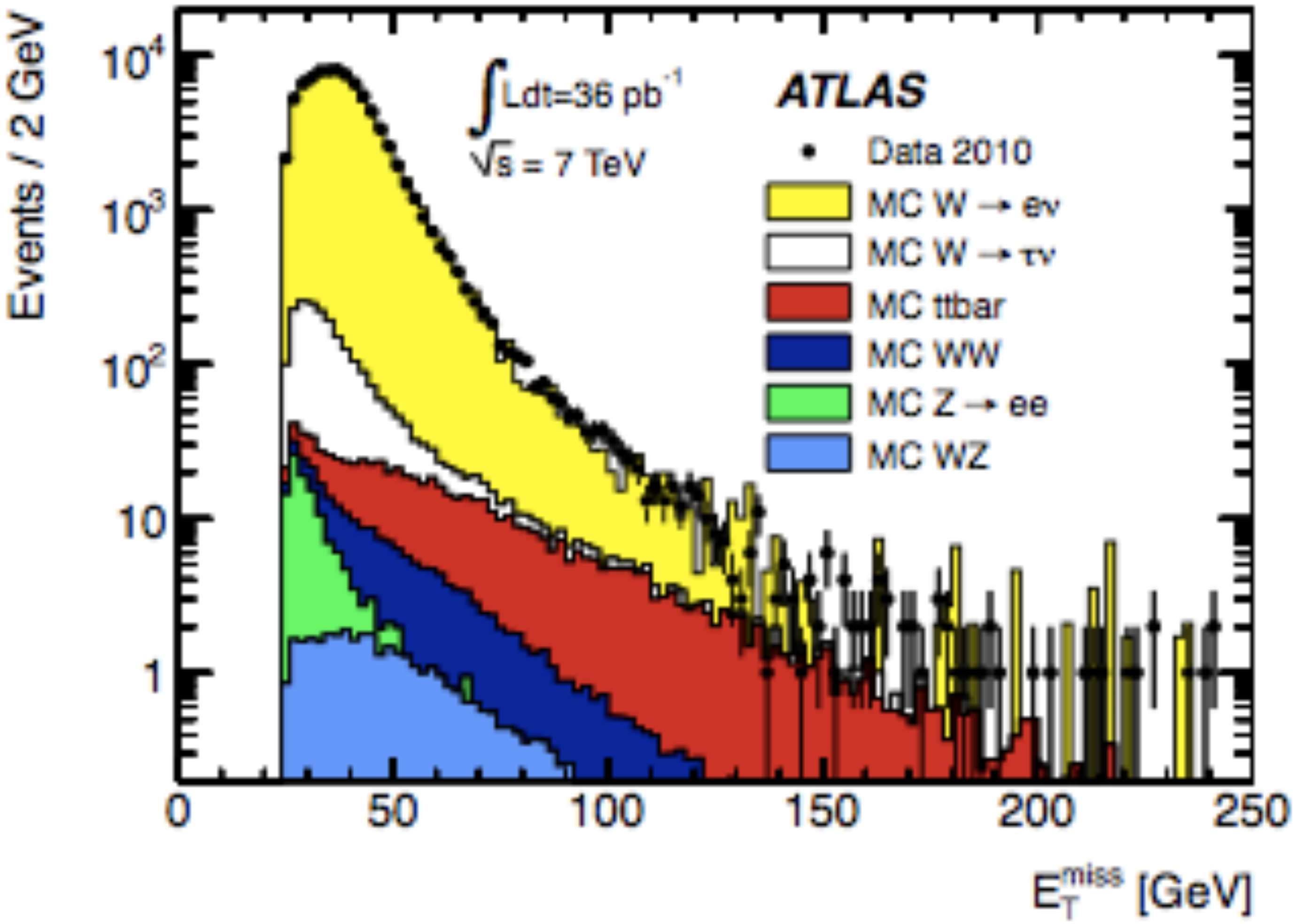
Exabytes of Data

Exabytes of Simulation



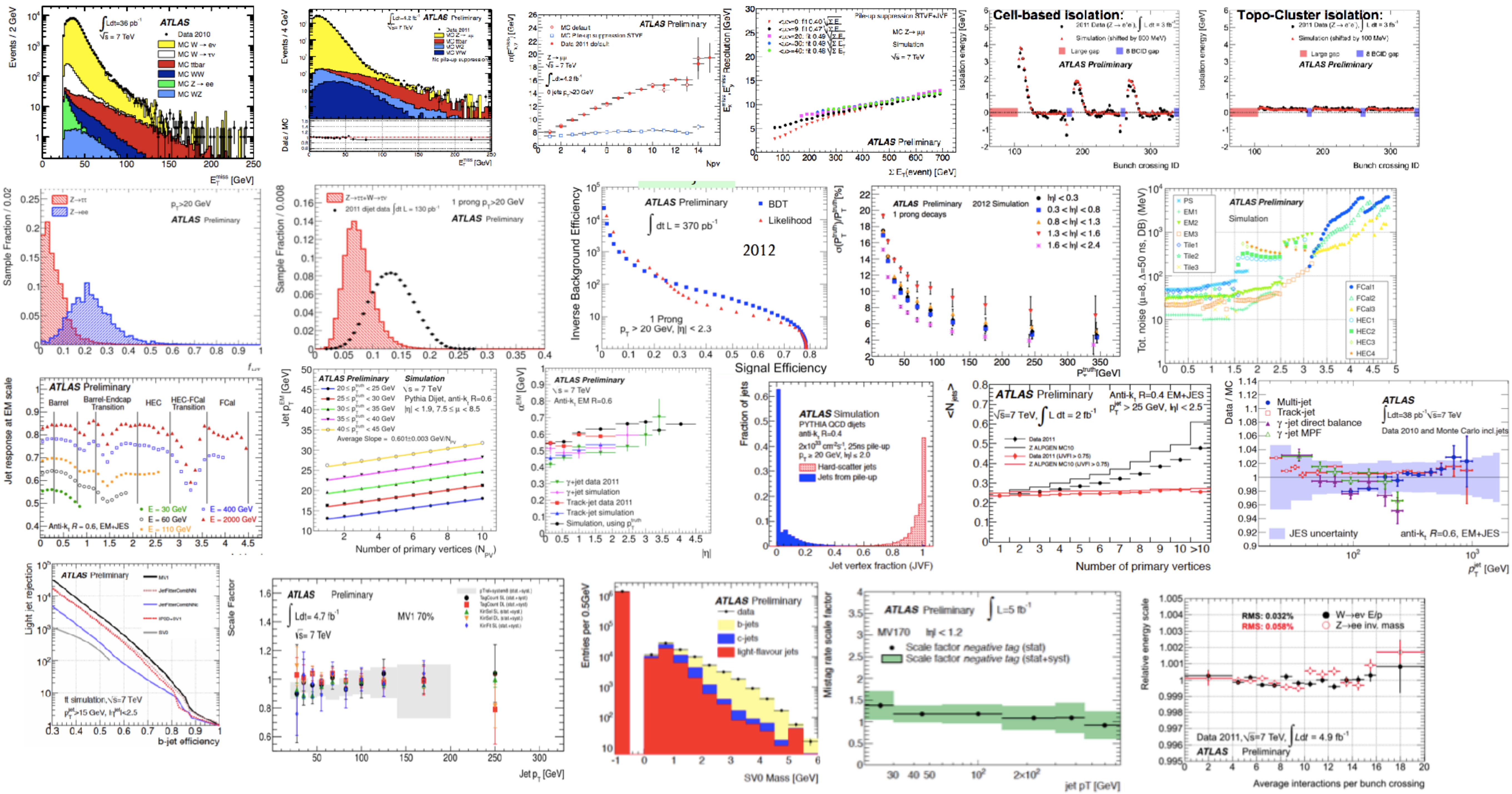
Publish!

Ingredients to the ATLAS physics program



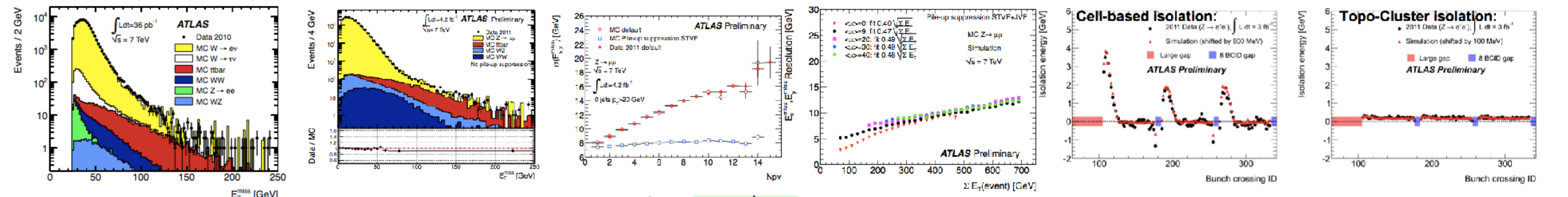
- We compare data with simulation

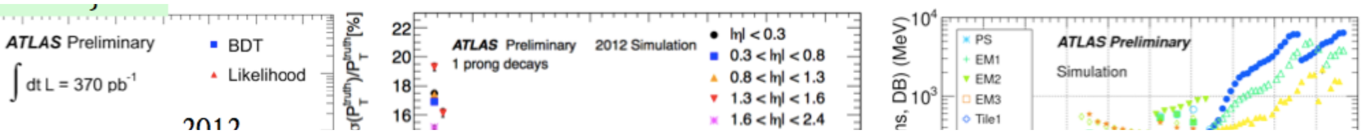
Ingredients to the ATLAS physics program

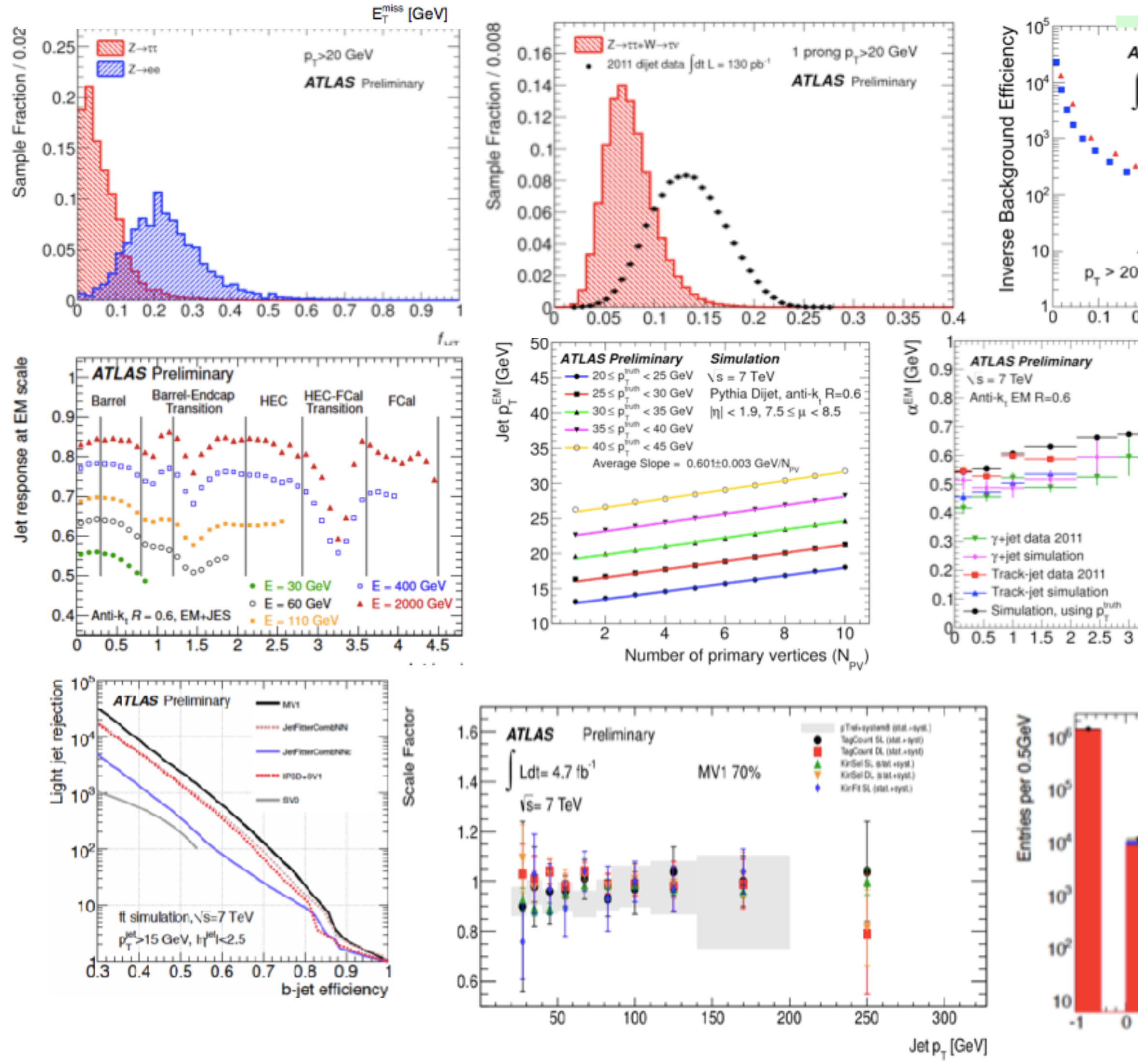


- We make a LOT of comparisons of data and simulation

Ingredients to the ATLAS physics program





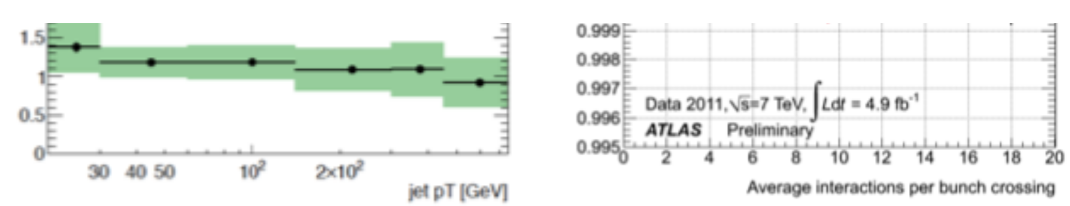


Compare data and simulation in every conceivable way !

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

That's more important than the underlying physics model, in fact it should be independent

WHY?



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

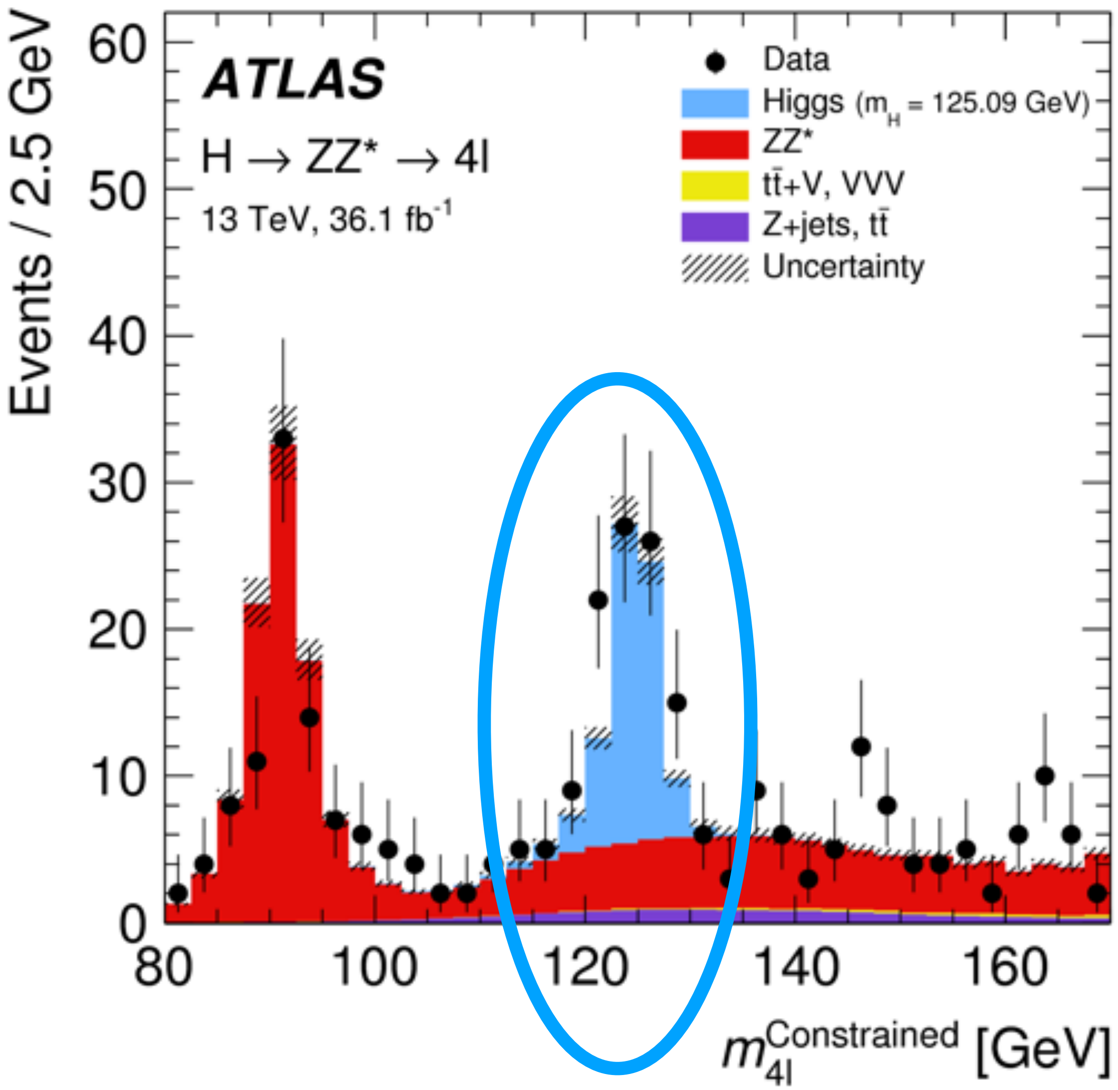
$$\sigma = \frac{N_{obs}}{A\,.\,\epsilon\,.\,L_{int}}$$

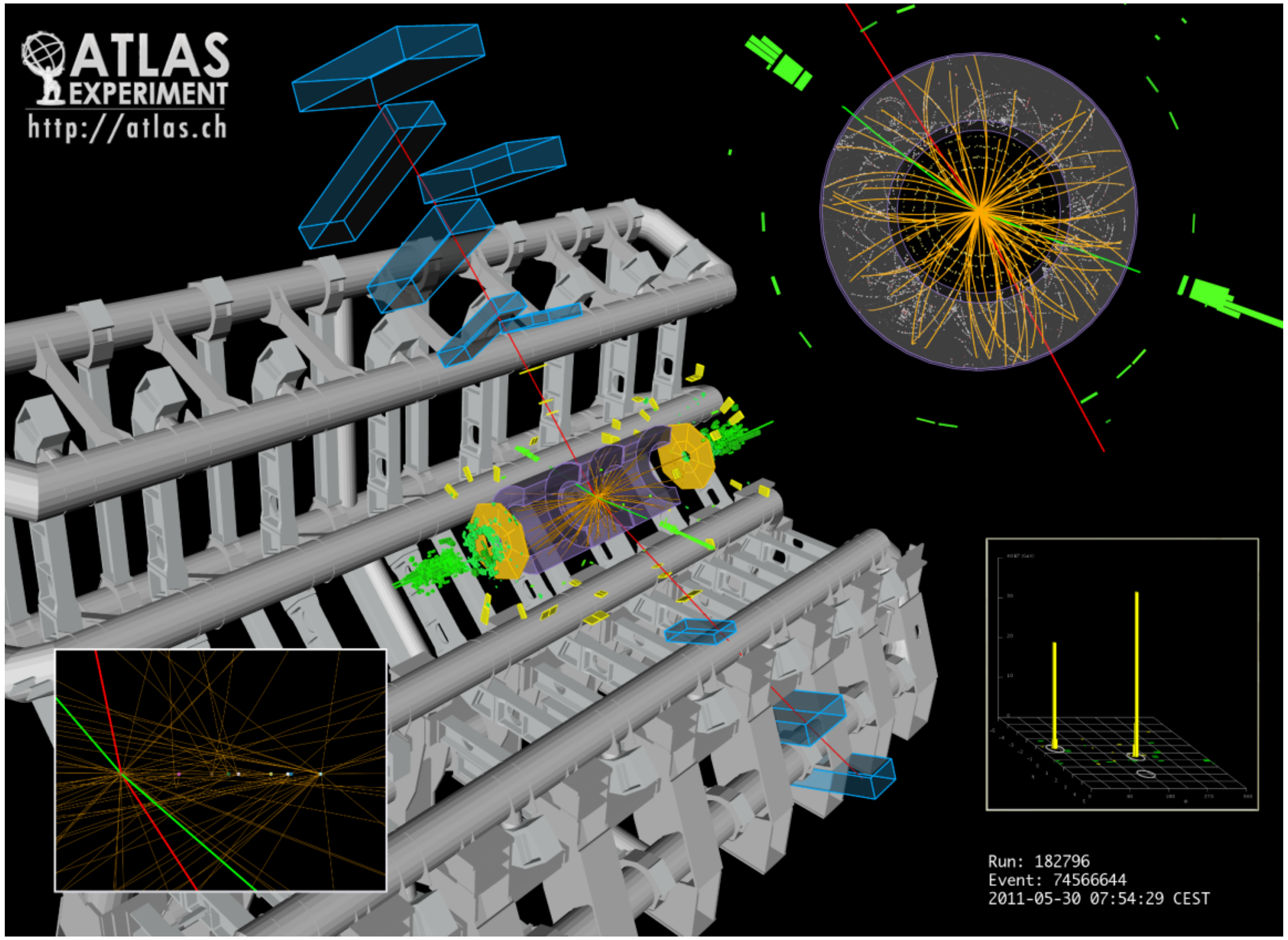
- $\boldsymbol{N_{obs}}$ in data needs to be corrected for the detector acceptance, $\boldsymbol{A}$, for selecting those events. The reconstruction efficiency, $\boldsymbol{\varepsilon}$, is a product of all of the efficiencies that we need to measure and ensure that they are the same in our data and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

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

- Finally, we need to measure and subtract background events that are not part of our signal process

Discovering the Higgs Boson: $H \to ZZ \to 4l$





- We will (nearly) always have some irreducible background to the signal process that we are trying to measure

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

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

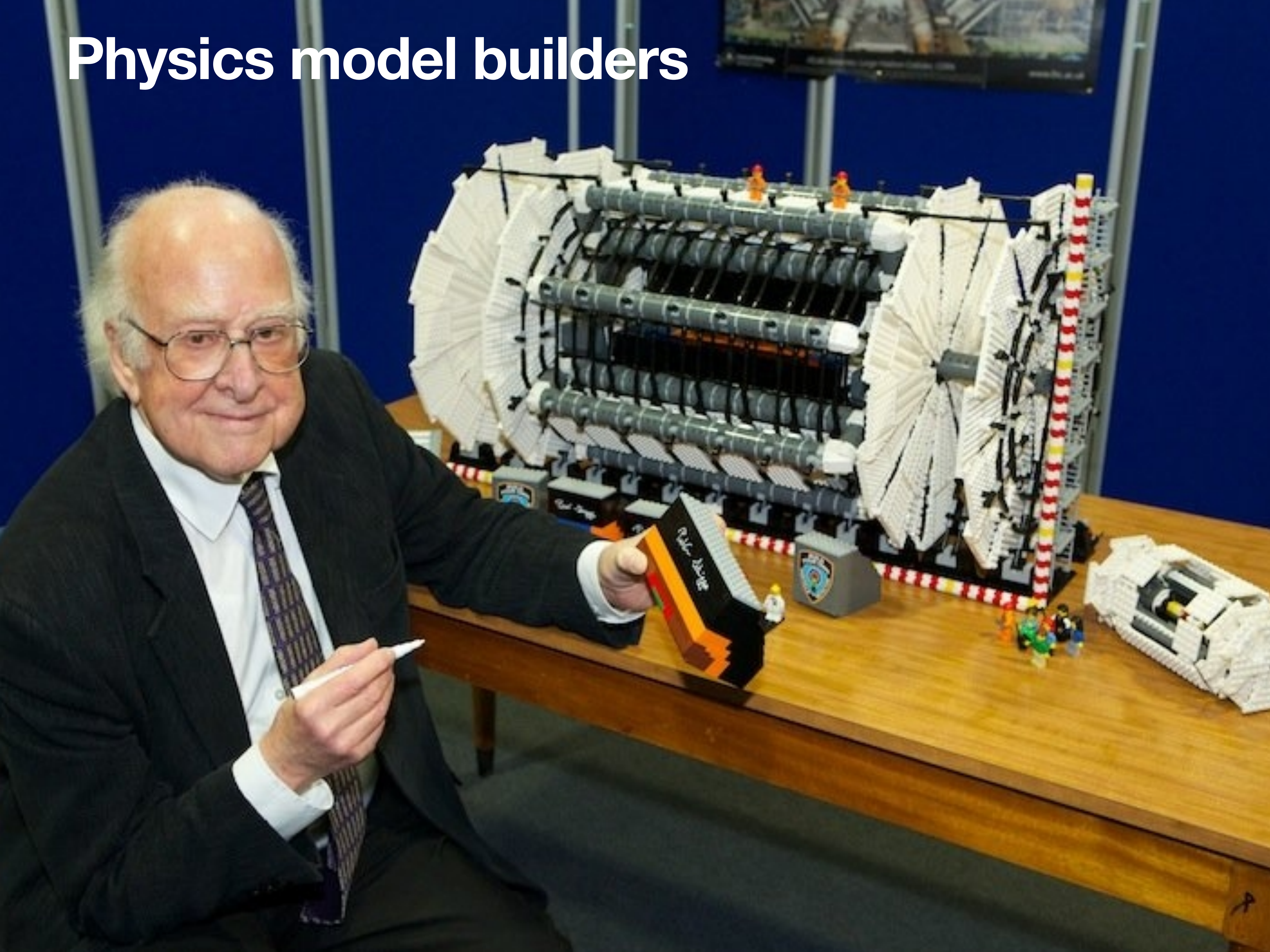
- $\boldsymbol{N_{obs}}$ in data needs to be corrected for the detector acceptance, $\boldsymbol{A}$, for selecting those events. The reconstruction efficiency, $\boldsymbol{\varepsilon}$, is a product of all of the efficiencies that we need to measure and ensure that they are the same in our data and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

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

- Finally, we need to measure and subtract background events that are not part of our signal process

Now we can compare this to the theoretical cross section!

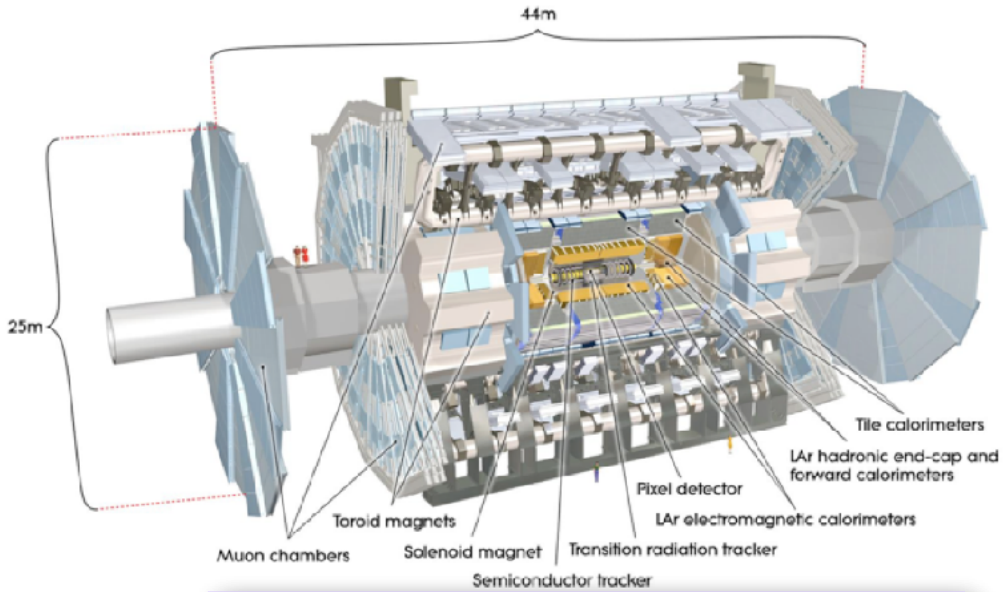
Physics model builders

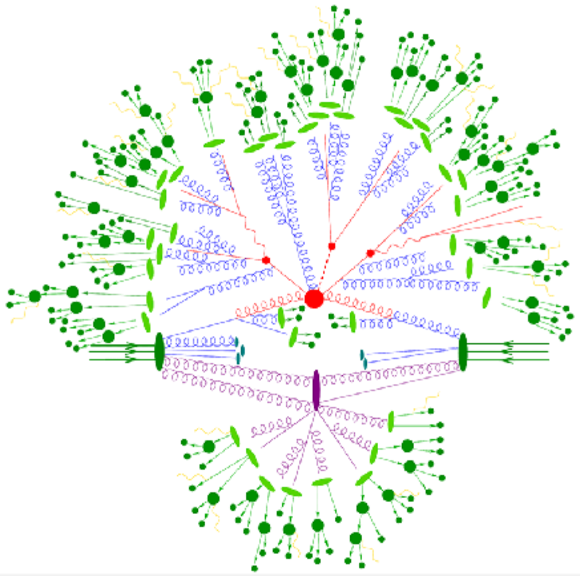


Physics event generators

- 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 !)

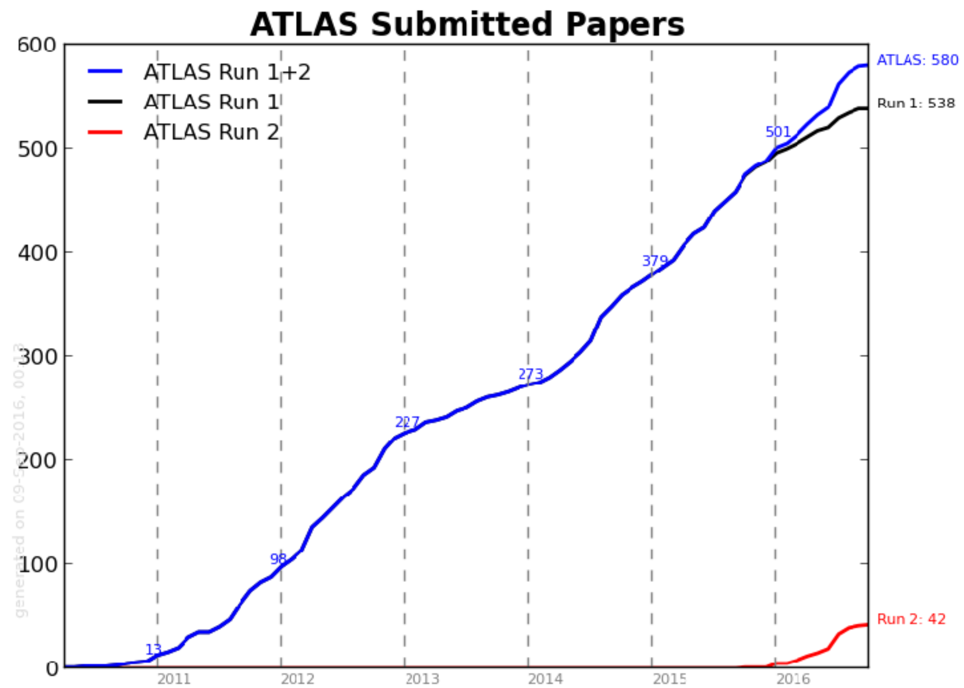
Are we ready to do some exabyte-scale physics analysis?





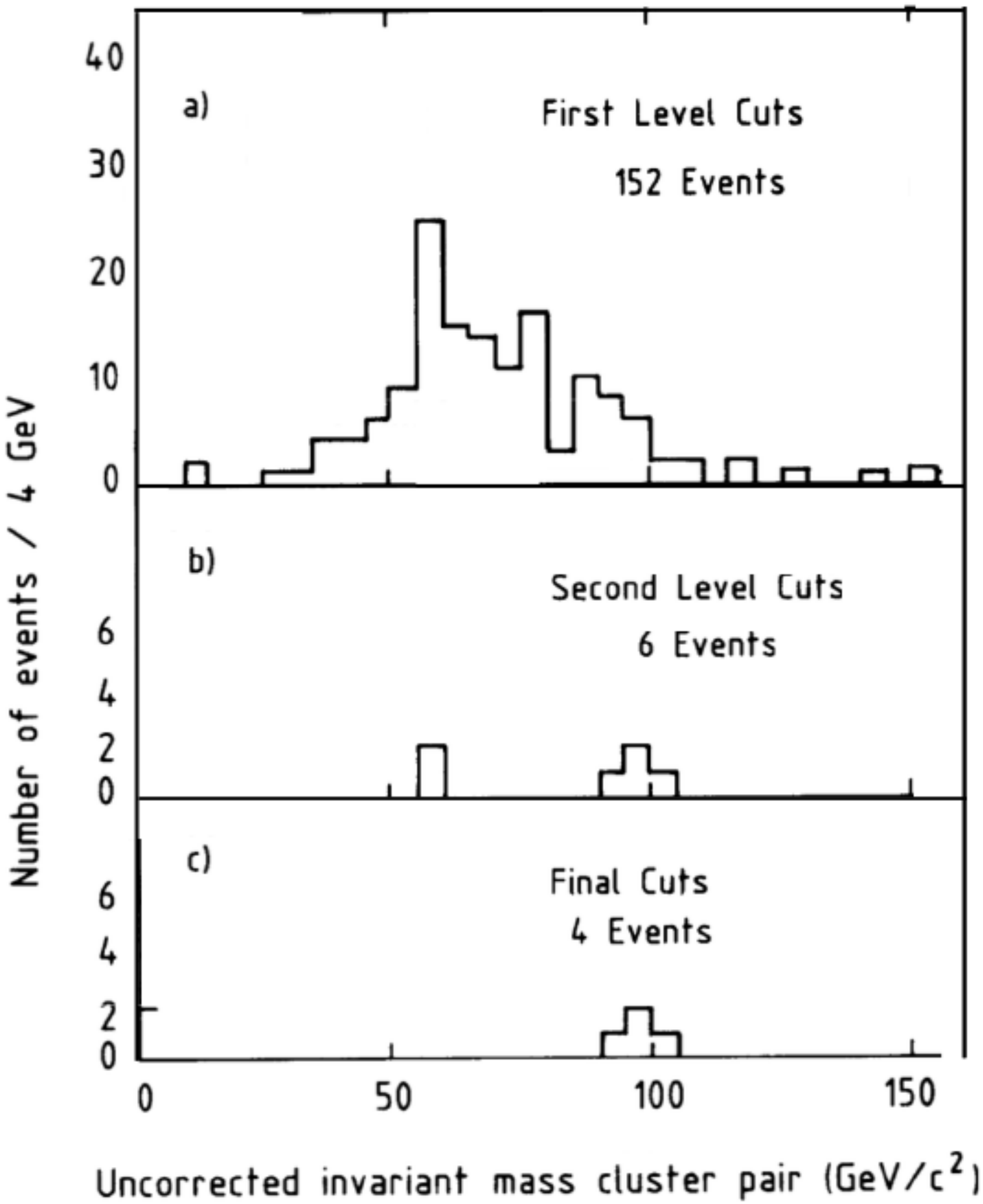
Exabytes of Data

Exabytes of Simulation



Publish!

First - measuring the Z boson



Z->ee in UA1

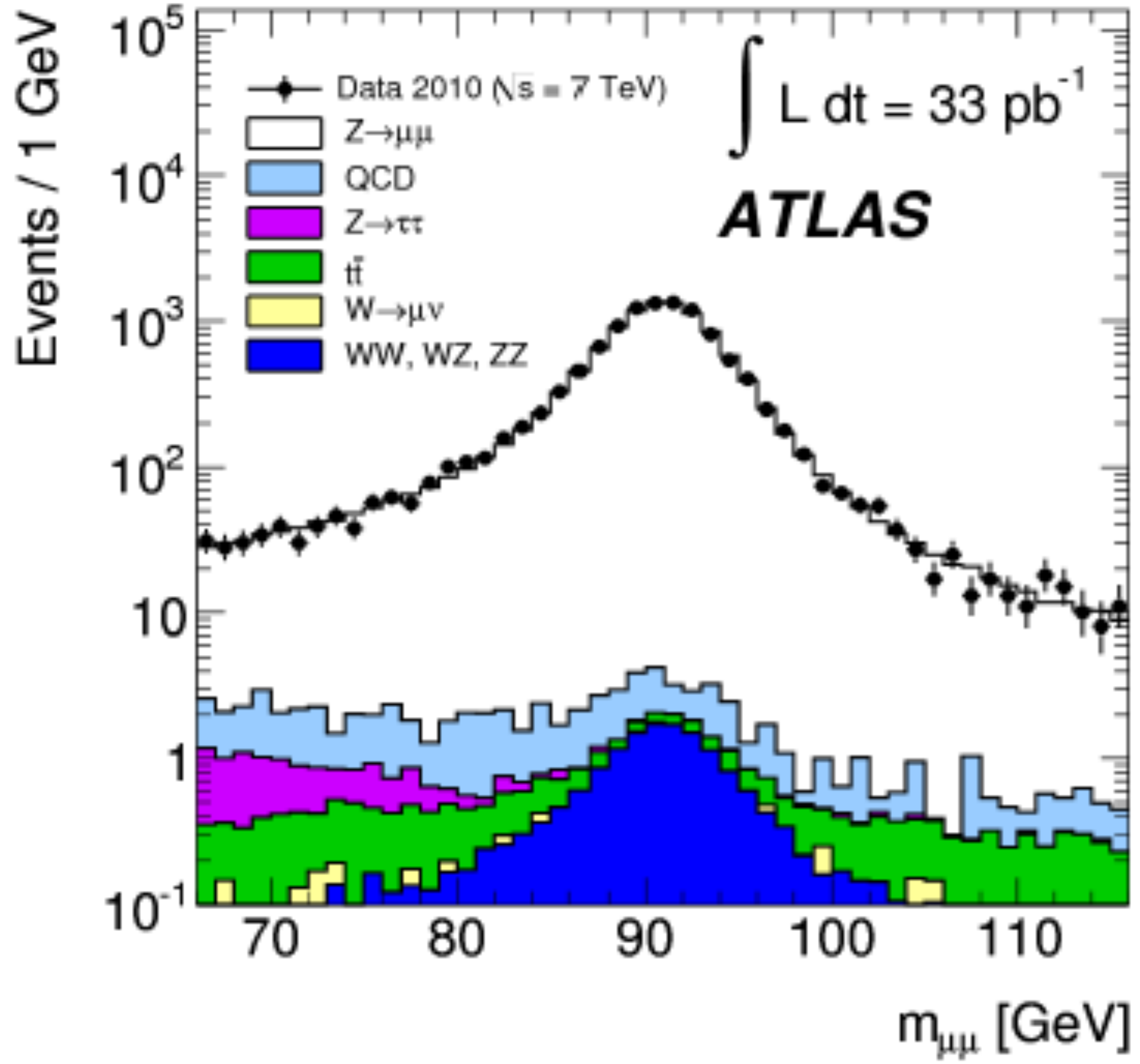
Two EM clusters with E_T>25GeV.

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.

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?**

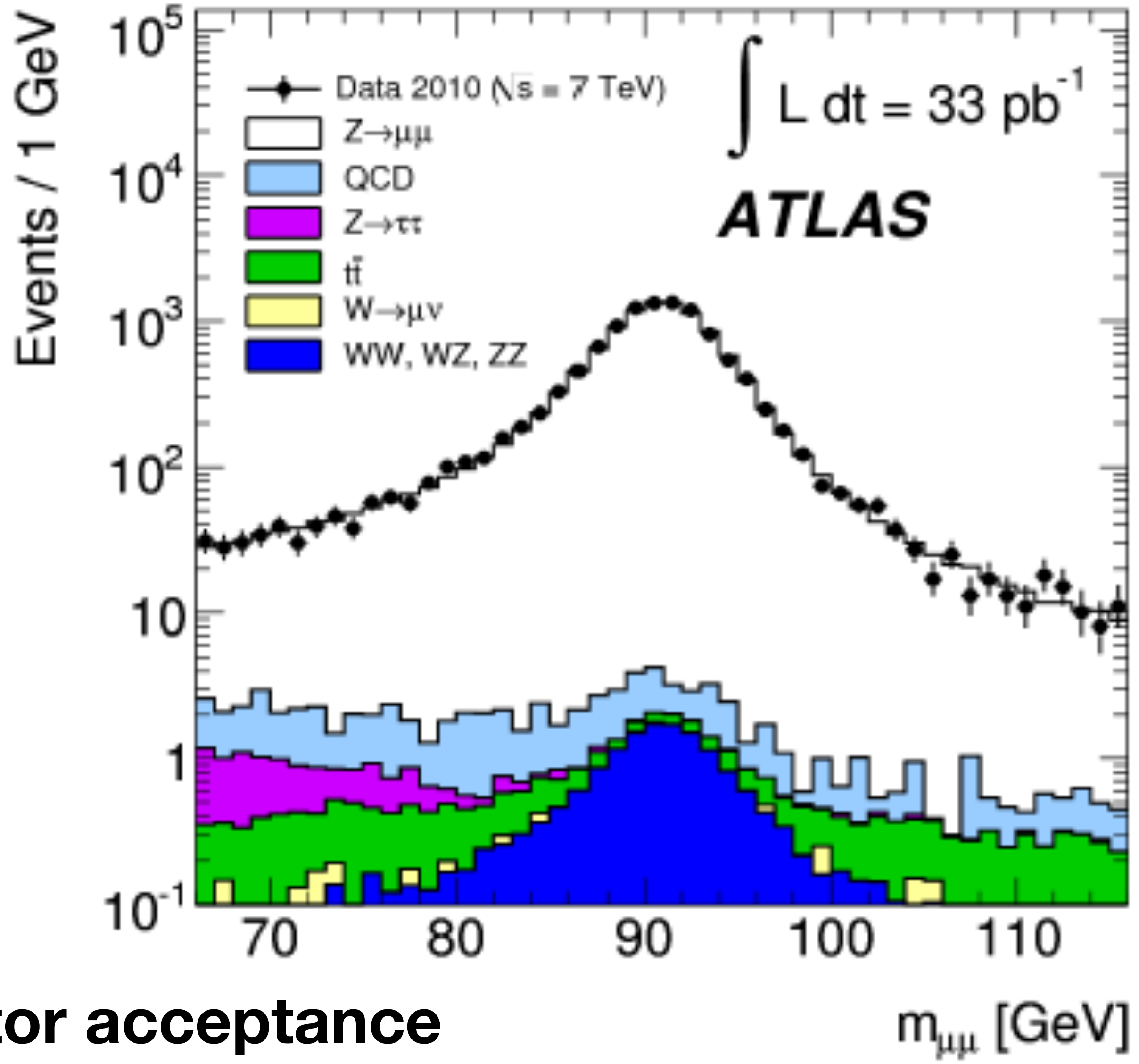
GRAVITATIONAL WAVE SCIENCE CENTER

UNIVERSITÉ DE GENÈVE

Faculté des Sciences
Département d'astronomie

Measuring the Z boson at ATLAS

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



- 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?***

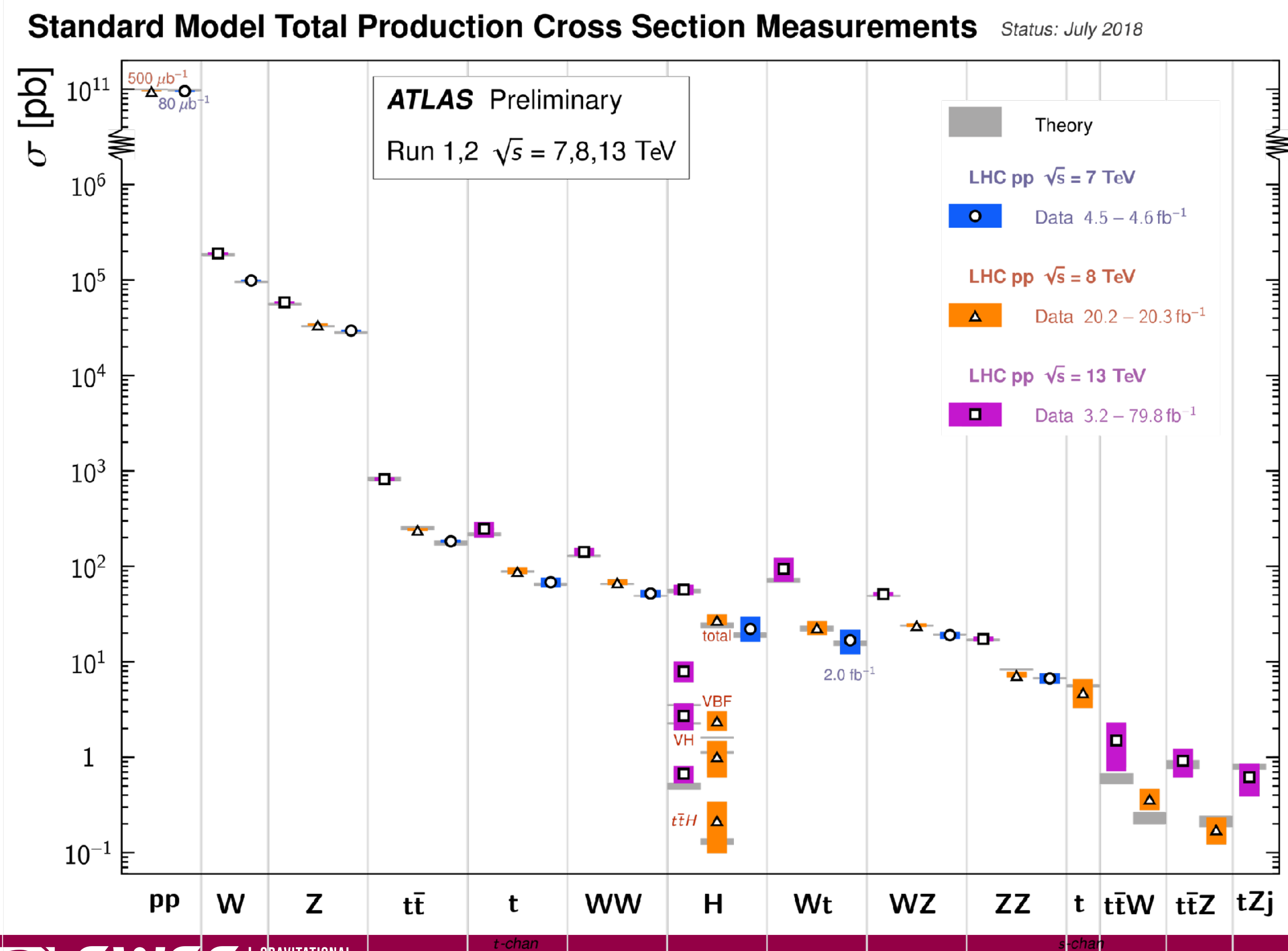
Measuring the Z boson at ATLAS

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

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

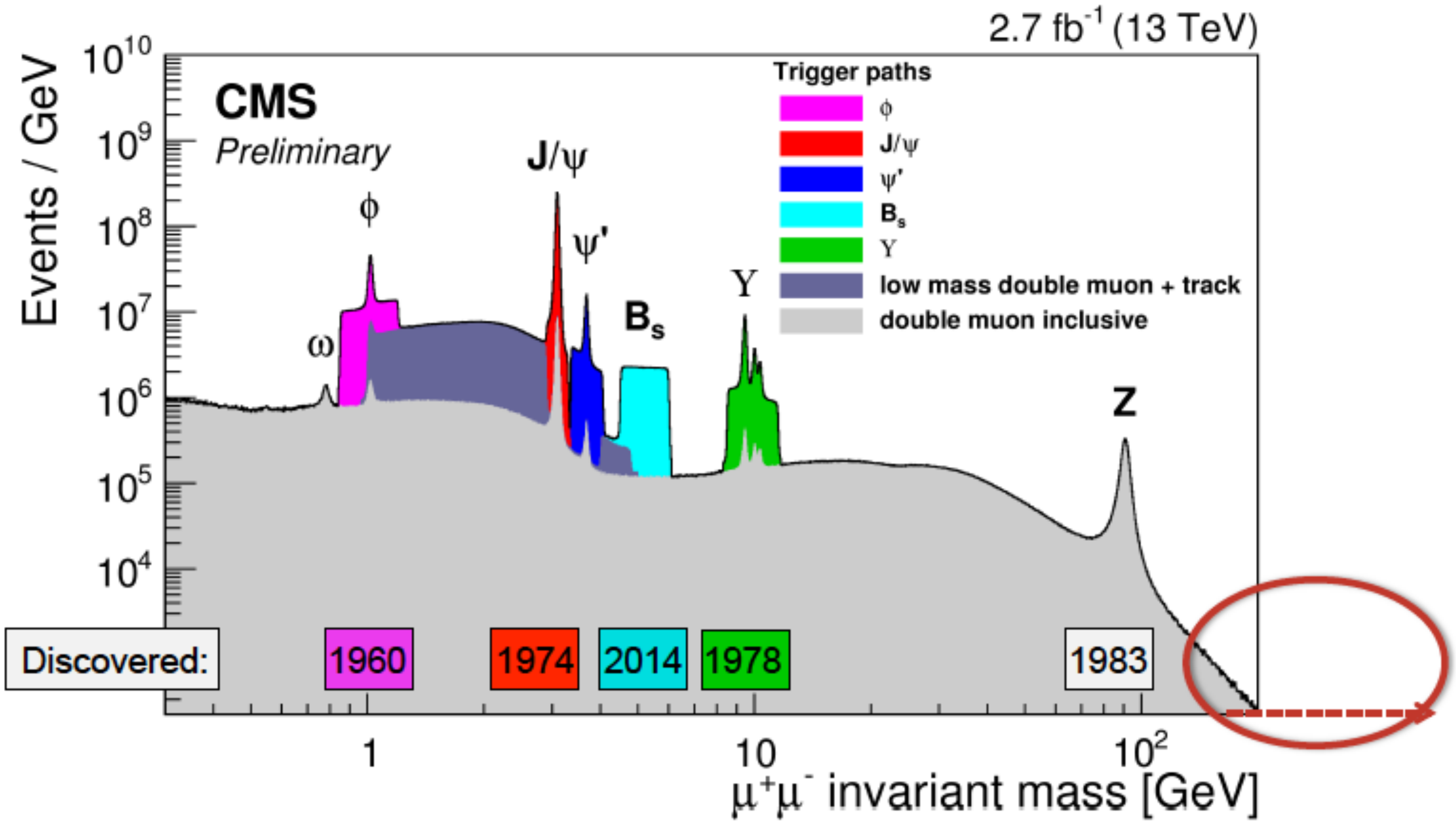
		$Z \to e^+e^-$				$Z \to \mu^+\mu^-$			
$\lvert y_{\ell\ell}\rvert^{min}$	$\lvert y_{\ell\ell}\rvert^{max}$	$d\sigma/d\lvert y_{\ell\ell}\rvert$ [pb]	$\delta\sigma_{stat}$ [pb]	$\delta\sigma_{syst}$ [pb]	$\delta\sigma_{lumi}$ [pb]	$d\sigma/d\lvert y_{\ell\ell}\rvert$ [pb]	$\delta\sigma_{stat}$ [pb]	$\delta\sigma_{syst}$ [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
- $\boldsymbol{N_{obs}}$ has a ***statistical*** uncertainty
- $\boldsymbol{N_{bkg}}$ 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 used
- Finally, $\boldsymbol{L_{int}}$ also has an uncertainty that dictates how well we know the absolute scale of the measurement - a ***normalisation*** uncertainty



Standard Model Total Production Cross Section Measurements
Status: July 2018
ATLAS Preliminary
Run 1,2 $\sqrt{s}$ = 7,8,13 TeV
σ [pb]
10^{11}
10^6
10^5
10^4
10^3
10^2
10^1
1
10^{-1}
500 μb^{-1}
80 μb^{-1}
Theory
LHC pp $\sqrt{s}$ = 7 TeV
Data 4.5 – 4.6 fb^{-1}
LHC pp $\sqrt{s}$ = 8 TeV
Data 20.2 – 20.3 fb^{-1}
LHC pp $\sqrt{s}$ = 13 TeV
Data 3.2 – 79.8 fb^{-1}
total
VBF
VH
$t\bar{t}H$
2.0 fb^{-1}
pp
W
Z
$t\bar{t}$
t
t-chan
WW
H
Wt
WZ
ZZ
t
s-chan
$t\bar{t}W$
$t\bar{t}Z$
tZj

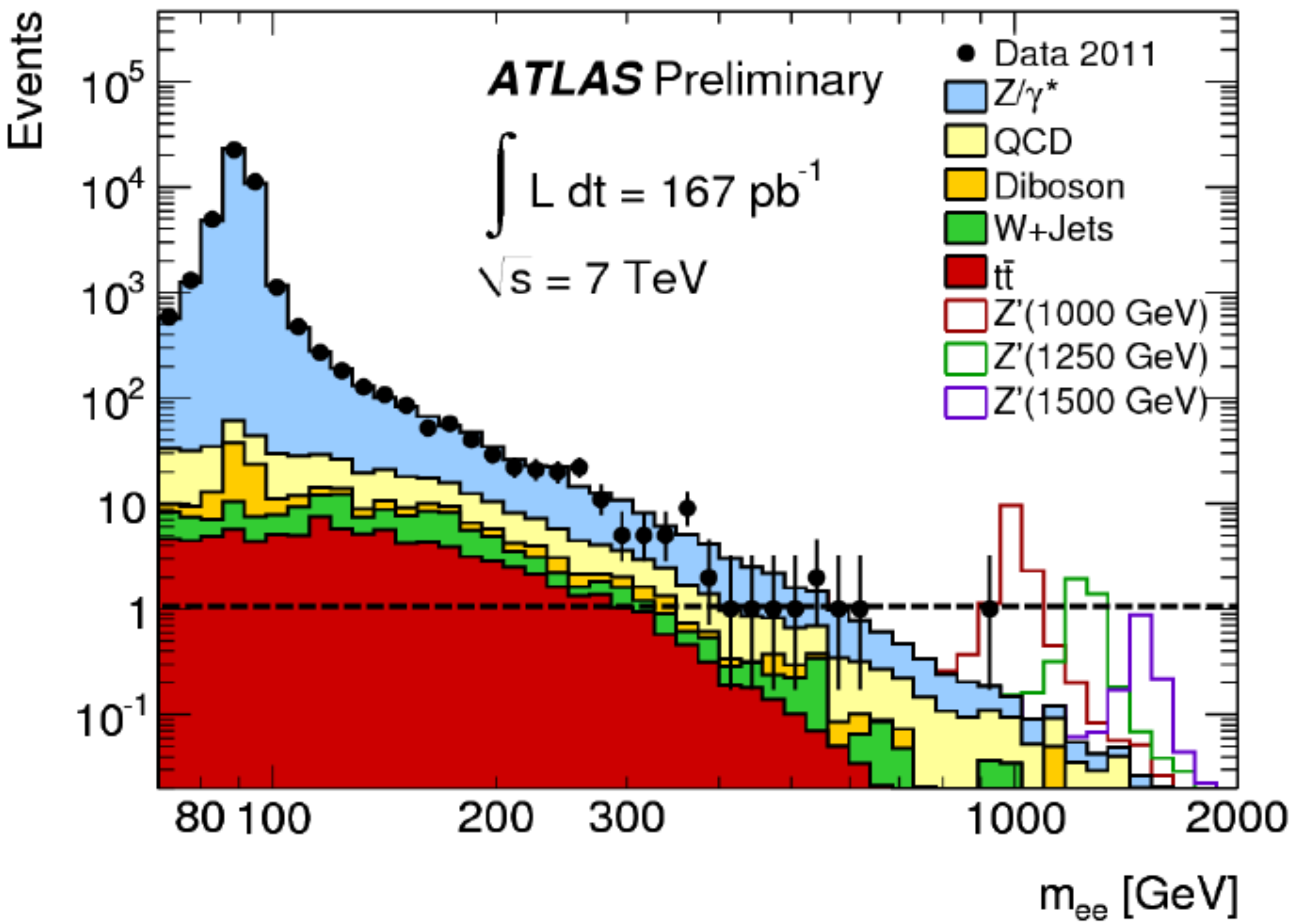
Elements of a search



What's out there?

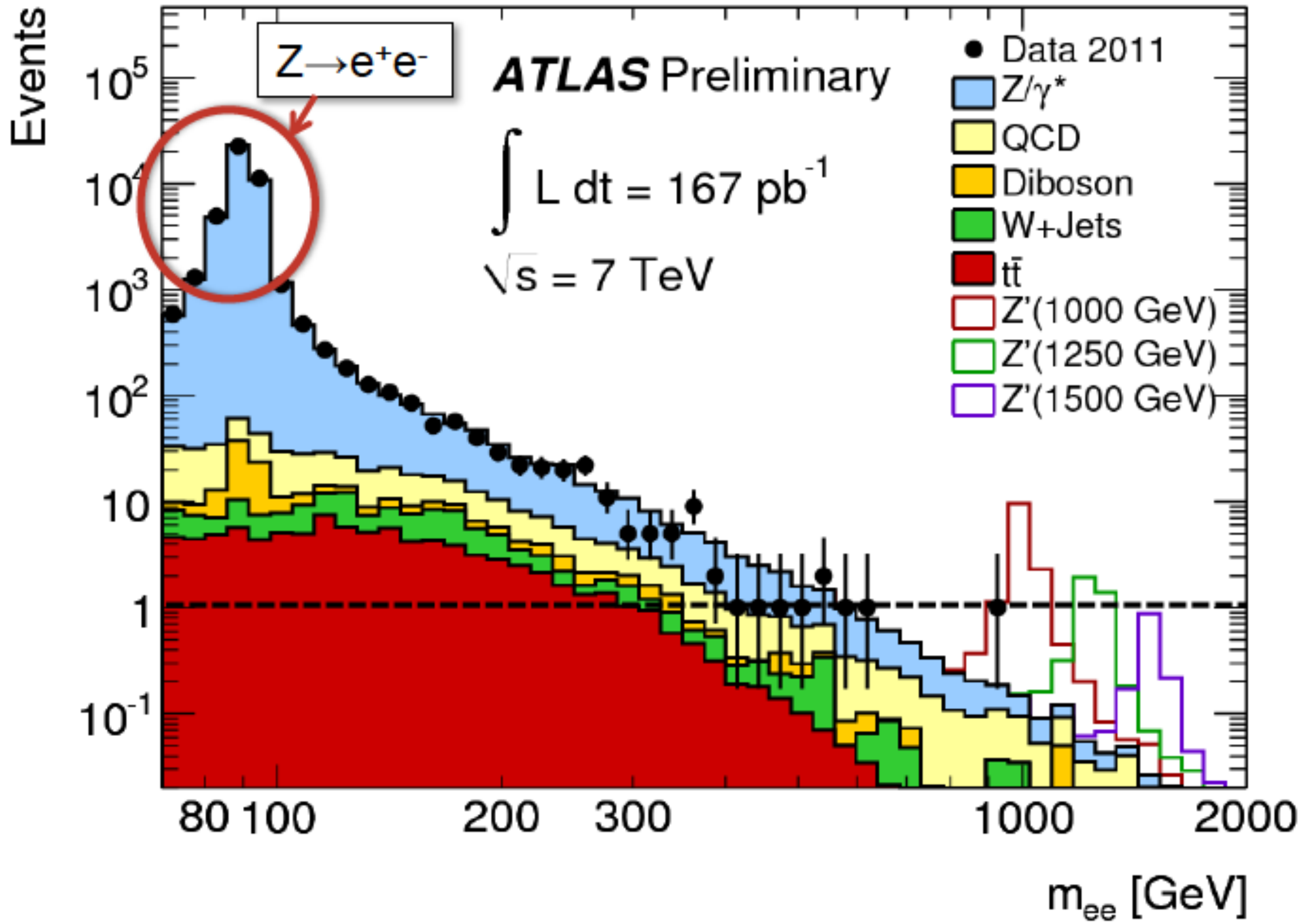
Elements of a search

- Like Z->ee but at higher mass.



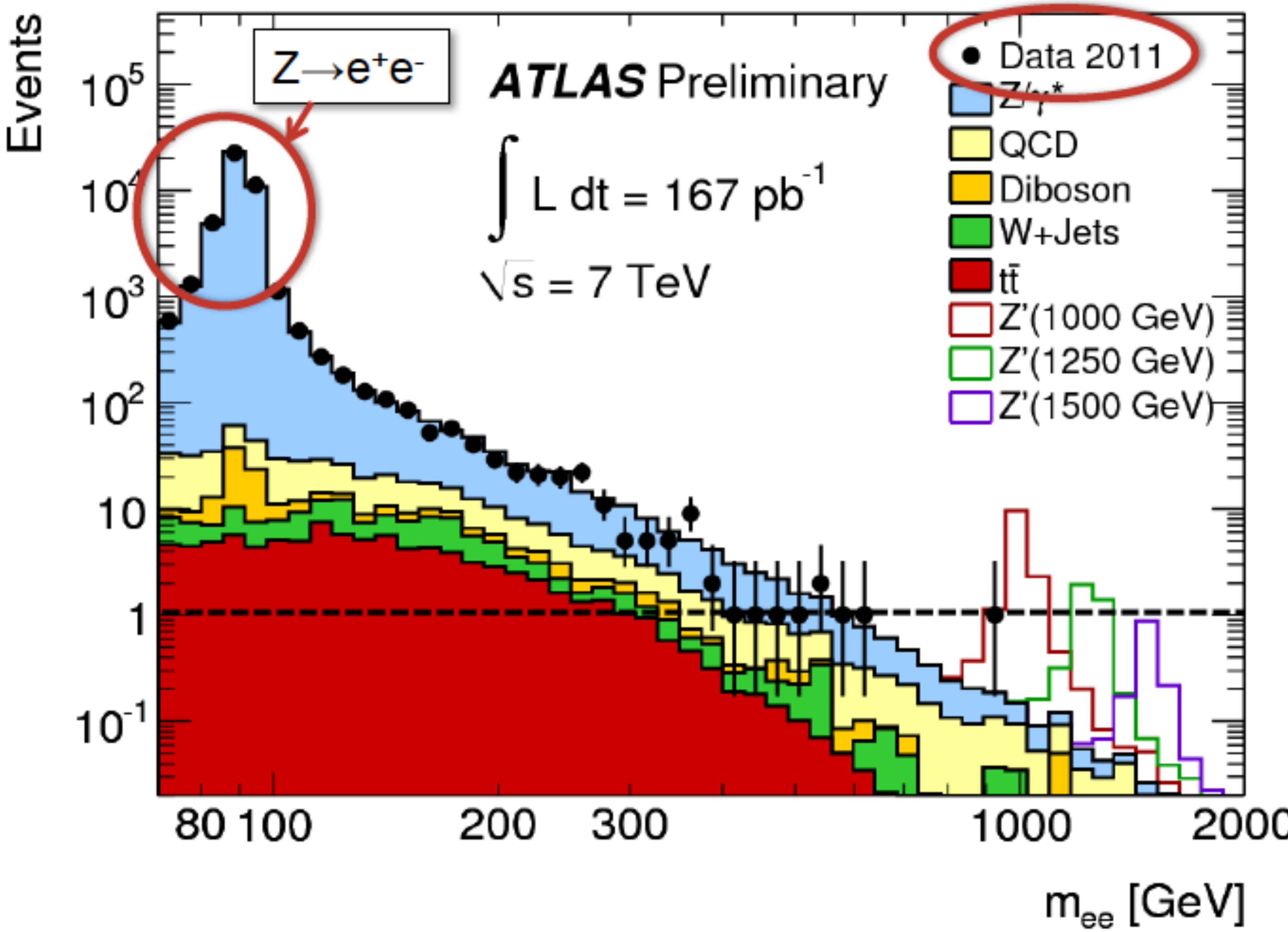
Elements of a search

- Like Z->ee but at higher mass.



Elements of a search

- Like Z->ee but at higher mass.

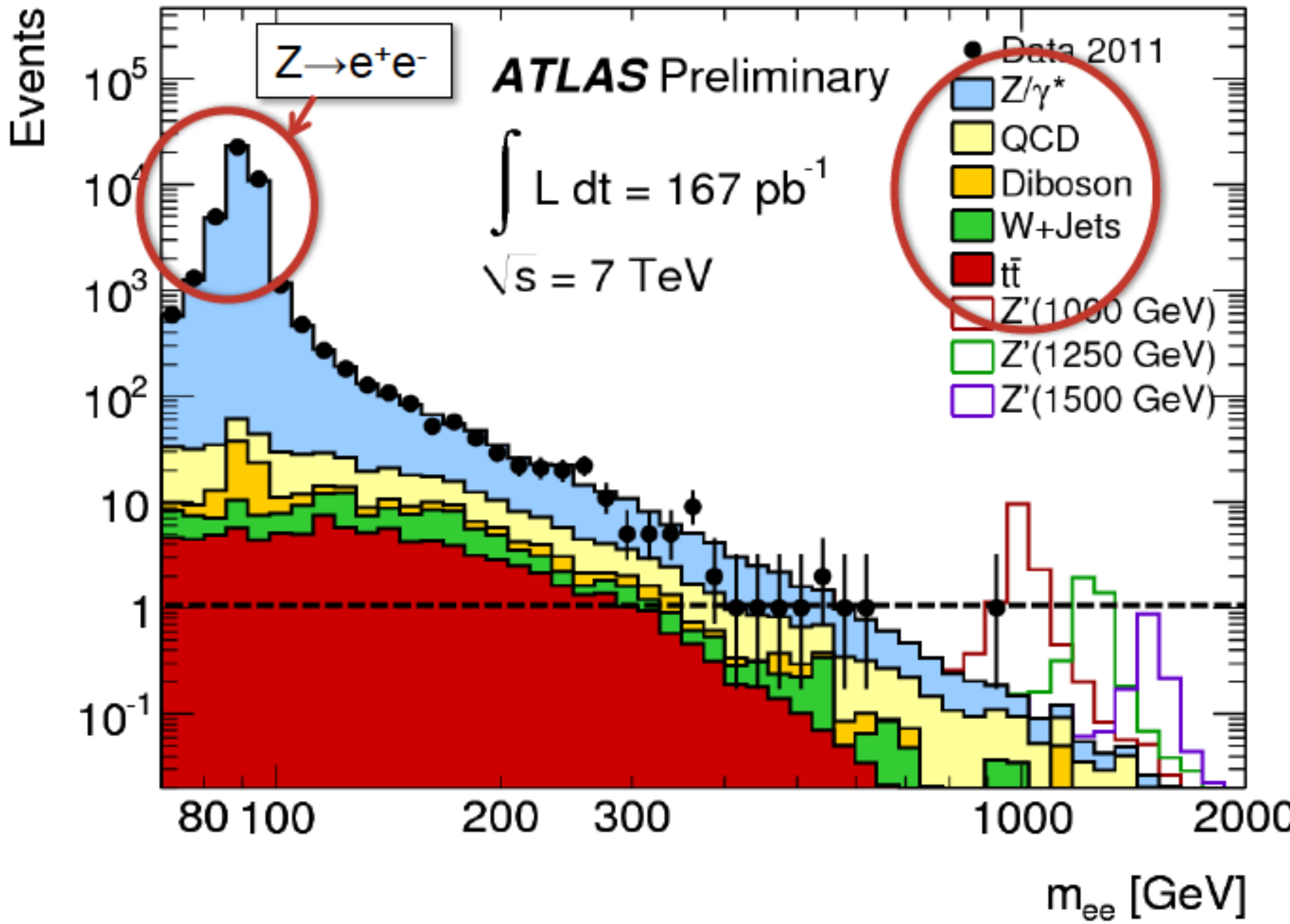


Select 2 electron candidates and plot their invariant mass for:

1. **Data**

Elements of a search

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

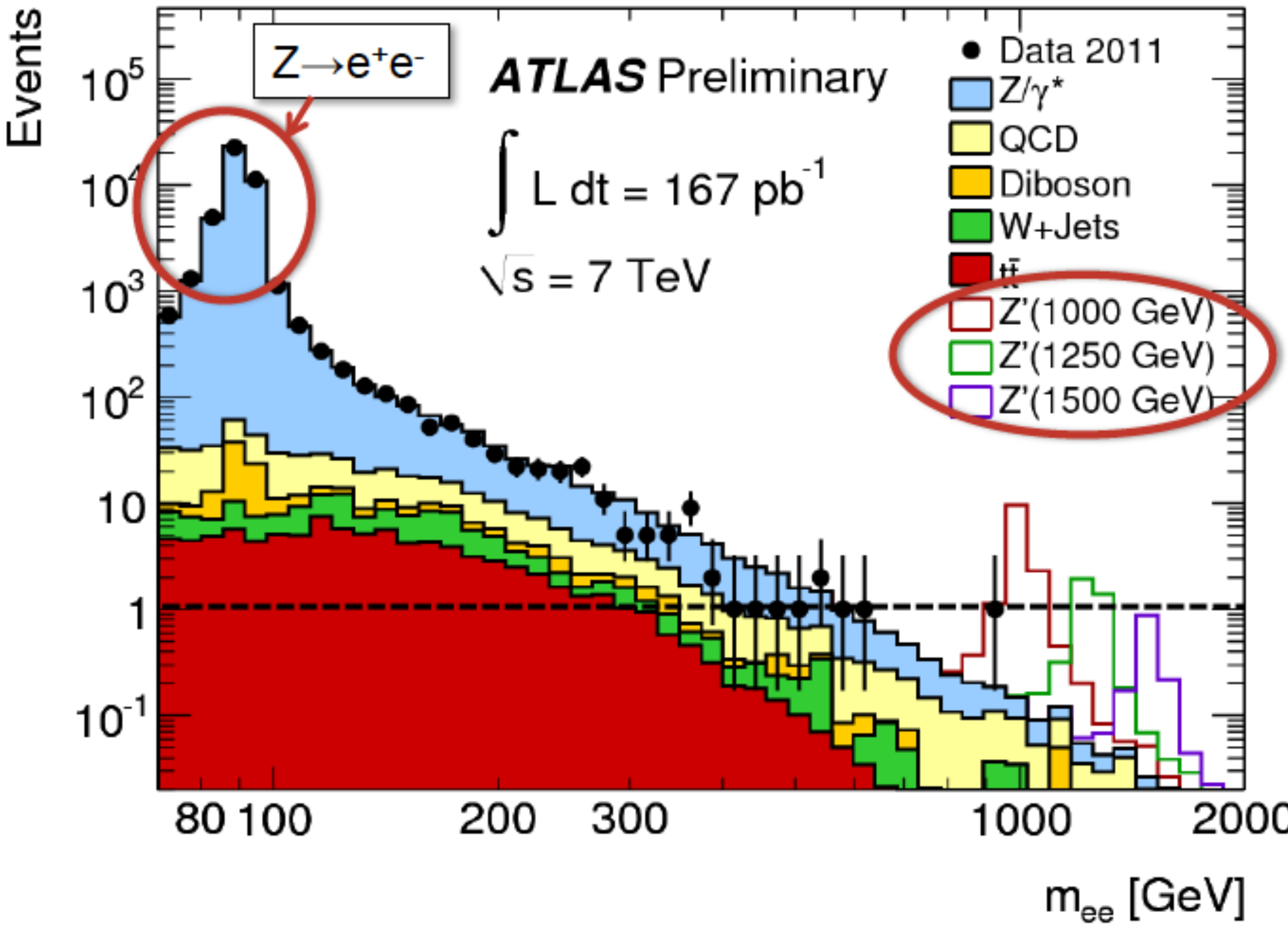


Select 2 electron candidates and plot their invariant mass for:

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

Elements of a search

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

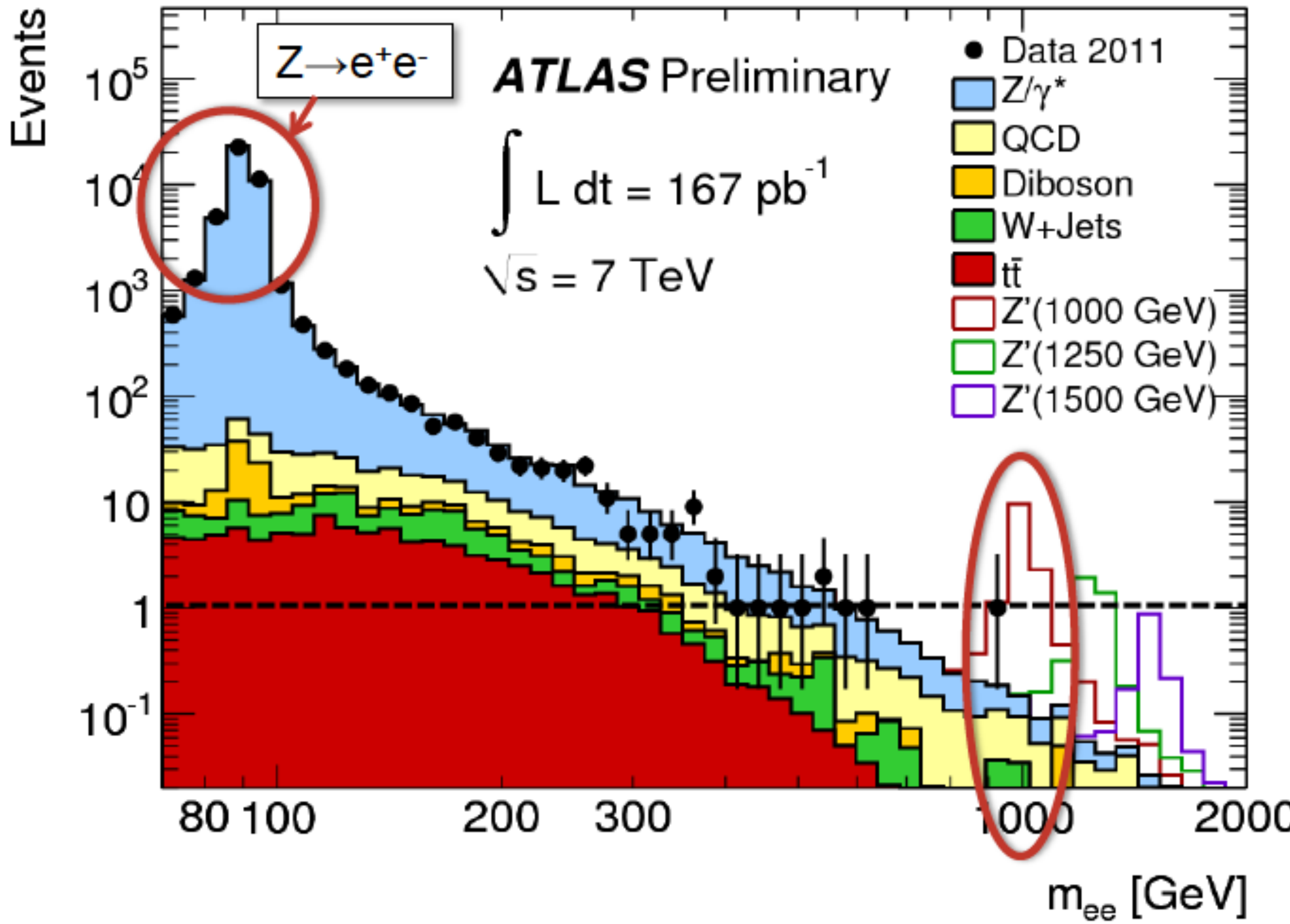


Select 2 electron candidates and plot their invariant mass for:

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

Elements of a search

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



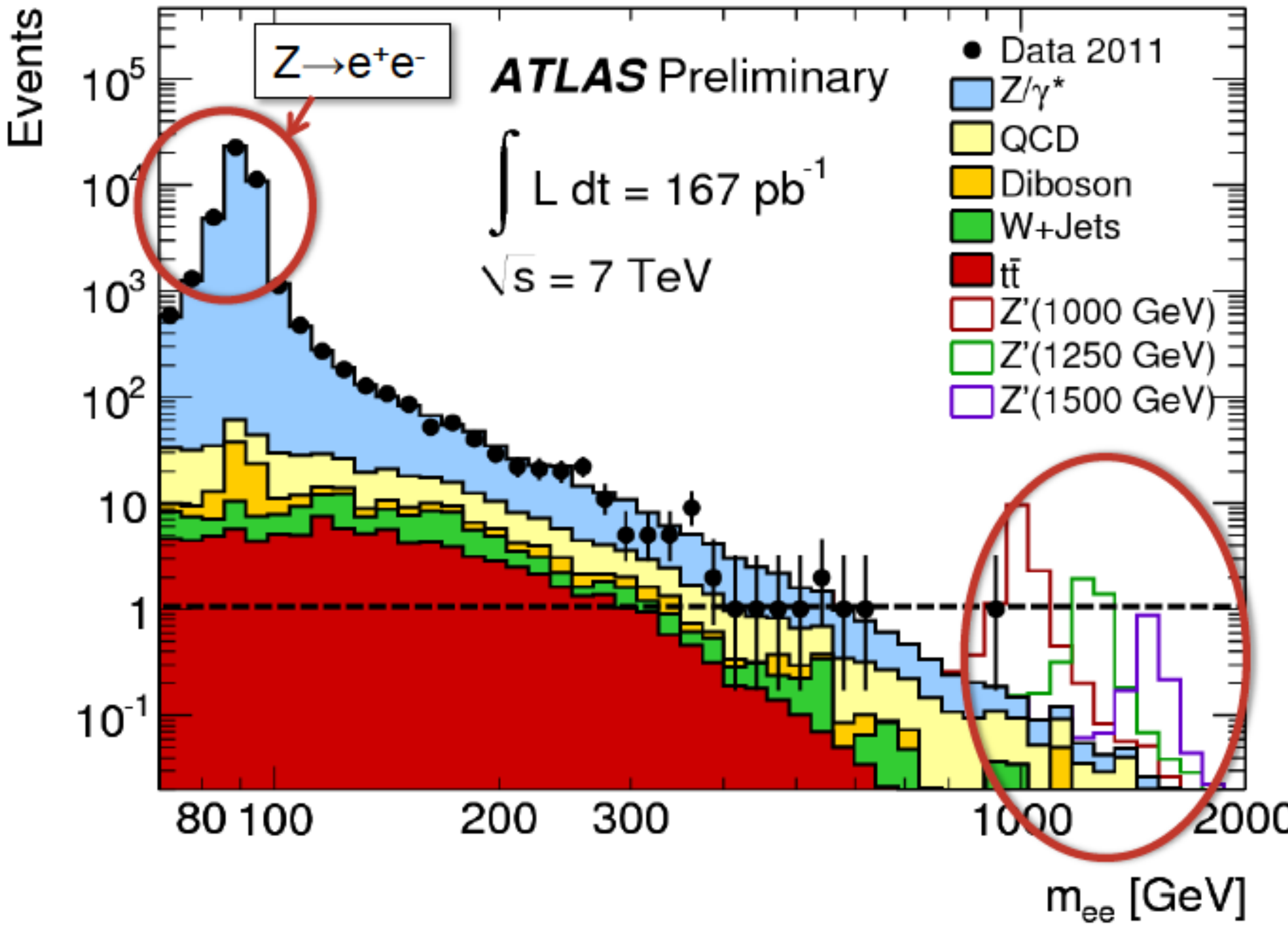
Select 2 electron candidates and plot their invariant mass for:

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

Data inconsistent with a 1TeV Z'

Elements of a search

◎ **Like Z->ee but at higher mass.**



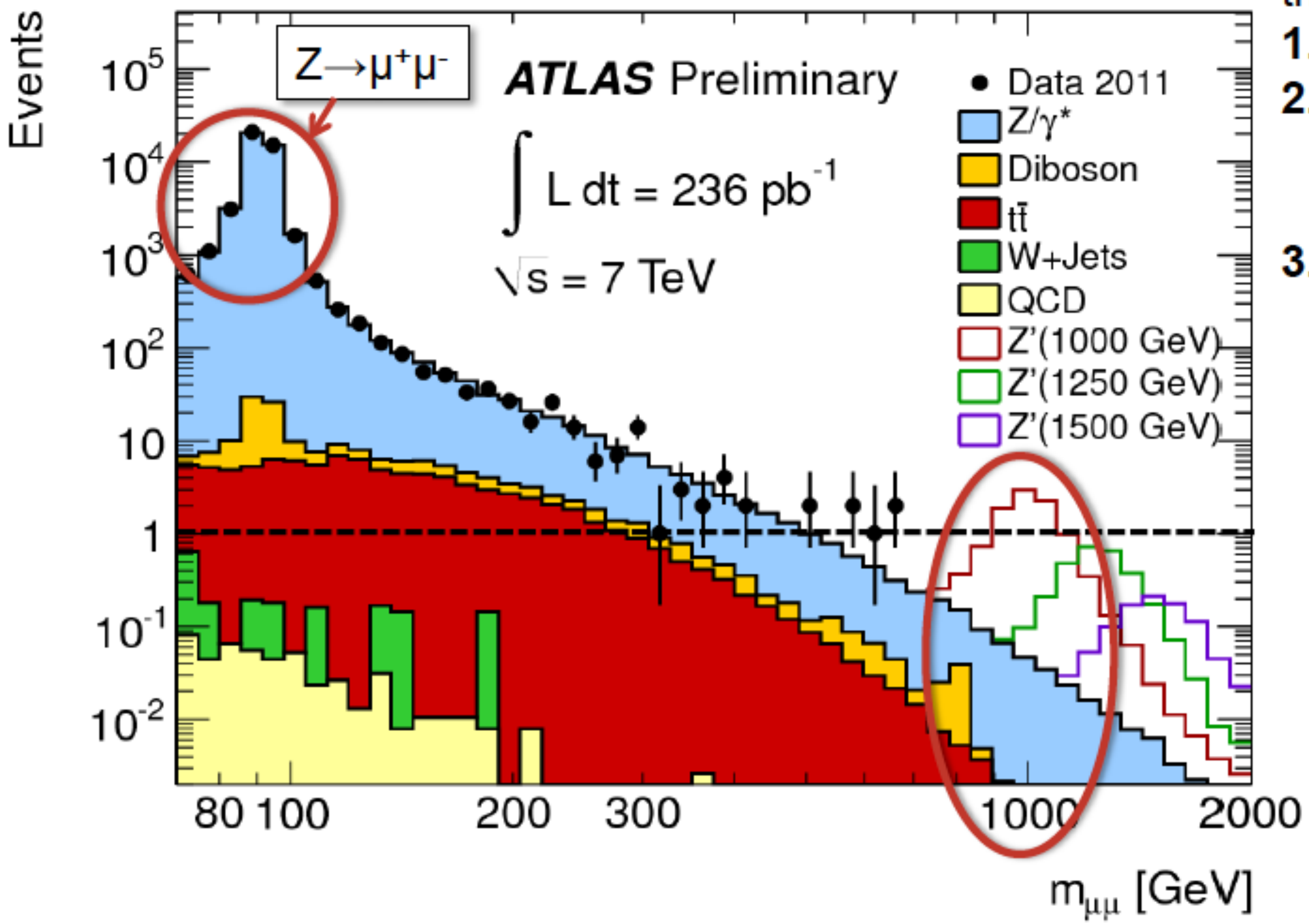
Select 2 electron candidates and plot their invariant mass for:

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

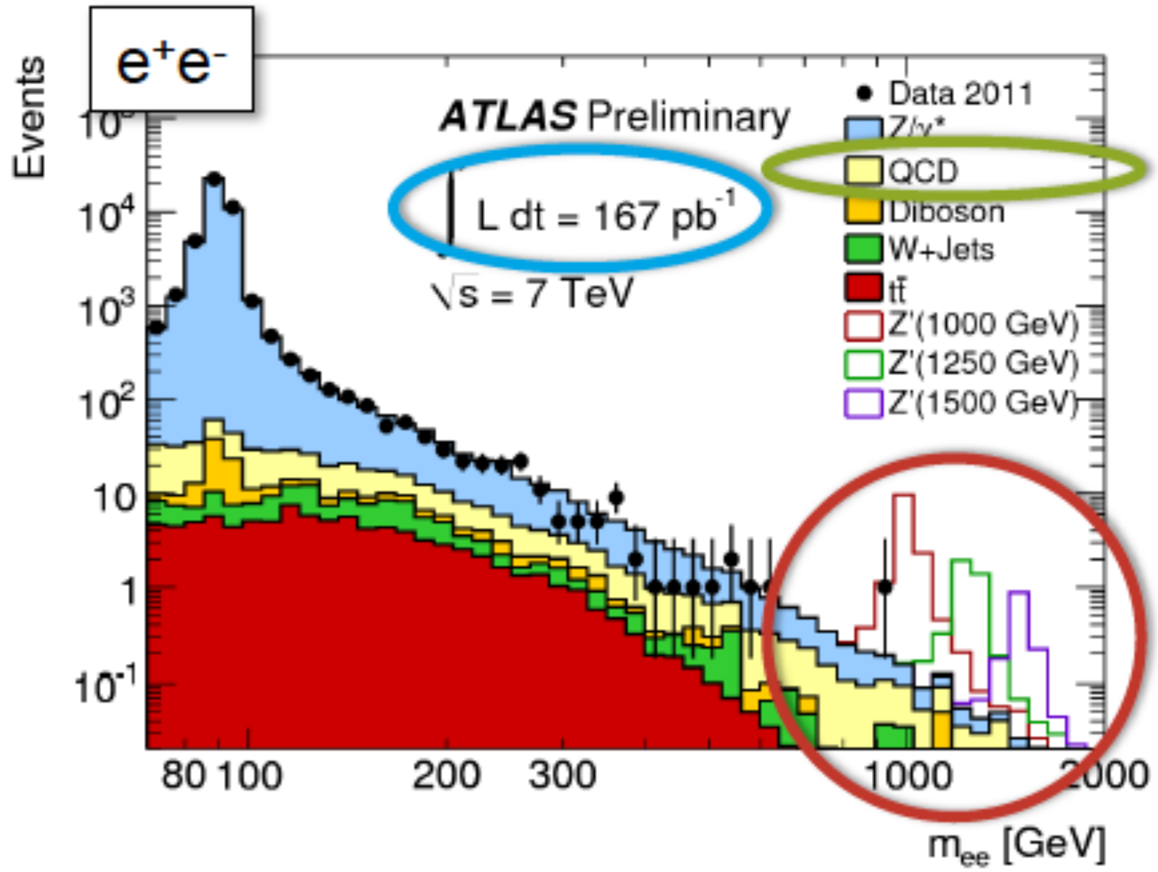


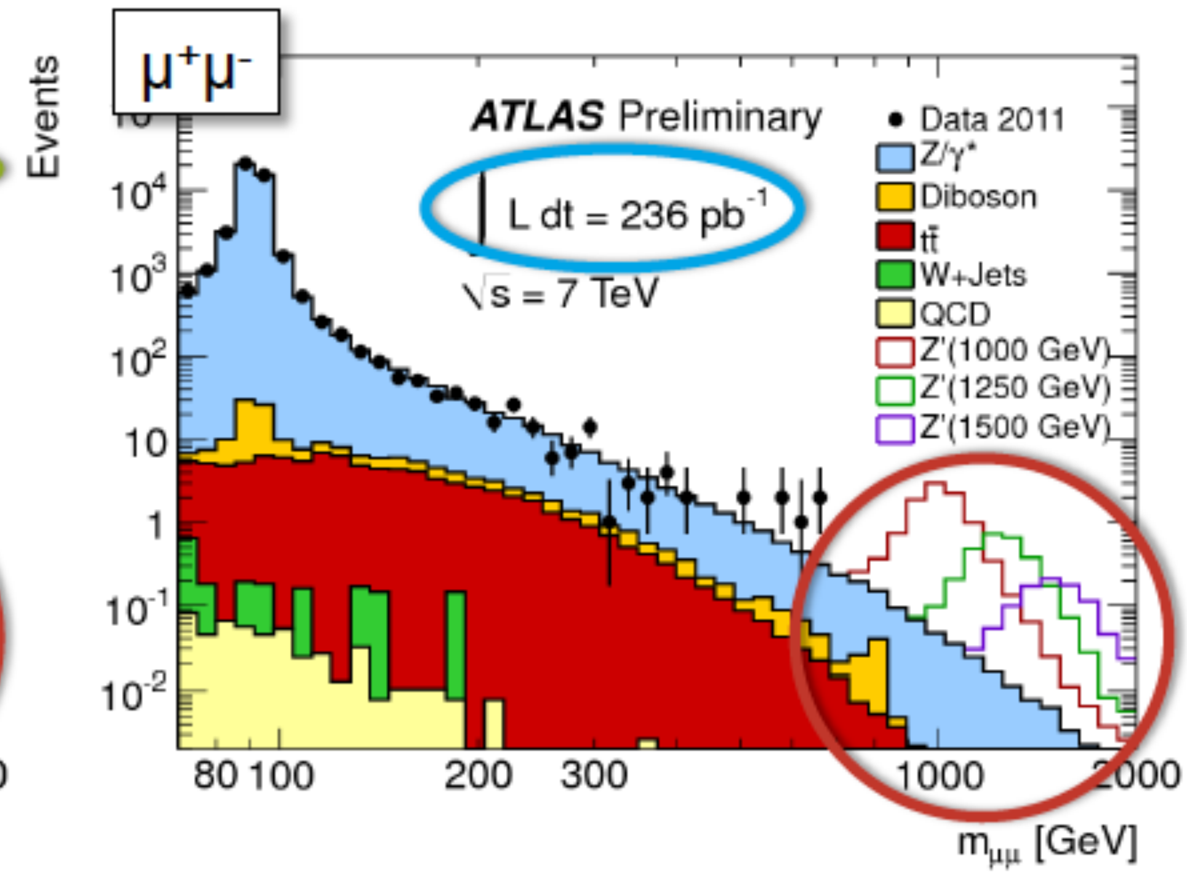
Select 2 muon candidates and plot their invariant mass for:

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

Data inconsistent with a 1TeV Z'

Elements of a search



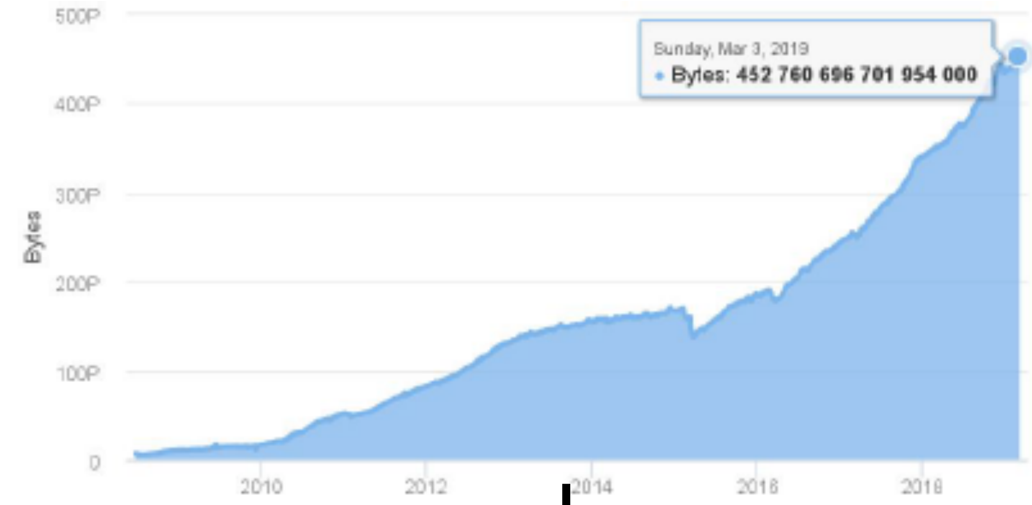


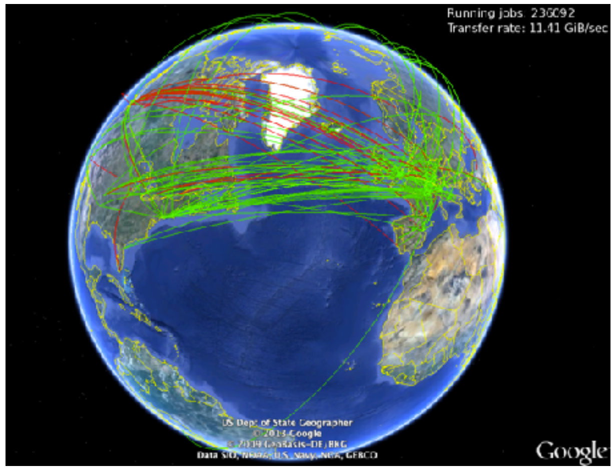
Why is the resolution worse in the muon channel?

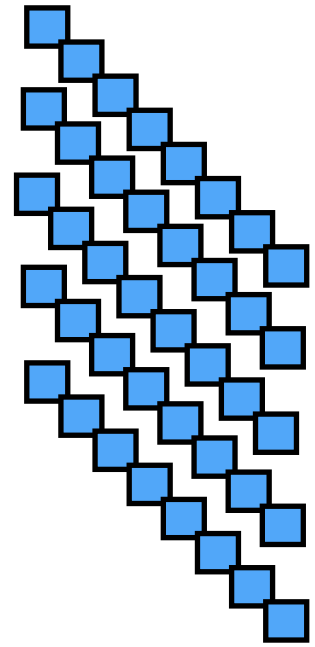
Differences in:

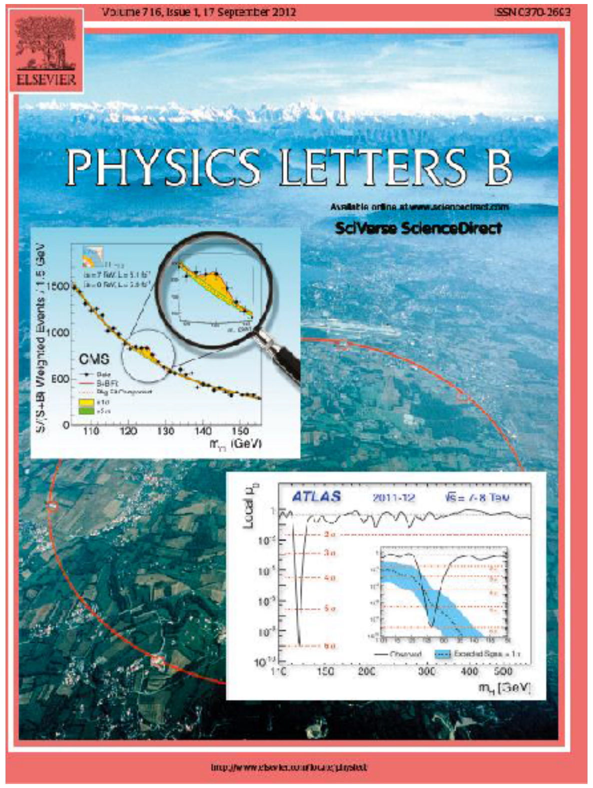
- **Resolution**
- **Background composition**
- **Dataset**

Data analysis









Needles in haystacks

- We record billions of events

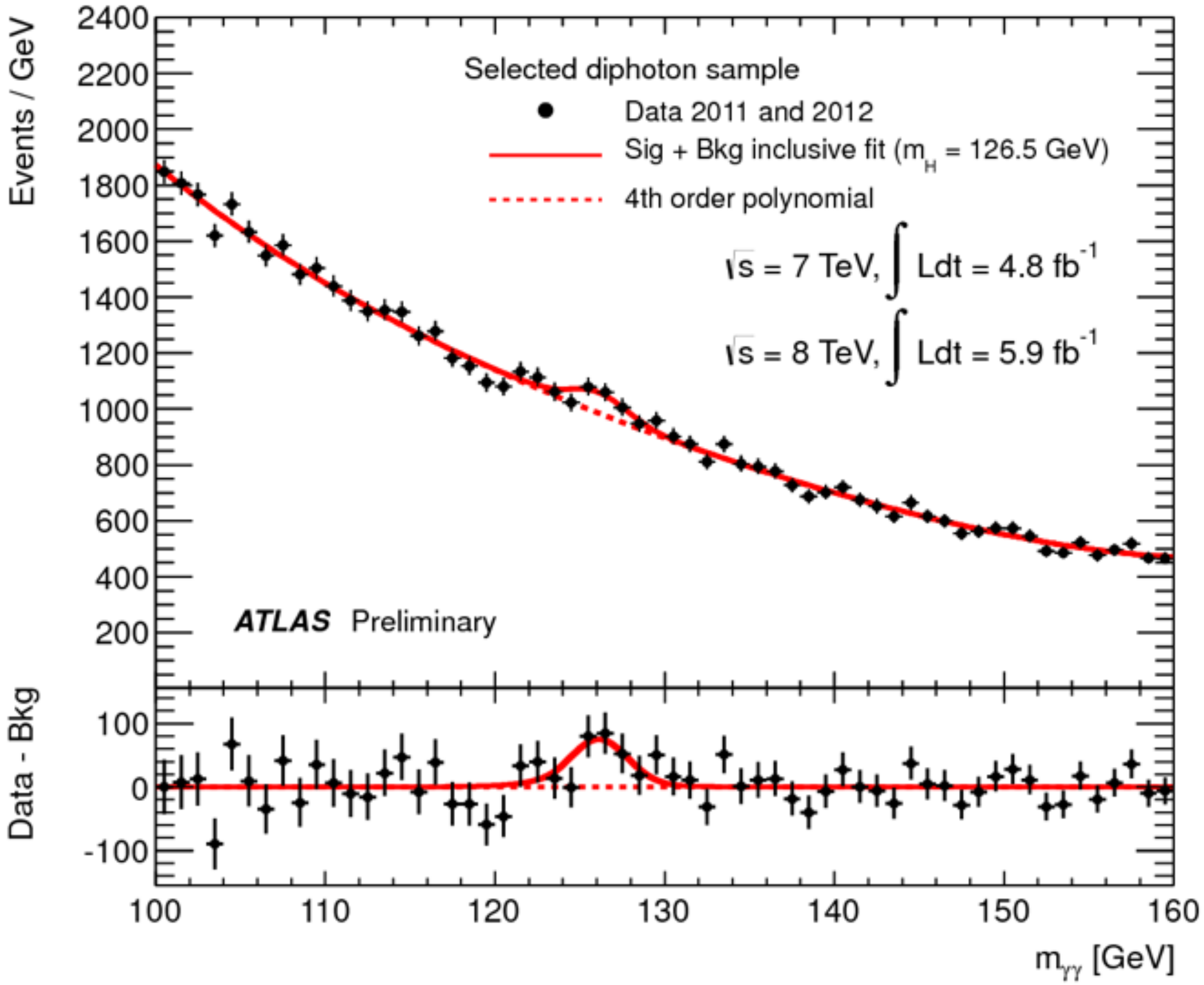
- The data are structured but each event is different - ***unique data science challenge***

- Data reduction proceeds via a two-pronged approach:

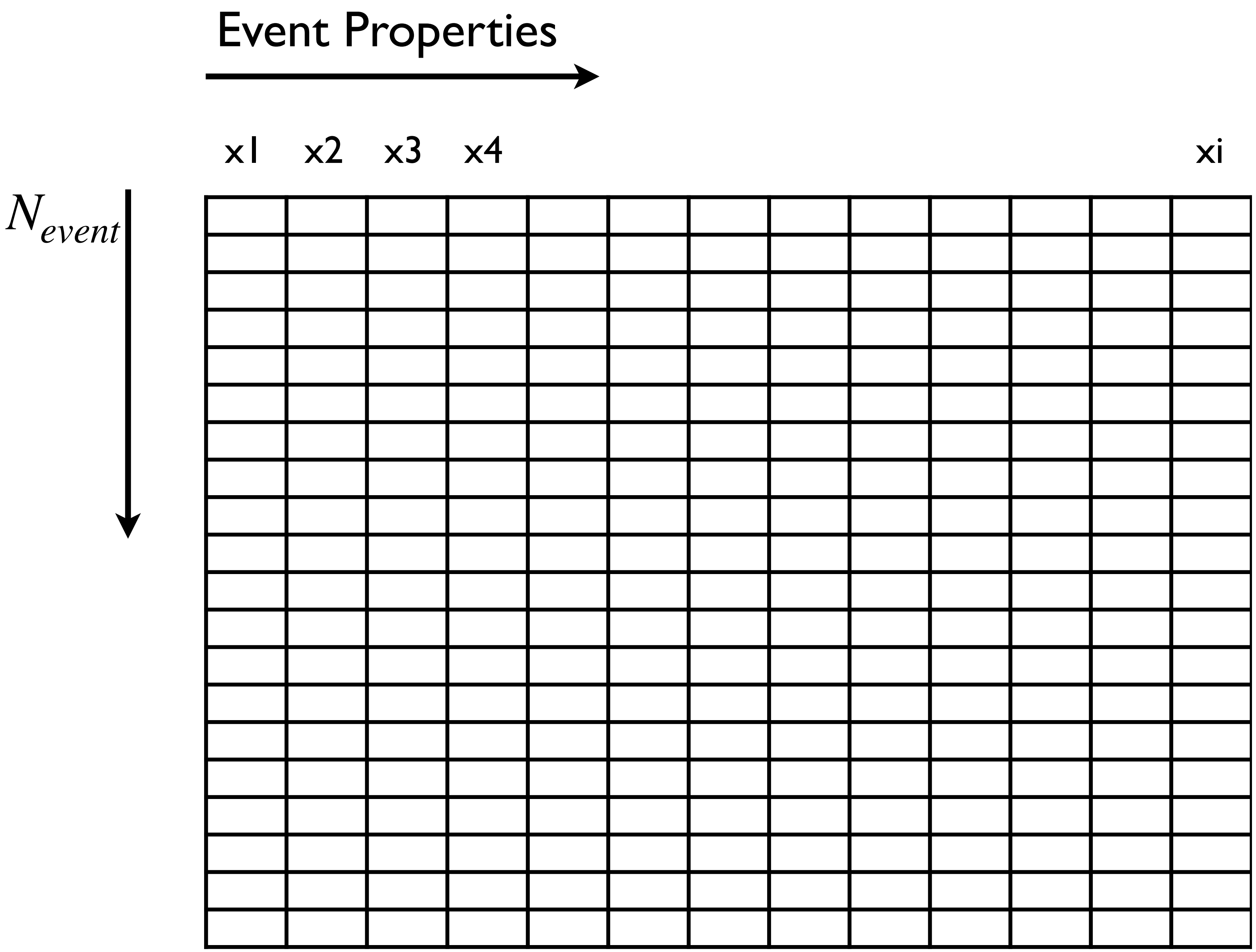
- Select only the events that you are interested in
    - ***e.g. events with two photons***

- Keep only the information you need
    - Throw away the rest !

- Final statistical inference is only performed on the reduced data



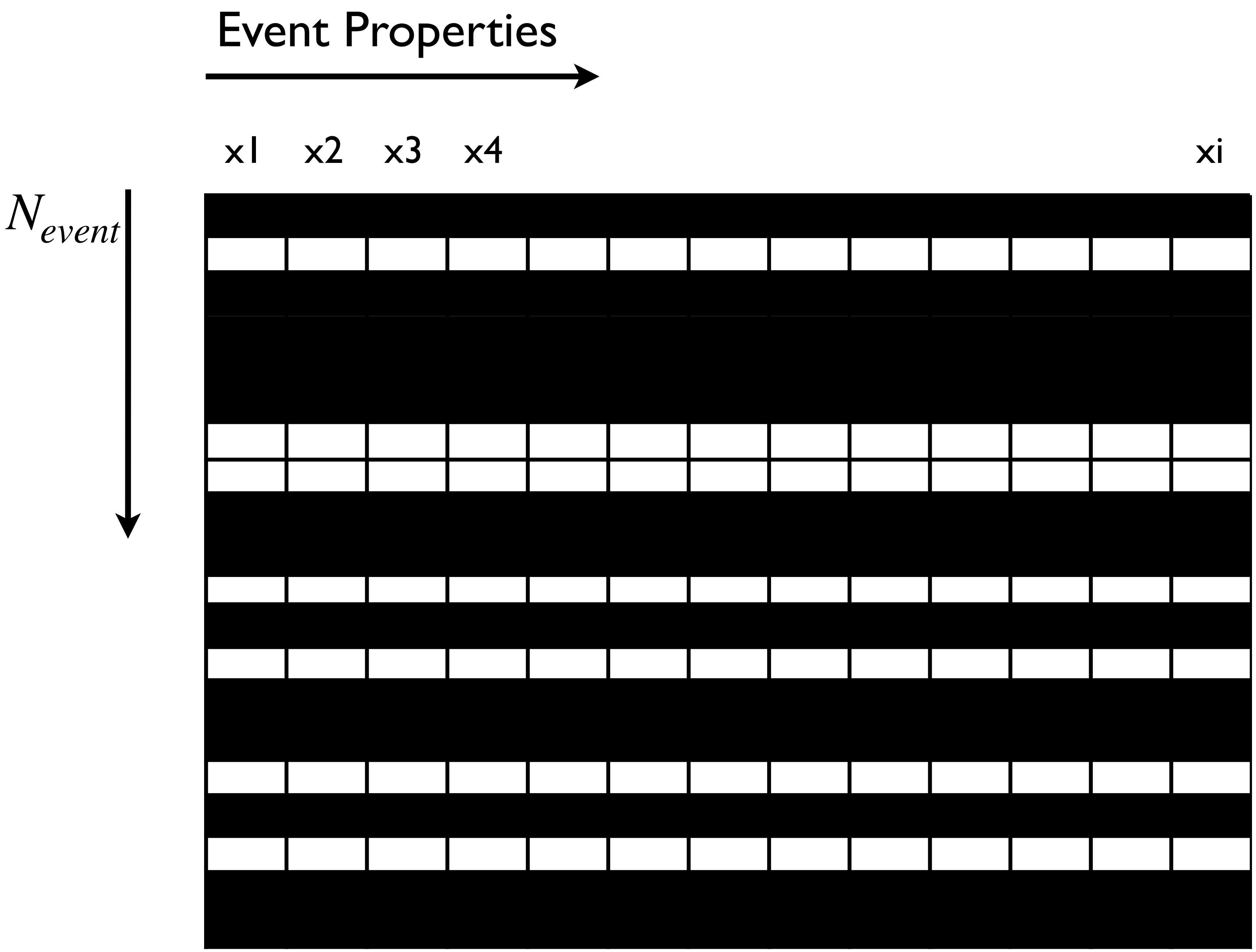
Slice and dice - data reduction



Simplified picture, in reality the event properties depend on the event content found by event reconstruction

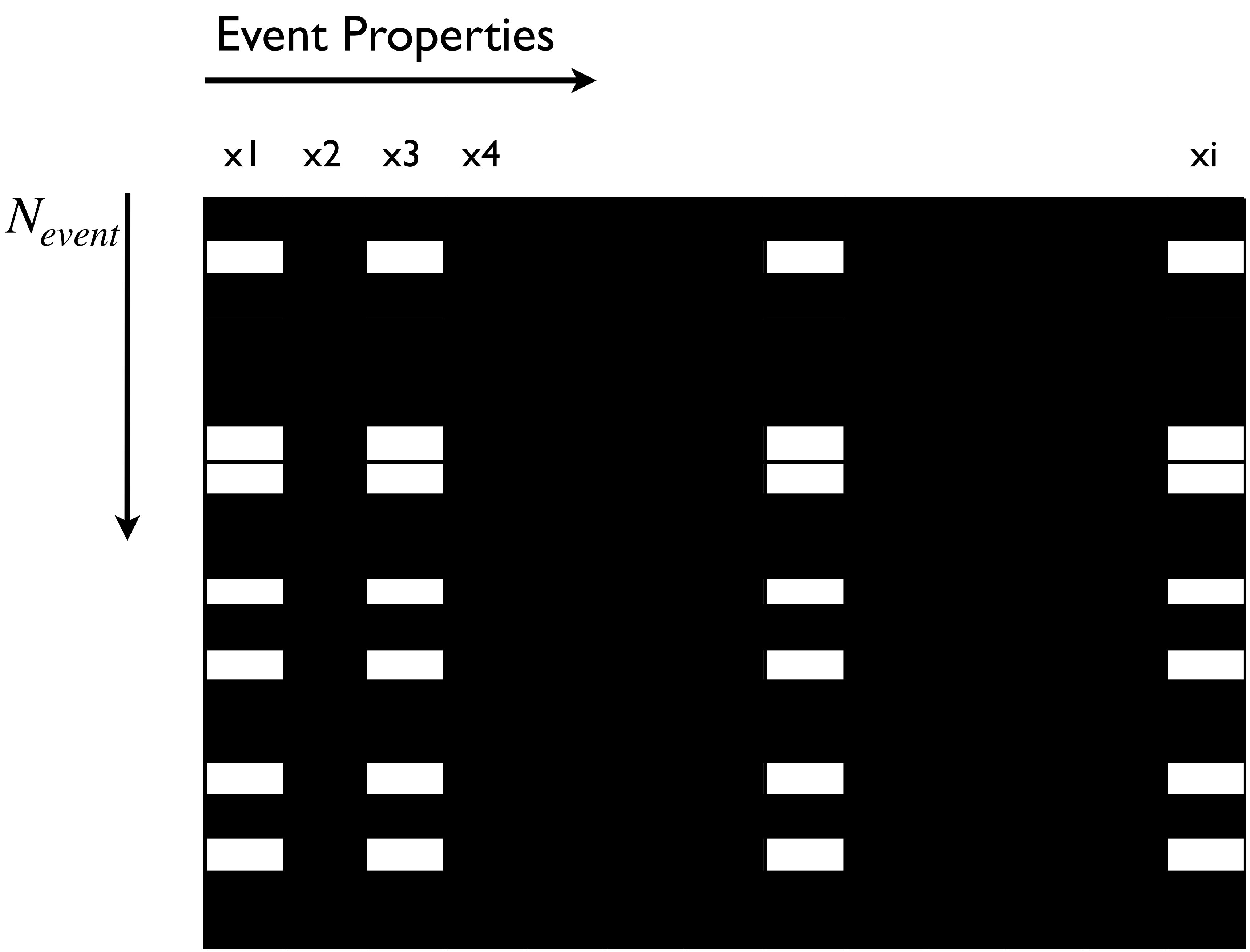
- There are two dimensions to our data challenge, one is the (billions) of individual events, the other is the properties of each event

Slice and dice - data reduction



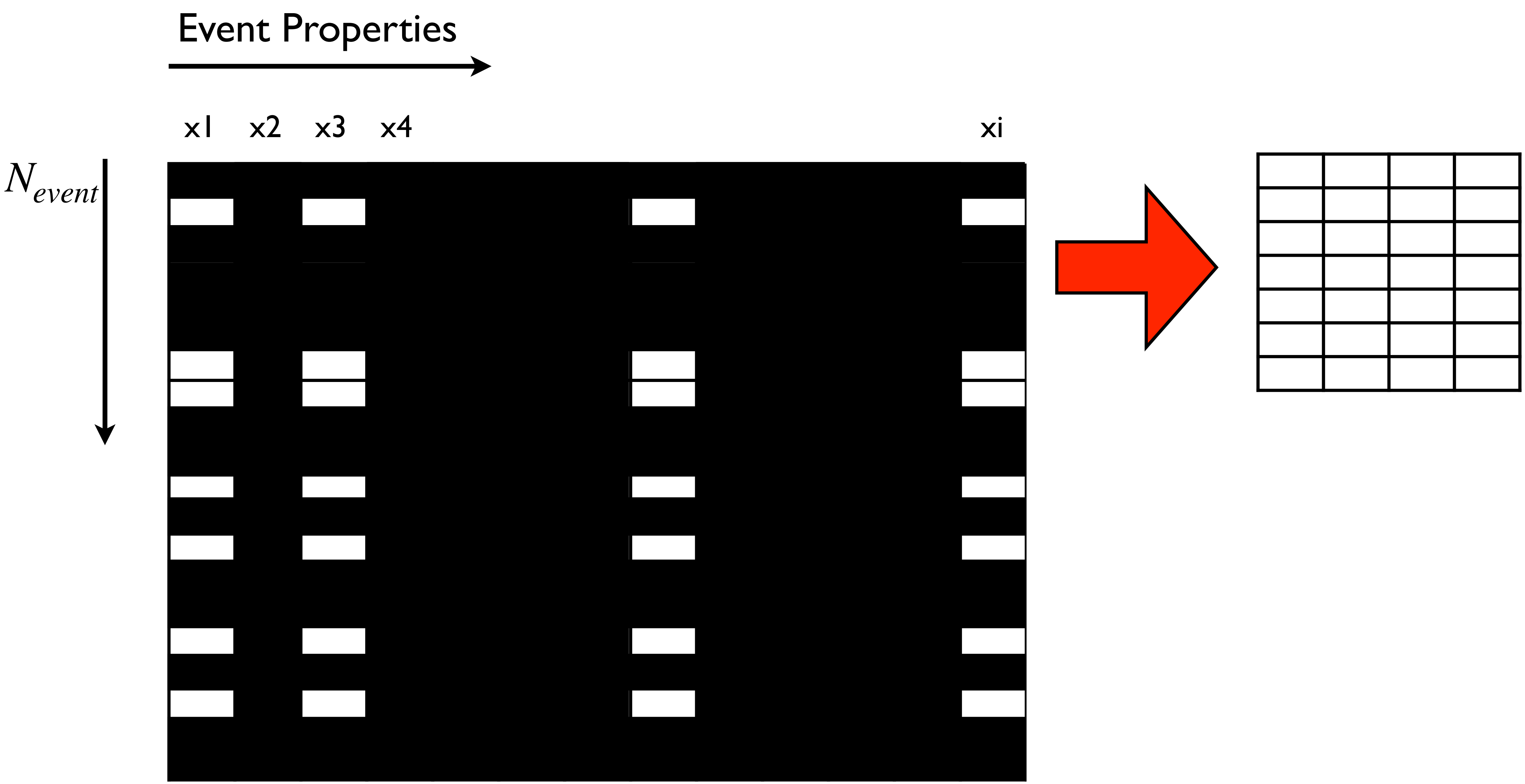
- We can reduce data by selecting only our interesting events

Slice and dice - data reduction



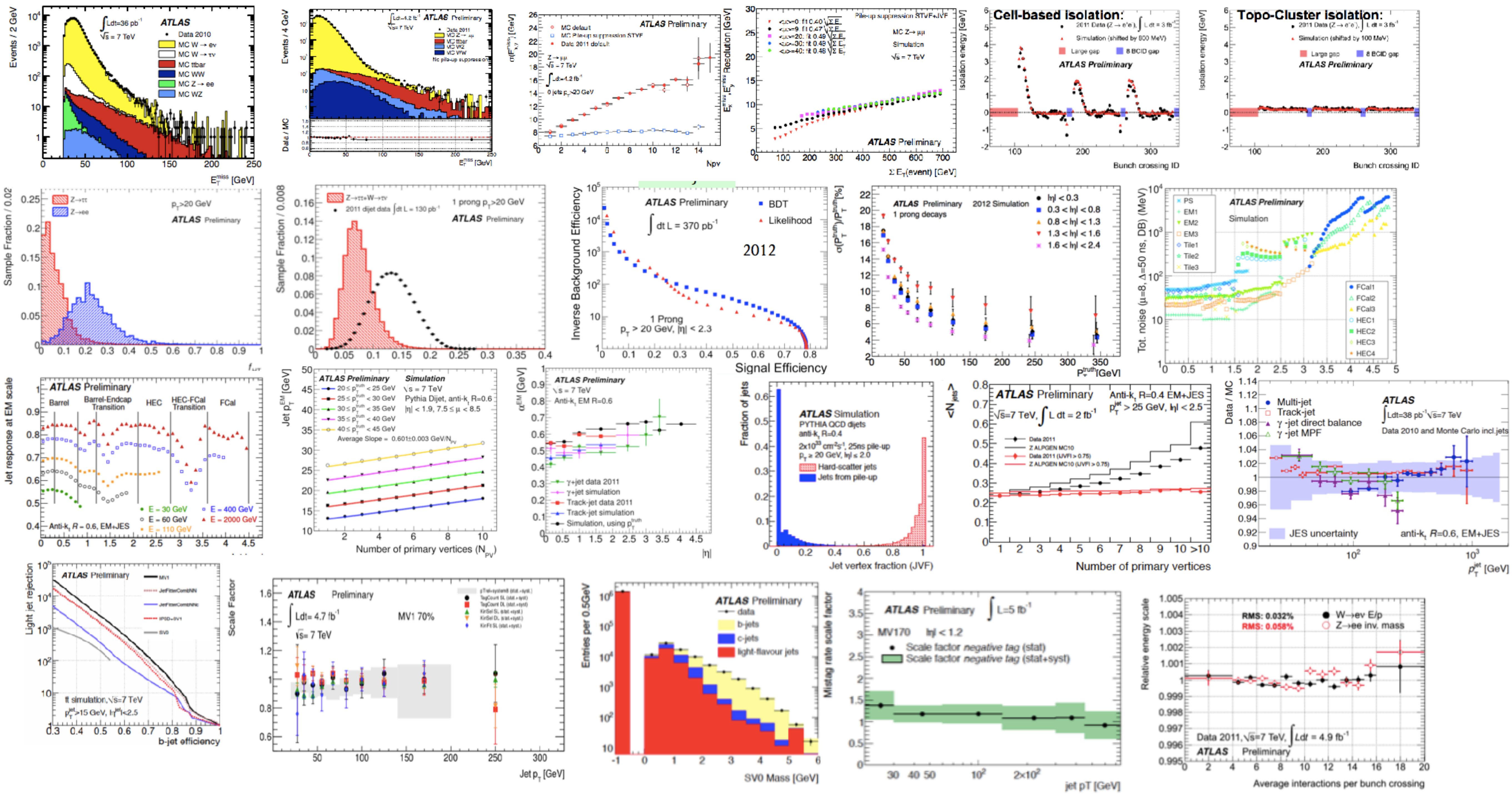
- And we can reduce data by selecting only the properties needed for our analysis

Slice and dice - data reduction



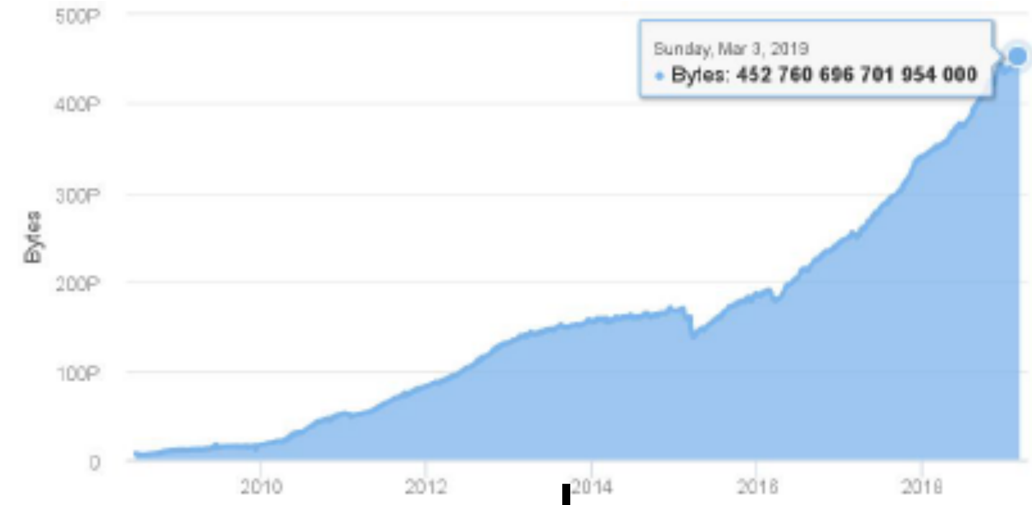
- Data reduction usually aims for factors of 100 or more (more than shown here !)

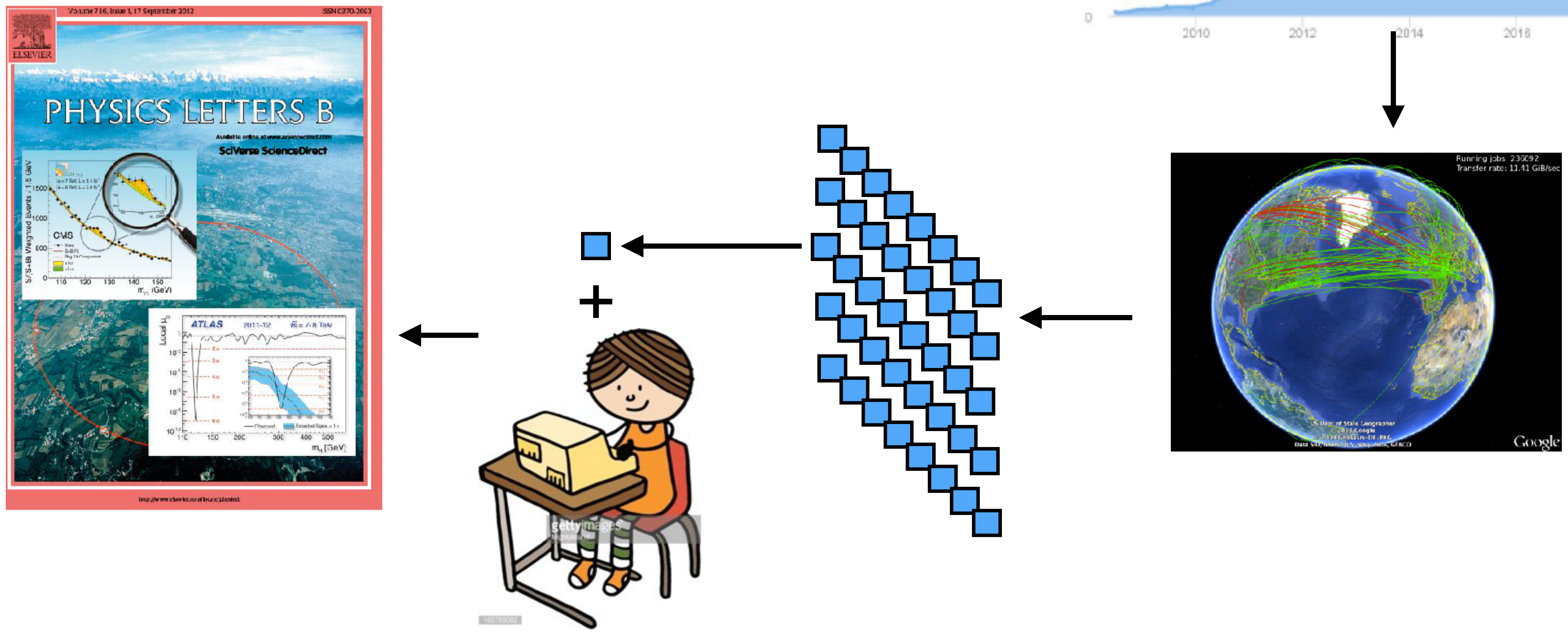
Ingredients to the ATLAS physics program



- We make lots of reduced samples of both data and simulation, which all need to be replicated around the world - a computing challenge !

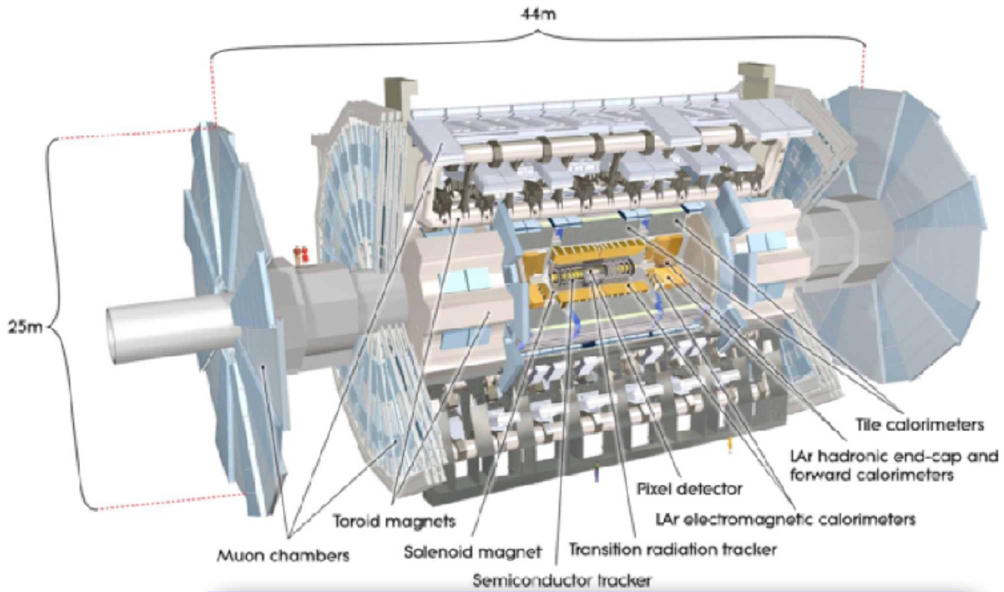
The best computing model

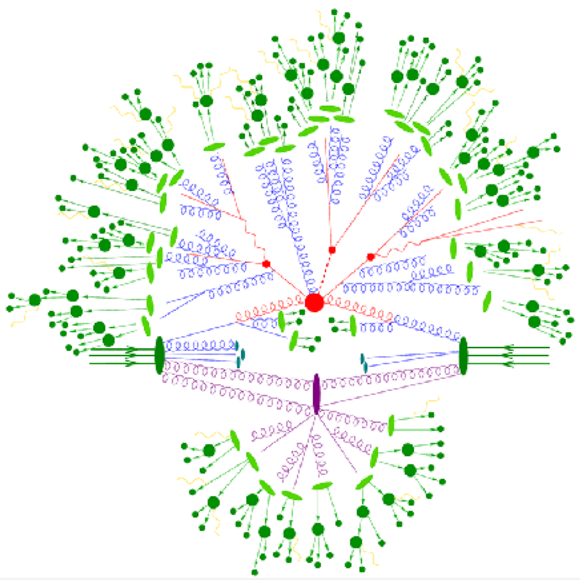




- How to most efficiently do this across the whole physics program making the best use of computing resources and the best use of people’s time is an important question

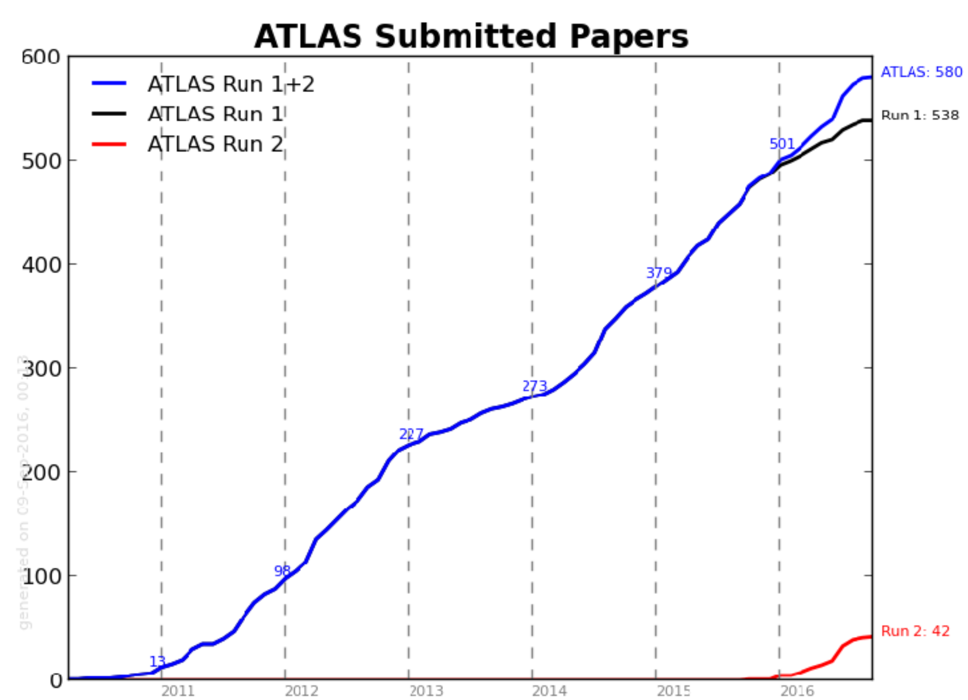
Now you know how to do exabyte-scale physics analysis!





Exabytes of Data

Exabytes of Simulation



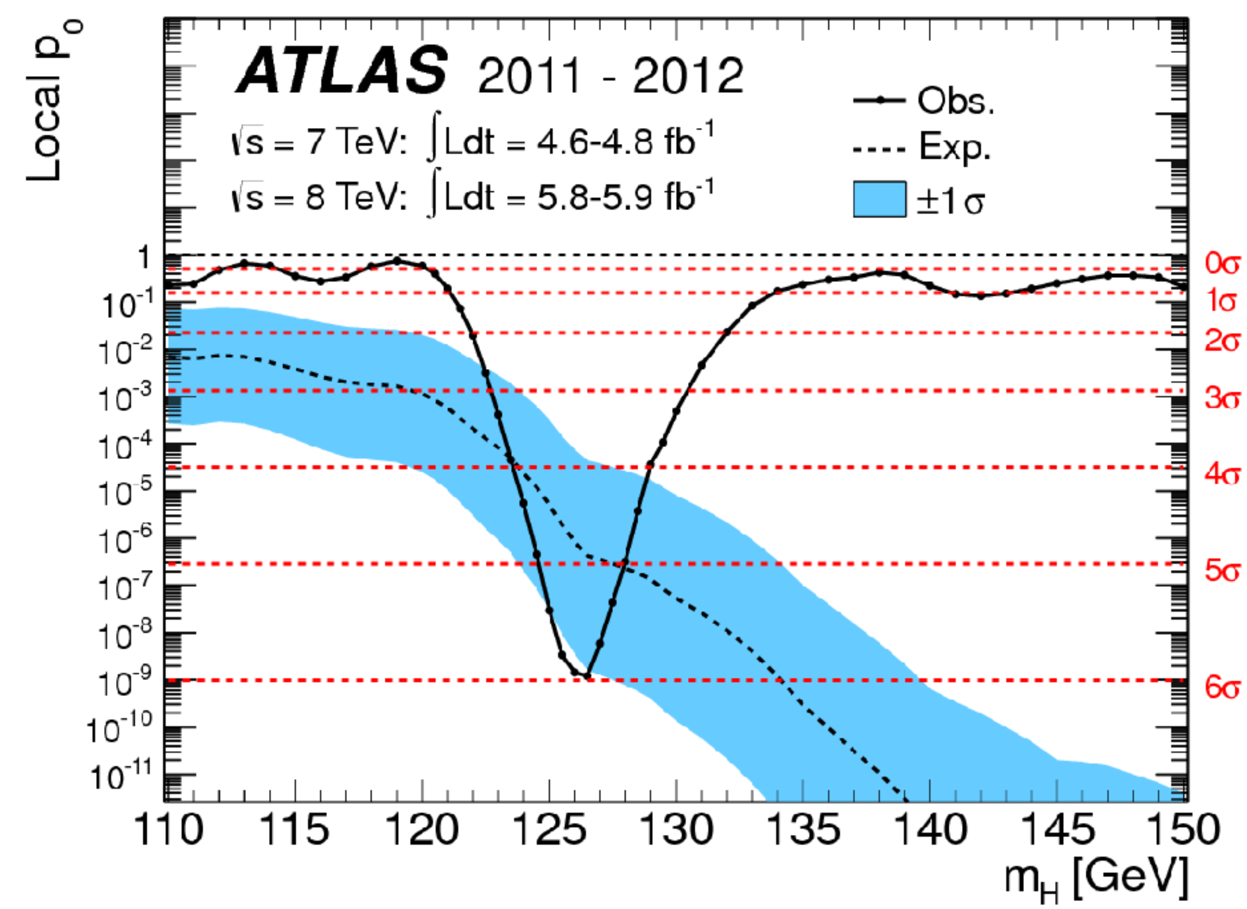
Publish!

Now it’s over to you !

Your

Photo

Here



- Our future computing needs outstrip our computing resources
  - and computing gets more heterogeneous and complicated
  - and we want to be as environmentally-responsible as possible
- So you have work to do - good luck and have fun!

Contact details

- I am usually based at Geneva Observatory in Versoix, but will be here at CERN Wednesday 28th through Friday 30th June.
    - I will be available for Q&A every afternoon from 3pm-4pm in restaurant 1, feel free to send questions to my email

- email: paul.laycock@unige.ch