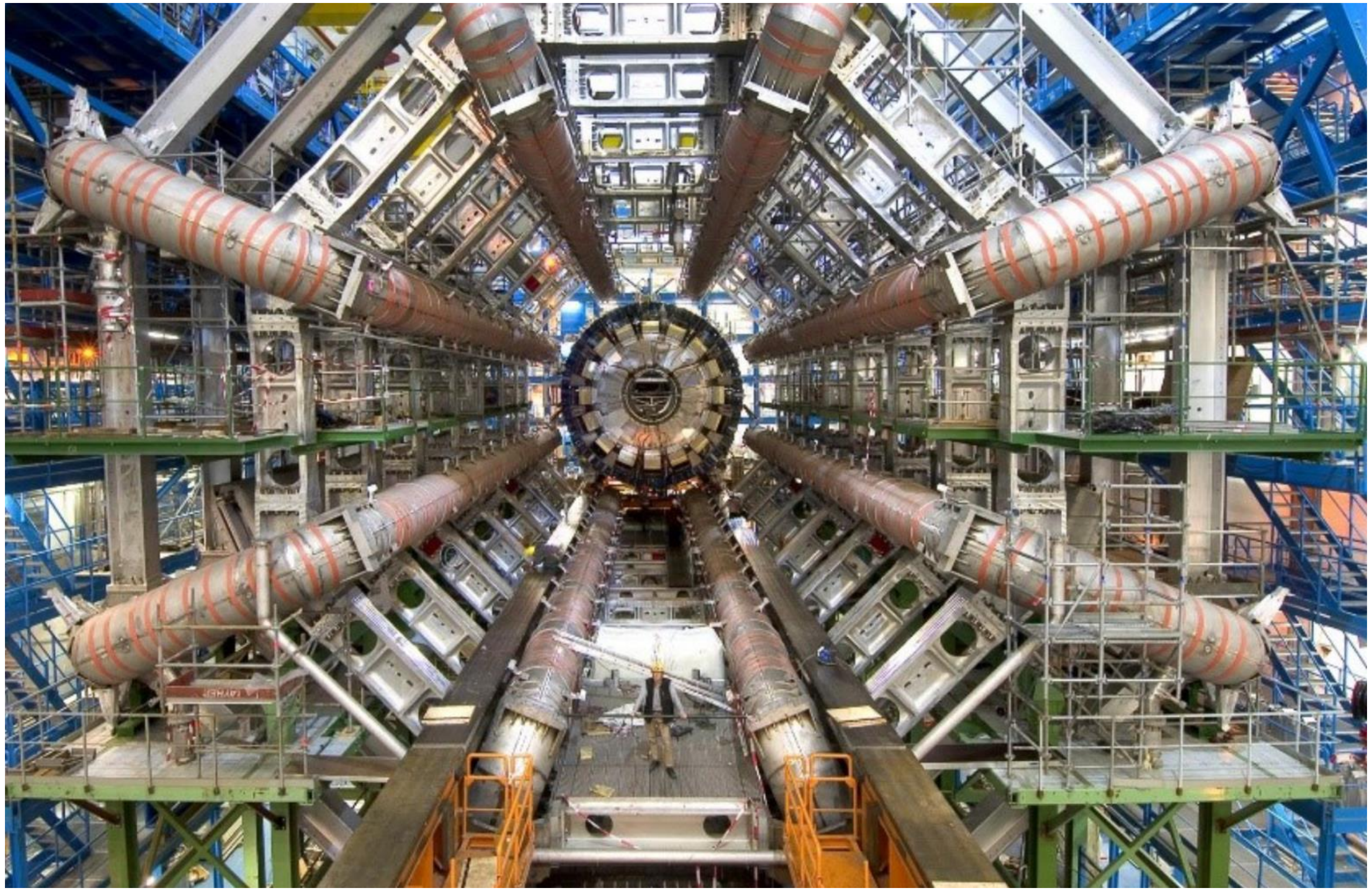
Data Analysis at the ATLAS experiment at the Large Hadron Collider



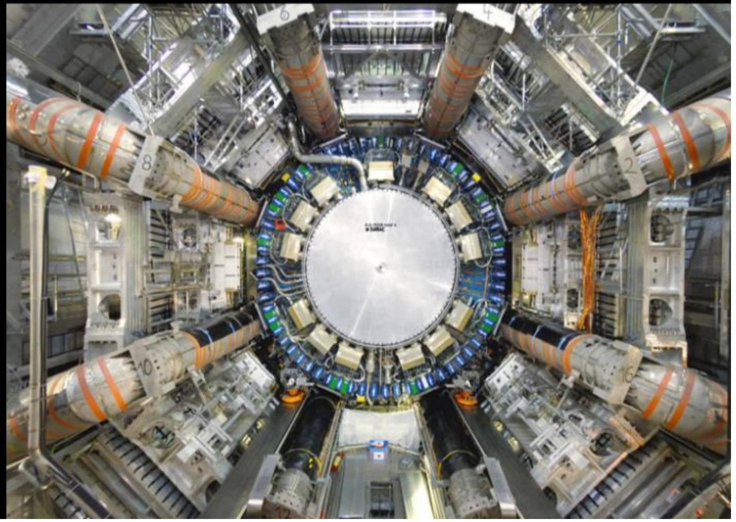
Dr Kate Shaw

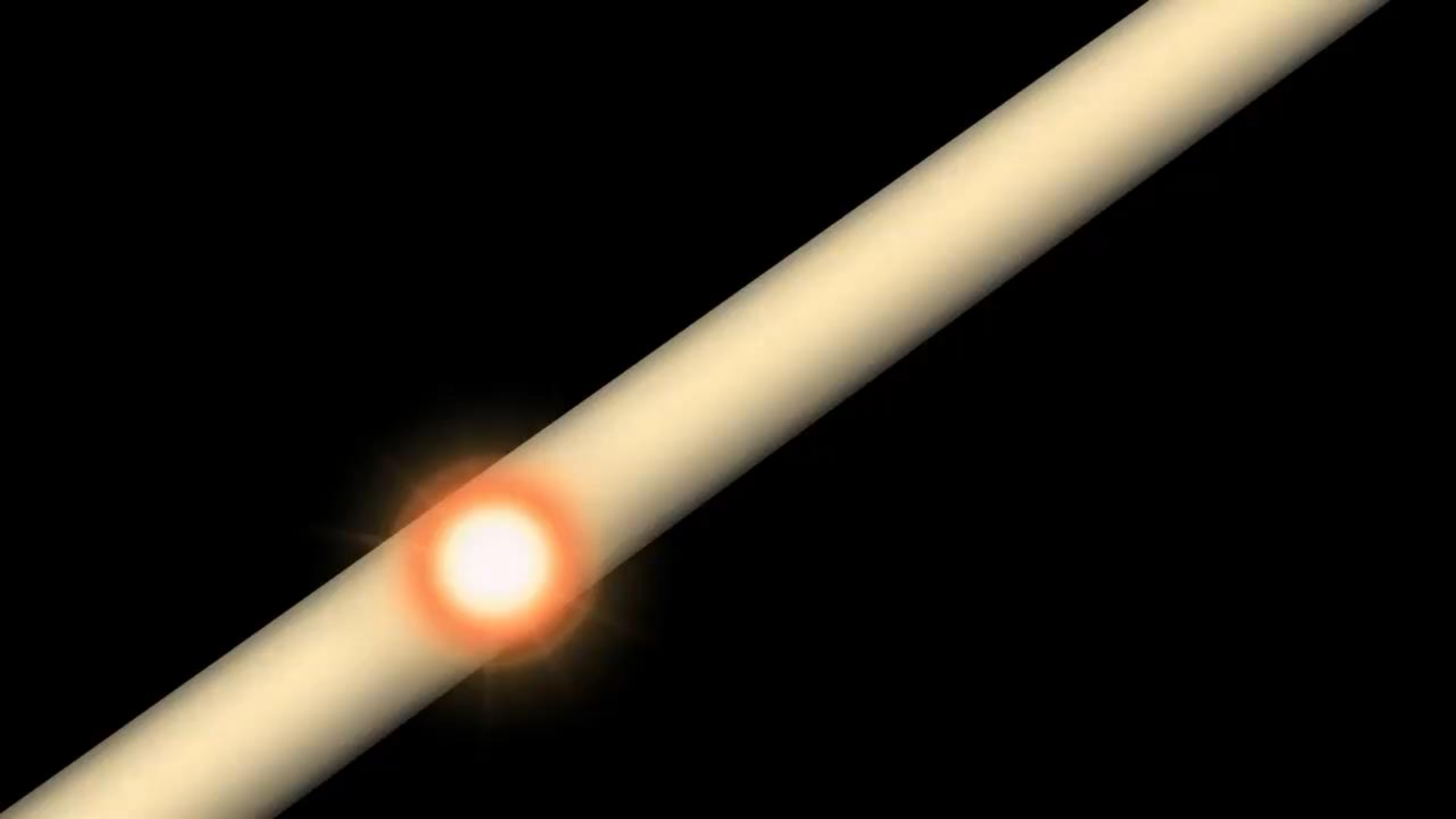
ICTP & University of Sussex

Physics Without Frontierts Thailand

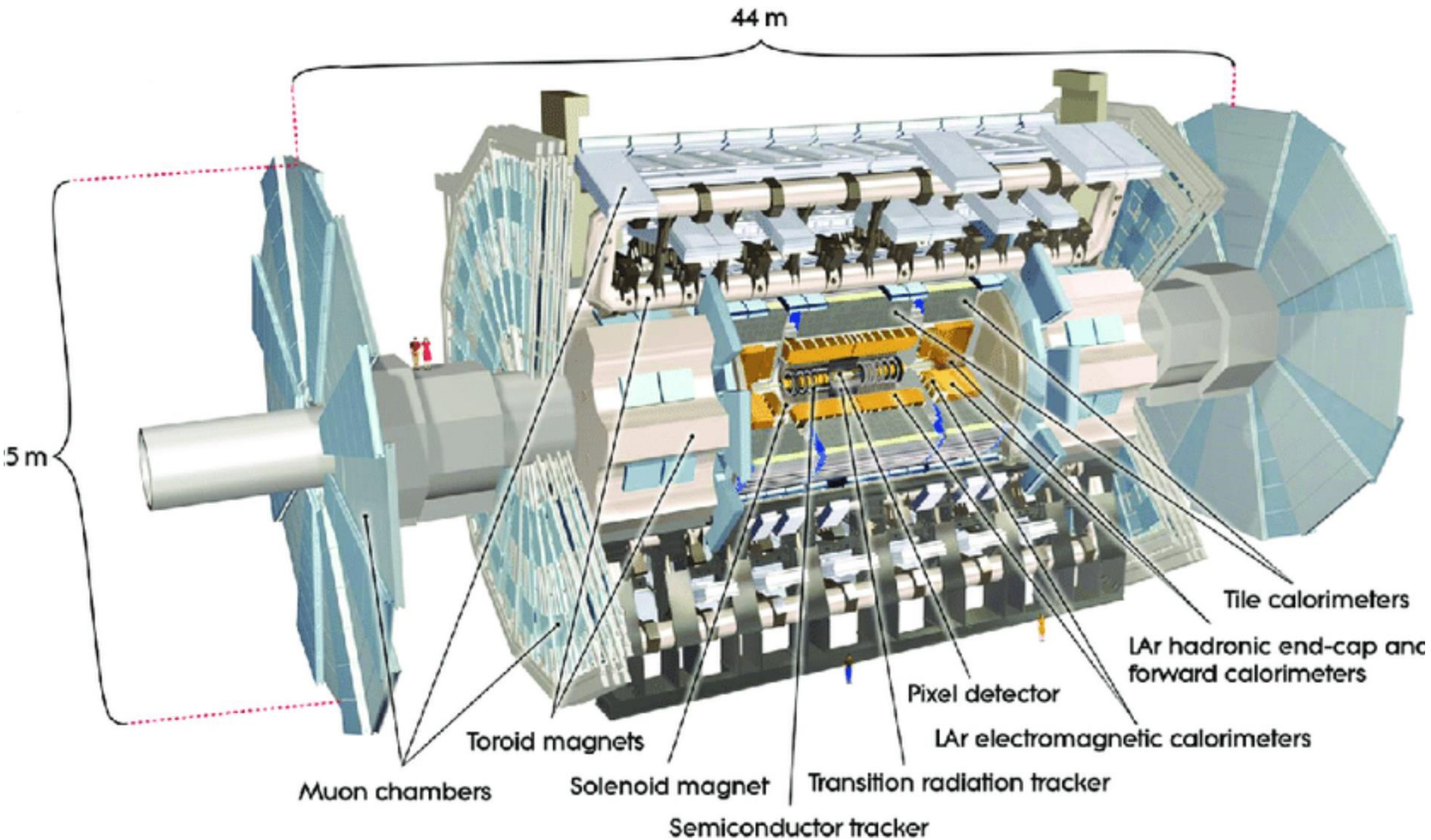
20-24 October 2025

The Large Hadron Collider at CERN



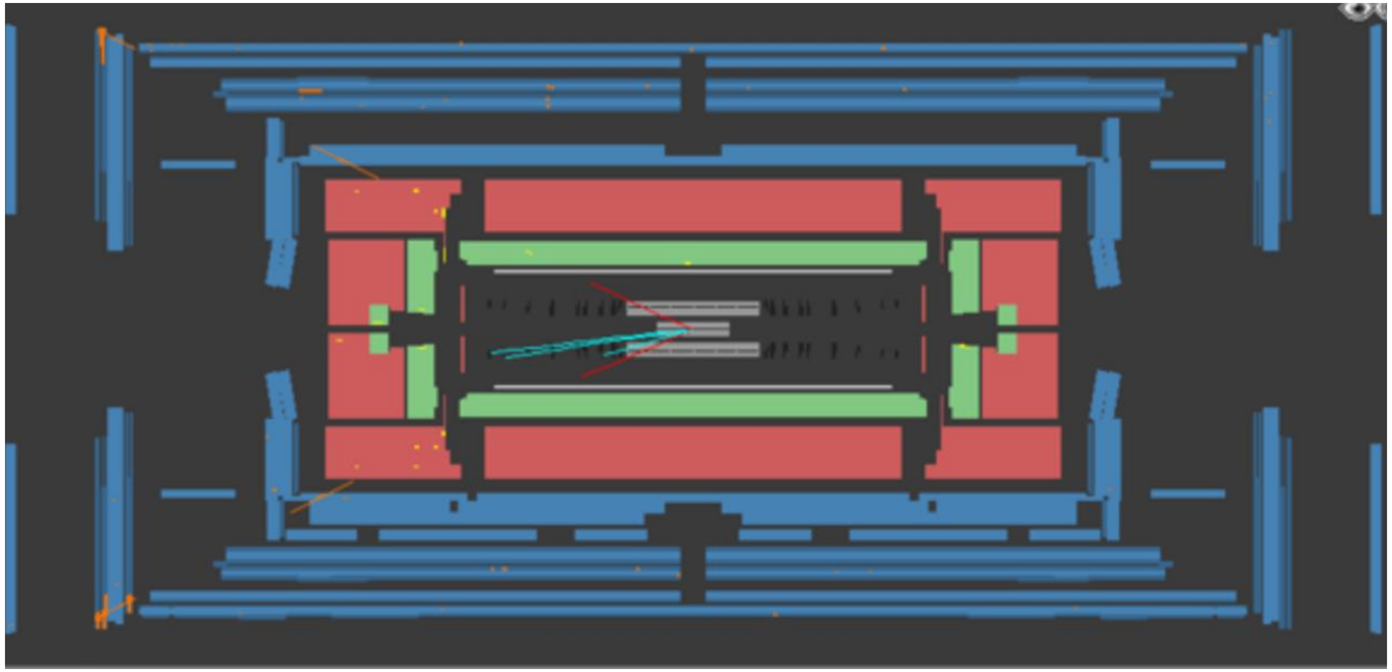


The ATLAS experiment

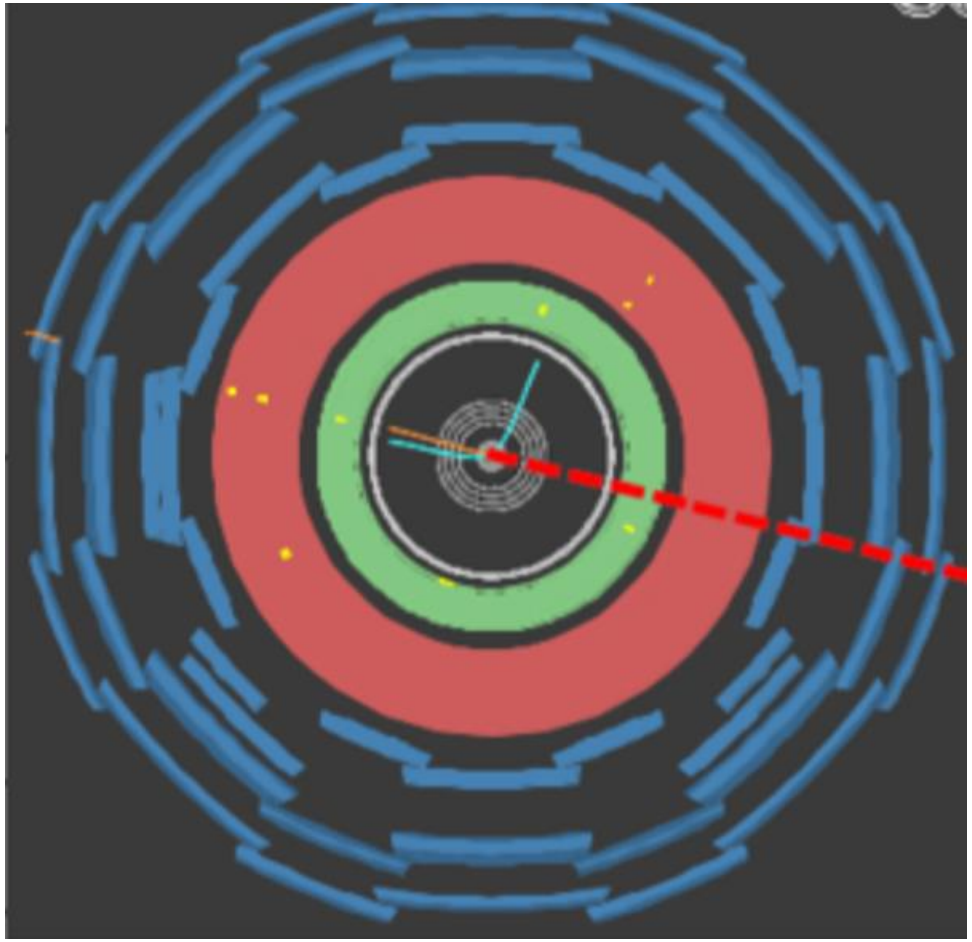


The ATLAS experiment

Longitudinal plane



Transverse plane



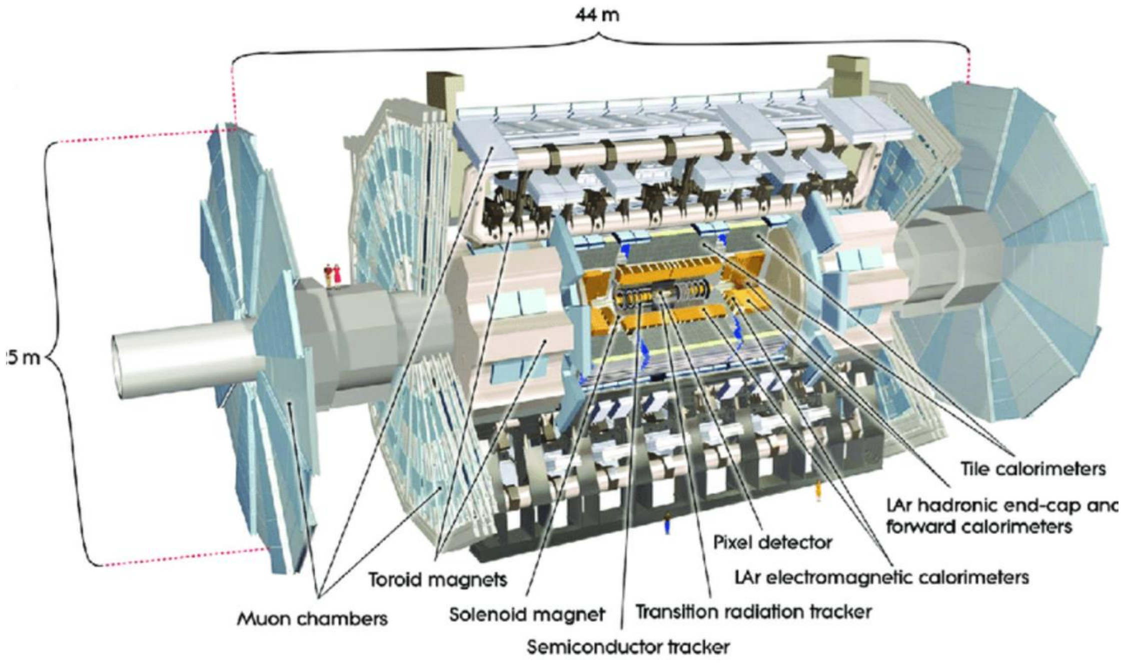
Inner Tracker : tracks the paths of <u>charged</u> particles, bent in solenoid magnetic field

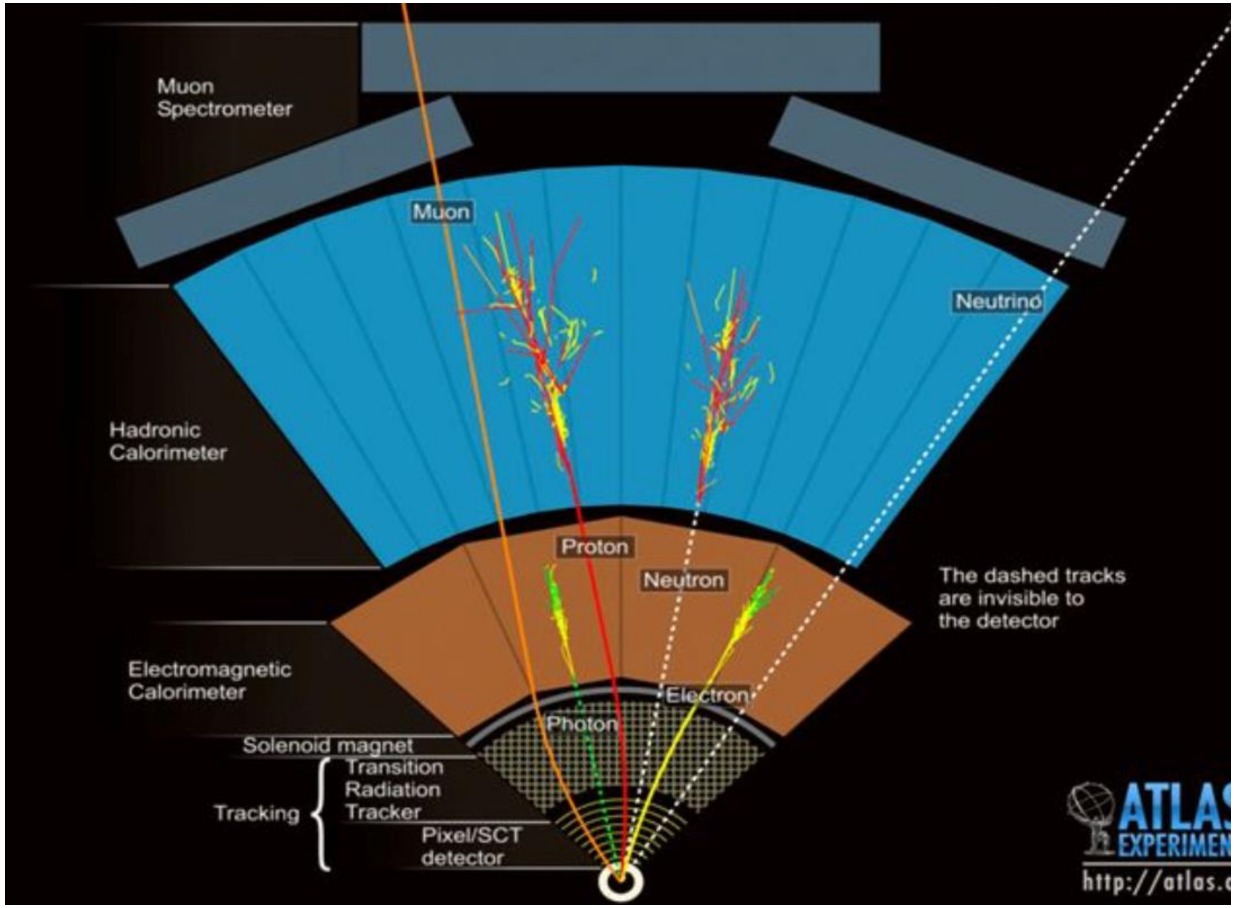
Electromagnetic calorimeter: measured energy of Electrons, Photons

Hadronic calorimeter: measures energy of hadronic jets

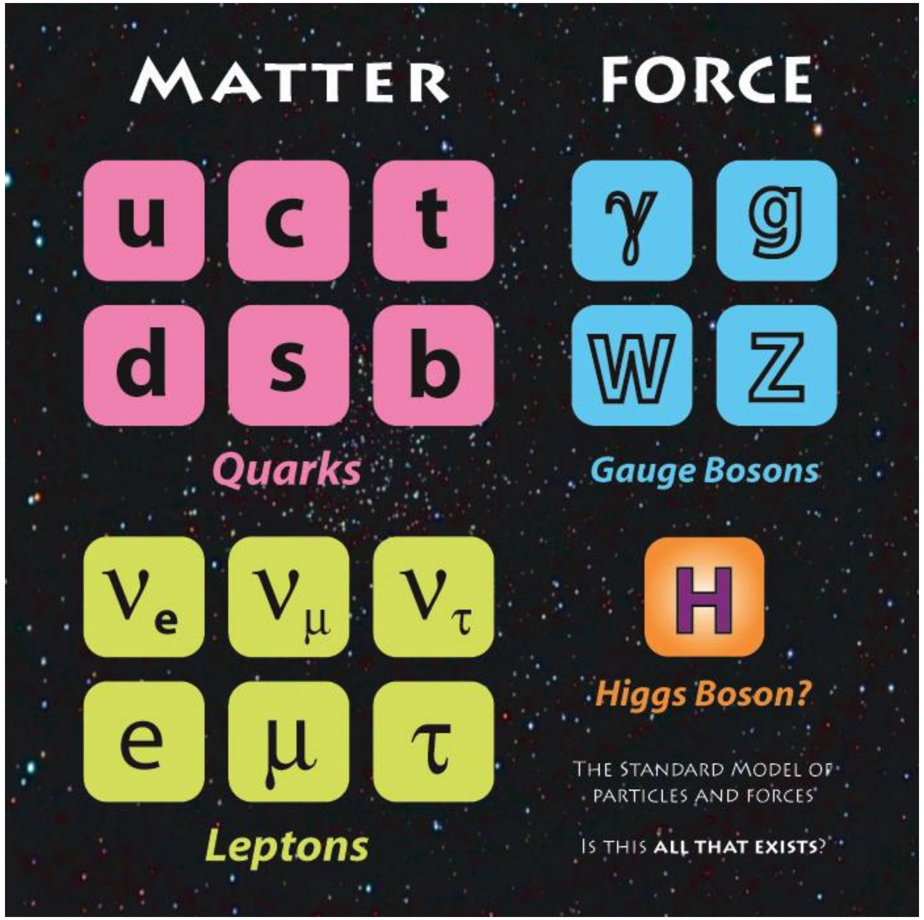
Muon spectrometer: Tracks the paths of muons, bent in toroidal magnetic field

The ATLAS experiment



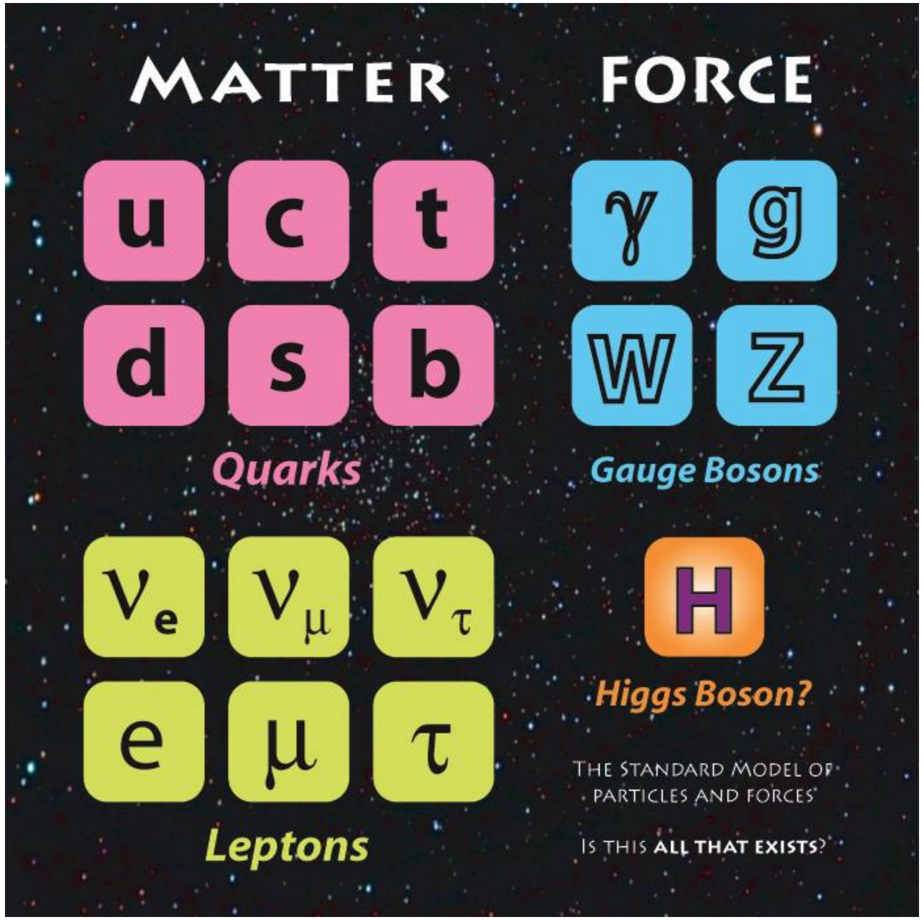


The ATLAS experiment



- **What Objects do we reconstruct?**
  - Electrons
  - Muons
  - Photons
  - Jets
  - Bjets
  - Missing Energy
- **What variables?**
  - Number of particles
  - Transverse momentum (PT)
  - Energy
  - Phi and eta
  - Charge
- **What do we also calculate with the variables?**
  - Invariant mass (of the two or more objects)
  - Angles between objects

The ATLAS experiment

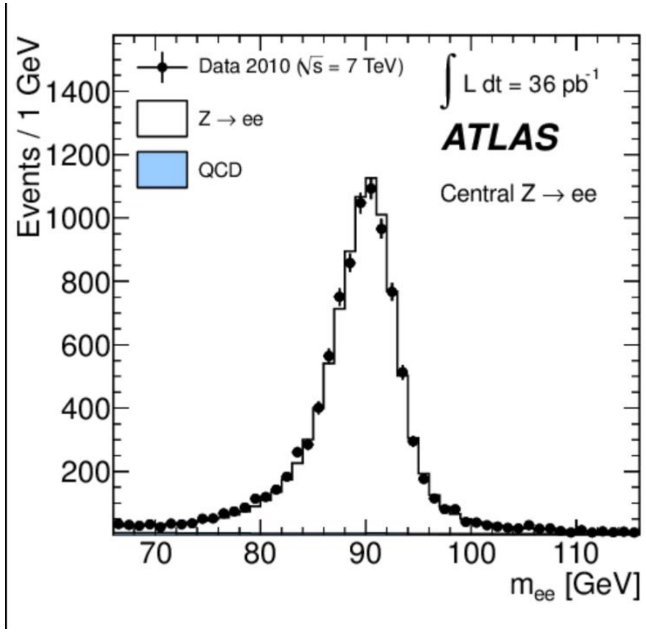


- **Quarks** – hadronise into jets (also gluons!)
  - Detected in the hadronic Calorimeter
  - Bjets can be tagged as they don't hadronise immediately!
  - Top quark does not turn into jets, it decays immediately!
- **Electrons** – stable
  - Track ID, energy absorbed in the EM calorimeter
- **Muons** – travel all through ATLAS, decay later!
  - Track ID, Track Muon Spectrometer
- **Taus** – decay very soon - Reconstruct these
- **Neutrinos** – don't interact - Leave missing Energy!
- **Photons** – No track in ID, energy reconstructed in EM calorimeter
- **Top, W and Z and Higgs** – all decay into other particles immediately, **tau** decays after a few meters.

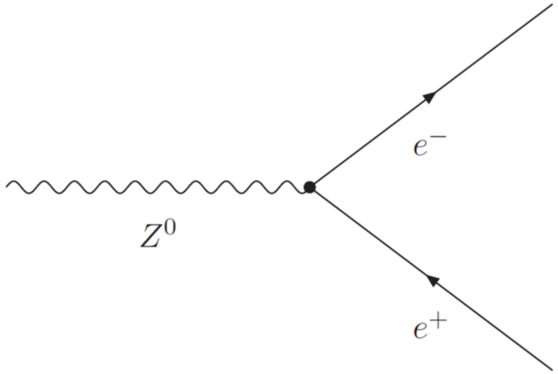
The ATLAS experiment

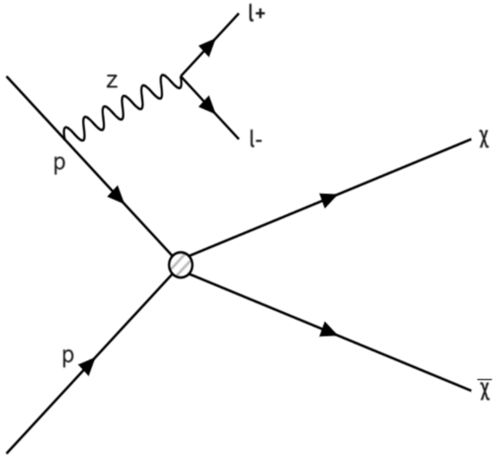
- **What do we also calculate with the variables?**
  - Invariant mass (of the two leptons)
  - Angles between leptons

$$m_0^{(Z)} = \sqrt{\left(\frac{(E_{e^-} + E_{e^+})}{c^2}\right)^2 - \left(\frac{\vec{p}_{e^-} + \vec{p}_{e^+}}{c}\right)^2}$$

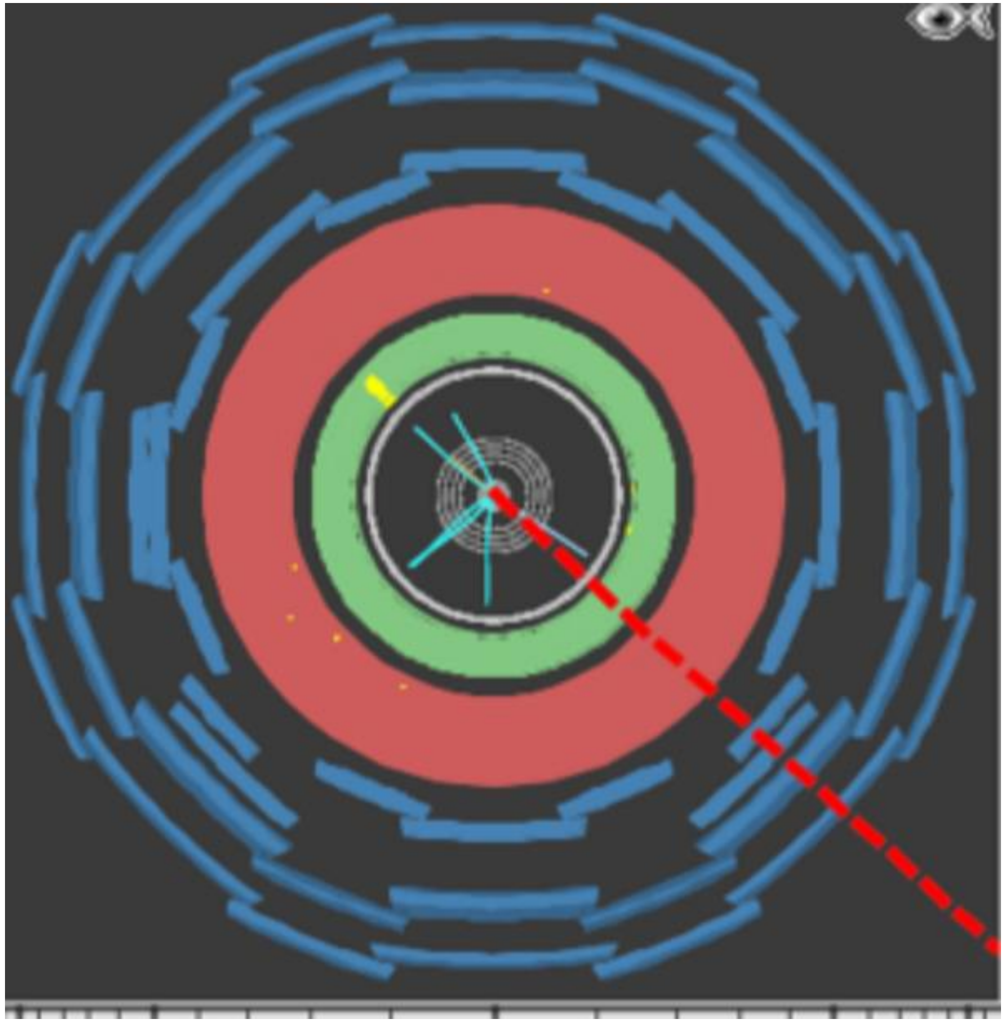


The angle between to leptons
Z-> ee = back to back, phi = ~180
DM+Z->ee = angle smaller~90





The ATLAS experiment

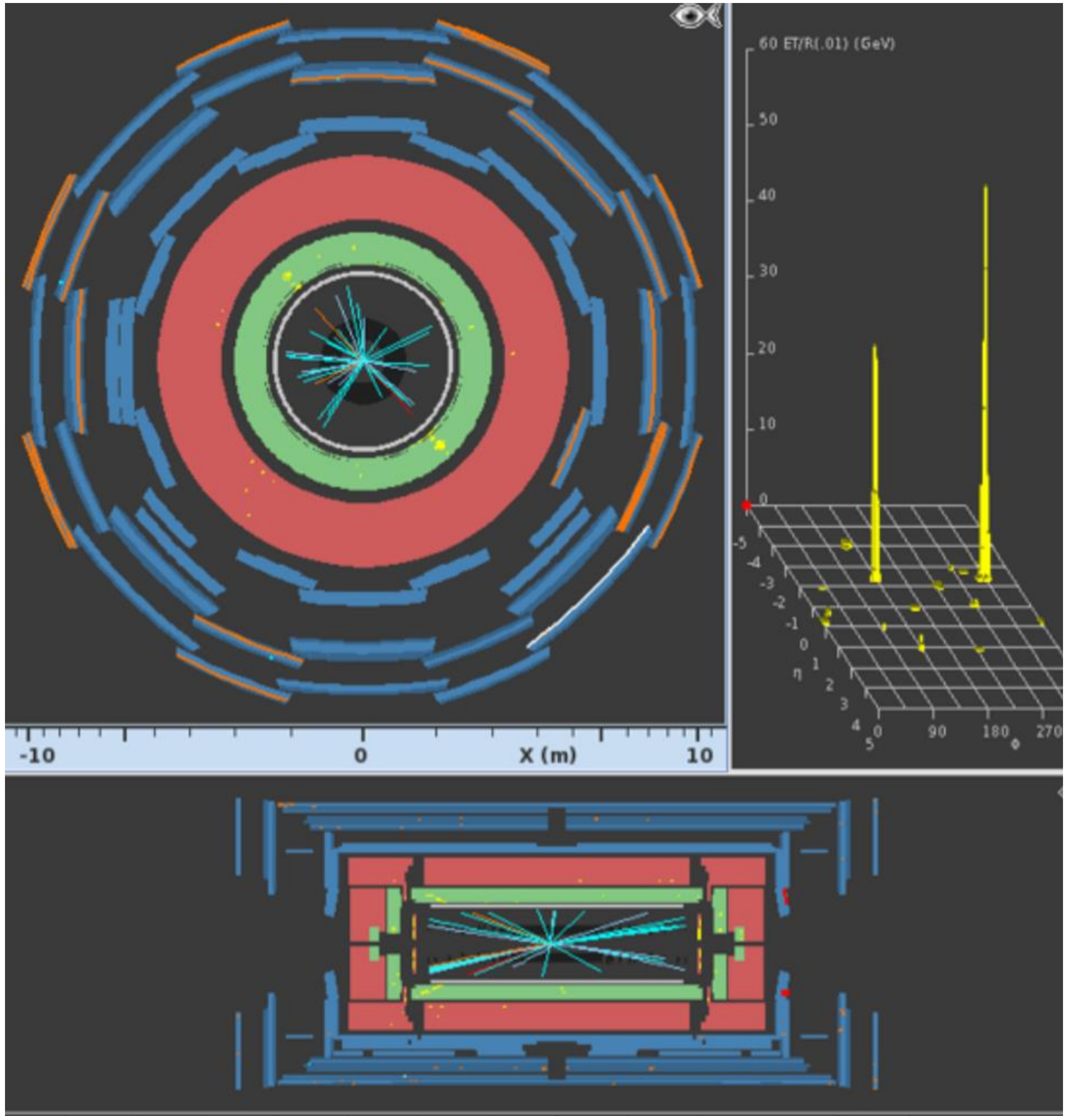


One track point to energy in EM calorimeter,
Track had charge = -1
This particle is an electron

The red dotted line = missing energy
If MissET > 25 GeV it could be a neutrino

Thus we think this is a W-> eν event!

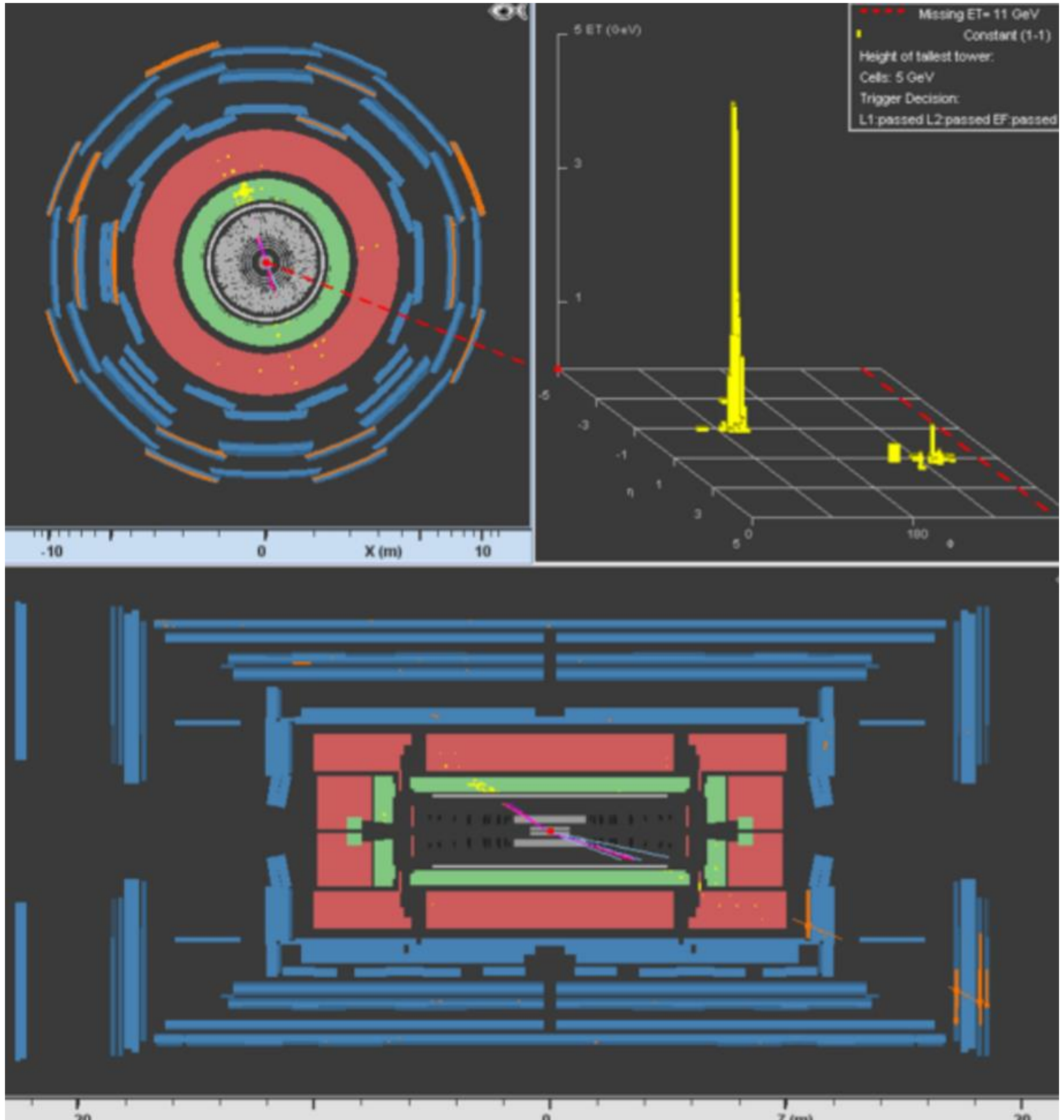
The ATLAS experiment



Two big energy deposits in EM calorimeter
-> Must be photons or electrons
Two tracks are pointing to the energy deposits
One track has charge -1 and one has +1
These are e+ and e-

We also could work out the invariant mass 9see later which ~ 88 GV
This event is Z-> ee

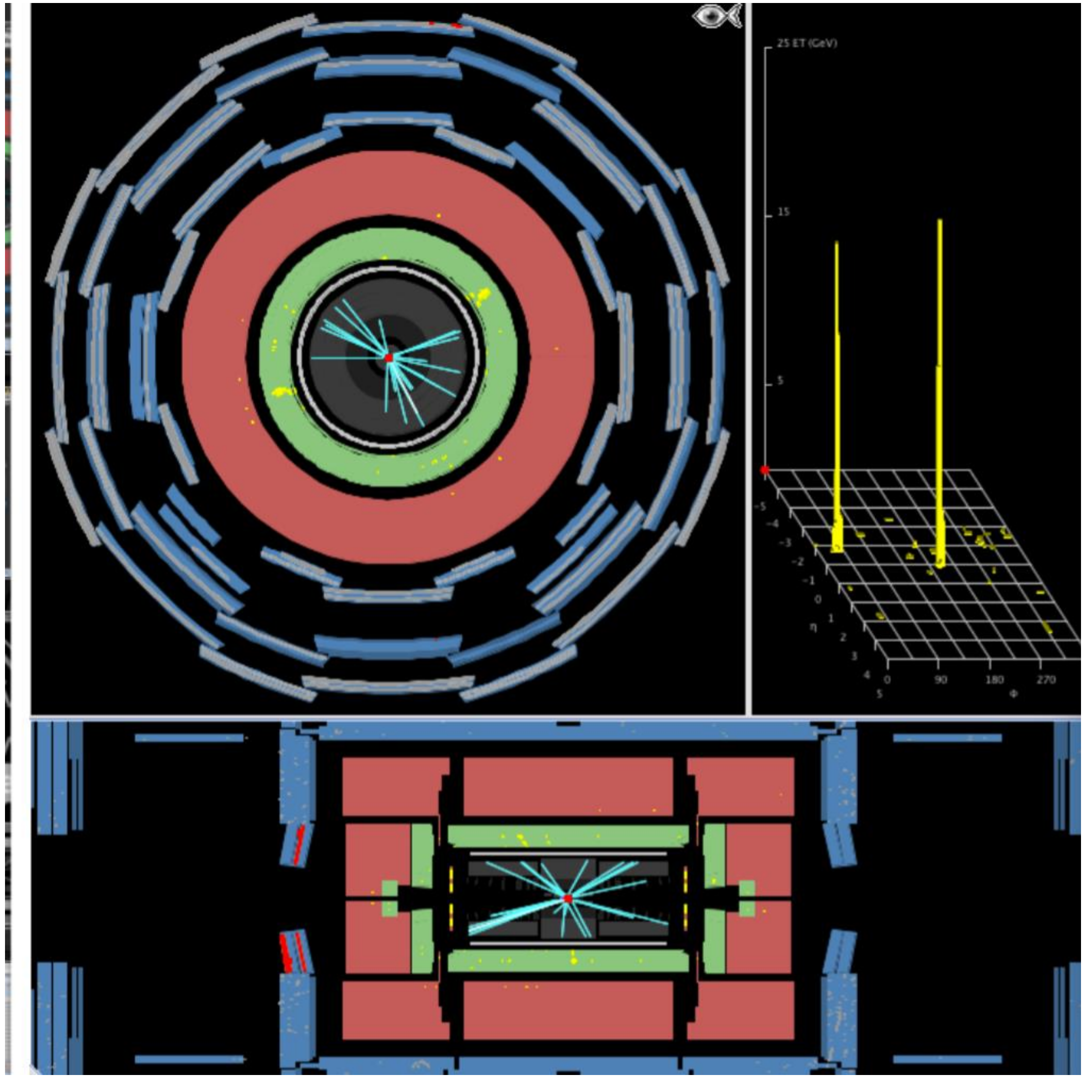
The ATLAS experiment



Missing ET is small < 25 GeV so we don't think this is from neutrino

We see lots of track together in bundles
These probably 2 jets from two quarks or two gluons

The ATLAS experiment



Two big deposits in the EM calorimeter
What particles are here?

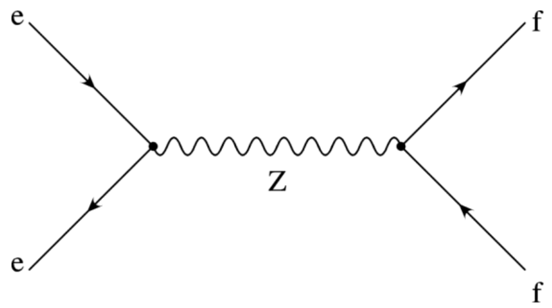
The invariant mass = 110 GeV

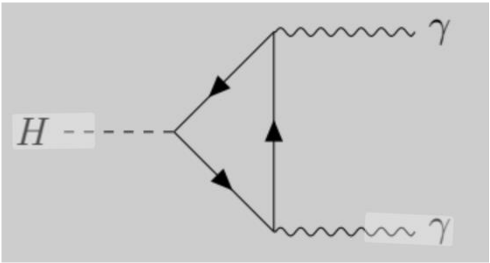
Hint : mass of Higgs = 125 GeV

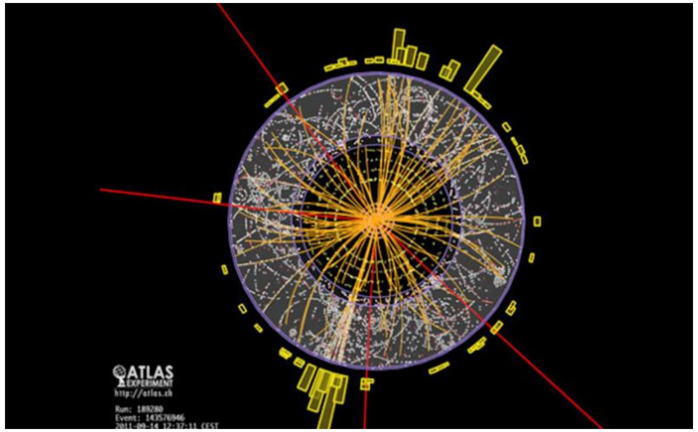
The ATLAS experiment

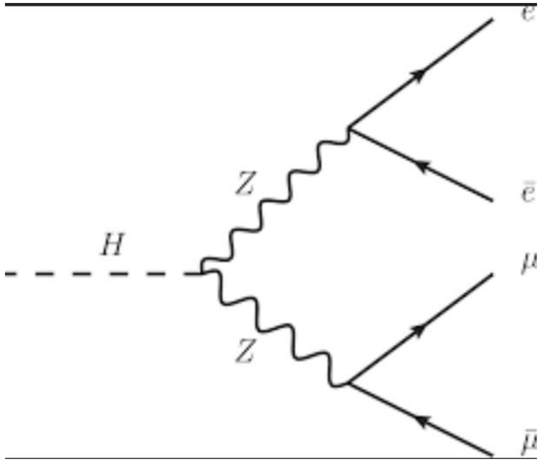
- Some example Processes:
- Z -> lepton (anti)lepton (Z-> ee, Z->μμ)
- Z-> quark (anti)quark (Z->qq, Z->bb)

- Higgs -> photon photon (H->γγ)
- Higgs -> Z-> 4 leptons (H->Z->llll)

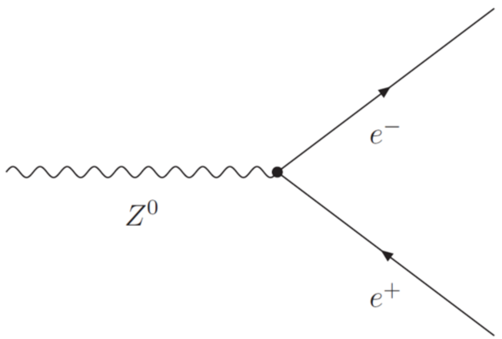




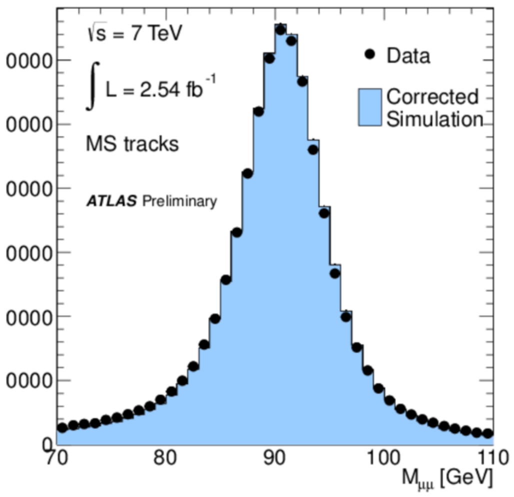




The ATLAS experiment

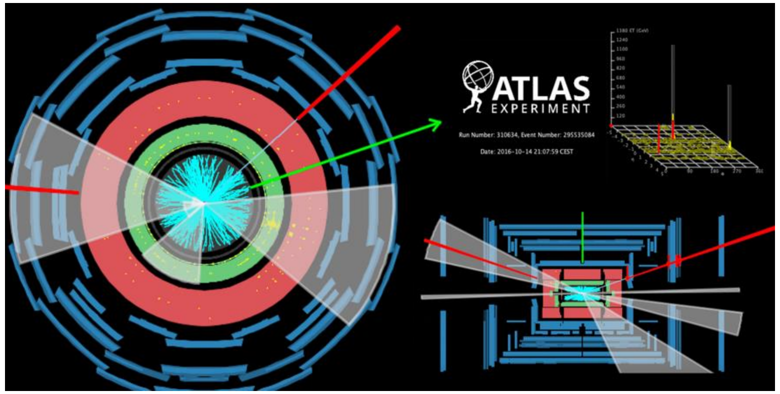


- Event reconstruction:
  - Each proton-proton collision we record the objects detected by ATLAS
  - We then collect all the data together, and look to reconstruct certain processes.
  - For example, if we look for Z bosons, want do we look for in ATLAS
    - Events with 2 leptons of opposite electric charge
    - Events with two quarks of opposite electric charge
    - NO jets, NO Bjets, No missing Energy
    - Another good (cut!) idea is
      - M$\upsilon\upsilon$ > 70, and M$\mu\mu$ < 110

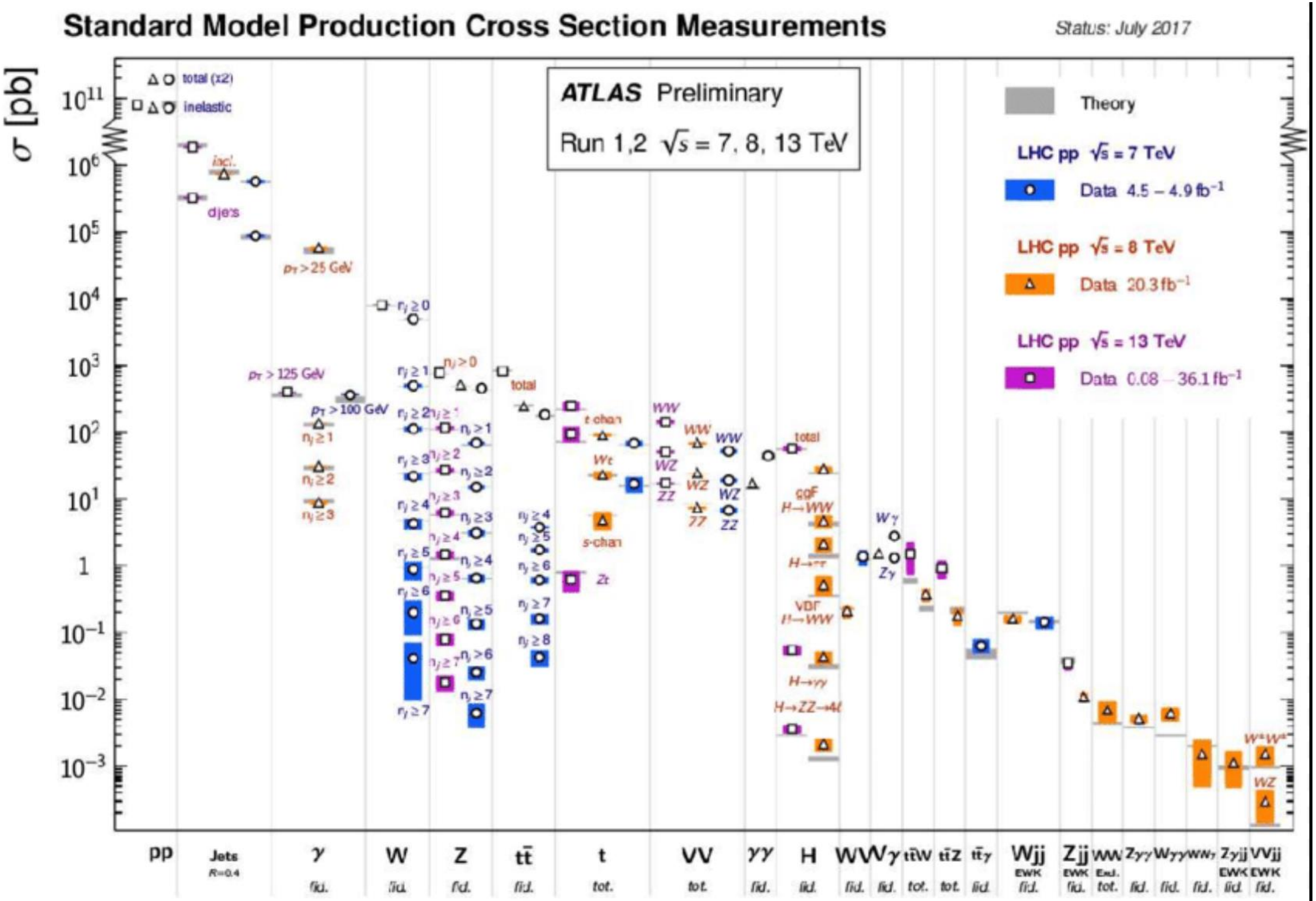


How we do analysis

- We collect the data from ATALS, and store all the information for each event, the objects (e.g. electron), its measured variables (e.g. PT, phi, eta), calculate reconstructed quantities (Mll, Delta Phi)
- We decide what physics we want to do!
  - Measurement of SM particles
  - Searches for New BSM particles
- This week = we will search for dark matter particles!



Standard Model Production Cross Section Measurements

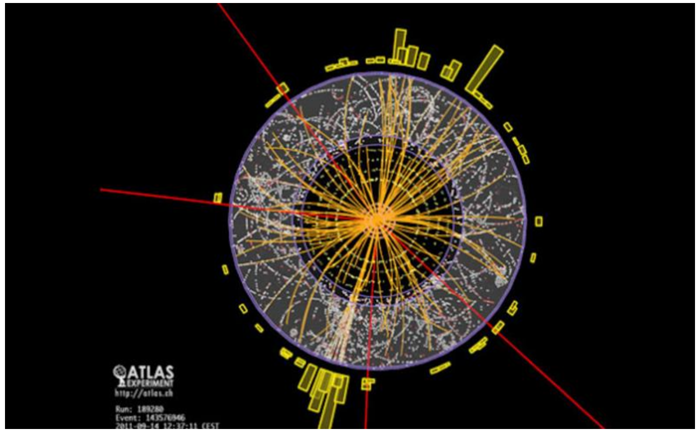


Needle in a haystack!

- Searches are often hard because we are looking for a rare event, so only a very few proton-proton collisions make this particle
  - They are often rare – because we have found all the common things!

- Searches are also often hard because the '**signature**' of the event we are searching for often looks just like many other SM events

Final States & Signatures of processes



- What is the **final state**?
  - **The final particles from the process**
  - E.g. for Z-> ee, the final state is two electrons
    - For W->$\mu\nu$, the final state is a muon and a neutrino
    - For H-> $\gamma\gamma$, the final state is two photons
    - For H-> bb, the final state is two bjets
- What is the **signature**?
  - **This is what we reconstruct in ATLAS, the objects we can see**
    - E.g. For Z-> ee, this is two electrons with opposite charge
      - For W->$\mu\nu$, this is a muon and <u>missing energy</u>
      - For H-> $\gamma\gamma$, this is two photons
      - For H-> bb, this is two jets that are bragged = 2 bjets

Final States & Signatures of processes

- Our **signal** is the process we are searching for

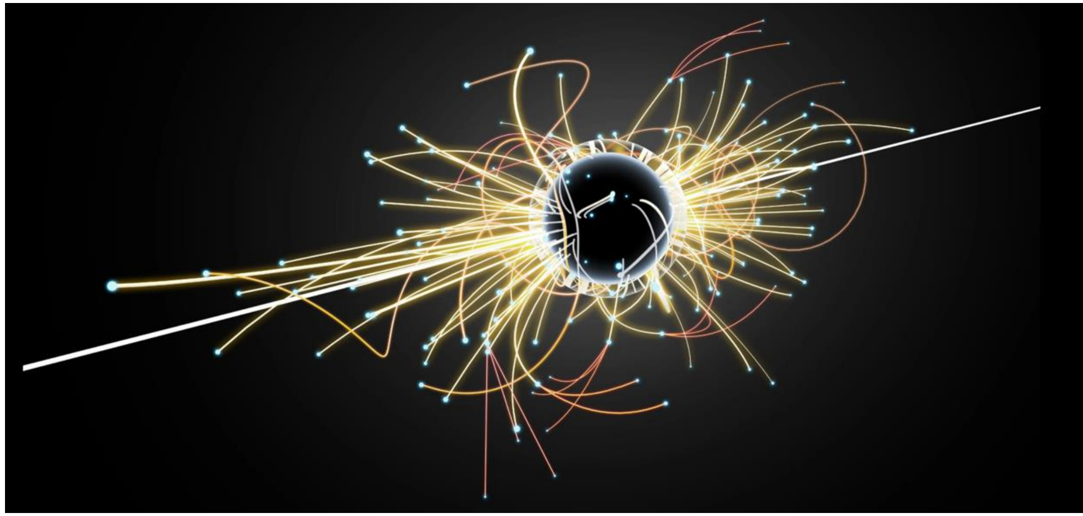
- The **signature** of our signal is what this looks like

- The **background** = any SM processes that have the same signature
  - = Processes that have the same signature
  - Often because of mis-reconstruction
  - For example, a b-jet not being reconstructed
  - Or a charge misidentification of a lepton

Search for Dark Matter!

- This week we are going to design a search analysis using Monte Carlo simulated data for dark Matter.
- The particle we will search for, which we are calling a dark matter candidate is a WIMP
  - **Weakly Interacting massive particle**

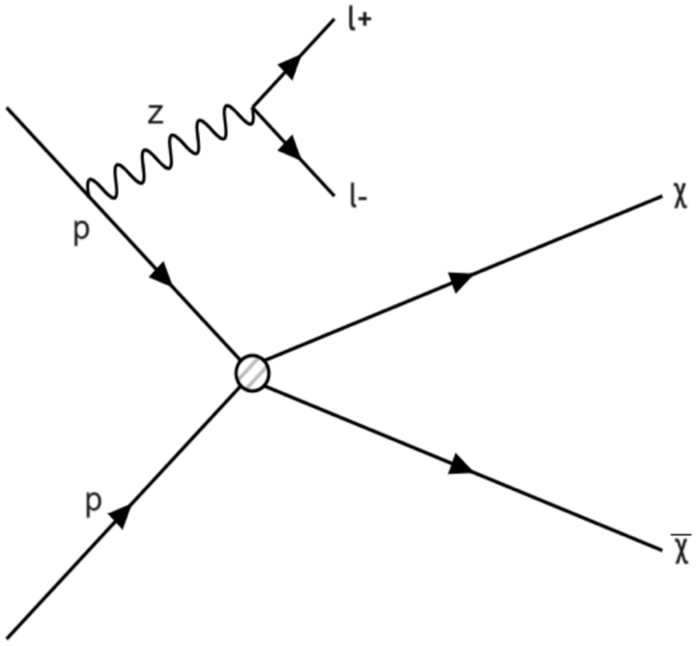
- What does it look like in ATLAS?



- This particles has no electric charge
- Leave the detector without being detected = <u>Missing ET</u>
- The dark matter particle is massive = very big for a fundamental particle = <u>LARGE missing ET</u>

Search for Dark Matter!

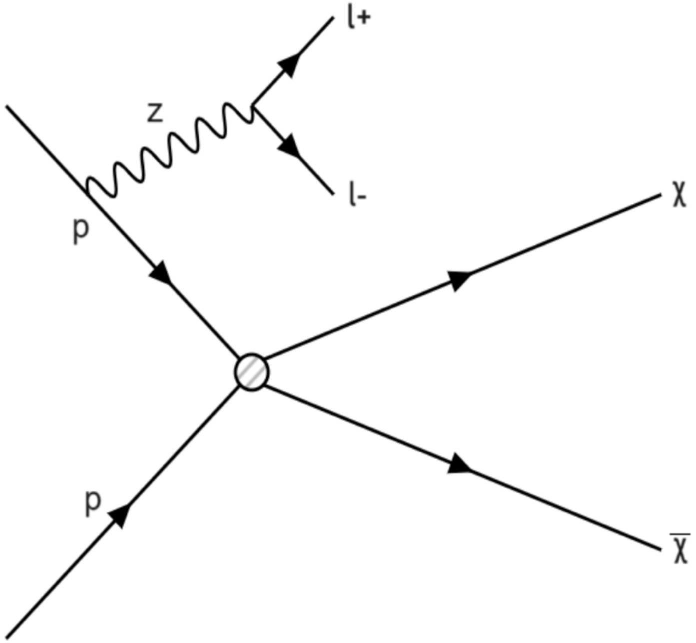
- We are going to search for a specific process, where the Dark Matter particles are created with a Z-boson
- The Z boson decays into two leptons
- This is our SIGNAL process:
  - pp-> X X+ Z->ll
  - **Written as DM+Z->ll**



Analysis: SIGNAL

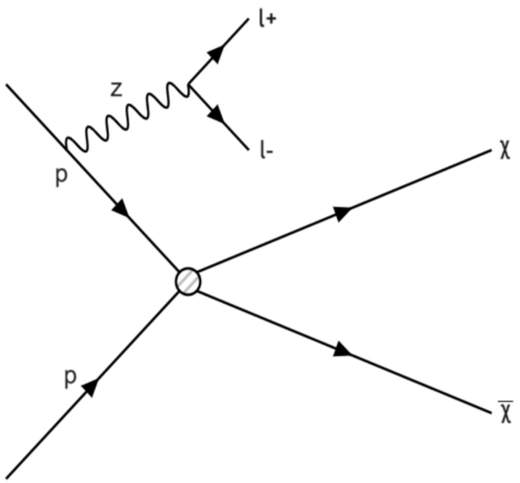
- Our signal is the process we are searching for
  - Signal = Dark Matter + Z boson -> lepton + lepton
  - Signal = DM +Z->ll

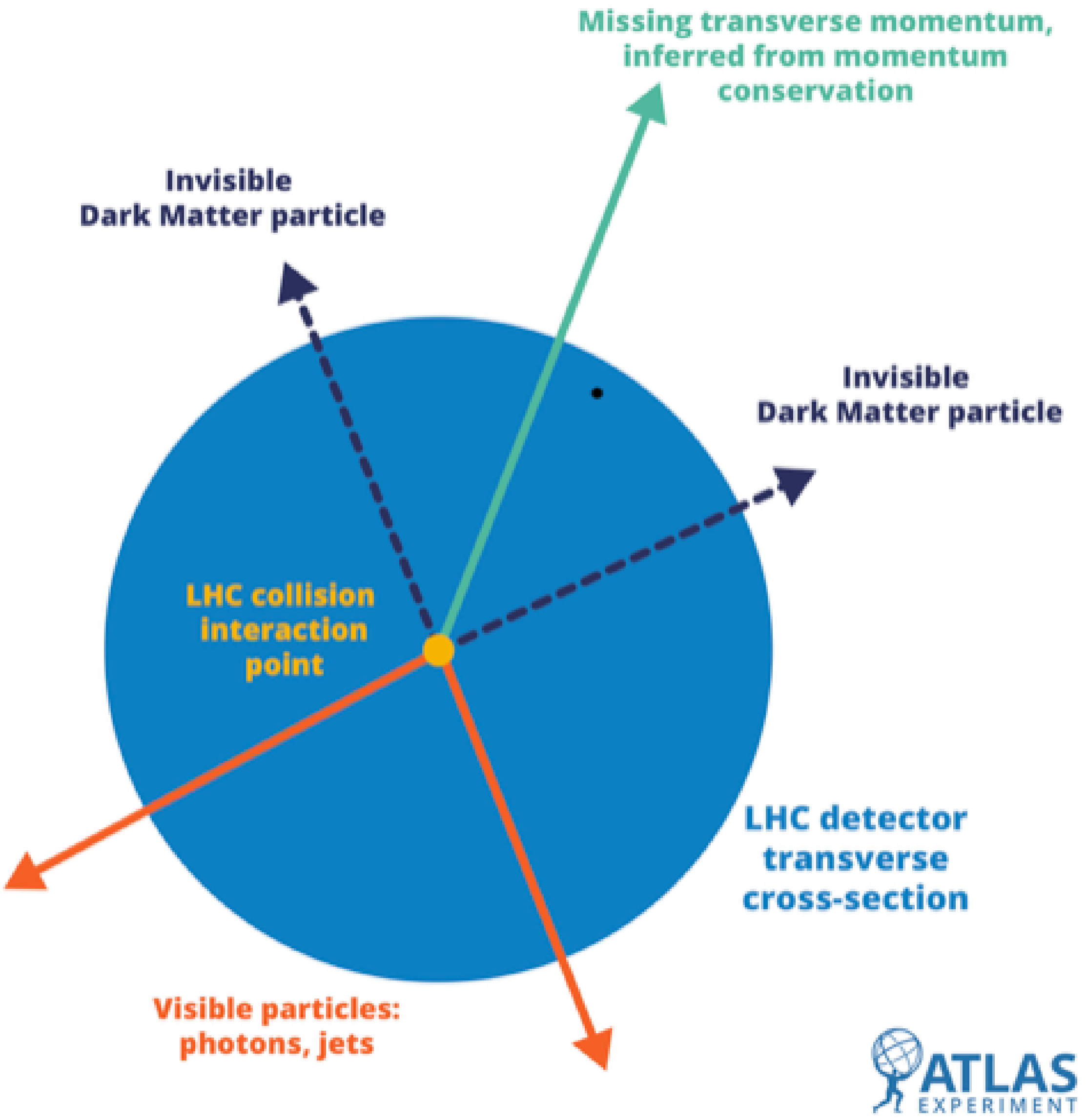
  - What will the signature be?



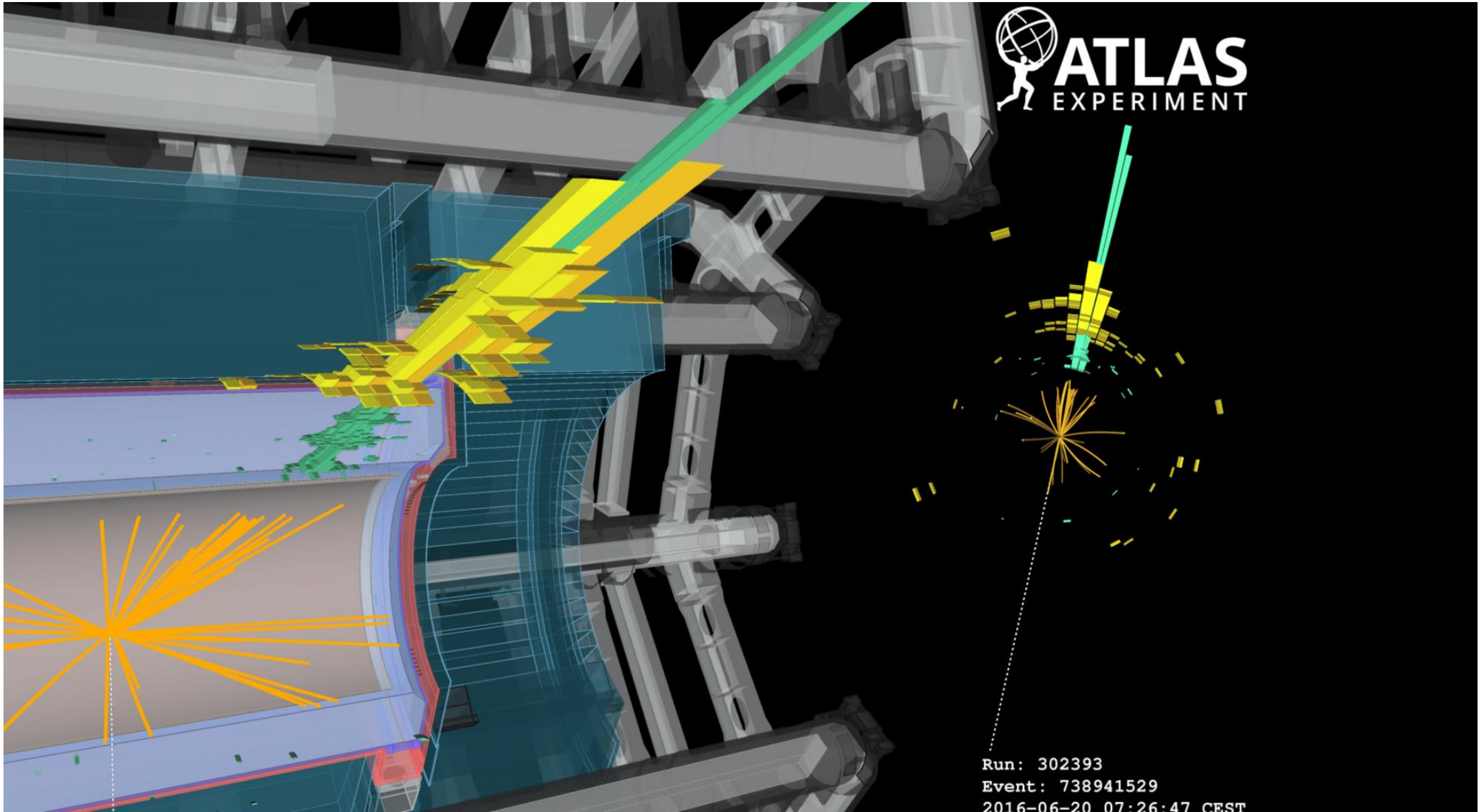
Analysis: Signature

- The Signature of our signal
  - Missing ET + two opposite sign leptons (electrons or muons)
  - Etmiss + 2 OS leptons (e,μ)



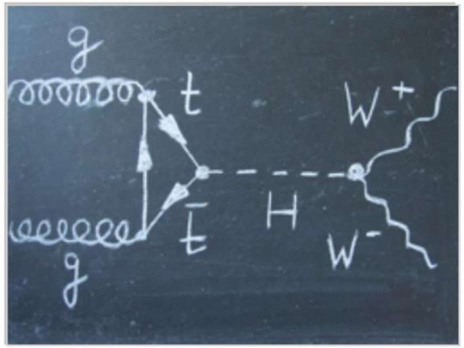


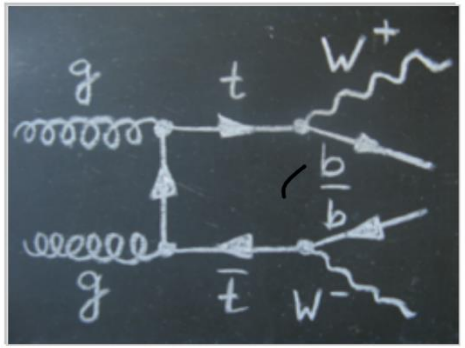
Analysis: Signature



Analysis: background

- The background = any SM process that looks like the Signal
  - = Processes that have the same signature
  - Often because of mis-reconstruction
  - For example, a b-jet not being reconstructed
  - Or a charge misidentification of a lepton



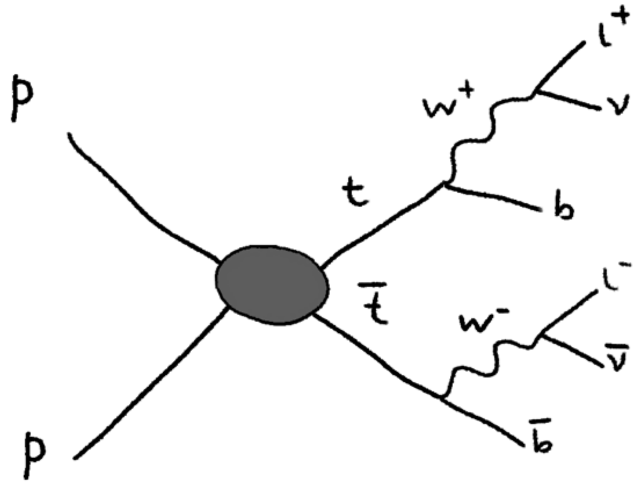


Signal and Background

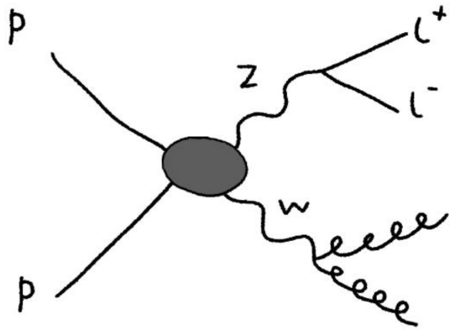
- Our **signal** is the process we are searching for
  - Signal = Dark Matter + Z boson -> lepton + lepton
  - Signal = DM +Z->ll
- The **signature** of our signal
  - Missing ET + two opposite sign leptons (electrons or muons)
  - Etmiss + 2 OS leptons (e,μ)
- The **background** = any SM processes that have the same signature
  - Signature of the backgrounds = Etmss + 2OSl

Analysis: Background

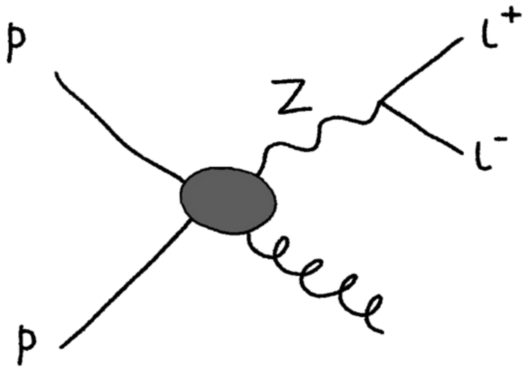
Non-resonant ll ($t\bar{t}$):



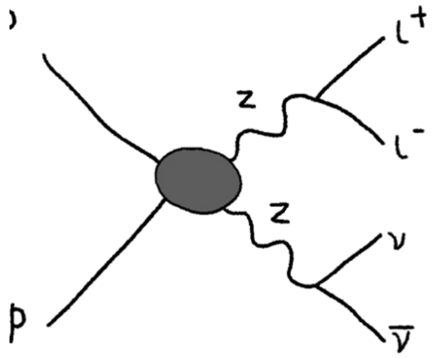
WZ:



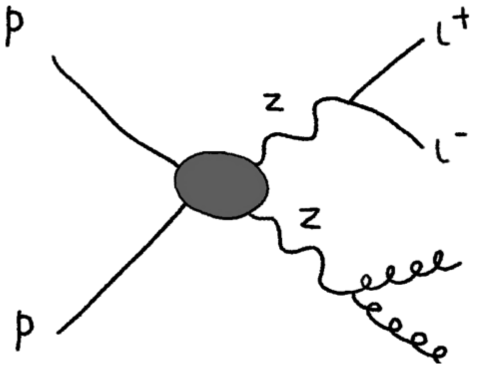
Z+jets:



ZZ:

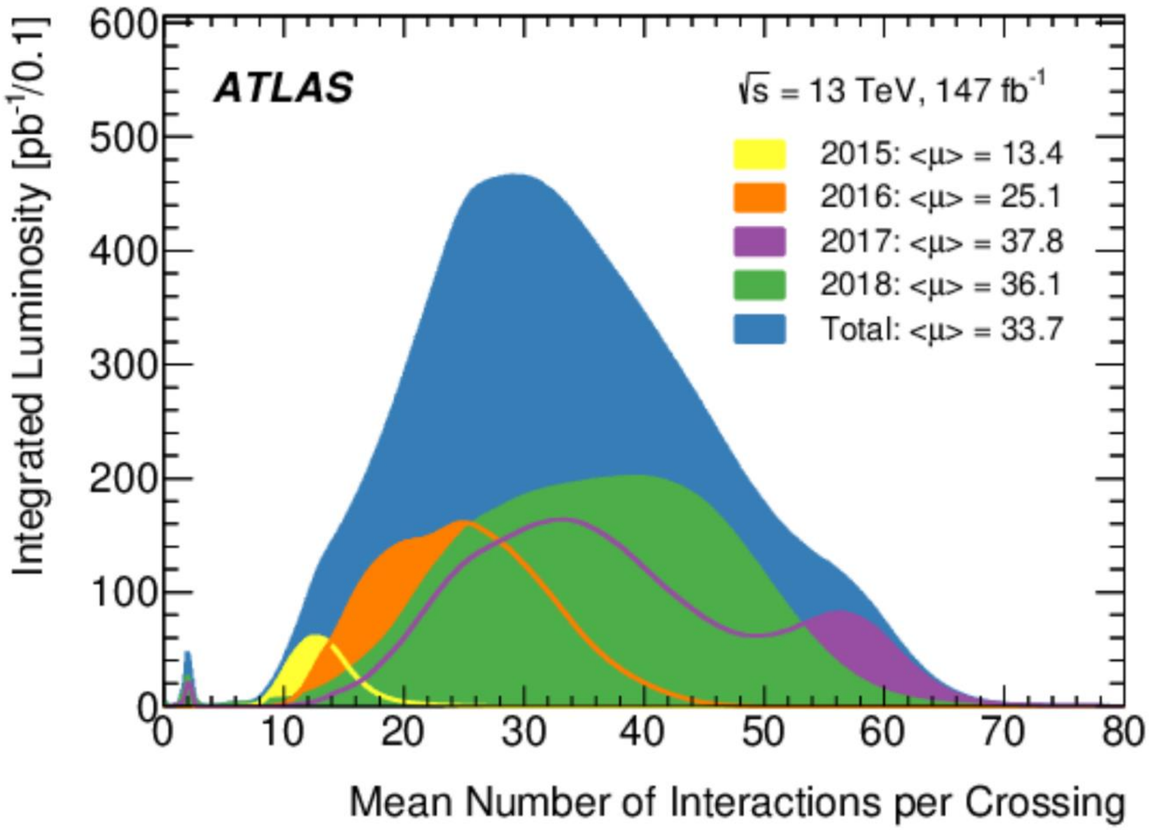


OR



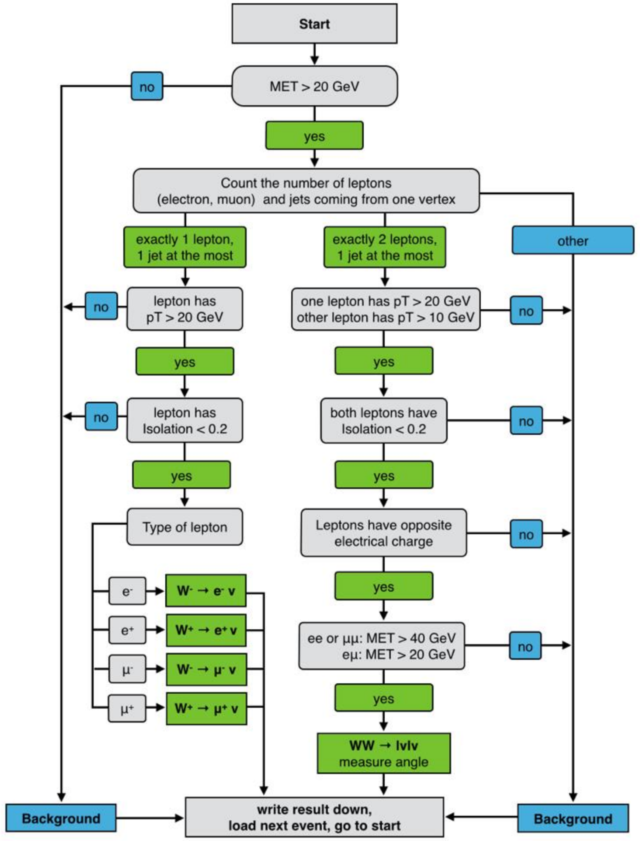
Search for Dark Matter

- How do we find it – how do we do the experiment
- We collide protons for an amount of time to get our dataset
  - Our data set has a number o events. In particle physics we measure the number of events in terms of luminosity
- This data set will have everything
  - How many Higgs, W and Z etc.
- <u>We simulate Monte call simulations of our SIGNAL and backgrounds to develop and optimise our analysis</u>



Search for Dark Matter

Example of a cut flow



- We make cuts on the data to find events that look like our signal!

- Some background events will also look like our signal (hey have a similar signature) so we need to be clever to think how to get rid of more background without getting rid of too much background
  - Our CUTS Would start:
    - Must have exactly TWO Opposite sign leptons
    - Must have large missing ET (lets say >25 GeV)

Search for Dark Matter

- How do we find it – how do we lets look at the background again
  - What cuts can we do to get rid of more of these events?
- This is doing a cut based optimisation analysis:

Table 1: Event selection criteria in the $\ell\ell + E_{\mathrm{T}}^{\mathrm{miss}}$ search.

	Selection criteria
Two leptons	Two opposite-sign leptons, leading (subleading) $p_{\mathrm{T}} > 30$ (20) GeV
Third lepton veto	Veto events if any additional lepton with $p_{\mathrm{T}} > 7$ GeV
$m_{\ell\ell}$	$76 < m_{\ell\ell} < 106$ GeV
$E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss}}/H_{\mathrm{T}}$	$E_{\mathrm{T}}^{\mathrm{miss}} > 90$ GeV and $E_{\mathrm{T}}^{\mathrm{miss}}/H_{\mathrm{T}} > 0.6$
$\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\ell\ell}, \vec{E}_{\mathrm{T}}^{\mathrm{miss}})$	$\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\ell\ell}, \vec{E}_{\mathrm{T}}^{\mathrm{miss}}) > 2.7$ radians
$\Delta R_{\ell\ell}$	$\Delta R_{\ell\ell} < 1.8$
Fractional p_{T} difference	$\left\| p_{\mathrm{T}}^{\ell\ell} - p_{\mathrm{T}}^{\mathrm{miss,jets}} \right\| / p_{\mathrm{T}}^{\ell\ell} < 0.2$
b-jets veto	$N(b\text{-jets}) = 0$ with b-jet $p_{\mathrm{T}} > 20$ GeV and $\|\eta\| < 2.5$

Search for Dark Matter

- We optimise the cuts as best we can to get most amount of background and least amount of signal
- We measure this as a number signal/sqrt (number background)

Final State	ee	$\mu\mu$
Observed Data	437	497
Signal		
$ZH \to \ell\ell + \text{inv}$ ($B_{H\to\text{inv}} = 30\%$)	$32 \pm 1 \pm 3$	$34 \pm 1 \pm 3$
DM ($m_{\text{med}} = 500$ GeV, $m_\chi = 100$ GeV) $\times$ 0.27	$10.8 \pm 0.3 \pm 0.8$	$11.1 \pm 0.3 \pm 0.8$
Backgrounds		
$qqZZ$	$212 \pm 3 \pm 15$	$221 \pm 3 \pm 17$
$ggZZ$	$18.9 \pm 0.3 \pm 11.2$	$19.3 \pm 0.3 \pm 11.4$
WZ	$106 \pm 2 \pm 6$	$113 \pm 3 \pm 5$
Z + jets	$30 \pm 1 \pm 28$	$37 \pm 1 \pm 19$
Non-resonant-$\ell\ell$	$30 \pm 4 \pm 2$	$33 \pm 4 \pm 2$
Others	$1.4 \pm 0.1 \pm 0.2$	$2.5 \pm 2.0 \pm 0.8$
Total Background	$399 \pm 6 \pm 34$	$426 \pm 6 \pm 28$

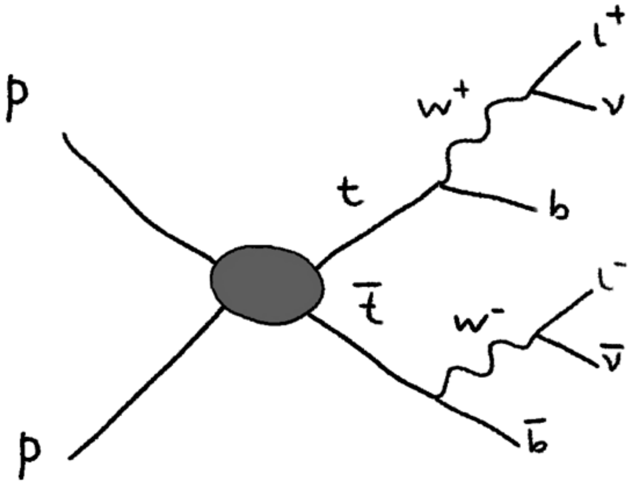
Search for Dark Matter

- We usually use hundreds of variables in our analyses.
- This week we have chosen a few important ones

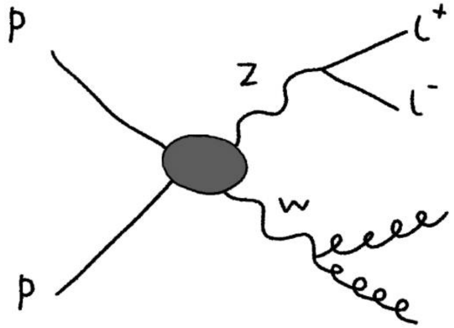
	totalWeight	sum_lep_charge	lead_lep_pt	sublead_lep_pt	mll	ETmiss	dRll	dphi_pTll_ETmiss	fractional_pT_difference	ETmiss_over_HT	N_bjets
0	0.041298	0	113.229602	37.342027	91.874195	124.311867	1.568130	3.036052	0.047148	0.599380	0
1	0.042212	0	89.615922	31.122283	87.832052	105.491891	1.697633	3.083360	0.032974	0.873724	0
2	0.061651	0	112.169008	65.336797	87.876299	170.239734	1.030940	3.133160	0.028335	0.959066	0
3	0.067919	0	108.710273	93.620266	89.113704	221.266453	0.872517	3.052979	0.005579	0.940918	0
4	0.085524	0	85.937289	40.265805	88.230021	132.649672	1.573029	2.554065	0.040406	0.613755	0

Analysis: Background

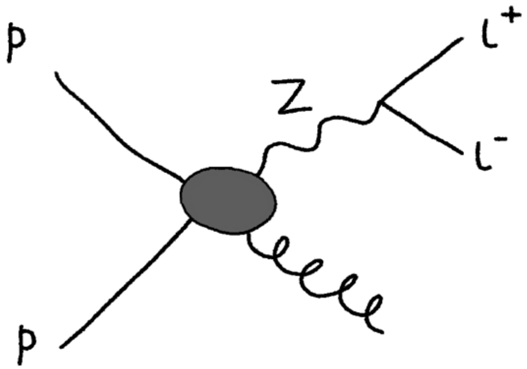
Non-resonant ll ($t\bar{t}$):



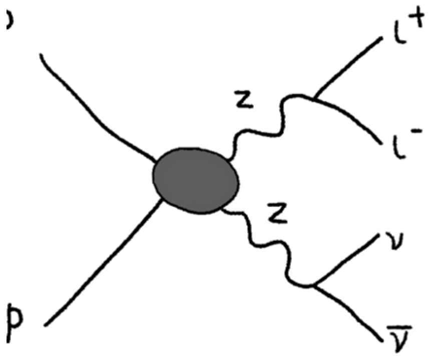
WZ:



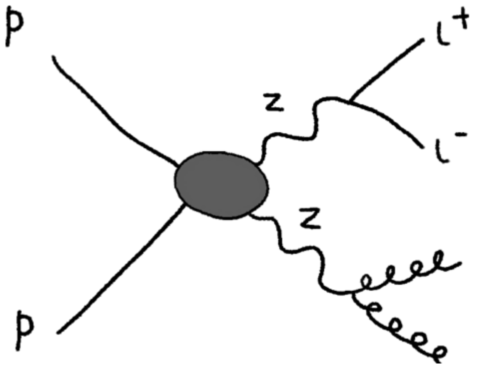
Z+jets:



ZZ:



OR



Search for Dark Matter

Signal dataset: DM_300, Background datasets: NR_ll, Z_jets, WZ or ZZ.
Each row represents a single event or proton-proton collision, inside ATLAS.
Each column represents a different type of measurement / variable of the proton-proton collisions.

1. **sum_lep_charge** represents the total charge for the leptons that ATLAS detects after the collision. Generally, these will have opposite charge, and so they add to zero.

2. **lead_lep_pt** represents the transverse momentum of the fastest lepton that ATLAS detects.

3. **sublead_lep_pt** represents the transverse momentum of the second-fastest lepton that ATLAS detects

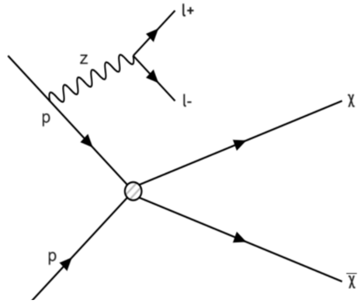
4. **Mll** represents the "dilepton invariant mass"

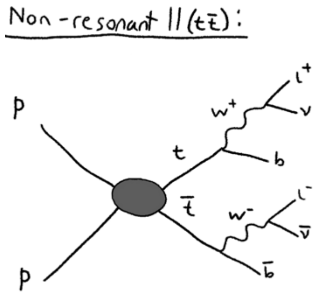
5. **Etmiss** represents the missing transverse momentum.

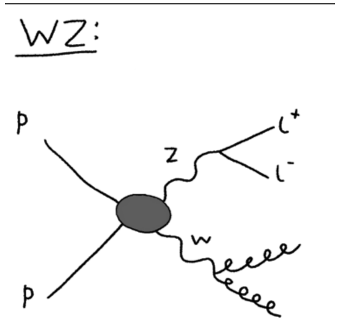
6. **HT** is the scalar sum of all the momentum after the collision; so ETmiss_over_HT is a measurement of how much of the momentum in a collision is missing.

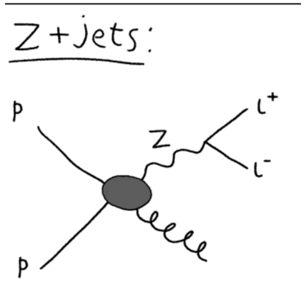
Search for Dark Matter

	totalWeight	sum_lep_charge	lead_lep_pt	sublead_lep_pt	mll	ETmiss	dRll	dphi_pTll_ETmiss	fractional_pT_difference	ETmiss_over_HT	N_bjets
0	0.041298	0	113.229602	37.342027	91.874195	124.311867	1.568130	3.036052	0.047148	0.599380	0
1	0.042212	0	89.615922	31.122283	87.832052	105.491891	1.697633	3.083360	0.032974	0.873724	0
2	0.061651	0	112.169008	65.336797	87.876299	170.239734	1.030940	3.133160	0.028335	0.959066	0
3	0.067919	0	108.710273	93.620266	89.113704	221.266453	0.872517	3.052979	0.005579	0.940918	0
4	0.085524	0	85.937289	40.265805	88.230021	132.649672	1.573029	2.554065	0.040406	0.613755	0









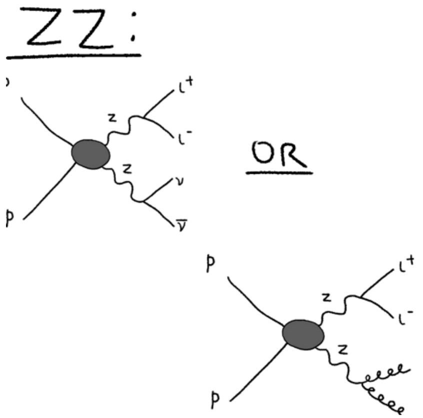


Table 1: Event selection criteria in the $\ell\ell + E_{\mathrm{T}}^{\mathrm{miss}}$ search.

	Selection criteria
Two leptons	Two opposite-sign leptons, leading (subleading) $p_{\mathrm{T}} > 30$ (20) GeV
Third lepton veto	Veto events if any additional lepton with $p_{\mathrm{T}} > 7$ GeV
$m_{\ell\ell}$	$76 < m_{\ell\ell} < 106$ GeV
$E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss}}/H_{\mathrm{T}}$	$E_{\mathrm{T}}^{\mathrm{miss}} > 90$ GeV and $E_{\mathrm{T}}^{\mathrm{miss}}/H_{\mathrm{T}} > 0.6$
$\Delta\phi(\vec{p}_{\mathrm{T}}^{\ell\ell}, \vec{E}_{\mathrm{T}}^{\mathrm{miss}})$	$\Delta\phi(\vec{p}_{\mathrm{T}}^{\ell\ell}, \vec{E}_{\mathrm{T}}^{\mathrm{miss}}) > 2.7$ radians
$\Delta R_{\ell\ell}$	$\Delta R_{\ell\ell} < 1.8$
Fractional p_{T} difference	$\left\|p_{\mathrm{T}}^{\ell\ell} - p_{\mathrm{T}}^{\mathrm{miss,jets}}\right\| / p_{\mathrm{T}}^{\ell\ell} < 0.2$
b-jets veto	$N(b\text{-jets}) = 0$ with b-jet $p_{\mathrm{T}} > 20$ GeV and $\|\eta\| < 2.5$

You are going to choose and optimise these cuts!

Statistical Significance

- Most particle physics analyses are based on hypothesis testing, meaning that the goal is to use data to support or disprove a given hypothesis.
- For example, if we are looking for a **new particle not predicted by the Standard Model**, then our hypothesis would be the presence of this new particle in our data.

- The **null hypothesis** would exactly correspond to the Standard Model prediction, assuming no new particle.

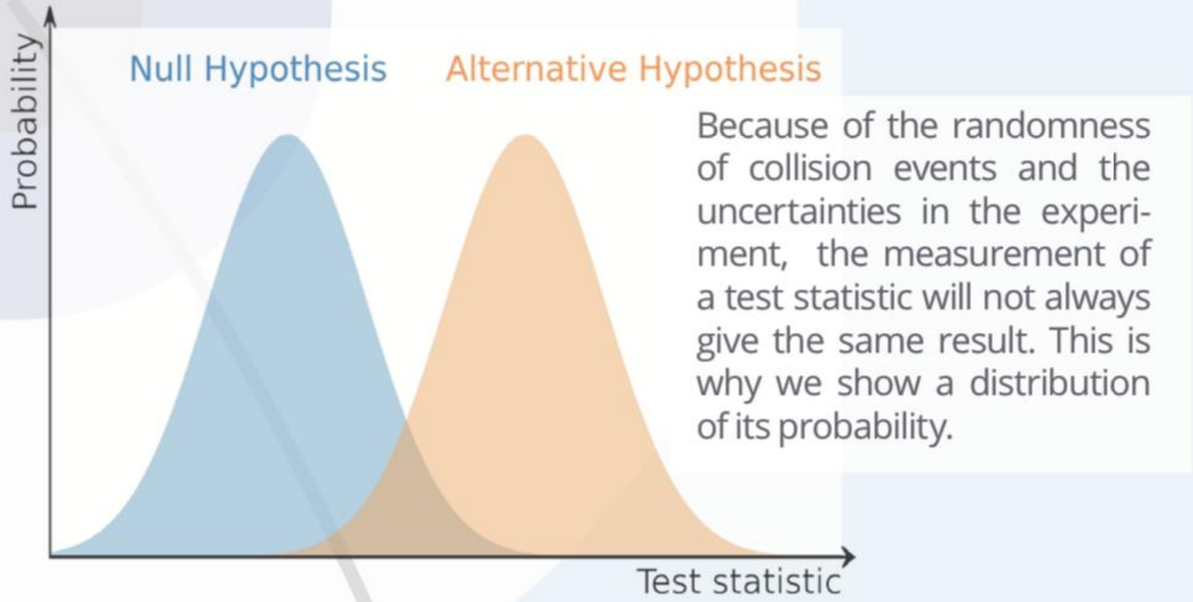
Statistical Significance

- The **p-value** corresponds to the likelihood that an observed test statistic can be explained by the null hypothesis, without the need for an alternative.
- More specifically, it equates to the **fraction of times you would get an observed test statistic showing worse agreement with the null hypothesis than that observed**, assuming the null hypothesis is true.
- The smaller the p-value, the more confidence we have that the null hypothesis can be excluded

Statistical Significance

WHAT IS A TEST STATISTIC?

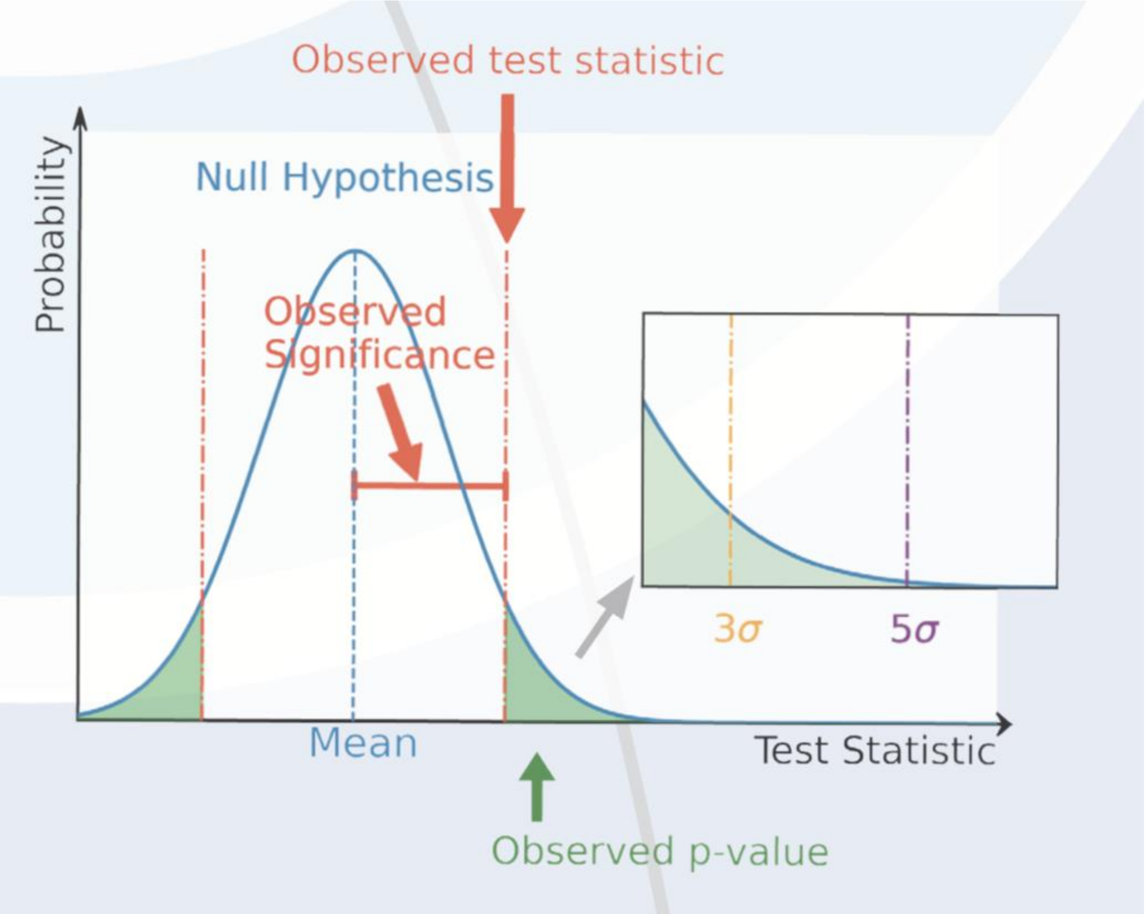
A test statistic is a measurable quantity that is derived from data and is used for hypothesis testing. It can indicate whether the data is closer to the hypothesis or the null hypothesis. An example of a test statistic is the number of observed collision events passing a set of criteria. If this is close to the expected number from the Standard Model, it would tend to support the null hypothesis of no new physics, whereas a larger number could indicate the presence of new physics.



Statistical Significance

- **Significance** is a quantity that expresses how certain we are that the null hypothesis can be excluded.

- Closely related to the **p-value**: the lower the p-value, the higher the significance, and vice-versa

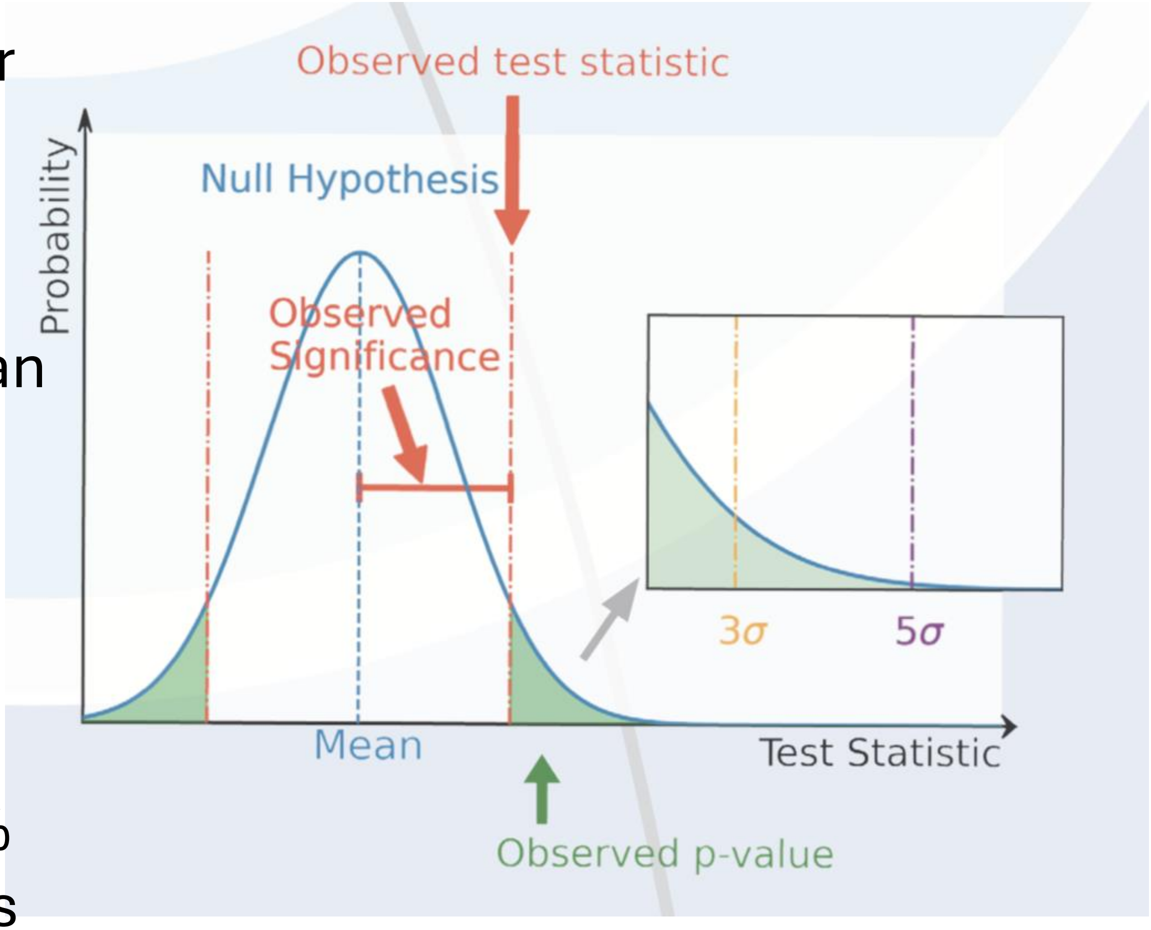


Statistical Significance

- Significance is often expressed as the number of "**standard deviations**" the observed test statistic is away from the null hypothesis.

- If the test statistic is distributed like a Gaussian (a bell shape) and the null hypothesis is true, then a randomly sampled test statistic will be within 1 standard deviation (1σ) of the mean 68% of the time.

- Therefore, for a 1σ significance, there is a 32% probability that the observed test statistic was obtained outside of this range by **random chance**, despite the null hypothesis being true.



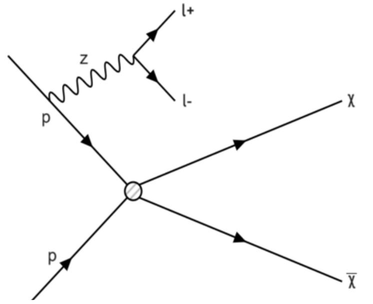
Search for an invisibly decaying Higgs boson or dark matter candidates produced in association with a Z boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

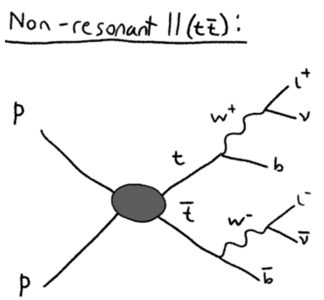
The ATLAS Collaboration

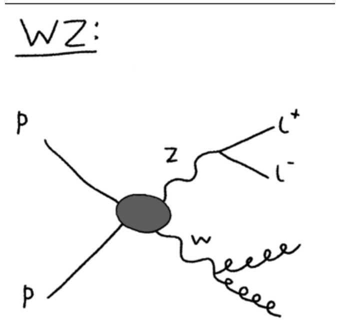
A search for an invisibly decaying Higgs boson or dark matter candidates produced in association with a leptonically decaying Z boson in proton–proton collisions at $\sqrt{s} = 13$ TeV is presented. This search uses 36.1 fb^{-1} of data collected by the ATLAS experiment at the Large Hadron Collider. No significant deviation from the expectation of the Standard Model backgrounds is observed. Assuming the Standard Model ZH production cross-section, an observed (expected) upper limit of 67% (39%) at the 95% confidence level is set on the branching ratio of invisible decays of the Higgs boson with mass $m_H = 125$ GeV. The corresponding limits on the production cross-section of the ZH process with the invisible Higgs boson decays are also presented. Furthermore, exclusion limits on the dark matter candidate and mediator masses are reported in the framework of simplified dark matter models.

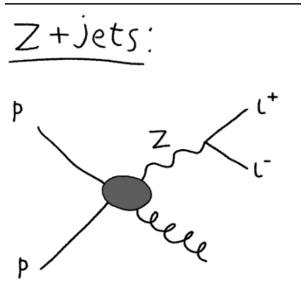
Search for Dark Matter

	totalWeight	sum_lep_charge	lead_lep_pt	sublead_lep_pt	mll	ETmiss	dRll	dphi_pTll_ETmiss	fractional_pT_difference	ETmiss_over_HT	N_bjets
0	0.041298	0	113.229602	37.342027	91.874195	124.311867	1.568130	3.036052	0.047148	0.599380	0
1	0.042212	0	89.615922	31.122283	87.832052	105.491891	1.697633	3.083360	0.032974	0.873724	0
2	0.061651	0	112.169008	65.336797	87.876299	170.239734	1.030940	3.133160	0.028335	0.959066	0
3	0.067919	0	108.710273	93.620266	89.113704	221.266453	0.872517	3.052979	0.005579	0.940918	0
4	0.085524	0	85.937289	40.265805	88.230021	132.649672	1.573029	2.554065	0.040406	0.613755	0









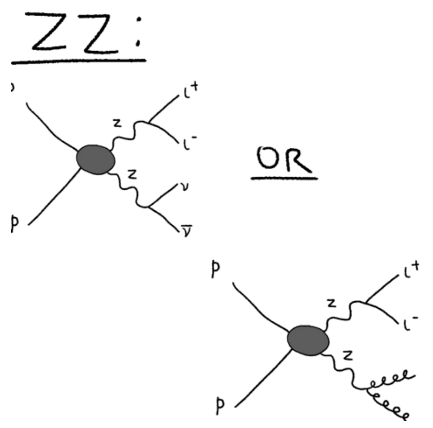


Table 1: Event selection criteria in the $\ell\ell + E_{\mathrm{T}}^{\mathrm{miss}}$ search.

	Selection criteria
Two leptons	Two opposite-sign leptons, leading (subleading) $p_{\mathrm{T}} > 30$ (20) GeV
Third lepton veto	Veto events if any additional lepton with $p_{\mathrm{T}} > 7$ GeV
$m_{\ell\ell}$	$76 < m_{\ell\ell} < 106$ GeV
$E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss}}/H_{\mathrm{T}}$	$E_{\mathrm{T}}^{\mathrm{miss}} > 90$ GeV and $E_{\mathrm{T}}^{\mathrm{miss}}/H_{\mathrm{T}} > 0.6$
$\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\ell\ell}, \vec{E}_{\mathrm{T}}^{\mathrm{miss}})$	$\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\ell\ell}, \vec{E}_{\mathrm{T}}^{\mathrm{miss}}) > 2.7$ radians
$\Delta R_{\ell\ell}$	$\Delta R_{\ell\ell} < 1.8$
Fractional p_{T} difference	$\left\| p_{\mathrm{T}}^{\ell\ell} - p_{\mathrm{T}}^{\mathrm{miss,jets}} \right\| / p_{\mathrm{T}}^{\ell\ell} < 0.2$
b-jets veto	$N(b\text{-jets}) = 0$ with b-jet $p_{\mathrm{T}} > 20$ GeV and $\|\eta\| < 2.5$

You are going to choose and optimise these cuts!