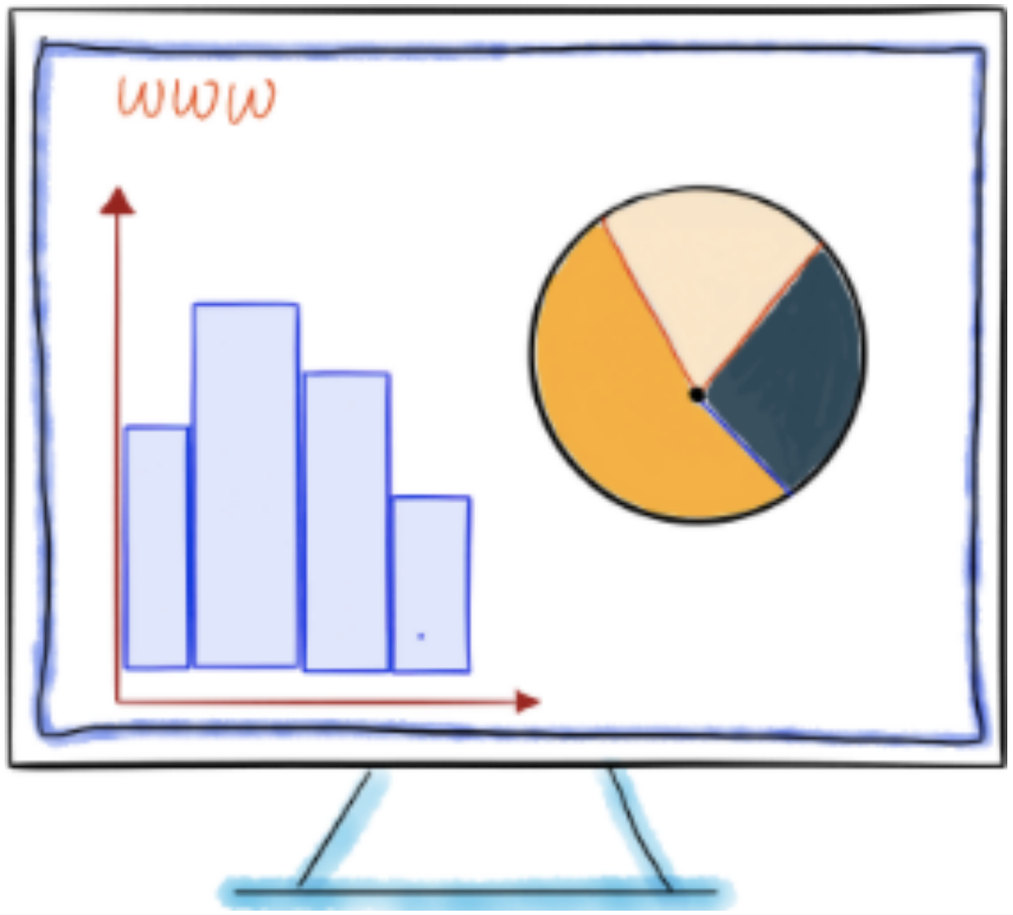
ATLAS Open Data: Particle physics for everyone

Meirin Oan Evans

Nuclear Physics Society, April 25th 2018



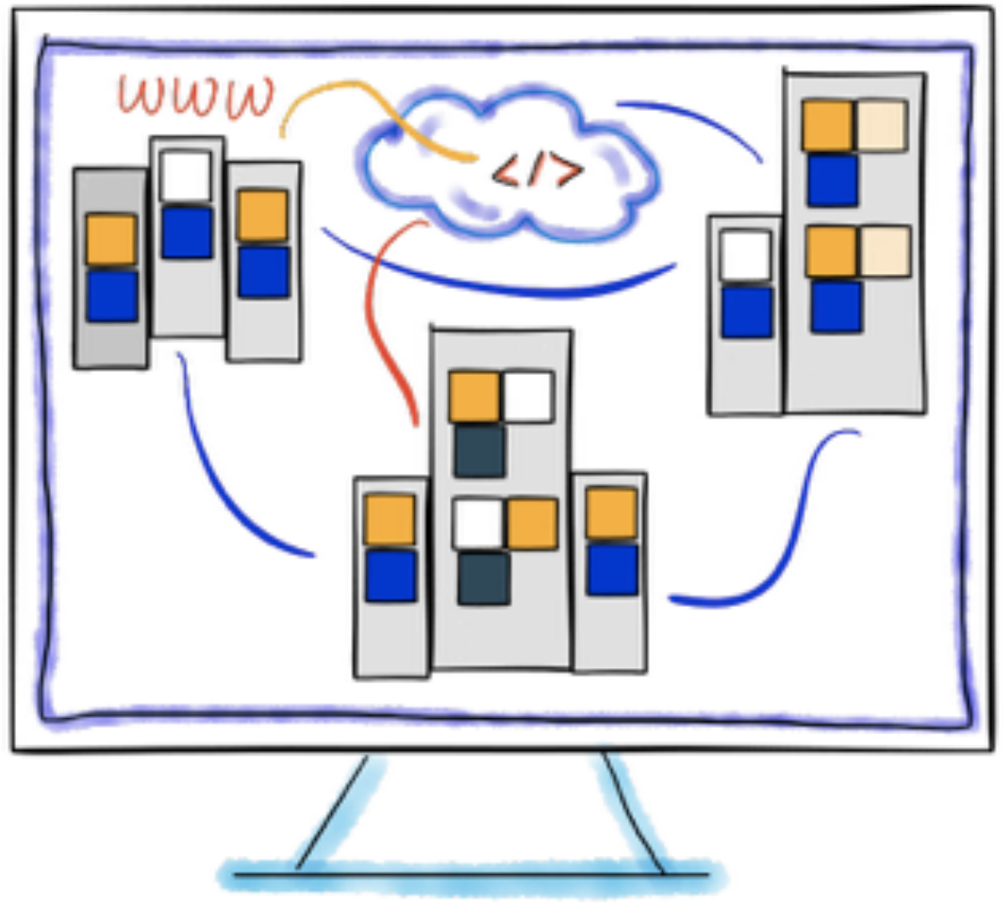
The University of Manchester

About me

- Final year MPhys Physics student
- 2 project days a week
- Since end of September

Plan

- What is ATLAS Open Data?
- How is this particle physics for everyone?
- Discuss some particle physics!
- Questions throughout!

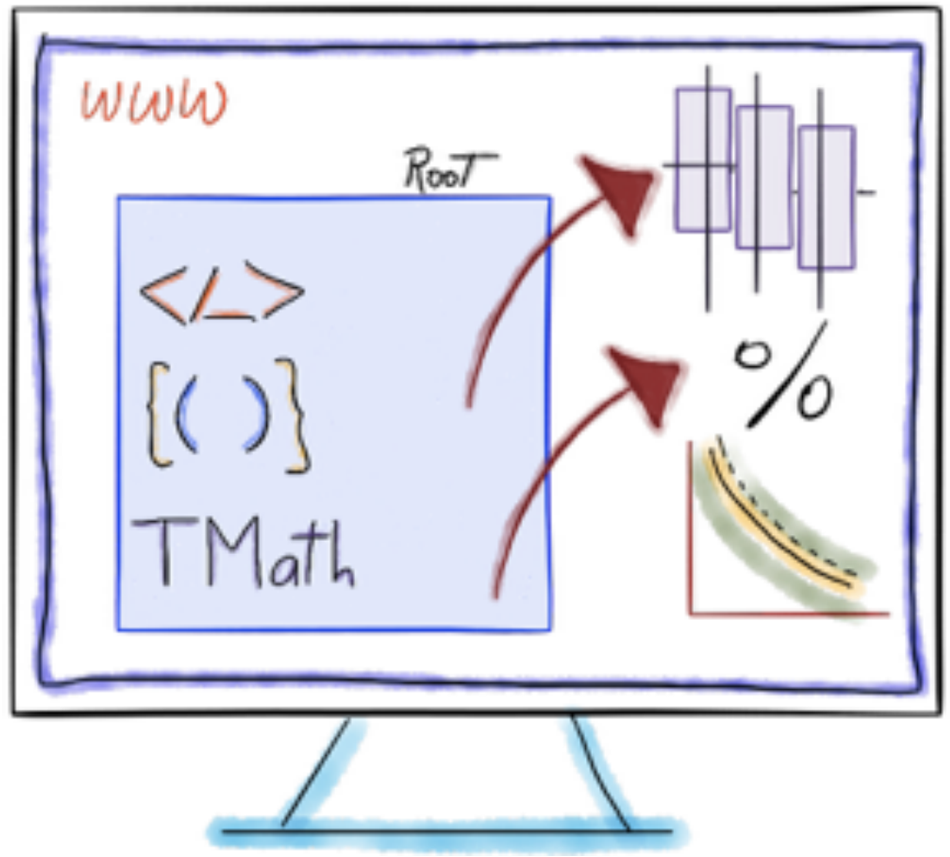


ATLAS


LHCb
ATLAS
CERN Meyrin
CERN Prévessin
SPS 7 km
SUISSE
FRANCE
ALICE
CMS
LHC 27 km

Open Data

- Definition: Public release of data
- ATLAS:
  - visualise data
  - download & use data
  - provide open-source software for you to make your own discoveries

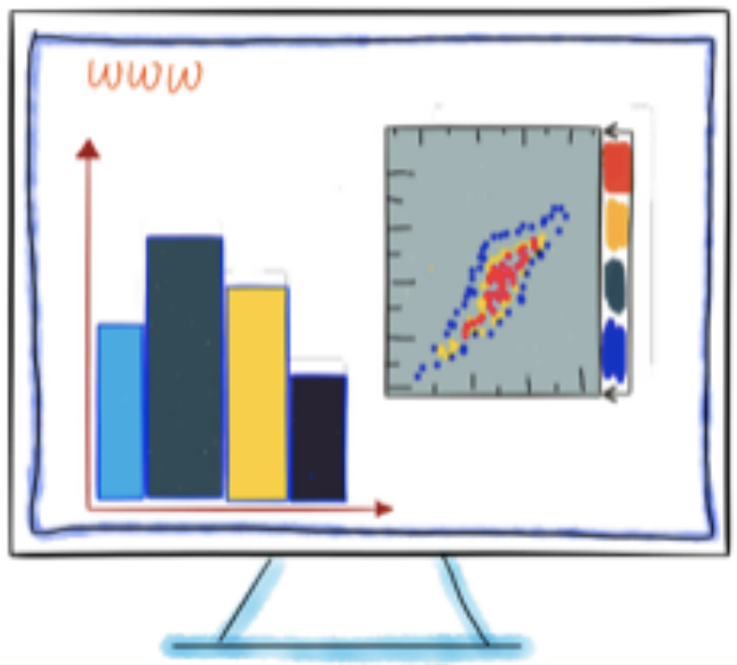


What & Who?

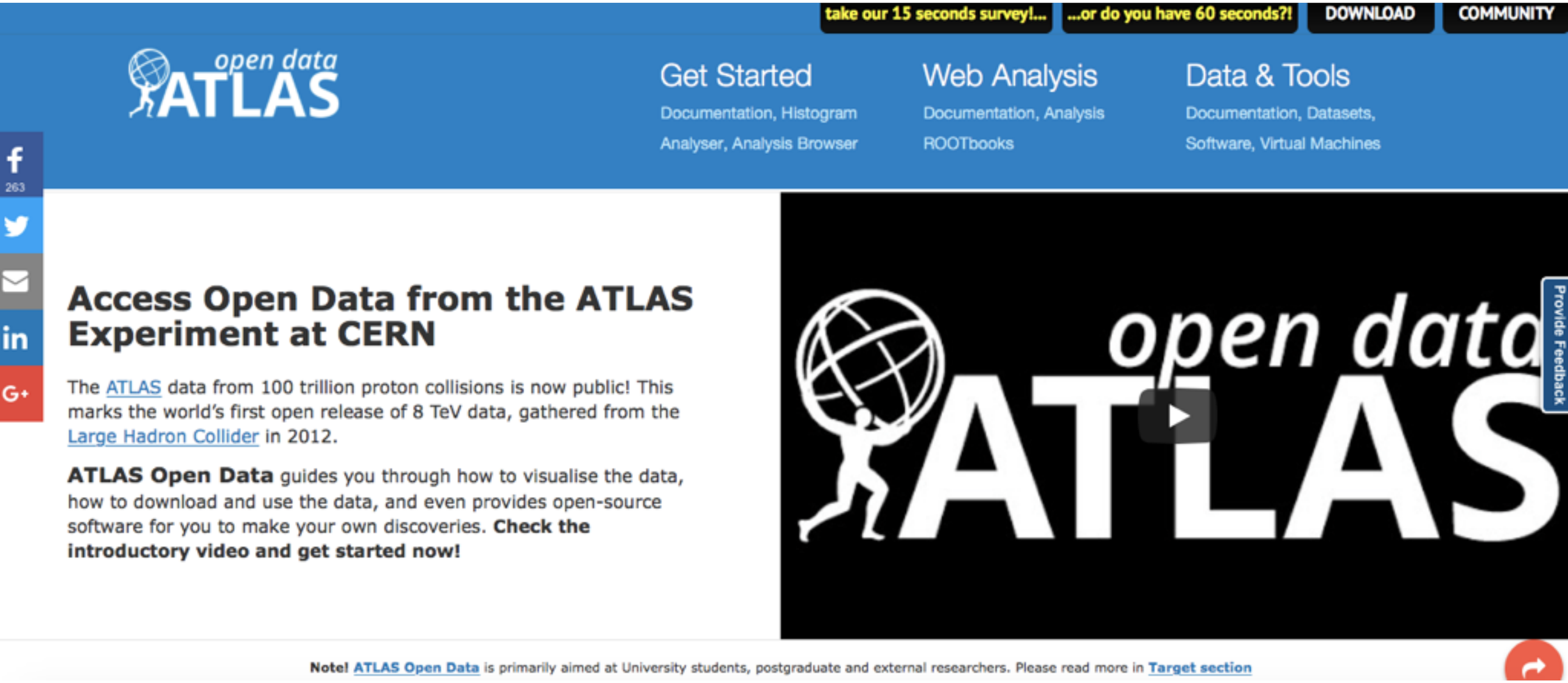
- **What is ATLAS Open Data about?**
- Making ATLAS data public (also data preservation)
- **What & who should Open Data be used for?**
  - Primarily aimed at university students, postgraduates, external researchers
  - Students can use in teaching labs & further research work
  - External researchers can use as training for new analyzers on data analysis techniques, especially outside ATLAS scientific community
  - Can use with general public for outreach & for curious people

8 TeV Open Data

- From proton-proton collisions in 2012
- Release includes simulations of Standard Model & Beyond Standard Model processes
- Use for education purposes



ATLAS Open Data



- opendata.atlas.cern website link

3rd year lab

- Intro to Particle Physics Data Analysis with ATLAS
- Aims: Give some appreciation of physics processes that occur in high energy proton-proton collisions at LHC. Introduce you to event selection & measurement methods used in particle physics data analysis

- Lab uses 8 TeV data from 2012
- I'm preparing 13 TeV data from 2015

Kaggle

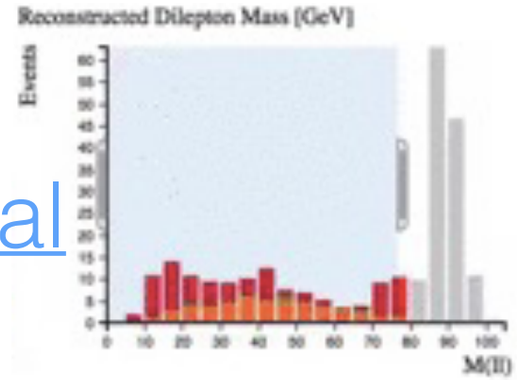
- Kaggle Higgs Boson Machine Learning Challenge
- Unanticipated use of 8 TeV Open Data
- Data used to study & develop data analysis techniques in wider scientific community
- “Use the ATLAS experiment to identify the Higgs Boson”

$13000 of prize money

Higgs challenge

CERN + ATLAS articles

- Articles appearing on CERN public website give updates on ATLAS Open Data



- Explore LHC Data on new ATLAS educational platform soon after 8 TeV dataset release

- Dortmund students work with ATLAS data
- Longer articles on ATLAS website for more detail
- Find more articles at atlas.cern/tags/open-data

ATLAS Publications

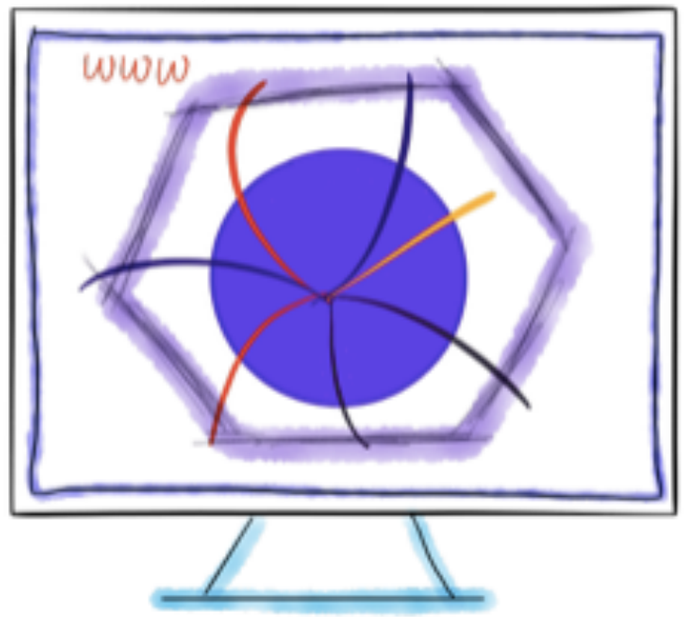
- On outreach for Open Data & associated resources

- '*Review of ATLAS Open Data Dataset'* (3rd Aug 16)
  - cds.cern.ch/record/2203649

- *'Review of ATLAS Open Data 8 TeV datasets, tools and activities'* - (after user feedback)

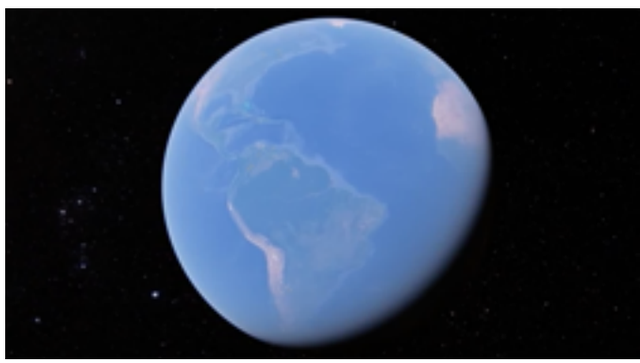
- To be fully approved by April 2018

- cds.cern.ch/record/2290911



Video link

- Produce several multimedia materials in collaboration with communications team



- To show the website & tutorials
- '*People using & creating ATLAS Open Data & Tools*'
- Documents in nice way the different groups using the resources worldwide
- From CERN to universities to conferences to high schools
- youtube.com/watch?v=pDS3SUeIRA8

Open Data Gitbooks

- Gitbook documentation now adopted in similar way by entire ATLAS collaboration

- opendata.atlas.cern/books/current/get-started/_book link to example of Gitbook produced for Open Data

- Constant contact & communication with other groups looking for public data, like ATLAS Machine Learning group

- Participated in 2016 & 17 ATLAS Machine Learning Workshops, plan to continue to be present there

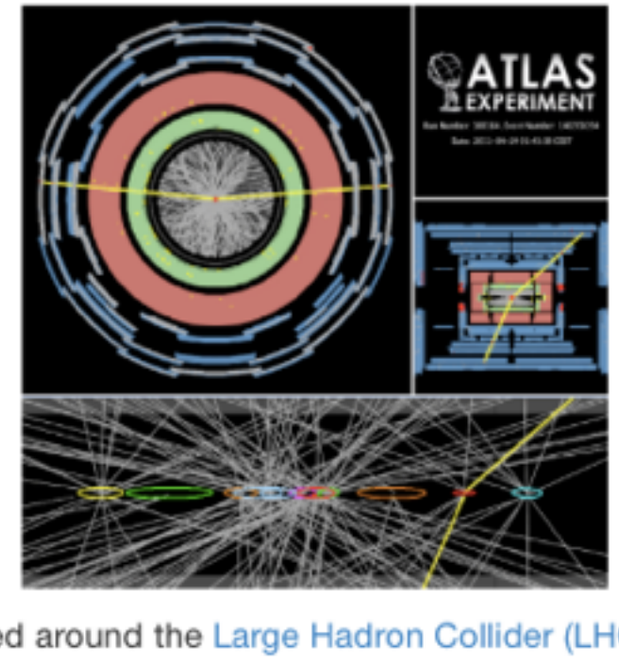
Gitbook

Type to search

- Get Started
- The Higgs Boson
- ATLAS events
- Analyses
- Data and Simulated Data
- Histogram Analyser
- Histogram Analyser 2
- Separate Signals
- Find the Higgs
- Analysis browser
- ROOTbrowser datasets
- ROOTbrowser Variable Names
- ROOTbrowser final plots
- Glossary
- Particle Physics Masterclasses

Published with GitBook

ATLAS events



Beams of protons are accelerated around the Large Hadron Collider (LHC) and are brought to collision at the centre of the ATLAS detector. The collisions produce debris in the form of new particles which fly out in all directions. Over a billion particle interactions take place in the ATLAS detector every second.

The protons within the two beams are grouped in bunches which are squeezed down in size to increase the chances of a collision. In the released data, the bunches crossed every 50 ns. There were about 30 collisions on average per **bunch-crossing**.

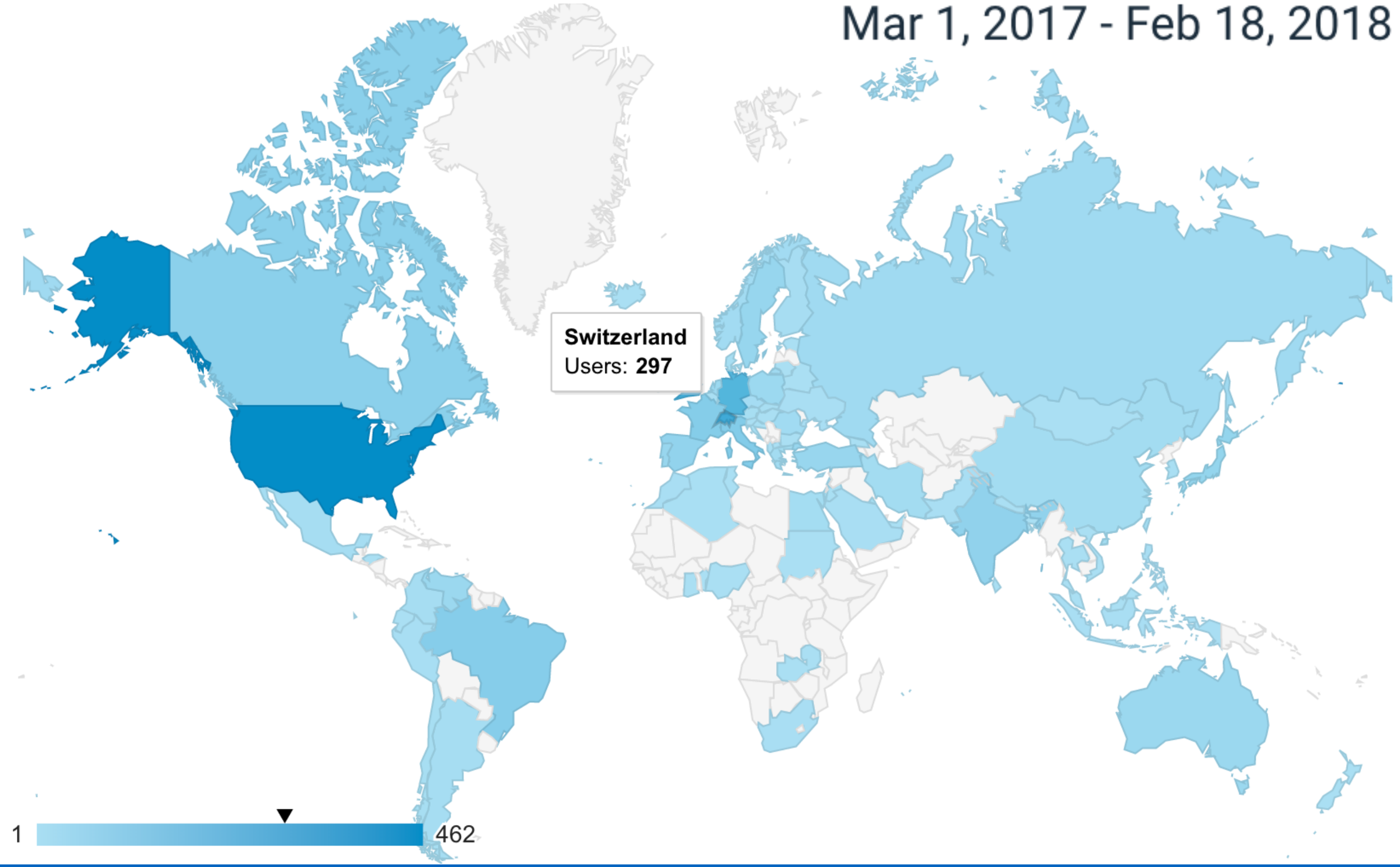
An **event** is the data resulting from a particular bunch-crossing.

Pile-up is defined as the average number of particle interactions per bunch-crossing. It is directly correlated with the instantaneous luminosity.

Luminosity is one of the most important parameters of the LHC. The higher the luminosity, the more data the experiments can gather to allow them to observe rare processes. However, increasing luminosity increases pile-up. This presents a challenge for physics analyses as it makes successfully identify collisions of interest harder.

Website use

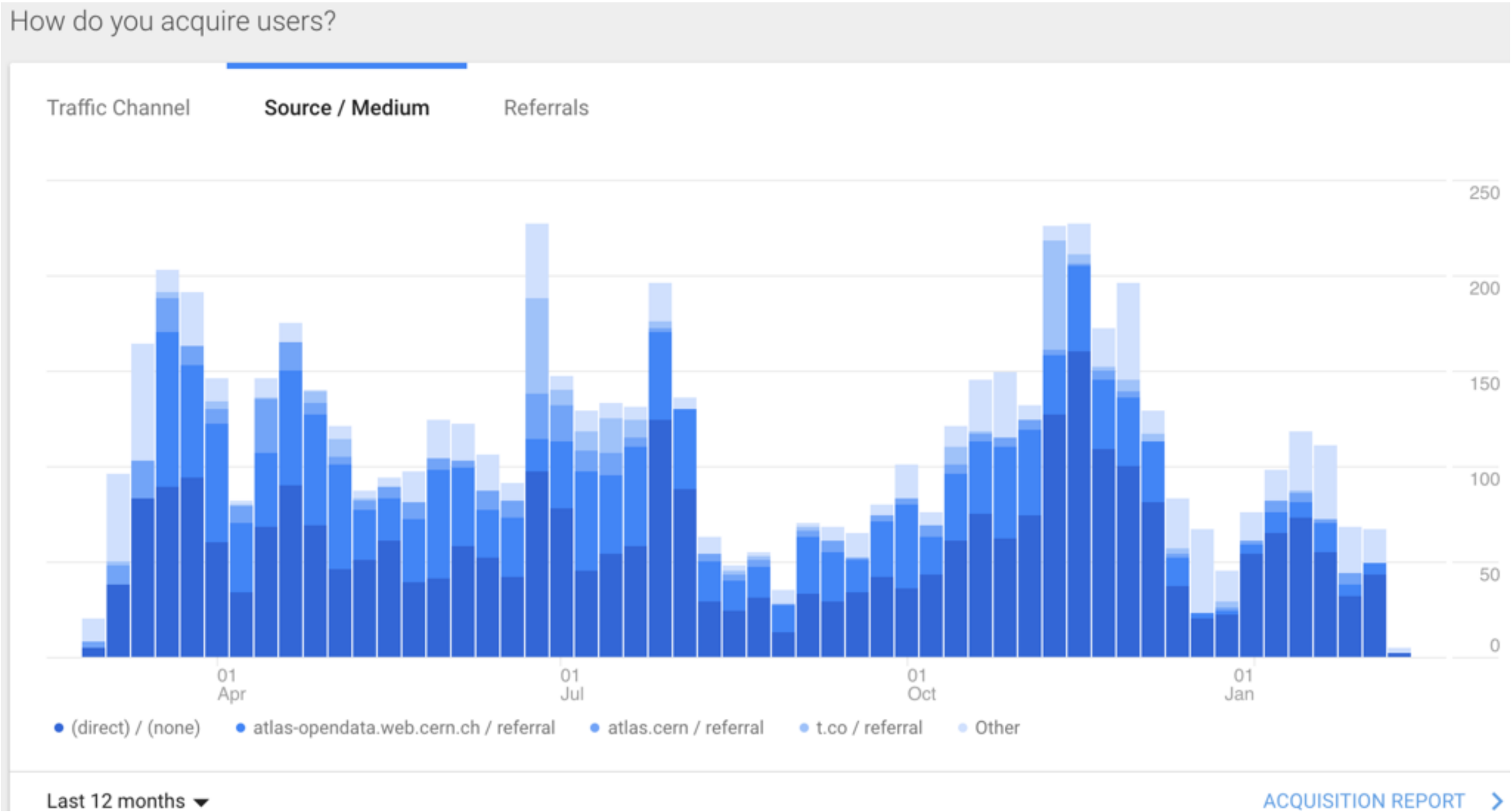
Mar 1, 2017 - Feb 18, 2018



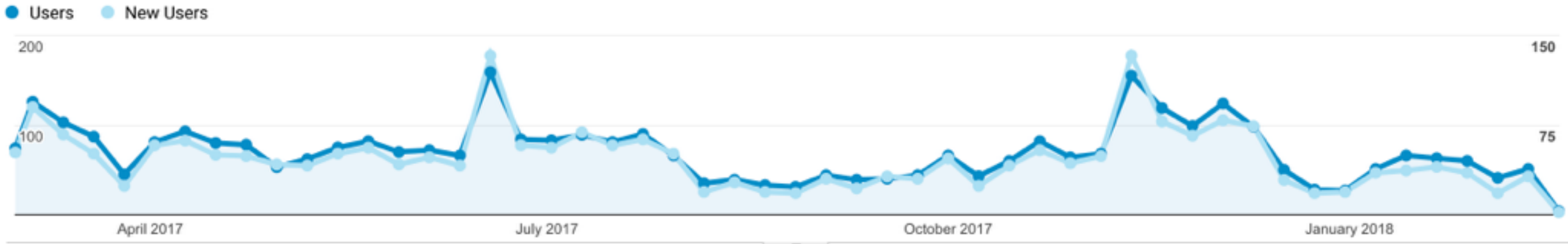
Website use

Country	Users	Sessions	Pages / Session
	2,534 % of Total: 100.00% (2,534)	6,035 % of Total: 100.00% (6,035)	3.24 Avg for View: 3.24 (0.00%)
1. United States	**462** (17.61%)	957 (15.86%)	3.20
2. Switzerland	**297** (11.32%)	771 (12.78%)	3.32
3. United Kingdom	**286** (10.90%)	548 (9.08%)	3.93
4. Germany	**238** (9.07%)	625 (10.36%)	3.45
5. Italy	**110** (4.19%)	289 (4.79%)	3.76
6. France	**101** (3.85%)	126 (2.09%)	3.32
7. Brazil	**100** (3.81%)	276 (4.57%)	3.58
8. Canada	**88** (3.35%)	271 (4.49%)	2.87
9. India	**72** (2.74%)	129 (2.14%)	3.71
10. Spain	**68** (2.59%)	150 (2.49%)	3.12

Website use


How do you acquire users?
Traffic Channel
Source / Medium
Referrals
250
200
150
100
50
0
01 Apr
01 Jul
01 Oct
01 Jan
(direct) / (none)
atlas-opendata.web.cern.ch / referral
atlas.cern / referral
t.co / referral
Other
Last 12 months
ACQUISITION REPORT

Website use



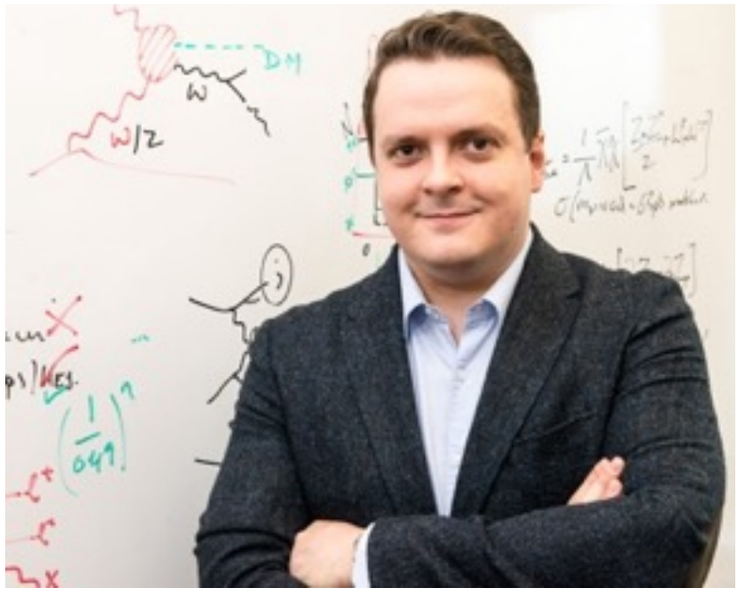
Users	New Users	Sessions	Number of Sessions per User
2,534	2,460	6,035	2.38

Pageviews	Pages / Session	Avg. Session Duration	Bounce Rate
19,562	3.24	00:04:15	43.96%

Enabling Open Science with the ATLAS Open Data Project at CERN

- Full year
- CE: Computational experimental
- Collaborate with colleagues in ATLAS ‘Outreach Data & Tools’ group, contributing to 1st public release of ATLAS 13 TeV datasets for education & science research as part of the ATLAS Open Data project: opendata.atlas.cern

Support



- Darren Price - lecturer at Manchester

- Agni Bethani - postdoc at Manchester
- Arturo Sanchez - at CERN, no link to Manchester

13 TeV

- First public release of ATLAS 13 TeV datasets including real & simulated data

- Want to go further than 8 TeV dataset release, to build on its success

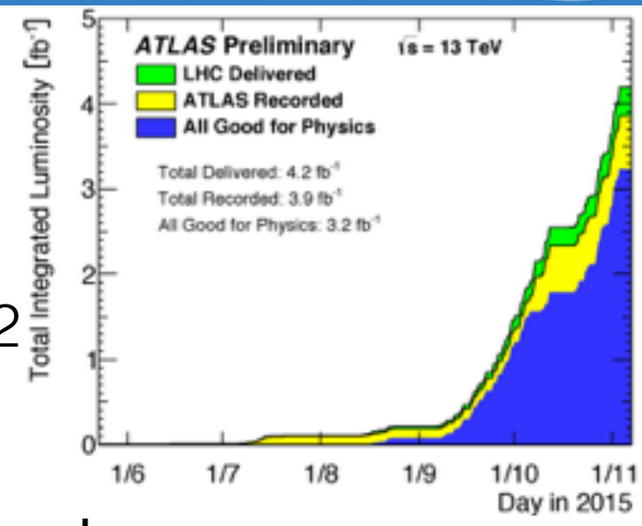
- Newly-collected ATLAS data from 13 TeV collisions

- Larger and more complete ATLAS 13 TeV datasets

13 TeV

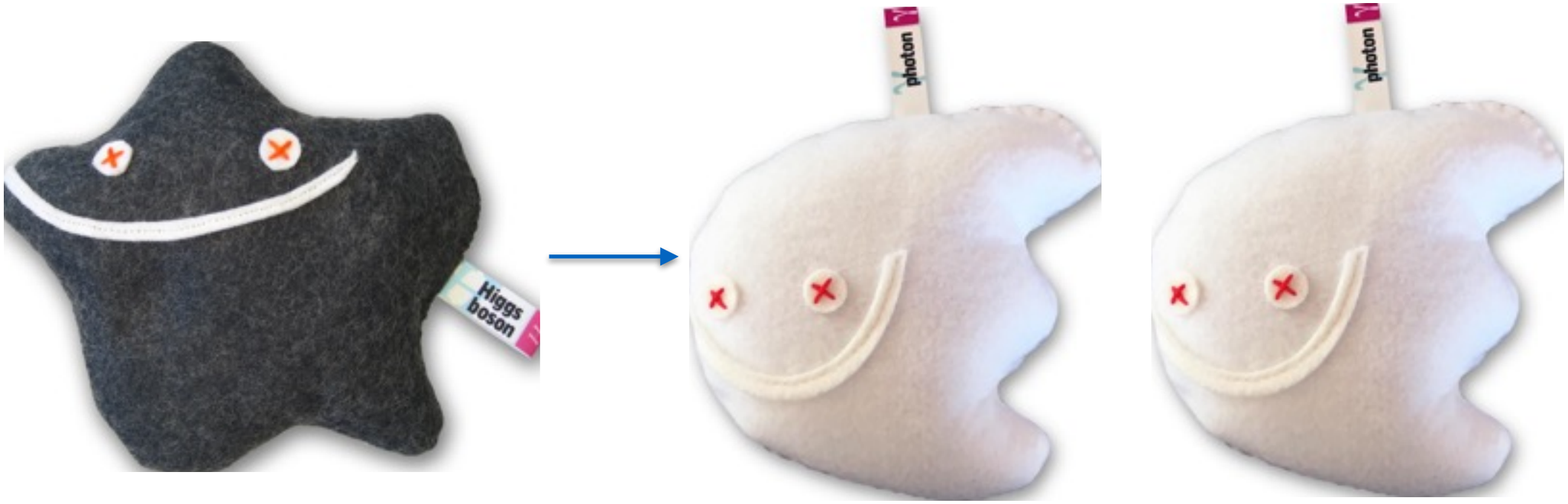
- Develop & promote dataset
- To be endorsed by ATLAS Collaboration
- Consist of measured & simulated data
- Needs to be flexible
- Will be made publicly available on Open Data portal
- Hope to release all 2015 data 3.2 fb^{-1}

Luminosity



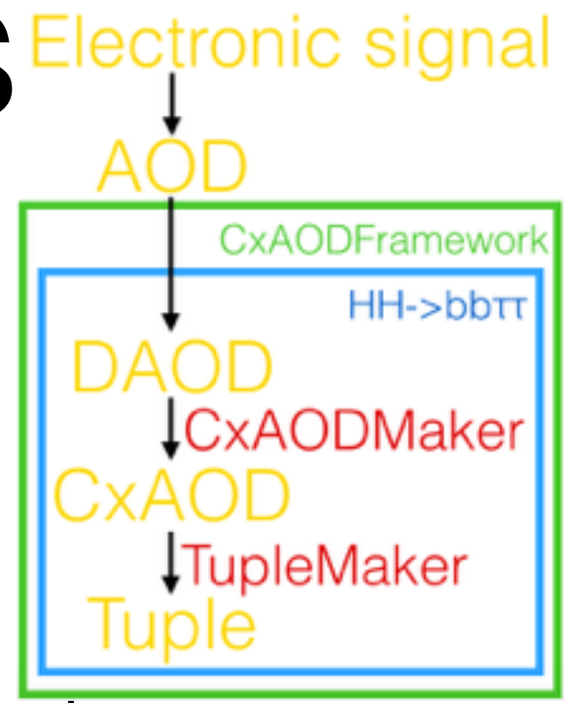
- 1 barn = 10^{-28} m^2, 1 femtobarn = 10^{-43} m^2
- Cross section (σ) for process measured in barn
- QM processes have different probabilities
- Integrated luminosity is basically amount of data
- Integrated luminosity ($\mathscr{L}$) measured in inverse barn
- Number of events for process = $\mathscr{L}\,\sigma$

13 TeV Plans



- Add *H->γγ* capability in data since this was the Higgs discovery channel!
- Include systematics to be able to quantify errors
- Include truth info to fully compare measured data with simulation

Project progress



- Design raw data -> simplified data workflow for 13 TeV
- Process 13 TeV data and Monte Carlo samples
- Validate pre-existing (8 TeV) selections in new data
- Add additional channels / selections as appropriate (other new processes open for discussion)

Tactile Collider

- twitter.com/tactilecollider
- Project led by Rob Appleby
- Wants to make 3D models similar to online Histogram Analyser
- For blind & partially sighted young people
- Could feel difference between signal & background

Programming

- Data production framework in **C++**
- User-facing analysis scripts in **Python**

- Data analysis framework used by CERN called **ROOT**, based on C++
- **Jupyter notebooks** for interactive code writing
- Data formats like **CSV, JSON, XML**

Extra talks

Outreach Data and Tools Group Meeting

Restricted

Wednesday 22 Nov 2017, 15:00 → 15:55 Europe/Zurich

Vidyo (CERN)

Arturos Sanchez Pineda (Universita e INFN, Napoli (IT)), Kate Shaw (INFN Gruppo Collegato di Udine and ICTP Trieste)

Description ATLAS outreach meeting relative to report and to work in the different Data & Tools projects currently carry on by the members of the team.

Videoconference Rooms Outreach_Data_and_Tools_Group — Join

15:15 → 15:35 Updates in developing of new ntuples for 13 TeV release

Speakers: Darren Price (University of Manchester (GB)), Meirin Oan Evans (University of Manchester (GB))

ATLAS Week Outreach Parallel Session

Tuesday 20 Feb 2018, 16:00 → 20:35 Europe/Zurich

40-S2-C01 - Salle Curie (CERN)

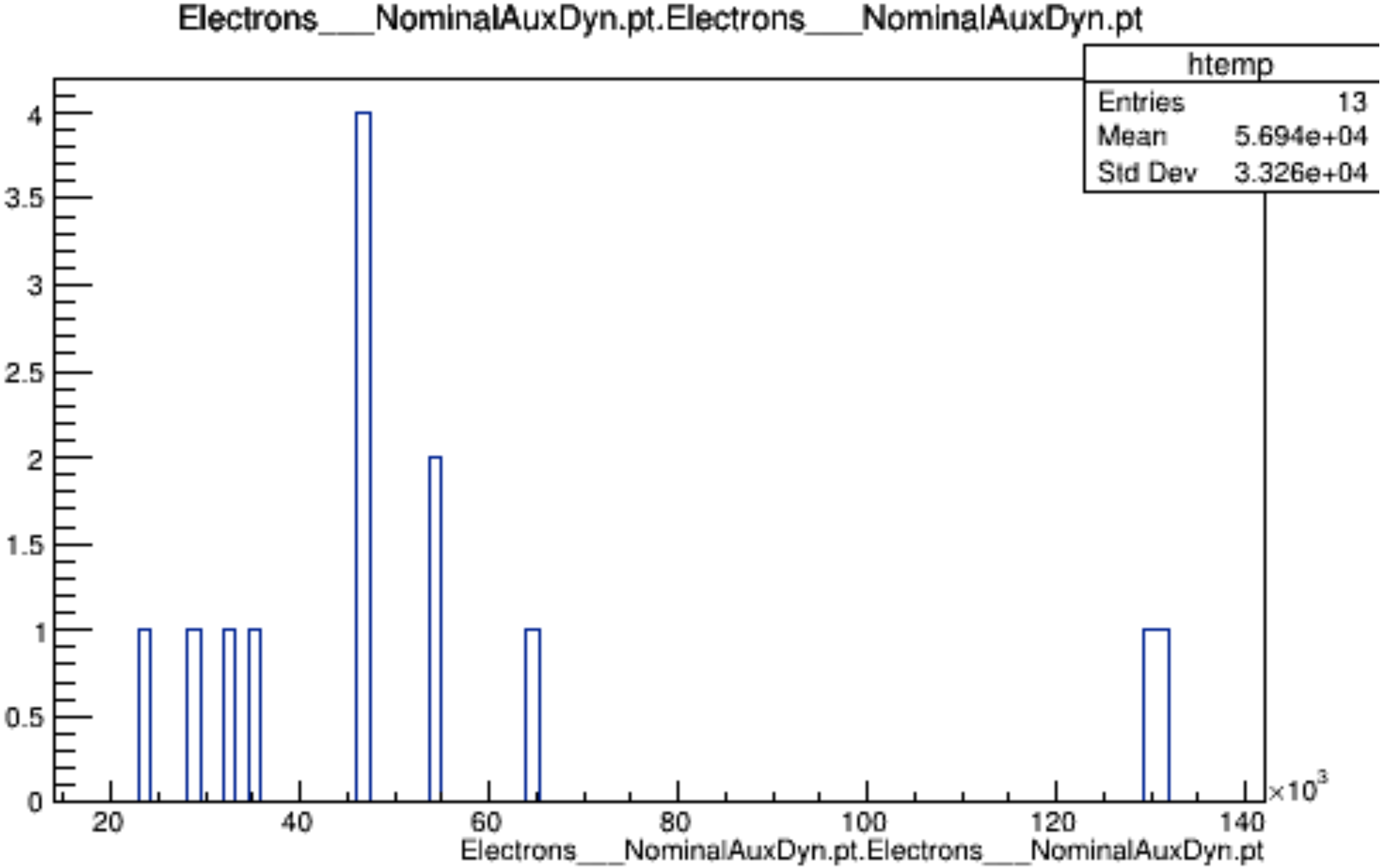
Kate Shaw (INFN Gruppo Collegato di Udine and ICTP Trieste), Sascha Mehlhase (Ludwig-Maximilians-Univ. Muenchen (DE))

Videoconference Rooms ATLAS_Week_Outreach_Parallel_Session

16:45 → 17:00 ATLAS Open Data

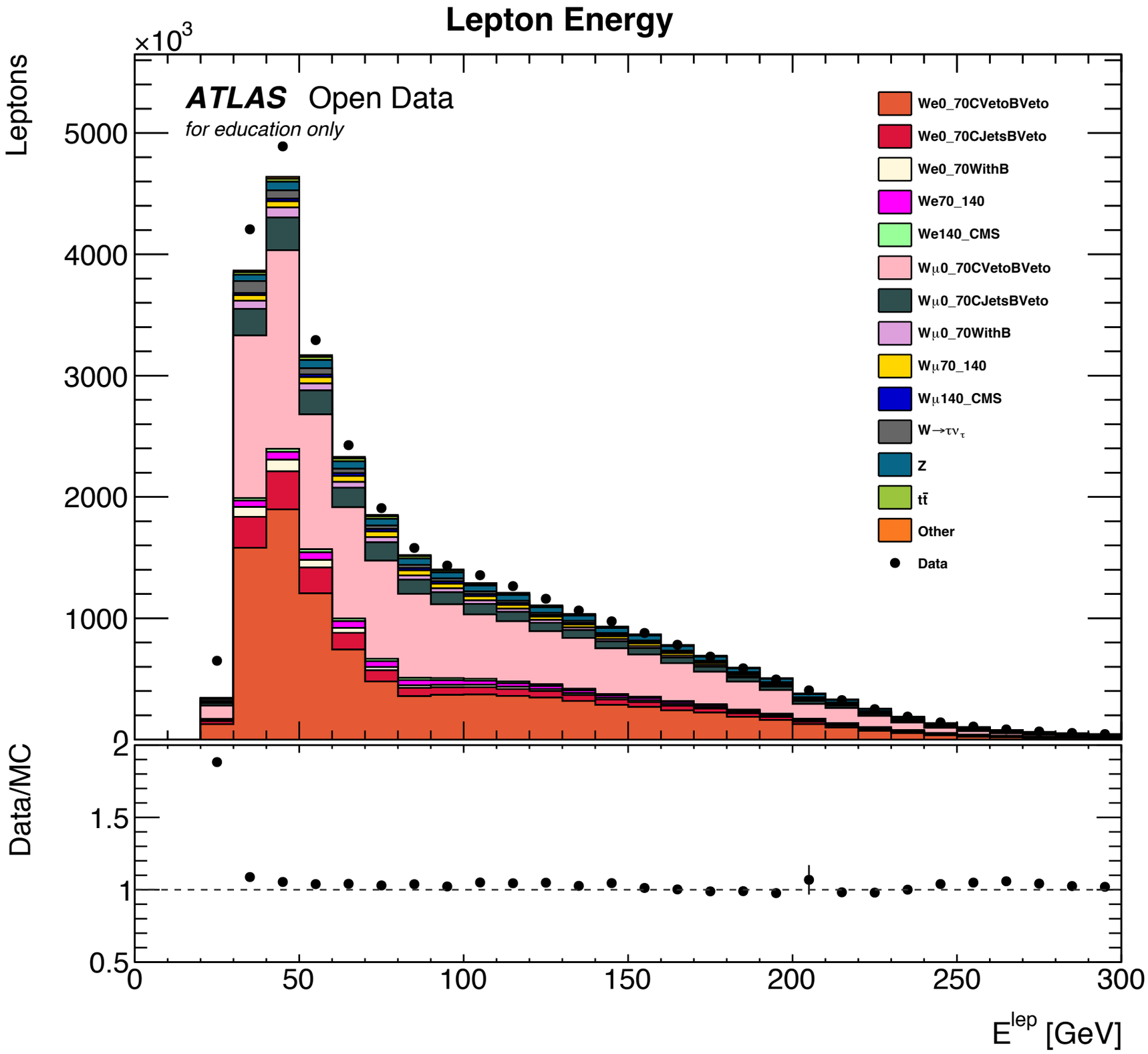
Speaker: Meirin Oan Evans (University of Manchester (GB))

First plot


Electrons___NominalAuxDyn.pt.Electrons___NominalAuxDyn.pt
htemp
Entries 13
Mean 5.694e+04
Std Dev 3.326e+04
4
3.5
3
2.5
2
1.5
1
0.5
0
20
40
60
80
100
120
140
×10^3
Electrons___NominalAuxDyn.pt.Electrons___NominalAuxDyn.pt

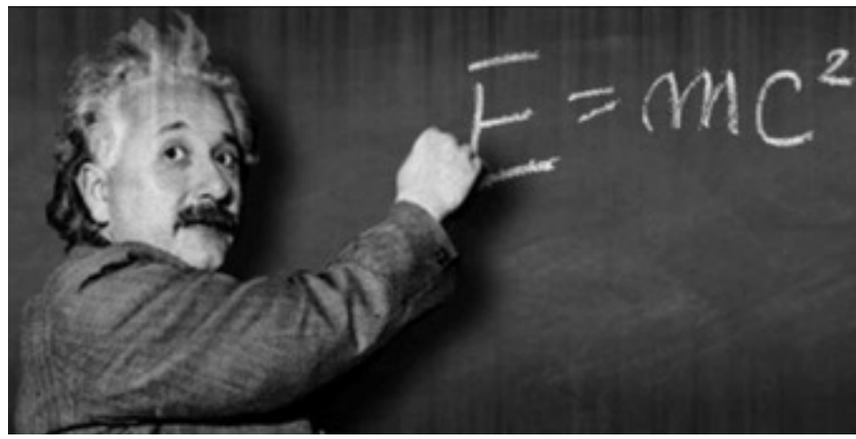
Lepton energy

Data 2015



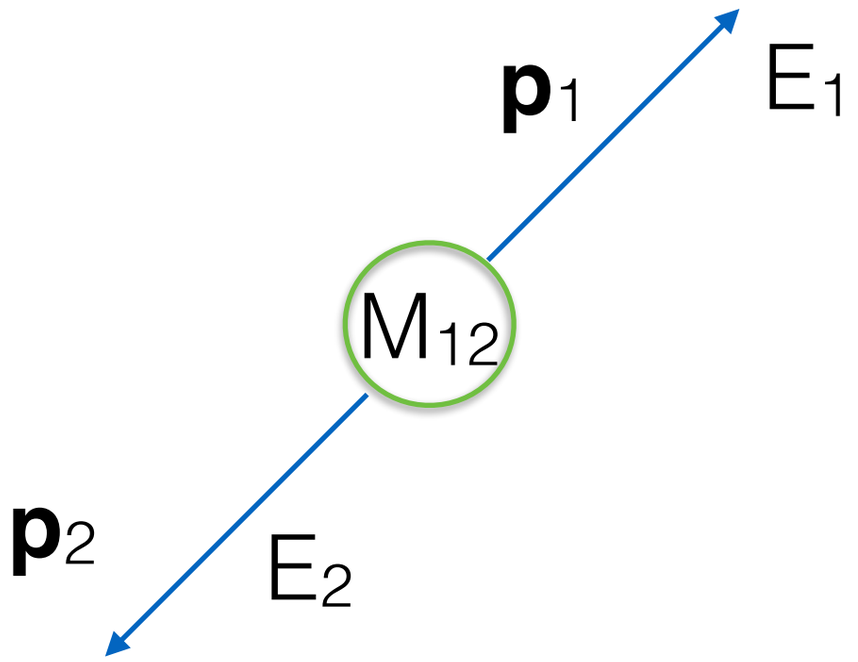
Invariant mass

$$E^2 = (pc)^2 + (Mc^2)^2$$

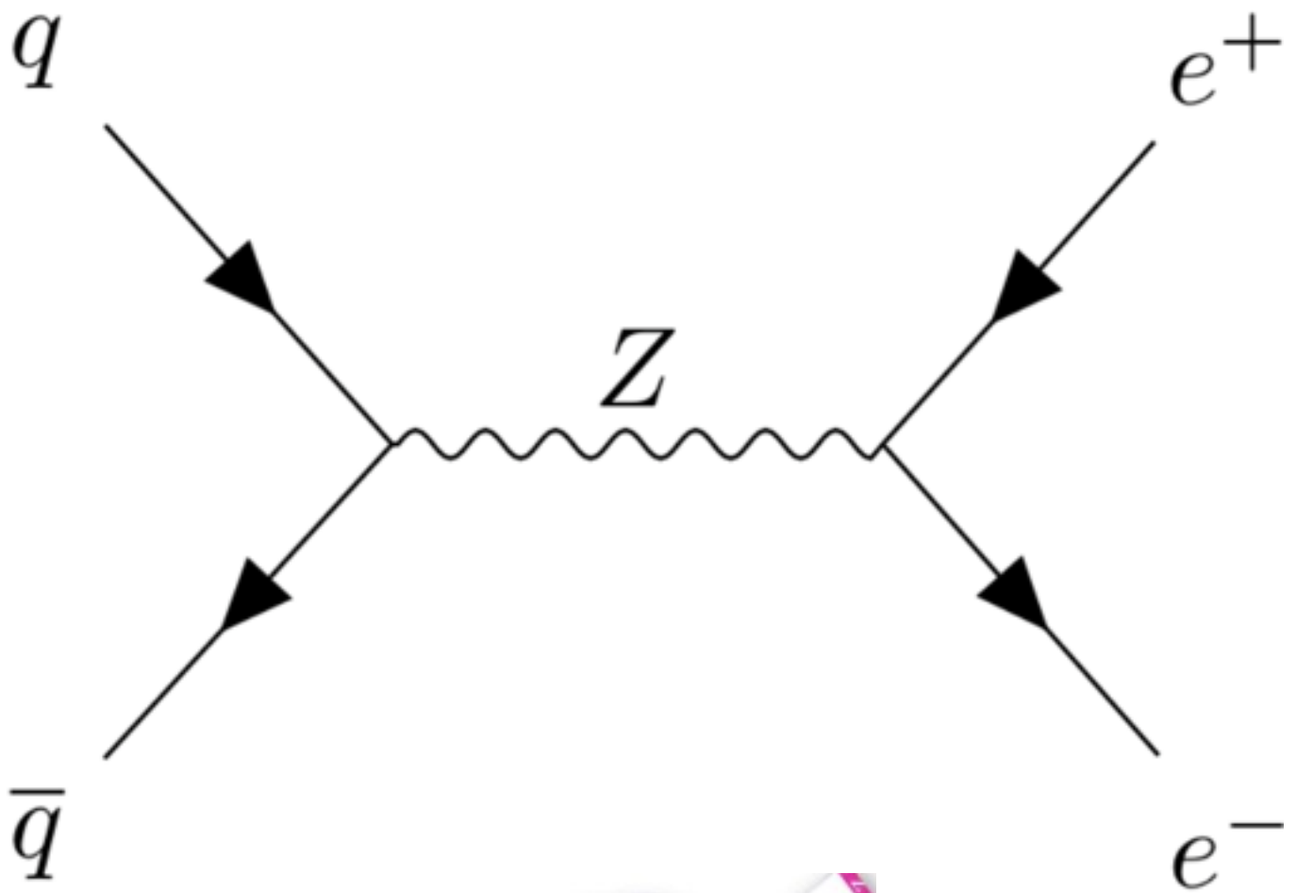


$$M = \sqrt{E^2 - p^2}$$

$$M_{12} = \sqrt{(E_1 + E_2)^2 - (\mathbf{p}_1 + \mathbf{p}_2)^2}$$

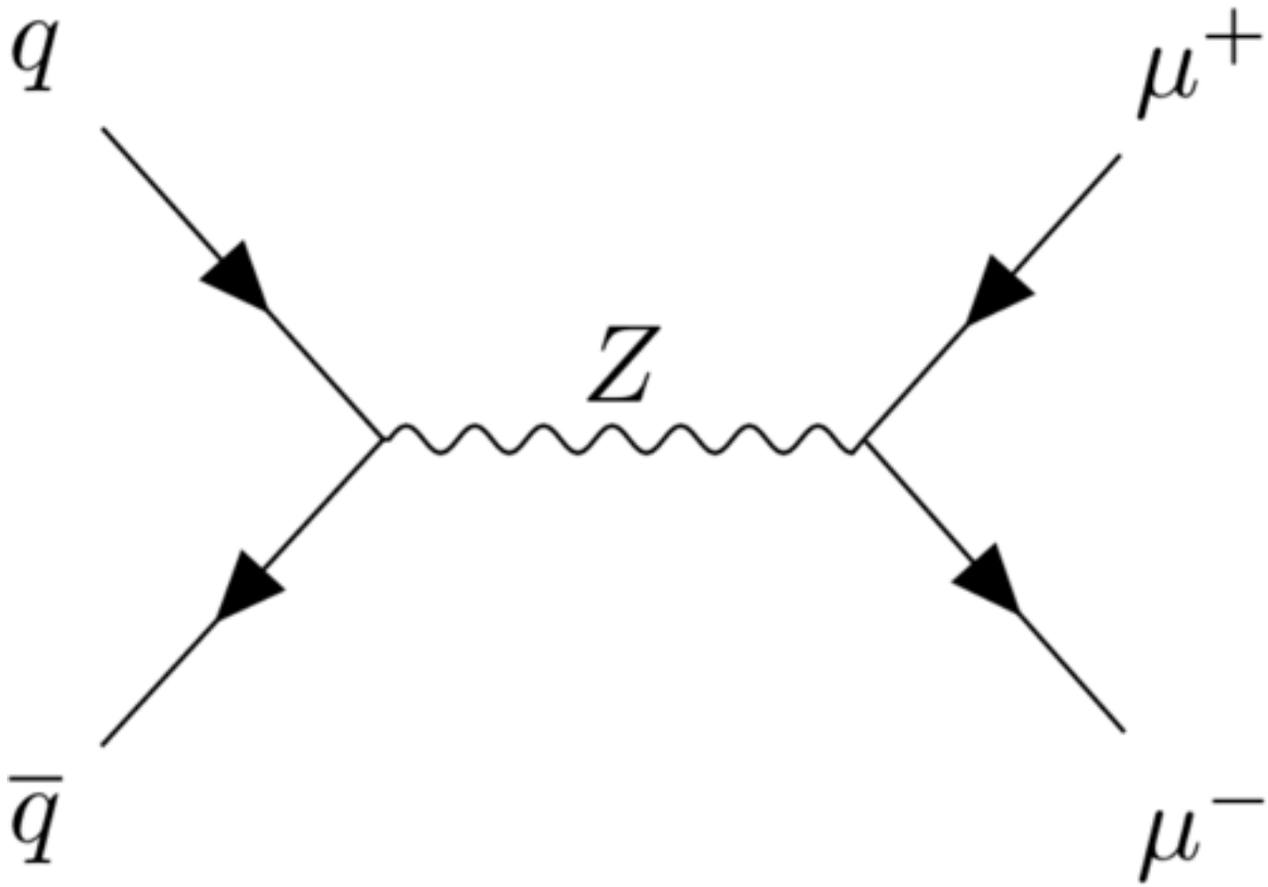


Z decay

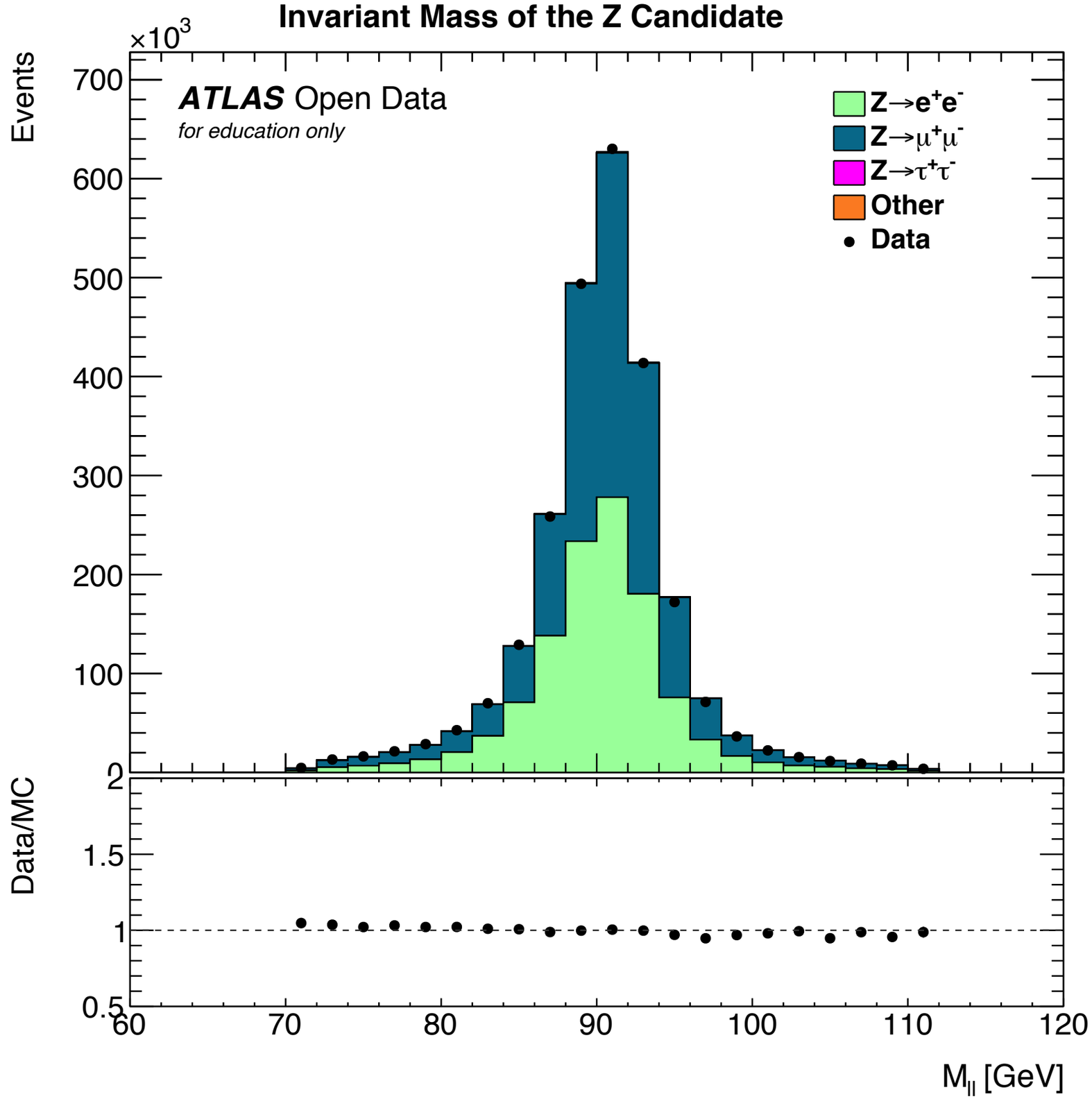


$$q : u, d, s, c, b, t$$

$$\overline{q} : \overline{u}, \overline{d}, \overline{s}, \overline{c}, \overline{b}, \overline{t}$$



Z Invariant mass

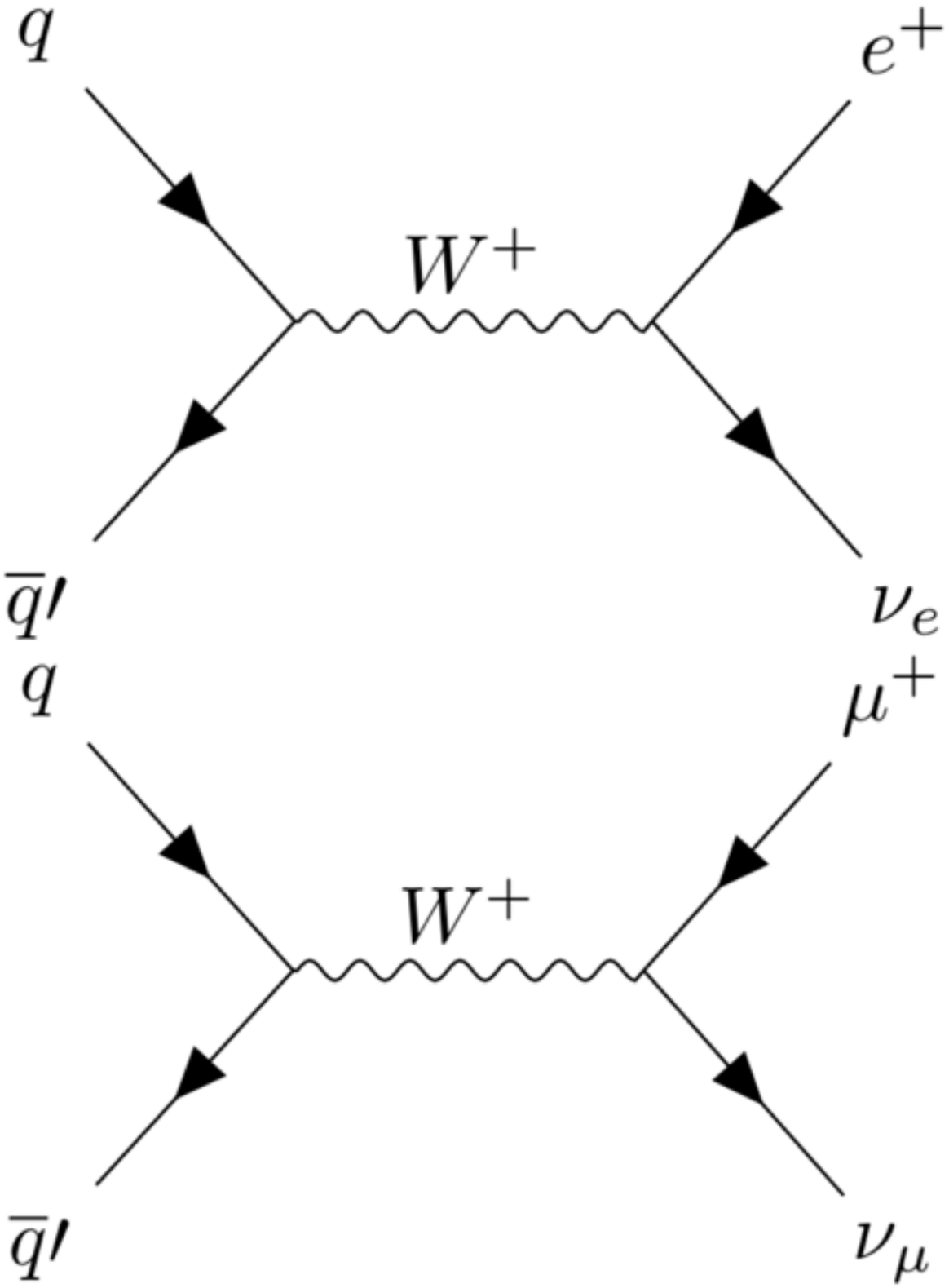


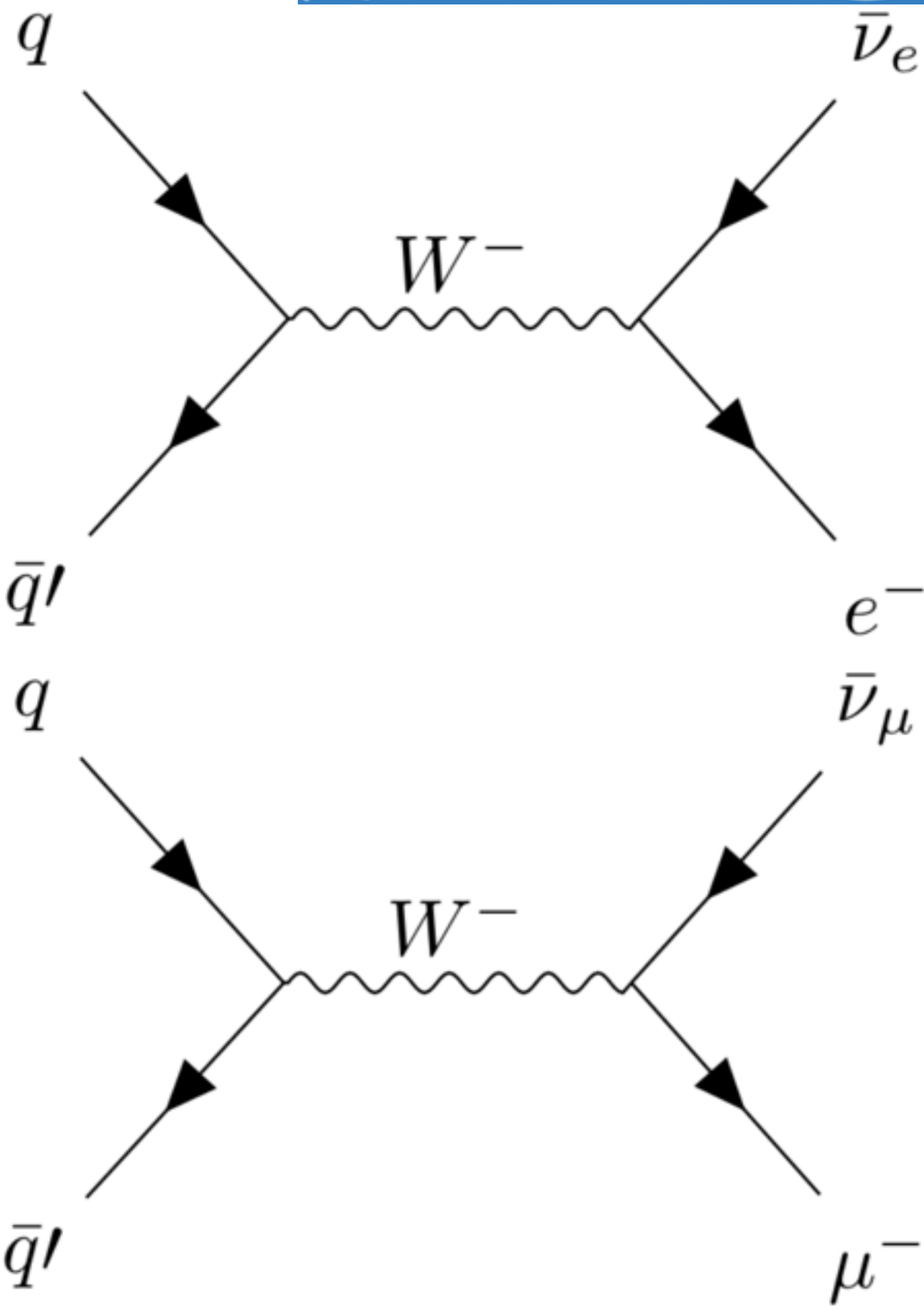
$Z \to e^+ e^-$

$Z \to \mu^+ \mu^-$

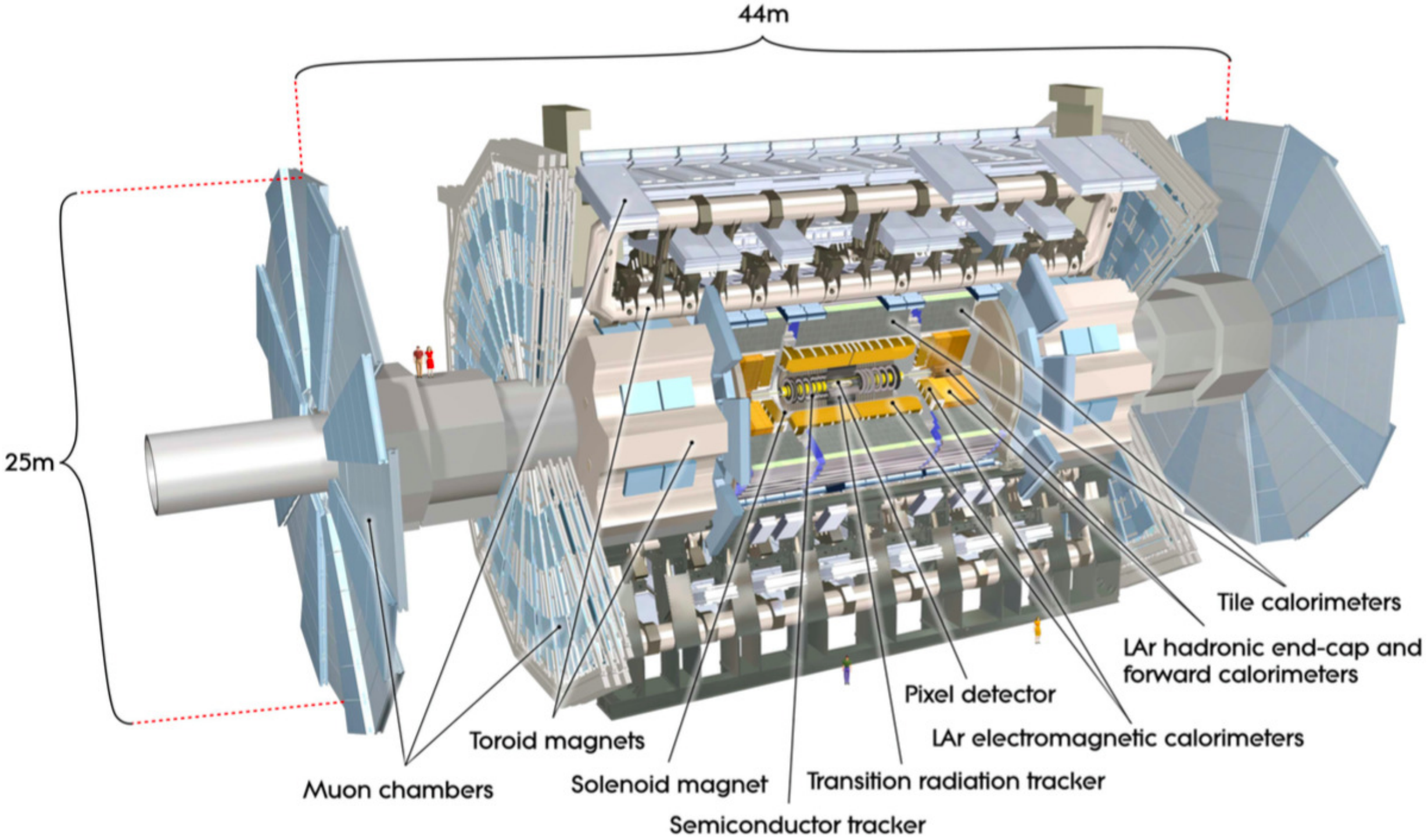
Data 2015

W decay


q
e^+
W^+
$\bar{q}'$
ν_e
q
μ^+
W^+
$\bar{q}'$
ν_μ

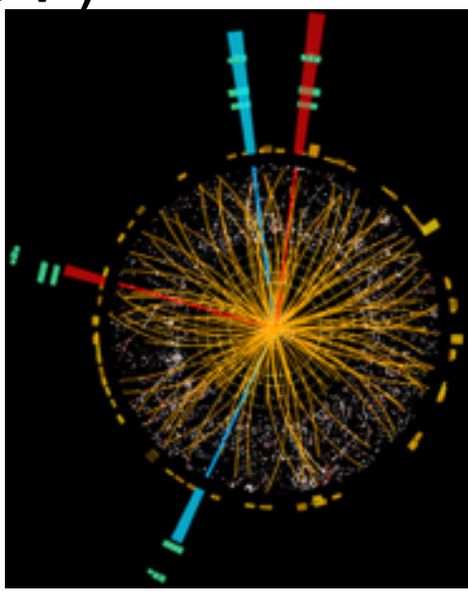

q
$\bar{\nu}_e$
W^-
$\bar{q}'$
e^-
q
$\bar{\nu}_\mu$
W^-
$\bar{q}'$
μ^-

ATLAS detector



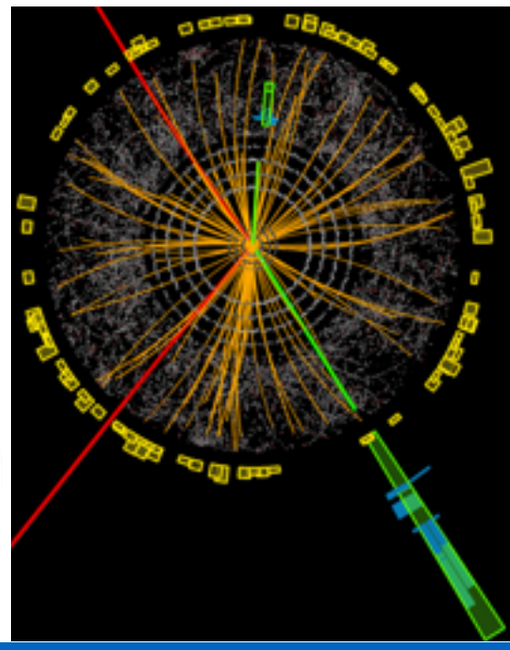
Transverse plane

- Know the momentum each proton has (6.5 TeV)
- But it's not the proton that collides
- It's the proton's constituents (quarks, gluons)
- Don't know the momentum each constituent has
- Initial constituents' momentum is along beam
- Initial constituents have no transverse momentum



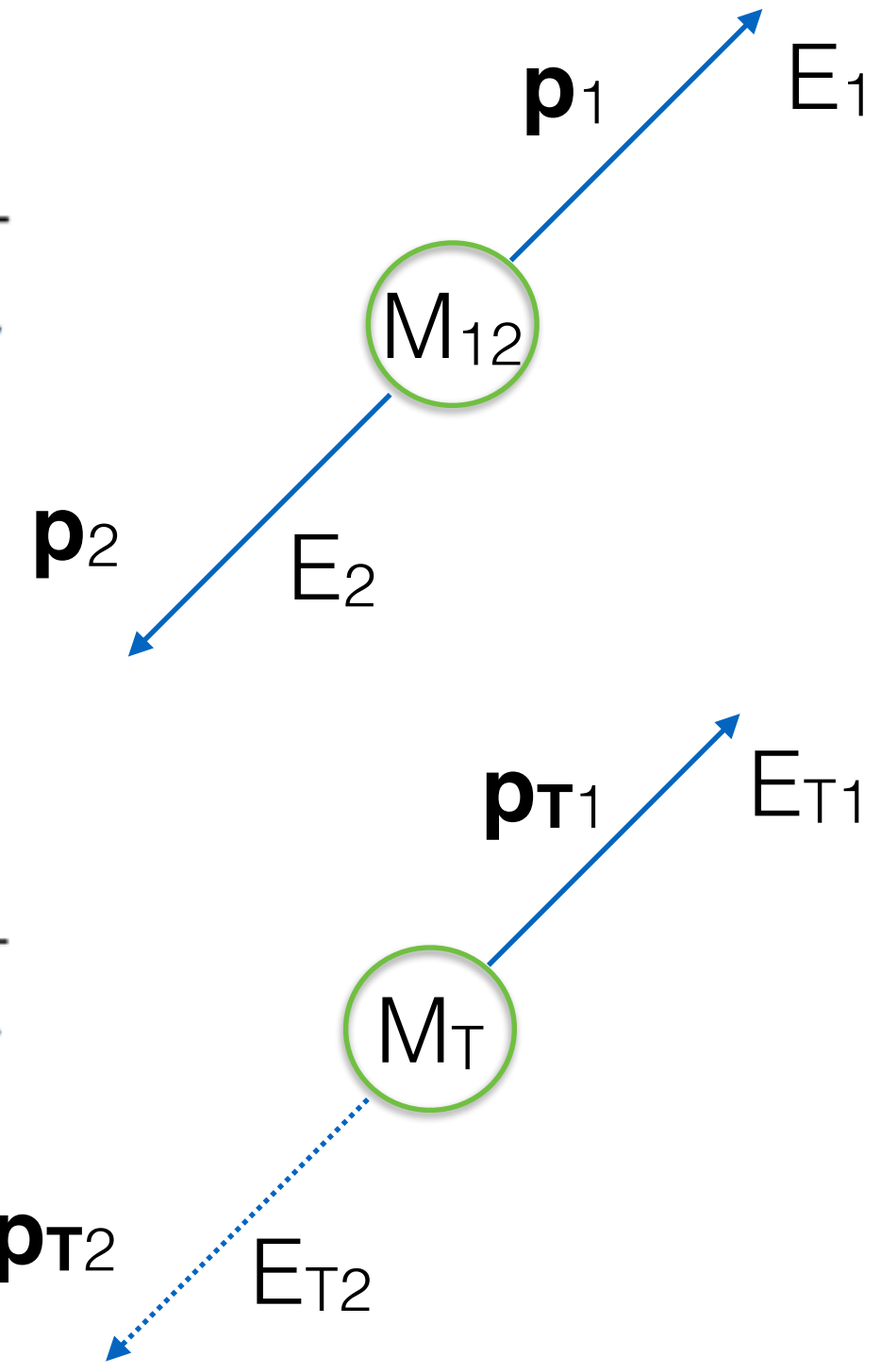
Missing momentum

- Measure transverse momentum p_T of final particles
- Use momentum conservation in transverse plane (transverse/normal/orthogonal to beam line)
- Can calculate Missing Transverse Momentum
- Longitudinal momentum p_L of final particles disappears down beam line
- Can't know Missing Longitudinal Momentum



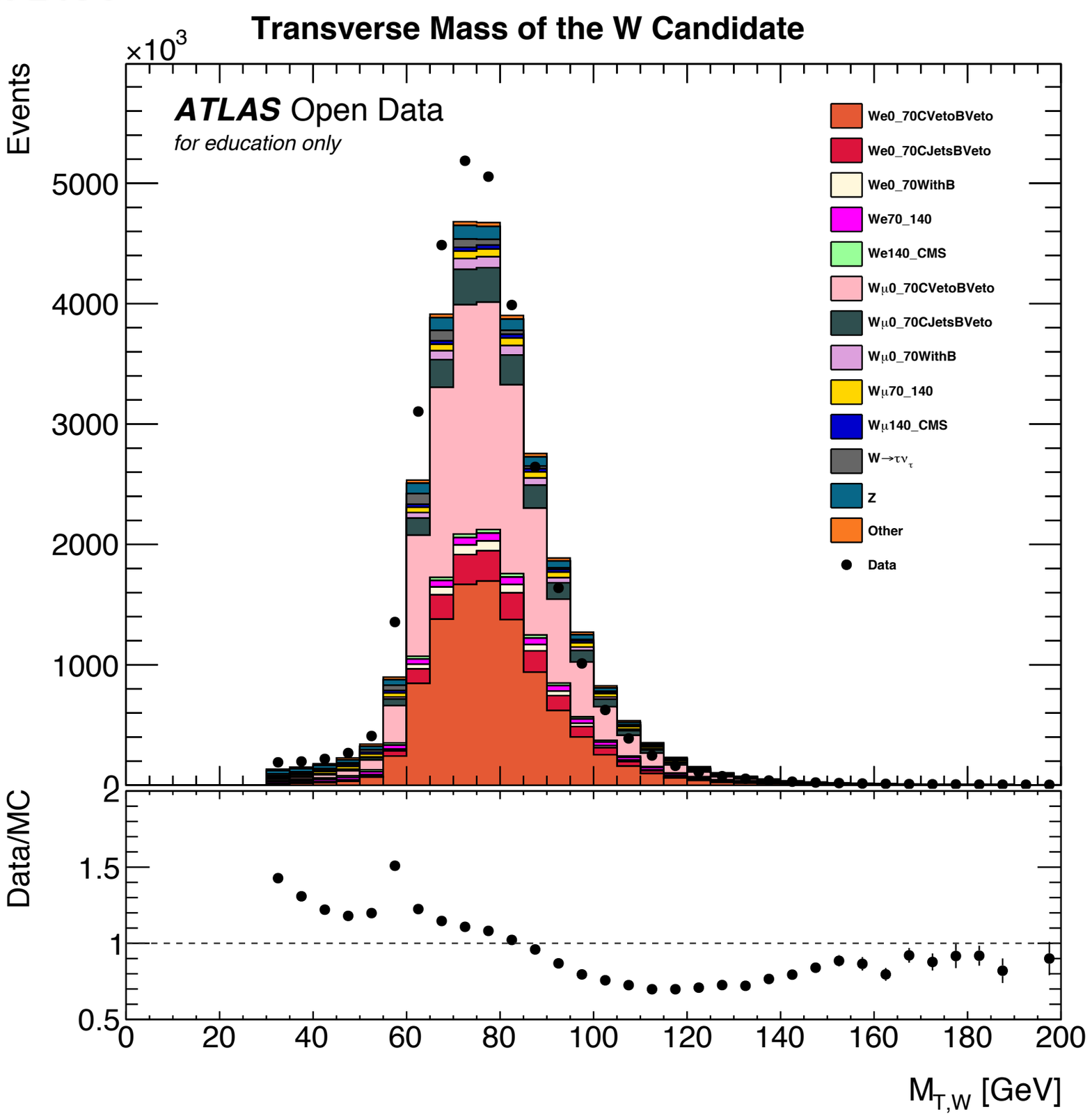
Transverse mass

$$M_{12} = \sqrt{(E_1 + E_2)^2 - (\mathbf{p}_1 + \mathbf{p}_2)^2}$$



$$M_T = \sqrt{(E_{T,1} + E_{T,2})^2 - (\mathbf{p}_{T,1} + \mathbf{p}_{T,2})^2}$$

W Transverse mass

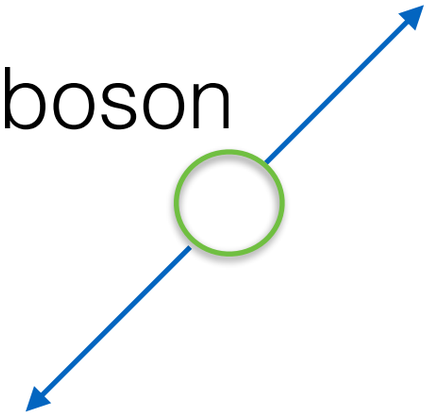


Data 2015

Transverse momentum

- Not only useful in inferring presence of neutrinos
- Z, W, Higgs bosons ~100 times heavier than proton
- Need large fraction of proton constituent energy to form this much mass
- Little energy left as kinetic (momentum) $E = E_K + M$
- Heavy bosons produced nearly at rest

Transverse momentum



- Heavy boson decay products sent out at 180^0 in boson rest frame

- In random direction in boson rest frame

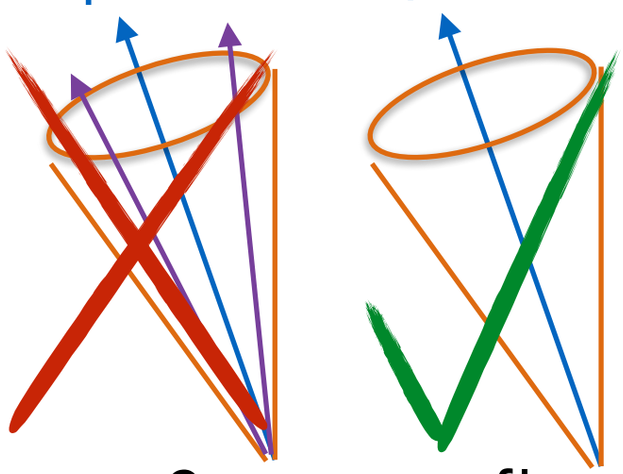
- Low chance products will move along beam line

- So high chance that products will have transverse momentum in boson rest frame

- Since boson produced nearly at rest, its rest frame is similar to the lab frame

Z Event selections

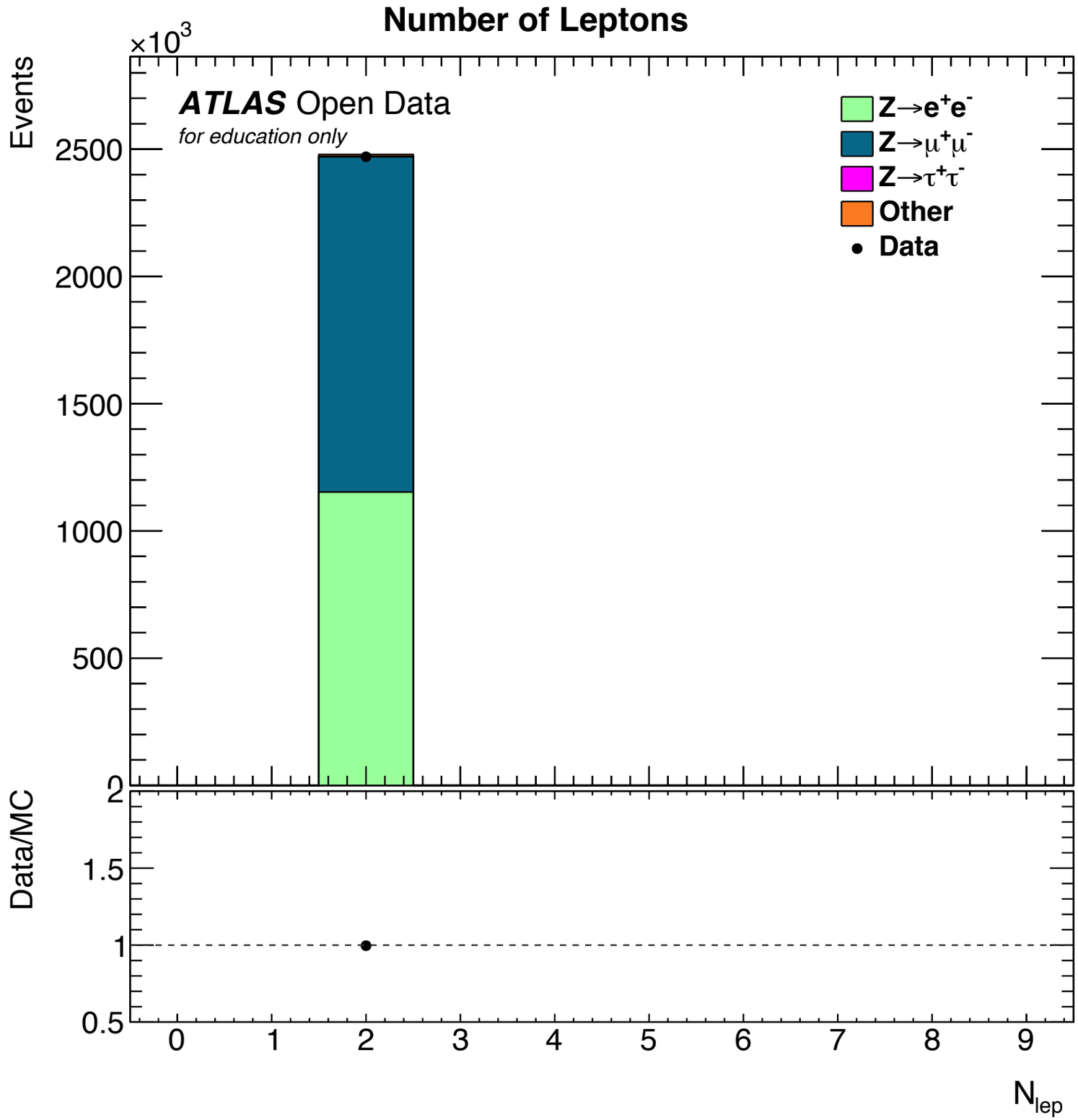
- Exactly 2 good leptons with p_T > 25 GeV
- Lepton ptcone30/p_T < 0.15
- Lepton etcone20/p_T < 0.15
- Leptons have opposite charge & same flavour
- Reconstructed Z mass within 20 GeV of quoted value

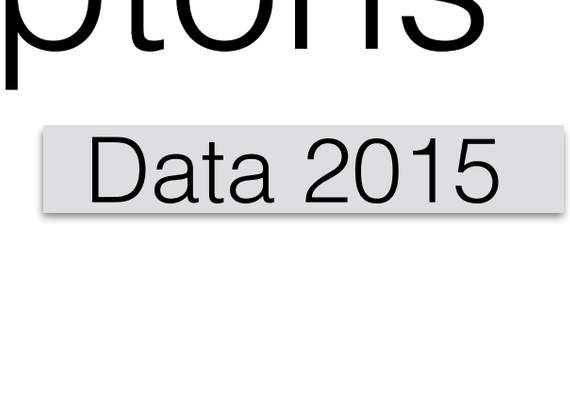


isolation

Z number of leptons

Data 2015

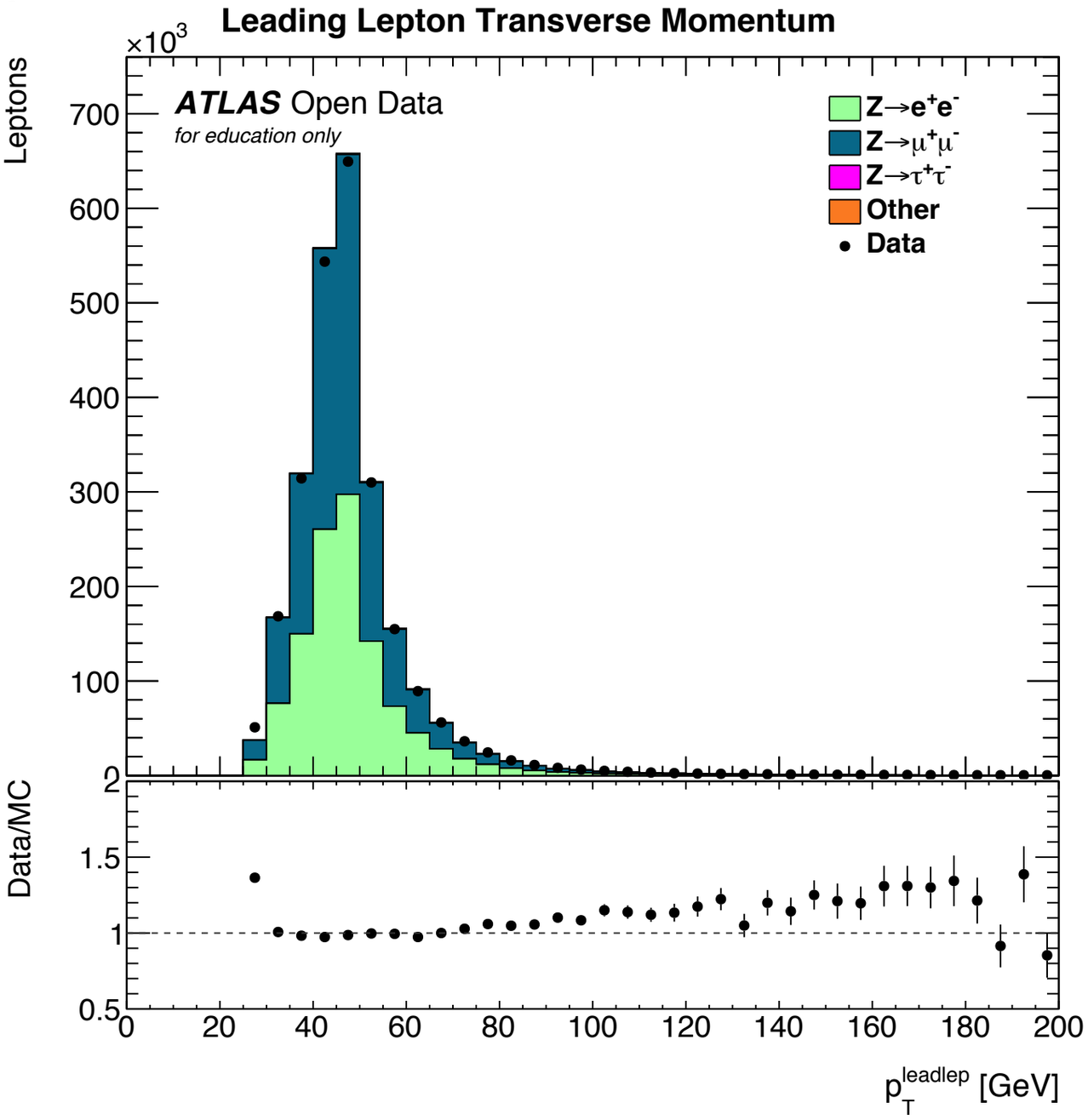




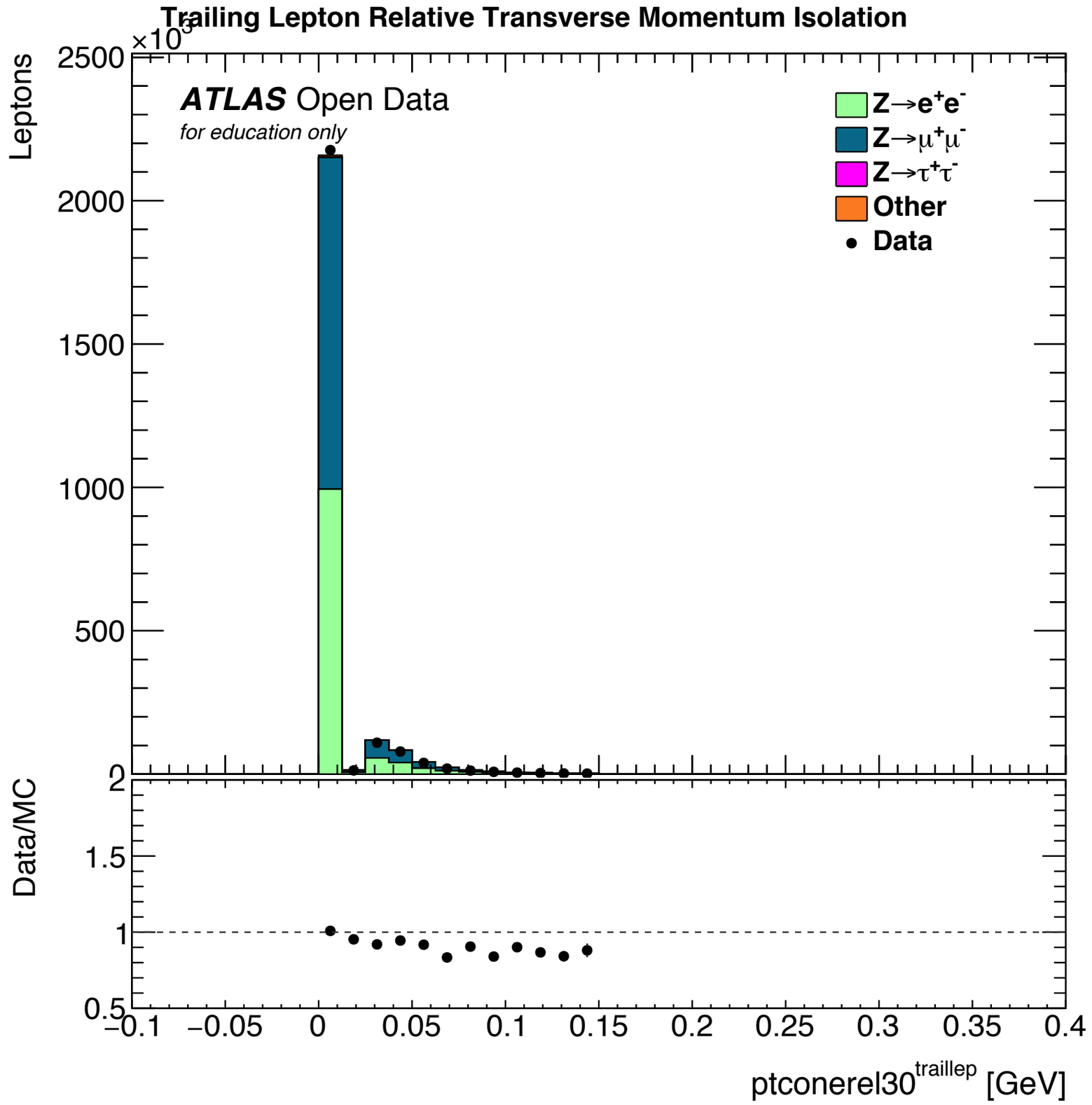
Z lepton p_T

Data 2015

cut at p_T > 25 GeV



Z lepton ptcone

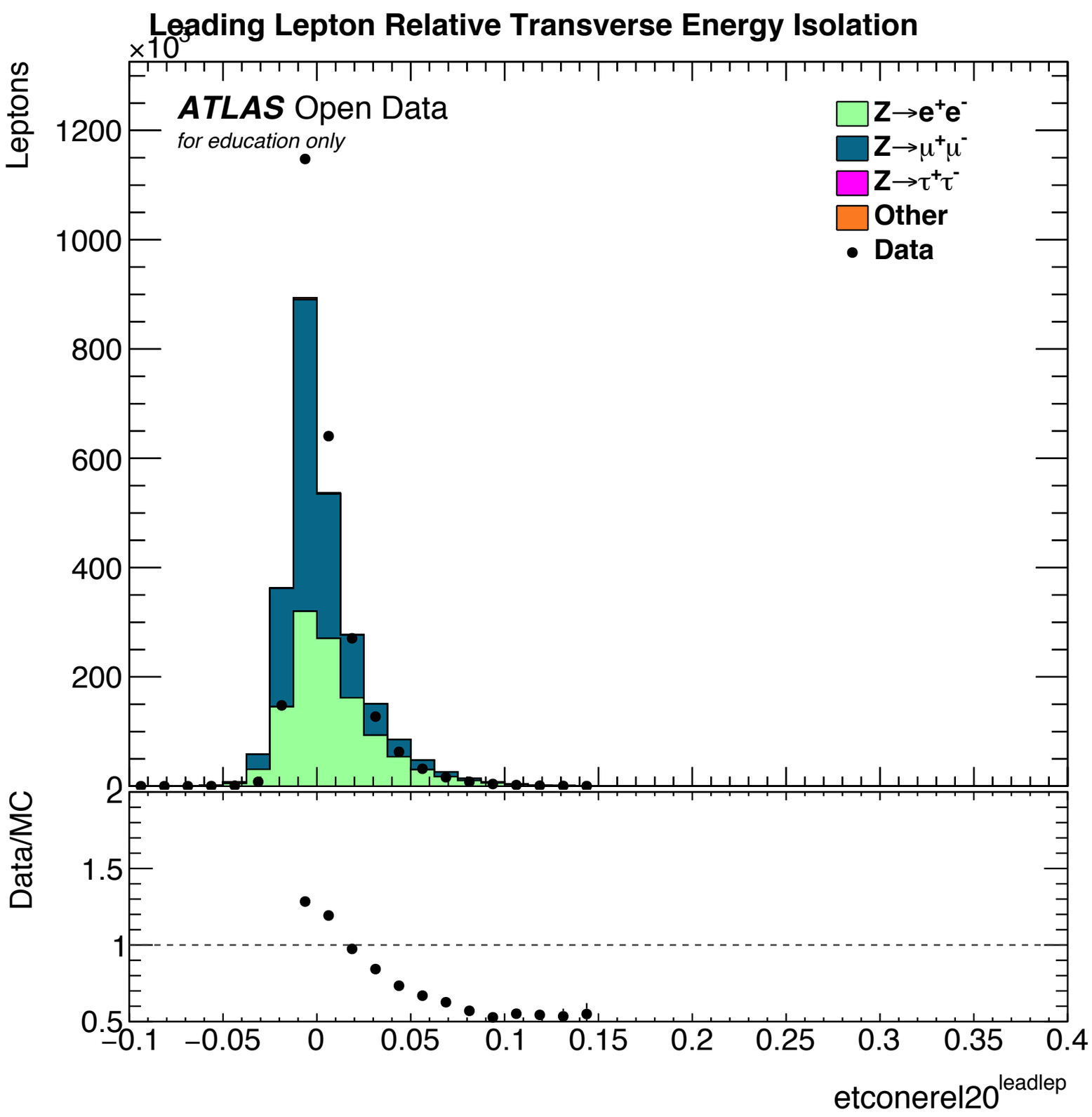


Data 2015

cut at

ptcone/p_T < 0.15

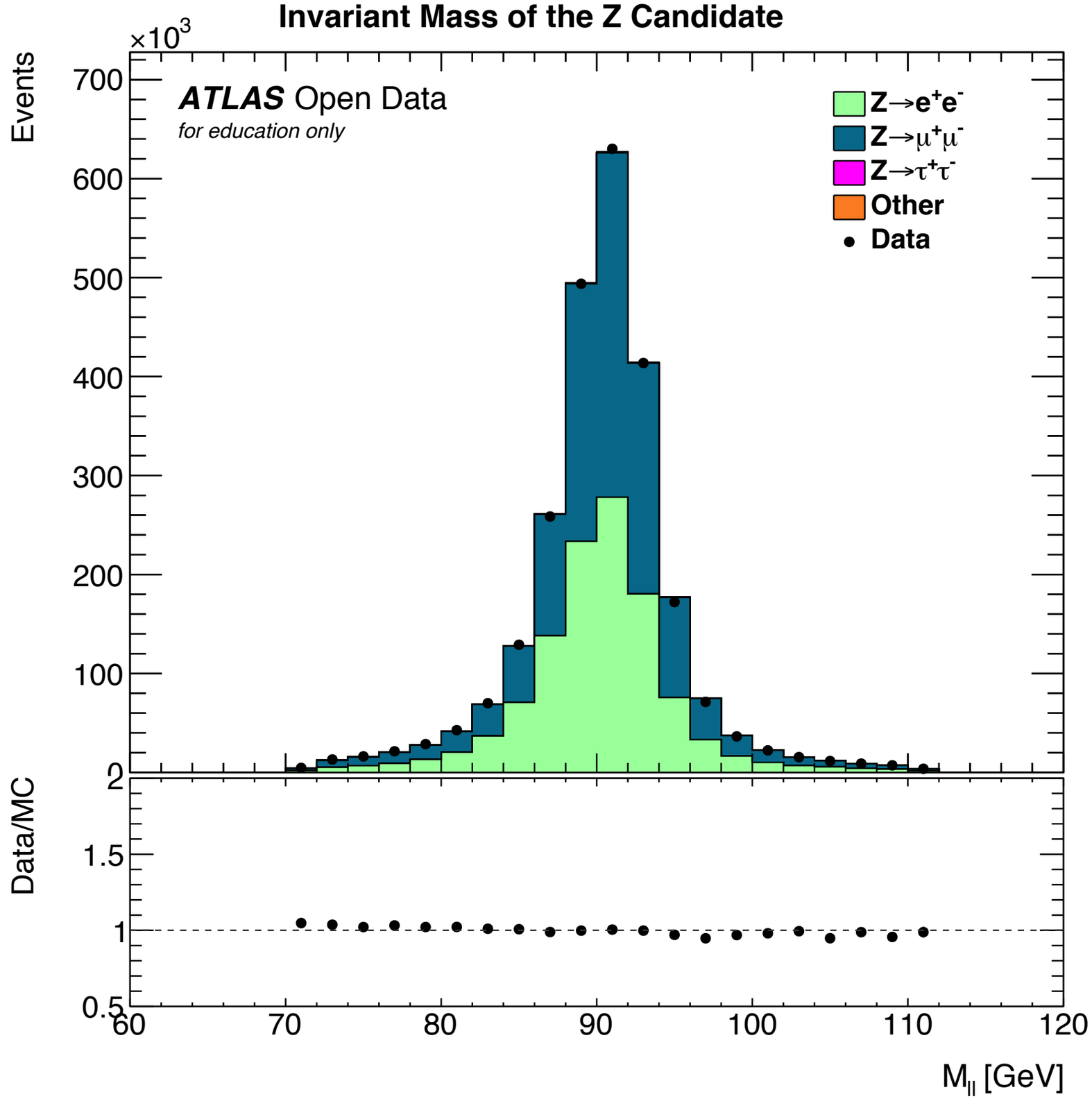
Z lepton etcone



Data 2015

cut at etcone/p_T < 0.15

Z Invariant mass



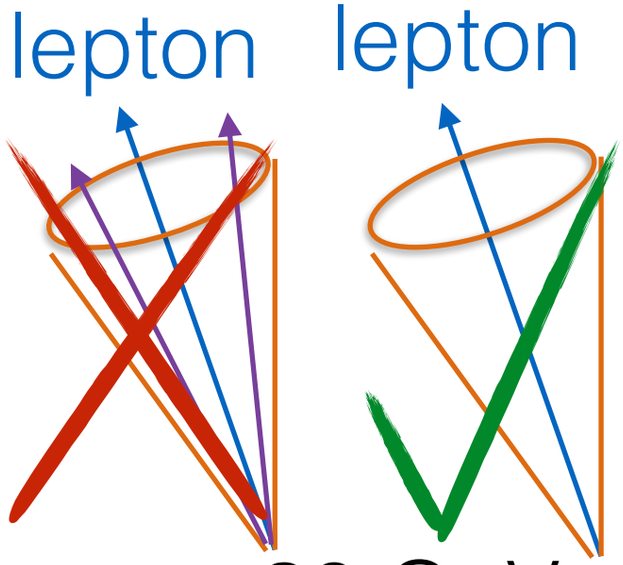
$Z \to e^+ e^-$

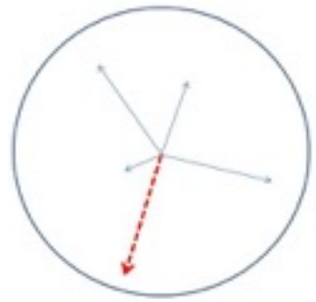
$Z \to \mu^+ \mu^-$

Data 2015

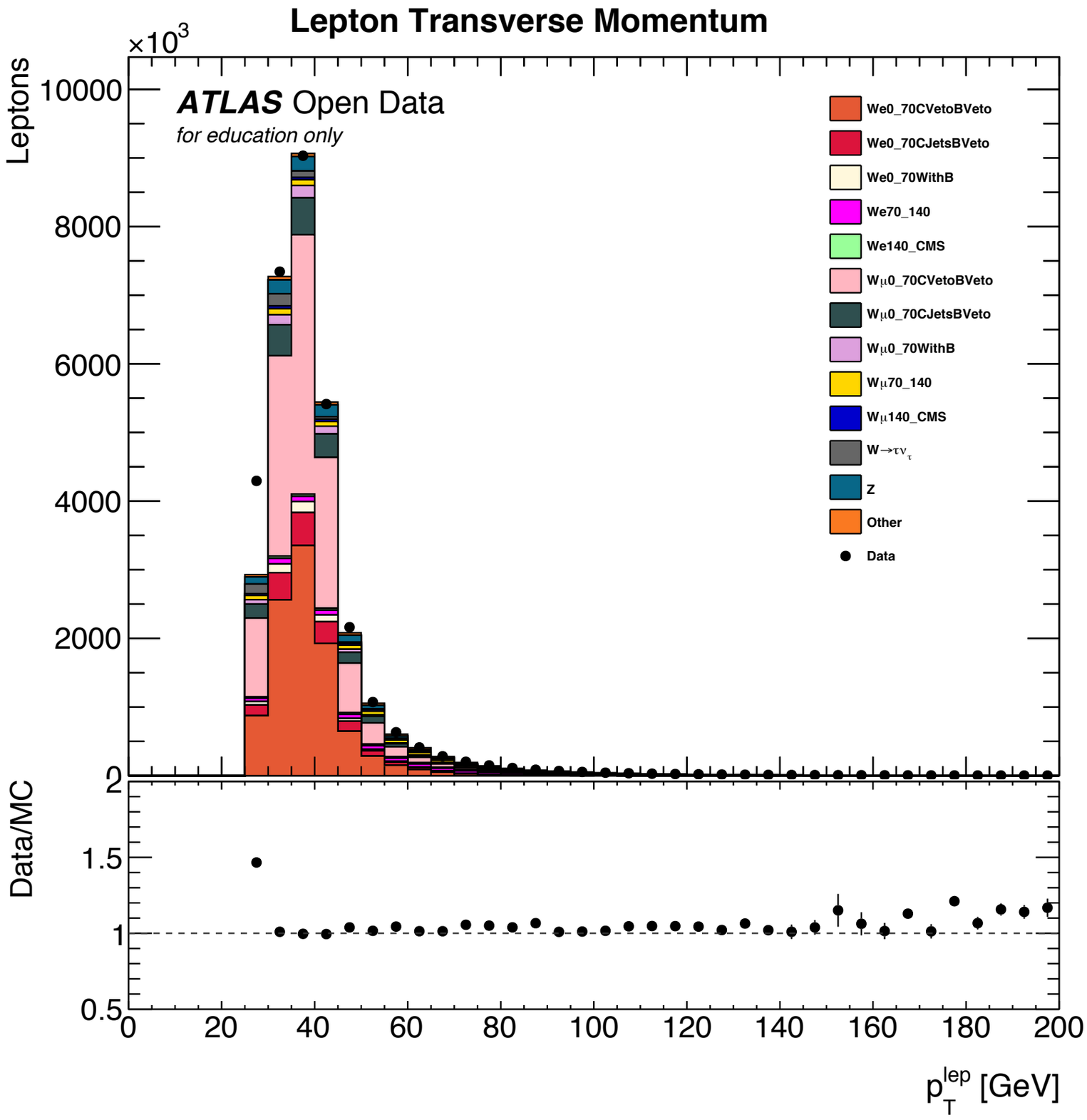
W Event selections

- Exactly 1 good lepton with p_T > 25 GeV
- Lepton ptcone30/p_T < 0.15
- Lepton etcone20/p_T < 0.15
- Missing Transverse Momentum > 30 GeV
- Reconstructed *W* transverse mass > 30 GeV





W lepton p_T

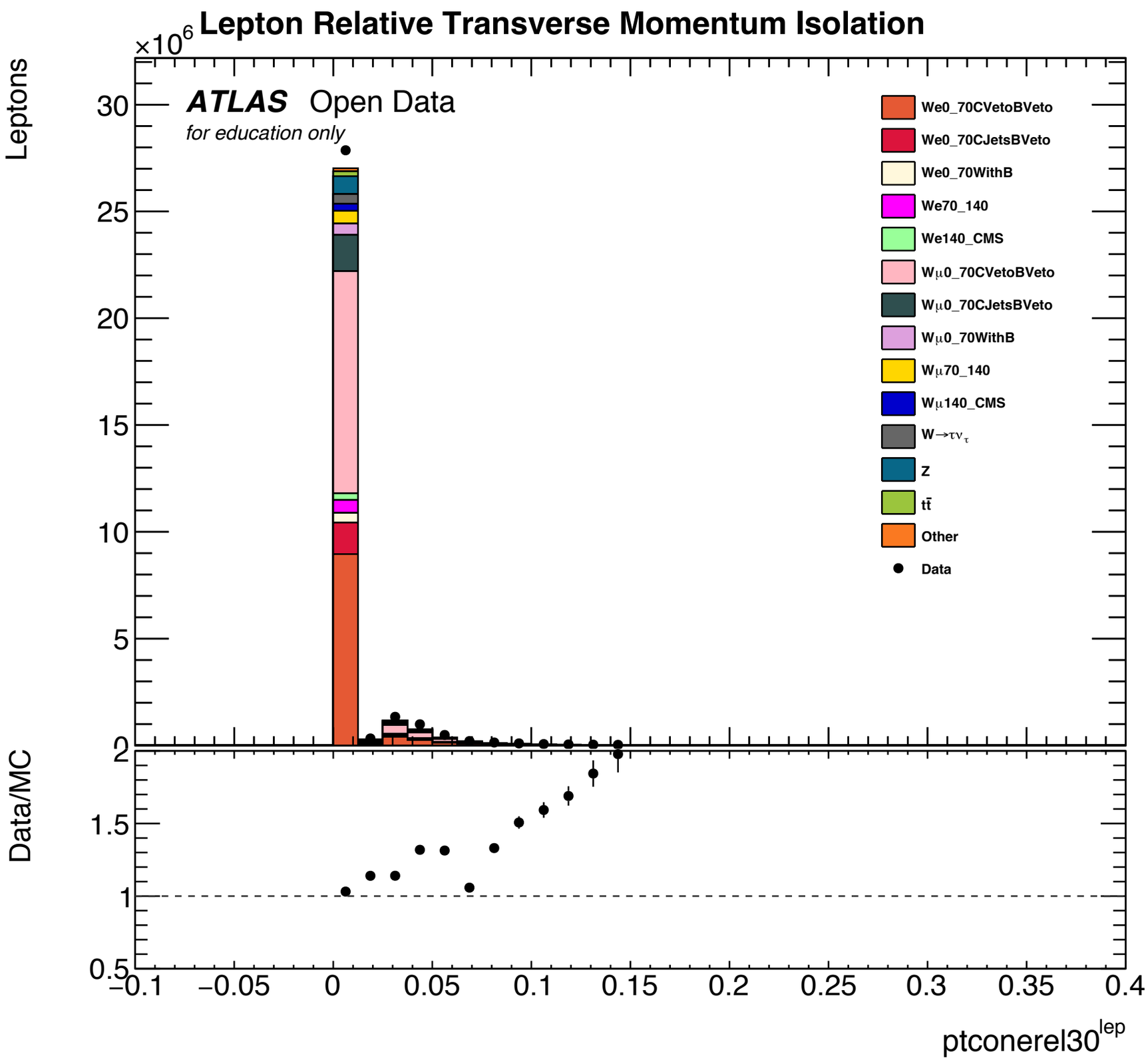


Data 2015

cut at

p_T > 25 GeV

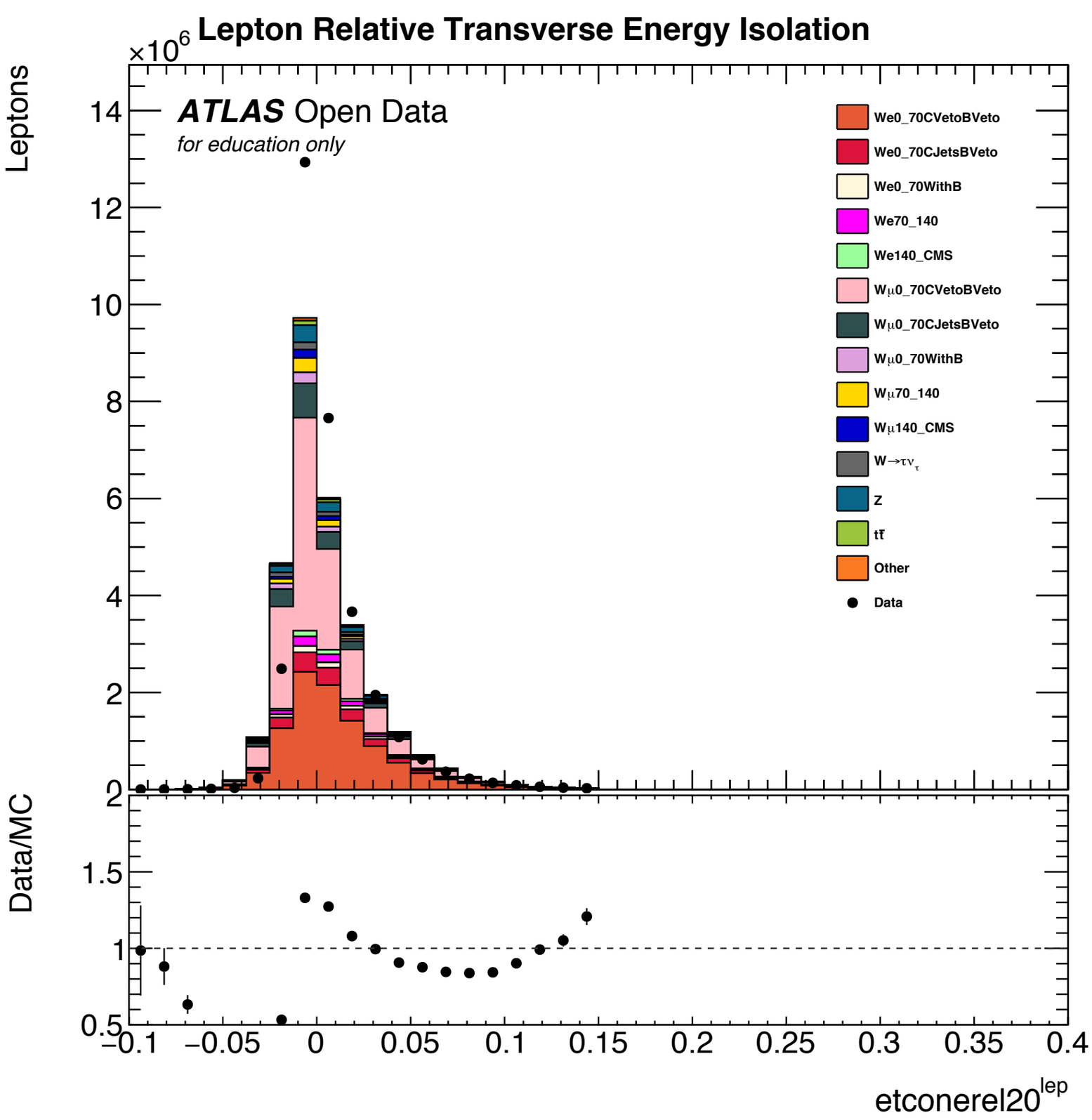
W lepton ptcone



Data 2015

cut at ptcone/p_T < 0.15

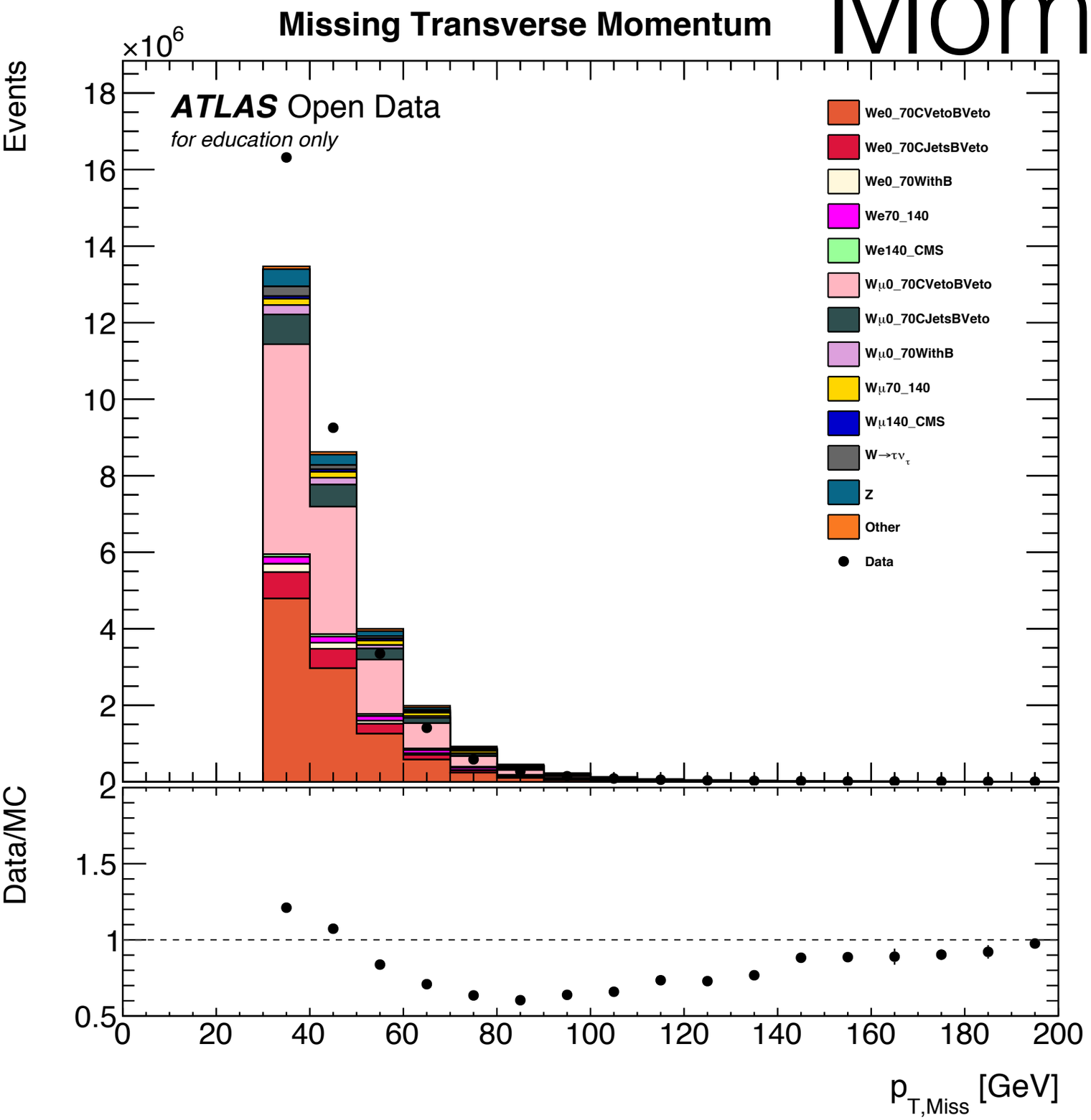
W lepton etcone



Data 2015

cut at etcone/p_T < 0.15

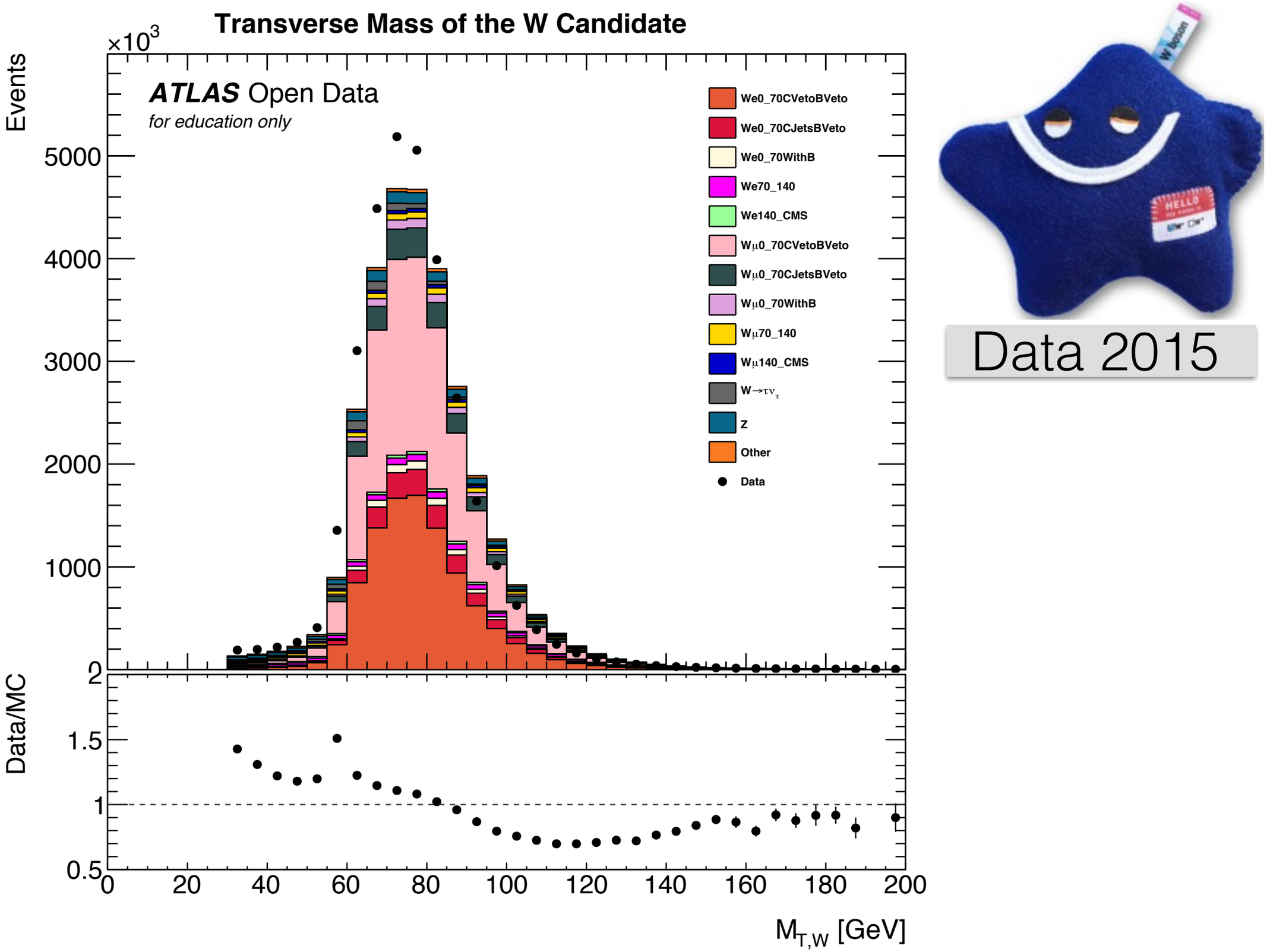
W Missing Transverse Momentum



Data 2015

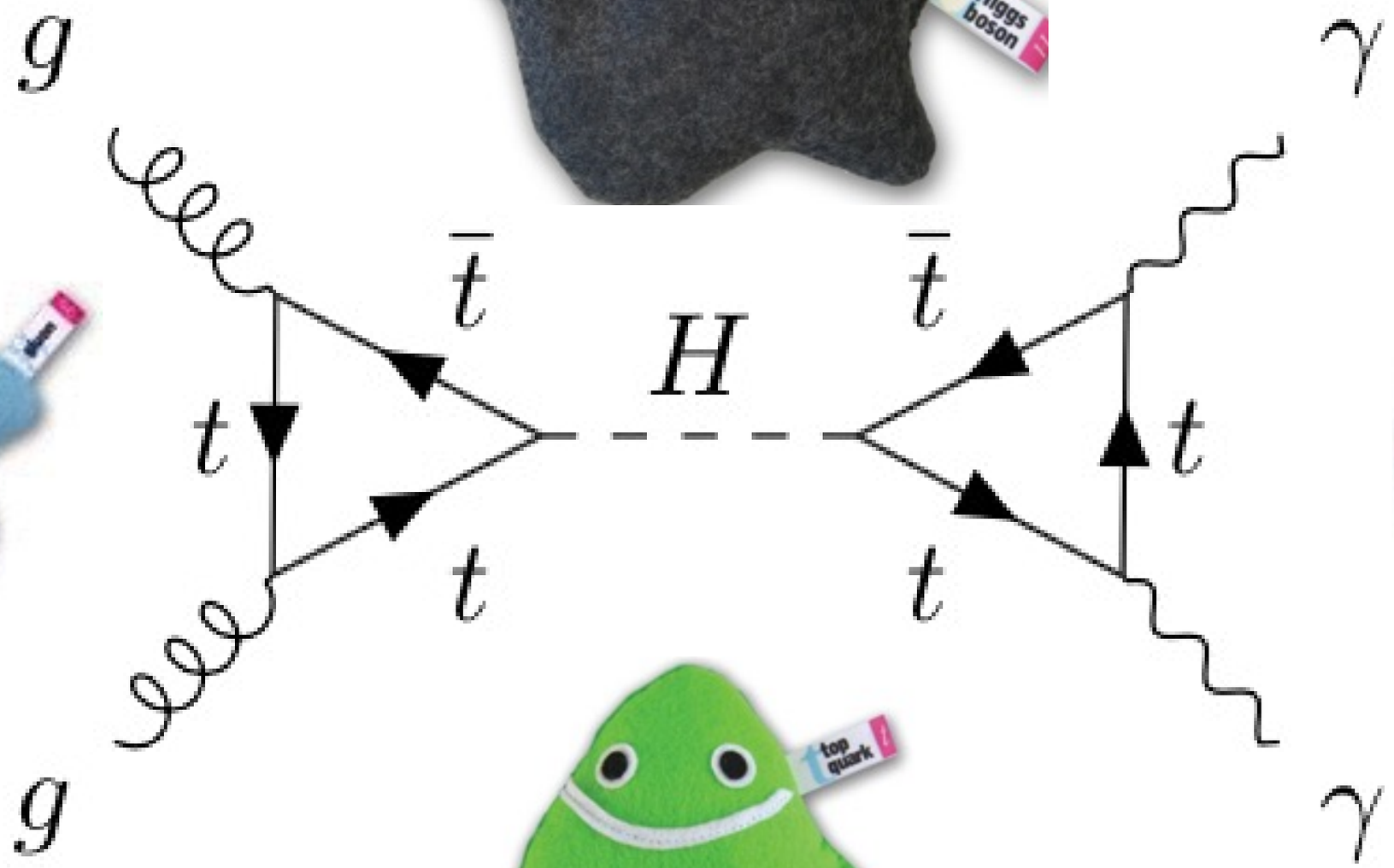
cut at Missing Transverse Momentum >30GeV

W Transverse mass

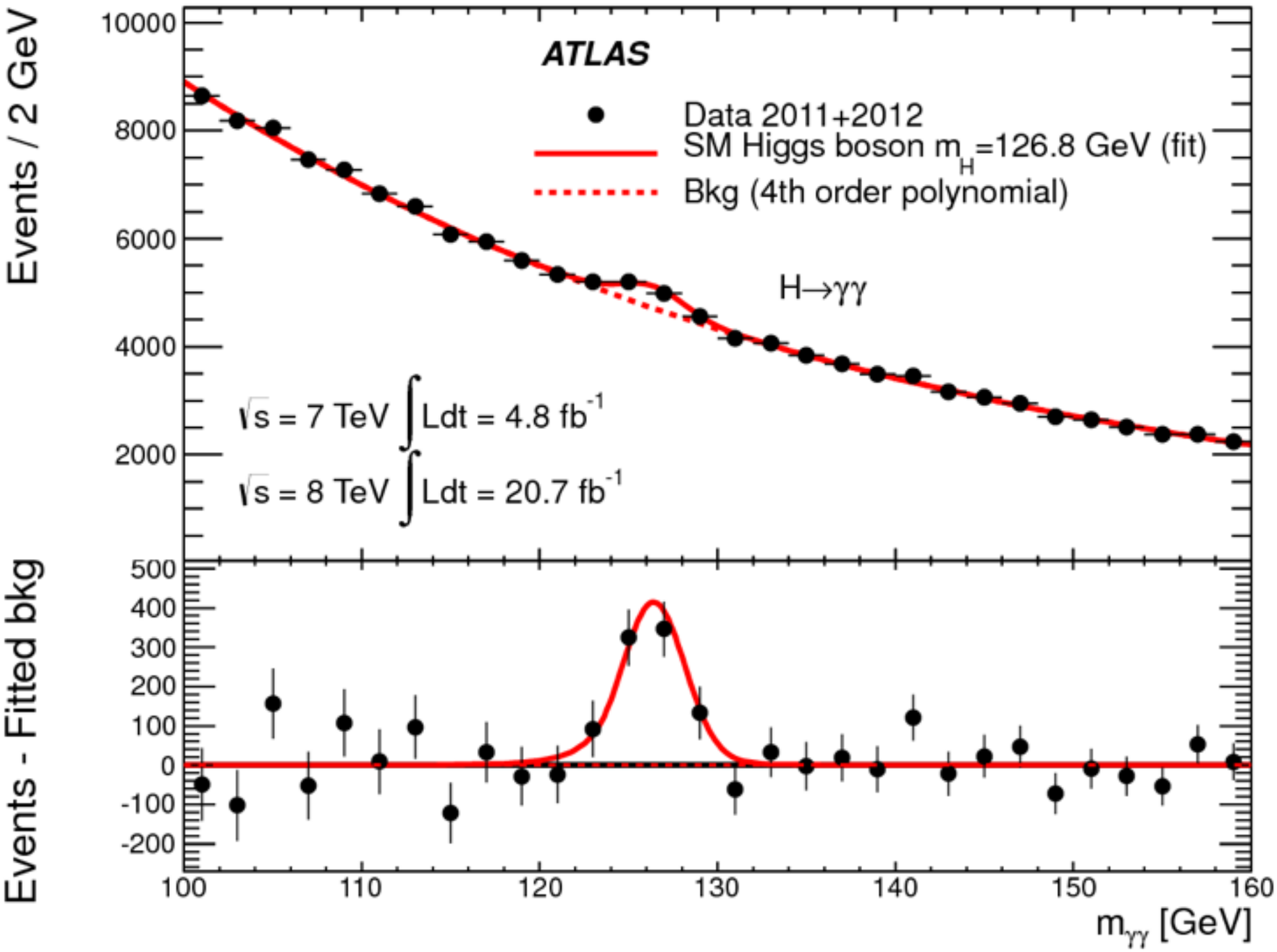


Data 2015

H decay



H peak



- Need photon energies & momenta for Higgs peak

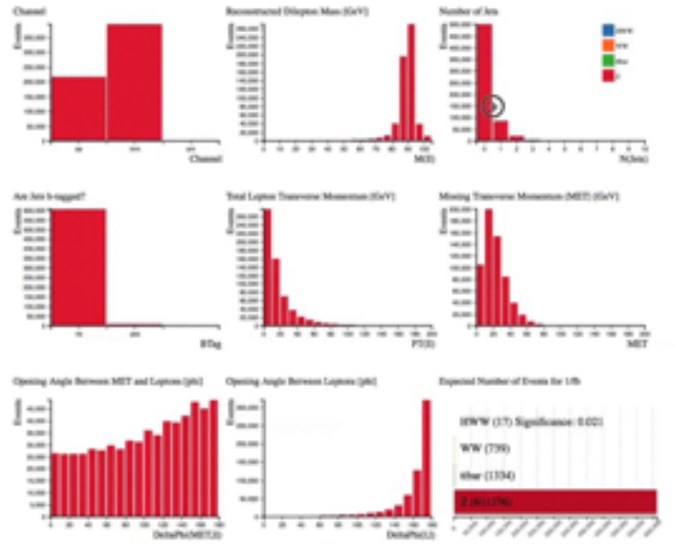
Current status

- Contribution to first public release of ATLAS 13 TeV Open Data

- Lot of work to be done before release

- Important steps made

- Hope 13 TeV Open Data will be used to improve current 8 TeV uses & find new ones

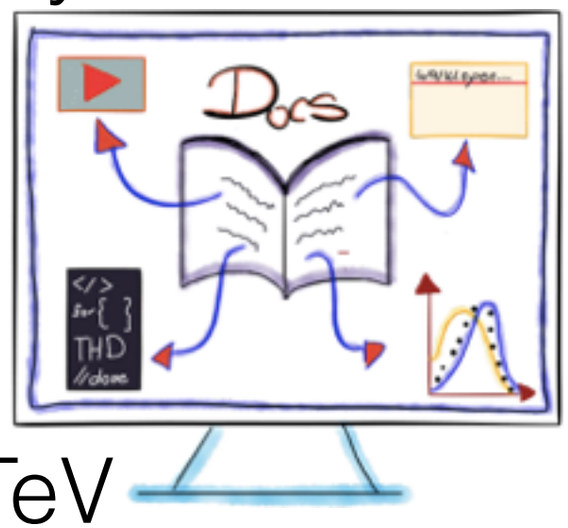


Summary

- There's lots **YOU** can do with 8 TeV ATLAS Open Data already

- ATLAS Open Data is aimed at students like you to use

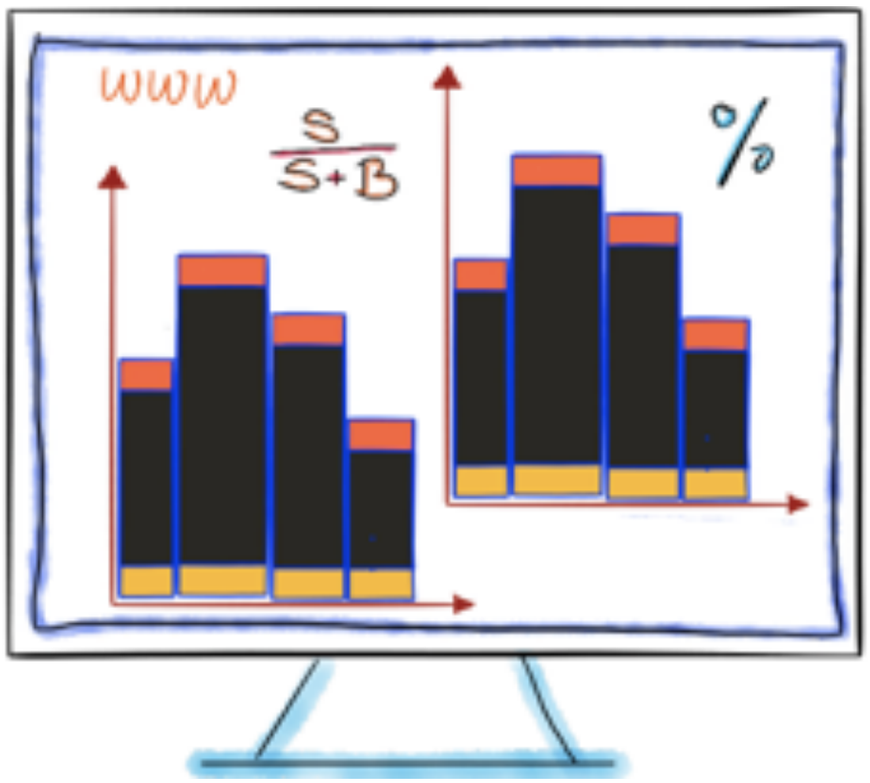
- opendata.atlas.cern to get started



- There'll be even more you can do with 13 TeV ATLAS Open Data

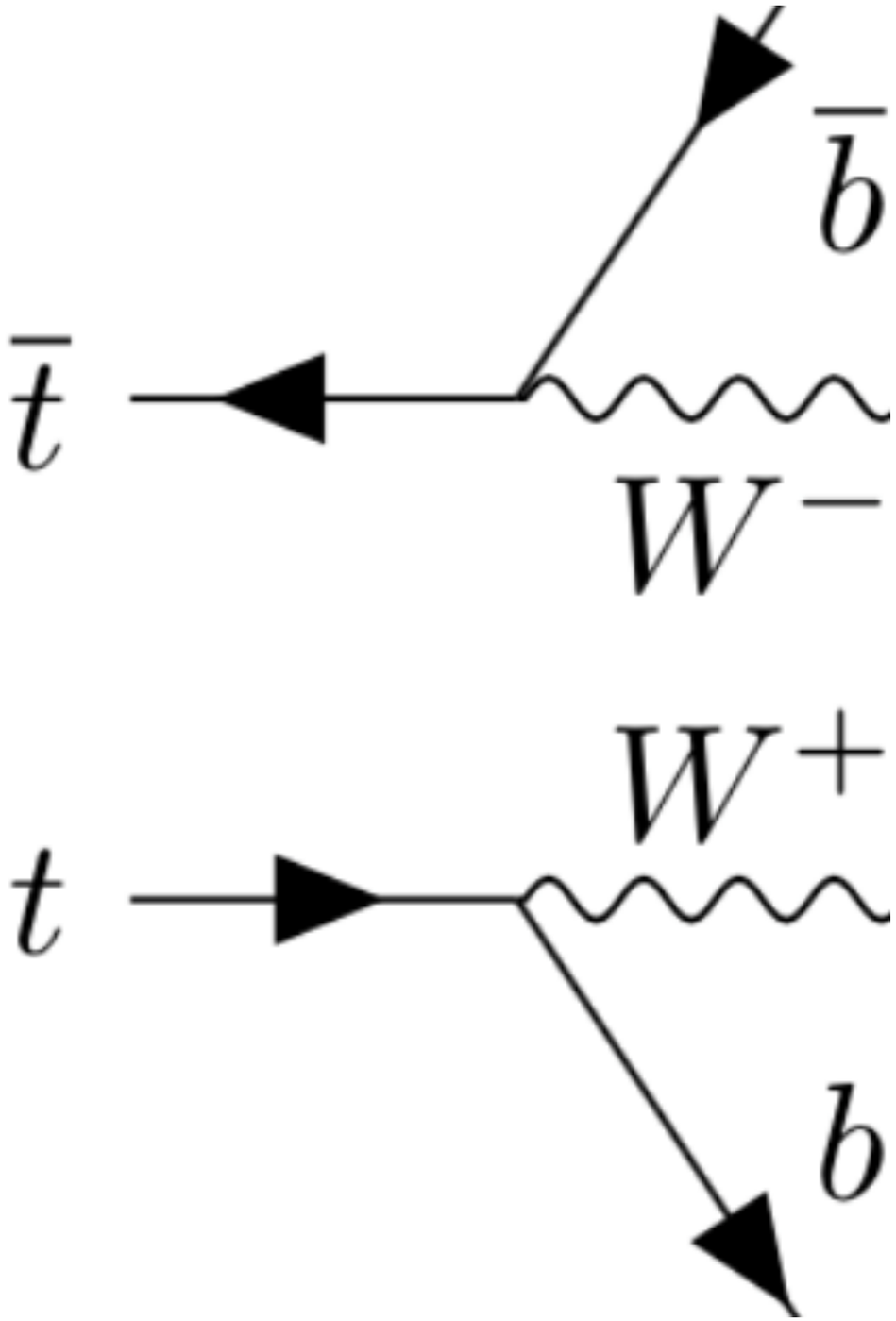
Thanks!

- meirin.oan.evans@cern.ch
- darren.price@manchester.ac.uk ask him to run another similar project



Backup

$t\bar{t}$



Creative Commons

- Article on interview about ATLAS Open Data

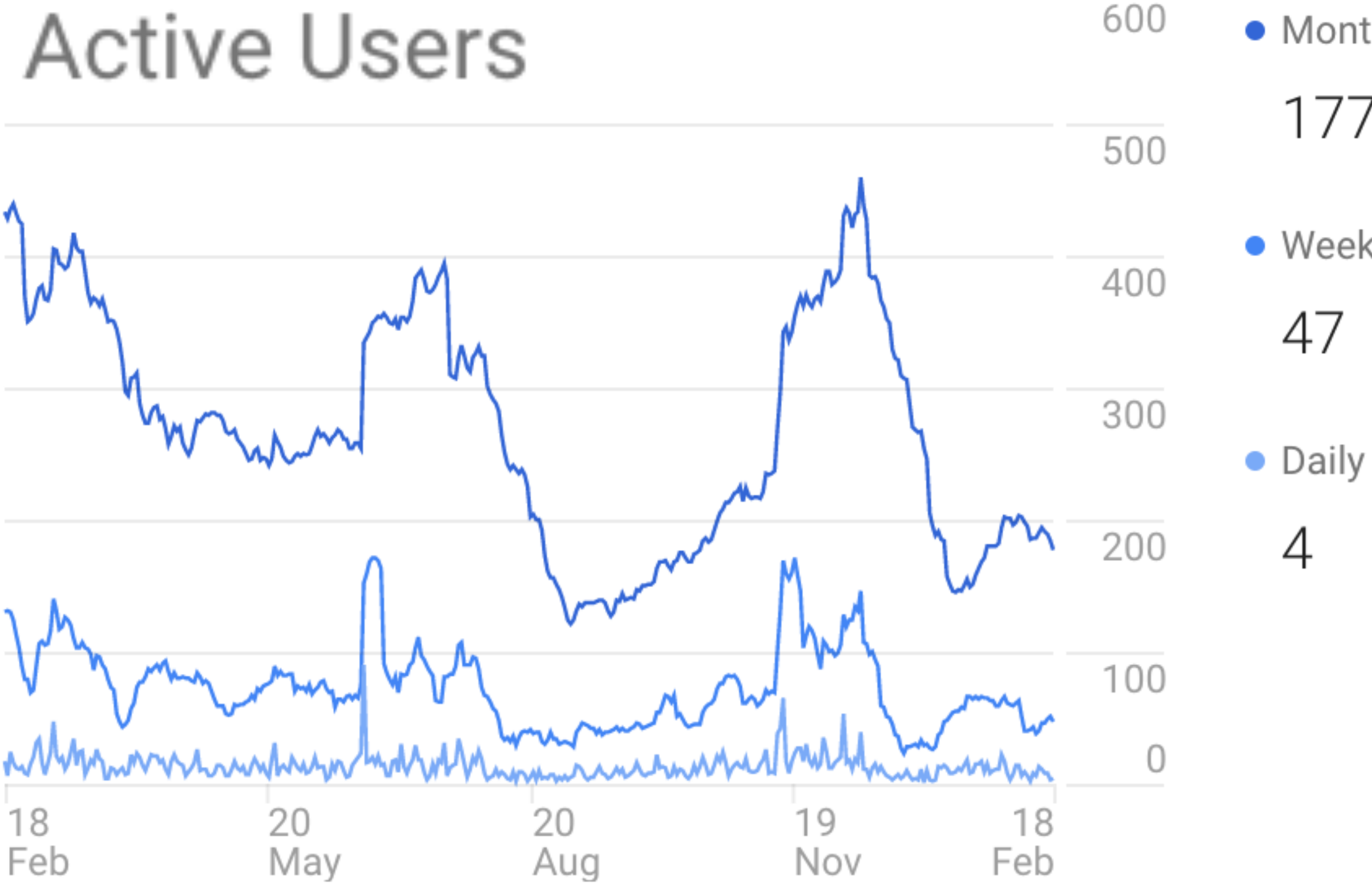
- Title: Making data and tools available for the world to see: Arturo Sanchez of CERN on why ATLAS uses CC0 data (link)

- Appeared on Creative Commons website November 2016 (soon after 8 TeV dataset release)

- Successful in reaching non-physics audience interested in open data & software (potential users)

- Want to follow the right licenses & procedures to release Open Source resources

Website use


Active Users
600
500
400
300
200
100
0
18 Feb
20 May
20 Aug
19 Nov
18 Feb
Monthly
177
Weekly
47
Daily
4

W peak

- Contributions from rarer processes not in inclusive sample (*W* from *t* production, di-boson production *WW*;*WZ*)

- Mis modeling problems / deficiencies in MC

- “fake” *W* candidates (events which pass applied selection criteria but don’t have a real *W*)

- Main contribution is multijet production

- Jet might be misreconstructed as lepton

- Leads to mismeasurement / mis-calibration of measured energy

- Leads to mismeasurement & creation of “fake” MET

Detector materials

- Calorimeters are layers of passive or absorbing high-density material
- e.g. lead
- interleaved with layers of an active medium like solid lead-glass or LAr
- HCAL always sampling (alternate between absorber & readout)
- ECAL might be homogenous
- Muon chambers are gas tubes

ATLAS detector

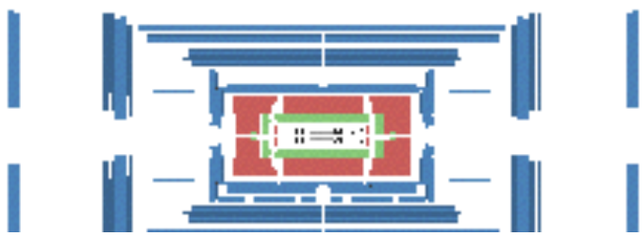
display instantly

Electron
Proton
Neutrino
Photon
Positron
Anti-proton
Jets
Muon
Neutron
Anti-muon

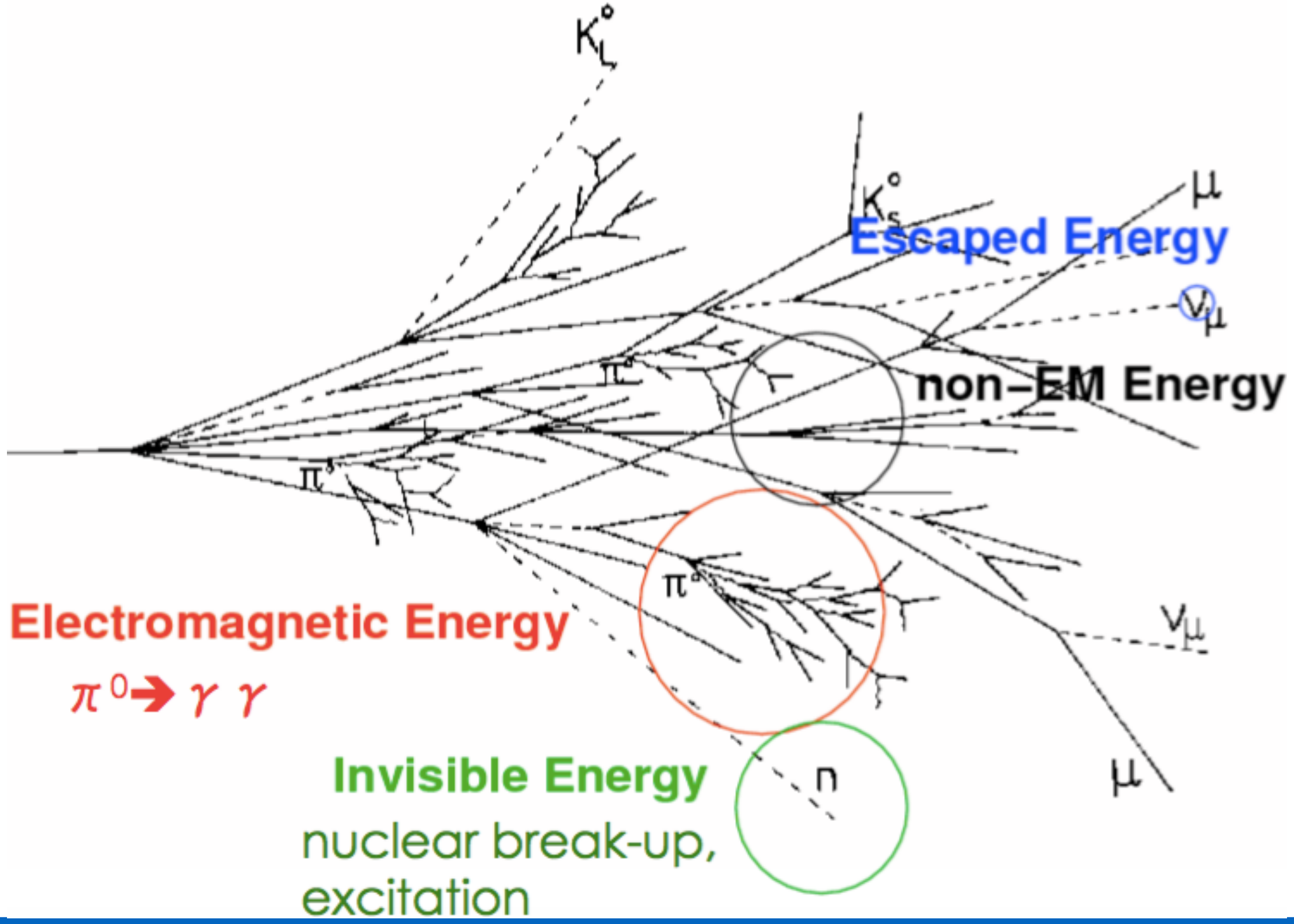
Magnification 3x

Pixel/SCT detector
Transradiation tracker
Electromagnetic Calorimeter
Inner magnetic field 2 T
Solenoid magnet

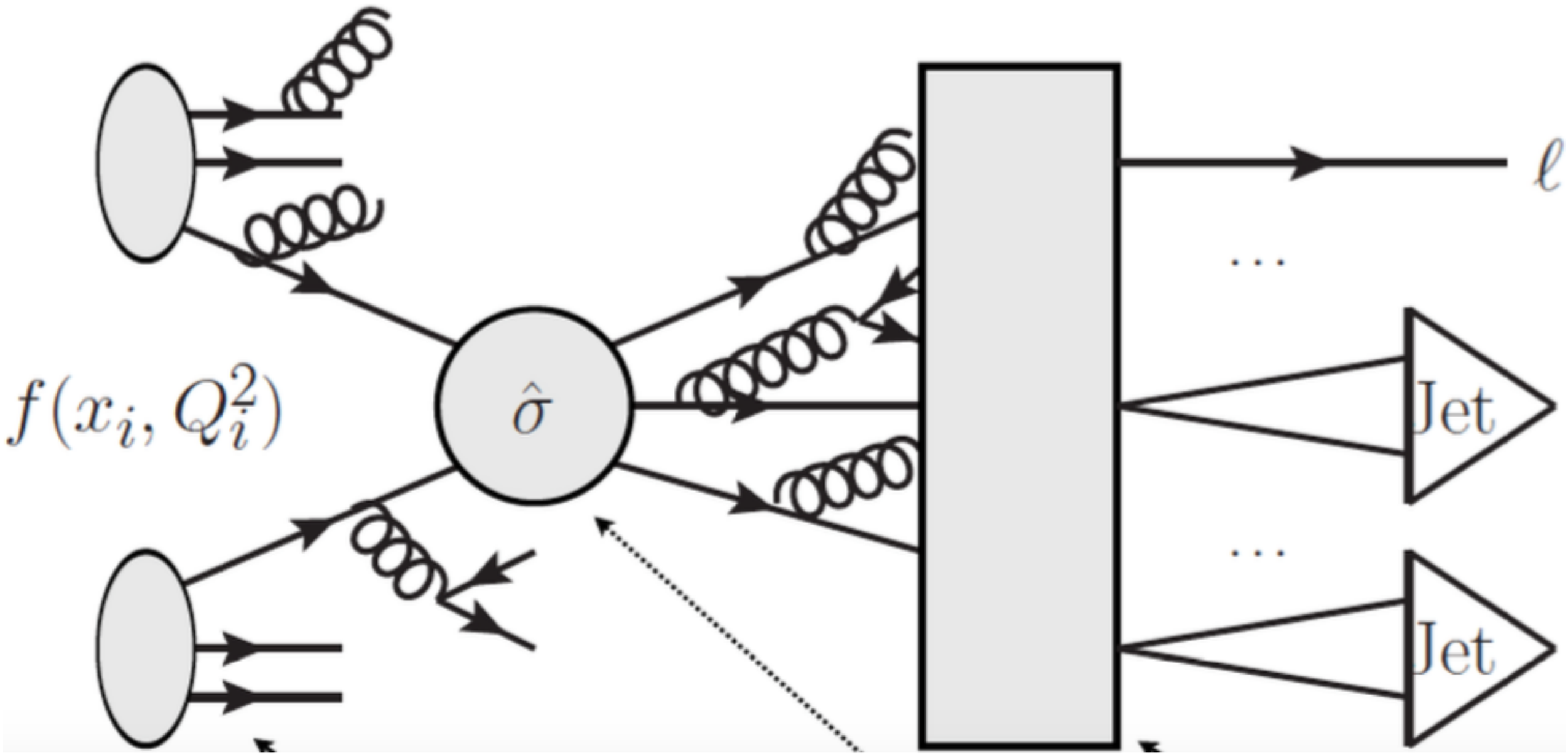
Hadronic Calorimeter
Muon Spectrometer
0 m
1.4 m
2.28 m
4.25 m
Toroid magnets
Outer magnetic field 0.5 T
10.569 m



Hadronic shower

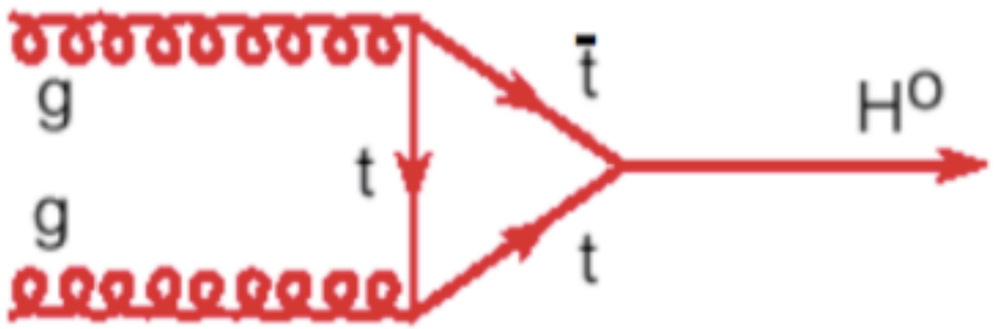


Hadronisation

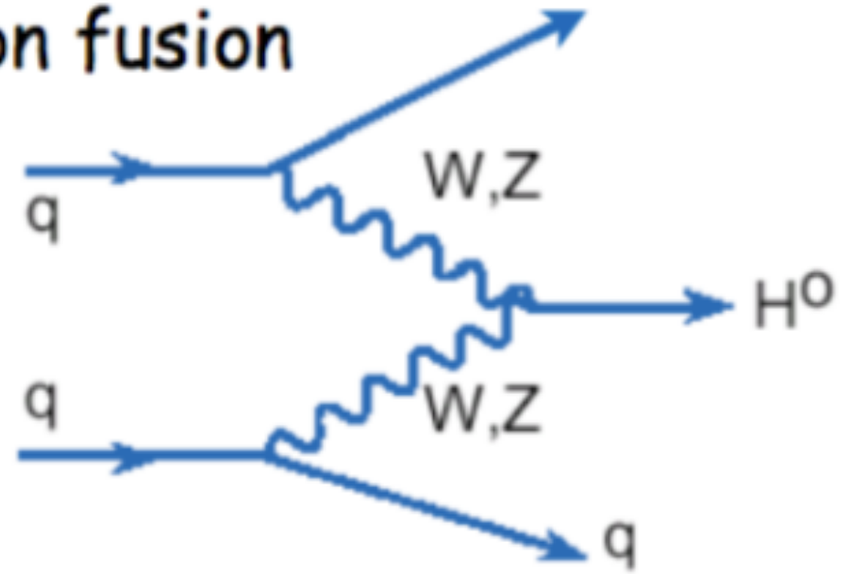


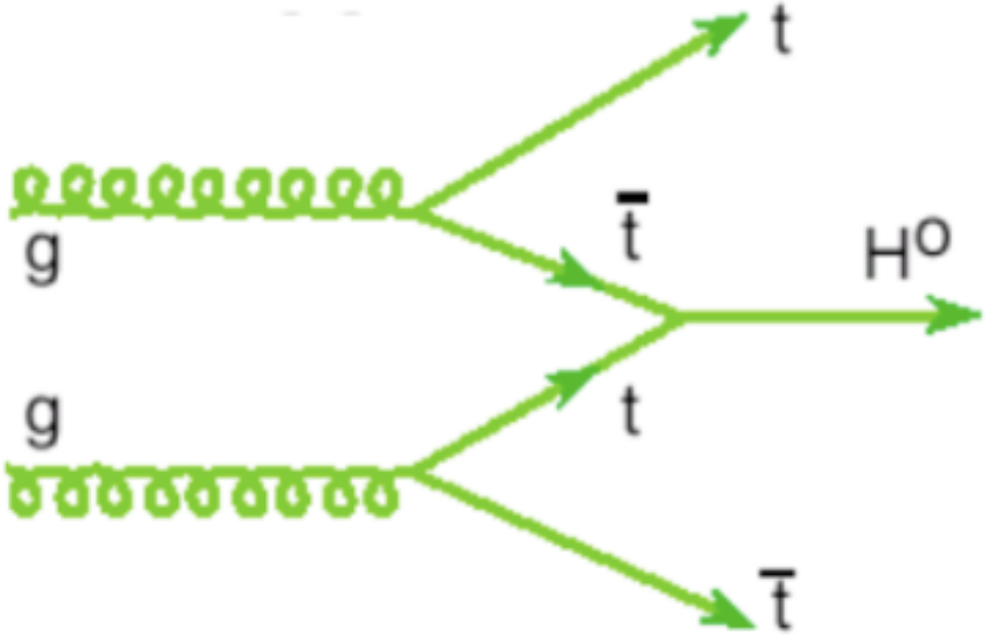
Higgs production

1. Gluon fusion

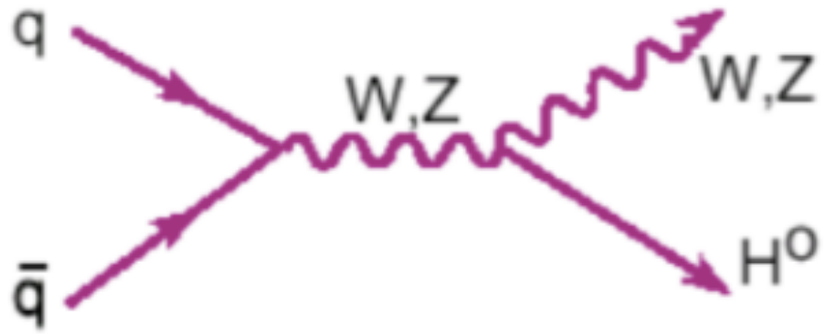


2. Vector boson fusion



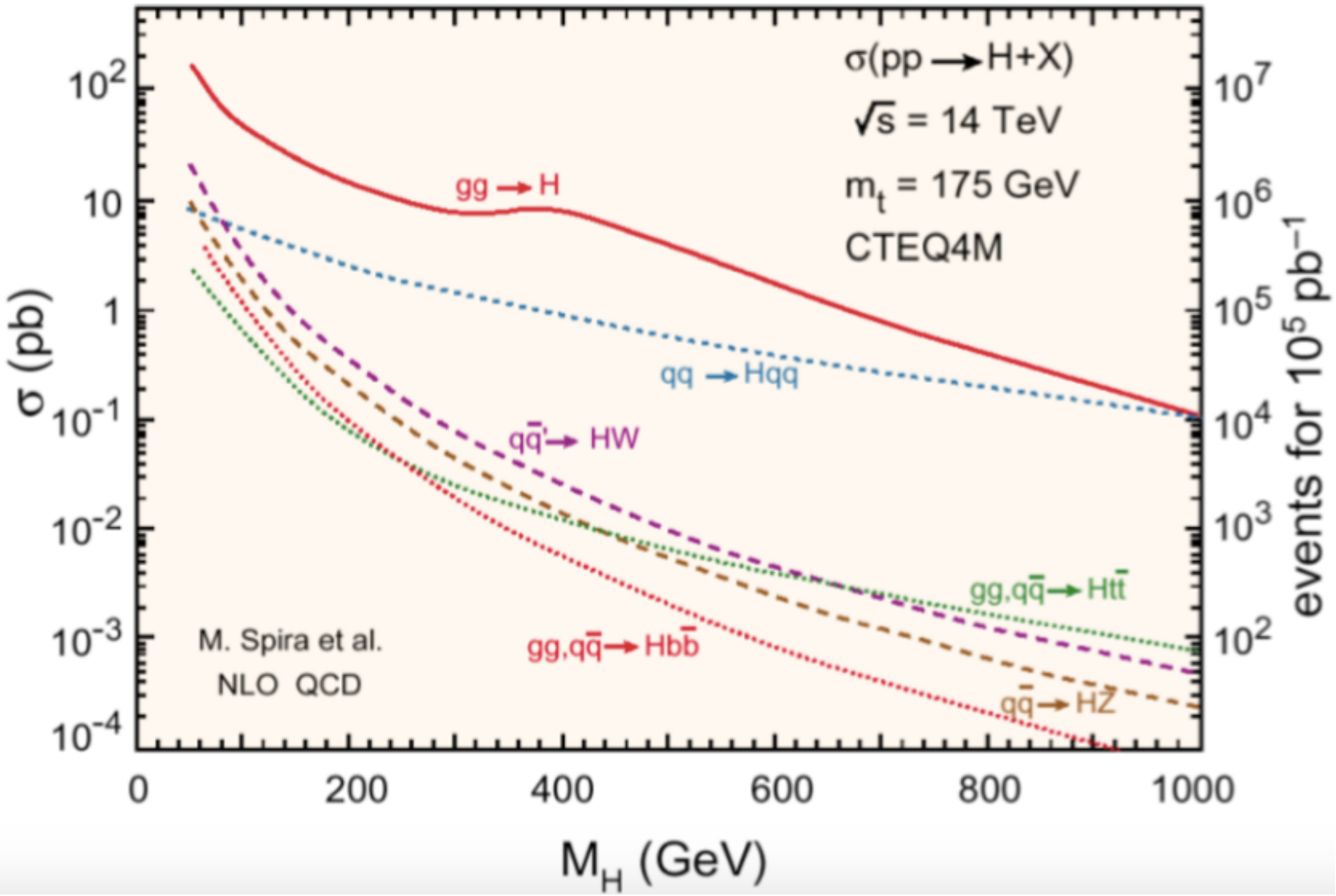


3. $t\bar{t}$-fusion

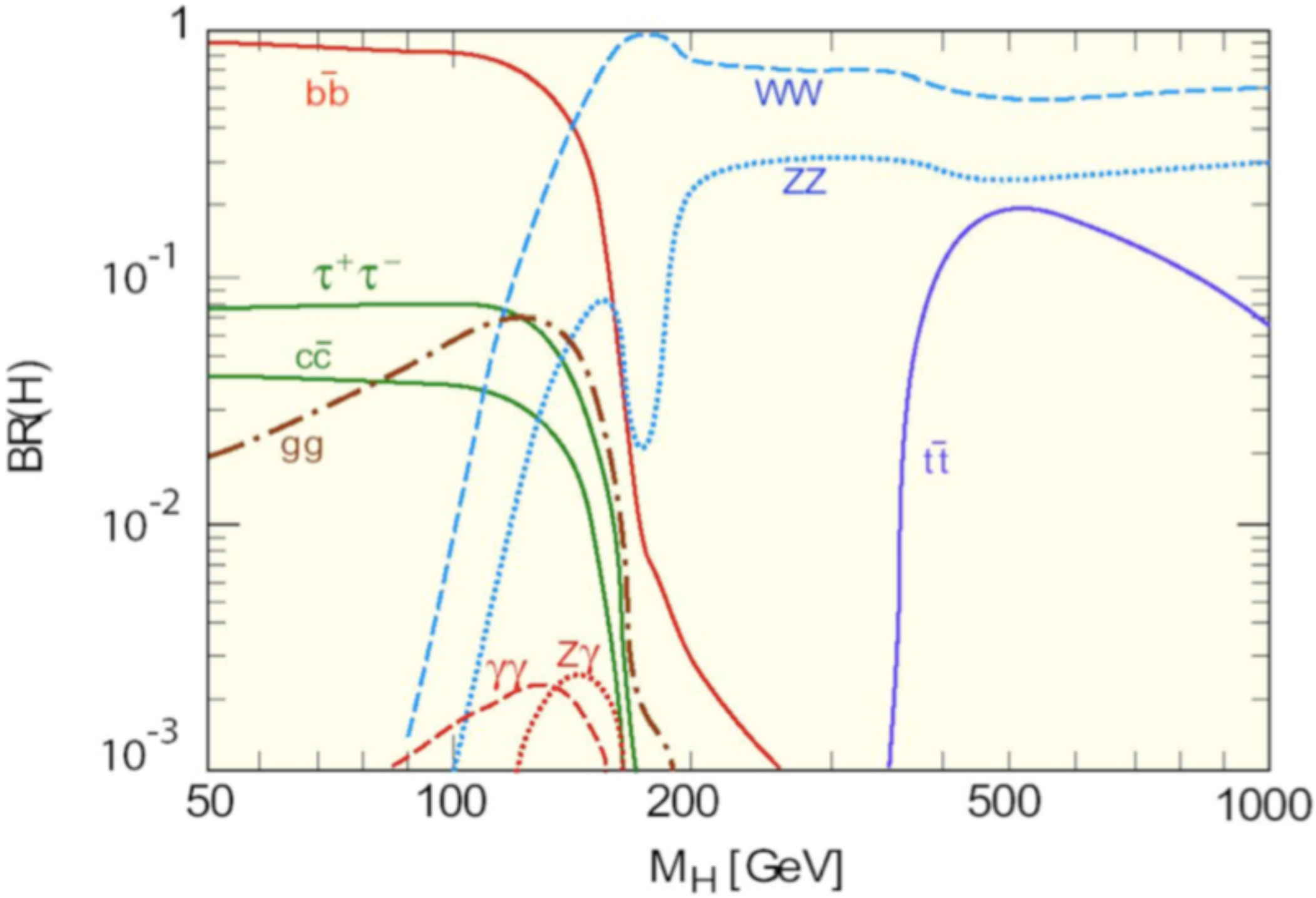


4. Associated production

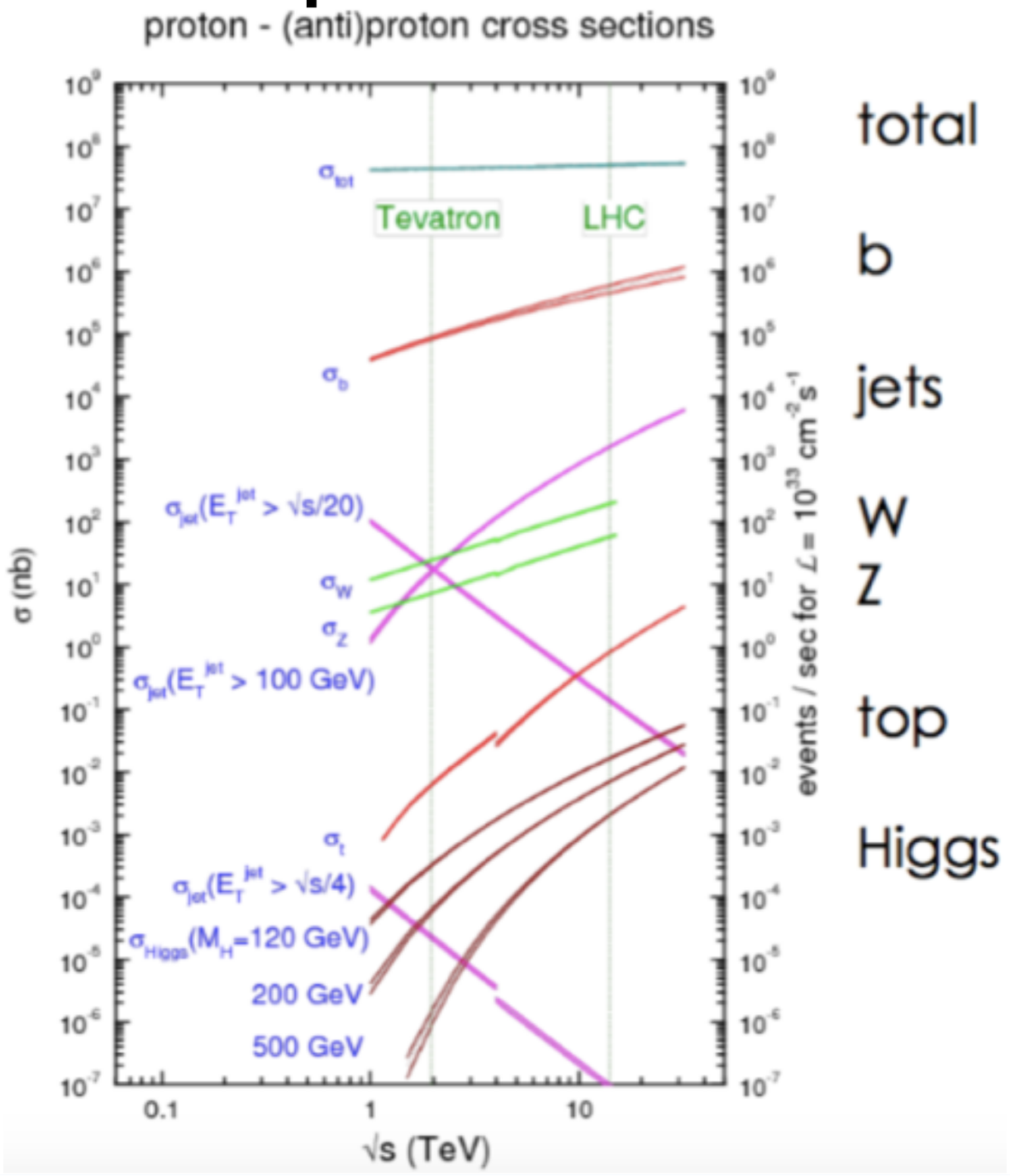
Higgs production



Higgs decay


1
10^{-1}
10^{-2}
10^{-3}
BR(H)
$b\bar{b}$
WW
ZZ
$\tau^+\tau^-$
$c\bar{c}$
gg
$\gamma\gamma$
$Z\gamma$
$t\bar{t}$
50
100
200
500
1000
M_H [GeV]

LHC production



LHC production