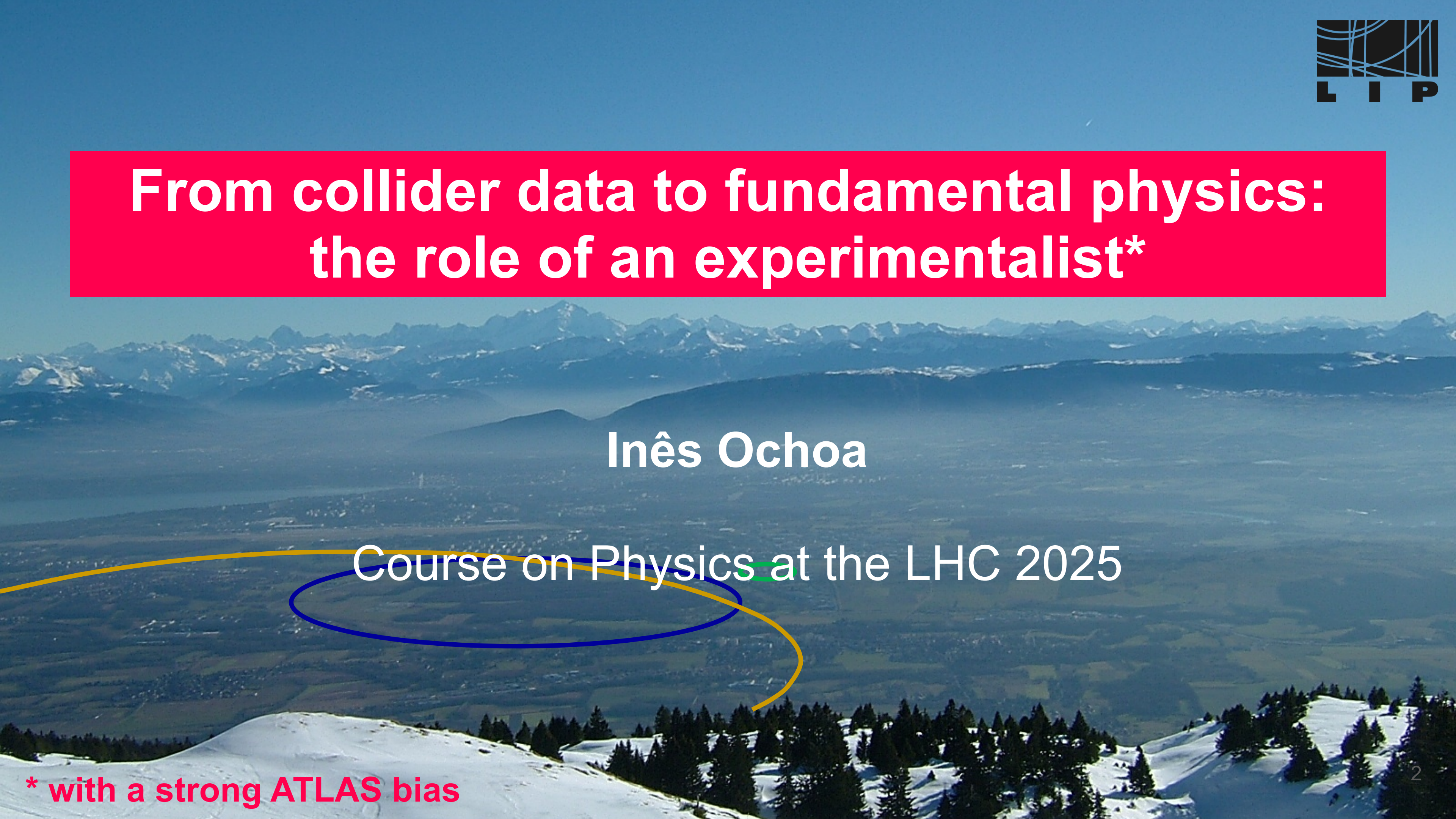
Analysis Methods

Inês Ochoa

Course on Physics at the LHC 2025



From collider data to fundamental physics: the role of an experimentalist*

Inês Ochoa

Course on Physics at the LHC 2025

*** with a strong ATLAS bias**

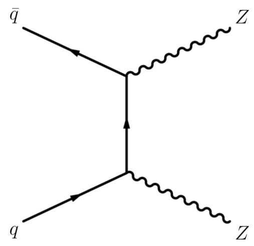
Introduction

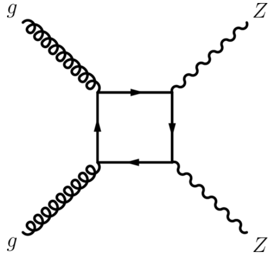
- The role of an experimentalist is to piece together all the elements in the chain that links theory and data.

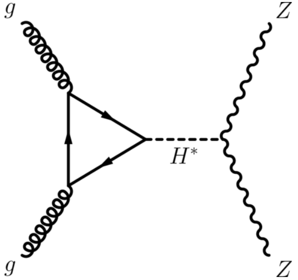
- Main topics:
  1. Event reconstruction
  2. Trigger and detector operations
  3. Simulation
  4. Calibrations
  5. Computing & Software
  6. Case study: measuring the WH cross-section
     1. Simulation-based backgrounds, global fit,
  7. Case study: searching for new physics resonances
     1. Data-driven background estimation methods
     2. Anomaly detection

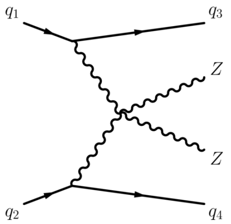
What physics comes out of LHC data?

ZZ

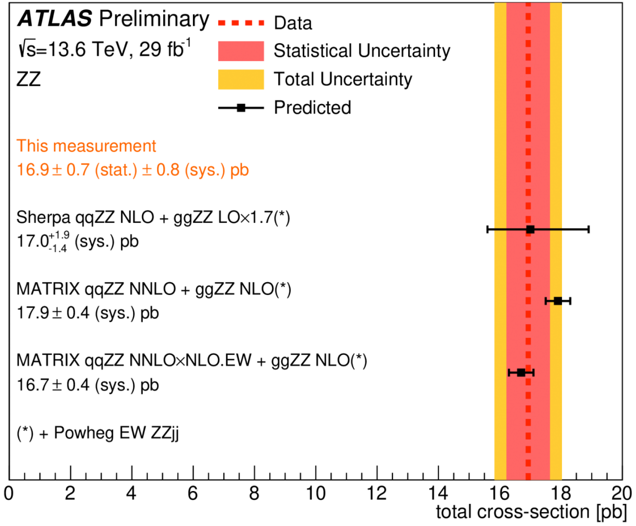






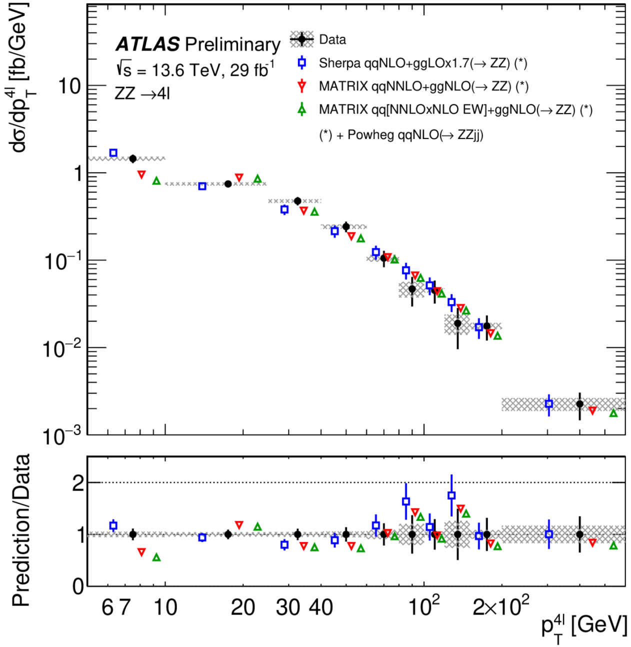


Total, fiducial* cross-sections



*in the detector's acceptance

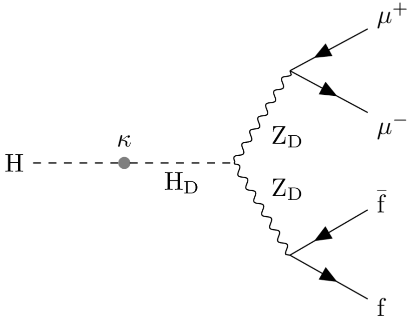
Unfolded** differential cross-sections



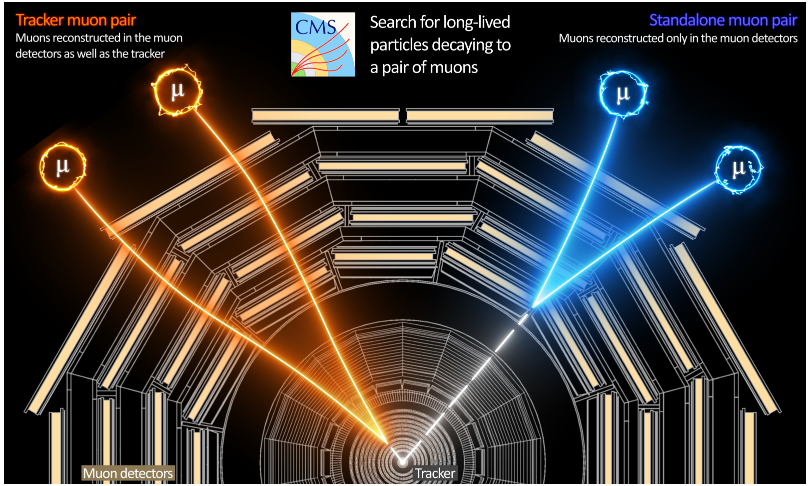
** correcting for detector effects

STDM-2022-17

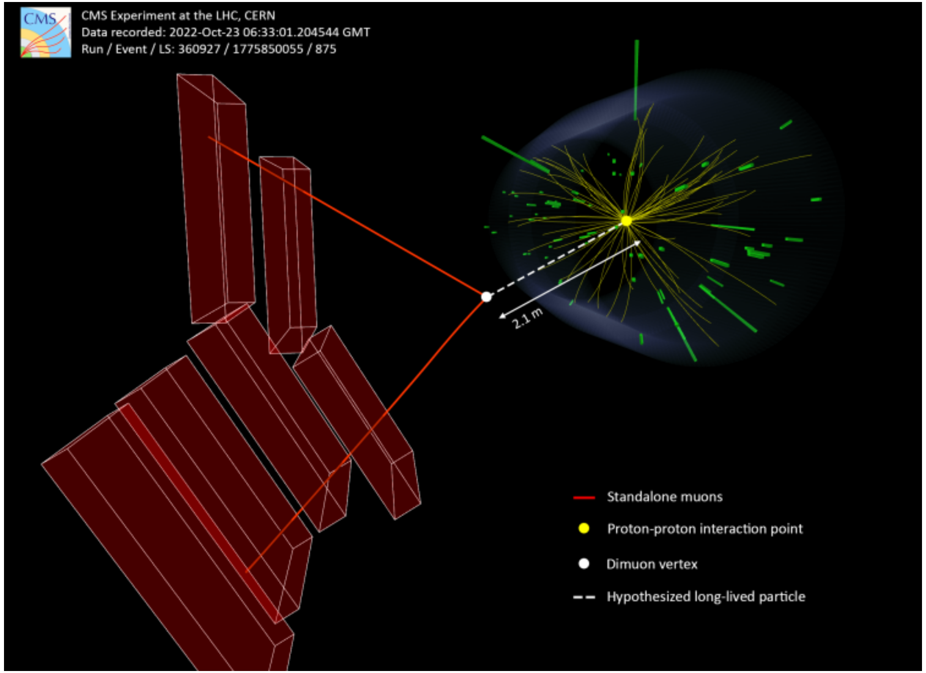
What physics comes out of LHC data?



Upper limits on the production of new particles

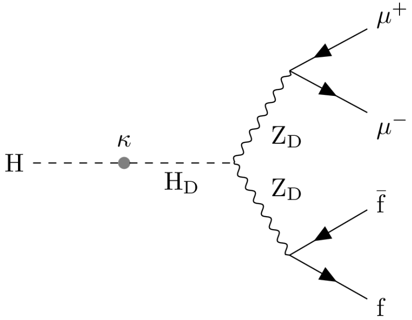


Long-lived particles

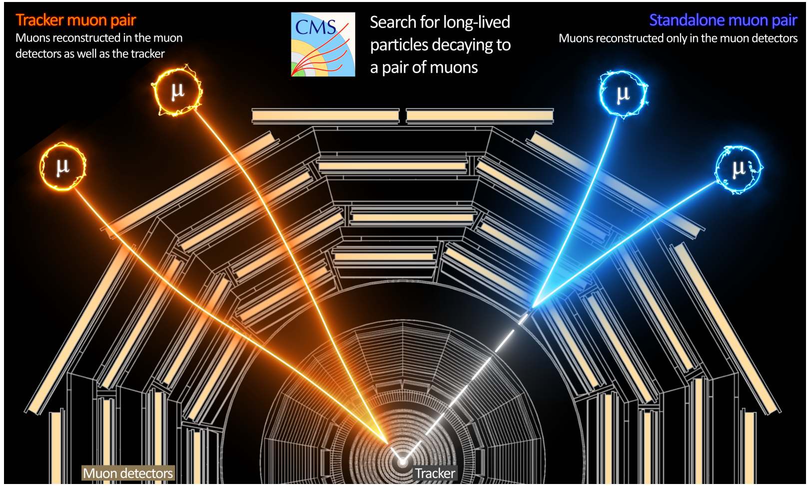


[long-lived-particles-light-lhc-run-3-data]

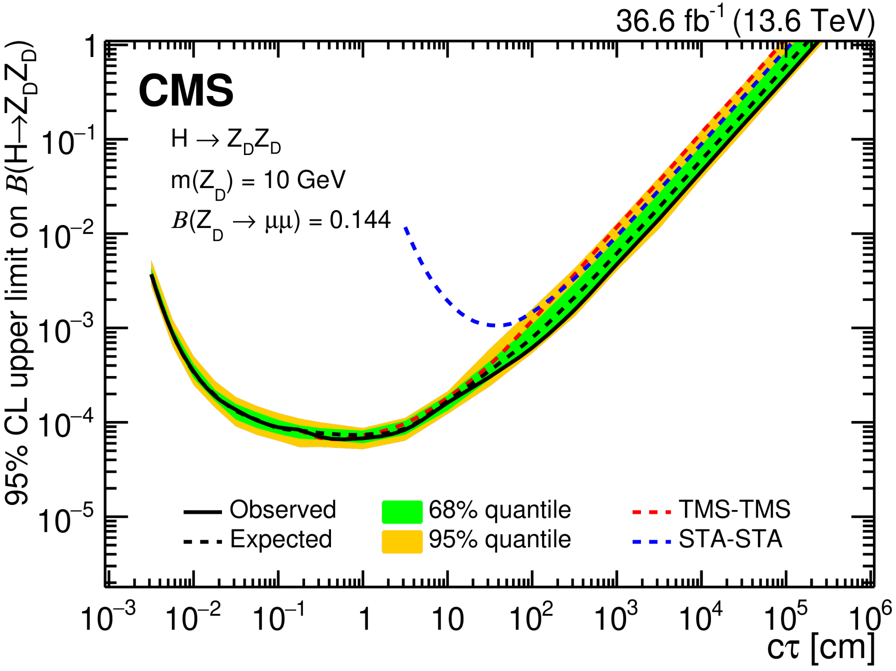
What physics comes out of LHC data?



Upper limits on the production of new particles

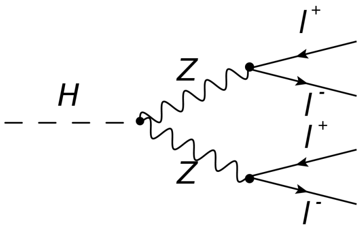


Long-lived particles

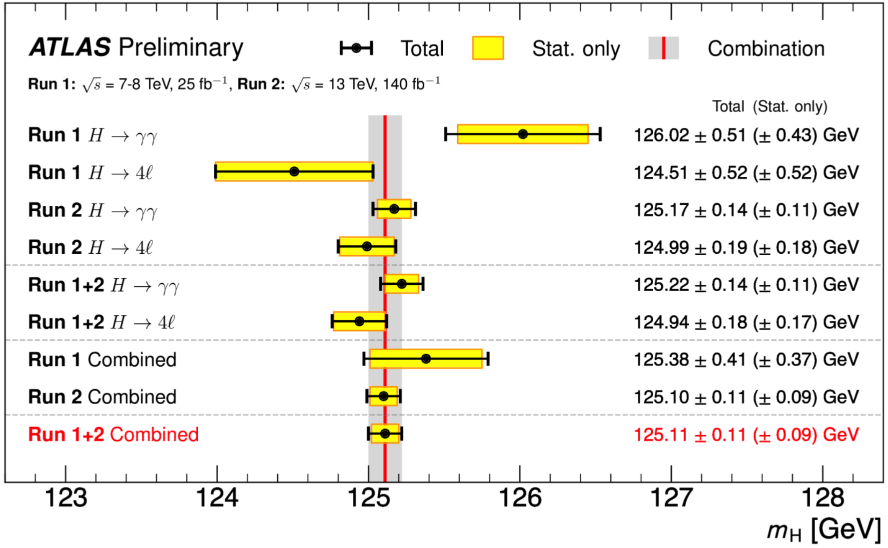


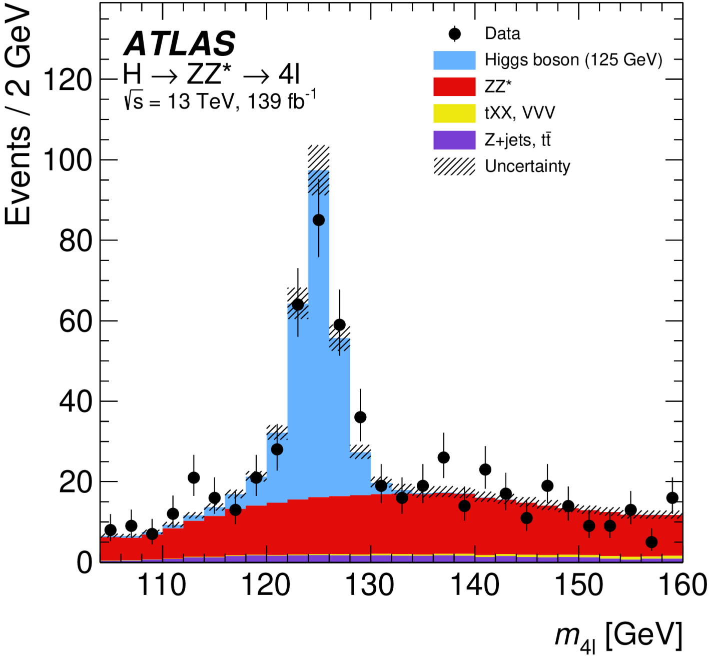
long-lived-particles-light-lhc-run-3-data

What physics comes out of LHC data?



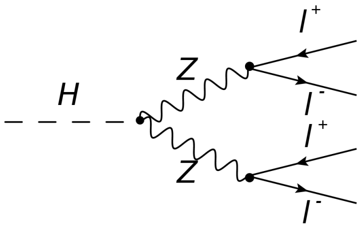
Fundamental properties of particles, e.g. Higgs boson mass

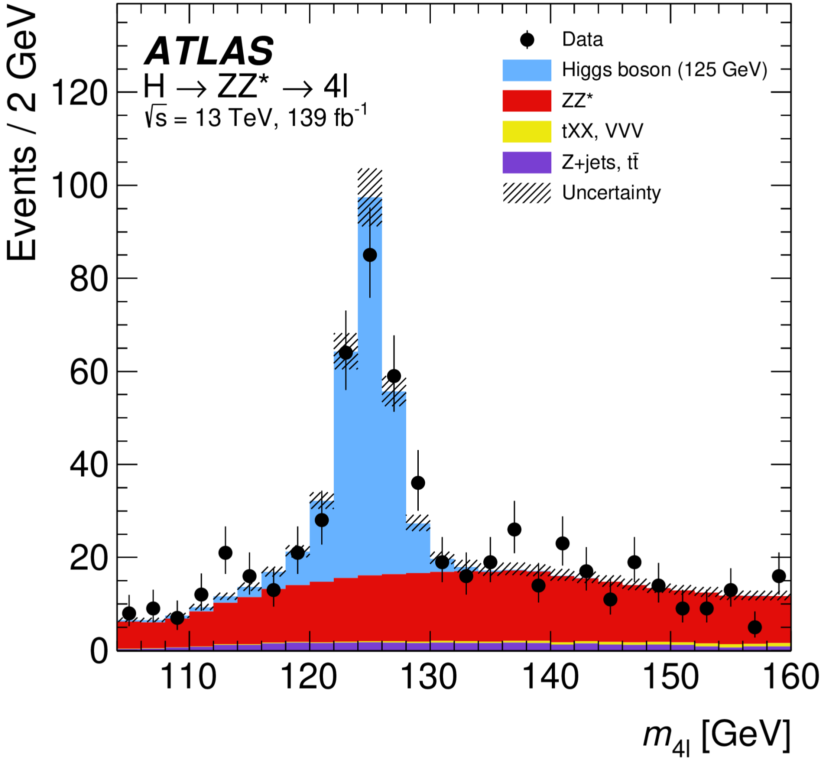




HIGG-2019-16

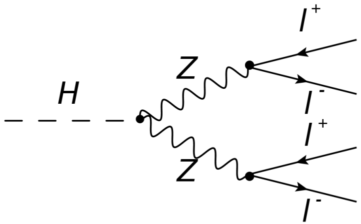
What's in a plot?

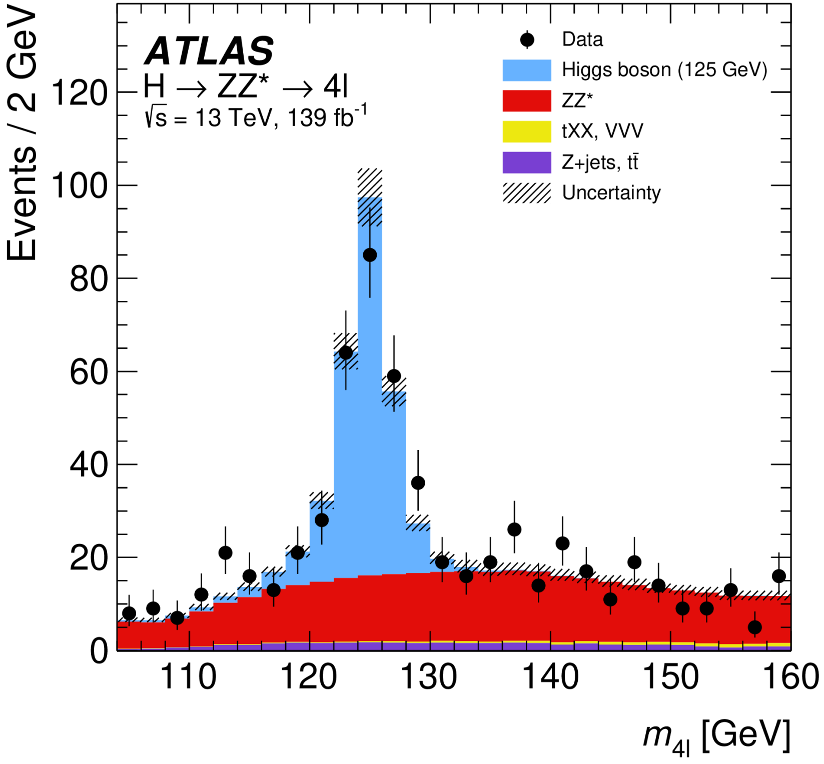




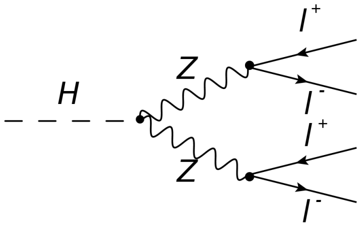
What's in a plot?

- Data
- Simulation





What’s in a plot?

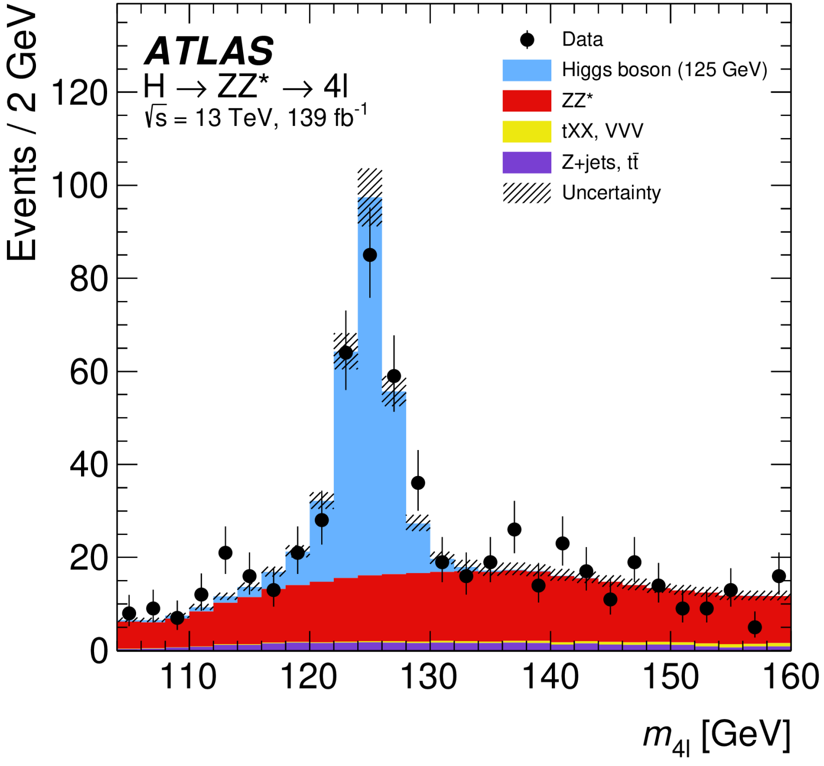


- Data

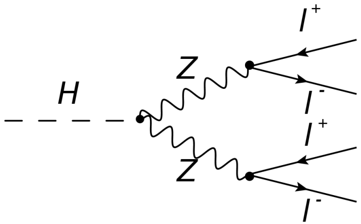
● Simulation

★ an invariant mass

- Lepton reconstruction & identification



What’s in a plot?



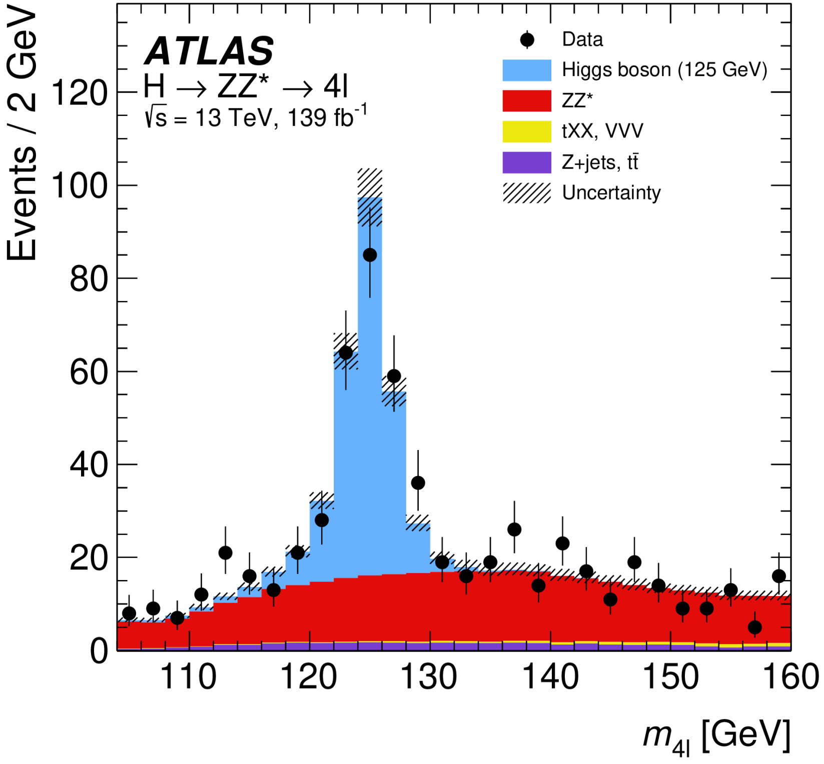
- Data
- Simulation

★ an invariant mass

- Lepton reconstruction & identification

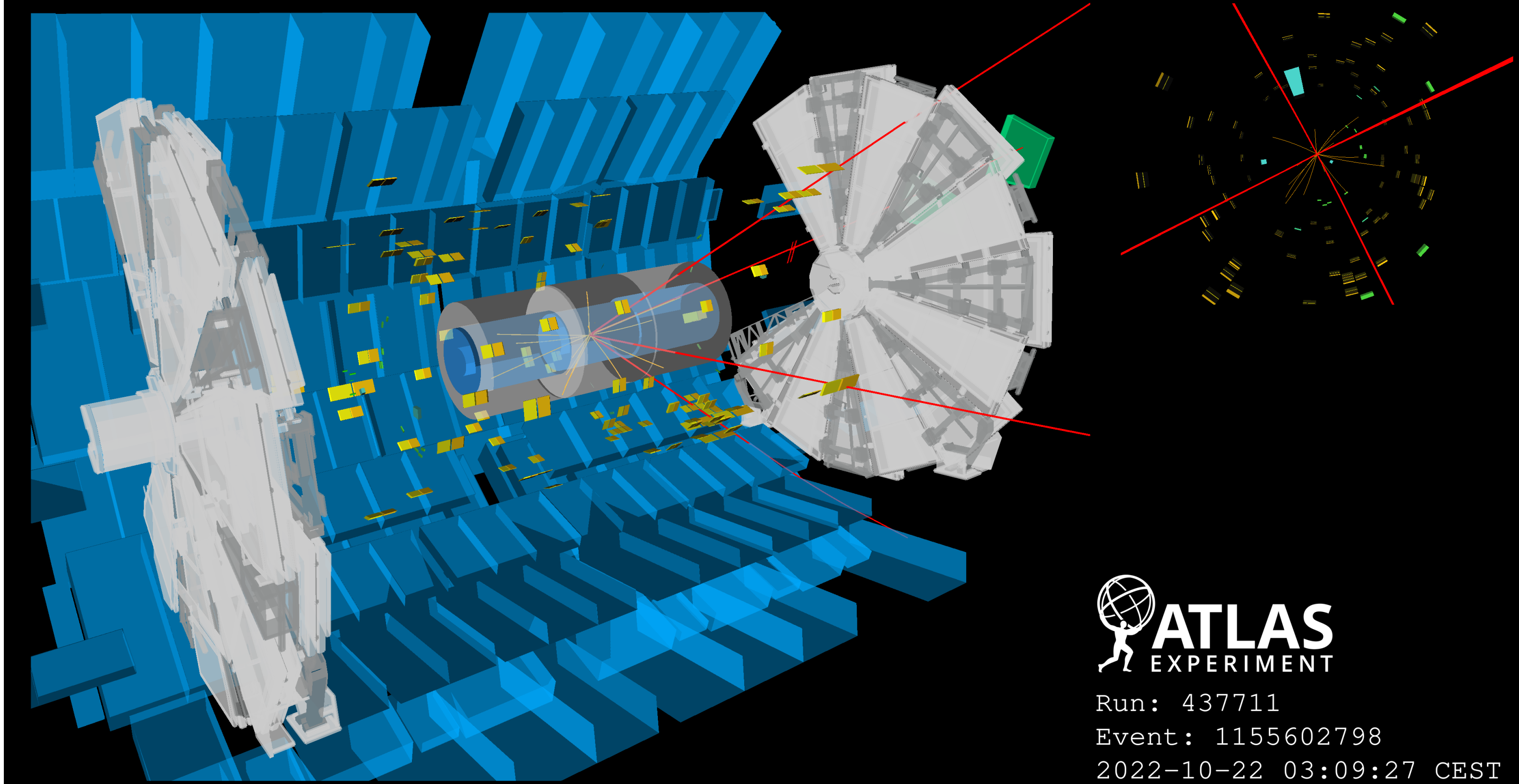
◉ Calibrations, detector alignment, pile-up, much more…

How do we get to this plot?



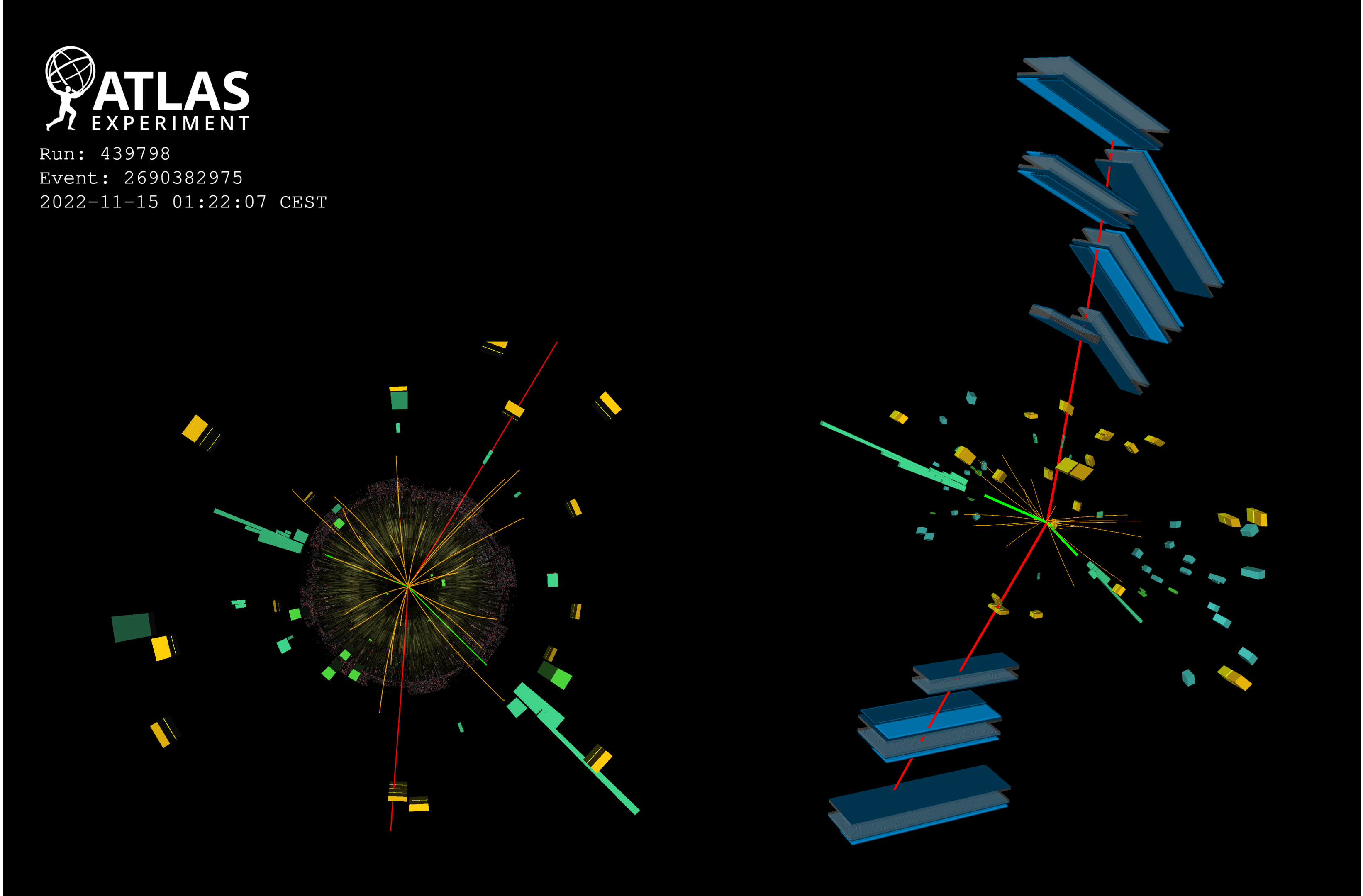
The data

A H to 4μ candidate



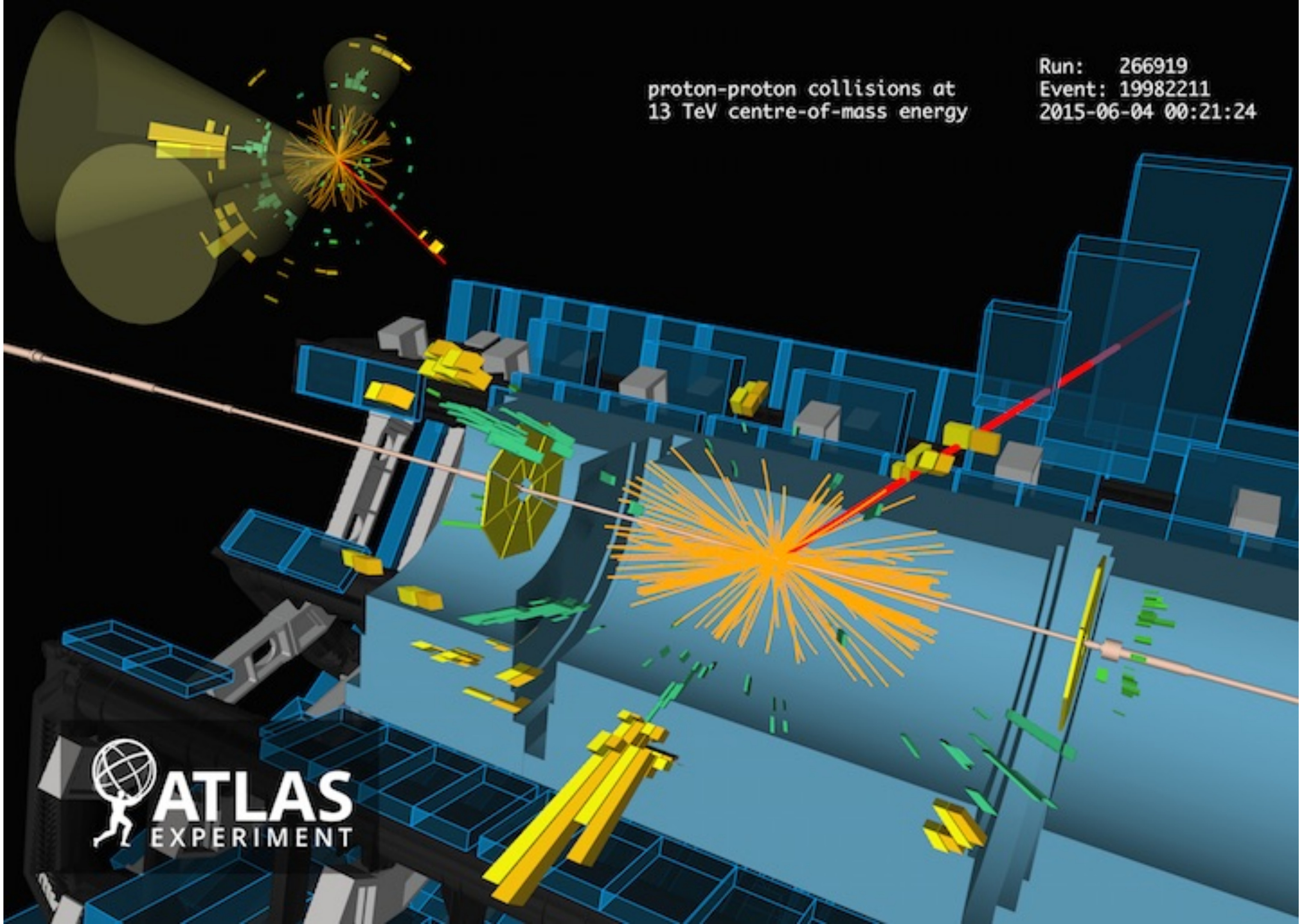
Run: 437711
Event: 1155602798
2022-10-22 03:09:27 CEST

A H to 2μ2e candidate

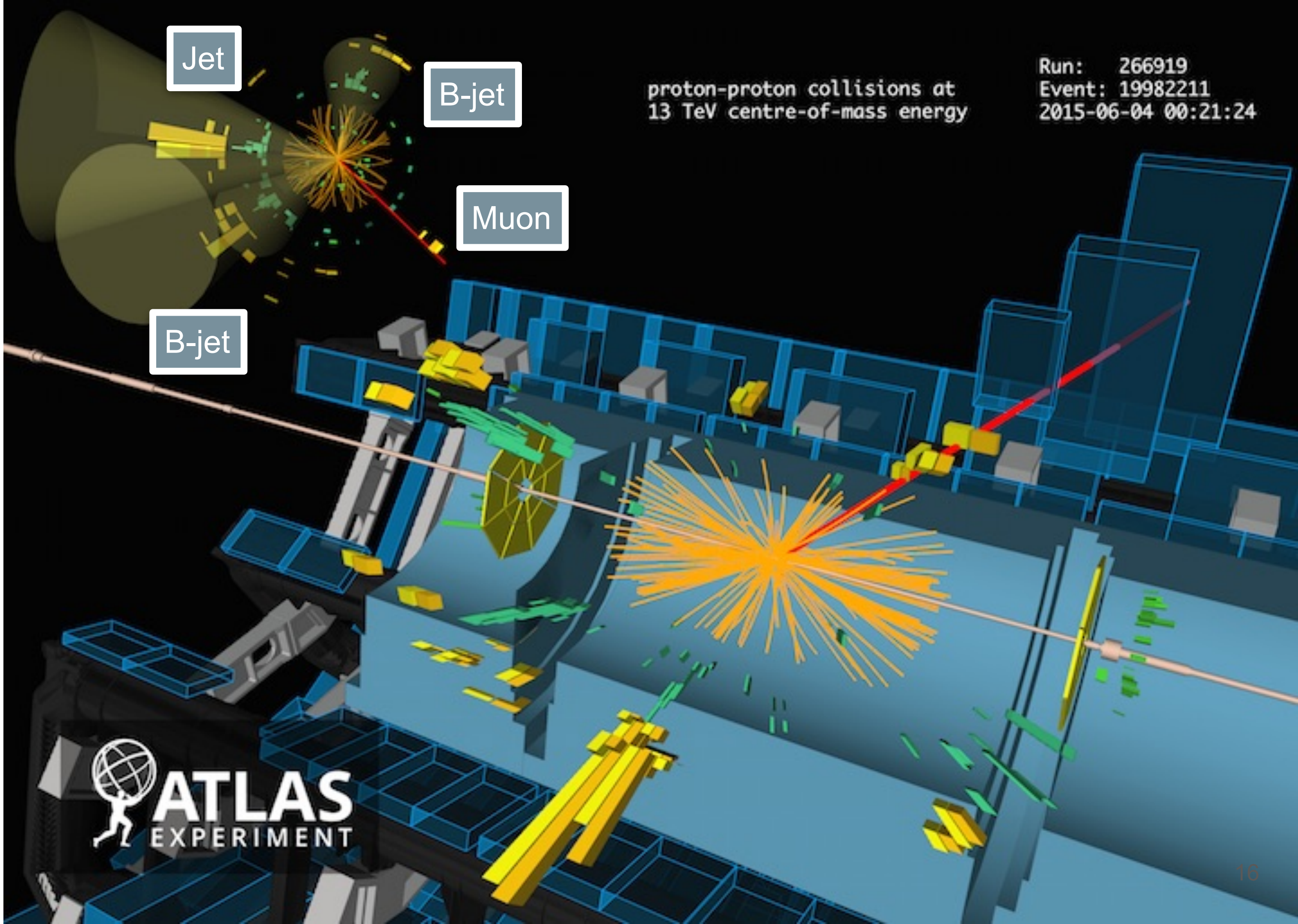


Run: 439798
Event: 2690382975
2022-11-15 01:22:07 CEST

A pair of top-quarks produced in ATLAS

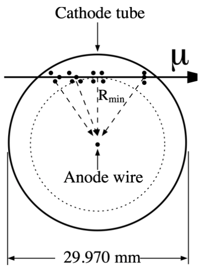


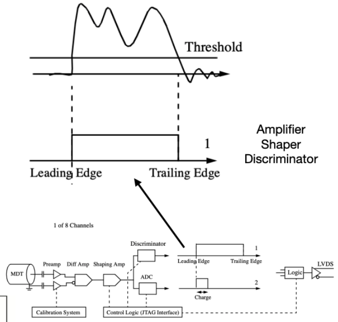
A pair of top-quarks produced in ATLAS

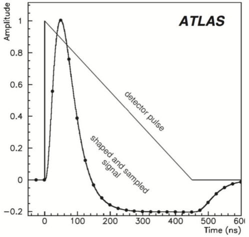


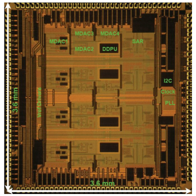
A simplified picture

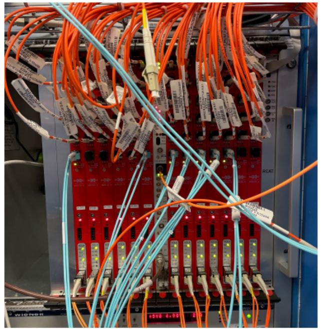
Data Acquisition

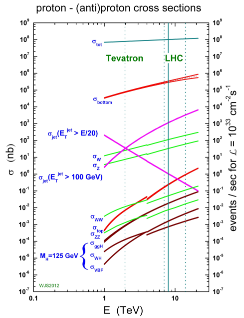




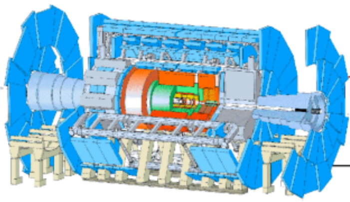


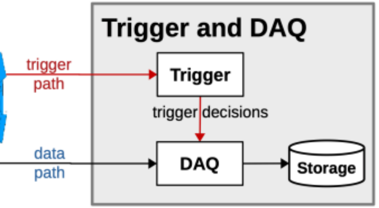






Trigger system





trigger path

data path

Trigger and DAQ

Trigger

trigger decisions

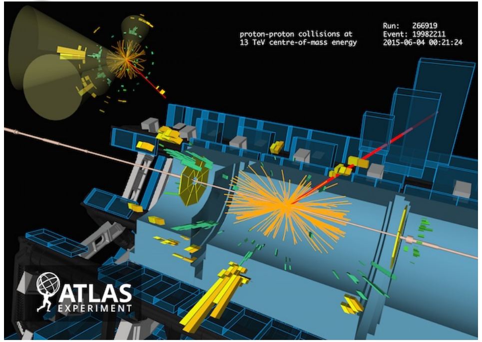
DAQ

Storage

Raw data

```
00000004 00000001 0000c89c aa1234aa 00003227 0000001c 04000000 00793c29 00000001 00000000
00000000 50753e27 0ab16f70 00097a2b 00000000 00033dac 00000063 920117d5 00000aa8 00000081
00000018 00020000 40000000 00000000 00000000 00000000 00000000 00000000 00000000 00020000
00000000 dd1234dd 0000002d 00000009 04000000 00210000 00000002 00000000 92011d7f 00000001
ee1234ee 00000009 03010000 00210000 00033dac 920117d5 00000aa8 00000081 00000000 2003e766
2013e282 201490d2 9c122017 ef322018 9d562023 dfa22039 c2224000 2040aa82 2041c3a2 204282b3
20489082 2057efb2 205a8616 2063cce2 2066aee2 2068a0c2 20768ff7 99522077 de72207b d8224000
00000000 00000000 00000002 00000015 00000001 d04326b2 dd1234dd 0000002d 00000009 04000000
00210001 00000002 00000000 92011d80 00000001 ee1234ee 00000009 03010000 00210001 00033dac
920117d5 00000aa8 00000081 00000000 2004af72 2010a3f2 20128ec2 2017c212 202083c2 9ec22025
c6c22026 a3022034 afb74000 20488602 2053c7c2 20548512 95829672 2063c2e2 e512ee02 20648fb2
2074a5e2 2075d5b2 207aa892 ad32207b ed72ee32 00000000 00000000 00000002 00000015 00000001
3de510d4 dd1234dd 00000031 00000009 04000000 00210002 00000002 00000000 92011d80 00000001
ee1234ee 00000009 03010000 00210002 00033dac 920117d5 00000aa8 00000081 00000000 20109ef2
2011ee42 efc22012 93222013 e2822014 97022017 e182201b e0222025 eaa22027 cab22028 80d3202a
84b22035 c5c2ccb2 2036ebc2 20389672 20508002 95a22051 d3172056 9ee22057 ef42205b cee2eca2
2060ad62 2061c4a2 2063ddb7 20649542 00000000 00000000 00000002 00000019 00000001 f631054a
dd1234dd 00000029 00000009 04000000 00210003 00000002 00000000 92011d80 00000001 ee1234ee
00000009 03010000 00210003 00033dac 920117d5 00000aa8 00000081 00000000 2027d422 203088a2
2031d692 20369542 2037ed92 20409c92 ace22044 9a822046 a9e22047 d3422048 8fb2204a 8a12204b
e172205b c4872060 8f822065 ea222067 c3f24000 00000000 00000000 00000002 00000011 00000001
aeaa0e15 dd1234dd 00000039 00000009 04000000 00210004 00000002 00000000 92011d80 00000001
ee1234ee 00000009 03010000 00210004 00033dac 920117d5 00000aa8 00000081 00000000 2006af12
2017eb47 201a8e76 2025e6d2 20268fa2 a292202b dff74000 2040a152 20469122 20529182 2060aea2
2061c4c2 d722d942 2063c5e2 2064a772 206aa152 206bc322 c7c22070 89d22072 8ad22073 c0b7800f
c187c1a7 c1f7c227 c287c2c7 c2e7c3a7 c3c7800f c3f7c417 c497c4d7 c547c5b7 c5e7c637 c657c677
c6b7c727 c767c7a7 00000000 00000000 00000002 00000021 00000001 a1feebf3 dd1234dd 0000002d
00000009 04000000 00210005 00000002 00000000 92011d80 00000001 ee1234ee 00000009 03010000
```

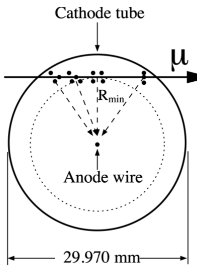
Reconstruction

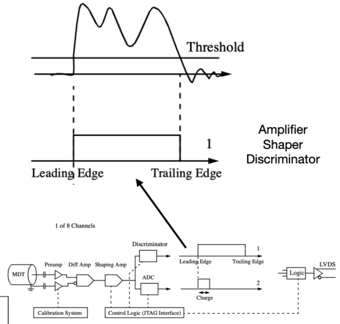


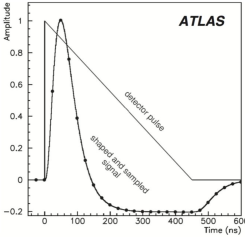
Taken from ATLAS Lectures on DAQ, Trigger, Data Processing Workflow: slides, slides, slides

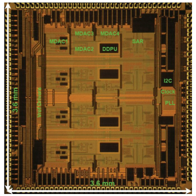
A simplified picture

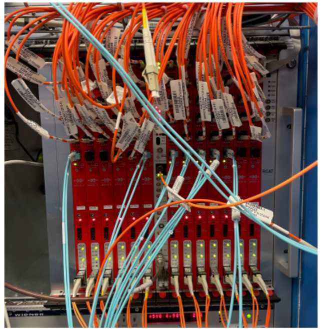
Data Acquisition

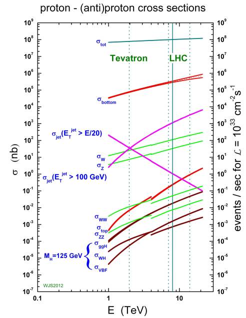




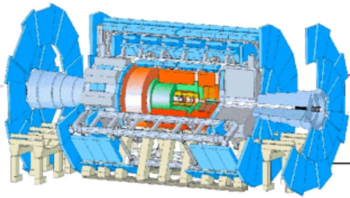


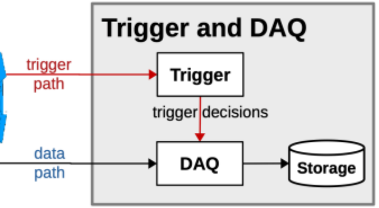






Trigger system





trigger path

data path

Trigger and DAQ

Trigger

trigger decisions

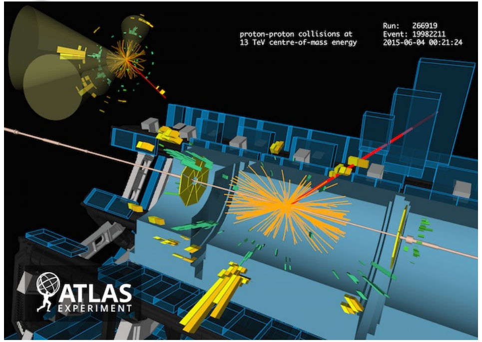
DAQ

Storage

Raw data

```
00000004 00000001 0000c89c aa1234aa 00003227 0000001c 04000000 00793c29 00000001 00000000
00000000 50753e27 0ab16f70 00097a2b 00000000 00033dac 00000063 920117d5 00000aa8 00000081
00000018 00020000 40000000 00000000 00000000 00000000 00000000 00000000 00000000 00020000
00000000 dd1234dd 0000002d 00000009 04000000 00210000 00000002 00000000 92011d7f 00000001
ee1234ee 00000009 03010000 00210000 00033dac 920117d5 00000aa8 00000081 00000000 2003e766
2013e282 201490d2 9c122017 ef322018 9d562023 dfa22039 c2224000 2040aa82 2041c3a2 204282b3
20489082 2057efb2 205a8616 2063cce2 2066aee2 2068a0c2 20768ff7 99522077 de72207b d8224000
00000000 00000000 00000002 00000015 00000001 d04326b2 dd1234dd 0000002d 00000009 04000000
00210001 00000002 00000000 92011d80 00000001 ee1234ee 00000009 03010000 00210001 00033dac
920117d5 00000aa8 00000081 00000000 2004af72 2010a3f2 20128ec2 2017c212 202083c2 9ec22025
c6c22026 a3022034 afb74000 20488602 2053c7c2 20548512 95829672 2063c2e2 e512ee02 20648fb2
2074a5e2 2075d5b2 207aa892 ad32207b ed72ee32 00000000 00000000 00000002 00000015 00000001
3de510d4 dd1234dd 00000031 00000009 04000000 00210002 00000002 00000000 92011d80 00000001
ee1234ee 00000009 03010000 00210002 00033dac 920117d5 00000aa8 00000081 00000000 20109ef2
2011ee42 efc22012 93222013 e2822014 97022017 e182201b e0222025 eaa22027 cab22028 80d3202a
84b22035 c5c2ccb2 2036ebc2 20389672 20508002 95a22051 d3172056 9ee22057 ef42205b cee2eca2
2060ad62 2061c4a2 2063ddb7 20649542 00000000 00000000 00000002 00000019 00000001 f631054a
dd1234dd 00000029 00000009 04000000 00210003 00000002 00000000 92011d80 00000001 ee1234ee
00000009 03010000 00210003 00033dac 920117d5 00000aa8 00000081 00000000 2027d422 203088a2
2031d692 20369542 2037ed92 20409c92 ace22044 9a822046 a9e22047 d3422048 8fb2204a 8a12204b
e172205b c4872060 8f822065 ea222067 c3f24000 00000000 00000000 00000002 00000011 00000001
aeaa0e15 dd1234dd 00000039 00000009 04000000 00210004 00000002 00000000 92011d80 00000001
ee1234ee 00000009 03010000 00210004 00033dac 920117d5 00000aa8 00000081 00000000 2006af12
2017eb47 201a8e76 2025e6d2 20268fa2 a292202b dff74000 2040a152 20469122 20529182 2060aea2
2061c4c2 d722d942 2063c5e2 2064a772 206aa152 206bc322 c7c22070 89d22072 8ad22073 c0b7800f
c187c1a7 c1f7c227 c287c2c7 c2e7c3a7 c3c7800f c3f7c417 c497c4d7 c547c5b7 c5e7c637 c657c677
c6b7c727 c767c7a7 00000000 00000000 00000002 00000021 00000001 a1feebf3 dd1234dd 0000002d
00000009 04000000 00210005 00000002 00000000 92011d80 00000001 ee1234ee 00000009 03010000
```

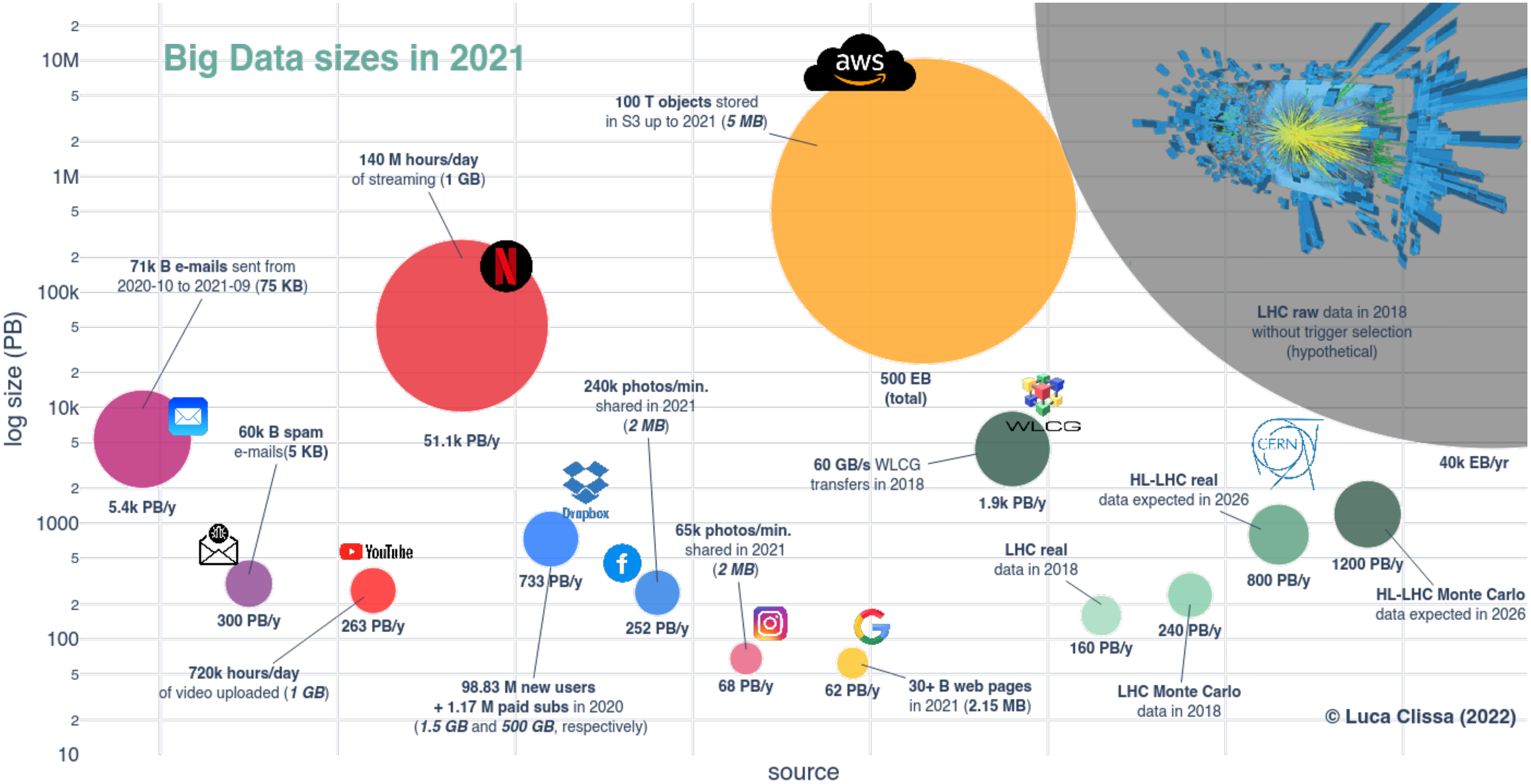
Reconstruction



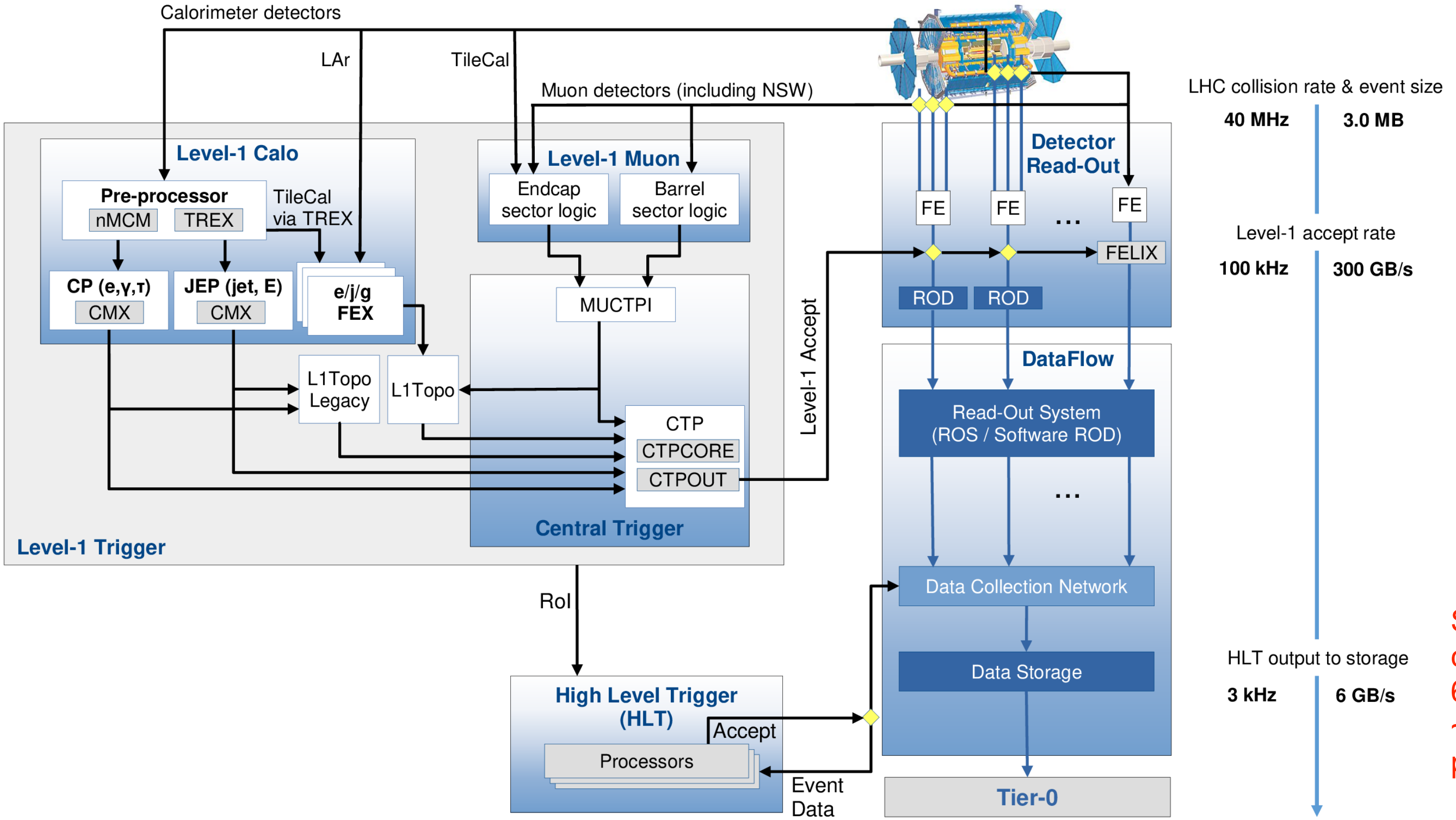
Taken from ATLAS Lectures on DAQ, Trigger, Data Processing Workflow: slides, slides, slides

Data volumes at the LHC

- Up to 40 million collisions per second (MHz)
- ~1 MB of data per collision
- 40 MHz * 1 MB = 40 TB/s
- 40 TB/s * 1e6 s/year = 40 000 EB/year



The ATLAS trigger system



Custom Hardware
Latency <2.5 us

Software-based, computing farm of 60k CPU cores, ~600 ms processing time

Poster by Marco Montella

THE ATLAS RUN 3 TRIGGER

MENU

MARCO MONTELLA [1]
ON BEHALF OF THE
ATLAS TRIGGER GROUP

OUR SIGNATURES

ELECTRONS	270 HZ
MUONS	290 HZ
TAUS	160 HZ
JETS & MET	630 HZ

ABOUT US

ATLAS runs a two-level triggering strategy:

- **L1** → Hardware-based, coarse reconstruction, total accepted rate is 90-100 kHz
- **HLT** → Software-based, reconstruction precision approaching offline reconstruction

The Trigger Menu is limited by:

- **L1 RATE** → Constrained by dead time
  Impacts the range of physics accessible
- **HLT CPU** → Limits the execution rate of high-precision reconstruction algorithms
- **T0 CPU** → Limits the data volume that can be reconstructed promptly for endpoint analysis

L1/HLT limitations scale with luminosity:

- **END OF FILL** → Enhance signatures limited by L1 rate and/or HLT rate & CPU

OUR STREAMS

MAIN
For prompt reconstruction
Rate limited by
L1, CPU & T0 Resources
1.7 KHZ

DELAYED Hadronic B-Physics
Delayed Processing
when T0 CPU available
900 HZ

LOW Calorie

TRIGGER-LEVEL ANALYSIS (TLA) Jets, Photons, b-tag
Reduced event content,
HLT Objects only.
Minimal burden on bandwidth
5+ KHZ

PARTIAL EVENT BUILDING (PEB) Jets, Photons, flavour-tag
Regional data around near
physics objects identified by trigger
0.5 KHZ

Specialty TRIGGERS

EMERGING JETS	10 HZ	DISAPPEARING TRACK	4 HZ
HIGHLY IONIZING TRACK	5 HZ	DISPLACED OBJECTS	40 HZ
ISOLATED TRACK	1 HZ	PARTIAL EVENT BUILDING	200 HZ

HEAVY IONS:
RUN 2 THRESHOLDS PRESERVED !

New trigger strategies: how to make the most out of the data

Bandwidth = event rate x event size

MAIN
For prompt reconstruction
Rate limited by
L1, CPU & T0 Resources
1.7 KHZ

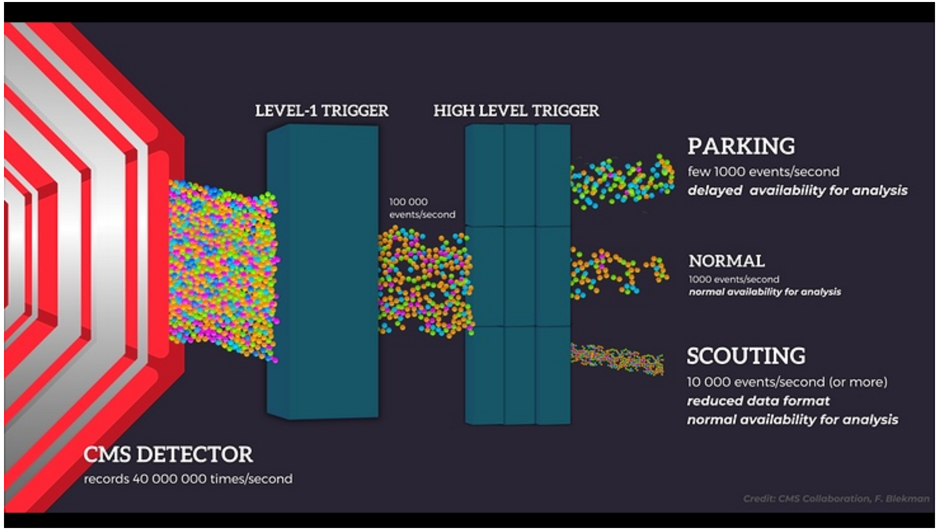
DELAYED Hadronic B-Physics
Delayed Processing
when T0 CPU available
900 HZ

TRIGGER-LEVEL ANALYSIS (TLA) Jets, Photons, b-tag
LOW Calorie
Reduced event content,
HLT Objects only.
Minimal burden on bandwidth
5+ KHZ

PARTIAL EVENT BUILDING (PEB) Jets, Photons, flavour-tag
Regional data around near
physics objects identified by trigger
0.5 KHZ

Delayed reconstruction (“Parking”)
- Park and wait.
- Process events when resources are available.

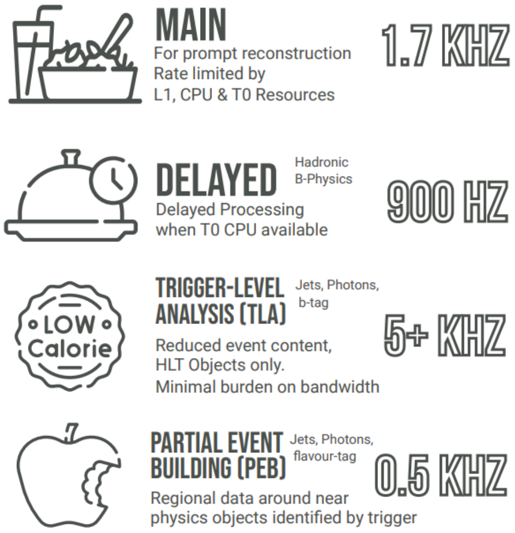
Do analysis in real-time (TLA, “Scouting”)
- Save only the trigger objects, smaller event size.



Credit: CMS Collaboration, F. Blekman

New trigger strategies: how to make the most out of the data

Bandwidth = event rate x event size



MAIN
For prompt reconstruction
Rate limited by
L1, CPU & T0 Resources
1.7 KHZ

DELAYED Hadronic B-Physics
Delayed Processing
when T0 CPU available
900 HZ

TRIGGER-LEVEL ANALYSIS (TLA) Jets, Photons, b-tag
Reduced event content,
HLT Objects only.
Minimal burden on bandwidth
5+ KHZ

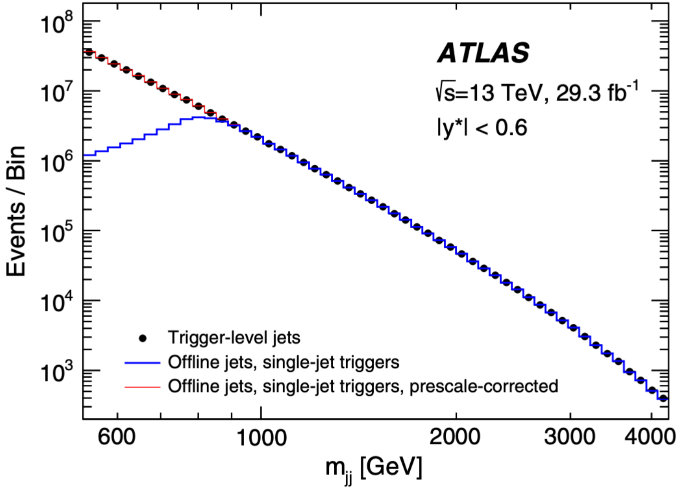
PARTIAL EVENT BUILDING (PEB) Jets, Photons, flavour-tag
Regional data around near
physics objects identified by trigger
0.5 KHZ

Delayed reconstruction (“Parking”)

- Park and wait.
- Process events when resources are available.

Do analysis in real-time (TLA, “Scouting”)

- Save only the trigger objects, smaller event size.

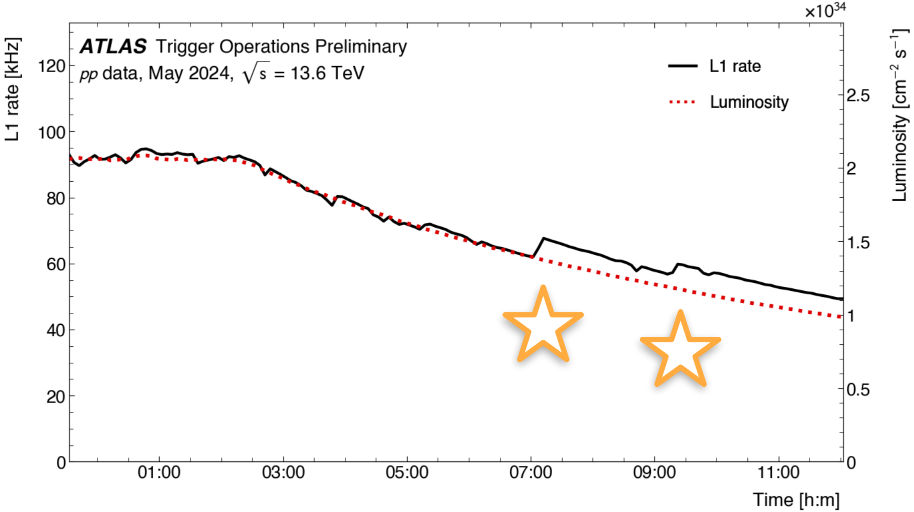


Trigger Operations

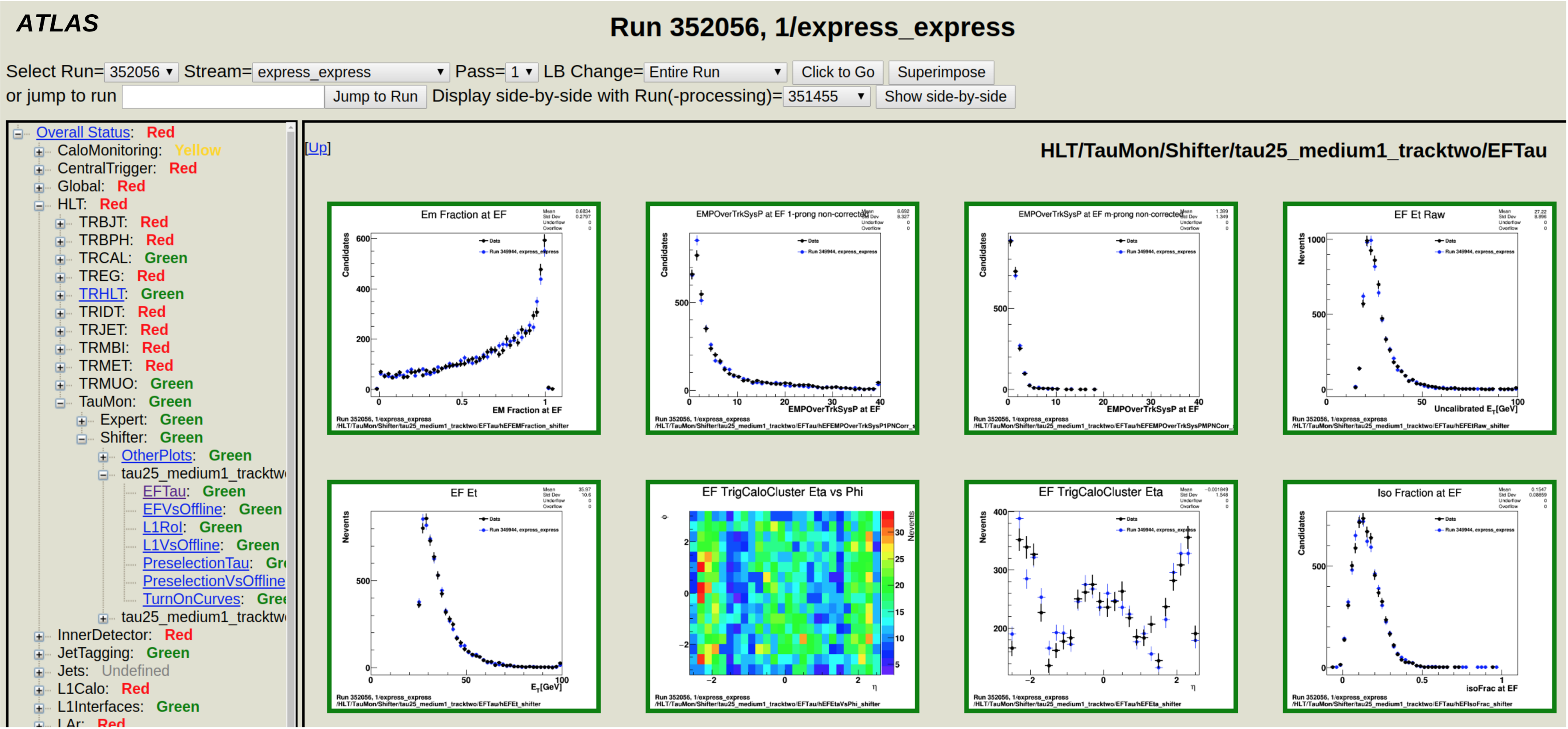
- Real-time constraints during any data-taking run:
  - At L1: maximum rate of 100 kHz (detector readout capability)
  - At HLT: bandwidth + CPU resources of the HLT farm
  - Offline prompt processing capabilities

Trigger menu and prescales are continually adjusted to optimise the available resources.

- Prescale factors are applied to L1 and HLT triggers
  - Triggers can be executed for a fraction of events and to be enabled/disabled; can be changed during data-taking to adapt to decreasing luminosity
- End-of-fill (EOF) strategies:
  - Enable/unprescale additional/resource-heavy triggers when luminosity declines



Data Quality shifts



ATLAS

Run 352056, 1/express_express

- Overall Status: Red
  - CaloMonitoring: Yellow
  - CentralTrigger: Red
  - Global: Red
  - HLT: Red
    - TRBJT: Red
    - TRBPH: Red
    - TRCAL: Green
    - TREG: Red
    - TRHLT: Green
    - TRIDT: Red
    - TRJET: Red
    - TRMBI: Red
    - TRMET: Red
    - TRMUO: Green
    - TauMon: Green
      - Expert: Green
      - Shifter: Green
        - OtherPlots: Green
        - tau25_medium1_tracktw
          - EFTau: Green
          - EFVsOffline: Green
          - L1RoI: Green
          - L1VsOffline: Green
          - PreselectionTau: Gr
          - PreselectionVsOffline
          - TurnOnCurves: Gre
        - tau25_medium1_tracktw
  - InnerDetector: Red
  - JetTagging: Green
  - Jets: Undefined
  - L1Calo: Red
  - L1Interfaces: Green
  - LAr: Red

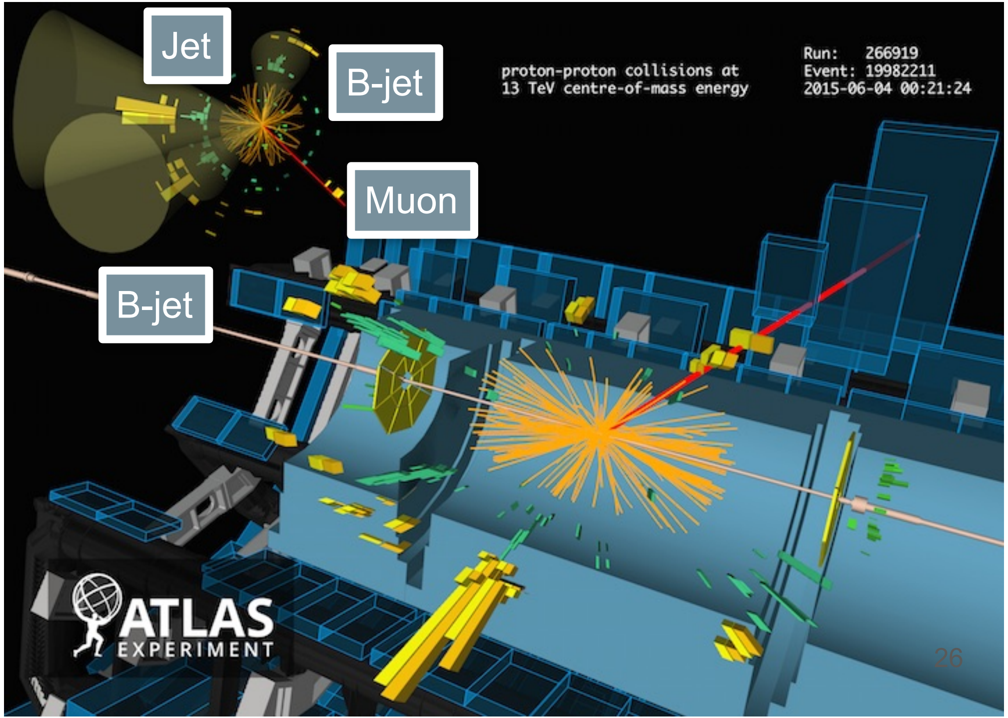
[Up]

HLT/TauMon/Shifter/tau25_medium1_tracktwo/EFTau

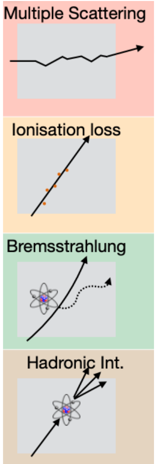
Event *Reconstruction*

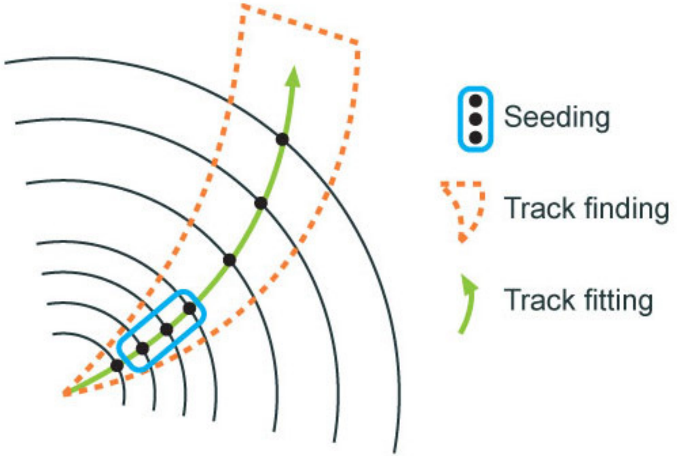
- Going from raw data to analysis objects.
- **Important:** data and simulation pass through the same reconstruction algorithms.

- Raw data reconstructed into:
    - **Tracks**
    - **Calorimeter deposits**
- Which are then reconstructed into "physics" objects:
    - Jets, electrons, muons, taus
    - Photons, missing transverse energy



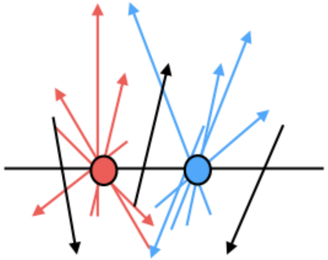
From *hits* to *physics*: **tracking**



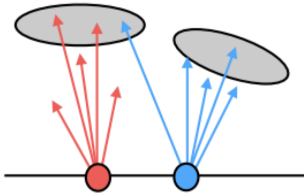


- Efficiently and precisely reconstructing charged particles:
  - Under a non-uniform magnetic field (equations of motion have to be solved numerically)
  - With hundreds to thousands of particles per event.
  - With tight CPU timing constraints.

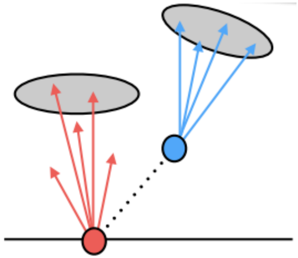
- Used in almost every element of reconstruction.



Vertex reconstruction

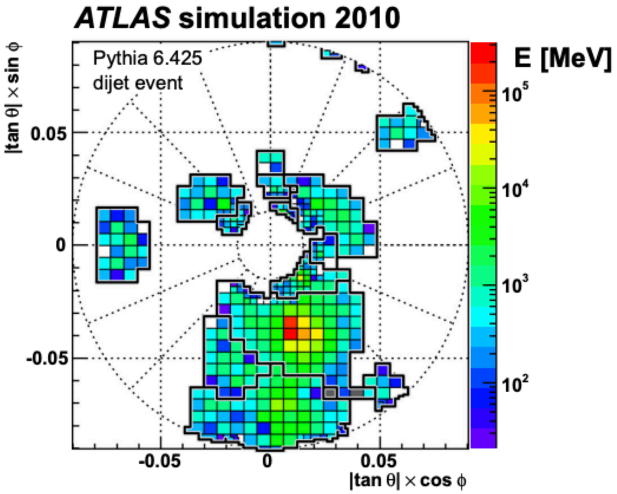


Pile up removal



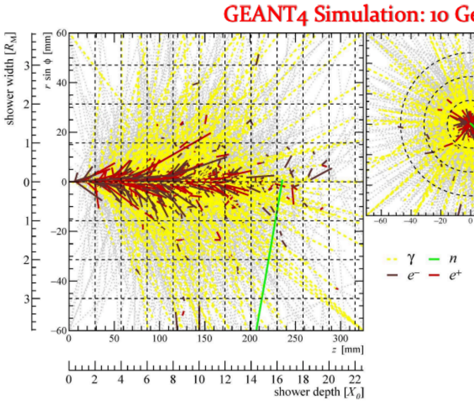
Jet flavour tagging

From *hits* to *physics*: **clustering**

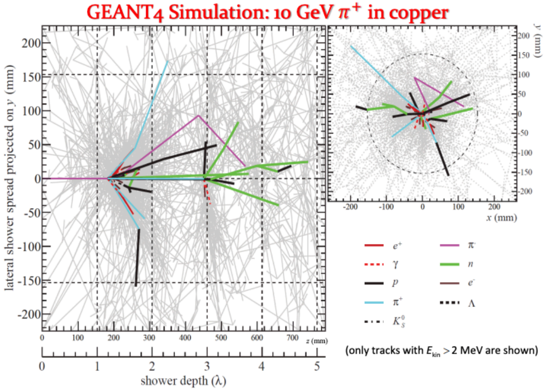


(c) All clustered cells

- Three-dimensional topological clustering (*topo-clustering*) of individual calorimeter cell signals.
- Algorithm sensitive to the nature of the shower producing the cluster signal:
  - EM showers are more compact, smaller intrinsic fluctuations
  - HADdronic shower have larger shower-by-shower fluctuations and are located deeper in the calorimeter.



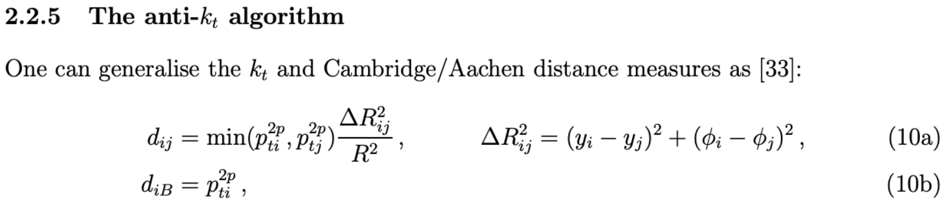
Interplay of bremsstrahlung and e^+e^- pair production up to shower maximum → ionizations ($e^\pm$), Compton scattering (γ), photo-effect (γ)



paper

From *hits* to *physics*: **clustering**

- Calorimeter topoclusters are one of the ingredients to **jet clustering**

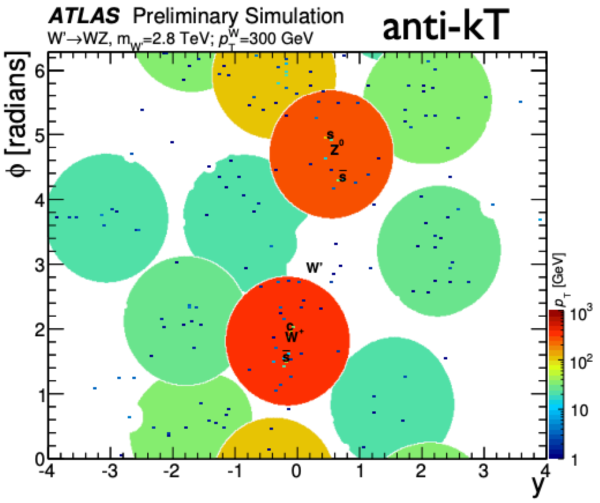


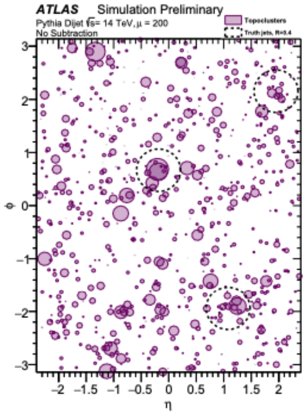
2.2.5 The anti-k_t algorithm

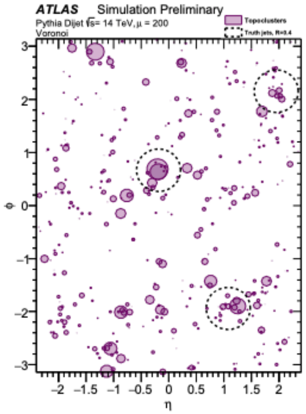
One can generalise the k_t and Cambridge/Aachen distance measures as [33]:

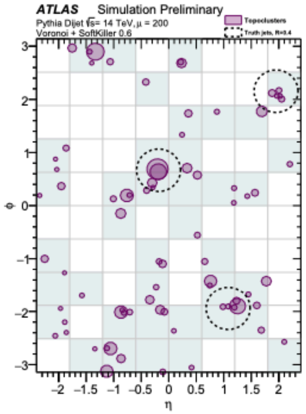
$$d_{ij} = \min(p_{ti}^{2p}, p_{tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2}\,, \qquad \Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2\,, \tag{10a}$$

$$d_{iB} = p_{ti}^{2p}\,, \tag{10b}$$





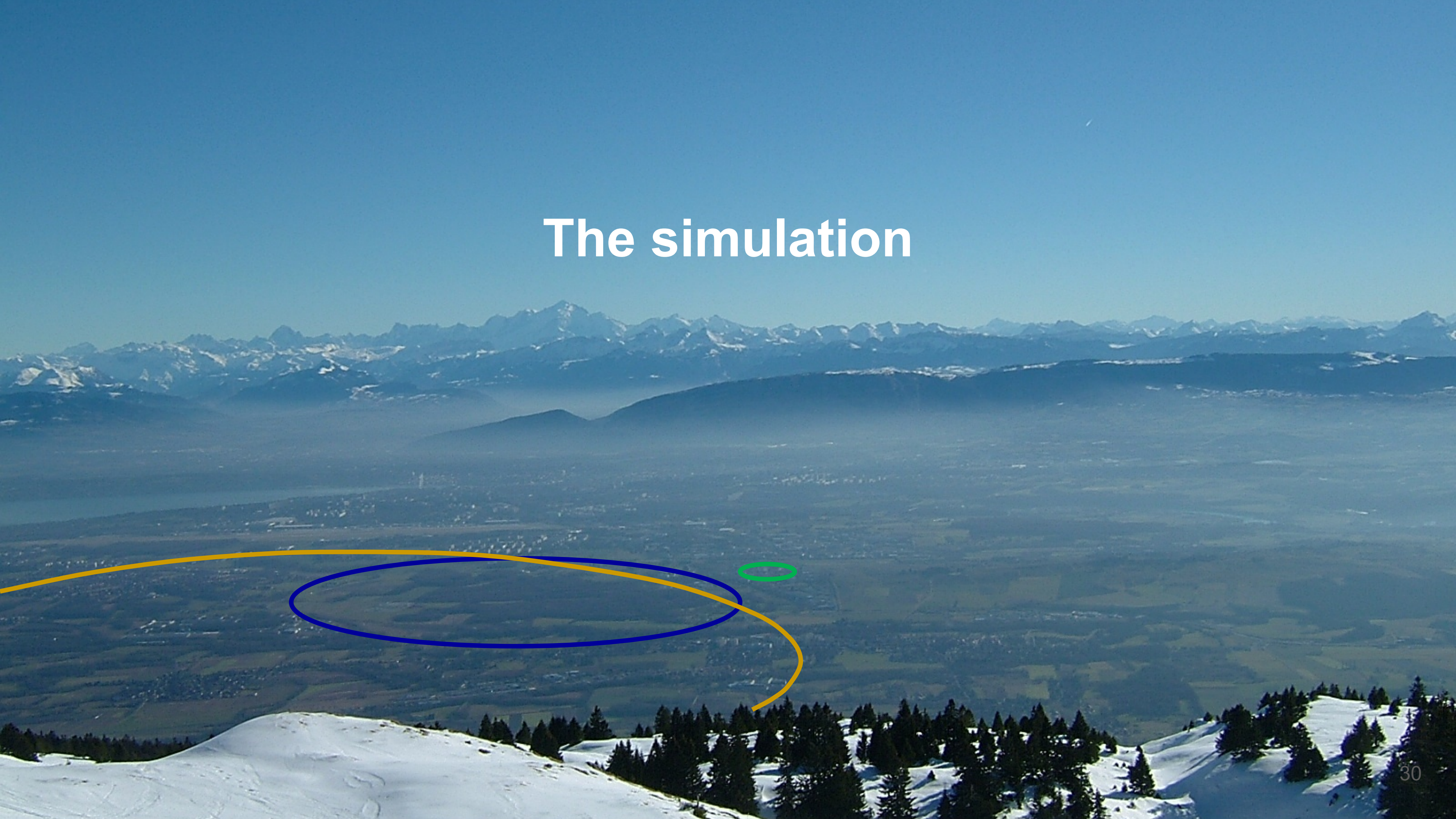




Different techniques to handle pile-up:

- At constituent-level, e.g. subtracting low-pt constituents
- At jet-level, subtracting energy density x jet area, cut on jet timing, etc…

The simulation



Physics analyses at the LHC

The power of factorisation of physics at different energy scales.

Inclusive* cross-section for the production of the final state X in the collision of hadrons h_1,h_2

Differential partonic cross-section

$$\sigma_{h_1,h_2\to X} = \sum_{a,b\in\{q,g\}} \int dx_a \int dx_b f_a^{h_1}(x_a,\mu_F^2) f_b^{h_2}(x_b,\mu_F^2) \int d\Phi_{ab\to X} \frac{d\hat{\sigma}_{ab}(\Phi_{ab\to X},\mu_F^2)}{d\Phi_{ab\to X}}$$

Partons a,b in the PDF with momentum fraction x_a,x_b

Parton distribution functions (PDFs) at factorisation scale μ_F^2

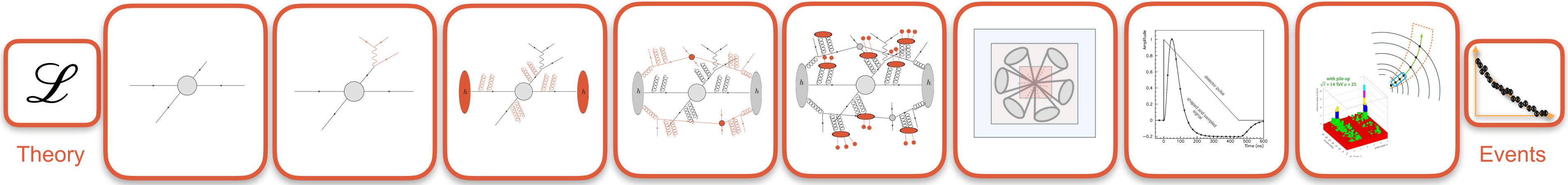
* inclusive since no specific kinematic configuration or particle multiplicity is specified

From collisions to physics results

Theory and data can be linked through **precise simulations** of:

- The hard scatter interaction
- The parton showering and hadronization
- The detector itself

The goal is to have a twin set of collision data to compare to the real collisions.



Theory

Hard scattering

Decay

Parton shower

Underlying event

Hadronisation

Simulation

Digitisation

Reconstruction

Events

Detector simulation (I)

No. of steps ∝ simulation time

Simulation of the passage of particles through the detectors.

Particle ionisation in the trackers

Energy deposition in the calorimeters

Intermediate particle decays, radiation and scattering…

Typically done using the *GEANT4* software, taking into account:

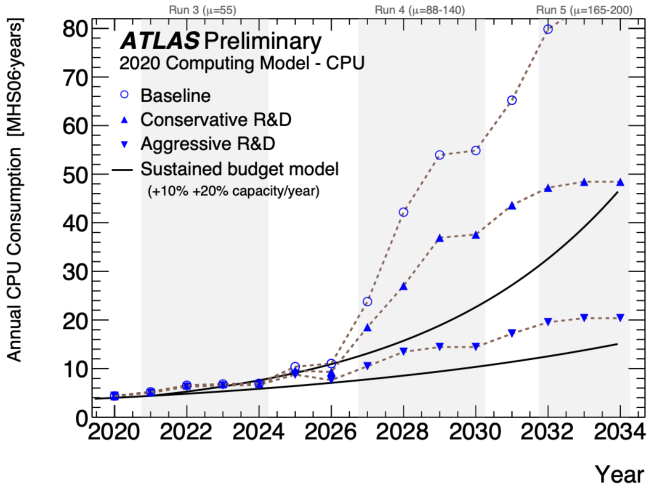
Dense hit content in the inner trackers.

Electromagnetic and hadronic shower development.

Effect of the magnetic fields.

Complex geometry with multiple sub-detectors, support structures, cooling pipes, cables, …

A problem…



One of the most computational expensive steps in the entire Monte Carlo generation chain: ~40% of ATLAS resources in Run 2.

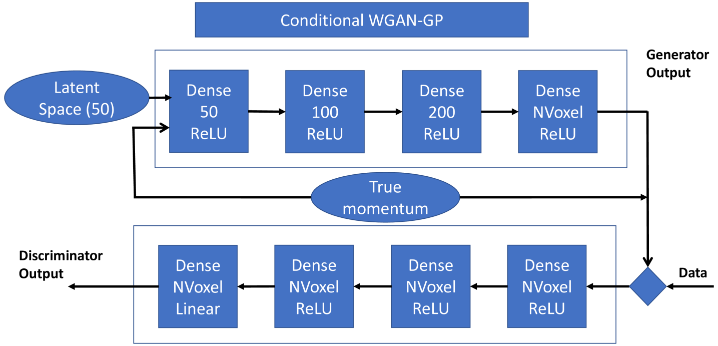
Detector simulation (II)

Can also be done using a *fast* simulation:

Parameterising the calorimeter response to single particles (smearing 4-vectors)

New: improved methods using machine learning!

e.g. Generative Adversarial Networks (GANs)



Conditional WGAN-GP

Latent Space (50)

Dense 50 ReLU

Dense 100 ReLU

Dense 200 ReLU

Dense NVoxel ReLU

Generator Output

True momentum

Data

Dense NVoxel ReLU

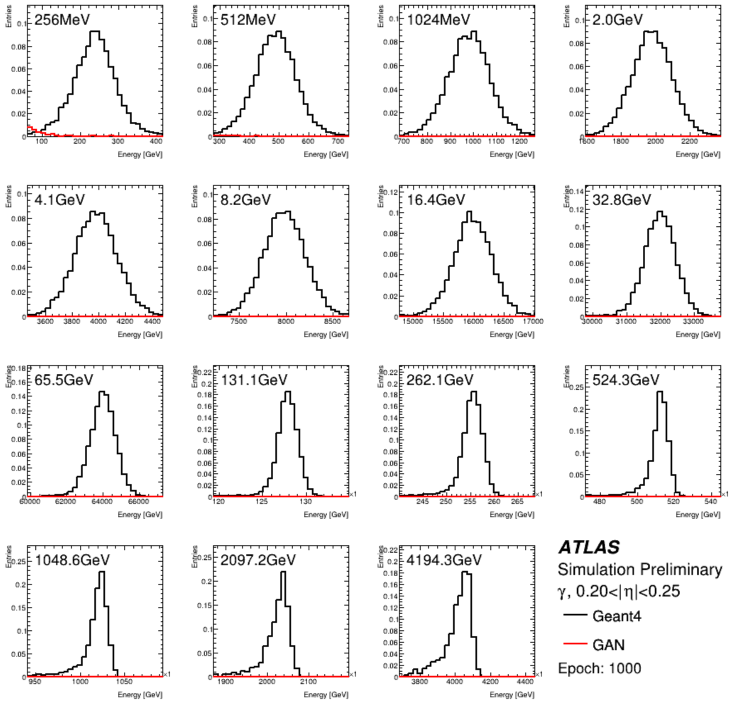
Dense NVoxel ReLU

Dense NVoxel ReLU

Dense NVoxel Linear

Discriminator Output

Calorimeter energy response to photons

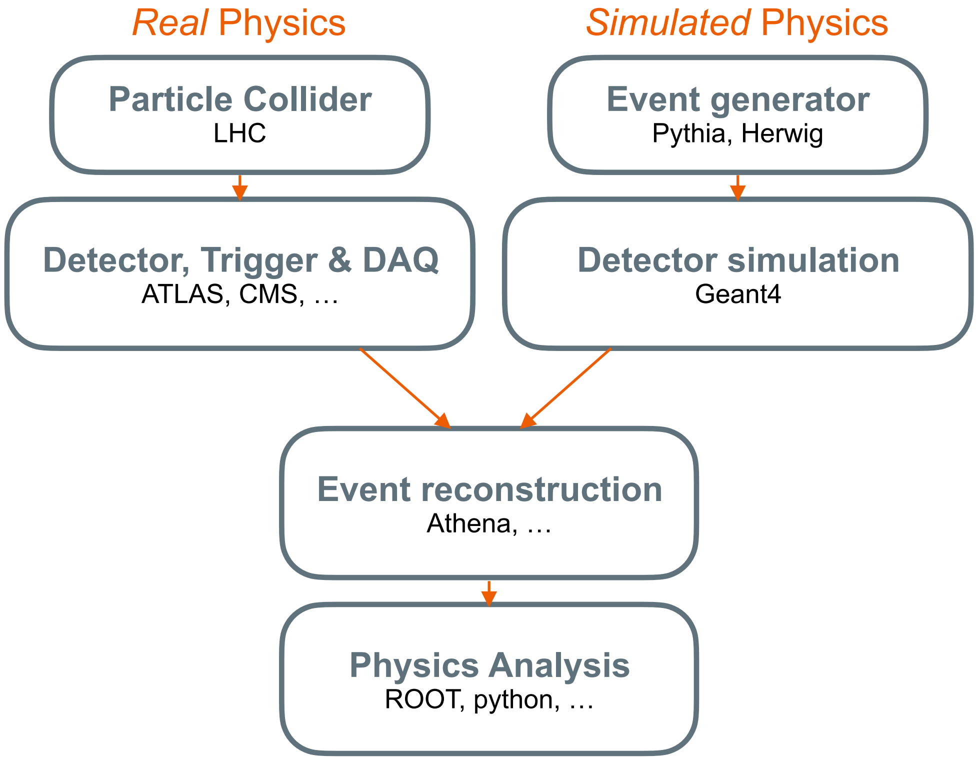


- Fast simulation provides speed gains of O(500) for calorimeter simulation!

FastCaloGAN

Bringing the two together

What’s in a plot?



Real Physics

Simulated Physics

Particle Collider
LHC

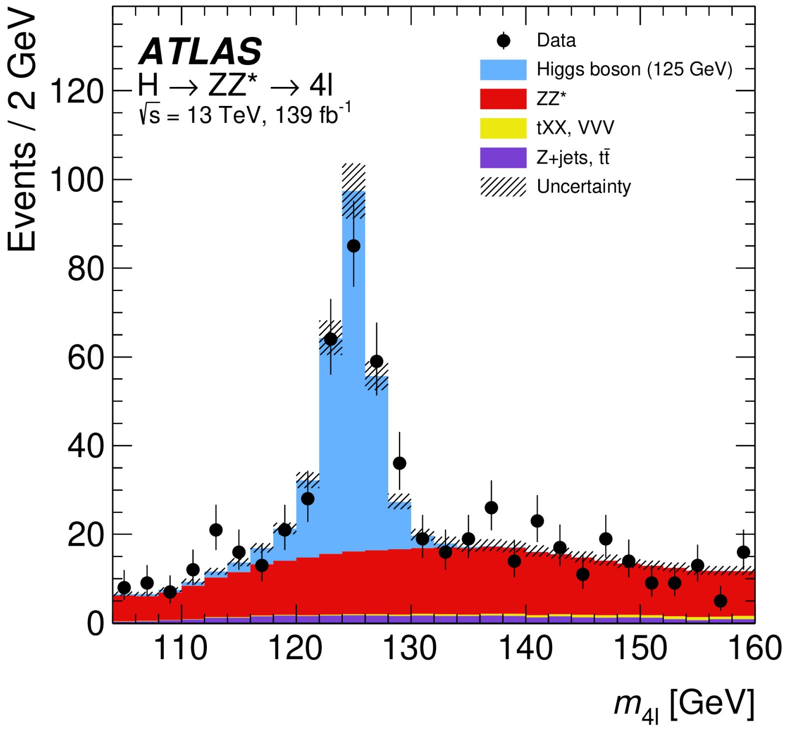
Event generator
Pythia, Herwig

Detector, Trigger & DAQ
ATLAS, CMS, …

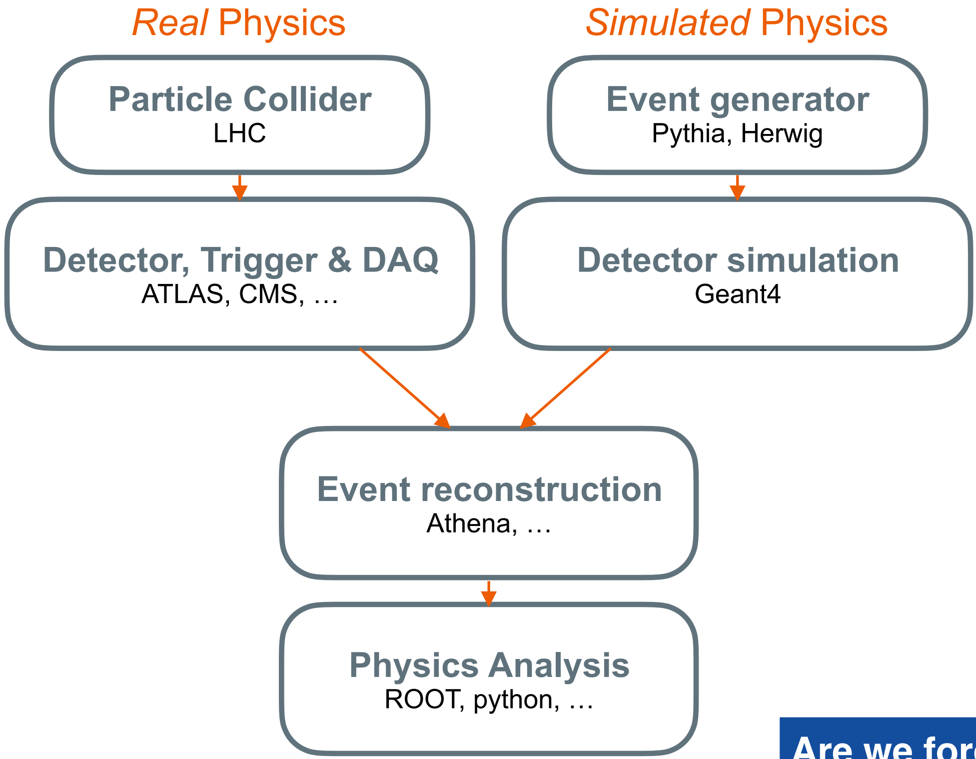
Detector simulation
Geant4

Event reconstruction
Athena, …

Physics Analysis
ROOT, python, …



What’s in a plot?



Real Physics

Particle Collider
LHC

Detector, Trigger & DAQ
ATLAS, CMS, …

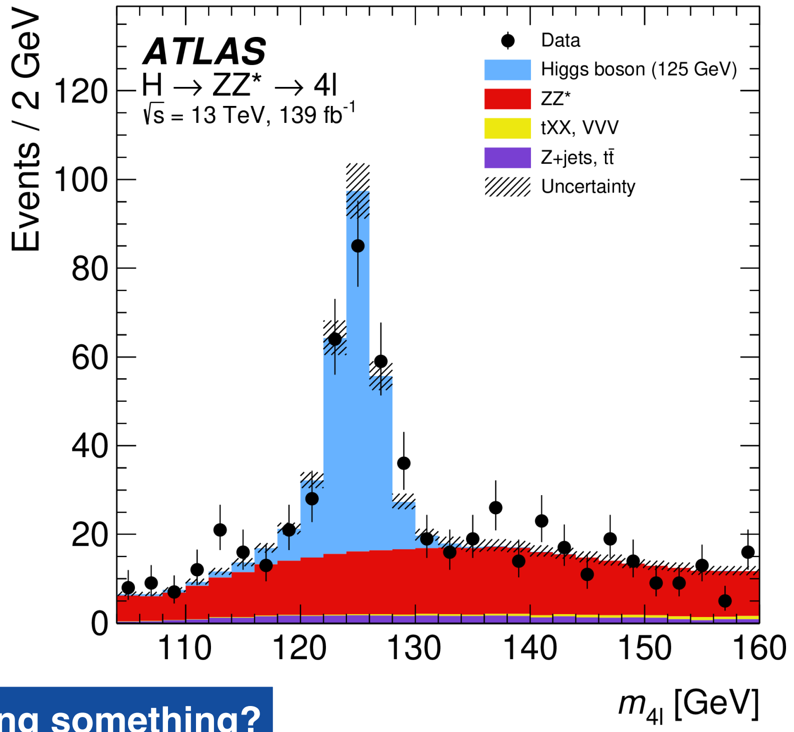
Simulated Physics

Event generator
Pythia, Herwig

Detector simulation
Geant4

Event reconstruction
Athena, …

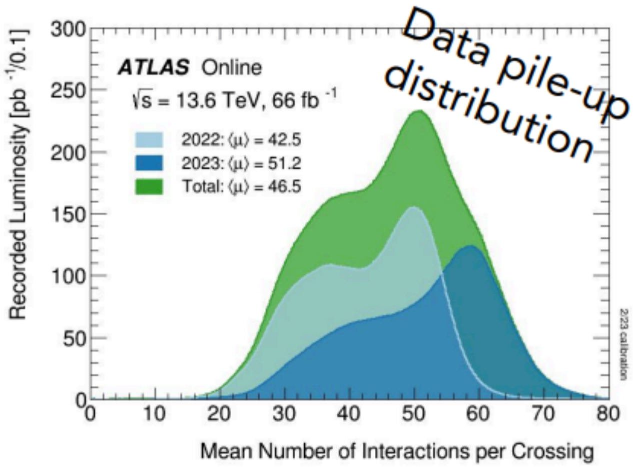
Physics Analysis
ROOT, python, …



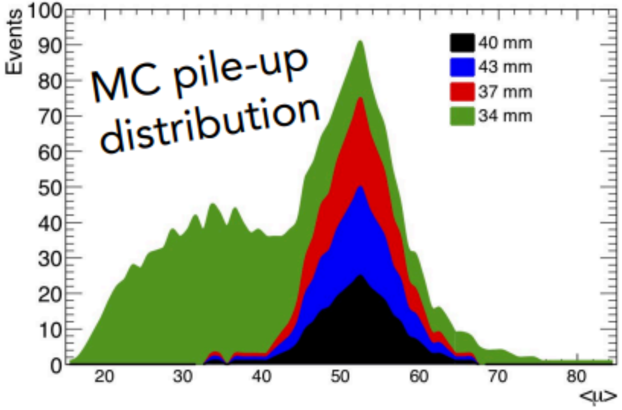
Are we forgetting something?

Overlaying pile-up collisions

- Pile-up comes “for free” in our data. It needs to be modelled in our Monte Carlo as well, for a fair comparison to data.
- In ATLAS Run 3, this is done by ***MC pile-up overlay***:
    - *Minimum-bias* and *single neutrino* events are generated using Pythia8.
    - GEANT4 is run on these events to simulate detector response.
    - Digitisation is then run on a combination of these events, including shifts in time to reproduce in-time and out-of-time pile-up.
    - The digitisation output of a pile-up event is then overlaid with the hard-scatter MC.
- **Proton bunch structure and luminosity based on that of real data.**



Data pile-up distribution

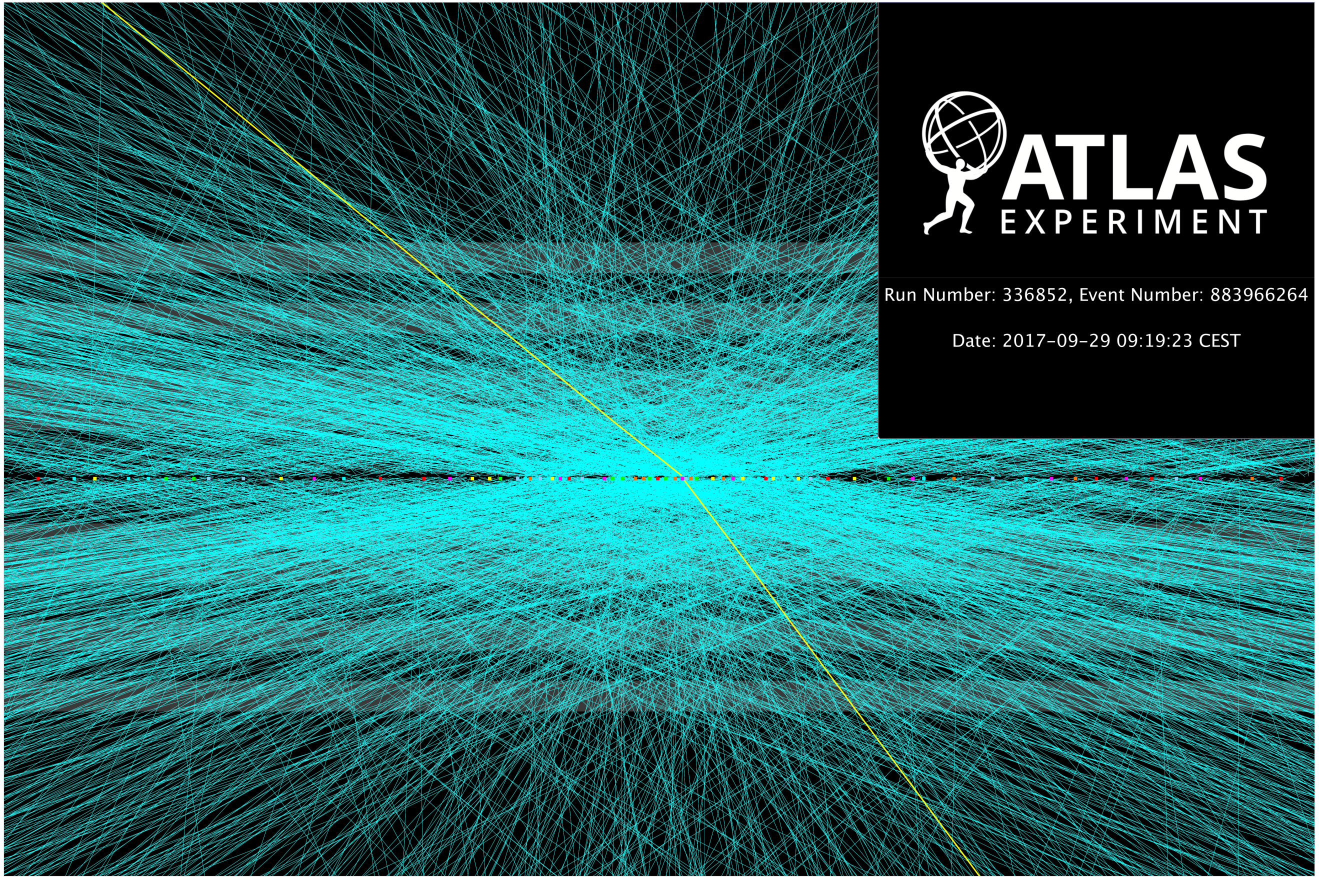


MC pile-up distribution

$< \mu >$ = mean number of interactions per crossing

Taking from ATLAS Lectures on Data Processing Workflow: slides

Challenge: pile-up



A Z boson decays to 2 muons in an event with 65 (!) additional pile-up collisions.

$\sqrt{\hat{s}} = 13$ TeV

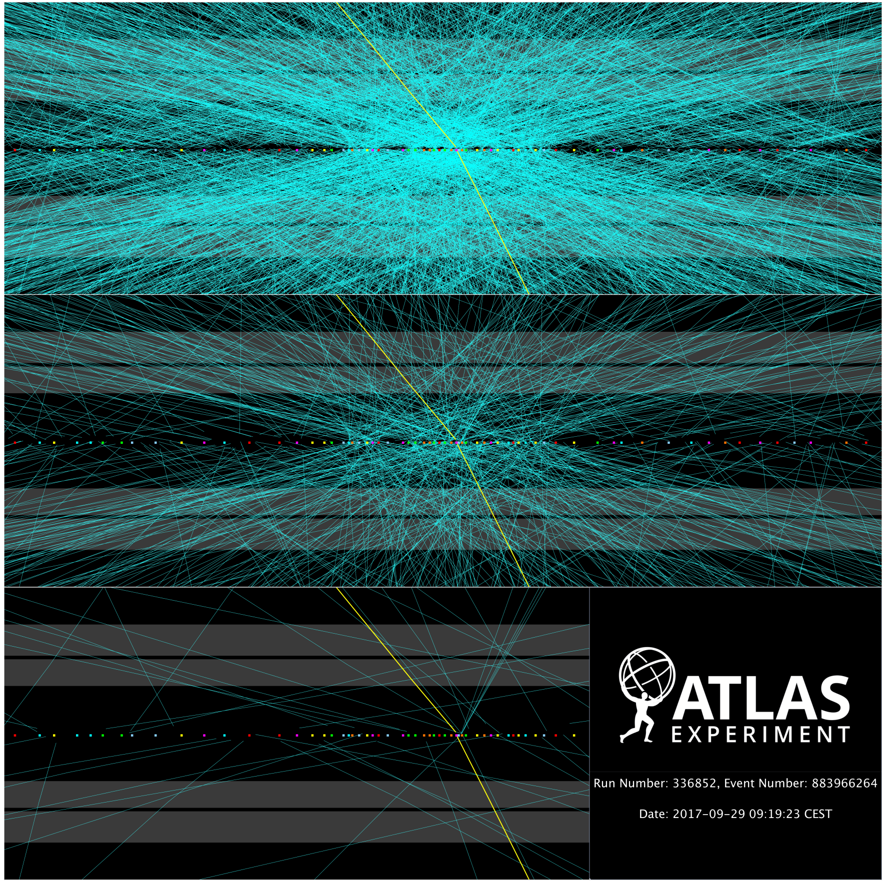
$<\mu>$ = mean number of interactions per crossing

Challenge: pile-up

Track p_T > 100 MeV

Track p_T > 1 GeV

Track p_T > 5 GeV



A Z boson decays to 2 muons in an event with 65 (!) additional pile-up collisions.

$\sqrt{\hat{s}} = 13$ TeV

$< \mu >$ = mean number of interactions per crossing

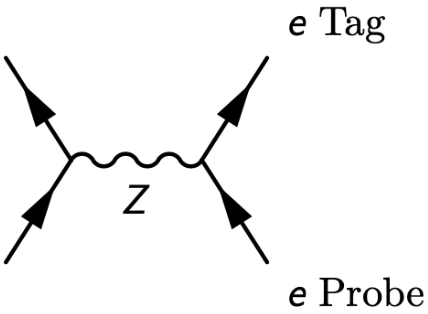
What else?

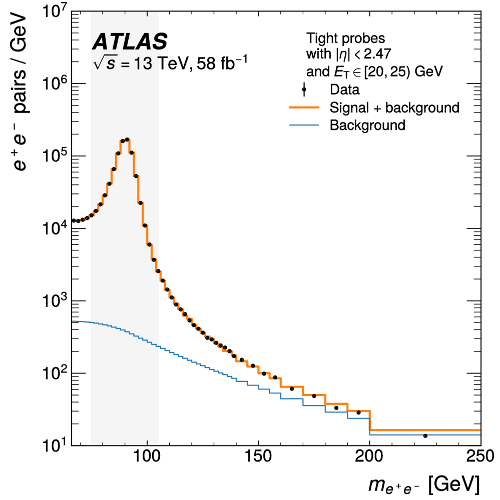
MC WARNING

- The *real world* need to be reflected in the Monte Carlo simulation.
  - E.g. a section of the calorimeter readout dies and cannot be repaired until the detector is opened during an LHC shutdown.
    - If this impacts x% of the data, we need a representative slice of the problem in our MC.
      - But x is usually hard to know until we know how much data we will collect until the shutdown. At that point we need to *reprocess* the MC.
- Even then, MC often doesn't describe the data.
  - Improving MC (e.g. via tuning of input parameters) is an ever on-going (and time consuming task.
  - Another way to deal with inaccurate modeling is to **correct / calibrate** the MC.
    - We can correct:
      - An efficiency (event-level correction).
      - An object's energy scale or resolution (object-level correction).

Example #1: tag & probe method

- **How efficiently do we identify an electron?**





Use a Standard Model candle like $Z \rightarrow ee$

- ✓ We know the Z decays to one electron and a positron
- ✓ We know the Z invariant mass very well
- ✓ We "tag" one electron and study the "probe" electron

For example, the identification efficiency can be calculated as:

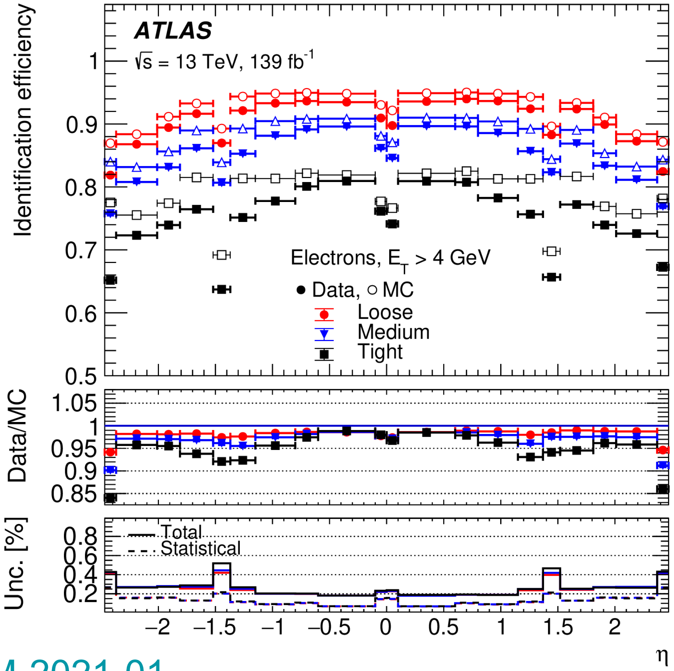
$$\epsilon = \frac{\text{probe is identified}}{\text{all probes}}$$

Only the tag electron is used to select events.

EGAM-2021-01

Example #1: tag & probe method

- **How efficiently do we identify an electron?**

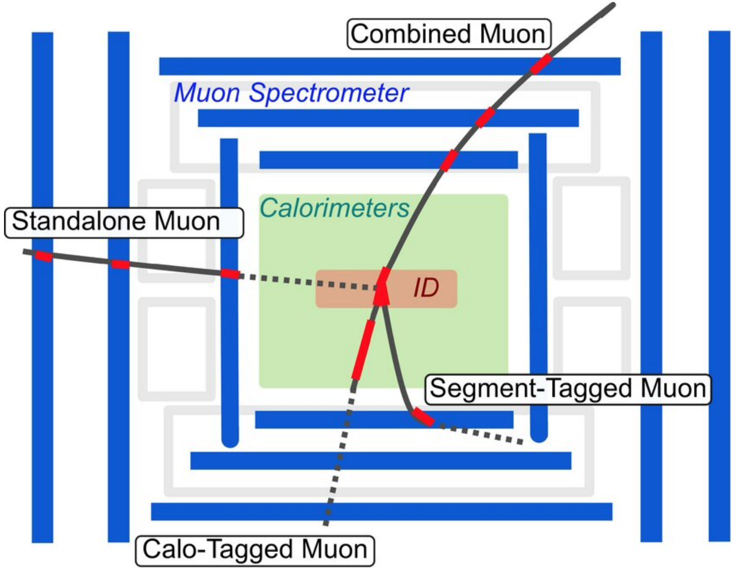


- **Electrons are identified with tracks and EM topo-clusters.**
- Id efficiency shown as a function of electron η:
  - Also studied as a function of electron E_T, pile-up…
- Data and simulation have different efficiencies, an approximately 5% effect.
  - *Weights* or *scale-factors* are derived as a function of η, E_T, …
  - We *reweight* the simulation to achieve the same efficiencies as in the data.

EGAM-2021-01

Example #2: smearing the MC

- **How well do we measure the momenta of muons?**



- Muons are typically reconstructed using the ATLAS inner detector and the muon spectrometer.
- Each detector has its own momentum resolution:

$$\frac{\sigma(p_T)}{p_T} = r_0/p_T \oplus r_1 \oplus r_2 \cdot p_T$$

Fluctuations of energy loss in material

Multiple scattering, B inhomogeneities, local radial displacements

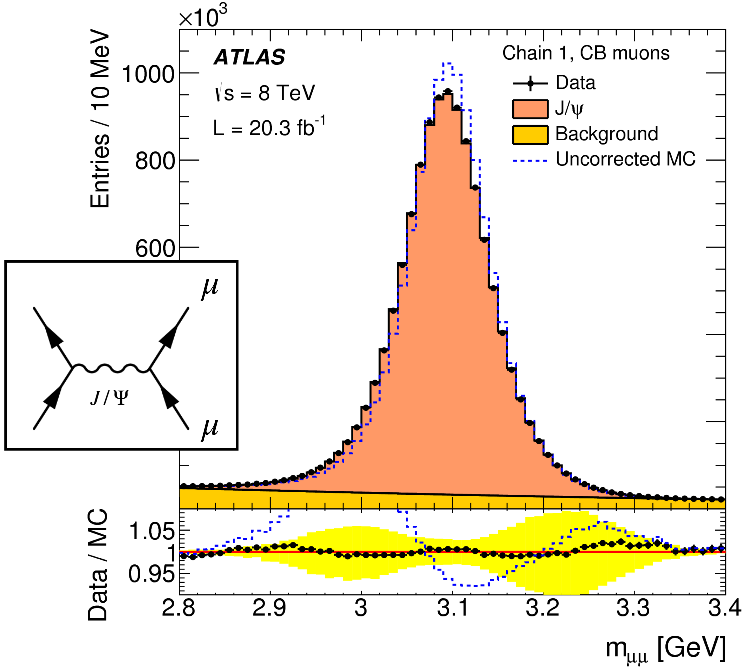
Intrinsic resolution effects from spatial resolution of hits and res. misalignment

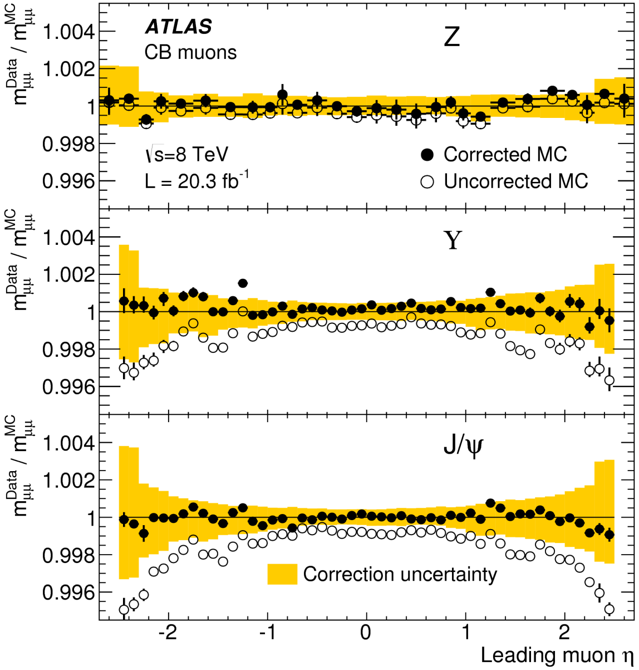
- We *smear the MC* (depending on detector region) to reproduce the muon momentum resolution and scale of data at high precision.

arXiv:1407.3935

Example #2: smearing the MC

- **How well do we measure the momenta of muons?**



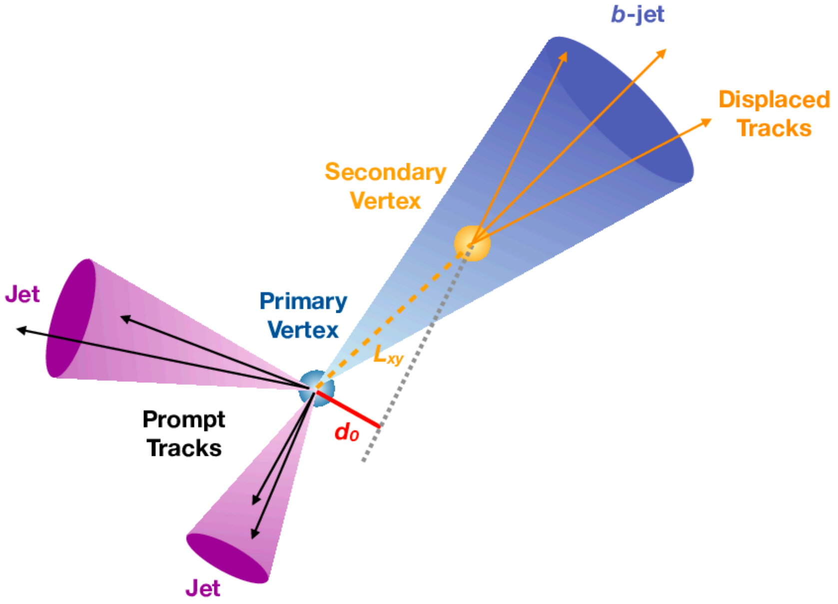


arXiv:1407.3935

Example #3: measuring tag rates, fake rates

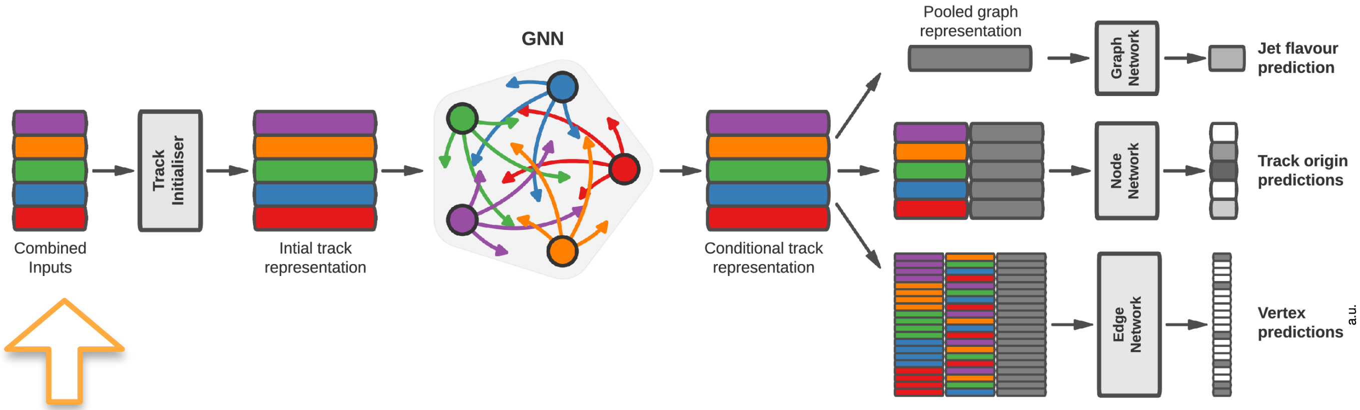
- **How do we identify b-jets?**
  - b-jets contain the decay particles of long-lived b-hadrons and some additional particles
  - Key properties:
    - Relatively large b-hadron mass ~5 GeV
    - Significant b-hadron lifetime ~1.5 ps

  - This leads to **unique characteristics** that distinguish them from light (u,d,s,g) and to a lesser extent charm (c) jets:
    - A secondary vertex
    - Tracks with large impact parameters
    - Leptons from the b-hadron decay

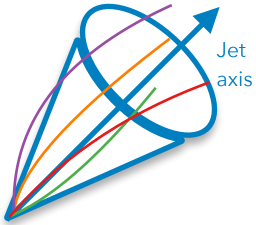


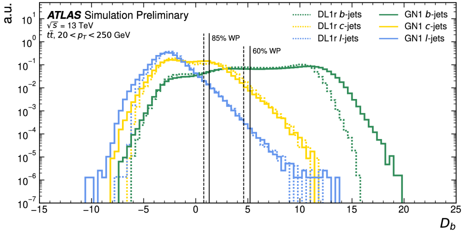
Example #3: measuring tag and mis-tag rates

- **State-of-the-art b-tagging in ATLAS**



Is it a b-jet?
Is it a c-jet?
Is it a light-jet?





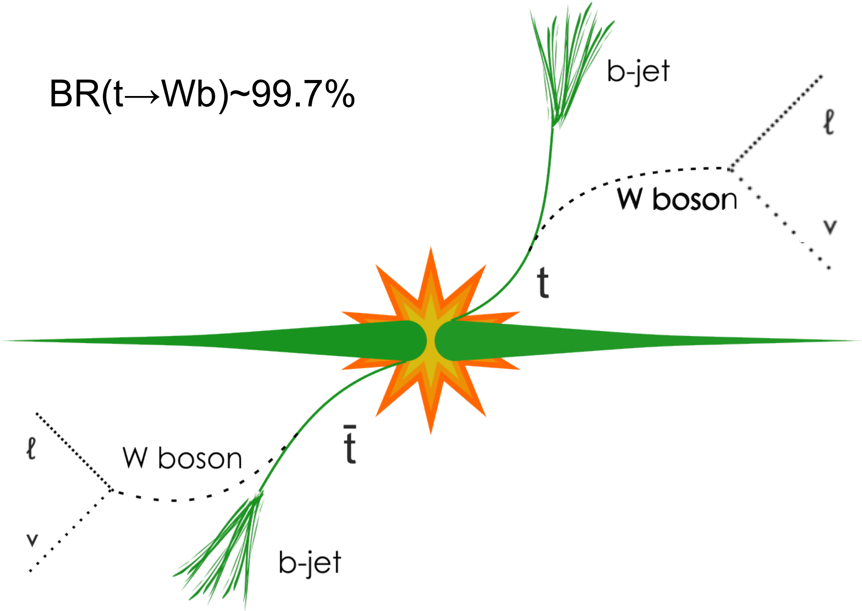
$$D_b = \log \left[\frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_u} \right]$$

ATL-PHYS-PUB-2022-027

Example #3: measuring tag and mis-tag rates

- **How efficiently do we tag a b-jet?**

BR(t→Wb)~99.7%



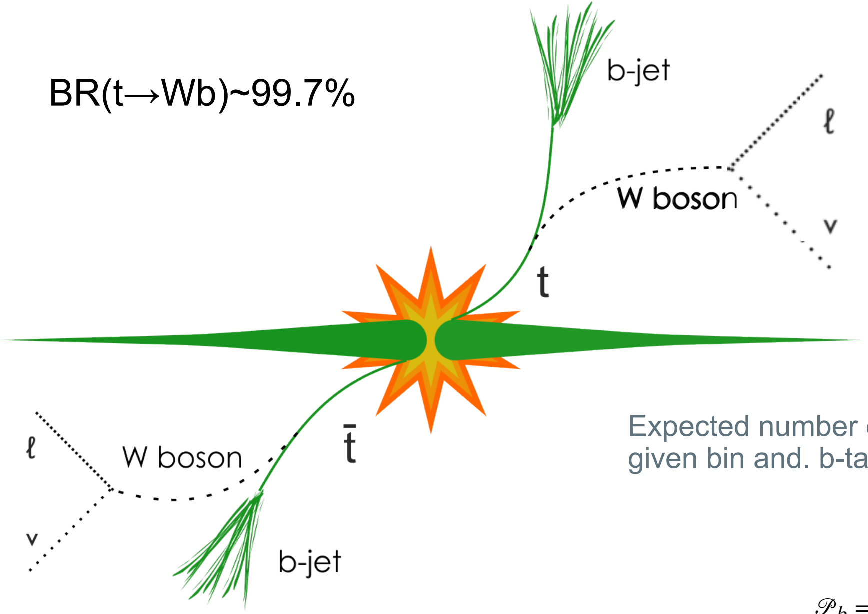
- Use a highly-enriched sample of top-pair events to:
  - Measure the jet flavour composition.
  - Measure the b-tagging efficiency vs jet p_{T}.
- Invariant mass of each of the top systems:
  - $m_{t1} = m_{j1,\ell}$
  - $m_{t2} = m_{j2,\ell}$
- Real top-pair events will have $m_{j1,\ell}, m_{j2,\ell}$ distributions with an upper limit around the top-quark mass of 172.5 GeV
  - In practice smaller due to the undetected neutrino.

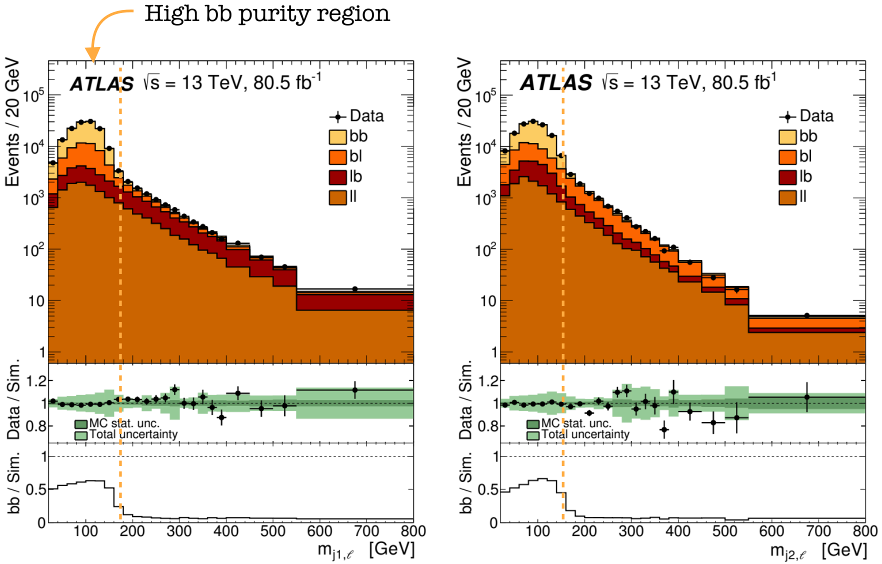
FTAG-2018-01

Example #3: measuring tag and mis-tag rates

- **How efficiently do we tag a b-jet?**

BR(t→Wb)~99.7%




High bb purity region

Expected number of events in a given bin and. b-tagging selection:

$$\begin{aligned}\nu_{SR}(T^m,T^n,O^k,O^p) = {} & c^{m,n}_{bb}\nu^{m,n}_{SR,bb}\cdot\mathcal{P}_b(O^k|T^m)\cdot\mathcal{P}_b(O^p|T^n)\\ &+ c^{m,n}_{bl}\nu^{m,n}_{SR,bl}\cdot\mathcal{P}_b(O^k|T^m)\cdot\mathcal{P}_l(O^p|T^n)\\ &+ c^{m,n}_{lb}\nu^{m,n}_{SR,lb}\cdot\mathcal{P}_l(O^k|T^m)\cdot\mathcal{P}_b(O^p|T^n)\\ &+ c^{m,n}_{ll}\nu^{m,n}_{SR,ll}\cdot\mathcal{P}_l(O^k|T^m)\cdot\mathcal{P}_l(O^p|T^n),\end{aligned}$$

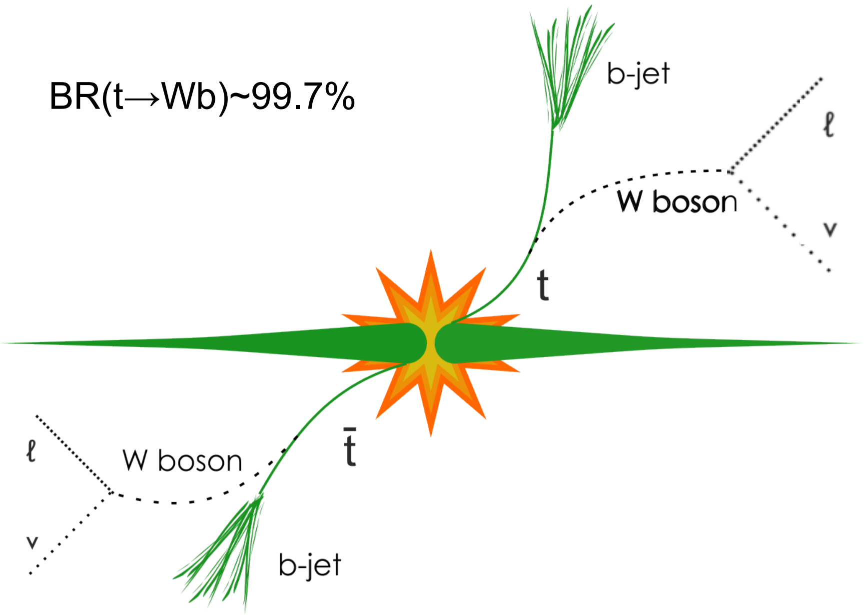
$\mathscr{P}_b$ = b-tagging probability

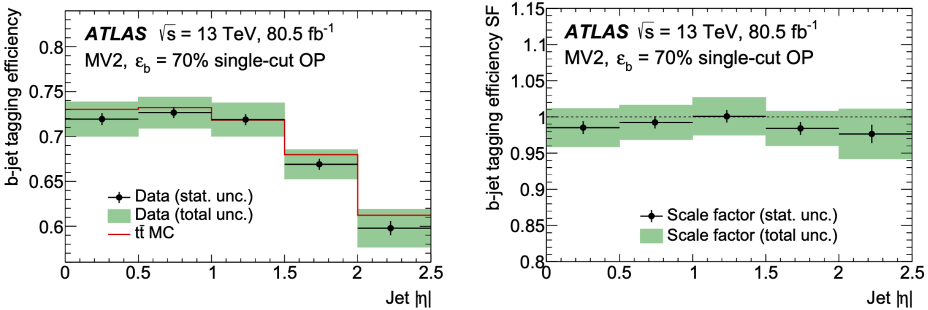
FTAG-2018-01

Example #3: measuring tag and mis-tag rates

- **How efficiently do we tag a b-jet?**

BR(t→Wb)~99.7%



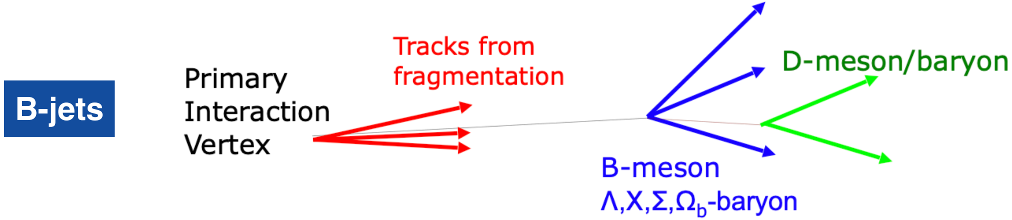


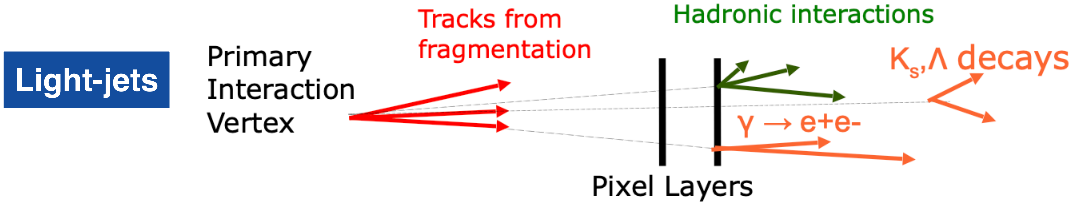
- Scale-factors to correct b-tagging efficiency in MC as a function of pseudo-rapidity (also transverse momentum).

FTAG-2018-01

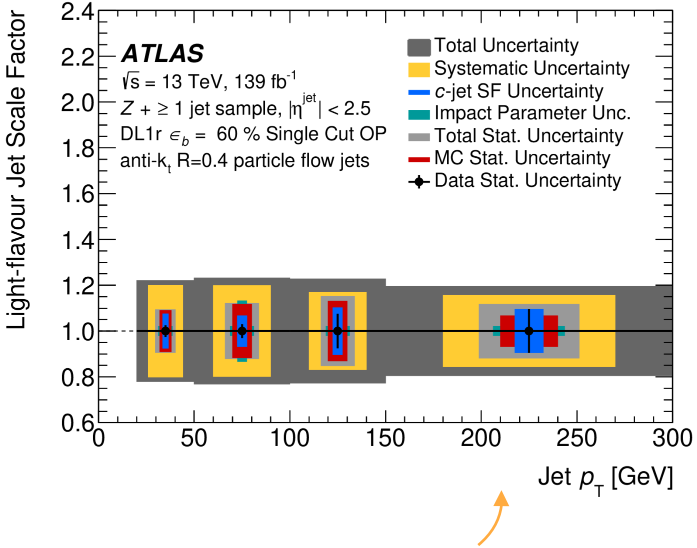
Example #3: measuring tag and mis-tag rates

- **And how often do we tag a light or a c-jet instead (*mis-tag*)?**

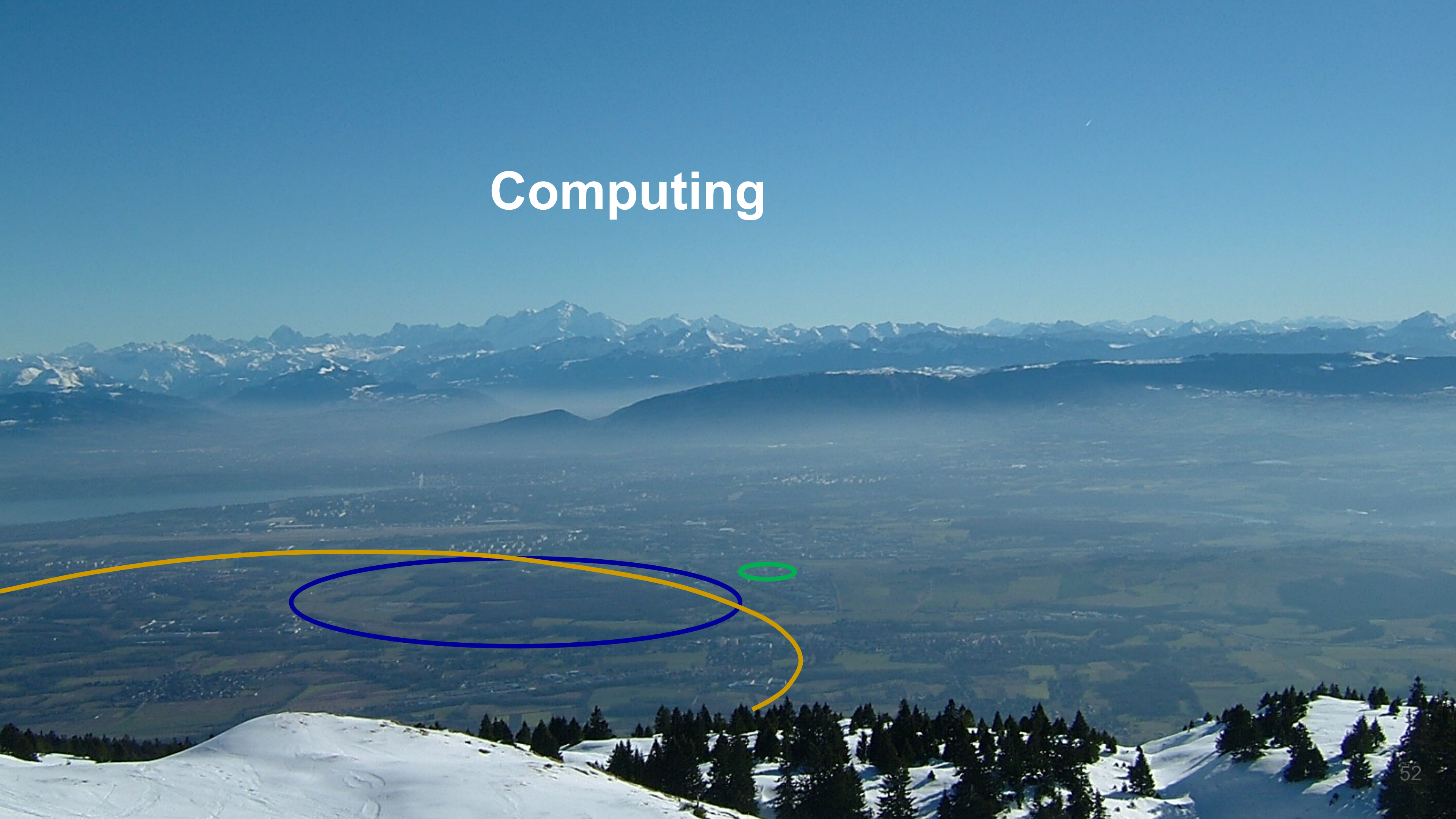




Diagrams by Giacinto Piacquadio



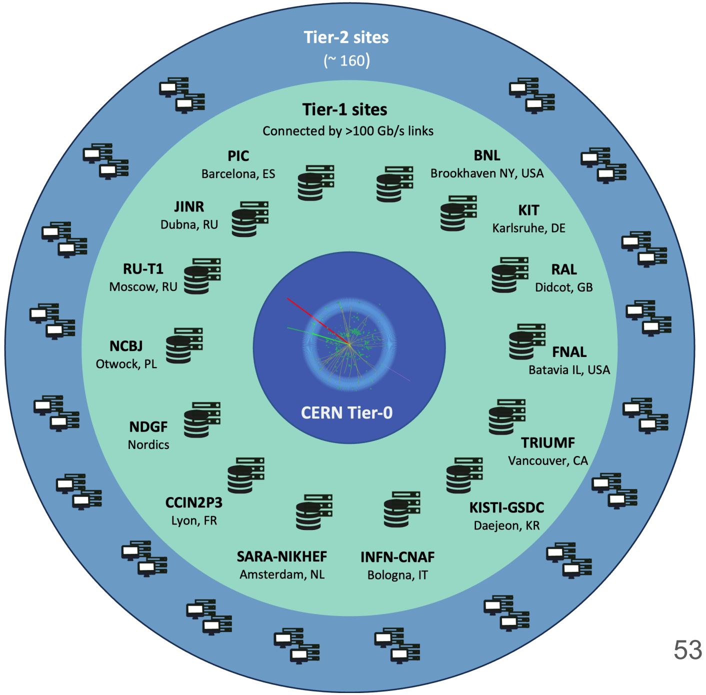
- Scale-factors to correct light mis-tag rate in MC as a function of jet transverse momentum.



Computing

The grid

- Processing data and simulation poses huge computing, storage and analysis challenges.
- We rely on the World LHC Computing Grid (WLCG), and international organisation of computing centres.
    - Tier-0: the CERN Data Centre where O(100) PB of data are stored on magnetic tapes.
    - Tier 1: 14 large data centres for intensive computing tasks and secondary storage.
    - Tier 2: ~160 smaller processing centres, like universities or labs that can provide storage and computing power for specific analysis tasks.



Tier-2 sites
(~ 160)

Tier-1 sites
Connected by >100 Gb/s links

PIC Barcelona, ES
BNL Brookhaven NY, USA
JINR Dubna, RU
KIT Karlsruhe, DE
RU-T1 Moscow, RU
RAL Didcot, GB
NCBJ Otwock, PL
FNAL Batavia IL, USA
NDGF Nordics
TRIUMF Vancouver, CA
CCIN2P3 Lyon, FR
KISTI-GSDC Daejeon, KR
SARA-NIKHEF Amsterdam, NL
INFN-CNAF Bologna, IT

CERN Tier-0

The grid

- Processing data and simulation poses huge computing, storage and analysis challenges.
- We rely on the World LHC Computing Grid (WLCG), and international organisation of computing centres.
    - Tier-0: the CERN Data Centre where O(100) PB of data are stored on magnetic tapes.
    - Tier 1: 14 large data centres for intensive computing tasks and secondary storage.
    - Tier 2: ~160 smaller processing centres, like universities or labs that can provide storage and computing power for specific analysis tasks.

An ATLAS example:

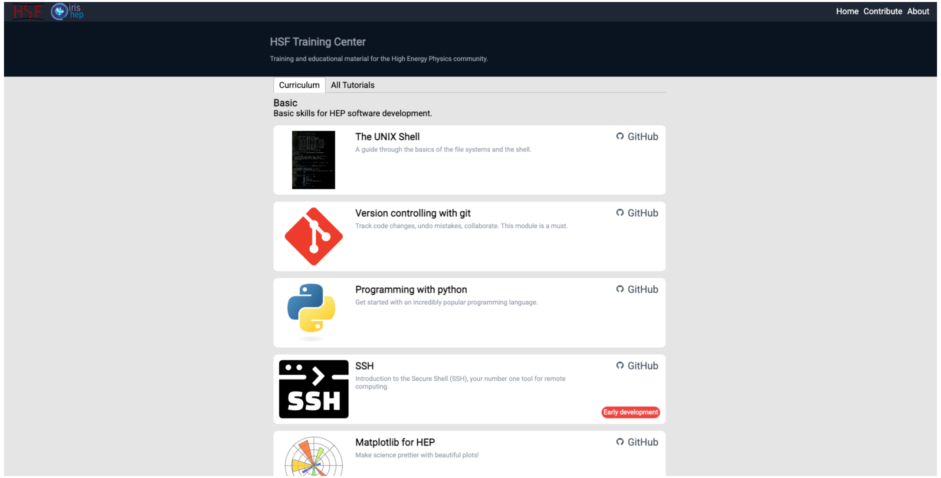
DAOD	129 PB
AOD	80.2 PB
HITS	51.4 PB
RDO	17.9 PB
EVNT	13.0 PB
RAW	11.9 PB
ESD	2.27 PB
DESD	1.52 PB
log	1.43 PB
no_name	1.24 PB
TXT	1.00 PB
HIST	886 TB
DRAW	787 TB
user	697 TB
NTUP	263 TB

HEP Software Foundation

The HEP Software Foundation facilitates cooperation and common efforts in High Energy Physics software and computing internationally.

Lots of training resources:

https://hsf-training.org/



HSF Training Center

Training and educational material for the High Energy Physics community.

Basic

Basic skills for HEP software development.

The UNIX Shell

A guide through the basics of the file systems and the shell.

Version controlling with git

Track code changes, undo mistakes, collaborate. This module is a must.

Programming with python

Get started with an incredibly popular programming language.

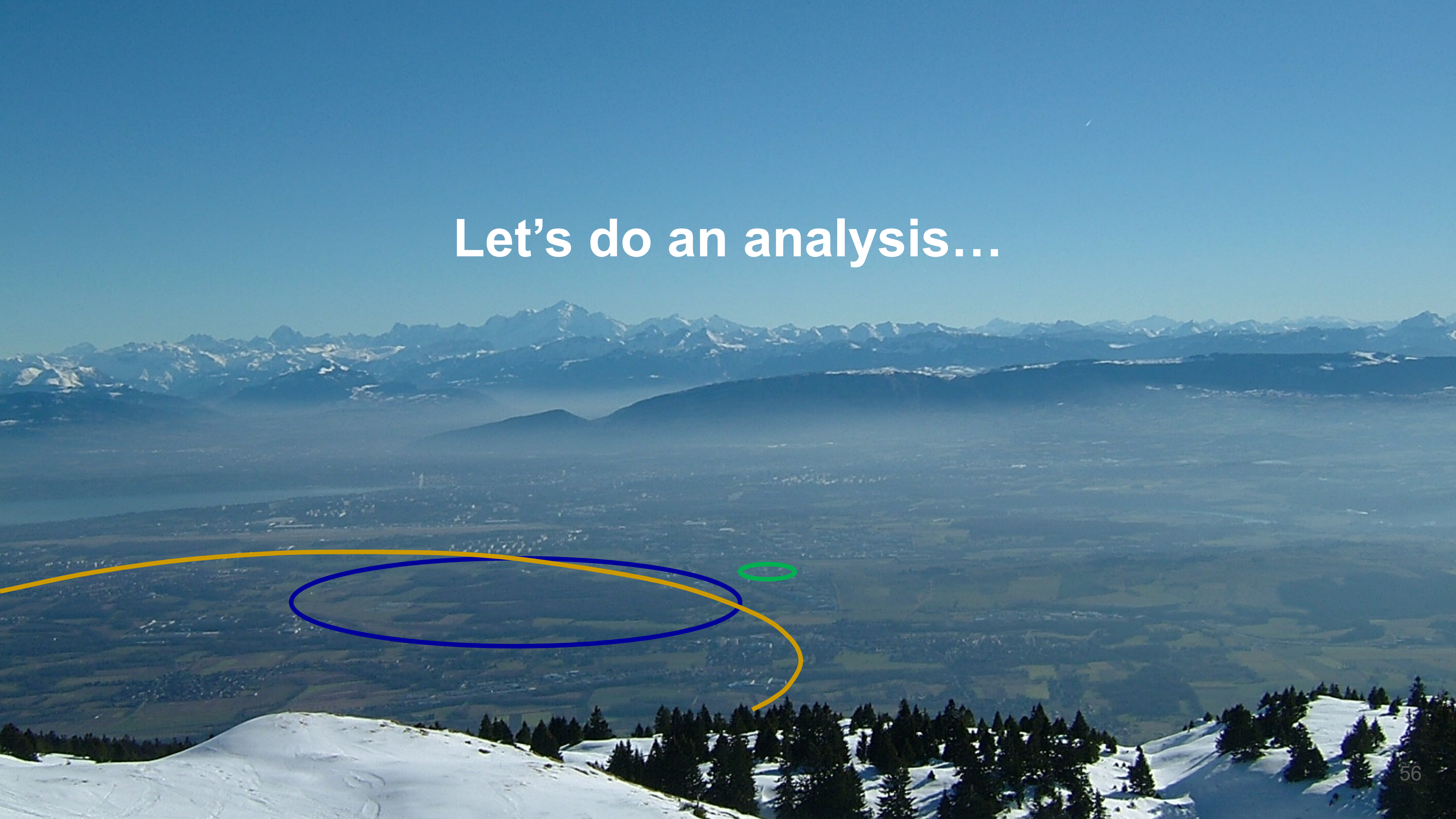
SSH

Introduction to the Secure Shell (SSH), your number one tool for remote computing

Matplotlib for HEP

Make science prettier with beautiful plots!

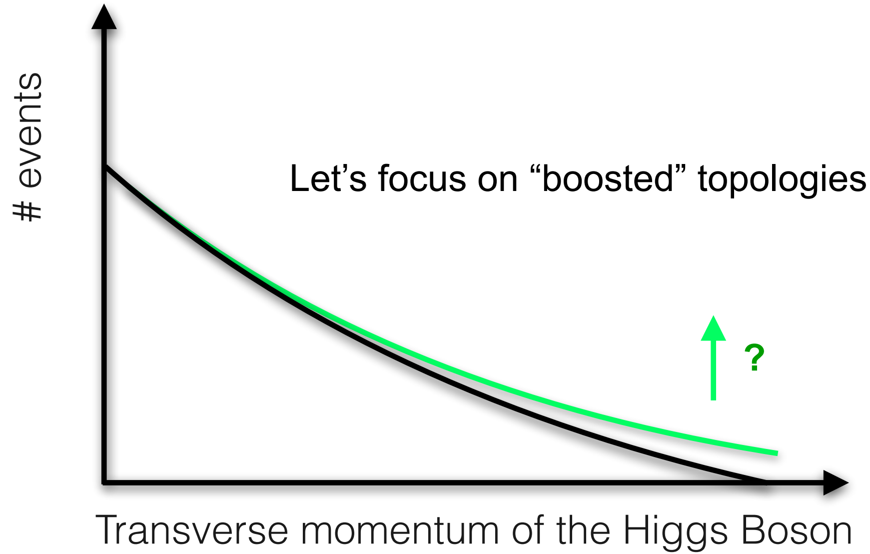
Let’s do an analysis…



Case study: let's measure the WH cross-section

☑ Physics motivation

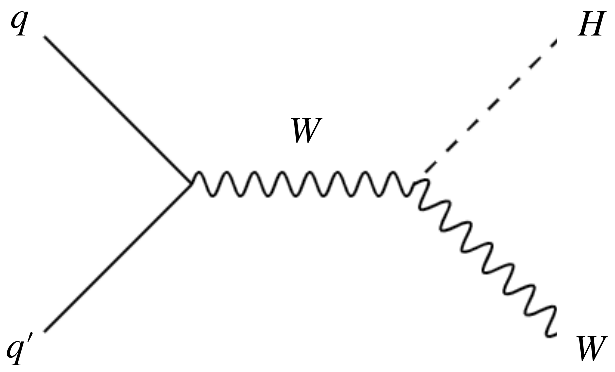
  - Precise measurements of the cross-section and decays of Higgs boson as a test of Standard Model predictions and probe of New Physics.



Let's focus on "boosted" topologies…

Case study: let's measure the WH cross-section

☑ Signal characterization



SM Higgs Boson decay modes

Decay channel	Branching ratio
$H \to \gamma\gamma$	2.27×10^{-3}
$H \to ZZ$	2.62×10^{-2}
$H \to W^+W^-$	2.14×10^{-1}
$H \to \tau^+\tau^-$	6.27×10^{-2}
$H \to b\bar{b}$	5.84×10^{-1}
$H \to Z\gamma$	1.53×10^{-3}
$H \to \mu^+\mu^-$	2.18×10^{-4}

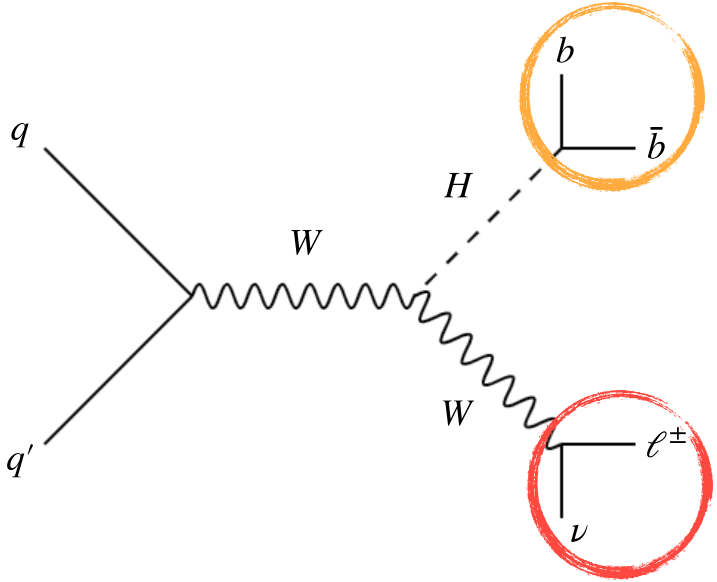
W boson decay modes

		Fraction (Γ_i/Γ)
$\ell^+\nu$	[b]	(10.86 ± 0.09) %
$e^+\nu$		(10.71 ± 0.16) %
$\mu^+\nu$		(10.63 ± 0.15) %
$\tau^+\nu$		(11.38 ± 0.21) %
hadrons		(67.41 ± 0.27) %

PDG

Case study: let's measure the WH cross-section

☑ Signal characterisation



✓ **Two b's**

🤩 Large branching ratio

😬 Lots of background

✓ **One charged lepton**

✓ **One neutrino**

🤩 Easy to trigger on

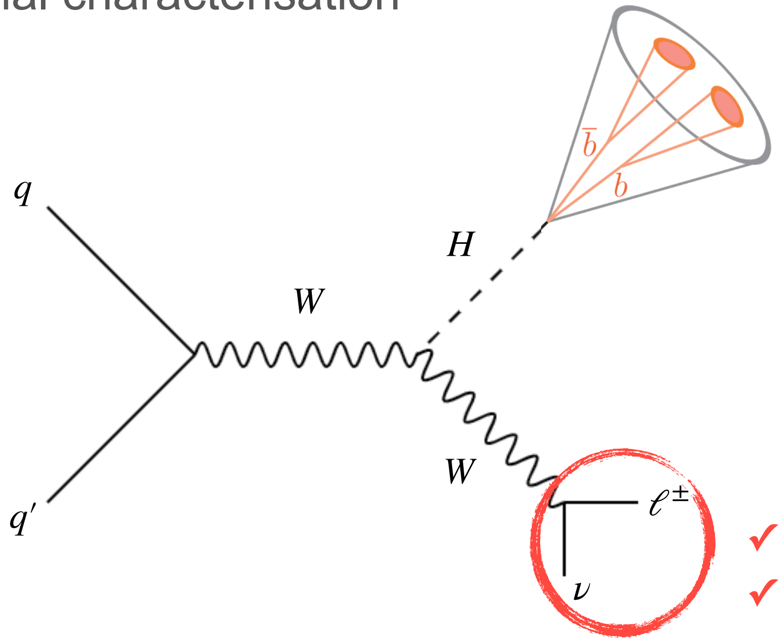
😬 Small branching ratio

😬 Invisible neutrino

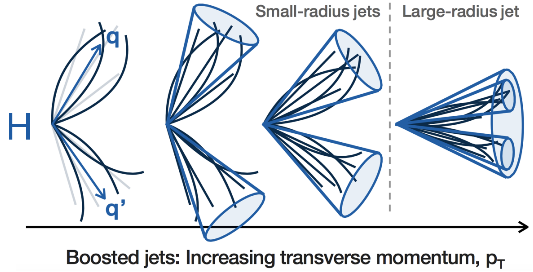
Let's explore other options later in the talk…

Case study: let's measure the WH cross-section

☑ Signal characterisation



✓ **1 large-radius jet containing 2 b-jets**



✓ **One charged lepton**

✓ **Missing transverse energy**

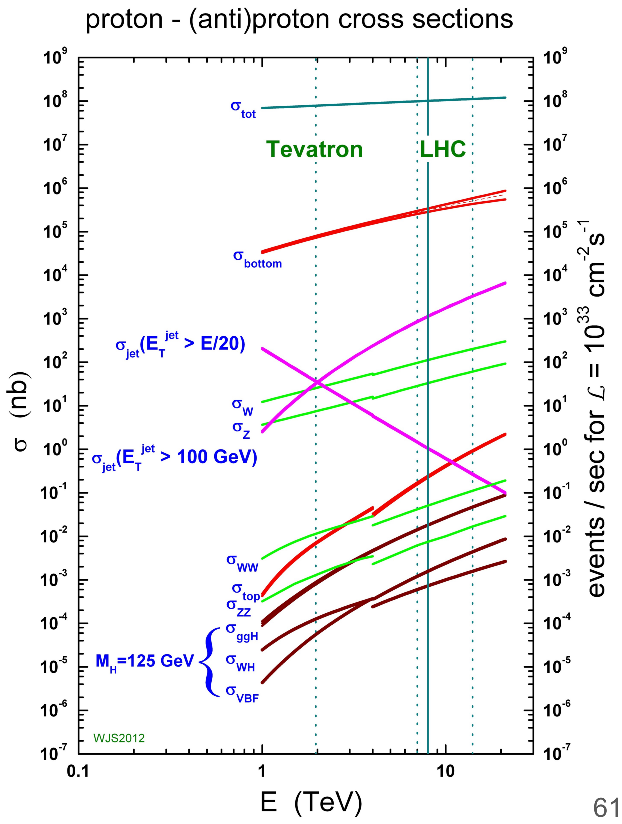
Case study: let's measure the WH cross-section

- **Background:** it is crucial to correctly estimate the expected background and its uncertainty.
- Common strategy (for many backgrounds):

1. Use Monte Carlo estimate (yields and shapes) during analysis optimisation.

2. Use data to correct and constrain MC estimate. *

* when there is an appropriate **control region**.

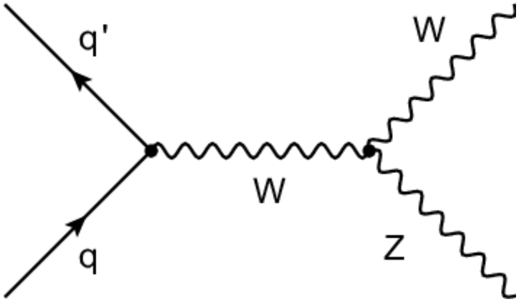


Case study: let's measure the WH cross-section

☑ What are the (dominant) backgrounds? How can we reduce them?

WZ production

- ✓ **1 large-R jet**
  - ✓ **W/ 2 b-tags**
- ✓ **1 $\ell^{\pm}$**
- ✓ **EtMiss**



W→ℓv

✓ **One charged lepton**

✓ **Missing transverse energy**

Z→bb

✓ **1 large-radius jet containing 2 b-jets**

😕 **Background from cc faking bb**

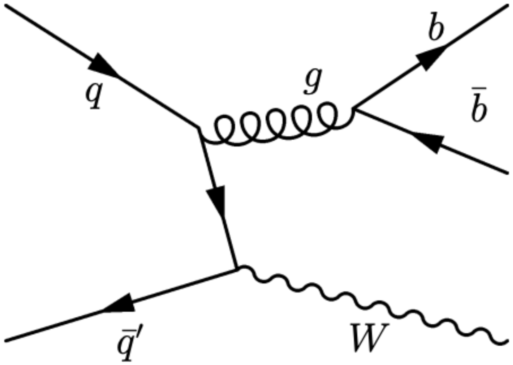
💡 Can take advantage of different invariant mass of Z and H (if mass resolution allows it)

Case study: let's measure the WH cross-section

☑ What are the (dominant) backgrounds? How can we reduce them?

- ✓ 1 large-R jet
  - ✓ W/ 2 b-tags
- ✓ 1 $\ell^\pm$
- ✓ EtMiss

W+jets production



✓ 1 large-radius jet containing 2 b-jets

😕 **Background from cc faking bb**

W→ℓv

✓ One charged lepton

✓ Missing transverse energy

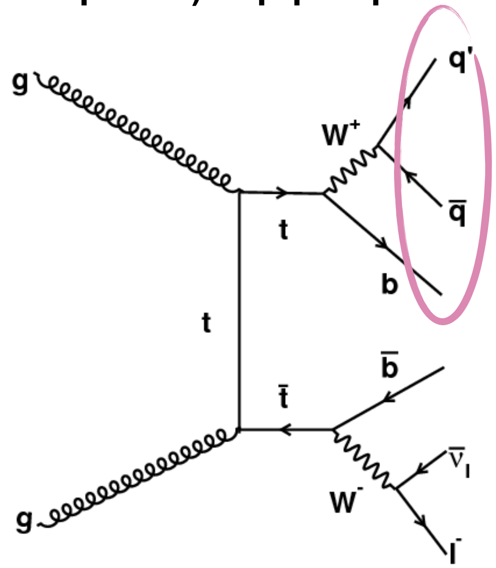
💡 Can take advantage of different kinematics of gluon vs Higgs decay

Case study: let's measure the WH cross-section

☑ What are the (dominant) backgrounds? How can we reduce them?

- ✓ 1 large-R jet
- ✓ 2 b-tags
- ✓ 1 $\ell^{\pm}$
- ✓ EtMiss

(Semi-leptonic) Top-pair production



➡ e.g. 1 large-radius jet with 1b+1c (where the c was *mis-tagged* as a b)

➡ 1 extra b (outside the large-R jet)

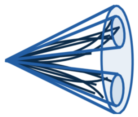
W→ℓv

✓ **One charged lepton**

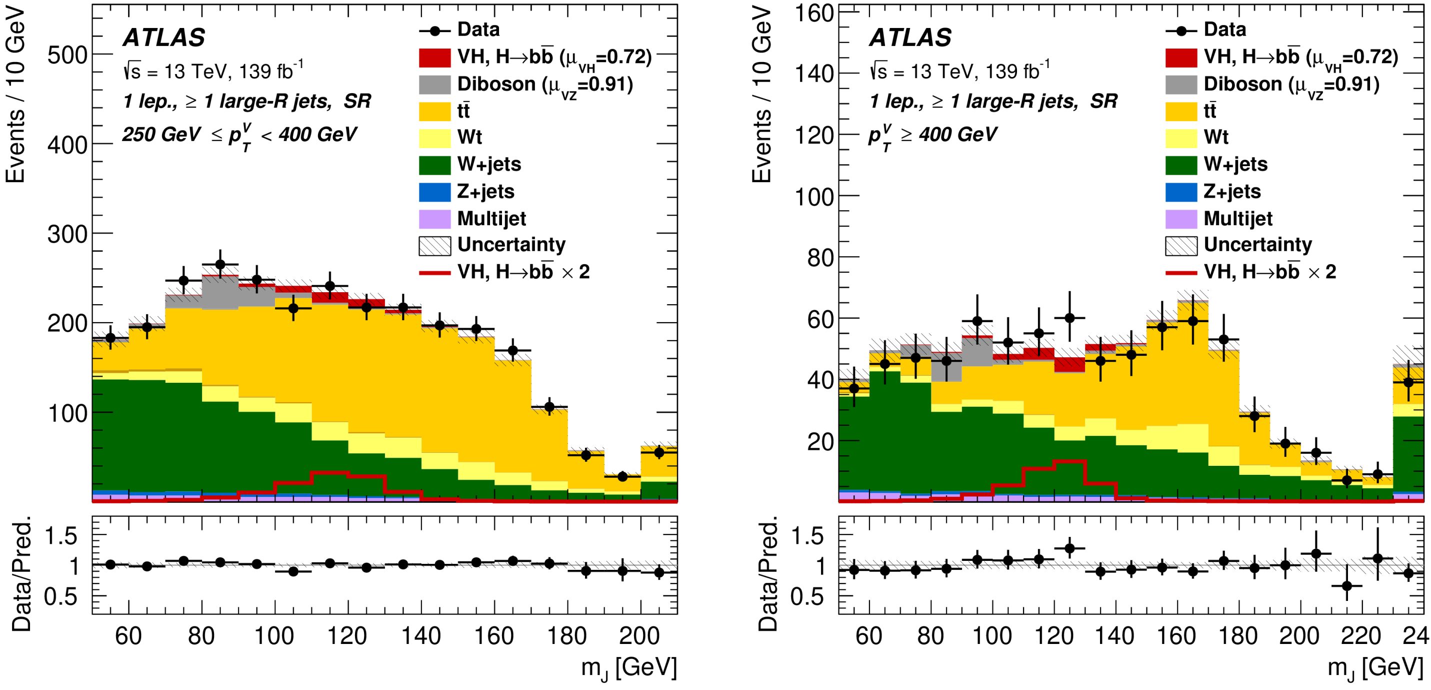
✓ **Missing transverse energy**

💡 Can take advantage of the extra b-jet.

Case study: let's measure the WH cross-section



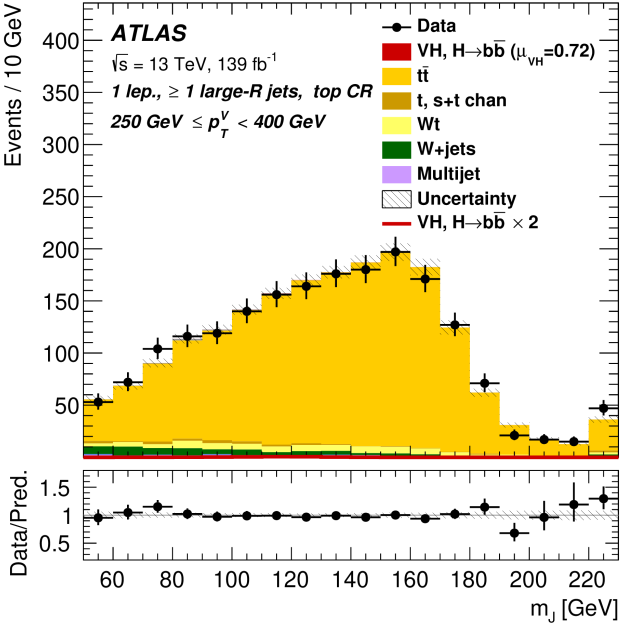
- ☑ Event selection: regions with high signal efficiency ➡ *signal region(s)*

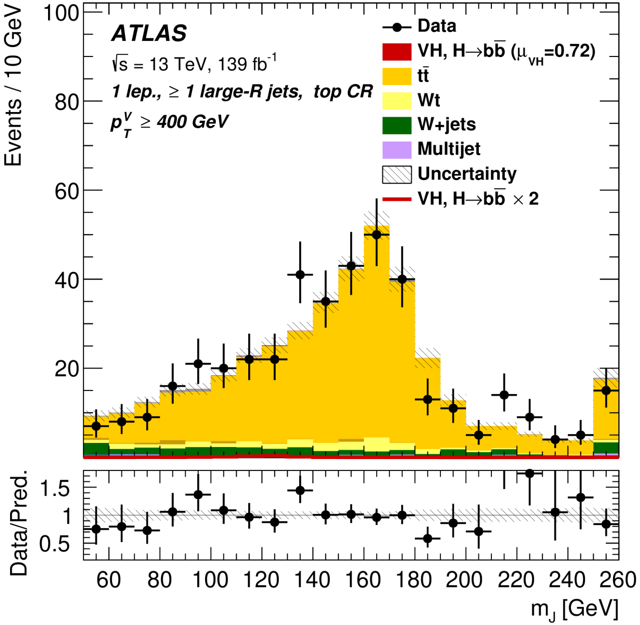


m_J = mass of the large-radius jet

Case study: let's measure the WH cross-section

☑ Event selection: regions with high background purity ➡ *control region(s)*



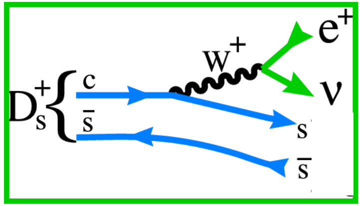


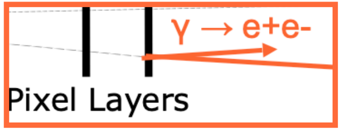
➡ **Top CR: events with 1 extra b (outside the large-R jet)**

Case study: let's measure the WH cross-section

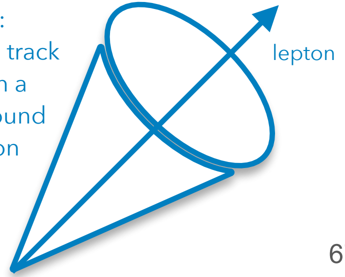
☑ Background from *non-prompt / fake* leptons:

- Non-prompt: from semi-leptonic decay of hadrons or photon conversions.
- Fake leptons from misidentified jets.
- Very challenging to model these processes in simulation:
  - Depend strongly on details of physics simulation, often in non-perturbative regions.
  - Depend on modeling of material composition and response.
  - Very low probability for hadronic jets to fake a lepton, yet multi-jet cross section is huge and simulating this effect would be prohibitive.





Isolation:
calo and track
activity in a
cone around
the lepton



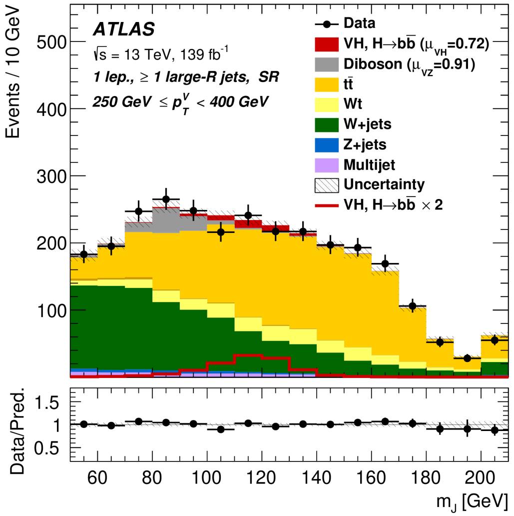
💡 Non-prompt leptons can be reduced by requiring *isolated* leptons.

ANA-EGAM-2019-01

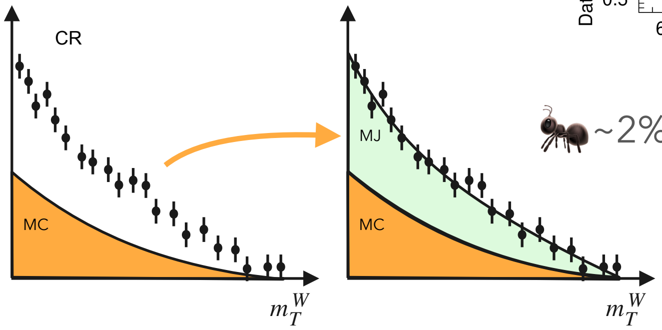
Case study: let’s measure the WH cross-section

☑ Background from *non-prompt / fake* leptons:

- **Use data-driven methods!**
- E.g. template method:
  - Extract a background *template* from a control region enriched in multi-jet events.
    - Built by inverting the lepton isolation and missing transverse energy requirements.
    - Assumption: shape SR = shape CR
  - Determine its normalisation in fits to the W transverse mass distribution in the signal region.



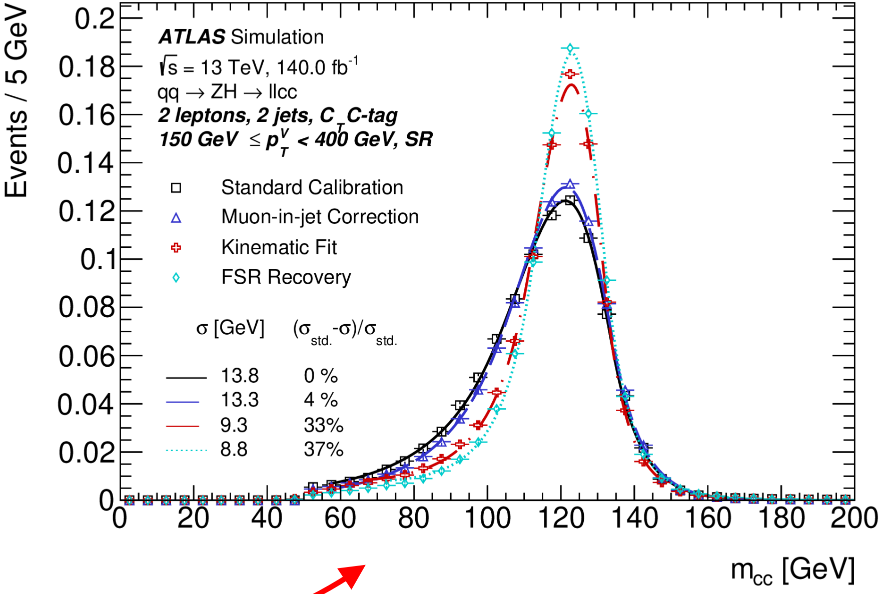
$$m_T^W = \sqrt{2p_T^\ell E_T^{miss}(1 - \cos\Delta\phi(\ell, E_T^{miss}))}$$

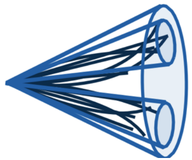


🐜~2% in the electron channel (with a 55% uncertainty)

Case study: let's measure the WH cross-section

☑ Analysis specific: improvements to the invariant mass resolution

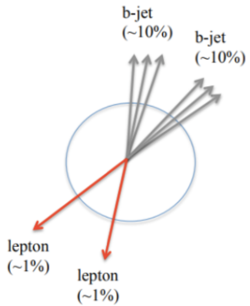




$m_J = m_{bb}$ = mass of the large-radius jet (Higgs candidate)

- ✓ Correct b-jets semi-leptonic decays with muon four-vector.
- ✓ Correct for missing energy from neutrinos.
- ✓ In the ZH(llbb) channel, a kinematic fit.

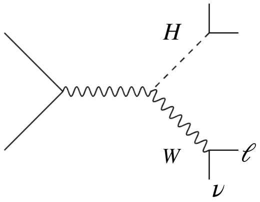
In the end, a 37% improvement.



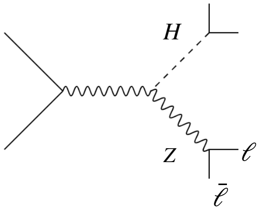
Case study: let’s measure the **VH** cross-section

☑ All together now!

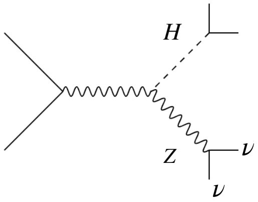
- The data, the simulated and data-driven backgrounds, as well as the Higgs boson signal go into a **likelihood fit** of the signal and control regions, considering theoretical and experimental uncertainties.



1-lepton



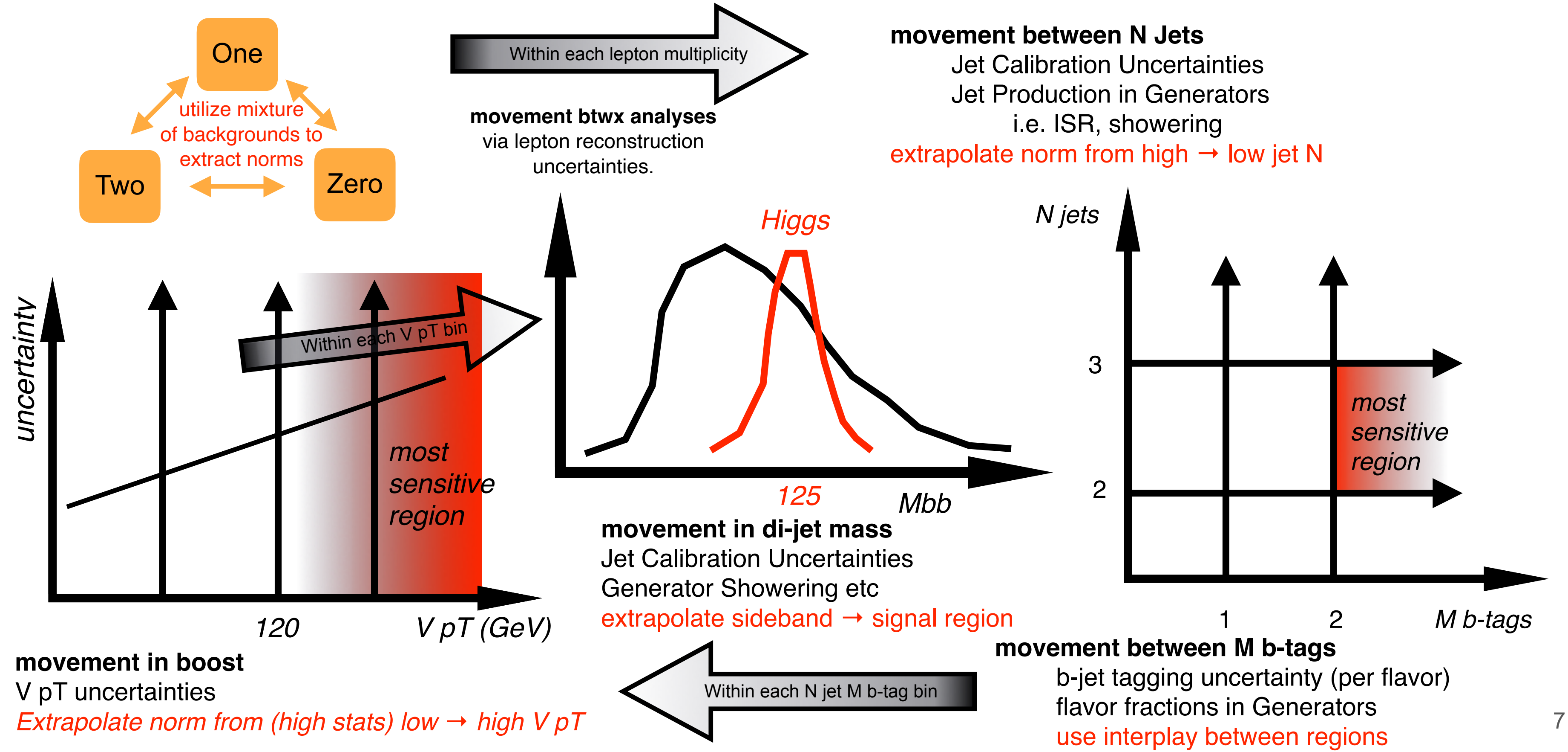
2-lepton



0-lepton

HIGG-2018-52

Case study: let's measure the **VH** cross-section



utilize mixture of backgrounds to extract norms

movement btwx analyses
via lepton reconstruction
uncertainties.

movement between N Jets
Jet Calibration Uncertainties
Jet Production in Generators
i.e. ISR, showering
extrapolate norm from high → low jet N

movement in boost
V pT uncertainties
Extrapolate norm from (high stats) low → high V pT

movement in di-jet mass
Jet Calibration Uncertainties
Generator Showering etc
extrapolate sideband → signal region

movement between M b-tags
b-jet tagging uncertainty (per flavor)
flavor fractions in Generators
use interplay between regions

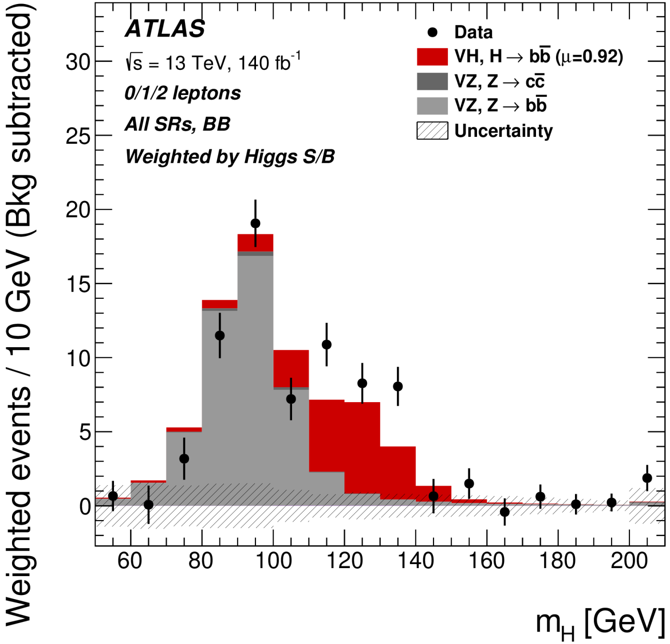
Case study: let’s measure the VH cross-section

☑ All together now

- In this case, a strength parameter of the signal is measured:

$$\mu_{VH} = \frac{\sigma_{\text{meas}}}{\sigma_{SM}} = 0.92 \pm (\text{stat.})^{+0.13}_{-0.11}(\text{syst.})$$

- We take advantage of the diboson peak for validation, before *unblinding* the data.



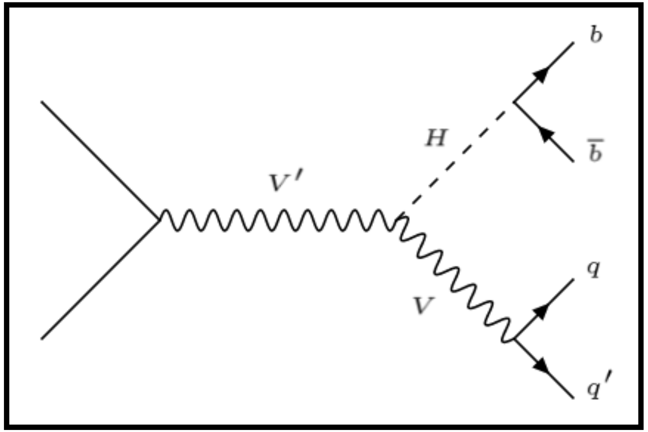
(Including resolved too)

HIGG-2018-52

https://arxiv.org/abs/1707.06958
https://arxiv.org/abs/2007.05293

Case study: searches for new resonances

- Let’s make it more interesting…
  - ✓ Make it a resonance: $W' \rightarrow WH$
  - ✓ Make it all hadronic 🤩



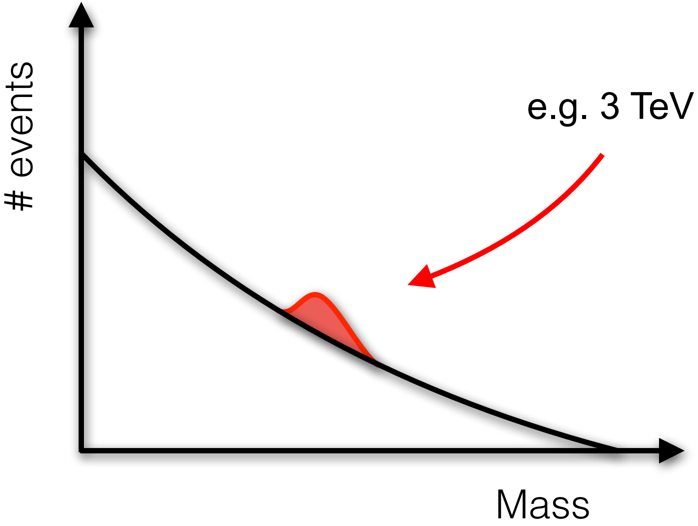
W’, Z’: heavier versions of the W and Z bosons

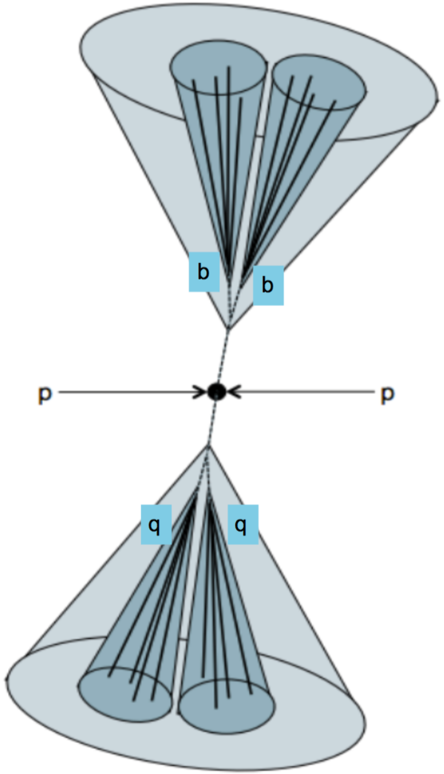
https://arxiv.org/abs/1707.06958
https://arxiv.org/abs/2007.05293

Case study: searches for new resonances

☑ Signal characterization:

- 2 large-radius jets: 1 boosted H→bb jet and 1 boosted W→qq jet
- A resonant peak above the multijet background, ~TeV scale





Case study: searches for new resonances

☑ Dominant background: multi-jet production

- Huge cross-section!
- Tagging of boosted Higgs and boosted W bosons rejects a lot of background, but what remains is tricky and expensive to simulate precisely.

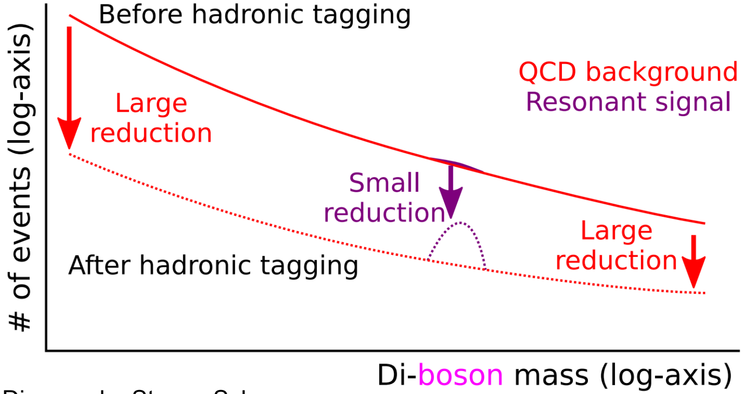


Diagram by Steven Schramm

We do a fully data-driven background estimation ⭐

- In other words, we *interpolate* or *extrapolate* from a background dominated CR into a SR.
- We use data directly (in some cases using MC but only in defining regions or checking assumptions).

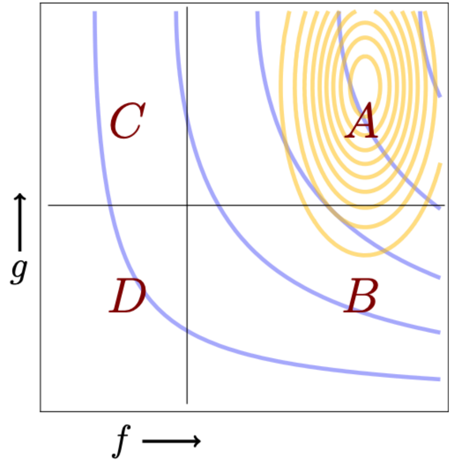
Example #1: the *ABCD* method

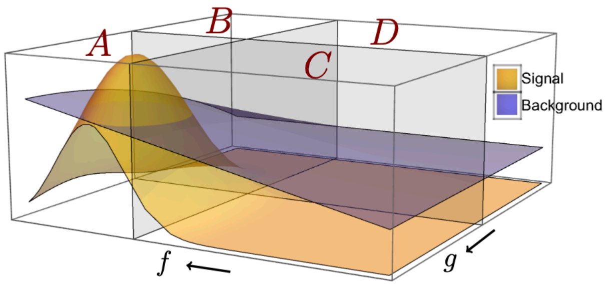
E.g. ***f,g*** = mass, missing ET, …

1. Pick two observables ***f*** and ***g*** which are:
    - ☑ Approximately statistically independent for the background.
    - ☑ Effective discriminators of signal vs background.
2. Apply thresholds on these observables to define 4 regions:
    - A: signal region
    - B, C, D: background regions
3. If ***f*** and ***g*** are independent then the background in A can be predicted from the other three regions:

$$N_A = \frac{N_B N_C}{N_D}$$

$N_i =$ number of events in region i





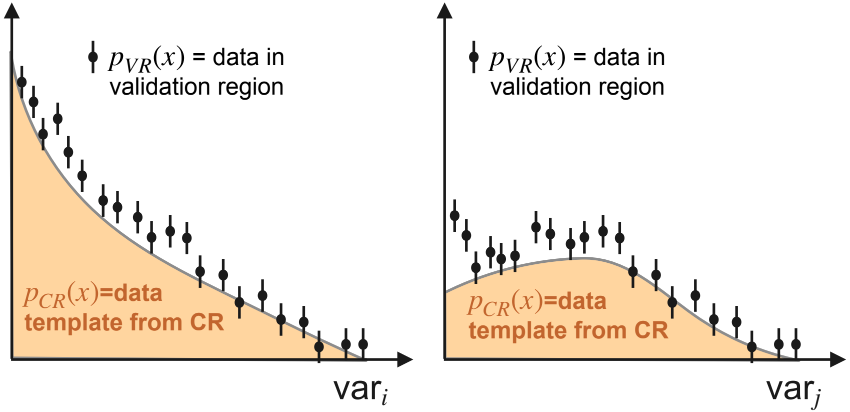
See arXiv:2007.14400 for a ML+ABCD method

Example #2: multi-dimensional reweighting with ML

- Let’s say we extrapolate the background from control region to signal region.
- We cross-check modelling in validation region and observe discrepancies.
- Then, do a **reweighting**: use one sample with distribution $p_{CR}(x)$ to model sample $p_{VR}(x)$, via a density ratio $r(x)$

$$p_{VR}(x) = r(x)p_{CR}(x)$$

How do we determine $r(x)$?

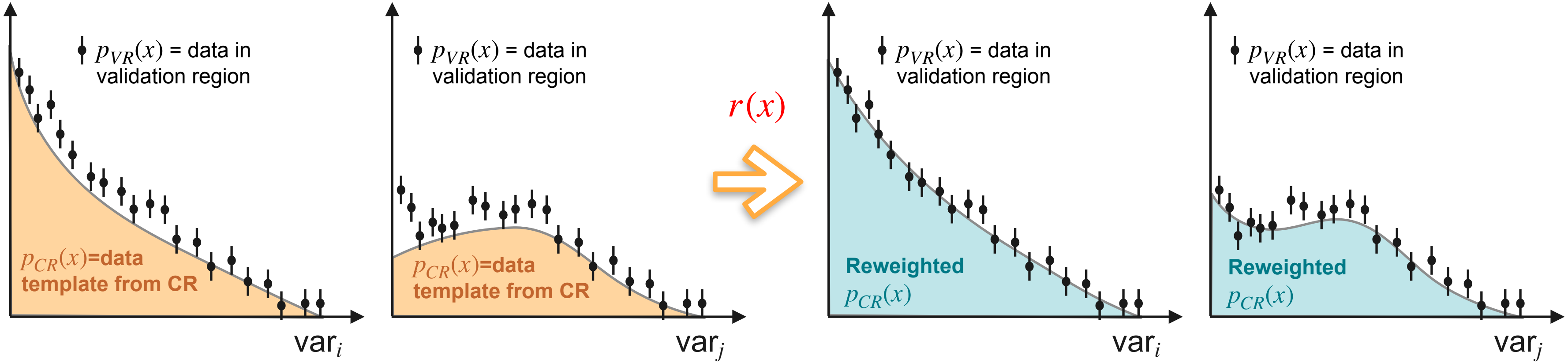


Example #2: multi-dimensional reweighting with ML

- ***Likelihood-ratio trick:*** a classification model (NN, BDT, …) trained to discriminate between samples A and B can also estimate their probabilities.

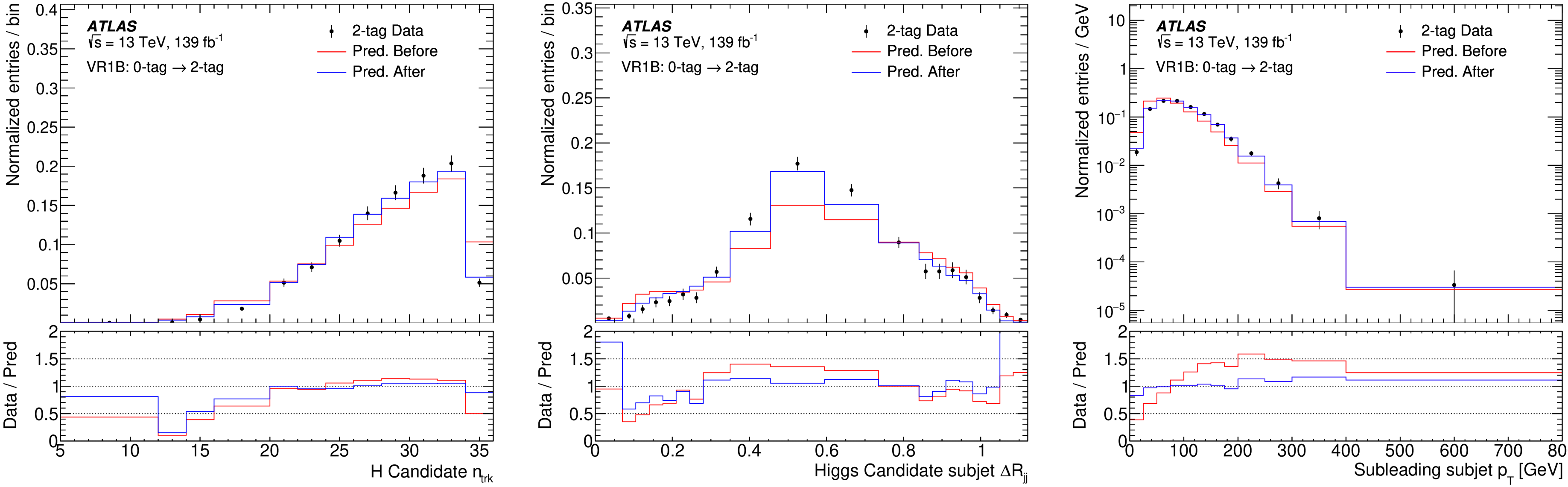
$$r(x) = \frac{p_A(x)}{p_B(x)} = \frac{p(x|A)}{p(x|B)} = \frac{p(A|x)}{p(B|x)} = \frac{h(x)}{1-h(x)}$$

h(x) is the classifier output



Example #2: multi-dimensional reweighting with ML

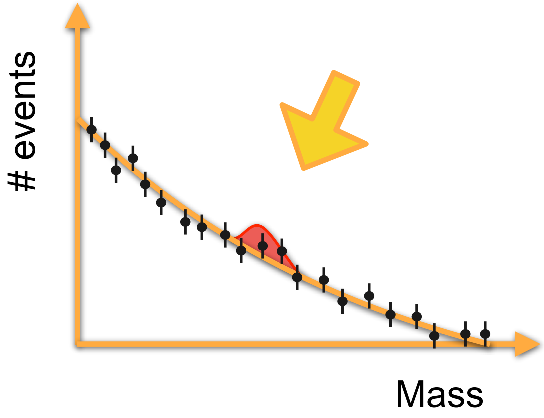
Before BDT reweighing
After BDT reweighting

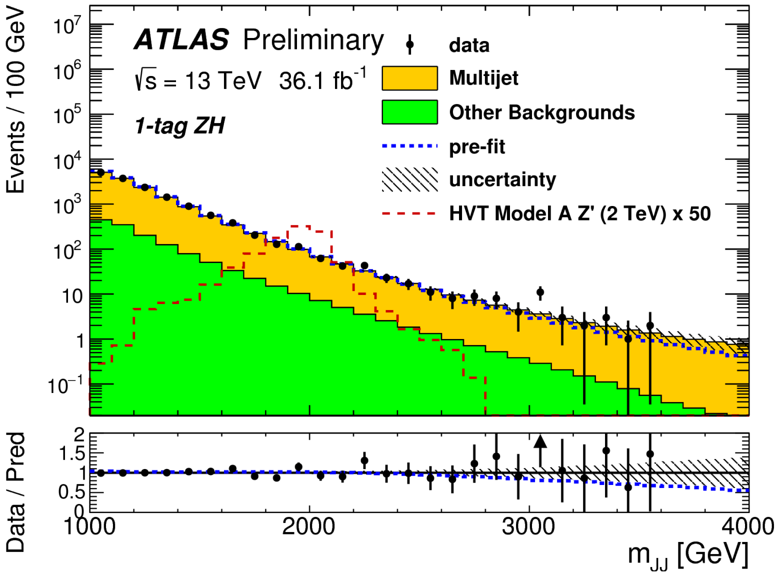


https://arxiv.org/abs/1707.06958
https://arxiv.org/abs/2007.05293

Case study: searches for new resonances

☑ Results:





Finally, we have a background model for the signal region (shape and yields) and systematic uncertainties on that model, to compare against data.

Signs of new physics / statistical fluctuations?

https://arxiv.org/abs/1707.06958
https://arxiv.org/abs/2007.05293

When do announce a discovery?

- To find out if our data is compatible with the presence of new physics, we can compute the probability for our observation with no signal present.

- Suppose we observe $\mathbf{n_b}$ background events and $\mathbf{n_s}$ signal events.
- Suppose $\mathbf{n=n_b+n_s}$ is distributed according to a Poisson distribution with mean **s+b**:

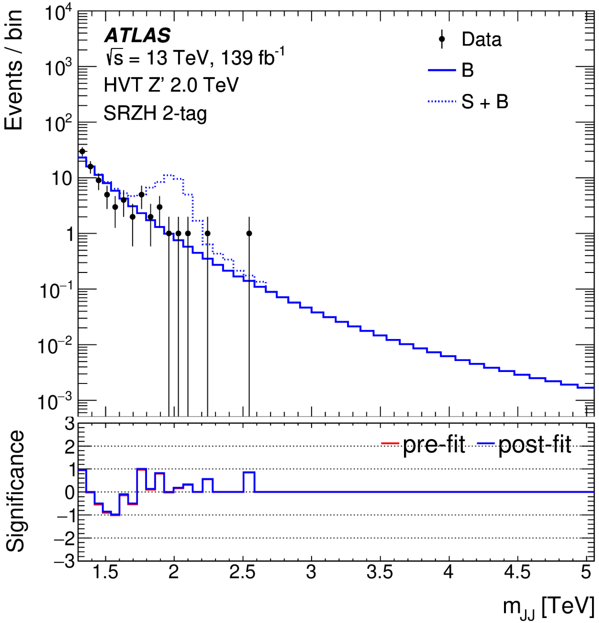
$$P(n;s,b) = \frac{(s+b)^n}{n!}e^{-(s+b)}$$

- If b=0.5 and n=5, do we claim discovery?

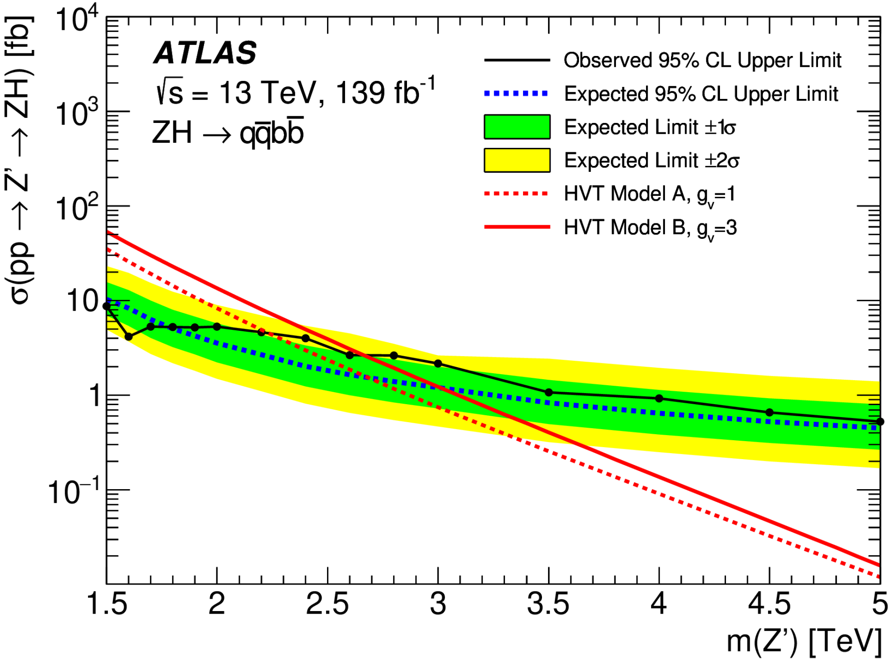
$$\text{p-value} = P(n \geq 5; b = 0.5, s = 0) = 1.7 \times 10^{-4}$$

Case study: searches for new resonances

☑ Interpreting the results in the absence of an excess:



Upper limits on the cross-section of the signal



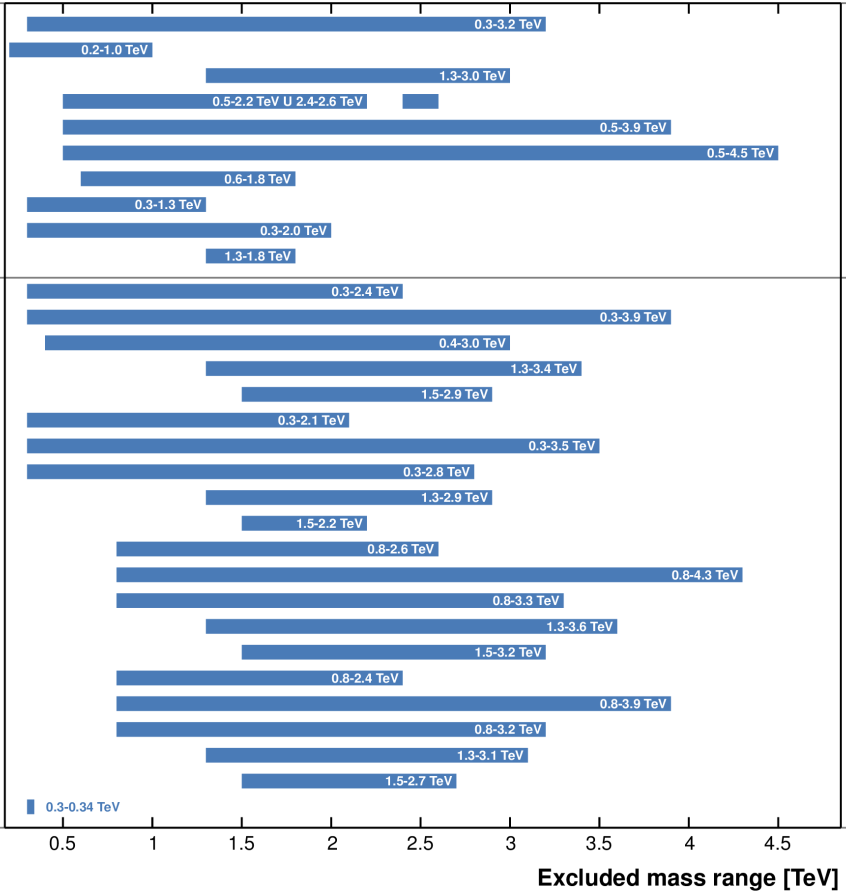
ATLAS Diboson Searches - 95% CL Exclusion Limits

Status: March 2023

$\mathcal{L}$ = 139 fb^{-1}

ATLAS Preliminary

$\sqrt{s}$ = 13 TeV



	Model	Channel†	Strategy*	Limit	Reference
Extra dimensions	Bulk RS ($k\pi r_c = 35$, $\Lambda_R = 3\,$TeV)	$R \to WW, ZZ \to \nu\nu qq, \ell\nu qq, \ell\ell qq$	resolved, boosted	0.3-3.2 TeV	Eur. Phys. J. C 80 (2020) 1165
	Bulk RS ($k\pi r_c = 35$, $\Lambda_R = 3\,$TeV)	$R \to WW \to e\nu\mu\nu$	resolved	0.2-1.0 TeV	ATLAS-CONF-2022-066
	Bulk RS ($k\pi r_c = 35$, $\Lambda_R = 3\,$TeV)	$R \to WW, ZZ \to qqqq$	boosted	1.3-3.0 TeV	JHEP 06 (2020) 042
	RS1 ($k/\overline{M}_{Pl} = 0.01$)	$G_{KK} \to \gamma\gamma$	resolved	0.5-2.2 TeV U 2.4-2.6 TeV	Phys. Lett. B 822 (2021) 136651
	RS1 ($k/\overline{M}_{Pl} = 0.05$)	$G_{KK} \to \gamma\gamma$	resolved	0.5-3.9 TeV	Phys. Lett. B 822 (2021) 136651
	RS1 ($k/\overline{M}_{Pl} = 0.1$)	$G_{KK} \to \gamma\gamma$	resolved	0.5-4.5 TeV	Phys. Lett. B 822 (2021) 136651
	Bulk RS ($k/\overline{M}_{Pl} = 1.0$)	$G_{KK} \to ZZ \to \ell\ell\ell'\ell', \nu\nu\ell\ell$	resolved	0.6-1.8 TeV	Eur. Phys. J. C 81 (2021) 332
	Bulk RS ($k/\overline{M}_{Pl} = 1.0$)	$G_{KK} \to WW \to e\nu\mu\nu$	resolved	0.3-1.3 TeV	ATLAS-CONF-2022-066
	Bulk RS ($k/\overline{M}_{Pl} = 1.0$)	$G_{KK} \to WW, ZZ \to \nu\nu qq, \ell\nu qq, \ell\ell qq$	resolved, boosted	0.3-2.0 TeV	Eur. Phys. J. C 80 (2020) 1165
	Bulk RS ($k/\overline{M}_{Pl} = 1.0$)	$G_{KK} \to WW, ZZ \to qqqq$	boosted	1.3-1.8 TeV	JHEP 06 (2020) 042
Gauge bosons	HVT model A	$W' \to WZ \to \ell\nu\ell'\ell'$	resolved	0.3-2.4 TeV	arXiv:2207.03925
	HVT model A	$W' \to WZ \to \nu\nu qq, \ell\nu qq, \ell\ell qq$	resolved, boosted	0.3-3.9 TeV	Eur. Phys. J. C 80 (2020) 1165
	HVT model A	$W' \to WH \to \ell\nu bb$	resolved, boosted	0.4-3.0 TeV	arXiv:2207.00230
	HVT model A	$W' \to WZ \to qqqq$	boosted	1.3-3.4 TeV	JHEP 06 (2020) 042
	HVT model A	$W' \to WH \to qqbb$	boosted	1.5-2.9 TeV	Phys. Rev. D 102 (2020) 112008
	HVT model A	$Z' \to WW \to e\nu\mu\nu$	resolved	0.3-2.1 TeV	ATLAS-CONF-2022-066
	HVT model A	$Z' \to WW \to \ell\nu qq$	resolved, boosted	0.3-3.5 TeV	Eur. Phys. J. C 80 (2020) 1165
	HVT model A	$Z' \to ZH \to \nu\nu bb, \ell\ell bb$	resolved, boosted	0.3-2.8 TeV	arXiv:2207.00230
	HVT model A	$Z' \to WW \to qqqq$	boosted	1.3-2.9 TeV	JHEP 06 (2020) 042
	HVT model A	$Z' \to ZH \to qqbb$	boosted	1.5-2.2 TeV	Phys. Rev. D 102 (2020) 112008
	HVT model B	$W' \to WZ \to \ell\nu\ell'\ell'$	resolved	0.8-2.6 TeV	arXiv:2207.03925
	HVT model B	$W' \to WZ \to \nu\nu qq, \ell\nu qq, \ell\ell qq$	resolved, boosted	0.8-4.3 TeV	Eur. Phys. J. C 80 (2020) 1165
	HVT model B	$W' \to WH \to \ell\nu bb$	resolved, boosted	0.8-3.3 TeV	arXiv:2207.00230
	HVT model B	$W' \to WZ \to qqqq$	boosted	1.3-3.6 TeV	JHEP 06 (2020) 042
	HVT model B	$W' \to WH \to qqbb$	boosted	1.5-3.2 TeV	Phys. Rev. D 102 (2020) 112008
	HVT model B	$Z' \to WW \to e\nu\mu\nu$	resolved	0.8-2.4 TeV	ATLAS-CONF-2022-066
	HVT model B	$Z' \to WW \to \ell\nu qq$	resolved, boosted	0.8-3.9 TeV	Eur. Phys. J. C 80 (2020) 1165
	HVT model B	$Z' \to ZH \to \nu\nu bb, \ell\ell bb$	resolved, boosted	0.8-3.2 TeV	arXiv:2207.00230
	HVT model B	$Z' \to WW \to qqqq$	boosted	1.3-3.1 TeV	JHEP 06 (2020) 042
	HVT model B	$Z' \to ZH \to qqbb$	boosted	1.5-2.7 TeV	Phys. Rev. D 102 (2020) 112008
	HVT model C	$W' \to WZ \to \ell\nu\ell'\ell'$	resolved	0.3-0.34 TeV	arXiv:2207.03925

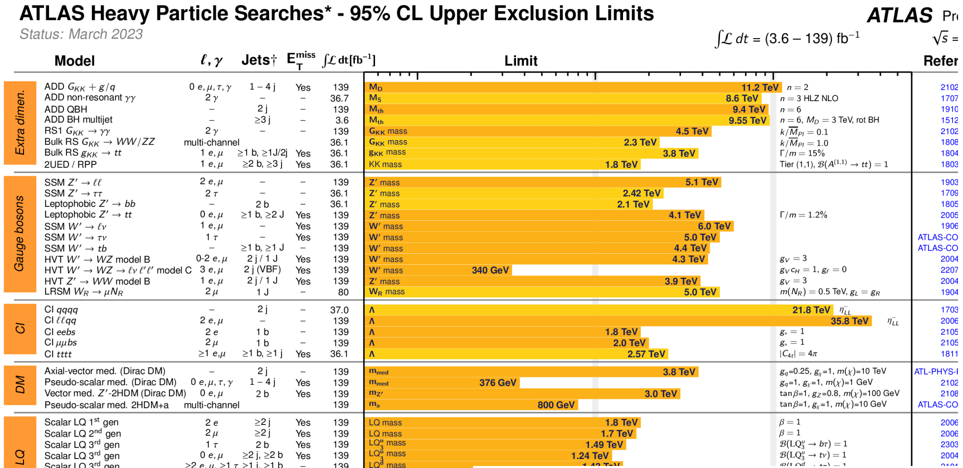
HVT model A: $g_F = -0.55$, $g_H = -0.56$

HVT model B: $g_F = 0.14$, $g_H = -2.9$

HVT model C: $g_F = 0$, $g_H = 1$

**small-radius (large-radius) jets are used in resolved (boosted) events*

†*with* $\ell = \mu, e$



ATLAS Heavy Particle Searches* - 95% CL Upper Exclusion Limits

Status: March 2023

ATLAS Pr

$\int\mathcal{L}\,dt$ = (3.6 – 139) fb^{-1}

$\sqrt{s}$ =

	Model	ℓ, γ	Jets†	E_T^{miss}	$\int\mathcal{L}\,dt[fb^{-1}]$	Limit			Refer
Extra dimen.	ADD $G_{KK} + g/q$	0 e, μ, τ, γ	1 – 4 j	Yes	139	M_D	11.2 TeV	$n = 2$	2102
	ADD non-resonant $\gamma\gamma$	2 γ	–	–	36.7	M_S	8.6 TeV	$n = 3$ HLZ NLO	1707
	ADD QBH	–	2 j	–	139	M_{th}	9.4 TeV	$n = 6$	1910
	ADD BH multijet	–	≥3 j	–	3.6	M_{th}	9.55 TeV	$n = 6$, $M_D = 3$ TeV, rot BH	1512
	RS1 $G_{KK} \to \gamma\gamma$	2 γ	–	–	139	G_{KK} mass	4.5 TeV	$k/\overline{M}_{Pl} = 0.1$	2102
	Bulk RS $G_{KK} \to WW/ZZ$	multi-channel			36.1	G_{KK} mass	2.3 TeV	$k/\overline{M}_{Pl} = 1.0$	1808
	Bulk RS $g_{KK} \to tt$	1 e, μ	≥1 b, ≥1J/2j	Yes	36.1	g_{KK} mass	3.8 TeV	$\Gamma/m = 15\%$	1804
	2UED / RPP	1 e, μ	≥2 b, ≥3 j	Yes	36.1	KK mass	1.8 TeV	Tier (1,1), $\mathcal{B}(A^{(1,1)} \to tt) = 1$	1803
Gauge bosons	SSM $Z' \to \ell\ell$	2 e, μ	–	–	139	Z' mass	5.1 TeV		1903
	SSM $Z' \to \tau\tau$	2 τ	–	–	36.1	Z' mass	2.42 TeV		1709
	Leptophobic $Z' \to bb$	–	2 b	–	36.1	Z' mass	2.1 TeV		1805
	Leptophobic $Z' \to tt$	0 e, μ	≥1 b, ≥2 J	Yes	139	Z' mass	4.1 TeV	$\Gamma/m = 1.2\%$	2005
	SSM $W' \to \ell\nu$	1 e, μ	–	Yes	139	W' mass	6.0 TeV		1906
	SSM $W' \to \tau\nu$	1 τ	–	Yes	139	W' mass	5.0 TeV		ATLAS-CO
	SSM $W' \to tb$	–	≥1 b, ≥1 J	–	139	W' mass	4.4 TeV		ATLAS-CO
	HVT $W' \to WZ$ model B	0-2 e, μ	2 j / 1 J	Yes	139	W' mass	4.3 TeV	$g_V = 3$	2004
	HVT $W' \to WZ \to \ell\nu\,\ell'\ell'$ model C	3 e, μ	2 j (VBF)	Yes	139	W' mass	340 GeV	$g_V c_H = 1$, $g_f = 0$	2207
	HVT $Z' \to WW$ model B	1 e, μ	2 j / 1 J	Yes	139	Z' mass	3.9 TeV	$g_V = 3$	2004
	LRSM $W_R \to \mu N_R$	2 μ	1 J	–	80	W_R mass	5.0 TeV	$m(N_R) = 0.5$ TeV, $g_L = g_R$	1904
CI	CI $qqqq$	–	2 j	–	37.0	Λ	21.8 TeV	η_{LL}^{-}	1703
	CI $\ell\ell qq$	2 e, μ	–	–	139	Λ	35.8 TeV	η_{LL}^{-}	2006
	CI $eebs$	2 e	1 b	–	139	Λ	1.8 TeV	$g_* = 1$	2105
	CI $\mu\mu bs$	2 μ	1 b	–	139	Λ	2.0 TeV	$g_* = 1$	2105
	CI $tttt$	≥1 e,μ	≥1 b, ≥1 j	Yes	36.1	Λ	2.57 TeV	$\lvert C_{4t}\rvert = 4\pi$	1811
DM	Axial-vector med. (Dirac DM)	–	2 j	–	139	m_{med}	3.8 TeV	g_q=0.25, g_χ=1, $m(\chi)$=10 TeV	ATL-PHYS-I
	Pseudo-scalar med. (Dirac DM)	0 e, μ, τ, γ	1 – 4 j	Yes	139	m_{med}	376 GeV	g_q=1, g_χ=1, $m(\chi)$=1 GeV	2102
	Vector med. Z'-2HDM (Dirac DM)	0 e, μ	2 b	Yes	139	$m_{Z'}$	3.0 TeV	tanβ=1, g_Z=0.8, $m(\chi)$=100 GeV	2108
	Pseudo-scalar med. 2HDM+a	multi-channel			139	m_a	800 GeV	tanβ=1, g_χ=1, $m(\chi)$=10 GeV	ATLAS-CO
LQ	Scalar LQ 1st gen	2 e	≥2 j	Yes	139	LQ mass	1.8 TeV	$\beta = 1$	2006
	Scalar LQ 2nd gen	2 μ	≥2 j	Yes	139	LQ mass	1.7 TeV	$\beta = 1$	2006
	Scalar LQ 3rd gen	1 τ	2 b	Yes	139	LQ_3^u mass	1.49 TeV	$\mathcal{B}(LQ_3^u \to b\tau) = 1$	2303
	Scalar LQ 3rd gen	0 e, μ	≥2 j, ≥2 b	Yes	139	LQ_3^u mass	1.24 TeV	$\mathcal{B}(LQ_3^u \to t\nu) = 1$	2004
	Scalar LQ 3rd gen	≥2 e, μ, ≥1 τ	≥1 j, ≥1 b	–	139	LQ_3^d mass			
	Scalar LQ 3rd gen	0 e, μ, ≥1 τ	0 – 2 j, 2 b	Yes	139	LQ_3^d mass			
	Vector LQ mix gen	multi-channel	≥1 j, ≥1 b	Yes	139	LQ_3^V mass			
	Vector LQ 3rd gen	2 e, μ, τ	≥1 b	Yes	139	LQ_3^V mass			
Vector-like fermions	VLQ $TT \to Zt + X$	2e/2μ/≥3e,μ	≥1 b, ≥1 j	–	139	T mass			
	VLQ $BB \to Wt/Zb + X$	multi-channel			36.1	B mass			
	VLQ $T_{5/3}T_{5/3}\vert T_{5/3} \to Wt + X$	2(SS)/≥3 e,μ	≥1 b, ≥1 j	Yes	36.1	$T_{5/3}$ mass			
	VLQ $T \to Ht/Zt$	1 e, μ	≥1 b, ≥3 j	Yes	139	T mass			
	VLQ $Y \to Wb$	1 e, μ	≥1 b, ≥1 j	Yes	36.1	Y mass			
	VLQ $B \to Hb$	0 e,μ	≥2b, ≥1j, ≥1J	–	139	B mass			
	VLL $\tau' \to Z\tau/H\tau$	multi-channel	≥1 j	Yes	139	τ' mass			
Exctd ferm.	Excited quark $q^* \to qg$	–	2 j	–	139	q^* mass			
	Excited quark $q^* \to q\gamma$	1 γ	1 j	–	36.7	q^* mass			
	Excited quark $b^* \to bg$	–	1 b, 1 j	–	139	b^* mass			
	Excited lepton τ^*	2 τ	≥2 j	–	139	τ^* mass			
Other	Type III Seesaw	2,3,4 e, μ	≥2 j	Yes	139	N^0 mass			
	LRSM Majorana ν	2 μ	2 j	–	36.1	N_R mass			
	Higgs triplet $H^{\pm\pm} \to W^\pm W^\pm$	2,3,4 e, μ (SS)	various	Yes	139	$H^{\pm\pm}$ mass			
	Higgs triplet $H^{\pm\pm} \to \ell\ell$	2,3,4 e, μ (SS)	–	–	139	$H^{\pm\pm}$ mass			
	Multi-charged particles	–	–	–	139	multi-charged particle mass			
	Magnetic monopoles	–	–	–	34.4	monopole mass			

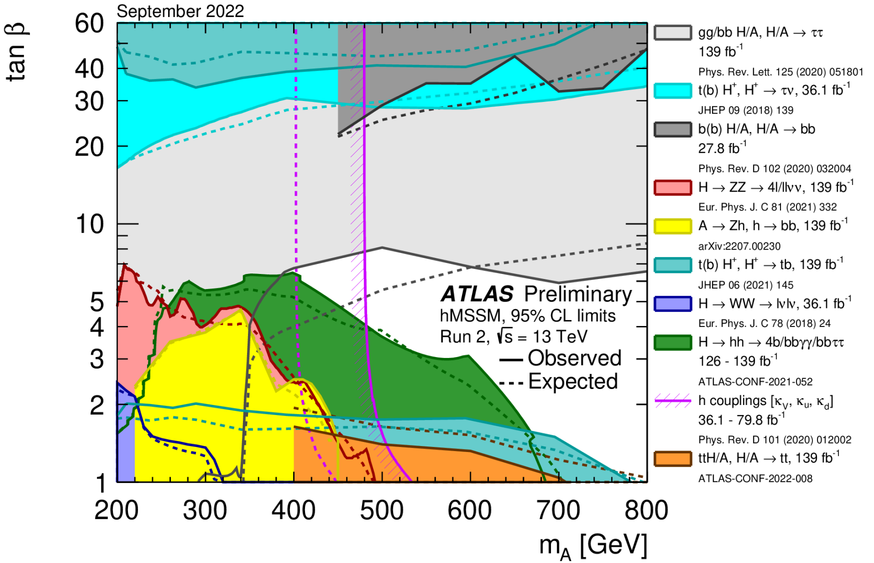
$\sqrt{s}$ = 13 TeV partial data

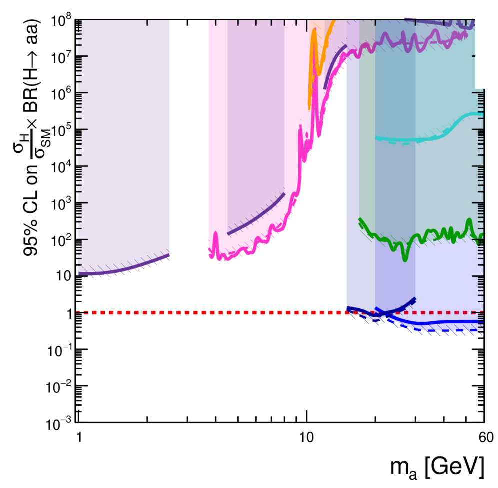
$\sqrt{s}$ = 13 TeV full data

10^{-1}

*Only a selection of the available mass limits on new states or phenomena is shown.

†Small-radius (large-radius) jets are denoted by the letter j (J).





Run 2: $\sqrt{s}$ = 13 TeV

2HDM+S Type-IV, tan*β *= 5

expected ± **1** σ

observed

Run 1 20.3 fb^{-1} H→ aa–
PRD 92 (2015) 052002

Run 1 20.3 fb^{-1} H→ aa–
EPJC 76 (2016) 210

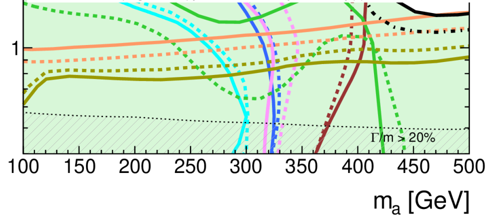
Run 2 36.1 fb^{-1} H→ aa–
JHEP 06 (2018) 166

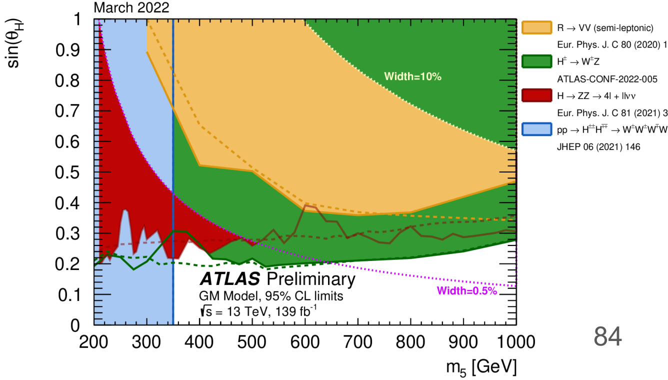
Run 2 36.1 fb^{-1} H→ aa–
JHEP 10 (2018) 031

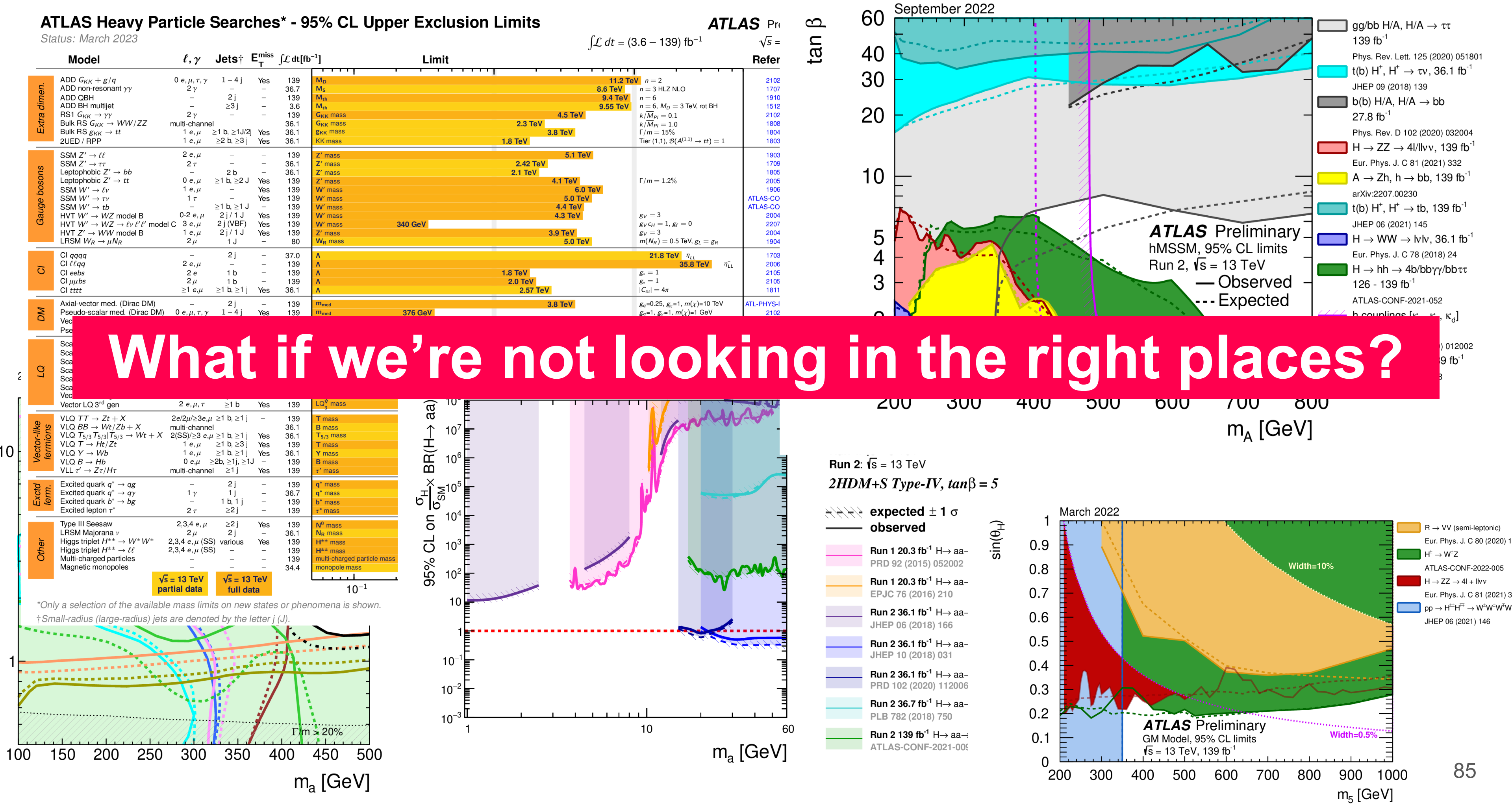
Run 2 36.1 fb^{-1} H→ aa–
PRD 102 (2020) 112006

Run 2 36.7 fb^{-1} H→ aa–
PLB 782 (2018) 750

Run 2 139 fb^{-1} H→ aa→
ATLAS-CONF-2021-00







What if we’re not looking in the right places?

ATLAS Heavy Particle Searches* - 95% CL Upper Exclusion Limits

Status: March 2023

ATLAS Pr

$\int\mathcal{L}\,dt$ = (3.6 – 139) fb^{-1}

$\sqrt{s}$ =

	Model	ℓ,γ	Jets†	E_T^{miss}	$\int\mathcal{L}\,dt[fb^{-1}]$	Limit			Refer
Extra dimen.	ADD $G_{KK}+g/q$	0 e,μ,τ,γ	1 – 4 j	Yes	139	M_D	11.2 TeV	$n=2$	2102
	ADD non-resonant $\gamma\gamma$	2 γ	–	–	36.7	M_S	8.6 TeV	$n=3$ HLZ NLO	1707
	ADD QBH	–	2 j	–	139	M_{th}	9.4 TeV	$n=6$	1910
	ADD BH multijet	–	≥3 j	–	3.6	M_{th}	9.55 TeV	$n=6$, $M_D=3$ TeV, rot BH	1512
	RS1 $G_{KK}\to\gamma\gamma$	2 γ	–	–	139	G_{KK} mass	4.5 TeV	$k/\overline{M}_{Pl}=0.1$	2102
	Bulk RS $G_{KK}\to WW/ZZ$	multi-channel			36.1	G_{KK} mass	2.3 TeV	$k/\overline{M}_{Pl}=1.0$	1808
	Bulk RS $g_{KK}\to tt$	1 e,μ	≥1 b, ≥1J/2j	Yes	36.1	g_{KK} mass	3.8 TeV	$\Gamma/m=15\%$	1804
	2UED / RPP	1 e,μ	≥2 b, ≥3 j	Yes	36.1	KK mass	1.8 TeV	Tier (1,1), $\mathcal{B}(A^{(1,1)}\to tt)=1$	1803
Gauge bosons	SSM $Z'\to\ell\ell$	2 e,μ	–	–	139	Z' mass	5.1 TeV		1903
	SSM $Z'\to\tau\tau$	2 τ	–	–	36.1	Z' mass	2.42 TeV		1709
	Leptophobic $Z'\to bb$	–	2 b	–	36.1	Z' mass	2.1 TeV		1805
	Leptophobic $Z'\to tt$	0 e,μ	≥1 b, ≥2 J	Yes	139	Z' mass	4.1 TeV	$\Gamma/m=1.2\%$	2005
	SSM $W'\to\ell\nu$	1 e,μ	–	Yes	139	W' mass	6.0 TeV		1906
	SSM $W'\to\tau\nu$	1 τ	–	Yes	139	W' mass	5.0 TeV		ATLAS-CO
	SSM $W'\to tb$	–	≥1 b, ≥1 J	–	139	W' mass	4.4 TeV		ATLAS-CO
	HVT $W'\to WZ$ model B	0-2 e,μ	2 j / 1 J	Yes	139	W' mass	4.3 TeV	$g_V=3$	2004
	HVT $W'\to WZ\to\ell\nu\,\ell'\ell'$ model C	3 e,μ	2 j (VBF)	Yes	139	W' mass	340 GeV	$g_V c_H=1$, $g_f=0$	2207
	HVT $Z'\to WW$ model B	1 e,μ	2 j / 1 J	Yes	139	Z' mass	3.9 TeV	$g_V=3$	2004
	LRSM $W_R\to\mu N_R$	2 μ	1 J	–	80	W_R mass	5.0 TeV	$m(N_R)=0.5$ TeV, $g_L=g_R$	1904
CI	CI $qqqq$	–	2 j	–	37.0	Λ	21.8 TeV	η^-_{LL}	1703
	CI $\ell\ell qq$	2 e,μ	–	–	139	Λ	35.8 TeV	η^-_{LL}	2006
	CI $eebs$	2 e	1 b	–	139	Λ	1.8 TeV	$g_*=1$	2105
	CI $\mu\mu bs$	2 μ	1 b	–	139	Λ	2.0 TeV	$g_*=1$	2105
	CI $tttt$	≥1 e,μ	≥1 b, ≥1 j	Yes	36.1	Λ	2.57 TeV	$\lvert C_{4t}\rvert=4\pi$	1811
DM	Axial-vector med. (Dirac DM)	–	2 j	–	139	m_{med}	3.8 TeV	g_q=0.25, g_χ=1, $m(\chi)$=10 TeV	ATL-PHYS-I
	Pseudo-scalar med. (Dirac DM)	0 e,μ,τ,γ	1 – 4 j	Yes	139	m_{med}	376 GeV	g_q=1, g_χ=1, $m(\chi)$=1 GeV	2102
	Vec								
	Pse								
LQ	Sca								
	Sca								
	Sca								
	Sca								
	Sca								
	Sca								
	Vec								
	Vector LQ 3rd gen	2 e,μ,τ	≥1 b	Yes	139	LQ_3^V mass			
Vector-like fermions	VLQ $TT\to Zt+X$	$2e/2\mu/\geq3e,\mu$	≥1 b, ≥1 j	–	139	T mass			
	VLQ $BB\to Wt/Zb+X$	multi-channel			36.1	B mass			
	VLQ $T_{5/3}T_{5/3}\lvert T_{5/3}\to Wt+X$	2(SS)/≥3 e,μ	≥1 b, ≥1 j	Yes	36.1	$T_{5/3}$ mass			
	VLQ $T\to Ht/Zt$	1 e,μ	≥1 b, ≥3 j	Yes	139	T mass			
	VLQ $Y\to Wb$	1 e,μ	≥1 b, ≥1 j	Yes	36.1	Y mass			
	VLQ $B\to Hb$	0 e,μ	≥2b, ≥1j, ≥1J	–	139	B mass			
	VLL $\tau'\to Z\tau/H\tau$	multi-channel	≥1 j	Yes	139	τ' mass			
Exctd ferm.	Excited quark $q^*\to qg$	–	2 j	–	139	q^* mass			
	Excited quark $q^*\to q\gamma$	1 γ	1 j	–	36.7	q^* mass			
	Excited quark $b^*\to bg$	–	1 b, 1 j	–	139	b^* mass			
	Excited lepton τ^*	2 τ	≥2 j	–	139	τ^* mass			
Other	Type III Seesaw	2,3,4 e,μ	≥2 j	Yes	139	N^0 mass			
	LRSM Majorana ν	2 μ	2 j	–	36.1	N_R mass			
	Higgs triplet $H^{\pm\pm}\to W^\pm W^\pm$	2,3,4 e,μ (SS)	various	Yes	139	$H^{\pm\pm}$ mass			
	Higgs triplet $H^{\pm\pm}\to\ell\ell$	2,3,4 e,μ (SS)	–	–	139	$H^{\pm\pm}$ mass			
	Multi-charged particles	–	–	–	139	multi-charged particle mass			
	Magnetic monopoles	–	–	–	34.4	monopole mass			

$\sqrt{s}$ = 13 TeV partial data

$\sqrt{s}$ = 13 TeV full data

10^{-1}

*Only a selection of the available mass limits on new states or phenomena is shown.

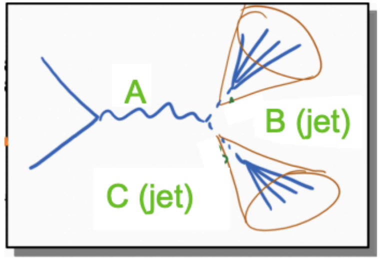
†Small-radius (large-radius) jets are denoted by the letter j (J).

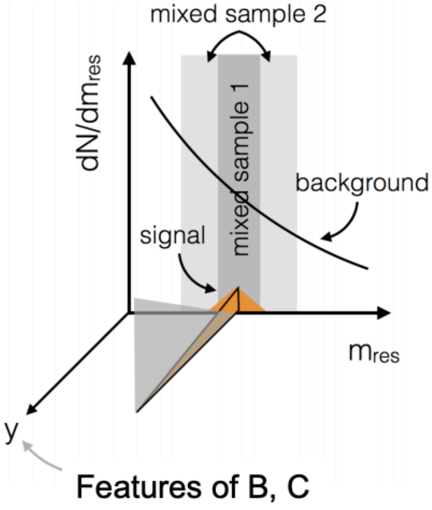
Bonus: anomaly detection

https://arxiv.org/abs/2005.02983

CERN seminar

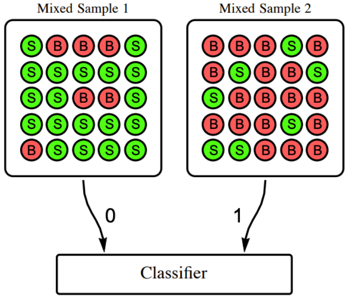
Weak supervision





CWoLa

Classification Without Labels



Mixed Sample 1

S	B	B	B	S
S	S	S	S	S
S	S	B	B	S
S	S	S	S	S
B	S	S	S	S

Mixed Sample 2

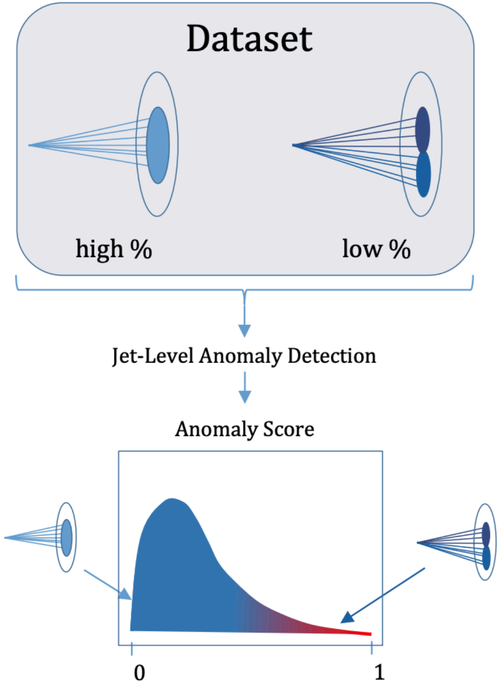
B	B	B	S	B
B	S	B	B	S
S	B	B	B	B
B	B	B	S	B
S	S	B	B	B

Bonus: anomaly detection

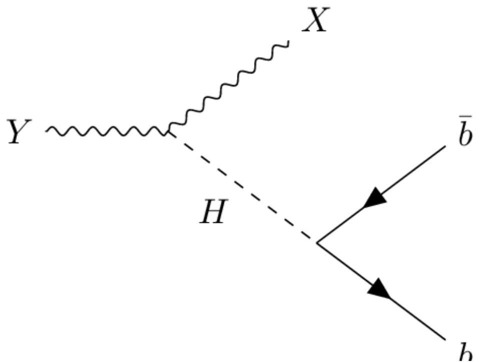
https://arxiv.org/abs/2105.09274v1

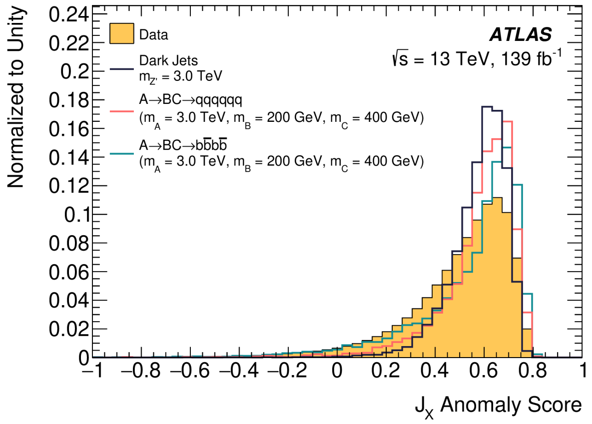
https://arxiv.org/abs/2306.03637

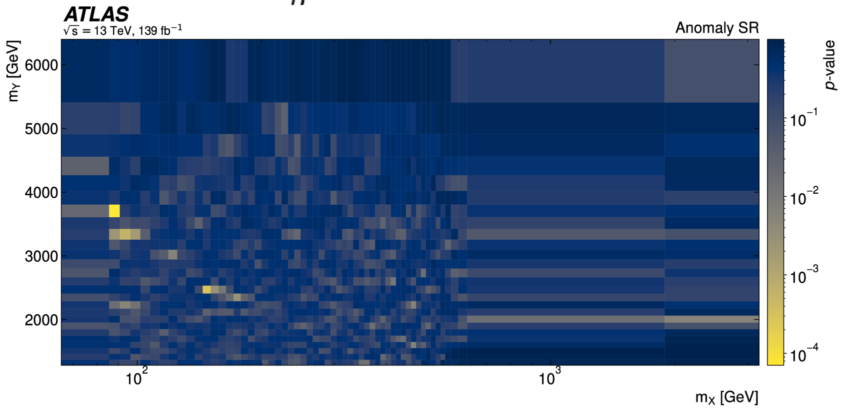
Unsupervised learning



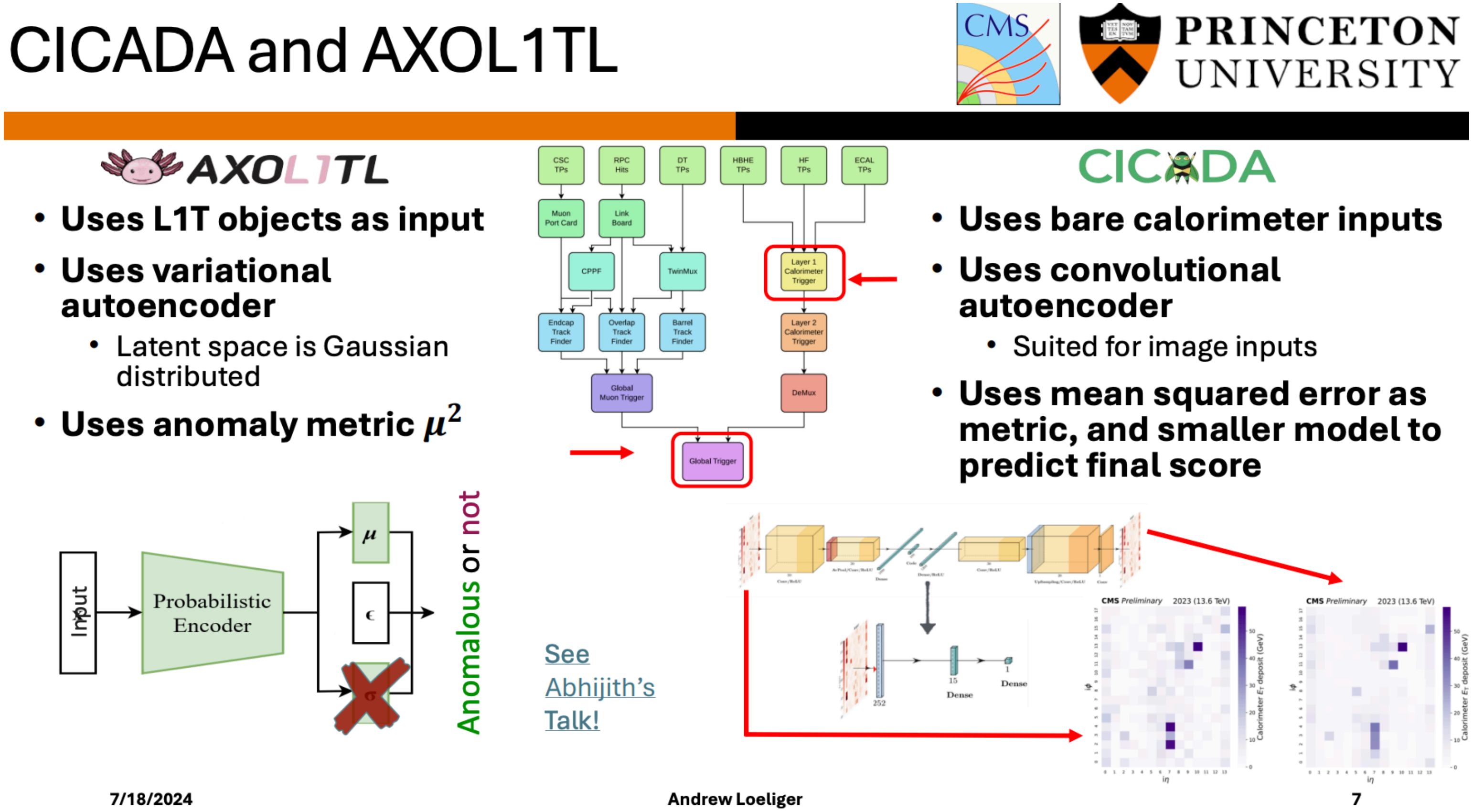
2D search mX,mY
Anomaly score for identifying X







Bonus: anomaly detection in the *trigger*



CICADA and AXOL1TL

AXOL1TL

- **Uses L1T objects as input**
- **Uses variational autoencoder**
  - Latent space is Gaussian distributed
- **Uses anomaly metric μ^2**

CICADA

- **Uses bare calorimeter inputs**
- **Uses convolutional autoencoder**
  - Suited for image inputs
- **Uses mean squared error as metric, and smaller model to predict final score**

See Abhijith's Talk!

LHC / HL-LHC Plan

You are here

LHC

HL-LHC

Run 1 | Run 2 | Run 3 | Run 4 - 5...

LS1

splice consolidation
button collimators
R2E project

EYETS

LS2

Diodes Consolidation
LIU Installation
Civil Eng. P1-P5

LS3

HL-LHC
installation

7 TeV

8 TeV

13 TeV

13.6 TeV

13.6 - 14 TeV

energy

cryolimit
interaction
regions

pilot beam

inner triplet
radiation limit

2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2041

experiment
beam pipes

ATLAS - CMS
upgrade *phase 1*

ALICE - LHCb
upgrade

ATLAS - CMS
HL upgrade

75% nominal Lumi

nominal Lumi

2 x nominal Lumi

2 x nominal Lumi

5 to 7.5
x nominal Lumi

$30\ fb^{-1}$

$190\ fb^{-1}$

$500\ fb^{-1}$

integrated
luminosity

$3000\ fb^{-1}$
$4000\ fb^{-1}$

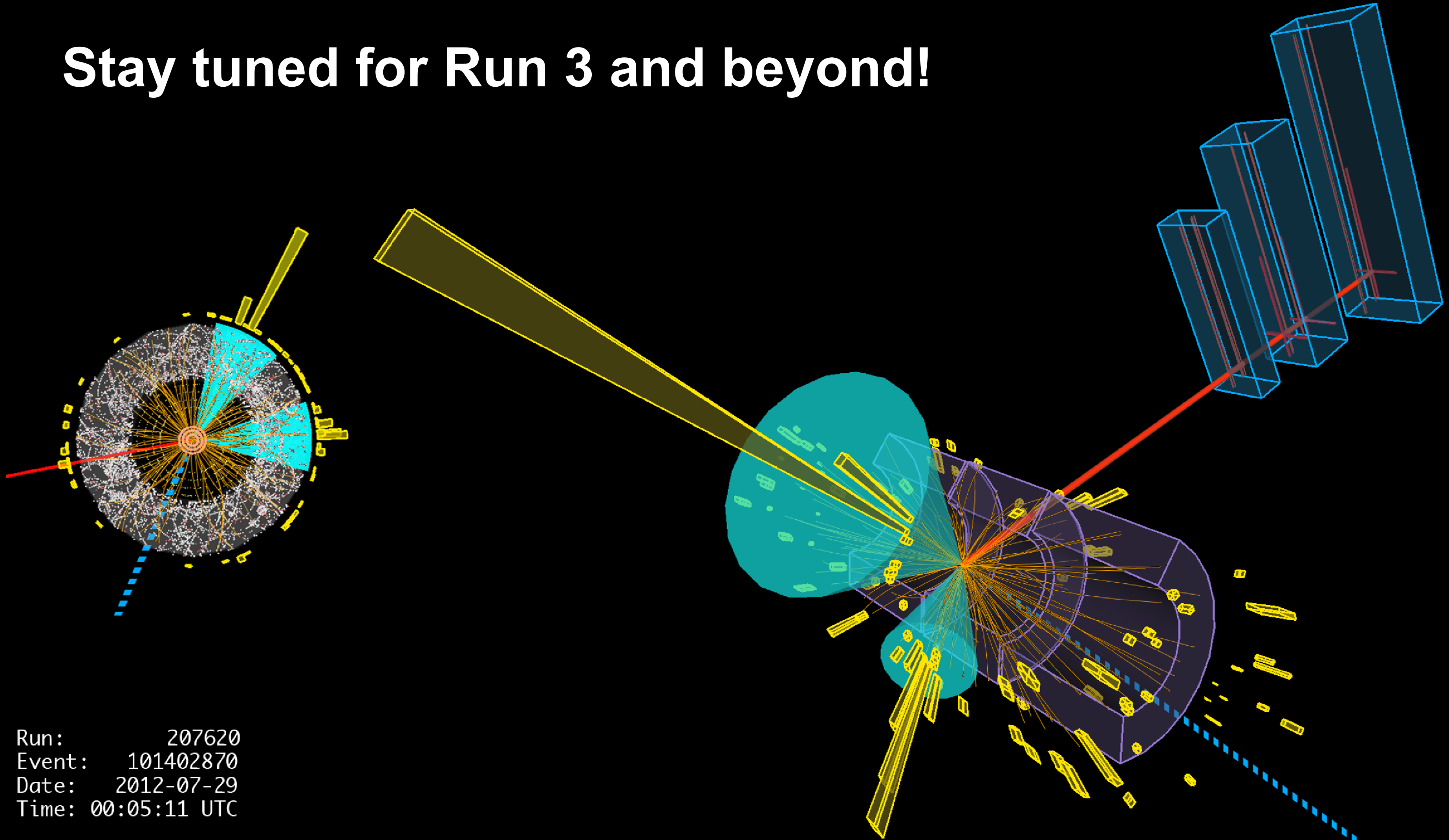
HL-LHC TECHNICAL EQUIPMENT:

DESIGN STUDY | PROTOTYPES | CONSTRUCTION | INSTALLATION & COMM. | PHYSICS

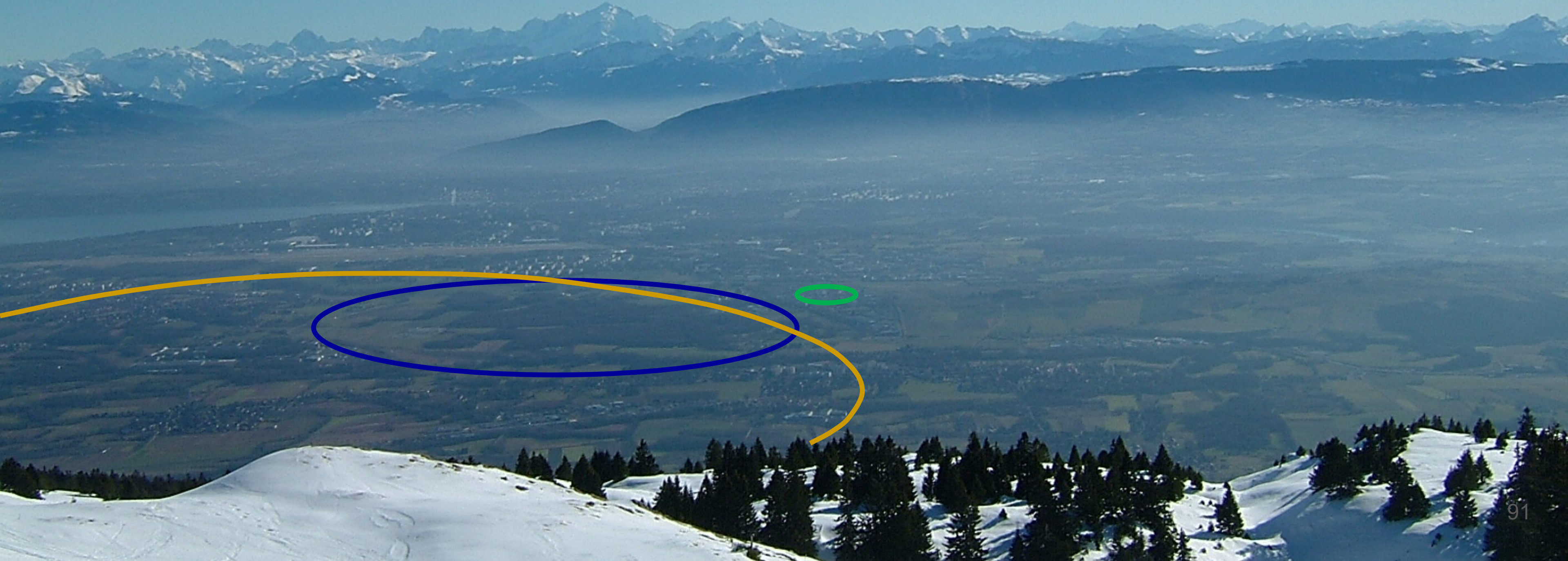
HL-LHC CIVIL ENGINEERING:

DEFINITION | EXCAVATION | BUILDINGS | CORES

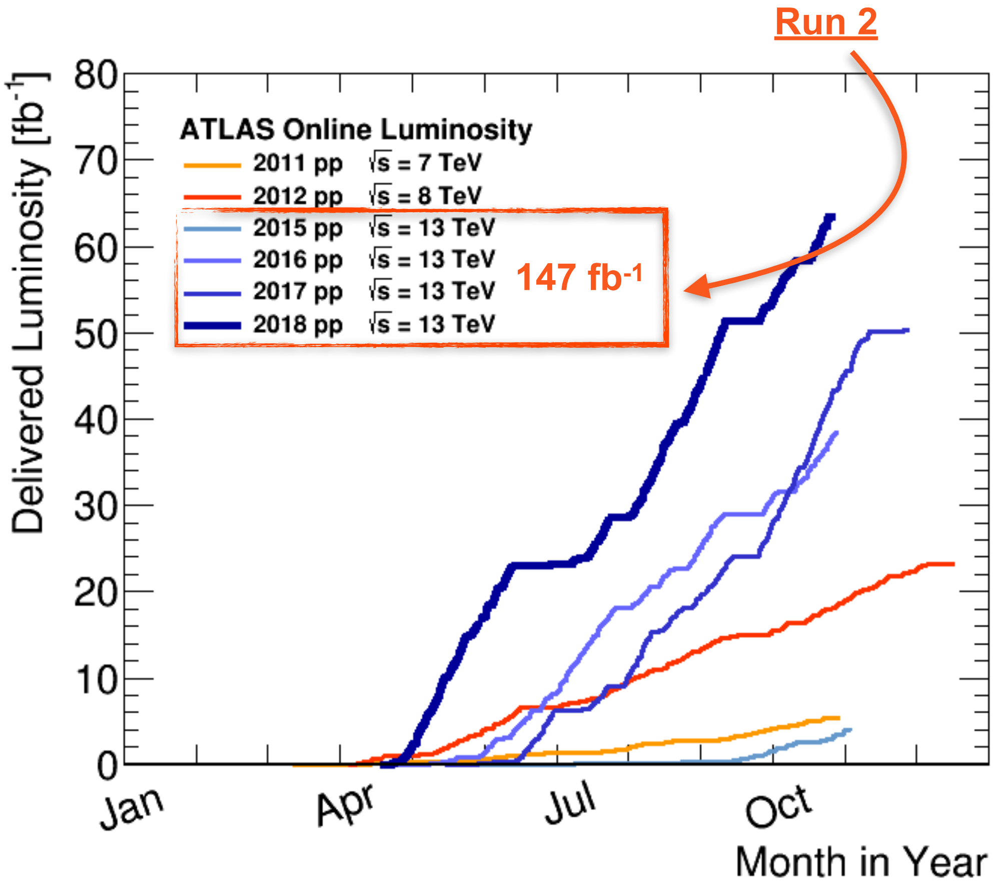
Stay tuned for Run 3 and beyond!

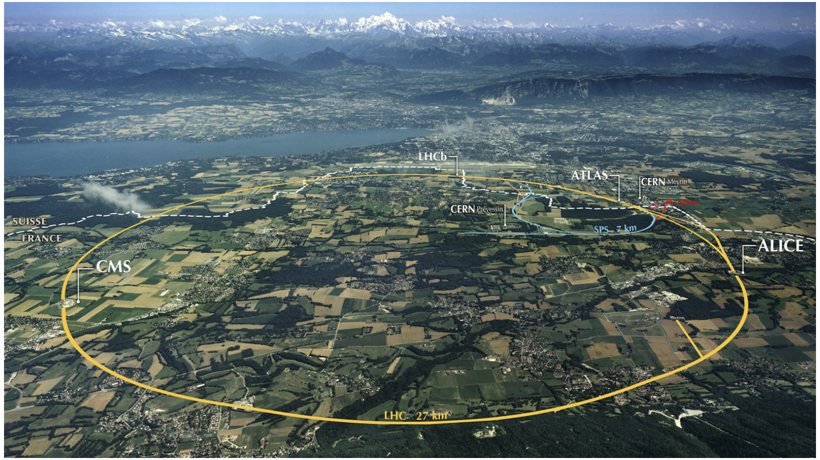


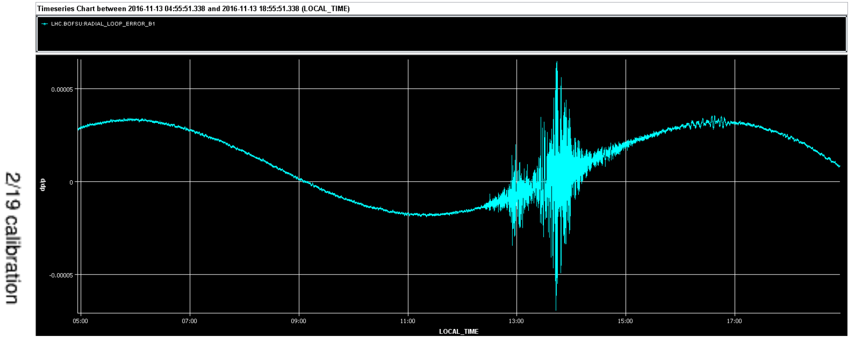
Backup



The LHC datasets







The New Zealand earthquake (2016).

LHC beam orbit displacement.