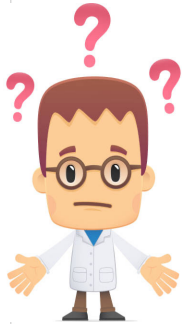
UNIVERSITÀ DEGLI STUDI DI NAPOLI
FEDERICO II

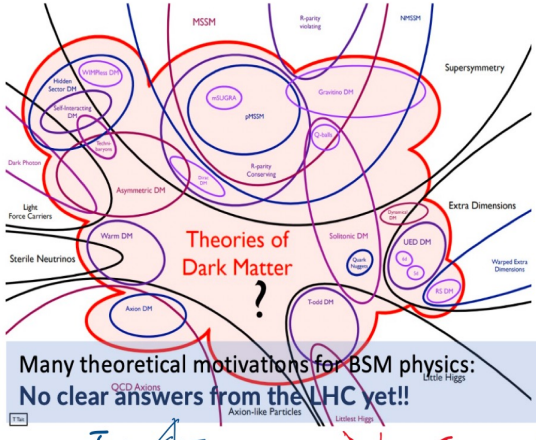
# MODEL AGNOSTIC SEARCHES IN FINAL STATES WITH JETS AT ATLAS

ANTONIO D'AVANZO, on behalf of the ATLAS Collaboration

29° Symposium on Particles, String and Cosmology (PASCOS 2024), 09/07/2024, Quy Nhon

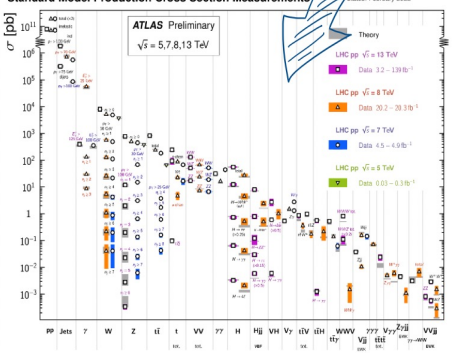
# Introduction



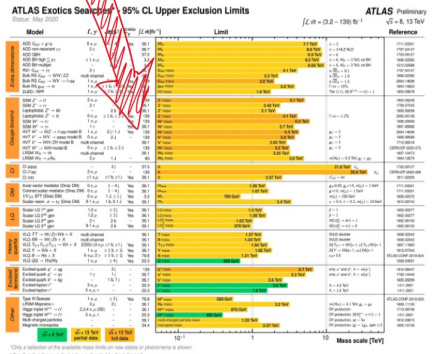


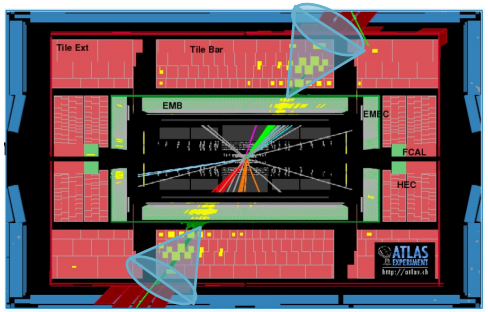

- Standard Model (SM) remarkably predictive of experimental results
  - discovery of the Higgs boson in 2012 by ATLAS and CMS
- Open questions: many Beyond Standard Model theories (Dark Matter, Gravity, Hierarchy problem ecc.)
- Search for new resonances decaying into hadronic final states jj (jets) → localized excesses (bumps) over expected background $m_{jj}$

measurements



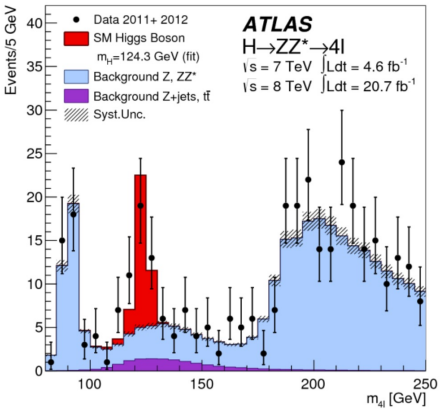

Searches




Hadronization scheme of quarks/gluons

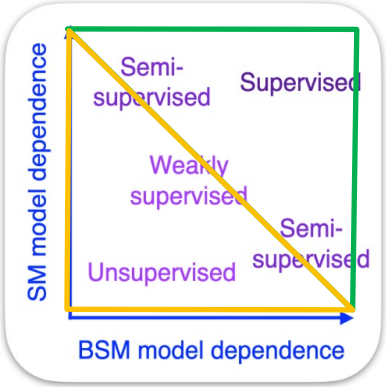
https://cds.cern.ch/record/1605823



# To be or not to be model-dependent?

Model dependent approach:

- A new well motivated physics-scenario is chosen
- The search is maximized based on signal signatures (supervised machine learning methods)
- Unlikely to be sensitive to different process




Model independent approach:

- Minimal assumptions of signal properties
- Deviations from background-only hypothesis (methods often provided by Machine Learning)
- Not optimal as model-dependent, but more prone to generality

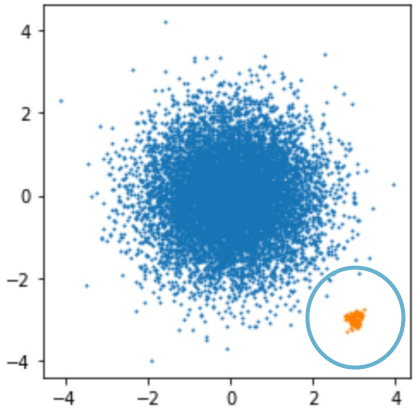
## In this review

- Full Run 2 (2015-2018, 140 $fb^{-1}$) of LHC data (beside n. 4), pp centre of mass energy 13 TeV
  - Results interpreted with 95% Confidence Levels

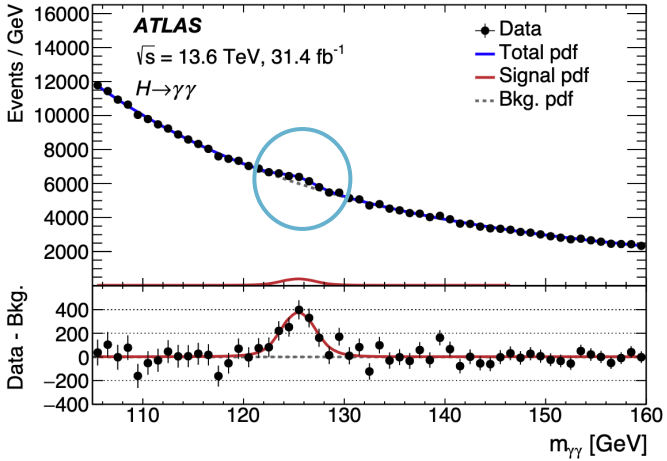
1. Search for new phenomena in dijet events using **quark tagging**
2. **Weakly-supervised anomaly detection** for resonant new physics in the dijet final state
3. **Anomaly detection** search for new resonances decaying into a Higgs boson and a generic new particle X in hadronic final states
4. Search for Low-Mass Dijet Resonances Using **Trigger Level Analysis**

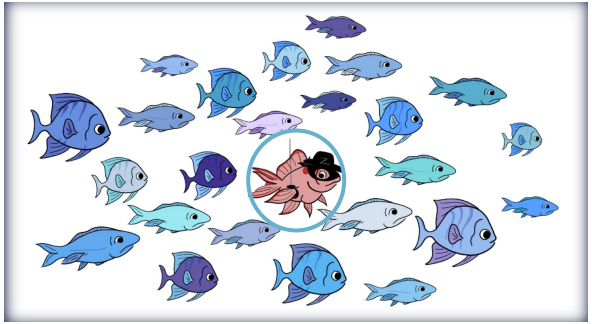
# Non supervised Anomaly Detection

- Anomaly Detection (AD) refers to Machine Learning (ML) techniques used to spot these outliers.
- Particle physics → Identification of features of detector data inconsistent with the expected background.
- Machine learning techniques exploited: semi-supervised (partial labels), weakly-supervised (noisy labels) and unsupervised (no labels)




Eur.Phys.J.C 84 (2024) 1, 78



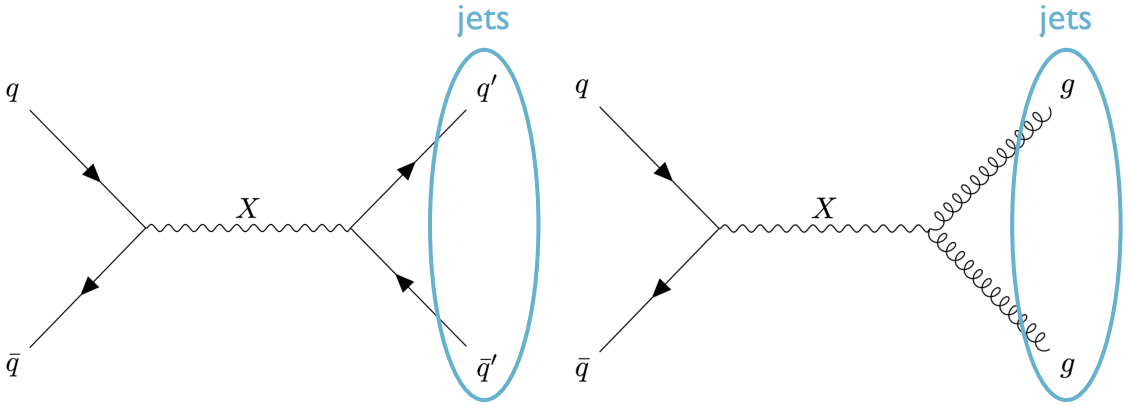
# Search for new phenomena in dijet events using quark tagging

# Search of new resonances in jet pairs

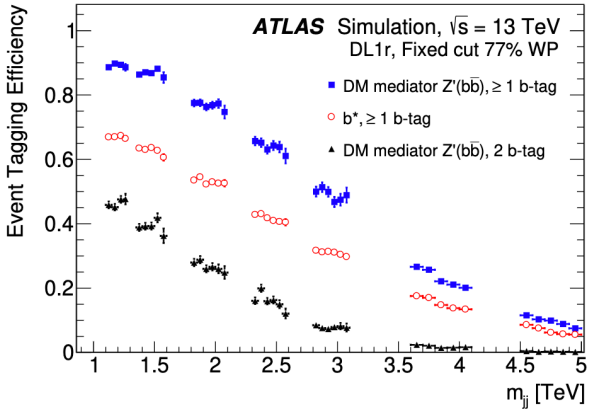
JHEP03(2020)145

- Search for resonant decays of heavy BSM particles strongly coupled to quarks/gluons
  - $m_{jj}$ spectrum ranges from 1.1 to 8 TeV
  - **3 signal regions**: Inclusive jets content and 1 or 2 b-jets required
  - Trigger efficiency cuts on jets kinematics, invariant mass and $y^* = \frac{y_1 - y_2}{2}$

- Results interpreted with many new physics scenarios, but also generic Gaussian-shaped narrow-resonance G($m_X, \sigma_X$)



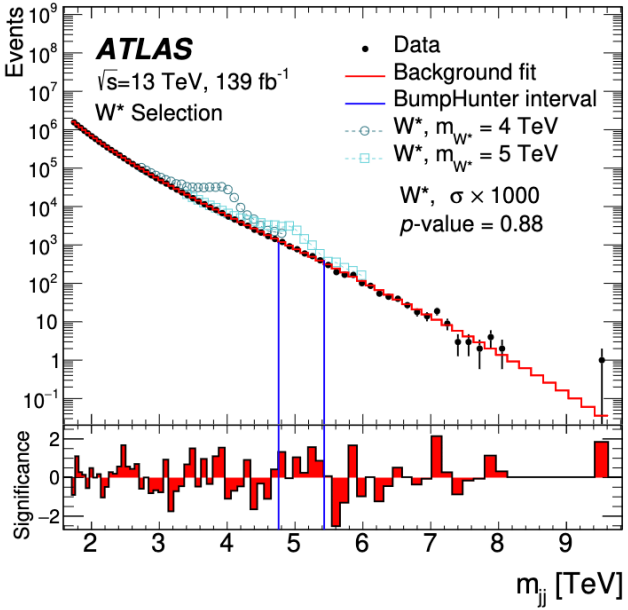

b-tagging: selection on discriminant variable from NN output

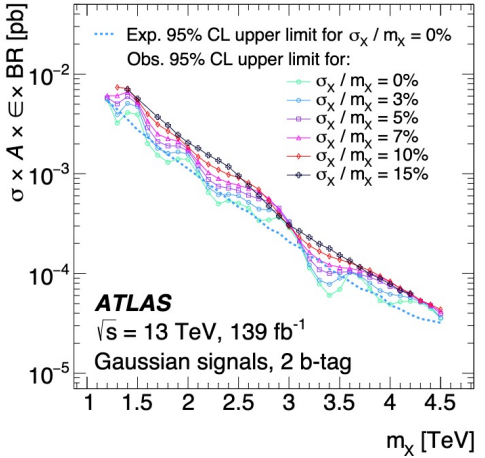


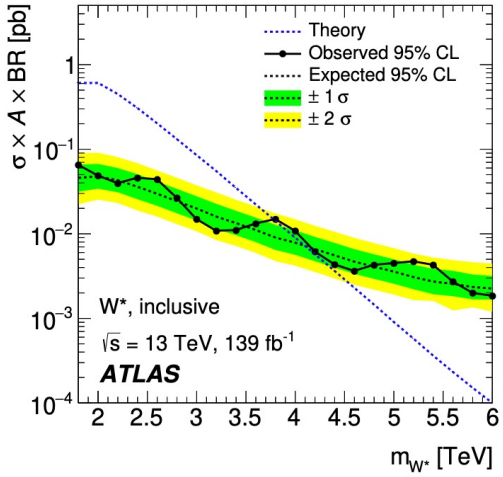
# Results

JHEP03(2020)145

- Main QCD background estimated with smoothly falling fit functions on the $m_{jj}$ distribution
- No significant deviation from background
  - Upper limits on cross sections estimated from fit considering the several signal hypothesis



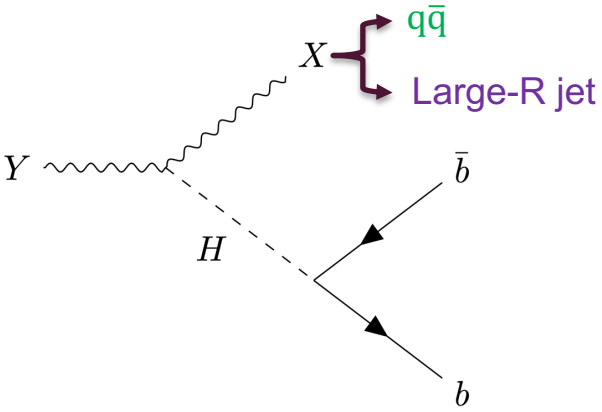



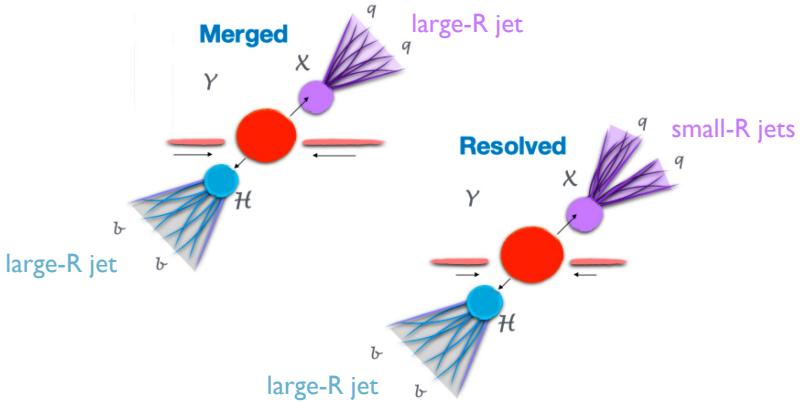



# Anomaly detection search for new resonances decaying into a Higgs boson and a generic new particle X in hadronic final states

# Y → XH overview

- Search for a heavy-mass resonance **Y** decaying in a Higgs boson (**H** → $b\bar{b}$) and a new particle **X** in the fully hadronic channel
- Mass range: $m_Y$ in 1 - 6 TeV range, $m_X$ in 65 - 3000 GeV range → boosted regime for H boson
- Signal regions:
  - Model dependent: 2-prong (X → $q\bar{q}$) boosted ($m_X/m_Y$ < 0.3) and resolved ($m_X/m_Y$ > 0.3)
  - Model independent: anomalous X hadronic decay in large-R jet



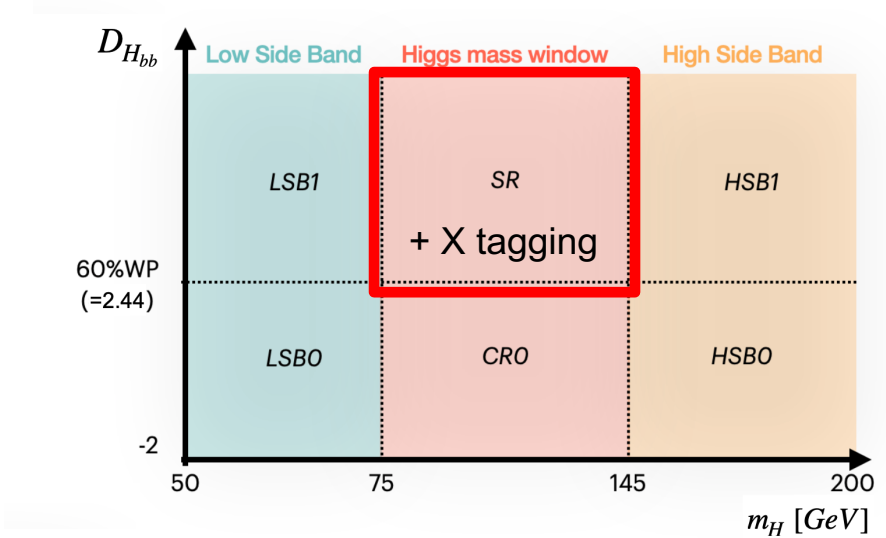



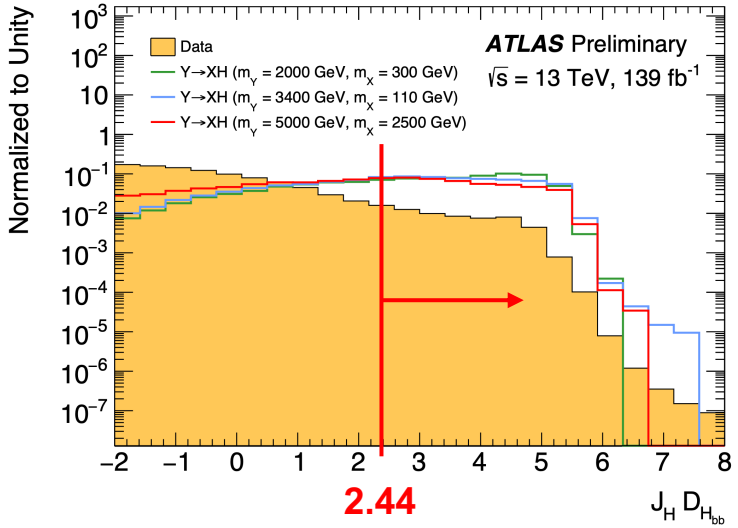

- Background is mainly composed of **QCD dijet events** (~97%), estimated fully data-driven (Machine Learning approach) → more in backup

# Model independent signal region

Phys.Rev.D 108 (2023) 052009

- X and H candidate associated to pT-leading and –subleading jets, ambiguity resolved by $H \rightarrow b\bar{b}$ tagger based on Deep Neural Network
  - Discriminant $D_{H_{bb}}$ score computed from NN outputs per jet → H candidate chosen by highest score criteria
- H candidate is further tagged if $D_{H_{bb}} > 2.44$
- X candidate tagged with discriminant from fully data-driven anomaly detection

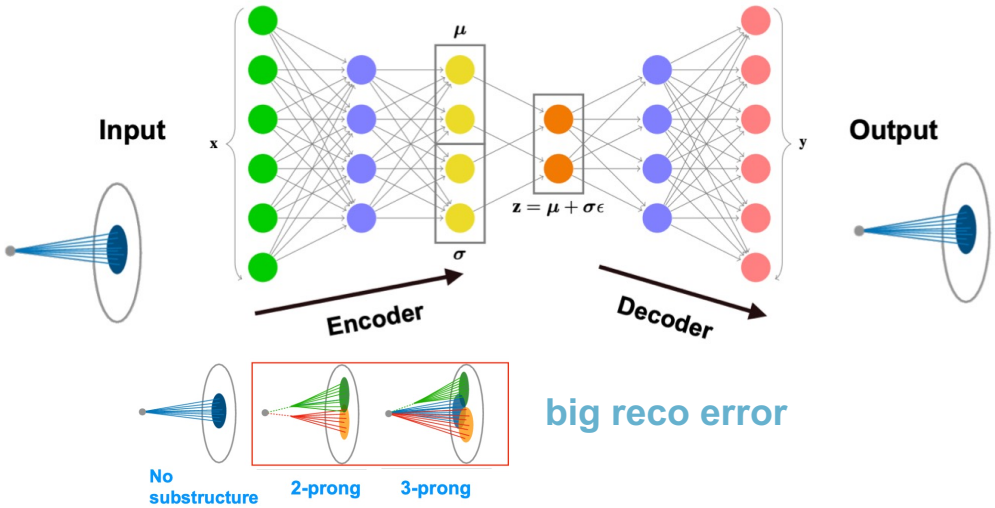


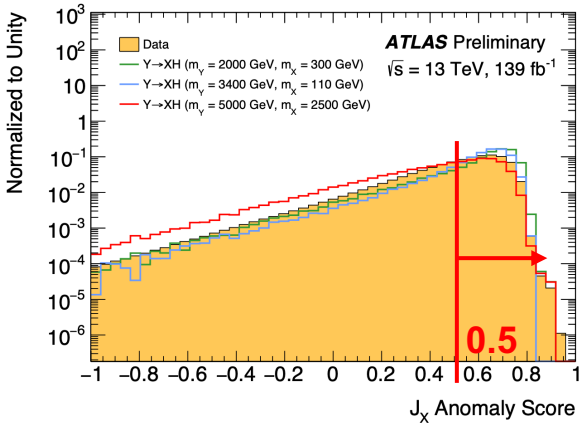



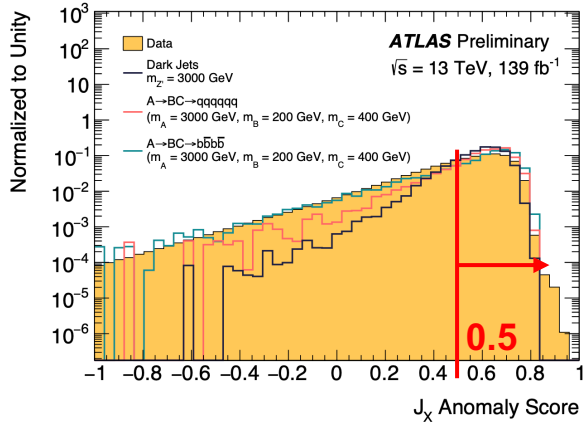
# Anomaly detection X tagging

Phys.Rev.D 108 (2023) 052009

- Fully unsupervised (first in ATLAS) variational recurrent neural network (VRNN)
  - Trained over **constituents of jets** with $p_T$ > 1.2 TeV modeled as sequence of four-vectors
- Anomaly score computed from VRNN output
  - Sensitive to alternative X decay hypothesis other than 2-prong (e.g. heavy flavor, three-prong and dark jet)



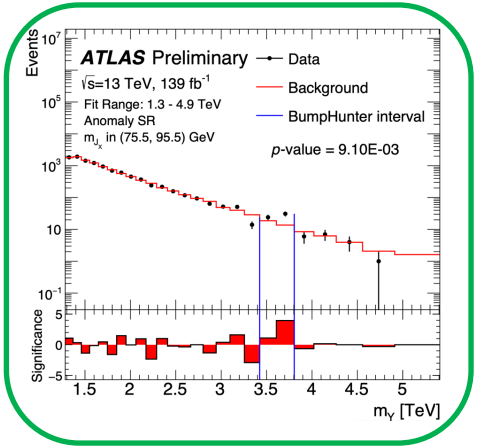



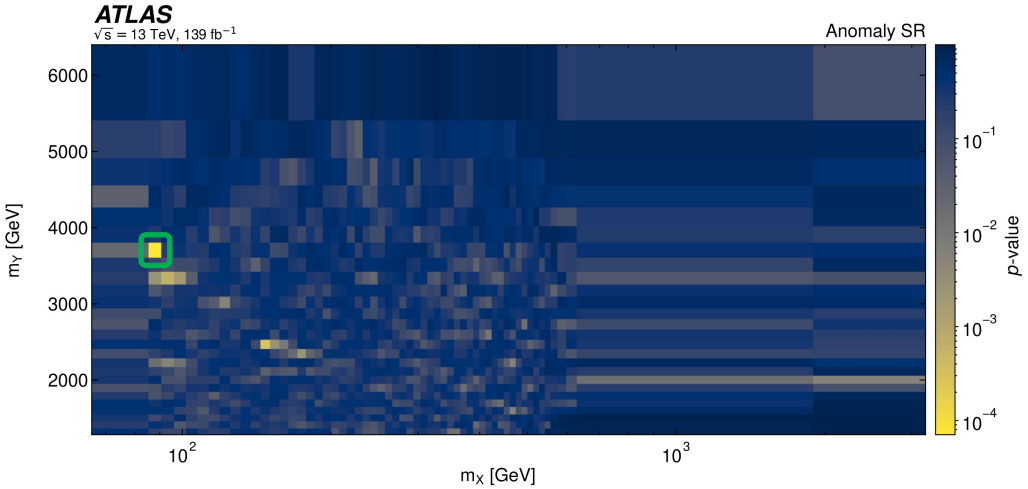



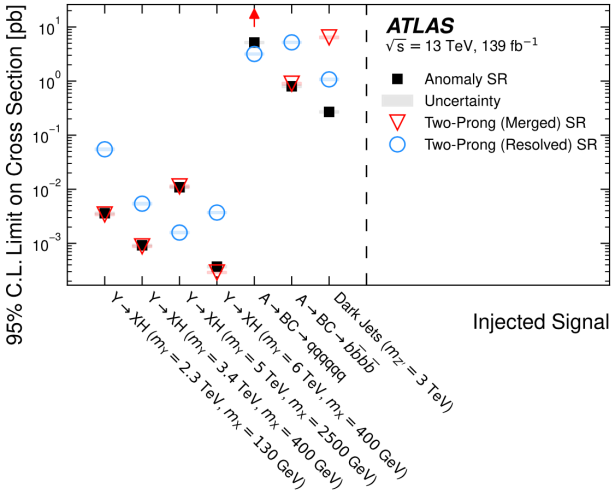
# Results

Phys.Rev.D 108 (2023) 052009

- Fit performed on final state invariant mass distribution $m_{jj}$ in SR of data, repeated several times in overlapping bins of the X candidate mass
- Calculated stat-only p-value to test compatibility with background only hypothesis
- Max deviation: 1.43$\sigma$ global significance due to the several search regions defined



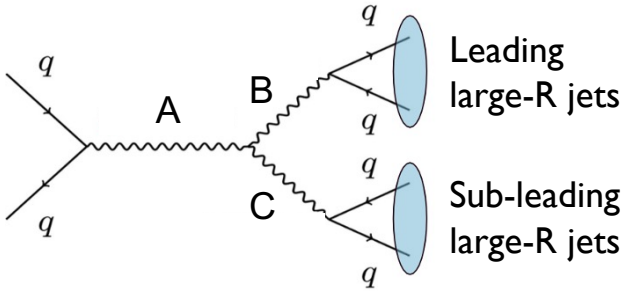



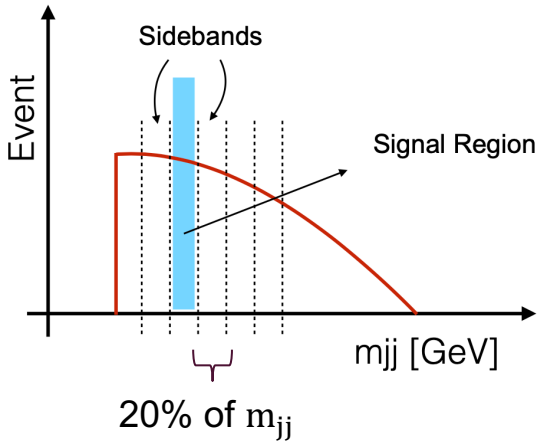
# Weakly-supervised anomaly detection for resonant new physics in the dijet final state

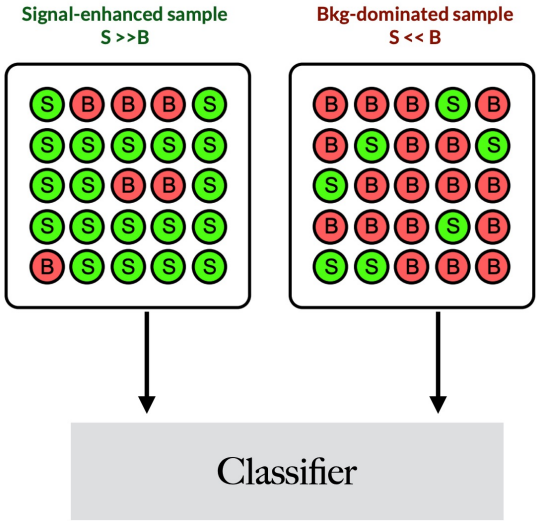
# CWoLa hunting

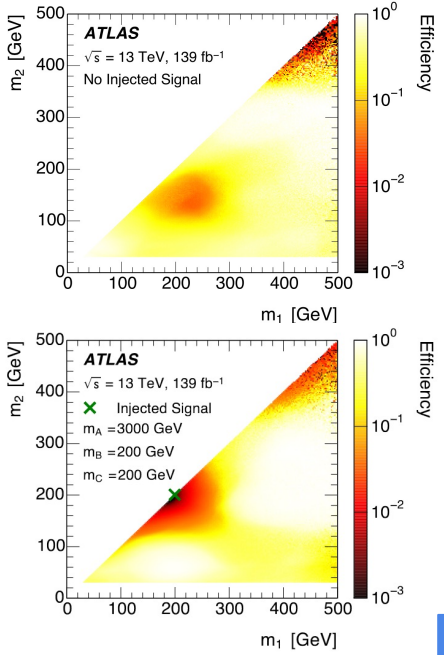
Phys. Rev. Lett. 125 (2020) 131801

- Classification Without Labels (CWoLa) method used for A → BC search
  - mass range: 1.1 - ~8 TeV
- 6 signal regions by $m_{jj}$ splitting, jets mass > 30 and < 500 GeV, $|\Delta y| < 1.2$
- Classifier trained on two samples D1 and D2, mixtures of signal and background, to produce discriminant output
  - Input variables: $m_1, m_2$ (pT leading jets)



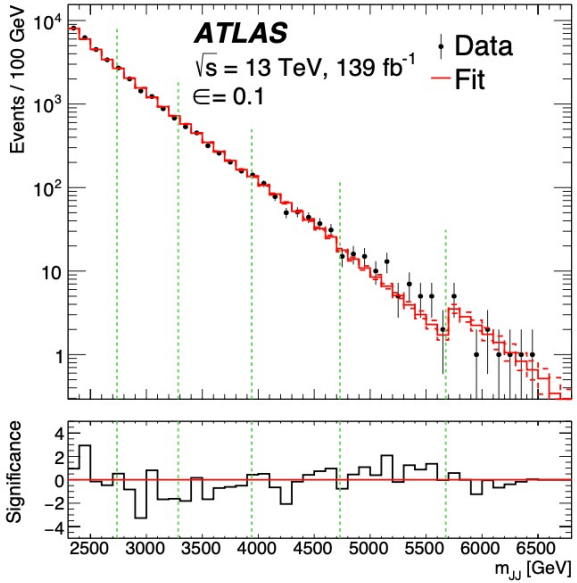



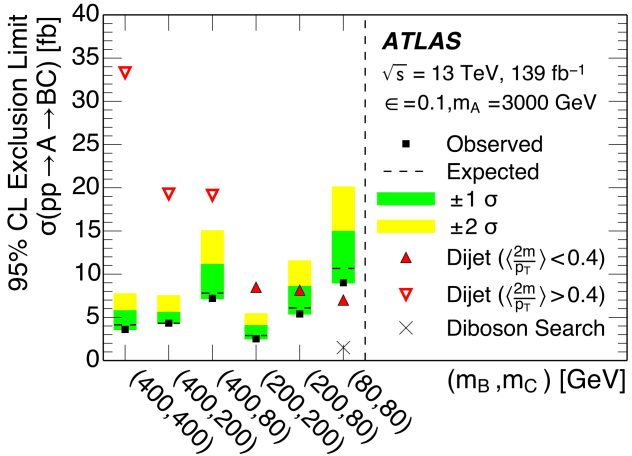
# CWoLa hunting results

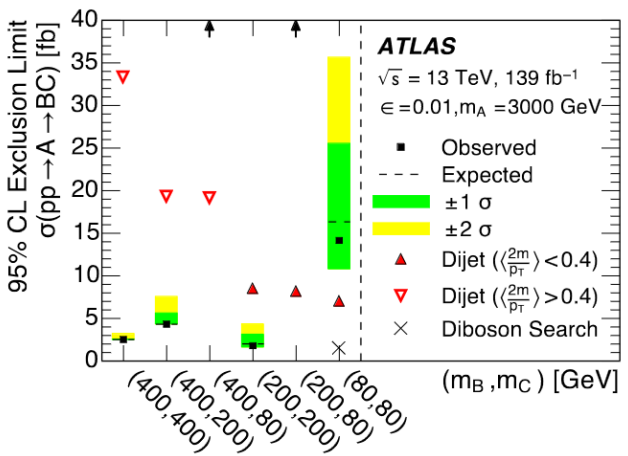
Phys. Rev. Lett. 125 (2020) 131801

- Upper limits on signal cross section, benchmark models compared with other diboson searches
  - Different values of signal selection efficiency, 0.1 and 0.01
  - QCD background estimation in SR done with functional fits
- CWoLa performs better when local signal-to-background ratio is high




Signal hypothesis: W' → WZ

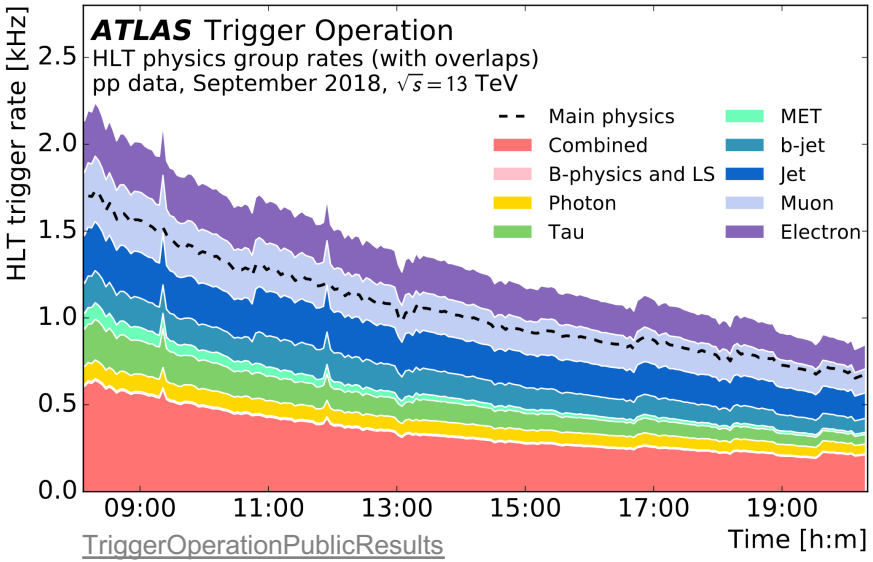


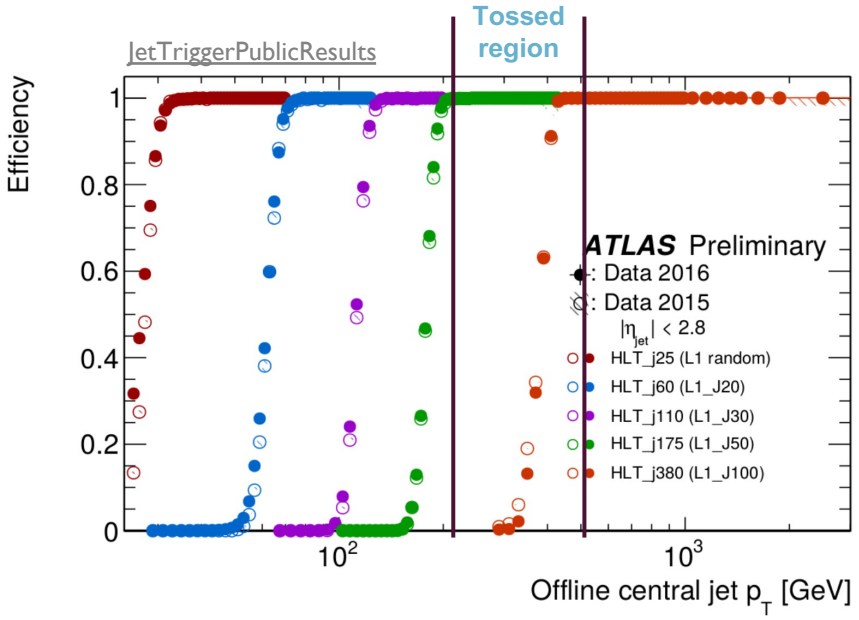



# Search for Low-Mass Dijet Resonances Using Trigger Level Analysis

# Trigger Level Analysis (TLA)

- Low pT jets physics (200 – 440 GeV) is tossed in ATLAS due to **trigger** limitations
- ATLAS normally stores the entire detector output for triggered events, limiting the rate at which events can saved
- Trigger Level Analysis chains record only the output of HLT reconstruction o(3kB/event) at extremely high rate o(3kHz)
  - Jets included (~15% of total trigger decisions)

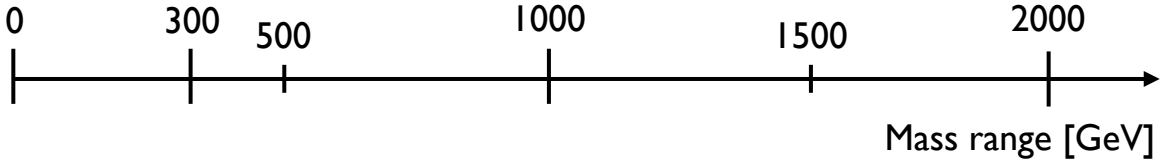


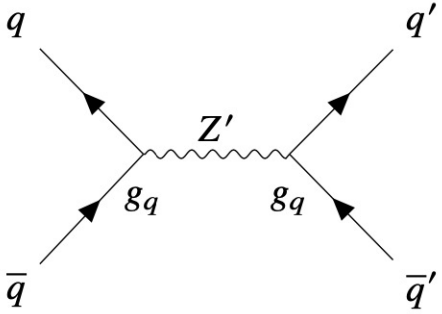
# TLA search in fully hadronic final states

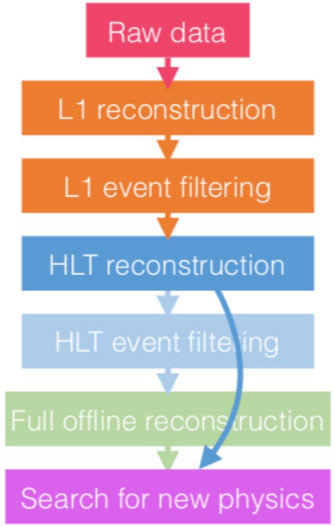
PRL 121 081801

- Electroweak-TeV scale should be studied throughtly, as W, Z, Higgs boson and top are all found there
  - Current single jet HLT trigger (pT > 440 GeV) constraints $m_{jj} \gtrsim 1.5$ TeV
- TLA can be used to recover sensitivity at the TeV scale! → HLT reconstructed jets and event header
  - No calorimeter cells, constituents, hits or tracks are saved, no offline reconstruction
  - TLA jets calibrated to match offline reconstructed jets
- Model independent, benchmark model used to set upper limits on coupling constant $g_q$ (29.3 $fb^{-1}$)




**Benchmark model: DM mediator**

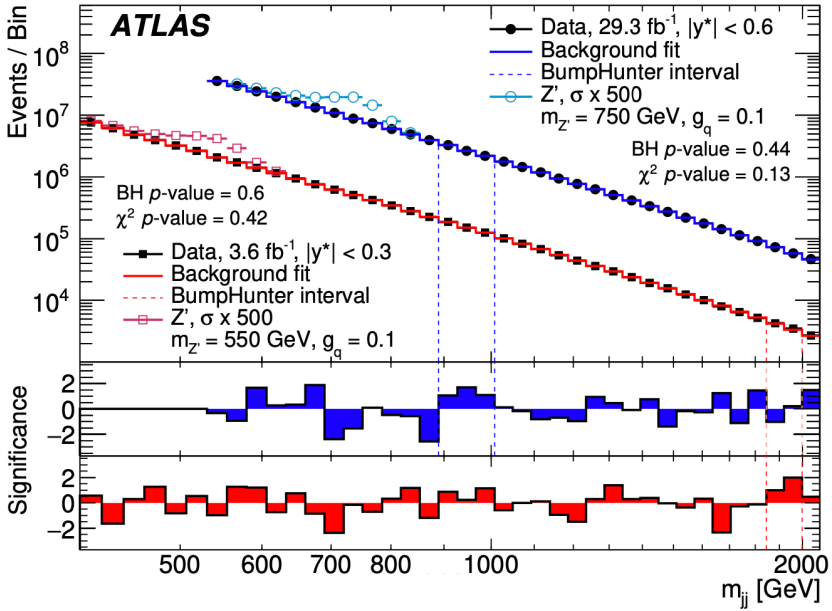


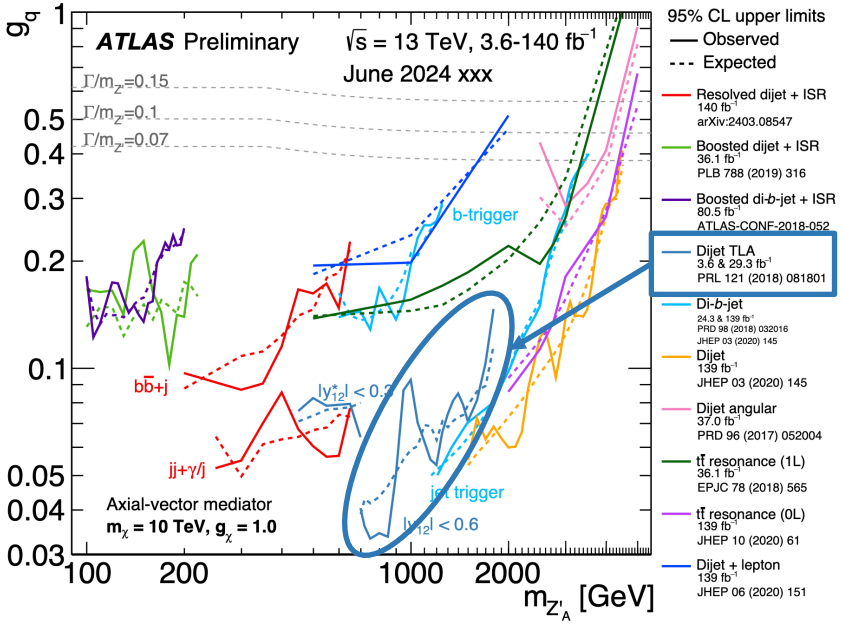
# TLA search results

PRL 121 081801

- Background estimated with functional fit of subranges with sliding window
- No bump found → factor 2-5x improvement in coupling constant limits w.r.t. other searches for lower masses




ATL-PHYS-PUB-2024-010



# Conclusions

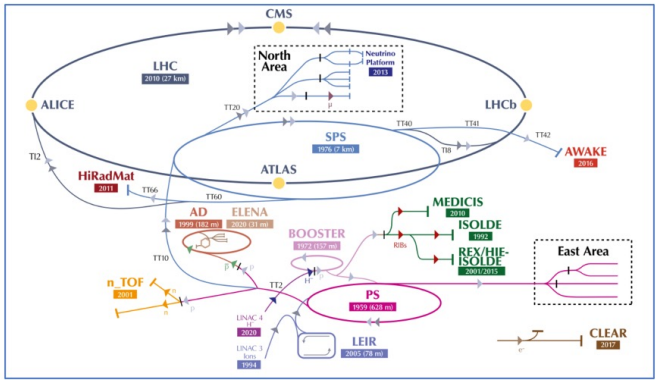
- No new interactions and particles since the Higgs boson's discovery → more generic searches opposed to the existing model-dependent analysis standard
- Model agnostic searches with jets in final state becoming a main topic in the ATLAS collaboration
- Exploited LHC Run 2 data collected by ATLAS, also moving on to Run 3 data
  - Run 2: TLA analysis, CWoLa, search for resonances with quark tagging, YXH
  - Run 3: Anomaly Detection with Graph Neural Networks
- Honorable mentions: Anomaly Detection search with Run 2 data (Phys. Rev. Lett. 132, 081801), search for signatures of Soft Unclustered Energy Patterns
- Take home message: **Model agnostic searches can be a powerful tool that is complementary to beyond standard model dependent searches approach**

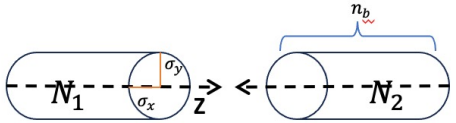
*Stay tuned and thank you for your attention!*

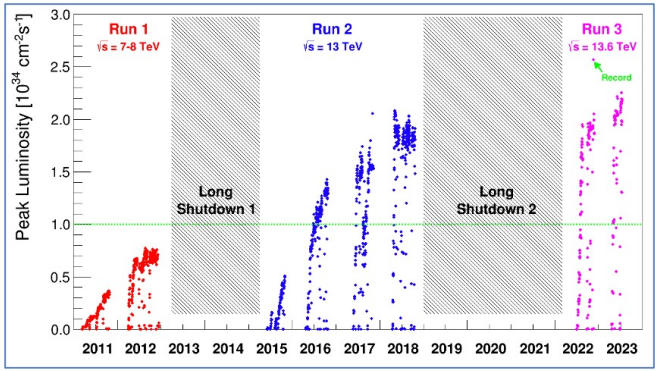
# BACKUP

# Large Hadron Collider

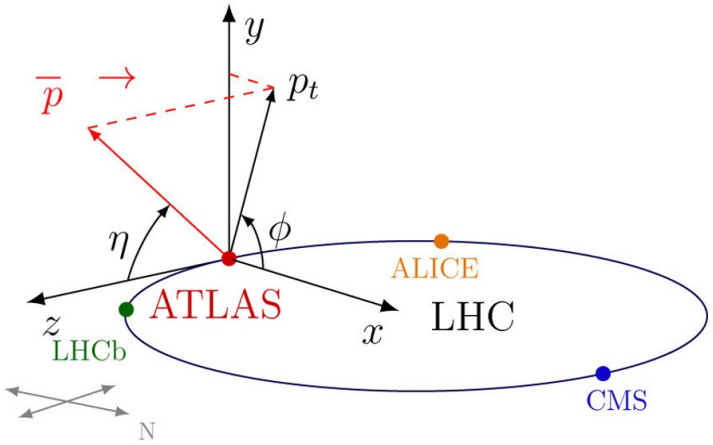
- Largest and most powerful particle accelerator worldwide;
- $27\ km$ long tunnel underground provided by superconductive magnets to bend and accelerate particles;
- $13.6\ TeV$ center of mass energy (July 2022);
- 4 interaction points where main detectors are located: ATLAS, LHCb, CMS and ALICE;
- Investigate fundamental particles and forces of the universe; explore dark matter, SUSY and Higgs boson physics;
- Luminosity, defined as $L = \frac{N_1 N_2 f n_b}{4\pi\sigma_x\sigma_y}$, is a geometric parameter used to measure the number of collisions that can be produced in a detector per $cm^2$ and per second.



# The ATLAS experiment




Multipurpose detectors arranged in concentric layers around the **collision point**:

Muon spectrometers;

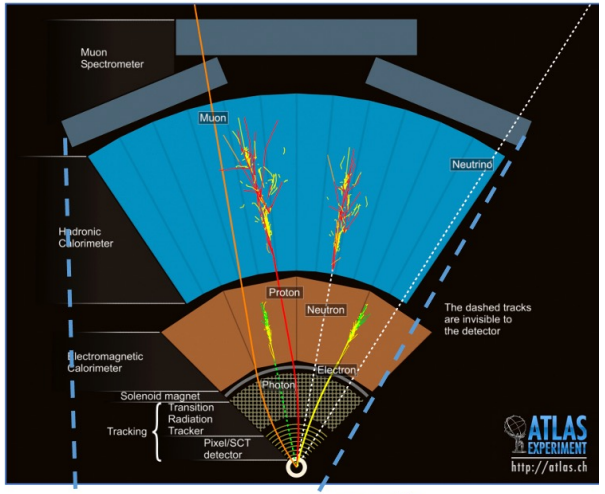
Magnetic system;

Hadronic calorimeter;

Electromagnetic calorimeter;

Inner detectors (trackers).

ATLAS adopts a complex **2-level trigger system** for data recording




ATLAS coordinate system

$(z, \eta, \phi)$

Pseudorapidity

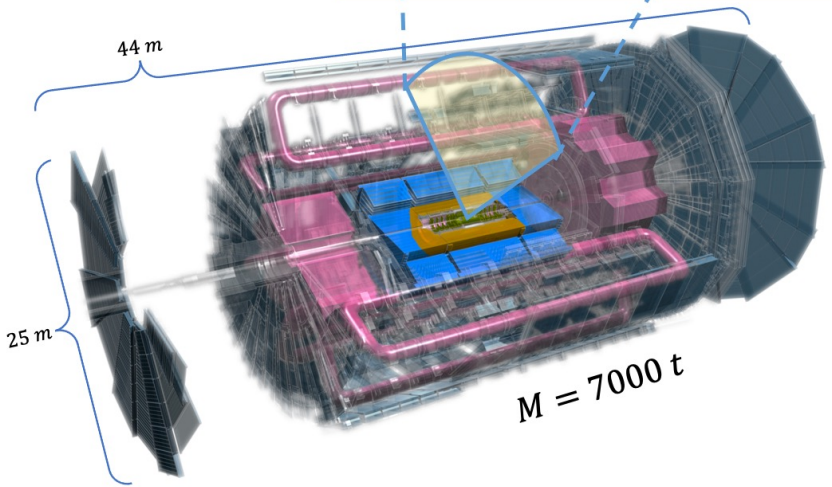
$$\eta = \frac{1}{2}\ln\left(\tan\left(\frac{\theta}{2}\right)\right)$$

Transverse momentum

$$p_T = \vec{p}\cos(\phi)$$

Angular separation in $\eta$ - $\phi$ plane

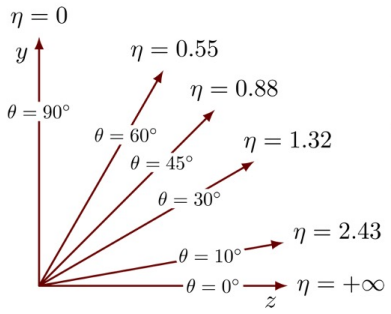
$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$$



# Coordinate system

Jets reconstructed using tracks in ID, calorimeter deposits and anti-$k_T$ algorithm.

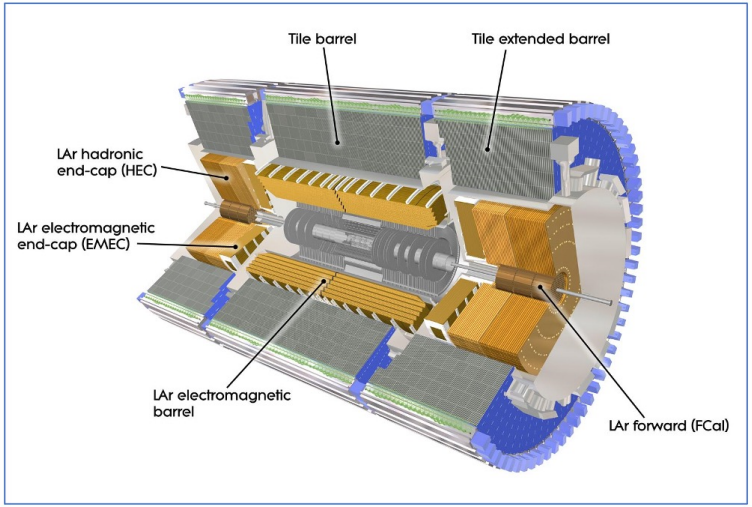
- Tile hadronic calorimeter: 14 mm of iron absorber alternated to a 3 mm sparkling plates, in bunches;
- Liquid Argon end-cap hadronic calorimeter: copper and tungsten as absorbers and LAr as active component.

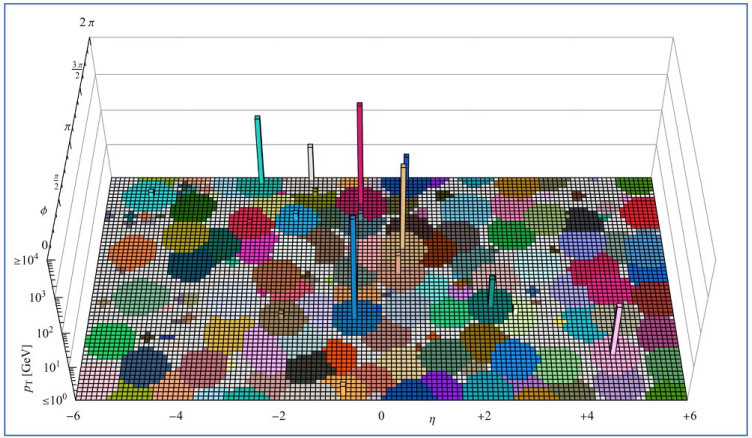



**Energy resolution**

| Pseudorapidity range | Energy resolution $\frac{\sigma_E}{E}$ |
|---|---|
| $\lvert\eta\rvert < 3.0$ | $\frac{50\%}{\sqrt{E}} \oplus 3\%$ |
| $3.0 < \lvert\eta\rvert < 4.9$ | $\frac{100\%}{\sqrt{E}} \oplus 10\%$ |

Anti-$k_T$ reconstruction algorithm takes topoclusters (clusters of energy deposits in the calorimeters) as input and combine them to form jet cones with characteristic radius R using a distance parameter



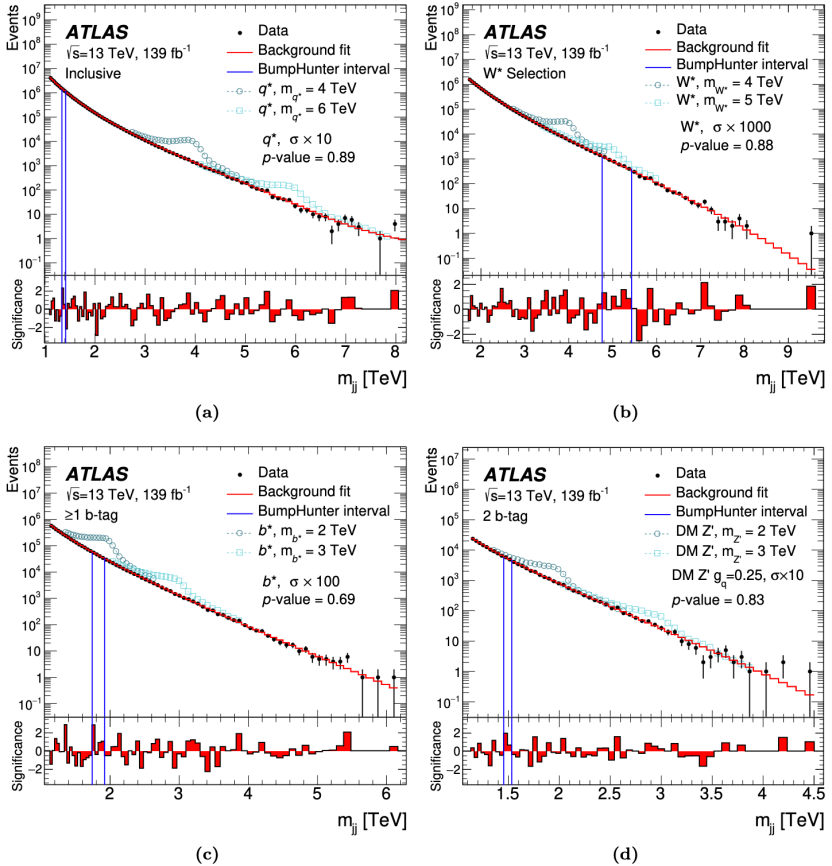



# Event Selection in Analysis I

$$y^* = \frac{y_1 - y_2}{2}$$

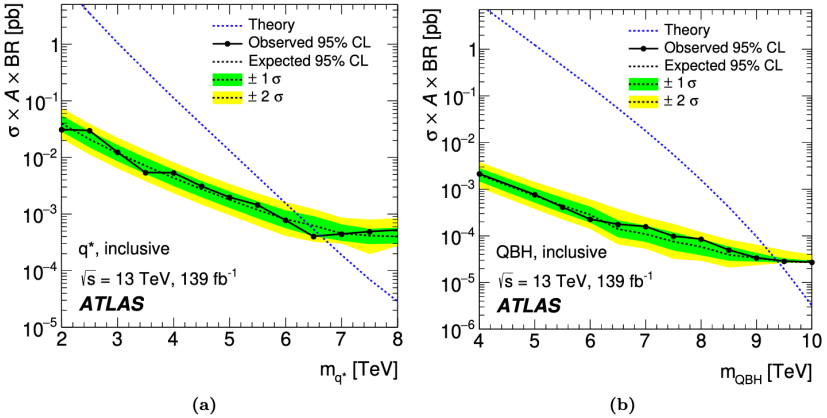
| Category | Inclusive | | 1$b$ | 2$b$ |
|---|---|---|---|---|
| Jet $p_{\mathrm{T}}$ | $> 150\,\mathrm{GeV}$ | | | |
| Jet $\phi$ | $\lvert\Delta\phi(jj)\rvert > 1.0$ | | | |
| Jet $\lvert\eta\rvert$ | — | | $< 2.0$ | |
| $\lvert y^*\rvert$ | $< 0.6$ | $< 1.2$ | $< 0.8$ | |
| $m_{\mathrm{jj}}$ | $> 1100\,\mathrm{GeV}$ | $> 1717\,\mathrm{GeV}$ | $> 1133\,\mathrm{GeV}$ | |
| $b$-tagging | no requirement | | $\geqslant 1$ $b$-tagged jet | 2 $b$-tagged jets |
| Signal | DM mediator $Z'$<br>$W'$<br>$q^*$<br>QBH<br>Generic Gaussian | $W^*$ | $b^*$<br>Generic Gaussian | DM mediator $Z'$ $(b\bar{b})$<br>SSM $Z'$ $(b\bar{b})$<br>graviton $(b\bar{b})$<br>Generic Gaussian |

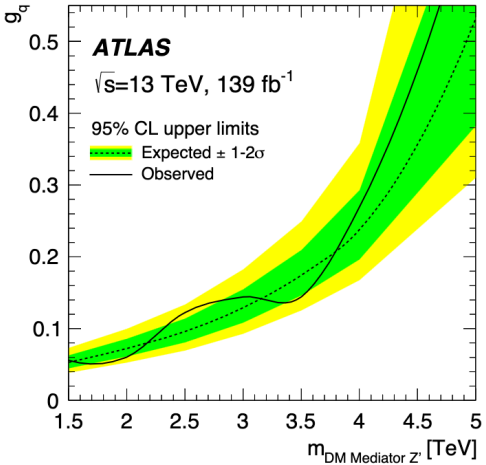
# Further results of Analysis I

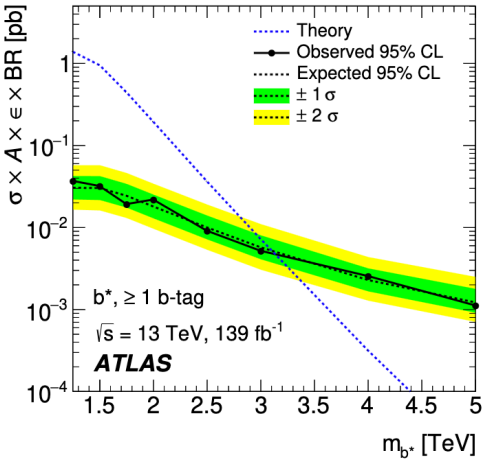



| Category | Model | Lower limit on signal mass at 95% CL | |
|---|---|---|---|
| | | Observed | Expected |
| Inclusive | $q^*$ | 6.7 TeV | 6.4 TeV |
| | QBH | 9.4 TeV | 9.4 TeV |
| | $W'$ | 4.0 TeV | 4.2 TeV |
| | $W^*$ | 3.9 TeV | 4.1 TeV |
| | DM mediator $Z'$, $g_q = 0.20$ | 3.8 TeV | 3.8 TeV |
| | DM mediator $Z'$, $g_q = 0.50$ | 4.6 TeV | 4.9 TeV |
| $1b$ | $b^*$ | 3.2 TeV | 3.1 TeV |
| $2b$ | DM mediator $Z'$ $g_q = 0.20$ | 2.8 TeV | 2.8 TeV |
| | DM mediator $Z'$, $g_q = 0.25$ | 2.9 TeV | 3.0 TeV |
| | SSM $Z'$, | 2.7 TeV | 2.7 TeV |
| | graviton, $k/\overline{M}_{\mathrm{PL}} = 0.2$ | 2.8 TeV | 2.9 TeV |

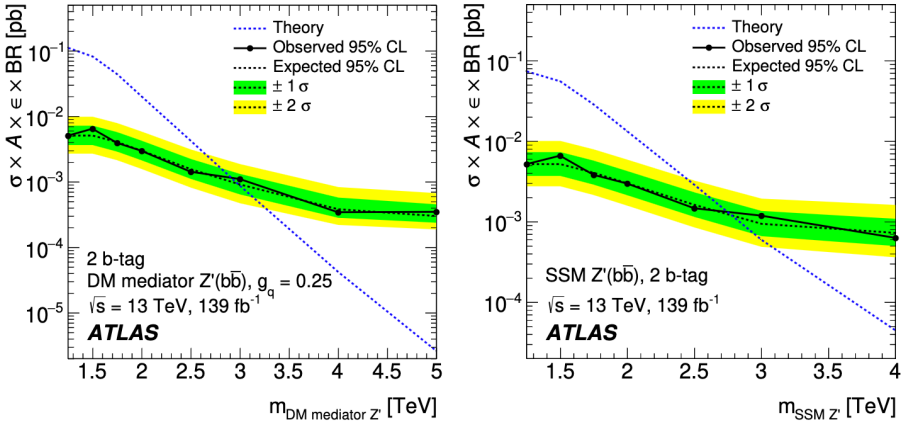
# Further results of Analysis I



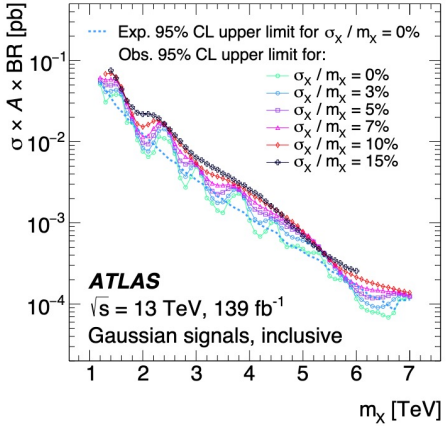
# Further results of Analysis I



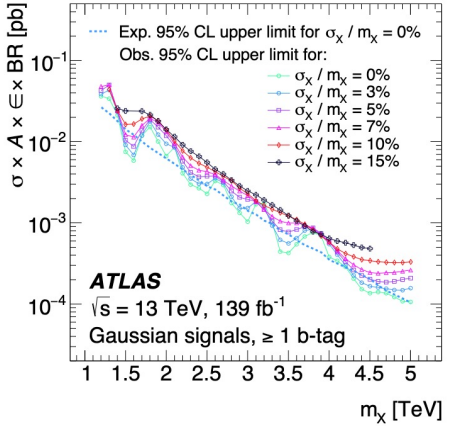

(a)



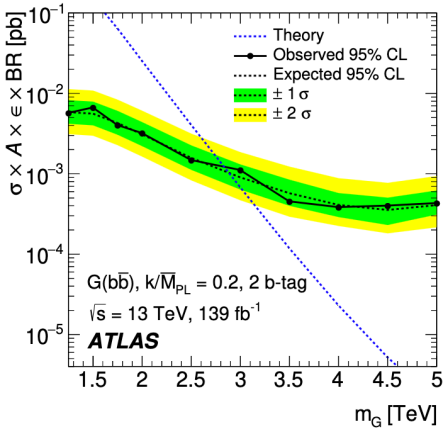

(b)



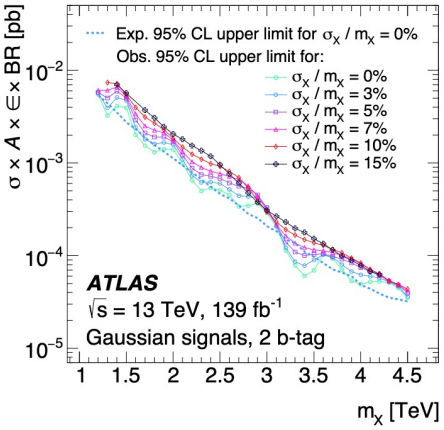

(a)




(b)




(c)




(c)

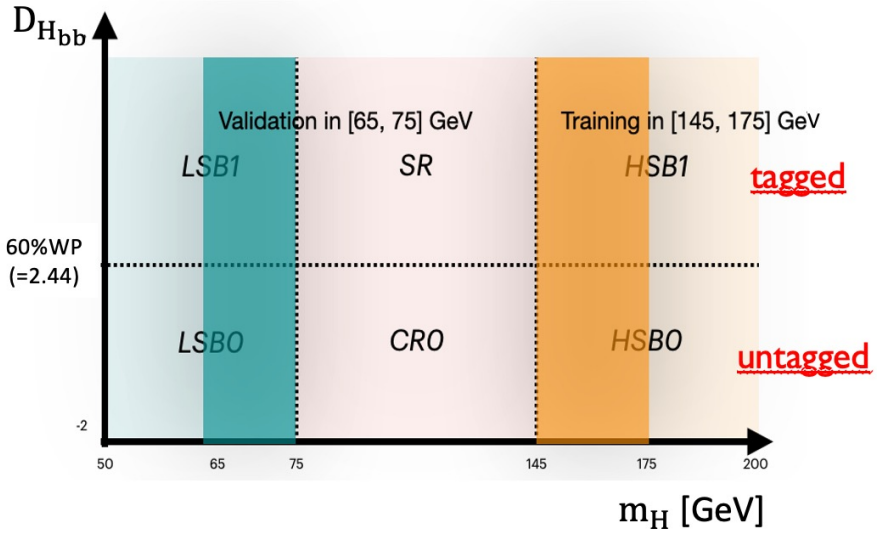
# YXH background estimation

- Data-driven background estimation since ~97% from QCD di-jet
  - MC simulations are not precise enough!
- Performed by reweighting events with a function w(x) from CR0 to SR data:

$$w(\vec{x}) = \frac{pdf_1(\vec{x})}{pdf_0(\vec{x})}$$

- w(x) is learned by a Deep Neural Network (DNN) in the training region HSB, validated in LSB and finally extrapolated in the Higgs mass window
- Training performed on data before $D^2_{Tracks}$ and AS categorization
- DNN with 3 fully-connected inner layers, 20 neurons each, implemented with Keras (Tensorflow backend)

Totally innovative background estimation technique based on DNN data-driven reweighting

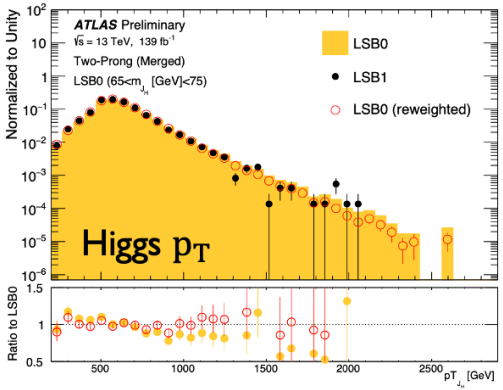



| DNN input features |
| --- |
| H four-momentum |
| Ntracks associated to H |
| $p_T, \phi, \eta$, m of $p_T$ leading and sub-leading small-R jets associated to H |

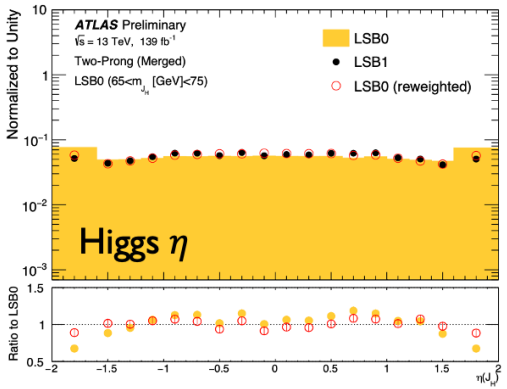
# Background validation

Merged Region

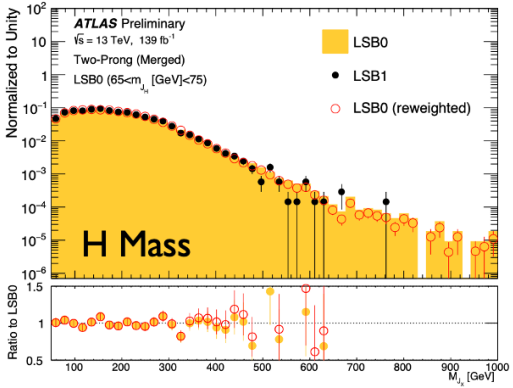
- Since the DNN training is inclusive in X tagging selections, reweighting is applied in AS and $D^2_{Tracks}$ regions without retraining
- Generally good closure of the background prediction to data is observed in validation region (LSB) for each scenario
  - Occasional non-closure is taken as a systematic uncertainty of the background estimation method



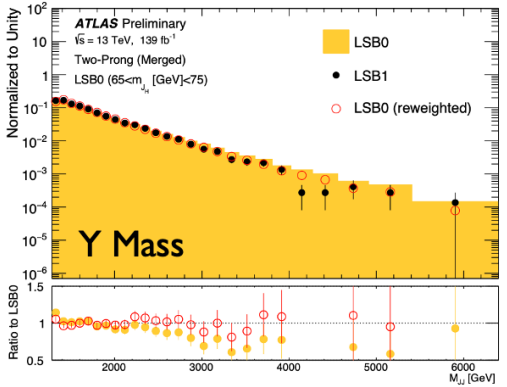

(a)




(b)



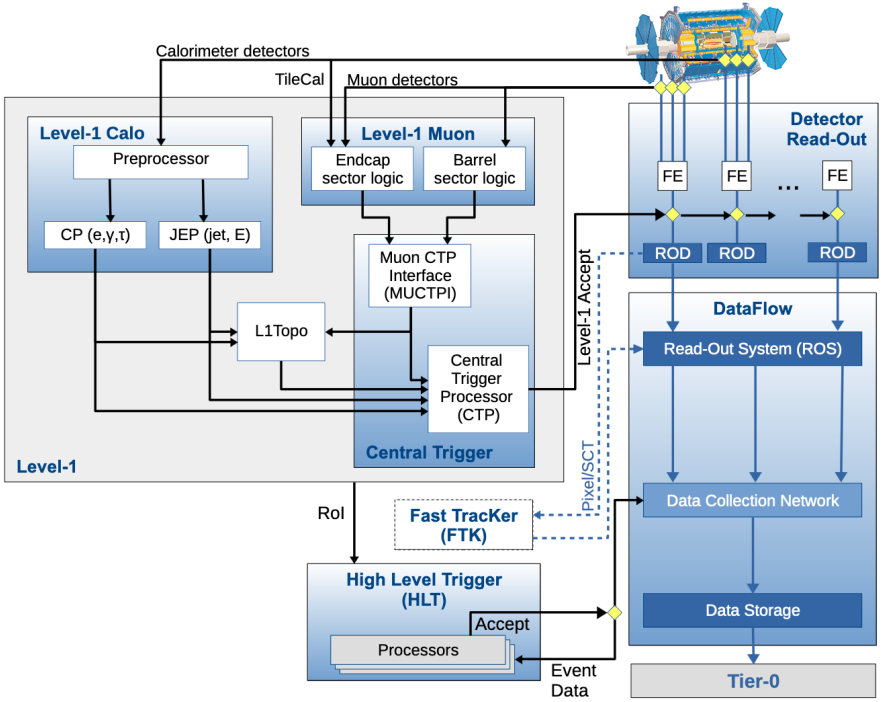

(c)




(d)

before reweighting

after reweighting

# ATLAS trigger system

- 23 collisions per bunch crossing every 25 ns → **60TB/s** to store everything!!!
- Selection applied to store only interesting physics; decision took in two steps:
  - L1 trigger, hardware based (100 kHz)
  - High Level Trigger (HLT), software based (~1 kHz)
- Decisions taken based on calorimeter and muon detectors