ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Measurements of Z+b-jets with the ATLAS experiment at the LHC

Giuseppina Santoro (g.santoro@cern.ch)

UNIVERSITÀ DELLA CALABRIA

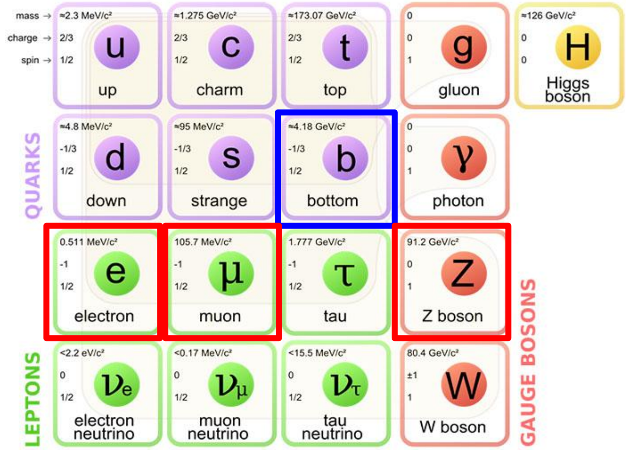
Measurements of Z+b-jets with the ATLAS experiment at the LHC

Outline

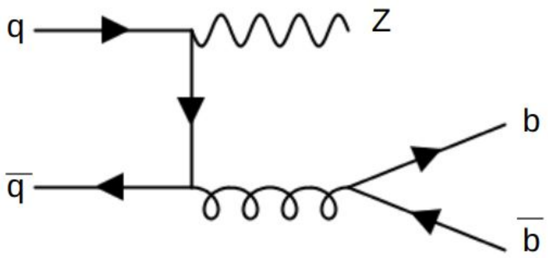
- Introduction
- ATLAS detector at LHC
- Z+ b-jets measurements:
    - Latest ATLAS published results
    - Ongoing analysis

Z + b-jets in SM

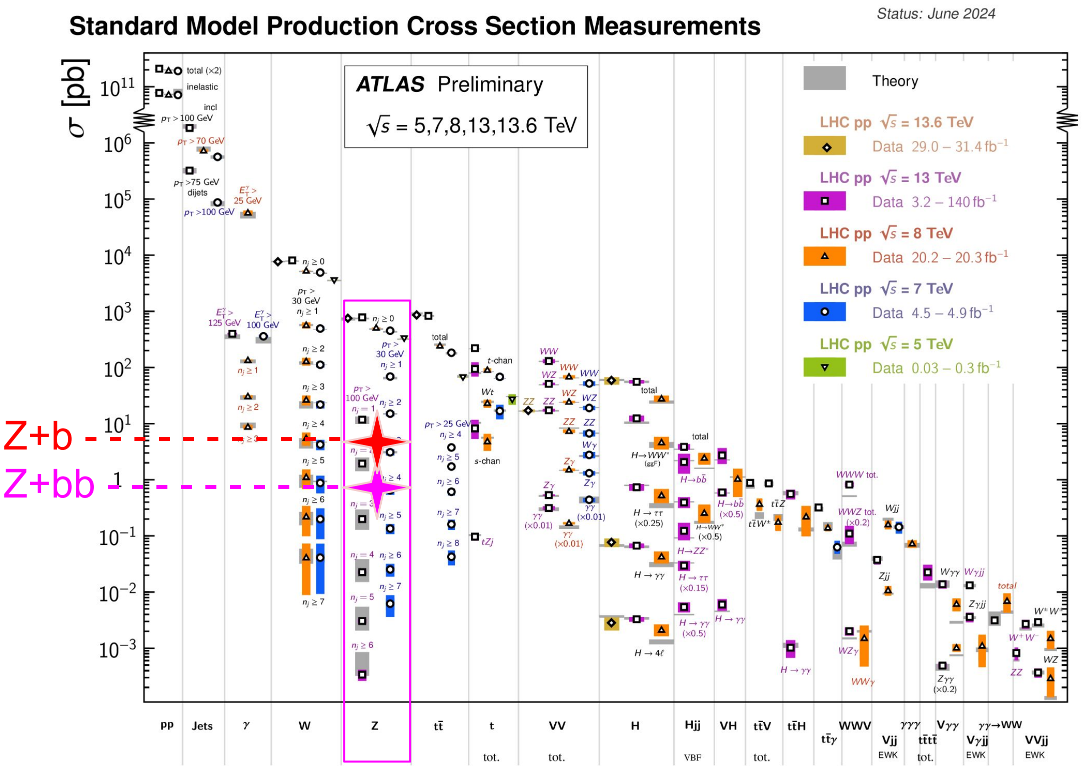
- The Standard Model theory describes EW and strong interaction between elementary particles



- Z + b-jets :
  - Z studied through decay in leptons
  - b quark identified as heavy flavour jet



Z+jets production



- Z + (inclusive) jets abundantly produced at the LHC and measured up to a large jet multiplicities and in extreme phase spaces
- Z + b-jets more challenging to measure and to predict

Focus of this seminar

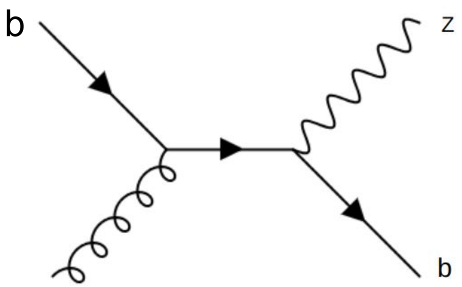
- Production of a Z boson in association with b-jets (Eur. Phys. J. C 84 (2024) 984, CERN-THESIS-2024-256)
- Production of a Z boson in association with 2 b-jets in the boosted regime (on-going analysis)

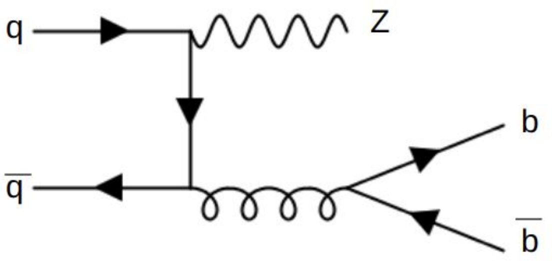
Full Run 2 ATLAS dataset at $\sqrt{s}$=13 TeV used

Z + b-jets measurements: motivations

The Z +jets originating from the hadronization of heavy particles is important because:

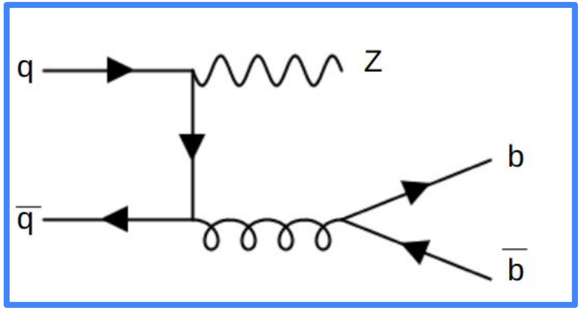
- important test of perturbative quantum chromodynamics (pQCD);
  - fixed order calculations up to NNLO
  - MCs with multileg NLO matrix elements (ME) + Parton Showers (PS)
  - various flavour schemes (FS) with different b-mass treatments
- sensitivity to PDFs.

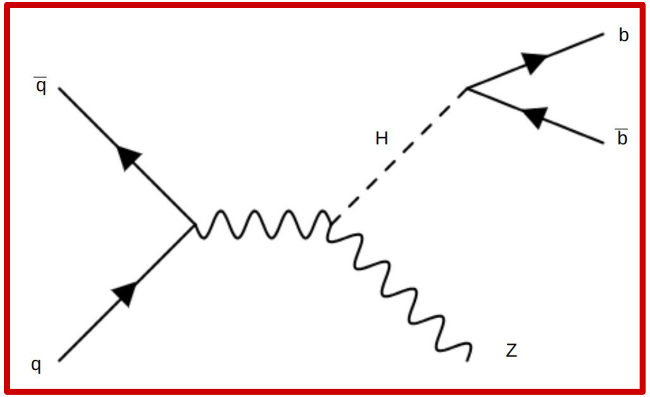


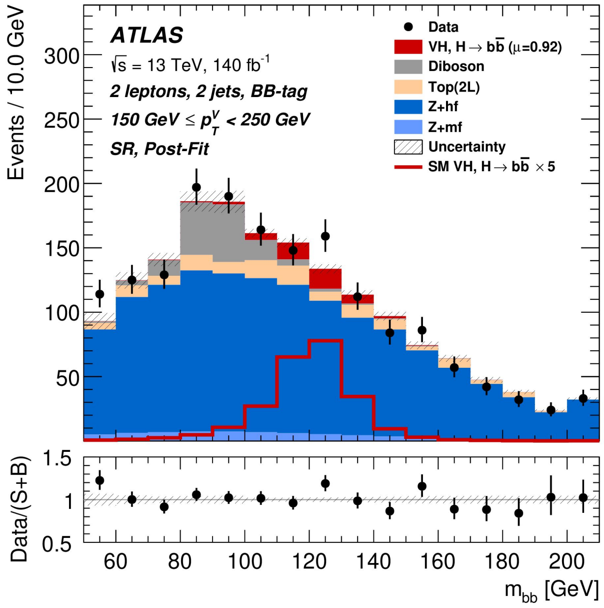


Z + b-jets measurements: motivations

- Inputs to improve the background modelling in MCs for Higgs boson measurements and searches of New Physics
  - Z+ b-jets significant background for ZH (--> bb)



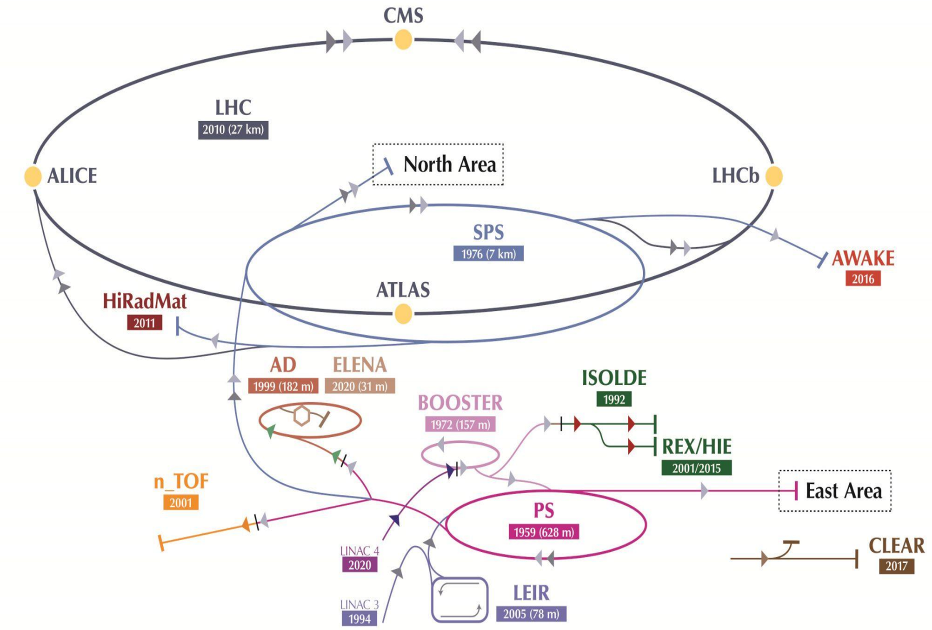




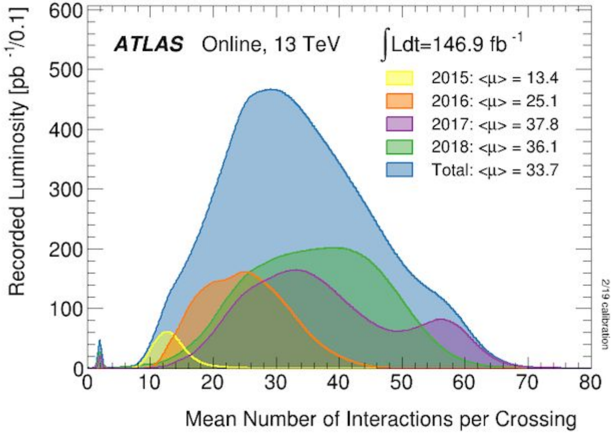
Paper: [Submitted to: JHEP]

ATLAS at LHC

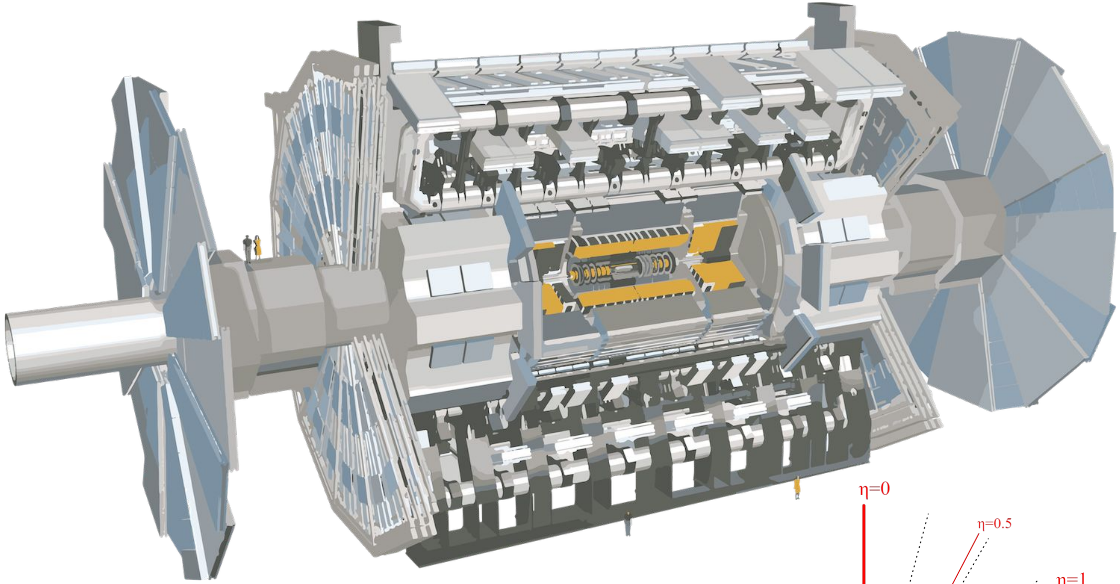
The Large Hadron Collider

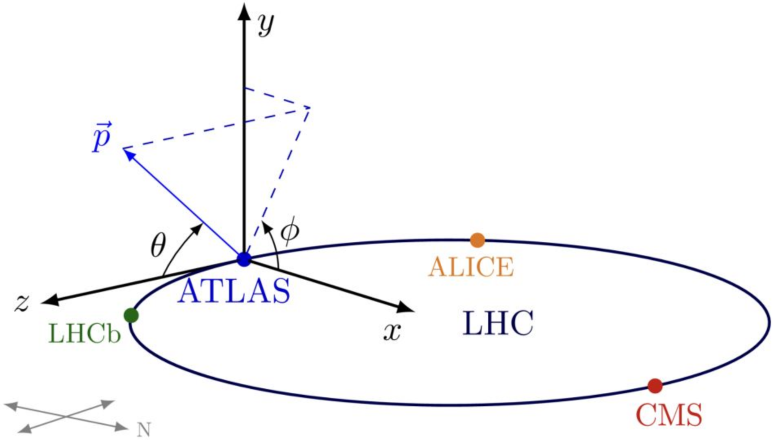


pp collision	Run 1 (2010-2012)	Run 2 (2015-2018)	Run 3 (2022-ongoing)
Center of mass energy	7-8 TeV	13 TeV	13,6 TeV
Luminosity	20,3 fb^{-1}	140 fb^{-1}	300 fb^{-1} (expected)



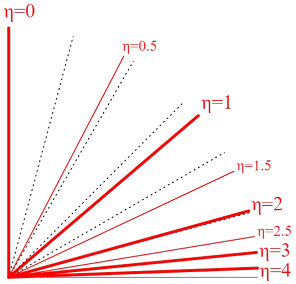
The ATLAS experiment at LHC





$$p_T = \sqrt{p_x^2 + p_y^2}$$

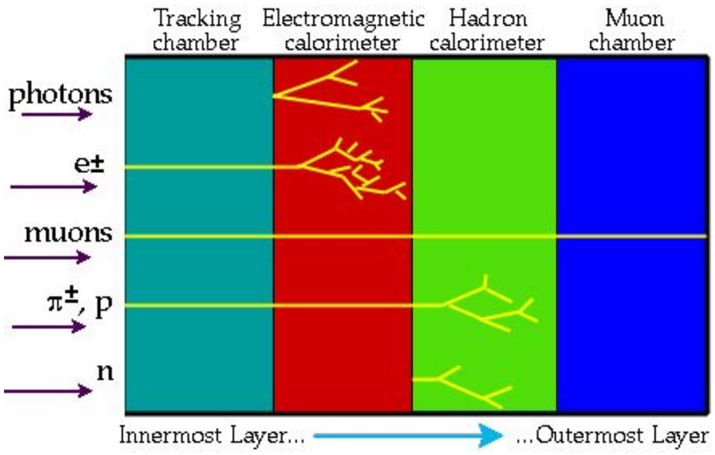
$$\eta = -\ln\left(\tan\frac{\theta}{2}\right)$$

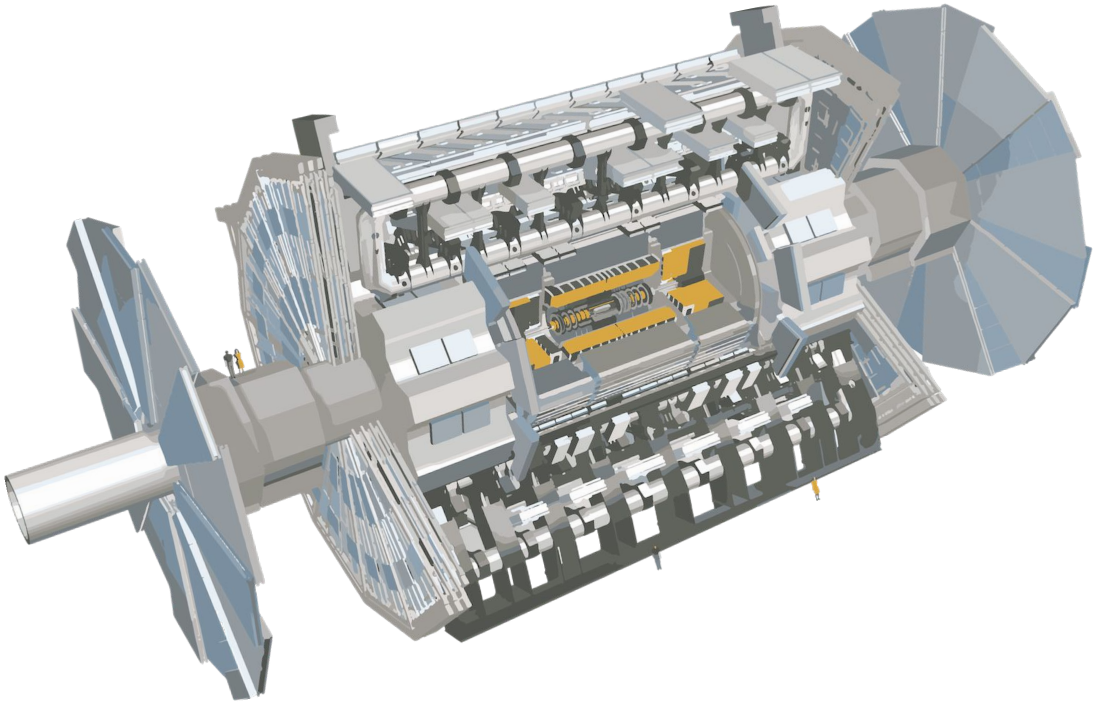


$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$$

Object detection and recognition at ATLAS

- ATLAS is a cylinder, 46m long, 25m in diameter.
- ATLAS sub-detectors: inner tracker,calorimeters (electromagnetic and hadronic) and muon spectrometer.

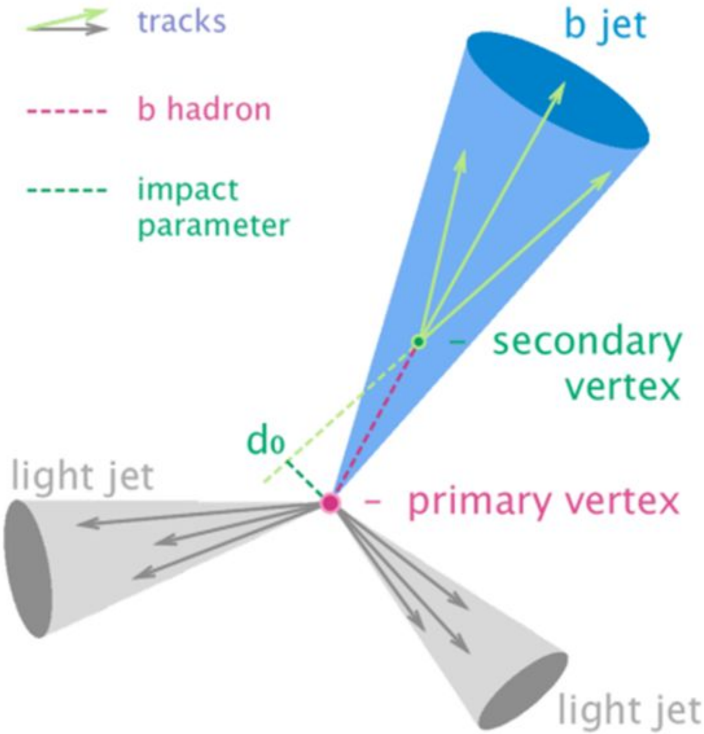




Flavour Tagging

- Flavour of the jets determined with DL1r algorithm
- DL1r is an high level algorithm operating on outputs from intermediate track and vertex algorithms
- DL1r exploits b-quark properties (i.e. long lifetime and heavy mass) relying on track-based observables (displaced tracks, secondary vertex, longitudinal impact parameter, decay topologies)

ϵ_b	DL1		
	Selection	Rejection	
		c-jet	Light-flavour jet
60%	> 2.74	27	1300
70%	> 2.02	9.4	390
77%	> 1.45	4.9	130
85%	> 0.46	2.6	29



Measurements of the production cross-section for a Z boson in association with b-jets in proton-proton collisions at 13 TeV with the ATLAS detector

- Eur. Phys. J. C 84 (2024) 984
- CERN-THESIS-2024-256

Analysis goal

Inclusive and differential cross-sections of Z+≥1 b-jet, Z+≥2 b-jets in a fiducial phase space with 140 fb^{-1} at 13 TeV

- update of the paper at 36 fb^{-1} (JHEP 07 (2020) 044)

$\mathbf{Z} + \geq 1b - jet$
Z boson p_T
Lead b-jet p_T
ΔR_{Zb}

$\mathbf{Z} + \geq 2b - jets$
Z boson p_T
Z boson $\lvert Y \rvert$
Lead b-jet p_T
Lead b-jet $\lvert Y \rvert$
Dibjets $\Delta\Phi_{bb}$
Dibjets ΔY_{bb}
Dibjets ΔR_{bb}
Dibjets M_{bb}

- Eur. Phys. J. C 84 (2024) 984
- CERN-THESIS-2024-256

Analysis Strategy

Detector level

- Select the events
- Estimate backgrounds through MC simulations and data-driven techniques
- Comparison at detector level of the data and the total (signal+background) predictions.

↓

Correct for detector effects (trigger and reconstruction efficiency, resolution) (UNFOLDING with MCs)

↓

Particle Level

Comparison: unfolded data - ME+PS MCs
unfolded data- Fixed order calc.

↑

Correct for various effect such as non perturbative effects (fragmentation and underlying events) with MCs

↑

Parton level

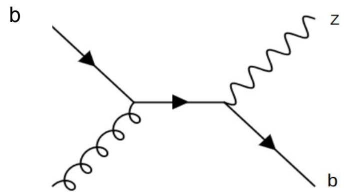
Event selection

N Leptons	
Exactly 2 leptons (2 muons or 2 electrons) with same flavour, opposite charge, and isolated	
Muons:	Single muon Trigger, Medium, $p_T > 27$ GeV, $\|\eta\| < 2.5$
Electrons:	Single electron Trigger, Tight, $p_T > 27$ GeV, $\|\eta\| < 1.37$ or $1.52 < \|\eta\| < 2.47$
Z mass window	
$76\,\text{GeV} < m_{\ell\ell} < 106\,\text{GeV}$	
Missing Transverse energy	
$E_T^{\text{miss}} < 60$ GeV if $p_T^Z < 150$ GeV	
Z + ≥ 1 b-jet (Z+b)	
Events with a Z-boson candidate, ≥ 1 b-tagged (85 % WP) jets (Anti-k_T with $R = 0.4$) with $p_T > 20$ GeV, $\|y\| < 2.5$, $\Delta R(\text{jet}, l) > 0.4$	
Z + ≥ 2 b-jets (Z+bb)	
Events with a Z-boson candidate, ≥ 2 b-tagged (85 % WP) jets (Anti-k_T with $R = 0.4$) with $p_T > 20$ GeV, $\|y\| < 2.5$, $\Delta R(\text{jet}, l) > 0.4$	

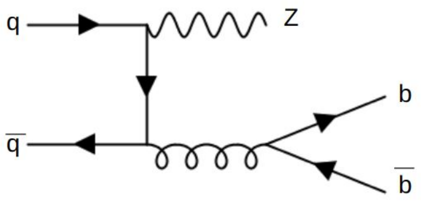
Signal predictions

The Z + b-jets signal predictions are derived in:

- 4-flavour number scheme (4FNS): in the 4FNS, b-quarks do not contribute to the parton distribution functions (PDFs) of the proton and, in QCD, they only appear in a massive final state due to gluon splitting g→ bb.
- 5-flavour number scheme (5FNS). In the 5FNS b-quarks contribute to the parton distribution functions of the proton.



Only in 5FNS



Both in 5FNS and 4FNS

	Generator/settings	Flav. scheme
MCs	MGaMC+PY8 FxFx	5FS
	SHERPA 2.2.11	5FS
	MGaMC+PY8	5FS
	MGaMC+PY8 Zbb	4FS
Fixed order calc.	NLO	5FS
	NNLO	5FS

- MGaMC+PY8 FxFx and Sherpa 2.2.11 fully simulated
- MGaMC+PY8 FxFx up to 3p NLO in ME
- Sherpa 2.2.11 up to 2p NLO and up to 5p LO in ME

Monte Carlo simulation

Process	Generator	Order of pQCD in ME (FS)
$Z \to \ell\ell$	MGaMC+Py8 FxFx Sherpa 2.2.11	0-3p NLO (5FS) 0-2p NLO, 3-5p LO (5FS)
$t\bar{t}$	NNLO+NNLL Powheg+Py8	NLO NLO
single top ($s/t/Wt$-channel)	 Powheg+Py8	NLO NLO
$qg/q\bar{q} \to VV \to \ell\ell/\nu\nu + q\bar{q}$	Sherpa 2.2.1	1p NLO, 2-3p LO
$qq \to ZH \to \ell\ell/\nu\nu + b\bar{b}$	Powheg+Py8	NLO
$gg \to ZH \to \ell\ell/\nu\nu + b\bar{b}$	Powheg+Py8	NLO

The background processes are: Diboson, single top production, production of H in association with Z, tt.

Background estimation

- **ttbar events** are estimated via data driven techniques
- Minor backgrounds (**Diboson processes, Single top quark production, Higgs boson in association with a Z boson)** are estimated via Monte Carlo simulations
- **Z+light or c-jets** dominant in Z+ b-jets events

Signal Z + ≥ 1 b-jet	
$Z + b$, $Z + bb$	34%
Backgrounds	
$Z + c$	29%
$Z + l$	35%
Top	2%
Others	1%
Total predicted	4 294 900 ± 2100
Data	4 145 168

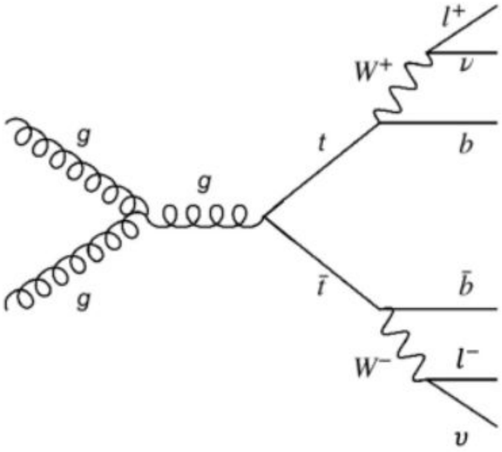
} 64% ($Z + c$, $Z + l$)

Signal Z + ≥ 2 b-jets	
$Z + bb$	46%
Backgrounds	
$Z + b$	11%
$Z + c$	23%
$Z + l$	7%
Top	12%
Others	2%
Total predicted	325 300 ± 600
Data	309 199

} 41% ($Z + b$, $Z + c$, $Z + l$)

Data-driven techniques for ttbar simulation

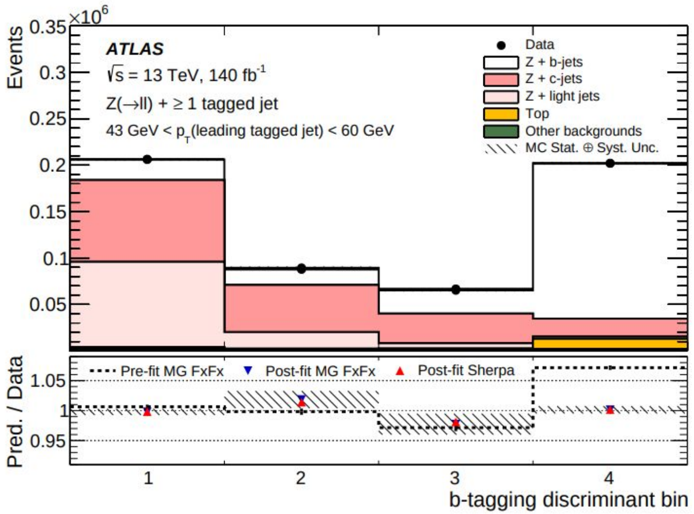
- **ttbar events** are the second largest background, simulated with data-driven techniques
- Events with opposite flavor leptons (eμ) are selected to define the control region (CR).
- The small background contributions in CR are subtracted from data
- The events in the signal region are obtained from the ratio between the distribution of the given observable in the signal region and in the CR.

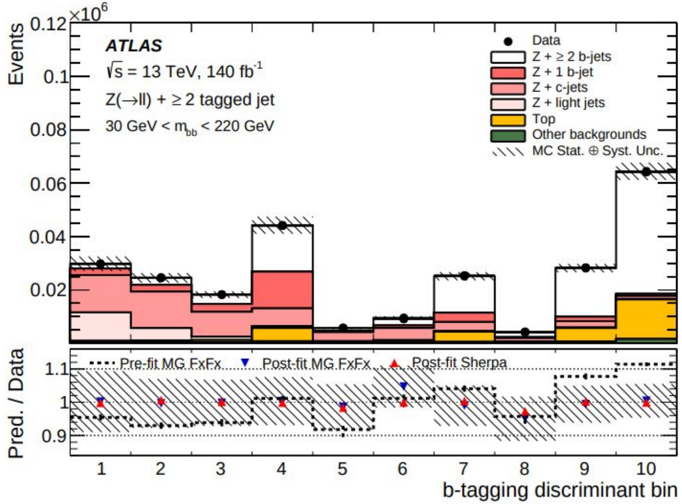


$$t\bar{t}^{SR} = t\bar{t}^{CR}_{\text{Data}} \cdot TF^{CR \to SR} \qquad TF^{CR \to SR} = \frac{t\bar{t}^{SR}_{MC}(ee/\mu\mu)}{t\bar{t}^{CR}_{MC}(e\mu)}$$

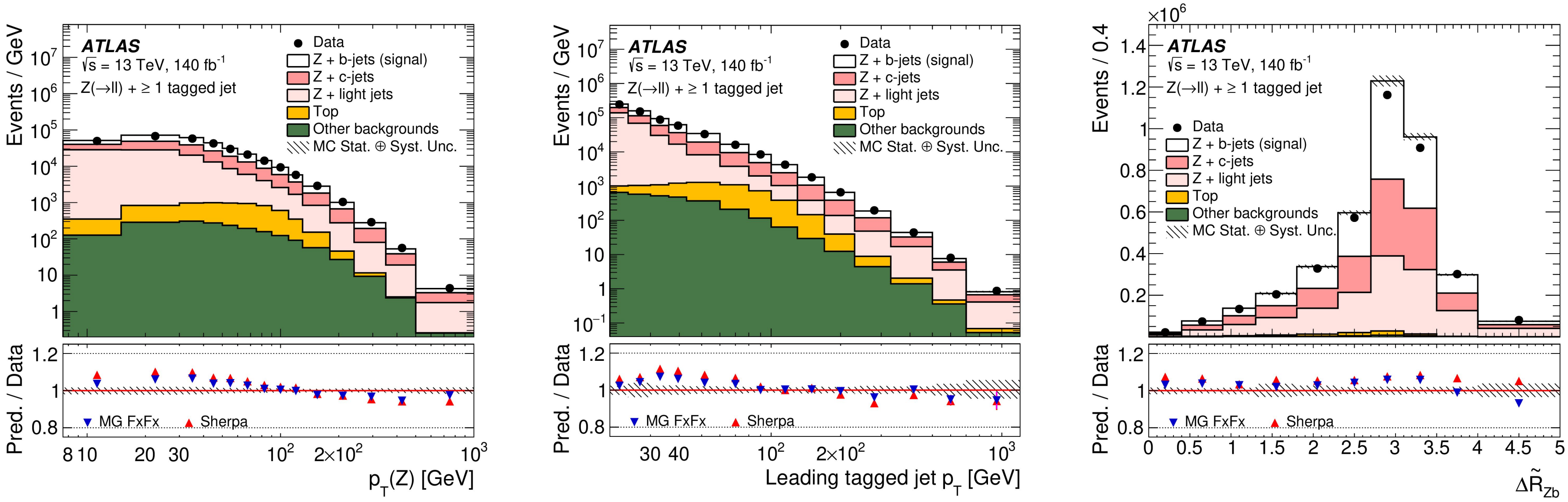
Flavor fit

- Z+jets background from a fit to data of a flavour sensitive variable, templates from MC, fit performed in individual bins of each observable.
- For Z + ⩾ 1 b-jet the DL1r b-tagging discriminant output is used.
- For Z + ⩾ 2 b-jets two DL1r discriminant outputs are combined.

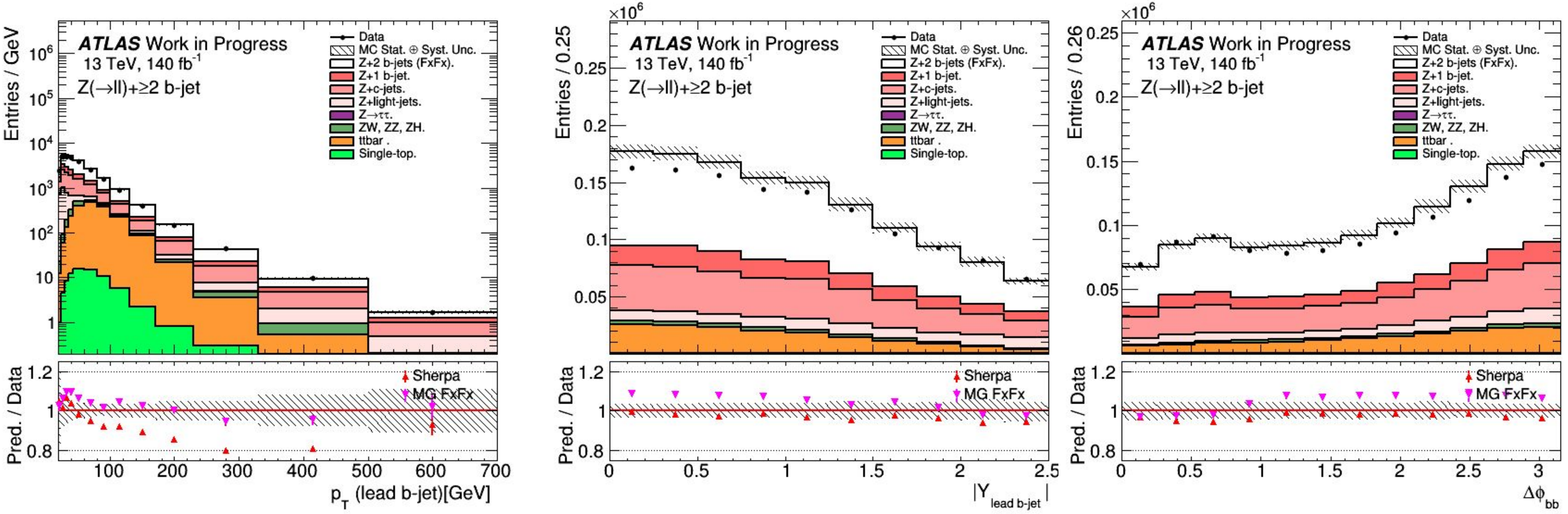




Detector level comparison (Z + ⩾ 1 b-jet)



Detector level comparison (Z + ⩾ 2 b-jets)



Fiducial phase space for the cross-section measurements

- In this analysis the fiducial phase-space is very close to the detector-level ones, this implies that the unfolding procedure relies at minimum on MC-based acceptance corrections and therefore the modeling uncertainties related to this corrections are small.

Object Selection	Acceptance cuts
$Z \to \mu\mu$	2 OS μ with $p_T > 27\,\text{GeV}$, $\lvert\eta\rvert < 2.5$, $m_{\mu\mu} = 91\,\text{GeV} \pm 15\,\text{GeV}$
$Z \to ee$	2 OS e with $p_T > 27\,\text{GeV}$, $\lvert\eta\rvert < 2.5$, $m_{ee} = 91\,\text{GeV} \pm 15\,\text{GeV}$
bjet	$p_T > 20\,\text{GeV}$ and $\lvert y\rvert < 2.5$, $\Delta R(\text{bjet}, \ell) > 0.4$

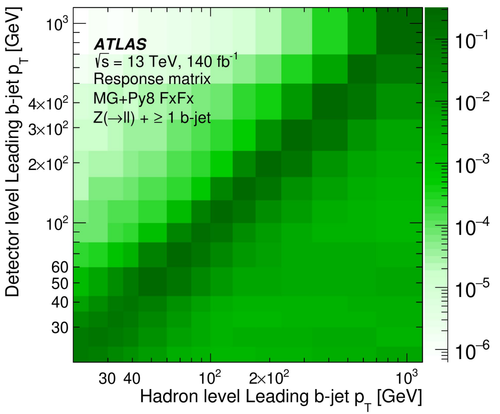
Unfolding

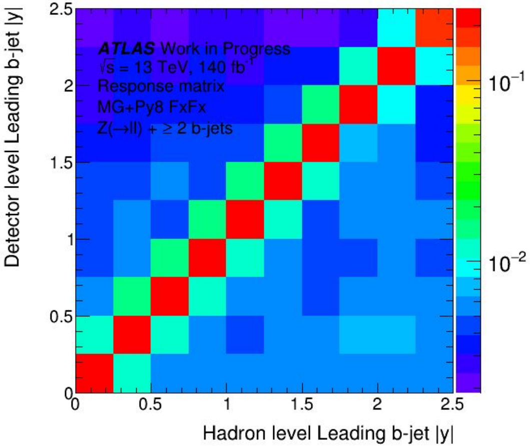
- (Data-Bkg) corrected for selection efficiency, resolution effects and small differences between detector level and fiducial phase spaces

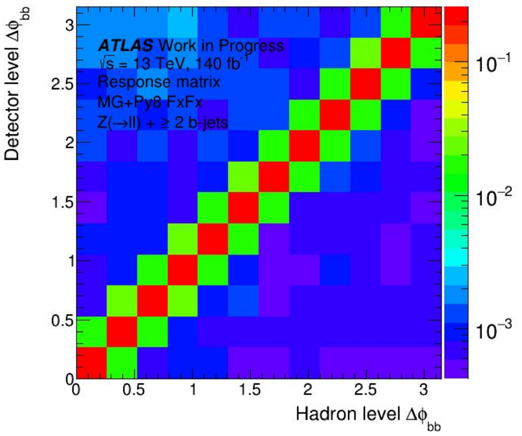
$$\frac{d\sigma_i}{dX_i} = \frac{M_{ij}^{-1}(N_j^{\mathrm{obs}} - B_j)f_j}{\epsilon_i L \Delta X_i}$$

- where M_{ij} is the migration matrix
- f_j corrects for events that pass the reco selection but fail the truth one
- ε_i corrects for events that pass the truth selection but fail the reco one

- Iterative Bayesian unfolding method as implemented in RooUnfold
- For prior (the initial approximation of the cross section), migration matrix, efficiency and fake factor → MG+FxFx
- To estimate the systematic uncertainties on the unfolding method →Sherpa

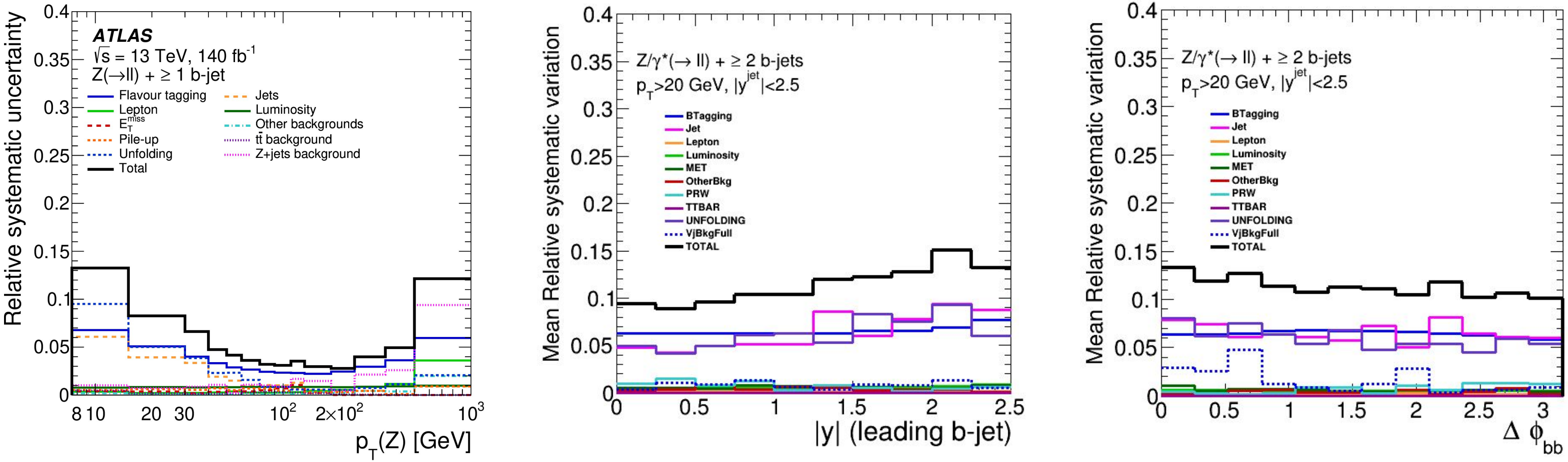






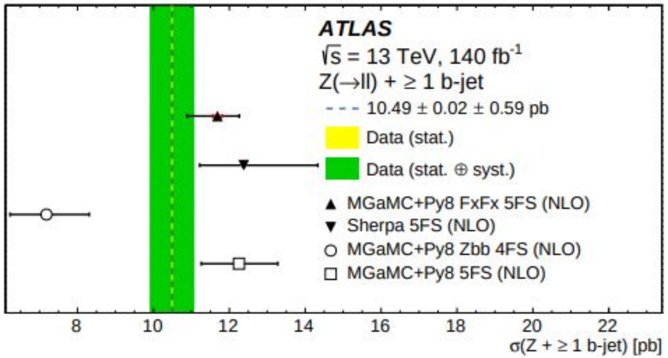
Uncertainty on the differential cross-section

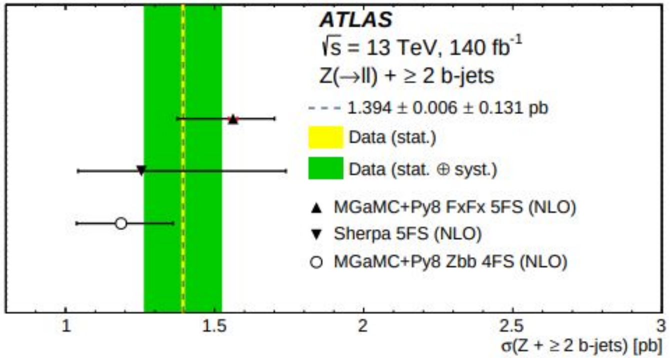
- The uncertainties at detector level (i.e. Lepton, b-tagging, jets, etc.), uncertainties on the backgrounds are propagated to the unfolding procedure.
- Uncertainty on the unfolding procedure also accounted for.
- The total uncertainty for Z + ≥ 1 b-jet ranges between 5 % and 10 %.
- The total uncertainty for Z + ≥ 2 b-jets ranges between 10 % and 15 % in most of the variables



Inclusive fiducial cross sections

σ(Z+≥1 b-jet) = 10.49 ± 0.02 (stat.) ± 0.59 (syst.) pb
σ(Z+≥2 b-jets) = 1.39 ± 0.01 (stat.) ± 0.13 (syst.) pb

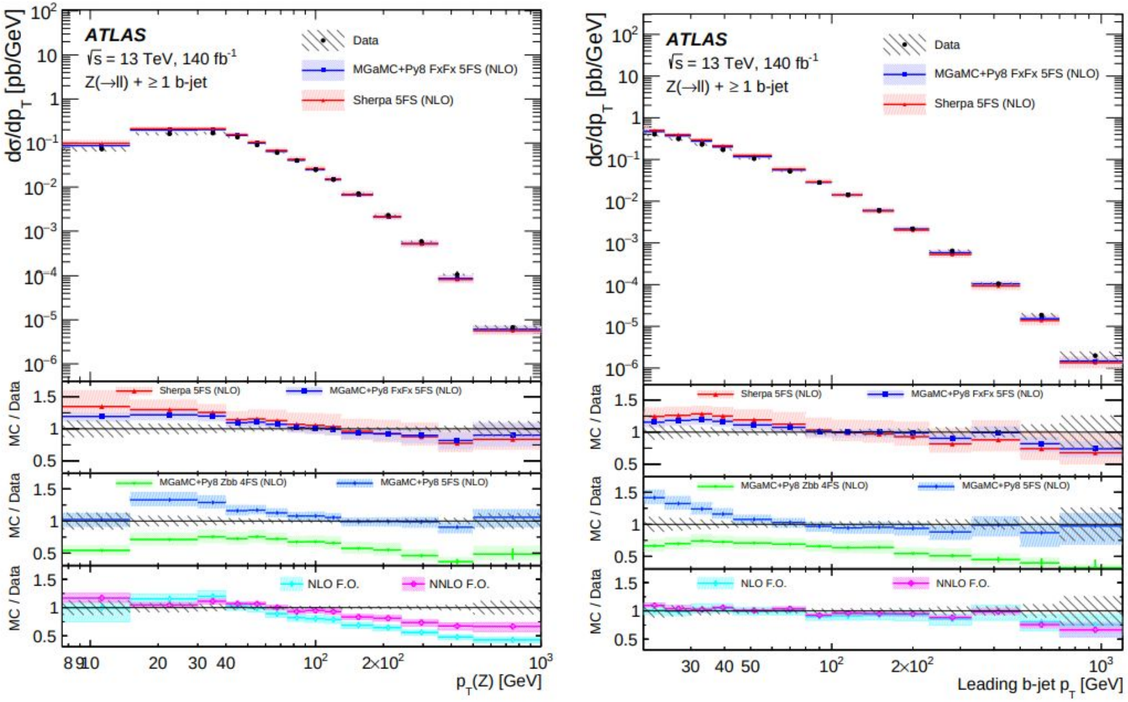




Uncertainty on the measurement: ~6% (Z+ ≥ 1 b-jet), ~10% (Z+ ≥ 2 b- jets)

- 5FS NLO ME+PS MCs (MGaMC+Py8 FxFx and Sherpa) describe Z+b and Z+bb, while 4FS Zbb NLO (MGaMC+Py8 Zbb) describes only Z+bb largely underestimating Z+b

Differential cross section measurements (Z+ ⩾ 1 b-jet)

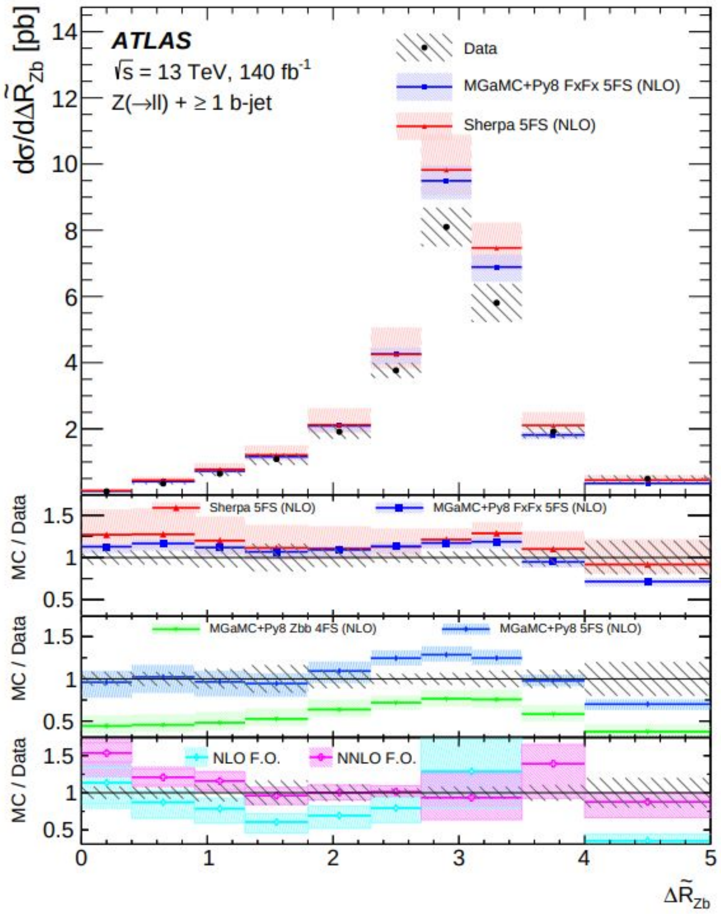


Important for test pQCD and for background prediction for other processes

4FS ME+PS MC largely underestimate data

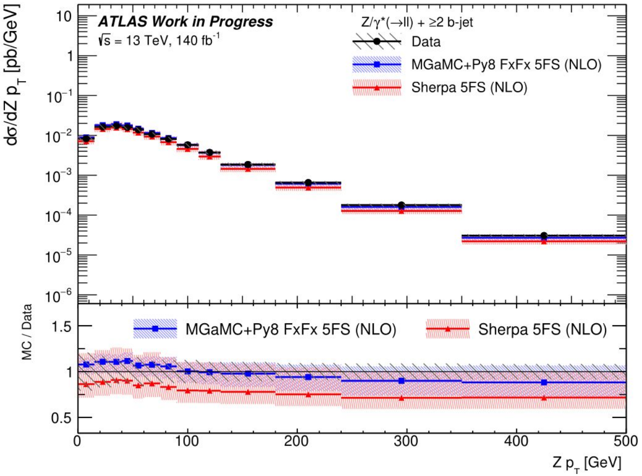
Fixed-order calculations show discrepancy with data in the high pT (Z) region

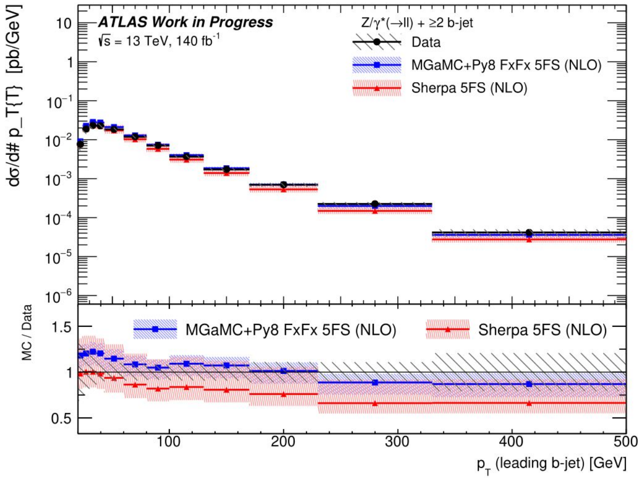
Differential cross section measurements (Z+ ⩾ 1 b-jet)



- Sensitive to additional radiation
- Good description for 5FS
- Mismodelling of 4FS for collinear and large Δ R(Z,b)

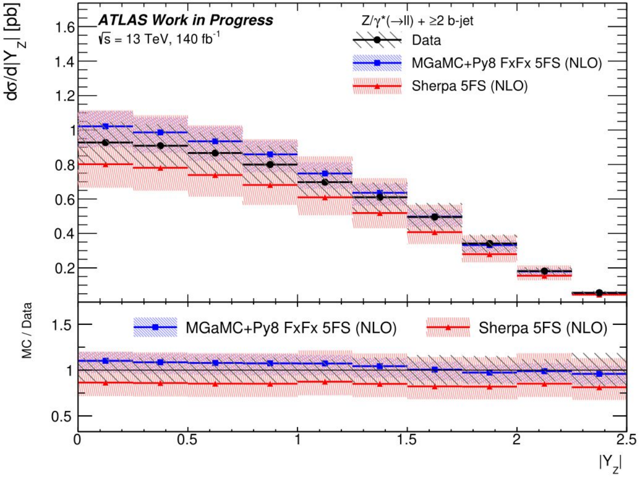
Differential cross section measurements (Z+ ⩾ 2 b-jets)

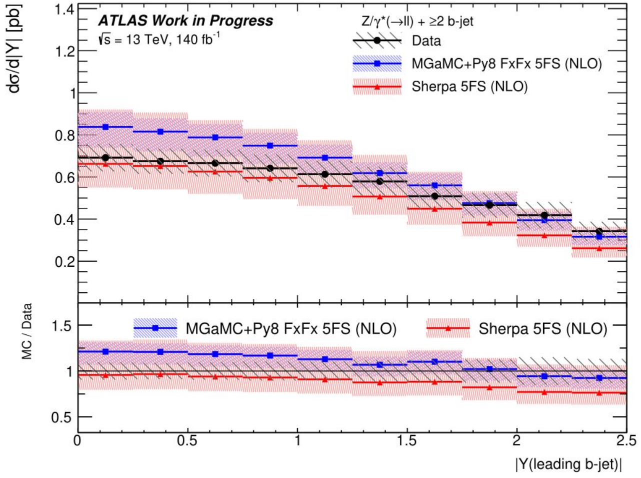




- Interesting for pQCD and MC modeling.
- Data slightly harder than predictions.
- MG+Py8 FxFx demonstrates the best agreement with data.

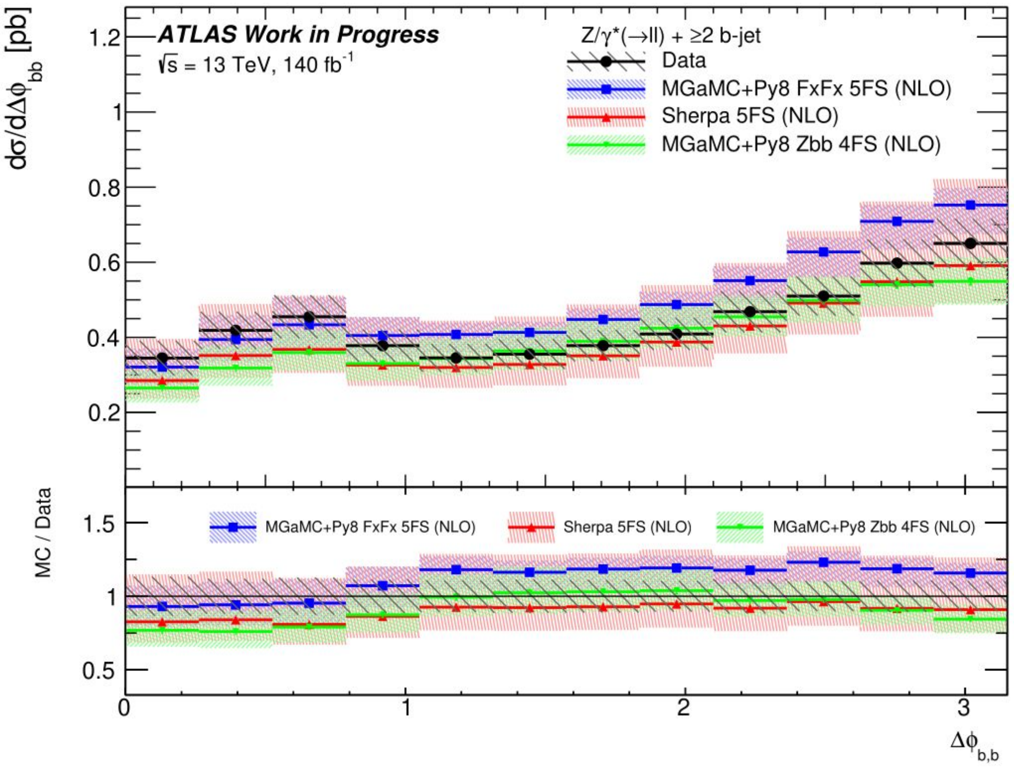
Differential cross section measurements (Z+ 2 b-jets)





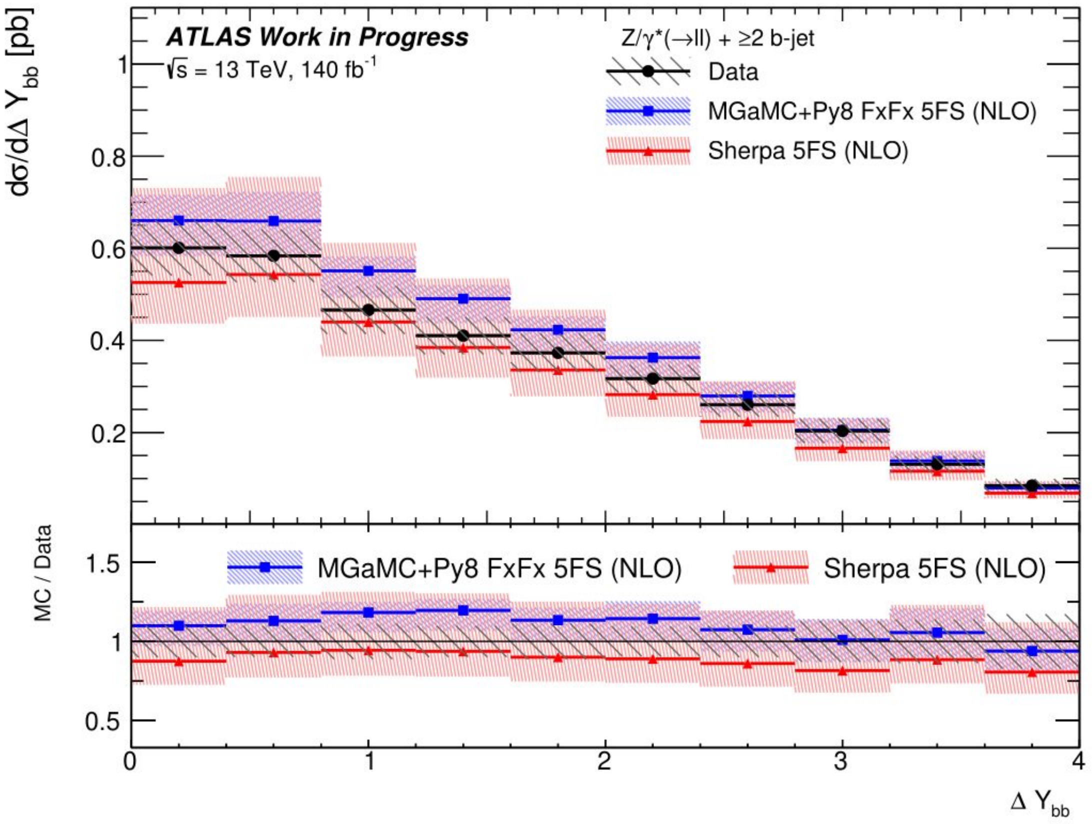
- Sensitive to PDFs in addition to probing pQCD.
- Good agreement between data and predictions.
- Sherpa describes very well the shape of the data.

Differential cross section measurements (Z+ 2 b-jets)



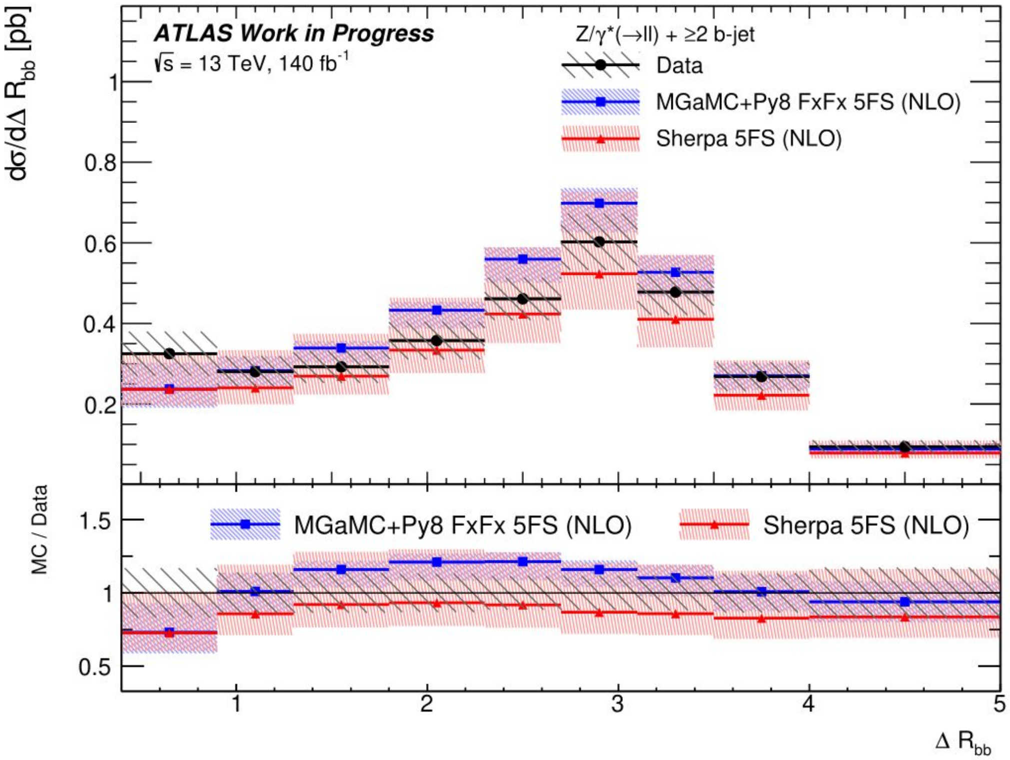
- Characterize soft and hard radiation
- MG+Py8 FxFx and Sherpa describe the data.
- MG+Py8 (4FS) slightly underestimate small and large values.

Differential cross section measurements (Z+ 2 b-jets)



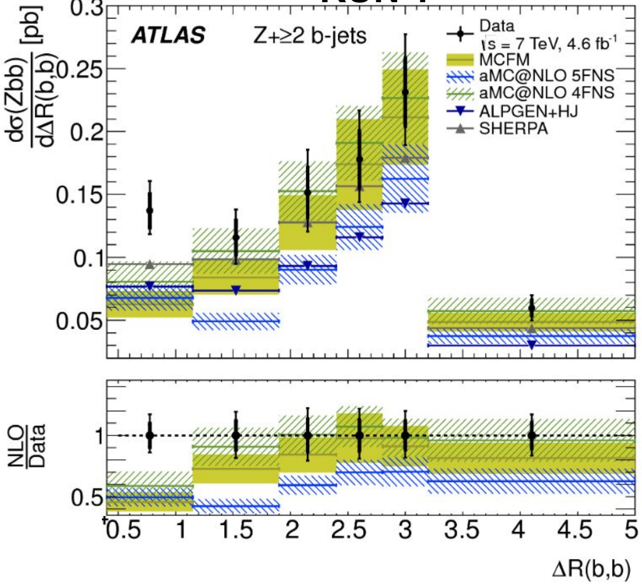
- Sensitive to PDFs.
- MG+Py8 FxFx and Sherpa describe the data shape well

Differential cross section measurements

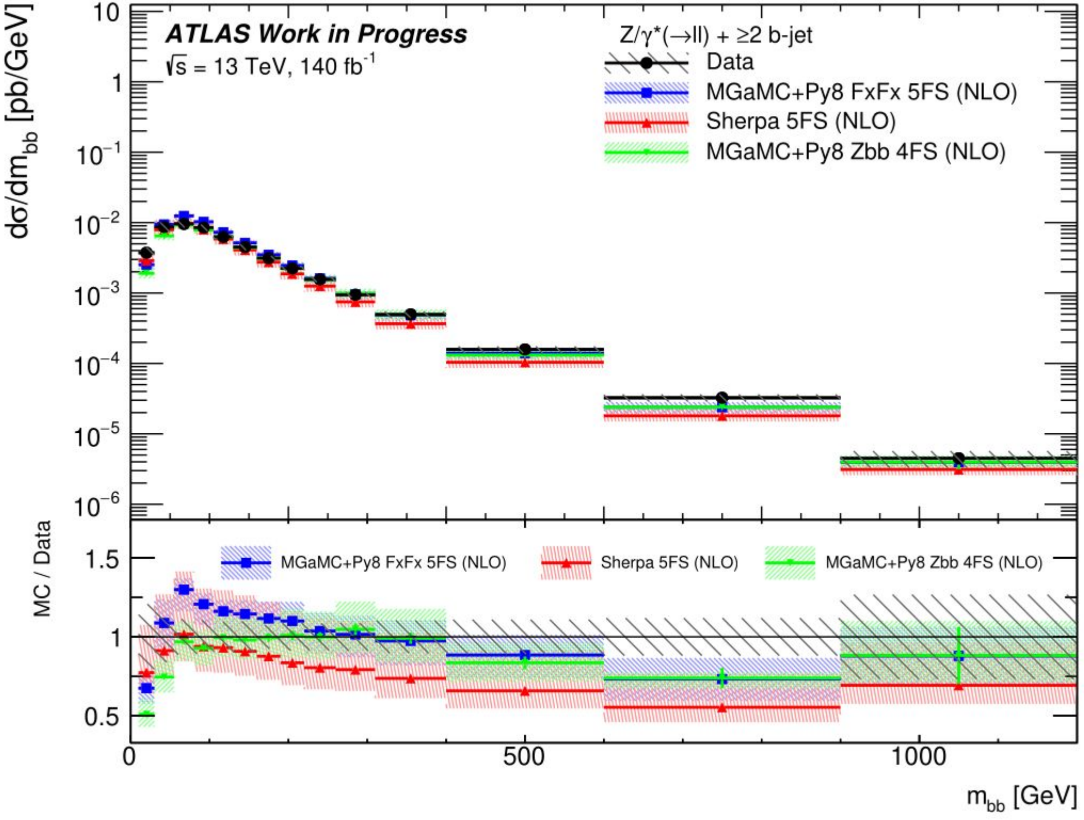


- Sherpa and MG+Py8 FxFx agree with data within the uncertainty.
- Sherpa describe the shape of this observable quite well.
- MG+Py8 FxFx shows some difference in shape.

RUN 1



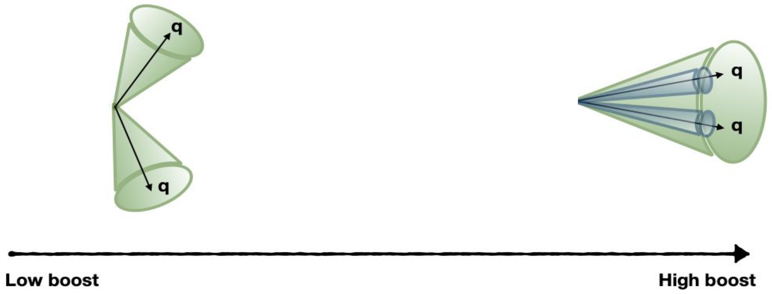
Differential cross section measurements (Z+ 2 b-jets)



- Important observable ZH (H→bb) production.
- They are still in agreement with data within the uncertainties.
- None of the predictions describe the shape of the data.

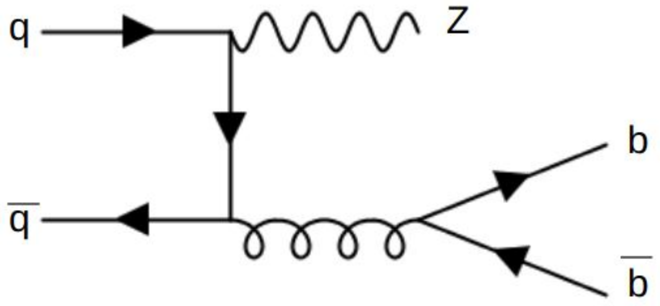
Study of the production of a Z boson in association with 2 b-jets in the boosted regime exploring also Jet Substructures

On-going analysis



Z+bb JSS analysis : motivations and strategy

- The radiation pattern around b-quarks largely unknown
- JSS observables never observed on HF jets
- Explore JSS sensitive to the color structure of the final state.
- High-momentum regime is particularly sensitive to modifications to SM dynamics by new physics.



- 2 kinematic regions (resolved and boosted) x 2 signals (Z+light jets, Z + b-jets):
  - Z + light jets measurements in resolved and boosted regions
  - Z + b-jets measurements in resolved and boosted regions

Proposed observables

resolved
Lund Jet Plane
p_T^{bb}
p_T^{bb}/m_{bb}
$\Delta R(bb)$

boosted
Lund Jet Plane
m_{FATJET}
$p_{T,FATJET}$
$\Delta R(bb)$

Paper at 13 TeV with 36 fb^{-1} :
Phys. Rev. D 108 (2023) 1, 012022

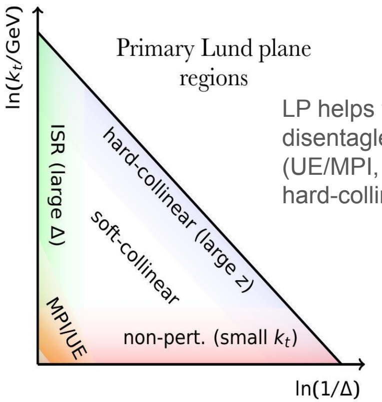
- Lund Plane measured both for Z + light jets and Z+ b-jets
- JSS under study (Color ring, ECFs)

Lund Jet Plane

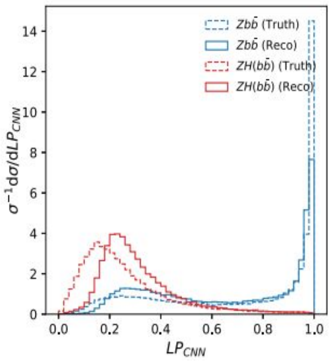
A jet can be considered as a constitute by softs emissions (b and c) around a hard core (a, the originating q/g)

$\Delta^2 = (y_a - y_b)^2 + (\phi_a - \phi_b)^2$ = angle of the emission wrt hard core

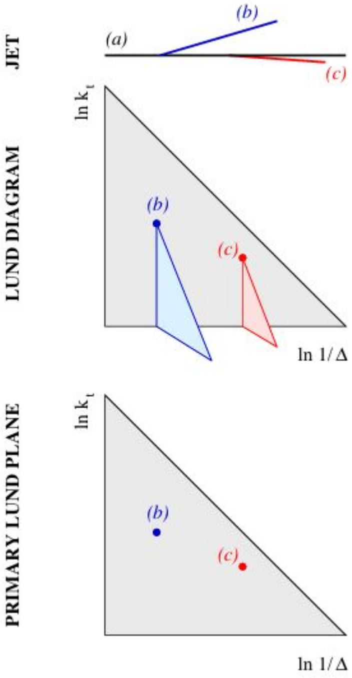
$k_t = p_{Tb}\Delta$ = transverse momentum of the emission w.r.t the jet axis



LP helps to disentangle different effects (UE/MPI, hard&wide-angles, hard-collinear)



LP might be used to develop taggers to discriminate H→bb and g→bb
[Eur. Phys. J. C (2022) 82:493]



Technically:

- Take anti-k_t clustered jet
- Re-cluster with Cambridge-Aachen
- “De-cluster” C/A jets (undo the last clustering step) and record the kinematics of the splitting.
- LP obtained by iterating the procedure, always following the hardest branch in each splitting

Event selection

N Leptons	
Exactly 2 leptons (2 muons or 2 electrons) with same flavour, opposite charge and isolated	
Leptons	$p_T > 27$ GeV, $\|\eta\| < 2.5$

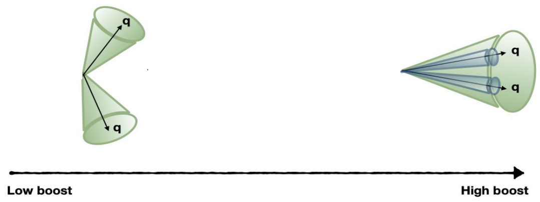
Z mass window
76 GeV $< m_{\ell\ell} <$ 106 GeV

Resolved Topology

- ≥2 b-tagged (70 % WP) jets (Anti-kt with R=0.4) with p_T> 20 GeV and |η| < 2.5

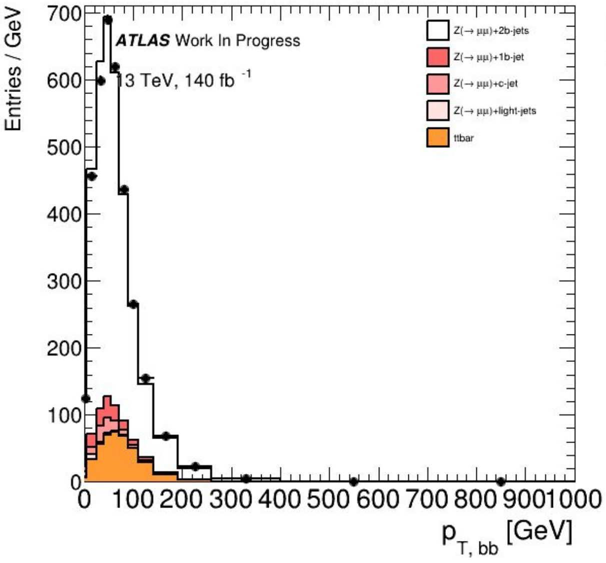
Boosted Topology

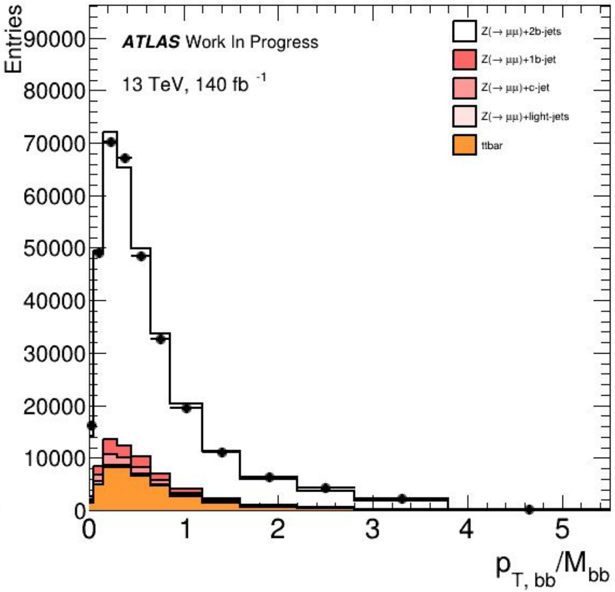
- ≥1 large R jet (R=1.0) with and p_T> 200 GeV and |η| < 1.5
- ≥2 b-tag (70 % WP) jets matched to a fatjet

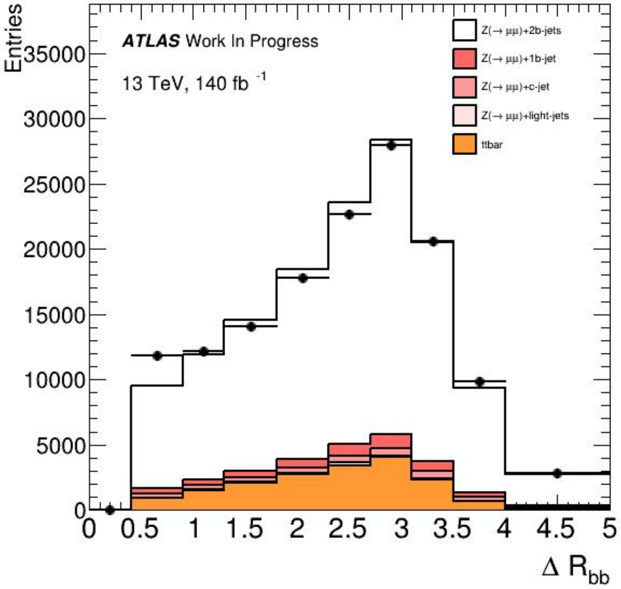


Detector level plots (resolved region)

Very Preliminary



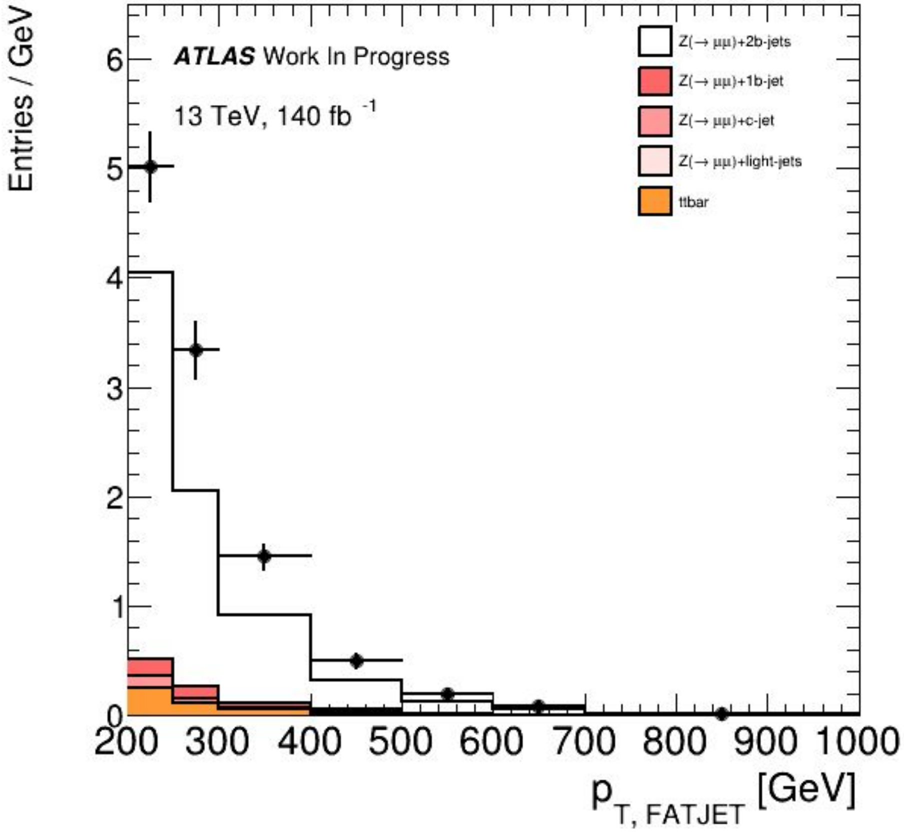




- Low background contamination (difference in WP)
- Good agreement with data

Detector level plots (boosted region)

Very Preliminary



- Distribution doesn't fully agree with data
- Low statistics (plan to move the η cut)

Conclusions

- Z+b-jets measurements provide essential inputs for the improvement of MCs modeling, allowing a better understanding of pQCD and useful inputs for PDFs fits.
- Latest ATLAS Z+b-jets measurement with 140 fb^{-1} at 13 TeV in resolved regime:
  - 4 FS MCs underestimate Z+b data
  - 5 FS MCs describe better the data, but discrepancies for some observables in specific phase phases observed (MG FxFx describes better pT observables, Sherpa describes better angular variables in Z+bb topology)
- New high profiled measurements in this sector are coming through:
  - the investigation of the boosted regime and jets substructure observables in Z+bb (on-going)
  - W+bb (not yet started), ratio measurements (not yet started)

Thanks for the attention!

BACK-UP SLIDES

Sequential Algorithm k_T

- For each pair of particles i,j the distance is defined as:

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

- and for each particle the distance with respect to the beam is defined as

$$d_{iB} = p_{ti}^2$$

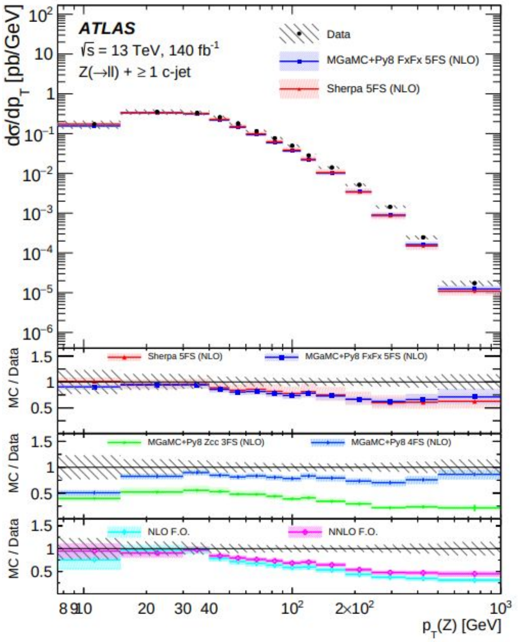
- Looking for small distances:
  - If is d_{ij} two particles are combined in a new one
  - If it is d_{iB} declare i as jet and remove it from the particle list

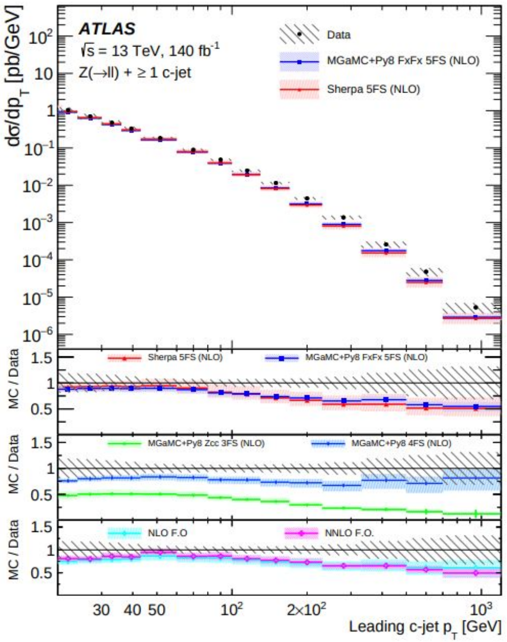
- The procedure is repeated until there are no more unexamined particles

Uncertainties

Source of uncertainty	$Z(\to \ell\ell) + \geq 1$ b-jet [%]	$Z(\to \ell\ell) + \geq 2$ b-jets [%]	$Z(\to \ell\ell) + \geq 1$ c-jet
Flavour tagging	3.6	5.7	10.3
Jet	2.4	4.3	6.5
Lepton	0.3	0.3	0.4
$E_{\mathrm{T}}^{\mathrm{miss}}$	0.4	0.5	0.3
Z+jets background	0.6	1.5	1.6
Top background	0.1	0.3	<0.1
Other backgrounds	<0.1	0.2	0.1
Pile-up	0.6	0.6	0.2
Unfolding	3.3	5.8	5.0
Luminosity	0.8	0.9	0.7
Total [%]	5.6	9.4	13.2

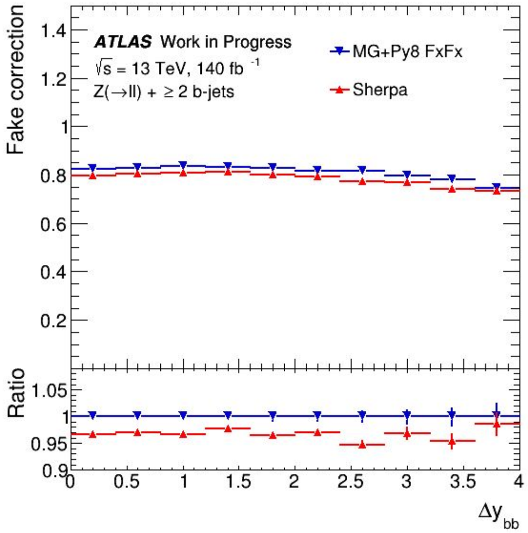
Differential cross section measurements (Z+1 c-jet)

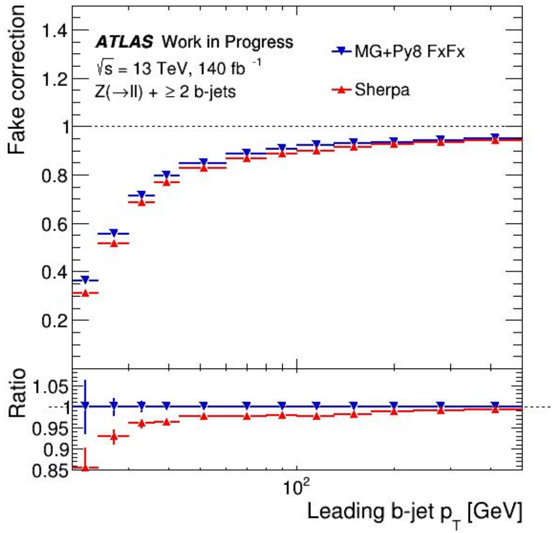




Fake corrections

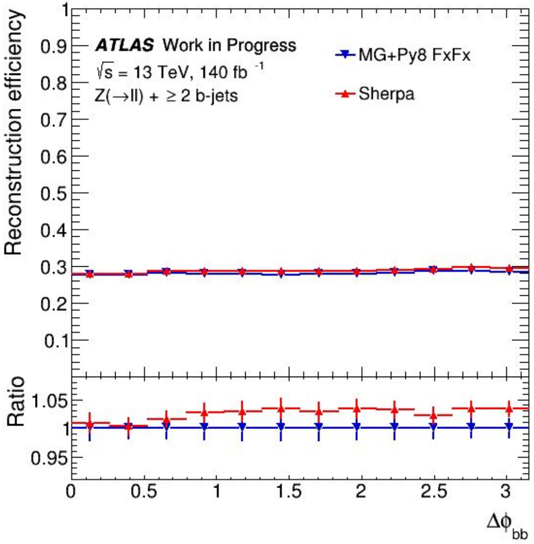
- Corrections are applied for fake events, so events that fall outside the considered phase space at the particle level but migrate into the studied region at the detector level.
- It is the ratio between the number of reconstructed events that are matched to a truth event (i.e., events that pass both the reconstruction and truth-level selections) to the total number of reconstructed events (i.e. events that pass the reconstruction regardless the truth-level selection).

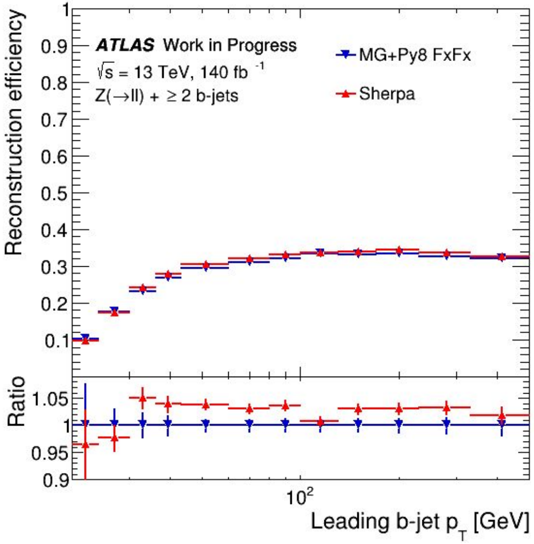


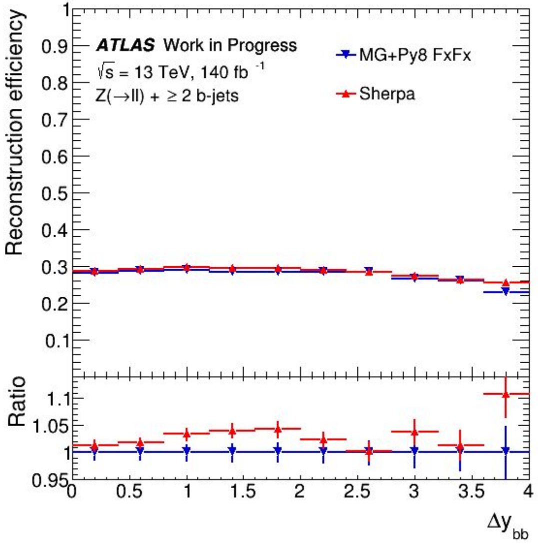


Reconstruction efficiencies

- It is the ratio between the number of reconstructed events that are matched to a truth event (i.e., events that pass both the reconstruction and truth-level selections) to the total number of truth events.







Bayesian iterative procedure

- In the Bayesian iterative procedure, the relation between the distribution of an observable α at particle level T(α) and at detector level R(α) is:

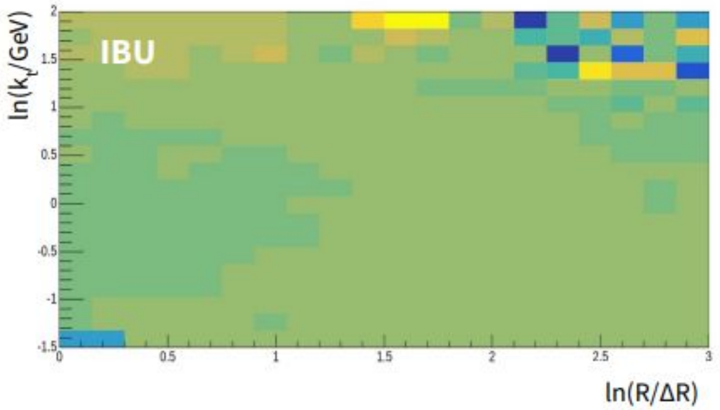
$$R(\alpha_i) = \sum_{j=1}^{N_T} M_{ij} T(\alpha_j) = \sum_{j=1}^{N_T} M(R(\alpha_i)|T(\alpha_j))T(\alpha_j)$$

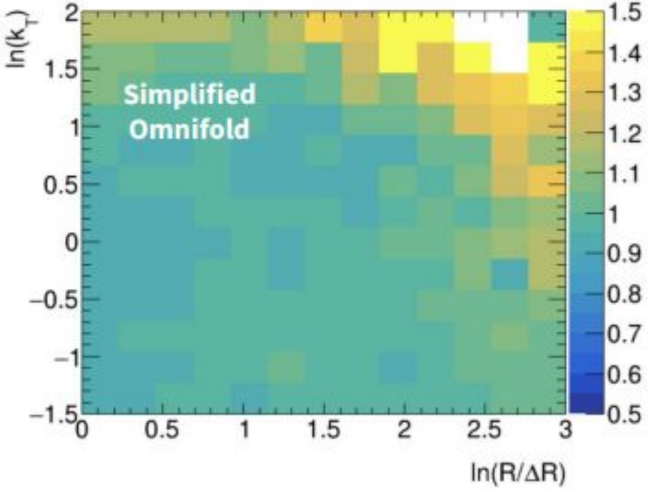
- To produce the unfolding matrix a MC simulation is chosen as a priori, so the probability set of the hypothesis T(α).
- For the first iteration the prior is taken from the particle-level prediction, for the second iteration from the unfolded distribution of the previous iteration.

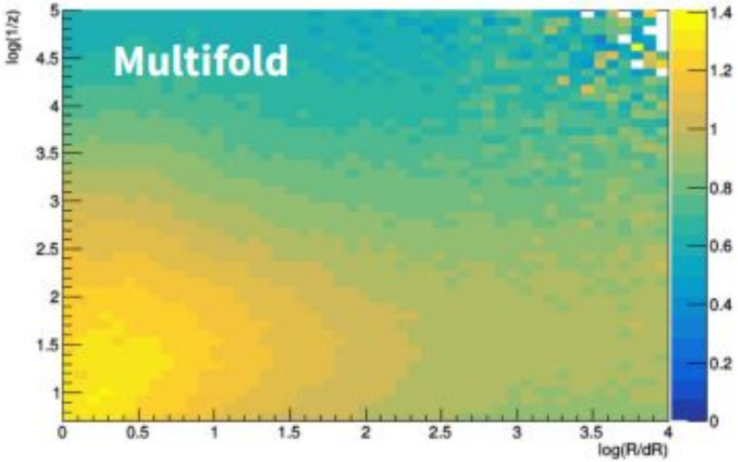
Unfolding

Different unfolding strategies:

- IBU
- Multifold with DNN
- Omnifold with LundNet GNN



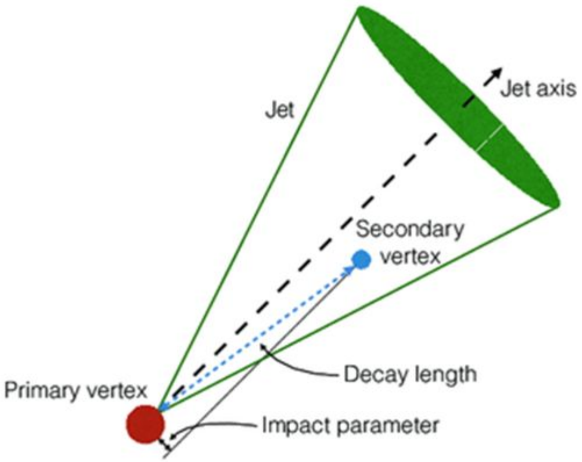




Flavour tagging calibration and algorithm

Two level algorithms:

- Low Level Algorithm: inputs → impact parameters and traces
- High Level Algorithms: inputs → outputs of the low level algorithm, outputs → probability that a jet derives from the adronization of a specific quark. The most common are: MV2c10 and DL1r.



Calibrations techniques correct for efficiency differences between data and simulations and this is done through a scale factor

$$\mathrm{SF}^f = \frac{\varepsilon^f_{\mathrm{data}}}{\varepsilon^f_{\mathrm{MC}}}$$

$$\varepsilon^f = \frac{N^f_{\mathrm{pass}}}{N^f_{\mathrm{all}}}$$

Muons and electrons identification efficiency

- The electrons are reconstructed matching the tracks in the inner detector and the energy deposit in the electromagnetic calorimeter.
- Electrons are identified using a likelihood (LH) method. The LH identification method distinguishes between signal and background by calculating likelihoods for both signal (LS) and background (LB).
- Three operating points ordered in decreasing identification efficiency with an inversely proportional increase in electron signal purity : LooseLH, MediumLH and TightLH

- Muons are charged particles that leave tracks in the inner detector (ID), muon spectrometer (MS) and occasionally small energy deposits in the calorimeters.
- Similar to electron identification, muons are identified using certain discriminating variables with four identification working points: Loose, Medium, Tight and High-p_T.

$$L_{S(B)}(x) = \prod_{i=1}^{n} P_{S(B),i}(x_i)$$

$$d_L = \frac{L_S}{L_S + L_B}$$

JES and JER

- Jet Energy Scale (JES) uncertainties are measured by calibrating jet energies using in-situ techniques and comparing data to Monte Carlo simulations.
- Jet Energy Resolution (JER) uncertainties are determined by evaluating the resolution in data using similar techniques and applying smearing corrections to match Monte Carlo truth-level predictions.
- Both JES and JER uncertainties depend on factors like jet p_T, η, pile-up, calorimeter response, and modeling differences between data and simulations.
- These uncertainties are systematically propagated to analyses, ensuring their impact on key observables and reconstructed objects is well accounted for.