

UNIVERSITÀ DEGLI STUDI DI PAVIA

DOCTORAL THESIS

Testing the Standard Model with precision flavour measurements at the ATLAS experiment

Author:
Giulia MANCO

Supervisor:
Prof. Daniela M. REBUZZI

*A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Physics*

in the

ATLAS experiment
Dipartimento di Fisica

March 26, 2024



Contents

Introduction	1
1 Introduction to particle physics	5
1.1 The Standard Model	5
1.1.1 Particle content	5
1.1.2 The fundamental interactions	6
1.2 Quantum Chromodynamics	7
1.2.1 The QCD Lagrangian	7
1.2.2 Asymptotic freedom and confinement	10
1.2.3 Parton model and PDFs	12
1.3 Electroweak interactions	16
1.3.1 The Gauge field term	16
1.3.2 The Higgs term	17
1.3.3 The fermion term	20
2 Physics at the ATLAS experiment	23
2.1 Introduction	23
2.2 Physics at colliders	23
2.2.1 Center-Of-Mass Energy, Cross Section and Luminosity	24
2.2.2 Kinematic variables	25
2.2.3 Jets and event shape variables	26
2.3 The ATLAS detector	27
2.3.1 The ATLAS detector coordinate system	28
2.3.2 The Inner detector	28
2.3.3 Calorimeters	30
2.3.4 Muon System	30
2.4 Selection and Data	37
2.4.1 Trigger and data acquisition system	37
2.4.2 Data processing	39
2.5 Reconstruction and identification of physics objects	40
2.5.1 Muon reconstruction and identification	40
2.5.2 Electron reconstruction and identification	41
2.5.3 Jet reconstruction	42
3 The ATLAS Online Monitoring system	45
3.1 Introduction	45
3.2 Monitoring framework	46
3.3 Monitoring Tools	47
3.3.1 Muon Spectrometer DQMF	50
3.3.2 Muon Spectrometer OHP	60

4	Flavour tagging in ATLAS and new techniques	69
4.1	Flavour tagging in ATLAS	69
4.1.1	Low-level b -taggers	71
4.1.2	High-level flavour-taggers	73
4.1.3	Flavour-tagging performance	75
4.1.4	Double b -tagging in boosted Higgs $b\bar{b}$ decay	77
4.2	$Hb\bar{b}$ tagging and colour sensitive variables	79
4.2.1	Colour sensitive variables	80
4.2.2	$VHb\bar{b}$ simulation and selection	84
4.2.3	Performances	86
5	$W+c$ analysis	91
5.1	The motivation	91
5.2	Measurement strategy	93
5.2.1	Background sources	95
5.3	Data and Simulation	96
5.3.1	W +jets samples	99
	Generator level filters	99
	Baseline sample	100
	Alternative sample	100
5.3.2	Background Samples	103
	Z +jets	103
	Di-top ($t\bar{t}$)	103
	Single top	103
	Dibosons (Db)	103
5.4	Selections	103
5.4.1	Primary Vertices	104
5.4.2	Electrons	104
5.4.3	Muons	104
5.4.4	Jets	104
5.4.5	Overlap removal	105
5.4.6	Soft muon tagging	105
5.4.7	b -tagging	105
5.4.8	Missing transverse momentum	106
5.4.9	Trigger	106
5.4.10	Calibrations and corrections	106
5.4.11	Event Selection	107
	DL1r cut	108
5.4.12	Definition of particle-level objects	109
5.4.13	Fiducial phase space for the measurement and particle level selection	110
5.4.14	W +jets composition studies	110
5.4.15	Yields and control plots	111
6	Statistic analysis and results	117
6.1	Maximum Likelihood Fit	117
6.1.1	Fit Strategy	118
6.2	Multijet estimation	120
	ABCD Method	121
	ABCD Likelihood-based Method	123
6.3	W +light background estimation	124

6.4	Fit results	125
	Fit tests	135
6.5	Unfolding procedure	136
	6.5.1 The mathematical problem	137
	6.5.2 The Iterative Bayesian Unfolding	138
	6.5.3 Algorithm validation and choice of iterations	139
	6.5.4 Post-Unfolding Results	141
6.6	Systematic Uncertainties	141
	6.6.1 Systematic uncertainty evaluation	143
	6.6.2 Detector-level systematic uncertainties	143
	6.6.3 Signal and background systematic uncertainties	145
6.7	Results	146
Conclusions		151
A Detector-level systematic uncertainties		153
B MDT Algorithms		155
Bibliography		159

Introduction

The Large Hadron Collider (LHC) at CERN is a 27 km long accelerator, situated between France and Switzerland. It is currently operative and collides protons at the center of mass energy of $\sqrt{s} = 13.6$ TeV, just below the design energy of 14 TeV. It has been created in order to search for new physics and to consolidate or disprove the Standard Model (SM) theory by means of high resolution physics measurements.

ATLAS is one of the biggest experiments operating at LHC and has been optimized in order to search for the Higgs boson and to favour new phenomena discovery. In thirteen years, ATLAS achieved many important physics results, among which the discovery of the Higgs boson in 2012. Nevertheless, tests of the SM are still ongoing in order to obtain more precise measurements which could act as portal for new physics. In this phase of precision physics, the quark flavour discrimination plays an important role: the access to the flavour information allows to isolate and identify particles in processes.

The precise measurement of the W boson cross section, when produced in association with charm quark, is one of the main topics covered by this manuscript and during my three-years PhD work. This process is extremely important because it provides direct information on the parton distribution function (PDF) of the strange sea-quark in the proton. This is matter of interest since long time and several experiments have put efforts in its determination. Past DIS nucleon-neutrino experiments at Tevatron, residing at Fermilab, found a suppression of the strange quark in comparison with the other sea-quarks of the proton. ATLAS experiment, in the QCD fit with HERA and W, Z 7 TeV data, found instead an unsuppressed strange PDF both at low and high momentum fraction x . The last ATLAS PDF fit confirms the unsuppressed PDF strange at low x , but observes a suppression at high x . The goal of this analysis is to provide the global fits with additional input data, in order to refine the strange-PDF results.

The $W+c$ final state is reconstructed through the W leptonic decay, while the c -jet is identified through a Soft-Muon Tagger. The latter consists of selecting a muon with low p_T inside the jet, produced by the c -hadron semi-leptonic decay. The measurement is performed in two channels according to the lepton, electron or muon, coming from the W -lepton decay.

The analysis strategy developed during my PhD work consists in exploiting the charge correlation between the W -lepton and the c -soft-muon: the signal is characterized by an anti-correlation between the charge sign of the two leptons, while the background is almost symmetric between opposite and same sign. The subtraction of these two regions removes large part of backgrounds, except for the multijet and W +jets backgrounds, which are slightly asymmetric. For these two backgrounds the estimation is provided by a data driven methods instead of by Monte Carlo simulations. The multijet background is estimated by the ABCD method, which consists of defining control regions with as little signal-contamination as possible where deduce the background in the signal region. For the W +light jets, a correction factor is assessed by a simultaneous fit. The signal estimation is also provided by the fit, which include the background estimation as well. The fit has an innovative structure:

it is in 2D dimensions. The first dimension is the interesting variable, as default the lepton rapidity. The second dimension is a two bins pseudo-continuous- b -tagging variable (PCBT): PCBT bins correspond to different b -tagging efficiencies. The first bin is light-type, which means that selects jets with light flavours, while the last bins select jets from b -quarks. This kind of selection allows to have one region where the W +light quark jets has the largest contribution.

As a last step, an unfolding technique is applied on the signal, in order to reduce the detector effects and obtain a result independent from experiments.

As ATLAS analysis contact of this analysis, during my PhD period I covered the role, together with an another colleague, to guide the working group to fulfill the analysis duties.

A second part of my PhD work deals with flavour quark processes. In particular, it addresses the problem of b -jet identification, when they are coming from the decay of a colour singlet, i.e., the Higgs boson, or from the QCD background (colour octets). This study started as a collaboration between experimental and theoretical physicists, who aimed at studying a $Hb\bar{b}$ tagging algorithm by means of colour sensitive variables. This study focuses on the associated production with a vector boson of the Higgs boson, then decaying into a pair of b -jets in the boosted regime, exploiting theory-driven variables sensitive to the colour. The idea is to create an innovative $Hb\bar{b}$ tagger, using the combination of different colour sensitive variables embedded in a machine learning algorithm. The variables chosen in the tagger are: *jet pull* variables, *jet colour ring*, D_2 and *Lund jet plane*. The result shows good discrimination performances, in particular thanks to the Lund jet plane variable. The tagger was developed using standalone Monte Carlo samples, with a fast simulation for the ATLAS detector response estimation. The goal of this study is to prove the effectiveness of the colour sensitive variables with the aim of adapting them in the general ATLAS tagger.

During my three-years of PhD, I have also worked in the data quality (DQ) field. More precisely, starting from my ATLAS qualification, I have been involved in the Online Data Quality Monitoring activity for the MDT and NSW, muon spectrometer sub-detectors. My activity consists in creating and maintaining the monitoring tools in the ATLAS Control Room, which are the Online Histogram Presenter (OHP) and the Data Quality Monitoring Framework (DQMF). This activity allowed me to become a DQ Muon Online expert and to take responsibilities in this field.

This manuscript is organised as follows. Chapter 1 is dedicated to shortly review the particle physics framework. In particular, Section 1.1 is a general introduction of the Standard Model, the theory of elementary particles and their interactions. In Section 1.2, the Quantum Chromodynamics is presented, including the Parton model and PDFs (1.2.3), useful to understand the strange sea-quark PDF in the $W+c$ analysis. A general introduction of the Electroweak interactions is given in Section 1.3, with an overview on the Higgs sector (1.3.2), and, in the end, some ingredients for the physics at colliders are discussed in Section 2.2.

Chapter 2 is focused on the ATLAS detector (2.3). The overview of the selection and data in the experiment follows in Section 2.4, while the last Section explains the reconstruction and identification of the majority of physics objects (2.5) necessary in the $W+c$ analysis.

Chapter 3 is dedicated to the ATLAS Online Monitoring system. After a brief introduction in Section 3.1 and a description about the DQ Online frameworks (3.2), the Monitoring Tools are shown in 3.3, focusing on the presentation of the algorithms I developed during my PhD work.

In Chapter 4, the flavour tagging techniques in ATLAS and the new $Hb\bar{b}$ tagger are

presented. In particular, Section 4.1 is an overview of the taggers used by ATLAS for flavour identification, while Section 4.2 presents the study about the new $Hb\bar{b}$ tagger I have contributed to develop, exploiting colour sensitive variables.

In Chapter 5, the $W+c$ analysis is fully presented. Section 5.1 details the motivation behind the analysis and its state of the art, then Section 5.2 describes the signal and background estimation strategy. Section 5.3 specifies the MC samples and data chosen for this analysis, while in Section 5.4 all the selections tested and tuned for improving the background rejection and the signal purity are explained in details. Moving to Chapter 6, in Section 6.1, a brief introduction about the maximum likelihood fit is presented, along with the fit strategy of the measurement. The multijet and W +jets background estimation are exposed in Sections 6.2 and 6.3 respectively. The unfolding procedure is explained in Section 6.5, where, after a general overview on the technique, the post-unfolding results of the analysis are shown.

Chapter 1

Introduction to particle physics

1.1 The Standard Model

The Standard Model (SM), in particle physics, is the theory developed in the second half of the 20th century, which provides a picture of the universe in terms of the elementary known particles and their possible interactions. The SM gives a successful description of almost all the experimental data, nevertheless it does not include the gravity force and it presents unsolved open questions. Theorists are working on alternative theories, tested by several experiments, globally known as beyond Standard Model (BSM) theories, which incorporates the SM in the low energy regime. In this chapter, I will give an introduction about the SM without any reference to the BSM theories. More in details, in section 1.1.1, I will give an overview about the particles included in the SM, while in section 1.1.2, I will present the fundamental forces.

1.1.1 Particle content

In particle physics, particles are divided in two categories: fermions and bosons.

Fermions follow the Fermi-Dirac statistics and they have a half-integer spin. They include *quarks* and *leptons* and each group consists of six particles together with the correspondent anti-particles, which, as consequence of the Dirac equation, have exactly the same masses, but opposite charges.

As shown in the SM picture shown in Figure 1.1, there are three families or *generations* of quarks and leptons. Inside each generation, fermions differ only in masses, while they have the same proprieties and interactions.

The *first generation* includes the electron (e^-) and the electron neutrino (ν_e) as leptons and includes the up (u) and down (d) as quarks. The *second generation* includes muons (μ^-) and muon neutrinos (ν_μ) as leptons, strange (s) and charm (c) as quarks. Finally in the *third generation* tau (τ^-) and tau-neutrino (ν_τ) as leptons, top (t) and bottom (b) as quarks.

Fermions, with the exception of neutrinos, are electrically charged and feel the electromagnetic interaction. Only quarks are sensitive to the strong force, since they also carry *colour charge*. Quarks are never observed as free particles, but they are always confined in bound states, called *hadrons*, as protons and neutrons. The hadronic states so far observed are the *mesons* and *baryons*, including some exotic states. Mesons consist of a quark (q) and antiquark ($\bar{q}$) pairs, while baryons consist of three quarks (qqq) or antiquarks ($\bar{q}\bar{q}\bar{q}$). The only stable hadrons are the protons and the neutrons. The latter, as free particle, decays in 15 minutes, while all the other hadrons decay usually very rapidly.

Bosons are integer-spin particles which obey to the Bose-Einstein statistics. The bosons are further classified in *gauge bosons* and *scalar bosons*. The gauge bosons

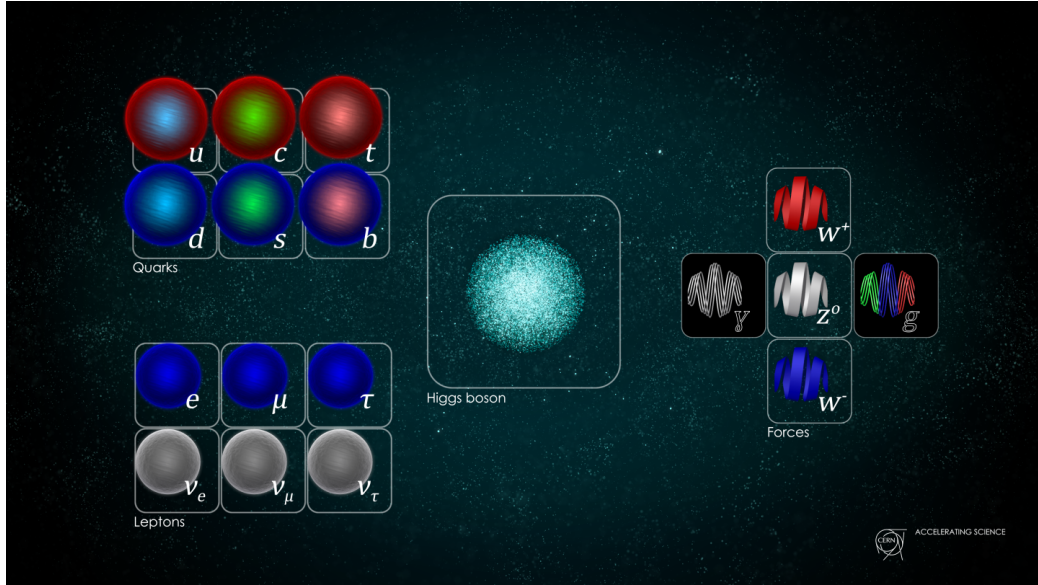


FIGURE 1.1: Particles of the Standard Model [1].

are responsible for the force interactions and they consist in the photon γ , for the electromagnetic interaction, $W^\pm$ and Z^0 bosons for the weak interaction and gluons g for the strong force. Among the scalar bosons, there is only the Higgs boson (H), which, differently from the others which are spin-1 particles, is the only SM particle carrying spin-0.

The Higgs is the last SM piece discovered in 2012 by the ATLAS and CMS experiments at the Large Hadron Collider (LHC) at CERN [2, 3]. The discovery was important since the Higgs boson provides an explanation for the electroweak symmetry breaking and it is part of a mechanism by which all other particles acquire mass. More details are given in the *Electroweak interactions* section 1.3.

1.1.2 The fundamental interactions

In nature, four forces exist: electromagnetic, strong, weak and gravity. The SM provides an accurate description of the first three, while it does not include the fourth one. Although the exclusion of the gravity is a weakness of the model, the gravity strength is negligible at the short scale at which particle physics generally takes place. In modern particle physics, each interaction is modelled by a Quantum Field Theory, which considers particles as excited states of their quantum fields. The force or interaction is described by a Lagrangian density, function of particles and fields.

Each force is mediated by the exchange of a spin-1 gauge boson. For the electromagnetic interactions, the theory is the **Quantum Electrodynamics** (QED), in which charged particles interact via by exchange of photons. The strong force theory is the **Quantum Chromodynamics** (QCD), where colour-charged particles interact by the exchange of gluons. The QCD theory is briefly described in section 1.2.

The weak force is responsible for the β -decay and nuclear fusion and, differently from the other forces, can generate either charged- or neutral-current. The weak interaction is generally described within the **Electroweak Theory** (EW), the unification of the weak and electromagnetic forces. The *electroweak* theory is summarised in section 1.3. The weak charged-current interactions are mediated by the W bosons, while the weak neutral-current interactions have the Z bosons as propagator.

A particle can couple with a gauge boson only if it carries the charge of its specific interaction. Only electric charged particles can in fact couple with the photon, which excludes only the neutrinos. Regarding the QCD, only quarks can interact with gluons, since they possess the colour charge of the strong interaction.

All the twelve fermions can interact via the weak force, carrying the isospin, the charge of the weak interaction. Since the W boson can have either negative or positive electric charge, the weak charged-current can couple only fermions which differs by one unit of electric charge. This means that, for leptons, each charged particle can interact with its corresponding neutrino. For quarks, the interaction is similarly possible when it couples quarks differing by one unit of charge: this means link between up-type and down-type quarks. The weak interaction is the only known force in which the quarks can swap their flavour into another.

1.2 Quantum Chromodynamics

In this section, a brief overview of the Quantum Chromodynamics is provided, based mainly on the *QCD and Collider Physics* book of R.K. Ellis, W.J. Stirling and B.R. Webber [4]. The QCD theory is summarised in section 1.2.1, where the Lagrangian density and the Feynman diagrams are presented. In the following section 1.2.2, the QCD keys concepts of asymptotic freedom and confinement are exposed, introducing the running of the strong coupling. Other QCD-related topics, as quark model, PDF and DGLAP equations are mentioned in the last section 1.2.3 of this chapter.

1.2.1 The QCD Lagrangian

The Quantum Chromodynamics is the strong interaction field theory and describes the interaction between quarks and gluons. This theory is based on a non-Abelian gauge group, with $SU(3)$ as symmetry group, which implies that the gluon fields can interact with each other. At high energies the strong coupling becomes small and the theory becomes perturbative: the quantities can be written as small corrections respect the leading order term.

In field theory, analogously as for mechanical theory, a particle system is described through the *Lagrangian* formalism. The Lagrangian density describes a system of particles and fields with an infinite number of degrees of freedom. For the sake of simplicity, we will shortly call Lagrangian the Lagrangian density.

Free field propagation, conserved quantities, interactions and the Feynman rules are all derived from the Lagrangian description. The rules allow to show all the possible particle interactions of the system as diagrams, which are called *Feynman diagrams*.

The QCD Lagrangian has three terms: classical, gauge-fixing and ghost:

$$\mathcal{L}_{\text{QCD}} = \mathcal{L}_{\text{classical}} + \mathcal{L}_{\text{gauge-fixing}} + \mathcal{L}_{\text{ghost}}. \quad (1.1)$$

The classical Lagrangian describes the interaction of spin- $\frac{1}{2}$ quarks of mass m and massless spin-1 gluons:

$$\mathcal{L}_{\text{classical}} = -\frac{1}{4}F_{\alpha\beta}^A F_A^{\alpha\beta} + \sum_{\text{flavours}} \bar{q}_a (i\mathcal{D} - m)_{ab} q_b. \quad (1.2)$$

In the equation 1.2, $\mathcal{D}$ is a symbolic notation for $\gamma_\mu D^\mu$, where D_μ is the gauge covariant derivative and γ_μ are the Gamma matrices, which satisfy the anticommutation

relations:

$$(\gamma^\mu, \gamma^\nu) = 2g^{\mu\nu}. \quad (1.3)$$

$F_{\alpha\beta}^A$ is the field strength tensor derived from the gluon field $\mathcal{A}_\alpha^A$:

$$F_{\alpha\beta}^A = [\partial_\alpha \mathcal{A}_\beta^A - \partial_\beta \mathcal{A}_\alpha^A - gf^{ABC} \mathcal{A}_\alpha^B \mathcal{A}_\beta^C] \quad (1.4)$$

The ABC indices run over the 8 *colour* degrees of freedom, while f^{ABC} are the structure constants of the SU(3) colour group¹.

The third term in equation 1.4 derives from the *non-abelian* structure of SU(3) group and this also implies that spin-1 propagators can self-interact. Quarks q_a are written in the triplet representation of the colour group ($a=1,2,3$).

The covariant derivative takes the form of:

$$(D_\alpha)_{ab} = \partial_\alpha \delta_{ab} + ig(t^C A_\alpha^C)_{ab} \quad (1.5)$$

when acting on the triplet and

$$(D_\alpha)_{AB} = \partial_\alpha \delta_{AB} + ig(T^C A_\alpha^C)_{AC} \quad (1.6)$$

if it acts on the octet fields, where t and T are matrices in the fundamental and adjoint representations of SU(3):

$$[t^A, t^B] = if^{ABC} t^C, \quad [T^A, T^B] = if^{ABC} T^C, \quad (T^A)_{BC} = -if^{ABC}. \quad (1.7)$$

A representation for the generators t^A is provided by the eight Gell-Mann matrices, following the relation: $t^A = \frac{1}{2}\lambda^A$.

By convention, the normalisation of the SU(N) matrices is chosen to be:

$$\text{Tr}(t^A t^B) = T_R \delta^{AB}, \quad T_R = \frac{1}{2}. \quad (1.8)$$

Consequently, the colour matrices obey the following relations:

$$\sum_A t_{ab}^A t_{bc}^A = C_F \delta_{ac}, \quad C_F = \frac{N^2 - 1}{2N} \quad (1.9)$$

$$\text{Tr}(T^C T^D) = \sum_{A,B} f^{ABC} f^{ABD} = C_A \delta^{CD}, \quad C_A = N. \quad (1.10)$$

For the specific case of SU(3), the C_A and C_F factors become:

$$C_F = \frac{4}{3}, \quad C_A = 3. \quad (1.11)$$

For the anti-commutator of the fundamental representation t matrices, the following relations are determined:

$$\{t^A, t^B\} = \frac{1}{N} \delta^{AB} I + d^{ABC} t^C \quad (1.12)$$

$$\sum_{A,B} d^{ABC} d^{ABD} = \frac{N^2 - 1}{N} \delta^{CD}, \quad d^{AAC} = 0. \quad (1.13)$$

¹The colour charge has been introduced for satisfying the Fermi-Dirac statistics which imply the total anti-symmetry of the wave function in space, spin and SU(3)_f degrees of freedom.

The classical Lagrangian has the propriety of being invariant under local gauge transformations, which implies that it is possible to redefine the quark fields at every space and time point, without changing the physical content. The covariant derivative has this name since it transforms under a local gauge transformation in the same way as the field. The quark-mass term is invariant under local gauge transformation, while a mass term for the gluon cannot be included, in order to preserve the gauge invariance.

In addition to the gauge symmetry, the Lagrangian of equation 1.1 is invariant under other discrete global symmetries. These are the parity (P), charge conjugation (C) and time reversal transformation (T). Unlike weak interactions, there are no experimental evidences of CP violation in the strong sector, where CP stands for the combination of C- and P-symmetry.

The QCD Lagrangian also has several approximate symmetries, related to the quark masses. If the *up* and *down* quarks are considered with the same masses, the Lagrangian is symmetric under $U(1)_V \otimes SU(2)_V$ ². The $U(1)$ symmetry implies the conservation of the quark number, which is always valid. The $SU(2)_V$ is the approximate *isospin symmetry*, which can also be enlarged if considering the strange quark at the same mass as the other light quarks. This approximation brings to a $SU(3)_V$ flavour symmetry, which gives the classification of baryons and mesons into flavour octets and decuplets.

If we consider the mass of the *up* and *down* quarks negligible, the Lagrangian has in addition $SU(2)$ *chiral symmetry*, i.e., it acts separately on left and right handed fields. If this symmetry would be observed in nature, each hadron should have a pair with the same mass and opposite parity. The chiral $SU(2)$ is considered spontaneously broken: this means that there are massless spin-zero Goldstone bosons. They are identified in nature with pions: π^+ , π^- and π^0 . Since the symmetry is not exact, they are not exactly Goldstone bosons, but still their masses are smaller compared with those of the other hadrons.

The same argument can be developed, if considering the strange-quark as massless. In this case, the Goldstone bosons, arisen from the chiral $SU(3)$ spontaneously broken, are the flavour octet: $\pi^\pm$, π^0 , $K^\pm$, K^0 , $\bar{K}^0$ and η . The breaking symmetry is an important key for the Higgs boson theory formulation in the electroweak sector.

The second term of the Lagrangian 1.1 is the gauge-fixing term. In perturbation theory is not possible to define the propagator for the gluon field without making a gauge choice. The chosen term is covariant gauge and is given by:

$$\mathcal{L}_{gauge-fixing} = -\frac{1}{2\lambda}(\partial^\alpha A_\alpha^A)^2 \quad (1.14)$$

where λ is the gauge parameter.

In order to cancel the unphysical degree of freedom, another term is added. The term is given by ghost fields, which obey at the Fermi statistic and enter in the third term of the QCD Lagrangian 1.1,:

$$\mathcal{L}_{ghost} = \partial_\alpha \eta^{A\dagger} (D_{AB}^\alpha \eta^B). \quad (1.15)$$

Starting from equations 1.2, 1.14 and 1.15 is possible to derive the Feynman rules of the theory. An overview of these rules is shown in Table 1.1.

²The V letter is a reminder that the symmetry is associated to a vector current.

$A, \alpha \text{ } \text{-----} \text{ } B, \beta$ 	$\delta^{AB} [-g^{\alpha\beta} + (1 - \lambda) \frac{p^\alpha p^\beta}{p^2 + i\epsilon}] \frac{i}{p^2 + i\epsilon}$
$a, i \text{ } \text{---} \text{ } b, j$ 	$\delta^{AB} \frac{i}{p^2 + i\epsilon}$
$a, i \text{ } \cdots \text{ } b, j$ 	$\delta^{AB} \frac{i}{p^2 - m^2 + i\epsilon}$
B, β 	$-gf^{ABC} [(p - q)^\gamma g^{\alpha\beta} + (q - r)^\alpha g^{\beta\gamma} + (r - p)^\beta g^{\gamma\alpha}]$
A, α B, β 	$-ig^2 f^{XAC} f^{XBD} [g^{\alpha\beta} g^{\gamma\delta} - g^{\alpha\delta} g^{\beta\gamma}]$ $-ig^2 f^{XAD} f^{XBC} [g^{\alpha\beta} g^{\gamma\delta} - g^{\alpha\gamma} g^{\beta\delta}]$ $-ig^2 f^{XAB} f^{XCD} [g^{\alpha\gamma} g^{\beta\delta} - g^{\alpha\delta} g^{\beta\gamma}]$
A, α 	$gf^{ABC} q^\alpha$
A, α 	$-ig(t^A)_{cb} (\gamma^\alpha)_{ij}$

TABLE 1.1: Feynman rules for QCD, in curly lines gluons, in solid lines fermions and in dotted lines ghosts.

1.2.2 Asymptotic freedom and confinement

As mentioned before, the QCD is a perturbative theory, i.e., the observables can be written in series of the strong coupling $\alpha_s = g^2/4\pi$. The coupling is not constant, but it is a function of the energy Q , therefore it is defined as a *running coupling* $\alpha(Q^2)$.

As an example of observable, a generic dimensionless physical observable R is considered, which depends on the energy Q . If R is expanded in perturbation series of α_s , a second mass scale μ is introduced by the renormalisation process in order to absorb the ultraviolet divergence. The mass scale corresponds to the energy at which the divergence occurs. Consequently also the renormalised coupling α_s becomes dependent to μ . Since R is dimensionless, it depends on the ratio Q^2/μ^2 and the renormalised coupling α_s , following the equation:

$$\mu^2 \frac{d}{d\mu^2} R(Q^2/\mu^2, \alpha_s) = [\mu^2 \frac{\partial}{\partial \mu^2} + \mu^2 \frac{\partial \alpha_s}{\partial \mu^2} \frac{\partial}{\partial \alpha_s}] R = 0 \quad (1.16)$$

In order to write the equation in a more compact way, it is possible to introduce $t = \ln(Q^2/\mu^2)$ and the β function $\beta(\alpha_s) = \mu^2 \frac{\partial \alpha_s}{\partial \mu^2}$. The equation 1.16 is solved by introducing

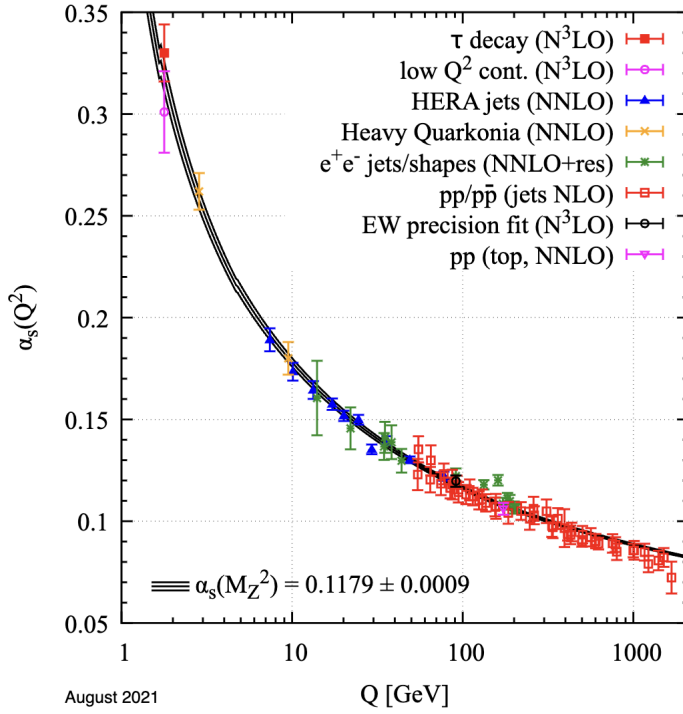


FIGURE 1.2: Summary of measurements of α_s as function of the energy scale Q [5].

the running coupling $\alpha_s(Q^2)$:

$$t = \int_{\alpha_s}^{\alpha_s(Q^2)} \frac{dx}{\beta(x)}, \quad \alpha_s(\mu^2) = \alpha_s \quad (1.17)$$

or by differential:

$$\frac{\partial \alpha_s(Q^2)}{\partial t} = \beta(\alpha_s(Q^2)), \quad \frac{\partial \alpha_s(Q^2)}{\partial \alpha_s} = \frac{\beta(\alpha_s(Q^2))}{\beta(\alpha_s)}. \quad (1.18)$$

In QCD the β function has the perturbative expansion in series of α_s . The coefficients are extracted from the high order loops correction to the vertices of the theory. Using the series expansion, it is possible to rewrite the relation between the running coupling and the β function in this way:

$$Q^2 \frac{\partial \alpha_s(Q^2)}{\partial Q^2} = -b \alpha_s^2(Q^2) [1 + b' \alpha_s(Q^2) + O(\alpha_s^2(Q^2))], \quad (1.19)$$

where $b = \frac{(33-2n_f)}{12\pi}$ and $b' = \frac{(153-19n_f)}{2\pi(33-2n_f)}$, with n_f the number of active light flavours. In perturbative regime, i.e., α_s is small, it is possible to solve the equation at LO and explicitly obtain the relation between $\alpha_s(Q^2)$ and $\alpha_s(\mu^2)$:

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu^2)}{1 + \alpha_s(\mu^2) b t}, \quad t = \ln(Q^2/\mu^2) \quad (1.20)$$

If the energy Q or, similarly, t increases, the running coupling $\alpha_s(Q^2)$ decreases. The running coupling behaviour as a function of Q^2 is shown in Figure 1.2, where the most

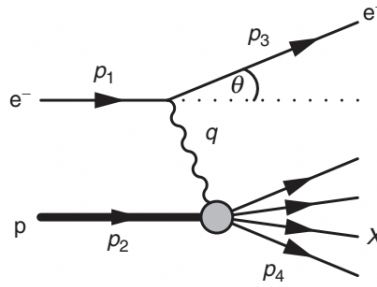


FIGURE 1.3: Electron-proton inelastic scattering[6].

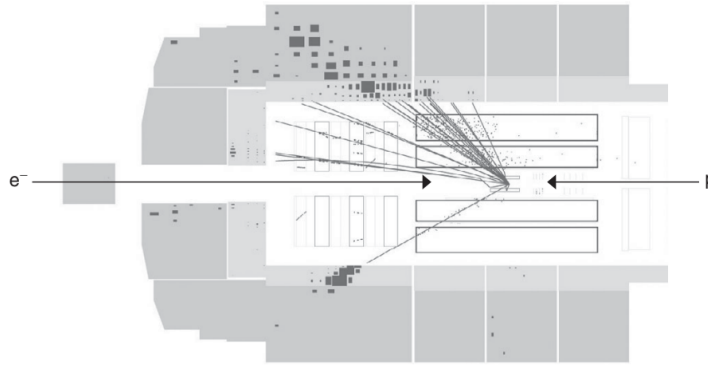


FIGURE 1.4: A electron-proton deep inelastic collision of H1 experiment at HERA. In figure the electron is scattered at large angle and the proton is broken up, producing a jet of particles [6].

recent measurements are summarised. The property of decreasing at high values of t is called *asymptotic freedom*. This means that at high energies or small distances, the force is weak and particles behave almost as free objects. This can be explained qualitatively with the anti-screening effect of the virtual gluons, which enforces the colour of the system at higher distances. The energy scale at which the coupling becomes big, which marks the transition from perturbative and non-perturbative regime, is called Λ parameter, also called QCD perturbative border.

Together with the asymptotic freedom, another important concept in QCD is the *confinement*: the feature of quarks and gluons of not being observed as free particles at macroscopic distances. As two colour charges are separated, at some point it is favoured for them to produce a couple of quark-anti-quark or, for quarks and gluons, to clump together in hadrons. This is the reason why, when quarks are produced in particle accelerators, they are observed as *jets* or showers of many colour-neutral particles. The *hadronization* process occurs in a time scale of $1/\Lambda$, i.e., when mesons or baryons are formed. The only exception is provided by the top quark, which decays before having time to hadronize.

1.2.3 Parton model and PDFs

In this section, a short overview, based on this [6] reference, is given, starting from the deep inelastic scattering (DIS) and the parton model, arriving to the parton distribution functions (PDFs) definition. The PDF concept is important in order to understand the motivation of the analysis presented in Chapter 5.

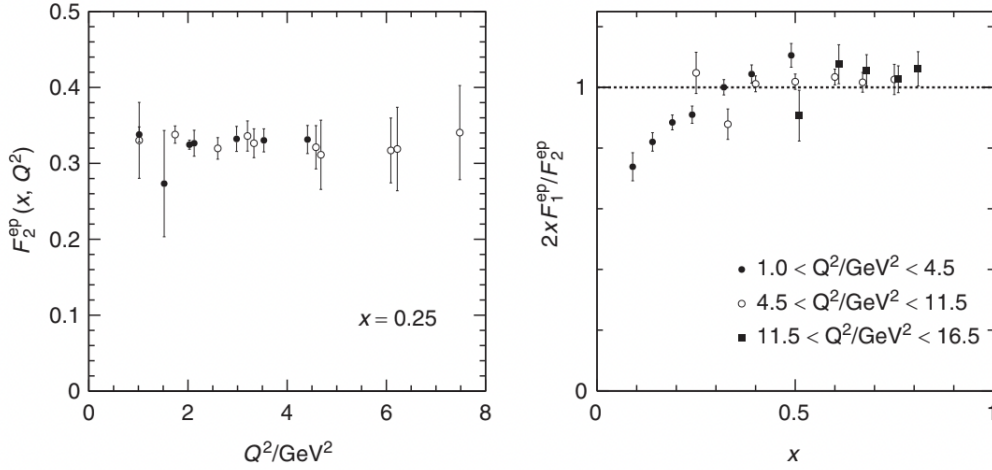


FIGURE 1.5: Structure function measurements from fixed target electron-proton inelastic scattering at SLAC. On left the $F_2(x, Q^2)$ function, on right the measurement of $2xF_1/F_2$ [6].

Electron-proton DIS

The electron-proton deep inelastic scattering interaction occurs when the energy is high enough for the proton to break up and to produce many particles in the final state. An example of the process is shown in Figure 1.3. Studies of deep inelastic scattering have been performed at the HERA proton-electron collider, which was operative at the DESY laboratory from 1991 to 2007. In this collider, electrons (or positrons) of 27.5 GeV collided with protons of 820/920 GeV, running in a ring of 3 km. H1 and ZEUS were the two detectors which collected inelastic collision data from HERA at very high energies. An event display of a deep-inelastic event recorded by the H1 detector is shown in Figure 1.4. The results obtained by H1 and ZEUS experiments provide precise determination of the proton structure functions, which are explained in this section.

In order to study the inelastic scattering, it is useful to define some kinematic variables. Q^2 is defined as the negative four momentum of the virtual photon:

$$Q^2 = -q^2 \quad (1.21)$$

which at very high energies, when the electron mass can be neglected, is always positive. The Bjorken x is a dimensionless and Lorentz-invariant parameter, defined as:

$$x = \frac{Q^2}{2p_2 \cdot q}, \quad (1.22)$$

where p_2 is the momentum of the incoming proton. The Bjorken x varies between 0 and 1 and suggests how much a scattering is elastic: the limit $x=1$ corresponds to a totally elastic scattering.

A second dimensionless Lorentz-invariant variable is the inelasticity y , defined with respect to the incoming electron and proton momentum p_2 and p_1 respectively and the virtual photon, as:

$$y = \frac{p_2 \cdot q}{p_2 \cdot p_1}. \quad (1.23)$$

It can be interpreted as the fraction of energy lost by the electron during the interaction.

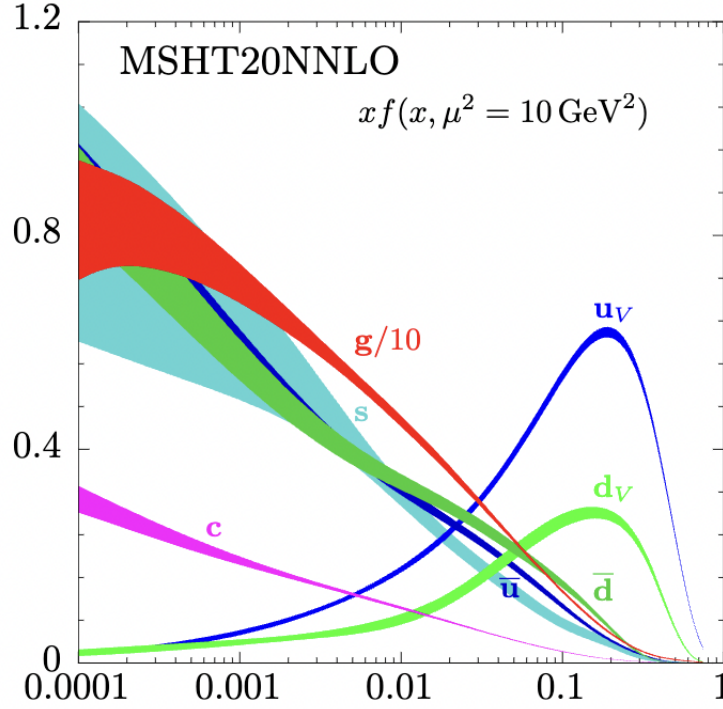


FIGURE 1.6: Distribution of $xf(x)$ parton distribution obtained in the NNLO MSHT20 global analysis [7].

It can be demonstrated [6], starting from the Rosenbluth formula of the elastic scattering, that the differential cross section for deep inelastic electron-proton scattering ($Q^2 \gg m_p^2 y^2$) is:

$$\frac{d^2\sigma}{dx dQ^2} = \frac{4\pi\alpha^2}{Q^4} \left[(1-y) \frac{F_2(x, Q^2)}{x} + y^2 F_1(x, Q^2) \right]. \quad (1.24)$$

In equation 1.24, α is the electromagnetic coupling (the analogous of α_s but for the electromagnetic interactions) and $F_1(x, Q^2)$ and $F_2(x, Q^2)$ are the proton *structure functions*. They indicate the interaction between the photon and the proton, where $F_1(x, Q^2)$ is associated to the magnetic interaction, while $F_2(x, Q^2)$ has both magnetic and electric contributions. The structure functions are not calculable in perturbative QCD, so they have to be experimentally measured.

The structure functions are related to the *parton distribution functions*, which have been introduced in the *quark-parton model*. The model was proposed by Richard Feynman and draws the hadrons as composed by point-like particles, the partons. In the quark-parton model, the deep inelastic scattering between electron and proton can be seen as the sum of the elastic contributions of electron-quark scattering. The differential cross section for the electron-quark scattering, where the quark carries a fraction x of the momentum of the proton is:

$$\frac{d\sigma}{dQ^2} = \frac{4\pi\alpha^2 Q_q^2}{Q^4} \left[(1-y) + \frac{y^2}{2} \right] \quad (1.25)$$

where Q_q is the quark charge.

Parton Distribution Functions

Quarks inside the proton are allowed to interact each other by exchanging gluons. Quarks are then described through a distribution of their momenta inside the protons. The PDFs are the parton-distribution functions $f(x)$, defined in order to have that $f_i(x)\delta x$ represents the number of quarks of flavour i in the proton with momentum fraction between x and $x + \delta x$.

The differential cross section in 1.25, summed over all the possible flavours in the proton and derived for the momentum fraction x , becomes:

$$\frac{d^2\sigma}{dx dQ^2} = \frac{4\pi\alpha^2}{Q^4} \left[(1-y) + \frac{y^2}{2} \right] \sum_i Q_i^2 f_i(x). \quad (1.26)$$

Comparing 1.26 to 1.24, the general expression for the deep inelastic scattering in terms of structure functions, we obtain:

$$F_2(x, Q^2) = 2xF_1(x, Q^2) = x \sum_i Q_i^2 f_i(x). \quad (1.27)$$

The first relation is called *Callan-Gross relation*, which leads to the observation that the structure functions are not independent. The second relation is called *Bjorken scale*, which suggest that, at least in first approximation, the structure function are independent from the energy scale Q . The first experimental evidence of the relations is shown in Figure 1.5.

Proton constituents are not only quarks but also gluons, that can generate quark-antiquark pairs. It is possible to split the contribution of quarks in *valence* quarks and the contribution from the *sea* of quarks pair-produced from the virtual gluons. Since gluons with large momenta are suppressed by the gluon propagator as $1/q^2$, the sea of virtual quarks and antiquarks are most probably produced at low values of x . An example of the most recent (2022) MSHT20 set of PDFs at NLO in α_s are reported in Figure 1.6, obtained by global analyses of all available hard scattering dataset, namely the final datasets from HERA and Tevatron.

DGLAP equation

In the previous section, we have seen that the structure functions or the PDFs, at first approximation, follow the Bjorken scaling, i.e., they are independent from the energy scale. Considering higher orders in perturbative QCD, a scaling violation for the structure functions is expected, they are observed to have Q^2 -dependence.

Regarding PDFs, it is not possible to calculate the distributions from first principles, but the energy dependence is calculable thanks to parton evolution equations known as Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations.

DGLAP equation at leading order has the following structure:

$$\mu_F^2 \frac{\partial f_i(x, \mu_F^2)}{\partial \mu_F^2} = \sum_j \frac{\alpha_s(\mu_F^2)}{2\pi} \int_x^1 \frac{dz}{z} P_{ij}^{(1)}(z) f_j(x/z, \mu_F^2) \quad (1.28)$$

where z is the ratio of the parton momentum between before- and after-splitting emission, μ_F is the *factorization scale* (usually chosen $\mu_F = Q$) and $P_{ij}^{(1)}(z)$ are the splitting functions, which encode the splitting probability. The strength of this equation, which is proven also at higher QCD orders, is that PDFs measured at

one scale, can be perturbatively extrapolated at any other energy scales. The non-perturbative parton distribution functions are obtained from data using a parametric fit.

Hadron-hadron collisions

At hadron colliders, like LHC, the particles or jets in the final state are directly linked to the partonic scattering of the initial state. The cross section of the inelastic scattering between the two hadrons is the convolution of the partonic distribution functions and the hard scattering cross section:

$$\sigma(P_1, P_2) = \sum_{ij} \int dx_1 dx_2 f_i(x_1, \mu^2) f_j(x_2, \mu^2) \hat{\sigma}_{ij}(p_1, p_2, \alpha_s(\mu^2), Q^2/\mu^2) \quad (1.29)$$

The two partons p_1 and p_2 participate to the hard scattering process with a fraction x_1 and x_2 of the hadron momenta. The cross section is summed over the possible i, j possible parton distributions of quarks and gluons and integrated on the possible fraction momenta. Q is the characteristic scale of the hard scattering and could be the mass of the final state particle or jet.

The cross section σ_{ij} between partons i, j , since it occurs at very high energy interaction or, in terms of distances, at very short distance between the particles, is called the short distance cross section. For this reason, it is possible to calculate it as a perturbation series in the running coupling α_s . At leading order the short-distance cross section coincides with the total parton cross section, but at higher orders there are some long-distance factors. The *factorisation theorem* allows the parton distribution functions to absorb the distance contributions and to separate the short-distance part, calculable in perturbative theory, from the long-distance one. The factorisation scale μ corresponds to the scale which separates the two contributions.

1.3 Electroweak interactions

The electroweak theory is the unified description of the electromagnetic and weak interactions. At low energies, the two forces are separated, while at the unification energy of about 246 GeV they are combined in terms of a renormalizable, gauge invariant field theory.

The Lagrangian of the electroweak sector is divided in three parts:

$$\mathcal{L} = \mathcal{L}_g + \mathcal{L}_H + \mathcal{L}_f. \quad (1.30)$$

$\mathcal{L}_g$ represents the gauge field interactions, which is described in section 1.3.1. The second term, $\mathcal{L}_H$, is the Higgs boson interactions, briefly described in section 1.3.2. The last term, $\mathcal{L}_f$, includes the fermionic part, which describes the interaction between the fermion multiplets and the Higgs boson, presented in section 1.3.3.

The present EW brief description, as for the QCD part, follows the *QCD and Collider Physics* of R.K. Ellis, W.J. Stirling and B.R. Webber [4] by reference.

1.3.1 The Gauge field term

The Standard electroweak model is based on $SU(2) \times U(1)$ symmetry group. The Lagrangian includes three massless gauge bosons W^i ($i=1,2,3$) for the gauge group $SU(2)$ and a boson B for the gauge group $U(1)$. The Lagrangian of the gauge bosons

is similar to the QED Lagrangian in the field propagation part ($\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu}$ with $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ for the photon field A^μ):

$$\mathcal{L}_g = -\frac{1}{4}W^{i\mu\nu}W_{\mu\nu}^i - \frac{1}{4}B^{\mu\nu}B_{\mu\nu}, \quad (1.31)$$

$$W_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i - g_W \epsilon^{ijk} W_\mu^j W_\nu^k, \quad (1.32)$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \quad (1.33)$$

where g_W is the coupling of SU(2) group and ϵ^{ijk} are the antisymmetric structure constants. In 1.32, the vector bosons W^i have a self-interaction term, which comes from the non-Abelian nature of SU(2) symmetry. The gauge interaction with the fermionic matter fields is described by the covariant derivative:

$$(D^\mu)_{ij} = \delta_{ij}\partial^\mu + ig_W(T \cdot W^\mu)_{ij} + iY\delta_{ij}g'_W B^\mu \quad (1.34)$$

where g'_W is the U(1) gauge coupling. T are the representation matrices of SU(2), based on *weak isospin*, while Y is the charge of U(1), the *weak hypercharge*.

It is possible to re-define the vector bosons W_μ^i in two new vector bosons $W_\mu^\pm = (W_\mu^1 \mp iW_\mu^2)/\sqrt{2}$, and likewise the T matrices as $T^\pm = T^1 \pm T^2$, where T^+ and T^- now becomes the rising and lowering operators of the weak isospin.

This model is not complete, since it does not correspond to the real world, where the gauge bosons have masses. However, adding a mass term would violate the local gauge symmetry. It is necessary, then, to introduce a mechanism of symmetry breaking, called *the Higgs mechanism*.

1.3.2 The Higgs term

In order to have mass terms for gauge bosons, as mentioned in the previous section, it is necessary to introduce a mechanism which breaks the SU(2)×U(1) symmetry, keeping the electric charge conserved.

Let us consider a new scalar field and its Lagrangian term, which is SU(2)×U(1) symmetric. The scalar Higgs field ϕ is a weak-isospin doublet which leaves invariant the U(1)_{em} symmetry, defined as:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}. \quad (1.35)$$

The invariant Lagrangian under SU(2)×U(1) transformation for the Higgs part is:

$$\mathcal{L}_H = (D_\mu \phi)^\dagger (D_\mu \phi) - V(\phi^\dagger \phi) \quad (1.36)$$

where V is the Higgs potential and is chosen to be:

$$V(\phi^\dagger \phi) = \lambda(\phi^\dagger \phi)^2 - \mu^2 \phi^\dagger \phi, \quad \mu^2, \lambda > 0 \quad (1.37)$$

The potential shows a set of non trivial local minima for:

$$|\phi|^2 = \frac{v^2(\mu^2, \lambda)}{2} \neq 0. \quad (1.38)$$

	$+ig_v[(p-q)_\lambda g_{\mu\nu} + (q-r)_\mu g_{\nu\lambda} + (r-p)_\nu g_{\lambda\mu}]$
	$+ig_{VH}M_W g_{\mu\nu}$
	$+ig_W^2[2g_{\mu\nu}g_{\lambda\rho} - g_{\mu\lambda}g_{\nu\rho} - g_{\mu\rho}g_{\nu\lambda}]$
	$-ig_W^2 \cos^2 \theta_W [2g_{\mu\nu}g_{\lambda\rho} - g_{\mu\lambda}g_{\nu\rho} - g_{\mu\rho}g_{\nu\lambda}]$
	$-ie^2[2g_{\mu\nu}g_{\lambda\rho} - g_{\mu\lambda}g_{\nu\rho} - g_{\mu\rho}g_{\nu\lambda}]$
	$-ieg_W \cos \theta_W [2g_{\mu\nu}g_{\lambda\rho} - g_{\mu\lambda}g_{\nu\rho} - g_{\mu\rho}g_{\nu\lambda}]$

TABLE 1.2: Feynman rules for boson interactions.

It is possible to choose a vector $\langle\phi\rangle$, corresponding to the vacuum expectation value, such that it is annihilated by the electric charge operator $Q = T^3 + Y$:

$$(T^3 + Y)\langle\phi\rangle = 0. \quad (1.39)$$

These conditions bring to a unique solution for the $\langle\phi\rangle$ vector:

$$\langle\phi\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \quad (1.40)$$

A general Higgs field parametrization is obtained considering a fluctuation around the minimum:

$$\phi = U^{-1}(\xi) \begin{pmatrix} 0 \\ (H+v)/\sqrt{2} \end{pmatrix} = U^{-1}(H+v)/\sqrt{2} \chi \quad (1.41)$$

where $\chi = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ is a unit vector along the direction of vacuum expectation and the matrix $U(\xi) = \exp(-iT \cdot \xi/v)$. The Lagrangian in 1.36 becomes:

$$\begin{aligned} \mathcal{L}_H = & \frac{\partial_\mu H \partial^\mu H}{2} - V\left(\frac{(v+H)^2}{2}\right) \\ & + \frac{(v+H)^2}{8} \chi^\dagger (2g_W T \cdot W_\mu + g'_W B_\mu) (2g_W T \cdot W^\mu + g'_W B^\mu) \chi \end{aligned} \quad (1.42)$$

This Lagrangian breaks spontaneously the $SU(2) \times U(1)$ symmetry, because it keeps the group symmetry, while the vacuum does not.

If we examine only the vector boson field term, we notice that the propagator for B and W_μ^3 fields is not diagonal. So it is possible to redefine the fields in function of the Weinberg or electroweak mixing angle θ_W and of the two physical bosonic fields, Z_μ and A_μ :

$$\begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix}. \quad (1.43)$$

The mixing angle is fixed by the coupling constants: $\sin \theta_W^2 = \frac{g_W'^2}{g_W^2 + g_W'^2} \simeq 0.23$. With this 1.43 redefinition, the vector boson contribution of the Lagrangian becomes:

$$\mathcal{L}_{H,VB} = \frac{g_W^2 v^2}{4} W_\mu^+ W^{-\mu} + \frac{(g_W^2 + g_W'^2) v^2}{8} Z_\mu Z^\mu. \quad (1.44)$$

With this last Lagrangian term, the W and Z bosons acquire masses equal to $M_W = \frac{1}{2} v g_W$ and $M_Z = \frac{1}{2} v \sqrt{g_W^2 + g_W'^2} = \frac{M_W}{\cos \theta_W}$ respectively. The photon is still massless, since there are not quadratic terms of the field A . Experimentally, the vacuum expectation value of the Higgs field has been found to be 246 GeV.

The remaining term in the Lagrangian is the Higgs boson one, which includes the Higgs boson kinetic and self interaction terms:

$$\mathcal{L}_{H,Higgs} = \frac{1}{2} \partial_\mu H \partial^\mu H - \mu^2 H^2 - \lambda v H^3 - \frac{1}{4} \lambda H^4. \quad (1.45)$$

It describes the mass term of the Higgs boson of $M_H = \sqrt{2}\mu$, the kinetic field propagation and the Higgs tri- and four-linear couplings. While v can already be experimentally determined from the weak boson measurements, λ still needs to be measured

$\mu \sim\!\!\sim\!\!\sim A \nu$	$[-g^{\mu\nu} + \frac{q^\mu q^\nu}{q^2}] \frac{i}{q^2}$
$\mu \sim\!\!\sim\!\!\sim W, Z \nu$	$[-g^{\mu\nu} + \frac{q^\mu q^\nu}{M^2}] \frac{i}{(q^2 - M^2)}$
$\longrightarrow$ -----	$\frac{i}{\not{q} - m}$ $\frac{i}{q^2 - M_H^2}$
A_μ 	$-ieQ_f\gamma_\mu$
W_μ 	$\frac{-ig_W}{2\sqrt{2}}\gamma^\mu(1 - \gamma_5)(T^+)_{f'f}$
Z_μ 	$\frac{-ig_W}{2\cos\theta_W}\gamma^\mu(V_f - A_f\gamma_5)$
H 	$\frac{-ig_W m_f}{2M_W}$

TABLE 1.3: Propagators and Feynman rules for fermion interactions.

directly.

In Table 1.2 the Feynman rules for the $SU(2) \times U(1)$ boson interactions, as obtained from the Lagrangian, are shown.

1.3.3 The fermion term

The Lagrangian fermion term is defined by the fermion interactions with the gauge bosons. This term is expected by the gauge symmetry, since the interaction is described by the covariant derivative, defined in section 1.3.1. If the right- and left-handed components of the fermions are made explicit, the Lagrangian term, before the symmetry breaking, is given by:

$$\mathcal{L}_f = \bar{\psi}_R i(\not{\partial} + ig'_W Y_R \mathcal{B})\psi_R + \bar{\psi}_L i(\not{\partial} + ig_W \cdot T\mathcal{W} + ig'_W Y_L \mathcal{B})\psi_L. \quad (1.46)$$

Since the electroweak theory is not symmetric respect the parity, i.e., it is a chiral theory, the left and right-handed fields transform with different representation of the symmetry group $SU(2) \times U(1)$. The left-handed fields $\psi_L = \gamma_L \psi$, where $\gamma_L = \frac{1}{2}(1 - \gamma_5)$, are organized into doublets of the fundamental representation of the weak isospin group $SU(2)$. The right-handed fields, instead, follow the one dimensional

representation of SU(2) $\psi_R = \gamma_R \psi$, where $\gamma_R = \frac{1}{2}(1 + \gamma_5)$. Consequently the g_W term in the right-handed part is not present in 1.46. The hypercharges of the fermions are fixed in order to satisfy the relation $Q = T^3 + Y$.

Fermions fall in three families or *generations*, each made up of a left-handed doublet of lepton and quark fields:

$$\psi_L = \gamma_L \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}, \gamma_L \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}, \gamma_L \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix} \quad (1.47)$$

$$\psi_L = \gamma_L \begin{pmatrix} u \\ d' \end{pmatrix}, \gamma_L \begin{pmatrix} c \\ s' \end{pmatrix}, \gamma_L \begin{pmatrix} t \\ b' \end{pmatrix} \quad (1.48)$$

plus all the corresponding one dimensional right-handed fields.

The primed down-type quarks which interact with the up-type quarks are a superposition of the down, strange and bottom quarks, by a unitary transformation and they are the gauge auto-states. The unitary matrix V responsible of this transformation in the quark sector is known as Cabibbo-Kobayashi-Maskawa (CKM) matrix:

$$V = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \quad (1.49)$$

All values of the CKM matrix elements, obtained from experiments, can be found here [5].

After the spontaneous symmetry breaking, the Higgs field interacts with the fermions in the form $g_f \bar{\psi}_f \psi_f \phi$, called the *Yukawa* interaction term. Consequently, fermions have a mass of $m_f = g_f v / \sqrt{2}$ and $\bar{\psi}_f \psi_f H$ interactions of strength $g_f = \sqrt{2} m_f / v$.

The Lagrangian 1.46 becomes:

$$\begin{aligned} \mathcal{L}_f = & \sum_f \bar{\psi}_f \left(i \not{\partial} - m_f - g_W \frac{m_f H}{2M_W} \right) \\ & - \frac{g_W}{2\sqrt{2}} \sum_f \bar{\psi}_f (\gamma^\mu (1 - \gamma_5) T^+ W_\mu^+ + \gamma^\mu (1 - \gamma_5 T^- W_\mu^-)) \psi_f \\ & - e \sum_f q_f \bar{\psi}_f \not{A} \psi_f - \frac{g_W}{2 \cos \theta_W} \sum_f \bar{\psi}_f \gamma^\mu (V_f - A_f \gamma_5) \psi_f Z_\mu \end{aligned} \quad (1.50)$$

where the couplings with the Z boson V_f and A_f are defined as $V_f = T_f^3 - 2Q_f \sin^2 \theta_W$ and $A_f = T_f^3$.

From the Lagrangian 1.50 it is possible to obtain the Feynman rules for the fermions, which are shown in Table 1.3. It is evident from this last interaction terms, that fermions with different flavours can interact each other and this is a peculiarity of the weak interaction.

Chapter 2

Physics at the ATLAS experiment

2.1 Introduction

The Large Hadron Collider (LHC), situated at CERN between Switzerland and France, is the most powerful and largest particle accelerator in the world. LHC consists in a 27 kilometer ring, equipped with both superconducting magnets and accelerating structures in order to bend and boost the particles up to high energies.

Bunches of up to 10^{11} protons (p) collide in the ring, designed to reach 14 TeV of center of mass energy at the instantaneous luminosity of a few $10^{34} \text{ cm}^{-2}\text{s}^{-1}$.

In the interaction points of LHC, different detectors have been enrolled, among which ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid), for exploiting p - p and ion-ion collisions in the high energy regime.

In this chapter, some ingredients of the physics at colliders are provided in Section 2.2. Consequently, an overview of the ATLAS detector is presented, deepening on the Muon Spectrometer, framework of my PhD work in the Data Quality (DQ) Online Monitoring. The DQ monitoring is the subject of the next chapter 3.

Section 2.4 shows briefly the processes of triggering and data acquisition. In the last Section 2.5, the reconstruction and identification of physics objects are shortly explained.

2.2 Physics at colliders

A particle accelerator is a machine which produces high energy particle beams, bending their trajectory with electromagnetic fields. Particle accelerators are usually of two types: collider accelerators and the fixed target accelerators. In the first case, two beams of particles are accelerated at a chosen energy and they collide one a more interaction points, where the detectors are installed for measuring the products of the collisions. In the latter case, instead, there is only one particle beam accelerated towards a fixed target.

In this thesis, I will refer only to the first case, since my studies are performed on data collected by the Large Hadron Collider (LHC) at CERN, but fixed target experiments are indirectly referred as they are involved in PDF measurements.

The detector system is built in order to measure each particle interaction result, in energy and momentum. In jargon, each interaction is usually called *event*. The primary particle interactions are measured by the detectors, exploiting different technologies.

In general, a typical collider detector consists in a cylinder, rolling around the beam axis. In the innermost region, close to the interaction point, a tracking system is placed, to measure track momenta. It is surrounded by the electromagnetic (ECAL) and hadronic calorimeter (HCAL) systems, to measure particle energies. The ECAL

can detect electrons and photons, while the HCAL is tailored for detecting and measuring the energy of hadrons outside the ECAL. The most external part of the detector is projected to detect high energy muons, which survive after the passage into the HCAL.

Particles can be identified using a combination of track and calorimeter information, to measure momentum and energy. Neutrinos, instead, are not directly detected, since they interact little with matter and leave no signal in the collider. Their presence can be derived from the *missing momentum*, defined as:

$$\mathbf{p}_{mis} = - \sum_i \mathbf{p}_i \quad (2.1)$$

where the sum is meant over all the observed particles in an event.

As explained in the previous chapter, due to the QCD nature, quarks are never observed as free particle. They are usually bond together with gluons forming hadrons. The hadrons formed by the quark-gluon interaction quickly decay in lighter particles, producing the hadronic showers. The result of this process is the production of several energy deposits in the calorimeters. In particle physics they are referred as *jets* of particles, as further described in section 2.2.3.

2.2.1 Center-Of-Mass Energy, Cross Section and Luminosity

One of the purposes of the modern colliders is to detect massive particles, as the Higgs boson and gauge bosons (W, Z). In order to obtain this result, colliders must reach high center-of-mass energy $\sqrt{s}$. This is the square root of the Lorentz invariant quantity s defined by the two particles in the initial state ¹:

$$s = \left(\sum_i E_i \right)^2 - \left(\sum_i \mathbf{p}_i \right)^2 \quad (2.2)$$

The two important parameters for a collider are the center-of-mass energy and the instantaneous luminosity L . If the center-of-mass energy determine the type of particles that can be studied, the L is related to the discovery of rare events. The instantaneous luminosity, integrated over the time machine, is directly proportional to the number of events collected following this equation:

$$N = \sigma \int L dt \quad (2.3)$$

where σ is the interaction *cross section*. The cross section is the quantum probability for a certain interaction and it is directly related to the amplitude of the process involved. The cross section has the dimension of an area and the usual unit is the barn b ($1 b = 10^{-24} \text{cm}^2$), while the instantaneous luminosity has the inverted unit of the cross section b^{-1} .

At colliders, particles travel in bunches, which are groups of particles with the same momentum. In the case of LHC, the bunches are separated by 25 ns, corresponding to a collision frequency f of 40 MHz. If n_1 and n_2 are the number of particles of the two beams and, assuming that the beams have a Gaussian profile, with σ_x and σ_y the horizontal and transverse beam sizes, the instantaneous luminosity is defined as:

$$L = f \frac{n_1 n_2}{4\pi\sigma_x\sigma_y}. \quad (2.4)$$

¹The definition is done in natural units, with speed velocity $c=1$.

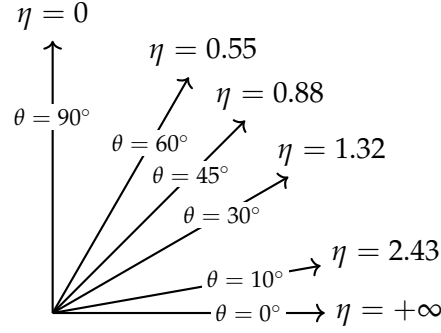


FIGURE 2.1: The schematic representation of the pseudorapidity, if the polar angle goes to zero, η tends to infinity.

Events collected by the detector are the input parameters to infer cross sections of physically interesting processes. For this reason, the luminosity has to be measured with high precision. At LHC usually there are specific runs, called Van Der Meer Scan [8], designed to have the most precise luminosity measurement as possible. In this procedure, the location of the beams are varying one respect each other and the rate is observed changing. The rates are calculated for well known processes, namely *standard candles*, from which the instantaneous luminosity of the accelerator is obtained.

ATLAS has several detectors sensitive to the number of particles produced in the proton interactions, which are used to measure the instantaneous and integrated luminosity. One of the main sub-detector is LUCID-2 [9] (Luminosity Cherenkov Integrating Detector) located at 17m from the interaction point, consisting in two sets of photomultiplier tubes (PMT), which detect the Cherenkov light emitted by charged particles coming from the p - p collisions. The Inner Detector and the Calorimeter are also used to measure the luminosity in a complementary way: counting the number of charged particle tracks and monitoring the energies deposited, respectively.

2.2.2 Kinematic variables

When protons collide, partons are those which actually interact and their longitudinal momenta is a spectrum determined by the parton distribution functions. Since the center of mass of the partons is boosted with respect the hadrons, it is useful to define some kinematic variables which transform simply under longitudinal boosts. In the center of mass of the two protons, the transverse momentum is zero, and, for the conservation of the momentum, it has to be zero also after the interaction. It is hence useful to define kinematic variables related to the transverse projection.

The variables are the rapidity y , the transverse momentum p_T and the azimuthal angle ϕ .

The four-momentum of a particle of mass m can be written as:

$$\begin{aligned} p^\mu &= (E, p_x, p_y, p_z) \\ &= (m_T \cosh y, p_T \sin \phi, p_T \cos \phi, m_T \sinh y) \end{aligned} \quad (2.5)$$

giving $|p| = p_T \cosh y$, where the transverse mass m_T is defined as $m_T = \sqrt{p_T^2 + m^2}$ and the rapidity y as

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \quad (2.6)$$

where p_z is the momentum along the beam axis. The rapidity difference is Lorentz invariant. When the mass is negligible, the rapidity can be approximated by the *pseudorapidity* η :

$$\eta = -\ln \tan \theta/2. \quad (2.7)$$

The pseudorapidity depends on the angle θ relative to the beam axis, which is directly measured by detectors.

In Figure 2.1, the schematic representation of the pseudorapidity values for a typical LHC detector are shown.

Following the transverse momentum definition, the transverse energy E_T can be also defined as:

$$E_T = E \sin \theta \quad (2.8)$$

The pseudorapidity is used to define the angular separation between particles

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}, \quad (2.9)$$

which is Lorentz invariant under a longitudinal boost. Sometimes, also for quasi-massless particles, the rapidity is replaced by the pseudorapidity. This variable is used in the jet definition, as explained in the next Section 2.2.3.

2.2.3 Jets and event shape variables

As mentioned before, in the accelerator detectors, quarks are not revealed as single particles, but they are observed as *jets*, i.e., a spread of colour-neutral particles, as hadrons, clustered together.

Since jets are collections of particles, it is necessary to have a *jet definition*, in order to cluster together the correct particles. First of all, a good jet definition has to be *infrared and collinear (IRC) safe*. In a nutshell, IRC safety is the property of being insensitive to the emission of soft or collinear gluons, i.e., gluons of low energy or very close to the particle that emits. In other words, if $\mathbf{p}_i$ is the momentum of the particle, it must be invariant under the branching: $\mathbf{p}_i \rightarrow \mathbf{p}_j + \mathbf{p}_k$. This happens if $\mathbf{p}_j$ and $\mathbf{p}_k$ are parallel or one of the two is soft. The lack of IRC safety can cause divergences in the observable theoretical definition, which are not experimentally observed resulting in complications in theory/experiment comparison.

There are two classes of jet finding algorithms: the *cone algorithm* and the *sequential recombination algorithms*. The first one is based on finding a cone of radius R of particles around a seed particle i . The direction of the jet is given by the sum of the momenta of all particles. The iteration is carried on until the direction of the cone is stable and it is called a jet if overcome a certain momentum threshold. Even if they are easy to implement, seed algorithms are not generally compatible with QCD calculations at higher orders. They are therefore deprecated. They were used at older hadron colliders, like at LEP e^+e^- collider where the final jet states are not contaminated by other QCD processes ??, and still use at Fermilab Tevatron.

Sequential recombination algorithms are instead used at LHC and they are based on a distance definition between pairs of particles i and j :

$$d_{ij} = \min(k_{t,i}^2, k_{t,j}^2) \Delta_{ij}^2 / R^2, \quad (2.10)$$

where Δ_{ij} is the particle separation in function of rapidity and azimuthal angle, $k_{t,i}$ the transverse momentum respect the incoming beams and R is a free parameter. For each i particle the d_{ij} is calculated respect another j particle and evaluated in

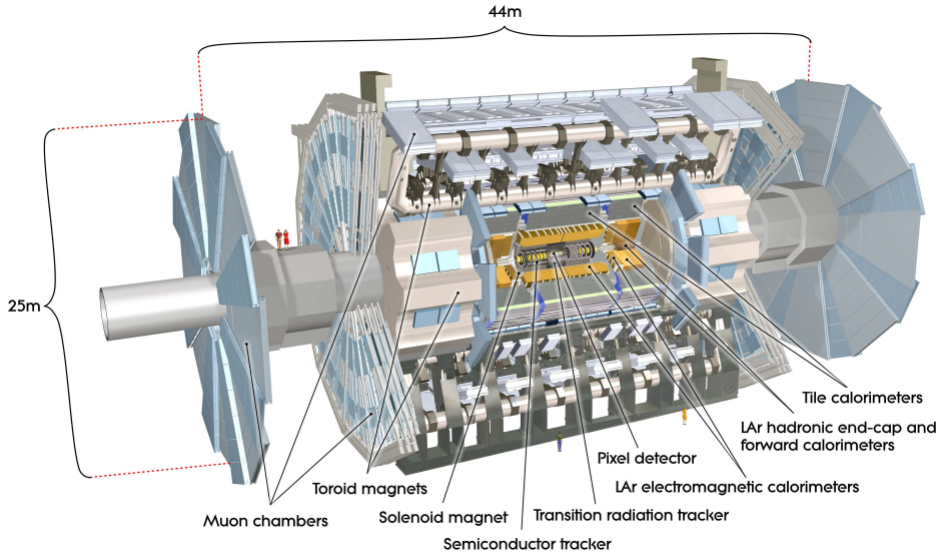


FIGURE 2.2: Cut-view of the ATLAS detector, 25 m high, 44 m long and approximately of 7000 tonnes in weight [16].

comparison with $d_{iB} = k_{t,i}^{2p}$ the beam distance. If the beam distance d_{iB} is smaller than d_{ij} , the i particle is defined as a jet and removed from the list of particles. If the d_{ij} distance is then smaller, the two particles are merged into a new pseudo-particle. The iteration goes on with the new pseudo-particle until the beam distance becomes the smallest. It is considered a jet if satisfies a given threshold.

The p parameter determines different kind of algorithms: $p = 1$ corresponds to the k_t algorithm [10, 11, 12], which prefers to cluster soft particles and it does a good job at resolving sub-jets, $p = 0$ to the *Cambridge – Aachen* algorithm [13, 14], which has a purely spatial character and it has the best de-cluster features, and $p = -1$ to the *anti- k_t* algorithm [15]. The last one is very good at resolving jets and it is actually the standard algorithm at the LHC experiments.

To classify the shape of the final state particles in a jets, variables can be defined, which characterize the topology of an event. They are *event shape* variables. They are usually dependent on the four-momenta of the particles in the final state and they are hence sensitive to the QCD radiation. Some variables are obtained by the jet substructure, i.e., sub-jets. They are sensitive to the QCD radiation and have been recently used to study the properties of quarks and gluons inside the jets. Another use of the jet substructures is for distinguishing QCD jets from those coming from boosted massive particles, as Higgs boson. Some event shape variables connected with jet substructures will be investigated in Chapter 4.

2.3 The ATLAS detector

The ATLAS experiment [16] is a large scientific collaboration born in 1992, composed by 5952 participants from 263 institutes in 47 different countries (data of 2023, here updated numbers [18]).

The ATLAS experiment has been designed in order to search for the Higgs boson and to favour new phenomena discovery. Furthermore, high values of luminosity and energy enable high precision tests on QCD, electroweak interactions and flavour

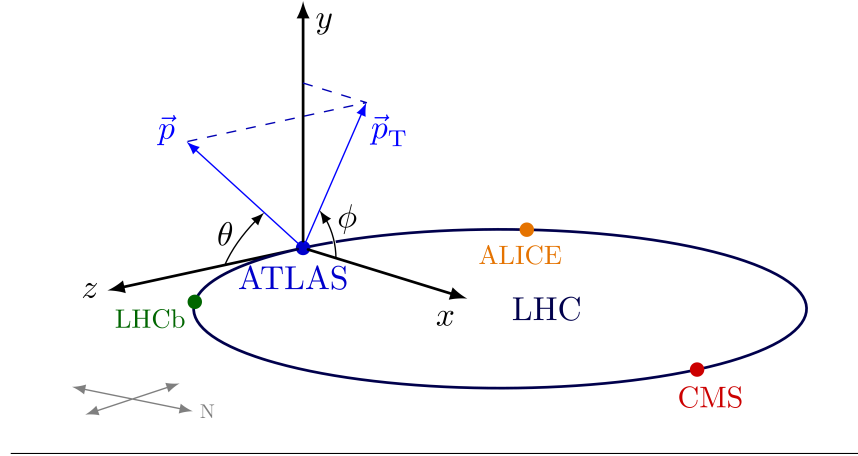


FIGURE 2.3: The global ATLAS coordinate system [17].

physics. It is the largest detector ever constructed for a particle collider: 44 m long and 25 m in diameter.

In Figure 2.2, a representation of the ATLAS detector is shown.

The magnetic field is provided by one thin superconducting solenoid placed around the inner-detector cavity and by three large superconducting toroids (one barrel and two end-caps) located around the calorimeters. The calorimeter system is surrounded by the muon spectrometer, which dominates the external dimensions of the entire detector. The selection and collection of data is supported by a trigger system, organised in a two level structure.

In this section, after a quick ATLAS coordinate system description in Section 2.3.1 a brief overview of all subsystems is presented: in Section 2.3.2 the inner detector, 2.3.3 the calorimetry and 2.3.4 the muon system.

2.3.1 The ATLAS detector coordinate system

The ATLAS global right-handed Cartesian coordinate system is shown in Figure 2.3. The z -axis is defined along the beam direction, with positive direction from Point 1 to Point 8, where ATLAS and LHCb are located respectively, while the x -axis points towards the center of the LHC ring. The y -axis is defined perpendicular to x - z plane directed upwards. The θ angle is defined respect the z -axis.

In cylindrical coordinates, the radius r is the distance from the beam line in the x - z plane and ϕ the angle between the radius to the x -axis.

The A side of the detector corresponds to positive z values, and the C side to the negative values.

For each detector unit is defined an independent coordinate system, where the x - z plane is parallel to the strip planes of the detector, with the z -axis along the primary coordinate and x -axis along the second coordinate. The origin is located in the middle strip on the closest to the interaction point edge.

2.3.2 The Inner detector

The Inner detector (ID) is the first detector in proximity of the interaction point. Since approximately 1000 particles are produced from the collisions every 25 ns within $|\eta| < 2.5$, it needs to be very sensitive in order to measure, with high resolution, the particle curvature and the vertex position. It consists of three sub-systems, which collaborate to measure the direction, momentum and charge of particles. The benchmark is

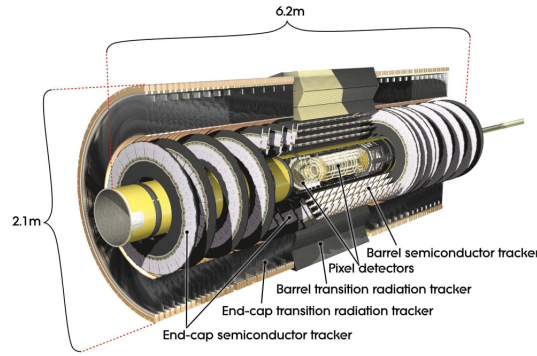


FIGURE 2.4: Cut-view of the ATLAS inner detector [19].

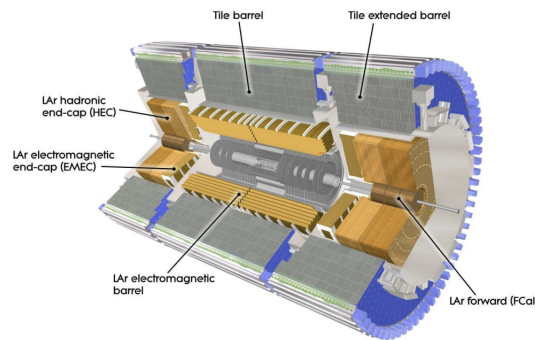


FIGURE 2.5: Cut-view of the ATLAS Calorimeters [19].

achieved with the interplay of two precision tracking detectors, pixel [20] and silicon microstrips (SCT) [21], together with the straw tubes for the measurement of the Transition Radiation Tracker (TRT) [22].

The layout of the ID is illustrated in Figure 2.4.

The ID has dimensions of 6.2 m in length and 2.1 m of diameter in high and it is immersed in a 2T magnetic field, generated by the central solenoid. The Pixel and SCT precision trackers cover the region of $|\eta| < 2.5$ and they are located in the barrel, arranged as concentric cylinders around the beam pipe, and in the end-cap regions, placed on disks perpendicular to the beam axis. The pixel provides the highest granularity, as they are located just 3.3 cm from the beam line. It consists of three layers of silicon pixels in the barrel and three disks in the end-cap regions. The pixel detector has approximately 80.4 million readout channels. It is used to determine the origin and momentum of the particles with a precision up to $10 \mu\text{m}$ in the $R\text{-}\phi$ plane. The SCT is the middle detector of the ID. It is also based on silicon detector technique, but the cell size is determined by the strip. This detector is designed to measure and detect the tracks of the charged particles.

The final and outer part of the ID is the TRT. It is made of thin drift tubes of 4 mm of diameter, which provides only the $R\text{-}\phi$ information up to $|\eta| = 2$. It is not as precise as the other two ID components, but gives information about the particle type.

2.3.3 Calorimeters

If Inner Detector measures the tracks and momentum of the particles, calorimeters measure their energies. The ATLAS calorimeter system consists of two different sampling calorimeters, optimized for electromagnetic and hadronic showers measurements, respectively. They both consist of layers of absorbing high density material that slow the incoming particles down followed by layers of active medium which actually measure their energy.

The layout of the Calorimeters is illustrated in Figure 2.5.

The Liquid Argon (LAr) Calorimeter [23] is the electromagnetic calorimeter, which surrounds the Inner Detector. It consists in layers of metal (tungsten, copper or lead depending on the region) which allows to generate the showers and reduces particle energy. Particles ionise in the liquid argon volume placed between accordion-shaped layers kept at -184°C . The current produced is conducted in a warmer area and measured.

The Tile Calorimeter [24] surrounds the LAr and it is the main part of ATLAS hadronic calorimeter. It is made of layers of steel and plastic scintillating tiles. The forward part of hadronic calorimeter is in the same EM cryostat. It is built in liquid argon and copper and tungsten as absorbers.

2.3.4 Muon System

The ATLAS Muon System [25] [26] provides track and momentum information about muon particles.

The muon spectrometer (MS) measures muon tracks deflected by the magnetic field using high-precision precision track chambers and a separated muon-trigger system is assigned to the event selection.

The magnetic field is provided by three large air-core toroids, two in the end-cap and one in the barrel regions, the field is approximately 0.5 T. Over the range of $|\eta| < 1.4$, muon tracks are bent by the barrel toroid, over the range $1.6 < |\eta| < 2.7$ by two smaller end-cap magnets insert in the ends of barrel toroid. The remain region, $1.4 < |\eta| < 1.6$, is called transition region and the magnetic field is provided by a combination of barrel and end-cap toroids. The magnetic field strength is such that the *sagitta*, i.e., the maximum deviation of the track from a straight line between the endpoints in the spectrometer, is on average $500\text{ }\mu\text{m}$ for a transverse momentum of $p_T = 1\text{ TeV}$. The system is also equipped with an optical alignment system, which monitors constantly the chambers relative position and deformation.

In the barrel region, the muon system consists in three cylindrical layers around the beam axis, while in the transition and end-cap regions, the three layers can be also perpendicular.

The muon trajectories are measured by the high-precision track chambers, i.e., Monitored drift tubes (MDT) which cover the region up to $|\eta| < 2.7$ and, since Phase-I, Micromegas (MMG), which cover the innermost end-cap wheels, in the region $2.0 < |\eta| < 2.7$. The MMG technology replaces the old cathode strip chambers (CSC) and the innermost MDT chambers.

The trigger system is composed by three layers of resistive plate chambers (RPC) in the barrel region ($|\eta| < 1.05$), three or four layers of thin gap chambers (TGC) and, since Phase-I, correspondently to the MMG, eight layers of the small-strip thin gap chambers (STG) in the end-cap regions ($1.05 < |\eta| < 2.4$). They also measure the muon trajectory in non-bending direction of the toroid magnets.

The hit coincidences between layers of RPC or TGC provide the muon first level

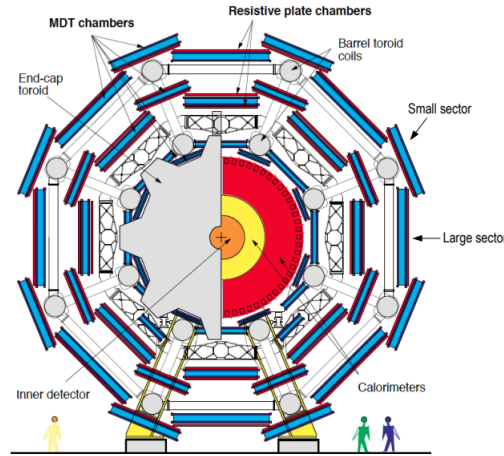


FIGURE 2.6: View of the Run1/2 ATLAS muon spectrometer layout in the X-Y (transverse) plane [26].

trigger (L1), with a maximum rate of 100 kHz and latency of $2.5 \mu\text{s}$. The second step of triggering is provided by the High Level Trigger (HLT) system, which allows a fully object reconstruction, based on a software technology. The recording rate of HLT is few kHz. More details about the trigger system are provided in Section 2.4.1.

The following sections are meant to provide all the information useful for a better understanding of my DQ work. In the first Section 2.3.4 the MS layout will be depicted, in the consequently sections, the MS sub-detectors MDT (2.3.4), RPC (2.3.4), NSW (2.3.4) and TGC (2.3.4) will be shortly treated.

Layout

The three muon system layers are called inner (I), middle (M) and outer (O) and are labelled BI, BM and BO in the barrel and EI, EM and EO in the end-cap regions. The middle and outer end-cap layers are called *Big Wheels*, while the innermost is the old *Small Wheel*, now replaced by the *New Small Wheel*. In the barrel-end-cap transition region ($1.05 < |\eta| < 1.3$), where the magnetic field change sign, there are also BIS7/8 or EIL4, EE (extra), and BEE layers.

The subsystem is divided into 16 sectors, respect the azimuthal ϕ direction, eight large (L) and eight small (S). Large sectors are located between the eight barrel toroid coils, while the small sectors overlap them in ϕ direction. The TGC system has instead 24 or 48 sectors in azimuthal direction. In the Phase-I upgrade, the old Small Wheel, the end-cap part made by CSC and EI1-3, is replaced by the New Small Wheel (NSW), composed by small strip TGC (sTGC) and MMG. The BIS7 and BIS8 MDT chambers will be replaced by integrated BIS78 stations or BIX/Y7 of new RPC and small diameter MDT (sMDT) chambers, in order to enlarge the trigger coverage in this region.

The barrel layout in the transverse plane is illustrated in Figure 2.6 and the MS in the Y-Z view after Phase-II for small and large sectors in Figure 2.7.

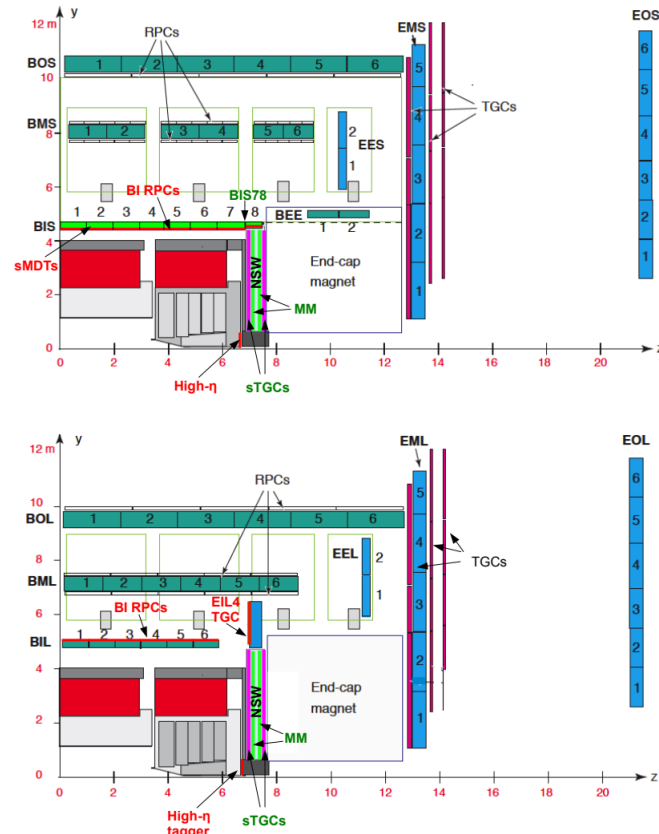


FIGURE 2.7: The Y-Z view of the ATLAS spectrometer layout for the small (top) and large (bottom) sectors [26].

MDTs

Monitored Drift Tubes (MDTs) are made by drift tubes with a diameter of 30 mm, operating with Ar/CO₂ gas mixture at 3 bar. Muons ionise the gas and generate electrons collected by the central tungsten-rhenium wire with a diameter of 50 μm . The main parameters of the MDT chambers are shown in Table 2.1. In order to optimise the solid angle coverage, the chambers are rectangular in the barrel and trapezoidal in the end-cap. The tubes have the same length in the barrel, while vary along R in steps of 24 tubes.

Figure 2.8 shows the structure of a MDT barrel chamber. An MDT chamber consists in two groups of tube layers, called multi-layers, separated by a mechanical spacer of aluminium. In the innermost layer of MS, the multi-layer is four-tube layers, while in the middle and outer three-tube layers only. The lateral parts, at the end of the tubes, are designed for readout (RO) and for the high voltage supply (HV). Since chamber deformations are expected, an internal chamber alignment system is implemented. Due to the continued monitoring of the MDT chambers, the relative position of the wires are well known, ensuring a good track resolution. For a track crossing three MDT chambers, a sagitta resolution of 45 μm is expected, for a momentum of units of TeV.

The structure of the MDT readout electronics is shown in Figure 2.9. The first stage of the readout is the mezzanine card, a printed circuit board directly connected to 24 tubes via a signal distribution card. There are about 15000 mezzanine cards for 354000

TABLE 2.1: Main MDT chamber parameters

Parameter	Designed value
Tube material	Al
Outer tube diameter	29.970 mm
Tube wall thickness	0.4 mm
Wire material	gold-plated W/Re (97/3)
Wire diameter	50 μm
Gas mixture	Ar/CO ₂ /H ₂ O (93/7/ ≤ 1000 ppm)
Gas pressure	3 bar (absolute)
Gas gain	2×10^4
Wire potential	3080 V
Maximum drift time	~ 700 ns
Average resolution per tube	~ 80 μm

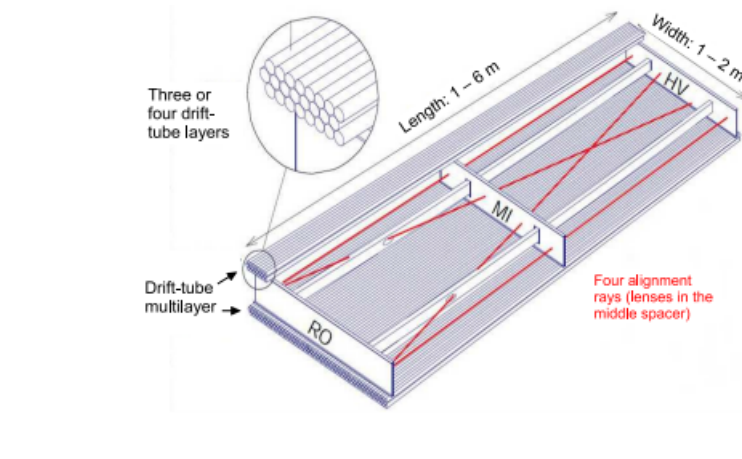


FIGURE 2.8: Mechanical structure of a MDT chamber [19].

MDT tubes. On the card, 8 tubes are connected to a ASD chip, which implement Amplification, Shaping and Discrimination functions. The binary differential output of the ASD is sent to the Time Digital Converter (TDC), where the arrival time of leading and trailing edges are saved with an identifier indicating the tube. The time is measured in units of TTC (Timing, Trigger and Control) clock of 40.08 MHz. The ADS allows also to measure the pulse height of the signal by an ADC (Analogue-to-Digital Converter), encoding the delay between trailing and leading edges. A delay of 150 ns corresponds to the maximum pulse height.

JTAG protocol is used for programming the ASD and TDC via a shift register, allowing a string of 54 bits for each ASD and 180 bits for the TDC.

One CSM (Chamber Service Module) controls all the MDT mezzanines, which are up to 18. The CSM broadcasts the TTC signals to the TDC and collects the hits to send to the L1 trigger. Then the data are formatted, stored and sent via optical link to the MDT Readout Driver (ROD) in the USA15 service cavern. The ROD is a VME module, which serves six CSMs.

RPCs

In the barrel region, corresponding to $|\eta| < 1.05$, Resistive Plate Chambers (RPCs) are chosen, due their good spatial and time resolution and their adequate rate capability.

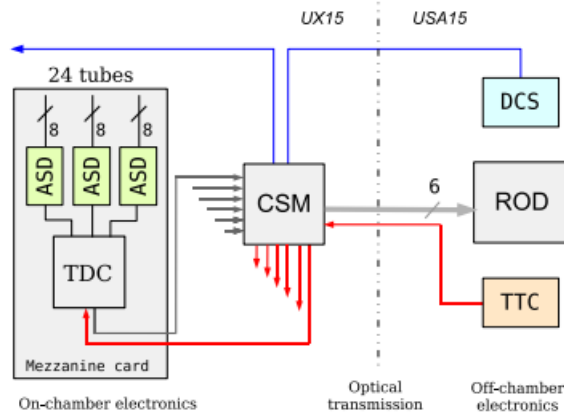


FIGURE 2.9: Schematic diagram of the MDT readout electronics [19].

In the barrel, the trigger system consists of three layers around the beam axis, called trigger stations. The Figure 2.10 shows a barrel muon sector and the position of RPCs respect the MDTs.

Each station consists of two detector layers, measuring independently η and ϕ . A track going through all the three stations has six measurements in η and ϕ . This allow to use the RPC1 and RPC2 coincidences for the low- p_T trigger and RPC3 coincidences for high- p_T .

The RPCs consist in two units, made of two gas volumes supported by spacers of 2 mm, four resistive electrodes and four readout planes.

The signal obtained from a RPC is compared with the other two RPCs along the path of the particle. This task is accomplished by a system of fast coincidence units.

TGCs

The end-cap region has different constraints respect the barrel region. A first difference is the muon momenta, since the p_T increases with η . This leads to an increased and η dependent granularity in the system. Furthermore, the radiation level increase up to a factor 10 in the end-caps. For these reasons, in the end-cap regions described by $1.05 \leq |\eta| \leq 2.4$, Thin Gap Chambers (TGCs) have been selected, which provide good time resolution and high rate capability.

TGCs provide the same RPCs functions but in the end-caps: the trigger and the determination of the second, azimuthal coordinates, complementing the radial MDT direction. In the Big Wheel there are seven layers of TGCs. The azimuthal coordinate in the MDT outer wheel is obtained by extrapolation of the middle layer, since in that region no magnetic field is present.

TGCs are multi-wire proportional chambers, with the characteristic of having the wire-cathode distance smaller than wire-wire one. The seven detector layers are arranged in one triplet and one doublets. The structure consists of wire planes as anode, cathode planes, strip planes, shields and honeycomb support structures, as shown in Figure 2.11. A chamber consists in a gas volume containing a wire plane and two cathodes, while a unit is the entire of three (two) chambers in a triplet (doublet) arrangement.

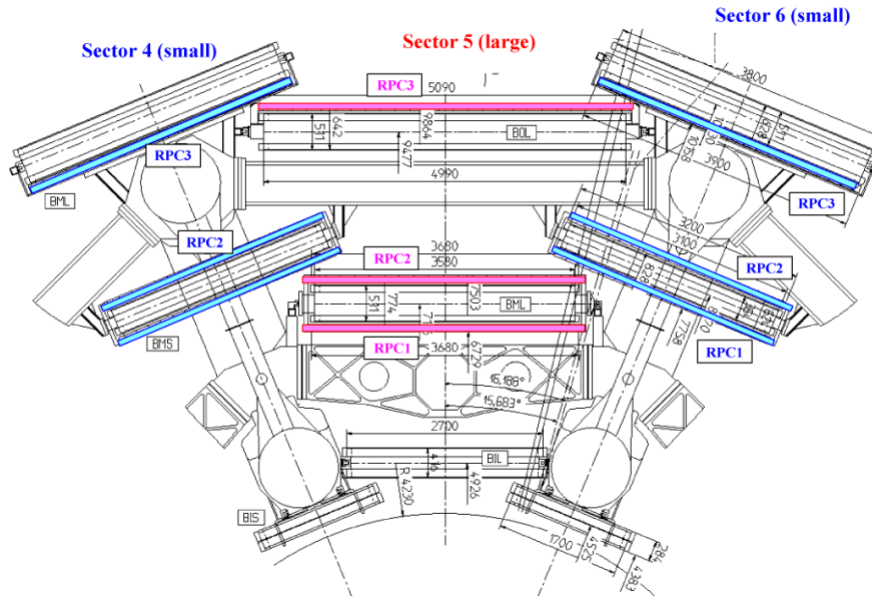


FIGURE 2.10: Upper part of the barrel of the muon spectrometer, with the RPCs in color. In the middle layer, RPC1 and RPC2 are upstream and downstream MDTs, in the outer RPC3 is only in the inner side for the small sectors and in the outer side for the large sectors [19].

NSW

The New Small Wheels (NSW) [27] are the part of the muon system which replaces the "old" Small Wheels, which consists of the Cathode Strip Chambers and the innermost end-cap MDT chambers with Micromegas and small-strip TGC (sTGC) chambers, in order to fulfill the increasing performances of LHC.

The layout is the same as EM and EO, with eight small and eight large detectors.

The basic design of NSW consists in sTGC-MMG-MMG-sTGC order in z direction, in which each sTGC is a quadruplet. A quadruplet contains four sTGCs, with pad, wire and strip readout.

The internal structure of the sTGC is shown in Figure 2.12. It consists of a grid of gold-plated tungsten wires sandwiched between two cathode planes, behind which on one side there are strips and on the other side pads. The strips have 3.2 mm pitch, much smaller than the TGC, hence the name small TGC.

The **micromegas**, an abbreviation for Micro-Mesh Gaseous Structure, are used as a tracking system of the NSW, fulfilling the requirements of good position resolution independent of the particle incident angle and high efficiency in high background conditions.

MMG detectors consist of a planar electrode, a gas gap acting both as conversion and drift region, and a thin metallic mesh which separates the amplification region from the drift one and the read-out strip planes. A schematic representation of the MMG principle is shown in Figure 2.13. The drift space is crossed by charged particles which ionise the gas. The electrons produced drift towards the mesh. The electron avalanche takes place in the amplification region, while the ions produced in this step move in the opposite direction back and are neutralised at the mesh.

MMG detectors are provided by a spark protection system.

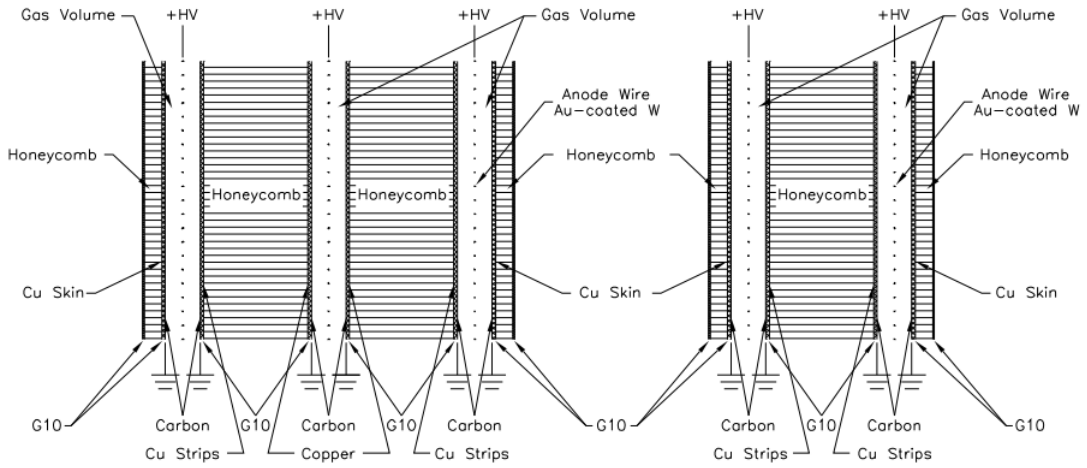


FIGURE 2.11: Cross section of a TGC triplet and doublet module. The triplet has three wire layers and two strip layers [19].

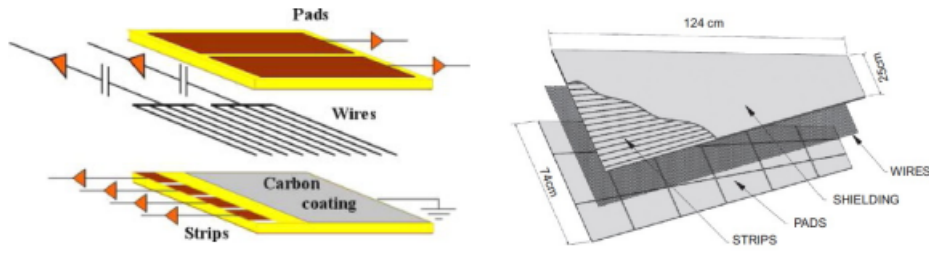


FIGURE 2.12: The sTGC internal structure [27].

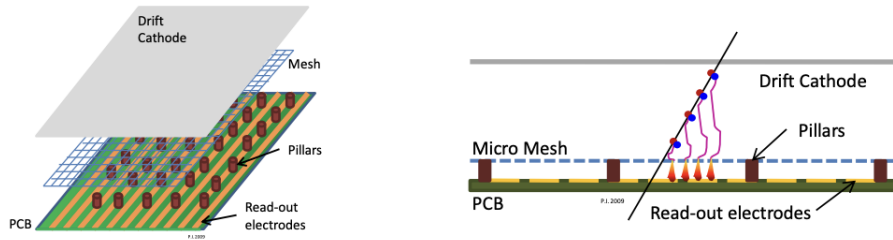


FIGURE 2.13: The sketch of the internal structure of a MMG [27].

Differently by a traditional MMG, in the MMG detectors for NSW, the resistive strips are taken at positive HV, while the amplification mesh is at ground.

In the MMG detectors the amplification mesh is not integrated in the readout structure. The readout strips are implemented on thick printed-circuit-boards (PCBs) and then covered by a $64 \mu\text{m}$ layer of insulator, followed by the resistive strips. Each sector has 8 PCBs, with 1024 read-out strips each. One PCB board is read out by two readout boards, carrying eight front-end chips VMM of 64 channels each. In Figure 2.14 the internal structure of a PCB is shown.

The readout electronics is located on the side of the chambers, which requires low

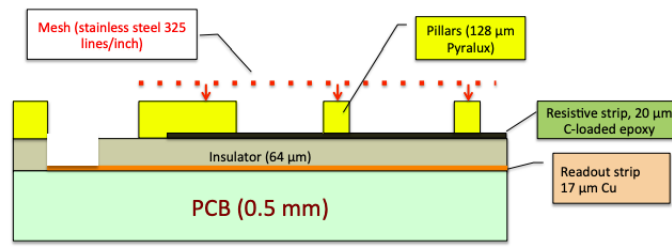


FIGURE 2.14: Internal structure of the MMG readout boards [27].

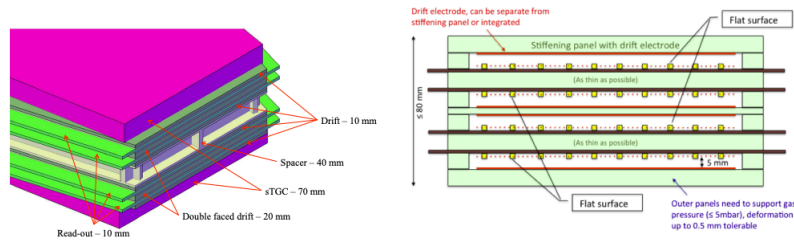


FIGURE 2.15: On left arrangement of the MMG and STG detectors in a sector and on right arrangement of MMG planes in a multiplet [27].

voltage and optical fibre connections using GBT link. The conversion of the low voltage is done on the corresponding front-end or GBT boards. There are other two types of electronic boards for transmitting the data via optical links to USA15. For the trigger, one trigger card takes data from 32 VMM chips, while for the readout data are transmitted by a single GBT link per 16 front-end cards or per plane. Each sector has eight MMG detector layers, grouped in two multiplets of four layers each, separated by a 40 mm spacer. Each multiplet has four active layers, grouped into two pairs, one pair is equipped with η strips and the other with stereo strips. In Figure 2.15, MMG arrangement is shown for a sector and for a multiplet.

2.4 Selection and Data

In this section, some information about the ATLAS data acquisition system are provided.

In particular, Section 2.4.1 summarizes trigger and data acquisition system, while Section 2.4.2 gives an overview on the data processing after the trigger selection.

2.4.1 Trigger and data acquisition system

High-energy experiments, as the ATLAS experiment, typically collect a large number of data every beam interaction. Therefore, it is necessary to have an efficient trigger system, in order to remove uninteresting events and to save the physical meaningful ones.

The ATLAS experiment uses a complex Trigger and Data Acquisition System (referred as TDAQ [29] in the following) to select events, to reduce the event rate from 40 MHz to 1 kHz, number which is constrained by the maximum data storage of a

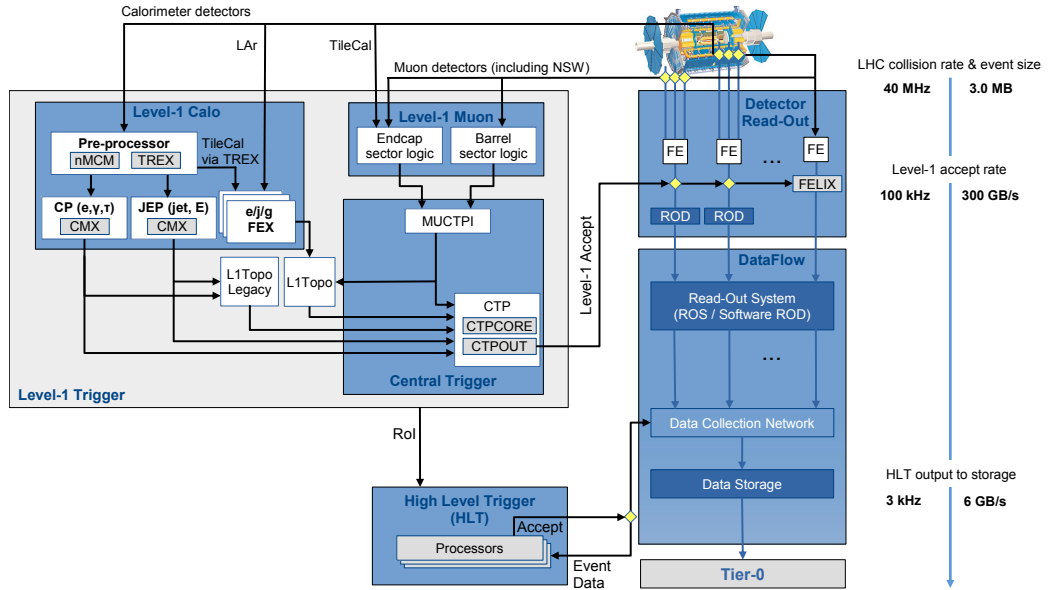


FIGURE 2.16: The ATLAS TDAQ system, with the latest updates for Run 3 [28].

few hundred MB/s.

The ATLAS trigger system is composed by two levels. The first is the L1 hardware-based trigger, which uses a subset of detector information in order to reduce the rate below 100 kHz, which corresponds to the maximum detector readout rate. The second stage consists of the software-based High Level Trigger (HLT), which further reduces the rate to 1 kHz (on average). In Figure 2.16, a schematic representation of the ATLAS TDAQ system with the latest updates for Run 3 and beyond is shown.

The Level-1 trigger uses calorimetric and muonic information in order to determine the Region of Interest (RoIs) in the detector, which represent the position of the triggering object in η - ϕ plane. Among the updates of the Run 3, the Level-1 Muon is also making use of the NSW detector information for triggering. The final decision is taken by the Central Trigger Processor (CTP), which addresses the information to the following stage, the HLT.

The HLT is responsible for the final physics selection of the events for off-line analyses. The decision is an OR of many independent trigger chains, defined by the L1 decisions followed by a series of HLT algorithms.

The DAQ system, in Run 1 and Run 2, consists of front-end electronics (FE) read out via point-to-point links, called Read-out Drivers (RODs). Each detector implements the design of its ROD, usually using the VMEbus technology. RODs have several roles, from checking errors to monitoring the data. The FrontEnd Link eXchange (FELIX) is a new interface between the FE and DAQ system recently developed, with the aim of improving the quality and the flexibility of the readout system.

Up to now the data processing, formatting and monitoring is provided by an hardware component, which instead, with the FELIX system, is going to take place in software running on server farms, a facility known as the Software Readout Driver

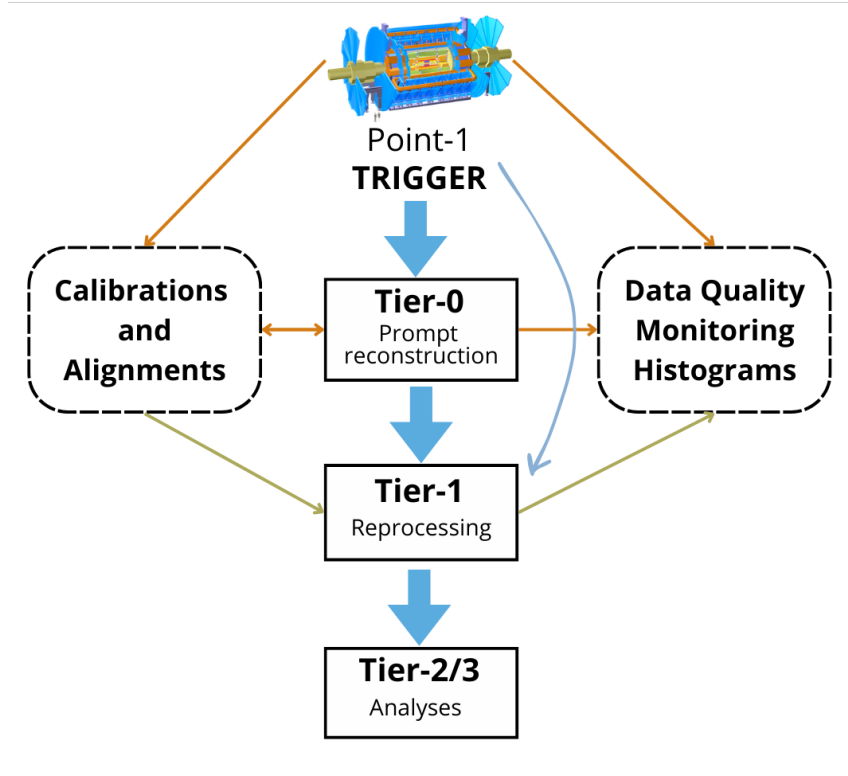


FIGURE 2.17: The data-flow chain from the trigger to the analysis.

(SW ROD). During the long Shutdown 2 (2019-2022), the new system has been installed and commissioned by some detectors, i.e., the NSW, LAr digital readout, BIS7/8 and the calorimeter hardware trigger electronics (L1Calo) [30].

In Figure 2.16 is also shown the DataFlow (DF) as final stage of the TDAQ system. At the end of the L1 data processing, in fact, the selected events are moved and buffered to the Read-out system (ROS). The final HLT selection is based on ROS storage data.

2.4.2 Data processing

The data collected and recorded after trigger selections are still in a *raw* format, which means that they need to pass further steps before being ready for the analyses.

The data are stored in *streams*: collections of events or event fragments in the same dataset. Each stream depends on HLT decisions. In Run 3, some of the stream definitions are:

- **Main stream:** the physic stream used for data analyses, almost used during stable beam data taking.
Other physic streams are: MinBias (special runs), CosmicCalo, Standby, Late, L1Calo;
- **Express stream:** made by events submitted to a prompt reconstruction, which overlaps exactly the Main Stream;
- **Calibration stream:** recorded events with the minimum amount of information necessary for detector calibration.

From the raw data there are several reconstruction steps, necessary to provide objects suited for the analyses, such as tracks or jets.

The data-flow chain, from the trigger to the analyses, is shown in Figure 2.17. The first step of processing data is the bulk or prompt reconstruction, which is done after the calibration loop, usually after 48 hours from the end of the run. This step is done in the Tier-0 facility, a big computer farm located at CERN [31, 32].

Subsequently, a reprocessing is done, in order to improve the reconstruction and fix some bugs. Since the Tier-0 is dedicated to the bulk reconstruction, this reprocessing step is performed on the Tier-1 grid, computer farms located around the world. In the end, Tier-2/3 are arranged for the last analyses usage.

The analyses need also a Good Run List (GRL) for selecting interesting events: a list of data which fulfills certain criteria provided by the Data Quality Monitoring infrastructure. The Data Quality Monitoring Online and the tools used in the ATLAS Control Room are described in the next Chapter 3.

2.5 Reconstruction and identification of physics objects

Electrons, muons and photons are the main final state particles observed in the p - p collisions at LHC, since they are stable particles. They can be the result of a particle decay as prompt particles or they can be part of a jet of particles, generated by the hadronization of a quark. For this reason, developing and performing techniques is really important in order to optimize the correct reconstruction and identification of these objects.

In this section, I will focus on muon 2.5.1 and electron 2.5.2 reconstruction and identification, and jet reconstruction 2.5.3, since they are matter of interest of the analysis described in the last Chapters 5 and 6.

2.5.1 Muon reconstruction and identification

Muons are reconstructed using information from both the ID and the muon spectrometer tracking detectors [33]. Calorimeter information can also be exploited in the determination of track parameters or in the energy loss quantification.

Muon reconstruction starts with the standalone muon spectrometer information, identifying segments from hits in the individual MS station. The global muon reconstruction is done using the complete detector information and implementing five different strategies, which differentiate in the combination of ID, MS and calorimeter signals.

The high-quality muon candidates are selected by a set of requirements on the number of hits in the ID and MS detectors, on the track fit proprieties and on the compatibility of the individual measurements in the two detector system. The set of requirements are referred as *working points* (WPs). The optimization of the WPs is developed with the selection efficiency and purity in simulation. In particular, the prompt muon efficiency of a selected WP represents the probability that a prompt muon is reconstructed as a muon and satisfy the WP criteria. The purity of a selection WP is one minus the hadron misidentification rate, as the fraction of light hadrons reconstructed as muons and satisfying the WP. In order to increase the purity, trying to not reduce much the efficiency, three WPs are designed: *Loose*, *Medium* and *Tight*. The *Medium* provides a good compromise for several analyses, while keeping the systematic uncertainties in the prompt muon efficiency and background rejection

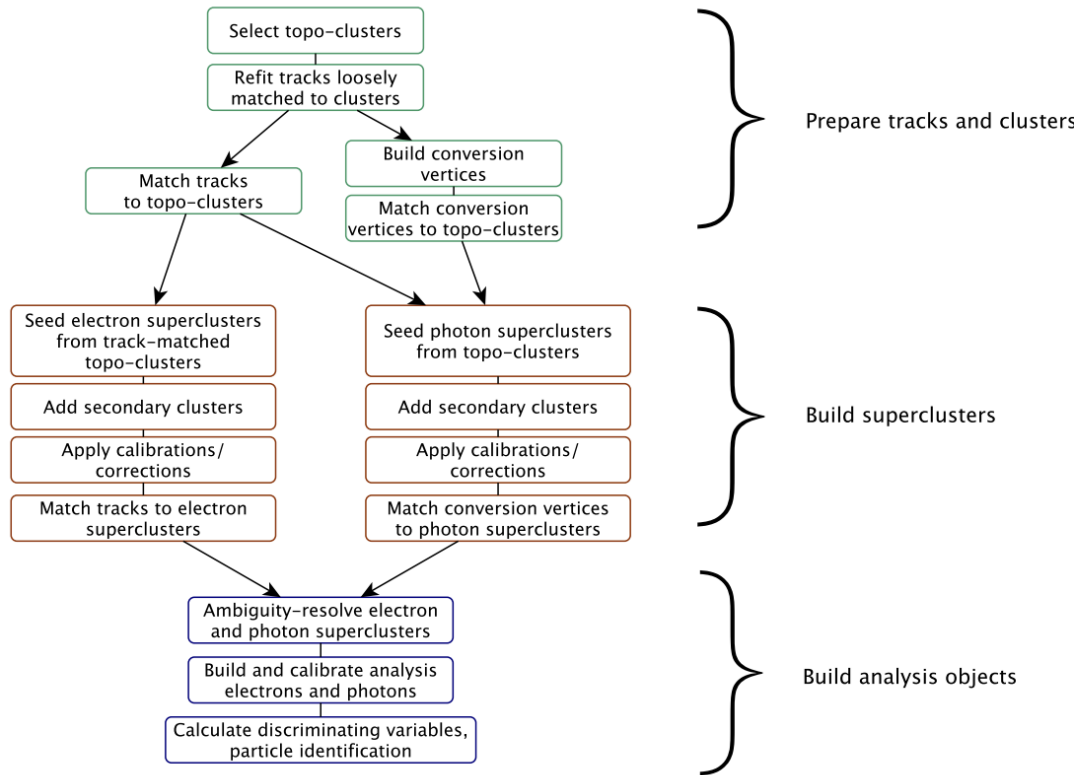


FIGURE 2.18: The algorithm flow for photon and electron reconstruction [34].

small. In the *Loose* case, the efficiency is higher at cost of less purity and larger systematic uncertainties. Finally, the *Tight* WP provides the highest purity, offering an improved background rejection, at cost of a few percent efficiency loss for prompt muon compared to *Medium*.

2.5.2 Electron reconstruction and identification

Electrons and photons are defined as clusters built from energy deposits in the calorimeter and matched track or converted vertex. Photons are defined *converted* if the energy deposits match with converted vertices or tracks, otherwise are *unconverted*, i.e., they have interacted upstream the calorimeter and have produced an electron-positron pair.

The reconstruction of both electron and photons is performed by an algorithm, which firstly prepare the tracks and clusters. It then selects the *topo-clusters*: clusters of energy topologically connected with EM and hadronic calorimeter cells. These clusters are matched to ID tracks, which are re-fitted as electron candidates accounting for the energy losses for bremsstrahlung. The topo-clusters are also matched with converted vertices by the algorithm.

At this point, *super-clusters* are obtained through the algorithm, then corrected with initial position and energy calibrations.

Consequently, the electron super-clusters are matched with tracks, while the photon super-clusters with the conversion vertices. In the last stage, the physics objects needed by the analysis are built and, in order to distinguish electrons from photons, the discriminant variables are obtained.

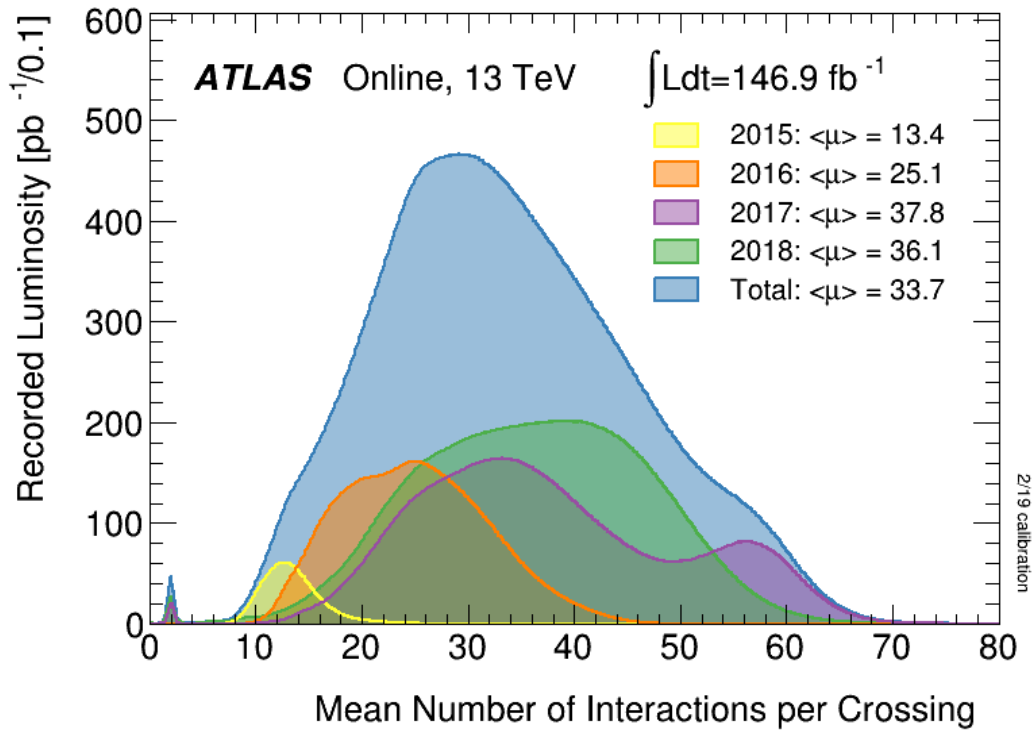


FIGURE 2.19: The mean number of interactions per crossing for the 2015 – 2018 pp collision data at 13 TeV centre-of-mass energy [36].

A scheme of the algorithm flow is shown in Figure 2.18. The algorithm and the most recent performances of Run 2 are reported here [34].

In order to improve the purity for the selection of prompt electrons, further quality criteria, based on identification selections, are applied on the reconstructed candidate electrons. The identification is based on a likelihood discriminant variable, built from quantities chosen for their capability to distinguish prompt electrons from misidentified particles, as hadrons or electrons from photon conversion, or not isolated electrons from heavy flavour decays. Like muons, also for electrons operating points are defined, as *Loose*, *Medium* and *Tight* at which the electrons satisfy different values of the likelihood discriminant variable, correspondingly at some predefined efficiencies. The values of the WPs are presented in ref. [35].

2.5.3 Jet reconstruction

Jet reconstruction in ATLAS is usually divided in two steps: the definition of the input objects and the jet algorithm implementation.

For the input objects the topo-clusters are used, formed by seed calorimeter cells with more than 4σ of energy, where σ is the expected noise. Iteratively, all cells adjacent to the seed cells in 3-dimensions are grouped together. Topo-clusters are then calibrated both at the electromagnetic scale, considering the isolated topo-cluster as the sum of the cell energies, and at hadronic scale, accounting for the difference between the electromagnetic and hadronic interaction in the ATLAS calorimeter.

In Run 1 of the LHC, the majority of the ATLAS analyses used only topo-clusters as input for the jet reconstruction, formed the hadronic jets. Recently also the ID tracks associated with the jets are taken into account, as input for the *particle-flow* jet algorithm [37]. The algorithm consists in a cell by cell subtraction energy, in order to

remove any overlaps between the ID and calorimeter energy measurements. After a selection of the *well-measured* tracks, in fact, the algorithm attempts to match each track to a single topo-cluster in the calorimeter. For each track/topo-cluster system, the algorithm evaluates the probability that the energy of the particle is deposited in more than one topo-cluster. If this is the case, the particle-flow algorithm adds more topo-clusters in order to reconstruct the full shower. The energy deposited by the particle which formed the track is subtracted cell-by-cell from the selected topo-clusters. The particle-flow algorithm provides a list of the tracks, all the topo-clusters and the new topo-clusters after the energy subtraction.

The topo-clusters and the tracks are matched to the hard-scatter primary vertex.

The particle-flow and calorimeter jets are then reconstructed using the anti-kt algorithm [15] with radius parameter 0.4, briefly explained in section 2.2.3. In case of large-R jet reconstruction, such as boosted Higgs decay, the radius parameter is required to be 1.0. The algorithm needs to consider the *pile-up* events, which are the interactions inside each bunch. Pile-up is measured by the mean number of interactions per bunch crossing $\langle\mu\rangle$, the distribution for the Run 2 data is shown in Figure 2.19. A jet calibration is then applied on the reconstructed jet, in order to correct for the pile-up interactions and to correct the reconstructed energy for the mean value of the truth jet energy, using a jet energy scale (JES) correction factor [38, 39, 40, 41].

Chapter 3

The ATLAS Online Monitoring system



FIGURE 3.1: Members of the ATLAS Collaboration celebrate the first Run 3 beam splashes in the ATLAS control room, 22 April 2022. I was present in the ACR (close to the rightmost window) acting as a muon shifter [42].

3.1 Introduction

Online monitoring is an essential tool in order to verify whether the huge number of electric channels work properly during the high event rate seen by LHC. During data taking, the monitoring is performed constantly in order to spot immediately any problems and react promptly. The monitoring has also the meaning of constantly testing the quality of the collected data, procedure known as Data Quality (DQ) monitoring.

The DQ, during the data taking, is referred as DQ Online and mainly monitors data obtained after the first step of the trigger (L1) or, in some cases, after the high level trigger (HLT). In the so called DQ Offline, instead, the data quality is evaluated after the HLT and the reconstruction provided by the Athena software [43]. In this case, the meaning of the data quality is to assure a good physic reconstruction.

The DQ is part of the *operation* activities of the experiment. The operations are the sum of all the activities that checks the experiment is working properly. During data taking, the experts for each sub-detector of the experiment discuss the status of the detector at least once a week in a working meeting.

During these years, I have worked in the DQ Muon Online team as *expert*, I joined the operation activities for the Muon Spectrometer, attending the meetings in order to discuss and report about the status of the Data Quality monitoring online of the muon system. I have also worked at CERN during data taking as support for the data quality activities in the ATLAS Control Room, located at Point 1 (P1) of the LHC ring. In this chapter, my contributes to the Online Data Quality are shortly presented. In section 3.2, an overview of the monitoring frameworks is presented, then in Section 3.3 the monitoring tools used for the DQ are described. Then in Sections 3.3.1 and 3.3.2, the Muon Spectrometer DQ will be detailed, since it conveys part of my PhD work.

In Figure 3.1 some members of the ATLAS collaboration, among which me, are celebrating in control room during the beam splashes in 2022. The splash beam [44] consists in a single proton beam hitting a collimator placed ahead of the ATLAS detector.

3.2 Monitoring framework

In the TDAQ system, there are many parameters which are constantly monitored during data taking: physics data quality and integrity, consistency of trigger information, status of the software and hardware, etc. The error reporting is another important aspect of the monitoring, in order to notify quickly any malfunctioning of the experiment.

The Monitoring Framework has the task of transporting the information from the place where it is produced to the places where it can be processed.

In order to optimize the performances of the monitoring, several services are created for different types of information:

- the **Event Monitoring** (Emon) Service is responsible for transportation of events to the applications for the monitoring. User programs can ask sampling events with selected proprieties, like trigger or sub-detector type;
- the **Information Service** (IS) is responsible for archiving and sharing any kind of information. It allows TDAQ applications to exchange information with the user during the run;
- the **Online Histogram Service** (OHS) manages histograms, providing a temporary storage between histogram producers and displays. OHS is based on IS extending its functions to handle histograms;
- the **Message Reporting System** (MRS) transports error messages among TDAQ applications.

All the services are based on the same communication service, the Inter Process Communication (IPC) for object implementation and remote location. It is built on the top of the CORBA (Common Object Request Broker Architecture) Architecture, which provides the inter-object communication and it is a standard designed in order to facilitate the communication of systems deployed on different platforms. CORBA

enables the exchange of information by the hardware platforms, programming languages and operating systems.

Figure 3.2 shows the architecture of the Monitoring framework.

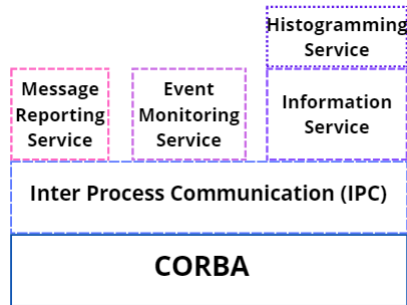


FIGURE 3.2: Monitoring framework architecture.

3.3 Monitoring Tools

Many monitoring tools have been developed during last years [45] in ATLAS. They are customised, written specifically for the experiment and not for commercial use. They communicate with the services described above and have the task to monitor the information received and spot any malfunctioning.

The **GNAM** is a light framework for detector monitoring, developed in a detector-independent way. It is modular by construction, in order to separate the common actions from the detector dependent ones, controllable either by TDAQ or by the users and the histogram production and visualization independent of one another. A schematic representation of GNAM and its possible interactions with the other software is shown in Figure 3.3. The data source can be either an online sampler from Emon (ROD, ROS, or SFI data) or a file sampler. In the case of a file sampler, an application is developed to read a raw data file and fill the Event Monitoring Service, emulating the online data source. The histogramming library brings data related to the detectors, applies a simple analysis and fills histograms. The core handles the histograms, publishing on OHS server, saving them into a file and executing user commands on request.

OHP is the **Online Histogram Presenter**, designed to be an interactive presenter in the framework of the Online Monitoring System. OHP has two functions: it can both browse the OHS and show a set of online histograms to be checked by both shifters and experts.

In Figure 3.5, a screenshot of the GUI for the OHP in the browser mode and in Figure 3.6, the OHP tabs of a muon subsystem are shown.

OHP has been configured in order to perform operations like fitting, zooming and changing the appearance of the plots. OHP is divided in the following subsystems, as shown in Figure 3.4:

- the Core Subsystem which provides access to the histograms from the Graphical User Interface, it is also responsible to coordinate all the other components, as the Histogram receiver, which receives notifications and retrieves histograms from OHS request;

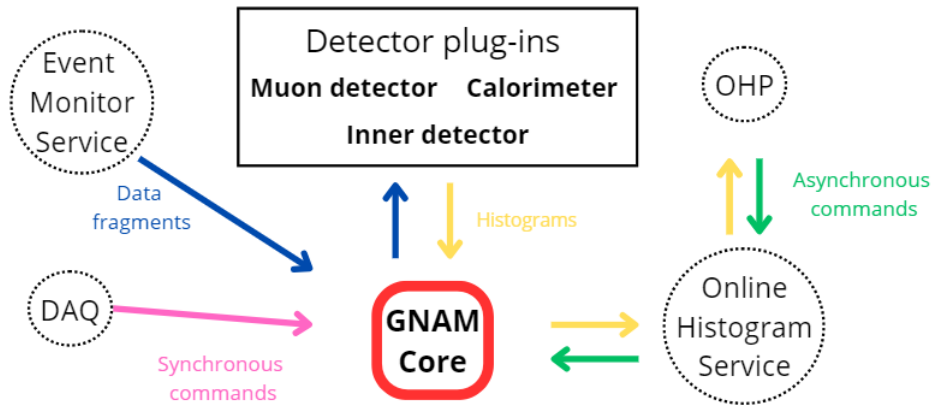


FIGURE 3.3: GNAM framework scheme and its relation with the ATLAS Online Monitoring environment.

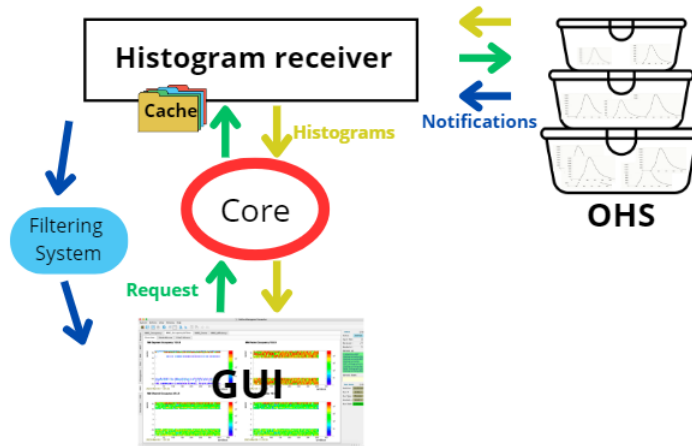


FIGURE 3.4: Schema of the OHP architecture.

- the Cache which is used to store the received notification and histograms, each Histogram receiver has its own cache, aimed to minimize the network traffic;
- the Filtering system which filters notifications to the related canvas;
- the Graphical User Interface (GUI), which uses as underlying technology ROOT, providing functionalities, as fitting, zooming to the users.

The **Data Quality Monitoring Framework (DQMF)** [46] is the ATLAS framework capable of monitoring the data quality for both online and offline. In the DQMF, a set of data are checked by *DQParameters*, which specify which histogram and algorithm are using. The *DQAlgorithm* provides the algorithm and the parameter to apply to each histogram. The algorithm gives a result (*DQResult*) about the quality of the data, comparing the histogram with some predefined thresholds. The *DQParameters* are grouped in *DQRegions*, which can be organized in other sub-*DQRegions*, creating a *DQTree*. The schema and the logic of how to combine the regions and the parameters are defined in a configuration database. Propagation of *DQ Results* through the tree

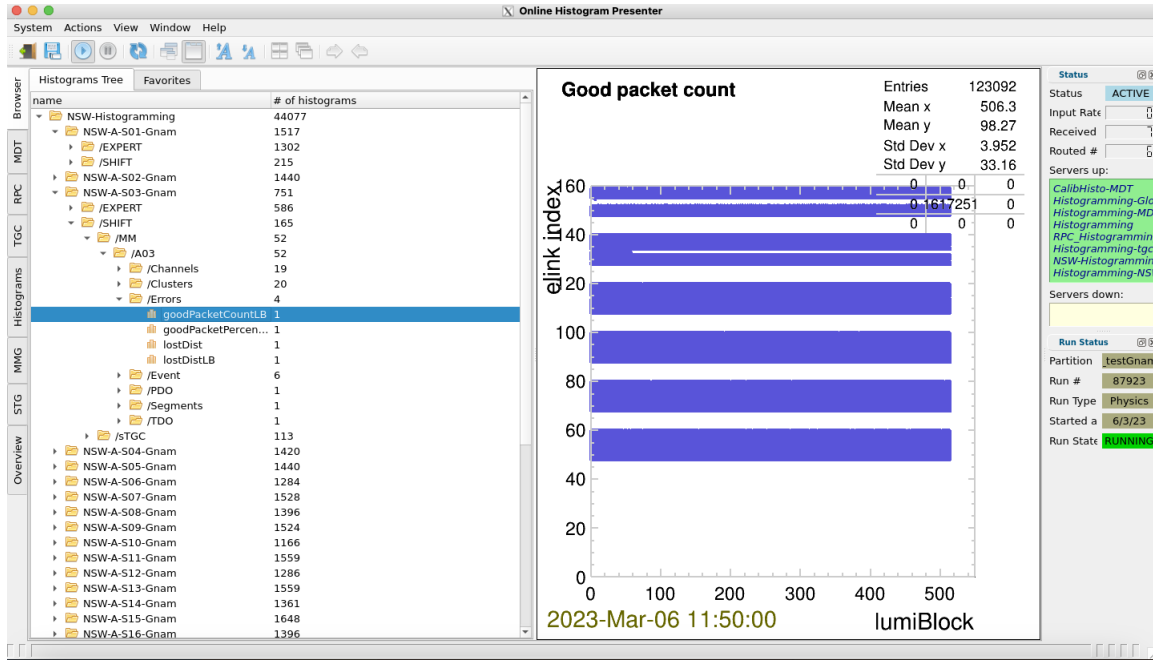


FIGURE 3.5: Screenshot of the OHP GUI, in the browser mode.

of *DQ Region* is also possible to ease the observation of problems.

The result of the checks from DQMF are published to the online Information Service and archived. The information can be accessed by the **Data Quality Online Display (DQMD)** for visualization. The tool has been implemented using the QTROOT library of ROOT program. The display has been built using CORBA technology.

The DQM display allows easily the navigation through the *DQRegions* and *DQParameters*, providing a quick view of the Results produced by the DQMF.

In Figure 3.7, the ATLAS DQMD available at P1 is shown. In the part above the DQMD provides histograms for any ATLAS sub-detectors, while the part below shows summary plots of the DQMF results. Results and information are continuously refreshed when new data are available.

The summary panel consists in the *History* of the DQMF ATLAS results and the *Alarms & Logs*, which show the errors originating from reading the configuration database or general information by DQMD. The detailed panel is organized in a tree of regions and parameters for the sub-systems. On the right, the sub-systems are organized in tabs which contains a tree of sub-tabs for specific detector parts. For each part several histograms are chosen for being checked by the DQMF. On the left, the same information reported as DQTree.

The DQMF layout is written in a configuration language of XML files and converting in a IGUI interface to facilitate use.

The DQMD and OHP can be opened using the **TDAQ Panel**, which allows experts and shifters to setup the partition configurations and open utilities for the data monitoring. The partition consists of a finite state machine, which allows to configure the detector for the data taking, and supports several plugin panels for different sub-detector tasks. In Figure 3.8, a screenshot of the panel is shown, where different utilities are visible, included the OHP and DQM Display.

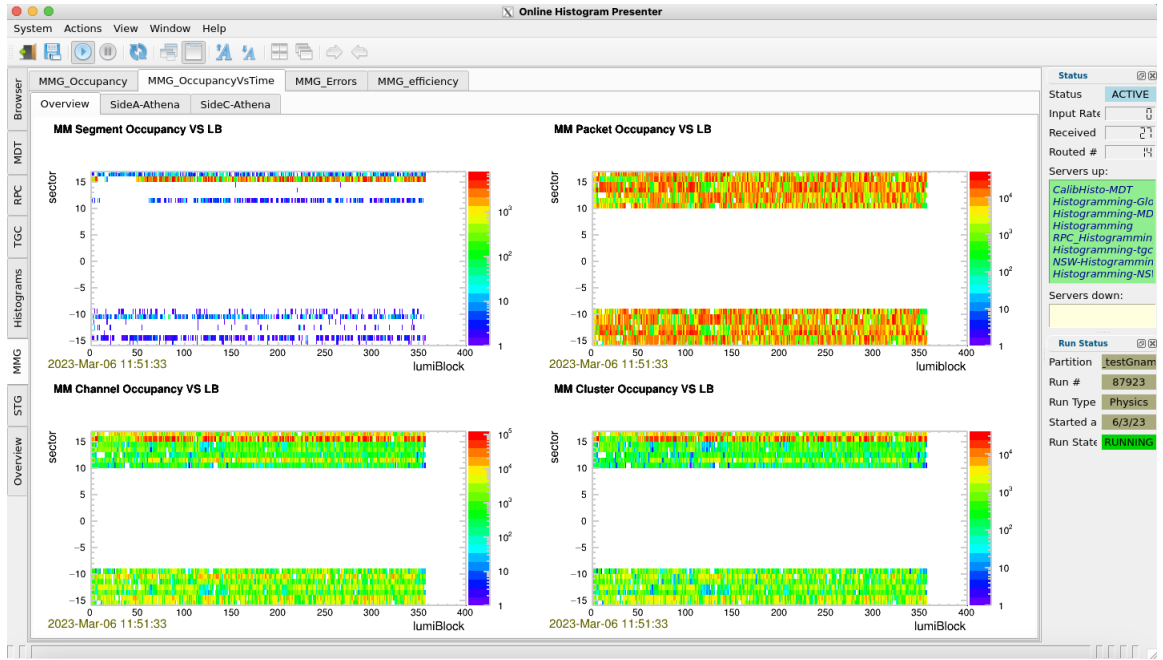


FIGURE 3.6: Screenshot of the OHP GUI, in the tabs mode.

3.3.1 Muon Spectrometer DQMF

The screenshot of the all muon spectrometer DQMF is reported in Figure 3.9. The muon spectrometer layout is composed by five groups, one for each sub-detector: MDT, TGC, RPC, MMG and STG.

In the rest of the present chapter, I will focus on the MDT and MMG DQMF since it was the subject of my qualification task and part of my PhD work.

MDT DQMF

In Figure 3.10, a screenshot of the MDT DQMD available in ATLAS Control Room (ACR) since Run 3 is shown. The layout is chosen in order to mimic the actual position of the MDTs in the spectrometer. The chambers are shown as small squares, divided in three blocks for the outer, middle and inner layer respect the vertical line; while divided in end cap and barrel for side A and C, on the horizontal line.

For each chamber, browsing down through the tree, five type of histograms are available [47]: ADC, TDC and Occupancy for each multi-layer, and Layer and Error for both layers together. Figures 3.11 and 3.12 show the histograms, as an example, relative to the BIL1A01 chamber. Histograms are selected from GNAM and MDTCalibGnam, which includes histograms produced by the Calibration Stream [48], depending on the source which is more suitable. Calibration stream extracts single muon track data from the HLT trigger, selecting hits from isolated muons for calibration. Plots from the calibration stream contain only muon hits, differently from the regular online histograms, offering more details since they have less background hits.

In the following, the histograms of the MDT DQMF and the relative algorithms are explained in details.

The **ADC distribution** (Figure 3.11a) is the total charge generated by the particle

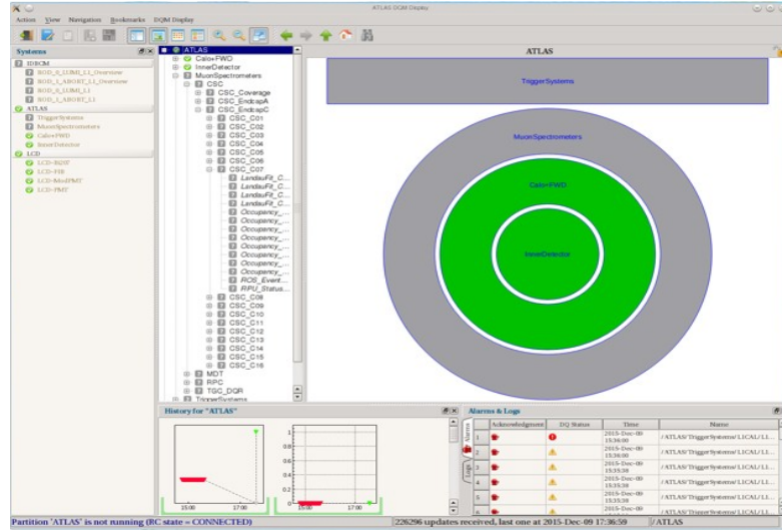


FIGURE 3.7: Screenshot of the ATLAS DQMD. The window consists of a detailed part above and a summary panel below. The detailed part shows histograms and algorithm results for each sub detector, the summary part shows the results for the ATLAS detector.

as it passed through an MDT tube. The histogram is provided per MDT chamber multi-layer by the Calibration Stream. The charge distribution is shown in ADC counts with a cut on 50 where is usually located a noise spike. The algorithm used for testing the histogram is the *ADCAAlgorithm*.

The **TDC distribution** (Figure 3.11b) is the travel time of secondary electrons in the tube after a particle passes through it. The width of the spectrum therefore illustrates the range of travel times of these freed electrons. As for the ADC, the histogram is provided per MDT chamber multi-layer by the Calibration Stream and only actual tracks are plotted, with a cut on 50 ADC. The two main features of this histograms are the location of the rising edge t_0 and the position where the spectrum drops off (t_{max}). However, it is important that the shape of the distribution below t_0 and above t_{max} is flat, and roughly the same value on either side. The algorithm, *TDCAlgorithm*, finds t_0 and t_{max} , computes the difference, and compares this Δt to some expected value.

In the Occupancy section, the **hits per event vs LB** shows hits per each multi-layer as a function of LBs. The histogram comes directly from the L1 trigger from the GNAM server, showing the hits not cleaned from background tracks. An example for the BIL1A01 is shown in Figure 3.11c.

In the Layer section, the multi-layers and single layers hits are checked. The **hits per tube and per layer** is a 2D histogram, showing the layers occupancy vs tubes per layer and taken from the Calibration Stream, as in Figure 3.12a. The algorithm used is the *LayerAlgorithm*.

The DAQ system can propagate a variety of error signals for each chamber. The **Error Rate** (Figure 3.12b) histogram collects the rate (that is, the number of times the error was sent divided by the total number of hits) at which these errors stand



FIGURE 3.8: Screenshot of the TDAQ panel at P1.

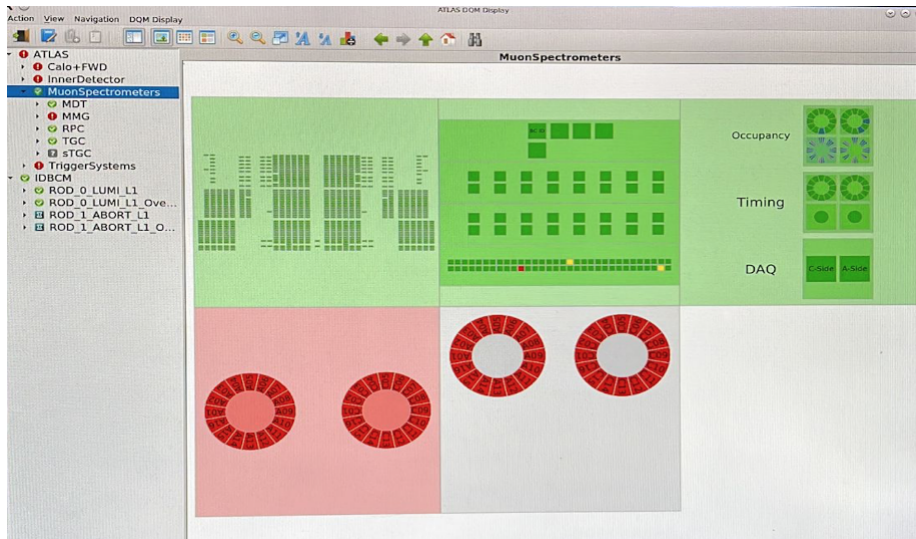


FIGURE 3.9: Screenshot of the ATLAS Muon DQMD. MDT (up, left), RPC (up, middle), TGC (up, right), MMG (down, left) and STG (down, right),

for lumiblocks (LBs). Each error has its own tolerable ranges, set individually in the configuration file. The *ChamberErrorsAgo2D* compares each kind of error rate with the correspondent threshold.

More details about the algorithms, the parameters and the thresholds are reported in Appendix B. These algorithms have been implemented in the past, my main contribution is to maintain them, modify them when needed.

MMG DQMF

At the end of my qualification task, I started working for the MMG online data quality. The framework configuration was at a first stage and the algorithms not implemented. I had then the opportunity, with other collaborators and experts, thanks also to my experience in control room, to choose the suitable histograms for being checked by the shifters, to create the best shape for the framework and to develop and test the



FIGURE 3.10: Screenshot of the MDT DQMD available in ATLAS Control Room since Run 3. The screenshot shows events from 13.6 TeV pp collision, run number 364292.

algorithms.

For the MMG data quality check, the layout of two wheels, one for side A and one for side C, is chosen for the DQMD. Together with the two wheels, a square shows global histograms: the entries are for all the MMG system.

For the per-sector part, each chamber has three groups of histograms: **Errors**, **Event** and **Layers**.

Details about the algorithms are given, since I implemented, then tested and chose the most suitable threshold values after discussions with the experts.

In the **Errors** folder there are three histograms which are checked: the Good Packet Percentage vs LB, the Missing VMM vs LB and the lost e-links vs LB (Figure 3.13). The three plots are linked with the electronic signal provided by the MMG channels.

Good Packet Percentage vs LB: for each sector the histogram is filled with the percent of good packets, i.e., e-links which are sending data. The algorithm I implemented is called *GoodPacksVsLB* and checks for each e-link and per a certain interval of LBs if the data are above certain thresholds. The algorithm has one parameter and two thresholds:

- **Parameter:**
 - **nLBChecked:** currently set at 5 LB. It corresponds to the interval of LBs checked by the algorithm. It is also the minimum value of LBs at the beginning of the run for starting the algorithm checking. Before that value the histogram is flagged grey.
- **Thresholds:**
 - **RateThreshold:** for each e-link is tested if the percent is above the RateThreshold. The value is averaged on the number of LBs checked. One e-link is considered working properly (green flag) if the value is above 85%, red below 80%, orange in between;

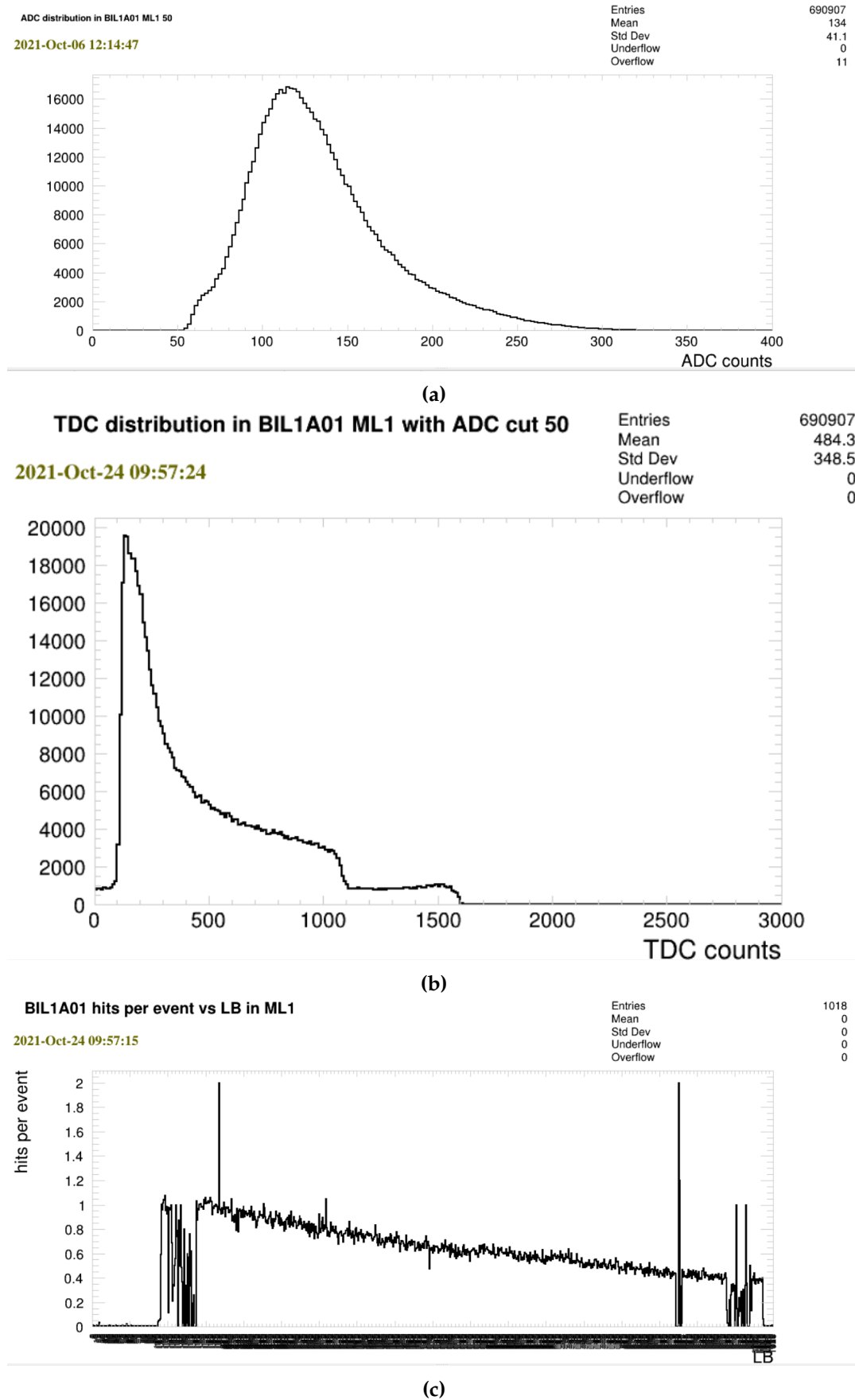
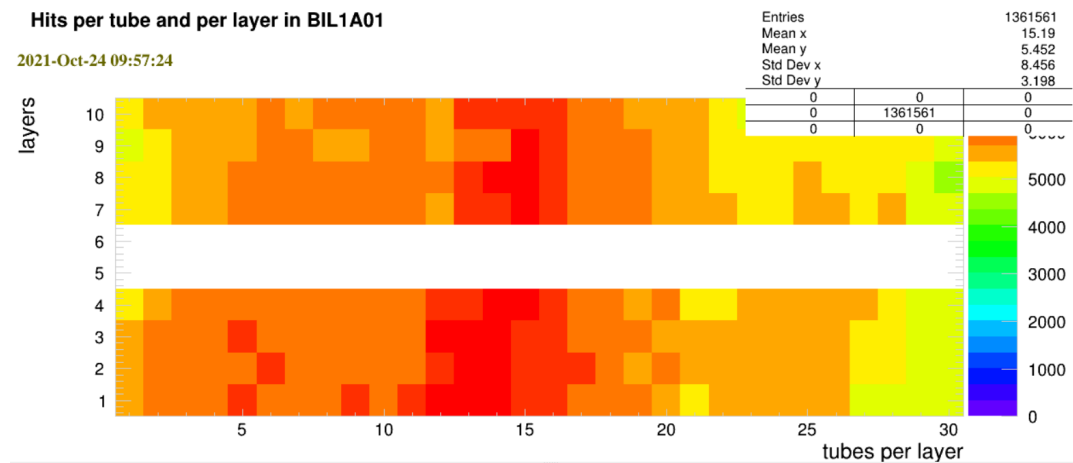


FIGURE 3.11: Three of the five histograms available in the MDT DQMF, for ML1 and for BIL1A01 chamber: a) ADC charge distribution for ML1, b) TDC charge distribution for ML1, c) Hits per even vs LB, Run 2.



(a)



(b)

FIGURE 3.12: The two other histograms available in the MDT DQMF, for BIL1A01 chamber: a) Hits per tube and per layer, b) Error rate vs LB, Run 2.

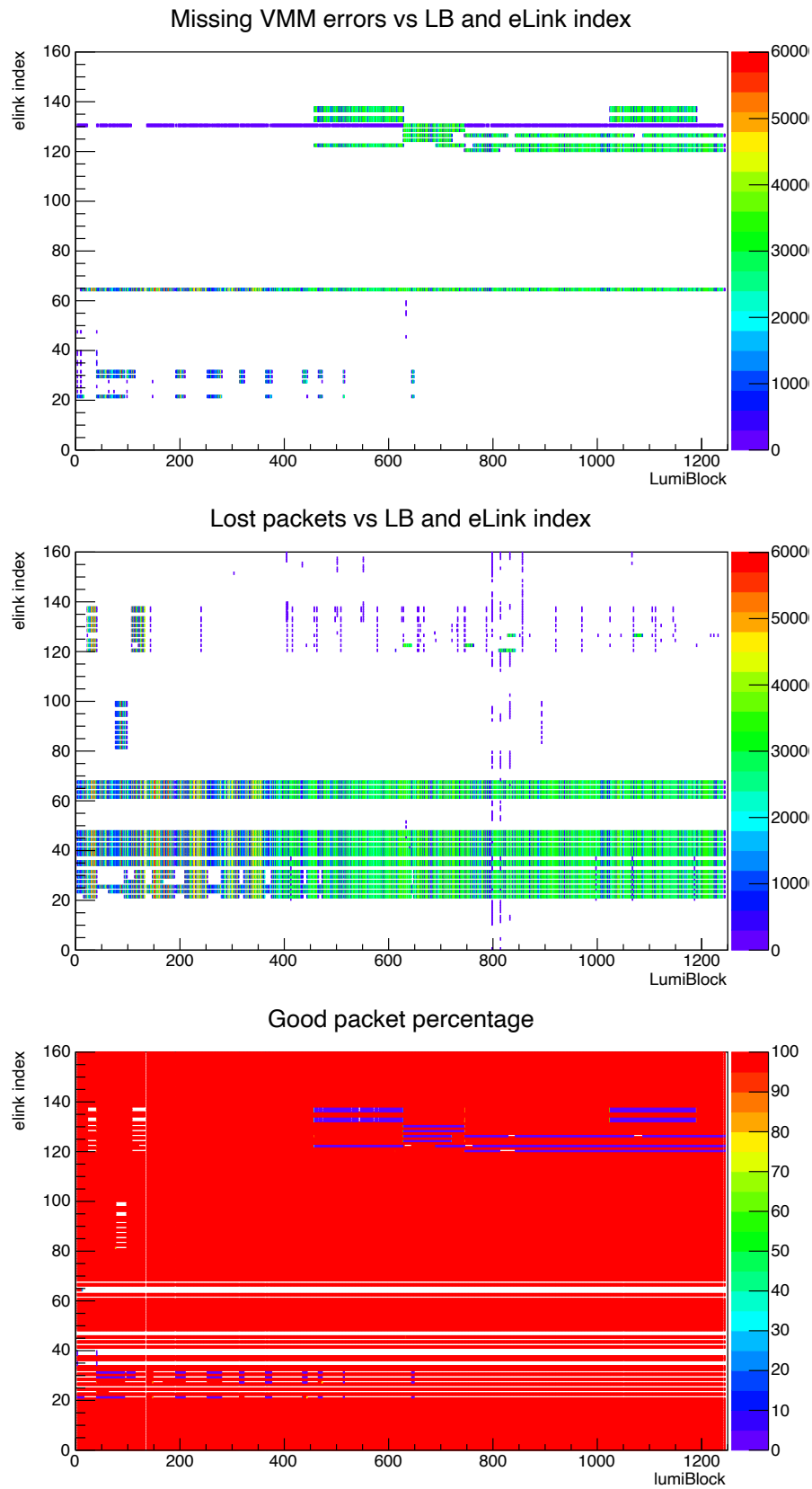


FIGURE 3.13: MMG Missing VMM, Lost Elinks and Good Packet Percentage VS LB per sector A07, run 460348.

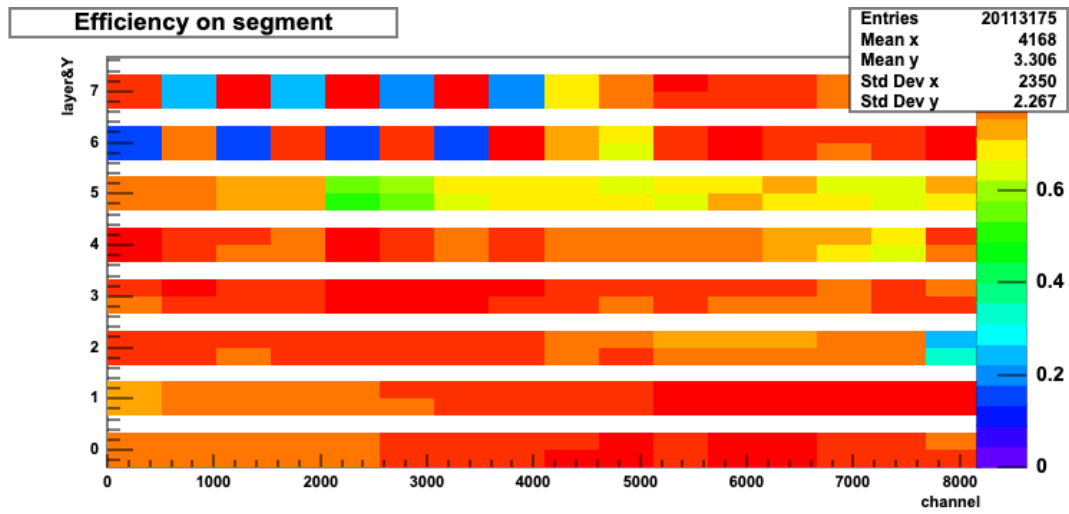


FIGURE 3.14: MMG efficiency on segment, Run 3.

Cluster charge layer 1, region a-

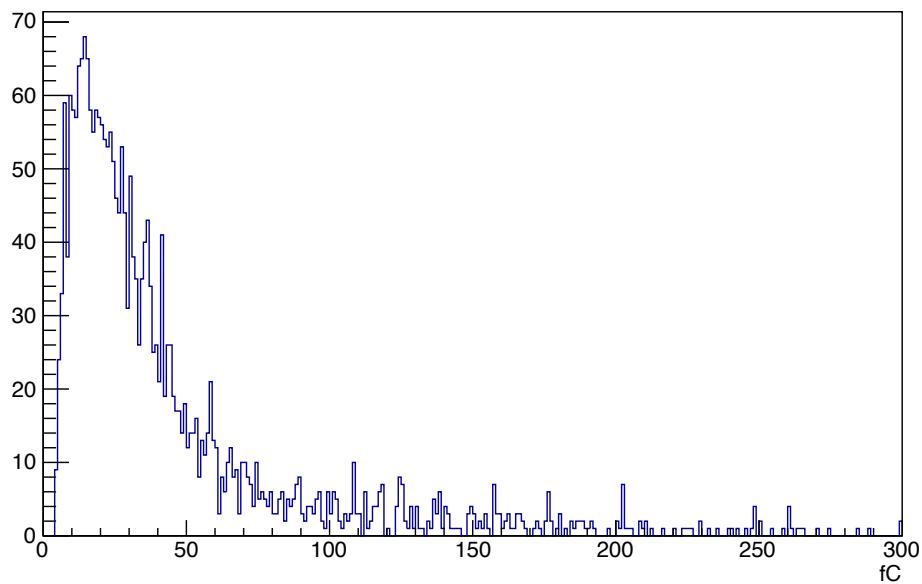


FIGURE 3.15: The Landau distribution for the Layer 1, a- HV region of sector A07, run 460348.

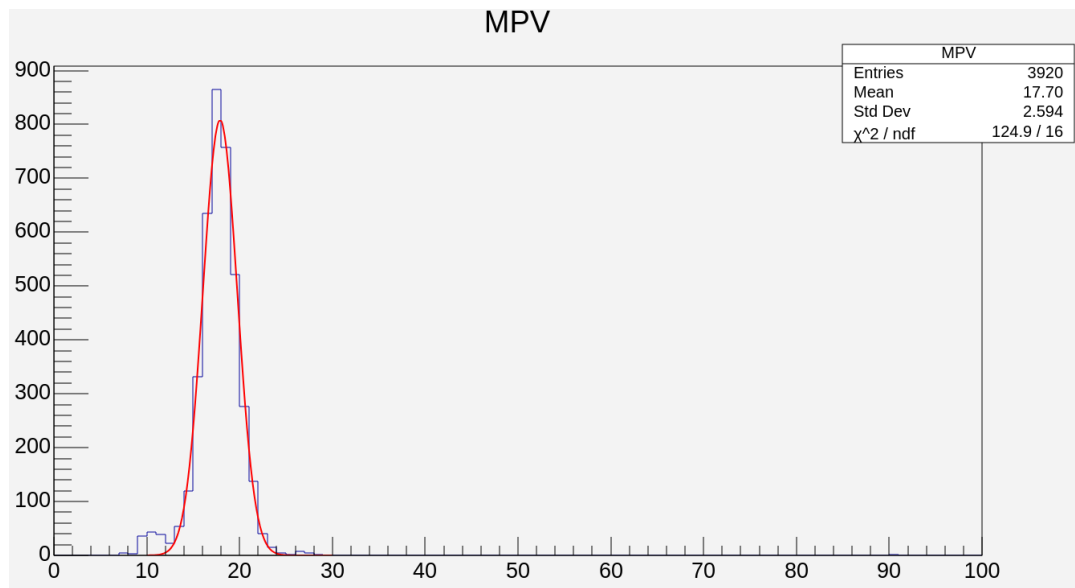


FIGURE 3.16: The most probable values (MPVs) obtained from Landau fits on Landau distributions of run r440199 [49].

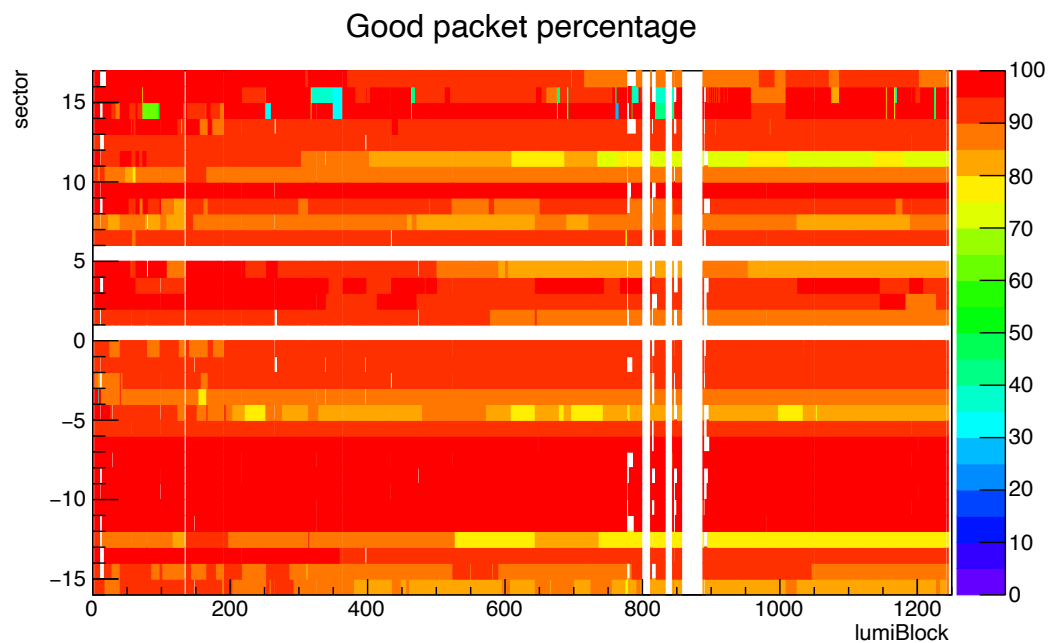


FIGURE 3.17: MMG Good Packet Percentage VS LB global, run 460348.

- **CountThreshold:** it counts how many e-links are considered *bad* for the RateThreshold. If they are less than the 15% the histogram is flagged green, if they are more than the 20% red, orange in between.

In addition, the *GoodPacksVsLB* algorithm does not consider LBs empty for all the e-links: sometimes, due to the hold of the trigger, some LBs are completely empty but they are not flag of detector problems. Another implementation, which is still under development, is about the disabled e-links. Some e-links, in fact, are not working, so they are disabled from the beginning. It is useful if the algorithm already knows it and does not consider them in the algorithm computation.

The **Missing VMM vs LB** histogram is filled to the corresponding e-link for each event with has one missing VMM in function of the LBs. VMM are front-end chips places on the readout boards on the MMG. Ideally the histogram should be empty. The algorithm implemented for checking this histogram in the DQMD is *ErrorsVsLB*, very similar to the *GoodPacksVsLB* and consists of one parameter and two thresholds:

- **Parameter:**
 - **nLBChecked:** currently set at 5 LB. It corresponds to the interval of LBs checked by the algorithm. It is also the minimum value of LBs at the beginning of the run for starting the algorithm checking. Before that value the histogram is flagged grey.
- **Thresholds:**
 - **RateThreshold:** for each e-link it is tested if the percent is below the RateThreshold. The value is averaged on the number of LBs checked. One e-link is considered green if the value is below 15%, red above 20%, orange in between;
 - **CountThreshold:** it counts how many e-links are considered *bad* for the RateThreshold. If they are less than the 15% the histogram is flagged green, if they are more than the 20% red, orange in between.

The **Lost Elinks vs LB** plot shows the lost e-links in function of time. The algorithm applied is *ErrorsVsLB* also here with the same parameter and threshold values as for the missing VMM vs LB one.

The **Event** folder includes the **Efficiency On Segment** histogram (Figure 3.14), which shows the efficiency calculated on segments per layer vs the channels of the sectors. The algorithm implemented is the *EffiOnSegment* and has one parameter and two thresholds:

- **Parameter:**
 - **MinStat:** currently set at 1000 events. This is the minimum number of events which needs to be integrated for checking the histogram; if the number of events are less than MinStat, then the histogram is flagged grey.
- **Thresholds:**
 - **LayerThreshold:** for each layer it is counted the percent of channels with efficiency less than 0.9. One layer is considered green if the percent of *bad* channels is below 5%, red above 15%, orange in between;

- **CountThreshold:** it counts how many layers are considered *bad* for the LayerThreshold. If they are less than the 5% the histogram is flagged green, if they are more than the 10% red, orange in between.

The *EffOnSegment* algorithm is still in commissioning phase, in particular we need to properly tune the thresholds. Therefore, the algorithm is applied but the result is not propagated to the region where it is contained.

In the **Layers** folder there are 8 sub-folder for each sector. In each layer, there are 16 histograms which show the charge distribution for each HV channel, an example is reported in Figure 3.15. The algorithm is the *LandauFit_MMG* and computes a Landau fit on the distribution and check if the most probable value (MPV) obtained falls in a defined range of values. The range of the MPVs are tuned considering the landau distributions for all the sectors and layers for the run 440199, on which the landau template is fitted and the MPV obtained. In Figure 3.16 the distribution of the MPVs obtained from the Landau fit of run 440199 is shown. Through Gaussian fit, the mean and the sigma of 18 and 3 respectively are obtained. The algorithm has 3 parameters and 1 threshold:

- **Parameters:**

- **MinStat:** currently set at 1000 events. This is the minimum number of events for checking the histogram, if the number of events are less than MinStat, then the histogram is flagged grey;
- **Mean:** is the mean of the MPV values, currently set at 18;
- **Sigma:** is the standard deviation of the MPV distribution, currently set at 3.

- **Thresholds:**

- **Multi-sigma:** is the number of sigma within which the MPV has to be. If the MPV is less than 3 sigma the histogram is flagged green, if it is above 5 is flagged red, orange in between.

In the MMG DQ check is added also a global tab with the **Good Packet Percentage** histograms for all the sectors (Figure 3.17). The algorithm used is the same as for the per-sector one, with only one modification. In the global histogram, if one sector is empty, the histogram is immediately flagged red, since it is a severe problem.

The Muon spectrometer DQ flag is the result of the worse case of the sub-detector flag: if, for example, the MMG is red and STG, MDT, TGC and RPC are green, the muon spectrometer becomes red, because it takes the worse case. The same is for the MMG, if at least one of its folder is red also MMG becomes red. For now we are propagating in the MMG only the flag of the Good Packet Percentage plot. The reason is that the plot is significant: if there is a problem in the detector is easy to spot from it. The other histograms that I have explained are also checked, but the resulting flag is not propagated up to the MMG detector level.

3.3.2 Muon Spectrometer OHP

The Online Histogram Presenter can be used to display histograms to be monitored, without any algorithm applied. The tool is very useful for shifters in order to have a quick overview of any sub-detectors.

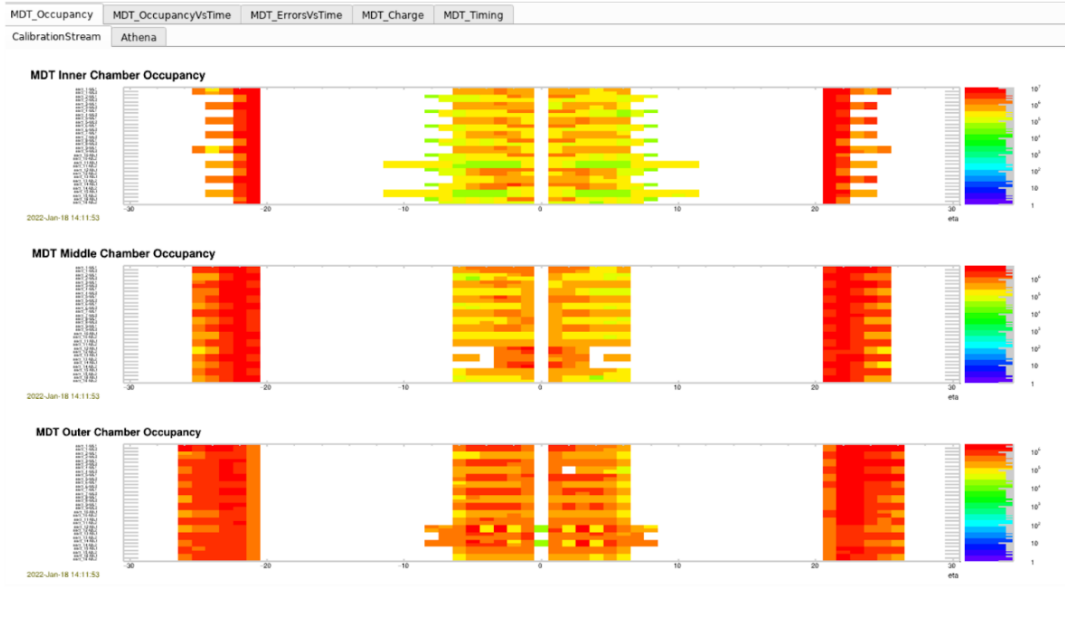


FIGURE 3.18: MDT_Occupancy plots for Calibration Stream, run 430036.

Muon spectrometer OHP has a main configuration file *muon.ohp.xml* which includes the configuration files of every plugin.

The tool provides two different utilities: the Browser, which allows direct access to any histograms published in OHS, and the individual sub-detector tab.

MDT OHP

The MDT tab is divided in 5 sections: MDT_Occupancy, MDT_OccupancyVsTime, MDT_ErrorsVsTime, MDT_Charge, MDT_Timing.

- **MDT_Occupancy** consists of two tabs containing histograms showing the accumulated occupancy per chamber: **Calibration Stream** (Figure 3.18) shows the accumulated occupancy for each chamber (3 histograms, one for inner layer, one for middle and one for the outer layer) and **Athena** (Figure 3.19) shows the accumulated occupancy for each chamber (2 histograms, one for barrel and one for end cap chambers). These histograms are filled by Athena global monitoring after full event reconstruction;
- **MDT_OccupancyVsTime** includes five tabs. One is for the overview and four tabs (one per detector region) containing histograms showing the occupancy per chamber as a function of time. For each ROD crate 2 histograms are displayed, which are based on different data sources (GNAM and Athena);
- **MDT_ErrorsVsTime** presents four tabs (one per detector region) containing histograms showing errors reported by GNAM;
- **MDT_Charge**: shows the ADC distribution for all chambers connected to a given ROD crate (16 histograms). These histograms are filled by GNAM running on the calibration stream data. A reference is shown in blue on top of the histogram. In Figure 3.20 the histogram is presented;

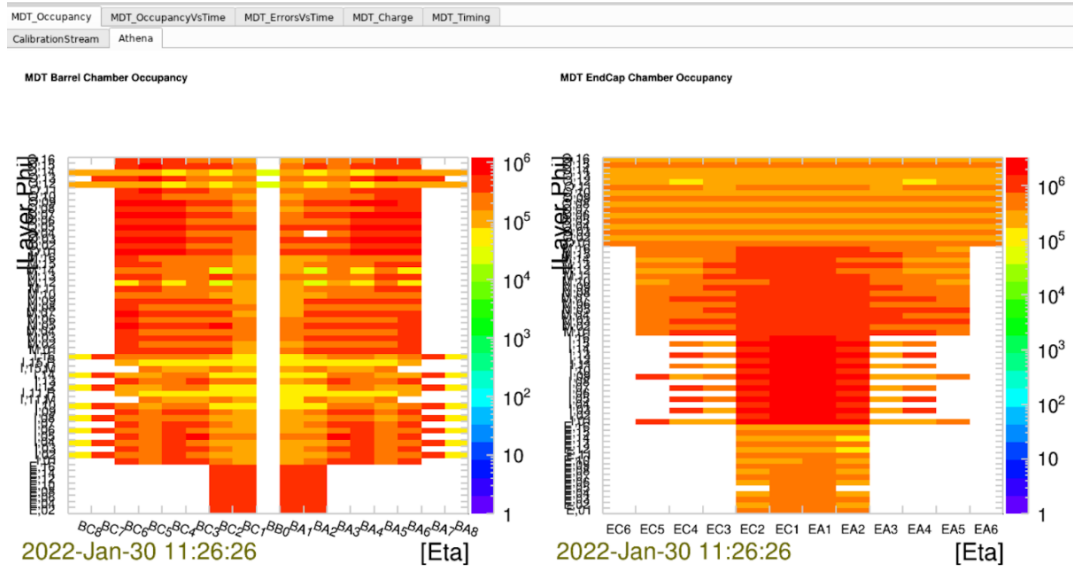


FIGURE 3.19: MDT_Occupancy plots for Athena, run 437079.

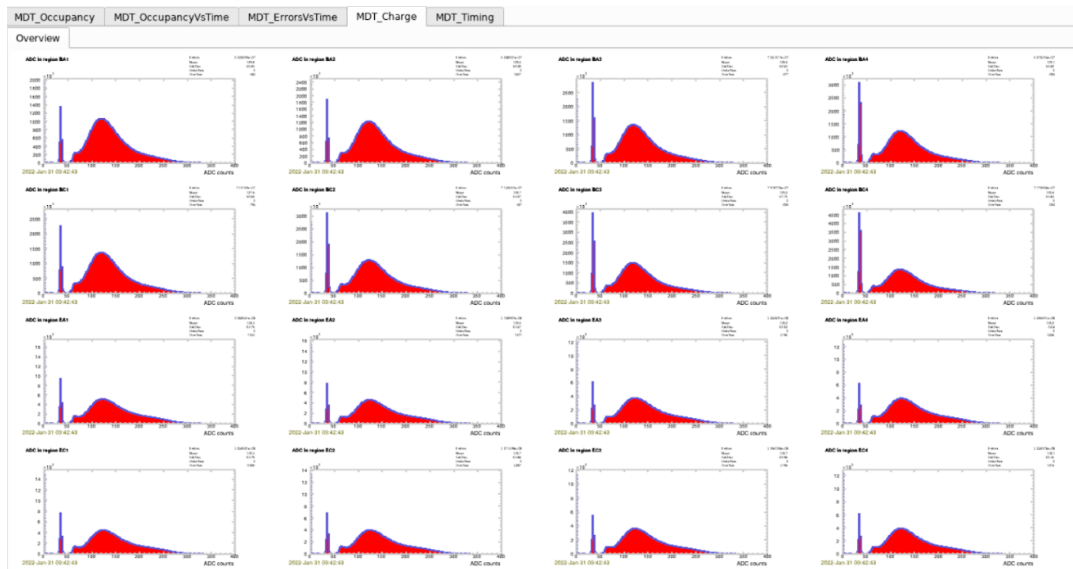


FIGURE 3.20: MDT_Charge plots, Run 3.

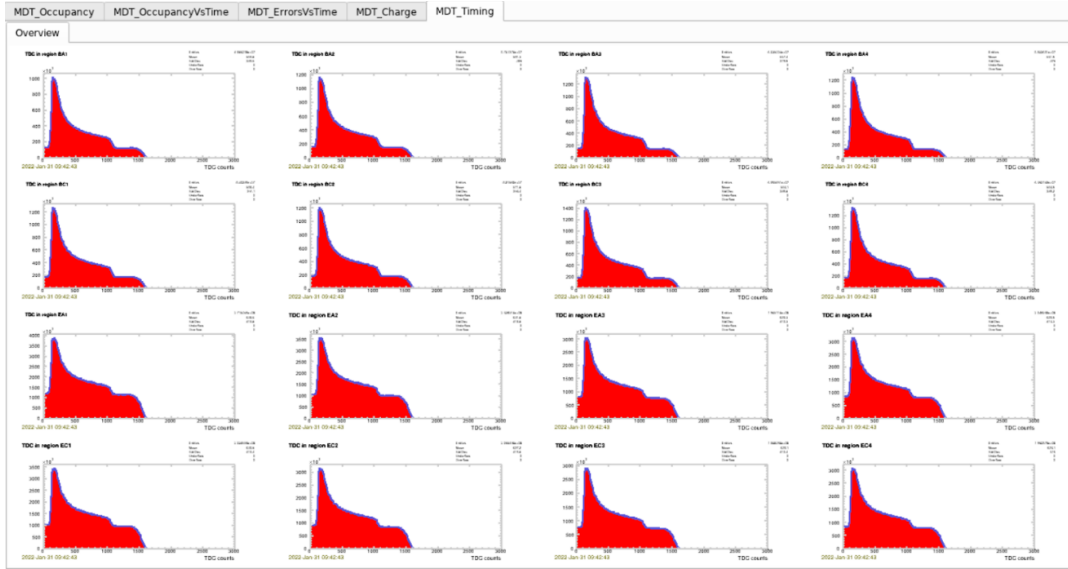


FIGURE 3.21: MDT_Timing plots, Run3.

- **MDT_Timing:** shows the TDC distribution for all chambers connected to a given ROD crate (16 histograms). These histograms are filled by GNAM running on the calibration stream. A reference is shown in blue on top of the histogram. In Figure 3.21 an example distribution is shown.

MMG OHP

Similarly for the DMQF, the MMG OHP includes histograms which we chose in coordination with the MMG and STG detector experts. For this reason, I will go into details in the histogram presentation.

The MMG tab is divided in 4 sections [50]: MMG_Occupancy, MMG_Occupancy_VS_LB, MMG_Errors, MMG_Efficiency.

- **MMG_Occupancy:**
 - **Overview:** in this section global plots for all sectors are presented. In particular, for all sectors, occupancy distribution plots are shown for segments, packets, clusters and the time of clusters on segments. Sectors are numbered from 0 to 15, for convention positive numbers are side A, negative side C sectors. The plots are expected to be flat, with all bin filled, see Figure 3.22.
 - * **Packet Occupancy:** packets are electronic read-out blocks from each link for every hit;
 - * **Cluster Occupancy:** a cluster is the charge collection of contiguous strips of read-out;
 - * **Segment Occupancy:** segments are reconstructed tracks (within a chamber), using a linear fit on charge collected by contiguous strips (cluster);
 - * **Time of Cluster on segment:** the plot shows the peaking time of a cluster on a segment. The plot is expected to have entries among -25 to 175 ns, with all bin filled;

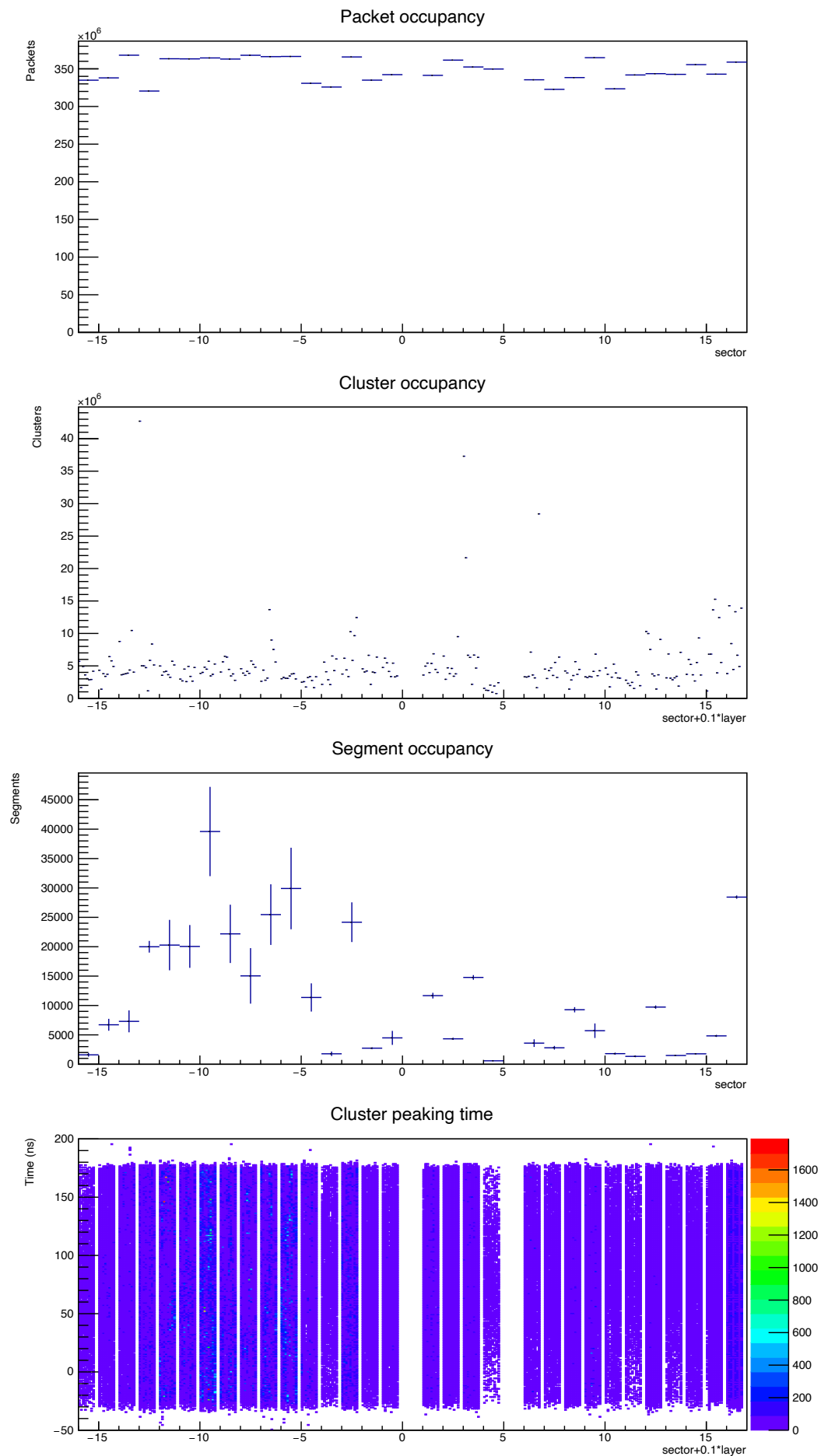


FIGURE 3.22: MMG packet, cluster, segment occupancy and the time cluster on segment in function of sector number, run 460348.

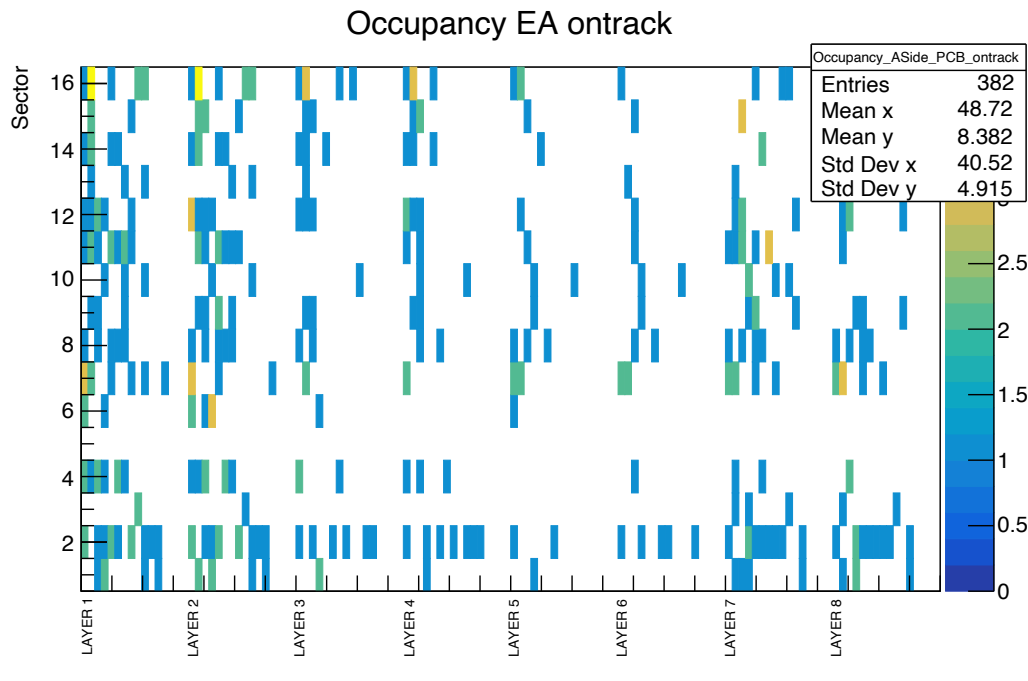


FIGURE 3.23: MMG Athena occupancy side A, run 460348.

- **Athena:** shows for all sectors, for each side, the occupancy on track reconstructed by Athena. The 2D plot shows the occupancy per sector as function of layers. An example in Figure 3.23;
- **MMG_Occupancy_VS_LB:**
 - **Overview:** in this section global plots for all sectors are presented. In particular, for all sectors, occupancy distribution in function of LB plots are shown for segments, packets, channels and clusters. Sectors are numbered from 1 to 16, for convention positive numbers are side A, negative side C sectors, see Figure 3.24.
 - * **Packet Occupancy VS LB;**
 - * **Channel Occupancy VS LB;**
 - * **Cluster Occupancy VS LB;**
 - * **Segment Occupancy VS LB;**
 - **Athena:** in this section, global plots for all sectors are presented, Figure 3.25. In particular, for all sectors, occupancy distribution in function of LB plots are shown for segments, packets, channels and clusters. Sectors are numbered from 1 to 16, for convention positive numbers are side A, negative side C sectors. These plots are selected from the Global Monitoring after Athena reconstruction and show the layer occupancy for each sector in function of LB. Each layer has 8 PCB and 16 FEB;
- **MMG_Errors:**
 - **Overview:**
 - * **Missing VMM VS LB:** VMM are front-end chips placed on readout boards on the MM. In an ideal data taking conditions no VMM is missing;

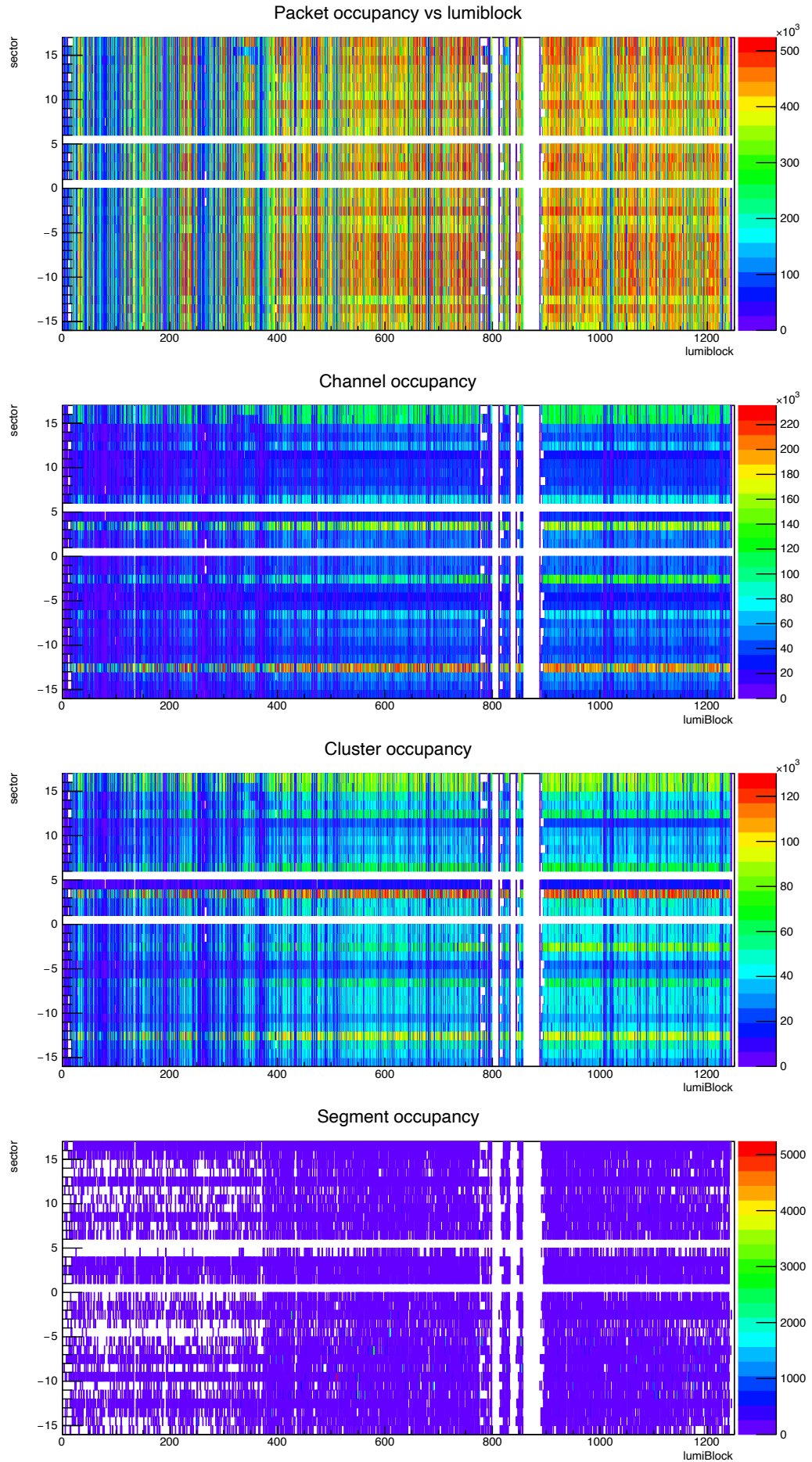


FIGURE 3.24: MMG packet, channel, cluster and segment occupancy VS LB, run 460348.

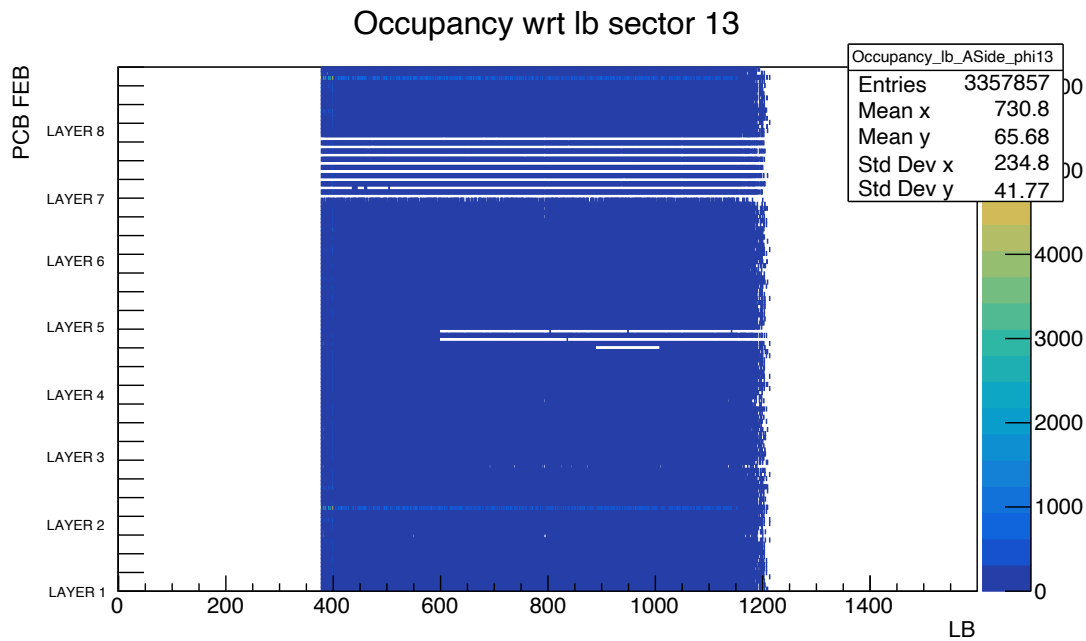


FIGURE 3.25: MMG Athena occupancy VS LB for sector 13, run 460348.

- * **Lost Elinks VS LB:** e-links per sector are 160 and correspond to data packets, if e-links are missing for a specific sector, data are lost. In a perfect data taking condition, the plot should be empty;
- * **Good Package Percentage VS LB:** this plot shows the percentage of sectors vs luminosity block. All packets should be at 100%, see Figure 3.17;
- **Good Packet Percentage VS LB side A and C:** This plot shows the percentage of good e-links (packets) vs luminosity block for each sector;
- **MMG_Efficiency:** the efficiency on segment reconstruction is defined as the number of hits in a channel divided by the number of total hits per PCB. Every PCB has 1024 channels and every layer has 8 PCB, then in total 8192 channels per layer. The histogram represents the efficiency per layer vs channels per sector. Low efficiency for few channels are acceptable. The efficiency plot is published only if a certain number of layers of a sector have reconstructed a segment. In Figure 3.14 reports an example of the histogram.

Chapter 4

Flavour tagging in ATLAS and new techniques

In this chapter a new tagger for $Hb\bar{b}$ identification is presented, exploiting some variables sensitive to the color of the mother particle of two the b quarks. This an important part of my PhD work, which converged in a paper [51], in collaboration with theoretical and ATLAS experimental physicists.

Before going into details of the tagger in section 4.2, an overview of the general flavour tagging in ATLAS is reported in 4.1, also showing which $Hb\bar{b}$ tagger the collaboration is using now.

4.1 Flavour tagging in ATLAS

The flavour tagging is the capability to distinguish if a jet is originated from b , c quarks instead of coming from *light* (neither b or c) quarks. Jets are identified as b -jet, c -jet and *light*-jets correspondingly. Flavour tagging has been decisive in ATLAS strategy for the observations of Higgs boson decay into bottom quarks [53] and its production in association with a top-quark pair [54]. Moreover, it is crucial for many Standard Model precision measurements and searches for new phenomena.

ATLAS uses various algorithms to identify b - and c -jets, called flavour-tagging algorithms, based on several objects reconstructed as key ingredients. The most important objects are the charge particle tracks, the primary vertices (PVs) and the hadronic jets. A short description will be provided here, while more information can be found at this reference [55] and here [56].

Charge-particle tracks are reconstructed in the ID (see Section 2.3.2). For the flavour tagging tracks with $p_T > 500$ MeV are used, with further selections for rejecting fake and poorly measured tracks. The total tracking efficiency for charged pions of $p_T > 4$ GeV ranges from 90% for $|\eta| < 1$ to 70% in the forward region, $2.3 < |\eta| < 2.5$.

The **primary vertex** (PV) is particularly important for the b -tagging, since it defines the reference point from which track and vertex displacements are computed. Flavour tagging requires at least one PV in each event. The *hard-scatter vertex* or primary interaction point is the PV with the highest sum of the squared transverse momenta of the selected tracks. Transverse parameter d_0 is the distance of closest approach to the track to the PV in the transverse plane, the longitudinal parameter z_0 is the longitudinal separation between the PV and the point of the track where d_0 is measured, referred as *perigee*.

Figure 4.1 shows a schematic diagram of the b -jet production with two light jets, shown in the transverse plane.

The **hadron jets** are particle-flow objects which are constructed in ATLAS ID and

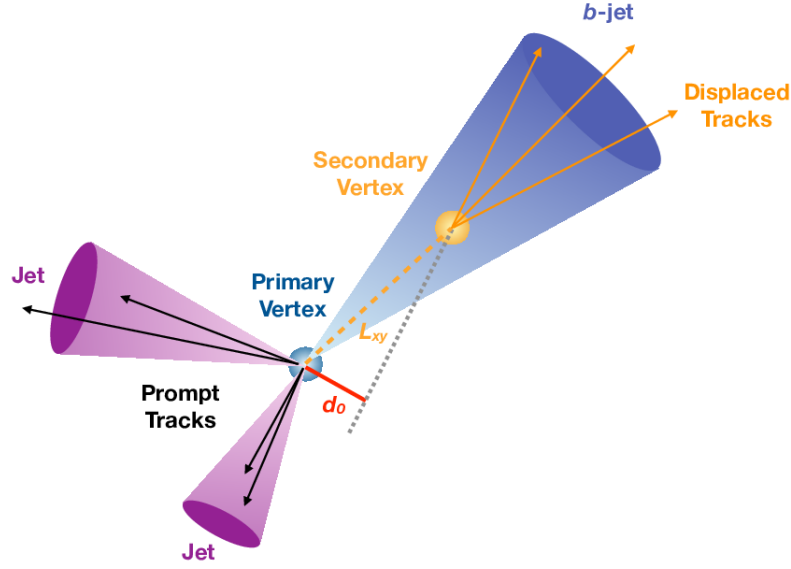


FIGURE 4.1: A schematic representation of a *light* and *b* jet production. The *b* jet is the blue cone one, and presents a lifetime of the *b*-hadron which corresponds to a transverse decay length L_{xy} and produces displaced tracks from the secondary vertex. The distance d_0 is the transverse impact parameter, measured from the primary to the secondary vertex and it is typically large for *b*-jets [52].

calorimeters, using the anti-kt algorithm [15] with a radius parameter R of 0.4, implemented in Fastjet [57]. Corrections are also applied on jet in order to consider the correspondence with data and the particle level jet. In order to reduce the number of jets with large energy fractions from pile-up collision vertices, the *jet vertex tagger* (JVT) algorithm is used [58]. Tracks are matched with jets using the angular separation ΔR between the track momenta at the perigee and the jet axis. The angular requirement varies as a function of the jet p_T , since the decay products of the *b*-hadrons are more collimated for higher p_T . If more than one jet fulfill the requirements, the closest one is chosen.

The **τ lepton** is the heaviest lepton known in the SM, with a life time of $2.9 \cdot 10^{-13}$ s, and can decays both leptonically and hadronically. The hadron decay is the most probable, which contains one or three charged pions and at least one neutral pion, in the 68% of cases. The major background is composed by energetic jets, produced via fragmentation of quarks and gluons. The discrimination is performed exploiting the different narrow shower formed in the calorimeter, the track multiplicity and the displaced decay vertex.

The *true* flavour of a jet in a simulation is selected correspondingly to the hadron contained in it. The jet is labeled as *b*-jet if contains at least one weakly decaying *b*-hadron of $p_T \geq 5$ GeV within a cone of ΔR 0.3 around the jet axis. If no *b*-hadrons are found, *c*-hadrons or τ -leptons are searched with the same criteria. The remaining jets are labelled as *light*-flavour jets. If more the one heavy flavour hadron is found inside a jet, e.g., after gluon splitting, the procedure follows the same steps as above

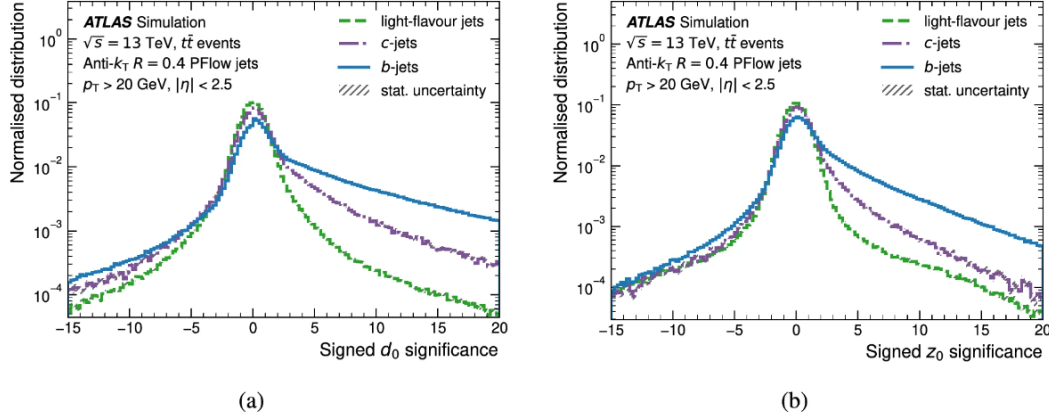


FIGURE 4.2: In Figure (a) the transverse and in (b) the longitudinal impact parameter significance of tracks for b -, c - and $light$ - jets in $t\bar{t}$ events [59].

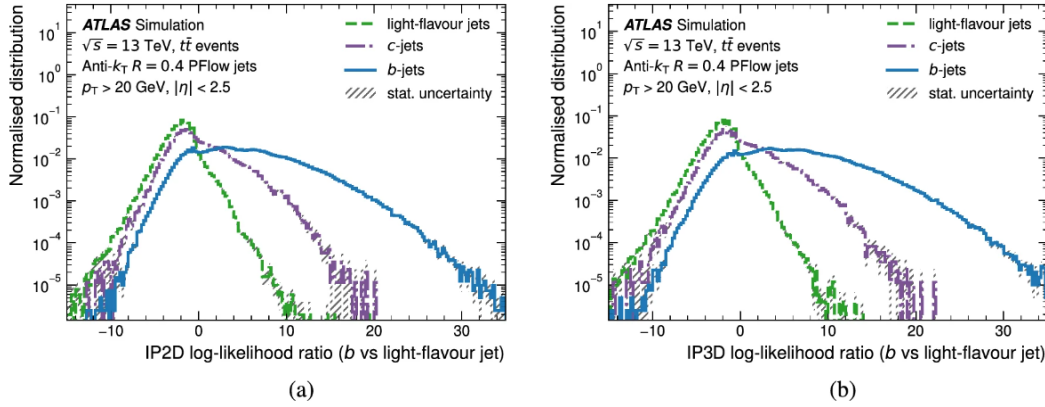


FIGURE 4.3: In Figure, the log-likelihood ratio for (a) IP2D and for (b) IP3D b -tagging algorithms for b -jets, c -jets and $light$ - jets in $t\bar{t}$ events. The log-likelihood ratio shown is computed as the ratio of the b -jets to $light$ -flavour jet hypothesis probabilities. [59].

and $b\bar{b}$ jets will receive a b label, the same for c -jets.

The ATLAS Collaboration uses two stages for the flavour tagging. The first is based on **low-level algorithms** using the proprieties of tracks or the reconstructed displaced vertices. The second is the combination of the results of the low-level with **high-level algorithms**, based on multivariate classifiers.

4.1.1 Low-level b -taggers

Algorithms based on impact parameters

The first low-level algorithms are those based on exploiting the large impact parameters of the tracks originating from the b -hadron decays. The transverse impact parameter d_0 is defined as the minimum distance in the r - ϕ plane between the track and the primary vertex. The algorithms used in ATLAS are the IP2D and IP3D [60]. The former uses the signed transverse impact parameter significance of tracks to

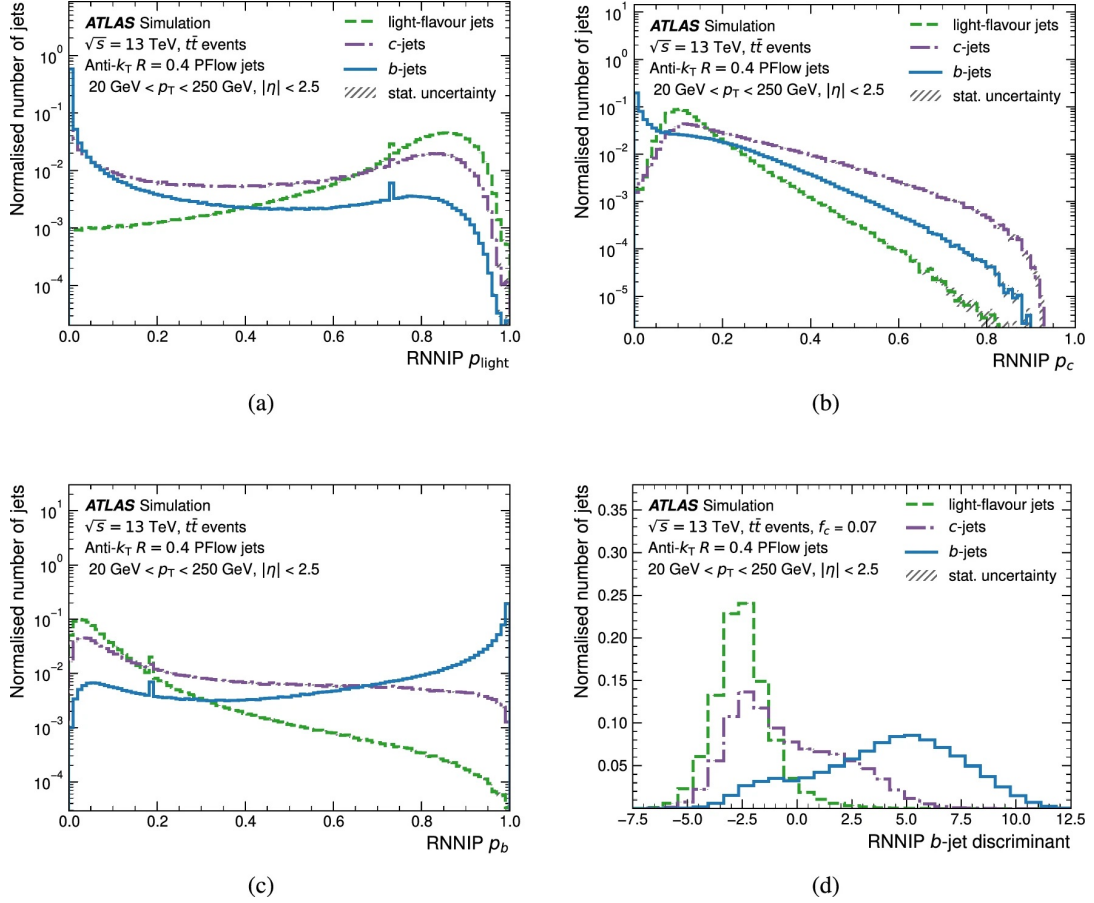


FIGURE 4.4: In Figure, distributions of the outputs of the RNNIP b -tagging algorithm for b -jets, c -jets and light-flavour jets in the baseline $t\bar{t}$ simulated events: (a) p_{light} , (b) p_c , (c) p_b and (d) the RNNIP discriminant [56]. The spikes are originated from jets with no tracks.

construct the discriminant variable, while the latter uses both longitudinal and transverse impact parameter significance in a two dimensional template to account for their correlations. The significance S_i is defined as the ratio of the track impact parameter on its uncertainty. The signed transverse and longitudinal impact parameter significance are shown in Figure 4.2.

The discriminant log-likelihood distribution ratio (LLR) is defined as the sum of the logarithms of the per-track probability ratios for each jet-flavour hypothesis. The probabilities are calculated from probability density functions (pdfs) derived from MC simulation. The log-likelihood ratios separating b -jets from light-jets for the IP2D and IP3D algorithms are shown in Figure 4.3.

Two extra LLR are computed for both IP2D and IP3D to separate b -jets from c -jets and c -jets from light-flavour jets. These LLRs are used as inputs for the high-level taggers.

Track-based recurrent neural network tagger

Very lately, ATLAS improved the flavour-tagging procedure by introducing a new algorithm, based on the Recurrent Neural Network (RNN) [59, 60]. In the track-based recurrent neural network tagger, the sequence dependencies between variables in sequence of arbitrary length is learnt by the network [61], as the case of the track

jets. RNN, in fact, can easily overcome the IP-based tagger assumption of each track independence from all the other tracks in the same jet, which does not correctly describe its proprieties.

The RNN uses as input each selected track variables. They are the lifetime-correlated signed transverse (S_{d_0}) and longitudinal (S_{z_0}) impact parameter significances, the angular distance between the track and the jet-axis ($\Delta R(\text{track}, \text{jet-axis})$), the fraction of transverse momentum carried by the track relative to the jet p_T (p_T^{frac}) and the hit multiplicities of the tracks in the different ID layers. The tracks are ordered by the $|S_{d_0}|$ values, then passed to the neural network as a vector. The RNN output provides the b -jet, c -jet and light-flavour jet probabilities, p_b , p_c and p_{light} respectively and are used as input variables for the high-level taggers. The outputs are then combined into the b -tagging discriminant function $D_{\text{RNN}} = \ln(p_b / (f_c \cdot p_c + (1-f_c)p_{\text{light}}))$, where f_c is the c -jet fraction. In Figure 4.4 the flavour-probabilities and the RNN b -tagging algorithm for b -jets, c -jets and light-flavour jets.

Secondary-vertex-tagging algorithm

The secondary-vertex-tagging **SV1** [62] reconstructs a single displaced secondary vertex in a jet. The algorithm starts selecting any possible two-track vertices, excluding those compatible with K_S^0 or Λ decays, photon conversions or hadronic interactions with the detector material. The **SV1** algorithm run on all tracks of the selected vertices, resulting in only one fitted secondary vertex. In each iteration, the match between track and vertex is evaluated with the χ^2 test.

Some **SV1**-related variables are used as input to the high-level taggers, including the number of tracks associated with the **SV1** vertex, the invariant mass of the secondary vertex, its energy fraction and the 3D decay length significance. They are all shown in Figure 4.5.

Topological multi-vertex finding algorithm

JetFitter [63] is the topological multi-vertex finding algorithm used by ATLAS. It exploits the topological structure of the weak decays of heavy flavour hadrons inside the jet for reconstructing the full b -hadron decay chain. The algorithm starts from a modified Kalman filter [64] for finding approximately the b -hadron flight path and the vertex positions. Eight discriminating variables, including the track multiplicity at the **JetFitter** displaced vertices, the invariant mass of tracks associated with these vertices, their energy fraction and their average 3D decay length significance are used as inputs to the high-level taggers.

4.1.2 High-level flavour-taggers

The outputs of the low-level algorithms are used as input for a deep-learning classifiers. The classifiers are based on a fully connected neural-network (NN), called **DL1** algorithm series. The algorithms were introduced at the beginning of Run 2 [55] and exploits as input **IP2D**, **IP3D**, **SV1** and **JetFitter** algorithm outputs, while **DL1r** includes also the **RMNIP** output probabilities. In addition, kinematic jet variables, as jet p_T and η , are included.

The output of the **DL1r** NN corresponds to the probabilities for a jet to be a b -jet, c -jet or light-flavour jet. The output probabilities are shown in Figure 4.6. Since the flavours are treated equally during training, the trained network can be used for both

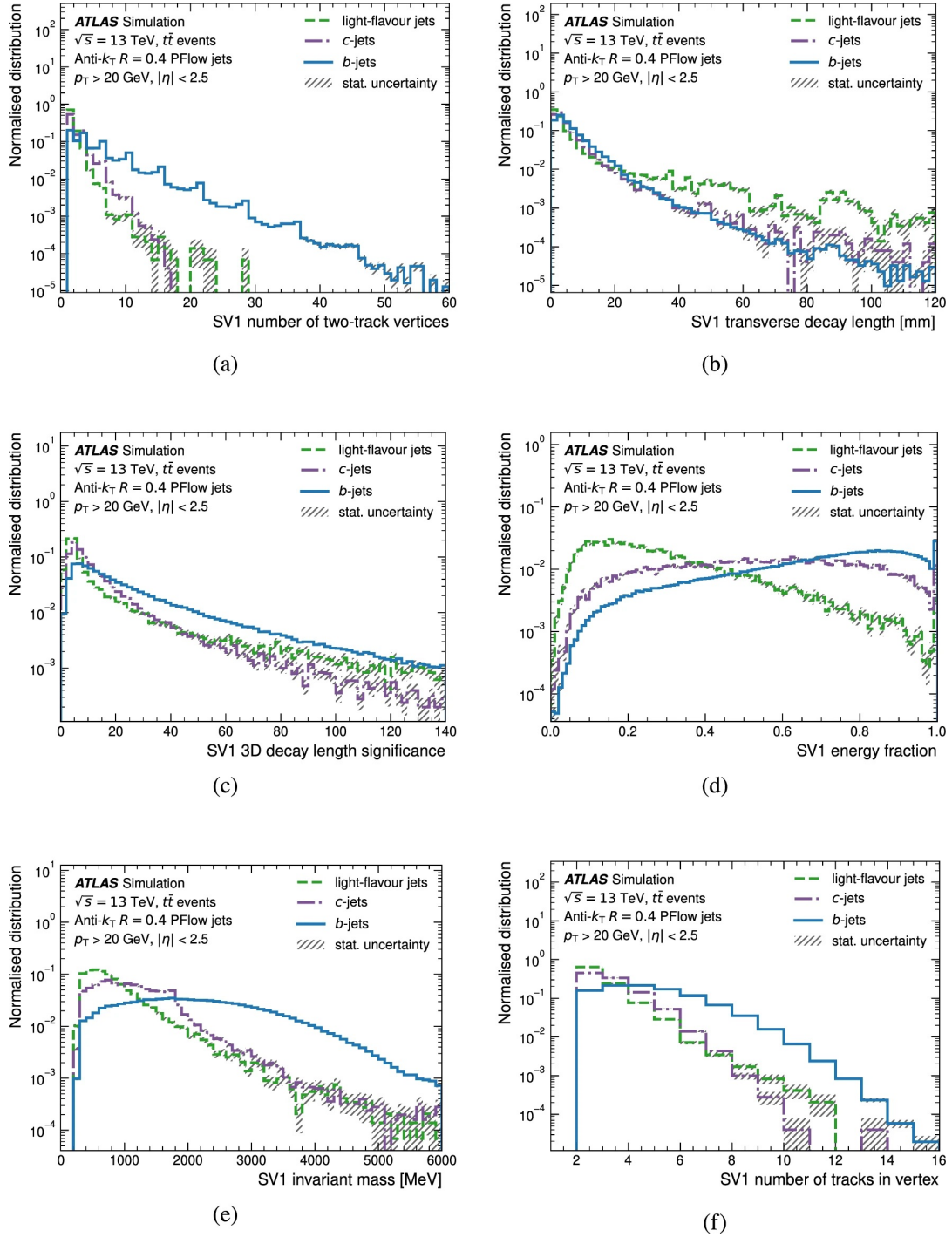


FIGURE 4.5: In Figure, proprieties of secondary vertices reconstructed by SV1 algorithm for b -jets, c -jets and light-flavour jets in the baseline $t\bar{t}$ simulated events; (a) the number of two-track vertices reconstructed within the jet, (b) the transverse decay length, (c) the 3D decay length significance of the distance between the primary and displaced vertex, (d) the energy fraction, (e) the invariant mass and (f) the number of tracks in vertex [56].

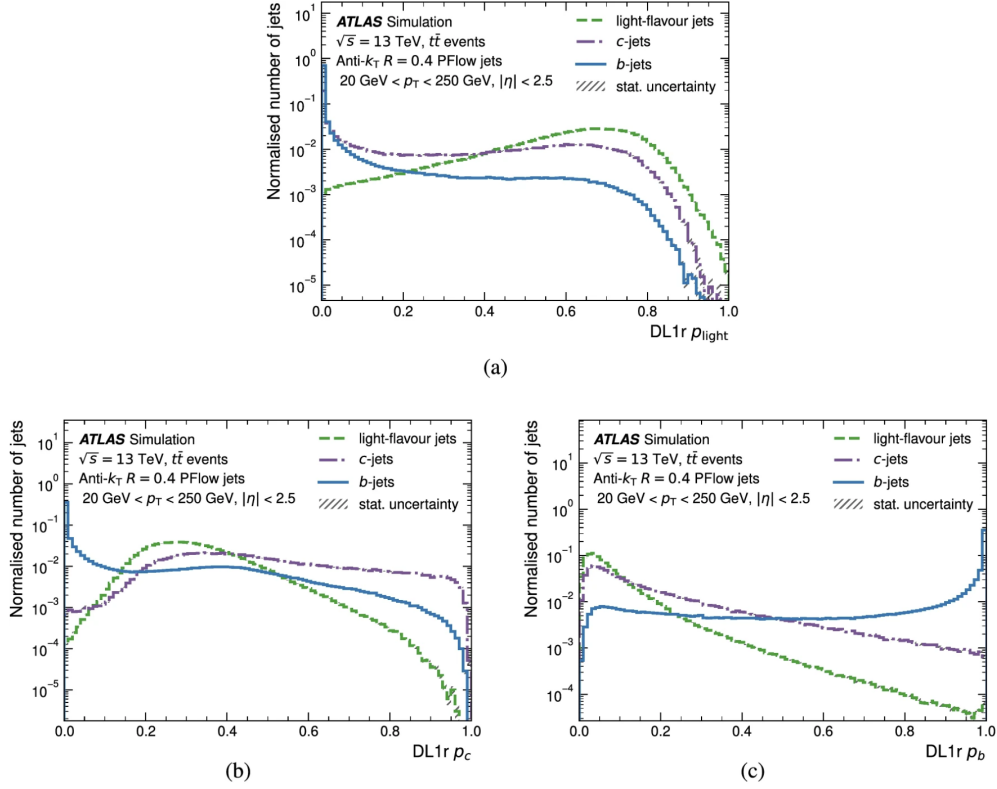


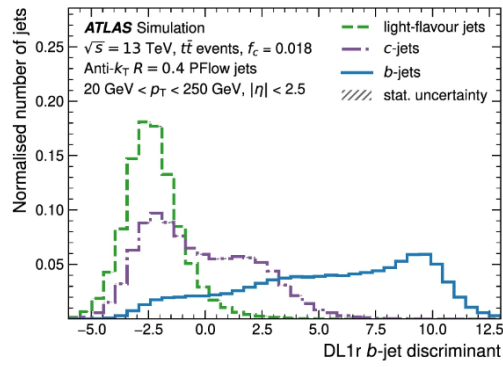
FIGURE 4.6: In Figure, the outputs of the DL1r b -tagging algorithm for b -jets, c -jets and light-flavour jets in $t\bar{t}$ simulated events [56].

b - and c -tagging. The final DL1r b -tagging discriminant variable is:

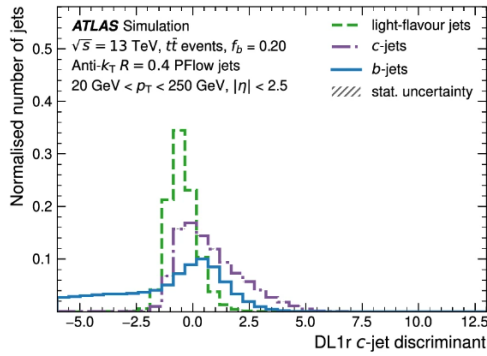
$$D_{DL1r} = \ln \left(\frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_{light}} \right) \quad (4.1)$$

where p_b , p_c and p_{light} are the correspondent flavour probabilities and f_c is the effective c -jet fraction in the background hypothesis. The fraction can be chosen a posteriori on simulated events in order to optimize the algorithm performance at physics analysis level. Similarly can be defined the DL1r c -tagging discriminant variable, swapping in the equation 4.1 the b contribution with c . The final discriminant distributions for b - and c - jets are shown in Figure 4.7.

4.1.3 Flavour-tagging performance



(a)



(c)

(b)

FIGURE 4.7: The final DL1r discriminant variables for b -jets, c -jets and light-flavour jets in $t\bar{t}$ simulated events [56].

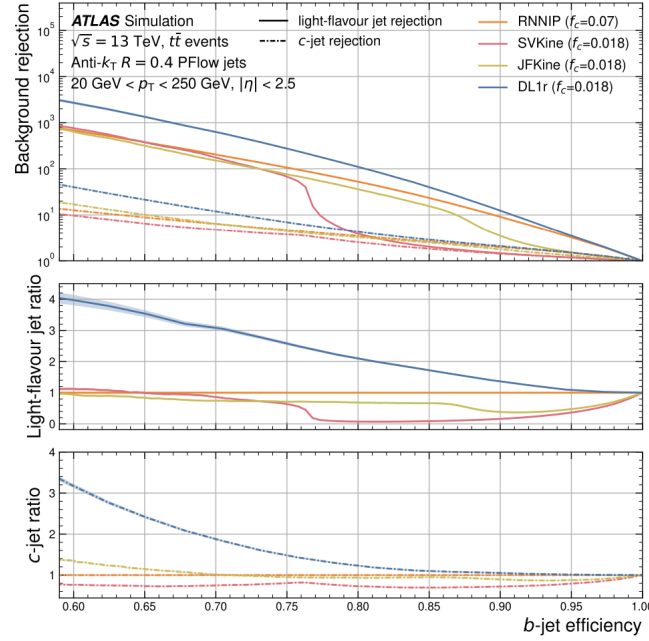


FIGURE 4.8: The light-flavour jet and c -jet rejection factors as function of ϵ_b for the high level b -tagger DL1r and low level algorithms RNNIP, SVKine and JFKine. [56].

Flavour-tagging performances are evaluated by the efficiency of tagging a signal jet, ϵ , combined with the probability of a wrong identification of a background jet, referred as mis-tag rate. The evaluation of the performances of each algorithm is based on the b -jet tagging single-cut operating points (OPs). They consist of different cuts on the discriminant variable of the algorithm correspondingly to different b -tagging efficiencies ϵ_b in the $t\bar{t}$ sample used to train DL1r. The discriminant distributions are also divided in five pseudo-continuous bins (PCBT), delimited by the selections used for the OPs for [100%,85%], [85%,77%], [77%,70%], [70%,60%] and [60%,0%] efficiency.

Figure 4.8 and 4.9 show the light and c rejection factors as function of the b -jet efficiency for several low- and high-level b -taggers. SVKine and JFKine are simple feed-forward NNs trained with, respectively, the SV1 and JetFitter outputs and some kinematic variables, defined only to illustrate the comparison with the other algorithms. Among the low-level taggers, RNNIP provides a good background rejection at high efficiency, while SVKine performs better at low efficiencies. It is evident that the DL1r algorithm performs better than the other algorithms.

Concerning the charm tagging, since the charmed hadrons have a shorter life time and smaller mass than the b -hadrons, the efficiencies of the OPs tends to be much lower.

4.1.4 Double b -tagging in boosted Higgs $b\bar{b}$ decay

The identification of the Higgs boson in its $b\bar{b}$ quark decay plays an important role in the ATLAS physics program. This decay, in fact, is used to probe the Higgs coupling with fermions and possesses the highest branching ratio in the low-medium mass range. In order to improve the sensitivity on the channel decay measurement and then, more in general, on the total Higgs cross section, flavour tagging algorithms are

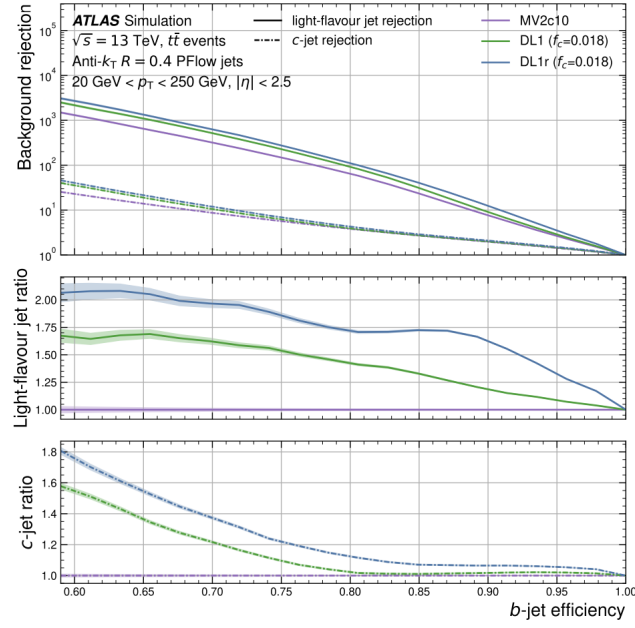


FIGURE 4.9: The light-flavour jet and c -jet rejection factors as function of ϵ_b for the high level b -taggers MV2c10, DL1 and DL1r [56].

exploited.

In ATLAS usually the Higgs $b\bar{b}$ boosted decay is identified with a common recipe. For each candidate, a large- R jet of $R = 1.0$ is reconstructed in order to include all the decay products, furthermore two smaller radius jets, sub-jets, of $R = 0.2$ inside the large- R jet, are used to identify the individual b -hadrons.

As reported in this reference [65], the general application of the b -tagging is extended at the large- R level, defining a tagging algorithm optimized specifically for the Higgs in two b -quarks boosted decay process. In the boosted regime, in fact, the sub-jets are very close to each other and they can be reconstructed as a single large- R jet. In this case, variable-radius (VR) track-jets are considered. In this view, a neural network is built counting on the strong tagging performance of the individual sub-jets and their discriminant correlations, and the large- R kinematic. The flavour tagging information are provided by the DL1r algorithm by the three outputs corresponding to the jet probability of being a b -, c -, or light-jet. Discriminant values for up to three leading VR track jets are passed to the double b -tagging algorithm. In addition the p_T and η of the large- R jets are used. In summary, up to eleven input variables for each jet are given to the NN. The model used for the double b -tagging is a feed-forward neural network, with six fully connected 250 dimensional hidden layers and a rectified linear unit activation functions.

The algorithm performance is tested in this reference [65] using two simulated samples of backgrounds: jets coming from multijet interactions and jets from processes resulting in high- p_T top quarks through $Z' \rightarrow t\bar{t}$ production. The discriminant variable D_{Xbb} is shown in Figure 4.10. The discriminant depends on the f_{top} factor, which is the fraction of the top background, set here at 0.25.

For the full range of signal efficiencies, D_{Xbb} achieves an equal or higher multijet and top jet background rejection if compared with an DL1r. The improvements are more significant when jet p_T increases.

The multijet and top jet rejection as a function of the $H \rightarrow b\bar{b}$ tagging efficiency, for

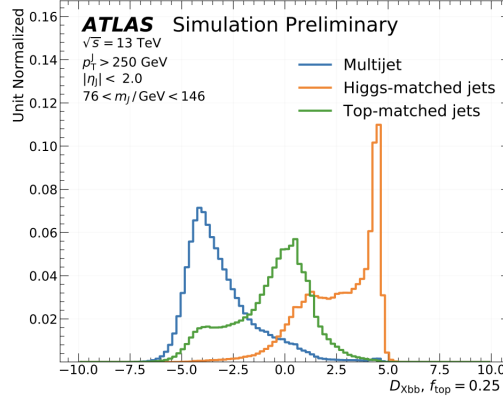


FIGURE 4.10: In Figure, the normalized discriminant distribution of the double b -tagging algorithm D_{Xbb} with a top fraction $f_{top} = 0.25$ [65].

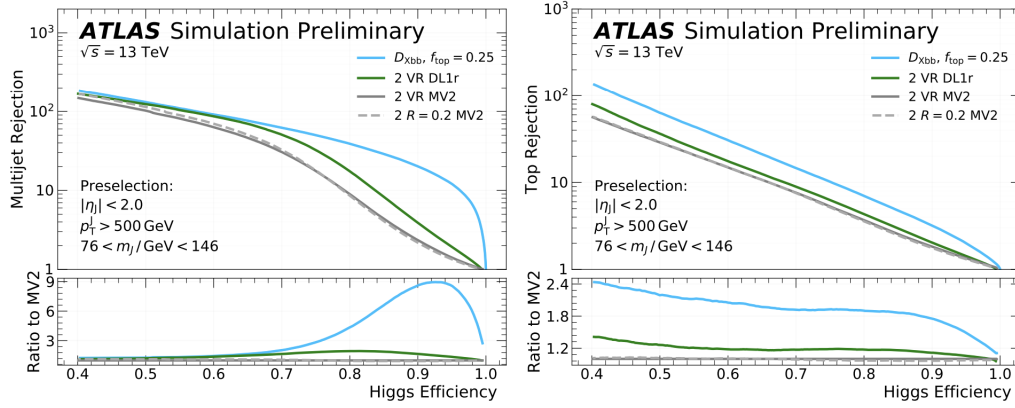


FIGURE 4.11: In Figure, multijet (left) and top jet (right) rejection as function of the $H \rightarrow b\bar{b}$ tagging efficiency, for large- R jet with $p_T > 500$ GeV [65].

large- R jet with $p_T > 500$ GeV is shown in Figure 4.11 where is evident the relative improvement of the rejection of D_{Xbb} over the MV2 algorithm for both backgrounds.

4.2 $Hb\bar{b}$ tagging and colour sensitive variables

This section is largely based on a paper written with collaborators and published in [51]. This work started as a collaboration between experimental and theoretical physicists, which aimed of studying a $Hb\bar{b}$ tagging algorithm by means of colour sensitive variables.

This paper addresses the problem of identification of b -jets coming from the decay of a colour singlet, i.e., the Higgs boson, and those from the QCD background. In particular the study is focused on the associate production of a vector boson and a Higgs boson decaying into a pair of b -jets, recently observed at LHC. The combination of existing theory-driven observables and the Lund jet plane images are considered, in order to design an innovative $Hb\bar{b}$ tagger. The combination is done by machine learning algorithms, trained on events obtained with fast detector simulation

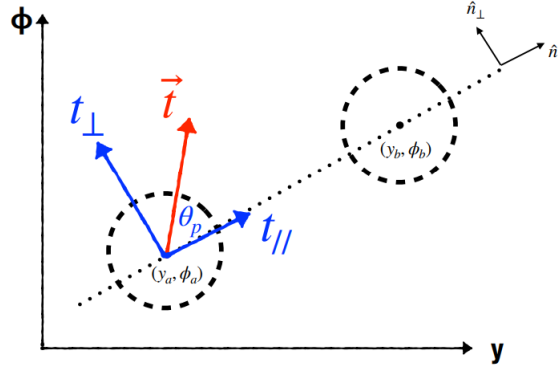


FIGURE 4.12: A pull vector representation with the parallel and perpendicular components and the pull angle of the hard jet J_a [69]. See text for details

techniques. The result shows good discrimination performances. Moreover, also the mass-dependence of the tagger is investigated.

In the section 4.2.1, the colour sensitive observables used in the tagger are described, in section 4.2.2 the analysis simulation and selection are presented, and in the end 4.2.3 shows the tagger performances are shown.

4.2.1 Colour sensitive variables

Let us consider an event in which there are two hard jets. The QCD radiation emitted by the jets has different configuration in case the two jets are formed from a decay of a colour singlet, i.e., Higgs boson, or of a colour octet, i.e., a gluon. The jets from a colour singlet, in fact, produce radiation between the two jets, while those from a colour octet emit with higher probability in the other direction. The colour sensitive variables are created in order to exploit this feature of the QCD emission.

The variables chosen in the Hbb tagger, Jet pull and its projections, namely the pull angle, θ_p [66, 67], and the parallel and perpendicular components of the pull vector, $t_{\parallel}$ and $t_{\perp}$ [68, 69]; the colour ring $\mathcal{O}$ [70]; D_2 [71, 72] and the Lund jet plane [73] are shortly described in the following.

Jet pull

The pull vector $\vec{t}$ is the observable defined as:

$$\vec{t} = \frac{1}{p_{Ta}} \sum_{i \in J_a} p_{Ti} |\vec{r}_i|^2 \hat{r}_i, \quad (4.2)$$

where J_a is the hard jet, and p_{Ta} the transverse momentum of the jet. The sum runs over the constituents of the jet and $\vec{r}_i$ is the distance between the i -th constituent and the jet axis, in the rapidity y and azimuthal-angle ϕ plane:

$$\vec{r}_i = (y_i - y_a, \phi_i - \phi_a). \quad (4.3)$$

If the hard jet is formed paired with another J_b hard jet, the pull vector direction is sensitive to the emitted radiation. It points towards the second jet, in case they are

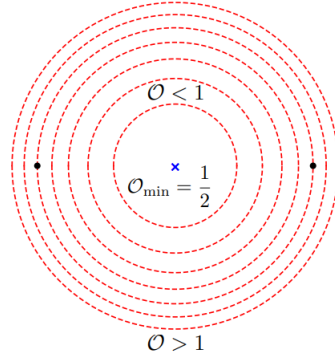


FIGURE 4.13: In Figure, an illustration of the colour ring geometry [70]. The location of the final-state hard partons are denoted by the black dots, and the $\mathcal{O} = 1$ contour passes through the black dots. The region inside the $\mathcal{O} = 1$ circle has observable value less than 1 and the region outside is greater than 1. The point at the center of the circle, directly between the bottom quarks, is where the observable takes its minimum value, as labeled.

from a colour-singlet decay and in another direction in case of colour-octet decay. The pull vector is considered in its projections along two directions: $\hat{n}_{\parallel}$ between the center of the jet J_a to the center of the second jet J_b , and the perpendicular direction respect the previous one $\hat{n}_{\perp}$:

$$t_{\parallel} = \vec{t} \cdot \hat{n}_{\parallel}, \quad \text{with} \quad \hat{n}_{\parallel} = \frac{1}{\sqrt{\Delta y^2 + \Delta \phi^2}} (\Delta y, \Delta \phi), \quad (4.4)$$

$$t_{\perp} = \vec{t} \cdot \hat{n}_{\perp}, \quad \text{with} \quad \hat{n}_{\perp} = \frac{1}{\sqrt{\Delta y^2 + \Delta \phi^2}} (-\Delta \phi, \Delta y), \quad (4.5)$$

where $\Delta y = y_a - y_b$ and $\Delta \phi = \phi_a - \phi_b$.

The angle between the vector and one direction, called **pull angle**, is also considered, defined as:

$$\theta_p = \arccos \frac{t_{\parallel}}{|\vec{t}|}. \quad (4.6)$$

In Figure 4.12 a schematic representation of the pull vector, its component and the pull angle is presented.

Among the observables described above, the pull angle resulted the most discriminant one. However, it is demonstrated that the pull vector and its component are infrared and collinear safe (IRC), while the pull angle is only Sudakov safe [67]. The IRC is important in order to take the variable under control, from experimental-theory comparison point of view, as explained in Section 2.2.3. For this reason, the parallel and perpendicular components are considered, even though their less discriminant power. In the ML algorithm, $t_{\parallel}$, $t_{\perp}$ and θ_p are included. It is also interesting studying the pull vector for both hard jets, they can provide further information about the QCD radiation. For this reason, also the pull vector variables of J_b are given to the ML algorithm.

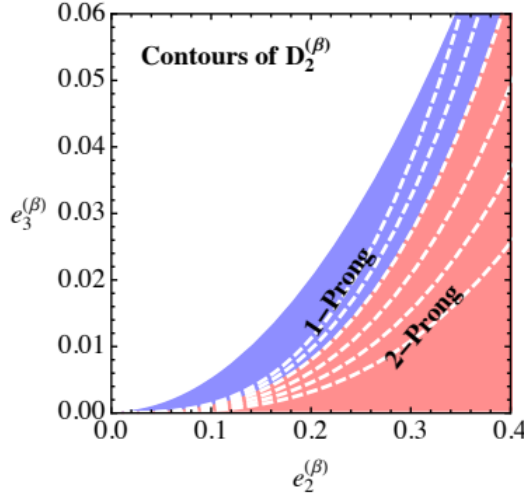


FIGURE 4.14: The contours of constant $D_2^{(\beta)}$ in the the phase space of $e_2^{(\beta)}$ and $e_3^{(\beta)}$ [71] are shown.

Jet colour ring

The jet colour ring is another observable defined as matrix element ratio between signal and background processes in the soft limit. In this case the signal is referred to the production of two partons a and b from a Higgs boson and background to their production from a gluon. For the colour ring one soft gluon emission k from both signal and background is also considered. For the Neyman-Pearson lemma, the optimal discriminant observable is monotonic in their likelihood ratio.

In the boosted regime the ratio results in:

$$\frac{|\mathcal{M}_S|^2}{|\mathcal{M}_B|^2} \simeq \frac{\theta_{ak}^2 + \theta_{bk}^2}{\theta_{ab}^2}, \quad (4.7)$$

where θ_{ij} is the angle between the two partons.

Starting from the ratio above, the jet colour ring can be defined as:

$$\mathcal{O} = \frac{\Delta_{ak}^2 + \Delta_{bk}^2}{\Delta_{ab}^2}, \quad (4.8)$$

where now Δ_{ij} are distances between jets (or sub-jets) in the azimuth-rapidity plane. The values of the colour ring reflect the emission direction: if the gluon tends to fall between the two jets the colour ring is $\mathcal{O} < 1$, otherwise it has values of $\mathcal{O} > 1$. A geometrical representation is shown in Figure 4.13.

D_2

D_2 variable is defined starting from (normalized) n -point energy correlation functions $e_k^{(\beta)}$:

$$D_2^{(\beta)} = \frac{e_3^{(\beta)}}{(e_2^{(\beta)})^3}. \quad (4.9)$$

The $e_2^{(\beta)}$ and $e_3^{(\beta)}$ functions are defined as:

$$e_2^{(\beta)} = \frac{1}{p_{TJ}^2} \sum_{1 \leq i < j < n_j} p_{Ti} p_{Tj} R_{ij}^\beta, \quad (4.10)$$

$$e_3^{(\beta)} = \frac{1}{p_{TJ}^3} \sum_{1 \leq i < j < k < n_j} p_{Ti} p_{Tj} p_{Tk} R_{ij}^\beta R_{ik}^\beta R_{jk}^\beta, \quad (4.11)$$

where p_{TJ} is the transverse momentum of the jet with respect the beam, p_{Ti} the transverse momentum of the particle i and n_j is the number of the particles in the jet. The boost invariant angle $R_{ij}^2 = (\phi_i - \phi_j)^2 + (y_i - y_j)^2$ is the Euclidean distance in the azimuth-rapidity plane and for IRC, the angular exponent $\beta > 0$.

The correlation functions are good candidates for defining new observables because of their proprieties. First they tend to zero in any soft or collinear limits in a configuration of n particles. Second, since they are defined from the angle between pairs of particles, they are not sensitive to recoil. This means insensitive to the angular displacement of the hardest particle inside the jet.

Concerning the discrimination potential, the two-point energy correlation function is sensitive to radiation about a single hard core, i.e., quark versus gluon discrimination. The three-point energy correlation function is useful in two-prong jet identification for the boosted electroweak bosons, i.e., Higgs boson, while the QCD background from the two-point one.

The definition of $D_2^{(\beta)}$ can exploit the 1- and 2-prong jet discrimination. Signal jets are characterized by a small value of $D_2^{(\beta)}$, while background jets predominantly have large $D_2^{(\beta)}$. Contours of constant $D_2^{(\beta)}$ lie in the signal or background region, as is shown schematically in Figure 4.14.

Lund jet Plane

The variable is inspired by Lund diagrams, a theoretical representation of the phase-space of jets. The diagram is a triangle in a two dimensional logarithmic plane that shows the transverse momentum and the angle of any given emission with respect to its emitter. The Lund Jet Plane can be built for a single jet, following the clustering tree of the Cambridge/Aachen (C/A) algorithm [13, 14].

In Figure 4.15, a schematic representation which can help to understand how the primary Lund plane is constructed. On the top of the Figure, the (a) black particle is the primary particle of the jet, from which (b) and (c) are emitted. In the middle, the Lund diagram is represented. The diagram space is formed by $\ln 1/\Delta$ and $\ln k_t$ plane, where Δ is the angle and k_t the transverse momentum of a emission respect to its emitter. In the bottom the primary Lund Plane shows the emission points in the phase-space if it comes from the primary particle. In the right case, in fact, only particle (b) is represented since (c) is not emitted from the primary (a) particle, but from (b).

The primary Lund plane can be obtained by the re-cluster of a jet constituents with the C/A algorithm, recording the kinematics of the splitting following the hardest branch.

The Lund jet plane has been used in vector boson [73], top [74] and Higgs [75] studies for tagging. Furthermore, it has been successfully measured at ATLAS [76] and first-principle theoretical predictions have been performed in Ref. [77].

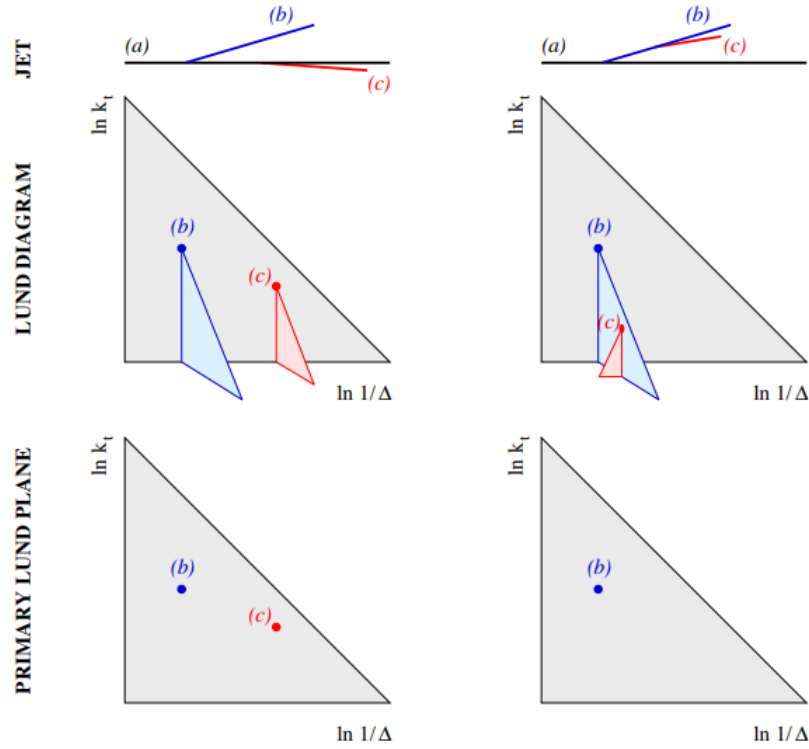


FIGURE 4.15: The representation for different emissions: on the left two emissions from the particle (a), on the right one emission (b) from the particle (a) and one emission (c) from the emission (b). On the top the jet representation, in the middle the Lund diagram and in the bottom the primary Lund plane [73].

TABLE 4.1: The efficiencies of the analysis after cuts.

	Truth	Reco
Signal	20%	17%
Background	1.6%	1.3%

4.2.2 $VHbb$ simulation and selection

The simulation studies are based on datasets which count 300k events for the signal and 4M events for the background. The simulation has been done privately, outside the ATLAS simulation infrastructure. The produced processes are $pp \rightarrow H(b\bar{b})Z(\nu_\ell\bar{\nu}_\ell)$ for the signal and $pp \rightarrow b\bar{b}\nu_\ell\bar{\nu}_\ell$ for the background. The number of events are chosen in order to have about 50k events for both signal and background after the analysis cuts, reported here in Table 4.1.

The parton-level generation is done using MG5_AMC@NLO v2.8.3.2 [78], imposing a cut of 200 GeV on the p_T for the neutrino pair in the final state, for ensuring the boosted regime. The process is then showered in PYTHIA v8.305 [79], including MPI and underlying events, to produce particle-level events.

In order to take into account the detector effects, a fast detector simulation is performed with DELPHES v3.5.0 [80, 81]. The fast simulation provides an evaluation on the reconstruction can affect the analysis, without having a computing-expensive full simulation and, in addition, without a strong dependence on one specific detector.

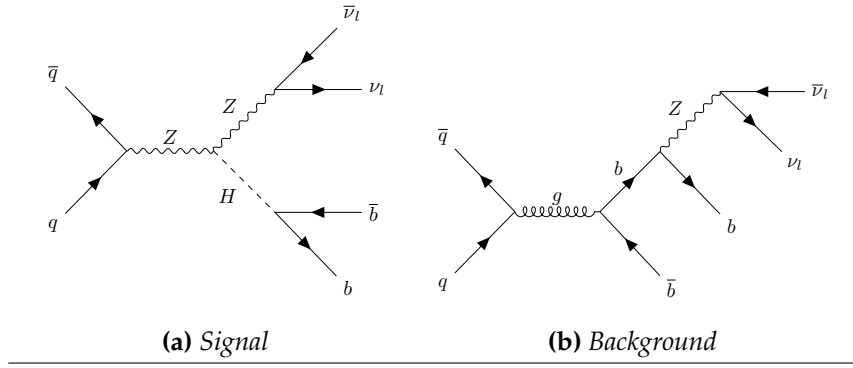


FIGURE 4.16: In Figure the main Feynman diagrams of signal and background processes are shown.

From DELPHES, both Monte Carlo *truth* and *reco* are obtained. Truth events are referred as those containing the particle-level information from a perfectly efficient detector, while reco events are those reconstructed including the detector effects. The fast simulation is run on the ATLAS card, customized for our usage. For the truth case, all visible, stable particles with $p_T > 0.5$ GeV are considered. Instead, in the reco case, the jets are built using the simulated calorimeter towers and tracks. All electromagnetic calorimeter towers with energy $E > 0.5$ GeV and significance $S > 2.0$ and all hadronic calorimeter towers with energy $E > 1.0$ GeV and significance $S > 2.0$ are considered. Tracks are required to have $p_T > 0.5$ GeV. DELPHES uses FASTJET v3.3.4 [57] to perform the jet clustering.

The analysis cuts are the same for truth and reco. The constituents are clustered into jets with radius $R = 1.0$ using anti- k_t algorithm [15]. The leading- p_T jet with $p_T > 200$ GeV, for the boosted regime and $|y| < 1.5$, because of the tracking detector acceptance is chosen for each event, called *large radius jet*. The constituents are also clustered into smaller jets with radius $R = 0.2$. Those jets which have the $\Delta R < 0.8$ from the large radius jet are considered as its sub-jets. Since the studied process considers two b -jets in the final state, a procedure of b -identification is performed, starting from b -partons, known as true b -labelling. First, the b -partons originated from the hard scattering with $p_T > 5.0$ GeV are considered. Second the closest jet of the event to the parton is identified as b -jet, if the distance is below 0.2.

The event is accepted if there are at least two b sub-jets with $p_T > 10$ GeV. Pull jet variables are calculated on the sub-jets, while D_2 on the larger jet. For the colour ring a third non- b sub-jet is needed. Since in many cases the third sub-jet is not present, in order to avoid discarding too many events, a default value of -1 is assigned to the colour ring.

For the Lund Plane the constituents of the large jet are re-clustered using the C/A algorithm. The Lund Jet Plane is built starting from the declustering history of the jet. The Lund Plane image is constructed, which is a 2D Lund Plane histogram, with 25×25 pixels. In each pixel a value of 0 or 1 is put if the splitting obtained from the declustering falls in that pixel or not. The implementation is based on the Fastjet-contrib implementation, it can be found in this repository [82].

The efficiencies of signal and background events are reported in Table 4.1. The p_T and rapidity cut on the large jet are those that affects more the efficiency, contributing respectively with 60% and 10% on the discard events.

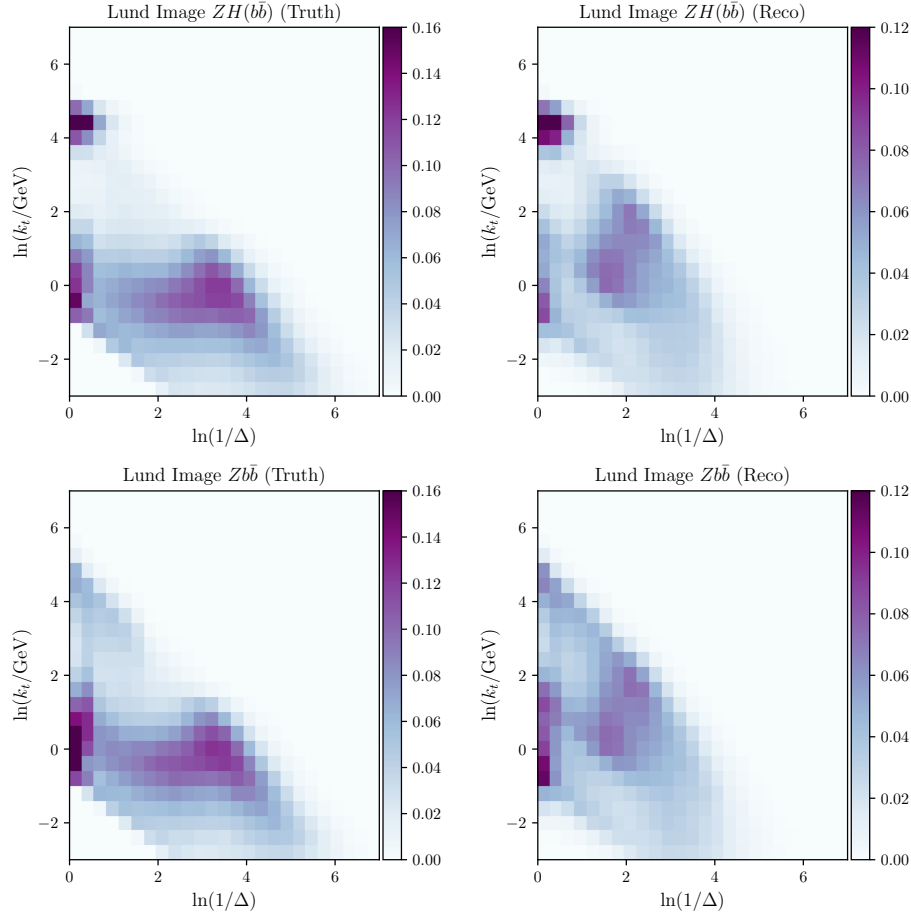


FIGURE 4.17: Averaged primary Lund jet plane images for $ZH(b\bar{b})$ and $Zb\bar{b}$ in the truth and reco case [51].

4.2.3 Performances

The normalized Lund images for signal and background for both truth and reco cases are shown in Figure 4.17. The quality of the image is worse in the reco case, but the different pattern at high k_t and large Δ values is still visible. The Lund images are classified with a CNN, implemented using Keras [83]. The data set is distributed in 70% for the training, 15% for the validation and 15% for the test. The tuned architecture used for the CNN is reported in Table 4.2. As activation functions *relu* and *softmax* are used for the intermediate and last layer, and *adam* as optimiser. After the second and fourth layer the MaxPooling is used for image downsizing.

The normalized colour sensitive variable distributions (CS) and the output of the CNN Lund jet Plane classifier (LP_{CNN}) after the analysis cuts are reported in Figure 4.18. The variables are shown for both truth and reco cases, respectively in continuous and dashed lines, and for signal and background, in red and blue colours. From the figure, it is evident that the most discriminant variables among the CS are the $\mathcal{O}$ and D_2 . The pull vector jet-related variables are less powerful, in particular in comparison between truth and reco case. Those variables seem more affected by the detector effects. The pull angle, in fact, shows a good discriminant power between signal and background at truth level, which loses at reco level.

All the variables, CS + LP_{CNN} are used as input to ML algorithms in order to build combined classifiers, mainly a BDT is trained.

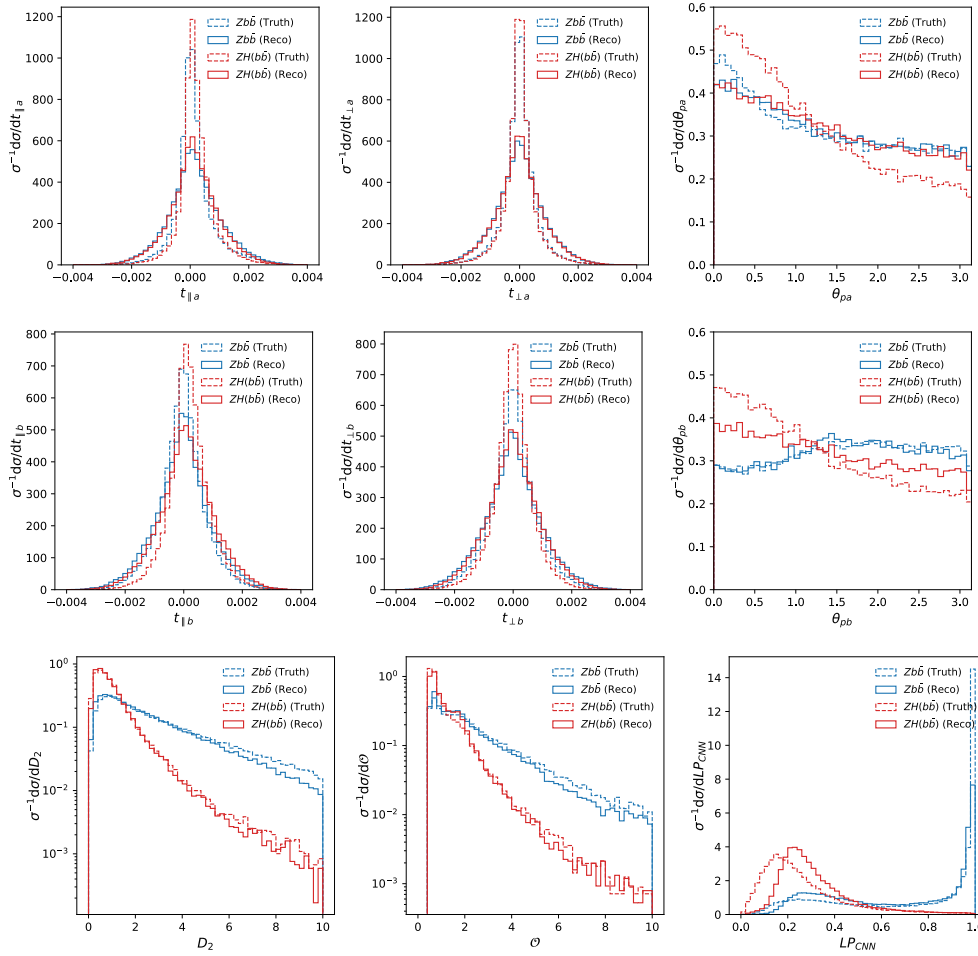


FIGURE 4.18: The distributions of the colour sensitive variables and Lund jet plane after the selection cuts. Signal ($ZH(b\bar{b})$) and background ($Zb\bar{b}$) distributions are shown in red and blue, respectively, for both truth and reco cases [51].

In Figure 4.19, the receiver operating characteristic (ROC) curves for different combinations of the observables are shown. Namely, the colour sensitive variables CS, without the LP_{CNN} , D_2 and the colour ring (D_2+CR), only the LP_{CNN} and all the CS observables with the LP_{CNN} (CS+ LP_{CNN}). For each ROC curve, the area under it (AUC) is calculated in Table 4.3 for both truth and reco cases.

As expected, it can be observed a worsening of the discrimination power by moving from truth to reco case. The performances are still good, reaching 0.85 for the combination CS + LP_{CNN} in the reco case. The most discriminant variables among the CS, as evident from the Table 4.3, are D_2 and CR. This is expected even looking at the observable distributions, where the pull jet variables show less discriminant power than the other variables, with a significant worsening at reco level. From the Table it is evident that the LP_{CNN} alone performs better than CS. On the other hand, the combination of the CS and LP_{CNN} results in a good classification power with a value of AUC equal to 0.893 in the truth case and 0.846 in the reco case.

In Table 4.4 the rank of variables based on their importance in the BDT are shown. As expected, the most important is the LP_{CNN} for both truth and reco, followed by D_2 and colour ring $\mathcal{O}$. In order to create a tagger which can be used in other channels,

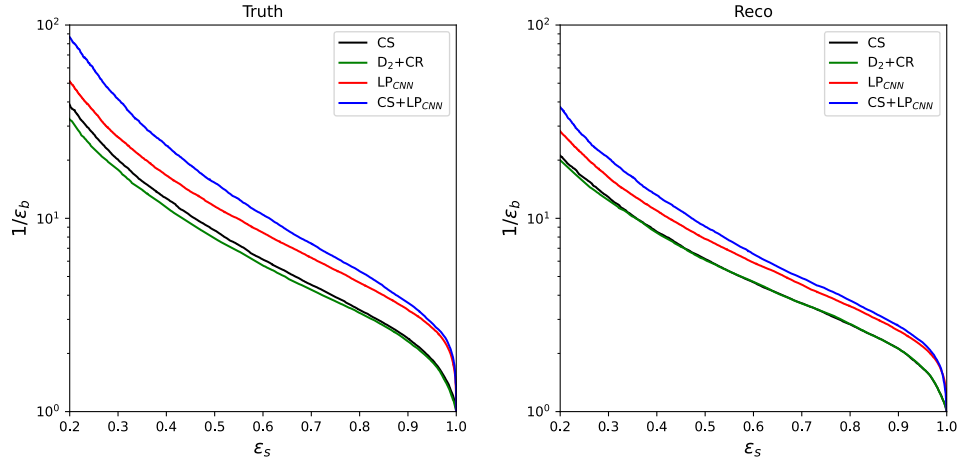


FIGURE 4.19: The ROC curves showing background rejection as a function of signal efficiency for the truth (left) and reco case (right) for CS variables, LP_{CNN} and the combined cases.

such as the Z or W boson hadronic decays, the discrimination has to be uncorrelated with the invariant mass of the system, specifically to the invariant mass of the b -jet pairs. In Figure 4.20, the invariant masses are shown for D_2 only 4.20a, CS without D_2 4.20b and the LP_{CNN} only 4.20c. The normalized distributions are shown for three different cuts on the BDT discriminant variable and without any cuts. The distributions are shown in the reco case, since the truth case is similar. If the tagger is mass-independent the three distributions should be overlapped. The analysis of each variable lead to find two most mass-correlated variables. In particular, from figure 5a, it is evident that D_2 , among the CS, is the high correlated with mass. Removing it from the CS variables of BDT, in fact, the correlation disappears (Figure 4.20b). The mass-correlation of D_2 has been already investigated in literature [84, 85]. The other correlated variable is the LP_{CNN} , as visible in Figure 4.20c.

Unfortunately both mass-correlated variables, D_2 and LP_{CNN} , are the most important in the BDT rank, which means that they can not be removed without a consistent loss in performances. Further studies can be done in order to find a de-correlation method.

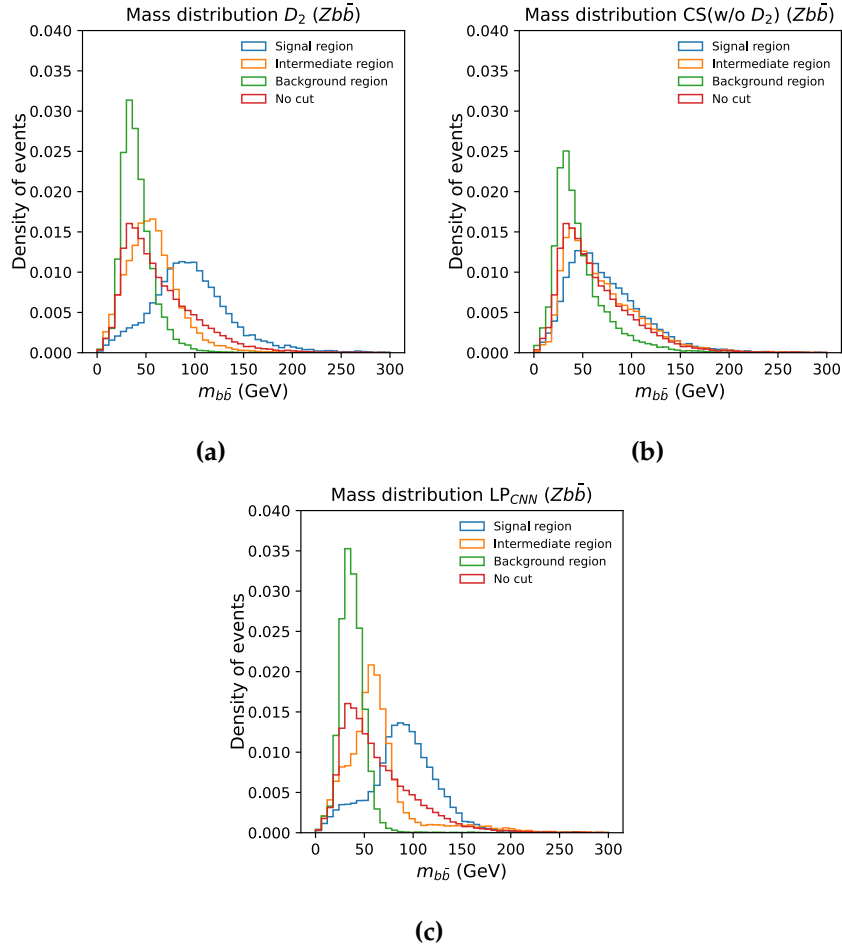


FIGURE 4.20: The distribution of the invariant mass of the b -jets pair in the reco background sample for (a) different cuts on D_2 , (b) different ranges of the BDT output for the CS without the inclusion of D_2 , (c) and for different cuts on Lund plane CNN output.

TABLE 4.2: CNN architectures used for truth and reco cases of the Lund Plane. N_i are the number of filters in the corresponding i -convolutional layer. The filter size is always 3×3 for each layer. The numbers mentioned in the brackets are the dropouts strength used only for the reco data set.

Parameter	Value
N_1 Conv2D	30
N_2 Conv2D	30
Dropouts	– (0.3)
N_3 Conv2D	30
Dropouts	– (0.3)
N_4 Conv2D	10
Dropouts	– (0.1)
Flat Layer	150
Epochs	30
Batch Size	800

TABLE 4.3: Area under the ROC curves for different combination of observables.

	AUC - Test Sample	
	Truth	Reco
CS observables	0.826	0.788
D_2 +CR	0.817	0.787
LP_{CNN}	0.876	0.828
CS + LP_{CNN}	0.893	0.846

TABLE 4.4: BDT observable ranking for the truth and reco cases.

Observable Ranking				
Rank	Truth		Reco	
	Obs.	Importance	Obs.	Importance
1	LP_{CNN}	6.6×10^{-1}	LP_{CNN}	4.8×10^{-1}
2	D_2	1.4×10^{-1}	$\mathcal{O}$	1.0×10^{-1}
3	$\mathcal{O}$	5.7×10^{-2}	D_2	9.3×10^{-2}
4	θ_{pb}	3.0×10^{-2}	θ_{pb}	7.0×10^{-2}
5	θ_{pa}	2.9×10^{-2}	θ_{pa}	6.5×10^{-2}
6	$t_{\parallel b}$	2.6×10^{-2}	$t_{\perp b}$	6.0×10^{-2}
7	$t_{\parallel a}$	2.4×10^{-2}	$t_{\parallel a}$	4.5×10^{-2}
8	$t_{\perp b}$	1.9×10^{-2}	$t_{\perp a}$	4.3×10^{-2}
9	$t_{\perp a}$	1.0×10^{-3}	$t_{\parallel a}$	3.3×10^{-2}

Chapter 5

$W+c$ analysis

This chapter is dedicated to the analysis of a particular process foreseen by the Standard Model, namely the production of a W boson in association with a single charm quark, and to the measurement of its production rate at the center of mass energy of 13 TeV. The data used for this analysis are those collected by ATLAS during Run 2, between 2015 and 2018, corresponding to a total integrated luminosity of 140.1 fb⁻¹.

This process is interesting since is sensitive to the strange quark PDF in the proton, as further clarified in the following Section 5.1. It represents, to some extent, the core work of this manuscript, which maps the three years of my PhD work.

The chapter is distributed as following: in Section 5.2 the measurement strategy is discussed; Section 5.3 provides information about the data and the MC samples used in the analysis; in Section 5.4 are detailed all the selections applied and the definition of the fiducial phase space.

5.1 The motivation

The strange sea-quark PDF is matter of interest since long time and several experiments have spent efforts in its determination.

Precision measurements by HERA collaborations of neutral current (NC) and charged current (CC) cross sections in $e^\pm p$ scattering were used to determine a PDF set (referred as HERAPDF2.0 [96]). They do have limitations, for instance they cannot distinguish quark flavour between the down-type sea quarks, $\bar{s}$ and $\bar{d}$. Some constraints at high x and low Q^2 are provided by fixed-target experiments, i.e., NuTeV [97] and Hermes [98]. However these results are affected by large uncertainties, due to the nuclear correction factors in the theoretical predictions of the PDFs.

Differential W and Z distribution measurements are studied by LHC and Tevatron in order to obtain other constraints on the light sea quark distributions at low x . The $W^\pm$ asymmetry measurements, performed using $p\bar{p}$ collisions at Tevatron, have less uncertainties, but there is an historical tension between the results of CDF [99] and DØ[100] collaborations.

In this recent ATLAS result [86], a combination of neutral and charge current DIS data from HERA are simultaneously fit with several ATLAS cross section measurements: the inclusive W , Z production at 7 TeV [89], W +jet production at 8 TeV [101] and the Z +jet production at 8 TeV [55]. The resulting PDF sets [86], ATLASepWZVjet20 and ATLASepWZ16, improved the determination of the sea-quark densities at high Bjorken x .

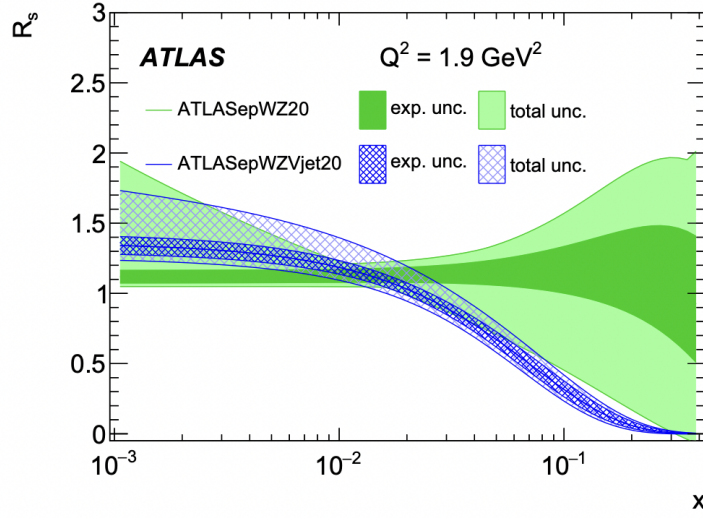


FIGURE 5.1: The strangeness suppression factor as a function of x at the factorization scale of 1.9 GeV^2 extracted from the ATLASepWZV-jet20 (green) and ATLASepWZVjet20 (blue) fit [86].

The fraction of the strange-quark density can be quantified by the *strangeness suppression fraction* r_s (or R_s), defined as:

$$r_s = \frac{s + \bar{s}}{\bar{u} + \bar{d}} \quad (5.1)$$

where s , u and d are the corresponding quark density functions.

Previous results obtained from dimuon production data from muon-neutrino CC DIS with associated charm-quark production [102] supported the assumption of strange quark suppression relative to the up and down quark sea over the full range of x . Best fits to this neutrino scattering data resulted in a value of $r_s \simeq 0.5$ at $Q^2 = 1.9 \text{ GeV}^2$ [90, 95, 103].

The latest ATLAS results, contained in the ATLASepWZVjet20 and ATLASepWZ20 fits, instead found out a better data description by the unsuppressed strange quark sea at low x , while at high x the suppression is restored. ATLASepWZVjet20 in particular found a ratio near to unity at $x \sim 0.01$ and approximately 0.5 at $x=0.1$, as shown in the strangeness suppression fraction distribution in Figure 5.1. This result is also supported by the previous ATLAS $W + c$ analysis at 7 TeV.

The recent analysis of CMS $W + c$ data at 7 and 13 TeV [87, 104] has found a suppressed strange-quark density relative to the light sea at high values. The strangeness suppression factor as function of x obtained by the CMS analysis [87] is shown in Figure 5.2. The distribution is shown in comparison to other NLO predictions from ABMP16nlo [88], which uses the most recent data on charm quark production in charge-current neutrino-nucleon DIS collected by NOMAD and CHORUS experiments together with the other DIS, DY and top quark production results, and ATLASepWZ16nnlo [86]. This result can be interpreted as a tension with the ATLAS results, as seen in Figure 5.2, but considering the total uncertainty, as in Figure 5.1, the ATLAS and CMS results are compatible.

Due to the intriguing topic, several studies have been performed including all the experimental results and theoretical predictions, in order to provide a global fit for

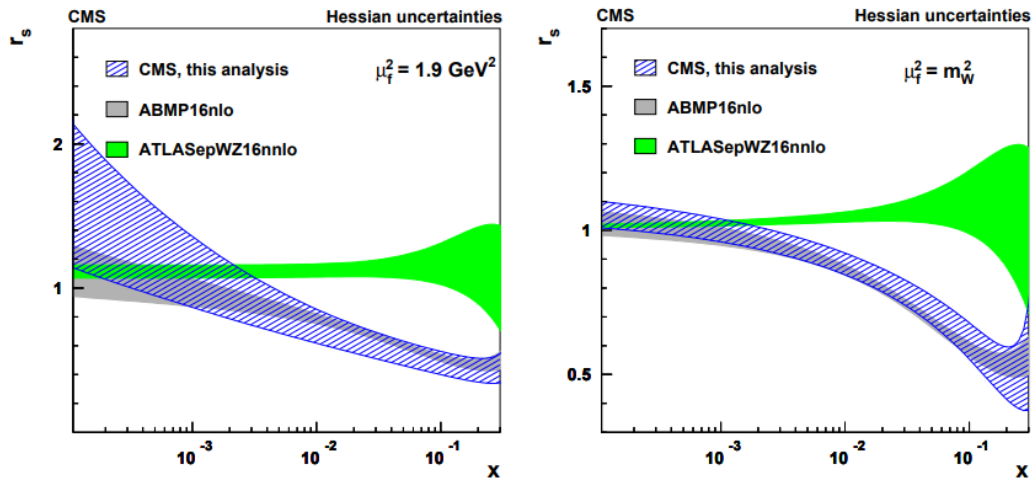


FIGURE 5.2: The strangeness suppression factor as a function of x at the factorization scale of 1.9 GeV^2 (left) and m_W^2 (right). The CMS results obtained by the analysis [87] is compared with NLO prediction calculated using ABMP16nlo [88] and ATLASepWZ16nnlo [89] PDF sets.

the proton-quark PDF and the strangeness suppression factor. Among several studies the CT, MMHT and NNPDF global fit provide consistent results including also the recent ATLAS W, Z dataset [89]. In particular, NNPDF3.1 [105] found out that, including the ATLAS W, Z dataset, the larger strangeness is favoured, but the χ^2 is still not optimal if the neutrino dimuon dataset are included. CT18 [92] compared the fit results including or not the neutrino dimuon dataset, founding a difference of one-sigma for both the gluon and the total strangeness. Finally also the MMHT collaboration [106] revealed an enhanced strangeness driven from the ATLAS W, Z dataset.

The most recent ATLAS set of proton parton distribution functions [107] is the ATLASpdf21 and confirms the ATLAS previous results. This analysis includes the most recent ATLAS measurements with the final combined inclusive HERA data, in order to evaluate the impact of the ATLAS data with the other data.

The most recent PDF results are compared in Figure 5.3. It is evident that including the ATLAS result in NNPDF, the NNPDF_strange, the ratio is shifted from 0.5 to 0.7.

5.2 Measurement strategy

The SM process which leads to the production of a W boson in association with charm quark is mainly induced by gluon-strange quark fusion, with a lower probability of about 10% of gluon-down quark fusion, suppressed by the CKM coupling.

At lowest order in the $W+c$ production, the W boson is produced in association with exactly one charm jet. An illustration of the diagram for the LO process is shown in Figure 5.4.

At higher orders, it is possible to have more than one jet in the final state. This can be due to initial or final radiation or other initial states other than strange-gluon fusion. In figure 5.5, an example of $W+c$ production with more than one jet via gluon-gluon fusion.

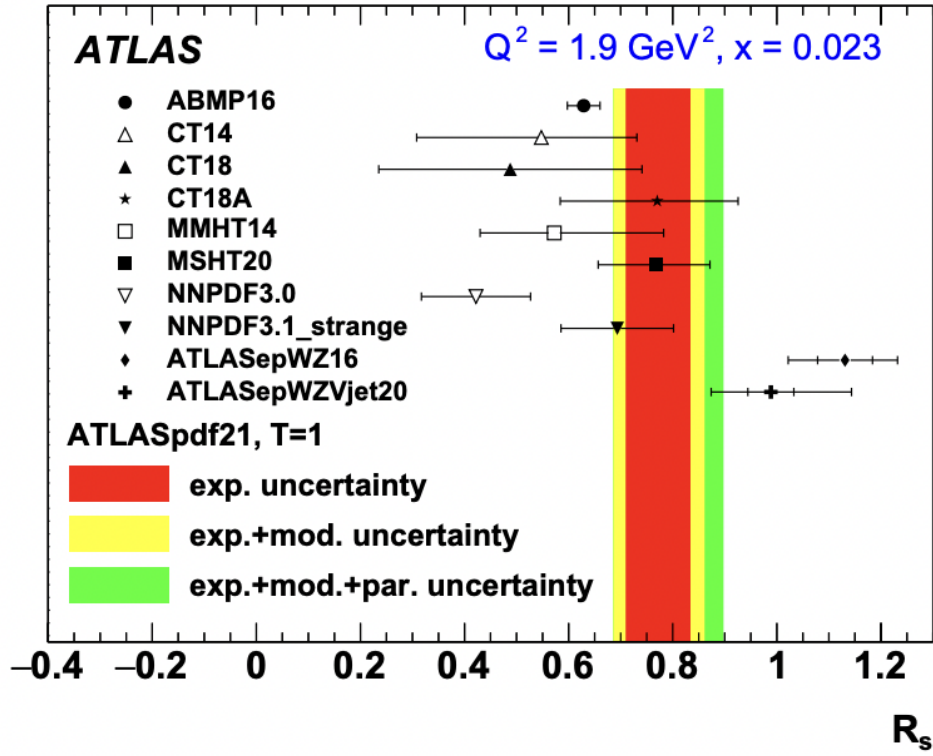


FIGURE 5.3: The strangeness suppression factor computer at $x=0.023$, $Q^2=1.9 \text{ GeV}^2$ from ATLASpdf21, compared with other PDFs: ABMP16[90], CT14[91], CT18, CT18A[92], MMHT14[93], MSHT20[7], NNP3.0[94], NNP3.1_strange[95], ATLASepWZ16 [89] and ATLASepWZVjets20 [86].

For this reason this analysis is done requiring 1 or 2 jets. More than two jets are not considered, due to the large background from $t\bar{t}$ sample.

The W boson is always produced with opposite charge sign with respect the charm quark ($W^+ + \bar{c}$ or $W^- + c$), therefore this feature can be exploited for signal isolation with respect to other background processes for which this sign anti-correlation does not hold. Indeed, except for some cases that will be treated a part, the backgrounds do not have correlation in W boson-charm charge sign. It can be said that they are not opposite/same sign (OS/SS) symmetric. For this reason, the procedure of sample subtraction between opposite and same sign (OS-SS) provides a high purity signal sample.

The W boson is selected via its lepton decay. In this analysis only muons and electrons are considered as W -lepton, while the tau decay is considered as background, due to its high hadron decay rate.

The charm quark can be generally reconstructed in two ways. The first is through its charmed hadrons D^{*+} and D^+ , the second, with a branching fraction of 9.6%, is through its semileptonic decay into a muon, called *soft*.

In this analysis, only the second reconstruction is taken into account, while, in ATLAS, a separate analysis has been recently published [108] which considers the first case.

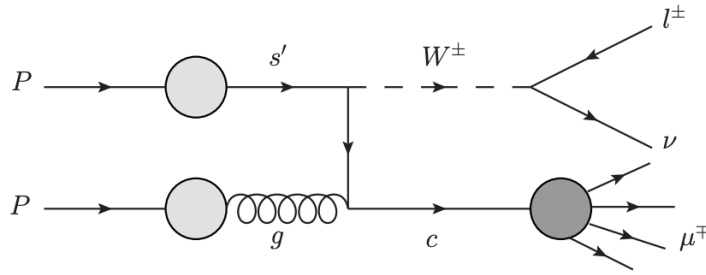


FIGURE 5.4: The strange-gluon fusion associated production of a W boson and a single charm quark.

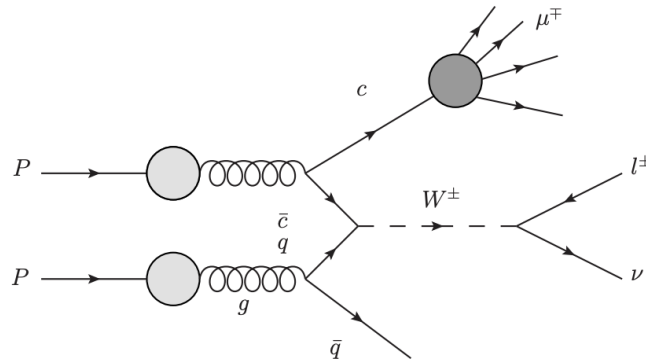


FIGURE 5.5: The gluon-gluon fusion associated production of a W boson and a single charm quark.

5.2.1 Background sources

The background sources of the $W+c$ signal come from W or Z bosons or top quarks, which mimic the same final state as the signal. In particular, the most relevant backgrounds are:

- **Z +jets**, where Z decays leptonically, $Z \rightarrow ll$. In this case also the τ lepton is considered, as shown in Figure 5.6. This process can be identified as signal, since one lepton can mimic the one coming from the W decay while the other is not identified;
- **Dibosons**, i.e., the couple W - W , Z - Z , W - Z bosons, in their lepton and hadron decays, which can mimic the W decay and jets final states. Possible Feynman diagrams are shown in Figure 5.7;
- **single top quark**, where its most probable decay is the W +jets. The three channels of top production, t -, s -channel and Wt , are reported in Figure 5.8;
- **$t\bar{t}$** , where both top decays in W +jets and this can be reduced when only one jet is required. The most probable production way is shown in Figure 5.9;

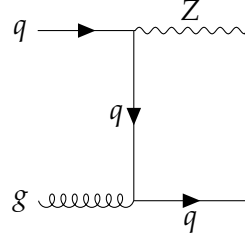
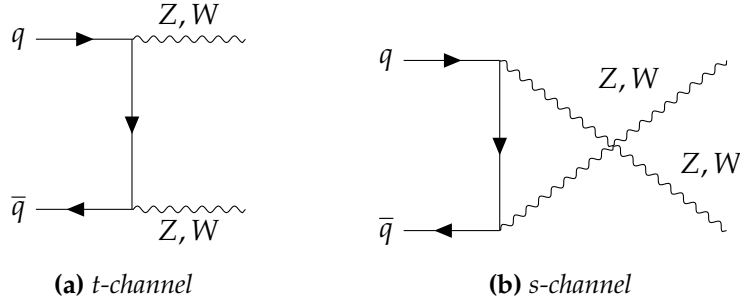


FIGURE 5.6: Feynman diagram as example for Z+jets background

FIGURE 5.7: *t*- and *s*-channel Feynman diagrams for the Diboson background source.

- **multijet background**, which comes from the wrong identification of the lepton as coming from the W decay. More details will be provided in the multijet dedicated Section 6.2, while possible channels of this background and production are shown in Figure 5.10.

5.3 Data and Simulation

This analysis is based on data recorded between 2015 and 2018 (Run 2) by the ATLAS detector in LHC pp collisions at the center of mass energy $\sqrt{s} = 13$ TeV. The detector conditions evolved significantly during Run 2, with the LHC instantaneous luminosity $\mathcal{L}_{peak}$ at the start of fills increasing from 5 to $19 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ and the number colliding bunch pairs n_b from 2232 in 2015 to 2544 in 2018, separated by 25 ns of bunch spacing on average. The integrated luminosities for each year are reported in Table 5.1, corresponding to a total integrated luminosity of 140.1 fb^{-1} . The uncer-

TABLE 5.1: The integrated luminosities for Run 2 data-taking (2015-2018).

Year	Integrated luminosity [pb^{-1}]
2015	3244.54
2016	33402.2
2017	44630.6
2018	58791.6

tainty in the combined 2015-2018 integrated luminosity amounts to 0.83% [110], as obtained using the LUCID-2 detector [111] for the primary luminosity measurements,

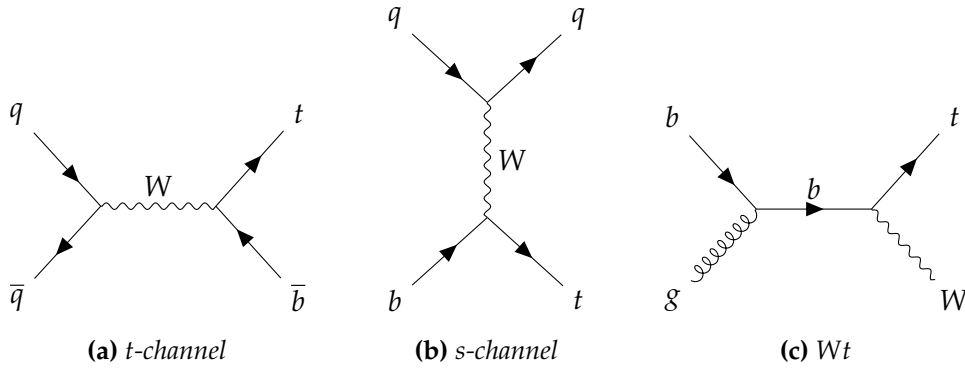


FIGURE 5.8: Example of Feynman diagrams for the top background production.

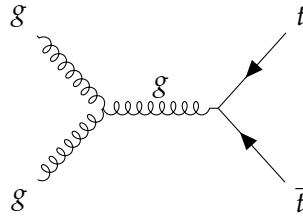


FIGURE 5.9: Example of Feynman diagram for the $t\bar{t}$ background production.

complemented by measurements using the inner detector and calorimeters. The data-taking in ATLAS is organized in *runs*, which normally correspond to a LHC fill, including beam injection, acceleration and preparation for collision. ATLAS prepares a list of *good runs*, i.e., a list of runs when the beam is declared *stable* and the detector is working efficiently. For data, the Good Run List (GRL), explained in 2.4.2 adopted are:

- 2015: *data15_13TeV.periodAllYear_DetStatus-v89-pro21-02_Unknown_PHYS_StandardGRL_All_Good_25ns.xml*;
- 2016: *data16_13TeV.periodAllYear_DetStatus-v89-pro21-01_DQDefects-00-02-04_PHYS_StandardGRL_All_Good_25ns.xml*;
- 2017: *data17_13TeV.periodAllYear_DetStatus-v99-pro22-01_Unknown_PHYS_StandardGRL_All_Good_25ns_TriggerNo17e33prim.xml*;
- 2018: *data18_13TeV.periodAllYear_DetStatus-v102-pro2204_Unknown_PHYS_StandardGRL_All_Good_25ns_TriggerNo17e33prim.xml*.

The Derivation Framework [109] is used by the ATLAS analyses in order to reduce the data file obtained after the detector reconstruction, the Analysis Object Data (AOD), in a more usable size. The files obtained are the derived AOD (DAODs), written in xAOD data structure. This means that they can directly be read by the analysis framework, which makes use of the ROOT interface. The DAODs are obtained from the AOD reducing the size format, via four filter operations: *skimming*, which removes the not-interesting events; *thinning*, which removes the not useful reconstructed objects from an event, but keeping the rest of the event; *slimming*, which

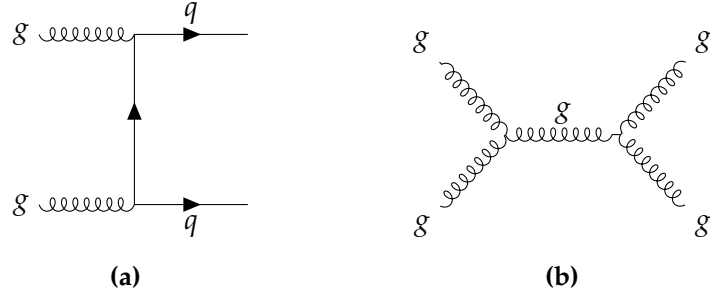


FIGURE 5.10: Example of Feynman diagrams for multijet background source.

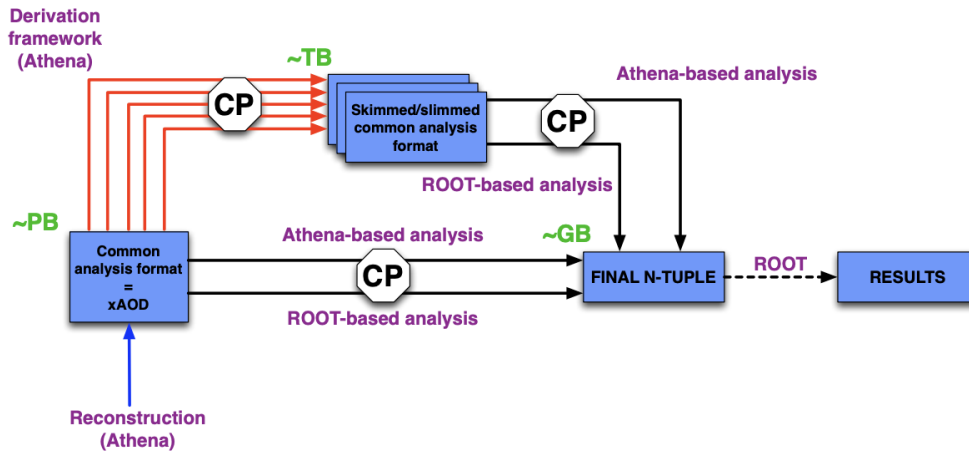


FIGURE 5.11: Schematic representation of the derivation framework chain [109].

removes information within objects, but keeping the rest of the object, and finally *augmentation*, which adds data not found in the input.

The filters are done following selections, which are different for each analysis group, accordingly with the specific needs for the analyses. In this thesis, for the $W+c$ analysis, the STDM4 and TOPQ1 DxAODs are chosen for simulated samples and data. The STDM4 are produced for Standard Model analyses, in which at least one lepton (electron or muon) is required in each event, passing the single lepton trigger selection [112, 113]. The TOPQ1 derivations are produced from the ATLAS top group, which consists in the single lepton selection, with at least one muon or electron, both at trigger and offline level, as for the STDM4 production, [114].

A scheme of the ATLAS Derivation Framework is shown in Figure 5.11. The Derivation Framework is based on Athena [43], which is based on the common Gaudi framework [115], used by ATLAS, LHCb and FCC experiments. It includes all the components and interfaces needed to build data processing applications [116].

For this analysis Athena release 21 is used, in particular the AnalysisBase, 21.2.248, which is the release chosen for the analysis tools. For the electron channel, the CxAODFramework is also used for ntuple production, where a calibration is applied to physics objects and the events are saved in xAOD format. In the muon channel, the AnalysisTop framework has been used. These ATLAS guidelines are detailed in [117, 118, 119].

The $W+c$ signal and all backgrounds sources containing W or Z bosons or top quarks are modeled using simulated event samples, as discussed in more detail in the following sections. All generated samples are passed through the simulation of the ATLAS detector [120], based on GEANT4 [121]. The same offline reconstruction methods used with data events are applied to the simulated events samples. Minimum-bias events, which are the soft inelastic interactions selected with a loose trigger requirement, are generated by Pythia 8.165 [122, 123] and are used to simulate multiple pp interactions in the same and nearby bunch crossings (pile-up). The additional multijet background, important due to the presence of misidentified reconstructed leptons, is estimated with a data-driven method described in Section 6.2.

5.3.1 W +jets samples

The simulation of $W + c$ -jet signal is part of a wider W +jets MC production, which includes both light and heavy flavour jets.

Generator level filters

In order to reduce the resources requested for the production of the samples, some filters are required at the generator level:

- **B-Filter:** selects events with at least one b-hadron with $p_T > 5$ GeV and $|\eta| < 2.9$ matched within $\Delta R < 0.4$ with a *AntiKt4TruthJets* jet with $p_T > 10$ GeV and $|\eta| < 2.9$;
- **C-Filter B-Veto:** selects events failing the filter above but with a c -hadron with $p_T > 4$ GeV and $|\eta| < 3.0$;
- **C-Veto-B-Veto:** selects events failing both filters above;
- **Soft Muon Tagging (SMT) filter:**
 - electron channel: selects events with at least ≥ 1 electron with $p_T > 15$ GeV and $|\eta| < 2.7$ and ≥ 1 muon with $p_T > 3.25$ GeV and $|\eta| < 2.7$;
 - muon channel: selects events with ≥ 2 muons with $p_T > 3.25$ GeV and $|\eta| < 2.7$, at least one of which must have $p_T > 15$ GeV;
 - τ channel: selects events passing the OR of the previous two.

The first three filters are the standard ones defined by the ATLAS Physics Modelling Group, while the SMT filter was developed for this analysis.

The SMT filter helps increasing the effective statistics of the MC samples for our analysis, which request to have one muon within a jet in addition to the prompt, isolated W -lepton. However, this filter has an important caveat: it correctly selects the full component of heavy-hadron $\rightarrow \mu + X$ decays, but introduces an important bias in the light-hadrons $\rightarrow \mu + X$ component. Light-hadrons with long lifetimes (e.g. pions and kaons) are considered stable at the generator level, and are only later decayed (possibly to muons) during the detector simulation step, thus their decays to muons cannot be selected by the generator level filters and they are therefore discarded. The light-hadrons $\rightarrow \mu + X$ component must therefore be obtained from MC samples without the SMT filter. Furthermore, an overlap removal procedure has to be applied in order to avoid double-counting of the components: for this reason, the W +jets

sample is defined by combining samples with the SMT filter, retaining only the events for which the soft muon in the jet is originated by the decay of a HF-hadron, and samples without the SMT filter, for which only events with a soft muon not originated by a HF-hadron decay, are considered. A more detailed description of the overlap removal procedure can be found in section 5.4.5.

Baseline sample

The baseline samples used for W +jets are produced with the event generator Madgraph MG5_aMC@NLO [78, 124], which is a framework used for cross-section computing and event generation, including NLO QCD corrections. It can combine the matrix element (ME) processes at leading-order (LO) and at next-to-leading-order (NLO) with different multiplicities. With the $FxFx$ procedure [125], in fact, it is possible to match the ME at NLO to the parton showers when they have different parton final states, i.e., different final-state hard process multiplicities, avoiding the double counting. The showering part is provided by PYTHIA 8.245 (P8) event generator [126]. PYTHIA is a program for high-energy physic collision events, which includes also hard and soft interactions, parton distributions, initial and final state parton showers, multi-parton interactions, fragmentation and decay. In the end, MG5_aMC@NLO+P8 are interfaced with the EvtGen [127], the event generator suited for B physics, with detailed models for semileptonic decays [128].

The dataset identification number (DSID) of the samples, with the correspondent channel, generator and filter applied are reported in Table 5.2. The samples are produced with a ME calculation at NLO order for up to 3 jets, i.e., with ME covering the phase space of up to 4 jets. As discussed above for the SMT filter, the W +jets channels are defined combining events from the SMT filter and the other filters. For example, the electron channel (W_{ev}) is defined by combining events from DSID 510718 for which the soft muon inside the jet is originated by a HF-hadron decay with events from samples corresponding to 508979-508980-508981 DSID for which that muon is not originated by a HF-hadron decay, and similarly for the other channels.

Alternative sample

An alternative W +jets sample is defined in a similar way combining the SHERPA generator, version 2.2.11 [129, 130] samples of Table 5.3. SHERPA 2.2.11 is an alternative MC event generator, combining the ME and parton shower implementation.

In this case, samples used for the HF-component (700409-700411) have been produced (due to limited resources) with NLO precision up to 1 jet (i.e., with ME covering the phase space up to 2 jets), which is still sufficient for our analysis, interested in events with ≤ 2 jets for the signal. However, SHERPA has been in this case interfaced with EvtGen in order to have a more precise description of the HF-hadron decays. The light-hadron component is drawn from the standard SHERPA 2.2.11 samples recommended in ATLAS, 700338-700346, which have NLO precision up to 3 jets.

DSID	Channel	Generator	Filter	Comments
501718	$W_{\nu\nu}$	MG5_aMC@NLO +P8 +EvtGen	SMT filter	0,1,2,3 jets @NLO
501719	$W_{\mu\nu}$			
508979	$W_{\nu\nu}$	MG5_aMC@NLO +P8 +EvtGen	BFilter	0,1,2,3 jets @NLO
508980			CFilterBVeto	
508981			CVetoBVeto	
508982	$W_{\mu\nu}$	MG5_aMC@NLO +P8 +EvtGen	BFilter	0,1,2,3 jets @NLO
508983			CFilterBVeto	
508984			CVetoBVeto	
509751	$W\tau\nu$ with $\tau \rightarrow \ell\nu\nu$	MG5_aMC@NLO+P8+EvtGen	BFilter	0,1,2,3 jets @NLO
509752			CFilterBVeto	
509753			CVetoBVeto	

TABLE 5.2: List of W +jets MG5_aMC@NLO+P8+EvtGen FxFx samples used in the analysis

DSID	Channel	Generator	Filter	Comments
700409	$W_{\nu\nu}$	Sherpa 2.2.11+EvtGen	SMT filter	0,1 jets @NLO
700410	$W_{\mu\nu}$			
700411	$W\tau\nu$ with $\tau \rightarrow \ell\nu\nu$			
700338	$W_{\nu\nu}$	Sherpa 2.2.11	BFilter	0,1,2,3 jets @NLO
700339			CFilterBVeto	
700340			CVetoBVeto	
700341	$W_{\mu\nu}$	Sherpa 2.2.11	BFilter	0,1,2,3 jets @NLO
700342			CFilterBVeto	
700343			CVetoBVeto	
700344	$W\tau\nu$ with $\tau \rightarrow \ell\nu\nu$	Sherpa 2.2.11	BFilter	0,1,2,3 jets @NLO
700345			CFilterBVeto	
700346			CVetoBVeto	

TABLE 5.3: List of W +jets SHERPA samples used in the analysis

Process	Summary Name	DSID	$\sigma \cdot \text{BR}$ [pb]	ϵ_{filter}	k-factor
Zee BFilter	MGPpy8EG_Zee_FxFx_3jets_HT2bias_BFilter	506193	2270.9	2.175350E-02	1.0
Zee CFilterBVeto	MGPpy8EG_Zee_FxFx_3jets_HT2bias_CFilterBVeto	506194	2265.9	1.209887E-01	1.0
Zee CVetoBVeto	MGPpy8EG_Zee_FxFx_3jets_HT2bias_CVetoBVeto	506195	2267.9	8.571174E-01	1.0
$Z\mu\mu$ BFilter	MGPpy8EG_Zmumu_FxFx_3jets_HT2bias_BFilter	506196	2274.3	2.143442E-02	1.0
$Z\mu\mu$ CFilterBVeto	MGPpy8EG_Zmumu_FxFx_3jets_HT2bias_CFilterBVeto	506197	2271.9	1.208696E-01	1.0
$Z\mu\mu$ CVetoBVeto	MGPpy8EG_Zmumu_FxFx_3jets_HT2bias_CVetoBVeto	506198	2267.8	8.574356E-01	1.0
$Z\tau\tau$ BFilter	MGPpy8EG_FxFx_Ztautau_3jets_HT2bias_BFilter	512198	2270.0	2.175909E-02	1.0
$Z\tau\tau$ CFilterBVeto	MGPpy8EG_FxFx_Ztautau_3jets_HT2bias_CFilterBVeto	512199	2270.0	1.218835E-01	1.0
$Z\tau\tau$ CVetoBVeto	MGPpy8EG_FxFx_Ztautau_3jets_HT2bias_CVetoBVeto	512200	2270.0	8.562539E-01	1.0
$Z \rightarrow \ell\ell + Z \rightarrow qq$	Sh_221_NNPDF30NNLO_ZqqZll	363356	15.563	1.0	0.1396
$W^- \rightarrow \ell\nu + W^+ \rightarrow qq$	Sh_221_NNPDF30NNLO_WpqqWmlv	363359	24.717	1.0	1.0
$W^+ \rightarrow \ell\nu + W^- \rightarrow qq$	Sh_221_NNPDF30NNLO_WplvWmqq	363360	24.725	1.0	1.0
$Z \rightarrow \ell\ell + W \rightarrow qq$	Sh_221_NNPDF30NNLO_WqqZll	363358	3.437	1.0	1.0
$W \rightarrow \ell\nu + Z \rightarrow qqZ$	Sh_221_NNPDF30NNLO_WlvZqq	363489	11.413	1.0	1.0
$t\bar{t}$, lep-jets	PwPy8EG_A14_ttbar_hdamp258p75_nonallhad	410501	831.76	0.543	1.0
top t -channel	PowhegPythiaEvtGen_P2012_singlelep_top	410011	43.739	1.0094	1.0
anti-top t -channel	PowhegPythiaEvtGen_P2012_singlelep_top	410012	25.778	1.0193	1.0
top Wt -channel	PowhegPythiaEvtGen_P2012_Wt_inclusive_top	410013	34.009	1.054	1.0
anti-top Wt -channel	PowhegPythiaEvtGen_P2012_Wt_inclusive_antitop	410014	33.989	1.054	1.0
top s -channel	PowhegPythiaEvtGen_P2012_SingleTopSchan_noAllHad_top	410025	2.0517	1.0046	1.0
anti-top s -channel	PowhegPythiaEvtGen_P2012_SingleTopSchan_noAllHad_antitop	410026	1.2615	1.0215	1.0

TABLE 5.4: Simulated background samples.

5.3.2 Background Samples

Several background processes are modelled using MC simulation, they are: Z+jets, di-top, single-top quark and Dibosons.

A summary of event generators, sample identification number (DSID) and cross section used the normalization of the various processes is provided in Table 5.4.

Z+jets

The Z+jets background samples have been produced with the same setup of the W+jets MG5_aMC@NLO+P8+EvtGen FxFx samples reported in previous section.

Di-top ($t\bar{t}$)

The $t\bar{t}$ background was produced using the POWHEG-BOX v2.2 [131, 132, 133, 134] generator at NLO with the NNPDF 3.0 NNLO PDF set and the h_{damp} parameter¹ set to $1.5 m_{\text{top}}$ [135]. This sample was interfaced to PYTHIA 8.230, using the A14 tune and the NNPDF 2.3 LO PDF set POWHEG+PYTHIA8. This sample is normalized to the NNLO cross-section, including the resummation of next-to-next-to-leading logarithmic (NNLL) soft gluon terms, calculated with TOP++2.0 [136, 137, 138, 139, 140, 141, 142].

Single top

Single top quark production, in association with W bosons (tW) as well as in the s- and t- channels, was also simulated using POWHEG-BOX v2.2 interfaced with PYTHIA 8.230 at NLO using the NNPDF2.3 LO PDF set. To remove overlap with the $t\bar{t}$ sample, the diagram-removal [143] procedure was used.

Dibosons (Db)

Semi-leptonic Diboson (VV) final states, where one W or Z boson decays leptonically and the other decays hadronically, were simulated using SHERPA v2.2.1 [144], with NLO MEs for up to one parton emission and LO MEs for up to three parton emissions calculated using the NNPDF 3.0 NNLO PDF set. The ME calculations are matched to the SHERPA parton shower using the MEPS@NLO [145, 146] prescription.

5.4 Selections

The final state for this analysis is defined by the presence of an isolated lepton (electron or muon), coming from the W boson decay, accompanied by sizeable missing transverse energy E_T^{miss} due to the presence of the neutrino from the decay, and a jet containing a muon coming from the decay of a c-hadron. The criteria used to reconstruct and select candidates for each of these objects and the final analysis selection are discussed below.

¹The h_{damp} parameter is a resummation damping factor and one of the parameters that controls the matching of POWHEG matrix elements to the parton shower and thus effectively regulates the high- p_T radiation against which the $t\bar{t}$ system recoils.

5.4.1 Primary Vertices

Primary vertices (PV) are reconstructed using tracks with p_T higher than 400 MeV, and must be associated with at least two tracks. If more than one vertex satisfying this selection is found, the one with the highest $\sum p_T^2$ of associated tracks is used as the PV of the event.

5.4.2 Electrons

Electron candidates are reconstructed from energy deposits (clusters) in the electromagnetic calorimeter associated to reconstructed tracks in the Inner Detector. Candidates in the calorimetry region with $|\eta_{cluster}| < 2.47$ is requested, with the exclusion of the transition region $1.37 < |\eta_{cluster}| < 1.52$. Electrons must pass the *TightLH* identification criterion and further selections on the transverse and longitudinal impact parameters: $|z_0 \sin \theta| < 0.5$ mm and $|\frac{d_0}{\sigma(d_0)}| < 5$. To further reduce the background from non-prompt electrons, conversions and hadrons, electron candidates are also required to be isolated and must pass the *Tight_VarRad* isolation working point [147]. For the di-lepton veto (see Sec. 5.4.11) electrons passing the *MediumLH* identification WP and the *Loose_VarRad* isolation WP are used.

In simulation, when the truth origin of the electron is needed, it is determined using the *MCTruthClassifier* tool [148].

5.4.3 Muons

In this section we discuss the reconstruction of muons coming from W boson decays. Muon candidates are reconstructed from track segments in the various layers of the muon spectrometer, and matched with tracks found in the Inner Detector. The final muon candidates are refitted using the complete track information from both detector systems.

Muons must pass the *Medium* identification working point and further selections on the transverse and longitudinal impact parameters: $|z_0 \sin \theta| < 0.5$ mm and $|\frac{d_0}{\sigma(d_0)}| < 3$. To further reduce the background from non-prompt muons, conversions and hadrons, muon candidates are also required to be isolated and must pass the *PflowTight_FixedRad* isolation working point [147]; the *PflowLoose_FixedRad* WP is used for muons considered in the dilepton veto (see Sec. 5.4.11).

Also for the muon, in simulation, when needed the truth origin of the muon is determined using the *MCTruthClassifier* tool [148].

5.4.4 Jets

Candidate jets are reconstructed clustering Particle Flow objects [149] using the anti- k_T jet algorithm [150] with a radius parameter of 0.4. The Particle Flow jet collection used in the analysis is the *AntiKt4EMPFJet_BTagging201903*. A jet energy scale derived from 13 TeV data and simulation [151] is applied as calibration on the jets. Only jets satisfying $p_T > 20$ GeV and $|\eta| < 2.5$ are considered.

Quality criteria (also called *jet cleaning procedure*) are imposed to identify jets arising from non-collision sources or detector noise (using the LooseBad operating points) and any event containing at least one such jet is removed, which causes a negligible loss of efficiency [152].

To avoid selecting jets from secondary pp -interactions, an additional requirement on the Jet Vertex Tagger is made for low p_T ($p_T < 60$ GeV) jets in the central ($|\eta| < 2.4$)

region of the detector: such jets must have $JVT > 0.5$. This requirement, applied both to data and simulation, correspond to the Tight WP, as recommended by the JetEtMiss group [153]. Since the recommendation are not explicit for $|\eta| > 2.4$, we decide to not apply any JVT cut in this region.

In simulation, in cases where the particle-level flavour of the jet must be assessed, this is determined using particle-level information, as discussed in Sec. 5.4.12.

5.4.5 Overlap removal

The following overlap removal procedure is applied for electrons, muons and jets defined as discussed above:

- any electron found with a track overlapping with any other electron is removed;
- any calorimeter muon found to share a track with an electron is removed;
- any electron found to share a track with a muon is removed;
- any jet found within a ΔR of 0.2 of an electron is removed;
- any electron subsequently found within ΔR of 0.4 of a jet is removed;
- any jet with less than 3 tracks associated to it found within ΔR of 0.2 of a muon is removed;
- any jet with less than 3 tracks associated to it, which has a muon inner-detector track ghost-associated to it, is removed;
- any muon subsequently found within ΔR of 0.4 of a jet is removed.

An additional muon-jet overlap removal procedure is applied due to a bug in the PF jet reconstruction in release 21, see [154, 155, 156]. Following ATLAS conventions, for the overlap removal ΔR quantities are calculated using rapidities.

5.4.6 Soft muon tagging

Soft muon tagging is a heavy flavour jet tagging technique based on selecting soft muons associated with semileptonic decays of heavy flavour hadrons. Muons passing the *Tight* identification WP, with $p_T > 4$ GeV matched with a jet within $\Delta R < 0.4$ (with ΔR defined using the rapidity) are selected. Of course, no overlap removal procedure and isolation criteria are applied for these muons, since they have to be inside and hence close to the jets. Very loose cuts are applied on the transverse and longitudinal impact parameters, to remove bad quality muons and muons produced in pileup collisions: $|z_0 \sin \theta| < 3$ mm and $|d_0| < 3$ mm.

5.4.7 *b*-tagging

Jets are identified as originating from the hadronization of a *b* quark (*b*-tagged) via an algorithm using multivariate techniques to combine information from the impact parameters of displaced tracks as well as topological properties of secondary and tertiary decay vertices reconstructed within the jet.

In this analysis, *b*-tagging is performed by placing a selection requirement on the weight from the DL1r tagger, which is trained on simulated $t\bar{t}$ events to discriminate *b*-jets from a background consisting of light-jets and *c*-jets.

5.4.8 Missing transverse momentum

The missing transverse momentum in the event, E_T^{miss} , is defined as the magnitude of the negative vector sum p_T of all selected and calibrated physics objects in the event, with an extra term added to account for soft energy in the event that is not associated to any of the selected objects. This soft term is calculated from inner detector tracks matched to the primary vertex to make it more resilient to pile-up contamination.

5.4.9 Trigger

The following triggers are used in the electron channel [157]:

- 2015: HLT_e24_lhmedium_L1EM20VH OR HLT_e60_lhmedium OR HLT_e120_lhloose;
- 2016-18: HLT_e26_lhtight_nod0_ivarloose OR HLT_e60_lhmedium_nod0 OR HLT_e140_lhloose_nod0 OR HLT_e300_etcut.

The muon channel uses the following triggers [158]:

- 2015: mu20_iloose_L1MU15 OR mu50;
- 2016-18: mu26_ivarmedium OR mu50.

5.4.10 Calibrations and corrections

Jets momenta are calibrated both in data and simulations following the official prescriptions from the JetEtMiss group [159]. The following corrections and calibrations are applied in simulated events only:

- event samples are reweighted so that the distribution of number of pile-up interactions per bunch crossing, μ , reproduces the one measured in data;
- lepton trigger efficiencies are corrected using scale factors to match the performance measured in data using $Z \rightarrow ee$ [160] and $Z \rightarrow \mu\mu$ [161] events, following official recommendations [162, 163] using the [164];
- muon efficiencies (reconstruction+identification, isolation, track-to-vertex association for W -leptons; identification for the soft muons) are corrected using scale factors, and their momenta are calibrated, in order to match the performance measured in data in $Z \rightarrow \mu\mu$ and $J/\Psi \rightarrow \mu\mu$ events [165, 166]; official corrections from the Muon Combined Performance (MCP) group are applied [117];
- electron efficiencies (reconstruction, identification, isolation) are corrected using scale factors, and their momenta are calibrated, in order to match the performance measured in data in $Z \rightarrow ee$, $W \rightarrow ev$ and $J/\Psi \rightarrow ee$ events [167]; official corrections from the EGamma group are applied [162];
- the efficiency of the JVT cut on jets is corrected applying the official scale factor recommended in [153];
- when the DL1r tagging algorithm is applied, dedicated scale factors (SFs) from the Flavour Tagging (FTAG) group are used to correct its performances to match those measured in data [168].

For the soft muon calibration, past studies [169, 170] show that the standard MCP SFs for the muons can be applied also to soft muons. The mistag rate of the SMT needs a data-driven calibration, but since the most relevant fake-dominated backgrounds (multijet and V +light-flavour jets) are estimated with data-driven techniques, any mistag rate correction that may be needed is calculated within the background estimation procedure.

5.4.11 Event Selection

Using the object definitions mentioned above, the following event selection is applied for the signal region:

- the event must have a PV and satisfy the trigger requirements;
- W -lepton selection:
 - electron channel: exactly one electron with $p_T > 27$ GeV and $|\eta| < 2.5$ passing the *TightLH* identification (id) working point (WP) and *Tight_VarRad* isolation (iso) WP ;
 - muon channel: exactly one isolated muon with $p_T > 27$ GeV and $|\eta| < 2.5$ the *Medium* id WP and *PflowTight_FixedRad* iso WP;
 - events are required to have no additional electron with $p_T > 15$ GeV and $|\eta| < 2.5$ passing the *MediumLH* id WP and *Loose_VarRad* iso WP;
 - events are required to have no additional muon with $p_T > 10$ GeV and $|\eta| < 2.5$ passing the *Medium* id WP and *PflowTight_FixedRad* iso WP (of course, these are muons passing the OR procedure, so they will be all at $\Delta R > 0.4$ from jets, and these are not those considered for the SMT tagging later);
- $E_T^{miss} > 30$ GeV² and $m_T^W > 45$ GeV;
- one or two hadronic jets in the event;
- exactly one SMT muon must be identified in the event (and therefore the event will contain exactly 1 SMT jet);
- the SMT jet is required to have a b -tagging weight $DL1r < 4.56$ (more details in the following section 5.4.11).

In the muon channel, in order to further reduce the impact of a bug in the P-Flow jet reconstruction in release 21 (see [152]), which can create a fake jet around the sub-leading muon from Z decay and results in a large contamination from the Z + jets background, the following cuts are applied:

- we explicitly required the soft muon to fail the loose isolation WP;
- $p_T^{balance} = \frac{p_T^Z}{p_T^{lep} + p_T^{soft-\mu}} > 0.4$;
- $(EMF < 0.8 \parallel Ntracks > 2)$;

²The value has been changed recently, so some final plots have 20 GeV instead of 30 GeV. In case, this will be indicated.

- the invariant mass of the hard-lepton and the soft muon $m_{\mu\mu}$ is requested to be greater than 4 GeV and outside the 8-11 GeV and 80-100 GeV windows, in order to remove the Y and Z mass contribution, which peaks are around 9.5 GeV and 90 GeV respectively.

Where “EMF” represents the electromagnetic fraction and “Ntracks” the number of ID tracks with $p_T > 500$ MeV associated with the jet.

Selecting only events passing the above-mentioned cuts rejects 68% of $Z + \text{jets}$ events, with an overall inefficiency on the $W + c$ signal of 3%.

Events are divided into Opposite Sign (OS) and Same Sign (SS) categories depending on the relative sign of the charge of the W -lepton and of the soft muon. MC $W + \text{jets}$ events are split in the following categories using truth-level information:

- $W + c$ signal: the SMT-tagged jet is c -jet ($flavor = 4$); soft muon originated by c -hadron or tau ($origin = 9, 25, 27, 28, 32$);
- $W + b$ bkg: SMT-tagged jet is a b -jet ($flavor = 5$);
- $W + l$ bkg: all remaining events.

l stands for light-flavour, and here we intend that the origin of the soft muon is from a light-hadron decay. The $W + l$ bkg can actually be split in two subcategories: $W + l$ c -jet bkg and $W + l$ light-jet bkg, with the latter one by far dominating. This is because in few cases the SMT jet can be a c -jet, but the soft muon is actually coming from a light-hadron decay inside the jet. We consider both in the same categories in our measurement because the soft muons have all the same properties (they are coming mostly from pion and kaon decays, so they have low quality tracks mostly removed by the usage of the Tight quality WP), and because these events cannot be modelled using SMT-filtered MC samples (see more details in Sec. 5.3.1).

A similar classification is used for $Z + \text{jets}$, but in that case for most uses we consider $Z + c$ and $Z + b$ together in a joint $Z + HF$ category, where HF stands for Heavy Flavour.

DL1r cut

In this section, some details on the DL1r cut are provided.

In summary, we refer with DL1r cut the following requirements:

- the SMT jet is required to have a b -tagging weight of $Jet_DL1r_w < 4.56$. This selection corresponds to a cut on the last bin of PCBT;
- the nominal FTag SF is applied on the jet, which corrects the flavour tagging efficiency independently from the production process (e.g. b jet from top or $W+b$)

The DL1r variable and the pseudo continuous b -tagging PCBT has been described more in details in Chapter 4, Section 4.1.3. In the PCBT distribution the last bin corresponds to 60%-0% of b tagging efficiency. Removing this bin means to remove the background processes with b jet in the final state, so mainly $W+b$, $t\bar{t}$, Single Top and Dibosons. These backgrounds have the biggest contribution in this bin: in the OS+SS selection, in fact, the $W+b$ has the 23%, the $t\bar{t}$ the 31% and the Single Top the 22%, as shown in Table 5.5, where the percentage of each sample in the last bin of PCBT are reported.

In the end, in Table 5.6 the sample lost efficiencies after the cut application are

TABLE 5.5: Percentage of background and signal samples for OS, SS and OS+SS selections, for the electron channel for 2016 in the last bin of PBCT.

Sample	OS[%]	SS[%]	OS+SS[%]
$W+c$	18	4	12
$W+l$	0.6	0.3	0.4
$W+b$	20	26	23
$W\tau$	1	0.7	0.8
$Z\tau$	0.4	0.6	0.5
$t\bar{t}$	29	35	31
<i>SingleTop</i>	22	21	22
Z	9	12	10
<i>Dibosons</i>	0.6	0.7	0.6

calculated, where the advantage of the DL1r cut is clear. The lost efficiency are calculated as the difference between the sample integral before and after the cut, normalized with the sample integral before the cut. With a signal loss of 5%, the background rejection corresponds to 57%, 55% and 57% for $W+b$, $t\bar{t}$ and *Single Top*, respectively. Notice that the $W+l$ is a very small fraction of the last PCBT bin (0.4%) with large fluctuations, for this reason the negative lost efficiency does not have to be considered pathological.

TABLE 5.6: Lost efficiencies of background and signal samples for the OS+SS selection, in the electron channel 2016, after the DL1r cut.

Sample	Lost Efficiency[%]
$W+c$	5
$W+l$	-0.08
$W+b$	57
$W\tau$	8
$Z\tau$	19
$t\bar{t}$	55
<i>SingleTop</i>	57
Z	28
<i>Dibosons</i>	15

5.4.12 Definition of particle-level objects

The following setup is used for the definition of particle-level objects, which are needed for the unfolding:

- particle-level jets from the *AntiKt4TruthWZJets* collection are used. Particle-level jets with $p_T > 20$ GeV and $|\eta| < 2.5$ are considered. The *truth flavour* of jet, using the standard ATLAS labelling, is defined by selecting weakly decaying b - and c -hadrons with $p_T > 5$ GeV within $\Delta R < 0.3$ from the jet: if at least one b -hadron is found, the jet is labelled as b -jet; otherwise, if at least a c -hadron is found, it is labelled as c -jet; otherwise, the jet is labeled as light-jet;

- for the W -lepton, electrons and muons are dressed using final state radiation (FSR) photons in a $\Delta R < 0.1$ cone around them; the origin of the lepton is identified by the MCTruthClassifier tool [148]. To mimic the OR procedure and the lepton isolation, and to avoid double-counting of muons in the muon channel, electrons and muons within $\Delta R < 0.4$ (using rapidity) from any selected jet are not considered for the W -lepton selection;
- neutrinos originated by W bosons, as determined by the MCTruthClassifier tool, are considered;
- muons within $\Delta R < 0.4$ (using rapidity) from selected jets are used for the particle-level SMT.

5.4.13 Fiducial phase space for the measurement and particle level selection

Applying the definitions of particle level objects defined in section 5.4.12, the fiducial phase space ³ is defined by the following particle level selection:

- for the W -lepton, after the OR procedure with jets, exactly one electron or muon with $p_T > 27$ GeV and $|\eta| < 2.5$ in the event; it must be identified by the MCClassifier to be originated from a W -boson (*origin* = 12);
- the W -originated neutrino (*origin* = 12) with $p_T > 20$ GeV;
- $m_T^W > 45$ GeV calculated using the W -lepton and W -neutrino kinematics;
- exactly 1 soft muon within $\Delta R < 0.4$ (using rapidity) from a selected jet;
- the SMT-tagged jet must be identified as c -jet (*flavor* = 4);
- the MCTruthClassifier must identify the muon as originated by a c -hadron or a τ (likely to be originated by a HF-hadron decay) (*origin* = 9, 25, 27, 28, 32)
- particle-level quantities are defined with a OS-SS subtraction, where OS and SS are defined using the charge of the W -lepton and the soft muon. This removes additional not really physical contribution from Wcc events, which cannot be easily identified using parton-level quantities in MC.

The origin codes are defined in Ref. [171], and correspond to: 9 (τ), 25 (c -meson), 26 (b -meson), 27 ($c\bar{c}$ -meson), 28 (J/Ψ), 29 ($b\bar{b}$ -meson), 32 (c -baryon), 33 (b -baryon).

5.4.14 W +jets composition studies

The selection outlined in Sec. 5.4.11 is here studied on W +jets MC samples. In Table 5.7 we run on the FxFx non-SMT filtered samples 508982-508984 in the muon channel after the final analysis selection, and we classify the events based on the truth flavour of the SMT-jet and of the origin of the soft muon. These inclusive samples are used to study the composition to avoid bias from the SMT-filter applied at generator level, that would remove long-lived light-hadron decays, effectively biasing the phase space covered by the sample for this study. The soft muon is considered as from a

³The fiducial region or phase space is the space at particle level which is close enough to what is experimentally accessible by the detector.

HF hadron if the MCTruthClassifier classify its origin as 25 (c -meson), 26 (b -meson), 27 ($c\bar{c}$ -meson), 28 (J/Ψ), 29 ($b\bar{b}$ -meson), 32 (c -baryon), 33 (b -baryon). Muons with a truth-origin of 9 (τ) are in a separate category in this study, because in most cases in HF-jets these τ s will be originated by HF-hadron decays. In all other cases the muon is classified as from a light-hadron.

smt jet flavour: bjet

sm-origin	OS	SS	OS/SS
sm from HF	0.13799 ± 0.00479	0.13728 ± 0.00504	1.00522 ± 0.05080
sm from l	0.00417 ± 0.00077	0.00438 ± 0.00078	0.95293 ± 0.24474
sm from Tau	0.00439 ± 0.00086	0.00564 ± 0.00091	0.77770 ± 0.19760

smt jet flavour: cjet

sm-origin	OS	SS	OS/SS
sm from HF	4.73628 ± 0.04541	0.35856 ± 0.01141	13.20916 ± 0.43894
sm from l	0.17430 ± 0.00834	0.18516 ± 0.00893	0.94132 ± 0.06394
sm from Tau	0.04258 ± 0.00423	0.00299 ± 0.00118	14.25818 ± 5.82711

smt jet flavour: ljet

sm-origin	OS	SS	OS/SS
sm from HF	0.04102 ± 0.00451	0.01542 ± 0.00284	2.65990 ± 0.57070
sm from l	1.28954 ± 0.02532	1.09140 ± 0.02269	1.18154 ± 0.03378
sm from Tau	0.00263 ± 0.00131	0.00017 ± 0.00048	15.72693 ± 45.63077

TABLE 5.7: Muon channel composition obtained on FxFx non-SMT filtered W +jets samples 508982-508984 after the final event selection. Yields are normalised to 1 pb^{-1} .

5.4.15 Yields and control plots

In this section, the MC sample yields for 2015-16, 2017, 2018, OS and SS regions, with the associated statistic uncertainties, are reported in Table 5.8 for the electron channel and in Table 5.9 for the muon channel. The yields are shown at reco level after the selections described in Section 5.4.

In Figure 5.12 and 5.13 are shown control plots for muon and electron channel, respectively. The variables shown are: electron rapidity, electron p_T , soft-mu jet rapidity, soft-mu jet p_T , soft-mu rapidity, soft-mu p_T , for OS and SS selections. The disagreement between data and MC are due mainly to the multijet background, which is estimated through the likelihood fit, described in Section 6.1 and 6.2.

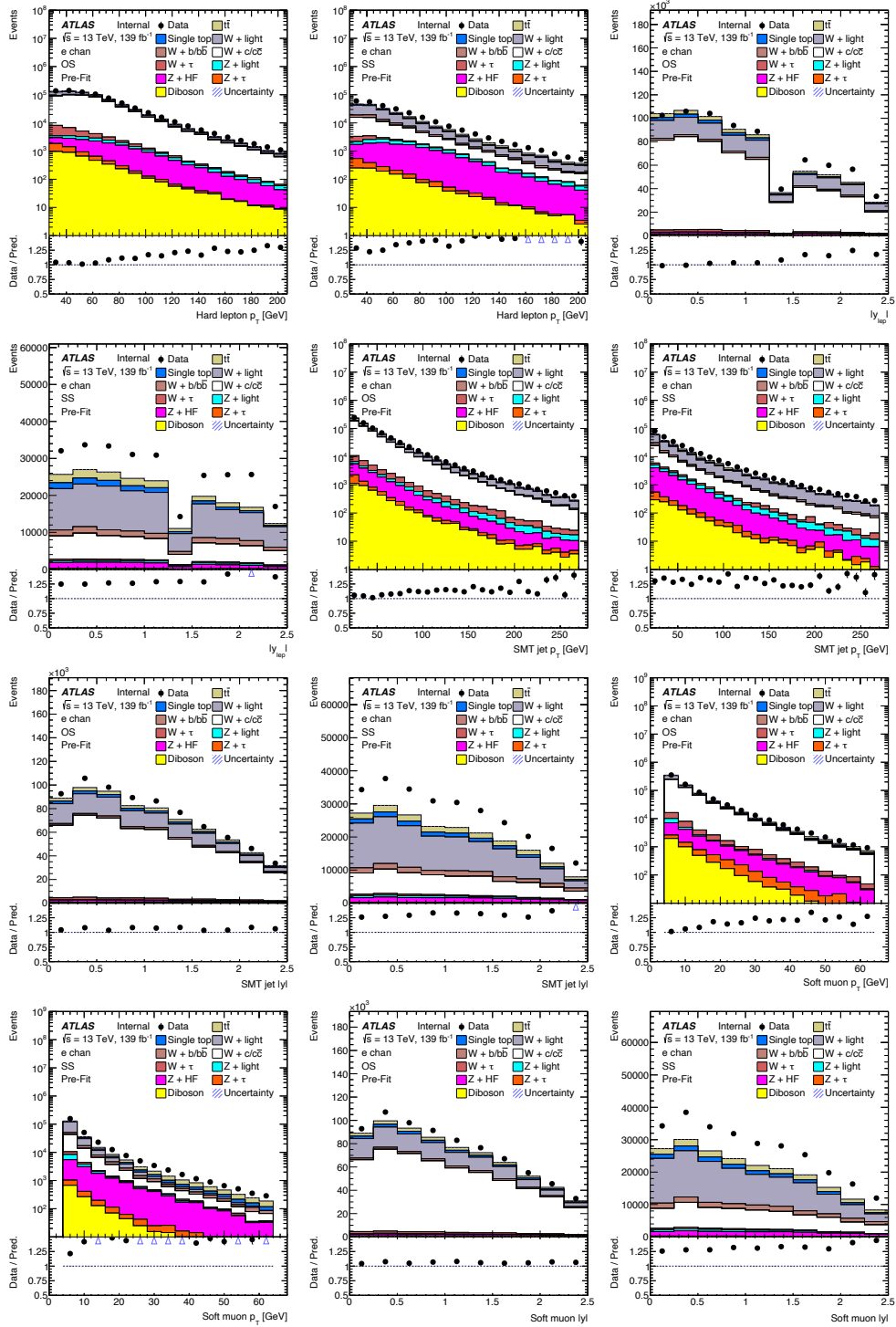


FIGURE 5.12: Control pre-Fit plots for the electron channel, full Run 2 statistic, for OS and SS selections. The variables are: electron rapidity, electron p_T , soft-mu jet rapidity, soft-mu jet p_T , soft-mu rapidity. The disagreement is mainly due to the lack of the multijet background, estimated in the fit.

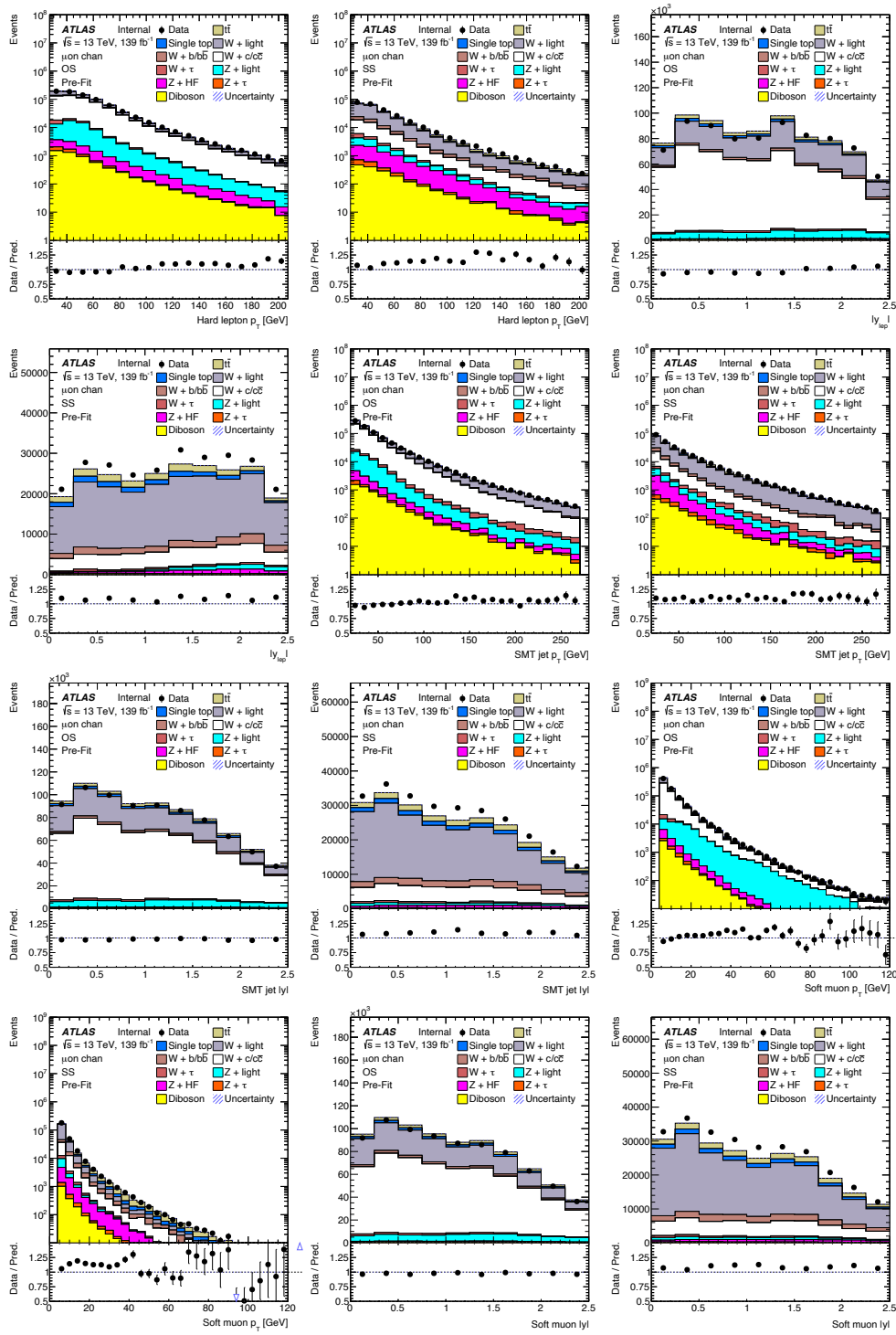


FIGURE 5.13: Control plots for the muon channel, full Run 2 statistic, for OS and SS selections. The variables are: muon rapidity, muon p_T , soft-mu jet rapidity, soft-mu jet p_T , soft-mu p_T , soft-mu rapidity.

TABLE 5.8: Yields for the electron channel, 2015-16, 2017 and 2018, OS and SS regions. The yields are calculated as the integral of the lepton rapidity distribution.

OS	2015-16	2017	2018
Data	199115 ± 446	240039 ± 490	310336 ± 557
$W+c$	134180 ± 1215	161295 ± 1377	210686 ± 2159
$W+l$	28520 ± 736	35586 ± 854	47304 ± 1038
$W+b$	3465 ± 146	4137 ± 167	6013 ± 204
$W\tau+jets$	3036 ± 205	4189 ± 223	5787 ± 257
$Z\tau+jets$	448 ± 30	719 ± 42	987 ± 48
ZHF	3247 ± 84	4013 ± 94	5166 ± 131
$Z+l$	1104 ± 49	1430 ± 57	1895 ± 70
top	4409 ± 18	5167 ± 20	6923 ± 23
$t\bar{t}$	7042 ± 31	7918 ± 34	7044 ± 32
Db	1155 ± 18	1294 ± 14	1735 ± 23
SS	2015-16	2017	2018
Data	71417 ± 267	86230 ± 294	111352 ± 334
$W+c$	13559 ± 368	17039 ± 418	22412 ± 757
$W+l$	24430 ± 663	28932 ± 752	37969 ± 896
$W+b$	3583 ± 151	4163 ± 164	5854 ± 187
$W\tau+jets$	823 ± 98	1043 ± 102	1275 ± 126
$Z\tau+jets$	190 ± 18	207 ± 22	315 ± 25
ZHF	3121 ± 82	3983 ± 95	5284 ± 112
$Z+l$	1029 ± 47	1431 ± 56	1863 ± 69
top	2786 ± 15	3216 ± 16	4322 ± 19
$t\bar{t}$	4944 ± 26	5567 ± 28	4964 ± 27
Db	345 ± 9	406 ± 6	516 ± 11

TABLE 5.9: Yields for the muon channel, 2015-16, 2017 and 2018, OS and SS regions. The yields are calculated as the integral of the lepton rapidity distribution.

OS	2015-16	2017	2018
Data	213132 ± 462	250727 ± 501	329366 ± 574
$W+c$	135287 ± 1293	161144 ± 1395	217919 ± 2324
$W+l$	41537 ± 974	53498 ± 1084	71569 ± 1398
$W+b$	4409 ± 199	5539 ± 288	6918 ± 261
$W\tau+jets$	3196 ± 216	3468 ± 230	5322 ± 263
$Z\tau+jets$	300 ± 28	475 ± 36	555 ± 39
ZHF	809 ± 52	2697 ± 108	3145 ± 106
$Z+l$	10838 ± 196	18100 ± 256	23913 ± 296
top	4464 ± 18	5089 ± 19	6853 ± 22
$t\bar{t}$	5966 ± 29	6463 ± 30	8785 ± 36
Db	1355 ± 18	1823 ± 15	2420 ± 25
SS	2015-16	2017	2018
Data	70593 ± 266	84267 ± 290	110225 ± 332
$W+c$	135287 ± 1293	161144 ± 1395	217919 ± 2324
$W+l$	41537 ± 974	53498 ± 1084	71569 ± 1398
$W+b$	4409 ± 199	5539 ± 288	6918 ± 261
$W\tau+jets$	3196 ± 216	3468 ± 230	5322 ± 263
$Z\tau+jets$	300 ± 28	475 ± 36	555 ± 39
ZHF	809 ± 52	2697 ± 108	3145 ± 106
$Z+l$	10838 ± 196	18100 ± 256	23913 ± 296
top	2885 ± 15	3278 ± 15	4380 ± 18
$t\bar{t}$	4184 ± 25	4456 ± 25	6096 ± 30
Db	434 ± 9	591 ± 10	769 ± 12

Chapter 6

Statistic analysis and results

This chapter is dedicated to the description of the statistic $W+c$ analysis and shows the latest results. More specifically, Section 6.1 presents the Likelihood Fit theory and strategy; Section 6.2 and 6.3 are dedicated to the multijet and $W+l$ background estimation; the fit results on the full Run 2 statistic are then shown in Section 6.4; Section 6.5 is dedicated to the Unfolding procedure, with a brief introduction of the theory and then showing the post-unfolding results; in the end, Section 6.7 presents the first results of the lepton rapidity distribution post-unfolding with statistic and systematic uncertainties and the integrated cross section.

6.1 Maximum Likelihood Fit

Most analyses, especially those have more data available, are *shape analyses*: they are based on binned variable distributions, which depend on several parameters. Usually, the parametrization consists in: the Parameter Of Interest (POI) μ , i.e., the strength of the signal, and other parameters, the Nuisance Parameters (NPs), i.e., the background normalization.

In this analysis, the global likelihood function $\mathcal{L}(n|\mu, \theta)$ has been chosen for the estimate of those parameters. Imagine to have a sample, $X=(X_1, \dots, X_n)$, which have a joint probability $f(x|\theta)$, where θ is a parameter and x the observation. The likelihood function is defined as $\mathcal{L}(\theta|x)=f(x|\theta)$. Differently from the probable distribution, the sample x is fixed and the parameter varies over possible values.

In a *binned likelihood distribution*, each bin is considered independent and the global distribution is the products of the probability density functions (pdfs) of each bin:

$$\mathcal{L}(n|\mu, \theta) = \prod_{i \in \text{bins}} P(n_i|\mu S_i(\theta) + B_i(\theta)) \quad (6.1)$$

Where P , in this case, has a Poisson distribution and θ is a nuisance parameter. Other NPs can be added on the background as *free parameters*, in the same implementation as the POI: $B_i(k, \theta) = k B_i(\theta)$. The best fit result is provided by a procedure of maximization of the likelihood respect the POIs and the nuisance parameters k . More statistical information about the likelihoods and the statistic tests can be found at this reference [172].

In this analysis, the TRExFitter framework [173] has been chosen for the binned likelihood fit implementation. The TRExFitter framework is built using the *HistFactory* format [174], which creates statistic models starting from binned histograms. The statistical framework and utilities are implemented by RooFit [175] and RooStats [176], which are ROOT libraries.

The TRExFitter asks as input n-tuples or histograms (in our case we are using histograms) and a *configuration* (config) file. The config file has a block-structure: the *Job*

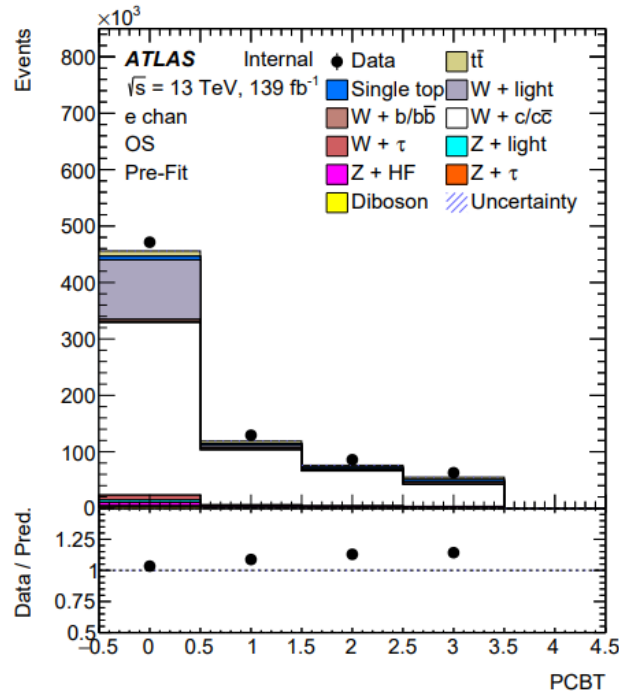


FIGURE 6.1: The PCBT distribution for the electron channel, OS region, Run 2 statistic.

block, which has the general option of the fit; the *Fit* block, which manages the fit options; the samples, MC and data, are indicated in the *Sample* block; the signal and control regions are defined in the *Region* block. All the details about the settings can be found in this reference [173].

6.1.1 Fit Strategy

The fit strategy consists in a simultaneous signal-background fit. The fit goal is to estimate the signal shape, the multijet background and the W +light normalization. The fit is developed in two dimensions:

- the variable of interest for the unfolding, by default the lepton rapidity ($\text{lep } y$);
- the pseudo-continuous b -tagging variable (PCBT), which consists in 5 bins corresponding to different requirement on the b -tagging algorithm discriminant distribution (DL1r), ensuring a specific b -tagging efficiency ϵ_b ([100%,85%], [85%,77%], [77%,70%], [70%,60%], [60%,0%]). The variable is described in this reference [177] and the details of the calibration in this TWiki page [178].

In this fit, PCBT is divided in two bins: the first corresponds to [100%,85%] of b -tagging efficiency and the second to [85%,60%]. The bin corresponding to [60%,0%] is excluded, for improving background rejection as explained in Section 5.4.11. From now on, the two bins chosen are called BIN 0 and BIN X. An example of the PCBT distribution after the selections, for the electron channel, OS region, is shown in Figure 6.1.

In Figure 6.2 is reported the DL1r b -jet discriminant variable, which shows the separation between the light-flavour jets at low values from the b -jets, which instead pick at

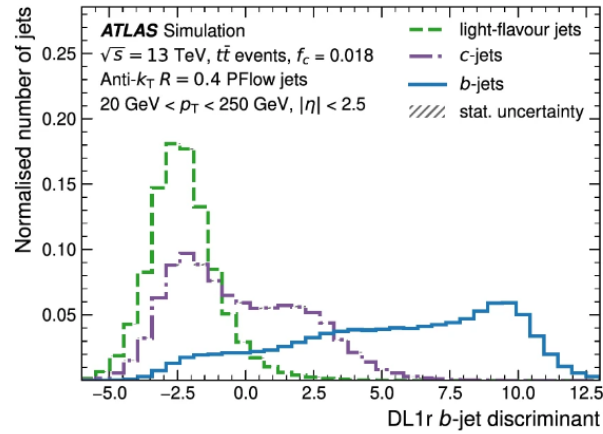


FIGURE 6.2: The DL1r b -jet discriminant variable in $t\bar{t}$ simulated event [56].

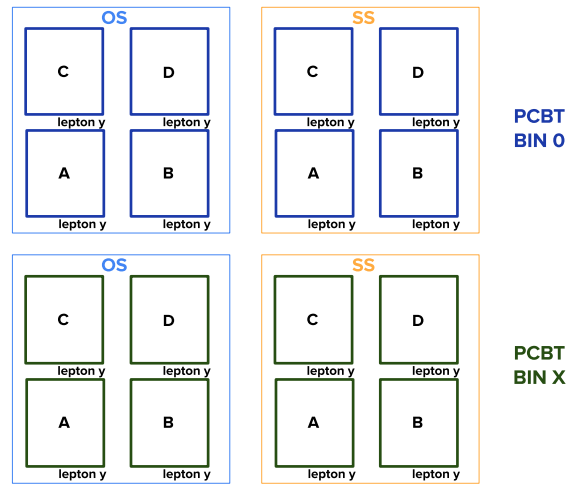


FIGURE 6.3: The schematic representation of the 16 regions composing the fit structure.

higher values. The idea behind the BIN 0 and BIN X division is to separate the light jets from the heavy flavour (HF) contribution. $W+l$ contribution is mainly for values of $DL1r < 0.665$, which corresponds to the first bin of PCBT, as notable in Figure 6.1. In this manner, the fit has a light-region, where it is possible estimate the $W+light$ normalization.

In total, 16 regions are created:

- 2 regions for the PCBT bins (BIN 0 and BIN X);
- 4 regions for the ABCD method for multijet estimation (A,B,C,D);
- 2 regions for the OS and SS division, required by the analysis strategy for signal estimation.

A schematic representation of the regions is shown in Figure 6.3.

In the TRExFitter config file, the normalization factors that have to be evaluated, NPs or POIs, are inserted as *NormFactors* Block. In this part, it is possible to indicate the

samples and the regions on which has to be applied, a nominal value and a range of max and min values in which fluctuate.

For this analysis, the following parameters are defined:

- *Norm0*: normalization factor in the BIN 0, for the SS Signal, $W+l$, $W+b$, $W\tau+jets$;
- *NormX*: normalization factor in the BIN X, for the SS Signal, $W+l$, $W+b$, $W\tau+jets$;
- *SignalNormX*: normalization factor in the OS-SS signal in BIN X.

Notice that the *Norm0* and *NormX*, together with the $W+l$, are applied also on the SS $W+c$, $W+b$ and $W\tau+jets$. The SS region, as it is possible to observe in a background only fit, shows an incomprehensible gap between data and MC, which the $W+l$ alone is not enough to cover it. Since the sample $W+jets$ include $W+l$, $W+b$, $W+c$ and $W\tau+jets$, we decided to include all in the factors.

The *SignalNormX* is added in the BIN X region, for helping the systematic estimation for the b -tagging variation.

In order to estimate a distribution, in the config file it is possible to add *ShapeFactors*. They, in practice, are a series of NormFactors applied on each bin. In this config file, 4 Shape Factors are added for the multijet shape estimation and 8 Shape Factors for the extrapolation factors (see Section 6.2).

In the end, in the fit the signal sample is already divided in the $W+c\bar{c}$ component, present at both OS and SS, and the other component which derives from the OS/SS subtraction. The OS-SS signal fit is obtained as a Shape Factor.

6.2 Multijet estimation

The multijet background results from the wrong identification of the prompt lepton from the W decay, also called *fake lepton*. The multijet background has different contributes:

- fake electrons:
 - heavy flavour quarks which decay semi-leptonically;
 - jets misidentified as electrons, usually it happens when they have few charged tracks and low energy deposits in the hadronic calorimeter;
 - photon conversion;
- fake muons:
 - heavy flavour quarks which decay semi-leptonically;
 - pions or kaons which decay in flight in the tracker;
 - punch-through hadrons, the shower of which exits the calorimeter and enter the muon spectrometer.

Monte Carlo generators can have difficulties in the multijet estimation for several reasons, such as the limited understanding of the QCD processes, simulation of hadronization with low precision, and consequently they have large uncertainties. An alternative approach involves a combination of Monte Carlo and experimental data and the technique is called *data-driven* method. Roughly, the data driven methods use regions dominated by the background which has to be determined, namely control

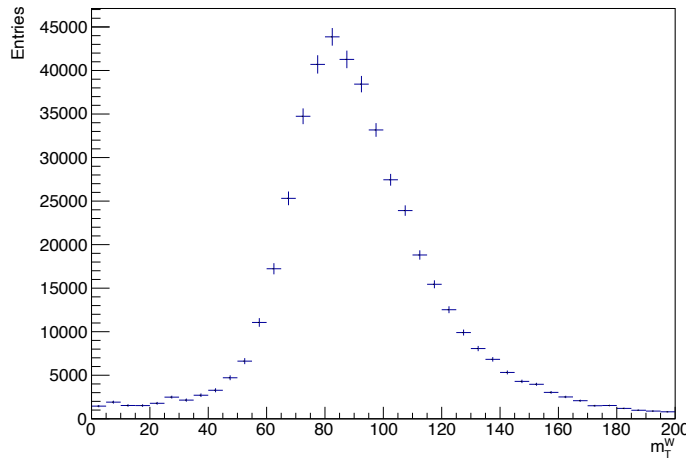


FIGURE 6.4: Transverse W mass distribution for the signal OS-SS MC sample, electron channel, full Run 2 statistic. The mean of the distribution is around 90 GeV, so the selection of 45 GeV guarantees to have high signal efficiency.

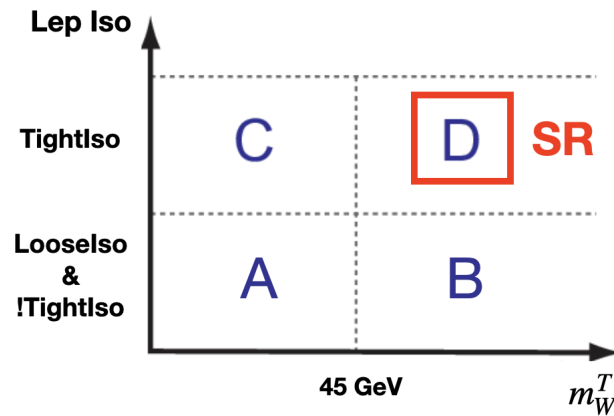


FIGURE 6.5: Schematic representation of the signal region D and the control regions A,B,C defined by the lepton isolation and m_W^T selections.

regions (CRs), to extrapolate the background in the region of interest, the signal region (SR).

The data-driven approach implemented in this analysis is the likelihood-base ABCD method.

ABCD Method

In this section, the canonical ABCD method [179] is presented, while in the next section the likelihood-base method, which is used in this analysis, is treated.

The ABCD method is largely used for the background estimation in analyses at high energy physics, due to its efficiency and the easy implementation.

The method requires the definition of one signal region, *region D*, and three orthogonal regions, *region A*, *B* and *C*. The control regions are chosen because enriched of

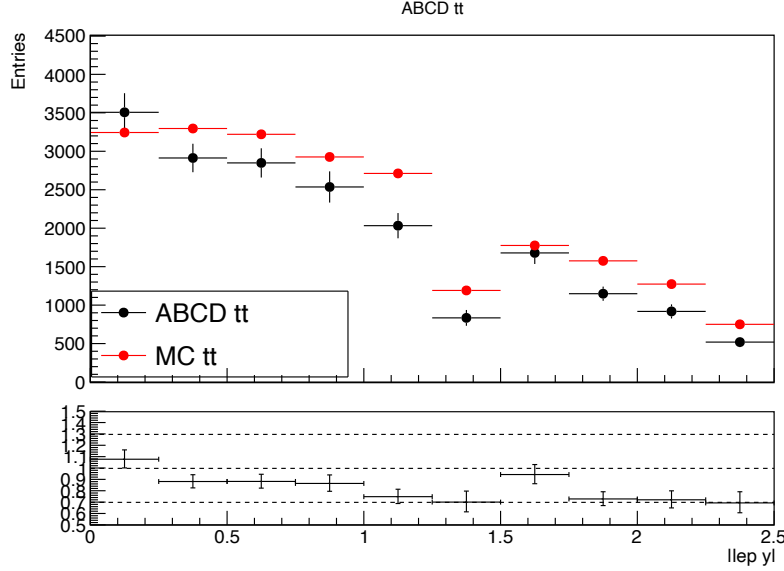


FIGURE 6.6: The $t\bar{t}$ lepton rapidity distribution for electron channel, using full Run 2 MC samples. In black the distribution obtained in the D region from the ABCD method, in red the MC sample.

background.

The regions can be defined from selections on continuous variables or on discrete binary requirements. In this analysis, the transverse mass of the W , m_W^T , and the quality of the lepton-isolation, iso , are chosen, which are continuous and discrete binary variables respectively. The m_W^T variable is chosen in order to separate the region where the lepton is from a W decay, above 45 GeV, from those regions where it comes from background sources, below 45 GeV. In figure 6.4, the transverse W mass for the W +jets MC sample is shown, where 97% of the events are distributed above 45 GeV.

Since the multijet background is the result of a wrong identification of jets or converted photons as prompt leptons, we expect the multijet objects to have some deposits around. The requirement of loose but not tight isolation should identify those events. The region with low transverse W mass and loose but not tight isolation, corresponding to region A, is the one enriched of multijet background.

The regions are then defined as following:

- Region D (Signal region): $m_W^T > 45$ GeV and `Ele_isFixedCutTightIso == 1`;
- Region A : $m_W^T < 45$ GeV and `Ele_isFixedCutLooseIso == 1 && Ele_isFixedCutTightIso == 0`;
- Region B: $m_W^T > 45$ GeV and `Ele_isFixedCutLooseIso == 1 && Ele_isFixedCutTightIso == 0`,
- Region C: $m_W^T < 45$ GeV and `Ele_isFixedCutTightIso == 1`.

A schematic representation of the four regions is shown in Figure 6.5. The method is based on the variable-independence assumption, which means that the multijet background distribution in the signal region is a linear combination of the multijet in

the control regions. Explicitly:

$$mj(D) = mj(A) \frac{mj(B)}{mj(A)} \frac{mj(C)}{mj(A)} \quad (6.2)$$

where $\frac{mj(B)}{mj(A)}$ and $\frac{mj(C)}{mj(A)}$ are called *extrapolation factors* A-to-B and A-to-C respectively. The A region, in fact, is ideally the pure-multijet one, while the multijet in the signal region is obtained from the A region multiplied by the two extrapolation factors. See next section for more details.

The assumption of the independence of the chosen variables is not straightforward to demonstrate. The naive idea is to test the ABCD on one MC sample, i.e., the $t\bar{t}$ background, and compare the background obtained by the ABCD method in the signal region with the MC one. This is not a conclusive test, since the different statistic of the control regions respect the signal one and also because the behaviour of this sample could be different from the multijet one. In Figure 6.6, the two $t\bar{t}$ sample distributions in the signal region are shown. In first approximation, it is possible to conclude that the method reproduce the $t\bar{t}$ MC sample, since the ratio showed in the plot is almost compatible with 1 within the uncertainty bars.

ABCD Likelihood-based Method

The ABCD method is implemented in the likelihood-base version using TRExFitter, as explained in the Fit strategy 6.1.1. The multijet in the signal region D is estimated starting from the likelihood fit in the control regions A, B, C.

The implementation has been inspired from this [180], which are provided by the TRExFitter community. In order to estimate the multijet in the signal region bin-by-bin, are defined 3 **shape-factors**:

- overall shape-factor, **dd_bkg** applied in all the 4 regions;
- one shape-factor as extrapolation factor from A to B region, **EFAtoB**, applied on the multijet in B and D regions;
- another shape-factor as extrapolation factor from A to C, **EFAtoC** applied on the multijet in C and D regions.

As multijet templates, four flat histograms, with the same normalization, are provided to the fit. The multijet in D region is estimated by the contribution of the **dd_bkg** shape-factor, mainly estimated in the A region, which is suppose to be the multijet-dominated one; the **EFAtoB** shape-factor, estimated simultaneously in the D and B region and the **EFAtoC** simultaneously in C and D region. The extrapolation factors are estimated starting from the template fitted in the A region. In this complex implementation, it is possible to find the ABCD formula:

$$D = A(B/A) \times (C/A), \quad (6.3)$$

under the hypothesis of variable independence:

$$\frac{D}{B} = \frac{C}{A} \quad (6.4)$$

An example of a pre-fit configuration is shown in Figure 6.7. The multijet sample in green is provided as flat template of 1000 events as fit input.

The multijet estimation test is provided by a simultaneous background only fit in

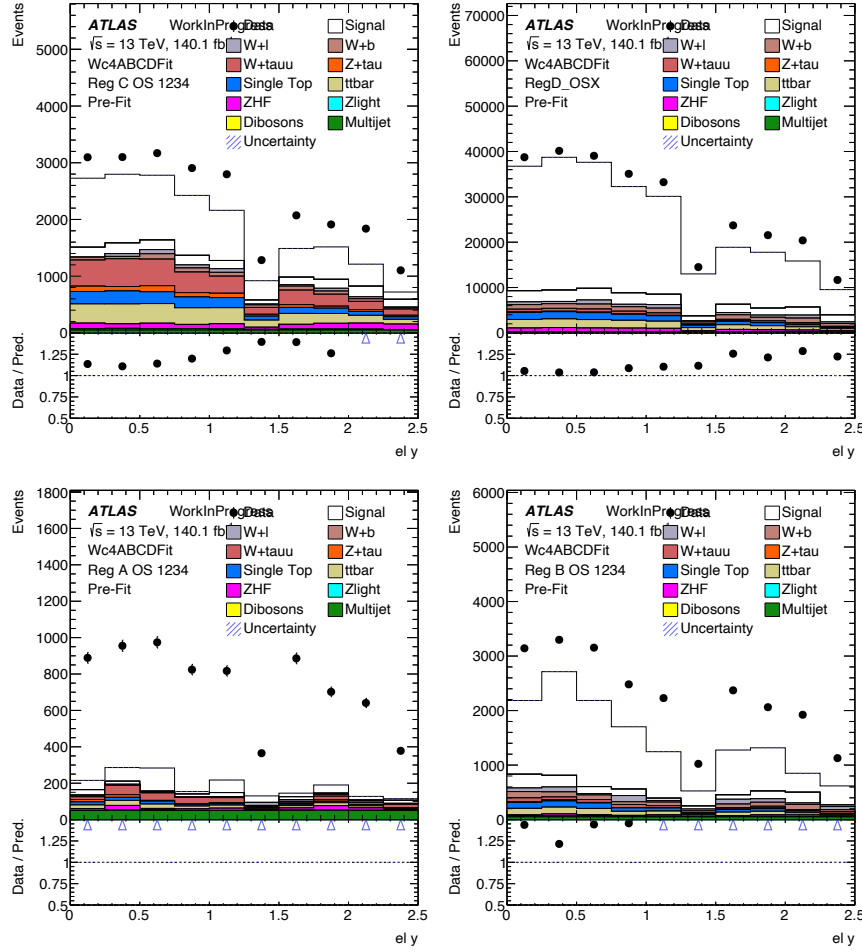


FIGURE 6.7: The four ABCD regions in the pre-fit configuration is reported for the electron channel, for the electron rapidity variable. The regions are for the opposite sign (OS) selection for the PCBT bin X. In this case the full Run 2 dataset is used.

A,B and C regions in order to have a prediction in D region, set in this case as *VALIDATION* region.

6.3 W+light background estimation

The $W+l$ events could pass the selection because the light jet is mistagged as a c -jet. For the charge conservation, the fake muon from the light jet has high probability to be correlated with the lepton from the W decay. Then, after the OS-SS subtraction, the $W+l$ has not negligible contribution. The W +light contribution is estimated by MC sample, but, as already mentioned in the section *Calibrations and corrections* 5.4.10, the normalization and the asymmetry are data-driven estimated as factors in the combined fit. Also in the previous $W+c$ analysis of ATLAS experiment, with Run I data, both normalization and asymmetry were estimated with a data driven method [108]. In this case a pre-tag level sample was used, with two different approaches for the normalization estimation in the electron and muon channel.

In this analysis, instead, the same approach is used for both channels.

The $W+l$ normalization is estimated through NormFactors (NFs), which account

for both the SMT muon data/MC mis-modelling and the SMT fake rate, i.e., the misidentification rate due to light-hadron decays and detector background. Since there is a MC/data different disagreement for BIN 0 and BIN X, two different NFs are applied in the bins, namely $Norm0$ and $NormX$. This difference cannot be only absorbed by the signal, since is supposed to be independent from the PCBT bins. From the background only fit, a not understandable data/MC hole in SS appears, more evident in the BIN 0, which can affect the OS/SS signal subtraction. Also for this reason, a normalization factor is added on the $W+c\bar{c}$, $W+b$ and $W\tau+jets$ samples. Since it is not the purpose on the analysis to estimate the precise value of each normalization, a common factor is defined.

For the $W+l$ asymmetry assessment, a validation region has been defined. The validation region has the same selections as the signal region, except for the E_T^{miss} , which is $20 \text{ GeV} < E_T^{miss} < 30 \text{ GeV}$, instead of $E_T^{miss} > 30 \text{ GeV}$. The fit has been done in this region, fixing the NFs and signal shape obtained from fit and adding a NF on the $W+l$ sample, SS region, for the asymmetry calculation, namely $W+l \text{ NF}_{Asym}$.

The values obtained from the fit are reported in Table 6.1.

From the $W+l \text{ NF}_{Asym}$, it is possible to derive the asymmetry A , which is defined as:

$$A = \frac{N_{Wl,OS} - NF_{Asym} N_{Wl,SS}}{N_{Wl,OS} + NF_{Asym} N_{Wl,SS}} \quad (6.5)$$

where $N_{Wl,OS/SS}$ are the number of $W+l$ events for OS or SS region in the validation region in the post fit. The values obtained are reported in Table 6.1 as well.

The results are not too precise, the uncertainty of 20% for the electron and 9% for the muon channel, but smaller in comparison with the Run I ATLAS results.

TABLE 6.1: Fit result for the Normalization Factor $W+l$ SS for muon and electron channel, full Run 2, obtained in the validation region of $20 \text{ GeV} < E_T^{miss} < 30 \text{ GeV}$.

Parameter	Electron	Muon
$W+l \text{ NF}_{Asym}$	1.136 ± 0.014	1.077 ± 0.008
Asymmetry	0.029 ± 0.006	0.053 ± 0.005

6.4 Fit results

In this section, the fit results are shown for the electron and muon channel separately, for the full Run 2 statistic. In particular, the NFs, $Norm0$, $NormX$, $SignalNormX$, and the Shape Factors obtained from the fit are presented.

In Figures 6.8, 6.9, 6.10, 6.11, the post fit regions are shown for the electron channel. In Figures 6.12, 6.13, 6.14, 6.15 the same plots for the muon channel. The fit has a good convergence as seen in the ratio and the indicative χ^2 below each plot.

In table 6.2 the number obtained by the fit of the Norm Factors are reported, for both electron and muon channels. It is possible to notice that electron and muon results are not compatible, for this reason we decided to keep the fit separate for the two channels. The uncertainty of the parameters, which is only statistical, is quite small, around 0.3-0.5%, index also of the goodness of the fit.

In the end, the signal OS-SS post fit is shown in Figure 6.16, in comparison with the MC sample distribution.

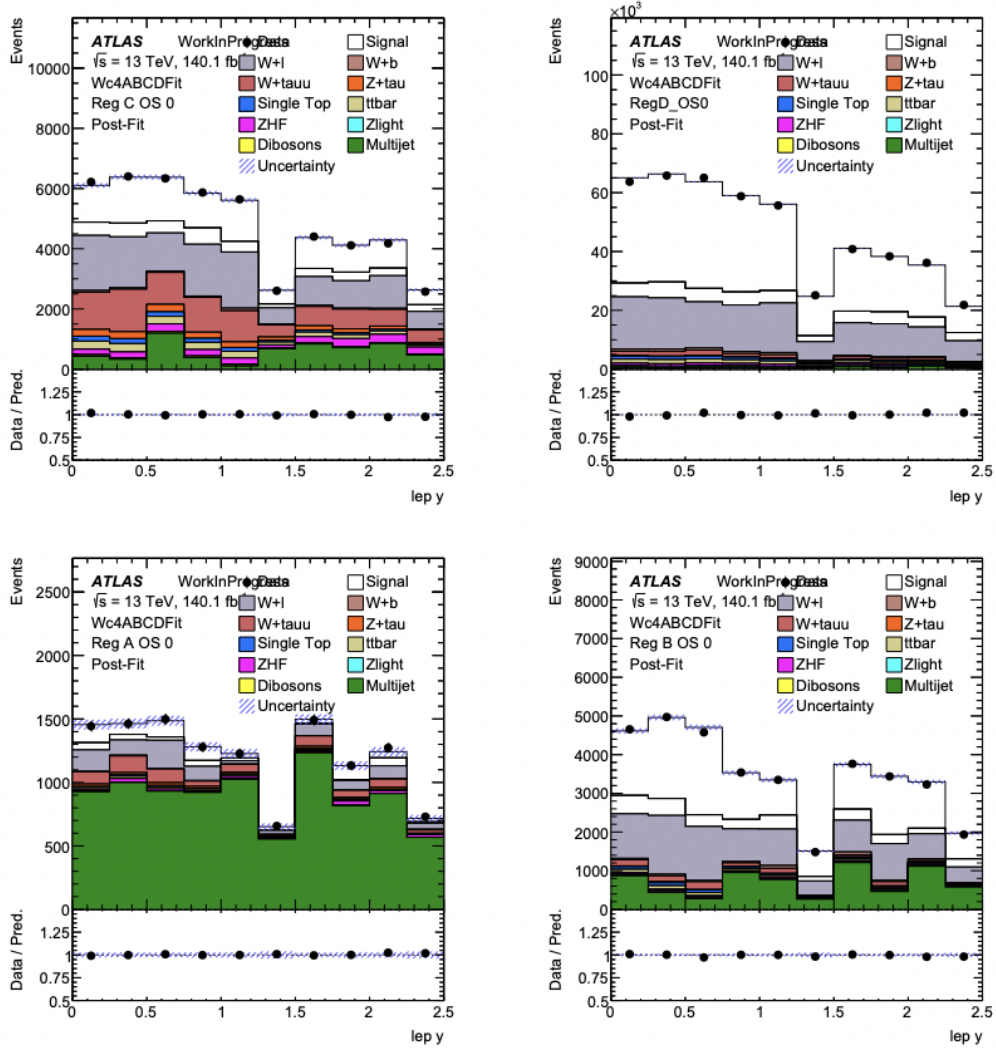


FIGURE 6.8: Post Fit sample distribution for the electron channel full Run 2. The region shown is BIN 0 OS divided in 4 (A,B,C,D) regions for the multijet estimation. The D region is the signal one, the white sample is the OS-SS signal, the grey is the $W+l$ and the green one the multijet.

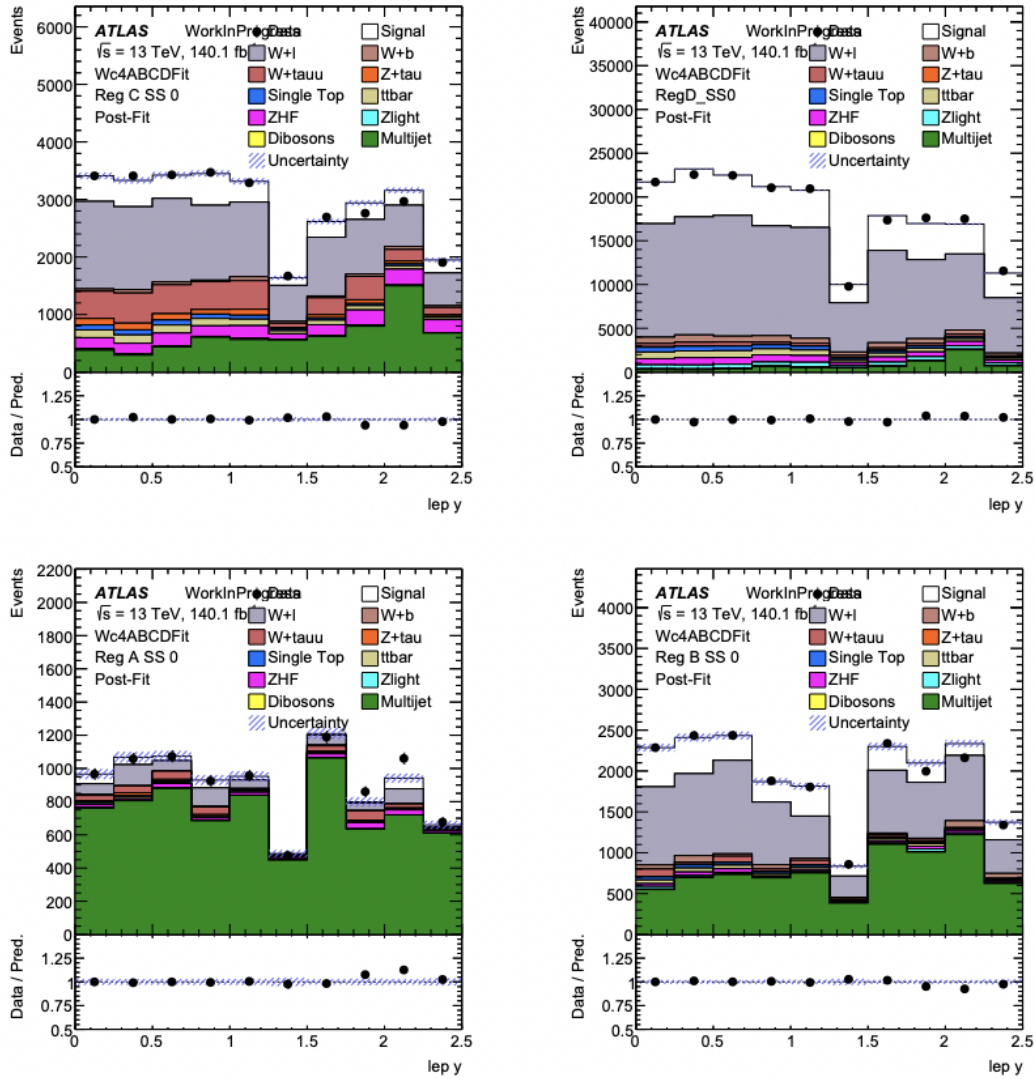


FIGURE 6.9: Post Fit sample distribution for the electron channel full Run 2. The region shown is BIN 0 SS divided in 4 (A,B,C,D) regions for the multijet estimation. The D region is the signal one, the white sample is the OS-SS signal, the grey is the $W+l$ and the green one the multijet.

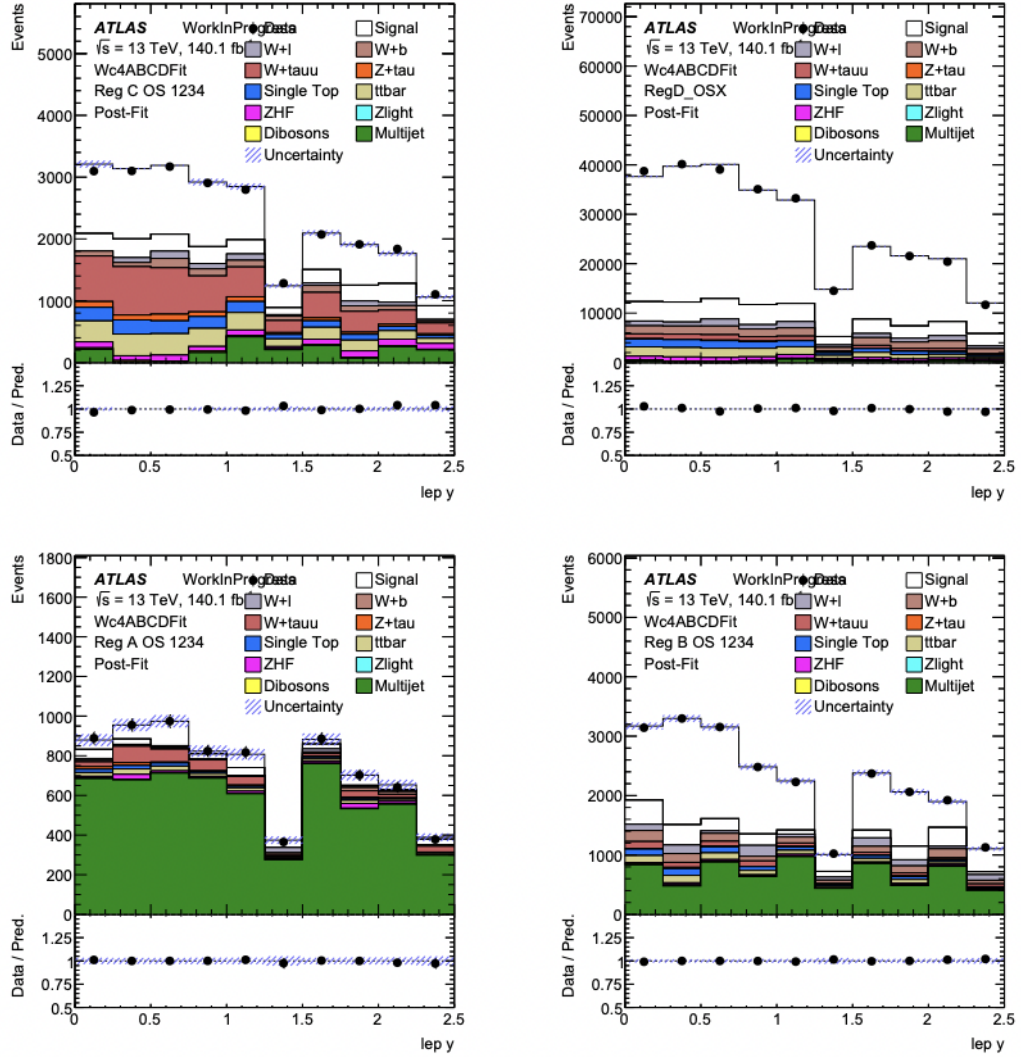


FIGURE 6.10: Post Fit sample distribution for the electron channel full Run 2. The region shown is BIN X OS divided in 4 (A,B,C,D) regions for the multijet estimation. The D region is the signal one, the white sample is the OS-SS signal, the grey is the $W+l$ and the green one the multijet.

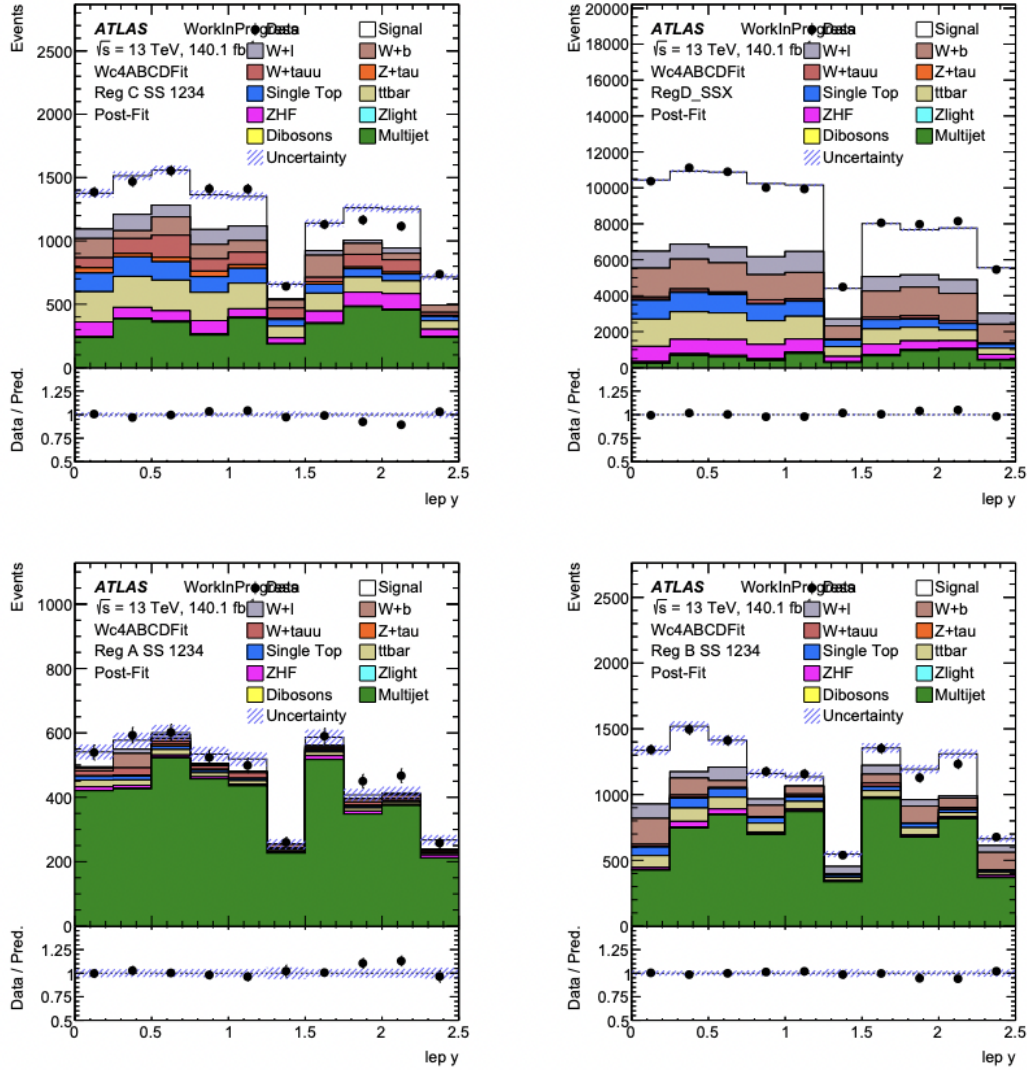


FIGURE 6.11: Post Fit sample distribution for the electron channel full Run 2. The region shown is BIN X SS divided in 4 (A,B,C,D) regions for the multijet estimation. The D region is the signal one, the white sample is the OS-SS signal, the grey is the $W+l$ and the green one the multijet.

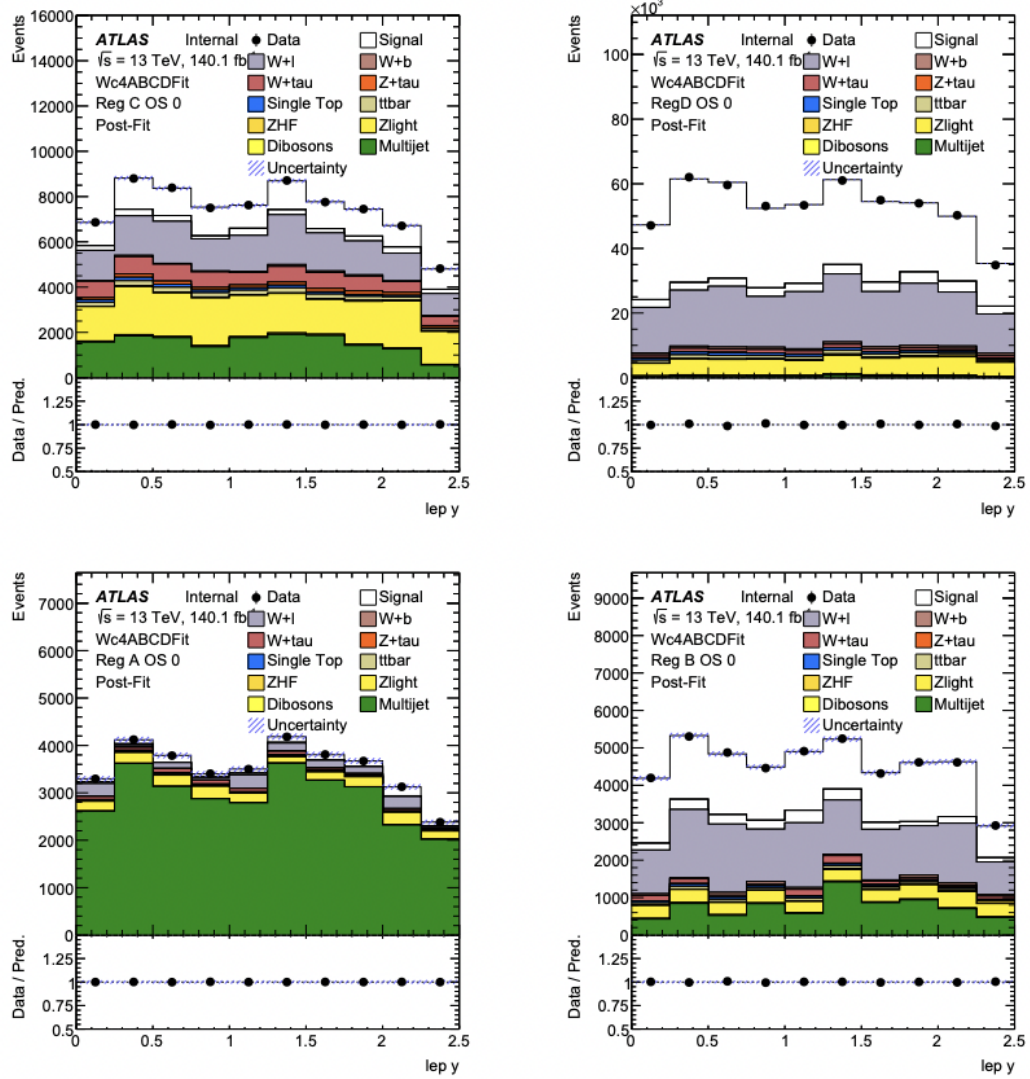


FIGURE 6.12: Post Fit sample distribution for the muon channel full Run 2. The region shown is BIN 0 OS divided in 4 (A,B,C,D) regions for the multijet estimation. The D region is the signal one, the white sample is the OS-SS signal, the grey is the $W+l$ and the green one the multijet.

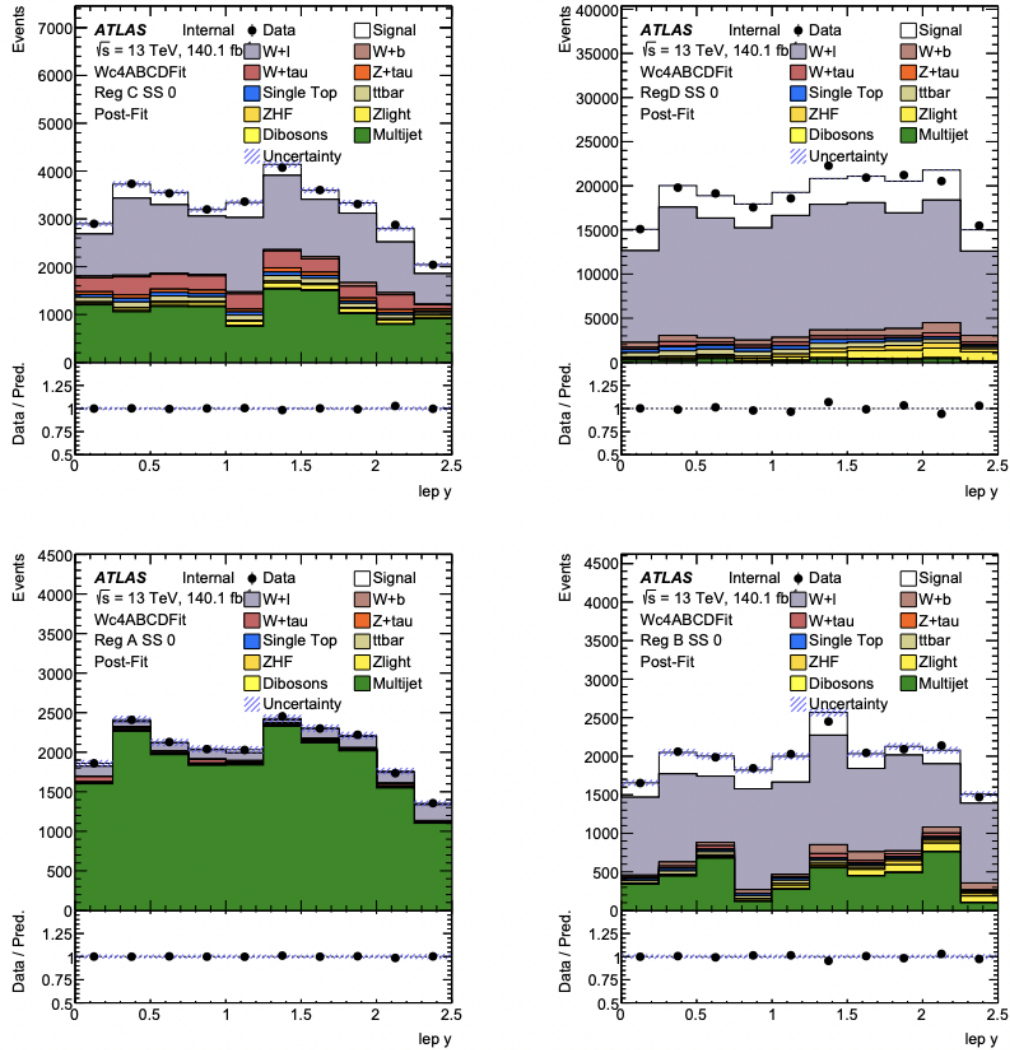


FIGURE 6.13: Post Fit sample distribution for the muon channel full Run 2. The region shown is BIN 0 SS divided in 4 (A,B,C,D) regions for the multijet estimation. The D region is the signal one, the white sample is the OS-SS signal, the grey is the $W+l$ and the green one the multijet.

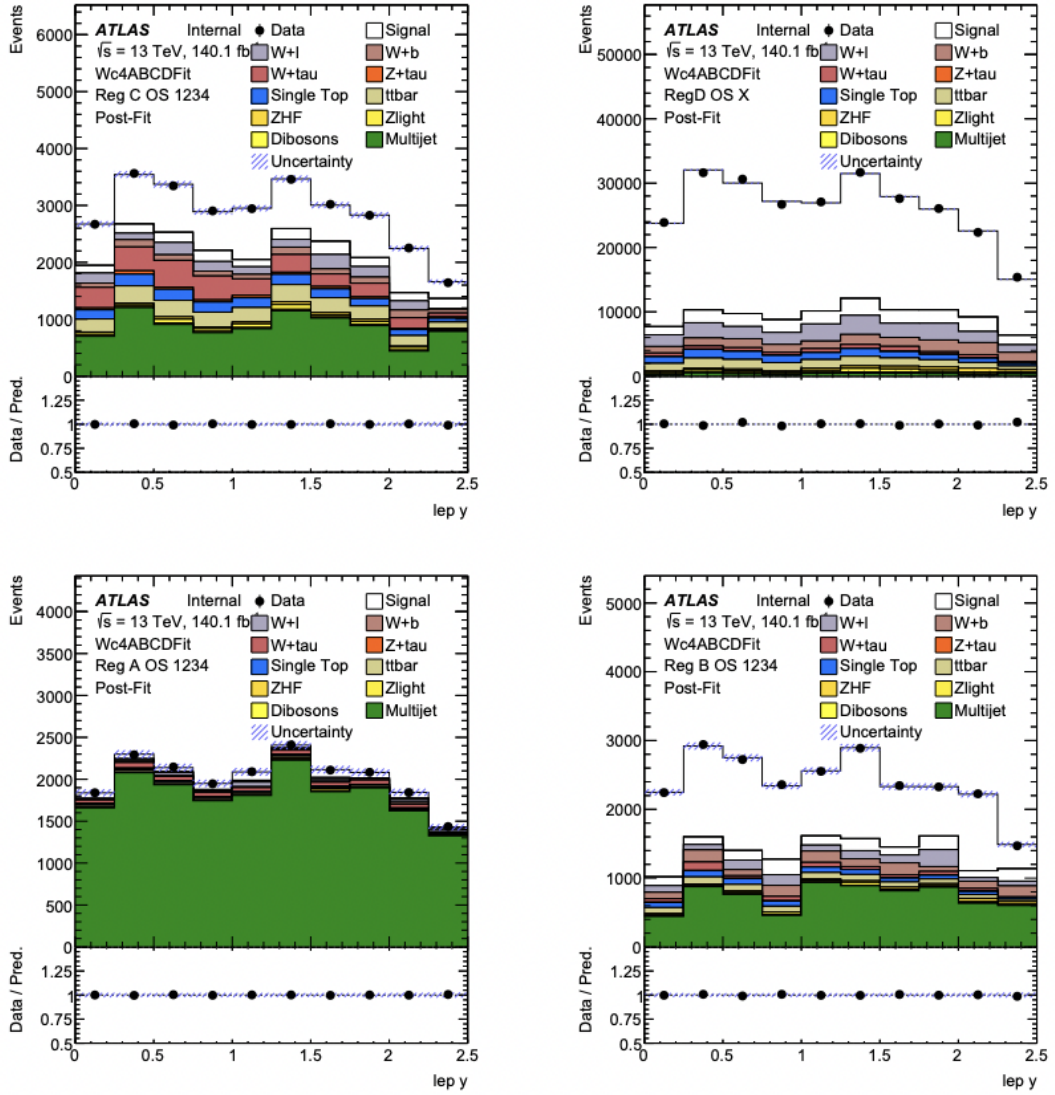


FIGURE 6.14: Post Fit sample distribution for the muo channel full Run 2. The region shown is BIN X OS divided in 4 (A,B,C,D) regions for the multijet estimation. The D region is the signal one, the white sample is the OS-SS signal, the grey is the $W+l$ and the green one the multijet.

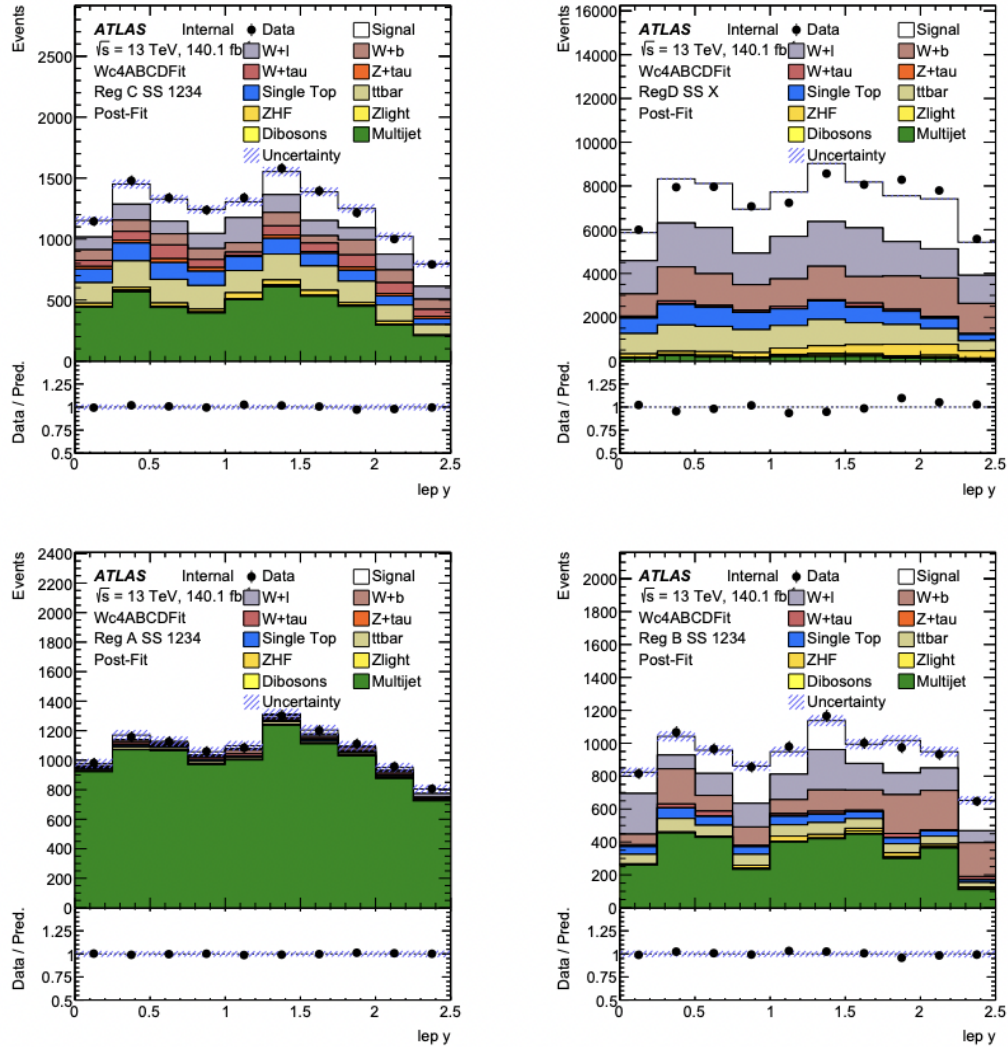


FIGURE 6.15: Post Fit sample distribution for the muon channel full Run 2. The region shown is BIN X SS divided in 4 (A,B,C,D) regions for the multijet estimation. The D region is the signal one, the white sample is the OS-SS signal, the grey is the $W+l$ and the green one the multijet.

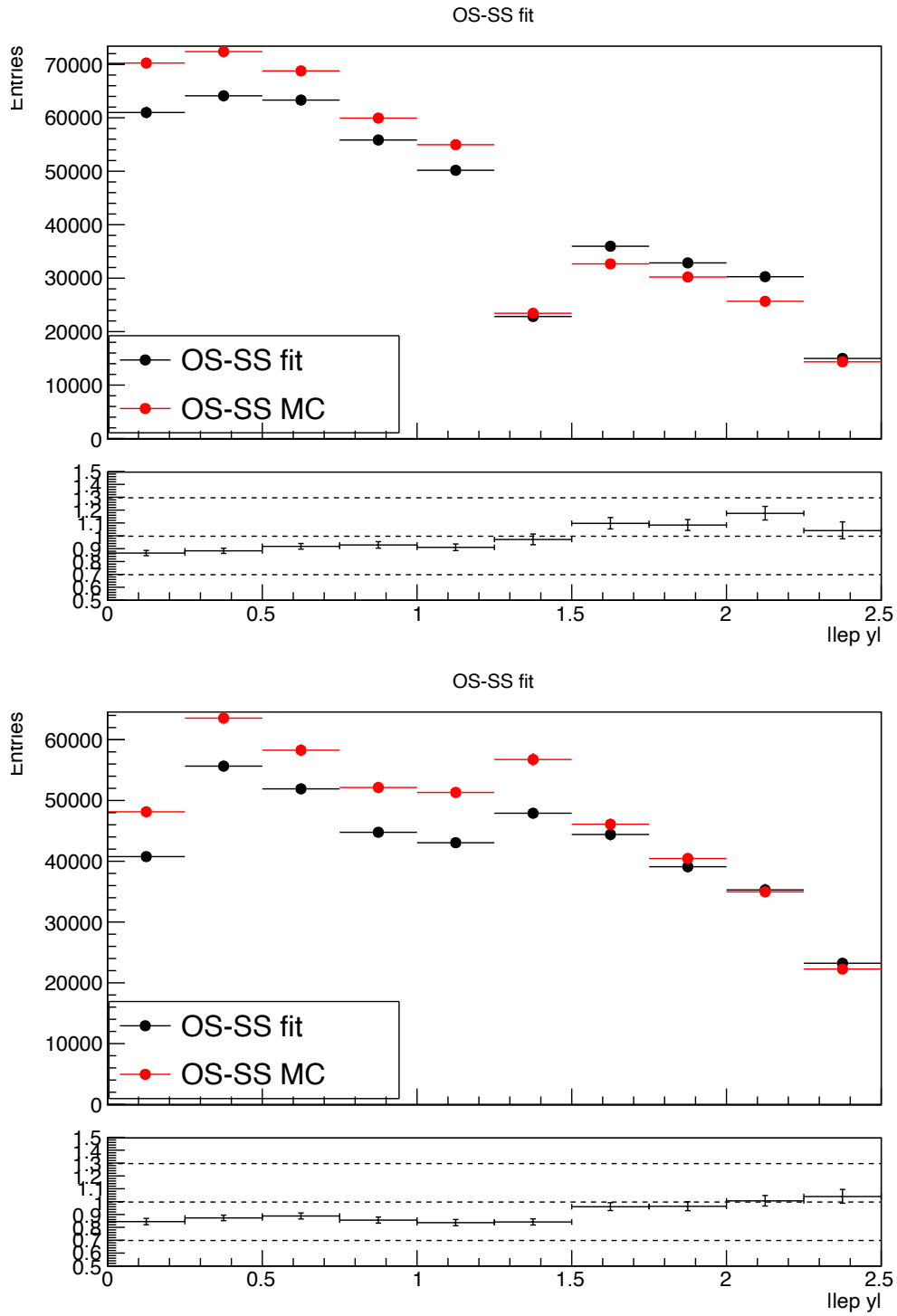


FIGURE 6.16: Post Fit distribution for the OS-SS signal in comparison with the MC sample for electron (up) and muon (down) channel, Run 2 statistic.

TABLE 6.2: Fit result parameters for muon and electron channel, full Run 2.

Parameter	Electron	Muon
Norm0	1.224 ± 0.005	1.036 ± 0.003
NormX	1.603 ± 0.013	1.268 ± 0.007
SignalNormX	1.105 ± 0.006	1.033 ± 0.005

TABLE 6.3: Fit result parameters for muon and electron channel, full Run 2, for the c -jet rapidity variable.

Parameter	Electron	Muon
Norm0	1.231 ± 0.005	1.036 ± 0.003
NormX	1.623 ± 0.012	1.263 ± 0.007
SignalNormX	1.108 ± 0.006	1.033 ± 0.005

Fit tests

TABLE 6.4: Fit result parameters for muon and electron channel, for each year of Run 2 (2015-16, 2017, 2018).

Ele Parameter	2015-16	2017	2018
Norm0	1.266 ± 0.009	1.189 ± 0.009	1.203 ± 0.007
NormX	1.39 ± 0.03	1.597 ± 0.003	1.624 ± 0.02
SignalNormX	1.100 ± 0.012	1.096 ± 0.012	1.128 ± 0.011
Muo Parameter	2015-16	2017	2018
Norm0	1.106 ± 0.006	1.017 ± 0.005	0.999 ± 0.004
NormX	1.330 ± 0.014	1.227 ± 0.013	1.251 ± 0.011
SignalNormX	0.999 ± 0.009	1.02 ± 0.01	1.065 ± 0.009

In order to test whether the fit works as expected, as closure test a Toy-Fit has been implemented.

In our case, the Toy-Fit is a procedure in which the output is supposed to be compatible with the input. More specifically, the data input are built as the sum of the MC samples and the multijet, where the multijet is simply the difference between data and MC. Since the data fits exactly on the samples by construction, the result should be 1 for the NFs and the post-fit OS-SS signal Shape Factor should give the MC OS-SS signal.

Form Table 6.5, it is evident that all the normalization factors obtained by the toy fit are compatible with 1. In Figure 6.17 the signal OS-SS obtained from the Toy-Fit is

TABLE 6.5: Fit result parameters for muon and electron channel, full Run 2.

Parameter	Electron	Muon
Norm0	1.000 ± 0.005	1.000 ± 0.003
NormX	1.000 ± 0.014	1.000 ± 0.007
SignalNormX	1.000 ± 0.006	1.000 ± 0.005

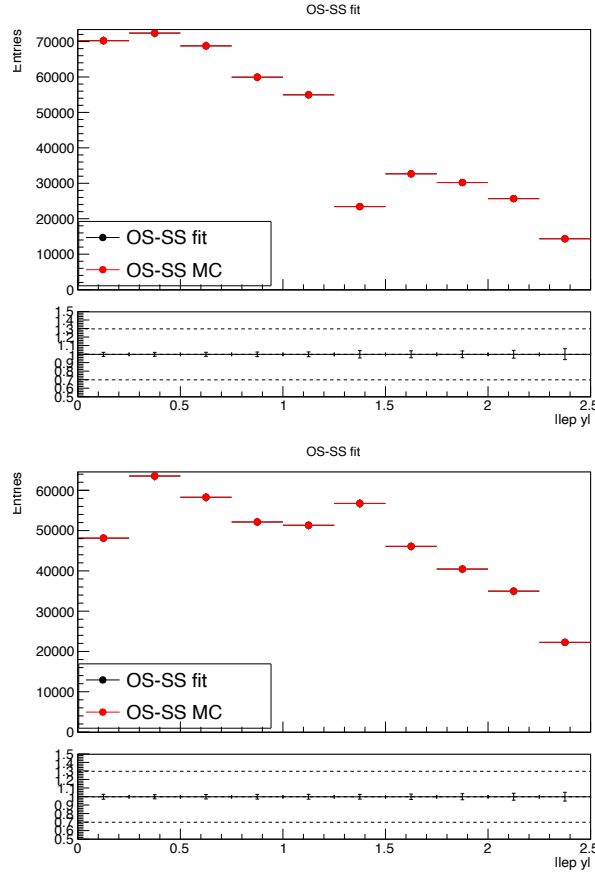


FIGURE 6.17: Post Toy-Fit distribution for the OS-SS signal in comparison with the MC sample for electron (up) and muon (down) channel.

compared with the MC sample and they are completely overlapped. It is possible to conclude that the fit is working properly and with good precision. We are not considering the error obtained from the Toy as source of systematic uncertainty, since it is negligible respect the others (see Section 6.6).

For testing the stability of the fit, it is possible to compare the fit results of the default variable, lepton rapidity, and those of an alternative variable, c -jet rapidity. In Table 6.3 the parameters for the alternative variable fit are reported and in Figure 6.18 and 6.19 the D regions for electron and muon channel respectively for the c -jet rapidity. If we compare the values with those shown in Table 6.2 with the lepton rapidity, they are compatible.

In the end, we can compare the fit results for each Run 2 year (2015-16, 2017 and 2018) in Table 6.4 and observe the stability of the values as well.

6.5 Unfolding procedure

The unfolding is the procedure widely used in the high energy physic experiments in order to obtain the *true* distribution from the *reco* or *measured* one. The *true* distribution of a physic quantity is the one that would be observed under ideal conditions with an ideal detector. The *measured* or *reco* distribution is a noisy distortion of the true one, measured with a detector with finite resolution. The idea to infer an unknown

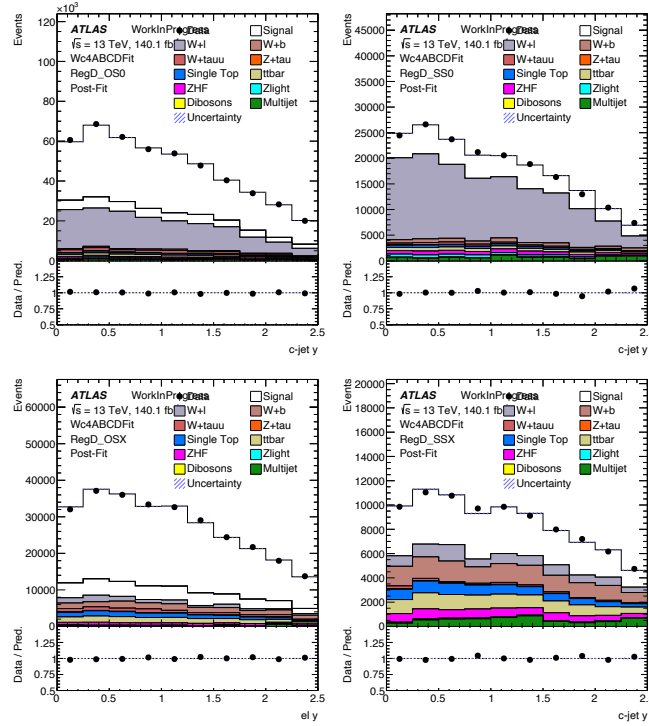


FIGURE 6.18: Region D for OS, SS bin 0 and bin X, electron channel, Run 2, c -jet rapidity variable.

distribution of a certain variable from the measured distribution by using knowledge and assumptions on the probability relation between the two is called *inference problem*. The unfolding technique is meant to solve this problem, factorizing out all the detector effects. The reason behind the unfolding is producing data that are independent from the detector, to make it possible to compare experimental data with theory predictions and combine results from different experiments.

6.5.1 The mathematical problem

If we consider a variable y , we wish to determine its probability density function $f(y)$. In an experiment, the measured values x of the variable are subject to random fluctuations because of measurement errors. The probability density functions of the true and measured value are related by a convolution relation:

$$f_{meas}(x) = \int R(x|y)f_{true}(y)dy \quad (6.6)$$

where $R(x|y)$ is the response function, supposed to be dependent only on the measuring apparatus. If the distributions of x and y are represented by histograms, the equation 6.6 becomes:

$$v_i = \sum_{j=1}^M R_{ij}\mu_j, \quad i = 1, \dots, N \quad (6.7)$$

where $\mu = (\mu_1, \dots, \mu_M)$ are the expectation values for the histogram of y distributed in M bins and $v = (v_1, \dots, v_N)$ are the expected number of events for the observed variable x . The *response or migration matrix* R is defined as the following conditional

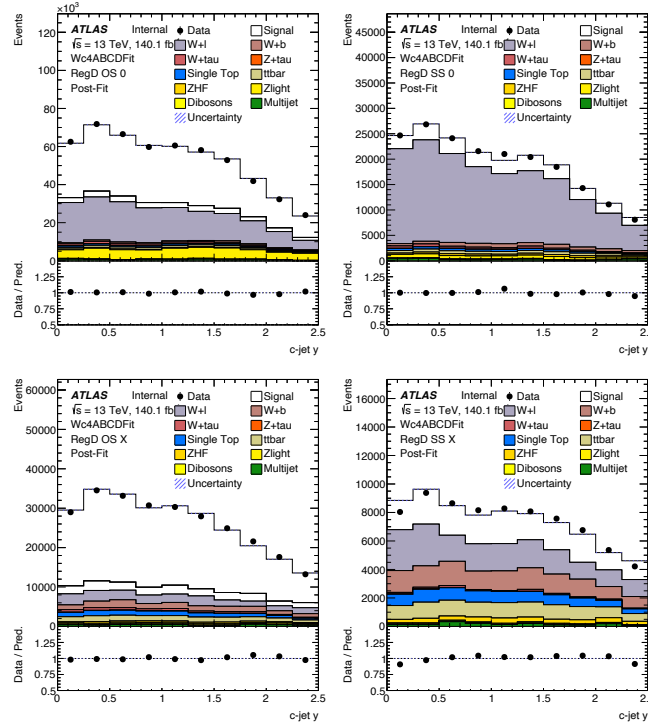


FIGURE 6.19: Region D for OS, SS bin 0 and bin X, muon channel, Run 2, c -jet rapidity variable.

probability:

$$R_{ij} = P(\text{observed in bin } i \mid \text{true value in bin } j). \quad (6.8)$$

The efficiency of the true bin j is defined by the R summing over all the possible bins of the observed value i :

$$\sum_{i=1}^N R_{ij} = P(\text{observed anywhere} \mid \text{true value in bin } j) = \epsilon_j. \quad (6.9)$$

Since the goal of the unfolding is to obtain the true distribution, it is necessary to invert the equation 6.7. There are several methods, the one investigated and used in this analysis is the Iterative Bayesian Unfolding, subject of the next section 6.5.2.

6.5.2 The Iterative Bayesian Unfolding

The *Iterative Bayesian Unfolding* is the standard algorithm for the unfolding in High Energy Physics, implemented by D'Agostini [181] and based on the Bayes theorem. Also in this case, the problem is considered discrete, so the distributions are binned. Each bin in the true distribution is considered as independent from the others. For these reasons, the algorithm can theoretically deal with any smearing and can be extended to multi-dimensional problems: in fact the core algorithm does not know the location of the bins in the configuration space of the problem, meaning that any multi-dimensional binned variable can be interpreted as an ensemble of independent bins.

Bayesian unfolding allows also for background estimation, in particular if the experimental data is affected by background events, they can be handled directly in the unfolding procedure: this is done by adding one cause bin for every source of

background. Inefficiencies, on the other hand, are taken into account during the unfolding by adding one extra effect bin .

In the probabilistic prospective, unfolding a particular measured spectrum means to estimate the conditional probability of having a true level spectrum $\bar{x}$, given the reco level one $\bar{y}$, $P(\bar{x}|\bar{y})$. Applying the Bayes' theorem directly to the bins, the conditional probabilities can be written as:

$$P(x_i | y_j, I) = \frac{P(y_j | x_i, I)P(x_i, I)}{\sum_i P(y_j | x_i, I)P(x_i, I)} \quad (6.10)$$

where I represents the prior information about the analysis and $P(x_i, I)$ the prior probability of finding the true value in the bin i . Using the definition of the response matrix 6.8, the equation 6.10 can be rewritten:

$$P(x_i | y_j, I) = \frac{R_{ji}P(x_i, I)}{\sum_i R_{ji}P(x_i, I)}. \quad (6.11)$$

In this way, the Bayes theorem allows to invert the probabilistic relation between true and reco and obtain the true probable distribution.

The evaluation of the matrix can be done through MC simulations, as well as the prior probabilities. The posterior probability results strongly affected by the choice of the prior. For this reason, on the *iterative procedure* is introduced, which consists of using, as prior probability, the posterior probability obtained in the previous iteration. The number of iterations has to be tuned, usually two or three steps are sufficient. Anyways a *regularization* procedure is necessary in order to choose the optimal number of iterations: too many iterations can compromise the result because the statistical fluctuations become dominant.

6.5.3 Algorithm validation and choice of iterations

In this analysis, for the unfolding problem, the Iterative Bayesian method is chosen as the standard one, as explained in the previous section 6.5.2. The algorithm is implemented using the library RooUnfold [183].

The unfolding is applied on the post-fit signal sample, while the response matrix is built from MC reco and truth samples. From the MC samples, it is also possible to infer the detection efficiencies for each bin, alongside the so called *fakes*. With the latter term we refer to events which have a reco level value but not the corresponding true level one.

Before applying the procedure to the data, it is necessary to test the correct implementation of the method. For this reason, the first step is to perform a *closure test*, which checks that the unfolding implementation is working properly, comparing the unfolded distribution with the truth level one. In this test, in fact, the same reco level distribution is used to construct the response matrix, making them completely correlated. If everything is working as expected, the algorithm should converge exactly at the true level distribution of the MC sample.

After the closure test mentioned above, it is possible to move on to the choice of hyper-parameters for the algorithm that, in case of the Iterative Bayesian Unfolding, is just the number of iterations.

As discussed in the previous section 6.5.2, iterations are introduced as a way to reduce the dependence of the unfolding on the prior distribution and Monte Carlo simulations. In order to find the correct number of iterations needed to have good convergence for the method, two quantities can be studied:

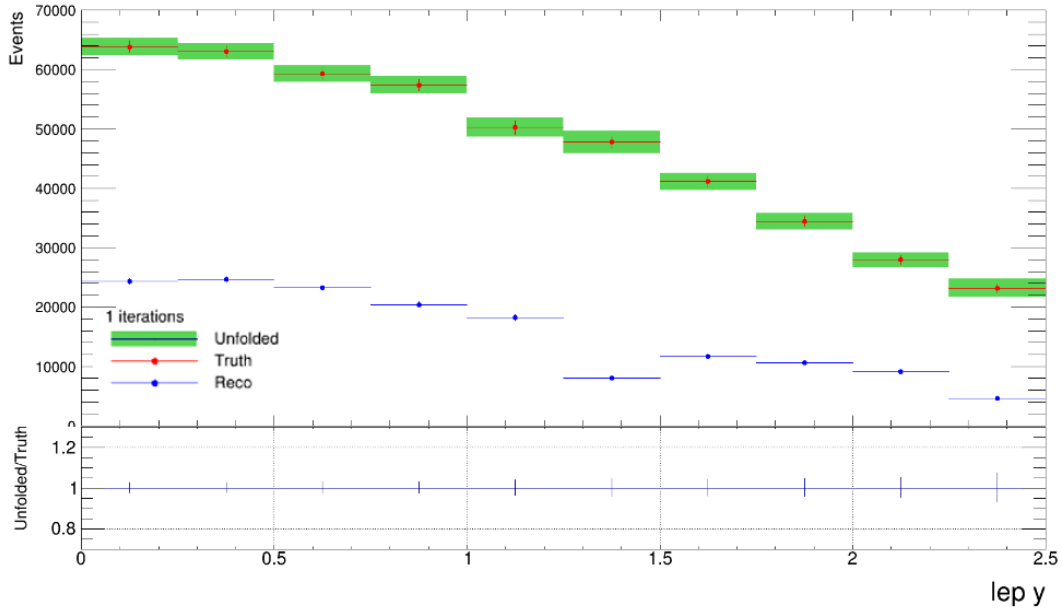


FIGURE 6.20: Closure test for the OS-SS lepton pseudorapidity distribution in the electron channel. In blue the reco level MC data, in red the true level MC distribution, in green the unfolded distribution [182].

- the statistical error propagated after each iteration. In particular it is interesting to look at the error generated from the finite size of the Monte Carlo sample that enters the unfolding through the response matrix. In fact, after each iteration, the response matrix is re-used to calculate the unfolded spectrum, meaning that its associated error gets propagated multiple times. Looking at this particular source of error can give some information about the correct number of iterations: convergence is found when the uncertainty remains stable with the respect to the number of iterations;
- a pseudo χ^2 quantity can be defined as:

$$f = \sum_{i=1}^{n_{bins}} \frac{(unf_i - truth_i)^2}{\sigma_i} \quad (6.12)$$

where unf_i and $truth_i$ represent the number of events in that particular bin i for the unfolded and Monte Carlo truth distribution respectively; while σ_i represents the statistical error in that bin. While this quantity is defined just like a regular χ^2 , it does not carry the same statistical meaning. In fact equation 6.12 just represents how the unfolded distribution approximates the true level one: an ideal unfolding would yield $f = 0$, in contrast to what is expected from a standard χ^2 test where the ideal result would be 1.

In order to have both quantities available, it is necessary to work with Monte Carlo samples because the true level distribution needs to be known. It is also interesting to note that, if a single Monte Carlo sample is used, the quantity f becomes trivial, since it will be equal to 0 for every iteration. It is then necessary to work with two different Monte Carlo samples or it is sufficient to divide a sample in two sub-datasets.

In Figure 6.21 the results for the lepton rapidity in the electron channel are shown.

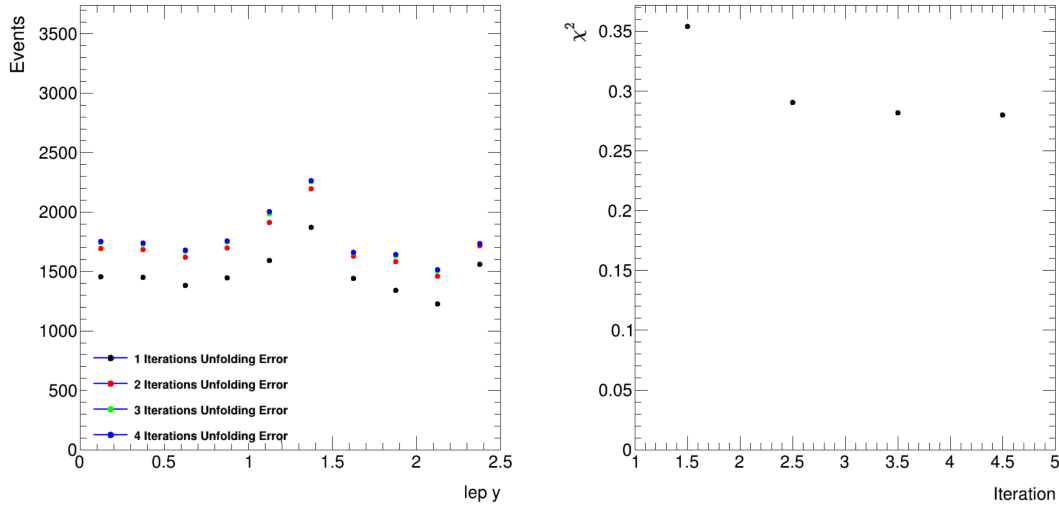


FIGURE 6.21: Distribution of statistical error in function of lepton rapidity, different colour represents different iteration (left). Pseudo χ^2 quantity plotted in function of the number of iterations (right) [182].

It is evident that the statistical error increases after each iteration but it starts to stabilize already after the second iteration. A similar observation can be made also for the pseudo χ^2 : it reaches a plateau almost instantly. From these results, one can extract the optimal number of iterations for rapidity, which here has been chosen to be 3. It is interesting to see that going beyond this optimal choice the unfolded distribution remains stable.

6.5.4 Post-Unfolding Results

After performing the closure test and choosing the number of iterations, we are now ready to implement the unfolding on the post-fit signal.

The Monte Carlo signal sample is used to construct the response matrix, fakes and inefficiency for the lepton rapidity variable. The unfolding is then performed via the Iterative Bayesian Unfolding where the prior distribution for the first iteration step corresponds to the Monte Carlo truth. The post-unfolding results for electron and muon channel separately are shown in Figure 6.22.

The unfolded distribution is shaped rather differently if compared to the Monte Carlo truth distribution but this difference is caused by the different reco level spectrum. It is also interesting to note that, in the electron channel, that the unfolded spectrum seems to show a 3-step distribution with some discontinuities, similarly to the reco level one. For this variable, in fact, the response matrix is almost completely diagonal: the unfolding in this case performs mostly an efficiency correction and makes little to no adjustments for the smearing.

For the muon channel, the unfolded distribution appears to be smoother if compared to the electron channel, but it still different when compared to the truth distribution.

6.6 Systematic Uncertainties

In the previous sections, only the statistical uncertainties are considered, however, also the systematic uncertainty needs to be addressed. The *systematic uncertainty* refers to all the sources of errors related to parameters that affect the measurements,

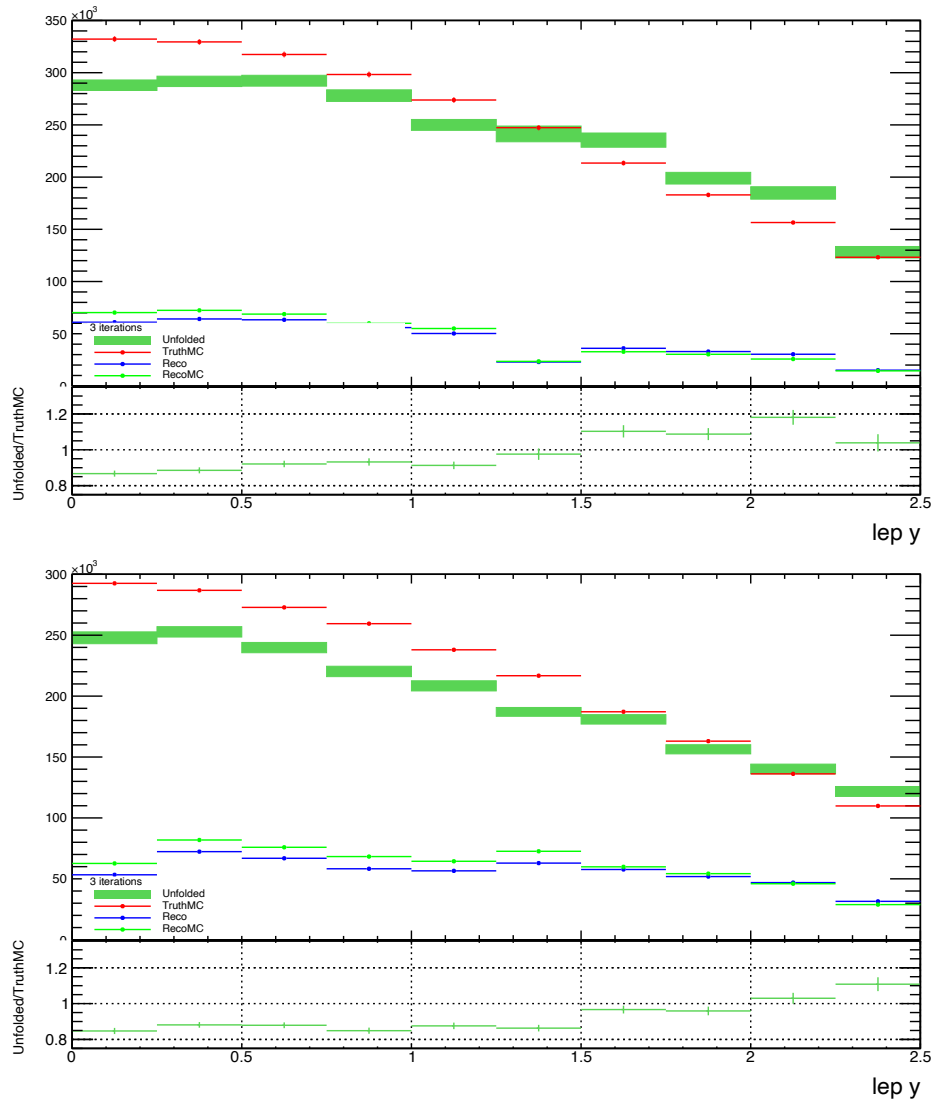


FIGURE 6.22: Unfolded distribution for electron (upper figure) and the muon (lower figure) channel Run 2 statistic (green rectangles), in red the MC truth, in blue the post-fit signal, in light green the MC reco. The uncertainties are statistic only [182].

as for example the error associated to the energy calibration or to the jet energy scale. The systematic uncertainties can also be associated to the model chosen for signal and background estimation.

In this analysis, for each uncertainty, background and signal MC samples fluctuate from the nominal by $\pm 1\sigma$. The fluctuations are kept totally correlated to the nominal, in order to eliminate every source of statistic uncertainties. Usually the process is done by adding weights to the associated events. In this analysis, the approach chosen is to consider each systematic as an independent nuisance parameter. The analysis is performed for each fluctuation, producing the post-unfolding result. The systematic error is calculated as the integral difference between the nominal and the fluctuation and then square-summed for the total systematic, as explained in the next section 6.6.1.

Systematic error sources considered in this thesis are the *detector-level uncertainty*, which are mainly related to the detector reconstruction of the events, and the *model uncertainty*. The latter is the error connected to the parametrization chosen in the physic model, as well as the background estimation procedure.

More details about the detector uncertainties are provided in Appendix A, while for the model ones in the following sections.

6.6.1 Systematic uncertainty evaluation

The fit explained in Section 6.1.1 is performed for each upward ($+1\sigma$) and downward (-1σ) variation from the nominal samples, obtaining the signal shape OS-SS in function of lepton rapidity. The signal distribution is then unfolded as explained in Section 6.5, independently for each systematic variation. In order to create the response matrices, both truth and reco MC samples are created for each systematic variation.

The systematic uncertainty on the signal yield for each systematic is obtained as following:

$$\begin{aligned}\sigma_{up,i} &= \mu_{up,i} - \mu_{nominal}, \\ \sigma_{down,i} &= \mu_{down,i} - \mu_{nominal}\end{aligned}\tag{6.13}$$

where μ is the integral of the OS-SS signal sample, obtained from the fit. The total systematic uncertainty is then obtained as squared sum:

$$\begin{aligned}\sigma_{up,tot} &= \sum_i (\sigma_{up,i})^2 \\ \sigma_{down,tot} &= \sum_i (\sigma_{down,i})^2.\end{aligned}\tag{6.14}$$

The signal systematic uncertainty is then estimated respect the lepton rapidity binned distribution, where the values are estimated in the same manner as in Equation 6.13 for each bin. The uncertainty is treated asymmetrically, means that the up and down variations are not combined in one single medium value. In this manner, it is possible to keep track of the impact and the sign of every systematic variation. The symmetric approach, instead, although it is more conservative, does not preserve the information about the relevance of the up and down variation of each systematic.

6.6.2 Detector-level systematic uncertainties

The most relevant systematic uncertainties for both electron and muon channel, for the full Run 2 dataset, are reported in Table 6.6 and 6.7 respectively. For these

evaluation, the samples have the $E_T^W > 20$ GeV cut.

TABLE 6.6: Fractional uncertainty in percentage for the most relevant systematics separated for up and down variation, for the electron channel full Run 2.

Systematic name	Uncertainty Up[%]	Uncertainty Down[%]
JET JER EffectiveNP	2.88	2.88
JET Flavor Composition	1.78	1.53
FT EFF Eigen C	1.47	1.40
JET BJETS Response	1.05	0.99
PRW DATASF	0.91	0.89
JET Flavor Response	0.83	0.66
Others	1.39	1.21
Total	4.11	3.91

TABLE 6.7: Fractional uncertainty in percentage for the most relevant systematics separated for up and down variation, for the muon channel full Run 2.

Systematic name	Uncertainty Up[%]	Uncertainty Down[%]
MUON EFF SYS LOWPT	3.14	2.95
JET JER EffectiveNP	1.84	0.31
FT EFF Eigen C	1.33	1.31
MUON SCALE	0.51	0.74
MUON EFF TrigSystUncertainty	0.67	0.67
Others	1.25	0.91
Total	4.15	3.51

From these tables, it is possible to observe the following items:

- the total uncertainty from the detector-level contribution are 4.11% and 4.15% for the up variation, and 3.91% and 3.51% for the down variation, for electron and muon channel respectively. They are very similar for the both channels, even if the systematic contribution is slightly different;
- for the electron channel the most relevant systematics are related to the Jet Energy Resolution, the Jet Flavor Composition and the efficiency of jet Flavor Tagging for the c -jets. For the muon channel, the Jet Energy Resolution and the efficiency of jet Flavor Tagging for the c -jets are also in the top 3, but in the first position there is the correction factors for low- p_T muons;
- the upwards and downwards variations are different for the majority of the systematics, reflecting in the total uncertainty. For this reason, it is important to keep the upwards and downwards variation separated, in order to appreciate the different contribution.

In Table 6.8 the systematic uncertainties for the electron channel with the default selection $E_T^W > 30$ GeV are reported. It is clear that the uncertainty has decreased by $\sim 35\%$ as a result of an increase in signal purity and a decrease in background. In particular the systematics most reduced are those related with the energy resolution.

TABLE 6.8: Fractional uncertainty for the most relevant systematics, for electron channel Run 2, the default selection $E_T^W > 30$ GeV is applied.

Systematic name	Uncertainty Up[%]	Uncertainty Down[%]
FT EFF Eigen C	1.43	1.37
JET Flavor Composition	1.24	1.38
JET BJETS Response	0.93	0.98
SysPRW DATASF down	0.85	0.90
JET JER EffectiveNP	0.84	0.84
JET Flavor Response	0.099	0.74
EG SCALE ALL down	0.50	0.53
Total	2.64	2.81

TABLE 6.9: Fractional uncertainty in percentage for the multijet and $W+l$ asymmetry estimation for the muon and electron channel full Run 2 post-Fit.

Systematic name	Electron [%]	Muon[%]
Multijet	2.18	1.38
$W+l$ asym	3.67	2.90

6.6.3 Signal and background systematic uncertainties

Other important systematic uncertainties are related to the signal and background modeling. For the signal modeling systematic evaluation, an alternative MC event generator for the $W+jets$ sample, has to be considered. In our case, as described in Section 5.3.1, the alternative sample is SHERPA 2.2.11, in order to estimate the error of choosing the Madgraph MG5_aMCNLO as the baseline one.

Another source of systematic uncertainty is the $W+light$ asymmetry estimation, which has been done in the validation region of $20 < E_T^{miss} < 30$, as explained in Section 6.3. The systematic evaluation is done, performing another time the fit fixing the $W+light$ value at that one obtained in the validation region. The difference between the signal OS-SS from the nominal one is considered as a systematic uncertainty.

The multijet background is estimated directly in the likelihood fit, for this reason the uncertainty from its evaluation on the signal is already considered. Moreover, it is necessary to consider the uncertainty in the ABCD method itself and the choice of the variables. For this job, we explored alternative variables: instead of the transverse mass of the W boson vs lepton isolation, we considered the transverse energy missing vs the lepton isolation. The difference of the signal distribution is considered as systematic.

Other systematics that need to be considered are those related to the charm hadronization, in particular the uncertainty on the c -quark fragmentation and the c -hadron decay models, which affect the soft- μ p_T spectrum and the number of hadrons decaying into muons. Moreover the branching ratios of c -hadron semileptonic decays to muons is affected by uncertainty in the event generator and has to be taken into account.

The contribution of multijet and $W+l$ asymmetry estimation for electron and muon channels are reported in Table 6.10. The different shape for the nominal and systematics are also shown in Figure 6.23 and 6.24. The results are shown in the post-fit version, regarding the post-unfolding, for the electron channel, the values are compatible with

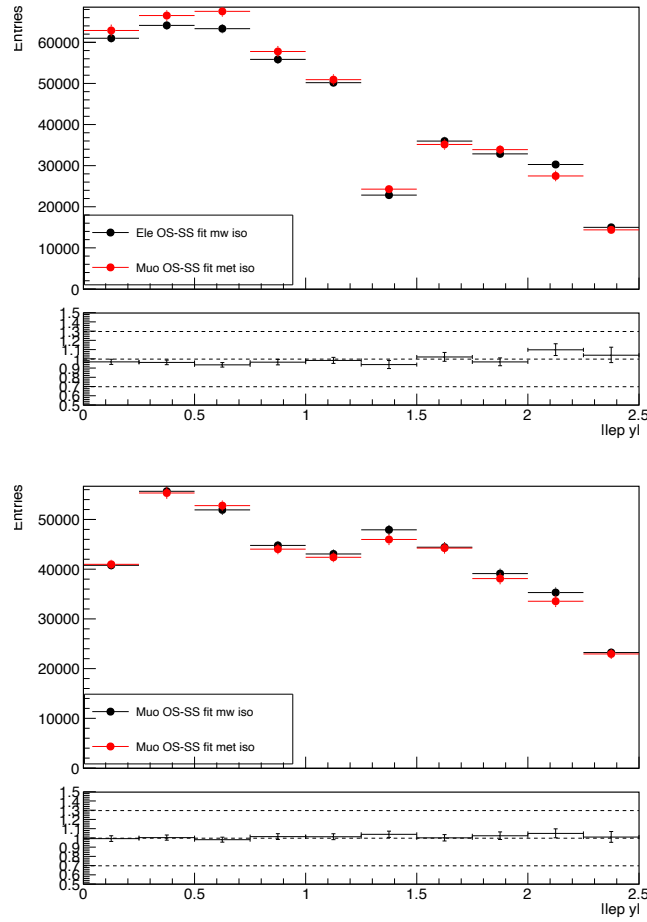


FIGURE 6.23: Comparison of the signal OS-SS post fit for the m_W^T and iso vs MET and iso for the ABCD method, for the electron (up) and muon (down) channels, full Run 2.

the post-fit those.

It is evident that these two systematics are dominant respect the others, in particular the one related with the $W+l$ asymmetry estimation. This last will be investigated in the near future.

6.7 Results

In this section, I present the latest results post-unfolding and discuss the next steps. In the previous Section 6.6 the systematic estimation is explained, giving an idea on which contribution is more relevant. The lepton rapidity post-unfolding distributions for both channels are shown in Figure 6.25, reporting both statistic and systematic uncertainties bin by bin. The statistic uncertainty is figured in red, the systematic one in black tilted lines, while the green box is the sum squared of the two. It is evident that the systematic uncertainty is dominant respect the statistic one. In the figures, for the electron channel, the default $E_T^{miss} > 30$ GeV cut is applied and the present systematics are shown, for the muon channel, instead, the 20 GeV cut is applied and the systematics related to the ABCD method and the $W+l$ asymmetry estimation are included at post-fit level.

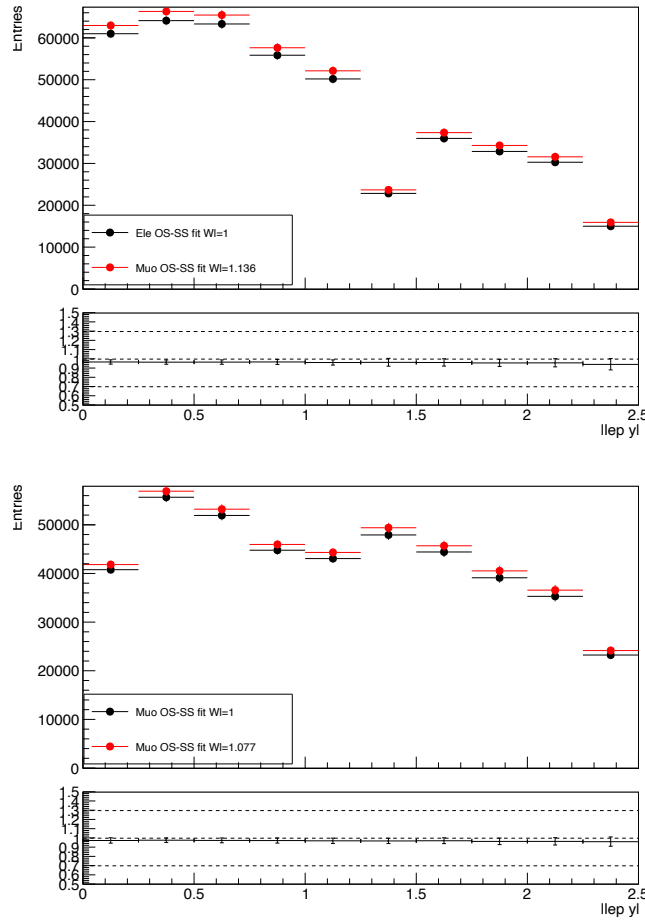


FIGURE 6.24: Comparison of the signal OS-SS post fit for the nominal and for the $W+l$ asymmetry fixed at the values obtained in the validation region, for the electron (up) and muon (down) channels, full Run 2.

For the sake of completeness, the cross section σ for the electron and muon channels are reported, evaluated in the fiducial phase space, defined in Section 5.4.13. The cross section is calculated as the total integral of the lepton rapidity distribution divided for the integrated luminosity, following the equation 2.3. The total uncertainty is around the 6% for the electron channel and 5-6% for the muon channel.

TABLE 6.10: Total cross section for the muon and electron channel full Run 2. The two channels have different truth phase space, the electron has the requirement of 1 or 2 jets, while the muon channel has an inclusive selection.

Channel	$\sigma_{W+c}^{OS-SS,fid}$ (pb)
Electron	$17.05^{+0.99}_{-1.02}$
Muon	$13.96^{+0.74}_{-0.83}$

The next steps are to include the remaining systematics uncertainties, e.g., the charm fragmentation, the signal modelling and the PDF ones. In the next future the combination between the two channels will be investigated in order to provide a complete

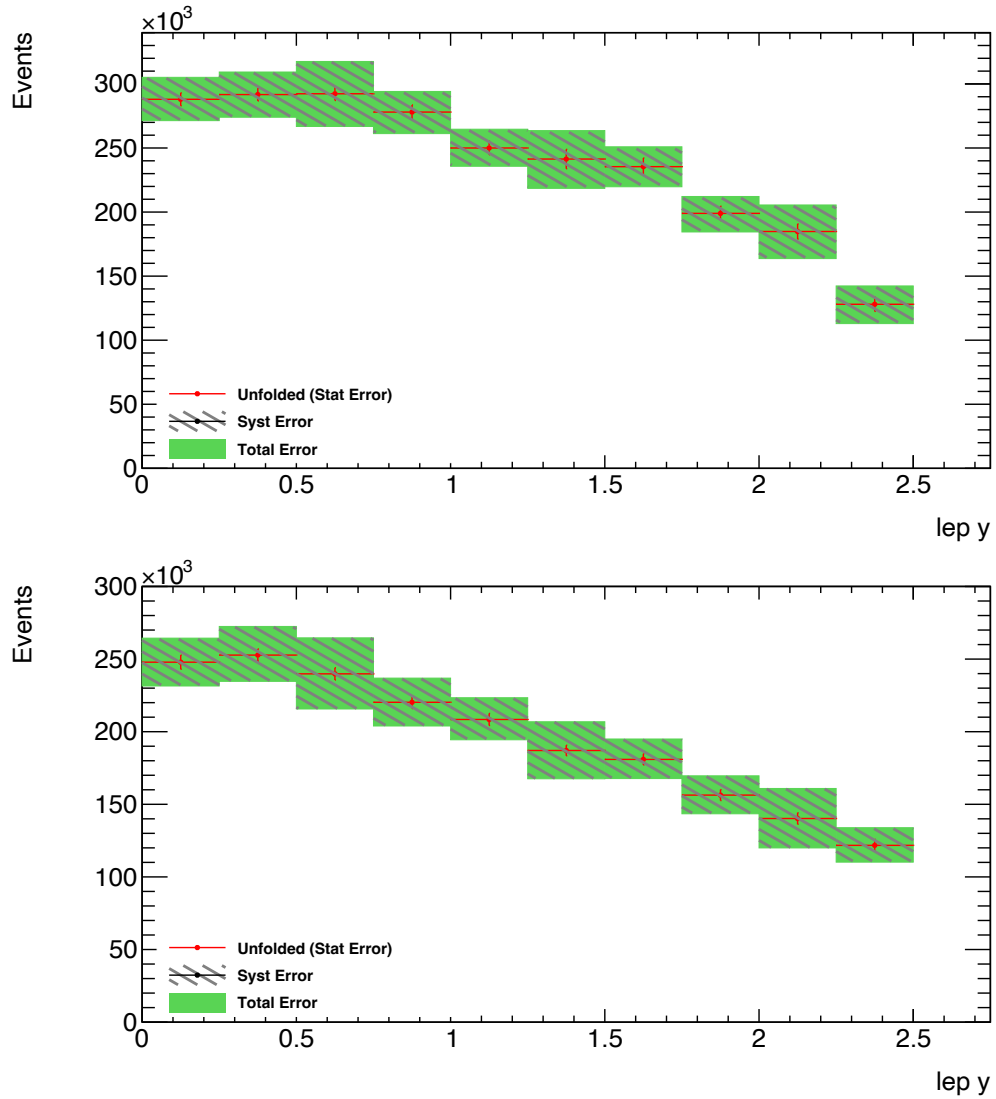


FIGURE 6.25: Post unfolding distributions for lepton rapidity for electron (above) and muon (below) channels, using full Run 2 statistic. In red the unfolded distribution and the statistic error associated. The black diagonal lines are the systematic error and in green the sum of the systematic and statistic error.

measurement of the $W+c$ process.

Other variables will be considered, as the c -jet rapidity. Also the sign asymmetry will be investigated: the procedure will be done for W^- and W^+ separately, in order to test any possible asymmetries between s and $\bar{s}$ respectively.

In the end, the result will be compared with different strange PDF predictions, in order to obtain information regarding the suppression of the strange quark respect the other sea quarks in the proton. The result will be also include in future global PDF fits.

Conclusions

In this thesis, several experimental aspects related to the quark-flavour identification have been investigated.

In particular, I have contributed to the implementation of a new $Hb\bar{b}$ tagger, which makes use of colour information of two b -flavour jets, in order to distinguish if they originate from a singlet of colour or from pure QCD background. The study is focused on the $pp \rightarrow H(b\bar{b})Z(\nu_l\bar{\nu}_l)$ process and introduces eight high-level colour sensitive variables as training for a BDT algorithm. One of these variables is the output of a CNN, trained with the Lund jet plane images. We included in the study the detector effect using a fast simulation, comparing the results in the truth and reconstructed cases. The performances are promising, showing a good discrimination power of the combination of the colour sensitive observables (AUC = 0.893), slightly worsening with the inclusion of the detector effects (AUC = 0.846). The Lund jet plane was already proven to be powerful alone, even in presence of detector effects and, when combined with phenomenological observables (such as D_2 and colour ring), the performances improve. We concluded, therefore, that the colour sensitive observables are not only effective in particle theory, but they also seem to be promising in experimental physics.

In order to use the $Hb\bar{b}$ tagger in a more general contest, the decaying b -quarks mass correlation has been investigated. As known in literature, D_2 shows a mass bias. We discover that also the Lund plane CNN presents a large mass correlation. Since these two observables are fundamental for the tagger performances, we cannot exclude them. We should think about some decorrelation techniques, as those presented here [84, 85], based on debiasing a posteriori.

Although we included the detector effects by means of a fast simulation, the algorithm performances should be tested in a more realistic scenario. For example, for this tagger we are considering a perfect b -identification, while this is not the case in ATLAS. Further, the ML architecture is trained with the high level observables from the primary Lund plane, while we are not considering the secondary splittings. In order to use the tagger in the ATLAS analyses, we should evaluate its effectiveness in a real scenario.

The charm- and strange-flavour quarks are investigated in the measurement of W boson in association with charm quark production, using data collected during Run 2 at energy of center of mass $\sqrt{s}=13$ TeV with the ATLAS detector. The analysis strategy consists in signal and background estimation using an innovative simultaneous 2D fit, which is able to estimate the multijet background with the ABCD method implementation, to extrapolate a W +light correction factor and to obtain the signal OS-SS distribution. The default variable is the lepton rapidity, but also the soft-muon-jet rapidity has been taken into account. The latter variable is also used as proof of the fit stability, since the results are compatible with the lepton rapidity ones. For the ABCD method the default variables are the transverse W mass, m_W^T , and the lepton isolation, which are shown to be almost independent.

The post-fit OS-SS signal is then unfolded, in order to remove the detector effects.

The iterative Bayesian unfolding has been chosen as method and provide good performances after two iterations.

In order to estimate the uncertainties on the post-fit signal sample, the systematic uncertainties are evaluated, along with the statistic ones. For each systematic variation, the fit is performed and consequently the post-fit signal obtained is unfolded. The uncertainty is then considered as difference from the nominal and square-summed integral. The total uncertainty on the integral is about 10% for both electron and muon channel.

The last step of the analysis, which will be investigated in the next months, is to obtain the final information about the strange PDF. The unfolding provides the cross-section differential in lepton rapidity, which can be used to derive the strange PDF. Via the DGLAP equations and the factorisation theorem, it is possible to directly connect the parton distribution functions with the final-state cross sections.

Finally, in order to have good data for the Good Run Lists used in the analyses, a monitoring of the quality of the data during the data-taking is mandatory. The task is performed by the Data Quality Online Monitoring, via several tools available in the ATLAS Control Room. My contribution to the Data Quality during the last three years has been the developing and maintaining of algorithms used in the tools for the Muon Spectrometer, for the MDT and NSW sub-detectors. In the future, I plan to carry on with this activity, improving the existing algorithms and creating new even more performative ones.

Appendix A

Detector-level systematic uncertainties

An overview of the detector-level uncertainties is given in this section.

ATLAS combined performance groups produced uncertainty variations and they are applied to all the MC signal and background samples. They are then propagated through the reco-level plots and are summed in quadrature to obtain the total uncertainty. Each uncertainty, unless otherwise specified, is defined using $\pm 1\sigma$ “up/down” variations. An overview of their purpose is depicted in the following:

- **MUON:** This includes uncertainties in the muon measurements, including the smearing of tracks for “combined” muons (MUON_CB) and mismodelling of the momentum scale (MUON_SCALE). A charge-dependent scale correction (MUON_SAGITTA_RHO) is included, along with the residual charge dependency after the correction (MUON_SAGITTA_RESBIAS) and its statistical error (MUON_SAGITTA_DATASTAT). Finally, both the statistical and systematic uncertainties in the isolation measurement (MUON_ISO_STAT/MUON_ISO_SYST) and the track-to-vertex association (MUON_TTVA_STAT/MUON_TTVA_SYST) are included.
- **MUON_EFF:** The uncertainties on muon efficiency measurements include per-object correction factors (MUON_EFF_STAT/ MUON_EFF_SYS) and uncertainties in the trigger efficiencies (MUON_EFF_TrigStatUncertainty/ MUON_EFF_TrigSystUncertainty). Correction factors are also included specifically for low- p_T muons (MUON_EFF_STAT_LOWPT/ MUON_EFF_SYS_LOWPT).
- **EG:** This group encompasses uncertainties due to imperfect electron calibration. This includes mis-modelling of the electron energy (EG) scale (EG_SCALE) and resolution (EG_RESOLUTION_ALL). Finally, both the statistical and systematic uncertainties due to electron charge identification (EL_CHARGEID_STAT/ EL_CHARGEID_SYST) are included.
- **EL_EFF:** These uncertainties address the effects of imprecise electron efficiency measurements. This includes measurements of the reconstruction (EL_EFF_Reco_TOTAL_1NPCOR_PLUS_UNCOR), identification (EL_EFF_ID_TOTAL_1NPCOR_PLUS_UNCOR), and isolation (EL_EFF_Iso_TOTAL_1NPCOR_PLUS_UNCOR) efficiencies, as well as the determination of the trigger scale factor (EL_EFF_Trigger_TOTAL_1NPCOR_PLUS_UNCOR).
- **MET:** The missing transverse energy uncertainties are on the resolution (MET_SoftTrk_ResoPerp) and scale (MET_SoftTrk_Scale) of the track-based soft-term contributions. Hard-term uncertainties are covered by the electron, muon, and jet uncertainties.

- **JES:** Uncertainties in the Jet Energy Scale (JES) are handled with correction factors measured in-situ with respect to well-understood reference objects; these include Zs (relevant for low- p_T (< 100 GeV) jets), photons ($100 \text{ GeV} < p_T < 1 \text{ TeV}$), and systems of previously-calibrated low- p_T jets ($1 < p_T < 2 \text{ TeV}$). A correction factor is then the ratio of the jet p_T to the p_T of the object the jet is being balanced against. All of these nuisance parameters (NPs) are combined into a smaller set of statistical (JET_EffectiveNP_Statistical), modelling (JET_EffectiveNP_Modelling), detector (JET_EffectiveNP_Detector), and mixed (JET_EffectiveNP_Mixed) variations. Additional in-situ calibration are done to balance forward jets against central jets (JET_EtaIntercalibration) and very high (1-2 TeV) p_T jets against single particle propagation (JET_SingleParticle_HighPt). Other uncertainties accounted for include pileup (JET_Pileup), which is relevant for very low- p_T jets; punch-through jets which escape the calorimeters (JET_PunchThrough); the mismodelling of the jet flavor response (JET_Flavor_Response) and composition (JET_Flavor_Composition), with b -jets handled individually (JET_BJETS_Response); and the mismodelling of the jet vertex tagging (JVT) requirement efficiency (JET_JvtEfficiency).
- **JET:** Uncertainties in the Jet Energy Resolution (JER) are handled with a set of eight NPs (JET_JER_DataVsMC_MC16 and JET_JER_EffectiveNP)
- **FT_EFF:** The uncertainty in the jet Flavor Tagging (FT) efficiency is handled with a set of nuisance parameters determined using an eigenvector variation method. 44 variations are used for b -jets (FT_EFF_Eigen_B) and 19 are used each for c -jets (FT_EFF_Eigen_C) and light jets (FT_EFF_Eigen_Light).

The uncertainty in the combined 2015-2018 integrated luminosity is 0.83% [184], which was obtained using the LUCID-2 detector [9] for the primary luminosity measurements.

Appendix B

MDT Algorithms

The **ADC distribution** is the total charge generated by the particle as it passed through an MDT tube. The histogram is provided per MDT chamber multi-layer by the Calibration Stream. The charge distribution is shown in ADC counts with a cut on 50 where is usually located a noise spike. The algorithm used for testing the histogram is the *ADCAlgorithm* which includes three parameters and two thresholds:

- **Parameters:**
 - **StatisticsThreshold:** minimum number of entries to actually check the other parameters. If this threshold is not met, the histogram is simply marked grey. Currently set to 500.
 - **ADCSignalRegionBegin:** the ADC value for which any hits below are considered noise, and any hits above are considered signal. Currently set to 80.
 - **ADCTargetAverage:** expected value for the average signal ADC. Currently set to 155.
- **Thresholds:**
 - **ADCAvgSigTolerance:** the histogram is marked green/red if its average value is different from the expected value by more than this threshold. Currently set to 35/40.5 respectively.
 - **ADCNoiseToSigTolerance:** the histogram is marked green/red if its noise to signal ratio is above this threshold. Currently set to 0.15/0.3.

The **TDC distribution** is the travel time of secondary electrons in the tube after a particle passes through it. The width of the spectrum therefore illustrates the range of travel times of these freed electrons. As for the ADC, the histogram is provided per MDT chamber multi-layer by the Calibration Stream and only actual tracks are plotted, with a cut on 50 ADC. The two main features of this histograms are the location of the rising edge t_0 and the position where the spectrum drops off t_{max} . The x axis in TDC counts, which are 0.7 ns wide. In few cases, the entire spectrum is shifted up and the "baseline" from which the main distribution rises and to which it falls again is actually significantly nonzero. However, it is important that the shape of the distribution below t_0 and above t_{max} is flat, and roughly the same value on either side. The algorithm, *TDCAAlgorithm*, finds t_0 and t_{max} , computes the difference, and compares this delta t to some expected value. For the sMDT (BIX/Y7) the distribution is narrower, since the time drift is shorter: 240 counts instead of 940 counts. The TDC algorithm takes four parameters and four thresholds:

- **Parameters:**

- **StatisticsThreshold**: if a histogram has less than this number of events, we can't meaningfully analyze it, so it is simply marked in grey. Currently set to 3000 by default.
- **T0EdgeThreshold**: currently set to 0.4.
- **TdriftTarget**: currently set to 940.
- **sMDT_TdriftTarget**: currently set to 260.
- **Thresholds**:
 - **BaselineThreshold**: currently set to 0.9.
 - **TpeakThreshold**: currently set to 300.
 - **T0Threshold**: currently set to 0.
 - **BumpThreshold**: currently set to 0.25.

In the Occupancy section, the **hits per event vs LB** shows hits per each multi-layer over the course of LBs. The histogram comes directly from the L1 trigger from the GNAM server, showing the hits not cleaned from background tracks. The *OccupancyPerLBAlgo1D* provides three parameters and one threshold:

- **Parameters**:
 - **nLBChecked**: number of consecutive LBs checked. Currently set at 20.
 - **SigmasAboveMedianIsNoise**: number of sigma above median which hits are considered noise. Currently set at 5.
 - **StartByLB**: minimum bin number at which to expect all chambers should be sending data and the algorithm start to test. Currently set at 20.
- **Thresholds**:
 - **nLBEmptyThreshold**: number of LBs with no hits. Green value for less than 5, yellow between 5 and 10, red for values greater than 10.

In the Layer section, the multi-layers and single layers hits are checked. The **hits per tube and per layer** is a 2D histogram, showing the layers occupancy vs tubes per layer and taken from the Calibration Stream. The *LayerAlgorithm* requires one parameter and two thresholds:

- **Parameters**:
 - **StatisticsThreshold**: currently set at 1000.
- **Thresholds**:
 - **MLThreshold**: for every multi-layer checks if the number of hits per tube are different of a certain value. Green threshold for values greater than 0.9, yellow between 0.9 and 0.7, red less than 0.7.
 - **LayerThreshold**: same check as ML but single layer. Green value for greater than 0.9, yellow between 0.9 and 0.5, red less than 0.5.

The DAQ system for each chamber can send a variety of error signals. The **Error Rate** histogram collects the rate (that is, the number of times the error was sent divided by the total number of hits) at which these errors stand for lumiblocks (LBs). Each error has its own tolerable ranges, set individually in the configuration file.

The *ChamberErrorsAgo2D* compares each kind of error rate with the correspondent threshold. They are reported in the Table B.1.

TABLE B.1: Error description and threshold values

Errors	Description	Green	Red
BOT	TDC start word BOT missing	0.001	0.01
BOTL1ID	TDC start word has wrong L1ID	0.001	0.01
BOTBCID	TDC start word has wrong BCID		
EOTL1ID	TDC end word has wrong L1ID	0.001	0.01
TDCSIZE	TDC block size out of range	0	0.01
CSMMROD	Transmission error from CSM to MROD	0.001	0.01
TDCCSM	Transmission error from TDC to CSM	0.001	0.01
COA	Parity coarse	0.005	0.01
CH	Channel selection error	0.0001	0.01
L1B	L1 buffer	0.0001	0.01
TRI_FIFO	Trigger FIFO error	0.0001	0.01
MATCH	Matching state error	0.0001	0.01
RO_FIFO	Read-out FIFO error	0.00005	0.02
RO_STA	Read-out state error	0.0001	0.01
CONTROL	Control parity error	0.0001	0.01
JTAG	JTAG instruction error	0.0001	0.01
RO_FIFO_OF	Readout FIFO overflow		
TSM	Wrong channel read in the TSM word	0.01	0.05
TSMERR	Error flag in the TSM word	0.9	1.5
TCM	Wrong channel read in the TCM word	0	0.01
TDCWORD	Wrong TDC word	0	0.01
BOTTLP	Error in TLP word	0.05	0.01
TLP	Missing or bad TLP	0	0.01
TLPLINK	Bad TDC link in TLP word	0	0.01
TWC	Missing TWC	0	0.01
TWCCNT	Wrong TWC counter	0	0.0001
TWCL1ID	TWC counter does not match L1ID	0	0.001
EOT	Missing EOT	0.01	0.01
EOTWCNT	TDC size not equal to EOT_WCNT	0.01	0.01
LWCL1ID	Wrong LWC event id (one nibble)		
EOTBCID		0.001	0.01

Bibliography

- [1] CERN. *Standard Model*. URL: <https://home.cern/science/physics/standard-model>.
- [2] “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Phys. Lett. B* 716 (2012). Comments: 24 pages plus author list (38 pages total), 12 figures, 7 tables, revised author list, pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214. URL: <https://cds.cern.ch/record/1471031>.
- [3] “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235. URL: <https://cds.cern.ch/record/1471016>.
- [4] R. Keith Ellis, W. James Stirling, and B. R. Webber. *QCD and collider physics*. Vol. 8. Cambridge University Press, Feb. 2011. ISBN: 978-0-511-82328-2, 978-0-521-54589-1. DOI: 10.1017/CB09780511628788.
- [5] R. L. Workman and Others. “Review of Particle Physics”. In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.
- [6] Mark Thomson. *Modern particle physics*. New York: Cambridge University Press, 2013. ISBN: 978-1-107-03426-6. DOI: 10.1017/CB09781139525367.
- [7] S. Bailey et al. “Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs”. In: *The European Physical Journal C* 81.4 (Apr. 2021). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-021-09057-0. URL: <http://dx.doi.org/10.1140/epjc/s10052-021-09057-0>.
- [8] Vladislav Balagura. “Van der Meer scan luminosity measurement and beam-beam correction”. In: *The European Physical Journal C* 81.1 (2021). DOI: 10.1140/epjc/s10052-021-08837-y. URL: <https://doi.org/10.1140/epjc/s10052-021-08837-y>.
- [9] G Avoni et al. “The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS”. In: *JINST* 13.07 (2018), P07017. DOI: 10.1088/1748-0221/13/07/P07017. URL: <https://cds.cern.ch/record/2633501>.
- [10] S. Catani et al. “New clustering algorithm for multi-jet cross-sections in e^+e^- annihilation”. In: *Phys. Lett. B* 269 (1991), pp. 432–438. DOI: 10.1016/0370-2693(91)90196-W.
- [11] S. Catani et al. “Longitudinally invariant K_t clustering algorithms for hadron hadron collisions”. In: *Nucl. Phys. B* 406 (1993), pp. 187–224. DOI: 10.1016/0550-3213(93)90166-M.

- [12] Stephen D. Ellis and Davison E. Soper. “Successive combination jet algorithm for hadron collisions”. In: *Physical Review D* 48.7 (1993), pp. 3160–3166. DOI: 10.1103/physrevd.48.3160. URL: <https://doi.org/10.1103/physrevd.48.3160>.
- [13] Yu.L Dokshitzer et al. “Better jet clustering algorithms”. In: *Journal of High Energy Physics* 1997.08 (1997), pp. 001–001. DOI: 10.1088/1126-6708/1997/08/001. URL: <https://doi.org/10.1088/1126-6708/1997/08/001>.
- [14] M. Wobisch and T. Wengler. *Hadronization Corrections to Jet Cross Sections in Deep-Inelastic Scattering*. 1999. arXiv: hep-ph/9907280 [hep-ph].
- [15] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. “The anti- k_t jet clustering algorithm”. In: *Journal of High Energy Physics* 2008.04 (2008), pp. 063–063. DOI: 10.1088/1126-6708/2008/04/063. URL: <https://doi.org/10.1088/1126-6708/2008/04/063>.
- [16] The ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *Journal of Instrumentation* 3.08 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003. URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08003>.
- [17] <https://tikz.net/>. Accessed: 2023-11-23.
- [18] CERN. *Grey Book*. URL: <https://greybook.cern.ch/experiment/detail?id=ATLAS>.
- [19] A. Airapetian and Grabsky et al. *ATLAS detector and physics performance: Technical Design Report, 1*. Technical design report. ATLAS. Geneva: CERN, 1999. URL: <http://cds.cern.ch/record/391176>.
- [20] *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*. Tech. rep. Geneva: CERN, 2017. DOI: 10.17181/CERN.F0ZZ.ZP3Q. URL: <https://cds.cern.ch/record/2285585>.
- [21] *Technical Design Report for the ATLAS Inner Tracker Strip Detector*. Tech. rep. Geneva: CERN, 2017. URL: <https://cds.cern.ch/record/2257755>.
- [22] A Romaniouk. *TRT straw counting rate from photon background at LHC*. Tech. rep. Geneva: CERN, 1996. URL: <http://cds.cern.ch/record/686000>.
- [23] *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical design report. ATLAS. Geneva: CERN, 1996. DOI: 10.17181/CERN.FWRW.F00Q. URL: <https://cds.cern.ch/record/331061>.
- [24] *ATLAS tile calorimeter: Technical Design Report*. Technical design report. ATLAS. Geneva: CERN, 1996. DOI: 10.17181/CERN.JRBJ.7028. URL: <https://cds.cern.ch/record/331062>.
- [25] *ATLAS muon spectrometer: Technical Design Report*. Technical design report. ATLAS. Geneva: CERN, 1997. URL: <https://cds.cern.ch/record/331068>.
- [26] *Technical Design Report for the Phase-II Upgrade of the ATLAS Muon Spectrometer*. Tech. rep. Geneva: CERN, 2017. URL: <https://cds.cern.ch/record/2285580>.
- [27] T Kawamoto et al. *New Small Wheel Technical Design Report*. Tech. rep. ATLAS New Small Wheel Technical Design Report. 2013. URL: <https://cds.cern.ch/record/1552862>.
- [28] <https://twiki.cern.ch/twiki/pub/AtlasPublic/ApprovedPlotsDAQRun3/tdaq-run3-schematic-dec22.pdf>. Accessed: 2023-11-23.

- [29] Peter Jenni et al. *ATLAS high-level trigger, data-acquisition and controls: Technical Design Report*. Technical design report. ATLAS. Geneva: CERN, 2003. URL: <https://cds.cern.ch/record/616089>.
- [30] William Vazquez. "FELIX: the new detector interface for ATLAS". In: *EPJ Web of Conferences* 245 (Jan. 2020), p. 01037. DOI: 10.1051/epjconf/202024501037.
- [31] A V Vaniachine. *ATLAS Detector Data Processing on the Grid*. Tech. rep. Geneva: CERN, 2011. URL: <https://cds.cern.ch/record/1403169>.
- [32] Markus Elsing et al. "The ATLAS Tier-0: Overview and operational experience". In: *Journal of Physics: Conference Series* 219.7 (2010), p. 072011. DOI: 10.1088/1742-6596/219/7/072011. URL: <https://dx.doi.org/10.1088/1742-6596/219/7/072011>.
- [33] "Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J., C* 81 (2021). 64 pages in total, author list starting page 42, auxiliary material starting at page 59, 34 figures, 3 tables. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/MUON-2018-03/>, p. 578. DOI: 10.1140/epjc/s10052-021-09233-2. arXiv: 2012.00578. URL: <http://cds.cern.ch/record/2746302>.
- [34] ATLAS experiment. "Electron and photon performance measurements with the ATLAS detector using the 2015-2017 LHC proton-proton collision data". In: *Journal of Instrumentation* 14.12 (2019), P12006–P12006. DOI: 10.1088/1748-0221/14/12/p12006. URL: <https://doi.org/10.1088/1748-0221/14/12/p12006>.
- [35] ATLAS Collaboration. "Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV". In: *The European Physical Journal C* 79.8 (2019). DOI: 10.1140/epjc/s10052-019-7140-6. URL: <https://doi.org/10.1140/epjc/s10052-019-7140-6>.
- [36] <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>. Accessed: 2024-03-10.
- [37] "Jet reconstruction and performance using particle flow with the ATLAS Detector. Jet reconstruction and performance using particle flow with the ATLAS Detector". In: *Eur. Phys. J. C* 77.7 (2017). 67 pages in total, author list starting page 51, 37 figures, 1 table, final version published in *Eur. Phys. J C*, all figures including auxiliary figures are available at <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/PERF-2015-09/>, p. 466. DOI: 10.1140/epjc/s10052-017-5031-2. arXiv: 1703.10485. URL: <https://cds.cern.ch/record/2257597>.
- [38] *Jet global sequential corrections with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2015-002>. Geneva: CERN, 2015. URL: <https://cds.cern.ch/record/2001682>.
- [39] *Data-driven determination of the energy scale and resolution of jets reconstructed in the ATLAS calorimeters using dijet and multijet events at $\sqrt{s} = 8$ TeV*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2015-017>. Geneva: CERN, 2015. URL: <https://cds.cern.ch/record/2008678>.

- [40] *Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2015-037>. Geneva: CERN, 2015. URL: <https://cds.cern.ch/record/2044941>.
- [41] *Determination of the jet energy scale and resolution at ATLAS using Z/γ -jet events in data at $\sqrt{s} = 8$ TeV*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2015-057>. Geneva: CERN, 2015. URL: <https://cds.cern.ch/record/2059846>.
- [42] Rafal Bielski. “ATLAS Control Room celebrates first Run 3 beam splashes”. General Photo. 2022. URL: <https://cds.cern.ch/record/2809871>.
- [43] P. Calafiura et al. “The Athena control framework in production, new developments and lessons learned”. In: *14th International Conference on Computing in High-Energy and Nuclear Physics*. 2005, pp. 456–458.
- [44] CERN. *Countdown to physics: Beams splash in the ATLAS experiment*. URL: <https://atlas.cern/Updates/News/Run3-beams-splash>.
- [45] P. Adragna et al. “GNAM and OHP: Monitoring Tools for ATLAS experiment at LHC”. In: *2006 IEEE Nuclear Science Symposium Conference Record*. Vol. 1. 2006, pp. 114–118. DOI: 10.1109/NSSMIC.2006.356119.
- [46] A Corso-Radu et al. “Data Quality Monitoring Framework for the ATLAS experiment: Performance achieved with colliding beams at the LHC”. In: *Journal of Physics: Conference Series* 331 (2011), p. 022027. DOI: 10.1088/1742-6596/331/2/022027.
- [47] Twiki. *MDTOnlineDQMDocumentation*. (Internal document). URL: <https://twiki.cern.ch/twiki/bin/viewauth/Atlas/MDTOnlineDQMDocumentation>.
- [48] E Diehl. *The ATLAS Muon Calibration Stream: Design and Performance*. Tech. rep. Geneva: CERN, 2012. URL: <https://cds.cern.ch/record/1493565>.
- [49] Andrea Pareti. personal communication. Feb. 17, 2023.
- [50] Twiki. *MMOnlineHistogramPresenterDocumentation*. (Internal document). URL: <https://twiki.cern.ch/twiki/bin/view/Atlas/MMOnlineHistogramPresenterDocumentation>.
- [51] Luca Cavallini et al. “Tagging the Higgs boson decay to bottom quarks with colour-sensitive observables and the Lund jet plane”. In: *The European Physical Journal C* 82.5 (2022). DOI: 10.1140/epjc/s10052-022-10447-1. URL: <https://doi.org/10.1140/epjc/s10052-022-10447-1>.
- [52] “Configuration and performance of the ATLAS b -jet triggers in Run 2”. In: *Eur. Phys. J. C* 81.12 (2021). 61 pages in total, author list starting page 46, 17 figures, 10 tables, to be submitted to EPJC. All figures including auxiliary figures are available at <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/TRIG-2018-08>, p. 1087. DOI: 10.1140/epjc/s10052-021-09775-5. arXiv: 2106.03584. URL: <https://cds.cern.ch/record/2771727>.
- [53] M. Aaboud and G. Aad et.al. “Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector”. In: *Physics Letters B* 786 (2018), pp. 59–86. DOI: 10.1016/j.physletb.2018.09.013.

- [54] M. Aaboud and G. Aad et al. “Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector”. In: *Physics Letters B* 784 (2018), pp. 173–191. DOI: 10.1016/j.physletb.2018.07.035.
- [55] ATLAS Collaboration. “Measurement of the inclusive cross-section for the production of jets in association with a Z boson in proton–proton collisions at 8 TeV using the ATLAS detector”. In: *The European Physical Journal C* 79.10 (2019). DOI: 10.1140/epjc/s10052-019-7321-3. URL: <https://doi.org/10.1140/epjc/s10052-019-7321-3>.
- [56] ATLAS Collaboration. *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset*. 2022. arXiv: 2211.16345 [physics.data-an].
- [57] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “FastJet user manual”. In: *The European Physical Journal C* 72.3 (2012). DOI: 10.1140/epjc/s10052-012-1896-2. URL: <https://doi.org/10.1140/epjc/s10052-012-1896-2>.
- [58] *Tagging and suppression of pileup jets with the ATLAS detector*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2014-018>. Geneva: CERN, 2014. URL: <https://cds.cern.ch/record/1700870>.
- [59] Daniel Hay Guest. “Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment. IML Machine Learning Workshop”. In: (2017). URL: <https://cds.cern.ch/record/2256687>.
- [60] *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-013>. Geneva: CERN, 2017. URL: <https://cds.cern.ch/record/2273281>.
- [61] Alex Graves. “Supervised Sequence Labelling with Recurrent Neural Networks”. In: *Studies in Computational Intelligence*. 2012. URL: <https://api.semanticscholar.org/CorpusID:2118350>.
- [62] *Secondary vertex finding for jet flavour identification with the ATLAS detector*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-011>. Geneva: CERN, 2017. URL: <https://cds.cern.ch/record/2270366>.
- [63] *Topological b-hadron decay reconstruction and identification of b-jets with the JetFitter package in the ATLAS experiment at the LHC*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2018-025>. Geneva: CERN, 2018. URL: <https://cds.cern.ch/record/2645405>.
- [64] R Fruehwirth. *Application of Kalman filtering to track and vertex fitting*. 1987. DOI: 10.1016/0168-9002(87)90887-4.
- [65] *Identification of Boosted Higgs Bosons Decaying Into $b\bar{b}$ With Neural Networks and Variable Radius Subjets in ATLAS*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2020-019>. Geneva: CERN, 2020. URL: <http://cds.cern.ch/record/2724739>.
- [66] Jason Gallicchio and Matthew D. Schwartz. “Seeing in Color: Jet Superstructure”. In: *Phys. Rev. Lett.* 105 (2010), p. 022001. DOI: 10.1103/PhysRevLett.105.022001. arXiv: 1001.5027 [hep-ph].

- [67] Andrew J. Larkoski, Simone Marzani, and Chang Wu. “Theory Predictions for the Pull Angle”. In: *Phys. Rev. D* 99.9 (2019), p. 091502. DOI: 10.1103/PhysRevD.99.091502. arXiv: 1903.02275 [hep-ph].
- [68] Yunjia Bao and Andrew J. Larkoski. “Calculating Pull for Non-Singlet Jets”. In: *JHEP* 12 (2019), p. 035. DOI: 10.1007/JHEP12(2019)035. arXiv: 1910.02085 [hep-ph].
- [69] Andrew Larkoski, Simone Marzani, and Chang Wu. “Safe Use of Jet Pull”. In: *JHEP* 01 (2020), p. 104. DOI: 10.1007/JHEP01(2020)104. arXiv: 1911.05090 [hep-ph].
- [70] Andy Buckley et al. “An Optimal Observable for Color Singlet Identification”. In: *SciPost Phys.* 9 (2020), p. 026. DOI: 10.21468/SciPostPhys.9.2.026. arXiv: 2006.10480 [hep-ph].
- [71] Andrew J. Larkoski, Ian Mould, and Duff Neill. “Power Counting to Better Jet Observables”. In: *JHEP* 12 (2014), p. 009. DOI: 10.1007/JHEP12(2014)009. arXiv: 1409.6298 [hep-ph].
- [72] Ian Mould, Lina Necib, and Jesse Thaler. “New Angles on Energy Correlation Functions”. In: *JHEP* 12 (2016), p. 153. DOI: 10.1007/JHEP12(2016)153. arXiv: 1609.07483 [hep-ph].
- [73] Frédéric A. Dreyer, Gavin P. Salam, and Grégory Soyez. “The Lund Jet Plane”. In: *JHEP* 12 (2018), p. 064. DOI: 10.1007/JHEP12(2018)064. arXiv: 1807.04758 [hep-ph].
- [74] Frédéric A. Dreyer and Huilin Qu. *Jet tagging in the Lund plane with graph networks*. 2021. arXiv: 2012.08526 [hep-ph].
- [75] Charanjit K. Khosa and Simone Marzani. “Higgs boson tagging with the Lund jet plane”. In: *Physical Review D* 104.5 (2021). DOI: 10.1103/physrevd.104.055043. URL: <https://doi.org/10.1103/physrevd.104.055043>.
- [76] ATLAS Collaboration. “Measurement of the Lund Jet Plane Using Charged Particles in 13 TeV Proton-Proton Collisions with the ATLAS Detector”. In: *Physical Review Letters* 124.22 (2020). DOI: 10.1103/physrevlett.124.222002. URL: <https://doi.org/10.1103/physrevlett.124.222002>.
- [77] Andrew Lifson, Gavin P. Salam, and Grégory Soyez. “Calculating the primary Lund Jet Plane density”. In: *Journal of High Energy Physics* 2020.10 (2020). DOI: 10.1007/jhep10(2020)170. URL: [https://doi.org/10.1007/jhep10\(2020\)170](https://doi.org/10.1007/jhep10(2020)170).
- [78] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *Journal of High Energy Physics* 2014.7 (2014). DOI: 10.1007/jhep07(2014)079. URL: [https://doi.org/10.1007/jhep07\(2014\)079](https://doi.org/10.1007/jhep07(2014)079).
- [79] Torbjörn Sjöstrand et al. “An introduction to PYTHIA 8.2”. In: *Computer Physics Communications* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. URL: <https://doi.org/10.1016/j.cpc.2015.01.024>.
- [80] S. Ovyn, X. Rouby, and V. Lemaitre. *Delphes, a framework for fast simulation of a generic collider experiment*. 2010. arXiv: 0903.2225 [hep-ph].
- [81] J. de Favereau et al. “DELPHES 3: a modular framework for fast simulation of a generic collider experiment”. In: *Journal of High Energy Physics* 2014.2 (2014). DOI: 10.1007/jhep02(2014)057.

- [82] *fastjet-contrib*. URL: <https://fastjet.hepforge.org/contrib/>.
- [83] François Chollet et al. *Keras*. <https://keras.io>. 2015.
- [84] James Dolen et al. “Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure”. In: *Journal of High Energy Physics* 2016.5 (May 2016). ISSN: 1029-8479. DOI: 10.1007/jhep05(2016)156. URL: [http://dx.doi.org/10.1007/JHEP05\(2016\)156](http://dx.doi.org/10.1007/JHEP05(2016)156).
- [85] Ian Moutl, Benjamin Nachman, and Duff Neill. “Convolved substructure: analytically decorrelating jet substructure observables”. In: *Journal of High Energy Physics* 2018.5 (May 2018). ISSN: 1029-8479. DOI: 10.1007/jhep05(2018)002. URL: [http://dx.doi.org/10.1007/JHEP05\(2018\)002](http://dx.doi.org/10.1007/JHEP05(2018)002).
- [86] Georges Aad et al. “Determination of the parton distribution functions of the proton from ATLAS measurements of differential W and Z boson production in association with jets”. In: *JHEP* 07 (2021), p. 223. DOI: 10.1007/JHEP07(2021)223. arXiv: 2101.05095 [hep-ex].
- [87] CMS Collaboration. “Measurement of associated production of a W boson and a charm quark in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *The European Physical Journal C* 79.3 (2019). DOI: 10.1140/epjc/s10052-019-6752-1.
- [88] S. Alekhin, J. Blümlein, and S. Moch. “NLO PDFs from the ABMP16 fit”. In: *The European Physical Journal C* 78.6 (2018). DOI: 10.1140/epjc/s10052-018-5947-1.
- [89] ATLAS Collaboration. “Precision measurement and interpretation of inclusive W^+ , W^- and Z/γ^* production cross sections with the ATLAS detector”. In: *The European Physical Journal C* 77.6 (2017). DOI: 10.1140/epjc/s10052-017-4911-9.
- [90] S. Alekhin et al. “Parton distribution functions, α_s , and heavy-quark masses for LHC Run II”. In: *Phys. Rev. D* 96 (1 2017), p. 014011. DOI: 10.1103/PhysRevD.96.014011. URL: <https://link.aps.org/doi/10.1103/PhysRevD.96.014011>.
- [91] Sayipjamal Dulat et al. “New parton distribution functions from a global analysis of quantum chromodynamics”. In: *Physical Review D* 93.3 (Feb. 2016). ISSN: 2470-0029. DOI: 10.1103/physrevd.93.033006. URL: <http://dx.doi.org/10.1103/PhysRevD.93.033006>.
- [92] Tie-Jiun Hou et al. “New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC”. In: *Physical Review D* 103.1 (Jan. 2021). ISSN: 2470-0029. DOI: 10.1103/physrevd.103.014013. URL: <http://dx.doi.org/10.1103/PhysRevD.103.014013>.
- [93] L. A. Harland-Lang et al. “Parton distributions in the LHC era: MMHT 2014 PDFs”. In: *The European Physical Journal C* 75.5 (May 2015). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-015-3397-6. URL: <http://dx.doi.org/10.1140/epjc/s10052-015-3397-6>.
- [94] Richard D. Ball et al. “Parton distributions for the LHC run II”. In: *Journal of High Energy Physics* 2015.4 (Apr. 2015). ISSN: 1029-8479. DOI: 10.1007/jhep04(2015)040. URL: [http://dx.doi.org/10.1007/JHEP04\(2015\)040](http://dx.doi.org/10.1007/JHEP04(2015)040).
- [95] Ferran Faura et al. “The strangest proton?” In: *The European Physical Journal C* 80.12 (2020). DOI: 10.1140/epjc/s10052-020-08749-3. URL: <https://doi.org/10.1140/epjc/s10052-020-08749-3>.

- [96] H1 and ZEUS Collaborations. *Combination of Measurements of Inclusive Deep Inelastic $e^\pm p$ Scattering Cross Sections and QCD Analysis of HERA Data*. 2015. arXiv: 1506.06042 [hep-ex].
- [97] G. P. Zeller et al. "Precise Determination of Electroweak Parameters in Neutrino-Nucleon Scattering". In: *Phys. Rev. Lett.* 88 (9 2002), p. 091802. DOI: 10.1103/PhysRevLett.88.091802. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.88.091802>.
- [98] Martin Kueckes. "The HERMES experiment". In: *AIP Conf. Proc.* 338 (1995). Ed. by S. J. Seestrom, pp. 665–668. DOI: 10.1063/1.48484.
- [99] CDF Collaboration. In: *Physical Review Letters* 102.18 (2009). DOI: 10.1103/physrevlett.102.181801. URL: <https://doi.org/10.1103/physrevlett.102.181801>.
- [100] Timo Antero Aaltonen et al. "Measurement of the charge asymmetry of electrons from the decays of W bosons produced in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV". In: *Phys. Rev. D* 104.9 (2021), p. 092002. DOI: 10.1103/PhysRevD.104.092002. arXiv: 2107.04678 [hep-ex].
- [101] ATLAS Collaboration. "Measurement of differential cross sections and W^+/W^- cross-section ratios for W boson production in association with jets at $\sqrt{s} = 8$ TeV with the ATLAS detector". In: *Journal of High Energy Physics* 2018.5 (2018). DOI: 10.1007/jhep05(2018)077. URL: [https://doi.org/10.1007/jhep05\(2018\)077](https://doi.org/10.1007/jhep05(2018)077).
- [102] R. S. Towell et al. "Improved measurement of the $\bar{d}/\bar{u}$ asymmetry in the nucleon sea". In: *Phys. Rev. D* 64 (5 2001), p. 052002. DOI: 10.1103/PhysRevD.64.052002. URL: <https://link.aps.org/doi/10.1103/PhysRevD.64.052002>.
- [103] Tie-Jiun Hou et al. "New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC". In: *Phys. Rev. D* 103 (1 2021), p. 014013. DOI: 10.1103/PhysRevD.103.014013. URL: <https://link.aps.org/doi/10.1103/PhysRevD.103.014013>.
- [104] *Measurement of the production cross section of a W boson in association with a charm quark in proton-proton collisions at $\sqrt{s} = 13$ TeV*. Tech. rep. Geneva: CERN, 2022. URL: <https://cds.cern.ch/record/2815162>.
- [105] Richard D. Ball et al. "Parton distributions from high-precision collider data". In: *The European Physical Journal C* 77.10 (2017). DOI: 10.1140/epjc/s10052-017-5199-5. URL: <https://doi.org/10.1140/epjc/s10052-017-5199-5>.
- [106] R. S. Thorne et al. *Updates of PDFs in the MMHT framework*. 2019. arXiv: 1907.08147 [hep-ph].
- [107] Aad et al. "Determination of the parton distribution functions of the proton using diverse ATLAS data from pp collisions at $\sqrt{s} = 7, 8$ and 13 TeV." In: *Eur. Phys. J. C* 82.5 (2022). 92 pages in total, author list starting page 77, 41 figures, 7 tables, published in EPJC. All figures including auxiliary figures are available at <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/STDM-2020-32>, p. 438. DOI: 10.1140/epjc/s10052-022-10217-z. arXiv: 2112.11266. URL: <https://cds.cern.ch/record/2798650>.
- [108] ATLAS Collaboration. "Measurement of the production of a W boson in association with a charm quark in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector". In: *Journal of High Energy Physics* 2014.5 (2014). DOI: 10.1007/jhep05(2014)068. URL: [https://doi.org/10.1007/jhep05\(2014\)068](https://doi.org/10.1007/jhep05(2014)068).

- [109] James Catmore et al. *A New Petabyte-scale Data Derivation Framework for ATLAS*. Tech. rep. 7. Geneva: CERN, 2015. DOI: 10.1088/1742-6596/664/7/072007. URL: <https://cds.cern.ch/record/2016628>.
- [110] ATLAS Collaboration. “Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC”. In: (2022). DOI: 10.48550/ARXIV.2212.09379. arXiv: 2212.09379 [hep-ex].
- [111] G. Avoni et al. “The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS”. In: *JINST* 13.07 (2018), P07017. DOI: 10.1088/1748-0221/13/07/P07017.
- [112] <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/StandardMode1ASG>. Accessed: 2023-10-30. (Internal document).
- [113] GridPP. *The new Analysis Model for ATLAS*. URL: <https://indico.cern.ch/event/321806/contributions/746281/attachments/621903/855701/GridPP2014.pdf>.
- [114] <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TopDerivations>. Accessed: 2023-10-30. (Internal document).
- [115] <https://gaudi.web.cern.ch/gaudi/>. Accessed: 2023-10-28.
- [116] <https://atlassoftwaredocs.web.cern.ch/athena/athena-intro/>. Accessed: 2023-10-28.
- [117] <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MCPAnalysisGuidelinesMC16>. Accessed: 2021-07-21. (Internal document).
- [118] <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/EGammaRecommendationsR21>. Accessed: 2023-10-28. (Internal document).
- [119] <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/JetEtmissRecommendationsR21>. Accessed: 2023-10-28. (Internal document).
- [120] ATLAS Collaboration. “The ATLAS Simulation Infrastructure”. In: *Eur. Phys. J. C* 70 (2010), p. 823. DOI: 10.1140/epjc/s10052-010-1429-9. arXiv: 1005.4568 [physics.ins-det].
- [121] S. Agostinelli and al. “GEANT4 - a simulation toolkit”. In: *Nucl. Instr. Meth. A* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: [http://dx.doi.org/10.1016/S0168-9002\(03\)01368-8](http://dx.doi.org/10.1016/S0168-9002(03)01368-8). URL: <http://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- [122] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. “A Brief Introduction to PYTHIA 8.1”. In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].
- [123] *Summary of ATLAS Pythia 8 tunes*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2012-003>. Geneva: CERN, 2012. URL: <https://cds.cern.ch/record/1474107>.
- [124] R. Frederix et al. “The automation of next-to-leading order electroweak calculations”. In: *JHEP* 07 (2018). [Erratum: *JHEP* 11, 085 (2021)], p. 185. DOI: 10.1007/JHEP11(2021)085. arXiv: 1804.10017 [hep-ph].
- [125] Rikkert Frederix and Stefano Frixione. “Merging meets matching in MC@NLO”. In: *Journal of High Energy Physics* 2012.12 (2012). DOI: 10.1007/jhep12(2012)061. URL: [https://doi.org/10.1007/jhep12\(2012\)061](https://doi.org/10.1007/jhep12(2012)061).

- [126] Christian Bierlich et al. *A comprehensive guide to the physics and usage of PYTHIA 8.3*. 2022. arXiv: 2203.11601 [hep-ph].
- [127] Anders Ryd et al. “EvtGen: A Monte Carlo Generator for B-Physics”. In: (May 2005).
- [128] “Modelling and computational improvements to the simulation of single vector-boson plus jet processes for the ATLAS experiment”. In: *JHEP* 2208 (2022). 59 pages in total, author list starting page 44, 15 figures, 3 tables, published in *JHEP* 08 (2022) 089. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/PMGR-2021-01/>, p. 089. DOI: 10.1007/JHEP08(2022)089. arXiv: 2112.09588. URL: <https://cds.cern.ch/record/2798348>.
- [129] T Gleisberg et al. “SHERPA 1.α, a proof-of-concept version”. In: *Journal of High Energy Physics* 2004.02 (2004), pp. 056–056. DOI: 10.1088/1126-6708/2004/02/056. URL: <https://doi.org/10.1088/1126-6708/2004/02/056>.
- [130] Enrico Bothmann et al. “Event generation with Sherpa 2.2”. In: *SciPost Physics* 7.3 (2019). DOI: 10.21468/scipostphys.7.3.034. URL: <https://doi.org/10.21468/scipostphys.7.3.034>.
- [131] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. “A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction”. In: *JHEP* 09 (2007), p. 126. DOI: 10.1088/1126-6708/2007/09/126. arXiv: 0707.3088 [hep-ph].
- [132] Paolo Nason. “A new method for combining NLO QCD with shower Monte Carlo algorithms”. In: *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146.
- [133] Stefano Frixione, Paolo Nason, and Carlo Oleari. “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”. In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].
- [134] Simone Alioli et al. “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”. In: *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].
- [135] ATLAS Collaboration. *Studies on top-quark Monte Carlo modelling for Top2016*. ATL-PHYS-PUB-2016-020. 2016. URL: <https://cds.cern.ch/record/2216168>.
- [136] M. Beneke et al. “Hadronic top-quark pair production with NNLL threshold resummation”. In: *Nucl. Phys. B* 855 (2012), pp. 695–741. DOI: 10.1016/j.nuclphysb.2011.10.021. arXiv: 1109.1536 [hep-ph].
- [137] Matteo Cacciari et al. “Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation”. In: *Phys. Lett. B* 710 (2012), pp. 612–622. DOI: 10.1016/j.physletb.2012.03.013. arXiv: 1111.5869 [hep-ph].
- [138] Peter Bärnreuther, Michal Czakon, and Alexander Mitov. “Percent-Level-Precision Physics at the Tevatron: Next-to-Next-to-Leading Order QCD Corrections to $q\bar{q} \rightarrow t\bar{t} + X$ ”. In: *Phys. Rev. Lett.* 109 (2012), p. 132001. DOI: 10.1103/PhysRevLett.109.132001. arXiv: 1204.5201 [hep-ph].

- [139] Michal Czakon and Alexander Mitov. “NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels”. In: *JHEP* 12 (2012), p. 054. DOI: 10.1007/JHEP12(2012)054. arXiv: 1207.0236 [hep-ph].
- [140] Michal Czakon and Alexander Mitov. “NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction”. In: *JHEP* 01 (2013), p. 080. DOI: 10.1007/JHEP01(2013)080. arXiv: 1210.6832 [hep-ph].
- [141] Michal Czakon, Paul Fiedler, and Alexander Mitov. “Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_s^4)$ ”. In: *Phys. Rev. Lett.* 110 (2013), p. 252004. DOI: 10.1103/PhysRevLett.110.252004. arXiv: 1303.6254 [hep-ph].
- [142] Michal Czakon and Alexander Mitov. “Top++: A program for the calculation of the top-pair cross-section at hadron colliders”. In: *Comput. Phys. Commun.* 185 (2014), p. 2930. DOI: 10.1016/j.cpc.2014.06.021. arXiv: 1112.5675 [hep-ph].
- [143] Stefano Frixione et al. “Single-top hadroproduction in association with a W boson”. In: *JHEP* 07 (2008), p. 029. DOI: 10.1088/1126-6708/2008/07/029. arXiv: 0805.3067 [hep-ph].
- [144] Enrico Bothmann et al. “Event Generation with Sherpa 2.2”. In: *SciPost Phys.* 7.3 (2019), p. 034. DOI: 10.21468/SciPostPhys.7.3.034. arXiv: 1905.09127 [hep-ph].
- [145] Stefan Höche et al. “QCD matrix elements parton showers. The NLO case”. In: *Journal of High Energy Physics* 2013.4 (2013). DOI: 10.1007/jhep04(2013)027. URL: [https://doi.org/10.1007/jhep04\(2013\)027](https://doi.org/10.1007/jhep04(2013)027).
- [146] Thomas Gehrmann et al. “NLO QCD matrix elements parton showers in $ee \rightarrow$ hadrons”. In: *Journal of High Energy Physics* 2013.1 (2013). DOI: 10.1007/jhep01(2013)144. URL: [https://doi.org/10.1007/jhep01\(2013\)144](https://doi.org/10.1007/jhep01(2013)144).
- [147] <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/RecommendedIsolationWPs>. Accessed: 2021-07-21. (Internal document).
- [148] <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MCTruthClassifier>. Accessed: 2021-07-21. (Internal document).
- [149] ATLAS Collaboration. “Jet reconstruction and performance using particle flow with the ATLAS Detector”. In: *Eur. Phys. J. C* 77 (2017), p. 466. DOI: 10.1140/epjc/s10052-017-5031-2. arXiv: 1703.10485 [hep-ex].
- [150] Matteo Cacciari et al. “The anti-ktjet clustering algorithm”. In: *Journal of High Energy Physics* 2008.04 (2008), pp. 063–063. DOI: 10.1088/1126-6708/2008/04/063. URL: <https://doi.org/10.1088/1126-6708/2008/04/063>.
- [151] ATLAS Collaboration. “Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Eur. Phys. J. C* 81 (2020), p. 689. DOI: 10.1140/epjc/s10052-021-09402-3. arXiv: 2007.02645 [hep-ex].
- [152] ATLAS. *HowToCleanJetsR21*. (Internal document). URL: <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/HowToCleanJetsR21>.
- [153] <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/PileupJetRecommendations>. Accessed: 2021-07-21. (Internal document).

- [154] *Selection of jets produced in 13TeV proton-proton collisions with the ATLAS detector*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2015-029>. Geneva: CERN, 2015. URL: <https://cds.cern.ch/record/2037702>.
- [155] ATLAS. *Jet Cleaning in 2016*. URL: https://indico.cern.ch/event/506093/contributions/2307104/attachments/1338786/2023953/hcw16_jgonski.pdf.
- [156] Julia Gonski. *Jet Cleaning in 2016 and the Event Level Cleaning Tool*. Tech. rep. Geneva: CERN, 2017. URL: <https://cds.cern.ch/record/2272136>.
- [157] “Performance of electron and photon triggers in ATLAS during LHC Run 2”. In: *Eur. Phys. J. C* 80.1 (2020). 56 pages in total, author list starting page 40, 26 figures, 10 tables, published in EPJC. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/TRIG-2018-05>, p. 47. DOI: 10.1140/epjc/s10052-019-7500-2. arXiv: 1909.00761. URL: <https://cds.cern.ch/record/2688248>.
- [158] Georges Aad et al. “Performance of the ATLAS muon triggers in Run 2”. In: *JINST* 15.09 (2020), P09015. DOI: 10.1088/1748-0221/15/09/p09015. arXiv: 2004.13447 [physics.ins-det].
- [159] <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/JetEtmisRecommendationsR21>. Accessed: 2021-07-21. (Internal document).
- [160] ATLAS Collaboration. “Performance of electron and photon triggers in ATLAS during LHC Run 2”. In: *Eur. Phys. J. C* 80 (2020), p. 47. DOI: 10.1140/epjc/s10052-019-7500-2. arXiv: 1909.00761 [hep-ex].
- [161] ATLAS Collaboration. “Performance of the ATLAS muon triggers in Run 2”. In: *JINST* 15 (2020), P09015. DOI: 10.1088/1748-0221/15/09/p09015. arXiv: 2004.13447 [hep-ex].
- [162] <https://twiki.cern.ch/twiki/bin/view/Atlas/TrigEgammaRecommendations2015>. Accessed: 2021-07-21. (Internal document).
- [163] <https://twiki.cern.ch/twiki/bin/view/Atlas/MuonTriggerPhysicsRecommendationsRel212017>. Accessed: 2021-07-21. (Internal document).
- [164] <https://twiki.cern.ch/twiki/bin/view/Atlas/TrigGlobalEfficiencyCorrectionTool>. Accessed: 2021-07-21. (Internal document).
- [165] ATLAS Collaboration. “Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV”. In: *Eur. Phys. J. C* 81 (2021), p. 578. DOI: 10.1140/epjc/s10052-021-09233-2. arXiv: 2012.00578 [hep-ex].
- [166] ATLAS Collaboration. “Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV”. In: *Eur. Phys. J. C* 76 (2016), p. 292. DOI: 10.1140/epjc/s10052-016-4120-y. arXiv: 1603.05598 [hep-ex].
- [167] ATLAS Collaboration. “Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data”. In: *JINST* 14 (2019), P12006. DOI: 10.1088/1748-0221/14/12/P12006. arXiv: 1908.00005 [hep-ex].
- [168] <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/BTagCalibrationRecommendationsRelease21>. Accessed: 2021-07-21. (Internal document).

- [169] <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/MUON-2018-004/index.html>. Accessed: 2023-03-08.
- [170] ATLAS Collaboration. “Measurement of the top-quark mass using a leptonic invariant mass in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: (2022). arXiv: 2209.00583 [hep-ex].
- [171] <https://gitlab.cern.ch/atlas/athena/-/blob/21.2/PhysicsAnalysis/MCTruthClassifier/MCTruthClassifier/MCTruthClassifierDefs.h>. Accessed: 2023-03-08. (Internal document).
- [172] Glen Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. In: *The European Physical Journal C* 71.2 (2011). DOI: 10.1140/epjc/s10052-011-1554-0. URL: <https://doi.org/10.1140/epjc/s10052-011-1554-0>.
- [173] ATLAS Michele Pinamonti Loic Valery. *TRExFitter documentation*. URL: <https://trexfitter-docs.web.cern.ch/trexfitter-docs/>.
- [174] Kyle Cranmer et al. *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*. Tech. rep. New York: New York U., 2012. URL: <https://cds.cern.ch/record/1456844>.
- [175] Wouter Verkerke and David Kirkby. *The RooFit toolkit for data modeling*. 2003. arXiv: physics/0306116 [physics.data-an].
- [176] Lorenzo Moneta et al. *The RooStats Project*. 2011. arXiv: 1009.1003 [physics.data-an].
- [177] “ATLAS b -jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV”. In: *Eur. Phys. J. C* 79.11 (2019). 52 pages in total, author list starting page 36, 12 figures, 6 tables, submitted to EPJC. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/FTAG-2018-01/>, p. 970. DOI: 10.1140/epjc/s10052-019-7450-8. arXiv: 1907.05120. URL: <https://cds.cern.ch/record/2682119>.
- [178] ATLAS Adrea Cocco. *Instructions for calibration analyzers or for creating user CDI information*. (Internal document). URL: <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/BTaggingCalibrationDataInterfaceForCalibrators>.
- [179] https://twiki.cern.ch/twiki/pub/Main/ABCDMethod/ABCDGuide_draft18Oct18.pdf. Accessed: 2024-01-05.
- [180] *TRExFitter Tutorial - part 2: Data-driven Backgrounds and Unfolding*. 2020. URL: https://docs.google.com/presentation/d/1q_947Mya_rDXHkVUbima5dJS7nHrsK7R5PI3ZEsuwms/edit#slide=id.p.
- [181] G. D’Agostini. *Improved iterative Bayesian unfolding*. 2010. arXiv: 1010.0632 [physics.data-an].
- [182] Tommaso Cresta. personal communication. Dec. 6, 2023.
- [183] Tim Adye. *Unfolding algorithms and tests using RooUnfold*. 2011. arXiv: 1105.1160 [physics.data-an].

- [184] “Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC. Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC”. In: *Eur. Phys. J. C* 83.10 (2023). All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/DAPR-2021-01>, p. 982. DOI: 10.1140/epjc/s10052-023-11747-w. arXiv: 2212.09379. URL: <https://cds.cern.ch/record/2844887>.