



UNIVERSIDAD NACIONAL DE LA PLATA
FACULTAD DE CIENCIAS EXACTAS
INSTITUTO DE FÍSICA LA PLATA

Trabajo de Tesis Doctoral

My thesis title

My subtitle

Autor:
Francisco SILI

Directores:

Prof. Dr. Maria Teresa DOVA
Dr. Francisco ALONSO

12th October 2024

ACKNOWLEDGEMENTS

akwnoelasfsa

I, Francisco SILI, hereby declare that this thesis has not been and will not be,
submitted in whole or in part to another university for the award of any other degree.

La Plata,
12th October 2024

Francisco SILI

Universidad Nacional de La Plata
Facultad de Ciencias Exactas
Instituto de Física La Plata

TESIS DE DOCTORADO

My thesis title

by Francisco SILI

ABSTRACT

some text

CONTENTS

Contents	vi
List of Figures	ix
List of Tables	xii
Introduction	2
I Theory Motivation	3
1 The Standard Model and Beyond	4
II Experimental setup	5
2 The ATLAS Experiment	6
2.1 LHC	6
2.2 ATLAS	8
2.2.1 ATLAS Coordinate system	9
2.2.2 Inner Detector	10
2.2.3 Calorimeters	11
2.2.4 Muon spectrometer	15
2.2.5 The Trigger System	16
2.3 Data-taking during Run-2	17
3 Reconstruction and identification of physical objects	20
3.1 Track and vertex reconstruction	20
3.2 Photons and electrons	22
3.2.1 Reconstruction	22
3.2.2 Identification	23
3.2.3 Isolation	24
3.3 Muons	25

3.4	Jets	26
3.4.1	Anti- k_t jet clustering algorithm	27
3.4.2	Calorimeter Jets	27
3.4.3	PFlow Jets	28
3.4.4	Jet calibration	29
3.5	Jet flavor tagging	30
3.5.1	b -jet identification performance	31
3.5.2	c -jet identification performance	32
 III Photon shower shape corrections		 33
 4	Shower shapes and photon identification	 34
4.1	Shower shapes	34
4.2	Photon Identification	37
4.2.1	Processes and event selection	38
4.2.2	Optimisation	40
4.2.3	Efficiency measurements	40
4.3	Shower shapes variables differences between data and MC	42
 5	Shower shapes corrections	 45
5.1	Fudge Factors (FFs)	45
5.1.1	Data and simulated samples	45
5.1.2	Calculation	46
5.1.3	Uncertainties	49
5.1.4	Results	50
5.2	Cell-based energy reweighting	53
5.2.1	Event selection	54
5.2.2	Calculation	55
5.2.3	Results	57
5.3	Conclusions and future work	58
5.3.1	Future work	59
 IV New Physics		 60
 6	Analysis motivation and strategy	 61
 7	Signal and Background samples	 62
 8	Event selection	 63
 9	Background estimation	 64
 10	Background modeling	 65

11 Systematic uncertainties	66
12 Statistical analysis	67
13 Results	68
Conclusions	69
A Ellipse fitting formulae	70
Glossary	71
Bibliography	74

LIST OF FIGURES

2.1	Overview of the Large Hadron Collider (LHC) complex where all the accelerators that lead up to the LHC are shown [3].	7
2.2	Overview of the A Toroidal LHC ApparatuS (ATLAS) detector and all its sub-detectors, including the systems added during the Long Shut down 2 (LS2) [11].	8
2.3	ATLAS coordinate system [11].	9
2.4	ATLAS Inner Detector (ID) diagrams showing the different submodules, with their corresponding dimensions.	10
2.5	ATLAS calorimeter system, showing the Electromagnetic Calorimeter (ECAL) and the Hadronic Calorimeter (HCAL) [21].	12
2.6	Segment of the ECAL showing the layer arrangement and cell dimensions in each layer [7].	13
2.7	Radiation lengths as a function of $ \eta $ for the ECAL, separated for each sublayer [7].	14
2.8	ATLAS Muon Spectrometer (MS) [7].	15
2.9	Luminosity delivered by the LHC and recorded by ATLAS during the Run-2 [31] and Run-3 data taking periods. For Run-2, the fraction of data good for physics analyses is also displayed.	18
2.10	Pileup conditions during Run-2 and Run-3.	19
3.1	Schematic showing the tracking parametrization [36].	21
3.2	Diagrama showing the reconstruction algortihm workflow for electrons and photons, extacted from Ref. [41]	23
3.3	Diagram showing the calculation of the calorimetric isolation variable. When $R = 0.4$, $E_T^{\text{cone}40}$ is computed.	24
3.4	Schematic representation of the anti- k_t algorithm for jet clustering [45].	28
3.5	PFLow 4-momemtum jet calibration steps [48].	29
3.6	GN2 tagger discriminant comparison between data and single-lepton $t\bar{t}$ Monte Carlo (MC) simulation. The l -, b - and c -jets are contributions are shown with different colors, and the 5 b -tag Working Points (WPs) shown with the dashed vertical lines. From left to right, the dashed lines represent the 90, 85, 77, 70 and 65% efficiency WPs. The lower pad shows the ratio between data and the stacked MC [57].	31

4.1	Schematic representation of the photon Shower Shapes (SSs). The values $E_C^{S_N}$ represent the energy in layer N of the ECAL in a cluster C	35
4.2	Characteristic energy deposits by an isolated photon (left), and a $\pi^0 \rightarrow \gamma\gamma$ event (right), which is possible to distinguish thanks to the granularity of the first ECAL layer [58].	37
4.3	Feynman diagrams for radiative Z decays $Z \rightarrow \ell\ell\gamma$ for the initial-state (left) and final-state (right) radiation.	38
4.4	Two-dimensional invariant mass distribution of the two and three body systems $m_{\ell\ell}$ and $m_{\ell\ell\gamma}$, respectively, for (a) data, (b) background and (c) signal. The region in which there is a high concentration of events for $m_{\ell\ell} \sim m_Z$ corresponds to Initial State Radiation (ISR) events, while Final State Radiation (FSR) events can be identified when $m_{\ell\ell\gamma} \sim m_Z$	39
4.5	Normalised signal (blue) and background (orange) distributions for different SSs, using Radiative Z (RZ) event samples, passing the event selection detailed in Section 4.2.1.	40
4.6	Comparison of the measurements between data and MC of the three distinct data-driven methods to compute efficiencies. In both figures, for each method, two different set of MC measurements are displayed: the nominal one and the corrected one. The bottom panels show the ratio of data efficiencies to MC predictions (referred as Scale Factors (SFs) in the text). The figures are taken from Ref. [61].	42
4.7	Photon identification SFs in the different p_T - $ \eta $ bins for both converted (left) and unconverted photons (right). Ask fran for the plots in his presentation!	43
4.8	Example of SSs comparisons between data (black dots) against nominal (red line) and corrected (blue line) MC simulation, using the full Run-2 RZ photon sample [64].	43
5.1	Adaptive and non-adaptive Kernel Density Estimator (KDE) smoothing.	47
5.2	KDE smoothing applied to the R_{had} for photons in $0.8 < \eta < 1.15$ in two scenarios: low and high statistics. Original histograms are shown with the blue points and their Probability Density Functions (PDFs) with the orange lines. The fine factors used for each case are displayed in the figure.	47
5.3	Calculation of shift+stretch FFs for f_{side} with unconverted photons.	49
5.4	Shift and stretch FFs for R_η (left) and $w_{\eta 2}$ (right) SSs for converted photons as a function of the photon p_T with $1.15 < \eta < 1.37$. Results from RZ (black) and Single Photon (SP) (blue) are displayed, where the shaded regions denote the total uncertainty, while the error bars only the statistical one. Stretch values are shown in the bottom pad. The shift, on the other hand, is shown in the top pad, which are normalised by the standard deviation of the SS after applying the stretch, as indicated in the text.	51

5.5 Shift and stretch FFs for $w_{s\text{ tot}}$ SSs using converted photons as a function of $ \eta $ and photons from the SP with $50 < p_T < 60$ GeV. Normalised shift values are displayed in Figure a while original, or raw, values are shown in Figure b. Points with the uncertainty bar show the statistical uncertainty only, and the shaded regions represent the total uncertainty. Stretch values are shown in the bottom pad. Shift values are displayed in the top pad, which are normalised by the standard deviation of the SS after applying the stretch.	52
5.6 Selected SSs distributions using the RZ samples for converted photons after applying the FFs corrections event-by-event to the MC simulation. The SSs are shown in the barrel region (top) and in the endcap (bottom). Data points are represented by the black points and uncorrected (corrected) MC simulation by the blue (green) lines. The bottom pads show the ratio of data to the corresponding MC simulation.	52
5.7 Same as Figure 5.6 but with the SP samples.	53
5.8 Cells arrangement and average energy distribution amongst the clusters.	54
5.9 Energy reweights to be applied to MC cells, for unconverted photons with $0.4 < \eta < 0.6$. fix luminosity	56
5.10 Example of two reweights matrices, using shift (left) and stretches (right) corrections. Shown results correspond to unconverted photons. The shift values are multiplied by a factor of 100 to improve visualisation.	57
5.11 Normalized cell energy distributions for cells 28, 39 and 50 of the cluster using unconverted photons. Blue and red points correspond to the reweighted and original MC distributions, respectively, while the grey histogram represents data.	58
5.12 SSs distributions for unconverted (top row) and converted (bottom row) photons in $ \eta < 0.6$. Data is represented by the grey histogram. Uncorrected MC simulation is shown with the red points and reweighted (fudged) MC with the blue (green) line.	59
A.1 Parameters of the most general case of an ellipse.	70

LIST OF TABLES

2.1	General performance goals of the ATLAS detector. The units of p_T and E are in GeV. Extracted from Ref. [7]	9
3.1	Summary of electron and photon isolation WPs use throughout this thesis.	25
3.2	Measured b -tagging efficiencies and c -jets, l -jets and τ -jets rejections in the low and high- p_T regime.	32
3.3	Measured c -tagging efficiencies and b -jets, l -jets and τ -jets rejections in the low and high- p_T regime. The values shown correspond to those after applying the b -tagging 77% WP veto and the 50% c -tagging WP. rejection values not correct!	32
4.1	Discriminative SSs used for photon identification. The three columns on the right denote whether the variable is used for the <i>loose</i> (L), <i>medium</i> (M) or <i>tight</i> (T) identification WP or not.	35

TO DO'S AND NOTES TO KEEP IN MIND

use **orange** to highlight that there needs to be made sure that there is a discussion in previous chapters - in editing clarify where that discussion should happen!

purple: this needs a reference, have used from memory or notes

red: open question

Fixes, to dos

- test
- need to include a definition on met in the object definition part?

Thoughts to work with

- have to be consistent with times in the description - discuss with Fab about it

Might be good to answer for viva preps

- how is the reconstruction considered overall? are there different

INTRODUCTION

This thesis presents a search for new phenomena in high-mass final states with a photon and a jet in proton-proton (pp) collisions at a centre-of-mass energy of 13 TeV using data collected by the [ATLAS](#) detector. [ATLAS](#) (“A Toroidal LHC ApparatuS”) is one of the two general-purpose detectors at the Large Hadron Collider ([LHC](#)) and the biggest multi-purpose particle detector ever built. It is used to investigate a wide range of physics, from Standard Model ([SM](#)) measurements, such as precision tests of quantum chromodynamics or study of the properties of the Higgs boson, to the search of new phenomena like extra dimensions and dark matter candidates. The LHC, built by the European Organization for Nuclear Research ([CERN](#)) and installed in a 27-kilometre circular tunnel, is the world’s largest and most powerful particle collider. This machine is capable of colliding energetic beams of protons (or heavier nuclei) at rates upward of millions per second. The precision and high beam energy of the LHC allow to explore the tera-electronvolt scale, an energy range never before achieved in a particle collider.

The thesis is divided into three parts. The first part describes the theory background and motivations of the work, in which the first chapter describes the Standard Model ([SM](#)) showing the excellent agreement there is between the theory and the experimental data. It also shows the actual problems of the [SM](#), which motivates the search for New Physics.

The second part describes the experiment. CHAPTER describes the [ATLAS](#) experiment, with each sub-detector in detail, to finally describes how the samples for data analysis are produced. The reconstruction of the different physics objects is explained in CHAPTER. At the end of the second part, a special chapter is dedicated to the photon identification, where a new method for correcting the shower shapes is explained and detailed.

The third part of the thesis is dedicated to the photon+jet resonance search. In CHAPTER, the general strategy and motivation of the search is discussed. The samples generation is discussed in CHAPTER, where the final event selection is given in CHAPTER. The methods for the background estimation as well as background modeling is given in CHAPTER AND CCHAPETER. The systematic uncertainties are discussed in CHAPTER, and finally in CHAPTER and CHAPTER, the statistical analysis with the results are shown.

Finally, the conclusions of the work are present in CHAPTER.

Part I

Theory Motivation

THE STANDARD MODEL AND BEYOND

1

*"Nothing in life is to be feared. It is
only to be understood. Now is the
time to understand more, so that we
may fear less"*

Marie Curie

another template text

Part II

Experimental setup

THE ATLAS EXPERIMENT

2

Something.

Someone

The work in this thesis has been performed using data from the [ATLAS](#) detector, one of the particle detectors recording collisions of protons accelerated by the [LHC](#) particle accelerator at European Organization for Nuclear Research ([CERN](#)). In the following chapter, an introduction to the [LHC](#) is given in Section 2.1, followed by a discussion of the [ATLAS](#) detector in Section 2.2. The discussion is focused on aspects important to the analyses of this thesis.

[update chapter, see version in spanish](#)

2.1 LHC

The [LHC](#) [1, 2] is the largest hadron accelerator in the world, located at [CERN](#), in the French-Swiss border. It has a longitude of 27 km, located between 50 and 174 meters underground. The [LHC](#) is designed to collide protons (and heavy ions) at a center of mass energy of 14 TeV. To keep the protons and heavy ions on the accelerator ring, overall 9593 magnets are used. These magnets include superconducting dipole and quadrupole magnets, cooled down to 1.9 K (-271 °C). The dipole magnets generate a magnetic field of 8.3 T.

In Figure 2.1 a general overview of the [LHC](#) accelerator facilities is shown. The protons are sourced from hydrogen gas by stripping its electrons and are accelerated in a first linear accelerator (LINAC2) to 50 MeV. Subsequently, the protons are successively accelerated in the Proton Synchrotron Booster ([PSB](#)), the Parton Shower ([PS](#)), and the Super Proton Synchrotron ([SPS](#)), where they reach an energy of 450 GeV before being injected into the [LHC](#). Overall 8 radiofrequency cavities can push the energy of the protons in the [LHC](#) up to 14 TeV.

The protons are injected as bunches of $\mathcal{O}(10^{11})$ protons into the [LHC](#) with a spacing of 25 ns (7.5 m). These bunches are later brought to collision in so-called bunch crossings. The filling scheme of the pre-accelerator chain, in combination with finite switching times of the injection and dumping magnets, results in regular patterns of filled and empty bunches.

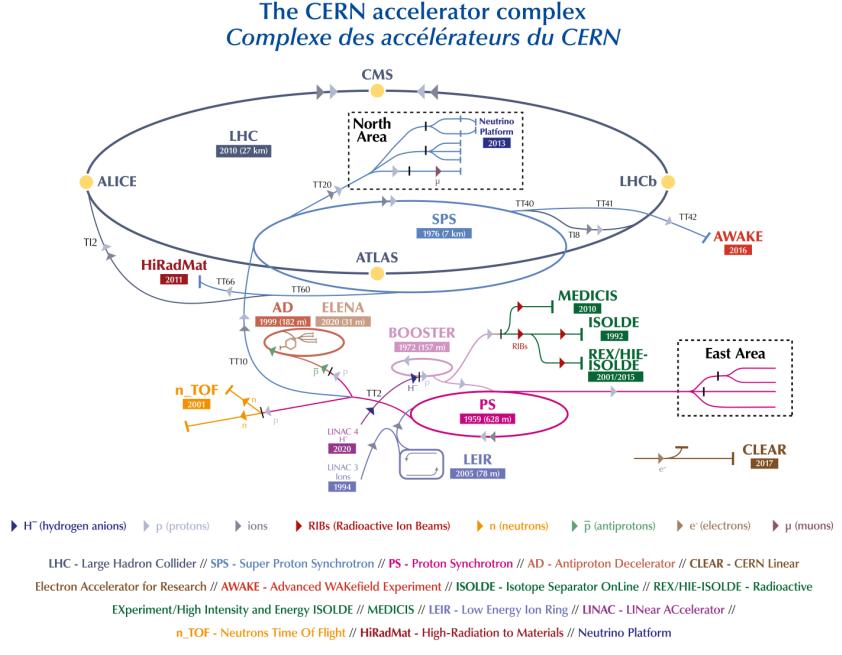


Figure 2.1: Overview of the LHC complex where all the accelerators that lead up to the LHC are shown [3].

The LHC so far provided proton and heavy ion beams for two data-taking periods, and is undergoing a third. Between 2009 and 2013 (known as Run 1), the LHC operated with centre-of-mass energy (\sqrt{s}) of 7 TeV and 8 TeV. After a long shutdown (LS1), the second run (Run 2) started in 2015 and ended in 2018, providing 13 TeV collisions to the experiments around the LHC ring. In 2022 the Run-3 began, at which $p\bar{p}$ collisions happen at an energy of 13.6 TeV, estimated to run until 2026. The four yellow dots in Figure 2.1 there are four interaction points, housing the ALICE [4], LHCb [5], CMS [6], ATLAS [7], LHCf [8], TOTEM [9], MoEDAL [10] experiments, among many other experiments.

One of the most important parameters to characterize the functioning of the accelerator is the instantaneous luminosity \mathcal{L} , defined as the number of particles per unit time per unit area, and can be calculated from the relation

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} F \quad (2.1)$$

where N_b is the number of particles per bunch, n_b the bunches per beam, γ_r is the relativistic gamma factor, ϵ_n is the normalised transverse beam emittance and β^* being the beta function at the collision point which determining the transverse spread of the particle beam. The correction term F takes into account the beam crossing angle. The revolution frequency is represented by f_{rev} which is ~ 11 kHz, and with the bunch-spacing of 25 ns, allows for beam crossing at the four interaction points with a frequency of ~ 40 MHz.

A measure for the total recorded data is the integrated luminosity over time is given by

$$N_{event} = L_{int} \sigma_{event} = \sigma_{event} \int \mathcal{L} dt, \quad (2.2)$$

connecting the luminosity with the number of events. More details regarding the luminosity measurements in ATLAS are shown in Section 2.3.

2.2 ATLAS

[ATLAS](#) is one of the multi-purpose detectors of the [LHC](#), located at Point-1 along the [LHC](#). It was designed and built to study the pp (and heavy ion) collisions at the TeV scale.

The overall shape of the detector is that of a cylinder as shown in Figure 2.2. It has a length of 44m and 25m in diameter, being the largest particle detector built so far. The [ATLAS](#) detector is divided geometrically in two parts: the central part called *barrel*, and the outer caps called *end-caps*.

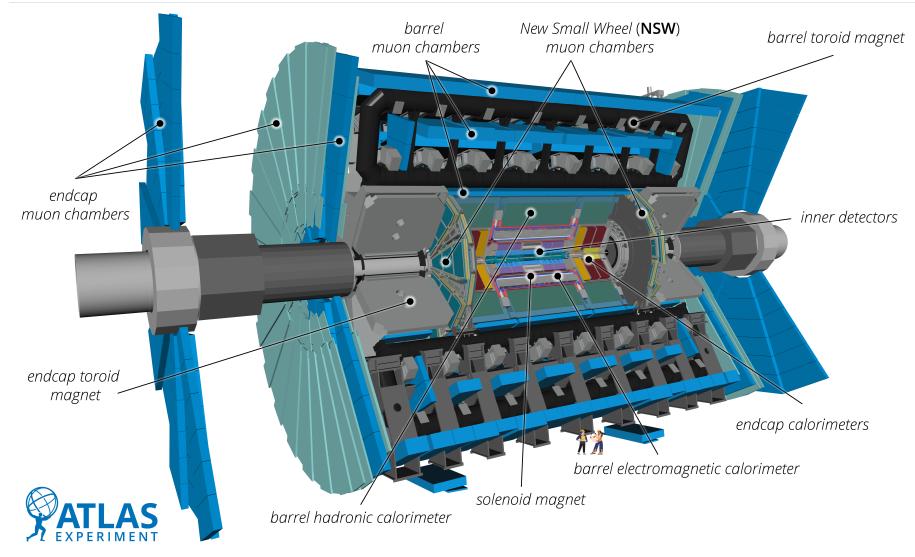


Figure 2.2: Overview of the [ATLAS](#) detector and all its sub-detectors, including the systems added during the LS2 [11].

[ATLAS](#) is built in layers of sub-detectors, each of which designed to have a different role on the identification and reconstruction of particles produced in the collisions. [ATLAS](#) provides hermetic coverage around the beam axis, enabling detection of all charged particles generated in the collisions in the plane orthogonal to the beam axis. This is particularly important in searches for new physics, relying on analyses of momentum balances in the orthogonal plane.

It is built up of multiple layers, starting from the innermost component, the Inner Detector ([ID](#)), providing tracking hits close to the beam pipe. Around the [ID](#), there is a superconductor solenoid which creates an axial magnetic field of ~ 2 T to curve the [ID](#) tracks of charged particles. After the first magnet, there is a system of two calorimeters: the [ECAL](#) and [HCAL](#). The former is in charge of measuring the kinetic energy of photons and electrons, and the latter measures the energy of the jets. The outermost parts of [ATLAS](#) are built by the muon spectrometer, providing momentum reconstruction for muons passing through the inner detector layers. Intertwined with the muon spectrometer, there are a total of 8 barrel toroid coils, providing a total magnetic field of 4 T (0.5 T per coil) to measure the momentum of muons. The toroid magnetic field is completed by the end-cap toroids, also generating a magnetic field up to 4T for muons leaving [ATLAS](#) close to the beam pipe.

Every component in [ATLAS](#) working together enables the reconstruction and identification of a

Table 2.1: General performance goals of the [ATLAS](#) detector. The units of p_T and E are in GeV. Extracted from Ref. [7]

Detector Component	Required resolution	η coverage	
		Measurement	Trigger
Tracking	$\sigma_{p_T}/p_T = 0.05\% p_T \oplus 1\%$	± 2.5	
EM calorimetry	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$	± 3.2	± 2.5
Hadronic calorimetry (jets)			
barrel and end-cap	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$	± 3.2	± 3.2
forward	$\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$	$3.1 < \eta < 4.9$	$3.1 < \eta < 4.9$
Muon spectrometer	$\sigma_{p_T}/p_T = 10\% \text{ at } p_T = 1 \text{ TeV}$	± 2.7	± 2.4

variety of particles with high precision. An overview of the design capabilities of [ATLAS](#) in terms of the momentum and energy resolution is given in Table 2.1, adapted from Ref. [7]. Here the resolution given lists first a stochastic term, measuring the uncertainty based on the statistically dominated interaction of a particle with the material, followed by a noise term, which accounts for uncertainties due to electronic noise in the readout process.

2.2.1 ATLAS Coordinate system

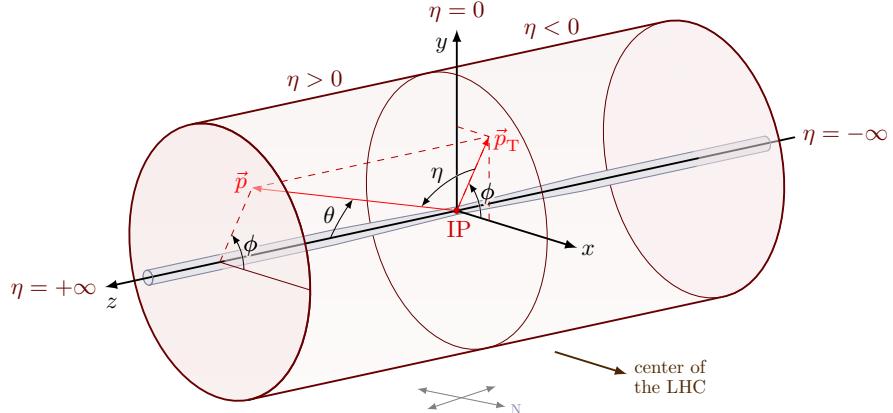


Figure 2.3: [ATLAS](#) coordinate system [11].

The coordinate system used within [ATLAS](#), displayed in Figure 2.3, is used throughout this thesis and shortly described in the following [7]. The origin of the right-handed coordinate system is at the nominal interaction point, with the positive x-axis pointing towards the centre of the [LHC](#). The x-y plane is perpendicular to the beam axis, defining the z-axis. Towards the surface defines the positive y-axis. An azimuthal angle ϕ is defined around the beam axis, and a polar angle θ is the angle from the beam axis. Instead of θ the rapidity y is used for heavy objects:

$$y = \frac{1}{2} \ln[(E + p_z)/(E - p_z)]. \quad (2.3)$$

Differences in rapidity are invariant under boosts along the beam axis. For massless objects or relativistic objects ($m \ll p$), the pseudorapidity is used instead:

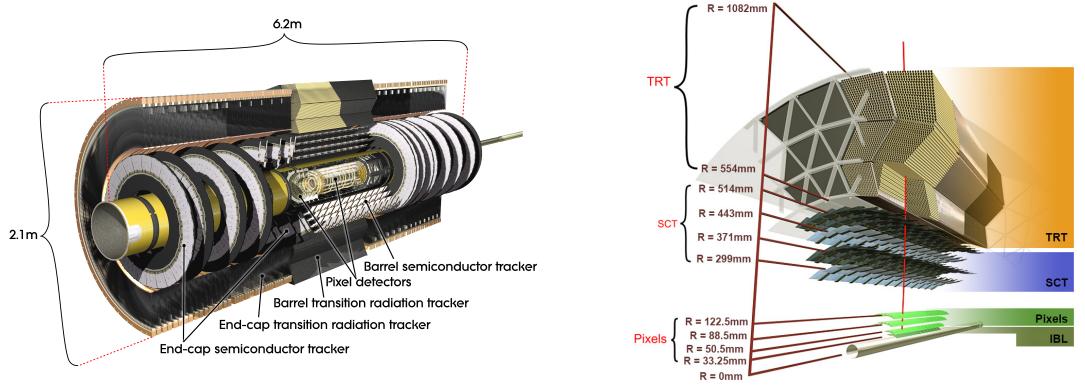
$$\eta = -\ln(\tan(\theta/2)). \quad (2.4)$$

To quantify the distance between two objects, ΔR is defined:

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} \quad (2.5)$$

The transverse momentum and energy are defined in the x-y plane, with the transverse momentum given as $p_T = \sqrt{p_x^2 + p_y^2}$.

2.2.2 Inner Detector



(a) ID with all its submodules in the barrel and endcap regions [12]. (b) Layers of the ID showing their distance to the beam [13].

Figure 2.4: ATLAS ID diagrams showing the different submodules, with their corresponding dimensions.

A cross-section of the Inner Detector (ID) system [14] is shown in Figure 2.4, highlighting the distance of each subsystem from the beampipe. The innermost part of the ID is the Insertable B-Layer (IBL), followed by three layers of pixel detectors. At 299 mm radial distance from the beam pipe, four layers of SemiConductor Tracker (SCT) modules are located before the Transition Radiation Tracker (TRT), which extends the overall ID detector size to a radius of 1082 mm. The ID allows for particle track reconstruction within $|\eta| < 2.5$.

The role of the ID is the trajectory tracking of charged particles to determine their charge and momentum. It is immersed in a 2 T magnet field generated by the ATLAS solenoid magnet system, that bends the trajectories of charged particles. The curvature radius is proportional to the particle momentum and its direction distinguishes positive from negative charges. The detected particle tracks allow for the reconstruction of primary collision vertices, which is important to distinguish pile-up collisions from the collision of interest, and of secondary decay vertices of longer-lived particles, which is crucial for the identification of e.g. B mesons or τ leptons.

IBL - Insertable B-layer After Run-1, during a long shutdown in 2013-2014, the pixel detector system was subject to maintenance and upgrades. Within this set of upgrades, a 4th pixel layer at a 3.3 cm distance from a new, smaller beam pipe (33 mm outer radius, originally 36 mm), which was the first in particle physics experiments [15, 16] and has led to significant improvements in interaction vertex reconstruction and identification of b-hadron jets.

Pixel Detector The innermost pixel layer, the IBL, is surrounded by three layers of pixel detectors, arranged in barrels around the beam pipe [17, 18]. The method of detection of charged particles is the measurement of deposited induced charges in a silicon layer, product of ionization. The first layer is at a distance of 50.5 mm from the beam pipe's centre. As can be seen in Figure 2.4a, the end caps of the pixel layer consist of 3 disks around the beampipe, stretching the length of the pixel component of the ID to 1.4 m length along the beam axis. The pixel detector consists of overall 1744 pixel modules with a nominal size of $50\mu\text{m} \times 400\mu\text{m}$ in the (ϕ, z) plane (ϕ, r for the disk panels), comprising over 80 million readout channels. The pixel and IBL part of the ATLAS detector is crucial for tracking, providing 4 pixel hits over the entire ID pseudorapidity coverage ($|\eta| < 2.5$.).

Semiconductor Tracker The pixel detector and IBL are located within SCT modules [19]. Similar to the pixel detector modules, the SCT modules are semiconductor-based, arranged into cylindrical layers around the beampipe in the barrel region, forming disks in the endcap. Since the SCT modules only provide precise location along one axis, two modules are combined back-to-back and rotated against each other to gain two dimensional spacial information. Four layers are arranged in the barrel, nine disks in each endcap side (see Figure 2.4a). Including the endcap disks, the SCT extends up to $|z| < 2735\text{mm}$.

Transition Radiation Tracker The last part of the ID is the TRT [20], in the barrel stretching from 554 mm to 1082 mm radial distance. This detector is composed of 4 mm diameter straw tubes, arranged in parallel to the beam pipe or radially in the barrel and end-cap, respectively. Within $|\eta| < 2.0$, three barrel rings and 18 end-cap units provide typically 36 hits per track. The straws are intertwined with polypropylene fibres for passing through particles to create transition radiation. Inside the straws is a thin tungsten wire, collecting charges drifting through the straws gas mixture (Xe, CO₂ and O₂). The level of radiation and collected charges in each straw can be used to discriminate between electrons and charged pions. The TRT only offers spatial information in the $(R - \phi)$ plane, no information in the z-direction can be extracted due to the straws orientation. There is a total of 50000 tubes in the barrel region, while the end-caps contain 320000 tubes.

2.2.3 Calorimeters

As previously mentioned, the ID system is surrounded by two calorimeters: the Electromagnetic Calorimeter (ECAL) and the Hadronic Calorimeter (HCAL), as shown in Figure 2.5. These calorimeters are designed to measure the energy and position of the incident particles, via the deposited energy by the secondary particle cascades produced by the incident ones. It covers the whole ϕ range and up to $|\eta| < 4.9$, with a finer granularity in the region that coincides with the ID. The calorimeter system allows for the discrimination between photons and electrons from hadrons (jets). Furthermore, it allows to measure the energetic imbalance (thanks to its

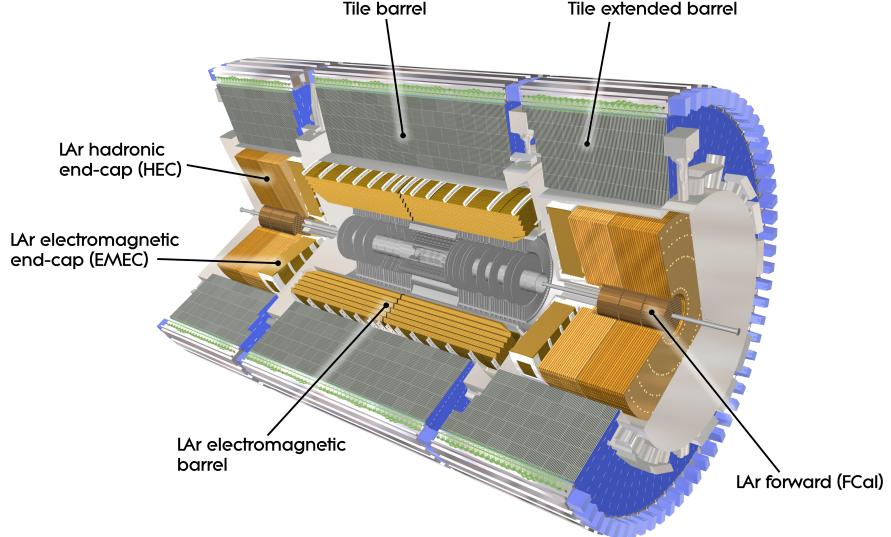


Figure 2.5: ATLAS calorimeter system, showing the Electromagnetic Calorimeter (ECAL) and the Hadronic Calorimeter (HCAL) [21].

total coverage and hermiticity) and it provides the trigger system with the necessary information for the event selection.

Both calorimeters are so-called sampling calorimeters with alternating layers of absorber and active material. The absorber layer triggers a shower development of consecutive interactions with the detector material, the active layer detects the signal. The shower development and properties are of vital importance for the particle identification, as it will be shown later. Two important quantities in connection with the calorimeters are the radiation length, X_0 , and the interaction length λ . The radiation length refers to the distance after which an particle (electrons for example) energy has been reduced to $1/e$ of its initial energy. The interaction length describes the mean free path before the occurrence of an hadronic interaction.

The design resolution of the system on the calorimetric energy is given by

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E}} \oplus b \oplus \frac{c}{E} \quad (2.6)$$

where \oplus means that the terms are summed in quadrature. The stochastic term $\frac{a}{\sqrt{E}}$ is related with the fluctuations on the shower developments, the constant term b takes into account the inhomogeneities of the detector, and the last term is associated with the electronic noise and is proportional to $\frac{1}{E}$. The value of the coefficients a and b depend on the incident objects. For the electrons' case in the ECAL, $a \sim 10\% \text{ GeV}^{1/2}$ and $b \sim 0.7\%$, while those for charged pions in the center of the detector are $a \sim 50\% \text{ GeV}^{1/2}$ and $b \sim 5\%$ [22].

Electromagnetic Calorimeter - ECAL

The ECAL specializes on the detection of electrons, positrons and photons, which deposit their energy in relatively dense showers: energetic electrons that radiate Bremsstrahlung photons,

while energetic photons convert to electron-positron pairs when traversing the dense material. The absorber is made of lead (Pb) with stainless steel sheets, while Liquid Argon ([LAr](#)) is used as the active material with copper and kapton electrodes for readout.

The calorimeter has an accordion geometry which provides complete ϕ symmetry without azimuthal cracks. It is divided into two half barrels covering the central detector region ($|\eta| < 1.475$), with a small (4 mm) gap at $z = 0$ and one end-cap on each side of the beamline ($1.375 < |\eta| < 3.2$). The transition region between the barrel and end-cap is referred as the *crack* region, and the majority of physics analysis using the [ECAL](#) require that the photons and electrons are outside of it. Additionally, the [LAr](#) technology is used for the hadronic calorimeters end-caps as well as a Forward Calorimeter ([FCAL](#)) ($3.1 < \eta < 4.9$).

The thickness of the [ECAL](#) is over $22 X_0$ in the barrel region, while over $24 X_0$ in the end-cap region. For photons, the distance at which the energy dropped to $1/e$ is $9/7 X_0$, therefore all the photon's electromagnetic energy is deposited in the [ECAL](#), and only a small part reaches the [HCAL](#).

The mode of measurement is as follows. The incident particles interact with the absorbent medium (Pb), initiating a shower of charged and neutral particles. The charged particles ionize the [LAr](#) medium and the electrodes, with the help of an applied magnetic field, collect the electrons produced in the ionization process. The total signal of the active medium is then proportional to the total real energy of the incident particle.

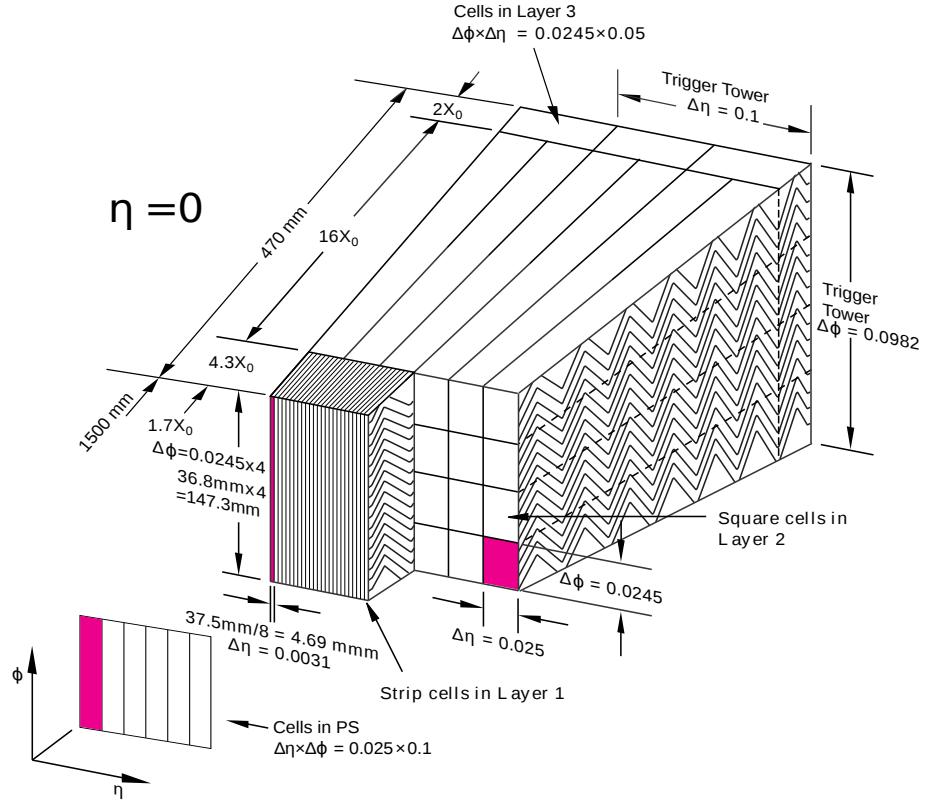


Figure 2.6: Segment of the [ECAL](#) showing the layer arrangement and cell dimensions in each layer [7].

Within the region accepted for precision measurements ($|\eta| < 2.5$ excluding the crack), the

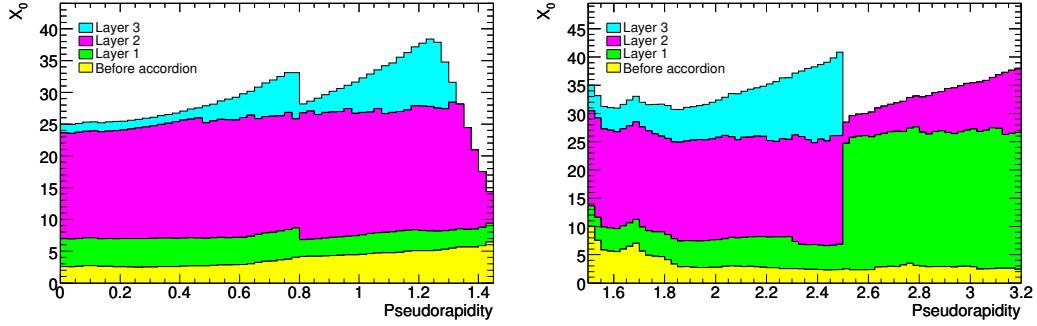


Figure 2.7: Radiation lengths as a function of $|\eta|$ for the [ECAL](#), separated for each sublayer [7].

[ECAL](#) is segmented in three longitudinal layers, displayed in Figure 2.6. The first layer consists on fine-granularity bands (also called the strip layer) which helps with the discrimination between isolated photons and pairs of photons spacialy closed originating from $\pi^0 \rightarrow \gamma\gamma$ decays. This layer has a constant thickness of $\sim 6X_0$ as a function of η (see Figure 2.7), and provides a precise measurement of this variable. For high energy photons and electrons, the majority of their energy is collected in the second layer, which has a lateral granularity of 0.025×0.025 in (η, ϕ) and a thickness of $\sim 24X_0$. The third layer collects the energy deposited by the tails of the electromagnetic shower, with thickness that varies between 2 and $12 X_0$. There is also a presampler (not shown in figures), that covers the region $|\eta| < 1.8$ that improves the energy measurement for particles that start showering before entering the calorimeter.

Hadronic Calorimeter - HCAL

Three hadronic calorimeter layers surround the [ECAL](#) and provide additional discrimination for electrons and photons when measuring the hadronic energy. The [HCAL](#) extends in pseudorapidity up to $|\eta| < 4.9$, allowing virtually the entirety of the solid angle to be covered from the interaction point. In the barrel region ($|\eta| < 1.7$), the tile calorimeter, a sampling calorimeter using steel as absorbing material and plastic scintillator tiles as active material [23], is located. It is divided into two parts ($|\eta| < 1.0$ and $0.8 < |\eta| < 1.7$). The scintillators, arranged in a periodic array, are connected to an optical fiber that carries the light produced by the passing particles to a photomultiplier tube. This array extends, in R , from 2.28 to 4.25 m. In the endcap region ($1.5 < |\eta| < 3.2$) there is an hadronic sampling calorimeter, the Hadronic End-Cap Calorimeter ([HEC](#)), with copper plates as absorber and liquid argon as active material. Each side of the endcap consists of two wheels, one behind the other with the flat Cu plates arranged perpendicular to the beam axis, with a radius of 2.3 m. Finally there is the [FCAL](#), a sampling calorimeter that extends the coverage of the system to $|\eta| < 4.9$, coaxial to the beam axis and located 4.7 m on either side of the interaction point. The main material of the modules is [LAr](#) (with copper or tungsten), and while not used for precision measurements, it provides information for computation of the missing transverse energy and reconstruction of jets in regions very close to the beam axis.

The [HCAL](#) has a thickness greater than 7.7λ in the barrel region (9.7λ in total if the [ECAL](#) is

counted). Analogous to the radiation length mentioned for the [ECAL](#), a hadronic interaction length is defined as the average distance over which the energy of a hadron is reduced to $1/e$ of its initial energy. Thus, all the energy with which the hadrons arrive at the [HCAL](#) is deposited there.

2.2.4 Muon spectrometer

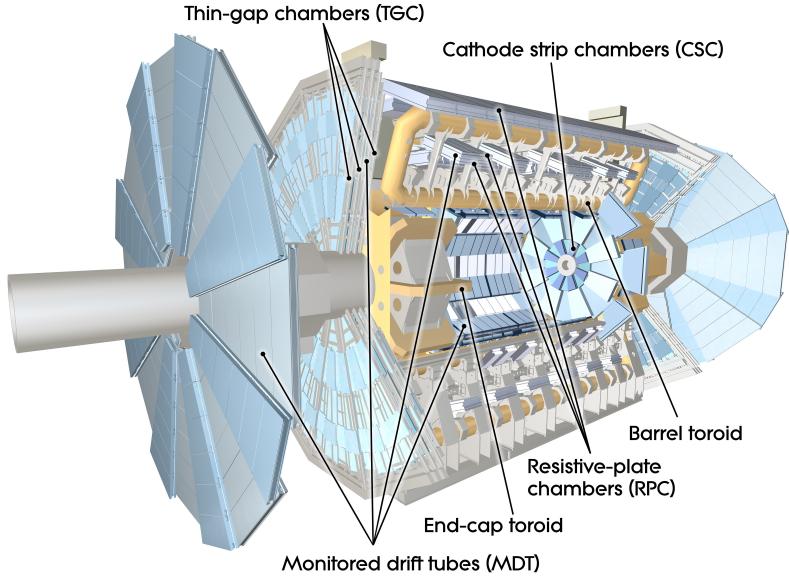


Figure 2.8: [ATLAS](#) Muon Spectrometer ([MS](#)) [7].

The high p_T muons generated at the interaction point have very high penetrating power and are poorly interacting. Therefore, the [MS](#) [24] is located in the outermost part of the [ATLAS](#) detector, embedded within the 4 T magnetic field generated by the barrel and endcap toroid magnets, and is designed to obtain high precision position and momentum measurements of high p_T muons. This is the largest subdetector and the one that gives [ATLAS](#) its size.

The [MS](#) is designed to precisely measure muons within $|\eta| < 2.7$ and to provide muon trigger information up to $|\eta| < 2.4$, shown in Figure 2.8, highlighting the different subsystems

The [MS](#) is composed of different types of muon detection chambers (see Figure 2.8). Monitored Drift Tubes ([MDTs](#)) are responsible for most of the precision measurements and cover the range of $|\eta| < 2.7$. They operate similarly to the [TRT](#), with tubes filled with an ionising gas and a central anode collecting the electrons produced, and the drift time is associated with the distance to the track. In the endcap region there are Cathode Strip Chamber ([CSC](#)) that have high spatiotemporal resolution and a coverage of $|\eta| > 2.0$. These chambers work by measuring the charge deposited on an anode as a result of the cascade of electrons created near the anode. Resistive-Plate Chambers ([RPCs](#)) provide a fast estimation of the muon momentum at trigger-level with a coverage of $|\eta| < 1.05$ ¹. [RPCs](#) measure the discharge between two parallel

¹ The innermost End-Cap layer has been replaced with the [ATLAS](#) New Small Wheel ([NSW](#)) after Run 2 [25]. It features MicroMegas as precision trackers as they provide better performance at the high rates expected in future LHC operations.

resistive plates subjected to a high potential difference, following the ionisation of the internal gas volume caused by the passage of energetic muons. Finally, in the endcap region, there are Thin-Gap Chambers (**TGCs**), similar in function to **CSCs**. They also provide information to the trigger system in this region and have a coverage of $|\eta| < 2.4$.

If hits in the **ID** and the **MS** can be associated with a single muon, a very good momentum resolution of up to

$$\frac{\sigma(p_T)}{p_T} = 0.02\% \cdot p_T [\text{GeV}] \oplus 2\% \quad (2.7)$$

is achieved. The momentum resolution degrades accordingly if a track is identified in only one of the two systems.

2.2.5 The Trigger System

The **ATLAS** trigger system [26–28] uses information from the detector to reject events that do not possess interesting physics (physics already known for example), reducing the event frequency from 40 MHz (bunch-crossing frequency mentioned in Section 2.1) to around 1.5 kHz. It is necessary to emphasize here the central role of the trigger system for the proper functioning of the whole experiment, being responsible for deciding which events are saved and, ultimately, which physics will be encountered (or not) during the event analysis. Without an efficient trigger system, all the subdetectors described above would be wasted. To achieve such a reduction in event frequency and, at the same time, have a high efficiency in selecting those of interest, the trigger system is composed of two consecutive levels capable of performing increasingly complex particle identification; a first hardware-based trigger level, Level-1 (**L1**), and then a high-level software-based trigger, the High Level Trigger (**HLT**). Each level allows events to be analyzed in greater detail, increasing the accuracy of the selection criteria and the complexity of the algorithms used.

Level-1 trigger

The trigger decision starts with the hardware-based **L1** trigger [29], which identifies Region of Interests (**ROIs**). These **ROI** consist of neighbouring cells in the **ECAL** and **HCAL** and are defined from the position in the calorimeter of each object found in a potentially interesting event, which extends as a cone from the interaction point along the detector. Regarding muons, it takes the information read by the **MS**, more specifically the **TGC** and **RPC** and allows to obtain a fast estimate of the muon p_T . The **L1** trigger also has a component that allows for topological requirements such as invariant mass selections and distance measures to be taken into account in the **L1** decision, referred as the Level-1 Topological (**L1Topo**).

The design of the **L1** allows to have an acceptability in the range of $|\eta| < 2.5$ for electrons, photons, muons and taus, up to $|\eta| < 3.2$ for jets, and $|\eta| < 4.9$ for the missing transverse momentum calculation. Using the **ROIs**, the **L1** trigger must make the decision to keep or discard the event, reducing the event rate from 40 MHz to less than 100 kHz in approximately 2.5 μ s,

time determined in part by the limited size of the memory buffers and in part by the time it takes for the muons produced in the event to reach the [MS](#). This final decision is done by the Central Trigger Processor ([CTP](#)), and then passes the [ROIs](#) to the next trigger level: the [HLT](#).

The High Level Trigger

When an event is accepted by the [L1](#), the [HLT](#) [30] executes a sequence of algorithms starting from the [ROIs](#) defined by the [L1](#), and allows to reduce the event rate that is stored at 1.5 kHz in 0.2 s. The reconstruction and identification of candidate particles in the [HLT](#) is evaluated in a sequence of steps where different algorithms are applied. If the selection fails in a certain step, the following steps are no longer executed to save execution time. In [HLT](#), the algorithms are grouped into sets of fast reconstruction algorithms executed first, and then a set of precision reconstruction algorithms similar to those used offline are executed, thanks to the latency time available. The fast reconstruction algorithms use the calorimeter and track information from the [ID](#) only within the [ROI](#) to perform candidate selection and identification, and perform background rejection as quickly and early as possible. If the candidate particle passes the criteria defined by the fast reconstruction selection, precision selection algorithms are run. These have access to detector information outside the RoI, at the highest granularity and including details on calorimeter energy calibration, sub-detector alignment and magnetic field mapping.

The exact sequence and type of algorithms considered at the [HLT](#) are defined in the trigger *menu*. This comprises a database of triggers, each of one defining a sequence of algorithms and requirements on these algorithms for an event to pass the [HLT](#). The trigger requirements are designed and budgeted in a way that the overall [HLT](#) rate does not exceed 1 kHz. In some cases, even the reduction in event rate achieved through the [HLT](#) algorithms for desired trigger requirements, such as low momentum triggers, is too high. To keep the overall [HLT](#) rate below 1 kHz in these cases, triggers can still be included in the menu, but with a prescale. A prescale is an artificial scaling of the trigger, only accepting every Nth trigger decision if the prescale factor is N. This allows triggers with an otherwise high rate to still collect events.

The [HLT](#) algorithms run on approximately 40 thousand CPU cores. In addition, partial event construction is used for trigger-level analysis, detector monitoring, and detector subsystem calibrations. Finally, the accepted events by the [HLT](#) are saved to a disk and distributed, available *offline* for any study or analysis.

2.3 Data-taking during Run-2

The operation of the [LHC](#) is organized into distinct periods known as data-taking Runs. Each run typically spans several years and is characterized by specific experimental conditions, including the energy at which the protons are collided and the intensity of the beams. Since its commissioning, the [LHC](#) has undergone multiple data-taking runs: Run-1 (2010-2013) oper-

ated at collision energies up to 8 TeV, Run-2 (2015-2018) at 13 TeV, and Run-3 (2022-present) at 13.6 TeV. Each data taking period, once the LHC announces stable beams, is divided into Luminosity-blocks (LBs) of approximately two minutes. At each LB, the instantaneous luminosity is practically constant and the beam conditions are stable. Due to the high complexity of the LHC and the ATLAS detector, it is expected to have inefficiencies in the detectors and sub-detectors and/or in the data acquisition chain. During each Run, each part of ATLAS is monitored and any failure or problem is registered, including inactive components, or problems on the LHC beam.

In order to guarantee the high-quality data, free from significant defects, the LBs and ranges within them that pass all the quality criteria are compiled into Good Runs List (GRL). The lists are produced and distributed in a centralized manner, in order to provide any ATLAS group with the same collection of LBs. Since during the runs different parts of the detector are available (in an optimal run, all of the subdetectors are available), there are multiple GRLs available to use. Each analysis, then, selects which GRL to use depending their tolerance to the subdetectors' faults.

The present thesis uses ATLAS data recollected from pp collisions during the Run-2 (2015-2018), at a centre of mass energy of $\sqrt{s} = 13$ TeV. During this run, the LHC delivered a total of 156 fb^{-1} , from which ATLAS collected 147 fb^{-1} . The total integrated luminosity available for Physics analysis is 140.07 fb^{-1} ², as seen from Figure 2.9a. The uncertainty in the combined integrated luminosity for Run-2 is 0.83% [31], obtained using the LUCID-2 detector [32]. Combining the 2022, 2023 and 2024 years of data taking for Run-3, 159 fb^{-1} of data was recollected, shown in Figure 2.9b [33, 34].

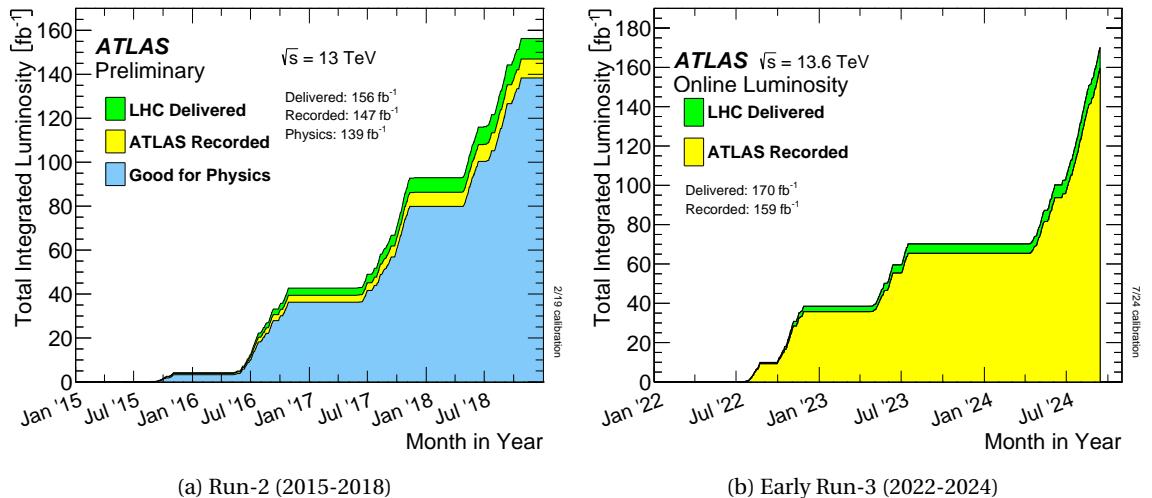


Figure 2.9: Luminosity delivered by the LHC and recorded by ATLAS during the Run-2 [31] and Run-3 data taking periods. For Run-2, the fraction of data good for physics analyses is also displayed.

Another important concept in ATLAS data acquisition is pileup, which occurs when particles produced in more than one pp collision arrive at the detector at the same time, or more generally, when signals overlap in a way that cannot be separated. When bunches of protons collide,

² First measurements and initial GRLs led to a total of 139 fb^{-1}

the probability of an interaction is proportional to the particle density, or better, to the particle flux, which is expressed by the instantaneous luminosity. The actual number of particle collisions that take place when two bunches intersect is a random variable that follows a Poisson distribution. For low luminosities, in most beam crossings, no collisions occur, but for high instantaneous luminosities, in most crossings many particle collisions occur at the same time. Depending on the subdetector and the type of measurement, it may or may not be possible to distinguish between particles coming from different simultaneous interactions. This is called in-time pile-up. In contrast, out-of-time pile-up includes the effects that arise when the time the detector needs to return to its waiting state is longer than the time between bunches crossing. A quantitative measure of pile-up and event activity is the mean value of pp inelastic interactions per bunches crossing, $\langle\mu\rangle$.

The maximum instantaneous luminosities increased by a factor of four over the four years of Run-2, resulting in an increase of $\langle\mu\rangle$ from 10 up to 60, as shown in Figure 2.10a. For Run-3, pileup was drastically increase up to values of 57 for year 2024, in average increasing up to 52 interactions per bunch crossing, displayed in Figure 2.10b.

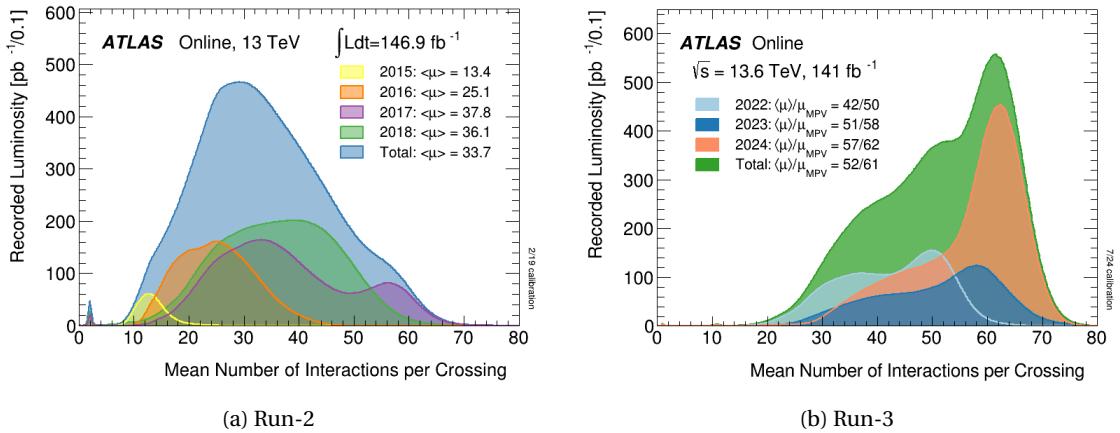


Figure 2.10: Pileup conditions during Run-2 and Run-3.

RECONSTRUCTION AND IDENTIFICATION OF PHYSICAL OBJECTS

3

“Champions keep playing until they get it right.”

Billie Jean King

The particles (and products of their decays) produced at every collision, interact with the detector in a particular manner according to their nature. The information recollected by all the sub-detectors described in the previous chapter allow for the reconstruction and identification of the physical objects present in each accepted event by the trigger system. Two types of reconstruction and identification exist. The *online* one, is carried out at the same time the pp collisions take place, and the *offline* one, done after the events are saved to storage. The reconstruction is done event by event, and is carried in the same way for events recorded by the [ATLAS](#) detector and for simulated [MC](#) events. In the following, a brief overview of the offline reconstruction and identification of the objects used in this thesis is given.

[update chapter, see version in spanish](#)

3.1 Track and vertex reconstruction

In a high-pile-up event, there can be of the order of 1000 charged particles passing through the [ATLAS](#) detector. The information from the [ID](#) (Section 2.2.2) is used to reconstruct the trajectories of charged particles, called *tracks*.

Tracking charged particles is a critical step in reconstruction. Tracks encode charged particles' momentum and trajectory, playing an essential role in particle identification and primary vertex reconstruction. As the inner detector is closest to the beamline and comprises minimally ionizing detector material with high granularity, it plays the main role in track reconstruction. A charged particle passing through different layers of [ID](#) leaves a signal via ionization. As the [ID](#) solenoidal field is homogenous, the resulting trajectory is circular in the xy plane. Five parameters shown in Figure 3.1 define charged particle tracks:

- q/p_T : the ratio of the charge and transverse momentum defining the curvature

- d_0 : the distance of the closest approach to the primary vertex in xy -plane defining the transverse impact parameter
- z_0 : the longitudinal impact parameter along the z -axis
- ϕ_0 : the azimuthal angle
- θ_0 : the polar angle of the particle direction at the closest point of approach [35].

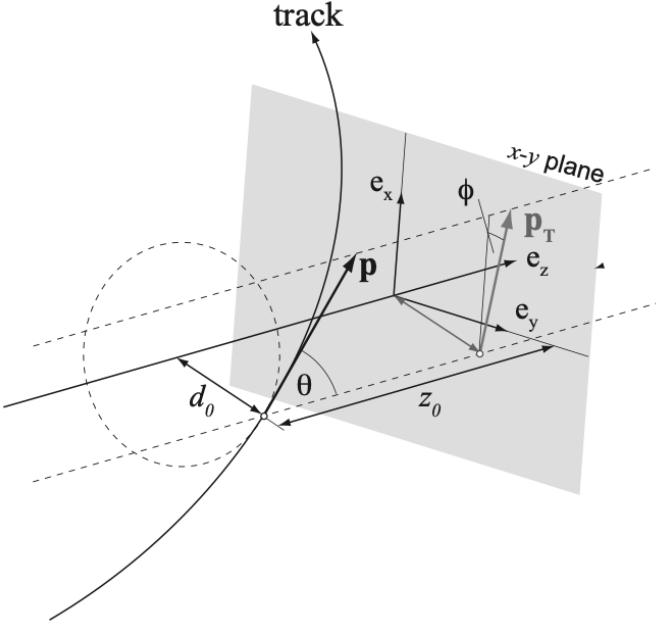


Figure 3.1: Schematic showing the tracking parametrization [36].

The track reconstruction used in Run-2 uses two complementary approaches: the *inside-out* approach, and the *outside-in* [37].

The first step in the inside-out track reconstruction is the seed-finding, where three hits in the silicon detector are searched for to seed the track reconstruction. Using these three hits and assuming an uniform magnetic field, a first estimate of the track parameters is obtained. Using the track seeds, the track is extrapolated to the other silicon layers, from which a combinatorial Kalman filter is used to estimate the track parameters. At this stage of the process there can be several track candidates for each track seed. Once the track is formed, an ambiguity resolution algorithm is applied to reassign shared clusters to the track with a better match [38], and the final track candidate is fitted using a global χ^2 method. The last part of the inside-out method consists on extending the tracks to the TRT, and including the TRT hits to the track, to improve the track's momentum resolution.

To improve the efficiency for tracks from decays displaced from the original collision point, an outside-in track reconstruction algorithm is also used. The track is seeded with hits from the TRT. The track is extended to include hits from the silicon detector, with an ambiguity solver again applied to mitigate the hit sharing between two tracks.

Primary and secondary vertices are of vital importance for the subsequent object reconstruction in **ATLAS**. In this step, the tracks found as explained previously are used as input to the vertexing algorithm [39, 40]. First of all, the Primary Vertex (**PV**) is defined as the location where two protons collide. **PVs** are reconstructed by matching up intersecting tracks, which proceeds in three main steps: seeding, track assignment, and fitting. The vertex with the largest $\sum p_T^2$ for all associated tracks is labeled as the hard-scatter vertex. There are some particles that decay rapidly after their production, such as τ leptons or heavier quarks (b or c -quarks), and their decay position can be measured. From the remaining tracks originated from these decays, it is possible to identify secondary vertices. All the remaining reconstructed vertices are considered to be pile-up.

3.2 Photons and electrons

The reconstruction of electrons and photons in **ATLAS** is based on the energy deposition in the **ECAL**. Since electrons and photons leave similar signals in the **ECAL**, their reconstruction is done simultaneously, distinguishing between them by the reconstructed track information left in the **ID**.

3.2.1 Reconstruction

The *offline* photon and electron reconstruction [41, 42] makes use of dynamic, variable-size clusters, connected topologically between the **ECAL** and **HCAL** cells [43], called topo-clusters. This approach allows for the clusters to recover energy from bremsstrahlung photons or from electrons from photon conversions. With this approach, there are three types of objects:

- Electrons: consists of a cluster built from the energy deposits in the **ECAL** and a matched track.
- Converted photons: consists of a cluster matched to a conversion vertex (or vertices)
- Unconverted photons: cluster matched to neither an electron track nor a conversion vertex.

The algorithm for the reconstruction of electrons and photons proceeds as shown in Figure 3.2. The reconstruction process begins with the topo-cluster formation. First, proto-clusters are formed in the **ECAL** and **HCAL** by grouping cells that have a required energy, and by subsequently adding neighbouring cells in four consecutive steps, obtaining the topo-cluster. Reconstructions starts only in those cases where the topo-clusters energy in the **ECAL** is greater than 400 MeV.

The algorithm also builds conversion vertices out of the refitted tracks and matches them to the selected topo-clusters. After the initial track-cluster matching and conversion building, the

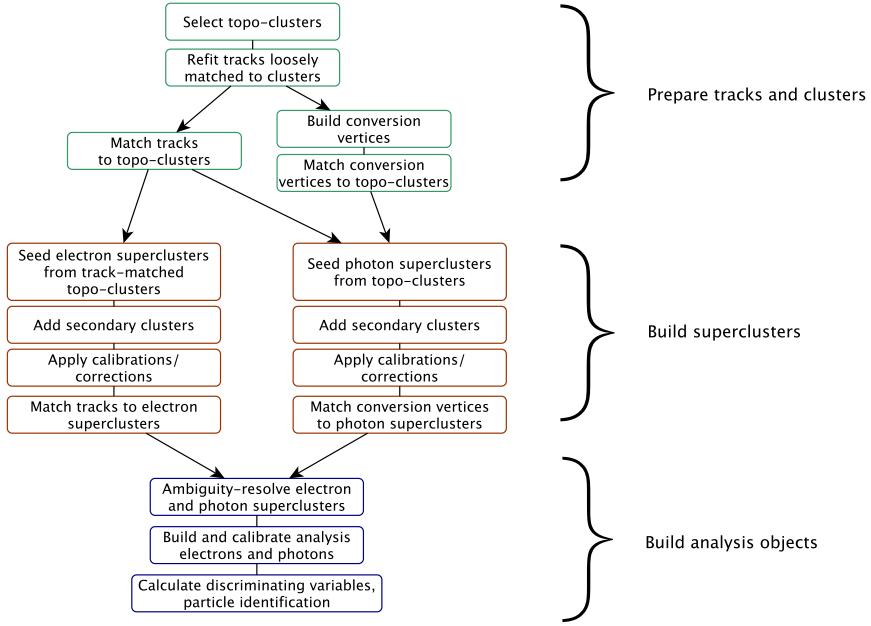


Figure 3.2: Diagram showing the reconstruction algorithm workflow for electrons and photons, extracted from Ref. [41]

electron and photon supercluster algorithms run separately in parallel. In the first stage, topo-clusters are evaluated for use as seed cluster candidates, which form the basis of superclusters; in the second stage, clusters near the seed candidates are identified as satellite cluster candidates, which may emerge from bremsstrahlung radiation or topo-cluster splitting. Satellite clusters are added to the seed candidates to form the final superclusters, if they pass the necessary selection criteria. After applying initial position corrections to the resultant superclusters, the reconstruction algorithm matches tracks to the electron superclusters and conversion vertices to the photon superclusters.

Since one object may be reconstructed as both an electron and a photon, an ambiguity resolution is performed to remove part of the overlap. However, some overlap is allowed in order to maintain a high reconstruction efficiency for electrons and photons, to which physics analyses may apply their own criteria. The final electrons and photons are then built and calibrated, facilitating the calculation of additional variables used for quality cuts and ambiguity resolution

3.2.2 Identification

In order to distinguish real photons (those coming from the collision) from background photons which have much larger production cross sections (coming from hadrons decays, also called fake photons), it is necessary to rely on a algorithm of identification with high signal efficiency and background rejection, for photon candidates with $p_T \sim 10$ GeV up to the TeV scale. Currently, photon identification in ATLAS is based on a set of rectangular cuts on Shower Shape Variables (SSVs) computed from the energy deposited in the cells of the cluster in the first and second layer of the **ECAL**, and from the leakage to the **HCAL**. These variables describe the pas-

sage of the photons through the calorimeters, characterizing the lateral and longitudinal electromagnetic showers. The full photon identification process is presented in Chapter 4, where the **SSVs** are explained one by one. Also, Chapters ?? and ?? present two approaches to correct the differences seen in these **SSVs** between data and **MC**.

Shower shape variable corrections

Due to the imperfection of the **ATLAS** simulation to model the **SSVs**, and that these variables are used as input to the identification step, it is important that they are corrected. Historically, the corrections are called **FFs**, and they comprised modifications to the mean value of the **SSVs**, calculated by minimizing the χ^2 value between the data and **MC SSVs**. More details on the **SSV** corrections are given in GIVE SECTION!!!

Additional corrections to all the reconstructed and identified photons in the simulation are applied event-by-event in the form of scale factors. These values, provided centrally by the EGAMMA group to **ATLAS**, represent the residual differences on the efficiencies between actual data and **MC**, computed as a function of the photon p_T and pseudo-rapidity η and separately for converted and unconverted photons.

3.2.3 Isolation

To further reduce the backgrounds of jets misidentified as photons and of hadron decay within the jets (such as the case of neutral pions), two isolation variables are defined: $E_T^{\text{cone}40}$ and $p_T^{\text{cone}20}$.

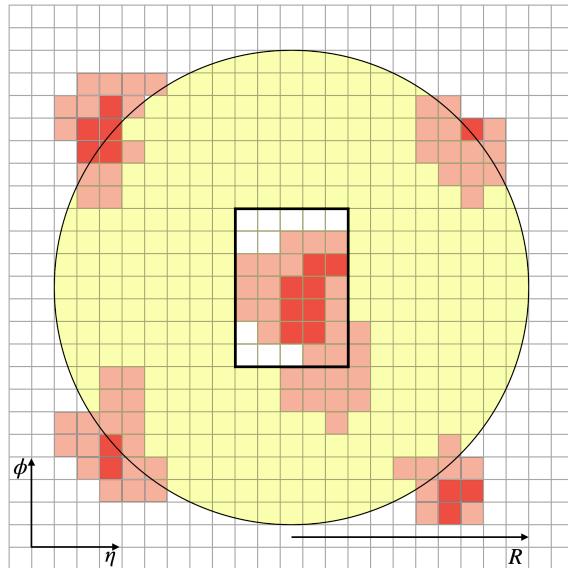


Figure 3.3: Diagram showing the calculation of the calorimetric isolation variable. When $R = 0.4$, $E_T^{\text{cone}40}$ is computed.

The procedure to compute the isolation energy $E_T^{\text{cone}40}$ is as follows, and showed in Figure 3.3. First, a cone of radius $\Delta R < 0.4$ is built around the photon or electron candidate, and the energies of all the cells in the topo-clusters (introduced in Section 3.2.1) whose bary-centers are

located inside the cone, are added together. Then, to this computed energy, the energy of all the cells in a 5×7 window (in units of $\eta \times \phi$ in the second layer of the [ECAL](#)) centered around the candidate are subtracted, in order to remove the energy of the candidate itself. Pile-up contributions and energy leakages outside the cone are also taken into account.

The track isolation variable $p_T^{\text{cone}20}$ is obtained by adding the p_T of the good-quality tracks in a cone of radius $\Delta R < 0.2$ around the electron candidate or in the direction of the converted photon cluster. The track associated to the track or to the converted photon are excluded from this computation, as well as those tracks which do not pass the *good-quality* track requirement. A *good-quality* track is defined as one in which the p_T is $p_T > 1$ GeV, and it has a minimum distance to the primary vertex along the z -axis of $|z_0 \sin\theta| < 3$ mm.

In general, for photons and electrons, there is no other energy deposited in the cone around the candidate, apart from the low-energy objects originating from the remnants of the collision, multiple interactions and pile-up. On the other hand, for fake photon candidates and non-direct photons, additional energy is observed within the cone, originating from objects accompanying the jet.

Table 3.1: Summary of electron and photon isolation [WPs](#) use throughout this thesis.

Object	WP	Calorimetric Isolation	Track Isolation
Photon	FixedCutLoose	$E_T^{\text{cone}20} < 0.065 \times p_T$	-
	FixedCutTightCaloOnly	$E_T^{\text{cone}40} < 0.022 \times p_T + 2.45$ GeV	-
	FixedCutTight	$E_T^{\text{cone}40} < 0.022 \times p_T + 2.45$ GeV	$p_T^{\text{cone}20} / p_T < 0.05$
Electron	Loose_VarRad	$E_T^{\text{cone}20} < 0.2 \times p_T$	$p_T^{\text{cone}30} / p_T < 0.15$
	HighPtCaloOnly	$E_T^{\text{cone}20} < \max(0.015 \times p_T, 3.5$ GeV)	-

From the calorimetric and track isolation different [WPs](#) can be defined separately for both electrons and photons. For electrons, two strategies are defined: either to achieve a fixed efficiency, or to apply fixed cuts on the isolation variables. In the case of photons, there are [WPs](#) which do not use both the isolation variable, as is the case of the [FixedCutTightCaloOnly WP](#), which only uses calorimetric isolation. The definitions of the different [WPs](#) used throughout this thesis is shown in Table 3.1. Also, it is common to define the following variables for the photon [FixedCutTight WP](#):

$$E_T^{\text{iso}} = E_T^{\text{cone}40} - 0.022 \times E_T - 2.45 \text{ GeV} \quad (3.1)$$

$$p_T^{\text{iso}} = p_T^{\text{cone}20} / E_T \quad (3.2)$$

therefore leaving the [FixedCutTight WP](#) defined as

$$E_T^{\text{iso}} < 0 \text{ GeV} \quad (3.3)$$

$$p_T^{\text{iso}} < 0.05 \quad (3.4)$$

3.3 Muons

The rate of bremsstrahlung radiation is inversely proportional to the square of a particle's mass. Since muons are about 200 times heavier than electrons, they primarily interact with the de-

tector material through ionization. Therefore, muons are minimally ionizing particles that do not create electromagnetic shower in the calorimeters and pass through all layers of the [ATLAS](#) detector. Hence, muon detection relies on track measurements from the [ID](#) and [MS](#). The combination of the two subdetectors define four types of muons, depending on the used information for the reconstruction:

- Combined Muons ([CBs](#)): muons reconstructed from a global refit of [ID](#) and [MS](#) tracks
- Segmented Muons ([STs](#)): muons reconstructed from a fitted [ID](#) track and [MS](#) segment track
- Calorimetric Muons ([CTs](#)): muons reconstructed using [ID](#) track matched to the minimum ionizing energy deposits in the calorimeters
- Standalone/Extrapolated Muons ([MEs](#)): muons reconstructed solely from [MS](#) tracks.

The overlap between different types of muons is resolved as follows. When two muon types share the same [ID](#) track, the order of preference is: first [CB](#), then [ST](#) and finally [CTs](#). The overlap with [MEs](#) is solved by analyzing the hits of the tracks, selecting those tracks with the best fit and the highest number of hits.

For the muon identification, quality cuts are applied to distinguish isolated muons from those coming from background processes, mainly from pion and kaon decay. The variables with good discriminating power used are described in Ref. [44]. Four identification selections are defined: Loose, Medium, Tight, and High- p_T . The first three categories are inclusive, and Medium being the default selection in [ATLAS](#). Finally, the muon candidates to be used by the analyses are asked to satisfy the isolation requirements, both at track and calorimetric levels, analogously to what was detailed for photons in the previous section. For the first case, a variable similar to that used for photons is used, but with a variable-radius cone $\Delta R = \min(10 \text{ GeV}/p_T, 0.3)$ around the muon p_T , excluding the muon track. For calorimetric isolation the same variable $E_T^{\text{cone}40}$ is used, with the difference of using a radius of $R = 0.2$, instead of 0.4 as before. Based on these variables, 7 isolation selection criteria (7 [WPs](#)), optimized for different analyses, are defined.

3.4 Jets

Due to color confinement in Quantum Chromodynamics ([QCD](#)), a quark or gluon cannot exist on its own and goes through hadronization to form a collimated color-neutral stream of particles, *jets*. Generally, jets penetrate through the [ECAL](#) and get fully absorbed by the material in the hadronic calorimeter. In the following, a brief description of the typical clustering method adopted by [ATLAS](#) is given. Also, the two existing types jet reconstruction are described.

3.4.1 Anti- k_t jet clustering algorithm

Given that jets are constituted by a high number of particles that leave energy depositions in the **ECAL** and **HCAL** and tracks in the **ID**, a clustering algorithm groups together constituents in the event to define the jets. Said algorithm is called the anti- k_t algorithm [45]. In the same way as for electrons and photons, **ATLAS** jet reconstruction relies on the formation of topo-clusters: grouped energy depositions in the calorimeters cells using a sequential combination algorithm. Then, the anti- k_t algorithm combines the topo-clusters with the following steps:

- Measure the distance between all topo-clusters between themselves, and of each topo-cluster with the beam:

$$d_{ij} = \min(p_{T,i}^{-2}, p_{T,j}^{-2}) \frac{\Delta_{i,j}^2}{R^2} \quad (3.5)$$

$$d_{iB} = p_{T,i}^{-2} \quad (3.6)$$

where $\Delta_{i,j}^2 = \Delta\phi_{i,j}^2 + \Delta\eta_{i,j}^2$ and R is the jet-radius.

- If the minimum of all the distances computed previously is d_{iB} , the topo-cluster i is classified as a jet, and is discarded in successive iterations.
- If the minimum of all the distances is d_{ij} , topo-clusters i and j are combined, all the distances are computed again with this new topo-cluster and the iteration is carried all over again.

This process is repeated until all the particles in the event have been clustered.

The anti- k_t algorithm starts by clustering the radiation around the hardest particle in the event since the leading p_T particle will define the $\min\left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2}\right)$ term in the d_{ij} definition. This allows jets in the event to have a stable direction early on the combination process. The anti- k_t algorithm is preferred to other sequential jet algorithms since jets have regular boundaries which are approximately conical, shown in Figure 3.4. Jets originating from quarks or gluons in general are called small- R jets and a radius of $R = 0.4$ is used for their reconstruction. On the other hand, jets representing massive particles which decay hadronically are called large- R jets, and use $R = 1.0$. The usage of a wider cone helps to include the majority of the particles product of the decay.

3.4.2 Calorimeter Jets

One way to reconstruct jets is based on energy deposits in the calorimeter. In a similar way to what has been explained for electrons and photons in Section 3.2.1, energy depositions on the cells of the **ECAL** and **HCAL** are used to build topo-clusters, which approximates the energy deposits of individual hadrons [42, 43]. Jets reconstructed in this manner and clustered with the anti- k_t algorithm with a radius of $R = 0.4$ are referred as EMTOPO jets, and are the proxies for the individual quarks and gluons. In the jet reconstruction, only the topo-clusters with positive net energy are included.

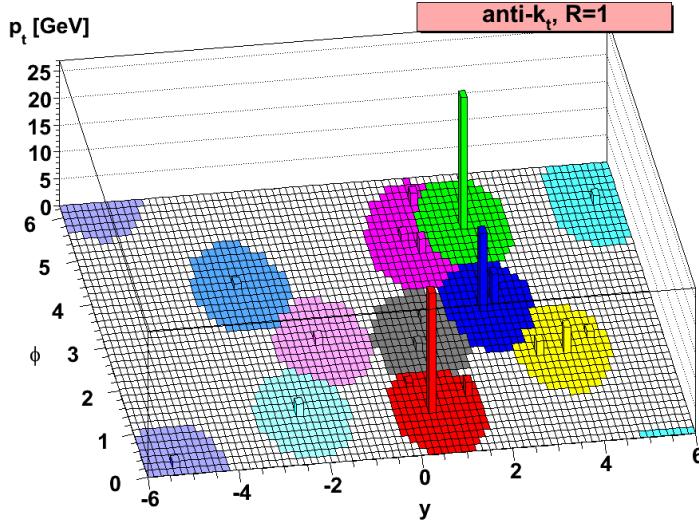


Figure 3.4: Schematic representation of the anti- k_t algorithm for jet clustering [45].

3.4.3 PFlow Jets

Another approach for jet reconstruction is taken in the particle flow algorithm, in which measurements from both the tracker and the calorimeter are combined to form the signals, which ideally represent individual particles. In this algorithm, tracks are matched to topo-clusters using proximity in (η, ϕ) space, also accounting for the size of the topo-cluster. The tracks are only “matched” if the cluster carries more than 10% of the track’s momentum. Sometimes the topo-cluster fails to cluster all of the hadron’s energy in a single topo-cluster. In cases where the expected energy of the track is less than the expected track’s energy, a “split shower recovery” combines nearby topo-clusters to form a topo-cluster set. From this topo-cluster set, the expected energy of the track is subtracted from the topo-cluster’s cells, starting with high-energy density cells. If the residual energy is consistent with the resolution of the expected track energy, the residual energy is also subtracted in the last step called “remnant removal”.

The result of this algorithm is a set of tracks, modified and unmodified topo-clusters which are the Particle Flow (**PFlow**) objects. The **PFlow** objects can also be clustered with the anti- k_t algorithm and the same $R = 0.4$ to form **PFlow** jets.

There quite a lot of benefits of using the **PFlow** algorithm over the EMTOPO one:

- The momentum resolution of the tracker is significantly better than the calorimeter’s energy resolution for low-energy charged particles.
- Allows for a higher acceptance for softer particles. Tracks are reconstructed for charged particles with a minimum p_T of 400 MeV, and oftentimes these particles’ energy deposits do not pass the thresholds to seed topo-clusters.
- Improved angular resolution of a single charged particle as it uses the tracker information instead of the calorimeter’s.

- Low- p_T charged particles originating within a hadronic jet are swept out of the jet cone by the magnetic field by the time they reach the calorimeter. By using the tracks azimuthal coordinate at the perigee, these particles are clustered into the jet.
- It is possible to remove those tracks originating from pile-up, knowing that these do not originate from the PV.

However, particle flow introduces a complication. For any particle whose track measurement ought to be used, it is necessary to correctly identify and subtract its signal in the calorimeter to avoid double-counting. In the particle flow algorithm, a boolean decision is made as to whether to use the tracker or calorimeter measurement. The ability to accurately subtract all of a single particle energy, without removing energy deposited by other particles, forms the key performance criterion upon which the algorithm is optimised.

In this thesis, PFlow jets are considered, as they have proven to provide better jet reconstruction [46], principally for those with low p_T and in the E_T^{miss} reconstruction [47].

3.4.4 Jet calibration

Once the jets are reconstructed, their 4-momentum is corrected to match the kinematics of a truth jet¹, as shown in Figure 3.5. The first three corrections account for contamination from the underlying pile-up distribution and fluctuations due to the origin of the jet [48]. The Global Sequential Calibration improves the jets p_T resolution (and associated uncertainties) by sequentially removing the dependence of the reconstructed jet response ($R = E^{\text{reco}}/E^{\text{truth}}$) on key event observables. Finally, the residual differences between data and MC are accounted for by measuring the momentum imbalance in $Z + \text{jets}$, $\gamma + \text{jet}$ and multi-jet events.

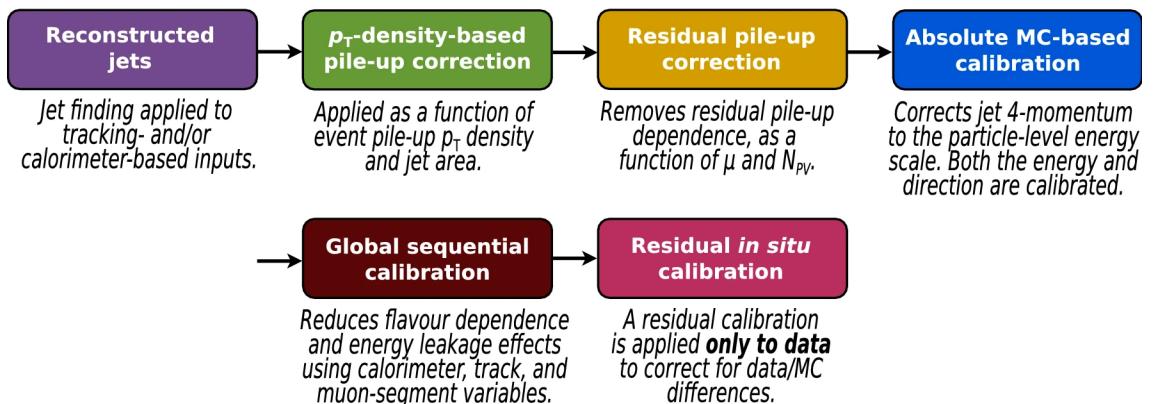


Figure 3.5: PFlow 4-momentum jet calibration steps [48].

To reduce the number of jets with a considerable fraction of energy coming from pile-up, the Jet vertex Tagger (JVT) algorithm is used. This algorithm update to NNJVT reconstructs a multivariate discriminant that combines, among other quantities, the Jet vertex Fraction (JVF) (fraction of the tracks' p_T associated to a jet originating from the PV, and the total number of tracks) and

¹ The truth jets come from the anti- k_t clustering of the stable final state truth particles (hadrons and charged leptons) in simulation.

the number of [PVs](#) in the event N_{PV} . As the jets that do not originate from the hard-scatter interaction are generally softer, the [JVT](#) cut is applied only to jets with $p_{\text{T}} < 60 \text{ GeV}$ and $|\eta| < 2.4$. The default [JVT WP](#) is 96% efficient for hard-scatter jets.

3.5 Jet flavor tagging

Heavy hadrons decays are governed mainly by the heaviest hadron in the decay cascade. A b -hadron generally decays through a cascade to a c -hadron, which in turn decays to an s -hadron, etc, which leads to the existence of multiple vertices.

Flavor Tagging ([FTAG](#)) is the classification of jets containing b -hadrons (b -jets), c -hadrons (c -jets) or neither b - or c -hadrons (light-flavour jets, or l -jets) by using algorithms sensitive to the distinctive properties of the respective classes. These complex algorithms rely on the multiple vertices, on the high mass, high decay multiplicity and characteristic decay modes of the b - and c -hadrons, as well as on the properties of heavy-quark fragmentation.

In [ATLAS](#) a two-step approach is employed to reconstruct key characteristics of heavy-flavour jets. In the first stage, low-level algorithms use complementary methods to extract track information from the charged particles linked to the jet. Some algorithms focus on the properties of individual tracks, while others analyse their correlations or combine them to explicitly reconstruct displaced vertices. In the second stage, the outputs from these algorithms are integrated into a high-level algorithm using multivariate classifiers to optimize performance. Over time, the algorithms have evolved significantly, starting with likelihood-based discriminants and boosted decision trees during [LHC](#) Run-1, and progressing to more advanced methods like recurrent and deep neural networks, resulting in notable improvements on the identification performance [49–52].

Starting in Run-3, a novel Transformer-based "GN2" algorithm is developed by the [FTAG](#) combined performance group in [ATLAS](#). The GN2 algorithm is a single trained model which supersedes DL1d [53] and the low level algorithms that feed it. It is based on GN1 [54], and was quickly refined into GN2. GN2 replaces the Graph Attention Network [55] used by GN1 with a Transformer [56], and also benefits from several other architectural optimisations and from an order of magnitude more training statistics.

GN2 directly accepts information about the jet and associated tracks and as such does not depend on other flavour tagging algorithms. GN2 retains the two auxiliary training objectives that were introduced with GN1: the grouping of tracks originating from a common vertex, and the prediction of the underlying physics process from which each track originated.

This new algorithm is also prepared to provide identification of c -jets and jets originating from τ decays. Outputs of this tagger comprise the probabilities of a jet to be tagged as a b -, c -, τ - and light-flavor jet, labeled as p_b , p_c , p_τ and p_w , respectively.

3.5.1 b -jet identification performance

In order to evaluate the performance of the tagger of identifying b -jets at a constant efficiency, the ability to reject c -, τ - and light-flavor jets is measured. The tagger output probabilities are combined to build a single discriminant GN2_b , defined as

$$\text{GN2}_b = \log\left(\frac{p_b}{f_c p_c + f_\tau p_\tau + (1 - f_c - f_\tau) p_u}\right). \quad (3.7)$$

The parameters $f_{c(\tau)}$ are free and determine the weighting between $p_{c(\tau)}$ and p_u in the discriminant. The specific values of these parameters are determined through an optimisation procedure aimed at maximising the rejection of c -jets (τ -jets) and l -jets and found to be 0.2 (0.01).

From the tagger discriminant score, several **WPs** can be defined, simply by requiring the GN2_b score to be above a certain threshold. The **FTAG** working group provides centrally to the whole **ATLAS** collaboration 5 different **WPs** to achieve a fixed overall b -tagging efficiency: 65, 70, 77, 85 and 90% efficiency, and are shown in Figure 3.6. In said figure, the data and **MC** GN2 tagger distributions are compared, where the different flavour contributions are shown with different colors.

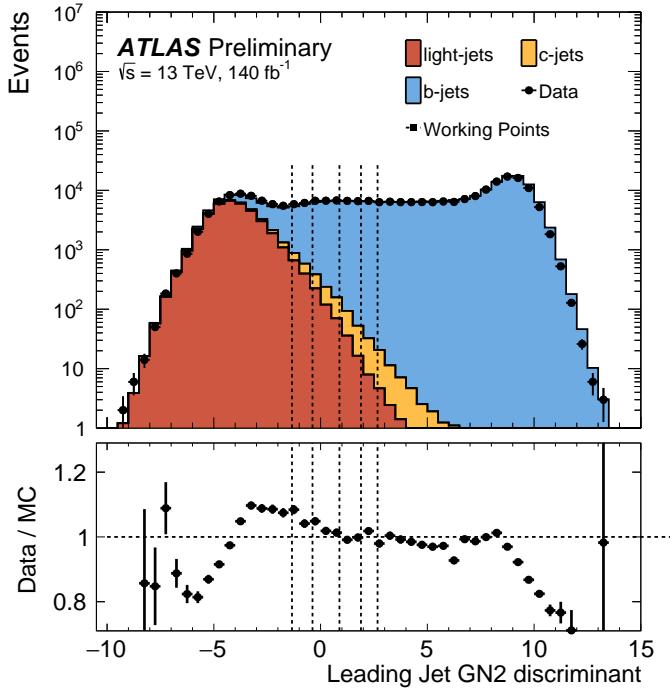


Figure 3.6: GN2 tagger discriminant comparison between data and single-lepton $t\bar{t}$ **MC** simulation. The l -, b - and c -jets are contributions shown with different colors, and the 5 b -tag **WPs** shown with the dashed vertical lines. From left to right, the dashed lines represent the 90, 85, 77, 70 and 65% efficiency **WPs**. The lower pad shows the ratio between data and the stacked **MC** [57].

One key challenge of b -tagging is the decrease in efficiency at higher p_T . In this high- p_T regime, particles become more collimated and they tend to travel further in the **ID** before decaying, potentially leading to a decay track with spurious hits. The degraded efficiency is visualised

in Table 3.2, where tagging efficiencies are shown for b -jets, along with c -jets, l -jets and τ -jets rejections, in the low and high- p_T regimes. The values shown are computed by using different samples, where $t\bar{t}$ is used at low- p_T and Z' decay events² are used in the high- p_T region. It can be seen that the b -tag efficiency drops by 30% for higher p_T jets.

Table 3.2: Measured b -tagging efficiencies and c -jets, l -jets and τ -jets rejections in the low and high- p_T regime.

Sample	p_T range [GeV]	b -efficiency	c -rejection	light-flavor rejection	τ -rejection
$t\bar{t}$	$20 < p_T < 250$	0.76	17.52	448.61	71.15
Z'	$250 < p_T < 6000$	0.41	20.27	179.99	452.94

3.5.2 c -jet identification performance

Similar to b -tagging, a single discriminant can be built from the output probabilities of the tagger in order to identify c -jets against b -jets, τ -jets and l -jets:

$$\text{GN2}_c = \log\left(\frac{p_c}{f_b p_b + f_\tau p_\tau + (1 - f_b - f_\tau) p_u}\right) \quad (3.8)$$

where now the $f_{b(\tau)}$ are the free parameters that control the rejection between b -, τ - and light-flavor jets. Using the same optimisation procedure as for b -tagging, the values for $f_{b(\tau)}$ are found to be 0.3 (0.05).

Thanks to the great b -tagging efficiency achieved by GN2, it is possible to design a c -tagging [WP](#) after applying b -tagging-veto, further separating c -jets from l -jets. By building this simultaneous tagging [WP](#) and assuming the fraction of τ -jets to be negligible, one can separate light-, c - and b -jets in three orthogonal regions. Starting from requiring a jet to *not* pass the 77% b -tagging [WP](#) (b -tag veto), three different c -tagging [WPs](#) are defined by fixing the GN2_c score: 10, 30 and 50% c -tag efficiency. The efficiency and rejection measurements for the both samples described above, after applying the 50% c -tag [WP](#) are shown in Table 3.3.

Table 3.3: Measured c -tagging efficiencies and b -jets, l -jets and τ -jets rejections in the low and high- p_T regime. The values shown correspond to those after applying the b -tagging 77% [WP](#) veto and the 50% c -tagging [WP](#). **rejection values not correct!**

Sample	p_T range [GeV]	c -efficiency	b -rejection	light-flavor rejection	τ -rejection
$t\bar{t}$	$20 < p_T < 250$	0.467	17.52	448.61	71.15
Z'	$250 < p_T < 6000$	0.344	20.27	179.99	452.94

Add plots with GN2 distributions?

² The leptophobic axial-vector Z' model is a simplified Dark-Matter model in which one of the theorised decay products are a pair of quarks.

Part III

Photon shower shape corrections

SHOWER SHAPES AND PHOTON IDENTIFICATION

4

“Champions keep playing until they get it right.”

Billie Jean King

The [ECAL](#) was presented briefly in Section 2.2.3, where the measurement mechanism and all the layers it has were described. In this subdetector, photons deposit their energy via electron-positron pair creation and bremsstrahlung radiation, creating an electromagnetic ([EM](#)) shower. The [ECAL](#) does a great job to compute the energy of the [EM](#) shower, but identifying the initiating particle remains a challenging task. However, by virtue of the different layers and granularities in the [ECAL](#), different characteristics of these [EM](#) showers can be studied, and are encoded by different variables called Shower Shapes ([SSs](#)).

This chapter presents all the [SSs](#) that are used to identify real photons from fakes in Section 4.1. The mentioned variables are heavily used in the process of photon identification, as they provide the separation needed between real and fake photons. For this reason, in Section 4.2, the optimisation of the photon identification (based on the [SSs](#)) is briefly described, as well as how the photon identification efficiencies are measured in data, using three distinct methods. Finally, Section 4.3 describes one key problematic arising in the simulation of the [SSs](#) and how this is handled, for which a more thorough explanation will be given in the Chapter 5.

4.1 Shower shapes

As mentioned in Section 3.2.2, photon identification relies on rectangular cuts applied to [SSs](#) that lead to an excellent separation power between real isolated photons from fake photons originating from hadrons. These [SSVs](#) are computed from the photon candidates' energy deposits in the [ECAL](#) and [HCAL](#) cells, and serve to describe the passage of the photons candidates throughout the calorimeters, characterizing the lateral and longitudinal [EM](#) showers.

In general, real photons produce narrower energy deposits in the [ECAL](#), and have lower leakages to the [HCAL](#), compared to those photons provenient from hadrons, where the presence

of additional neighbouring hadrons close to the fake photon tend to widen the showers. Furthermore, since the first layer of the **ECAL** consists on fine strips, it is possible to discriminate photon candidates coming from $\pi^0 \rightarrow \gamma\gamma$ decays, characterized by two local maxima due to the presence of two nearby photons.

Table 4.1: Discriminative **SSs** used for photon identification. The three columns on the right denote whether the variable is used for the *loose* (L), *medium* (M) or *tight* (T) identification **WP** or not.

Category	Description	Name	L	M	T
Hadronic leakage	Ratio of E_T in the first sampling layer of the HCAL to E_T of the EM cluster (used over the ranges $ \eta < 0.8$ and $ \eta > 1.52$)	$R_{\text{had}1}$	✓	✓	✓
	Ratio of E_T in the HCAL to E_T of the EM cluster (used over the range $0.8 < \eta < 1.37$)	R_{had}	✓	✓	✓
EM second layer	Ratio of the energy in $3 \times 7 \eta \times \phi$ cells over the energy in 7×7 cells centered around the photon cluster position	R_η	✓	✓	✓
	Lateral shower width in η	$w_{\eta 2}$	✓	✓	✓
	Ratio of the energy in $3 \times 3 \eta \times \phi$ cells over the energy of 3×7 cells centered around the photon cluster position	R_ϕ	✓	✓	
EM first layer	Lateral shower width in 3 strips around the maximum	$w_{\eta 1}$ or w_1	✓	✓	
	Total lateral width	$w_{\text{s tot}}$	✓	✓	
	Energy outside the core of the three central cells, within seven cells divided by the energy within the three central strips	f_{side}	✓	✓	
	Difference between the energy associated with the second maximum in the strip layer with the minimum value found between the first and second maxima	ΔE	✓	✓	
	Ratio of the energy difference between the maximum energy deposit and the energy deposit in the secondary maximum in the cluster to the sum of these energies	E_{ratio}	✓	✓	
	Ratio of the energy in the first layer to the total energy of the EM cluster	f_1	✓	✓	

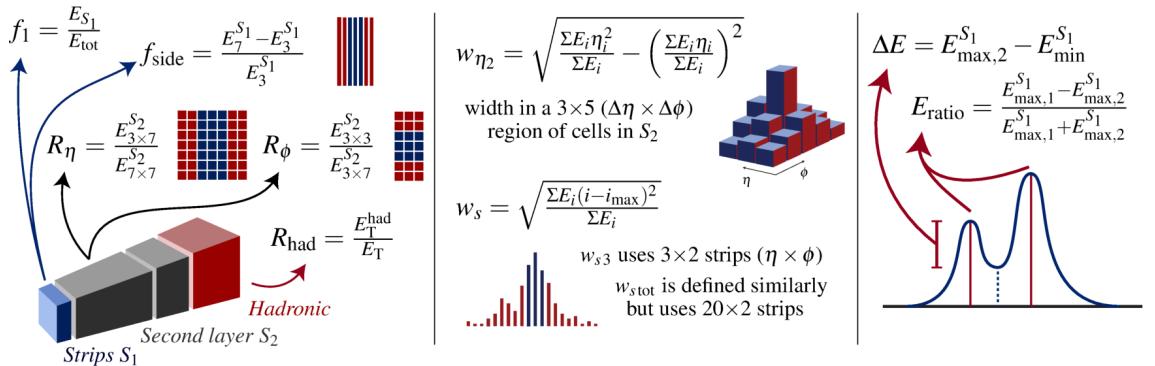


Figure 4.1: Schematic representation of the photon **SSs**. The values $E_C^{S_N}$ represent the energy in layer N of the **ECAL** in a cluster C .

In the following, the **SSVs** used for photon identification are detailed, shown summarised in Table 4.1 and a scheme of how they are calculated is shown in Figure 4.1. The first variables make use of the energy measured in the **HCAL**:

- Hadronic leakage: is the transverse energy deposited in the **HCAL**, normalized to the energy deposited in the **ECAL**:

$$R_{\text{had}(1)} = \frac{E_{\text{T}}^{\text{had}}}{E_{\text{T}}^{\text{EM}}} \quad (4.1)$$

In order to minimize the effects of resolution degradation, in the barrel-endcap transition region of the **HCAL** ($0.8 \leq |\eta| \leq 1.37$) the energy deposit in the whole **HCAL** is used (R_{had}). On the remaining of the detector, only the energy deposited in first layer of the **HCAL** is used ($R_{\text{had } 1}$).

The following variables use the second-layer information of the **ECAL**:

- Lateral energy profile in η :

$$R_\eta = \frac{E_{3 \times 7}^{s2}}{E_{7 \times 7}^{s2}} \quad (4.2)$$

where $E_{i \times j}^{s2}$ is the energy sum in the second calorimeter layer contained in a window of $i \times j$ cells (units of $\eta \times \phi$ cells), centered at the most energetic cell. This variable gives a measure of the showers' width in the η direction.

- Lateral energy profile in ϕ :

$$R_\phi = \frac{E_{3 \times 3}^{s2}}{E_{3 \times 7}^{s2}} \quad (4.3)$$

defined in a similar way as R_η . However, this variable behaves very different for converted and unconverted photons. Due to the action of the magnetic field, the electrons and positrons are curved into opposite directions in ϕ , therefore leading to wider **EM** showers for converted photons compared to those from unconverted ones.

- Lateral shower width in η :

$$w_{\eta^2} = \sqrt{\frac{\sum E_i \eta_i^2}{\sum E_i} - \left(\frac{\sum E_i \eta_i}{\sum E_i} \right)^2} \quad (4.4)$$

measures the proper width of the **EM** shower, where E_i is the energy in the i -th cell of the **ECAL**, measured in a window of 3×5 cells in $\eta \times \phi$.

The following variables use the information from the first **ECAL** layer, composed of the strip cells that allow for a high η resolution and allows for a good separation between isolated photons from photons product of the π^0 decay. Figure 4.2 shows the difference in the energy deposited in the **ECAL** between the two cases mentioned previously.

- Lateral energy profile in η

$$f_{\text{side}} = \frac{E_7^{s1} - E_3^{s1}}{E_3^{s1}} \quad (4.5)$$

measures the energy outside the core of the three central strips within a window of 7 cells, divided by the energy in the three central cells.

- Lateral shower width in η (3 strips)

$$w_1 = \sqrt{\frac{\sum E_i(i - i_{\max})^2}{\sum E_i}} \quad (4.6)$$

where i runs over all cells in a window of 3 cells around the highest-energy-cell. This variable measures the width of the EM shower in the first layer of the calorimeter.

- Lateral shower width in η (full). It is defined in a similar way as w_1 , but uses all the cells in a window of $\Delta\eta \times \Delta\phi = 0.0625 \times 0.2$, corresponding to approximately to 20×2 strips $\eta \times \phi$.
- Energy difference

$$\Delta E = E_{\max,2}^{s1} - E_{\min}^{s1} \quad (4.7)$$

represents the energy difference between the second maximum and the minimum reconstructed energy between the two maxima in the strip layer.

- Energy ratio

$$E_{\text{ratio}} = \frac{E_{\max,1}^{s1} - E_{\max,2}^{s1}}{E_{\max,1}^{s1} + E_{\max,2}^{s1}} \quad (4.8)$$

is the ratio of energy difference between the two maxima, normalized to the sum of those energies, in the strip layer.

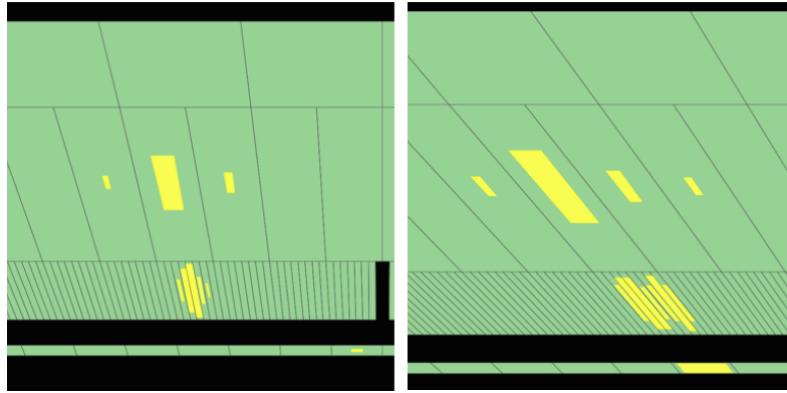


Figure 4.2: Characteristic energy deposits by an isolated photon (left), and a $\pi^0 \rightarrow \gamma\gamma$ event (right), which is possible to distinguish thanks to the granularity of the first ECAL layer [58].

4.2 Photon Identification

The identification of prompt photons over fake photons in hadronic collisions is particularly challenging. Fake photons are vastly dominated by reconstructed photon candidates arising

from hadron decays in jets, while a smaller fraction of fake candidates are associated with hadrons that deposit significant energy in the **ECAL**, mimicking that of real photons. Processes with prompt photons in the final state, occurring in pp collisions at the **LHC**, play a central role in the **ATLAS** physics programme. Either for searches or precision measurements, it is important to rely on excellent algorithms and techniques to identify the real photons over the fake ones. These searches or measurements are performed in a very wide range of the photon p_T , starting from very light Higgs resonances like to a pair of axion-like particles to 4 photons ($H \rightarrow aa \rightarrow 4\gamma$) [59], where the photon p_T is ~ 25 GeV, up to very high- p_T photons in searches for $\gamma + \text{jet}$ resonances, with photons having $p_T > 1$ TeV. In this section, the procedures for optimising the identification **WPs** and measuring the corresponding efficiencies are described.

4.2.1 Processes and event selection

Given the very wide range in which photons are used in **ATLAS**, the optimisation of the tight identification **WP** relies on two different processes that eventually allow for clean photon samples in the low and high p_T regimes. In the low- p_T case, a very clean source of photons from radiative Z decays are used. On the other hand, although with higher background contamination, Single Photon (**SP**) events are employed for high p_T photons. In the following paragraphs, a description of each photon sample is given.

Radiative Z decays In the low- p_T range, photons from radiative decays of the Z boson ($Z \rightarrow \ell\ell\gamma$) are selected as signal photons. There are two different production modes possible for the Standard Model (**SM**) $pp \rightarrow Z(\ell^+\ell^-)\gamma$ processes, where ℓ is either an electron or a muon. These are: Initial State Radiation (**ISR**) where the photon is radiated from the quarks, and Final State Radiation (**FSR**) (hereinafter also referred as Radiative Z (**RZ**) decays), where the photon is radiated from one of the final-state leptons through bremsstrahlung. Both production modes are shown in Figure 4.3.

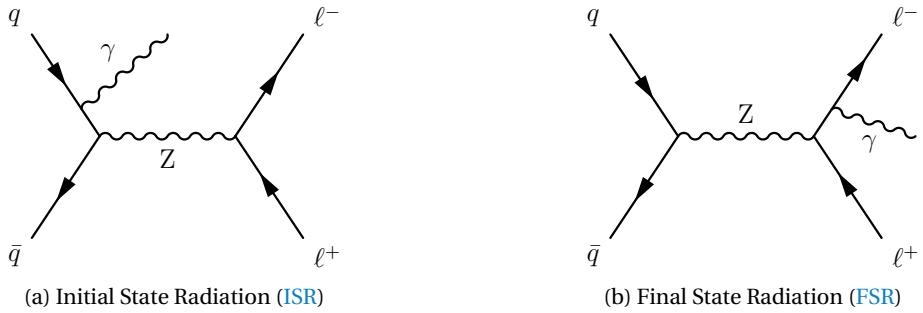


Figure 4.3: Feynman diagrams for radiative Z decays $Z \rightarrow \ell\ell\gamma$ for the initial-state (left) and final-state (right) radiation.

Both the **FSR** and **ISR** processes can be easily identified by comparing the two-body invariant mass $m_{\ell\ell}$ distribution to the three-body invariant mass $m_{\ell\ell\gamma}$ distribution. For **ISR** events, $m_{\ell\ell}$ follows the Z line-shape, and the photon simply adds to the invariant mass making it larger than 91 GeV. In the **FSR** case, the three-body invariant mass $m_{\ell\ell\gamma}$ follows the Z line-shape,

which is seen from Figure 4.4a. For photon identification studies, only photons from **FSR** events are considered. The reason behind the selection of **FSR** over **ISR** is driven by the following. **ISR** events also suffer from $Z + \text{jets}$ background contamination, where the jet is misidentified as a photon, and the $Z + \text{jets}$ cross-section is magnitudes higher than for $Z + \gamma$. From Figures 4.4b and 4.4c, where the $m_{\ell\ell} - m_{\ell\ell\gamma}$ distributions are shown for $Z \rightarrow \ell\ell$ and $Z \rightarrow \ell\ell\gamma$ processes, respectively, the separation between these two can be appreciated when selecting **FSR** photons.

With the **RZ** sample, the photons are required to have a transverse momentum $p_T > 7 \text{ GeV}$ and a pseudorapidity in the range $|\eta| < 1.37$ or $1.52 < |\eta| < 2.37$, to avoid the crack region. For the optimisation studies no photon isolation is applied, but loose photon isolation, described in Section 3.2.3, is used for the efficiency measurements. To avoid any biases on the photon footprints in the calorimeter, no other selection is applied for them. Leptons are required to have $E_T > 10 \text{ GeV}$, muons pseudorapidity to be $|\eta| < 2.5$ and for electrons $|\eta| < 2.47$, excluding the crack. Both electrons and muons need to pass loose isolation requirements and to pass medium identification.

FSR photon then are selected by requiring $80 < m_{\ell\ell\gamma} < 100 \text{ GeV}$, and $40 < m_{\ell\ell} < 83 \text{ GeV}$. Finally, to avoid any biases on the photon **SS** and isolation variables, a minimum distance of $\Delta R > 0.4$ is required between the photon and the closest lepton.

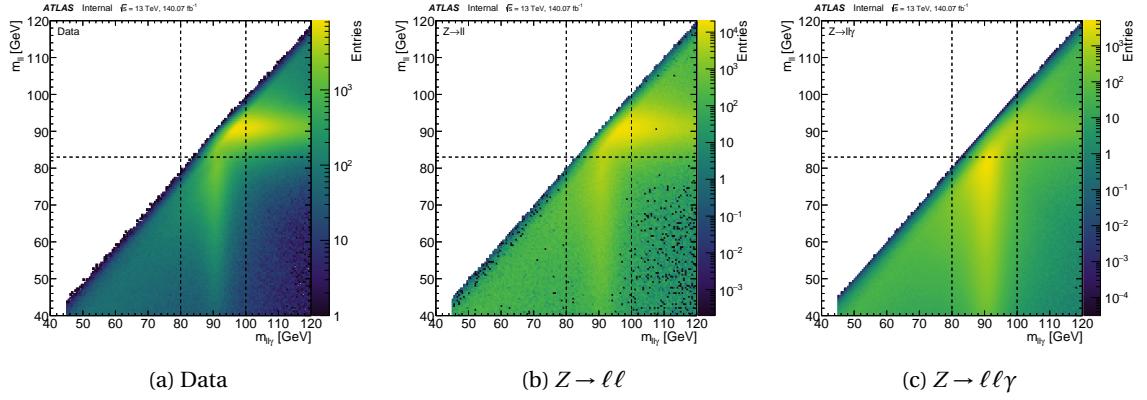


Figure 4.4: Two-dimensional invariant mass distribution of the two and three body systems $m_{\ell\ell}$ and $m_{\ell\ell\gamma}$, respectively, for (a) data, (b) background and (c) signal. The region in which there is a high concentration of events for $m_{\ell\ell} \sim m_Z$ corresponds to **ISR** events, while **FSR** events can be identified when $m_{\ell\ell\gamma} \sim m_Z$.

Single Photon (SP) The inclusive photon, or Single Photon (**SP**) sample is collected by single-photon triggers, whose thresholds range from 10 GeV up to 140 GeV and require loose photon identification. Although the triggers used to obtain this sample are prescaled (with the exception of the 140 GeV one) the provide very large photon datasets for high p_T . These processes include leading-order $\gamma + \text{jet}$ events from $qg \rightarrow q\gamma$ and $q\bar{q} \rightarrow g\gamma$ hard scattering, as well as prompt photons from quark fragmentation in **QCD** dijet events. Photons from these events are required to satisfy $|\eta| < 2.37$ excluding the crack, and to pass the loose isolation requirement. The **SP** samples are used for both the optimisation and measurements studies.

4.2.2 Optimisation

Starting from these discriminating **SSs**, three **WPs** can be defined for photons: *loose*, *medium* and *tight* [60]. The loose **WP** employs cuts to the variables defined in the second layer and to the hadronic leakage variable, used primarily by the trigger. The medium and tight **WPs** use all the previously defined variables. The former is optimised to have a flat 95% efficiency, while the latter provides an excellent background rejection. Table 4.1 shows which variables are used for each **WP**.

The optimisation of the **WPs** uses two different samples: **RZ** events for photons with $10 < p_T < 25$ GeV as signals and $Z \rightarrow \ell\ell$ as backgrounds; and for the high- p_T regime ($p_T > 25$ GeV) **SP** events are considered signals while dijet events are the backgrounds.

As previously mentioned, photon identification uses cuts to photon Shower Shapes (**SSs**). Examples of the R_η , E_{ratio} and $w_{\eta 2}$ **SS** comparing signal and background events using the **RZ** samples are shown in Figure 4.5, where excellent discriminating power is seen. The cuts on all the **SSs**, for each identification **WP**, are optimised as a function of the transverse energy and the pseudo-rapidity of the photon candidate, to account for the shape of the variables for different η and for variations in the amount of material and the geometry of the calorimeter. The medium and tight **WPs** are also optimised separately for converted and unconverted photons. The cuts are optimised using a Multivariate (**MV**) approach, where signal efficiencies are scanned between 0% and 100% while trying to maximise the background rejection. The resulting, optimised, cut values are subject to fluctuations and therefore they are manually smoothed.

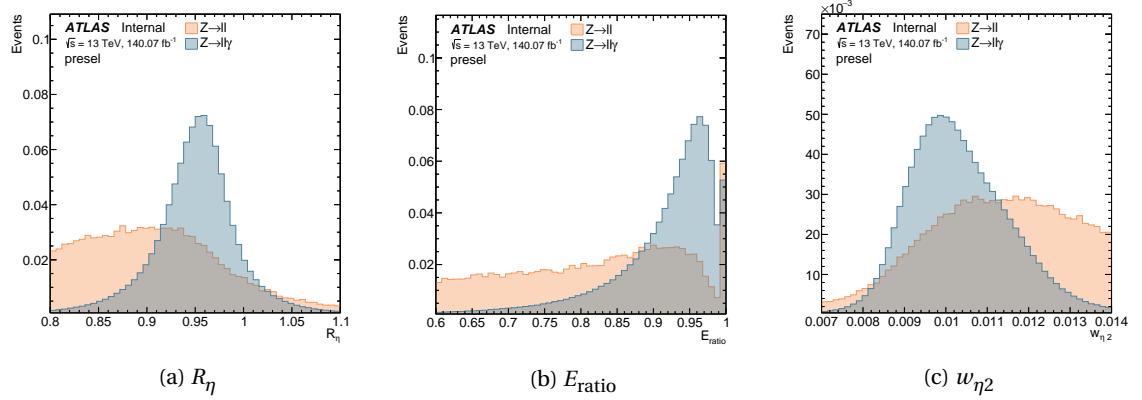


Figure 4.5: Normalised signal (blue) and background (orange) distributions for different **SSs**, using **RZ** event samples, passing the event selection detailed in Section 4.2.1.

4.2.3 Efficiency measurements

The most used **WP** for precision measurements and general searches in **ATLAS** is the tight **WP**. Once the cuts to the **SSs** are optimised as previously explained, data and **MC** identification efficiencies are calculated. In all cases, photons are required to satisfy the loose isolation criterion defined in Section 3.2.3 and therefore the photon efficiencies are measured relative to this isolation criterion. These measurements are carried out using three different methods that

are detailed in Ref. [61] and in the following paragraphs, a brief description of each method is given.

For the lower p_T range ($7 < p_T < 100$ GeV), photons from radiative Z decays are used as signal photons. The process to estimate the efficiencies relies on using template fits to the observed three-body invariant-mass ($m_{\ell\ell\gamma}$) distribution before and after applying the tight identification criteria. The number of signal and background events can then be counted from the fits, and signal purities are computed each time: P^{total} before applying tight identification and P^{pass} after. The final efficiency in data is then given by

$$\varepsilon_{ID} = \frac{P^{\text{pass}} N_{\text{data}}^{\text{pass}}}{P^{\text{total}} N_{\text{data}}^{\text{total}}}. \quad (4.9)$$

The second method to compute efficiencies relies on Smirnov transformations [62] to the electrons' **SSs** to resemble those of photons'. The samples used in this approach are $Z \rightarrow ee$ decays, in which the electrons are required to pass loose photon isolation. The candidate electrons in data contain a small background from $W + \text{jets}$ and multijet production; this background is subtracted by fitting simulated signal samples and background templates derived from data control regions to the m_{ee} data distributions. The electron candidates are counted from events in the range $70 < m_{ee} < 110$ GeV, and the efficiencies are measured using the tag-and-probe method described in Ref. [41]. The p_T range in which this method is implemented is $25 < p_T < 250$ GeV.

The final and third method uses **SP** samples with photons in the range $50 < p_T < 1500$ GeV. The matrix method [61] is used in this case, which constructs four orthogonal regions that either pass or fail the tight identification **WP**, and pass or fail track-isolation (described in Section 3.2.3). For each region, two unknowns arise: the number of signal and background events. If the track isolation efficiencies are known for the signal and background components, then it is possible to estimate the efficiency for loose photons passing the tight identification criteria. The isolation efficiencies for signal photons are estimated using **MC** samples, and the ones for backgrounds are obtained in a jet-enriched control region constructed by inverting the identification criteria. The efficiency measurements in data for the tight identification **WP** then reads:

$$\varepsilon^{\text{tight-ID}} = \frac{\frac{\hat{\varepsilon}_{ID} - \hat{\varepsilon}_{ID}^b}{\hat{\varepsilon}_{ID}^s - \hat{\varepsilon}_{ID}^b} \cdot N_{ID}^T}{\frac{\hat{\varepsilon} - \hat{\varepsilon}^b}{\hat{\varepsilon}^s - \hat{\varepsilon}^b} \cdot N^T}, \quad (4.10)$$

where N^T accounts for the totality of photons in the inclusive sample which consists on N^s prompt photons (or signal photons) and N^b fake photons (background photons). The number N_{ID}^T is the subset of N^T that pass the identification requirement. Data, signal and background track isolation efficiencies are represented by $\hat{\varepsilon}$, $\hat{\varepsilon}^s$ and $\hat{\varepsilon}^b$, respectively. Similarly, the track isolation efficiencies for those photons passing tight identification are shown as $\hat{\varepsilon}_{ID}$, $\hat{\varepsilon}_{ID}^s$ and $\hat{\varepsilon}_{ID}^b$, respectively. The measured efficiencies for photons with $p_T > 150$ GeV is between 90 and 96%.

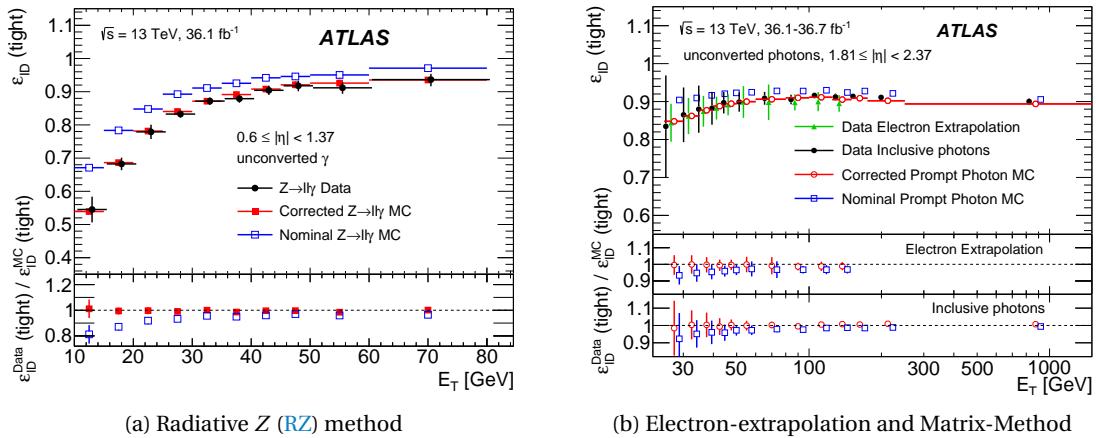


Figure 4.6: Comparison of the measurements between data and MC of the three distinct data-driven methods to compute efficiencies. In both figures, for each method, two different set of MC measurements are displayed: the nominal one and the corrected one. The bottom panels show the ratio of data efficiencies to MC predictions (referred as Scale Factors (SFs) in the text). The figures are taken from Ref. [61].

Example of the photon identification efficiencies as a function of the photon p_T using the **RZ** method is shown in Figure 4.6a. Data efficiencies are represented by the black points, while nominal **MC** is shown with blue empty squares. The ratios of data to **MC** (also referred as Scale Factors (**SFs**)) shown in the bottom pad using the nominal **MC** vary up to 20%, showing that the simulation is not correctly describing the data. However, another set of **MC** efficiencies is displayed, in this case using corrected **MC**, which drastically improves the agreement between data and simulation, as seen from the **SFs**. The reason why these corrections are needed and how they were implemented in **ATLAS** is explained in the following section (Section 4.3), and how they are currently corrected in Chapter 5. Figure 4.6b shows the efficiency measurements using the two remaining methods (electron-extrapolation and matrix-method), where the same improvements on the **SFs** is obtained when using the corrected **MC** simulation.

As mentioned above, these ratios between data and MC efficiencies are referred as Scale Factors (SFs) and they encapsulate the differences between data and simulation. They are computed separately for each one of the three methods and are later combined using a weighted average [63] in each bin and assuming the statistical and systematic uncertainties to be uncorrelated between the methods. Current results of these SFs, computed using the full Run-2 dataset, are shown in Figure 4.7.

4.3 Shower shapes variables differences between data and MC

The **ATLAS MC** simulation does not perfectly describes data. This is clearly seen when computing the previously mentioned **SFs**, whose values were different from 1, meaning that different efficiencies are obtained between data and in **MC**. In particular, when comparing the **SSs** distributions, it is seen that **MC** distributions are shifted or even the whole shape differs, as shown in Figure 4.8, by comparing data (black dots) against the red line histogram corresponding to **MC**.

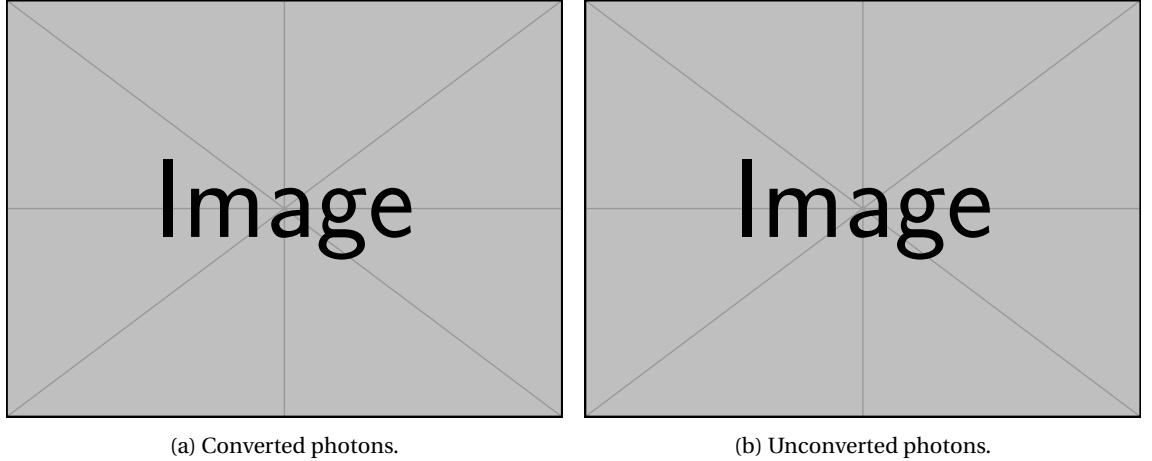


Figure 4.7: Photon identification SFs in the different p_T - $|\eta|$ bins for both converted (left) and unconverted photons (right). **Ask fran for the plots in his presentation!**

The main differences on the distributions arise for the η shower profiles, where broader distributions are seen in data compared to MC. Part of the effect was corrected in 2010 after moving to detailed description of the material composition in the accordion absorbers in GEANT4. However, the remaining data-MC disagreements are still under study and could be due to several potential effects:

- Detector geometry description of the lead thickness (including possible variations due to gravity).
- Mismodeling of the electric field in the LAr gaps.
- Mismodeling of the cross-talk effect (energy sharing between calorimeter cells due to electronics possible in η direction).

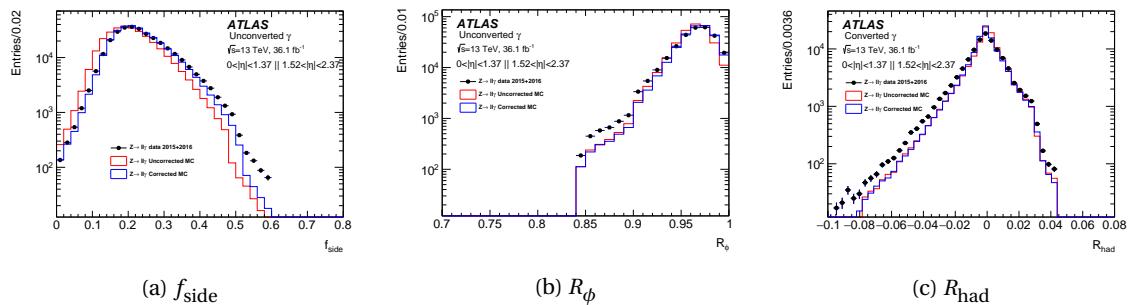


Figure 4.8: Example of SSs comparisons between data (black dots) against nominal (red line) and corrected (blue line) MC simulation, using the full Run-2 RZ photon sample [64].

To account for the differences in the SSs, historically, corrections were made in the form of shifts to each one of the MC distributions. These shifts comprised the so-called Fudge Factors (FFs), and were determined using a χ^2 minimisation on the comparison of data and MC SSs [41, 61]. Even though the mean value differences decreased substantially after these corrections as seen for example in the case of f_{side} in Figure 4.8a, residual but notable differences remained. It is seen from the distributions that the main differences that remained are related to the shape

of them, therefore needing for higher order corrections. In the following chapter, a detailed description of newly derived corrections is presented. Moreover, since **SSs** are built from energy deposits on the **ECAL** cells, another possible way of correcting the current disagreement between data and **MC SSs** is to directly correct the energies on **MC** at a cell-level, fixing the differences in all **SSVs** at once. This new approach is described as well in the following chapter.

SHOWER SHAPES CORRECTIONS

5

“Champions keep playing until they get it right.”

Billie Jean King

In the previous chapter, it was seen that **SFs** (ratio of data efficiencies to simulated ones) deviate from unity. Given that photon identification relies on cuts to the photon **SSs**, it was found that the differences in fact appear on these variables. Since Run-1, they have been corrected with what is known as Fudge Factors (**FFs**), which, historically, they have been computed as simple shifts to the **MC** distributions and have been found to provide very good improvements on the **SFs**. However, as seen before, there are still discrepancies between the distributions that need to be addressed in order to rely on even a better simulation. In Section 5.1, a more sophisticated approach based on a higher order computation to correct the **SSs** is presented. Also, a novel approach using directly the cells energies is studied and addressed in Section 5.2. The studies presented in this chapter comprise one of the main topics of work for the current thesis.

5.1 Fudge Factors (**FFs**)

5.1.1 Data and simulated samples

FFs are computed using full Run-2 dataset, collected at $\sqrt{s} = 13$ TeV and with a corresponding integrated luminosity of 140 fb^{-1} . Both the **RZ** and **SP** simulated samples are used for this study, as their represent complementary p_T -ranges. **RZ** events are generated with SHERPA 2.2.11 [65], while SHERPA 2.2.1 is used for $Z \rightarrow \ell\ell$ background events. Respecting the **SP** samples, events are generated with PYTHIA 8.186 [66], which includes leading-order $\gamma + \text{jet}$ events from both direct processes ($qg \rightarrow q\gamma$ and $q\bar{q} \rightarrow g\gamma$) and photon fragmentation from **QCD** dijet events.

In both cases, the **ATLAS** detector is simulated using GEANT4 4 [67] and the **MC** events are reweighted so that their pileup distributions resembles the one in data, for each year of the Run-2 data-taking period.

5.1.2 Calculation

The calculation is performed separately for the two considered samples: **RZ** for photons with $7 \leq p_T \leq 50$ GeV and **SP** for photons with $p_T > 50$ GeV, which were already discussed in Section 4.2.1. Since **SSs** distributions vary as a function of p_T and $|\eta|$, the computation is done in bins of these variables:

$$p_T^\gamma : \begin{cases} \text{RZ} : [7, 15, 20, 30, 50] \text{ GeV} \\ \text{SP} : (50, 60, 80, 100, 150, 300, 600, \infty] \text{ GeV} \end{cases} \quad (5.1)$$

$$|\eta| : [0, 0.6, 0.8, 1.15, 1.37, 1.52, 1.81, 2.01, 2.37]. \quad (5.2)$$

Furthermore, as mentioned in Section 4.1, there are variables very sensitive to the photon's conversion status, that is, whether if the photons are converted or unconverted. For this reason, the calculation is done separately for converted and unconverted photons. A total of nine variables are corrected using this method: E_{ratio} , f_{side} , R_η , R_ϕ , R_{had} , $R_{\text{had}\,1}$, w_1 , $w_{\eta\,2}$ and $w_{\text{s tot}}$; as they are the ones in which the largest discrepancies are observed between data and **MC**.

For each **SS**, histograms of **MC** and data of 100 bins are created. The choice of the binning is done based on having sufficient statistics at each bin and also to capture all the features of the variables. After that, each histogram is smoothed using the Kernel Density Estimator (**KDE**) tool from TMVA [68]. The **KDE** method consists of estimating the shape of a Probability Density Function (**PDF**) by the sum over smeared events. The **PDF** $p(x)$ of a variable x is

$$p(x) = \frac{1}{N} \sum_{i=1}^N K_h(x - x_i) \quad (5.3)$$

where N is the number of events, $K_h(t) = K(t/h)/h$ is the kernel function, and h is the bandwidth of the kernel. The basic idea is that each event is considered as a Dirac- δ -function, which is replaced by a Kernel function (Gaussian) and finally they are summed altogether to form the final **PDF**. The **KDE** smoothing can be applied in two forms: non-adaptive **KDE** or adaptive **KDE**, as seen in Figure 5.1. In the former, the bandwidth is constant for the entire sample h_{NA} , while in the latter, it uses the value from non-adaptive **KDE**, but it varies as a function of $p(x)$ as

$$h_A = \frac{h_{NA}}{\sqrt{p(x)}} \quad (5.4)$$

Adaptive **KDE** improves the shape of the estimated **PDF** in regions of low statistics, however, in high statistics regions it can give rise to over-smoothing. The degree of smoothing is tuned by multiplying the bandwidth h by fine factors. These fine factors are user-defined parameter which are tuned to allow the **PDF** to retain the important features of the **SS** but to also avoid statistical fluctuations. Higher values indicate broader Kernel functions and therefore de **PDF** catches less statistical fluctuations. Examples of the smoothing procedure applied to R_{had} are shown in Figure 5.2 for cases in which original histograms have low and high statistics.

Once the data and **MC PDFs** are created for a given variable, p_T , $|\eta|$ and conversion type, the **MC PDF** is normalised to data's and a χ^2 value is computed between both, excluding the underflow

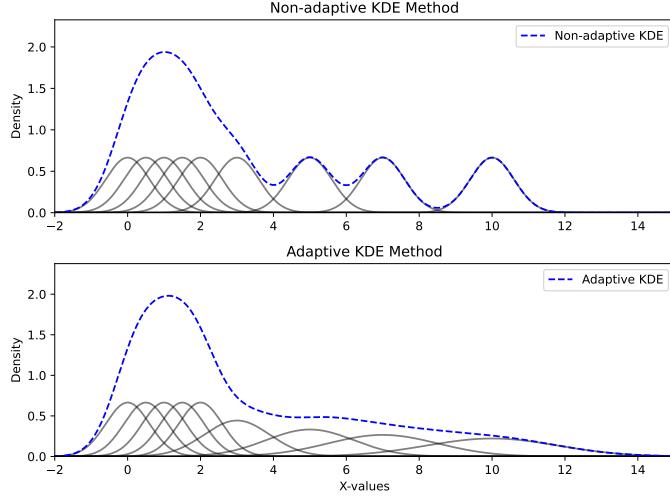
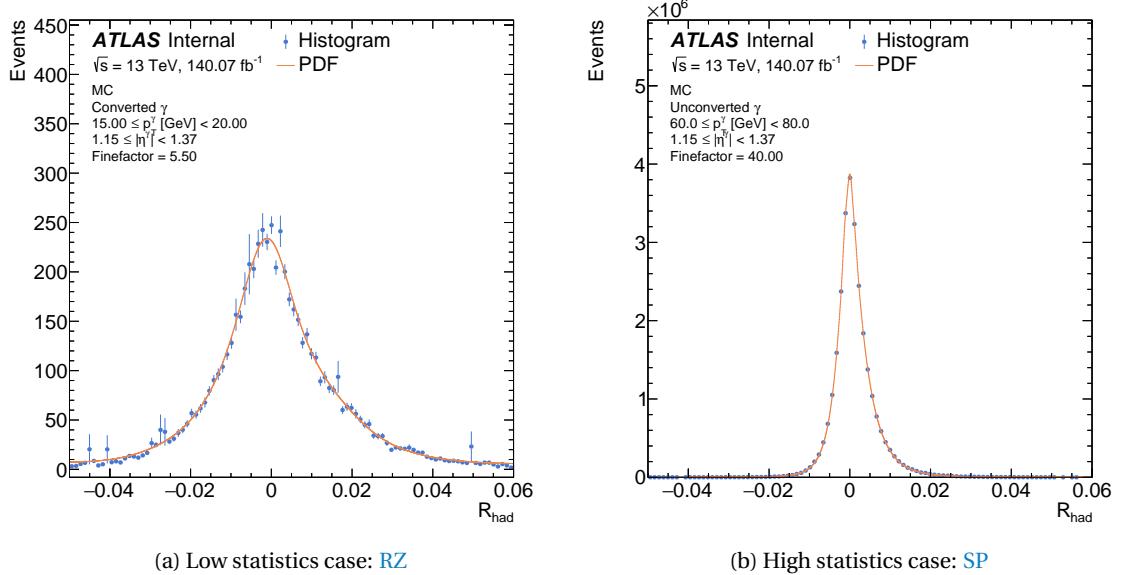


Figure 5.1: Adaptive and non-adaptive KDE smoothing.

Figure 5.2: KDE smoothing applied to the R_{had} for photons in $0.8 < |\eta| < 1.15$ in two scenarios: low and high statistics. Original histograms are shown with the blue points and their PDFs with the orange lines. The fine factors used for each case are displayed in the figure.

and overflow bins, as:

$$\chi^2 = \sum_{i=1}^N \frac{(w_{\text{MC},i} W_{\text{data}} - w_{\text{data},i} W_{\text{MC}})^2}{s_{\text{MC},i}^2 W_{\text{data}}^2 + s_{\text{data},i}^2 W_{\text{MC}}^2}. \quad (5.5)$$

N is the number of bins in the PDFs, $w_{\text{MC},i}$ and $w_{\text{data},i}$ are the event numbers of MC and data at each bin, respectively, $s_{\text{MC},i}$ and $s_{\text{data},i}$ are the bin errors and finally W_{data} and W_{MC} are the sum of weights for data and MC, respectively.

Shift-only corrections

Taking into account only the mean's correction of the SSs, the MC PDF is shifted to the left and right one bin at a time. As a consequence of this procedure, the shift FF resolution directly depends on the bin-width of the PDFs, therefore having smaller bin-widths mean to obtain

a better resolution on the shift value. Given that histograms, in the first place, are built with relative wide bins, the **PDFs** can be constructed using high-accuracy narrow bin to ensure high resolution. After convergence tests on the **FFs**, the **PDFs** are constructed with 5000 bins. The starting number of bins that the **MC** distribution needs to be shifted is estimated by computing the difference on the means between data and simulation. From this starting value, shifts of 100 bins to each side are considered.

For each bin the distribution has been shifted, the aforementioned χ^2 value is computed and recorded. Assuming that the measurements errors $s_{MC,i}$ and $s_{data,i}$ have a normal gaussian distribution¹, and that the parameters for each χ^2 value are independent, it is expected that the shape followed by the χ^2 values is approximately paraboloidal.

To extract the **FFs**, the χ^2 scan near the minimum is fitted with a parabolic function (5 bins to each side of the minimum bin) and the shift **FF** is obtained from the fit minimum. Finally, the **SSs** can be corrected as

$$SS_{\text{new}} = SS_{\text{old}} + \text{shift}.$$

Shift+stretch corrections

It was seen that even after applying corrections to the means of the **MC SSs** distributions, differences remained on the shapes of them, and in some cases these can be quite substantial. One way to continue improving the agreement between data and **MC** is to include another correction which is referred as *stretching*. The two corrections, called shift+stretch corrections, are meant to fix both the mean and the widths of the **MC** distributions simultaneously.

The shift+stretch correction method starts by finding the maximum of the **MC PDF**. The **PDF** is then stretched around it by calculating the new position of each bin by the product: stretch \times (x – stretch point). In this manner, each bin's center conserves the initial distance to the center of the distribution, multiplied by the stretch factor. In the scenario where the shift is > 1 , there might be cases in which it is big enough to give rise to empty bins inbetween. The content of these empty bins are then linearly interpolated from the neighbouring non-zero bins. Once the **PDF** is stretched, it is then shifted left and right following the same procedure as for the shift-only case, calculating χ^2 values for each shift_{*i*} after applying stretch_{*j*}. As a result of the procedure, now, a two-dimensional grid on the shift-stretch plane of χ^2 values is obtained. The pair of shift-stretch **FFs** is now retrieved from the minimum bin's center, and the corrections are applied to the **MC SSs** as:

$$SS_{\text{new}} = \text{stretch} \times (SS_{\text{old}} - \text{stretch point}) + \text{shift} + \text{stretch point}. \quad (5.6)$$

An example of the resulting χ^2 values for the f_{side} variable for unconverted photons with $15 < p_T < 20$ GeV in $2.01 < |\eta| < 2.37$ is shown in Figure 5.3a, where the shift is on the x -axis and the stretch on the y -axis. The optimal shift-stretch value is given be the position of the minimum

¹ This requirement is satisfied as long as the bin contents of both **PDFs** are greater than 10, which is also satisfied since histograms are built with relatively wide bins.

bin, which corresponds to shift = 0.03 and stretch = 1.09. A visualisation of the PDFs before and after applying the corrections is shown in Figure 5.3b, where they are compared with the data distribution. As seen from the figure, there is a huge improvement and the distributions match almost perfectly.

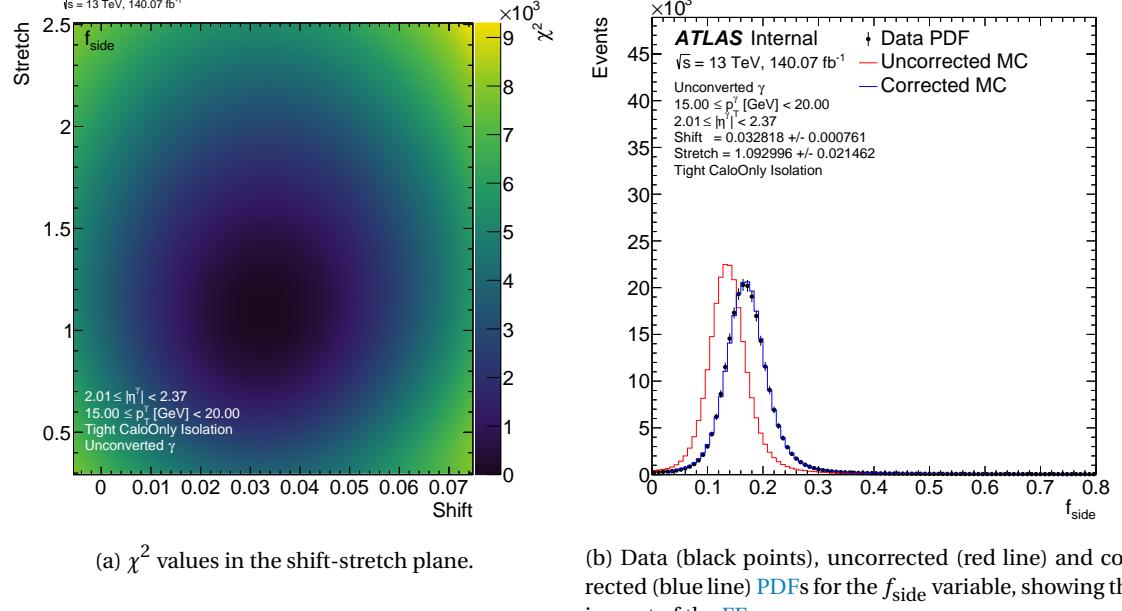


Figure 5.3: Calculation of shift+stretch FFs for f_{side} with unconverted photons.

5.1.3 Uncertainties

Statistical uncertainties

To extract the statistical uncertainties on the shift and stretch FFs, a fit to the 1σ (68.3% confidence level) contour on the χ^2 values is performed. This contour represents an ellipse in the large sample limit which takes the following form:

$$\chi^2 = \chi^2_{\min} + \frac{1}{1-\rho^2} \left[\left(\frac{x-x_0}{\sigma_x} \right)^2 + \left(\frac{y-y_0}{\sigma_y} \right)^2 - 2\rho \left(\frac{x-x_0}{\sigma_x} \right) \left(\frac{y-y_0}{\sigma_y} \right) \right], \quad (5.7)$$

where ρ is the correlation coefficient between both variables, σ_x and σ_y the uncertainties on x and y , respectively, (x_0, y_0) is the location of the ellipse's center, and χ^2_{\min} is the χ^2 minimum value obtained from the 2D histogram.

By extracting the semi-major and semi-minor axes of the fitted ellipse, and with the tilt angle of it, the statistical uncertainties on two variables x and y (in this case representing the shift and stretch, respectively) are (see Appendix A):

$$\sigma_x = \sqrt{a^2 \cos^2 \theta + b^2 \sin^2 \theta} \quad (5.8)$$

$$\sigma_y = \sqrt{a^2 \sin^2 \theta + b^2 \cos^2 \theta}. \quad (5.9)$$

Systematic uncertainties

The systematic uncertainties are derived by varying the preselection criteria, that is, photon identification and photon isolation. Changing different preselection criteria allows the **SSs** to vary depending on the amount of background contamination, and in consequence so do the **FFs**. The different selections are:

- Radiative Z (**RZ**) sample:
 - Nominal: No ID, `FixedCutTightCaloOnly` isolation.
 - Loose ID, no isolation.
 - Loose ID, `FixedCutTightCaloOnly` isolation.
 - No ID, `FixedCutLoose` isolation.
- Single Photon (**SP**) sample:
 - Nominal: Tight ID, `FixedCutLoose` isolation.
 - Tight ID, `FixedCutTight` isolation.

All other combinations (or lack thereof) of selection criteria would result in either a sample with too low statistics, or too low purity.

FFs are derived for each one of the previous selections, and the difference between the nominal and the varied ones is calculated. The maximum difference is taken as the systematic uncertainty, as the most conservative approach.

5.1.4 Results

Due to the fact that **FFs** are calculated in different p_T using two different samples which span complementary regions, the results are concatenated at 50 GeV. in the following, shift and stretch values are reported for different **SS** variables.

The reported shift values in the figures are normalised by the standard deviation of the **SS** after stretching, as this quantity allows to understand how much each variable is shifted with respect to its width. Also, it provides an unique measure for all considered variables, since they span different ranges. Nevertheless, the variables' width vary for different p_T - $|\eta|$ bins, leading to eventual large differences between neighbouring bins. In such cases, there are no drastic differences differences on the original shift values.

In Figure 5.4, examples of resulting **FFs** for the R_η and $w_{\eta 2}$ variables using converted photons are presented. It can be seen that for both variables the **FFs** depend on p_T , specially towards higher momenta and converted photons. This behaviour is also repeated on all variables. By inspecting the behaviours and trends of the **FFs**, it is also possible to retrieve information on the **MC** mismodelling of the **SSs**. As said in Section 4.3, broader η widths and profiles were observed for data compared to the simulation. This is, in fact, still observed to present day,

since the stretch values increase towards higher p_T , stretching the MC simulations as much as twice theirs initial width. In the case of R_η ($w_{\eta 2}$) (for the displayed $|\eta|$ bin and converted photons), the MC simulation overestimated (underestimated) the central value by almost a standard deviation after fixing the width, meaning that huge differences are still present in the uncorrected MC distribution.

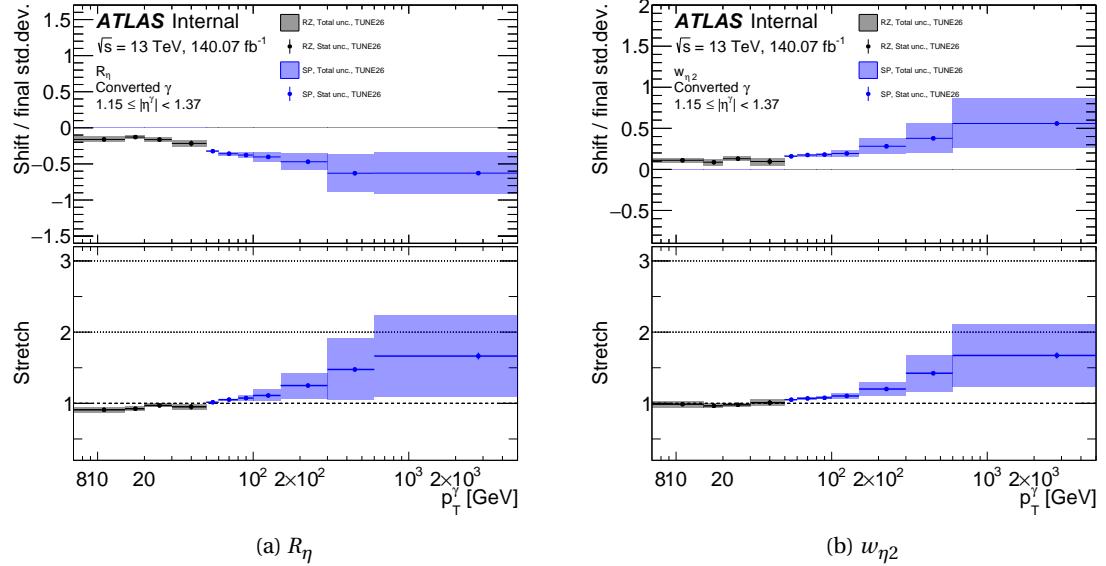


Figure 5.4: Shift and stretch FFs for R_η (left) and $w_{\eta 2}$ (right) SSs for converted photons as a function of the photon p_T with $1.15 < |\eta| < 1.37$. Results from RZ (black) and SP (blue) are displayed, where the shaded regions denote the total uncertainty, while the error bars only the statistical one. Stretch values are shown in the bottom pad. The shift, on the other hand, is shown in the top pad, which are normalised by the standard deviation of the SS after applying the stretch, as indicated in the text.

It is also useful to visualise the FFs for a fixed p_T -bin and as a function of $|\eta|$. This is shown for $w_{s \text{ tot}}$ using converted photons with $50 < p_T < 60$ GeV in Figure 5.5 where the normalised and raw shift values are displayed. For $|\eta| > 1.81$ (last two bins), normalised shift values are greater than previous bins by, at least, a factor of 2. However, the raw shift values shown in Figure 5.5b do not present such an abrupt change, indicating that this is a direct consequence of the change of width of the distributions.

In order to validate the obtained FFs, the corrections are applied to the SSs event-by-event. Figures 5.6 and 5.7 show the application of the FFs to some of the SS distributions using the RZ and SP samples, respectively, divided into the barrel and endcap regions in $|\eta|$. In the barrel region, the corrections indeed improve the agreement, but the magnitudes of these corrections are not as dramatic as in the endcap region, where huge improvements are seen. Taking the w_1 and $w_{s \text{ tot}}$ variables as an example, major shape differences are observed between the nominal simulation and data, which the shift+stretch methods manage to fix. The same behaviour is observed using the SP samples, where these variables present two or more peaks, and they are correctly fixed by the FFs. In all the shown cases, the corrected MC and data are almost undistinguishable showing the importance of these corrections and how they were improved.

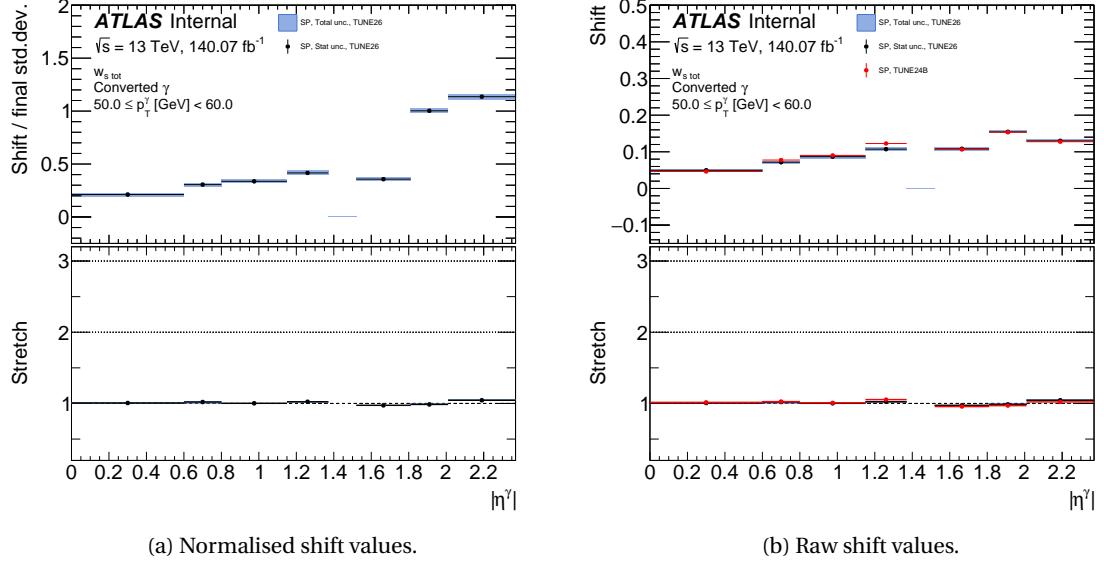


Figure 5.5: Shift and stretch FFs for w_s^{tot} SSs using converted photons as a function of $|\eta|$ and photons from the SP with $50 < p_T < 60$ GeV. Normalised shift values are displayed in Figure a while original, or raw, values are shown in Figure b. Points with the uncertainty bar show the statistical uncertainty only, and the shaded regions represent the total uncertainty. Stretch values are shown in the bottom pad. Shift values are displayed in the top pad, which are normalised by the standard deviation of the SS after applying the stretch.

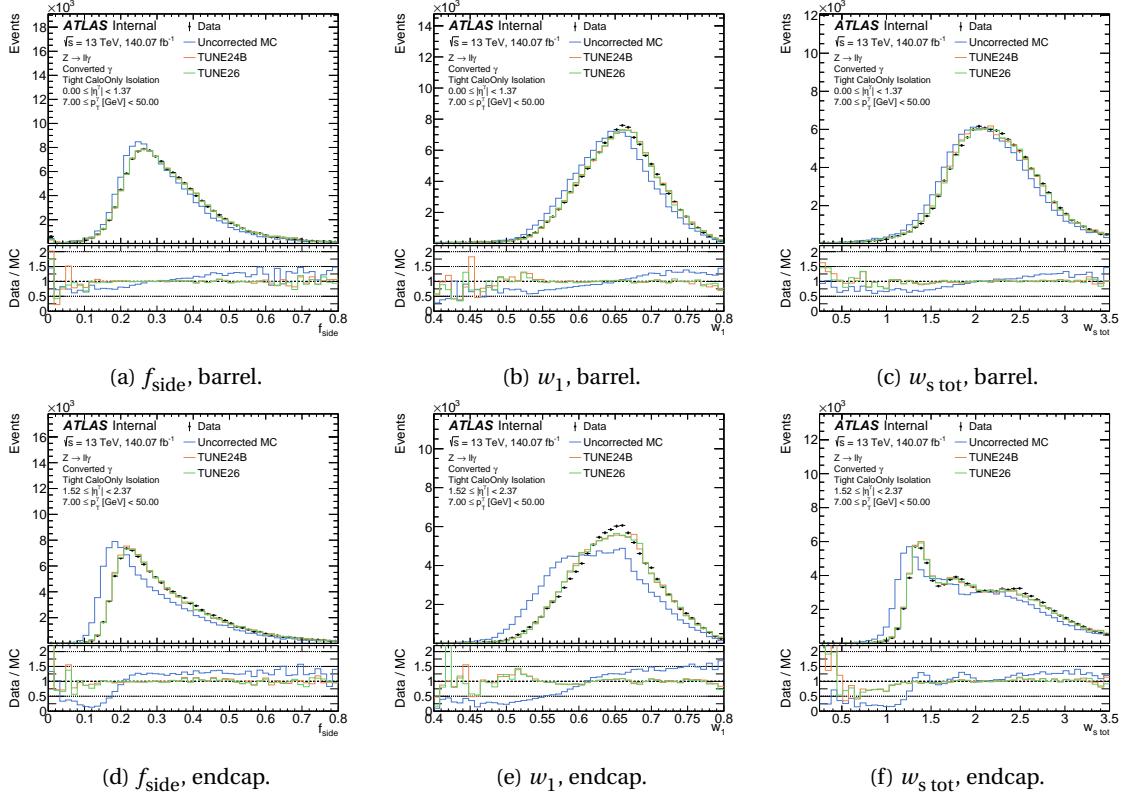


Figure 5.6: Selected SSs distributions using the RZ samples for converted photons after applying the FFs corrections event-by-event to the MC simulation. The SSs are shown in the barrel region (top) and in the endcap (bottom). Data points are represented by the black points and uncorrected (corrected) MC simulation by the blue (green) lines. The bottom pads show the ratio of data to the corresponding MC simulation.

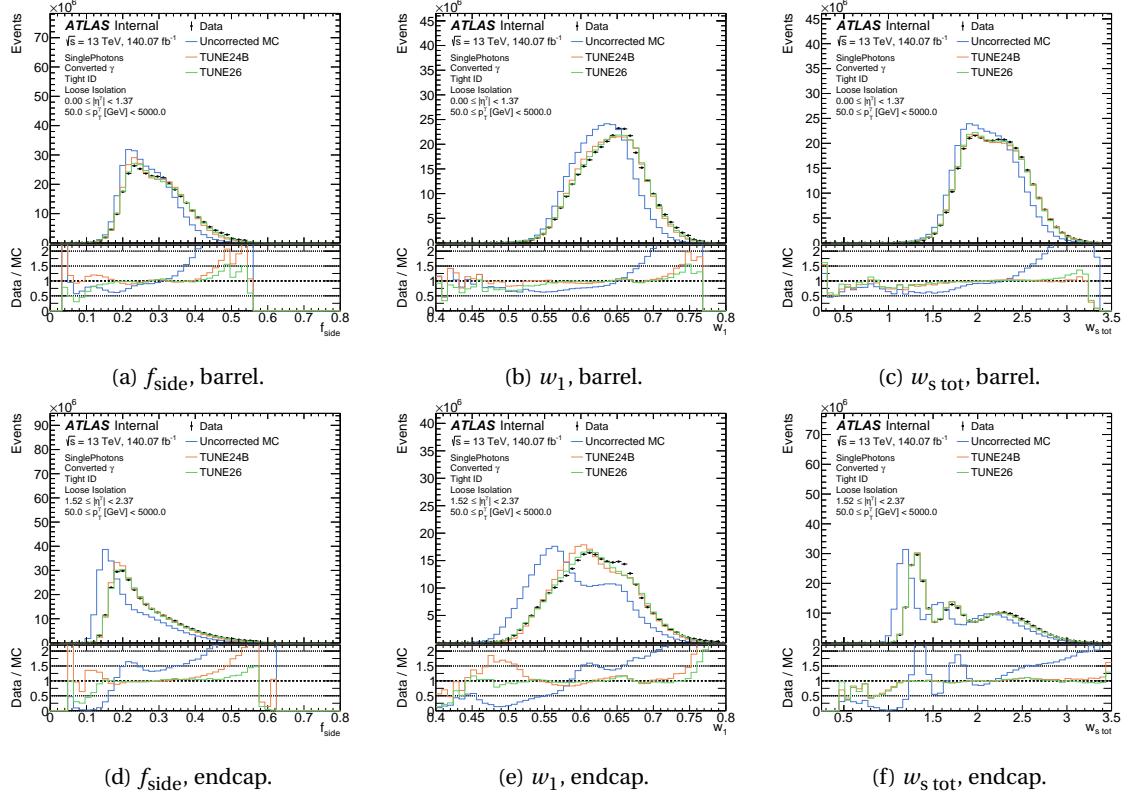


Figure 5.7: Same as Figure 5.6 but with the SP samples.

5.2 Cell-based energy reweighting

The design and functionality of the **ATLAS ECAL** was described in Section 2.2.3 as well as the process from which electrons and photons deposit their energies in the **ECAL**: pair creation and bremsstrahlung radiation. Then, from these energy depositions in the **ECAL** the **SSs** are built and used for photon identification. However, the fact that **MC** and data **SSs** do not match, means that the energy depositions are different between these two, leading to a lower-level disagreement.

Although the Fudge Factor (**FF**) method described before led to an excellent improvement on the agreement between data and **MC** distributions, it is still based on modifying high-level variables and all independently of each other. On the other hand, by directly correcting the cells' energy depositions in the simulation, a simultaneous fix to all the Shower Shapes (**SSs**) and any other variable computed from the energies would be acquired. This processes is what is known as cell-based energy reweighting.

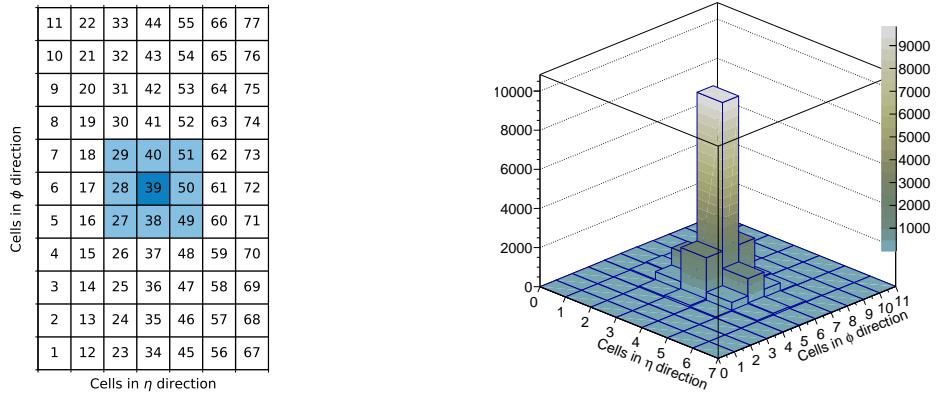
The cell-based reweighting approach has been developed and tested initialy for electrons [69], and further tested for photons [70]. For the electron's case, results have been very promising, where the second-layer **SSs** were substantially corrected. However, for photons, the same method that was used for electrons struggled, only working on average. Another approach to correct the simulation was based on matching data and simulated events, tested only on pseudo-data and technically complicated, but leading to better improvements [70].

In the current section, a new way of correcting the cell energies in MC is studied, using only the second layer of the ECAL, for simplicity. The method shares similarities with the FF method, making it easy to understand. First, the special event selection used for this study is presented. An overview of the old method to correct energies is briefly discussed, and then a study on how this method is improved is presented.

5.2.1 Event selection

The studies presented in this section are carried out with the same dataset as used for the FF calculation, described in Section 5.1.1. However, in this case, only the RZ samples are used. Events are selected as described in Section 4.2.1, using loose-isolated photons. Nevertheless, given that these studies rely on information on the second layer of the ECAL, special selection on the cells needs to be taken into account.

When an electron or photon enters the calorimeter, its footprint in the second layer is a visible cluster of cells surrounding the most energetic and central one (also referred as *hottest cell*). Clusters of 7×11 cells in $\eta \times \phi$ are considered, shown in Figure 5.8a with the current cell arrangement used. Approximately, 90% of the energy of the cluster is shared amongst the 9 central cells, which are highlighted in blue in Figure 5.8a, and the average normalised energy for data is shown in Figure 5.8b, visualising how the energy is distributed.



(a) Cell arrangement showing the cell number. The hottest cell is cell number 39, and the neighbouring ones are highlighted in light blue.

(b) Average energy of cells in 7×11 clusters for data.

Figure 5.8: Cells arrangement and average energy distribution amongst the clusters.

In this work, only events in which the clusters have the total of 77 cells are considered. Also, the events are required to have the central cell to be the most energetic.

5.2.2 Calculation

Early developments

All events that pass the selection will have a cluster associated to it, each of one having N cells and each cell has an energy E_i , for $i = 1, \dots, N$. For each event, in the first place, cluster energies are obtained as

$$E = \sum_{i=1}^N E_i.$$

which then are used to compute normalized cell energies $e_i = E_i/E$. Distributions of these normalized energies are then obtained when considering all the events that pass selection, and their means are used to compute the corrections at each cell i :

$$\Delta_i = \overline{\left(\frac{E_i^{\text{data}}}{E^{\text{data}}} \right)} - \overline{\left(\frac{E_i^{\text{MC}}}{E^{\text{MC}}} \right)} = \bar{e}_i^{\text{data}} - \bar{e}_i^{\text{MC}} \quad (5.10)$$

where $E^{\text{data/MC}}$ are the cluster energies for data and MC and, as mentioned above, the mean is done over all events that pass the selection. By definition these corrections coefficients sum to 0 over the whole cluster:

$$\sum_i \Delta_i = \sum_i \overline{\left(\frac{E_i^{\text{data}}}{E^{\text{data}}} \right)} - \sum_i \overline{\left(\frac{E_i^{\text{MC}}}{E^{\text{MC}}} \right)} = \sum_i \frac{E_i^{\text{data}}}{E^{\text{data}}} - \sum_i \frac{E_i^{\text{MC}}}{E^{\text{MC}}} = 1 - 1 = 0$$

Then a cell energy is corrected as:

$$E_i^{\text{MC-RW}} = E_i^{\text{MC}} + \Delta_i E^{\text{MC}}, \quad (5.11)$$

which is translated into shifting the cell energy divided by the cluster energy E^{MC} (from now on called normalized cell energy, e_i^{MC}) by an amount of Δ , so data and MC distributions' means match. For the early studies on photons, reweights were calculated separately for unconverted and converted photons, and they were also binned in $|\eta|$:

$$|\eta| : [0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.3, 1.37, 1.52, 1.6, 1.8, 2.0, 2.2, 2.37]$$

Because the corrections coefficients sum to zero, this method also implies that the cluster energy remains constant throughout the reweighting procedure:

$$E^{\text{MC-RW}} \equiv \sum_i E_i^{\text{MC-RW}} = \sum_i E_i^{\text{MC}} + \sum_i \Delta_i E^{\text{MC}} = E^{\text{MC}} + E^{\text{MC}} \sum_i \Delta_i = E^{\text{MC}}$$

The corrections can be visualized as 2-dimensional matrices, as seen in Figure 5.9, and therefore the reweighting procedure is developed using 2-dimensional histograms. For the case of working with the calorimeter layer 2, clusters of 77 cells ($\eta \times \phi = 7 \times 11$) are used.

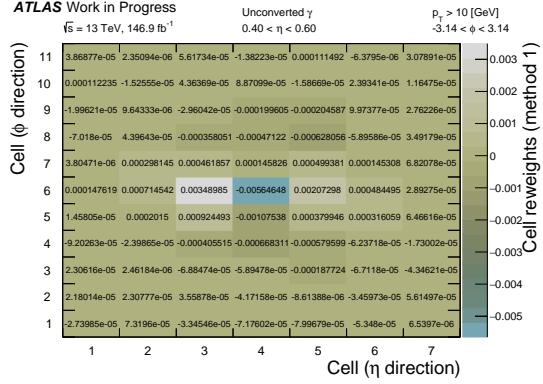


Figure 5.9: Energy reweights to be applied to MC cells, for unconverted photons with $0.4 < |\eta| < 0.6$. [fix luminosity](#)

The SSs computed using the second-layer information are R_η , R_ϕ and $w_{\eta 2}$, which can be calculated from energy deposits from the relations:

$$R_\eta = \frac{E_{3 \times 7}}{E_{7 \times 7}}$$

$$R_\phi = \frac{E_{3 \times 3}}{E_{7 \times 3}}$$

$$w_{\eta 2} = \sqrt{\frac{\sum_i E_i \eta_i^2}{\sum_i E_i} - \left(\frac{\sum_i E_i \eta_i}{\sum_i E_i} \right)^2}$$

where $E_{i \times j}$ is the summed cell energy in a region $\eta \times \phi = i \times j$ cells around the central cell. It was shown in the previous studies [70] that this method only corrects the shower shapes in average, but differences on the shape remain. This is due to the fact that this method only corrects the means of the energies distributions, by redistributing the energies amongst the cells. However, these energy distributions still present differences, specially regarding the shapes, leading to a very similar situation to what has been seen for the FFs. In this manner, a very similar approach of further correcting the means and widths of the normalised energies distributions can be employed.

New re-weighting method

This new method aims to correct both the mean and variance of normalized cell energies distributions by applying shifts and stretches. Similar to the approach followed for SSs, a first approximation for shift and stretch values of the energy distributions is to compute means and root mean squares (RMS) of them at each cell, respectively. Then, reweighted normalized cell energies can be obtained as

$$e_i^{\text{MC-RW}} = \frac{\text{RMS}_{e,i}^{\text{data}}}{\text{RMS}_{e,i}^{\text{MC}}} \left(e_i^{\text{MC}} - \bar{e}_i^{\text{MC}} \right) + \bar{e}_i^{\text{data}}, \quad (5.12)$$

where the subindex e indicates that the RMS is calculated from normalized cell energy distributions and i runs over all cells in the cluster.

Since the normalized cell energy at cell i can be calculated as $e_i^j = E_i^j / E^j$, for $j = \text{MC-RW}, \text{MC}$ and data, and it is required to have the same cluster energy after the reweighting procedure

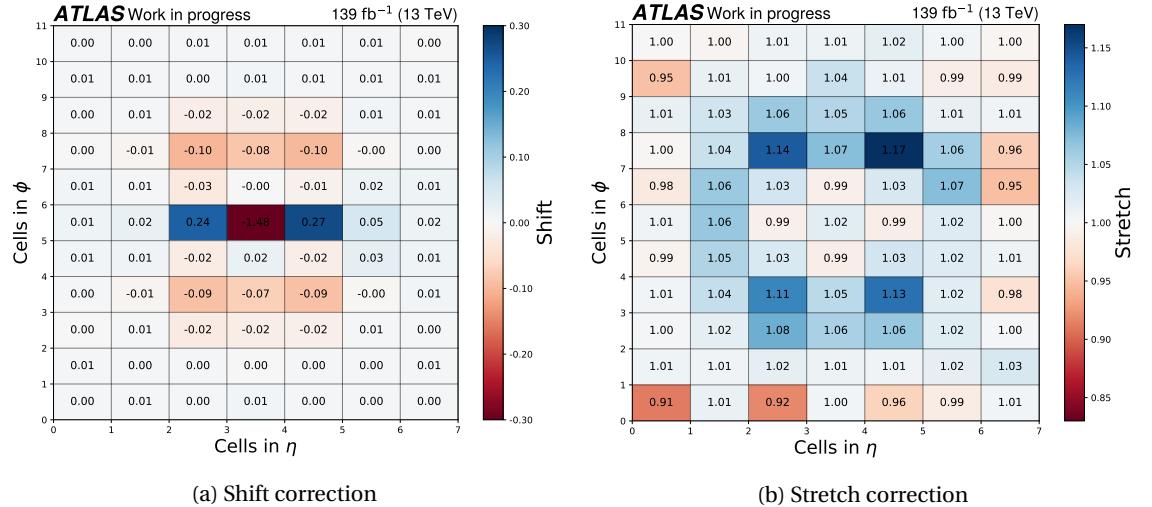
$(E_i^{\text{MC-RW}} = E_i^{\text{MC}})$. Multiplying Eq. 5.12 by $E_i^{\text{MC-RW}}$ it is possible to arrive at an expression for $E_i^{\text{MC-RW}}$:

$$E_i^{\text{MC-RW}} = \frac{\text{RMS}_{e,i}^{\text{data}}}{\text{RMS}_{e,i}^{\text{MC}}} E_i^{\text{MC}} + \left(\bar{e}_i^{\text{data}} - \frac{\text{RMS}_{e,i}^{\text{data}}}{\text{RMS}_{e,i}^{\text{MC}}} \bar{e}_i^{\text{MC}} \right) E_i^{\text{MC}}. \quad (5.13)$$

Comparing Eq. 5.13 to Eq. 5.11, it is observed that E_i^{MC} is now multiplied by a stretch factor, and a similar shift is applied to the normalized cell energy distribution but with modifications introduced by the stretch:

$$\begin{aligned} \text{stretch: } & 1 \rightarrow \frac{\text{RMS}_{e,i}^{\text{data}}}{\text{RMS}_{e,i}^{\text{MC}}} \\ \text{shift: } & \bar{e}_i^{\text{data}} - \bar{e}_i^{\text{MC}} \rightarrow \bar{e}_i^{\text{data}} - \frac{\text{RMS}_{e,i}^{\text{data}}}{\text{RMS}_{e,i}^{\text{MC}}} \bar{e}_i^{\text{MC}} \end{aligned}$$

Finally, to guarantee that the cluster energy remains constant, cell energies are re-scaled by $\sum_i E_i^{\text{MC}} / \sum_i E_i^{\text{MC-RW}}$. As a result of this procedure, two reweights matrices are needed, and one example of them is presented in Figure 5.10. For what follows, this new method is applied to correct the cell energies, and it is computed inclusively in p_T and $|\eta|$, only separating between unconverted and converted photons.



Using these correction factors to the normalised energies to each cell, in Figure 5.11, the resulting cell energy distributions are displayed for cells 28, 39 and 50². The new reweighting method achieves greats improvements on the agreement between data and MC. The reweighting method does well correcting tails of the distributions for all cells, as well as the peaks of them, which can be seen specially for cell 28.

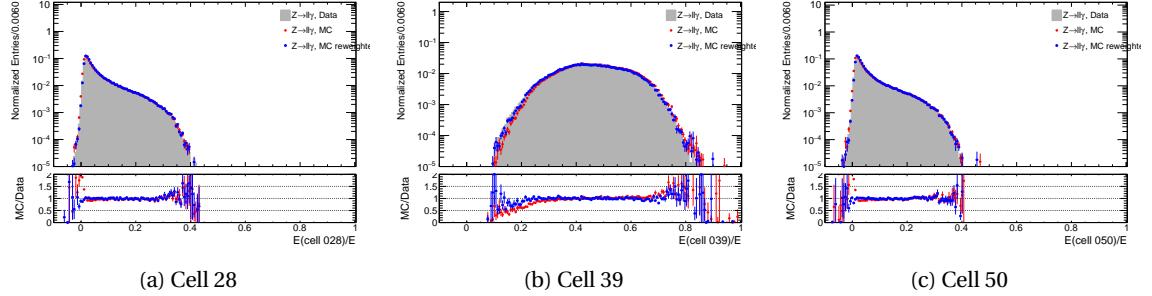


Figure 5.11: Normalized cell energy distributions for cells 28, 39 and 50 of the cluster using unconverted photons. Blue and red points correspond to the reweighted and original MC distributions, respectively, while the grey histogram represents data.

To assess the behaviour of the reweighting procedure applied to the SSs of the second ECAL layer, in Figure 5.12 the comparison of the methods for correcting the R_η , R_ϕ and $w_{\eta 2}$ SSs are shown. In the three cases, an improvement over the uncorrected MC is observed, specially for the R_ϕ and $w_{\eta 2}$ variables, but the cell-based energy reweighting method for unconverted photons does not reach the level of agreement with data as the FFs, which has been shown to provide an excellent match with data. Nevertheless, there is almost no difference seen between the reweighting and the FF method for converted photons, indicating that there is room for improvement on the corrections.

5.3 Conclusions and future work

In the current chapter two methods to correct the observed disagreement between data and simulated Shower Shapes (SSs) were studied.

The Fudge Factor (FF) method has been historically used in the collaboration, which in the beginnings it was only based on simple shifts of the distributions by adding a constant term to the variable. Even though the corrections led to good improvements and therefore obtaining Scale Factors (SFs) closer to one, notable shape differences remained between data and simulation. In the context of this work, by adding a linear term to the variable transformation, the widths of the MC distributions are fixed leading to even better agreement. This new method of correcting the SSs using FFs is referred as the shift+stretch method and is now used throughout the ATLAS collaboration.

A novel and lower-level method that aims to modify the energies in the ECAL cells has also been developed. By using the energy distributions at each cell in clusters around the most

² As shown in Figure 5.8a, cell number 39 is the central one, while cell 28 and 50 are to the left and right, respectively, in the η direction

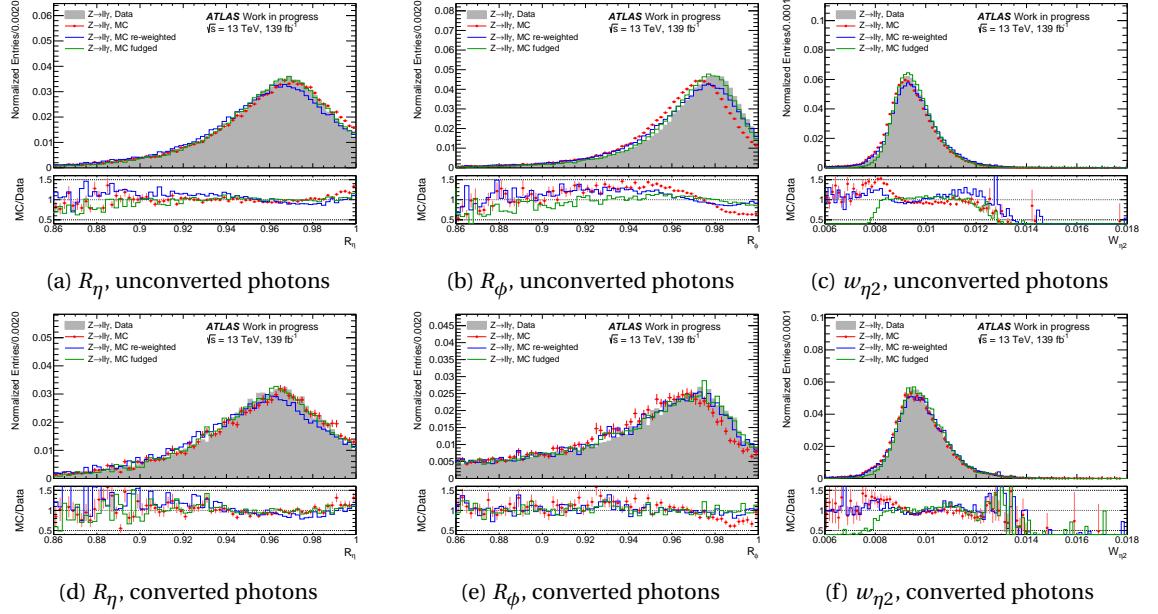


Figure 5.12: SSs distributions for unconverted (top row) and converted (bottom row) photons in $|\eta| < 0.6$. Data is represented by the grey histogram. Uncorrected MC simulation is shown with the red points and reweighted (fudged) MC with the blue (green) line.

energetic cell, it is possible to correct all the SSs at once. This method applies the same strategy of shifting and stretching the normalised energy from MC simulation to match the distribution found in data. Although the method is new and still needs polishing and its extension to other layers of the ECAL, it led to promising results in which some variables are corrected in the same way as with FFs. The cell-based energy reweighting method shows a great potential in the collaboration, not only in the context of *offline* photon identification, but also at trigger level.

5.3.1 Future work

One of the most exciting and promising approaches to correct the SSs is the cell-based method. As previously mentioned, this approach could be employed on different steps on the photon identification process, such as at trigger level, or event *offline* to correct all the SSs simultaneously. Another potential, and important usage, is to use the corrected clusters to directly compute photon identification, for example as considering the clusters as images and using a Convolutional Neural Network (CNN) to perform the photon identification [70].

Shower Shape Variables (SSVs) can be easily interpreted in terms of physical terms. For this reason, maintaining these physical quantities serve to understand the underlying physics of the processes. To continue correcting these variables is of great interest and there are several ways in which these can be achieved. The current method of transforming the variable but using higher-order terms remains a challenging task, but yet not explored. By making use of the novel Machine Learning (ML) techniques, it is possible to obtain correction factors for higher-order terms in the expansion, further correcting higher-order momenta of the distributions (skewness, kurtosis, etc.). Other interesting approach is using multivariate reweighting, which was explored in Ref. [71], showing very promising results.

Part IV

New Physics

ANALYSIS MOTIVATION AND STRATEGY

6

“Champions keep playing until they get it right.”

Billie Jean King

yet another template (yat)

SIGNAL AND BACKGROUND SAMPLES

7

“Champions keep playing until they get it right.”

Billie Jean King

yet another template (yat)

EVENT SELECTION

8

“Champions keep playing until they get it right.”

Billie Jean King

yet another template (yat)

BACKGROUND ESTIMATION

9

“Champions keep playing until they get it right.”

Billie Jean King

yet another template (yat)

BACKGROUND MODELING**10**

*“Champions keep playing until they
get it right.”*

Billie Jean King

yet another template (yat)

SYSTEMATIC UNCERTAINTIES**11**

*“Champions keep playing until they
get it right.”*

Billie Jean King

yet another template (yat)

STATISTICAL ANALYSIS

12

“Champions keep playing until they get it right.”

Billie Jean King

yet another template (yat)

RESULTS

13

“Champions keep playing until they get it right.”

Billie Jean King

yet another template (yat)

CONCLUSIONS

Don't let anyone rob you of your imagination, your creativity, or your curiosity.

Mae Jemison

ELLIPSE FITTING FORMULAE

A

The ellipse fit to obtain the statistical uncertainties on the shift and stretch Fudge Factors (FFs) is implemented following the procedure suggested in Ref. [72]. The fit returns the set of parameters $\{A, B, C, D, E, F\}$ that parametrise the conic:

$$F(x, y) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$$

with $B^2 - 4AC < 0$ for ellipses. The variables x, y shown are general, but in the case of FFs they represent the shift and stretch parameters, respectively.

$$\frac{((x - x_0) \cos \theta + (y - y_0) \sin \theta)^2}{a^2} + \frac{((x - x_0) \sin \theta - (y - y_0) \cos \theta)^2}{b^2} = 1, \quad (\text{A.1})$$

it is possible to extract the ellipse center (x_0, y_0) , its tilt angle θ and its semi-major and semi-minor axis, a and b , respectively. Then, the desired uncertainties, and the correlation between x and y can be extracted using the following relations (see Figure A.1):

$$\sigma_x = \sqrt{a^2 \cos^2 \theta + b^2 \sin^2 \theta} \quad (\text{A.2})$$

$$\sigma_y = \sqrt{a^2 \sin^2 \theta + b^2 \cos^2 \theta} \quad (\text{A.3})$$

$$\rho = \tan(2\theta) \frac{\sigma_x^2 - \sigma_y^2}{2\sigma_x \sigma_y}. \quad (\text{A.4})$$

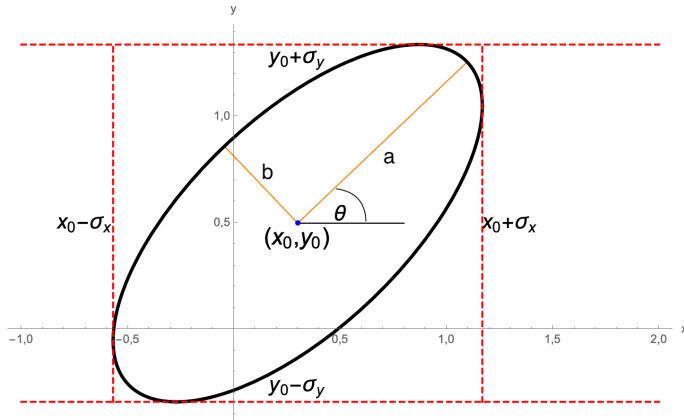


Figure A.1: Parameters of the most general case of an ellipse.

GLOSSARY

CERN European Organization for Nuclear Research	6
ALICE A Large Ion Collider Experiment	
ATLAS A Toroidal LHC ApparatuS	ix
CMS Compact Muon Solenoid	
LHC Large Hadron Collider	ix
LHCb Large Hadron Collider beauty	
LHCf Large Hadron Collider forward	
LS2 Long Shut down 2	ix
TOTEM TOTal cross section, Elastic scattering and diffraction dissociation Measurement at the LHC	
MoEDAL Monopole & Exotics Detector At the LHC	
SPS Super Proton Synchrotron	6
L1 Level-1	16
L1Topo Level-1 Topological	16
LAr Liquid Argon	13
ECAL Electromagnetic Calorimeter	ix
CSC Cathode Strip Chamber	15
CTP Central Trigger Processor	17
FCAL Forward Calorimeter	13
HEC Hadronic End-Cap Calorimeter	14
HCAL Hadronic Calorimeter	ix
HLT High Level Trigger	16
IBL Insertable B-Layer	10
ID Inner Detector	ix

Glossary	72
SCT SemiConductor Tracker	10
RPC Resistive-Plate Chamber	15
TGC Thin-Gap Chamber	16
TRT Transition Radiation Tracker	10
NSW New Small Wheel	15
MDT Monitored Drift Tube	15
MS Muon Spectrometer	ix
GRL Good Runs List	18
LB Luminosity-block	18
PSB Proton Synchrotron Booster	6
MC Monte Carlo	ix
MV Multivariate	40
KDE Kernel Density Estimator	x
ROI Region of Interest	16
FF Fudge Factor	vii
SF Scale Factor	x
SS Shower Shape	x
SSV Shower Shape Variable	23
WP Working Point	ix
PDF Probability Density Function	x
EM electromagnetic	34
FSR Final State Radiation	x
RZ Radiative Z	x
SP Single Photon	x
ISR Initial State Radiation	x
ME Matrix Element	26
PS Parton Shower	6
QCD Quantum Chromodynamics	26
SM Standard Model	38
JVT Jet vertex Tagger	29
JVF Jet vertex Fraction	29
PV Primary Vertex	22
CB Combined Muon	26

ST	Segmented Muon	26
CT	Calorimetric Muon	26
ME	Standalone/Extrapolated Muon	26
PFlow	Particle Flow	28
FTAG	Flavor Tagging	30

BIBLIOGRAPHY

- [1] O. S. Brüning, J. Poole, P. Collier, P. Lebrun, R. Ostojic, S. Myers, and P. Proudlock, *LHC Design Report*, CERN, Geneva **1** (2004) 548. <https://cds.cern.ch/record/782076>. 6
- [2] L. Evans and P. Bryant, *LHC Machine*, *Journal of Instrumentation* **3** no. 08, (2008) S08001. <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08001><https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08001/meta>. 6
- [3] E. Lopienska, *The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022*, <https://cds.cern.ch/record/2800984>. ix, 7
- [4] The ALICE Collaboration, *The ALICE experiment at the CERN LHC*, *Journal of Instrumentation* **3** no. 08, (2008) S08002. <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08002>. 7
- [5] The LHCb Collaboration, *The LHCb Detector at the LHC*, *Journal of Instrumentation* **3** no. 08, (2008) S08005. <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08005><https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08005/meta>. 7
- [6] The CMS Collaboration, *The CMS experiment at the CERN LHC*, *Journal of Instrumentation* **3** no. 08, (2008) S08004. <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08004><https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08004/meta>. 7
- [7] The ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *Journal of Instrumentation* **3** no. 08, (2008) S08003. <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08003>. ix, xii, 7, 9, 13, 14, 15
- [8] The LHCf Collaboration, *The LHCf detector at the CERN Large Hadron Collider*, *Journal of Instrumentation* **3** no. 08, (2008) S08006. <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08006><https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08006/meta>. 7

- [9] The TOTEM Collaboration, *The TOTEM Experiment at the CERN Large Hadron Collider*, *Journal of Instrumentation* **3** no. 08, (2008) S08007.
<https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08007>
<https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08007/meta>. 7
- [10] MoEDAL Collaboration, *Technical Design Report of the MoEDAL Experiment*. 7
- [11] R. M. Bianchi and A. Collaboration, “ATLAS experiment schematic or layout illustration.” 2022. ix, 8, 9
- [12] J. Pequenao, “Computer generated image of the ATLAS inner detector.” 2008. 10
- [13] J. Pequenao, “Computer generated image of the ATLAS inner detector.” 2008. 10
- [14] The ATLAS Collaboration, *Inner Detector Forward SCT Barrel SCT TRT Pixel Detectors Technical Design Report*, tech. rep., 1997.
<https://cds.cern.ch/record/331063/files/ATLAS-TDR-4-Volume-I.pdf>. 10
- [15] T. A. Collaboration, *ATLAS Insertable B-Layer Technical Design Report Addendum*, Tech. Rep. May, Sep, 2012. <https://cds.cern.ch/record/1291633>
<https://cdsweb.cern.ch/record/1451888>. 10
- [16] K. Potamianos, *The upgraded Pixel detector and the commissioning of the Inner Detector tracking of the ATLAS experiment for Run-2 at the Large Hadron Collider*, [arXiv:1608.07850](https://arxiv.org/abs/1608.07850). <https://arxiv.org/abs/1608.07850v1>. 10
- [17] The ATLAS Collaboration, *ATLAS pixel detector electronics and sensors*, *Journal of Instrumentation* **3** no. 07, (2008) P07007.
<https://iopscience.iop.org/article/10.1088/1748-0221/3/07/P07007>
<https://iopscience.iop.org/article/10.1088/1748-0221/3/07/P07007/meta>. 11
- [18] T. Heim, *Status and performance of the ATLAS Pixel Detector after 3 years of operation*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **765** (2014) 227–231. 11
- [19] A. Ahmad, Z. Albrechtskirchinger, P. P. Allport, J. Alonso, et al., *The silicon microstrip sensors of the ATLAS semiconductor tracker*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **578** no. 1, (2007) 98–118. 11
- [20] The ATLAS Collaboration, *The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: design and performance*, *Journal of Instrumentation* **3** no. 02, (2008) P02013.
<https://iopscience.iop.org/article/10.1088/1748-0221/3/02/P02013>
<https://iopscience.iop.org/article/10.1088/1748-0221/3/02/P02013/meta>. 11
- [21] J. Pequenao, “Computer Generated image of the ATLAS calorimeter.” 2008. ix, 12

- [22] V. Rossetti, *Performance of the ATLAS Calorimeters and Commissioning for LHC Run-2,* <https://cds.cern.ch/record/2037117>. 12
- [23] The ATLAS Collaboration, *ATLAS tile calorimeter: Technical Design Report,* CERN/LHCC96-42(1996) (1996). <https://cds.cern.ch/record/331062>. 14
- [24] *ATLAS muon spectrometer : Technical Design Report,* <https://cds.cern.ch/record/331068>. 15
- [25] B. Stelzer, *The New Small Wheel Upgrade Project of the ATLAS Experiment,* Nuclear and Particle Physics Proceedings **273-275** (2016) 1160–1165. <https://www.sciencedirect.com/science/article/pii/S2405601415006719>. 15
- [26] The ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2010,* The European Physical Journal C **72** no. 1, (2012) 1849. 16
- [27] A. ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2015,* [arXiv:1611.09661](https://arxiv.org/abs/1611.09661). <http://arxiv.org/abs/1611.09661><http://dx.doi.org/10.1140/epjc/s10052-017-4852-3>. 16
- [28] The ATLAS Collaboration, *Operation of the ATLAS trigger system in Run 2,* *Journal of Instrumentation* **15** no. 10, (2020), [arXiv:2007.12539v2](https://arxiv.org/abs/2007.12539v2). 16
- [29] R. Achenbach, P. Adragna, V. Andrei, P. Apostologlou, B. Åsman, C. Ay, B. M. Barnett, B. Bauss, et al., *The ATLAS Level-1 Calorimeter Trigger,* *Journal of Instrumentation* **3** no. 03, (2008) P03001. <https://dx.doi.org/10.1088/1748-0221/3/03/P03001>. 16
- [30] ATLAS Collaboration, P. Jenni, M. Nessi, M. Nordberg, and K. Smith, *ATLAS high-level trigger, data-acquisition and controls: Technical Design Report.* Technical design report. ATLAS. CERN, Geneva, 2003. <https://cds.cern.ch/record/616089>. 17
- [31] The ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC,* The European Physical Journal C **83** no. 10, (2023) 982. ix, 18
- [32] G. Avoni, M. Bruschi, G. Cabras, D. Caforio, N. Dehghanian, A. Floderus, B. Giacobbe, F. Giannuzzi, et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS,* *Journal of Instrumentation* **13** no. 07, (2018) P07017. <https://dx.doi.org/10.1088/1748-0221/13/07/P07017>. 18
- [33] ATLAS Collaboration, *Preliminary analysis of the luminosity calibration of the ATLAS 13.6 TeV data recorded in 2022,* tech. rep., CERN, Geneva, 2023. <https://cds.cern.ch/record/2853525>. 18
- [34] ATLAS Collaboration, *Preliminary analysis of the luminosity calibration for the ATLAS 13.6 TeV data recorded in 2023,* tech. rep., CERN, Geneva, 2024. <https://cds.cern.ch/record/2900949>. 18

- [35] The ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, The European Physical Journal C **77** no. 10, (2017) 673. [21](#)
- [36] T. G. Cornelissen, N. Van Eldik, M. Elsing, W. Liebig, E. Moyse, N. Piacquadio, K. Prokofiev, A. Salzburger, and A. Wildauer, *Updates of the ATLAS Tracking Event Data Model (Release 13)*, tech. rep., CERN, Geneva, 2007. <https://cds.cern.ch/record/1038095>. [ix](#), [21](#)
- [37] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig, E. Moyse, and A. Salzburger, *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, tech. rep., CERN, Geneva, 2007. <https://cds.cern.ch/record/1020106>. [21](#)
- [38] T. A. collaboration, *A neural network clustering algorithm for the ATLAS silicon pixel detector*, Journal of Instrumentation **9** no. 09, (2014) P09009. <https://dx.doi.org/10.1088/1748-0221/9/09/P09009>. [21](#)
- [39] The ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, The European Physical Journal C **77** no. 5, (2017) 332. [22](#)
- [40] *Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13$ TeV*, tech. rep., CERN, Geneva, 2015. <https://cds.cern.ch/record/2037717>. [22](#)
- [41] The ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data*, Journal of Instrumentation **14** no. 12, (2019) P12006. <https://dx.doi.org/10.1088/1748-0221/14/12/P12006>. [ix](#), [22](#), [23](#), [41](#), [43](#)
- [42] ATLAS Collaboration, *Electron and photon reconstruction and performance in ATLAS using a dynamical, topological cell clustering-based approach*, tech. rep., CERN, Geneva, 2017. <https://cds.cern.ch/record/2298955>. [22](#), [27](#)
- [43] The ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, The European Physical Journal C **77** no. 7, (2017) 490. [22](#), [27](#)
- [44] The ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}=13$ TeV*, The European Physical Journal C **76** no. 5, (2016) 292. [26](#)
- [45] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- k_T jet clustering algorithm*, Journal of High Energy Physics **2008** no. 04, (2008) 063. <https://dx.doi.org/10.1088/1126-6708/2008/04/063>. [ix](#), [27](#), [28](#)
- [46] The ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, The European Physical Journal C **77** no. 7, (2017) 466. [29](#)
- [47] ATLAS Collaboration, *E_T^{miss} performance in the ATLAS detector using 2015-2016 LHC p-p collisions*, tech. rep., CERN, Geneva, 2018. <https://cds.cern.ch/record/2625233>. [29](#)

- [48] The ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, The European Physical Journal C **81** no. 8, (2021) 689. [ix](#), [29](#)
- [49] *Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data*, tech. rep., CERN, Geneva, 2014. <https://cds.cern.ch/record/1741020>. [30](#)
- [50] ATLAS Collaboration, *Measurement of the b-tag Efficiency in a Sample of Jets Containing Muons with 5 fb^{-1} of Data from the ATLAS Detector*, tech. rep., CERN, Geneva, 2012. <https://cds.cern.ch/record/1435197>. [30](#)
- [51] A. Lavrenov, *MV2-algorithm's clones*, CoRR **cs.CR/0208003** (2002). <https://arxiv.org/abs/cs/0208003>. [30](#)
- [52] M. Lanfermann, *Deep Learning in Flavour Tagging at the ATLAS experiment*, PoS **EPS-HEP2017** (2018) 764. [30](#)
- [53] The ATLAS Collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset*, The European Physical Journal C **83** no. 7, (2023) 681. [30](#)
- [54] ATLAS Collaboration, *Graph Neural Network Jet Flavour Tagging with the ATLAS Detector*, tech. rep., CERN, Geneva, 2022. <https://cds.cern.ch/record/2811135>. [30](#)
- [55] S. Brody, U. Alon, and E. Yahav, *How Attentive are Graph Attention Networks?*, CoRR **abs/2105.14491** (2021), [2105.14491](#). <https://arxiv.org/abs/2105.14491>. [30](#)
- [56] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, *Attention Is All You Need*, CoRR **abs/1706.03762** (2017), [1706.03762](#). <http://arxiv.org/abs/1706.03762>. [30](#)
- [57] Z. Wu, “b-jets calibration in atlas.” ATLAS-CMS Flavour Tagging Workshop, 2024. [ix](#), [31](#)
- [58] I. Wingerter-Seez, *Particle Physics Instrumentation*, arXiv:1804.11246. <https://cds.cern.ch/record/2315747>. [x](#), [37](#)
- [59] ATLAS Collaboration, *Search for short- and long-lived axion-like particles in $H \rightarrow aa \rightarrow 4\gamma$ decays with the ATLAS experiment at the LHC*, The European Physical Journal C **84** no. 7, (2024) 742. <https://doi.org/10.1140/epjc/s10052-024-12979-0>. [38](#)
- [60] The ATLAS Collaboration, *Electron and photon efficiencies in LHC Run 2 with the ATLAS experiment*, Journal of High Energy Physics **2024** no. 5, (2024) 162. [40](#)
- [61] The ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016*, The European Physical Journal C **79** no. 3, (2019) 205. [x](#), [41](#), [42](#), [43](#)
- [62] L. Devroye, *Non-Uniform Random Variate Generation*. Springer New York, NY, 1986. [41](#)

- [63] L. Lyons, D. Gibaut, and P. Clifford, *How to Combine Correlated Estimates of a Single Physical Quantity*, *Nucl. Instrum. Meth.* **A270** (1988) 110. 42
- [64] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton–proton collision data*, *JINST* **14** (2019) P03017, [arXiv:1812.03848 \[hep-ex\]](https://arxiv.org/abs/1812.03848). x, 43
- [65] E. Bothmann, G. S. Chahal, S. Höche, J. Krause, F. Krauss, S. Kuttimalai, S. Liebschner, D. Napoletano, M. Schönher, H. Schulz, S. Schumann, and F. Siegert, *Event generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) 034. <https://scipost.org/10.21468/SciPostPhys.7.3.034>. 45
- [66] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, *Computer Physics Communications* **178** no. 11, (2008) 852–867. <https://www.sciencedirect.com/science/article/pii/S0010465508000441>. 45
- [67] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, F. Behner, L. Bellagamba, J. Boudreau, L. Broglia, A. Brunengo, H. Burkhardt, S. Chauvie, J. Chuma, R. Chytracek, G. Cooperman, G. Cosmo, P. Degtyarenko, A. Dell’Acqua, G. Depaola, D. Dietrich, R. Enami, A. Feliciello, C. Ferguson, H. Fesefeldt, G. Folger, F. Foppiano, A. Forti, S. Garelli, S. Giani, R. Giannitrapani, D. Gibin, J. Gómez Cadenas, I. González, G. Gracia Abril, G. Greeniaus, W. Greiner, V. Grichine, A. Grossheim, S. Guatelli, P. Gumplinger, R. Hamatsu, K. Hashimoto, H. Hasui, A. Heikkinen, A. Howard, V. Ivanchenko, A. Johnson, F. Jones, J. Kallenbach, N. Kanaya, M. Kawabata, Y. Kawabata, M. Kawaguti, S. Kelner, P. Kent, A. Kimura, T. Kodama, R. Kokoulin, M. Kossov, H. Kurashige, E. Lamanna, T. Lampén, V. Lara, V. Lefebure, F. Lei, M. Liendl, W. Lockman, F. Longo, S. Magni, M. Maire, E. Medernach, K. Minamimoto, P. Mora de Freitas, Y. Morita, K. Murakami, M. Nagamatu, R. Nartallo, P. Nieminen, T. Nishimura, K. Ohtsubo, M. Okamura, S. O’Neale, Y. Oohata, K. Paech, J. Perl, A. Pfeiffer, M. Pia, F. Ranjard, A. Rybin, S. Sadilov, E. Di Salvo, G. Santin, T. Sasaki, N. Savvas, Y. Sawada, S. Scherer, S. Sei, V. Sirotenko, D. Smith, N. Starkov, H. Stoecker, J. Sulkimo, M. Takahata, S. Tanaka, E. Tcherniaev, E. Safai Tehrani, M. Tropeano, P. Truscott, H. Uno, L. Urban, P. Urban, M. Verderi, A. Walkden, W. Wander, H. Weber, J. Wellisch, T. Wenaus, D. Williams, D. Wright, T. Yamada, H. Yoshida, and D. Zschiesche, *Geant4—a simulation toolkit*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** no. 3, (2003) 250–303. <https://www.sciencedirect.com/science/article/pii/S0168900203013688>. 45
- [68] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, et al., *TMVA - Toolkit for Multivariate Data Analysis*, 2007. [arXiv:physics/0703039 \[physics.data-an\]](https://arxiv.org/abs/physics/0703039). 46
- [69] M. Khandoga, *Calibration of electron shower shapes, hadronic recoil reconstruction using deep learning algorithms and the measurement of W boson transverse momentum distribution with the ATLAS detector*, 2020. <https://cds.cern.ch/record/2752635>. 53

- [70] M. Belfkir, *Search for Higgs pair production at LHC collider (CERN): The first measurement for Higgs potential and search for new physics*, 2021.
<https://cds.cern.ch/record/2792268>. Presented 15 Sep 2021. 53, 56, 59
- [71] J. L. Spah, *Data-driven corrections to shower shape variables for photon identification at the ATLAS experiment*, 2021. <https://kroeninger-group.physik.tu-dortmund.de/en/research/particle-physics/theses/#c135884>. 59
- [72] R. H. oy and J. Flusser, *Numerically Stable Direct Least Squares Fitting of Ellipses*, 1998.
<https://api.semanticscholar.org/CorpusID:15772208>. 70