



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

---

Trabajo de Tesis Doctoral

## The excitement to search the light

*Photon variables corrections and Search  
for photon+jet high-mass resonances*

---

Autor:  
Francisco SILI

Directores:  
Prof. Dr. Maria Teresa DOVA  
Dr. Francisco ALONSO

8th November 2024



## **ACKNOWLEDGEMENTS**

---

akwnoelasdfa

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,*  
*8th November 2024*

---

Francisco SILI

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

## TESIS DE DOCTORADO

---

The excitement to search the light

---

by Francisco SILI

## ABSTRACT

some text

# CONTENTS

|  |            |
|--|------------|
| <b>Contents</b>  | <b>vi</b>  |
| <b>List of Figures</b>   | <b>x</b>   |
| <b>List of Tables</b>  | <b>xix</b> |
| <b>Introduction</b>  | <b>1</b>   |
| <br>   |            |
| <b>I Theory Motivation</b>   | <b>5</b>   |
| <br>   |            |
| <b>1 The Standard Model and Beyond</b>                               | <b>6</b>   |
| 1.1 The Standard Model (SM) . . . . .                                | 6          |
| 1.1.1 Elementary particles and their interactions . . . . .          | 7          |
| 1.1.2 Brief mathematical description of the Standard Model . . . . . | 8          |
| 1.1.3 Hadron interactions in proton-proton colliders . . . . .       | 14         |
| 1.1.4 Theory of prompt-photon production . . . . .                   | 16         |
| 1.2 Physics Beyond Standard Model (BSM) . . . . .                    | 20         |
| 1.2.1 Quark compositeness theories . . . . .                         | 21         |
| 1.2.2 Higher dimensional theories . . . . .                          | 22         |
| 1.3 Monte Carlo (MC) simulations . . . . .                           | 24         |
| 1.3.1 Hard interactions and parton shower . . . . .                  | 24         |
| 1.3.2 Hadronisation . . . . .  | 25         |
| 1.3.3 Underlying Event (UE) . . . . .                                | 26         |
| 1.3.4 Tunes . . . . .  | 26         |
| 1.3.5 ATLAS detector simulation . . . . .                            | 26         |
| <br>   |            |
| <b>II Experimental setup</b>   | <b>28</b>  |
| <br>   |            |
| <b>2 The ATLAS Experiment</b>  | <b>29</b>  |
| 2.1 LHC . . . . .  | 29         |

---

|  |    |
|--|----|
| <b>2.2 ATLAS</b>   | 31 |
| 2.2.1 ATLAS Coordinate system                                  | 32 |
| 2.2.2 Inner Detector   | 33 |
| 2.2.3 Calorimeters   | 35 |
| 2.2.4 Muon spectrometer  | 38 |
| 2.2.5 The Trigger System                                       | 39 |
| <b>2.3 Data-taking during Run-2</b>                            | 41 |
| <br>   |    |
| <b>3 Reconstruction and identification of physical objects</b> | 44 |
| 3.1 Track and vertex reconstruction                            | 44 |
| 3.2 Photons and electrons                                      | 46 |
| 3.2.1 Reconstruction   | 46 |
| 3.2.2 Identification   | 47 |
| 3.2.3 Isolation  | 48 |
| 3.3 Muons  | 50 |
| 3.4 Jets   | 50 |
| 3.4.1 Anti- $k_t$ jet clustering algorithm                     | 51 |
| 3.4.2 Calorimeter Jets   | 51 |
| 3.4.3 PFlow Jets   | 52 |
| 3.4.4 Jet calibration  | 53 |
| 3.5 Jet flavor tagging   | 54 |
| 3.5.1 $b$ -jet identification performance                      | 55 |
| 3.5.2 $c$ -jet identification performance                      | 56 |
| <br>   |    |
| <b>III Photon shower shape corrections</b>                     | 58 |
| <br>   |    |
| <b>4 Shower shapes and photon identification</b>               | 59 |
| 4.1 Shower shapes  | 59 |
| 4.2 Photon Identification                                      | 63 |
| 4.2.1 Processes and event selection                            | 63 |
| 4.2.2 Optimisation   | 65 |
| 4.2.3 Efficiency measurements                                  | 66 |
| 4.3 Shower shapes variables differences between data and MC    | 68 |
| <br>   |    |
| <b>5 Shower shapes corrections</b>                             | 70 |
| 5.1 Fudge Factors (FFs)  | 70 |
| 5.1.1 Data and simulated samples                               | 70 |
| 5.1.2 Calculation  | 71 |
| 5.1.3 Uncertainties  | 74 |
| 5.1.4 Results  | 75 |
| 5.2 Cell-based energy reweighting                              | 77 |
| 5.2.1 Event selection  | 79 |

---

|  |   |        |
|--|---|--------|
| 5.2.2  | Calculation . . . . .                               | 80     |
| 5.2.3  | Results . . . . .                                   | 82     |
| 5.3  | Conclusions and future work . . . . .               | 83     |
| 5.3.1  | Future work . . . . .                               | 84     |
| <br>IV Search for new physics in high-mass photon+jet final states |   | <br>86 |
| 6  | Analysis strategy and Statistical Treatment         | 87     |
| 6.1  | Strategy, signals and SM's backgrounds . . . . .    | 87     |
| 6.2  | Statistical Treatment . . . . .                     | 88     |
| 6.2.1  | Statistical Model . . . . .                         | 88     |
| 6.2.2  | Systematic uncertainties . . . . .                  | 89     |
| 6.2.3  | Simultaneous fits . . . . .                         | 90     |
| 6.2.4  | Hypothesis test . . . . .                           | 90     |
| 6.2.5  | The BumpHunter algorithm . . . . .                  | 92     |
| 6.3  | Fits and results . . . . .                          | 94     |
| 7  | Signal models and simulated samples                 | 95     |
| 7.1  | Signals . . . . .                                   | 95     |
| 7.1.1  | Excited quarks . . . . .                            | 95     |
| 7.1.2  | Quantum Black Holes . . . . .                       | 96     |
| 7.2  | Backgrounds . . . . .                               | 98     |
| 8  | Event selection and signal regions definitions      | 100    |
| 8.1  | Trigger . . . . .                                   | 100    |
| 8.2  | Preselection . . . . .                              | 101    |
| 8.2.1  | Objects . . . . .                                   | 101    |
| 8.2.2  | Overlap Removal . . . . .                           | 102    |
| 8.3  | Signal regions optimisation . . . . .               | 103    |
| 8.3.1  | Photon and jet angular selections . . . . .         | 104    |
| 8.3.2  | Extended isolation . . . . .                        | 106    |
| 8.3.3  | Jet $p_T$ . . . . .                                 | 108    |
| 8.4  | Signal regions . . . . .                            | 110    |
| 9  | Signal modeling and systematic uncertainties        | 113    |
| 9.1  | Signal modeling . . . . .                           | 113    |
| 9.2  | Acceptances and efficiencies . . . . .              | 115    |
| 9.3  | Systematic uncertainties . . . . .                  | 117    |
| 9.3.1  | Experimental uncertainties . . . . .                | 118    |
| 9.3.2  | Theory uncertainties . . . . .                      | 119    |
| 10   | Background estimate                                 | 121    |
| 10.1   | Background estimation: jet faking photons . . . . . | 121    |

---

|  |            |
|--|------------|
| 10.1.1 ABCD method . . . . .                     | 122        |
| 10.1.2 Corrections to the ABCD method . . . . .  | 124        |
| 10.1.3 Template fits procedure . . . . .         | 125        |
| 10.1.4 Results . . . . .                         | 127        |
| 10.2 Background modeling . . . . .               | 127        |
| 10.2.1 Fit functions . . . . .                   | 128        |
| 10.2.2 Datasets preparation . . . . .            | 129        |
| 10.2.3 Modeling strategy . . . . .               | 132        |
| 10.2.4 Fit strategy validation . . . . .         | 134        |
| 10.2.5 Summary of modeling strategies . . . . .  | 141        |
| <b>11 Results</b>                                | <b>144</b> |
| 11.1 Photon+jet invariant mass binning . . . . . | 144        |
| 11.2 Results . . . . .                           | 146        |
| 11.2.1 Background-only interpretation . . . . .  | 146        |
| 11.2.2 Signal interpretation . . . . .           | 150        |
| <b>Conclusions</b>                               | <b>159</b> |
| <b>A Ellipse fitting formulae</b>                | <b>160</b> |
| <b>B Signal injection results</b>                | <b>161</b> |
| <b>Glossary</b>                                  | <b>163</b> |
| <b>Bibliography</b>                              | <b>167</b> |

# LIST OF FIGURES

|     |  |    |
|-----|--|----|
| 1.1 | Overview of the particles of the SM. All fermions participate in the weak interaction, but only the quarks interact with gluons, whereas both quarks and charged leptons interact with via the electromagnetic (EM) force. Neutrinos, being neutral and colourless, only interact with the $W$ and $Z$ bosons via the weak force. Finally, the graviton, although it has not been discovered yet, should be the corresponding force carrier of the gravity force. Extracted from Ref. [2]. . . . . | 7  |
| 1.2 | Experimental measurements of the coupling constant of Quantum Chromodynamics (QCD) compared to the coupling computed at five loops [13]. . . .   | 13 |
| 1.3 | Parton momentum fraction $x$ times the unpolarized parton distributions $f_i(x, Q^2)$ (where $i = u_v = u - \bar{u}$ , $d_v = d - \bar{d}$ , $\bar{u}$ , $\bar{d}$ , $s \simeq \bar{s}$ , $c = \bar{c}$ , $b = \bar{b}$ , $g$ ) obtained in the NNLO NNPDF3.0 global analysis [26] at scales $\mu^2 = 10 \text{ GeV}^2$ (left) and $\mu^2 = 10^4 \text{ GeV}^2$ (right) with $\alpha_s(M_Z^2) = 0.118$ . Figures extracted from Ref. [27].   | 15 |
| 1.4 | Illustration of the stages of a hadron-hadron collision. The red circle in the center of the figure represents the hard collision, surrounded by a tree-like structure representing bremmstrahlung radiation as simulated by parton showers. The purple blob at the bottom represents the UE. The hadronisation process is represented by the light green blobs, dark green blobs indicate hadron decays, while yellow lines signal soft photon radiation [28]. . . . .                            | 16 |
| 1.5 | Summary of several Standard Model total and fiducial production cross-section measurements, compared against their theoretical predictions [29]. .   | 17 |
| 1.6 | Feynman diagrams for the Leading Order (LO) direct-photon production in $pp$ collisions. . . . .   | 18 |
| 1.7 | Feynman diagrams for direct-photon production at Next-to-Leading Order (NLO) in $pp$ collisions. . . . .   | 18 |
| 1.8 | Feynman diagrams for the LO fragmentation-photon processes in $pp$ collisions (a) $qg \rightarrow qg(\gamma)$ and (b) $qg \rightarrow qg(\gamma)$ . . . . .  | 19 |
| 1.9 | Feynman diagrams of the Excited Quark (EQ) production in $pp$ collisions and decay into a quark and a photon in the $s$ -channel (left) and $t$ -channel (right).  | 22 |

|      |  |    |
|------|--|----|
| 2.1  | Overview of the Large Hadron Collider (LHC) complex where all the accelerators that lead up to the LHC are shown [81]. . . . .   | 30 |
| 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) [89]. . . . .   | 31 |
| 2.3  | ATLAS coordinate system [89]. . . . .  | 32 |
| 2.4  | ATLAS Inner Detector (ID) diagrams showing the different submodules, with their corresponding dimensions. . . . .  | 33 |
| 2.5  | ATLAS calorimeter system, showing the Electromagnetic Calorimeter (ECAL) and the Hadronic Calorimeter (HCAL) [98]. . . . .   | 35 |
| 2.6  | Segment of the ECAL showing the layer arrangement and cell dimensions in each layer [85]. . . . .  | 37 |
| 2.7  | Radiation lengths as a function of $ \eta $ for the ECAL, separated for each sub-layer [85]. . . . .   | 37 |
| 2.8  | ATLAS Muon Spectrometer (MS) [85]. . . . .   | 39 |
| 2.9  | Luminosity delivered by the LHC and recorded by ATLAS during the Run-2 [108] and Run-3 data taking periods. For Run-2, the fraction of data good for physics analyses is also displayed. . . . .   | 42 |
| 2.10 | Pileup conditions during Run-2 and Run-3. . . . .  | 43 |
| 3.1  | Schematic showing the tracking parametrization [113]. . . . .  | 45 |
| 3.2  | Diagrama showing the reconstruction algortihm workflow for electrons and photons, extacted from Ref. [118] . . . . .   | 47 |
| 3.3  | Diagram showing the calculation of the calorimetric isolation variable. When $R = 0.4$ , $E_T^{\text{cone}40}$ is computed. . . . .  | 49 |
| 3.4  | Schematic representation of the anti- $k_t$ algorithm for jet clustering [122]. . . . .  | 52 |
| 3.5  | PFLow 4-momemtum jet calibration steps [125]. . . . .  | 54 |
| 3.6  | GN2 tagger discriminant comparison between data and single-lepton $t\bar{t}$ 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 [134]. . . . . | 56 |
| 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 $C$ . . . . .  | 61 |
| 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 [135]. . . . .   | 62 |
| 4.3  | Feynman diagrams for radiative $Z$ decays $Z \rightarrow \ell\ell\gamma$ for the initial-state (left) and final-state (right) radiation. . . . .   | 64 |

|     |  |    |
|-----|--|----|
| 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 $m_{\ell\ell\gamma} \sim m_Z$ . . . . .  | 64 |
| 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. . . . .   | 65 |
| 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. [138]. . . . .   | 67 |
| 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!  | 68 |
| 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 [141]. . . . .  | 69 |
| 5.1 | Adaptive and non-adaptive Kernel Density Estimator (KDE) smoothing. . . . .  | 72 |
| 5.2 | KDE smoothing applied to the $R_{\text{had}}$ 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. . . . .  | 72 |
| 5.3 | Calculation of shift+stretch FFs for $f_{\text{side}}$ with unconverted photons. . . . .   | 74 |
| 5.4 | Shift and stretch FFs for $R_\eta$ (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. . . . .                | 76 |
| 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. . . . . | 77 |

|      |  |     |
|------|--|-----|
| 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. . . . .   | 78  |
| 5.7  | Same as Figure 5.6 but with the SP samples. . . . .  | 78  |
| 5.8  | Cells arrangement and average energy distribution amongst the clusters. . . .  | 79  |
| 5.9  | Energy reweights to be applied to MC cells, for unconverted photons with $0.4 <  \eta  < 0.6$ . *** fix luminosity . . . . .   | 81  |
| 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. . . . .   | 82  |
| 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. . . . .   | 83  |
| 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. . . . .  | 84  |
| 7.1  | Cross-section times branching ratio of the different EQ production modes as a function of the EQ mass at $\sqrt{s} = 13$ TeV. The figure shows the comparison between $q^*$ (blue), $c^*$ (orange) and $b^*$ (green) signals using three different couplings. . . . .  | 97  |
| 7.2  | Cross-section times branching ratio of the RS1 (orange) and ADD (blue) Quantum Black Hole (QBH) models, for a center-of-mass energy of $\sqrt{s} = 13$ TeV. . . . .  | 97  |
| 8.1  | Trigger efficiency for <code>HLT_g140_loose</code> trigger as a function of photon $p_T$ (left), $\eta$ (middle) and $\langle \mu \rangle$ (right) as measured from data for every year between 2015 and 2018. . . . .   | 101 |
| 8.2  | $ \Delta\eta(\gamma, j)  - m_{\gamma+j}$ 2D distribution for the $\gamma + \text{jet}$ PYTHIA background, separating into direct (left) and fragmentation (right) photons. . . . .   | 104 |
| 8.3  | $ \Delta\eta(\gamma, j) $ distribution in two $m_{\gamma+j}$ windows comparison between background and signal models. The bottom pads in Figures 8.3a and 8.3b show the signals efficiencies (coloured lines) and background rejection (hashed histogram) if a cut of the type $ \Delta\eta  < X$ is applied. On the contrary, on Figures 8.3c and 8.3d, the bottom pad shows the signal significance. The $\gamma + \text{jet}$ events were modelled with PYTHIA. . . . . | 105 |

|      |   |     |
|------|---|-----|
| 8.4  | $\eta^\gamma$ and $\eta^j$ distributions in a window of $1000 \text{ GeV} < m_{\gamma+j} < 3000 \text{ GeV}$ comparing signals with the main $\gamma+\text{jet}$ PYTHIA background. Bottom pads show signals significances over the background. . . . .   | 106 |
| 8.5  | $\Delta R_{\min}$ distribution for the $\gamma+\text{jet}$ PYTHIA background. This variable shows the minimum distance found between the leading photon and any of the jets. . . . .  | 107 |
| 8.6  | $E_T^{\text{iso}}$ contribution as a function of $\Delta R_{\min}$ . The $E_T^{\text{iso}}$ values are obtained from the mean of the $E_T^{\text{iso}}$ distribution. . . . .   | 107 |
| 8.7  | $p_T^\gamma \cdot p_T^{\text{jet}}$ 2D distribution for direct and fragmentation PYTHIA $\gamma+\text{jet}$ background. . . . .   | 108 |
| 8.8  | $p_T^\gamma \cdot p_T^{\text{jet}}$ 2D distribution for direct and fragmentation PYTHIA $\gamma+\text{jet}$ background, selecting events in which the leading jet $p_T$ satisfies Eq. 8.1 with $X = 0.5$ . . . . .  | 109 |
| 8.9  | $p_T^\gamma \cdot p_T^{\text{jet}}$ 2D distribution for different $q^*$ signal samples with $f = 1.0$ , selecting jets according to formula Eq. 8.1 with $X = 0.5$ . The first three figures shown the 2D distribution without the application of the $p_T^{\text{jet}}$ cut, while the three at the bottom contain it. . . . . | 109 |
| 8.10 | $p_T^{\text{jet}}$ distribution before and after the cut to remove fragmentation events. The contributions for direct and fragmentation photons are shown separately. On the third plot, the orange (blue) line represents the distribution after (before) the cut is applied. . . . .  | 110 |
| 8.11 | $m_{\gamma+j}$ distribution before and after the $p_T^{\text{jet}}$ cut. The orange (blue) line represents the distribution after (before) the cut is applied. . . . .  | 111 |
| 8.12 | Two-dimensional sequential tagger scheme. . . . .   | 112 |
| 9.1  | Fine factors optimization for the $q^*$ and $c^*$ signal models. Points represents original distribution, while coloured lines represent the estimated PDFs. . . . .  | 114 |
| 9.2  | Illustration of the moment morph method to interpolate between PDFs, for the case of $q^*$ signals with $f = 1$ . The original signal distributions are shown with the solid lines with different colors, while all the interpolated shapes are shown with the dashed red line. . . . .   | 114 |
| 9.3  | Measured acceptances for the EQ (left) and QBH (right). . . . .   | 117 |
| 9.4  | Measured efficiencies for the EQ signals. Figure a: $q^*$ signal when no tagging is applied whatsoever; Figure b: $c^*$ signal with $c$ -tagging applied; and Figure c: $b^*$ signal with $b$ -tagging applied. . . . .   | 118 |
| 10.1 | Identification vs. $E_T^{\text{iso}}$ two-dimensional distribution from data events for the Full Run-2 dataset. . . . .   | 123 |
| 10.2 | $R'$ values computed as a function of $p_T^\gamma$ . The error bars shown correspond to the statistical uncertainties. . . . .  | 125 |

|       |  |     |
|-------|--|-----|
| 10.3  | Composite fit to data. The red curve represents the real photons component which is represented by a Double-sided Asymmetric Crystal-Ball (DSACB) distribution, computed in the first step. The yellow histogram is the fake photon contribution, obtained via leak-photon subtraction in the previous step. The lower pad of the figures represent the normalised residuals (or pull) of the fits. . . . .                    | 126 |
| 10.4  | Measured $\gamma$ +jet $P_A$ values (left) and $FF_{ISO}$ (right) as a function of $p_T^\gamma$ obtained using the ABCD method. The measurements are shown in black (statistical uncertainty only), the shaded blue rectangles show the total uncertainty on the measurements (systematic and statistical added in quadrature), and the red points and line show the smoothed measurements with the total uncertainty. . . . . | 128 |
| 10.5  | Datasets generation for background modeling studies. . . . .   | 129 |
| 10.6  | $m_{\gamma+j}$ distribution showing the effect of the sparsed jet-fakes events at high $m_{\gamma+j}$ , for the SR and SRB regions. . . . .  | 130 |
| 10.7  | Fits to the jet-faking photons backgrounds using the <i>dof8</i> functional model for different analysis regions. . . . .  | 130 |
| 10.8  | Background only fits using different functional models and fit ranges for the inclusive SR region. . . . .   | 131 |
| 10.9  | Background-only (B-only) fits to different toys in the SR region . . . . .   | 132 |
| 10.10 | $m_{\gamma+j}$ maximum distribution for the toys sets for each signal region considered. The different sets of toys, corresponding to different initial <i>dofn</i> functions, are shown with different colors. For each one of them, the mean and the width of the distribution is shown. . . . .   | 133 |
| 10.11 | Flowchart for the fit function and fit-range validation done using MC samples. . . . .   | 133 |
| 10.12 | Scheme of the fit function selection based on the <i>F</i> -test. . . . .  | 134 |
| 10.13 | Signal+Background (S+B) fit to the Asimov background distribution for the calculation of Spurious Signal (SS). The signal model, with floating normalization, is shown with the orange line, the functional model in this case the <i>dof2</i> , represented by the blue line, and the composite fit is shown in green. The figure on the right shows a case in which the SS is negative. . . . .                              | 135 |
| 10.14 | Same as Figure 10.13 but using background toys distributions . . . . .   | 135 |
| 10.15 | $S_{spur}$ distribution (blue line) in the SR region, with a gaussian fit to it. The final $S_{spur}$ value is directly taken from the distribution, not from the fit. The functional model is <i>dof2</i> and the background is fitted within the range [800, 1000] GeV. The EQ signal model is used contemplating only $q^*$ with coupling $f = 1.0$ . . . . .   | 136 |
| 10.16 | SS tests results for the inclusive SR region. The different figures correspond to the fit-range that yields the lowest SS for the given signal model and function's Degree of Freedom (dof). Different functional forms are shown with different colors. . . . .   | 137 |

|       |  |     |
|-------|--|-----|
| 10.17 | Idem Figure 10.16 but using toy samples. . . . .   | 137 |
| 10.18 | Signal Injection (SI) tests for the EQ model in regions SR, SRL, SRB and SRC, respectively. The fit-range and functional model used are the ones that yield the lowest $S_{\text{spur}}$ , taken from the SS tests, and are displayed in each plot. Each considered mass is shown with a different color. The $x$ -axis shows the injected amplitude in units of $\sqrt{N_B}$ , while the $y$ -axis represents the extracted signal in terms of $\sqrt{N_B}$ . . . . . | 139 |
| 10.19 | $\chi^2/\text{ndof}$ distribution for each functional model in different signal regions. . . . .   | 140 |
| 10.20 | $F$ -score distribution. The tests are done by comparing two functional models at the same time, shown by each of the filled histograms. For each one of them, the number of matching and converged, toys are shown, as are the means and widths of the $F$ -score distributions. . . . .  | 141 |
| 10.21 | p-value score distribution comparing the performance of the different pair of functions. The reported means and widths correspond to the $F$ -score distributions, and for each p-value distribution, the average p-value is shown. . . . .  | 142 |
| 11.1  | Selected $m_{\gamma+j}^{\text{reco}}/m_{\gamma+j}^{\text{truth}}$ fits for the $m_{\gamma+j}$ binning studies. The distribution is shown with the blue dots while the gaussian fit with the red line. In all cases, the mean and standard deviation of the gaussian post-fit is shown. . . . .   | 145 |
| 11.2  | $m_{\gamma+j}$ resolution values and fit using Eq. 11.1. The post-fit parameters are indicated in the figure where $p_0 \equiv a$ , $p_1 \equiv b$ and $p_2 \equiv c$ . The fit has been performed excluding the last measured resolution value due to lack of statistics.   | 145 |
| 11.3  | $m_{\gamma+j}$ binning optimisation results showing the bin-widths as a function of $m_{\gamma+j}$ (left) and the comparison between the binning resolution and detector resolution (right). . . . .   | 146 |
| 11.4  | Scheme of the B-only fit worflow to data. . . . .  | 147 |
| 11.5  | B-only fits to full Run-2 data in the four considered analysis regions. The data distribution and the background fit are shown in the top pad, while the normalised residues (or pull) in the lower pad. Moreover, fit $\chi^2$ and $p$ -value are displayed for each region. . . . .  | 148 |
| 11.6  | BumpHunter results for region SR. The figure on the left shows the most significant bump, marked with red dashed lines. On the right, the local $p$ -value for different tested windows is presented, indicating where the most significant excess is located. The global $p(\text{BH})$ as calculated by Eq. 6.25 is 0.3938, while the global significance $-0.009$ . . . . .   | 149 |
| 11.7  | Same as Figure 11.6 but in region SRL, where the resulting global $p(\text{BH}) = 0.8790$ and significance $-1.170$ . . . . .  | 149 |
| 11.8  | Same as Figure 11.6 but in region SRC, where the resulting global $p(\text{BH}) = 0.6438$ and significance $-0.368$ . . . . .  | 150 |
| 11.9  | Same as Figure 11.6 but in region SRC50, where the resulting global $p(\text{BH}) = 0.7941$ and significance $-0.820$ . . . . .  | 150 |
| 11.10 | Scheme of the S+B fit worflow to data. . . . .   | 151 |

|       |  |     |
|-------|--|-----|
| 11.11 | Observed and expected limits on $q^*$ , $c^*$ and $b^*$ signals with coupling $f = 1$ , using the full Run-2 dataset. Expected upper limits are shown with the black dashed lines, while the line with marks represent observed limits. Red solid lines indicate the predicted cross section. Marked in the figures, the expected and observed upper limits on the mass are shown. . . . .           | 152 |
| 11.12 | Upper limit comparison between the SRC (blue lines) and the SRC+SRB+SRL (orange lines) regions using $c^*$ signals with coupling $f = 1$ . For both regions, the expected and observed upper limit on the $m_{q^*}$ mass is displayed in the figure. . . . .   | 153 |
| 11.13 | Observed and expected limits comparison between different couplings for EQ signals. Blue (orange) lines represent the limits for coupling $f = 1$ ( $f = 0.5$ ), and the expected model cross sections are shown with brown (pink) solid lines for $f = 1$ ( $f = 0.5$ ) signals. . . . .  | 154 |
| 11.14 | Expected and observed upper limits in the mass-coupling plane for the $q^*$ (SR), $b^*$ (SRB) and $c^*$ (SRC50+SRB+SRL50) signals. The expected (observed) upper limits are displayed with dashed (solid) lines, and the $1\sigma$ uncertainty band is represented by the colored shaded regions. . . . .  | 155 |
| 11.15 | Observed and expected upper limits on the QBH signal models. The RS1 model ( $n = 1$ ) is shown on the left, while the ADD model ( $n = 6$ ) is presented on the right figure. Observed limits are shown with the black dots and solid line, meanwhile, expected upper limits are represented with the black dashed line, with the correponding $\pm 1$ and $\pm 2\sigma$ uncertainty bands. . . . . | 156 |
| 11.16 | Observed and expected upper lmts on general gaussian signals with three different widths: $\sigma_G/m_G = 0.02, 0.07$ and $0.15$ in regions SR, SRL, SRC and SRB. . . . .  | 157 |
| 11.17 | Comparison of the observed and expected limits on Gaussian-shaped resonances of 2% width obtained in this analysis (blue and orange for the SR and SRL regions, respectively) against the previous ATLAS result using 2015–2016 data (grey). The observed limits are shown with the solid line and the expected ones with dashed lines. . . . .  | 158 |
| A.1   | Parameters of the most general case of an ellipse. . . . .   | 160 |
| B.1   | SI tests results using Gaussians signals in the inclusive SR region. The fit is performed in the range $1200 - 8000$ GeV using the <i>dof3</i> function. Each considered mass is shown with a different color. The $x$ -axis shows the injected amplitude in units of $\sqrt{N_B}$ , while the $y$ -axis represents the extracted signal in terms of $\sqrt{N_B}$ . . . . .                          | 161 |
| B.2   | Same as Figure B.1 in the SRL region performing fits in the range $1500 - 8000$ GeV with the <i>dpf3</i> function. . . . .   | 161 |
| B.3   | Same as Figure B.1 in the SRC region performing fits in the range $1000 - 7000$ GeV with the <i>dpf4</i> function. . . . .   | 162 |

|     |  |     |
|-----|--|-----|
| B.4 | Same as Figure B.1 in the SRB region performing fits in the range 900 – 7000 GeV with the <i>dpf5</i> function. . . . .                                  | 162 |
| B.5 | SI tests results using QBH signals in the inclusive SR region. The fit is performed in the range 1000 – 8000 GeV using the <i>dof5</i> function. . . . . | 162 |

# 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. [85] . . . . .   | 32  |
| 3.1 | Summary of electron and photon isolation WPs use throughout this thesis. . . . .   | 49  |
| 3.2 | Measured $b$ -tagging efficiencies and $c$ -jets, $l$ -jets and $\tau$ -jets rejections in the low and high- $p_T$ regime. . . . .   | 56  |
| 3.3 | Measured $c$ -tagging efficiencies and $b$ -jets, $l$ -jets and $\tau$ -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! . . . . . | 57  |
| 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. . . . .  | 60  |
| 7.1 | Cross-section times branching ratio [fb] of the EQ model, for the three considered flavours using different coupling values. . . . .   | 96  |
| 7.2 | Sum of cross section times branching ratio [fb] for the $\gamma$ +jet final state for the six non-thermal black holes states for each threshold mass $m_{\text{th}}$ . . . . .   | 98  |
| 7.3 | Details of the background $\gamma$ +jet MC samples . . . . .   | 99  |
| 8.1 | Overlap removal procedure. . . . .   | 103 |
| 8.2 | Background and signal efficiency on the $p_T^{\text{jet}}$ cut defined above. . . . .  | 110 |
| 8.3 | Signal regions definitions. *** for the moment, all plots have the old naming SRC50/SRL50, but in the text I used SRC/SRL. Should I keep the old naming until I re-do the plots with the correct text? . . . . .   | 111 |
| 9.1 | EQ signals significances over the total background in the SR, SRB, SRC and SRL regions. The considered signals for each flavor have coupling $f = 1.0$ . . . . .   | 115 |

|      |   |     |
|------|---|-----|
| 9.2  | Acceptance measurements for two benchmark EQ $q^*$ signals with $f = 1.0$ .<br>The second and fifth columns denote the cutflow for all the cuts, the third<br>and sixth the absolute efficiencies, and the fourth and seventh the relative<br>efficiencies of each cut. . . . .   | 117 |
| 10.1 | Baseline event selection used for the jet fakes estimation using the template<br>fit procedure. $p_T^{\text{iso}}$ is defined as shown in Eq. 3.1.*** revise selection on jet pt  | 122 |
| 10.2 | Smoothed photon fake factors $\text{FF}_{\text{ISO}}$ and $\gamma + \text{jet}$ purity as a function of $p_T^\gamma$ . . .  | 127 |
| 10.3 | Summary of fit strategies for each analysis region and signal model. The<br>table shows the functional form used, as well as the $m_{\gamma+j}$ fit-range in GeV.<br>The last column indicates if the fit is performed simultaneously for the SRC,<br>SRB and SRL regions, in case each one of them are used for the interpreta-<br>tion/model. . . . . | 143 |

1

## 2 INTRODUCTION

3 Our best current understanding of particle physics is given by the Standard Model ([SM](#)),  
 4 a theory that successfully explains a wide range of experimental results and precisely  
 5 predicted many different physics phenomena and new particles, such as the Higgs boson  
 6 discovered in 2012 by the [ATLAS](#) and Compact Muon Solenoid ([CMS](#)) collaborations. This  
 7 discovery, lead to the Nobel Prize Award to François Englert and Peter Higgs the following  
 8 year. Despite its remarkable success, the [SM](#) is known to be incomplete as it cannot ex-  
 9 plain a number of experimental observations, such as the overwhelming astrophysical and  
 10 cosmological evidence for dark matter, the hierarchy problem, why there are only three  
 11 families of fermions, among others. In the past decades, many theories for new physics

12 Beyond Standard Model ([BSM](#)), such as Supersymmetry ([SUSY](#)), emerged and provided  
 13 well-motivated and promising theoretical frameworks to extend our fundamental under-  
 14 standing of particle physics and improve upon the shortcomings of the [SM](#). However,  
 15 none of the numerous searches for new physics signatures that have been conducted at  
 16 particle colliders in the past years could provide any direct evidence for the existence of  
 17 new particles or forces as predicted by these theories.

18 A wide variety of these new theoretical models predict the existence of particles at high  
 19 energies. In order to explore these regions, specially the TeV scale, in the [CERN](#) laboratory  
 20 the Large Hadron Collider ([LHC](#)) [1] was built. Installed in a 27-kilometre circular tunnel,  
 21 it is the world's largest and most powerful particle collider. This machine is capable of  
 22 colliding energetic beams of protons at rates upward of millions per second. The precision  
 23 and high beam energy of the [LHC](#) allow to explore energies above 7 TeV, an energy range  
 24 never before achieved in a particle collider. The [LHC](#) collides protons in four interaction  
 25 points, where the 4 [LHC](#) experiments are located: [ATLAS](#), [CMS](#), [LHCb](#) and [ALICE](#). Between  
 26 the years 2015 and 2018 a data-taking period called Run-2 took place, where protons were  
 27 collided at  $\sqrt{s} = 13$  TeV, and collecting a total of integrated luminosity of  $140.01 \text{ fb}^{-1}$ . In  
 28 2022, the [LHC](#) Run-3 started, where the center-of-mass energy was increased to  $\sqrt{s} =$   
 29  $13.6$  TeV, and by the end of 2024 [ATLAS](#) could collect  $183 \text{ fb}^{-1}$  of data.

30 One of the most important experiments at the [LHC](#) is [ATLAS](#) (A Toroidal LHC ApparatuS),  
 31 a general-purpose detector designed to perform both precision measurements within the

32 **SM** and searches for new phenomena associated with physics **BSM**. The **ATLAS** detector  
33 is composed of different subdetectors that play different roles in the reconstruction of the  
34 colliding particles. The Inner Detector is in charge of measuring the tracks of charged  
35 particles, the calorimeters measure the energetic depositions of photons, electrons and  
36 different hadrons, and finally the Muon Spectrometer (**MS**) allows to measure the muon  
37 trajectories. Intertwined between them there is a powerful magnet system, which bends  
38 the trajectory of charged particles. Finally, the **ATLAS** detector has a precise Trigger Sys-  
39 tem that filters out events of little interest, thus reducing the frequency of the data flow.  
40 Typically, any high energy physics experiment relies on two different types of samples: ac-  
41 tual data and simulated data. The latter are heavily used by the collaborations, as they  
42 allow to understand the behaviour any different processes would have, so that one expects  
43 what the actual data would behave like.

44 Prompt photon production from  $pp$  collisions at the **LHC** constitutes a key part on the  
45 **ATLAS** physics program, either for precise measurements of Quantum Chromodynam-  
46 ics (**QCD**) observables, or because several **BSM** scenarios involve having isolated prompt  
47 photons in the final state. However, the main process emerging from  $pp$  collisions is  
48 the dijet production, and sometimes one of these jets has a very similar signature as a  
49 photon, therefore this particle being mis-identified. The process of correctly identifying  
50 the particles that enter the **ATLAS** detector is called identification, and it constitutes one  
51 of the main ingredients in a physics analysis. For photons, this identification is carried  
52 out by studying the electromagnetic (**EM**) shower initiated by the particles in the calori-  
53 meter using several shape variables, and as anticipated, it is carried out using the actual  
54 recollected data and the simulations. However, it was seen that the simulation did not  
55 correctly predict data leading to incongruent results. One of the main tasks for this thesis  
56 is the correction of the shape variables used for photon identification. The current method  
57 to correct the variables is called Fudge Factor (**FF**), which was drastically improved in  
58 this work. Furthermore, another approach was studied, where modifications to lower-  
59 level variables can be carried out to simultaneously fix all the variables used to identify  
60 photons.

61 It was mentioned in the previous paragraph that prompt photons are of great importance  
62 for **BSM** searches. In particular, in the photon+jet final state, the invariant mass follows  
63 a very smoothly falling shape, providing an excellent scenario for bump-searches, where  
64 different particles that decay to a photon+jet pair can be searched for. Two of the the-  
65 oretical models that aim to answer different **SM**'s shortcomings predict the existence of  
66 these types of particles. The first theoretical model gives an explanation of why there  
67 are three fermion families, and propose that the quarks are not fundamental particles but  
68 bound states of more fundamental ones that experiment an unknown force. Then, Excited  
69 Quark (**EQ**) states ( $q^*$ ) should be observed in  $pp$  collisions at the **LHC** depending if the  
70 value of compositeness scale  $\Lambda$  is lower than the center-of-mass energy. These **EQs** would  
71 decay into a pair of photon and jet, leaving a bump on the  $m_{\gamma+j}$  around the mass of the  
72 **EQ**. The second model, with the introduction of extra dimensions, attempts to propose a

73 solution for the hierarchy problem. Certain types of extra-dimension models predict the  
74 fundamental Planck scale  $m_P$  in the  $4+n$  dimensions ( $n$  being the number of extra spatial  
75 dimensions) to be at the TeV scale, and thus accessible in  $pp$  collisions at  $\sqrt{s} = 13$  TeV at  
76 the [LHC](#). In such a TeV-scale  $m_P$ , Quantum Black Holes ([QBHs](#)) may be produced at the  
77 [LHC](#) as a continuum above the threshold mass  $m_{\text{th}}$  and then decay into a small number of  
78 final-state particles including photon-quark/gluon pairs before they are able to thermalize.  
79 In this case a broad resonance-like structure could be observed just above  $m_{\text{th}}$  on top of the  
80 [SM](#)  $m_{\gamma+j}$  distribution. Two particular models are studied under this theory, which propose  
81 different number of extra dimensions: the Randall-Sundrum RS1 model proposes a total  
82 of 5 space-time dimensions, and the Arkani-Hamed, Dimopoulos and Dvali ADD model,  
83 which counts with a total of 10 space-time dimensions. Finally, given the smoothness of  
84 the  $m_{\gamma+j}$  distribution, it is possible to make a model-agnostic search on this background,  
85 where the signal is considered to follow a Gaussian-shape resonance. This type of search  
86 provides a more general interpretation of the study, as it allows to compare with any  
87 theory model that proposes a Gauss-shape resonance.

88 Similar for  $\gamma+\text{jet}$  resonances considering the theoretical models previously discussed have  
89 been carried out previously, setting upper limits on the theories. [ATLAS](#) latest result used  
90  $36.7 \text{ fb}^{-1}$  and excluded  $q^*$  models with masses up to 5.3 TeV, RS1-type [QBH](#) models up  
91 to 4.4 TeV and ADD-type ones up to 7.1 TeV. On the other hand, [CMS](#) also performed  
92 similar studies using  $138 \text{ fb}^{-1}$ , where they studied [EQ](#) models separating into light- ( $q^*$ )  
93 and heavy-flavour ( $b^*$ ) [EQs](#), setting stringent limits of 6.0 TeV and 2.2 TeV, respectively.  
94 Similarly, the [CMS](#) collaboration studied the ADD and RS1 [QBH](#) models, where upper  
95 limits on the masses extend up to 7.5 TeV and 5.2 TeV, respectively.

96 Then, the main work for this thesis is the search for high-mass resonances in the photon+jet  
97 final state. The search is performed using the full Run-2 dataset collected at a center-of-  
98 mass energy of  $\sqrt{s} = 13$  TeV, using a total of  $140.01 \text{ fb}^{-1}$ . Both the [EQ](#) and [QBH](#) models  
99 are studied, as well as the search for generic Gaussian-shaped signals. In this work, [EQs](#)  
100 models are studied separating into  $q^*$  ( $u^*/d^*$ ),  $c^*$  and  $b^*$  signals, being this work the first  
101 one at the [LHC](#) considering the charm flavour, thanks to a novel and excellent-performing  
102 flavour tagger algorithm.

103 The thesis is divided into four parts. Part I describes the theory background and motiva-  
104 tions of the work, where [SM](#) is briefly described and the two theory models for physics  
105 [BSM](#) are discussed.

106 Part II contains two chapters in which the first one (Chapter 2) describes the [LHC](#) and the  
107 [ATLAS](#) detector in detail, making emphasis on the different parts of the [ATLAS](#) detector.  
108 In Chapter 3, the methods used for the different object reconstruction and identification  
109 are discussed.

110 In Part III, the different algorithms used for photon identification are explained. First,  
111 in Chapter 4, the variables used for photon identification, how it is optimised, and how  
112 the identification efficiency measurements are performed are discussed. Then, Chapter 5

113 presents the different methods derived in this thesis to fix the photon shape variables in  
114 the simulation.

115 Finally, Part IV presents the search for  $\gamma$ +jet resonances in the high-mass range. Chapter 6  
116 starts discussing the general analysis strategy and sets up the statistical methods to be  
117 used. Next, in Chapter 7, the samples that are used for both the theory models as for  
118 the SM backgrounds are described. The event selection and the definitions of the signal  
119 regions that are used in the search is presented in Chapter 8. Signal modeling and the  
120 experimental and theoretical systematic uncertainties are described in Chapter 9, which  
121 comprise one of the most important and challenging points in any search. It is crucial to  
122 have an excellent understanding of the SM backgrounds. For this, in Chapter 10, the two  
123 main backgrounds are studied, where a functional model is used to model the final, total,  
124 background. Finally, the results obtained from this search are presented and discussed in  
125 Chapter 11.

126

## Part I

127

# Theory Motivation

128

# <sup>129</sup> THE STANDARD MODEL AND BEYOND

130

*"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

<sup>131</sup> This thesis covers a search for new particles predicted by different scenarios beyond the  
<sup>132</sup> Standard Model (**SM**). In this chapter, the foundations for this search will be laid. The  
<sup>133</sup> chapter starts with a summary of the main concepts of the **SM** used throughout this thesis  
<sup>134</sup> in Section 1.1. In said section, special focus is given on the Strong-force theory, on the  
<sup>135</sup> hadron interactions in a  $pp$  collision, and on the prompt-photon production process. Then,  
<sup>136</sup> a brief overview of the current limitations of the **SM** is given in Section 1.2, to then present  
<sup>137</sup> two different forms of New Physics that aim to solve the **SM** shortcomings. Finally, in  
<sup>138</sup> Section 1.3, the chapter ends with how these **SM** processes are simulated using Monte  
<sup>139</sup> Carlo (**MC**), showing the different steps and tools to do so.

## <sup>140</sup> 1.1 The Standard Model (**SM**)

<sup>141</sup> The Standard Model (**SM**) of particle physics is the mathematical theory that describes  
<sup>142</sup> all the known elementary particles and their interactions. The theory has been developed  
<sup>143</sup> through the end of the 20<sup>th</sup> century, being finalised by the mid-1970s after the experimental  
<sup>144</sup> confirmation of the quarks. Over time, after many experiments backing its predictions,  
<sup>145</sup> it has become the most complete and precise theory in particle physics.

<sup>146</sup> The **SM** managed to describe, to present day, three of the four fundamental forces in  
<sup>147</sup> nature: the electromagnetic (**EM**), the weak and the strong interactions. These interactions  
<sup>148</sup> work over different ranges and have different strengths. Gravity, the fourth force,  
<sup>149</sup> although not included in the **SM**, is the weakest of the interactions and has an infinite  
<sup>150</sup> range. The **EM** interaction also has an infinite range but is much stronger than gravity. On  
<sup>151</sup> the other hand, the weak and strong forces act on very short distances, and only domin-

ate in the subatomic range. The weak interaction is weaker than the EM and the strong, but still much stronger than gravity. Finally, the strong force is the strongest of them all. The three forces described by the SM arise from the exchange of mediator particles called *bosons* between all matter particles, called *fermions*.

### 1.1.1 Elementary particles and their interactions

According to the SM, all matter is made out of fermions, which are particles following the Fermi-Dirac statistics and have half-integer spin. These fermions interact between themselves by the exchange of the aforementioned bosons, which are particles of integer spin, following the Bose-Einstein statistics. Up to date, there has not been any single experiment capable of finding evidence that these fermions have internal structure.

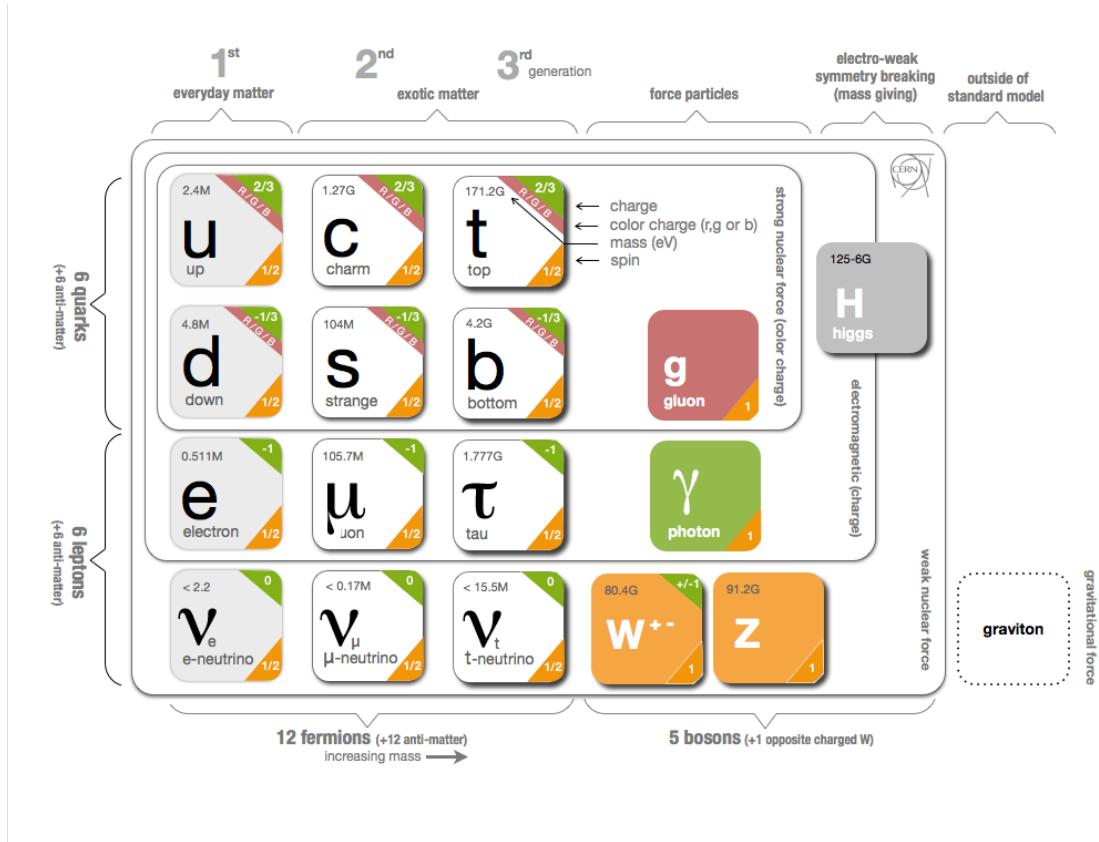


Figure 1.1: Overview of the particles of the SM. All fermions participate in the weak interaction, but only the quarks interact with gluons, whereas both quarks and charged leptons interact with via the EM force. Neutrinos, being neutral and colourless, only interact with the  $W$  and  $Z$  bosons via the weak force. Finally, the graviton, although it has not been discovered yet, should be the corresponding force carrier of the gravity force. Extracted from Ref. [2].

Fermions are divided into two kinds of elementary particles: leptons and quarks. There are six leptons classified according to their charge, and are divided into three families or generations, ordered based on their mass. Particles in higher generations have higher mass and are highly unstable, decaying into lower generation leptons. For this reason, matter is built on first generation leptons. The leptons are: electron ( $e$ ), muon ( $\mu$ ), and tau ( $\tau$ ), with their respective neutrinos: electron neutrino ( $\nu_e$ ), muon neutrino ( $\nu_\mu$ ), and tau neutrino

168 ( $\nu_\tau$ ), and properties of each are shown in Figure 1.1. There are also six antileptons, which  
 169 have the opposite charge as the leptons, therefore increasing the number of leptons in the  
 170 SM up to 12. The electron, muon and tau, all have electric charge and sizable mass, while  
 171 the neutrinos are electrically neutral and have very small mass.

172 Similarly, there are six flavours of quarks (also having their respective antiparticle): up  
 173 ( $u$ ), down ( $d$ ), charm ( $c$ ), strange ( $s$ ), top ( $t$ ), and bottom ( $b$ ). Quarks also come in  
 174 three different colours giving a total of 36 quarks, and only mix in such a way as to form  
 175 colourless objects. An overview of the quarks and their properties are shown in Figure 1.1.

176 Each of the three forces unified in the SM is described by a Quantum Field Theory (QFT),  
 177 corresponding to the exchange of a boson mediator. The strong force, mediated by mass-  
 178 less gluons, is responsible for binding quarks together. While gluons do not carry electric  
 179 charge, they possess color charge, which leads to the phenomenon of *color confinement*.  
 180 Despite being massless, the strong interaction becomes stronger at low energies, confining  
 181 quarks and gluons within hadrons due to the property asymptotic freedom and the afore-  
 182 mentioned property of color confinement. The EM force is mediated between charged  
 183 particles by photons. Photons do not have mass, and, as a consequence, the interaction  
 184 has infinite range. Finally, the weak interaction is mediated by the massive  $W$  and  $Z$  bo-  
 185 sons, leading to short-range interactions. The fundamental properties of these bosons are  
 186 also displayed in Figure 1.1.

### 187 1.1.2 Brief mathematical description of the Standard Model

188 The SM is a renormalizable field theory based on local symmetries, providing a description  
 189 of the fundamental particles and their interactions: the strong, the weak and the EM.  
 190 These interactions span by the requirement that the theory is invariant under local gauge  
 191 transformations of the symmetry group:

$$SU(3)_C \times SU(2)_L \times U(1)_Y,$$

192 where  $Y$  is the hypercharge,  $L$  the left-handed helicity and  $C$  the colour charge, and  
 193 they represent the conserved quantities of the symmetry group. Every local gauge trans-  
 194 formation can be absorbed within a gauge field, with the excitations of the gauge fields  
 195 called gauge bosons. The Electroweak (EW) sector of the SM  $SU(2)_L \times U(1)_Y \rightarrow U(1)_{\text{EM}}$   
 196 describes the weak and EM interactions, after the spontaneous symmetry breaking mech-  
 197 anism by virtue of the Higgs potential. The non-abelian group  $SU(3)_C$  with colour charge  
 198 describes the strong interactions between quarks and gluons, and the theory is known as  
 199 Quantum Chromodynamics (QCD) [3].

200 In principle, the particles included in the SM are massless, unlike the particles observed  
 201 in nature. While the equations for the EW interactions correctly describe particles like  
 202 the photon,  $W$ , and  $Z$  bosons, they fail to account for their masses. To address this, the  
 203 concept of ElectroWeak Symmetry Breaking (EWSB) was introduced, known as the Brout-  
 204 Englert-Higgs mechanism [4–7]. This mechanism explains how the  $W$  and  $Z$  bosons

205 acquire mass through the spontaneous breaking of the **EW** symmetry, caused by the Higgs  
 206 scalar field obtaining a non-zero vacuum expectation value. Furthermore, it predicts the  
 207 existence of a new scalar particle, leading to a new massive boson with spin 0, called  
 208 the Higgs bosons. This particle was experimentally confirmed in 2012 by the A Toroidal  
 209 LHC ApparatuS ([ATLAS](#)) and Compact Muon Solenoid ([CMS](#)) collaborations at the Large  
 210 Hadron Collider ([LHC](#)), with a measured mass of 125.25 GeV [8, 9].

## 211 The Electroweak interaction

212 The **EW** interactions satisfy the  $SU(2)_L \times U(1)_Y$  group gauge symmetry. The  $SU(2)_L$   
 213 group, called weak isospin, acts only on the left-handed chiral fermions, and  $U(1)_Y$  is the  
 214 hypercharge group and acts on both chiralities in a vectorial way.

215 The  $SU(2)_L \times U(1)_Y$  group has four generators, of which three belong to the weak isospin:  
 216  $T_i = \frac{\sigma_i}{2}$ , with  $i = 1, 2, 3$  and  $\sigma_i$  being the Pauli matrices, and one to the hypercharge  
 217 group:  $\frac{Y}{2}$ . Left-handed fermions transform as doublets under  $SU(2)_L$ ,  $f_L \rightarrow e^{iT_i\theta_i} f_L$  with

$$f_L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}, \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \dots, \quad (1.1)$$

218 meanwhile right-handed fermions transform as singlets  $f_R \rightarrow f_R$  with

$$f_R = e_R, u_R, d_R, \dots \quad (1.2)$$

219 This distinction between left- and right-handedness implies that the  $SU(2)_L \times U(1)_Y$  is a  
 220 chiral theory.

221 The electric charge is related to the third component of the weak isospin  $T_3$  and the  
 222 hypercharge  $Y$ , according to the Gell-Mann Nishijima formula:

$$Q = T_3 + \frac{Y}{2} \quad (1.3)$$

223 The number of associated gauge bosons coincides with the number of generators of the  
 224 symmetry group. For the weak isospin there are 3  $SU(2)_L$  bosons:  $W_\mu^1, W_\mu^2, W_\mu^3$ , and for  
 225  $U(1)$  we have one hypercharge boson:  $B_\mu$ . The global  $SU(2)_L \times U(1)_Y$  symmetry becomes  
 226 local, replacing in the lagrangian the derivative of the fields by the covariant derivative:

$$D_\mu = \partial_\mu - ig \frac{\tau^i}{2} W_\mu^i - ig' \frac{Y}{2} B_\mu, \quad (1.4)$$

227 where  $g$  is the coupling constant of  $SU(2)_L$  and  $g'$  of  $U(1)_Y$ .

228 The **EW** lagrangian density can then be written as the sum of the fermionic lagrangian  
 229 with the gauge interactions and the kinetic terms for the introduced gauge fields:

$$\mathcal{L}_{\text{EW}} = -\frac{1}{4} F_{\mu\nu}^i F_i^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \sum_{f=\ell,q} \bar{f} i \gamma^\mu D_\mu f, \quad (1.5)$$

230 where the Yang-Mills field tensors  $F_{\mu\nu}^i$  for  $SU(2)_L$  and  $B_{\mu\nu}^i$  for  $U(1)_Y$  are defined as:

$$F_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i + g \epsilon_{ijk} W_\mu^j W_\nu^k \quad (1.6)$$

$$B_{\mu\nu} = \partial_\mu W_\nu - \partial_\nu W_\mu. \quad (1.7)$$

231 **The Higgs mechanism**

232 The photon and the gluons have zero masses as a consequence of the exact conservation of  
 233 the corresponding symmetry generators: the electric charge and the eight colour charges,  
 234 respectively. On the other hand, the weak bosons have large masses signalling that the  
 235 corresponding symmetries are largely broken. The SM predicts in principle massless fer-  
 236 mions and gauge bosons, in contradiction with the observations. To fix this problem, a  
 237 mechanism is introduced in order these masses are generated. A scalar field is then ad-  
 238 deded, and the symmetry is considered as a latent symmetry, present in the Lagrangian but  
 239 not respected by the Vacuum Expectation Value (VEV) of the field. The degenerate va-  
 240 cum states respect the  $SU(2)_L \times U(1)_Y$  symmetry, and it is the particular choice of a  
 241 VEV that generates the  $SU(2)_L \times U(1)_Y \rightarrow U(1)_{\text{EM}}$  breaking. This, is what is known as  
 242 the Higgs mechanism, and it was theorised by Brout, Englert, Higgs, Guralnik, Hagen and  
 243 Kibble in 1964.

244 To break the EW symmetry, a complex scalar field is added, called the Higgs field, that  
 245 adds another boson to the theory which is an isospin doublet that couples to the  $W_{\mu\nu}^i$  and  
 246  $B_{\mu\nu}$  fields. The Higgs sector of the Lagrangian is given by:

$$\mathcal{L}_{\text{Higgs}} = (D^\mu \phi)^\dagger (D_\mu \phi) - V(\phi) \quad (1.8)$$

247 where  $\phi$  is a complex scalar field in the  $SU(2)$  representation:

$$\Phi = \begin{pmatrix} \phi_+ \\ \phi_0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \quad (1.9)$$

248 with hypercharge  $U(1) Y = +1$ , where  $\phi_+$  and  $\phi_0$  are complex with electric charge +1  
 249 and 0, respectively, and the covariant derivative is given by Eq. 1.4. The reason to have  
 250 the additional  $U(1)_Y$  symmetry is so that the theory produces a massless gauge boson  
 251 associated to the photon. The Higgs potential  $V(\phi)$  is required to have the form:

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \quad (1.10)$$

252 in order to guarantee the renormalisability of the theory, and the invariance of  $SU(2)$  and  
 253  $U(1)$ . The parameter  $\lambda$  needs to be positive so that the potential has a minimum, so that the  
 254 behaviour of the field is then determined by  $\mu$ . For  $\mu^2 > 0$ , the field generates a non-zero  
 255 VEV ( $v := \phi^\dagger \phi$ ) that spontaneously breaks the symmetry. The potential  $V(\phi)$  samples the  
 256 very well-known shape of a Mexican hat, and has infinite numbers of degenerate states  
 257 with minimum energy satisfying  $v = \sqrt{-\mu^2/\lambda}$ . From these infinite states, it is usual to  
 258 use:

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad (1.11)$$

259 with  $\phi_1 = \phi_2 = \phi_4 = 0$ ,  $\phi_3 = v$ .

260 To study the particle spectrum, the field around the minimum is studied using an expan-  
 261 sion in the radial direction:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \quad (1.12)$$

262 where  $H(x)$  are ground-state excitations around the minimum, but in the radial direction  
 263 of the potential. Due to the gauge invariance of the potential for excitations around the  
 264 circumference minima, and according to the Goldstone theorem, one should in principle  
 265 have three non-mass scalar bosons associated to the non-radial degrees of freedom of the  
 266 field. The gauge arbitrariness allows the Goldstone bosons to be absorbed by the  $W$  and  
 267  $Z$  bosons (they provide the longitudinal polarisations acquired by the gauge fields). The  
 268 development in the radial direction gives the mass of the  $H$  excitation,  $\sqrt{2\lambda}v$  which is the  
 269 mass of the Higgs boson, and cubic and quartic couplings of this boson. In this way, the  
 270 **SM** boson's masses have the form:

$$m_\gamma = 0 \quad (1.13)$$

$$m_W = \frac{gv}{2} \quad (1.14)$$

$$m_Z = \frac{g}{2} \sqrt{g^2 + g'^2} \quad (1.15)$$

$$m_H = \sqrt{2\lambda}v \quad (1.16)$$

271 Yukawa-type coupling terms to the Higgs field give masses to the **SM** fermions:

$$\mathcal{L}_{\text{Yuk}} = g_f \left( \bar{\psi}_L \phi \psi_R + \phi^\dagger \bar{\psi}_R \psi_L \right), \quad (1.17)$$

272 being this a  $SU(2)$  invariant. The coupling constant  $g_f$  describes the coupling between  
 273 the Higgs doublet to the fermions. By doing a similar expansion of the field as done  
 274 previously and substituting into the Yukawa Lagrangian, one obtains terms indicating the  
 275 fermion masses as

$$m_f = \frac{g_f v}{\sqrt{2}}. \quad (1.18)$$

## 276 Quantum Chromodynamics (QCD)

277 The huge effort to describe the rich spectrum of mesons and baryons resonances that  
 278 were discovered during the 1950s, prompted Gell-Mann and Zweig to propose in 1964  
 279 the quark model [10–12], which asserts that hadrons are in fact composites of smaller  
 280 constituents. Zweig called the elementary particles *aces* while Gell-Mann called them  
 281 *quarks*, but finally the theory came to be called the quark model.

282 The quark model was formalised into the theory of **QCD** with quarks carrying an additional  
 283 quantum number called the colour charge,  $C = R, G, B$ . Without colour charge, it would  
 284 seem that the quarks inside some hadrons exist in symmetric quantum states, in violation  
 285 of the Pauli exclusion principle. The theory satisfies the gauge symmetry of the group  
 286  $SU(3)_C$ , which has eight generators  $T^a = \frac{\lambda_{\alpha\beta}^a}{2}$ , with  $\alpha$  and  $\beta$  being the color indices,  $\lambda_{\alpha\beta}^a$   
 287 the eight Gell-Man matrices ( $a = 1, 2, \dots, 8$ ). These eight generators introduce eight new  
 288 physical gauge fields: the gluons. Mesons and baryons, hadrons composed of two and  
 289 three quarks respectively, are *white* singles (neutral color charge) of  $SU(3)_C$ .

290 The local  $SU(3)_C$  symmetry is obtained by replacing in the lagrangian the covariant  
 291 derivatives

$$D_\mu = \partial_\mu - ig_s \sum_{a=1}^8 \frac{\lambda_{\alpha\beta}^a}{2} G_\mu^a,$$

292 where  $g_s$  is the bare **QCD** coupling constant and is usually replaced by  $\alpha_s = g_s^2/4\pi$ . The  
 293 Yang-Mills field tensor  $G_{\mu\nu}^a$  for the group  $SU(3)_C$  can be written as

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_s f_{abc} G_\mu^b G_\nu^c,$$

294 where  $f_{abc}$  are the structure constants of  $SU(3)$ . It is important to note that the last term in  
 295 the previous equation describes the gluon auto-interaction, responsible of the non-abelian  
 296 nature of **QCD**. The **QCD** Lagrangian density is then given by:

$$\begin{aligned} \mathcal{L}_{\text{SM}} \supset \mathcal{L}_{\text{QCD}} &= -\frac{1}{2} \text{Tr} \{ G_{\mu\nu} G^{\mu\nu} \} + \sum_{\text{flavours}} i\bar{q}_f \gamma^\mu D_\mu q_f \\ &= -\frac{1}{4} \sum_{a=1}^8 G_{\mu\nu}^a G_a^{\mu\nu} + \sum_{\text{flavours}} i\bar{q}_f \gamma^\mu D_\mu q_f \end{aligned}$$

297 **Renormalisation** As mentioned, the **SM** is a renormalisable **QFT**. What this term refers  
 298 to is briefly detailed in the following. Higher-order effects introduce quantum corrections,  
 299 e.g., in the calculation of couplings in the **SM**, which must be taken into account. At the  
 300 same time, the particles in these loops have unbounded momenta, therefore divergences  
 301 arise in the calculations for both low (Infrared (**IR**)) and high (Ultraviolet (**UV**)) momenta,  
 302 which must be eliminated for the theory to be consistent with experimental measure-  
 303 ments. The process by which divergences disappear or are 'absorbed' by adding a scale  
 304 dependence to parameters such as couplings or particle masses, is known as renormalisa-  
 305 tion. In this way the physical lagrangian, with couplings comparable to experiments, can  
 306 be written as a bare lagrangian, minus a lagrangian containing the divergence-removing  
 307 terms, at the cost of introducing a scale dependence  $\mu$  of the momentum. Therefore, the  
 308 renormalisation results in the couplings (and other observables) being non-consistent and  
 309 varying with  $\mu$ . The phenomenon of asymptotic freedom and colour confinement in **QCD**  
 310 are consequences of this renormalisation process, which is in turn a property of gauge  
 311 theories.

312 **The running coupling constant  $\alpha_s$**  One of the consequences of the non-abelian nature  
 313 of **QCD** appears on the renormalisation of the coupling constant  $\alpha_s$  via the vacuum polar-  
 314 isation diagrams, which ends up depending on the scale  $Q$  of interaction. For Quantum  
 315 Electrodynamics (**QED**), the vacuum polarisation is induced by virtual  $e^+e^-$  pairs, which  
 316 (shield) the electric charge and result in the coupling decreasing with distance. In con-  
 317 trast, gluons not only produce  $q\bar{q}$ -pairs (which cause an effect similar to **QED**) but also  
 318 create additional gluon pairs, which tend to anti-screen the apparent colour charge. In the  
 319 high-energy regime (small distances), the coupling constant can be approximated with a

320 1-loop calculation in perturbative QCD, as follows:

$$\alpha_s(Q^2) = \frac{\alpha_s(Q_0^2)}{1 + (11N_C - 2N_f) \frac{\alpha_s(Q_0^2)}{12\pi} \log\left(\frac{Q^2}{Q_0^2}\right)} = \frac{12\pi}{(33 - 2N_f) \log\left(\frac{Q^2}{\Lambda_{\text{QCD}}^2}\right)}, \quad (1.19)$$

321 where  $N_C$  is the numbers of colors in the theory (3),  $N_f$  is the number of active flavours<sup>1</sup>,  
 322  $\alpha_s(Q_0)$  is the value of the coupling constant at a fixed scale  $Q_0$ , determined experimentally  
 323 at the  $Z$  mass value squared, and  $\Lambda_{\text{QCD}}$  is the cut-off IR scale, where the perturbative  
 324 approximation in  $\alpha_s$  stops being valid. Experimental measurements, compared to the  
 325 theory prediction, of the running coupling constant  $\alpha_s$  is shown in Figure 1.2, showing  
 326 the excellent agreement between both.

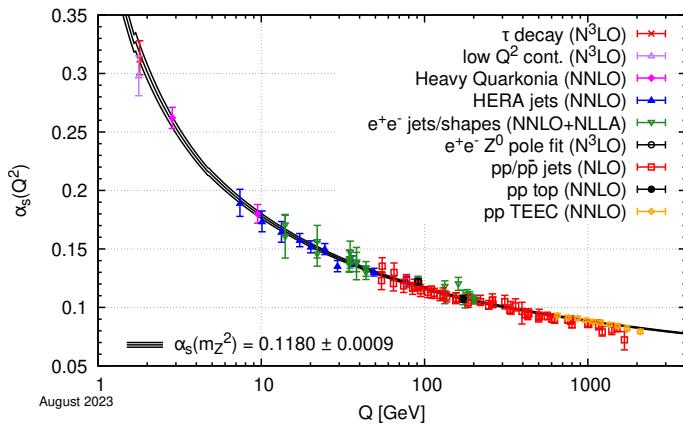


Figure 1.2: Experimental measurements of the coupling constant of QCD compared to the coupling computed at five loops [13].

327 **Asymptotic freedom and confinement** The coupling constant is said to run, being large  
 328 at low energy and becoming smaller at high energy. From Eq. 1.19, at high energies  
 329  $\alpha_s \rightarrow 0$  therefore QCD interacts weakly, allowing the quarks as unbounded particles, phe-  
 330 nomenon known as asymptotic freedom [14, 15]. On the other hand, for low energies  
 331 ( $Q^2 \rightarrow 0$ ), the coupling  $\alpha_s$  increases divergently, and therefore QCD is strongly interacting  
 332 leading to the confinement of quarks and gluons [16]. Confinement implies that neither  
 333 quarks nor gluons can appear in isolation, they can only exist within colourless composite  
 334 "partons", called hadrons. Moreover, starting from the infrared cut-off scale  $\Lambda_{\text{QCD}}$ , where  
 335 perturbative approximation at  $\alpha_s$  is no longer valid, the creation of quark-antiquark pairs  
 336 in the vacuum is more energetically favourable than the separation of a pair of bound  
 337 quarks. For this reason, as they lose energy, the quarks and gluons produced in a proton  
 338 collider undergo a repetitive process known as hadronisation, in which collimated cas-  
 339 cades of hadrons, called jets, are created, forming a cone from the initial quark or gluon  
 340 to the calorimeters, where all their energy is deposited.

<sup>1</sup> Those quarks with  $m_q \ll Q$ , where  $m_q$  is the quark mass after the process of EWSB produced by the Higgs boson.

### 341 1.1.3 Hadron interactions in proton-proton colliders

342 As discussed in Section 1.1.2, the coupling constant  $\alpha_s$ , which governs the strong inter-  
 343 actions between quarks, has a strong dependence on the energy scale of each interaction,  
 344 radically modifying the nature of the processes. The modelling of a proton-proton colli-  
 345 sion in an experiment like ATLAS, where it is necessary to know its evolution from the  
 346 interaction between the protons at  $\sqrt{s} \sim$  TeV, to the interaction of the particles in the  
 347 final state with the active and passive materials of the detector at a few GeV, represents a  
 348 huge challenge, as it covers very different behaving QCD regimes. Given that the LHC is a  
 349 proton collider, it is mandatory to have a very precise description of the proton structure,  
 350 as a  $pp$  collision at very high energies is basically to collide the constituents of them.

351 At very high energies, but within the perturbative regime, the collision between two pro-  
 352 tons can be studied via the Parton Model. This model has been introduced by Feyn-  
 353 man [17] and Bjorken [18] in the late-1960s, to interpret electron-nucleon deep inelastic  
 354 scattering at SLAC. This description has proven to be a good approximation for parton-  
 355 parton interactions with large momentum transfer (i.e. Bjorken scaling [19]) but is not  
 356 appropriate for modelling the interaction at low energies. Under this abstraction, the par-  
 357 tons include not only the valence quarks ( $u$ ,  $\bar{u}$  and  $d$  in the case of the proton), but also  
 358 the pairs of particles and antiparticles in the quark sea, and the gluons that mediate the  
 359 interactions between them. The model assumes a permanent interaction between partons,  
 360 so their individual momentum is unknown, although their fraction of momentum with re-  
 361 spect to the total hadron momentum can be modelled as a random variable. Furthermore,  
 362 in the case of experimental verification, the quarks and gluons in the final state are not  
 363 directly observed due to hadronisation (concept discussed in Section 1.3.2). Instead, an  
 364 effective hadronic cross section,  $\sigma(pp \rightarrow jj)$ , is calculated between the incident protons  
 365 and the final state jets. To perform this passage, the factorisation theorem [17, 20–22]  
 366 is used, which allows a systematic separation between the short-distance interactions (of  
 367 the partons), and the long-distance interactions (responsible for colour confinement and  
 368 hadron formation). This theorem states that the total cross-section for two hadrons can  
 369 be obtained by weighting and combining the cross-sections for two particular partons.  
 370 This weighting is done using  $f_i(x, Q^2)$ , the Parton Distribution Functions (PDFs), which  
 371 describe the parton density for a parton of species  $i$  in a hadron, with a fraction  $x$  of the  
 372 hadron energy-momentum when the hadron is probed at a resolution scale  $Q^2$ . The cross-  
 373 section for a hard scattering process  $pp \rightarrow X$ , initiated by two hadrons with four-momenta  
 374  $P_1$  and  $P_2$  can be written as:

$$\sigma_{pp \rightarrow X} = \sum_{ij} \int_0^1 dx_1 dx_2 f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \hat{\sigma}_{ij} \left( p_1, p_2, \alpha_s(\mu_R^2), Q^2/\mu_R^2, Q^2/\mu_F^2 \right), \quad (1.20)$$

375 where  $x_1$  and  $x_2$  are the momentum fractions carried by the interacting partons, and  
 376  $p_1 = x_1 P_1$  and  $p_2 = x_2 P_2$  are the interacting parton momenta. The partonic cross-section  
 377  $\hat{\sigma}_{ij}$ , corresponding to the interaction of partons  $i$  and  $j$ , is calculated at a fixed order in

$\alpha_s$ , which is evaluated at some renormalisation scale,  $\mu_R$  and factorisation scale  $\mu_F$ . The renormalisation scale  $\mu_R$  is important to absorb UV divergences in calculations at higher orders. The total cross-section is obtained by summing over all possible parton flavours and integrating over all possible momentum fractions. The parton distribution functions,  $f_i$  and  $f_j$ , are evaluated at a factorisation scale,  $\mu_F$ , which can be thought of as the scale that separates short-distance, perturbative physics, from long-distance, non-perturbative physics (i.e., separates hard and soft processes).

If the perturbative expansion were carried to all orders, the cross-section in Eq. 1.20 would be independent of  $\mu_F$  and  $\mu_R$ . However, in actual finite order calculation this does not hold. They are usually both taken to be equal,  $\mu_F = \mu_R = \mu$ , chosen at the typical scale  $Q^2$  of the process, in order to minimise the contribution of uncalculated higher order terms, whose forms are logarithmic  $\log(Q^2/\mu_R^2)$  and  $\log(Q^2/\mu_F^2)$ . The dependence of the prediction on  $\mu_R$  and  $\mu_F$  is assigned as a theoretical uncertainty. The fact that the cross-section of a process should be independent of the factorisation scale  $\mu_F$  led to the DGLAP equations (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) [23–25]. These equations determine the evolution of the PDF with  $Q^2$ . For the case of the proton, Figure 1.3 shows the PDFs evaluated at two different factorisation scales for all possible partons.

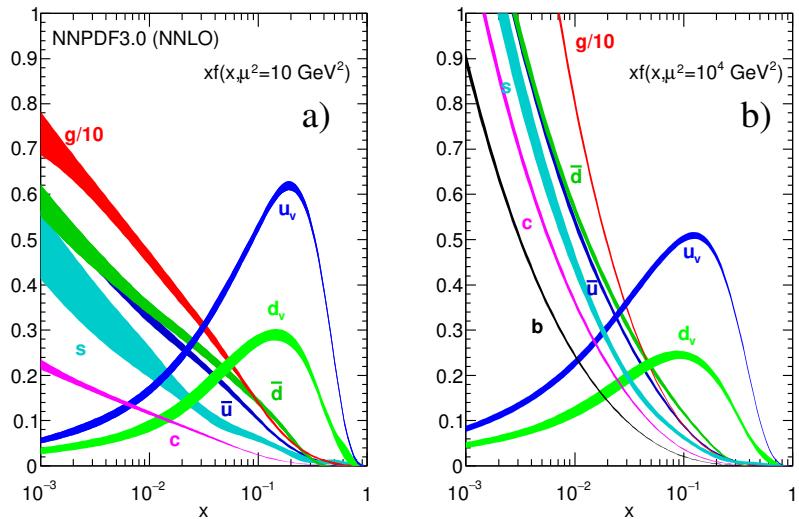


Figure 1.3: Parton momentum fraction  $x$  times the unpolarized parton distributions  $f_i(x, Q^2)$  (where  $i = u_v = u - \bar{u}$ ,  $d_v = d - \bar{d}$ ,  $\bar{u}$ ,  $\bar{d}$ ,  $s \simeq \bar{s}$ ,  $c = \bar{c}$ ,  $b = \bar{b}$ ,  $g$ ) obtained in the NNLO NNPDF3.0 global analysis [26] at scales  $\mu^2 = 10 \text{ GeV}^2$  (left) and  $\mu^2 = 10^4 \text{ GeV}^2$  (right) with  $\alpha_s(M_Z^2) = 0.118$ . Figures extracted from Ref. [27].

### 395 Process description

Initially two hadrons are coming in on a collision course, where each hadron can be thought as a group of essentially collinear partons quantitatively characterised by the parton distributions. In a collision scenario with accelerated particles carrying EM and colour charges, bremsstrahlung can occur, e.g. as gluon radiation such as  $q \rightarrow qg$ . A collision between two partons, one from each side, takes place producing the hard process of interest, that can be calculated by a perturbative approach to some order in  $\alpha_s$ , which

402 corresponds to the number of outgoing partons. Emissions that are started off from the  
 403 two incoming colliding partons are called Initial State Radiation (ISR), while radiations  
 404 from the outgoing partons are referred as Final State Radiation (FSR). With the parton  
 405 shower development, the colour field strength increases as partons loose energy and they  
 406 can break up by the production of quark-antiquark pairs. Thus, quarks and antiquarks may  
 407 combine to produce a primary hadron. The creation of hadrons as a consequence of the  
 408 confinement phenomenon is referred to as “hadronisation”. The additional products of the  
 409 collision that are not explicitly related to the hard process (radiation, hadron remnants,  
 410 products of multiple parton interactions, etc.), are generally grouped altogether and called  
 411 Underlying Event (UE). A visualisation of the  $pp$  collision is shown in Figure 1.4.

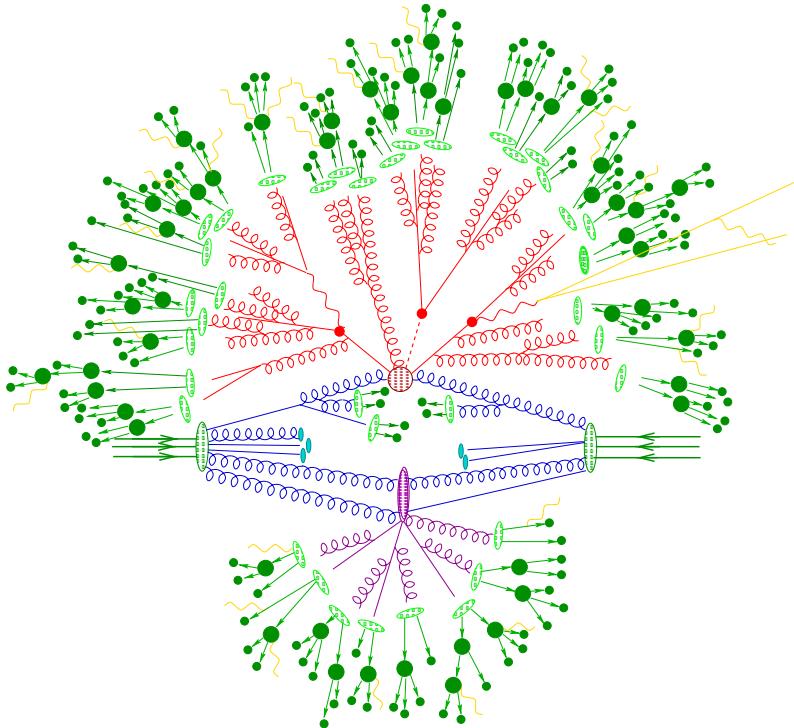


Figure 1.4: Illustration of the stages of a hadron-hadron collision. The red circle in the center of the figure represents the hard collision, surrounded by a tree-like structure representing bremmstrahlung radiation as simulated by parton showers. The purple blob at the bottom represents the UE. The hadronisation process is represented by the light green blobs, dark green blobs indicate hadron decays, while yellow lines signal soft photon radiation [28].

412 Over the years, different LHC experiments have measured cross sections of different SM  
 413 processes. Figure 1.5 shows the good agreement between the ATLAS-measured cross sec-  
 414 tions of some processes and their theoretical predictions.

#### 415 1.1.4 Theory of prompt-photon production

416 High transverse momentum (“prompt”) photons constitute colourless probes of the hard  
 417 interaction and their production in proton-proton collisions,  $pp \rightarrow \gamma + X$ , provides a  
 418 testing ground for QCD, whose measurement offers certain advantages over other analyses  
 419 in jet production events, the most abundant process in single hadron colliders. In this case,

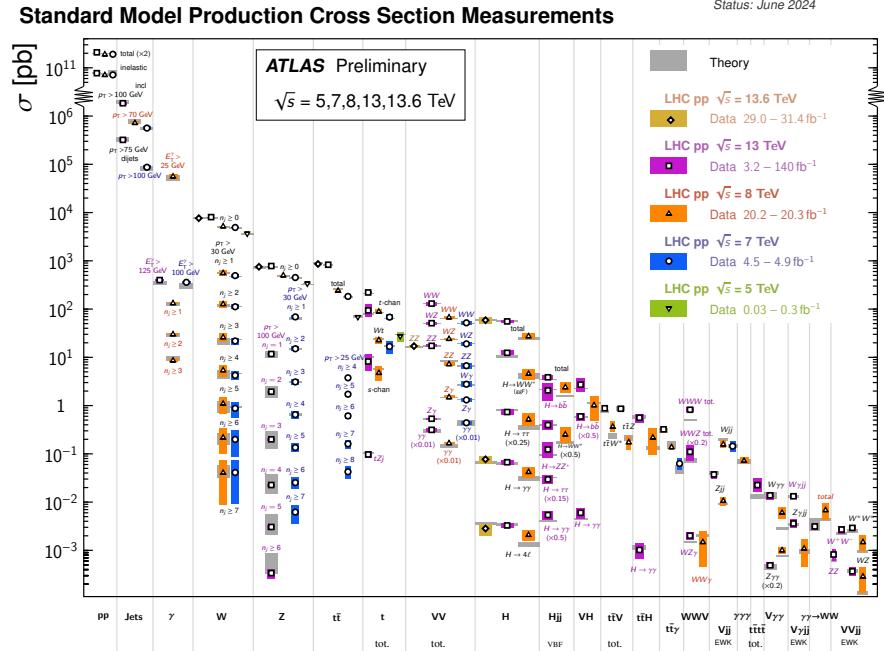
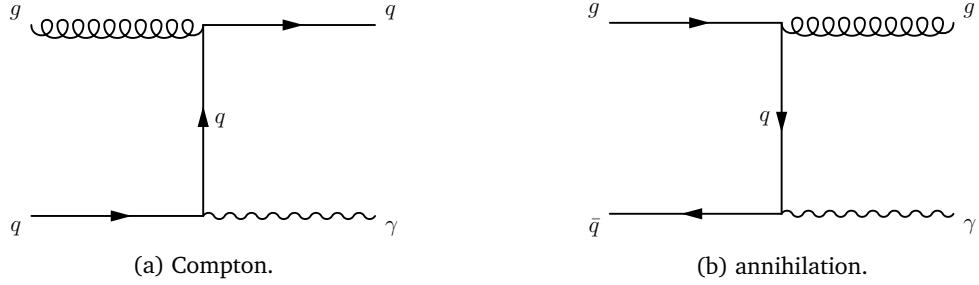


Figure 1.5: Summary of several Standard Model total and fiducial production cross-section measurements, compared against their theoretical predictions [29].

- 420 the presence of a QED vertex at Leading Order (LO) makes the theoretical calculations  
 421 more reliable and gives access to a lower range of  $p_T$ . Moreover, the energy resolution of  
 422 electromagnetic calorimeters are in general better than those of the hadronic calorimeter<sup>2</sup>,  
 423 and systematic uncertainties in the photon energy scale are smaller. Due to the fact that  
 424 photons do not hadronise (see Section 1.3.2), the direction and energy of photons is  
 425 straightforwardly measured in the calorimeter without the need for a jet algorithm to  
 426 reconstruct a jet.
- 427 Prompt-photon production proceeds via two processes: the direct-photon process (D), in  
 428 which the photon arises directly from the hard interaction, and the fragmentation-photon  
 429 process (F), in which the photon is emitted in the fragmentation of a high transverse  
 430 momentum parton [30, 31]. From a topological point of view, when a direct photon is  
 431 produced, it is most likely that it will be separated from the hadronic activity, whereas a  
 432 photon produced from a fragmentation process, is most probably accompanied by had-  
 433 rons.
- 434 At LO in perturbation theory, there are two subprocesses that contribute to the direct-  
 435 photon production: (a) the Compton process  $qg \rightarrow \gamma q$ , and (b) the annihilation process  
 436  $q\bar{q} \rightarrow \gamma + g$ , shown in Figures 1.6a and 1.6b. At medium and large  $x$ , there is a natural  
 437 hierarchy of parton distributions in the proton,  $q \gg g \gg \bar{q}$ , while at small  $x$ ,  $g \gg q, \bar{q}$ .  
 438 As a consequence, in proton-proton collisions, the  $qg$  Compton process dominates over  
 439 essentially all the  $p_T$  range. This makes direct photon production particularly useful for  
 440 constraining the gluon distribution.

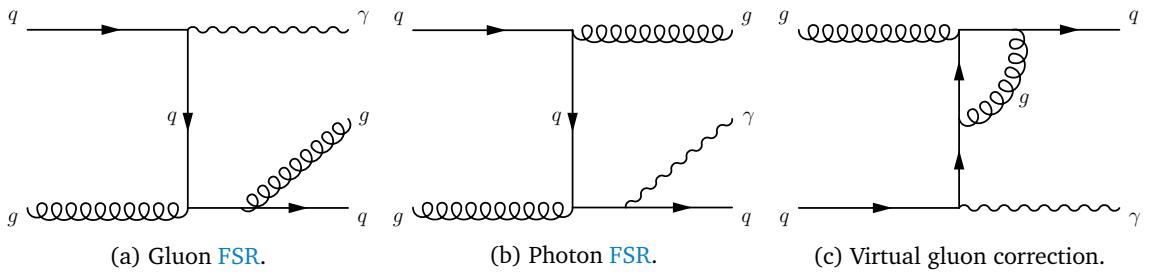
<sup>2</sup> A description of both calorimeters is given in Chapter 2.

Figure 1.6: Feynman diagrams for the LO direct-photon production in  $pp$  collisions.

441 Next-to-Leading Order (NLO) corrections to this process are represented in Figure 1.7. In  
 442 Figure 1.7a, there is a collinear singularity when the momenta of the final-state quark and  
 443 gluon are parallel. This divergence cancels when real and virtual gluon contributions (see  
 444 Figure 1.7c) are summed, and the net effect is a finite  $\mathcal{O}(\alpha_s)$  correction to the LO process.  
 445 On the other hand, in the diagram of Figure 1.7b there is another collinear singularity, this  
 446 time, when the photon and quark momenta are parallel. This singularity, however, does  
 447 not cancel, but has to be absorbed into a photon fragmentation function  $D_q^\gamma(z, \mu_f^2)$  that  
 448 represents the probability of finding a photon carrying longitudinal momentum fraction  
 449  $z$  in a quark jet at scale  $\mu_f$ . This fragmentation function is not calculable in perturbation  
 450 theory, and obeys a DGLAP evolution equation similar to that for the hadron fragmentation  
 451 functions. The contribution to the cross section from Figure 1.7b contains a piece of the  
 452 form

$$\hat{\sigma}(qg \rightarrow qg) \oplus D_q^\gamma(z, \mu_f^2). \quad (1.21)$$

453 The photon-fragmentation contribution appears when a final-state quark-photon collin-  
 454 ear singularity occurs in the calculation of the contribution from subprocesses such as  
 455  $qg \rightarrow qg\gamma$ . At higher orders, multiple final-state collinear singularities appear in any  
 456 subprocess where a high- $p_T$  parton undergoes a cascade of successive collinear splittings  
 457 ending up with a quark-photon splitting. These singularities are factorised to all orders  
 458 in  $\alpha_s$  according to the factorisation theorem, and are absorbed into quark and gluon frag-  
 459 mentation functions of the photon,  $D_q^\gamma(z, \mu_f^2)$  and  $D_g^\gamma(z, \mu_f^2)$ , respectively.

Figure 1.7: Feynman diagrams for direct-photon production at NLO in  $pp$  collisions.

460 The photon fragmentation function increases uniformly with the scale over the whole  $z$   
 461 range, i.e.  $D_k^\gamma(z, \mu_f^2) \sim d^\gamma(z) \ln(\mu^2)$  as  $\mu^2 \rightarrow \infty$ . When the  $p_T$  is large with respect to  
 462  $\sim 1$  GeV, the  $\ln p_T^2$  growth of the fragmentation function in Eq. 1.21 compensates one of  
 463 the  $\alpha_s(p_T^2)$  couplings in the subprocess cross section, and the contribution is effectively of

464 order  $\alpha_s(p_T^2) \alpha_{EM}$ , i.e. the same as the LO contribution. Feynman diagrams correspond-  
 465 ing to the LO fragmentation component are shown in Figure 1.8.

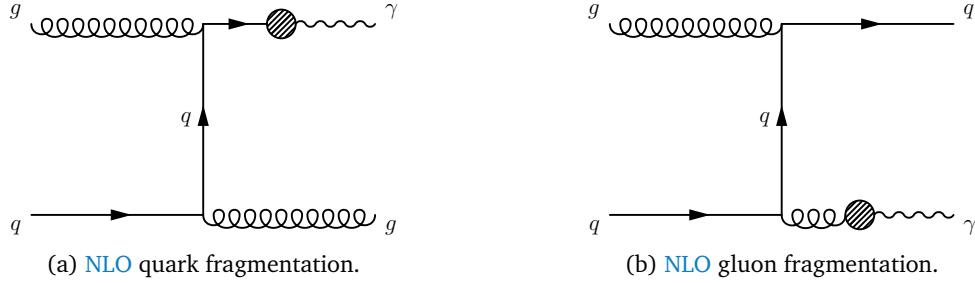


Figure 1.8: Feynman diagrams for the LO fragmentation-photon processes in  $pp$  collisions (a)  $qg \rightarrow gq(\gamma)$  and (b)  $qg \rightarrow qg(\gamma)$ .

466 The inclusive differential cross section in  $E_T^\gamma$  for the production of a non-isolated photon  
 467 is given by the sum of the direct and fragmentation contributions

$$\begin{aligned} \frac{d\sigma}{dE_T^\gamma} &= \frac{d\sigma_{\text{dir}}}{dE_T^\gamma} + \frac{d\sigma_{\text{frag}}}{dE_T^\gamma} \\ &= \sum_{a,b=q,\bar{q},g} dx_a dx_b f_a(x_a, \mu_F^2) f_b(x_b, \mu_F^2) \times \\ &\quad \left[ d\hat{\sigma}_{ab}^\gamma(p^\gamma; x_a, x_b, \mu_R, \mu_F, \mu_f) + \sum_{c=q,\bar{q},g} \int_{z_{\min}}^1 \frac{dz}{z^2} d\hat{\sigma}_{ab}^c(p^\gamma; x_a, x_b, z, \mu_R, \mu_F, \mu_f) D_c^\gamma(z, \mu_f^2) \right] \end{aligned} \quad (1.22)$$

468 where  $D_c^\gamma(z, \mu_f^2)$  is the fragmentation function of a parton  $c$  to a photon carrying mo-  
 469 mentum fraction  $z$ ,  $f_a(x_a, \mu_F^2)$  is the PDF of a parton  $a$ ,  $\mu_R$  and  $\mu_F$  are the standard  
 470 renormalisation and factorisation scales, and  $\mu_f$  is the fragmentation scale. Corrections  
 471 to the direct component of the partonic cross section  $\hat{\sigma}_{ab}^\gamma$  are known up to the Next-to-  
 472 Next-to-Leading Order (NNLO) in Perturbative QCD (pQCD), while the fragmentation  
 473 component  $\hat{\sigma}_{ab}^c$  is only known at NLO.

474 At LO, the theory calculations for the direct and fragmentation processes converge sep-  
 475 arately, and can be considered independently. However, this distinction has no physical  
 476 meaning beyond the LO, since both kinds of processes need to be considered at the same  
 477 time to cancel the final-state infrared and collinear singularities. Therefore, beyond the  
 478 LO, both direct and fragmentation processes cannot be considered separately. From a the-  
 479 oretical point of view, the distinction is defined by an arbitrary choice. It follows from the  
 480 necessity of factorising the final-state collinear singularities and absorbing them into the  
 481 fragmentation functions. This factorisation requires the introduction of an arbitrary frag-  
 482 mentation scale  $\mu_f$ , which is a non-physical parameter. More generally, it relies on the ar-  
 483bitrary choice of the factorisation scheme, which defines the finite part of the higher-order  
 484 corrections that is absorbed in the fragmentation functions together with the singularities;  
 485 the remaining finite part is then included in the higher-order contributions to the partonic

486 cross sections. The dependence on this arbitrariness, and in particular, on  $\mu_f$ , cancels only  
 487 in the sum of the direct and fragmentation contributions, so only this sum is a physical  
 488 observable.

## 489 1.2 Physics Beyond Standard Model (BSM)

490 The previous section briefly described most of the properties of the **SM**, together with  
 491 **ATLAS** results showing how well the **SM** agrees with experimental data. Despite being  
 492 one of the most successful theories in physics in general, the model naturally has a range  
 493 of validity. However, it cannot be considered the final theory (the one that could "explain  
 494 everything"), as it has certain limitations, both from a theoretical and an experiential  
 495 point of view. The **SM** is still regarded as an effective theory, a low-energy approximation  
 496 of a more fundamental theory. There are three popular types of new physics theories:  
 497 (i) models with an extended (family) symmetry or scalar sector, (ii) higher dimensional  
 498 theory, and (iii) quark-lepton compositeness (namely, the **SM** fermions are not elementary  
 499 anymore [32–37]). In the following, a general overview of the main shortcomings of the  
 500 **SM** are presented. After that, the two last types of new physics theories are discussed,  
 501 enumerating the theoretical models used in the search carried out in this thesis.

- 502     • Gravity: One of the main limitations of the **SM** is the impossibility of including gravity  
 503       in the same way as other interactions. Not only is including gravity in the theory  
 504       not enough to explain the observations, but the mathematics used in the **SM** is prac-  
 505       tically incompatible with the formulation of General Relativity.
- 506     • Hierarchy Problem: In the context of high energy physics, a hierarchy problem occurs  
 507       when the fundamental value of some physical parameter (such as a coupling  
 508       constant or a mass), in some Lagrangian is vastly different from its effective value,  
 509       which is the value that gets measured in an experiment. Typically the renormalised  
 510       value of parameters are close to their fundamental values, but in some cases, it ap-  
 511       pears that there has been a delicate cancellation between the fundamental quantity  
 512       and the quantum corrections. In general, hierarchy problems are related to fine-  
 513       tuning of the parameters in the theory. The most well-known case in particle physics  
 514       is the difference on the **EW** scale  $M_W \sim 10^2$  GeV and Planck scale, where quantum  
 515       gravity effects start to take over  $M_P \sim 10^{19}$  GeV, whose ratio is  $M_W/M_P \sim 10^{-17}$ .
- 516     • Dark Matter (DM): A hint towards the incompleteness of the **SM** is the presence of  
 517       **DM**. Based on astrophysical measurements and cosmological considerations [38–  
 518       42], known matter accounts only for 4% of the total of the universe. On the other  
 519       hand, 23% of the total matter is associated with a type of unknown matter, referred  
 520       as **DM**, since it does not emit **EM** radiation, but is massive as it has considerable  
 521       gravitational effects on visible matter. The only **SM** particle that could be a viable **DM**

522 candidate is the neutrino, but as its mass is too small to explain these phenomena,  
 523 it has been discarded.

524 • Neutrino's masses: The observation of neutrino oscillation implies that although  
 525 neutrinos have a very small mass, it is not zero, in contrast to the **SM** prediction.  
 526 Although there are several mechanisms for including them in the **SM**, there is insuf-  
 527 ficient evidence to know which is the correct form, and some models propose the  
 528 existence of new, yet unobserved, heavy particles [43–45].

### 529 1.2.1 Quark compositeness theories

530 In quark compositeness theories, the quarks are no longer the fundamental constituents  
 531 of matter, but rather are bound states of particles often termed *preons* [46]. The latter  
 532 are postulated to experience a hitherto unknown force on account of an asymptotically  
 533 free but confining gauge interaction [47], which becomes very strong at a characteristic  
 534 scale  $\Lambda$ , thereby leading to the aforementioned composites. In many such models [48–  
 535 50], though not all, quarks and leptons share at least some common constituents. Such  
 536 a hypothesis naturally leads to the existence of excited fermion states at a mass scale  
 537 comparable to the dynamics of the new binding force.

538 As the "excited states" do undergo the **SM** gauge interactions, they may be produced at  
 539 colliders operating at high enough energies. On production, they would decay into **SM**  
 540 particles, with a particularly favorable channel being the radiative decay into an ordinary  
 541 fermion and a gauge boson (photon,  $W$ ,  $Z$ , or gluon). If quarks and leptons are not fun-  
 542 damental constituents but only composites, this fact could, in principle, be revealed either  
 543 through an accumulation of statistics at energy scales comparable to the compositeness  
 544 scale  $\Lambda$  at the **LHC**. If  $\Lambda$  is not too high then Excited Quark (**EQs**) can be produced on  
 545 shell, while at energies well below  $\Lambda$ , such excitations could manifest themselves through  
 546 an effective four fermion contact interaction involving **SM** particles alone.

547 In general, the interactions between the **EQs** ( $q^*$ ) and gauge bosons can be written as [37]:

$$\mathcal{L}_{\text{gauge}} = \frac{1}{2\Lambda} \overline{q_R^*} \sigma^{\mu\nu} \left[ g_s f_s \frac{\lambda_a}{2} G_{\mu\nu}^a + g f \frac{\tau}{2} W_{\mu\nu} + g' f' \frac{Y}{2} B_{\mu\nu} + \right] q_L + \text{H.c.} \quad (1.23)$$

548 where  $G_{\mu\nu}^a$ ,  $W_{\mu\nu}$  and  $B_{\mu\nu}$  are the field strength tensors of the SU(3), SU(2) and U(1)  
 549 gauge fields, respectively. The coefficients  $g_s$ ,  $g = e/\sin\theta$ ,  $g' = e/\cos\theta$  are the strong  
 550 and electroweak gauge couplings,  $\lambda_a$  is the Gell-Mann matrix,  $\tau$  is the Pauli matrix, and  
 551 the weak hypercharge is  $Y = 1/3$ , respectively.  $\Lambda$  is compositeness scale and  $f_s$ ,  $f$ ,  $f'$  are  
 552 parameters determined by composite dynamics, which represent the strength of the inter-  
 553 actions between the **EQs** and their **SM** partners. The  $s$  and  $t$ -channel Feynman diagrams  
 554 for such process are presented in Figure 1.9. Finally, the decay width of **EQs** to a photon  
 555 and a quark can be calculated at **LO** [37]:

$$\Gamma(q^* \rightarrow q\gamma) = \frac{1}{4} \alpha \left( f \tau_3 + f' \frac{Y}{2} \right)^2 \frac{m_q^{3*}}{\Lambda^2}. \quad (1.24)$$

556 which increases with the **EQ** mass  $m_{q^*}$  if one considers  $\Lambda = m_{q^*}$ .

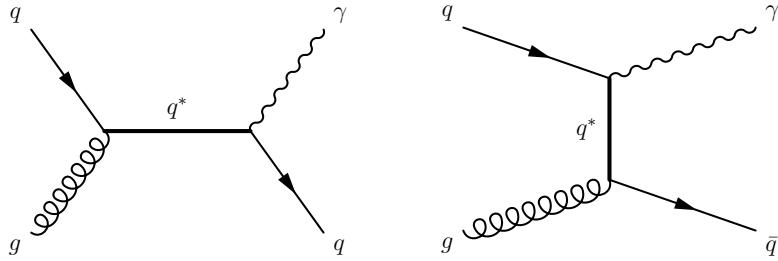


Figure 1.9: Feynman diagrams of the **EQ** production in  $pp$  collisions and decay into a quark and a photon in the  $s$ -channel (left) and  $t$ -channel (right).

557 In the **SM** there is not a resonance production process decaying into a photon+jet pair in  
 558  $pp$  collisions, and direct photon+jet production at tree level occurs via Compton scattering  
 559 or  $q\bar{q}$  annihilation, as described in Section 1.1.4. As a result, the  $\gamma$ +jet invariant mass  
 560 ( $m_{\gamma+j}$ ) distribution is rapidly falling; thus, the  $\gamma$ +jet production mediated by a heavy **EQ**  
 561 may be discovered if it exists. Hereinafter, in the context of this thesis, **EQ** models are only  
 562 studied with  $\gamma$ +jet decays. In Chapter 7, information regarding the signals cross sections  
 563 is given.

### 564 1.2.2 Higher dimensional theories

565 There are at least two seemingly fundamental energy scales in nature, the electroweak  
 566 scale  $m_W \sim 10^3$  GeV and the Planck scale  $m_P = G^{-1/2} \sim 10^{18}$  GeV, where  $G$  is the  
 567 Gravitational constant. Explaining the enormity of the ratio  $m_P/m_W$  has been the prime  
 568 motivation for constructing extensions of the **SM** such as models with technicolor or low-  
 569 energy supersymmetry. It is remarkable that these rich theoretical structures have been  
 570 built on the assumption of the existence of two very disparate fundamental energy scales.  
 571 However, there is an important difference between these scales. While electroweak inter-  
 572 actions have been probed at distances approaching  $\sim m_W^{-1}$ , gravitational forces have not  
 573 remotely been probed at distances  $\sim m_P^{-1}$ .

574 Proposals for a spacetime with more than three spatial dimensions date back to the  
 575 1920s, mainly through the work of Kaluza and Klein, in an attempt to unify the forces  
 576 of nature [51]. Although their initial idea failed, the formalism that they and others de-  
 577 veloped is still useful nowadays. Around 1980, string theory proposed again to enlarge  
 578 the number of space dimensions, this time as a requirement for describing a consistent  
 579 theory of quantum gravity. The extra dimensions were supposed to be compactified at a  
 580 scale close to the Planck scale, and thus not testable experimentally in the near future.

581 A different approach was given by Arkani-Hamed, Dimopoulos, and Dvali (ADD) [52],  
 582 where they showed that the weakness of gravity could be explained by postulating two  
 583 or more flat extra dimensions in which only gravity could propagate. The size of these  
 584 extra dimensions should range between roughly a millimeter and  $\sim 1/\text{TeV}$ , leading to  
 585 possible observable consequences in current and future experiments. Another approach,

586 by Randall and Sundrum (RS) [53, 54], postulates a five-dimensional Anti-deSitter (AdS)  
 587 spacetime with warped geometry, where the compactification is of the scale of  $1/\text{TeV}$ .

588 These low-scale gravity models [52–57] allow for the production of small black holes  
 589 (Quantum Black Holes (QBHs)) in particle collisions [58–60]. QBHs, unlike semiclassical  
 590 ones, show significant differences as their mass approaches the Planck scale. Semiclassical  
 591 black holes decay thermally, losing mass at the Hawking temperature with minimal effect  
 592 on the surrounding spacetime. However, as the black hole's mass decreases and nears the  
 593 Planck scale, the influence of back-reaction on the spacetime becomes substantial, and the  
 594 black hole can no longer maintain thermal equilibrium with its radiation. Microcanonical-  
 595 ical corrections help refine the decay model, but eventually quantum mechanical effects  
 596 dominate. When the black hole's Compton wavelength surpasses its Schwarzschild radius,  
 597 quantum behavior begins to emerge, potentially giving the black hole particle-like prop-  
 598 erties. At this point, the concepts of a well-defined temperature and entropy break down,  
 599 making it unlikely that these black holes will decay thermally [61–63].

600 Focusing on black holes with mass slightly above the Planck scale, it's expected that QBH  
 601 decays will not follow a thermal pattern. Instead, decays into only a few particles will  
 602 likely dominate, and these processes will take place in a small region of spacetime. The  
 603 QBH might behave like a strongly coupled resonance or a gravitationally bound state.  
 604 After the black hole decays, the QCD hadronization process will take place, given the  
 605 involvement of color charges.

606 In  $pp$  collisions, only a fraction of the total center of mass energy  $\sqrt{s}$  is available in the  
 607 hard-scattering process. By defining  $sx_a x_b \equiv s\tau \equiv \hat{s}$ , where  $x_a$  and  $x_b$  are the fractional  
 608 energies of the two colliding partons (see Section 1.1.3), the full cross section  $\sigma$  reads [64]:

$$\sigma_{pp \rightarrow \text{BH}+X}(s) = \sum_{a,b} \int_{m^2/s}^1 d\tau \int_\tau^1 \frac{dx}{x} f_a\left(\frac{\tau}{x}\right) f_b(x) \Theta(m - m_{\text{th}}) \hat{\sigma}_{ab \rightarrow \text{BH}}(\hat{s} = m^2),$$

609 where  $a$  and  $b$  go through all the partons, and  $f_a$  and  $f_b$  are the PDFs of them. The  
 610 Heaviside step function  $\Theta$  marks the minimum mass threshold  $m_{\text{th}}$  at which QBHs could  
 611 be produced. The threshold is typically taken to be the Planck scale  $m_P$  for QBHs, or a few  
 612 times  $m_P$  for classical black holes. For QBHs the overall range in which they are considered  
 613 to be produced is  $m_P \leq m \leq 3m_P$  [65]. The parton-level cross section  $\hat{\sigma}$  is most often  
 614 taken to be the geometrical cross-section  $\sigma \sim \pi r_g^2$  with

$$r_g = k(D) \frac{1}{m_P} \left( \frac{m}{m_P} \right)^{\frac{1}{D-3}},$$

615 where  $k(D)$  is a numerical coefficient depending only on the number of dimensions and  
 616 the definition of the fundamental Planck scale:

$$k(D) = \left( 2^{D-4} (\sqrt{\pi})^{D-7} \frac{\Gamma(\frac{D-1}{2})}{D-2} \right)^{\frac{1}{D-3}}$$

617 Based on current experimental and phenomenological limits on the Planck scale, it is  
 618 unlikely that semiclassical black holes will be accessible at energies produced by the LHC.

619 However, if the Planck scale is low enough, [QBHs](#) may be produced in abundance at  
 620 the [LHC](#), and these would appear as resonances in the invariant mass of the final state  
 621 particles. Concerning only the  $\gamma$ +jet final state, there are six non-thermal black hole  
 622 states:

$$\begin{aligned} u + g &\rightarrow QBH^{2/3} \rightarrow u + \gamma \\ \bar{d} + g &\rightarrow QBH^{1/3} \rightarrow \bar{d} + \gamma \\ q + \bar{q} &\rightarrow QBH^0 \rightarrow g + \gamma \\ q + g &\rightarrow QBH^0 \rightarrow g + \gamma \\ d + g &\rightarrow QBH^{-1/3} \rightarrow d + \gamma \\ \bar{u} + g &\rightarrow QBH^{-2/3} \rightarrow \bar{u} + \gamma, \end{aligned}$$

623 where  $u$  represents all up-type quarks,  $d$  all down-type quarks and  $q$  all quark flavours.  
 624 Similarly as the [EQ](#) model, a more details on the samples is given in Chapter 7.

## 625 1.3 Monte Carlo ([MC](#)) simulations

626 The [MC](#) technique is a way of calculating difficult integrals that may be hard to solve  
 627 by ordinary numerical interpolation methods. High-energy collisions between elementary  
 628 particles normally produce complex final states, which are populated by many hadrons,  
 629 leptons, photons and neutrinos. The relation between the final states and the underlying  
 630 physics description is not simple due to the lack of understanding of the physics and the  
 631 fact that any analytical approach is not feasible due to the large particle multiplicities.  
 632 An additional difficulty is related to the need to simulate complicated geometrical factors  
 633 that represent detectors, a routine situation for experimenters. [MC](#) methods allow the  
 634 generation of complete events with final particles (i.e. hadrons, leptons and photons)  
 635 together with their momenta, with the same average behaviour and the same fluctuations  
 636 as the data. Whereas in the data the fluctuations arise from the quantum mechanical  
 637 character of the underlying theory, in generators these fluctuations are the result of the  
 638 (quasi-)randomness of the [MC](#) approach.

639 The main aspects of the simulated events are: Hard process, Parton Shower, Hadronisation  
 640 and [UEs](#), and it follows the schematic representation shown in Figure 1.4 The main [MC](#)  
 641 event generators used in this thesis are PYTHIA 8.1 [66], PYTHIA 8.2 [67], PYTHIA 8.3 [68]  
 642 and SHERPA 2.2.2 [69].

### 643 1.3.1 Hard interactions and parton shower

644 In order to describe a  $2 \rightarrow n$  process from the Lagrangian of the theory (where  $n$  repres-  
 645 ents a given number of partons in the final state), Feynman diagrams are drawn and eval-  
 646 uated using their specific rules in order to compute the Matrix Elements ([MEs](#)) in powers  
 647 of  $\alpha_s$ . As the number of partons in the final state increases, the number of Feynman

648 diagrams grows factorially, making higher-order calculations challenging. However, com-  
 649 plex processes can be simplified by factoring them into core  $2 \rightarrow 2$  processes, which are  
 650 convoluted with parton splitting probabilities to approximate higher-order effects. Simu-  
 651 lation programs implementing this approach are for instance PYTHIA and HERWIG. These  
 652 use [LO](#) perturbative calculations of matrix elements of  $2 \rightarrow 2$  processes and implement  
 653 higher-order [QCD](#) processes approximately via the so-called initial- and final-state Parton  
 654 Showers ([PSs](#)) [70, 71] to produce the equivalent of multi-parton final states.

655 In a hard process with virtuality  $Q^2$ , incoming and outgoing partons emit gluons in a pat-  
 656 tern where emissions diverge when gluons become collinear with quarks or when their  
 657 energy vanishes. Gluon branchings ( $g \rightarrow gg$ ) exhibit similar divergences, while  $g \rightarrow q\bar{q}$   
 658 does not. [NLO QCD](#) programs, such as SHERPA and POWHEG, must match [PSs](#) to the [ME](#)  
 659 calculation to avoid double-counting emissions. These emissions, ordered by increasing  
 660 virtuality, continue until they match the hard process's  $Q^2$ . [FSR](#) similarly decreases parton  
 661 virtuality until a lower cut-off ( $Q_0^2 \equiv \Lambda_{\text{QCD}} \sim 1 \text{ GeV}$ ) is reached, beyond which perturba-  
 662 tion theory loses relevance, and hadronization takes over.

### 663 1.3.2 Hadronisation

664 As the evolution reaches  $Q_0^2 = \Lambda_{\text{QCD}}$ , the [PS](#) phase is truncated since the coupling forces  
 665 become significant and confinement takes place. This phenomenon cannot still be de-  
 666 scribed from first principles, and therefore, it involves some modelling to transform all  
 667 the outgoing coloured partons into colourless hadrons of a typical 1 GeV mass scale. The  
 668 dynamics of this evolution is generally absorbed in fragmentation functions that repres-  
 669 ents the probability of a parton to fragment into a certain hadron of the final state. Many  
 670 of these primary hadrons are unstable and decay further at various timescales. Those  
 671 that are sufficiently long-lived have their decays visible in the detector, or they are stable.  
 672 There are several models of the hadronisation process, that attempt to connect the results  
 673 of the [PS](#) and the final particle spectrum observed. These models can be complemented  
 674 and tuned using experimental observations. The hadronisation is commonly described  
 675 by either the Lund string fragmentation model [72] (as implemented in PYTHIA), or the  
 676 cluster fragmentation model [73] (as implemented in HERWIG and SHERPA). Essentially,  
 677 the Lund string fragmentation model assumes a linear confinement, where the energy  
 678 stored in the colour field between quarks and antiquarks is assumed to increase linearly  
 679 with the separation of colour charges. Thus, it depicts the colour force by means of a  
 680 linearly rising potential as charges separate. The potential energy stored increases as  
 681 partons recede, so it may break up by the production of new quark-antiquark pairs that  
 682 screen the endpoint colours. Then, quarks and antiquarks may combine to produce had-  
 683 rons. The cluster fragmentation model is based on the colour preconfinement property of  
 684 the branching processes, which assumes that the separation of the colour charges forming  
 685 a singlet are inhibited. After the perturbative parton branching process, the remaining  
 686 gluons are split into light  $q\bar{q}$  pairs, and then neighbouring quarks and antiquarks can

687 be combined into colour singlets (colourless “clusters”), with masses distributions peaking  
 688 at low values and asymptotically independent of the hard subprocess scale.

### 689 1.3.3 Underlying Event ([UE](#))

690 In addition to the hard interaction that is generated by the [MC](#) simulation, it is also ne-  
 691 cessary to account for the interactions between the incoming proton remnants. This is  
 692 usually modelled through multiple extra  $2 \rightarrow 2$  scattering, occurring at a scale of a few  
 693 GeV. The modelling of the [UE](#) is crucial in order to give an accurate reproduction of the  
 694 energy flow that accompanies hard scatterings in hadron colliders. The [UE](#) can include  
 695 additional hard interactions and soft processes which can not be calculated perturbatively.  
 696 These are modelled with adjustable parameters which are tuned to experimental data.

### 697 1.3.4 Tunes

698 Due to the non-perturbative, and therefore incalculable, nature of much of the soft physics  
 699 processes, like the shower approximations, hadronisation and [UE](#), [MC](#) generators inevit-  
 700 ably contain a number of free parameters. These different parameters are usually tuned  
 701 with data from colliders. A specific set of chosen parameters for a [MC](#) generator is referred  
 702 to as a “tune”. In general, the [ATLAS](#) PYTHIA A14 tune [74] is used throughout this thesis.  
 703 The A14 tune is based on the MONASH tune [75] of the PYTHIA authors which uses  $e^+e^-$   
 704 collision data for the hadronisation parameters, and minimum bias  $pp$  collision data at  
 705 [LHC](#) to constrain parameters sensitive to initial state radiation and the [UE](#). The A14 tune  
 706 uses in addition a large variety of [ATLAS](#) data sensitive to multiple parton interactions  
 707 and [ISR/FSR](#), and includes jets built from tracks and variables sensitive to the internal jet  
 708 structure.

### 709 1.3.5 [ATLAS](#) detector simulation

710 To directly compare the data collected with the [ATLAS](#) detector with the prediction of [SM](#)  
 711 and [BSM](#) events in simulation, the interaction of the produced particles with the detector  
 712 material has to be simulated. The GEANT4 [76] software package is used to simulate the  
 713 interaction of particles produced in  $pp$  collisions with the different parts of the detector  
 714 (the [ATLAS](#) detector is described in Chapter 2). GEANT4 is an extensive particle simu-  
 715 lation package that governs all aspects of the propagation of particles through detectors,  
 716 based on a description of the geometry of the detector components and the magnetic field.  
 717 The physics processes include, among others, ionisation, Bremsstrahlung, photon conver-  
 718 sions, multiple scattering, scintillation, absorption and transition radiation. The last step  
 719 involves the digitalisation, which simulates the detector outputs in the same format as the  
 720 actual raw data. Due to the detailed and complicated geometry of [ATLAS](#) and the diversity

721 and complexity of the physics processes involved, the consumed computing time per event  
 722 is large ( $\mathcal{O}(1 \text{ hour})$ ).

723 The simulation of a large number of interactions necessary to mimick the [ATLAS](#) recon-  
 724 struction is computationally extensive. Especially the simulation of shower developments  
 725 in the calorimeters consumes a large amount of CPU and computing time. For many [BSM](#)  
 726 searches, a large number of parameters affecting the predicted particle masses and inter-  
 727 actions have to be simulated, therefore, a "fast" parameterised detector simulation has  
 728 been developed to cope with this high simulation demand.

729 A so-called AtlFast3 or AF3 [77] (built upon AltFast2 [78]) setup simulation chain uses  
 730 [GEANT4](#) [76] simulation for the interactions in the Inner Detector ([ID](#)) and Muon Spectro-  
 731 meter ([MS](#)) (described in Chapter 2), and two parametrised simulations of the Electromagnetic  
 732 Calorimeter ([ECAL](#)) and Hadronic Calorimeter ([HCAL](#)) are used: FastCaloSim V2<sup>3</sup>, and  
 733 FastCaloGAN. Parametric simulations of the calorimeter response simulate the energy of a  
 734 particle shower as a single step based on an underlying parametrization instead of simu-  
 735 lating how every particle propagates and interacts inside the calorimeter volume.

736 AtlFast3 introduces several key improvements compared to AtlFast2. Specifically, AtlFast3  
 737 enhances the handling of calorimeter showers, significantly improving how energy depos-  
 738 its in the detector cells are simulated. These improvements address limitations in AtlFast2,  
 739 where sub-cluster structures and lateral shower shapes were not fully described. This new  
 740 generation also integrates enhanced parameterised simulations and a more precise calor-  
 741 imeter model, leading to improved reconstruction of physics objects like jets and missing  
 742 transverse energy. These changes lead to better agreement between fast simulation and  
 743 full simulation results. Moreover, AtlFast3 supports more advanced algorithms for tracking  
 744 and calorimeter simulation, ensuring that discrepancies seen in AtlFast2 are minimized,  
 745 such as inaccuracies in shower shapes and fluctuations.

---

<sup>3</sup> The previous version of AtlFast, called AtlFast2 used FastCaloSim [79] to simulate the pasage of particles through the calorimeters.

746

## Part II

747

# Experimental setup

748

# 749 THE ATLAS EXPERIMENT

# 2

750

*Something.*

---

Someone

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

757 \*\*\* update chapter, see version in spanish

## 758 2.1 LHC

759 The [LHC](#) [1, 80] is the largest hadron accelerator in the world, located at [CERN](#), in the  
 760 French-Swiss border. It has a longitude of 27 km, located between 50 and 174 meters  
 761 underground. The [LHC](#) is designed to collide protons (and heavy ions) at a center of mass  
 762 energy of 14 TeV. To keep the protons and heavy ions on the accelerator ring, overall  
 763 9593 magnets are used. These magnets include superconducting dipole and quadrupole  
 764 magnets, cooled down to 1.9 K (-271 °C). The dipole magnets generate a magnetic field  
 765 of 8.3 T.

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

773 The protons are injected as bunches of  $\mathcal{O}(10^{11})$  protons into the [LHC](#) with a spacing of

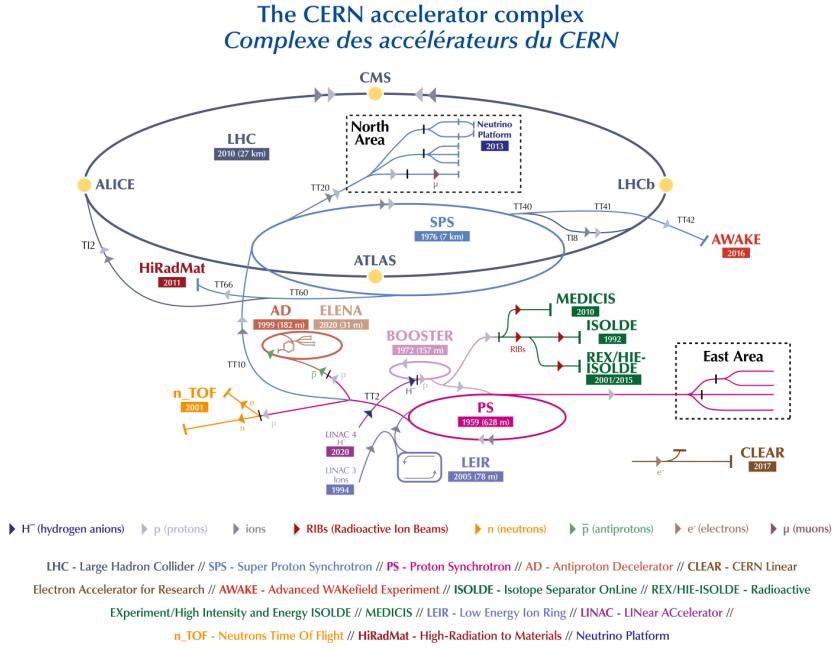


Figure 2.1: Overview of the LHC complex where all the accelerators that lead up to the LHC are shown [81].

774 25 ns (7.5 m). These bunches are later brought to collision in so-called bunch crossings.  
 775 The filling scheme of the pre-accelerator chain, in combination with finite switching times  
 776 of the injection and dumping magnets, results in regular patterns of filled and empty  
 777 bunches.

778 The LHC so far provided proton and heavy ion beams for two data-taking periods, and  
 779 is undergoing a third. Between 2009 and 2013 (known as Run 1), the LHC operated  
 780 with centre-of-mass energy ( $\sqrt{s}$ ) of 7 TeV and 8 TeV. After a long shutdown (LS1), the  
 781 second run (Run 2) started in 2015 and ended in 2018, providing 13 TeV collisions to  
 782 the experiments around the LHC ring. In 2022 the Run-3 began, at which  $p\bar{p}$  collisions  
 783 happen at an energy of 13.6 TeV, estimated to run until 2026. The four yellow dots in  
 784 shown Figure 2.1 there are four interaction points, housing the ALICE [82], LHCb [83],  
 785 CMS [84], ATLAS [85], LHCf [86], TOTEM [87], MoEDAL [88] experiments, among many  
 786 other experiments.

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

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

790 where  $N_b$  is the number of particles per bunch,  $n_b$  the bunches per beam,  $\gamma_r$  is the re-  
 791 lativistic gamma factor,  $\epsilon_n$  is the normalised transverse beam emittance and  $\beta^*$  being the  
 792 beta function at the collision point which determining the transverse spread of the particle  
 793 beam. The correction term  $F$  takes into account the beam crossing angle. The revolution  
 794 frequency is represented by  $f_{rev}$  which is  $\sim 11$  kHz, and with the bunch-spacing of 25 ns,

795 allows for beam crossing at the four interaction points with a frequency of  $\sim 40$  MHz.

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

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

797 connecting the luminosity with the number of events. More details regarding the lumin-  
798 osity measurements in **ATLAS** are shown in Section 2.3.

## 799 2.2 ATLAS

800 **ATLAS** is one of the multi-purpose detectors of the **LHC**, located at Point-1 along the **LHC**.

801 It was designed and built to study the  $pp$  (and heavy ion) collisions at the TeV scale.

802 The overall shape of the detector is that of a cylinder as shown in Figure 2.2. It has a

803 length of 44m and 25m in diameter, being the largest particle detector built so far. The

804 **ATLAS** detector is divided geometrically in two parts: the central part called *barrel*, and

805 the outer caps called *end-caps*.

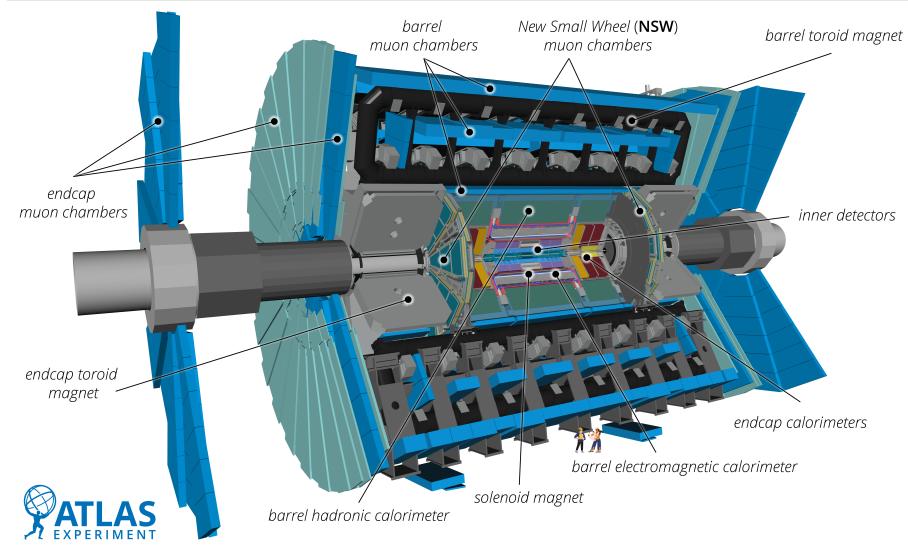


Figure 2.2: Overview of the **ATLAS** detector and all its sub-detectors, including the systems added during the Long Shut down 2 (**LS2**) [89].

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

812 It is built up of multiple layers, starting from the innermost component, the Inner De-  
813 tector (**ID**), providing tracking hits close to the beam pipe. Around the **ID**, there is a  
814 superconductor solenoid which creates an axial magnetic field of  $\sim 2$  T to curve the **ID**

815 tracks of charged particles. After the first magnet, there is a system of two calorimeters:  
 816 the **ECAL** and **HCAL**. The former is in charge of measuring the kinetic energy of  
 817 photons and electrons, and the latter measures the energy of the jets. The outermost  
 818 parts of **ATLAS** are built by the muon spectrometer, providing momentum reconstruction  
 819 for muons passing through the inner detector layers. Intertwined with the muon spectro-  
 820 meter, there are a total of 8 barrel toroid coils, providing a total magnetic field of 4 T (0.5  
 821 T per coil) to measure the momentum of muons. The toroid magnetic field is completed  
 822 by the end-cap toroids, also generating a magnetic field up to 4T for muons leaving **ATLAS**  
 823 close to the beam pipe.

824 Every component in **ATLAS** working together enables the reconstruction and identifica-  
 825 tion of a variety of particles with high precision. An overview of the design capabilities  
 826 of **ATLAS** in terms of the momentum and energy resolution is given in Table 2.1, adapted  
 827 from Ref. [85]. Here the resolution given lists first a stochastic term, measuring the un-  
 828 certainty based on the statistically dominated interaction of a particle with the material,  
 829 followed by a noise term, which accounts for uncertainties due to electronic noise in the  
 830 readout process.

Table 2.1: General performance goals of the **ATLAS** detector. The units of  $p_T$  and  $E$  are in GeV. Extracted from Ref. [85]

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

### 831 2.2.1 ATLAS Coordinate system

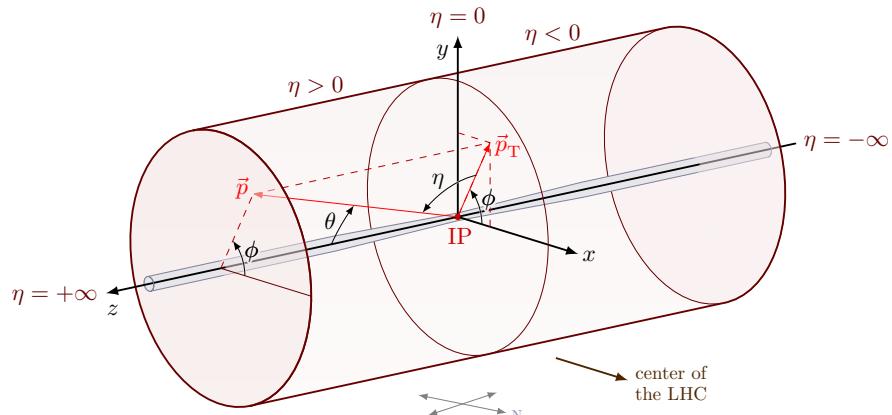


Figure 2.3: **ATLAS** coordinate system [89].

832 The coordinate system used within **ATLAS**, displayed in Figure 2.3, is used throughout  
 833 this thesis and shortly described in the following [85]. The origin of the right-handed  
 834 coordinate system is at the nominal interaction point, with the positive x-axis pointing  
 835 towards the centre of the **LHC**. The x-y plane is perpendicular to the beam axis, defining  
 836 the z-axis. Towards the surface defines the positive y-axis. An azimuthal angle  $\phi$  is defined  
 837 around the beam axis, and a polar angle  $\theta$  is the angle from the beam axis. Instead of  $\theta$   
 838 the rapidity  $y$  is used for heavy objects:

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

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

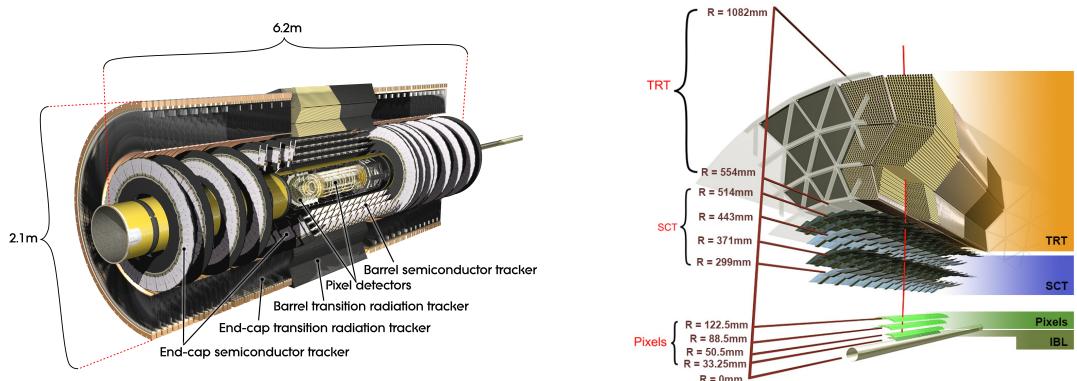
$$\eta = -\ln(\tan(\theta/2)). \quad (2.4)$$

841 To quantify the distance between two objects,  $\Delta R$  is defined:

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

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

### 844 2.2.2 Inner Detector



(a) ID with all its submodules in the barrel and end-cap regions [90].

(b) Layers of the ID showing their distance to the beam [90].

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

845 A cross-section of the Inner Detector (ID) system [91] is shown in Figure 2.4, highlighting  
 846 the distance of each subsystem from the beampipe. The innermost part of the ID is the  
 847 Insertable B-Layer (IBL), followed by three layers of pixel detectors. At 299 mm radial  
 848 distance from the beam pipe, four layers of SemiConductor Tracker (SCT) modules are  
 849 located before the Transition Radiation Tracker (TRT), which extends the overall ID de-  
 850 tector size to a radius of 1082 mm. The ID allows for particle track reconstruction within  
 851  $|\eta| < 2.5$ .

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

860 **IBL - Insertable B-layer** After Run-1, during a long shutdown in 2013-2014, the pixel  
 861 detector system was subject to maintenance and upgrades. Within this set of upgrades, a  
 862 4th pixel layer at a 3.3 cm distance from a new, smaller beam pipe (33 mm outer radius,  
 863 originally 36 mm), whic was the first in particle physics experiments [92, 93] and has  
 864 led to significant improvements in interaction vertex reconstruction and identification of  
 865 b-hadron jets.

866 **Pixel Detector** The innermost pixel layer, the IBL, is surrounded by three layers of pixel  
 867 detectors, arranged in barrels around the beam pipe [94, 95]. The method of detection  
 868 of charged particles is the measurement of deposited induced charges in a silicon layer,  
 869 product of ionization. The first layer is at a distance of 50.5 mm from the beam pipe's  
 870 centre. As can be seen in Figure 2.4a, the end caps of the pixel layer consist of 3 disks  
 871 around the beampipe, stretching the length of the pixel component of the **ID** to 1.4 m  
 872 length along the beam axis. The pixel detector consists of overall 1744 pixel modules with  
 873 a nominal size of  $50\mu\text{m} \times 400\mu\text{m}$  in the  $(\phi, z)$  plane ( $\phi, r$  for the disk panels), comprising  
 874 over 80 million readout channels. The pixel and **IBL** part of the **ATLAS** detector is crucial  
 875 for tracking, providing 4 pixel hits over the entire **ID** pseudorapidity coverage ( $|\eta| < 2.5$ .).

876 **Semiconductor Tracker** The pixel detector and **IBL** are located within **SCT** modules  
 877 [96]. Similar to the pixel detector modules, the **SCT** modules are semiconductor-based,  
 878 arranged into cylindrical layers around the beampipe in the barrel region, forming disks in  
 879 the endcap. Since the **SCT** modules only provide precise location along one axis, two mod-  
 880 ules are combined back-to-back and rotated against each other to gain two dimensional  
 881 spacial information. Four layers are arranged in the barrel, nine disks in each endcap side  
 882 (see Figure 2.4a). Including the endcap disks, the **SCT** extends up to  $|z| < 2735\text{mm}$ .

883 **Transition Radiation Tracker** The last part of the **ID** is the **TRT** [97], in the barrel  
 884 stretching from 554 mm to 1082 mm radial distance. This detector is composed of 4 mm  
 885 diameter straw tubes, arranged in parallel to the beam pipe or radially in the barrel and  
 886 end-cap, respectively. Within  $|\eta| < 2.0$ , three barrel rings and 18 end-cap units provide typ-  
 887 ically 36 hits per track. The straws are intertwined with polypropylene fibres for passing  
 888 through particles to create transition radiation. Inside the straws is a thin tungsten wire,

889 collecting charges drifting through the straws gas mixture (Xe, CO<sub>2</sub> and O<sub>2</sub>). The level of  
 890 radiation and collected charges in each straw can be used to discriminate between elec-  
 891 trons and charged pions. The **TRT** only offers spatial information in the ( $R - \phi$ ) plane, no  
 892 information in the z-direction can be extracted due to the straws orientation. There is a  
 893 total of 50000 tubes in the barrel region, while the end-caps contain 320000 tubes.

894 **2.2.3 Calorimeters**

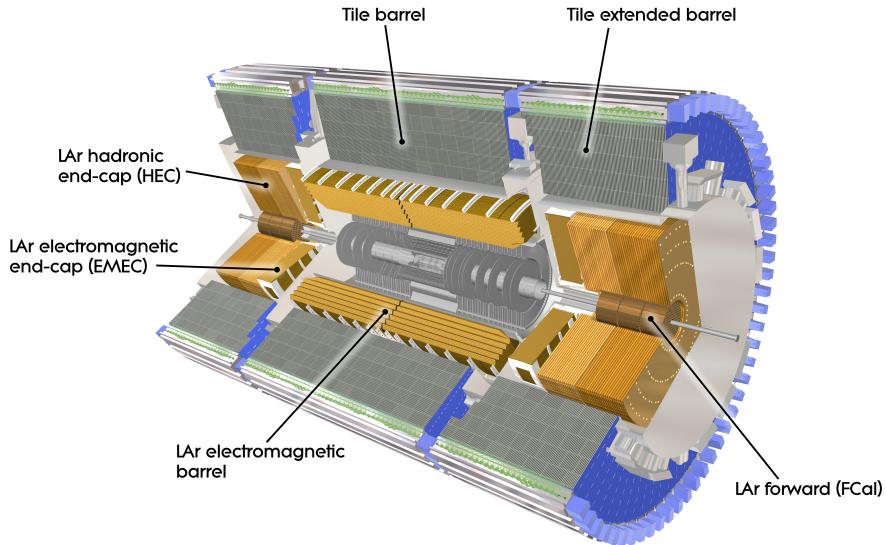


Figure 2.5: **ATLAS** calorimeter system, showing the Electromagnetic Calorimeter (**ECAL**) and the Hadronic Calorimeter (**HCal**) [98].

895 As previously mentioned, the **ID** system is surrounded by two calorimeters: the Electro-  
 896 magnetic Calorimeter (**ECAL**) and the Hadronic Calorimeter (**HCal**), as shown in Fig-  
 897 ure 2.5. These calorimeters are designed to measure the energy and position of the in-  
 898 cident particles, via the deposited energy by the secondary particle cascades produced by  
 899 the incident ones. It covers the whole  $\phi$  range and up to  $|\eta| < 4.9$ , with a finer gran-  
 900 ularity in the region that coincides with the **ID**. The calorimeter system allows for the  
 901 discrimination between photons and electrons from hadrons (jets). Furthermore, it allows  
 902 to measure the energetic imbalance (thanks to its total coverage and hermiticity) and it  
 903 provides the trigger system with the necessary information for the event selection.

904 Both calorimeters are so-called sampling calorimeters with alternating layers of absorber  
 905 and active material. The absorber layer triggers a shower development of consecutive  
 906 interactions with the detector material, the active layer detects the signal. The shower  
 907 development and properties are of vital importance for the particle identification, as it  
 908 will be shown later. Two important quantities in connection with the calorimeters are  
 909 the radiation length,  $X_0$ , and the interaction length  $\lambda$ . The radiation length refers to  
 910 the distance after which an particle (electrons for example) energy has been reduced to

911  $1/e$  of its initial energy. The interaction length describes the mean free path before the  
 912 occurrence of an hadronic interaction.

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

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

914 where  $\oplus$  means that the terms are summed in quadrature. The stochastic term  $\frac{a}{\sqrt{E}}$  is  
 915 related with the fluctuations on the shower developments, the constant term  $b$  takes into  
 916 account the inhomogeneities of the detector, and the last term is associated with the elec-  
 917 tronic noise and is proportional to  $\frac{1}{E}$ . The value of the coefficients  $a$  and  $b$  depend on the  
 918 incident objects. For the electrons' case in the [ECAL](#),  $a \sim 10\% \text{ GeV}^{1/2}$  and  $b \sim 0.7\%$ , while  
 919 those for charged pions in the center of the detector are  $a \sim 50\% \text{ GeV}^{1/2}$  and  $b \sim 5\%$  [99].

## 920 Electromagnetic Calorimeter - ECAL

921 The [ECAL](#) specializes on the detection of electrons, positrons and photons, which deposit  
 922 their energy in relatively dense showers: energetic electrons that radiate Bremsstrahlung  
 923 photons, while energetic photons convert to electron-positron pairs when traversing the  
 924 dense material. The absorber is made of lead (Pb) with stainless steel sheets, while Liquid  
 925 Argon ([LAr](#)) is used as the active material with copper and kapton electrodes for readout.

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

934 The thickness of the [ECAL](#) is over 22 radiation lengths ( $X_0$ ) in the barrel region, while over  
 935  $24X_0$  in the end-cap region. For photons, the distance at which the energy dropped to  $1/e$   
 936 is  $9/7X_0$ , therefore all the photon's electromagnetic energy is deposited in the [ECAL](#), and  
 937 only a small part reaches the [HCAL](#).

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

943 Within the region accepted for precision measurements ( $|\eta| < 2.5$  excluding the crack),  
 944 the [ECAL](#) is segmented in three longitudinal layers, displayed in Figure 2.6. The first  
 945 layer consists on fine-granularity bands (also called the strip layer) which helps with the  
 946 discrimination between isolated photons and pairs of photons spacialy closed originating

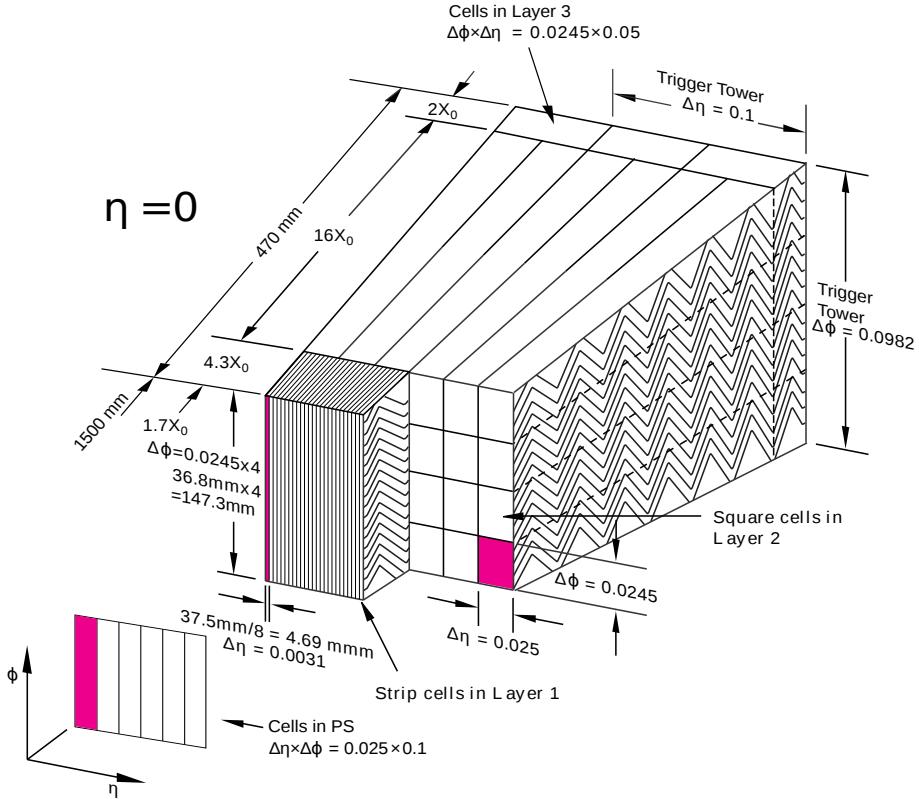


Figure 2.6: Segment of the ECAL showing the layer arrangement and cell dimensions in each layer [85].

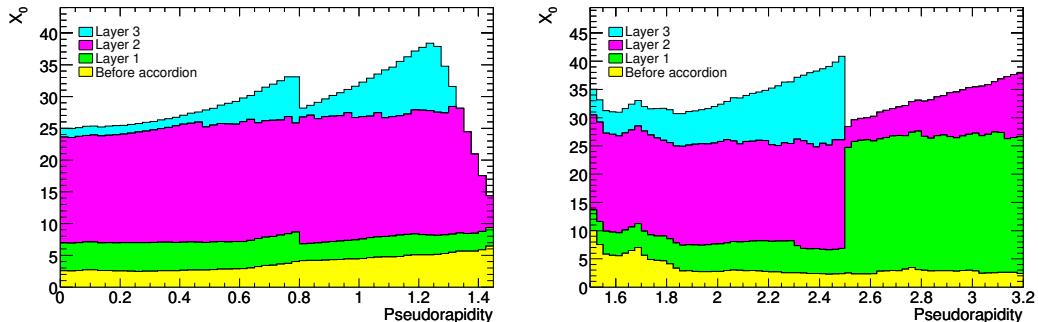


Figure 2.7: Radiation lengths as a function of  $|\eta|$  for the ECAL, separated for each sublayer [85].

from  $\pi^0 \rightarrow \gamma\gamma$  decays. This layer has a constant thickness of  $\sim 6X_0$  as a function of  $\eta$  (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  $0.025 \times 0.025$  in  $(\eta, \phi)$  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.

## 955 Hadronic Calorimeter - HCAL

956 Three hadronic calorimeter layers surround the [ECAL](#) and provide additional discrimination  
 957 for electrons and photons when measuring the hadronic energy. The [HCAL](#) extends  
 958 in pseudorapidity up to  $|\eta| < 4.9$ , allowing virtually the entirety of the solid angle to be  
 959 covered from the interaction point. In the barrel region ( $|\eta| < 1.7$ ), the tile calorimeter, a  
 960 sampling calorimeter using steel as absorbing material and plastic scintillator tiles as active  
 961 material [100], is located. It is divided into two parts ( $|\eta| < 1.0$  and  $0.8 < |\eta| < 1.7$ ).  
 962 The scintillators, arranged in a periodic array, are connected to an optical fiber that carries  
 963 the light produced by the passing particles to a photomultiplier tube. This array extends,  
 964 in  $R$ , from 2.28 to 4.25 m. In the endcap region ( $1.5 < |\eta| < 3.2$ ) there is an hadronic  
 965 sampling calorimeter, the Hadronic End-Cap Calorimeter ([HEC](#)), with copper plates as absorber  
 966 and liquid argon as active material. Each side of the endcap consists of two wheels,  
 967 one behind the other with the flat Cu plates arranged perpendicular to the beam axis,  
 968 with a radius of 2.3 m. Finally there is the [FCAL](#), a sampling calorimeter that extends  
 969 the coverage of the system to  $|\eta| < 4.9$ , coaxial to the beam axis and located 4.7 m on  
 970 either side of the interaction point. The main material of the modules is [LAr](#) (with copper  
 971 or tungsten), and while not used for precision measurements, it provides information for  
 972 computation of the missing transverse energy and reconstruction of jets in regions very  
 973 close to the beam axis.

974 The [HCAL](#) has a thickness greater than  $7.7 \lambda$  in the barrel region ( $9.7 \lambda$  in total if the  
 975 [ECAL](#) is counted). Analogous to the radiation length mentioned for the [ECAL](#), a hadronic  
 976 interaction length is defined as the average distance over which the energy of a hadron is  
 977 reduced to  $1/e$  of its initial energy. Thus, all the energy with which the hadrons arrive at  
 978 the [HCAL](#) is deposited there.

### 979 2.2.4 Muon spectrometer

980 The high  $p_T$  muons generated at the interaction point have very high penetrating power  
 981 and are poorly interacting. Therefore, the [MS](#) [101] is located in the outermost part of  
 982 the [ATLAS](#) detector, embedded within the 4 T magnetic field generated by the barrel and  
 983 endcap toroid magnets, and is designed to obtain high precision position and momentum  
 984 measurements of high  $p_T$  muons. This is the largest subdetector and the one that gives  
 985 [ATLAS](#) its size.

986 The [MS](#) is designed to precisely measure muons within  $|\eta| < 2.7$  and to provide muon trigger  
 987 information up to  $|\eta| < 2.4$ , shown in Figure 2.8, highlighting the different subsystems  
 988 The [MS](#) is composed of different types of muon detection chambers (see Figure 2.8).  
 989 Monitored Drift Tubes ([MDTs](#)) are responsible for most of the precision measurements  
 990 and cover the range of  $|\eta| < 2.7$ . They operate similarly to the [TRT](#), with tubes filled with  
 991 an ionising gas and a central anode collecting the electrons produced, and the drift time  
 992 is associated with the distance to the track. In the endcap region there are Cathode Strip

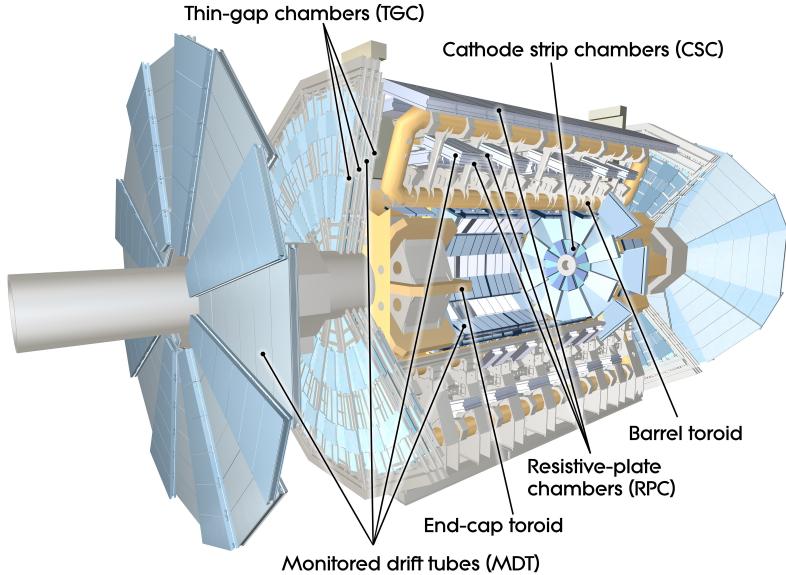


Figure 2.8: [ATLAS](#) Muon Spectrometer (MS) [85].

993 Chamber ([CSC](#)) that have high spatiotemporal resolution and a coverage of  $|\eta| > 2.0$ .  
 994 These chambers work by measuring the charge deposited on an anode as a result of the  
 995 cascade of electrons created near the anode. Resistive-Plate Chambers ([RPCs](#)) provide a  
 996 fast estimation of the muon momentum at trigger-level with a coverage of  $|\eta| < 1.05$ <sup>1</sup>.  
 997 [RPCs](#) measure the discharge between two parallel resistive plates subjected to a high  
 998 potential difference, following the ionisation of the internal gas volume caused by the  
 999 passage of energetic muons. Finally, in the endcap region, there are Thin-Gap Cham-  
 1000 bers ([TGCs](#)), similar in function to [CSCs](#). They also provide information to the trigger  
 1001 system in this region and have a coverage of  $|\eta| < 2.4$ .  
 1002 If hits in the [ID](#) and the [MS](#) can be associated with a single muon, a very good momentum  
 1003 resolution of up to

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

1004 is achieved. The momentum resolution degrades accordingly if a track is identified in only  
 1005 one of the two systems.

### 1006 2.2.5 The Trigger System

1007 The [ATLAS](#) trigger system [103–105] uses information from the detector to reject events  
 1008 that do not possess interesting physics (physics already known for example), reducing the  
 1009 event frequency from 40 MHz (bunch-crossing frequency mentioned in Section 2.1) to  
 1010 around 1.5 kHz. It is necessary to emphasize here the central role of the trigger system  
 1011 for the proper functioning of the whole experiment, being responsible for deciding which  
 1012 events are saved and, ultimately, which physics will be encountered (or not) during the

<sup>1</sup> The innermost End-Cap layer has been replaced with the [ATLAS](#) New Small Wheel ([NSW](#)) after Run 2 [102]. It features MicroMegas as precision trackers as they provide better performance at the high rates expected in future LHC operations.

1013 event analysis. Without an efficient trigger system, all the subdetectors described above  
 1014 would be wasted. To achieve such a reduction in event frequency and, at the same time,  
 1015 have a high efficiency in selecting those of interest, the trigger system is composed of two  
 1016 consecutive levels capable of performing increasingly complex particle identification; a  
 1017 first hardware-based trigger level, Level-1 ([L1](#)), and then a high-level software-based trig-  
 1018 ger, the High Level Trigger ([HLT](#)). Each level allows events to be analyzed in greater detail,  
 1019 increasing the accuracy of the selection criteria and the complexity of the algorithms used.

## 1020 **Level-1 trigger**

1021 The trigger decision starts with the hardware-based [L1](#) trigger [106], which identifies  
 1022 Region of Interests ([ROIs](#)). These [ROI](#) consist of neighbouring cells in the [ECAL](#) and [HCAL](#)  
 1023 and are defined from the position in the calorimeter of each object found in a potentially  
 1024 interesting event, which extends as a cone from the interaction point along the detector.  
 1025 Regarding muons, it takes the information read by the [MS](#), more specifically the [TGC](#)  
 1026 and [RPC](#) and allows to obtain a fast estimate of the muon  $p_T$ . The [L1](#) trigger also has  
 1027 a component that allows for topological requirements such as invariant mass selections  
 1028 and distance measures to be taken into account in the [L1](#) decision, referred as the Level-1  
 1029 Topological ([L1Topo](#)).

1030 The design of the [L1](#) allows to have an acceptability in the range of  $|\eta| < 2.5$  for electrons,  
 1031 photons, muons and taus, up to  $|\eta| < 3.2$  for jets, and  $|\eta| < 4.9$  for the missing transverse  
 1032 momentum calculation. Using the [ROIs](#), the [L1](#) trigger must make the decision to keep or  
 1033 discard the event, reducing the event rate from 40 MHz to less than 100 kHz in approx-  
 1034 imately 2.5  $\mu s$ , time determined in part by the limited size of the memory buffers and in  
 1035 part by the time it takes for the muons produced in the event to reach the [MS](#). This final  
 1036 decision is done by the Central Trigger Processor ([CTP](#)), and then passes the [ROIs](#) to the  
 1037 next trigger level: the [HLT](#).

## 1038 **The High Level Trigger**

1039 When an event is accepted by the [L1](#), the [HLT](#) [107] executes a sequence of algorithms  
 1040 starting from the [ROIs](#) defined by the [L1](#), and allows to reduce the event rate that is stored  
 1041 at 1.5 kHz in 0.2 s. The reconstruction and identification of candidate particles in the [HLT](#)  
 1042 is evaluated in a sequence of steps where different algorithms are applied. If the selection  
 1043 fails in a certain step, the following steps are no longer executed to save execution time.  
 1044 In [HLT](#), the algorithms are grouped into sets of fast reconstruction algorithms executed  
 1045 first, and then a set of precision reconstruction algorithms similar to those used offline  
 1046 are executed, thanks to the latency time available. The fast reconstruction algorithms use  
 1047 the calorimeter and track information from the [ID](#) only within the [ROI](#) to perform can-  
 1048 didate selection and identification, and perform background rejection as quickly and early  
 1049 as possible. If the candidate particle passes the criteria defined by the fast reconstruction

1050 selection, precision selection algorithms are run. These have access to detector information  
1051 outside the RoI, at the highest granularity and including details on calorimeter energy  
1052 calibration, sub-detector alignment and magnetic field mapping.

1053 The exact sequence and type of algorithms considered at the [HLT](#) are defined in the trigger  
1054 *menu*. This comprises a database of triggers, each of one defining a sequence of algorithms  
1055 and requirements on these algorithms for an event to pass the [HLT](#). The trigger require-  
1056 ments are designed and budgeted in a way that the overall [HLT](#) rate does not exceed 1  
1057 kHz. In some cases, even the reduction in event rate achieved through the [HLT](#) algorithms  
1058 for desired trigger requirements, such as low momentum triggers, is too high. To keep the  
1059 overall [HLT](#) rate below 1 kHz in these cases, triggers can still be included in the menu, but  
1060 with a prescale. A prescale is an artificial scaling of the trigger, only accepting every Nth  
1061 trigger decision if the prescale factor is N. This allows triggers with an otherwise high rate  
1062 to still collect events.

1063 The [HLT](#) algorithms run on approximately 40 thousand CPU cores. In addition, partial  
1064 event construction is used for trigger-level analysis, detector monitoring, and detector  
1065 subsystem calibrations. Finally, the accepted events by the [HLT](#) are saved to a disk and  
1066 distributed, available *offline* for any study or analysis.

## 1067 2.3 Data-taking during Run-2

1068 The operation of the [LHC](#) is organized into distinct periods known as data-taking Runs.  
1069 Each run typically spans several years and is characterized by specific experimental con-  
1070 ditions, including the energy at which the protons are collided and the intensity of the  
1071 beams. Since its commissioning, the [LHC](#) has undergone multiple data-taking runs: Run-1  
1072 (2010-2013) operated at collision energies up to 8 TeV, Run-2 (2015-2018) at 13 TeV,  
1073 and Run-3 (2022-present) at 13.6 TeV. Each data taking period, once the [LHC](#) announces  
1074 stable beams, is divided into Luminosity-blocks ([LBs](#)) of approximately two minutes. At  
1075 each [LB](#), the instantaneous luminosity is practically constant and the beam conditions are  
1076 stable. Due to the high complexity of the [LHC](#) and the [ATLAS](#) detector, it is expected to  
1077 have inefficiencies in the detectors and sub-detectors and/or in the data acquisition chain.  
1078 During each Run, each part of [ATLAS](#) is monitored and any failure or problem is registered,  
1079 including inactive components, or problems on the [LHC](#) beam.

1080 In order to guarantee the high-quality data, free from significant defects, the [LBs](#) and  
1081 ranges within them that pass all the quality criteria are compiled into Good Runs List  
1082 ([GRL](#)). The lists are produced and distributed in a centralized manner, in order to provide  
1083 any [ATLAS](#) group with the same collection of [LBs](#). Since during the runs different parts  
1084 of the detector are available (in an optimal run, all of the subdetectors are available),  
1085 there are multiple [GRLs](#) available to use. Each analysis, then, selects which [GRL](#) to use  
1086 depending their tolerance to the subdetectors' faults.

1087 The present thesis uses [ATLAS](#) data recollected from  $pp$  collisions during the Run-2 (2015-

1088 2018), at a centre of mass energy of  $\sqrt{s} = 13$  TeV. During this run, the LHC delivered a  
 1089 total of  $156 \text{ fb}^{-1}$ , from which ATLAS collected  $147 \text{ fb}^{-1}$ . The total integrated luminosity  
 1090 available for Physics analysis is  $140.07 \text{ fb}^{-1}$ <sup>2</sup>, as seen from Figure 2.9a. The uncertainty in  
 1091 the combined integrated luminosity for Run-2 is 0.83% [108], obtained using the LUCID-2  
 1092 detector [109]. Combining the 2022, 2023 and 2024 years of data taking for Run-3, 159  
 1093  $\text{fb}^{-1}$  of data was recollected, shown in Figure 2.9b [110, 111].

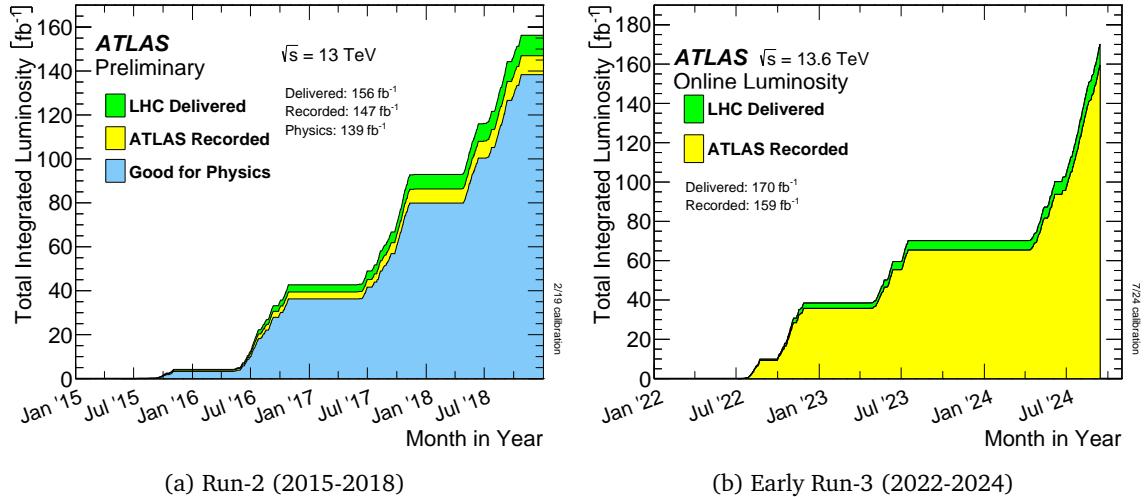


Figure 2.9: Luminosity delivered by the LHC and recorded by ATLAS during the Run-2 [108] and Run-3 data taking periods. For Run-2, the fraction of data good for physics analyses is also displayed.

1094 Another important concept in ATLAS data acquisition is pileup, which occurs when particles  
 1095 produced in more than one  $pp$  collision arrive at the detector at the same time, or more  
 1096 generally, when signals overlap in a way that cannot be separated. When bunches of  
 1097 protons collide, the probability of an interaction is proportional to the particle density, or  
 1098 better, to the particle flux, which is expressed by the instantaneous luminosity. The actual  
 1099 number of particle collisions that take place when two bunches intersect is a random vari-  
 1100 able that follows a Poisson distribution. For low luminosities, in most beam crossings, no  
 1101 collisions occur, but for high instantaneous luminosities, in most crossings many particle  
 1102 collisions occur at the same time. Depending on the subdetector and the type of measure-  
 1103 ment, it may or may not be possible to distinguish between particles coming from different  
 1104 simultaneous interactions. This is called in-time pile-up. In contrast, out-of-time pile-up  
 1105 includes the effects that arise when the time the detector needs to return to its waiting  
 1106 state is longer than the time between bunches crossing. A quantitative measure of pile-up  
 1107 and event activity is the mean value of  $pp$  inelastic interactions per bunches crossing,  $\langle\mu\rangle$ .  
 1108 The maximum instantaneous luminosities increased by a factor of four over the four years  
 1109 of Run-2, resulting in an increase of  $\langle\mu\rangle$  from 10 up to 60, as shown in Figure 2.10a.  
 1110 For Run-3, pileup was drastically increase up to values of 57 for year 2024, in average  
 1111 increasing up to 52 interactions per bunch crossing, displayed in Figure 2.10b.

<sup>2</sup> First measurements and initial GRLs led to a total of  $139 \text{ fb}^{-1}$

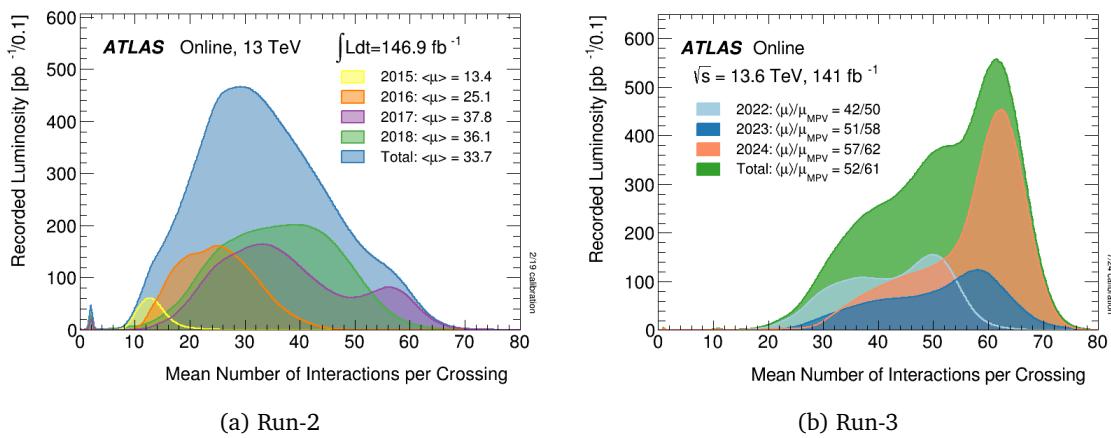


Figure 2.10: Pileup conditions during Run-2 and Run-3.

1112

1113 **RECONSTRUCTION AND  
IDENTIFICATION OF  
PHYSICAL OBJECTS**

# 3

1114

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

---

Billie Jean King

1115 The particles (and products of their decays) produced at every collision, interact with the  
 1116 detector in a particular manner according to their nature. The information recollected by  
 1117 all the sub-detectors described in the previous chapter allow for the reconstruction and  
 1118 identification of the physical objects present in each accepted event by the trigger system.  
 1119 Two types of reconstruction and identification exist. The *online* one, is carried out at the  
 1120 same time the  $pp$  collisions take place, and the *offline* one, done after the events are saved  
 1121 to storage. The reconstruction is done event by event, and is carried in the same way for  
 1122 events recorded by the [ATLAS](#) detector and for simulated [MC](#) events. In the following, a  
 1123 brief overview of the offline reconstruction and identification of the objects used in this  
 1124 thesis is given.

1125 \*\*\* update chapter, see version in spanish

1126 **3.1 Track and vertex reconstruction**

1127 In a high-pile-up event, there can be of the order of 1000 charged particles passing through  
 1128 the [ATLAS](#) detector. The information from the [ID](#) (Section 2.2.2) is used to reconstruct the  
 1129 trajectories of charged particles, called *tracks*.

1130 Tracking charged particles is a critical step in reconstruction. Tracks encode charged  
 1131 particles' momentum and trajectory, playing an essential role in particle identification and  
 1132 primary vertex reconstruction. As the inner detector is closest to the beamline and com-  
 1133 prises minimally ionizing detector material with high granularity, it plays the main role  
 1134 in track reconstruction. A charged particle passing through different layers of [ID](#) leaves  
 1135 a signal via ionization. As the [ID](#) solenoidal field is homogenous, the resulting trajectory  
 1136 is circular in the  $xy$  plane. Five parameters shown in Figure 3.1 define charged particle

1137 tracks:

- 1138 •  $q/p_T$ : the ratio of the charge and transverse momentum defining the curvature
- 1139 •  $d_0$ : the distance of the closest approach to the primary vertex in  $xy$ -plane defining
- 1140     the transverse impact parameter
- 1141 •  $z_0$ : the longitudinal impact parameter along the  $z$ -axis
- 1142 •  $\phi_0$ : the azimuthal angle
- 1143 •  $\theta_0$ : the polar angle of the particle direction at the closest point of approach [112].

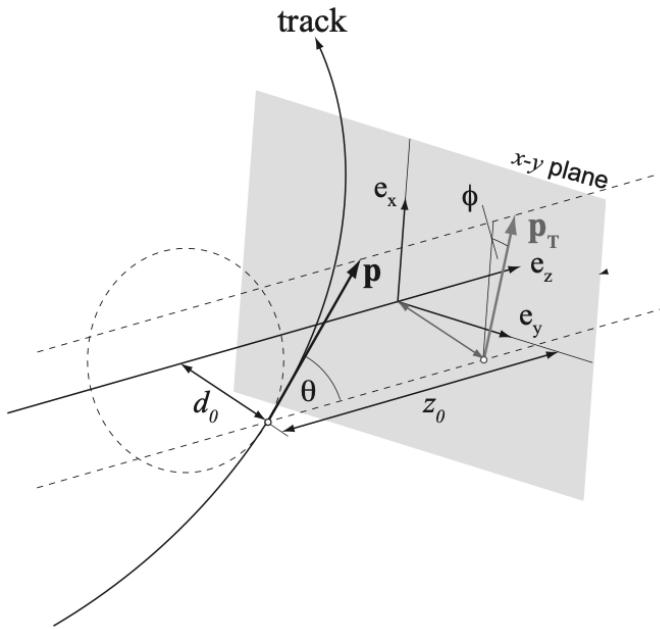


Figure 3.1: Schematic showing the tracking parametrization [113].

1144 The track reconstruction used in Run-2 uses two complementary approaches: the *inside-out*  
 1145 approach, and the *outside-in* [114].

1146 The first step in the inside-out track reconstruction is the seed-finding, where three hits in  
 1147 the silicon detector are searched for to seed the track reconstruction. Using these three  
 1148 hits and assuming an uniform magnetic field, a first estimate of the track parameters is  
 1149 obtained. Using the track seeds, the track is extrapolated to the other silicon layers, from  
 1150 which a combinatorial Kalman filter is used to estimate the track parameters. At this stage  
 1151 of the process there can be several track candidates for each track seed. Once the track  
 1152 is formed, an ambiguity resolution algorithm is applied to reassign shared clusters to the  
 1153 track with a better match [115], and the final track candidate is fitted using a global  
 1154  $\chi^2$  method. The last part of the inside-out method consists on extending the tracks to the  
 1155 **TRT**, and including the **TRT** hits to the track, to improve the track's momentum resolution.  
 1156 To improve the efficiency for tracks from decays displaced from the original collision point,  
 1157 an outside-in track reconstruction algorithm is also used. The track is seeded with hits

1158 from the [TRT](#). The track is extended to include hits from the silicon detector, with an  
 1159 ambiguity solver again applied to mitigate the hit sharing between two tracks.

1160 Primary and secondary vertices are of vital importance for the subsequent object recon-  
 1161 struction in [ATLAS](#). In this step, the tracks found as explained previously are used as input  
 1162 to the vertexing algorithm [116, 117]. First of all, the Primary Vertex ([PV](#)) is defined as  
 1163 the location where two protons collide. [PV](#)s are reconstructed by matching up intersecting  
 1164 tracks, which proceeds in three main steps: seeding, track assignment, and fitting. The  
 1165 vertex with the largest  $\sum p_T^2$  for all associated tracks is labeled as the hard-scatter vertex.  
 1166 There are some particles that decay rapidly after their production, such as  $\tau$  leptons or  
 1167 heavier quarks ( $b$  or  $c$ -quarks), and their decay position can be measured. From the re-  
 1168 maining tracks originated from these decays, it is possible to identify secondary vertices.  
 1169 All the remaining reconstructed vertices are considered to be pile-up.

## 1170 3.2 Photons and electrons

1171 The reconstruction of electrons and photons in [ATLAS](#) is based on the energy deposition  
 1172 in the [ECAL](#). Since electrons and photons leave similar signals in the [ECAL](#), their recon-  
 1173 strucion is done simultaneously, distinguishing between them by the reconstructed track  
 1174 information left in the [ID](#).

### 1175 3.2.1 Reconstruction

1176 The *offline* photon and electron reconstruction [118, 119] makes use of dynamic, variable-  
 1177 size clusters, connected topologically between the [ECAL](#) and [HCAL](#) cells [120], called  
 1178 topo-clusters. This approach allows for the clusters to recover energy from bremsstrahlung  
 1179 photons or from electrons from photon conversions. With this approach, there are three  
 1180 types of objects:

- 1181   • Electrons: consists of a cluster built from the energy deposits in the [ECAL](#) and a  
   1182    matched track.
- 1183   • Converted photons: consists of a cluster mathed to a conversion vertex (or vertices)
- 1184   • Unconverted photons: cluster matched to neither an electron track nor a conversion  
   1185    vertex.

1186 The algorithm for the reconstruction of electrons and photons proceeds as shown in Fig-  
 1187 ure 3.2. The reconstruction process begins with the topo-cluster formation. First, proto-  
 1188 clusters are formed in the [ECAL](#) and [HCAL](#) by grouping cells that have a required energy,  
 1189 and by subsequently adding neighbouring cells in four consecutive steps, obtaining the  
 1190 topo-cluster. Reconstructions starts only in those cases where the topo-clusters energy in  
 1191 the [ECAL](#) is greater than 400 MeV.

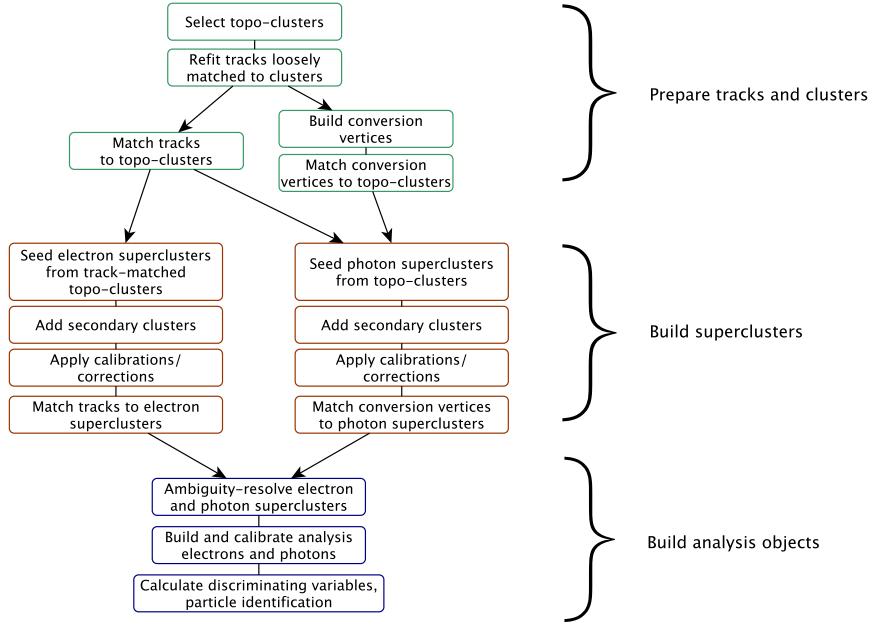


Figure 3.2: Diagram showing the reconstruction algorithm workflow for electrons and photons, extracted from Ref. [118]

1192 The algorithm also builds conversion vertices out of the refitted tracks and matches them  
 1193 to the selected topo-clusters. After the initial track-cluster matching and conversion build-  
 1194 ing, the electron and photon supercluster algorithms run separately in parallel. In the  
 1195 first stage, topo-clusters are evaluated for use as seed cluster candidates, which form the  
 1196 basis of superclusters; in the second stage, clusters near the seed candidates are identi-  
 1197 fied as satellite cluster candidates, which may emerge from bremsstrahlung radiation or  
 1198 topo-cluster splitting. Satellite clusters are added to the seed candidates to form the final  
 1199 superclusters, if they pass the necessary selection criteria. After applying initial position  
 1200 corrections to the resultant superclusters, the reconstruction algorithm matches tracks to  
 1201 the electron superclusters and conversion vertices to the photon superclusters.

1202 Since one object may be reconstructed as both an electron and a photon, an ambiguity  
 1203 resolution is performed to remove part of the overlap. However, some overlap is allowed  
 1204 in order to maintain a high reconstruction efficiency for electrons and photons, to which  
 1205 physics analyses may apply their own criteria. The final electrons and photons are then  
 1206 built and calibrated, facilitating the calculation of additional variables used for quality  
 1207 cuts and ambiguity resolution

### 1208 3.2.2 Identification

1209 In order to distinguish real photons (those coming from the collision) from background  
 1210 photons which have much larger production cross sections (coming from hadrons decays,  
 1211 also called fake photons), it is necessary to rely on a algorithm of identification with high  
 1212 signal efficiency and background rejection, for photon candidates with  $p_T \sim 10$  GeV up

1213 to the TeV scale. Currently, photon identification in ATLAS is based on a set of rectangular  
 1214 cuts on Shower Shape Variables (**SSVs**) computed from the energy deposited in the  
 1215 cells of the cluster in the first and second layer of the **ECAL**, and from the leakage to  
 1216 the **HCAL**. These variables describe the passage of the photons through the calorimeters,  
 1217 characterizing the lateral and longitudinal electromagnetic showers. The full photon  
 1218 identification process is presented in Chapter 4, where the **SSVs** are explained one by  
 1219 one. Also, Chapter 5 present two approaches to correct the differences seen in these **SSVs**  
 1220 between data and **MC**.

### 1221 Shower shape variable corrections

1222 Due to the imperfection of the **ATLAS** simulation to model the **SSVs**, and that these vari-  
 1223 ables are used as input to the identification step, it is important that they are corrected.  
 1224 Historically, the corrections are called Fudge Factors (**FFs**), and they comprised modifica-  
 1225 tions to the mean value of the **SSVs**, calculated by minimizing the  $\chi^2$  value between the  
 1226 data and **MC SSVs**. More details on the **SSV** corrections are given in GIVE SECTION!!!  
 1227 Additional corrections to all the reconstructed and identified photons in the simulation are  
 1228 applied event-by-event in the form of scale factors. These values, provided centrally by the  
 1229 **EGAMMA** group to **ATLAS**, represent the residual differences on the efficiencies between  
 1230 actual data and **MC**, computed as a function of the photon  $p_T$  and pseudo-rapidity  $\eta$  and  
 1231 separately for converted and unconverted photons.

#### 1232 3.2.3 Isolation

1233 To further reduce the backgrounds of jets misidentified as photons and of hadron decay  
 1234 within the jets (such as the case of neutral pions), two isolation variables are defined:  
 1235  $E_T^{\text{cone}40}$  and  $p_T^{\text{cone}20}$ .

1236 The procedure to compute the isolation energy  $E_T^{\text{cone}40}$  is as follows, and showed in Fig-  
 1237 ure 3.3. First, a cone of radius  $\Delta R < 0.4$  is built around the photon or electron candidate,  
 1238 and the energies of all the cells in the topo-clusters (introduced in Section 3.2.1) whose  
 1239 bary-centers are located inside the cone, are added together. Then, to this computed en-  
 1240 ergy, the energy of all the cells in a  $5 \times 7$  window (in units of  $\eta \times \phi$  in the second layer of  
 1241 the **ECAL**) centered around the candidate are subtracted, in order to remove the energy of  
 1242 the candidate itself. Pile-up contributions and energy leakages outside the cone are also  
 1243 taken into account.

1244 The track isolation variable  $p_T^{\text{cone}20}$  is obtained by adding the  $p_T$  of the good-quality tracks  
 1245 in a cone of radius  $\Delta R < 0.2$  around the electron candidate or in the direction of the  
 1246 converted photon cluster. The track associated to the track or to the converted photon are  
 1247 excluded from this computation, as well as those tracks which do not pass the *good-quality*  
 1248 track requirement. A *good-quality* track is defined as one in which the  $p_T$  is  $p_T > 1$  GeV,  
 1249 and it has a minimum distance to the primary vertex along the  $z$ -axis of  $|z_0 \sin \theta| < 3$  mm.

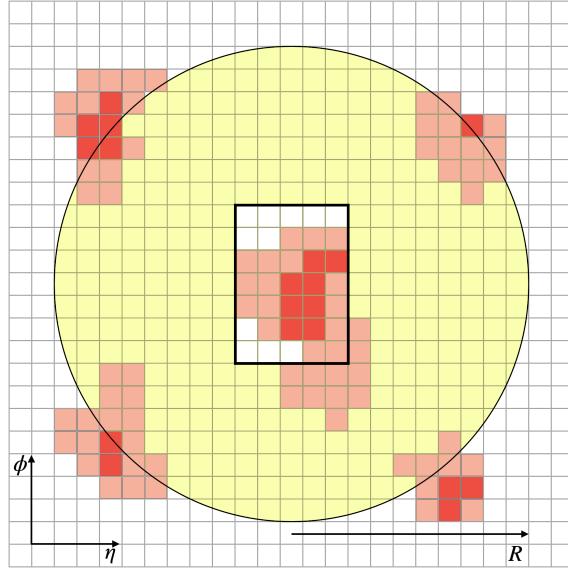


Figure 3.3: Diagram showing the calculation of the calorimetric isolation variable. When  $R = 0.4$ ,  $E_T^{\text{cone}40}$  is computed.

1250 In general, for photons and electrons, there is no other energy deposited in the cone  
 1251 around the candidate, apart from the low-energy objects originating from the remnants of  
 1252 the collision, multiple interactions and pile-up. On the other hand, for fake photon can-  
 1253 didates and non-direct photons, additional energy is observed within the cone, originating  
 1254 from objects accompanying the jet.

Table 3.1: Summary of electron and photon isolation Working Points (WPs) use throughout this thesis.

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

1255 From the calorimetric and track isolation different WPs can be defined separately for both  
 1256 electrons and photons. For electrons, two strategies are defined: either to achieve a fixed  
 1257 efficiency, or to apply fixed cuts on the isolation variables. In the case of photons, there are  
 1258 WPs which do not use both the isolation variable, as is the case of the FixedCutTightCaloOnly  
 1259 WP, which only uses calorimetric isolation. The definitions of the different WPs used  
 1260 throughout this thesis is shown in Table 3.1. Also, it is common to define the following  
 1261 variables for the photon FixedCutTight WP:

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

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

1262 therefore leaving the FixedCutTight WP defined as

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

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

### 1263 3.3 Muons

1264 The rate of bremsstrahlung radiation is inversely proportional to the square of a particle's  
 1265 mass. Since muons are about 200 times heavier than electrons, they primarily interact  
 1266 with the detector material through ionization. Therefore, muons are minimally ionizing  
 1267 particles that do not create electromagnetic shower in the calorimeters and pass through  
 1268 all layers of the **ATLAS** detector. Hence, muon detection relies on track measurements  
 1269 from the **ID** and **MS**. The combination of the two subdetectors define four types of muons,  
 1270 depending on the used information for the reconstruction:

- 1271     • Combined Muons (**CBs**): muons reconstructed from a global refit of **ID** and **MS**  
     1272       tracks
- 1273     • Segmented Muons (**STs**): muons reconstructed from a fitted **ID** track and **MS** seg-  
     1274       ment track
- 1275     • Calorimetric Muons (**CTs**): muons reconstructed using **ID** track matched to the min-  
     1276       imum ionizing energy deposits in the calorimeters
- 1277     • Standalone/Extrapolated Muons (**MES**s): muons reconstructed solely from **MS** tracks.

1278 The overlap between different types of muons is resolved as follows. When two muon  
 1279 types share the same **ID** track, the order of preference is: first **CB**, then **ST** and finally **CTs**.  
 1280 The overlap with **MES**s is solved by analyzing the hits of the tracks, selecting those tracks  
 1281 with the best fit and the highest number of hits.

1282 For the muon identification, quality cuts are applied to distinguish isolated muons from  
 1283 those coming from background processes, mainly from pion and kaon decay. The variables  
 1284 with good discriminating power used are described in Ref. [121]. Four identification  
 1285 selections are defined: Loose, Medium, Tight, and High- $p_T$ . The first three categories are  
 1286 inclusive, and Medium being the default selection in **ATLAS**. Finally, the muon candidates  
 1287 to be used by the analyses are asked to satisfy the isolation requirements, both at traack  
 1288 and calorimetric levels, analogously to what was detailed for photons in the previous  
 1289 section. For the first case, a variable similar to that used for photons is used, but with a  
 1290 variable-radius cone  $\Delta R = \min(10 \text{ GeV}/p_T, 0.3)$  around the muon  $p_T$ , excluding the muon  
 1291 track. For calorimetric isolation the same variable  $E_T^{\text{cone}40}$  is used, with the difference of  
 1292 using a radius of  $R = 0.2$ , instead of 0.4 as before. Based on these variables, 7 isolation  
 1293 selection criteria (7 **WPs**), optimized for different analyses, are defined.

### 1294 3.4 Jets

1295 Due to color confinement in **QCD**, a quark or gluon cannot exist on its own and goes  
 1296 through hadronization to form a collimated color-neutral stream of particles, *jets*. Gener-  
 1297 ally, jets penetrate through the **ECAL** and get fully absorbed by the material in the hadronic

1298 calorimeter. In the following, a brief description of the typical clustering method adopted  
 1299 by [ATLAS](#) is given. Also, the two existing types jet reconstruction are described.

### 1300 3.4.1 Anti- $k_t$ jet clustering algorithm

1301 Given that jets are constituted by a high number of particles that leave energy depos-  
 1302 itions in the [ECAL](#) and [HCAL](#) and tracks in the [ID](#), a clustering algorithm groups to-  
 1303 gether constituents in the event to define the jets. Said algorithm is called the anti- $k_t$   
 1304 algorithm [122]. In the same way as for electrons and photons, [ATLAS](#) jet reconstruction  
 1305 relies on the formation of topo-clusters: grouped energy depositions in the calorimeters  
 1306 cells using a sequential combination algorithm. Then, the anti- $k_t$  algorithm combines the  
 1307 topo-clusters with the following steps:

- 1308 • Measure the distance between all topo-clusters between themselves, and of each  
 1309 topo-cluster with the beam:

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

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

1310 where  $\Delta_{ij}^2 = \Delta\phi_{ij}^2 + \Delta\eta_{ij}^2$  and  $R$  is the jet-radius.

- 1311 • If the minimum of all the distances computed previously is  $d_{iB}$ , the topo-cluster  $i$  is  
 1312 classified as a jet, and is discarded in successive iterations.  
 1313 • If the minimum of all the distances is  $d_{ij}$ , topo-clusters  $i$  and  $j$  are combined, all the  
 1314 distances are computed again with this new topo-cluster and the iteration is carried  
 1315 all over again.

1316 This process is repeated until all the particles in the event have been clustered.

1317 The anti- $k_t$  algorithm starts by clustering the radiation around the hardest particle in the  
 1318 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.  
 1319 This allows jets in the event to have a stable direction early on the combination process.  
 1320 The anti- $k_t$  algorithm is preferred to other sequential jet algorithms since jets have regular  
 1321 boundaries which are approximately conical, shown in Figure 3.4. Jets originating from  
 1322 quarks or gluons in general are called small- $R$  jets and a radius of  $R = 0.4$  is used for  
 1323 their reconstruction. On the other hand, jets representing massive particles which decay  
 1324 hadronically are called large- $R$  jets, and use  $R = 1.0$ . The usage of a wider cone helps to  
 1325 include the majority of the particles product of the decay.

### 1326 3.4.2 Calorimeter Jets

1327 One way to reconstruct jets is based on energy deposits in the calorimeter. In a similar way  
 1328 to what has been explained for electrons and photons in Section 3.2.1, energy depositions

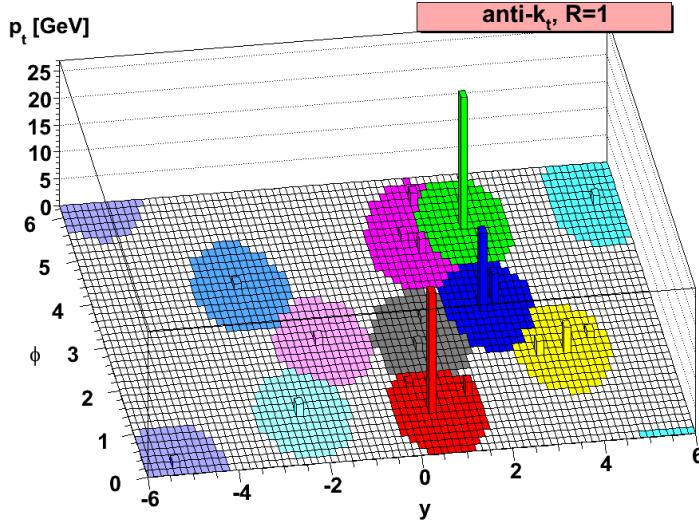


Figure 3.4: Schematic representation of the anti- $k_t$  algorithm for jet clustering [122].

1329 on the cells of the **ECAL** and **HCAL** are used to build topo-clusters, which approximates  
 1330 the energy deposits of individual hadrons [119, 120]. Jets reconstructed in this manner  
 1331 and clustered with the anti- $k_t$  algorithm with a radius of  $R = 0.4$  are referred as EMTOPO  
 1332 jets, and are the proxies for the individual quarks and gluons. In the jet reconstruction,  
 1333 only the topo-clusters with positive net energy are included.

### 1334 3.4.3 PFlow Jets

1335 Another approach for jet reconstruction is taken in the particle flow algorithm, in which  
 1336 measurements from both the tracker and the calorimeter are combined to form the signals,  
 1337 which ideally represent individual particles. In this alorithm, tracks are matched to topo-  
 1338 clusters using proximity in  $(\eta, \phi)$  space, also accounting for the size of the topo-cluster. The  
 1339 tracks are only “matched” if the cluster carries more than 10% of the track’s momentum.  
 1340 Sometimes the topo-cluster fails to cluster all of the hadron’s energy in a single topo-  
 1341 cluster. In cases where the expected energy of the track is less than the expected track’s  
 1342 energy, a “split shower recovery” combines nearby topo-clusters to form a topo-cluster set.  
 1343 From this topo-cluster set, the expected energy of the track is subtracted from the topo-  
 1344 cluster’s cells, starting with high-energy density cells. If the residual energy is consistent  
 1345 with the resolution of the expected track energy, the residual energy is also subtracted in  
 1346 the last step called “remnant removal”.

1347 The result of this algorithm is a set of tracks, modified and unmodified topo-clusters which  
 1348 are the Particle Flow (**PFlow**) objects. The **PFlow** objects can also be clustered with the  
 1349 anti- $k_t$  algorithm and the same  $R = 0.4$  to form **PFlow** jets.

1350 There quite a lot of benefits of using the **PFlow** algorithm over the EMTOPO one:

- 1351     • The momentum resolution of the tracker is significantly better than the calorimeter’s

- 1352        energy resolution for low-energy charged particles.
- 1353     • Allows for a higher acceptance for softer particles. Tracks are reconstructed for  
 1354        charged particles with a minimum  $p_T$  of 400 MeV, and oftentimes these particles'  
 1355        energy deposits do not pass the thresholds to seed topo-clusters.
- 1356     • Improved angular resolution of a single charged particle as it uses the tracker in-  
 1357        formation instead of the calorimeter's.
- 1358     • Low- $p_T$  charged particles originating within a hadronic jet are swept out of the jet  
 1359        cone by the magnetic field by the time they reach the calorimeter. By using the tracks  
 1360        azimuthal coordinate at the perigee, these particles are clustered into the jet.
- 1361     • It is possible to remove those tracks originating from pile-up, knowing that these do  
 1362        not originate from the **PV**.
- 1363        However, particle flow introduces a complication. For any particle whose track measure-  
 1364        ment ought to be used, it is necessary to correctly identify and subtract its signal in the  
 1365        calorimeter to avoid double-counting. In the particle flow algorithm, a boolean decision is  
 1366        made as to whether to use the tracker or calorimeter measurement. The ability to accur-  
 1367        ately subtract all of a single particle energy, without removing energy deposited by other  
 1368        particles, forms the key performance criterion upon which the algorithm is optimised.
- 1369        In this thesis, **PFlow** jets are considered, as they have proven to provide better jet recon-  
 1370        struction [123], principally for those with low  $p_T$  and in the  $E_T^{\text{miss}}$  reconstruction [124].

#### 1371    3.4.4 Jet calibration

1372        Once the jets are reconstructed, their 4-momentum is corrected to match the kinematics of  
 1373        a truth jet<sup>1</sup>, as shown in Figure 3.5. The first three corrections account for contamination  
 1374        from the underlying pile-up distribution and fluctuations due to the origin of the jet [125].  
 1375        The Global Sequential Calibration improves the jets  $p_T$  resolution (and associated un-  
 1376        certainties) by sequentially removing the dependence of the reconstructed jet response  
 1377        ( $R = E^{\text{reco}}/E^{\text{truth}}$ ) on key event observables. Finally, the residual differences between  
 1378        data and **MC** are accounted for by measuring the momentum imbalance in  $Z + \text{jets}$ ,  $\gamma + \text{jet}$   
 1379        and multi-jet events.

1380        To reduce the number of jets with a considerable fraction of energy coming from pile-  
 1381        up, the Jet vertex Tagger (**JVT**) algorithm is used. This algorithm \*\*\* update to **NNJVT**  
 1382        reconstructs a multivariate discriminant that combines, among other quantities, the Jet  
 1383        vertex Fraction (**JVF**) (fraction of the tracks'  $p_T$  associated to a jet originating from the  
 1384        **PV**, and the total number of tracks) and the number of **PVs** in the event  $N_{\text{PV}}$ . As the jets  
 1385        that do not originate from the hard-scatter interaction are generally softer, the **JVT** cut is

---

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

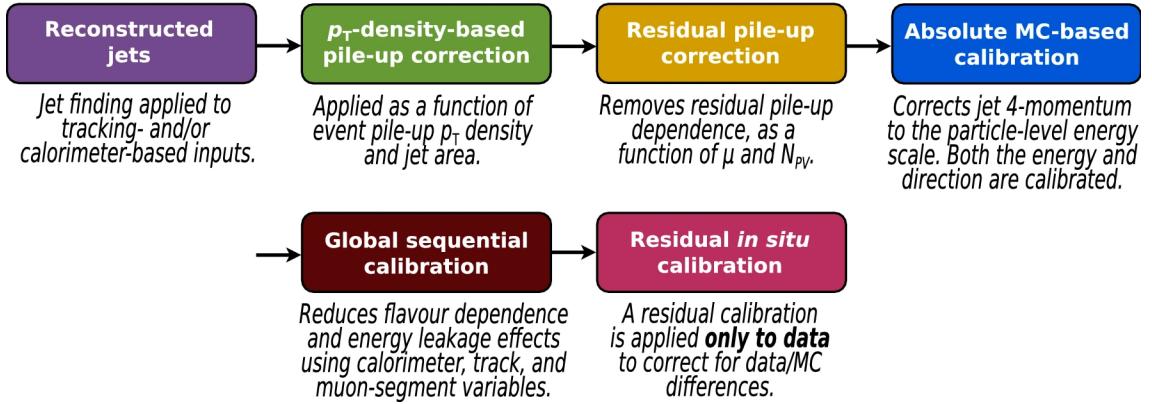


Figure 3.5: PFLow 4-momentum jet calibration steps [125].

1386 applied only to jets with  $p_T < 60$  GeV and  $|\eta| < 2.4$ . The default **JVT WP** is 96% efficient  
 1387 for hard-scatter jets.

### 1388 3.5 Jet flavor tagging

1389 Heavy hadrons decays are governed mainly by the heaviest hadron in the decay cascade.  
 1390 A  $b$ -hadron generally decays through a cascade to a  $c$ -hadron, which in turn decays to an  
 1391  $s$ -hadron, etc, which leads to the existence of multiple vertices.

1392 Flavor Tagging (**FTAG**) is the classification of jets containing  $b$ -hadrons ( $b$ -jets),  $c$ -hadrons  
 1393 ( $c$ -jets) or neither  $b$ - or  $c$ -hadrons (light-flavour jets, or  $l$ -jets) by using algorithms sensitive  
 1394 to the distinctive properties of the respective classes. These complex algorithms rely on  
 1395 the multiple vertices, on the high mass, high decay multiplicity and characteristic decay  
 1396 modes of the  $b$ - and  $c$ -hadrons, as well as on the properties of heavy-quark fragmentation.

1397 In **ATLAS** a two-step approach is employed to reconstruct key characteristics of heavy-  
 1398 flavour jets. In the first stage, low-level algorithms use complementary methods to extract  
 1399 track information from the charged particles linked to the jet. Some algorithms focus  
 1400 on the properties of individual tracks, while others analyse their correlations or combine  
 1401 them to explicitly reconstruct displaced vertices. In the second stage, the outputs from  
 1402 these algorithms are integrated into a high-level algorithm using multivariate classifiers  
 1403 to optimize performance. Over time, the algorithms have evolved significantly, starting  
 1404 with likelihood-based discriminants and boosted decision trees during **LHC** Run-1, and  
 1405 progressing to more advanced methods like recurrent and deep neural networks, resulting  
 1406 in notable improvements on the identification performance [126–129].

1407 Starting in Run-3, a novel Transformer-based "GN2" algorithm is developed by the **FTAG**  
 1408 combined performance group in **ATLAS**. The GN2 algorithm is a single trained model  
 1409 which supersedes DL1d [130] and the low level algorithms that feed it. It is based on  
 1410 GN1 [131], and was quickly refied into GN2. GN2 replaces the Graph Attention Net-  
 1411 work [132] used by GN1 with a Transformer [133], and also benefits from several other  
 1412 architectural optimisations and from an order of magnitude more training statistics.

1413 GN2 directly accepts information about the jet and associated tracks and as such does not  
 1414 depend on other flavour tagging algorithms. GN2 retains the two auxiliary training  
 1415 objectives that were introduced with GN1: the grouping of tracks originating from a com-  
 1416 mon vertex, and the prediction of the underlying physics process from which each track  
 1417 originated.

1418 This new algorithm is also prepared to provide identification of  $c$ -jets and jets originating  
 1419 from  $\tau$  decays. Outputs of this tagger comprise the probabilities of a jet to be tagged as a  
 1420  $b$ -,  $c$ -,  $\tau$ - and light-flavor jet, labeled as  $p_b$ ,  $p_c$ ,  $p_\tau$  and  $p_u$ , respectively.

### 1421 3.5.1 $b$ -jet identification performance

1422 In order to evaluate the performance of the tagger of identifying  $b$ -jets at a constant ef-  
 1423 ficiency, the ability to reject  $c$ -,  $\tau$ - and light-flavor jets is measured. The tagger output  
 1424 probabilities are combined to build a single discriminant  $\text{GN2}_b$ , defined as

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

1425 The parameters  $f_{c(\tau)}$  are free and determine the weighting between  $p_{c(\tau)}$  and  $p_u$  in the dis-  
 1426 criminant. The specific values of these parameters are determined through an optimisation  
 1427 procedure aimed at maximising the rejection of  $c$ -jets ( $\tau$ -jets) and  $l$ -jets and found to be  
 1428 0.2 (0.01).

1429 From the tagger discriminant score, several WPs can be defined, simply by requiring the  
 1430  $\text{GN2}_b$  score to be above a certain threshold. The FTAG working group provides centrally  
 1431 to the whole ATLAS collaboration 5 different WPs to achieve a fixed overall  $b$ -tagging ef-  
 1432 ficiency: 65, 70, 77, 85 and 90% efficiency, and are shown in Figure 3.6. In said figure, the  
 1433 data and MC GN2 tagger distributions are compared, where the different flavour contrib-  
 1434 utions are shown with different colors.

1435 One key challenge of  $b$ -tagging is the decrease in efficiency at higher  $p_T$ . In this high-  
 1436  $p_T$  regime, particles become more collimated and they tend to travel further in the ID  
 1437 before decaying, potentially leading to a decay track with spurious hits. The degraded  
 1438 efficiency is visualised in Table 3.2, where tagging efficiencies are shown for  $b$ -jets, along  
 1439 with  $c$ -jets,  $l$ -jets and  $\tau$ -jets rejections, in the low and high- $p_T$  regimes. The values shown  
 1440 are computed by using different samples, where  $t\bar{t}$  is used at low- $p_T$  and  $Z'$  decay events <sup>2</sup>  
 1441 are used in the high- $p_T$  region. It can be seen that the  $b$ -tag efficiency drops by 30% for  
 1442 higher  $p_T$  jets.

---

<sup>2</sup> 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.

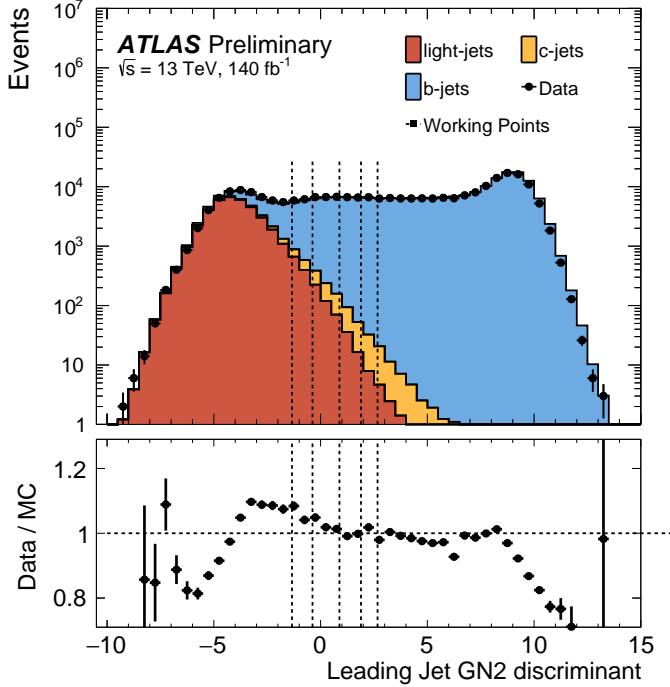


Figure 3.6: GN2 tagger discriminant comparison between data and single-lepton  $t\bar{t}$  MC simulation. The  $l$ -,  $b$ - and  $c$ -jets are contributions shown with different colors, and the 5  $b$ -tag WPs shown with the dashed vertical lines. From left to right, the dashed lines represent the 90, 85, 77, 70 and 65% efficiency WPs. The lower pad shows the ratio between data and the stacked MC [134].

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

| Sample     | $p_T$ range [GeV]  | $b$ -efficiency | $c$ -rejection | light-flavor rejection | $\tau$ -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,  $\tau$ -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  $b$ -,  $\tau$ - 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  $\tau$ -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  $\text{GN2}_c$  score: 10, 30 and 50%  $c$ -tag efficiency. The efficiency and rejection

<sup>1455</sup> measurements for the both samples described above, after applying the 50% *c*-tag **WP** are  
<sup>1456</sup> shown in Table 3.3.

Table 3.3: Measured *c*-tagging efficiencies and *b*-jets, *l*-jets and  $\tau$ -jets rejections in the low and high- $p_{\text{T}}$  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     | $p_{\text{T}}$ range [GeV ] | <i>c</i> -efficiency | <i>b</i> -rejection | light-flavor rejection | $\tau$ -rejection |
|------------|-----------------------------|----------------------|---------------------|------------------------|-------------------|
| $t\bar{t}$ | $20 < p_{\text{T}} < 250$   | 0.467                | 17.52               | 448.61                 | 71.15             |
| $Z'$       | $250 < p_{\text{T}} < 6000$ | 0.344                | 20.27               | 179.99                 | 452.94            |

<sup>1457</sup> \*\*\* Add plots with GN2 distributions?

1458

## Part III

1459

# Photon shower shape corrections

1460

1461 **SHOWER SHAPES AND  
PHOTON IDENTIFICATION**



1462

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

---

Billie Jean King

1463 The [ECAL](#) was presented briefly in Section 2.2.3, where the measurement mechanism and  
 1464 all the layers it has were described. In this subdetector, photons deposit their energy  
 1465 via electron-positron pair creation and bremsstrahlung radiation, creating an [EM](#) shower.  
 1466 The [ECAL](#) does a great job to compute the energy of the [EM](#) shower, but identifying the  
 1467 initiating particle remains a challenging task. However, by virtue of the different layers  
 1468 and granularities in the [ECAL](#), different characteristics of these [EM](#) showers can be studied,  
 1469 and are encoded by different variables called Shower Shapes ([SSs](#)).

1470 This chapter presents all the [SSs](#) that are used to identify real photons from fakes in Sec-  
 1471 tion 4.1. The mentioned variables are heavily used in the process of photon identification,  
 1472 as they provide the separation needed between real and fake photons. For this reason, in  
 1473 Section 4.2, the optimisation of the photon identification (based on the [SSs](#)) is briefly de-  
 1474 scribed, as well as how the photon identification efficiencies are measured in data, using  
 1475 three distinct methods. Finally, Section 4.3 describes one key problematic arising in the  
 1476 simulation of the [SSs](#) and how this is handled, for which a more thorough explanation will  
 1477 be given in the Chapter 5.

1478 **4.1 Shower shapes**

1479 As mentioned in Section 3.2.2, photon identification relies on rectangular cuts applied to  
 1480 [SSs](#) that lead to an excellent separation power between real isolated photons from fake  
 1481 photons originating from hadrons. These [SSVs](#) are computed from the photon candidates'  
 1482 energy deposits in the [ECAL](#) and [HCAL](#) cells, and serve to describe the passage of the  
 1483 photons candidates throughout the calorimeters, characterizing the lateral and longitud-  
 1484 inal [EM](#) showers.

1485 In general, real photons produce narrower energy deposits in the **ECAL**, and have lower  
 1486 leakages to the **HCAL**, compared to those photons provenient from hadrons, where the  
 1487 presence of additional neighbouring hadrons close to the fake photon tend to widen the  
 1488 showers. Furthermore, since the first layer of the **ECAL** consists on fine strips, it is possible  
 1489 to discriminate photon candidates coming from  $\pi^0 \rightarrow \gamma\gamma$  decays, characterized by two  
 1490 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 <b>HCAL</b> to $E_T$ of the <b>EM</b> cluster (used over the ranges $ \eta  < 0.8$ and $ \eta  > 1.52$ )    | $R_{\text{had}1}$     | ✓ | ✓ | ✓ |
|                  | Ratio of $E_T$ in the <b>HCAL</b> to $E_T$ of the <b>EM</b> cluster (used over the range $0.8 <  \eta  < 1.37$ )  | $R_{\text{had}}$      | ✓ | ✓ | ✓ |
| EM second layer  | Ratio of the energy in $3 \times 7 \eta \times \phi$ cells over the energy in $7 \times 7$ cells centered around the photon cluster position                  | $R_\eta$              | ✓ | ✓ | ✓ |
|                  | Lateral shower width in $\eta$  | $w_{\eta 2}$          | ✓ | ✓ | ✓ |
|                  | Ratio of the energy in $3 \times 3 \eta \times \phi$ cells over the energy of $3 \times 7$ cells centered around the photon cluster position                  | $R_\phi$              | ✓ | ✓ |   |
|                  | Lateral shower width in 3 strips around the maximum   | $w_{\eta 1}$ or $w_1$ | ✓ | ✓ |   |
| EM first layer   | 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                                  | $f_{\text{side}}$     | ✓ | ✓ |   |
|                  | Difference between the energy associated with the second maximum in the strip layer with the minimum value found between the first and second maxima.         | $\Delta E$            | ✓ | ✓ |   |
|                  | 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 | $E_{\text{ratio}}$    | ✓ | ✓ |   |
|                  | Ratio of the energy in the first layer to the total energy of the <b>EM</b> cluster   | $f_1$                 | ✓ | ✓ |   |

1491 In the following, the **SSVs** used for photon identification are detailed, shown summarised  
 1492 in Table 4.1 and a scheme of how they are calculated is shown in Figure 4.1. The first  
 1493 variables make use of the energy measured in the **HCAL**:

- 1494 • Hadronic leakage: is the transverse energy deposited in the **HCAL**, normalized to  
 1495 the energy deposited in the **ECAL**:

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

1496 In order to minimize the effects of resolution degradation, in the barrel-endcap transition  
 1497 region of the **HCAL** ( $0.8 \leq |\eta| \leq 1.37$ ) the energy deposit in the whole **HCAL**

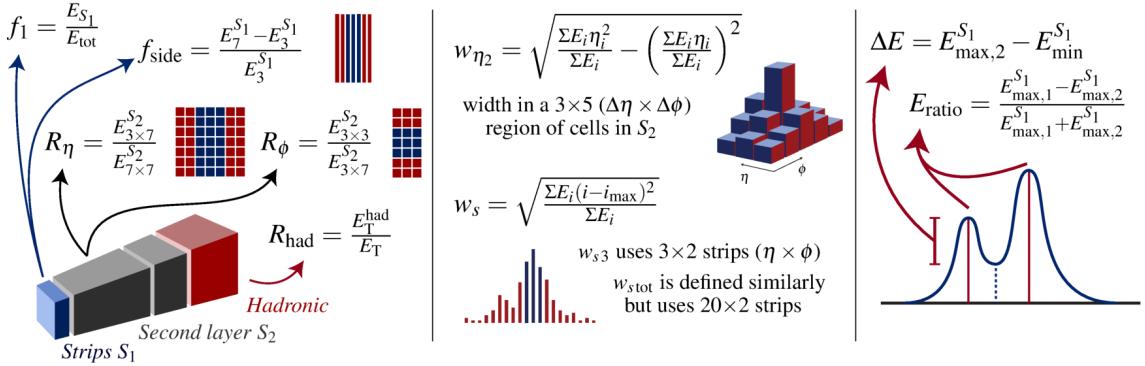


Figure 4.1: Schematic representation of the photon SSSs. The values  $E_C^{S_N}$  represent the energy in layer  $N$  of the ECAL in a cluster  $C$ .

1498        is used ( $R_{\text{had}}$ ). On the remaining of the detector, only the energy deposited in first  
 1499        layer of the HCAL is used ( $R_{\text{had}1}$ ).

1500        The following variables use the second-layer information of the ECAL:

- 1501        • Lateral energy profile in  $\eta$ :

$$R_\eta = \frac{E_{3 \times 7}^{S2}}{E_{7 \times 7}^{S2}} \quad (4.2)$$

1502        where  $E_{i \times j}^{S2}$  is the energy sum in the second calorimeter layer contained in a window  
 1503        of  $i \times j$  cells (units of  $\eta \times \phi$  cells), centered at the most energetic cell. This variable  
 1504        gives a measure of the showers' width in the  $\eta$  direction.

- 1505        • Lateral energy profile in  $\phi$ :

$$R_\phi = \frac{E_{3 \times 3}^{S2}}{E_{7 \times 7}^{S2}} \quad (4.3)$$

1506        defined in a similar way as  $R_\eta$ . However, this variable behaves very different for  
 1507        converted and unconverted photons. Due to the action of the magnetic field, the  
 1508        electrons and positrons are curved into opposite directions in  $\phi$ , therefore leading  
 1509        to wider EM showers for converted photons compared to those from unconverted  
 1510        ones.

- 1511        • Lateral shower width in  $\eta$ :

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

1512        measures the proper width of the EM shower, where  $E_i$  is the energy in the  $i$ -th cell  
 1513        of the ECAL, measured in a window of  $3 \times 5$  cells in  $\eta \times \phi$ .

1514        The following variables use the information from the first ECAL layer, composed of the  
 1515        strip cells that allow for a high  $\eta$  resolution and allows for a good separation between  
 1516        isolated photons from photons product of the  $\pi^0$  decay. Figure 4.2 shows the difference in  
 1517        the energy deposited in the ECAL between the two cases mentioned previously.

- 1518 • Lateral energy profile in  $\eta$

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

1519 measures the energy outside the core of the three central strips within a window of  
1520 7 cells, divided by the energy in the three central cells.

- 1521 • Lateral shower width in  $\eta$  (3 strips)

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

1522 where  $i$  runs over all cells in a window of 3 cells around the highest-energy-cell. This  
1523 variable measures the width of the EM shower in the first layer of the calorimeter.

- 1524 • Lateral shower width in  $\eta$  (full). It is defined in a similar way as  $w_1$ , but uses all the  
1525 cells in a window of  $\Delta\eta \times \Delta\phi = 0.0625 \times 0.2$ , corresponding to approximately to  
1526 20  $\times$  2 strips  $\eta \times \phi$ .

- 1527 • Energy difference

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

1528 represents the energy difference between the second maximum and the minimum  
1529 reconstructed energy between the two maxima in the strip layer.

- 1530 • Energy ratio

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

1531 is the ratio of energy difference between the two maxima, normalized to the sum of  
1532 those energies, in the strip layer.

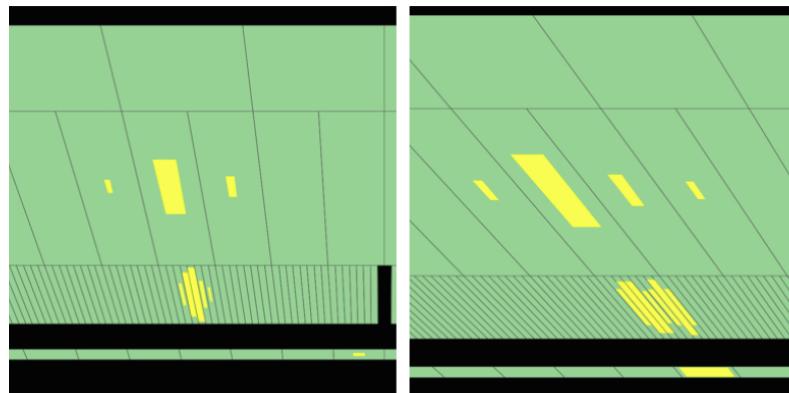


Figure 4.2: Characteristic energy deposits by an isolated photon (left), and a  $\pi^0 \rightarrow \gamma\gamma$  event (right), which is possible to distinguish thanks to the granularity of the first ECAL layer [135].

## 1533 4.2 Photon Identification

1534 The identification of prompt photons over fake photons in hadronic collisions is particu-  
 1535 larly challenging. Fake photons are vastly dominated by reconstructed photon candidates  
 1536 arising from hadron decays in jets, while a smaller fraction of fake candidates are asso-  
 1537 ciated with hadrons that deposit significant energy in the **ECAL**, mimicking that of real  
 1538 photons. Processes with prompt photons in the final state, occurring in  $p\bar{p}$  collisions at the  
 1539 **LHC**, play a central role in the **ATLAS** physics programme. Either for searches or precision  
 1540 measurements, it is important to rely on excellent algorithms and techniques to identify  
 1541 the real photons over the fake ones. These searches or measurements are performed in  
 1542 a very wide range of the photon  $p_T$ , starting from very light Higgs resonances like to a  
 1543 pair of axion-like particles to 4 photons ( $H \rightarrow aa \rightarrow 4\gamma$ ) [136], where the photon  $p_T$   
 1544 is  $\sim 25$  GeV, up to very high- $p_T$  photons in searches for  $\gamma +$ jet resonances, with photons  
 1545 having  $p_T > 1$  TeV. In this section, the procedures for optimising the identification **WPs**  
 1546 and measuring the corresponding efficiencies are described.

### 1547 4.2.1 Processes and event selection

1548 Given the very wide range in which photons are used in **ATLAS**, the optimisation of the  
 1549 tight identification **WP** relies on two different processes that eventually allow for clean  
 1550 photon samples in the low and high  $p_T$  regimes. In the low- $p_T$  case, a very clean source  
 1551 of photons from radiative  $Z$  decays are used. On the other hand, although with higher  
 1552 background contamination, Single Photon (**SP**) events are employed for high  $p_T$  photons.  
 1553 In the following paragraphs, a description of each photon sample is given.

1554 **Radiative  $Z$  decays** In the low- $p_T$  range, photons from radiative decays of the  $Z$  boson  
 1555 ( $Z \rightarrow \ell\ell\gamma$ ) are selected as signal photons. There are two different production modes  
 1556 possible for the **SM**  $p\bar{p} \rightarrow Z(\ell^+\ell^-)\gamma$  processes, where  $\ell$  is either an electron or a muon.  
 1557 These are: Initial State Radiation (**ISR**) where the photon is radiated from the quarks, and  
 1558 Final State Radiation (**FSR**) (hereinafter also referred as Radiative  $Z$  (**RZ**) decays), where  
 1559 the photon is radiated from one of the final-state leptons through bremsstrahlung. Both  
 1560 production modes are shown in Figure 4.3.

1561 Both the **FSR** and **ISR** processes can be easily identified by comparing the two-body in-  
 1562 variant mass  $m_{\ell\ell}$  distribution to the three-body invariant mass  $m_{\ell\ell\gamma}$  distribution. For **ISR**  
 1563 events,  $m_{\ell\ell}$  follows the  $Z$  line-shape, and the photon simply adds to the invariant mass  
 1564 making it larger than 91 GeV. In the **FSR** case, the three-body invariant mass  $m_{\ell\ell\gamma}$  follows  
 1565 the  $Z$  line-shape, which is seen from Figure 4.4a. For photon identification studies, only  
 1566 photons from **FSR** events are considered. The reason behind the selection of **FSR** over **ISR**  
 1567 is driven by the following. **ISR** events also suffer from  $Z +$  jets background contamination,  
 1568 where the jet is misidentified as a photon, and the  $Z +$  jets cross-section is magnitudes  
 1569 higher than for  $Z + \gamma$ . From Figures 4.4b and 4.4c, where the  $m_{\ell\ell}\text{-}m_{\ell\ell\gamma}$  distributions are

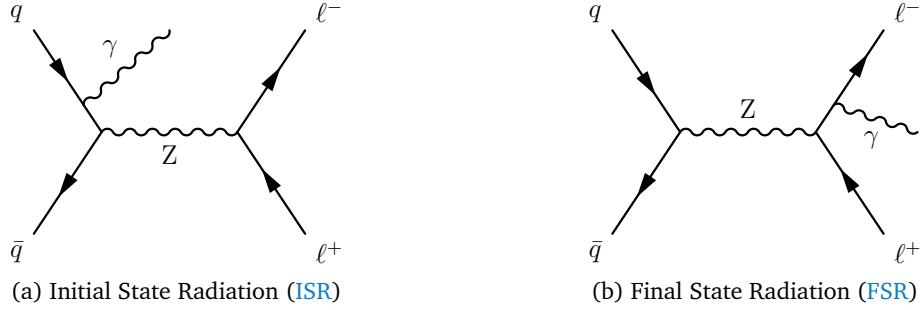


Figure 4.3: Feynman diagrams for radiative  $Z$  decays  $Z \rightarrow \ell\ell\gamma$  for the initial-state (left) and final-state (right) radiation.

1570 shown for  $Z \rightarrow \ell\ell$  and  $Z \rightarrow \ell\ell\gamma$  processes, respectively, the separation between these two  
1571 can be appreciated when selecting FSR photons.

1572 With the RZ sample, the photons are required to have a transverse momentum  $p_T > 7$  GeV  
1573 and a pseudorapidity in the range  $|\eta| < 1.37$  or  $1.52 < |\eta| < 2.37$ , to avoid the crack  
1574 region. For the optimisation studies no photon isolation is applied, but loose photon  
1575 isolation, described in Section 3.2.3, is used for the efficiency measurements. To avoid  
1576 any biases on the photon footprints in the calorimeter, no other selection is applied for  
1577 them. Leptons are required to have  $E_T > 10$  GeV, muons pseudorapidity to be  $|\eta| < 2.5$   
1578 and for electrons  $|\eta| < 2.47$ , excluding the crack. Both electrons and muons need to pass  
1579 loose isolation requirements and to pass medium identification.

1580 FSR photon then are selected by requiring  $80 < m_{\ell\ell\gamma} < 100$  GeV, and  $40 < m_{\ell\ell} < 83$  GeV.  
1581 Finally, to avoid any biases on the photon SS and isolation variables, a minimum distance  
1582 of  $\Delta R > 0.4$  is required between the photon and the closest lepton.

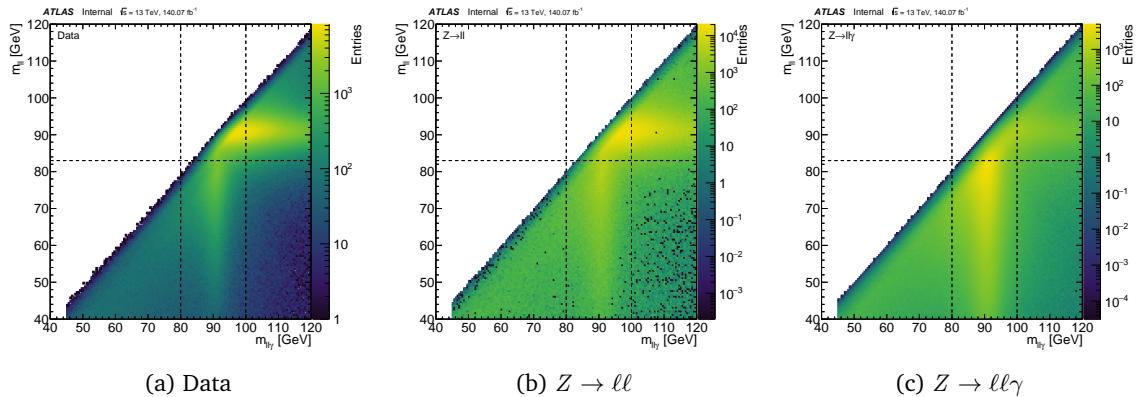


Figure 4.4: Two-dimensional invariant mass distribution of the two and three body systems  $m_{\ell\ell}$  and  $m_{\ell\ell\gamma}$ , respectively, for (a) data, (b) background and (c) signal. The region in which there is a high concentration of events for  $m_{\ell\ell} \sim m_Z$  corresponds to ISR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

1583 **Single Photon (SP)** The inclusive photon, or Single Photon (SP) sample is collected by  
1584 single-photon triggers, whose thresholds range from 10 GeV up to 140 GeV and require  
1585 loose photon identification. Although the triggers used to obtain this sample are prescaled  
1586 (with the exception of the 140 GeV one) the provide very large photon datasets for high  
1587  $p_T$ . These processes include leading-order  $\gamma + \text{jet}$  events from  $qg \rightarrow q\gamma$  and  $q\bar{q} \rightarrow g\gamma$

1588 hard scattering, as well as prompt photons from quark fragmentation in QCD dijet events.  
 1589 Photons from these events are required to satisfy  $|\eta| < 2.37$  excluding the crack, and to  
 1590 pass the loose isolation requirement. The SP samples are used for both the optimisation  
 1591 and measurements studies.

### 1592 4.2.2 Optimisation

1593 Starting from these discriminating SSs, three WPs can be defined for photons: *loose*, *me-  
 1594 dium* and *tight* [137]. The loose WP employs cuts to the variables defined in the second  
 1595 layer and to the hadronic leakage variable, used primarily by the trigger. The medium  
 1596 and tight WPs use all the previously defined variables. The former is optimised to have a  
 1597 flat 95% efficiency, while the latter provides an excellent background rejection. Table 4.1  
 1598 shows which variables are used for each WP.

1599 The optimisation of the WPs uses two different samples: RZ events for photons with  
 1600  $10 < p_T < 25$  GeV as signals and  $Z \rightarrow ll$  as backgrounds; and for the high- $p_T$  regime  
 1601 ( $p_T > 25$  GeV) SP events are considered signals while dijet events are the backgrounds.

1602 As previously mentioned, photon identification uses cuts to photon Shower Shapes (SSs).  
 1603 Examples of the  $R_\eta$ ,  $E_{\text{ratio}}$  and  $w_{\eta 2}$  SS comparing signal and background events using the  
 1604 RZ samples are shown in Figure 4.5, where excellent discriminating power is seen. The  
 1605 cuts on all the SSs, for each identification WP, are optimised as a function of the transverse  
 1606 energy and the pseudo-rapidity of the photon candidate, to account for the shape of the  
 1607 variables for different  $\eta$  and for variations in the amount of material and the geometry of  
 1608 the calorimeter. The medium and tight WPs are also optimised separately for converted  
 1609 and unconverted photons. The cuts are optimised using a Multivariate (MV) approach,  
 1610 where signal efficiencies are scanned between 0% and 100% while trying to maximise the  
 1611 background rejection. The resulting, optimised, cut values are subject to fluctuations and  
 1612 therefore they are manually smoothed.

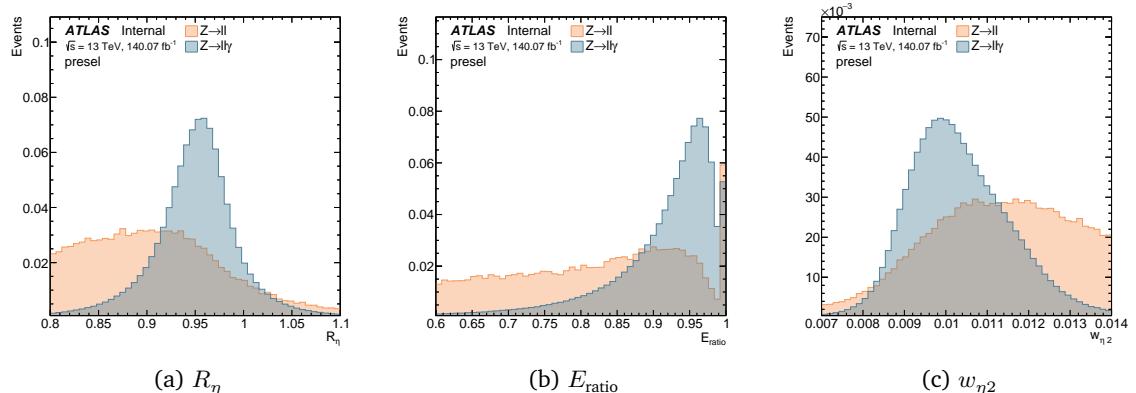


Figure 4.5: Normalised signal (blue) and background (orange) distributions for different SSs, using RZ event samples, passing the event selection detailed in Section 4.2.1.

### 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. [138] 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:  $P^{\text{total}}$  before applying tight identification and  $P^{\text{pass}}$  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 [139] 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. [118]. 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 [138] 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%.

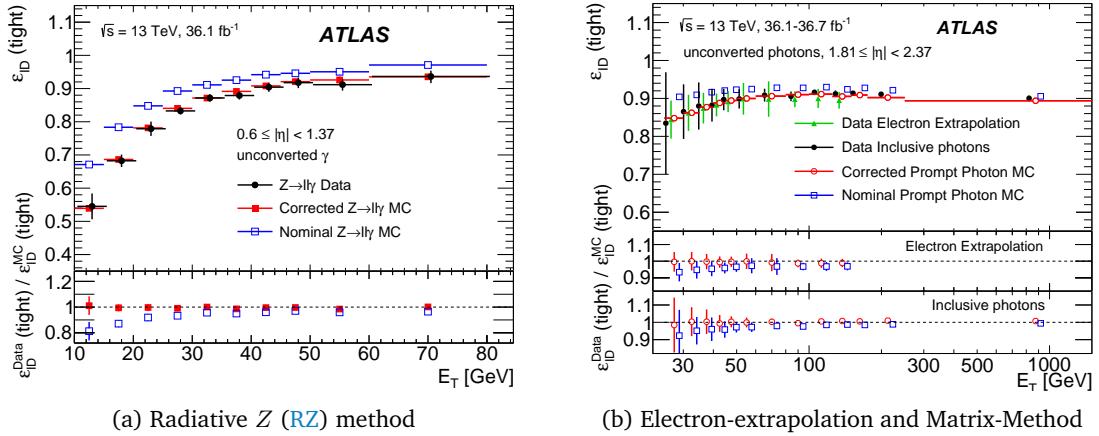


Figure 4.6: Comparison of the measurements between data and MC of the three distinct data-driven methods to compute efficiencies. In both figures, for each method, two different set of MC measurements are displayed: the nominal one and the corrected one. The bottom panels show the ratio of data efficiencies to MC predictions (referred as Scale Factors (SFs) in the text). The figures are taken from Ref. [138].

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 [140] 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.

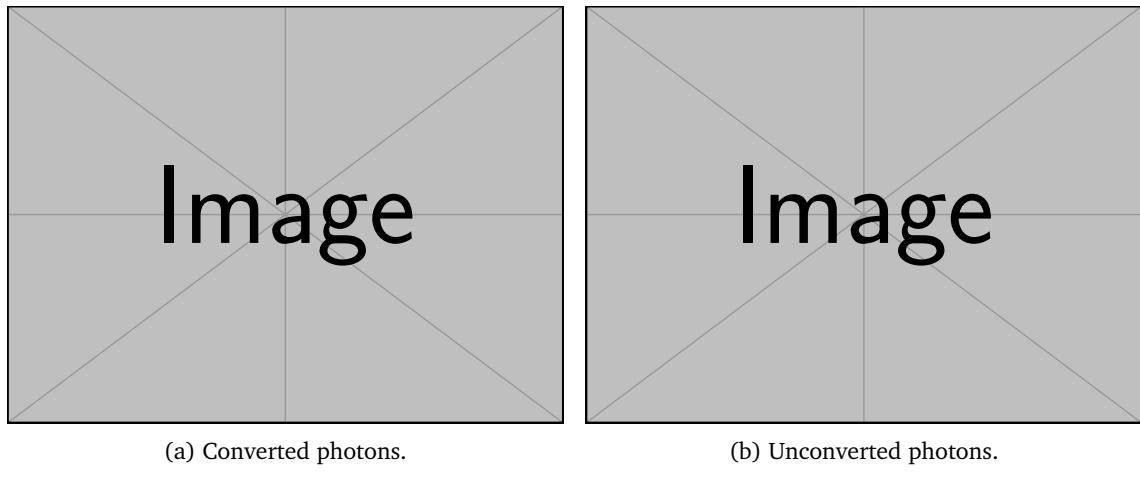


Figure 4.7: Photon identification SFs in the different  $p_T$ - $|\eta|$  bins for both converted (left) and unconverted photons (right). \*\*\* Ask fran for the plots in his presentation!

## <sup>1671</sup> 4.3 Shower shapes variables differences between data and MC

1672 The [ATLAS MC](#) simulation does not perfectly describes data. This is clearly seen when  
1673 computing the previsously mentioned [SFs](#), whose values were different from 1, meaning  
1674 that different efficiencies are obtained between data and in [MC](#). In particular, when com-  
1675 paring the [SSs](#) distributions, it is seen that [MC](#) distributions are shifted or even the whole  
1676 shape differs, as shown in Figure 4.8, by comparing data (black dots) against the red line  
1677 histogram corresponding to [MC](#).

1678 The main differences on the distributions arise for the  $\eta$  shower profiles, where broader  
1679 distributions are seen in data compared to MC. Part of the effect was corrected in 2010  
1680 after moving to detailed description of the material composition in the accordion absorbers  
1681 in GEANT4. However, the remaining data-MC disagreements are still under study and  
1682 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  $n$  direction).

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  $\chi^2$  minimisation on the comparison of data and MC SSs [118, 138]. Even though the mean value differences decreased substantially after these corrections as seen for example in the case of  $f_{\text{side}}$  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 cor-

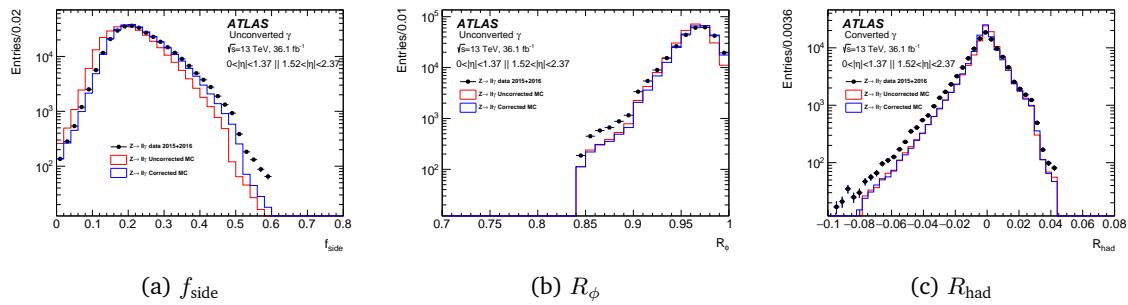


Figure 4.8: Example of **SSs** comparisons between data (black dots) against nominal (red line) and corrected (blue line) **MC** simulation, using the full Run-2 **RZ** photon sample [141].

rections. 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.

1700

# 1701 SHOWER SHAPES CORRECTIONS



1702

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

---

Billie Jean King

1703 In the previous chapter, it was seen that **SFs** (ratio of data efficiencies to simulated ones)  
 1704 deviate from unity. Given that photon identification relies on cuts to the photon **SSs**,  
 1705 it was found that the differences in fact appear on these variables. Since Run-1, they  
 1706 have been corrected with what is known as Fudge Factors (**FFs**), which, historically, they  
 1707 have been computed as simple shifts to the **MC** distributions and have been found to  
 1708 provide very good improvements on the **SFs**. However, as seen before, there are still  
 1709 discrepancies between the distributions that need to be addressed in order to rely on even  
 1710 a better simulation. In Section 5.1, a more sophisticated approach based on a higher  
 1711 order computation to correct the **SSs** is presented. Also, a novel approach using directly  
 1712 the cells energies is studied and addressed in Section 5.2. The studies presented in this  
 1713 chapter comprise one of the main topics of work for the current thesis.

## 1714 5.1 Fudge Factors (**FFs**)

### 1715 5.1.1 Data and simulated samples

1716 **FFs** are computed using full Run-2 dataset, collected at  $\sqrt{s} = 13$  TeV and with a corres-  
 1717 ponding integrated luminosity of  $140 \text{ fb}^{-1}$ . Both the **RZ** and **SP** simulated samples are  
 1718 used for this study, as their represent complementary  $p_T$ -ranges. **RZ** events are generated  
 1719 with SHERPA 2.2.11 [69], while SHERPA 2.2.1 is used for  $Z \rightarrow \ell\ell$  background events.  
 1720 Respecting the **SP** samples, events are generated with PYTHIA 8.186 [66], which includes  
 1721 leading-order  $\gamma + \text{jet}$  events from both direct processes ( $qg \rightarrow q\gamma$  and  $q\bar{q} \rightarrow g\gamma$ ) and photon  
 1722 fragmentation from **QCD** dijet events.

1723 In both cases, the **ATLAS** detector is simulated using GEANT4 4 [76] and the **MC** events  
 1724 are reweighted so that their pileup distributions resembles the one in data, for each year

1725 of the Run-2 data-taking period.

1726 **5.1.2 Calculation**

1727 The calculation is performed separately for the two considered samples: **RZ** for photons  
 1728 with  $7 \leq p_T \leq 50$  GeV and **SP** for photons with  $p_T > 50$  GeV, which were already discussed  
 1729 in Section 4.2.1. Since **SSs** distributions vary as a function of  $p_T$  and  $|\eta|$ , the computation  
 1730 is done in bins of these variables:

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

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

1731 Furthermore, as mentioned in Section 4.1, there are variables very sensitive to the photon's  
 1732 conversion status, that is, whether if the photons are converted or unconverted. For this  
 1733 reason, the calculation is done separately for converted and unconverted photons. A total  
 1734 of nine variables are corrected using this method:  $E_{\text{ratio}}$ ,  $f_{\text{side}}$ ,  $R_\eta$ ,  $R_\phi$ ,  $R_{\text{had}}$ ,  $R_{\text{had}\,1}$ ,  $w_1$ ,  $w_{\eta 2}$   
 1735 and  $w_{\text{s tot}}$ ; as they are the ones in which the largest discrepancies are observed between  
 1736 data and **MC**.

1737 For each **SS**, histograms of **MC** and data of 100 bins are created. The choice of the binning  
 1738 is done based on having sufficient statistics at each bin and also to capture all the features  
 1739 of the variables. After that, each histogram is smoothed using the Kernel Density Estimator  
 1740 (**KDE**) tool from TMVA [142]. The **KDE** method consists of estimating the shape of a  
 1741 Probability Density Function (**PDF**) by the sum over smeared events. The **PDF**  $p(x)$  of a  
 1742 variable  $x$  is

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

1743 where  $N$  is the number of events,  $K_h(t) = K(t/h)/h$  is the kernel function, and  $h$  is the  
 1744 bandwidth of the kernel. The basic idea is that each event is considered as a Dirac- $\delta$ -  
 1745 function, which is replaced by a Kernel function (Gaussian) and finally they are summed  
 1746 altogether to form the final **PDF**. The **KDE** smoothing can be applied in two forms: non-  
 1747 adaptive **KDE** or adaptive **KDE**, as seen in Figure 5.1. In the former, the bandwidth is  
 1748 constant for the entire sample  $h_{NA}$ , while in the latter, it uses the value from non-adaptive  
 1749 **KDE**, but it varies as a function of  $p(x)$  as

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

1750 Adaptive **KDE** improves the shape of the estimated **PDF** in regions of low statistics, how-  
 1751 ever, in high statistics regions it can give rise to over-smoothing. The degree of smoothing  
 1752 is tuned by multiplying the bandwidth  $h$  by fine factors. These fine factors are user-defined  
 1753 parameter which are tuned to allow the **PDF** to retain the important features of the **SS** but  
 1754 to also avoid statistical fluctuations. Higher values indicate broader Kernel functions and

1755 therefore de PDF catches less statistical fluctuations. Examples of the smoothing procedure  
 1756 applied to  $R_{\text{had}}$  are shown in Figure 5.2 for cases in which original histograms have  
 1757 low and high statistics.

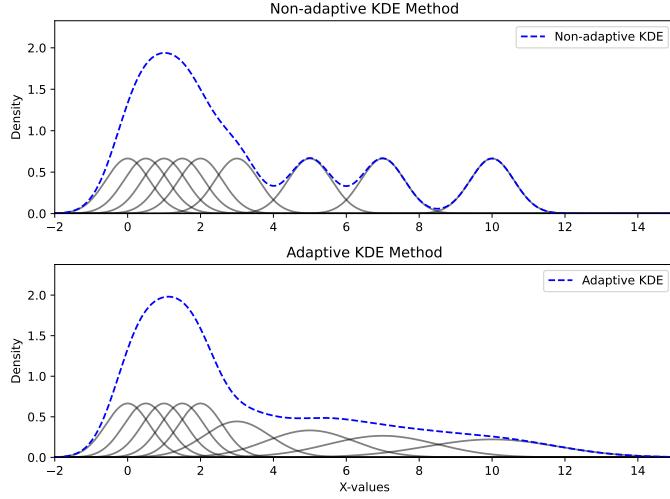


Figure 5.1: Adaptive and non-adaptive KDE smoothing.

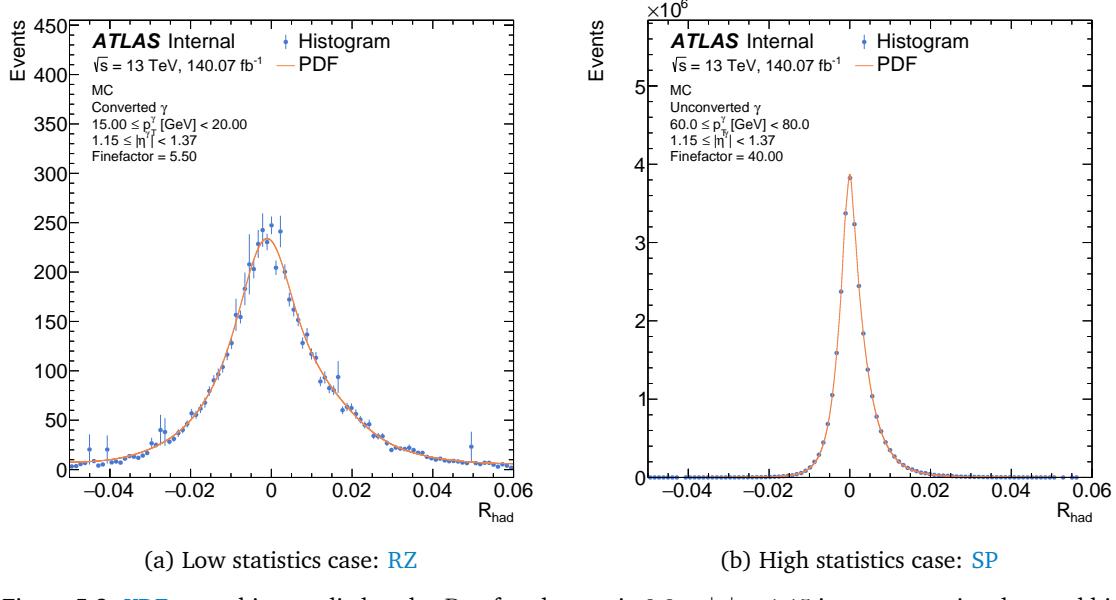


Figure 5.2: KDE smoothing applied to the  $R_{\text{had}}$  for photons in  $0.8 < |\eta| < 1.15$  in two scenarios: low and high statistics. Original histograms are shown with the blue points and their PDFs with the orange lines. The fine factors used for each case are displayed in the figure.

1758 Once the data and MC PDFs are created for a given variable,  $p_T$ ,  $|\eta|$  and conversion type,  
 1759 the MC PDF is normalised to data's and a  $\chi^2$  value is computed between both, excluding  
 1760 the underflow and overflow bins, as:

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

1761  $N$  is the number of bins in the PDFs,  $w_{\text{MC},i}$  and  $w_{\text{data},i}$  are the event numbers of MC and  
 1762 data at each bin, respectively,  $s_{\text{MC},i}$  and  $s_{\text{data},i}$  are the bin errors and finally  $W_{\text{data}}$  and  $W_{\text{MC}}$   
 1763 are the sum of weights for data and MC, respectively.

1764 **Shift-only corrections**

1765 Taking into account only the mean's correction of the **SSs**, the **MC PDF** is shifted to the left  
 1766 and right one bin at a time. As a consequence of this procedure, the shift **FF** resolution  
 1767 directly depends on the bin-width of the **PDFs**, therefore having smaller bin-widths mean  
 1768 to obtain a better resolution on the shift value. Given that histograms, in the first place,  
 1769 are built with relative wide bins, the **PDFs** can be constructed using high-accuracy narrow  
 1770 bin to ensure high resolution. After convergence tests on the **FFs**, the **PDFs** are constructed  
 1771 with 5000 bins. The starting number of bins that the **MC** distribution needs to be shifted is  
 1772 estimated by computing the difference on the means between data and simulation. From  
 1773 this starting value, shifts of 100 bins to each side are considered.

1774 For each bin the distribution has been shifted, the aforementioned  $\chi^2$  value is computed  
 1775 and recorded. Assuming that the measurements errors  $s_{MC,i}$  and  $s_{data,i}$  have a normal  
 1776 gaussian distribution <sup>1</sup>, and that the parameters for each  $\chi^2$  value are independent, it is  
 1777 expected that the shape followed by the  $\chi^2$  values is approximately paraboloidal.

1778 To extract the **FFs**, the  $\chi^2$  scan near the minimum is fitted with a parabolic function (5  
 1779 bins to each side of the minimum bin) and the shift **FF** is obtained from the fit minimum.  
 1780 Finally, the **SSs** can be corrected as

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

1781 **Shift+stretch corrections**

1782 It was seen that even after applying corrections to the means of the **MC SSs** distribu-  
 1783 tions, differences remained on the shapes of them, and in some cases these can be quite  
 1784 substantial. One way to continue improving the agreement between data and **MC** is to  
 1785 include another correction which is referred as *stretching*. The two corrections, called  
 1786 shift+stretch corrections, are meant to fix both the mean and the widths of the **MC** distri-  
 1787 butions simultaneously.

1788 The shift+stretch correction method starts by finding the maximum of the **MC PDF**. The  
 1789 **PDF** is then stretched around it by calculating the new position of each bin by the product:  
 1790  $\text{stretch} \times (x - \text{stretch point})$ . In this manner, each bin's center conserves the initial distance  
 1791 to the center of the distribution, multiplied by the stretch factor. In the scenario where  
 1792 the shift is  $> 1$ , there might be cases in which it is big enough to give rise to empty  
 1793 bins inbetween. The content of these empty bins are then linearly interpolated from the  
 1794 neighbouring non-zero bins. Once the **PDF** is stretched, it is then shifted left and right  
 1795 following the same procedure as for the shift-only case, calculating  $\chi^2$  values for each  
 1796 shift<sub>i</sub> after applying stretch<sub>j</sub>. As a result of the procedure, now, a two-dimensional grid  
 1797 on the shift-stretch plane of  $\chi^2$  values is obtained. The pair of shift-stretch **FFs** is now

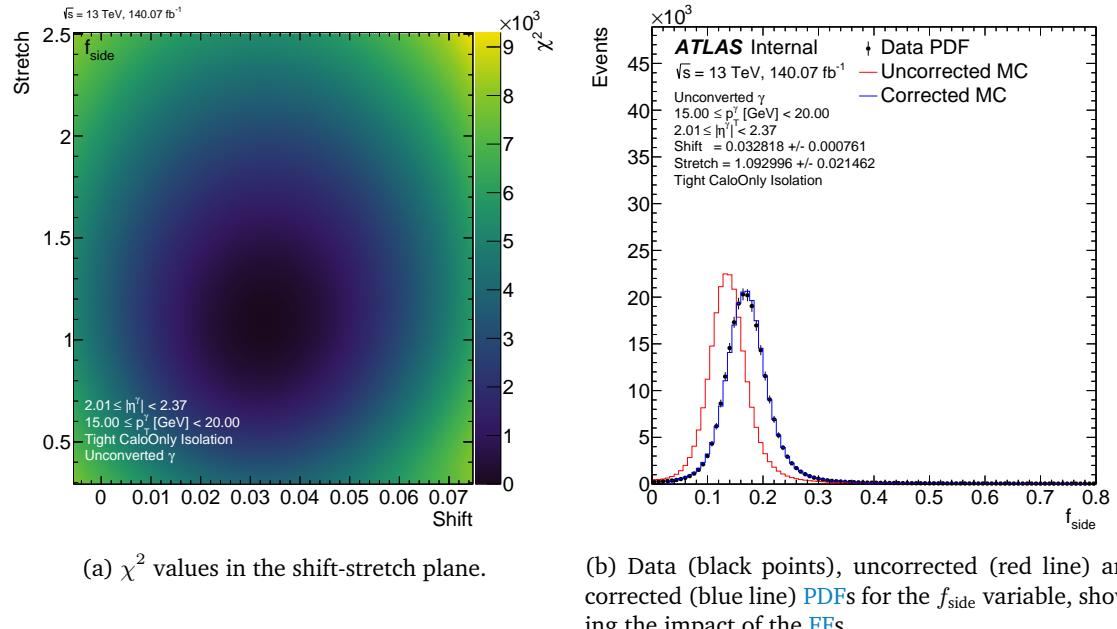
---

<sup>1</sup> 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.

1798 retrieved from the minimum bin's center, and the corrections are applied to the MC SSs  
 1799 as:

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

1800 An example of the resulting  $\chi^2$  values for the  $f_{\text{side}}$  variable for unconverted photons with  
 1801  $15 < p_T < 20 \text{ GeV}$  in  $2.01 < |\eta| < 2.37$  is shown in Figure 5.3a, where the shift is on the  $x$ -  
 1802 axis and the stretch on the  $y$ -axis. The optimal shift-stretch value is given by the position of  
 1803 the minimum bin, which corresponds to shift = 0.03 and stretch = 1.09. A visualisation of  
 1804 the PDFs before and after applying the corrections is shown in Figure 5.3b, where they are  
 1805 compared with the data distribution. As seen from the figure, there is a huge improvement  
 1806 and the distributions match almost perfectly.



(a)  $\chi^2$  values in the shift-stretch plane.

(b) Data (black points), uncorrected (red line) and corrected (blue line) PDFs for the  $f_{\text{side}}$  variable, showing the impact of the FFs.

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

### 1807 5.1.3 Uncertainties

#### 1808 Statistical uncertainties

1809 To extract the statistical uncertainties on the shift and stretch FFs, a fit to the  $1\sigma$  (68.3%)  
 1810 confidence level) contour on the  $\chi^2$  values is performed. This contour represents an ellipse  
 1811 in the large sample limit which takes the following form:

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

1812 where  $\rho$  is the correlation coefficient between both variables,  $\sigma_x$  and  $\sigma_y$  the uncertainties  
 1813 on  $x$  and  $y$ , respectively,  $(x_0, y_0)$  is the location of the ellipse's center, and  $\chi^2_{\min}$  is the  $\chi^2$   
 1814 minimum value obtained from the 2D histogram.

1815 By extracting the semi-major and semi-minor axes of the fitted ellipse, and with the tilt  
 1816 angle of it, the statistical uncertainties on two variables  $x$  and  $y$  (in this case representing  
 1817 the shift and stretch, respectively) are (see Appendix A):

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

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

1818 **Systematic uncertainties**

1819 The systematic uncertainties are derived by varying the preselection criteria, that is, photon  
 1820 identification and photon isolation. Changing different preselection criteria allows the SSs  
 1821 to vary depending on the amount of background contamination, and in consequence so  
 1822 do the FFs. The different selections are:

- 1823 • Radiative  $Z$  (RZ) sample:

- 1824 – Nominal: No ID, FixedCutTightCaloOnly isolation.  
 1825 – Loose ID, no isolation.  
 1826 – Loose ID, FixedCutTightCaloOnly isolation.  
 1827 – No ID, FixedCutLoose isolation.

- 1828 • Single Photon (SP) sample:

- 1829 – Nominal: Tight ID, FixedCutLoose isolation.  
 1830 – Tight ID, FixedCutTight isolation.

1831 All other combinations (or lack thereof) of selection criteria would result in either a sample  
 1832 with too low statistics, or too low purity.

1833 FFs are derived for each one of the previous selections, and the difference between the  
 1834 nominal and the varied ones is calculated. The maximum difference is taken as the sys-  
 1835 tematic uncertainty, as the most conservative approach.

1836 **5.1.4 Results**

1837 Due to the fact that FFs are calculated in different  $p_T$  using two different samples which  
 1838 span complementary regions, the results are concatenated at 50 GeV. in the following,  
 1839 shift and stretch values are reported for different SS variables.

1840 The reported shift values in the figures are normalised by the standard deviation of the SS  
 1841 after stretching, as this quantity allows to understand how much each variable is shifted  
 1842 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  $R_\eta$  and  $w_{\eta 2}$  variables using converted photons are presented. It can be seen that for both variables the FFs depend on  $p_T$ , specifically 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  $\eta$  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  $R_\eta$  ( $w_{\eta 2}$ ) (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.

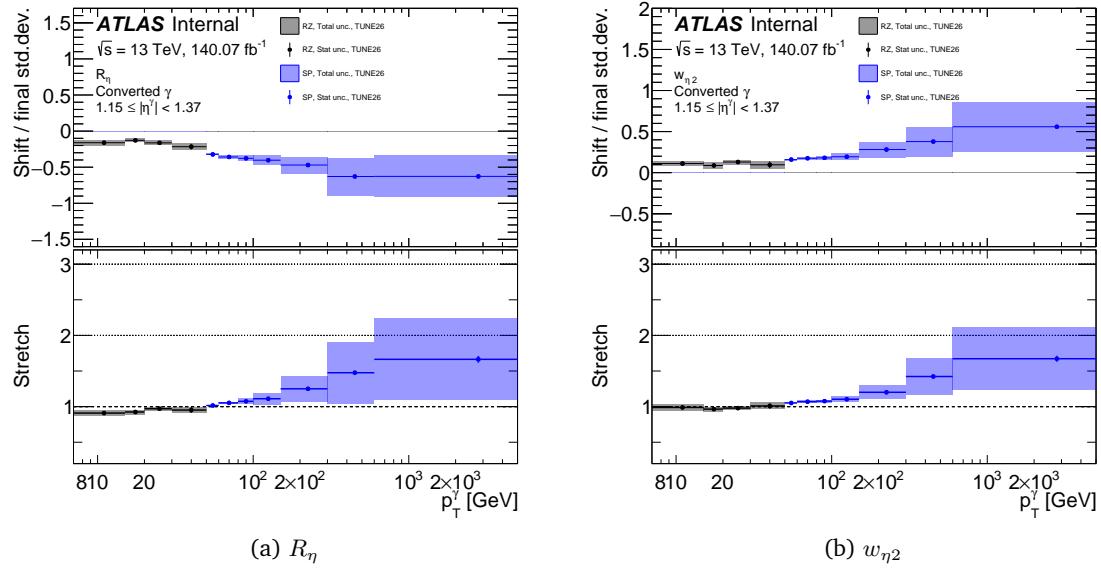


Figure 5.4: Shift and stretch FFs for  $R_\eta$  (left) and  $w_{\eta 2}$  (right) SSs for converted photons as a function of the photon  $p_T$  with  $1.15 < |\eta| < 1.37$ . Results from RZ (black) and SP (blue) are displayed, where the shaded regions denote the total uncertainty, while the error bars 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  $|\eta|$ .

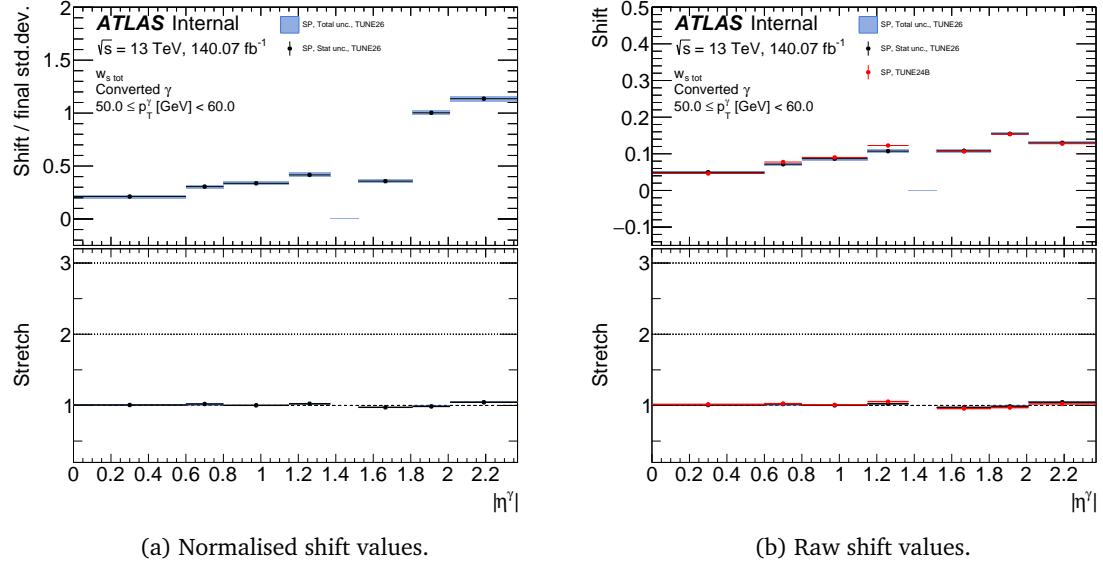


Figure 5.5: Shift and stretch FFs for  $w_{s\text{ tot}}$  SSs using converted photons as a function of  $|\eta|$  and photons from the SP with  $50 < p_T < 60$  GeV. Normalised shift values are displayed in Figure a while original, or raw, values are shown in Figure b. Points with the uncertainty bar 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.

1866 In the barrel region, the corrections indeed improve the agreement, but the magnitudes  
 1867 of these corrections are not as dramatic as in the endcap region, where huge improve-  
 1868 ments are seen. Taking the  $w_1$  and  $w_{s\text{ tot}}$  variables as an example, major shape differences  
 1869 are observed between the nominal simulation and data, which the shift+stretch methods  
 1870 manage to fix. The same behaviour is observed using the SP samples, where these vari-  
 1871 ables present two or more peaks, and they are correctly fixed by the FFs. In all the shown  
 1872 cases, the corrected MC and data are almost undistinguishable showing the importance of  
 1873 these corrections and how they were improved.

## 1874 5.2 Cell-based energy reweighting

1875 The design and functionality of the ATLAS ECAL was described in Section 2.2.3 as well  
 1876 as the process from which electrons and photons deposit their energies in the ECAL: pair  
 1877 creation and bremsstrahlung radiation. Then, from these energy depositions in the ECAL  
 1878 the SSs are built and used for photon identification. However, the fact that MC and data  
 1879 SSs do not match, means that the energy depositions are different between these two,  
 1880 leading to a lower-level disagreement.

1881 Although the Fudge Factor (FF) method described before led to an excellent improvement  
 1882 on the agreement between data and MC distributions, it is still based on modifying high-  
 1883 level variables and all independently of each other. On the other hand, by directly correct-  
 1884 ing the cells' energy depositions in the simulation, a simultaneous fix to all the Shower  
 1885 Shapes (SSs) and any other variable computed from the energies would be acquired. This  
 1886 processes is what is known as cell-based energy reweighting.

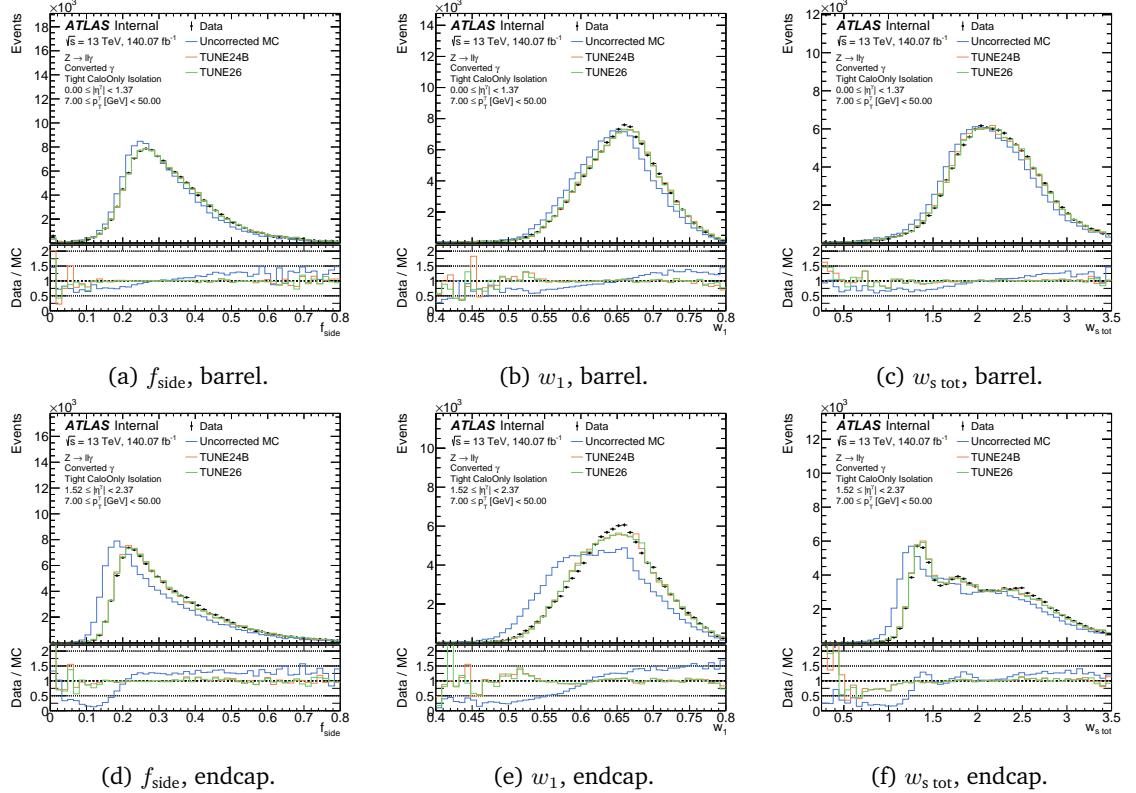


Figure 5.6: Selected SSs distributions using the RZ samples for converted photons after applying the FFs corrections event-by-event to the MC simulation. The SSs are shown in the barrel region (top) and in the endcap (bottom). Data points are represented by the black points and uncorrected (corrected)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

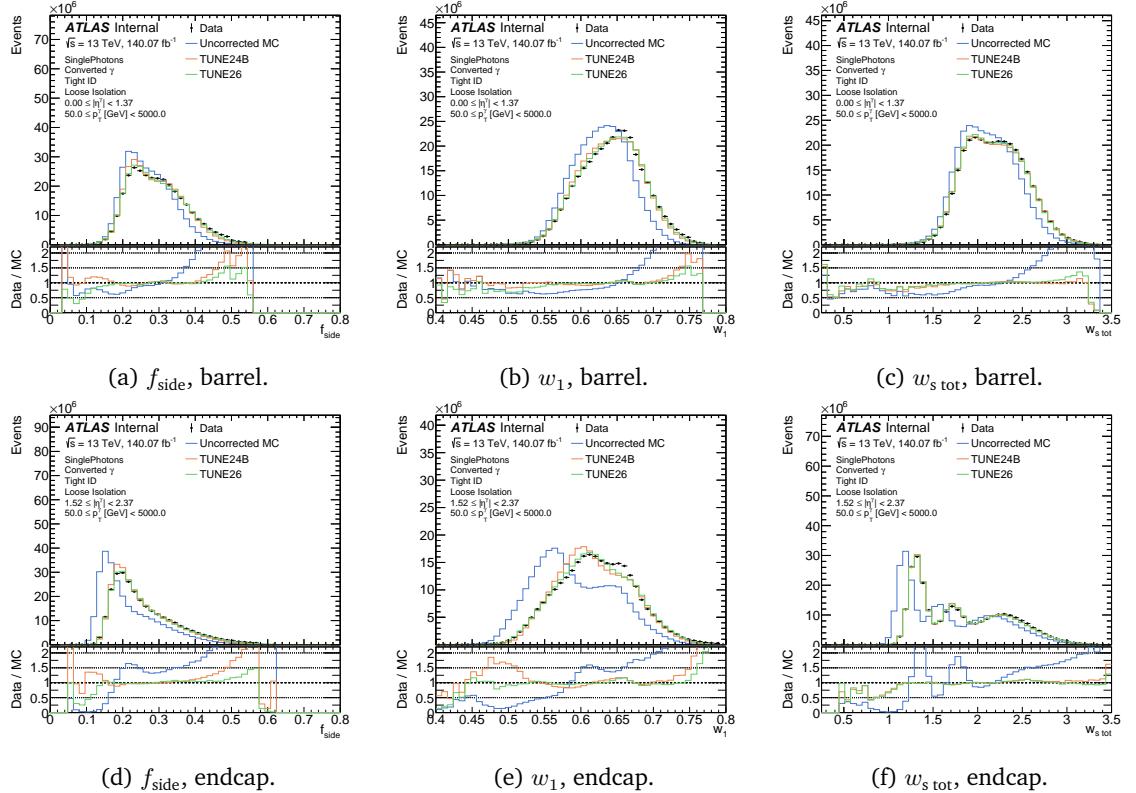


Figure 5.7: Same as Figure 5.6 but with the SP samples.

1887 The cell-based reweighting approach has been developed and tested initially for elec-  
 1888 trons [143], and further tested for photons [144]. For the electron's case, results have  
 1889 been very promising, where the second-layer **SSs** were substantially corrected. However,  
 1890 for photons, the same method that was used for electrons struggled, only working on  
 1891 average. Another approach to correct the simulation was based on matching data and  
 1892 simulated events, tested only on pseudo-data and technically complicated, but leading to  
 1893 better improvements [144].

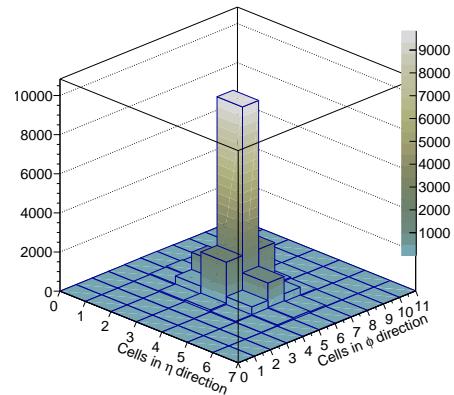
1894 In the current section, a new way of correcting the cell energies in **MC** is studied, using  
 1895 only the second layer of the **ECAL**, for simplicity. The method shares similarities with the  
 1896 **FF** method, making it easy to understand. First, the special event selection used for this  
 1897 study is presented. An overview of the old method to correct energies is briefly discussed,  
 1898 and then a study on how this method is improved is presented.

### 1899 5.2.1 Event selection

1900 The studies presented in this section are carried out with the same dataset as used for  
 1901 the **FF** calculation, described in Section 5.1.1. However, in this case, only the **RZ** samples  
 1902 are used. Events are selected as described in Section 4.2.1, using loose-isolated photons.  
 1903 Nevertheless, given that these studies rely on information on the second layer of the **ECAL**,  
 1904 special selection on the cells needs to be taken into account.

1905 When an electron or photon enters the calorimeter, its footprint in the second layer is a  
 1906 visible cluster of cells surrounding the most energetic and central one (also referred as  
 1907 *hottest cell*). Clusters of  $7 \times 11$  cells in  $\eta \times \phi$  are considered, shown in Figure 5.8a with  
 1908 the current cell arrangement used. Approximately, 90% of the energy of the cluster is  
 1909 shared amongst the 9 central cells, which are highlighted in blue in Figure 5.8a, and the  
 1910 average normalised energy for data is shown in Figure 5.8b, visualising how the energy is  
 1911 distributed.

| Cells in $\phi$ direction |    |    |    |    |    |    |
|---------------------------|----|----|----|----|----|----|
| 11                        | 22 | 33 | 44 | 55 | 66 | 77 |
| 10                        | 21 | 32 | 43 | 54 | 65 | 76 |
| 9                         | 20 | 31 | 42 | 53 | 64 | 75 |
| 8                         | 19 | 30 | 41 | 52 | 63 | 74 |
| 7                         | 18 | 29 | 40 | 51 | 62 | 73 |
| 6                         | 17 | 28 | 39 | 50 | 61 | 72 |
| 5                         | 16 | 27 | 38 | 49 | 60 | 71 |
| 4                         | 15 | 26 | 37 | 48 | 59 | 70 |
| 3                         | 14 | 25 | 36 | 47 | 58 | 69 |
| 2                         | 13 | 24 | 35 | 46 | 57 | 68 |
| 1                         | 12 | 23 | 34 | 45 | 56 | 67 |



(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  $7 \times 11$  clusters for data.

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

- 1912 In this work, only events in which the clusters have the total of 77 cells are considered.  
 1913 Also, the events are required to have the central cell to be the most energetic.

1914 **5.2.2 Calculation**

1915 **Early developments**

- 1916 All events that pass the selection will have a cluster associated to it, each of one having  $N$   
 1917 cells and each cell has an energy  $E_i$ , for  $i = 1, \dots, N$ . For each event, in the first place,  
 1918 cluster energies are obtained as

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

- 1919 which then are used to compute normalized cell energies  $e_i = E_i/E$ . Distributions of these  
 1920 normalized energies are then obtained when considering all the events that pass selection,  
 1921 and their means are used to compute the corrections at each cell  $i$ :

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

- 1922 where  $E^{\text{data/MC}}$  are the cluster energies for data and MC and, as mentioned above, the  
 1923 mean is done over all events that pass the selection. By definition these corrections coefficients sum to 0 over the whole cluster:

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

- 1925 Then a cell energy is corrected as:

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

- 1926 which is translated into shifting the cell energy divided by the cluster energy  $E^{\text{MC}}$  (from  
 1927 now on called normalized cell energy,  $e_i^{\text{MC}}$ ) by an amount of  $\Delta$ , so data and MC distributions' means match. For the early studies on photons, reweights were calculated separately  
 1928 for unconverted and converted photons, and they were also binned in  $|\eta|$ :

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

- 1930 Because the corrections coefficients sum to zero, this method also implies that the cluster  
 1931 energy remains constant throughout the reweighting procedure:

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

- 1932 The corrections can be visualized as 2-dimensional matrices, as seen in Figure 5.9, and  
 1933 therefore the reweighting procedure is developed using 2-dimensional histograms. For

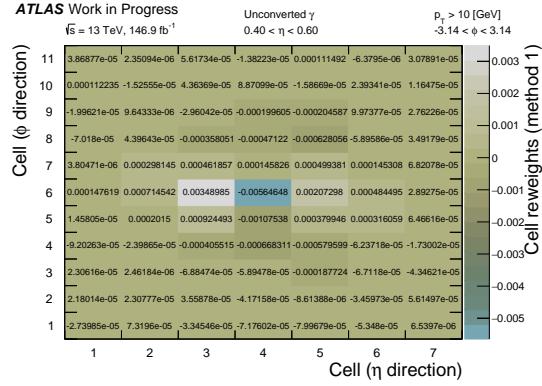


Figure 5.9: Energy reweights to be applied to MC cells, for unconverted photons with  $0.4 < |\eta| < 0.6$ . \*\*\* fix luminosity

1934 the case of working with the calorimeter layer 2, clusters of 77 cells ( $\eta \times \phi = 7 \times 11$ ) are  
 1935 used.

1936 The SSs computed using the second-layer information are  $R_\eta$ ,  $R_\phi$  and  $w_{\eta 2}$ , which can be  
 1937 calculated from energy deposits from the relations:

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

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

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

1938 where  $E_{i \times j}$  is the summed cell energy in a region  $\eta \times \phi = i \times j$  cells around the cent-  
 1939 ral cell. It was shown in the previous studies [144] that this method only corrects the  
 1940 shower shapes in average, but differences on the shape remain. This is due to the fact that  
 1941 this method only corrects the means of the energies distributions, by redistributing the  
 1942 energies amongst the cells. However, these energy distributions still present differences,  
 1943 specially regarding the shapes, leading to a very similar situation to what has been seen  
 1944 for the FFs. In this manner, a very similar approach of further correcting the means and  
 1945 widths of the normalised energies distributions can be employed.

#### 1946 New re-weighting method

1947 This new method aims to correct both the mean and variance of normalized cell energies  
 1948 distributions by applying shifts and stretches. Similar to the approach followed for SSs, a  
 1949 first approximation for shift and stretch values of the energy distributions is to compute  
 1950 means and root mean squares (RMS) of them at each cell, respectively. Then, reweighted  
 1951 normalized cell energies can be obtained as

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

1952 where the subindex  $e$  indicates that the RMS is calculated from normalized cell energy  
 1953 distributions and  $i$  runs over all cells in the cluster.

1954 Since the normalized cell energy at cell  $i$  can be calculated as  $e_i^j = E_i^j / E^j$ , for  $j = \text{MC-RW}$ ,  
 1955 MC and data, and it is required to have the same cluster energy after the reweighting  
 1956 procedure ( $E^{\text{MC-RW}} = E^{\text{MC}}$ ). Multiplying Eq. 5.12 by  $E^{\text{MC-RW}}$  it is possible to arrive at an  
 1957 expression for  $E_i^{\text{MC-RW}}$ :

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

1958 Comparing Eq. 5.13 to Eq. 5.11, it is observed that  $E_i^{\text{MC}}$  is now multiplied by a stretch  
 1959 factor, and a similar shift is applied to the normalized cell energy distribution but with  
 1960 modifications introduced by the stretch:

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

1961 Finally, to guarantee that the cluster energy remains constant, cell energies are re-scaled  
 1962 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,  
 1963 and one example of them is presented in Figure 5.10. For what follows, this new method  
 1964 is applied to correct the cell energies, and it is computed inclusively in  $p_T$  and  $|\eta|$ , only  
 1965 separating between unconverted and converted photons.

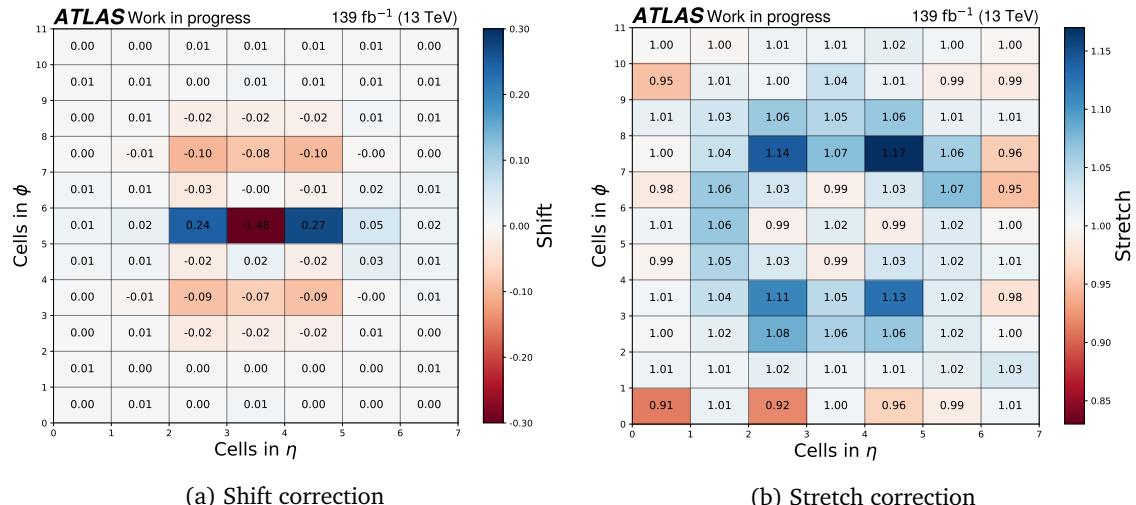


Figure 5.10: Example of two reweights matrices, using shift (left) and stretches (right) corrections. Shown results correspond to unconverted photons. The shift values are multiplied by a factor of 100 to improve visualisation.

### 1966 5.2.3 Results

1967 Figure 5.10 shows the shift and stretch reweighting matrices for unconverted photons. It  
 1968 can be noted that the major shift correction is done in the central cell, where the shift

1969 corresponds to a negative value. Negative shifts to the normalised cell energy means that  
 1970 energy must be removed from the central cell, and distributed to the neighbouring ones,  
 1971 as seen from the positive shifts values in the closest cells. Regarding stretch values, a  
 1972 rather symmetric distribution of values is observed with respect to the central cell.  
 1973 Using these correction factors to the normalised energies to each cell, in Figure 5.11,  
 1974 the resulting cell energy distributions are displayed for cells 28, 39 and 50<sup>2</sup>. The new  
 1975 reweighting method achieves greats improvements on the agreement between data and  
 1976 MC. The reweighting method does well correcting tails of the distributions for all cells, as  
 1977 well as the peaks of them, which can be seen specially for cell 28.

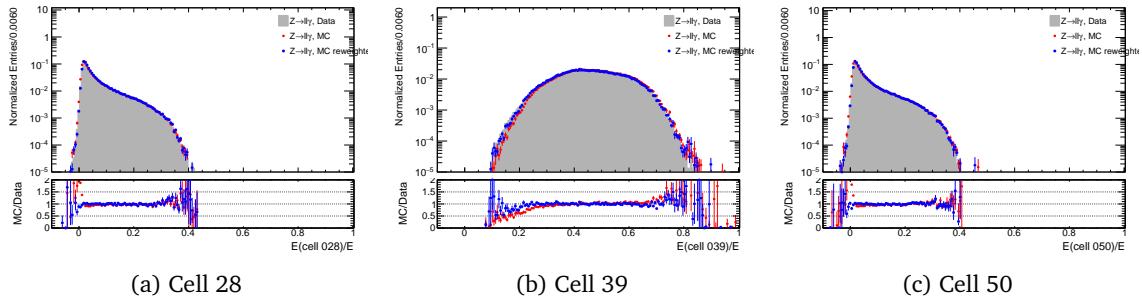


Figure 5.11: Normalized cell energy distributions for cells 28, 39 and 50 of the cluster using unconverted photons. Blue and red points correspond to the reweighted and original MC distributions, respectively, while the grey histogram represents data.

1978 To assess the behaviour of the reweighting procedure applied to the SSs of the second  
 1979 ECAL layer, in Figure 5.12 the comparison of the methods for correcting the  $R_\eta$ ,  $R_\phi$   
 1980 and  $w_{\eta 2}$  SSs are shown. In the three cases, an improvement over the uncorrected MC  
 1981 is observed, specially for the  $R_\phi$  and  $w_{\eta 2}$  variables, but the cell-based energy reweighting  
 1982 method for unconverted photons does not reach the level of agreement with data as the  
 1983 FFs, which has been shown to provide an excellent match with data. Nevertheless, there  
 1984 is almost no difference seen between the reweighting and the FF method for converted  
 1985 photons, indicating that there is room for improvement on the corrections.

### 1986 5.3 Conclusions and future work

1987 In the current chapter two methods to correct the observed disagreement between data  
 1988 and simulated Shower Shapes (SSs) were studied.  
 1989 The Fudge Factor (FF) method has been historically used in the collaboration, which in  
 1990 the beginnings it was only based on simple shifts of the distributions by adding a constant  
 1991 term to the variable. Even though the corrections led to good improvements and therefore  
 1992 obtaining Scale Factors (SFs) closer to one, notable shape differences remained between  
 1993 data and simulation. In the context of this work, by adding a linear term to the variable  
 1994 transformation, the widths of the MC distributions are fixed leading to even better agree-

<sup>2</sup> 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  $\eta$  direction.

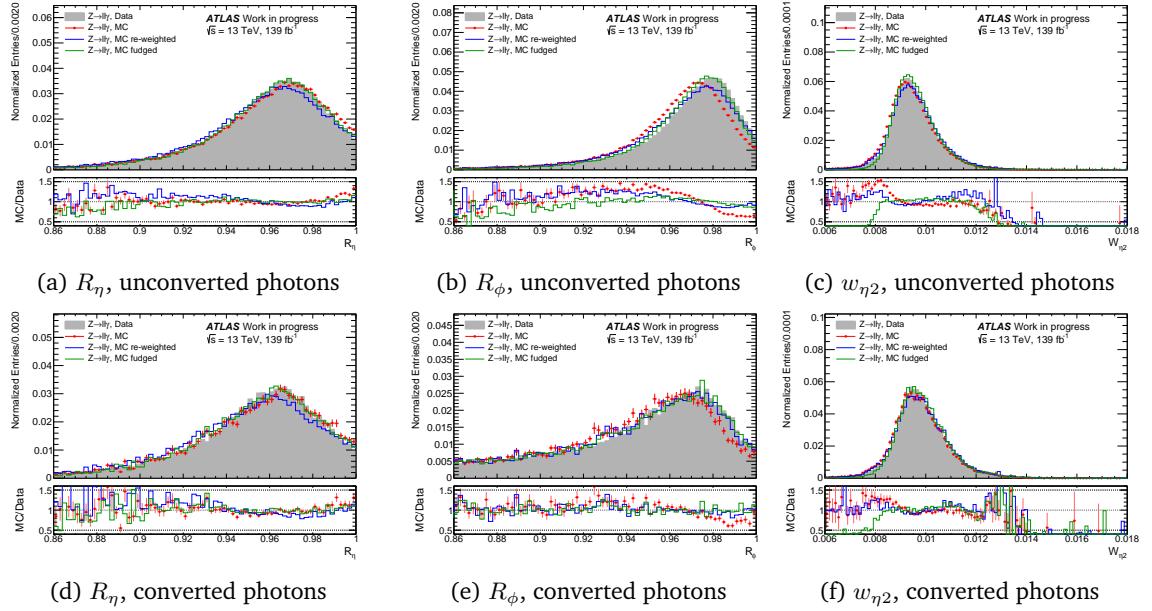


Figure 5.12: SSs distributions for unconverted (top row) and converted (bottom row) photons in  $|\eta| < 0.6$ . Data is represented by the grey histogram. Uncorrected MC simulation is shown with the red points and reweighted (fudged) MC with the blue (green) line.

1995    ment. This new method of correcting the SSs using FFs is referred as the shift+stretch  
 1996    method and is now used throughout the ATLAS collaboration.

1997    A novel and lower-level method that aims to modify the energies in the ECAL cells has also  
 1998    been developed. By using the energy distributions at each cell in clusters around the most  
 1999    energetic cell, it is possible to correct all the SSs at once. This method applies the same  
 2000    strategy of shifting and stretching the normalised energy from MC simulation to match the  
 2001    distribution found in data. Although the method is new and still needs polishing and its  
 2002    extension to other layers of the ECAL, it led to promising results in which some variables  
 2003    are corrected in the same way as with FFs. The cell-based energy reweighting method  
 2004    shows a great potential in the collaboration, not only in the context of *offline* photon  
 2005    identification, but also at trigger level.

### 2006    5.3.1 Future work

2007    One of the most exciting and promising approaches to correct the SSs is the cell-based  
 2008    method. As previously mentioned, this approach could be employed on different steps on  
 2009    the photon identification process, such as at trigger level, or event *offline* to correct all  
 2010    the SSs simultaneously. Another potential, and important usage, is to use the corrected  
 2011    clusters to directly compute photon identification, for example as considering the clusters  
 2012    as images and using a Convolutional Neural Network (CNN) to perform the photon iden-  
 2013    tification [144].

2014    Shower Shape Variables (SSVs) can be easily interpreted in terms of physical terms. For  
 2015    this reason, maintaining these physical quantities serve to understand the underlying phys-  
 2016    ics of the processes. To continue correcting these variables is of great interest and there

2017 are several ways in which these can be achieved. The current method of transforming the  
2018 variable but using higher-order terms remains a challenging task, but yet not explored.  
2019 By making use of the novel Machine Learning (ML) techniques, it is possible to obtain  
2020 correction factors for higher-order terms in the expansion, further correcting higher-order  
2021 momenta of the distributions (skewness, kurtosis, etc.). Other interesting approach is us-  
2022 ing multivariate reweighting, which was explored in Ref. [145], showing very promising  
2023 results.

2024

## Part IV

2025

# Search for new physics in high-mass photon+jet final states

2026

2027

2028 **ANALYSIS STRATEGY AND  
STATISTICAL TREATMENT**



2029

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

---

Billie Jean King

2030 The main work of this thesis is the search for exotic particles that decay in photon+jet final  
 2031 state. In this chapter, an introduction to the general strategy of the analysis is given, enu-  
 2032 merating the backgrounds and what signals are used. Additionally, the statistical model  
 2033 used in this work is built, starting from a simple likelihood function. In the next chapters,  
 2034 more in-detail aspects of the analysis will be described.

2035 **6.1 Strategy, signals and SM’s backgrounds**

2036 The analysis done for this thesis consists on exotic searches with an isolated, highly ener-  
 2037 getic photon, in association with jets. The strategy is based on finding bumps on the final  
 2038 state invariant mass over the **SM** predictions. Prompt photons in association with jets are  
 2039 copiously produced at the **LHC**. The production modes were described in detail in Sec-  
 2040 tion 1.1.4 with the main one being quark-gluon Compton scattering  $qg \rightarrow q\gamma$ . By looking  
 2041 at the invariant mass  $m_{\gamma+j}$  of the  $\gamma$ +jet system, one finds that it follows a smoothly decay-  
 2042 ing shape, leading to an ideal scenario for searches of physics **BSM**. The exotic particles  
 2043 that are searched for decay resonantly in a photon and a jet, hence leading to a bump on  
 2044 the  $m_{\gamma+j}$  distribution over the **SM** prediction.

2045 Different benchmark models are considered in this search. The first model is based on gen-  
 2046 eric Gaussian-shaped mass distributions, with different values of its mean and standard  
 2047 deviation. This model provides a generic interpretation for the presence of a signal with  
 2048 different Gaussian widths, ranging from resonances with a width similar to the reconstruc-  
 2049 ted  $m_{\gamma+j}$  resolution of  $\sim 2\%$  to wide resonances with a width up to 15%. Secondly, the  
 2050 Excited Quarks (**EQs**) and Quantum Black Holes (**QBHs**) theoretical models introduced  
 2051 in Section 1.2 are studied. These models depend on one or more parameters, like for  
 2052 example, the **EQ** particle mass and its coupling to the **SM**. That being the case, a multi-

2053 dimensional grid of parameters is built, where each point on the grid represents a set of  
 2054 unique parameter values that define the signal model. In the case no significant excess is  
 2055 found over the smooth **SM** background, exclusion limits to the values on the grid are set,  
 2056 excluding a subset of the possible parameters of the theory.

2057 The major challenge in any search for new physics is the excellent understanding of the  
 2058 **SM** background. For this search, the dominant background is prompt photon production.  
 2059 This process, as discussed in Section 1.1.4, can be understood as two separate processes,  
 2060 direct and fragmentation, in which the latter can be highly reduced. The second largest  
 2061 background arises from **QCD** multijet events, where one of the jets is rich in **EM** energy  
 2062 and is mis-identified as a photon. This background also includes events in which a jet  
 2063 fragments into an energetic  $\pi^0$  or  $\eta$  meson, decaying into two overlapping photons that are  
 2064 reconstructed as a single photon in the detector. Finally, the third background, although  
 2065 not considered as its contribution is almost negligible, arise from **EW** processes via the  
 2066 production of  $W/Z + \text{jet}$ , with the heavy boson decaying leptonically and an electron  
 2067 mimics a photon because missing hits in the **ID**.

2068 In resonance searches it is common to model the background using a functional form. This  
 2069 function usually is a particular case of a family of functions, and the selection of it is based  
 2070 on several statistical tests, being this one of the most crucial steps of the analysis. Although  
 2071 the background is estimated directly from data, **MC** simulations of the background are  
 2072 necessary in order to perform optimisations of the event selection and to carry out the  
 2073 different tests to select the function.

## 2074 6.2 Statistical Treatment

### 2075 6.2.1 Statistical Model

2076 Fits to of the  $m_{\gamma+j}$  spectra are performed separately for each analysis signal region using  
 2077 a log-likelihood minimization method. The definition of the likelihood function is given  
 2078 in the following step-by-step. The foundation is a counting experiment in each bin using  
 2079 a set of Poissonian probability terms:

$$\mathcal{L} = \prod_{i=1}^{N_{\text{bins}}} \text{Pois}(N_i | \mu s_i + b_i) = \prod_{i=1}^{N_{\text{bins}}} \frac{(\mu s_i + b_i)^{N_i}}{N_i!} e^{-(\mu s_i + b_i)} \quad (6.1)$$

2080 where  $N_i$  is the number of measured data events in bin  $i$  and  $s_i$  and  $b_i$  are expected signal  
 2081 and background counts, estimated from **MC** simulation and the functional fit, respectively.  
 2082 The parameter  $\mu$  is referred as the signal strength, and is a common scale factor to the  
 2083 signal yield in all bins. One of the goals of the fit is the determination of this parameter,  
 2084 also known as the Parameter of Interest (**POI**).

2085 In the case of a parametrized background,  $b_i$  could be re-written as  $b_i = N_b f_b(x, \vec{\theta})$ ,  
 2086 where  $x$  represents the observable ( $m_{\gamma+j}$  in the case of this analysis),  $f_b$  is the back-  
 2087 ground parametrization function, which depends on nuisance parameters  $\vec{\theta}$  that describe

2088 the shape of the function (known as nuisance parameters), and  $N_b$  is the total background  
 2089 normalization, another nuisance parameter. The likelihood function then reads

$$\mathcal{L} = \prod_{i=1}^{N_{\text{bins}}} \frac{\left(\mu s_i + N_b f_b(x, \vec{\theta})\right)^{N_i}}{N_i!} e^{-(\mu s_i + N_b f_b(x, \vec{\theta}))}. \quad (6.2)$$

2090 The number of signal events at bin  $i$  is given by the product of the signal PDF  $f_s^i(x)$ , the  
 2091 cross-section times branching ratio  $\sigma_s \times \text{Br}$ , the total integrated luminosity  $L$ , and the  
 2092 signal acceptance times efficiency  $A \times \varepsilon$ . Then, the signal term can be re-written as

$$\mu s_i = (\sigma_S \times \text{Br}) \times L \times (A \times \varepsilon) \times f_s^i(x). \quad (6.3)$$

2093 The special case when  $\mu = 0$  corresponds to a background-only hypothesis.

### 2094 6.2.2 Systematic uncertainties

2095 Systematic uncertainties are parameterized as a set of Nuisance Parameter (NP)  $\vec{\theta}$  that  
 2096 modify the expected signal and background yields, i.e.  $\{s_i, b_i\} \rightarrow \{s_i(\vec{\theta}), b_i(\vec{\theta})\}$ . They are  
 2097 implemented in the likelihood by multiplying it by

$$\prod_k G_k(0|\theta_k, 1) \quad (6.4)$$

2098 where  $G_k$  are the NPs constraint PDFs, and  $k$  ranges over all the systematic variations.

2099 Moreover, the relevant parameter is multiplied by a factor of

$$1 + \theta_k \delta_k \quad (6.5)$$

2100 for the case of Gaussian response functions, and

$$\exp(\theta_k \ln(1 + \delta_k)) = (1 + \delta_k)^{\theta_k} \quad (6.6)$$

2101 for log-normal response functions.

2102 Given that the background function is selected in an arbitrary way, but by undergoing  
 2103 different statistical tests, a special systematic uncertainty is added to take this into account,  
 2104 called Spurious Signal (SS). The SS is computed by means of the SS test discussed in  
 2105 Section 10.2.4, and is implemented into the statistical model by adding another term to the  
 2106 total number of events:

$$\mu s_i + N_b f_b \rightarrow \mu s_i + \sigma_{\text{spur},i} \theta_{\text{spur}} + N_b f_b. \quad (6.7)$$

2107  $\sigma_{\text{spur}}$  represents the number of spurious events in bin  $i$  of the distribution following the  
 2108 same PDF as the signal, and  $\theta_{\text{spur}}$  is the NP associated to this other source of signal.

2109 **6.2.3 Simultaneous fits**

2110 It is useful to discuss as well the procedure for doing simultaneous fits in different  $m_{\gamma+j}$   
 2111 shapes simultaneously. These simultaneous fits test the same signal model with same  
 2112 signal strength  $\mu$  in different regions. The total likelihood is simply the product of the  
 2113 individual likelihoods given by Eq. 6.2, and it reads:

$$\mathcal{L}_{\text{total}} = \prod_{c \in \text{categories}} \mathcal{L}_c \quad (6.8)$$

$$= \prod_{c \in \text{categories}} \left[ \prod_{i=1}^{N_{\text{bins}}^c} \frac{(n_{\text{sig}} + b_{i,c})^{N_{i,c}}}{N_{i,c}!} e^{-(n_{\text{sig}} + b_{i,c})} \right] \quad (6.9)$$

2114 The implementation of the systematic uncertainties follows the same recipe as explained  
 2115 above. The experimental systematics affecting only the signal are 100% correlated amongst  
 2116 the signal regions. On the other hand, each signal region has its own spurious signal un-  
 2117 certainty, as it is associated to a different background function and fit-range<sup>1</sup>.

2118 **6.2.4 Hypothesis test**

2119 The objective of a search such as the one presented in this thesis is to be able to specify  
 2120 how good the agreement is between the observed data and a given hypothesis, which  
 2121 typically is the “null”, Background-only ([B-only](#)), or “0-signal” hypothesis ( $H_0$ ). One could  
 2122 test the consistency of the data with any hypothesis, but  $H_0$  is usually chosen because  
 2123 typically a discovery can be claimed by establishing that the data is inconsistent with  
 2124 the “standard” theory, without having to show that it is consistent with some alternative  
 2125 theory. Once inconsistency with  $H_0$  is established, several alternative signal hypotheses  
 2126 can be tested to characterize the discovery, denoted  $H_1$ . To distinguish between these  
 2127 theories the parameter  $\mu$  is used, which is 0 for the  $H_0$  and 1 for the nominal signal.

2128 When comparing the data against the hypotheses, their differences are quantified by a  
 2129 single number, called a “test statistic”, and are functions that depend on the data. There-  
 2130 fore, for each test statistic there is a [PDF](#) associated to it.

2131 To test a hypothesized value of  $\mu$ , consider the *profile likelihood*

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\theta})}{\mathcal{L}(\hat{\mu}, \hat{\theta})}. \quad (6.10)$$

2132 The numerator of this ratio is the *profile likelihood function*. The quantity  $\hat{\theta}$  denotes the  
 2133 value of  $\theta$  that maximises  $\mathcal{L}$  for the specified  $\mu$ . The denominator is the maximised likeli-  
 2134 hood function, that is,  $\hat{\mu}$  and  $\hat{\theta}$  are their maximum-likelihood estimators. From Eq. 6.10,

---

<sup>1</sup> As it will be discussed in Chapter 10, there are different signal regions in the analysis and for each one a different functional form is selected.

2135 it is possible to see that  $0 \leq \lambda \leq 1$ , with  $\lambda$  near 1 implying good agreement between the  
 2136 data and the hypothesized value of  $\mu$ . It is convenient to define the test statistic  $q_\mu$  as

$$q_\mu = -2 \ln \lambda(\mu), \quad (6.11)$$

2137 such that good agreement is seen as  $q_\mu \rightarrow 0$ , while high  $q_\mu$  values indicate disagreement  
 2138 between data and the hypothesized value of  $\mu$ .

2139 To measure the discrepancy between data and the hypothesis  $H$ , that is, the probability  
 2140 that a given test statistic is as big as its observed value assuming  $H$  is correct, is defined  
 2141 by the *p*-value  $p_\mu$ :

$$p_\mu = P(q_\mu > q_{\mu,\text{obs}}) = \int_{q_{\mu,\text{obs}}}^{\infty} f(q_\mu | H) dq_\mu, \quad (6.12)$$

2142 where  $q_{\mu,\text{obs}}$  is the value of the test statistic observed from data, and  $f(q_\mu | H)$  is the  $q_\mu$  PDF  
 2143 under the hypothesis  $H$ . A small *p*-value means that the hypothesis is not in agreement  
 2144 with the data, and therefore  $H$  is excluded when the *p*-value is lower than a defined  
 2145 value  $\alpha$ . It is common to convert the *p*-value to a significance  $Z$ , defined such that a  
 2146 Gaussian distributed variable from  $Z$  standard deviations above its mean has an upper-  
 2147 tail probability equal to  $p_\mu$ :

$$Z = \Phi^{-1}(1 - p_\mu). \quad (6.13)$$

2148 In the context of search for new physics, the data is tested against the  $H_0$  hypothesis, since  
 2149 a rejection of  $H_0$  may mean the discovery of a new signal. For this, the test statistic takes  
 2150 the form:

$$q_0 = \begin{cases} 0 & \text{if } \hat{\mu} < 0, \\ -2 \ln \lambda(0) & \text{if } \hat{\mu} \geq 0. \end{cases} \quad (6.14)$$

2151 If the observed data turn out to be smaller than the background predictions, one has  $\hat{\mu} < 0$ .  
 2152 This could be evidence against the **B-only** hypothesis, but it does not actually show that  
 2153 the data are composed of signal events. With this definition then, the possibility of ruling  
 2154 out the **B-only** hypothesis occurs only when  $\hat{\mu} \geq 0$ , and otherwise  $q_0 = 0$ . The *p*-value for  
 2155 this test statistic is then:

$$p_0 = P(q_0 > q_{0,\text{obs}}) = \int_{q_{0,\text{obs}}}^{\infty} f(q_0 | 0) dq_0. \quad (6.15)$$

2156 The particle physics community defines a rejection of the **B-only** hypothesis with a signifi-  
 2157 ance greater than  $5\sigma$  ( $p_0 = 2.86 \times 10^{-7}$ ) as the appropriate level to constitute a discovery.  
 2158 It should be emphasized that in an actual scientific context, rejecting the **B-only** hypothesis  
 2159 in a statistical sense is only part of discovering a new phenomenon. One's degree of belief  
 2160 that a new process is present will depend in general on other factors as well, such as the  
 2161 plausibility of the new signal hypothesis and the degree to which it can describe the data.  
 2162 When the *p*-value obtained is greater than the limit defined for a discovery, it is not possible  
 2163 to reject the **B-only** hypothesis, and in that case it is desired to establish limits on the tested  
 2164 model. To do so, one seeks instead to reject the Signal+Background (**S+B**) hypothesis,

and to find the upper value of  $\mu$  for which such rejection is not possible (upper limit). A new test statistic is then defined, where the roles of the **B-only** and **S+B** hypotheses are swapped:

$$\tilde{q}_\mu = \begin{cases} -2 \ln \tilde{\lambda}(\mu) & \text{if } \hat{\mu} \leq \mu \\ 0 & \text{if } \hat{\mu} > \mu \end{cases} = \begin{cases} -2 \ln \left( \frac{\mathcal{L}(\mu, \hat{\theta}(\mu))}{\mathcal{L}(0, \hat{\theta}(0))} \right) & \text{if } \hat{\mu} \leq 0 \\ -2 \ln \left( \frac{\mathcal{L}(\mu, \hat{\theta}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \right) & \text{if } 0 \leq \hat{\mu} \leq \mu \\ 0 & \text{if } \hat{\mu} \geq \mu \end{cases} \quad (6.16)$$

he reason for setting  $q_\mu = 0$  for  $\hat{\mu} > \mu$  is that when setting an upper limit, one would not regard data with  $\hat{\mu} > \mu$  as representing less compatibility with  $\mu$  than the data obtained, and therefore this is not taken as part of the rejection region of the test. That is, the upper limit is obtained by testing  $\mu$  against the alternative hypothesis consisting of lower values of  $\mu$ . From the definition of the test statistic one sees that higher values of  $q_\mu$  represent greater incompatibility between the data and the hypothesized value of  $\mu$ . One should note that  $q_0$  is not simply a special case of  $q_\mu$  with  $\mu = 0$ , but rather has a different definition. That is,  $q_0$  is zero if the data fluctuate downward ( $\hat{\mu} < 0$ ), but  $q_\mu$  is zero if the data fluctuate upward ( $\hat{\mu} > \mu$ ).

\*\*\* revise and discuss with Tere. With this test statistic it is aimed to find the highest  $\mu$  at which the **S+B** hypothesis is not compatible with data anymore. To accomplish this, the following Confidence Level (**CL**) is defined:

$$\text{CL}_s = \frac{p_\mu}{1 - p_b} \equiv \frac{\text{CL}_{s+b}}{\text{CL}_b}, \quad (6.17)$$

where

$$p_\mu = \int_{q_{\mu,\text{obs}}}^{\infty} f(q_\mu | \mu) dq_\mu \equiv \text{CL}_{s+b} \quad (6.18)$$

$$1 - p_b = \int_{q_{\mu,\text{obs}}}^{\infty} f(q_\mu | 0) dq_\mu \equiv \text{CL}_b \quad (6.19)$$

being  $f(q_\mu | \mu)$  the PDF of the test statistic  $q_\mu$  and  $f(q_\mu | 0)$  the PDF under the **B-only** hypothesis. For smaller values of  $\text{CL}_s$ , there is lower compatibility between the data and the **S+B** hypothesis. The upper limit on  $\mu$ ,  $\mu_{\text{up}}$ , is defined as the  $\mu$  when  $\text{CL}_s = 0.05$ , and the models with  $\mu < \mu_{\text{up}}$  are rejected at 95% **CL**.

## 6.2.5 The BumpHunter algorithm

The BumpHunter algorithm [146, 147] constitutes a hypertest that combines the result of many individual hypothesis tests into a single test statistic. It iterates over all possible windows of adjacent bins in the spectrum, starting with windows one bin wide and increasing in width up to a configurable upper threshold for the window size (typically half of the fit

range). In each of these windows, the observed number of events  $N_d$  and the estimated background events  $N_b$  are summed as if the window were a single bin:

$$N_d = \sum_{\substack{\text{bins } i \\ \text{in window}}} N_d^i \quad (6.20)$$

$$N_b = \sum_{\substack{\text{bins } i \\ \text{in window}}} N_b^i \quad (6.21)$$

Each of these windows is assigned a local  $p$ -value based on the Poissonian probability to observe at least as many events as seen in data:

$$p_{\text{local}}(N_d, N_b) = \begin{cases} \Gamma(N_d, N_b) & \text{if } N_d > N_b, \\ 1 & \text{if } N_d \leq N_b, \end{cases} \quad (6.22)$$

where:

$$\Gamma(N_d, N_b) = \sum_{k=N}^{\infty} \frac{N_b^k}{N_d!} e^{-N_b} \quad (6.23)$$

is the lower incomplete gamma function.

Since many statistically independent windows are tested for excesses, the look-elsewhere effect [148] must be accounted for. It describes that if e.g. 100 independent tests are performed, on average one of them will find a  $p$ -value below 0.01 due to statistical fluctuations alone. The local  $p$ -value for an excess in a specific window must hence be translated into a global  $p$ -value to find an excess in any of the windows. For that purpose, the BumpHunter algorithm defines the test statistic:

$$t = \min_{\text{windows}} (-\log p_{\text{local}}) \quad (6.24)$$

which identifies the most significant excess of all considered windows. The expected distribution of  $t$  is determined numerically to high accuracy by drawing 2000 Poissonian fluctuated toy distributions from the background expectation and computing  $t$  for each of these toys. The global BumpHunter  $p$ -value is then given by the probability for a toy distribution to exhibit a more significant excess than the data. This corresponds to the fraction of toys for which  $t$  exceeds the observed value  $t_{\text{obs}}$ :

$$p(\text{BH}) = \frac{\# \text{ toys with } t > t_{\text{obs}}}{\# \text{ toys}}. \quad (6.25)$$

One should note that, with this definition, the global  $p$ -value can be higher than 0.5, producing a negative global significance. In this case, this negative value should not be interpreted as a deficit in the data, but as a deviation which is less significant than the median deviation observed in the [B-only](#) pseudo-data distributions.

This way,  $p(\text{BH})$  by definition follows a uniform distribution between 0 and 1 if excesses over the background estimate are due to statistical fluctuations alone. Observing, for example,  $p(\text{BH}) < 0.05$  then means that less than 5% of background-like toys exhibit a larger excess than the most significant one in data. Such an observation can be interpreted as evidence for a resonance causing a deviation of the data from the background expectation.

## 2217 6.3 Fits and results

2218 Starting from Eq. 6.2, different kind of fits can be performed to the background distribu-  
2219 tion, depending on the used regions and if signal samples are included or not.

2220 In order to select the background function, both **B-only** and **S+B** fits are performed. **B-only**  
2221 fits are characterised by not having any signal included, therefore the signal strength  
2222 parameter  $\mu$  is kept at zero, and the background-shape **NPs** are computed. In **S+B** fits,  
2223  $\mu$  is allowed to float and the final signal yield is used for the studies. In Section 10.2,  
2224 these two kind of fits are used thoroughly using **MC** simulation in order to determine the  
2225 different fitting strategies to be used with the actual data.

2226 To perform the actual search, as a first step a **B-only** fit is carried out and different metrics  
2227 are evaluated to assess if a significant excess over the **SM** is observed. In case there is  
2228 none, **S+B** fits are done for each signal model, in order to set exclusion limits on the  
2229 signal grid.

2230

2231 **SIGNAL MODELS AND  
SIMULATED SAMPLES**



2232

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

---

Billie Jean King

2233 Simulation of the different physical processes and the response of the detector to them is  
 2234 necessary to optimise and estimate the performance of the different analyses. In addition,  
 2235 it allows the strategies used in particle identification to be developed prior to the data-  
 2236 taking and the efficiencies of the algorithms to be tested. The preparation of searches  
 2237 for new physics requires a detailed simulation of the detector to estimate its discovery  
 2238 potential and to develop optimal methods for measuring particle properties. A correct un-  
 2239 derstanding of the signal and background processes is essential to distinguish between the  
 2240 two. Once real collision data is available, simulated data is also needed to find deviations  
 2241 from the **SM**. All the steps in the **MC** simulation was described in detail in Section 1.3.  
 2242 Similarly to any physics analysis at the **LHC**, this analysis makes use of simulated samples,  
 2243 both to understand the possible signals to be discovered, and to do a correct modeling of  
 2244 the background with  $\gamma + \text{jet}$  final states. In this chapter, details on the signal and back-  
 2245 ground samples generation and simulation are given.

2246 **7.1 Signals**

2247 **7.1.1 Excited quarks**

2248 One of the two benchmark theories to be tested correspond to **EQ**, which were explained  
 2249 in detail in Section 1.2.1. If quarks are composed of more fundamental constituents bound  
 2250 together by some unknown interaction, new effects should appear depending on the value  
 2251 of the compositeness scale  $\Lambda$ . It was also seen that **EQs** couple to the **SM** bosons, whose  
 2252 strengths are determined by the coupling constants  $f$ ,  $f'$  and  $f_s$ .  
 2253 In this theory, there are a total of 5 parameters: the compositeness scale  $\Lambda$ , the three  
 2254 couplings and the **EQ** mass. In order to reduce the number of parameters, it is usual to set

2255  $\Lambda = m_{q^*}$  [37], and to take the three coupling constants to be the same. In this way, only  
 2256 the mass and the coupling are the free parameters of the theory.

2257 Samples of **EQ** events are produced using PYTHIA 8.245 [67] with the **LO** NNPDF 2.3 [149]  
 2258 **PDF** set and the A14 set of tuned parameters for the underlying event. For the first time at  
 2259 the **LHC** in  $\gamma + \text{jet}$  final states, three different flavors are considered: light **EQs**  $q^*$  ( $u^*/d^*$ ),  
 2260 and heavy resonances, separating between  $c^*$  and  $b^*$ . Moreover, different couplings are  
 2261 studied in this work, where the following values are used:  $f = 0.01, 0.1, 0.5, 0.75, 1.0$ .  
 2262 The **EQ** masses and couplings used are listed in Table 7.1, where the processes cross-  
 2263 section times branching ratio are shown. Moreover, all the cross-section values are shown  
 2264 in Figure 7.1.

Table 7.1: Cross-section times branching ratio [fb] of the **EQ** model, for the three considered flavours using different coupling values.

| Excited quark                  | Mass [GeV] | Couplings $f = f' = f_s$ |           |            |        |             |
|--------------------------------|------------|--------------------------|-----------|------------|--------|-------------|
|                                |            | 0.01                     | 0.10      | 0.50       | 0.75   | 1.00        |
| $q^* \rightarrow \gamma + u/d$ | 500        | 29.9400                  | 3007.0000 | 75240.0000 | -      | 304200.0000 |
|                                | 1000       | 1.6560                   | 165.0000  | 4153.0000  | -      | 16490.0000  |
|                                | 2000       | 0.0536                   | 5.3800    | 133.2000   | -      | 520.8000    |
|                                | 3000       | -                        | 0.4435    | 11.0200    | -      | 43.0600     |
|                                | 4000       | -                        | 0.0488    | 1.2270     | -      | 4.8240      |
|                                | 5000       | -                        | -         | 0.1450     | -      | 0.5877      |
|                                | 5500       | -                        | -         | -          | -      | 0.2064      |
|                                | 6000       | -                        | -         | 0.0163     | 0.0384 | 0.0719      |
|                                | 6500       | -                        | -         | -          | -      | 0.0250      |
|                                | 7000       | -                        | -         | -          | 0.0043 | 0.0088      |
| $c^* \rightarrow \gamma + c$   | 500        | 3.6540                   | 362.2000  | 9051.0000  | -      | 36290.0000  |
|                                | 1000       | 0.1333                   | 13.3400   | 332.4000   | -      | 1297.0000   |
|                                | 2000       | -                        | 0.2434    | 6.0190     | -      | 23.6800     |
|                                | 3000       | -                        | -         | 0.3135     | -      | 1.2450      |
|                                | 4000       | -                        | -         | -          | -      | 0.0906      |
| $b^* \rightarrow \gamma + b$   | 500        | 0.6381                   | 63.7700   | 1588.0000  | -      | 6324.0000   |
|                                | 1000       | 0.0220                   | 2.2080    | 54.7600    | -      | 215.4000    |
|                                | 2000       | -                        | 0.0372    | 0.9249     | -      | 3.6200      |
|                                | 3000       | -                        | -         | 0.0446     | -      | 0.1770      |
|                                | 4000       | -                        | -         | -          | -      | 0.0121      |

### 2265 7.1.2 Quantum Black Holes

2266 **QBH** are predicted to be produced at the **LHC** in  $pp$  collisions, providing a solution to  
 2267 the hierarchy problem through the existence of extra dimensions. The production cross  
 2268 section is determined by the gravitational radius which depends on the Planck scale and  
 2269 number of dimensions, formulated as a classical cross section.

2270 Samples of QBH decaying into a photon and a parton are generated with the **QBH** 3.01  
 2271 software described in Ref. [150], and PYTHIA 8.3 for hadronisation and underlying event.  
 2272 The CTEQ6L1 **PDF** set has been used together with the standard A14 tuning of the under-

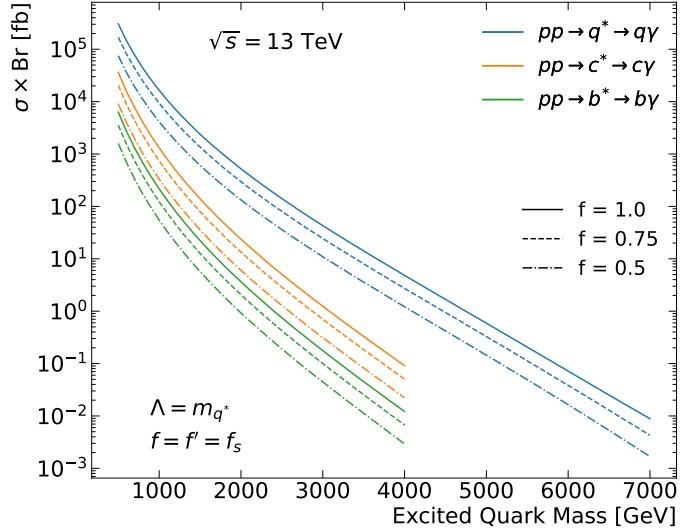


Figure 7.1: Cross-section times branching ratio of the different EQ production modes as a function of the EQ mass at  $\sqrt{s} = 13$  TeV. The figure shows the comparison between  $q^*$  (blue),  $c^*$  (orange) and  $b^*$  (green) signals using three different couplings.

lying events. The black hole production mass threshold  $m_{\text{th}}$  is set to be equal to the Planck scale  $m_p$  and the maximum black hole production mass to  $3m_p$  or the center-of-mass energy  $\sqrt{s}$ , whichever is less. Two different models are taken into account, depending on the number of extra dimensions  $n$ . The ADD model, as discussed in Section 1.2.2, considers 6 extra dimensions, leading to a total number 10 space-time dimensions. On the other hand, the RS1 model proposes only one warped extra dimension. Since only the  $\gamma + \text{jet}$  final state are of interest for this analysis, the 6 non-thermal black hole states shown in Section 7.1.2 are considered. The cross-sections times branching ratio of the two models are plotted in Figure 7.2, and the exact numbers shown in Table 7.2.

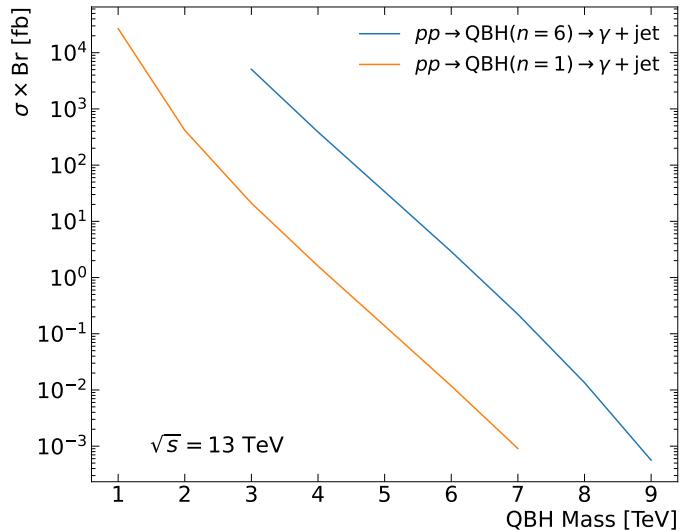


Figure 7.2: Cross-section times branching ratio of the RS1 (orange) and ADD (blue) QBH models, for a center-of-mass energy of  $\sqrt{s} = 13$  TeV.

Table 7.2: Sum of cross section times branching ratio [fb] for the  $\gamma$ +jet final state for the six non-thermal black holes states for each threshold mass  $m_{\text{th}}$ .

| $m_{\text{th}}$ [GeV] | ADD ( $n = 6$ )       | RS1 ( $n = 1$ )       |
|-----------------------|-----------------------|-----------------------|
| 1000                  |                       | $2.69 \times 10^{+4}$ |
| 2000                  |                       | $4.17 \times 10^{+2}$ |
| 3000                  | $5.07 \times 10^{+3}$ | $2.11 \times 10^{+1}$ |
| 4000                  | $3.88 \times 10^{+2}$ | $1.60 \times 10^{+0}$ |
| 5000                  | $3.37 \times 10^{+1}$ | $1.38 \times 10^{-1}$ |
| 6000                  | $2.90 \times 10^{+0}$ | $1.18 \times 10^{-2}$ |
| 7000                  | $2.22 \times 10^{-1}$ | $9.08 \times 10^{-4}$ |
| 8000                  | $1.35 \times 10^{-2}$ |                       |
| 9000                  | $5.64 \times 10^{-4}$ |                       |

## 2282 7.2 Backgrounds

- 2283 In the final state of interest, in which there are at least a photon and a jet, QCD prompt  
 2284 photon production is the main SM process that cannot be reduced. To study this particular  
 2285 background contribution, MC samples are used. Moreover, due to mis-identification and  
 2286 mis-reconstruction of jets, there is a considerable fraction of events in which there is at  
 2287 least one jet being mis-identified as a photon. To study this particular contribution, a  
 2288 data-driven approach has been used as described in Chapter 10.
- 2289 Samples of  $\gamma$ +jet with high statistics are generated using the PYTHIA 8.186 [66] pro-  
 2290 gram. The partonic processes are simulated using LO ME, with the inclusion of initial-  
 2291 and final-state parton showers, where the parametrisation of the proton structure is given  
 2292 by the LO NNPDF2.3 [149] PDFs. The hadronisation process is modelled by the Lund  
 2293 string model [72], briefly discussed in Section 1.3.2, and the sample also counts with a  
 2294 simulation of the UE. The event generator parameters are set according to the “A14” tune  
 2295 for PYTHIA [74]. The PYTHIA sample, given that it is generated at LO, allows for the  
 2296 separation between direct and fragmentation components (see Section 1.1.4).
- 2297 The samples are sliced into different regions of  $p_T$  to optimise the event generation. Details  
 2298 of each sample, including cross-section and filter efficiencies, are shown in Table 7.3.
- 2299 As discussed in Section 6.1, the background in the final data samples is estimated in a  
 2300 data-driven way using different functional forms. MC background samples, nevertheless,  
 2301 are used to filter and rank the most optimal functions.

Table 7.3: Details of the background  $\gamma$ +jet MC samples

| Sample name          | Generator name                | $p_T$ slice [GeV] | Cross section [pb] | Filter Eff. |
|----------------------|-------------------------------|-------------------|--------------------|-------------|
| $\gamma$ +jet direct | PYTHIA 8.244.3+EVTGEN v.1.7.0 | [70, 140]         | 28396.0            | 7.2863E-02  |
| $\gamma$ +jet direct | PYTHIA 8.244.3+EVTGEN v.1.7.0 | [140, 280]        | 2625.5             | 7.0598E-02  |
| $\gamma$ +jet direct | PYTHIA 8.244.3+EVTGEN v.1.7.0 | [280, 500]        | 198.39             | 6.0369E-02  |
| $\gamma$ +jet direct | PYTHIA 8.244.3+EVTGEN v.1.7.0 | [500, 800]        | 18.846             | 4.4596E-02  |
| $\gamma$ +jet direct | PYTHIA 8.244.3+EVTGEN v.1.7.0 | [800, 1000]       | 2.3312             | 2.4130E-02  |
| $\gamma$ +jet direct | PYTHIA 8.244.3+EVTGEN v.1.7.0 | [1000, 1500]      | 0.79945            | 2.3667E-02  |
| $\gamma$ +jet direct | PYTHIA 8.244.3+EVTGEN v.1.7.0 | [1500, 2000]      | 0.055512           | 1.9632E-02  |
| $\gamma$ +jet direct | PYTHIA 8.244.3+EVTGEN v.1.7.0 | [2000, 2500]      | 0.0052361          | 1.6644E-02  |
| $\gamma$ +jet direct | PYTHIA 8.244.3+EVTGEN v.1.7.0 | [2500, 3000]      | 0.00052733         | 1.4446E-02  |
| $\gamma$ +jet direct | PYTHIA 8.244.3+EVTGEN v.1.7.0 | [3000, $\infty$ ] | 4.8856e-05         | 1.4371E-02  |
| $\gamma$ +jet frag   | PYTHIA 8.244.3+EVTGEN v.1.7.0 | [70, 140]         | 106180000.0        | 1.9271E-05  |
| $\gamma$ +jet frag   | PYTHIA 8.244.3+EVTGEN v.1.7.0 | [140, 280]        | 6702000.0          | 2.0959E-05  |
| $\gamma$ +jet frag   | PYTHIA 8.244.3+EVTGEN v.1.7.0 | [280, 500]        | 344070.0           | 2.0507E-05  |
| $\gamma$ +jet frag   | PYTHIA 8.244.3+EVTGEN v.1.7.0 | [500, 800]        | 23711.0            | 1.6991E-05  |
| $\gamma$ +jet frag   | PYTHIA 8.244.3+EVTGEN v.1.7.0 | [800, 1000]       | 2284.6             | 1.0123E-05  |
| $\gamma$ +jet frag   | PYTHIA 8.244.3+EVTGEN v.1.7.0 | [1000, 1500]      | 701.22             | 1.0074E-05  |
| $\gamma$ +jet frag   | PYTHIA 8.244.3+EVTGEN v.1.7.0 | [1500, 2000]      | 70.086             | 5.0238E-06  |
| $\gamma$ +jet frag   | PYTHIA 8.244.3+EVTGEN v.1.7.0 | [2000, $\infty$ ] | 11.548             | 2.4464E-06  |

2302

2303 **EVENT SELECTION AND  
SIGNAL REGIONS  
DEFINITIONS**

# 8

2304

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

---

Billie Jean King

2305 The photon+jet high-mass final state represents an ideal scenario for searches of phys-  
 2306 ics **BSM**. As mentioned in Chapter 6, bumps on the  $m_{\gamma+j}$  distribution over the **SM** are  
 2307 searched for by modeling the background with a functional form. In order to be able to  
 2308 visualise, in case of it exists, any signal (physics **BSM**) over the background, it is indis-  
 2309 pensable to have an excellent knowledge of it, and to be able to separate correctly both.  
 2310 Unlike count-searches, where one needs to define signal, validation and control regions to  
 2311 estimate the background, in resonance searches it is usual to design a signal region that is  
 2312 able to provide a smooth background, but also, a clean signal over it. In this chapter, all  
 2313 the event selection that leads to the signal regions used in this analysis are presented. In  
 2314 Chapter 10, the estimation and modeling of the background are explained in detail.

2315 **8.1 Trigger**

2316 Events are collected by a single photon trigger (`HLT_g140_loose`) with a transverse mo-  
 2317 mentum threshold of 140 GeV which is the lowest unprescaled photon trigger for the most  
 2318 part of the data taking periods<sup>1</sup>, selecting events with at least one photon passing the  
 2319 loose identification criteria. This trigger has been kept unprescaled and is fully efficient  
 2320 for photons with 145 GeV with uncertainties less than 0.1% above 150 GeV. It has been  
 2321 computed with a bootstrap method following the prescriptions in [151] and it is shown in  
 2322 Figure 8.1 as a function of photon  $p_T$ ,  $\eta$  and  $\langle \mu \rangle$  for the different years of data taking.

---

<sup>1</sup> For 2015 the lowest unprescaled trigger was the `HLT_g120_loose`, but the `HLT_g140_loose` was also in the menu

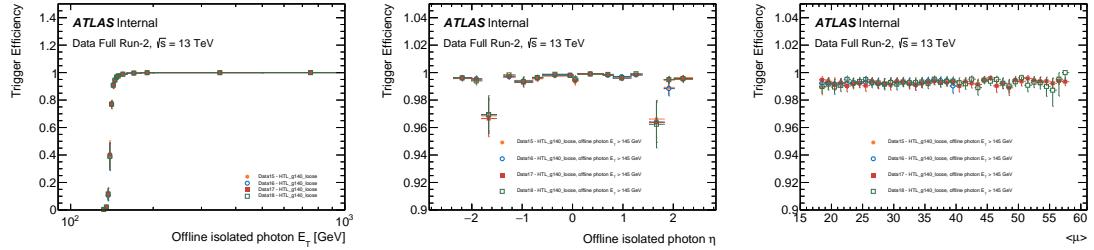


Figure 8.1: Trigger efficiency for HLT\_g140\_loose trigger as a function of photon  $p_T$  (left),  $\eta$  (middle) and  $\langle\mu\rangle$  (right) as measured from data for every year between 2015 and 2018.

## 8.2 Preselection

As discussed in Section 2.3, during a data-taking period, events recollected by the ATLAS detector is grouped into LBs to then compile them into GRLs. These GRLs guarantees that the data used is free of inefficiencies in the detector or sub-detectors, or in the LHC beam. This analysis uses the Full Run-2 dataset collected at  $\sqrt{s} = 13$  TeV, which leads to a total of  $140.1 \pm 1.26 \text{ fb}^{-1}$  after selecting the good-quality events from the GRLs. Additionally, events with different kinds of problems are removed, such as those with calorimeter noise or corrupted events, or those with non-working calorimeter cells.

### 8.2.1 Objects

First, photon, lepton and jet candidates are selected with a set of general requirements, called *baseline*. After this initial selection, an overlap removal procedure is applied to deal with the case of the same particle being reconstructed as different objects. The missing transverse momentum is calculated from these baseline objects. Finally, the photon, lepton and jet candidates used to define the different signal regions must fulfil additional requirements and are hereafter referred to as "signal candidates". In the following paragraphs, a brief description of the objects used in the analysis is given, based on the description shown in Chapter 3.

**Photons** In the offline selection, photon candidates must pass the Tight identification criteria based on the shape of the lateral and longitudinal shower they leave in the calorimeter, pass a  $p_T$  threshold greater than 25 GeV and be contained in an angular range of  $|\eta| < 2.37$ , discarding the barrel-endcap transition region ( $1.37 < |\eta| < 1.52$ ). An additional requirement of  $p_T > 150$  GeV is required for the candidate signal photons so as to ensure that it was selected by the trigger. The photon is also required to be isolated, satisfying the requirements of the Tight WP (defined in Section 3.2.3), which applies cuts on both the calorimetric isolation energy and the track isolation, thus reducing the background from jets misidentified as photons.

2349 **Electrons** The baseline electrons are selected with  $p_T > 10 \text{ GeV}$ ,  $|\eta| < 2.47$ , and originate  
 2350 from the primary vertex. The Loose identification requirement is applied. The signal  
 2351 electrons are further selected applying the Tight identification and the Loose\_VarRad  
 2352 isolation requirement, or HighPtCaloOnly if they have  $p_T > 200 \text{ GeV}$ . After baseline and  
 2353 signal electrons are identified, it is required that there is no electrons in the event.

2354 **Muons** Baseline muons are selected with the Medium identification, have  $p_T > 10 \text{ GeV}$ ,  
 2355  $|\eta| < 2.7$  and originate from the primary vertex. Signal muons are additionally required to  
 2356 pass the Loose\_VarRad isolation [WP](#). A further selection of requiring 0 muons is imposed,  
 2357 as in the final state of this search only photons and jets are expected.

2358 **Jets** [PFlow](#) jets are reconstructed using the anti- $k_t$  algorithm with  $R = 0.4$  as described  
 2359 in Section 3.4, and the baseline selection is defined as those which have  $p_T > 20 \text{ GeV}$  and  
 2360  $|\eta| < 2.8$ . The Neural Network [JVT](#) ([NNJvt](#)) algorithm is used to remove jets originating  
 2361 from pileup interactions for jets with  $p_T < 60 \text{ GeV}$ .

2362 Heavy flavour jets are of great importance for this analysis, since the search will be con-  
 2363 ducted for three different flavours: light,  $c$ , and  $b$ -flavour jets. For this reason, the novel  
 2364 GN2 tagger, defined in Section 3.5, is used to discriminate between these three flavours.  
 2365 Flavour tagging is only applied to the leading jet, and only if it has  $|\eta| < 2.5$ . To accom-  
 2366 plish the simultaneous two-dimensional tagging, first, jets are identified as  $b$ -jets if they  
 2367 pass the 77% tagging efficiency [WP](#). In the second step, those events in which the leading  
 2368 jet is not tagged as a  $b$ -jet are identified as a  $c$ -jet if they pass the loose 50% tagging effi-  
 2369 ciency [WP](#). Those events which fail both the  $b$ -tag and  $c$ -tag [WPs](#) are defined as untagged,  
 2370 or containing a light-jet.

### 2371 8.2.2 Overlap Removal

2372 Due to final-state object misidentification, a single object could be reconstructed as more  
 2373 than one object, being therefore effectively counted multiple times. For this purpose,  
 2374 an overlap removal procedure is taken in to account to eliminate overlaps between the  
 2375 selected objects. The strategy and order of removal is shown in Table 8.1. Two types of  
 2376 baseline objects are compared with each other, based on their closeness in terms of  $\Delta R$   
 2377 as well as other criteria. In each step, if the overlap criteria highlighted under Condition  
 2378 is met, the object listed under the *Object removed* column gets discarded, whereas the  
 2379 *Object compared* to is kept in the event. This way ambiguities in the object reconstruction  
 2380 are resolved and double counting of detector signals as two different types of objects is  
 2381 avoided.

Table 8.1: Overlap removal procedure.

| Step | Object removed | Object compared to | Condition                                    |
|------|----------------|--------------------|--|
| 1    | muon           | electron           | is <b>CT</b> muon and shared <b>ID</b> track |
| 2    | photon         | electron           | $\Delta R < 0.4$                             |
| 3    | photon         | muon               | $\Delta R < 0.4$                             |
| 4    | jet            | electron           | $\Delta R < 0.2$                             |
| 5    | electron       | jet                | $\Delta R < 0.4$                             |
| 6    | jet            | muon               | $\Delta R < 0.2$ and $N_{\text{tracks}} < 3$ |
| 7    | muon           | jet                | $\Delta R < 0.4$                             |
| 8    | jet            | photon             | $\Delta R < 0.4$                             |

## 2382 8.3 Signal regions optimisation

2383 This works aims to search for resonances in the  $\gamma + \text{jet}$  system invariant mass spectrum,  
 2384 therefore the main goal of the event selection is to achieve:

- 2385 • a clean and smoothly falling background distribution, containing mainly direct  $s$ -  
 2386 channel  $\gamma + \text{jet}$  events, rejecting fragmentation,  $t$ -channel and jet faking photons  
 2387 events, and
- 2388 • high signal efficiency and significance.

2389 To determine the basic event selection, studies on basic kinematic variables have been  
 2390 carried out, optimizing in all cases for a high signal significance, and are presented in the  
 2391 following. However, in the first place, several basic cuts are defined below from where all  
 2392 the studies are based upon:

- 2393 • Require at least one tight and isolated photon:  $N_\gamma > 0$ .
- 2394 • At least one jet:  $N_{\text{jets}} > 0$ .
- 2395 • In the hard-scatter interaction for prompt photon production, the photon and the  
 2396 jet carry approximately the same momentum. For this reason, the jet is required to  
 2397 have  $p_T^{\text{jet}} > 150$  GeV.
- 2398 • Remove any non-leading jet with  $p_T^{\text{jet}} < 60$  GeV.
- 2399 • Require same pseudorapidity selection for the jet as for the photon, to avoid photon  
 2400 faking jet events:  $|\eta^j| < 1.37$  or  $1.52 < |\eta^j| < 2.37$ .
- 2401 • Avoid turn-on region of the mass spectrum:  $m_{\gamma+j} > 500$  GeV.

2402 **8.3.1 Photon and jet angular selections**

2403 **Pseudorapidity separation**

2404 The dynamics of the underlying processes in  $2 \rightarrow 2$  hard collinear scattering can be in-  
 2405 vestigated using the variable  $\theta^*$ , where  $\cos \theta^* \equiv \tanh(\Delta y/2)$  and  $\Delta y$  is the difference  
 2406 between the rapidities of the two final-state particles. The variable  $\theta^*$  coincides with the  
 2407 scattering angle in the centre-of-mass frame, and its distribution is sensitive to the spin  
 2408 of the exchanged particle. For processes dominated by  $t$ -channel gluon exchange, such  
 2409 as dijet production in  $pp$  collisions (and therefore fragmentation photon production), the  
 2410 differential cross section behaves as  $(1 - |\cos \theta^*|)^{-2}$  when  $|\cos \theta^*| \rightarrow 1$ . In contrast, pro-  
 2411 cesses dominated by  $t$ -channel quark exchange, such as direct photon production (see  
 2412 Figure 1.6), are expected to have an  $(1 - |\cos \theta^*|)^{-1}$  behaviour when  $|\cos \theta^*| \rightarrow 1$ . For  
 2413 both processes, there are also  $s$ -channel contributions which are, however, non-singular  
 2414 when  $|\cos \theta^*| \rightarrow 1$ . This behaviour on the cross-section has been measured in Ref. [152].  
 2415 Since the analysis considers highly energetic jets and the fact that photons are massless,  
 2416 it is possible to approximate  $\Delta\eta \sim \Delta y$ , and accomplish a removal of non-direct photons.  
 2417 Therefore, for this reason, by removing events with high  $|\Delta\eta(\gamma, j)|$ -values, fragmentation  
 2418 photon and  $t$ -chanel events can be removed.

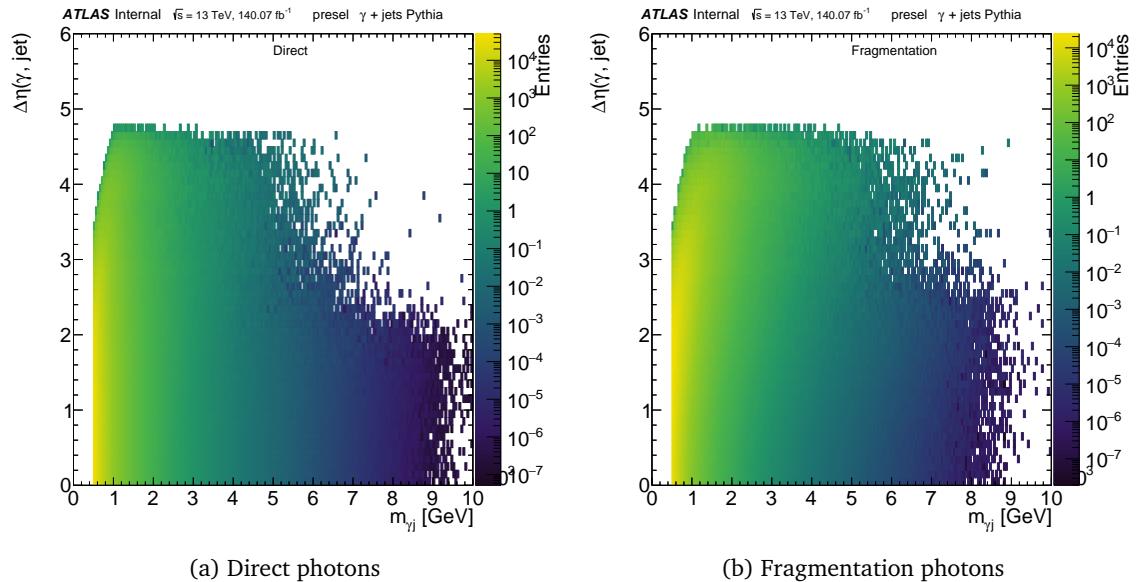


Figure 8.2:  $|\Delta\eta(\gamma, j)| - m_{\gamma+j}$  2D distribution for the  $\gamma + \text{jet}$  PYTHIA background, separating into direct (left) and fragmentation (right) photons.

2419 Figure 8.2 shows the  $|\Delta\eta|$ - $m_{\gamma+j}$  two-dimensional distribution of the  $\gamma + \text{jet}$  background,  
 2420 separating into direct and fragmentation events. It can be seen from these distributions  
 2421 there is a higher concentration of high- $|\Delta\eta(\gamma, j)|$  events for fragmentation photons com-  
 2422 pared to direct photons. This scenario es true regardless of the  $m_{\gamma+j}$  value, but more  
 2423 prominent in the  $1 < m_{\gamma+j} < 5$  TeV region. By selecting events with low  $|\Delta\eta(\gamma, j)|$  it is  
 2424 possible to reject a high proportion of fragmentation photons and  $t$ -channel events.

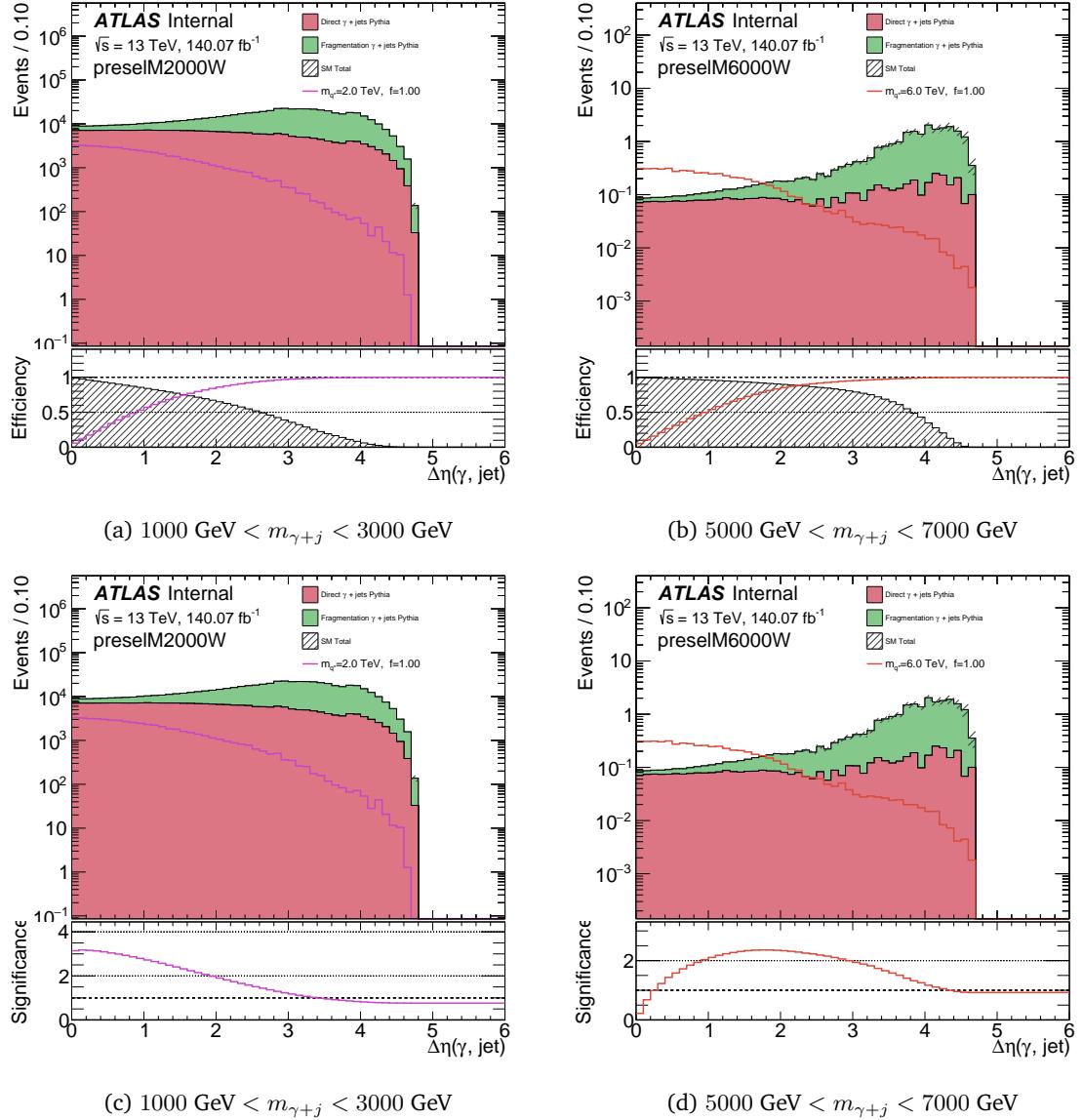


Figure 8.3:  $|\Delta\eta(\gamma, \text{jet})|$  distribution in two  $m_{\gamma+j}$  windows comparison between background and signal models. The bottom pads in Figures 8.3a and 8.3b show the signals efficiencies (coloured lines) and background rejection (hashed histogram) if a cut of the type  $|\Delta\eta| < X$  is applied. On the contrary, on Figures 8.3c and 8.3d, the bottom pad shows the signal significance. The  $\gamma + \text{jet}$  events were modelled with PYTHIA.

2425 Comparisons of background with signal models using this variable are shown in Figure 8.3.  
 2426 The selection in these figures corresponds on selecting events with  $m_{q^*} - 1000 < m_{\gamma+j} <$   
 2427  $m_{q^*} + 1000 \text{ GeV}$ , which consists on a 2 TeV mass window around the signal model's mass.  
 2428 The bottom pad in Figures 8.3a and 8.3b show the efficiency on the signals for the coloured  
 2429 lines, while the hashed histogram shows the background rejection if a cut of the type  
 2430  $|\Delta\eta| < X$  is applied. From these, it can be seen that a cut to this variable at  $|\Delta\eta| \approx 1.6$   
 2431 would help reduce background considerably ( $\sim 80\%$ ) while keeping most of the signal  
 2432 events (efficiency 60 – 80%). Further studies have been carried out as well computing  
 2433 the significance of this cut. Figures 8.3c and 8.3d, show in the bottom pads the signal  
 2434 significance for the same cases. Excepting the trivial cut on  $|\Delta\eta| = 0$ , it can be noted that  
 2435 for higher  $m_{\gamma+j}$  values maximum significance can be achieved by selecting events with

2436  $|\Delta\eta(\gamma, j)| \lesssim 1.6$ , and for this reason, it has been opted to apply the cut  $|\Delta\eta(\gamma, j)| < 1.6$  for  
 2437 the rest of the analysis. Another important aspect that can be seen from these figures is  
 2438 the fact that the majority of fragmentation photon events “live” in the high- $|\Delta\eta|$  region,  
 2439 as anticipated from Figure 8.2b. By applying this cut, fragmentation photon events are  
 2440 highly reduced.

### 2441 Photon and jet pseudorapidity

2442 In the low mass region  $m_{\gamma+j} \lesssim 3$  TeV, there is a high background concentration of events  
 2443 with  $\eta^\gamma > 1.37$  and  $\eta^j > 1.37$ , compared to the benchmark signals, as seen from Figure 8.4.  
 2444 For this reason, by applying a cut on these two variables the signal-to-background ratio  
 2445 would increase with almost no cost on the signal significance nor efficiency, as shown in  
 2446 the bottom pads of the figures. Therefore, it is decided that events with  $\eta^\gamma > 1.37$  and  
 2447  $\eta^j > 1.37$  are removed.

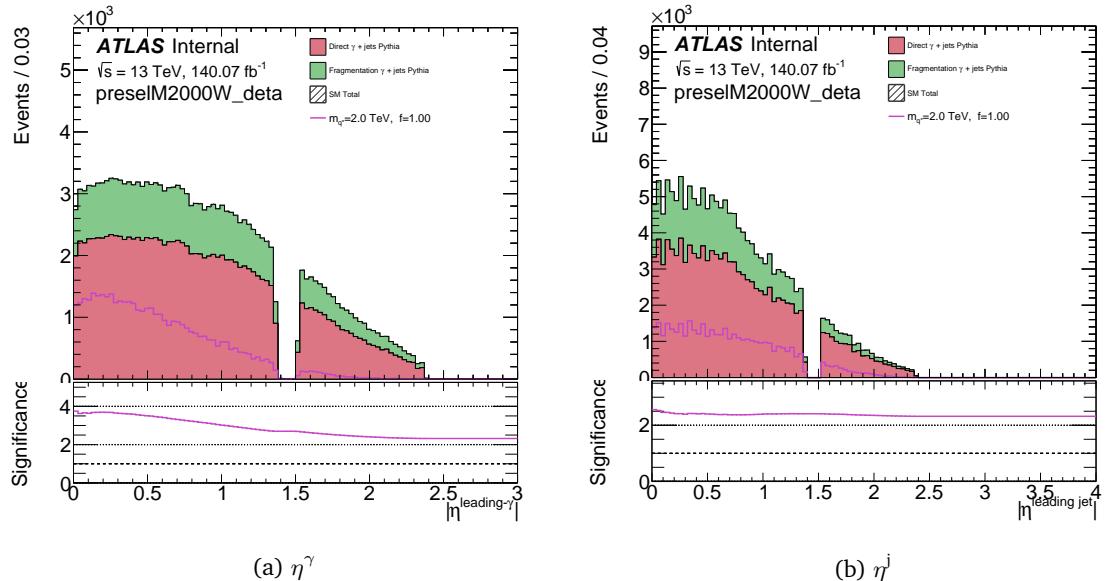
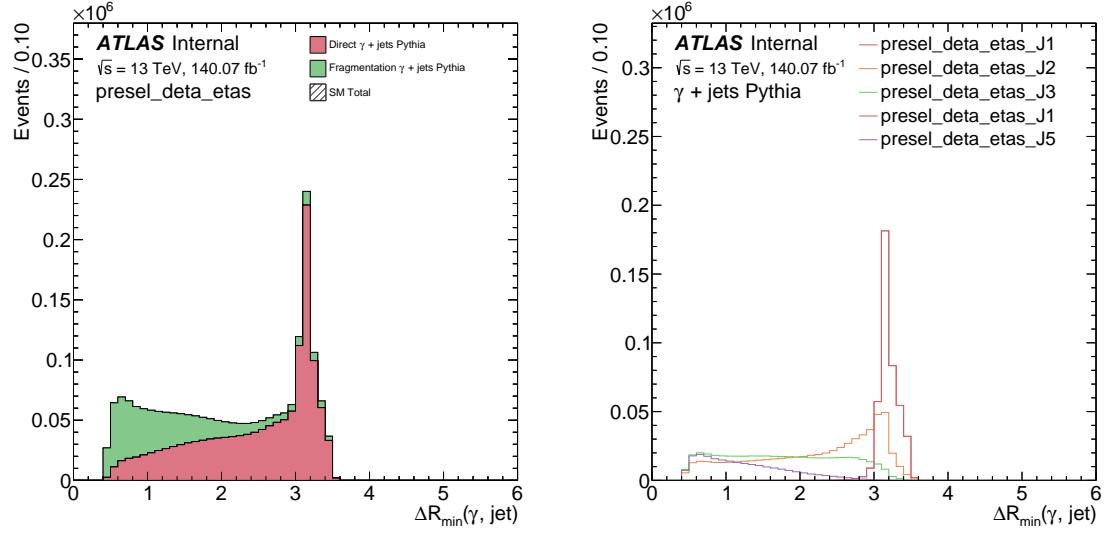


Figure 8.4:  $\eta^\gamma$  and  $\eta^j$  distributions in a window of  $1000 \text{ GeV} < m_{\gamma+j} < 3000 \text{ GeV}$  comparing signals with the main  $\gamma+\text{jet}$  PYTHIA background. Bottom pads show signals significances over the background.

### 2448 8.3.2 Extended isolation

2449 The photon reconstruction, identification and isolation cuts on photons act to reduce in-  
 2450 strumental backgrounds (misidentified hadrons) to a negligible level, but some of the  
 2451 substantial background from secondary (fragmentation) photons remains. To further re-  
 2452 duce this background, the contribution to the leading photon isolation energy from jets  
 2453 close to it is investigated, after applying all the selections shown above.

2454 In Figure 8.5a, the distribution of  $\Delta R_{\min}$  is shown, separating between fragmentation and  
 2455 direct photon events. Said variable, measures the angular distance between the leading  
 2456 photon and the closest jet to the photon. From this distribution it is possible to note that

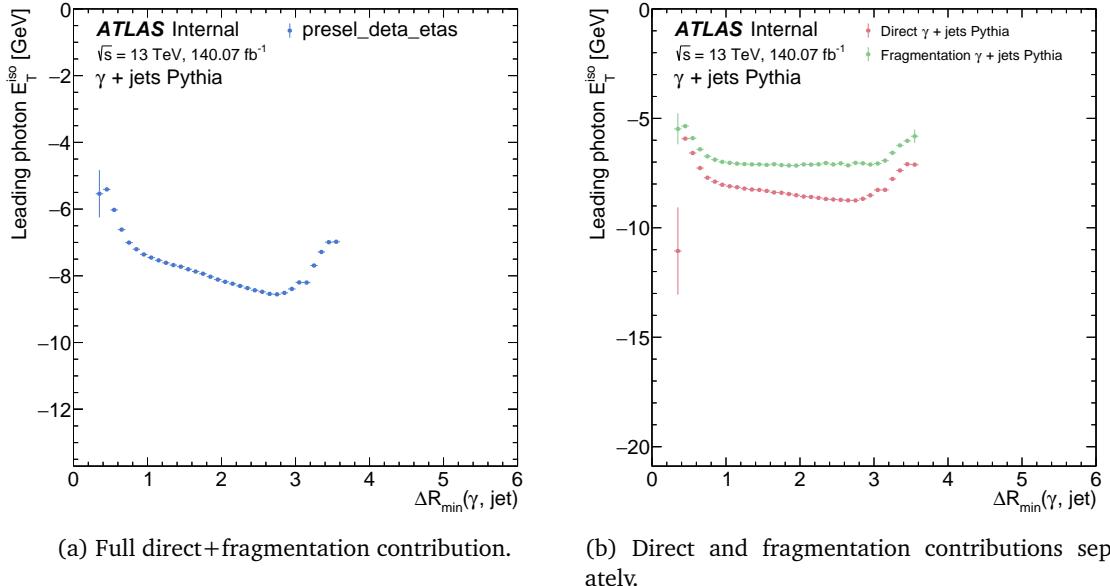


(a) Separation into direct and fragmentation photon events.

(b) Separation into jet multiplicity.

Figure 8.5:  $\Delta R_{\min}$  distribution for the  $\gamma + \text{jet}$  PYTHIA background. This variable shows the minimum distance found between the leading photon and any of the jets.

the majority of events very close to the photon (small  $\Delta R_{\min}$ ) are fragmentation events. Furthermore, the contribution to the said distribution can be seen for events which have different jet multiplicity ( $N_{\text{jet}}$ ), shown in Figure 8.5b. It is known that fragmentation photon events contain more jets than those events coming from direct photon production, and with higher jet multiplicity, the higher the probabilities are that a jet is very close to the photon, contributing to the isolation energy of it.



(a) Full direct+fragmentation contribution.

(b) Direct and fragmentation contributions separately.

Figure 8.6:  $E_T^{\text{iso}}$  contribution as a function of  $\Delta R_{\min}$ . The  $E_T^{\text{iso}}$  values are obtained from the mean of the  $E_T^{\text{iso}}$  distribution.

To study the contribution to the isolation energy of the photon from these jets, Figure 8.6a shows the mean photon isolation energy contribution from the closest jet to the photon.

2465 Jets lying very close to the photon contribute highly to this energy, specially jets with  
 2466  $\Delta R(\gamma, j) < 1.0$ , and in this particular value, the energy starts to drastically increase. This  
 2467 behaviour is also presented separately for direct and fragmentation photon production in  
 2468 Figure 8.6b, from which the higher contribution to the energy is given by fragmentation  
 2469 photon events. From hereinafter, a cut to this variable at  $\Delta R(\gamma, j) \geq 1.0$  is going to be  
 2470 considered to reduce fragmentation events and to obtain a cleaner photon signature.

### 2471 8.3.3 Jet $p_T$

2472 After applying all the aforementioned cuts, one observed, key, feature of fragmentation  
 2473 events is that there is a big proportion of events in which the leading jet carries much  
 2474 more momentum than the leading photon. In an ideal direct photon production, which is  
 2475 what is aimed to using this event selection, both the photon and jet carry approximately  
 2476 the same  $p_T$ . In order to study if a selection based on the jet and photon  $p_T$  is feasible,  
 2477  $p_T^\gamma$  vs  $p_T^{\text{jet}}$  distributions are shown in Figure 8.7 for both direct and fragmentation photons.  
 2478 From these figures, it can be clearly seen that the fragmentation photon production is the  
 2479 one contributing to having events with  $p_T^{\text{jet}} \gg p_T^\gamma$ .

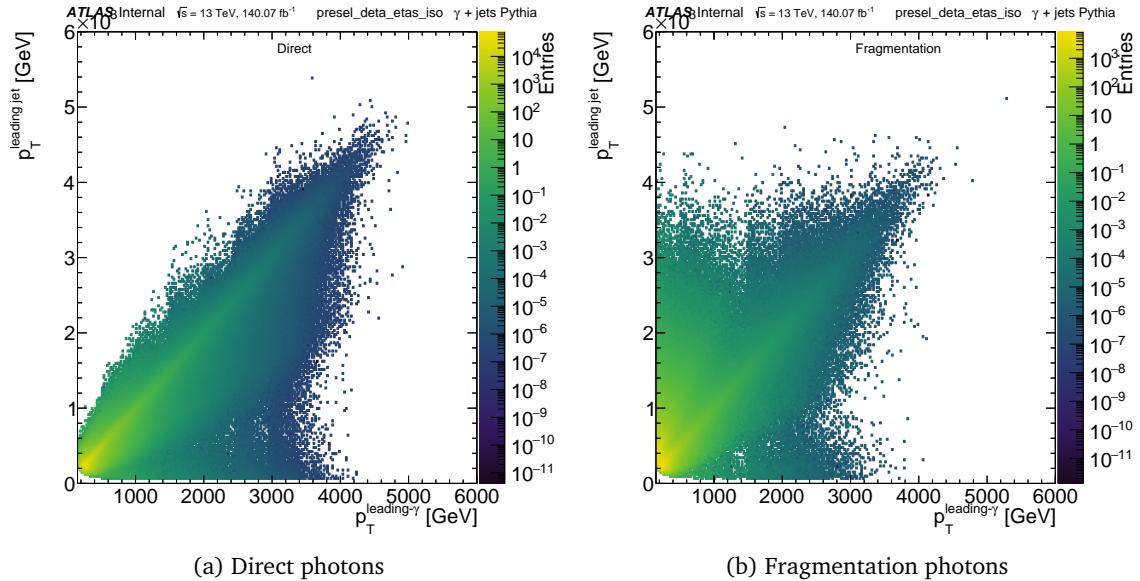


Figure 8.7:  $p_T^\gamma$ - $p_T^{\text{jet}}$  2D distribution for direct and fragmentation PYTHIA  $\gamma +$ jet background.

2480 In order to clean the sample from fragmentation photon contributions even more, the  
 2481 events selected are those that satisfy:

$$\frac{p_T^{\text{jet}} - p_T^\gamma}{p_T^\gamma} < X, \quad X \in [0, 1] \quad (8.1)$$

2482 where  $X$  is the allowed fraction of  $p_T^\gamma$  the jet has to have, hence defining an upper value for  
 2483  $p_T^{\text{jet}}$  for a given  $p_T^\gamma$ . The optimal value is found to be  $X = 0.5$ , at which the signal efficiency  
 2484 is very high, while the rejected background consists solely of fragmentation events.

2485 In Figure 8.8, the background  $p_T^\gamma$  vs  $p_T^{\text{jet}}$  distribution is shown again separately for dir-  
 2486 ect and fragmentation photons, from which it is noted that the vast majority of events

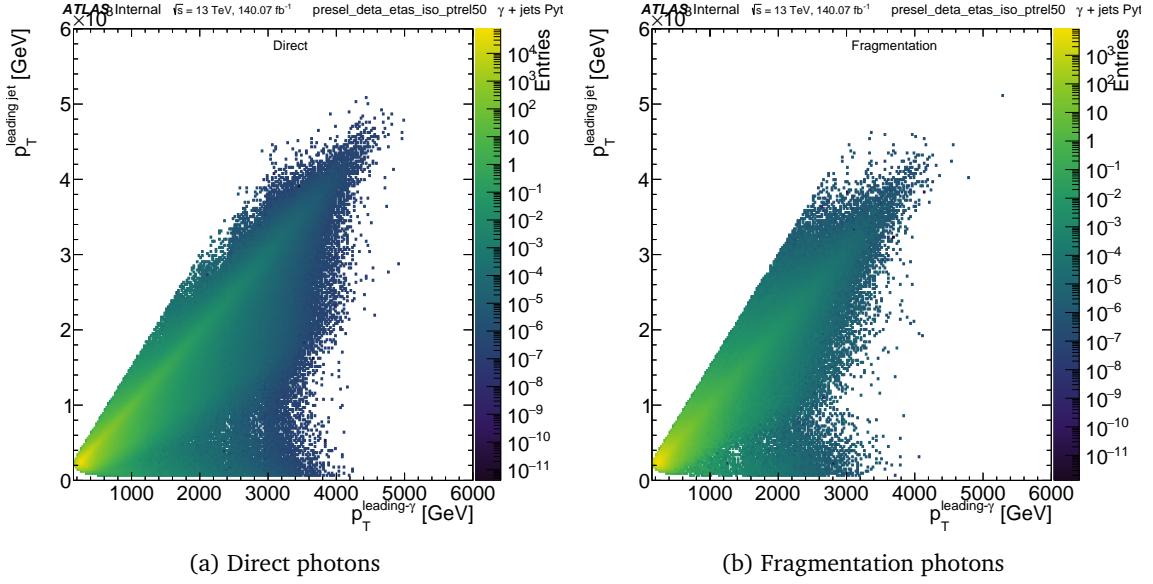


Figure 8.8:  $p_T^\gamma - p_T^{\text{jet}}$  2D distribution for direct and fragmentation PYTHIA  $\gamma + \text{jet}$  background, selecting events in which the leading jet  $p_T$  satisfies Eq. 8.1 with  $X = 0.5$ .

removed are fragmentation. The same distribution for different **EQ** signals is shown in Figure 8.9, having efficiencies greater than 95%. The cut efficiency for background and some  $q^*$  signals are shown in Table 8.2.

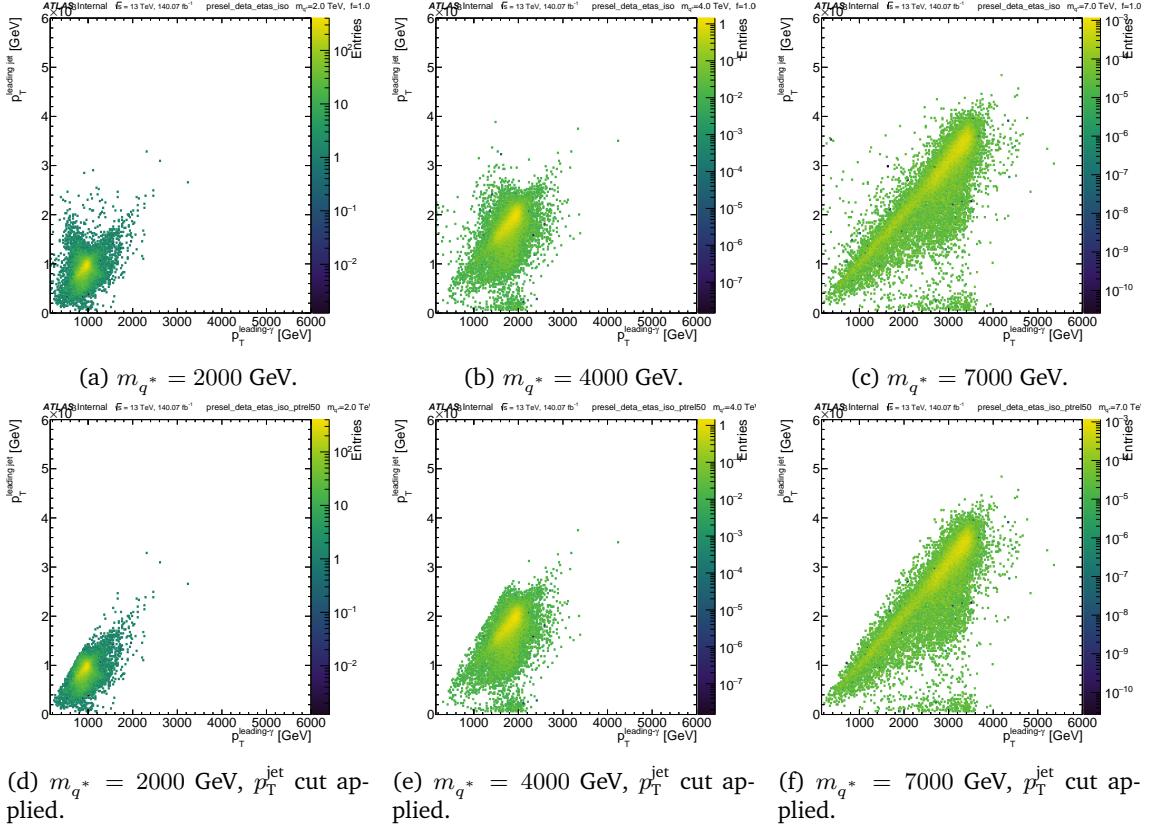


Figure 8.9:  $p_T^\gamma - p_T^{\text{jet}}$  2D distribution for different  $q^*$  signal samples with  $f = 1.0$ , selecting jets according to formula Eq. 8.1 with  $X = 0.5$ . The first three figures shown the 2D distribution without the application of the  $p_T^{\text{jet}}$  cut, while the three at the bottom contain it.

Table 8.2: Background and signal efficiency on the  $p_T^{\text{jet}}$  cut defined above.

|  | $\varepsilon_{\text{rel}}$ |
|--|----------------------------|
| $\gamma + \text{jet Pythia}$           | 0.8827                     |
| $m_{q^*} = 2000 \text{ GeV}, f = 1.00$ | 0.9707                     |
| $m_{q^*} = 4000 \text{ GeV}, f = 1.00$ | 0.9918                     |
| $m_{q^*} = 7000 \text{ GeV}, f = 1.00$ | 0.9942                     |

2490 The variable that is affected the most by this particular cut is, as expected,  $p_T^{\text{jet}}$ . This  
 2491 distribution is shown in Figure 8.10 before and after the cut is applied. It is observed how  
 2492 the fragmentation events contribution drastically decrease, making the  $p_T^{\text{jet}}$  distribution  
 2493 much smoother. Moreover, since the observable of interest is  $m_{\gamma+j}$ , a comparison of this  
 2494 distribution is shown in Figure 8.11, to assess the changes in the spectrum. Applying  
 2495 this particular cut on  $p_T^{\text{jet}}$  has almost not effect on the final  $m_{\gamma+j}$  distribution. Very small  
 2496 differences are observed but at very low  $m_{\gamma+j}$ . This is due to the asymmetry in  $p_T$  that the  
 2497 jet and photon had ( $p_T^{\text{jet}} \gg p_T^\gamma$ ).

2498 In the final selection, the cut  $(p_T^{\text{jet}} - p_T^\gamma)/p_T^\gamma < 0.5$  is applied.

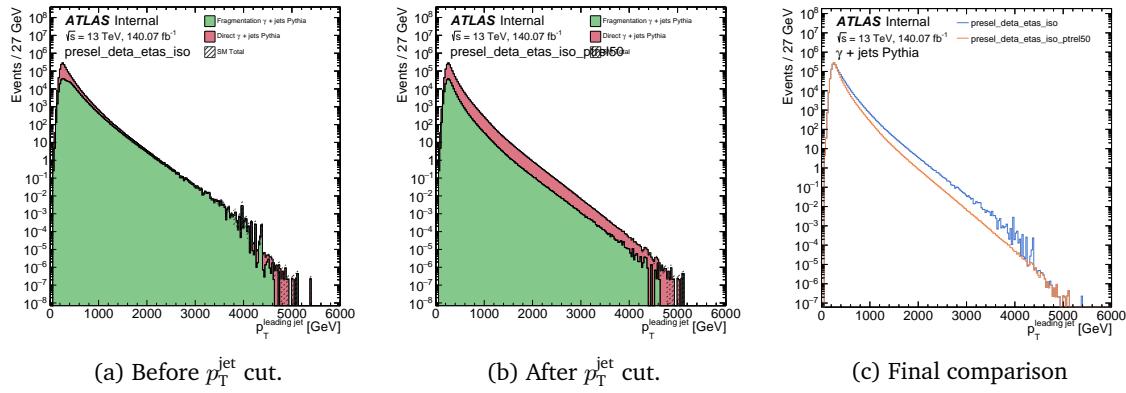


Figure 8.10:  $p_T^{\text{jet}}$  distribution before and after the cut to remove fragmentation events. The contributions for direct and fragmentation photons are shown separately. On the third plot, the orange (blue) line represents the distribution after (before) the cut is applied.

## 2499 8.4 Signal regions

2500 By applying the previously defined cuts (summarised in Table 8.12), it is possible to obtain  
 2501 a clean  $m_{\gamma+j}$  distribution, where the vast majority of fragmentation photon and  $t$ -channel  
 2502 events are removed, but still having high signal efficiency.

2503 This work will benefit from the further separation that can be achieved by classifying  
 2504 the leading jet into three possible flavours: light-,  $c$ - or  $b$ -tagged jets. Making use of  
 2505 the current ATLAS  $b$  and  $c$ -jet GN2 tagger, the signal regions SRB SRC and SRL can be  
 2506 defined. A scheme of how this separation occurs is presented in Figure 8.12. First,  $b$ -jets  
 2507 are discriminated against light- and  $c$ -jets by means of the 77%  $b$ -tag efficiency WP.  
 2508 By selecting those jets that fail to enter region SRB, a  $c$ -tagger WP of 50%  $c$ -tag efficiency is

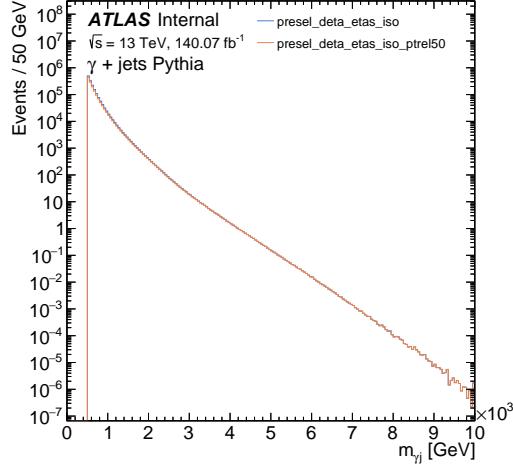


Figure 8.11:  $m_{\gamma+j}$  distribution before and after the  $p_T^{\text{jet}}$  cut. The orange (blue) line represents the distribution after (before) the cut is applied.

2509 applied to select  $c$ -jets, and those do not pass this  $c$ -tagger, are classified as untagged, or,  
 2510 simply  $l$ -jets.

Table 8.3: Signal regions definitions. \*\*\* for the moment, all plots have the old naming SRC50/SRL50, but  
 in the text I used SRC/SRL. Should I keep the old naming until I re-do the plots with the correct text?

| Cut  | SR | SRB  | SRC                                     | SRL  |
|--|----|------|---|------|
| $N_\gamma$                                   |    |      | $> 0$                                   |      |
| $p_T^\gamma$ [GeV]                           |    |      | $> 150$                                 |      |
| $N_{\text{jet}}$                             |    |      | $> 0$                                   |      |
| $p_T^{\text{jet}}$ [GeV]                     |    |      | $> 150$                                 |      |
| $\eta^\gamma$                                |    |      | $< 1.37 \parallel (> 1.52 \&\& < 2.37)$ |      |
| $m_{\gamma+j}$ [GeV]                         |    |      | $> 500.$                                |      |
| $ \Delta\eta(\gamma, j) $                    |    |      | $< 1.6$                                 |      |
| $\eta^j$                                     |    |      | $< 1.37$                                |      |
| $\Delta R_{\min}(\gamma, j)$                 |    |      | $\geq 1.0$                              |      |
| $(p_T^{\text{jet}} - p_T^\gamma)/p_T^\gamma$ |    |      | $< 0.5$                                 |      |
| $b$ -tag 77%                                 | -  | Pass |   | Fail |
| $c$ -tag 50%                                 | -  | -    | Pass                                    | Fail |

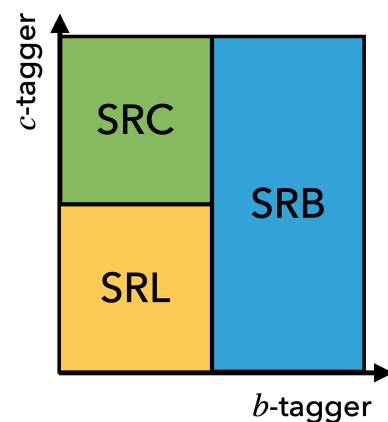


Figure 8.12: Two-dimensional sequential tagger scheme.

2511

2512 **SIGNAL MODELING AND  
SYSTEMATIC  
UNCERTAINTIES**

# 9

2513 This search targets three different types of  $\gamma$ +jet resonance signals: [EQs](#), [QBHs](#) and gen-  
 2514 eric, Gaussian-shaped resonances. To perform hypothesis tests for these signals, their  
 2515 expected  $m_{\gamma+j}$  spectra must be known. These predicted signal shapes are modeled by a  
 2516 continuous [PDF](#) that are later included in the [S+B](#) fits. Also, the theoretical and exper-  
 2517 imental systematic uncertainties are studied in this chapter, as these enter the final fits  
 2518 performed to data.

2519 **9.1 Signal modeling**

2520 The fully-simulated signal model describes both the invariant mass shape, the efficiency  
 2521 and the signal acceptances. However, these signals are generated in a limited set of masses  
 2522 (and couplings in the case of [EQs](#) signals), but one would like to evaluate the signal  
 2523 predictions at any intermediate parameter, that is, rely on interpolated signals.

2524 To accomplish this interpolation, a two step process is followed. First, it is necessary that  
 2525 each one of the simulated signals follow a continuous distribution rather than a histogram.  
 2526 This process of transforming a histogram to a [PDF](#) was employed in the calculation of [FFs](#),  
 2527 in Section 5.1.2, but with the purpose of smoothing the histograms. In this case, the [KDE](#)  
 2528 algorithm is used to retrieve a [PDF](#) from the signal histogram.

2529 The fine factors are optimized by eye for parameter of the theory, both for the [EQ](#) model  
 2530 and [QBH](#) models. Examples of the optimization can be found on Figure 9.1 for  $q^*$  and  $c^*$   
 2531 signals. As the number of entries in the main core of the distributions span multiple orders  
 2532 of magnitude, it is most important to model correctly the peak of the distributions rather  
 2533 than the tails, therefore in some cases the tails are not perfectly modeled for all entries of  
 2534 the original distributions.

2535 Once all the [PDFs](#) are obtained, signals models for any intermediate mass and/or coupling  
 2536 are obtained with a moment morphing method as described in Ref. [153]. The interpola-  
 2537 tion is done in a pair-wise manner: to obtain any intermediate [PDF](#), the closest two to  
 2538 the desired mass/coupling value are used. For instance, to obtain an interpolated signal  
 2539 with  $m_{q^*} = 3200$  GeV,  $f = 1$ , the input signals to the interpolation are the ones with

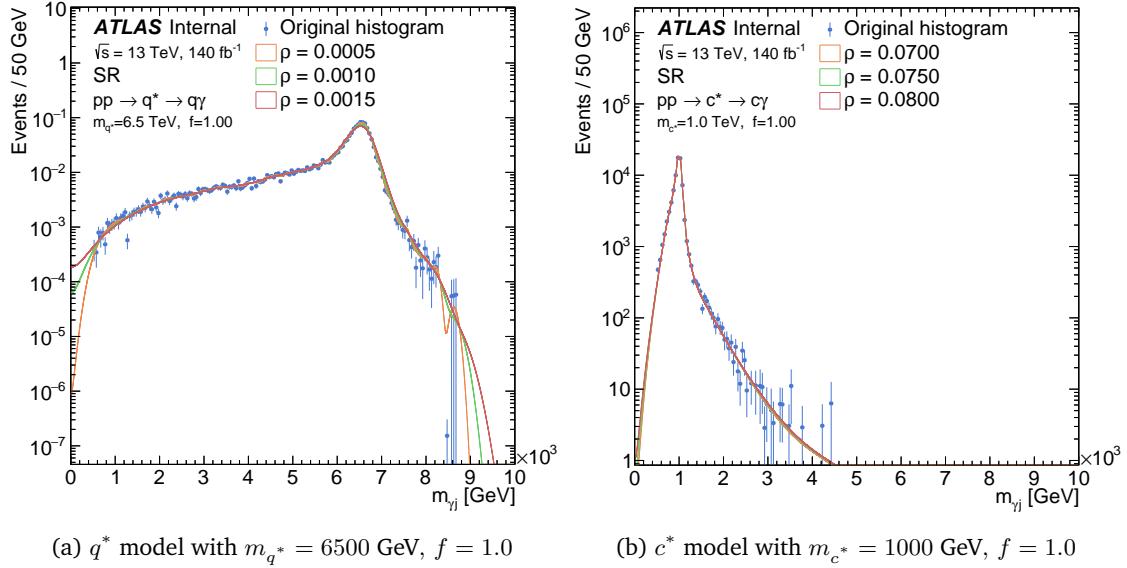


Figure 9.1: Fine factors optimization for the  $q^*$  and  $c^*$  signal models. Points represent original distribution, while colored lines represent the estimated PDFs.

2540  $m_{q^*} = 3000$  GeV and  $m_{q^*} = 4000$  GeV. This method of obtaining the signal samples is  
 2541 shown in Figure 9.2, which contains the interpolated signals with dashed lines and the  
 2542 original ones with solid lines for the  $q^*$  model. For all the interpolated signals, shapes are  
 2543 sensible and lead to a smooth transition between original signal samples.

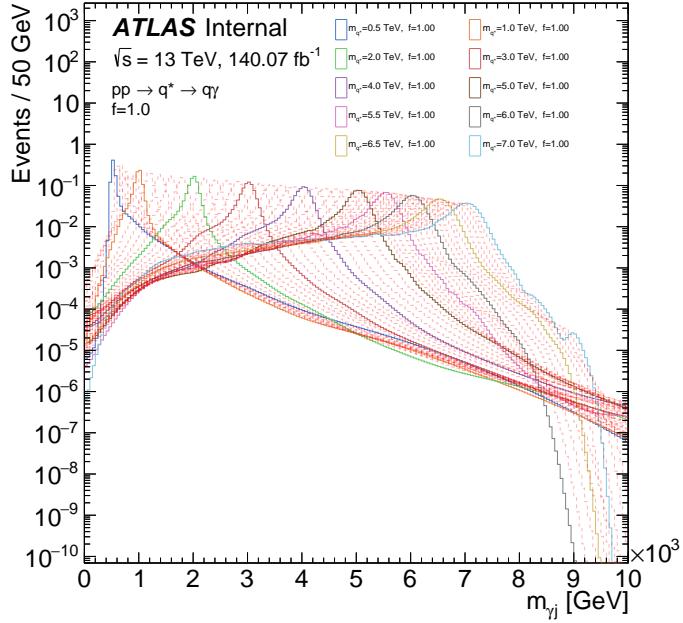


Figure 9.2: Illustration of the moment morph method to interpolate between PDFs, for the case of  $q^*$  signals with  $f = 1$ . The original signal distributions are shown with the solid lines with different colors, while all the interpolated shapes are shown with the dashed red line.

## 2544 9.2 Acceptances and efficiencies

2545 The event selection described in the previous chapter is aimed to reduce non- $s$ -channel  
 2546  $\gamma$ +jet backgrounds while maintaining a high signal efficiency. In addition to that, the  
 2547 novel GN2 tagger is used to separate the inclusive signal region SR into three orthogonal  
 2548 ones: SRB for  $b$ -jets, SRC for  $c$ -jets, and SRL for  $l$ -jets. This is a crucial aspect in the  
 2549 analysis as it allows, for the first time at the LHC, to study EQ resonances initiated by a  $c^*$ ,  
 2550 orthogonal to other flavors.

Table 9.1: EQ signals significances over the total background in the SR, SRB, SRC and SRL regions. The considered signals for each flavor have coupling  $f = 1.0$ .

(a)  $m = 2000$  GeV in a window of  $1000 < m_{\gamma+j} < 3000$  GeV

| Signal channel                | SR                      | SRB                   | SRC                    | SRL                     |
|-------------------------------|-------------------------|-----------------------|------------------------|-------------------------|
| Total Bkg. events             | 89030.428               | 2570.406              | 14911.838              | 71548.184               |
| jet $\rightarrow \gamma$ fake | 3649.795                | 180.319               | 673.710                | 2795.767                |
| $\gamma +$ jets PYTHIA        | 85380.633               | 2390.087              | 14238.128              | 68752.418               |
| $b^*$                         | 242.824, $Z = 0.810$    | 147.637, $Z = 2.880$  | 40.605, $Z = 0.330$    | 54.582, $Z = 0.200$     |
| $c^*$                         | 1576.968, $Z = 5.270$   | 108.852, $Z = 2.130$  | 729.760, $Z = 5.930$   | 738.356, $Z = 2.760$    |
| $q^*$                         | 31183.749, $Z = 99.160$ | 625.209, $Z = 11.880$ | 4335.836, $Z = 33.970$ | 26222.705, $Z = 92.810$ |

(b)  $m = 4000$  GeV in a window of  $3000 < m_{\gamma+j} < 5000$  GeV

| Signal channel                | SR                    | SRB                | SRC                 | SRL                   |
|-------------------------------|-----------------------|--------------------|---------------------|-----------------------|
| Total Bkg. events             | 200.888               | 5.999              | 27.677              | 167.212               |
| jet $\rightarrow \gamma$ fake | 7.014                 | 0.334              | 1.419               | 5.260                 |
| $\gamma +$ jets PYTHIA        | 193.874               | 5.665              | 26.258              | 161.951               |
| $b^*$                         | 0.708, $Z = 0.050$    | 0.281, $Z = 0.110$ | 0.136, $Z = 0.030$  | 0.291, $Z = 0.020$    |
| $c^*$                         | 5.514, $Z = 0.390$    | 0.337, $Z = 0.140$ | 1.668, $Z = 0.310$  | 3.509, $Z = 0.270$    |
| $q^*$                         | 295.673, $Z = 17.530$ | 6.803, $Z = 2.410$ | 33.739, $Z = 5.520$ | 255.131, $Z = 16.500$ |

2551 In high-energy physics it is fundamental to first study the sensitivity of the signals that are  
 2552 searched for. In Table 9.1, the number of events for each considered background<sup>1</sup>, as well  
 2553 as the number of signal events in each of one of the signal regions, for two benchmark  
 2554 signals. The signal regions in the tables, however, have an additional cut on  $m_{\gamma+j}$ , such  
 2555 that only a window around the hypothesized signal is covered. The three considered  
 2556 flavours are shown in the table, and for each one the signal significance is calculated  
 2557 according to

$$Z = \sqrt{2 \left( (s+b) \ln \left( 1 + \frac{s}{b} \right) - s \right)}. \quad (9.1)$$

2558 In both cases, the  $q^*$  signal has the highest significance, regardless of the signal region.  
 2559 This is a consequence of the high cross-section this process has, and for this reason, the  
 2560 inclusive SR region is used to search for  $q^*$  signals.

2561 One important feature to note is the very high improvement of the  $b^*$  signals in the dedi-  
 2562 cated  $b$ -tagging regions, when compared to the inclusive one. In the case of  $b^*$  with  
 2563  $m_{b^*} = 2000$  GeV, the significance is increased by a factor of 3.5. This improvement is due  
 2564 to the impressive performance of the GN2 tagger, which allows a great separation of  $b$ -jets  
 2565 against other flavors.

<sup>1</sup> The jet-faking-photon background is studied in Chapter 10.

2566 Respecting the  $c^*$  signals, the dedicated  $c$ -tagging region SRC leads to an increase on the  
 2567 significance over the background, compared to the inclusive SR region. This behaviour,  
 2568 however, is prominent at lower  $m_{c^*}$ , where the performance of the GN2 tagger is optimal.  
 2569 Nevertheless, since no drastic separation between  $c$ -quarks and  $l$ -quarks can be obtained  
 2570 with the current tagging algorithm, a non-negligible quantity of  $c^*$  events remain in the  
 2571  $l$ -tagged region SRL. In order to achieve greater sensitivity for this particular signal, the  
 2572 search for the  $c^*$  signals is carried out in three orthogonal regions simultaneously, herein-  
 2573 after referred as SRC+SRB+SRL.

2574 The signals production cross sections are corrected to the visible cross section by multiplying  
 2575 them by  $A \times \varepsilon$ , where  $A$  is the probability of the event selection criteria to accept the  
 2576 signal event (**EQ** or **QBH** in this work), referred as the *acceptance*, and  $\varepsilon$  is the reconstruc-  
 2577 tion and identification efficiency. The factor  $A \times \varepsilon$  is of crucial interest for theorists as well,  
 2578 as it allows them to compare theoretical results with the experiments.

2579 The acceptance is computed for each signal as the ratio

$$A = \frac{N_{\text{pass}}^{\text{truth}}}{N_{\text{total}}} \quad (9.2)$$

2580 where  $N_{\text{total}}$  is the total number of generated events and  $N_{\text{pass}}^{\text{truth}}$  is the number of events  
 2581 passing the event selection criteria at particle (truth) level (that is, all the even selection  
 2582 based on kinematic cuts only). On the other hand, the selection efficiency is calculated as

$$\varepsilon = \frac{N_{\text{pass}}^{\text{reco}/\text{ID}}}{N_{\text{pass}}^{\text{truth}}}, \quad (9.3)$$

2583 and  $N_{\text{pass}}^{\text{reco}/\text{ID}}$  is the number of events that pass all the reconstruction and identification  
 2584 requirements, such as photon identification and isolation, jet cleaning and trigger. \*\*\*  
 2585 **look at definitions, need to change results?**

2586 Table 9.2 shows the measured acceptance for two benchmark **EQ** signals with different  
 2587 mass, and are all displayed in Figure 9.3a, for each flavour and with unit coupling. The  
 2588 same results for **QBH** signals are presented in Figure 9.3b. For the  $c^*$  and  $q^*$  signals, it can  
 2589 be seen that lower acceptances are obtained, compared to the  $q^*$  signals. This reason for  
 2590 this behaviour is due to the lepton-veto. Heavier quarks are likely to decay into lighter  
 2591 quarks accompanied by a  $W$  that decays into a pair of leptons  $\ell\bar{\nu}$ , or to a pair of quarks  
 2592 that hadronise, then it is most likely that a lepton is present in the event in this case. In  
 2593 the  $q^*$  case, only  $\sim 10\%$  of the events contain a lepton, while this number increases to  
 2594 almost  $\sim 30\%$  for the  $b^*$  signal, as seen from Table 9.2.

2595 Similarly, in Figure 9.4, reconstruction and identification efficiencies are shown for the  
 2596 three flavors of the **EQ** signals. As it is expected, best performance is obtained for  $q^*$   
 2597 signals when neither of the  $b$  or  $c$ -tagging is applied. Moreover, as it was discussed in  
 2598 Section 3.5, a significant decrease on the efficiency is observed on the performance of the  
 2599 GN2 tagger, and this is reflected on the measured efficiencies in Figures 9.4b and 9.4c,  
 2600 where for higher masses the tagger efficiency decreases to almost half its initial value.

Table 9.2: Acceptance measurements for two benchmark **EQ**  $q^*$  signals with  $f = 1.0$ . The second and fifth columns denote the cutflow for all the cuts, the third and sixth the absolute efficiencies, and the fourth and seventh the relative efficiencies of each cut.

| Cut  | $m_{q^*} = 4000 \text{ GeV}$ |                            |                            | $m_{b^*} = 4000 \text{ GeV}$ |                            |                            |
|--|------------------------------|----------------------------|----------------------------|------------------------------|----------------------------|----------------------------|
|  | Events                       | $\varepsilon_{\text{abs}}$ | $\varepsilon_{\text{rel}}$ | Events                       | $\varepsilon_{\text{abs}}$ | $\varepsilon_{\text{rel}}$ |
| Total Events   | 50000                        | 1.0000                     |                            | 50000                        | 1.0000                     |                            |
| $N_\gamma > 0$ and $N_{\text{jets}} > 0$ after baseline sel.   | 47702                        | 0.9540                     |                            | 47318                        | 0.9464                     |                            |
| $N_\gamma > 0$ and $N_{\text{jets}} > 0$ after OR              | 47356                        | 0.9471                     | 0.9927                     | 45304                        | 0.9061                     | 0.9574                     |
| $N_\gamma > 0$ and $N_{\text{jets}} > 0$ after signal sel.     | 45811                        | 0.9162                     | 0.9674                     | 42505                        | 0.8501                     | 0.9382                     |
| Skim $N_\gamma > 0$  | 45811                        | 0.9162                     | 1.0000                     | 42505                        | 0.8501                     | 1.0000                     |
| Skim $p_T^\gamma > 145 \text{ GeV}$                            | 45796                        | 0.9159                     | 0.9997                     | 42500                        | 0.8500                     | 0.9999                     |
| Skim $ \eta^\gamma  < 1.37$ or $(1.52 <  \eta^\gamma  < 2.37)$ | 45796                        | 0.9159                     | 1.0000                     | 42500                        | 0.8500                     | 1.0000                     |
| Skim $N_{\text{jets}} > 0$                                     | 45796                        | 0.9159                     | 1.0000                     | 42500                        | 0.8500                     | 1.0000                     |
| Skim $N_{\text{leptons}} = 0$                                  | 44304                        | 0.8861                     | 0.9674                     | 29659                        | 0.5932                     | 0.6979                     |
| $p_T^\gamma > 150 \text{ GeV}$                                 | 44304                        | 0.8861                     | 1.0000                     | 29658                        | 0.5932                     | 1.0000                     |
| $p_T^{\text{jet}} > 150 \text{ GeV}$                           | 43612                        | 0.8722                     | 0.9844                     | 28504                        | 0.5701                     | 0.9611                     |
| $ \eta^j  < 1.37$ or $(1.52 <  \eta^j  < 2.37)$                | 41197                        | 0.8239                     | 0.9446                     | 27362                        | 0.5472                     | 0.9599                     |
| $m_{\gamma j} > 500 \text{ GeV}$                               | 41180                        | 0.8236                     | 0.9996                     | 27291                        | 0.5458                     | 0.9974                     |
| $ \Delta\eta(\gamma, j)  < 1.6$                                | 30219                        | 0.6044                     | 0.7338                     | 19821                        | 0.3964                     | 0.7263                     |
| $ \eta^\gamma  < 1.37$   | 30056                        | 0.6011                     | 0.9946                     | 19603                        | 0.3921                     | 0.9890                     |
| $ \eta^j  < 1.37$  | 29880                        | 0.5976                     | 0.9941                     | 19472                        | 0.3894                     | 0.9933                     |
| $\Delta R_{\min}(\gamma, j) \geq 1.0$                          | 27795                        | 0.5559                     | 0.9302                     | 18073                        | 0.3615                     | 0.9282                     |
| $(p_T^{\text{jet}} - p_T^\gamma)/p_T^\gamma \leq 0.5$          | 27795                        | 0.5559                     | 1.0000                     | 18073                        | 0.3615                     | 1.0000                     |
| Acceptance   |                              | 0.5559                     |                            |                              | 0.3615                     |                            |

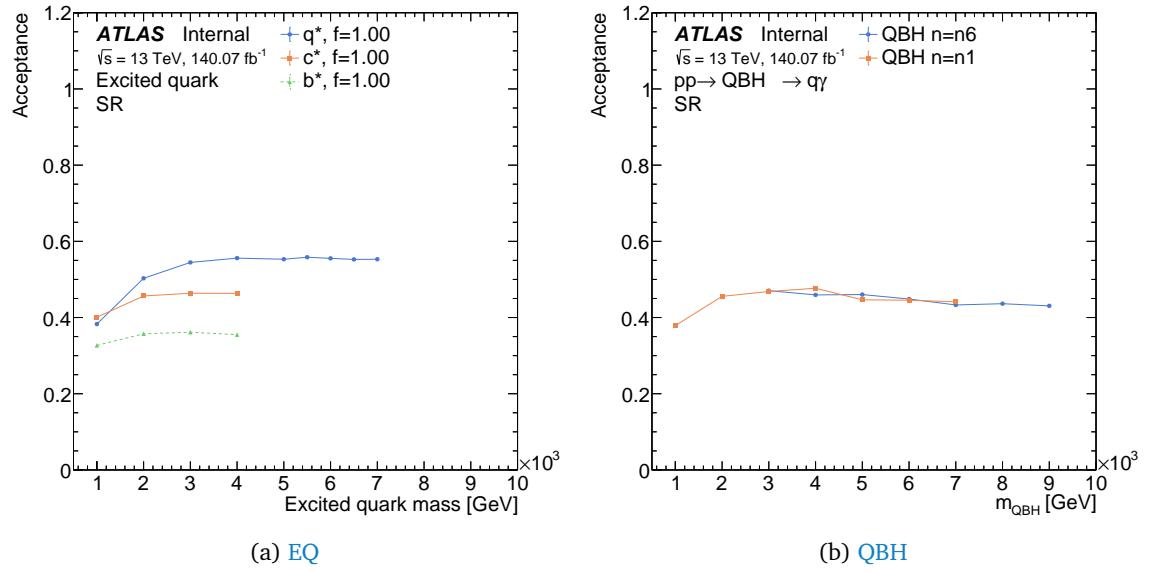


Figure 9.3: Measured acceptances for the **EQ** (left) and **QBH** (right).

## 2601 9.3 Systematic uncertainties

2602 Systematic uncertainties are one of the most crucial aspects in a search, given that they  
 2603 modify the signal expected values in the different signal regions of the analysis. These  
 2604 uncertainties are included in the statistical model as nuisance parameters, explained in  
 2605 Section 6.2.2. As in this analysis the background is obtained directly from data, systematic

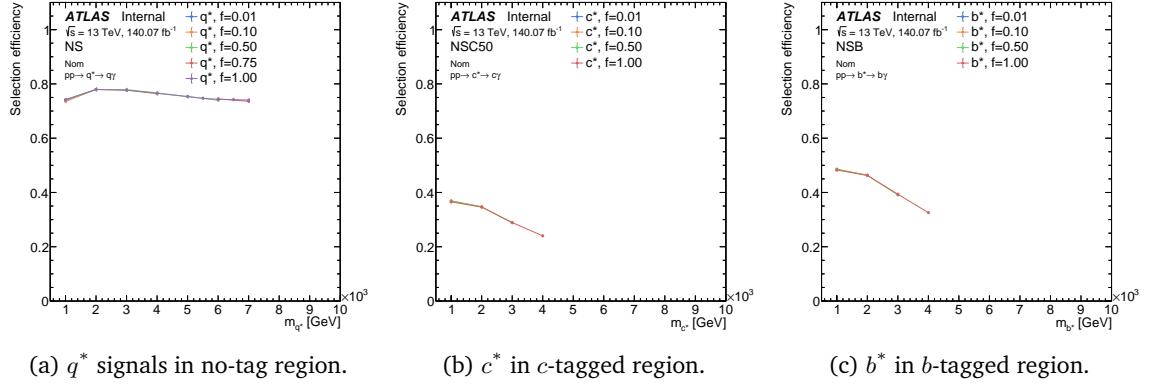


Figure 9.4: Measured efficiencies for the EQ signals. Figure a:  $q^*$  signal when no tagging is applied whatsoever; Figure b:  $c^*$  signal with  $c$ -tagging applied; and Figure c:  $b^*$  signal with  $b$ -tagging applied.

uncertainties only enter the statistical model through the signals. However, uncertainties related to the background modeling are still taken into account, but a specific estimation of those is given in Chapter 10. There are two types of systematic uncertainties affecting the signals: experimental and theoretical, and both are described in the following.

### 9.3.1 Experimental uncertainties

Experimental uncertainties arise from uncertainties on the simulation of the detector, reconstruction or calibration of physical objects, corrections due to pileup or luminosity. For each one of the considered sources of uncertainties, with the exception of the luminosity ones, their effect is studied on the relative difference seen on the selection efficiency of the signals.

#### Luminosity and pileup reweighting uncertainty

The uncertainty on the combined full Run-2 integrated luminosity is obtained using the LUCID-2 detector [109] for the primary luminosity measurements, complemented by measurements using the ID and calorimeters. The obtained uncertainty is of 0.83% [108]. Moreover, uncertainties related to the reweighting of MC events to match the pile-up distributions found in data need to be taken into account for which a nuisance parameter is added to the fit, allowing for variations of the applied pile-up event weight [154].

#### Photon identification, isolation, trigger, energy scale and resolution

The uncertainty on the photon identification is evaluated by computing the difference between identification efficiencies by applying the uncertainties on the photon FF, computed in Chapter 5. Regarding the uncertainties on photon isolation, they are computed by applying data-driven shifts to the simulated  $E_T^{\text{iso}}$  shapes. The effect of these uncertainties on the selection efficiency is found to be, in all cases,  $< 0.2\%$ . The photon energy scale and resolution is considered as well, and the uncertainties are found to be  $< 0.5\%$ .

2630 **Jet energy scale (JER) and resolution (JES)**

2631 The uncertainties associated with jets are estimated following the methodology described  
 2632 in Refs. [155] and [156], which come from multiple sources and provide a large number  
 2633 of nuisance parameters. Among them are the uncertainty associated with the energy  
 2634 resolution (JER), obtained from the variation in the smearing of the MC jets to the data  
 2635 to correct the MC resolution. These uncertainties are evaluated from dijet events and zero  
 2636 bias data using random cones. The uncertainties related to the energy scale (JES) arise  
 2637 from dijet eta-intercalibration across eta,  $Z \rightarrow ee, \mu\mu$ +jets,  $\gamma$ +jet and multijet balance,  
 2638 and the propagation of single particles in high- $p_T$  regimes. The uncertainties due to the  
 2639 mistag rate of the pileup suppression algorithm (mistakenly tagging hard scattering jets)  
 2640 and the impact of the MC generator are also taken into account. It is found that the jet  
 2641 uncertainties are no bigger than 0.1%.

2642 **Jet flavour tagging**

2643 This analysis heavily relies on the identification of  $b$ -jets and  $c$ -jets. All of the employed  
 2644 heavy-flavor tagging algorithms entail systematic uncertainties which need to be con-  
 2645 sidered. The  $b$ -tagging uncertainties are given as a reduced nuisance parameter set con-  
 2646 sisting of scale factor uncertainties for  $b$ -tagging efficiency,  $c$ -jet mistag rate,  $l$ -jet mistag  
 2647 rate and high- $p_T$  extrapolation. Given that the final calculations are not ready yet, a fixed  
 2648 and very conservative 30% uncertainty is considered in every flavour-tagged region of the  
 2649 analysis.

2650 **9.3.2 Theory uncertainties**

2651 The EQ models determined from MC simulations are subject to theoretical uncertainties.  
 2652 Three sources of uncertainty are evaluated, and described in the following.

2653 The effect of missing higher orders in the calculation of the hard interaction is estimated  
 2654 by varying the factorization and renormalization scales for QCD emissions in FSR and  
 2655 ISR by  $\{0.5, 0.625, 0.75, 0.875, 1.0, 1.25, 1.75, 2.0\}$  \*\*\* change to 7?, leading to a total of 20  
 2656 combinations. The scale uncertainty is then approximated by adding in quadrature each  
 2657 one of these variations \*\*\* need to change to envelope?. This corresponds to the biggest  
 2658 theoretical uncertainty, having a magnitude of at most 2%.

2659 The PDF uncertainty is estimated using 100 alternative replicas of the PDF. These are  
 2660 provided by the NNPDF group [149] and encode uncertainties on the measurements en-  
 2661 tering the PDF fit and the fitting method. The PDF uncertainty is given by the standard  
 2662 deviation over the replica sample. In this analysis, it is found to be < 1%.

2663 Finally, \*\*\* finish with alphaS uncertainties, asked about how it's done

2664 Regarding theoretical uncertainties for the QBH samples is less motivated than that for

2665 EQs, given that none of the PDFs includes the effects of gravity and the QCD scale for  
2666 strong gravity is unknown.

2667

2668

## BACKGROUND ESTIMATE

# 10

2669

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

---

Billie Jean King

2670 With the event selection described in Chapter 8, events with fragmentation photons and  
 2671  $t$ -channel events were highly reduced and the signal significance was increased as well,  
 2672 providing an excellent scenario for bump searches. Now, in this chapter in Section 10.1,  
 2673 the background estimation in these signal regions is described, in particular the ones from  
 2674 where a jet fakes a photon. The background in this type of resonance search is modeled  
 2675 with an analytical function, and several statistical tests are needed in order to define it. In  
 2676 Section 10.2, the selection of the optimal function(s) is presented, where all the statistical  
 2677 tests to achieve so are discussed in detail.

2678 **10.1 Background estimation: jet faking photons**

2679 It was discussed in Chapter 6 that the main backgrounds encountered for this analysis  
 2680 are those in which there is at least a photon and a jet in the final state. Although SM  
 2681  $\gamma$ +jet events (prompt photons discussed in Section 1.1.4) is the dominant background, jet  
 2682 faking photons events is another important source of background that need to be taken  
 2683 into account.

2684 Jets can be misidentified as photons (fake photons) if they fluctuate to  $\pi^0$ 's (in a dominant  
 2685 way), resulting in an EM object indistinguishable from a single, real, highly energetic  
 2686 photon (also called prompt photon). To cope with the large jet backgrounds, the Tight  
 2687 identification criteria is applied on photon candidates. This selection is expected to contain  
 2688 prompt photons with moderate jet contamination. As this misidentification rate is not  
 2689 expected to be accurately modeled in MC, a data-driven determination has been used.  
 2690 The Tight offline identification is by design tighter than the photon trigger used to collect  
 2691 the data, so there are quite a few photon candidates from jets that will fail the Tight WP  
 2692 but satisfy some intermediate selection. These photon-like jets, from hereinafter called

2693 pseudo-photons (or Non-Tight), are defined as those passing the Loose identification but  
 2694 failing (at least) one of  $w_1$ ,  $f_{\text{side}}$ ,  $\Delta E$ ,  $E_{\text{ratio}}$  Tight selection cuts [118].

2695 To estimate the number of jet faking photons in the signal regions of the present analysis  
 2696 a combination of two methods is employed. Using the so-called ABCD method with the  
 2697 different expected isolation profiles for both real and fake photons, it is possible to estim-  
 2698 ate fake factors that allow the calculation of the fakes in signal regions [157–160]. The  
 2699 second method, by making use of a sequential template-fit procedure to the photon isol-  
 2700 ation distribution in data and MC, allows to correctly count the number of real and fake  
 2701 photons in the regions delimited by the ABCD method.

Table 10.1: Baseline event selection used for the jet fakes estimation using the template fit procedure.  $p_T^{\text{iso}}$  is defined as shown in Eq. 3.1.\*\*\* **revise selection on jet pt**

|                                   | Selection  |
|-----------------------------------|--|
| Trigger                           | HLT_g140_loose   |
| $N_\gamma$                        | $\geq 1$   |
| $p_T^\gamma$ [GeV]                | $> 150$  |
| $p_T^{\text{jet}}$ [GeV]          | $> 60$   |
| $N_{\text{jets}}$                 | $> 0$  |
| $N_{\text{leptons}}$              | 0  |
| Track isolation                   | $p_T^{\text{iso}} < 0.05$                                      |
| $ \eta^\gamma $ acceptance region | $ \eta^\gamma  < 2.37$ excluding $1.37 <  \eta^\gamma  < 1.52$ |
| $m_{\gamma+j}$ [GeV]              | $m_{\gamma+j} > 500$   |

2702 The study uses Loose identification and non-isolated photons and the event selection tak-  
 2703 ing part in this study is shown in Table 10.1. It is important to notice that the Tight and  
 2704 isolated photons used in the search are only a sub-set of those used in this background-  
 2705 estimation study. By requiring the Loose, non-isolated, photons to pass the selection  
 2706 shown in Table 10.1, Tight identification and  $E_T^{\text{iso}} < 0$  GeV (see Eq. 3.1), one recov-  
 2707 ers the Tight and isolated photons used in the search. Finally, a manual Overlap-Removal  
 2708 procedure is carried out between the photons and jets, to remove jet overlapping with the  
 2709 leading Loose photon if  $\Delta R(\gamma, j) < 0.4$ . This method, as previously mentioned, is data-  
 2710 driven. Given that unblinding of the full Run-2 data is not performed up to the last analysis  
 2711 stage, only the 2015+2016 dataset is used, as it was already unblinded in a previous work  
 2712 by the ATLAS collaboration [161].

### 2713 10.1.1 ABCD method

2714 The ABCD method defines a signal region  $A$  and three control regions, namely  $B$ ,  $C$  and  
 2715  $D$ . These regions are defined by varying the identification status between Tight and  
 2716 Non-Tight, and also by changing the calorimetric isolation requirements (isolated and  
 2717 non-isolated) [162]. The complete definition of the ABCD regions is given as:

- 2718     • Region  $A$ : Tight photons and  $-20 < E_T^{\text{iso}} < 0$  GeV
- 2719     • Region  $B$ : Tight photons and  $8 < E_T^{\text{iso}} < 80$  GeV

- 2720     • Region *C*: Non-Tight photons and  $-20 < E_T^{\text{iso}} < 0$  GeV  
 2721     • Region *D*: Non-Tight photon and  $8 < E_T^{\text{iso}} < 80$  GeV

2722 where  $E_T^{\text{iso}}$  was defined in Eq. 3.1. Figure 10.1 shows the resulting four different regions.

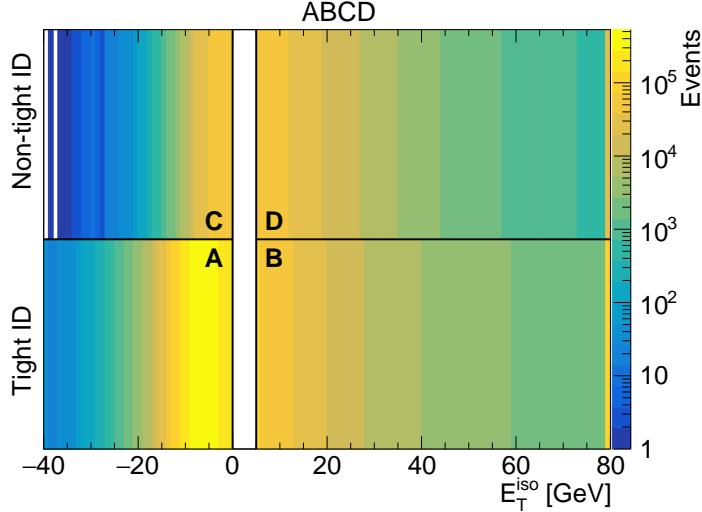


Figure 10.1: Identification vs.  $E_T^{\text{iso}}$  two-dimensional distribution from data events for the Full Run-2 dataset.

2723 Assuming that there is no signal contamination in any control region, the *B*, *C* and *D*  
 2724 regions are only composed of background  $N_{(B,C,D)} = N_{(B,C,D)}^b$ . In addition assuming no  
 2725 correlation between isolation and the considered shape variables, the following relation  
 2726 holds:  $N_B^b/N_A^b = N_D^b/N_C^b$ . Moreover, two different Fake Factors (FFs) could be defined:

$$\text{FF}_{\text{ISO}} = \frac{N_C}{N_D} \quad \text{FF}_{\text{ID}} = \frac{N_B}{N_D}$$

2727 Therefore, the number of jets faking photons can be estimated using the FF as:

$$N_{j \rightarrow \gamma} = N_A = \text{FF}_{\text{ISO}} \times N_B = \text{FF}_{\text{ID}} \times N_C. \quad (10.1)$$

2728 So two different approaches could be used: to model the fake photons using the Tight  
 2729 but non isolated photons from region *B* using  $\text{FF}_{\text{ISO}}$  or model the fake photons using  
 2730 the Non-Tight but isolated photons using the  $\text{FF}_{\text{ID}}$ . Even though both approaches give  
 2731 equivalent results, the  $\text{FF}_{\text{ISO}}$  approach is used as it leads to higher statistics.

2732 Using the FF, now it is possible to estimate the jet-faking photons background contribution  
 2733 in each region of the analysis (*R*). To this end, a jet control region (CRJ-R) is defined  
 2734 equally to the region *R* but replacing the isolation requirements with the one used in  
 2735 region *B*, and weighted by the corresponding  $\text{FF}_{\text{ISO}}$ :

$$N_{j \rightarrow \gamma}^R(p_T) = \text{FF}_{\text{ISO}}(p_T) \cdot N_{\text{CRJ-R}}(p_T)$$

2736 **10.1.2 Corrections to the ABCD method**

2737 Several corrections can be applied to the ABCD method. The first one is to consider the  
 2738 possibility of a signal contamination in any of the control regions  $B$ ,  $C$  or  $D$ . By subtracting  
 2739 the amount of signal events in these regions, Eq. 10.1 becomes:

$$N_{j \rightarrow \gamma} = \frac{N_B - N_B^s}{N_D - N_D^s} \times (N_C - N_C^s) \quad (10.2)$$

2740 where  $N_{(B,C,D)}^s$  is the number of real photons in each region. The estimation of these  
 2741 numbers is a complicated task, since it is needed to have a correct description of the real  
 2742 photons in data, and it is highly contaminated with fake photons. The calculation of the  
 2743 number of real photons in data is performed with an iterative template fit method to the  
 2744 calorimetric isolation distribution in data, explained in more detail below.

2745 The presence of a residual correlation of the background across the four regions, which  
 2746 may manifest itself as a difference in the background distributions for the Tight and  
 2747 Non-Tight regions could be taken into account by calculating:

$$R = \frac{N_A^b N_D^b}{N_B^b N_C^b} \neq 1.$$

2748 However, since  $R$  can not be found in data because that would mean to obtain  $N_A$ , an  
 2749 equivalent parameter is calculated, which can also be written using real photons (leakage  
 2750 photons) subtraction:

$$R' = \frac{N_{A'} N_{D'}}{N_{B'} N_{C'}} = \frac{(N_{A'} - N_{A'}^s)(N_{D'} - N_{D'}^s)}{(N_{B'} - N_{B'}^s)(N_{C'} - N_{C'}^s)}$$

2751 with the definition for each primed region being:

- 2752 • region  $A'$ : Tight photons and  $8 < E_T^{\text{iso}} < 15$  GeV.
- 2753 • region  $B'$ : Tight photons and  $16 < E_T^{\text{iso}} < 80$  GeV.
- 2754 • region  $C'$ : Non-Tight photons and  $8 < E_T^{\text{iso}} < 15$  GeV.
- 2755 • region  $D'$ : Non-Tight photons and  $16 < E_T^{\text{iso}} < 80$  GeV.

2756 The particular selection of 8 GeV aims to define a background only region, but keeping  
 2757 enough statistics to compute the  $R'$  values. In Figure 10.2 the  $R'$  values are shown as a  
 2758 function of  $p_T^\gamma$ . The values are very close to 1 with some exceptions in which the values  
 2759 deviate from 1 by a maximum amount of  $\approx 20\%$  at low  $p_T^\gamma$ .

2760 Finally, taking into account the leakage and possible correlations, the Eq. 10.2 for the  
 2761 expected number of jet fakes results:

$$N_{j \rightarrow \gamma}(p_T) = N_A^b = \left[ R' \frac{N_C - N_C^s}{N_D - N_D^s} \left( 1 - \frac{N_B^s}{N_B} \right) \right] \times N_B = \text{FF}_{\text{ISO}}(p_T) \times N_B(p_T). \quad (10.3)$$

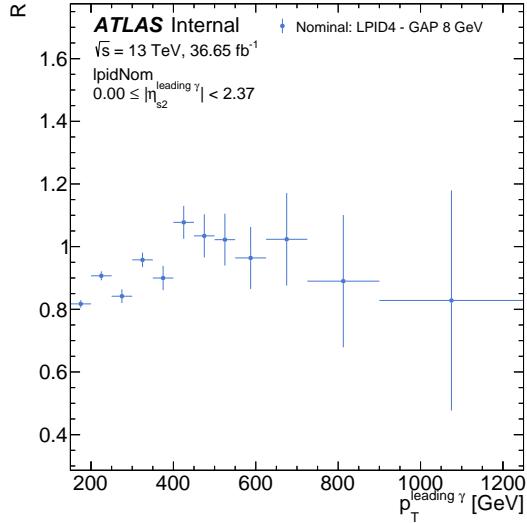


Figure 10.2:  $R'$  values computed as a function of  $p_T^\gamma$ . The error bars shown correspond to the statistical uncertainties.

### 2762 10.1.3 Template fits procedure

2763 In order to estimate the number of jet faking photons in the signal regions of the analysis,  
 2764 it is necessary to have an estimate on the number of events with real photons in the ABCD  
 2765 control regions  $B$ ,  $C$  and  $D$ . To achieve this, a series of fits to the data and real-photon  
 2766 MC photon isolation distribution is carried out for both Tight and Non-Tight photons.  
 2767 The final objective of the iterative fits, is to have real- and fake-photon components fits to  
 2768 the data distribution, that will then be used to compute the number of real photons in the  
 2769 ABCD control regions. The procedure uses the MC PYTHIA samples, which are expected  
 2770 to be real photons, hence the leak photon contribution will be very small. The calculation  
 2771 is performed in 11  $p_T^\gamma$  bins:

$$p_T^\gamma : [150, 200, 250, 300, 350, 400, 450, 500, 550, 625, 725, 900, \infty] \text{ GeV.}$$

2772 The shape of the calorimetric isolation  $E_T^{\text{iso}}$  is fitted in an iterative fashion, using photons  
 2773 passing the Tight and loose-prime Non-Tight identification criteria, as explained above.  
 2774 By definition, events with  $E_T^{\text{iso}} < 0$  GeV pass the calorimetric isolation requirement and  
 2775 correspond to photons falling in region  $A$  if they are Tight, or  $C$  if they are Non-Tight.  
 2776 On the other hand, events with  $E_T^{\text{iso}} > 0$  GeV define regions  $B$  (Tight photons) and  $D$   
 2777 (Non-Tight photons). In what follows, real photons leaked into the Non-Tight regions  
 2778 will be referred as leaked photons. Both Tight and Non-Tight photons will contain a fake  
 2779 component, which will dominate over the leaked photons in the latter case [163, 164].

2780 The sequence of fits proceeds as follows:

- 2781 1. Tight MC photons fit: Given prompt photon samples provides a good description of  
 2782 Tight photons, its  $E_T^{\text{iso}}$  distribution is fitted with a Crystal-Ball (CB) function<sup>1</sup>. It was

<sup>1</sup> This function consists of a Gaussian core but one of the tails of it follows a power-law form. This function was named after the Crystal Ball Collaboration.

2783 found that the simple **CB** description does not accomodate well in the whole range,  
 2784 specially in the bulk of the  $E_T^{\text{iso}}$  distribution, therefore a more flexible function is  
 2785 employed. In this way, an improved version of the **CB** is used, namely the Double-  
 2786 sided Asymmetric Crystal-Ball (**DSACB**) function<sup>2</sup>. Although the description is much  
 2787 better in this case, the Gaussian core struggles to model correctly the peak of the  
 2788 distribution.

- 2789 2. Non-Tight data leak subtraction: Using the leak photon component from **MC**, the  
 2790 fake photon shape is estimated by subtracting the leak photons histogram to data  
 2791 in the whole  $E_T^{\text{iso}}$  range. This provides a very good description of the fake photon  
 2792 component in the Non-Tight region.
- 2793 3. Tight data composite fit: Using the real photons **DSACB** shape estimated in the first  
 2794 step and the fake photons shape estimated in the previous step, a composite fit  
 2795 is performed to the Tight photons  $E_T^{\text{iso}}$  distribution in data. An example of the  
 2796 resulting fit in three different  $p_T$ -bins is shown in Figure 10.3.

2797 The final composite distribution is in good agreement with data, indicating the cor-  
 2798 rect selection of the distributions for each component. The real photons component  
 2799 is the responsible for the high peak at low isolation values, while the fake compo-  
 2800 nents contributes mainly in the range  $0 \text{ GeV} < E_T^{\text{iso}} < 40 \text{ GeV}$ , but having a much  
 2801 smaller yield, as expected. Some differences between the model and data can be  
 2802 seen near the peak of the distribution, as indicate by the fit pull. This difference was  
 2803 also seen in the first step of the calculation when modeling the real photon compo-  
 2804 nents, and it is directly originating from the gaussian modeling of the real photons  
 2805 peak.

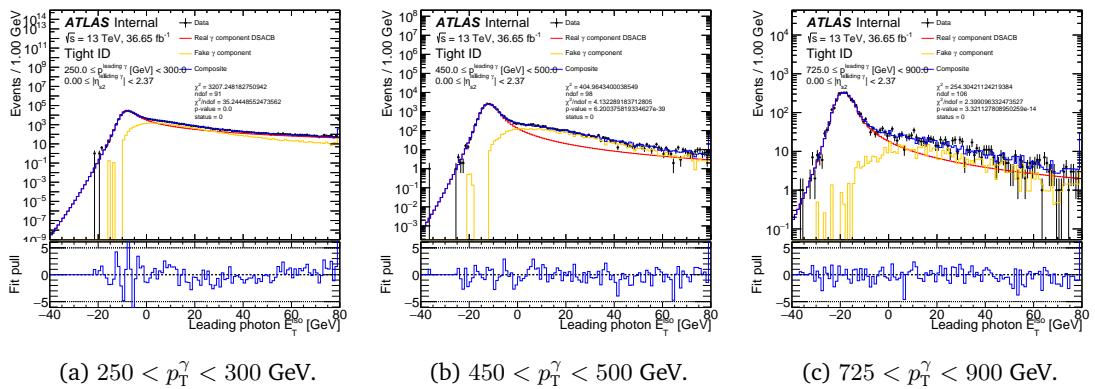


Figure 10.3: Composite fit to data. The red curve represents the real photons component which is represented by a **DSACB** distribution, computed in the first step. The yellow histogram is the fake photon contribution, obtained via leak-photon subtraction in the previous step. The lower pad of the figures represent the normalised residuals (or pull) of the fits.

<sup>2</sup> As it names suggests, in the **DSACB** function, the two tails are modeled by power-law functions, and the core of the Gaussian distribution has two different standard deviations, hence the asymmetry.

2806 **10.1.4 Results**

2807 From the above procedure using the ABCD and the template-fit methods, several key  
 2808 figures can be extracted. One important variable, to gain understanding of the physics  
 2809 process, is the  $\gamma$ +jet purity, calculated as:

$$P_A = \frac{N_{\text{real}\gamma,\text{postfit}}^A}{N_{\text{real}\gamma,\text{postfit}}^A + N_{\text{fake}\gamma,\text{postfit}}^A}.$$

2810 These purities are shown in Figure 10.4a and their numerical values in Table 10.2. As it  
 2811 can be observed, purities of  $> 92\%$  are achieved throughout the whole  $p_T^\gamma$  range, indicating  
 2812 that processes containing a real photon and a jet encompass the majority of the sample. The  
 2813 purity measurements are smoothed using a 3<sup>rd</sup> order spline, shown with the red line in  
 2814 the figure.

Table 10.2: Smoothed photon fake factors  $FF_{\text{ISO}}$  and  $\gamma$ +jet purity as a function of  $p_T^\gamma$ .

| $p_T^{\text{leading}\gamma}$ [GeV] | $FF_{\text{ISO}}$ | Purity real $\gamma$ in A |
|------------------------------------|-------------------|---------------------------|
| 150 – 200                          | 0.1873            | 0.9201                    |
| 200 – 250                          | 0.1885            | 0.9321                    |
| 250 – 300                          | 0.1901            | 0.9418                    |
| 300 – 350                          | 0.1918            | 0.9494                    |
| 350 – 400                          | 0.1934            | 0.9552                    |
| 400 – 450                          | 0.1948            | 0.9593                    |
| 450 – 500                          | 0.1956            | 0.9620                    |
| 500 – 550                          | 0.1956            | 0.9636                    |
| 550 – 625                          | 0.1943            | 0.9642                    |
| 625 – 725                          | 0.1891            | 0.9633                    |
| 725 – 900                          | 0.1703            | 0.9604                    |
| 900 – $\infty$                     | 0.0835            | 0.9650                    |

2815 **FFs**, and particularly  $FF_{\text{ISO}}$ , are applied to data events in a control region CRJ-R, which  
 2816 only differs from any signal region  $R$  in the analysis by requiring non-isolated photons.  
 2817  $FF_{\text{ISO}}$  values, then, can be interpreted as the probability that a jet fakes a photon in region  
 2818  $R$ , given the fake-rate in CRJ-R. The results are shown with the black dots in Figure 10.4b  
 2819 and the total uncertainties with the shaded blue areas. As it can be seen from these results,  
 2820 unstable  $FF_{\text{ISO}}$  value are seen for  $p_T^\gamma < 400$  GeV. A smoothing using a 3<sup>rd</sup> order spline is  
 2821 employed in this case to avoid possible bumps in the final distributions. The numerical  
 2822 values of the  $FF_{\text{ISO}}$  are shown in Table 10.2.

2823 **10.2 Background modeling**

2824 The most challenging task in a resonance search is the correct modeling of the background.  
 2825 As mentioned previously, this analysis only makes use the background simulated samples  
 2826 to optimise the event selection and to select the possible functional forms to model the  
 2827 background. At the same time of selecting the optimal functional form, the range at which

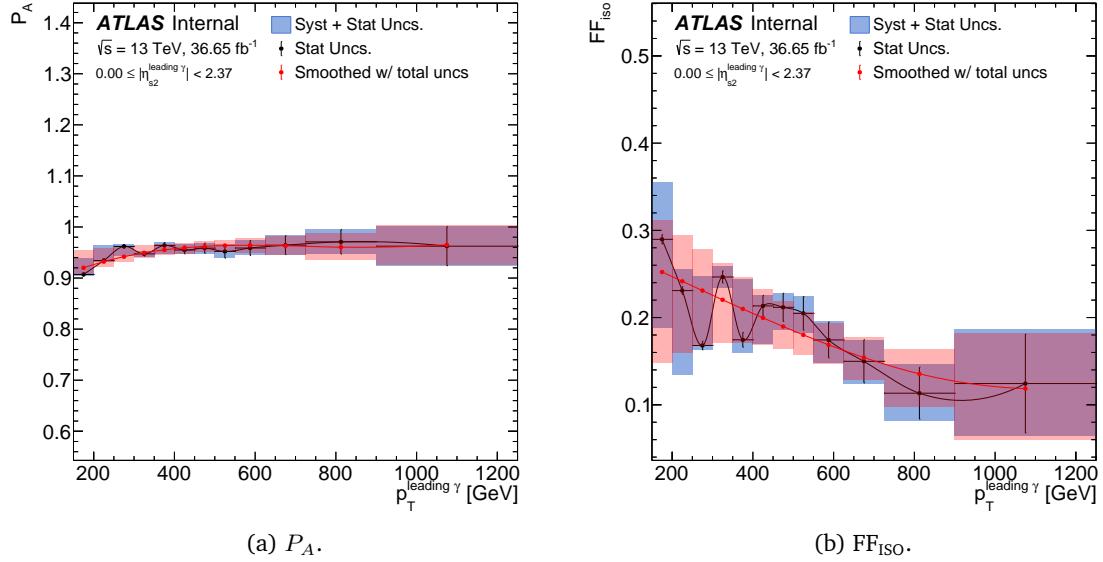


Figure 10.4: Measured  $\gamma$ +jet  $P_A$  values (left) and  $FF_{ISO}$  (right) as a function of  $p_T^\gamma$  obtained using the ABCD method. The measurements are shown in black (statistical uncertainty only), the shaded blue rectangles show the total uncertainty on the measurements (systematic and statistical added in quadrature), and the red points and line show the smoothed measurements with the total uncertainty.

2828 the fits will be performed are selected, to then rank the combinations of functional models  
 2829 and fit-ranges based on the **SS** value. After data unblinding, the background function and  
 2830 fit-range which gives the lowest **SS** is used to fit the data, therefore the exact shape of the  
 2831 background is estimated from data.

### 2832 10.2.1 Fit functions

2833 To model irreducible and reducible backgrounds inclusively, the following family of smoothly-  
 2834 falling functions is used:

$$f_b(x \equiv m_{\gamma+j}/\sqrt{s}) = (1-x)^{p_0} x^{-\sum_{i=1} p_i (\ln x)^{i-1}} \quad (10.4)$$

2835 This family of functions is commonly used in searches of bumps in a smoothly falling  
 2836 background spectrum, such as dijet searches as well as the previous  $\gamma$ +jet search using the  
 2837 partial Run-2 dataset [161, 165]. This family of functions allows to modify the functional  
 2838 form by adding or removing Degree of Freedom (**dof**). There are multiple ways to add  
 2839 additional **dofs**, each of which has a different effect on the fit. The background function  
 2840 will always be scaled by its normalization, although in Eq. 10.4 the parameter is skipped.  
 2841 The final background function reads:

$$f_b(x \equiv m_{\gamma+j}/\sqrt{s}) = n_{\text{bkg}} (1-x)^{p_0} x^{-\sum_{i=1} p_i (\ln x)^{i-1}} \quad (10.5)$$

2842 In what follows, when counting the number of parameters the normalization is included.  
 2843 For example, a function **dof3** will have 3 parameters that control its shape and one con-  
 2844 troling the normalization, hence having a total of 4 parameters.

2845 These functions are tested on MC predictions, in order to determine which functions  
 2846 should be considered for fitting the data distributions. The choice is done with respect  
 2847 to the statistics expected in the  $m_{\gamma+j}$  fit, SSs tests, signal injection tests and F-tests.

### 2848 10.2.2 Datasets preparation

2849 Two different types of samples are used for the background modeling studies: toys and  
 2850 Asimov datasets. These samples are derived directly from the smooth MC background  
 2851 distribution, which consists of  $\gamma+j$  and jet-faking photon events. The strategy to gen-  
 2852 erate these samples is summarized in the flowchart in Figure 10.5 and in the following a  
 2853 detailed description of each step is outlined.

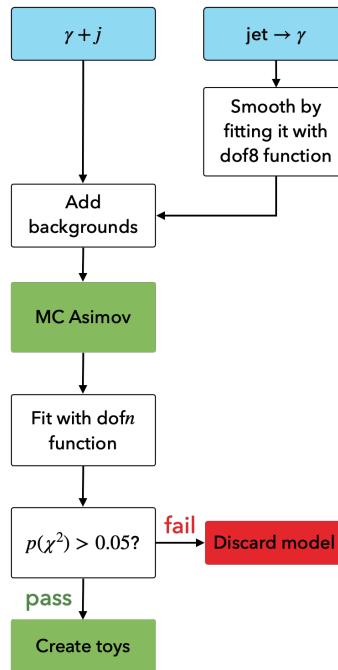


Figure 10.5: Datasets generation for background modeling studies.

### 2854 Smoothing of the jet-faking photon background

2855 It was seen that approximately  $\sim 5\%$  of the  $\gamma+j$  sample is populated by jet faking photon  
 2856 events. This background is estimated directly from data in control regions which fail calor-  
 2857 imetric isolation (as described before), and then weighted by the corresponding FF as a  
 2858 function of  $p_T^\gamma$ . However, specially at very high  $m_{\gamma+j}$ , there are very few but sparse events  
 2859 that when they are added to the other main background ( $\gamma+j$ ), they distort the smooth  
 2860 distribution and artificial bumps start to appear, shown in Figure 10.6. The jet-fakes con-  
 2861 tribution, moreover, is always an order of magnitude less than the  $\gamma+j$  background.

2862 Taking into account these reasons, a smoothing of the jet-fakes background is performed  
 2863 by fitting the  $m_{\gamma+j}$  shape with a *dof8* function that will not be used to model the combined  
 2864 background. The fits are done in the range [500–10000] GeV, to avoid the peak of the  $m_{\gamma+j}$

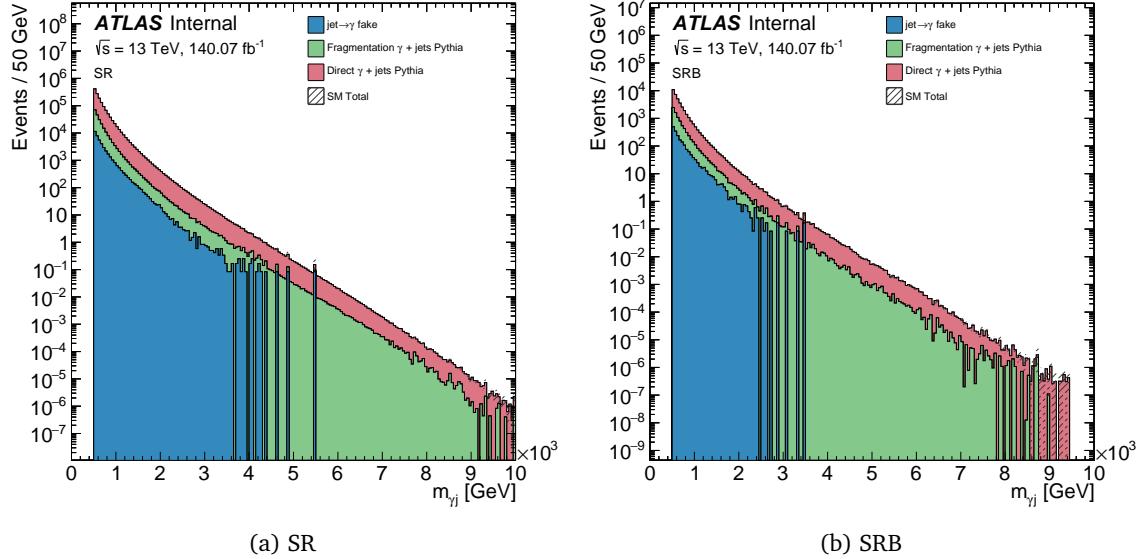


Figure 10.6:  $m_{\gamma+j}$  distribution showing the effect of the sparsed jet-fakes events at high  $m_{\gamma+j}$ , for the SR and SRB regions.

2865 distribution. Examples of these fits for different signal regions are shown in Figure 10.7,  
2866 where in all cases a converged fit result is obtained.

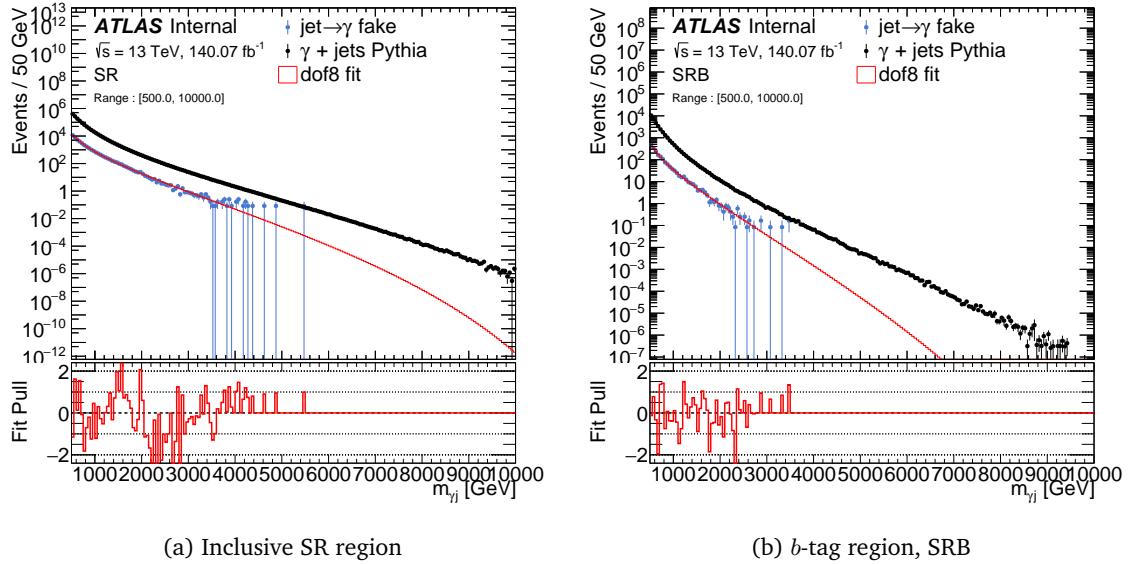


Figure 10.7: Fits to the jet-faking photons backgrounds using the *dof8* functional model for different analysis regions.

2867 Now that the jet-faking photon background contribution is smoothed, the fit is added  
2868 directly to the MC histogram of the  $\gamma+jet$  background.

2869 **Asimov datasets and background only fits**

2870 An Asimov dataset is defined such that when one uses it to evaluate the estimators for all  
 2871 parameters, one obtains the true parameter values:

$$\hat{x} = x_0 \quad (10.6)$$

2872 for all parameters  $x$ , where  $x_0$  is the true value of the parameter. These datasets, are built  
 2873 as binned datasets with very fine binning, in which the event count at each bin is set to  
 2874 the expected event yield.

2875 The PYTHIA  $\gamma$ +jet background benefits from large statistics, making it the best choice for a  
 2876 template of the  $\gamma$ +jet invariant mass spectrum. Therefore, the PYTHIA  $\gamma$ +jet sample, with  
 2877 the addition of the smoothed jet-faking photon distribution, is fitted to create Asimov  
 2878 datasets. First, the bin content at each one of the bins is set to be:

$$n_i = \begin{cases} n_i, & n_i > 0 \\ 0, & n_i \leq 0 \end{cases}, \quad (10.7)$$

2879 and bin errors set to  $\sqrt{n_i}$ , where  $n_i$  is the bin content at bin  $i$ .

2880 Numerous combinations of functional models (number of dof) and fit ranges are defined  
 2881 per signal region. Moreover, in order to produce toy experiments (described below) for  
 2882 each one of the functional models and ranges, it is important to guarantee that a simple  
 2883 **B-only** fit can be achieved with the previously created datasets. To this end, the prepared  
 2884  $m_{\gamma+j}$  distributions are used to create Asimov datasets by fitting them with the different  
 2885 models, while using a big enough range to accommodate for all signals that are going to be  
 2886 used in the analysis in each region.

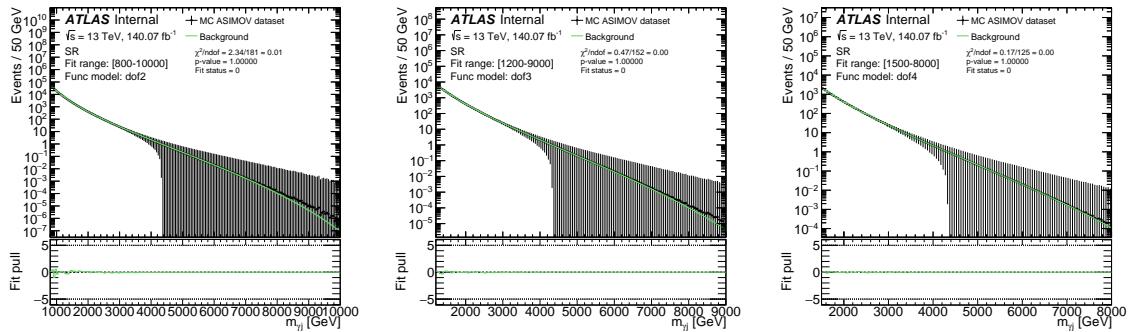


Figure 10.8: Background only fits using different functional models and fit ranges for the inclusive SR region.

2887 Figure 10.8 shows some examples of **B-only** fits done in the inclusive SR. As it can be seen,  
 2888 all functions succeed to represent the background distribution.

2889 **Pseudo-data generation**

2890 Pseudo-data, or also called *toys* distributions, are essentially background distributions used  
 2891 to mimic real data. Toys are computed from **B-only** fits to the Asimov datasets, and each

- 2892 bin of the distribution is computed as a Poisson random number with mean  $v_i = n_i$ ,  
 2893 where  $n_i$  is the fit evaluated at bin  $i$ . Different sets of toys distributions are computed  
 2894 depending on the number of parameters the [B-only](#) fit has been made. A total of 500 toys  
 2895 are generated per signal region and per functional model.
- 2896 To test fits using the  $dofn$  model, pseudo-data distributions generated from a  $dof(n+1)$   
 2897 model are used. In Figure 10.9, two fit examples to toys are shown. In this case, the toys  
 2898 were generated from a  $dof3$  fit to the MC, and are now being fitted with a  $dof2$  model.

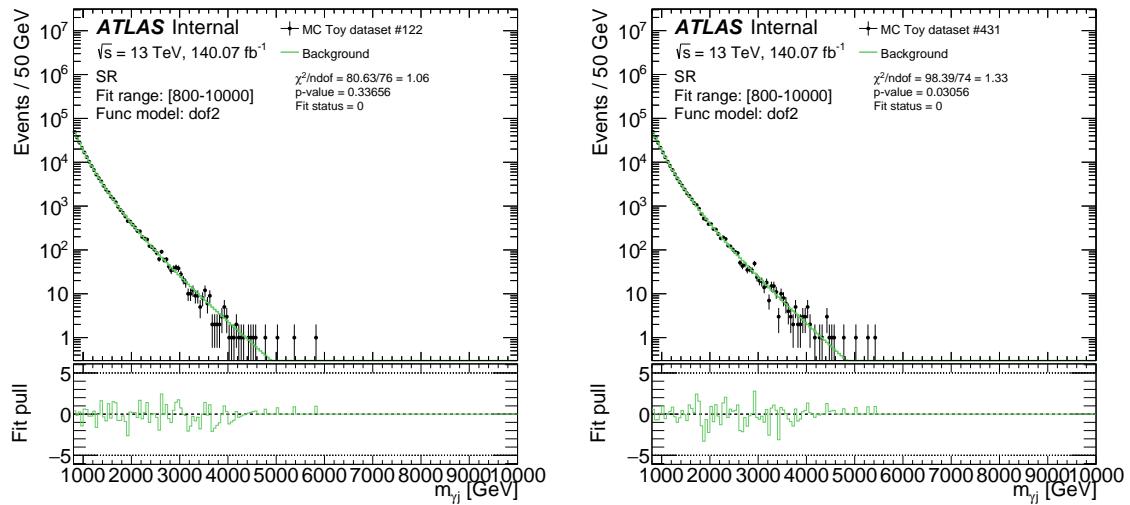


Figure 10.9: [B-only](#) fits to different toys in the SR region

- 2899 As previously stated, toys distributions are made to mimic the behaviour real data would  
 2900 have. In order to have a first guess on the maximum  $m_{\gamma+j}$  value that can be obtained  
 2901 from data, in Figure 10.10 the maximum  $m_{\gamma+j}$  value distribution is shown, for each set of  
 2902 toy generated from the different functional models. For any of the regions dominated by  
 2903  $l$ -jets, the last toy event is approximately at  $\sim 5500$  GeV and the right tails extends up to  
 2904  $\sim 8000$  GeV. Similarly, for the  $c$ -jets and  $b$ -jets regions, there are toy events up to  $\sim 6$  TeV,  
 2905 with the mean at  $\sim 4$  TeV. For these reasons, it is important that the upper limit on the  
 2906 fit-ranges are always higher than the last present event.

### 2907 10.2.3 Modeling strategy

- 2908 The process to select an optimal fit-range and functional model combination is schemat-  
 2909 ised in Figure 10.11. The validation process is carried out with [MC](#), for both, toys and  
 2910 Asimov datasets. As a first step, [B-only](#) fits to the Asimov dataset need to pass the require-  
 2911 ments shown in Figure 10.5 ( $p(\chi^2) > 0.05$ ) and then toys sets are computed. The first test  
 2912 to which the datasets are subjected to is the Spurious Signal ([SS](#)) test. Those models and  
 2913 ranges passing this particular test are then the input to Signal Injection ([SI](#)) tests. In case  
 2914 all these pass, the fit function is retained and will be subjected to another test, discussed  
 2915 below. There are cases in which the functional models do not pass these requiremenets.  
 2916 To treat those cases, the first step is to increase the lower limit of the fit, and the whole

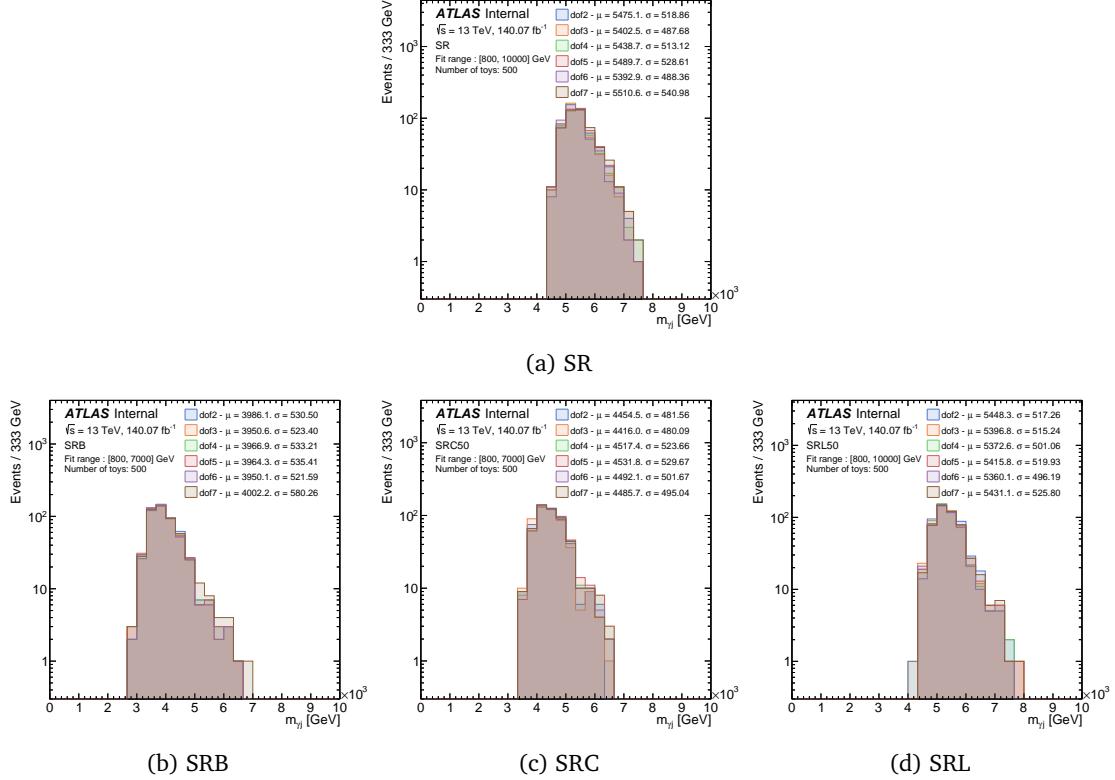


Figure 10.10:  $m_{\gamma+j}$  maximum distribution for the toys sets for each signal region considered. The different sets of toys, corresponding to different initial  $dofn$  functions, are shown with different colors. For each one of them, the mean and the width of the distribution is shown.

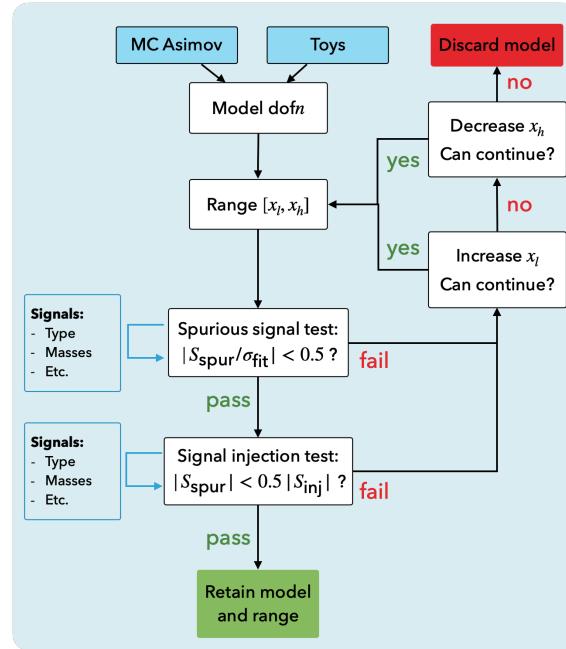


Figure 10.11: Flowchart for the fit function and fit-range validation done using MC samples.

procedure is carried out again. When the lower limit cannot be increased anymore, the higher limit can be decreased, taking into account that all the signal models need to be accommodated. For each one of the tests said above, a complete explanation with results is presented in the following sections.

2921 With the selected functions and fit ranges, the function choice is performed based on  $F$ -  
 2922 test, and a ranking of the fit models and ranges is created based on the  $\sigma_{\text{fit}}$  obtained from  
 2923 **SS** tests, schematized in Figure 10.12

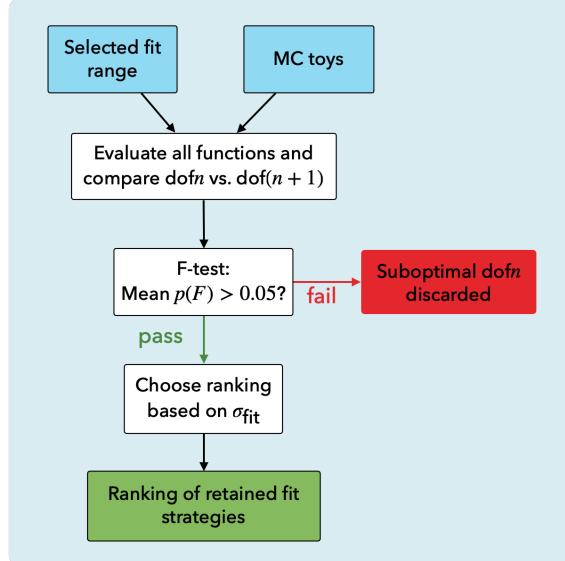


Figure 10.12: Scheme of the fit function selection based on the  $F$ -test.

#### 2924 10.2.4 Fit strategy validation

2925 In the following sections, a detailed discussion of each part of the modeling strategy for  
 2926 the background is presented. As shown in Figure 10.11, the process starts by computing  
 2927 **SS** and **SI** tests, in oder to first gather the possible functions and fit-ranges to use in the  
 2928 final fits to data. These fit-range/fit-function combination pairs are then ranked based on  
 2929 the **SS** value. Finally, the  $F$ -test is used to reorder these combinations.

#### 2930 Spurious signal tests

2931 The fit bias, or Spurious Signal (**SS**), is estimated by fitting a **B-only** distribution with a  
 2932 **S+B** composite model, in which one of the components is the signal template obtained  
 2933 from the interpolated signals (or a gaussian), and the background component being the  
 2934 different functional models evaluated throughout this section. The **SS** ( $S_{\text{spur}}$ ) and its un-  
 2935 certainty ( $\sigma_{\text{fit}}$ ) are obtained from the resulting signal yield ( $N_{\text{sig}}$ ) after the fit to the **B-only**  
 2936 distribution. In an ideal case,  $S_{\text{spur}}$  should approximate to zero, indicating that the func-  
 2937 tional model selected for the background correctly captures the whole distribution. The  
 2938 **SS** can also be computed from fits to the toys distributions. In such cases,  $S_{\text{spur}}$  is the  
 2939 mean value of all the experiments carried out, and  $\sigma_{\text{fit}}$  is computed as the RMS of the  $S_{\text{spur}}$   
 2940 distribution. The two definitions of the **SS** value therefore are:

$$S_{\text{spur}} = \begin{cases} N_{\text{sig}} & \text{if Asimov dataset,} \\ \langle N_{\text{sig}} \rangle_{\text{toys}} & \text{if Toys dataset.} \end{cases} \quad (10.8)$$

2941 The calculation is done for each one of the signal masses for each model, at each one of  
 2942 the signal regions. The **B-only** distribution can either be the asimov dataset or the whole  
 2943 set of toys distributions. In Figure 10.13, two examples of fits to the asimov dataset is  
 2944 shown for the inclusive region, SR. The signal models taken into account are both  $q^*$  with  
 2945  $f = 1.0$  and using masses  $m_{q^*} = 2000$  GeV and  $m_{q^*} = 6000$  GeV. The functional model  
 2946 used to model the background is the *dof2* model and the fit is done in the  $m_{\gamma+j}$  range of  
 2947  $800 - 10000$  GeV.

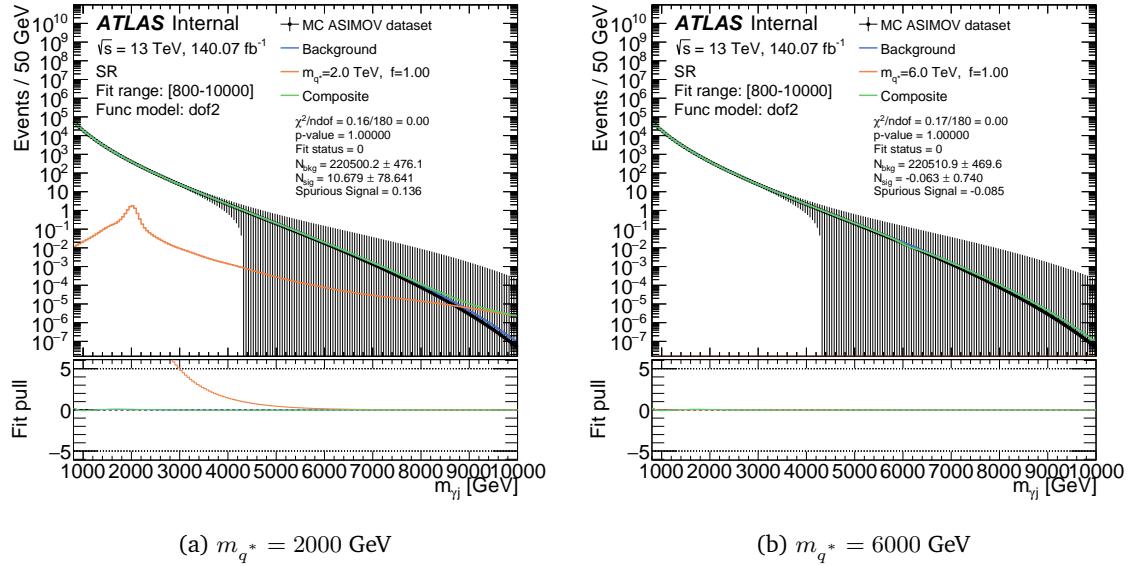


Figure 10.13: **S+B** fit to the Asimov background distribution for the calculation of **SS**. The signal model, with floating normalization, is shown with the orange line, the functional model in this case the *dof2*, represented by the blue line, and the composite fit is shown in green. The figure on the right shows a case in which the **SS** is negative.

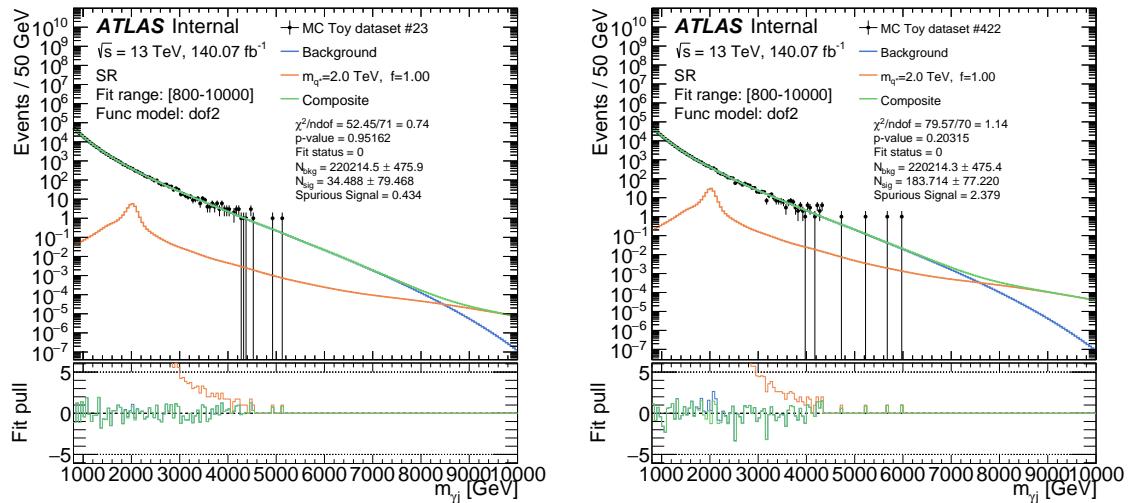


Figure 10.14: Same as Figure 10.13 but using background toys distributions

2948 Fit to toys, on the other hand, are shown in Figure 10.14, in the same conditions as shown  
 2949 in Figure 10.13, only for the  $m_{q^*} = 2000$  GeV model. As mentioned above, the final result  
 2950 of **SS** for the toys is obtained from the  $S_{\text{spur}}$  distribution of all the fits that are successful.

2951 As an example, fits to the inclusive SR using  $q^*$  signals in the range 800 – 10000 GeV  
 2952 and with the function `dof2` are shown in Figure 10.15. From the figures, two different  $q^*$   
 2953 masses are shown: 2000 and 4000 GeV. In the first case, the mean value of the  $S_{\text{spur}}$  is  
 much higher than in the second case, but with higher standard deviation as well.

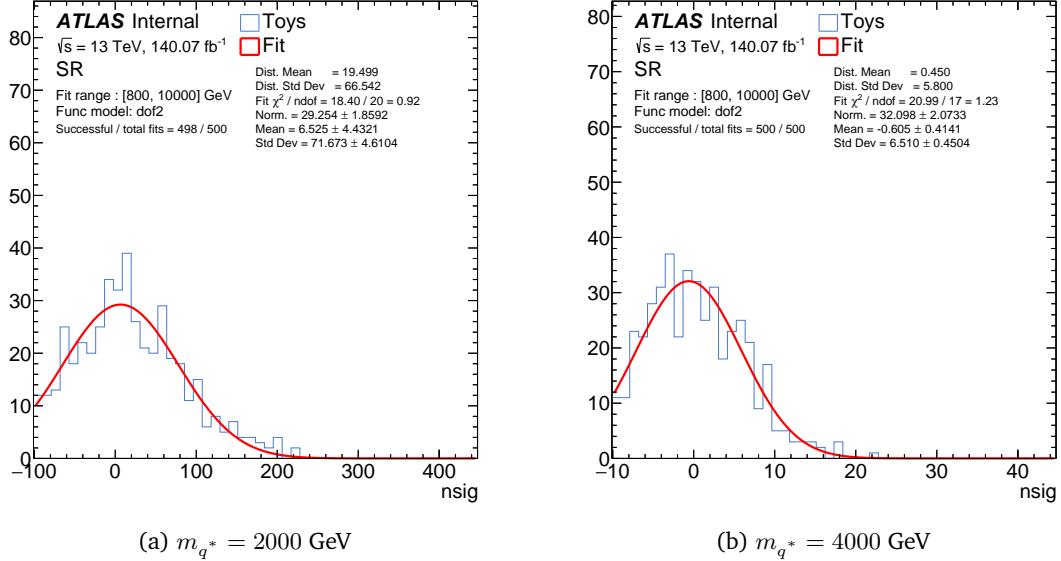


Figure 10.15:  $S_{\text{spur}}$  distribution (blue line) in the SR region, with a gaussian fit to it. The final  $S_{\text{spur}}$  value is directly taken from the distribution, not from the fit. The functional model is `dof2` and the background is fitted within the range [800, 1000] GeV. The `EQ` signal model is used contemplating only  $q^*$  with coupling  $f = 1.0$ .

2954  
 2955 In general, for a model to pass the test, the  $S_{\text{spur}}$  needs to be within the acceptable bounds  
 2956 of:

$$|S_{\text{spur}}| < 0.5\sigma_{\text{fit}}$$

2957 and after that they can be used for further studies. In both cases shown in Figure 10.15,  
 2958 the `SS` test is passed, where the  $S_{\text{spur}}/\sigma_{\text{fit}}$  ratios are  $19.5/66.5 = 0.29$  and  $0.45/5.80 = 0.07$   
 2959 for the  $m_{q^*} = 2000$  and  $m_{q^*} = 4000$  GeV masses, respectively. The main purpose of these  
 2960 tests is to find a group of appropriate functions to be used, as well as the optimal fit  
 2961 range for these functional models. Different lower and upper bounds for the fit ranges are  
 2962 considered depending on the signal region being studied.

2963 Figure 10.16 shows the `SS` for the different considered signal models in the inclusive  
 2964 region SR. The selected results correspond to using the `MC` asimov dataset and the fit-  
 2965 range is the one that yields the lowest `SS`. Each figure shows different functional models  
 2966 (different `dof`). Similarly, `SS` results from toys datasets are shown in Figure 10.17.

2967 It can be noted from the comparison of Figure 10.16 (asimov) and Figure 10.17 (toys),  
 2968 that for toys, the `SS` tests have been carried out for signals with masses up to 5000 GeV.  
 2969 The reason behind this is the absence of events in the  $m_{\gamma+j}$  distributions for  $m_{\gamma+j} \gtrsim$   
 2970 5.5 TeV, therefore a `S+B` fit to zero events lacks of physical meaning. A discussion on the  
 2971 maximum  $m_{\gamma+j}$  distribution was presented in Section 10.2.2, where the maximum  $m_{\gamma+j}$   
 2972 value for each signal region was shown.

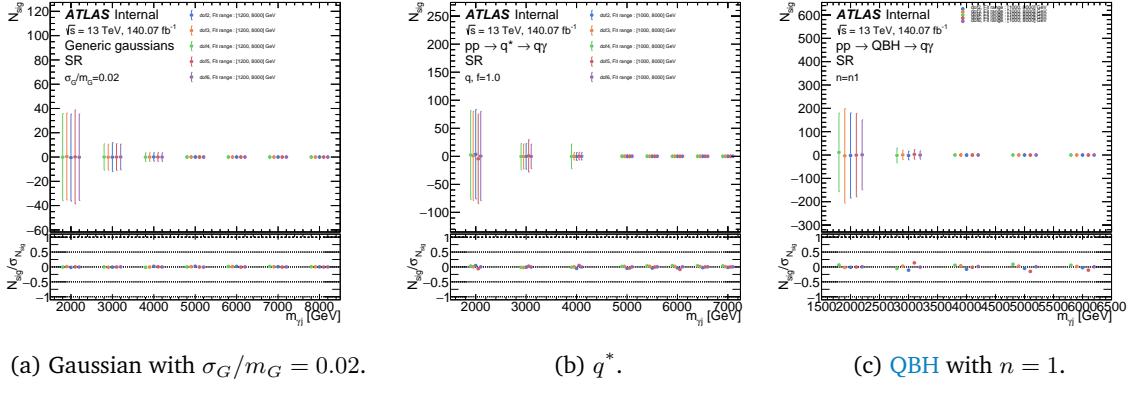


Figure 10.16: **SS** tests results for the inclusive SR region. The different figures correspond to the fit-range that yields the lowest **SS** for the given signal model and function's **dof**. Different functional forms are shown with different colors.

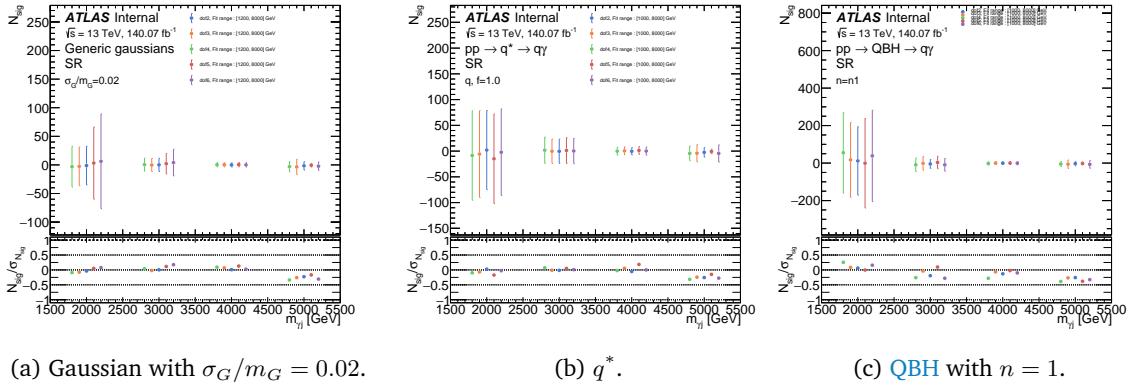


Figure 10.17: Idem Figure 10.16 but using toy samples.

2973 Due to the two methodologies (toys and asimov datasets) used to compute the **SS** test, res-  
 2974 ults from both calculations are combined giving preference to toys, as they are calculated  
 2975 in a more robust form. The combination is performed as follows:

- 2976 1. Use toys **SS** results up to the mass point where it is expected to have a considerable  
 2977 amount of background events. The maximum mass point considered for all those  
 2978 regions that are dominated by  $l$ -jets ( $b$ -jets and  $c$ -jets) is  $< 5$  TeV ( $< 4$  TeV). Toy **SS**  
 2979 results therefore only contribute on the low mass regime.
- 2980 2. Asimov **SS** results are used for the rest of the mass range, hence contributing only to  
 2981 the high mass region of the spectrum.
- 2982 The **SS**, or fit bias, is used in the final limit calculation as one of the systematic uncer-  
 2983 tainties associated with the background modeling, therefore it is then desired that this  
 2984 uncertainty is as small as possible. For each signal region and signal model, after the pre-  
 2985 viously mentioned combination is performed, a ranking of the fit-ranges and functional  
 2986 models is created, based on the average absolute **SS**. This quantity indicates which com-  
 2987 bination of fit-range and functional model provides the lowest **SS** in average for each mass  
 2988 point, giving equal importance to each mass.

2989 **Signal injection tests**

2990 A Signal Injection (**SI**) test probes the ability of a fit strategy to correctly determine the  
 2991 amount of signal present in a dataset. It is performed similarly to the **SS** test, with the  
 2992 difference that a signal of the expected hypothesis is injected with a certain amplitude  
 2993 on top of the **B-only** pseudo-data or toys. While the definition of  $S_{\text{spur}}$  for the **SS** tests  
 2994 was simply the number of extracted signal events, shown by Eq. 10.8, in these tests it is  
 2995 generalised such that  $S_{\text{spur}}$  is now the difference between the extracted and the injected  
 2996 signals:

$$S_{\text{spur}} = \begin{cases} N_{\text{fit}} - N_{\text{inj}} & \text{if Asimov dataset,} \\ \langle N_{\text{fit}} - N_{\text{inj}} \rangle_{\text{toys}} & \text{if Toys dataset.} \end{cases} \quad (10.9)$$

2997 The amplitude of the injection is denoted in units of  $\sqrt{B}$ , which takes into account the  
 2998 square root of the number of background events in the Full Width at Half Maximum  
 2999 (**FWHM**) range of the signal in question, corrected by the number of signal events in the  
 3000 range:

$$\sqrt{B} \equiv \frac{\sqrt{\sum_{\substack{\text{bins in FWHM}}} b_i}}{\sum_{\substack{\text{bins in FWHM}}} s_i}$$

3001 The inclusion of this correction aims to have the reported amplitudes in units of  $\sqrt{B}$  to  
 3002 always correspond to  $S/\sqrt{B}$  ratios.

3003 To evaluate the **SI** tests, 300 toy experiments are run per signal hypothesis, per signal  
 3004 region and injection amplitude, in the same ranges as in the **SS** tests, and using different  
 3005 functional models. The recommended criterion for passing the signal injection tests is  
 3006  $|S_{\text{spur}}| < 0.5N_{\text{inj}}$ . In Figure 10.18, examples of the test for the **EQ** signal model are shown.  
 3007 The upper pad of the figures show, as a function of  $N_{\text{sig}}^{\text{inj}}/\sqrt{B}$ , the extracted signal in the  
 3008 **FWHM** range  $N_{\text{sig}}^{\text{fit}}/\sqrt{B}$  for different masses. The bottom pad of the plots show the  $S_{\text{spur}}$   
 3009 calculated as in Eq. 10.9, divided by the injected signal amplitude  $N_{\text{inj}}$ . In all cases, the  
 3010 **SI** test is passed as the ratio is  $< 0.5$ . Moreover, it can be seen that the results show  
 3011 very good linearity. Results for the other signal models are shown in Appendix B where  
 3012 good linearity is seen for all cases. \*\*\* revise definition of ratio. Should it be with the  
 3013 uncertainty, or with the injected signal?

3014 **F-tests**

3015 Once the optimal fit-range and a function ranking has been set by virtue of the **SS** tests,  
 3016 the final statistical test to select the function is the *F*-test. The test compares fits done  
 3017 with a function *a* with  $p_a = n$  parameters with another function *b* which has  $p_b = n + 1$   
 3018 parameters. For the *F*-test, model *a* is a subset of model *b*. In this test, a test statistic *F* is

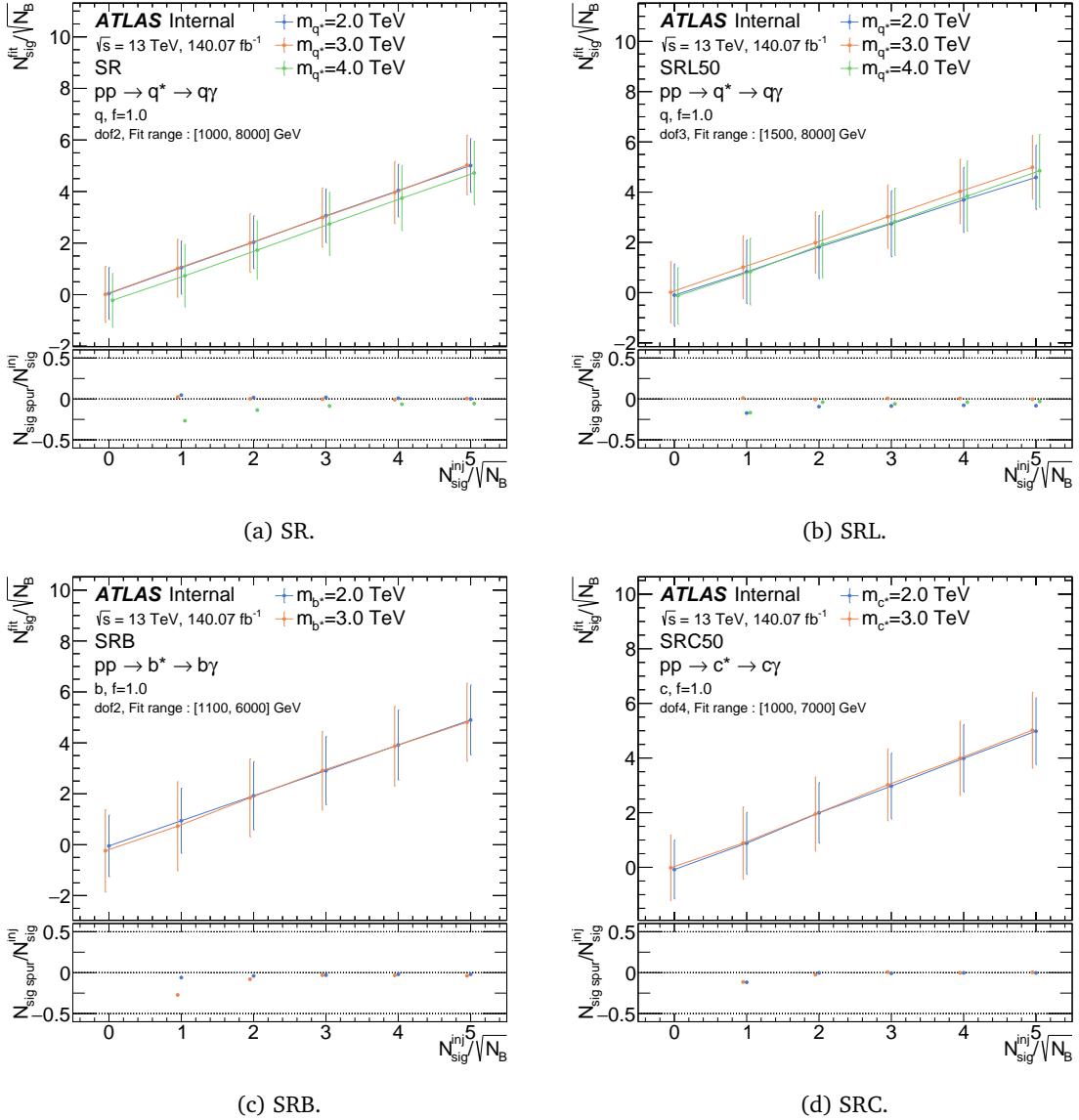


Figure 10.18: SI tests for the EQ model in regions SR, SRL, SRB and SRC, respectively. The fit-range and functional model used are the ones that yield the lowest  $S_{\text{spur}}$ , taken from the SS tests, and are displayed in each plot. Each considered mass is shown with a different color. The  $x$ -axis shows the injected amplitude in units of  $\sqrt{N_B}$ , while the  $y$ -axis represents the extracted signal in terms of  $\sqrt{N_B}$ .

3019 computed from the resulting  $\chi^2$  values::

$$F = \frac{\frac{\chi_a^2 - \chi_b^2}{p_b - p_a}}{\frac{\chi_b^2}{N - p_b}}, \quad (10.10)$$

- 3020 where  $N$  is the total number of bins in the sample. High values of  $F$  mean that the model  
 3021 with  $n$  parameters should be discarded, while cases in which  $F \rightarrow 0$  ( $\chi_a^2 - \chi_b^2 \rightarrow 0$ ),  
 3022 the difference between the models is not significant, leading to select the model with the  
 3023 lowest amount of parameters, i.e model  $a$  with  $n$  parameters.  
 3024 The null hypothesis is defined as the one in which model  $b$  does not provide a better  
 3025 significant difference compared to model  $a$  (small  $F$ ), and in this situation,  $F$  will have an

3026  $F$ -distribution with  $(p_b - p_a, N - p_b)$  dof. The null hypothesis is rejected if the  $F$ -score is  
 3027 higher than a critical value, usually set to that leading to a p-value  $p(F(p_b - p_a, N - p_b)) <$   
 3028 0.05. In summary, if the p-value of comparing model  $a$  and  $b$  is  $> 0.05$ , the null hypothesis  
 3029 is not rejected, and the two models are considered similar, while p-values  $< 0.05$  mean  
 3030 that there is evidence against the null hypothesis, and the more simple model  $a$  is rejected.  
 3031 For the study of selecting the best fit function, model  $a$  is the nominal function with dof $n$   
 3032 and model  $b$  is the alternate function dof( $n+1$ ).  
 3033 Fits are done to toys drawn from the dof7 fit to the B-only MC Asimov distribution. The  
 3034 selected fit ranges for these studies are the ones decided based on the previous SS test.  
 3035 Model  $a$  and  $b$  are fitted to the same toy, and they are compared if and only if both of the  
 3036 fits converge. For each one of the toys, the  $F$ -value is calculated according to Eq. 10.10,  
 3037 and finally, the p-value  $p_0$ .

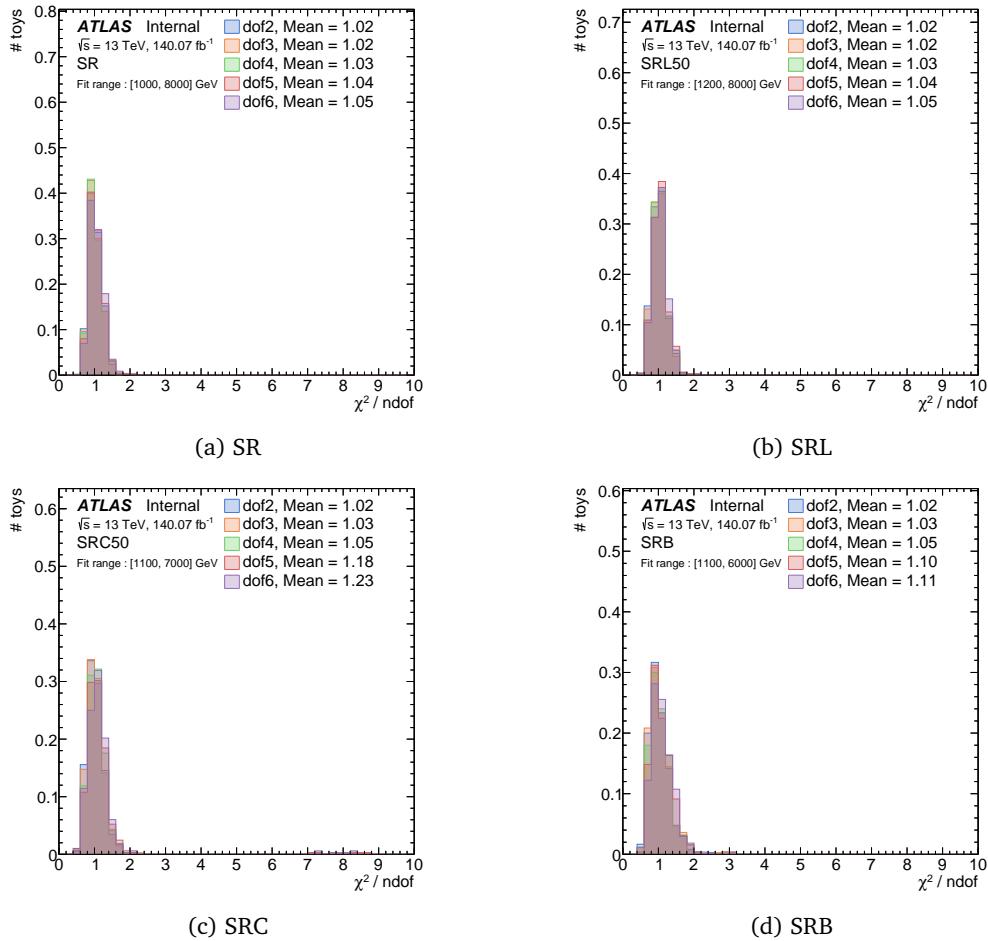


Figure 10.19:  $\chi^2/\text{ndof}$  distribution for each functional model in different signal regions.

3038 Figure 10.19 shows the  $\chi^2/\text{ndof}$  distributions for each one of the models for different  
 3039 signal regions computed from toys. It can be seen that all the distributions are centered  
 3040 at  $\sim 1$ , meaning that in general all models describe the B-only distribution correctly. To  
 3041 further compare the different models among them, the  $F$ -score distributions are plotted  
 3042 in Figure 10.20 for different signal regions, comparing two models at a time. For each

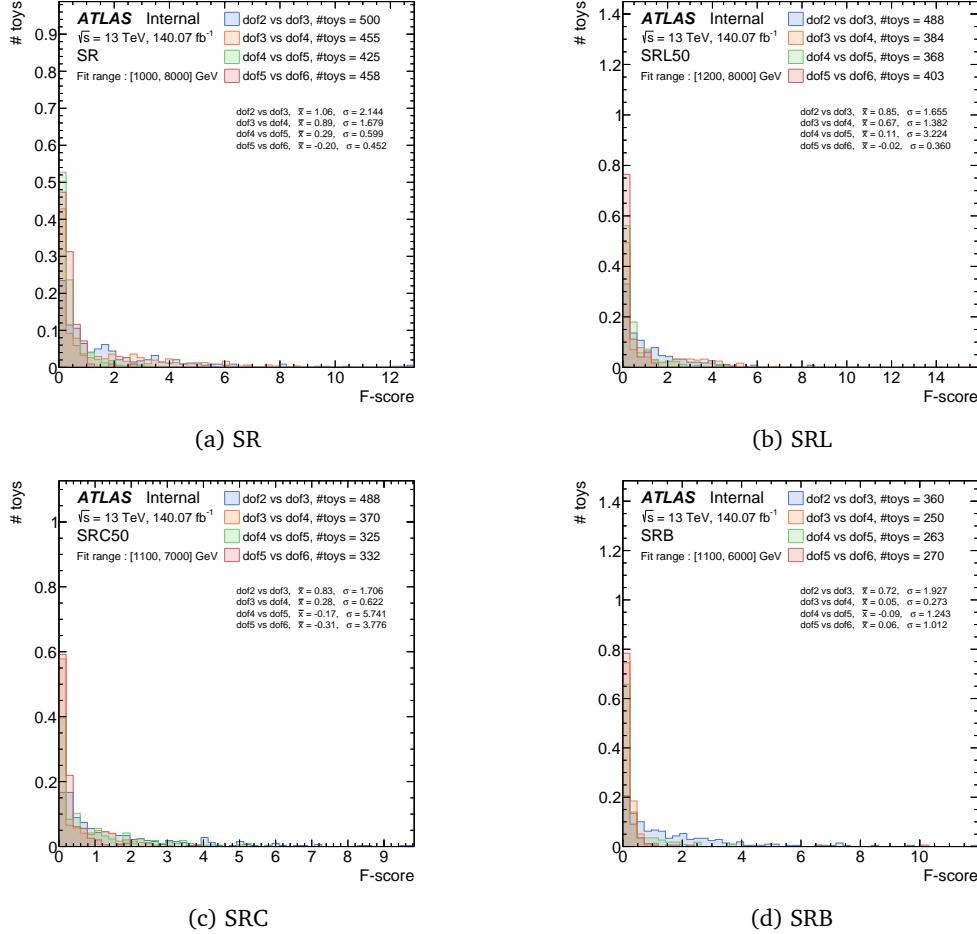


Figure 10.20:  $F$ -score distribution. The tests are done by comparing two functional models at the same time, shown by each of the filled histograms. For each one of them, the number of matching and converged, toys are shown, as are the means and widths of the  $F$ -score distributions.

pair, the number of toys that converge for both models is shown, as well as the means and widths of the distributions. Likewise, in Figure 10.21 the p-value distribution for the  $F$ -test can be observed, which shows that all the comparison between models lead to  $p(F)$  values very close to 1, meaning that no model fits the distributions better than another. For this reason, no functional model is discarded using the  $F$ -tests.

### 10.2.5 Summary of modeling strategies

Throughout this section different statistical tests have been carried out to determine the optimal fit-range and functional form to model the background distribution in data. These tests have been done for each one of the signal regions considered in the analysis, and also for each signal model.

In the first step, SS tests were performed in order to decide which of these ranges and functions combinations minimise the appearance of any signal when doing B-only fits. Moreover, the background modeling uncertainty results from this test, meaning that using the combination that lead to the lowest SS, will also influence the final uncertainties when

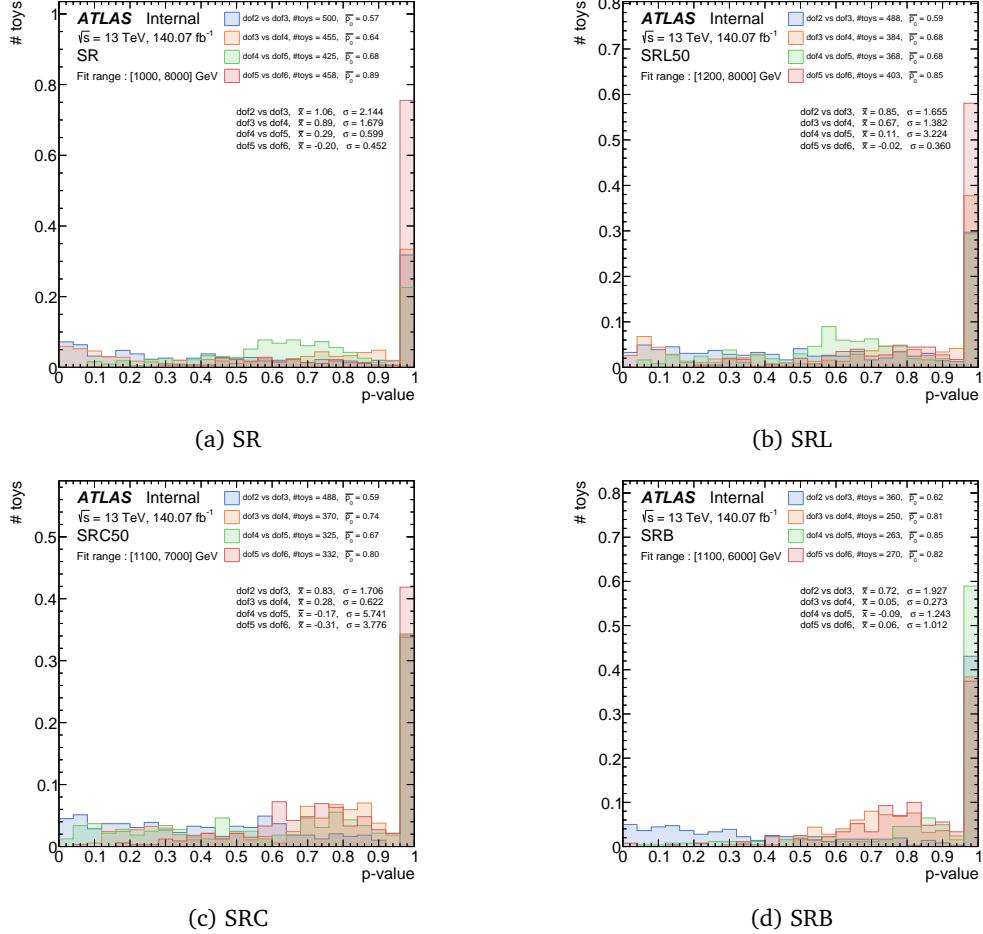


Figure 10.21: p-value score distribution comparing the performance of the different pair of functions. The reported means and widths correspond to the  $F$ -score distributions, and for each p-value distribution, the average p-value is shown.

3057 performing the search on data. For each fit-range, rankings of the functional models are  
 3058 built and are all passed through the **SI** and  **$F$** -tests. The former checks if the function is  
 3059 able to capture all the injected signal to the background, and the second tests if by adding  
 3060 another **dof** to the function a significant improvement on the fits is seen.

3061 As discussed in Section 6.3, two types of interpretations are studied. In the **B-only** in-  
 3062 terpretation, the fits are done to data with a **B-only** model, that is, the signal strength in  
 3063 Eq. 6.2 is fixed at 0 and all the nuisance parameters are fitted. These fits are performed  
 3064 in regions SR, SRB and SRC, using the functions and ranges given in Table 10.3. On the  
 3065 other hand, for the **S+B** interpretation the signal strength  $\mu$  is allowed to float and the  
 3066 yielded signal is quantified. The fits to data are then performed with a **S+B** model, where  
 3067 the signal component is a **PDF** and the background function differs for each signal model  
 3068 and signal region. In Table 10.3, a summary of the functional forms and the fit-ranges for  
 3069 each signal model and region are displayed.

Table 10.3: Summary of fit strategies for each analysis region and signal model. The table shows the functional form used, as well as the  $m_{\gamma+j}$  fit-range in GeV. The last column indicates if the fit is performed simultaneously for the SRC, SRB and SRL regions, in case each one of them are used for the interpretation/model.

|                   | Inclusive SR               | SRL                        | SRC                        | SRB                        | Do simult.<br>SRC+SRB+SRL? |
|-------------------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
| <b>B-only</b>     | <i>dof2</i> , [1000, 8000] | <i>dof2</i> , [1200, 8000] | <i>dof4</i> , [1100, 7000] | <i>dof2</i> , [1100, 6000] | No                         |
| Gaussians         | <i>dof3</i> , [1200, 8000] | <i>dof3</i> , [1500, 8000] | <i>dof4</i> , [1000, 7000] | <i>dof5</i> , [900, 7000]  | No                         |
| <b>QBH</b>        | <i>dof5</i> , [1000, 8000] | -                          | -                          | -                          | -                          |
| <b>EQ</b> , $q^*$ | <i>dof2</i> , [1000, 8000] | -                          | -                          | -                          | -                          |
| <b>EQ</b> , $c^*$ | -                          | <i>dof2</i> , [1200, 8000] | <i>dof4</i> , [1100, 7000] | <i>dof2</i> , [1100, 6000] | Yes                        |
| <b>EQ</b> , $b^*$ | -                          | -                          | -                          | <i>dof2</i> , [1100, 6000] | -                          |

3070

3071 **RESULTS**

# 11

3072

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

---

Billie Jean King

3073 It was introduced in Section 6.1, in a brief way, the two different types of fits that can be  
 3074 performed to the recollected ATLAS Run-2 data. Moreover, the numerous fitting strategies  
 3075 that need to be applied for each signal and in each signal region were derived and dis-  
 3076 cussed in Section 10.2.

3077 In this chapter, the results obtained using the full Run-2 dataset are presented. First, in  
 3078 Section 11.1, in order to correctly present the results and to compute the fit p-values, the  
 3079 optimal binning is derived using MC simulation. Finally, in Section 11.2, the different  
 3080 fitting strategies are applied based on the B-only and S+B interpretations.

3081 **11.1 Photon+jet invariant mass binning**

3082 The  $\gamma$ +jet invariant mass,  $m_{\gamma+j}$ , is the observable in the current analysis. In order to  
 3083 correctly display the results, a suitable binning needs to be chosen, being not too wide  
 3084 where bumps might not be visible, but also not too narrow that there is bin migration,  
 3085 due to the detector resolution on the mass.

3086 The binning should be wider than the detector resolution. In order to estimate the detector  
 3087 resolution, the PYTHIA  $\gamma$ +jet MC simulation normalised to the 2015+2106 dataset is used.  
 3088 As a first step, the ratios  $m_{\gamma+j}^{reco}/m_{\gamma+j}^{truth}$  are computed in bins of  $m_{\gamma+j}$  and for each bin a  
 3089 gaussian function  $g(\mu, \sigma)$  is fitted to the ratio, as shown in Figure 11.1, and the ratio  
 3090  $\sigma/\mu$  corresponds to the resolution. The resulting set of resolution values are plotted as a  
 3091 function of  $m_{\gamma+j}$  and a fit to it of the form

$$\frac{\sigma}{m_{\gamma+j}} = \sqrt{\frac{a}{m_{\gamma+j}} + \left(\frac{b}{m_{\gamma+j}}\right)^2 + c} \quad (11.1)$$

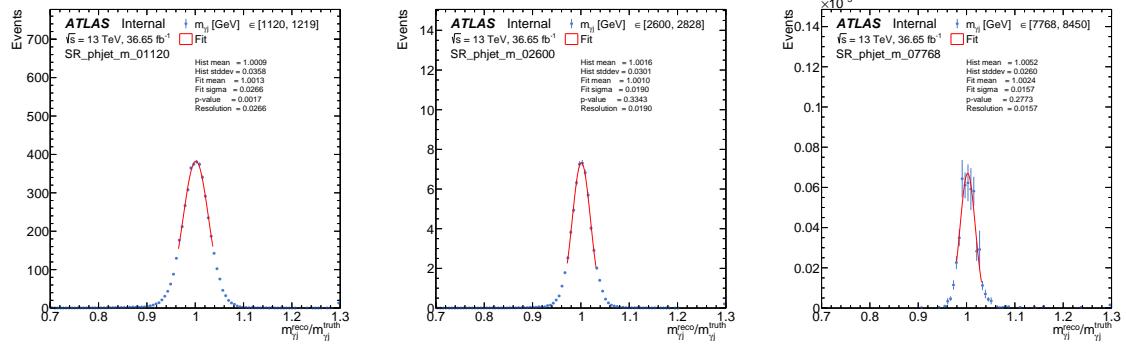


Figure 11.1: Selected  $m_{\gamma+j}^{\text{reco}}/m_{\gamma+j}^{\text{truth}}$  fits for the  $m_{\gamma+j}$  binning studies. The distribution is shown with the blue dots while the gaussian fit with the red line. In all cases, the mean and standard deviation of the gaussian post-fit is shown.

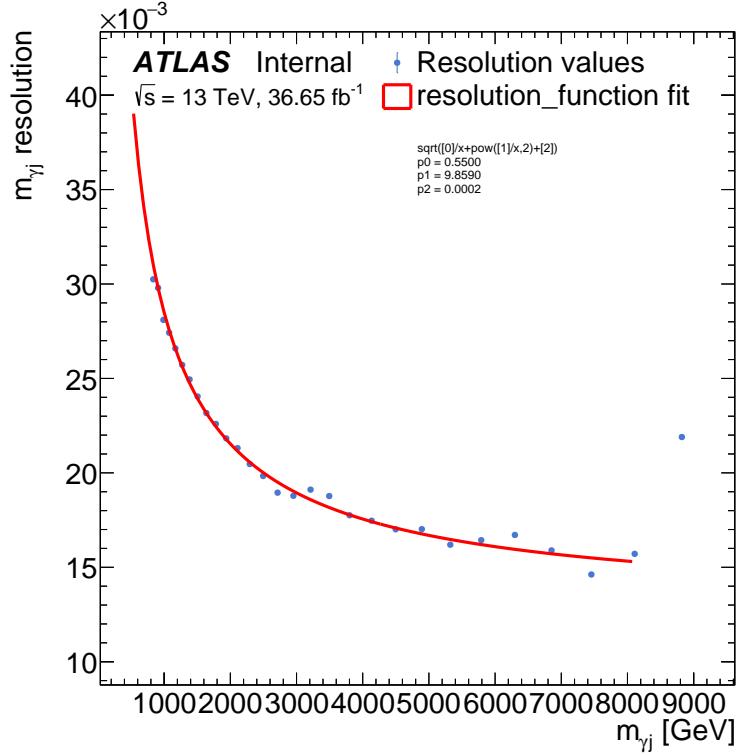


Figure 11.2:  $m_{\gamma+j}$  resolution values and fit using Eq. 11.1. The post-fit parameters are indicated in the figure where  $p_0 \equiv a$ ,  $p_1 \equiv b$  and  $p_2 \equiv c$ . The fit has been performed excluding the last measured resolution value due to lack of statistics.

3092 is performed, shown in Figure 11.2. Finally, the  $m_{\gamma+j}$  binning is calculated, starting at  
 3093 500 GeV up to 10 TeV, in an iterative fashion so that the bin-width is compatible with the  
 3094 resolution. The obtained bin-width as a function of  $m_{\gamma+j}$  is shown in Figure 11.3a, which  
 3095 shows a monotonically increasing nature, as expected. Finally, to validate that the binning  
 3096 resolution is in agreement with the detector one, in Figure 11.3b, a comparison between  
 3097 two is shown, from which excellent agreement is observed.

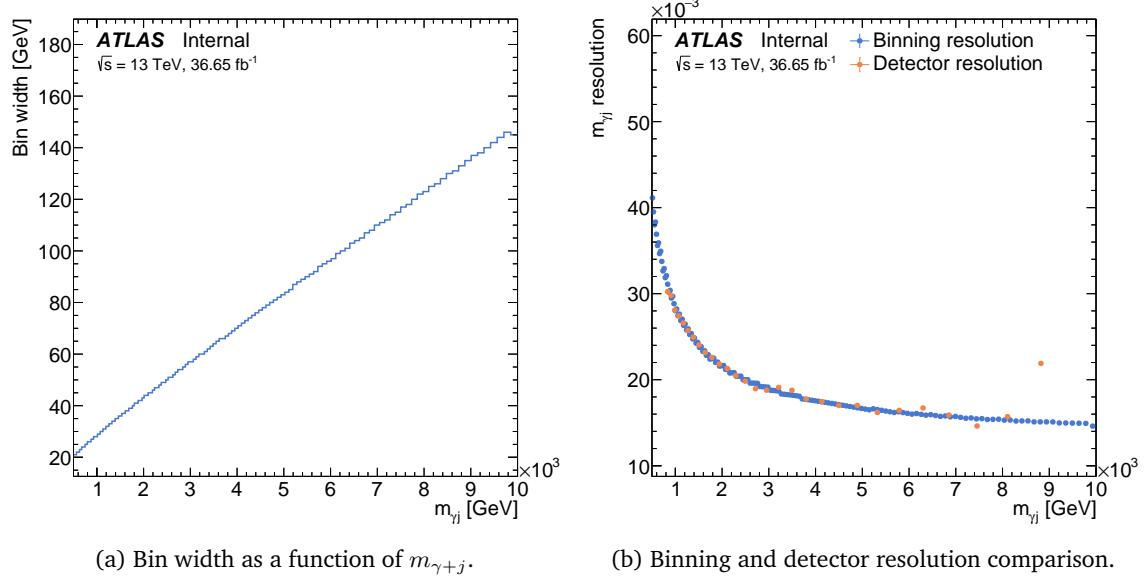


Figure 11.3:  $m_{\gamma+j}$  binning optimisation results showing the bin-widths as a function of  $m_{\gamma+j}$  (left) and the comparison between the binning resolution and detector resolution (right).

## 3098 11.2 Results

3099 Now that the fitting strategies are defined for each one of the signal regions and signal  
 3100 models considered (see Table 10.3), the fits are now applied to the observed  $m_{\gamma+j}$  in data.  
 3101 Two different types of interpretations are studied:

- 3102 • **Background-only (B-only) interpretation:** A **B-only** fit is performed to data in  
 3103 each one of the signal regions, in order to test the compatibility of the data with a  
 3104 smoothly falling spectrum. In case any excess is found, it is quantified by virtue of  
 3105 the BumpHunter algorithm previously discussed.
- 3106 • **Signal interpretation:** The compatibility of the data with three different signal mod-  
 3107 els is studied. One of the models correspond to generic Gaussian shapes, which al-  
 3108 low to interpret the results in a more general way. The other two benchmark models  
 3109 correspond to **QBHs** and **EQs** models, where their resonance shapes are different  
 3110 between them, and also for different parameters of the theories. In case no signifi-  
 3111 cant excess is found with the **B-only** interpretation, exclusion limits on the parameter  
 3112 values of these theories are derived.

### 3113 11.2.1 Background-only interpretation

3114 The fit strategies derived in the previous chapter are now applied to the observed  $m_{\gamma+j}$   
 3115 distribution in data. This allows to perform an interpretation of the invariant mass dis-  
 3116 tribution without any assumption on the potential signal other than it being a localised  
 3117 effect. However, the consistency of the observed  $\gamma$ +jet mass spectrum with a smoothly  
 3118 falling **SM** background can not be quantified with the  $p(\chi^2)$ -value alone. This, does not

3119 provide the optimal sensitivity to a localized resonance since it considers all  $m_{\gamma+j}$  bins  
 3120 simultaneously and independently of each other. A true resonance, on the other hand,  
 3121 would likely result in a correlated excess in one or several adjacent  $m_{\gamma+j}$  bins, and here is  
 3122 when the BumpHunter algorithm comes to place.

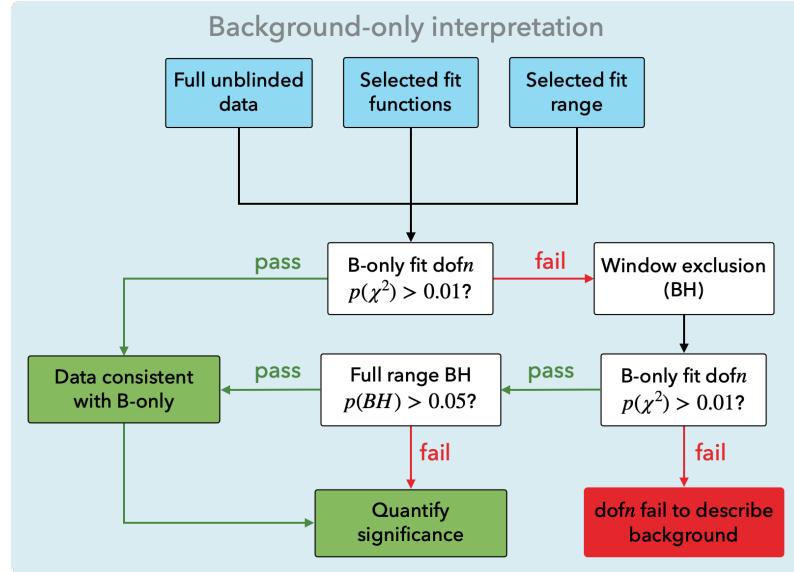


Figure 11.4: Scheme of the **B-only** fit workflow to data.

3123 The **B-only** fit workflow is shown in Figure 11.4. For each signal region, the first step is  
 3124 to perform a **B-only** fit to observed  $m_{\gamma+j}$  spectrum. If the  $\chi^2$   $p$ -value of this fit exceeds a  
 3125 threshold of 0.01, the **B-only** hypothesis succeeds to describe the data and BumpHunter  
 3126 is run to identify the most significant deviation and assign a  $p$ -value to it. On the other  
 3127 hand, if the  $p(\chi^2)$ -threshold is not passed, that can either hint at the fit not being able  
 3128 to globally describe the data spectrum or at one or more localized deviations. In order  
 3129 to test the latter case, BumpHunter is executed to identify the most significant deviation.  
 3130 This region is subsequently masked and the **B-only** fit is repeated and the  $p(\chi^2)$ -value is  
 3131 checked again. If it does not pass the threshold, a single local effect cannot explain the  
 3132 bad fit and no clear interpretation of the observed spectrum can be given. If on the other  
 3133 hand the threshold is passed, this means that indeed the local effect was preventing a  
 3134 good fit. The BumpHunter is run again and a  $p$ -value is assigned to this most significant  
 3135 deviation. A  $p$ -value larger than 0.05 is then interpreted as no significant BumpHunter  
 3136 discovery. If on the other hand this threshold is not exceeded, a local effect is assumed  
 3137 and the significance of hypothetical signal is evaluated.

## 3138 Results

3139 The resulting fits in each signal region are shown in Figure 11.5. It can be seen that no  
 3140 significant excess is found in neither of the regions. The lower pad of each one of the  
 3141 figures shows the normalised residues of the fit, where differences between data and the  
 3142 **B-only** hypothesis can be more easily identified. The major differences observed between

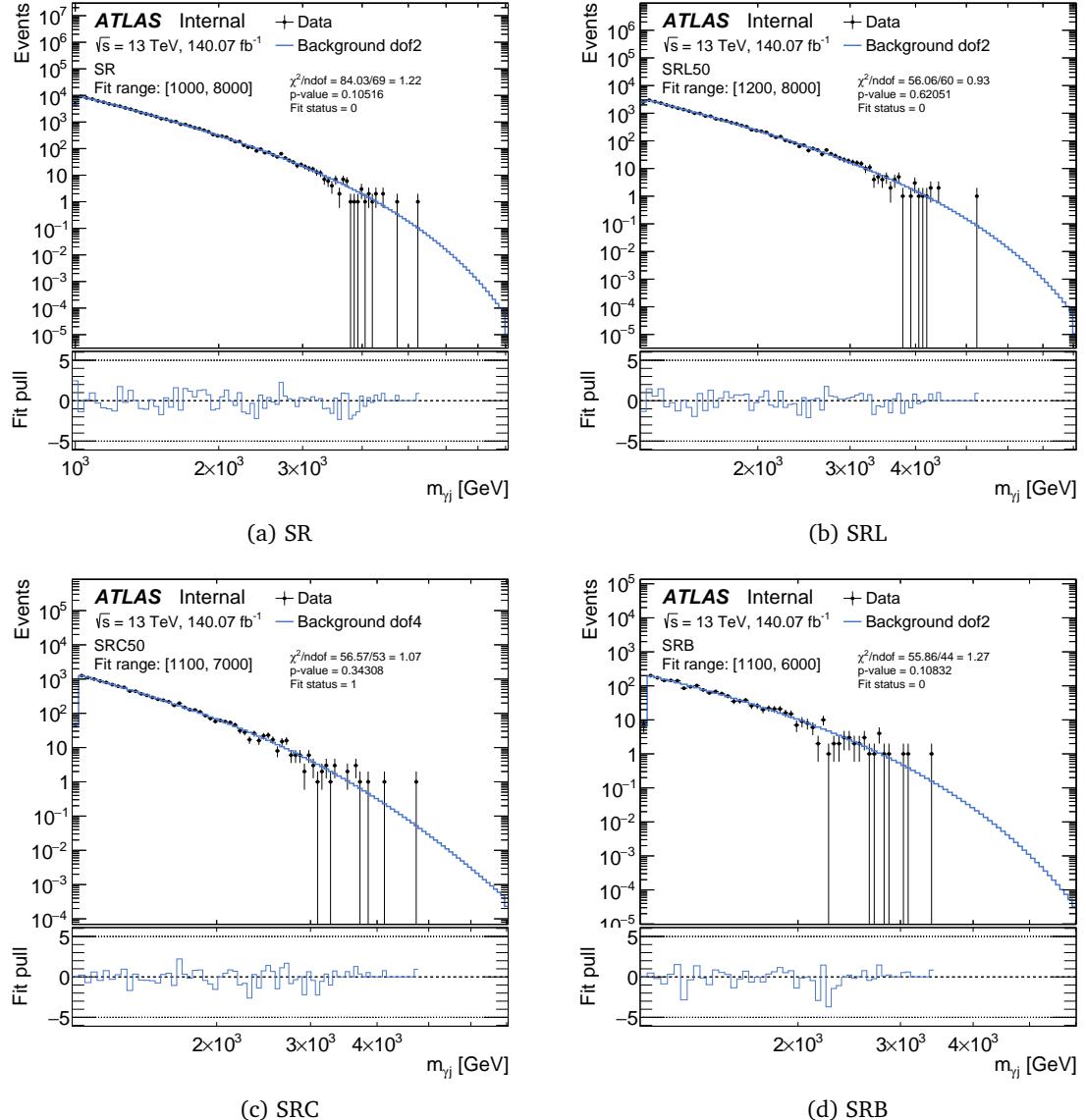


Figure 11.5: **B-only** fits to full Run-2 data in the four considered analysis regions. The data distribution and the background fit are shown in the top pad, while the normalised residues (or pull) in the lower pad. Moreover, fit  $\chi^2$  and  $p$ -value are displayed for each region.

3143 the background estimate and the actual data are downward data fluctuations, specially  
 3144 seen for the SRB region at  $\sim 2100$  GeV. These type of downwards fluctuations make  
 3145 the quality of the fit to decrease, but not drastically. Although there are bins in which  
 3146 the difference is close to a  $+2\sigma$  deviation, these deviations are better identified by the  
 3147 BumpHunter algorithm in the following.

3148 Considering that the  $p(\chi^2)$ -values in the **B-only** fits did not present any sign of devi-  
 3149 ation from the **B-only** hypothesis, that is, no significant difference between the back-  
 3150 ground model and the data, the dedicated BumpHunter algorithm is run to identify where  
 3151 the most significant deviation is. Moreover, the deviation is quantified in terms of the  
 3152 BumpHunter  $p$ -value, calculated as indicated by Eq. 6.25. These results are shown in Fig-  
 3153 ures 11.6–11.9, where the most significant bump is marked with the red dashed lines in

3154 the  $m_{\gamma+j}$  spectrum, obtained where the local  $p$ -value is found to be maximum, as indicated by the figures on the right-hand side for each case. In all cases, the global  $p(\text{BH})$  and  
 3155 significance is indicated in the caption.  
 3156

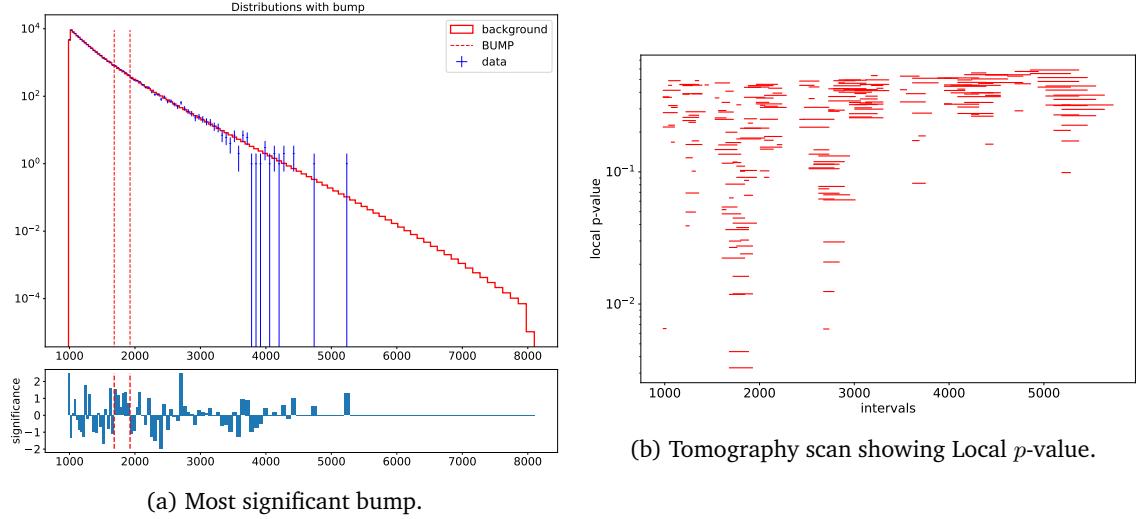


Figure 11.6: BumpHunter results for region SR. The figure on the left shows the most significant bump, marked with red dashed lines. On the right, the local  $p$ -value for different tested windows is presented, indicating where the most significant excess is located. The global  $p(\text{BH})$  as calculated by Eq. 6.25 is 0.3938, while the global significance  $-0.009$ .

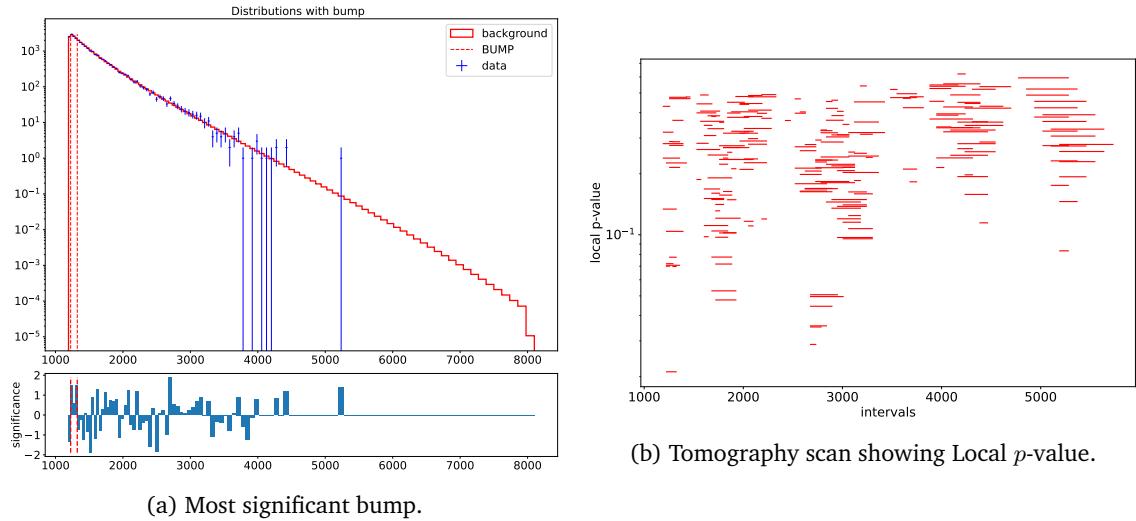


Figure 11.7: Same as Figure 11.6 but in region SRL, where the resulting global  $p(\text{BH}) = 0.8790$  and significance  $-1.170$ .

3157 In all the signal regions, the BumpHunter algorithm manages to find the most significant  
 3158 window in data, but in all cases the significance is either 0 or negative. As discussed in  
 3159 Section 6.2.5, a negative significance means the deviation seen in data is less significant  
 3160 than the one observed from pseudo-data.  
 3161 The fact that the obtained  $p$ -values are greater than the thresholds, both when looking  
 3162 in individual windows or in the whole distribution, when comparing the data and **B-only**  
 3163 functions, means that an excellent modeling of the background is achieved. Only small  
 3164 differences are seen in the residues of the fits, due to data downwards fluctuations in

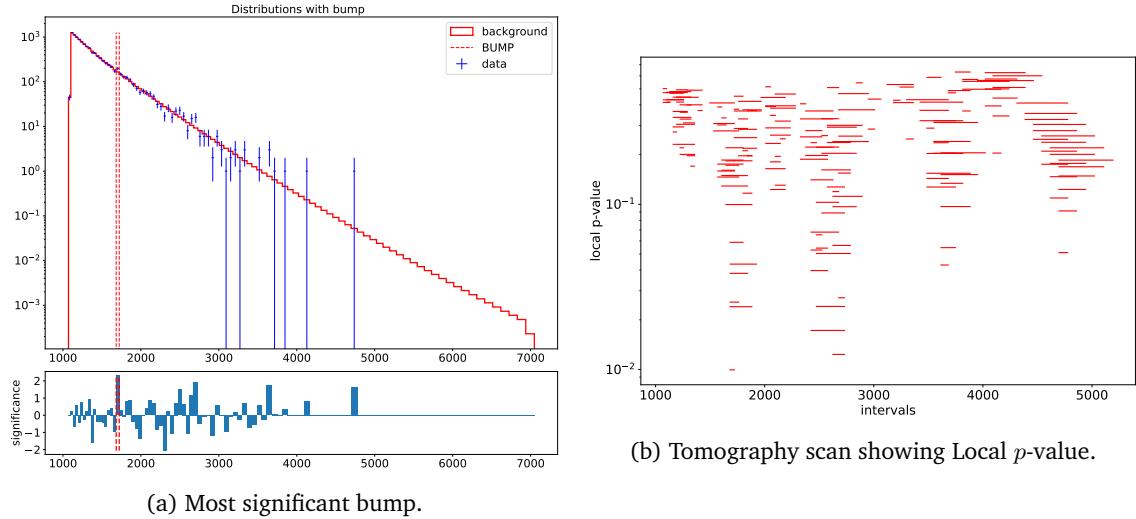


Figure 11.8: Same as Figure 11.6 but in region SRC, where the resulting global  $p(\text{BH}) = 0.6438$  and significance  $-0.368$ .

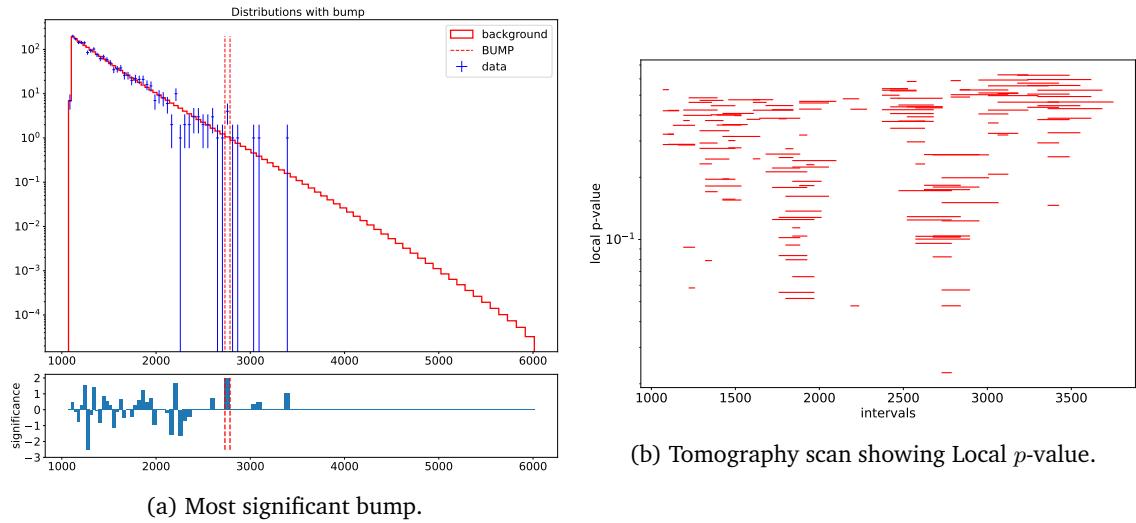
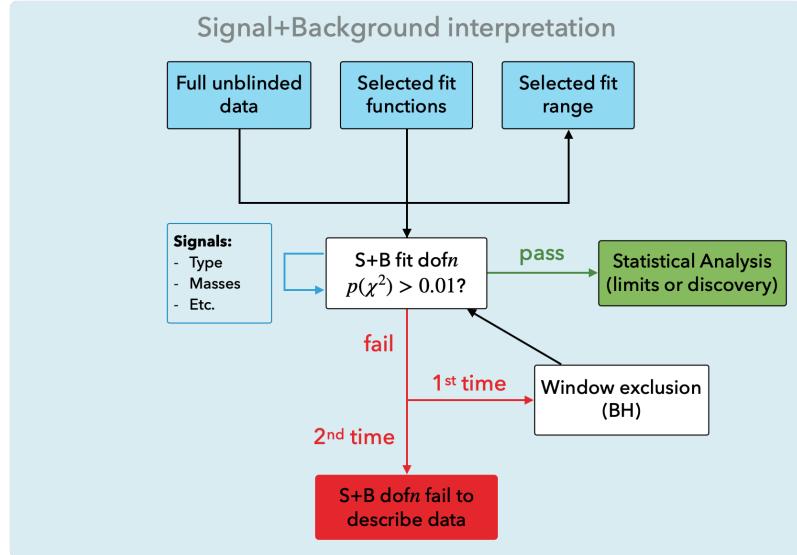


Figure 11.9: Same as Figure 11.6 but in region SRC50, where the resulting global  $p(\text{BH}) = 0.7941$  and significance  $-0.820$ .

3165 high-mass regions.

### 3166 11.2.2 Signal interpretation

3167 The results shown above indicate that no significant bump is observed on the data distri-  
3168 butions. For this reason, now a signal-dependent interpretation is presented.  
3169 The fitting workflow follows a similar pattern as the one used for the **B-only** interpretation  
3170 (schematised in Figure 11.10), but is repeated for each one of the considered signal mod-  
3171 els: Excited Quarks (EQs), Quantum Black Holes (QBHs), and generic Gaussians. First,  
3172 a **S+B** fit is performed on the chosen signal region, followed by a  $\chi^2$ -test. If the  $p$ -value  
3173 of it passes the threshold of 0.01, a good fit across the entire spectrum is provided and  
3174 the limit-setting procedure is started. If the threshold is not passed, a local effect other  
3175 than the signal hypothesis might be the reason. In that case a window exclusion using

Figure 11.10: Scheme of the  $S+B$  fit workflow to data.

3176 BumpHunter is applied as described above and the  $\chi^2$ -test repeated. This path is neces-  
 3177 sary in order to be able to set limits across the entire  $m_{\gamma+j}$  range even if a single localized  
 3178 effect would prevent good fits for all other signal hypothesis at different masses. If in this  
 3179 second test  $p(\chi^2)$  exceeds the threshold of 0.01, the statistical analysis is continued like in  
 3180 the other path, just with a window masked. If the threshold is not exceeded, the  $S+B$  fit  
 3181 fails to describe the data and no interpretation can be presented.  
 3182 The goal is to establish limits on the theory's parameters, that is, determine the maximal  
 3183 amount of potential signal for which the  $S+B$  hypothesis  $H_1$  still provides a good descrip-  
 3184 tion of the data in comparison with the  $B$ -only hypothesis  $H_0$ . The limits are computed at  
 3185 95% CL following the procedure described in Section 6.2.4.

## 3186 Results

3187 In the following paragraphs, a description of the obtained results is provided for each  
 3188 signal model under consideration in the present thesis.

3189 **Excited quarks** The  $EQ$  model proposes a scenario where quarks are not fundamental  
 3190 particles, but rather bound states which decay promptly to a photon and a quark of the  
 3191 same flavor. In this sense, for the first time at the LHC, limits on the  $EQ$  models are  
 3192 separated into three different flavors:  $q^*$ ,  $c^*$  and  $b^*$ . Moreover, the coupling of the  $EQ$  to  
 3193 the  $SM$  partners,  $f = f' = f_s$ , is kept as a free parameter, therefore limits are computed in  
 3194 the coupling-mass plane, leading to two-dimensional limits.

3195 Exclusion limits are set on the signal cross section  $\sigma_s$  instead of the signal strength  $\mu$ . A  
 3196 relation between the two can be found in Eq. 6.3. However, for this model, it is possible  
 3197 to report upper limits on the  $\sigma_s \times Br$ , as the other factors entering Eq. 6.3 are known for  
 3198 each signal.

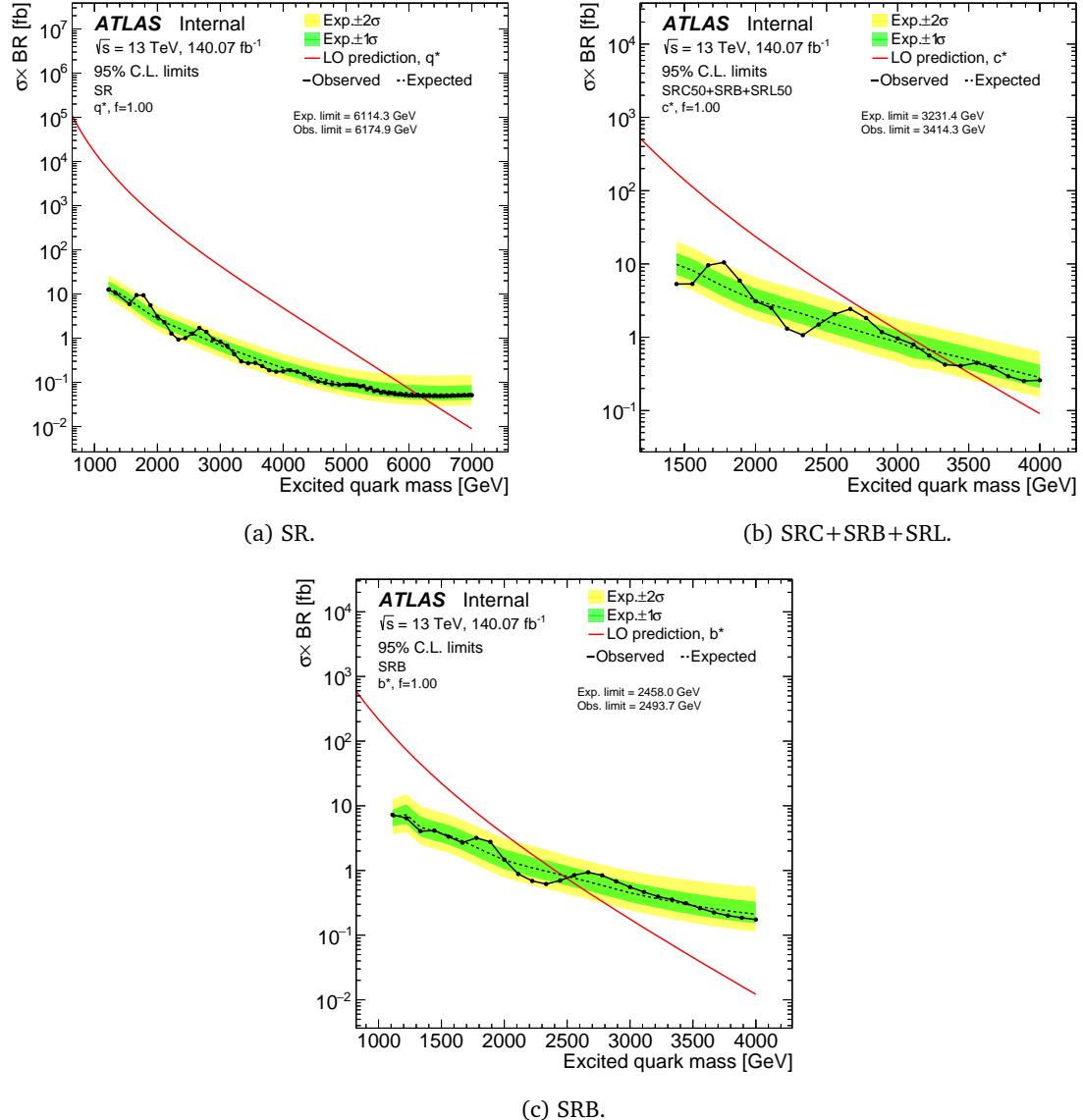


Figure 11.11: Observed and expected limits on  $q^*$ ,  $c^*$  and  $b^*$  signals with coupling  $f = 1$ , using the full Run-2 dataset. Expected upper limits are shown with the black dashed lines, while the line with marks represent observed limits. Red solid lines indicate the predicted cross section. Marked in the figures, the expected and observed upper limits on the mass are shown.

- 3199 In the case of  $q^*$  signals, the upper limits are obtained in the inclusive region SR, and  
 3200 computed for different coupling values. Results of expected and observed upper limits on  
 3201  $\sigma_s \times \text{Br}$  are shown in Figure 11.11a for the case of unit coupling. Local observed excesses  
 3202 can be noted at lower masses ( $\sim 1800$  GeV and  $\sim 2600$  GeV), but are always within the  $2\sigma$   
 3203 expected limit uncertainty. These excesses also match exactly the same locations at where  
 3204 the most significant bumps were located by BumpHunter in Figure 11.6.
- 3205 For  $b^*$  signals, the specialised SRB region is used. The upper limits are shown in Fig-  
 3206 ure 11.11c, where again the observed limits behave in the same way as for the SR region:  
 3207 highest deviations away from the expected limits is at masses where the most significant  
 3208 bumps were found using BumpHunter.
- 3209 Finally, for the first time at the LHC, limits on  $c^*$  signals in the  $\gamma + \text{jet}$  final state are obtained

in this work. The upper limits on  $c^*$  signals are computed in three orthogonal regions simultaneously: SRC, SRB and SRL, and shown for coupling  $f = 1$  in Figure 11.11b. A similar interpretation of the deviations from the expected limits is more difficult in this case, as the simultaneous fit combines the contributions in the three orthogonal regions. However, it is useful to visualise the limits one would get in case of using the  $c$ -tagging region SRC alone, compared to the simultaneous limits. This comparison is shown in Figure 11.12 where a clear improvement is noted, leading to approximately  $\sim 200$  GeV increase on the upper limits on the  $m_{c^*}$  mass.

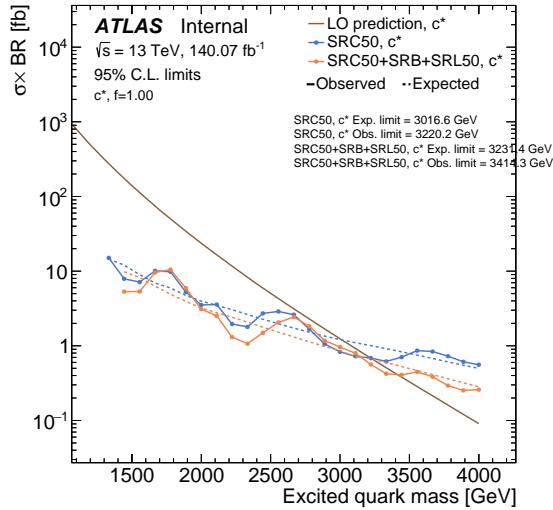


Figure 11.12: Upper limit comparison between the SRC (blue lines) and the SRC+SRB+SRL (orange lines) regions using  $c^*$  signals with coupling  $f = 1$ . For both regions, the expected and observed upper limit on the  $m_{q^*}$  mass is displayed in the figure.

It is also important to test that the behaviour of the limits for different couplings remains mostly unchanged. In Figure 11.13, comparisons of the upper limits by using two different coupling values are shown. For all three flavors, the signal with lower coupling always gives the best limits, by a small and (almost) constant factor. This behaviour is in perfect agreement with what is expected given the different acceptances. The EQ signals' shapes remain mostly unchanged by varying the coupling, but for  $f \rightarrow 1$ , a secondary, off-shell peak appears at lower  $m_{\gamma+j}$ , leading to slightly lower acceptances.

As previously mentioned, EQ upper limits can be represented in the coupling-mass plane, showing the exclusion in these two parameters simultaneously. In the 2D case, the upper limits on the  $\sigma_s \times \text{Br}$  are represented by a three-dimensional surface, and the upper limits on the parameters are obtained. In this space, each point representing the expected and observed upper limit is obtained as the intersection of the theory prediction and the calculated upper limit, for a given coupling and mass value. However, the grid granularity is not fine enough to provide a smooth prediction on these values, so a two-dimensional interpolation is necessary.

However, by virtue of the signal morphing technique presented in Section 9.1, it is possible to obtain a much finer 2D-grid from where more accurate limits are computed. After that,

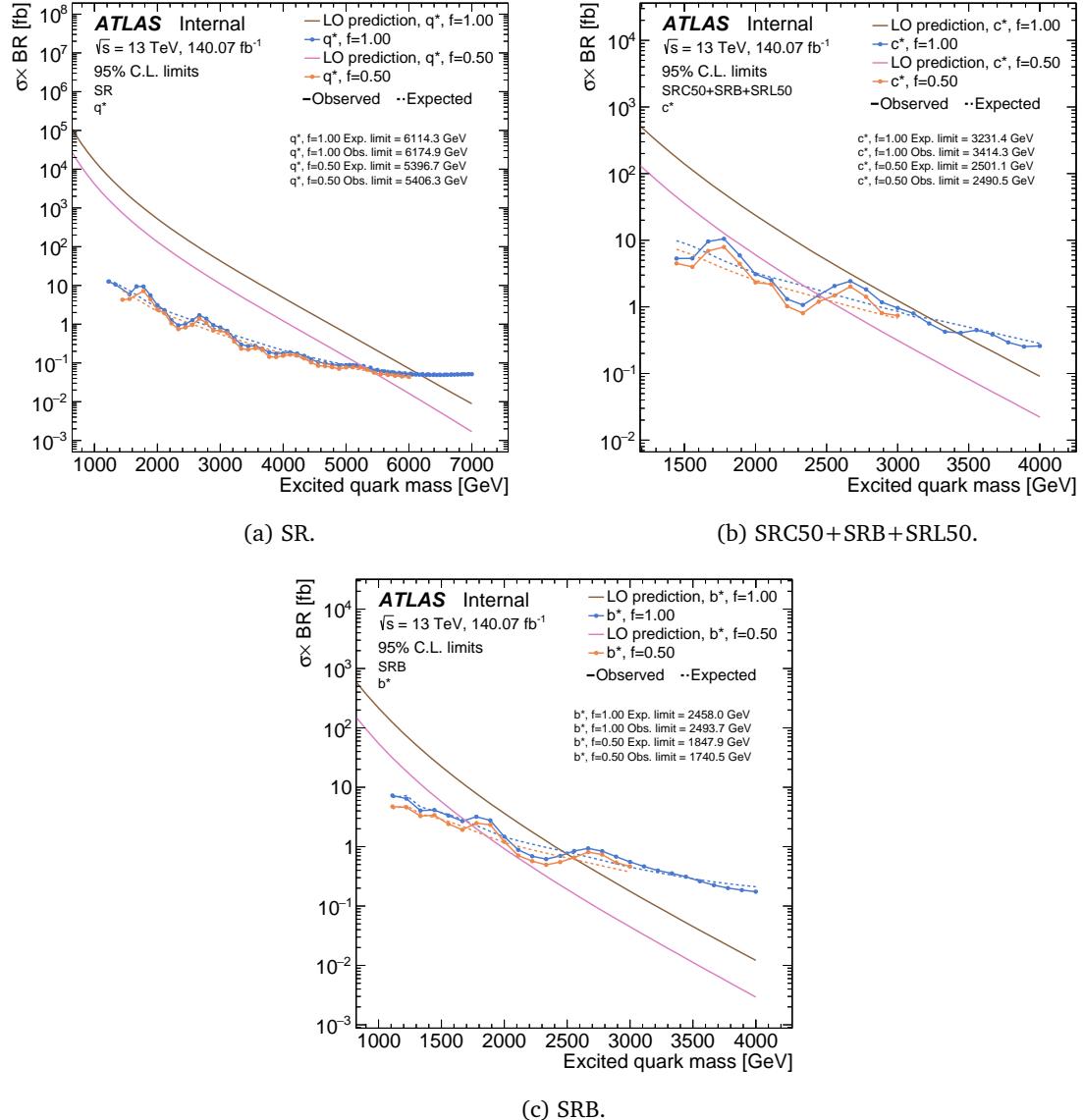


Figure 11.13: Observed and expected limits comparison between different couplings for EQ signals. Blue (orange) lines represent the limits for coupling  $f = 1$  ( $f = 0.5$ ), and the expected model cross sections are shown with brown (pink) solid lines for  $f = 1$  ( $f = 0.5$ ) signals.

3235 a three-dimensional surface is obtained by interpolating the  
 3236 In the two-dimensional case, the upper limit on the two parameters is found as the inter-  
 3237 section of the theory prediction with the calculated upper limit ( $\sigma_s \times \text{Br}$ ).  
 3238 In spite of that, since the upper limit value is found at the intersection of the theory pre-  
 3239 diction with the  $\mu$  value, the obtained granularity is not sufficient to obtain a smooth  
 3240 prediction on the upper limits. For this reason, it is necessary to interpolate the upper lim-  
 3241 its, obtaining a 3D surface, while doing the same for the signal prediction. The intersec-  
 3242 tion of both surfaces will indicate the upper limit value of the theory.  
 3243 The upper limits are also shown in the coupling-mass two-dimensional plane in Fig-  
 3244 ure 11.14. The figure shows the observed and expected 95% CL upper limits (with the  
 3245  $\pm 1\sigma$  uncertainty band) for the three considered flavors, in their respective regions. As

3246 expected the limits on the  $q^*$  model are approximately two times more strict than for the  
 3247  $c^*$  model. In turn, the  $c^*$  upper limits are stricter than for the  $b^*$  signal. This hierarchy  
 3248 on the limits is a consequence primarily on the cross-section of the processes, as lower  
 3249 cross-sections are expected for the heavier resonances.

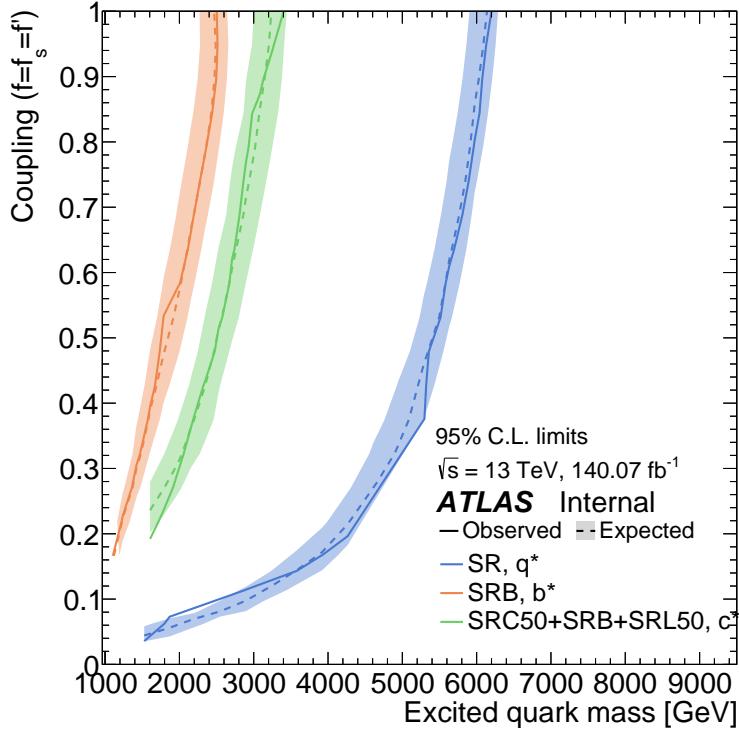


Figure 11.14: Expected and observed upper limits in the mass-coupling plane for the  $q^*$  (SR),  $b^*$  (SRB) and  $c^*$  (SRC50+SRB+SRL50) signals. The expected (observed) upper limits are displayed with dashed (solid) lines, and the  $1\sigma$  uncertainty band is represented by the colored shaded regions.

3250 As of now, these upper limits are the most stringent up to date for the  $\gamma$ +jet final state. At  
 3251 unit coupling, the data exclude  $q^*$ ,  $c^*$  and  $b^*$  models with masses up to 6174.9, 3414.3 and  
 3252 2493.7 GeV, respectively.

3253 **Quantum Black Holes** Similarly as for [EQ](#) exclusion limits, for [QBH](#), the upper limits are  
 3254 reported on the cross-section times branching ratio  $\sigma_s \times \text{Br}$ . The results on the observed  
 3255 and expected limits for the two [QBH](#) models (RS1 and ADD) are shown in Figure 11.15.  
 3256 Following the fitting strategies summary presented in Table 10.3, [QBH](#) signals are only  
 3257 searched for in the inclusive SR. Similarly to what was seen in the case of [EQ](#) signals,  
 3258 observed limits at  $\sim 1800$  GeV and  $\sim 2600$  GeV present deviations up to  $2\sigma$ , due to the  
 3259 local, but not significant, excesses found on data with BumpHunter.

3260 Again, the present upper limits on the [QBH](#) masses are the most stringent up to date with  
 3261 the  $\gamma$ +jet final state. Models proposing extra dimensions, whether with warped geometry  
 3262 as proposed by the RS1 model, or with at least two flat extra dimensions as proposed by  
 3263 the ADD model, suggest the formation of [QBH](#), and these can be excluded at 95% up to  
 3264 masses of 5347.4 GeV and 7590.1 GeV, respectively.

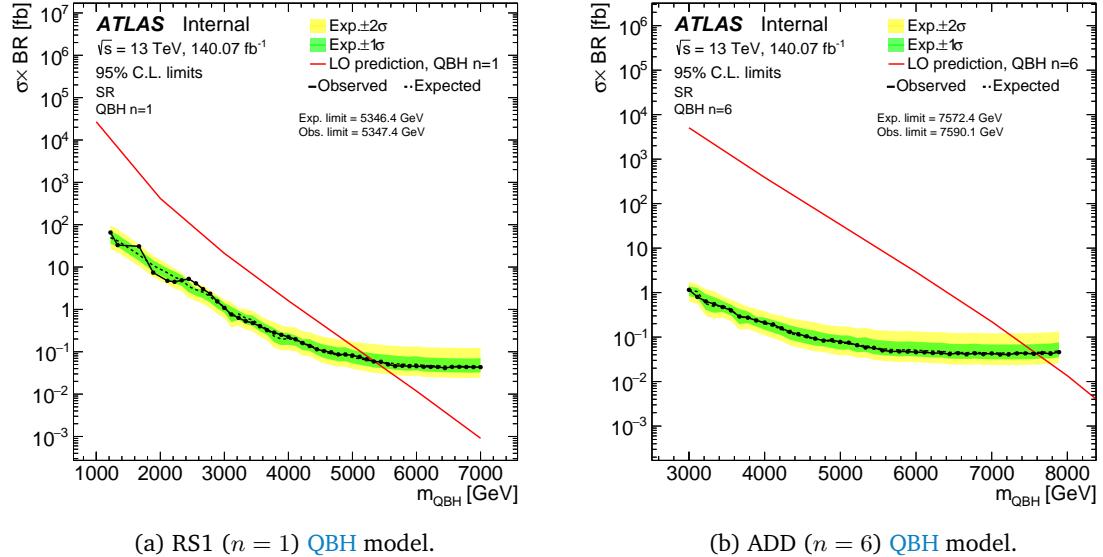


Figure 11.15: Observed and expected upper limits on the **QBH** signal models. The RS1 model ( $n = 1$ ) is shown on the left, while the ADD model ( $n = 6$ ) is presented on the right figure. Observed limits are shown with the black dots and solid line, meanwhile, expected upper limits are represented with the black dashed line, with the corresponding  $\pm 1$  and  $\pm 2\sigma$  uncertainty bands.

3265 **General Gaussian signals** The final signal model considered are general Gaussian shapes.  
 3266 These signals allows for a general and model-agnostic interpretation in this final state. Ex-  
 3267 pected and observed upper limits are computed with Gaussian-shaped resonances, whose  
 3268 width is chosen to take the three following values: 2%, 7% and 15% the mass resolution.  
 3269 Unlike the previous models, the reported exclusion limits are on the visible cross section,  
 3270 namely

$$\sigma_s \times \text{Br} \times A \times \varepsilon, \quad (11.2)$$

3271 as they can be interpreted for any arbitrary signal model that produce a  $\gamma + \text{jet}$  resonance  
 3272 of approximately a Gaussian shape in  $m_{\gamma+j}$ .  
 3273 For each one of the signal regions considered (SR, SRL, SRC and SRB), the results limits on  
 3274 generic Gaussians with the aforementioned widths are displayed in Figure 11.16. In the  
 3275 inclusive SR, the obtained limits show exactly the same structure at low  $m_{\gamma+j}$  as the other  
 3276 signal models, where a small excess of data events is seen at  $\sim 1800$  GeV and  $\sim 2500$  GeV.  
 3277 The sensitivity for narrower resonances is better than that for wider resonances for two  
 3278 reasons: for wide resonances, the same number of signal events is distributed across a  
 3279 larger  $m_{\gamma+j}$  range, which lowers the signal-to-background ratio in this range. Additionally,  
 3280 the flexibility of the background fit allows it to better adapt to wide signals than to narrow  
 3281 signals, resulting in a larger ambiguity between signal and background events.  
 3282 From the comparison of different widths, it is also important to note that narrow res-  
 3283 onances lead to much more fluctuations on the limits. The reason behind this is that  
 3284 narrow peaks can capture more easily any statistical fluctuation on the data, and these  
 3285 fluctuations are then translated into the limits. Conversely, wider signals lead to much  
 3286 smoother shapes on the observed limits.

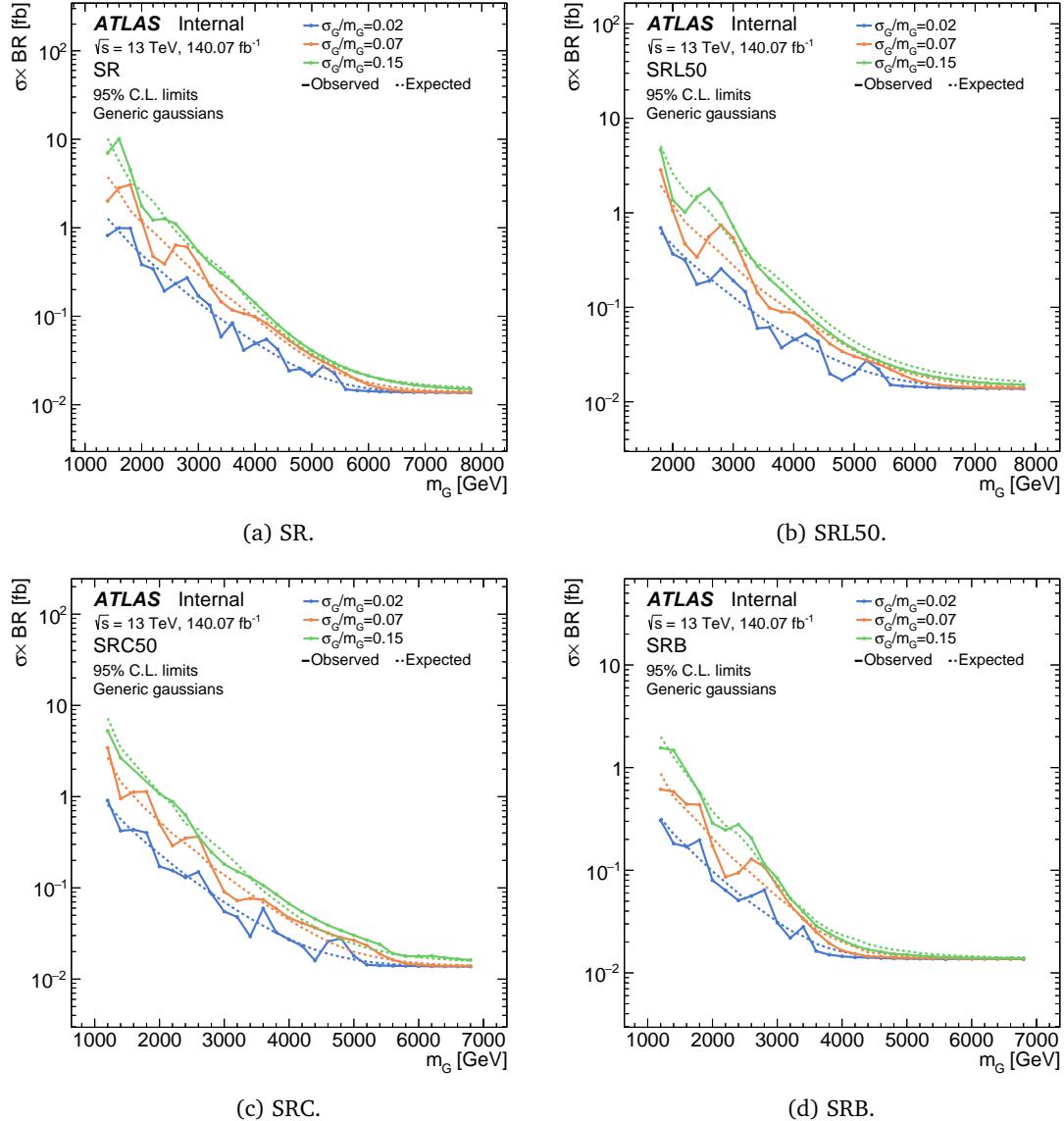


Figure 11.16: Observed and expected upper limits on general gaussian signals with three different widths:  $\sigma_G/m_G = 0.02, 0.07$  and  $0.15$  in regions SR, SRL, SRC and SRB.

Finally, in Figure 11.17, a comparison of the current results with the latest results on gaussian resonances in the  $\gamma + \text{jet}$  final state are presentd. The previous search was done using only the 2015+2016 ATLAS dataset, leading to a total integrated luminosity of  $36.1 \text{ fb}^{-1}$ . Since the upper limit on the cross-section scales with the luminosity as  $1/\sqrt{L}$ , one would expect a  $\sqrt{140.07}/\sqrt{36.1} \approx 1.96$  improvement on the upper limit. However, as seen from the figures, a higher improvement on the sensitivity is obtained in the current analysis, where the upper limit at 5500 GeV is improved by a factor of approximately 5.3. In consequence, with the current results, by virtue of a more sophisticated background modeling, it is possible to increase the improvements one would get by using more data alone by a factor of  $\sim 2.7$ .

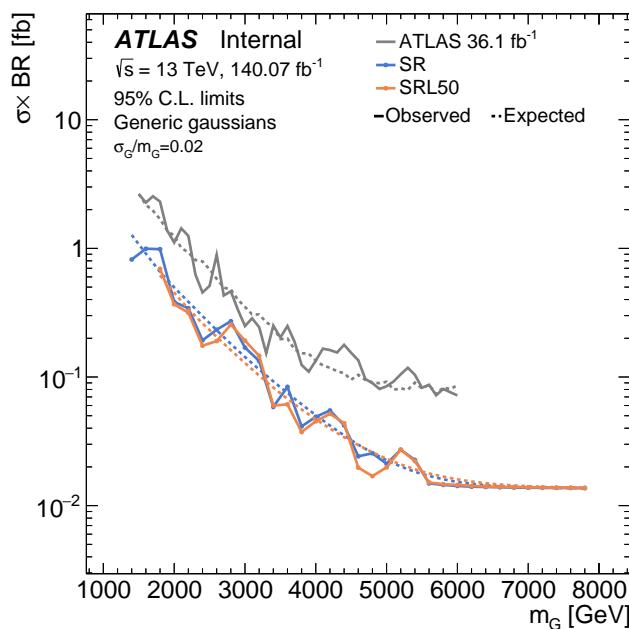


Figure 11.17: Comparison of the observed and expected limits on Gaussian-shaped resonances of 2% width obtained in this analysis (blue and orange for the SR and SRL regions, respectively) against the previous [ATLAS](#) result using 2015-2016 data (grey). The observed limits are shown with the solid line and the expected ones with dashed lines.

3297

## 3298 CONCLUSIONS

3299

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

---

Mae Jemison

- 3300 Physics processes with photons in the final state play a central role in the [ATLAS](#) physics
- 3301 program. Ranging from photons with  $p_T \sim 10$  GeV, such as the ones produced from
- 3302 radiative  $Z$  decays, up to prompt photon with transverse momentum at the TeV scale,
- 3303 these are widely used either for [QCD](#) precision measurements, or for searches of new
- 3304 physics [BSM](#).
- 3305 Photon identification in [ATLAS](#) relies on the photon signature in the calorimeters systems.
- 3306 When photons pass through the [ECAL](#), they undergo processes such as electron-positron
- 3307 pair creation, or Bremsstrahlung radiation, leaving an [EM](#) shower. However, not only
- 3308 photons leave this trace on the [ECAL](#). Jets, specially highly energy ones, can produce
- 3309 The correct identification of photons in [ATLAS](#) is one of the most crucial
- 3310 Photon identification in [ATLAS](#) is one of the most challenging tasks For the Higgs boson
- 3311 discovery, amongst the different decay channels, the one with the cleanest signature and
- 3312 the lowest uncertainty was the  $H \rightarrow \gamma\gamma$ , evidencing the excellent photon identification
- 3313 The process with highest cross-section in  $pp$  collisions is, by far, [QCD](#) dijet. The final state
- 3314 jets,

3315

# 3316 ELLIPSE FITTING FORMULAE



3317 The ellipse fit to obtain the statistical uncertainties on the shift and stretch Fudge Factors  
 3318 (FFs) is implemented following the procedure suggested in Ref. [166]. The fit returns the  
 3319 set of parameters  $\{A, B, C, D, E, F\}$  that parametrise the conic:

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

3320 with  $B^2 - 4AC < 0$  for ellipses. The variables  $x, y$  shown are general, but in the case of  
 3321 FFs they represent the shift and stretch parameters, respectively.

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

3322 it is possible to extract the ellipse center  $(x_0, y_0)$ , its tilt angle  $\theta$  and its semi-major and  
 3323 semi-minor axis,  $a$  and  $b$ , respectively. Then, the desired uncertainties, and the correlation  
 3324 between  $x$  and  $y$  can be extracted using the following relations (see Figure A.1):

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

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

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

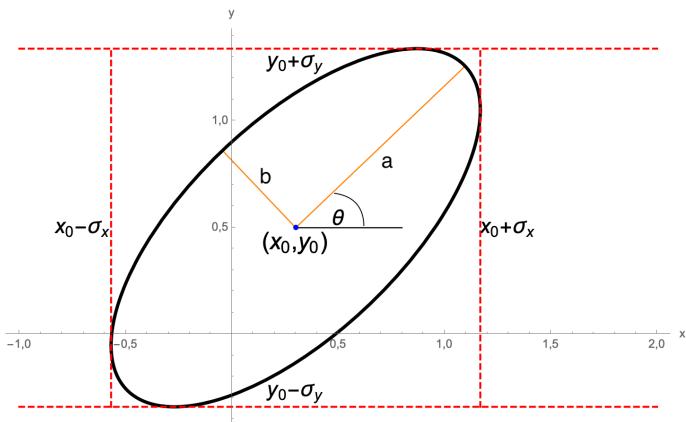


Figure A.1: Parameters of the most general case of an ellipse.

3325

# 3326 SIGNAL INJECTION RESULTS

# B

3327 **Gaussian signals** In Figures B.1–B.4, results of the SI test with Gaussian signals are  
 3328 shown for the different gaussian widths of  $\sigma_G/m_G = 0.02, 0.05$  and  $0.20$ . The tests have  
 3329 been carried out in each one of the analysis regions using the Gaussian signal model: SR,  
 3330 SRL, SRC and SRB. For all widths hypothesis the test is passed as linear behaviours are  
 3331 seen, and the ratio of  $|S_{\text{spur}}|/N_{\text{inj}} < 0.5$

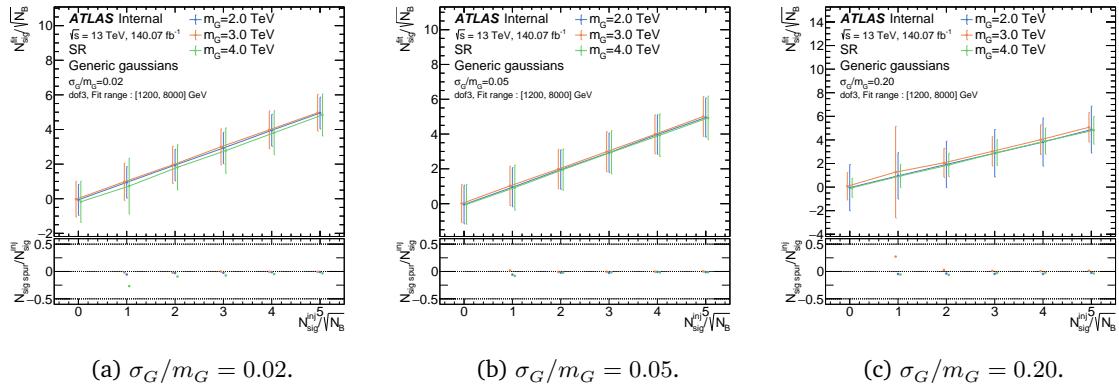


Figure B.1: SI tests results using Gaussians signals in the inclusive SR region. The fit is performed in the range  $1200 - 8000$  GeV using the *dof3* function. Each considered mass is shown with a different color. The  $x$ -axis shows the injected amplitude in units of  $\sqrt{N_B}$ , while the  $y$ -axis represents the extracted signal in terms of  $\sqrt{N_B}$ .

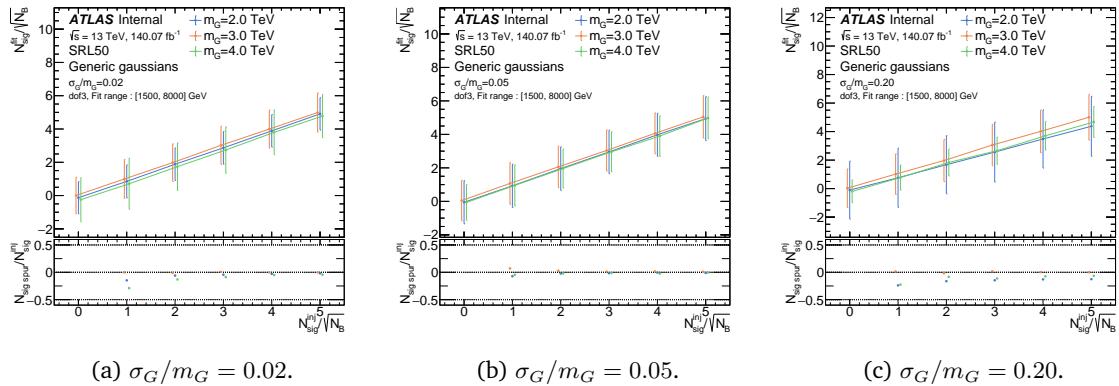


Figure B.2: Same as Figure B.1 in the SRL region performing fits in the range  $1500 - 8000$  GeV with the *dpf3* function.

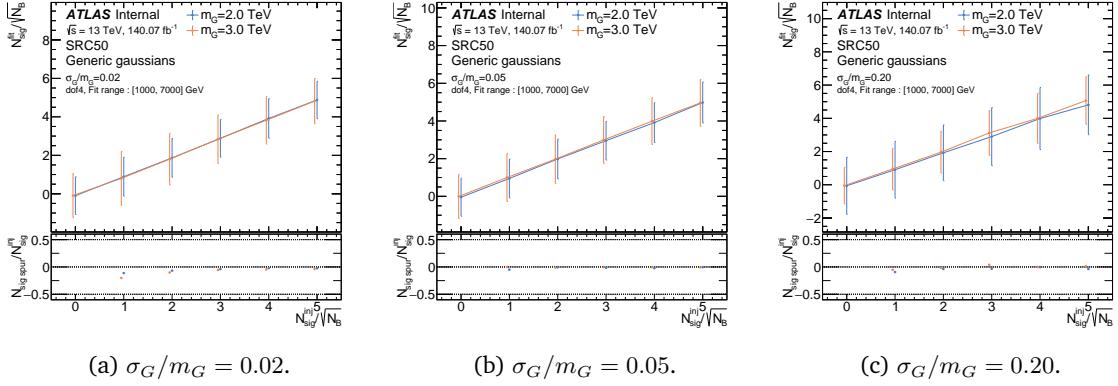


Figure B.3: Same as Figure B.1 in the SRC region performing fits in the range  $1000 - 7000$  GeV with the  $dpf4$  function.

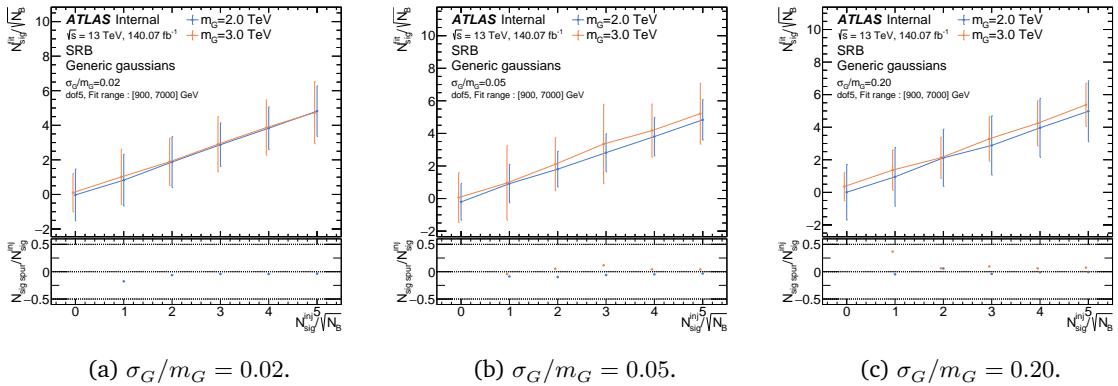


Figure B.4: Same as Figure B.1 in the SRB region performing fits in the range  $900 - 7000$  GeV with the  $dpf5$  function.

3332 **Quantum Black Holes** Similarly to what has been seen for the gaussian signals, in the  
3333 **QBHs** case good linearity is followed by all the considered masses within the fit-range.

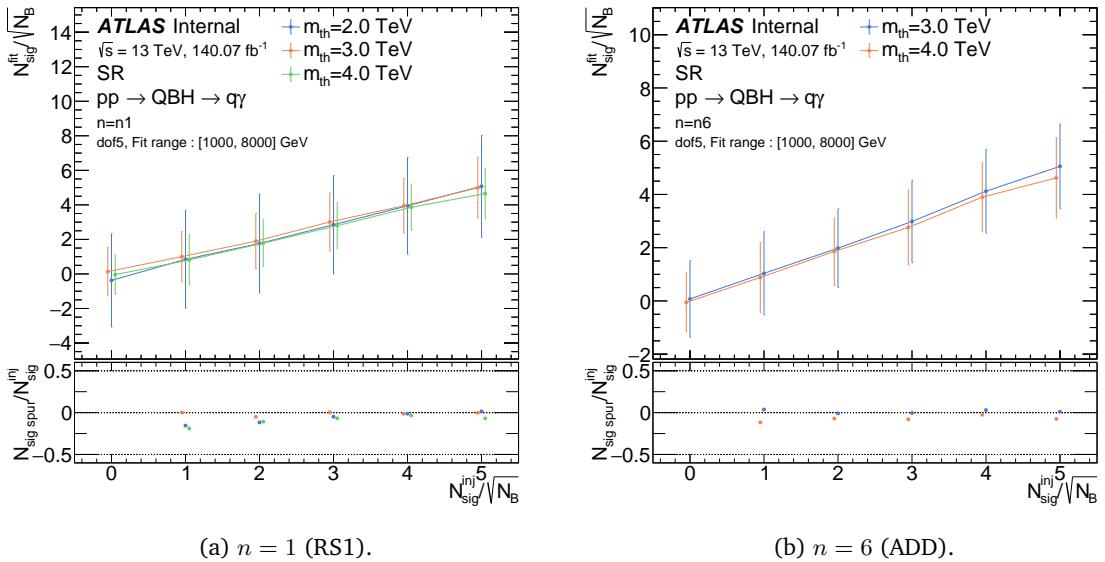


Figure B.5: SI tests results using QBH signals in the inclusive SR region. The fit is performed in the range  $1000 - 8000$  GeV using the  $dof5$  function.

3334

3335

# GLOSSARY

|      |   |    |
|------|---|----|
| 3336 | <b>CERN</b> European Organization for Nuclear Research . . . . .                              | 29 |
| 3337 | <b>ALICE</b> A Large Ion Collider Experiment  |    |
| 3338 | <b>ATLAS</b> A Toroidal LHC ApparatuS . . . . .   | 9  |
| 3339 | <b>CMS</b> Compact Muon Solenoid . . . . .  | 9  |
| 3340 | <b>LHC</b> Large Hadron Collider . . . . .  | 9  |
| 3341 | <b>LHCb</b> Large Hadron Collider beauty  |    |
| 3342 | <b>LHCf</b> Large Hadron Collider forward   |    |
| 3343 | <b>LS2</b> Long Shut down 2 . . . . .   | 31 |
| 3344 | <b>TOTEM</b> TOTal cross section, Elastic scattering and diffraction dissociation Measurement |    |
| 3345 | at the <b>LHC</b>   |    |
| 3346 | <b>MoEDAL</b> Monopole & Exotics Detector At the <b>LHC</b>                                   |    |
| 3347 | <b>PS</b> Proton Synchrotron . . . . .  | 29 |
| 3348 | <b>SPS</b> Super Proton Synchrotron . . . . .   | 29 |
| 3349 | <b>L1</b> Level-1 . . . . .   | 40 |
| 3350 | <b>L1Topo</b> Level-1 Topological . . . . .   | 40 |
| 3351 | <b>LAr</b> Liquid Argon . . . . .   | 36 |
| 3352 | <b>ECAL</b> Electromagnetic Calorimeter . . . . .   | 27 |
| 3353 | <b>CSC</b> Cathode Strip Chamber . . . . .  | 38 |
| 3354 | <b>CTP</b> Central Trigger Processor . . . . .  | 40 |
| 3355 | <b>FCAL</b> Forward Calorimeter . . . . .   | 36 |
| 3356 | <b>HEC</b> Hadronic End-Cap Calorimeter . . . . .   | 38 |
| 3357 | <b>HCAL</b> Hadronic Calorimeter . . . . .  | 27 |
| 3358 | <b>HLT</b> High Level Trigger . . . . .   | 40 |
| 3359 | <b>IBL</b> Insertable B-Layer . . . . .   | 33 |

|      |               |                              |     |
|------|---------------|------------------------------|-----|
| 3360 | <b>ID</b>     | Inner Detector               | 27  |
| 3361 | <b>SCT</b>    | SemiConductor Tracker        | 33  |
| 3362 | <b>RPC</b>    | Resistive-Plate Chamber      | 39  |
| 3363 | <b>TGC</b>    | Thin-Gap Chamber             | 39  |
| 3364 | <b>TRT</b>    | Transition Radiation Tracker | 33  |
| 3365 | <b>NSW</b>    | New Small Wheel              | 39  |
| 3366 | <b>MDT</b>    | Monitored Drift Tube         | 38  |
| 3367 | <b>MS</b>     | Muon Spectrometer            | 27  |
| 3368 | <b>GRL</b>    | Good Runs List               | 41  |
| 3369 | <b>LB</b>     | Luminosity-block             | 41  |
| 3370 | <b>PSB</b>    | Proton Synchrotron Booster   | 29  |
| 3371 | <b>MC</b>     | Monte Carlo                  | 6   |
| 3372 | <b>MV</b>     | Multivariate                 | 65  |
| 3373 | <b>KDE</b>    | Kernel Density Estimator     | 71  |
| 3374 | <b>ROI</b>    | Region of Interest           | 40  |
| 3375 | <b>B-only</b> | Background-only              | 90  |
| 3376 | <b>S+B</b>    | Signal+Background            | 91  |
| 3377 | <b>CL</b>     | Confidence Level             | 92  |
| 3378 | <b>FWHM</b>   | Full Width at Half Maximum   | 138 |
| 3379 | <b>NP</b>     | Nuisance Parameter           | 89  |
| 3380 | <b>FF</b>     | Fudge Factor                 | 48  |
| 3381 | <b>FF</b>     | Fake Factor                  | 123 |
| 3382 | <b>SF</b>     | Scale Factor                 | 67  |
| 3383 | <b>SS</b>     | Shower Shape                 | 59  |
| 3384 | <b>SSV</b>    | Shower Shape Variable        | 48  |
| 3385 | <b>SS</b>     | Spurious Signal              | 89  |
| 3386 | <b>SI</b>     | Signal Injection             | 132 |
| 3387 | <b>WP</b>     | Working Point                | 49  |
| 3388 | <b>POI</b>    | Parameter of Interest        | 88  |
| 3389 | <b>PDF</b>    | Probability Density Function | 71  |
| 3390 | <b>BSM</b>    | Beyond Standard Model        | 20  |
| 3391 | <b>DM</b>     | Dark Matter                  | 20  |
| 3392 | <b>EM</b>     | electromagnetic              | 6   |

|      |              |  |     |
|------|--------------|--|-----|
| 3393 | <b>EW</b>    | Electroweak . . . . .                          | 8   |
| 3394 | <b>IR</b>    | Infrared . . . . .                             | 12  |
| 3395 | <b>UV</b>    | Ultraviolet . . . . .                          | 12  |
| 3396 | <b>FSR</b>   | Final State Radiation . . . . .                | 16  |
| 3397 | <b>RZ</b>    | Radiative $Z$ . . . . .                        | 63  |
| 3398 | <b>SP</b>    | Single Photon . . . . .                        | 63  |
| 3399 | <b>ISR</b>   | Initial State Radiation . . . . .              | 16  |
| 3400 | <b>LO</b>    | Leading Order . . . . .                        | 17  |
| 3401 | <b>ME</b>    | Matrix Element . . . . .                       | 24  |
| 3402 | <b>NLO</b>   | Next-to-Leading Order . . . . .                | 17  |
| 3403 | <b>NNLO</b>  | Next-to-Next-to-Leading Order . . . . .        | 19  |
| 3404 | <b>PDF</b>   | Parton Distribution Function . . . . .         | 14  |
| 3405 | <b>PS</b>    | Parton Shower . . . . .                        | 25  |
| 3406 | <b>EWSB</b>  | ElectroWeak Symmetry Breaking . . . . .        | 8   |
| 3407 | <b>QCD</b>   | Quantum Chromodynamics . . . . .               | 8   |
| 3408 | <b>pQCD</b>  | Perturbative QCD . . . . .                     | 19  |
| 3409 | <b>QED</b>   | Quantum Electrodynamics . . . . .              | 12  |
| 3410 | <b>QFT</b>   | Quantum Field Theory . . . . .                 | 8   |
| 3411 | <b>QBH</b>   | Quantum Black Hole . . . . .                   | 23  |
| 3412 | <b>EQ</b>    | Excited Quark . . . . .                        | 21  |
| 3413 | <b>SM</b>    | Standard Model . . . . .                       | 6   |
| 3414 | <b>UE</b>    | Underlying Event . . . . .                     | 16  |
| 3415 | <b>VEV</b>   | Vacuum Expectation Value . . . . .             | 10  |
| 3416 | <b>SUSY</b>  | Supersymmetry . . . . .                        | 1   |
| 3417 | <b>CB</b>    | Crystal-Ball . . . . .                         | 125 |
| 3418 | <b>DSACB</b> | Double-sided Asymmetric Crystal-Ball . . . . . | 126 |
| 3419 | <b>dof</b>   | Degree of Freedom . . . . .                    | 128 |
| 3420 | <b>NNJvt</b> | Neural Network JVT . . . . .                   | 102 |
| 3421 | <b>JVT</b>   | Jet vertex Tagger . . . . .                    | 53  |
| 3422 | <b>JVF</b>   | Jet vertex Fraction . . . . .                  | 53  |
| 3423 | <b>PV</b>    | Primary Vertex . . . . .                       | 46  |
| 3424 | <b>CB</b>    | Combined Muon . . . . .                        | 50  |
| 3425 | <b>ST</b>    | Segmented Muon . . . . .                       | 50  |

|      |  |    |
|------|--|----|
| 3426 | <b>CT</b> Calorimetric Muon . . . . .            | 50 |
| 3427 | <b>ME</b> Standalone/Extrapolated Muon . . . . . | 50 |
| 3428 | <b>PFlow</b> Particle Flow . . . . .             | 52 |
| 3429 | <b>FTAG</b> Flavor Tagging . . . . .             | 54 |

3430

## 3431 BIBLIOGRAPHY

- 3432 [1] L. Evans and P. Bryant, *LHC Machine*, *Journal of Instrumentation* **3** no. 08, (2008)  
 3433 [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>  
 3434 <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08001/meta>. 1, 29
- 3436 [2] A. Purcell, *Go on a particle quest at the first CERN webfest. Le premier webfest du  
 3437 CERN se lance à la conquête des particules.*,  
 3438 <https://cds.cern.ch/record/1473657>. x, 7
- 3439 [3] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and Collider Physics*. Cambridge  
 3440 Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge  
 3441 University Press, 1996. 8
- 3442 [4] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.* **13**  
 3443 (1964) 508–509. <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>. 8
- 3444 [5] P. Higgs, *Broken symmetries, massless particles and gauge fields*, *Physics Letters* **12**  
 3445 no. 2, (1964) 132–133.  
 3446 <https://www.sciencedirect.com/science/article/pii/0031916364911369>. 8
- 3447 [6] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, *Phys. Rev.*  
 3448 **145** (1966) 1156–1163.  
 3449 <https://link.aps.org/doi/10.1103/PhysRev.145.1156>. 8
- 3450 [7] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*,  
 3451 *Phys. Rev. Lett.* **13** (1964) 321–323.  
 3452 <https://link.aps.org/doi/10.1103/PhysRevLett.13.321>. 8
- 3453 [8] ATLAS Collaboration, *Observation of a new particle in the search for the Standard  
 3454 Model Higgs boson with the ATLAS detector at the LHC*, *Physics Letters B* **716** no. 1,  
 3455 (2012) 1–29.  
 3456 <https://www.sciencedirect.com/science/article/pii/S037026931200857X>.  
 3457 9

- 3458 [9] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS*  
3459 *experiment at the LHC*, Physics Letters B **716** no. 1, (2012) 30–61.  
3460 <https://www.sciencedirect.com/science/article/pii/S0370269312008581>.  
3461 9
- 3462 [10] M. Gell-Mann, *A schematic model of baryons and mesons*, Physics Letters **8** no. 3,  
3463 (1964) 214–215.  
3464 <https://www.sciencedirect.com/science/article/pii/S0031916364920013>.  
3465 11
- 3466 [11] G. Zweig, *An  $SU_3$  model for strong interaction symmetry and its breaking; Version 1*,  
3467 tech. rep., CERN, Geneva, 1964. <https://cds.cern.ch/record/352337>. 11
- 3468 [12] G. Zweig, *An  $SU_3$  model for strong interaction symmetry and its breaking; Version*  
3469 *2*,. <https://cds.cern.ch/record/570209>. 11
- 3470 [13] Particle Data Group Collaboration Collaboration, S. Navas et al., *Review of Particle*  
3471 *Physics*, Phys. Rev. D **110** (2024) 030001.  
3472 <https://link.aps.org/doi/10.1103/PhysRevD.110.030001>. x, 13
- 3473 [14] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Non-Abelian Gauge Theories*,  
3474 Phys. Rev. Lett. **30** (1973) 1343–1346.  
3475 <https://link.aps.org/doi/10.1103/PhysRevLett.30.1343>. 13
- 3476 [15] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Phys. Rev. Lett.  
3477 **30** (1973) 1346–1349.  
3478 <https://link.aps.org/doi/10.1103/PhysRevLett.30.1346>. 13
- 3479 [16] H. Georgi and S. L. Glashow, *Unity of All Elementary-Particle Forces*, Phys. Rev. Lett.  
3480 **32** (1974) 438–441.  
3481 <https://link.aps.org/doi/10.1103/PhysRevLett.32.438>. 13
- 3482 [17] R. P. Feynman, *Very High-Energy Collisions of Hadrons*, Phys. Rev. Lett. **23** (1969)  
3483 **1415–1417**. <https://link.aps.org/doi/10.1103/PhysRevLett.23.1415>. 14
- 3484 [18] J. D. Bjorken and E. A. Paschos, *Inelastic Electron-Proton and  $\gamma$ -Proton Scattering*  
3485 *and the Structure of the Nucleon*, Phys. Rev. **185** (1969) 1975–1982.  
3486 <https://link.aps.org/doi/10.1103/PhysRev.185.1975>. 14
- 3487 [19] J. D. Bjorken, *Asymptotic Sum Rules at Infinite Momentum*, Phys. Rev. **179** (1969)  
3488 **1547–1553**. <https://link.aps.org/doi/10.1103/PhysRev.179.1547>. 14
- 3489 [20] R. Ellis, H. Georgi, M. Machacek, H. Politzer, and G. G. Ross, *Factorization and the*  
3490 *parton model in QCD*, Physics Letters B **78** no. 2, (1978) 281–284.  
3491 <https://www.sciencedirect.com/science/article/pii/0370269378900230>.  
3492 14

- 3493 [21] J. C. Collins, D. E. Soper, and G. Sterman, “Factorization of hard processes in qcd.”  
3494 [https://www.worldscientific.com/doi/abs/10.1142/9789814503266\\_0001](https://www.worldscientific.com/doi/abs/10.1142/9789814503266_0001).  
3495 14
- 3496 [22] J. C. Collins and D. E. Soper, *The Theorems of Perturbative QCD*, Annual Review of  
3497 Nuclear and Particle Science **37** no. Volume 37,, (1987) 383–409.  
3498 <https://www.annualreviews.org/content/journals/10.1146/annurev.ns.37.120187.002123>. 14
- 3500 [23] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering*  
3501 and  $e^+ e^-$  Annihilation by Perturbation Theory in Quantum Chromodynamics., Sov.  
3502 Phys. JETP **46** (1977) 641–653. 15
- 3503 [24] V. Gribov and L. Lipatov, *Deep inelastic electron scattering in perturbation theory*,  
3504 Physics Letters B **37** no. 1, (1971) 78–80.  
3505 <https://www.sciencedirect.com/science/article/pii/0370269371905764>.  
3506 15
- 3507 [25] G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*, Nuclear Physics B  
3508 **126** no. 2, (1977) 298–318.  
3509 <https://www.sciencedirect.com/science/article/pii/0550321377903844>.  
3510 15
- 3511 [26] NNPDF Collaboration, *Parton distributions for the LHC run II*, Journal of High  
3512 Energy Physics **2015** no. 4, (2015) 40.  
3513 [https://doi.org/10.1007/JHEP04\(2015\)040](https://doi.org/10.1007/JHEP04(2015)040). x, 15
- 3514 [27] P. D. Group, *Review of Particle Physics*, Progress of Theoretical and Experimental  
3515 Physics **2020** no. 8, (2020) 083C01,  
3516 <https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf>.  
3517 <https://doi.org/10.1093/ptep/ptaa104>. x, 15
- 3518 [28] S. Höche, *Introduction to parton-shower event generators*, 2015.  
3519 <https://arxiv.org/abs/1411.4085>. x, 16
- 3520 [29] ATLAS Collaboration, T. A. collaboration, *Standard Model Summary Plots June*  
3521 *2024*, tech. rep., CERN, Geneva, 2024. <https://cds.cern.ch/record/2903866>.  
3522 x, 17
- 3523 [30] T. Pietrycki and A. Szczurek, *Photon-jet correlations in  $pp$  and  $p\bar{p}$  collisions*, Phys.  
3524 Rev. D **76** (2007) 034003.  
3525 <https://link.aps.org/doi/10.1103/PhysRevD.76.034003>. 17
- 3526 [31] Z. Belghobsi, M. Fontannaz, J.-P. Guillet, G. Heinrich, E. Pilon, and M. Werlen,  
3527 *Photon-jet correlations and constraints on fragmentation functions*, Phys. Rev. D **79**  
3528 (2009) 114024. <https://link.aps.org/doi/10.1103/PhysRevD.79.114024>. 17

- 3529 [32] J. Kühn and P. Zerwas, *Excited quarks and leptons*, Physics Letters B **147** no. 1,  
3530 (1984) 189–196.  
3531 <https://www.sciencedirect.com/science/article/pii/037026938490618X>.  
3532 20
- 3533 [33] N. Cabibbo, L. Maiani, and Y. Srivastava, *Anomalous Z decays: excited leptons?*,  
3534 Physics Letters B **139** no. 5, (1984) 459–463.  
3535 <https://www.sciencedirect.com/science/article/pii/0370269384918501>.  
3536 20
- 3537 [34] A. De Rújula, L. Maiani, and R. Petronzio, *Search for excited quarks*, Phys. Lett. B  
3538 **140** (1984) 253–258. <https://cds.cern.ch/record/149662>. 20
- 3539 [35] U. Baur, M. Spira, and P. M. Zerwas, *Excited-quark and -lepton production at*  
3540 *hadron colliders*, Phys. Rev. D **42** (1990) 815–824.  
3541 <https://link.aps.org/doi/10.1103/PhysRevD.42.815>. 20
- 3542 [36] S. Bhattacharya, S. S. Chauhan, B. C. Choudhary, and D. Choudhury, *Quark*  
3543 *excitations through the prism of direct photon plus jet at the LHC*, Phys. Rev. D **80**  
3544 (2009) 015014. <https://link.aps.org/doi/10.1103/PhysRevD.80.015014>. 20
- 3545 [37] Y. C. Zhan, C. S. Li, Z. L. Liu, and S. A. Li, *Signature of the  $\gamma$  + jet and dijet*  
3546 *production mediated by an excited quark with QCD next-to-leading order accuracy at*  
3547 *the LHC*, Phys. Rev. D **93** (2016) 014018.  
3548 <https://link.aps.org/doi/10.1103/PhysRevD.93.014018>. 20, 21, 96
- 3549 [38] F. Zwicky, *On the Masses of Nebulae and of Clusters of Nebulae*, ApJ **86** (1937) 217.  
3550 <https://ui.adsabs.harvard.edu/abs/1937ApJ....86..217Z/abstract>. 20
- 3551 [39] V. C. Rubin and J. Ford, W. Kent, *Rotation of the Andromeda Nebula from a*  
3552 *Spectroscopic Survey of Emission Regions*, ApJ **159** (1970) 379.  
3553 <https://ui.adsabs.harvard.edu/abs/1970ApJ...159..379R/abstract>. 20
- 3554 [40] Planck Collaboration, *Planck 2013 results. XVI. Cosmological parameters*,  
3555 *Astronomy and Astrophysics* **571** (2014), arXiv:1303.5076.  
3556 <https://arxiv.org/abs/1303.5076v3>. 20
- 3557 [41] D. Clowe, M. Bradač, A. H. Gonzalez, M. Markevitch, S. W. Randall, et al., *A Direct*  
3558 *Empirical Proof of the Existence of Dark Matter*, The Astrophysical Journal **648**  
3559 *no. 2, (2006) L109–L113*, arXiv:0608407 [astro-ph].  
3560 <https://iopscience.iop.org/article/10.1088/0004-637X/648/2/L109>. 20
- 3561 [42] M. Bradač, S. W. Allen, T. Treu, H. Ebeling, R. Massey, R. G. Morris, A. von der  
3562 Linden, D. Applegate, and ;, *Revealing the properties of dark matter in the merging*  
3563 *cluster MACSJ0025.4-1222*, The Astrophysical Journal **687** no. 2, (2008)  
3564 *959–967*, arXiv:0806.2320v2. <http://dx.doi.org/10.1088/0004-637X/687/2/959>. 20

- 3565 [43] M. Gell-Mann, P. Ramond, and R. Slansky, *Complex spinors and unified theories*, in  
3566 *Murray Gell-Mann: Selected Papers*. World Scientific, 2010. [21](#)
- 3567 [44] S. L. Glashow, *The future of elementary particle physics*, in *Quarks and Leptons: Cargese 1979*. Springer, 1980. [21](#)
- 3569 [45] P. Ramond, *The family group in grand unified theories*, in *SEESAW 25*. World  
3570 Scientific, 2005. [21](#)
- 3571 [46] *Proceedings of the 1981 international symposium on lepton and photon interactions at high energies*. 1981. [21](#)
- 3573 [47] G. Hooft, *Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking*, in *Recent developments in gauge theories*. Springer, 1980. [21](#)
- 3575 [48] J. Pati, A. Salam, and J. Strathdee, *Are quarks composite?*, *Physics Letters B* **59**  
3576 no. 3, (1975) 265–268.  
3577 <https://www.sciencedirect.com/science/article/pii/0370269375900428>.  
3578 [21](#)
- 3579 [49] H. Fritzsch and G. Mandelbaum, *Weak interactions as manifestations of the substructure of leptons and quarks*, *Physics Letters B* **102** no. 5, (1981) 319–322.  
3580 <https://www.sciencedirect.com/science/article/pii/0370269381906262>.  
3581 [21](#)
- 3583 [50] U. Baur and H. Fritzsch, *The masses of composite quarks and leptons as electromagnetic self energies*, *Physics Letters B* **134** no. 1, (1984) 105–110.  
3584 <https://www.sciencedirect.com/science/article/pii/037026938490995X>.  
3585 [21](#)
- 3587 [51] D. Bailin and A. Love, *Kaluza-klein theories*, *Reports on Progress in Physics* **50**  
3588 no. 9, (1987) 1087. [22](#)
- 3589 [52] N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, *The hierarchy problem and new dimensions at a millimeter*, *Physics Letters B* **429** no. 3-4, (1998) 263–272. [22](#), [23](#)
- 3591 [53] L. Randall and R. Sundrum, *An alternative to compactification*, *Physical Review Letters* **83** no. 23, (1999) 4690. [23](#)
- 3593 [54] L. Randall and R. Sundrum, *Large mass hierarchy from a small extra dimension*, *Physical review letters* **83** no. 17, (1999) 3370. [23](#)
- 3595 [55] I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, *New dimensions at a millimeter to a Fermi and superstrings at a TeV*, *Physics Letters B* **436** no. 3-4, (1998) 257–263. [23](#)

- 3598 [56] G. Dvali and M. Redi, *Black hole bound on the number of species and quantum*  
3599 *gravity at CERN LHC*, Physical Review D—Particles, Fields, Gravitation, and  
3600 Cosmology **77** no. 4, (2008) 045027. [23](#)
- 3601 [57] G. Dvali, *Black holes and large N species solution to the hierarchy problem*,  
3602 *Fortschritte der Physik* **58** no. 6, (2010) 528–536.  
3603 <http://dx.doi.org/10.1002/prop.201000009>. [23](#)
- 3604 [58] P. C. Argyres, S. Dimopoulos, and J. March-Russell, *Black holes and sub-millimeter*  
3605 *dimensions*, Physics Letters B **441** no. 1-4, (1998) 96–104. [23](#)
- 3606 [59] T. Banks and W. Fischler, *A Model for High Energy Scattering in Quantum Gravity*,  
3607 1999. <https://arxiv.org/abs/hep-th/9906038>. [23](#)
- 3608 [60] S. B. Giddings and S. Thomas, *High energy colliders as black hole factories: The end*  
3609 *of short distance physics*, Physical Review D **65** no. 5, (2002) 056010. [23](#)
- 3610 [61] P. Meade and L. Randall, *Black holes and quantum gravity at the LHC*, Journal of  
3611 High Energy Physics **2008** no. 05, (2008) 003. [23](#)
- 3612 [62] G. L. Alberghi, R. Casadio, D. Galli, D. Gregori, A. Tronconi, and V. Vagnoni,  
3613 *Probing quantum gravity effects in black holes at LHC*, 2006.  
3614 <https://arxiv.org/abs/hep-ph/0601243>. [23](#)
- 3615 [63] G. L. Alberghi, R. Casadio, and A. Tronconi, *Quantum gravity effects in black holes*  
3616 *at the LHC*, Journal Of Physics G: Nuclear and Particle Physics **34** no. 4, (2007)  
3617 767. [23](#)
- 3618 [64] D. M. Gingrich and B. Undseth, *Quantum black holes in the horizon quantum*  
3619 *mechanics model at the Large Hadron Collider*, Phys. Rev. D **102** (2020) 095020.  
3620 <https://link.aps.org/doi/10.1103/PhysRevD.102.095020>. [23](#)
- 3621 [65] D. M. Gingrich, *Quantum black holes with charge, color and spin at the LHC*,  
3622 Journal of Physics G: Nuclear and Particle Physics **37** no. 10, (2010) 105008. [23](#)
- 3623 [66] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*,  
3624 *Computer Physics Communications* **178** no. 11, (2008) 852–867.  
3625 <https://www.sciencedirect.com/science/article/pii/S0010465508000441>.  
3626 [24](#), [70](#), [98](#)
- 3627 [67] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna,  
3628 S. Prestel, C. O. Rasmussen, and P. Z. Skands, *An introduction to PYTHIA 8.2*,  
3629 *Computer Physics Communications* **191** (2015) 159–177.  
3630 <https://www.sciencedirect.com/science/article/pii/S0010465515000442>.  
3631 [24](#), [96](#)

- 3632 [68] C. Bierlich, S. Chakraborty, N. Desai, L. Gellersen, I. Helenius, P. Ilten,  
3633 L. Lönnblad, S. Mrenna, S. Prestel, C. T. Preuss, T. Sjöstrand, P. Skands,  
3634 M. Utheim, and R. Verheyen, *A comprehensive guide to the physics and usage of*  
3635 *PYTHIA 8.3*, 2022. <https://arxiv.org/abs/2203.11601>. 24
- 3636 [69] E. Bothmann, G. S. Chahal, S. Höche, J. Krause, F. Krauss, S. Kuttimalai,  
3637 S. Liebschner, D. Napoletano, M. Schönher, H. Schulz, S. Schumann, and  
3638 F. Siegert, *Event generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) 034.  
3639 <https://scipost.org/10.21468/SciPostPhys.7.3.034>. 24, 70
- 3640 [70] T. Sjöstrand, *Monte Carlo Generators*, 2006.  
3641 <https://arxiv.org/abs/hep-ph/0611247>. 25
- 3642 [71] M. A. Dobbs, S. Frixione, E. Laenen, et al., *Les Houches Guidebook to Monte Carlo*  
3643 *Generators for Hadron Collider Physics*, 2004.  
3644 <https://arxiv.org/abs/hep-ph/0403045>. 25
- 3645 [72] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, *Parton fragmentation*  
3646 *and string dynamics*, *Physics Reports* **97** no. 2, (1983) 31–145.  
3647 <https://www.sciencedirect.com/science/article/pii/0370157383900807>.  
3648 25, 98
- 3649 [73] B. Webber, *A QCD model for jet fragmentation including soft gluon interference*,  
3650 *Nuclear Physics B* **238** no. 3, (1984) 492–528.  
3651 <https://www.sciencedirect.com/science/article/pii/055032138490333X>.  
3652 25
- 3653 [74] T. A. collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, tech. rep., CERN, Geneva,  
3654 2014. <https://cds.cern.ch/record/1966419>. 26, 98
- 3655 [75] P. Skands, S. Carrazza, and J. Rojo, *Tuning PYTHIA 8.1: the Monash 2013 tune*,  
3656 *The European Physical Journal C* **74** no. 8, (2014) 3024.  
3657 <https://doi.org/10.1140/epjc/s10052-014-3024-y>. 26
- 3658 [76] S. Agostinelli and J. Allison and K. Amako and J. Apostolakis and H. Araujo and  
3659 others, *Geant4—a simulation toolkit*, *Nuclear Instruments and Methods in Physics*  
3660 *Research Section A: Accelerators, Spectrometers, Detectors and Associated*  
3661 *Equipment* **506** no. 3, (2003) 250–303.  
3662 <https://www.sciencedirect.com/science/article/pii/S0168900203013688>.  
3663 26, 27, 70
- 3664 [77] ATLAS Collaboration, *AtFast3: The Next Generation of Fast Simulation in ATLAS*,  
3665 *Computing and Software for Big Science* **6** no. 1, (2022) 7.  
3666 <https://doi.org/10.1007/s41781-021-00079-7>. 27

- 3667 [78] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, The European Physical  
3668 Journal C **70** no. 3, (2010) 823–874.  
3669 <https://doi.org/10.1140/epjc/s10052-010-1429-9>. 27
- 3670 [79] C. ATLAS, M. Beckingham, M. Duehrssen, E. Schmidt, M. Shapiro, M. Venturi,  
3671 J. Virzi, I. Vivarelli, M. Werner, S. Yamamoto, and T. Yamanaka, *The simulation*  
3672 *principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*,  
3673 tech. rep., CERN, Geneva, 2010. <https://cds.cern.ch/record/1300517>. 27
- 3674 [80] O. S. Brüning, J. Poole, P. Collier, P. Lebrun, R. Ostojic, S. Myers, and P. Proudlock,  
3675 *LHC Design Report*, CERN, Geneva **1** (2004) 548.  
3676 <https://cds.cern.ch/record/782076>. 29
- 3677 [81] E. Lopienska, *The CERN accelerator complex, layout in 2022. Complexe des*  
3678 *accélérateurs du CERN en janvier 2022.*, <https://cds.cern.ch/record/2800984>.  
3679 xi, 30
- 3680 [82] The ALICE Collaboration, *The ALICE experiment at the CERN LHC*, Journal of  
3681 Instrumentation **3** no. 08, (2008) S08002.  
3682 <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08002>. 30
- 3683 [83] The LHCb Collaboration, *The LHCb Detector at the LHC*, Journal of  
3684 Instrumentation **3** no. 08, (2008) S08005. <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08005>  
3685 <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08005/meta>. 30
- 3687 [84] The CMS Collaboration, *The CMS experiment at the CERN LHC*, Journal of  
3688 Instrumentation **3** no. 08, (2008) S08004. <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08004>  
3689 <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08004/meta>. 30
- 3691 [85] The ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron  
3692 Collider*, Journal of Instrumentation **3** no. 08, (2008) S08003.  
3693 <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08003>. xi,  
3694 xix, 30, 32, 33, 37, 39
- 3695 [86] The LHCf Collaboration, *The LHCf detector at the CERN Large Hadron Collider*,  
3696 Journal of Instrumentation **3** no. 08, (2008) S08006. <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08006>  
3697 <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08006/meta>. 30
- 3699 [87] The TOTEM Collaboration, *The TOTEM Experiment at the CERN Large Hadron  
3700 Collider*, Journal of Instrumentation **3** no. 08, (2008) S08007. <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08007>  
3701 <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08007/meta>. 30

- 3703 [88] MoEDAL Collaboration, *Technical Design Report of the MoEDAL Experiment.*, 30
- 3704 [89] R. M. Bianchi and A. Collaboration, “ATLAS experiment schematic or layout  
3705 illustration.” General Photo, 2022. xi, 31, 32
- 3706 [90] J. Pequenao, “Computer generated image of the ATLAS inner detector.” General  
3707 Photo, 2008. 33
- 3708 [91] The ATLAS Collaboration, *Inner Detector Forward SCT Barrel SCT TRT Pixel*  
3709 *Detectors Technical Design Report*, tech. rep., CERN, 1997.  
3710 <https://cds.cern.ch/record/331063/files/ATLAS-TDR-4-Volume-I.pdf>. 33
- 3711 [92] T. A. Collaboration, *ATLAS Insertable B-Layer Technical Design Report Addendum*,  
3712 Tech. Rep. May, CERN, Sep, 2012. <https://cds.cern.ch/record/1291633>  
<https://cdsweb.cern.ch/record/1451888>. 34
- 3714 [93] K. Potamianos, *The upgraded Pixel detector and the commissioning of the Inner*  
3715 *Detector tracking of the ATLAS experiment for Run-2 at the Large Hadron Collider*,  
3716 [arXiv:1608.07850](https://arxiv.org/abs/1608.07850). <https://arxiv.org/abs/1608.07850v1>. 34
- 3717 [94] The ATLAS Collaboration, *ATLAS pixel detector electronics and sensors*, *Journal of*  
3718 *Instrumentation* 3 no. 07, (2008) P07007. <https://iopscience.iop.org/article/10.1088/1748-0221/3/07/P07007>  
3719 <https://iopscience.iop.org/article/10.1088/1748-0221/3/07/P07007/meta>. 34
- 3721 [95] T. Heim, *Status and performance of the ATLAS Pixel Detector after 3 years of*  
3722 *operation*, *Nuclear Instruments and Methods in Physics Research Section A:*  
3723 *Accelerators, Spectrometers, Detectors and Associated Equipment* 765 (2014)  
3724 227–231. 34
- 3725 [96] A. Ahmad, Z. Albrechtskirchinger, P. P. Allport, J. Alonso, et al., *The silicon*  
3726 *microstrip sensors of the ATLAS semiconductor tracker*, *Nuclear Instruments and*  
3727 *Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors*  
3728 *and Associated Equipment* 578 no. 1, (2007) 98–118. 34
- 3729 [97] The ATLAS Collaboration, *The ATLAS Transition Radiation Tracker (TRT)*  
3730 *proportional drift tube: design and performance*, *Journal of Instrumentation* 3  
3731 no. 02, (2008) P02013. <https://iopscience.iop.org/article/10.1088/1748-0221/3/02/P02013>  
3732 <https://iopscience.iop.org/article/10.1088/1748-0221/3/02/P02013/meta>. 34
- 3734 [98] J. Pequenao, “Computer Generated image of the ATLAS calorimeter.” General  
3735 Photo, 2008. xi, 35
- 3736 [99] V. Rossetti, *Performance of the ATLAS Calorimeters and Commissioning for LHC*  
3737 *Run-2*, <https://cds.cern.ch/record/2037117>. 36

- 3738 [100] The ATLAS Collaboration, *ATLAS tile calorimeter: Technical Design Report*,  
3739 CERN/LHCC96-42(1996) (1996) 354. <https://cds.cern.ch/record/331062>.  
3740 38
- 3741 [101] T. A. collaboration, *ATLAS muon spectrometer : Technical Design Report*,  
3742 <https://cds.cern.ch/record/331068>. 38
- 3743 [102] B. Stelzer, *The New Small Wheel Upgrade Project of the ATLAS Experiment*, Nuclear  
3744 and Particle Physics Proceedings 273-275 (2016) 1160–1165.  
3745 <https://www.sciencedirect.com/science/article/pii/S2405601415006719>.  
3746 39
- 3747 [103] The ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2010*, The  
3748 European Physical Journal C 72 no. 1, (2012) 1849. 39
- 3749 [104] A. ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2015*,  
3750 arXiv:1611.09661. <http://arxiv.org/abs/1611.09661> http:  
3751 //dx.doi.org/10.1140/epjc/s10052-017-4852-3. 39
- 3752 [105] The ATLAS Collaboration, *Operation of the ATLAS trigger system in Run 2*, Journal  
3753 of Instrumentation 15 no. 10, (2020) P10004, arXiv:2007.12539v2. 39
- 3754 [106] R. Achenbach, P. Adragna, V. Andrei, P. Apostologlou, B. Åsman, C. Ay, B. M.  
3755 Barnett, B. Bauss, et al., *The ATLAS Level-1 Calorimeter Trigger*, Journal of  
3756 Instrumentation 3 no. 03, (2008) P03001.  
3757 <https://dx.doi.org/10.1088/1748-0221/3/03/P03001>. 40
- 3758 [107] ATLAS Collaboration, P. Jenni, M. Nessi, M. Nordberg, and K. Smith, *ATLAS*  
3759 *high-level trigger, data-acquisition and controls: Technical Design Report*. Technical  
3760 design report. ATLAS. CERN, Geneva, 2003.  
3761 <https://cds.cern.ch/record/616089>. 40
- 3762 [108] The ATLAS Collaboration, *Luminosity determination in pp collisions at  $\sqrt{s} = 13$  TeV*  
3763 *using the ATLAS detector at the LHC*, The European Physical Journal C 83 no. 10,  
3764 (2023) 982. xi, 42, 118
- 3765 [109] G. Avoni, M. Bruschi, G. Cabras, D. Caforio, N. Dehghanian, A. Floderus,  
3766 B. Giacobbe, F. Giannuzzi, et al., *The new LUCID-2 detector for luminosity*  
3767 *measurement and monitoring in ATLAS*, Journal of Instrumentation 13 no. 07,  
3768 (2018) P07017. <https://dx.doi.org/10.1088/1748-0221/13/07/P07017>. 42,  
3769 118
- 3770 [110] ATLAS Collaboration, T. A. collaboration, *Preliminary analysis of the luminosity*  
3771 *calibration of the ATLAS 13.6 TeV data recorded in 2022*, tech. rep., CERN, Geneva,  
3772 2023. <https://cds.cern.ch/record/2853525>. 42

- 3773 [111] ATLAS Collaboration, T. A. collaboration, *Preliminary analysis of the luminosity*  
3774 *calibration for the ATLAS 13.6 TeV data recorded in 2023*, tech. rep., CERN,  
3775 Geneva, 2024. <https://cds.cern.ch/record/2900949>. 42
- 3776 [112] The ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms*  
3777 *in dense environments in LHC Run 2*, The European Physical Journal C **77** no. 10,  
3778 (2017) 673. 45
- 3779 [113] T. G. Cornelissen, N. Van Eldik, M. Elsing, W. Liebig, E. Moyse, N. Piacquadio,  
3780 K. Prokofiev, A. Salzburger, and A. Wildauer, *Updates of the ATLAS Tracking Event*  
3781 *Data Model (Release 13)*, tech. rep., CERN, Geneva, 2007.  
3782 <https://cds.cern.ch/record/1038095>. xi, 45
- 3783 [114] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig, E. Moyse, and A. Salzburger,  
3784 *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, tech.  
3785 rep., CERN, Geneva, 2007. <https://cds.cern.ch/record/1020106>. 45
- 3786 [115] T. A. collaboration, *A neural network clustering algorithm for the ATLAS silicon*  
3787 *pixel detector*, Journal of Instrumentation **9** no. 09, (2014) P09009.  
3788 <https://dx.doi.org/10.1088/1748-0221/9/09/P09009>. 45
- 3789 [116] The ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS*  
3790 *experiment in Run 1 proton–proton collisions at the LHC*, The European Physical  
3791 Journal C **77** no. 5, (2017) 332. 46
- 3792 [117] T. A. collaboration, *Vertex Reconstruction Performance of the ATLAS Detector at*  
3793  $\sqrt{s} = 13 \text{ TeV}$ , tech. rep., CERN, Geneva, 2015.  
3794 <https://cds.cern.ch/record/2037717>. 46
- 3795 [118] The ATLAS Collaboration, *Electron and photon performance measurements with the*  
3796 *ATLAS detector using the 2015–2017 LHC proton-proton collision data*, Journal of  
3797 Instrumentation **14** no. 12, (2019) P12006.  
3798 <https://dx.doi.org/10.1088/1748-0221/14/12/P12006>. xi, 46, 47, 66, 68,  
3799 122
- 3800 [119] ATLAS Collaboration, T. A. collaboration, *Electron and photon reconstruction and*  
3801 *performance in ATLAS using a dynamical, topological cell clustering-based approach*,  
3802 tech. rep., CERN, Geneva, 2017. <https://cds.cern.ch/record/2298955>. 46, 52
- 3803 [120] The ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and*  
3804 *its performance in LHC Run 1*, The European Physical Journal C **77** no. 7, (2017)  
3805 490. 46, 52
- 3806 [121] The ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector*  
3807 *in proton–proton collision data at  $\sqrt{s}=13 \text{ TeV}$* , The European Physical Journal C **76**  
3808 no. 5, (2016) 292. 50

- 3809 [122] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- $k_T$  jet clustering algorithm*, *Journal*  
3810 *of High Energy Physics* **2008** no. 04, (2008) 063.  
3811 <https://dx.doi.org/10.1088/1126-6708/2008/04/063>. **xi**, **51**, **52**
- 3812 [123] The ATLAS Collaboration, *Jet reconstruction and performance using particle flow*  
3813 *with the ATLAS Detector*, *The European Physical Journal C* **77** no. 7, (2017) 466.  
3814 **53**
- 3815 [124] ATLAS Collaboration, T. A. collaboration,  *$E_T^{\text{miss}}$  performance in the ATLAS detector*  
3816 *using 2015-2016 LHC p-p collisions*, tech. rep., CERN, Geneva, 2018.  
3817 <https://cds.cern.ch/record/2625233>. **53**
- 3818 [125] The ATLAS Collaboration, *Jet energy scale and resolution measured in*  
3819 *proton–proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *The European*  
3820 *Physical Journal C* **81** no. 8, (2021) 689. **xi**, **53**, **54**
- 3821 [126] T. A. collaboration, *Calibration of the performance of b-tagging for c and*  
3822 *light-flavour jets in the 2012 ATLAS data*, tech. rep., CERN, Geneva, 2014.  
3823 <https://cds.cern.ch/record/1741020>. **54**
- 3824 [127] ATLAS Collaboration, T. A. collaboration, *Measurement of the b-tag Efficiency in a*  
3825 *Sample of Jets Containing Muons with  $5 \text{ fb}^{-1}$  of Data from the ATLAS Detector*, tech.  
3826 rep., CERN, Geneva, 2012. <https://cds.cern.ch/record/1435197>. **54**
- 3827 [128] A. Lavrenov, *MV2-algorithm’s clones*, CoRR **cs.CR/0208003** (2002).  
3828 <https://arxiv.org/abs/cs/0208003>. **54**
- 3829 [129] M. Lanfermann, *Deep Learning in Flavour Tagging at the ATLAS experiment*, PoS  
3830 **EPS-HEP2017** (2018) **764**. **54**
- 3831 [130] The ATLAS Collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp*  
3832 *collision dataset*, *The European Physical Journal C* **83** no. 7, (2023) 681. **54**
- 3833 [131] ATLAS Collaboration, T. A. collaboration, *Graph Neural Network Jet Flavour*  
3834 *Tagging with the ATLAS Detector*, tech. rep., CERN, Geneva, 2022.  
3835 <https://cds.cern.ch/record/2811135>. **54**
- 3836 [132] S. Brody, U. Alon, and E. Yahav, *How Attentive are Graph Attention Networks?*,  
3837 CoRR **abs/2105.14491** (2021), **2105.14491**.  
3838 <https://arxiv.org/abs/2105.14491>. **54**
- 3839 [133] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser,  
3840 and I. Polosukhin, *Attention Is All You Need*, CoRR **abs/1706.03762** (2017),  
3841 **1706.03762**. <http://arxiv.org/abs/1706.03762>. **54**
- 3842 [134] Z. Wu, “b-jets calibration in atlas.” ATLAS-CMS Flavour Tagging Workshop, 2024.  
3843 **xi**, **56**

- 3844 [135] I. Wingerter-Seez, *Particle Physics Instrumentation*, arXiv:1804.11246.  
3845 <https://cds.cern.ch/record/2315747>. xi, 62
- 3846 [136] ATLAS Collaboration, *Search for short- and long-lived axion-like particles in*  
3847  *$H \rightarrow aa \rightarrow 4\gamma$  decays with the ATLAS experiment at the LHC*, The European  
3848 Physical Journal C **84** no. 7, (2024) 742.  
3849 <https://doi.org/10.1140/epjc/s10052-024-12979-0>. 63
- 3850 [137] The ATLAS Collaboration, *Electron and photon efficiencies in LHC Run 2 with the*  
3851 *ATLAS experiment*, Journal of High Energy Physics **2024** no. 5, (2024) 162. 65
- 3852 [138] The ATLAS Collaboration, *Measurement of the photon identification efficiencies with*  
3853 *the ATLAS detector using LHC Run 2 data collected in 2015 and 2016*, The  
3854 European Physical Journal C **79** no. 3, (2019) 205. xii, 66, 67, 68
- 3855 [139] L. Devroye, *Non-Uniform Random Variate Generation*. Springer New York, NY,  
3856 1986. 66
- 3857 [140] L. Lyons, D. Gibaut, and P. Clifford, *How to Combine Correlated Estimates of a*  
3858 *Single Physical Quantity*, Nucl. Instrum. Meth. **A270** (1988) 110. 67
- 3859 [141] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS*  
3860 *detector using 2015–2016 LHC proton–proton collision data*, JINST **14** (2019)  
3861 P03017, arXiv:1812.03848 [hep-ex]. xii, 69
- 3862 [142] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, et al., *TMVA -*  
3863 *Toolkit for Multivariate Data Analysis*, 2007. arXiv:physics/0703039  
3864 [physics.data-an]. 71
- 3865 [143] M. Khandoga, *Calibration of electron shower shapes, hadronic recoil reconstruction*  
3866 *using deep learning algorithms and the measurement of  $W$  boson transverse*  
3867 *momentum distribution with the ATLAS detector*. PhD thesis, U. Paris-Saclay, 2020.  
3868 <https://cds.cern.ch/record/2752635>. 79
- 3869 [144] M. Belfkir, *Search for Higgs pair production at LHC collider (CERN): The first*  
3870 *measurement for Higgs potential and search for new physics*. PhD thesis, Savoie  
3871 Mont Blanc University, 2021. <https://cds.cern.ch/record/2792268>. 79, 81, 84
- 3872 [145] J. L. Spah, *Data-driven corrections to shower shape variables for photon*  
3873 *identification at the ATLAS experiment*, Master's thesis, TU Dortmund University,  
3874 Dortmund, 2021. <https://kroeninger-group.physik.tu-dortmund.de/en/research/particle-physics/theses/#c135884>. 85
- 3876 [146] G. Choudalakis, *On hypothesis testing, trials factor, hypertests and the BumpHunter*,  
3877 2011. <https://arxiv.org/abs/1101.0390>. 92

- 3878 [147] L. Vaslin, S. Calvet, V. Barra, and J. Donini, *pyBumpHunter: A model independent*  
3879 *bump hunting tool in Python for high energy physics analyses*, *SciPost Phys.*  
3880 *Codebases* (2023) **15**. <https://scipost.org/10.21468/SciPostPhysCodeb.15.92>
- 3882 [148] L. Lyons, *Open statistical issues in Particle Physics*, *The Annals of Applied Statistics*  
3883 **2** no. 3, (2008) 887 – 915. <https://doi.org/10.1214/08-AOAS163>. 93
- 3884 [149] NNPDF Collaboration, R. D. Ball, et al., *Parton distributions with LHC data*, *Nucl.*  
3885 *Phys. B* **867** (2013) 244, [arXiv:1207.1303 \[hep-ph\]](https://arxiv.org/abs/1207.1303). 96, 98, 119
- 3886 [150] D. M. Gingrich, *Monte Carlo event generator for black hole production and decay in*  
3887 *proton-proton collisions*, *Comput.Phys.Commun.* **181** (2010) 1917–1924,  
3888 [arXiv:0911.5370 \[hep-ph\]](https://arxiv.org/abs/0911.5370). 96
- 3889 [151] G. Pasztor, R. White, F. Monticelli, J. Hoya, and H. Wahlberg, *Photon trigger*  
3890 *performance in 2015 ATLAS data*, Tech. Rep. ATL-COM-DAQ-2015-101, CERN,  
3891 Geneva, Jul, 2015. <https://cds.cern.ch/record/2034866>. For upcoming  
3892 conferences: EPS, DPF and LISHEP. 100
- 3893 [152] ATLAS Collaboration, *Measurement of the cross section for isolated-photon plus jet*  
3894 *production in pp collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector*, *Phys. Lett. B*  
3895 **780** (2018) 578, [arXiv:1801.00112 \[hep-ex\]](https://arxiv.org/abs/1801.00112). 104
- 3896 [153] M. Baak, S. Gadatsch, R. Harrington, and W. Verkerke, *Interpolation between*  
3897 *multi-dimensional histograms using a new non-linear moment morphing method*,  
3898 *Nuclear Instruments and Methods in Physics Research Section A: Accelerators,*  
3899 *Spectrometers, Detectors and Associated Equipment* **771** (2015) 39–48.  
3900 <https://www.sciencedirect.com/science/article/pii/S0168900214011814>.  
3901 113
- 3902 [154] W. Buttinger, *Using Event Weights to account for differences in Instantaneous*  
3903 *Luminosity and Trigger Prescale in Monte Carlo and Data*, tech. rep., CERN,  
3904 Geneva, 2015. <https://cds.cern.ch/record/2014726>. 118
- 3905 [155] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in*  
3906 *proton–proton collisions at  $\sqrt{s} = 7$  TeV*, *Eur. Phys. J. C* **73** (2013) 2304,  
3907 [arXiv:1112.6426 \[hep-ex\]](https://arxiv.org/abs/1112.6426). 119
- 3908 [156] ATLAS Collaboration, “Monte Carlo Calibration and Combination of In-situ  
3909 Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS.”  
3910 ATLAS-CONF-2015-037, 2015. <https://cds.cern.ch/record/2044941>. 119
- 3911 [157] M. Aaboud, G. Aad, B. Abbott, O. Abdinov, B. Abeloos, S. H. Abidi, O. AbouZeid,  
3912 N. L. Abraham, H. Abramowicz, H. Abreu, et al., *Search for photonic signatures of*  
3913 *gauge-mediated supersymmetry in 13 TeV pp collisions with the ATLAS detector*,  
3914 *Physical review D* **97** no. 9, (2018) 092006. 122

- 3915 [158] F. Alonso, F. A. Arduh, J. Hoya, M.-T. Dova, H. Wahlberg, and G. E. Orellana,  
3916 *Search for supersymmetry in events with photons, jets and missing transverse*  
3917 *momentum with  $36 \text{ fb}^{-1}$  of data at  $13 \text{ TeV}$* , Tech. Rep. ATL-COM-PHYS-2016-1662,  
3918 CERN, Geneva, Nov, 2016. <https://cds.cern.ch/record/2233238>. 122
- 3919 [159] G. Aad, B. Abbott, D. Abbott, A. Abed Abud, K. Abeling, D. K. Abhayasinghe,  
3920 S. Abidi, A. Aboulhorma, H. Abramowicz, H. Abreu, et al., *Search for new*  
3921 *phenomena in final states with photons, jets and missing transverse momentum in  $pp$*   
3922 *collisions at  $\sqrt{s} = 13 \text{ TeV}$  with the ATLAS detector*, Journal of High Energy Physics  
3923 **2023** no. 7, (2023) 1–42. 122
- 3924 [160] F. Alonso, F. A. Arduh, M.-T. Dova, J. Hoya, G. E. Orellana, and H. Wahlberg,  
3925 *Search for SUSY signatures in events with photons, missing transverse energy and jets*  
3926 *using  $140 \text{ fb}^{-1}$* , tech. rep., CERN, Geneva, Oct, 2018.  
3927 <https://cds.cern.ch/record/2641602>. 122
- 3928 [161] ATLAS Collaboration, *Search for new phenomena in high-mass final states with a*  
3929 *photon and a jet from  $pp$  collisions at  $\sqrt{s} = 13 \text{ TeV}$  with the ATLAS detector*, Eur.  
3930 Phys. J. C **78** (2018) 102, arXiv:1709.10440 [hep-ex]. 122, 128
- 3931 [162] A. Collaboration et al., *Search for dark matter at  $\sqrt{s} = 13 \text{ TeV}$  in final states*  
3932 *containing an energetic photon and large missing transverse momentum with the*  
3933 *ATLAS detector*, Eur. Phys. J. C **77** (2017) 393. 122
- 3934 [163] S. Laplace, M. A. Chelstowska, S. M. Mazza, J. Ocariz, J. Qian, J. Poveda,  
3935 N. Proklova, Y.-T. Shen, G. Unal, M. Villaplana Perez, J.-B. De Vivie De Regie,  
3936 Z. Wang, and K. Todome, *Search for new phenomena in diphoton events with the*  
3937 *ATLAS detector at  $\sqrt{s} = 13 \text{ TeV}$ : Isolation studies*, tech. rep., CERN, Geneva,  
3938 2016. <https://cds.cern.ch/record/2162758>. 125
- 3939 [164] R. Hyneman, S. Laplace, J. Ocariz, J. Poveda, N. Proklova, J.-B. De Vivie De Regie,  
3940 M. Zgubic, O. A. Ducu, M. H. Klein, M. J. Alconada Verzini, F. Alonso, A. J. Taylor,  
3941 C. Merlassino, L. Pascual Dominguez, E. Varnes, J. Maurer, and J. Reichert,  
3942 *Electron, Muon, and Photon Isolation in 2015-2018*, tech. rep., CERN, Geneva,  
3943 2019. <https://cds.cern.ch/record/2672803>. 125
- 3944 [165] ATLAS Collaboration, *Search for new resonances in mass distributions of jet pairs*  
3945 *using  $139 \text{ fb}^{-1}$  of  $pp$  collisions at  $\sqrt{s} = 13 \text{ TeV}$  with the ATLAS detector*, JHEP **03**  
3946 (2020) 145, arXiv:1910.08447 [hep-ex]. 128
- 3947 [166] R. H. oy and J. Flusser, *Numerically Stable Direct Least Squares Fitting of Ellipses*,  
3948 1998. <https://api.semanticscholar.org/CorpusID:15772208>. 160