Studies of photon isolation in a large-pileup environment and search for axion-like particles decaying into two photons with the ATLAS detector at the LHC

Luis Pascual Dominguez

UNIVERSITÉ DE PARIS

Thèse de doctorat
pour obtenir le grade de Docteur en Sciences de l'Université de Paris
Mention: Physique de l'Univers

École doctorale des Sciences de la Terre et de l'environnement et Physique de l'Univers, Paris - ED 560
Laboratoire de Physique Nucléaire et de Hautes Énergies - UMR 7585

Études sur l'isolement des photons dans l'environnement à haut empilement au LHC, et recherche de particules de type axion se désintégrant en deux photons avec le détecteur ATLAS

Thèse présentée par

Luis Pascual Domínguez

Dirigée par
José Ocariz

Soutenue publiquement le 29 janvier 2021 devant le jury composé de:

Pierre Antilogus Directeur de Recherche (LPNHE - Sorbonne Université)	Examinateur
María Luisa Cepeda Chercheuse (CIEMAT)	Examinatrice
Fabrice Couderc Directeur de Recherche (CEA - Université Paris-Saclay)	Rapporteur
François Le Diberder Professeur (IJCLab - Université de Paris)	Examinateur
José Ocariz Professeur (LPNHE - Université de Paris)	Directeur de thèse
Elisabeth Petit Chargée de Recherche (CPPM - Université Aix-Marseille)	Examinatrice
Joaquín Poveda Chercheur (IFIC - Universitat de Valencia)	Rapporteur
Gregory Soyez Directeur de Recherche (IPhT - Université Paris-Saclay)	Examinateur

UNIVERSITÉ DE PARIS

PhD Thesis

Submitted in fulfillment of the requirements for the degree of
Docteur en Sciences de l'Université de Paris
Specialty: Physics of the Universe

École doctorale des Sciences de la Terre et de l'environnement et Physique de l'Univers, Paris - ED 560
Laboratoire de Physique Nucléaire et de Hautes Énergies - UMR 7585

Studies of photon isolation in a large-pileup environment and search for axion-like particles decaying into two photons with the ATLAS detector at the LHC

Presented by

Luis Pascual Domínguez

Supervised by José Ocariz

Defended on the 29^{th} of January 2021 in front of the jury:

Pierre Antilogus	Examiner
Research Director (LPNHE - Sorbonne Université)	
María Luisa Cepeda	Examiner
Researcher (CIEMAT)	
Fabrice Couderc	Referee
Research Director (CEA - Université Paris-Saclay)	
François Le Diberder	Examiner
Professor (IJCLab - Université de Paris)	
José Ocariz	Supervisor
Professor (LPNHE - Université de Paris)	
Elisabeth Petit	Examiner
Researcher (CPPM - Université Aix-Marseille)	
Joaquín Poveda	Referee
Researcher (IFIC - Universitat de Valencia)	
Gregory Soyez	Examiner
Research Director (IPhT - Université Paris-Saclay)	

Résumé

Cette thèse présente mes contributions à l'expérience ATLAS, qui portent principalement sur deux sujets: une étude sur l'isolation des photons, et une recherche de nouvelles résonances dans le canal de désintégration en deux photons. Le travail décrit est basé sur les données de collisions proton-proton au LHC à une énergie au centre de masse de 13 TeV, enregistrées avec le détecteur ATLAS de 2015 à 2018, et correspondant à une luminosité intégrée de 139 fb^{-1}.

Les photons "directs", produits en collisions hadroniques, sont une sonde importante pour étudier la chromodynamique quantique, et aussi une signature puissante pour les recherches de nouvelle physique. Ces deux objectifs nécessitent d'une compréhension approfondie des processus de production de photons et des sources de bruit de fond. Les candidats photons sont reconstruits et identifiés à l'aide d'informations provenant des sous-détecteurs ATLAS, notamment le calorimètre électromagnétique. Les photons directs sont aussi identifiés à l'aide d'un critère d'isolation, basé sur un véto sur les dépôts d'énergie autour du candidat. Les efficacités d'isolation sont mesurées pour des photons d'énergies transverses (E_T^γ) comprises entre 20 GeV et 1.5 TeV, et pour différentes régions de pseudorapidité (η^γ) du détecteur. Le bruit de fond, composé principalement de mésons neutres dans les jets hadroniques, est estimé puis soustrait à l'aide de techniques basées sur les données.

Les performances d'isolation des photons sont fortement affectées par l'augmentation du bruit d'empilement provenant de collisions proton-proton supplémentaires, car des dépôts d'énergie issus de processus non corrélés avec la production du photon, peuvent être trouvés autour du photon candidat. Une étude prospective porte sur la façon d'améliorer l'isolation des photons en face d'une augmentation de l'empilement, basée sur des informations provenant des dépôts d'énergie dans les calorimètres. La mise en œuvre de cette étude permettra d'améliorer l'identification des photons directs dans les périodes futures de prise de données.

La deuxième partie de cette thèse présente une recherche de nouvelle physique avec deux photons dans l'état final. Des particules légères et faiblement interagissantes sont prédites dans une variété de modèles théoriques. Phénoménologiquement, de telles résonances suscitent un intérêt en tant que possibles médiateurs de la matière noire. Avant ces travaux, les expériences ATLAS et CMS avaient cherché de telles résonances dans le canal diphoton à des masses inférieures à la masse du boson de Higgs jusqu'à 65 GeV.

L'analyse décrite dans cette thèse utilise une nouvelle approche, basée sur la topologie particulière des événements avec des paires collimatées de photons. Cette approche permet d'atteindre des masses invariantes inférieures à 65 GeV, une région peu explorée dans les collisionneurs de hadrons en raison du défi expérimental que cela pose pour l'étude des photons de faible énergie. Les limites attendues fournissent la plus forte contrainte sur les couplages de particules de type axion aux gluons et aux photons pour des masses inférieures à 65 GeV. Mes contributions à cette analyse couvrent presque toutes les étapes de l'étude, allant de la génération et de la validation des échantillons de simulation de signal et de fond, la stratégie d'analyse, l'estimation de fond, l'évaluation des incertitudes systématiques et l'évaluation de la sensibilité attendue. J'ai également contribué à la validation de la prochaine génération du déclenchement en diphoton. Les futures périodes de prise de données bénéficieront de seuils en énergie plus bas, ce qui permettra d'étendre encore la recherche vers des plus faibles masses. Une étude prospective évalue l'amélioration attendue de la sensibilité de cette future stratégie.

Mots clés: LHC, ATLAS, photons, empilement, isolement, particules de type axion

Abstract

This thesis presents my contributions to the ATLAS experiment, consisting mainly on two topics: studies of photon isolation, and a search for new resonances in the diphoton decay channel. The work described is based on proton-proton collision data from the Large Hadron Collider at a center-of-mass energy of 13 TeV recorded with the ATLAS detector from 2015 to 2018, corresponding to an integrated luminosity of 139 fb^{-1}.

Prompt photons at hadron colliders are an important probe of the theory of Quantum Chromodynamics and a powerful signature in searches for new physics. Both objectives require a deep understanding of photon production processes and possible background sources. Photon candidates are reconstructed and identified using information from ATLAS calorimetric and tracking systems. For a precise prompt photon identification, the isolation of photons from nearby energy deposits is required. Measurements of photon isolation efficiency are presented for photons with transverse energies (E_T^γ) in the range from 20 GeV up to 1.5 TeV in different pseudorapidity (η^γ) regions. The methodology presented requires the subtraction of background events, primarily from neutral mesons in hadronic jets, estimated using data-driven techniques.

Photon isolation performances are heavily affected by the increase of pileup noise from additional proton-proton collisions, since energy from uncorrelated processes may be found around the photon candidate. A prospective study on how to improve photon isolation against increasing pileup using additional information from energy deposits in the calorimeters is presented. The implementation of this study will allow, in future data-taking periods, to improve prompt photon identification at various steps in photon reconstruction.

The second part of this thesis presents a search for new physics with two photons in the final state. Light and weakly interacting particles are predicted in a variety of theoretical models which would appear as resonances in the low mass part of the spectrum. Phenomenologically, such resonances gather interest as possible Dark Matter mediators. Previous to this work, the ATLAS and CMS experiments have searched for diphoton resonances at masses below the Higgs boson mass down to 65 GeV with no observed evidence for new physics.

The analysis described in this thesis uses a novel approach, exploiting the particular topology of events with collimated pairs of photons reconstructed in the detector. This topology allows to reach invariant masses below 65 GeV, a region poorly explored in hadron colliders due to the experimental challenge it poses for recording low energy photons. The expected limits provide the strongest bound on couplings of axion-like-particles to gluons and photons for masses below 65 GeV. My contributions to this analysis span over almost all the steps of the study, being responsible for the generation and validation of signal and background simulation samples, analysis strategy, background estimation, evaluation of systematic uncertainties and computation of the expected limit. I also contributed to the validation of the next generation of diphoton triggers. Future data-taking periods will benefit from novel energy thresholds allowing to further extend the search towards lower masses. A prospective study is presented with the expected improvement in analyses targetting low invariant diphoton masses.

Keywords: LHC, ATLAS, photons, pileup, isolation, axion-like particles

Publications and posters

- **Paper in preparation**: Search for axion-like particles in boosted diphoton events with the ATLAS detector at $\sqrt{s} = 13$ TeV (title subject to change).
- **ATLAS internal document**: Very-low-mass diphoton resonance search using 139 fb^{-1} of 13 TeV pp collision data with the ATLAS detector. ATL-COM-PHYS-2020-745
- **Talk**: Search for new phenomena in low-mass diphoton final states with proton–proton collisions collected at $\sqrt{s}$=13 TeV with the ATLAS detector. Journeés de Rencontre des Jeunes Cherheurs (2019)
- **ATLAS internal document**: Prospective studies on photon isolation performances using cluster-level information ATL-COM-PHYS-2020-822
- **ATLAS internal document**: Documentation on isolation performances in the ATLAS experiment. ATL-COM-PHYS-2019-432
- **Poster**: Photon Calorimetric isolation with the ATLAS detector: robustness against pileup, IV^{th} Asia-Europe-Pacific School of High-Energy Physics: CERN Summer School in Vietnam (2018)

Contents

Acronyms

ACB Asymmetric Crystal Ball.

ALP axion-like-particle.

ATLAS **A** **T**oroidal **L**HC **A**pparatu**S**.

BSM Beyond the Standard Model.

CMS **C**ompact **M**uon **S**olenoid.

DAQ Data Acquisition system.

DE dark energy.

DM dark matter.

DSCB Double-sided Crystal Ball.

EFT Effective Field Theory.

EM electromagnetic.

ggF gluon-fusion.

GPR Gaussian Process Regression.

HC Higgs Characterization framework.

HEP High Energy Physics.

HLT High-Level Trigger.

ID Inner Detector.

IP interaction point.

L1 Level 1 Trigger.

LAr Liquid-Argon.

LHC Large Hadron Collider.

LO leading order.

MPV Most probable value.

MS Muon Spectrometer.

NLO next-to-leading order.

NP New Physics.

NP nuissance parameters.

PDF parton distribution functions.

PDF Probability Density Function.

pNGB pseudo Nambu-Goldstone boson.

PoI parameters of interest.

QCD Quantum Chromodynamics.

QED Quantum Electrodynamics.

RoI Regions of Interest.

SM Standard Model.

Overview of the manuscript

This work described in this manuscript is divided into two main parts: studies on photon isolation and a search for resonances with masses below 65 GeV in the diphoton channel.

Chapter 1 provides the underlying theoretical framework needed in this dissertation. The chapter begins with an overview of the Standard Model of the fundamental interactions and fields, including a brief description of the spontaneous symmetry breaking mechanism which introduces the Higgs field to the field content of the theory. After presenting the current limitations of the Standard Model, emphasis is given on models predicting new fields that show up as resonances in the diphoton channel. This chapter also focuses on the phenomenology of photon production in hadron colliders with a discussion on how photon isolation measurements can be matched with theoretical predictions. The chapter finishes with a discussion on the simulation techniques used in particle physics, indispensable for the interpretation of experimental measurements.

Chapter 2 describes the experimental apparatus used to record the data used in this thesis, namely the Large Hadron Collider and the ATLAS experiment. This chapter also provides details on the trigger and data acquisition system.

Chapter 3 is devoted to the reconstruction of photons in the ATLAS experiment. An overview is given for all the steps in photon reconstruction: starting from the energy deposits in the calorimeters, passing through their association to photon candidates, and finally their identification. Special emphasis is given on the clustering algorithms of the calorimeter cells since they play a relevant role in the coming chapters.

Chapter 4 describes the studies performed on photon isolation. The chapter starts with a survey of the observables used in ATLAS to define whether a photon is isolated or not, and it is followed by an overview of the methodology used to extract the photon signal using data-driven techniques to estimate the background. Corrections applied to the simulation are also extracted from data and implemented. Both efficiencies and corrections are of importance in a wide variety of physics analyses with photons in the final state in the ATLAS collaboration. This chapter concludes with an alternative proposal on the implementation of the corrections to the simulation.

Chapter 5 presents a prospective study whose aim is the improvement of photon isolation performances using additional information from the energy deposits around photon candidates. This approach benefits from the longitudinal segmentation of the electromagnetic calorimeter and the timing of each energy deposit to discriminate clusters that arise from pileup interactions from those related to the photon candidate. This chapter describes the aforementioned observables and the methodology used to construct a likelihood weight indicating the probability for a cluster to come from a pileup interaction. The relevance of this study is directly related with the increase in instantaneous luminosity to come, aiming at rendering the photon isolation performances more pileup resilient.

Chapter 6 is devoted to the search of resonances with masses below 65 GeV in the diphoton channel. A novel analysis strategy is presented that allows to reach low diphoton masses, unexplored before with proton-proton collisions, based on the restriction of the phase space requiring collimated photons in the detector. This chapter describes in detail the generation of simulated samples, the composition of the background present in the data and the expected sensitivity on a resonance produced in gluon-fusion that decays to photon pairs. A comparison between past and future sensitivities is shown at the end to provide a complete pictures of the diphoton resonance searches performed and to come in the region below 100 GeV.

Finally, five appendices include information and results related to the presented studies. Appendices A and B provide the statistical framework relevant for the last chapter of the manuscript and the mathematical formulation for particular distributions widely used in the studies presented in Chapters 4, 5 and 6. Appendix C provides complementary information to Chapter 5. In Appendix D, crosstalk effects observed in the electromagnetic calorimeter are described. This appendix in particular presents an original strategy that allows for identifying features or issues in the reconstruction of electromagnetic objects using the tools presented in Chapter 5. The last Appendix E describes a prospective study on future resonance searches in the diphoton channel.

Chapter 1

Theoretical context

This chapter reviews a selected set of theoretical topics, relevant to understand the content of this thesis. First, a brief introduction on the Standard Model of fundamental interactions and fields is given, with a focus on the phenomenology of photon production in hadron colliders. It is followed by a discussion to motivate the physics analysis presented in this thesis: a search for an axion-like-particle decaying to two photons for masses between 10 and 65 GeV. This chapter finishes with a section describing the tools used to simulate events, necessary for data-theory comparisons.

1.1 The Standard Model

The Standard Model (SM) [1–4] is a relativistic quantum field theory that describes the fundamental interactions (aside from gravity) between the elementary fields that constitute matter and radiation. It was developed during the second half of the 20th century and it has successfully explained a very wide variety of phenomena; for essentially all physical observables accessible to current experiments, the predictions of the SM have been verified, sometimes with incredibly high accuracy.

1.1.1 Overview of the theory

The SM is based on the concept of local symmetries, referring to the set of transformations under which a physical system remains invariant. These transformations are denominated gauge transformations belonging to a symmetry group of a gauge theory, if the lagrangian $\mathcal{L}$ that describes the theory is left invariant under their action. Each of the fields present in a gauge field theory, like the SM, are classified according to the representations of the group to which they belong, and hence transform differently.

A quantum field theory is required to be invariant under both gauge transformations and under the action of Lorentz transformations. The former requirement implies the introduction of extra gauge spin-1 fields, besides the field content of the theory, which will ensure invariance under gauge transformations and mediate the interactions between the various particles. These fields are called gauge bosons.

The fundamental interactions embedded in the SM are the weak, the electromagnetic and the strong interactions. The SM is described by the gauge symmetry group $SU(3)_c \times SU(2)_L \times U(1)_Y$ whose components are briefly explained in the following. The field content of the SM contains 12 fermions, classified in quarks and leptons (and their antiparticle counterparts), gauge bosons which mediate their interactions and the Higgs field, the only

fundamental spin 0 boson in the model. Their quantum numbers and masses are shown in Figure 1.1.

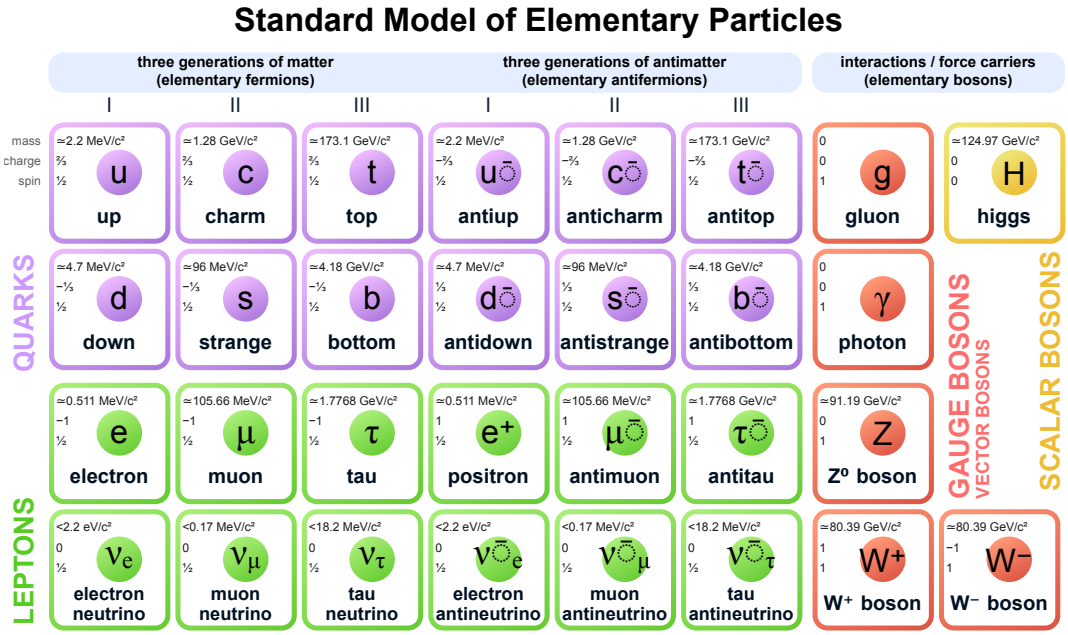


Figure 1.1: *Scheme of the field content of the Standard Model.*

The Glashow, Weinberg and Salam (GWS) electroweak theory (EW) [1–3], based on the symmetry group $SU(2)_L \times U(1)_Y$, unifies the electromagnetic and the weak interactions. It describes the interaction of fermions with the electroweak gauge bosons. These interactions are mediated through four vector bosons: 3 generators associated to the $SU(2)_L$ symmetry, W^a_μ; and one from $U(1)_Y$, B_μ. Under the action of $U(1)_Y$, fermion fields are affected by a phase shift, changing their weak hypercharge value Y. All irreducible representations of this symmetry group are one-dimensional and hence every field is a singlet under this symmetry group. Under the action of $SU(2)_L$, fermions are classified in two different categories: left-handed and right-handed. Left-handed fermions are those that transform as doublets under $SU(2)_L$ transformations, while right-handed fermions are scalars under $SU(2)_L$. This structure that discriminates between different chirality states, forbids fermions with masses.

The theory of Quantum Chromodynamics (QCD) [4] describes the strong interactions based on $SU(3)_c$, the colour symmetry group. The strong interaction is mediated by 8 massless vector bosons, $G^a_{\mu\nu}$, generators of the $SU(3)_c$ symmetry, called gluons. This symmetry allows to classify the fermions in two different categories: leptons and quarks. Leptons are scalars under the action of $SU(3)_c$ with no colour charge. Quarks transform as triplets under the action of $SU(3)_c$ and thus are sensitive to the strong interaction.

Numerical predictions of physical observables from these two theories is extremely difficult. These can be performed perturbatively in energy ranges for which the coupling constant α, parameter that measures the intensity of a given interaction, is small enough. However, restricting the phase space to where this condition is valid leads to mathematical divergences. Historically, these divergences are dealt with through a procedure called *renormalization*, in which the divergences are absorbed through a reparametrization of the coupling constants and mass normalizations of the field operators, in terms of an extra parameter called the renormalization scale μ_R. It was initially thought that the

renormalization scale had no physical meaning, and provided only the energy scale at which the renormalization procedure applied is valid.

This point of view has changed across history, leading to the introduction of Effective Field Theory (EFT) by Wilson in the 1970s [5]. Effective Field Theories are approximate theories used to describe physical phenomena below a certain energy scale. They are based on the concept of decoupling, meaning that the non understood higher-energy processes do not play a significant role when considering lower-energy scale interactions. This notion of EFT addresses the fact that the divergences appearing when computing physical observables arise from the inclusion of energy ranges that are beyond our current understanding of the theory. In this framework, the renormalization scale denotes the energy cut-off beyond which our understanding of the physical processes is limited.

Different parametrizations of the theory must lead to the same observable quantities. The renormalization group (RG) comprises the transformations of the parameters of the theory when considering different energy scales [6]. The equations describing the invariance of observable under changes of μ_R are the Renormalization Group Equations (RGE). For massless particles, the RGE of QCD leads to the evolution of the strong interaction coupling α_s with the energy scale Q^2, described by the following expression and its solution at leading-order (LO) in α_s:

$$\mu_R^2 \frac{\partial \alpha_s}{\partial \mu_R^2} = \beta(\alpha_s) = -(b_0 \alpha_s^2 + \mathcal{O}(\alpha_s^3)) \xrightarrow{\text{sol. at LO}} \alpha_s(Q^2) = \frac{\alpha_s(\mu_R^2)}{1 + b_0 \alpha_s(\mu_R^2) \log\left(\frac{Q^2}{\mu_R^2}\right)}, \quad (1.1)$$

where $b_0 = (33 - 2n_f)/12\pi$, with n_f the number of quark flavours. The non-Abelian nature of QCD leads to the minus sign of the beta function in Equation 1.1, and this feature has two important phenomenological consequences: asymptotic freedom and quark confinement.

When the transferred momentum, Q^2, in a given strong interaction increases, the coupling constant α_s decreases, as well as the intensity of the interaction. Therefore in the very high energy regime, perturbation theory methods can be used to compute observable quantities. This behaviour is called asymptotic freedom and it is currently accessible using state of the art accelerators.

When the energy transferred in the interaction becomes smaller, close to the fundamental QCD scale $\Lambda_{QCD} \sim 220$ MeV, the theory becomes strongly interacting, leading to quark confinement in hadrons. This is related to the fact that quarks and gluons are not observable states that can propagate over macroscopic distances. This behaviour is illustrated in Figure 1.2, where the coupling of the strong interaction α_s is shown as a function of the energy scale of the interaction Q.

1.1.2 Spontaneous Electroweak Symmetry Breaking

The SM is only gauge invariant for massless particles, a feature in contradiction with many obvious facts, like fermion masses (i.e the electron mass is measured to be non-zero since the XIX century), or the time scale of measured weak interactions, incompatible with massless W_μ^a mediators. These contradictions motivated the development of the Spontaneous Electroweak Symmetry Breaking (EWSB) theory, proposed by Brout and Englert [8], Higgs [9], and Guralnik, Hagen and Kibble (among others) [10] in 1964, providing a mechanism to solve this issue.

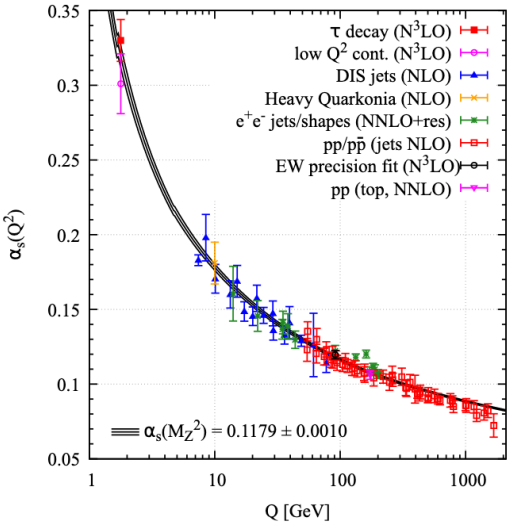


Figure 1.2: *Summary of measurements of α_s as a function of the energy scale Q [7].*

The original concept of spontaneous symmetry breaking is taken from condensed matter physics, in particular from Landau's theory of phase transitions applied to ferromagnetic materials [11]. A non-magnetized ferromagnetic material is invariant under any rotation since all its spins are randomly oriented and the total magnetic field is zero. If the temperature of the material decreases below a critical energy T_C, the spins line up in a particular direction, resulting in a non-zero magnetic field of the system. The rotational symmetry is then said to be spontaneously broken because the system is no longer invariant under arbitrary rotations, but only around the magnetic field axis.

A similar approach is used in particle physics to give masses to the different fermions and mediators. A complex scalar field doublet is introduced in the theory, different from other fields already present, since one neutral component of this field is required to have a non-zero vacuum expectation value (or vev) below a certain energy. Any action from the $SU(2)_L$ symmetry group on the doublet field changes the ground state; the symmetry $SU(2)_L \times U(1)_Y$ is therefore spontaneously broken. From the Goldstone theorem [12], for each spontaneously broken global symmetry, a new massless boson appears in the particle spectrum. However, if instead of a global symmetry a gauge symmetry is broken (as it is the case here) the extra degrees of freedom become longitudinal polarizations of the existing gauge bosons, and these gain an effective mass. The four mentioned generators of the electroweak symmetry then mix with each other, resulting in four different eigenstates: a massless vector boson, the photon; and 3 massive vector bosons: $W^{\pm}$ and Z^0. The massive states inherit from the broken $SU(2)_L$ symmetry generators, while the massless photon is related to an unbroken $U(1)_Q$ symmetry that leads to the theory of quantum electro-dynamics (QED).

The breaking of the electroweak symmetry predicts a new massive scalar boson that couples to particles with a strength proportionally to their mass, the Higgs boson. Moreover, by the inclusion of the Higgs potential it also provides an explanation for the fermion masses (Yukawa couplings).

This mechanism receives the name of the Brout-Englert-Higgs mechanism [8–10] and it has been confirmed with, out of a large list of discoveries and observations [13, 14], the discovery by the UA1 and UA2 collaborations of the massive $W^{\pm}$ [15] and Z bosons [16] and more recently, the Higgs boson by the ATLAS and CMS Collaborations [17, 18].

1.2 Photon physics at hadron colliders

While the SM provides a framework for precise calculations under certain conditions, physics at hadron colliders covers a wide energy range and phenomena, including regions where quantitative theoretical predictions are not easy to produce. This requires a detailed understanding of the dynamics between the proton consituents and their interaction with other composite objects by means of the strong interaction.

1.2.1 Parton distribution functions, hard processes and fragmentation functions.

The Parton Model was proposed by Bjorken and Feynman [19, 20] to describe deep inelastic experiments in 1960s. It is based on the concept of asymptotic freedom[1] in which a quark or a gluon ("parton") in the hadron is assumed to not interact with the other partons; the parton is therefore a "free" particle.

Under this assumption, the inner structure of the proton is described using form factors $f(x, Q^2)$ for each parton, called parton distribution functions (PDF). Here, x denotes the fraction of the proton momentum carried by a parton, and Q^2 denotes the momentum transfer of the interaction. The PDFs encapsulate the processes that cannot be computed perturbatively due to the character of the interactions that take place inside coloured-composite objects.

Each hadron is composed of valence quarks, which determine its charge and flavour. However, the sole presence of valence quarks is not sufficient to describe the inner structure of hadrons. Due to the QCD interactions occurring among them, a strong gluon field is present within the proton, leading to a "sea" of virtual quark-antiquarks pairs. As a result, all hadrons are made of quarks, anti-quarks and gluons and any of those partons can potentially interact with an incoming particle.

A cross-section computation of a hadron-hadron process can then be divided into two components: the hard process which can be computed on perturbative grounds in orders of the strong coupling α_s and the non-perturbative regime, which is described by the parton model and encapsulated in the PDFs. The separation between these two regimes is done through the introduction of an additional non-physical parameter called the factorization scale μ_F. By means of the QCD factorization theorem [22, 23], the cross-section of producing an arbitrary final state X from the interaction of hadrons A and B can be written as such:

$$\sigma_{AB\to X} = \int dx_a f_{a/A}(x_a, \mu_F^2) \int dx_b f_{b/B}(x_b, \mu_F^2) \times \left[\hat{\sigma}_0 + \alpha_s(\mu_R^2)\hat{\sigma}_1 + ...\right]_{ab\to X}, \tag{1.2}$$

where the first two integrals are integrated across all possible momentum fractions $x_a(x_b)$ of each of the partons $a(b)$ inside hadron $A(B)$, and the hard process is encoded in the expansion in terms of fixed order cross-section calculations up to $\mathcal{O}(\alpha_s^n)$. Since PDFs cannot be obtained perturbatively, they are fit from deep inelastic scatering measurements [24] and their evolution as a function of Q^2 is obtained from the DGLAP (Dokshitzer–Gribov–Lipatov–Altarelli–Parisi) equations [25–27]. The PDFs inside a proton obtained at next-to-next-to-leading order in α_s (NNLO) in the NNPDF analysis [28] at two different energy scales μ_R are shown in Figure 1.3.

[1] Though the parton model was developed before asymptotic freedom was discovered in 1973 by David Gross and Frank Wilczek [19] and separately by David Politzer [21].

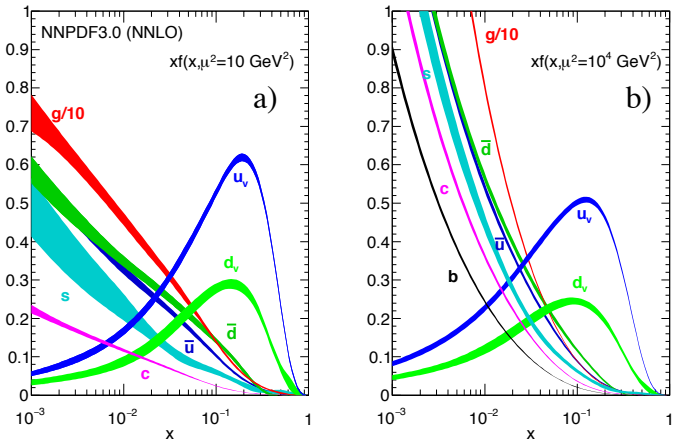


Figure 1.3: *Parton distribution functions for the proton obtained in the NNLO NNPDF3.0 global analysis at energy scales $\mu^2 = 10$ and 10^4 GeV^2 [28]. The gluon contribution clearly dominates in high energy collisions.*

In interactions in which a final-state parton is produced, a bunch of collimated partons and hadrons collinear to the initial parton is formed. This fragmentation and collinear showering arises from the iterative gluon radiation and quark-antiquark production since both are colourful particles and the strong coupling α_s increases while it "distances" from the proton local colour field. The collection of collimated final state hadrons is called a "jet".

While the production of the initial colourful parton can be described in terms of perturbative QCD processes, its hadronization crosses again to the non-perturbative regime. Following the same scheme as with previous transitions between perturbative and non-perturbative regimes, an extra parameter, again with non-physical meaning, called the fragmentation scale, μ_f, is included to distinguish these two regimes. The full cross-section accounting for a particular final state $C+X$[2] arising specifically from parton fragmentation can be written as follows:

$$\sigma_{AB\to C+X} = \int dz_C D_{C/k}(z_C, \mu_f^2)\sigma_{AB\to k+X}(\mu_R^2, \mu_F^2), \tag{1.3}$$

where z_C is the momentum fraction carried out by C, $D_{C/k}$ is the fragmentation function that defines the probability of a parton k to produce a certain final state C, and $\sigma_{AB\to k+X}(\mu_R^2, \mu_F^2)$ is the cross-section of a general hadron-hadron process, defined as in Equation 1.2, with a parton k in the final state. The evolution of the fragmentation functions with Q^2 is also defined by the DGLAP equations. Regarding the various spurious scales introduced along this section, each of them, μ_R, μ_F and μ_f, are usually set to the same value $\sqrt{Q^2}$, characteristic of the interaction under study. It is common practice in LHC measurements to set the value of the scales at the mass of the Z boson. Systematic uncertainties related to this choice are conventionally obtained through variations between twice and half this value.

[2]being C a specific stable particle.

1.2.2 Prompt photon production at the LHC

In the context of this thesis, only proton-proton (pp) interactions with prompt photons in the final state are of interest to us. Prompt photons are defined as those not produced in hadron decays. A theoretically motivated classification (at LO in α_s) divides final state photons into two categories: direct photons and fragmentation photons.

Direct photons are those produced directly in a hard process. At LO they are produced via two Born processes: annihilation, $q\overline{q} \to g\gamma$; and Compton, $qg \to q\gamma$. Fragmentation photons are those produced in the collinear fragmentation of a final state parton, essentially radiated from quarks. Diagrams for both direct and fragmentation photon production are shown in Figure 1.4.

However, at next-to-leading order (NLO) this classification is no longer valid since non-trivial divergences appear and only the combined contributions from both direct and fragmentation photons yield a physical observable [29].

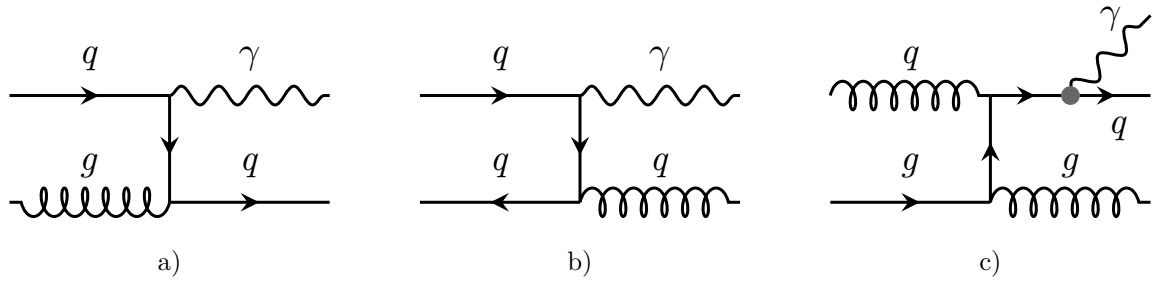


Figure 1.4: *Leading order diagrams for photon production at hadron colliders. (a) Quark-gluon Compton scattering (direct). (b) Quark-antiquark annihilation (direct). (c) Bremsstrahlung photon radiated by a final state quark (fragmentation).*

1.2.3 From theory to experiment: photon isolation

Measurements related to prompt photon production at hadron colliders are an important probe for QCD. They also provide a clean final state in many analyses, like Higgs boson mass or coupling measurements [30, 31] and many others. The overwhelming background from non-prompt photons arise from the decay of light mesons ($\pi^0, \eta, \omega ...$) to photons, produced in large quantities inside jets. Photons produced in such decays are usually very collimated and can be identified as a single prompt photon signal surrounded by other hadron energy deposits. In order to separate prompt from non-prompt photons, photon candidates are required to be isolated from hadronic energy flow.

A photon is considered isolated if it is surrounded by relatively low levels of hadronic energy (see Figure 1.5). The challenge is to provide a valid definition of isolation, for both experiment and theory. At the experimental level, the estimate of energy flow around a photon candidate is diluted by many factors like energy deposits arising from uncorrelated processes or the finite granularity of the detector. The most common technique to implement isolation at the detector level is the *cone criterion*: events are vetoed if the hadronic energy flow around a cone of fixed radius around the photon exceeds a certain threshold.

The cone criterion poses difficulties at theory level since this isolation definition imposes phase-space restrictions and hence limits the usage of photon fragmentation functions

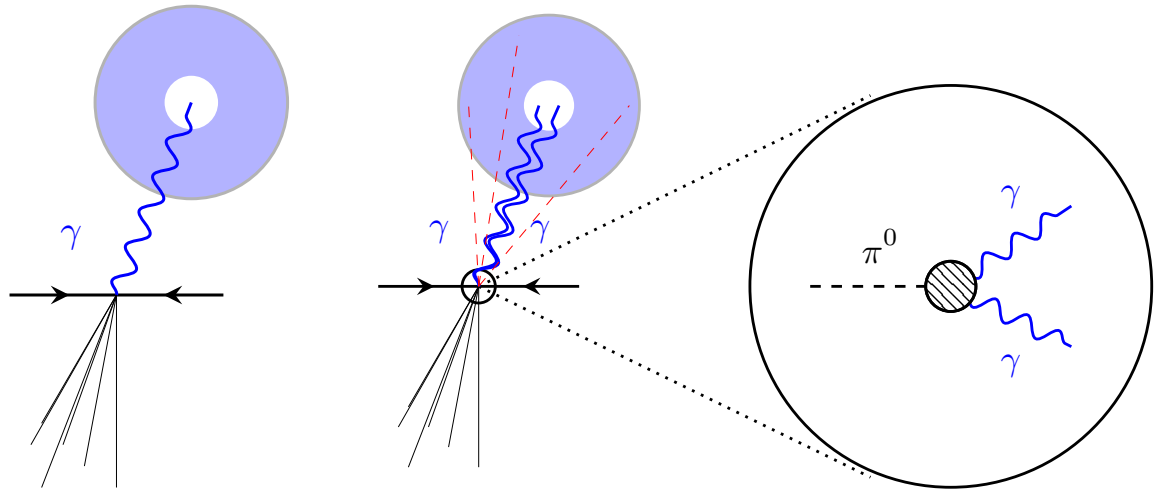


Figure 1.5: *Diagrams illustrating a comparison between prompt (left) and non-prompt (right) photons. On the left, a prompt photon recoils against a jet, illustrated as a collection of lines in the picture. On the right, a dijet event is shown in which within one of the jets contains a neutral pion, π^0, that decays to two collimated photons. The circular crown in light blue represents the isolation cone, in which the hadronic energy flow is estimated and used to discriminate prompt photons from non-prompt photons.*

$D_{\gamma/k}$. This obstacle is overcome computationally by applying, in fixed-order calculations, "smooth-cone" or "Frixione" isolation [32], in which the energy threshold is a function of the angular distance between the photon and the nearby parton. Unfortunately, due to the finite granularity of the detectors this condition is basically impossible to apply experimentally. However, other approaches, applicable to both theoretical and experimental levels, are under development, being the most popular ones "soft-drop isolation" [33] and "democratic isolation" [34].

Up to now, all experimental measurements with photons in the final state performed at the LHC use a fixed size cone while most QCD at NNLO [35] computations use smooth-cone isolation. A large fraction of the work dedicated in this thesis is devoted to this subject, with the studies shown in Chapters 4 and 5.

1.2.4 Non-resonant diphoton backgrounds at the LHC

Non-resonant QCD diphoton production contributes significantly to the expected background in diphoton resonance searches, like the one presented in this thesis in Chapter 6. Several processes arise from combinations of prompt and non-prompt photon production. Three purely prompt processes contribute to the non-resonant background diphoton production (see Figure 1.6):

- Born (tree-level) diphoton production through quark-antiquark annihilation, $q\bar{q} \to \gamma\gamma$.
- Box diphoton production from gluon fusion, through a loop of quarks, $gg \to \gamma\gamma$. Despite being of order $\mathcal{O}(\alpha^2\alpha_s^2)$, this process has a non-negligible cross-section at the LHC due to the large gg interaction cross-section at the LHC.

- Bremsstrahlung processes, in which the outgoing parton radiates a photon, $gq \to q\gamma\gamma$.

These three processes contribute to what will be called the "irreducible" background of the analysis, since they are indistinguishable in a per-event basis from a resonant diphoton event.

The background arising from non-prompt processes is called "reducible" background because any improvement in identification or isolation techniques at the experimental level would reduce its contribution. The main contribution comes from quark/gluon hadronization (jet) into neutral hadrons, which then decay into collimated photons (like π^0 decays). Dijet events are by far the dominant contribution since they have significantly larger cross-sections ($\sigma(pp \to jj) \sim 10^5$pb [36]) compared to diphoton production ($\sigma(pp \to \gamma\gamma) = 30$pb [30]), hence the importance of achieving excellent photon identification and isolation performances.

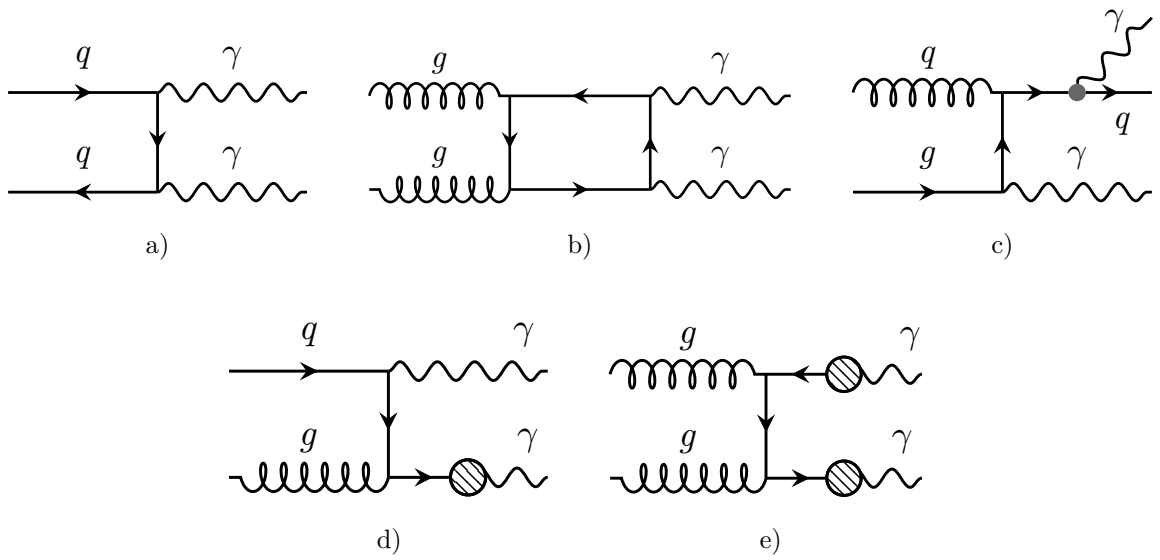


Figure 1.6: *Diagrams of dominant diphoton production processes at the LHC. The upper row (a,b and c) shows the prompt diphoton processes being from left to right: Born, box and bremsstrahlung processes. The bottom row (d and e) shows jet and dijet hadronization with photons in the final state. Bremsstrahlung from parton fragmentation is denoted with gray circles and hadronizations and hadron decays are denoted with dash filled circles.*

1.3 Beyond the Standard Model

The Standard Model has proven to be an incredibly successful theory, being able to describe a large variety of phenomena in a wide range of energies and processes. However, it is conventionally considered to be an incomplete theory. A number of outstanding experimental and theoretical results require the addition of new elements to the theory. A set of phenomena and theoretical results not fitting within the framework of the SM is listed below.

The first item in this reduced list is the naturalness problem, related to the consistency between the parameters found in a particular theory. According to this criterion, the value of all fundamental parameters should be of the order of unity, when expressed in the

appropriate units. An "unnatural" theory has parameters with large differences in value between them, like for example the difference between the strength of weak interactions and gravity. An example of a naturalness problem relevant in particle physics is the value for the bare Higgs mass [3].

Another structural problem of the SM is the strong CP problem. Compared to weak interactions, strong interactions are observed to preserve P and CP symmetries, while theoretically their violation is not forbidden. This behaviour is ruled by the parameter θ_{QCD} which if different from 0, would directly imply the violation of the aforementioned symmetries. However, this parameter is measured to be small ($\theta < 10^{-9}$) and the reason why it takes this precise value is understood as the strong CP problem. Peccei and Quinn [37] proposed a solution to the strong CP problem by introducing an additional global symmetry $U(1)_{PQ}$ to the theory which after breaking would make θ_{QCD} a dynamic variable. Despite being ruled out by experimental results [38, 39], the introduction of additional global symmetries has remained as a common mechanism to create light particles in many theories attempting to solve this problem [40, 41].

From the cosmological evolution of the universe and its present state, problematic experimental signatures have been observed regarding its composition. First, matter and anti-matter should have been produced in similar quantities during the Big Bang, despite the striking observed abundance of matter. The observed differences cannot be explained with purely SM processes [42] and additional asymmetry sources are being searched for.

The last conflict (in this non-exhaustive list) with the SM presented is the the strong observational evidence for the existence of Dark Matter (DM) and Dark Energy (DE). Ordinary matter (described by the SM baryonic fields) is measured to be around 5% of the observed universe [43]. The remaining contribution, which adds up to 95%, comes from additional matter and energy sources which cannot be explained by the SM.

The existence of DE provides a solution to the observation of the accelerated expansion of the universe, from supernovae measurements [44–46]. Matter alone (ordinary or dark) cannot explain this phenomenon since both contributions have positive pressures present in the stress-energy tensor of the Einstein Equations. An inclusion of an additional energy contribution with large negative pressure, sufficient to overcome the gravitational pull caused by its energy, predicts an accelerated universe expansion. No evidence of the interaction of DE with SM fields has been observed.

The DM hypothesis arose first as an explanation to several astrophysical and cosmological observations, such as galaxy rotation curves [47], gravitational lensing [48] and the measurements on cosmological scales of anisotropies in the Cosmic Microwave Background [49]. It constitutes, together with DE, an essential part of the Standard Model of Cosmology Λ_{CDM}. Besides the simple hypothesis of an additional type of matter that could explain the previously mentioned observations, a variety of explanations exists such as modified gravity [50], (which in fact excludes the existence of DM); SM extensions which include not yet found DM candidates [51, 52]; or hypotheses on the existence of large composite objects emitting little or no radiation, like primordial black holes [53]. Despite large efforts made by experiments covering a wide range in energies and with significant gains in sensitivity accross the years, DM remains poorly understood, being yet unclear if it is made or not by particles as such and, in the case it is, if it is connected with the SM fields. As of today, the quest for clarifying the particle nature of DM and its relation with the SM fields continues.

[3] After renormalization, the observed Higgs boson mass is parametrized in terms of the bare Higgs boson mass (see Section 1.1.1).

The analysis presented in this thesis is a search for resonances in the diphoton channel which probes the lowest diphoton invariant mass range accessible with proton-proton collisions. The following sections describe the framework used to interpret the results together with an overview of the experimental context.

1.4 Effective Field Theories and axion-like particles

Effective Field Theories, introduced in Section 1.1.1 in the context of renormalization, are a powerful tool often used to interpret results. Analogies to this approach are the description of macroscopic motion without dealing with quantum mechanics or the description of larger composite objects (atoms and molecules) without knowing the details of SM interactions. EFTs are model-independent to a large extent, despite the fact that their predictions are only valid within a given range of energies. They are usually expressed in a Lagrangian form, in which all possible operators fulfilling the symmetries of the theory are "written down":

$$\mathcal{L}_{EFT} = \sum_i \frac{c_i^D}{\Lambda_*^{D-4}} \mathcal{O}_i^D, \tag{1.4}$$

where $\mathcal{O}_i^D$ is an arbitrary operator i of dimension D, and c_i^D is the Wilson coefficient of such operator and specifies the intensity of such interaction. The relevance of a given operator is determined by the cut-off energy scale Λ_*^{D-4}. Operators beyond LO are usually not considered since they are suppressed by higher orders in Λ_*^{-1}.

Since no field in the SM can explain the observed conceptual and phenomenological difficulties described previously, developing higher order terms in EFT with only SM fields would lead to a similar result. Instead, new fields are included in the theory that could describe a variety of New Physics (NP) processes, including the DM hypothesis[4]. Axion-like particles gather special interest in this context as possible DM mediators and are described in the following.

The so called axion-like particles (ALP) appear in any theory with a spontaneously broken global symmetry. If such a symmetry $U(1)_{PQ}$ is only approximate, due to perturbative effects or explicit symmetry breaking mechanisms, the Nambu-Goldstone boson arising from the symmetry breaking acquires a mass, becoming a pseudo-Nambu-Goldstone boson (pNGB). Its mass is naturally light and the strength of its couplings with SM fields are small, since they scale with Λ_*^{-1}, being in this case Λ_* the energy scale of the spontaneous symmetry breaking. The energy scale can be described in terms of $\Lambda_* \sim g_* f_a$, where f_a is the decay constant controlling the coupling strength of the ALP with the SM, and g_* is the adimensional constant relating f_a with the symmetry breaking scale.

An ALP is a scalar and a singlet under SM gauge symmetries. The CP-odd choice is usually made since it predicts a strong suppression of event rates in direct DM detection experiments [54, 55]. The interactions between an ALP and SM fields are usually described

[4] Assuming two things: DM is a particle, and it couples with the SM fields.

in an EFT framework. The following dimension 5 operators[5] [38] represent the interactions up to LO in f_a^{-1}:

$$\mathcal{L}_{ALP} = \frac{1}{2}(\partial_\mu a)(\partial^\mu a) - c_{\tilde{B}} B_{\mu\nu}\tilde{B}^{\mu\nu}\frac{a}{f_a} - c_{\tilde{W}} W^A_{\mu\nu}\tilde{W}^{A\mu\nu}\frac{a}{f_a} - c_{\tilde{G}} G^A_{\mu\nu}\tilde{G}^{A\mu\nu}\frac{a}{f_a} + i(\phi^\dagger \overleftrightarrow{D}_\mu \phi)\frac{\partial^\mu a}{f_a}, \quad (1.5)$$

where a and ϕ are the ALP and Higgs boson fields respectively, $B_{\mu\nu}$, $W^A_{\mu\nu}$ and $G^A_{\mu\nu}$ are the gauge field strength tensors corresponding to $U(1)_Y$, $SU(2)_L$ and $SU(3)_c$ respectively. Analogously to equation 1.4, the constants $c_{\tilde{i}}$ encode the coupling strength of each of the operators.

Depending on the region under study in the ALP parameter space, defined by its mass m_a and the scale f_a, ALPs can be non-thermal DM candidates [56] or if they decay, mediators connecting the SM to a dark sector, which englobes a set of fields not included in the SM which could provide a particle DM candidate. In the context of this thesis, emphasis is given on the latter.

1.4.1 Resonant axion-like particle phenomenology at hadron colliders

Phenomenologically, ALPs gather interest at particle colliders not only as DM mediators but also for particular values of the decay constant f_a. Decay constants in the range between $f_a \sim 0.1 - 10$ TeV are ubiquitous in theoretical models in which additional pNGBs are naturally present in their spectrum and accessible at LHC energies. A small review is given in the following on the relevant physics cases in particle colliders for resonances in the diphoton channel:

- Heavy axion hypothesis: this point in the theory space assesses the strong CP problem with a solution inspired in the original QCD axion from Peccei and Quinn, with additional mechanisms to make the ALP heavier [57–59]. This ALP couples through loops to heavy fermions, charged under the $U(1)_{PQ}$ symmetry, with masses beyond the LHC reach. Its coupling to gluons introduces an extra parameter in the QCD lagrangian, leading to a dynamical θ parameter and thus not technically fine-tuned.

- R-axion and composite Higgs models: the R-axion is a pNGB arising from the symmetry breaking of the R-parity in Supersymmetric models [60]. Similarly, composite Higgs models also include naturally additional pNGBs in their spectrum [61]. Both frameworks provide the necessary elements to potentially solve the naturalness problem of the scale of EW interactions.

- Dark Matter freeze-out: the existence of an ALP for the mentioned decay constant range gains relevance by requiring it to be a mediator to fermionic DM [62, 63]. In particular, an ALP of these characteristics could be the mediator of the DM annihilation into SM particles, leading to the freeze-out of DM abundance.

The non-zero coupling to gluons is present in all the previously mentioned models, favouring a large rate of ALPs in proton-proton collisions at the LHC. At high-energy

[5] The explicit mass term is assumed to be small and it is not considered in the following.

hadron colliders, ALPs can be produced with no additional assumptions via two mechanisms: gluon-fusion $gg \to a$ and photon-fusion $\gamma\gamma \to a$. Other vector boson fusion ($W^{\pm}$ or Z) production mechanisms are not considered here, since they would only be dominant if the coupling of the ALP to gluons is set to zero [64]. Both gluon-fusion and photon-fusion mechanisms scale with the mass of the ALP, m_a, over the decay constant f_a. In proton-proton collisions, the contribution from photon-fusion is negligible, being the ALP essentially produced in gluon-fusion.

For m_a below the mass of the Higgs boson, the ALP dominant two-body decays are:

$$\Gamma_{gg} = K_g \frac{(\alpha_s c_{\tilde{G},eff})^2}{8\pi^3} \frac{m_a^3}{f_a^2}, \quad \Gamma_{\gamma\gamma} = \frac{(\alpha_{em} c_{\gamma,eff})^2}{64\pi^3} \frac{m_a^3}{f_a^2}; \tag{1.6}$$

where $c_\gamma = c_{\tilde{W}} + c_{\tilde{B}}$ and the labels $c_{i,eff}$ denote that these parameters encode the anomalies [65, 66] of the global $U(1)_{PQ}$ symmetry with the SM symmetries, which are not included in the analogous parameters in Equation 1.5, and $K_g = 2.1$ encodes higher-order QCD corrections [67]. The dominant contribution to the total width is the decay into gluons, unless $c_{\tilde{W},\tilde{B}} \leq 10^2 c_{\tilde{G}}$. In the mass range under consideration, the total width of the resonance is typically narrow ($\Gamma_{tot}/m_a = 10^{-3}$) leading to prompt ALP decays within the acceptance of the detector.

1.4.2 Existing constraints on axion-like particles

Search strategies vary significantly depending on the region under study in the ALP parameter space. Light ALPs with masses below twice the electron mass are only allowed to decay to photons, and are usually assumed to be long-lived and escaping the acceptance of common collider detectors. These are constrained from helioscope [39] and beam-dump experiments [68]. Helioscopes aim to detect the axions produced in the solar core by enhancing its conversion into a photon through the Primakoff effect[69]. Beam-dump experiments target long-lived particles which would traverse the absorber and decay in a detector placed beneath it. For intermediate masses up to the GeV scale, constraints were obtained in lower energy colliders such as LEP, by the L3 and OPAL experiments [70–72], which searched for enhancements in $Z \to 3\gamma$ suppressed decays or associated production with an additional photon in the events.

The strongest constraints up to date from GeV levels up to 50 GeV have been recently set at the LHC by the ATLAS and CMS experiments in light-by-light scattering in ultra-peripheral Heavy-Ion collisions [73, 74]. The Pb ions interact electromagnetically, producing two photons in the final state through the Dellbrück process $\gamma\gamma \to \gamma\gamma$ [75]. This process is exclusively mediated by QED and by measuring its cross section, limits can be set on additional BSM contributions. However, these limits are only sensitive to ALP couplings to photons only, and this restriction reduces significantly its constraining power by not considering gluon-fusion. Moreover, Heavy Ion analyses suffer from very limited statistics compared to those using pp collisions.

At higher masses up to the Higgs boson mass, the strongest constraints on the ALP parameter space are derived from dijet and diphoton resonance searches. Dijet resonances have been performed by CMS, reaching invariant masses down to 50 GeV. Beyond 65 GeV, diphoton searches set the strongest constraints since diphoton final states benefit from a cleaner signature in the detector compared to dijet searches, despite its smaller branching ratio[6].

[6] $\mathcal{B}(a \to gg) > \mathcal{B}(a \to \gamma\gamma)$

The lowest masses that can be reached in these searches with pp collisions are limited by the energy required to trigger the various objects (either jets or photons) plus various other experimental requirements. For diphotons, an additional limitation arises from isolation requirements. In the searches shown in Figure 1.7, the diphoton energy thresholds required by ATLAS are looser compared to those applied in CMS (20 GeV for both photons compared to 30 and 18 GeV for each photon respectively), leading to a slightly lower reach in the invariant mass (65 GeV compared to 70 GeV). Any decrease in the energy thresholds or isolation requirements would increase the sensitivities at lower invariant masses.

The current results in this mass range have been obtained with data recorded at $\sqrt{s} = 8$ TeV for ATLAS and CMS, representing an integrated luminosity of $\sim 20\text{fb}^{-1}$ and $\sqrt{s} = 13$ TeV for CMS with an integrated luminosity of $\sim 36\text{fb}^{-1}$. The work presented in this thesis is performed with data recorded by the ATLAS experiment at $\sqrt{s} = 13$ TeV with 139fb^{-1}, almost a factor 4 increase in the available statistics.

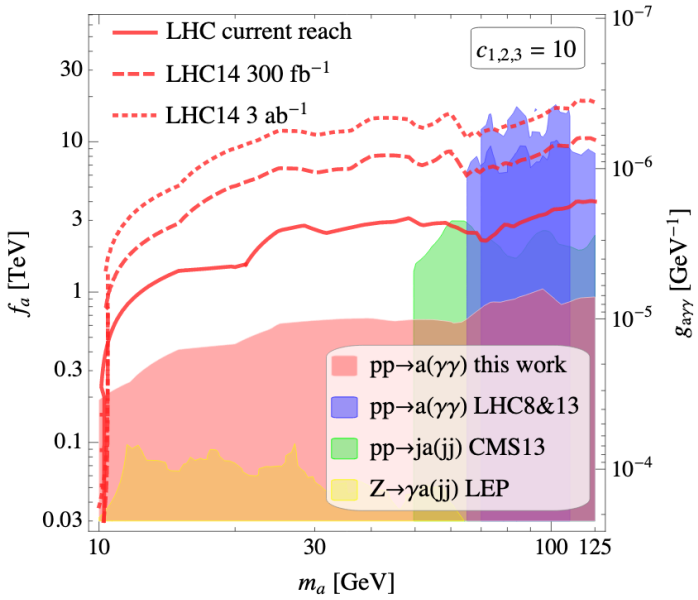


Figure 1.7: *Constraints on the ALP parameter space from collider searches. The constraints extracted from LEP [70, 71] are shown in yellow, while the constraints obtained at the LHC from dijet and diphoton searches are shown in green and blue respectively. The red shaded region denotes the expected theoretical bound extrapolated from ATLAS and CMS measurements in [76]. The lines denote the expected LHC sensitivities derived in [76].*

1.5 Monte Carlo Event Generators

The main aspects of high energy collisions event generation are discussed in this section. Monte Carlo (MC) methods and phenomenological models are used to generate events. The processes involved between the incoming protons and the final stable particles in the event cover a wide range of energies and interactions. For this reason, the event generation

is usually performed in various separate steps which are described in the following and illustrated in Figure 1.8.

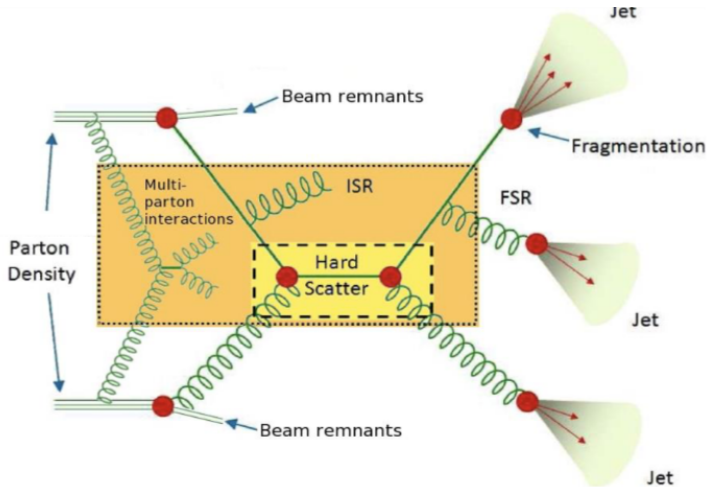


Figure 1.8: *Schematic picture of a pp collision.*

1.5.1 Hard process

Hard processes are generated perturbatively. The event generator computes the cross-section (matrix element), at fixed order in perturbation theory, of a given process provided a set of PDFs, encoding the inner structure of the proton and the initial state partons, and the available phase space. A practical limit is imposed on the maximum order in perturbation theory that can be achieved, since the number of diagrams contributing to the process grows exponentially with increasing powers in the coupling constant, and so does the computation time. State-of-the-art generators provide matrix element computations up to NLO.

1.5.2 Parton showering

Parton showering provides a connection between the hard process and the final state particles. Each charged (either electrically or colour charged) particle involved in the hard process can radiate while they are still in the perturbative regime. Though these radiations are indeed higher order contributions to LO processes, computing their matrix elements requires significant computational resources. The connection between the hard process and the parton showering needs to be carefully treated, since diagrams appearing in the original hard process can re-appear after parton showering. A matching between both steps has to be performed in order to avoid double-counting parton emissions in the same phase-space.

In some generators it is convenient to identify two contributions for computational purposes: initial-state radiation (ISR) for radiation emitted by the incoming partons involved in the hard process, and final-state radiation (FSR) for radiation emitted by the outgoing partons involved in the hard process.

1.5.3 Hadronization

The strong interaction at long distances, beyond the perturbative regime, is described with phenomenological models of hadronization. The most used event generators include the Lund string fragmentation model and the cluster fragmentation model. These will not be developed here, but details for both models can be found in [77] and [78].

1.5.4 Underlying event

The underlying event contains all the particles in the event that are not coming from the main energy transfer, the hard process. It is composed of interactions from the beam remnants, additional parton-parton interactions, and initial and final state radiations. It is dominated by soft particles. The description of this part of the event is difficult, since the topology of the beam remnants is highly dependent on the energy transferred in the hard process. All these events are tuned in the simulation with "minimum bias" events recorded from data using very loose trigger requirements.

1.5.5 Decays of unstable particles

Upstream the hadronization step, the event generator ensures the decay of any unstable particle produced. This is performed for a large number of hadron decays through decay amplitudes provided to the generator by external specialized packages, like EvtGen [79].

1.5.6 Monte Carlo generators used in this analysis

Various MC generators are available to reproduce pp collisions at the LHC. Since the event generation is divided in several steps, some event generators are specialized in particular parts of the process, like for example, providing precise NLO computations of the hard process. These generators are then be interfaced with another multipurpose generator in order to include the rest of the elements, like the parton showering. Though convenient for a faster processing, matching needs to be performed afterwards to avoid double-counting. Hypothetical resonances and Standard Model processes are generated in this analysis with the following generators:

1.5.6.1 PYTHIA

PYTHIA [80] is a multipurpose LO event generator that provides tools for a full event generation: hard process, underlying event and parton showering. Matrix elements are computed for $2 \to 2$ processes as double differential cross-sections in transverse energy E_T and rapidity y.The fragmentation component of photons is modelled as FSR, with no knowledge of the fragmentation functions. No isolation criterion is applied for radiated photons. It is usually interfaced with other generators with more precise matrix element calculations. All the PYTHIA samples used in this thesis are generated with the A14 parameter tune [81] along with the NNPDF23LO PDF set [28].

1.5.6.2 SHERPA

SHERPA [82] is another multipurpose event generator that can only be used standalone, without options for interfacing it with other generators. The modelling of fragmentation

photons differs from the treatment given by PYTHIA, since it models the fragmentation functions as part of the parton showering describing it through $2 \to n$ matrix elements. To avoid collinear singularities in this scheme, smooth-cone isolation is implemented in opposition to PYTHIA.

1.5.6.3 MadGraph

MadGraph [83] provides matrix element calculations of the hard processes at LO and NLO accuracy. In the samples used in this analysis, it is interfaced in a later step with Pythia for the parton showering, and the matching is performed under the CKKW-L merging scheme [84].

1.5.7 Detector simulation

The final state particles produced by the event generators are passed through a detailed simulation of the ATLAS detector. The simulation of the ATLAS detector response is performed using the GEANT4 package [85]. The GEANT4 package is a toolkit used to simulate the passage of particles and radiation through matter and it is widely used in high energy physics.

The samples used in this thesis are simulated with two different approaches regarding the simulation of particle interactions in the calorimeters. They can be classified as fully simulated samples and fast-simulated samples.

The fully simulated samples use a very detailed description of the ATLAS detector built with the GEANT4 package, in which the electromagnetic and hadronic interactions in the calorimeters are simulated using MC techniques. Fully simulated samples are used in the studies which require a high quality calorimeter response like the isolation studies shown in Sections 4 and 5, and the signal samples characterization in Section 6.2.2

The fast simulation samples use a parametrized version of the calorimeter response, which reduces the simulation time by large factors, depending on the type of event [86]. The tracking detectors and muon chambers remain fully simulated. In the document presented in this thesis, this simulation setup is only used for the large-statistics samples describing the non-resonant QCD diphoton background in Section 6.2.3.

Chapter 2

The Large Hadron Collider and the ATLAS experiment

The following chapter describes the experimental setup used to produce the data used in this thesis. A description of the Large Hadron Collider and the ATLAS experiment is given, with special emphasis placed on the most relevant subsystems used for photon detection.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a circular proton-proton (pp) and heavy ion collider at the European Laboratory for Particle Physics (CERN). It is placed in a tunnel previously used by the Large Electron-Positron (LEP) collider, built between 1984 and 1989, with approximately 27 kilometers of circumference and buried ~ 100 meters underground, underneath the French-Swiss border. The construction of the LHC began in 1998 and it was designed to produce a wide variety of physical processes at very high center-of-mass energies, to push further the existing energy frontier of previous colliders.

2.1.1 The LHC acceleration chain

Before injecting the proton beams in the LHC, they are accelerated by means of a series of lower-energy particle accelerators. The starting point is a single bottle of hydrogen. After stripping the hydrogen molecules of their electrons with an intense electric field, protons are then injected in the LINAC2, a LINear ACcelerator. The LINAC2 increases the energy of the proton beam up to 50 MeV using oscillating electric fields and collimated with magnetic quadrupoles. In the following stages, the beam is further accelerated by three synchrotrons: the Proton Synchrotron Booster (PSB) , the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS) reaching 1.4 GeV, 26 GeV and 450 GeV respectively. Each of them has a larger radius than the previous one, as it is depicted in Figure 2.1 together with the full injection chain.

The proton beams are finally injected in the LHC pipes where they are accelerated up to TeV-energies. Unlike particle-antiparticle colliders, two pipes are necessary as opposite rotating beams with the same charge require opposite magnetic fields. In this last stage

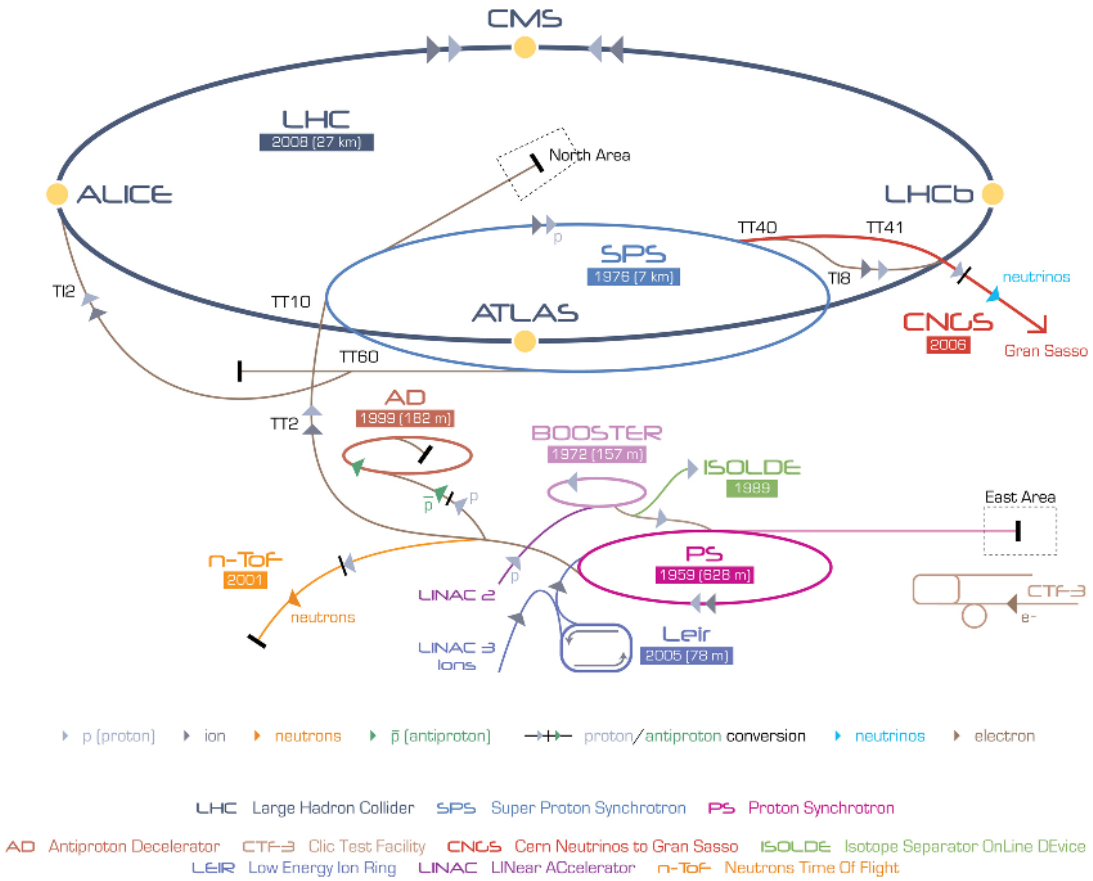


Figure 2.1: *Sketch of the CERN accelerators complex. [87]*

protons are accelerated up to 6.5[1] TeV using radiofrequency cavities that provide an oscillating electric field of $5MV/m$ which oscillates with a frequency of 400 MHz.Each beam reaches an energy of 6.5 TeV after 20 min (4 trillion revolutions) of acceleration, providing a centre-of-mass energy of 13 TeV. These energies require very strong magnetic fields in order to bend the beams to match the curvature of the LHC and maintain them focused inside the pipe. This is achieved with superconducting magnets operating at 1.4 K, cooled with liquid Helium. The LHC magnetic system consists of 1232 dipoles that bend the beam and 392 quadrupoles that focus the beam [88].

After the acceleration chain, the two proton beams collide in 4 collision points, where detectors with different purposes are located. In each of the collision points, different crossing angles are used depending on the objectives of each experiment. ATLAS and CMS are multipurpose detectors designed to be sensitive to a wide spectrum of possible physics signatures. The ALICE experiment is a heavy-ion detector designed to study states of matter in extreme environments, such as the quark-gluon plasma. Finally and unlike the previous detectors, LHCb is designed to study, in forward collisions, flavour physics and search for NP through precision measurements of CP violation or rare B decays.

[1]The design center of mass energies was 7 TeV, but technical problems with the magnetic systems allowed only 6.5TeV.

2.1.2 LHC beam characterization

In this subsection, beam-related parameters and quantities are described: necessary concepts to understand how data are collected at the LHC.

2.1.2.1 Beam structure: bunches

The LHC beam is not a continuous flow of particles. Instead, it is divided in "bunches" of approximately 10^{11} protons spaced by 25 ns. A bunch in this context is a collection of protons grouped together using radio frequency cavities during the LINAC2 acceleration. Assuming the spacing between bunches has negligible length and given its circumference, the LHC tunnel could potentially store ∼3600 bunches. However, some constraints reduce this number down to 2808 in the nominal configuration, such as the necessary time to inject and dump the beams across the different acceleration stages. The bunch structure in the nominal configuration is shown in Figure 2.2

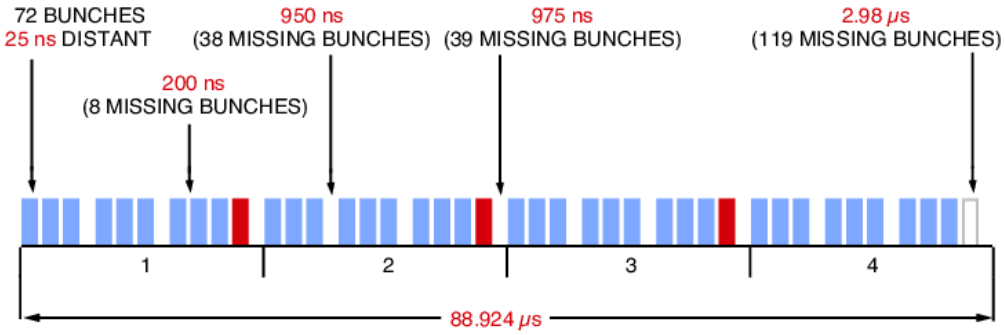


Figure 2.2: *The LHC nominal bunch structure. [89]*

2.1.2.2 Luminosity

Besides the energy of the colliding beams, the rate of useful interactions produced is an important parameter. The *instantaneous luminosity* $\mathcal{L}_i$ is defined as the proportionality factor between the event rate $\dot{N}$ and the production cross-section σ_p of a certain process,

$$\frac{dN}{dt} = \mathcal{L}_i \sigma_p. \tag{2.1}$$

Instantaneous luminosity is therefore measured in $cm^{-2}s^{-1}$ and it only depends on the beam parameters and not on the physical process under study, being the information of the latter contained in σ_p.

Intuitively, the instantaneous luminosity can be understood as the rate of overlap of two particle bunches in the same region of space. Mathematically, in the case of two colliding bunched beams, the instantaneous luminosity can be expressed as the overlap integral over space and time of the density functions of two proton beams:

$$\mathcal{L}_i = N_1 N_2 f_{rev} n_b M_{KLF} \int \rho_1(x,y,z,-z_0)\rho_2(x,y,z,z_0)dxdydzdz_0, \tag{2.2}$$

where N_i is the number of particles per beam bunch, ρ_i is the time-dependent distribution function of each of the beams, f_{rev} is the revolution frequency of a bunch in the

LHC ring, z_0 is the distance of each beam to the interaction point, n_b is the number of bunches stored in the ring and M_{KLF} is the Møller Kinematic Luminosity Factor [90], which accounts for the relative velocity of both beams.

A practical result is obtained assuming two Gaussian profiled beams with the same number of particles per bunch:

$$\mathcal{L}_i = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi\sigma_x\sigma_y} F \quad \text{with} \quad F = \left(1 + \left(\frac{\theta_c \sigma_s}{2\sigma^*}\right)^2\right)^{-1/2}, \tag{2.3}$$

where γ_r is the Lorentz factor of the proton beam, σ_i are the horizontal and vertical beam sizes and F is the geometric luminosity reduction factor due to the crossing angle at the interaction point (IP). The geometric reduction factor F is a function of parameters and θ_c being the crossing angle at the IP, σ_s is the RMS bunch length and σ_* is the transverse RMS beam size at the IP. The largest delivered luminosity is achieved with head-on colliding bunches but this is often not optimal when the distance between bunches is small to avoid extra unwanted interactions. LHC luminosity parameters are shown in Table 2.1 and a sketch of two colliding beams is shown in Figure 2.3a.

While the instantaneous luminosity describes how much data is being provided per second, in practice one is often interested in the *integrated luminosity*, which reprent the total collected data over a period of time:

$$L = \int \mathcal{L}_i dt, \tag{2.4}$$

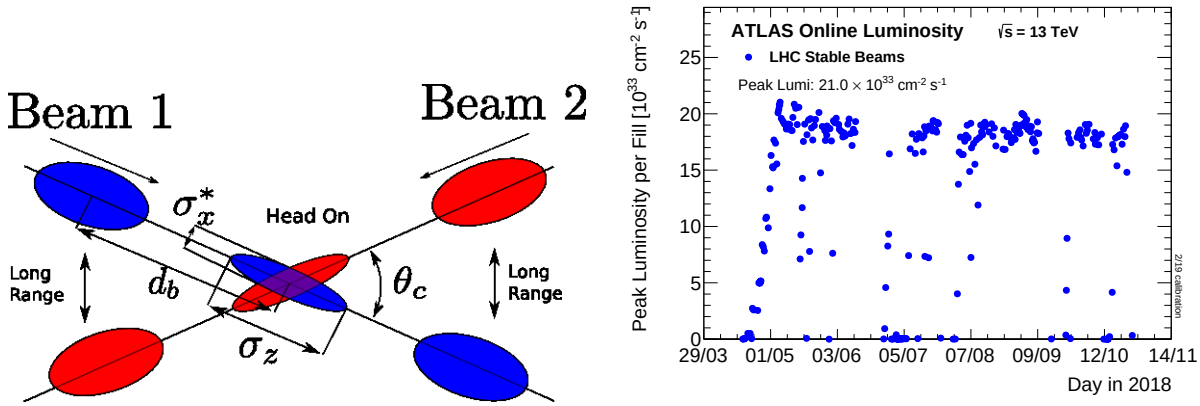
where L is measured in units of m^{-2} and expressed typically in b^{-2}, being $1b = 10^{-28}m^2$. The design instantaneous luminosity of the LHC is $\sim 10^{34} cm^{-2}s^{-1}$. After successful years of operation, this value has been overcome delivering up to 2 times more instantaneous luminosity. Figure 2.3b shows how the instantaneous luminosity peak per day evolved during the 2018 data-taking period. After several pp collisions, the initial instantaneous luminosity decreases due to the degradation of both intensities and emittances of the beams after collisions in each of the IPs. Together with other effects (intra-beam scattering, residual gas in the pipe,...) that degrade the beam, an average estimate of the luminosity lifetime is $\sim$ 15h in the LHC.

2.1.2.3 Pileup

As a consequence of the continuous increase of the instantaneous luminosity to gather more data, more than one inelastic pp collision takes place when two bunches collide. Multiple simultaneous interactions pose a challenge for the detectors that must resolve particles from different pp collisions. The additional inelastic collisions besides the main hard-scatter pp collision are referred to as *pileup*. Two types of pileup can be identified:

Table 2.1: *Summary of nominal LHC parameters.*

Parameter	Description	Nominal LHC Value
N_b	Number of particles per bunch	$10^{10} - 10^{11}$
n_b	Number of bunches per beam	2808
f_{rev}	Revolution frequency	11.2 kHz
$\sigma_{x,y}$	Horizontal and vertical beam sizes	$2.5\mu m$
F	Geometric reduction factor	150-250 μrad



a) b)

Figure 2.3: *(a) Sketch of a collision of two pp beams. Varying the parameters depicted in the figure and the number of protons per bunch, luminosity and machine specifications can be tuned to suitable values for operation. [91] (b) The peak instantaneous luminosity as a function of time, delivered to ATLAS for pp collisions at 13 TeV center-of-mass energy in 2018. [92]*

- In-time pileup refers to the additional *pp* collisions that take place within the same bunch crossing.
- Out-of-time pileup refers to additional *pp* collisions occurring before or after the bunch-crossing under study. This type of pileup is sensitive to the bunch time separation of 25 ns, since the recovery time for many subdetectors is larger than this quantity.

Pileup can be characterized by means of the average number of interactions per bunch crossing $\langle\mu\rangle$, which is directly related to the instantaneous luminosity $\mathcal{L}$:

$$\langle\mu\rangle = \frac{\sigma_{\text{inelastic}}\,\mathcal{L}_{\text{i,bunch}}}{f_r}, \tag{2.5}$$

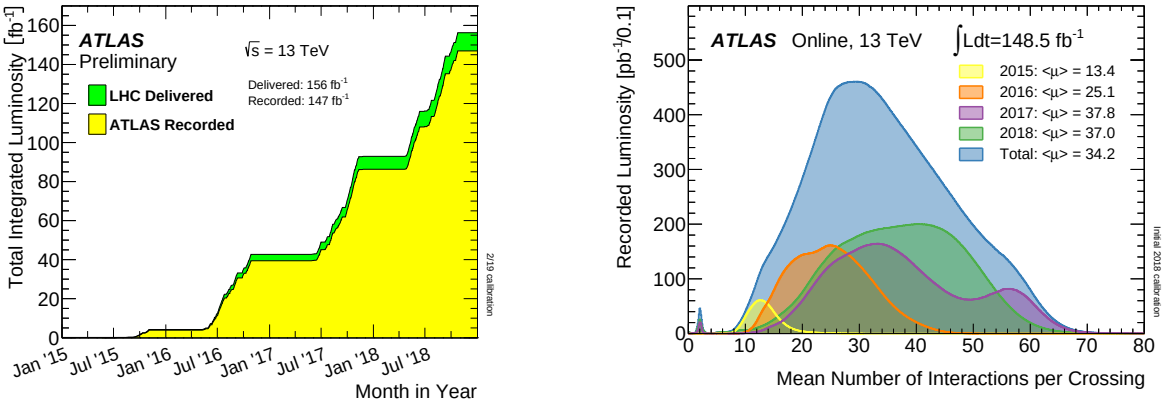
where $\mathcal{L}_{\text{i,bunch}}$ is the per-bunch instantaneous luminosity, f_r is the LHC revolution frequency and $\sigma_{\text{inelastic}}$ is the inelastic *pp* cross-section (78.1 mb at $\sqrt{s} = 13$ TeV [93]).

2.1.3 LHC performance during Run 2

The LHC has proven to be a very successful machine delivering stable *pp* beams to various experiments, exceeding design instantaneous luminosity but slightly below its design energy of $\sqrt{s} = 14$ TeV [94, 95]. During Run 2, defined as the data-taking period between 2015-2018, the LHC has delivered 156 fb^{-1} from which the ATLAS experiment has recorded 139 fb^{-1} at $\sqrt{s} = 13$ TeV (see Figure 2.4a). These numbers can be compared to the ones obtained during Run 1 (from 2010-2012), during which 25 fb^{-1} of data at 7 and 8 TeV where recorded.

The instantaneous luminosity has changed along the years, increasing with it the average number of interactions per bunch crossing. Run 2 can be characterized by approximately 35 additional *pp* simultaneous interactions on average. Mitigation of the effects caused by additional interactions in particle identification is a crucial part in performance programs

and prepares the experiments for the next data-taking period, where $\langle\mu\rangle$ is expected to be of the order of ~ 60 [96].



a) b)

Figure 2.4: *(a) Cumulative luminosity as a function of time delivered to ATLAS by the LHC (green) and recorded by ATLAS (yellow) at 13 TeV center-of-mass energy in Run 2. (b) Recorded luminosity by ATLAS during Run 2 as a function of the average number of interactions per bunch crossing $\langle\mu\rangle$. Different years are shaded with different colors, showing an average $\langle\mu\rangle$ of 34.2 [92]*

.

2.2 The ATLAS detector

ATLAS is a multipurpose detector designed to cover a wide spectrum of different physics signatures from pp collisions at the LHC. It is the largest detector at the LHC, with a cylindrical shape of 46 meters long, 25 m of diameter and weighing 7000 tonnes. The cylindrical shape of the detector is divided into three parts: the barrel, which comprises the center part of the detector; and the two endcaps, referring to the forward parts of the cylinder and closer to the beam. Following a common strategy in particle detectors, it is composed by various sub-detectors, placed coaxially to the proton beam, that measure energy and momentum of charged particles, electromagnetically interacting particles, hadronic jets and muons. Three main ingredients provide discrimination between the most common physics objects: tracking detectors, calorimeters, and muon spectrometers. Closer to the beam pipe, tracking detectors are immersed in a 2 T magnetic field allowing the reconstruction of trajectories from charged particles (tracks). This first step allows for charged and neutral particle discrimination. A calorimeter system is placed downstream the tracking systems to measure the energy deposited by incident particles while they traverse material. An electromagnetic sampling calorimeter is placed to absorb almost all the energy from electrons and photons. Hadronic jets are fully contained in the hadronic calorimeter. Muons being capable of escaping the calorimeter layers, the outermost layer is a muon spectrometer composed by tracking systems and toroidal magnets that allow their momentum reconstruction. Surrounding to the tracking system, a toroidal magnet delivering a magnetic field of 0.5 and 1 T in the barrel and endcaps respectively, is used to bend muon trajectories and provide an independent measurement of muon momenta.

Together, all detector sub-systems provide a close to 4π geometrical acceptance. A general view of the ATLAS detector is shown in Figure 2.5.

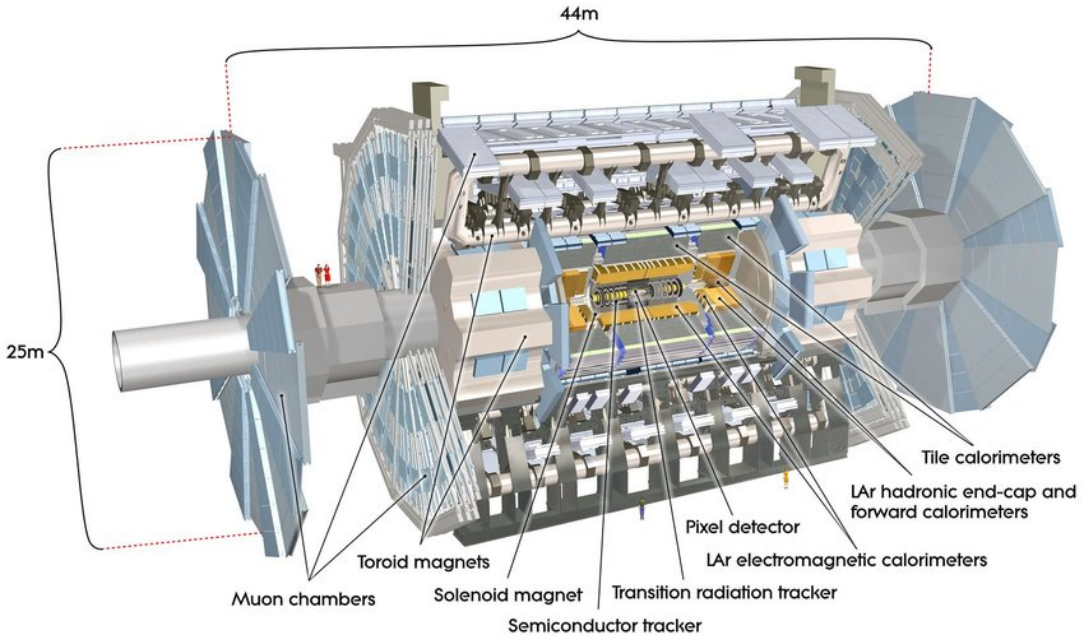


Figure 2.5: *A general view of the ATLAS detector and its sub-systems [97]*

The various sub-detectors are described in more detail, emphasizing those which play an important role in the development of this thesis.

2.2.1 The ATLAS coordinate system

ATLAS uses a right-handed coordinate system with its origin at the nominal IP, where protons collide. The x-y plane is transversal to the detector with the x-axis pointing towards the center of the LHC and the z-axis placed along the beam pipe. Being ATLAS shaped as a cylinder, it makes cylindrical coordinates useful to describe the geometry of the interactions. The azimuthal angle ϕ is defined around the beam axis and instead of using the usual polar angle θ, particle physicists generally use the *pseudorapidity* η,

$$\eta = -\ln \tan\left(\frac{\theta}{2}\right). \tag{2.6}$$

This definition is an approximation valid in the massless limit of the rapidity y,

$$y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right), \tag{2.7}$$

where E is the energy of the particle and p_z is the projection of its momentum over the z-axis. Differences in rapidity Δy (and hence in $\Delta\eta$) are Lorentz invariant under boosts along the longitudinal axis, making this variable suitable for hadron colliders since the colliding partons carry different longitudinal momenta, providing a boost to the rest frame of the collision with respect to the laboratory frame. Angular distances between two objects in the detector are defined as $\Delta R^2 = \Delta\eta^2 + \Delta\phi^2$, which thanks to the definition of the pseudorapidity is also Lorentz invariant under boosts along the z-axis.

2.2.2 Inner detector

The Inner Detector (ID) aims to provide robust trajectory reconstruction together with excellent momentum and vertex resolution of charged particles within $|\eta| < 2.5$. The necessary magnetic field to bend particle trajectories is provided by a solenoid coil aligned with the beam axis, providing a 2T magnetic field. Its design is driven by the required excellent calorimeter performance which can be severely affected by insensitive material upstream the ID. The total thickness of the solenoid contributes with ~ 0.66 radiation lengths X_0 (defined later in Section 2.2.3) at normal incidence to the traversed material before particles reached the calorimeters.

Three independent subdetectors compose the ID: the pixel detector, the silicon microstrip detector and the transition radiation detector. Front and lateral cut-offs of the ID are shown in Figures 2.6 and 2.7, where the various subdetectors and dimensions are indicated.

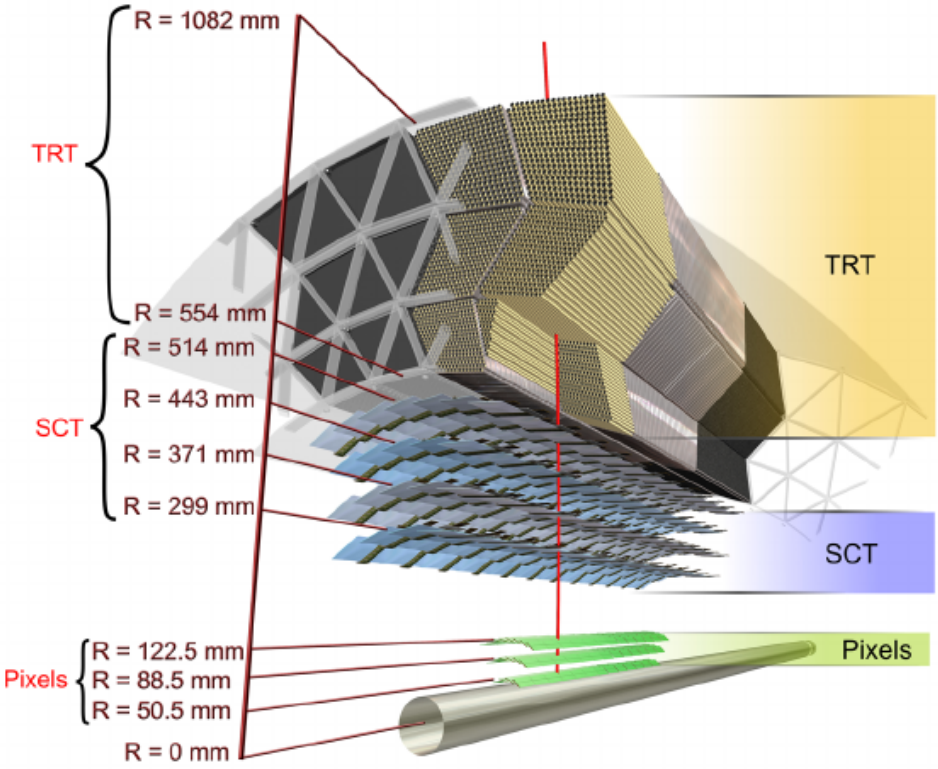


Figure 2.6: *Detail of sensors and subsystems of the Inner detector in the barrel. Dimensions of each component are shown in the vertical axis. A hypothetic particle trajectory is shown in red traversing the detector. [97]*

2.2.2.1 Pixel detector

The pixel detector is the closest sub-system to the beam axis and makes use of pixel sensors to detect charged particles. It was originally designed with three coaxial pixel layers in the barrel and four perpendicular disks in the endcaps.

Each of the layers is composed by 1744 pixel sensors with a total of ~ 80 million channels. Each pixel sensor is a semiconductor silicon sensor with 50×400 μm in size, providing a single pixel spatial resolution of 10 μm in the $(R-\phi)$ direction and 115 μm

in the z direction. Similar resolutions are achieved in barrel and endcaps of the pixel detector.

A fourth layer was installed in the barrel between 2012 and 2015, during the LHC long shutdown 1, to provide measurements closer to the IP and better radiation damage resilience. This layer is called the Insertable B-Layer (IBL) and it has a smaller pixel size of 25×400 μm. It is a crucial element for b-jets identification, since it provides a robust reconstruction of the B-hadron vertex decay.

2.2.2.2 Semiconductor tracker (SCT)

Covering the pixel detector, four layers in the barrel and nine in the endcaps of silicon microstrip detectors provide a second independent measurement of charged particle trajectories. The SemiConductor Tracker (SCT) covers a similar range in pseudorapidity as the pixel detector up to $|\eta| < 2.5$. Each silicon microstrip detector is composed by two sensor strips arranged back-to-back with a relative angle of 40 $mrad$ that provide precise position measurements with a resolution of 17 μm in the $(R-\phi)$ direction and 580 μm in the z direction. The total number of readout channels in the SCT is approximately 6.3 million.

2.2.2.3 Transition radiation tracker (TRT)

The last subsystem located inside the solenoid coil is the TRT. Out of the three components in the ID, the TRT is the only detector capable of providing both tracking information and particle identification, in particular electron and charged pion identification for particle energies between 1 and 200 GeV. The detector consists of an assembly of almost 300.000 proportional drift tubes ("straws") with a diameter of $4mm$ covering up to $|\eta| < 2.0$. In the barrel the straws are divided in layers coaxially to the beam axis, while radially in the endcaps. The space between layers is filled with polyethylene fiber (barrel) or polypropylene foil (endcap). Each drift tube is filled with a mixture of gases (Xe, CO_2, O_2)[2] and made of aluminum with a gold-plated tungsten wire running along its center. The wire is used to read out the signal from particles producing transition radiation emissions.

Transition radiation is emitted when charged particles go through inhomogeneous media with different dielectrics. This radiation lies in the X-ray range with energies between 5-100 keV and is strictly proportional to the relativistic γ factor of the incident particle, providing mass discrimination for different particles. Lighter particles are more likely to emit transition radiation than heavier particles, allowing the discrimination of electrons with respect to heavier particles as for example muons or pions.

The TRT measures the $(R-\phi)$ position of the trajectory, with a resolution of 130 μm per drift tube. This resolution is worse compared to those of the SCT or pixel detector, but this is compensated a larger number of hits per track and longer measured track length.

2.2.3 The calorimeters

Calorimeters in high-energy physics are blocks of material used to completely stop incoming particles. Incoming particles interact with the material of the calorimeter and

[2]Other gases, like Ar may be included during nominal LHC operation to clean impurities on the anode wires.

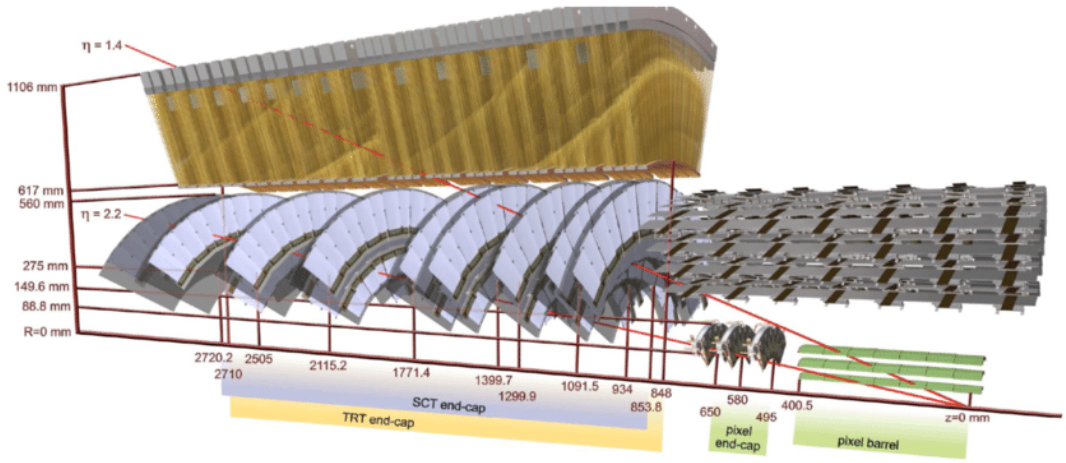


Figure 2.7: *Cut-off of the Inner Detector illustrating barrel and endcap differences in the arrangement of sensors for each component. The TRT barrel is not shown.[97]*

deposit all of their energy, which is then transformed into a measurable signal. Particles entering the calorimeter material initiate a particle shower of daughter particles from which properties of the primary particle can be inferred. Not only the energy of the incident particle can be provided from this measurement, but also the spatial location and sometimes the direction and timing depending on the design.

Depending on which particles the calorimeters are designed to detect, calorimeters can be broadly classified in two categories:

- Electromagnetic (EM) calorimeters: used to measure the energy of particles which interact with the material of the calorimeter through the electromagnetic interaction.
- Hadronic calorimeters: used to measure the energy of particles which interact with the material of the calorimeter through the electromagnetic or strong interaction.

Two elements are defined for the different parts of a calorimeter: absorber and active layers. Absorber layers are used to stop the incoming particles and initiate the creation of secondary particles. They are usually made of a dense material but since they have no sensitive material, no measurement can be performed in these layers. Active media layers are the actual detector of the calorimeter in which through the interaction with the particles from the shower a signal is measured.

Calorimeters are usually classified according to their amount of absorber material:

- Homogeneous calorimeters: constructed solely with one material that plays both roles of absorber and active medium. Only used for EM calorimeters, since very large size detectors would be needed to stop particles interacting hadronically.
- Sampling calorimeters: constructed alternating layers of absorber and active material. Only part of the energy is measured since a fraction of it is lost in the absorbers but their construction is usually cheaper compared to homogeneous calorimeters. This design is used for both hadronic and electromagnetic calorimeters.

Before describing in detail the experimetal calorimetric setup of the ATLAS experiment, a brief description of how EM particles interact with matter and some notions on precision calorimetry are given, since both points are fundamental to understand the design and operation of the calorimeters.

2.2.3.1 Interaction of electromagnetic particles with matter

Electromagnetic interactions between particles and matter vary drastically with the energy of the particle, from a few eV where the photoelectric effect dominates up to GeV levels while pair production is the main contribution for photon interaction cross-sections. Among all those processes, this section reviews the ones responsible for particle showering, being pair production for photons and bremsstrahlung for electrons; and those involved in the measurement of signals in the active layers of the calorimeter.

Two main processes contribute to energy losses of electrons and positrons when passing through matter: elastic collisions with other charged particles in the media and radiative processes. Both processes contribute to the emission of radiation of some kind. The latter ionizes electrons from the surrounding atoms while the former, clasically, radiates due to the deceleration induced on the electron when interacting with an un-screened nucleus. The total energy loss E_{tot} per unit distance x from electrons when traversing matter can be described as:

$$\frac{dE_{tot}}{dx} = \frac{dE_{ion}}{dx} + \frac{dE_{rad}}{dx}, \tag{2.8}$$

where x is usually expressed in units of mass by traversed surface, g/cm^2.

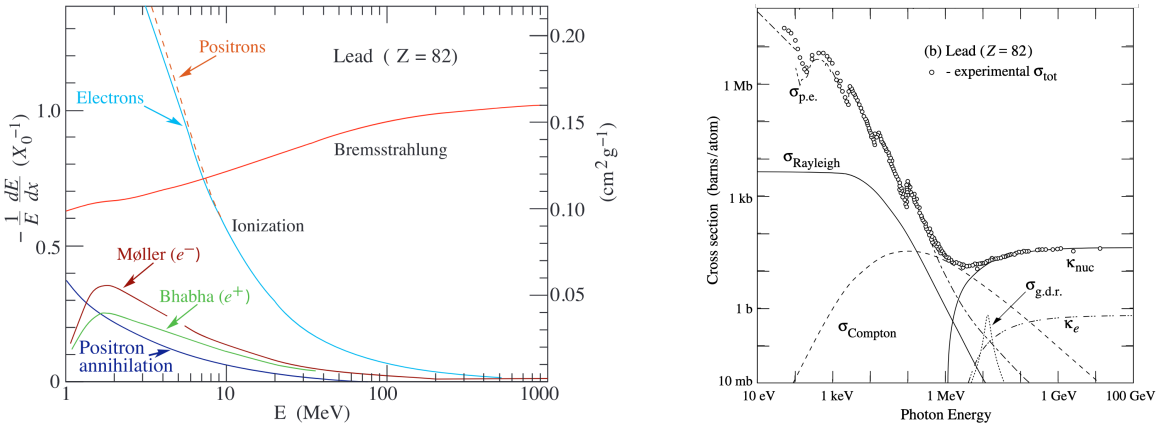
At low energies, incoming electrons interact with atomic electrons through ionization processes, losing a fraction of their energies while traversing the material. This process is described by the Bethe-Bloch formula [7, 98]. It is the mechanism through which charged low energy particles deposit energy in the detectors. Moreover, it is in principle used to measure the energy of the incident particle since it is proportional to the number of ionized atomic electrons.

At higher energies radiation is emitted by accelerated charged particles when passing through matter. This radiation is called *bremsstrahlung* and it is the primary source of energy losses for electrons at high energies. It is caused by the atomic electric field that the electron traverses, which is a function of the screening of the valence electrons around the nuclei. This makes the process dependent on the material and also on the distance electron-nucleus. The different contributions to the total energy loss of electrons interacting with matter are shown in Figure 2.8a.

Since photons have no electric charge, their energy loss processes differ significantly from those of electrons. Although Compton scattering and photoelectric effect contribute to the total photon interaction cross section, in the GeV level a process is dominant: *pair production*. Figure 2.8b shows the photon interaction total cross section as a function of the energy with the different contributions from different processes.

Photon pair production involves the production of an electron-positron pair through its interaction with a nearby nucleus. The presence of a nucleus is crucial since otherwise momentum is no conserved. Moreover, the energy of the photon is required to be larger that the sum of masses of both electrons (1.022 MeV).

When entering the calorimeter, a high energy photon interacts with matter producing electron-positron pairs, which then emit lower energy photons through bremsstrahlung. The iteration of these two processes creates what is called an *electromagnetic shower*, that spans in the calorimeter until the energy of the electrons falls below the critical energy E_c, defined as the energy at which the losses from ionization and bremsstrahlung are the same (equal contributions in Equation 2.8). From this point onwards, electrons lose their energy through ionization processes without further photon emission. Once the end of the shower development has been reached, the total number of particles created in the shower is proportional to E_0/E_c, being E_0 the energy of the incident particle. Out



a) b)

Figure 2.8: *(a) Fractional energy loss per radiation length in lead as a function of electron or positron energy. The value at which ionization losses and radiation losses are equivalent defines the critical energy. [7] (b) Photon total cross sections as a function of energy in lead (Z=82), showing the contributions of different processes. In the high energy regime (above 1 GeV) the two relevant contributions are photon pair production: κ_{nuc} and κ_e stand for pair production of photons interacting with either the nuclear field or the electron field respectively. At lower energies, Compton effect ($\sigma_{Compton}$) and photo-electric effect ($\sigma_{p.e}$) dominate. [7]*

of those, charged particles, through ionization, create a measurable signal. The signal measurement procedure is described in the following section.

Electromagnetic showers can be characterized by their longitudinal and transversal development. The longitudinal profile of the shower can be parametrized as a function of the *radiation length*, X_0, defined as the distance of traversed material over which an electron has lost on average $1/e$ of its energy due to radiative processes. It is usually expressed in units of mass by traversed surface, g/cm^2, and it clearly depends on the properties of the material.

The longitudinal profile is fairly described by the following analytical function:

$$\frac{dE}{dt} = E_0 \beta \frac{(\beta t)^{\alpha-1} e^{-\beta t}}{\Gamma(\alpha)}; \;\; t = \frac{x}{X_0}, \tag{2.9}$$

where α and β depend on the material and incoming particle. Two useful quantities arise from this expression:

- First, the distance in radiation lengths at which the energy loss is maximal can be described as $t_{max} = \frac{\alpha-1}{\beta} = ln\left(\frac{E_0}{E_c}\right) - C_{e\gamma}$, with $C_{e\gamma}$ being 1.0 for photons and 0.5 for electrons. This result states that the position of the core of the EM shower increases logarithmically with E, which implies that the penetration power barely increases with the energy of the incident particle (a factor 10 in energy translates to $\sim 2X_0$ in depth).

- Second, the distance in radiation lengths which contains 95% of the longitudinal shower development is $t_{95\%} = t_{max} + 0.08Z + 9.6$. This results explains why for very energetic EM particles, if a calorimeter is not deep enough, some energy may leak outside of it.

Figure 2.9 shows the relative energy loss in a calorimeter for different incident electron energies.

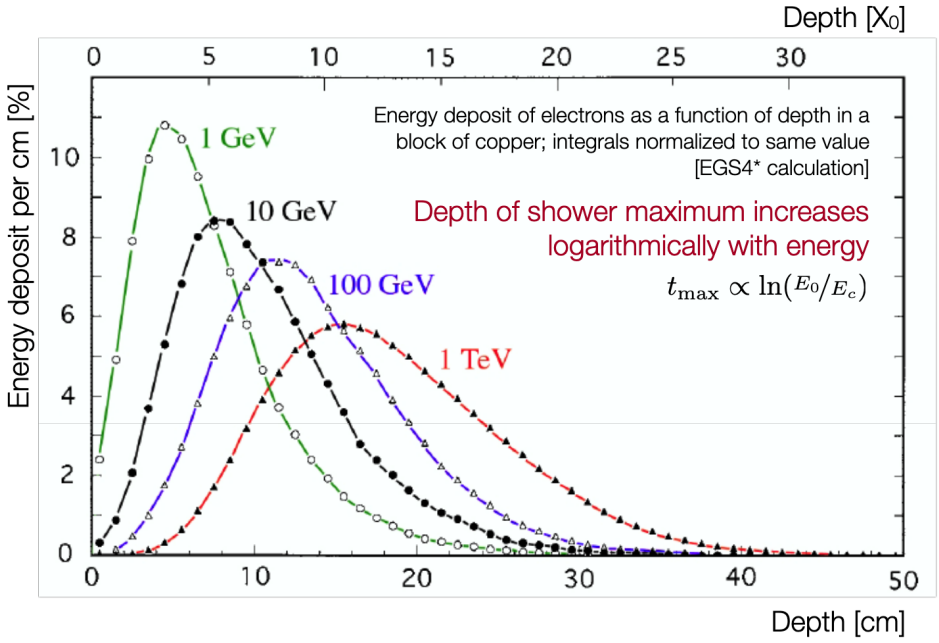


Figure 2.9: *Relative energy loss by incident electrons of various energies in copper as a function of the depth in the calorimeter. t_{max} (see text) describes the evolution of the position of the peak with increasing energy. Larger energy electrons have broader showers in the longitudinal direction. [99]*

The lateral profile is dominated by low energetic multiple-scattering effects between pair-produced electrons and nuclei in the calorimeter. This type of development makes the transversal size of an EM shower almost independent of the energy. Ignoring large angle scatterings, the core of the lateral development can be fairly described by the *Molière Radius*, given by:

$$R_M = X_0 \frac{E_s}{E_c}, \tag{2.10}$$

with $E_s \approx 21$ MeV. The Molière Radius is a parameter that characterizes the lateral profile of the shower since it is the transversal distance that contains 90% of the energy of a shower, respectively. Moreover, $2R_m$ and $3.5R_m$ contain 95% of 99% of the energy of the shower. Even if there is always some fraction of the energy of the EM shower that is not contained within various R_M, it is nevertheless a negligible fraction of the energy. In Table 2.2 relevant values for the characterization of EM showers in Liquid-Argon and Lead are shown, since these two materials are used in ATLAS EM calorimeters.

Table 2.2: *Lead and Liquid-Argon properties.*

	Lead (Pb)	Liquid-Argon (LAr)
Radiation length (X_0) [cm]	0.56	14.0
Molière radius (R_M) [cm]	1.6	9.0
Critical energy (E_c) [MeV]	7.4	32.8

2.2.3.2 Calorimetric energy resolution

Energy measurements using calorimeters are based on the principle that the initial energy of the incoming particle is proportional to the energy measured through the ionization/excitation of the material by the secondary particles. Since these are stochastic processes, an ideal calorimeter would be already limited by fluctuations when performing energy measurements. However, real life calorimeter resolutions are also subject to other contributions. In general, the energy-dependence of the resolution of a calorimeter can be parametrized as:

$$\frac{\sigma_E}{E} = \underbrace{\frac{a}{\sqrt{E}}}_{\text{Stochastic term}} \oplus \underbrace{\frac{b}{E}}_{\text{Noise term}} \oplus \underbrace{c}_{\text{Constant term}} . \tag{2.11}$$

Three different sources are identified:

- Stochastic term: this term comprises the fluctuations in the number of secondary particle signals detected to reconstruct the energy of the initial particle. For our purposes, two sources of fluctuations can be identified: intrinsic fluctuations in the number of ionizations/excitations along the shower development and sampling fluctuations. The first source is present in any calorimeter, being $E \propto \left\langle N_{\text{particles}} \right\rangle \rightarrow \sigma_E \propto \sqrt{\left\langle N_{\text{particles}} \right\rangle}$. In a homogeneous calorimeter, this is the intrinsic resolution arising from fluctuations since the particles, if fully stopped, deposit all their energy in the active medium. It usually contributes with a few per-cent to the total energy resolution (3% in CMS). The second source affects only sampling calorimeters and arises from fluctuations in the energy deposited in the absorbers (non-active layers). With this second contribution, the stochastic term in sampling calorimeters can be between $5-20\%$ (10% in ATLAS).

- Noise term: this term comprises the contribution of electronic noise in the readout chain and pileup noise. The contribution of this term is in general negligible at higher energies.

- Constant term: the last term includes contributions that are energy independent like instrumental defects or non-linearities. Since they arise from unexpected defects, corrections are difficult to implement. It is the dominant contribution at higher energies.

2.2.3.3 ATLAS Electromagnetic Calorimeter

The ATLAS EM calorimeter is a sampling calorimeter alternating layers of lead as absorber and Liquid-Argon (LAr) as active media. The overall resolution of the EM calorimeter is:

$$\frac{\sigma_E}{E} = \frac{10\%\sqrt{\text{ GeV}}}{\sqrt{E}} \oplus \frac{0.25 \text{ GeV}}{E} \oplus 0.7\%, \tag{2.12}$$

where E is measured in GeV. Each of the coefficients may vary depending on the region of the detector. Designed with an accordion shape the EM calorimeter is divided in three different sectors: one barrel calorimeter and two endcaps, as shown in Figure 2.10.

The barrel region is 6.4m long and it is divided into two half-barrels covering $|\eta|$: $[0, 1.475]$ with 2.8m and 4m of inner and outer diameters. A single half-barrel is composed

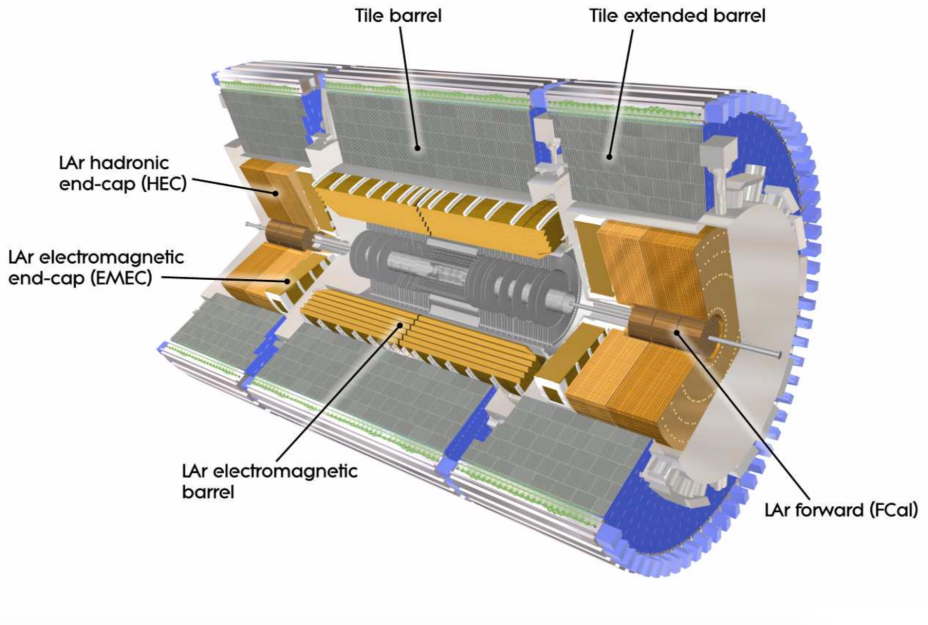


Figure 2.10: *Scheme of ATLAS electromagnetic and hadronic calorimeters. [97]*

by 1024 accordion-shaped lead absorbers interleaved with electrodes to gather the ionized charges. The total amount of radiation lengths in the barrel EM calorimeter is at least $22X_0$, from $22X_0$ to $30X_0$ below $|\eta| < 0.8$ and from $24X_0$ to $33X_0$ until the end of the barrel (Figure 2.11). The electromagnetic endcaps (EMEC) consist of two wheels of $\sim 2m$ of diameter placed at each side of the barrel, extending from $|\eta| = 1.375$ up to $|\eta| = 3.2$. A single EMEC wheel is composed by two co-axial wheels: an outer wheel further from the beam pipe covering $|\eta| : [1.475, 2.5]$ and an inner wheel closer to the beam pipe covering up to $|\eta| = 3.2$. Each outer wheel contains 768 lead absorbers compared to the inner wheels composed by only 256. Compared to the barrel, slightly larger amounts of material are found in the EMECs, increasing the total radiation lengths which in this region span from $24X_0$ up to $36X_0$ (Figure 2.11).

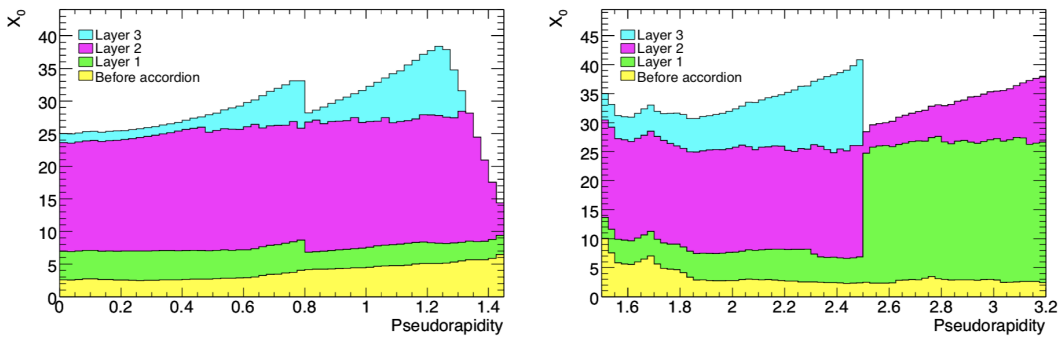


Figure 2.11: *Cumulative amounts of material, in X_0 units, present upstream and in the calorimeters as a function of $|\eta|$ for the barrel (left) and endcaps (right). The discontinuity at $|\eta| = 0.8$ is due to a change in the thickness of the lead absorbers, being 1.53 mm for $|\eta| < 0.8$ and 1.13 mm for $|\eta| > 0.8$. [97]*

The *accordion geometry* of the calorimeter is designed to accommodate LAr as the active media. Successfully used in previous calorimeters (DØ [100], NA31, ...), LAr posed

a challenge in ATLAS due to its cylindrical design and the desire to have a full ϕ coverage. Standard LAr calorimeters interlay absorbers and active media gaps perpendicular to the direction of the particle. The ionized electrons produced by the shower are drifted by a high voltage (HV) electric field and collected by an electrode placed in the middle of the LAr gap of 2mm size. This structure needs ground lines and wires to transport the charge collected in the electrodes on both sides of each layer, in conflict with a cylindrical geometry design. Using an accordion geometry, readouts can be placed at the back or in front of the calorimeter providing full coverage in ϕ. A section of the barrel calorimeter showing the accordion geometry is illustrated in Figure 2.12a.

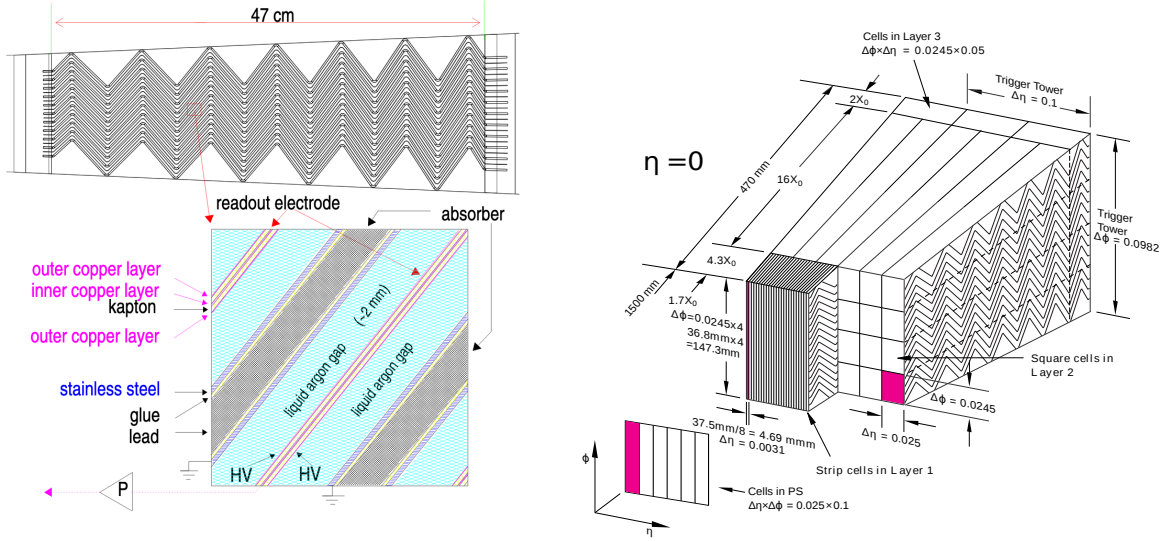


Figure 2.12: *(a) Accordion structure of the barrel EM calorimeter. [101] (b) Sketch of the barrel electromagnetic calorimeter at $|\eta| = 0$. The granularity in η and ϕ is shown together with the sizes of each. [97]*

Waves (or bend direction) of the accordion layers are placed differently along the detector in order to provide an homogeneous LAr shape with increasing $|\eta|$. In the barrel, the accordion waves are axial and for a given ϕ, the same angle is used. On the contrary, in the EMECs waves are parallel to the radial directions and run axially. With increasing $|\eta|$ the bending angle changes, as illustrated in Figure 2.13 for the barrel.

Sampling calorimetry benefits from the possibility of adding extra longitudinal segmentation as opposed to homogeneous calorimeters. The ATLAS EM calorimeters are divided in three different layers[3] in the barrel and in the outer wheels of the EMECs. The inner wheels are only divided into two layers. The layers are designed with a particular purpose each:

- The first layer (L1 or *strip* layer as referred later in the text) is designed with a high granularity in order to identify photon energy deposits compared to $\pi^0 \to \gamma\gamma$, which leaves two energy deposits. This extra care is needed since usually the π^0's are boosted and hence the products of their decay are collimated in the detector and can be mis-identified as a single photon.

[3]Also denominated samplings in ATLAS. Both terms are used interchangeably in the following

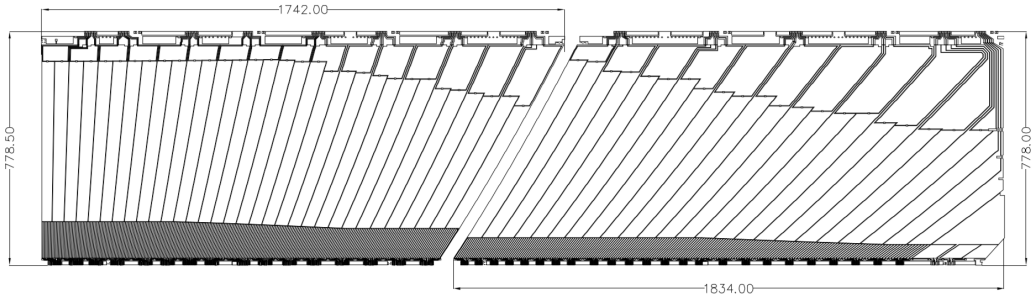


Figure 2.13: *Layout of the two calorimeter electrodes in the barrel before folding (no accordion structure is visible). Since the bending angle changes with $|\eta|$, the orientation of the cells also changes. The three longitudinal layers of the calorimeter are clearly visible. [97]*

- The second layer (L2) aims to stop (almost) completely the incident particles, and hence measure the largest fraction of their energy compared to the other samplings.

- The third layer (L3) is used to better contain the energy of electromagnetic particles, collecting the tail of the electromagnetic shower.

The three layers are often referred to in the following as L1,L2,L3 respectively. A detail of the longitudinal layers is shown in Figure 2.12b.

The LAr calorimeters are immersed in a cryostat to keep the necessary temperature for the detectors to operate; which together with all the material from the inner detectors degrade the energy measurement. A presampler is a thin layer (11mm in the barrel and 8mm in the endcaps) of active material added before the calorimeters to correct for the energy loss in the cryostat and other elements found upstream the calorimeter. It covers up to $|\eta| = 1.8$. To mitigate energy resolution degradation in the gap between barrel and endcap cryostats due to large amounts of material, scintillators are placed in order to provide a complementary energy measurement. These scintillators, referred to as "E4" scintillators cover the $|\eta| : [1.4, 1.6]$ region, also denominated "crack region".

2.2.3.4 ATLAS Hadronic Calorimeter

While electrons and photons are completely stopped in the EM calorimeter, hadrons have not traversed through enough material to interact and lose all their energy. Hadronic showers are also developed through the interaction of hadrons with the material of the calorimeters. Hadronic interactions are parametrized in terms of the interaction length λ_{int}, defined as the mean free path an hadron travels through matter before interacting with the nuclei, analogously to X_0 in EM showers. However, λ_{int} is in general larger than X_0, and this implies that hadronic calorimeters are usually "thicker" compared to EM calorimeters (Figure 2.14).

In order to provide precise jet measurements, a hadronic calorimeter is placed downstream the EM calorimeters to completely stop hadrons and other particles escaping the EM calorimeters (except muons and neutrinos). Figure 2.10 shows the complete calorimetric system layout.

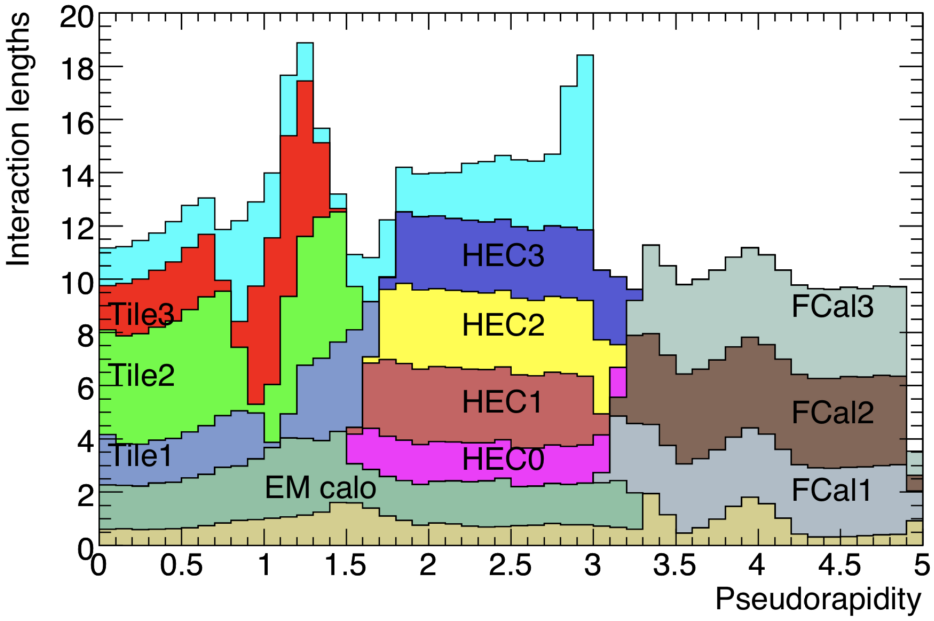


Figure 2.14: *Cumulative amount of material, in units of interaction lengths λ_{int}, as a function of $|\eta|$ in front of the calorimeters, in the calorimeters and in the first layer of the muon spectrometer. [97]*

Tile calorimeter

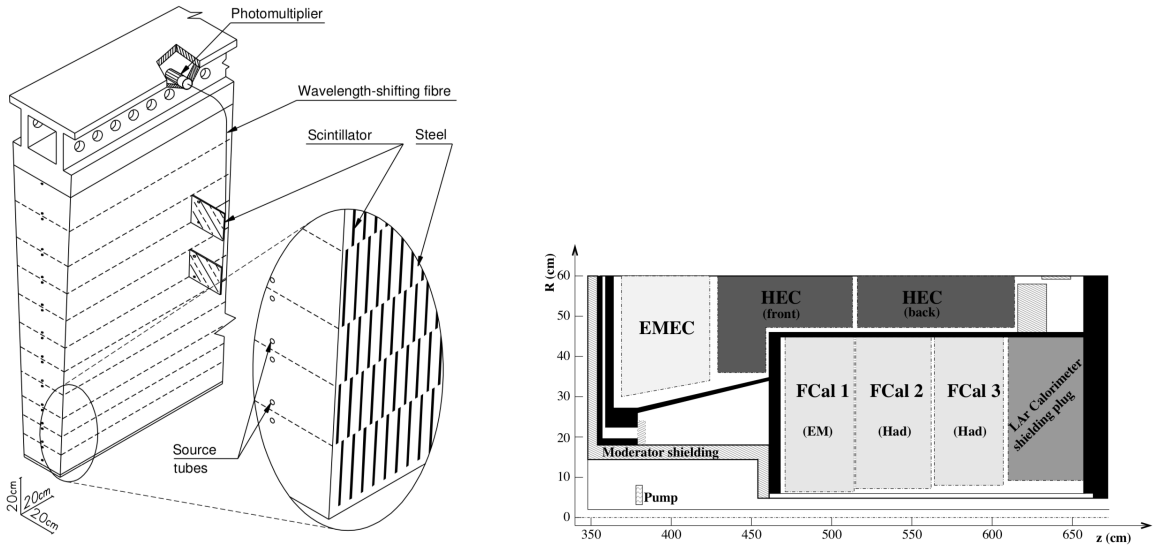
The tile calorimeter is a sampling hadronic calorimeter with steel absorbers and plastic scintillator tiles as active media divided in two sections: the tile barrel calorimeter which covers up to $|\eta| = 1$ and the extended barrel up to $|\eta| = 1.7$. Charged particles traversing the active media produce light in the ultraviolet spectrum. The collected radiation is then shifted to the visible range through a wavelength-shifting optical fiber connected to a photomultiplier where the signal is measured. The absorbers and active layers are placed radial and perpendicular to the beam pipe (Figure 2.15a) since no electrodes are present compared to the LAr calorimeter. Similarly to the LAr calorimeter, it is also segmented longitudinally into three layers.

Hadronic endcap calorimeters

The hadronic endcap calorimeters (HEC) are placed on both sides of the tile calorimeter, covering $|\eta| : [1.5, 3.2]$. Contrary to the tile calorimeter, the HEC are sampling calorimeters using the LAr technology, detailed in the previous section, but using copper instead of lead as absorber and using flat-shapped layers. Each HEC is composed of two wheels sharing the same cryostat as the LAr EMECs.

2.2.3.5 Forward Calorimeters

Forward calorimeters (FCal) cover the region closest to the beam, $|\eta| : [3.1, 4.9]$. The FCals are sampling calorimeters using LAr as active medium and they are segmented longitudinally in three layers: the first one designed to detect forward EM particles uses copper absorbers; while the other two use tungsten absorbers. LAr gaps are thinner compared to the EM calorimeter(2mm), increasing from 0.21mm in the first layer to



a) b)

Figure 2.15: *(a) Sketch of the assembly of absorbers and scintillators in the tile barrel and extended calorimeter orthogonal to the beam pipe. The necessary elements for signal readout (photo-multiplier and wavelength-shiftter) are also shown. [97] (b) Schematic diagram of the elements in the forward region of the calorimeters in the R-z plane. Black regions are structural parts of the cryostat, necessary for the systems using Liquid-Argon. [97]*

0.57mm in the third one. Figure 2.15b shows how all elements in the endcaps are arranged and the cryostat disposition.

2.2.4 The Muon Spectrometer

The outermost layer of detectors is the muon spectrometer (MS), a tracking detector which aims to measure muon momenta using gaseous detectors and air-core toroidal magnets. It can be considered as an almost standalone muon detector separated from the rest of ATLAS, since it has its own magnetic, tracking and trigger systems (Figure 2.16).

The magnetic system is composed by three toroidal magnets (one in the barrel and one per endcap) that provide the necessary bending power to precisely measure the momenta of muons. A conceptual layout of the magnetic system used in ATLAS is shown in Figure 2.17, which includes the previously mentioned solenoid used in the tracking system.

The MS tracking system covers up to $|\eta| < 2.7$ with specific technologies for different regions of the detector. Up to $|\eta| < 2$, three layers of Monitored Drift Tubes (MDTs) provide precise measurements in the principal direction of the magnetic field. An MDT chamber consists of 6 layers of straw tubes filled with a gas mixture. When ionized by an incident muon, it produces electrons that are then collected by a wire placed at the center of the tube. For larger pseudorapidities up to $|\eta| < 2.7$, the MDTs are replaced by multiwire proportional chambers with cathodes segmented into strips called Cathode Strip Chambers (CSCs). This change in technology is driven by the increase of incoming radiation at larger pseudorapidities, as the sensitive material is placed closer to the beam pipe.

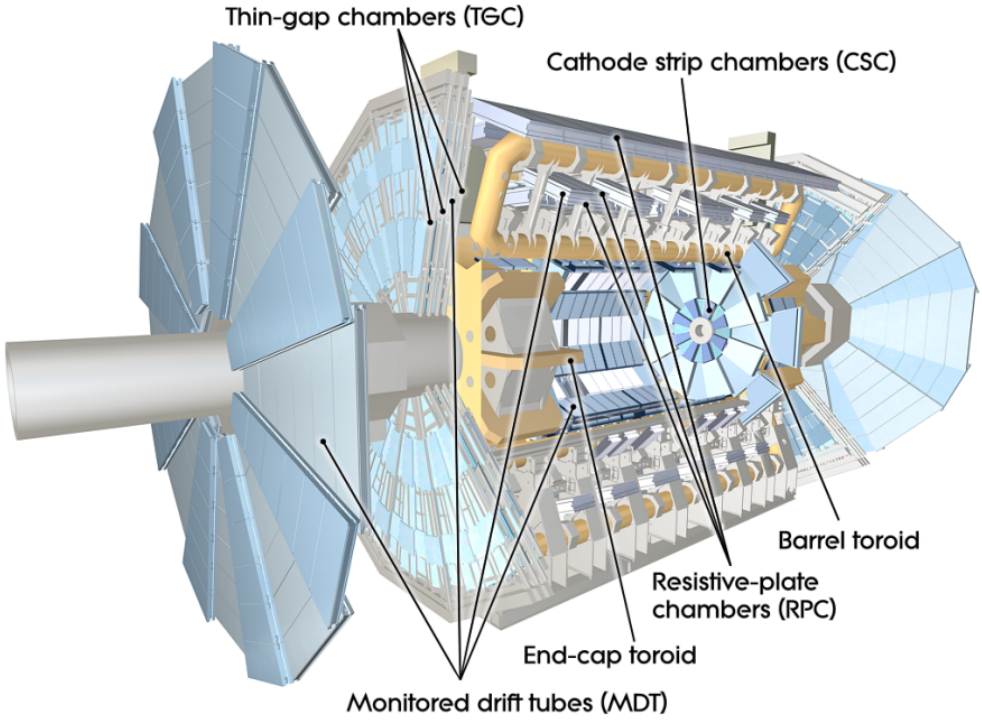


Figure 2.16: *Structure of the ATLAS muon spectrometer. [97]*

The muon trigger system is composed by two different technologies: Resistive Plate Chambers for the barrel regions and Thin Gap Chambers for the endcaps. These chambers are based on coincidences of muon hits in 2 or more layers that match with those from a muon track above a minimum p_T threshold. Besides trigger information, they also measure the coordinate in the orthogonal direction to the one provided by the precision tracking chambers.

2.2.5 Trigger and Data Acquisition System

The pp collision rate is 40 MHz, thus making it impossible to record and reconstruct every single event with the current technologies. Instead, a real-time selection of potentially interesting events is performed using software and hardware triggers that reduce the rate down to kHz levels.

A first level (L1) hardware trigger [102] uses coarser-granularity signals from calorimeters and muon chambers to skim the incoming data flow and reduce the rate a factor ∼400. The L1 spends 2.5 μs to decide which events are selected. Simultaneously, Regions of Interest (RoI) in the detector are defined with the energy deposits identified in the L1 trigger which are then processed by a High-Level Trigger (HLT). HLT uses software implemented algorithms which reduce the rate down to acceptable kHz levels. At this stage, full detector information is available from each of the systems: tracking information from the ID, fine-granularity from the calorimeters and precision measurements from the muon spectrometer.

The reduction factors are achieved by the L1 and HLT triggers in combination with what are called prescales. The prescale of a given trigger is the number of times that it needs to be fired in order to select and record an event with that trigger. The large rate of certain final states (like a photon with very low energy) leads to prescales of the

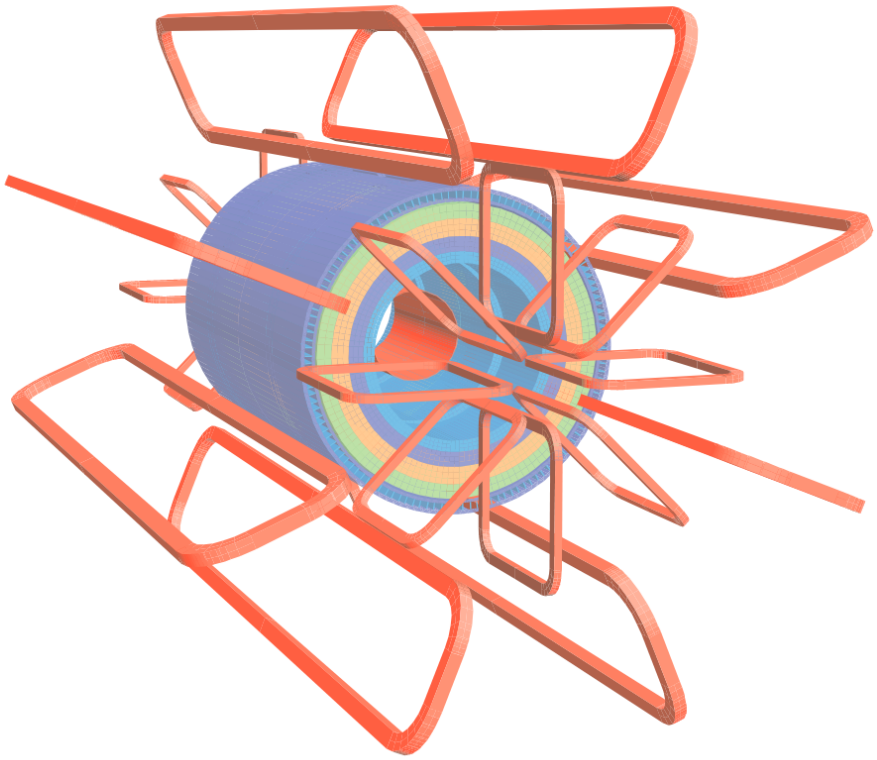


Figure 2.17: *Conceptual layout of the magnet system in ATLAS. [97]*

order of 1000 or even higher. Unprescaled triggers are those that when fired are always recorded (equivalent to have a prescale equal to unity).

A trigger chain in ATLAS is defined as the combination of an L1 item and an HLT item, built in order to identify interesting physics signatures with one or more objects in the final state, fulfilling certain conditions and over a certain energy threshold. The structure of both L1 and HLT items can be sketched by the following expressions:

```
L1_[number of seeds][type seed][energy threshold]_[conditions],
HLT_[number of objects][object type][energy threshold]_[conditions].
```

As an example, since photon (and diphoton) triggers play an important role in the rest of this manuscript, one particular diphoton trigger item is detailed in the following for the diphoton trigger chain: `HLT_2g20_tight_icalovloose_L1_2EM15VHI`.

First, a L12EMXX trigger is fired if two L1EMXX seeds are fired at L1, where XX corresponds to the energy threshold of the L1 item. A L1EM15XX seed fires a trigger when scanning the calorimeter with coarser cells, denominated towers of size $\eta \times \phi = 0.1 \times 0.1$, registers an analog sum of transverse energy E_T larger than the energy threshold, 15 GeV in the example. The sum is computed on windows of 2×2 towers from the four possible pairs of nearest-neighbour towers in the EM calorimeter. Additionally, other requirements are applied in order to further reduce the event rate. Three conditions are applied in the example at L1:

- V: refers to a small **v**ariation of the energy threshold as a function of $|\eta|$, since different regions of the detector are calibrated differently. This variation can modify the baseline energy threshold between -2 and +3 GeV.

- H: in order to reject **h**adronic activity, electron and photon candidates can be rejected if the sum of transverse energy in the hadronic towers matched to the 2×2 central region in the EM is larger than 1 GeV and exceeds $E_T/23.0 - 0.2$ GeV. This requirement is only applied for seeds with $E_T < 50$ GeV.
- I: an EM **i**solation requirement can also be applied to reject electron and photon candidates if the transverse energy found in the 12 EM towers surrounding the 2×2 central region is larger than 2 GeV and exceeds $E_T/8.0 - 1.8$ GeV. This requirement is only applied for seeds with $E_T < 50$ GeV.

A visual description of an L1EM seed is shown in Figure 2.18 to complete the explanation.

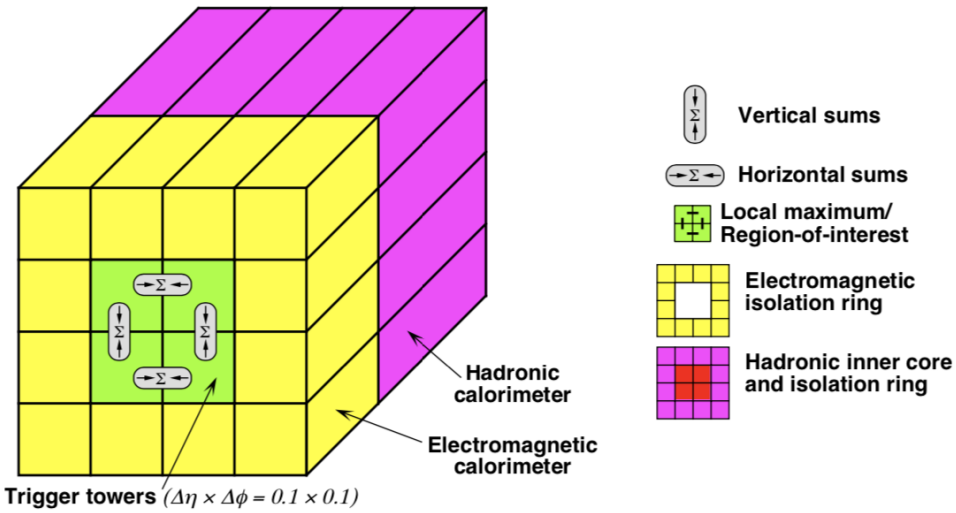


Figure 2.18: *Components of an L1EM seed [103]*

Over the RoIs identified by the L1 trigger, the HLT performs a refined selection over the candidates. The first part of the HLT item `2g20` corresponds to the number of particles of a certain type with E_T larger than a given threshold that are required to fire the trigger: 2 photons (gammas) of at least 20 GeV in the example. Then, other selections/criteria can be applied to further reduce the rate and/or backgrounds. Since this is a particular trigger, only two elements are required: `tight_icalovloose`. The first refers to the quality of the identification the photon candidate is required to attain in order to be triggered. A more detailed description on photon identification is provided in Chapter 3.3. The second requires the photons to be isolated from hadronic activity in the calorimeter, using a similar approach as explained in the L1 trigger. Since photon isolation corresponds to a significant contribution in this thesis, precise details can be found in Section 4.

A more exhaustive list of HLT items can be found in [102].

The output of the trigger system is passed at several stages to the Data Acquisition system (DAQ) that finally records events in the permanent storage [104]. A sketch of the ATLAS trigger and DAQ is shown in Figure 2.19. Accepted events by the L1 hardware trigger arrive to the Readout Drivers (RODs) which are temporarily stored in order to be available for the HLT, simultaneous to the sending of RoIs of the L1 to the HLT. If the events are also accepted by the HLT, the events are finally send to the Data Logger, where the full information of the event is stored.

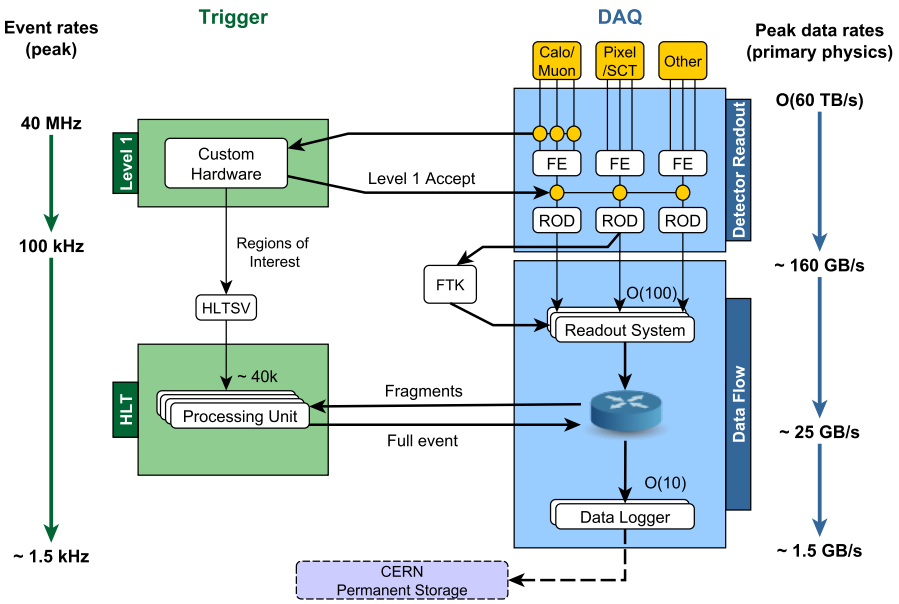


Figure 2.19: *Diagram of the Run 2 Data Acquisition system showing the different peak rates and bandwidths for each step along the data flow. [102]*

Chapter 3

Photon reconstruction and identification in ATLAS

Particles produced in *pp* collisions are reconstructed in ATLAS by converting raw signals from the detector into meaningful information through a set of techniques and algorithms. The two main building blocks used as input for the reconstruction are *tracks* built from hits in the ID and MS and *clusters* constructed from energy deposits in the calorimeters.

This chapter provides an overview of the reconstruction and identification of photons. Special emphasis is given to clustering algorithms, since clusters are widely used in the sections of this thesis devoted to performance studies of photon isolation.

3.1 Photon reconstruction

Electromagnetic showers developed by electrons and photons in the LAr calorimeter leave similar signatures and so their reconstruction proceeds simultaneously. Photons undergoing pair production before reaching the calorimeters are called converted photons, and called unconverted photons if they reach the calorimeter. Depending on the different signals in the detector, a particle hypothesis is made. An *electron* is defined as an object built from a cluster and a matched track. A *converted photon* is defined as a cluster matched to a conversion vertex. An *unconverted photon* is defined as a cluster without tracks or conversion vertices matched.

In this section, the reconstruction of tracks and clusters is described, as main building blocks for electron and photon reconstruction.

3.1.1 Energy measurement in the LAr calorimeter

Ionized charges produced in EM showers are drifted towards the electrodes located in the LAr gaps, read-out by the Frond End Boards (FEB). The main role of the FEB is to amplify and shape the signal collected by the electrodes. The signal is first pre-amplified, to suppress electronic noise downstream the electronic chain, and then shaped. The shaper is a bipolar $CR-(RC)^2$ filter which provides a null-integral signal. This feature helps suppressing the impact of out-of-time pileup on average, only degrading the signal due to fluctuations on the number of pileup interactions in the event. The shape of the ionization signal before and after the shaping is shown in Figure 3.1.

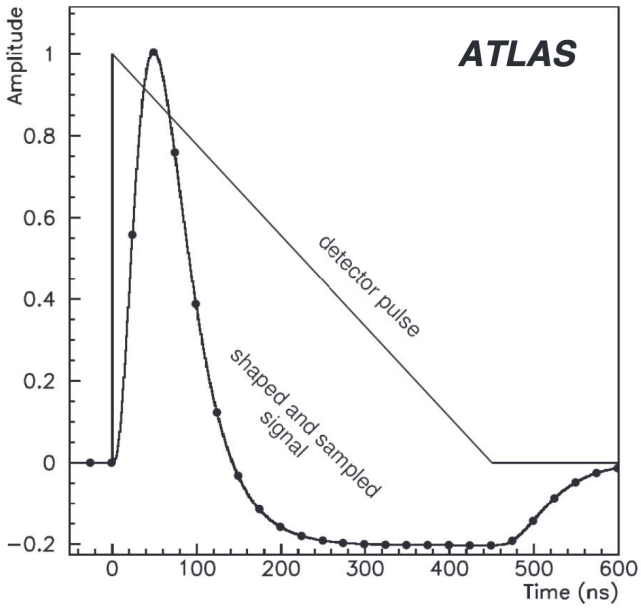


Figure 3.1: *The ionization signal from the ATLAS LAr calorimeter before (triangular) and after (curved) the shaper. Each dot denotes an ideal position of samples separated by 25 ns [105].*

The resulting signal is then sampled every 25 *ns*, amplified with three different gains (low, medium and high), and stored in a switch capacitor array until the Level-1 trigger decision is made. If the event is saved, the samples are digitized with an Analog-to-Digital Converter (ADC)[1]. The third sample of the ionization pulse, expected to be the maximum amplitude of the pulse, is used to estimate the optimal gain for the cell depending on the value of the ADC counts. The ADC counts saturates at an energy close to 3 TeV, so low signals are amplified with high gains and vice-versa to avoid saturations. The output ADC counts are used to reconstruct the energy and timing of the cell, together with information from simulations and test-beams. Energy (A) and time (t) are reconstructed in ADC counts, at cell level, according to the following expressions:

$$A = \sum_{j=1}^{N_{Samples}} a_j(s_j - p), \quad t = \frac{1}{A}\sum_{j=1}^{N_{Samples}} b_j(s_j - p). \tag{3.1}$$

In both expressions, p denotes the electronic pedestal corresponding to the mean value in ADC counts of the samples when no signal is present. The coefficients a_j and b_j are calculated according to an optimal filtering procedure which optimizes energy and timing resolution, referred to in the literature as Optimal Filtering Coefficients (OFC). These are determined using the expected shape of the signal. In Run 2, not all the samples are used to measure the energy/time of the pulse but only the first 4 contiguous are kept. The complete expression that converts ADC counts to energy in MeV is:

$$E_{cell} = F_{\mu A \to MeV} \times F_{DAC \to \mu A} \times \frac{1}{\frac{Mphys}{Mcali}} \times G_{ADC \to DAC} \times A, \tag{3.2}$$

where:

[1] A Digital-to-Analog Converter (DAC) is used in later steps to retrieve the input current.

- $F_{\mu A \to \text{ MeV}}$: conversion factor from the ionization current to the total energy deposited in the cell. Measured in test-beams [106], it depends on factors such as the sampling fraction, which is the energy deposited in the absorbers which is not measured.
- $F_{DAC \to \mu A}$: conversion factor that relates the digital ADC counts back to the injected current.
- $G_{ADC \to DAC}$: cell gain measured in calibration runs where a known signal is injected and reconstructed. It is expressed in ADC to DAC units.
- $\frac{1}{\frac{Mphys}{Mcali}}$: correction factor accounting for differences in the response produced by a physics signal and a calibration signal with the same input current.
- A: defined in Equation 3.1, is the energy in ADC counts.

The measured cell energies are the elementary units used for cluster building. The next subsections describe the different algorithms used for cluster reconstruction.

3.1.2 Clustering algorithms

A given object usually deposits energy in several cells of the calorimeter. A cluster is defined as a three-dimensional collection of cells that are clustered together if certain criteria are fulfilled. Two cell-clustering algorithms have been mainly used in electron and photon reconstruction: the *sliding-window* algorithm and the *dynamical topological cell clustering* algorithm.

3.1.2.1 Sliding-window algorithm

Electrons and photons (referred collectively to as *egamma objects* in the following) develop showers of transversal size defined by the Molière radius R_M. This fact makes feasible a tool that builds fixed-size windows to describe both egamma objects, since 90% of the energy is contained within a cylinder of radius 1 R_M. Moreover, the fact that sliding-window clusters have a fixed-size makes their energy calibration very precise compared to dynamically sized clustering algorithms, which need advanced multivariate calibration techniques.

The sliding-window algorithm [107] can be summarized in three steps:

1. The $\eta \times \phi$ space of cells in the calorimeter is described by a grid of $(N_\eta \times N_\phi)$ towers of size 0.025×0.025 $(\Delta\eta \times \Delta\phi)$. The energy of each tower comprises the cells from the three different samplings of the LAr calorimeter and the presampler.
2. A scan is performed with a window of size 5×5 towers looking for transverse energy local maxima. If a local maximum has $E_T > 2.5$ GeV, a precluster is formed. The position of the precluster is computed as the energy-weighted η and ϕ of the barycentres of each cell in the precluster. Then, duplicate preclusters are removed.
3. Clusters are built from the preclusters found in the previous step. The transverse energy of each cluster is computed as the sum of transverse energies of the cells in the fixed-size tower windows and the cluster position is the barycentre of the same group of cells. The possible window sizes are shown in Table 3.1 and depend on η and on the hypothesized particle that left the energy deposit. Converted photons

and electrons deposit energy in a larger windows in ϕ due to electron trajectories being bent by the magnetic field.

Particle Type	Barrel	Endcap
Electron	3×7	5×5
Unconverted photon	3×7	5×5
Converted photon	3×5	5×5

Table 3.1: *Cluster size in $N_{\eta}^{cluster} \times N_{\phi}^{cluster}$ for different particle types and regions of the EM calorimeter.*

This algorithm has been used in 2015 and 2016 data taking years to reconstruct electromagnetic particles. An illustration of the performance of the sliding window algorithm for simulated electrons is shown in Figure 3.2.

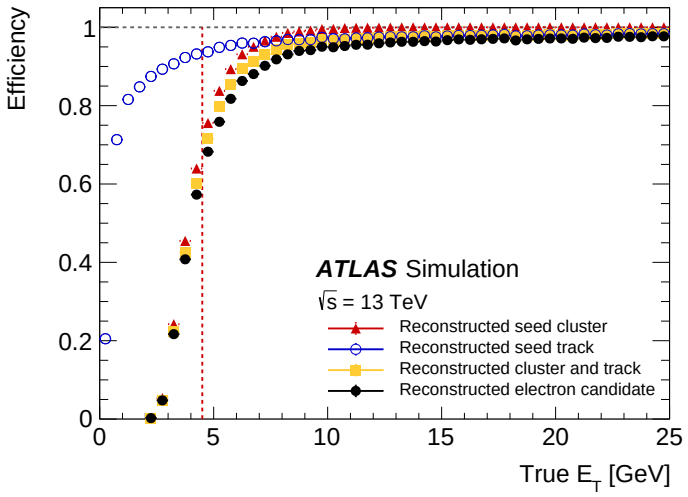


Figure 3.2: *The total reconstruction efficiency for simulated electrons as a function of the true generated transverse energy E_T, for several reconstruction steps. Red markers denote seed cluster reconstruction efficiency. Since cluster reconstruction requires uncalibrated cluster seeds with $E_T > 2.5$ GeV , the total cluster reconstruction efficiency is below 65% at $E_T = 4.5$ GeV (dashed line). Seed cluster reconstruction efficiency ranges from 65% at $E_T = 4.5$ GeV, to 96% at $E_T = 7$ GeV and more than 99% for $E_T = 15$ GeV [108].*

3.1.2.2 Dynamical topological cell clustering

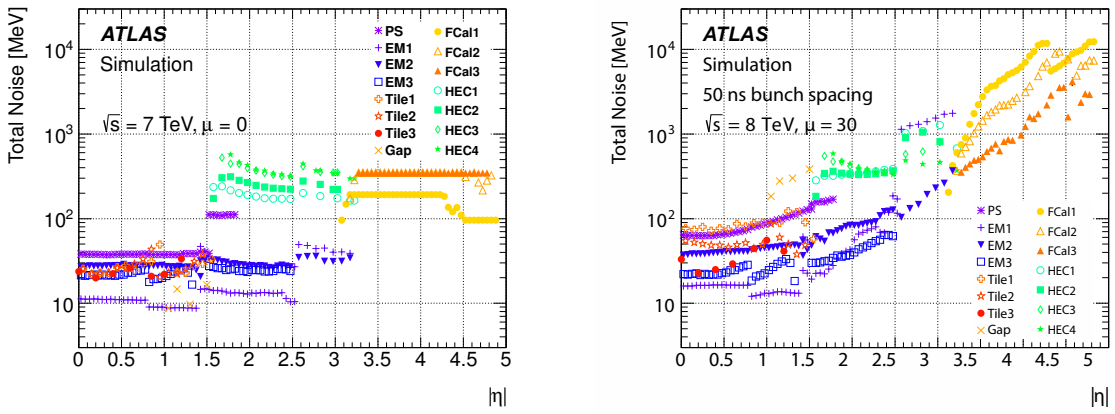
The development of multivariate calibration techniques [109] enables the calibration of dynamically-sized topological clusters (*topoclusters*) [107], hence the reconstruction of egamma objects using these techniques. Dynamical clustering algorithms offer multiple advantages compared to fixed-size ones.

The primary advantage of dynamical clustering is the ability to recover low energy photons produced by electrons interacting with matter upstream the calorimeters through bremsstrahlung processes. The collection of additional energy deposits from these processes together with the main deposit of the electron is called a *supercluster* and its construction is detailed in the following subsections [110].

Another motivation to use topoclusters compared to fixed-size algorithms is the fact that the growth of the cluster is similar to the development of the shower in the calorimeter. Topoclusters grow from the highest energy cell outwards, capturing the surrounding of the EM shower. They also collect more energy on average than sliding-windows, requiring smaller corrections and hence improving the energy resolution. The construction of topoclusters relies on a quantity denominated *cell significance* ζ_{cell}^{EM}, defined as:

$$\zeta_{cell}^{EM} = \left| \frac{E_{cell}^{EM}}{\sigma_{noise,cell}^{EM}} \right|, \tag{3.3}$$

where $|E_{cell}^{EM}|$ is the absolute energy at the scale of the EM calorimeter[2] and $\sigma_{noise,cell}^{EM}$ is the expected cell noise coming from both pileup σ_{noise}^{pileup} and the electronic chain $\sigma_{noise}^{electronic}$. The total expected cell noise rises with increasing pileup activity, by several orders of magnitude in some regions of the detector. The noise-per-cell values are extracted from minimum bias simulated samples as a function of $|\eta|$ for the expected average $\langle\mu\rangle$ prior to the recollection of data to allow for an online fast reconstruction. A validation is performed in a later step once data is recorded with events recorded with minimum bias triggers.



a) b)

Figure 3.3: *The total noise-per-cell in the ATLAS calorimeters at the EM scale for two different expected $\langle\mu\rangle$ conditions ($\mu = 0$ on the left and $\mu = 30$ on the right) as a function of $|\eta|$. The different colours denote the different layers of both LAr and hadronic calorimeters. These two maps were extracted for the topocluster reconstruction performed in Run 1 [111].*

The standard ATLAS topoclustering algorithm proceeds by first finding *proto-clusters*, cells with significant energy in the calorimeter, to which neighbouring cells are attached to form topoclusters. First, proto-clusters are seeded in the LAr calorimeter with cells from L2 and L3 having $\zeta_{cell}^{EM} > 4$. Presampler and L1 layers, with lower noise thresholds, are excluded from proto-cluster seeds collection in order to suppress noise proto-clusters. Then, proto-clusters start collecting iteratively neighbouring cells with $\zeta_{cell}^{EM} \geq 2$. A neighbouring cell is defined as the eight surrounding cells found in the same sampling, plus the cells in contact from other layers in the calorimeter. The addition of cells to proto-clusters is

[2] also referred to as EM scale.

done in order of decreasing energies to favour the growth of proto-clusters with larger energy cells. In case two or more proto-clusters share cells, the proto-clusters are merged and can be separated in a later step if more than one local maximum is found. Each new cell added to the proto-cluster is a new seed for the next iteration. When no more cells are found with $\zeta_{cell}^{EM} \geq 2$, a final set of neighbouring cells is added to the proto-cluster fulfilling $\zeta_{cell}^{EM} \geq 0$. This set of thresholds is known as "4-2-0" topocluster reconstruction, and a visual representation is shown in Figure 3.4.

The output of the topocluster algorithm is a list of topoclusters all over the detector with cells from both LAr and Tile calorimeters. A selection is performed on the built topoclusters in order to remove clusters from hadronic activity and keep the ones resulting from EM showers, referred to in the future as EM topoclusters. An initial cut is made on the EM fraction, f_{EM}, for each cluster:

$$f_{EM} = \frac{E_{L1} + E_{L2} + E_{L3} + w\,(E_{PS} + E_{E4})}{E_{clus}}, \quad w = \begin{cases} 1, & \text{if } 1.37 < |\,\eta\,| < 1.63 \\ 0, & \text{otherwise} \end{cases} \tag{3.4}$$

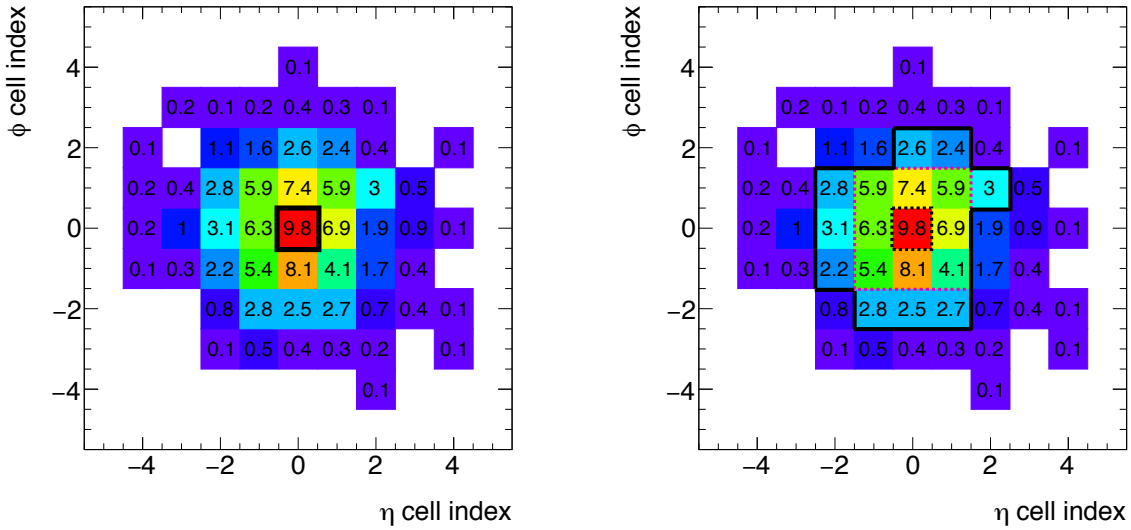
where E_{Li} is the cluster energy in the layer i. The last term $w\,(E_{PS} + E_{E4})$ is only considered if the topocluster is found in $|\eta| : [1.37, 1.63]$, the "crack region". The first term E_{PS} is the energy of the cluster in the presampler and E_{E4} is the energy of the cluster in "E4" scintillators. Topoclusters originated from hadronic contributions are naturally removed when requiring large enough values of f_{EM}. Thus topoclusters are required to have $f_{EM} > 0.5$ which does not limit the selection efficiency while rejecting 60% of pileup clusters [110]. The threshold in f_{EM} is determined using simulated electron samples with and without simulated pileup effects.

3.1.3 Track reconstruction

Electron and photon reconstruction also relies on the performance of the Inner Detector. Electrons, as charged particles, describe a trajectory which can be reconstructed using the hits in each of the subsystems in the ID. Unlike them, photons have no charge, but they can undergo pair production and convert into an electron-positron pair which may leave a signal in the ID. An illustration of an electron trajectory is shown in Figure 3.5 with pedagogical purposes.

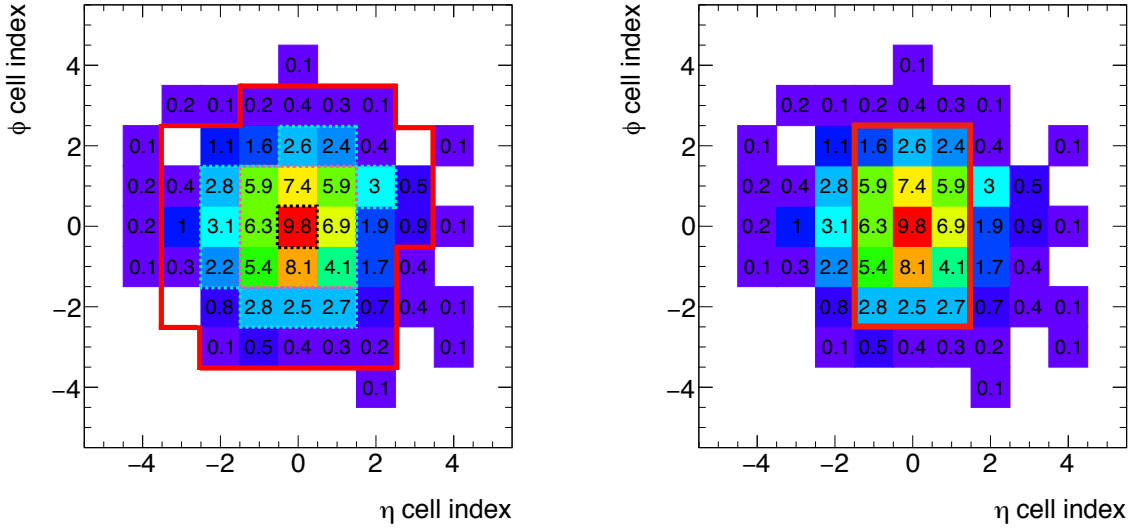
Track reconstruction relies on a standard pattern recognition algorithm [110], which reconstructs tracks in all the ID. First, three silicon hits in the SCT passing p_T and spatial requirements are used as input to a Kalman filter [112] and extrapolated outwards to the full tracker, followed first by an extrapolation to the pixel detector and finally to the TRT. The pion hypothesis is used by the Kalman filter, meaning that minimal energy loss is expected when traversing the material of the ID. Track candidates with $p_T > 400$ MeV are fit with a global χ^2 fitter [113] to determine the full trajectory of the charged particle. For electrons though, this procedure leads to inefficiencies since energy losses are significant due to bremsstrahlung processes compared to heavier charged particles. A Gaussian-sum Filter (GSF) [114] is a non-linear generalization of the Kalman Filter, that re-fits electron track candidates, accounting for significant losses when traversing the material of the ID.

A complementary back-tracking algorithm is implemented to reconstruct secondary particles, not coming from the interaction point. These tracks are reconstructed from seeds in the TRT and extrapolated inwards the ID to aid, among other particles, in the reconstruction of converted photons.



a) *First step: most significant cell marked as proto-cluster seed with* $\zeta_{cell}^{EM} > 4$

b) *Second step: adding neighbours with* $\zeta_{cell}^{EM} \geq 2$



c) *Final step: adding neighbours with* $\zeta_{cell}^{EM} \geq 0$

d) *Sliding-window cluster*

Figure 3.4: *Illustration of the topocluster construction. Histograms show the cell significance* ζ_{cell}^{EM} *for an EM shower in the middle layers of the LAr as a function of* $\eta \times \phi$*. In (a), the most significant cell is marked with a black box. In fact, every cell with* $\zeta_{cell}^{EM} > 4$ *is a proto-cluster seed, but since only one local maximum exists, only one topocluster is obtained at the end. In (b), all neighbouring cells with* $\zeta_{cell}^{EM} > 2$ *are attached to the proto-cluster. Previous iterations are shown with dashed lines. In (c), the final topocluster is shown in red lines to be compared with (d), which shows the cluster obtained with the sliding-window algorithm. A larger energy is usually reconstructed with the topocluster algorithm.*

3.1.4 Matching tracks and clusters

At this point in the reconstruction chain, the raw detector information is composed of tracks and topoclusters. In the next step, reconstructed tracks, output of the χ^2 Global

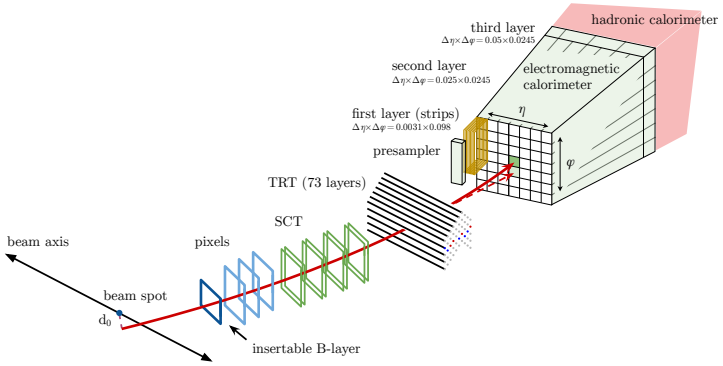


Figure 3.5: *A schematic illustration of the path of an electron through the detector. The red solid trajectory shows a hypothetical path of an electron. First, it traverses the inner detector (pixel detectors, then the SCT and lastly the TRT) and then enters the electromagnetic calorimeter. The dashed red trajectory indicates the path of a photon produced by the interaction of the electron with the material in the tracking system [108].*

fits and the GSF, are matched to EM topoclusters in order to classify each object as an electron, unconverted photon or converted photon.

First, tracks falling within an angular distance of $\Delta R = 0.3$ around the L2 barycentre of an EM topocluster, are considered. Tracks are loosely matched to topoclusters in the η, ϕ plane by requiring them to have $|\eta_{track} - \eta_{cluster}| < 0.05$ and $-0.10 < q(\phi_{track} - \phi_{cluster}) < 0.05$, where q is the charge of the reconstructed track, η, ϕ_{track} are the extrapolation of the track in the second layer of the calorimeter and $\eta, \phi_{cluster}$ are the coordinates of the barycentre of the cluster in the same layer. The asymmetry in the $q(\phi_{track} - \phi_{cluster})$ interval arises from possible energy losses of the charged particle that may modify significantly its trajectory. Also, to improve track-cluster matching performances arising from bremsstrahlung processes, a rescaling of the momentum of the track to match the momentum of the cluster candidate is done. If multiple tracks are associated to a single topocluster, tracks with hits only in the pixel detector are preferred before tracks with hits in the SCT but not in the pixel detector. Within these two possibilities, tracks are ordered according to their angular distance ΔR with respect to the cluster.

Loosely matched tracks are also used as input for the reconstruction of conversion vertices. Conversion vertices can be reconstructed with either one or two tracks. Also, two types of tracks are used to reconstruct conversion vertices: tracks with silicon hits ("Si tracks") and tracks reconstructed only in the TRT ("TRT tracks").

Double-track conversion vertices are reconstructed from a pair of oppositely charged tracks with an invariant mass compatible with 0, since photons are massless. Single-track conversion vertices are usually reconstructed from tracks without hits in the innermost layers of the ID, since these are mostly due to late conversions. The requirements to build conversions vary depending on the type of tracks used to build them, being loose for Si tracks compared to TRT tracks in the case of double-track conversions, and tighter for single-track conversions. The fraction of converted photons as a function of η is shown in Figure 3.6 for the different reconstructed track cases.

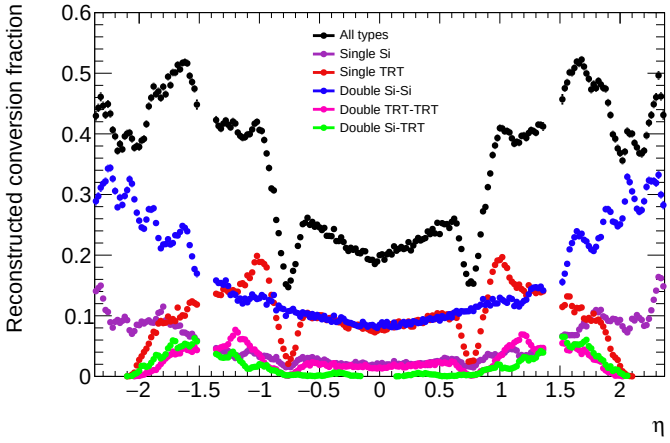


Figure 3.6: *The fraction of reconstructed converted photons as a function of η, extracted from a simulated samples with photons without pileup. The amount of material photons pass through increases with η, so does the conversion fraction. The TRT acceptance covers up to $|\eta| = 2.0$, which explains the abrupt decrease in reconstructed vertices using TRT tracks.*

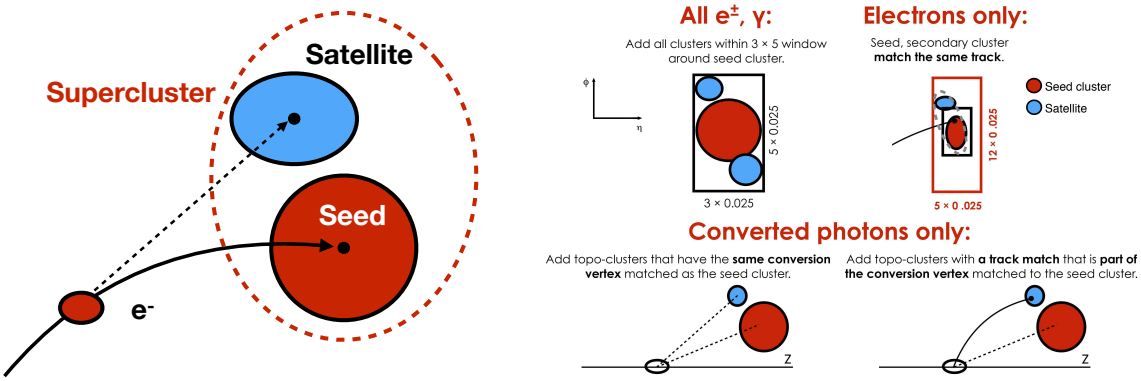
3.1.5 Superclusters

As mentioned previously, topological clusters are powerful tools that allow to capture within the same object (supercluster) the radiated energy deposits from electrons. Figure 3.7a illustrates the "supercluster concept", which essentially consists in looking for extra topoclusters in a nearby region and attach them to a higher energy topocluster. The supercluster construction runs simultaneously for both electron and photon hypotheses using all the information detected with the trackers and EM calorimeter.

The first step is to consider the EM topoclusters in descending order of E_T and to evaluate each of them as a possible supercluster seed. Electron supercluster seeds are required to have $E_T > 1$ GeV and a matched track with at least 4 silicon hits. Photon supercluster seeds are required to have $E_T > 1.5$ GeV, since no requirement is made on matched tracks or conversion vertices and lower energy thresholds would be much too sensitive to noise.

Once all the supercluster seeds have been identified, a search for satellite clusters starts. Though the only requirement for a cluster to be a satellite cluster is to have a smaller E_T than the seed cluster, an already identified satellite cluster cannot be used as seed cluster. In both electron and photon cases, a cluster is considered to be a satellite if it falls within a window of $\Delta\eta \times \Delta\phi = 0.075 \times 0.125$ around the seed cluster barycentre. For electrons, clusters within a larger window ($\Delta\eta \times \Delta\phi = 0.125 \times 0.300$) with the same best-matched track of the seed clusters are also attached as satellite clusters. For photons with conversion vertices made only with Si tracks, a cluster is added if the best-matched track also belongs to the conversion vertex of the seed cluster. The different cases are illustrated in Figure 3.7b.

The final step in the superclustering algorithm is to assign which cells are considered to belong to the supercluster among the seed plus satellite cluster cells. When considering the addition of different topoclusters, there may be cases in which two or more topoclusters share cells. The total amount of energy in such cells E_{cell} is only considered once in the supercluster reconstruction. In order to obtain a more linear response with the current



a) b)

Figure 3.7: *(a) Diagram of an example supercluster showing a seed electron cluster and a satellite photon cluster from bremsstrahlung processes. The dashed red line has no meaning regarding the construction of the supercluster, explained in the text. (b) Diagram of the superclustering algorithm for electrons and photons where red ellipses denote seed topoclusters and blue ellipses denote satellite topoclusters. Larger size ellipses denote larger energy deposits. [110]*

calibration procedure, the number of cells is restricted in $|\eta|$. Only three (five) cells in $|\eta|$ with respect to the supercluster barycentre are finally used while no restrictions are applied in the ϕ direction.

The quantity studied to compare performances between both algorithms (fixed-size windows and superclusters) is the energy response E_{raw}/E_{true}, where E_{raw} is the raw, uncalibrated energy of the clustering algorithm; and E_{true} is the true simulated energy of the object. A ratio equal to 1 would imply a perfect energy reconstruction. In particular, the *effective interquartile range* (IQE) of the energy response E_{raw}/E_{true} is computed:

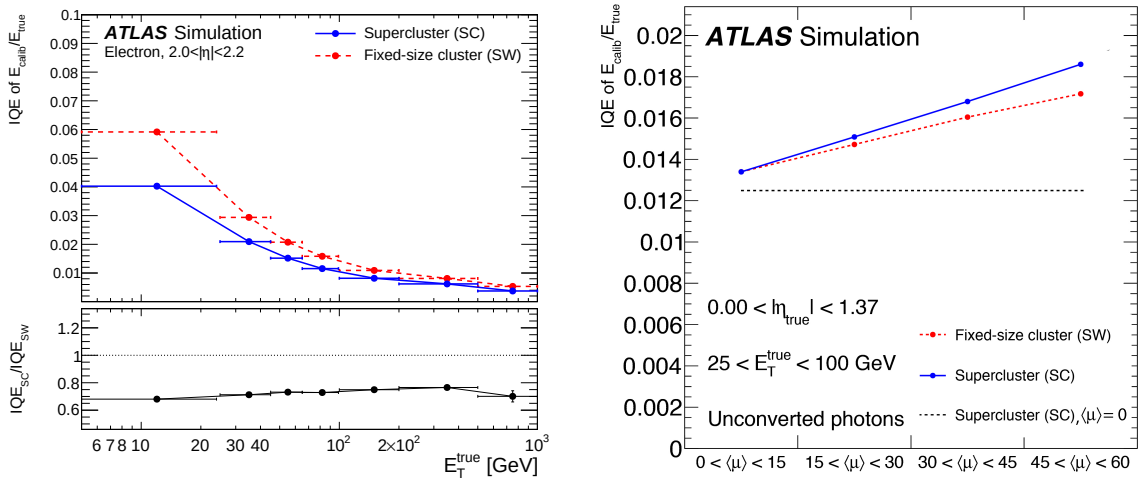
$$\text{IQE} = \frac{Q_3 - Q_1}{2\Phi^{-1}(0.75)}, \tag{3.5}$$

where $Q_{1,3}$ are the first and third quartiles respectively and the denominator $2\Phi^{-1}(0.75)$[3] is chosen such that in case of Gaussian distributions the IQE would be equal to its standard deviation. A comparison of the performances of the sliding-window algorithm and the superclustering algorithm is shown for electrons and photons as a function of E_T^{true} in Figure 3.8a and, as a function of the average number of pileup interactions, $\langle\mu\rangle$, in Figure 3.8b.

3.1.6 Photon and electron candidates

The next step is to create *analysis-level* electrons and photons. Since both hypotheses are followed when building superclusters, the same reconstructed object can be identified as both an electron and photon. An unconverted photon candidate is reconstructed if no cluster in the supercluster has a matched track. An electron candidate is reconstructed if the seed cluster has a good track with no conversion vertex. Superclusters matched to a conversion vertex are considered converted photon candidates. If no clear decision can be made, a series of decisions based on the quality and energy of the tracks, described in

[3] Φ is the cumulant of the Standard Normal distribution.



a) b)

Figure 3.8: *Comparison between fixed-size and superclustering algorithms. The fixed-size algorithm (SW) is shown in red and the superclustering algorithm (SC) is shown in blue. (a) shows for electrons the performance of both algorithms as a function of the true generated transverse energy E_T. A significant improvement is observed at lower energies, since the lost contribution of radiated photons in the SW algorithm is captured in the SC. (b) shows for unconverted photons the performance of both algorithms as a function of the average number of pileup interactions $\langle\mu\rangle$. The dashed black line denotes the SC algorithm IQE for photons in simulation without pileup. Increasing pileup worsens the SC performance, since nearby energy deposits make the energy response wider compared to SW algorithms.*

detail in [110], are followed in order to decide whether the supercluster is a photon, an electron or an ambiguous object, for which both electron and photon objects are kept for later use at analysis level.

3.2 Photon energy calibration

Similar to the material upstream the calorimeters, detector imperfections and inhomogeneities modify the development of the EM shower and impact the reconstructed energy, worsening its resolution or biasing the measurement. The energy measured in the cells of a cluster needs to be calibrated in order to produce the best possible estimate of the true energy deposed by the incident particle. The energy response calibration procedure consists on a series of steps, involving both data and simulation samples, that aims to recover the energy of the incident particle while achieving the best energy resolution possible. The steps are sketched in Figure 3.9 and briefly described in the following.

3.2.1 A first estimate of the energy from simulation: multivariate algorithms

The energy of electrons and photons is first estimated from the reconstructed EM cluster and corrected for energy losses upstream the calorimeters, energy deposited beyond

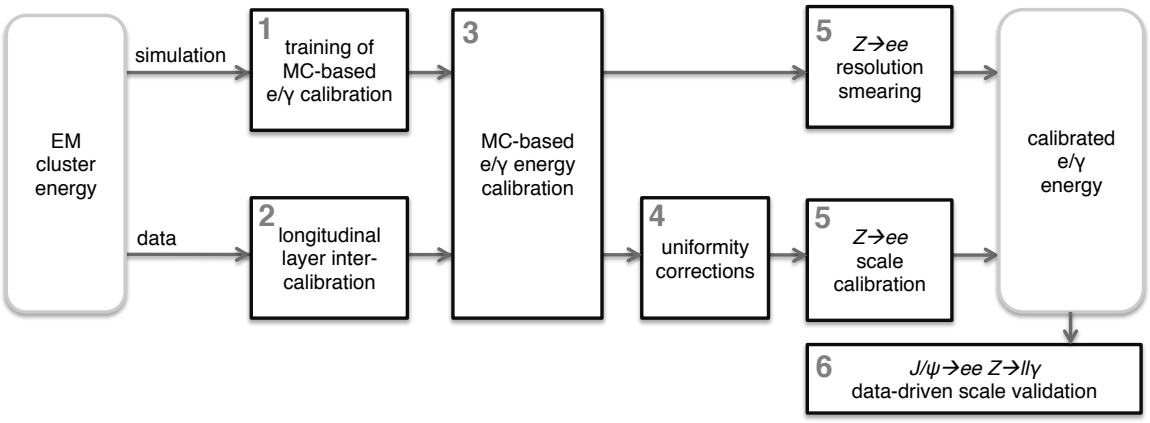


Figure 3.9: *Schematic overview of the procedure used to calibrate the energy response of electrons and photons in ATLAS [109].*

the LAr and also neighbouring cells outside the cluster definition. These effects are accounted for by using a multivariate regression algorithm (MVA) tuned on simulated samples (electrons, unconverted and converted photons) without pileup. The training of the algorithm is performed with Boosted Decision Trees (BDTs) using a set of variables that describe the development of the shower in the calorimeter, such as the total energy measured in the cells or ratios of energy between the different layers. The full list of variables is detailed in [109]. For unconverted photons without any present pileup, the resolution follows the expected sampling term of the calorimeter ($\sim 10\%\sqrt{E/\text{ GeV}}$ in the barrel and $\sim 15\%\sqrt{E/\text{ GeV}}$ in the endcaps). For converted photons some degradation is observed at low energies, since additional interactions occur upstream the calorimeters. The same MVA is applied on both data and simulation in order to obtain a first calibrated energy.

3.2.2 Corrections on data: uniformity corrections and longitudinal intercalibration

Residual differences from features not present or not properly described by the simulation are accounted for before determining the absolute energy scale in data.

Uniformity corrections are first applied to ensure an homogeneous energy reconstruction across the detector. Common sources of inhomogeneities are high-voltage (HV) short-circuits, azimuthal non-uniformities and gain differences.

In a small number of sectors of the LAr calorimeter, the applied HV is set to a non-nominal value to avoid short circuits occurring in specific LAr gaps. This feature is not present in the simulation and requires per-cell specific calibration of typical values from 1% to 7% [109].

The EM LAr barrel is divided in 16 modules across ϕ, producing wider gaps in between two contiguous modules. Since the gap is wider, a weaker electric field is present reducing the energy response. This effect is not present in the simulation and a dedicated correction is applied to data.

As mentioned in Section 3.1.1, the output signal collected by cells in the calorimeter can be readout with three different gains. The linearity of each of these modes is measured to be better than 0.1% in the usual energy ranges in data. A relative mis-calibration between different gains may modify the relation between the energy response and the

energy of the particle, with differences found to be typically 0.1%, increasing up to 1% at higher energies [105].

On top of the uniformity corrections, special attention is given to the relative energy scales of each of the layers of the calorimeter, since it is longitudinally segmented. A layer intercalibration is performed to ensure an equalized energy response between the different samplings of the calorimeter. Only the presampler, first and second samplings of the accordion calorimeter are intercalibrated, since only a small fraction of the energy is deposited in the third sampling, except for egamma objects with energies below several hundreds of GeV.

The intercalibration is first performed between the first two samplings of the LAr by measuring the ratio of the energies deposited by muons in both data and simulation. Muons from Z decays are used because they are almost insensitive to the amount of material in front of the calorimeter, since they are minimum ionizing particles in this energy range, below ~ 100GeV. The Most Probable Value (MPV) of the energy deposited is extracted for the first and second samplings to define the quantity $< E_{1/2} >= MPV_1/MPV_2$. Finally, the intercalibration result is expressed in terms of $\alpha_{1/2}$, defined as the ratio of $< E_{1/2} >$ in data and simulation. This calibration is applied as a function of $|\eta|$ on the energy of the second sampling $L2$ measured in data, with a resulting uncertainty between 1% and 1.5%.

The presampler energy scale α_{PS} is defined as the ratio of measured energies in data over simulation using electrons from W and Z decays. Measurements of the electron energy are affected from material mismodelling upstream the presampler, causing earlier shower developments and thus larger energy deposits in the PS. The correlation of the energy deposited in the PS and the first two layers of the calorimeters is exploited to correct this effect. The measured energy scale α_{PS} is a correction applied on data with uncertainties ranging from 2 to 3% depending on $|\eta|$.

3.2.3 Energy scale and resolution with $Z \to ee$ decays

Once all previously listed corrections are applied, a remaining discrepancy between data and simulation is observed. The difference is encapsulated in coefficients α and c, used to correct the data so that it agrees with the simulation. The coefficients are measured using electrons from $Z \to ee$ decays to benefit from very large statistics and a clean signature in the detector. Small corrections are applied in a later step to account for differences in the reconstruction that could affect differently electrons and photons.

The energy scale factors α and resolution scale factors c are defined as:

$$E^{data} = E^{MC}(1+\alpha_i), \quad \left(\frac{\sigma_E}{E}\right)^{data} = \left(\frac{\sigma_E}{E}\right)^{MC} \oplus c_i, \tag{3.6}$$

where E and σ_E are the electron energy and its resolution measured in both data and simulation, α_i is the energy scale factor in pseudorapidity $|\eta|$ bins, and c_i is a constant term in the energy resolution in pseudorapidity $|\eta|$ bins. Given the existing correlation between the energy scale and the energy resolution, both parameters are simultaneously extracted from fits to the dielectron invariant mass distribution.

First, the dielectron invariant mass m_{ee}^{data} measured in data is expressed as a function of the individual electron energy scales $\alpha_{i(j)}$ and the measured mass in simulation m_{ee}^{MC}:

$$m_{ij}^{data} = m_{ij}^{MC}\sqrt{(1+\alpha_i)(1+\alpha_j)} \approx m_{ij}^{MC}(1+\alpha_{ij}), \text{ with } \quad \alpha_{ij} = \frac{\alpha_i+\alpha_j}{2}, \tag{3.7}$$

where the indices i, j denote the pseudorapidity bin (η_i, η_j) of each electron and second order terms in α are neglected. [4]

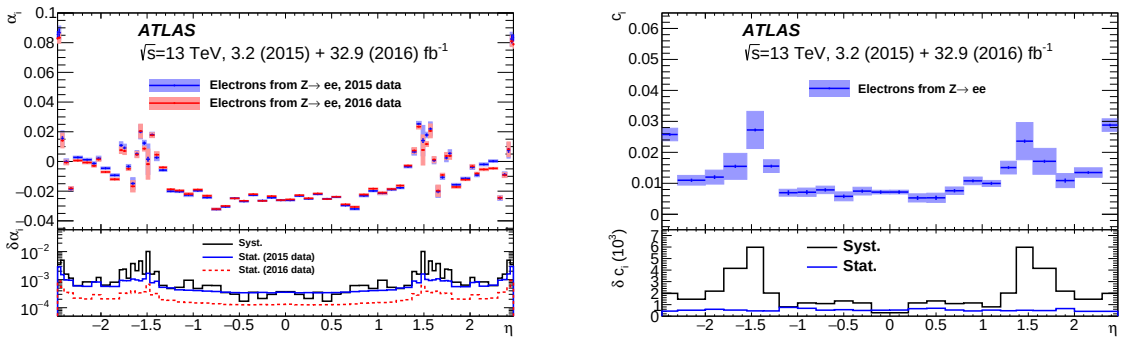
The discrepancy in simulation and data between the energy resolution in the dielectron mass distribution is characterized by an additional constant term c_{ij} added in quadrature to the resolution measured in the simulation:

$$\left(\frac{\sigma_m}{m}\right)_{ij}^{data} = \frac{1}{2}\left[\left(\frac{\sigma_E}{E}\right)_i^{data} \oplus \left(\frac{\sigma_E}{E}\right)_j^{data}\right] = \left(\frac{\sigma_m}{m}\right)_{ij}^{MC} \oplus c_{ij}, \text{ with } \quad c_{ij} = \frac{c_i \oplus c_j}{2}, \quad (3.8)$$

where c_{ij} are the relative invariant mass resolution scale factors for electrons in the i, j pseudorapidity bins (η_i, η_j).

To perform a simultaneous subtraction of both energy scale and energy resolution scale, a grid of templates for each (η_i, η_j) bin is built in the (α_{ij}, c_{ij}) space. The optimal values for α_{ij} and c_{ij} are obtained through a χ^2 minimization, from which the individual electron energy scale and resolution factors are computed using Equations 3.7 and 3.8.

The final result obtained for α_i and c_i is shown in Figure 3.10. The energy scale factor α_i ranges across η from -3% up to 2% while the energy resolution constant term c_i is roughly stable in the barrel, typically within less than 1%, and increases towards the endcaps, ranging between 1% and 2%.



a) b)

Figure 3.10: *Results of the data to MC calibration from $Z \to ee$ events. (a) Energy scale corrections α_i as a function of η. (b) Energy resolution constant term c_i as a function of η. The systematic and statistical uncertainties are shown separately in the bottom panels. [115]*

The energy scale factors extracted from electrons are expected to be valid also for photons despite the differences in shower development, since the EM calorimeter response is effectively corrected. However, a cross-check is performed to verify that the effects in the energy response caused by electron-photon differences are under control. Events with $Z \to ll$ boson decays (so called "radiative Z events") are used to check the photon energy scale, in particular in the low energy range. The residual difference between the electron energy scale and the one extracted from radiative Z events is parametrized and included as an additional correction only to photons. The residual difference is consistent with zero within uncertainties.

[4] To obtain this expression the impact of the resolution on the angular separation between the two electrons is neglected, since its resolution is significantly better compared to that of the energies of the electrons.

3.2.4 Summary of uncertainties and calibration results

Besides the uncertainties obtained in the template fit performed to the dielectron invariant mass, further effects are considered for the total uncertainty on both photon energy scale and resolution.

Systematic uncertainties on the photon energy scale are described by a set of 64 independent uncertainty variations, some of which consist on the same variation for different $|\eta|$ bins. The largest contribution to the total uncertainty for photons below 100 GeV arises from the intercalibration between the first and second layer, followed by the uncertainty associated to the cell energy non-linearity. Nevertheless, all these effects add up to a small energy scale uncertainty spanning from 0.2% to 0.3% in the barrel and from 0.45% to 0.8% in the endcap [115].

Systematic uncertainties on the photon energy resolution arise from various sources: shower and sampling fluctuations, electronics and pileup noise, and residual non-uniformities in the energy measurements in data. All these lead to the result presented in Figure 3.11, which shows the energy resolution σ_E/E as a function of the transverse energy of unconverted photons in the barrel and in the endcap. The energy resolution improves with increasing transverse energy of the photon candidate from 5% at 10 GeV down to less than 1% beyond 100 GeV. For unconverted photons with transverse energies from 20 to 60 GeV, which is the dominant contribution in the analysis presented in this thesis (Section 6), the energy resolution spans from 3% to 1.5% at higher energies and is known with a precision better than 10% [115].

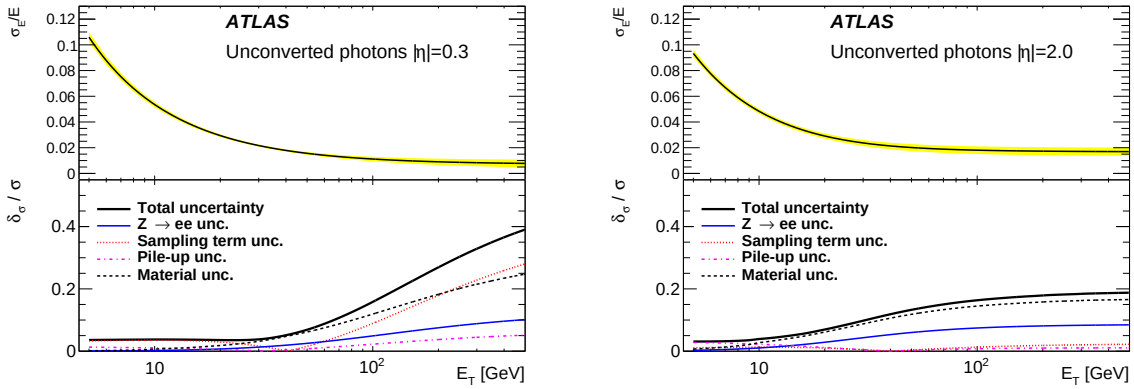


Figure 3.11: *The energy resolution, σ_E/E, as a function of E_T for unconverted photons at $|\eta| = 0.3$ (left) and $|\eta| = 2.0$(right). The yellow band shows the total uncertainty in the resolution. The breakdown of the relative uncertainty, δ_σ/σ is shown in the bottom panel. [115]*

3.3 Photon identification

In physics analyses an overwhelming amount of photons product of hadronic decays is present together with the prompt photons from of the hard scatter. A fake photon in this context is an object different from a prompt photon which is reconstructed as a photon candidate. Photons and electrons produce narrow electromagnetic showers in the calorimeter that are contained almost completely in the LAr calorimeter, while the

majority of fakes are broader and also leave a significant amount of energy in the hadronic calorimeter.

The identification of prompt photons over fake photons is performed by exploiting the characteristic shape of electromagnetic showers in the calorimeter. The description of the shape is based on a set of variables called *shower shape* variables, that characterize the lateral and longitudinal development of the EM shower. The identification of photons is performed through a set of cuts on the shower shape variables to select candidates whose shape is consistent with that expected from an EM shower. The definitions of the photon shower shape variables used in photon identification are shown in Table 3.2 and their distributions for signal and fakes are illustrated in Figure 3.12.

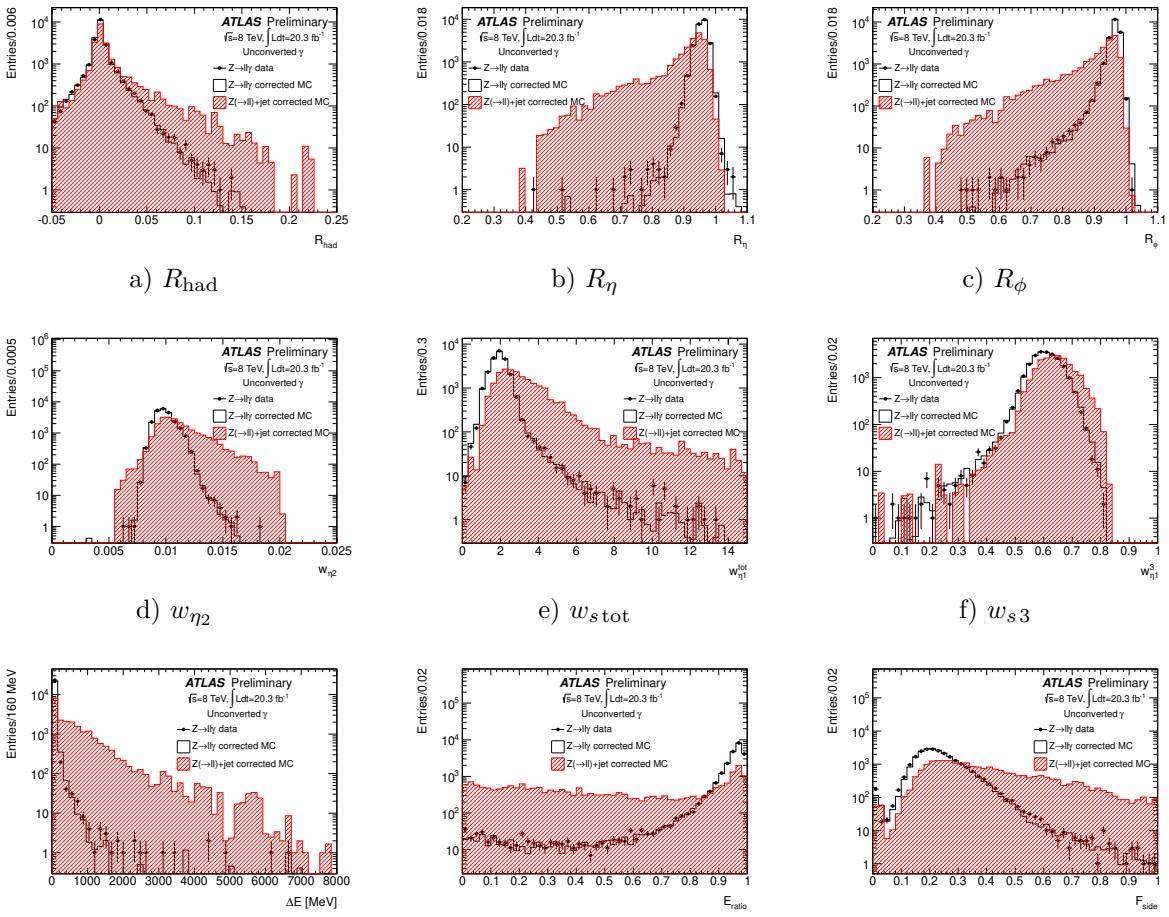


Figure 3.12: *Distribution of the shower shape variables for unconverted photon candidates with $E_T^\gamma > 20$ GeVselected from $Z \to ll\gamma$ events [116].*

Identification selections, referred to in the following as working points, can be built to provide different purities and efficiencies. Three working points are defined by applying cuts on subsets of the shower shape variables: *loose*, *medium* and *tight* identification working points.

The *loose* working point is typically used at trigger level and provides an initial background rejection. This working point bases its selection on the second layer shower shape variables and the energy leaking in the hadronic calorimeter, since photons are not

Table 3.2: *The set of discriminating variables used for photon identification. The last three columns indicate which working point applies cuts on that variable being L for loose, M for medium and T for tight. For variables calculated in the first EM layer, if the cluster has more than one cell in the ϕ direction at a given η, the two cells closest in ϕ to the cluster barycentre are merged and the definitions below are given in terms of this merged cell. Variables for which no working points are indicated, are used in electron identification and are included here for completeness.*

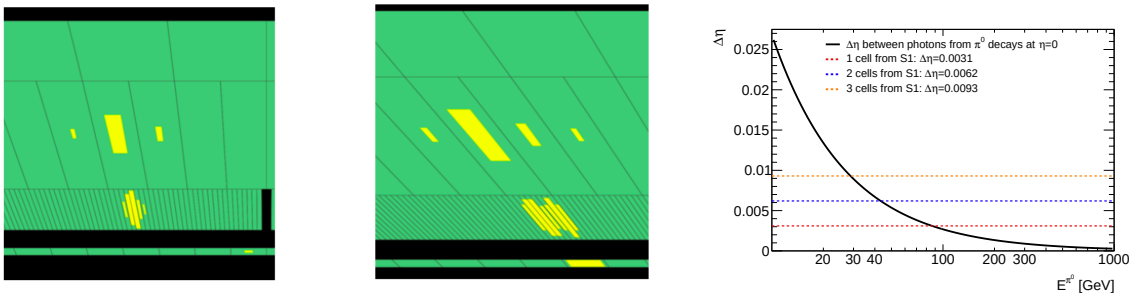
Category	Description	Name	Loose	Medium	Tight
Hadronic leakage	Ratio of E_{T} in the first layer of the hadronic calorimeter to E_{T} of the EM cluster (used over the ranges $\lvert\eta\rvert < 0.8$ and $\lvert\eta\rvert > 1.37$)	R_{had_1}	✓	✓	✓
	Ratio of E_{T} in the hadronic calorimeter to E_{T} of the EM cluster (used over the range $0.8 < \lvert\eta\rvert < 1.37$)	R_{had}	✓	✓	✓
EM third layer	Ratio of the energy in the third layer to the total energy in the EM calorimeter	f_3	-	-	-
EM second layer	Ratio of the sum of the energies of the cells contained in a 3×7 $\eta\times\phi$ rectangle (measured in cell units) to the sum of the cell energies in a 7×7 rectangle, both centred around the most energetic cell	R_η	✓	✓	✓
	Lateral shower width, $\sqrt{(\Sigma E_i\eta_i^2)/(\Sigma E_i) - ((\Sigma E_i\eta_i)/(\Sigma E_i))^2}$, where E_i is the energy and η_i is the pseudorapidity of cell i and the sum is calculated within a window of 3×5 cells	w_{η_2}	✓	✓	✓
	Ratio of the sum of the energies of the cells contained in a 3×3 $\eta\times\phi$ rectangle (measured in cell units) to the sum of the cell energies in a 3×7 rectangle, both centred around the most energetic cell	R_ϕ	-	-	✓
EM first layer	Total lateral shower width, $\sqrt{(\Sigma E_i(i-i_{\mathrm{max}})^2)/(\Sigma E_i)}$, where i runs over all cells in a window of $\Delta\eta\approx0.0625$ and i_{max} is the index of the highest-energy cell	$w_{s\,\mathrm{tot}}$	-	-	✓
	Lateral shower width, $\sqrt{(\Sigma E_i(i-i_{\mathrm{max}})^2)/(\Sigma E_i)}$, where i runs over all cells in a window of 3 cells around the highest-energy cell	$w_{s\,3}$	-	-	✓
	Energy fraction outside core of three central cells, within seven cells	f_{side}	-	-	✓
	Difference between the energy of the cell associated with the second maximum, and the energy reconstructed in the cell with the smallest value found between the first and second maxima	ΔE_s	-	-	✓
	Ratio of the energy difference between the maximum energy deposit and the energy deposit in a secondary maximum in the cluster to the sum of these energies	E_{ratio}	-	✓	✓
	Ratio of the energy measured in the first layer of the electromagnetic calorimeter to the total energy of the EM cluster	f_1	-	-	✓

expected to deposit energy that deep. The same selection is used for both converted and unconverted photons. It provides a photon efficiency higher than 99% for $E_T^\gamma > 40$ GeV.

The *medium* working point is based on the loose working point but including a selection on the shower shape variable E_{ratio}. This selection requires photons to have, qualitatively, a clear maximum in the energy deposits of the first layer of the calorimeter, in comparison to fakes, from which two maxima are expected for example from neutral pion decays.

It is typically used at trigger level and was defined to cope with the increasing pile-up conditions after 2017.

The *tight* working point includes all the shower shape variables used in the medium working point, and includes the so-called *strip variables*, shower shape variables based on the information provided by the first sampling. The fine granularity of the strip layer is exploited to discriminate prompt photons from jets inside which light hadrons (like π^0s) may decay into collimated pairs of photons. The latter can be discriminated from prompt photons by building shower shape variables sensitive to differences produced by the two photons from the decay compared to the signal left by the prompt photon, as it is illustrated in Figure 3.13.



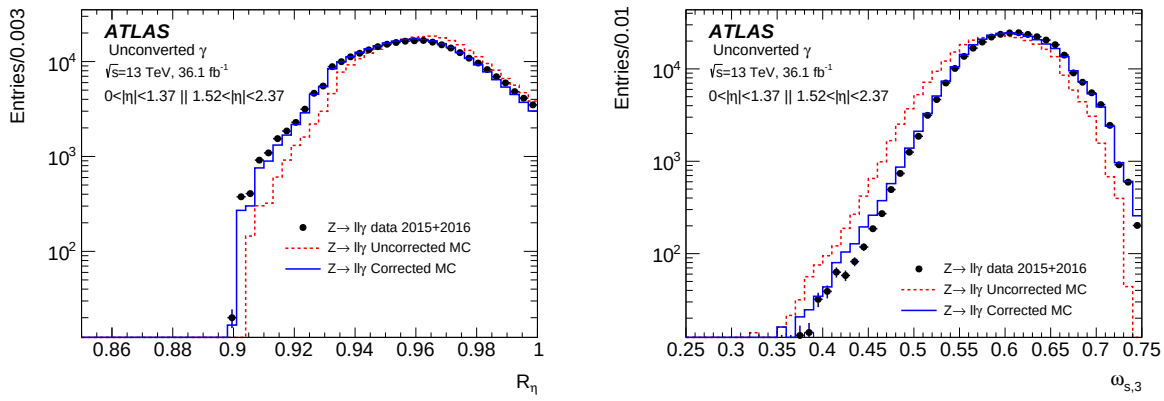
a) b) c)

Figure 3.13: *(a) and (b) Zoom on the first and second layer of the EM calorimeter in two different event displays. The picture on the left shows the energy deposited by a prompt photon while the picture on the right shows (possibly) a π^0 decaying to two photons, and hence leaving two clear deposits in the first layer.*
(c) Minimum separation in $\Delta\eta$ for two photons from a neutral pion decay at $\eta = 0$, as a function of the energy of the pion. The dashed lines in red, blue and orange denote the size in $\Delta\eta$ for 1,2 and 3 cells respectively in the first layer. Photons product from π^0 with energies larger than values below one of the dashed lines cannot be resolved with that number of cells.

The values of the cuts of the tight working point are optimized with a multivariate algorithm for unconverted and converted photons using the variables detailed in Table 3.2. Since shower shape distributions are $|\eta|$ dependent, different sets of cuts are obtained in separate $|\eta|$ bins for the three working points. From 2017 onwards, the tight working point is optimized also in E_T^γ bins, improving identification efficiencies for low-energy photons by 20% [117].

Prior to efficiency measurements, photon shower shapes are corrected due to a systematical mismodelling in the ATLAS MC simulation. While the description of the longitudinal component of the electromagnetic shower seems to be properly described by the simulation, the transversal component description is imperfect. This mismodelling reflects in differences in the shower shape distributions when comparing data and simulation. This

difference is corrected by applying shifts to the lateral shower shape variables obtained by minimizing the χ^2 between simulation and data distributions. Some examples are shown in Figure 3.14 for shower shapes obtained with radiative $Z \to ll\gamma$ decays.



a) b)

Figure 3.14: *Two examples of distributions of the shower shape variables R_η and w_{s3} for unconverted photons candidates with $E_T^\gamma \in [10, 50]$ GeV and $|\eta| < 2.37$ from $Z \to ll\gamma$ data and simulation. The uncorrected (dashed red line) and corrected (solid blue line) distributions from simulation are also shown [118].*

Photon identification efficiencies are measured in data using three different samples. The first one covers the lowest part of the photon energy spectrum using radiative Z events. The second sample covers a broader part of the spectrum and uses an inclusive-photon sample while the third sample uses electrons from $Z \to ee$ decays, with a method that transforms the electron shower shapes to resemble photon shower shapes [118]. The results from the three different methods are then combined to provide the best possible efficiency determination in data, which is compared in a later step with efficiencies obtained from simulation. The observed differences are used as a correction factor in the simulation called scale factor, consisting on the ratio of the measured efficiencies in data and simulation. The measured efficiencies in data and their difference with respect to the simulation are presented in Figure 3.15.

While tight identification rejects the majority of reconstructed fake photons, a remaining fraction of them still passes the identification selection, since the granularity of the first layer is not enough to reject the heavily boosted neutral mesons decays (see Figure 3.13c). An additional selection is performed based on the hadronic energy flow present around the photon candidates to improve the purity of the selected photons. This procedure is called photon isolation and it is described in detail in the next chapter.

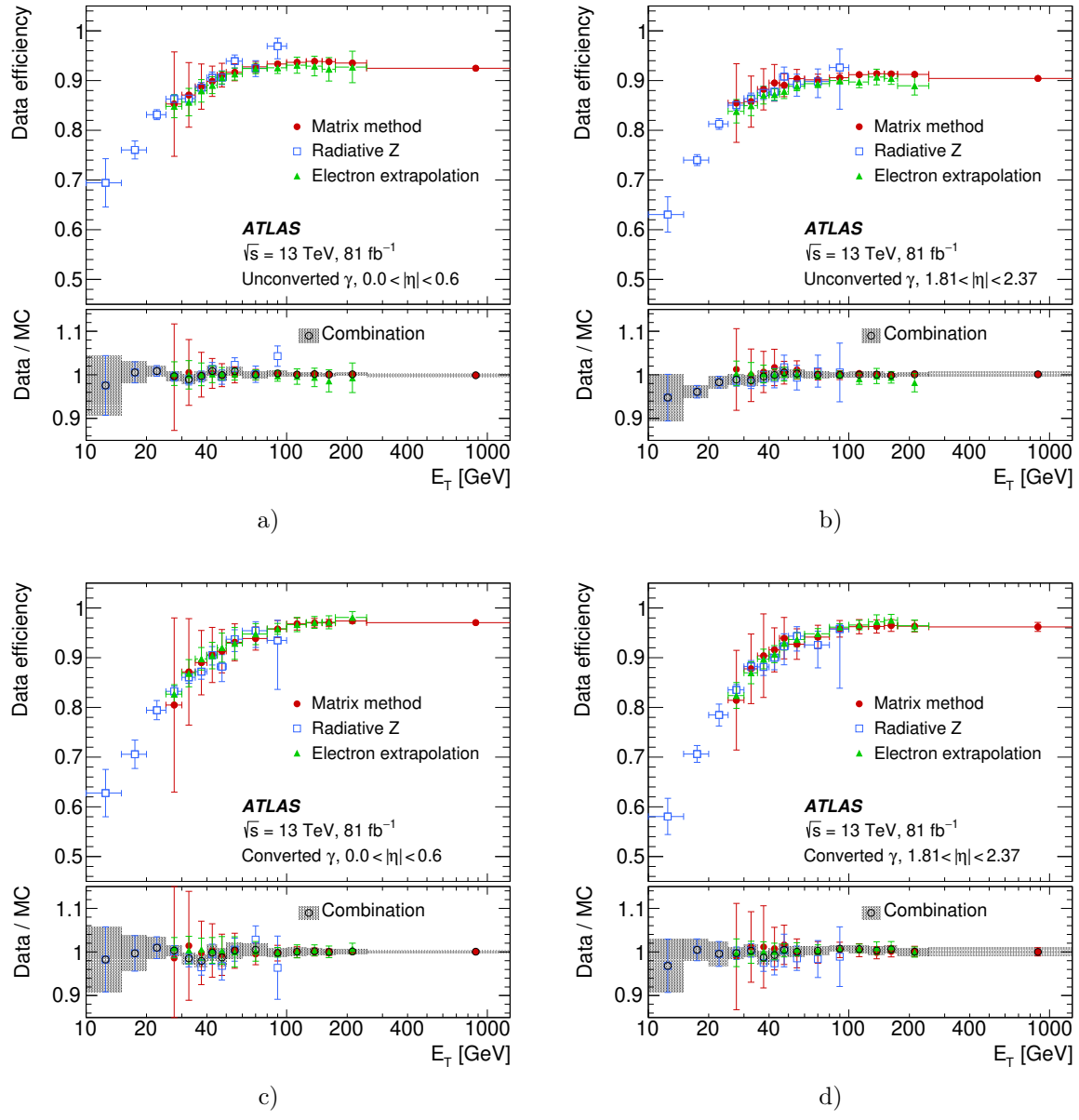


Figure 3.15: *The photon identification efficiency, and the ratio of data to simulation efficiencies, for unconverted and converted photons as a function of E_T^γ in two different $|\eta^\gamma|$ regions. The combined scale factor, obtained using a weighted average of scale factors from the individual measurements, is also presented; the band represents the total uncertainty [117].*

Chapter 4

Photon isolation

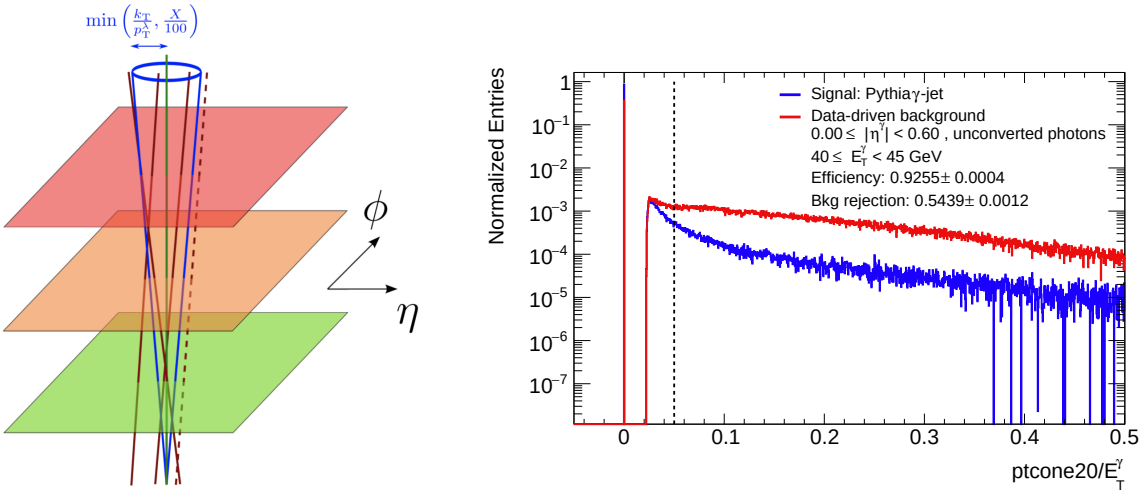
Isolation is a concept defined to classify objects regarding the amount of hadronic activity around them. Hadronic activity in this context refers to the energy flow of hadrons present around photon candidates. Prompt photons are expected to be isolated from nearby hadronic activity while fakes, being produced from decays of light mesons inside jets, are expected to be surrounded by certain level of hadronic activity. This information is exploited to reject photon candidates with large hadronic energy deposits around, which are likely fakes, enriching the photon candidate sample in prompt photons. As mentioned in Section 1.2.3, the experimental challenge is to construct a criterium which can be safely translated into a theoretical prescription.

This chapter provides a description of photon isolation performance in the ATLAS detector using the full Run 2 dataset, with emphasis on the measurements using information from the calorimeters. In the first section, the observables used to estimate the hadronic activity around photon candidates are defined. The next two sections detail the procedure followed to perform the isolation efficiency measurements and the results obtained. The final section discuss innovative correction techniques regarding possible data-simulation mismodellings. The discussion presented in this chapter has contributions from various ATLAS colleagues.

4.1 Isolation energy estimators

Information from tracking and calorimetric systems is used to determine whether a photon candidate is isolated or not. The track isolation energy is defined as the sum of transverse momenta, p_T, of tracks falling within a cone centered around the photon candidate. The size of the cone can be fixed at $R = 0.2$, or vary as a function of E_T^γ to improve efficiencies in boosted topologies. This selection is highly efficient for signal photons (higher than 90% at low E_T^γ up to almost 100% at large E_T^γ) and rejects roughly half of the background (see Figure 4.1). However, track isolation is insensitive to the neutral component of hadronic showers.

The calorimetric isolation energy is built from energy deposits in the calorimeter around the photon candidate. A sketch illustrating how the calorimetric isolation variable is computed is shown in Figure 4.2. This variable is called **topoetcone** since it is built from the transverse energy of topoclusters. It is first built from the "raw" calorimetric isolation energy, defined as the sum of the positive transverse energies of the topoclusters whose barycenter lie within a cone centered around the photon candidate, including the energy of the photon candidate. The size of the cone is fixed in this case since the development



a) b)

Figure 4.1: *(a) Schema of the track isolation variable with variable radius* `ptvarcone` *and analogously for fixed radius* `ptcone`*. All the good tracks (brown) allocated in a cone (blue) around the object are selected. The dotted lines show tracks which are too far from the object and therefore are not considered. The p_T of selected tracks is summed to calculate the isolation variable, and the core energy is subtracted (for muons, this corresponds to the transverse momentum of the object's track, in green). (b) Track isolation energy* `ptcone20` *over E_T^γ distribution for signal and background. The discontinuity arises from tracks being required to have $p_{T,track} > 1$ GeV.*

of EM showers in the calorimeter is almost invariant with the energy, besides a mild logarithmic dependence with E_T^γ. Calorimetric isolation is more difficult to deal with compared to track isolation, since the photon itself deposits energy in the calorimeter which needs to be subtracted on top of other contributions to the isolation energy, like pileup or electronic noise. The various corrections applied in order to improve the hadronic energy estimate are described in the following.

4.1.1 Photon energy subtraction

The photon shower lateral development in the calorimeter, as detailed in Section 2.2.3.1, can be described in terms of the Molière Radius. For the ATLAS EM calorimeter, the Molière Radius is approximately 5.0 cm [111] corresponding to ~ 1.3 cells at $\eta = 0$. This implies that 90% of the energy of the photon should be contained in a grid of 3×3 cells, and 95% in a grid of 5×7, missing a few % of the energy.

Two different methodologies were implemented during Run 2 to remove the energy of the photon included in the raw calorimetric isolation energy. They are referred to in this text as fixed-mask core subtraction and supercluster core subtraction.

The fixed-mask core subtraction removes the energy of the photon candidate in two steps: first, the energy collected by the cells in a fixed-mask of size 5×7 is subtracted. Then, an additional correction accounting for the energy leaking outside the mask is applied and usually amounts from 1.5% up to 5% of the photon energy. While the fixed-mask subtraction is computed on an event-by-event basis, the leakage correction is applied as an average as a function of E_T^γ in bins of $|\eta^\gamma|$ and conversion status, being different for unconverted and converted photons. The leakage correction is estimated

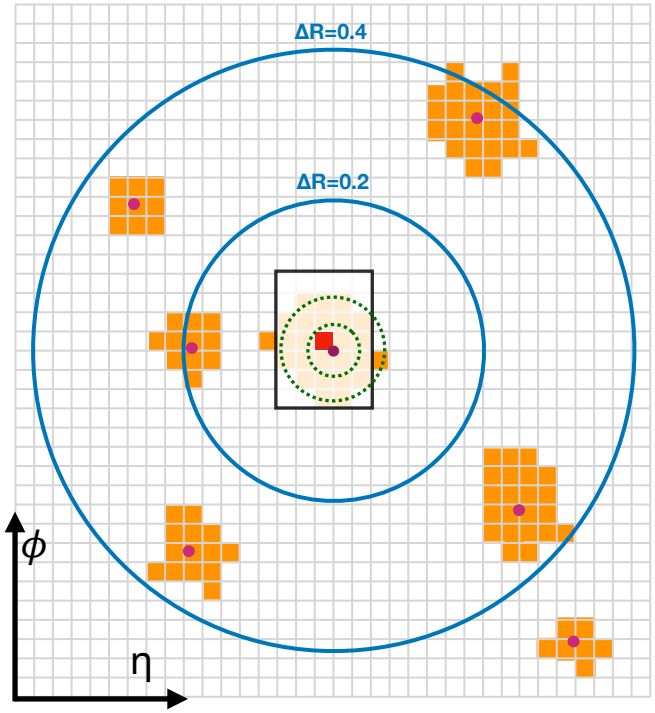
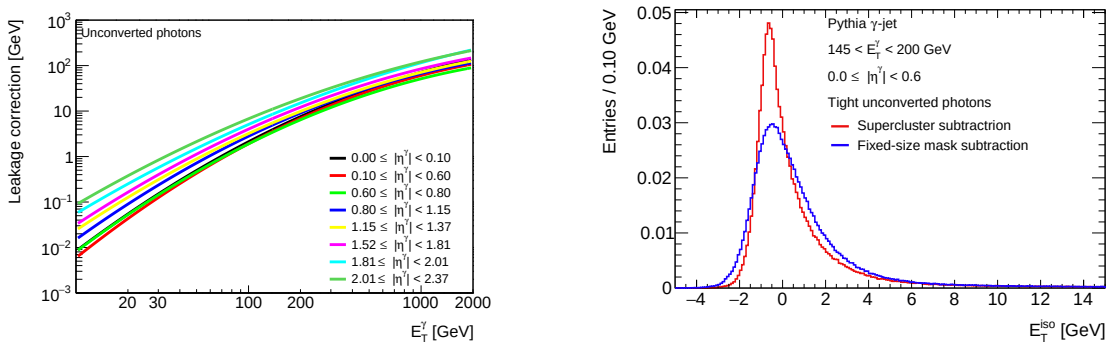


Figure 4.2: *Schema of the* ***topoetcone*** *variable: the grid represents the second layer of the calorimeter in the η-ϕ plane. The photon is in the center of the grid with its hottest cell denoted by a red square. All topological clusters, represented in orange, whose barycenter falls inside the isolation cone are included in the isolation computation. The 5×7 fixed-size mask over the cluster of the photon is used to subtract the energy of the photon. The 4 circles centered around the photon candidate cluster barycenter denote the following in order of increasing radius: 1 and 2 Molière Radius; and two cones of 0.2 and 0.4 opening.*

from simulated samples in which only a photon is present in the detector, without pileup. The distribution of the energy leakage can be qualitatively described by a peak with a small contribution from higher energy tails. The parametrization of the peal position of the energy leaking outside the mask is performed in the $\log_{10}(\mu_{\text{iso}}) - \log_{10}(E_T^{\gamma})$ plane, where μ_{iso} corresponds to the peak position of the energy leakage. The leakage correction is fairly independent of the cone size, since the EM shower of the photon is well contained within the smallest cone of size $\Delta R = 0.2$, with differences of a few tenths of MeV between the smallest and largest cone. The applied leakage correction for different $|\eta^{\gamma}|$ bins is shown in Figure 4.3a.

The supercluster core subtraction, still in development, consists removing directly the supercluster of the photon candidate. It is by construction expected to be more robust than the fixed-mask core subtraction, since each correction contributes to widening the calorimetric isolation energy distribution (see Figure 4.3b).

However, two issues were identified when evaluating the performance of the supercluster core subtraction. The subtraction relies on the reconstruction performance of the supercluster of the egamma object: identify and collect all topoclusters belonging to the egamma object. If a topocluster belonging to the photon is not attached to the supercluster, it will not be subtracted from the isolation energy and will systematically bias the estimate of the calorimetric isolation energy towards larger values. This is observed



a) b)

Figure 4.3: *(a) The leakage correction as a function of E_T^γ. The correction is a function of the geometry and material in the detector, reflected in the definition of the $|\eta^\gamma|$ bins. A similar correction is applied to corrected photons. The correction is larger in the endcaps compared to the barrel at a given E_T^γ. (b) The isolation energy distribution for the two different core subtraction techniques: in blue the distribution with the fixed-size mask subtraction and in red with the supercluster core subtraction.*

to occur mostly for energetic photons ($E_T^\gamma > 100$ GeV) in the endcap regions, where a secondary peak appears in the isolation energy distribution from clusters not correctly subtracted.

The second issue observed when evaluating the supercluster core subtraction is related to the performance of the isolation variable regarding prompt-fake discrimination. The hadronic energy flow surrounding a fake photon has an electromagnetic component which can create a topocluster in the EM calorimeter. These topoclusters can be included in the supercluster reconstruction and then subtracted with the energy of photon candidate (a fake in this case). Then, it would decrease its calorimetric isolation energy, decreasing the prompt-fake discrimination and hence worsening the performance. Since the supercluster core subtraction presents these issues, despite its robustness and resolution improvements, the fixed-mask core subtraction is used in the following for all the presented results.

4.1.2 Pileup subtraction

The underlying event and pileup interactions deposit energy in the calorimeters. As these energy deposits overlap with the photon signal, they reduce the prompt-fake discrimination to widening the calorimetric isolation energy distribution. To subtract the contribution from pileup to the isolation energy a procedure is followed to estimate, in an event-by-event basis, the ambient energy density ρ [119]. It is computed with the FastJet [120] package which reconstructs all the jets in the event using a $k_\perp$ algorithm [121, 122]. Each jet is in correspondence with a certain area via Voronoi tessellation [123] in the $\eta - \phi$ plane, which is used to assign an individual energy density $\rho_i = p_{T,i}/A_{\text{Voronoi},i}$. The median of the distribution of the individual energy density is used as an estimate of the ambient energy density, since the mean is sensitive to high energy photons/jets compared to the jets produced in the underlying event.

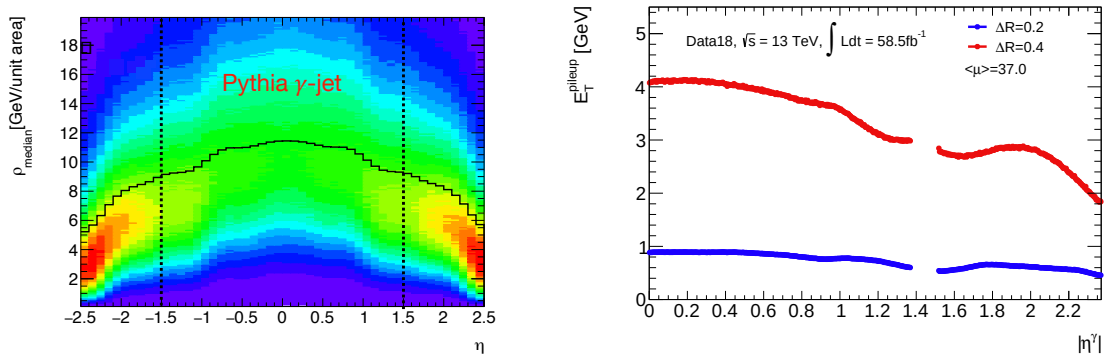
In Run 1, the ambient energy density was estimated for two regions of the detector, the barrel region covering up to $|\eta| < 1.5$ and the endcap region up to $|\eta| < 3.0$ (see

Figure 4.4a). Then, the pileup correction was computed by multiplying the effective area of the isolation cone by the median ambient energy density ρ_{median}:

$$E_T^{\text{pileup}}(|\eta^\gamma|, \langle\mu\rangle) = \rho_{\text{median}}(A_{\text{cone}} - A_{\text{mask}}). \tag{4.1}$$

where A_{cone} is the size of the isolation cone and A_{mask} is the 5×7 cells mask size.

However, the fact that only two values for ρ_{median} were used, lead to over-subtract (under-subtract) the pileup contribution to the isolation energy in the regions in the barrel close to the crack (in the endcap close to the crack). With the aim of providing a more homogeneous transition with $|\eta^\gamma|$ and a better estimation of the pileup contribution to the calorimetric isolation energy, a parametrization of the ambient energy density as a function of $|\eta^\gamma|$ is computed. This correction varies depending on the size of the cone, pileup conditions and region of the detector, since a larger energy flow is expected towards the forward regions of the detector. A comparison of the estimation of the energy arising from pileup for the two different cone sizes is shown in Figure 4.4b.



a) b)

Figure 4.4: *(a) The ambient energy density of γj events in simulation. The solid black line denotes the mean value as a function of η. The dashed black lines denote the two regions used for the barrel and endcap pileup corrections. (b) The average pileup correction measured in data as a function of $|\eta^\gamma|$ for two different cone sizes: $\Delta R = 0.2$ in blue and $\Delta R = 0.4$ in red. As expected, a larger correction is applied for the larger cone.*

4.1.3 Corrected calorimetric isolation energy

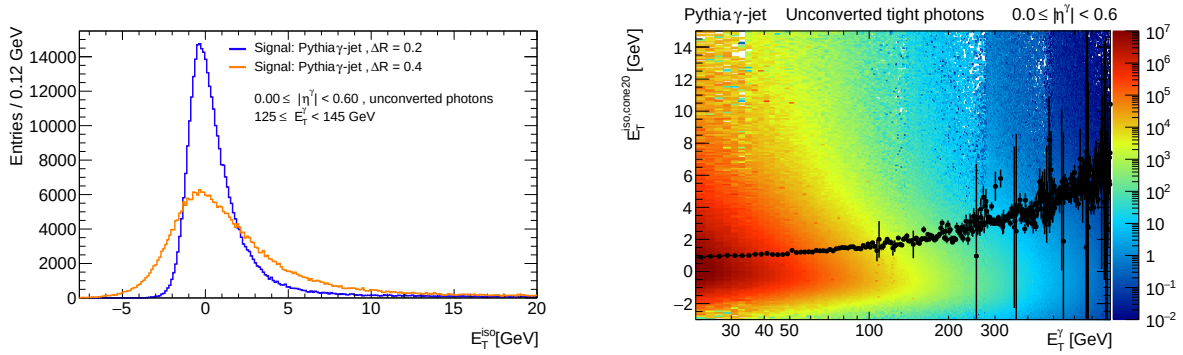
Accounting for all the corrections, the calorimetric isolation energy is computed as follows:

$$E_T^{\text{iso}}(R) = \left(\sum_{r_i<R}^{\text{clusters}} E_T^i\right) - E_T^{\text{core}} - E_T^{\text{leakage}}(E_T^\gamma, |\eta^\gamma|) - E_T^{\text{pileup}}(|\eta^\gamma|, \langle\mu\rangle), \tag{4.2}$$

where R is the radius of the cone, E_T^i is the transverse energy of each topocluster found inside the cone, E_T^{core} represents the fixed-size mask correction, $E_T^{\text{leakage}}(E_T^\gamma, |\eta^\gamma|)$ denotes the leakage correction and $E_T^{\text{pileup}}(|\eta^\gamma|)$ is the pileup correction defined in Equation 4.1.

Once all the corrections are aplied, the calorimetric isolation variable should be E_T^γ-independent and centered at values close to zero for prompt photons. However, a remaining

dependence remains, increasing the calorimetric isolation energy with increasing E_T^γ (see Figure 4.5b).



a) b)

Figure 4.5: *(a) The distribution of isolation energy for simulated unconverted photons with E_T^γ between 125 and 145 GeV in the barrel for the two different cone sizes. The distribution is wider for the larger cone since more topoclusters fall inside it. (b) The isolation energy $E_T^{iso,cone20}$ as a function of the transverse energy of photons E_T^γ. The mean of the isolation energy distribution is overlaid in black dots, showing how it increases at larger E_T^γ.*

This remaining dependence justifies the choice of E_T^γ-dependent photon isolation criteria to avoid efficiency drops at higher energies. This is of course possible because the background behaves similarly, and E_T^γ-dependent cuts allow to keep the same efficiencies without worsening the purities. Three isolation working points are defined to provide different purities and efficiencies using both tracking and calorimeter information and different cone sizes: `FixedCutLoose`, `FixedCutTight` and `FixedCutTightCaloOnly`. Their definitions are presented in Table 4.1.

Table 4.1: *The definitions of the three standard photon isolation working points.*

Working Point	Calorimetric Isolation Requirement	Track Isolation Requirement
`FixedCutLoose`	$E_T^{\text{iso,cone20}} - 0.065 \times E_T^\gamma < 0.0$ GeV	$p_T^{\text{cone20}} - 0.05 \times E_T^\gamma < 0.0$ GeV
`FixedCutTight`	$E_T^{\text{iso,cone40}} - 0.022 \times E_T^\gamma < 2.45$ GeV	$p_T^{\text{cone20}} - 0.05 \times E_T^\gamma < 0.0$ GeV
`FixedCutTight(CaloOnly)`	$E_T^{\text{iso,cone40}} - 0.022 \times E_T^\gamma < 2.45$ GeV	–

4.2 Studies of photon isolation efficiency: methodology

An accurate assessment of photon isolation efficiencies is an important ingredient in, for instance, cross-section measurements in analyses involving photons in the final state. However, the separation between prompt photons and fakes is not perfect and after tight photon identification, a large fraction of non-prompt photons remains in the sample of photon candidates. The prompt photon background is dominated by jets in which a highly

energetic π^0 decays to a collimated pair of photons and thus the larger isolation energy is expected compared to prompt photons. The separation using the isolation energy is not perfect and an event-by-event discrimination is not possible. However, the distributions are different enough to perform a statistical subtraction of the background.

Photon isolation efficiencies are defined as the fraction of reconstructed tight identified photons with an isolation energy E_T^{iso} below a certain value provided by the working point under study:

$$\varepsilon_{WP} = \frac{N_{\text{Tight, pass WP}}}{N_{\text{Tight, pass WP}} + N_{\text{Tight, fails WP}}} \tag{4.3}$$

where WP stands for one of the three working points defined in Table 4.1.

Efficiencies are measured in 6 pseudorapidity $|\eta^\gamma|$ bins, 17 E_T^γ bins and separately for converted and unconverted photons (see Table 4.2). The E_T^γ-binning used in this measurement matches the different trigger prescales, which vary as a function of E_T^γ. The binning in $|\eta^\gamma|$ follows the differences in the geometry of the detector and material upstream the calorimeters (see for example Figure 3.6). For practical purposes regarding the fit quality, 5 additional pileup, $\langle\mu\rangle$, bins are used, since increasing pileup widens the isolation energy distribution.

Table 4.2: *Binning used in photon isolation efficiency measurements.*

Variable	Binning
E_T^γ [GeV]	$[22,25)$, $[25,30)$, $[30,35)$, $[35,40)$, $[40,45)$, $[45,50)$, $[50,55)$, $[55,65)$, $[65,85)$, $[85,105)$, $[105,125)$, $[125,145)$, $[145,175)$, $[175,280)$, $[280,500)$, $[500,800)$, $[800,1500)$
$\|\eta^\gamma\|$	$[0.00,0.60)$, $[0.60,0.82)$, $[0.82,1.15)$, $[1.15,1.37)$, $[1.52,1.81)$, $[1.81,2.37)$
$\langle\mu\rangle$	$[0,20)$, $[20,28)$, $[28,36)$, $[36,44)$, $[44,\infty)$
Conversion	Unconverted, Converted

In this section, measurements of photon isolation efficiencies are presented for the previously defined isolation working points. A selection of plots is shown in the introductory parts where the method is explained. The full collection of results is shown at the end in Section 4.3.

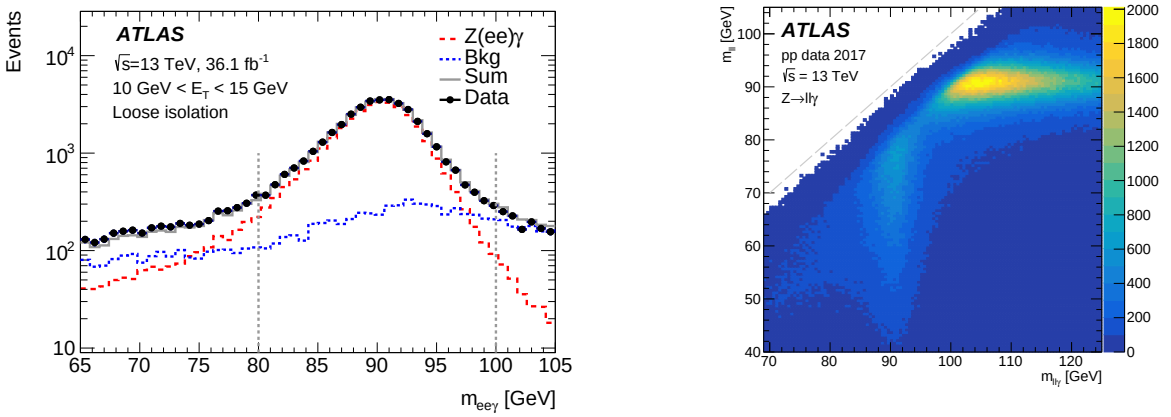
4.2.1 Signal and background samples

Isolation efficiency measurements are performed in a wide photon energy range, from E_T^γ=10 GeV up to TeV energies.

A low energies ($10 < E_T^\gamma < 100$ GeV), $Z \to ll\gamma$ events are used. The invariant mass of the three-object system is necessarilly the mass of the Z boson, leaving a very clean signature in the detector. This type of events (briefly introduced in Section 3.2 for photon identification measurements) constitute a powerful sample to evaluate photon performances.

Radiative $Z \to ll\gamma$ candidates are selected from events triggered with single or dilepton unprescaled triggers with two reconstructed electrons and one reconstructed photon with more than E_T^γ>10 GeV. Selections on the combined dilepton invariant mass m_{ll} : $[40,83]$ GeV and the three-body invariant mass $m_{ll\gamma}$: $[80,100]$ GeV are used in order to enrich the sample on final state radiation photons (see Figure 4.6b). This selection ensures a photon purity between 95% and 98% with increasing transverse momentum. However,

it covers only the low part of the energy spectrum and it suffers from low statistics (see Figure 4.7).



a) b)

Figure 4.6: *(a) Distribution of $m_{ee\gamma}$ events satisfying the selection criteria described in the text except for the $m_{ee\gamma}$ selection denoted by the two dotted vertical lines. Photons are selected to have $10 < E_T^\gamma < 15$ GeV. The solid gray line represents the fit to the data described as a sum of the signal invariant mass template in red and background in blue. (b) Dielectron invariant mass as a function of the 3-body $ll\gamma$ mass. Final state radiation photons are observed in the region around the Z mass in the horizontal axis while initial state radiation photons are observed in the vertical axis around the Z mass. [118]*

At higher energies ($22 < E_T^\gamma < 1000$ GeV), events with a triggered photon are used. Data are recorded using single photon triggers of the form `HLT_gX_loose`, with `X` denoting the E_T^γ-threshold. These triggers are prescaled for $E_T^\gamma < 140$ GeV in which only loose identification and the E_T^γ requirement is applied. The recorded events are predominantly QCD γ-jet events, in which the recorded photon is recoiling against a jet. The amount of available data is shown in Figure 4.7, in which the impact of the trigger prescales is also shown.

Compared to the radiative Z samples, γ-jet samples suffer from larger backgrounds, especially at low E_T^γ, and contributions from other photon production mechanisms, like bremsstrahlung photons (see Figure 1.4c), which modify the shape of the isolation energy distribution. The isolation energy from bremsstrahlung photons has in general larger tails compared to those from direct photons, since the nearby jet may deposit extra energy in the isolation cone. The studies shown here are almost exclusively performed with γ-jet samples, with occasional mentions to the results from radiative Z samples.

Signal simulated samples are generated with PYTHIA8, including parton showering and hadronization, and consist of events in which a prompt photon is produced in the final state recoiling usually against a jet. The processes generated are the ones illustrated in Figure 1.4, including dijet events in which one of the outgoing partons radiates a photon. To populate all the E_T^γ spectrum, the samples are generated in E_T^γ slices. These samples are referred to in the following as single photon samples.

Background samples are extracted from data by relaxing photon tight identification, in particular not requiring the cuts on the strip shower shape variables (see Section 3.3). The strip variables are used in photon isolation studies because they are assumed to be not correlated with the calorimetric isolation energy, as the desposits of the first layer

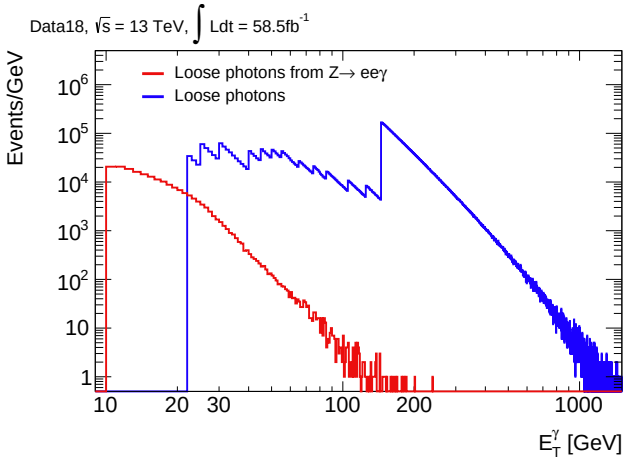


Figure 4.7: *The transverse momenta E_T^γ of photon candidates recorded during the 2018 data-taking period for events passing the radiative Z selection (red line) and single photon triggers (blue line). At lower transverse energies, larger prescales are needed to cope with the higher rates for the γ-jet sample.*

are usually subtracted within the energy of the photon. Reverting the cuts in a subset of the photon identification strip variables allows to define background control regions from data.

Five strip variables are used to define background-enriched samples: $w_{s\,\mathrm{tot}}$, w_{s3}, ΔE_s, f_{side} and E_{ratio}. These samples are composed of photon candidates failing tight identification but passing a subset of the strip variables. Five different identification criteria for each background sample are defined in Table 4.3 and used in data. These criteria are denominated Loose' selections. Unless otherwise stated, when referring to a Loose' selection, **tight identification is assumed to be failed**.

Despite using data-driven background samples, simulated samples enriched in fake photons are available. These samples, denominated "jet-filtered" samples, are built from PYTHIA8 dijet events, filtered with a grid-based algorithm that enhances the contribution of fakes. However, the available statistics after photon reconstruction and identification is limited.

The isolation energy distribution for the jet-filtered samples is compared with the background-enriched samples obtained from data in Figure 4.8. Fair agreement, within

Table 4.3: *The variables used to determine the definitions of the Loose, Tight and Loose' selections. Here, Loose'4 indicates that the tight identification criteria are applied to all shower variables, with the exception that the shower shapes indicated are not subject to the nominal tight identification selections.*

ID name	Cuts
Loose	R_{had_1}, R_{had}, R_η, w_{η_2}
Tight	Loose + R_ϕ, w_{s3}, f_{side}, ΔE_s, E_{ratio}, $w_{s\,\mathrm{tot}}$, f_1
Loose'2 (LP2)	Tight – w_{s3}, f_{side}
Loose'3 (LP3)	Tight – w_{s3}, f_{side}, ΔE_s
Loose'4 (LP4)	Tight – w_{s3}, f_{side}, ΔE_s, E_{ratio}
Loose'5 (LP5)	Tight – w_{s3}, f_{side}, ΔE_s, E_{ratio}, $w_{s\,\mathrm{tot}}$

statistical uncertainties, between the shapes of background simulated events passing tight and passing Loose'4 is shown in Figure 4.8 to support the assumption of poor correlation between the strip variables and E_T^{iso}. The lack of realistic simulated background samples and with sufficient statistics motivates the usage of data-driven control regions.

Although the Loose' photon selection provides a fakes-enriched sample, a non negligible fraction of true prompt photons leaks into it. The fraction of signal events passing Loose' selections varies with E_T^{γ} and $|\eta^{\gamma}|$ and it needs to be subtracted from the fakes enriches subsamples to obtain the correct background shape.

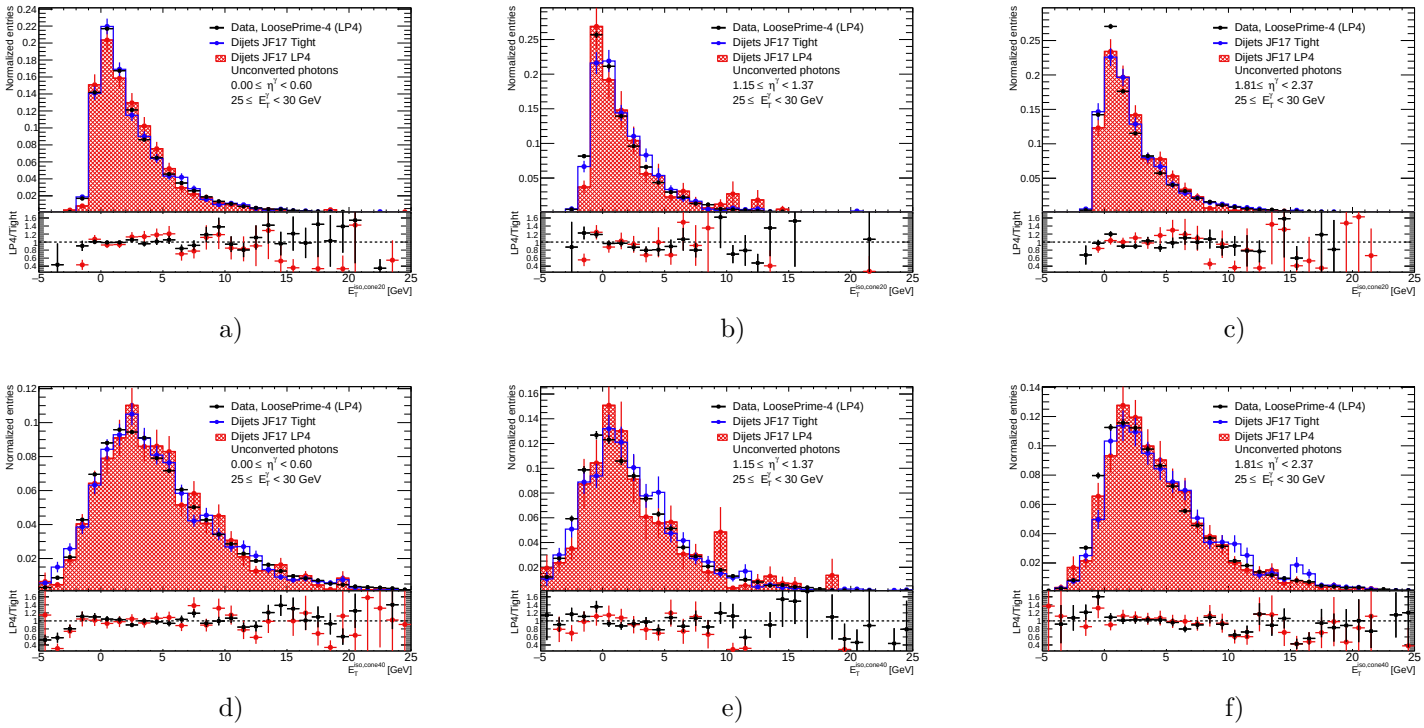


Figure 4.8: *Isolation energy distributions for background simulated events passing tight identification (blue solid line) and Loose'4 identification (red filled dashed) and for data events passing Loose'4 identification (black solid dots).*

4.2.2 Fit methodology

This subsection describes how the isolation energy shape for signal photons is extracted from data, accounting for the background contribution. This is performed through a "template fit"[1], in which the shape for both components is obtained from fits to either simulated signal or data-driven samples. Four samples are defined to be used in the template fit: simulated events passing tight identification, simulated events passing a given Loose' identification, data events passing tight identification and data events passing a given Loose' identification. The control region built with photons passing Loose'4 identification is defined in the following as the nominal control region. A complete example of the template fit in a single bin is shown in Figure 4.9 as a guide to understand the methodology explained in the following.

[1]The procedure described here inherits the basis from the existing methodology used for previous photon isolation measurement.

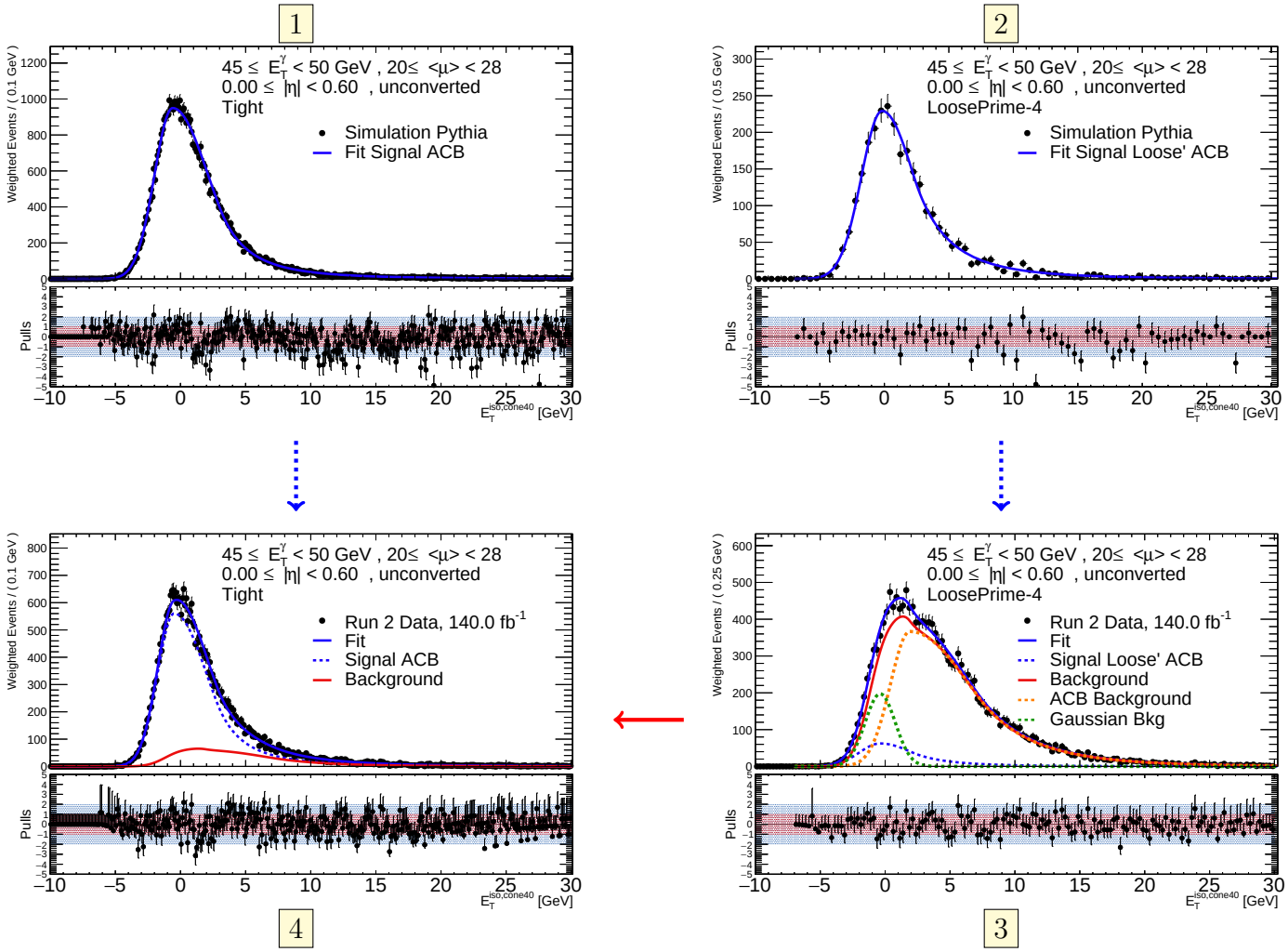


Figure 4.9: *Steps of the photon isolation template fit method. In step 1, the isolation energy distribution of simulated photons passing tight identification is fit with an ACB. In step 2, the isolation energy distribution of simulated photons passing Loose'4 identification is fit with another ACB. In step 3, the isolation energy distribution of data events passing Loose'4 identification is fit with a composite model formed by an ACBfor the bulk of the background shape (orange), a Gaussian for the background without clusters (green) and an ACB to describe the signal component present in the control regions, which inherits all the parameters, except the peak position, from step 2. In step 4, the isolation energy distribution of data events passing tight identification is fit with two components: the signal shape obtained from step 1 fixed and the background shape obtained from step 3.*

First, the signal template (step 1) for each bin is built from a fit to the isolation energy distribution of simulated signal photons passing tight identification with an Asymmetric Crystal Ball[2] (ACB), which consists of a bifurcated Gaussian core (a Gaussian with two different widths at each side of the mean value) with a power law tail in one of the sides. The mathematical expression is given in Appendix B. This template is referred to in the following as the **ACB signal tight template**.

[2]Details on Crystal Ball-like functions is given in Appendix B

In order to account for the signal photons present in the background control regions, a fit with an ACB is performed on simulated photons passing Loose'4 identification (step 2). This template is referred to in the following as the **ACB signal loose' template**.

The background template is obtained from a composite fit to the isolation energy distribution of events from data passing the Loose'4 identification (step 3). The model is composed by four different analytical functions:

- An ACB is used to describe the contribution of signal photons in the control regions, inheriting the shape from the **ACB signal loose' template** of the simulation.
- An ACB to describe the bulk of the background distribution at large E_T^{iso}. This component is referred to in the following as the **ACB background template**.
- A Gaussian distribution to describe fakes without clusters in the isolation cone. This component is referred to in the following as the **Gaussian background component**.

The total background template is the sum of the ACB distribution and the Gaussian, without the signal component, and it is referred to in the following as the **background template**.

With the signal and background templates, the isolation energy distribution of events from data passing tight identification is fitted (step 4). The signal component in this fit is referred to as the **data signal template**. In this fit, the parameters of the data background template are fixed to those obtained from the Loose'4 sample. The parameters of interest of the fit are:

- Parameters that describe the shape of the signal in data: the peak position μ_{ACB} and left and right widths respectively $\sigma_{\text{ACB,left}}$ and $\sigma_{\text{ACB,right}}$; where ACB stands for Asymmetric Crystal-Ball (defined in Appendix B).
- Signal and background yields N_{signal} and N_{bkg} in data signal region.

This methodology provides good results in all considered bins. Also, it allows to integrate in the statistical uncertainty several sources of systematic errors, such as the fraction of signal photons in the background enriched control regions. The results are presented on plots in the following subsections.

4.2.3 Corrections to the isolation energy in the simulation

As shown in Equation 4.2, the calorimetric isolation energy is corrected for uncorrelated elements like the core leakage correction and pileup subtraction. A difference in the behaviour of these corrections in data and simulation would directly translate in differences in efficiencies. Two sets of parameters encapsulate those differences: data-driven shifts and scale factors.

Data-driven shifts are a correction that accounts for differences between data and simulation most probable value ("peak position"):

$$DD_{\text{shift}} = \mu_{ACB,Data} - \mu_{ACB,Simu}, \tag{4.4}$$

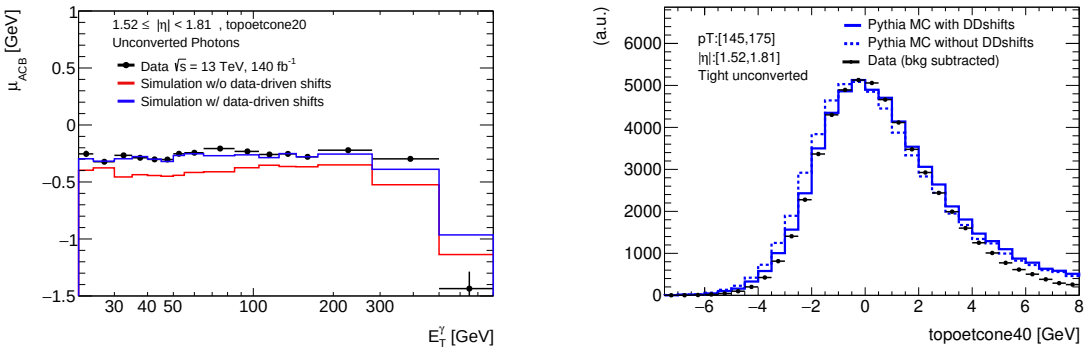
where $\mu_{ACB,Data}$ and $\mu_{ACB,Simu}$ are the peak positions of the fitted ACB distributions in data and simulation respectively. The most likely candidate to account for these

differences is the core leakage subtraction (see Section 4.1.1). The difference in the peak position is larger in the endcaps and at higher E_T^γ, ranging from a few MeV up to a few GeVs. The difference between the peak positions between data and simulation is obtained with the procedure explained in the next section before performing the efficiency measurements. The data-driven shift is applied to the simulation and it is computed from the combination with the same measurement performed with radiative Z events to cover the full E_T^γ spectrum. An illustration of this effect for a single $|\eta^\gamma|$ bin is shown in Figure 4.10a. This correction is in general systematically implemented in every physics analyses, except those evaluating independently isolation performances for particular reasons.

Once the data-driven shifts are applied, the remaining difference observed between data and simulation efficiencies is encapsulated in a scale factor, defined as:

$$SF = \frac{\varepsilon_{\text{Data}}}{\varepsilon_{\text{Simu}}}, \tag{4.5}$$

where $\varepsilon_{\text{Data}}$ and $\varepsilon_{\text{Simu}}$ are the measured efficiencies in data and simulation respectively. Various sources can produce differences between both samples (different fragmentation photon fraction, pileup mismodelling,..). An illustration of these two corrections on the isolation energy in the simulation are shown in Figure 4.10.



a) b)

Figure 4.10: *(a) The peak position of the isolation energy $E_T^{iso,cone20}$ (called μ_{ACB} in the plot) as a function of the photon transverse energy E_T^γ for unconverted photons in the region $1.52 \leq |\eta^\gamma| < 1.81$. The values measured in data are represented by black markers. The values measured in the simulation with and without data-driven corrections are denoted by blue and red lines respectively. (b) The isolation energy for tight unconverted photons in data (black dots) and simulation with (solid blue line) and without (dashed blue line) data-driven corrections applied. The observed width in data is slightly narrower compared to that observed in simulation.*

4.3 Studies of photon isolation efficiency: results

The results in this section are presented as follows. First, the parameters of the models obtained in the template fit are shown as a function of E_T^γ, compared between data and simulation for a single $|\eta^\gamma|$, $\langle\mu\rangle$ and conversion status bin. Then, the combination of data-driven shifts applied to the simulation from single photon and radiative Z events is

shown. Finally, the measured efficiencies in data are presented together with the scale factors.

4.3.1 Model parameters

The parameters of the signal model have no current use as an input to other photon measurements but they are a useful tool to verify the good behaviour of the fits and to understand the sources of possible remaining discrepancies between data and simulation after the data-driven shifts.

The peak position μ_{ACB} shows different trends as a function of E_T^γ across the various $|\eta^\gamma|$ bins (see Figure 4.11). The deviation from 0 is likely to arise from an inacurate estimation of pileup or core energy leakage correction. In general the values remain stable at low E_T^γ values and increase or decrease, depending on the $|\eta^\gamma|$ bin at larger E_T^γ values, pointing towards an incorrect leakage energy estimation.

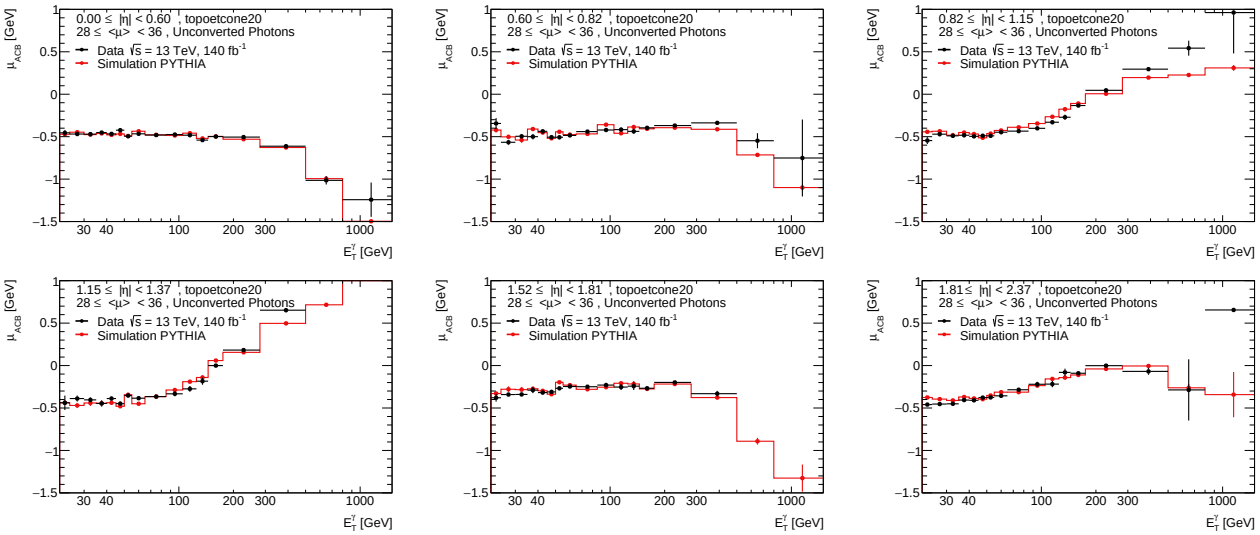


Figure 4.11: *The peak position of the isolation energy $E_T^{iso,cone20}$, μ_{ACB}, as a function of the photon transverse energy E_T^γ for unconverted photons and $28 \leq \langle\mu\rangle < 36$. The six different $|\eta^\gamma|$ bins are shown. The values measured in data are represented by black markers. The values measured in the simulation are represented by red markers.*

A discrepancy of approximately 10% is consistently observed between data and simulation widths at low E_T^γ(see Figure 4.12). The observed difference on the left side of the distribution, characterized by σ_{left}, is attributed to an overestimation of pileup present in the simulation. Further discussion on the pileup overestimation and its impact on the isolation energy is currently under investigation.

4.3.2 Purities

The purities are defined as the ratio of the signal yield over the total yield in the range of E_T^{iso} that contains 90% of the signal. The observed trend as a function of E_T^γ shows how the purity increases from less than 60% up to more than 95% for high energy photons (see Figure 4.13). Worse purities are observed in the regions close to the crack ($|\eta^\gamma|$: [1.37, 152]).

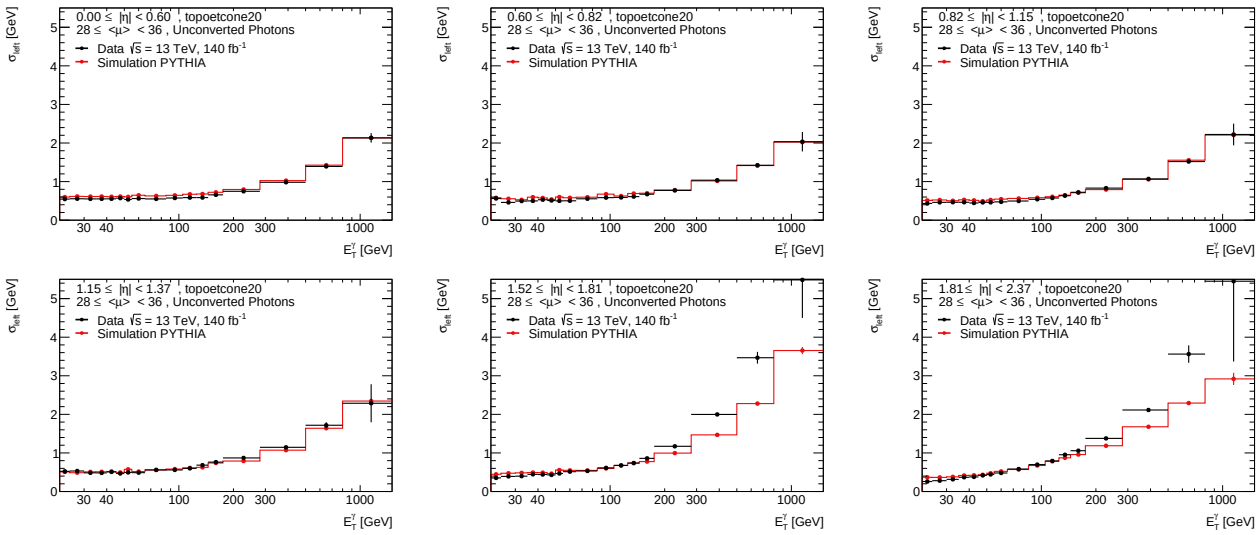


Figure 4.12: *Left widths of the ACB used to characterize the isolation energy $E_T^{iso,cone20}$ as a function of the photon transverse energy E_T^γ for unconverted photons and $28 \leq \langle\mu\rangle < 36$. The six different $|\eta^\gamma|$ bins are shown. The values measured in data are represented by black markers. The values measured in the simulation are represented by red markers.*

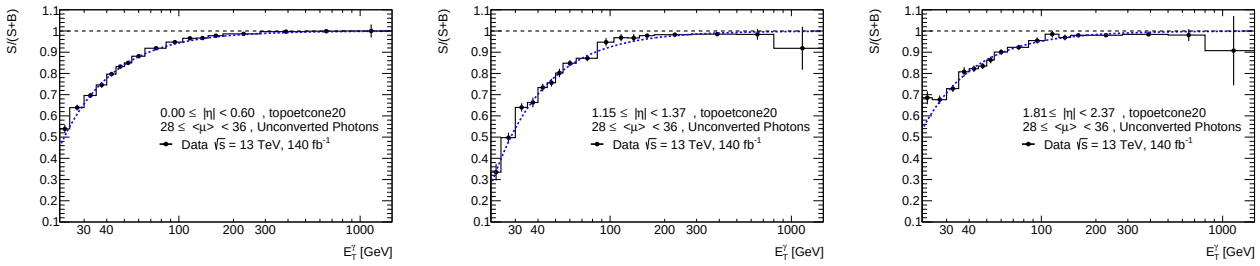


Figure 4.13: *Measured purities in data as a function of the photon transverse energy E_T^γ for tight unconverted photons and $28 \leq \langle\mu\rangle < 36$. Three different $|\eta^\gamma|$ bins are shown: $0.00 \leq |\eta^\gamma| < 0.60$, $1.15 \leq |\eta^\gamma| < 1.37$ and $1.81 \leq |\eta^\gamma| < 2.37$.*

4.3.3 Data-driven shifts

The data-driven shifts obtained from radiative Z and single photon samples are combined as a weighted average for each of the bins to take into account their statistical uncertainties. Fair agreement is observed between both measurements in Figure 4.14. The combined correction is obtained with Friedman's super smoother [124] to provide a continuous correction as a function of E_T^γ for each bin. Since the origin of the difference in the peak position it is caused by a poor core energy leakage subtraction, it is fairly independent of the cone size, and thus the same data-driven shifts are used for both cone sizes. As a function of $|\eta^\gamma|$ the correction varies, being larger close to the crack region and in the endcaps.

4.3.4 Systematic uncertainties

The dominant systematic uncertainty arises from the method used to extract the signal yield: either from the integral of the signal PDF or by statistically subtracting the background component. The difference in the signal yield obtained with both methods

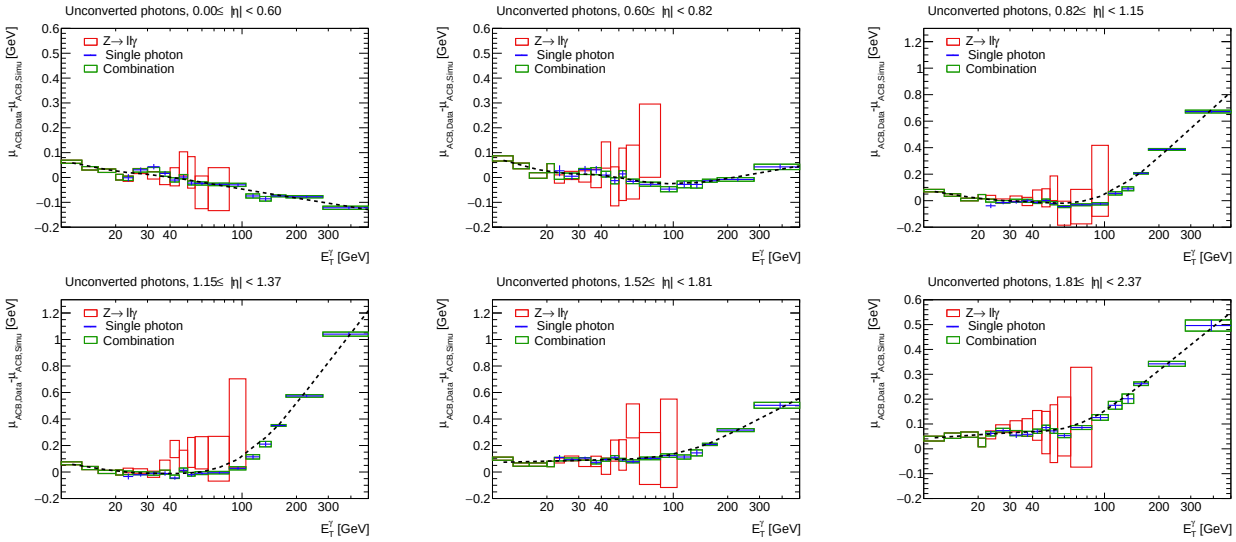


Figure 4.14: *Data-driven shifts as a function of the photon transverse energy E_T^γ for unconverted photons and each $|\eta^\gamma|$ bin. Blue markers denote the data-driven shift measured in single photon samples. Red markers denote the data-driven shift measured in radiative Z samples. Green markers denote the combined value of the two mentioned measurements. The dashed black line denotes the smoothed final correction applied to the simulation.*

is taken as systematic uncertainty, with values ranging from 2% to less than 0.1% with increasing energy.

Several improvements in the presented methodology with respect to previous measurements [117] have reduced the list of systematic uncertainties under consideration, since some of those are now absorbed in the statistical uncertainty of the fit.

- As explained in Section 4.2.2, the distribution in data of Loose' events is fit with three components, including one for fakes that have zero clusters in their isolation cone and thus have isolation energies peaking at zero. This component was not considered in previous iterations, and was partly absorbed in the component describing prompt photons failing tight identification but passing the Loose' requirement. As the event rate for this component was constrained to the value observed in Monte Carlo, this translated into a consistent overestimation of its value, thus biasing the fraction of ACB background. All three components are now free in the fit to the Loose' sample in data.

- The second dominant systematic effect arises from the quality of the fits. The χ^2/nDoF values in the fits to the tight photon candidates in data are typically close to 1 in the lower E_T^γ bins, and tend to increase in the higher E_T^γ bins, with values reaching χ^2/nDoF$\sim$2 to 5. A systematic uncertainty is assigned to cover for this, using the PDF recipe.

4.3.5 Efficiencies and scale factors

The signal yields for both data and simulation are obtained as the integral of the fitted isolation shape below a given threshold depending on the working point (see Table 4.1).

The measurement of isolation efficiencies with single photon events is sensitive to a precise estimation of the signal width and the background yield, in particular for

low E_T^γ photons, where the purity is low. The measured efficiencies are shown for the `FixedCutLoose` working point in Figures 4.15 for one of the $\langle\mu\rangle$ bins, $28 \leq \langle\mu\rangle < 36$. For `FixedCutLoose`, efficiencies are higher than 95% beyond E_T^γ =50GeV and in general in fair agreement with data, with scale factors below 1%. For lower energies, scale factors are larger up to 3-4%, depending on the $|\eta^\gamma|$ bin. This discrepancy is the consequence of the observed difference on the signal widths between data and simulation (see Figure 4.12).

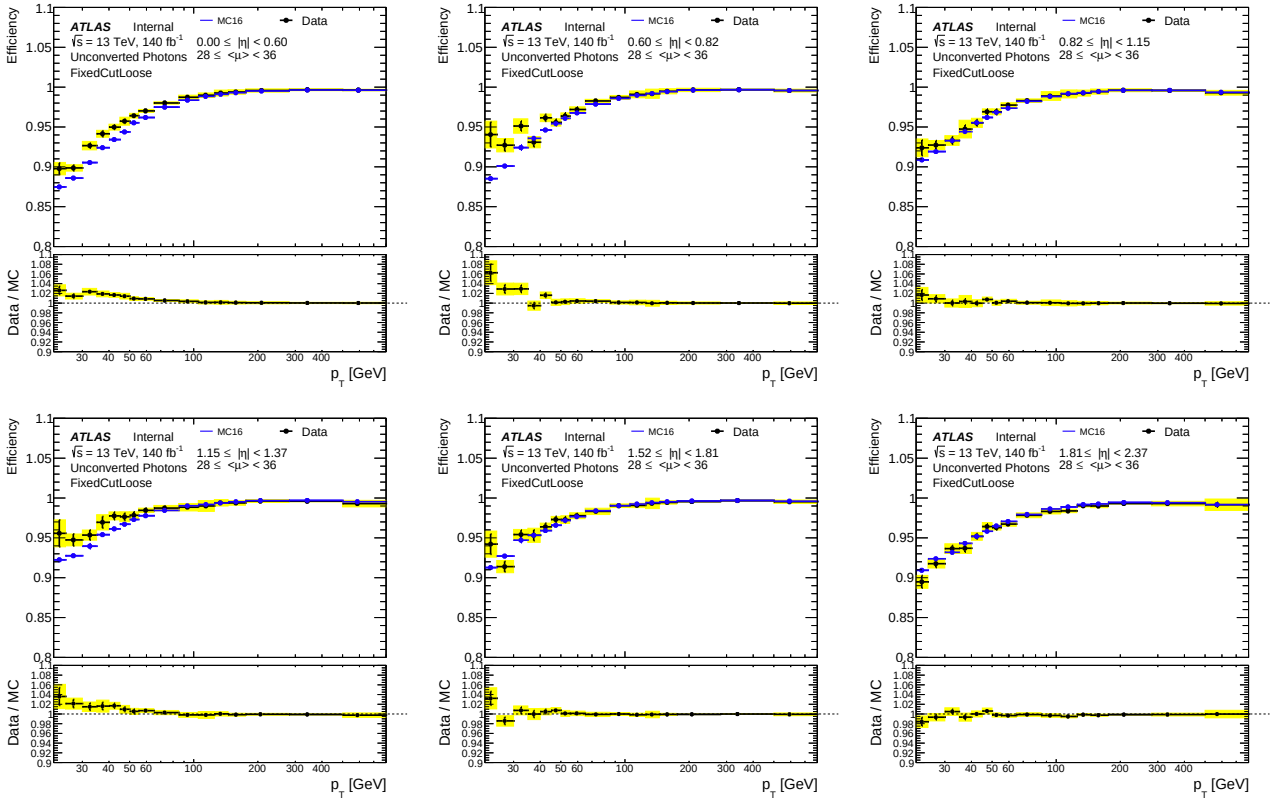


Figure 4.15: *Data (black markers) and simulation (blue markers) efficiencies of the* `FixedCutLoose` *working point for photons in γ-jet events. The pannels below show the scale factor defined as the ratio of data and simulation efficiencies. Solid error bars denote the statistical uncertainty and the yellow bands denote the total uncertainty.*

An important consequence of increasing $\langle\mu\rangle$ is the degradation of photon isolation efficiencies. This is further discussed in the next subsection.

4.3.6 Scale factors as continuous functions

The scale factors presented in the previous subsection are applied to the simulation integrated in $\langle\mu\rangle$ as a binned correction on E_T^γ for each $|\eta^\gamma|$ and conversion status bins. This procedure does not exploit the full knowledge acquired with the measurement presented, since the increase in $\langle\mu\rangle$ also plays an important role on isolation efficiencies. . The results presented until now can be summarized as follows:

- Efficiencies increase as a function of the E_T^γ of the photon, ranging from 80% up to more than 98% for `FixedCutLoose` and from 70% to more than 95% for `FixedCutTight`.
- Increasing pileup degrades the efficiencies for all energies and working points.

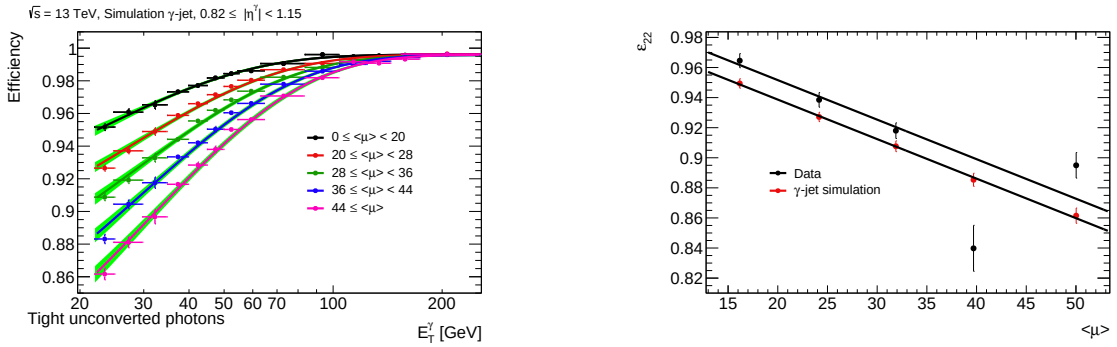
- No clear trend as a function of $|\eta^\gamma|$ is observed in the scale factors, since material dependent effects modify the isolation shape and can induce specific data-simulation discrepancies.

In this subsection a continuous description of the scale factors as a function of E_T^γ and $\langle\mu\rangle$ is presented for the `FixedCutLoose` working point as an alternative to binned corrections. For this, the efficiencies are described with a multidimensional model function of E_T^γ and $\langle\mu\rangle$. The efficiencies as a function of E_T^γ are parametrized by an exponential of the form:

$$f(E_T^\gamma;\varepsilon_{22},\varepsilon_\infty,\tau) = \varepsilon_\infty + (\varepsilon_{22}-\varepsilon_\infty)exp\left(-\frac{E_T^\gamma - x_0}{\tau}\right), \tag{4.6}$$

where ε_∞ is the the efficiency measured in the limit of infinite E_T^γ, ε_{22} is the efficiency measured at x_0 and τ represents the decay length of the exponential. In the following, x_0 is fixed to 22 GeV, and thus the value of ε_{22} corresponds to the value of the efficiency at 22GeV.

To motivate the choice of this expression, Figure 4.16a, shows the efficiencies measured in the simulation under different pileup conditions for unconverted photons in $0.82 \leq |\eta^\gamma| < 1.15$. From fits to the efficiencies in data and simulation, the evolution of each parameter as a function of $\langle\mu\rangle$ is obtained. The isolation efficiency measured for photons with 22 GeV, ε_{22}, is shown as a function of $\langle\mu\rangle$ in Figure 4.16b for data and simulation. Both are described with linear trends.



a) b)

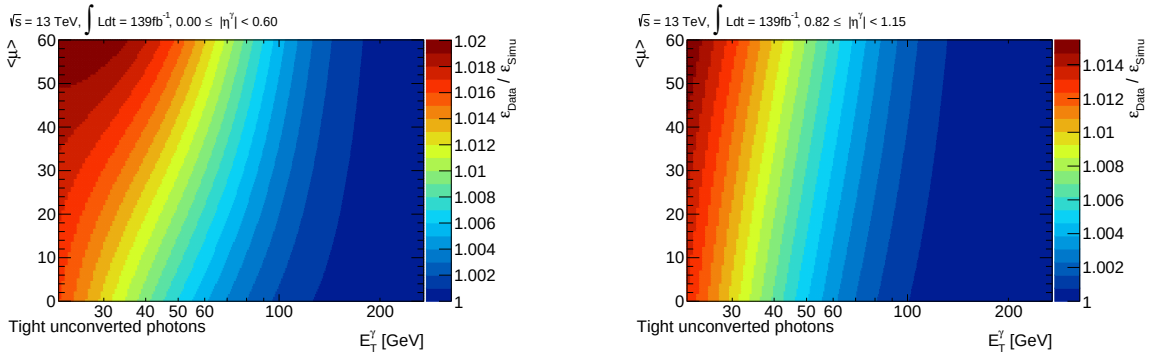
Figure 4.16: *(a) Photon isolation efficiencies measured on γ-jet simulated samples under different pileup conditions. The green bands denote the $1\pm\sigma$ fit uncertainties. (b) The parameter ε_{22}, that denotes the efficiency measured at 22 GeV, as a function of $\langle\mu\rangle$for data and simulation. The fit to the data has the same slope as the one obtained in the simulation.*

The efficiency at large E_T^γ, ε_∞, and τ do not show a clear dependence with $\langle\mu\rangle$, and thus are set constant with $\langle\mu\rangle$. Under these hypotheses, the full model used to describe the measured efficiencies in data and simulation is built as follows:

$$f(E_T^\gamma,\mu;\varepsilon_{22},\varepsilon_\infty,\tau) = \varepsilon_\infty + (\varepsilon_{22}(\mu)-\varepsilon_\infty)exp\left(-\frac{E_T^\gamma - x_0}{\tau}\right); \quad \text{with} \quad \varepsilon_{22}(\mu;\varepsilon_0,\alpha) = \varepsilon_0 + \alpha\mu. \tag{4.7}$$

where α and ε are the offset and slope of the linear parametrization of ε_{22} with $\langle\mu\rangle$. In particular, ε_0 represents the efficiency of photons with $E_T^\gamma = x_0$ at $\langle\mu\rangle$=0.

The ratio of fitted models to data and simulation efficiencies describes a continuous scale factor that encapsulates the difference between the parametrized efficiencies as a function of E_T^γ and $\langle\mu\rangle$. The choice of the model of course induces an additional systematic uncertainty to be considered on top to those described in the subsection 4.3.4. The result for unconverted photons in the ranges between $0.0 \leq |\eta^\gamma| < 0.6$ and $0.82 \leq |\eta^\gamma| < 1.15$ are illustrated in Figure 4.17, to be compared with the same bin in Figure 4.15. A fair agreement is observed between the continuous and binned scale factors.



a) b)

Figure 4.17: *Scale factors measured on unconverted photons as a function of E_T^γ and $\langle\mu\rangle$ in two different $|\eta^\gamma|$ bins: $0.0 \leq |\eta^\gamma| < 0.6$ (left) and $0.82 \leq |\eta^\gamma| < 1.15$ (right).*

4.4 Conclusions

In this chapter, photon isolation performances have been described using the full Run 2 dataset. Using γ-jet samples, the study covers a wide range in the photon E_T^γ spectrum, up to the TeV energy range.

Based on the framework used for the previous round of measurements, the description of the different components of the template fit has been improved and allows for a more robust estimation of the signal shape in data. This methodology deals differently with the systematic uncertainties compared to previous measurements, since sources such as the fraction of signal present in the control regions are absorbed in the statistical uncertainty of the fit. Moreover, an updated estimation of the leakage corrections directly translates into an efficiency improvement of up 7% at low E_T^γ [3] with respect to the last published result [117].

The results presented show a small discrepancy, at the percent level, between data and simulation, specially for low E_T^γ photons, related to an overestimation of the pileup in the simulation. The dominant systematic in the scale factors used to correct the simulation arises from the estimation of the signal yield in data after the subtraction of the background and it ranges from 2% down to less than 0.1%. Future plans to asses this discrepancy include dedicated studies based on correcting directly the isolation energy shape, specifically the observed difference on its width between data and simulation.

An alternative approach for future corrections to the simulation compared to the binned setup has been presented. Using a multidimensional model, a more robust estimation of the scale factors is obtained, based on their dependence on E_T^γ and $\langle\mu\rangle$.

[3] Depending on the $|\eta^\gamma|$ bin and working point.

Chapter 5

Prospective studies on photon isolation using cluster-level information

As detailed in the previous chapter, photon isolation makes use of information from the inner detector and the calorimeters. In particular, the energy and position of the topoclusters reconstructed in the calorimeters are used to estimate the overall energy flow found around a given photon candidate. This procedure rejects further backgrounds on top of photon tight identification. While track isolation is quite robust against pileup, as it is easier to identify tracks coming from pileup particles, calorimetric isolation is severely affected by increasing pileup conditions. The spurious energy deposits from pileup particles widen the isolation energy distribution, directly decreasing signal efficiencies. The degradation in isolation performances motivates the study of new observables with potentially discriminating power between energy deposits created by pileup particles from those produced by fakes.

This chapter explores new observables related to the topoclusters, used to estimate the isolation energy. After introducing the samples used in this study, Section 5.2 defines a set of cluster-level observables studied along the chapter, followed by a discussion on how these observables yield a classification of topoclusters into "pileup-like" and "fake-like" categories. Instead of feeding these inputs directly to a state-of-the-art machine learning algorithm, a different approach is followed to better understand the discrimination process. This has proven to be a very good decision, as it allows to promptly identify issues directly related with topocluster reconstruction or instrumental artefacts. Section 5.4 describes the methodology used to compute a pileup-likelihood weight, assigned to each topocluster and extracted from a multidimensional simultaneous fit, to be finally used in a proposal for a new, pileup resilient, calorimetric isolation variable.

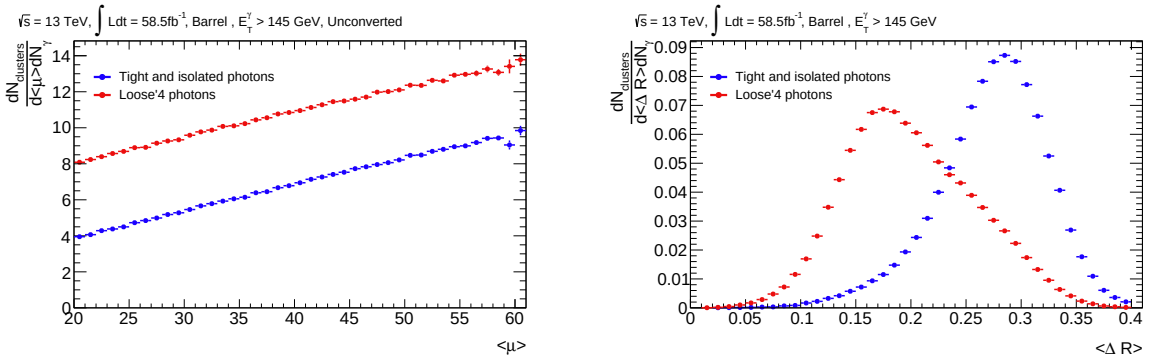
5.1 Topocluster samples

The γ-jet simulation samples and single photon triggered events in data described in the previous chapter are also used in this study (see Section 4.2.1) . In order to characterize topoclusters arising from pileup interactions, an additional selection is applied in order to build two different and orthogonal subsamples: clusters from a pileup-enriched subsample, and clusters from a fake-enriched subsample.

Pileup topoclusters originate from additional pp collisions, creating a uniform layer of energy deposits over the calorimeter. For prompt photons, expected to be isolated in the

detector, it is safe to assume that the majority of these energy deposits is uncorrelated with the photon itself. This means that for a prompt photon in a low pileup environment, few or none clusters would be found around it. The number of clusters found around a prompt photon increases to first order linearly with increasing $\langle\mu\rangle$ (see Figure 5.1a).

Fakes on the other hand are expected to have a larger energy flow around them, implying that a given fake is naturally surrounded by topoclusters, that carry a non-negligible fraction of the energy of the jet producing the photon fake. Thus, topoclusters from fakes are expected to be reconstructed closer to the photon candidate rather than far from it (see Figure 5.1b). These are to be found together with those arising from pileup interactions (see Figure 5.1a), with the difference that the former contribution does not depend on the average number of pileup interactions. These two behaviours are illustrated in Figure 5.1 and motivate the selection for clusters from pileup and clusters from fakes explained in the following.



a) b)

Figure 5.1: *(a) The average number of clusters per photon around tight and isolated (blue markers) and Loose'4 (red markers) photon candidates in data. (b) The average energy-weighted distance between topoclusters and photon candidates in data for two different selections: tight and isolated photons are shown in blue while Loose'4 photon candidates are shown in red.*

On photon candidates, a first selection is applied to make a prompt-enriched subsample and a fake-enriched subsample, from which topoclusters are selected depending on their angular distance to the photon candidate $\Delta R_{cl\gamma}$:

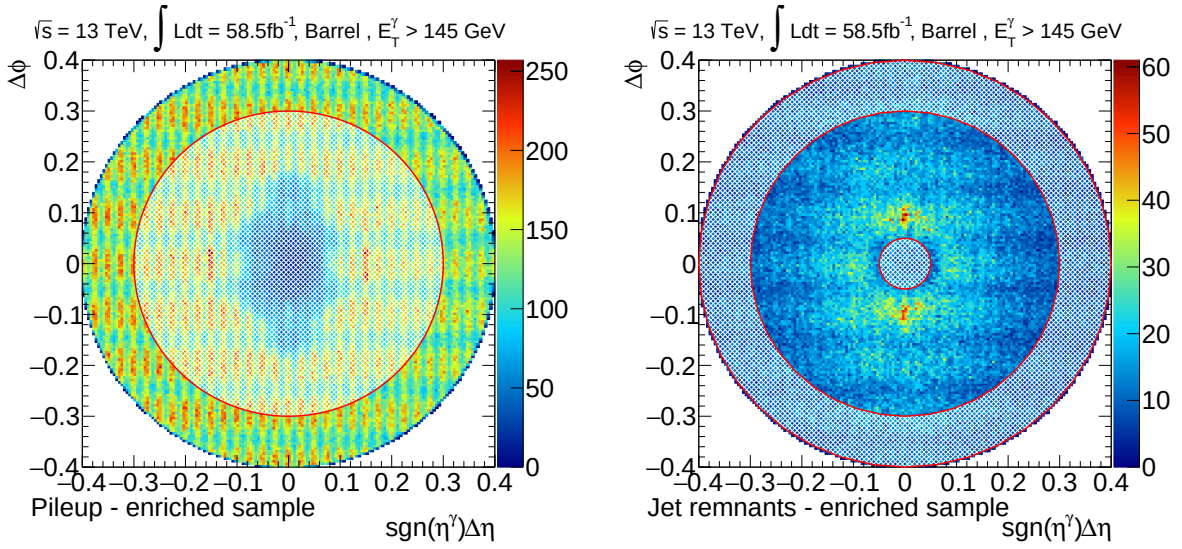
- A subsample enriched in topoclusters from pileup, is built from events with tight and isolated photon candidates, with topoclusters far from the photon candidate, $0.3 \leq \Delta R_{cl\gamma} < 0.4$ in high pileup environments ($\langle\mu\rangle > 40$). Photons are required to pass tight identification and isolation requirements to reduce the contribution of clusters from fakes. This sample is referred to in the following as the pileup-enriched subsample.

- Opposite to the previous requirements, the selection aiming to select topocluster from fakes is intended to reject prompt photons in order to select events enriched in topoclusters from fakes. Two subsamples enriched in topoclusters from fakes are built from events with photon candidates that fail either tight identification or isolation requirements. More precisely, photons are required to either pass Loose'4 identification or to fail the isolation criteria that makes use only of calorimetric information (`FixedTightCaloOnly`, see Table 4.1). The selected topoclusters are

those whose barycentre is located between $0.05 \leq \Delta R_{cl\gamma} < 0.3$ with respect to the photon candidate. Moreover, a low pileup environment ($\langle\mu\rangle < 25$) is chosen, in order to reduce the number of clusters from pileup inside the cone. These two samples are referred to in the following as the fake-enriched subsamples[1].

These selections are illustrated in Figure 5.2. These two-dimensional plots represent the spatial distribution of clusters around photon candidates in the $(sgn(\eta^\gamma)\Delta\eta, \Delta\phi)$ plane, where $\Delta\eta$ and $\Delta\phi$ are defined as $\eta_{cl} - \eta_\gamma$ and $\phi_{cl} - \phi_\gamma$ respectively, and $sgn(\eta^\gamma)$ is the sign of the photon pseudorapidity. This definition sets a reference for positive and negative values of $sgn(\eta^\gamma)\Delta\eta$, being positive for clusters located closer to the centre of the detector, and negative for topoclusters pointing towards the forward regions of the detector.

The figure on the left shows the spatial distribution of topoclusters from the pileup-enriched subsample. A homogeneous distribution of clusters is observed (the regular oscillations illustrate the granularity of the detector). The figure on the right shows the spatial distribution of topoclusters from the fake-enriched subsample. A slightly larger topocluster density is observed around the photon candidate, mostly along $|\Delta\phi|$, probably arising from charged particles bent by the magnetic field.



a) b)

Figure 5.2: *Spatial distribution of topoclusters in the $(sgn(\eta^\gamma)\Delta\eta, \Delta\phi)$ plane around unconverted photon candidates with $E_T^\gamma > 145GeV$ in the barrel. (a) Topoclusters from the pileup-enriched subsample are selected on the outer ring at $0.3 < \Delta R < 0.4$ (b) Topoclusters from the fake-enriched subsample are selected on the central ring at $0.05 < \Delta R < 0.3$. Hatched regions denote the regions **excluded** from both pileup-enriched (left) and fake-enriched (right) topocluster subsamples.*

The studies described in this Chapter are almost exclusively data-driven and performed with high energy photons ($E_T^\gamma > 145$ GeV) to benefit from the large statistics of the unprescaled trigger `HLT_g140_loose` and better purities compared to lower E_T^γ photons. Moreover, this is performed to avoid effects arising from pileup mismodelling in the simulation (discussed in Chapter 4), in which the calorimetric isolation energy is shown

[1]Distinction will be made when necessary but, as it will be shown, both samples share almost identical features.

to be slightly overestimated compared to the one observed in data. Thus, simulation samples are mainly used for validation purposes.

To avoid conflicts with the performance of the supercluster reconstruction algorithm (see Section 4.1), each cluster within the cone around the photon candidate is considered as an individual cluster.

5.2 Cluster-level observables

A cluster-level observable is defined in the context of this Chapter as an observable extracted from reconstructed topoclusters in the calorimeters. The transverse energy E_T^{γ} of each cluster as well as its $\Delta R_{cl\gamma}$ angular separation in the calorimeter relative to the photon are cluster-level observables already introduced and used for calorimetric photon isolation.

As explained in Section 3.1.2.2, topoclusters are three-dimensional objects made up of cells from possibly all layers of both hadronic and EM calorimeters. In the same section, the EM fraction, f_{EM}, of a topocluster is defined to select energy deposits likely left by photons. Using an analogous strategy, the same variable can also be used to discriminate energy deposits left from pileup from those arising from hadrons around fakes.

Figure 5.3 shows the distribution of f_{EM} for topoclusters from both pileup-enriched and fake-enriched subsamples. Both distributions present similar features, with two peaks: one at $f_{EM} = 0$ and $f_{EM} = 1$; and a continuous distribution in between. The distribution for both samples can then be separated in 4 independent categories:

- Topoclusters with $f_{EM} = 1$, implying that they are reconstructed exclusively from energy deposits in the EM calorimeter. It corresponds to the majority of clusters (80%) found around photon candidates. This category is referred to in the following as **pure EM clusters**.

- Topoclusters with $f_{EM} = 0$, implying that they are reconstructed exclusively from energy deposits in the hadronic calorimeter. It is the second largest contribution to clusters found around photon candidates. This category is referred to in the following as **pure Had clusters**.

- Topoclusters with $0.5 \leq f_{EM} < 1$, implying that at least 50% of the energy is reconstructed from cells found in the EM calorimeter. This category is referred to in the following as **mainly EM clusters**.

- Topoclusters with $0 < f_{EM} \leq 0.5$, implying that at least 50% of the energy is reconstructed from cells found in the hadronic calorimeter. This category is referred to in the following as **mainly Had clusters**.

A powerful cluster-level information is the timing. As a collection of cells, each topocluster is assigned a time, t_{cluster}, computed as an energy squared-weighted average:

$$t_{\text{cluster}} = \frac{\sum_{i}^{N_{\text{cells}}} t_i E_i^2}{\sum_{i}^{N_{\text{cells}}} E_i^2}, \tag{5.1}$$

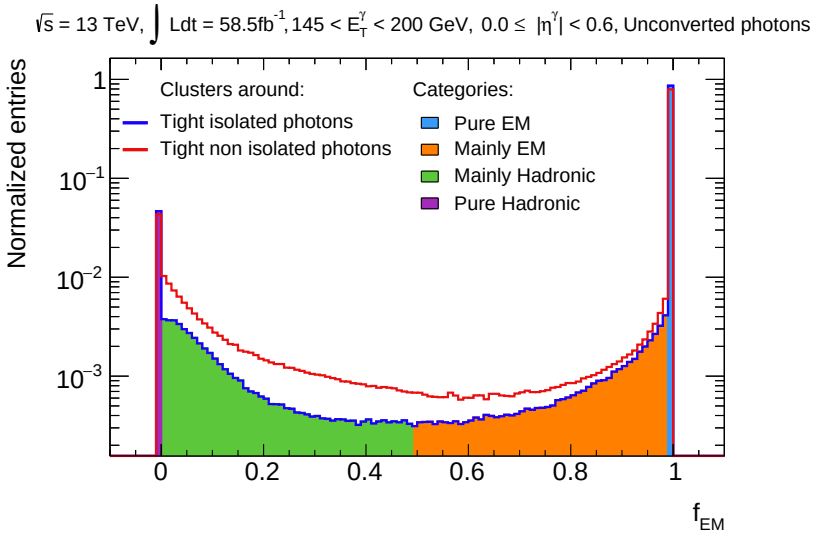


Figure 5.3: *The electromagnetic energy fraction f_{EM} for topoclusters from pileup-enriched (blue line) and fake-enriched (red line) subsamples. The different topocluster categories are also shown:* ***pure EM*** *in light blue,* ***mainly EM*** *in orange ,* ***mainly Had*** *in green and* ***pure Had*** *y purple.*

where t_i an E_i are the time and energy of the cell i, measured with the Optimal Filtering Method (Section 3.1.1)[2]. The difference in time with respect to the topoclusters found in the isolation cone is defined as:

$$\Delta t = t_{cluster} - t_{\gamma}. \tag{5.2}$$

The distribution of Δt is shown in Figure 5.4 for data and simulation. Two peaks from the previous and following bunch crossings can be identified a few tenths of nanoseconds before/after the photon time. A narrow peak at $\Delta t = 0$ corresponds to clusters from the same bunch-crossing; the peak is wider in data than in simulation. Since all these topoclusters contribute to the isolation energy (see Equation 4.2), a non-negligible fraction of these clusters are not related to the photon candidate and hence degrade the isolation performances.

5.3 Topocluster categories

The performance of calorimetric energy is η-dependent and differs significantly between the EM and the hadronic calorimeter. The energy resolution is, in general, largely superior in the EM calorimeter compared to the hadronic one. This is reflected in the per-cell time resolutions, which are for the LAr calorimeter better than 1 ns for all energies and gains, while for the Tile calorimeter the resolution spans from a few nanoseconds for low energy deposits and dominated by constant term of 0.5 ns beyond ~ 20 GeV. [3]

These differences motivate the separate study of the barrel and endcap as well as a categorization distinguishing both EM and hadronic calorimeters. Topoclusters are first classified according to their EM fraction, using the same categories identified in Figure 5.3.

[2]The time of each cell is calibrated in a way that a neutrino travelling from the interaction point would give a signal measured at $t = 0$ ns.

[3]These values are orientative, since they vary significantly between different energy gains.

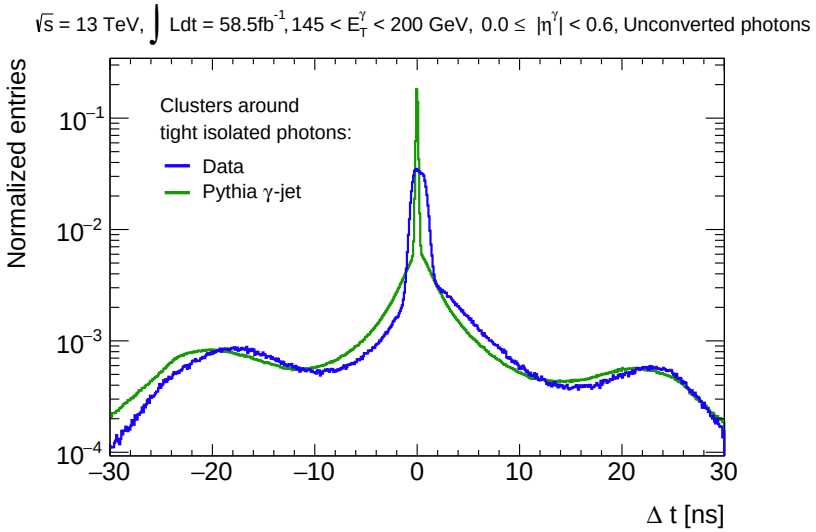


Figure 5.4: *The time difference Δt between topoclusters and photon candidates from tight, isolated photon in data and in simulation.*

This classification (detailed in Table 5.1) allows to identify regions in the $\left(\Delta t, E_T^\gamma\right)$ plane which are dominated by either topoclusters from pileup or from fakes. For the following discussion, only unconverted photons in the barrel are considered. Additional plots can be found in Appendix C.

Table 5.1: *Binning and categories used to characterize the topocluster subsamples.*

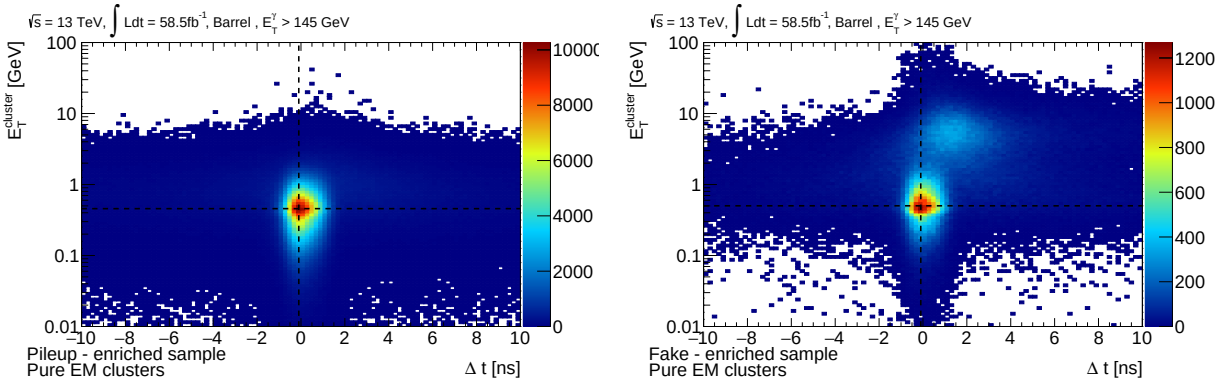
	Binning/Categories
Subsamples	Pileup-enriched, Fake-enriched (Loose'4), Fake-enriched (Tight,non isolated)
$\lvert\eta^\gamma\rvert$	Barrel: $[0.0, 1.37)$, Endcap: $(1.52, 181)$
f_{EM}	Pure EM, Mainly EM, Mainly Had, Pure Had
Conversion status	Unconverted, converted

5.3.1 Pure EM topoclusters

Pure EM topoclusters are characterized by a clear contribution of topoclusters in time ($\Delta t \sim 0$) with the photon candidate and transverse energies in the range of a few hundreds of MeV (see Figure 5.5). This contribution is present in both fake-enriched and pileup-enriched samples, since the fake-enriched samples are polluted with some clusters from pileup. In the fake-enriched samples, an additional wider local maximum can be identified at slightly higher transverse energies (~ 3 GeV). In conclusion, **pure EM** topoclusters are dominated by soft, in-time pileup.

5.3.2 Mainly EM topoclusters

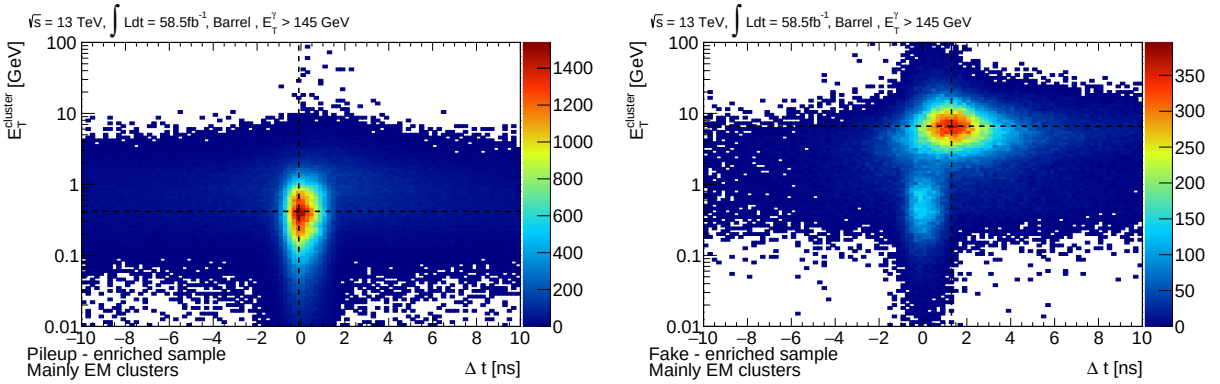
Mainly EM topoclusters present larger differences between the pileup-enriched and fake-enriched subsamples. The pileup-enriched topocluster sample (Figure 5.6a) is characterized similarly to the **pure EM** category, dominated by soft and in-time topoclusters. The fake-enriched topocluster sample is clearly dominated by harder clusters, reaching up to tens of GeV (Figure 5.6b). A small delay of 1-2 ns with respect to the photon



a) *Pileup enriched* b) *Fakes enriched (LP4)*

Figure 5.5: *Bidimensional distributions of* $\left(\Delta t, E_T^\gamma\right)$ *for* ***pure EM*** *topoclusters around unconverted photon candidates for the two subsamples defined in Table 5.1: pileup-enriched (left) and fake-enriched around Loose'4 photon candidates (right) in the barrel.*

candidate is also observed. In conclusion, the mainly EM topoclusters are a mixture of both soft, in-time pileup and clusters from fakes.

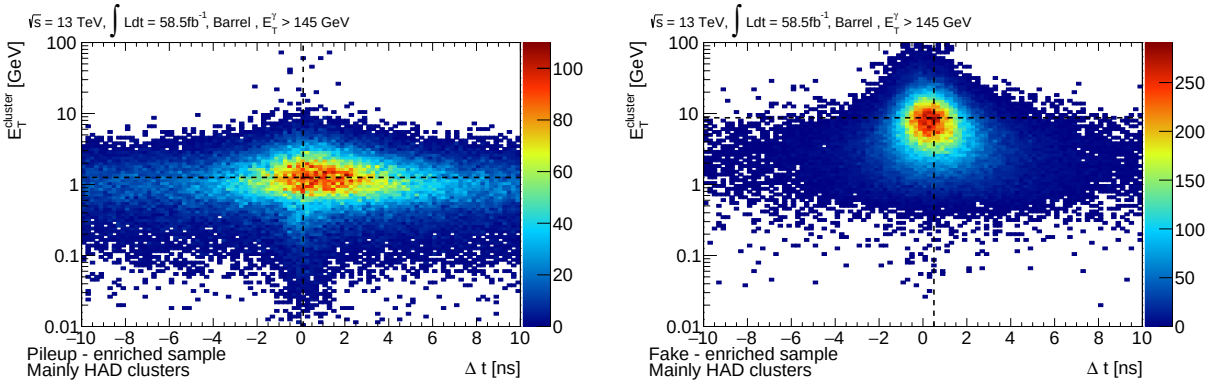


a) *Pileup enriched* b) *Fakes enriched (LP4)*

Figure 5.6: *Bidimensional distributions of* $\left(\Delta t, E_T^\gamma\right)$ *for* ***mainly EM*** *topoclusters around unconverted photon candidates for the different the two enriched subsamples defined in Table 5.1: pileup-enriched (left) and fake-enriched around Loose'4 photon candidates (right) in the barrel.*

5.3.3 Mainly Had topoclusters

Topoclusters with energy deposits mostly in the hadronic calorimeter can be characterized analogously, with the difference that worse energy and time resolutions widen the observed distributions. **Mainly Had** topoclusters in the pileup-enriched subsample are also characterized by larger transverse energy deposits compared to those observed in the **mainly EM** and **pure EM** categories. Similar features are observed in the fake-enriched subsamples to those observed in the **mainly EM** category: a localized contribution from clusters with approximately 10 GeV . (see Figure 5.7).

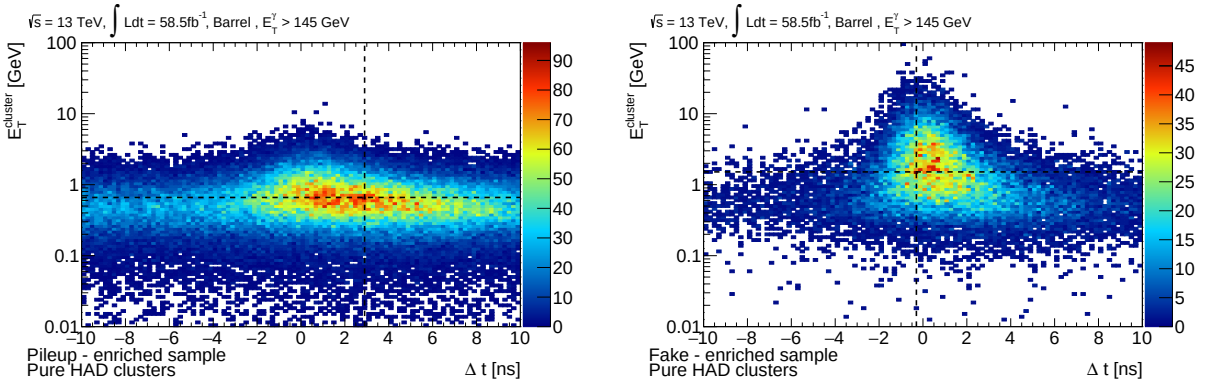


a) *Pileup enriched* b) *Fakes enriched (LP4)*

Figure 5.7: *Bidimensional distributions of* $\left(\Delta t, E_T^\gamma\right)$ *for* ***mainly Had*** *topoclusters around unconverted photon candidates for the two subsamples defined in Table 5.1: pileup-enriched (left) and fake-enriched around Loose'4 photon candidates (right) in the barrel.*

5.3.4 Pure Had topoclusters

In the **pure Had** category, topoclusters in the pileup-enriched subsample are spread around various tenths of ns in time with respect to the photon, and contain on average less than 1 GeV of transverse energy. Slightly harder topoclusters are found in the fake-enriched category.



a) *Pileup enriched* b) *Fakes enriched (LP4)*

Figure 5.8: *Bidimensional distributions of* $\left(\Delta t, E_T^\gamma\right)$ *for* ***pure Had*** *topoclusters around unconverted photon candidates for the different the two enriched subsamples defined in Table 5.1: pileup-enriched (left) and fake-enriched around Loose'4 photon candidates (right) in the barrel.*

A set of similar distributions for the endcap region can be found in Appendix C. When performing the same analysis in the endcap region, unexpected features were observed in the topocluster $(sgn(\eta^\gamma)\Delta\eta, \Delta\phi)$ distributions along photon candidates. These structures are very likely due to charge sharing from the photon candidate energy deposits. This observation was an unexpected output of this analysis, and led to a series of additional studies and discussions within the LAr community. More details are given in Appendix D.

5.3.5 Summary

Topoclusters around fake photons show properties different to those observed for reconstructed topoclusters from pileup objects. The former are in general harder than those expected from pileup, and at least some fraction of their energy is deposited in the hadronic calorimeter. Differences in the measured time of the topoclusters with respect to the photon candidate are also observed, being mostly in-time for pileup topoclusters and slightly delayed for those surrounding fakes. This discrimination is exploited in the following sections to evaluate the likelihood for a topocluster to come from pileup or from a fake.

5.4 Topocluster characterization

The two-dimensional distributions shown in the previous Section are parametrized by means of a simultaneous fit to the three enriched subsamples. The model describing the distributions is built with PDFs that are identified with regions populated by topoclusters from pileup or from fakes. This distinction is motivated by the aim of defining an observable sensitive to the likelihood of a topocluster to originate from a fake or from pileup. This section describes the procedure followed to perform the simultaneous fit.

5.4.1 Topocluster model

The distributions shown in the $\left(\Delta t, E_T^\gamma\right)$ plane are fit with bidimensional PDFs. The fit is performed by maximizing the joint likelihood of individual likelihoods built with three topocluster subsamples: the pileup-enriched subsample and the two fake-enriched subsamples:

$$\mathcal{L}_{\text{joint}}(\boldsymbol{\theta}) = \prod_{p}^{N_{\text{pileup}}} P(X_p|\boldsymbol{\theta}) \prod_{j}^{N_{\text{LP4}}} P(X_j|\boldsymbol{\theta}) \prod_{i}^{N_{\text{non-iso}}} P(X_i|\boldsymbol{\theta}), \tag{5.3}$$

where $\boldsymbol{\theta}$ is a set of analytical parameters of the model, P denotes the model with which the individual likelihoods are evaluated, and N_{sample} are the number of topoclusters found in each subsample.

The model P is based on the description of Figure 5.9 with 3 different species: one describes the contribution of soft topoclusters that are in-time with the photon; a second one describes a harder topocluster contribution with a larger spread in time ("out-of-time"); and the third one describes the hardest contribution of topoclusters, likely originating from fakes. Each species is described by 2-dimensional bifurcated Gaussian distributions[4], G, in the $\left(\Delta t, E_T^\gamma\right)$ space[5]:

- Soft and in-time pileup contribution (sp):
  $P_{sp} = G(\Delta t; \mu_{\Delta t}, \sigma^L_{\Delta t}, \sigma^R_{\Delta t}) \times G(E_T^{\text{cluster}}; \mu_{pt}, \sigma^L_{pt}, \sigma^R_{pt})$.
- Harder and spread in time pileup contribution (hp):
  $P_{hp} = G(\Delta t; \mu_{\Delta t}, \sigma^L_{\Delta t}, \sigma^R_{\Delta t}) \times G(E_T^{\text{cluster}}; \mu_{pt}, \sigma^L_{pt}, \sigma^R_{pt})$.

[4] A bifurcated Gaussian is described by a mean value, μ, and two widths, σ^L and σ^R, where L and R stand for left and right sides.

[5] All correlations between the two variables $\left(\Delta t, E_T^\gamma\right)$ are neglected

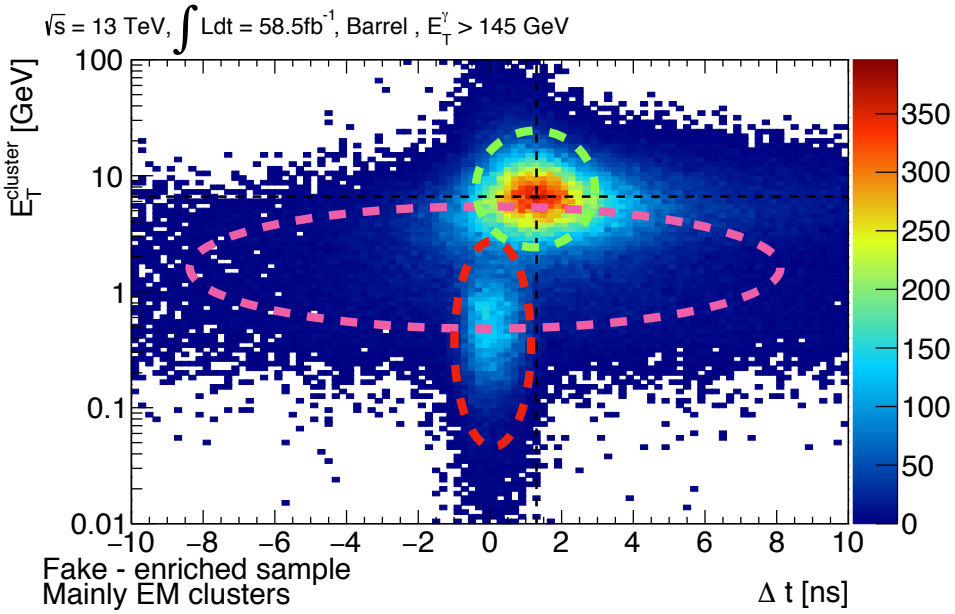


Figure 5.9: *Diagram showing the three different species defined for the simultaneous fit. The red dotted line represents the soft, in-time pileup species, the pink dotted line represents the hard, out-of-time pileup species and the green dotted line represents the topoclusters from fakes species.*

- Topoclusters from fakes contribution (fake):
  $P_{fake} = G(\Delta t; \mu_{\Delta t}, \sigma^L_{\Delta t}, \sigma^R_{\Delta t}) \times G(E_T^{\text{cluster}}; \mu_{pt}, \sigma^L_{pt}, \sigma^R_{pt})$.

where μ_i and $\sigma_i^{L,R}$ denote the mean and widths of each bifurcated Gaussian component, and are different for each P_i. The names describing the pileup contributions (soft and hard) are inspired on the shapes identified in the **mainly EM** category and play a different role in some of the categories, specially in the **mainly Had** and **pure Had** categories.

The full model is then defined as the normalized sum of the three species:

$$P(\Delta t, E_T^{\text{cluster}}) = f_{sp}P_{sp} + f_{hp}P_{hp} + f_{fake}P_{fake}, \tag{5.4}$$

where f_i corresponds to the relative fraction of each species present in the fitted sample, and $f_{sp} + f_{hp} + f_{fake} = 1$.

In all three topocluster subsamples (one pileup-enriched and two fake-enriched), the mean and the two widths of the bifurcated Gaussian distributions are common. The only non-common parameters between them are the relative fractions of them. For instance, a larger fraction of the soft-pileup species f_{sp} is expected in the pileup-enriched subsample compared to the one obtained in the fake-enriched subsample.

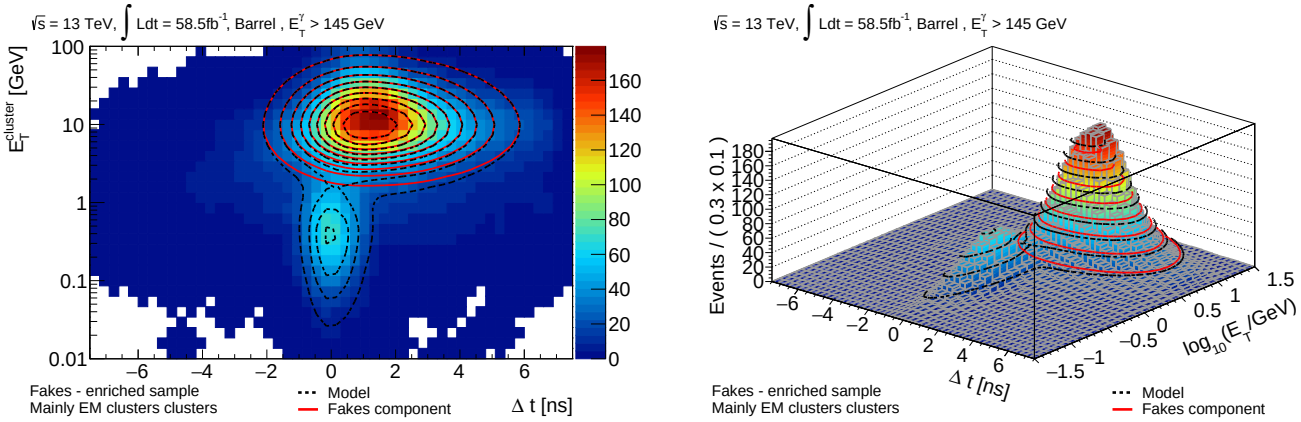
An example of 2-dimensional fit is shown in Figure 5.10 for the fake-enriched subsample. This fit illustrates how the simplified model is able to catch the contributions from pileup and fake clusters.

The parameters obtained with the simultaneous fits performed to each topocluster category are shown in Table 5.2. The pileup-enriched subsample is very pure, with barely a small contribution of topoclusters from fakes in the **pure Had** category (values in blue). The fake-enriched subsamples are all very similar (i.e. columns f_{LP4} and $f_{\text{non-iso}}$) and in general the largest contribution arises from topoclusters from fakes (values in red). However, a non-negligible fraction of pileup is present in some categories; for instance in the **pure EM** category, pileup remains the dominant contribution.

Table 5.2: *Measured parameters in each simultaneous fit to each topocluster category (**pure EM**, **mainly EM**, **mainly Had** and **pure Had**) and subsample (pileup-enriched and fake-enriched subsamples) in the barrel. The mean and the two widths of each species are also shown. The second table shows the fraction of each species (soft-pileup, hard-pileup and fakes) in each of the three subsamples (pileup-enriched subsample and two fakes-enriched subsamples). Values in blue highlight the small contribution of topoclusters from fakes in the pileup-enriched subsample, while values in red highlight the dominant contribution of topoclusters from fakes in the fake-enriched subsamples.*

Category	Species	$\mu_{\Delta t}[ns]$	$\sigma^L_{\Delta t}[ns]$	$\sigma^R_{\Delta t}[ns]$	μ_{pt} [GeV]	σ^L_{pt} [GeV]	σ^R_{pt} [GeV]
Pure EM	Soft-pileup	-0.063 ± 0.002	0.468 ± 0.002	0.673 ± 0.002	$0.554^{+0.001}_{-0.001}$	$2.730^{+0.005}_{-0.005}$	$0.216^{+0.000}_{-0.000}$
	Hard-pileup	1.504 ± 0.041	4.748 ± 0.039	3.767 ± 0.040	$1.039^{+0.006}_{-0.006}$	$2.835^{+0.012}_{-0.012}$	$0.254^{+0.001}_{-0.001}$
	Fakes	0.878 ± 0.024	1.300 ± 0.018	2.291 ± 0.019	$4.481^{+0.048}_{-0.047}$	$2.310^{+0.029}_{-0.028}$	$0.300^{+0.002}_{-0.002}$
Mainly EM	Soft-pileup	-0.055 ± 0.005	0.477 ± 0.004	0.677 ± 0.004	$0.486^{+0.004}_{-0.004}$	$3.421^{+0.024}_{-0.024}$	$0.272^{+0.002}_{-0.002}$
	Hard-pileup	1.661 ± 0.074	5.276 ± 0.078	3.938 ± 0.079	$0.942^{+0.011}_{-0.010}$	$2.682^{+0.020}_{-0.020}$	$0.306^{+0.002}_{-0.002}$
	Fakes	1.029 ± 0.025	1.455 ± 0.019	2.249 ± 0.019	$5.435^{+0.054}_{-0.054}$	$1.844^{+0.017}_{-0.016}$	$0.315^{+0.002}_{-0.002}$
Mainly Tile	Soft-pileup	0.050 ± 0.065	0.874 ± 0.074	1.222 ± 0.085	$1.981^{+0.116}_{-0.109}$	$5.148^{+0.252}_{-0.240}$	$0.211^{+0.009}_{-0.009}$
	Hard-pileup	2.542 ± 0.138	5.567 ± 0.142	3.439 ± 0.129	$1.454^{+0.017}_{-0.017}$	$2.452^{+0.022}_{-0.022}$	$0.285^{+0.002}_{-0.002}$
	Fakes	0.238 ± 0.023	1.230 ± 0.017	1.471 ± 0.019	$6.481^{+0.091}_{-0.090}$	$2.160^{+0.025}_{-0.024}$	$0.313^{+0.003}_{-0.003}$
Pure Tile	Soft-pileup	-7.489 ± 0.254	2.600 ± 4.339	2.734 ± 0.107	$0.726^{+0.015}_{-0.015}$	$3.340^{+0.115}_{-0.112}$	$0.061^{+0.001}_{-0.001}$
	Hard-pileup	2.352 ± 0.038	4.498 ± 0.036	4.702 ± 0.118	$0.739^{+0.003}_{-0.003}$	$2.594^{+0.012}_{-0.012}$	$0.254^{+0.001}_{-0.001}$
	Fakes	0.140 ± 0.025	1.180 ± 0.023	1.524 ± 0.029	$1.721^{+0.025}_{-0.025}$	$1.982^{+0.032}_{-0.032}$	$0.386^{+0.004}_{-0.004}$

Category	Species	Pileup-enriched [%]	Fakes-enriched [%]	Fakes-enriched (non-isolated photons)[%]
Pure EM	Soft-pileup	74.85 ± 0.05	39.11 ± 0.23	36.90 ± 0.28
	Hard-pileup	24.93 ± 0.05	15.75 ± 0.18	15.30 ± 0.21
	Fakes	0.22 ± 0.01	45.14 ± 0.24	47.80 ± 0.29
Mainly EM	Soft-pileup	66.15 ± 0.12	12.39 ± 0.22	10.85 ± 0.25
	Hard-pileup	33.73 ± 0.12	11.50 ± 0.21	11.03 ± 0.25
	Fakes	0.13 ± 0.01	76.11 ± 0.29	78.11 ± 0.33
Mainly Tile	Soft-pileup	14.93 ± 0.19	$< 0.01 \pm 0.00$	0.10 ± 0.03
	Hard-pileup	84.75 ± 0.19	14.61 ± 0.29	14.69 ± 0.36
	Fakes	0.32 ± 0.03	85.39 ± 0.29	85.20 ± 0.36
Pure Tile	Soft-pileup	5.89 ± 0.13	$< 0.01 \pm 0.00$	0.02 ± 0.03
	Hard-pileup	91.14 ± 0.16	41.82 ± 0.80	41.01 ± 0.96
	Fakes	2.97 ± 0.10	58.18 ± 0.80	58.97 ± 0.97



a) b)

Figure 5.10: *Bidimensional fit of the model (black dashed line) to the fake-enricheed subsample of mainly EM topoclusters. The red solid line denotes the fakes component of the model.*

5.4.2 Fake-likelihood weights

The PDFs P_i obtained from the previously described fits, are used to build a likelihood weight w, that denotes the probability for a topocluster to be originated from a fake. It is defined as follows:

$$w = \frac{P_{fake}}{P_{fake} + P_{pileup}}; \quad \text{with} \quad P_{pileup} = f P_{sp} + (1-f) P_{hp}, \tag{5.5}$$

where P_{pileup} is the joint pileup contribution and f is the soft pileup fraction. A topocluster with weight w close to 1 is very likely to originate from a fake, while a topocluster with w close to 0 is very likely to originate from pileup. The distribution of weights for the topoclusters found around tight and isolated photons and Loose'4 photons are shown in Figure 5.11.

5.5 Crosschecks and validations

In this section, a series of validations and tests are performed to study the behaviour of the topoclusters as a function of the weight w.

5.5.1 Data-driven characterization of pileup topoclusters

As mentioned in Section 5.1, the number of clusters around fakes is expected to be independent of the pileup $\langle\mu\rangle$ in the event, while the number of clusters from pileup per photon is expected to increase linearly with $\langle\mu\rangle$. This behaviour is illustrated in Figure 5.12, in which the number of clusters per photon with $w = 1$ shows a flat distribution independent of $\langle\mu\rangle$, coherent with the hypothesis that topoclusters with large weights originate from fakes; while the number of clusters per photon with $w = 0$ increases as a roughly linear function of $\langle\mu\rangle$, also consistent with the hypothesis that topoclusters with small w originate from pileup.

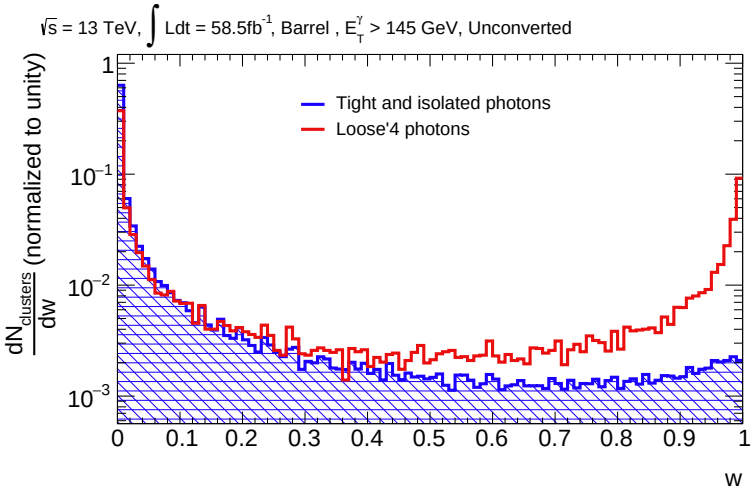
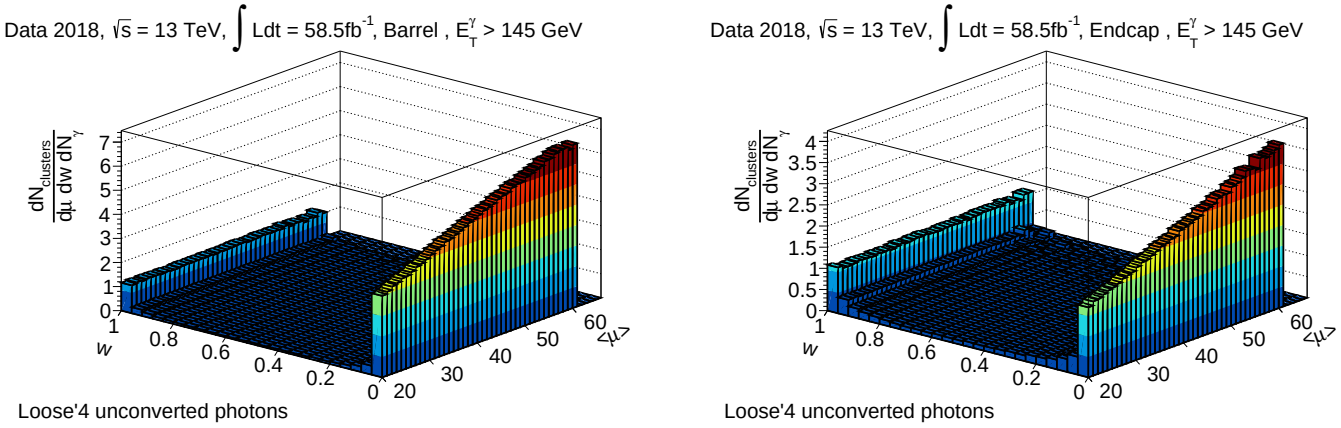


Figure 5.11: *The distributions of likelihood weights w for topoclusters around tight and isolated (blue) and Loose'4 (red) unconverted photons in the barrel.*



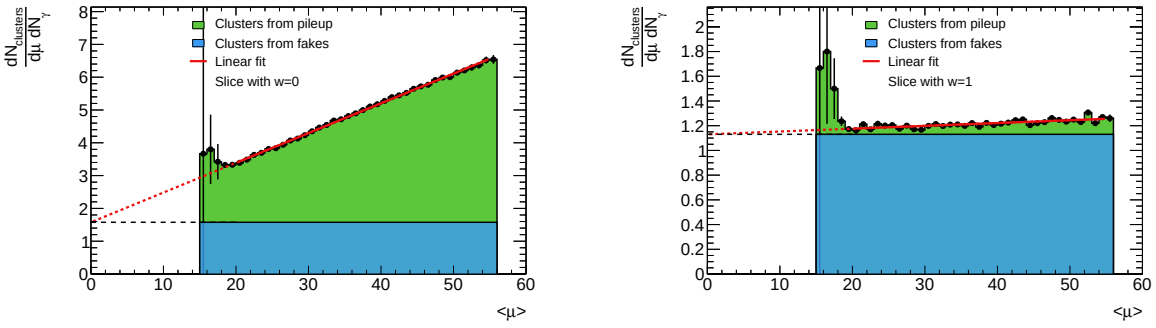
a) b)

Figure 5.12: *The average number of clusters per photon as a function of $\langle\mu\rangle$ and w. These two figures are produced with a subsample of unconverted photon candidates in the barrel, required to pass Loose'4 identification.*

The relation between the number of topoclusters arising from pileup or fakes as a function of $\langle\mu\rangle$ can be further exploited to validate the weights defined in Equation 5.5. This is done by studying the trend of the number of topoclusters per photon as a function of $\langle\mu\rangle$ for a given w. The parameters of a linear fit are interpreted as the relative contributions of topoclusters from pileup and topoclusters from fakes respectively.

To illustrate this in a pedagogical way, two limiting cases are considered in Figure 5.13: topoclusters with $w = 1$ and with $w = 0$. For topoclusters with large weights, $w = 1$, the number of topoclusters around photon candidates is almost constant with respect to $\langle\mu\rangle$. The additional topoclusters arising from the non-zero slope, are interpreted as a contribution of topoclusters from pileup. Analogously, topoclusters with weights close to $w = 0$ show an increasing linear trend with $\langle\mu\rangle$, and the extrapolation of this linear trend down to $\langle\mu\rangle = 0$ can be interpreted as the number of clusters per photon arising from fakes that are assigned a small weight. The measured slope for the latter case is

much larger, leading to the conclusion that the number of topoclusters with small weights increases faster with increasing $\langle\mu\rangle$ compared to those with larger weights.



a) b)

Figure 5.13: *The number of clusters per photon as a function of $\langle\mu\rangle$ for topoclusters with $w = 0$ (left) and with $w = 1$ (right). Both trends are fit with a linear function, and this fit is extrapolated down to $\langle\mu\rangle = 0$.*

Using this methodology, the contribution of clusters originated from fakes and pileup as a function of the weight w is evaluated in Figure 5.14. In fact, this figure shows the amount of topoclusters constant with $\langle\mu\rangle$, presumably from fakes, weighted in energy. This is motivated by the fact that topoclusters from fakes with small weights have usually small E_T^{cluster}, and hence barely contribute to the isolation energy measurement. The average contribution from topoclusters with $w = 0$ is observed to be 500 MeV, representing 7% of the energy contribution from topoclusters with $w = 1$. Under the hypothesis that the contribution of topoclusters that is independent of $\langle\mu\rangle$ arises from energy deposits around fakes, 7% of the total energy deposited by the fake is assigned a null weight.

The interplay between the slope and offset for each value of w can be used to estimate the amount of pileup interactions present on the simulated samples, since differences in the contribution from pileup in the calorimeters directly translate into larger/smaller fractions of pileup topoclusters with respect to that observed on data.

5.5.2 Spatial location of topoclusters from fakes

The number of clusters from pileup found at a given $\Delta R_{cl\gamma}$ is expected to increase with increasing $\Delta R_{cl\gamma}$. This is shown in Figure 5.15. Also, it is worth noticing that clusters originating from fakes are mostly located in a narrow region around $\Delta R_{cl\gamma} \sim 0.1$. This distribution of numbers of clusters per photon as a function of $\Delta R_{cl\gamma}$ is a completely data-driven extraction of the profile of isolation around fake photons. The sudden change in the shape at $\Delta R_{cl\gamma} \sim 0.07$ is related to the clustering algorithm, since energy deposits close to the photon candidate may be absorbed, and in those cases no additional topocluster is reconstructed.

As stated previously, additional topoclusters originating from pileup in the isolation cone dilute the discrimination power of the isolation energy between prompt photons and fakes. Any information rejecting such clusters, and in favour of topoclusters originated from fakes has the potential of improving isolation performances with increasing $\langle\mu\rangle$.

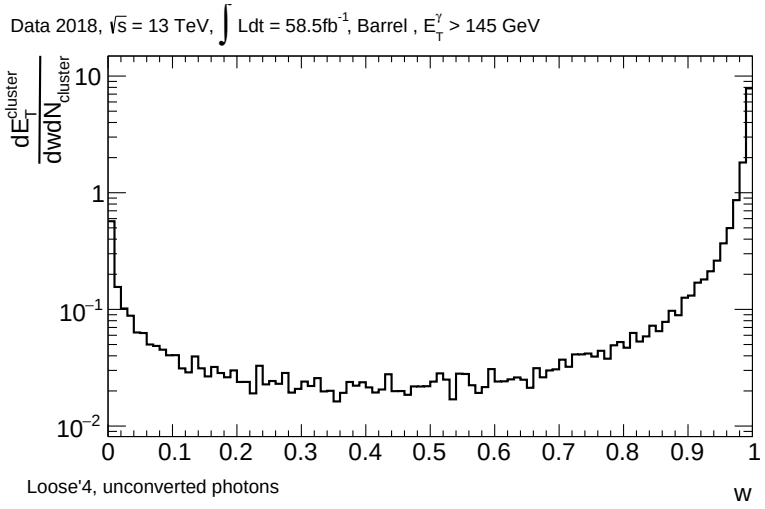


Figure 5.14: *Average energy contribution of topoclusters found in a cone around Loose'4 photon candidates as a function of w.*

5.6 Future prospectives

The methodology discussed in the previous sections, based on a data-driven extraction of a likelihood weight, illustrates how topoclusters can be effectively identified as "pileup-like", showing characteristics expected for pileup topoclusters from first principles. However, the information gathered in the previous sections can be used in many different ways with the aim of reducing the impact of busier pileup conditions on calorimeter isolation performances. In this section, a proposal for a new isolation variable, built from E_T-weighted topoclusters is presented as a study case.

5.6.1 A new proposal: weighted isolation energy

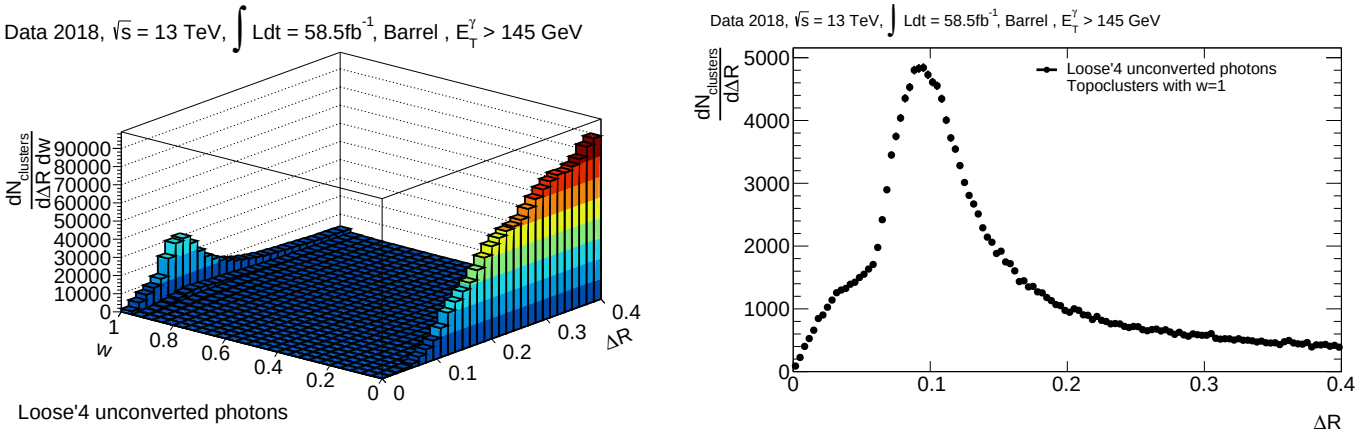
The *weighted isolation energy*, E_T^{iso-w}, is defined as the weighted sum of the transverse energy of the topoclusters found within the isolation cone. It is completely equivalent to the *nominal isolation energy*[6] (definition given in Section 4.2), $E_T^{nominal}$, if all topoclusters within the isolation cone had $w = 1$. The weights for each topocluster are obtained depending on the different categories described in the previous sections, effectively keeping in the sum those likely originating from a fake but reducing those likely coming from pileup. The weighted isolation variable is defined from the nominal isolation energy as follows:

$$E_T^{\text{nominal}} = \sum_{r_i<\Delta R}^{\text{clusters}} E_T^i \longrightarrow E_T^{\text{weighted}} = \sum_{r_i<\Delta R}^{\text{clusters}} w^i(|\eta|, f_{EM}, E_T^\gamma) E_T^i. \tag{5.6}$$

With the intention of providing a proper comparison, the energy of the photon candidate is subtracted identically in both variables by directly removing the topocluster of the photon.

A comparison of the shape for the nominal and weighted isolation variables is shown in Figure 5.16. The new isolation variable is significantly narrower than the nominal

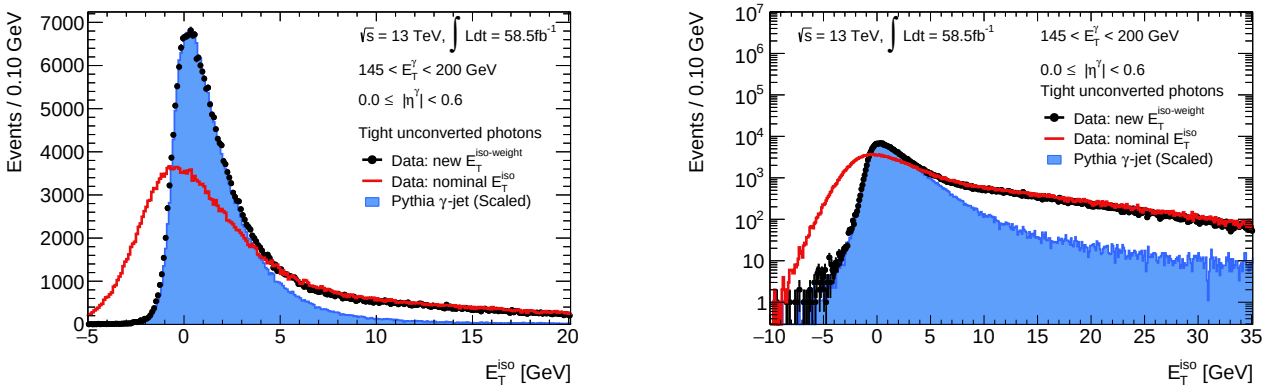
[6]The label "nominal" is included in this section to differentiate it from the weighted isolation energy



a) b)

Figure 5.15: *(a) The number of clusters as a function of ΔR and w around tight non isolated unconverted photon candidates in the barrel. (b) Slice of the histogram shown in (a) with $w = 1$, showing the density of topoclusters as a function of $\Delta R_{cl\gamma} \sim 0.1$.*

variable. The large isolation energy tail is almost identical for both variables in data, since topoclusters with larger transverse energies have by construction weights closer to unity.



a) b)

Figure 5.16: *The distributions of isolation energy for tight unconverted photons in the barrel with energies larger than $E_T^\gamma > 145$ GeV. The new isolation variable is shown for data (solid circles) and Pythia γ-jet simulation (blue solid histogram); to be compared with the nominal isolation variable for data (red histogram). Pileup is subtracted from the nominal variable, while the new proposal has no additional correction. No background is subtracted from data, which populates predominantly the large isolation energy tails.*

5.6.2 Pileup energy estimation

The aim of the studies shown in this chapter is to reduce the impact of pileup on photon isolation performances, with perspectives in future data-taking periods. To evaluate this, a comparison of the pileup mitigation achieved with this proposal with respect to the current pileup correction (described in Section 4.1) is shown in the following. The comparison

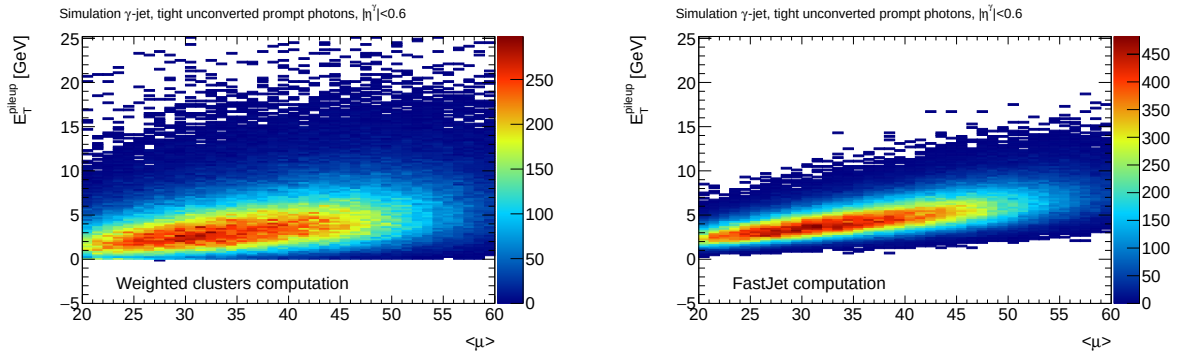
is performed on γ-jet simulated PYTHIA samples in which only prompt photons are selected. Eliminating the contribution of fragmentation photons, which may have non-negligible energy deposits around the photon, provides a clean estimation of the pileup energy in the isolation cone in the simulation.

The current pileup correction, from now on referred to as the FastJet correction, is estimated using as input all topoclusters in the calorimeters up to $|\eta^{\gamma}|<3.0$ of a given event. This allows to estimate the ambient energy density, which is correlated with the energy present in the isolation cone around a photon candidate. However, since all topoclusters are equally accounted for, the correction is barely sensitive to fluctuations in the number of clusters found around a photon candidate. The estimated energy from pileup using FastJet is defined in Equation 4.1, in Section 4.1.

The weighted isolation energy on the contrary evaluates independently each topocluster, reducing its contribution to the isolation energy depending on their properties. By construction, the weighted isolation energy is smaller than the nominal isolation energy without any corrections applied, since the topoclusters are either entirely considered ($w = 1$) or not ($w < 1$). The estimated energy from pileup in the weighted isolation energy is defined as follows:

$$E_T^{\text{pileup,w}} = \sum_{i}^{\text{clusters}} (1 - w^i(|\eta|, f_{EM}, E_T^{\gamma}))E_T^i. \tag{5.7}$$

Both distributions are shown in Figure 5.17 as a function of $\langle\mu\rangle$. A perfect correction would have exactly the same value as the isolation energy observed around each prompt photon candidate. The energy estimated with FastJet provides a robust estimation of the average energy arising from pileup in the calorimeter, and hence it is narrower compared to $E_T^{\text{pileup,w}}$. On the other hand, the weighted isolation energy is sensitive to per-photon fluctuations on the energy inside the isolation cone. Large energy deposits are not effectively accounted since they are assigned large weights.

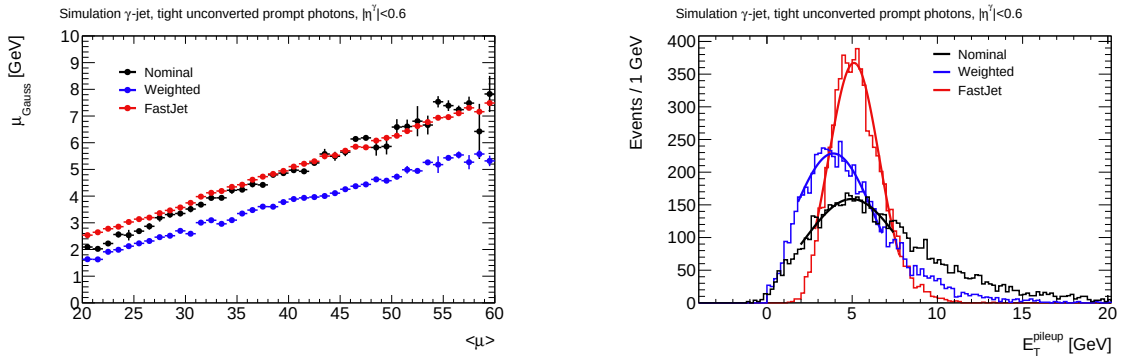


a) b)

Figure 5.17: *Bidimensional distributions of the pileup energy contribution in the isolation cone for prompt simulated photons. (a) Estimate using the weighted isolation energy computation (b) Estimate computed with FastJet.*

The pileup energy is then estimated using three observables: nominal isolation energy, the FastJet estimation and the estimation using the weighted isolation energy. The evolution of its most probable value (MPV) as a function of $\langle\mu\rangle$ is compared between the three samples in Figure 5.18a.

The MPV in the weighted pileup energy is smaller than the true pileup energy and the estimation provided by FastJet, which in turn are found to be in fair agreement. This is expected, since a perfect identification of pileup clusters would imply assigning integer weights (either 0 or 1) and by construction would lead to a subestimation of the pileup clusters, which in turn are considered with $w \neq 0$.



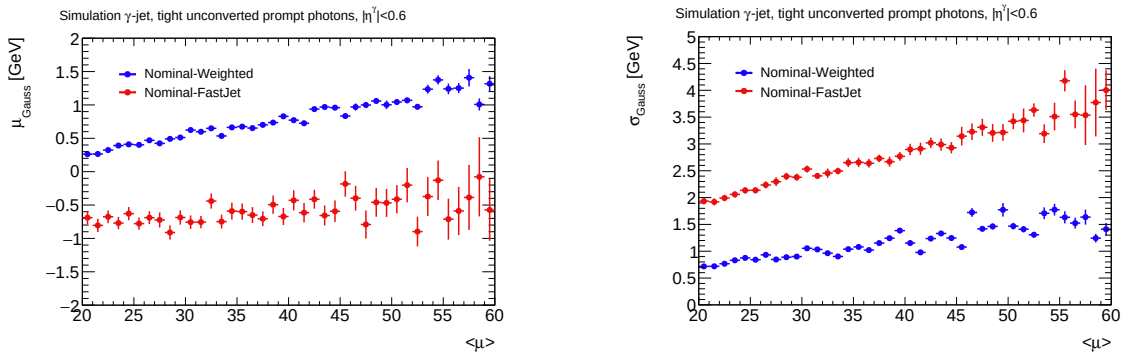
a) b)

Figure 5.18: *(a) MPV for the pileup energy for three different distributions computed for tight prompt simulated photons: true pileup energy (black,nominal in the label), weighted pileup energy (blue) and FastJet (red). (b) The distributions for the three pileup energy observables for prompt simulated photons.*

However, this comparison shows how different is the correction on average, when the strength of the weighted isolation variable arises precisely from the rejection of pileup clusters inside the isolation cone in a per-photon basis. While two photons would have similar pileup corrections estimated with FastJet (with a small difference arising from their $|\eta^{\gamma}|$ in the detector), the energy from pileup estimated with the weighting of topoclusters can vary greatly between photons in the same event.A direct comparison of the isolation energy variables is performed in the following.

5.6.3 Comparison of the isolation energy observables

For the sake of simplicity, the isolation energy distributions for the nominal and weighted variables are parametrized with their MPV, to avoid biases from the tails. The mean and the width are shown as a function of $\langle\mu\rangle$ in Figure 5.19a. The weighted isolation energy under-estimates the energy in the isolation cone for the aforementioned reasons. The pileup correction applied on the nominal isolation energy is on the contrary over-estimated, leading to a negative isolation energy. This effect was already observed in the previous Chapter, in particular in Figure 5.19b. More than a factor $\times 2$ of difference is observed between the widths measured in the nominal and weighted isolation energies. This improvement in the resolution arises from the correlation present in the pileup mitigation applied in the weighted isolation, since the pileup subtraction depends strictly on the same topoclusters used in the nominal isolation.



a) b)

Figure 5.19: *The MPV (left) and width (right) of the difference between the true pileup energy measured in prompt simulated photons and two estimates of it: weighted pileup energy (blue) and the one computed with FastJet (red).*

5.7 Conclusions

This new definition does not intend to be a direct replacement of the current calorimetric isolation variables, but introduces new features previously not used in photon isolation, such as timing information or the location of the clusters in the calorimeters. The results presented show that this alternative correction, exploiting the topoclusters found in the isolation cone, improves the resolution by more than a factor $\times 2$ compared to the current average pileup corrections.

A novel approach to estimate the pileup contribution to the isolation energy has been investigated in this chapter. Topoclusters, being the building blocks to estimate the isolation energy, have been shown to provide a larger discrimination compared to what is currently in use. The time and the EM fraction of the topoclusters are only a subset of the variables that have been discussed and provide some insight on the origin of the energy deposits in the isolation cone.

The study of variables related to the energy deposits around photon candidates has resulted in the definition of likelihood weights with sensitivity to discriminate pileup energy deposits from those relevant for isolation performances, related to fakes. Moreover, a more detailed study of the properties of such weights has led to a way of estimating the average contribution of energy in the isolation cone and the spatial distribution of the topoclusters originating from fakes.

Additional studies are required for a robust implementation of the work presented.

Chapter 6

Search for resonances with masses below 65 GeV in the diphoton channel

One of the primary goals of the experiments at the LHC is the search for any new phenomena, including unexpected ones, which manifest themselves in the high-energy regime accessible in hadron colliders. Diphoton final states, with a clean experimental signature and excellent invariant mass resolution, have previously been used in various New Physics searches, in particular, analyses looking for new phenomena in the form of narrow resonances over the invariant diphoton mass distribution. Both ATLAS and CMS experiments have presented results covering a wide range in the invariant diphoton mass, from $m_{\gamma\gamma} > 65$ GeV up to 2.5 TeV levels; no significant deviations with respect to the Standard Model predictions have been observed [125–127].

On the experimental side, going to lower masses poses a challenge on various aspects. The main issue is the unavoidable trigger E_T^{γ} thresholds, necessary to reduce the unmanageable rate of low energy diphotons. This naturally restricts the diphoton invariant mass phase space and sculpts its shape, that limits the lowest attainable mass, and is the main reason why previous analyses covered masses down to ∼65 GeV. This analysis overcomes the problem of the "turn on" of the trigger by selecting events with boosted photon pairs, recoiling usually against a hard jet. Requiring diphotons with large transverse momentum, $p_{\mathrm{T}}^{\gamma\gamma}$, flattens the background shape, by shifting the trigger-induced turn-on towards higher masses.

As most analyses in HEP nowadays, this analysis is first set up without looking at the signal region from data, what is called blinded. The main reason for this is to avoid biases from the human side. In this chapter the first expected result of diphoton resonance searches covering the region below 60 GeV using pp collisions with the ATLAS experiment is presented.

6.1 Analysis strategy

Resonance searches look for statistically significant deviations of the data with respect to a predefined model which supposedly describes SM processes. When no deviations are observed, upper limits are usually set on the production cross-section of a hypothetical non-SM signal. For this analysis, limits are set on the differential cross-section as a function of the invariant mass of a resonance. However, cross-sections (differential or total) can be strongly model dependent, since different theoretical assumptions can dramatically

modify the kinematics of a hypothetical signal. In order to avoid this model-dependency, limits are provided in a so-called fiducial phase space. The fiducial phase space is defined as the truth-level kinematic region that is directly accessible by the detector, minimizing the dependencies on the underlying model.

This strategy not only benefits from being more model-independent, but it also facilitates comparisons with theoretical models, since only the acceptance, defined as the fraction of the fiducial phase space with respect to the full phase space, is needed and can be computed with state-of-the-art MC generators. A diagram expressing the idea is shown in Figure 6.1.

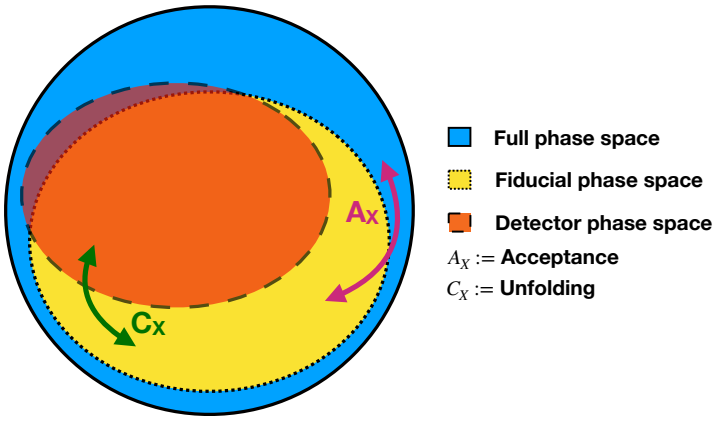


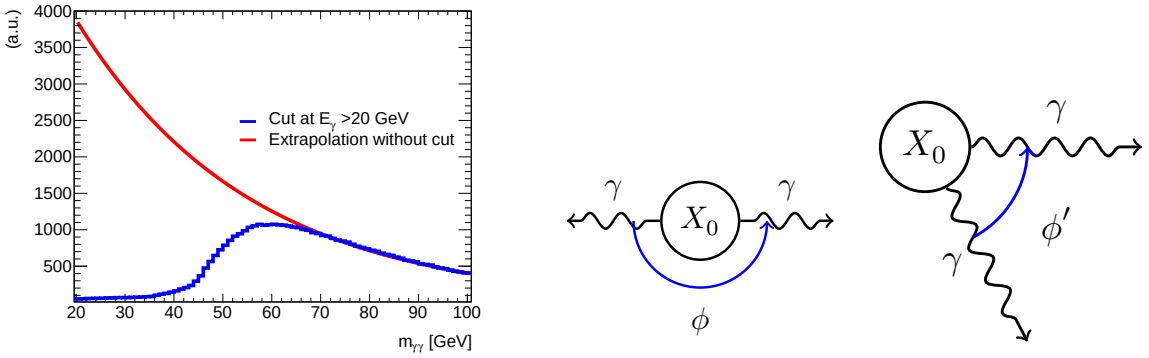
Figure 6.1: *Venn diagram showing the different phase spaces mentioned in the text. The total phase space is shown in blue and inside; it includes the fiducial phase space in yellow. The fiducial phase space is defined in such a way, that it overlaps as closely as possible with the detector phase space, in orange. Opposite to the situation between the total and fiducial phase spaces, some fraction of the detector level phase space may be composed by events not included in the fiducial phase space, due for example to resolution effects.*

A novel feature in this analysis is the diphoton selection. Even the loosest trigger requirements modify the diphoton invariant mass distribution, making its description with analytical functions complicated, due to steep changes in concavity (see Figure 6.2a). However, this fact does not forbid masses below twice the energy trigger requirements, since collimated photons can achieve much lower masses. A diphoton system is considered boosted if it has an important Lorentz boost with respect to the ATLAS detector rest frame. Lower masses are dominated by events showing this particular topology, depicted in Figure 6.2b. This analysis exploits this feature by selecting events with boosted diphotons, which in turn flattens out the diphoton invariant mass distribution.

This analysis uses data collected by the ATLAS experiment during Run 2, and simulated samples, all described in Section 6.2. Events are recorded in data using unprescaled triggers with at least two photons and fulfilling quality criteria during nominal operation of both ATLAS and the LHC. The selection of the events and the objects are described in Section 6.3, where particular emphasis on the novel diphoton boosted selection applied in the analysis is given.

Among the selected events, the diphoton invariant mass is used to discriminate known SM backgrounds from new physics resonances. The diphoton invariant mass is reconstructed as:

$$m_{\gamma\gamma} = \sqrt{2E_{T,1}^{\gamma}E_{T,2}^{\gamma}\left[\cosh\left(\eta_2-\eta_1\right)-\cos\left(\phi_2-\phi_1\right)\right]}. \tag{6.1}$$



a) b)

Figure 6.2: *Qualitative description of boosted topologies. (a) The invariant mass distribution for triggers requiring two photons with at least 20 GeV in blue, and a naive extrapolation of the same distribution without those cuts. The shape is significantly "sculpted". (b) A cartoon showing a two-body particle decay in the laboratory frame. On the left, the two photons appear back-to-back in the detector while on the right, the two photons appear collimated. In both topologies, the two photons have the same energy in the X_0 rest frame, $E^\gamma = m_{X_0}/2$. However, if the initial particle is boosted in a given direction of the detector, the energy of the photons could be large enough to pass the trigger requirements.*

Since the analysis strategy searches for excesses over the $m_{\gamma\gamma}$ distribution, models for the signal and background are built from simulation and control regions obtained from data. Sections 6.4 and 6.5 present the modelling of the signal and the background respectively.

The correction factor needed to unfold the detector effects and measure the fiducial cross-section is described in Section 6.6. Section 6.7 discusses the systematic uncertainties and section 6.8 describes the statistical framework used to extract the results. In section 6.10 the expected limits of the search and prospective studies of future Run 3 analyses are shown.

6.2 Simulation samples

Simulated samples are used to determine the shape of the diphoton mass spectrum for signal and SM diphoton background processes as well as signal efficiency and acceptance estimationss. All are simulated from *pp* collisions with a center-of-mass energy of 13 TeV. In this analysis, signal samples are generated for a scalar resonance with negligible width with respect to the calorimeter energy resolution. Both signal and diphoton background samples are then passed through a full GEANT4 simulation of the ATLAS detector.

The output of both simulations is then reconstructed using the same analysis chain used for collision data. Pileup is included in the simulation by overlying minimum bias events from inelastic proton-proton (*pp*) collisions, such that the average number of interactions per bunch-crossing reproduces the one observed in data.

Since this analysis targets a very small region of the available phase space, a short introduction on event filtering at various levels in the generation chain is given.

6.2.1 Event filtering in simulation

For several physics analyses, the accessible phase space with MC generators is very large compared to the targeted phase space. Moreover, restrictions are needed on the energy of some of the objects involved in the event to avoid divergences in cross-section computations. These restrictions are implemented at generation level in form of *filters*. A filter rejects events which do not fulfill certain kinematic requirements, like E_T^γ thresholds; or object criteria, like requiring 2 photons in the final state. In the context of this thesis, filters are applied at both parton and particle-level. Parton-level refers to the object characteristics (in simulation steps) before the parton showering takes place. Truth or particle-level refers to the information of the stable particles[1], produced as output of the MC simulation.

Parton-level filters are applied to avoid divergences and restrict the phase space from the first step in the generation chain. Generators like SHERPA, needing large amounts of CPU time to generate events, require in general a thorough study on the parton-level filters applied in order to optimize the CPU resources. Particle-level filters are commonly applied on generators for which the generation step is faster or parton-level filters. Of course, the speed at which a generator simulates events is process dependent, so these two statements may vary from case to case.

These two types of filters are used in the simulated samples used in this analysis.

6.2.2 Signal simulation samples

Signal samples are generated with MADGRAPH 2.6.5 and interfaced with PYTHIA8 with the A14 parameter tune and NNPDF2.3LO parton distribution functions set for parton-showering and hadronization simulation. These samples are used to characterize the shape of the signal diphoton invariant mass and measure the reconstruction efficiency. The production of a "Higgs boson-like" resonance that decays to two photons in association with 0, 1 or 2 jets in pp collisions at $\sqrt{s} = 13$ TeV is done for different masses in the [10 - 90] GeV range with a 5 GeV step. A Narrow Width Approximation (NWA) is assumed for all the signal samples (in practice a 4.07 MeV width is used). Samples are produced assuming a single production mode, gluon-fusion (ggF), since it is expected to be the dominant contribution compared to other production modes in the mass range of this analysis, namely $m_{\gamma\gamma}$<65 GeV (see Section 1.4.1). The interference between the $gg \to X \to \gamma\gamma$ process and the continuous QCD diphoton production associated to the $gg \to \gamma\gamma$ process is estimated to be small for the value of the width provided above [128], and it is neglected in the simulation. Since this analysis probes an unexplored range of the invariant diphoton mass, and the production of these samples was done specifically for this analysis, a detailed discussion on the generation is given in the following.

6.2.2.1 Effective field theory framework: scalar or pseudoscalar

Events are generated using an Effective Field Theory approach implemented in MADGRAPH within the Higgs Characterization HC framework [129]. Within this framework, both scalar and pseudoscalar resonances can be generated by including CP-even or CP-odd terms in the effective Lagrangian. The relevant interaction terms for this analysis are:

$$\mathcal{L}_{EFT} \supset c_\alpha \kappa_{H\gamma\gamma} g_{H\gamma\gamma} A_{\mu\nu} A^{\mu\nu} + s_\alpha \kappa_{A\gamma\gamma} g_{A\gamma\gamma} A_{\mu\nu} \tilde{A}^{\mu\nu} + c_\alpha \kappa_{Hgg} g_{Hgg} G^a_{\mu\nu} G^{a,\mu\nu} + s_\alpha \kappa_{Agg} g_{Agg} G^a_{\mu\nu} \tilde{G}^{a,\mu\nu}, \quad (6.2)$$

[1] A stable particle is defined as a particle with lifetime $c\tau_0 > 10$ mm and τ_0 the proper time.

where $H(A)$ stands for a scalar (pseudo-scalar) Higgs boson-like resonance that couples to photons $A^{\mu\nu}$ and gluons $G^{a,\mu\nu}$[2] with coupling strengths $g_{H\gamma\gamma}$ ($g_{A\gamma\gamma}$) and $g_{Hgg}(g_{Agg})$ respectively, c_α and s_α (abbreviations for $cos\alpha$ and $sin\alpha$) describe the mixing of the two CP states H and A, and the various κ_{Hii} (κ_{Aii}) take values of 0 or 1 to allow or forbid a particular coupling in the generation. In this way, by setting specific values for c_α (and therefore for s_α), a scalar, a pseudoscalar or a mixed state can be generated.

Prior to the generation of the samples, a comparison was made to evaluate possible particle-level differences. A comparison of both CP states is shown for various kinematic variables in Figure 6.3. As no significant differences are observed between them, and for the sake of simplicity, a scalar resonance is generated by setting $c_\alpha = 1$ (pure scalar); and $\kappa_{Hii} = 0$ except for gluons and photons.

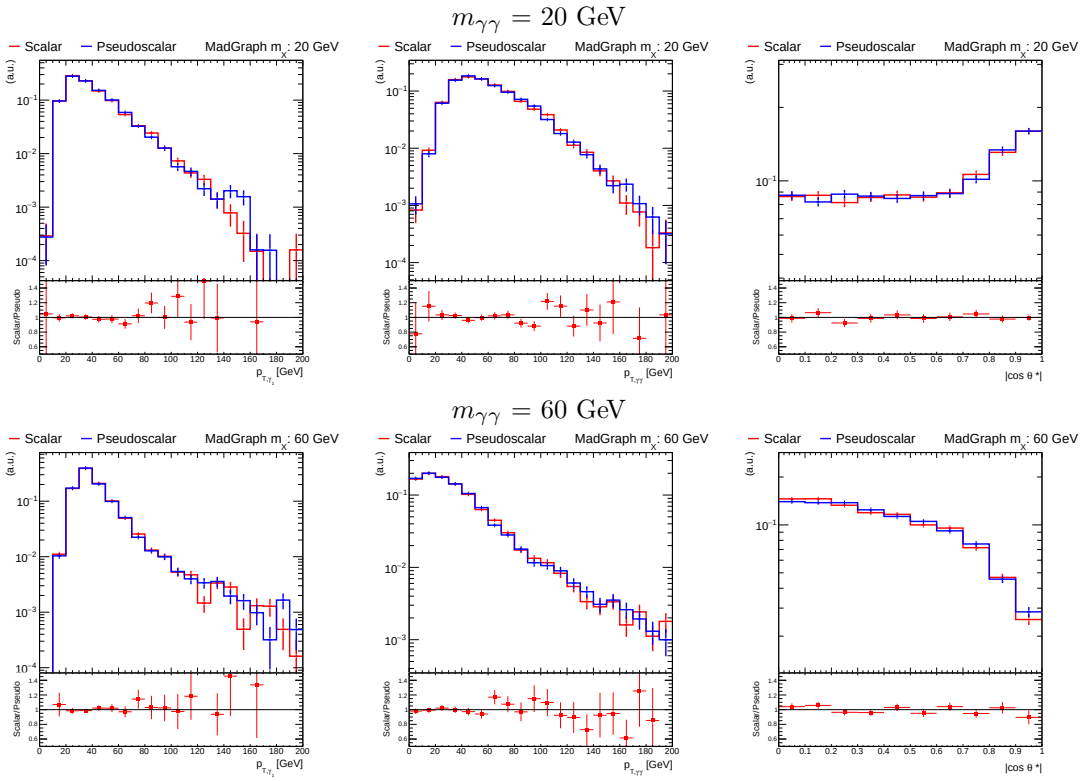


Figure 6.3: *Photon and diphoton kinematic variables from scalar and pseudoscalar resonances decaying to two photons + 0,1 jets without any additional filter for 20 (upper row) and 60 (lower row) GeV masses. From left to right: E_T^γ of the leading photon, $p_T^{\gamma\gamma}$ of the diphoton system and $cos\theta^*$ in the Collins-Sopper reference frame [130]. Changes in the distributions at different masses are due to differences in the generation phase space.*

6.2.2.2 Leading order or next-to-leading order

The signal samples used in this analysis are generated at LO with both parton and particle-level filters. The choice of a leading order generation is driven by the current filtering setup implemented in the HC model. This model can generate (among a wide variety of processes) Higgs boson-like resonances decaying to two photons, filtered out at parton-level if their transverse energy is below a certain threshold, being this threshold

[2]The dual tensors $\tilde{A}^{\mu\nu}$ and $\tilde{G}^{a,\mu\nu}$ use the same definition as in [129].

the same for both leading[3] and subleading photons. However, this setup with symmetric parton-level filtering is found to be unstable for masses m_X close to twice the value of this cut $\approx 2E_T^{\text{cut}}$ [131] leading to unphysical distributions due to large fractions of negative weights. Increasing the E_T^{γ} asymmetry of the photon filtering improves the quality of the generation, as shown in Figure 6.4, where two different behaviours are identified. At low masses with $m_X <\approx 2E_T^{\text{cut}}$, the resonance mass m_X is not sufficient to generate two photons passing the filter if the resonance is at rest, hence it is kinematically required that to recoil against a third object. This requirement is equivalent to degrading the generation from NLO to LO (only positive weights). When this kinematic barrier is overcome at larger masses, the fraction of negative weights increases to account for the NLO component.

Implementing an asymmetric filtering in the NLO setup implies significant technicalities, with mild or negligible impact in the signal generation, since at low masses the generation is mostly LO+jets. It was therefore decided to use a LO generation setup of a resonance in association with up to 2 jets.

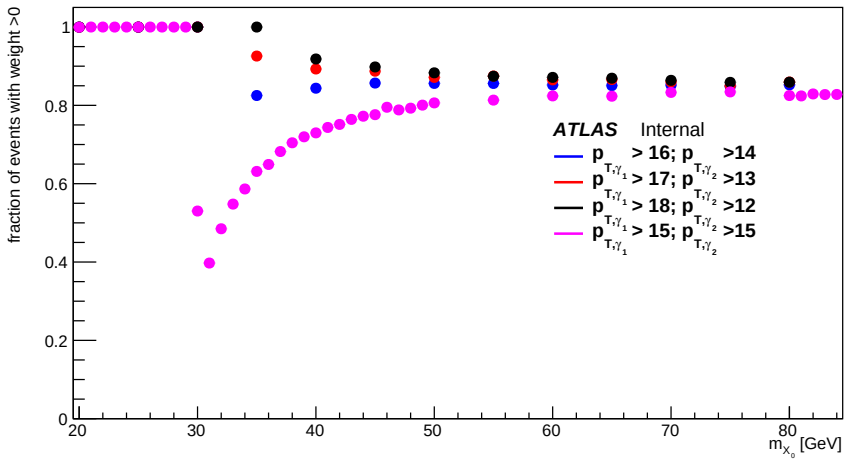


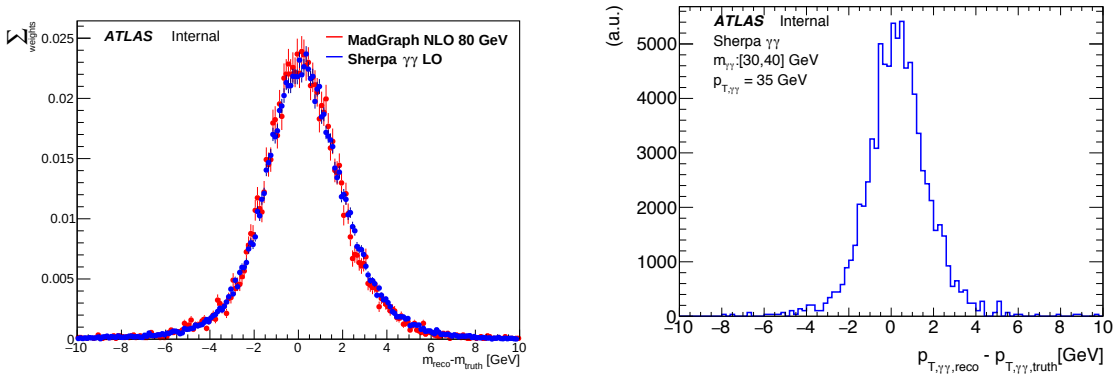
Figure 6.4: *Fraction of negative weights for several mass points generated with different filters on the transverse momentum of leading and subleading photon.*

6.2.2.3 Filtering at particle-level

Leading order generation is computationally much faster compared to the same process at NLO. For this reason, a loose filtering is applied at parton-level and then filtered at particle-level to increase the available number of events after reconstruction. The parton-level filter selects events with two photons with E_T^{γ}>2 GeV, with an angular separation of at least $\Delta R = 0.15$.

An implementation of a particle-level filter that selects photon pairs with large transverse momentum $p_{\text{T}}^{\gamma\gamma}$ was developed for this analysis. Since filtering before reconstruction rejects events, the bias caused by filtered-out events that would have otherwise been reconstructed needs to be estimated. The value of the bias of the filter is estimated from QCD diphoton simulated samples (detailed in Section 6.2.3), since the $m_{\gamma\gamma}$ resolution for signal and background diphotons is sensibly similar, as shown in Figure 6.5a.

[3]Leading and subleading photons in a diphoton pair, refer to the one with the largest and second largest transverse energy in the event.



a) b)

Figure 6.5: *(a) The detector resolution for a MadGraph signal point at 80 GeV and for diphoton events from the inclusive SHERPA $\gamma\gamma$ sample with $m_{\gamma\gamma}$:[78,82] GeV. (b) Diphoton $p_{\mathrm{T}}^{\gamma\gamma}$ resolution for the inclusive SHERPA $\gamma\gamma$ sample and diphoton pairs with $m_{\gamma\gamma}$:[30,40] GeV and $p_{\mathrm{T,particle}}^{\gamma\gamma}$> 35 GeV.*

The $p_{\mathrm{T}}^{\gamma\gamma}$ resolution is computed as a function of $m_{\gamma\gamma}$ for different $p_{\mathrm{T}}^{\gamma\gamma}$ cuts at particle-level. For each $p_{\mathrm{T}}^{\gamma\gamma}$ cut and $m_{\gamma\gamma}$ bin, the safety margin that keeps 95% of the signal events after reconstruction is computed. Figure 6.6 shows the needed safety margin in GeV, defined as the difference between reconstruction and particle level $p_{\mathrm{T,particle}}^{\gamma\gamma}$ from generation, that keeps the bias below 5% for each mass range and $p_{\mathrm{T,particle}}^{\gamma\gamma}$ cut. A 4 GeV safety margin appears to be sufficient, and thus photon pairs are required to satisfy $p_{\mathrm{T,particle}}^{\gamma\gamma}$> 40 GeV at particle level, to ensure a negligible effect after reconstruction.

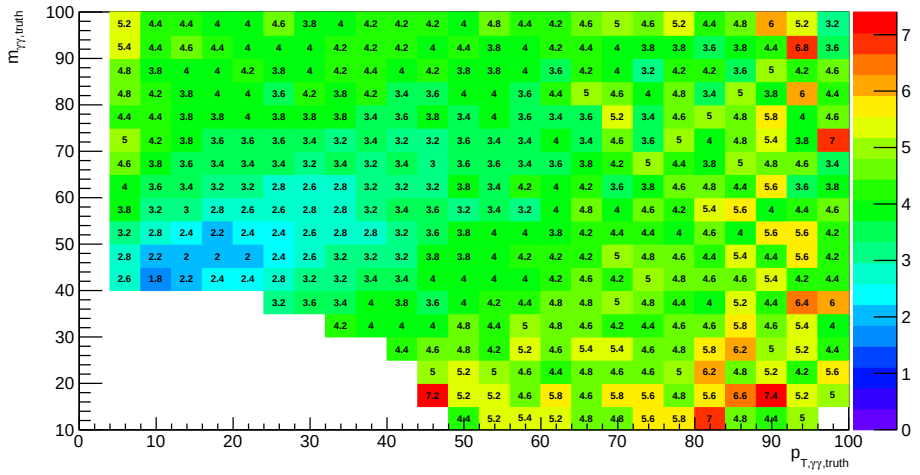


Figure 6.6: *Bidimensional scan showing the needed safety margins to avoid biases after applying $p_{\mathrm{T,particle}}^{\gamma\gamma}$ filters for each mass after requiring two photons of 20 GeV. Lower masses are almost exclusively generated with large $p_{\mathrm{T,particle}}^{\gamma\gamma}$. A 40 GeV filter at particle level with a final selection of at least 4-5 GeV difference keeps biases below 5%.*

6.2.3 Background samples

Background events from continuum $\gamma\gamma$ production are generated using SHERPA 2.2.4 [132, 133] and the SHERPA default tuning for the underlying event. The matrix elements are calculated at next-to-leading order (NLO) in the strong coupling constant α_s for up to one real emission of an additional parton and at leading order (LO) for two and three additional partons, and are merged with the SHERPA parton shower [134] according to the ME+PS@NLO prescription [135, 136]. The PDF set used is the NNPDF3.0 NNLO [28].

Two SHERPA diphoton samples are generated with different configurations. The first sample consists on a generation of photon pairs with invariant masses from 0 to 90 GeV, divided in two slices[4]. The first slice covers from 0 to 50 GeV and the second one from 50 to 90 GeV. The generation requires two photons with transverse energies larger than 20 GeV for the leading photon and 18 GeV for the subleading photon at parton-level. Isolation requirements at parton-level filter out photons which are closer than $\Delta R_{\gamma\gamma} = 0.2$. Unfortunately, the first slice is significantly more limited in statistics compared to the second slice. The generated events are simulated in both fast and full simulation. This sliced sample will be referred in the following as the inclusive SHERPA $\gamma\gamma$.

Table 6.1: *Prompt diphoton background samples.*

Generator	$m_{\gamma\gamma}$range [GeV]	Cross section [pb]	N_{events}
Inclusive Sherpa+Fullsim	0–50	93.5	3.9M
	50–90	139.0	173.6M
Inclusive Sherpa+Fastsim	0–50	93.5	3.9M
	50–90	139.0	361.2M
Boosted Sherpa+Fastsim	0–90	59.6	15.3M (+200M soon)

The second sample consists of a filtered diphoton generation with invariant masses from 0 to 90 GeV in a single slice. In order to optimize the usage of CPU resources and allow the production of sufficient statistics for the analysis, a filter at parton-level in the transverse momentum of the diphoton system is applied, requiring $p^{\gamma\gamma}_{\text{T,parton}}$>35 GeV. The same transverse energy per-photon filters are applied in this sample but the isolation requirement is relaxed, allowing photons to be close down to $\Delta R_{\gamma\gamma} = 0.15$. This sample will be referred in the following as the boosted SHERPA $\gamma\gamma$. A summary of the available statistics for both samples is shown in Table 6.1.

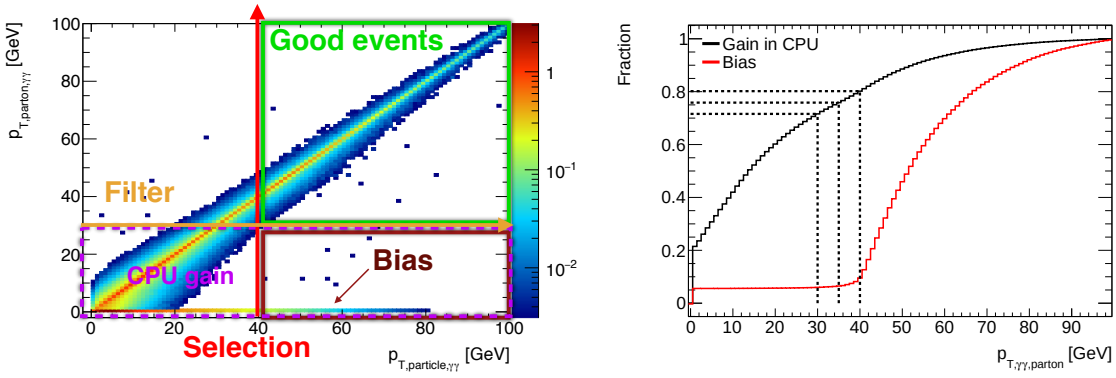
The inclusive SHERPA $\gamma\gamma$ sliced sample is statistically limited, for the selection of this analysis, compared to the recorded statistics in data. Moreover, the slicing may potentially create problems due to mismatches in the per-slice cross-sections [125]. The impact of the parton-level filter used in the filtered SHERPA $\gamma\gamma$ is evaluated in the following subsection.

6.2.3.1 Bias from $p^{\gamma\gamma}_{\text{T,parton}}$ filtering

A study was performed prior to the generation of this sample to evaluate the bias coming from events which would have passed the final selection but were not generated due to parton-level cuts.

[4]This sample existed before the start of this analysis and thus it was not generated accounting for the specific requirements of the search.

In this subsection, *bias* refers to the fraction of events that are not generated with a given parton-level filter, but that would have passed the particle-level selection; and *CPU gain* refers to the fraction of events that are not generated and would not have passed the final selection. The latter is not a proper estimation of the true CPU gain, but provides an idea on how efficient the filter is for the purposes of this analysis.



a) b)

Figure 6.7: *(a) The correlation between the initial $p_{\mathrm{T,parton}}^{\gamma\gamma}$ and final $p_{\mathrm{T,particle}}^{\gamma\gamma}$ after parton showering. The marked rectangles denote the different identified regions depending on the filter and final selection. (b) The expected bias and gain in CPU as a function of $p_{\mathrm{T,parton}}^{\gamma\gamma}$ for events with $m_{\gamma\gamma}:[0,90]$ GeV that pass $p_{\mathrm{T,particle}}^{\gamma\gamma}$>40 GeV. Increasing the cut at parton-level leads to larger biases.*

The final selection in this study is defined as $p_{\mathrm{T,particle}}^{\gamma\gamma}$>40 GeV, which is assumed to be a safe choice with respect to the intended offline selection at $p_{\mathrm{T}}^{\gamma\gamma}$>50 GeV, given that the $p_{\mathrm{T}}^{\gamma\gamma}$ resolution after reconstruction is about $\sim$3 GeV. The estimated bias for a final selection of $p_{\mathrm{T,particle}}^{\gamma\gamma}$>40 GeV is around 5% for filters below 40 GeV.

Figure 6.7 shows the correlation between $p_{\mathrm{T,parton}}^{\gamma\gamma}$ and $p_{\mathrm{T,particle}}^{\gamma\gamma}$, and the total bias expected as a function of $p_{\mathrm{T,parton}}^{\gamma\gamma}$. It is worth noticing the presence of a fraction of events with $p_{\mathrm{T,parton}}^{\gamma\gamma}$=0 GeV, which are then boosted via parton showering up to large $p_{\mathrm{T,particle}}^{\gamma\gamma}$ values. These events represent almost the totality of the bias.

The choice of the filter at $p_{\mathrm{T,parton}}^{\gamma\gamma}$>35 GeV is driven by the value that maximizes the gain in CPU before the bias grows to unmanageable values. The 5% observed bias is not homogeneously distributed across the different kinematic variables (see Figure 6.8), affecting the region above $m_{\gamma\gamma,\mathrm{particle}} = 40$ GeV, approximately twice the E_T^γ parton thresholds. It is understood to arise from the NLO generation without recoiling jets, which naturally has $p_{\mathrm{T,parton}}^{\gamma\gamma}$=0. If an event generated with small $m_{\gamma\gamma}$ passes the per-photon E_T^γ thresholds, it is naturally boosted i.e because it recoils against a hard jet. However, events with larger $m_{\gamma\gamma}$ are not necessarily recoiling against a jet and only acquire larger $p_{\mathrm{T}}^{\gamma\gamma}$ after parton showering.

A reweighting procedure is performed at particle-level in order to recover the contribution from non-generated events (NLO without jets). A 2D-reweighting map (see Figure 6.9) in the ($p_{\mathrm{T,particle}}^{\gamma\gamma}$,$\Delta R_{\gamma\gamma}$) space is computed with the ratio of events passing the filter of $p_{\mathrm{T,parton}}^{\gamma\gamma}$>35 GeV and $p_{\mathrm{T,particle}}^{\gamma\gamma}$>40 GeV over the number of events required to pass $p_{\mathrm{T,particle}}^{\gamma\gamma}$>40 GeV.

A fair agreement is observed at particle-level between the inclusive SHERPA $\gamma\gamma$ sample and the reweighted boosted SHERPA $\gamma\gamma$ in most of the kinematic variables. While

seemingly large differences remain uncorrected in the high-end domain of the subleading photon transverse momentum, these differences are present in only a small fraction of the complete distribution, and have only a negligible impact on the shape of the diphoton mass spectrum.

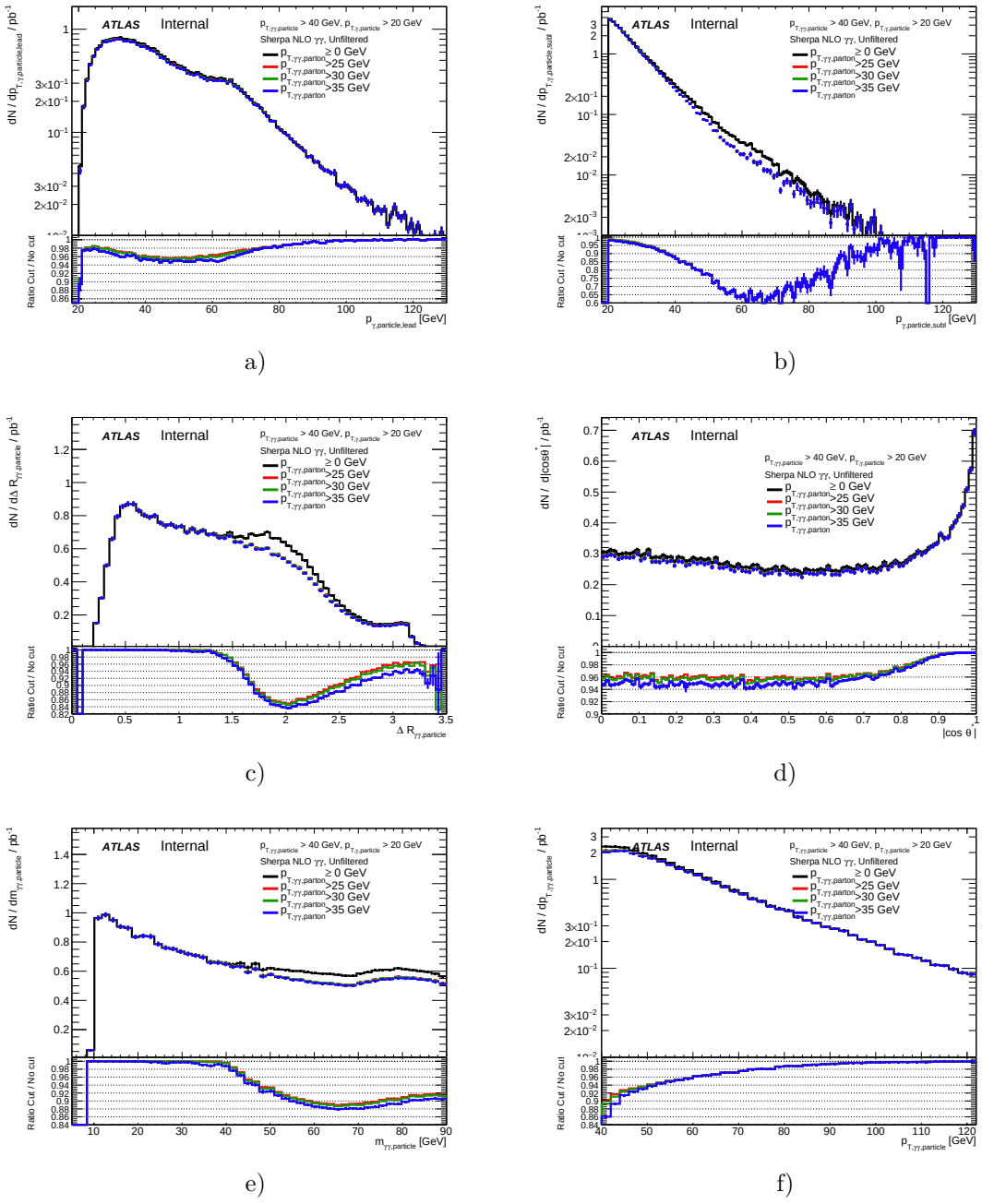


Figure 6.8: *The plots showing kinematic variable distributions before and after various* $p^{\gamma\gamma}_{\mathrm{T,parton}}$ *filters. (a)* $p_{T,particle,lead}$ *(b)* $p_{T,particle,subl}$ *(c)* $\Delta R_{\gamma\gamma,particle}$ *(d)* $|cos\theta^*|$ *(e)* $m_{\gamma\gamma,particle}$ *(e)* $p^{\gamma\gamma}_{\mathrm{T,particle}}$*.*

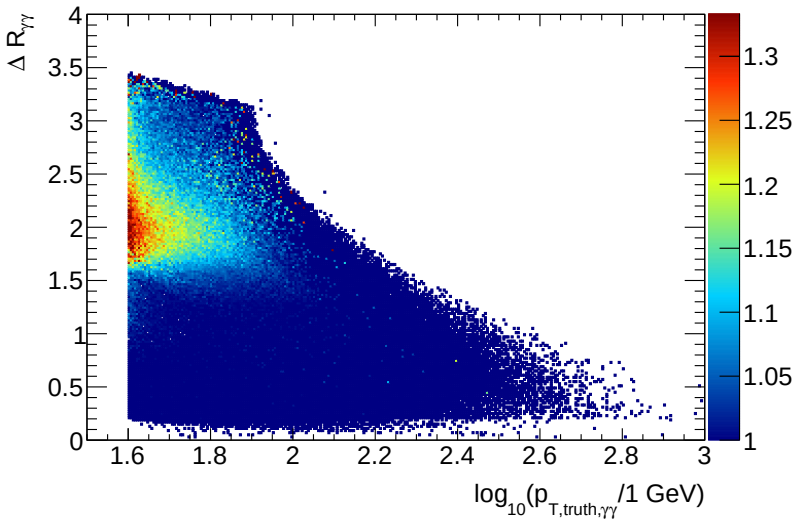


Figure 6.9: *Map of weights as a function of $p^{\gamma\gamma}_{\mathrm{T,particle}}$ and $\Delta R_{\gamma\gamma}$ used to reweight the boosted SHERPA $\gamma\gamma$ sample to account for the bias created from parton-level filtering. Boosted diphotons are barely affected by the filter, and the weights assigned to this kind of topology are essentially 1. Bins without content and isolated fluctuations are set to 1.*

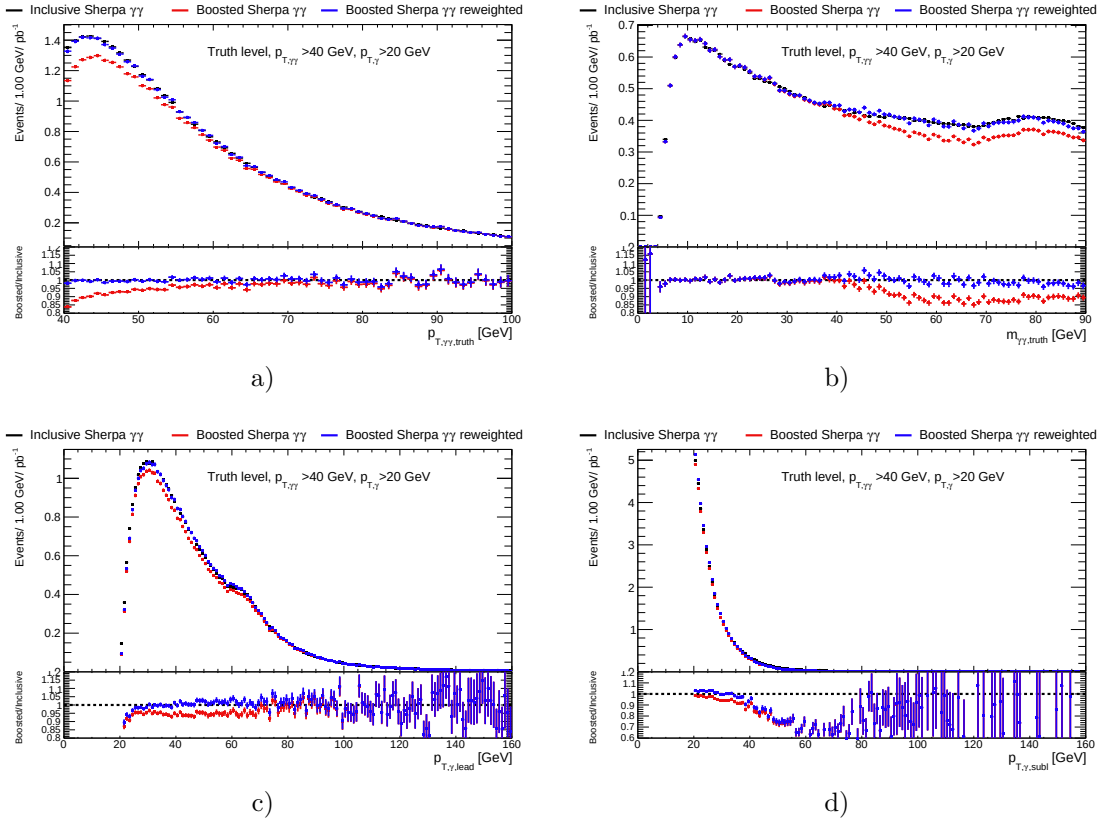


Figure 6.10: *Comparison between unfiltered SHERPA $\gamma\gamma$ sample and boosted SHERPA $\gamma\gamma$ unweighted and reweighted samples. Good improvement results from the the correction of the diphoton invariant mass shape. The agreement is lesser in the tail of the transverse momentum of the subleading photons.*

6.3 Event selection

The event and photon selections are described in this section. The triggers used to collect the data are described in Subsection 6.3.1. The photon selection is detailed in Subsection 6.3.2, where special emphasis is given to the boosted diphoton selection.

6.3.1 Collision data

This analysis is performed with pp collision data from the ATLAS detector collected during Run 2, operating with bunches separated by 25 ns and at a center-of-mass energy of $\sqrt{s} = 13$ TeV. Both prescaled and unprescaled diphoton triggers are used to record the events, the former used to build control regions for the background modelling.

Triggers have evolved in time, to cope with the increasing collision rates that imply changes at both L1 and HLT levels [102]. L1 seeds have had two main changes across Run 2: energy thresholds and isolation requirements. During 2015, deposits in the calorimeter towers (0.1×0.1 in $\eta \times \phi$) were required to be larger than 10 GeV. This threshold was increased from 2016 onwards to 15 GeV. During 2017 and 2018, isolation was required at L1 for seeds with transverse energy lower than 50 GeV and with at least 2 GeV of transverse energy, E_T^{iso}, in the 12 surrounding towers around 2×2 central region towers (see Section 2.2.5).

Energy thresholds and isolation requirements have been implemented in the HLT across the years for the unprescaled diphoton triggers. The 2015–2016 dataset was recorded with two diphoton E_T^γ thresholds, set at 20 GeV for 2015 and the first part of 2016, then increased to 22 GeV until end of that year, when the luminosity peak went above $2.10^{34} cm^{-2} s^{-1}$. The two associated electromagnetic clusters triggered by the L1 were required to pass tight photon identification at HLT level, with an equivalent definition to the offline level but different cuts applied [102]. No isolation in the HLT was applied for 2015 and 2016. For 2017 and 2018, photon isolation was applied in the HLT for unprescaled triggers. Analogous working points to those defined in Section 4 were applied at trigger level. For this analysis, which targets low diphoton masses, photons are required to pass at trigger level loose calorimetric isolation at trigger level, with a relaxed energy threshold: topoetcone20 $< 0.2 E_T^\gamma$. Both 2017 and 2018 datasets were obtained with a diphoton trigger with E_T^γ thresholds of 20 GeV, tight identification and loose isolation at HLT level. The different HLT and L1 triggers used per year are shown in Table 6.2.

Table 6.2: *The lowest unprescaled HTL items, depending on the data-taking period and their associated integrated luminosity. (*)Two different E_T^γ-thresholds are used in 2016 due the instantaneous luminosity reached values above $\mathcal{L}_i < 2.10^{34} cm^{-2} s^{-1}$.*

Year	2015	2016(*)	2016(*)	2017 & 2018
L1 item	`L1_2EM15VH`	`L1_2EM15VH`	`L1_2EM15VH`	`L1_2EM15VHI`
HLT item	`2g20_tight`	`2g20_tight`	`2g22_tight`	`2g20_tight_icalovloose`
luminosity [fb^{-1}]	3.2 fb^{-1}	11.5 fb^{-1}	21.5 fb^{-1}	43.6 fb^{-1} + 58.5 fb^{-1}

Data recorded with the prescaled trigger `HLT_2g20_loose` are used for background shape modelling since the 2017-2018 triggers apply isolation and *tight* identification at trigger level. The total prescale (L1*HLT) of the chain varies with the average number of interactions per bunch crossing $\langle\mu\rangle$ and year. Values range from 10 to 60, being the HLT

prescale constant at 10, and L1 prescale varying from 1.0 up to ∼6.0. For data recorded during 2015 and 2016 the only prescale is due to the HLT. Recorded luminosities per year are shown in Table 6.3.

Table 6.3: *Prescaled* `HLT_2g20_loose` *item, depending on the year, and associated integrated luminosity.*

Year	2015	2016	2017	2018
L1 item	`L1_2EM15VH`	`L1_2EM15VH`	`L1_2EM15VH`	`L1_2EM15VH`
HLT item	`2g20_loose`	`2g20_loose`	`2g20_loose`	`2g20_loose`
luminosity [fb^{-1}]	0.32	3.30	1.56	1.08

Data quality requirements that ensure stable operation of the ATLAS detector are applied, leading to a total recorded luminosity for the unprescaled diphoton triggers of 138fb^{-1}, and 6.3fb^{-1} for prescaled diphoton triggers, measured with a ±1.7% uncertainty [137].

Selected events have at least two reconstructed photons compatible with at least one reconstructed primary vertex (PV) candidate and firing one of the previously mentioned unprescaled diphoton triggers, which vary depending on the data-taking period. Events firing one of the prescaled triggers are also kept for background modelling purposes.

6.3.2 Photon selection

Photon candidates are reconstructed and calibrated as detailed in Sections 3.1 and 3.2. Given the E_T^γ thresholds at 20 GeV imposed by the triggers, an offline kinematic selection at 22 GeV is applied on photons reconstructed with $|\eta| < 2.37$. Photons reconstructed in the crack region are not included. At early stages of the selection, they are required to pass loose identification criteria. Tighter kinematic selections are applied at a later stage.

A custom selection of the PV, different to other analyses, is performed for diphoton final states. The determination of the diphoton vertex is based on the photon pointing method, which exploits the longitudinal segmentation of the LAr calorimeter to "point" at the vertex position, by combining the trajectories of both photons. This is performed with a Neural Network trained with diphoton events from $H \to \gamma\gamma$. The selected vertex position is used to recompute the four-momenta of the two photons, improving E_T^γ and $m_{\gamma\gamma}$ resolutions.

Photons are then required to pass tight identification criteria, with efficiencies increasing from 80% to 90% for unconverted photons with E_T^γ comprised between 20 and 40 GeV. Photons passing relaxed photon identification criteria, denominated Loose' identification and detailed in Section 3.3, are also kept for background estimations.

This analysis is affected by isolation performances, which degrade at low E_T^γ with efficiencies that decrease from 90% at 40 GeV down to 80% at 20 GeV (see Section 4.3.5). The loosest working point `FixedCutLoose` is chosen for this analysis due to the smaller size cone $\Delta R = 0.2$, which allows for searching for lower $m_{\gamma\gamma}$.

Except for the mass range required, the previously described selection is based on standard requirements in diphoton analyses. In what follows, the particular selection requirements and optimization studies relevant for this analysis are presented.

6.3.2.1 Boosted topologies

Requiring photons with an E_T^γ threshold (22 GeV in this analysis) sculpts the invariant mass distribution by creating a "turn-on" in the distribution, a shape that is difficult to describe with analytical functions. This has been the main limiting factor in previous resonance searches at low masses [125]. Since photon pairs with E_T^γ >22 GeV with masses below 40 GeV are essentially collimated in the detector, photon pairs are required to have a large $p_{\mathrm{T}}^{\gamma\gamma}$ with the aim of flattening the invariant diphoton mass distribution.

Figure 6.11 shows the effect of different $p_{\mathrm{T}}^{\gamma\gamma}$ cuts on the SHERPA $\gamma\gamma$ background sample described in Section 6.2.3. By simple inspection, the trigger "turn-on" shifts towards higher masses for increasingly tighter $p_{\mathrm{T}}^{\gamma\gamma}$ selections. From 40 GeV onwards, the description of the background with analytical functions becomes feasible.

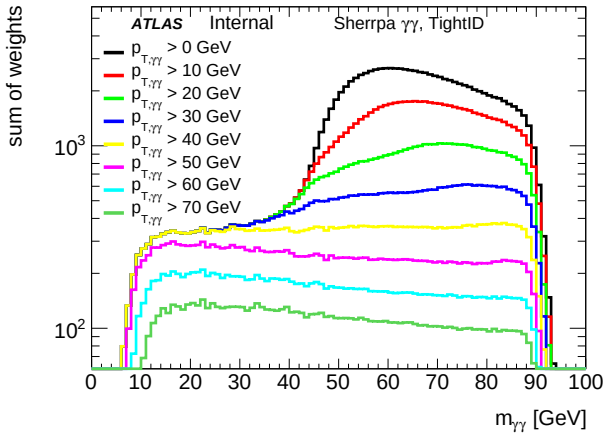


Figure 6.11: *The invariant mass distribution of the inclusive SHERPA $\gamma\gamma$ events for different $p_{\mathrm{T}}^{\gamma\gamma}$ cuts. All distributions are events passing the tight identification. Events passing $p_{\mathrm{T}}^{\gamma\gamma} > 40$ GeV displace the trigger "turn-on" towards higher masses.*

The choice of the $p_{\mathrm{T}}^{\gamma\gamma}$ cut for this analysis is evaluated by studying the significance of a given selection. To avoid being sensitive to model dependent artifacts, no cross-section value is used, as it is possible to perform a significance study by maximizing the ratio of significances for different selections. This ratio does not depend on the cross-sections of the processes nor on the luminosity if the triggers for both selections are the same. The reference significance, Z_{ref}, is the significance obtained for $p_{\mathrm{T}}^{\gamma\gamma} > 40$ GeV (on top of photon identification and isolation requirements), given that looser cuts would not completely flatten the turn-on of the background. The reference significance is compared to other cuts on the diphoton transverse energy with significance Z' to optimize the $p_{\mathrm{T}}^{\gamma\gamma}$ cut value. The ratio of significances can be expressed in terms of signal and background efficiencies as follows:

$$\begin{aligned} Z_i &= \frac{S_i}{\sqrt{B_i}}; \;\; S(B) = L\sigma_{S(B)}\varepsilon_{S(B)}; \\ \frac{Z}{Z_{ref}} &= \frac{\frac{S}{\sqrt{B}}}{\frac{S_{ref}}{\sqrt{B_{ref}}}} = \frac{\varepsilon_S}{\varepsilon_{S_{ref}}}\sqrt{\frac{\varepsilon_{B_{ref}}}{\varepsilon_B}}; \end{aligned} \tag{6.3}$$

where S_i and B_i are the number of signal events for signal and background respectively for a selection i; and ε_i denotes the signal and background efficiencies for each selection i.

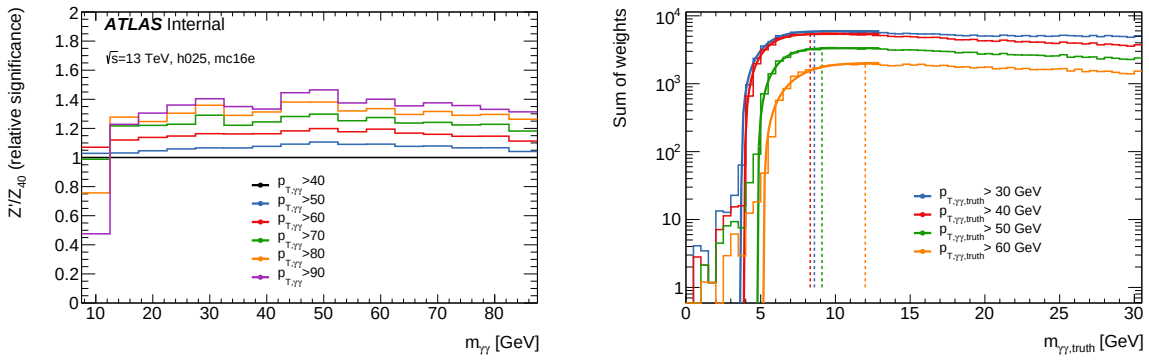
Signal efficiencies $\varepsilon_{S_{ref}(S)}$ are obtained from MadGraph events that pass the reference selection, $E_T^\gamma > 22$ GeV cuts, tight identification and loose isolation.

The number of background events B is obtained from the inclusive Sherpa $\gamma\gamma$ sample from events passing the same selection and falling within $\pm 2\sigma$ around the signal mass, where σ is the resolution of the signal peak. The value of the width of the resonance as a function of the mass is obtained also from the signal MadGraph samples, as will be described in Section 6.4.

The ratio of significances is shown in Figure 6.12a for different $p_{\mathrm{T}}^{\gamma\gamma}$ cuts. A small increase in relative significances is observed for larger $p_{\mathrm{T}}^{\gamma\gamma}$cuts, in detriment of the sensitivity for masses below 15 GeV for $p_{\mathrm{T}}^{\gamma\gamma} > 70$ GeV .

Finally, since this search aims at extending the range down to the lowest attainable mass, the selection aims at the loosest selection, since increasing the threshold also shifts the lower edge of the invariant mass distribution towards larger values (see Figure 6.12b). The shape of the background at lower masses could potentially limit the reach of the achievable mass.

For this analysis, diphotons are then required to have $p_{\mathrm{T}}^{\gamma\gamma} > 50$ GeV, a choice that is a compromise between three competing constraints: flattening out the background shape, improving the sensitivity in a wide mass range and reaching the lowest attainable masses.



a) b)

Figure 6.12: *(a) Relative significances for different $p_{\mathrm{T}}^{\gamma\gamma}$ cuts with respect to the reference selection with $p_{\mathrm{T}}^{\gamma\gamma}$>40 GeV (black straight line) (b) Truth invariant diphoton mass distributions for different $p_{T,\gamma\gamma,truth}$ cuts. Increasing the $p_{T,\gamma\gamma,truth}$ cut reduces the phase space and by shifting the edge of the distribution towards larger masses.*

6.3.3 Signal efficiencies

Total selection efficiencies are shown in Figure 6.13, where the cumulative effects of the preselection, trigger match, photon identification and isolation are presented. Efficiencies are measured on the signal simulated samples for different pileup conditions. Preselected events contain two loosely identified photons firing an unprescaled diphoton trigger, being the trigger the principal source of efficiency loss due to the per-photon E_T^γ requirement. The slightly higher efficiency in the simulation with the same pileup conditions as data collected in 2015 and 2016 (filled circles) compared to the other simulated samples (simulated to replicate data collected in 2017 and 2018) in the trigger selection efficiencies comes from the online loose isolation applied at trigger level in the last two years. The

drop in efficiencies is further discussed in the next lines. Efficiency losses arising from the boosted selection are larger at higher masses, since events recorded in the low mass part of the spectrum are necessarily boosted.

All the previous efficiencies are measured with respect to the total number of generated events, fulfilling all parton and particle-level filters. The black markers, denominated *Filter efficiency*, denote the total selection efficiency including the efficiency of the filter applied at on the generation. The overall efficiencies go from roughly $\sim 1\%$ at 10 GeV up to 10% at 90 GeV.

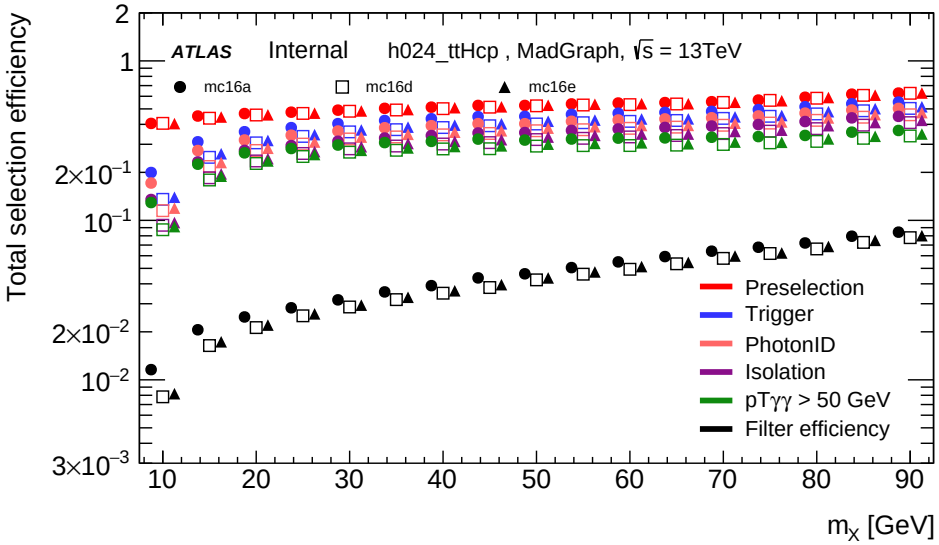


Figure 6.13: *Total selection efficiencies for simulated signal with data-taking conditions of 2015+2016 (solid circle), 2017 (empty square) and 2018 (solid triangle) for different mass points.*

6.3.3.1 Efficiency drop at low ΔR

In general, analyses are expected to probe regions of the phase space with approximately flat efficiencies at reconstruction, where the confidence that the simulation reproduces the data is high. The drop observed in the trigger efficiencies thus deserved additional attention.

In order to better illustrate the effect, Figure 6.14 shows the reconstruction efficiencies with and without requiring a trigger in the simulation, as a function of the angular distance $\Delta R_{\gamma\gamma}$ between the two reconstructed photons. For this set of plots, the efficiencies are defined with respect to the number of events in the fiducial phase space (Section 6.1). An efficiency drop is observed for photon pairs with $\Delta R_{\gamma\gamma} < 0.3$ after requiring a diphoton trigger. The drop in efficiency has a larger impact at lower masses, where a non negligible fraction of the events populates this region. Since this drop is observed for both diphoton triggers (with and without online isolation), its source is expected to be related to the topology of the photons. A good description of the trigger inefficiencies in the simulation needs to be ensured to extend the lowest attainable mass.

Single photon trigger efficiencies are measured in data using minimum bias events, with very loose or no trigger requirements; and with radiative $Z \to ll\gamma$ samples, where ll denotes either an electron or a muon pair. Since no clean diphoton control sample exists

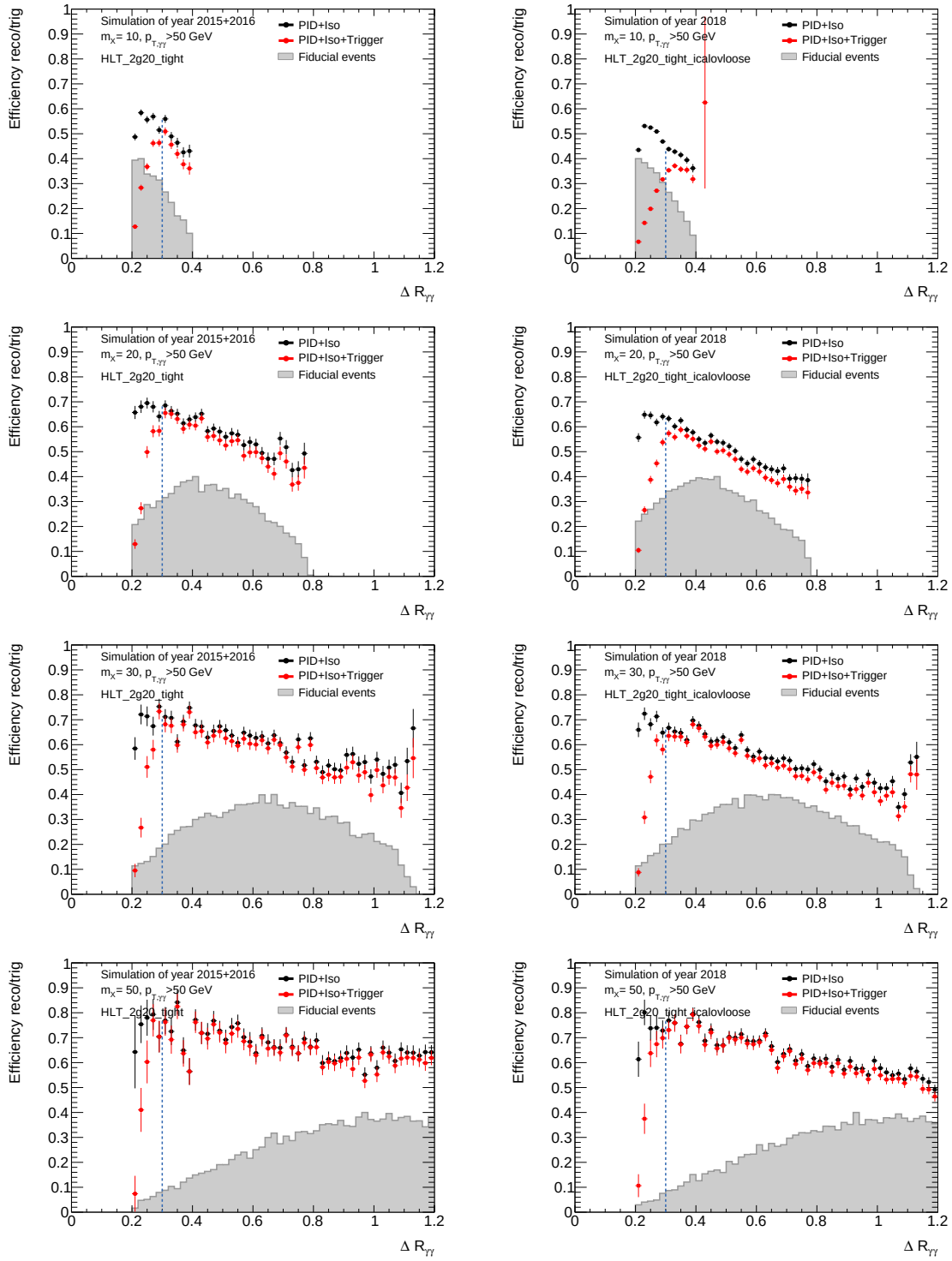
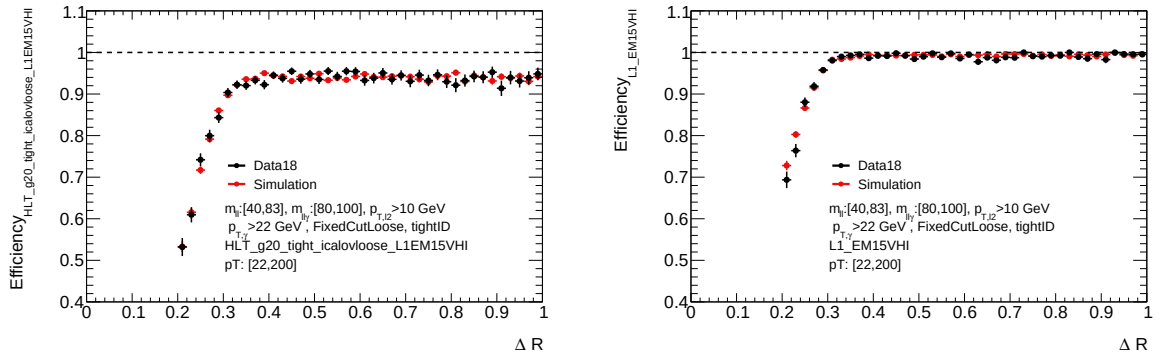


Figure 6.14: *Efficiencies with (red) and without (black) requiring a diphoton trigger matching in signal simulated samples for mass points of 10, 20, 30 and 50 GeV. Plots on the left show the efficiencies for the diphoton triggers used in 2015-2016 while plots on the right show the efficiencies for the diphoton triggers used in 2017-2018. The shaded region in grey shows the distribution of the events falling in the fiducial region (in arbitrary units). The fraction of the shaded region falling for $\Delta R < 0.3$ increases with decreasing mass.*

in data, diphoton trigger efficiencies can at best be estimated as the product of per-photon efficiencies. This procedure neglects any correlation between both triggered photons, an assumption that is not valid in this analysis, since diphotons are required to be boosted.

In order to evaluate the impact of possible differences between the simulation and recorded data, an alternative approach using radiative Z boson decays is followed. Radiative $Z \to ee\gamma$ candidates are selected from events with two reconstructed electrons and one reconstructed photon with combined $m_{ee} : [40, 83]$ GeV and $m_{ee\gamma} : [80, 100]$ GeV in order to enrich the sample with final state radiation photons. This selection ensures a photon purity between 95% and 98% with increasing transverse momentum. No additional background subtraction is performed.

Using the dielectron pair as a tag and the photon as a probe, the efficiency of the single photon trigger is measured in both data and simulation to evaluate the trigger performances for close egamma objects. A similar drop to the one observed in the signal simulated samples (Figure 6.14) is observed in Figure 6.15a for radiative $Z \to ee\gamma$ events. The drop in the trigger is understood to be caused by the L1 trigger, as shown in Figure 6.15b. A fair agreement between data and simulation is observed, and the remaining difference is absorbed in a correction factor, defined as the difference between data and simulation efficiencies. The correction factors are applied on the signal efficiencies for photon pairs with $\Delta R_{\gamma\gamma} < 0.35$, since data-simulation corrections already exist for non-collimated photon pairs.



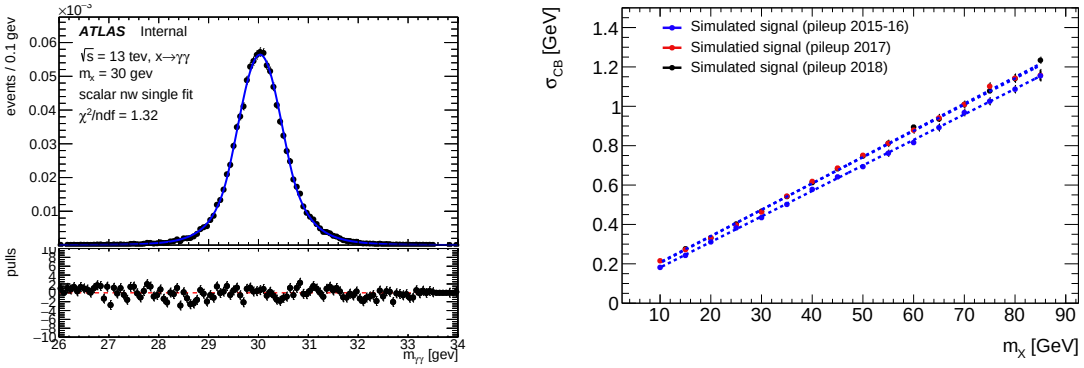
a) b)

Figure 6.15: *(a) Trigger efficiencies for the trigger chain* `HLT_g20_tight_icalovloose_L1EM15VHI` *measured in simulation and data recorded during 2018. (b)* `L1_EM15VHI` *item trigger efficiency measured in simulation and data recorded during 2018. The drop is already observed at L1 trigger.*

6.4 Signal modelling

This section describes the methodology used to extract a parametrization of the signal shape as a function of the diphoton invariant mass. The strategy for the signal modelling consists of describing the invariant mass distribution $m_{\gamma\gamma}$ of signal events with an analytical function. Since the signal was generated in the narrow-width-approximation, the width of the signal after reconstruction comes entirely from the resolution of the detector. A double-sided Crystal Ball (DSCB)function (see Appendix B) is used to model the detector resolution for each of the different signal mass points, obtaining the evolution

of its parameters as a function of $m_{\gamma\gamma}$. The DSCB function consists of a Gaussian core with two power law components to describe the tails. Instead of parametrizing the peak position μ, which should be close to the reference mass of each signal point, the parameter $\Delta m_X = \mu - m_X$, defined as the difference between the peak of the Gaussian and the reference mass value, is used. Figure 6.16a shows an example fit for a resonant signal with 30 GeV of mass. The evolution of the signal width with respect to the mass can be observed in Figure 6.16a. The resolution is ~ 200 MeV at 10 GeV and increases up to ~ 1.2 MeV at 90 GeV. The width follows to first order a linear trend as a function of $m_{\gamma\gamma}$ with a small dependence for different pileup conditions. The rest of the parameters of the DSCB do not vary significantly. The trends of all parameters as a function of $m_{\gamma\gamma}$ are shown in Table 6.4; altogether these trends provide the shape of the signal at any arbitrary mass in the 10-90 GeV range.



a) b)

Figure 6.16: *(a) Fit to the $m_{\gamma\gamma}$ distribution using a double-sided Crystal Ball function for a resonant signal of 30 GeV of mass. (b) Mass dependence of the width of the simulated signal samples for different pileup conditions. Blue markers denote the measured width under pileup conditions simulated to replicate those from years 2015 and 2016. Red and black markers are almost identical and denote the measured width under pileup conditions simulated to replicate those from years 2017 and 2018 respctively. The observed trend is well described with a linear dependence.*

Parameter	Parameterisation	a	b
Δm_X	$a+b(m_{\gamma\gamma}-50)$	50.8 ± 1.4 MeV	0.84 ± 0.05 MeV/GeV
σ_{CB}	$a+b(m_{\gamma\gamma}-50)$	734 ± 2 MeV	13.43 ± 0.06 MeV/GeV
α_{low}	$a+b(m_{\gamma\gamma}-50)$	1.44 ± 0.01	$(-0.07 \pm 0.27) \cdot 10^{-3}$
α_{high}	$a+b(m_{\gamma\gamma}-50)$	1.36 ± 0.01	$(1.5 \pm 0.3) \cdot 10^{-3}$
n_{low}	a	10.2	
n_{high}	a	11.6	

Table 6.4: *Parameterizations of the DSCB parameters describing the signal shape, as a function of the diphoton mass. The value $m_{\gamma\gamma}$=50 GeV is used as reference for the linear trends, i.e. the value of each parameter at 50 GeV is given by the value of the offset a, and its change as a function of $m_{\gamma\gamma}$is given by the slope b.*

The signal model parametrization is validated to evaluate potential biases arising from an imperfect description of the shape. This validation is performed with a pseudo-dataset, composed of a smooth background reproducing the shape in Figure 6.28, over which the simulated signal is added for different mass hypotheses. The amount of "injected" signal is reweighted to emulate the expected upper limit on the amount of events for a given mass hypotheses[5]. The bias is estimated as the difference between the weighted signal yield obtained in a signal plus background fit and the amount of "injected" signal. The result of this test is shown in Figure 6.17, in which the bias as a function of the mass is observed to be below 1% in the 10-90 GeV mass range.

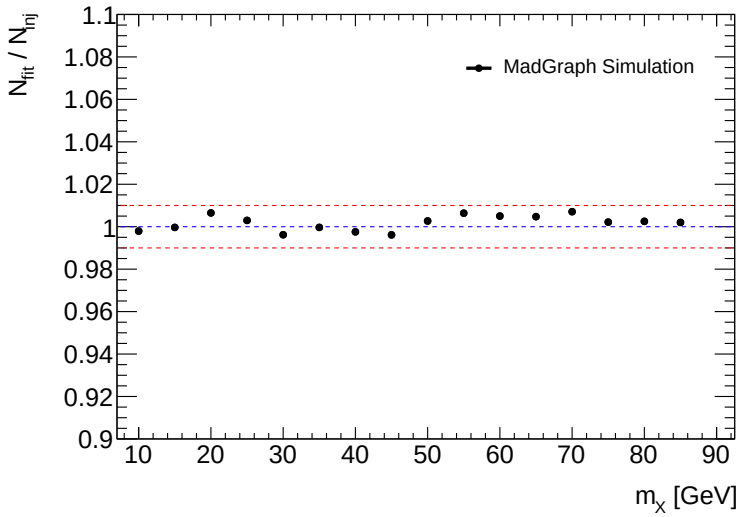


Figure 6.17: *Comparison between the fitted number of signal events N_{fit} with respect to the injected N_{Inj} one as a function of $m_{\gamma\gamma}$. The red dashed lines denote $\pm 1\%$ difference between them.*

6.5 Background modelling

Signal and background are separated in this analysis through differences in their diphoton invariant mass distributions. As explained for the signal model in the previous section, it is common practice to model the background distribution with analytical functions to describe the background shape. This analysis relies on simulated samples and events from data control regions, orthogonal to the signal region, to build background templates, which are then used to test and select a set of analytical functions. This section details the methodology followed to choose the analytical function used to describe the background shape.

6.5.1 Overview of the methodology

The construction of the background template is divided in several steps. In Subsection 6.5.2, the different components that are present in the background are explained, as well as the samples used to describe each of them. Each of the background components

[5]More precisely, the value corresponds to twice the uncertainty of the expected background in the region that contains 95% of the signal.

represents then a fraction of the data. These fractions are computed in Subsection 6.5.3 using a bidimensional ABCD method. The combination of the different background components is explained in Subsection 6.5.4, in which the simulated samples are reweighted to reproduce the total template to benefit from the large statistics. Finally, the model used to describe the shape of the background is presented in Subsection 6.5.5. Systematic variations of the background template are built in Subsection 6.5.6, and used to evaluate the flexibility of the model.

6.5.2 Background processes

The background template is built by considering the dominant background processes in the search range of the analysis. As briefly introduced in Subsection 1.2.4, the dominant background processes affecting this analysis can be classified in two different categories:

6.5.2.1 Diphoton component

The irreducible component, referred to in the following as the $\gamma\gamma$ component, arises from non-resonant QCD photon pairs produced in the processes described in Section 1.2.4. The invariant mass spectrum for this background is obtained from simulated $\gamma\gamma$ events generated with Sherpa (detailed in Section 6.2.3) and passing the full selection chain described in Section 6.3.

6.5.2.2 Reducible component

The reducible component is composed by events in which at least one of the two photon candidates is a misidentified non-prompt photon (a fake). Despite photon identification and isolation, which reject fakes by very large factors [138], a non-negligible fraction is still present after all the selection requirements. The reducible component is divided in three categories depending on which photon candidate(s) (either the leading or subleading or both) has been misidentified: γj, $j\gamma$ or jj.

The shape of the reducible component is extracted from data control regions recorded with the prescaled trigger `HLT_2g20_loose`, since all unprescaled triggers require online tight photon identification. This is a limiting factor in the estimation of the reducible background $m_{\gamma\gamma}$ shape. Data control regions are enriched in $\gamma j/j\gamma$ and jj events, by requiring the nominal kinematic and isolation criteria but inverting photon tight identification, as explained in Section 6.3. Any of both photon candidates is allowed to fail tight identification, this happening more often for the subleading photon. Among the different Loose' definitions, Loose'2 is the tightest requirement before *tight* identification. Loose'2 photons are then expected to provide a better estimation of the reducible background shape in the signal region compared to other Loose' selections. However, it is very limited in statistics. On the other hand, Loose'5 is the largest sample, but its shape may not necessarily be representative of that of the signal region. A comparison of the invariant mass distributions for the different Loose' selections is shown in Figure 6.18. The sharp turn-on at masses below 10 GeV shifts slightly towards larger masses when tightening the Loose' identification, probably due to identification inefficiencies in close-by photon topologies. Loose'4 is chosen as the nominal photon identification requirement for the data control regions, while the other Loose' variations are considered as systematic variations of the reducible background shape.

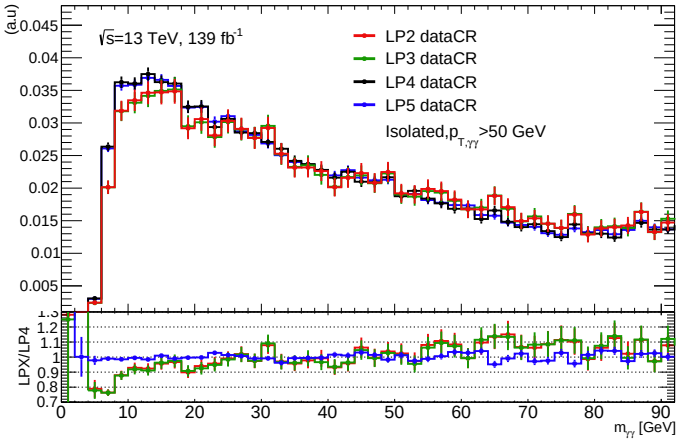
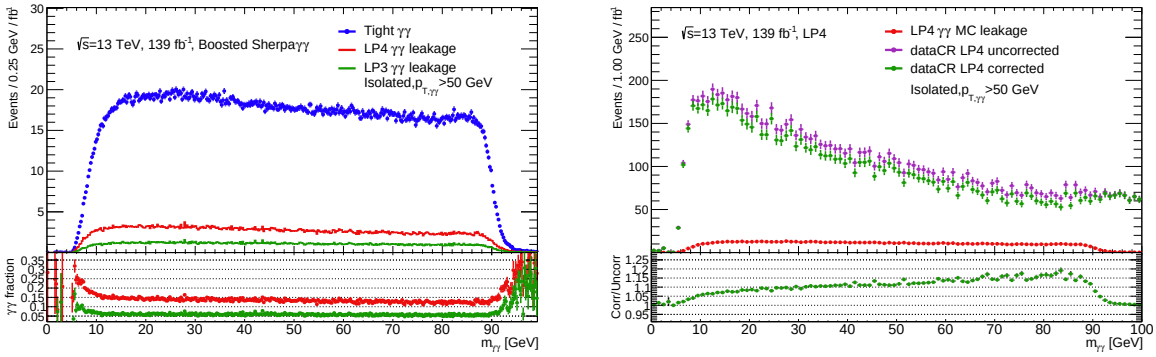


Figure 6.18: *Reducible background distributions obtained from data by inverting tight identification and requiring one of the Loose' selections. All distributions are normalized to unity. Differences in shape up to 10% are observed from 10 GeV up to 90 GeV between Loose'2 and Loose'3 compared to the nominal selection Loose'4. No significant differences are observed between Loose'5 and Loose'4.*

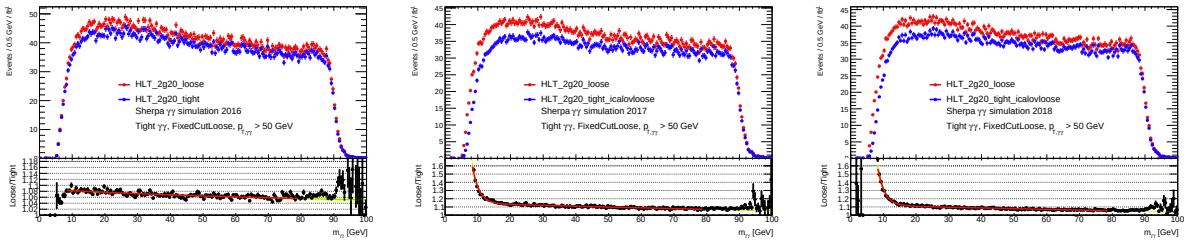
However, a non-negligible fraction of prompt photons is present in the control regions and needs to be subtracted. The contamination of true diphoton events is estimated from the simulated SHERPA samples. The fraction of $\gamma\gamma$ events in the data control regions is found to be between 5-20% as a function of the mass. The invariant mass distribution of these photons is statistically subtracted from the raw shape obtained from the data control regions, modifying the slope of the distribution by up to ±10% from 10 to 80 GeV (see Figure 6.19b).



a) b)

Figure 6.19: *(a) The diphoton invariant mass distributions for photon pairs in the boosted SHERPA $\gamma\gamma$ sample. Events with diphotons passing tight identification are shown in blue and photon pairs in which one or both photons fail tight identification but pass Loose'4 are shown in red. The fraction of true photons in the control region is shown below. (b) Diphoton invariant mass distribution for the different Loose' control regions in data shown together with the expected contribution from true photons. The ratio between both corrected and uncorrected shapes is shown below.*

Using different triggers in signal and control regions potentially induces an additional difference in the invariant mass distributions from the control regions obtained from data compared to those from the signal regions, as different L1 seeds are used in the unprescaled triggers as detailed in Section 6.3.1. The L1 seeds also varied across the different years, including an isolation requirement in the last two years. The largest difference between the shapes of the invariant mass distribution of events recorded with the prescaled and unprescaled triggers is observed at very-low masses for data recorded in 2017 and 2018, where differences larger than 20% are observed. No significant differences are observed for data recorded in 2015 and 2016. The observed effect in 2017 and 2018 is related to the HLT inefficiency illustrated in Figure 6.15a arising from the online isolation requirement. A correction for the shape is derived for each data-taking period from a parametrization of the ratio between the invariant mass distributions of events passing the full reconstruction chain of the selected events (tight identification and isolation) and triggered with `HLT_2g20_loose` over those triggered with the nominal unprescaled trigger used in each particular (see Figure 6.20).



a) b) c)

Figure 6.20: *The invariant mass distribution of SHERPA Monte Carlo $\gamma\gamma$ events triggered with the prescaled trigger `HLT_2g20_loose` in red and with the unprescaled trigger of each data-taking period in blue. The ratio below the plots indicates the changes of the invariant mass distribution due to the different triggers used in each data-taking period.*

6.5.2.3 Additional background sources

Other backgrounds are present in the considered mass range of this analysis from electrons faking photons in $Z \to ee$ decays and radiative $Z \to ee\gamma$ and $W \to e\nu\gamma$ decays. The contribution from these backgrounds is found to be small and their invariant mass distribution below the mass of the Z boson is similar to those observed from the $\gamma\gamma$ and reducible components (see Figure 6.21). These background sources are neglected in the following since their contributions do not change significantly the shape of the total background shape.

6.5.3 Background decomposition

To estimate the relative fraction of reducible and $\gamma\gamma$ background, the so-called 2x2D ABCD sideband decomposition method is used, described in detail in [139]. The 2x2D sideband method extracts the expected background contribution in the signal region from the observed yields in various data control regions. Four orthogonal regions are built per photon using two loosely correlated inputs: photon isolation and photon identification. Each region contains events failing/passing photon identification and/or photon isolation.

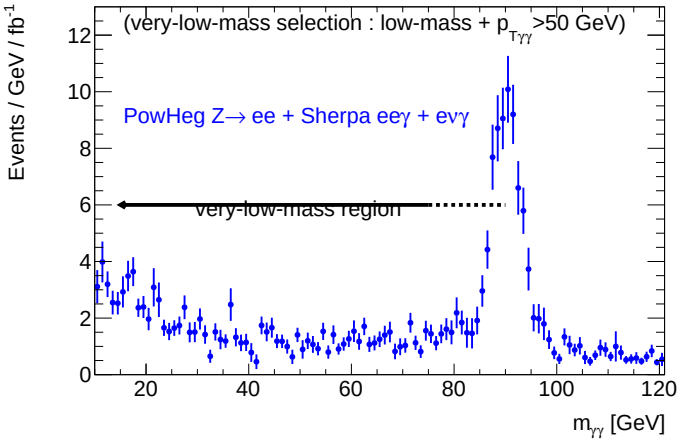


Figure 6.21: *Expected backgrounds from Z and W boson decays and radiative Z boson decays in which one or two electrons fakes a photon, passing the full selection chain.*

Since two photons are present, this procedure leads to 16 orthogonal regions. The signal region is composed of diphoton pairs, with both photons passing tight identification and isolation requirements. The control regions contain diphoton pairs, where are least one of the photons fails at least one of the two aforementioned criteria.

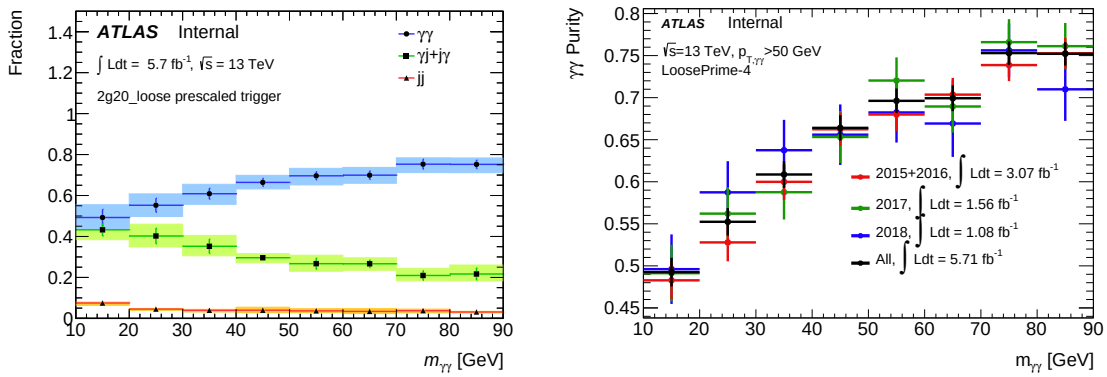
The fraction of $\gamma\gamma$ events obtained with the background decomposition 2x2D method is shown as a function of the diphoton mass on Figure 6.22a. The reducible background component increases towards lower masses, and becomes the dominant contribution to the total background at the lower edge of the range. The large uncertainties in the reducible components are caused by the limited statistics in the data control regions due to the use of prescaled triggers. The 10 GeV binning, significantly wider than the invariant mass resolution, is chosen to avoid a partial, indirect unblinding of the data.

The purity is also obtained separately for each data-taking year to evaluate the effects arising from changing pileup conditions and triggers. However, Figure 6.22b shows that the $\gamma\gamma$ purity does not change significantly between data-taking periods, ranging from 50% at low masses (10-20 GeV bin) up to 70% at larger masses. The mass-averaged purity is measured to be $0.642 \pm 0.006(stat)^{+0.028}_{-0.032}(syst)$.

6.5.4 Building the background template

Although the 2x2D decomposition method would allow to obtain the correct shape for the background, doing so would correspond to unblinding the data in the signal region. Moreover, the method breaks down when using finer binning due to the lack of statistics. In view of this, the complete background template is built by adding the $\gamma\gamma$ and reducible components, each of them scaled according to the luminosity and the $\gamma\gamma$ purity obtained with the 2x2D decomposition method.

The full template is built by adding the $\gamma\gamma$ component, obtained from the boosted Sherpa $\gamma\gamma$, and the reducible background component obtained from data control regions after applying the leakage and trigger corrections mentioned above. Since the statistics in the data control region is very limited, the $\gamma\gamma$ component is reweighted to reproduce the full template in order to protect the statistical power of the MC simulation. This procedure acts as a smoothing of the reducible control region, reducing the impact of the reduced statistics from data control regions on the final template. These weights are

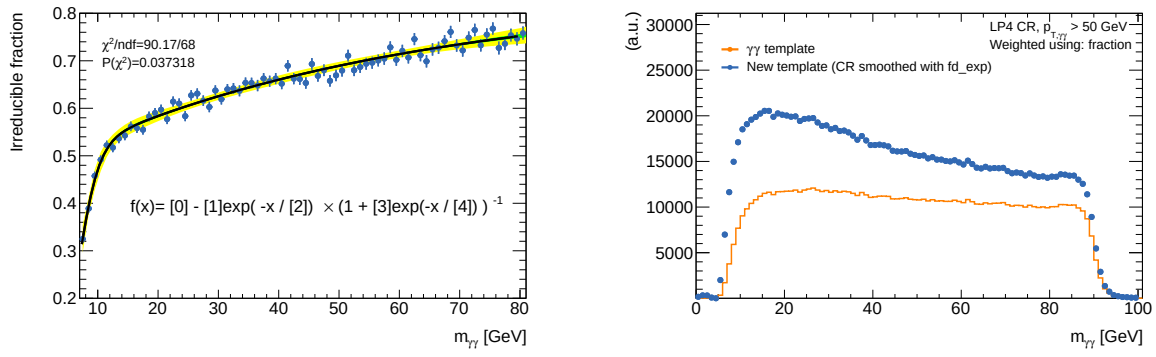


a) b)

Figure 6.22: *(a) The $\gamma\gamma$, $\gamma j/j\gamma$, and jj fractions determined by the 2x2D sideband method as a function of the diphoton mass. (b) Purity of the data sample in prompt isolated diphotons as obtained from the 2x2D sideband decomposition method. The result is obtained separately for the different data-taking periods in order to compare the purity for different pile-up and trigger conditions and good agreement is seen. The errors shown reflect statistical and systematic uncertainties coming from Loose' variations.*

extracted from an analytical parametrization of the $\gamma\gamma$ fraction as a function of $m_{\gamma\gamma}$, as shown in Figure 6.23a. Assuming that the fraction can be parametrized with a simple function implies that a certain knowledge of the shape of the two templates is added, which naturally reduces the impact of the poor statistics from the data control region in the final template. Given the aim of attaining the lowest mass possible, the ratio is described with two contributions. A "Fermi-Dirac like" function (as the one introduced in the next section) is used to describe the drop of the $\gamma\gamma$ component around 15 GeV, and a simple exponential is used to describe the increasing trend beyond this point up to 80 GeV.

The total template is shown in Figure 6.23b. The statistical uncertainties on the final template are the squared sum of the uncertainties on the $\gamma\gamma$ template and the uncertainties on the fit of the $\gamma\gamma$ purity as a function of $m_{\gamma\gamma}$.



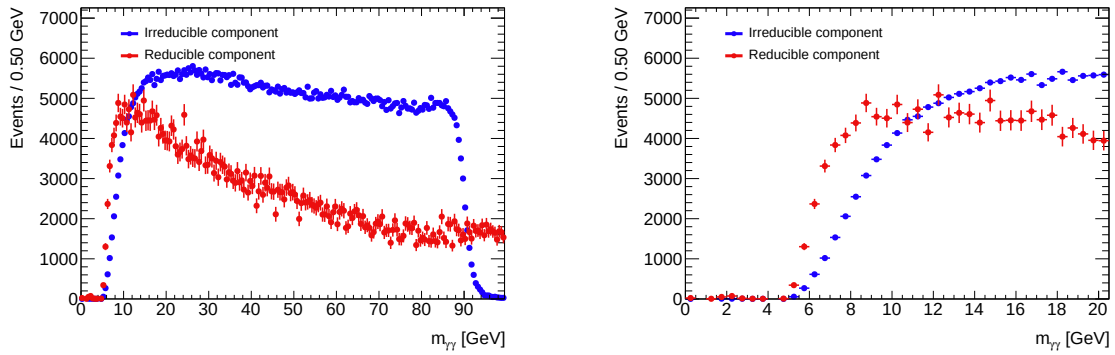
a) b)

Figure 6.23: *(a) The diphoton purity as a function of $m_{\gamma\gamma}$. The green and yellow bands around the parametrization denote the ± 1 and $\pm 2\sigma$ fit uncertainties respectively. (b) The total background template built using the parametrization shown on the left.*

6.5.5 Background model

The background shape is modeled using a PDF built with analytical functions whose parameters are to be left free in the fit to the background template. This methodology is known as *functional form modelling* and it is a common choice in several other resonance search analyses, in which the background distribution of the discriminant observable is smooth enough to be described empirically by analytical functions [140]. However, the unconventional shape of the background with a turn-on around 15 GeV requires a separate treatment, if one wants to extend the search range down to the lowest possible masses. Special emphasis on the modelling of the turn-on feature is given.

The strategy followed to describe the full template is based on the intrinsic characteristics of the template: both the $\gamma\gamma$ and the reducible templates are independently modeled with different analytical functions, and their features are then combined to describe the full template. This methodology relies on the fact that both templates have clear distinguishable features. The $\gamma\gamma$ component is qualitatively described by a turn-on feature spanning between 10 and 25 GeV, followed by a slowly falling component above 30 GeV. On the other hand, the reducible component is barely affected by a turn-on, since it only spans a few GeV above 5 GeV compared to that observed in the $\gamma\gamma$ component and it can be qualitatively described by a smoothly falling distribution. These statements are illustrated in Figure 6.24, where both templates are shown.



a) b)

Figure 6.24: *The irreducible and reducible background templates normalized to their expected contribution in data. A zoom is shown on the plot of the right, so that the differences in the turn-on region can be better illustrated.*

In the following, the description of the functional form describing the full template is detailed step by step.

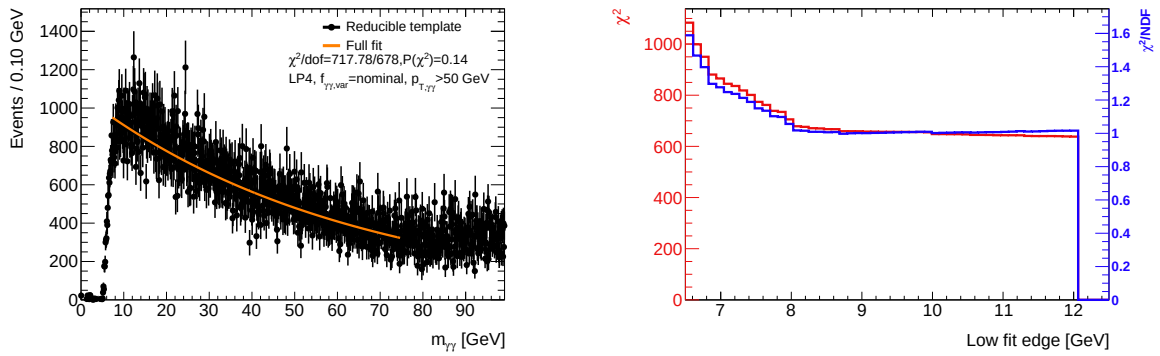
6.5.5.1 Reducible component shape

A simple decreasing exponential function is chosen since it describes the shape with a single free parameter:

$$Red(x;\{\tau_r\}) = e^{-\frac{x}{\tau_r}}, \quad \text{with} \quad x = m_{\gamma\gamma} - x_0 \tag{6.4}$$

where τ_r is the length scale and x_0 is a fixed quantity (arbitrarily set to 10GeV), that acts as a reference of the overall normalization. The definition of x in terms of $m_{\gamma\gamma}$ holds for the rest of the section. Figure 6.25a shows the fit of this functional form to the reducible

component. With the available statistics, adding a second exponential term does not increase the quality of the shape description. The lower edge of the fit range is determined by evaluating the χ^2 of fits starting at the different values from 7 up to 10 GeV. As shown in Figure 6.25b, the quality of the fit rapidly degrades for masses below 8 GeV.



a) b)

Figure 6.25: *(a) Fit to the reducible component from 8 and 75 GeV. (b) Evolution of the χ^2 (in red) and of χ^2/NDF (in blue) with the lower edge of the fit range. Beyond 8 GeV, the quality of the fit does not improve further.*

6.5.5.2 $\gamma\gamma$ component shape

Reaching masses below 15 GeV requires a robust description of the turn-on in the $\gamma\gamma$ component. For this, the $\gamma\gamma$ component is built with two different functional forms: a turn-on and a smoothly falling.

The turn-on is fairly well described by an "inverted" exponential:

$$f(x;\{f_0,\tau\}) = 1-(1-f_0)e^{-\frac{x}{\tau}}, \tag{6.5}$$

where f_0 corresponds to the value of the functional form for $x = x_0$; and τ is the length scale of the turn-on.

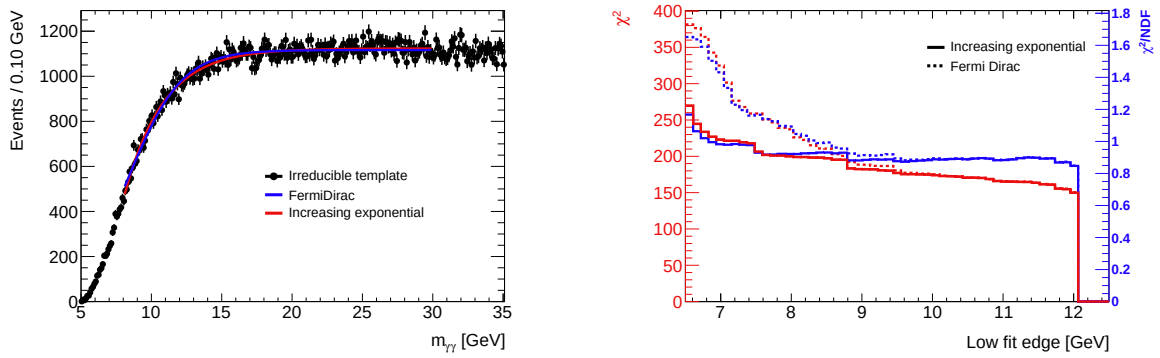
A different functional form used to evaluated the turn-on is a "Fermi-Dirac":

$$f(m_{\gamma\gamma};\{\delta,\tau\}) = \frac{1}{1+e^{-\frac{m_{\gamma\gamma}-\delta}{\tau}}}, \tag{6.6}$$

where the parameters δ and τ describe the mean and the increasing rate of the turn-on respectively.

Both functional forms are fit to the $\gamma\gamma$ template between 7 and 30 GeV in Figure 6.26a. The χ^2 is obtained for fits covering down to particular mass values to evaluate which functional form could be used to attain lower masses. While both provide identical descriptions above 8.5 GeV, the "inverted" exponential is observed to describe significantly better the turn-on down to lower mass values (see Figure 6.26b). Moreover, the parameters of the inverted exponential are more stable with varying fit ranges. For these reasons, the increasing exponential functional form is chosen as the nominal functional form.

The smoothly falling region beyond 30 GeV can be described by several smoothly decreasing functions, like exponentials or power laws. For the sake of simplicity, a simple



a) b)

Figure 6.26: *(a) Functional form fits to the $\gamma\gamma$ template, obtained from simulation, in the range between 7 and 30 GeV. Both Fermi-Dirac and the inverted exponential functional forms are shown in blue and red respectively. (b) Evolution of the χ^2 (in red) and of χ^2/NDF (in blue) with the lower edge of the fit range. The quality of the fit for the inverted exponential (solid line) functional form is observed to be better compared to the Fermi-Dirac (dashed line).*

exponential is used and it is shown to sufficiently describe the full template. The complete functional form of the $\gamma\gamma$ template is expressed as follows:

$$\begin{aligned} &Irred(x;\{f_0,\tau_t,\tau_{slow},f_{mix}\}) = \\ &f_{mix}\left(1-(1-f_0)exp\left(-\frac{x}{\tau_t}\right)\right)+(1-f_{mix})exp\left(-\frac{x}{\tau_{slow}}\right), \end{aligned} \tag{6.7}$$

where f_{mix} is the parameter describing the relative contribution of the turn-on and the smoothly falling component. An example of the model fitting the full $\gamma\gamma$ template is shown in Figure 6.27.

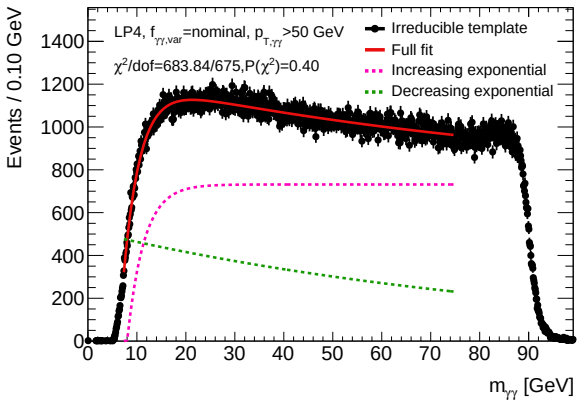


Figure 6.27: *Irreducible template described with the functional form (solid red) built as a combination of a turn-on (dashed pink) and a decreasing exponential (dashed green) functions.*

6.5.5.3 Full template description

The background model is constructed by combining the previous functional forms as:

$$Bkg(x;f_{\gamma\gamma},\boldsymbol{\theta}:\{f_0,\tau_t,\tau_{slow},f_{mix},\tau_r\}) = f_{\gamma\gamma}Irred(x;\{f_0,\tau_t,\tau_{slow},f_{mix}\})+(1-f_{\gamma\gamma})Red(x;\{\tau_r\}) \tag{6.8}$$

where $\boldsymbol{\theta}$ represents the parameters left free in the fit and $f_{\gamma\gamma}$ is the parameter describing the relative contribution of the $\gamma\gamma$ and reducible components. The latter symbol is chosen since it represents the $\gamma\gamma$ purity and it is expected by construction to be compatible with the value used to build the template, obtained with the 2x2D decomposition method (Section 6.5.3). The full template fit is shown in Figure 6.28. The agreement found between the model and the template is found to be sufficient enough with the current statistics to be chosen as main background functional form.

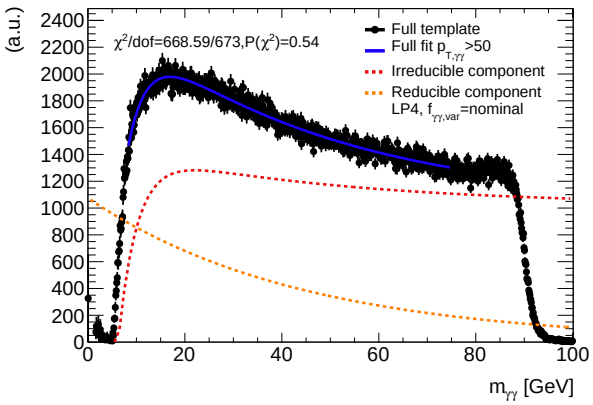


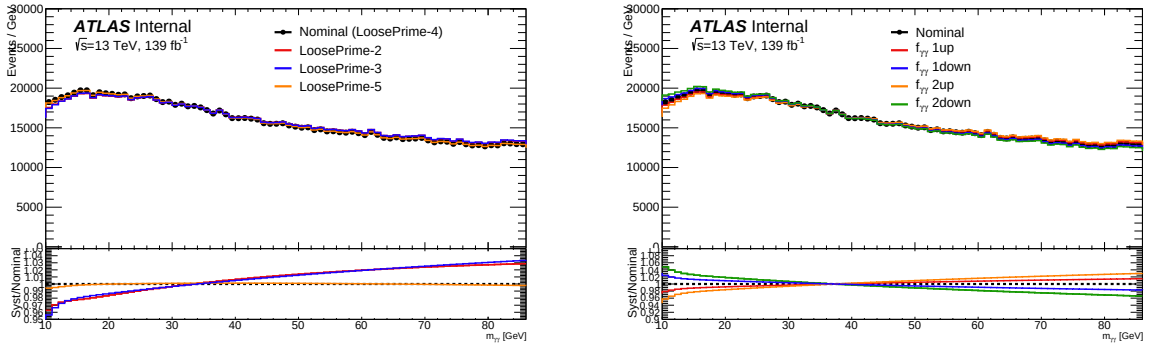
Figure 6.28: *Full template described with the functional form (solid blue) built as a combination of the $\gamma\gamma$ (dashed red) and the reducible (dashed orange) components.*

6.5.6 Systematic variations of the background template and consistency checks of the model

Systematic variations of the shape of the total background template are obtained by modifying the definition of the control region and the $\gamma\gamma$ purity. The Loose' requirement of the data control region is modified with respect to the nominal choice, resulting in a change in the background shape. Likewise, $\gamma\gamma$ purity variations are estimated by varying the purity according to the uncertainty provided by the decomposition method. Both variations are shown in Figures 6.29a and 6.29b respectively. A significant change on the shape is observed when considering different Loose' definitions, of the order of 3%. Changes of similar magnitude are observed when varying the normalization of the $\gamma\gamma$ component.

The robustness of the previously described methodology is evaluated by checking whether the existing systematic variations of the template can be described with the functional form in Equation 6.8. Normalization variations are not shown since, by construction, these imply only a variation in the $f_{\gamma\gamma}$ parameter without any impact on the remaining functional form parameters. However, Loose' variations modify the shape of the turn-on, being more prominent for the tightest definitions (Loose'2 and Loose'3). While a small detriment on the quality of the fit is observed, the description is found to be sufficient proof that the methodology used to build the functional form is robust.

Moreover, in order to evaluate possible variations of the turn-on, the performance of the functional form is evaluated for different kinematic selections. In particular, the cut value of $p_{\mathrm{T}}^{\gamma\gamma}$is shifted by ±5 GeV with respect to the nominal value of 50 GeV. These two variations modify significantly the shape of the turn-on, as shown in Figure 6.18. A worse description is found for the templates built requiring $p_{\mathrm{T}}^{\gamma\gamma} > 45$ GeV. However, tightening



a) b)

Figure 6.29: *Systematic variations of the total template, associated with both reducible and $\gamma\gamma$ components. (a) Reducible background variations compared to the nominal Loose'4. (b) $\gamma\gamma$ normalization variations for the irreducible component.*

the $p_{\mathrm{T}}^{\gamma\gamma}$ cut increases the quality of the fit as well as further reducing the contribution o r g
c I a e
d

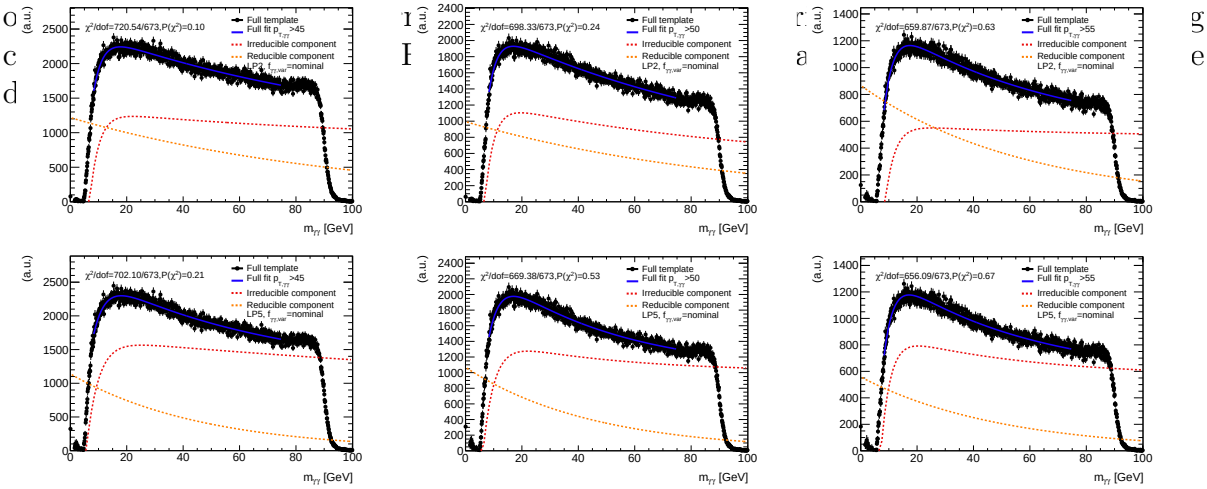


Figure 6.30: *Template fits to Loose' variations and different $p_{\mathrm{T}}^{\gamma\gamma}$ selections. From left to right, the $p_{\mathrm{T}}^{\gamma\gamma}$ cut increases from 45 up to 55 GeV. From top to bottom, the Loose' criteria is loosened, from Loose'2 (tightest) to Loose'5 (loosest). Loose'4 is not shown, since it is the nominal control region identification selection.*

6.6 Unfolding

The observed signal yield in data needs to be corrected for detector effects to retrieve the amount of signal in the fiducial volume. Following the strategy described in Section 6.1, detector effects are encapsulated in correction factors (C_X) obtained from the simulation, and applied to the measured yields to translate them in terms of fiducial cross section. The fiducial cross section is then defined as follows:

$$\sigma_{X,\text{fid}} \cdot \mathcal{B}(X \to \gamma\gamma) = \frac{N_{\text{sig}}^{\text{reco}}}{C_X \cdot \mathcal{L}_{\text{int}}} \quad , \tag{6.9}$$

where $\mathcal{B}$ is the branching ratio, $N_{\text{sig}}^{\text{reco}}$ is the measured signal yield, $\mathcal{L}_{\text{int}}$ is the integrated luminosity and C_X is the correction factor.

This approach benefits from the fact that cross sections computed in a precise fiducial volume, defined by the analysis selection, are more model independent compared to total cross sections measurements. For instance, the acceptance for a hypothetical resonance produced in vector boson fusion (VBF) compared to one produced in ggF should be larger, since the topology of the events is different, and a larger fraction of decay products of this resonance produced in VBF ends in the acceptance of the detector. However, this difference is not that large when computing fiducial observables, since the measurements only considers particles that are already in the acceptance of the detector.

This analysis aims at reaching both the highest sensitivity on the fiducial cross section, and covering the lowest attainable mass possible. In this search, only ggF production is assumed since other modes are suppressed in comparison, as explained in Section 1.4.1

6.6.1 Fiducial volume definition and correction factor

To extract the fiducial cross section, the number of fitted signal events is corrected with the correction factor C_X defined as follows:

$$C_X = \frac{N_{\text{selection}}}{N_{\text{fiducial}}} \quad , \tag{6.10}$$

where $N_{\text{selection}}$ is the number of reconstructed signal events passing all analysis cuts, and N_{fiducial} is the number of signal events at particle-level generated within the fiducial volume.

The fiducial volume is defined to match, as closely as possible, the experimental acceptance of the measurement, given by the geometrical acceptance of the detector, the kinematic thresholds of reconstruction, and any other analysis kinematic selections. The fiducial volume is therefore defined by kinematic cuts applied on the truth (di)photon variables in order to mimic the selections applied at the reconstruction level. An overview of these selections is given in Table 6.5. Both photons should be within $|\eta^\gamma| < 2.37$, including the crack region between $1.37 < |\eta^\gamma| < 1.52$ to ease a posterior total acceptance correction. The transverse momentum of the diphoton system is required to pass equivalent cuts to those applied at reconstruction level, that is $p_{\text{T}}^{\gamma\gamma} > 50$ GeV .

The kinematic acceptance cuts applied on E_T^γ and $|\eta^\gamma|$ are not enough to define a model-independent fiducial volume in this analysis, given the significant dependence of photon isolation on the ΔR distance between two reconstructed photons. To reduce this dependency, a particle-level isolation cut is applied to the photons as part of the fiducial definition. The particle-level isolation energy is defined as the scalar sum of the transverse

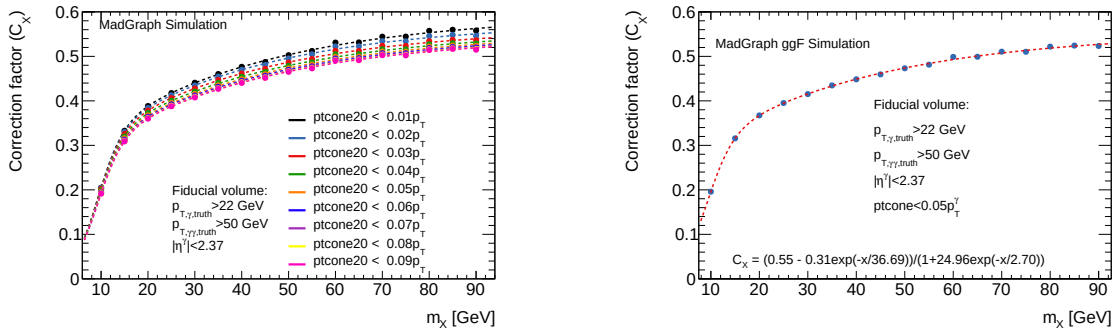
energy of all stable particles (except muons and neutrinos) found within a cone of given size around the truth photon. It is computed using a cone size of $\Delta R = 0.2$.

Table 6.5: *Particle-level selections for the fiducial measurements. The photon isolation, $\sum E_T^i$, is defined as the sum of the E_T of stable charged particles within $\Delta R < 0.2$ with respect to the photon.*

Object	Fiducial definition
Photons	$\lvert\eta^\gamma\rvert < 2.37$ (including crack)
	$E_T^\gamma > 22$ GeV
	$\sum E_T^i < 0.05 E_T^\gamma$ GeV
Diphotons	$p_T^{\gamma\gamma} > 50$ GeV

Unfolding studies between particle-level and reconstructed isolation were performed to find which particle-level selection matches best the applied working point used at reconstruction level [141]. Track isolation was shown to be tighter than calorimetric isolation, and almost equivalent to the track isolation applied at reconstruction, thus a similar cut is used to compute the events in the fiducial volume. Various particle-level isolation cuts are evaluated in Figure 6.31a. Tighter particle-level isolation cuts, compared to those applied at reconstruction level, show larger C_X factors since reducing the fiducial volume down to very small regions of the phase space reduces the kinematic phase space dependencies. Looser cuts compared to `ptcone20`$< 0.05 E_T^\gamma$, which correspond to the reconstruction level track isolation cut, saturate within few percent at similar C_X factors. The drop at lower masses is related to the drop in efficiency shown in Figure 6.13 from ID, isolation and trigger performances.

The C_X factor is computed from Equation 6.10, applying the full analysis selection to obtain the numerator and all fiducial cuts to obtain the denominator. The computation is performed for different values of m_X in the range from 10 GeV to 90 GeV. Because the results must be provided in finer mass steps than the resolution provided by the MC samples, an interpolation between these points is performed. The results are illustrated in Figure 6.31b where the C_X factor is shown as a function of $m_{\gamma\gamma}$.



a) b)

Figure 6.31: *(a) The correction factor, C_X, as a function of the resonance mass, for different particle-level isolation cuts. The cuts are shown in decreasing order from tightest in black to the loosest one in violet. (b) The correction factor, C_X, shown for ggF production mode as a function of the resonance mass.*

6.7 Systematic uncertainties

The sources of systematic uncertainties considered in this analysis can be separated into two groups: those which primarily affect the signal yield, like efficiencies, fit biases and acceptance related uncertainties; and those which mainly affect the shape of the signal parametrization, arising mainly from the photon energy calibration measurements. The systematics affecting the signal yield are furthermore divided into three groups: systematics with experimental sources, theoretical uncertainties and systematics arising background modeling. A special emphasis is given to the last one of this group, since it is the largest contribution to the total uncertainty. The various sources and measured values are discussed in this section.

6.7.1 Signal shape uncertainties

The photon energy scale and photon energy resolution uncertainties have a non negligible impact on the signal shape, and their impact vary with $m_{\gamma\gamma}$. Photon energy scale variations translate into shifts in the peak position while photon energy resolution variations contribute to shrinking/widening the distribution.

The effect of both uncertainties in the signal shape is determined from a reprocessed simulation, in which variations by $\pm 1\sigma$ of the energy scale and resolution (Section 3.2) are used instead of the nominal ones. The reconstructed invariant mass distributions are parametrized using the same signal model as the one explained in Section 6.4, fixing all the parameters of the model except for the peak position μ (the width σ) when evaluating the uncertainty related to the photon energy scale (resolution).

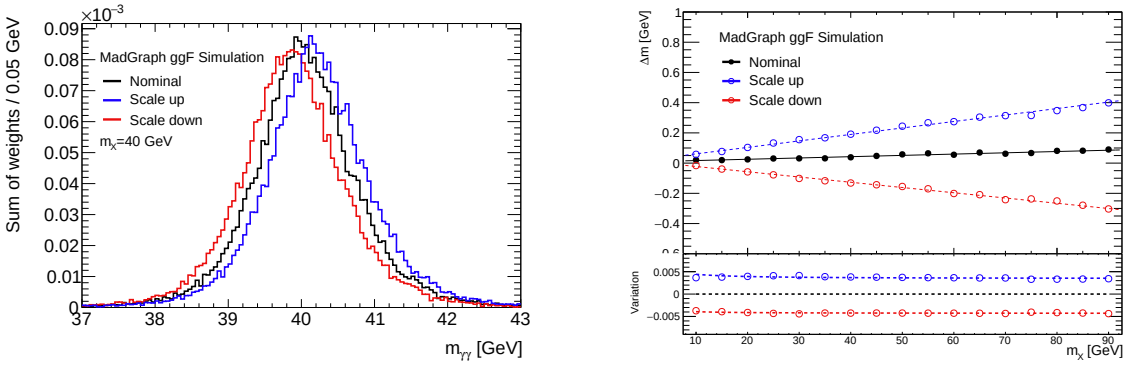
The measured peak position for the reprocessed samples, $\mu^{up/down}$, is compared to the values obtained in Section 6.4. The systematic effect on the "mass scale" is then estimated by quantifying the shift of the peak position with respect to the nominal value. The effect of changing the photon energy scale by $\pm 1\sigma$ for each simulated signal is shown in Figure 6.32a. The uncertainty in the signal shape peak position related to the photon energy scale is found to be below 0.5% and reasonably independent of the mass (Figure 6.32b). This systematic is neglected in the following.

The measured width for the reprocessed samples, $\sigma^{up/down}$, is also compared to the values extracted in Section 6.4. The systematic effect on the mass resolution is then estimated by quantifying the difference on the width for both variations with respect to the nominal value. The effect of $\pm 1\sigma$ on the photon energy resolution for a given mass point is shown in Figure 6.33a. The uncertainty related to the photon energy resolution is found to be of the order of 5%, slightly decreasing towards lower masses (Figure 6.33b).

The relative differences for both parameters are directly translated into systematic uncertainties on the μ and σ parameters of the model.

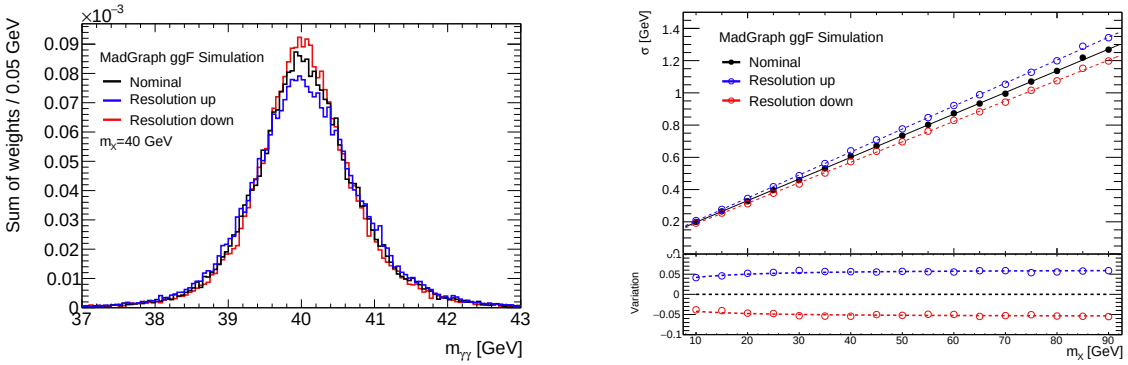
6.7.2 Signal yield uncertainties from experimental sources

The following systematic variations have a negligible effect on the signal shape and only its effect on the measured signal yield is considered.



a) b)

Figure 6.32: *(a) Invariant mass distribution for the simulated signalt of $m_{\gamma\gamma} = 40$ GeV obtained with the nominal (black line) and with systematically varied energy scale variations, being the upper variation in blue and the lower variation in red. (b) The shift Δm as a function of the mass m_X for the nominal (black markers) and systematically varied energy scale variations, being the upper variation in blue and the lower variation in red. The bottom panel shows the relative difference of the peak position $((\mu^{up/down} - \mu)/\mu)$ between the variations and the nominal sample distribution.*



a) b)

Figure 6.33: *(a)Invariant mass distribution for the signal point of mass 40 GeV obtained with the nominal (black line) and systematically varied energy resolution variations, being the upper variation in blue and the lower variation in red. (b)The width σ as a function of the mass m_X for the nominal (black markers) and systematically varied energy scale variations, being the upper variation in blue and the lower variation in red. The bottom panel shows the relative difference of the width $((\sigma^{up/down} - \sigma)/\sigma)$ between the variations and the nominal sample distribution.*

Luminosity measurement

The relative uncertainty on the full Run2 dataset luminosity is 1.7% [137]. This uncertainty is extracted from the calibration of the luminosity scale using Van Der Meer scans [142] with the LUCID [143] luminosity detectors.

Photon identification

The photon identification efficiency was measured on 2015–2018 data using three data-driven methods and is used to compute data-to-MC ratios (scale factors) separately for each method and also combined [110]. The associated up and down uncertainty of the scale factors is propagated to the analysis by applying the corresponding variation.

Photon isolation

The uncertainty of the photon isolation efficiency has contributions from both the calorimeter and charged-particle tracking portions of the requirement. The systematic uncertainty on the yield is obtained by applying (i) a data-driven shift to the calorimeter isolation and (ii) a E_T^γ-dependent shift to the track isolation [110] The shifts on the correction factor are calculated separately and combined in quadrature.

Photon trigger

The trigger efficiency is measured with data for single-photon triggers and is found to be well modeled in the MC simulation. For the diphoton triggers used in the analysis, efficiency variations detailed in [102] contribute to a total uncertainty of 0.5%. The systematic uncertainty assigned to the photon trigger is taken to be the uncertainty of this estimation.

Pile-up reweighting

A variation in the pileup reweighting of the simulation is performed to cover the difference between the predicted and measured inelastic pp cross-section in the fiducial volume of the detector [93]. This is achieved by shifting the data μ distribution by $\pm 3\%$ before reweighting the MC sample.

Figure 6.34 shows the evolution of the experimental uncertainties described above as a function of m_X, with each point corresponding to the uncertainty computed with a signal sample at the corresponding mass.

6.7.3 Theoretical uncertainties

For this analysis, only gluon-fusion simulated samples have been used, since ggF is the relevant production mode for the pursued theoretical model. This assumption implies that the signal follows a certain set of kinematic distributions, which then have a non-negligible effect on the signal efficiencies (i.e. a resonance that is produced with larger p_{T}^γ than what is expected in the simulation would translate into underestimated signal efficiencies). However, mismodellings on the kinematic variables of the simulated signal could induce biases on the correction factors. For this reason, a conservative uncertainty can be implemented by varying the $p_{\mathrm{T}}^{\gamma\gamma}$ selection and evaluating its effect on the correction factor. The impact on the C_X factor from an arbitrary variation of $\pm 10\%$ in the $p_{\mathrm{T}}^{\gamma\gamma}$ cut is shown in Figure 6.35. These conservative variations translate into differences of up to $\pm 5\%$ in the correction factor and are included as systematic uncertainties.

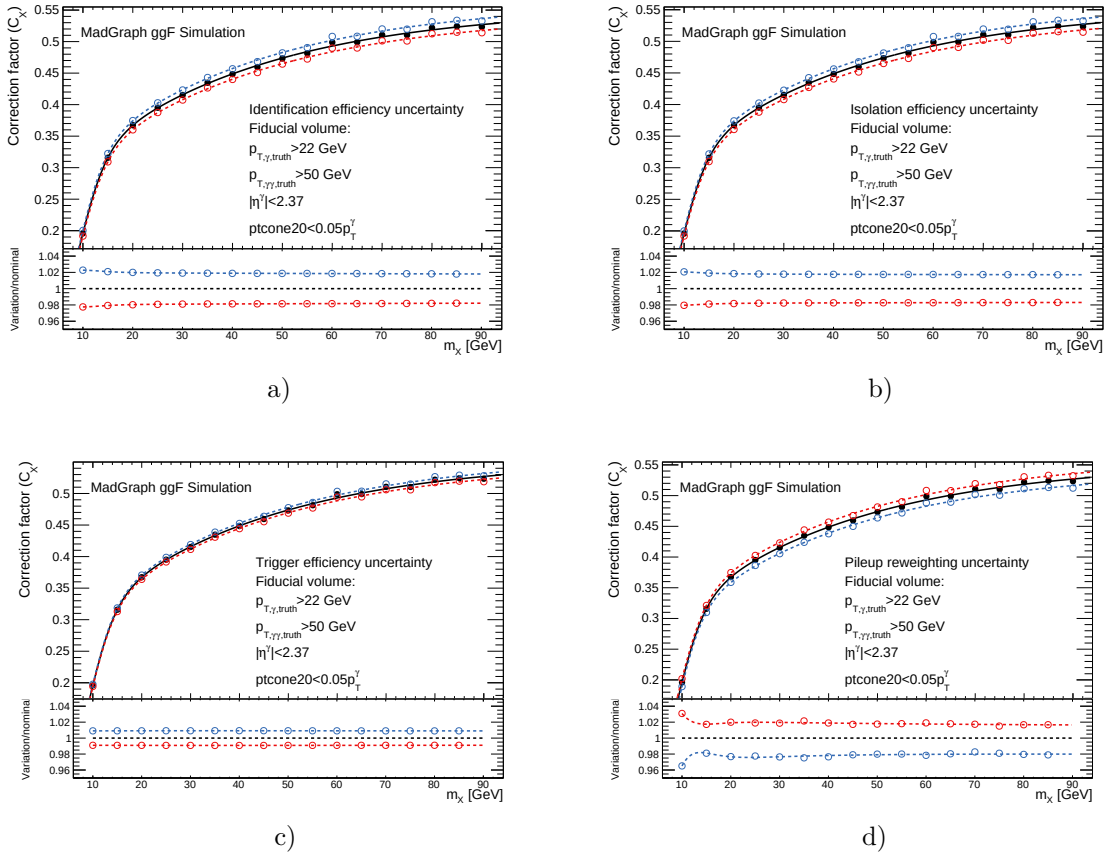


Figure 6.34: *Effect of several sources of experimental uncertainty (photon identification (upper left), isolation(upper right), trigger (lower left) and pile-up reweighting (upper right)) on the determination of the correction factor shown as a function of m_X.*

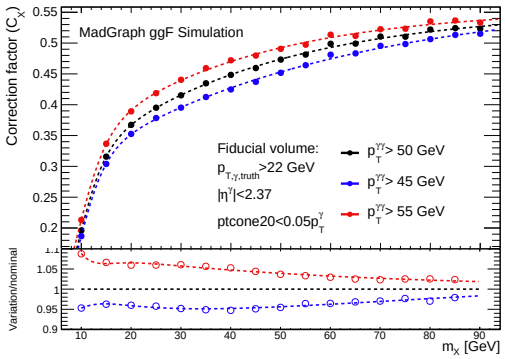


Figure 6.35: *Conservative systematic uncertainties on the correction factor arising from variations on the $p_{\mathrm{T}}^{\gamma\gamma}$ cut evaluated on signal simulated samples.*

6.7.4 Background modeling systematics

The true shape of the non-resonant background distribution in data is unknown and the strategy chosen in this analysis is to describe it with an analytical function with a finite number of parameters. However, a particular functional form may not provide a sufficiently good description of the background shape, leading to a fit bias denominated *spurious signal*. Prior to the estimation of this systematic, an introduction is given.

6.7.4.1 Spurious signal: definition

The fit bias or spurious signal can be defined as the difference between the median fitted signal yield and the expected signal yield. It is not exactly the difference between the true background shape and the functional form used to describe it since it also depends on the shape of the signal being looked for, though it is largely correlated. A precise estimation of this bias is important in order to account for the systematic error driven by a particular choice of a function to describe the background. The overall concept is illustrated in Figure 6.36.

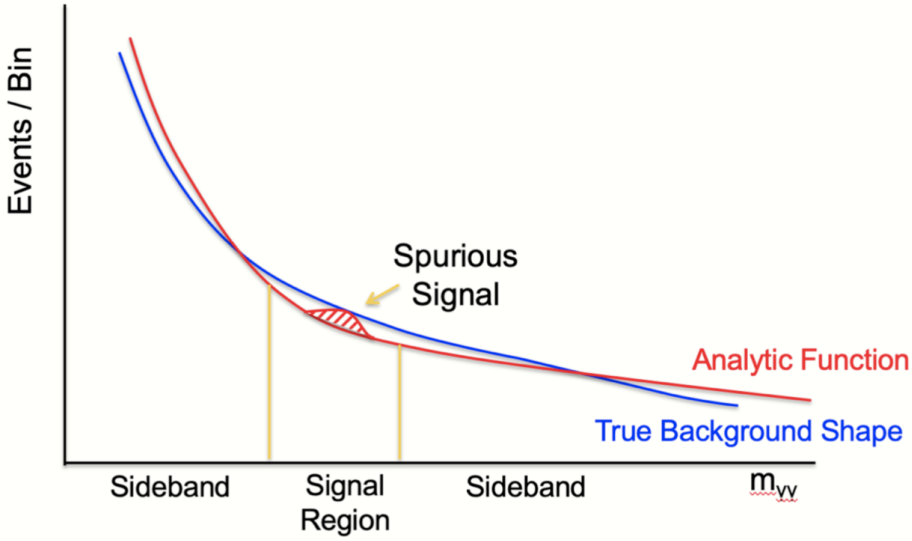


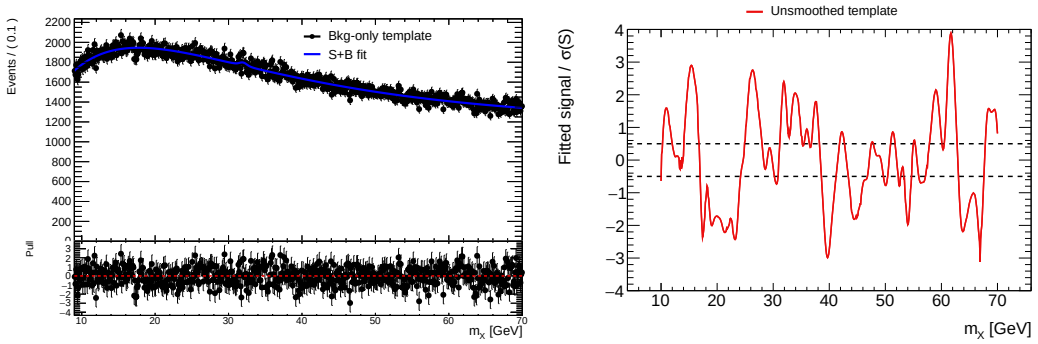
Figure 6.36: *Cartoon illustrating the concept of spurious signal. The blue line represents the true background shape while the red line is the functional form used to describe it. For a given signal hypothesis, with fixed mass and width, the spurious signal in this figure is the fitted signal yield, represented by the red dashed region [144].*

In order to estimate the spurious signal uncertainty assigned to a given functional form, signal plus background fits are performed on the background-only template for different mass hypotheses. The signal parametrization obtained in Section 6.4 is used for the signal model and the functional form detailed in Section 6.5.5 is used to describe the background. Since no signal is injected, the fitted signal yield as a function of the mass N_{SS} is our best estimate of the spurious signal systematic.

This procedure requires either very-large statistic templates or smoothed versions of them. Otherwise, the fitted signal yield would be heavily affected by statistical fluctuations and it would provide a poor (if any) estimate of the spurious signal. An example of this issue is shown in Figure 6.37. The smoothing techniques used in this analysis are explained in the following subsection.

6.7.4.2 Smoothing techniques

Since very-large statistic templates usually require expensive computational resources to simulate events, smoothing techniques have been used within various ATLAS analyses. In this analysis, the smoothing of the background template is performed using Gaussian Process Regression (GPR) [145]. This technique allows to smooth binned data by exploiting



a) b)

Figure 6.37: *(a) Signal plus background fit to the background-only template. This example shows a hypothetical signal at $m_X = 32$ GeV. The fitted signal yield arises from a statistical fluctuation. (b) Fitted signal yields over their uncertainties as a function of the mass m_X for the unsmoothed template. The statistical fluctuations from the template are reflected also in this figure. The dashed lines represent the target systematic of the analysis.*

the correlation between each pair of bins. For instance, since the error of each bin follows a Poisson distribution (in the case of unweighted events), for smoothly falling backgrounds, bins at higher masses will have smaller errors hence smaller bin contents. So the content of a given bin i can be estimated assuming the surrounding bins as constraints. The bin contents ν_i are constrained by the following Gaussian PDF:

$$G(\nu_i; r_i, C) = \frac{1}{\sqrt{(2\pi)^N |C|}} exp\left[-\frac{1}{2}\sum_{i,j=1}^{N}(r_i - \nu_i)C_{ij}^{-1}(r_j - \nu_j)\right], \tag{6.11}$$

where C is the covariance matrix between the bins of the template and r_i are the reference values (prior) for the bin contents (the content of the) The elements of C can be defined as a smooth function of the bin centers m_i, m_j (for example mass values in our case). The function $C_{ij} = K(m_i, m_j)$ is called the kernel function and essentially encodes the level of correlation between to distinct points. An intuitive example for a kernel function is the Radial Basis Function (RBF):

$$K(m_i, m_j) = exp\left(-\frac{1}{2}d(m_i/, m_j/l)^2\right), \tag{6.12}$$

where $d(a, b)$ represents the Euclidean distance between points a and b and l is called the length scale. The length scale defines the characteristic distance in m_i within which two bin contents are more or less correlated. Bins separated by a distance larger than the length scale have their contents uncorrelated. Sensible values of l should be larger than the width of the signal that is being searched for but smaller than the typical scale of the variations of the background shape. Other popular kernel function is the Gibbs kernel [146, 147]. An example of a smoothed template using GPR is illustrated in Figure 6.38.

6.7.4.3 Spurious signal estimation

Since the smoothed template is less sensitive to statistical fluctuations, it is used to provide a better estimate of the spurious signal systematic. The fitted signal yield as a

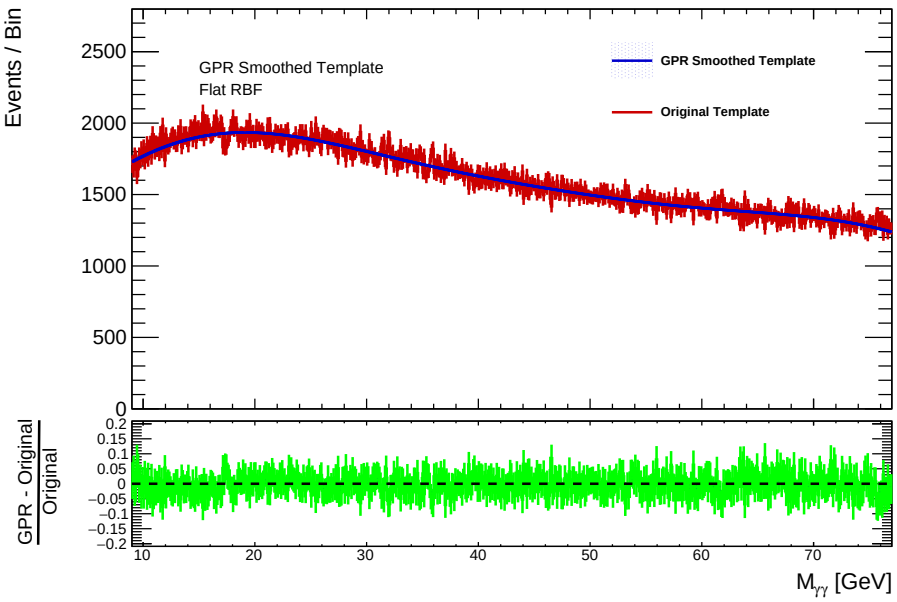


Figure 6.38: *Smoothed template using Gaussian Process Regression. The unsmoothed background template is shown in red and the smoothed result is shown in dark blue. The relative difference between them is shown in green below.*

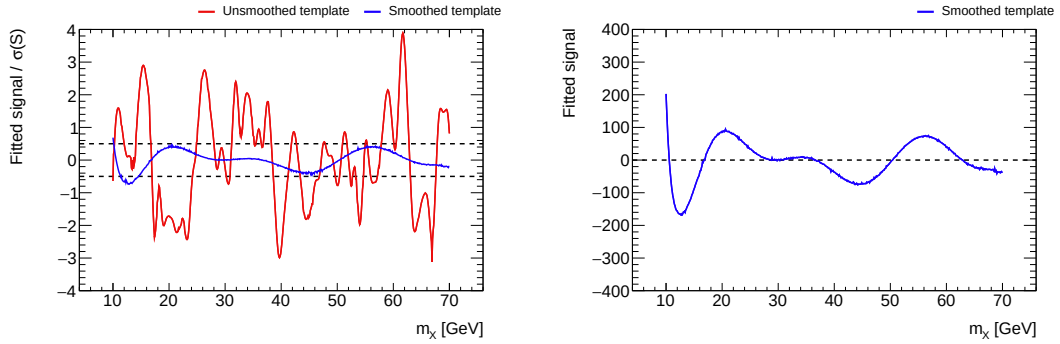
function of the mass of different signal hypotheses is shown in Figure 6.39a. Compared to Figure 6.37, the improvement on the estimation of the mismodelling arising from the choice of the function is evident. In general, the measured spurious signal (Figure 6.39b) for a given model can be used as a criterion to choose between different functional forms to describe the background. This procedure is referred to as spurious signal criterion and it consists on comparing the spurious signal of different models in order to choose the one that provides the lowest fit bias.

The value at which the measured spurious signal is sufficiently low is partly arbitrary and analysis-dependent, so no general spurious signal criterion exists to decide whether a model describes or not a given template. Usually, a compromise is found compared to the statistical error of the analysis. However, this is not always possible or easy since it depends strongly on the shape of the background. In this analysis, the aim is to achieve a functional form which leads to a spurious signal of less than 50% of the statistical uncertainty $N_{SS} < 0.5\sigma_{stat}$.

The increase of the statistics in the simulated sample (from 15M events to 200M) will improve the estimation of this systematic, since the contribution of statistical fluctuations will be reduced by almost a factor 4.

6.7.5 Systematics summary

Table 6.6 below shows a summary of the systematics accounted for in this analysis. Given the amount of statistics expected in data (10k events/GeV), leading to an statistical error of $\sigma_{stat} \sim 100$ events, this analysis is systematics-dominated due to the large uncertainty assigned to the spurious signal systematic $N_{SS} \sim 150$ events.



a) b)

Figure 6.39: *(a) Fitted signal yields over their uncertainties as a function of the mass m_X for the unsmoothed (red) and smoothed templates. (b) Fitted signal yields as a function of the mass m_X. The dashed lines represent the target systematic of the analysis.*

Table 6.6: *Summary of the main sources of systematic uncertainty for the measurement of σ_{fid} for the resonance search analysis. Unless written otherwise, numbers are mass independent.*

Source	Uncertainty
Signal yield	
Theoretical uncertainty	$^{+9}_{-5}$% (at 10 GeV) – 2.0% (beyond 85 GeV)
Pile-up reweighting	± 3.5% (at 10 GeV) – 2 % (beyond 15 GeV), mass dependent
Photon identification	±2.0%
Isolation efficiency	±2.0%
Trigger	±1.0%
Luminosity (2015–2018)	±1.7%
Photon energy scale/resolution	negligible
Signal modeling	
Photon energy resolution	$^{+6\%}_{-5\%}$
Photon energy scale	negligible
Pile-up reweighting	negligible
Background	
Spurious signal	$0.5\sigma_{\text{stat}}$ 160-110 events, NWA

6.8 Statistical analysis

An introduction to statistics in High Energy Physics and, together with the necessary statistical tools used in this analysis to extract results, is detailed in Appendix A.

Signal and background yields are estimated from an unbinned maximum-likelihood fit of the signal-plus-background model to the data. The signal parametrization is the one described in Section 6.4 and the background model is obtained with the procedure discussed in Section 6.5. The systematics were summarized in Section 6.7 and are included in the fits via nuisance parameters constrained by Gaussian penalty terms. The likelihood can be written as:

$$\begin{aligned}\mathcal{F}(m_{\gamma\gamma};\sigma_{fid},m_X,N_b,\boldsymbol{a},\boldsymbol{\theta}) &= N_X(\sigma_{fid},m_X,\boldsymbol{\theta}_{N_X})\cdot Sig(m_{\gamma\gamma};m_X,\theta_\sigma)\\ &+ N_b\cdot Bkg(m_{\gamma\gamma};\boldsymbol{a})\end{aligned} \tag{6.13}$$

where Sig, Bkg and N_X, N_b denote the PDFs and corresponding yields for the signal X and background components, σ_{fid} is the (fiducial) cross-section times branching ratio of the hypothetical resonance of mass m_X decaying to diphotons, $\boldsymbol{a}$ denotes the set of background shape parameters, $\boldsymbol{\alpha}_X$ denotes the set of signal shape parameters, σ_θ denotes the systematic uncertainty related to the energy resolution scale and $\boldsymbol{\theta}_{N_X}$ is the set of nuisance parameters for systematic uncertainties affecting N_X, as listed below:

- θ_{lumi}: uncertainty on the integrated luminosity of the data sample;
- θ_{eff}, θ_{isol} , θ_{PRW} : systematic uncertainties arising from photon identification and isolation efficiencies of the signal, and on the pileup reweighting;
- θ_{SS}: spurious signal systematic;
- θ_{ES}: photon energy scale systematics;

The quantities N_X and N_b are the number of events of the hypothetical resonant signal and the background components respectively. The background yield N_b is a free parameter in the fit, while the signal yield N_X is parametrized as:

$$N_X(\sigma, m_X, \boldsymbol{\theta}_{N_X}, \theta_{SS}) = \sigma \cdot \mathcal{L}_{\text{int}} \cdot C_X(m_X) \cdot \prod_{k}^{\dim \boldsymbol{\theta}_{N_x}} K_k(\theta_k) + \sigma_{SS} \cdot \theta_{SS}, \tag{6.14}$$

where $\mathcal{L}_{\text{int}}$ is the integrated luminosity of the sample; $C_X(m_X)$ is the value of the C_X factor for the considered mass m_X, as described in Section 6.6, K_k is a function characterizing the effect of the k-th normalization systematic and $\sigma_{SS}(= |N_{SS}|)$ and θ_{SS} are the values of the background modelling uncertainty and its associated nuisance parameter respectively.

The index k runs over the set of systematic uncertainties affecting N_X, so $K_i(\theta_i)$ factors implement each of the systematic uncertainties on the number of signal events as listed above. The expression $K_k(\theta_k)$ has the following form:

$$K_k(\theta_k) = [r_k(m_X)]^{\theta_k} \quad \text{with} \quad r_k(m_X) = \begin{cases} N_{X,+k}(m_X)/N_X(m_X), & \text{if } \theta_k > 0 \\ N_X(m_X)/N_{X,-k}(m_X), & \text{if } \theta_k < 0 \end{cases} \tag{6.15}$$

where $\pm k$ represent the positive/negative systematic variations on N_X for a given systematic i. This expression ensures that the modifications to the signal event yield for $\theta_k = \pm 1$ are exactly equal to the $\pm 1\sigma$ variations used to define the uncertainties. The actual expression for r_k is interpolated smoothly between the cases $\theta_k > 0$ and $\theta_k < 0$ to avoid numerical problems at $\theta_k = 0$.

The PDFs f_X and f_b describe the signal and background shape respectively. The signal PDF, $f_X(m_{\gamma\gamma}, \boldsymbol{x}_X(m_X, \alpha_X), \boldsymbol{\theta}_\sigma)$, is the double-sided Crystal Ball function with parameters $\boldsymbol{x}_X = \{m_{DSCB} = m_X + \Delta m, \sigma_{DSCB}, \alpha_{\text{low}}, \alpha_{\text{high}}, n_{\text{low}}, n_{\text{high}}\}$ computed as a function of m_X, as explained in the same section. The uncertainty on the DSCB width is implemented by using the expression $\sigma_{DSCB} = \sigma_0 \, K_\sigma(\theta_\sigma)$, where σ_0 is the nominal width, and $K_\sigma(\theta_\sigma)$ is defined similarly as for the event yield systematics above. The uncertainty on the position of the peak is defined analogously. The background PDF $f_b(m_{\gamma\gamma}; \boldsymbol{a})$ is the function from Eq. 6.8 in Section 6.5.5 described by the parameters $\boldsymbol{a}$ which are free in the fit.

The overall likelihood, including extended and constraint terms, reads:

$$\begin{aligned}&\mathcal{L}(\sigma, m_X, \alpha_X, N_b, \boldsymbol{a}, \boldsymbol{\theta}) = \\ &e^{-(N_X+N_b)} \left[\prod_{i=1}^{n} \mathcal{F}(m_{\gamma\gamma i}; \sigma, m_X, \alpha_X, N_b, \boldsymbol{a}, \boldsymbol{\theta})\right] \left[\prod_{k=1}^{\dim \boldsymbol{\theta}} \exp\left(-\tfrac{1}{2}(\theta_k - \theta_k^{\text{aux}})^2\right)\right]\end{aligned} \tag{6.16}$$

where n is the number of events in the dataset, and $\boldsymbol{\theta}^{\text{aux}}$ the set of auxiliary measurements used to constrain the systematic uncertainties.

6.9 Range of the search

The aim of this analysis is to extend as much as possible the lower mass edge of the limit on the production of new resonances extracted diphoton resonance searches. The range of the search is however a function of every previous step that characterize the shape of the background, like the fit with a functional form or the spurious signal estimation. In this section, the different elements that can limit the lowest attainable mass are explicitly discussed.

The first element under consideration is the estimation of the background shape from the available samples. In the presented strategy, the background template is built by combining simulated and data samples and it is expected to reproduce a similar shape to the one in data in the signal region. This statement, however, is not falsifiable without unblinding the data and thus the shape of the background is extracted from a fit with an analytical function. The particular shape of the invariant mass distribution at low masses, with the turn-on characterized as explained in the previous Section, is main limiting factor. The first range is then delimited by the capability of the proposed functional form to describe the background shape and in particular the turn-on region.

The second element under consideration is the level of spurious signal for different fit ranges. While a functional form can sufficiently describe a given range in $m_{\gamma\gamma}$, it can happen that the evaluation of the spurious signal systematic in the same range induces large uncertainties that would unavoidably worsen the results. The range is then determined by the trade of between sensible spurious signal systematics and the lowest mass possible.

Finally, the range of the search consists in a reduced range from the one provided by the spurious signal evaluation, to reduce "edge-effects". This analysis benefits from the good resolution at low masses (at 10 GeV the resolution is about $\sim$200 MeV) and can possibly set values a few GeV above the lower edge.

At higher masses, there are two effects that contribute to the worsening of the description of the background: the resonant Z boson and the remnants of the trigger turn-on. These two effects induce a change in the concavity of the background around 75 GeV, whose description would complicate significantly the strategy of the fit. With the aim of providing results in a common region in $m_{\gamma\gamma}$ with previous analyses, which set limits down to 65 GeV, the upper range of the fit is delimited by the quality of the fit without a special description of the aforementioned features.

6.10 Expected sensitivity

The expected limits on the fiducial production cross-section of a hypothetical resonance times its branching ratio to diphoton are shown in the following. The results of this

analysis are shown in this section, followed by an interpretation of the expected sensitivity, in terms of constraints on the decay constant of an ALP within the framework of an Effective Field Theory (more details explained in Section 1.4).

This expected upper limit on the number of signal events is computed from signal plus background fit of a background-only Asimov dataset (Appendix A). The computation is analogous to the example shown in Appendix A, Figure A.4, but performing a scan over a range in $m_{\gamma\gamma}$ (in Figure A.4, the Confidence Level limit on the allowed number of events is computed for a single value of horizontal axis $x = 2$, being in that case the limit around 47 events). The limit on the number of events is transformed into a limit on the fiducial cross-section times branching ration using Equation 6.9.

The expected sensitivity is evaluated for masses in the [10,75] GeV range. The expected Confidence Level at 95% on $\sigma_{fid} \times \mathcal{B}(X \to \gamma\gamma)$ are shown in Figure 6.40. Since the value of the spurious signal systematic will be re-evaluated with the extension of the SHERPA $\gamma\gamma$samples, a few spurious signal variations are overlaid.

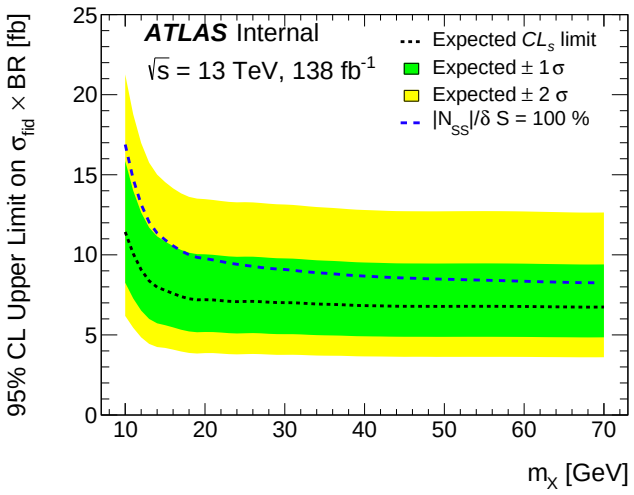


Figure 6.40: *The expected limit on the fiducial production cross-section times branching ratio to diphotons for a narrow scalar as a function of its mass m_X, considering all systematic uncertainties and an integrated luminosity of 138 fb^{-1}. The expected upper limit is shown as a dashed black line with its ± 1 and ± 2 standard deviation bands. The dashed blue line denotes the same limit but assuming a flat, 100% spurious signal systematic.*

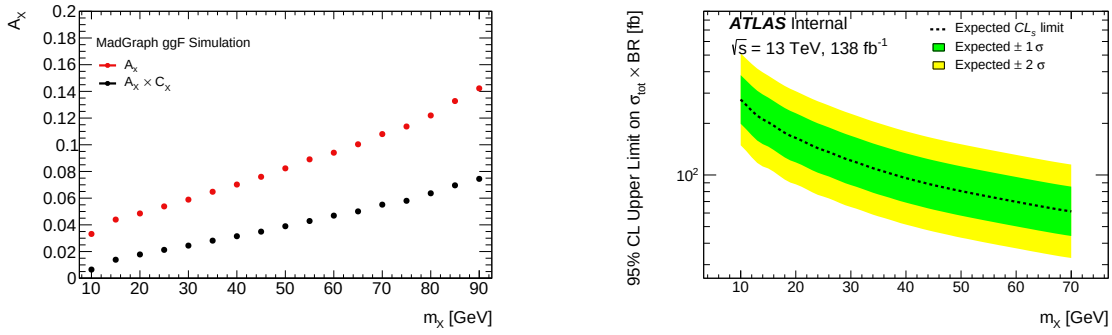
6.10.1 Phenomenological interpretation: limits on f_a

Various BSM models include resonances in the mass range of the spectrum below the mass of the Higgs which could decay to diphotons. The coupling strength of these resonances, arising from spontaneously broken symmetries, is governed by the decay constant f_a. An interpretation of the expected limit shown in Figure 6.40 is shown in this section. This is performed given that the production cross-section of such a resonance decaying to two photons scales with the inverse of the square of f_a:

$$\sigma(pp \to X \to \gamma\gamma) \propto \frac{f(m_X)}{f_a^2}, \tag{6.17}$$

where $f(m_X)$ is an arbitrary function of the mass of the resonance m_X , independent of f_a.

For this purpose, the total production cross-section instead of the fiducial cross-section is computed, by including the acceptance A_X in the denominator of Equation 6.9. On top of the correction factor used to compute the fiducial cross-section, the acceptance of the truth-level selection A_X is extracted as a function of the mass of the scalar m_X(see Figure 6.41a). The decrease in acceptance towards lower masses is driven by the boosted selection. Limits on the total cross-section times branching ratio are shown in Figure 6.41b.



a) b)

Figure 6.41: *(a) The acceptance factor, A_X, shown for ggF production mode as a function of the mass of the resonant signal. (b) The expected limit on the total production cross-section times branching ratio to diphotons for a scalar produced in ggF as a function of the resonance mass m_X, considering all systematic uncertainties and for an integrated luminosity of 138 fb^{-1}. The expected upper limit is shown as a dashed black line with it ± 1 and ± 2 standard deviation bands.*

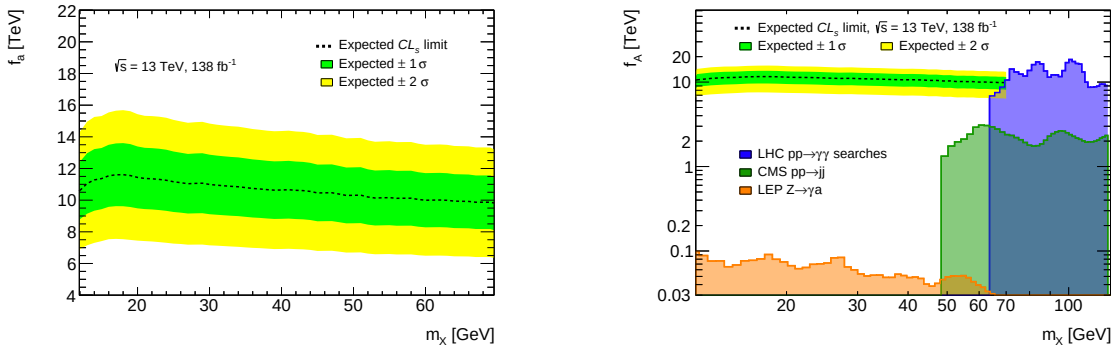
To compute the limit in terms of the decay constant $f_{a,exp}$, the total cross-section computed in [76] as a function of $m_X$for a given $f_{a,theo}$; $\sigma(pp \to X \to \gamma\gamma)_{theo}$, is rescaled by means of Equation 6.17:

$$\alpha_{COM}\frac{\sigma(pp \to X \to \gamma\gamma)_{exp,13}}{\sigma(pp \to X \to \gamma\gamma)_{theo,14}} = \frac{f^2_{a,theo}}{f^2_{a,exp}} \overset{f_{a,theo}=1}{\longrightarrow} f_{a,exp} = \sqrt{\alpha_{COM}\frac{\sigma(pp \to X \to \gamma\gamma)_{theo,14}}{\sigma(pp \to X \to \gamma\gamma)_{exp,13}}}, \quad (6.18)$$

where α_{COM} is the rescaling coefficient between the different center-of-mass energies of the data collected and the values computed in [76]. The expected limits on the decay constant f_a are shown in Figure 6.42a, where the excluded region is the one found **below** the curve (opposite to Figure 6.41b for instance). The derived constraints are compared in Figure 6.42b with those obtained from previous experiments [70, 125, 126, 148]. This analysis provides the strongest sensitivity to a hypothetical resonance produced in gluon fusion that decays to two photons in this mass region, whose precise limits are qualitatively described in Section 6.9.

6.11 Conclusion

In this chapter, the analysis strategy and sensitivity for a diphoton resonance search covering masses below 65 GeV are presented. The restriction of the phase space by



a) b)

Figure 6.42: *(a) The expected limit on the decay constant f_a as a function of the resonance mass m_X, considering all systematic uncertainties and an integrated luminosity of 138 fb^{-1}. (b) Compilation of exclusion limits at 95% CL on the decay constant f_a as a function of the resonance mass m_X obtained by different experiments. The existing limits, derived from [70, 125, 126, 148], are compared with the expected limits of this analysis. The excluded region in both plots is found below the curve.*

requiring exclusively collimated photon pairs eases the description of the background shape with analytical functions and thus allows to attain lower diphoton masses. Using simulated and data-driven samples, the shape of the background is estimated. The shape of the background is used to compute the expected sensitivity of this search, which is then compared to previous analyses. In the common part of the mass spectrum, the expected constraints are competitive with the ones obtained in more inclusive resonance searches [125, 126], despite the fact that the selection reduces the signal efficiencies by a factor 5.

A prospective study for future resonance searches in the diphoton channel during Run 3 is presented in Appendix E. The analyses to be performed in future data-taking periods will benefit from lower E_T^γ thresholds in exchange of topological restrictions, like requiring photon pairs to be back-to-back in the detector. This possibility completes the picture, providing an intermediate strategy between to the current results obtained with inclusive searches and the analysis presented in this manuscript.

The analysis strategy on its current state is undergoing an internal review within the ATLAS collaboration before unblinding the data. Since this analysis targets Moriond 2021, the approval and preparations towards a publication will hopefully rapidly follow.

Conclusions

The observation of the Higgs boson in 2012 by the ATLAS and CMS collaborations represents a milestone in the recent history of Science, and since then it has motivated a large number of searches aiming to describe the unexplained pieces beyond the Standard Model. The diphoton decay channel, together with $H \to 4l$, played a crucial role in its discovery, highlighting the outstanding performance and resolution of the electromagnetic LAr calorimeter of the ATLAS experiment. However, the worsening of photon reconstruction under the larger pileup environments foreseen in future data-taking periods, both Run 3 and High-Luminosity LHC, motivate different strategies that could help mitigating the impact on performances. This work covers both photon performances in large-pileup environments and a search for new resonances in the diphoton channel. The studies presented in this manuscript are performed with data collected by the ATLAS detector from 2015 to 2018 at $\sqrt{s} = 13$ TeV, representing 139 fb^{-1}.

The first study addresses the validation of photon isolation performances on data and simulation. The methodology used relies on the subtraction of fake photons using data-driven techniques to estimate the amount of signal in data. The results presented in this work entail the extraction of coefficients denominated "scale factors" that encapsulate data-simulation differences on the photon isolation efficiencies, necessary for correcting simulation predictions in every ATLAS analysis. The scale factors vary with the energy and region of the detector, being larger for photons with $E_T^\gamma < 60$ GeV. The systematic uncertainties decrease with increasing E_T^γ from 2% at $E_T^\gamma \sim 22$ GeV to less than 0.1% for $E_T^\gamma > 200$ GeV. The observed discrepancies for low energy photons between data and simulation have motivated a more precise study of the shape of the isolation distribution to disentangle the origins of such differences, leading in the close future to a more robust procedure to correct the simulation.

The second study explores new techniques to maintain an acceptable level of photon isolation performances with increasing pileup. Among the different photon identification techniques in place, photon calorimetric isolation is shown to be heavily affected by simultaneous *pp* interactions. Additional calorimetric information such as the longitudinal development of the shower and the time of each energy deposit in the calorimeter are exploited in a novel analysis to compute a likelihood weight that reflects the probability of an energy deposit to originate from pileup or from a fake. This original study is presented as a promising way of improving the resolution of the isolation energy for photons. From a broader point of view, the methodology presented can also be extrapolated to the trigger and reconstruction of other objects that deal with spurious energy deposits arising from pileup interactions. Future isolation studies can also benefit from the presented results motivating new techniques that span from the inclusion of additional variables with discriminating power against pileup energy deposits, such as timing information, to the usage of advanced machine learning techniques to identify prompt photons. The methodology shown in this manuscript has also proven to be extremely useful to under-

stand the reconstruction of particles with the electromagnetic calorimeter and identify instrumental defects with non-negligible impact on photon identification.

The final chapter describes an axion-like particle resonance search in the diphoton channel covering masses below 65 GeV. Experimentally, diphoton resonance searches covering lower diphoton masses at hadron colliders are limited by the trigger energy thresholds, complicating the description of the background shape. The selection of collimated photons in the detector allows to attain masses that are inaccessible to inclusive analyses without accounting for large systematic uncertainties. This search is naturally connected to the two previous studies presented in this manuscript since it deals with collimated pairs of photons and thus it is heavily dependent on photon isolation techniques. When evaluated on data, the results obtained with this analysis strategy will provide the strongest constraints on a poorly explored region with pp collisions at the LHC.

In the scenario of a resonance being observed, its impact in the theoretical landscape would be enormous, since such resonance could be explained by a light pNGB generated from the spontaneous breaking of a global U(1) symmetry, ubiquitous in common theoretical frameworks such as Supersymmetry, composite Higgs models or simplified dark matter models with light mediators connecting the Standard Model with dark sectors. Future analyses will benefit from triggers that include topological selections between the two photon candidates to reduce the trigger rates and thus giving the possibility to lower the E_T^γ thresholds. This strategy will allow to explore masses down to 45 GeV in non-collimated topologies, providing a complementary result to the one shown in this document.

Appendix A

Statistical tools in High Energy Physics analyses

Physicists, among others, aim to infer properties from Nature through observations and measurements combined with theoretical understanding. Theoretical input often comes in form of a model used to interpret a set of measured quantities and built in such a way that its parameters are sensitive to specific properties of Nature.

Statistical methods enter in the game as a tool that helps to interpret measurements and provide sensible values for parameters of the model, accounting for different uncertainties that could impact or bias the estimation of such parameters. In this chapter the necessary statistical tools to perform statistical analyses in HEP are described together with examples in to help the reader in this intricate matter.

A.1 Basic concepts, Probability Density Functions and Likelihoods

Random processes and probability are the two main building blocks of the statistical tools used to perform HEP analyses. A process is said to be random if its outcome cannot be predicted with complete certainty, only the probabilities of the possible outcomes can be known. The mathematical definition of probability was given in 1933 by Kolmogorov [149] by means of simple axioms based on set theory. However, these basic axioms are often not the best tool to perform probability measurements and more restrictive definitions of probability have been defined, the most popular ones are the *frequentist* probability and the *Bayesian* probability.

The *frequentist* interpretation of probability, widely used in HEP and in this thesis, defines the probability as the limiting ratio of the times an event X is obtained as outcome of N infinite and equal experiments:

$$\mathcal{P}(X) = \lim_{N\to\infty} \frac{N(X)}{N}. \tag{A.1}$$

This definition lies on the basis of *repeatable* experiments with *exact* conditions, which slightly restricts the phenomena to which it can be applied. Physicists take care that all relevant conditions in experiments remain as constant as possible in order to obtain reproducible results.

The *Bayesian* approach on the other hand, abandons the concept of frequency in order to account for non-reproducible experiments. The basis of the Bayesian probability is the *degree of belief* of a person on a particular event to occur. A definition of belief is given by Finetti based on the concept of coherent bet [150], which essentially consists in assigning a certain bet for a particular event to occur depending on the belief of the observer. An argument against this definition is that the degree of belief depends on the knowledge of the observer and will in general change as the observer obtains more knowledge.

An non-exhaustive bibliography on these subjects can be [7] and [151]

In the context of this thesis, the most common outcome of random processes (events) is stated in a numerical form, like a measurement of a particular physics observable or a triggered signal (event counting). From Equation A.1, each observed event X can then be assigned a certain probability $\mathcal{P}(X)$. This step is straight forward when dealing with discrete variables, since each discrete variable can be assigned a certain probability. The extrapolation to continuous variables is done through the definition of the Probability Density Function (PDF) *P(x)*. For an infinitely small number dx, the probability that x is found in the interval $\left[x-\frac{1}{2}dx, x+\frac{1}{2}dx\right]$ is given by:

$$\mathcal{P}\left(x-\frac{1}{2}dx < x < x+\frac{1}{2}dx\right) = P(x)dx, \tag{A.2}$$

where *P(x)* is positive-defined for any value of x and satisfies the following normalization condition:

$$\int_{-\infty}^{\infty} P(y)dy = 1 \tag{A.3}$$

The extrapolation of a PDF to a multi-dimensional one is called a joint PDF.

Random processes can be described with models, which are built out of PDFs. PDFs describing a model are, sometimes, parametrized with a finite number of parameters. A set of unspecified parameters in the PDF provides a full family of PDF. The aim would be to identify which is the correct set of parameters corresponding to the model that describes better a set of random processes (measurements). This is referred to in the literature as parameter estimation, being an estimate of a parameter our "best guess" of what the parameter value is.

Though several parameter estimation methods exist, a widely used in particle physics is the Maximum Likelihood Estimator (MLE), based on the Maximum Likelihood (ML) theorem. Consider a set of N independent observations of the same random variable X, $\boldsymbol{X} : \{X_1, X_2, ..., X_N\}$, whose probability distribution for each of the random variables is described by $f\left(X \mid \boldsymbol{\theta}\right)$, where $\boldsymbol{\theta}$ is a set of analytical parameters. Since all the observations are uncorrelated, the joint PDF of all the observations $\boldsymbol{X}$ is:

$$P(\boldsymbol{X} \mid \boldsymbol{\theta}) = P(X_1, X_2, ..., X_N \mid \boldsymbol{\theta}) = \prod_{i=1}^{N} f(X_i \mid \boldsymbol{\theta}) \tag{A.4}$$

When after an experiment, the random variables $\boldsymbol{X}$ are replaced by the observed data $\boldsymbol{X}^0$ in the expression above, the probability of obtaining *exactly* $\boldsymbol{X}^0$ is computed[1]. Then, $P(\boldsymbol{X} \mid \boldsymbol{\theta})$ is no longer a PDF but an analytical function of $\boldsymbol{X}^0$ called the *likelihood function* $\mathcal{L}$:

$$\mathcal{L}(\boldsymbol{\theta}) = P(\boldsymbol{X}^0 \mid \boldsymbol{\theta}) \tag{A.5}$$

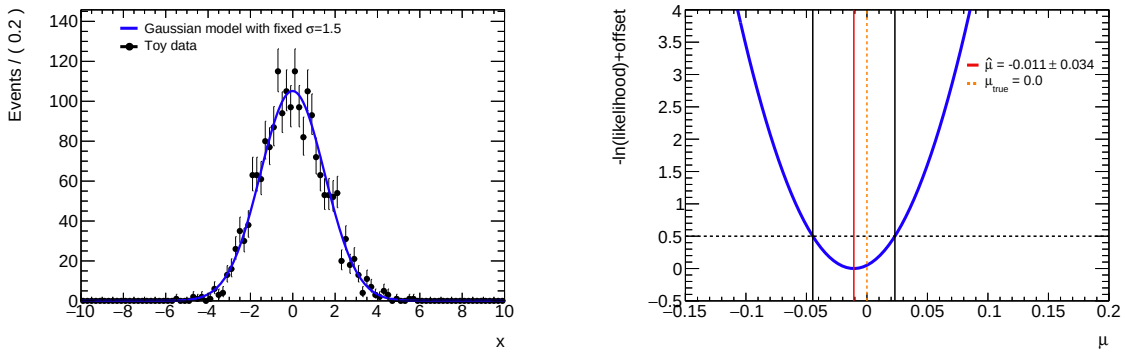
[1]This value is in general very small.

The Maximum Likehood theorem states that the set of $\hat{\boldsymbol{\theta}}$[2] that maximizes $\mathcal{L}$ are estimates of $\boldsymbol{\theta}$.

$$\mathcal{L}(\hat{\boldsymbol{\theta}}) = max\,|_{\boldsymbol{\theta}}\,(\mathcal{L}(\boldsymbol{\theta})) \tag{A.6}$$

The covariance of $\mathcal{L}(\boldsymbol{\theta})$ around its maximum can be used to extract the variance of $\sigma_{\hat{\boldsymbol{\theta}}}$ of each of the estimates.

In practice, the minimization of $-ln\mathcal{L}$ is preferred since it is easier to compute. A simple example is shown in Figure A.1 of parameter estimation of a gaussian PDF with fixed width given a set of randomly generated measurements. Further details are explained below the figure.



a) b)

Figure A.1: *(a) Toy data in black markers with gaussian PDF in blue. (b) Negative log-Likelihood function $-ln\mathcal{L}$ computed for the data and gaussian PDF shown in the plot on the left. The best estimate of the mean, $\hat{\mu}$, of the gaussian is found at the minimum of $-ln\mathcal{L}$ denoted with a red line. The uncertainty on the estimate, $\sigma_{\hat{\mu}}$, is given by the intersection of $-ln\mathcal{L}$ with $-ln\mathcal{L}(\hat{\mu})+0.5$ (horizontal dashed line) and marked in the plot with two black vertical lines.*

A.2 Parameters of interest and systematic uncertanties

Among the parameters used in a model necessary to describe a particular random process, two types can be identified:

- The parameters of interest (PoI) are the quantities to be estimated. These are usually denoted in the literature by $\boldsymbol{\mu} : \{\mu_1, \mu_2, ..., \mu_n\}$
- The nuissance parameters (NP) are other quantities needed to describe the random process but are not of interest in the measurement. The NP will be the source of systematic uncertainties in our estimates of PoI, since they are usually provided by the theory or auxiliary measurements. These are usually denoted in the literature by $\boldsymbol{\theta} : \{\theta_1, \theta_2, ..., \theta_k\}$

The fact that the NP are not of interest does not imply that they can be disregarded, since an incorrect estimation of them can lead to significant biases in the estimation of the

[2]The $\hat{}$ is used to make the difference between estimates and parameters.

PoI. A simple example of a problem with NP are counting experiments with signal and background species. The number of background events and also signal acceptance are both NP which, if badly estimated, would lead to biased results. Both can be estimated from *auxiliary measurements* $\boldsymbol{X}_{aux}$ of data sidebands, theory predictions or from simulations. When the analysis is simple enough, as it is this example, a simultaneous ML fit to data and auxiliary measurements would provide estimates of the PoI while including the effect of the NP in the uncertainty. In a more general example, the number of NP can be significantly larger which complicates the building of complete likelihood functions. A common practice then is to include these additional measurements by reparametrizing the likelihood and adding functional constraints:

$$\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\theta}) = \mathcal{L}_{\mu}(\boldsymbol{\mu}, \boldsymbol{\theta} \mid \boldsymbol{X}^0)\mathcal{L}_{\theta}(\boldsymbol{\theta} \mid \boldsymbol{X}_{aux}), \tag{A.7}$$

Each PoI affected by a given systematic θ, is reparametrized as $\mu' = \mu K(\theta)$, where $K(\theta)$ is a function characterizing a given systematic θ. The choice of this analytical function K_θ is arbitrary but usually a simple description is preferred, possibly one in which the relative magnitude of the uncertainty appears. The functional form of the constraint $\mathcal{L}_\theta(\boldsymbol{\theta})$ is usually chosen to be a unit Gaussian centered at 0 given θ.

Once the full likelihood function is built, a maximum likelihood fit provides estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\theta}}$. Even though both are equally treated statistically speaking, the NP should never be better constrained compared to the auxiliary measurement, since this could translate into an underestimation of the PoI uncertainties. Inspecting the pulls of the NP after the fit could hint where the issue comes from:

$$\text{Pull}(\theta_i) = \frac{\hat{\theta}_i - \theta_{i,obs}}{\sigma_{\hat{\theta}_i}} \tag{A.8}$$

This pull essentially quantifies how far the post-fit value is from the expected value obtained from auxiliary measurements. The central value is expected to be 0 with a pull of 1. Two types of issues can be identified by observing the post-fit pulls:

- A deviation from 0 would imply that the NP has absorbed a particular feature from data not accounted for by the model
- A smaller uncertainty than 1 would imply that the NP is over-constrained by the data, which may be possible but could also point towards a mismodelling issue.

A.3 On the problem of hypotheses testing

After describing the basic statistical tools to be used in this manuscript and methods to correctly estimate parameters with their uncertainties, the basics of hypothesis testing are described. Hypothesis testing in statistics provides an answer to the question: given a set of observed data $\boldsymbol{X}^0$, how robust is a particular hypothesis H_0 describing the data and how incompatible is another hypothesis H_1 with the data. The H_0 hypothesis under test is denominated the *null* hypothesis while H_1 is denominated the *alternative* hypothesis. As explained later, the denomination between *0-1* and *null-alternative* may change depending on the hypothesis under test.

The first necessary step is to condense the set of events $\boldsymbol{X}^0$ into a single numerical value. A test-statistics $T = T(\boldsymbol{X}^0)$ is a random variable defined as a function of the

observed data $\boldsymbol{X}^0$. Depending on the statistical analysis performed, it can be defined in different ways but always providing some discrimination between the null and alternative hypothesis. The discrimination is done by defining a subspace ω in the space of all possible values of T, W. This subspace is called the *critical* region and the complementary one, $W-\omega$, is denominated the *acceptance* region. If the T is contained in the acceptance region, then the null hypothesis is accepted, whereas if it falls outside it is rejected.

The critical region ω is usually defined in terms of the *size* α, defined as the probability of T falling in the critical region when H_0 is true:

$$\mathcal{P}(T \in \omega \mid H_0) = \alpha \tag{A.9}$$

In other words, α is the probability of discarding the null hypothesis while being true. Simultaneously, the *power* of the test-statistic, β can be defined as the probability of T to fall in the acceptance region when H_1 is true:

$$\mathcal{P}(T \in W-\omega \mid H_1) = \beta \tag{A.10}$$

Since both values α and β are somehow arbitrary and chosen by the analyser, it is practice to define one of them before performing the analysis in order to fix the size of the critical region. A physicist would then choose α (common practice in HEP is 0.05) and then look for the test-statistics with the largest β. This is a general problem with no solution unless only the two hypothesis that are being tested are fully determined, with no free parameters. This solution is given by the Neyman-Pearsons lemma [152] which states that the optimal test-statistics λ is the ratio of the likelihood functions of the two hypothesis:

$$\lambda = \frac{\mathcal{L}(\boldsymbol{X}^0 \mid H_0)}{\mathcal{L}(\boldsymbol{X}^0 \mid H_1)} \tag{A.11}$$

An example of how λ_{obs} varies for two input datasets with different amounts of signal is shown in Figure A.2. The shaded region represents the fraction of times the same or more extreme results would be obtained in infinite and equal experiments assuming true a certain hypothesis:

$$p\text{-value} = \int_{\lambda_{obs}}^{+\infty} P(\lambda \mid H_i) d\lambda \tag{A.12}$$

This probability is called *p-value* and is customary in particle physics to transform it into an equivalent significance Z, defined as the number of standard deviations with respect to the mean in a Gaussian distribution above which the integral of the distribution is equivalent to the p-value. As an example, a discovery is claimed with Z = 5 (convention) which is equivalent to $p = 2.87 \times 10^{-7}$.

For a parameter-dependent hypothesis, an intuitive generalization of the Neyman-Pearsons lemma is obtained by "profiling" out both likelihoods. This is called a *ratio of profile likelihoods* and it is obtained by maximizing both likelihoods with respect to all parameters except for the parameters of interest.

A.4 Confidence interval estimation

A similar approach as the one used to test hypothesis in the previous section can be used to estimate *confidence intervals*. A confidence interval $[\mu_a, \mu_b]$ contains the true value of the parameter μ_{true} with probability $1-\alpha$, also called *confidence level*. The

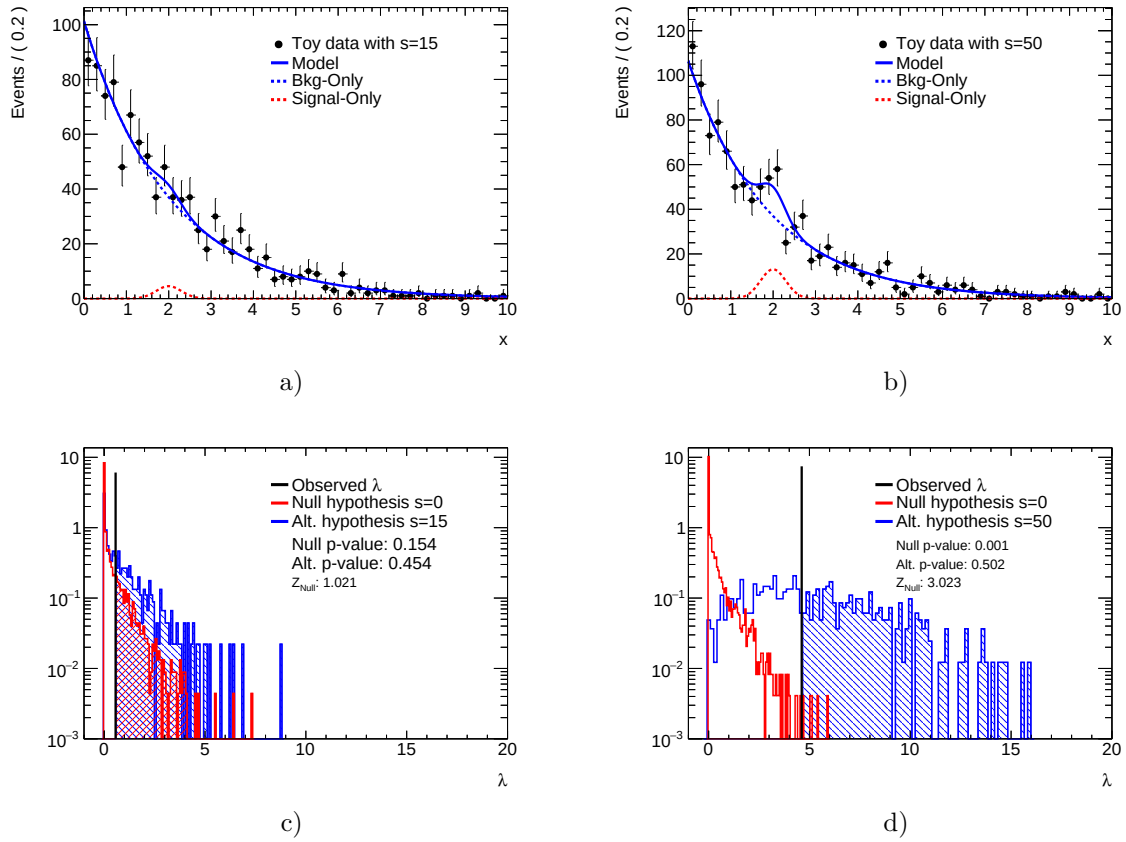


Figure A.2: *(a) and (b) Toy datasets with 15 and 50 signal events described by a falling exponential (background) and a gaussian peak (signal). (c) and (d) Distributions of the test-statistic λ for both input datasets shown in (a) and (b). Note how the observed λ_{obs} shifts towards larger values, reducing the p-value and hence increasing the significance of the deviations with respect to the null hypothesis (which in this is case is false). While in (a) the significance of the result is small and nothing can be claimed; in (b) larger significance is observed.*

general method to obtain confidence intervals is called the *Neyman construction* A.3 and essentially consists on "inverting" a probability in the data-space to the parameter-space.

$$1-\alpha = \mathcal{P}(x_1(\theta) < x < x_2(\theta)) \rightarrow \mathcal{P}(\theta_2(x) < \theta_0 < \theta_1(x)) \tag{A.13}$$

Note that the values x_1 and x_2 are not fixed and depend on the physicist, since many intervals in the data-space can contain a fraction $1-\alpha$ of the PDF. An illustration of the method is shown in Figure A.3 and the common frequentist approach used in particle physics is through the *profile likelihood ratio*[3]:

$$\lambda_\mu = \frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}})}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})} \tag{A.14}$$

with $\hat{\hat{\boldsymbol{\theta}}}$ referring to the best estimates of the NP for fixed values of μ and $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ the best estimates from a ML fit. The numerator can be understood as the null-hypothesis,

[3]In the following, only one-dimensional confidence intervals are considered. The multi-dimensional case is discussed in [151].

which depending on the confidence level required, may be accepted or not for a particular value of μ.

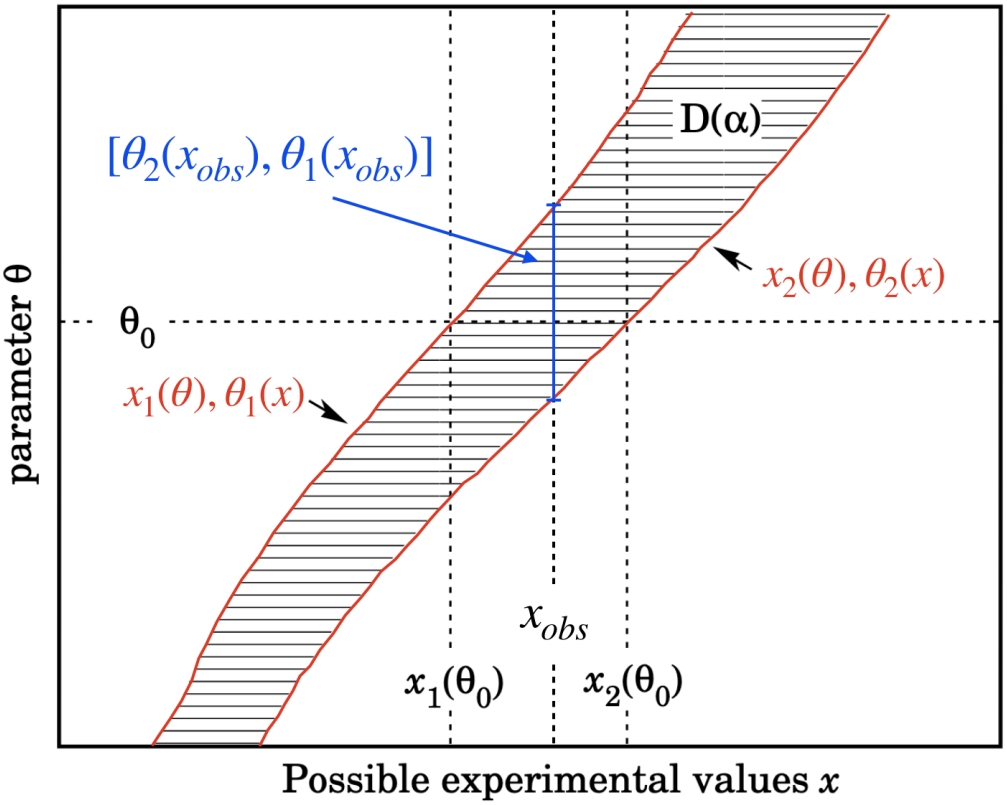


Figure A.3: *Illustration of the Neyman construction where θ is the parameter of interest, x can be any function of the measured data, $D(\alpha)$ is the so called confidence belt which contains all the possible(accepted) hypothesis of θ given a confidence level of 1-α. The confidence interval is then given by the interception of the measured data x_0 with the two red curves x_1 and x_2. Original figure from [7].*

The confidence interval is then the ensemble of all μ for which the null-hypothesis is accepted. This method provides exact *coverage* meaning that the confidence level matches with the fraction of times the true parameter will be included in the confidence interval in infinite and equal experiments.

Several confidence intervals can be defined depending on what the purpose of the analysis is. All of them arise from different choices of the critical region ω. In the following, common examples of confidence intervals used in HEP, and in particular in particle physics, are presented.

Estimates confidence intervals Measuring quantities is the most general case. Consider an estimate of a PoI, $\hat{\mu}$, obtained from a maximum likelihood fit like in Figure A.1b. The estimate could be off from the true value in two different ways: either being too high or too low. This reasoning justifies a symmetric critical region with respect to $\hat{\mu}$, since in this measurement no argument exists to favour positive values compared to negatives or vice-versa. Consider either a confidence level of $1-\alpha$ or the size of the test α. This ensures that if experiments were performed assuming μ true, a fraction of $\alpha/2$ experiments would be rejected because $\mu > \hat{\mu}$ and the same fraction for $\mu < \hat{\mu}$. The uncertainties in Figure A.1b correspond to a common confidence level 1 - α of 68%.

CL_{s+b} and CL_s intervals Consider an analysis in which a particle is searched for. The experimental setup for such analysis in which a model predicts a particular signal (an excess in the number of recorded events for example) could be not sensitive enough and a measurement on the number of events excess may not be possible. In those cases, a result to provide is an upper limit on the value of μ, which cannot be discarded with the observed data and the model given. The critical region would then be defined from the upper limit μ_{up} up to $+\infty$ and the acceptance region from the lowest value of μ up to μ_{up}. A good choice of test-statistics for computing upper limits is the profiled likelihood ratio A.14 with some modifications described below. In this case, the null hypothesis corresponds to the model with a tested value μ_{test} while the alternative hypothesis corresponds to the best fit to the data.

$$q_{\mu_{test}} = \frac{\mathcal{L}(\mu_{test}, \hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \quad \text{if} \quad \hat{\mu} < \mu_{test}. \tag{A.15}$$

Also, disagreement between the data and the model with $\hat{\mu} > \mu_{test}$ (values which the experiment is not sensitive to) should not be considered and so $q_{\mu_{test}}$ is set to 0 for these values. For each of the hypotheses $q_{\mu_{test}}$, if the computed p-value, p_μ, (Equation A.12) is below a certain size of critical region α, then that value is said to be excluded with a confidence level of $1-\alpha$. The upper limit is defined as the largest value of μ with a $p_\mu < \alpha$. The confidence level computed with this method is denoted CL_{s+b}, since in fact the computed interval is obtained over data which is a mixture of signal and background events.

It could happen, since both background and signal events are present in data, that both species suffer an under-fluctuation of events. This could lead to an overly aggressive 95% exclusion where the sensitivity is poor just due to the fact that none of the hypothesis (null and alternative) is suited for the observed data. In this context, an analogous definition of a confidence interval is defined to overcome issues caused by downward fluctuations of the background:

$$CL_s = \frac{CL_{s+b}}{CL_b} = \frac{p_\mu}{1-p_b}, \tag{A.16}$$

where p_b is the p-value obtained assuming the background only hypothesis to be true. Though this definition is not a probability (since it is a ratio of probabilities), it is protected from downward fluctuations of the background. It is in general stronger than the CL_{s+b} which implies that the computation of the CL_s over-covers the resulting confidence interval (the fraction of times an experiment would provide a limit inside the interval is 98% instead of 95%). An example of these two confidence intervals is shown in Figure A.4.

A.5 The missing piece: test-statistics distribution

In the last two sections, a small problem that has been kept unexplained, is illustrated by Equation A.9. This expression represents the probability of obtaining a test-statistic inside/outside the critical region, for which the distribution of the test-statistics is needed. The most general method to estimate the test-statistics distribution is to generate pseudo experiments. Though it is fairly simple to perform, since only a statistical model and data are required, it is often computationally heavy(increasing time with increasing complexity or number of parameters). Fortunately, under some assumptions and approximations, an analytical expression for the likelihood-based test-statistics distribution can be obtained.

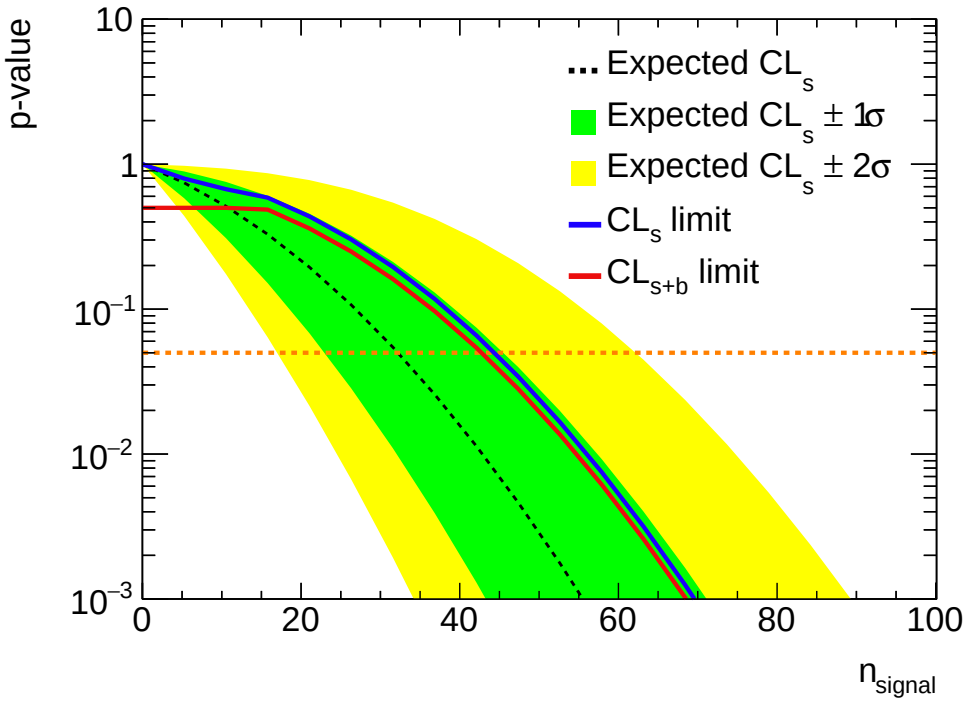


Figure A.4: *P-values as a function of different hypotheses of the parameter of interest for the data shown in Figure A.2a. The black dashed line shows the expected p-values and 1 and 2σ variations in green and yellow bands respectively. The upper limit $CL_{i,95\%}$ (with i being s or s+b) on the n_{signal}, with this dataset, is given by the intersection of the orange dashed line, which denotes $\alpha = 0.05$, with the two p-values computed (solid blue for CL_s and red for CL_{s+b} lines). As expected, CL_s is a slightly worse limit compared to CL_{s+b}.*

Both methodologies are described below since are of common use in particle physics measurements.

A.5.1 Pseudo-experiments

Pseudo-experiments are generated for each hypothesis (e.g. different values for μ) to construct a frequentist PDF. Once the distribution has enough statistics, p-values can be computed in order to reject/accept hypothesis. This methodology is used for Figure A.2. In that particular example, it is clear that the tails of the test-statistics distributions are not (yet) well described. The main drawback of this approach is that depending on the p-value of a particular hypothesis, very large pseudo-experiments may be needed increasing the generation time. This is the reason why the asymptotic approximation detailed in the next section is preferred if all the needed assumptions are valid.

A.5.2 Asymptotic formulae

The asymptotic formulae for likelihood-based tests [153] is derived using previous results from Wilks [154] and Wald [155]. These results show that given a large dataset

and a model with one PoI which follows a Gaussian distribution, the profile log-likelihood ratio test-statistics (Equation A.14) can be approximated to:

$$-2ln\lambda_\mu = \frac{(\mu-\hat{\mu})^2}{\sigma} + \mathcal{O}\left(1/\sqrt{N}\right) \tag{A.17}$$

where σ is the standard deviation of μ and $\hat{\mu}$ is the best estimate of μ. This shows that $t_\mu = -2ln\lambda_\mu$ follows a non-central χ^2 distribution of one degree of freedom:

$$f(t_\mu|\mu') = \frac{1}{2\sqrt{t_\mu}}\frac{1}{\sqrt{2\pi}}\left[exp\left(-\frac{1}{2}\left(\sqrt{t_\mu}+\frac{\mu-\mu'}{\sigma}\right)^2\right)+exp\left(-\frac{1}{2}\left(\sqrt{t_\mu}-\frac{\mu-\mu'}{\sigma}\right)^2\right)\right] \tag{A.18}$$

The missing piece in order to develop the formulae is precisely σ which can be estimated using an *Asimov* dataset [153].

A.5.2.1 The most representative dataset: the Asimov dataset

An Asimov dataset is such that, given a model, when one evaluates the estimates of the parameters $(\hat{\mu},\hat{\theta})$ through a ML fit, one would obtain *exactly* the true parameters of the model. Consider first a binned dataset. To construct this dataset, a fully determined model is needed describing the number of entries per bin in a histogram. This model is evaluated at a fixed parameters of interest, μ_{Asimov}, from which then the NPs, $\hat{\hat{\theta}}_{Asimov}$, are obtained by profiling them out on data. Then for a given μ_{Asimov}, the model is completely determined. The construction of the Asimov dataset is done by setting the number of entries per bin in a histogram to the number of expected number of events predicted by the model: $n_{bin,Asimov} = E[n_i](\mu_{Asimov},\hat{\hat{\theta}}_{Asimov})$. By construction, no statistical fluctuations exist in this binned dataset (Figure A.5)[4].

The interesting feature of this dataset is that, given a test-statistics λ_μ, the median of the $f(t_\mu \mid \mu_{Asimov})$ evaluated on the Asimov dataset would be equivalent to the median obtained generating pseudo-experiments. This implies that only one (Asimov) dataset is needed to evaluate the characteristics of a given test-statistics, which reduces significantly the number of pseudo-experiments needed to construct A.18. The Asimov dataset represents an ideal dataset (without statistical fluctuations) for a given model and it is of great use when computing expected sensitivities before performing the actual measurement in data.

[4]The extrapolation to an unbinned dataset is not exactly defined but a good approximation is obtained by setting the limit of the width of the bins to 0.

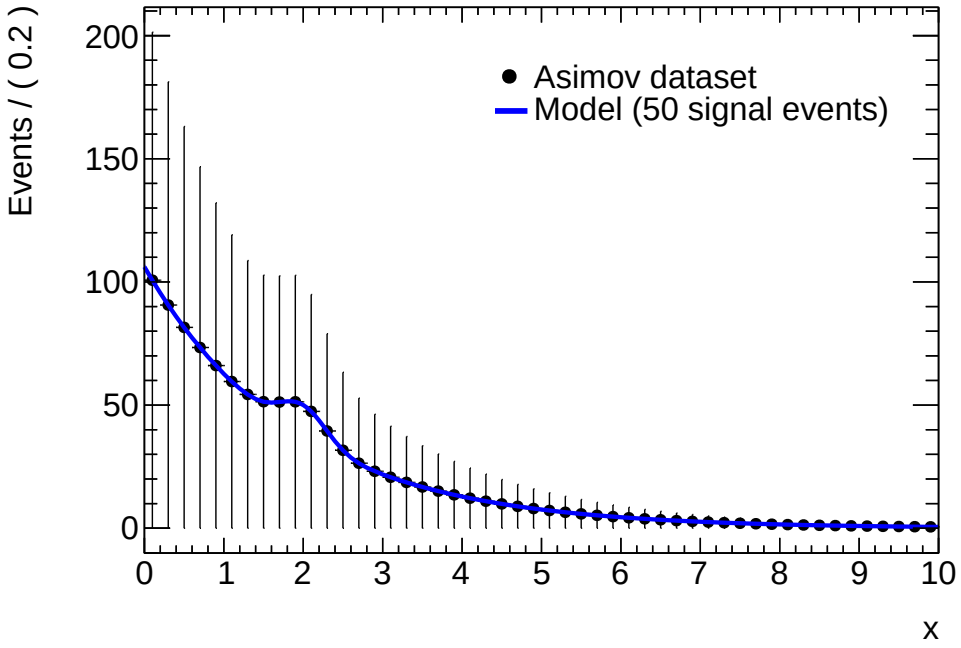


Figure A.5: *Asimov dataset built from the same model used in Figure A.2. Since no fluctuations exist by construction, the error bars shown in the histogram have no physical meaning (one entry per bin)*

Appendix B

Crystal Ball-like functions

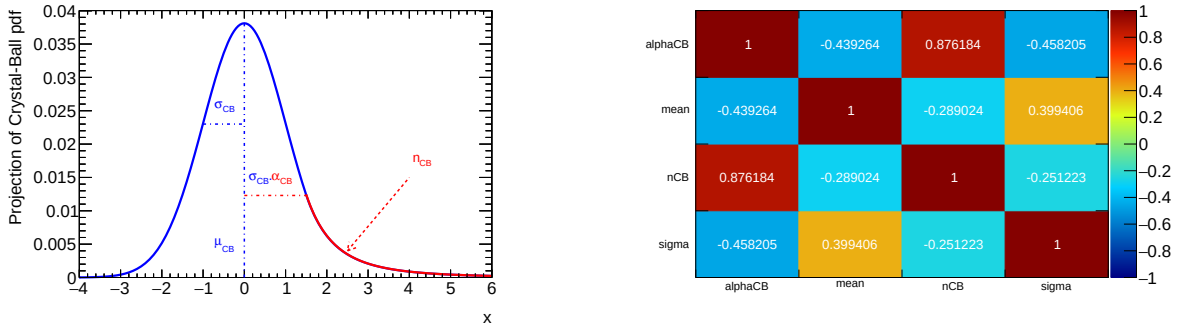
This appendix describes some of the properties of Crystal Ball functions, from which both Asymmetric Crystal Balls and double-sided Crystal Balls (Sections 4.2.2 and 6.4) inherit. This section intends to give a feeling of the correlations and behaviour of parameters to readers who have not used them before.

B.1 Crystal Ball function

A Crystal Ball is an analytical function composed by a Gaussian core and a single power law in one of the tails. It is defined as the following:

$$f_{CB}(\mu,\sigma,,\alpha,n) = N \cdot \begin{cases} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), & \text{if } x \geq \mu \text{ and } \frac{x-\mu}{\sigma} < \alpha; \\ \left(\frac{n}{n-\alpha^2+\alpha\sigma^{-1}(x-\mu)}\right)^n exp\left(-\frac{\alpha^2}{2}\right) & \text{if } \frac{x-\mu}{\sigma} \geq \alpha, \end{cases} \tag{B.1}$$

where x is the independent variable, μ and σ are the mean and width of the Gaussian core and α and n describe the power law tail. An example is shown in Figure B.1a, where the various parameters are also indicated.



a) b)

Figure B.1: *Crystal Ball distribution and correlation matrix of its parameters.*

The correlation matrix is shown in Figure B.1b, in which a large correlation can be observed between the parameters of the power law tail and the width σ of the distribution. To improve the stability of the fit, it is common practice to fix the parameter n, to values obtained from previous fit iterations. Moreover, large n values (approximately for $n > 15$)

barely affect the shape of the distribution and do not provide additional information on the model. This is illustrated in Figure B.2, where the Crystal Ball function is shown for different n values.

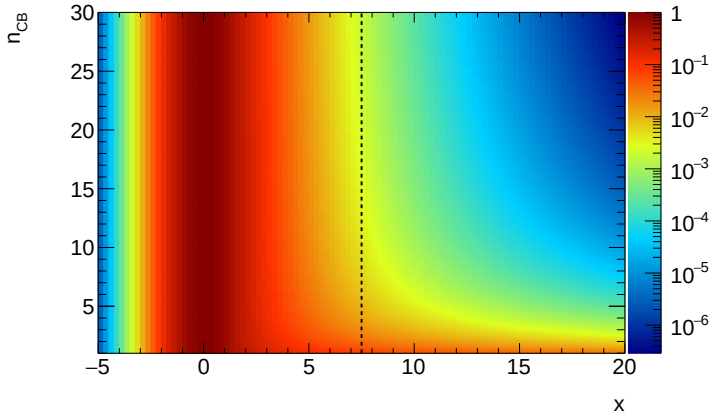


Figure B.2: *Evolution of the Crystal Ball function for different n_{CB} values. For each n_{CB} value, the Crystal Ball is shown in the horizontal axis normalized to the same value at the peak position (x=0) for every n_{CB}. The z-axis represents the probability as a function of x. The "width" of the distribution (understood for this plot as the size in x (by eye) with the largest probability and denoted by the dashed line) is relatively constant with increasing n_{CB} beyond $n_{CB} \approx 10$.*

B.2 Double-sided Crystal Ball function

It is defined as follows:

$$f_{DSCB}(\mu,\sigma,\alpha_{\text{low}},\alpha_{\text{high}},n_{\text{low}},n_{\text{high}}) = N \cdot \begin{cases} \frac{exp\left(-0.5\alpha_{\text{low}}^2\right)}{\left(\frac{n_{\text{low}}}{n_{\text{low}}-\alpha_{\text{low}}^2+\alpha_{\text{low}}t}\right)^{n_{\text{low}}}} & \text{if } t \geq -\alpha_{\text{low}}, \\ exp\left(-\frac{t^2}{2}\right), & \text{if } -\alpha_{\text{low}} < t < \alpha_{\text{high}}, \\ \frac{exp\left(-0.5\alpha_{\text{high}}^2\right)}{\left(\frac{n_{\text{high}}}{n-\alpha_{\text{high}}^2+\alpha_{\text{high}}t}\right)^{n_{\text{high}}}} & \text{if } t \geq \alpha_{\text{high}}; \end{cases} \quad \text{(B.2)}$$

where $t = (m_{\gamma\gamma} - \mu)/\sigma$ with μ being the position of the peak of the Gaussian distribution and σ represents the width of the Gaussian part of the function; N is a normalization parameter; α_{low} (α_{high}) is the position of the junction between the Gaussian and power law on the low (high) mass side in units of t; and n_{low} (n_{high}) is the exponent of this power law.

B.3 Asymmetric Crystal Ball function

It is defined as follows:

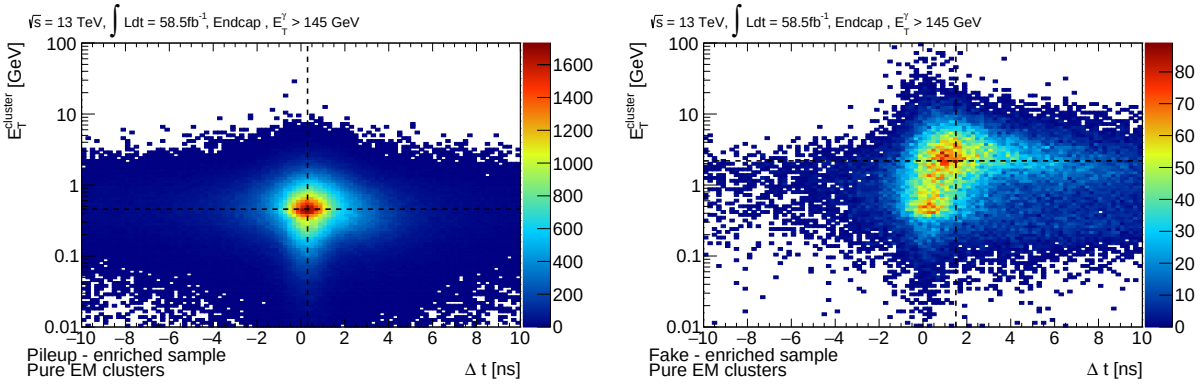
$$f_{ACB}(\mu,\sigma_L,\sigma_R,\alpha,n) = N \cdot \begin{cases} exp\left(-\frac{(x-\mu)^2}{2\sigma_L^2}\right), & \text{if } x < \mu; \\ exp\left(-\frac{(x-\mu)^2}{2\sigma_R^2}\right), & \text{if } x \geq \mu \text{ and } \frac{x-\mu}{\sigma_R} < \alpha; \\ \left(\frac{n}{n-\alpha^2+\alpha\sigma_R^{-1}(x-\mu)}\right)^n exp\left(-\frac{\alpha^2}{2}\right) & \text{if } \frac{x-\mu}{\sigma_R} \geq \alpha, \end{cases} \tag{B.3}$$

where x represents E_T^{iso}, μ is the mean of the bifurcated Gaussian, σ_L and σ_R are the Gaussian left and right widths and α and n describe the power law tail.

Appendix C

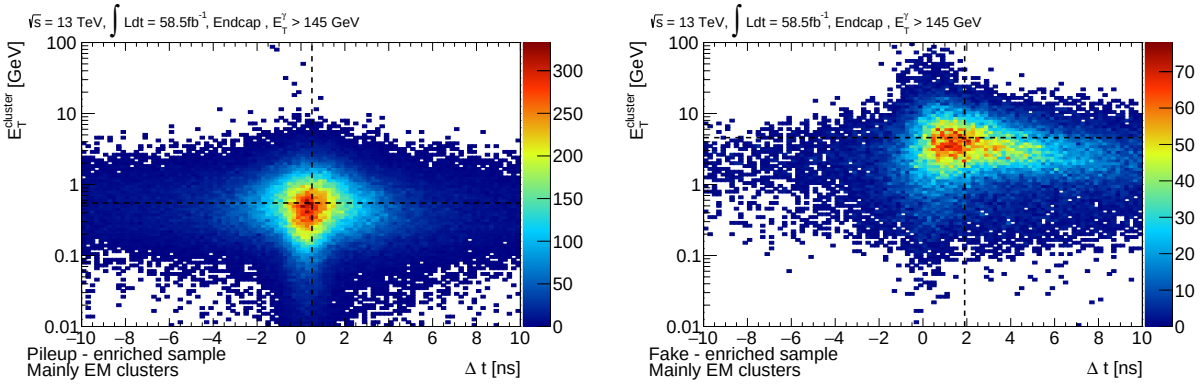
Additional topocluster plots

This appendix includes complementary plots to those shown in Chapter 5 for the endcap regions. Similar properties to those observed in the barrel are also present in the endcaps: two distinguishable regions in the $\left(\Delta t, E_T^\gamma\right)$ space that are associated to topoclusters from pileup or from fakes.



a) *Pileup enriched* b) *Fakes enriched (LP4)*

Figure C.1: *Bidimensional distributions of* $\left(\Delta t, E_T^\gamma\right)$ *for* ***pure EM*** *topoclusters around unconverted photon candidates for the two subsamples defined in Table 5.1: pileup-enriched (left) and fake-enriched around Loose'4 photon candidates (right) in the endcaps.*



a) *Pileup enriched* b) *Fakes enriched (LP4)*

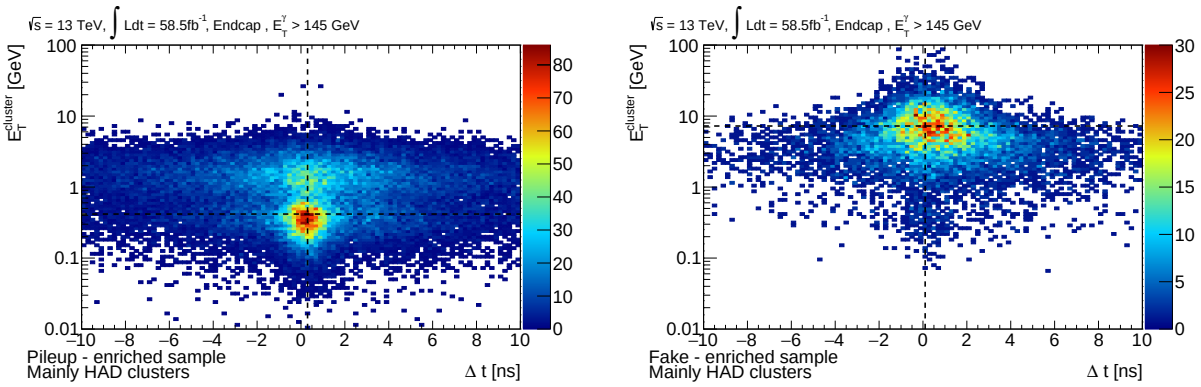
Figure C.2: *Bidimensional distributions of* $\left(\Delta t, E_T^\gamma\right)$ *for* ***mainly EM*** *topoclusters around unconverted photon candidates for the different the two enriched subsamples defined in Table 5.1: pileup-enriched (left) and fake-enriched around Loose'4 photon candidates (right) in the endcaps.*

Table C.1: *Measured parameters in each simultaneous fit to each topocluster category (**pure EM**, **mainly EM**, **mainly Had** and **pure Had**) and subsample (pileup-enriched and fake-enriched subsamples) in the endcaps. The mean and the two widths of each species are also shown. The second table shows the fraction of each species (soft-pileup, hard-pileup and fakes) in each of the three subsamples (pileup-enriched subsample and two fakes-enriched subsamples). Values in blue highlight the small contribution of topoclusters from fakes in the pileup-enriched subsample, while values in red highlight the dominant contribution of topoclusters from fakes in the fake-enriched subsamples.*

Category	Species/ Parameters	$\mu_{\Delta t}[ns]$	$\sigma_{L,\Delta t}[ns]$	$\sigma_{R,\Delta t}[ns]$	μ_{pt} [GeV]	$\sigma_{L,pt}$ [GeV]	$\sigma_{R,pt}$ [GeV]
Pure EM	Soft-pileup	0.244 ± 0.010	0.797 ± 0.009	0.975 ± 0.011	$0.545^{+0.004}_{-0.004}$	$2.800^{+0.017}_{-0.017}$	$0.291^{+0.001}_{-0.001}$
	Hard-pileup	1.604 ± 0.038	3.676 ± 0.034	2.358 ± 0.023	$0.521^{+0.004}_{-0.004}$	$2.474^{+0.014}_{-0.014}$	$0.264^{+0.001}_{-0.001}$
	Fakes	0.634 ± 0.050	1.152 ± 0.039	3.258 ± 0.050	$2.363^{+0.048}_{-0.047}$	$2.847^{+0.067}_{-0.065}$	$0.347^{+0.004}_{-0.004}$
Mainly EM	Soft-pileup	0.232 ± 0.022	0.832 ± 0.021	1.004 ± 0.025	$0.591^{+0.012}_{-0.012}$	$3.286^{+0.053}_{-0.052}$	$0.318^{+0.004}_{-0.004}$
	Hard-pileup	1.315 ± 0.074	3.737 ± 0.064	2.735 ± 0.052	$0.493^{+0.007}_{-0.007}$	$2.184^{+0.020}_{-0.020}$	$0.318^{+0.003}_{-0.003}$
	Fakes	1.364 ± 0.059	1.747 ± 0.047	3.149 ± 0.061	$2.914^{+0.055}_{-0.054}$	$1.836^{+0.030}_{-0.030}$	$0.378^{+0.004}_{-0.004}$
Mainly Tile	Soft-pileup	0.222 ± 0.044	0.668 ± 0.037	0.713 ± 0.038	$0.373^{+0.019}_{-0.018}$	$1.829^{+0.055}_{-0.053}$	$0.168^{+0.011}_{-0.011}$
	Hard-pileup	0.812 ± 0.121	4.232 ± 0.111	3.843 ± 0.120	$1.707^{+0.039}_{-0.038}$	$3.615^{+0.066}_{-0.065}$	$0.230^{+0.003}_{-0.003}$
	Fakes	0.440 ± 0.081	2.216 ± 0.060	2.259 ± 0.063	$5.417^{+0.165}_{-0.160}$	$1.863^{+0.050}_{-0.049}$	$0.330^{+0.006}_{-0.006}$
Pure Tile	Soft-pileup	-3.009 ± 0.006	19.847 ± 4.273	0.000 ± 0.011	$2.137^{+0.256}_{-0.229}$	$1.385^{+0.159}_{-0.143}$	$0.197^{+0.015}_{-0.014}$
	Hard-pileup	0.699 ± 0.110	4.051 ± 0.103	4.621 ± 0.172	$1.395^{+0.062}_{-0.060}$	$3.645^{+0.083}_{-0.081}$	$0.229^{+0.021}_{-0.020}$
	Fakes	-0.862 ± 0.120	1.582 ± 0.099	2.618 ± 0.099	$3.862^{+0.151}_{-0.145}$	$1.995^{+0.079}_{-0.076}$	$0.314^{+0.010}_{-0.010}$

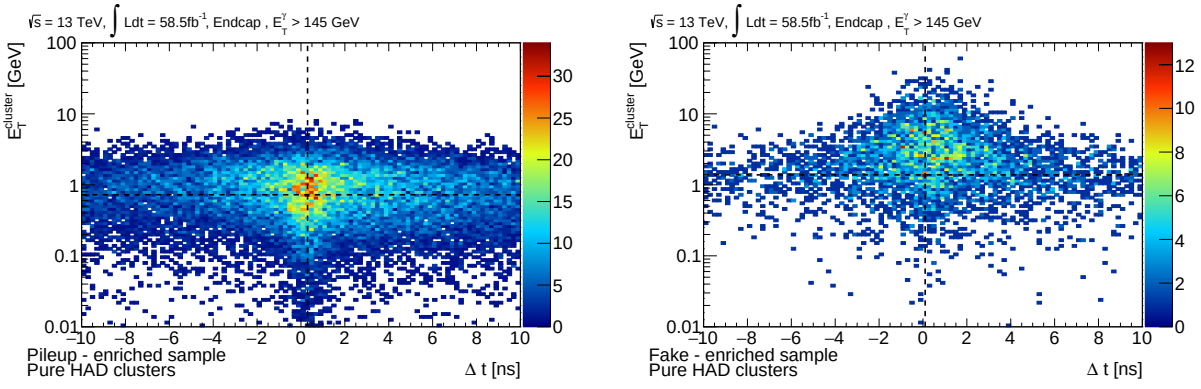
Category	Species	Pileup-enriched [%]	Fakes-enriched [%]	Fakes-enriched (non-isolated photons)[%]
Pure EM	Soft-pileup	50.17 ± 0.11	10.93 ± 0.38	11.80 ± 0.35
	Hard-pileup	49.78 ± 0.11	13.03 ± 0.41	12.16 ± 0.35
	Fakes	0.06 ± 0.01	76.05 ± 0.51	76.04 ± 0.46
Mainly EM	Soft-pileup	42.82 ± 0.21	2.80 ± 0.22	3.09 ± 0.21
	Hard-pileup	57.13 ± 0.21	10.95 ± 0.42	10.67 ± 0.37
	Fakes	0.05 ± 0.01	86.26 ± 0.46	86.24 ± 0.42
Mainly Tile	Soft-pileup	19.53 ± 0.32	0.83 ± 0.22	1.69 ± 0.27
	Hard-pileup	80.00 ± 0.32	10.46 ± 0.73	8.68 ± 0.59
	Fakes	0.46 ± 0.05	88.71 ± 0.75	89.64 ± 0.64
Pure Tile	Soft-pileup	0.47 ± 0.08	3.55 ± 0.57	4.79 ± 0.61
	Hard-pileup	99.51 ± 0.08	22.60 ± 1.29	22.88 ± 1.20
	Fakes	0.02 ± 0.02	73.85 ± 1.35	72.33 ± 1.28

182



a) *Pileup enriched* b) *Fakes enriched (LP4)*

Figure C.3: *Bidimensional distributions of* $\left(\Delta t, E_T^\gamma\right)$ *for* ***mainly Had*** *topoclusters around unconverted photon candidates for the two subsamples defined in Table 5.1: pileup-enriched (left) and fake-enriched around Loose'4 photon candidates (right) in the endcaps.*



a) *Pileup enriched* b) *Fakes enriched (LP4)*

Figure C.4: *Bidimensional distributions of* $\left(\Delta t, E_T^\gamma\right)$ *for* ***pure Had*** *topoclusters around unconverted photon candidates for the different the two enriched subsamples defined in Table 5.1: pileup-enriched (left) and fake-enriched around Loose'4 photon candidates (right) in the endcaps.*

Appendix D

Crosstalk in the Liquid Argon electromagnetic calorimeter

The energy measurement of particles with calorimeters is based on the transformation of the energy deposited in the active material into a measurable signal. An imprecise simulation or non-considered detector effects of data can bias the reconstruction and identification of an incident particle. Crosstalk between cells in the EM calorimeter can induce biases in the estimation of the energy and time of electromagnetic objects, as well as in their characterization through the shower shape variables.

This appendix discusses the observation of systematic low energy deposits located around photon candidates with energies larger than 145 GeV. In the first section, a brief introduction to crosstalk effects in ATLAS is given, as a possible explanation to the topoclusters observed around high energy photons. The latter are detailed in Section D.2. The last section summarizes the impact of the observed topoclusters in photon identification performances.

D.1 Sources of crosstalk in ATLAS

Crosstalk appears when the signal in a given cell or channel creates an undesired effect on another cell or channel. These effects can modify the pulse shape of a given cell (see Figure 3.1), and thereby its time and amplitude. The impact of crosstalk on the energy measured in a given cell varies with the source of crosstalk and cell granularity, ranging from a per-mille up to almost 8% in the strip layer of the LAr calorimeter. In the following, the input calibration signal in a given cell is referred to as the pulsed signal, while the perturbation arising from crosstalk in a different cell is referred to as the crosstalk signal.

This phenomenon can originate on the electrodes between two nearby cells or anywhere along the electronic readout chain. Depending on the source, three types of crosstalk can be identified: capacitive, inductive and resistive crosstalks. Each of them can be modeled by electrical circuits with resistive, capacitive and inductive couplings between cells. A few details relevant for the discussion on the different types of crosstalk are given[1], followed by a summary of their impact in the ATLAS LAr calorimeter.

[1] A more detailed review can be found in [156, 157]

D.1.1 Capacitive crosstalk

The capacitive crosstalk originates from charge sharing between the electrodes of nearby cells. The surrounding structure of an electrode is shown in Figure 2.12a with pink lines. The outer copper layers provide the necessary HV to drift the electrons produced by the LAr ionization, while the inner copper layer collects the induced signal. The three copper layers are separated by Kapton layers of 0.5 mm width, insulating each electrode from those in nearby cells. However, the insulation is not perfect and some charge transfer can occur by means of capacitive couplings, inducing signals in nearby electrodes.

The capacitive crosstalk is the dominant type of crosstalk in high granularity layers, since the electrodes are closer to each other. The crosstalk signal shape induced by capacitive coupling corresponds to the derivative of the pulsed signal shape [156], hence it modifies its time and amplitude (see Figure D.1).

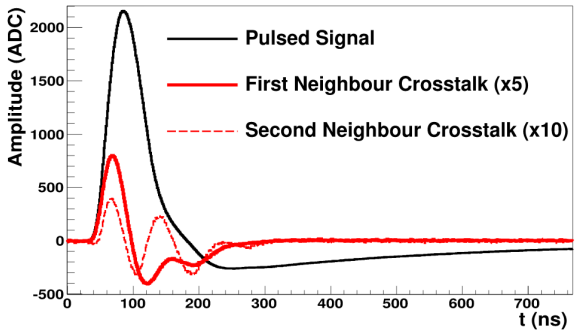


Figure D.1: *Pulsed (black) and capacitive crosstalk signals (red) in the first layer of the LAr calorimeter. The crosstalk signal for the first (solid) and second (dashed) neighbour cells are shown [157].*

D.1.2 Resistive crosstalk

The resistive crosstalk originates from the resistors placed on the boundaries between the first and second layers. The HV is supplied to the first layer by a series of resistors that physically connects cells with the second layer. Despite the fact that the resistors have a very large resistance (hundreds of $k\Omega$), 8 of them are placed in parallel between a single cell from the second layer and each of the adjacent cells in the first layer, inducing a non negligible crosstalk signal. The shape of the resistive crosstalk pulse is expected to be the same as the pulsed signal, since no dephasing exists between them (see Figure D.2).

D.1.3 Inductive crosstalk

The mutual inductance between elements of the readout chain can induce some coupling between different cells and is the source of the so-called inductive crosstalk. The fact that it can happen at any point in the readout chain implies the loss of spatial correlation between the cells in the calorimeter, leading to long distance crosstalk. Three main sources can lead to inductive crosstalk: from the ground return, from the mother boards or from the feedthroughs. Each of these sources induces a different crosstalk pulse shape, different to the pulsed signal; the actual shapes depend on which layers are involved (see Figure D.3).

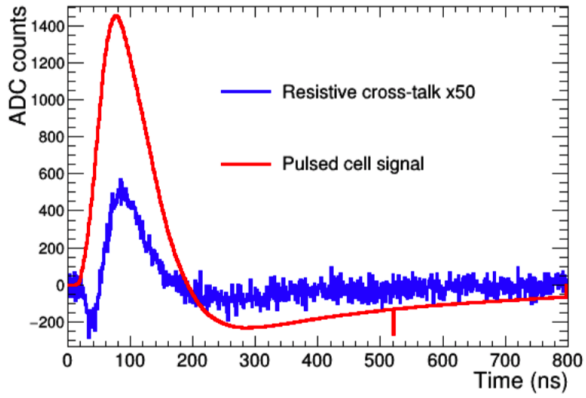
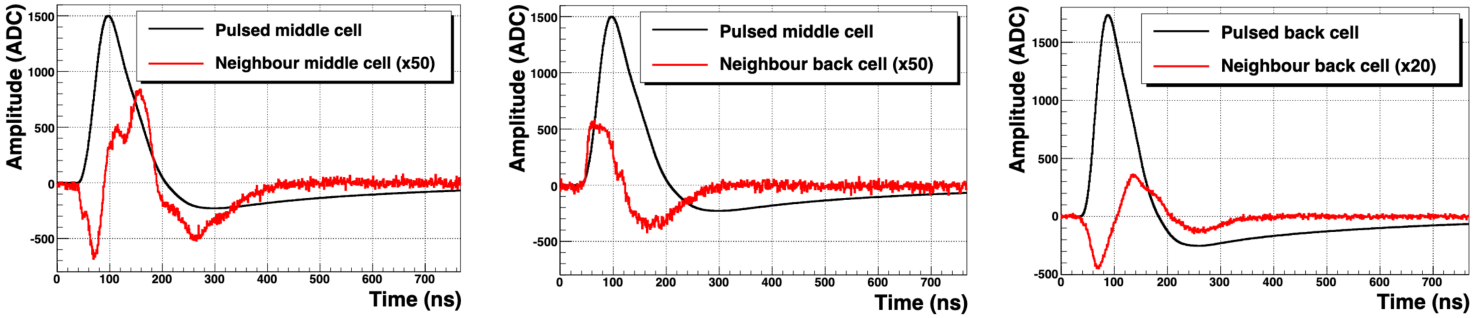


Figure D.2: *Pulsed (red) and resistive crosstalk signals (blue) between the first and second layers of the LAr calorimeter [158].*



a) *Middle to middle layer crosstalk* b) *Middle to back layer crosstalk* c) *Back to back layer crosstalk*

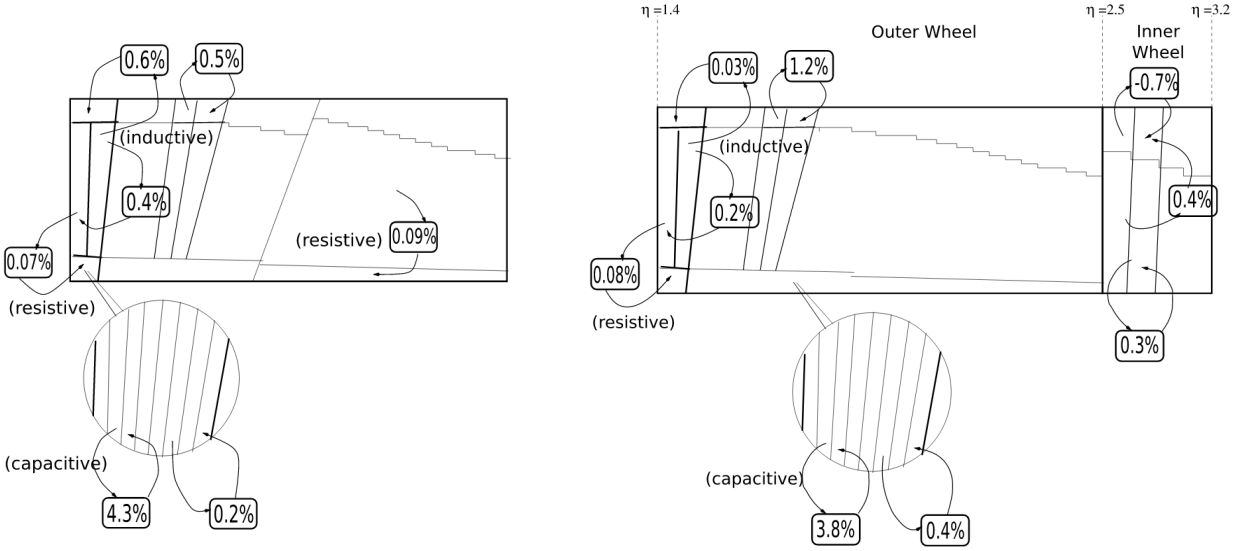
Figure D.3: *Pulsed signal (black) and inductive crosstalk signal (red) obtained from calibration data. (a) The signal is pulsed in a cell of the second layer and the crosstalk is measured in the nearby cell in the second layer. (b) The signal is pulsed in a cell of the second layer and the crosstalk is measured in the nearby cell in η in the third layer. (c) The signal is pulsed in a cell of the third layer and the crosstalk is measured in the nearby cell in the third layer [157].*

D.1.4 Crosstalk in ATLAS

All the aforementioned crosstalk sources were first studied before the construction and operation of the LAr calorimeter. Capacitive crosstalk is the dominant source of crosstalk in the first layer of the calorimeter, due to its high granularity. Resistive crosstalk is the dominant source of crosstalk between the first and second layers, and inductive crosstalk is the dominant source of crosstalk in the second and third layers. The study of crosstalk in the LAr calorimeter led to important modifications in the design of the motherboards [156] to reduce the impact of inductive crosstalk.

Similar studies were performed after the complete installation of the LAr calorimeter [157] to provide a more robust estimation of the impact from crosstalk in energy measurements. An estimate of the crosstalk is provided by the ratio of amplitudes $X(t)$ and $V(t)$ at a given time of the crosstalk and pulsed signals respectively. For the sake of simplicity, the time at which the pulsed signal has a maximum, $t_{\max}$, is used in the following, also referred to as under-peak to peak estimation. A summary of the measured crosstalk in the ATLAS LAr calorimeter for the barrel and endcaps is shown in Figure D.4,

ranging from per-cent level from capacitive crosstalk in the first layer to per-mille level for the rest of the couplings. These estimates are included in the ATLAS simulation.



a) *Barrel map* b) *Endcap map*

Figure D.4: *Crosstalk map in the LAr EM calorimeter in ATLAS for the (a) barrel and (b) endcaps [157].*

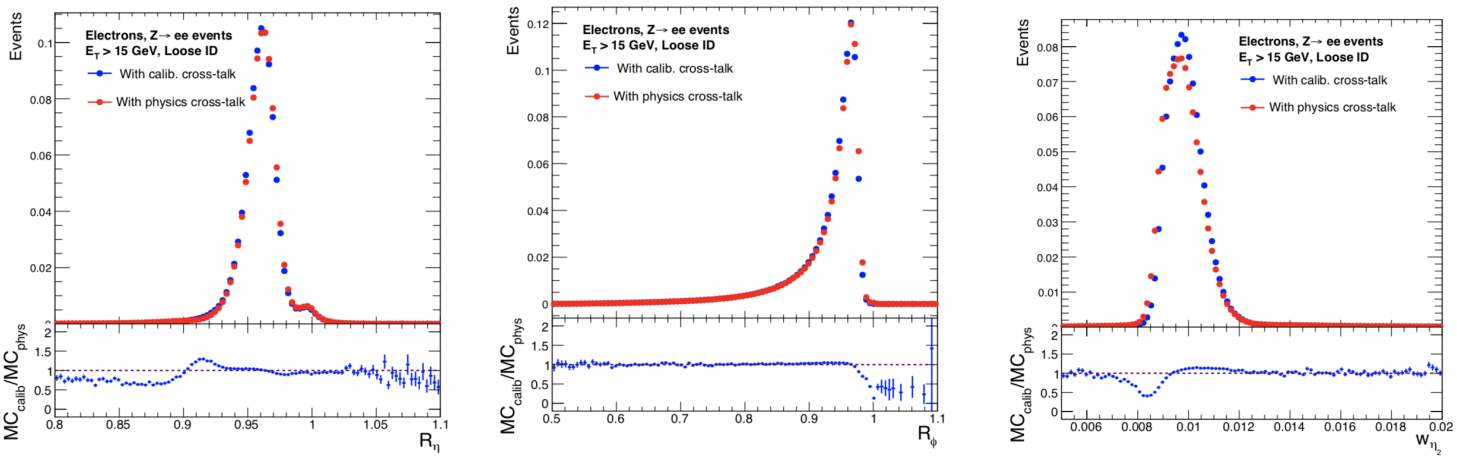
While the effects of crosstalk seem small, except from those observed in the first layer, they may play a non-negligible role in the characterization of EM showers through shower shape variables (defined in Section 3.3). An illustration of its impact is shown in Figure D.5, where the differences in various shower shapes of electrons from $Z \to ee$ in simulation arise from a different estimation of the crosstalk present in the LAr calorimeter[2].

D.2 Topoclusters from crosstalk around high energy photons

A pulsed signal in one of the calorimeter cells can induce crosstalk signals in the surrounding cells. The energy reconstructed in those cells is generally small, of the order of the per-mille level with respect to the energy of the particle. However, it can be sufficient to induce differences in the shower shape variables (defined in Section 3.3) and hence a precise modelling is required. Despite the fact that the crosstalk detailed in the previous section is already implemented in the simulation of the ATLAS detector, some effects may not be included, and could lead to data-simulation discrepancies.

In this section, the characteristics of energy deposits observed in data, but not in the simulation, around high energy photons are detailed. The same γ-jet samples used for data and simulation detailed in Chapters 4 and 5 are used in this study. Unconverted photons with $E_T^\gamma > 145$ GeV passing tight identification and `FixedCutTight` isolation requirements are selected in both data and simulation. This selection ensures purities above 95% in data (see Figure 4.13). To reduce the impact of pileup, events are required to have $\langle\mu\rangle < 30$. Regarding topoclusters, only those with substantial energy deposits in the

[2]Further details on both crosstalk estimation techniques are found in [158]



a) b) c)

Figure D.5: *Shower shape variables for EM showers initiated by electrons from $Z \to ee$ decays in simulations with two different crosstalk estimations. The blue markers denote the shower shape variables obtained in the simulation with the ATLAS nominal crosstalk implementation. The red markers denote the shower shape variables obtained in the simulation with the crosstalk implementation presented in [158].*

LAr calorimeter are chosen, with $f_{EM} > 0.9$ (see Sections 3.1.2.2 and 5.2 for more details on the definition of the EM fraction and how to use it to define topocluster samples).

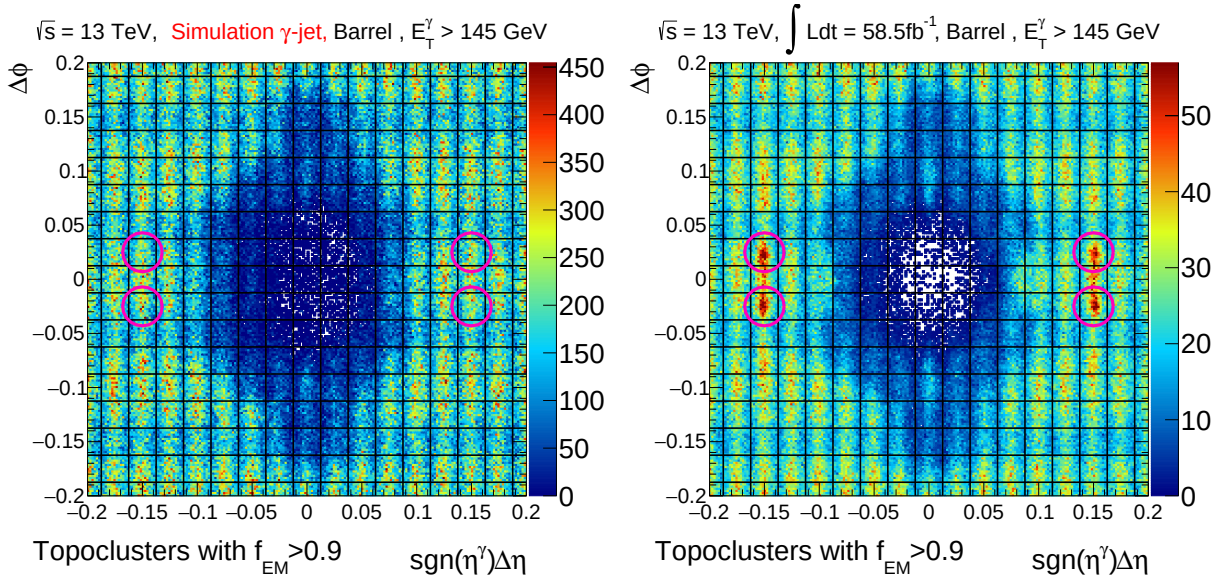
While the impact of crosstalk on the energy of a given cell is generally small, a systematic increase of the energy in a set of cells around a photon candidate could rise the probability of reconstructing a topocluster in such positions. This signature is observed in data, where topoclusters are systematically reconstructed in precise cells of the LAr calorimeter around the photon candidate. A comparison between data and simulation is shown in Figures D.6 and D.7 for the barrel and endcaps. These features are not present in the simulation in any of the cases.

Since the distance of these energy deposits with respect to the photon candidate is larger than several Molière radius and the fact that they are localized in single cells, a more precise study of their properties is presented in the following, to support the hypothesis that they originate from crosstalk.

D.2.1 Crosstalk in the barrel

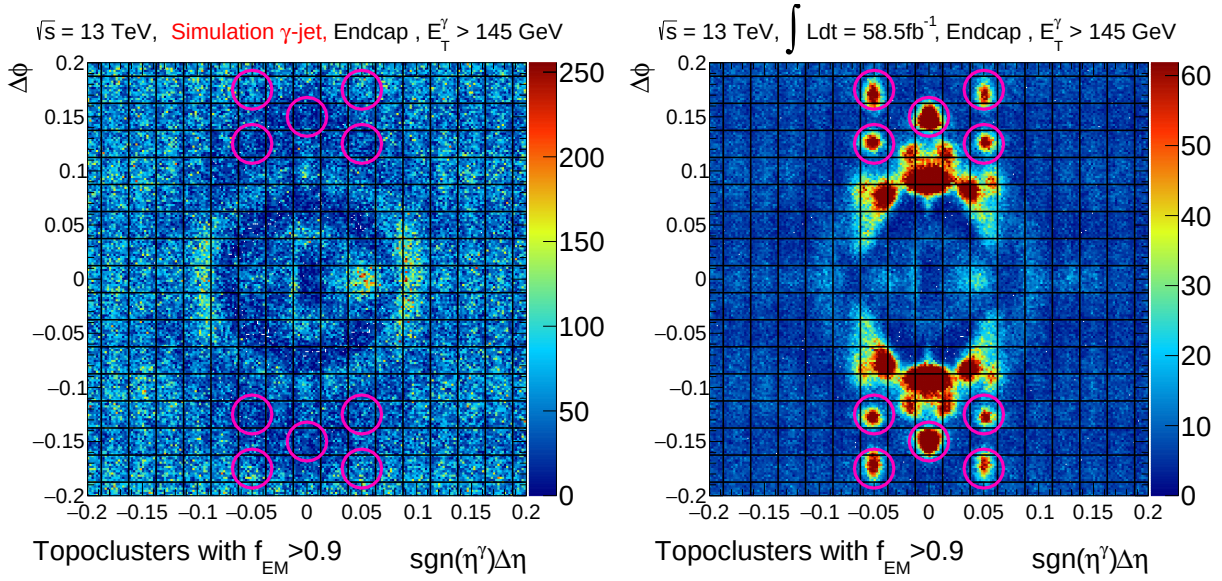
Figure D.6 shows the position of the reconstructed topoclusters around tight and isolated photons in the barrel. Four cells at a distance of $(\pm 6, \pm 1)$ cells in $(|\Delta\eta|, |\Delta\phi|)$ with respect to the photon are systematically found to have energies larger than the noise threshold, and hence these supposedly crosstalk-induced signals enter the seed list used for topocluster building. These signals are in contrast with the homogeneous topocluster density expected from pileup energy deposits.

The topoclusters reconstructed inside the circles in magenta in Figure D.6 arrive later to the calorimeter with respect to the photon candidate, between 5 and 20 ns, with approximately 0.1% of the photon energy. Moreover, the reconstruction rate of



a) *Simulation* b) *Data*

Figure D.6: *Topoclusters around tight isolated photon candidates in the barrel with $E_T^\gamma > 145$ GeV in the $(sgn(\eta^\gamma)\Delta\eta, \Delta\phi)$ plane. Simulation is shown in (a) and data in (b). The topoclusters inside the regions in magenta are selected for the presented analysis.*

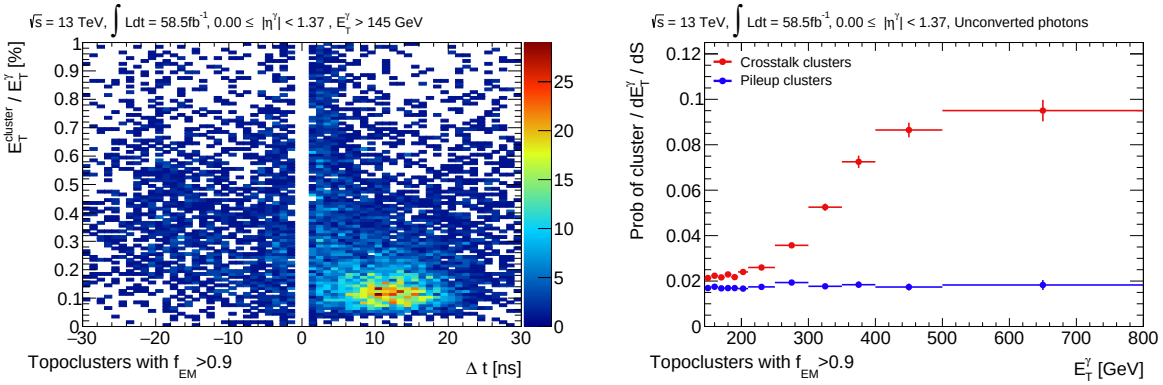


a) *Simulation* b) *Data*

Figure D.7: *Topoclusters around tight isolated photon candidates in the endcaps with $E_T^\gamma > 145$ GeV in the $(sgn(\eta^\gamma)\Delta\eta, \Delta\phi)$ plane. Simulation is shown in (a) and data in (b). The topoclusters inside the regions in magenta are selected for the presented analysis. The z-axis in the plot of data is set to 1/5 of the maximum value to visualize the features shown in the simulation.*

these topoclusters increases with the energy of the photon candidate (see Figure D.8b), compared to the reconstructing rate of pileup topoclusters, which remains stable[3].

[3]For a fair comparison with the rate of crosstalk topoclusters, the reconstruction rate of pileup topoclusters is computed in a surface S equivalent to the surface of the cells of the regions in magenta.



a) b)

Figure D.8: *(a) The bidimensional distributions of* $(\Delta t, E_T^{cluster}/E_T^{\gamma})$ *for topoclusters in data around tight isolated photon candidates with* $E_T^{\gamma} > 145$ *GeV in the barrel. Only topoclusters inside the circles in magenta of Figure D.6b are selected. (b) Reconstruction rate of crosstalk-induced topocluster as a function the transverse energy of the photon candidate in the barrel. The red markers denote the reconstruction rate of topoclusters in any of the regions in magenta of Figure D.6b while blue markers denote the reconstruction rate of topoclusters in a different region around the photon candidate.*

D.2.2 Crosstalk in the endcap

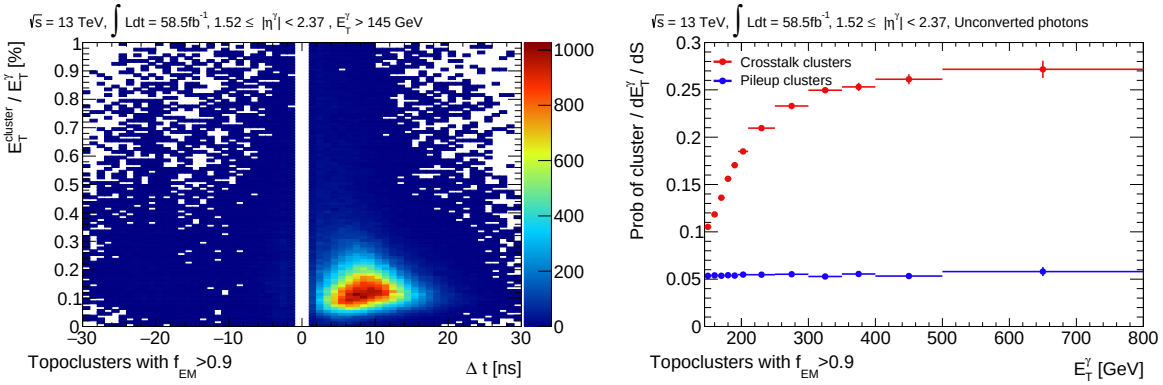
Figure D.7 shows the position of reconstructed topoclusters around tight and isolated photons in the endcaps. Regions with larger topocluster densities are located mainly around the $\Delta\phi$ direction, sometimes merging with other features around the photon candidate. In the following, only those topoclusters found inside the circles in magenta are considered, since these clearly belong to single cells in the calorimeter, and hence probably arising from crosstalk.

Similarly to those observed in the barrel, these topoclusters are found to arrive with a delay in time with respect to the photon candidate between 5 and 15 ns. The large majority of them are reconstructed with 0.1% of the energy of the photon candidate. Also, the reconstruction rate of these topoclusters increases with the energy of the photon candidate, being in this case much larger than in the barrel, reaching rates of 25% beyond $E_T^{\gamma} > 40$ GeV (see Figure D.9b).

D.2.3 Summary

The following facts extracted from the information on the topoclusters can be summarized as follows:

- The probability of reconstructing crosstalk topoclusters increases with the energy of the photon, as it is shown in Figures D.8b and D.9b.
- These topoclusters are localized in the second layer of the calorimeter.
- Topoclusters in these regions show a characteristic delay in time of $\Delta t > 7.5$ ns with respect to the photon candidate, reconstructed with approximately 0.1% of its energy (see Figures D.8a and D.9a).



a) b)

Figure D.9: *(a) The bidimensional distributions of* $(\Delta t, E_T^{cluster}/E_T^{\gamma})$ *for topoclusters in data around tight isolated photon candidates with* $E_T^{\gamma} > 145$ *GeV in the endcap. Only topoclusters in the regions in magenta of Figure D.7b are selected. (b) Reconstruction rate of topoclusters as a function the transverse energy of the photon candidate in the endcaps. The red markers denote the reconstruction rate of topoclusters in any of the regions in magenta of Figure D.7b while blue markers denote the reconstruction rate of topoclusters in a different region around the photon candidate.*

In particular, the last two points suggest the that inductive crosstalk in the second layer is at play, since the signal, supposedly from crosstalk, is delayed in time with respect to the pulsed signal (photon).

As explained in subsection D.1.4, crosstalk is included in the simulation for the first[4] neighbours around the pulsed cell. However, comparing Figures D.6 and D.7 for data and simulation, long distance crosstalk is not be present in the simulation in contrast to the large contribution of signals observed in data. The observed discrepancies, when located in the close to the photon candidate may induce non-negligible differences in the shower shape variables between data and simulation. This subject is further explored in the following section.

D.3 Impact on photon identification

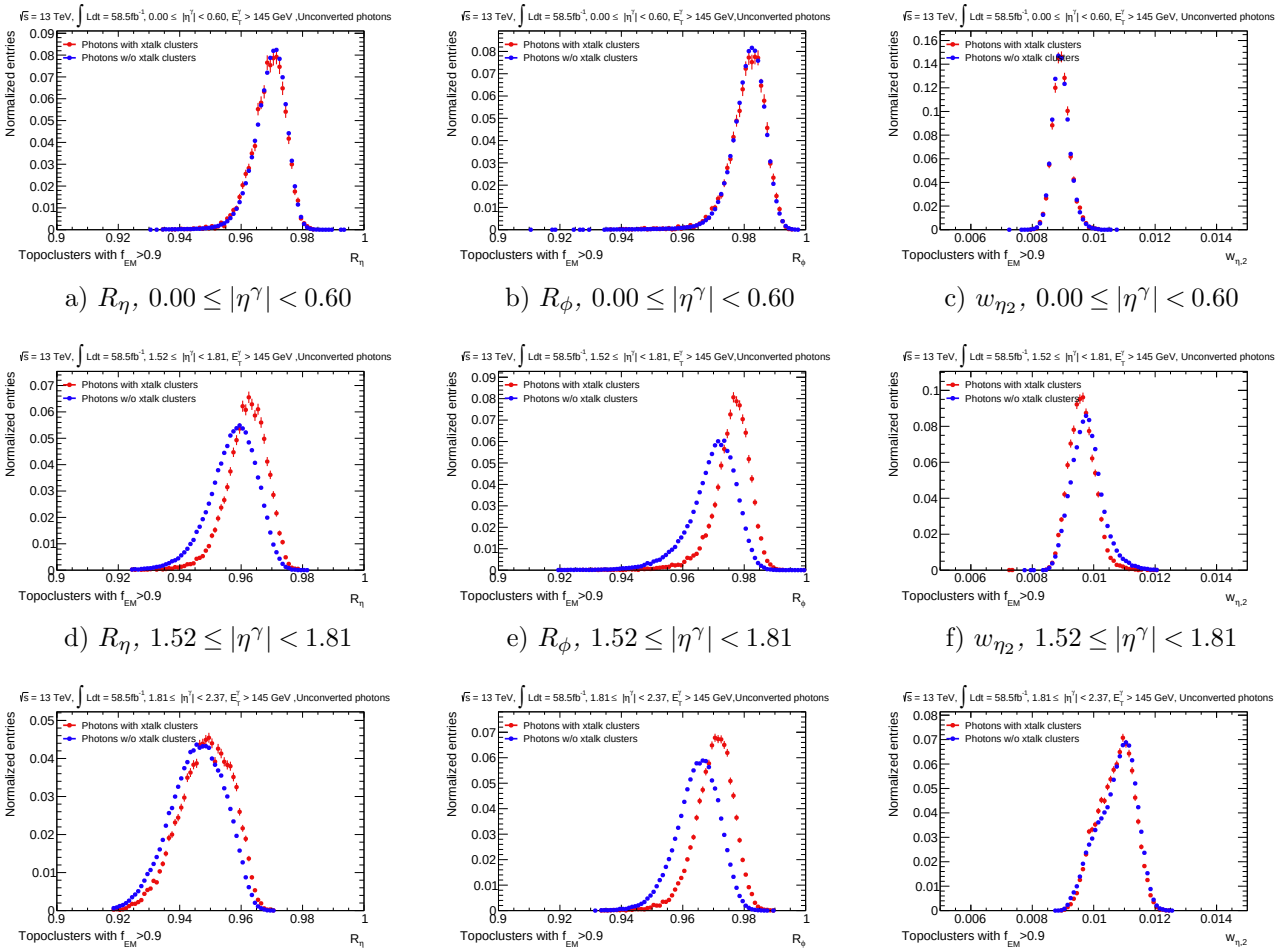
The crosstalk-induced topoclusters close to photon candidates can modify the shower shapes, mainly those related to the width of the photon shower. The induced differences can affect photon identification efficiencies depending on the presence or absence of these topoclusters, and since this effect is not predicted by the simulation, create additional data-simulation mismodellings. In this section, the impact of topoclusters from crosstalk on photon identification in data is investigated.

D.3.1 Shower shape variables and crosstalk clusters

Photons with crosstalk topoclusters are those with reconstructed topoclusters in one of the regions in magenta from Figures D.6 and D.7. Photons without crosstalk topoclusters are those without reconstructed topoclusters in the aforementioned magenta regions.

[4] And sometimes second, but with negligible impact.

The shower shape variables considered are R_ϕ, R_η and w_{η_2} (defined in Table 3.2), since they are built with information from the second layer of the LAr calorimeter. No large differences are observed in the barrel shower shape variables. Despite the distance of crosstalk induced signals of the regions in magenta with respect to the photon candidates, differences are observed in the shower shapes for photons in the endcaps. Photons with topoclusters from crosstalk in the endcaps show a narrower EM shower, illustrated by larger values of R_ϕ, R_η and a smaller value of w_{η_2} in Figure D.10.



a) R_η, $0.00 \leq |\eta^\gamma| < 0.60$ b) R_ϕ, $0.00 \leq |\eta^\gamma| < 0.60$ c) w_{η_2}, $0.00 \leq |\eta^\gamma| < 0.60$

d) R_η, $1.52 \leq |\eta^\gamma| < 1.81$ e) R_ϕ, $1.52 \leq |\eta^\gamma| < 1.81$ f) w_{η_2}, $1.52 \leq |\eta^\gamma| < 1.81$

g) R_η, $1.81 \leq |\eta^\gamma| < 2.37$ h) R_ϕ, $1.81 \leq |\eta^\gamma| < 2.37$ i) w_{η_2}, $1.81 \leq |\eta^\gamma| < 2.37$

Figure D.10: *Shower shape variables measured in data for unconverted photons with $E_T^\gamma > 145$GeV in three different $|\eta^\gamma|$ bins: $[0.00, 0.60)$, $[1.52, 1.81)$ and $[1.81, 2.37)$. The red markers denote the shower shapes measured on photons with crosstalk topoclusters around while blue markers denote those without crosstalk topoclusters around.*

The hypothesis that can explain this behaviour is described in the following. Energy deposits arising from crosstalk are always present in the LAr calorimeter with the characteristics described above. The reconstruction rate of topoclusters in those positions depends highly on the energy of the photon candidate, being higher at large E_T^γ. However, the reconstructed topoclusters can be absorbed by the topoclusters of the photon, leading to wider showers[5]. So, the fact that no topoclusters from crosstalk are found in those

[5]In terms of the shower shape variables.

positions it does not imply the-non existence of crosstalk, simply that it has been absorbed by the photon candidate topocluster.

The plots presented in Figure D.10 however, exploit only topoclusters present beyond ± 6 cells with respect to the photon candidate to reduce the impact of photon related features. The study of the mismodelling observed in the closest region around photon candidates could provide a more clear picture of the impact of crosstalk induced signals on the shower shape variables.

D.4 Conclusion

The study of topoclusters reconstructed around photon candidates presented in this appendix has shown that crosstalk mismodellings in the simulation can induce differences in the characterization of the EM shower with respect to data. Topoclusters are systematically reconstructed in data in regions that are sometimes beyond several Molière radius, up to 7 cells, with respect to the photon candidate. These signals are not present in the simulation and are attributed to inductive crosstalk in the second layer of the calorimeter, due to the observed time delay of several nanoseconds and their small fraction of energy relative to the photon candidate. The full impact of the observed crosstalk on the shower shapes in data presents a challenge given the proximity to the energy deposit of the photon candidate and its study is left for a future work.

Appendix E

Future of diphoton resonance searches with the ATLAS experiment

The analysis strategy presented in the Chapter 6 was driven by the limitation posed by the E_T^γ-thresholds required at trigger level during Run 2. Future data-taking periods are expected to record data with larger instantaneous luminosities, increasing the rates of the existing triggers. This would in principle imply a further increase in the trigger energy thresholds, thus drastically limiting the sensitivity at lower masses. In order to nevertheless protect the sensitivity at low masses, new topological criteria have been implemented at trigger level in order to reduce the rates in future data-taking periods without necessarily increasing the energy thresholds. In this appendix, a prospective study on future analyses probing masses below 65 GeV with the new triggers that are expected to be implemented for Run 3 is shown.

E.1 Diphoton triggers with topological selections

The largest contribution to the rates in diphoton triggers arises from the L1 triggers. As a reminder from Section 2.2.5, a trigger chain in ATLAS is defined as the combination of an L1 item and an HLT item. The structure of both L1 and HLT items can be sketched by the following expressions:

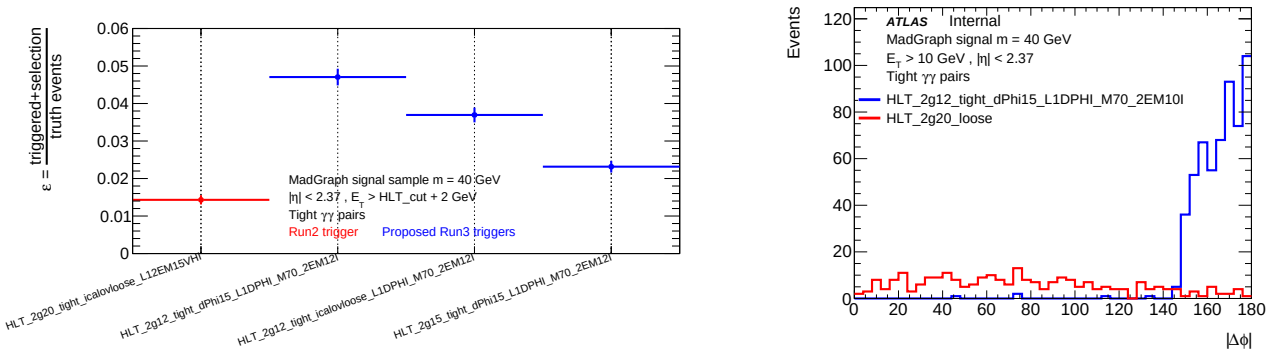
```
L1_[number of seeds][type seed][energy threshold]_[conditions],
HLT_[number of objects][object type][energy threshold]_[conditions].
```

With the aim of reducing the rate of L1 triggers and continue to probe very-low diphoton masses, two new topological items have been implemented:

- `DPHI`: this item requires two EM seeds to have $2.7 < |\Delta\phi| < \pi$.
- `M70`: this item requires the invariant mass of the two EM seeds to be lower than 70 GeV.

Equivalent definitions are implemented on the HLT. The reduction on the rates allows also for a small decrease in the E_T^γ-thresholds, down to 12 GeV at the L1 trigger. The `DPHI` selection on these new two topological items can potentially allow to probe low diphoton masses in not boosted topologies, an orthogonal phase space to the one studied in the analysis described in previous sections (see Figure E.1b).

The performance of several trigger proposals targetting very-low masses is evaluated on a simulated sample of a resonance decaying to diphotons of mass m_X=40 GeV, analogous to those detailed in Section 6.2.2, but without any filtering applied at particle-level. The trigger efficiencies (Figure E.1a) are larger, up to a factor 2.5, than the ones from the presented analysis. The chosen trigger for Run 3 is `HLT_2g15_tight_dPhi15_L1DPHI_M70_2EM12I`, which at HLT requires both photons to have more than E_T^γ>15 GeV and to have $|\Delta\phi| > 180 - 15$○, besides tight identification.



a) b)

Figure E.1: *(a) Efficiencies measured on different candidate triggers for Run 3 on a signal simulated sample of a narrow resonance decaying to diphotons. (b) Distribution of the angular distance in ϕ, $|\Delta\phi|$, between the two photons trigger by two different triggers. The blue distribution denotes the events recorded with a diphoton trigger requiring the diphoton pair to be closer than $|\Delta\phi| > 2.9$ and E_T^γ>12 GeV. The red distribution denotes the events recorded with a diphoton trigger requiring E_T^γ>20 GeV.*

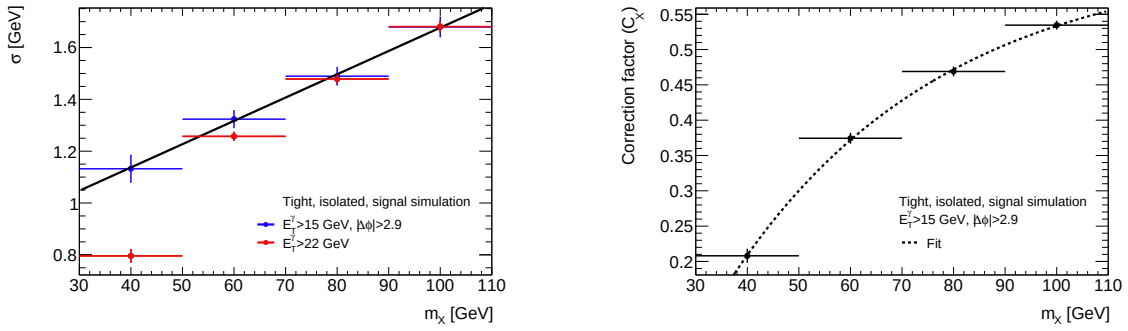
E.2 Prospective results

As described in Section 6.10, the expected limit on the observed cross-section of a narrow resonance over the diphoton background is a function of several ingredients: signal parametrization and efficiencies as a function of m_X, background shape and systematic uncertainties. In order to provide an estimate of the sensitivity of future analyses using the new diphoton triggers, isolation and identification performances are assumed to be exactly as in Run 2. This last assumption is probably not completely realistic, since pileup, among other factors, will worsen the performance of both steps in reconstruction. Also, systematic uncertainties are not considered.

The signal samples used in this study are unfiltered versions of samples with masses of 40,60,80 and 100 GeV, similar to the ones used in the presented analysis and detailed in Section 6.2.2. The diphoton background samples used in this study are filtered by requiring photons to have $E_{T,1}^\gamma$>20 and $E_{T,2}^\gamma$>18 GeV. This modifies the shape of the background around the turn-on with respect to what we expect for triggers requiring photons of E_T^γ>15 GeV, but its impact in the core of the distribution is expected to be minor.

First, the shape of the signal and its efficiencies are measured. The width of the signal shape is described with a Gaussian distribution and is parametrized as a function of m_X with a linear function. The mass resolution is worse compared to the inclusive search

since the selected photons have lower E_T^γ (see Figure E.2a). The performance of the photon reconstruction also degrades with decreasing E_T^γ and its effect is illustrated in the signal correction factors in Figure E.2b.



a) b)

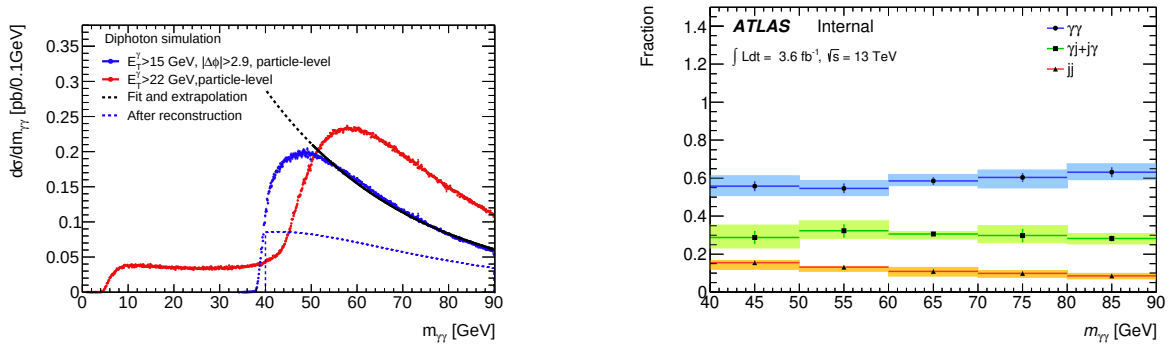
Figure E.2: *(a) The signal widths compared for two kinematic selections as a function of $m_{\gamma\gamma}$. Blue markers denote the width of the signal distribution measured for diphotons required to have $|\Delta\phi| > 2.9$ rad. Red markers denote the width of the signal distribution measured for diphotons required to have E_T^γ>22 GeV.*

Regarding the background, it is considered to be composed by irreducible and reducible backgrounds, neglecting the contribution of $Z \to ee$ events in which one or both electrons fake a photon. In the previous diphoton resonance search performed by ATLAS [127], the limit was provided down to 65 GeV, due to the difficulties posed to describe the trigger turn-on between 40 and 60 GeV. The threshold of the trigger decreasing from 20 GeV down to 15 GeV will push the turn-on towards lower masses. Moreover, some change is expected since the boosted contribution is not triggered due to the $|\Delta\phi| > 2.9$ requirement.

The shape of the irreducible background after requiring photons to have $|\Delta\phi > 2.9|$ and E_T^γ> 17 GeV is shown in Figure E.3a at particle-level, compared to the distribution obtained in the inclusive search with E_T^γ> 22 GeV. The $\Delta\phi$ cut rejects completely the boosted contribution, creating a cut-off in the lowest invariant mass. The distribution is modelled with a second order polynomial and extrapolated below the turn-on created by the filters applied at 18 and 20 GeV on the subleading and leading photons respectively.

Diphoton reconstruction efficiencies are extracted from the signal samples and assumed to be valid for the irreducible background events. The particle-level distribution is reweighted to account for the efficiency loss and obtain an estimate of the irreducible background shape after reconstruction. In order to estimate the contribution from the reducible component, the 2x2D method is used (Section 6.5.3). The shape of the irreducible and reducible components is observed to be similar with the aforementioned selection and the $\gamma\gamma$-purity is measured to be $f_{\gamma\gamma} = 60\%$, approximately constant with increasing mass. Thus, the total background distribution is obtained by reweighing the irreducible component, assuming the shape is identical and the purity constant with mass (see Figure E.3b).

A simplified estimation of the expected upper limit on the fiducial cross-section at 95% CL, σ_{up}, is estimated from the number of events equivalent to a 1.65σ fluctuation in the



a) b)

Figure E.3: *(a) Diphoton invariant mass distribution for the irreducible background under different kinematic selections at particle and reconstruction level. The red distribution corresponds to diphotons, at particle-level, selected by requiring E_T^γ>22 GeV. The blue distribution corresponds to diphotons, at particle-level, selected by requiring $|\Delta\phi| > 2.9$. The black fit and its extrapolation towards lower masses denote the distribution which would be obtained without any filtering at particle-level in the simulated diphoton sample. The dashed blue lines denote the blue invariant mass distribution after reconstruction. (b) Background decomposition computed with 3.6 fb^{-1} for diphotons selected requiring $|\Delta\phi| > 2.9$.*

number of background events:

$$\sigma_{up} = \frac{N_{s,up}}{\mathcal{L}_{int} C_X} = \frac{1.65 \times \sqrt{N_{bkg}}}{\mathcal{L}_{int} C_X}, \tag{E.1}$$

where $\mathcal{L}_{int}$ is the integrated luminosity, C_X is the correction factor and $N_{s,up}$ is the upper limit on the number of signal events, defined as $N_{s,up} = 1.65 \times \sqrt{N_{bkg}}$. Assuming an integrated luminosity expected by the end of Run 3 of $L \sim 180$ fb^{-1}, the expected limits are shown in Figure E.4, in which the result from the low-mass and very-low mass analyses are also shown for comparison.

E.3 Conclusions

This appendix presents a prospective study on future diphoton analyses covering masses below 100 GeVin a back-to-back topology. Future data-taking periods will benefit from improved triggering techniques with topological cuts, selecting photons which are predominantly back-to-back in the detector. The expected sensitivity is worse compared to the other analysis strategies, either the inclusive or the boosted case. At low masses, the boosted topology benefits from extraordinary narrow signals (factor ×2) while at higher masses the inclusive analysis has a factor ×2 in acceptance. However, more detailed studies, left for a future work, can be performed in order to benefit from the lower energy thresholds at trigger level, since topological cuts or allowing larger trigger rates seem the only possible ways to record low energy photons and thus explore lower masses.

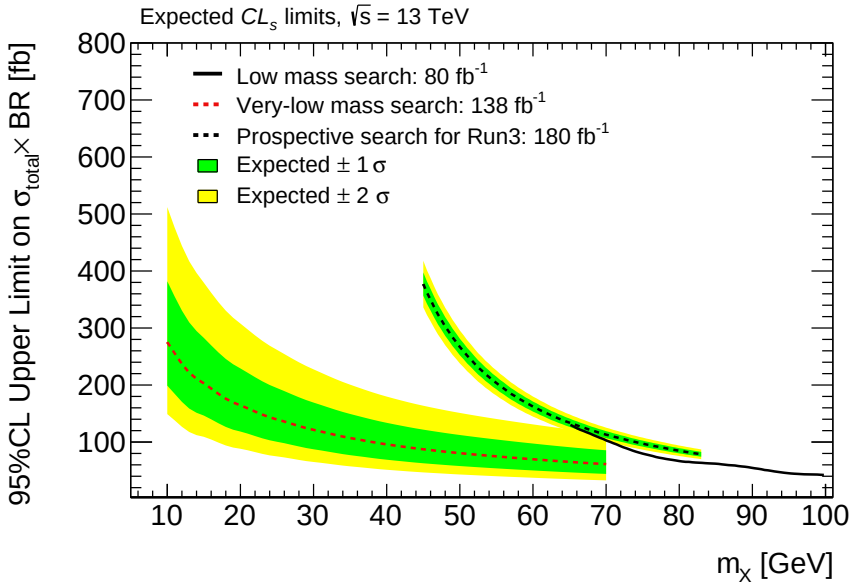


Figure E.4: *Compilation of exclusion limits at 95% CL on the total cross-section σ_{total} times its diphoton branching ratio $\mathcal{B}(X \to \gamma\gamma)$. The solid line denotes the public result by ATLAS without any systematic uncertainty obtained with an integrated luminosity of 80 fb^{-1}. This analysis was performed by selecting inclusive diphoton pairs with E_T^{γ}>22 GeV. The dashed red line denotes the expected limit obtained in the analysis presented in Chapter 6, requiring a boosted diphoton pair and using 138 fb^{-1}. The dashed black line denotes the expected limit computed with the prospective study detailed in this section. It assumes 180 fb^{-1} to be collected during Run 3. The green and yellow band denote ±1 and ±2 standard deviation bands.*

Bibliography

[1] Sheldon L. Glashow. "Partial-symmetries of weak interactions". In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: https://doi.org/10.1016/0029-5582(61)90469-2. URL: http://www.sciencedirect.com/science/article/pii/0029558261904692 (cit. on pp. 15, 16).

[2] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264. URL: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264 (cit. on pp. 15, 16).

[3] Abdus Salam. "Weak and Electromagnetic Interactions". In: *Conf. Proc. C* 680519 (1968), pp. 367–377. DOI: 10.1142/9789812795915_0034 (cit. on pp. 15, 16).

[4] M. Gell-Mann. "A schematic model of baryons and mesons". In: *Physics Letters* 8.3 (1964), pp. 214–215. ISSN: 0031-9163. DOI: https://doi.org/10.1016/S0031-9163(64)92001-3. URL: http://www.sciencedirect.com/science/article/pii/S0031916364920013 (cit. on pp. 15, 16).

[5] K.G. Wilson and W. Zimmermann. "Operator product expansions and composite field operators in the general framework of quantum field theory". In: *Commun. Math. Phys.* 24 (1972), pp. 87–106. DOI: 10.1007/BF01878448 (cit. on p. 17).

[6] E.C.G. Stueckelberg and A. Petermann. "The normalization group in quantum theory". In: *Helv. Phys. Acta* 24 (1951), pp. 317–319 (cit. on p. 17).

[7] Particle Data Group. "Review of Particle Physics". In: *Progress of Theoretical and Experimental Physics* 2020.8 (Aug. 2020). 083C01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptaa104. URL: https://doi.org/10.1093/ptep/ptaa104 (cit. on pp. 18, 43, 44, 166, 171).

[8] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321 (cit. on pp. 17, 18).

[9] P.W. Higgs. "Broken symmetries, massless particles and gauge fields". In: *Physics Letters* 12.2 (1964), pp. 132–133. ISSN: 0031-9163. DOI: https://doi.org/10.1016/0031-9163(64)91136-9. URL: http://www.sciencedirect.com/science/article/pii/0031916364911369 (cit. on pp. 17, 18).

[10] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles". In: *Phys. Rev. Lett.* 13 (20 Nov. 1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.585 (cit. on pp. 17, 18).

[11] L.D. Landau. "On the theory of phase transitions". In: *Zh. Eksp. Teor. Fiz.* 7 (1937), pp. 19–32 (cit. on p. 18).

[12] Jeffrey Goldstone, Abdus Salam, and Steven Weinberg. "Broken Symmetries". In: *Phys. Rev.* 127 (3 Aug. 1962), pp. 965–970. DOI: 10.1103/PhysRev.127.965. URL: https://link.aps.org/doi/10.1103/PhysRev.127.965 (cit. on p. 18).

[13] F.J. Hasert et al. "Observation of neutrino-like interactions without muon or electron in the gargamelle neutrino experiment". In: *Physics Letters B* 46.1 (1973), pp. 138–140. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(73)90499-1. URL: http://www.sciencedirect.com/science/article/pii/0370269373904991 (cit. on p. 18).

[14] G. Arnison et al. "Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s}$=540 GeV". In: *Physics Letters B* 122.1 (1983), pp. 103–116. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(83)91177-2. URL: http://www.sciencedirect.com/science/article/pii/0370269383911772 (cit. on p. 18).

[15] M. Banner et al. "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider". In: *Physics Letters B* 122.5 (1983), pp. 476–485. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(83)91605-2. URL: http://www.sciencedirect.com/science/article/pii/0370269383916052 (cit. on p. 18).

[16] P. Bagnaia et al. "Evidence for Z0e+e- at the CERN pp collider". In: *Physics Letters B* 129.1 (1983), pp. 130–140. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(83)90744-X. URL: http://www.sciencedirect.com/science/article/pii/037026938390744X (cit. on p. 18).

[17] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex] (cit. on p. 18).

[18] Serguei Chatrchyan et al. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex] (cit. on p. 18).

[19] J. D. Bjorken and E. A. Paschos. "Inelastic Electron-Proton and γ-Proton Scattering and the Structure of the Nucleon". In: *Phys. Rev.* 185 (5 Sept. 1969), pp. 1975–1982. DOI: 10.1103/PhysRev.185.1975. URL: https://link.aps.org/doi/10.1103/PhysRev.185.1975 (cit. on p. 19).

[20] R.P. Feynman. "The behavior of hadron collisions at extreme energies". In: *Conf. Proc. C* 690905 (1969), pp. 237–258 (cit. on p. 19).

[21] H. David Politzer. "Reliable Perturbative Results for Strong Interactions?" In: *Phys. Rev. Lett.* 30 (26 June 1973), pp. 1346–1349. DOI: 10.1103/PhysRevLett.30.1346. URL: https://link.aps.org/doi/10.1103/PhysRevLett.30.1346 (cit. on p. 19).

[22] John C. Collins, Davison E. Soper, and George F. Sterman. "Factorization of Hard Processes in QCD". In: vol. 5. 1989, pp. 1–91. DOI: 10.1142/9789814503266_0001. arXiv: hep-ph/0409313 (cit. on p. 19).

[23] J M Campbell, J W Huston, and W J Stirling. "Hard interactions of quarks and gluons: a primer for LHC physics". In: *Reports on Progress in Physics* 70.1 (Dec. 2006), pp. 89–193. DOI: 10.1088/0034-4885/70/1/r02. URL: https://doi.org/10.1088%2F0034-4885%2F70%2F1%2Fr02 (cit. on p. 19).

[24] Jon Butterworth et al. "PDF4LHC recommendations for LHC Run II". In: *Journal of Physics G: Nuclear and Particle Physics* 43.2 (Jan. 2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001. URL: https://doi.org/10.1088%2F0954-3899%2F43%2F2%2F023001 (cit. on p. 19).

[25] Yuri L. Dokshitzer. "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics." In: *Sov. Phys. JETP* 46 (1977), pp. 641–653 (cit. on p. 19).

[26] V.N. Gribov and L.N. Lipatov. "Deep inelastic e p scattering in perturbation theory". In: *Sov. J. Nucl. Phys.* 15 (1972), pp. 438–450 (cit. on p. 19).

[27] G. Altarelli and G. Parisi. "Asymptotic freedom in parton language". In: *Nuclear Physics B* 126.2 (1977), pp. 298–318. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(77)90384-4. URL: http://www.sciencedirect.com/science/article/pii/0550321377903844 (cit. on p. 19).

[28] Richard D. Ball et al. "Parton distributions for the LHC Run II". In: *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph] (cit. on pp. 19, 20, 30, 124).

[29] S. Catani et al. "Cross-section of isolated prompt photons in hadron hadron collisions". In: *JHEP* 05 (2002), p. 028. DOI: 10.1088/1126-6708/2002/05/028. arXiv: hep-ph/0204023 (cit. on p. 21).

[30] *Measurement of the production cross-section of isolated photon pairs in pp collisions at 13 TeV with the ATLAS detector*. Tech. rep. ATLAS-CONF-2020-024. Geneva: CERN, July 2020. URL: http://cds.cern.ch/record/2725552 (cit. on pp. 21, 23).

[31] *Measurements and interpretations of Higgs-boson fiducial cross sections in the diphoton decay channel using 139 fb^{-1} of pp collision data at $\sqrt{s}$ = 13 TeV with the ATLAS detector*. Tech. rep. ATLAS-CONF-2019-029. Geneva: CERN, July 2019. URL: http://cds.cern.ch/record/2682800 (cit. on p. 21).

[32] Stefano Frixione. "Isolated photons in perturbative QCD". In: *Phys. Lett. B* 429 (1998), pp. 369–374. DOI: 10.1016/S0370-2693(98)00454-7. arXiv: hep-ph/9801442 (cit. on p. 22).

[33] Zachary Hall and Jesse Thaler. "Photon isolation and jet substructure". In: *JHEP* 09 (2018), p. 164. DOI: 10.1007/JHEP09(2018)164. arXiv: 1805.11622 [hep-ph] (cit. on p. 22).

[34] E.W.Nigel Glover and A.G. Morgan. "Measuring the photon fragmentation function at LEP". In: *Z. Phys. C* 62 (1994), pp. 311–322. DOI: 10.1007/BF01560245 (cit. on p. 22).

[35] Stefano Catani et al. "Diphoton production at hadron colliders: a fully-differential QCD calculation at NNLO". In: *Phys. Rev. Lett.* 108 (2012). [Erratum: Phys.Rev.Lett. 117, 089901 (2016)], p. 072001. DOI: 10.1103/PhysRevLett.108.072001. arXiv: 1110.2375 [hep-ph] (cit. on p. 22).

[36] *Standard Model Summary Plots Spring 2020*. Tech. rep. ATL-PHYS-PUB-2020-010. Geneva: CERN, May 2020. URL: http://cds.cern.ch/record/2718937 (cit. on p. 23).

[37] R. D. Peccei and Helen R. Quinn. "CP Conservation in the Presence of Pseudoparticles". In: *Phys. Rev. Lett.* 38 (25 June 1977), pp. 1440–1443. DOI: 10.1103/PhysRevLett.38.1440. URL: https://link.aps.org/doi/10.1103/PhysRevLett.38.1440 (cit. on p. 24).

[38] Howard Georgi, David B. Kaplan, and Lisa Randall. "Manifesting the invisible axion at low energies". In: *Physics Letters B* 169.1 (1986), pp. 73–78. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(86)90688-X. URL: http://www.sciencedirect.com/science/article/pii/037026938690688X (cit. on pp. 24, 26).

[39] I.G. Irastorza et al. "Towards a new generation axion helioscope". In: *JCAP* 06 (2011), p. 013. DOI: 10.1088/1475-7516/2011/06/013. arXiv: 1103.5334 [hep-ex] (cit. on pp. 24, 27).

[40] Hyun Min Lee, Myeonghun Park, and Wan-Il Park. "Fermi Gamma Ray Line at 130 GeV from Axion-Mediated Dark Matter". In: *Phys. Rev. D* 86 (2012), p. 103502. DOI: 10.1103/PhysRevD.86.103502. arXiv: 1205.4675 [hep-ph] (cit. on p. 24).

[41] Peter Svrcek and Edward Witten. "Axions In String Theory". In: *JHEP* 06 (2006), p. 051. DOI: 10.1088/1126-6708/2006/06/051. arXiv: hep-th/0605206 (cit. on p. 24).

[42] Jonathan L. Rosner. "CP violation: A Brief review". In: *AIP Conf. Proc.* 540.1 (2000). Ed. by J.F. Nieves, pp. 283–304. DOI: 10.1063/1.1328890. arXiv: hep-ph/0005258 (cit. on p. 24).

[43] N. Aghanim et al. "Planck 2018 results. VI. Cosmological parameters". In: *Astron. Astrophys.* 641 (2020), A6. DOI: 10.1051/0004-6361/201833910. arXiv: 1807.06209 [astro-ph.CO] (cit. on p. 24).

[44] S. Perlmutter et al. "Measurements of Ω and Λ from 42 high redshift supernovae". In: *Astrophys. J.* 517 (1999), pp. 565–586. DOI: 10.1086/307221. arXiv: astro-ph/9812133 (cit. on p. 24).

[45] Adam G. Riess et al. "The risetime of nearby type ia supernovae". In: *Astron. J.* 118 (1999), pp. 2675–2688. DOI: 10.1086/301143. arXiv: astro-ph/9907037 (cit. on p. 24).

[46] Adam G. Riess et al. "Type Ia supernova discoveries at z $>$ 1 from the Hubble Space Telescope: Evidence for past deceleration and constraints on dark energy evolution". In: *Astrophys. J.* 607 (2004), pp. 665–687. DOI: 10.1086/383612. arXiv: astro-ph/0402512 (cit. on p. 24).

[47] V. C. Rubin, Jr. Ford W. K., and N. Thonnard. "Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R=4kpc) to UGC 2885 (R=122kpc)." In: *The Astrophysical Journal* 238 (June 1980), pp. 471–487. DOI: 10.1086/158003 (cit. on p. 24).

[48] M. J. Jee et al. "Discovery of a Ringlike Dark Matter Structure in the Core of the Galaxy Cluster Cl 0024+17". In: *The Astrophysical Journal* 661.2 (June 2007), pp. 728–749. DOI: 10.1086/517498. URL: https://doi.org/10.1086%2F517498 (cit. on p. 24).

[49] E. Komatsu et al. “Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation”. In: *Astrophys. J. Suppl.* 180 (2009), pp. 330–376. DOI: 10.1088/0067-0049/180/2/330. arXiv: 0803.0547 [astro-ph] (cit. on p. 24).

[50] M. Milgrom. “A Modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis”. In: *Astrophys. J.* 270 (1983), pp. 365–370. DOI: 10.1086/161130 (cit. on p. 24).

[51] Daniel Abercrombie et al. “Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum”. In: *Phys. Dark Univ.* 27 (2020). Ed. by Antonio Boveia et al., p. 100371. DOI: 10.1016/j.dark.2019.100371. arXiv: 1507.00966 [hep-ex] (cit. on p. 24).

[52] Stefan von Buddenbrock et al. “Phenomenological signatures of additional scalar bosons at the LHC”. In: *Eur. Phys. J. C* 76.10 (2016), p. 580. DOI: 10.1140/epjc/s10052-016-4435-8. arXiv: 1606.01674 [hep-ph] (cit. on p. 24).

[53] Juan García-Bellido. “Massive Primordial Black Holes as Dark Matter and their detection with Gravitational Waves”. In: *J. Phys. Conf. Ser.* 840.1 (2017). Ed. by Domencio Giardini and Philippe Jetzer, p. 012032. DOI: 10.1088/1742-6596/840/1/012032. arXiv: 1702.08275 [astro-ph.CO] (cit. on p. 24).

[54] Céline Boehm et al. “Extended gamma-ray emission from Coy Dark Matter”. In: *JCAP* 05 (2014), p. 009. DOI: 10.1088/1475-7516/2014/05/009. arXiv: 1401.6458 [hep-ph] (cit. on p. 25).

[55] Matthew J. Dolan et al. “A taste of dark matter: Flavour constraints on pseudoscalar mediators”. In: *JHEP* 03 (2015). [Erratum: JHEP 07, 103 (2015)], p. 171. DOI: 10.1007/JHEP03(2015)171. arXiv: 1412.5174 [hep-ph] (cit. on p. 25).

[56] John Preskill, Mark B. Wise, and Frank Wilczek. “Cosmology of the Invisible Axion”. In: *Phys. Lett. B* 120 (1983). Ed. by M.A. Srednicki, pp. 127–132. DOI: 10.1016/0370-2693(83)90637-8 (cit. on p. 26).

[57] Anson Hook et al. “High Quality QCD Axion and the LHC”. In: *Phys. Rev. Lett.* 124.22 (2020), p. 221801. DOI: 10.1103/PhysRevLett.124.221801. arXiv: 1911.12364 [hep-ph] (cit. on p. 26).

[58] V.A. Rubakov. “Grand unification and heavy axion”. In: *JETP Lett.* 65 (1997), pp. 621–624. DOI: 10.1134/1.567390. arXiv: hep-ph/9703409 (cit. on p. 26).

[59] Zurab Berezhiani, Leonida Gianfagna, and Maurizio Giannotti. “Strong CP problem and mirror world: The Weinberg-Wilczek axion revisited”. In: *Phys. Lett. B* 500 (2001), pp. 286–296. DOI: 10.1016/S0370-2693(00)01392-7. arXiv: hep-ph/0009290 (cit. on p. 26).

[60] Ann E. Nelson and Nathan Seiberg. “R symmetry breaking versus supersymmetry breaking”. In: *Nucl. Phys. B* 416 (1994), pp. 46–62. DOI: 10.1016/0550-3213(94)90577-0. arXiv: hep-ph/9309299 (cit. on p. 26).

[61] Gabriele Ferretti. “Gauge theories of Partial Compositeness: Scenarios for Run-II of the LHC”. In: *JHEP* 06 (2016), p. 107. DOI: 10.1007/JHEP06(2016)107. arXiv: 1604.06467 [hep-ph] (cit. on p. 26).

[62] Jared A. Evans, Stefania Gori, and Jessie Shelton. “Looking for the WIMP Next Door”. In: *JHEP* 02 (2018), p. 100. DOI: 10.1007/JHEP02(2018)100. arXiv: 1712.03974 [hep-ph] (cit. on p. 26).

[63] Xabier Cid Vidal et al. “New Axion Searches at Flavor Factories”. In: *JHEP* 01 (2019). [Erratum: JHEP 06, 141 (2020)], p. 113. DOI: 10.1007/JHEP01(2019)113. arXiv: 1810.09452 [hep-ph] (cit. on p. 26).

[64] Joerg Jaeckel, Martin Jankowiak, and Michael Spannowsky. “LHC probes the hidden sector”. In: *Phys. Dark Univ.* 2 (2013), pp. 111–117. DOI: 10.1016/j.dark.2013.06.001. arXiv: 1212.3620 [hep-ph] (cit. on p. 27).

[65] Stephen L. Adler. “Axial-Vector Vertex in Spinor Electrodynamics”. In: *Phys. Rev.* 177 (5 Jan. 1969), pp. 2426–2438. DOI: 10.1103/PhysRev.177.2426. URL: https://link.aps.org/doi/10.1103/PhysRev.177.2426 (cit. on p. 27).

[66] J.S. Bell and R. Jackiw. “A PCAC puzzle: $\pi^0 \to \gamma\gamma$ in the σ model”. In: *Nuovo Cim. A* 60 (1969), pp. 47–61. DOI: 10.1007/BF02823296 (cit. on p. 27).

[67] Abdelhak Djouadi. “The Anatomy of electro-weak symmetry breaking. II. The Higgs bosons in the minimal supersymmetric model”. In: *Phys. Rept.* 459 (2008), pp. 1–241. DOI: 10.1016/j.physrep.2007.10.005. arXiv: hep-ph/0503173 (cit. on p. 27).

[68] J. D. Bjorken et al. “Search for neutral metastable penetrating particles produced in the SLAC beam dump”. In: *Phys. Rev. D* 38 (11 Dec. 1988), pp. 3375–3386. DOI: 10.1103/PhysRevD.38.3375. URL: https://link.aps.org/doi/10.1103/PhysRevD.38.3375 (cit. on p. 27).

[69] H. Primakoff. “Photo-Production of Neutral Mesons in Nuclear Electric Fields and the Mean Life of the Neutral Meson”. In: *Phys. Rev.* 81 (5 Mar. 1951), pp. 899–899. DOI: 10.1103/PhysRev.81.899. URL: https://link.aps.org/doi/10.1103/PhysRev.81.899 (cit. on p. 27).

[70] O. Adriani et al. “Isolated hard photon emission in hadronic Z0 decays”. In: *Phys. Lett. B* 292 (1992), pp. 472–484. DOI: 10.1016/0370-2693(92)91205-N (cit. on pp. 27, 28, 160, 161).

[71] G. Abbiendi et al. “Multi-photon production in e+e- collisions at s= 189 GeV”. In: *Physics Letters B* 465.1 (1999), pp. 303–314. ISSN: 0370-2693. DOI: https://doi.org/10.1016/S0370-2693(99)01092-8. URL: http://www.sciencedirect.com/science/article/pii/S0370269399010928 (cit. on pp. 27, 28).

[72] Gautam Rupak and Elizabeth H. Simmons. “Limits on pseudoscalar bosons from rare Z decays at LEP”. In: *Phys. Lett. B* 362 (1995), pp. 155–163. DOI: 10.1016/0370-2693(95)01152-G. arXiv: hep-ph/9507438 (cit. on p. 27).

[73] Georges Aad et al. “Measurement of light-by-light scattering and search for axion-like particles with 2.2 nb^{-1} of Pb+Pb data with the ATLAS detector”. In: (Aug. 2020). arXiv: 2008.05355 [hep-ex] (cit. on p. 27).

[74] Albert M Sirunyan et al. “Evidence for light-by-light scattering and searches for axion-like particles in ultraperipheral PbPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *Phys. Lett. B* 797 (2019), p. 134826. DOI: 10.1016/j.physletb.2019.134826. arXiv: 1810.04602 [hep-ex] (cit. on p. 27).

[75] G. Jarlskog et al. “Measurement of Delbrück Scattering and Observation of Photon Splitting at High Energies”. In: *Phys. Rev. D* 8 (11 Dec. 1973), pp. 3813–3823. DOI: 10.1103/PhysRevD.8.3813. URL: https://link.aps.org/doi/10.1103/PhysRevD.8.3813 (cit. on p. 27).

[76] Alberto Mariotti et al. “New LHC bound on low-mass diphoton resonances”. In: *Phys. Lett. B* 783 (2018), pp. 13–18. DOI: 10.1016/j.physletb.2018.06.039. arXiv: 1710.01743 [hep-ph] (cit. on pp. 28, 160).

[77] B. Andersson et al. “Parton fragmentation and string dynamics”. In: *Physics Reports* 97.2 (1983), pp. 31–145. ISSN: 0370-1573. DOI: https://doi.org/10.1016/0370-1573(83)90080-7. URL: http://www.sciencedirect.com/science/article/pii/0370157383900807 (cit. on p. 30).

[78] B.R. Webber. “A QCD Model for Jet Fragmentation Including Soft Gluon Interference”. In: *Nucl. Phys. B* 238 (1984), pp. 492–528. DOI: 10.1016/0550-3213(84)90333-X (cit. on p. 30).

[79] David J. Lange. “The EvtGen particle decay simulation package”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 462.1 (2001). BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines, pp. 152–155. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(01)00089-4. URL: http://www.sciencedirect.com/science/article/pii/S0168900201000894 (cit. on p. 30).

[80] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. “A Brief Introduction to PYTHIA 8.1”. In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph] (cit. on p. 30).

[81] ATLAS Collaboration. *ATLAS Run 1 Pythia8 tunes*. Tech. rep. 2014. URL: https://cds.cern.ch/record/1966419 (cit. on p. 30).

[82] Frank Siegert. “A practical guide to event generation for prompt photon production with Sherpa”. In: *Journal of Physics G: Nuclear and Particle Physics* 44.4 (Mar. 2017), p. 044007. DOI: 10.1088/1361-6471/aa5f29. URL: https://doi.org/10.1088%2F1361-6471%2Faa5f29 (cit. on p. 30).

[83] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph] (cit. on p. 31).

[84] Leif Lonnblad. “Correcting the color dipole cascade model with fixed order matrix elements”. In: *JHEP* 05 (2002), p. 046. DOI: 10.1088/1126-6708/2002/05/046. arXiv: hep-ph/0112284 (cit. on p. 31).

[85] GEANT4 Collaboration, S. Agostinelli et al. “GEANT4: A simulation toolkit”. In: *Nucl. Instrum. Meth.* A506 (2003), p. 250 (cit. on p. 31).

[86] ATLAS Collaboration. *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*. ATL-PHYS-PUB-2010-013. 2010. URL: https://cds.cern.ch/record/1300517 (cit. on p. 31).

[87] Fabienne Marcastel. “CERN’s Accelerator Complex. La chaîne des accélérateurs du CERN”. In: (Nov. 2013). General Photo. URL: https://cds.cern.ch/record/1621583 (cit. on p. 34).

[88] V Parma and L Rossi. “Performance of the LHC magnet system”. In: CERN-ATS-2009-023 (Sept. 2009), 6 p. URL: https://cds.cern.ch/record/1204578 (cit. on p. 34).

[89] B G Taylor. "Timing distribution at the LHC". In: (2002). DOI: 10.5170/CERN-2002-003.63. URL: https://cds.cern.ch/record/592719 (cit. on p. 35).

[90] C. Møller. *General Properties of the Characteristic Matrix in the Theory of Elementary Particles I.* The Royal Danish Academy of Sciences and Letters, 1945 (cit. on p. 36).

[91] Riccardo De Maria. "LHC Interaction region upgrade". Presented on 14 Apr 2008. 2008. URL: https://cds.cern.ch/record/1127611 (cit. on p. 37).

[92] The ATLAS Collaboration. *ATLAS online luminosity public plots.* Tech. rep. Geneva: CERN. 2019. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2 (cit. on pp. 37, 38).

[93] M. Aaboud et al. "Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC". In: *Phys. Rev. Lett.* 117.18 (2016), p. 182002. DOI: 10.1103/PhysRevLett.117.182002. arXiv: 1606.02625 [hep-ex] (cit. on pp. 37, 151).

[94] Giulia Papotti et al. "Operation of the LHC with Protons at High Luminosity and High Energy". In: CERN-ACC-2016-229 (2016), WEOCA01. 4 p. DOI: 10.18429/JACoW-IPAC2016-WEOCA01. URL: https://cds.cern.ch/record/2207396 (cit. on p. 37).

[95] Rende Steerenberg et al. "Operation and performance of the CERN Large Hadron Collider during proton Run 2". In: (2019), MOPMP031. 4 p. DOI: 10.18429/JACoW-IPAC2019-MOPMP031. URL: https://cds.cern.ch/record/2696126 (cit. on p. 37).

[96] Xin Chen. *Prospects of LHC Higgs Physics at the end of Run III.* Tech. rep. ATL-PHYS-PROC-2017-019. Geneva: CERN, Mar. 2017. URL: https://cds.cern.ch/record/2256082 (cit. on p. 38).

[97] G Aad and Bentvelsen. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008). Also published by CERN Geneva in 2010, S08003. 437 p. DOI: 10.1088/1748-0221/3/08/S08003. URL: https://cds.cern.ch/record/1129811 (cit. on pp. 39, 40, 42, 47–53).

[98] H. Bethe. "Theory of the Passage of Fast Corpuscular Rays Through Matter". In: *Annalen Phys.* 5 (1930), pp. 325–400. DOI: 10.1002/andp.19303970303 (cit. on p. 43).

[99] I. Wingerter-Seez. "Particle Physics Instrumentation". In: *2nd Asia-Europe-Pacific School of High-Energy Physics.* 2017, pp. 295–314. DOI: 10.23730/CYRSP-2017-002.295. arXiv: 1804.11246 [physics.ins-det] (cit. on p. 45).

[100] V.M. Abazov et al. "D0 Run IIB Upgrade Technical Design Report". In: (Dec. 2002) (cit. on p. 47).

[101] C Gabaldon. *Drift Time measurement in the ATLAS Liquid Argon electromagnetic calorimeter using cosmic muons.* Tech. rep. ATL-LARG-PROC-2010-013. Geneva: CERN, Oct. 2010. URL: https://cds.cern.ch/record/1301524 (cit. on p. 48).

[102] Georges Aad et al. "Performance of electron and photon triggers in ATLAS during LHC Run 2". In: *Eur. Phys. J. C* 80.1 (2020), p. 47. DOI: 10.1140/epjc/s10052-019-7500-2. arXiv: 1909.00761 [hep-ex] (cit. on pp. 52, 54, 55, 128, 151).

[103] R. Achenbach et al. "The ATLAS level-1 calorimeter trigger". In: *JINST* 3 (2008), P03001. DOI: 10.1088/1748-0221/3/03/P03001 (cit. on p. 54).

[104] William Panduro Vazquez and ATLAS Collaboration. *The ATLAS Data Acquisition system in LHC Run 2*. Tech. rep. ATL-DAQ-PROC-2017-007. 3. Geneva: CERN, Feb. 2017. DOI: 10.1088/1742-6596/898/3/032017. URL: https://cds.cern.ch/record/2244345 (cit. on p. 54).

[105] H et al Abreu. "Performance of the Electronic Readout of the ATLAS Liquid Argon Calorimeters". In: *JINST* 5 (2010), P09003. DOI: 10.1088/1748-0221/5/09/P09003. URL: https://cds.cern.ch/record/1303004 (cit. on pp. 58, 69).

[106] M. Aharrouche et al. "Measurement of the response of the ATLAS liquid argon barrel calorimeter to electrons at the 2004 combined test-beam". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 614.3 (2010), pp. 400–432. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2009.12.055. URL: http://www.sciencedirect.com/science/article/pii/S0168900209023924 (cit. on p. 59).

[107] W Lampl et al. *Calorimeter Clustering Algorithms: Description and Performance*. Tech. rep. ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003. Geneva: CERN, Apr. 2008. URL: http://cds.cern.ch/record/1099735 (cit. on pp. 59, 60).

[108] Morad Aaboud et al. "Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 79.8 (2019), p. 639. DOI: 10.1140/epjc/s10052-019-7140-6. arXiv: 1902.04655 [physics.ins-det] (cit. on pp. 60, 64).

[109] Georges Aad et al. "Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data". In: *Eur. Phys. J. C* 74.10 (2014), p. 3071. DOI: 10.1140/epjc/s10052-014-3071-4. arXiv: 1407.5063 [hep-ex] (cit. on pp. 60, 68).

[110] *Electron and photon reconstruction and performance in ATLAS using a dynamical, topological cell clustering-based approach*. Tech. rep. ATL-PHYS-PUB-2017-022. Geneva: CERN, Dec. 2017. URL: https://cds.cern.ch/record/2298955 (cit. on pp. 60, 62, 66, 67, 151).

[111] Georges Aad et al. "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1". In: *Eur. Phys. J. C* 77 (2017), p. 490. DOI: 10.1140/epjc/s10052-017-5004-5. arXiv: 1603.02934 [hep-ex] (cit. on pp. 61, 78).

[112] R. Fruhwirth. "Application of Kalman filtering to track and vertex fitting". In: *Nucl. Instrum. Meth. A* 262 (1987), pp. 444–450. DOI: 10.1016/0168-9002(87)90887-4 (cit. on p. 62).

[113] T G Cornelissen et al. "The global χ^2 track fitter in ATLAS". In: *Journal of Physics: Conference Series* 119.3 (July 2008), p. 032013. DOI: 10.1088/1742-6596/119/3/032013. URL: https://doi.org/10.1088%2F1742-6596%2F119%2F3%2F032013 (cit. on p. 62).

[114] *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*. Tech. rep. ATLAS-CONF-2012-047. Geneva: CERN, May 2012. URL: https://cds.cern.ch/record/1449796 (cit. on p. 62).

[115] Morad Aaboud et al. “Electron and photon energy calibration with the ATLAS detector using 2015-2016 LHC proton-proton collision data”. In: *JINST* 14.03 (2019), P03017. DOI: 10.1088/1748-0221/14/03/P03017. arXiv: 1812.03848 [hep-ex] (cit. on pp. 70, 71).

[116] *Photon shower shapes data/mc comparisons from z+γ events*. Tech. rep. ATL-COM-PHYS-2013-600. Geneva: CERN, May 2012 (cit. on p. 72).

[117] ATLAS Collaboration. *Electron and photon performance measurements with the ATLAS detector using the 2015-2017 LHC proton-proton collision data*. 2019. arXiv: 1908.00005 [hep-ex] (cit. on pp. 74, 76, 92, 95).

[118] Morad Aaboud et al. “Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016”. In: *Eur. Phys. J. C* 79.3 (2019), p. 205. DOI: 10.1140/epjc/s10052-019-6650-6. arXiv: 1810.05087 [hep-ex] (cit. on pp. 75, 84).

[119] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. “The catchment area of jets”. In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 005–005. DOI: 10.1088/1126-6708/2008/04/005. URL: https://doi.org/10.1088%2F1126-6708%2F2008%2F04%2F005 (cit. on p. 80).

[120] Matteo Cacciari and Gavin P. Salam. “Dispelling the N3 myth for the kt jet-finder”. In: *Physics Letters B* 641.1 (2006), pp. 57–61. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2006.08.037. URL: http://www.sciencedirect.com/science/article/pii/S0370269306010094 (cit. on p. 80).

[121] Stephen D. Ellis and Davison E. Soper. “Successive combination jet algorithm for hadron collisions”. In: *Phys. Rev. D* 48 (7 Oct. 1993), pp. 3160–3166. DOI: 10.1103/PhysRevD.48.3160. URL: https://link.aps.org/doi/10.1103/PhysRevD.48.3160 (cit. on p. 80).

[122] S. Catani et al. “Longitudinally-invariant kt-clustering algorithms for hadron-hadron collisions”. In: *Nuclear Physics B* 406.1 (1993), pp. 187–224. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(93)90166-M. URL: http://www.sciencedirect.com/science/article/pii/055032139390166M (cit. on p. 80).

[123] Georges Voronoi. “Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Premier mémoire. Sur quelques propriétés des formes quadratiques positives parfaites.” In: *Journal für die reine und angewandte Mathematik (Crelles Journal)* 1908 (), pp. 102–97 (cit. on p. 80).

[124] Jerome H. Friedman. “A VARIABLE SPAN SMOOTHER”. In: (Oct. 1984) (cit. on p. 91).

[125] Georges Aad et al. “Search for Scalar Diphoton Resonances in the Mass Range $65-600$ GeV with the ATLAS Detector in pp Collision Data at $\sqrt{s} = 8\ TeV$”. In: *Phys. Rev. Lett.* 113.17 (2014), p. 171801. DOI: 10.1103/PhysRevLett.113.171801. arXiv: 1407.6583 [hep-ex] (cit. on pp. 117, 124, 130, 160, 161).

[126] *Search for new resonances in the diphoton final state in the mass range between 70 and 110 GeV in pp collisions at $\sqrt{s} = 8$ and 13 TeV*. Tech. rep. CMS-PAS-HIG-17-013. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2285326 (cit. on pp. 117, 160, 161).

[127] *Search for resonances in the 65 to 110 GeV diphoton invariant mass range using 80 fb^{-1} of pp collisions collected at $\sqrt{s} = 13$ TeV with the ATLAS detector*. Tech. rep. ATLAS-CONF-2018-025. Geneva: CERN, July 2018. URL: http://cds.cern.ch/record/2628760 (cit. on pp. 117, 197).

[128] *Estimate of the m_H shift due to interference between signal and background processes in the $H \to \gamma\gamma$ channel, for the $\sqrt{s} = 8$ TeV dataset recorded by ATLAS*. Tech. rep. ATL-PHYS-PUB-2016-009. Geneva: CERN, Apr. 2016. URL: http://cds.cern.ch/record/2146386 (cit. on p. 120).

[129] P. Artoisenet et al. "A framework for Higgs characterisation". In: *JHEP* 11 (2013), p. 043. DOI: 10.1007/JHEP11(2013)043. arXiv: 1306.6464 [hep-ph] (cit. on pp. 120, 121).

[130] J C Collins, Davison Eugene Soper, and George F Sterman. "Transverse momentum distribution in Drell-Yan pair and W and Z boson production". In: *Nucl. Phys. B* 250.CERN-TH-3923 (June 1984), 199–224. 37 p. DOI: 10.1016/0550-3213(85)90479-1. URL: https://cds.cern.ch/record/153460 (cit. on p. 121).

[131] Stefano Catani et al. "Diphoton production at the LHC: a QCD study up to NNLO". In: *JHEP* 04 (2018), p. 142. DOI: 10.1007/JHEP04(2018)142. arXiv: 1802.02095 [hep-ph] (cit. on p. 122).

[132] T. Gleisberg et al. "Event generation with SHERPA 1.1". In: *JHEP* 02 (2009), p. 007. DOI: 10.1088/1126-6708/2009/02/007. arXiv: 0811.4622 [hep-ph] (cit. on p. 124).

[133] Enrico Bothmann et al. "Event Generation with Sherpa 2.2". In: (2019). arXiv: 1905.09127 [hep-ph] (cit. on p. 124).

[134] Steffen Schumann and Frank Krauss. "A Parton shower algorithm based on Catani-Seymour dipole factorisation". In: *JHEP* 03 (2008), p. 038. DOI: 10.1088/1126-6708/2008/03/038. arXiv: 0709.1027 [hep-ph] (cit. on p. 124).

[135] Stefan Höche et al. "A critical appraisal of NLO+PS matching methods". In: *JHEP* 09 (2012), p. 049. DOI: 10.1007/JHEP09(2012)049. arXiv: 1111.1220 [hep-ph] (cit. on p. 124).

[136] Stefan Höche et al. "QCD matrix elements + parton showers: The NLO case". In: *JHEP* 04 (2013), p. 027. DOI: 10.1007/JHEP04(2013)027. arXiv: 1207.5030 [hep-ph] (cit. on p. 124).

[137] *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*. Tech. rep. ATLAS-CONF-2019-021. Geneva: CERN, June 2019. URL: https://cds.cern.ch/record/2677054 (cit. on pp. 129, 150).

[138] *Expected photon performance in the ATLAS experiment*. Tech. rep. ATL-PHYS-PUB-2011-007. Geneva: CERN, Apr. 2011. URL: http://cds.cern.ch/record/1345329 (cit. on p. 137).

[139] L Carminati et al. *Measurement of the isolated di-photon cross section in 4.9 fb-1 of pp collisions at sqrt(s) = 7 TeV with the ATLAS detector*. Tech. rep. ATL-COM-PHYS-2012-592. Geneva: CERN, May 2012. URL: https://cds.cern.ch/record/1450063 (cit. on p. 139).

[140] *Recommendations for the Modeling of Smooth Backgrounds*. Tech. rep. ATL-PHYS-PUB-2020-028. Geneva: CERN, Nov. 2020. URL: https://cds.cern.ch/record/2743717 (cit. on p. 142).

[141] Ahmed Tarek. "Measurement of Higgs boson production cross sections in the diphoton channel with the full ATLAS Run-2 data and constraints on anomalous Higgs boson interactions". Presented 30 Sep 2019. 2019. URL: https://cds.cern.ch/record/2696211 (cit. on p. 148).

[142] S van der Meer. *Calibration of the effective beam height in the ISR*. Tech. rep. CERN-ISR-PO-68-31. ISR-PO-68-31. Geneva: CERN, 1968. URL: https://cds.cern.ch/record/296752 (cit. on p. 150).

[143] G. Avoni et al. "The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS". In: *Journal of Instrumentation* 13.07 (July 2018), P07017–P07017. DOI: 10.1088/1748-0221/13/07/p07017. URL: https://doi.org/10.1088%2F1748-0221%2F13%2F07%2Fp07017 (cit. on p. 150).

[144] Rachel Jordan Hyneman. "Measuring Higgs Boson Couplings, including to the Top Quark, in the Diphoton Decay Channel with Run 2 Data Collected by the ATLAS Detector". Presented 04 Feb 2020. 2020. URL: https://cds.cern.ch/record/2712576 (cit. on p. 153).

[145] Mark Ebden. *Gaussian Processes: A Quick Introduction*. 2015. arXiv: 1505.02965 [math.ST] (cit. on p. 153).

[146] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006. ISBN: ISBN-13 978-0-262-18253-9 (cit. on p. 154).

[147] M. N. Gibbs. "Bayesian Gaussian Processes for Regression and Classification". PhD thesis. University of Cambridge, 1997 (cit. on p. 154).

[148] Albert M Sirunyan et al. "Search for low mass vector resonances decaying into quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 01 (2018), p. 097. DOI: 10.1007/JHEP01(2018)097. arXiv: 1710.00159 [hep-ex] (cit. on pp. 160, 161).

[149] A.N. Kolmogorov. *Foundations of the theory of probability*. New York: Chelsea Pub. Co., 1950 (cit. on p. 165).

[150] B. de Finetti. *Theory of Probability: A critical introductory treatment*. Wiley Series in Probability and Statistics. Wiley, 2017. ISBN: 9781119286370. URL: https://books.google.es/books?id=FzjoDQAAQBAJ (cit. on p. 166).

[151] F. James. *Statistical Methods in Experimental Physics*. World Scientific, 2006. ISBN: 9789812567956. URL: https://books.google.es/books?id=QbBm2VhV5TQC (cit. on pp. 166, 170).

[152] J. Neyman and E. S. Pearson. "On the Problem of the Most Efficient Tests of Statistical Hypotheses". In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 231 (1933), pp. 289–337. ISSN: 02643952. URL: http://www.jstor.org/stable/91247 (cit. on p. 169).

[153] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J. C* 71 (2011). [Erratum: Eur.Phys.J.C 73, 2501 (2013)], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an] (cit. on pp. 173, 174).

[154] S. S. Wilks. “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses”. In: *Ann. Math. Statist.* 9.1 (Mar. 1938), pp. 60–62. DOI: 10.1214/aoms/1177732360. URL: https://doi.org/10.1214/aoms/1177732360 (cit. on p. 173).

[155] Abraham Wald. “Tests of statistical hypotheses concerning several parameters when the number of observations is large”. In: *Transactions of the American Mathematical Society* 54 (1943), pp. 426–482 (cit. on p. 173).

[156] Jacques Colas et al. “Crosstalk in the ATLAS Electromagnetic Calorimeter”. In: (Oct. 1999) (cit. on pp. 185–187).

[157] J Labbe and R Ishmukhametov. *Crosstalk Measurements in the Electromagnetic Calorimeter during ATLAS Final Installation*. Tech. rep. ATL-LARG-INT-2009-004. ATL-COM-LARG-2008-012. Geneva: CERN, Dec. 2008. URL: https://cds.cern.ch/record/1143373 (cit. on pp. 185–188).

[158] Alvaro Lopez Solis. “Contribution to the energy and time measurement of electrons and photons in the ATLAS experiment and search for dark matter production in association to a Higgs boson.. Contribution à la mesure d’énergie et du temps des électrons et des photons dans l’expérience ATLAS et recherche de la production de matière noire en association avec un boson de Higgs.” Presented 28 Sep 2017. July 2017. URL: https://cds.cern.ch/record/2291698 (cit. on pp. 187–189).

Acknowledgements

Before expressing my gratitude to all the people that has been by my side in these 3 (well, 4) years of PhD, I would like to stress something that I myself find relevant: I had help. Help in every known form: at the scientific, emotional and physical levels. It seems impossible for me to verbalize how lucky and grateful I feel, but I will try.

In the first place, naturally, I would like to express my eternal gratitude to José Ocariz, for being the best supervisor I could have ever had: motivating, creative and always available for a nice discussion. From the first moment I arrived for my "stage", you have always guided me in the most pedagogical possible way, knowing exactly what I needed to hear to understand and solve a problem. This incredible ability is something I aspire and hope to learn with time and experience. You have also provided me with opportunities to show my (our) work in many different groups and on very diverse topics, always combined with large doses of coffee and good humour (from finding rhinoceroses in plots to exploring ways of avoiding the $\%\#^*$ spurious signal). For all this, and more, I will always be eternally grateful. Gracias José!

I want to express my gratitude to the members of the jury for their presence and dedication during my defence, specially to the rapporteurs Ximo Poveda and Fabrice Couderc, who in these turbulent times managed to find the time to thoroughly read my manuscript and provide me with numerous comments, but always constructive ones.

This work would not have been possible without the support of the ATLAS group at the LPNHE, in particular to my "photon people". I would like to specially thank Ioannis Nomidis, for all your accurate and useful comments; Alex and Yufeng, for all the discussions on random questions I occasionally posed you in your office; and Bertrand Laforge and Lydia Roos, for all the comments and suggestions in our weekly meetings.

On the same line, I would like to thank the e/γ and HGamma groups within the ATLAS collaboration for your feedback and guidance on the various studies I have performed. Also, I would like to explicitly express my gratitude to all the members of the very-low mass analysis team: Josefina Alconada, Liron Barak, Nicolas Berger, Yuval Frid, Rachel Hyneman, José Ocariz and Diego Redigolo; for all the discussions and meetings together (and the ones yet to come). We are almost there!

I would like to thank all the PhD students of the LPNHE, for joining me in this incredible journey: from the always interesting conversations in my honorary office 1222-2-13 with Jad, Jean-Philippe and Romain; to the almost mandatory coffee break with the "kittens-gang": indispensable each one of you guys. Surely not forget to explicitly mention the last acquisition of the LPNHE, Микола Хандога: the best binome, colleague and friend I could have ever wanted: always with stimulating discussion topics and your never ending disposition to have a beer, a cup of tea or some горілка.

During this journey, I had the chance to live in the Maison Heinrich Heine of Cité Universitaire, where I have met incredible people and, before this pandemic, established fantastic traditions like "Wednesday Wine". I would like to explicitly thank Yuliya, for

your endless support and disposition for tea-time at any time; and Samia, for being the perfect friend during those tequila nights (and also when stealing potatoes). I want also to express my gratitude to Anna and Yvonne who found the time to come and visit me in Spain, because friendship does not understand about distance.

In spite of being far from home, three supporting spots have always being there for me. Starting from the closest in distance to Paris, my friends from Valdunquillo: Mario, Elena, David, Álvaro and Rocío. It is fair to say that my life would have been incredibly different without you around: at a personal level but also as a Pokémon trainer. Following in distance, my volleyball team, "La Resistencia": Pareja, Félix, Héctor and Carol: your support and occasional beers when I was back in town were extremely appreciated. Also in Madrid, the people who ignited this journey, my colleagues from my bachelor in physics: Dani, Bris, Gonxo, Carlos, Iván and Alex. I am extremely grateful to you all.

Finally, I would like to express my gratitude to my family. I hope the non-Spanish speakers can forgive me for switching to my mother tongue for this last part.

Gracias, por todo el apoyo que me habéis proporcionado desde la distancia. Gracias a mi abuela, la más fuerte de todos. Gracias a mis tíos y primos, por todas esas risas y momentos. A mi padre, por tu orden y tu forma de pensar. A mi madre, por tu desorden y tus ganas de enseñar. A mi hermano, porque eres una parte fundamental de la persona que soy hoy. Y porque si de familias se trata, me siento profundamente afortunado.

Y por último, a María, por encontrar el orden dentro de mi caos, por ser mi piedra angular, mi fuente de inspiración y mi compañera de equipo. Gracias por estar siempre ahí.