

UNIVERSITÉ PARIS CITÉ

École doctorale des Sciences de la Terre et de l'Environnement et Physique
de l'Univers, Paris - ED 560

Laboratoire de Physique Nucléaire et Hautes Énergies - UMR 7585

Study of photon isolation and search for axion-like particles decaying into collimated photon pairs using machine learning with the ATLAS detector at the LHC

Par ROMAIN VAN DEN BROUCKE

Thèse de doctorat de PHYSIQUE DE L'UNIVERS

Dirigée par JOSÉ OCARIZ

Présentée et soutenue publiquement le 25 septembre 2025

Devant un jury composé de :

MATTEO CACCIARI

Professeur des universités, Université Paris Cité

Président du jury

JOSÉ OCARIZ

Professeur des universités, Université Paris Cité

Directeur de thèse

JEAN-BAPTISTE DE VIVIE DE REGIE

Directeur de recherche, CNRS

Rapporteur

LINDA FINCO

Senior research scientist, INFN Sezione di Torino

Rapporteuse

ANA CUETO GÓMEZ

Assistant professor, Universidad Autónoma de Madrid

Examinatrice

PAULO CÉSAR FARIAS

Full professor, Universidade Federal da Bahia

Examineur

Résumé

Titre : Étude de l’isolation des photons et recherche de particules de type axion se désintégrant en paires de photons collimatés par machine learning avec le détecteur ATLAS au LHC

Cette thèse propose une recherche de particules de type axion se désintégrant en paires de photons très collimatés, avec une sensibilité étendue, pour la première fois, à des masses inférieures à 10 GeV, dans le cadre de collisionneurs de protons. Elle s’accompagne d’une étude de l’isolation des photons. Ce travail s’inscrit dans le cadre de l’expérience ATLAS du Grand collisionneur de hadrons (LHC) au CERN, et s’appuie sur les données collectées entre 2015 et 2018 (Run 2), à une énergie au centre de masse de 13 TeV, ainsi que sur celles de la période 2022–2024 (Run 3), à une énergie de 13,6 TeV.

La première partie de ce travail constitue une contribution aux travaux de performance de l’isolation des photons dans l’expérience ATLAS. Les photons produits directement lors des collisions hadroniques sont une sonde précieuse pour l’étude de la chromodynamique quantique. Il est primordial de les distinguer des sources de bruit de fond, provenant principalement des mésons légers produits dans les jets hadroniques. Pour cela, l’isolation calorimétrique applique une sélection basée sur les dépôts d’énergie mesurés autour du candidat photon dans le détecteur, les photons provenant de jets étant généralement entourés d’une activité hadronique significative.

Cependant, l’énergie d’isolation calorimétrique des photons est imparfaitement modélisée dans les simulations, qui sont pourtant essentielles à la préparation des analyses et à la construction des modèles de signal et de bruit de fond. Pour y remédier, des corrections sont effectuées directement aux distributions d’énergie d’isolation dans les simulations, afin de les faire correspondre aux données expérimentales. Une nouvelle méthodologie est proposée pour corriger les simulations du Run 2, avec une précision supérieure à celle des approches précédentes. Ces corrections prennent notamment en compte l’énergie du photon, le chevauchement de collisions dans le détecteur (pile-up), la région du détecteur concernée, ainsi que le type de conversion, une fraction significative des photons se convertissant en paires électron-positon dans le détecteur.

La deuxième partie de ce travail propose une recherche de nouvelle résonance dans le canal de désintégration en deux photons collimatés, atteignant des masses inférieures à 10 GeV et étendant ainsi la portée des analyses précédentes menées dans les collisionneurs de protons. Cependant, les outils standard d’identification des photons, basés sur la forme des gerbes électromagnétiques dans le détecteur, ainsi que les processus d’isolation, fondés sur les dépôts d’énergie environnants, ne sont pas adaptés à des photons très proches, dont les gerbes et l’activité énergétique se chevauchent dans le détecteur. Pour surmonter cette limitation, des outils d’apprentissage automatique (machine learning) ont été mis en œuvre afin de développer de nouveaux critères d’identification et d’isolation, spécifiquement adaptés aux paires de photons très collimatés. Cette approche permet

d'étendre le potentiel de recherche de résonances se désintégrant en paires de photons. Plusieurs architectures de réseaux de neurones ont été explorées : un perceptron multi-couche (MLP), un réseau de neurones convolutif (CNN), un réseau de neurones à graphes (GNN) et un Set Transformer. Les performances de ces modèles sont exploitées pour évaluer la sensibilité à la découverte d'une nouvelle résonance et pour établir des limites supérieures attendues sur la section efficace multipliée par le rapport d'embranchement.

Enfin, dans le contexte actuel du réchauffement climatique, la recherche joue un rôle non négligeable. Un état des lieux des émissions de gaz à effet de serre de la recherche française, ainsi que du CERN, est proposé. Le travail mené au sein du LPNHE, laboratoire d'accueil de cette thèse, visant à quantifier ses émissions carbone et à proposer des mesures pour les réduire, est également présenté.

Mots-clés : LHC, ATLAS, photons, isolation, pile-up, particules de type axion, machine learning

Abstract

Title: Study of photon isolation and search for axion-like particles decaying into collimated photon pairs using machine learning with the ATLAS detector at the LHC

This thesis presents a search for axion-like particles decaying into closely collimated photon pairs, extending the sensitivity, for the first time, to masses below 10 GeV in the context of proton colliders. It is complemented by a study of photon isolation. This work is carried out within the ATLAS experiment at the Large Hadron Collider (LHC) at CERN and is based on data collected between 2015 and 2018 (Run 2), at a centre-of-mass energy of 13 TeV, as well as data from the 2022-2024 period (Run 3), at 13.6 TeV.

The first part of this work contributes to the study of photon isolation performance in the ATLAS experiment. Photons produced directly in hadronic collisions are a valuable probe for studying quantum chromodynamics. It is essential to distinguish them from background sources, which mainly originate from light mesons produced within hadronic jets. To achieve this, calorimetric isolation applies a selection based on the energy deposited around the photon candidate in the detector, as photons originating from jets are typically surrounded by significant hadronic activity.

However, the calorimetric isolation energy of photons is imperfectly modelled in simulations, which are essential for preparing analyses and constructing signal and background models. To address this, corrections are applied directly to the isolation energy distributions in the simulations to bring them into agreement with experimental data. A new methodology is proposed for correcting the Run 2 simulations, offering greater accuracy than previous approaches. These corrections account for the photon energy, the overlap of collisions in the detector (pile-up), the detector region involved, and the photon conversion type, as a significant fraction of photons convert into electron-positron pairs within the detector.

The second part of this work proposes a search for a new resonance in the collimated two-photon decay channel, reaching masses below 10 GeV and thus extending the scope of previous analyses carried out at proton colliders. However, standard photon identification tools, based on the shape of electromagnetic showers in the detector, and isolation processes, based on the surrounding energy deposits, are not suitable for very close photons, whose showers and energy deposits overlap in the detector. To overcome this limitation, machine learning tools have been implemented to develop new identification and isolation criteria specifically adapted to highly collimated photon pairs. This approach extends the potential to search for resonances decaying into photon pairs. Several neural network architectures have been explored: a multilayer perceptron (MLP), a convolutional neural network (CNN), a graph neural network (GNN), and a Set Transformer. The performance of these models is exploited to assess sensitivity to the discovery of a new resonance and to establish expected upper limits on the effective cross-section times the branching ratio.

Finally, in the current context of global warming, research plays a significant role.

An inventory of greenhouse gas emissions from French research institutions and CERN is presented. The work carried out at the LPNHE, the host laboratory for this thesis, to quantify its carbon emissions and propose measures to reduce them, is also described.

Keywords: LHC, ATLAS, photons, isolation, pile-up, axion-like particles, machine learning

Résumé subsantiel

Le Modèle Standard de la physique des particules constitue le cadre théorique décrivant les particules élémentaires et leurs interactions. Bien qu'il reproduise de nombreux phénomènes avec une très grande précision, il présente plusieurs lacunes, ce qui a conduit au développement d'extensions et de nouvelles théories. Parmi celles-ci, les particules de type axion (ALPs) pourraient, selon leur masse et leur couplage au Modèle Standard, fournir une explication à la matière noire ou encore au problème de CP fort. Cette thèse propose une recherche de particules de type axion se désintégrant en deux photons très collimatés, ainsi qu'une étude de l'isolation des photons. Elle s'inscrit dans le cadre de l'expérience ATLAS du Grand Collisionneur de Hadrons (LHC) au CERN et utilise les données enregistrées entre 2015 et 2018 (Run 2), à une énergie au centre de masse de 13 TeV, puis entre 2022 et 2024 (Run 3), à une énergie au centre de masse de 13,6 TeV.

Le cadre théorique de ce travail est tout d'abord présenté. Le Modèle Standard est une théorie quantique des champs, dans laquelle les particules fondamentales sont décrites comme des excitations de champs, susceptibles d'interagir entre eux. Un bref aperçu des motivations et des fondements mathématiques de la théorie quantique des champs est donné, puis le cas spécifique du Modèle Standard de la physique des particules est abordé. Expérimentalement, l'ALP recherchée dans ce travail se désintègre très rapidement en deux photons, avant même d'atteindre les détecteurs. Identifier les vrais photons, dits *prompt*, issus du processus dur ou de la désintégration de particules lourdes comme le boson de Higgs, des *faux* photons, provenant de la désintégration de mésons légers au sein des jets, est alors primordial. La physique du photon au LHC est brièvement présentée.

Le Modèle Standard de la physique des particules, bien que puissamment descriptif et prédictif, présente plusieurs lacunes : il n'explique par exemple ni l'asymétrie matière-antimatière, ni la masse du Higgs, ni celle des neutrinos. Différentes extensions du modèle sont brièvement passées en revue, avec un focus particulier sur l'ALP, motivation principale de cette thèse. L'état actuel des contraintes existantes sur l'ALP, ainsi que les différentes expériences contribuant à sa recherche, est présenté. Enfin, les générateurs Monte Carlo, outils indispensables en physique des particules pour la modélisation de la chromodynamique quantique (QCD), l'analyse de données ou la conception de nouvelles expériences, sont introduits.

Le dispositif expérimental utilisé dans ce travail est ensuite détaillé. Cette thèse s'inscrit dans le cadre du LHC au CERN, et plus précisément dans celui de l'expérience ATLAS. Au plus proche du point d'interaction, le détecteur interne est chargé de reconstruire les traces et les vertex des particules chargées, contribuant notamment à l'identification des électrons et des muons. Autour de celui-ci, les calorimètres électromagnétiques et hadroniques mesurent l'énergie des particules électromagnétiques (notamment les électrons et les photons) ainsi que des jets. Enfin, le spectromètre à muons mesure les traces et les impulsions des muons, qui ne déposent pas toute leur énergie dans les calorimètres. Un puissant système d'aimants permet enfin de courber la trajectoire des

particules chargées, rendant possible la mesure de leur charge et de leur impulsion.

L'identification des photons est au cœur de l'analyse présentée dans cette thèse. Pour reconstruire, à partir des dépôts d'énergie dans les détecteurs, des photons exploitables dans une analyse, plusieurs étapes sont nécessaires. En traversant les calorimètres, un photon dépose de l'énergie dans plusieurs cellules. Un algorithme de clustering regroupe ces cellules en structures tridimensionnelles appelées *clusters*. D'autres algorithmes permettent de reconstruire les trajectoires des particules chargées dans le détecteur, ce qui rend possible l'identification des photons convertis en paires électron-positron. Après une étape de calibration de l'énergie, un nouveau processus permet de sélectionner les photons *prompt* tout en rejetant le bruit de fond provenant des jets hadroniques. Ce processus, appelé *identification*, applique des coupures rectangulaires sur les variables décrivant le développement latéral et longitudinal de la gerbe électromagnétique initiée par le candidat photon dans le calorimètre. Plusieurs critères (Loose, Medium et Tight), correspondant à différents niveaux de pureté, sont définis.

Enfin, les candidats photons ainsi sélectionnés passent une dernière étape destinée à rejeter du bruit de fond supplémentaire en exploitant l'activité hadronique environnante, les photons *prompt* étant généralement isolés de ce flux d'énergie hadronique : l'isolation. Deux variables d'isolation sont couramment utilisées : l'énergie d'isolation calorimétrique et l'énergie d'isolation des traces. La première utilise les dépôts d'énergie dans le calorimètre autour du candidat photon, dans un cône de taille fixe. Après soustraction du dépôt d'énergie du candidat photon lui-même et de celui dû aux collisions de protons secondaires (pile-up), une coupure, dépendant de l'énergie du candidat photon, est appliquée sur la somme des dépôts d'énergie. La seconde applique une coupure sur la somme des traces dans un cône centré sur le candidat photon.

L'énergie d'isolation calorimétrique est imparfaitement modélisée dans les simulations Monte Carlo, notamment parce que la correction du pile-up repose sur un modèle lui-même imparfait. Cette divergence entraîne des différences d'efficacité d'isolation entre données et simulations. Afin de ne pas biaiser les analyses, les distributions d'énergie d'isolation calorimétrique dans les simulations sont corrigées pour correspondre aux données. Différents processus sont utilisés : à basse masse, les Z radiatifs offrent une signature très pure, et à plus haute masse, des photons QCD sont exploités. Dans ce travail, une nouvelle méthodologie est proposée pour les photons QCD. Deux critères d'identification, Tight et Loose⁴, sont employés pour estimer les composantes de vrais photons (*prompt*, provenant du processus dur) et de faux photons (issus de la désintégration de mésons légers dans les jets hadroniques). Dans la simulation, deux types de photons *prompt* sont considérés : les photons directs, directement issus du processus dur, et les photons de fragmentation, ou Bremsstrahlung, provenant de la radiation d'un produit du processus dur. Les distributions de photons simulées sont modélisées séparément dans ces quatre régions (Tight direct, Tight Bremsstrahlung, Loose⁴ direct et Loose⁴ Bremsstrahlung), puis les composantes directes et de Bremsstrahlung sont combinées en une seule distribution, en tenant compte des sections efficaces et des efficacités de sélection associées à ces deux catégories. Les données sont ensuite modélisées à partir des formes des distributions de photons obtenues dans les simulations. Enfin, l'écart de position des distributions entre données et simulation est mesuré, puis paramétrisé en fonction de l'énergie des photons et du pile-up. Cette quantité permet de corriger la position des distributions d'énergie d'isolation calorimétrique simulées afin qu'elles correspondent aux données. Cette correction continue représente une amélioration par rapport aux corrections précédentes, qui étaient données par intervalles d'énergie et de pile-up.

Les critères d'identification et d'isolation ne sont pas adaptés aux paires de photons très collimatés, issues de la désintégration d'hypothétiques ALPs légères. En effet, lorsque deux photons sont trop proches, leurs gerbes électromagnétiques se chevauchent, altérant ainsi les variables de forme utilisées par le processus d'identification. De plus, le critère d'isolation, qui consiste à sommer les dépôts d'énergie dans un cône centré sur le candidat photon, est affecté par la présence du second photon. Le chevauchement des cônes d'isolation des deux photons entraîne une augmentation artificielle de l'énergie d'isolation, ce qui réduit fortement l'efficacité du critère d'isolation. Ces effets dégradent significativement l'efficacité de sélection des paires de photons très collimatés, qui devient nulle pour des masses invariantes inférieures à 10 GeV.

Une approche alternative est proposée dans cette thèse. Plutôt que de considérer les deux photons individuellement, la paire de photons (le *diphoton*) est traitée comme un seul objet, avec de nouvelles définitions des critères d'identification et d'isolation. Des réseaux de neurones ont été développés, exploitant les mêmes variables que les critères standards.

D'une part, pour l'identification, les variables décrivant la forme des gerbes électromagnétiques sont utilisées, ainsi que la séparation angulaire entre les deux photons. Cette dernière permet au réseau de neurones d'exploiter la corrélation entre la séparation angulaire et la forme des gerbes. Des événements simulés, aussi bien pour les vrais que pour les faux diphotons, sont employés.

D'autre part, pour l'isolation, les dépôts d'énergie dans un cône centré sur le diphoton sont considérés. La position 3D de ces clusters dans le calorimètre, ainsi que le pile-up, sont pris en compte par le réseau de neurones. En raison d'un manque de statistiques disponibles, une région de contrôle dans les données a été utilisée pour les faux diphotons.

Différentes architectures de réseaux de neurones ont été développées. Pour l'identification des diphotons, un perceptron multicouche a été utilisé. Pour l'isolation, le nombre de clusters variant d'un diphoton à l'autre, des architectures plus flexibles ont dû être mises en place. Plusieurs topologies ont été testées : un CNN, un GNN et un Set Transformer. Le GNN offre des performances légèrement meilleures.

L'efficacité de sélection des diphotons *prompt* est supérieure à 65 % entre 5 et 20 GeV, et dépasse 90 % au-delà de 10 GeV. L'efficacité de sélection des faux diphotons reste inférieure à 50 %, et descend à 20 % aux plus basses masses.

Enfin, sur la base des performances d'identification et d'isolation des diphotons ainsi développées, la sensibilité à la découverte d'une nouvelle résonance en paires de photons entre 5 et 20 GeV est estimée. Tout d'abord, les composantes du bruit de fond sont évaluées. Les deux principales sont le bruit de fond irréductible, composé de vrais photons d'origine chromodynamique, et le bruit de fond réductible, constitué de fausses paires de photons provenant de jets. Une région de contrôle sur les données, définie à partir de l'énergie d'isolation des traces, est utilisée afin d'estimer les proportions des différentes composantes de bruit de fond dans la région de signal.

Le bruit de fond ainsi obtenu est ensuite modélisé en fonction de la masse invariante des paires de photons. D'autre part, la forme d'un signal potentiel est modélisée. La largeur intrinsèque d'une hypothétique nouvelle résonance est supposée négligeable devant la résolution du calorimètre. Le signal est donc décrit par la résolution du calorimètre, en fonction de la masse.

Le bruit de fond effectif, défini comme le bruit de fond sous l'hypothétique résonance, est estimé en fonction de la masse : des *toys* de bruit de fond sont générés, puis ajustés avec une fonction combinant signal et bruit de fond. L'incertitude sur la quantité de

signal renseigne alors directement sur le bruit de fond effectif.

Enfin, la limite supérieure attendue à 95 % de niveau de confiance sur la section efficace fiducielle multipliée par le rapport d'embranchement d'une hypothétique résonance en paire de photons est obtenue. Les nouveaux critères d'identification et d'isolation orientés diphotons développés dans cette thèse permettent de renforcer les contraintes existantes entre 10 et 20 GeV, ainsi que d'étendre le potentiel de recherche jusqu'à 5 GeV.

Enfin, dans le contexte actuel de réchauffement climatique, la recherche publique joue un rôle non négligeable. Un rappel de la situation planétaire est présenté, suivi d'un état des lieux des émissions de gaz à effet de serre de la recherche publique française, ainsi que du CERN, institution dans laquelle s'inscrit cette thèse. Le bilan carbone du LPNHE, laboratoire d'accueil de cette thèse, ainsi que la stratégie de réduction de ses émissions, développée au sein du groupe Développement Durable du laboratoire, sont également détaillés.

Au Campus.

Remerciements

Bien évidemment, cette thèse n'aurait pas pu se réaliser seule, et je souhaite remercier ici les personnes qui ont contribué à rendre ce travail possible au cours de ces trois années.

Mes premiers remerciements vont tout naturellement à mon encadrant de thèse, José. Je me souviens qu'en M2, nos professeurs nous ont inlassablement répété que choisir une thèse, c'est choisir un encadrant, avant de choisir un sujet. Il se trouve qu'ils avaient raison, et je suis heureux d'avoir fait le bon choix. Merci pour ces trois années passées ensemble, pour ton aide et ton accompagnement sans faille, pour toutes nos discussions sur la physique et tout autre sujet, pour ta bonne humeur ou encore ta pédagogie. Pour tout cela, merci !

Je remercie sincèrement les membres de mon jury d'avoir accepté d'évaluer ce manuscrit, pour leurs relectures attentives ainsi que pour leurs remarques et suggestions, qui ont grandement enrichi ce travail.

Je souhaite également remercier les "photon people" du LPNHE, ainsi que nos collègues brésiliens, pour nos nombreuses discussions et leurs conseils toujours pertinents lors de nos réunions. En particulier, merci Bertrand, pour ton aide précieuse et tes suggestions toujours utiles.

Plus largement, je remercie tout le groupe ATLAS du LPNHE, notamment Irena, Fred et Didier, pour leurs conseils, leur aide technique et nos échanges toujours enrichissants. Je tiens aussi à saluer le groupe Egamma d'ATLAS, au sein duquel s'est inscrit mon projet de recherche.

Je souhaite exprimer ma reconnaissance au groupe Développement Durable du LPNHE, pour son travail essentiel en faveur de la réduction de nos émissions de gaz à effet de serre, un effort indispensable et exemplaire au sein du laboratoire. Je ne doute pas que vous continuerez avec acharnement le travail que nous avons entamé, et j'espère qu'il portera ses fruits !

J'adresse aussi mes remerciements à toutes les personnes du LPNHE, chercheurs, ingénieurs, techniciens, personnels administratifs et doctorants, pour leur soutien et pour le cadre de travail stimulant qu'elles offrent au quotidien.

Je tiens à remercier Jean-Philippe, mon parrain au laboratoire, pour sa disponibilité, son écoute et ses conseils toujours justes.

Je souhaite adresser un mot particulier à Véronique, avec qui j'ai effectué mon stage de M1 (sans lequel je ne serai peut-être pas là aujourd'hui !) et avec qui j'ai eu la chance et le plaisir d'enseigner durant ma thèse.

Enfin, je remercie chaleureusement tous mes proches, amis et membres de ma famille, qui êtes venus à ma soutenance. J'espère avoir réussi à vous transmettre un peu de ma passion pour la physique ! Et merci Servane pour la thermodynamique.

Contents

Acronyms	1
Overview of the manuscript	3
1 Theoretical context	5
1.1 Quantum Field Theory	6
1.1.1 The need for a Quantum Field Theory	6
1.1.2 Lagrangian formalism and symmetries	7
1.1.3 Scattering amplitudes	8
1.1.4 Renormalisation	10
1.1.5 Effective Field Theory	12
1.2 The Standard Model of particle physics	13
1.2.1 Symmetries	14
1.2.2 Structure of the Standard Model	15
1.2.3 Electroweak symmetry breaking	16
1.2.4 The Standard Model Lagrangian	18
1.3 Photon physics at hadron colliders	19
1.3.1 Hadron interactions	19
1.3.2 Prompt photon production at the LHC	21
1.3.3 Photon isolation	22
1.3.4 Non-resonant diphoton backgrounds at the LHC	22
1.4 Beyond the Standard Model	23
1.4.1 Limitations of the Standard Model	23
1.4.2 Physics beyond the Standard Model	25
1.5 Axion-like particles	25
1.5.1 Resonant axion-like particle phenomenology at hadron colliders	26
1.5.2 Existing constraints on axion-like particles	26
1.6 Monte Carlo event generators	27
2 Experimental setup	31
2.1 CERN	32
2.2 The Large Hadron Collider	33
2.2.1 Accelerator overview	33
2.2.2 Luminosity and pile-up	34
2.3 The High-Luminosity Large Hadron Collider	35
2.4 The ATLAS experiment	36
2.4.1 Overview	36
2.4.2 Coordinate system	37
2.4.3 The Inner Detector	39

2.4.4	The calorimeter system	43
2.4.5	The muon spectrometer	50
2.4.6	The magnet system	51
2.4.7	Trigger and data acquisition system	53
3	Photon reconstruction and identification	55
3.1	Electron and photon reconstruction	56
3.1.1	Energy measurement in the LAr calorimeter	56
3.1.2	Clustering algorithms	59
3.1.3	Track reconstruction	63
3.1.4	Track-cluster matching and photon conversion reconstruction	64
3.1.5	Supercluster reconstruction	65
3.1.6	Electrons and photons at analysis-level	67
3.2	Electron and photon energy calibration	67
3.2.1	EM cluster energy	68
3.2.2	LAr layer calibration	69
3.2.3	Determination of the energy scale and resolution with $Z \rightarrow ee$ events	70
3.2.4	Uniformity corrections and ADC non-linearity correction	72
3.2.5	Photon leakage correction	73
3.2.6	Energy linearity and constraints on the calibration uncertainties	74
3.2.7	Calibration cross-checks	75
3.3	Photon identification	75
3.3.1	Photon identification optimisation	75
3.3.2	Photon identification efficiency	77
4	Photon isolation	83
4.1	Calorimetric isolation energy	84
4.2	Track-based isolation	87
4.3	Isolation criteria	87
4.4	Photon isolation corrections	88
4.4.1	Datasets and MC samples	89
4.4.2	Methodology	91
4.4.3	Function	91
4.4.4	Binning	92
4.4.5	Fit to MC distributions	93
4.4.6	Fit to MC PDF parameters	94
4.4.7	Combination of direct and Bremsstrahlung MC regions	96
4.4.8	Fit to data	100
4.4.9	Data-driven shift modelling	100
4.4.10	Conclusion	101
5	Neural network principles	105
5.1	Overview	106
5.2	Perceptron	107
5.3	Gradient descent	108
5.4	Training and validation	112
5.5	Optimisation	113
5.6	More complex architectures	114
5.6.1	Convolutional neural networks	114

5.6.2	Graph neural networks	117
5.6.3	Set Transformers	119
5.7	Machine Learning in this work	125
6	Diphoton identification and isolation	127
6.1	Closely collimated photon pairs	128
6.2	Event samples	129
6.2.1	Simulated prompt diphoton sample	130
6.2.2	Simulated dijet sample	132
6.2.3	Fake-enriched data sample	133
6.3	New diphoton identification criterion	133
6.3.1	Input preparation	134
6.3.2	Architecture of the neural network	138
6.3.3	Results and performance	139
6.3.4	Conclusion	140
6.4	New diphoton isolation criterion	142
6.4.1	Diphoton isolation with a CNN	144
6.4.2	Diphoton isolation with a GNN	150
6.4.3	Diphoton isolation with a Set Transformer	156
6.4.4	Conclusion	158
7	Sensitivity to a new diphoton resonance	161
7.1	Background components	162
7.2	Mass resolution	169
7.3	Effective background and expected limit	171
8	Carbon footprint of research	175
8.1	Global situation	176
8.1.1	Increase in greenhouse gas emissions	176
8.1.2	Global warming consequences	176
8.1.3	Scenarios for the future	179
8.2	Greenhouse gas emissions from research	179
8.2.1	French public research	179
8.2.2	CERN emissions	183
8.3	LPNHE's emissions	185
8.3.1	Laboratory presentation	185
8.3.2	GHG emissions assessment	187
8.3.3	Reduction measures	190
	Conclusions	197
A	Statistical tools	199
A.1	Probability interpretation	199
A.2	Probability densities and likelihood functions	200
A.3	Hypothesis testing	201
A.4	LHC statistical procedures	202
A.4.1	Test statistics	202
A.4.2	p -values	203
A.4.3	Expected sensitivity	205

A.4.4 Asymptotic formulas	205
List of Figures	207
List of Tables	219
Bibliography	221

Acronyms

ACB asymmetrical double-sided Crystal Ball

ALP axion-like particle

ATLAS **A** Toroidal **LHC** **A**pparatu**S**

BSM beyond the Standard Model

CMS Compact Muon Solenoid

CNN convolutional neural network

EFT Effective Field Theory

EM electromagnetic

GHG greenhouse gas

GNN graph neural network

HEP high-energy physics

HLT High-Level Trigger

ID identification

ID Inner Detector

IPCC Intergovernmental Panel on Climate Change

L1 Level-1 (trigger)

LAr liquid argon

LHC Large Hadron Collider

LO leading order

MC Monte Carlo

MS Muon Spectrometer

NLO next-to-leading order

NNLO next-to-next-to-leading order

PDF probability density function

PDF parton distribution function

pNGB pseudo-Nambu–Goldstone boson

PS presampler

QCD quantum chromodynamics

QED quantum electrodynamics

RoI Region of Interest

SCT Semiconductor Tracker

SM Standard Model

TRT Transition Radiation Tracker

Overview of the manuscript

The work presented in this manuscript is divided into two main parts: a study of photon isolation and a search for a new resonance with masses below 10 GeV in the collimated diphoton channel, a mass range still unexplored at the Large Hadron Collider (LHC).

Chapter 1 provides the theoretical background necessary for this work. It introduces the Standard Model of particle physics, beginning with a brief overview of the main ideas of quantum field theory. A short introduction to photon physics at hadron colliders is also included. The limitations of the Standard Model are outlined, along with possible extensions that address them, with a particular focus on axion-like particles, which serve as the primary motivation for this search. Monte Carlo generators, widely used in particle physics, are also discussed.

Chapter 2 describes the experimental setup used in this work: the LHC and the ATLAS experiment. The various ATLAS sub-detectors are presented, along with the trigger and data acquisition system.

Chapter 3 details the reconstruction of photons in the ATLAS software, starting from their energy deposits in the detector. Particular attention is paid to clustering algorithms. The energy calibration process is explained, and photon identification, the process of distinguishing photons based on the shape of their electromagnetic showers, is described in detail.

Chapter 4 presents the results of the first part of this work. It begins with an introduction to the concept of photon isolation, a method used to suppress background by analysing energy deposits around the photon candidate. A data-driven method is proposed to correct the simulated calorimetric isolation energy, which shows discrepancies compared to data. The methodology and results are presented.

Chapter 5 gives an overview of the main concepts of neural networks and machine learning, focusing on the topologies used in this work: the multilayer perceptron, convolutional and graph neural networks, and the Set Transformer.

Chapter 6 presents a performance study aimed at defining new identification and isolation criteria adapted to closely collimated photon pairs, as standard tools are unsuitable for low angular separations. A multilayer perceptron is used for the identification task, based on the photon pair angular separation and variables describing the photon electromagnetic showers. Different topologies are tested for the isolation task: a convolutional neural network, a graph neural network, and a Set Transformer, and their performances are compared.

Chapter 7 builds on the results of the previous chapter to extend the search potential for a diphoton resonance, with masses down to 5 GeV. The background event rates and the mass resolution are estimated. The effective background is then obtained, and an expected limit is set on the production of axion-like particles.

Chapter 8 gives an overview of the current global warming challenge, followed by a review of the status of the French research sector. The CERN carbon footprint is then

discussed, as this work is performed in the context of a CERN experiment. Finally, the work carried out at LPNHE to estimate its greenhouse gas emissions and propose ways to reduce them is presented.

Chapter 1

Theoretical context

Contents

1.1	Quantum Field Theory	6
1.1.1	The need for a Quantum Field Theory	6
1.1.2	Lagrangian formalism and symmetries	7
1.1.3	Scattering amplitudes	8
1.1.4	Renormalisation	10
1.1.5	Effective Field Theory	12
1.2	The Standard Model of particle physics	13
1.2.1	Symmetries	14
1.2.2	Structure of the Standard Model	15
1.2.3	Electroweak symmetry breaking	16
1.2.4	The Standard Model Lagrangian	18
1.3	Photon physics at hadron colliders	19
1.3.1	Hadron interactions	19
1.3.2	Prompt photon production at the LHC	21
1.3.3	Photon isolation	22
1.3.4	Non-resonant diphoton backgrounds at the LHC	22
1.4	Beyond the Standard Model	23
1.4.1	Limitations of the Standard Model	23
1.4.2	Physics beyond the Standard Model	25
1.5	Axion-like particles	25
1.5.1	Resonant axion-like particle phenomenology at hadron colliders	26
1.5.2	Existing constraints on axion-like particles	26
1.6	Monte Carlo event generators	27

1.1 Quantum Field Theory

This section provides a succinct presentation of quantum field theory, the theoretical and mathematical framework of the Standard Model of particle physics. It is primarily based on my readings of Ref. [1] and Ref. [2].

1.1.1 The need for a Quantum Field Theory

The concept of a *field* was introduced to replace the instantaneous action at a distance, characteristic of classical Newtonian and Coulombian laws, with a framework based on locality, which is philosophically more acceptable. However, there are other reasons for using fields in the quantum world.

First, due to special relativity, the number of particles is not fixed. When a particle of mass m is confined in a box of size L , the uncertainty in its momentum is $\Delta p \geq \hbar/L$. In special relativity, momentum and energy are treated on an equal footing, and a similar expression holds for the energy: $\Delta E \geq \hbar c/L$. When the energy uncertainty exceeds $2mc^2$, particle-antiparticle pairs can spontaneously emerge from the vacuum.

Second, particles of the same type are indistinguishable. For example, all protons observed on Earth or in cosmic rays are strictly identical, as if protons were “instances” of a more fundamental entity. It turns out that this primordial entity is the field (in the case of the proton, the quark and gluon fields), whose particles are localised excitations.

For instance, an elementary massive scalar field can be described by the Klein–Gordon equation:

$$\partial_\mu \partial^\mu \phi + m^2 \phi = 0, \quad (1.1)$$

where the Minkowski metric is used and the Einstein summation convention is applied, $\partial_\mu \partial^\mu = \partial_t^2 - \vec{\nabla}^2$. The general solution is a superposition of plane waves:

$$\phi(\vec{x}, t) = \int \frac{d^3 p}{(2\pi)^3} \left[a_p(t) e^{i\vec{p} \cdot \vec{x}} + a_p^*(t) e^{-i\vec{p} \cdot \vec{x}} \right], \quad (1.2)$$

where each mode $a_p(t)$ satisfies

$$(\partial_t^2 + \omega_p^2) a_p(t) = 0, \quad (1.3)$$

where $\omega_p = \sqrt{\vec{p} \cdot \vec{p} + m^2}$, which is the equation of a classical harmonic oscillator. Therefore, each momentum mode \vec{p} behaves like an independent oscillator. In quantum theory, the classical degrees of freedom are promoted to operators, a_p and a_p^\dagger , which are interpreted as annihilation and creation operators. The same procedure can be applied to the field—this is known as *second quantisation*. This leads to the Hamiltonian of the scalar field:

$$H_0 = \int \frac{d^3 p}{(2\pi)^3} \omega_p \left(a_p^\dagger a_p + \frac{1}{2} \right). \quad (1.4)$$

Each momentum mode \vec{p} corresponds to a quantum harmonic oscillator, and its excitations are interpreted as particles with momentum \vec{p} .

The field can be rewritten in a manifestly Lorentz-invariant form:

$$\phi(\vec{x}, t) = \int \frac{d^3 p}{(2\pi)^3} \frac{1}{\sqrt{2\omega_p}} \left[a_p e^{ip \cdot x} + a_p^\dagger e^{-ip \cdot x} \right]. \quad (1.5)$$

The time dependence of the a_p and a_p^\dagger operators has been extracted and placed in the exponentials, so that the Lorentz-invariant scalar product $p \cdot x$ appears explicitly. Additionally, the normalisation factor $\frac{1}{\sqrt{2\omega_p}}$ has been introduced to ensure that $\frac{d^3 p}{(2\pi)^3} \frac{1}{\sqrt{2\omega_p}}$ is Lorentz-invariant.

The full Hilbert space in QFT takes the form of a Fock space, defined as the direct sum of n -particle Hilbert spaces:

$$\mathcal{F} = \oplus_n \mathcal{H}_n. \quad (1.6)$$

The dynamics of quantum fields—such as the creation, evolution, and annihilation of particles—are described within this space.

1.1.2 Lagrangian formalism and symmetries

The Lagrangian formalism of classical mechanics can be applied to quantum fields, with a few minor adjustments: Poisson brackets are replaced by commutators (or anticommutators for fermionic fields), and care must be taken with the ordering of non-commuting operators.

The central quantity in this formalism is the Lagrangian density $\mathcal{L}(x)$, which is a function of the fields and their first derivatives. It encodes the fundamental properties of the fields and, by extremising the action,

$$S = \int dt L = \int d^4 x \mathcal{L}(x), \quad (1.7)$$

where L is the Lagrangian and $\mathcal{L}(x)$ the Lagrangian density, this leads to the Euler–Lagrange equations, which describe the dynamics of the fields:

$$\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \right) = 0. \quad (1.8)$$

In addition, the Lagrangian density is a key quantity for revealing the theory’s fundamental symmetries. Applying Noether’s theorem leads to conserved currents and, hence, conserved charges.

When constructing a quantum field theory, it is crucial to ensure that it respects the fundamental symmetries observed in nature. In particular, the laws of physics must be invariant under translations in space and time, reflecting the homogeneity of spacetime. They must also be invariant under spatial rotations due to the isotropy of space. Along with the Lorentz boosts, these symmetries form the Poincaré group, $\text{ISO}(1, 3)$, which is the group of isometries of Minkowski spacetime.

The Poincaré group has ten generators, with ten associated conserved quantities:

- Four generators correspond to spacetime translations (associated with the conservation of energy and momentum),
- Three to spatial rotations (conservation of angular momentum),
- And three to Lorentz boosts (related to the motion of the centre-of-mass frame).

Representations of the Poincaré group classify fields according to two fundamental quantities: their mass and spin. Spin is a purely quantum property that describes how a field transforms under rotations. Fundamental fields can have integer or half-integer spin:

- Spin-0 fields are called *scalars* (e.g., the Higgs field),
- Spin- $\frac{1}{2}$ fields are *spinors* (e.g., quarks and leptons),
- Spin-1 fields are *vectors* (e.g., the photon, W , and Z bosons).

Fields with integer spin follow Bose–Einstein statistics and are known as bosons: they are not subject to the Pauli exclusion principle and can occupy the same quantum state. Conversely, fields with half-integer spin follow Fermi–Dirac statistics and are referred to as fermions.

1.1.3 Scattering amplitudes

The calculation of scattering amplitudes in quantum field theory relies on the S -matrix. Assuming that, at asymptotic times ($t \rightarrow \pm\infty$), particles are in non-interacting (or free) states, and that interactions occur within a finite region of spacetime, the S -matrix encodes the probability for a given state $|i\rangle$ to evolve into a specific final state $|f\rangle$.

The differential probability for such a transition is given by:

$$dP = \frac{|\langle f|S|i\rangle|^2}{\langle f|f\rangle\langle i|i\rangle} d\Pi, \quad (1.9)$$

where $d\Pi$ is the differential phase space element corresponding to the final state. The normalisation factor in the denominator arises from the fact that the states are not normalised to one, but follow the relativistic normalisation condition:

$$\langle p|q\rangle = (2\pi)^3 2E_{\vec{p}} \delta^{(3)}(\vec{p} - \vec{q}). \quad (1.10)$$

The S -matrix is typically separated into two terms:

$$S = \mathbb{1} + i\mathcal{T}, \quad (1.11)$$

where the identity operator corresponds to the absence of interaction, and \mathcal{T} , referred to as the transfer matrix, encodes all the effects of the interaction.

Imposing four-momentum conservation on the process, the transfer matrix takes the form:

$$\mathcal{T} = (2\pi)^4 \delta^4 \left(\sum_i p_i^\mu - \sum_f p_f^\mu \right) \mathcal{M}, \quad (1.12)$$

where the sums run over the n_i initial and n_f final momenta, respectively, and \mathcal{M} is called the scattering amplitude. This is the fundamental quantity to compute for a given interaction process. If one restricts to the interacting part of the process:

$$\langle f|(S - \mathbb{1})|i\rangle = i(2\pi)^4 \delta^4 \left(\sum p^\mu \right) \langle f|\mathcal{M}|i\rangle, \quad (1.13)$$

where $\langle f|\mathcal{M}|i\rangle$, often referred to as the *matrix element*, governs the probability amplitude of the process.

In perturbation theory, the Lagrangian is divided into a non-interacting part \mathcal{L}_0 , in which the fields do not interact, and an interacting part \mathcal{L}_{int} , which is generally assumed to be small compared to the free part:

$$\mathcal{L}(\phi_1, \dots, \phi_n) = \mathcal{L}_0(\phi_1, \dots, \phi_n) + \mathcal{L}_{\text{int}}(\phi_1, \dots, \phi_n). \quad (1.14)$$

The solutions of the free part are assumed to be exactly known. The goal is to relate the solutions of the full interacting Lagrangian to those involving free fields. Several steps are needed.

First, the LSZ (Lehmann–Symanzik–Zimmermann) reduction formula allows one to relate these matrix elements to time-ordered products of fields:

$$\begin{aligned} \langle f | S | i \rangle &= \left[i \int d^4 x_1 e^{-ip_1 x_1} (\square_1 + m^2) \right] \dots \left[i \int d^4 x_n e^{ip_n x_n} (\square_n + m^2) \right] \\ &\times \langle \Omega | T \{ \phi(x_1) \dots \phi(x_n) \} | \Omega \rangle, \end{aligned} \quad (1.15)$$

where $|\Omega\rangle$ denotes the vacuum of the interacting theory, i.e. the lowest energy state with no particles; $T\{\phi(x_1) \dots \phi(x_n)\}$ are the time-ordered correlation functions, and the signs in the exponents indicate whether the particle is incoming ($-$) or outgoing ($+$). Here, the fields $\phi(x)$ are not free but fully interacting quantum fields.

Time-ordering is an operation that places operators at later times to the left of those at earlier times:

$$T\{\phi(x)\phi(y)\} = \begin{cases} \phi(x)\phi(y) & \text{if } x^0 > y^0, \\ \phi(y)\phi(x) & \text{otherwise,} \end{cases} \quad (1.16)$$

where $x^0 = t$ denotes the time coordinate.

From time-ordering, another important quantity can be defined, called the *Feynman propagator*. It is the vacuum expectation value (VEV) of the time-ordered product of two free fields:

$$D_F(x_1, x_2) = \langle 0 | T \{ \phi_0(x_1) \phi_0(x_2) \} | 0 \rangle = \lim_{\varepsilon \rightarrow 0} \int \frac{d^4 k}{(2\pi)^4} \frac{i}{k^2 - m^2 + i\varepsilon} e^{ik \cdot (x_1 - x_2)}, \quad (1.17)$$

where $|0\rangle$ is the vacuum of the free theory. The 0-index indicates that the fields are free. This Lorentz-invariant object has a pole at $k^2 = m^2$, cancelled by prefactors $(\square_i + m^2)$ in the LSZ reduction formula.

Then, interacting fields can be expressed in terms of free fields using perturbation theory:

$$\langle \Omega | T \{ \phi(x_1) \dots \phi(x_n) \} | \Omega \rangle = \frac{\langle 0 | T \left\{ \phi_0(x_1) \dots \phi_0(x_n) e^{i \int d^4 x \mathcal{L}_{\text{int}}[\phi_0]} \right\} | 0 \rangle}{\langle 0 | T \left\{ e^{i \int d^4 x \mathcal{L}_{\text{int}}[\phi_0]} \right\} | 0 \rangle}, \quad (1.18)$$

where $T \left\{ e^{i \int d^4 x \mathcal{L}_{\text{int}}[\phi_0]} \right\}$ is a Dyson series.

Finally, Wick's theorem allows us to express time-ordered products as sums of normal-ordered products and contractions:

$$T\{\phi_0(x_1)\dots\phi_0(x_n)\} = :\phi_0(x_1)\dots\phi_0(x_n): + \text{all possible contractions}, \quad (1.19)$$

where normal ordering $:\dots:$ consists of moving all creation operators to the left of annihilation operators. A contraction of two fields in a product of multiple fields involves replacing the pair with the corresponding Feynman propagator, leaving the other fields untouched.

This decomposition is particularly useful because vacuum expectation values of normal-ordered products vanish:

$$\langle 0| : \phi_0(x_1)\dots\phi_0(x_n) : |0\rangle = 0. \quad (1.20)$$

Thanks to all these tools—LSZ reduction, Feynman propagators, Wick’s theorem, and perturbation theory—the matrix elements can be calculated as a product of Feynman propagators, i.e., integrals over spacetime.

Although systematic, this procedure can become highly cumbersome due to the large number of contractions involved. Fortunately, Richard Feynman proposed in 1948 a pictorial representation of this perturbative expansion at different orders, with associated calculation rules, making it much easier to represent the process and perform the calculation. The construction of a diagram proceeds as follows:

- *External lines* are drawn for each incoming and outgoing particle: initial state particles from $|i\rangle$ are represented by incoming lines, while final state particles from $|f\rangle$ are represented by outgoing lines. A four-momentum p is assigned to each external line. Different conventions are used for the sense of time. In the following, the chosen convention is from left to right.
- *Arrows* are added on fermion lines, pointing to the right for fermions and to the left for anti-fermions. Spring-shaped lines are usually used for gluons, wave-shaped lines for electroweak bosons, and dashed lines for the Higgs boson.
- *Internal lines* connect vertices and represent virtual particles, which are not observed in the initial and final states. A momentum k is assigned to each internal line, and momentum conservation is imposed at each vertex.

To each Feynman diagram, a number is associated using the *Feynman rules*.

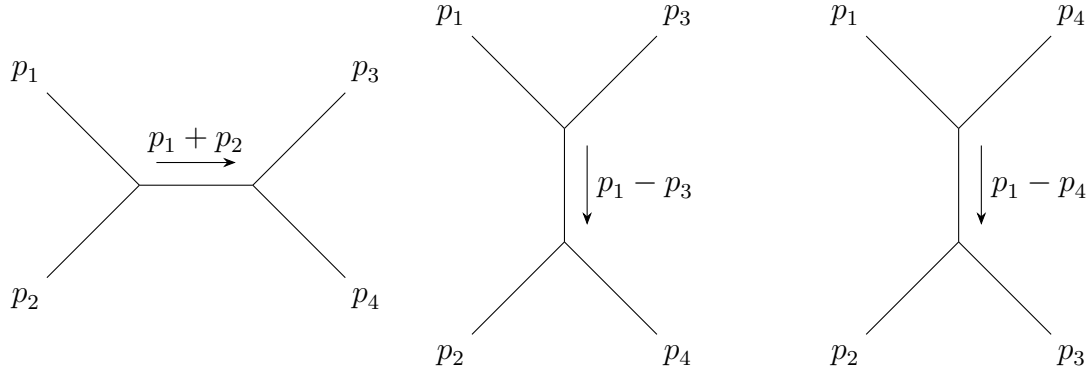
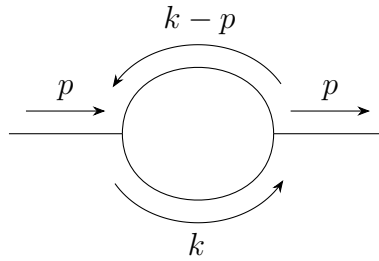
As an illustration, the Feynman diagrams at order λ^2 , associated with the Lagrangian density of a massive scalar particle with cubic self-interaction and coupling constant λ ,

$$\mathcal{L} = \frac{1}{2}\partial_\mu\partial^\mu\phi - m^2\phi^2 - \frac{\lambda}{3!}\phi^3, \quad (1.21)$$

are shown in Figure 1.1.

1.1.4 Renormalisation

Feynman diagrams are a very powerful tool to compute scattering amplitudes, but they do not come without difficulties. In particular, virtual particle four-momenta are not constrained and can take any arbitrary value ($k^2 \neq m^2$). They are said to be *off-shell*. When applying the Feynman rules, it is necessary to integrate over all possible momenta. This is a particular problem for loop diagrams when the calculation diverges. For example,


 Figure 1.1: Feynman diagrams at order λ^2 in the scalar ϕ^3 theory.

 Figure 1.2: One-loop in scalar ϕ^3 theory.

in the loop diagram of scalar ϕ^3 theory shown in Figure 1.2, the loop receives a factor of the form

$$\frac{1}{2}(i\lambda)^2 \int \frac{d^4k}{(2\pi)^4} \frac{i}{k^2 - m^2 + i\epsilon} \frac{i}{(k - p)^2 - m^2 + i\epsilon}. \quad (1.22)$$

At large k , this becomes

$$-\frac{1}{2}(i\lambda)^2 \int \frac{d^4k}{(2\pi)^4} \frac{1}{k^4}, \quad (1.23)$$

which is ultraviolet (UV) divergent.

Two techniques can be employed to avoid these infinities. The first is called *regularisation*, and it consists of using a cut-off Λ , beyond which the theory is considered no longer valid, and the integral is not performed. This allows one to avoid divergence, but the result is cut-off dependent.

The second, more fundamental technique is called *renormalisation*. It is based on the fact that the quantities considered here (e.g., mass and charge) are not physically observable since they do not take into account the “cloud” of virtual particles that interact with the considered particle. Bare quantities are thus re-scaled, or *renormalised*, to physical ones. The remaining parts of the bare quantities, and thus the corresponding infinities, are absorbed in *counterterms*, which cancel the divergences arising from other diagrams.

There are several ways to normalise the bare quantities, known as renormalisation schemes. Among the most common are the minimal subtraction scheme (MS), which consists of subtracting the divergent terms order by order; the modified minimal subtraction scheme ($\overline{\text{MS}}$), in which all terms independent of the renormalisation scale are

subtracted; and the on-shell scheme, which fixes the renormalised mass appearing in the Feynman propagator to the observed one.

The renormalisation procedure introduces an unphysical scale μ_R , known as the renormalisation scale. Physical quantities must be independent of the choice of this scale, leading to the renormalisation group (RG), which describes how the parameters transform under changes of parametrisation, through the renormalisation group equations (RGE). This is encoded in the “ β -function”:

$$\beta(g) = \mu_R \frac{dg}{d\mu_R}, \quad (1.24)$$

where g represents the coupling constant.

For quantum electrodynamics (QED), the β -function is positive:

$$\beta(e) = \mu_R \frac{de}{d\mu_R} = \frac{e^3}{12\pi^2}, \quad (1.25)$$

indicating that the coupling increases with increasing energy scale, which implies that the theory becomes non-interacting at large distance scales (i.e., low energy).

Conversely, for quantum chromodynamics (QCD), the β -function is negative, which has strong implications:

$$\beta(g_s) = -\frac{7}{16\pi^2} g_s^3, \quad (1.26)$$

where g_s is the strong coupling constant. Defining $\alpha_s = \frac{g_s^2}{4\pi}$ as the strong interaction coupling, it can be expressed as a function of the transferred momentum Q^2 :

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu_R^2)}{1 - \frac{\alpha_s(\mu_R^2)}{3\pi} \log\left(\frac{Q^2}{\mu_R^2}\right)}. \quad (1.27)$$

The minus sign in the β -function has major implications. At high transferred momentum Q^2 (i.e., at short distances or high energy scales), the coupling constant α_s decreases, and so does the intensity of the interaction. This is known as asymptotic freedom, and perturbation theory can be applied. At low Q^2 , or equivalently at large distances or low energy scales, the coupling increases and the theory becomes strongly interacting. Experimentally, quarks are never observed alone and always form bound states within hadrons, a phenomenon known as quark confinement. The energy dependence of the strong coupling constant, illustrated in Figure 1.3, vividly demonstrates this behaviour, with α_s decreasing at high energies.

1.1.5 Effective Field Theory

The introduction of an energy scale separating two different regimes has led to the development of Effective Field Theory (EFT). EFT is a framework used to describe processes below some energy scale, where the dynamics of fields can be neglected. In particular, the propagation of heavy mediators is ignored, and contact interactions between particles are assumed to occur. This idea is illustrated in Figure 1.4. The propagation of a heavy mediator, of mass M , is neglected compared to its mass ($p^2 \ll M^2$), and is replaced by a four-point interaction vertex with a single coupling strength G .

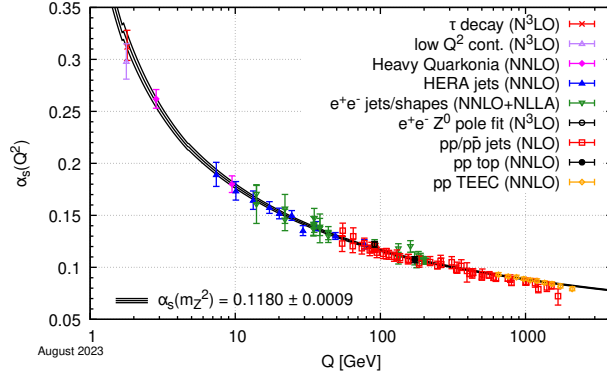


Figure 1.3: Summary of measurements of α_s as a function of the energy scale Q [3].

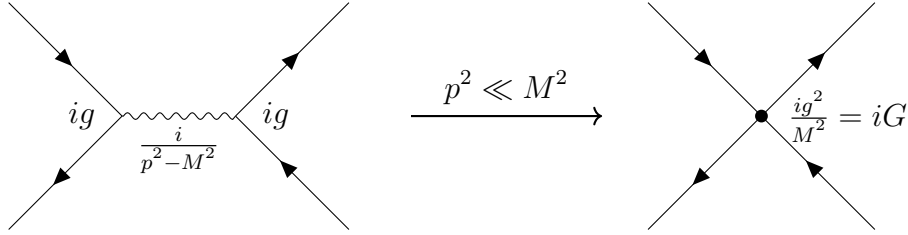


Figure 1.4: Illustration of an effective field theory, where a heavy mediator is replaced by a point-like interaction at low energy.

EFT has been successfully used in Fermi's theory of beta decay. At low energy scales, where the mass of the W boson is largely dominant, the interaction is well described by the four-fermion contact interaction:

$$\mathcal{L}_{\text{Fermi}} = -\frac{G_F}{\sqrt{2}}(\bar{\psi}_e \gamma^\mu \psi_\nu)(\bar{\psi}_p \gamma_\mu \psi_n), \quad (1.28)$$

where G_F is the Fermi constant. Historically, this effective Lagrangian successfully described neutron beta decay, long before the discovery of massive weak interaction mediators.

1.2 The Standard Model of particle physics

The Standard Model (SM) of particle physics¹ is the theoretical framework describing particles and their interactions, except for gravity. It is a quantum field theory in which particles are interpreted as local excitations of quantum fields.

Although general relativity is not part of it, both frameworks are compatible at energies much lower than the Planck scale, beyond which a theory of quantum gravity becomes necessary to describe physical phenomena.

The three fundamental interactions included in the SM—electromagnetic (EM), weak, and strong—arise from local symmetries of non-Abelian gauge groups. The complete gauge group of the SM is:

$$SU(3)_C \times SU(2)_L \times U(1)_Y, \quad (1.29)$$

¹This section is mostly based on my readings of Ref. [1], Ref. [4], and other sources.

where $SU(3)_C$ corresponds to the strong interaction, described by quantum chromodynamics (QCD), and $SU(2)_L \times U(1)_Y$ to the electroweak (EW) interaction, from which the weak and EM interactions emerge through electroweak symmetry breaking, explained in Section 1.2.3.

1.2.1 Symmetries

Instead of starting with the SM Lagrangian, it is particularly useful to study the symmetries of the theory, which play a central role by constraining the dynamics of the fields. A symmetry is a transformation under which the laws of physics remain invariant.

Spacetime symmetries In particle physics, spacetime is described by Minkowski space, $\mathbb{R}^{1,3}$, equipped with the metric $\eta_{\mu\nu} = \text{diag}(+1, -1, -1, -1)$. Spacetime symmetries include spacetime translations, spatial rotations, and Lorentz boosts, forming the 10-dimensional Poincaré group introduced in the previous section. These symmetries constrain the allowed types of fields, including scalars (spin-0), fermions (spin- $\frac{1}{2}$), and gauge fields (spin-1).

Gauge symmetries Also referred to as “gauge redundancies”, these constrain the interactions of the fields. A simple example of such a gauge redundancy arises in electromagnetism: the field strength $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is invariant under a transformation of the gauge potential of the form $A_\mu(x) \rightarrow A_\mu(x) + \partial_\mu \alpha(x)$, for any function $\alpha(x)$. This transformation does not correspond to two physically distinct situations: it represents the same physical configuration, much like a system described using two different coordinate systems. There is a *redundancy* in the description, and a specific choice must be made.

C, P, and T symmetries Discrete symmetries also play a fundamental role in the SM theory. Parity corresponds to a spatial inversion:

$$P : (t, \vec{x}) \mapsto (t, -\vec{x}). \quad (1.30)$$

Charge conjugation exchanges a particle with its antiparticle, replacing all of its charges with opposite ones:

$$C : Q_i \mapsto -Q_i. \quad (1.31)$$

Time reversal reverses the direction of time:

$$T : (t, \vec{x}) \mapsto (-t, \vec{x}). \quad (1.32)$$

Individually, each of these symmetries is violated in the Standard Model by the weak interaction, as is the combined CP symmetry. Nevertheless, the CPT theorem states that every quantum field theory must be invariant under the combined CPT symmetry. So far, no violation of CPT has been observed experimentally.

Global symmetries Global symmetries include baryon and lepton number conservation, as well as various approximate flavour symmetries.

Symmetry breaking Another important feature of the symmetries of a theory concerns how they can be *broken*. A broken symmetry can have two meanings:

- *Explicit symmetry breaking* occurs when a term in the Lagrangian does not respect the symmetry. Strictly speaking, the theory is not invariant under the symmetry, but if the corresponding term is small, or if the symmetry is only rarely broken, the symmetry can still provide a good approximation to describe phenomena.
- *Spontaneous symmetry breaking*, on the other hand, refers to a theory that is invariant under a symmetry, but whose ground state is not. Even if the field equations are invariant under the symmetry, small variations around the ground state do not reveal the underlying symmetry. This phenomenon has deep consequences, such as the appearance of Nambu–Goldstone bosons in the case of the spontaneous breaking of a continuous symmetry, or the generation of mass for the fundamental particles of the SM through the Higgs mechanism, explained in more detail in Section 1.2.3.

1.2.2 Structure of the Standard Model

As mentioned above, the SM theory is based on the local gauge symmetry group:

$$SU(3)_C \times SU(2)_L \times U(1)_Y, \quad (1.33)$$

which unifies the strong interaction ($SU(3)_C$) and the electroweak interaction ($SU(2)_L \times U(1)_Y$) within a common framework.

The fundamental fields in the SM are:

- *Gauge boson fields*: eight gluon fields G_μ^a ($1 \leq a \leq 8$) associated with $SU(3)_C$; three weak gauge bosons W_μ^i ($1 \leq i \leq 3$) associated with $SU(2)_L$; and a single gauge boson B_μ associated with $U(1)_Y$.
- *Fermion fields* ψ , which include quarks and leptons, separated into three generations, along with their antiparticles.
- *The Higgs field* H , a complex scalar doublet, which gives mass to the weak bosons through the Brout–Englert–Higgs (BEH) mechanism, and to the fermions via Yukawa interactions.

A conserved quantity is associated with each gauge group: colour charge with $SU(3)_C$, weak isospin T (whose third component is denoted T_3) with $SU(2)_L$, and weak hypercharge Y with $U(1)_Y$.

In the SM, fermions come in two chiralities: left and right. Left-handed fermions (and right-handed anti-fermions) transform as doublets under $SU(2)_L$, while right-handed fermions (and left-handed anti-fermions) transform as singlets under this gauge group and are thus not sensitive to the weak force. $SU(3)_C$, for its part, classifies fermions into quarks and leptons, depending on whether they experience the strong force: quarks, which carry a colour charge, transform as triplets under this group, while leptons, which are colour-neutral, transform as singlets. Finally, $U(1)_Y$ acts on each fermion carrying a non-zero hypercharge, regardless of their chirality.

Quarks and leptons come in three generations, with identical charges but different masses arising from their couplings to the Higgs field. The left-handed quark doublets Q_L^i have a weak hypercharge $Y = +\frac{1}{3}$ and are denoted, for the three generations:

$$\begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix}. \quad (1.34)$$

The different flavours are called up, down, charm, strange, top, and bottom. The left-handed lepton doublets L_L^i , which carry $Y = -1$, are denoted:

$$\begin{pmatrix} \nu_{e_L} \\ e_L \end{pmatrix}, \begin{pmatrix} \nu_{\mu_L} \\ \mu_L \end{pmatrix}, \begin{pmatrix} \nu_{\tau_L} \\ \tau_L \end{pmatrix}. \quad (1.35)$$

Each lepton flavour (electron, muon, and tau) comes with a corresponding neutrino.

In these doublets, $u_L, c_L, t_L, \nu_{e_L}, \nu_{\mu_L}$, and ν_{τ_L} carry weak isospin $T_3 = +\frac{1}{2}$ while $d_L, s_L, b_L, e_L, \mu_L$, and τ_L carry $T_3 = -\frac{1}{2}$.

The right-handed quark singlets are denoted similarly: u_R, c_R , and t_R (with $Y = +\frac{4}{3}$), and d_R, s_R , and b_R (with $Y = -\frac{2}{3}$). The right-handed lepton singlets are e_R, μ_R , and τ_R (with $Y = -2$), and ν_{e_R}, ν_{μ_R} , and ν_{τ_R} (with $Y = 0$), if such right-handed neutrinos exist. All right-handed particles carry $T_3 = 0$.

Since flavour is not conserved by the weak interaction, it is important to distinguish between mass eigenstates, which correspond to free states, and interaction eigenstates, which take part in interactions. For quarks, the strength of flavour-changing in weak interactions is given by the CKM (Cabibbo–Kobayashi–Maskawa) matrix. For neutrinos, this information is encoded in the PMNS (Pontecorvo–Maki–Nakagawa–Sakata) matrix and describes the phenomenon of neutrino oscillation.

1.2.3 Electroweak symmetry breaking

In the SM, gauge boson and fermion fields should be massless. However, this does not correspond to what is observed in nature. Fortunately, the Higgs boson, by breaking the electroweak symmetry, gives mass to the weak gauge bosons.

The Higgs field H is a complex scalar field transforming as a doublet under $SU(2)_L$:

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \quad (1.36)$$

It carries a weak isospin $T = \frac{1}{2}$ (with third component $T_3 = -\frac{1}{2}$) and weak hypercharge $Y = +1$.

Through electroweak symmetry breaking, the Higgs field acquires a non-zero vacuum expectation value (VEV):

$$\langle H \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \quad (1.37)$$

with v fixed by the Fermi constant: $v = 1/\sqrt{\sqrt{2}G_F}$ ($v \approx 246$ GeV). This VEV causes the spontaneous breaking of $SU(2)_L \times U(1)_Y$ down to the electromagnetic group $U(1)_{\text{EM}}$, with the electric charge being the conserved quantity, defined as:

$$Q = T_3 + \frac{1}{2}Y. \quad (1.38)$$

From the Goldstone theorem, if a global symmetry is broken, massless bosons appear. However, in the case of gauge symmetries, as for $SU(2)_L \times U(1)_Y$, the extra degrees of freedom become the longitudinal polarisations of the existing bosons. The electroweak bosons mix and give rise to the three massive bosons of the weak interaction:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \quad (1.39)$$

$$Z_\mu = \cos \theta_W W_\mu^3 - \sin \theta_W B_\mu, \quad (1.40)$$

and the massless photon:

$$A_\mu = \sin \theta_W W_\mu^3 + \cos \theta_W B_\mu. \quad (1.41)$$

Here, θ_W is the Weinberg angle, defined by the electroweak gauge couplings g and g' : $\tan \theta_W = g'/g$, where g and g' are the coupling constants for the weak and hypercharge interactions, respectively. This mechanism is known as the Brout–Englert–Higgs mechanism and led to the discovery of the Higgs boson.

The fourth degree of freedom of the Higgs field gives rise to a new particle: the Higgs boson.

Finally, fermions acquire mass via the Higgs field through a different mechanism, known as Yukawa interaction.

Figure 1.5 shows the fundamental particles of the Standard Model induced by the electroweak symmetry breaking, along with their quantum numbers and the corresponding interactions.

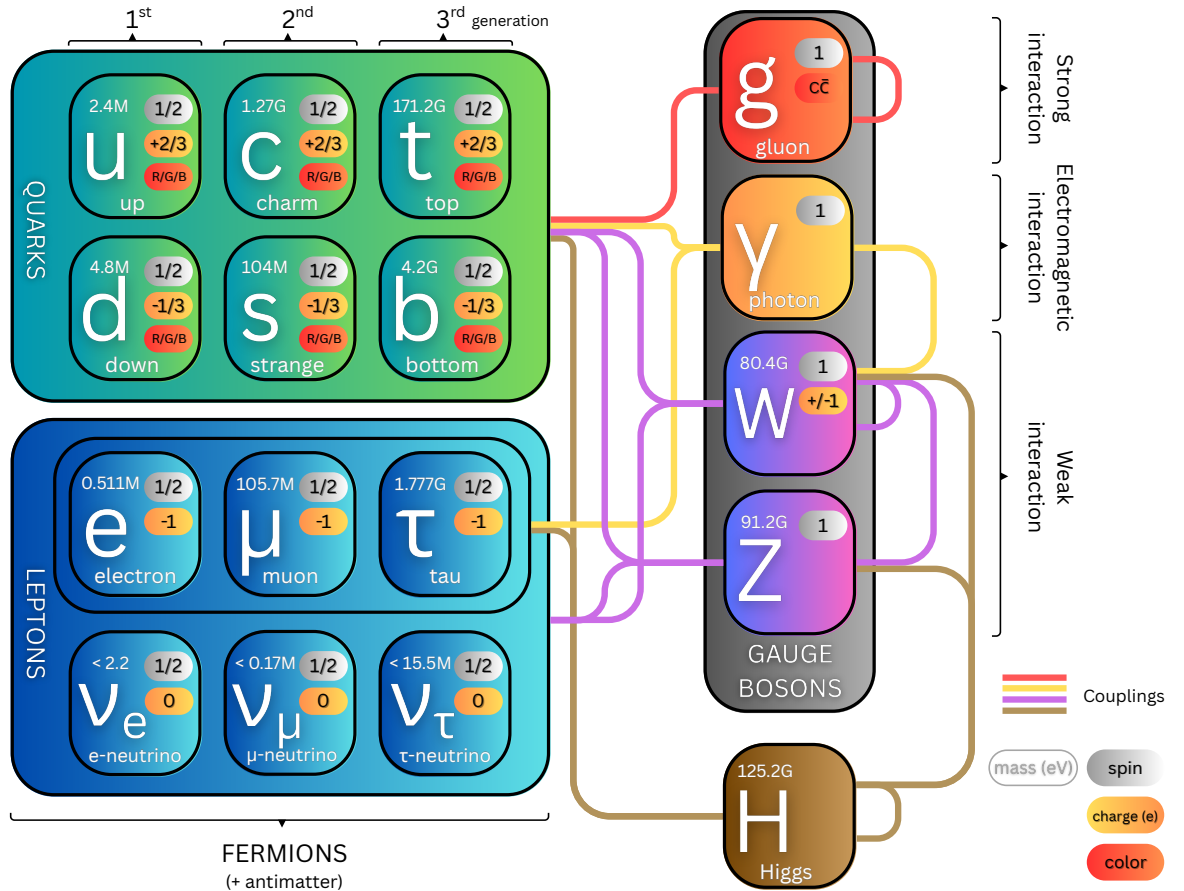


Figure 1.5: Fundamental particles of the Standard Model.

1.2.4 The Standard Model Lagrangian

As suggested previously, to construct the SM, it is more appropriate to begin by describing the symmetries of the theory, which constrain the Lagrangian and allow it to be determined almost completely. The full SM Lagrangian, in a summarised form, is given by:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{fermion}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}}. \quad (1.42)$$

The first two terms are the kinetic terms for the gauge and fermion fields. The last two terms describe the Higgs dynamics and its interaction with fermions.

Gauge sector The SM gauge fields correspond to the symmetry groups $SU(3)_C$, $SU(2)_L$, and $U(1)_Y$, with the associated field strengths $G_{\mu\nu}^a$, $W_{\mu\nu}^i$, and $B_{\mu\nu}$, respectively, where a runs over the eight generators of $SU(3)_C$ and i over the three generators of $SU(2)_L$. The gauge part of the Lagrangian takes the form:

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W_{\mu\nu}^iW^{i,\mu\nu} - \frac{1}{4}G_{\mu\nu}^aG^{a,\mu\nu}. \quad (1.43)$$

Fermionic sector The fermionic part is:

$$\begin{aligned} \mathcal{L}_{\text{fermion}} = -i \sum_{i=1}^3 & \left(\bar{Q}_L^i \bar{\sigma}^\mu \mathcal{D}_\mu Q_L^i + \bar{L}_L^i \bar{\sigma}^\mu \mathcal{D}_\mu L_L^i + \bar{u}_R^i \sigma^\mu \mathcal{D}_\mu u_R^i \right. \\ & \left. + \bar{d}_R^i \sigma^\mu \mathcal{D}_\mu d_R^i + \bar{e}_R^i \sigma^\mu \mathcal{D}_\mu e_R^i \right), \end{aligned} \quad (1.44)$$

where $\sigma^\mu = (\mathbb{1}, \sigma^i)$ and $\bar{\sigma}^\mu = (\mathbb{1}, -\sigma^i)$. The σ^i ($i = 1, 2, 3$) are the Pauli matrices, defined as:

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (1.45)$$

It includes the left- and right-handed quarks and leptons, along with their antiparticles, and the sum runs over the three generations. The covariant derivatives encode the interactions with the gauge sector and depend on the field quantum numbers. It can be expressed as:

$$\mathcal{D}_\mu = \partial_\mu - ig_S \frac{\lambda^a}{2} G_\mu^a - ig \frac{\sigma^j}{2} W_\mu^j - ig' \frac{Y}{2} B_\mu, \quad (1.46)$$

where g_S is the coupling constant for the strong interaction, λ^a ($a = 1, \dots, 8$) are the Gell-Mann matrices, and σ^j ($j = 1, \dots, 3$) the Pauli matrices.

Higgs sector The Higgs part consists of a kinetic and a potential term:

$$\mathcal{L}_{\text{Higgs}} = (\mathcal{D}_\mu H)^\dagger \mathcal{D}^\mu H - \lambda \left(H^\dagger H - \frac{v^2}{2} \right)^2, \quad (1.47)$$

where v is the Higgs VEV and λ is a positive real parameter. The covariant derivative is given by:

$$\mathcal{D}_\mu H = \partial_\mu H - ig W_\mu H - \frac{i}{2} g' B_\mu H. \quad (1.48)$$

Yukawa sector Fermions acquire mass through a Yukawa coupling to the Higgs field. The Lagrangian of the Yukawa sector is given by:

$$\mathcal{L}_{\text{Yukawa}} = -y_{ij}^d \bar{Q}_L^i H d_R^j - y_{ij}^u \bar{Q}_L^i \tilde{H} u_R^j - y_{ij}^e \bar{L}_L^i H e_R^j - y_{ij}^\nu \bar{L}_L^i \tilde{H} \nu_R^j + \text{h.c.}, \quad (1.49)$$

where $\tilde{H} = i\sigma^2 H^*$ (with σ^2 being the second Pauli matrix) is the charge-conjugated Higgs doublet, and the y^u , y^d , y^e , and y^ν are complex 3×3 coupling matrices, giving a different mass to each fermion. These matrices allow for mixing between generations, leading to flavour-changing phenomena described by the CKM and PMNS matrices induced by the electroweak symmetry breaking.

1.3 Photon physics at hadron colliders

1.3.1 Hadron interactions

The parton model, introduced by Richard Feynman in the 1960s [5] and later applied to electron–proton deep inelastic scattering by James Bjorken and Emmanuel Paschos [6], is a theoretical framework that describes the behaviour of hadron constituents (quarks and gluons), known as *partons*. It relies on the concept of asymptotic freedom, introduced in Section 1.1.4, and portrays partons as being approximately free inside hadrons.

Partons are characterised by functions $f(x, Q^2)$, which represent the probability density (or rather the number density, since they are normalised to the number of partons) of finding a parton carrying a fraction x of the hadron’s total momentum, for a given momentum transfer Q^2 . These functions are known as *parton distribution functions* (PDFs). They are typically extracted through global fits to data from deep inelastic scattering (DIS) [7, 8] and other hard processes, and their evolution with the energy scale Q^2 is governed by the DGLAP (Dokshitzer–Gribov–Lipatov–Altarelli–Parisi) equations [9–11].

The naive picture of a hadron composed solely of its two or three valence quarks (those that determine its quantum numbers) is incomplete, as it neglects the “sea” of virtual quark–antiquark pairs generated by the strong interaction. For example, in a proton, the total momentum is shared not only among the three valence quarks (two up and one down) but also among gluons and sea quarks.

In a hard scattering process, hadrons do not interact with each other as if they were fundamental particles: only their constituents, the partons, participate in the interactions.

In hadron colliders, two regimes can be distinguished. On one hand, hard processes, involving large momentum transfer, such as high- p_T jet production or Higgs boson production, can be computed perturbatively. On the other hand, soft processes, occurring at long-distance scales, such as underlying events, are less well understood and require phenomenological models for their description.

In this context, the factorisation theorem [13, 14] plays a central role in QCD calculations. It allows the separation of short-distance processes, which can be computed perturbatively, from long-distance non-perturbative effects in the cross section calculation. For a process involving two hadrons, A and B , producing a final state X , the total cross section is expressed as:

$$\sigma_{AB \rightarrow X} = \int dx_a dx_b f_{a/A}(x_a, \mu_F^2) f_{b/B}(x_b, \mu_F^2) [\hat{\sigma}_0 + \alpha_s(\mu_R^2) \hat{\sigma}_1 + \dots]_{ab \rightarrow X}, \quad (1.50)$$

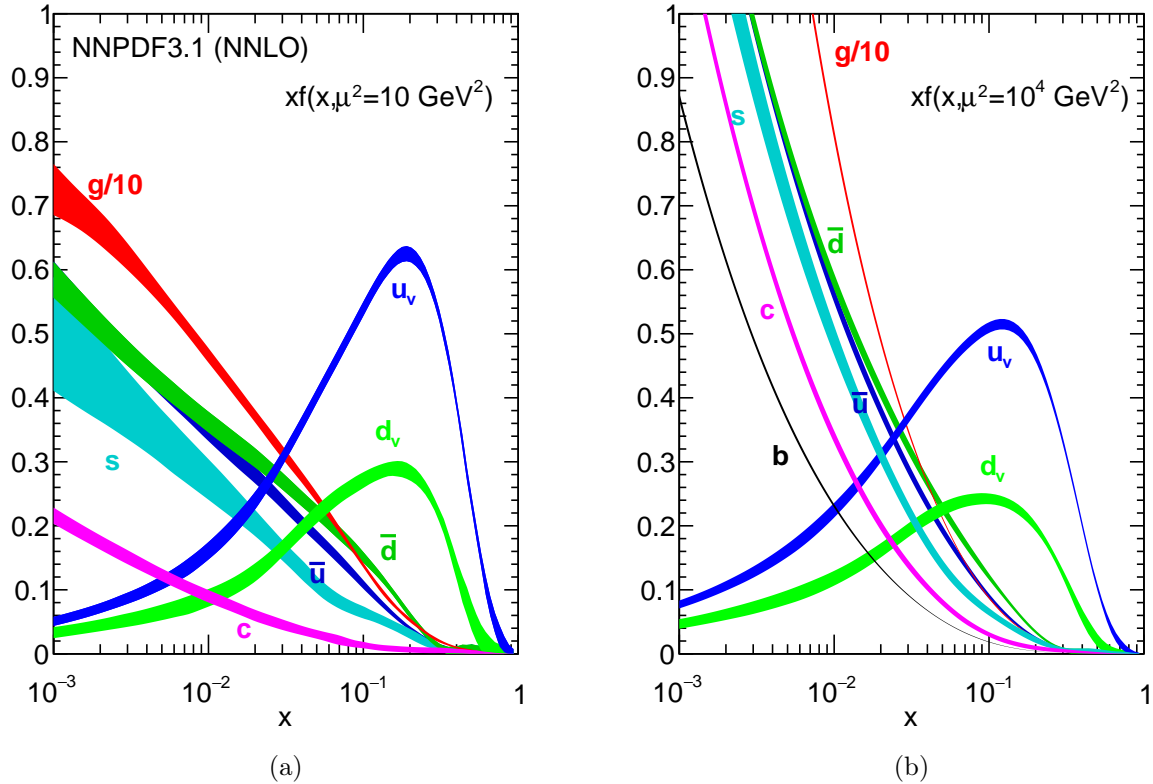


Figure 1.6: $xf(x, \mu^2)$ as a function of x , the fraction of momentum carried by the parton, with f the parton distribution functions (PDFs) of the proton obtained in the NNLO NNPDF3.1 global analysis, for two energy scales, (a) $\mu_F^2 = 10 \text{ GeV}^2$ and (b) $\mu_F^2 = 10^4 \text{ GeV}^2$ [12]. The valence quarks are denoted with the letter v . The sea quarks are also represented.

where x_a and x_b are the momentum fractions carried by partons a and b inside hadrons A and B . The $\hat{\sigma}_n$ terms in the second part of the expression denote the perturbative expansion of the hard scattering cross section. μ_R is the renormalisation scale introduced earlier, which governs the running of the strong coupling constant α_s . Another parameter, the factorisation scale μ_F , is introduced to separate the short- and long-distance regimes. Both parameters are unphysical and are typically chosen to be of the order of the characteristic energy scale of the process, often with $\mu_F = \mu_R$. Figure 1.6 shows the PDFs of the proton at next-to-next-to-leading order (NNLO) in α_s from the NNPDF analysis, evaluated at two energy scales: $\mu_F^2 = 10 \text{ GeV}^2$ and $\mu_F^2 = 10^4 \text{ GeV}^2$ [12].

When a parton is produced in a high-energy collision, it cannot exist freely due to colour confinement. As the characteristic energy scale of the parton decreases, the strong coupling constant α_s increases, raising the probability of gluon radiation. The parton subsequently emits gluons, which can produce quark–antiquark pairs, leading to a cascade that forms a shower of hadronic particles. These processes, known as *hadronisation* and *parton showers*, result in a collimated bunch of colourless particles referred to as a *jet*.

While the production of coloured partons can be described using perturbative QCD, hadronisation lies within the non-perturbative regime. The transition from partons to hadrons is modelled using *fragmentation functions*, $D_{k \rightarrow C}(z, \mu_f^2)$, which encode the prob-

ability that a parton k , carrying a fraction z of the hadron's momentum, forms a hadron C . In analogy to the factorisation scale used in PDFs, a new unphysical energy parameter, the *fragmentation scale* μ_f , is introduced to separate the two regimes.

The cross section for two colliding hadrons A and B to produce a hadron C , along with any other final state X , is then given by:

$$\sigma_{AB \rightarrow C+X} = \int dz_C D_{k \rightarrow C}(z_C, \mu_f^2) \sigma_{AB \rightarrow k+X}(\mu_R^2, \mu_F^2), \quad (1.51)$$

where $D_{k \rightarrow C}$ is the fragmentation function describing the probability for a parton k to form the hadron C , z_C is the fraction of the parton's momentum transferred to the hadron C , and $\sigma_{AB \rightarrow k+X}$ is the cross section for producing the parton k , as introduced in Eq. 1.50.

As with the PDFs, the energy-scale dependence of the fragmentation functions is governed by the DGLAP equations. The three energy scales— μ_R , μ_F , and μ_f —are typically chosen to be equal to the characteristic momentum transfer of the hard process, $\sqrt{Q^2}$.

1.3.2 Prompt photon production at the LHC

This thesis exclusively uses data from proton–proton (pp) collisions, with a particular focus on *prompt photons*. Prompt photons are high-transverse-momentum photons that do not originate from hadron decays, such as those of π^0 or η mesons. They are of primary interest as they provide a clean probe of short-distance QCD processes [15].

At leading order (LO), prompt photons are typically classified into two categories:

- **Direct photons**, produced in the hard scattering of two partons, e.g. $qg \rightarrow q\gamma$ and $q\bar{q} \rightarrow g\gamma$;
- **Fragmentation photons**, radiated by a parton produced in the hard scattering. These are also referred to as Bremsstrahlung photons.

These different processes are illustrated in Figure 1.7.

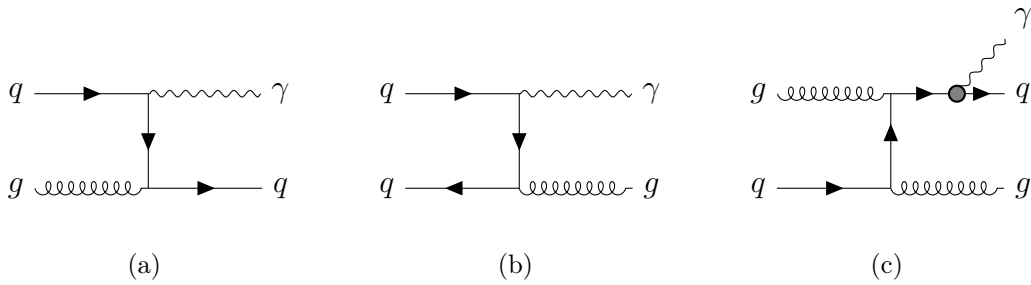


Figure 1.7: Prompt photon production at leading order in pp collisions. (a) Quark-gluon Compton scattering (direct). (b) Quark-antiquark annihilation (direct). (c) Bremsstrahlung photon radiated by a final state quark (fragmentation).

At next-to-leading order (NLO), the distinction between direct and fragmentation photons becomes ambiguous due to the presence of loop corrections involving gluons and photons, as well as interferences between different production mechanisms.

Photons originating from the decay of heavy particles, such as the Higgs boson, are also classified as prompt photons, since they do not result from hadron decays.

1.3.3 Photon isolation

Prompt photons serve as an important probe of QCD [16] and provide a clean signature for a wide range of analyses, such as measurements of Higgs boson mass and couplings [17]. It is therefore essential to identify them among the large number of non-prompt photons arising from hadron decays. A significant portion of these non-prompt photons is produced in highly collimated pairs from π^0 decays. Due to the limited resolution of detectors, such pairs can be misidentified as single prompt photons.

One way to distinguish them is by examining the energy flux around the photon candidates, with a prompt photon typically being isolated from the hadronic energy flow. A photon is considered isolated if the hadronic energy flow within a cone centred on the photon candidate falls below a given threshold. While this criterion can be implemented experimentally without major difficulties, it presents theoretical challenges: the fixed-cone isolation imposes constraints on the phase space for gluon emissions, which can prevent the cancellation of infrared divergences in higher-order QCD calculations.

This issue can be avoided using Frixione isolation [18], in which the hadronic energy threshold increases smoothly with the angular distance from the photon. However, this criterion cannot be easily applied experimentally due to the finite angular and energy resolutions of calorimeters.

Other criteria, both theoretically and experimentally consistent, are under development, such as “soft drop isolation” [19] and “democratic isolation” [20]. At the LHC, analyses involving photons use a fixed-size isolation cone, while NNLO QCD computations employ the theoretically smooth-cone isolation [21].

1.3.4 Non-resonant diphoton backgrounds at the LHC

Non-resonant photon pair (*diphoton*) production constitutes the main background in searches for new resonances decaying into two photons, such as the analysis presented in this thesis. All prompt and non-prompt photon sources must be taken into account.

The main processes contributing to prompt diphoton production are the following:

- Born (or tree-level) production via quark-antiquark annihilation: $q\bar{q} \rightarrow \gamma\gamma$
- Box diagram production from gluon-gluon fusion through a quark loop: $gg \rightarrow \gamma\gamma$. Although this process occurs at a higher order, $\mathcal{O}(\alpha^2\alpha_s^2)$, its contribution remains significant due to the large gluon luminosity at the LHC.
- Radiative processes, such as $gq \rightarrow q\gamma\gamma$, where one photon is radiated from the final-state quark.

These sources are often referred to as *irreducible* backgrounds, as they cannot be distinguished on an event-by-event basis from diphoton resonances. Only statistical analyses can separate their contribution from that of a potential diphoton signal.

Conversely, fake diphotons with a diphoton-like signature, arising from hadron decays such as those of the π^0 inside jets, form the *reducible* background, since any improvement in the identification and isolation tools can reduce their contribution.

Dijet events contribute much more to the total background than diphoton events due to their much higher cross section ($\sigma(pp \rightarrow jj) = 3.2 \times 10^5$ pb, for $\sqrt{s} = 13$ TeV and anti- k_t jets [22] with $R = 0.4$ [23]) compared to that of diphotons ($\sigma(pp \rightarrow \gamma\gamma) = 31$ pb,

for $\sqrt{s} = 13 \text{ TeV}$ and two isolated photons with $p_T^{\gamma_1} > 40 \text{ GeV}$ and $p_T^{\gamma_2} > 30 \text{ GeV}$ [24]). This highlights the importance of developing efficient tools to reject fake diphotons.

The main processes contributing to the diphoton background are illustrated in Figure 1.8.

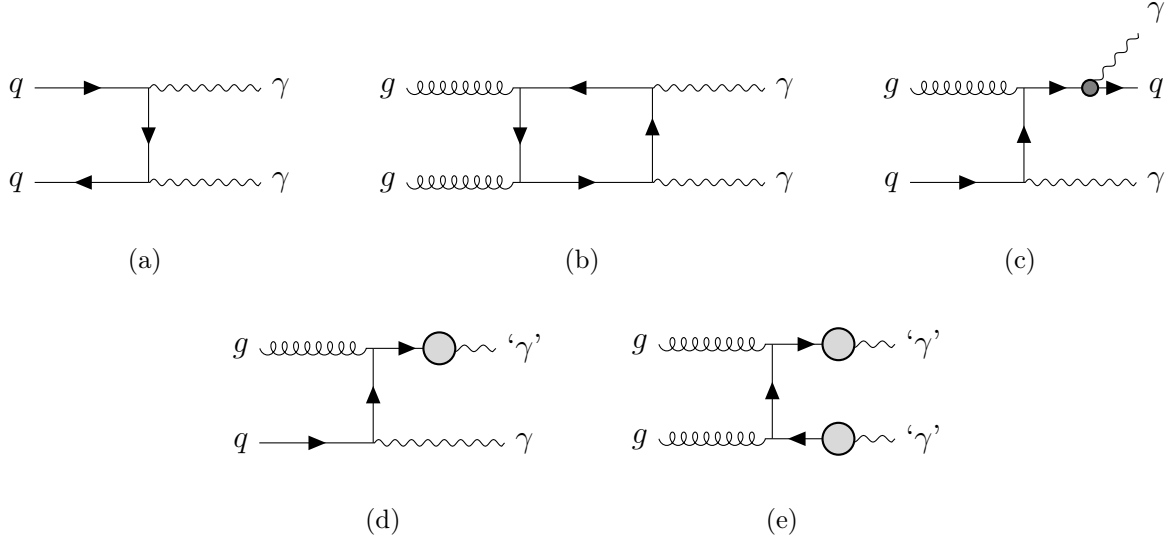


Figure 1.8: Diagrams of main processes of diphoton production at the LHC. The upper diagrams depict prompt diphoton processes: (a) Born, (b) box and (c) Bremsstrahlung processes. The bottom row ((d) and (e)) represents jet and dijet fragmentation with a diphoton-like final state. The ‘ γ ’ denotes a final state with a photon-like signature, such as the decay of a π^0 into two collimated photons. Hadron decays are represented by large grey circles.

1.4 Beyond the Standard Model

1.4.1 Limitations of the Standard Model

Despite its great success in describing and predicting phenomena, the Standard Model faces several difficulties, making it necessary to improve the theory. Some of these inconsistencies are briefly presented below.

Matter–antimatter asymmetry Matter appears to dominate the Universe (at least in our observable region). However, although the weak interaction violates CP symmetry, this violation is not sufficient to explain the observed asymmetry [25].

Higgs mass The Higgs boson mass arises from two contributions: its bare mass and quantum corrections. These radiative corrections are quadratically sensitive to the theory’s ultraviolet cut-off scale. If the Standard Model is assumed to be valid up to the Planck scale ($\Lambda \sim 10^{19} \text{ GeV}$), the corrections to the Higgs mass squared are of the order of $\delta m_H^2 \sim 10^{36} \text{ GeV}^2$, far exceeding the observed Higgs mass squared ($\sim 10^4 \text{ GeV}^2$) [26].

This implies that the bare mass must be huge and finely tuned to nearly cancel the quantum corrections — a situation known as the fine-tuning problem. This issue is referred to as the naturalness problem or the hierarchy problem.

Neutrino mass In its initial minimal formulation, the Standard Model includes only left-handed neutrinos, but not right-handed ones, preventing the construction of a mass term. However, neutrino oscillations, the phenomenon by which a neutrino changes flavour as it propagates, cannot occur if neutrinos are massless.

Dark matter Numerous cosmological and astrophysical observations suggest the existence of large amounts of invisible, massive matter, known as *dark matter*. It does not interact with light or with electromagnetic forces. Its presence is inferred through its gravitational effects on visible matter and the large-scale structure of the Universe. For instance, the rotation curves of spiral galaxies remain flat at large radii [27], which contradicts Kepler’s laws unless additional unseen mass is present. Similarly, the velocity dispersions observed in elliptical galaxies are inconsistent with predictions from the virial theorem unless dark matter is included. In galaxy clusters, independent mass estimates based on the radial velocity distributions of member galaxies, the X-ray emission from the intracluster medium, and gravitational lensing [28] all suggest that the total mass is far greater than the visible mass. The Standard Model offers no suitable candidate for dark matter. A dark matter candidate must be electrically neutral, stable, and sufficiently cold (i.e., its velocity dispersion should be negligible compared to that of galaxies at the epoch of their formation) to match observations. Beyond-the-Standard-Model theories propose various candidates, including Weakly Interacting Massive Particles (WIMPs), axions, and sterile neutrinos.

Dark energy Within the Λ CDM model, the accelerated expansion of the Universe (first revealed by Type Ia supernovae [29–31]) is explained by the cosmological constant Λ . The origin of this constant is unknown, and dark energy has been proposed to account for its appearance in Einstein’s field equations. Although its energy density is extremely low (on the order of $10^{-29} \text{ g cm}^{-3}$), it constitutes nearly 68% of the Universe. To drive accelerated expansion, dark energy must have negative pressure, which is incompatible with both ordinary and dark matter. Alternative explanations include dynamic scalar fields such as quintessence, but no evidence has yet been found for any specific model.

Strong CP problem Whereas the weak interaction violates CP symmetry, the strong interaction appears to preserve it, despite no known fundamental reason for doing so. CP violation in the strong interaction is governed by a parameter called θ_{QCD} , and any value different from zero would imply CP violation. However, experimental measurements have found it to be consistent with zero, with an upper bound of $\theta_{\text{QCD}} < 10^{-9}$. This extreme fine-tuning is known as the strong CP problem. A proposed solution, put forward by Peccei and Quinn [32], introduces a new global symmetry that is spontaneously broken at low energy, giving rise to a new pseudo-scalar particle called the *axion*. In this framework, axion dynamics drive θ_{QCD} to zero, thus restoring CP symmetry in the strong interaction. Moreover, axions are potential candidates for dark matter: they are neutral, very weakly coupled to ordinary matter, and, for large symmetry-breaking scales, they are very long-lived [33].

In addition, the Standard Model contains a large number of free parameters (19), which are not predicted and must be determined experimentally. The absence of gravity in its theoretical framework also motivates the search for new theories beyond the Standard Model (BSM).

1.4.2 Physics beyond the Standard Model

A wide range of theoretical models has been proposed to overcome the limitations of the Standard Model. Among them, supersymmetry (SUSY) is one of the most extensively studied. It postulates a fundamental link between bosons and fermions, associating with each a superpartner whose spin differs by half a unit. If these new particles have masses not too different from their Standard Model counterparts, they help reduce the fine-tuning required in quantum corrections.

Other theoretical approaches attempt to incorporate gravity into the quantum framework. Among the most prominent, string theory describes particles as one-dimensional strings, including the hypothetical graviton, whose vibrational modes determine their properties. In parallel, loop quantum gravity quantises spacetime itself, describing it as a network of discrete loops.

All these approaches face both theoretical and experimental challenges, and none has, so far, been experimentally confirmed.

1.5 Axion-like particles

A new diphoton resonance — the focus of the extended search proposed in this thesis — can be interpreted within the framework of axion-like particles (ALPs). An ALP is a spin-0, electrically neutral pseudo-Nambu–Goldstone boson (pNGB) [34], arising from the spontaneous breaking of an approximate accidental global $U(1)$ symmetry, often emerging at low energies as a remnant of exact discrete symmetries in more fundamental theories [33]. Its mass m_a can naturally be smaller than the Higgs boson mass without requiring fine-tuning.

Unlike the QCD axion, ALPs do not necessarily solve the strong CP problem. Their masses and couplings to other Standard Model particles are free parameters of the theory and can span several orders of magnitude. If sufficiently stable, they can serve as non-thermal dark matter candidates [35], i.e., whose present-day abundance was not set by thermal equilibrium in the early Universe. Otherwise, they could act as mediators between the Standard Model and a hidden dark sector.

ALPs are gauge singlets under the Standard Model. Their behaviour is well described within the effective field theory framework, introduced in Section 1.1.5. The energy scale associated with the spontaneous symmetry breaking can be expressed as

$$\Lambda \sim g_\star f_a, \quad (1.52)$$

with f_a the decay constant governing the strength of the ALP coupling to Standard Model fields, and g_\star a dimensionless constant relating f_a to the symmetry-breaking energy scale.

In the KSVZ model [36, 37], the ALP couples only to gluons and photons. At LO in the expansion in $1/f_a$, the effective Lagrangian describing the ALP interactions with the Standard Model can be written as [38]:

$$\mathcal{L}_{\text{ALP}} = \frac{1}{2}m_a^2 a^2 + \frac{a}{4\pi f_a} \left[\alpha_3 c_3 G\tilde{G} + \alpha_2 c_2 W\tilde{W} + \alpha_1 c_1 B\tilde{B} \right] \quad (1.53)$$

where a denotes the ALP field and m_a its mass. G , W , and B are the QCD and electroweak field strengths, with α_3 , α_2 , and α_1 the corresponding coupling constants. $\tilde{G}^{\mu\nu}$, $\tilde{W}^{\mu\nu}$, and $\tilde{B}^{\mu\nu}$ are the dual field-strength tensors, defined as $\tilde{G}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\alpha\beta}G_{\alpha\beta}$ (with $\epsilon^{0123} = 1$). The anomalies of the global symmetry, non-linearly realised by the ALP, are encoded in the coefficients c_3 , c_2 , and c_1 . The effective theory used here differs slightly from that introduced in Section 1.1.5, since the additional field has a low mass.

1.5.1 Resonant axion-like particle phenomenology at hadron colliders

Numerous theoretical models predict the existence of ALPs, spanning a wide range of masses m_a and decay constants f_a , some of which are suitable candidates for diphoton resonance searches at hadron colliders. Among them, heavy QCD axion models [39–41] solve the strong CP problem by following the original Peccei–Quinn mechanism, with additional features to make the axion heavier and accessible at hadron colliders. Other examples include the R-axion [42–44], which emerges from the spontaneous breaking of supersymmetry, and composite Higgs models [45, 46], which naturally predict a pNGB in their low-energy spectrum. These latter two models also propose solutions to the naturalness problem associated with the electroweak scale.

Other models propose ALPs as mediators between the Standard Model and the dark sector, allowing for the annihilation of dark matter into Standard Model particles and driving the thermal freeze-out of the dark matter abundance [47].

A common feature among all these models is the possible production of ALPs via gluon and photon fusion, making them suitable candidates for LHC searches. Both couplings scale as m_a^3/f_a^2 : for masses below the Higgs mass, the widths of the decays in dijets and diphotons are given by $\Gamma_{gg} = K_g \frac{\alpha_s^2 c_3^2}{8\pi^3} \frac{m_a^3}{f_a^2}$ and $\Gamma_{\gamma\gamma} = \frac{\alpha_{\text{em}}^2 c_\gamma^2}{64\pi^3} \frac{m_a^3}{f_a^2}$, where higher-order QCD corrections are encoded in $K_g = 2.1$, α_{em} is the fine-structure constant, $c_\gamma = c_2 + 5c_1/3$, and α_s and α_{em} are computed at the mass m_a [34]. In proton–proton collisions at the LHC, photon fusion is strongly suppressed compared to gluon fusion, making the latter the dominant production mechanism for ALPs. Figure 1.9 shows the ALP production via gluon fusion, in a pp collision, and its decay into two photons.

1.5.2 Existing constraints on axion-like particles

The ALP parameter space is constrained by a wide variety of experiments, from astrophysical to laboratory-based, depending on their mass and coupling constants.

Light ALPs with masses below twice the electron mass can only decay into photons. The lightest ALPs are primarily probed by astrophysical and cosmological observations. Among them, helioscopes [48, 49] such as CAST search for axions produced in the solar core. They aim to convert axions into X-rays through the Primakoff effect [50] using a strong magnetic field. The main advantage of this technique is that it exploits the well-known solar flux from a continuous and predictable source.

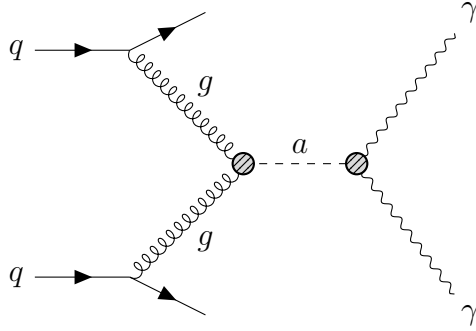


Figure 1.9: Feynman diagram of ALP production via gluon fusion and decay into photons. The dashed circles represent the effective couplings of the ALP to gluons and to photons.

Beam-dump experiments [51], such as those at SLAC, target long-lived, weakly interacting particles. They use a high-intensity beam whose passage through a thick target may produce axions, which are then detected via their decay into Standard Model particles.

Haloscope experiments, such as ADMX [52] or MADMAX [53], focus on ALPs originating from the dark matter halo. They aim to convert them into detectable microwave radiation using strong laboratory magnets.

At intermediate masses, up to the GeV scale, constraints come from lower-energy colliders such as LEP. Among others, the L3 and OPAL experiments [54–56] have searched for ALPs through the strongly suppressed $Z \rightarrow 3\gamma$ decay and other associated production modes involving an additional photon. In these multi-photon final states, any deviation from Standard Model predictions would indicate new physics.

From a few GeV up to around 100 GeV, constraints can be obtained from light-by-light scattering in ultra-peripheral heavy-ion collisions [57, 58], particularly at the LHC in the ATLAS and CMS experiments. Elastic light-by-light scattering is a process occurring at leading order in the QED coupling α , through virtual box diagrams involving charged particles. As it does not involve gluon fusion, the constraining power of this process remains limited.

From around 10 GeV up to the TeV scale, constraints are obtained from diphoton resonance searches. Dijet searches are also used, but above 65 GeV the strongest constraints come from diphoton searches [59], as they provide a cleaner signature than dijets. At the LHC, the diphoton resonance searches are divided into the high-mass range (above the Higgs mass) [60, 61], the low-mass range (from about 60 GeV up to the Higgs mass) [62, 63], and the very-low-mass range (from about 10 GeV to about 70 GeV). The lowest mass accessible with these techniques is limited by trigger thresholds and isolation requirements. Recently, the ATLAS and CMS experiments have extended the reach of diphoton resonance searches down to 10 GeV [38, 64]. The work presented in this thesis proposes tools to explore masses below 10 GeV.

Figure 1.10 shows the constraints on the ALP mass versus ALP–photon coupling parameter space, with a focus on the limits obtained from collider searches.

1.6 Monte Carlo event generators

Monte Carlo (MC) event generators provide precise simulations of high-energy particle collisions. They play a crucial role in QCD modelling, data analysis, and the design of new experiments [3]. When combined with detector simulations, they enable accurate

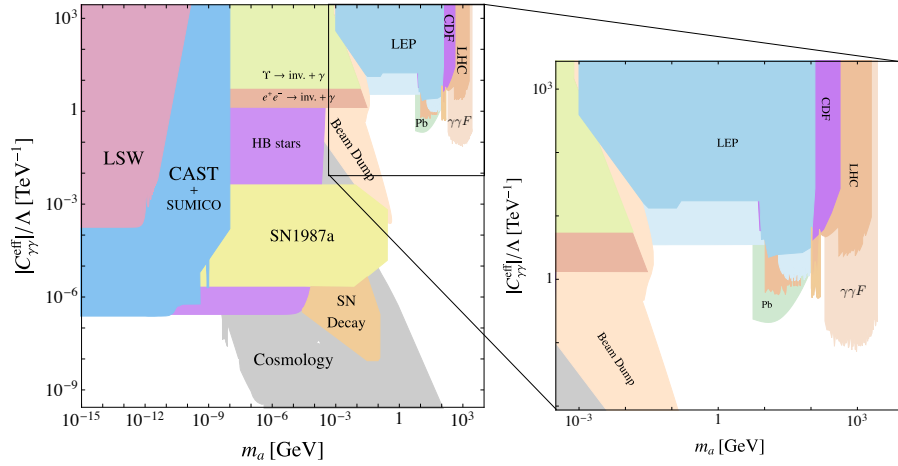


Figure 1.10: Left: constraints on the parameter space spanned by the ALP mass and ALP-photon coupling [65]. Right: enlarged display of the constraints from collider searches: LEP (light blue and blue), CDF (purple), LHC from associated production and Z decays (orange), LHC from photon fusion (light orange), and from heavy-ion collisions at the LHC (green).

estimations of both signal and background processes. The simulation proceeds in several stages, briefly summarised below [66].

Hard scattering process First, the generator computes the cross section (or matrix element) of the process with the highest momentum transfer in the event. This calculation relies on parton distribution functions to model the quarks and gluons within the incoming hadrons.

Parton shower Following the hard interaction, coloured partons radiate additional partons (primarily gluons), which in turn emit softer gluons. This shower is treated step by step, taking into account both initial- and final-state radiation.

Hadronisation As coloured partons cannot exist in isolation, the generator simulates their non-perturbative transition into colourless hadrons.

Underlying event In addition to the hard process, other softer activities may occur—such as secondary parton scatterings, beam remnants, and soft radiation. These give rise to soft hadrons that overlay the primary hard scattering and impact its reconstruction.

Decays of unstable particles Many of the particles produced during hadronisation are unstable. The generator simulates their decays into stable, observable final states.

Several MC event generators are used to model pp collisions at the LHC. Some are general-purpose tools that simulate all stages of the event, while others are specialised in specific aspects of the simulation. Typically, both types are combined to achieve a comprehensive and accurate description of the event. In this thesis, PYTHIA and SHERPA are employed. PYTHIA [67] is a leading-order event generator that simulates the full

evolution of a collision, including the hard scattering, underlying event, and parton shower. Matrix elements are calculated for $2 \rightarrow 2$ processes at LO using perturbative QCD. PYTHIA is often interfaced with external tools that provide higher-order matrix element calculations to improve accuracy. SHERPA [68] is another general-purpose generator, offering a modular framework that incorporates all components necessary for a factorised and probabilistic description of scattering events at hadron–hadron, lepton–hadron, and lepton–lepton colliders.

After the generation of stable final-state particles, their interactions with the detector are modelled using the GEANT4 framework [69], a toolkit widely used in high-energy physics to simulate the passage of particles through matter. It provides a detailed description of the ATLAS detector’s material and geometry.

Two main approaches are available for detector simulation:

- **Full simulation** offers highly accurate modelling of particle interactions, including detailed electromagnetic and hadronic showers in the calorimeters, simulated using MC techniques. It fully accounts for the complexity of the detector geometry and response and is therefore computationally intensive.
- **Fast simulation** relies on a parametrised response of the calorimeter systems and other simplified models, significantly reducing simulation time while maintaining reasonable accuracy for many physics studies. The speed-up factor depends on the event topology and complexity [70].

In the context of this thesis, all simulated samples use the full simulation chain, since a large part of this work relies on detailed GEANT4 simulation of photon showers and hadronic deposits in the calorimeters.

Chapter 2

Experimental setup

Contents

2.1	CERN	32
2.2	The Large Hadron Collider	33
2.2.1	Accelerator overview	33
2.2.2	Luminosity and pile-up	34
2.3	The High-Luminosity Large Hadron Collider	35
2.4	The ATLAS experiment	36
2.4.1	Overview	36
2.4.2	Coordinate system	37
2.4.3	The Inner Detector	39
2.4.4	The calorimeter system	43
2.4.5	The muon spectrometer	50
2.4.6	The magnet system	51
2.4.7	Trigger and data acquisition system	53

2.1 CERN

The idea of founding a European laboratory for research in particle and nuclear physics emerged in the aftermath of World War II, in a context where European science had lost its international prominence [71]. This initiative aimed to bring together scientists from across Europe and to share the costs of nuclear facilities. It was not until December 1951 that the first resolution for the creation of a European Council for Nuclear Research (*Conseil Européen pour la Recherche Nucléaire*) was adopted at an intergovernmental meeting at UNESCO in Paris, France, giving birth to the acronym CERN. Less than two years later, on June 29, 1953, the Convention for the Establishment of a European Organization for Nuclear Research was adopted by 12 countries, establishing fundamental research and the sharing of knowledge as core principles of the Organization, as stated in Article II, Paragraph 1 [72]:

The Organization shall provide for collaboration among European States in nuclear research of a pure scientific and fundamental character, and in research essentially related thereto. The Organization shall have no concern with work for military requirements and the results of its experimental and theoretical work shall be published or otherwise made generally available.

The first CERN accelerator, the Synchrocyclotron, was inaugurated in 1957. Several other accelerators were built in the following years, including the Proton Synchrotron (PS), launched in 1959, and the Super Proton Synchrotron (SPS) in 1976. The SPS enabled studies of matter-antimatter asymmetry and exotic forms of matter. Notably, it led to the discovery of the W and Z bosons, leading to a Nobel Prize in Physics in 1983. In 1988, a 27-kilometre circular tunnel was built for a new electron-positron collider, the Large Electron-Positron Collider (LEP). It operated from 1989 to 2000 and was then replaced by the Large Hadron Collider (LHC) in 2008, where the Higgs boson was famously discovered. However, CERN has not only enabled discoveries in fundamental physics but has also contributed to the development of tools that have revolutionised everyday life. It was a pioneer in touch-screen technology, developed for the Super Proton Synchrotron (SPS) control system, and it was at CERN that computer scientist Tim Berners-Lee created the World Wide Web in 1990 [73], whose first website is still online: <http://info.cern.ch/hypertext/WWW/TheProject.html>.



Figure 2.1: The first web server: this NeXT machine was used by Tim Berners-Lee in 1990 to develop and run the first WWW server, multimedia browser and web editor [74].

Today, CERN is a vast scientific organization, with more than 17 500 people collaborating to explore the fundamental laws of matter, including 12 200 scientists of 110 nationalities, from institutes in more than 70 countries [75].

2.2 The Large Hadron Collider

2.2.1 Accelerator overview

The Large Hadron Collider (LHC) [76] is a two-ring superconducting hadron accelerator and collider, of 26.7 km, the current longest one in the world. It is located under the French-Swiss border, near Geneva, within the tunnel of its predecessor, the LEP, which lies between 45 m and 170 m under the surface of the ground. The LHC is designed to accelerate and collide proton beams with a centre-of-mass energy of 14 TeV and heavy ions with an energy of 2.8 TeV per nucleon. During Run 1, centre-of-mass energies of 7 TeV and 8 TeV were employed, while Run 2 utilised 13 TeV. The LHC now operates at a centre-of-mass energy of 13.6 TeV for Run 3. Two beams of hadrons are accelerated up to 6.8 TeV in opposite directions in two pipes kept at ultra-high vacuum. The strong magnetic field necessary for this purpose is maintained by superconducting magnets, whose temperature is kept at -271.3°C , colder than outer space. More than 10 000 magnets are used to guide the hadrons in the accelerator.

Before entering the LHC, protons pass through a sequence of four accelerators, shown in Figure 2.2. H^{-} ions, extracted from a hydrogen bottle using an electric field, are injected into Linac4 [77] (Linac2 up to 2020), a linear accelerator designed to accelerate H^{-} ions up to 160 MeV. These ions are subsequently transmitted to the Proton Synchrotron Booster (or simply Booster), where the two electrons of the H^{-} ions are stripped away, and the resulting protons are accelerated up to 2 GeV. The protons are then injected into the Proton Synchrotron (PS), where they are further accelerated to 26 GeV, and then into the Super Proton Synchrotron (SPS), where they reach 450 GeV, before being injected into the LHC, where they get their final energy, 6.8 TeV.

Once in the LHC, protons collide at four interaction points, where four particle detectors (or rather, four sets of detectors) have been placed: ATLAS [79], CMS [80], ALICE [81], and LHCb [82].

- **ATLAS**, for **A Toroidal LHC ApparatuS**, is a general-purpose detector designed to investigate a variety of physics topics, including the Higgs boson, extra dimensions, and dark matter. The detector has colossal dimensions: 46 meters long and 25 meters in diameter, with a weight of about 7000 tons. It is located 100 meters below ground near the main CERN site. The ATLAS collaboration comprises over 6000 scientists from 177 institutions across 40 countries. As the work presented in this thesis is part of the ATLAS experiment, a detailed description of the detector is provided in Section 2.4.
- **CMS**, for **Compact Muon Solenoid**, is another general-purpose detector that shares common scientific objectives with ATLAS. It is built around a massive solenoid magnet, consisting of superconducting cables that produce a magnetic field of 4 teslas. The detector measures 21 meters in length and 15 meters in diameter, with a total weight of about 14 000 tons.

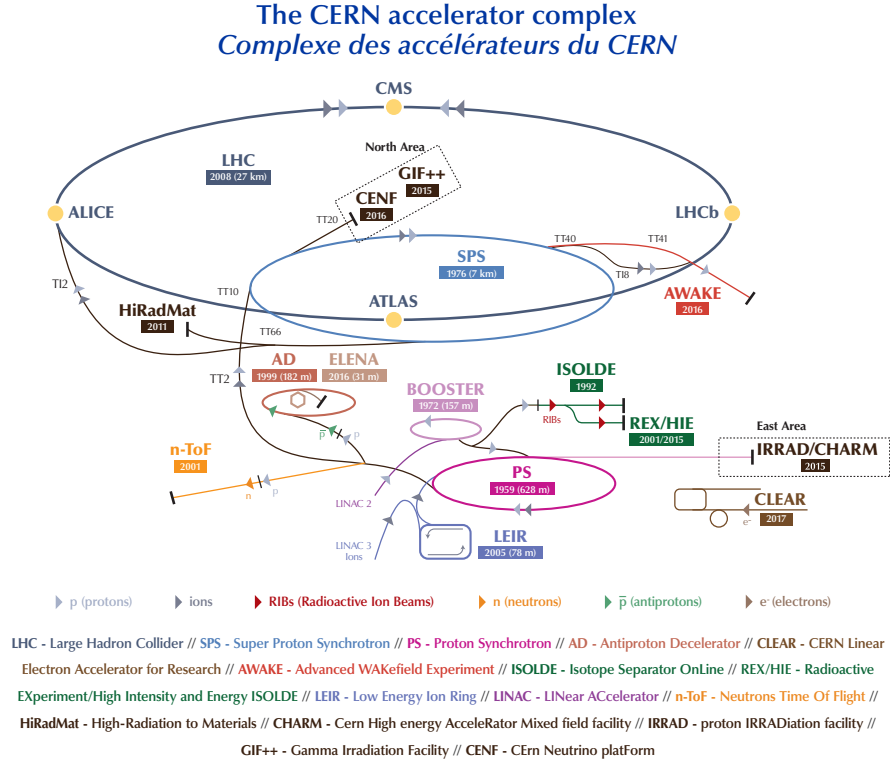


Figure 2.2: The CERN accelerator complex [78].

- **ALICE** (**A** **L**arge **I**on **C**ollider **E**xperiment) is a detector specialised in the physics of heavy ions. It studies quark-gluon plasma, a state of matter which existed immediately after the Big Bang. In this state, the temperature conditions are so extreme that quarks become unconfined from gluons, no longer remaining inside hadrons.
- **LHCb** (**L**arge **H**adron **C**ollider **b**eauty) explores the difference between matter and antimatter through the *beauty* (or *bottom*) quark. It is designed to detect mainly forward particles.

2.2.2 Luminosity and pile-up

At the LHC, protons travel in bunches of about 10^{11} , separated by 25 ns. The number of recorded events per unit time is related to the instantaneous luminosity, which quantifies the number of events that can be recorded, as follows:

$$\frac{dN}{dt} = \mathcal{L}\sigma, \quad (2.1)$$

where \mathcal{L} is the instantaneous luminosity, given in $\text{cm}^{-2} \text{s}^{-1}$, and σ is the cross section of the process.

The centre-of-mass energy is the energy of the system of particles in the centre-of-momentum frame. For two colliding particles, with four momenta $p_1 = (E_1, \vec{p}_1)$ and $p_2 = (E_2, \vec{p}_2)$, the centre-of-mass energy is simply determined by:

$$\begin{aligned}
 s &= (p_1 + p_2)^2, \\
 &= (E_1 + E_2)^2 - (\vec{p}_1 + \vec{p}_2)^2, \\
 &= (E_1 + E_2)^2,
 \end{aligned} \tag{2.2}$$

since $\vec{p}_1 + \vec{p}_2 = \vec{0}$ in the centre-of-mass frame.

So, for two protons colliding in the LHC, the centre-of-mass energy achieved in the collision is just the sum of the energies of the two particles:

$$\sqrt{s} = E_1 + E_2. \tag{2.3}$$

In the laboratory frame, for two relativistic particles, this equation becomes:

$$\sqrt{s} = 2\sqrt{E_1 E_2}. \tag{2.4}$$

For two beams with Gaussian transverse densities, the instantaneous luminosity is given by [83]:

$$\mathcal{L} = \frac{N_1 N_2 f N_b}{4\pi\sigma_x\sigma_y}, \tag{2.5}$$

where N_1 and N_2 are the numbers of particles in each bunch, f the revolution frequency, N_b the number of bunches in one beam, σ_x and σ_y the beam sizes.

From the instantaneous luminosity, we can calculate the integrated luminosity, given in fb^{-1} , which is the integral of the instantaneous luminosity over time:

$$L_{\text{int}} = \int \mathcal{L}(t) dt. \tag{2.6}$$

The number of recorded events is then given by:

$$N = L_{\text{int}} \sigma. \tag{2.7}$$

In total, LHC delivered 156 fb^{-1} during Run 2, and 147 fb^{-1} were recorded by ATLAS. Only 139 fb^{-1} were certified to be good quality data and were used in the analysis. Figure 2.3 shows integrated luminosity over time, for Run 2 and Run 3.

Another crucial measurement is pile-up. When two bunches intersect, multiple protons collide simultaneously, resulting in several nearly simultaneous collisions. This phenomenon is known as pile-up. Less energetic collisions are not of interest, we focus on the most energetic ones. It is essential to account for all vertices during the reconstruction process. To achieve this, we compute the average number of interactions per bunch crossing, denoted as $\langle\mu\rangle$. Figure 2.4 illustrates the evolution of pile-up over time for Run 2 and Run 3.

2.3 The High-Luminosity Large Hadron Collider

The High-Luminosity Large Hadron Collider (HL-LHC) [88] represents the next major upgrade of the Large Hadron Collider. It aims to reach a number of pp collision per bunch crossing of 140 and an integrated luminosity of 3000 fb^{-1} . This enhancement will facilitate the study of extremely rare phenomena, necessitating a vast quantity of data. The HL-LHC is expected to achieve an integrated luminosity of 250 fb^{-1} per year, with a goal of collecting 3000 fb^{-1} over twelve years of operation.

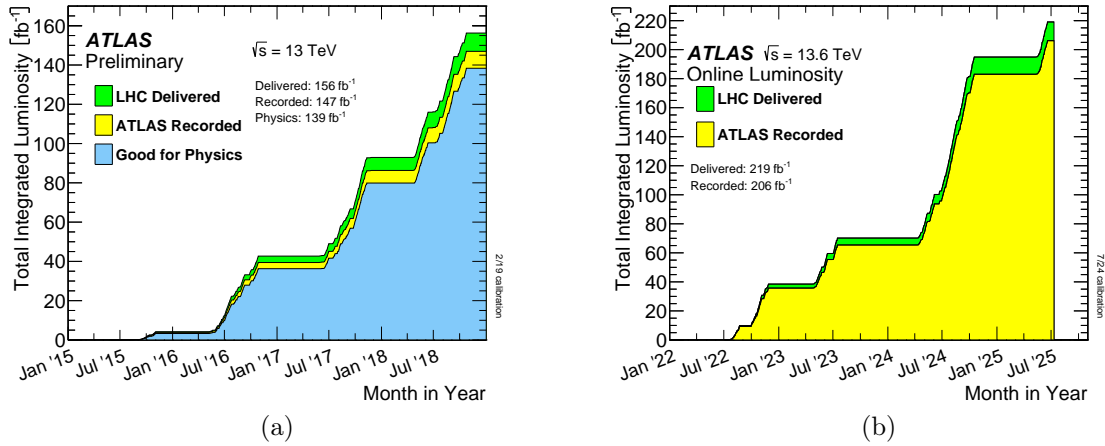


Figure 2.3: (a) Total integrated luminosity and data quality in 2015-2018 (Run 2) [84]. (b) Total integrated luminosity in Run 3 (13.6 TeV pp data only) [85].

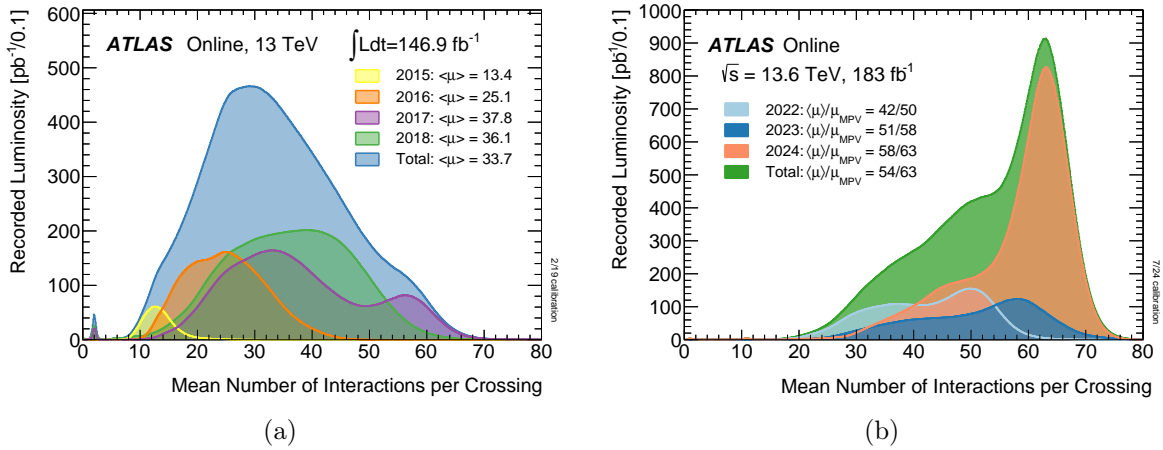


Figure 2.4: Recorded pile-up by ATLAS in (a) Run 2 (2015-2018) [86] and (b) Run 3 (2022-2024) [87].

The upgrade will incorporate several cutting-edge technologies, including 11 to 12 T superconducting magnets, novel magnet designs, and very compact superconducting RF cavities for beam rotation with ultra-precise phase control. Additionally, new technologies and materials will be used for beam collimation, and high-current superconducting links with almost zero energy dissipation will be implemented.

Data taking should start by 2030. The HL-LHC plan is shown in Figure 2.5.

2.4 The ATLAS experiment

2.4.1 Overview

The ATLAS detector [79], located approximately 90 metres below the surface at the Meyrin site, is designed to investigate a wide range of phenomena, including the Higgs boson, extra dimensions, and supersymmetry. The detector has a cylindrical geometry around the LHC beam line and is forward-backward symmetric with respect to the in-

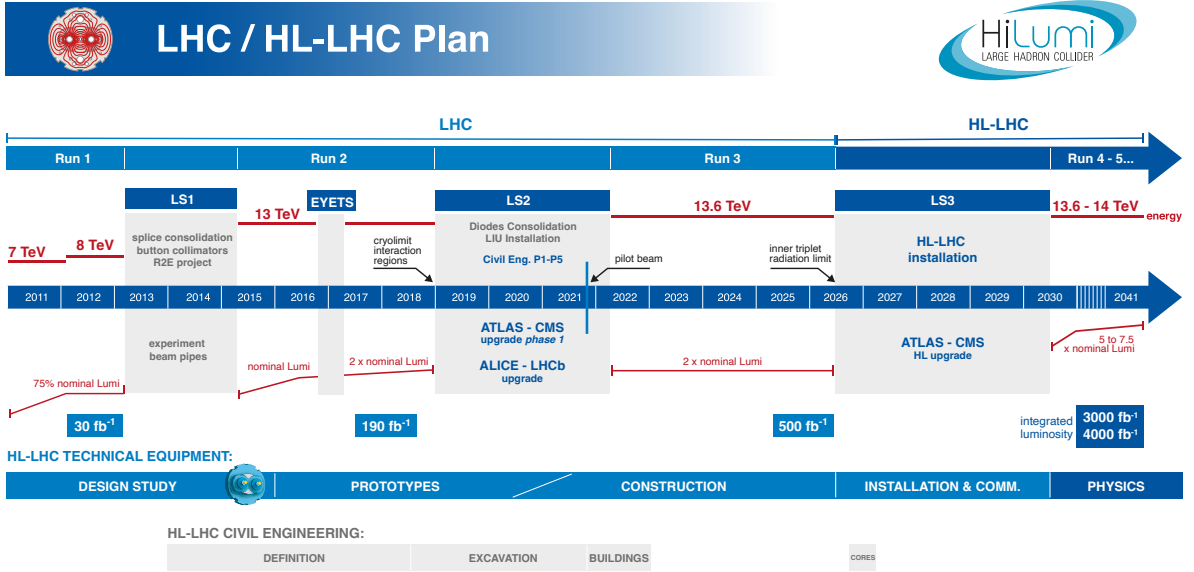


Figure 2.5: LHC and HL-LHC plan [89].

teraction point (IP). It consists of a central barrel and two end-caps on either side. It is shown in Figure 2.6.

It comprises several sub-detectors that measure the energy, momentum, position, and timing of particles interacting with the detector materials. Closest to the beam axis, the Inner Detector (ID) operates within a 2 T magnetic field and reconstructs the trajectories of charged particles. Surrounding the ID, the calorimetry systems measure the energy deposits from electromagnetic and hadronic particles. The muon spectrometers, consisting of tracking systems and toroidal magnets, reconstruct the momenta of muons, which escape from the calorimeters.

The magnet system bends charged particles, allowing for the measurement of their momenta. This system includes one solenoid and three toroidal (one barrel and two end-caps) superconducting magnet systems, storing an energy of 1.6 GJ. The solenoid is aligned with the beam axis and provides a 2 T magnetic field to the Inner Detector. The barrel and end-cap toroids generate magnetic fields of 0.5 T and 1 T, respectively, for the muon detectors in the central and end-cap regions. A cryogenic system delivers gaseous helium at 4.5 K and 60 K to cool the cold mass and the thermal shield.

A schematic transverse view of the sub-detectors, illustrating the paths of different types of particles as they traverse the detector, is shown in Figure 2.7.

2.4.2 Coordinate system

ATLAS employs a right-handed coordinate system where the x -axis points towards the centre of the accelerator, the y -axis points upwards towards the ground surface, and the z -axis runs along the beam pipe. The origin is located at the interaction point. Given the cylindrical geometry of the detector, cylindrical coordinates can be introduced:

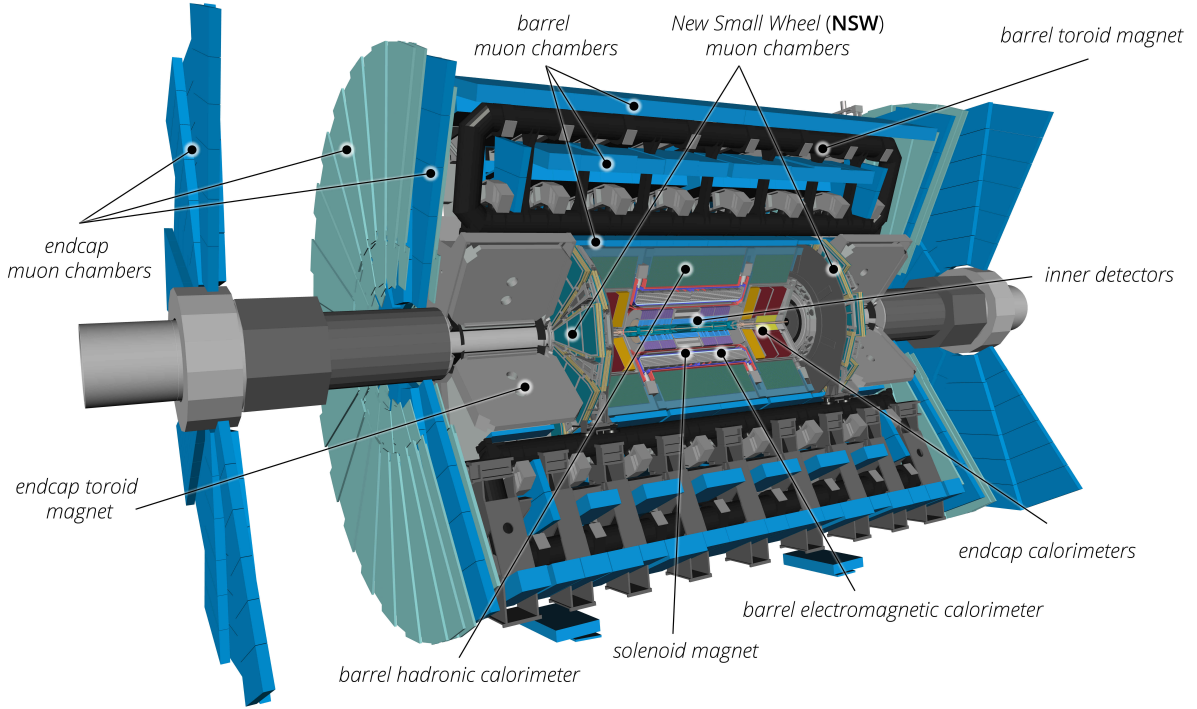


Figure 2.6: Run 3 configuration of the ATLAS detector [90].

$$\begin{cases} r = \sqrt{x^2 + y^2} \\ \phi = \arctan(y/x) \\ z \end{cases} \quad (2.8)$$

In most analyses, spherical coordinates are used. However, instead of the polar angle $\theta = \arccos(z/r)$, it is more convenient to use *rapidity*:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right), \quad (2.9)$$

or *pseudorapidity*:

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right], \quad (2.10)$$

as differences in rapidity, Δy , or in pseudorapidity, $\Delta \eta$, are Lorentz invariant under boosts along the beam axis. Even though the protons collide with the same energy, the net momentum along the z -axis is not zero due to the substructure of the protons, with each parton carrying a fraction of the total momentum. The rapidity y converges towards the pseudorapidity η in the massless limit, $m \ll E$. Table 2.1 shows η coverage for the different sub-detectors. Distances between objects are often measured using η and the azimuthal angle ϕ :

$$\Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2}. \quad (2.11)$$

Also defined in the x - y plane are the transverse momentum $p_T = \|\vec{p}\| \sin \theta$ and the transverse energy $E_T = E \sin \theta$.

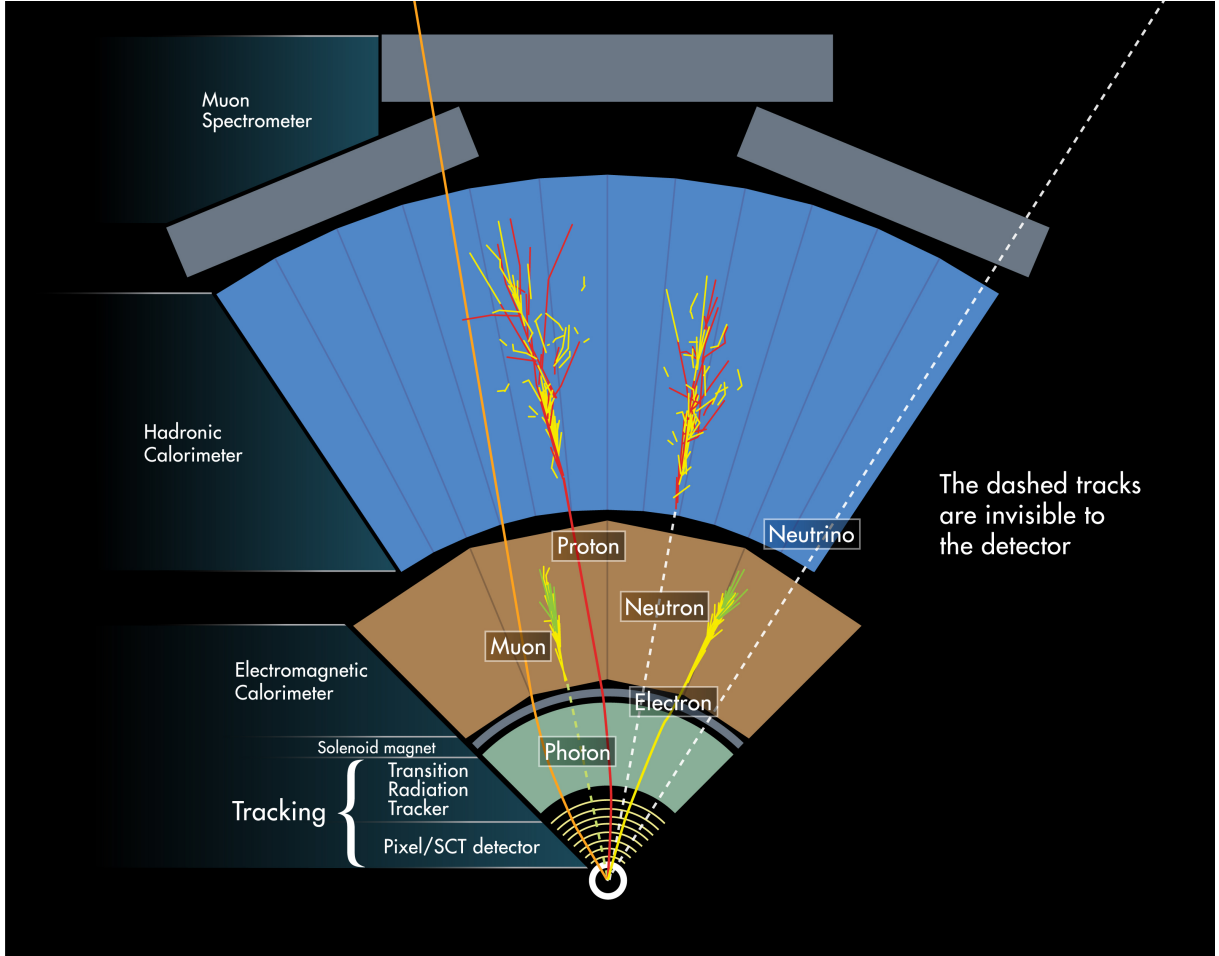


Figure 2.7: Diagram of particle paths in the detector [91].

2.4.3 The Inner Detector

The Inner Detector [92] is designed to operate within the high pile-up environment of the LHC, as close to the interaction point as possible. The high granularity of its detectors enables precise reconstruction of tracks and vertices, contributing to the identification of electrons and muons, as well as the detection of short-lived particle decay vertices. In its inner section, the Inner Detector features high-resolution detectors, while the outer section contains continuous tracking elements. The entire system operates within a 2 T magnetic field generated by the solenoidal magnet. A picture of the ID is shown in Figure 2.8. Front and lateral cuts of the ID are shown in Figure 2.9 and Figure 2.10.

2.4.3.1 The Pixel Detector

The Pixel Detector is the beam-closest subsystem. Its very high granularity enables high-precision charged particle tracking, as close to the interaction point as possible. It rests on silicon-based sensors, distributed in four barrel layers, located at distances $|z| < 400.5$ mm and radii from 33.25 mm to 122.5 mm, and two end-caps of three disks each, at distances z from 495 mm to 650 mm, between radii of 88.8 mm and 149.6 mm. These elements and their positions are depicted on Figure 2.11. It has an acceptance range in pseudorapidity η of ± 2.5 .

The three outer barrel layers and the end-cap disks were already in place during Run 1.

Detector component	η coverage
Inner Detector	± 2.5
Electromagnetic calorimeter	± 3.2
Hadronic calorimeter <i>barrel and end-cap</i> <i>forward</i>	± 3.2 $3.1 < \eta < 4.9$
Muon spectrometer	± 2.7

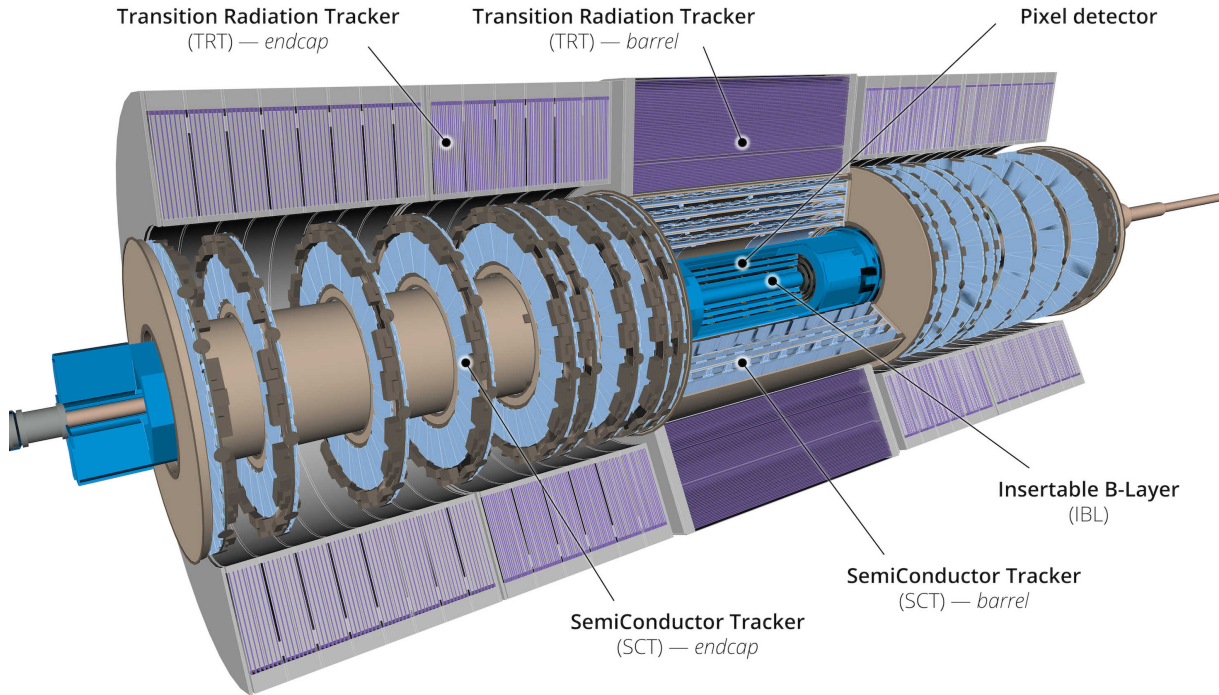
Table 2.1: η coverage of the ATLAS sub-detectors.

Figure 2.8: ATLAS Inner Detector [90].

They are made of 1744 pixel sensors, each of them formed of 47 232 pixels, including 46 080 readout channels. In total, about 80 million channels are operating, including about 67 million in the 3 cylindrical barrel layers and 13 million in the 3 forward and backward end-cap disk layers. The nominal pixel size is $50 \times 400 \mu\text{m}^2$.

The innermost barrel layer, known as the Insertable B-Layer (IBL) [94], was added in 2014 during the first long shutdown of the LHC at the end of Run 1. It was installed between the previous innermost layer, the B-layer, and a new, smaller-radius beam pipe. This addition was necessary to accommodate the increased luminosity, ensuring robust and precise tracking, particularly for vertexing and b -tagging. The IBL consists of approximately 6 million pixels, which are smaller than those in other layers, with dimensions of $50 \times 250 \mu\text{m}^2$.

2.4.3.2 The Semiconductor Tracker

Around the Pixel Detector, the Semiconductor Tracker (SCT) contributes to the measurement of momentum, impact parameter and vertex position, offering a good pattern recognition with its high granularity, up to $|\eta| < 2.5$. The SCT is made of four cylindrical

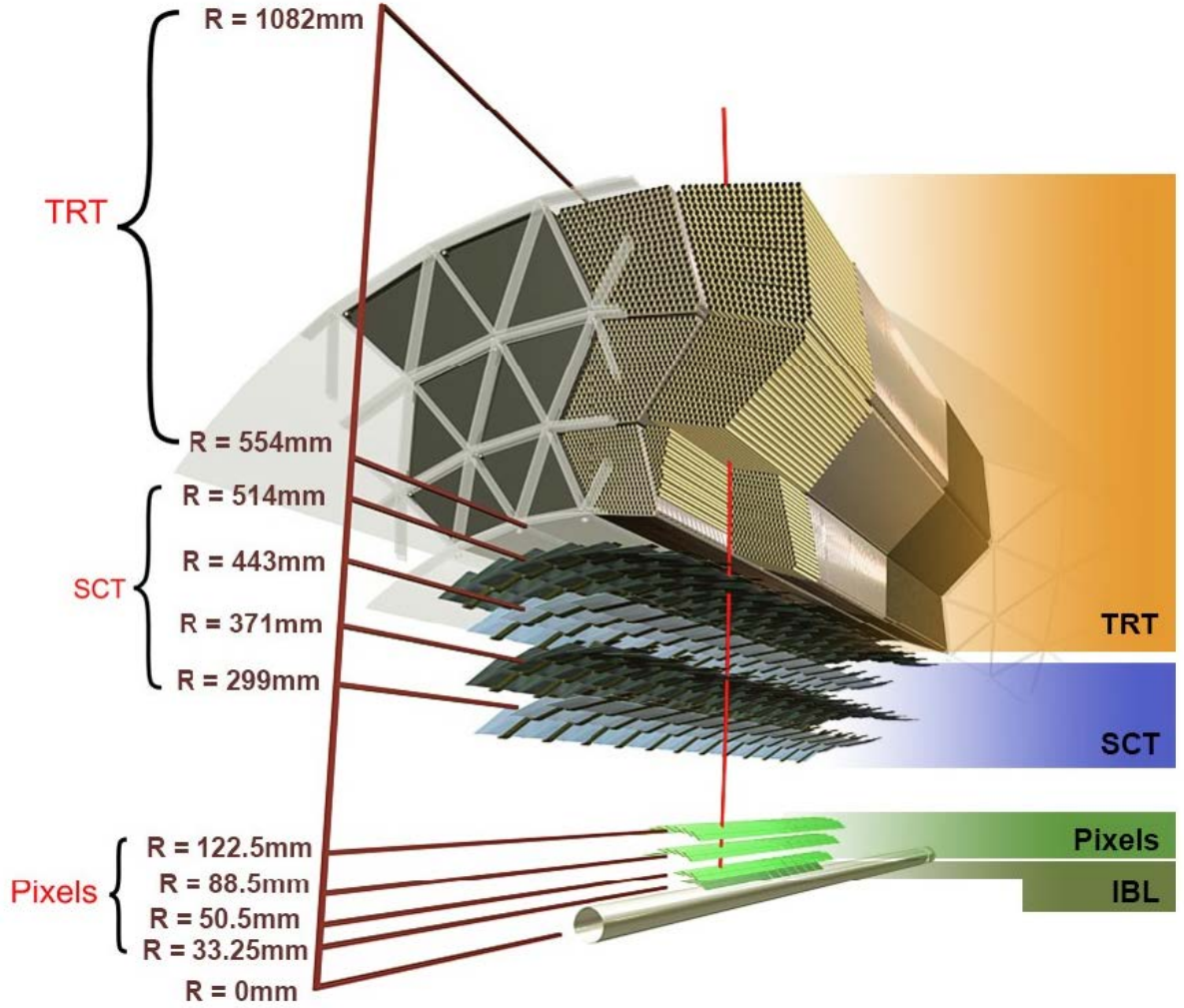


Figure 2.9: The structure of the Inner Detector, with the Pixel Detector, including the Insertable B-Layer, the Semiconducting Tracker, and the Transition Radiation Tracker [93]. The trajectory of a hypothetical particle is represented (red line).

layers in the barrel and nine end-cap disks on each side. Similarly to the Pixel Detector, the SCT is composed of silicon microstrip sensors, but longer in the z direction (126 mm in length and $80\mu\text{m}$ in pitch for the barrel layers). Each module is itself made of two layers of sensor strips, separated by an angle of 40 mrad . The strips are parallel to the beam line in the barrel and orthogonal in the end-caps. Its 6.3 million readout channels present a resolution of $17\mu\text{m}$ in the $R - \phi$ direction and $580\mu\text{m}$ in the z direction.

2.4.3.3 The Transition Radiation Tracker

The Transition Radiation Tracker (TRT) is the outermost sub-detector located inside the solenoid coil of the Inner Detector (ID), extending radially up to a distance of 1 m. It holds a unique position within the ID by providing both tracking information and particle identification. This feature is particularly useful for distinguishing between electrons and charged pions in the momentum range from 1 to 200 GeV. The TRT consists of approximately 300 000 drift tubes, or “straws”, each with a diameter of 4 mm. These straws are 144 cm long in the barrel region and 37 cm in the end-caps. The TRT covers

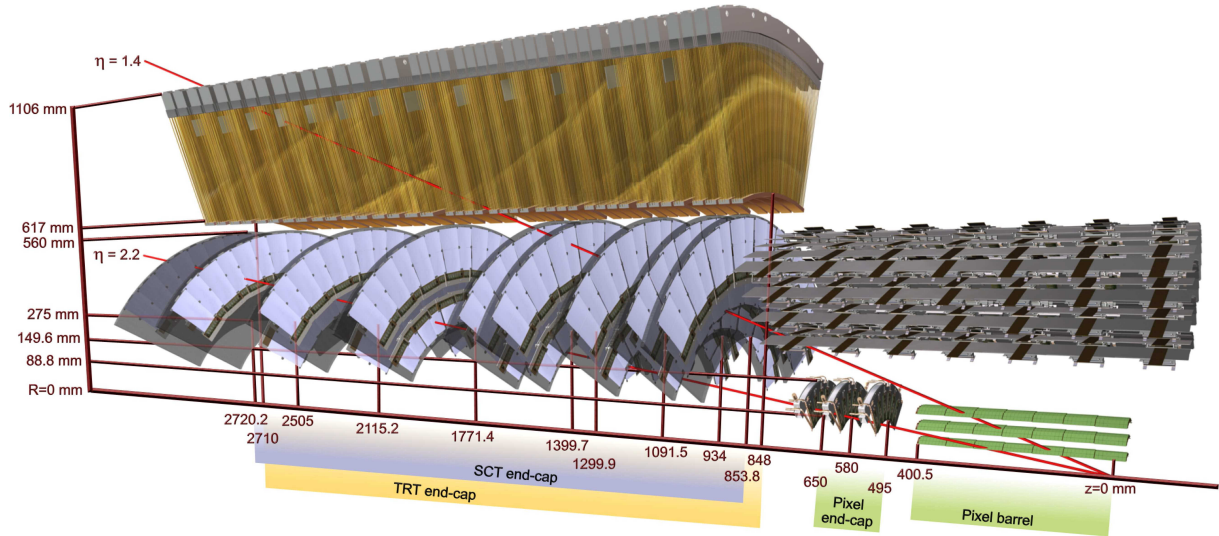


Figure 2.10: Lateral cut-off of the Inner Detector, showing the barrel and end-cap components [79]. The trajectories of two hypothetical charged particles are represented (red lines). The TRT barrel and the IBL are not represented.

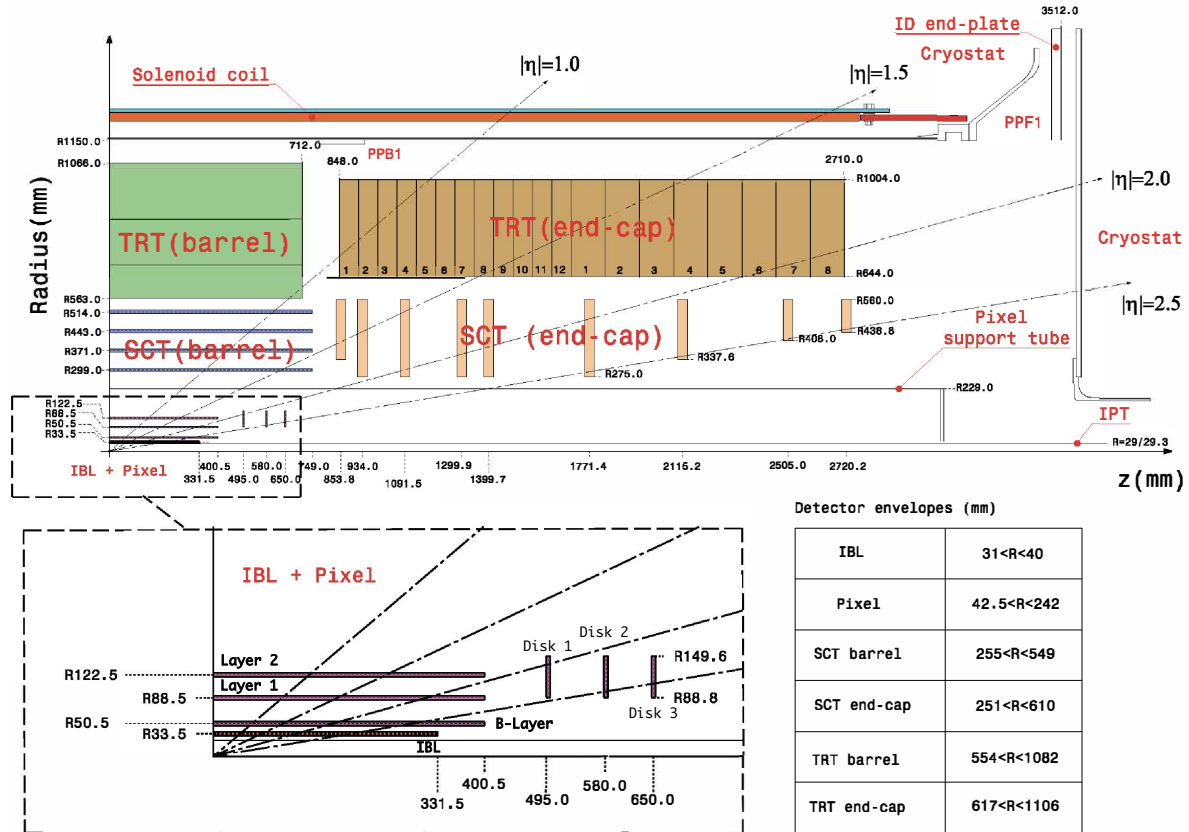


Figure 2.11: Layout of the Inner Detector in the Radius- z plan, showing the major elements and their active dimensions [90].

pseudorapidity values up to $\eta < 2.0$.

Each straw offers a position resolution of $130 \mu\text{m}$ in the $R - \phi$ plane of the trajectory. While this resolution is lower than that of the Pixel Detector or the Semiconductor Tracker

(SCT), the high number of hits per track and the longer tracks measured by the TRT compensate for this. In the barrel, the straws are arranged in 73 layers along the beam axis, while in the end-caps, they are organised radially into 160 layers. The spaces between the layers are filled with polyethylene fibre in the barrel and polypropylene foil in the end-caps.

The drift tubes are filled with a gas mixture of xenon (Xe), replaced by argon (Ar) in a large part of the detector during Run 3, carbon dioxide (CO₂), and dioxygen (O₂). Each straw is constructed from aluminium and contains a gold-plated tungsten wire running along its centre, which detects signals generated by particles passing through the detector. When a charged particle traverses a straw, it ionises the gas inside, producing electron-hole pairs that drift under the influence of the potential difference within the straw, creating an electrical current collected by the wire. This process also produces transition radiation, which aids in particle identification.

2.4.4 The calorimeter system

The calorimeter [95] is the next subsystem in the ATLAS detector after the Inner Detector. Its primary role is to measure the energy of electrons and photons (*electromagnetic particles*), as well as jets, hadrons, and missing transverse energy (E_T^{miss}). The ATLAS calorimeter system consists of two main types of calorimeters: electromagnetic and hadronic calorimeters. The former is designed to measure the energy of electromagnetic particles, while the latter measures the energy of particles that interact via the strong force (i.e. hadronic particles).

Two types of calorimeter designs are used. Sampling calorimeters, like the ones of ATLAS, alternate layers of absorber and active material. The absorber is a dense material used to stop incoming particles and initiate the creation of a particle shower. The secondary (daughter) particles then interact with the active medium, which generates a detectable signal. By analysing the signal produced by these daughter particles, the properties of the original incident particles can be inferred. Homogeneous calorimeters use a single material for both the absorber and the active medium. They are used only as EM calorimeters, since they would need to be extremely large to completely stop hadronic particles.

2.4.4.1 Interaction of electromagnetic particles with matter

Electromagnetic interactions between particles and matter differ substantially depending on the particle's type — whether it is electrically charged or neutral — and its energy.

Electrons can interact with matter in several ways, including ionisation, excitation, Bremsstrahlung, Cherenkov radiation, and Møller scattering. Figure 2.12a illustrates the various processes by which electrons lose energy as they pass through lead, plotted as a function of energy. In the ATLAS environment, the two main processes are ionisation and Bremsstrahlung. At low energies, ionisation dominates. Incoming electrons undergo elastic collisions with atomic electrons, knocking them out of their atoms. This process occurs as long as the energy of the incident electron is higher than the binding energy of the atomic electrons. The electron's energy can be easily deduced from this process, since the number of ejected electrons is directly proportional to its initial energy. This process is well described by the Bethe-Bloch formula [96]. At higher energies, Bremsstrahlung becomes the dominant effect. As introduced in the previous chapter, the Bremsstrahlung process refers to the emission of electromagnetic radiation (photons) as the electron decelerates

near an electric charge, typically an atomic nucleus. The rate of Bremsstrahlung depends on the material's atomic number (Z) and the distance between the electron and the nucleus. The total energy loss of the electron is given by:

$$\frac{dE_{tot}}{dx} = \frac{dE_{ion}}{dx} + \frac{dE_{rad}}{dx} \quad (2.12)$$

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

Since photons are electrically neutral, their interaction with matter differs substantially from that of electrons. Several processes can occur, including photoelectric absorption, Compton scattering, pair production, and Rayleigh scattering (see Figure 2.12b). At the GeV scale, the dominant process is electron-positron pair production (Compton scattering also plays a role), which occurs when the photon's energy exceeds the combined mass of the electron and the positron (1.022 MeV). In addition, the interaction must take place near an atomic nucleus to conserve momentum.

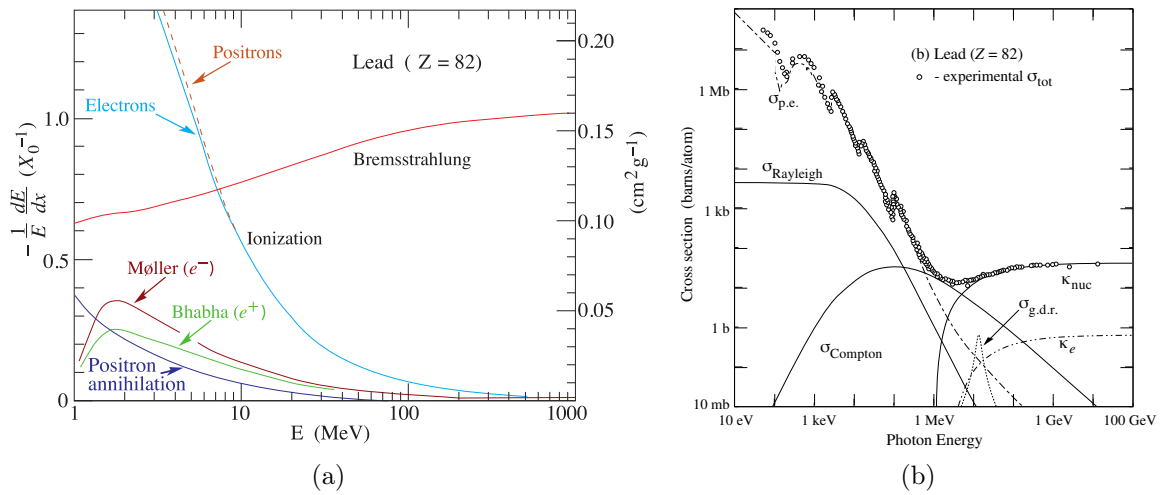


Figure 2.12: (a) Fractional energy loss per radiation length in lead as a function of electron or positron energy [97]. The value at which the energy losses by ionisation and by Bremsstrahlung are equal is called the critical energy (E_c). (b) Photon total cross sections as a function of energy in lead, showing the contributions of different processes [97]. At high energies, pair production dominates (κ_{nuc} stands for pair production in the nuclear field and κ_e in the electric field). At lower energies, Compton effect (σ_{Compton}) and photoelectric effect ($\sigma_{\text{p.e.}}$) are dominant.

When a photon enters the calorimeter, it initiates an electromagnetic (EM) shower by producing an electron-positron pair. These particles then emit photons via the Bremsstrahlung effect, which in turn generates additional electron-positron pairs, continuing the cascade. This process progresses until the energy of the electrons falls below the critical energy E_c , where energy losses due to ionisation and Bremsstrahlung become equal. At this point, the electrons cease radiating, and the shower development stops. The energy of the initial photon E_0 can be estimated from the shower, as the ratio E_0/E_c is proportional to the number of particles in the shower.

An electromagnetic shower is characterised by two parameters: its longitudinal development and the lateral one. The longitudinal development is described by the radiation length X_0 , function of the atomic number Z and weight A of the traversed material, mea-

sured in g/cm^2 or in cm , which is the mean distance over which a high-energy electron loses $1/e$ of its energy by Bremsstrahlung

$$\langle E(x) \rangle = E_0 e^{-\frac{x}{X_0}}, \quad (2.13)$$

and $7/9$ of the mean free path for pair production by a high-energy photon. For a photon beam with initial intensity I_0 , its intensity after a distance x is

$$\langle I(x) \rangle = I_0 e^{-\frac{7}{9} \frac{x}{X_0}}. \quad (2.14)$$

The mean longitudinal energy deposition by the incoming particle in the material is well described by the following function [97]:

$$\frac{dE}{dt} = E_0 \beta \frac{(\beta t)^{\alpha-1} e^{-\beta t}}{\Gamma(\alpha)}, \quad (2.15)$$

where $t = \frac{x}{X_0}$ (x being the distance traversed by the incoming particle), E_0 is the initial energy, and α and β depend on the material and the incoming particle. The gamma function is defined for complex numbers with a positive real part as:

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt. \quad (2.16)$$

It can be demonstrated that the maximum energy deposition occurs at $t_{\text{max}} = \frac{\alpha-1}{\beta} = \ln\left(\frac{E_0}{E_c}\right) - C_{e\gamma}$, where $C_{e\gamma} = 1.0$ for electron-induced showers and $C_{e\gamma} = 0.5$ for photon-induced showers. This shows that the maximum energy deposit, and consequently the particle's penetration depth, increases logarithmically with the energy of the incoming particle. Figure 2.13 shows the energy deposit of electrons with energies between 1 GeV and 1 TeV as a function of the depth in a block of copper. Additionally, the distance in radiation lengths X_0 that contains 95% of the shower energy is given by $t_{\text{max}} + 0.08Z + 9.6$. For highly energetic particles, this implies that a large calorimeter is required to fully contain the shower energy.

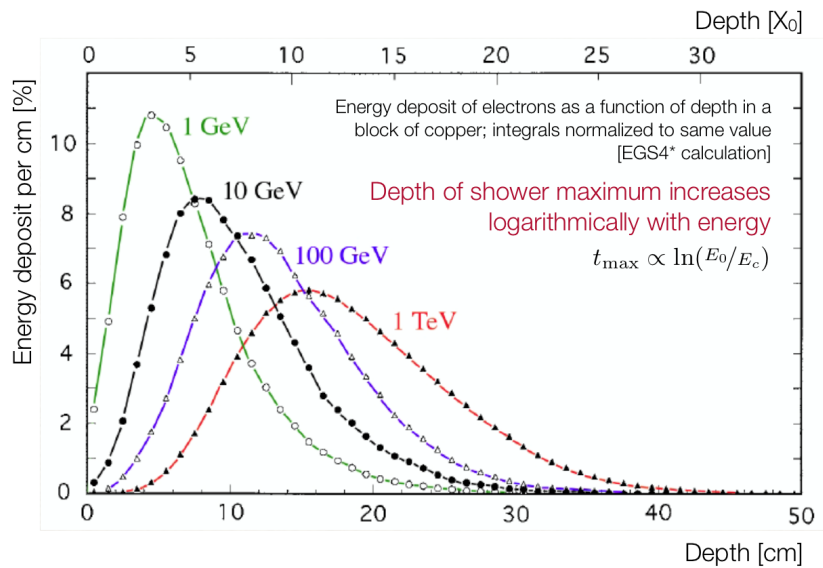


Figure 2.13: Energy deposit of electrons with energies between 1 GeV and 1 TeV as a function of the depth in a block of copper [98].

The lateral development of the EM shower is mainly due to multiple scattering of the electrons, and is well described by the Molière radius. By definition, the Molière radius R_M is the radius of a cylinder that contains 90 % of the shower, and 95 % being contained in $2R_M$. The Molière radius is approximately independent from the ongoing particle energy, and is given by $X_0 \frac{E_s}{E_c}$, with $E_s = m_e c^2 \sqrt{4\pi/\alpha_{\text{em}}} \approx 21 \text{ MeV}$, with m_e the electron mass and α_{em} the fine-structure constant.

Table 2.2 shows values that characterise EM showers in liquid argon and lead, two materials used in ATLAS EM calorimeters.

	Lead (Pb)	Liquid argon (LAr)
Radiation length (X_0) [cm]	0.56	14.2
Molière radius ($2R_M$) [cm]	1.6	10.1
Critical energy (E_c) [MeV]	7.4	30.5

Table 2.2: Lead and Liquid-Argon properties [95, 97].

2.4.4.2 Calorimetric energy resolution

Calorimeters rely on the principle that the energy of the incident particle is proportional to the energy released in the material by the secondary particles of the shower, mainly through ionisation and excitation [99]. More precisely, the total track length of the shower, T_0 , defined as the sum of all ionisation tracks due to the charged particles in the cascade, is proportional to the radiation length and to the number of particles in the cascade, E_0/E_c ,

$$T_0 \propto X_0 \frac{E_0}{E_c}. \quad (2.17)$$

This formula shows that a measurement of the signal produced by the charged tracks in the shower provides a measurement of the energy of the incident particle. The intrinsic resolution of an ideal calorimeter — a detector of infinite size, free of material defect — is due to fluctuations in the total track length T_0 , which is proportional to the number of track segments, a stochastic process. Therefore, the intrinsic resolution of a calorimeter is given by

$$\sigma_E \propto \sqrt{T_0}, \quad (2.18)$$

from which we can derive the fractional energy resolution

$$\frac{\sigma_E}{E} \propto \frac{1}{\sqrt{T_0}} \propto \frac{1}{\sqrt{E_0}}. \quad (2.19)$$

However, the energy resolution of a real calorimeter is deteriorated by other contributions and is given by

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \quad (2.20)$$

where \oplus indicates a quadratic sum. a is the *stochastic term*, b is called *noise term*, and c is the *constant term*.

- *Stochastic term*: As mentioned above, this term stands for the intrinsic fluctuations in the physical development of the shower, but also for what are called *sampling fluctuations*, which affect only sampling calorimeters and arise from the energy deposited in the absorber.
- *Noise term*: It arises from the electronic noise of the readout channel and the pile-up noise. It is generally negligible at higher energies.
- *Constant term*: This term includes contributions independent of the energy, such as instrumental defects or non-linearities. It is dominant at higher energies.

At low energies, the stochastic term dominates, with a design value of $9\text{--}10\%\sqrt{\text{GeV}}$ at low $|\eta|$, and it worsens at higher $|\eta|$ as the amount of material increases [100]. The noise term is about $350 \times \cosh \eta$ MeV for a 3×7 cluster in the $\eta \times \phi$ plane in the barrel, for $\langle \mu \rangle = 20$. At high $|\eta|$, it is dominated by pile-up noise. At high energies, the relative energy resolution approaches the constant term, which has a design value of 0.7% .

2.4.4.3 Electromagnetic calorimetry

The ATLAS electromagnetic calorimeters are sampling calorimeters, alternating layers of lead as absorbers and liquid argon (LAr) as active media. As shown in Figure 2.14, the calorimetry system is divided into three parts: a barrel and two end-caps. The barrel is itself separated into two half-barrels, 3.2 m long, of inner and outer diameters of 2.8 m and 4 m, and covering the region $|\eta| < 1.475$. A single half-barrel is made of 1024 accordion-shaped lead absorbers, interleaved with readout electrodes, placed in the middle of the liquid argon gaps. The total thickness of the module is at least $22 X_0$: from $22 X_0$ to $30 X_0$ for $|\eta| < 0.8$, and from $24 X_0$ to $33 X_0$ for $0.8 < |\eta| < 1.3$, as shown in Figure 2.15. The electromagnetic end-caps (EMECs) consist of two wheels, of internal and external radii of 330 mm and 2098 mm, placed on each side of the electromagnetic barrel, covering $1.375 < |\eta| < 3.2$. Each EMEC is itself made of two coaxial wheels, an outer wheel, up to $|\eta| = 2.5$, and an inner wheel, up to $|\eta| = 3.2$, made of 768 and 256 lead absorbers, respectively.

The calorimeter's cylindrical geometry presents challenges for achieving full coverage in the azimuthal angle ϕ , as ground lines and wires are needed to transport the electric charges collected by the electrodes in the liquid argon (LAr) gaps. The accordion geometry, shown in Figure 2.16a, addresses this issue by allowing the placement of readout channels either in front of or behind the calorimeter, enabling full ϕ coverage without cracks and maintaining excellent resolution. To ensure the LAr gaps remain constant, the direction and folding angle of the accordion waves vary depending on the part of the calorimeter. In the barrel region, the waves are axial and run along ϕ , while in the end-caps, they are parallel to the radial direction and run axially. Additionally, the bending direction of the waves changes with the radius.

The ATLAS calorimeters are divided into three longitudinal layers — also denominated as *samplings* — in the barrel (visible in Figure 2.16b) and in the outer wheel of the end-caps, and into two layers in the inner wheels. The first layer (L1, or strip layer) is finely granulated along η and allows identification of the π^0 , decaying into two photons. Since they are usually boosted, a fine segmentation is needed to identify the two decaying photons, which are collimated. The second layer (L2) measures the largest fraction of the EM cascade's energy, while the third layer (L3) contains only the tail of the shower.

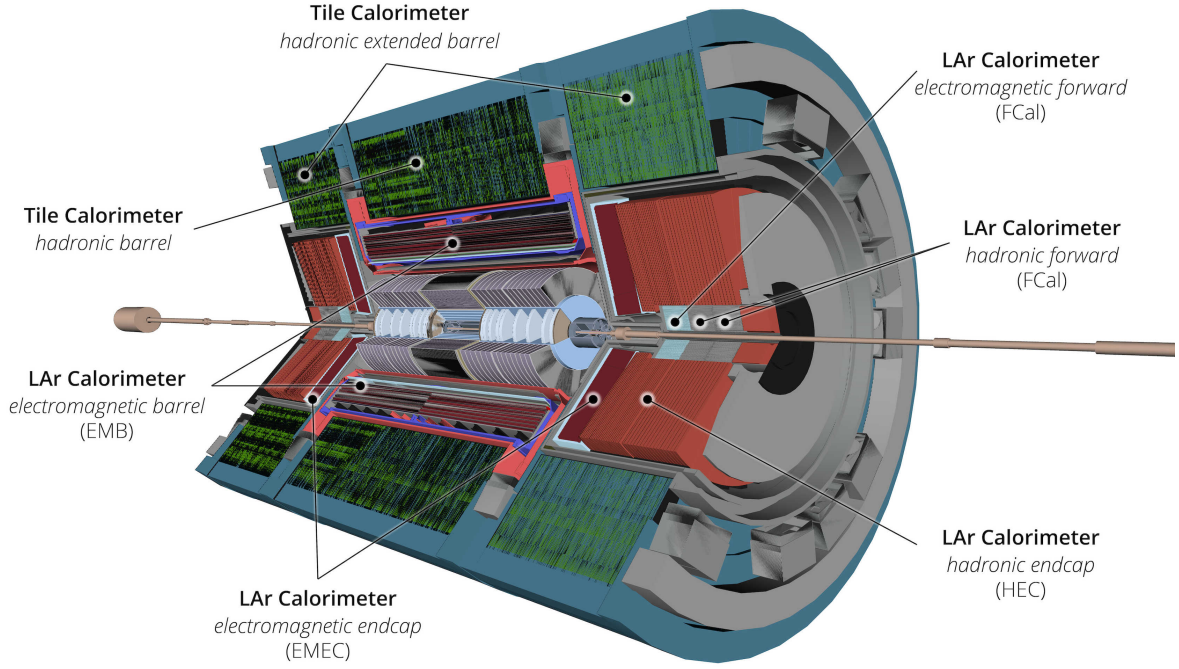


Figure 2.14: The ATLAS calorimeter systems [90].

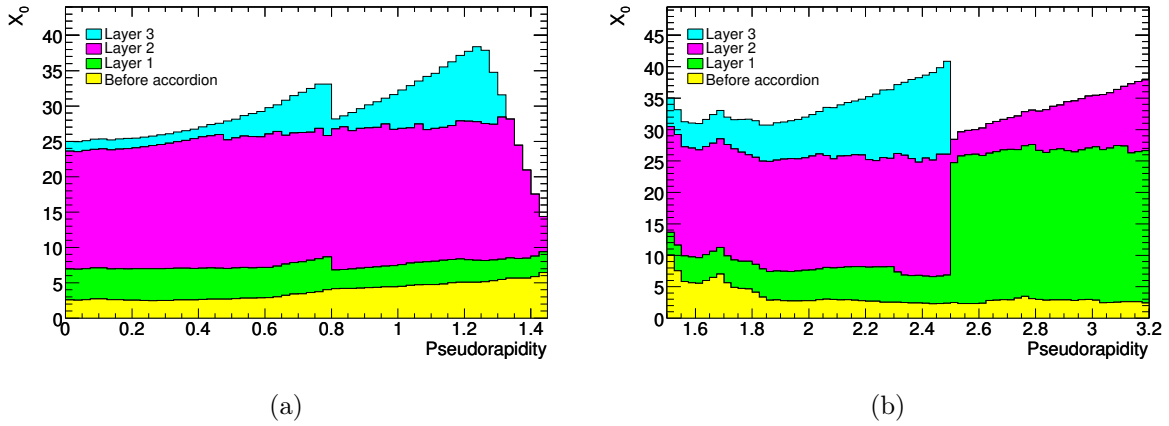


Figure 2.15: Cumulative amounts of material, in units of radiation length X_0 and as a function of $|\eta|$, in the electromagnetic calorimeters [79]. The two figures show, in contrast, separately for (a) the barrel and (b) end-cap (right), the thicknesses of each accordion layer as well as the amount of material in front of the accordion.

To keep the low temperature necessary for the LAr calorimeters to work, they are immersed in three different cryostats, one for the barrel and two for the end-caps (one shown in Figure 2.17). Each cryostat is made of two concentric aluminium vessels, an inner cold vessel and an outer warm vessel. A thin layer (11 mm in the barrel and 8 mm in the end-caps) of active media, called presampler (PS), is used upstream of the calorimeter to take into account the energy loss in the cryostat and other elements in the detector. It covers $|\eta| < 1.8$.

Additionally, scintillators are placed between the barrel and the end-caps to mitigate

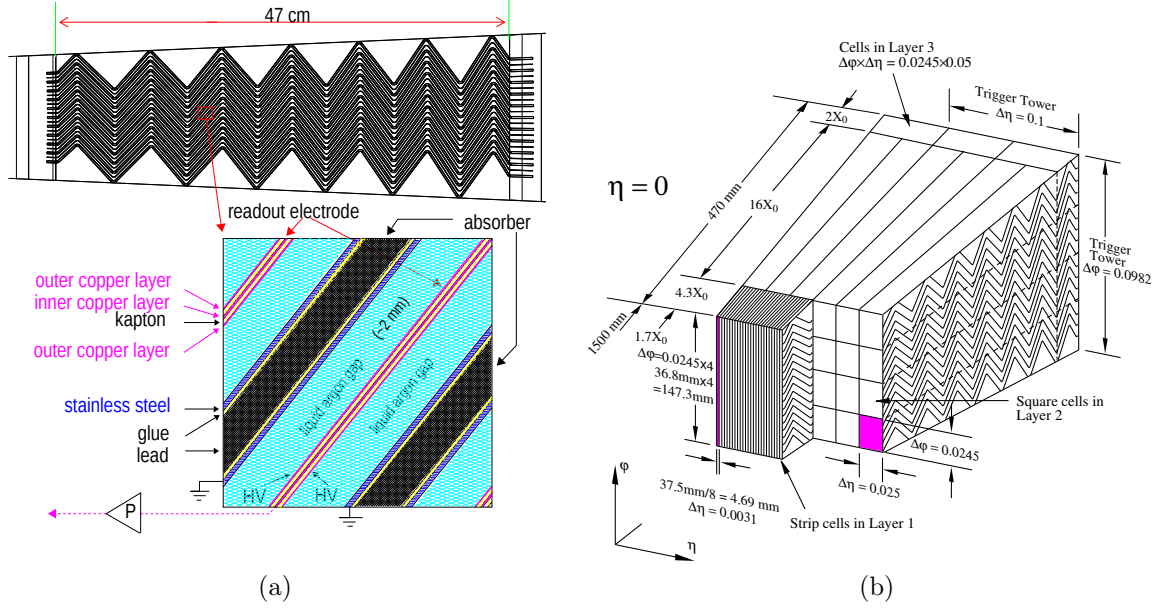


Figure 2.16: (a) Accordion structure of the barrel EM calorimeter [101]. (b) Sketch of a barrel module [79]. The different layers and the granularity in η and ϕ are visible.

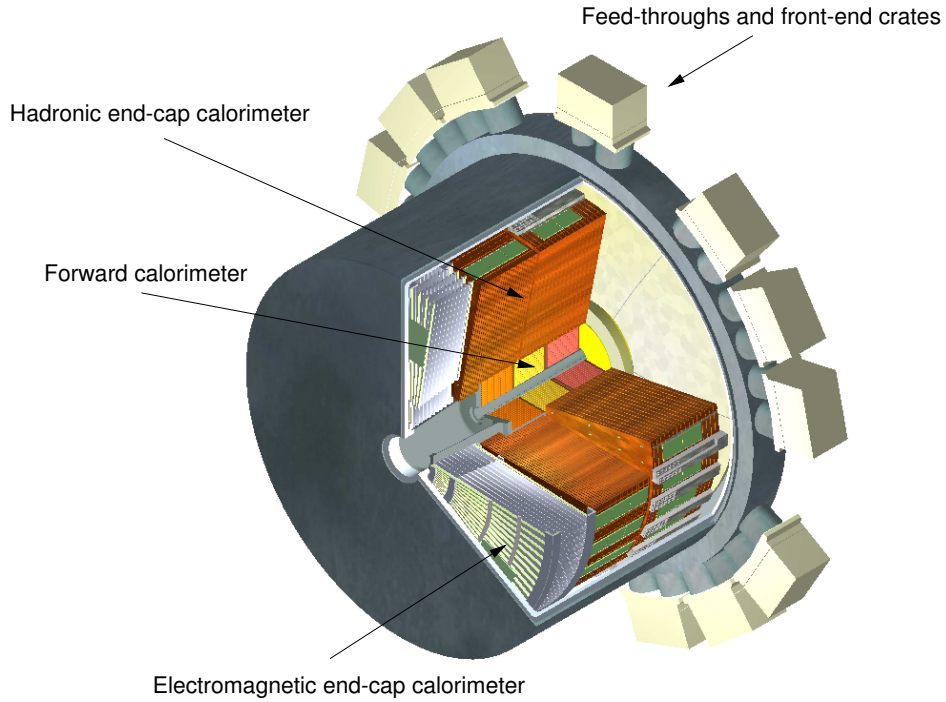


Figure 2.17: Cut-away view of an end-cap cryostat showing the positions of the three end-cap calorimeters [79].

the energy resolution degradation by performing a complementary energy measurement for $1.4 < |\eta| < 1.6$ (called the crack region). They are referred to as E4.

2.4.4.4 Hadronic calorimetry

While electrons and photons are fully absorbed in the EM calorimeters, hadrons do not entirely deposit their energy in these layers. To measure their energy, hadronic calorimeters are placed outside the EM calorimeters. Similar to EM particles, hadrons create hadronic showers, characterised by the interaction length λ_{int} , or mean free path, which is the typical distance a hadron travels before interacting with a nucleus in the medium. Typically, λ_{int} is greater than X_0 , so hadronic calorimeters must be thicker than EM calorimeters.

The ATLAS hadronic calorimetry system consists of three detectors: the Tile calorimeter, the liquid-argon hadronic end-cap calorimeter (HEC), and the liquid-argon forward calorimeter (FCal).

Tile calorimeter The Tile calorimeter is a sampling calorimeter that utilises steel as the absorber material and scintillators as the active medium. It covers the pseudorapidity range up to $|\eta| < 1.7$ and is situated behind the LAr EM calorimeter. The Tile calorimeter consists of a central barrel, which is 5.8 m long and extends up to $|\eta| = 1$, and two extended barrels, each 2.6 m long and reaching up to $|\eta| = 1.7$. The extended barrels have an inner radius of 2.28 m and an outer radius of 4.25 m. The calorimeter has a radial depth of approximately 7.4 radiation lengths (λ_{int}). The central and extended barrels are divided into 64 modules, each with an azimuthal size of $\Delta\phi \sim 0.1$, made of steel plates and scintillating tiles. When a charged particle is absorbed in the active medium, ultraviolet photons are emitted and transported through wavelength-shifting optical fibres to photomultiplier tubes (PMTs), where they are converted into measurable electric signals (see Figure 2.18). Similar to the EM calorimeters, the Tile calorimeter is segmented longitudinally into three layers.

LAr hadronic end-cap calorimeters The liquid argon hadronic end-cap calorimeters (HEC) consist of two detectors, placed on both sides of the Tile calorimeter, and covering $1.5 < |\eta| < 3.2$. It is a sampling calorimeter, also using liquid argon as active medium, but copper instead of lead as absorber, and flat-shaped layers. Each HEC is composed of two wheels that share the same cryostat as the LAr EMECs.

LAr forward calorimeter The LAr forward calorimeter (FCal) is integrated within the end-cap cryostats. Of size 10 radiation lengths approximately, it covers the region $3.1 < |\eta| < 4.9$. The FCal is a sampling calorimeter that uses LAr as the active medium and is segmented longitudinally into three layers. The first one employs copper as the absorber, while the second and third ones use tungsten. In comparison to the EM calorimeter, the LAr gaps are thinner, ranging from 0.25 mm in the first layer to 0.50 mm in the 3rd one.

2.4.5 The muon spectrometer

While most particles are completely stopped in the calorimeters, neutrinos and muons can escape detection. Neutrinos, being weakly interacting particles, traverse the detectors without depositing energy and thus remain undetected. Muons, although they interact with the detector materials, are not fully absorbed due to their high mass. To measure the tracks and momenta of these muons, the ATLAS experiment is equipped with the muon spectrometer (MS) [102], which is the outermost detector. It uses the magnetic

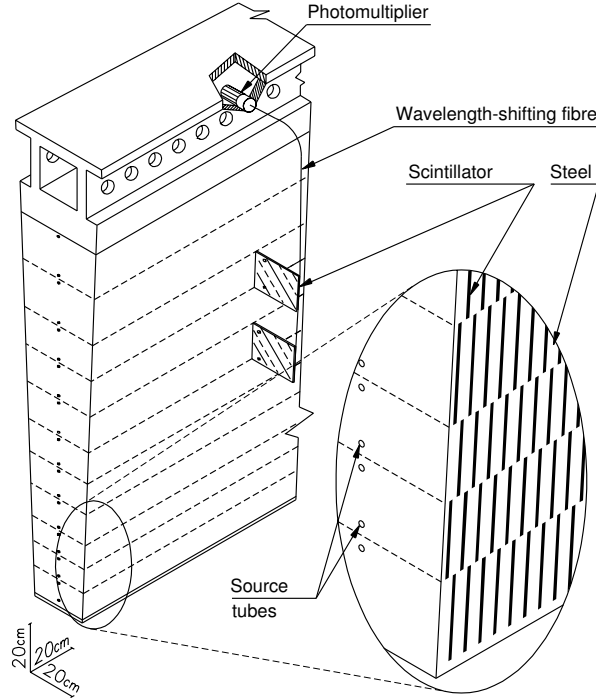


Figure 2.18: Sketch of the mechanical assembly, with the optical readout of the Tile calorimeter [79]. The various components of the optical readout (the tiles, the fibres and the photomultipliers) are shown.

deflection of muon tracks with large superconducting air-core toroidal magnets (one in the barrel and one per end-cap). It operates independently of the other ATLAS detectors, featuring its own magnetic, tracking and trigger systems. The muon system is depicted in Figure 2.19.

The MS operates up to $|\eta| < 2.7$, with different technologies. It is composed of three cylindrical layers, with radii of 5 m, 7.5 m and 10 m, in the barrel, and four disks at approximately 7, 10, 14, and 21–23 m from the interaction point, in each end-cap. Up to $|\eta| < 2$, three layers of monitored drift tubes (MDTs) chambers provide precise measurements of muon tracks in the principal direction of the magnetic field. The MDTs are made of 6 layers of straw tubes filled with a gas mixture, which is ionised when traversed by a muon. The produced electrons are collected by a wire at the centre of the tube and transformed into a measurable signal. In the forward region, for $2 < |\eta| < 2.7$, multiwire proportional chambers with cathodes segmented into strips called cathode strip chambers (CSCs) are used, due to the higher flux of particles closer to the beam pipe.

Two additional sub-detectors are included and form the muon trigger system: resistive plate chambers (RPCs) in the central region ($|\eta| < 1.05$) and thin gap chambers (TGCs) in the end-cap region ($1.05 < |\eta| < 2.4$). It selects candidates with transverse momentum p_T above a certain threshold, while requiring spatial and temporal coincidence on the hits in the RPCs and TGCs.

2.4.6 The magnet system

The ATLAS magnet system, depicted in Figure 2.20, consists of four large superconducting magnets, which measure the charge and momentum of charged particles by bending their tracks. The system spans a total of 22 m in diameter and 26 m in length, with a

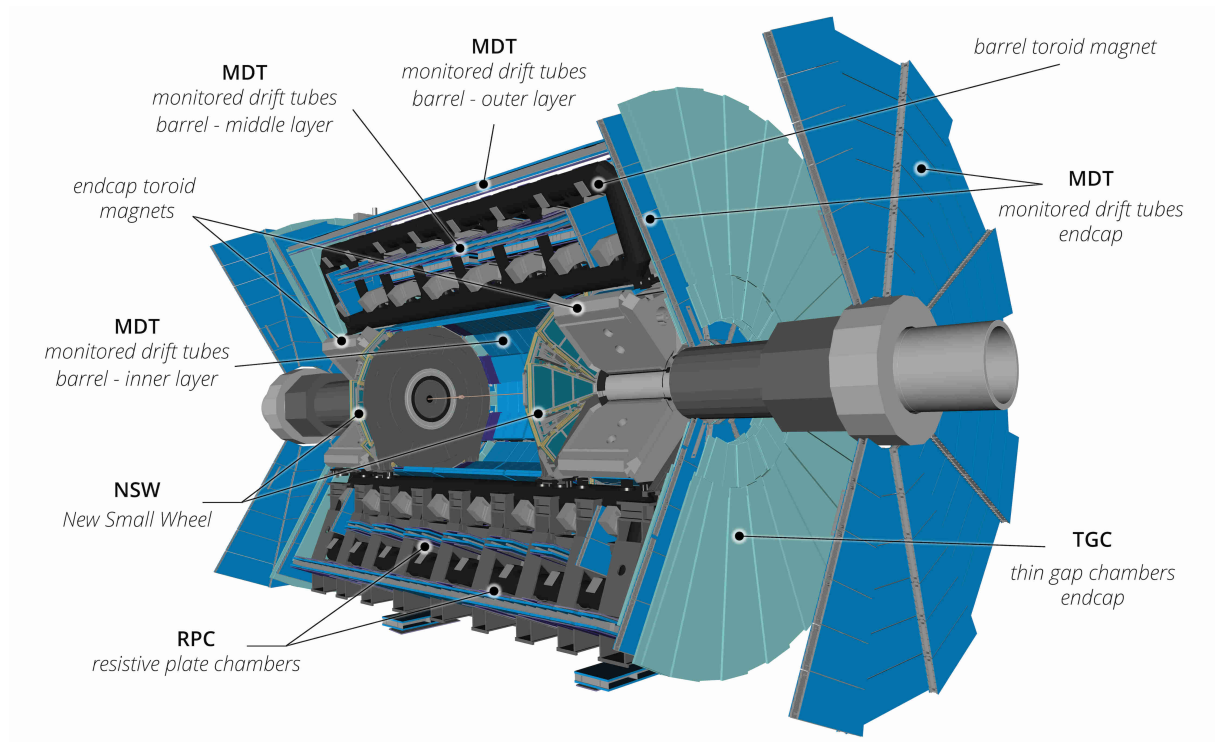


Figure 2.19: Representation of the muon system [90].

stored energy of 1.6 GJ, and is cooled to approximately 4.5 K to utilise superconductivity.

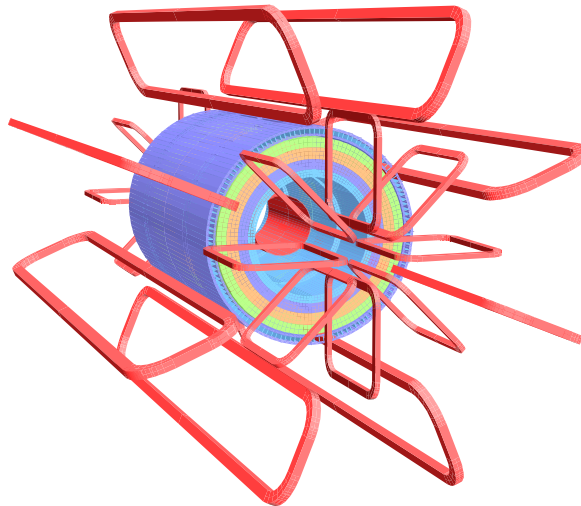


Figure 2.20: ATLAS magnet system [79].

One solenoid magnet surrounds the Inner Detector, providing a 2 T field along the beam direction (the magnetic field map is shown in Figure 2.21). It is constructed from an Al-Cu-NbTi coil with a current of 7.73 kA, running through 9 km of superconducting wire, with a stored energy of 38 MJ.

Three toroid magnets, one in the barrel and two in the end-caps, each with eight coils each, provide the necessary bending power for the muon spectrometer. These magnets also use an Al-Cu-NbTi coil, with a 20.5 kA current. The central toroid alone contains 56 km of superconducting wire. These magnets generate a magnetic field of up to 2.5 T

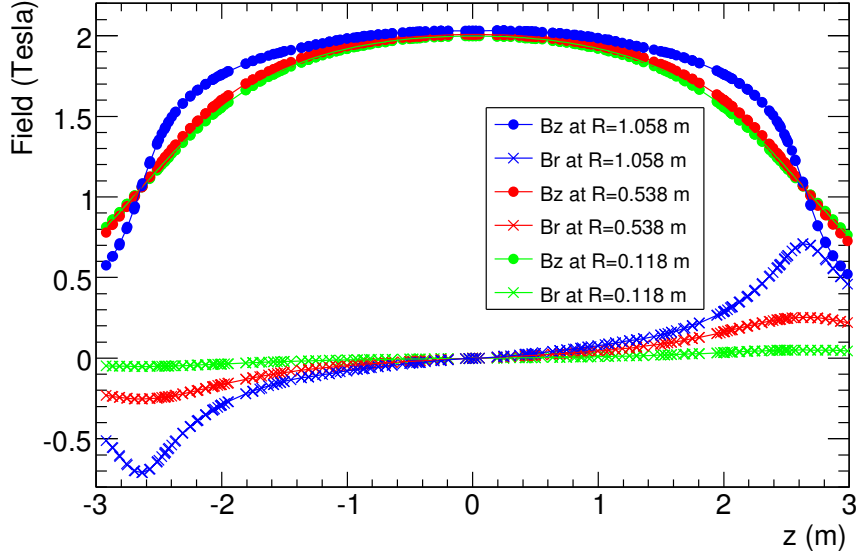


Figure 2.21: ATLAS solenoid magnetic field map at $\phi = 20\pi/16$ [79].

in the radial direction for the barrel and 3.5 T for the end-caps.

2.4.7 Trigger and data acquisition system

The very high rate of pp collisions (40 MHz, corresponding to 120 TB/s) makes it impossible to record and reconstruct every event, especially since many collisions, such as elastic interactions where protons collide without converting energy into mass, are not of interest. A real-time selection process, utilising hardware and software triggers, is employed to reduce the recorded event rate to the kHz level. This process is managed by the Trigger and Data Acquisition (TDAQ) system [103], which consists of three distinct stages: L1, L2, and the event filter, with the latter two forming part of the High-Level Trigger (HLT). Each stage progressively refines the selections made by the previous one by applying increasingly specific criteria.

The first-level (L1) trigger reduces the data rate by a factor of approximately 400, bringing it down to 100 kHz. Each event is either accepted or rejected within 2.5 μ s. The L1 Calorimeter Trigger (L1Calo) selects events containing an electron, photon, jet, or hadronically decaying τ -lepton with energy or momentum above a defined threshold, or events with significant transverse missing energy (E_T^{miss}). The L1 Muon Trigger (L1Muon) retains events with muons that surpass specific transverse momentum thresholds.

If accepted by the L1 trigger, events are forwarded to the High-Level Trigger (HLT), which comprises a Level-2 trigger and an event filter. *Online* algorithms are used to reconstruct events either in the full detector volume or in Regions of Interest (RoIs), which are localised areas identified by the L1 trigger as containing interesting features. The L2 trigger uses coordinates, energy, signature types, and RoI information to further reduce the event rate to 3.5 kHz, with an event processing time of approximately 40 ms. The event filter then uses reconstructed events and offline analysis techniques to bring the rate down to acceptable levels (around 1 kHz for Run 2 and 3 kHz for Run 3) within a few seconds, before sending the selected events to permanent storage in the Data Logger.

In addition to the selection criteria, prescales are employed, which define the number

of times a trigger must be fired before an event is recorded. Prescales are crucial for managing the data rate, especially for high-frequency events, such as those involving low-energy photons. For such high-rate events, prescales can reach values of 1000 or higher to prevent the system from being overwhelmed.

A “trigger chain” in ATLAS refers to the combination of an L1 trigger and an HLT, designed to identify specific physics signatures, passing certain conditions on energy and momentum. The structure follows this format [104]:

[Trigger level]_[object multiplicity][object type][min E_T value in GeV] and, optionally, with an optional string specifying additional requirements. The trigger level can be L1 or HLT. Object type can be e.g. “EM” for L1 EM clusters, “ g ” for HLT photons, and “ e ” for HLT electrons. Each HLT is linked to a corresponding L1 trigger, referred to as its “seed”. In Run 2, the associated L1 trigger is only mentioned if it differs from the default; however, in Run 3, it is consistently appended. The L1 trigger may specify a η -dependence of the E_T -threshold (denoted by the letter “V”), a hadronic veto (“H”), or an EM isolation requirement (“I”).

As an example, since photon triggers are crucial in this thesis, a specific photon trigger chain is detailed: `HLT_g140_loose_L1EM22VHI`. The L1 seed is `L1EM22VHI`. For an event to be selected and sent to the HLT, at least one 22 GeV EM RoI must be detected by the L1 trigger, with the following three conditions:

- V: Refers to a small variation of the energy threshold based on $|\eta|$, accounting for differences in detector calibration across regions. This variation can adjust the baseline energy threshold by -2 to $+3$ GeV.
- H: To reject hadronic activity, electron and photon candidates are discarded if the sum of transverse energy in the hadronic towers matched to the 2×2 central region in the EM calorimeter exceeds 1 GeV and surpasses the expression $E_T/23.0 - 0.2$ GeV. This criterion applies only to seeds with $E_T < 50$ GeV.
- I: An EM isolation requirement is imposed to reject electron and photon candidates if the transverse energy found in the 12 EM towers surrounding the 2×2 central region exceeds 2 GeV and is larger than $E_T/8.0 - 1.8$ GeV. This requirement is also limited to seeds with $E_T < 50$ GeV.

Another important L1 seed used in this study is `L1eEM26M`, written using the new naming conventions of Run 3 triggers. It is approximately equivalent to `L1EM22VHI`. The indicated energy threshold corresponds to the midpoint of the turn-on curve, and the M refers to a medium isolation requirement, which closely matches the VHI condition.

At the HLT level, at least one photon with $E_T > 140$ GeV is selected, which must satisfy the “loose” photon identification criteria.

Chapter 3

Photon reconstruction and identification

Contents

3.1	Electron and photon reconstruction	56
3.1.1	Energy measurement in the LAr calorimeter	56
3.1.2	Clustering algorithms	59
3.1.3	Track reconstruction	63
3.1.4	Track-cluster matching and photon conversion reconstruction	64
3.1.5	Supercluster reconstruction	65
3.1.6	Electrons and photons at analysis-level	67
3.2	Electron and photon energy calibration	67
3.2.1	EM cluster energy	68
3.2.2	LAr layer calibration	69
3.2.3	Determination of the energy scale and resolution with $Z \rightarrow ee$ events	70
3.2.4	Uniformity corrections and ADC non-linearity correction	72
3.2.5	Photon leakage correction	73
3.2.6	Energy linearity and constraints on the calibration uncertainties	74
3.2.7	Calibration cross-checks	75
3.3	Photon identification	75
3.3.1	Photon identification optimisation	75
3.3.2	Photon identification efficiency	77

Particles produced in pp collisions are reconstructed in ATLAS by processing the raw detector signals into physics objects using a series of dedicated algorithms. The reconstruction primarily relies on two types of inputs: *tracks*, formed from hits in the ID and the MS, and *clusters*, built from energy deposits in the calorimeter cells.

This chapter presents an overview of the reconstruction and identification procedures for photons.

3.1 Electron and photon reconstruction

Due to the similarity of their EM showers, photons and electrons are reconstructed using a common framework. Two categories of photons are distinguished and treated separately: *converted* photons, which undergo pair production before reaching the calorimeter, and *unconverted* photons, which do not. Converted photons are reconstructed by associating a cluster with a conversion vertex, typically formed by an electron–positron pair. In contrast, unconverted photons are identified as clusters without any matching tracks or associated vertices. Electrons are reconstructed by matching a calorimeter cluster to a charged-particle track originating from the interaction point.

3.1.1 Energy measurement in the LAr calorimeter

The electronic readout of the ATLAS LAr calorimeters is divided into two subsystems: the Front End (FE) and the Back End (BE) systems [105], as illustrated in Figure 3.1. When charged particles from an electromagnetic shower traverse the liquid argon gap between two absorber plates under high voltage, they ionise the medium, generating a triangular-shaped current pulse. This signal is processed by the FE system (Figure 3.2), where it is preamplified, shaped, and sampled by Front End Boards (FEBs). The LAr calorimeter is finely segmented into 182 468 channels, with each FEB handling 128 channels.

The preamplifiers amplify the raw signal to suppress electronic noise, and three versions of preamplifiers are used, each with different input impedance and maximum current specifications. The signal is then shaped by a bipolar $CR - (RC)^2$ shaping function, with a time constant $\tau = RC = 13$ ns, which ensures a null-integral output (see Figure 3.3). The shaping process helps to reduce the effect of out-of-time pile-up on average. The signal is subsequently sampled every 25 ns (corresponding to the LHC bunch crossing frequency of 40 MHz) by switched-capacitor array (SCA) pipeline chips using a 12-bit Analogue to Digital Converter (ADC). A number N_{samples} of samples are taken per channel, typically 4 in the EM calorimeter during Run 2, which ensures a good balance between noise suppression and real-time event processing capacity. Using multiple samples, rather than a single one, allows for better estimation of the signal amplitude, independent of the sample’s position on the pulse.

The sampled signal is stored until the L1 trigger makes a decision. Additionally, the FE system contains calibration boards to inject precision calibration signals, layer sum boards to produce analogue sums for the L1 trigger system, and control boards to distribute the 40 MHz LHC clock and other configuration and control signals.

The sampled signal is then transmitted to the Back End (BE) system via optical fibres operating at 1.6 Gbps. In the BE system, Read Out Driver (ROD) boards reconstruct the signal pulse amplitude (A) in ADC counts and determine the time offset of the energy deposition (t) by applying a digital filter to the recorded samples (s_j):

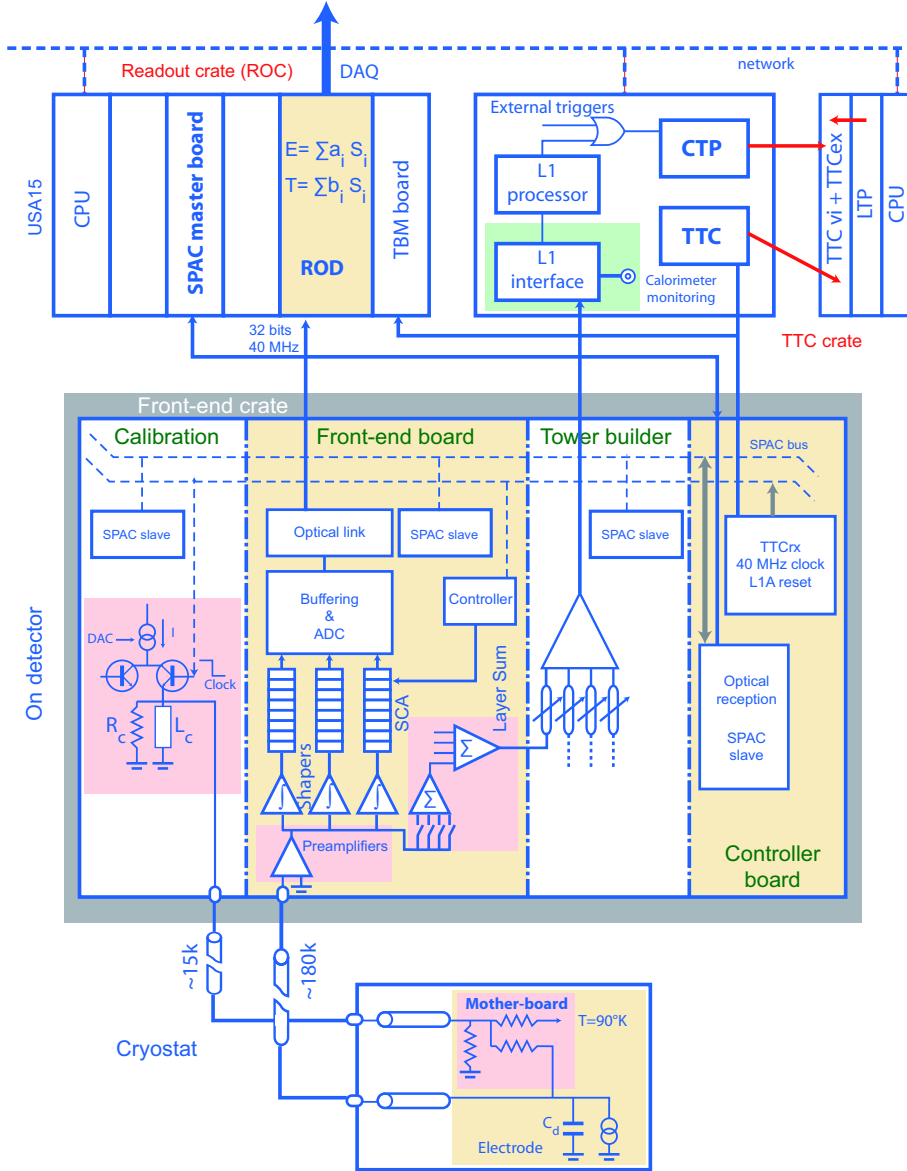


Figure 3.1: Overall architecture of the LAr readout electronics, for the EM calorimeters [106]. The LAr detectors are positioned at the bottom, and the signal proceeds upwards, through the Front End crates, then the Back End system.

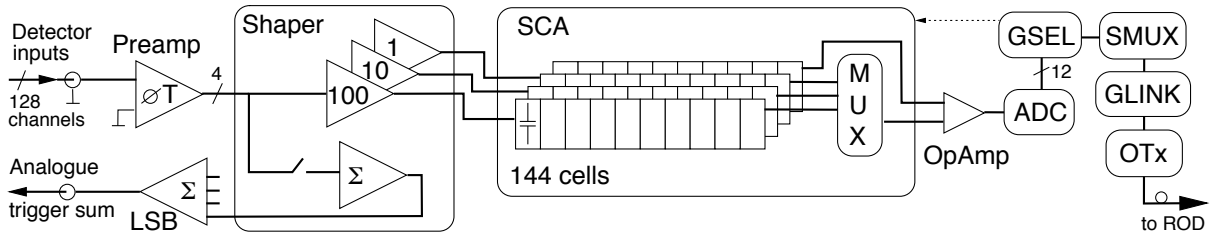


Figure 3.2: FEB architecture [107]. The data flow is depicted for four of the 128 channels per board.

$$A = \sum_{j=1}^{N_{\text{samples}}} a_j (s_j - p) \quad (3.1)$$

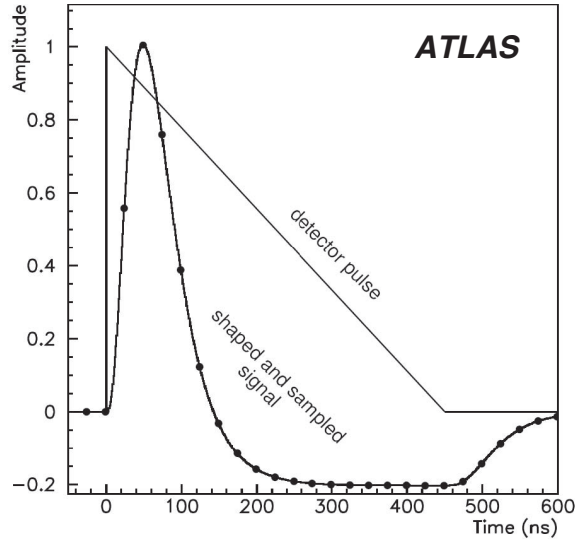


Figure 3.3: LAr current pulse before and after the shaper [106]. The dots denote ideal positions of samples separated by 25 ns.

and

$$t = \frac{1}{A} \sum_{j=1}^{N_{\text{samples}}} b_j (s_j - p), \quad (3.2)$$

where p represents the electronic pedestal (the mean ADC count when no signal is present), and a_j and b_j are coefficients derived from an optimal filtering procedure designed to maximise energy and timing resolution, known as Optimal Filtering Coefficients (OFC) [108]. Assuming the signal shape is known, it can be modelled as the convolution of the calorimeter's impulse response with the drift current induced by the ionisation in liquid argon. Under the assumption of zero-mean noise, constraints on a_j and b_j can be derived. By minimising the variances of the amplitude and time estimators under these constraints, the optimal coefficients are obtained.

The deposited energy E is computed from the reconstructed pulse amplitude A , according to the equation:

$$E = F_{\mu\text{A} \rightarrow \text{MeV}} \times F_{\text{DAC} \rightarrow \mu\text{A}} \times \frac{1}{\frac{M_{\text{phys}}}{M_{\text{cali}}}} \times \sum_{j=(0,1)}^{N_{\text{ramps}}} G_j A^j. \quad (3.3)$$

$F_{\mu\text{A} \rightarrow \text{MeV}}$ is the conversion factor that relates the ionisation current to the deposited energy. Determined through test beam experiments [109], it depends on parameters such as the calorimeter's sampling fraction, which represents the fraction of the energy deposited in the absorbers, which is not directly measured. $F_{\text{DAC} \rightarrow \mu\text{A}}$ is the conversion factor that translates the digital DAC counts into the injected current. It is derived from the known characteristics of the calibration boards and the injection resistors. $\frac{M_{\text{phys}}}{M_{\text{cali}}}$ is a factor that quantifies the difference in the response between a calibration signal and a physical ionisation pulse, both having the same input current. G_j is the electronic gain of the channel, determined through calibration runs. For medium and low gains, the sum over j starts at 0, and for high gains, it starts at 1. The ramp fit polynomial order

N_{ramps} is typically set to 1 for a linear fit, but can be increased to account for non-linear behaviour.

The measured cell energies serve as the fundamental building blocks for constructing clusters, a process that will be described in the following section.

3.1.2 Clustering algorithms

As a particle traverses the calorimeters, it deposits energy across multiple cells. Two primary algorithms have been employed to group these cells into a 3D collection, known as a cluster, corresponding to a single object: the *sliding-window* algorithm and the *dynamic topological cell clustering* algorithm. The sliding-window method was employed until 2016 and has since been replaced by the topological clustering algorithm. As this thesis makes use of data from both Run 2 and Run 3, both clustering approaches are described.

3.1.2.1 Sliding-window algorithm

Two types of sliding-window clusters are constructed: electromagnetic and combined. EM clusters are utilised for electrons and photons, referred to as *egamma* objects. This is justified by the fact that the transverse development of EM showers is well characterised by the Molière radius R_M , which contains 90% of the energy. Combined clusters are used for jet finding and tau identification. In the following, the sliding-window algorithm for EM clusters is described, as egamma objects are a central focus of this thesis. The sliding-window algorithm operates in three steps: tower building, precluster (seed) finding, and cluster filling [110].

Tower building The $\eta \times \phi$ space of cells (for $|\eta| < 2.5$) in the EM barrel and end-caps is divided into $N_\eta \times N_\phi$ (200×256) elements of size $\Delta\eta \times \Delta\phi$ (0.025×0.025). For each of these elements, the cell energy in the three layers and the presampler is summed into a *tower*. These towers serve as intermediate objects but are not part of the final reconstructed object.

Precluster (seed) finding A scan is performed over the grid using a window of size $N_\eta^{\text{window}} \times N_\phi^{\text{window}}$ (5×5), searching for local maxima in transverse energy. The window transverse energy is defined as the sum of the transverse energy of the towers within the window. If a local maximum with transverse energy $E_T > E_T^{\text{thresh}}$ (2.5 GeV) is found, a precluster is created. The window size and the threshold value are optimised to maximise the efficiency of finding preclusters while minimising the fraction of fake preclusters caused by noise.

The position of the precluster is then computed as the energy-weighted η and ϕ barycentres of each cell within a fixed-size window around the central tower of the sliding window. This window, of size $N_\eta^{\text{pos}} \times N_\phi^{\text{pos}}$ (3×3), is smaller than the one used to identify the central tower, to reduce sensitivity to noise. Special care is taken at the calorimeter edges, where the tower with the maximum energy may be located at the boundary.

Next, duplicate preclusters are removed: if two preclusters have positions within $\Delta\eta_{\text{dupl}} \times \Delta\phi_{\text{dupl}}$ (2×2), the one with the lower transverse energy is discarded.

Cluster filling Finally, clusters are constructed from the preclusters. Cells within a window of size $N_{\eta}^{\text{cluster}} \times N_{\phi}^{\text{cluster}}$, centred on a layer-dependent seed position, are included in the cluster. The processing is done layer by layer, as summarised in Table 3.1. The middle layer is processed first, using the precluster barycentre $(\eta_{\text{precl}}, \phi_{\text{precl}})$ as the seed. Next, the strips layer is processed, using the barycentre $(\eta_{\text{middle}}, \phi_{\text{middle}})$ of the middle layer cells as the seed position. The presampler is then processed, with the barycentre $(\eta_{\text{strips}}, \phi_{\text{strips}})$ of the strips layer cells as the seed. Finally, the back layer is processed using the cells from the middle layer.

Order	Layer	$\Delta\eta_{\text{cl}}$ (units of 0.025)	$\Delta\phi_{\text{cl}}$ (units of 0.025)	Seed
1	Middle (L2)	$N_{\eta}^{\text{cluster}}$	$N_{\phi}^{\text{cluster}}$	$(\eta_{\text{precl}}, \phi_{\text{precl}})$
2	Strips (L1)	$N_{\eta}^{\text{cluster}}$	6 or 8	$(\eta_{\text{middle}}, \phi_{\text{middle}})$
3	PS (L0)	$N_{\eta}^{\text{cluster}}$	6 or 8	$(\eta_{\text{strips}}, \phi_{\text{strips}})$
4	Back (L3)	$N_{\eta}^{\text{cluster}} + 1$	$N_{\phi}^{\text{cluster}}$	$(\eta_{\text{middle}}, \phi_{\text{middle}})$

Table 3.1: Summary of the cells included in the cluster. Column 1 gives the order the layers are processed. Columns 3 and 4 give the size $\Delta\eta_{\text{cl}} \times \Delta\phi_{\text{cl}}$ of the window centred on the position given in column 5, in which the cells are included in the cluster.

The possible cluster sizes are listed in Table 3.2. They must be sufficiently large to capture most of the energy deposited by the particle and mitigate fluctuations in the lateral development of the shower, while being small enough to minimise noise. The cluster size varies depending on the hypothesised particle and its location in the calorimeter. In the ϕ direction, the size is larger for electrons and converted photons due to the bending of their trajectories by the magnetic field (with electron-positron pairs in the latter case), resulting in broader energy deposits. In the end-cap region, the same ϕ size is maintained since the electromagnetic field is weaker, but a larger size is used in η to account for the smaller physical cell size.

Particle type	Barrel	End-cap
Electron	3×7	5×5
Converted photon	3×7	5×5
Unconverted photon	3×5	5×5

Table 3.2: Cluster size $N_{\eta}^{\text{cluster}} \times N_{\phi}^{\text{cluster}}$ for different hypothesized particles and regions of the detector.

The sliding-window algorithm performance for simulated electrons is shown in Figure 3.4.

3.1.2.2 Dynamical topological cell clustering

Reconstructing electrons and photons with variable-size clusters provides several benefits compared to the earlier fixed-size approach. Initially, fixed-size clusters were adopted due to limitations in the calibration techniques, which could not be reliably applied to irregularly shaped clusters. With the advent of advanced multivariate calibration methods [112], it is now possible to accurately calibrate dynamically sized topological clusters, or *topo-clusters*. These developments have enabled the use of topo-clusters in the reconstruction of egamma objects with significantly improved accuracy [113].

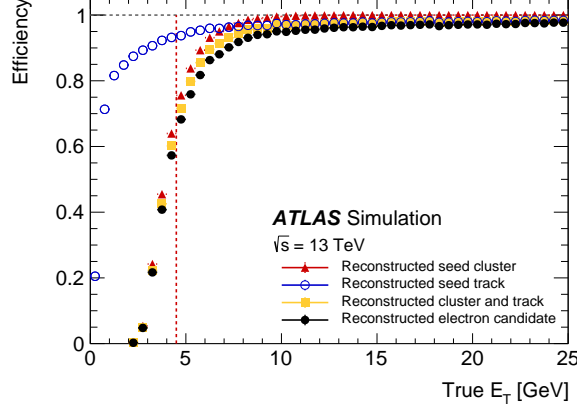


Figure 3.4: Total reconstruction efficiency for simulated electrons as a function of the true generated energy [111]. The efficiencies for the different steps of the reconstruction are depicted, in red for the seed cluster. Since the cluster reconstruction needs uncalibrated cluster seeds with $E_T > 2.5$ GeV, the total reconstruction efficiency is less than 65 % below 4.5 GeV (red dashed line). The seed cluster efficiency ranges from 65 % at $E_T = 4.5$ GeV, to 96 % at $E_T = 7$ GeV, to more than 99 % above $E_T = 15$ GeV.

Firstly, this approach enables the recovery of low-energy photons generated through Bremsstrahlung processes by electrons interacting with material located upstream of the calorimeters. The algorithm constructs a *supercluster* (see Figure 3.5) [114], linking the primary energy deposit of the electron (the seed cluster) to additional deposits from Bremsstrahlung photons (satellite clusters). This method provides a significant improvement in energy resolution and allows for the reconstruction of showers down to approximately $\mathcal{O}(100 \text{ MeV})$, for electrons and converted photons.

Another benefit of this approach is its alignment with the natural development of the shower, the cluster’s growth being similar to the shower’s development. The topo-clusters grow from the highest-energy cell outwards, capturing the full extent of the EM shower.

This technique also enables the capture of more energy compared to the sliding-window algorithm, thereby reducing the need for additional corrections.

The construction of topo-clusters relies on a quantity called *cell significance* [115], defined as

$$\zeta_{\text{cell}}^{\text{EM}} = \left| \frac{E_{\text{cell}}^{\text{EM}}}{\sigma_{\text{noise,cell}}^{\text{EM}}} \right|, \quad (3.4)$$

where $|E_{\text{cell}}^{\text{EM}}|$ is the absolute energy at the scale of the EM calorimeter, and $\sigma_{\text{noise,cell}}^{\text{EM}}$ is the expected cell noise coming from pile-up ($\sigma_{\text{noise}}^{\text{pile-up}}$) and from the electronic noise ($\sigma_{\text{noise}}^{\text{electronic}}$).

The topo-cluster reconstruction algorithm initiates by identifying cells with substantial energy, based on a set of noise thresholds, and subsequently attaches neighbouring cells to build the final cluster.

Initially, *proto-clusters* are seeded within the LAr calorimeter using cells from the second (L2) and third (L3) layers, requiring $\zeta_{\text{cell}}^{\text{EM}} > 4$. Cells from the PS and L1 layers are excluded to prevent the formation of noise-induced clusters.

Next, neighbouring cells with $\zeta_{\text{cell}}^{\text{EM}} \geq 2$ are iteratively collected from the eight adjacent cells in the same sampling layer, as well as from contiguous cells in other layers in contact

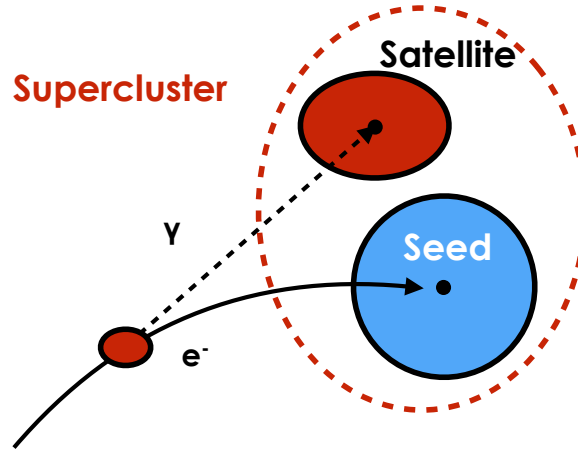


Figure 3.5: Representation of a supercluster showing a seed electron cluster and a satellite photon cluster [114].

with the proto-cluster. Cells with the highest energies are prioritised, enabling clusters to form around more energetic cells. If cells are shared between proto-clusters, they are merged initially, with later steps available to separate them if necessary. Each newly added cell becomes a seed for the next iteration.

Finally, neighbouring cells with $\zeta_{\text{cell}}^{\text{EM}} > 0$ are incorporated into the cluster. This entire process is commonly referred to as the “4-2-0” topo-cluster reconstruction.

After the initial topo-cluster collection, clusters containing two or more local maxima are divided. A local maximum is defined as a cell with $E_{\text{cell}}^{\text{EM}} > 500 \text{ MeV}$ that has at least four neighbouring cells, none with a larger signal. Following the division, a cell may be shared between proto-clusters that have been separated into distinct topo-clusters. A fractional weight is then allocated to the cell for each cluster, determined by both the clusters’ respective energies, $E_{\text{clus},1}^{\text{EM}}$ and $E_{\text{clus},2}^{\text{EM}}$, and the distances from the cell to the respective centres of gravity of the two clusters (d_1, d_2):

$$w_{\text{cell},1} = \frac{E_{\text{clus},1}^{\text{EM}}}{E_{\text{clus},1}^{\text{EM}} + r E_{\text{clus},2}^{\text{EM}}}, \quad (3.5)$$

where $r = \exp(d_1 - d_2)$, and the weight for the cell in the second split cluster is given by:

$$w_{\text{cell},2} = 1 - w_{\text{cell},1} = \frac{r E_{\text{clus},2}^{\text{EM}}}{E_{\text{clus},1}^{\text{EM}} + r E_{\text{clus},2}^{\text{EM}}}. \quad (3.6)$$

In practice, topo-clusters are constructed using cells from various regions of the detector, specifically from the LAr and Tile calorimeters. The same initial proto-clusters are employed to reconstruct other ATLAS objects, such as jets and τ leptons. This approach facilitates future enhancements to topo-cluster-based isolation calculations [116, 117] and enables the implementation of a detector-wide particle flow reconstruction algorithm [118, 119].

Photon and electron reconstruction begins from topo-clusters, using only the energy deposited in cells of the EM calorimeter. An exception is made in the transition region ($1.37 < |\eta| < 1.63$), where a significant fraction of the energy from electrons and photons is deposited in the PS and scintillator. In other detector regions, these contributions are

excluded to minimise noise. This is defined as the EM energy, and the ratio of this EM energy to the total cluster energy is referred to as the EM fraction:

$$f_{\text{EM}} = \frac{E_{\text{L1}} + E_{\text{L2}} + E_{\text{L3}} + w(E_{\text{PS}} + E_{\text{E4}})}{E_{\text{clus}}}, \quad w = \begin{cases} 1, & \text{if } 1.37 < |\eta| < 1.63 \\ 0, & \text{otherwise} \end{cases} \quad (3.7)$$

where $E_{\text{L}x}$ represents the cluster energy in layer x . The term $(E_{\text{PS}} + E_{\text{E4}})$ applies solely to the crack region. E_{clus} is the total cluster energy, comprising cells from both the electromagnetic and hadronic calorimeters, including the presampler and scintillators.

Only clusters with EM energy larger than 400 MeV are considered. The distribution of f_{EM} is shown in Figure 3.6a, and the electron reconstruction efficiency for different cuts on f_{EM} is shown in Figure 3.6b. To reconstruct photons and electrons, topo-clusters are required to have $f_{\text{EM}} > 0.5$. This threshold is chosen as it rejects $\sim 60\%$ of pile-up clusters without impacting true electron topo-clusters selection¹.

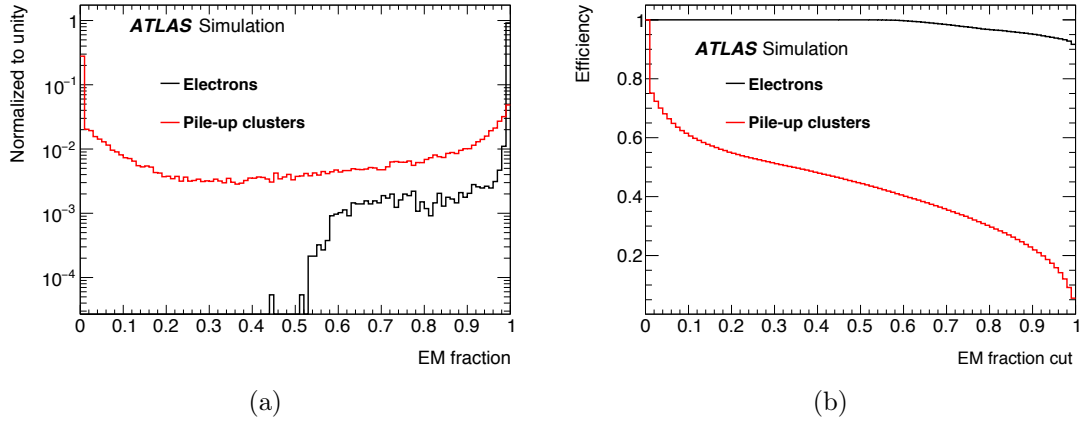


Figure 3.6: (a) f_{EM} distribution and (b) reconstruction efficiency as a function of the f_{EM} selection cut for simulated true electron (black) and pile-up (red) clusters [115].

3.1.3 Track reconstruction

The reconstruction of charged particle trajectories [111] relies on hits in the ID, which form the foundational elements of this process. A schematic illustration of the path of an electron through the detector is shown in Figure 3.7. While photons have no charge, they can produce electron-positron pairs, resulting in signals within the ID. Hits are grouped into clusters, and then three-dimensional measurements, known as *space-points*, are derived from these clusters. In the Pixel Detector, each cluster generates a single space-point, whereas in the SCT, clusters are combined to achieve a three-dimensional measurement. A track seed is constructed using three space-points from the silicon detector layers.

Track reconstruction involves three stages: pattern recognition, ambiguity resolution, and TRT extension [120].

For pattern recognition, an initial hypothesis is made regarding the particle type, assuming it to be a pion. If a track seed with $p_{\text{T}} > 1$ GeV cannot be extended to form a

¹In the transition region, some topo-clusters failing to pass the requirement on f_{EM} are also selected if they satisfy $E_{\text{T}} > 1$ GeV, to increase the reconstruction efficiency in that region.

track candidate with at least seven hits across different silicon layers, and if the associated EM cluster meets specified criteria for shower width and depth, a second reconstruction attempt is performed under the electron hypothesis. This second approach permits up to 30 % energy loss due to Bremsstrahlung at each interaction point between the track and detector material.

Track candidates with $p_T > 400$ MeV are subsequently fitted using the ATLAS Global χ^2 Track Fitter [121], based on the same particle hypothesis as in the pattern recognition step. At this stage, ambiguities from track candidates sharing hits are also resolved. If a fit fails under the pion hypothesis and the angular separation from the EM cluster is below a certain threshold, the track candidate is refitted under the electron hypothesis, an extra degree of freedom being added to the χ^2 calculation, to better account for the non-linear effects of Bremsstrahlung.

An additional fitting procedure, based on an optimised Gaussian-sum filter, is applied to clusters of raw measurements to more accurately account for Bremsstrahlung energy losses. Only tracks with at least four silicon hits and that are loosely matched to EM clusters are considered. Angular separation requirements are imposed between the extrapolated track position from the perigee to the calorimeter's second layer and the barycentre of the cluster. Specifically, the polar separation must meet $|\eta_{\text{track}} - \eta_{\text{cluster}}| < 0.05$, and the azimuthal separation must satisfy either $-0.20 < \Delta\phi < 0.05$ or $-0.10 < \Delta\phi_{\text{res}} < 0.05$, with $\Delta\phi$ and $\Delta\phi_{\text{res}}$ calculated as $q(\phi_{\text{track}} - \phi_{\text{cluster}})$, where $\Delta\phi_{\text{res}}$ uses the momentum rescaled to the cluster's energy, and q is the track charge. The asymmetric condition for matching in ϕ compensates for energy loss due to Bremsstrahlung, which affects charged particles as their trajectory bends in the magnetic field.

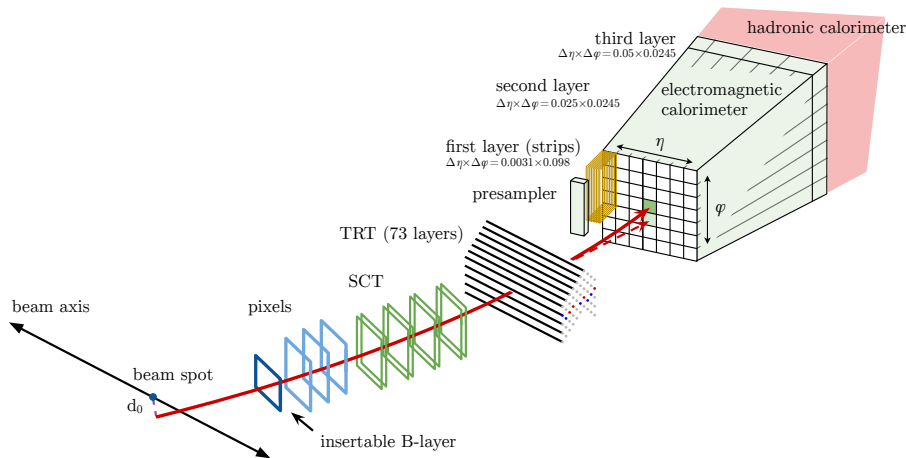


Figure 3.7: A schematic illustration of the path of an electron through the detector [111]. The red line shows the hypothetical path of an electron passing through the tracking system (pixel detectors, SCT, and TRT) and then entering the electromagnetic calorimeter. The dashed red line shows the trajectory of a photon produced by the interaction of the electron with the tracking system material.

3.1.4 Track-cluster matching and photon conversion reconstruction

Tracks are subsequently matched to EM clusters by extrapolating the track from the perigee to the second layer of the calorimeter [115]. The track momentum is either used

as measured or rescaled to match the cluster energy, improving the matching accuracy in cases of significant energy loss due to Bremsstrahlung radiation. A track is considered matched to an EM cluster if $|\Delta\eta| < 0.05$ and $-0.10 < q(\phi_{\text{track}} - \phi_{\text{cluster}}) < 0.05$, where q is the charge of the reconstructed track. This asymmetric condition accounts for the trajectory deflection caused by Bremsstrahlung radiation.

In cases where multiple tracks are matched to a single cluster, tracks with hits in the Pixel Detector are preferred over those with hits in the SCT but not in the Pixel Detector. Among competing tracks, priority is given to the one that exhibits a better match to the cluster in the second layer of the calorimeter, unless the difference is smaller than 0.01. Track extrapolation through the calorimeter is initially performed using the rescaled momentum and subsequently using the measured momentum. The extrapolation with the better match is retained unless the difference between the two is minimal. In such cases, the track with more pixel hits is selected, with additional weight given to hits in the innermost pixel layer.

Tracks loosely matched to fixed-size clusters serve as inputs for photon conversion reconstruction. Both tracks with silicon hits (*Si tracks*) and tracks reconstructed solely in the TRT (*TRT tracks*) are utilised. Two-track conversion vertices are identified by pairing oppositely charged tracks to form a vertex consistent with the decay of a massless particle. Single-track vertices are reconstructed from tracks without hits in the innermost sensitive layers of the detector. To enhance the purity of converted photons, tracks utilised for constructing conversion vertices must have a high probability of being electron tracks, as determined by the TRT [122]. The requirements are loose for Si tracks but are tight for TRT tracks involved in double-track conversions, and even tighter for those used in single-track conversions. Figure 3.8 shows the converted photon efficiencies as a function of E_T^{true} and $\langle\mu\rangle$, and the probability for an unconverted photon to be mistakenly reconstructed as a converted photon.

When multiple vertices are matched to a cluster, priority is given to double-track conversions involving two silicon tracks, followed by other double-track conversions, and lastly single-track conversions. Within each category, the vertex with the smallest conversion radius is selected.

3.1.5 Supercluster reconstruction

The supercluster reconstruction is performed in two steps: first, identifying seed clusters, and then associating satellite clusters, which may originate from Bremsstrahlung radiation or topo-cluster splitting [115].

Initially, topo-clusters are sorted in descending order of E_T using their EM energy². Each cluster is tested sequentially as a candidate for a supercluster seed. For electron superclusters, the seed cluster must have E_T greater than 1 GeV and must be matched to a track with at least four silicon hits. For photon superclusters, the seed cluster must have E_T greater than 1.5 GeV. Since photon superclusters impose no requirements on tracks or conversion vertices, the higher energy threshold helps reduce sensitivity to noise compared to electron seeds.

If a cluster satisfies the seed requirements, a search for satellite clusters is performed. For both electrons and photons, a cluster qualifies as a satellite if it lies within a window of size $\Delta\eta \times \Delta\phi = 0.075 \times 0.125$, which efficiently captures secondary EM showers originating

²An exception is made for clusters in the transition region, where a looser selection is applied. These clusters are added at the end.

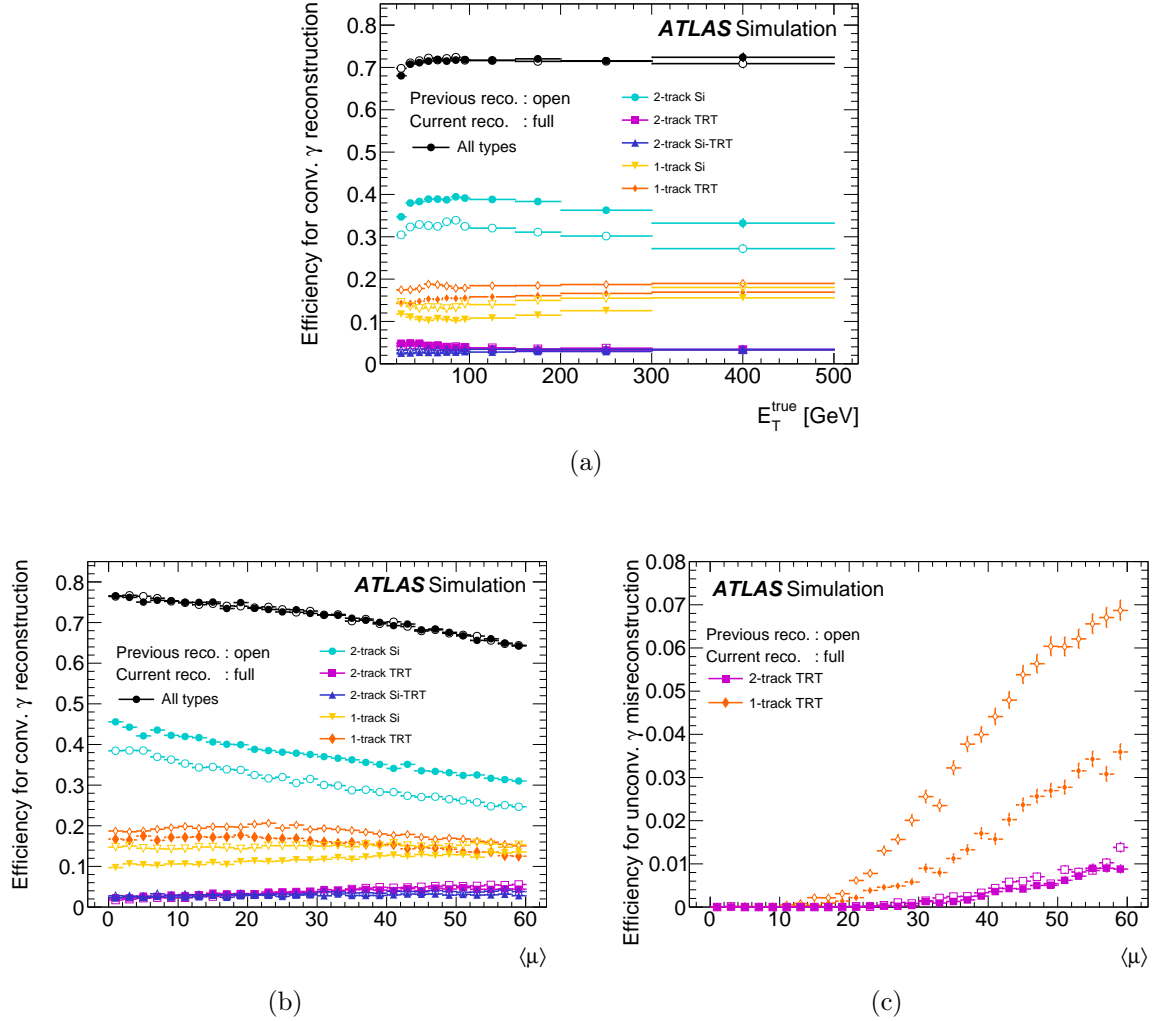


Figure 3.8: (a) Converted photon reconstruction efficiency as a function of E_T^{true} , averaged over $\langle\mu\rangle$ for a uniform $\langle\mu\rangle$ distribution between 0 and 60, (b) converted photon reconstruction efficiency as a function of $\langle\mu\rangle$, and (c) probability for an unconverted photon to be mistakenly reconstructed as a converted photon [115]. The contributions of the different conversion types are shown in each figure.

from the same initial electron or photon. For electrons, clusters located within a larger window of size $\Delta\eta \times \Delta\phi = 0.125 \times 0.300$ around the seed cluster barycentre are also added to the supercluster if their best-matched track is the same as the best-matched track of the seed cluster. For photons associated with conversion vertices built exclusively from Si tracks, additional clusters are included as satellite clusters if their best-matched (electron) track belongs to the conversion vertex matched to the seed cluster. Figure 3.9 recapitulates these cases.

The final step involves assigning calorimeter cells to superclusters. Only cells from the presampler and the first three layers of the calorimeter are included, except in the transition region ($1.4 < |\eta| < 1.6$), where energy measurements from the scintillators located between the calorimeter cryostats are also taken into account. To minimise the impact of pile-up on the sensitivity of superclusters, the size of each constituent topocluster is constrained to a maximum width of 0.075 in the barrel region and 0.125 in the end-cap region, both in the η direction. Due to the magnetic field, which is parallel to the

beam line in the ID, EM showers tend to spread in the ϕ direction. Thus, the restriction in η does not typically hinder the capture of the full electron or photon shower. No restrictions are applied in the ϕ direction.

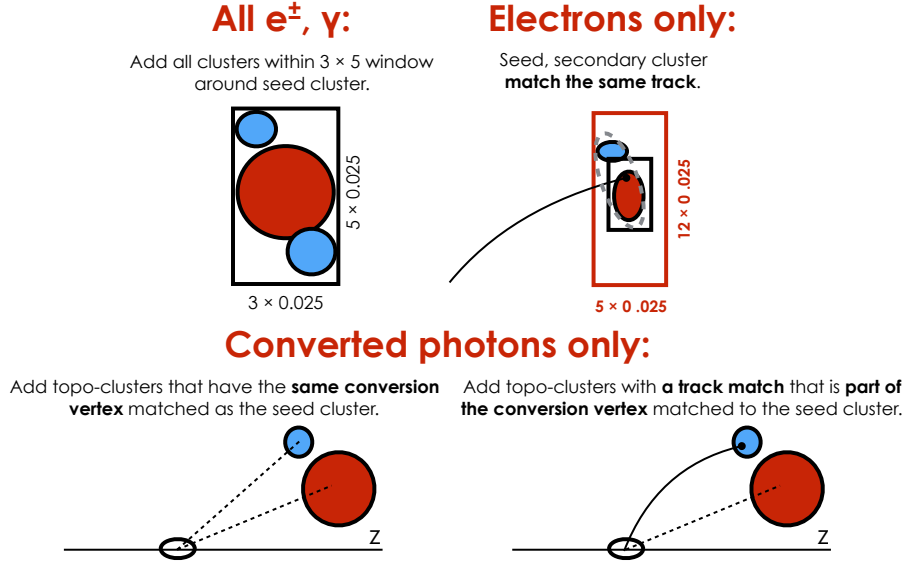


Figure 3.9: Diagram of the superclustering algorithm for photons and electrons [114]. Seed clusters are depicted in red and satellite clusters in blue.

3.1.6 Electrons and photons at analysis-level

The next step involves constructing analysis-level objects. An initial energy calibration and position correction are applied to the clusters. Tracks are matched to electron superclusters, and conversion vertices are matched to photon superclusters using the same matching criteria as for topo-clusters but applied at the supercluster level. Since electrons and photons are reconstructed independently, a single seed may generate both an electron and a photon. In such cases, the ambiguity resolution procedure described in Figure 3.10 is applied.

The primary principle is that if an object can be unambiguously identified as either an electron (a supercluster with a good track attached but no good photon conversion vertex) or a photon (a supercluster with no good track attached, or with a good conversion vertex), only one object is created. However, if ambiguity remains, both an electron and a photon object are constructed. These objects are explicitly flagged as ambiguous, with the final classification determined according to the analysis needs. Figure 3.11 shows the reconstruction efficiency for electrons.

As the energy calibration depends on the presence of matched tracks and conversion vertices, and the initial calibration is performed before the final matching, a recalibration is applied to the electron and photon objects after this step.

3.2 Electron and photon energy calibration

The steps to calibrate the energy response for electrons and photons are summarised in Figure 3.12. They are detailed below.

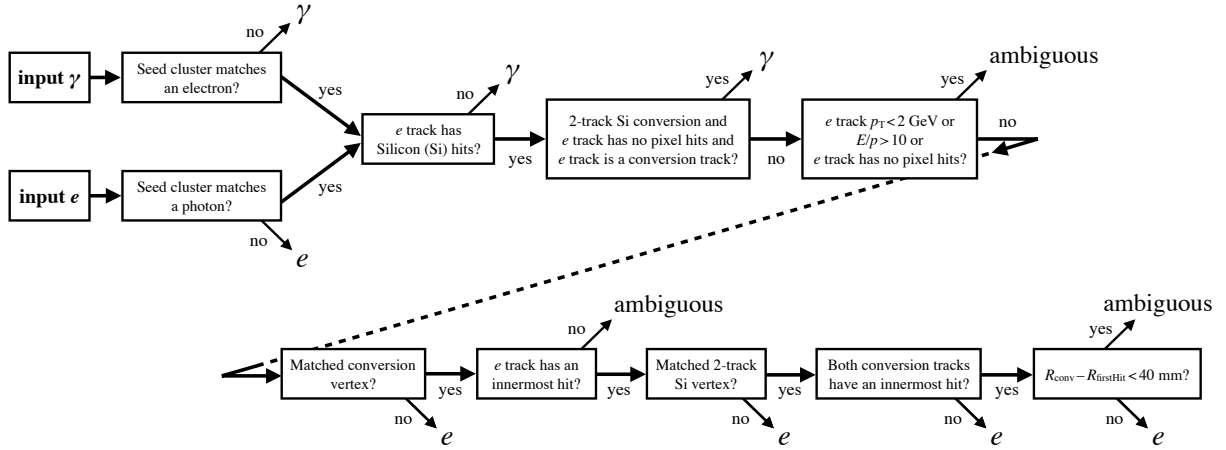


Figure 3.10: Ambiguity resolution for particles initially reconstructed both as electrons and photons. An “innermost hit” is a hit in the functioning pixel nearest to the beam-line along the track trajectory, E/p is the ratio of the supercluster energy to the measured momentum of the matched track, R_{conv} is the radial position of the conversion vertex, and R_{firstHit} is the smallest radial position of a hit in the track or tracks that make a conversion vertex [115].

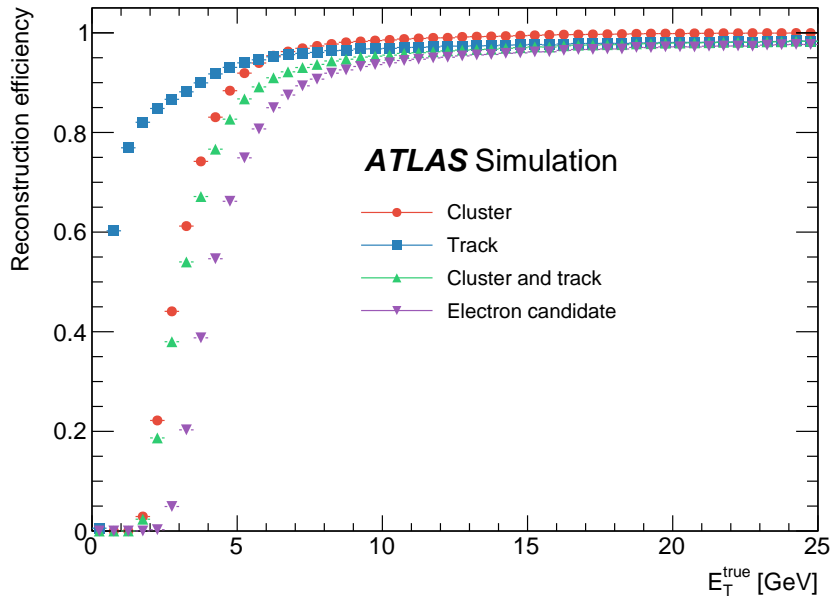


Figure 3.11: The cluster, track, cluster and track, and electron reconstruction efficiency, against the generated electron transverse energy E_T [115].

3.2.1 EM cluster energy

The energy of electrons and photons is determined based on the energy of the reconstructed cluster [123]. A correction is applied to account for energy lost in the material upstream of the calorimeter, energy deposited in cells neighbouring the cluster, and energy escaping beyond the LAr calorimeter [124]. These corrections are computed using multivariate regression algorithms.

The algorithms, trained separately for electrons, converted photons, and unconverted

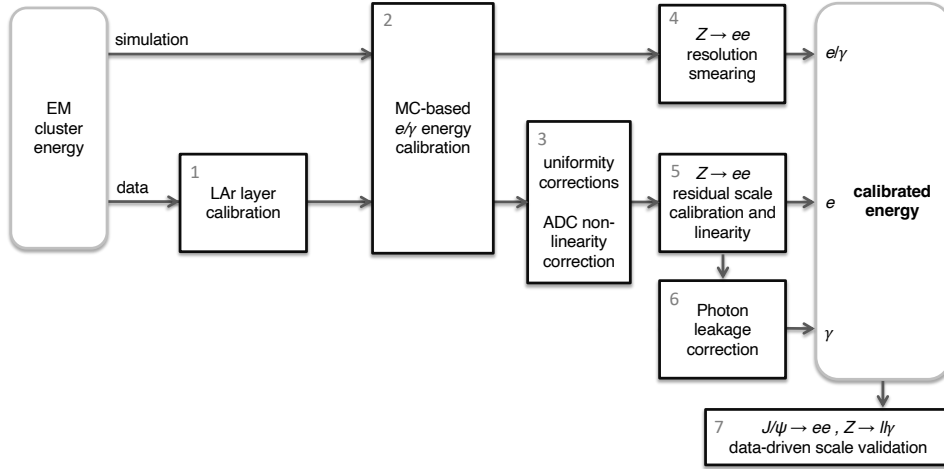


Figure 3.12: Overview of the electron and photon energy calibration [123].

photons, are optimised on simulated single particles without pile-up. They utilise Boosted Decision Trees (BDTs) and are trained in intervals of $|\eta|$ and transverse energy (E_T). The input variables for the training include the total energy deposited in the calorimeter, the energy deposited in the presampler, the ratio of energy deposits in the first and second layers of the EM calorimeter, the η coordinate of the shower's impact point in the calorimeter, and the distances in η and ϕ between the shower's impact point and the centre of the nearest cell in the second calorimeter layer.

3.2.2 LAr layer calibration

3.2.2.1 Presampler energy scale

The presampler energy scale, α_{PS} , is defined as the ratio of the presampler energies in the data to those in the simulation. Unlike in Run 1, where α_{PS} was determined using samples of electrons and photons, the current approach utilises muon candidates from a low-pile-up data sample. This choice is motivated by the insensitivity of muons to the amount of material in front of the presampler.

A high-purity sample of $W \rightarrow \mu\nu$ and $Z \rightarrow \mu\mu$ events is used for this calibration. Selection criteria include one or two isolated muons, with extra requirements on transverse energy and invariant mass. For the Z -boson sample, an additional condition on the dimuon invariant mass is imposed to ensure purity.

3.2.2.2 Intercalibration of the first and second calorimeter layers

The intercalibration of the first and second layers of the EM calorimeter ensures control over the linearity of the electron and photon energy responses. Unlike the approach used

in Run 1 [124], the current method employs muon and electron candidates from Z -boson decays to determine the relative calibration of the two layers.

The intercalibration using muons involves calculating the intercalibration factor, defined as $\alpha_{12} = (\langle E_1 \rangle / \langle E_2 \rangle)^{\text{data}} / (\langle E_1 \rangle / \langle E_2 \rangle)^{\text{MC}}$, where $\langle E_i \rangle$ represents the mean energy deposited in layer i . Two methods are employed and combined to estimate $\langle E_i \rangle$: the *most probable value* (MPV) method and the *truncated mean* (TM) method. The MPV method fits a Landau distribution convolved with a noise template, while the TM method calculates the mean of the distribution within a restricted window to reduce sensitivity to the tails. For both methods, $\langle E_i \rangle$ is estimated in intervals of $\langle \mu \rangle$ within the $[10, 40]$ range and extrapolated linearly to $\langle \mu \rangle = 0$. Excellent agreement is observed between the two measurements. The total uncertainty ranges from approximately 0.7% in the barrel to about 2% in the end-caps.

The intercalibration using electrons employs two methods to estimate α_{12} : one based on the ratio of the measured energy E and momentum p of electron candidates, and the other on the invariant mass distribution of electron pairs, m_{ee} . The E/p method models the distribution with a Crystal Ball function to account for the effects of Bremsstrahlung radiation caused by interactions in the material upstream of the calorimeter. In contrast, the m_{ee} method uses a convolution of a Crystal Ball function and a Breit-Wigner distribution, with a small background in data modelled by a second-order Chebyshev polynomial. In both approaches, the most probable value of the fitted Crystal Ball function serves as the estimator for the intercalibration factor.

The measurements obtained from the two-muon and two-electron methods are combined using the Best Linear Unbiased Estimator (BLUE) technique [125]. Statistical correlations between the two methods are neglected, as systematic uncertainties dominate and are largely uncorrelated across $|\eta|$. Incorporating the electron method alongside the muon measurements reduces the uncertainty by approximately a factor of 1.8 in the first half of the barrel. In the end-cap regions, however, the combined uncertainties are primarily driven by the discrepancies between the electron and muon results. Figure 3.13a shows α_{12} as a function of $|\eta|$, after the combination of the two-electron and two-muon measurements, and Figure 3.13b gives the total uncertainties.

3.2.3 Determination of the energy scale and resolution with $Z \rightarrow ee$ events

Energy scale and resolution corrections are determined using electrons from Z -boson decays. These corrections are evaluated in η regions, labelled i for the first electron and j for the second. The difference in the resonance positions between data and simulation is used to compute the invariant mass scale correction α_{ij} , defined such that $m_{ij}^{\text{corr}} = m_{ij} / (1 + \alpha_{ij})$. Similarly, the mass resolution correction is determined for simulation as $(\sigma_m/m)_{ij}^{\text{corr}} = (\sigma_m/m)_{ij} \oplus c_{ij}$. The parameters α_{ij} and c_{ij} are optimised to achieve the best agreement between the invariant mass distributions in data and simulation within each (i, j) category.

An absolute energy scale has been determined for the presampler, while only a relative response has been measured between the first and second layers of the accordion calorimeter. To establish the absolute energy scale, the overall scale of the accordion calorimeter is utilised. For electrons from Z -boson decays, the contribution from the third layer is negligible. The energy scale corrections for single electrons, defined as $E_i^{\text{acc,corr}} = E_i^{\text{acc}} / (1 + \alpha_i^{\text{acc}})$, are primarily linked to the invariant mass scale factors α_{ij}

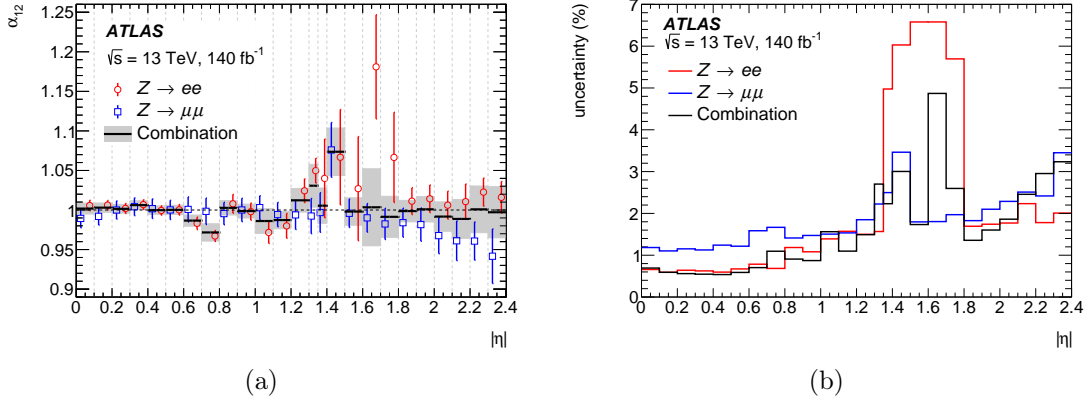


Figure 3.13: (a) Relative calibration scale factor α_{12} between the first and second layers of the EM calorimeter as a function of $|\eta|$ [123]. Open blue squares represent results derived from the study of muon energy deposits in $Z \rightarrow \mu\mu$ events, combining the *truncated mean* and *most probable value* methods. Open red circles correspond to the combined values extracted from the dependence of the dielectron invariant mass m_{ee} and the E/p ratio on E_1/E_2 in $Z \rightarrow ee$ events. The final scale factors, obtained by combining the results from both electron and muon studies, are depicted as solid black lines. (b) Corresponding total uncertainties, taking into account rescaling in the case of large χ^2 .

through the relation:

$$\alpha_{ij} = \frac{f_i^{\text{acc}} \alpha_i^{\text{acc}} + f_j^{\text{acc}} \alpha_j^{\text{acc}}}{2},$$

where f_i^{acc} , determined from simulation, quantifies the sensitivity of the total calibrated electron energy to the energy measured in the accordion calorimeter for electrons in the η bin i . The accordion energy scale corrections are applied only to the data.

The single-electron energy resolution corrections, denoted as c_i , can be extracted from the invariant mass resolution corrections c_{ij} using the relation $c_{ij} = (c_i \oplus c_j) / 2$. These corrections are subsequently applied to the reconstructed energy in the simulation.

The parameters α_i^{acc} and c_i are determined through simultaneous fits across all η regions. The α_i^{acc} values exhibit year-to-year variations due to changes in the LAr temperature and instantaneous luminosity. Year-specific energy scale corrections are provided, though the observed variations are minimal. The energy scale correction is applied uniformly across the three accordion layers, and the Z -based calibration fit is repeated to validate convergence, yielding the residual energy scale factor $\alpha_i^{\text{closure}}$.

The constant term resolution corrections, c_i , are also sensitive to the level of pile-up. A weighted average of c_i across different years is calculated to account for these variations. Systematic uncertainties in the determination of $\alpha_i^{\text{closure}}$ and c_i are evaluated by altering the event selection criteria and fitting range. Figure 3.14 presents the $\alpha_i^{\text{closure}}$ and c_i values as a function of η , separated by year.

Figure 3.15 shows the invariant mass distribution of $Z \rightarrow ee$ candidates in data and simulation once the energy scale correction has been applied to the data, and the resolution correction to the simulation. The background contamination, expected to be lower than 1 %, is not taken into account.

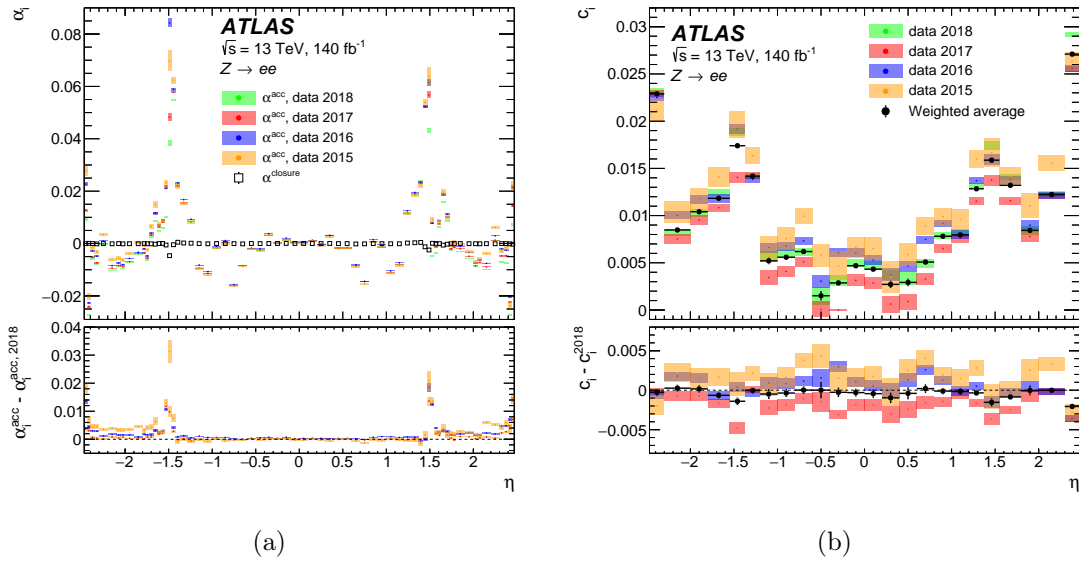


Figure 3.14: (a) Energy scale calibration factors α_i^{acc} and $\alpha_i^{\text{closure}}$, and (b) the additional constant term c_i , as a function of η [123]. The statistical uncertainties are depicted with the shaded areas. The bottom panels show the differences between (a) α_i^{acc} and (b) c_i measured in a given data-taking period and the measurement using 2018 only.

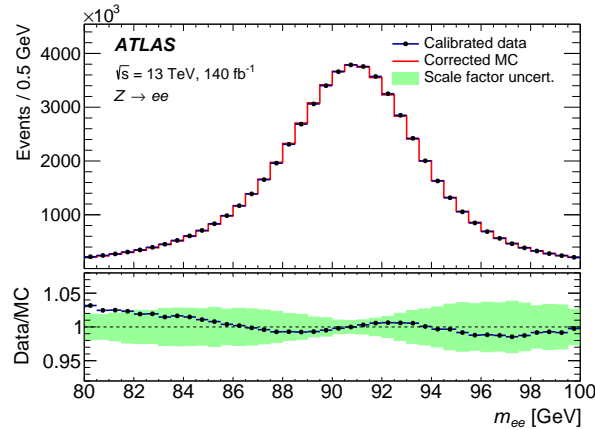


Figure 3.15: Comparison of the invariant mass distributions of electron pairs from selected $Z \rightarrow ee$ candidates in data and simulation, following the application of calibration and resolution corrections [123]. The simulation is normalised to match the total event count in the data. The bottom panel displays the data-to-simulation ratio, with the uncertainty band reflecting the effects of uncertainties in the calibration and resolution correction factors.

3.2.4 Uniformity corrections and ADC non-linearity correction

Additional corrections are applied to the data to account for response variations in specific detector regions that are not accurately modelled in the simulation. These include regions with non-optimal high-voltage settings, azimuthal non-uniformities, and biases related to the calibration of the LAr calorimeter's electronics. The stability of the calorimeter response over time and its dependence on pile-up conditions are systematically studied to ensure consistency. Furthermore, dedicated corrections are implemented to address

variations in energy reconstruction linked to the readout electronics.

Mechanical deformations caused by gravity result in variations in the size of the LAr gaps between the absorbers, leading to azimuthal-dependent changes in energy response. These variations are on the order of 0.1 %–0.2 % in the barrel region and can reach up to 1 % in the end-caps. To address these effects, energy corrections are applied in six $|\eta|$ intervals: 0 – 0.6, 0.6 – 1.0, 1.0 – 1.37, 1.37 – 1.55, 1.55 – 1.82, and 1.82 – 2.4.

The stability of the calorimeter response is evaluated using the reconstructed peak position of the dielectron invariant mass distribution, $m_{ee}/\langle m_{ee} \rangle$, in $Z \rightarrow ee$ events. Over the full 2015–2018 data-taking period, the stability is maintained at the level of 0.05 %. Corrections are derived to account for variations in the energy scale due to changes in the instantaneous luminosity on a bunch-by-bunch basis. After applying these corrections, the residual dependence of the energy scale on $\langle \mu \rangle$ is reduced to below 0.1 %. However, significant discrepancies are observed in the high- $\langle \mu \rangle$ region, particularly in data from 2017, necessitating the implementation of dedicated energy scale corrections for each data-taking year.

Additionally, corrections for ADC non-linearity are applied to account for intrinsic non-linear behaviour in the LAr calorimeter electronics. The energy reconstruction for each calorimeter cell relies on a linear conversion from ADC counts to current, followed by a conversion factor from current to energy. This linear relationship is established during dedicated electronics calibration runs, but residual deviations due to non-linearities in the electronics are observed. These deviations introduce non-linearities in the measured energy of electrons and photons. To mitigate this effect, corrections are applied using fifth-order polynomial fits to the residuals, calibrated separately for each calorimeter cell. The corrections result in an approximate 0.4 % increase in cluster energy at low E_T and a 0.2 % decrease at high E_T , with a moderate dependence on particle type and pseudorapidity. A relative uncertainty of 30 % is assigned to these corrections to account for potential systematic biases.

3.2.5 Photon leakage correction

The calibration factors derived from $Z \rightarrow ee$ events are designed to accurately represent the intrinsic response of the calorimeter and are therefore applied equally to electrons and photons. However, differences in the lateral development of electromagnetic showers between electrons and photons necessitate the implementation of photon-specific corrections to ensure precise energy measurements.

Electrons and photons deposit between 1 % and 6 % of their energy outside the reconstructed cluster, depending on their E_T , η , and particle type. This effect is corrected for using the MC-based energy response calibration. However, an imprecise simulation modelling could introduce a bias. To mitigate this, the lateral energy leakage is studied directly in both data and simulation, accounting for differences between electrons and photons. Event samples from $Z \rightarrow \ell\ell\gamma$ and $Z \rightarrow ee$ decays are utilised. The lateral energy leakage, l , is calculated by comparing the energy collected in the second-layer cells of the supercluster with the energy in the cells within a larger rectangular window of size 7×11 in $\eta \times \phi$, centred on the supercluster:

$$l = \frac{E_{7 \times 11}^{\text{L2}}}{E_{\text{nom}}^{\text{L2}}} - 1. \quad (3.8)$$

The difference between data and simulation is shown for electrons, as well as for

converted and unconverted photons, in two pseudorapidity regions in Figure 3.16.

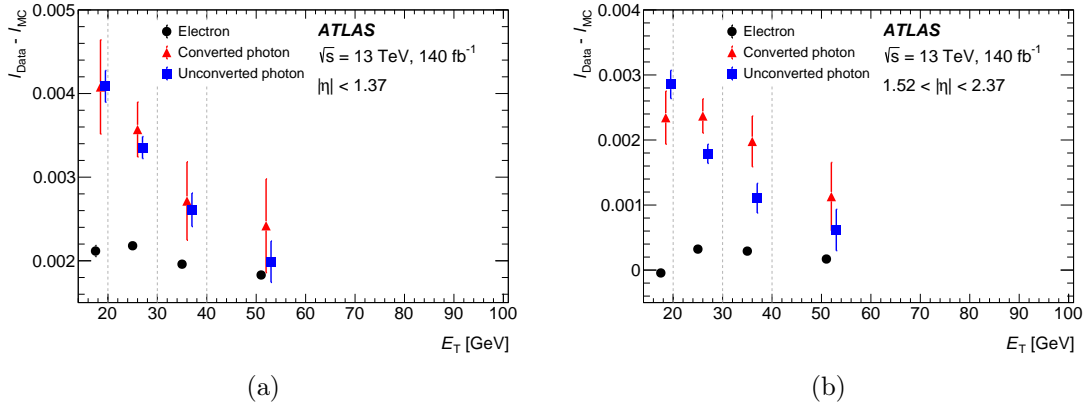


Figure 3.16: The difference between the leakage fractions in data and simulation, in (a) the barrel and (b) the end-caps as a function of E_T , for electrons, converted photons and unconverted photons [123]. A variable bin size is used, indicated by the dotted lines. The last bin corresponds to the range $E_T > 40$ GeV. The markers are placed near the average E_T value of the corresponding bin.

The double difference between electrons and photons in data and simulation is defined as:

$$\alpha_l = (l_e - l_\gamma)^{\text{data}} - (l_e - l_\gamma)^{\text{MC}}, \quad (3.9)$$

where l_e and l_γ represent the lateral energy leakage for electrons and photons, respectively. This calculation is performed separately for unconverted and converted photons and is used to correct the photon energy scale.

The value of α_l depends on the reconstruction of photon conversions and the classification into converted and unconverted photon categories. The impact of this dependence is evaluated by reweighting the simulation to match the data. The difference between the nominal α_l values and those obtained without reweighting is taken as a systematic uncertainty associated with the correction.

3.2.6 Energy linearity and constraints on the calibration uncertainties

To assess the E_T -dependence of the energy scale corrections α_i , the procedure detailed in Section 3.2.3 is repeated in bins of $|\eta|$ and E_T . An extended definition of the calibrated energy is introduced:

$$E^{\text{data,corr}} = \frac{E^{\text{data}}}{(1 + \alpha_i)(1 + \alpha'_j)}, \quad (3.10)$$

where the α_i values are fixed to their previously determined values, and α'_j quantifies the E_T -dependence of the energy scale ($\alpha'_j = 0$ if no dependence is observed). Here, the index j represents a two-dimensional mapping of the electron phase-space in $|\eta|$ and E_T .

The α'_j values are adjusted to achieve the best agreement between the invariant mass distributions in data and simulation for each bin j . This measurement of energy linearity

provides a constraint on the uncertainty model, improving the understanding of energy scale behaviour across different kinematic regions.

3.2.7 Calibration cross-checks

To validate the calibration chain in data, the complete procedure is applied to low- E_T electron candidates from $J/\psi \rightarrow ee$ decays and photon candidates from radiative Z -boson decays. The well-known mass of the J/ψ provides an independent check of the energy calibration for electrons with E_T in the range of 5 to 30 GeV. The difference between data and simulation in $J/\psi \rightarrow ee$ events is quantified through residual energy scale factors, α_i , which are extracted from the peak positions of the reconstructed invariant mass distribution. Figure 3.17 illustrates the variation of α as a function of η and E_T . The post-fit residual scale factors remain below 0.5 % and are consistent with zero within the total calibration uncertainty.

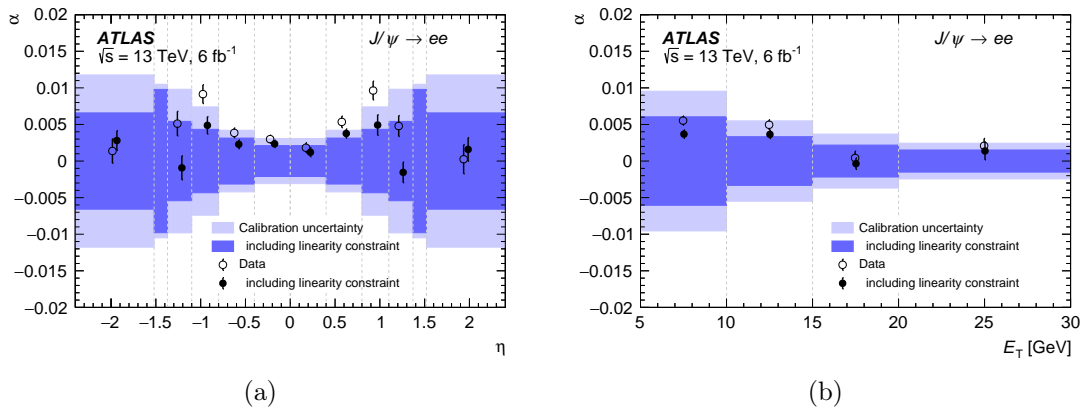


Figure 3.17: Residual energy scale as a function of (a) η and (b) E_T , as measured with $J/\psi \rightarrow ee$ events [123]. The data points and uncertainty bands are shown for both the pre- and post-linearity-fit energy scale models. The uncertainty bands correspond to the energy calibration uncertainty for the energy range of the $J/\psi \rightarrow ee$ decays.

The final photon energy scale is validated using radiative Z -boson decays into two oppositely charged leptons (electrons or muons) and a photon from final-state radiation. Inclusive residual energy scale factors are determined by comparing data distributions to MC templates via a χ^2 test. The measured scale factors are $(3.3 \pm 2.0) \times 10^{-3}$ and $(1.4 \pm 1.1) \times 10^{-3}$ for the $Z \rightarrow ee\gamma$ and $Z \rightarrow \mu\mu\gamma$ samples, respectively. Figure 3.18 displays the dependence of α on the photon's energy and η . The uncertainties combine statistical contributions with systematic effects originating from the lepton energy calibration. The residual scale factors agree within one standard deviation and lie within the expected calibration uncertainties.

3.3 Photon identification

3.3.1 Photon identification optimisation

The identification process establishes criteria to select prompt, isolated photons while efficiently rejecting background from hadronic jets [115]. These criteria rely on shower

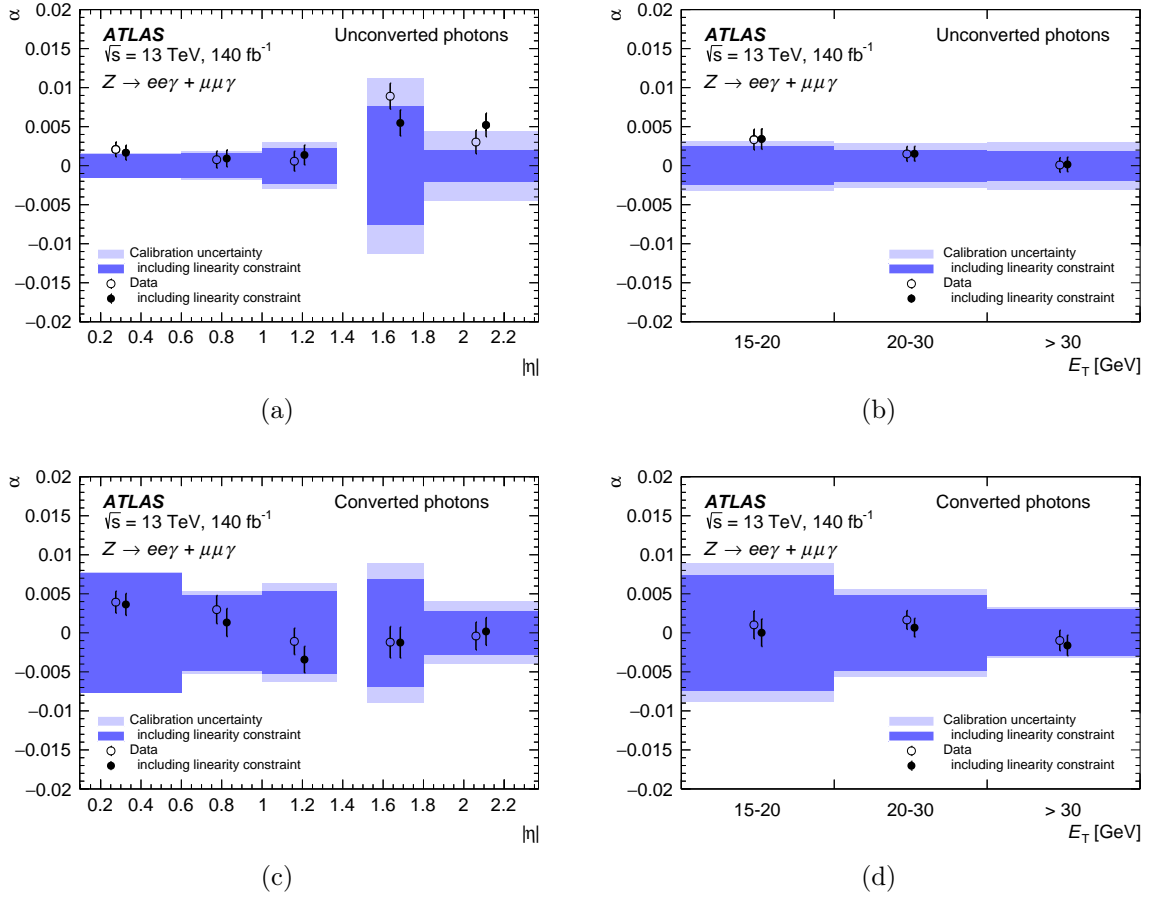


Figure 3.18: Variation of the residual energy scale for (a, b) unconverted and (c, d) converted photons as a function of (a, c) $|\eta|$ and (b, d) E_T , as measured with $Z \rightarrow \ell\ell\gamma$ events [123]. The data points and uncertainty bands are shown for both the pre- and post-linearity-fit energy scale models. The uncertainty bands correspond to the energy calibration uncertainty for photons from $Z \rightarrow \ell\ell\gamma$ decays.

shape variables, with rectangular selections applied in what is referred to as a *cut-based* selection. Table 3.3 provides a comprehensive list of the shower shape variables used for photon identification.

Three primary identification criteria are defined: *Loose*, *Medium*, and *Tight*.

- The Loose selection employs the variables R_{had} , R_{had_1} , R_η , and $w_{\eta 2}$. It serves as the primary selection for triggering photon and diphoton events during the 2015 and 2016 data-taking periods.
- The Medium selection builds upon the Loose selection by incorporating E_{ratio} as an additional variable. This criterion was predominantly used for triggering at the beginning of 2017.
- The Tight selection represents a refined subset of the Medium selection, optimised to enhance background rejection.

The shower shape variables exhibit an important dependence on the calorimeter geometry. As such, the three identification criteria are optimised separately across bins of

$|\eta|$. Furthermore, the Tight working point undergoes additional optimisation in separate bins of E_T to account for its energy dependence.

Beyond these criteria, four additional working points are defined: Loose'2, Loose'3, Loose'4, and Loose'5. The definitions of these working points are detailed in Table 3.4, tailored to specific analysis requirements.

The Tight identification is further optimised using TMVA, the machine learning library integrated into ROOT, with separate optimisations performed for converted and unconverted photons. The shower shape variables differ for converted photons due to the opening angle of the electron-positron conversion pairs, an effect magnified by the magnetic field, as well as additional interactions of the conversion pairs with the material upstream of the calorimeters. The optimisation leverages Monte Carlo samples that simulate prompt photons alongside a representative background across various transverse momentum ranges, ensuring robust performance in different kinematic regimes.

3.3.2 Photon identification efficiency

To evaluate the performance of the Tight photon identification on data, which exhibits an E_T -dependence, three distinct samples are utilised: an inclusive-photon production data sample, photons radiated from leptons in $Z \rightarrow \ell\ell\gamma$ decays, and electrons from $Z \rightarrow ee$ decays, with a method applied to transform the electron shower shapes to resemble those of photons. The efficiency measurements are performed on loosely isolated photons to ensure consistency across the datasets.

Imperfections in the modelling of shower shapes impact the efficiency measurement in data and must be accounted for. To address these discrepancies, simulated shower shape distributions are compared to those from data using a control region enriched in real photons. A correction is applied by shifting the MC distributions to align with the data through a χ^2 minimisation. However, certain data-MC discrepancies cannot be fully corrected by this method, particularly regarding the width of the distributions. To quantify these residual differences, the χ^2 minimisation is repeated on the tails of the distribution, defined as the region encompassing 30 % of the events closest to the identification cut value. The shift value obtained from this tail-based comparison is then used to define a systematic uncertainty associated with the modelling of the shower shapes. This ensures that the efficiency measurements incorporate both the central behaviour and the distribution tails to provide a comprehensive evaluation of the modelling accuracy.

Despite the application of shower shape corrections, a residual discrepancy between data and simulation efficiencies persists. The MC simulation is adjusted using scale factors derived in bins of E_T and $|\eta|$ to account for these differences. These scale factors are computed from the three distinct efficiency measurements and subsequently combined through a weighted average. Statistical and systematic uncertainties are treated as uncorrelated across the different methods.

The total uncertainty associated with the combined scale factors varies significantly, ranging from approximately 7 % at low E_T to 0.5 % at high E_T for unconverted photons. For converted photons, the uncertainty spans from 12 % at low E_T to below 1 % at high E_T .

Figures 3.19 and 3.20 illustrate the Tight identification efficiencies for unconverted and converted photons separately. The data/MC scale factors are also shown. Three distinct samples are utilised: an inclusive-photon production data sample (“matrix method”), photons radiated from leptons in $Z \rightarrow \ell\ell\gamma$ decays (“radiative Z”), and electrons from

Table 3.3: Shower-shape variables used for photon identification.

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 $ \eta < 0.8$ and $ \eta > 1.37$)	R_{had_1}	✓	✓	✓
	Ratio of E_T in the hadronic calorimeter to E_T of the EM cluster (used over the range $0.8 < \eta < 1.37$)	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 $3 \times 7 \eta \times \phi$ rectangle (measured in cell units) to the sum of the cell energies in a 7×7 rectangle, both centered around the most energetic cell	R_η	✓	✓	✓
	Lateral shower width, $\sqrt{\left(\sum E_i \eta_i^2\right) / \left(\sum E_i\right) - \left(\left(\sum E_i \eta_i\right) / \left(\sum E_i\right)\right)^2}$, 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	$w_{\eta 2}$	✓	✓	✓
	Ratio of the sum of the energies of the cells contained in a $3 \times 3 \eta \times \phi$ rectangle (measured in cell units) to the sum of the cell energies in a 3×7 rectangle, both centered around the most energetic cell	R_ϕ	-	-	✓
EM first layer	Total lateral shower width, $\sqrt{\left(\sum E_i (i - i_{\text{max}})^2\right) / \left(\sum E_i\right)}$, where i runs over all cells in a window of $\Delta\eta \approx 0.0625$ and i_{max} is the index of the highest-energy cell	w_{stot}	-	-	✓
	Lateral shower width, $\sqrt{\left(\sum E_i (i - i_{\text{max}})^2\right) / \left(\sum E_i\right)}$, where i runs over all cells in a window of 3 cells around the highest-energy cell	w_{s3}	-	-	✓
	Energy fraction outside the 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	-	-	✓

Table 3.4: Working points definitions. Loose'4 uses the same shower-shape variables as the ones used by the Tight working point, except for w_{s3} , f_{side} , ΔE_s , and E_{ratio} .

ID name	Cuts
Loose	$R_{\text{had}_1}, R_{\text{had}}, R_{\eta}, w_{\eta 2}$
Medium	Loose + E_{ratio}
Tight	Loose + $R_{\phi}, w_{s3}, f_{\text{side}}, \Delta E_s, E_{\text{ratio}}, w_{\text{stot}}, f_1$
Loose'2 (LP2)	Tight - w_{s3}, f_{side}
Loose'3 (LP3)	Tight - $w_{s3}, f_{\text{side}}, \Delta E_s$
Loose'4 (LP3)	Tight - $w_{s3}, f_{\text{side}}, \Delta E_s, E_{\text{ratio}}$
Loose'5 (LP3)	Tight - $w_{s3}, f_{\text{side}}, \Delta E_s, E_{\text{ratio}}, w_{\text{stot}}$

$Z \rightarrow ee$ decays (“electron extrapolation”). The efficiency measurements obtained using the different methods are consistent within their respective statistical and systematic uncertainties, demonstrating the robustness of the scale factor determination process.

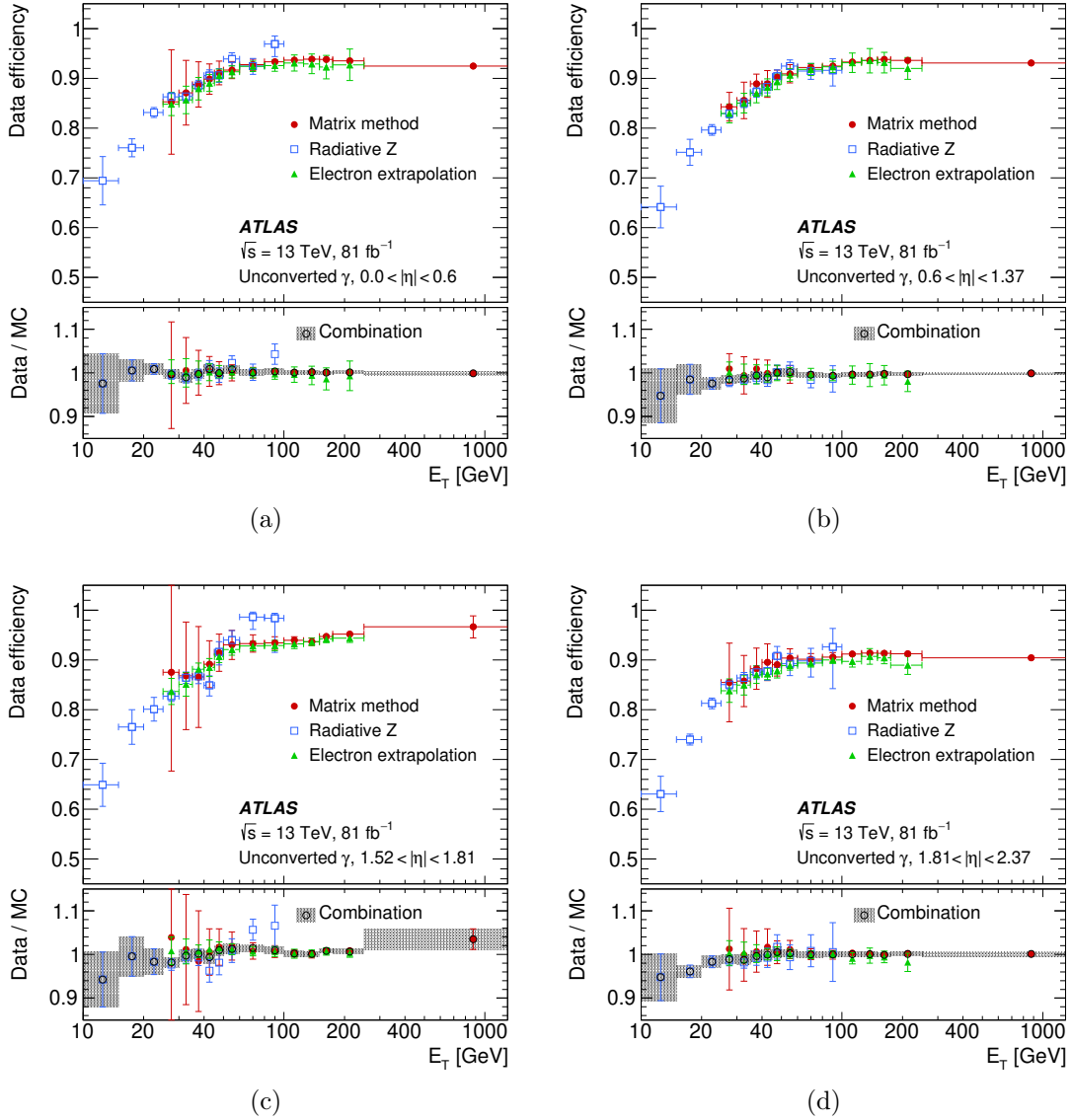


Figure 3.19: Efficiency of Tight identification and data-to-MC ratio for unconverted photons, preselected with a Loose isolation requirement, as a function of E_T , for (a) $0.0 < |\eta| < 0.6$, (b) $0.6 < |\eta| < 1.37$, (c) $1.52 < |\eta| < 1.81$ and (d) $1.81 < |\eta| < 2.37$ [115]. The combined scale factor, derived as a weighted average of the individual scale factors from different measurement methods, is also presented. The band represents the total uncertainty.

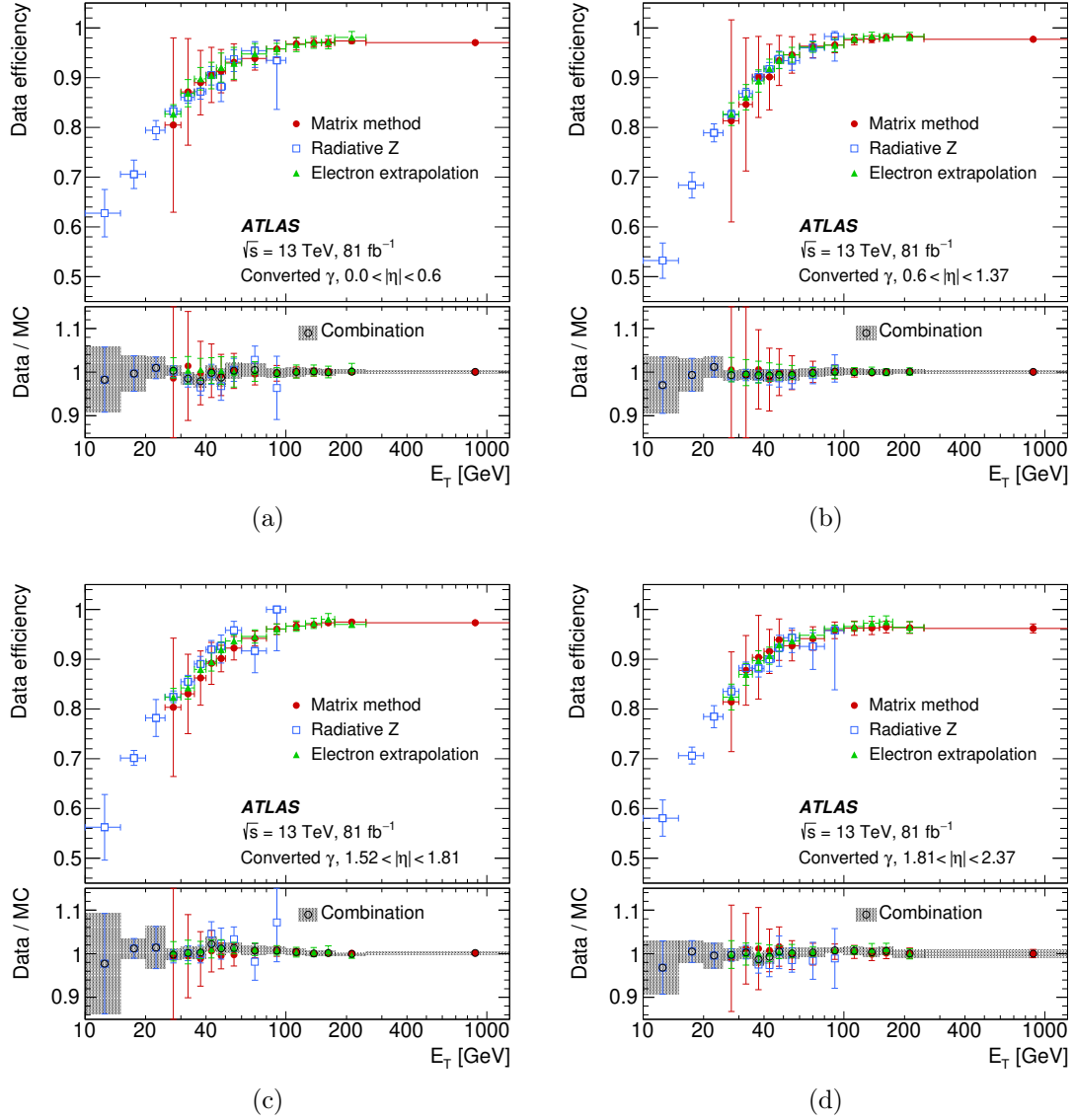


Figure 3.20: Efficiency of Tight identification and data-to-MC ratio for converted photons, preselected with a Loose isolation requirement, as a function of E_T , for (a) $0.0 < |\eta| < 0.6$, (b) $0.6 < |\eta| < 1.37$, (c) $1.52 < |\eta| < 1.81$ and (d) $1.81 < |\eta| < 2.37$ [115]. The combined scale factor, derived as a weighted average of the individual scale factors from different measurement methods, is also presented. The band represents the total uncertainty.

Chapter 4

Photon isolation

Contents

4.1	Calorimetric isolation energy	84
4.2	Track-based isolation	87
4.3	Isolation criteria	87
4.4	Photon isolation corrections	88
4.4.1	Datasets and MC samples	89
4.4.2	Methodology	91
4.4.3	Function	91
4.4.4	Binning	92
4.4.5	Fit to MC distributions	93
4.4.6	Fit to MC PDF parameters	94
4.4.7	Combination of direct and Bremsstrahlung MC regions	96
4.4.8	Fit to data	100
4.4.9	Data-driven shift modelling	100
4.4.10	Conclusion	101

Isolation is a criterion used to classify objects based on the hadronic activity surrounding them. Prompt photons, originating either directly from the hard-scattering vertex or from the decay of heavy resonances such as the Higgs, W , and Z bosons, are typically well-isolated, with little surrounding energy deposits in the calorimeter and the Inner Detector. In contrast, fake photons, such as those arising from hadronic decays, are often embedded in significant hadronic activity. This distinction is leveraged to further suppress the background.

Two isolation variables are commonly used: the *calorimetric isolation energy*, defined as the sum of energy deposits in the calorimeter around the photon candidate, after subtracting contributions from the photon itself and from pile-up, and the *track isolation energy*, calculated as the sum of the transverse momenta of tracks in a specified region around the photon candidate [126].

This chapter begins by introducing the concepts and methods of photon isolation within the ATLAS framework, and then presents my work on the correction of the simulated photon calorimetric isolation. This work, which served as my *qualification task* within the ATLAS collaboration, benefited from the contributions of several people: Luis Pascual Domínguez, who was my technical supervisor for this task, and Omar Medina, an intern I co-supervised.

4.1 Calorimetric isolation energy

The calorimetric isolation energy is constructed from the energy deposits surrounding the photon candidate within a fixed-size cone of radius $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, typically chosen as 0.2 or 0.4, as illustrated in Figure 4.1. It is based on the raw calorimetric isolation energy [127], defined as the sum of the transverse energies of all positive-energy topological clusters:

$$E_{\text{T,raw}}^{\text{iso}} = \sum_{r_i < R}^{\text{clusters}} E_{\text{T}}^i . \quad (4.1)$$

Since this sum includes the photon's own energy, a correction must be applied. To estimate the core energy ($E_{\text{T,core}}$), a fixed-size window of 5×7 calorimeter cells in EM-middle-layer units ($\Delta\eta \times \Delta\phi = 0.125 \times 0.175$) is used. This fixed masking approach is simple and stable across different transverse momenta and pile-up conditions, for both signal and background. However, it does not completely remove the photon's energy, as a fraction of the cluster extends beyond the mask, requiring an additional leakage correction.

The leakage correction ($E_{\text{T,leakage}}$) is derived from MC samples of single photons without pile-up. The energy leaking from the mask is fitted with a Crystal Ball function, and the most probable value μ_{CB} is parametrised as a function of E_{T} in bins of η .

Finally, a correction for pile-up and underlying-event contributions is applied using the ambient energy density ρ_{median} [129]. This correction is computed as the median energy density ρ_{median} multiplied by the area of the isolation cone (excluding the core region).

The median energy density ρ_{median} is estimated event-by-event using the k_t jet-clustering algorithm [130, 131] with a radius parameter of $\Delta R = 0.5$ and no transverse momentum threshold for jets. The energy density per jet is given by

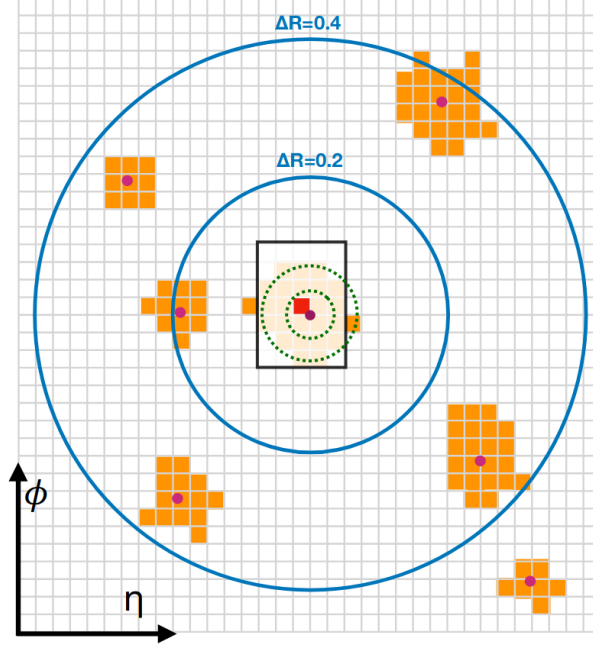


Figure 4.1: Schema of the calorimetric isolation energy variable [128]. The grid represents the second layer of the calorimeter in the $\eta - \phi$ plane. The two blue circles are the two cones of 0.2 and 0.4 opening used to compute the isolation energy. All topo-clusters, represented in orange, whose barycentre falls within the cone, are added in the calculation. The photon candidate is at the centre, with its hottest cell denoted with a red square. The 5×7 fixed-size mask (in EM-middle-layer units) used to subtract the photon energy is represented, with energy leaking outside. The two dotted circles represent the first and second Molière radius.

$$\rho^{\text{jet}} = \frac{p_T^{\text{jet}}}{A^{\text{jet}}}, \quad (4.2)$$

where p_T^{jet} is the jet transverse momentum and A^{jet} is its area. The median energy density ρ_{median} is then computed from all jets in the event.

Ideally, ρ_{median} should vary smoothly with $|\eta|$, but detector effects introduce non-monotonic variations. As $|\eta|$ increases, the calorimeter segmentation becomes coarser, with larger cell sizes and higher noise thresholds. This reduces the probability of soft pile-up energy deposits forming clusters, leading to a non-trivial dependence of ρ_{median} on $|\eta|$ [113].

Previously, ρ_{median} was computed in only two $|\eta|$ bins: $|\eta| < 1.5$ and $1.5 < |\eta| < 3.0$. However, this “two-region approach” fails to capture the smooth evolution of pile-up energy across η . A more refined parametrisation has since been introduced [132]:

$$E_{\text{T,pile-up}}(\eta) = \rho_{\text{median}}^{\text{central}} \times A \times \zeta(\eta), \quad (4.3)$$

where $\rho_{\text{median}}^{\text{central}}$ is the median energy density in the central region ($|\eta| < 1.5$), A is the effective isolation area, and $\zeta(\eta)$ is a correction function accounting for the smooth decrease of pile-up energy at larger $|\eta|$ and detector-specific inhomogeneities (e.g., reduced cluster reconstruction efficiency in the crack region).

Since detector effects vary with isolation area size, $\zeta(\eta)$ must be adapted accordingly.

For example, a small isolation cone near the calorimeter crack will have a larger fraction of its area inside the crack, whereas a larger isolation cone will extend beyond it. Consequently, $\zeta(\eta)$ should be higher for larger isolation cones to compensate for the difference in pile-up acceptance.

To account for arbitrary isolation area shapes, a set of parameters $f_{i_{\eta\text{bin}}}$ is introduced, allowing a more precise modelling of pile-up contributions:

$$\zeta(\eta) = \sum_{i_{\eta\text{bin}}} \frac{A_{i_{\eta\text{bin}}}}{A} f_{i_{\eta\text{bin}}} , \quad (4.4)$$

where the sum runs over a set of η -bins. $f_{i_{\eta\text{bin}}}$ are constant parameters in each η -bin, describing the average pile-up abundance and detector acceptance, A is the total isolation area, and $A_{i_{\eta\text{bin}}}$ is the isolation area within the i -th η -bin.

The parameters $f_{i_{\eta\text{bin}}}$ are extracted from $Z \rightarrow e^+ e^-$ events, where the median isolation energy is fitted simultaneously in two isolation rings: $E_T^{\text{cone30}} - E_T^{\text{cone20}}$ and $E_T^{\text{cone40}} - E_T^{\text{cone30}}$.

These rings are chosen because their area is less affected by energy leakage from the electron, making pile-up the dominant contribution to isolation on average.

$\zeta(\eta)$ obtained with this approach is shown in Figure 4.2 for two different isolation cones, along the fit parameters $f_{i_{\eta\text{bin}}}$. Compared to the two-region method, this “smooth correction” provides a significantly improved description of pile-up contributions, leading to a more accurate isolation correction.

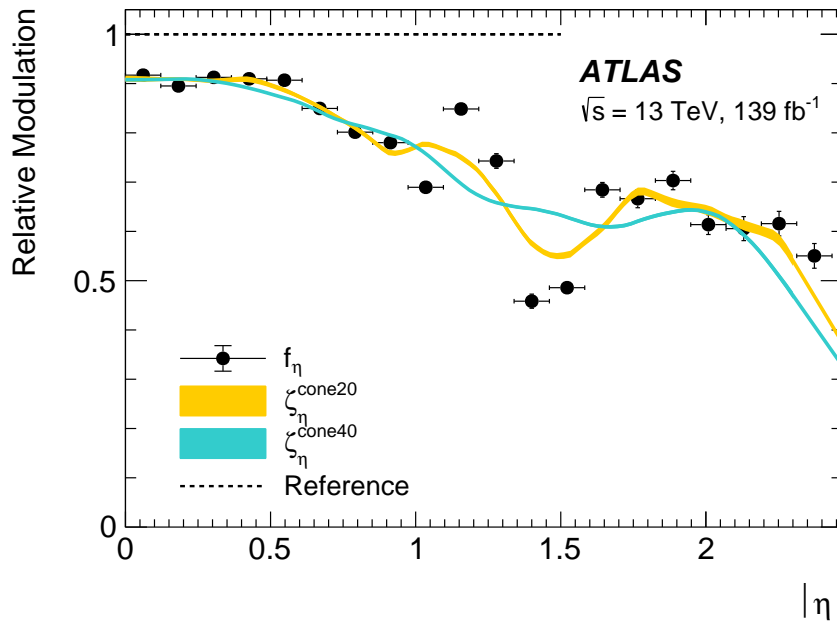


Figure 4.2: Fit parameters $f_{i_{\eta\text{bin}}}$ and $\zeta(\eta)$ for E_T^{cone20} and E_T^{cone40} versus $|\eta|$ [132]. The statistical uncertainties on the fit parameters $f_{i_{\eta\text{bin}}}$ and the propagated uncertainties on $\zeta(\eta)$ are shown. The range where the reference value of $\rho_{\text{median}}^{\text{central}}$ is measured on an event-by-event basis is shown as a dashed line.

Finally, the corrected calorimetric isolation energy is given by:

$$E_T^{\text{coneXX}} = E_{T,\text{raw}}^{\text{isoXX}} - E_{T,\text{core}} - E_{T,\text{leakage}}(E_T, \eta) - E_{T,\text{pile-up}}(\eta, \Delta R) \quad (4.5)$$

where XX refers to the cone size, with $\Delta R = \text{XX}/100$.

4.2 Track-based isolation

The track isolation variable (p_T^{coneXX}) is defined as the sum of the transverse momenta of selected tracks within a cone of radius $\Delta R = \text{XX}/100$, centred on the direction of the photon candidate, as illustrated in Figure 4.3 [126]. Only tracks with $p_T > 1 \text{ GeV}$ and $|\eta| < 2.5$, and that satisfy specific quality criteria [127], are considered.

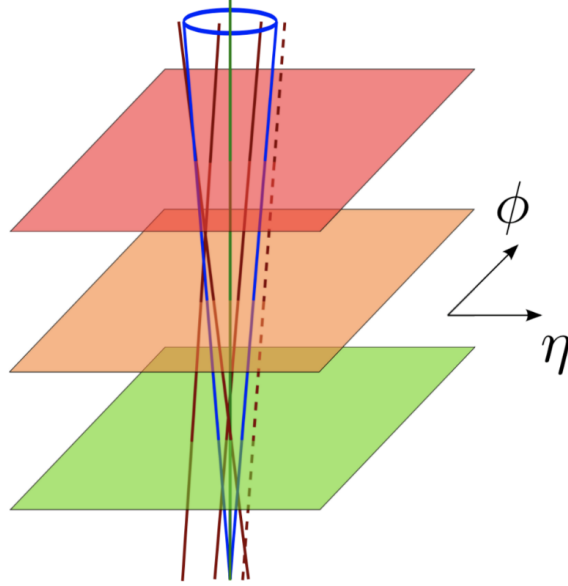


Figure 4.3: Schema of the track isolation variable. All good tracks (brown) within the cone (blue) centred on the photon candidate are selected. The dotted line shows a track outside the cone, which is not selected. The p_T of selected tracks is summed to compute the track isolation variable, and the core energy is subtracted. Image adapted from Ref. [128].

These requirements include a minimum number of hits in the silicon detectors (Pixel + SCT), as well as constraints on shared hits and detector holes. Specifically, a track must have at least seven silicon hits, no more than one shared hit—defined as $n_{\text{Pixel}}^{\text{sh}} + n_{\text{SCT}}^{\text{sh}}/2$, where $n_{\text{Pixel}}^{\text{sh}}$ and $n_{\text{SCT}}^{\text{sh}}$ are the numbers of hits shared between multiple tracks in the Pixel and SCT detectors—no more than two silicon holes, and at most one pixel hole.

To reduce the impact of pile-up, tracks are loosely associated with the primary vertex of interest, typically the hard-scattering vertex of the event. This vertex is defined as the one with the highest sum of squared transverse momenta of its associated tracks. For this purpose, tracks are selected if they satisfy $|\Delta z_0| \sin \theta < 3 \text{ mm}$, where Δz_0 is the longitudinal impact parameter relative to the selected primary vertex.

Finally, all tracks used to form conversion vertices associated with the photon are removed from the list.

4.3 Isolation criteria

The isolation criteria are tailored to the specific requirements of each analysis. They represent a trade-off between achieving high efficiency for identifying prompt, isolated photons and ensuring strong rejection of photons originating from hadronic decays. Different isolation selections, referred to as working points, are summarised in Table 4.1.

Working point	Calorimetric isolation	Track isolation
Loose	$E_T^{\text{cone20}} < 0.065 \times E_T$	$p_T^{\text{cone20}}/E_T < 0.05$
Tight	$E_T^{\text{cone40}} < 0.022 \times E_T + 2.45 \text{ GeV}$	$p_T^{\text{cone20}}/E_T < 0.05$
TightCaloOnly	$E_T^{\text{cone40}} < 0.022 \times E_T + 2.45 \text{ GeV}$	-

Table 4.1: Definition of the photon isolation working points.

4.4 Photon isolation corrections

The calorimetric isolation energy is not perfectly modelled in the MC simulation due to limitations in both the pile-up correction, based on an imperfect pile-up model, and the leakage correction, which is derived from simulations without pile-up. These modelling inaccuracies lead to discrepancies in isolation efficiencies between data and simulation. To ensure a consistent treatment across data and MC, a correction is applied to the calorimetric isolation energy in the simulation. This is done using data-driven shifts, which quantify the difference in the peak positions of the isolation energy distributions between data and MC. Figure 4.4 illustrates the concept of this data-driven shift. The methodology is explained in detail in the following sections.

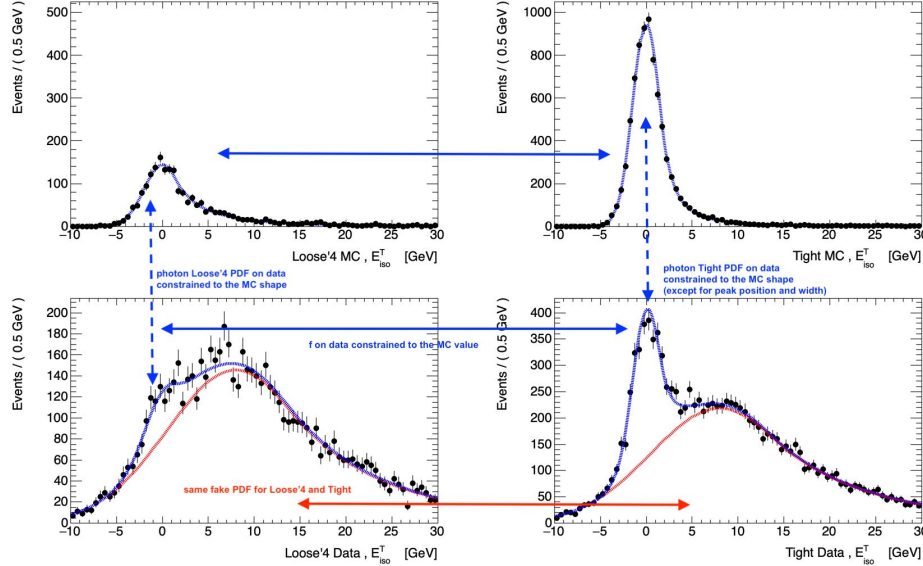


Figure 4.4: Isolation energy distributions for data and MC, with their fit, for a “toy” dataset, to illustrate the concept of the calculation. The data-driven shift is the difference in the peak positions of the distributions of Tight Data and Tight MC.

The shifts are extracted by fitting the calorimetric isolation energy distributions in regions enriched in real photons, and are derived separately for the E_T^{cone20} and E_T^{cone40} variables. As described in Section 4.4.2, the shifts are computed in discrete bins and then combined into a continuous two-dimensional correction function of E_T^γ and $\langle\mu\rangle$.

4.4.1 Datasets and MC samples

Photon isolation corrections are derived from two primary event topologies: radiative Z -boson decays at low photon transverse energy ($10 < E_T^\gamma < 100$ GeV), and inclusive photon production at higher energies ($25 \text{ GeV} < E_T^\gamma < \sim 1.5 \text{ TeV}$) [126].

4.4.1.1 Event selection

Radiative Z -boson decay events $Z \rightarrow \ell\ell\gamma$ events offer a clean signature, with the invariant mass of the three-body system reconstructing the Z -boson mass. Photon candidates are selected from events passing unprescaled single- or dilepton triggers, requiring two reconstructed leptons and one reconstructed photon. Kinematic selections are applied on the lepton and photon transverse energies, along with identification and isolation criteria. To enrich the sample in final-state radiation photons, events are selected with dilepton invariant mass in the range $40 < m_{\ell\ell} < 80$ GeV and three-body mass $80 < m_{\ell\ell\gamma} < 100$ GeV. These radiative Z -boson decays provide high photon purity but are limited to the low- E_T region.

Inclusive photon events Inclusive photon events are selected using single-photon triggers of the form `HLT_gX_loose`, where `X` denotes the photon E_T^γ threshold. For $E_T^\gamma < 140$ GeV, these triggers are prescaled and require only loose identification and a minimum photon transverse energy. The full list of triggers used to select photons in this work is shown in Table 4.2. The efficiencies of `HLT_g120_loose` and `HLT_g140_loose` are shown in Figure 4.5, for the 2015 data. Selected events are predominantly γ -jet events, where the photon recoils against a jet. While this selection provides access to higher E_T^γ photons, it suffers from a higher background contamination, particularly at low E_T^γ . Additional contributions arise from fragmentation (or Bremsstrahlung) photons, which tend to increase the tails of the isolation energy distribution due to energy deposited in the isolation cone by nearby jets. The distinction between direct photons (from the hard scatter) and Bremsstrahlung photons is illustrated in Figure 1.7.

Trigger	p_T threshold [GeV]
<code>HLT_g10_loose</code>	10
<code>HLT_g15_loose_L1EM7</code>	15
<code>HLT_g20_loose_L1EM12</code>	20
<code>HLT_g25_loose_L1EM15</code>	25
<code>HLT_g35_loose_L1EM15</code>	35
<code>HLT_g40_loose_L1EM15</code>	40
<code>HLT_g45_loose_L1EM15</code>	45
<code>HLT_g50_loose_L1EM15</code>	50
<code>HLT_g60_loose</code>	60
<code>HLT_g70_loose</code>	70
<code>HLT_g80_loose</code>	80
<code>HLT_g100_loose</code>	100
<code>HLT_g120_loose</code>	120
<code>HLT_g140_loose</code>	140

Table 4.2: List of single photon triggers, along with their p_T thresholds.

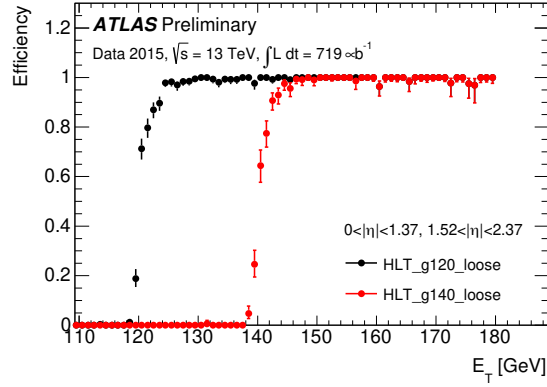


Figure 4.5: Efficiencies of HLT_g120_loose and HLT_g140_loose triggers with respect to photons passing Tight identification criteria as a function of the photon transverse energy, excluding the crack region ($1.37 < |\eta| < 1.52$) [133]. The efficiency is calculated using events recorded with a L1 trigger requiring an EM cluster with $E_T > 7$ GeV, and no background subtraction is applied.

Signal and background samples True photons are selected by requiring Tight identification. Fake photons are estimated from data by selecting photons that fail the Tight identification but pass a looser criterion, referred to as Loose’4 (see Section 3.3.1). Although MC samples with “jet-filtered” PYTHIA 8 [134] events (filtered with an algorithm that enhances the contribution of fake photons) are also available, they suffer from limited statistics. The Loose’4 selection applied on data yields a background-enriched sample, though it retains a non-negligible contamination from prompt photons, which depends on both E_T^γ and $|\eta^\gamma|$. This contamination must be subtracted to accurately extract the shape of the background distribution. Similarly, the Tight selection retains a non-negligible contamination from fake photons, which also has to be subtracted.

4.4.1.2 Samples

For the analysis, the full Run 2 dataset is used. Simulated samples include the two categories presented above.

Radiative $Z \rightarrow \ell\ell\gamma$ events with photon transverse energy above 10 GeV are generated using the SHERPA event generator [135], version 2.1.1. The simulation employs leading-order QCD matrix elements with up to three additional partons in the final state. The CT10 parton distribution function (PDF) set [136] is used for the proton structure.

Inclusive photon events are generated with PYTHIA 8, which simulates both direct and fragmentation photon production. The direct component includes leading-order γ -jet processes such as $qg \rightarrow q\gamma$ and $q\bar{q} \rightarrow g\gamma$, while the fragmentation component arises from QED radiation in all $2 \rightarrow 2$ QCD processes involving light quarks and gluons. To cover the full photon E_T^γ spectrum, samples are generated in bins of photon transverse energy. A weight is applied to events in each bin to reconstruct the physical E_T^γ distribution.

All generated events are processed through the full ATLAS detector simulation framework [137], which is based on GEANT4 [69]. Simulated events include additional pp interactions, matched to the pile-up conditions observed during Run 2 data-taking. These additional interactions are overlaid using PYTHIA 8 soft QCD processes with the A3 tune [138] and the NNPDF23LO PDF set.

The present thesis focuses exclusively on the high- E_T^γ regime, corresponding to inclu-

sive QCD photon production. Other work has been carried out in ATLAS in the low- E_T^γ regime. Unless otherwise specified, in the following, Loose'4 refers to a photon that meets the Loose'4 criteria but does not meet the Tight requirements. Since the transverse energy and transverse momentum of a photon are identical, the notations p_T^γ and E_T^γ are used interchangeably. When there is no ambiguity, the simplified forms p_T or E_T may also be used.

4.4.2 Methodology

A brief summary of the methodology used to estimate the shift in peak position between the data and MC calorimetric isolation energy is provided here; a more detailed explanation is given in the following sections. The procedure involves performing fits in several distinct data and MC regions. Six photon categories are defined: Tight direct MC, Tight Bremsstrahlung MC, Loose'4 direct MC, Loose'4 Bremsstrahlung MC, Tight data, and Loose'4 data.

In bins of $|\eta|$, E_T^γ , pile-up $\langle\mu\rangle$, and photon conversion type, the calorimetric isolation energy distributions are fitted independently in each MC region. The resulting fit parameters are then parametrised as continuous two-dimensional functions of E_T^γ and $\langle\mu\rangle$. These fits provide a set of smooth models for each photon type and region.

To build a single signal model, the relative contributions of direct and Bremsstrahlung photons are estimated and used to combine their corresponding MC components. Since the Tight and Loose'4 MC regions contain only true photons, it is assumed that their shapes can be transferred to the corresponding data regions. The calorimetric isolation distributions in the Tight and Loose'4 data regions are then fitted simultaneously, using the previously determined MC signal shapes. This is done under the additional assumption that the background shape is identical in both data regions.

The data-driven shift is defined as the difference in the peak position between the Tight data distribution and the combined Tight MC model. This shift is finally parametrised as a smooth two-dimensional function of E_T^γ and $\langle\mu\rangle$.

4.4.3 Function

The isolation energy distributions are modelled using Asymmetrical Double-Sided Crystal Ball (ACB) functions. An ACB consists of a Gaussian core with a peak position μ_P and different widths on each side: σ_L on the left and σ_R on the right. This core is smoothly connected to a power-law tail, which accounts for non-Gaussian behaviour in the high-energy region:

$$f(x; \mu_P, \sigma_L, \sigma_R, \alpha, n) = N \cdot \begin{cases} \exp\left(-\frac{(x-\mu_P)^2}{2\sigma_L^2}\right) & \text{for } x < \mu_P \\ \exp\left(-\frac{(x-\mu_P)^2}{2\sigma_R^2}\right) & \text{for } x \geq \mu_P \text{ and } \frac{x-\mu_P}{\sigma_R} < \alpha \\ \frac{A}{\left(B + \frac{x-\mu_P}{\sigma_R}\right)^n} & \text{for } \frac{x-\mu_P}{\sigma_R} > \alpha \end{cases} \quad (4.6)$$

with $A = \left(\frac{n}{\alpha}\right)^n \exp\left(-\frac{\alpha^2}{2}\right)$ and $B = \frac{n}{\alpha} - \alpha$

Figure 4.6 presents an ACB with arbitrary parameters.

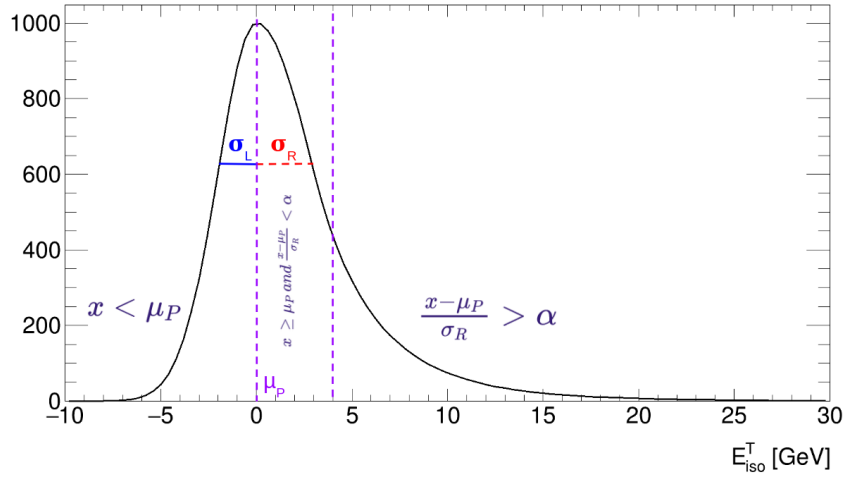


Figure 4.6: An ACB with $N = 1000$, $\mu_P = 0$ GeV, $\sigma_L = 2$ GeV, $\sigma_R = 3$ GeV, $\alpha = 1.3$, $n = 10$.

4.4.4 Binning

Fits of calorimetric isolation energy are performed in bins of $|\eta|$, pile-up $\langle\mu\rangle$, photon transverse energy p_T , and conversion type (converted or unconverted). As previously discussed, each MC sample covers a specific range of p_T , requiring event reweighting to reconstruct the physical distribution. In data, a similar correction must be applied to account for the trigger prescales associated with the HLT_gX_loose triggers. Because these weights introduce additional complexity to the fitting procedure, the fits are instead performed in p_T bins defined to be exclusive to specific datasets and trigger selections.

Because of the trigger thresholds, the p_T spectra display turn-on regions at different energies, which do not reflect the physical distribution. An example of such a turn-on is shown in Figure 4.7, between 120 GeV and 125 GeV, for the HLT_g120_loose trigger. To avoid biases introduced by these effects, the p_T binning is chosen to exclude the turn-on regions. The unweighted p_T distribution of Tight photons collected between 2015 and 2018 is shown in Figure 4.8.

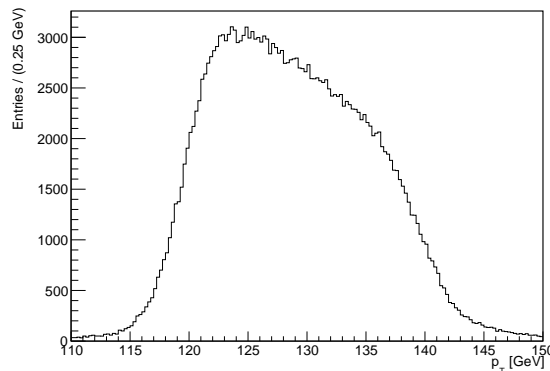


Figure 4.7: p_T distribution for 2018 data, Tight photons, passing HLT_g120_loose only. The turn-on is visible between 120 GeV and 125 GeV. A “turn-down” is also visible above 135 GeV, corresponding to events firing HLT_g140_loose.

Due to the variations in available statistics across different categories and ranges of

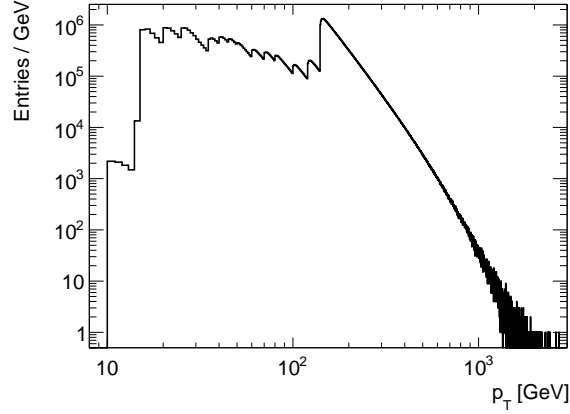


Figure 4.8: E_T^γ spectrum of inclusive photons for the full Run 2 dataset. Discontinuities in the distribution arise from trigger prescales, while smooth transitions are observed in the turn-on regions corresponding to trigger thresholds. The p_T thresholds of the triggers are listed in Table 4.2.

p_T and $\langle\mu\rangle$, a variable binning strategy is used for both variables. The bin widths are adjusted to ensure a minimum of 1000 events in each bin, providing sufficient statistical precision for the fits. The bins are chosen to be as narrow as possible in $\langle\mu\rangle$, increasing the width in p_T if necessary. This explains why the trends in the following are highly dependent on $\langle\mu\rangle$, but only mildly on p_T . Table 4.3 summarises the binning choices.

Variable	Binning
$ \eta $	[0, 0.6[, [0.6, 0.8[, [0.8, 1.15[, [1.15, 1.37[, [1.52, 1.81[, [1.81, 2.37[(crack region excluded)
Conversion type	Converted, unconverted
Pile-up $\langle\mu\rangle$	12 to 58 (variable binning to ensure sufficient statistics)
Transverse energy p_T	From 9 GeV to 1 TeV (variable binning)

Table 4.3: Binning used to fit the calorimetric isolation energy.

4.4.5 Fit to MC distributions

The MC distributions are first fitted individually for each category: Tight direct, Tight Bremsstrahlung, Loose'4 direct, and Loose'4 Bremsstrahlung. The two identification working points, Loose'4 and Tight, are used to model the signal components in the corresponding data regions. The direct and Bremsstrahlung components are fitted separately and later combined into a single probability density function (see Section 4.4.7). The data-driven shifts are ultimately computed by comparing the Tight data to the combined Tight MC signal model.

Depending on the photon identification, conversion status, and available statistics, the granularity of the binning in p_T and $\langle\mu\rangle$ varies significantly. In favourable cases, fine binning is achievable, allowing precise modelling. In less populated regions, larger bins are

required to ensure stable fits. The most challenging case is the Bremsstrahlung Loose'4 converted category, where the ACB function does not perform well. Fortunately, this category contributes only a small fraction to the overall signal model.

Because of the strong correlations between the ACB parameters, the tail parameter n is held fixed during the fits; it is set to 10 for direct photons and 3 for Bremsstrahlung photons, as these values enable the correct modelling of the two photon categories.

Figure 4.9 and Figure 4.10 show examples of fits to MC distributions in the bin $1.81 < |\eta| < 2.37$ for unconverted and converted photons, respectively, for direct and Bremsstrahlung photons, and for representative p_T and $\langle\mu\rangle$ intervals.

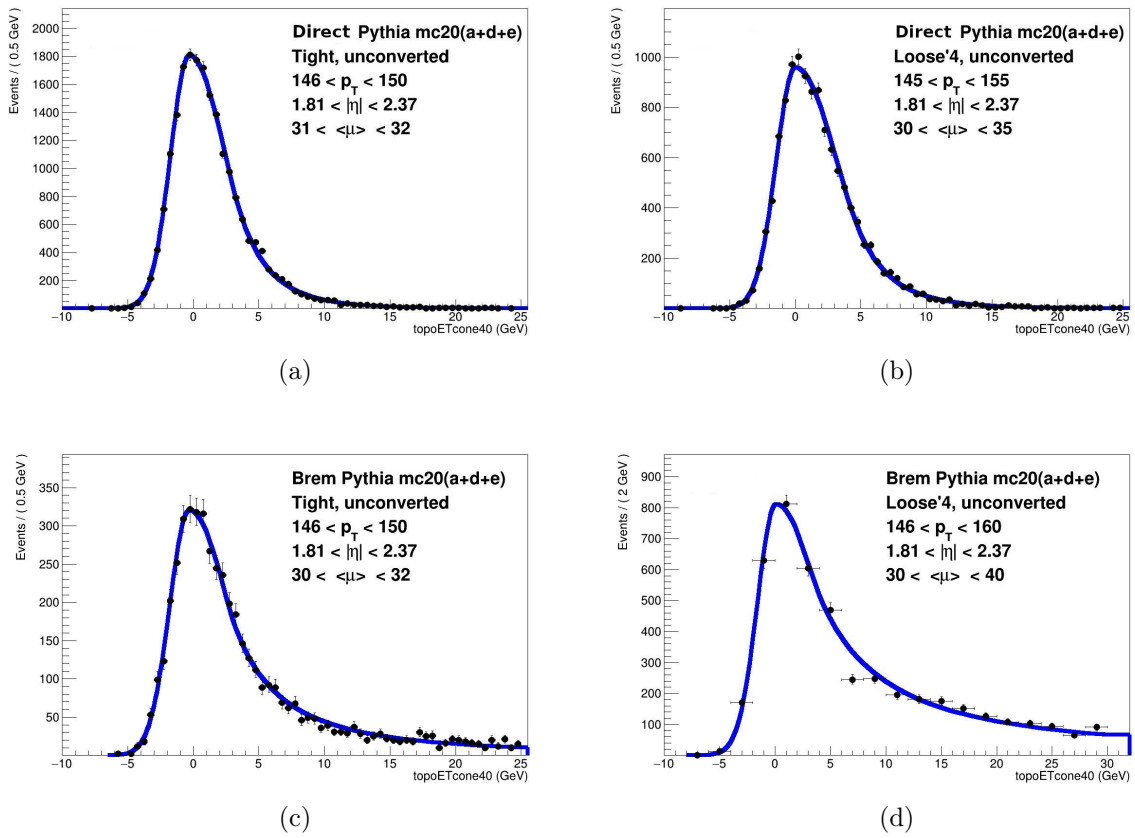


Figure 4.9: Fits to MC simulated calorimetric isolation energy distributions, with a cone size of $\Delta R = 0.4$, for $1.81 < |\eta| < 2.37$ and unconverted photons, shown for (a) Tight direct, (b) Loose'4 direct, (c) Tight Bremsstrahlung and (d) Loose'4 Bremsstrahlung photons. The p_T and $\langle\mu\rangle$ bins are indicated on the plots.

4.4.6 Fit to MC PDF parameters

Once the MC distributions have been fitted individually using ACB functions, the resulting PDF (probability density function) parameters are modelled as continuous functions of p_T and $\langle\mu\rangle$. This is achieved using a series expansion around a chosen reference point f_{00} .

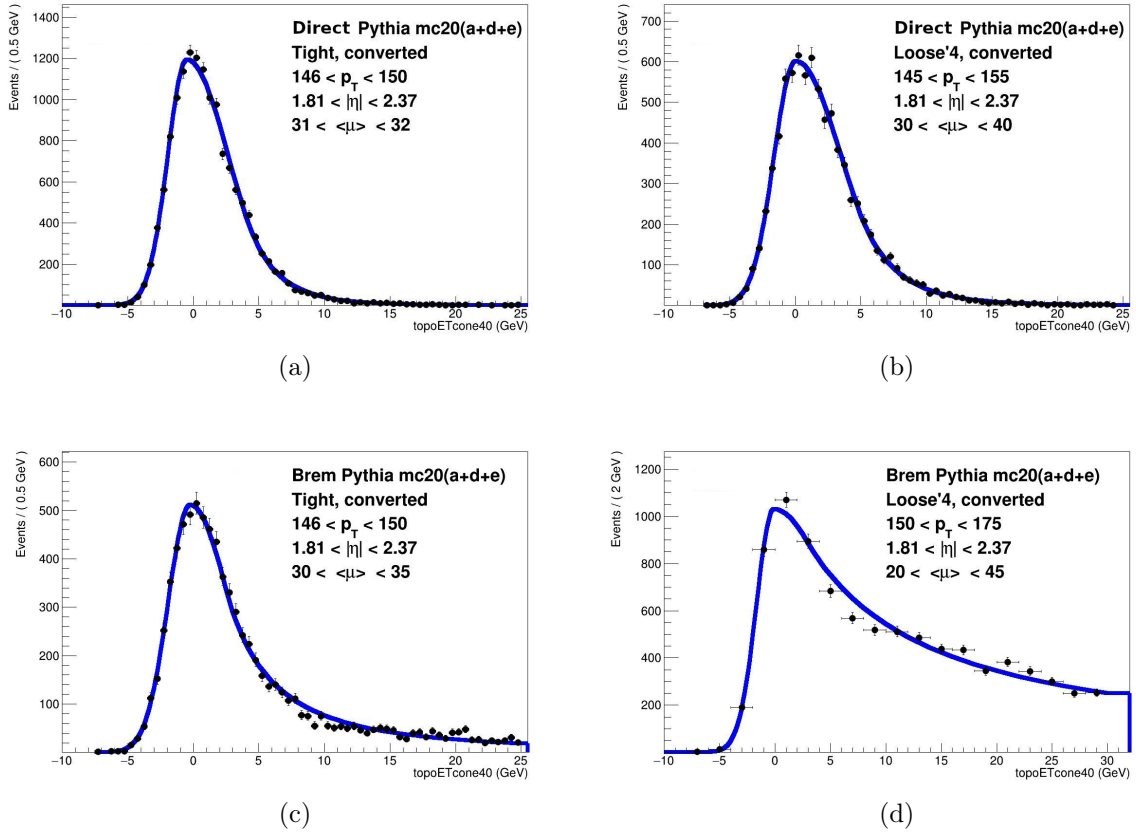


Figure 4.10: Fits to MC simulated calorimetric isolation energy distributions for a cone size of $\Delta R = 0.4$, in the region $1.81 < |\eta| < 2.37$, for converted photons. Results are shown for (a) Tight direct, (b) Loose'4 direct, (c) Tight Bremsstrahlung and (d) Loose'4 Bremsstrahlung photons. The p_T and $\langle \mu \rangle$ bins are indicated on the plots.

$$\begin{aligned}
 f(x, y) = & f_{00} \\
 & + c_x x + c_y y \\
 & + c_{xx} x^2 + c_{xy} xy + c_{yy} y^2 \\
 & + c_{xxx} x^3 + c_{xxy} x^2 y + c_{xyy} xy^2 + c_{yyy} y^3 \\
 & + c_{xxxx} x^4 + c_{yyyy} y^4, \\
 \text{with } x = & p_T - p_T^0 \text{ and } y = \langle \mu \rangle - \langle \mu \rangle^0.
 \end{aligned} \tag{4.7}$$

Depending on the parameter considered and the available statistics, the cubic and quartic terms are sometimes fixed to zero. p_T^0 and $\langle \mu \rangle^0$ are the mean p_T and $\langle \mu \rangle$ values: $p_T^0 = \langle p_T \rangle$ and $\langle \mu \rangle^0 = \langle \langle \mu \rangle \rangle$. The reference point f_{00} is the value of f evaluated at $p_T = p_T^0$ and $\langle \mu \rangle = \langle \mu \rangle^0$ (i.e., $x = 0$ and $y = 0$).

Figure 4.11 illustrates the fits to the right-hand width parameter σ_R in two representative $|\eta|$ bins, along with the corresponding pull distributions. As expected, σ_R increases with both p_T and $\langle \mu \rangle$, reflecting a broadening of the distribution: higher pile-up and more energetic photons both contribute to an overall increase in the spread of the isolation energy.

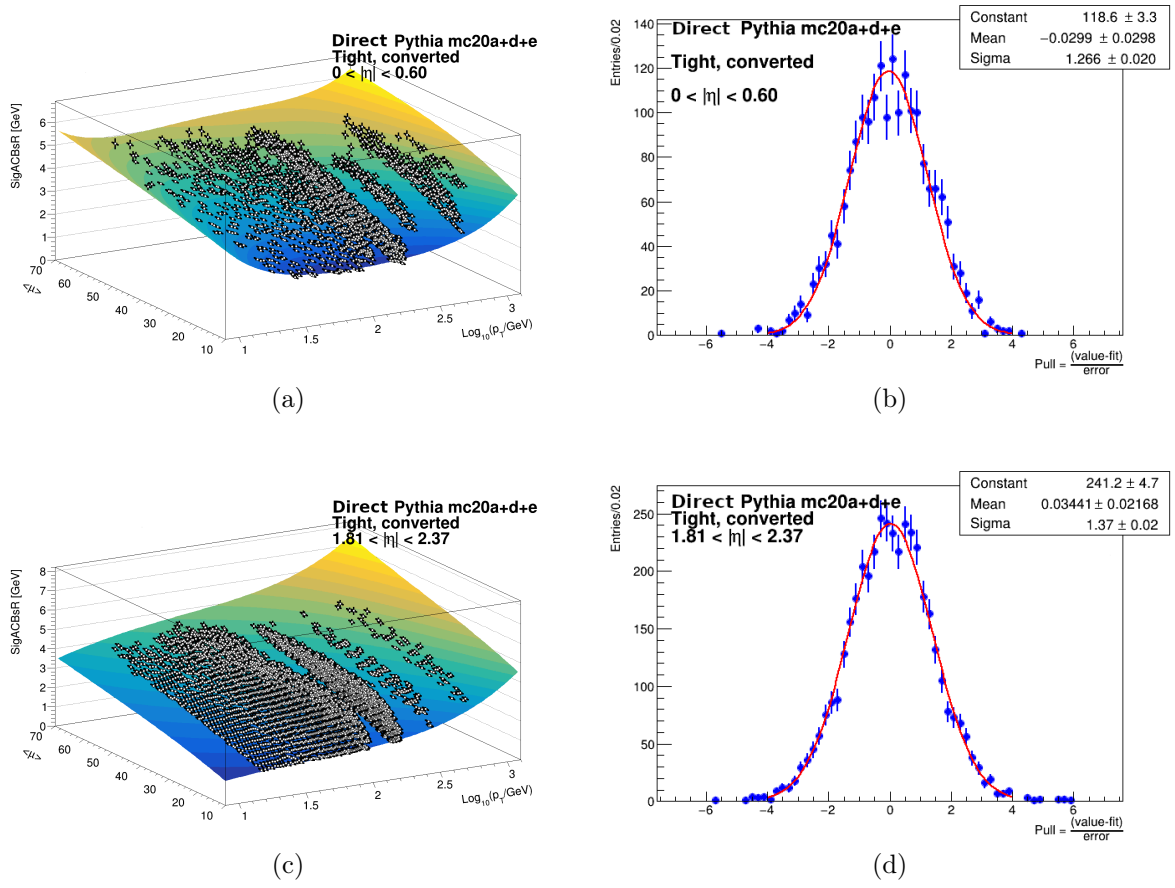


Figure 4.11: Fits to σ_R against p_T and $\langle \mu \rangle$, for Tight direct converted photons, for (a) $0 < |\eta| < 0.60$ and (c) $1.81 < |\eta| < 2.37$, along with (b), (d) their pull distributions. An isolation cone size of $\Delta R = 0.4$ was used. Since the widths of the pull distributions are slightly larger than one, the uncertainties on σ_R can be increased by approximately 27 % and 37 %, respectively, following the procedure described in Ref. [3].

Similarly, Figure 4.12 presents the fits to the left-hand width parameter σ_L in the same $|\eta|$ bins. The observed trends and conclusions are consistent with those for σ_R .

Figure 4.13 shows the dependence of the peak position μ_P on p_T and $\langle \mu \rangle$ for unconverted photons in the same $|\eta|$ bin. The fitted peak positions are generally centred around zero, as expected after correction, although small residual deviations remain due to modelling imperfections.

4.4.7 Combination of direct and Bremsstrahlung MC regions

As discussed above, the signal shape is derived from MC samples. However, to construct realistic signal templates, the direct and Bremsstrahlung photon contributions must be combined with appropriate weights. This requires accounting for both the relative cross sections of the two processes and the identification efficiencies for each photon type.

In what follows, ID denotes the photon identification working point (Loose'4 or Tight), and $type$ refers to the photon type (direct or Bremsstrahlung, abbreviated as Brem). The fraction of photons with a given identification working point within a given type, or conversely, the fraction of photons of a given type within a given identification working

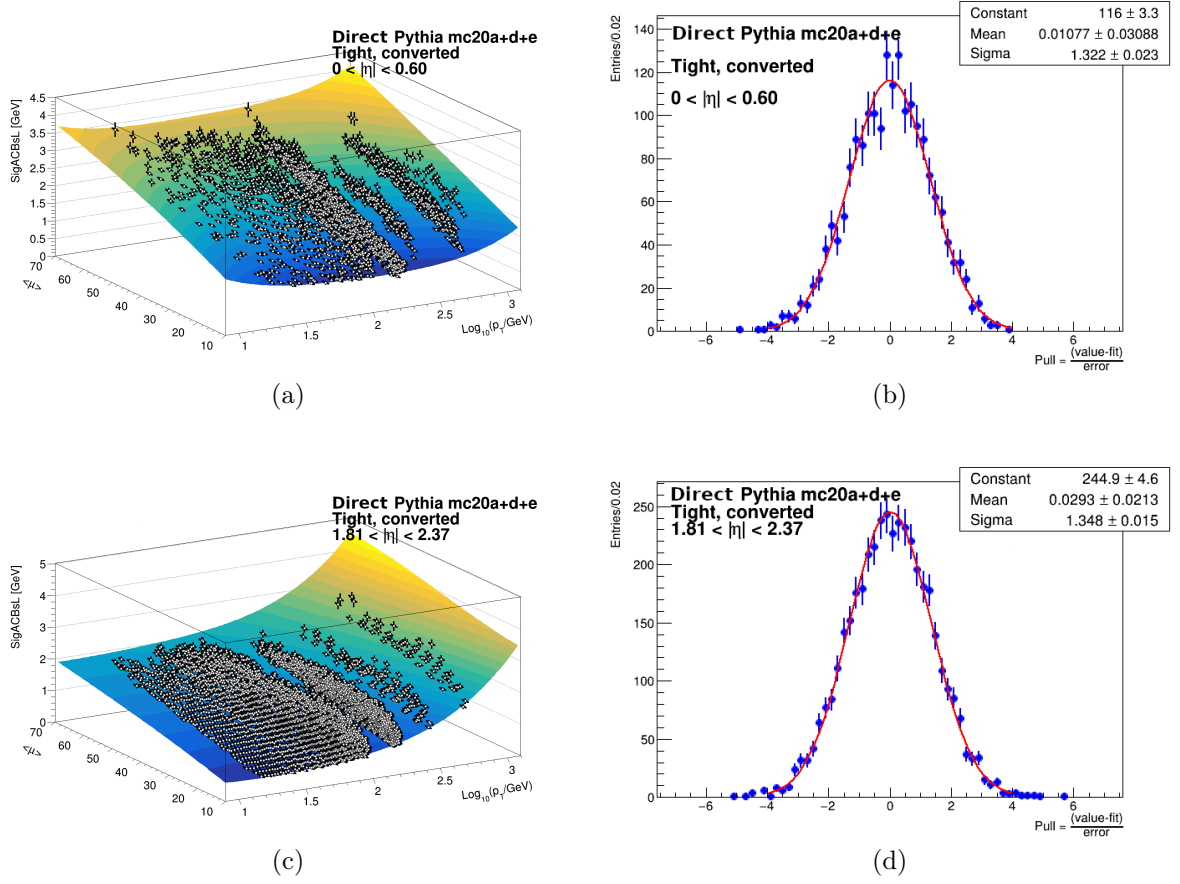


Figure 4.12: Fits to σ_L against p_T and $\langle\mu\rangle$, for Tight direct converted photons, for (a) $0 < |\eta| < 0.60$ and (c) $1.81 < |\eta| < 2.37$, along with (b), (d) their pull distributions. An isolation cone size of $\Delta R = 0.4$ was used. Since the widths of the pull distributions are slightly larger than one, the uncertainties on σ_L can be increased by approximately 32 % and 35 %, respectively, following the procedure described in Ref. [3].

point, is denoted by $c_{\text{subset}}^{\text{set}}$, where *subset* denotes a subset of photons (e.g. Loose'4) within a larger *set* of photons (e.g. direct).

4.4.7.1 Direct vs Bremsstrahlung

To determine the relative fractions of direct and Bremsstrahlung photons, their respective cross sections and filter efficiencies are used, which are functions of p_T . For simulated events, the filter efficiency is defined as the fraction of generated events that pass the selection of interest and are retained. It is used to reweight the nominal cross section provided by the generator, yielding the effective cross section. The ratio of cross section and filter efficiency products,

$$\frac{\sigma_{\text{direct}} \cdot \varepsilon_{\text{direct}}^{\text{filter}}}{\sigma_{\text{Brem}} \cdot \varepsilon_{\text{Brem}}^{\text{filter}}}, \quad (4.8)$$

is fitted as a function of p_T using the parametric form:

$$f(p_T) = a \times \left(\frac{p_T}{\text{GeV}} \right)^b. \quad (4.9)$$

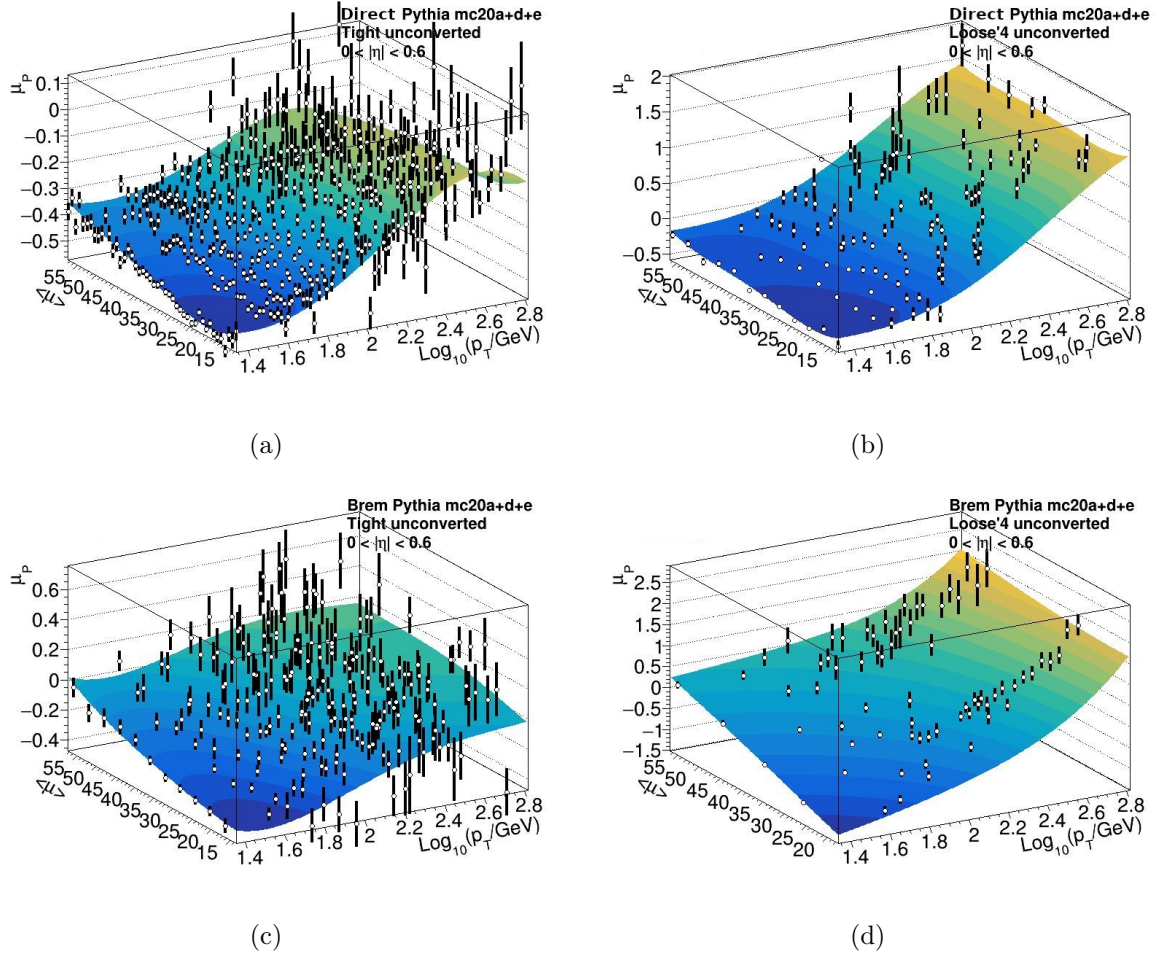


Figure 4.13: Fits to μ_P against p_T and $\langle\mu\rangle$, for $0 < |\eta| < 0.60$ and unconverted photons, for (a) Tight direct, (b) Loose'4 direct, (c) Tight Bremsstrahlung and (d) Loose'4 Bremsstrahlung photons.

From the fit, the following parameters are obtained: $a = 0.149 \pm 0.015$ and $b = 0.411 \pm 0.017$, with the resulting function shown in Figure 4.14.

The fraction of direct photons is then:

$$c_{\text{direct}}^{\text{all}} = \frac{f(p_T)}{1 + f(p_T)} \quad (4.10)$$

$$\text{and } c_{\text{Brem}}^{\text{all}} = 1 - c_{\text{direct}}^{\text{all}}.$$

4.4.7.2 Loose'4 vs Tight, within each photon type

To determine how each identification working point (Loose'4 or Tight) selects photons of a given type, the fractions of Loose'4 and Tight photons among photons of type *type* are computed:

$$c_{\text{Loose'4}}^{\text{type}} = \frac{N_{\text{Loose'4,type}}}{N_{\text{Loose'4,type}} + N_{\text{Tight,type}}} \quad (4.11)$$

and

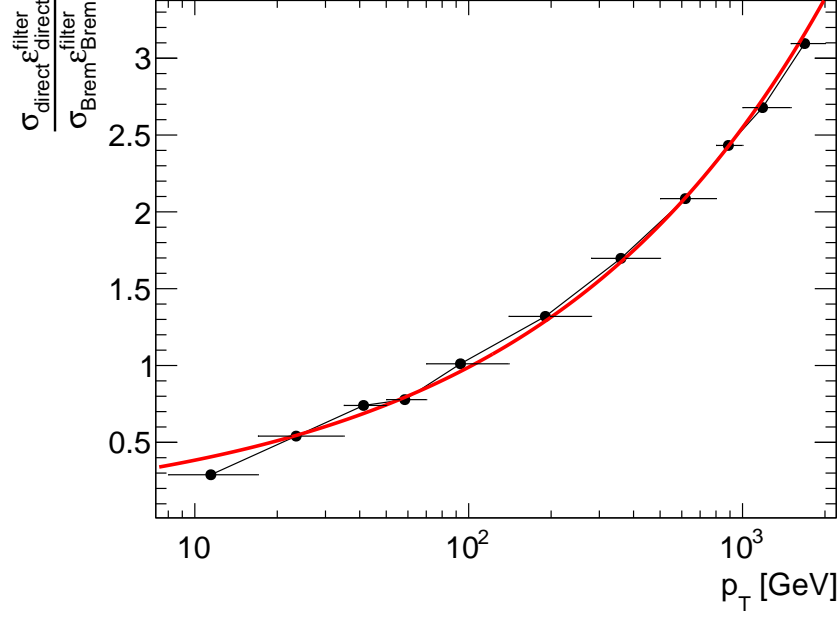


Figure 4.14: Fit to $\frac{\sigma_{\text{direct}}^{\text{filter}}}{\sigma_{\text{Brem}}^{\text{filter}}}$, with $f(p_T) = a \times \left(\frac{p_T}{\text{GeV}}\right)^b$. Fitted parameters: $a = 0.149 \pm 0.015$ and $b = 0.411 \pm 0.017$. $\chi^2/\text{ndf} = 4.59/9$.

$$c_{\text{Tight}}^{\text{type}} = 1 - c_{\text{Loose}'4}^{\text{type}} = \frac{N_{\text{Tight,type}}}{N_{\text{Loose}'4,\text{type}} + N_{\text{Tight,type}}}, \quad (4.12)$$

with $N_{\text{ID,type}}$ the number of photons of type *type* (direct or Bremsstrahlung) passing identification *ID* (Loose'4 or Tight).

4.4.7.3 Combination of direct and Bremsstrahlung within Loose'4 and Tight categories

With the above ingredients, the fractions of direct and Bremsstrahlung photons within each ID category are computed. For a given identification working point *ID*, the direct and Bremsstrahlung fractions are given by:

$$c_{\text{direct}}^{\text{ID}} = \frac{c_{\text{direct}}^{\text{all}} \cdot c_{\text{ID}}^{\text{direct}}}{c_{\text{direct}}^{\text{all}} \cdot c_{\text{ID}}^{\text{direct}} + c_{\text{Brem}}^{\text{all}} \cdot c_{\text{ID}}^{\text{Brem}}}, \quad c_{\text{Brem}}^{\text{ID}} = 1 - c_{\text{direct}}^{\text{ID}} \quad (4.13)$$

$c_{\text{direct}}^{\text{Tight}}$ and $c_{\text{direct}}^{\text{Loose}'4}$ are fitted as two-dimensional functions of p_T and $\langle\mu\rangle$, yielding four 2D continuous functions: $c_{\text{direct}}^{\text{Tight}}$, $c_{\text{Brem}}^{\text{Tight}}$, $c_{\text{direct}}^{\text{Loose}'4}$, and $c_{\text{Brem}}^{\text{Loose}'4}$. Although the underlying dependence is primarily on p_T , a 2D parametrisation is adopted for consistency across the analysis. Figure 4.15 shows the fitted distributions for the direct component.

Finally, in each $p_T\text{--}\langle\mu\rangle$ bin, the full MC probability density function for a given ID working point is constructed as:

$$\begin{aligned} \text{PDF}_{\text{MC,ID}}(p_T, \langle\mu\rangle) = & c_{\text{direct}}^{\text{ID}}(p_T, \langle\mu\rangle) \cdot \text{ACB}_{\text{MC,ID,direct}}(p_T, \langle\mu\rangle) \\ & + c_{\text{Brem}}^{\text{ID}}(p_T, \langle\mu\rangle) \cdot \text{ACB}_{\text{MC,ID,Brem}}(p_T, \langle\mu\rangle), \end{aligned} \quad (4.14)$$

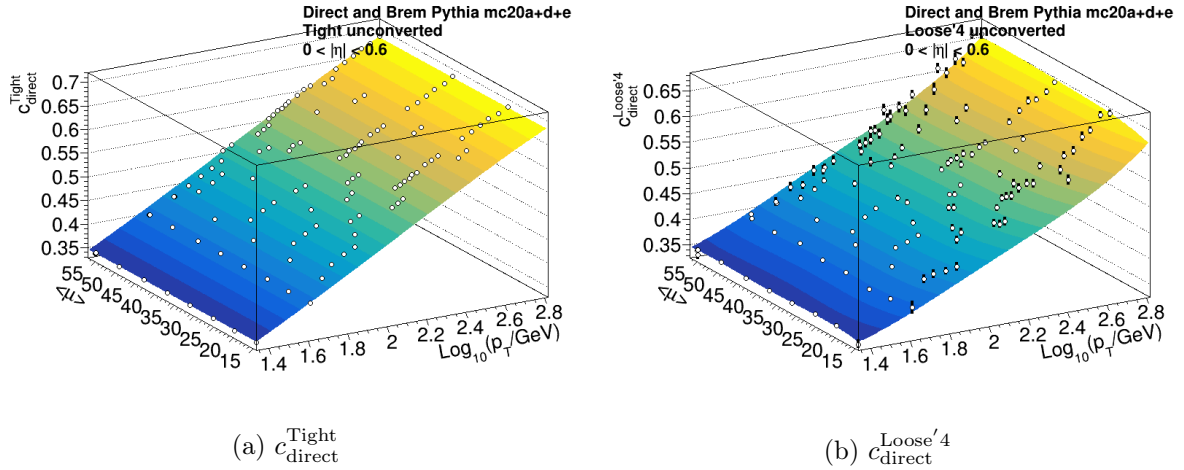


Figure 4.15: (a) $c_{\text{direct}}^{\text{Tight}}$ and (b) $c_{\text{direct}}^{\text{Loose'4}}$ fits against p_T and $\langle\mu\rangle$ for unconverted photons, $0 < |\eta| < 0.6$.

where $\text{ACB}_{\text{MC,ID,type}}(p_T, \langle\mu\rangle)$ denotes the ACB model fitted to MC photons of type *type* in the identification category *ID*, within the given bin.

4.4.8 Fit to data

The Tight and Loose'4 data distributions are fitted simultaneously. The relative signal yield between the Tight and Loose'4 selections is constrained using the corresponding ratio derived from simulation. The fraction of Tight photons, $N_{\text{Tight}}/(N_{\text{Tight}} + N_{\text{Loose'4}})$, is modelled as a smooth function of p_T and $\langle\mu\rangle$ via a series expansion. An example of the 2D distribution and its fit is shown in Figure 4.16, for unconverted photons in the central detector region.

The signal shapes in data are constrained to follow those obtained from the corresponding MC distributions: the parameters of the Tight (Loose'4) signal in data are fixed to the Tight (Loose'4) ones in MC, except for the peak position in the Tight region. This free parameter allows the extraction of the data-driven shift. The background component is modelled with a single PDF (an ACB), common to both the Loose'4 and Tight data regions. A schematic overview of the full procedure is presented in Figure 4.4.

Examples of fitted distributions for selected bins are shown in Figure 4.17.

Some bins show varying levels of background contamination. In a few cases, the peak position appears to be underestimated. This may stem from an imperfect modelling of the resolution in MC, a mismatch in pile-up conditions between data and simulation, or intrinsic limitations of the fitting model. However, the use of narrower bins generally improves the agreement.

4.4.9 Data-driven shift modelling

The data-driven shifts are parametrised as a continuous two-dimensional analytical function of p_T and $\langle\mu\rangle$, using a series expansion. The resulting fit, for unconverted photons in the central region of the detector, along with the pull distribution, defined as the difference between the data and the fitted model, are shown in Figure 4.18. The error on the fitted data-driven shift varies with p_T and $\langle\mu\rangle$. For instance, at $p_T = 100$ GeV and

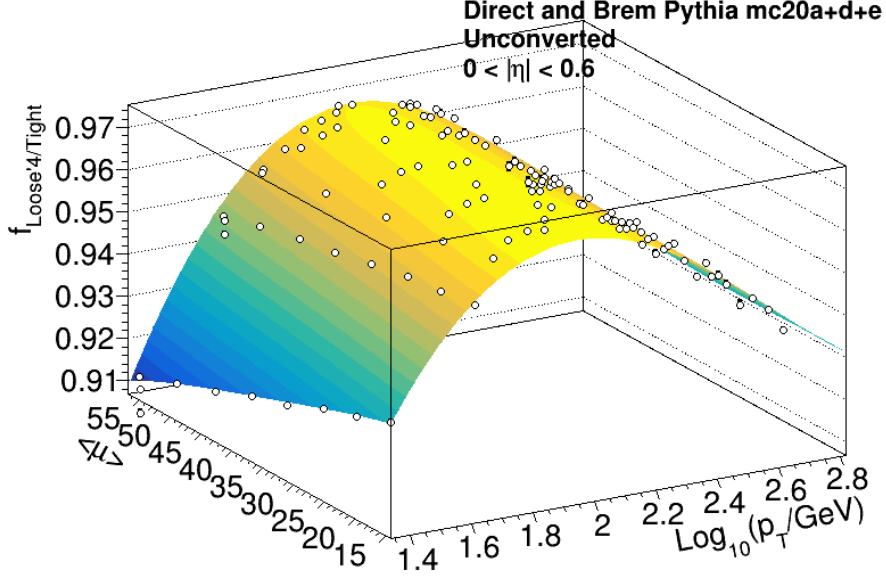


Figure 4.16: Fit to $N_{\text{Tight}}/(N_{\text{Tight}} + N_{\text{Loose'4}})$ for unconverted photons, and $0 < |\eta| < 0.6$, with a series expansion.

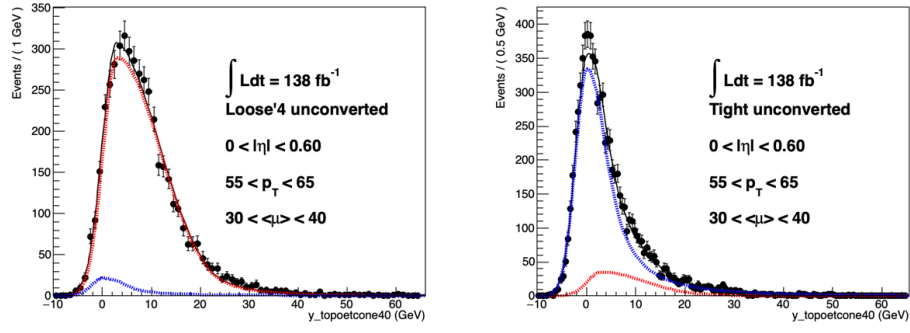
$\langle\mu\rangle = 35.5$, the error is 11 MeV. At low p_T (25 GeV) and low $\langle\mu\rangle$, the error increases to 26 MeV. At low p_T and high $\langle\mu\rangle$, the error increases further, reaching 52 MeV.

This approach, which provides continuous data-driven shifts, represents a notable improvement over previous methods. In earlier approaches, separate fits were performed in a limited number of coarse bins, often leading to large statistical uncertainties. In contrast, the current strategy enables a global fit across the full phase space, significantly reducing statistical fluctuations and yielding a smooth, continuous correction function in both p_T and pile-up. For reference, previous data-driven shift estimates are shown in Figure 4.19.

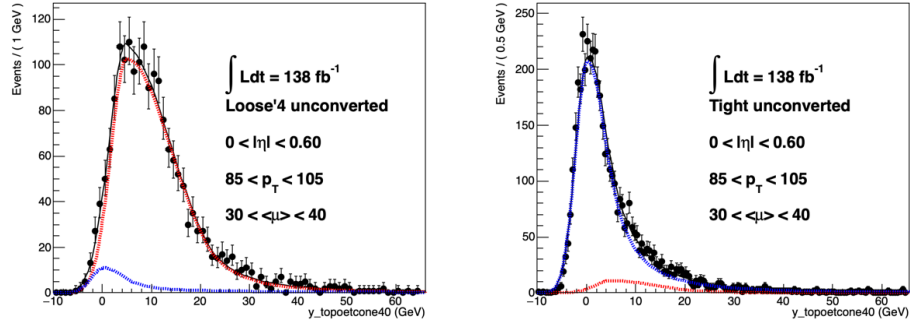
The final correction must combine the results obtained from this method (using QCD photons) with those derived from radiative Z -boson decays.

4.4.10 Conclusion

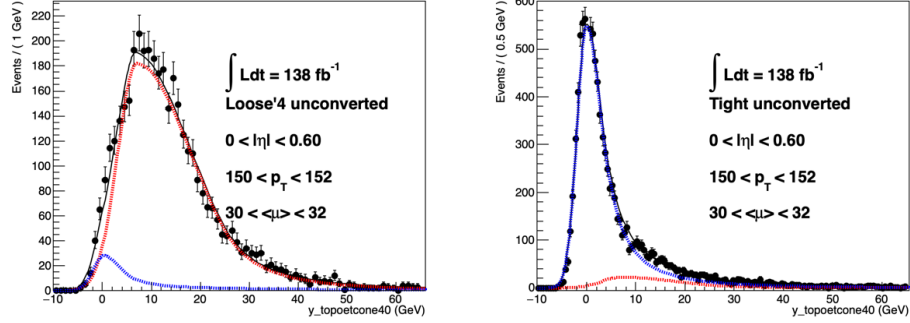
Calorimetric isolation energy is an essential quantity to identify prompt photons and reject fake ones. Due to discrepancies between simulation and data, the isolation energy distribution must be corrected in the simulation. The methodology developed in this thesis enables the continuous correction of the isolation energy as a function of p_T and $\langle\mu\rangle$, in bins of $|\eta|$ and photon conversion type, taking into account the two photon origins, direct or Bremsstrahlung, offering greater accuracy for Run 2 than previous approaches. This work is still ongoing, but it is close to being finalised.



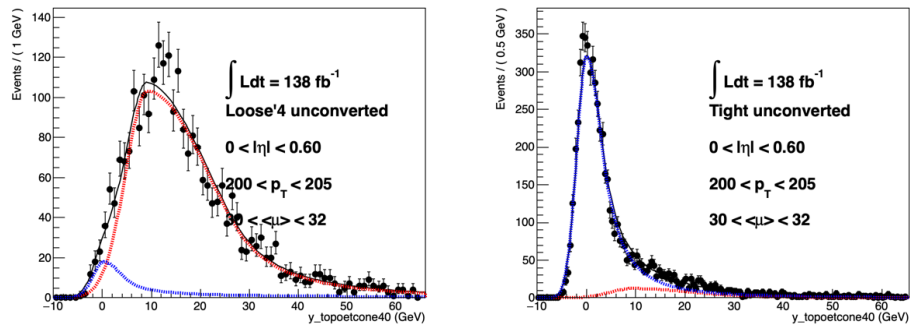
(a)



(b)



(c)



(d)

Figure 4.17: Fits to data for unconverted photons, $0 < |\eta| < 0.6$, for several p_T and $\langle \mu \rangle$ bins. The signal (blue line) and the background (red line) components are shown.

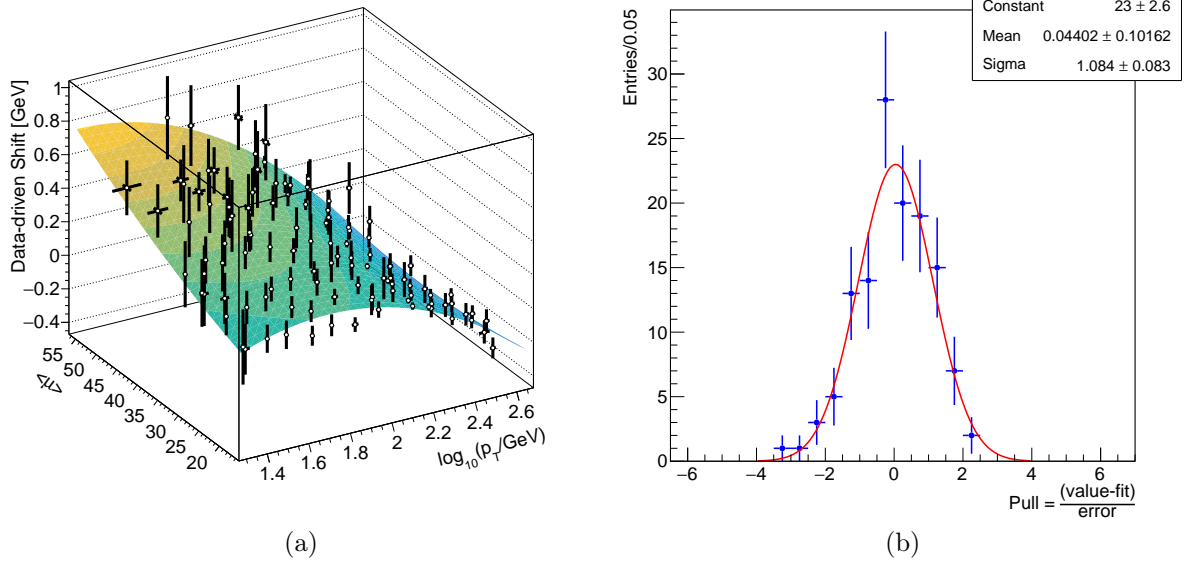


Figure 4.18: (a) Fit to data-driven shift against p_T and $\langle\mu\rangle$ and (b) pull distribution, for unconverted photons in $|\eta| < 0.6$.

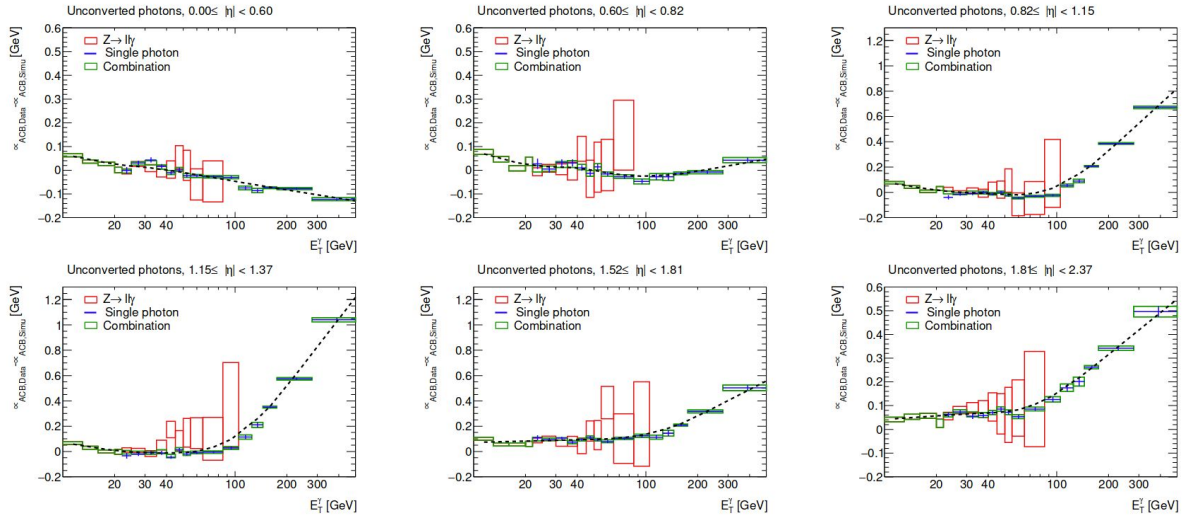


Figure 4.19: Former data-driven shifts [128].

Chapter 5

Neural network principles

Contents

5.1	Overview	106
5.2	Perceptron	107
5.3	Gradient descent	108
5.4	Training and validation	112
5.5	Optimisation	113
5.6	More complex architectures	114
5.6.1	Convolutional neural networks	114
5.6.2	Graph neural networks	117
5.6.3	Set Transformers	119
5.7	Machine Learning in this work	125

This chapter introduces the basics of neural networks and machine learning. It begins with the fundamentals, defining the perceptron, fully connected neural networks, and gradient descent. It then describes more advanced architectures used in this thesis: convolutional neural networks, graph neural networks, and Set Transformers.

5.1 Overview

Neural networks are computational models designed to emulate certain aspects of brain function by interconnecting simple processing units, often referred to as “neurons” or “nodes”. These models are particularly effective for complex pattern recognition and decision-making tasks. Simon Haykin provides the following definition [139]:

A neural network is a massively parallel distributed processor made up of simple processing units that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects:

1. Knowledge is acquired by the network from its environment through a learning process.
2. Interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge.

Neural networks are used for a large variety of tasks, such as regression, classification, or clustering, through a process called *learning*. In high-energy physics (HEP), they are mainly employed for classification tasks, such as separating signal and background or identifying processes beyond the SM.

The learning process consists of adjusting the node weights to improve the performance of the network. Most models rely on a fixed architecture, i.e., a fixed structure of nodes, connections, and functions, only modifying the node weights; however, some more complex models can dynamically adapt their topology during the learning process. Learning can be classified as *supervised*, in which the training inputs are labelled with their corresponding category, or *unsupervised* otherwise. In the latter case, the neural network tries to identify patterns or structure within the data. Midway between both, *semi-supervised* learning combines labelled and unlabelled data. In HEP, experimental data often lack inherent labels. Thus, most analyses rely on simulated data, with events typically labelled as 1 and 0 for signal and background events, respectively.

The structure of a simple neural network comprises several layers of interconnected nodes, as depicted in Figure 5.1. The first layer, or *input layer*, receives the input features¹ used in the learning process. The last layer, or *output layer*, returns a prediction as close as possible to the input label, if the inputs are labelled. Between them, *hidden layers* allow the network to learn complex relationships between the inputs and outputs. When all nodes in one layer are connected to all nodes in the subsequent layer, the network is referred to as a *fully connected* or *dense neural network*. A network with several hidden layers is called a *deep neural network* (DNN); this architecture allows the network to capture complex patterns and correlations in the data.

¹Input features are also often referred to as input variables.

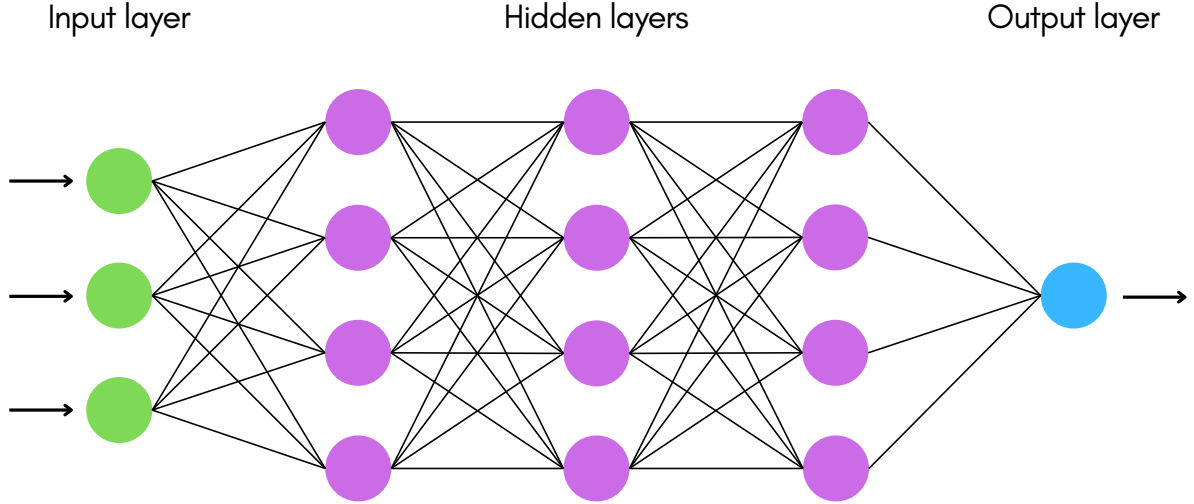


Figure 5.1: Fully connected neural network representation.

5.2 Perceptron

In supervised binary classification, the fundamental building block of a neural network is the *perceptron*, an algorithm introduced by Rosenblatt in 1958 [140]. It comprises a set of weights, a bias term, and an activation function, as illustrated in Figure 5.2. Each input feature x_j is associated with a weight w_j , which encodes the importance assigned to that feature during the process. The perceptron computes a weighted sum of the input features, similar to the activation potential of a biological neuron:

$$s = \mathbf{w}^T \mathbf{x} + b \quad (5.1)$$

$$= \sum_j^{N_{\text{features}}} w_j x_j + b \quad (5.2)$$

where $\mathbf{w} = (w_j)_{1 \leq j \leq N_{\text{features}}}$ is the vector of weights, $\mathbf{x} = (x_j)_{1 \leq j \leq N_{\text{features}}}$ the input features, and b the bias term. The final output \hat{y} of the perceptron is obtained by applying an activation function to this weighted sum:

$$\hat{y} = \varphi(s). \quad (5.3)$$

Several activation functions are commonly used, each with its own advantages and drawbacks. In particular, the step function, the sigmoid function, the hyperbolic tangent (tanh), the Rectified Linear Unit (ReLU), the leaky Rectified Linear Unit (leaky ReLU), and the Exponential Linear Unit (ELU) are popular choices. They are shown in Figure 5.3.

The perceptron can be considered the simplest neural network. However, it is often necessary to connect multiple perceptrons across several layers, in a *multi-layer perceptron* (MLP), to capture the full complexity of the data. The output o_k^l of the k -th node in the l -th layer is given by:

$$o_k^l = \varphi_k^l \left(\sum_j w_{j,k}^l o_j^{l-1} + b_k^l \right), \quad (5.4)$$

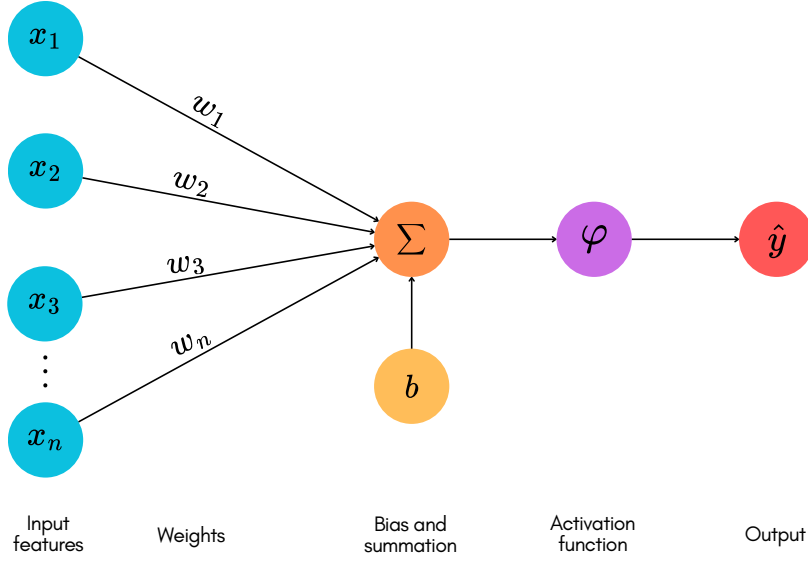


Figure 5.2: Representation of a perceptron.

where the summation runs over the nodes in the previous layer $l - 1$, φ_k^l denotes the activation function of the perceptron (which is usually identical for all perceptrons in the network, $\varphi_k^l = \varphi$), $w_{j,k}^l$ represents the weights associated with the outputs o_j^{l-1} from the previous layer, and b_k^l is the corresponding bias term. This expression highlights the structure of a feedforward neural network, in which information propagates layer by layer, each node applying a non-linear transformation to a weighted sum of the outputs of the previous layer.

In classification tasks, the output $\hat{\mathbf{y}}$ of the MLP is a vector of dimension N_{classes} , each component representing the probability that the input belongs to the corresponding class. To interpret these values as probabilities, they must be positive and sum to one. A *softmax* function is usually applied to the output vector:

$$\text{softmax}(\hat{\mathbf{y}})_c = \frac{e^{\hat{y}_c}}{\sum_{j=1}^{N_{\text{classes}}} e^{\hat{y}_j}} \text{ for } 1 \leq c \leq N_{\text{classes}}, \quad (5.5)$$

where $\text{softmax}(\hat{\mathbf{y}})_c$ represents the predicted probability that the input belongs to class c .

For binary classification, a similar structure can be used, with a single output node that directly represents the probability for the input to belong to the class of interest. In HEP, this typically represents the probability of being a signal event. In this configuration, the single output must be a real number in the range $[0, 1]$. A sigmoid function is generally applied to satisfy this requirement.

5.3 Gradient descent

The training phase of a neural network involves iteratively updating the weights to make predictions as close as possible to the true labels. The discrepancy between the prediction and the true label (i.e. the class, in classification tasks) is measured by a *loss function*.

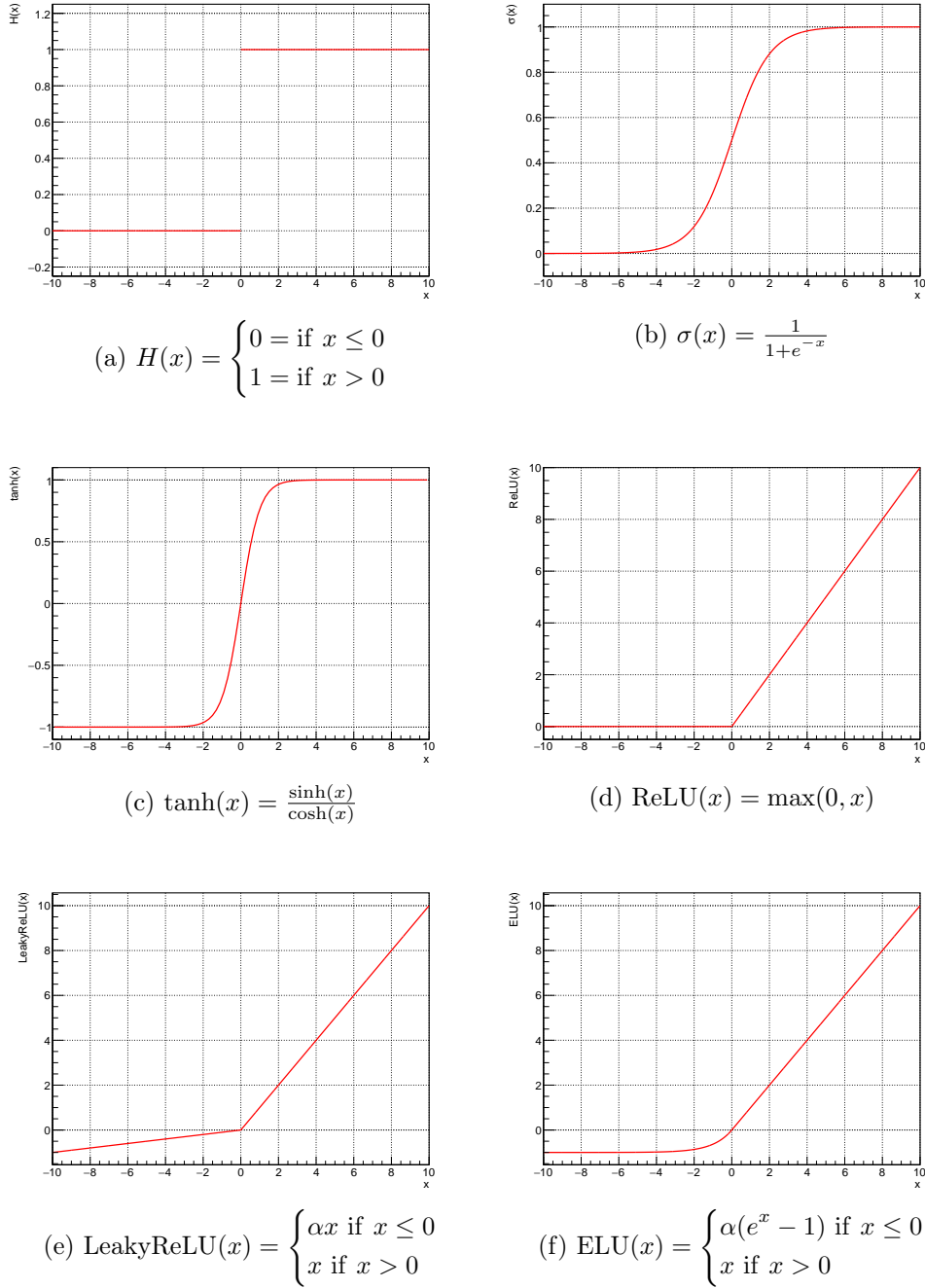


Figure 5.3: Usual activation functions, including (a) the Heavyside step function, (b) the sigmoid function, (c) the hyperbolic tangent, (d) the Rectified Linear Unit, (e), the leaky Rectified Linear Unit (with $\alpha = 0.1$), and (f) the Exponential Linear Unit (with $\alpha = 1$).

Before introducing widely used loss functions, the necessary notations are defined. In a dataset containing N inputs (or samples), the i -th sample is represented as \mathbf{x}^i , a vector with N_{features} components:

$$\mathbf{x}^i = (x_j^i)_{1 \leq j \leq N_{\text{features}}} . \quad (5.6)$$

The following notations are used: an upper i index refers to one input in a dataset of N inputs, while a lower j index refers to a specific feature of the inputs (or a specific

class for \mathbf{y}^i and $\hat{\mathbf{y}}^i$, see below).

Each input \mathbf{x}^i is associated with a true label \mathbf{y}^i , and a predicted output $\hat{\mathbf{y}}^i$ from the network. These last two are also vectors, whose dimension corresponds to the number of output nodes in the final layer (equivalent to the number of classes, in a classification task):

$$\mathbf{y}^i = (y_j^i)_{1 \leq j \leq N_{\text{classes}}} \quad (5.7)$$

and

$$\hat{\mathbf{y}}^i = (\hat{y}_j^i)_{1 \leq j \leq N_{\text{classes}}} . \quad (5.8)$$

For a single input feature or a single output class, these notations can be simplified by omitting the corresponding index.

A simple example of a loss function is the squared difference between the predicted and true values:

$$L(\hat{y}, y) = (\hat{y} - y)^2. \quad (5.9)$$

Averaging this quantity over all training samples, for a single output network, gives the *mean squared error* (MSE) commonly used for regression tasks:

$$\text{MSE}(\hat{\mathbf{y}}, \mathbf{y}) = \frac{1}{N} \sum_{i=1}^N (\hat{y}^i - y^i)^2. \quad (5.10)$$

It is important to note that the predictions, and consequently the loss function, depend on the trainable parameters of the neural network (namely, the weights and biases) collectively denoted as $\boldsymbol{\theta}$. More explicitly, if the model output is denoted as $h_{\boldsymbol{\theta}}(\mathbf{x}^i)$, the loss function can be written as:

$$\text{MSE}_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{y}) = \frac{1}{N} \sum_{i=1}^N (h_{\boldsymbol{\theta}}(\mathbf{x}^i) - y^i)^2. \quad (5.11)$$

For classification tasks, the *categorical cross-entropy* loss function is typically employed:

$$L_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{y}) = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{N_{\text{classes}}} y_j^i \log(\hat{y}_j^i), \quad (5.12)$$

where y_j^i represents the true label (encoded as a one-hot vector) and \hat{y}_j^i is the predicted probability for class j . The dependence of predictions on $\boldsymbol{\theta}$ and on the input is omitted here for brevity.

For binary classification, the problem simplifies to a single output node representing the probability of a given sample belonging to the positive class. In this case, the appropriate loss function is the *binary cross-entropy* (BCE):

$$\text{BCE}_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{y}) = -\frac{1}{N} \sum_{i=1}^N [y^i \log(\hat{y}^i) + (1 - y^i) \log(1 - \hat{y}^i)]. \quad (5.13)$$

The training process involves two main phases:

1. **Forward pass** – The input data are propagated through the network, generating predictions, and the loss is computed.
2. **Backward pass** – The gradient of the loss function with respect to the network parameters is computed using backpropagation (defined just below). This gradient is then used to update the node weights. An optimisation algorithm is usually employed, such as stochastic gradient descent (SGD) [141] or Adam [142].

By iteratively performing these updates, the neural network gradually refines its parameters to improve predictive performance on the given task.

The backpropagation process is explained step by step below, following the definition of the notations. The weight connecting the j -th node of the $(l - 1)$ -th layer to the k -th node of the l -th layer is denoted as $w_{j,k}^l$, with $1 < l < L$, $1 < j < N_{l-1}$, and $1 < k < N_l$, where N_l represents the number of nodes in the l -th layer. The weighted sum computed by the k -th node in the l -th layer, along with its final output, is given by:

$$\begin{cases} s_k^l = \sum_{j=1}^{N_l} w_{j,k}^l o_j^{l-1} + b_k^l \\ o_k^l = \varphi(s_k^l) \end{cases} \quad (5.14)$$

where φ is the activation function of the node. The index i of the input sample is omitted for clarity. For the first layer ($l = 1$), s_k^1 is the weighted sum of the input features:

$$s_k^1 = \sum_{j=1}^{N_{\text{features}}} w_{j,k}^1 x_j + b_k^1 \quad (5.15)$$

The gradient of the loss function with respect to all network weights, considering a single input, is then given by:

$$\nabla L_{\theta} = \left(\frac{\partial L_{\theta}}{\partial w_{j,k}^l} \right)_{\substack{1 < l < L, \\ 1 < j < N_{l-1}, \\ 1 < k < N_l}}. \quad (5.16)$$

These derivatives involve composed functions across multiple layers. Using the chain rule, they can be expanded as:

$$\frac{\partial L_{\theta}}{\partial w_{j,k}^l} = \frac{\partial L_{\theta}}{\partial o_k^l} \frac{\partial o_k^l}{\partial s_k^l} \frac{\partial s_k^l}{\partial w_{j,k}^l}. \quad (5.17)$$

The last term is straightforward:

$$\frac{\partial s_k^l}{\partial w_{j,k}^l} = \frac{\partial}{\partial w_{j,k}^l} \left[\sum_{j'=1}^{N_l} w_{j',k}^l o_{j'}^{l-1} + b_k^l \right] \quad (5.18)$$

$$= o_j^{l-1}. \quad (5.19)$$

The second term corresponds to the derivative of the activation function:

$$\frac{\partial o_k^l}{\partial s_k^l} = \varphi' \left(s_k^l \right). \quad (5.20)$$

The first term requires further decomposition and can be directly computed only for the output layer, where the outputs of the nodes correspond to the final predictions:

$$\frac{\partial L_{\theta}}{\partial o_k^L} = \frac{\partial L_{\theta}}{\partial \hat{y}_k}. \quad (5.21)$$

Here, k refers to the k -th output node, which represents a class in a classification problem.

For the MSE, this derivative simplifies to:

$$\frac{\partial L_{\theta}}{\partial \hat{y}_k} = \frac{\partial}{\partial \hat{y}_k} (\hat{y}_k - y_k)^2 = 2 (\hat{y}_k - y_k). \quad (5.22)$$

For $l \neq L$, another application of the chain rule gives:

$$\frac{\partial L_{\theta}}{\partial o_k^l} = \sum_{p=1}^{N_{l+1}} \frac{\partial L_{\theta}}{\partial o_p^{l+1}} \frac{\partial o_p^{l+1}}{\partial s_p^{l+1}} \frac{\partial s_p^{l+1}}{\partial o_k^l}. \quad (5.23)$$

Again, the last two derivatives can be computed explicitly:

$$\frac{\partial o_p^{l+1}}{\partial s_p^{l+1}} = \varphi' \left(s_p^{l+1} \right), \quad (5.24)$$

and

$$\frac{\partial s_p^{l+1}}{\partial o_k^l} = w_{k,p}^{l+1}. \quad (5.25)$$

Since the derivative of the loss function with respect to the output layer has already been computed, the gradients for the previous layer ($L - 1$) can be determined, allowing for the calculation of the derivatives with respect to the weights $w_{j,k}^{L-1}$. By propagating this process layer by layer from the output to the input, the full set of gradients with respect to all network parameters can be obtained.

Once the gradient has been computed, its value is averaged over all inputs, and the network weights are updated according to the following rule:

$$w_{j,k}^l \leftarrow w_{j,k}^l - \eta \frac{\partial L_{\theta}}{\partial w_{j,k}^l}, \quad (5.26)$$

where η is a positive real number known as the *learning rate*, and represents the magnitude of the weight update. It is explained in more detail in Section 5.5.

5.4 Training and validation

Typically, data are divided into three independent subsets:

- The **training dataset** is used to update the neural network's parameters. Data are processed in subsets called *batches*: the loss function is computed for each input in the batch, and then the network's weights are adjusted.

- The **validation dataset** is used to evaluate the model's performance after each pass through the full training dataset, referred to as an *epoch*. It provides an independent measurement of the model's performance and allows for stopping the training when the model starts overfitting on the training inputs, i.e., becoming too adapted to the training set and losing generality.
- The **test dataset** is kept completely separate from the training process and is used once, at the end of the process, to assess the final model's performance.

Training and validation steps are typically repeated until the performance of the model starts to decrease, as indicated by an increase in the validation loss. The final selected model is the one with the lowest validation loss.

5.5 Optimisation

The performance and convergence speed of a neural network can be significantly improved by tuning various *hyperparameters*. The most important ones are presented below.

Learning rate As introduced above, the learning rate is the magnitude by which the network weights are updated at each iteration, i.e. each time all inputs in a batch have been processed. It is typically chosen in the range $[0, 1]$. This parameter plays a crucial role in the training process: it must be small enough to ensure convergence to the minimum without overshooting it, yet large enough to prevent excessively slow training or getting trapped in a local minimum. Some optimisation algorithms, like Adam and RMSprop, use an adaptive learning rate, which decreases as the loss approaches the minimum.

Momentum In the gradient descent, momentum helps speed up convergence and prevents being stuck in a local minimum. It is analogous to physical momentum, where the velocity of a system depends on both external forces and its previous velocity. At each iteration t , the network weights are updated as:

$$\Delta \mathbf{w}_t = -\eta \nabla L_{\theta} + p \Delta \mathbf{w}_{t-1}, \quad (5.27)$$

where p is the momentum coefficient.

Batch size The batch size N_{batch} determines how many inputs are processed before updating the weights. It has a direct impact on convergence speed and generalisation:

- **Small batch sizes** result in more frequent updates and can improve generalisation by introducing stochasticity into the process. However, they can lead to noisy gradient updates and prevent convergence.
- **Large batch sizes** lead to more stable convergence and allow for greater parallelisation, but can be computationally expensive. They may also generalise poorly, as they approximate the gradient more closely and do not benefit from the stochasticity of small batches.

Regularisation Regularisation techniques aim to improve performance by avoiding overfitting during training, often without reducing the training error [143]. Among several approaches, weight regularisation is commonly used.

L^2 regularisation, or weight decay, penalises large weight values by adding a quadratic penalty term to the loss function:

$$L_{\theta} \rightarrow L_{\theta} + \lambda \|\mathbf{w}\|_2^2 \quad (5.28)$$

where \mathbf{w} represents the vector of weights, λ is a hyperparameter that controls the regularisation strength, and $\|\cdot\|_2$ is the Euclidean norm: $\|\mathbf{x}\|_2 = \sqrt{\sum_i x_i^2}$. This technique effectively reduces overfitting by discouraging excessively large weights, promoting smoother decision boundaries and improving generalisation.

L^1 regularisation instead penalises the absolute values of the weights, modifying the loss function as follows:

$$L_{\theta} \rightarrow L_{\theta} + \lambda \|\mathbf{w}\|_1 \quad (5.29)$$

where $\|\cdot\|_1$ is the “taxicab norm”, defined as: $\|\mathbf{x}\|_1 = \sum_i |x_i|$.

Unlike L^2 regularisation, which scales linearly in the gradient calculation, L^1 regularisation induces sparsity in the model by forcing some weights to zero. This property is especially useful for selecting features in the data and for obtaining a compressed representation in the case of high-dimensional data.

Dropout mitigates overfitting by randomly deactivating neurons during training, setting their outputs to zero.

A useful tool to assess the performance of a model is the Receiver Operating Characteristic (ROC) curve. In the case of a signal-background classification task, it shows the signal selection efficiency as a function of the background selection efficiency. It provides information about the model’s discrimination power: the larger the area under the curve, the better the classifier’s performance.

5.6 More complex architectures

Beyond fully connected neural networks, more complex architectures have been developed for specific tasks. This section presents the architectures used in this work.

5.6.1 Convolutional neural networks

Convolutional neural networks (CNNs) [144] are particularly well-suited to data with a grid-like topology [143], such as time-series data (which can be represented as a 1D grid), images (which form a 2D grid of pixels), and even higher-dimensional structured data.

The key feature of CNNs is the convolutional operation. This mathematical operation allows for capturing correlations between neighbouring elements. Being translation-invariant, it can identify structures and patterns regardless of their position in the data. Using a limited set of weights instead of a weight for each pixel in the image, CNNs significantly reduce computational costs.

A CNN consists of multiple processing steps, which are illustrated in Figure 5.4 and detailed below, using a 2D image as an example.

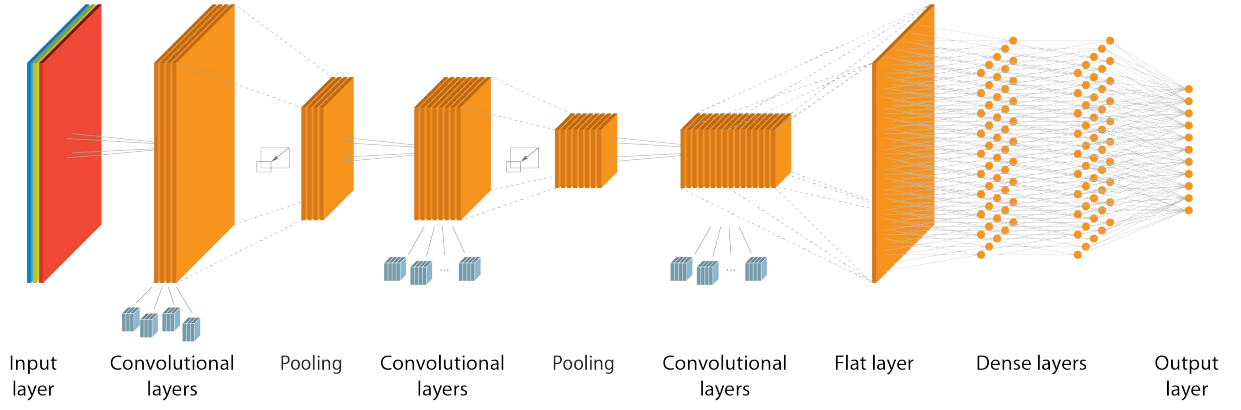


Figure 5.4: A CNN representation, from Ref. [145].

Convolution The convolutional operation transforms the initial image of size $N_x \times N_y$ by sliding a *convolutional window* (also called *kernel* or *filter*) of size $k_x \times k_y$ over the image, applying the following operation:

$$x'_{i,j} = \sum_{r,s} W_{r,s} x_{i+r,j+s} + b \quad (5.30)$$

where W is the convolution kernel and b is the bias term. The summation runs over the kernel dimensions. This process is illustrated in Figure 5.5.

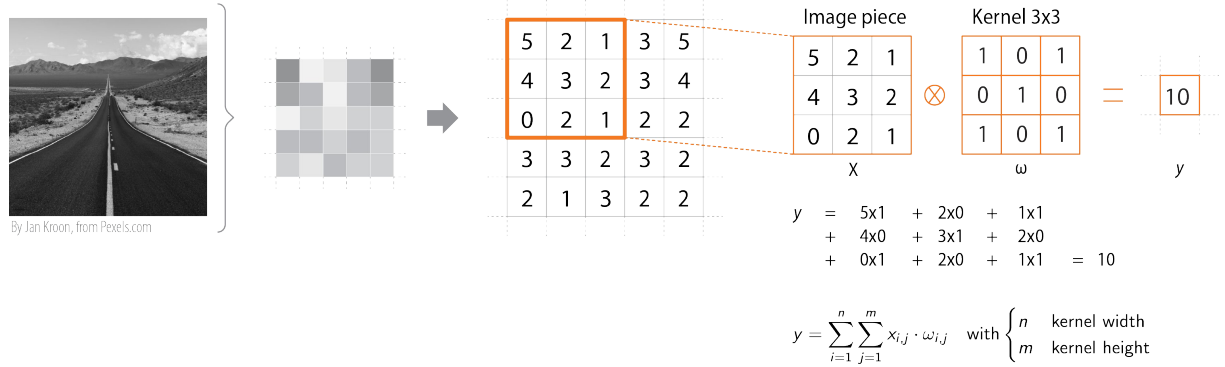


Figure 5.5: A convolution example, from Ref. [145].

A kernel acts as a feature detector, capturing specific patterns in the data. Multiple kernels can be used to identify different structures. An activation function is typically applied after the convolutional operation to add non-linearity to the process. A convolutional layer is illustrated in Figure 5.6.

Stride The stride defines how many rows and columns the kernel is shifted by between two convolution operations. For instance, strides of (s_x, s_y) move the kernel of s_x pixels horizontally and s_y pixels vertically. A stride larger than one decreases the size of the feature map compared to the initial image. This concept is illustrated in Figure 5.7.

Padding Padding allows control over the size of the output by surrounding the input image with additional zeros. It prevents excessive shrinking and enables convolution operations to be applied at the edges of the image. If N_{padding} pixels are added around the image, the horizontal size of the output feature map is given by:

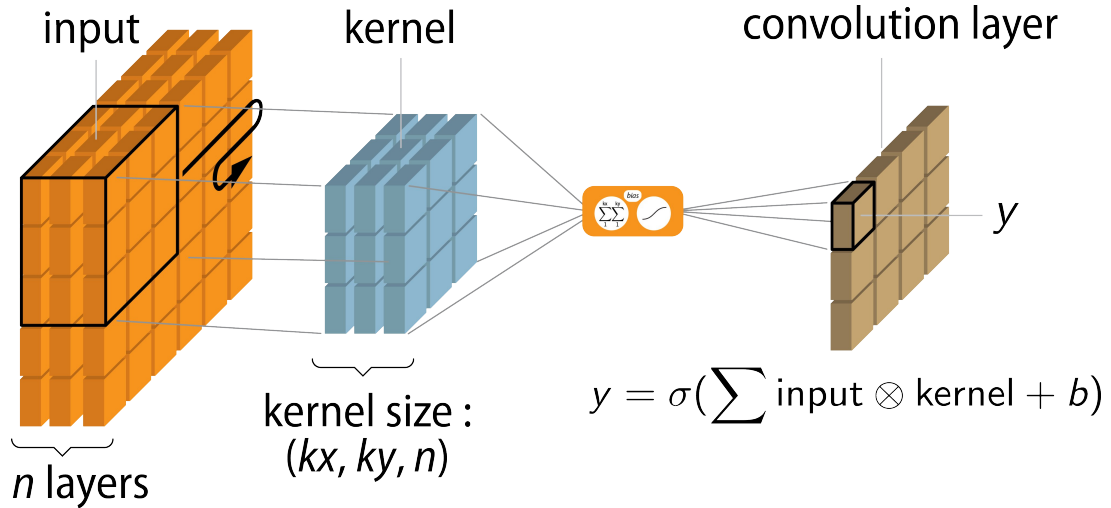
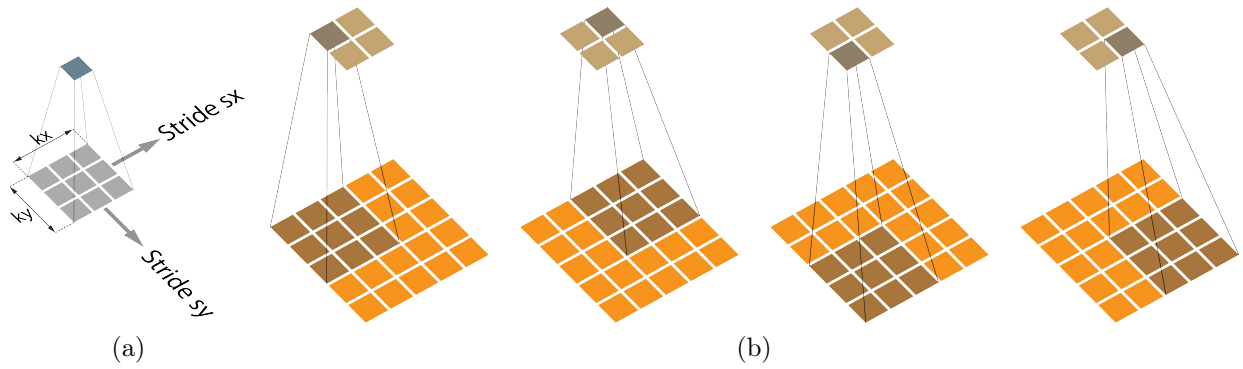


Figure 5.6: Representation of a convolutional layer with a 3D kernel, from Ref. [145].


 Figure 5.7: (a) Stride representation and (b) a convolutional kernel sliding over an image with a stride of 2×2 . Images from Ref. [145].

$$N'_x = \left\lfloor \frac{N_x - k_x + 2N_{\text{padding}}}{s_x} \right\rfloor + 1. \quad (5.31)$$

A similar formula applies to the vertical size. Several padding strategies can be used, such as reducing the size of the feature map compared to the input image (no padding) or maintaining the original size. This is illustrated in Figure 5.8.

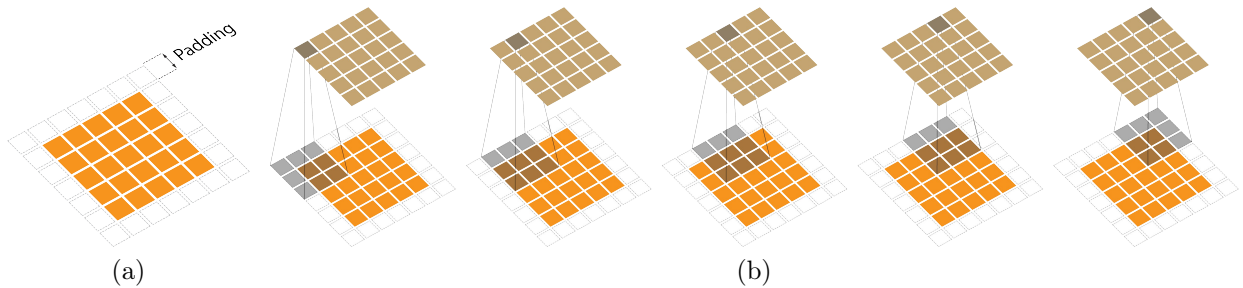


Figure 5.8: (a) Padding representation and (b) a convolutional kernel sliding over an image with a padding of 1. Images from Ref. [145].

Pooling After the convolution operations, pooling is typically applied. It allows for both further reducing the dimensionality, thus reducing the computational complexity, and improving translation invariance. To perform a pooling operation, the feature map is divided into patches of size $p \times p$, and a single value is assigned to each of them. Different criteria can be used:

- *Max pooling*, which keeps the maximum value within each patch, emphasises the strongest features.
- *Average pooling*, which computes the mean value within each patch, leads to a smoother representation.

Pooling techniques are of primary interest for capturing the most important features in the data by reducing the number of parameters.

5.6.2 Graph neural networks

Graph neural networks (GNNs) [146] have been designed to process *graphs*, which are non-Euclidean structures modelling objects as *nodes* and their relationships as *edges*. GNNs are widely used for tasks such as classification, clustering, and link prediction.

Classical neural networks have difficulty handling irregular structures such as graphs, in which the number of edges can vary between nodes, and the number of nodes is not constant across graphs. GNNs address this challenge by leveraging message-passing mechanisms, which allow for extracting meaningful representations from graph data.

5.6.2.1 Graphs

A graph G is defined as a pair (V, E) , where $V = \{v_i\}_{1 \leq i \leq N}$ is a set of N *vertices* (or *nodes*) and $E = \{e_k\}_{1 \leq k \leq M}$ is a set of M *edges*. Each edge e_k connects two nodes, forming a pair (v_i, v_j) with $v_i, v_j \in V$ [147].

A graph is *weighted* if a function $w : E \rightarrow \mathbb{R}$ assigns a numerical weight w_{ij} to each edge (v_i, v_j) . A graph is *undirected* if every edge $e_k = (v_i, v_j)$ is bidirectional, meaning that there also exists an edge $e_l = (v_j, v_i)$ with the same weight, $w_{ij} = w_{ji}$. A directed weighted graph is illustrated in Figure 5.9.

A graph $G = (V, E)$ can be represented using a square *adjacency matrix* A of size $N \times N$. Each element of A encodes the presence and weight of an edge between two nodes:

- In an *unweighted* graph, the adjacency matrix is binary, $A \in \{0, 1\}^{N \times N}$, where $A_{ij} = 1$ if there is an edge between v_i and v_j , and 0 otherwise.
- In a *weighted* graph, $A \in \mathbb{R}^{N \times N}$, where A_{ij} holds the weight of the edge.
- In an *undirected* graph, the adjacency matrix is symmetric, i.e., $A_{ij} = A_{ji}$.

The adjacency matrix corresponding to the graph in Figure 5.9 is:

$$\begin{pmatrix} 0 & 0 & 0.4 & 0 & 0 & 0 \\ 1.3 & 0 & 5.7 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.2 \\ 0 & 0 & 3.3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4.2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (5.32)$$

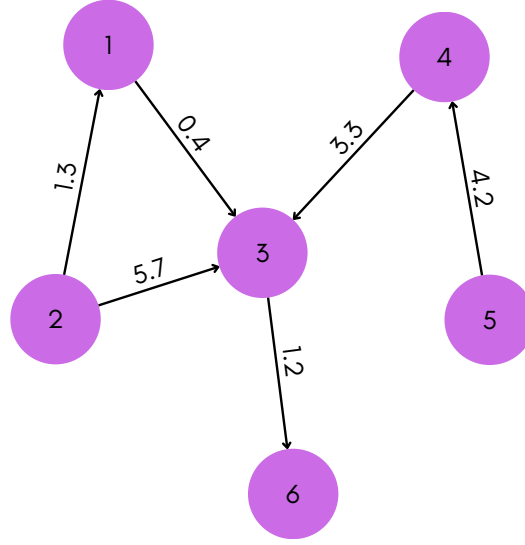


Figure 5.9: An oriented and weighted graph.

Another common representation is the *adjacency list*, where each node is associated with a list of its connected neighbours. The adjacency list corresponding to the graph in Figure 5.9 is:

$$((1, 3), (2, 1), (2, 3), (3, 6), (4, 3), (5, 4)) \quad (5.33)$$

If the graph is weighted, an additional list stores the corresponding edge weights. For the graph in Figure 5.9, this list is:

$$(0.4, 1.3, 5.7, 1.2, 3.3, 4.2). \quad (5.34)$$

Each node can also store additional *features*, which are attributes used by graph neural networks to learn meaningful representations of the data.

5.6.2.2 Graph embedding

Raw node, edge, and graph features are often high-dimensional and difficult to process directly. *Graph embedding* addresses this challenge by mapping each node v_i of a graph $G = (V, E)$ to a vector $u_i \in \mathbb{R}^d$ in a continuous latent space, while preserving important graph properties such as node neighbourhoods, structural roles, and feature similarities [147]. Nodes with similar connectivity patterns in the original graph should be mapped to close vectors in the embedding space.

For scalability, embeddings are typically low-dimensional ($d \ll N$). Beyond nodes, edges or entire graphs can also be embedded. Graph embedding can thus be seen as a dimensionality reduction technique adapted to non-Euclidean, discrete structures.

A good embedding should allow for information reconstruction, enabling the model to learn from the reconstruction error [148]. The general process of graph embedding consists of four key components:

- A *mapping function* that projects nodes, edges, or graphs into the embedding space.
- An *information extractor* that determines the graph properties I to be preserved.

- A *reconstructor* that attempts to recover the extracted graph information from the learned embeddings (the extracted information is denoted I').
- An *objective* that optimises the embeddings by minimising the difference between I and I' .

5.6.2.3 Networks

GNN tasks are typically categorised into node-focused and graph-focused tasks. In the former, a GNN learns informative features for individual nodes, while in the latter, learning node features is an intermediate step toward representing the entire graph [148].

For node-focused tasks, GNNs leverage both the node features and the graph structure. This process can be formalised as:

$$F^{(\text{of})} = h\left(A, F^{(\text{if})}\right) \quad (5.35)$$

where $A \in \mathbb{R}^{N \times N}$ is the adjacency matrix representing the graph structure, and $F^{(\text{if})} \in \mathbb{R}^{N \times d_{\text{if}}}$ and $F^{(\text{of})} \in \mathbb{R}^{N \times d_{\text{of}}}$ are the input and output feature matrices, with d_{if} and d_{of} their respective feature dimensions.

The transformation that takes node features and the graph structure as input and outputs an updated set of node features is known as graph filtering.

For node-focused tasks, stacking multiple graph filtering operations enables the extraction of increasingly abstract node representations. For graph-focused tasks, additional steps are required to aggregate node features into a single graph representation. Similar to pooling in CNNs, *graph pooling* operations summarise node features while preserving structural information. These operations generate a coarsened graph with fewer nodes, modifying both the structure and features. The graph pooling process can be expressed as:

$$A^{(\text{op})}, F^{(\text{op})} = \text{pool}\left(A^{(\text{ip})}, F^{(\text{ip})}\right) \quad (5.36)$$

where $A^{(\text{ip})} \in \mathbb{R}^{N_{\text{ip}} \times N_{\text{ip}}}$ and $F^{(\text{ip})} \in \mathbb{R}^{N_{\text{ip}} \times d_{\text{ip}}}$ are the adjacency and feature matrices before pooling, and $A^{(\text{op})} \in \mathbb{R}^{N_{\text{op}} \times N_{\text{op}}}$ and $F^{(\text{op})} \in \mathbb{R}^{N_{\text{op}} \times d_{\text{op}}}$ are the corresponding matrices after pooling, with $N_{\text{op}} < N_{\text{ip}}$.

5.6.3 Set Transformers

5.6.3.1 Transformers

The Transformer [149] is a neural network architecture designed for sequence transduction, i.e., the task of converting an input sequence into an output sequence, such as in translation. It has been developed to overcome the limitations faced by models such as Recurrent Neural Networks (RNNs).

RNNs have been widely used to process sequential data, such as text, images or time series, due to their ability to retain information during the process. Especially, they are well-suited for machine translation. However, they are confronted with several issues: information tends to degrade for long sequences, and their sequential structure prevents parallelisation, causing high computational costs.

The Transformer overcomes these issues by enabling each element in a sequence to attend to all others, determining the importance of each element relative to all others,

thus preserving long-range dependencies more effectively. Unlike RNNs, it does not rely on recurrence or convolutional structures, but is built upon an encoder-decoder architecture combined with attention mechanisms [150], defined mathematically below, allowing for efficient parallel processing and improved performance on a wide range of sequence-based tasks.

The encoder maps a variable-length input sequence (x_1, \dots, x_n) to a sequence of continuous representations of the same length (z_1, \dots, z_n) , which the decoder then processes to generate an output sequence (y_1, \dots, y_m) , with $m \neq n$ in general. It produces one element at a time, in an autoregressive manner: at each step, the model takes the previously generated symbols as input for the next prediction.

The overall structure of the Transformer is illustrated in Figure 5.10. The encoder, shown on the left, consists of N_x stacked identical layers, each comprising two sub-layers: a multi-head attention mechanism, and a fully connected feed-forward network. Each layer employs residual connections followed by normalisation, such that the output of each sub-layer is given by:

$$\text{LayerNorm}(x + \text{Sublayer}(x)) \quad (5.37)$$

where Sublayer represents the function implemented by the respective sub-layer.

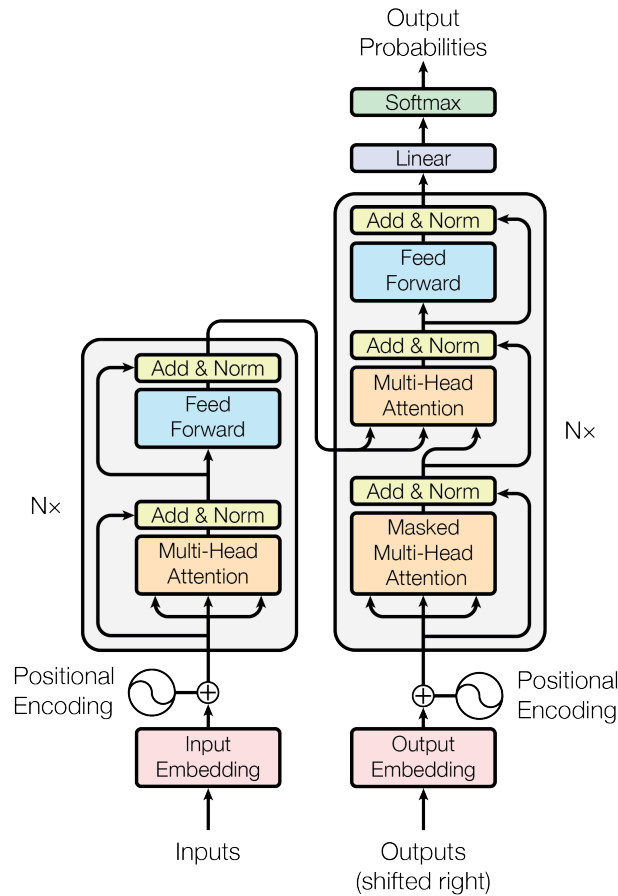


Figure 5.10: Transformer architecture, from Ref. [149].

The decoder follows a similar structure. It also uses residual connections and N_x stacked identical layers. Two main differences with the encoder have to be noted: it includes an additional sub-layer that applies multi-head attention to the encoder's outputs,

and a masking mechanism is incorporated to prevent each position from attending to future tokens, ensuring causal autoregressive generation.

As in other transduction models, raw sequences have to be transformed into a processable format, since raw sequences are not directly exploitable by neural networks. For instance, for transduction tasks, words are represented by indices in a vocabulary, which do not encode meaningful relationships between them. *Embedding* is applied to inputs and outputs, transforming them into continuous vectors in a lower-dimensional latent space, where proximity reflects semantic similarity.

Since the meaning of a sequence depends on the order of its elements, a positional encoding is added to the input embeddings of both the encoder and decoder to compensate for the absence of recurrence, which would otherwise convey positional information. The positional encoding has the same dimension d_{model} as the embeddings and is constructed using cos and sin functions with different frequencies.

The attention mechanism enables each element in a sequence to interact with all others. It avoids processing the data sequentially and allows for giving more importance to some parts of the sequence or to certain relationships between elements. For example, in the sentence “The cat lies on the sofa because it is tired”, the attention mechanism learns that “it” is more strongly related to “the cat” than to “the sofa”.

Attention weights are computed between each element and all others in the sequence. Formally, the mechanism maps a *query* and a set of *key-value* pairs to an output, where queries, keys, values, and outputs are vectors. The queries and keys have dimension d_k , while the values have dimension d_v . In practice, these vectors are stacked into matrices, and the attention mechanism is expressed as:

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^\top}{\sqrt{d_k}} \right) V \quad (5.38)$$

where Q , K , and V are the matrices of queries, keys, and values, respectively. The attention weight between input i and input j is encoded in the matrix product QK^\top . This product is scaled by $\sqrt{d_k}$ to prevent excessively large values, and a softmax function is applied to normalise the weights so that they sum to one across the sequence. The use of separate query and key matrices makes attention asymmetric: if input i strongly attends to input j , it does not necessarily mean that input j attends strongly to input i . This attention calculation, called “scaled dot-product attention”, is described in Figure 5.11a.

Instead of computing attention once using queries, keys, and values of dimension d_{model} , they are projected h times. Running several attention computations in parallel has several advantages: it allows the model to capture different types of relationships between elements, and reduces its sensitivity to weight initialisation.

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O \quad (5.39)$$

$$\text{where head}_i = \text{Attention} \left(QW_i^Q, KW_i^K, VW_i^V \right). \quad (5.40)$$

Here, the projections are learned parameter matrices: $W_i^Q \in \mathbb{R}^{d_{\text{model}} \times d_k}$ for the queries, $W_i^K \in \mathbb{R}^{d_{\text{model}} \times d_k}$ for the keys, $W_i^V \in \mathbb{R}^{d_{\text{model}} \times d_v}$ for the values, and $W^O \in \mathbb{R}^{hd_v \times d_{\text{model}}}$ for the final output projection. The model learns complementary representations by combining multiple attention heads, enhancing its ability to extract complex dependencies in the input sequence. The mechanism is shown in Figure 5.11b.

The Transformer makes use of multi-head attention in three distinct ways:

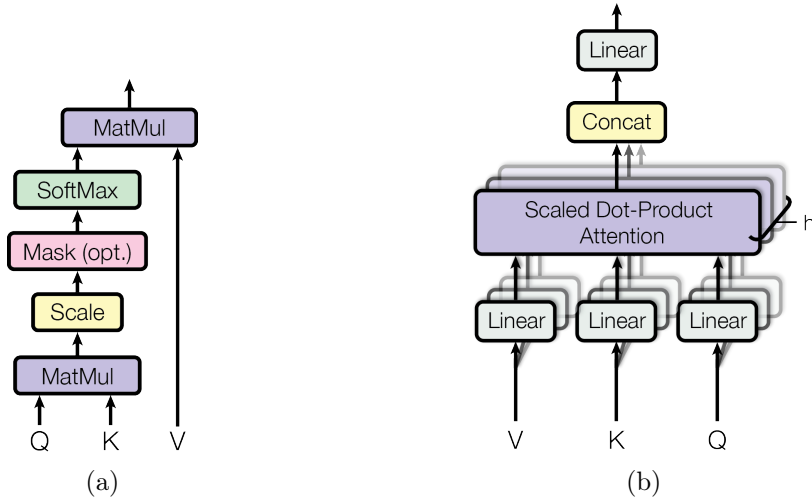


Figure 5.11: (a) Scaled dot-product attention, and (b) multi-head attention, consisting of several attention layers running in parallel. Images from Ref. [149].

- *Encoder-decoder attention*: In these layers, queries come from the previous decoder layer, while keys and values come from the encoder output. It enables each decoder output to attend to all positions in the input sequence.
- *Self-attention in the encoder*: The encoder includes self-attention layers. In these layers, all queries, keys, and values come from the output of the previous layer in the encoder. It allows each position in the encoder to attend to all positions in the previous layer.
- *Self-attention in the decoder*: It allows each position in the decoder to attend to all positions in the decoder up to itself, but not to positions beyond. This *masking* mechanism prevents elements from attending to future tokens, thus maintaining auto-regressivity.

Self-attention offers several benefits. Unlike RNNs, which tend to lose information when elements are far from each other, self-attention creates connections between elements regardless of the distance between them. Moreover, computational complexity is drastically reduced, and, as elements can be processed in parallel, computation time is significantly decreased.

5.6.3.2 Set Transformers

The Set Transformer [151] is a neural network architecture specifically designed to process set-structured data, which are permutation-invariant (i.e., swapping two elements in the set does not affect the model output) and variable in size (i.e., the model must handle sets of different cardinalities). Standard feed-forward neural networks do not satisfy these constraints, while RNNs violate permutation invariance.

Inspired by the Transformer, the Set Transformer employs self-attention to capture pairwise or higher-order interactions between elements in the set. It also relies on self-attention mechanisms for feature aggregation, making it particularly useful for tasks when multiple outputs depend on each other. A major improvement compared to the classical

Transformer is that it reduces the typical $\mathcal{O}(n^2)$ complexity of self-attention to $\mathcal{O}(nm)$, where m is a fixed hyperparameter, chosen lower than n .

The Set Transformer retains the multi-head attention mechanism of the standard Transformer but generalises the attention function to any function ω :

$$\text{Att}(Q, K, V; \omega) = \omega(QK^\top)V. \quad (5.41)$$

In practice, the scaled softmax used in the Transformer remains the most robust choice in many scenarios.

Like the Transformer, the Set Transformer relies on an encoder-decoder structure, where each layer in both the encoder and the decoder attends to its inputs. However, its pooling process differs from traditional models: rather than applying a fixed function, such as mean or max, it uses a learnable aggregation function that can adapt to the specific problem at hand.

Permutation-Equivariant (Induced) Set Attention Blocks The Set Transformer relies on a series of attention-based set operations, called Multihead Attention Block (MAB), Set Attention Block (SAB), and Induced Set Attention Block (ISAB). These are illustrated in Figure 5.12. These operations enable the model to capture pairwise and higher-order interactions between elements during the encoding process using self-attention mechanisms.

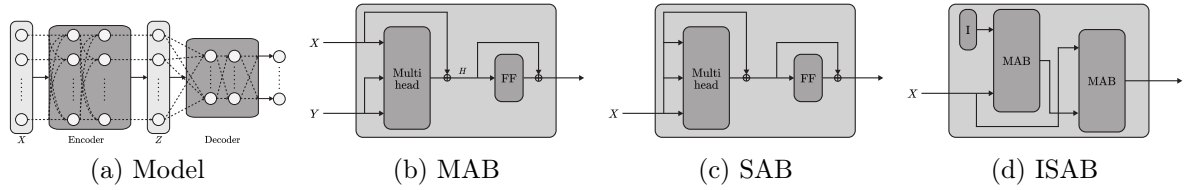


Figure 5.12: Diagrams of the attention-based set operations of the Set Transformer, from Ref. [151].

Given two sets of d -dimensional vectors represented as matrices $X, Y \in \mathbb{R}^{n \times d}$, the Multihead Attention Block (MAB) with parameters ω is defined as:

$$\text{MAB}(X, Y) = \text{LayerNorm}(H + \text{rFF}(H)), \quad (5.42)$$

where

$$H = \text{LayerNorm}(X + \text{Multihead}(X, Y, Y; \omega)). \quad (5.43)$$

Here, rFF represents a row-wise feedforward layer, meaning that it processes each instance independently and identically. The MAB is essentially an adaptation of the encoder block from the Transformer [149], but without positional encoding and dropout.

Using the MAB, the Set Attention Block (SAB) is defined as:

$$\text{SAB}(X) = \text{MAB}(X, X). \quad (5.44)$$

Pairwise interactions between elements in the input set X are thus captured in the SAB output. Consequently, several SABs can be stacked to encode higher-order relationships between elements. However, a major limitation of the SAB is its quadratic time complexity ($\mathcal{O}(n^2)$), making it computationally expensive for large sets.

To mitigate this issue, the Induced Set Attention Block (ISAB) is introduced. In addition to the input set $X \in \mathbb{R}^{n \times d}$, a set of m learnable inducing points, $I \in \mathbb{R}^{m \times d}$, is added as trainable parameters. The ISAB with m inducing points is then defined as:

$$\text{ISAB}_m(X) = \text{MAB}(X, H) \in \mathbb{R}^{n \times d}, \quad (5.45)$$

$$\text{where } H = \text{MAB}(I, X) \in \mathbb{R}^{m \times d}. \quad (5.46)$$

In this formulation, the inducing points I first attend to the input set X , producing a transformed representation H ; then, the input set X attends to H to produce the final output. This approach is analogous to low-rank projection or autoencoder models, where the input set X is first mapped onto a low-dimensional representation H , which is then used to reconstruct a more informative output.

However, unlike autoencoders, whose primary goal is reconstruction, ISAB aims to learn meaningful representations for the final task. The learned inducing points are expected to encode global features and structures, which help to understand the input set X , while reducing the computational complexity to $\mathcal{O}(nm)$.

In Equations 5.45 and 5.46, attention is computed between sets of sizes n and m , resulting in a computational complexity of $\mathcal{O}(nm)$, where m is typically small, representing a significant improvement over the $\mathcal{O}(n^2)$ complexity of the SAB.

Pooling by Multihead Attention Instead of averaging or summing the feature vectors, the Set Transformer aggregates features by applying multihead attention to a learnable set of k seed vectors $S \in \mathbb{R}^{k \times d}$. Given a set of feature representations $Z \in \mathbb{R}^{n \times d}$ produced by an encoder, Pooling by Multihead Attention (PMA) with k seed vectors is defined as

$$\text{PMA}_k(Z) = \text{MAB}(S, \text{rFF}(Z)). \quad (5.47)$$

To further capture dependencies between the k pooled representations, a SAB is applied to the output of PMA_k :

$$H = \text{SAB}(\text{PMA}_k(Z)). \quad (5.48)$$

Feature aggregation via attention is advantageous, as it allows the model to assign different levels of importance to individual instances, rather than treating them uniformly as in standard averaging- or summation-based pooling.

Overall architecture The Set Transformer consists of an encoder and a decoder, both built upon the attention-based set operations described above.

The encoder, $\text{Encoder} : X \mapsto Z \in \mathbb{R}^{n \times d}$, is constructed as a stack of SABs or ISABs, for example:

$$\text{Encoder}(X) = \text{SAB}(\text{SAB}(X)), \quad (5.49)$$

$$\text{Encoder}(X) = \text{ISAB}_m(\text{ISAB}_m(X)). \quad (5.50)$$

The computational complexity for l stacked SABs is $\mathcal{O}(ln^2)$, whereas for ISABs, it is reduced to $\mathcal{O}(lnm)$, offering a significant efficiency gain when using ISABs, especially for large sets.

Once the encoder transforms the input set $X \in \mathbb{R}^{n \times d_x}$ into feature representations $Z \in \mathbb{R}^{n \times d}$, the decoder aggregates them into a fixed-size output representation. This output can consist of a single vector or a set of vectors, which are subsequently processed by a feed-forward network to produce the final predictions. If PMA is used with $k > 1$ seed vectors, it should be followed by SABs to capture correlations between the k outputs:

$$\text{Decoder}(Z) = \text{rFF}(\text{SAB}(\text{PMA}_k(Z))) \in \mathbb{R}^{k \times d}, \quad (5.51)$$

where $\text{PMA}_k(Z) \in \mathbb{R}^{k \times d}$.

Since both SAB and ISAB are *permutation equivariant*, the encoder mapping $X \mapsto Z$ preserves permutation symmetry. Moreover, the PMA used in the decoder is a permutation-invariant transformation, ensuring that the Set Transformer as a whole remains invariant under input permutation.

5.7 Machine Learning in this work

Machine learning techniques are employed in the search for a new resonance decaying into two collimated photons. More specifically, in the next chapter, a fully connected neural network is used to develop an identification criterion for diphotons based on the shower shape variables of the two photons. In addition, a CNN, a GNN, and a Set Transformer are explored for diphoton isolation, using the energy deposits in the vicinity of the photon pair.

Chapter 6

Diphoton identification and isolation

Contents

6.1	Closely collimated photon pairs	128
6.2	Event samples	129
6.2.1	Simulated prompt diphoton sample	130
6.2.2	Simulated dijet sample	132
6.2.3	Fake-enriched data sample	133
6.3	New diphoton identification criterion	133
6.3.1	Input preparation	134
6.3.2	Architecture of the neural network	138
6.3.3	Results and performance	139
6.3.4	Conclusion	140
6.4	New diphoton isolation criterion	142
6.4.1	Diphoton isolation with a CNN	144
6.4.2	Diphoton isolation with a GNN	150
6.4.3	Diphoton isolation with a Set Transformer	156
6.4.4	Conclusion	158

As discussed in Section 1.5, the search for diphoton resonances is well-motivated in the context of axion-like particles, which can be produced via gluon fusion and decay into two photons, making them key targets for LHC experiments. Previous searches at the LHC [64, 152–156] have explored diphoton resonances down to masses of approximately 10 GeV. However, technical constraints, due to trigger thresholds and the limited resolution of the detectors, impose significant limitations, making it challenging to reconstruct photon pairs below this threshold using standard ATLAS tools.

The work presented here proposes a new approach. Rather than considering the two photons independently, they are treated as a single system, a *diphoton*, to which new definitions of identification and isolation are applied. This chapter begins by explaining the challenges encountered when studying low-mass diphoton resonances and then presents my work on the development of diphoton-oriented identification and isolation definitions.

This work has benefited from the contributions of several interns I co-supervised during my PhD: Anatole Saurier, Lucas Le Gall and Joël Abecassis, for their involvement in the identification task; and Emil Bark Kvamme and Amshuman Rousselot for their contributions to the isolation task.

6.1 Closely collimated photon pairs

The diphoton invariant mass is related to the transverse energies and angular separation of the photons by

$$m_{\gamma\gamma} = \sqrt{2p_{T,1}^{\gamma}p_{T,2}^{\gamma} [\cosh(\eta_2 - \eta_1) - \cos(\phi_2 - \phi_1)]}, \quad (6.1)$$

where the indices refer to the two photons. For closely collimated photon pairs, this relation becomes:

$$m_{\gamma\gamma} \approx \sqrt{p_{T,1}^{\gamma}p_{T,2}^{\gamma}}\Delta R_{\gamma\gamma}. \quad (6.2)$$

This expression illustrates that the reconstruction of low-mass diphoton resonances requires selecting either photons with low transverse energy or closely collimated photon pairs. However, both scenarios face experimental challenges. The trigger system imposes stringent constraints on photon transverse momenta, with thresholds of $p_T > 7$ GeV at Level-1 (L1) and 10 GeV at the High-Level Trigger (HLT), in addition to high prescales at such low energies. Moreover, the photon identification and isolation criteria introduce further limitations.

When two photons are in close proximity, their electromagnetic showers overlap, altering the shower shape variables used for photon identification. As a result, the efficiency of standard photon selection criteria is reduced. Furthermore, the isolation requirement, which involves summing the energy deposits within a predefined cone around the photon candidate, is affected by the presence of a nearby photon. The overlap of the isolation cones leads to an artificial increase in the energy deposit, which in turn decreases the likelihood of passing the isolation criteria. These effects significantly impact the identification efficiency of low-mass diphoton signals, becoming essentially zero below 10 GeV, as shown in Figure 6.1, and necessitate dedicated strategies to overcome these limitations.

The diphoton-trigger strategy, which has been employed in previous LHC studies [156, 157], has reached its intrinsic limitations, making further improvements within this framework challenging. The photon candidates are in the same trigger Region of Interest (RoI) at such low angles.

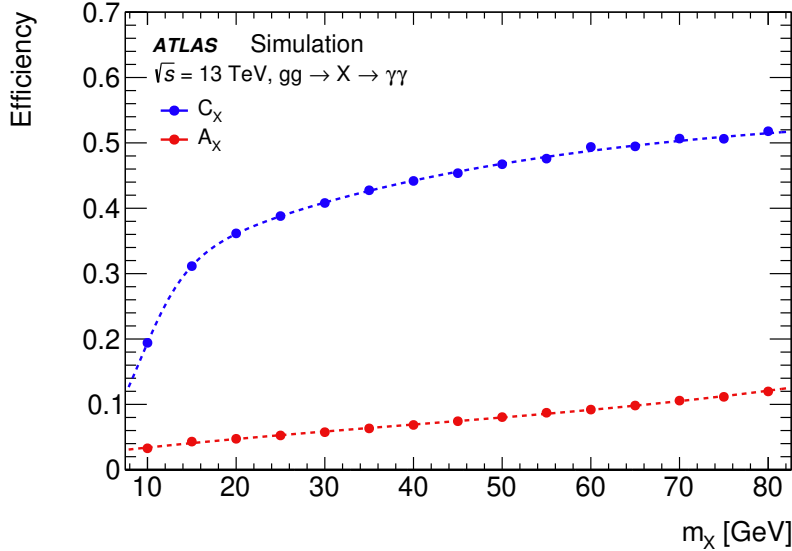


Figure 6.1: Selection efficiency in a search for boosted diphoton resonances in the 10 to 70 GeV mass range, using the full Run 2 dataset, with the ATLAS detector [156]. Events with at least two photon candidates, with transverse energies above 22 GeV, $|\eta| < 2.37$, and satisfying the Tight identification criteria were selected. The correction factor, C_X , is defined as the ratio between the number of reconstructed signal events passing the analysis cuts and the number of signal events at the particle level generated within the fiducial volume, and the acceptance correction factor, A_X , is defined as the fraction of diphoton resonances satisfying the fiducial acceptance at the particle level. The diphoton selection efficiency is the product of these two quantities. The fiducial volume is defined by requiring two photons at particle level with $E_T > 22$ GeV, $|\eta| < 2.37$ and $p_T^{\gamma\gamma} > 50$ GeV.

This work proposes an alternative approach that leverages a loose single-photon trigger to select events containing two loosely identified photon candidates. Instead of applying the standard Tight photon identification and isolation criteria, new requirements are defined using machine learning techniques and applied to reconstructed, closely collimated photon pairs. These new criteria, analogous to the standard ones, make use of shower shape variables and the energy deposits surrounding the diphoton candidates.

The following sections detail the development and results of this performance study, which aims to identify pairs of closely collimated photons. The samples used as control for true and fake diphotons are described, along with the event selection procedure. In contrast to the photon isolation correction study, which was based on Run 2 data, the present development relies on the new Run 3 data. An overview is then given of the neural networks developed for diphoton identification and isolation, along with their associated efficiencies. All models have been designed with PyTorch [158].

6.2 Event samples

This performance study relies on simulated samples for both true and fake diphotons and a fake-enriched control region on data. In the search for a diphoton resonance, the background is composed of two main components. The first is the *reducible* background, arising from events in which at least one of the photon candidates is not a true prompt photon. These fake diphotons arise from three main sources: jets leading to detector

signatures passing the diphoton selection criteria, and radiative decays of the Z and W bosons ($Z \rightarrow ee\gamma$ and $W^\pm \rightarrow e^\pm \nu_e \gamma$), in which an electron is misidentified as a photon, and an electron-photon pair passes the diphoton selection. As shown in more detail in the next chapter, the contributions from Z and W decays are negligible compared to the jet background. The second main background component is the *irreducible* background. It consists of prompt photon pairs from QCD originating from the hard scatter (see Figure 1.8) but unrelated to a new diphoton resonance, such as an ALP decay. Unlike the reducible background, the irreducible component cannot be suppressed without also suppressing the signal from a new resonance. The neural networks developed in this work focus on distinguishing prompt QCD diphotons from jet-induced diphotons.

The primary goal of this study is to identify pairs of closely collimated photons, with angular separation $\Delta R_{\gamma\gamma} < 0.2$, a regime in which standard photon identification and isolation tools are no longer effective. However, this work focuses exclusively on *resolved* photon pairs, cases in which two separate topo-clusters are reconstructed. *Merged* topologies, where a single topo-cluster is reconstructed for the photon pair, require a different treatment [159] and are not considered in this analysis. Cell-level information would be useful for these topologies, but it is not fully available. For this reason, only photon pairs with $\Delta R_{\gamma\gamma} > 0.1$ are retained.

Moreover, the analysis targets diphoton resonances with invariant masses below 10 GeV, making events with $m_{\gamma\gamma} \sim 10$ GeV particularly relevant.

6.2.1 Simulated prompt diphoton sample

Prompt diphoton events are generated using the PARTICLEGUN (PG) framework, designed to produce pairs of photons with specific kinematic properties. The generation is configured to populate the desired phase space in angular separation and invariant mass, ensuring good coverage of the target signal region. It should be noted that this setup does not correspond to a physical event generator. However, for this initial performance study, it provides a sufficiently accurate and controlled environment to evaluate the algorithm's response to closely collimated prompt photon pairs.

Two photons are generated, with transverse energy p_T^γ above 5 GeV, and the sum of their transverse energies $p_T^{\gamma 1} + p_T^{\gamma 2}$ above 130 GeV, to match the unrescaled trigger HLT_g140_loose threshold. The first photon is uniformly generated in the $\eta - \phi$ plane, and the second one in its vicinity, with an angular separation $\Delta R_{\gamma\gamma}$ between 0.05 and 0.25, as flat as possible within this range, using a wide Gaussian sampling.

A total of 1 000 000 events have been generated. All are subsequently processed through the full GEANT4-based ATLAS detector simulation. The effect of pile-up is included using simulated events that reproduce the nominal pile-up distribution of 2023.

Figure 6.2 shows the p_T^γ and $\Delta R_{\gamma\gamma}$ distributions at the generator level, before applying the selections at the reconstruction level, presented in the following.

To obtain a clean and representative sample of diphoton candidates for the analysis, the following selection criteria are applied at the reconstruction level:

- **Photon transverse momentum:** A minimum transverse momentum of $p_T^\gamma > 5$ GeV is required on the softest photon in the pair.
- **Trigger selection:** Events for which HLT_g140_loose_L1EM22VHI or HLT_g140_loose_L1eEM26M is fired have been selected. The L1 trigger L1EM22VHI is used for simulation of the years 2022 and 2023, while L1eEM26M is used for 2024. HLT_g140

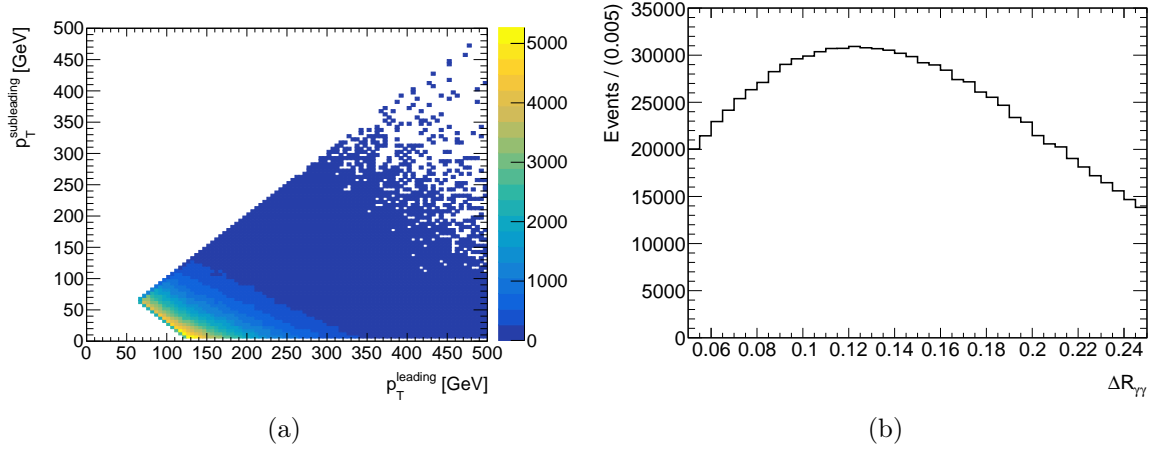


Figure 6.2: Distributions at generator level from the prompt diphoton sample, generated with ParticleGun: (a) leading vs subleading photon p_T and (b) diphoton angular separation $\Delta R_{\gamma\gamma}$.

_loose_L1EM22VHI and HLT_g140_loose_L1eEM26M are unprescaled. In the following, the L1 trigger chain is omitted; HLT_g140_loose refers to both HLT_g140_loose_L1EM22VHI and HLT_g140_loose_L1eEM26M.

- **Photon categories:** Only unconverted-unconverted photon pairs in the barrel region ($|\eta^\gamma| < 1.37$) are considered, to ensure optimal reconstruction quality and avoid complications related to conversions or end-cap geometry.
- **Identification criteria:** Both photons must satisfy the Loose identification criteria.
- **Angular separation:** The angular distance between the two photons is required to satisfy $\left(\frac{\Delta\eta_{\gamma\gamma}}{0.1}\right)^2 + \left(\frac{\Delta\phi_{\gamma\gamma}}{0.15}\right)^2 > 1$. This elliptical selection slightly restricts the region of interest ($0.1 < \Delta R_{\gamma\gamma} < 0.2$), but was applied to exclude two clusters of events observed in the data at $\Delta\phi_{\gamma\gamma} \sim \pm 0.1$, as shown in Figure 6.3. These events are not understood and have been removed from the selection¹.
- **Invariant mass:** Events are selected in the diphoton mass range $5 < m_{\gamma\gamma} < 35$ GeV, targeting the mass region relevant to the analysis.
- **Truth matching:** Truth matching is used to identify true diphoton events. Reconstructed photons must be matched to true prompt photons with $|\Delta\eta^{\text{reco-true}}| < 0.025$ and $|\Delta\phi^{\text{reco-true}}| < 0.025$, corresponding to the size of a second-layer cell of the EM calorimeter.

These selection criteria are summarised in Table 6.1. Out of the one million generated prompt diphoton events, 36 936 are selected. This is due to the fiducial volume and the chosen selection efficiency, particularly when selecting unconverted–unconverted events.

¹The most likely explanation is an artefact associated with the subleading photon. Tests have been performed to verify whether it was caused by cross-talk, following those in Ref. [160], but they were not conclusive.

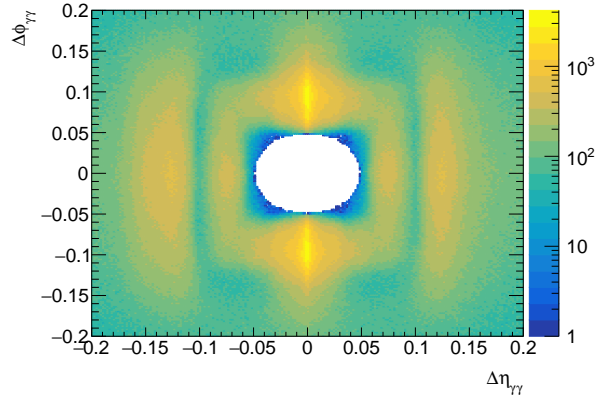


Figure 6.3: $\Delta\eta_{\gamma\gamma}$ vs $\Delta\phi_{\gamma\gamma}$ for 2022 to 2024 data, showing two clusters of events at $\Delta\phi \sim \pm 0.1$.

<i>Parameter</i>	<i>Selection criterion</i>
Transverse energy	$p_T^\gamma > 5 \text{ GeV}$
Trigger	HLT_g140_loose
Detector region	Barrel ($ \eta^\gamma < 1.37$)
Category	Both unconverted
Identification	Both Loose
Angular separation	$\left(\frac{\Delta\eta_{\gamma\gamma}}{0.1}\right)^2 + \left(\frac{\Delta\phi_{\gamma\gamma}}{0.15}\right)^2 > 1$
Invariant mass ($m_{\gamma\gamma}$)	[5 GeV, 35 GeV]
Truth matching	$ \Delta\eta^{\text{reco-true}} < 0.025$ $ \Delta\phi^{\text{reco-true}} < 0.025$

Table 6.1: Selection criteria applied to diphoton candidates in the simulated prompt diphoton sample.

6.2.2 Simulated dijet sample

Simulated dijet samples are employed in this analysis. The jet background events are generated using PYTHIA 8, simulating QCD multijet production. The samples are sliced in jet transverse momentum (p_T^{jet}) and filtered with parton-level momentum (p_T^{hat}) thresholds to ensure efficient coverage of the relevant phase space. These events are then processed through the same full GEANT4 detector simulation chain as the signal sample. The effect of pile-up is included using simulated events that reproduce the nominal pile-up distribution from 2022 to 2024. This sample is referred to as the “JZ” sample in the following.

A total of 2 098 750 000 events have been generated. The same selection criteria as for the prompt diphoton sample are applied, except for truth matching, which is replaced by jet matching: at least one of the reconstructed photons must be matched to a jet within $\Delta R_{\gamma j} < 0.15$, ensuring a reliable identification of fake diphoton candidates. The selection criteria are summarised in Table 6.2. 7771 events pass these selections. This is also due to the chosen fiducial volume and the applied selections, particularly the requirement of unconverted–unconverted candidates.

<i>Parameter</i>	<i>Selection criterion</i>
Transverse energy	$p_T^\gamma > 5 \text{ GeV}$
Trigger	<code>HLT_g140_loose</code>
Detector region	Barrel ($ \eta^\gamma < 1.37$)
Category	Both unconverted
Identification	Both Loose
Angular separation	$\left(\frac{\Delta\eta_{\gamma\gamma}}{0.1}\right)^2 + \left(\frac{\Delta\phi_{\gamma\gamma}}{0.15}\right)^2 > 1$
Invariant mass ($m_{\gamma\gamma}$)	[5 GeV, 35 GeV]
Jet matching	$\Delta R_{\gamma j} < 0.15$

Table 6.2: Selection criteria applied to diphoton candidates in the simulated dijet sample.

6.2.3 Fake-enriched data sample

Due to the limited statistics of fake diphotons in the simulated dijet dataset, a fake-enriched control region has been constructed using data. All available Run 3 data, from 2022 to 2024, have been used. The same selection criteria as for the simulated dijet sample are applied, except for the jet matching, which is replaced by a cut on the track isolation variable p_T^{cone30} : the minimum value between the two photon candidates, $\min(p_{T,\text{leading}}^{\text{cone30}}, p_{T,\text{subleading}}^{\text{cone30}})$, must be greater than 5 GeV. This cut strongly suppresses prompt diphoton events. Distributions of track isolation for the simulated prompt and fake diphotons can be seen in the next chapter, in Figure 7.2 and Figure 7.3.

The selection criteria are summarised in Table 6.3. A total of 106 955 events pass these selections.

<i>Parameter</i>	<i>Selection criterion</i>
Transverse energy	$p_T^\gamma > 5 \text{ GeV}$
Trigger	<code>HLT_g140_loose</code>
Detector region	Barrel ($ \eta^\gamma < 1.37$)
Category	Both unconverted
Identification	Both Loose
Angular separation	$\left(\frac{\Delta\eta_{\gamma\gamma}}{0.1}\right)^2 + \left(\frac{\Delta\phi_{\gamma\gamma}}{0.15}\right)^2 > 1$
Invariant mass ($m_{\gamma\gamma}$)	[5 GeV, 35 GeV]
Track isolation	$\min(p_{T,\text{leading}}^{\text{cone30}}, p_{T,\text{subleading}}^{\text{cone30}}) > 5 \text{ GeV}$

Table 6.3: Selection criteria applied to diphoton candidates in the data sample.

6.3 New diphoton identification criterion

Standard single-photon identification relies on rectangular cuts applied to shower shape variables (see Section 3.3). However, these variables evolve with the angular separation between the two photon candidates, changing the efficiency at low $\Delta R_{\gamma\gamma}$, as shown in Figures 6.4 and 6.5, for candidates passing the selections. Moreover, the correlations among the shower shape variables vary with $\Delta R_{\gamma\gamma}$ in different ways for true and fake diphotons, as illustrated in Figure 6.6. This behaviour underscores the advantage of using a neural network to capture the interplay between angular separation and shower shape

variables, allowing for a more robust discrimination between prompt and fake photon pairs. This study uses simulated samples for both prompt diphoton and dijet events. The fake-enriched sample was not used due to discrepancies in the shower shape variables between data and simulation, which would require the application of correction factors.

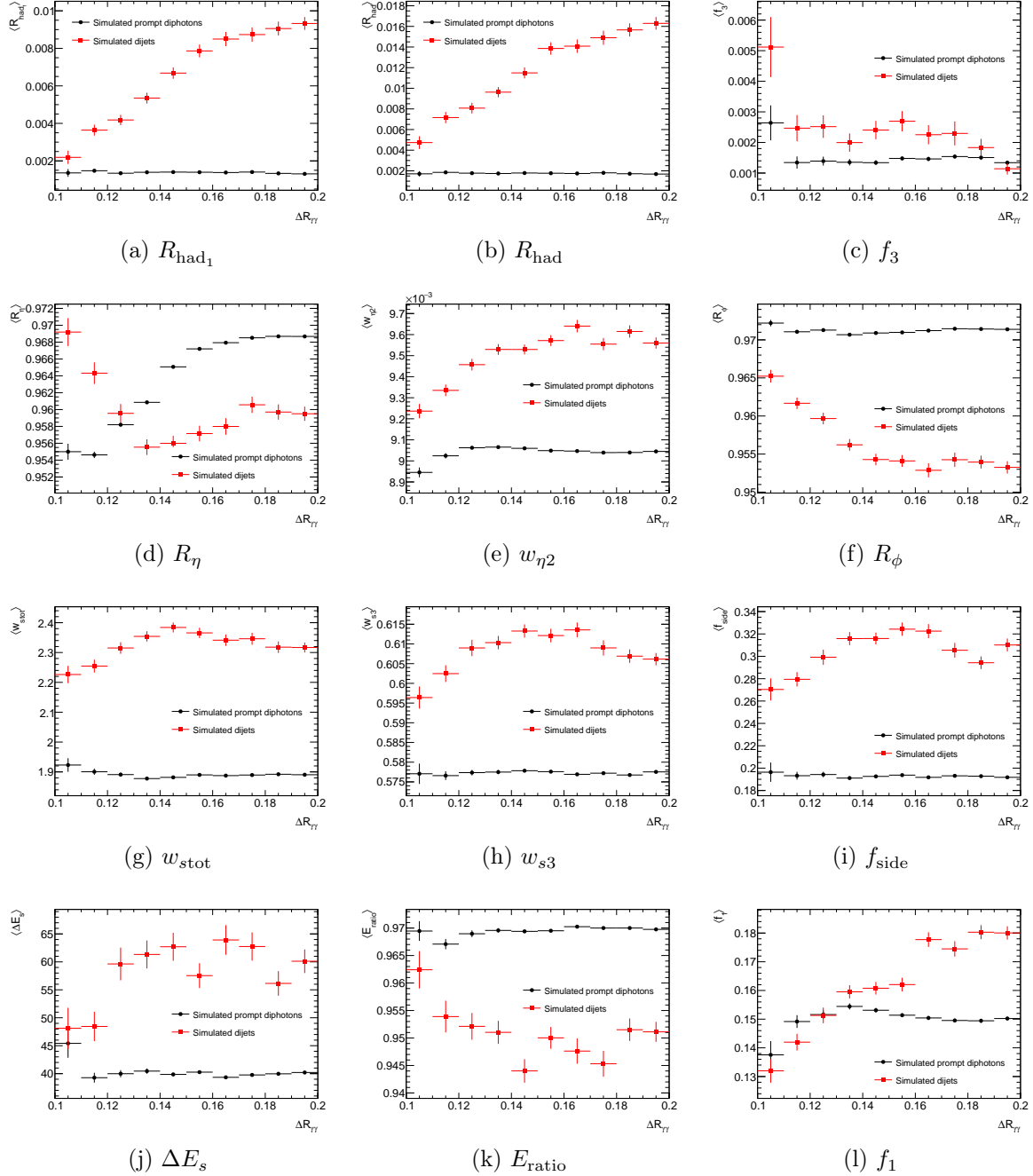


Figure 6.4: Mean shower shape variables as a function of the diphoton angular separation, shown for the leading photon, separately for prompt and fake diphoton events.

6.3.1 Input preparation

In addition to the selection criteria in Table 6.1 and Table 6.2, ranges have been defined for each shower shape variable to eliminate a few outliers while retaining almost all candidates.

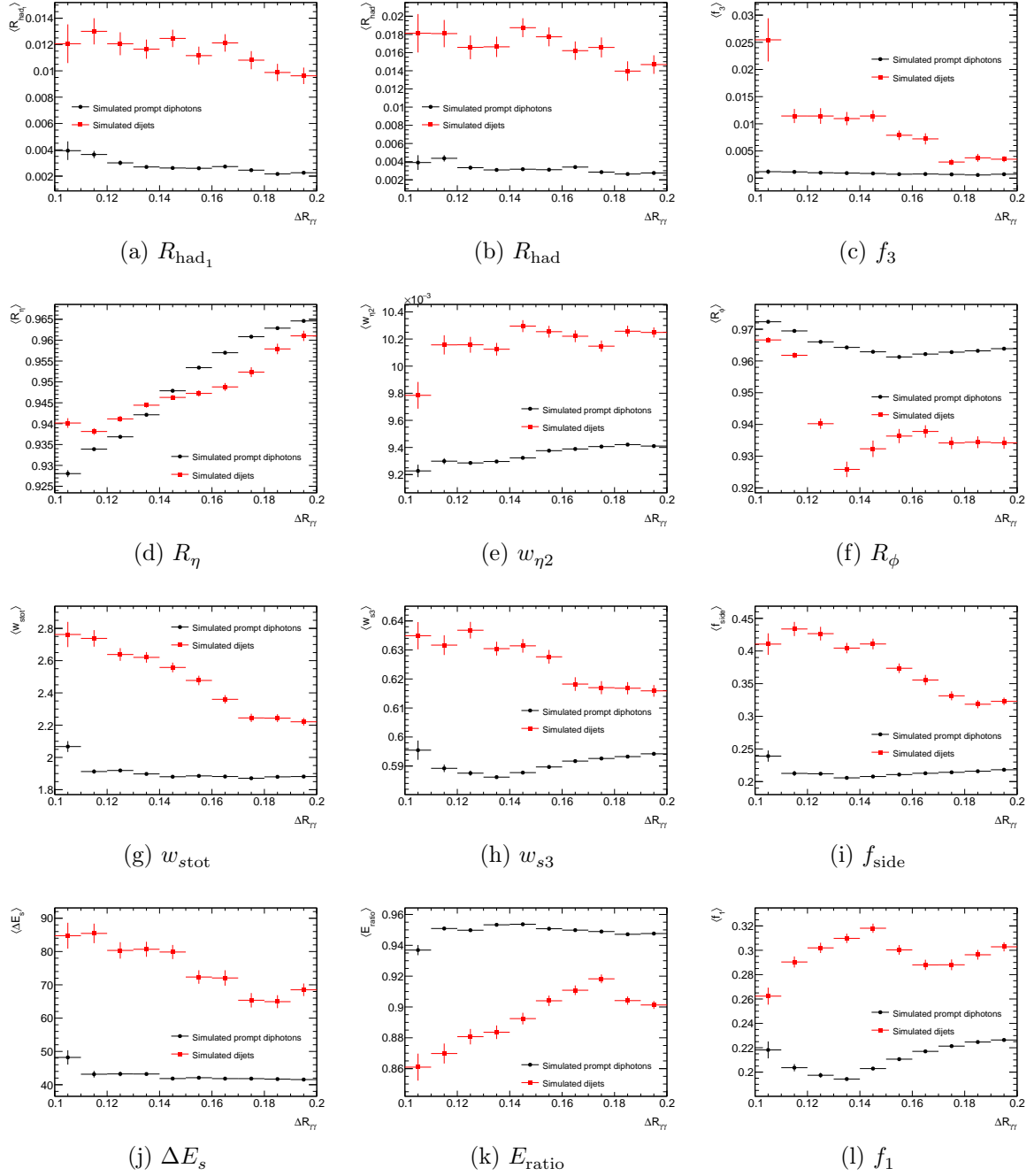


Figure 6.5: Mean shower shape variables as a function of the diphoton angular separation, shown for the subleading photon, separately for prompt and fake diphoton events.

After these additional selections, 35 639 prompt and 7261 fake diphotons are used for training, validation, and testing. Inputs are reweighted to account for these imbalanced datasets.

The inputs to the neural networks include the angular separation between the two photons ($\Delta R_{\gamma\gamma}$), the shower shape variables for each photon ($R_{\text{had}1}$, R_{had} , f_3 , R_η , $w_{\eta2}$, R_ϕ , w_{stot} , w_{s3} , f_{side} , ΔE_s , E_{ratio} , f_1), and the energy deposited in the PS and in each EM calorimeter layer (E_0 , E_1 , E_2 , E_3), for each photon.

In addition, the true and fake diphoton distributions in $\Delta R_{\gamma\gamma}$ were reweighted to achieve uniformity. The goal is to prevent the neural network from exploiting the $\Delta R_{\gamma\gamma}$

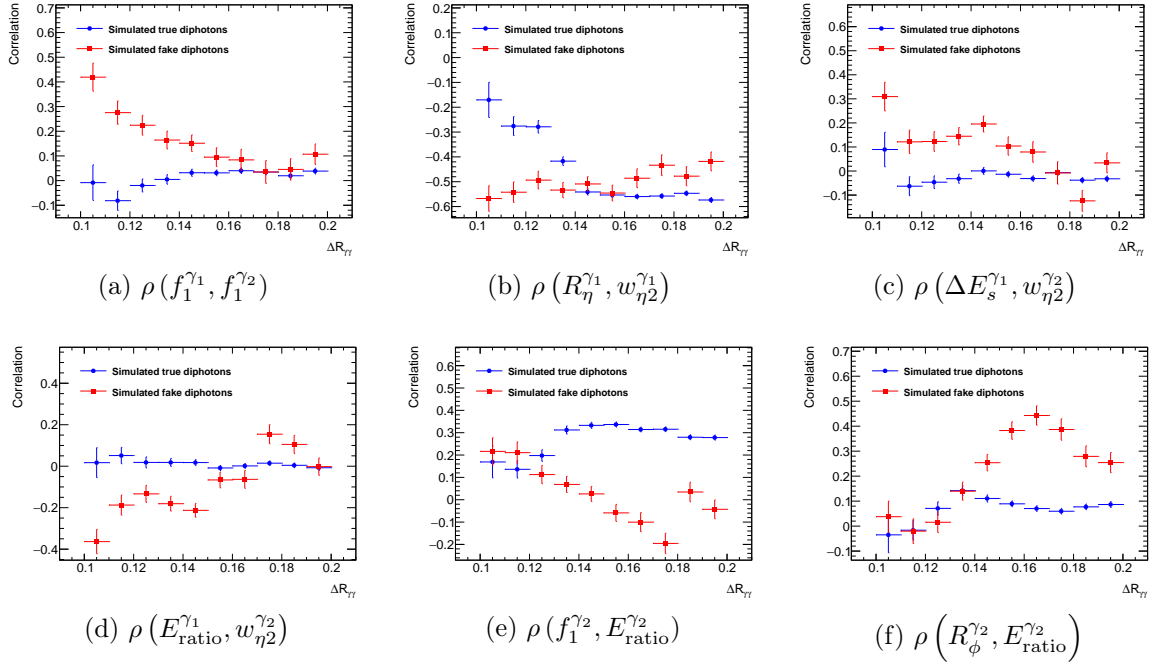


Figure 6.6: Correlations among pairs of shower shape variables as a function of the diphoton angular separation, shown separately for prompt and fake diphotons.

distribution itself, and instead to learn the correlations between angular separation and the shower shape variables. The inputs were divided into bins of $\Delta R_{\gamma\gamma}$, and weights were applied to each bin to flatten the $\Delta R_{\gamma\gamma}$ distributions. The reweighted $\Delta R_{\gamma\gamma}$ distributions are shown in Figure 6.7.

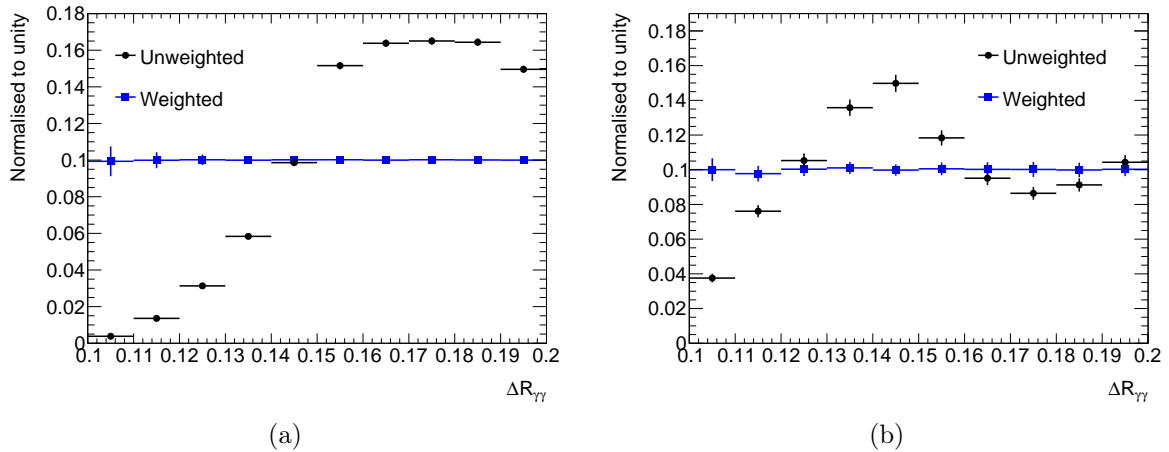
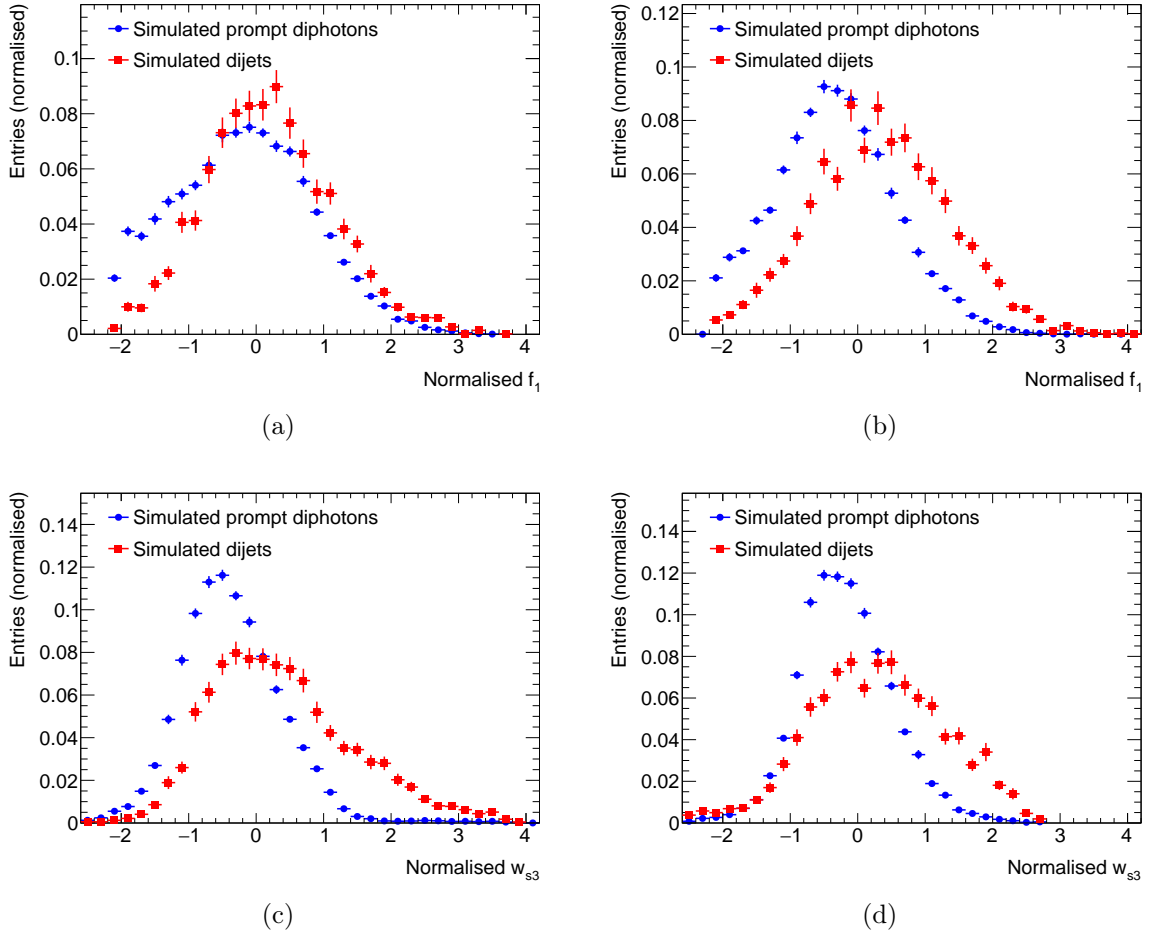
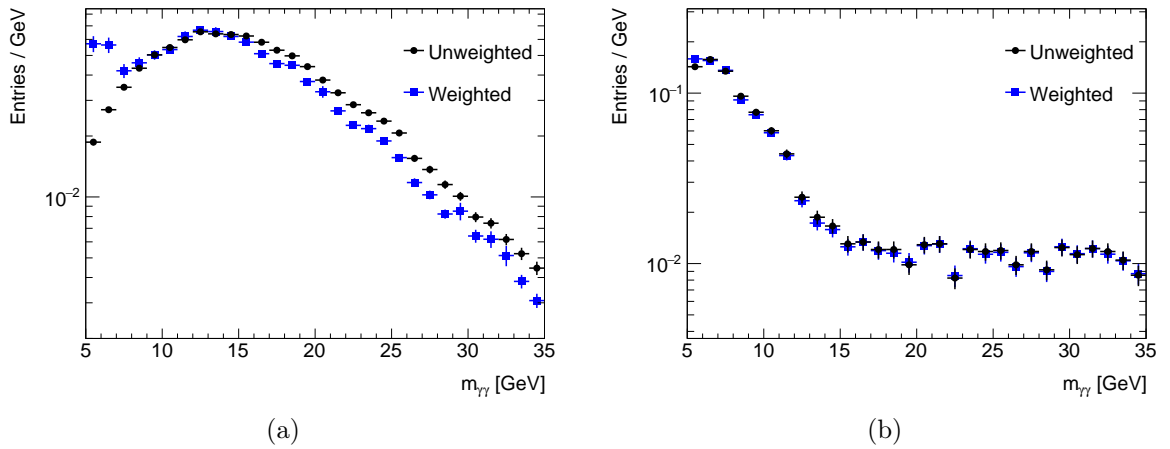


Figure 6.7: $\Delta R_{\gamma\gamma}$ distributions, for (a) simulated prompt diphotons and (b) simulated dijets, before and after reweighting.

Inputs used for training and validation are normalised to improve training stability by subtracting their mean and dividing by their standard deviation. The test inputs, used to evaluate the model's performance independently of training, are normalised using the mean and standard deviation computed from the training and validation datasets. The final reweighted and normalised distributions are shown for several shower shapes in

Figure 6.8: The mass distributions are illustrated in Figure 6.9.

Figure 6.8: Reweighted and normalised distributions of f_1 and w_{s3} for the leading (left) and the subleading (right) photons.Figure 6.9: Invariant mass distributions, for (a) simulated prompt diphotons and (b) simulated dijets, before and after reweighting. The first two bins in the prompt diphoton distribution are due to the large weights applied to low- $\Delta R_{\gamma\gamma}$ events (see Figure 6.7).

About 15 % of the inputs were set aside as a test dataset. The remaining data were used for training and validation. As discussed in the previous chapter, the training set is used to optimise the model parameters, the validation set monitors the model’s performance during training to prevent overfitting, and the test set provides an unbiased evaluation of the final model’s performance.

The model was trained independently 50 times. For each training run, the network weights were reinitialised to random values, and the inputs not reserved for testing were shuffled and split into training and validation datasets—70 % and 15 % of all inputs, respectively. Training was stopped if the validation loss did not decrease after 60 consecutive epochs. The final model selected is the one with the lowest validation loss, and its performance is evaluated on the test dataset.

6.3.2 Architecture of the neural network

A fully connected neural network has been designed to discriminate between real and fake diphotons, using the diphoton candidates’ angular separation and shower shape variables as input features. The architecture consists of an input layer, 20 hidden layers, each containing 40 nodes, and a single-node output layer. The rectified linear unit (ReLU) activation function is used in the hidden layers, while a sigmoid activation function is employed in the output layer to produce a probability score of being a prompt diphoton.

The binary cross-entropy (BCE) loss function is used for training, with additional weighting applied to account for two effects: the reweighting factors introduced to flatten the $\Delta R_{\gamma\gamma}$ distributions, and the imbalance between the true and fake diphoton samples:

$$\text{BCE}(\hat{y}, y) = -w[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})], \quad (6.3)$$

where \hat{y} is the output of the network for a certain input, y is its label, and w is its corresponding weight.

The model is trained with a batch size of 64 and a learning rate of 0.001. To mitigate overfitting, L2 regularisation is applied with a weight decay of 10^{-4} . Optimisation is performed using the Adam algorithm. Table 6.4 summarises the architecture and hyperparameters of the model, which were chosen for their good performance.

<i>Parameter</i>	<i>Value</i>
Number of hidden layers	20
Number of nodes in the hidden layers	40
Number of nodes in the output layer	1
Activation function in the hidden layers	ReLU
Activation function in the output layer	Sigmoid
Loss function	Binary cross-entropy
Batch size	64
Learning rate	0.001
Weight decay (L2 regularisation)	0.0001
Optimiser	Adam

Table 6.4: Diphoton identification model characteristics.

6.3.3 Results and performance

The convergence is very fast, as can be seen from the training and validation losses shown in Figure 6.10.

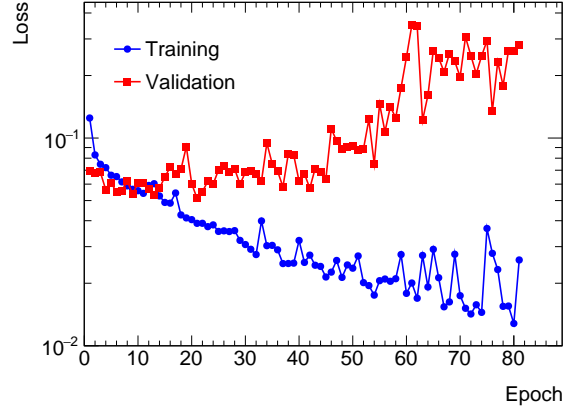


Figure 6.10: Training and validation losses vs. epochs for the best training run.

The distributions of predictions for prompt and fake diphotons are shown in Figure 6.11.

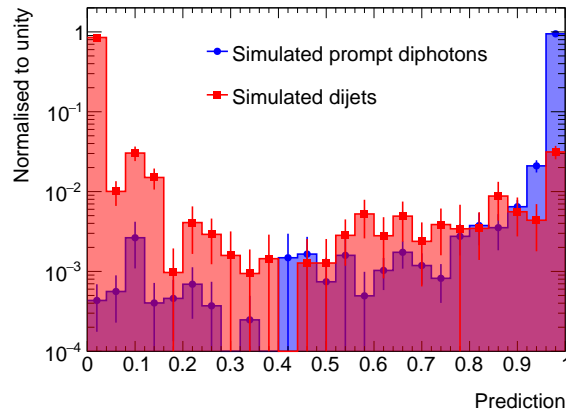
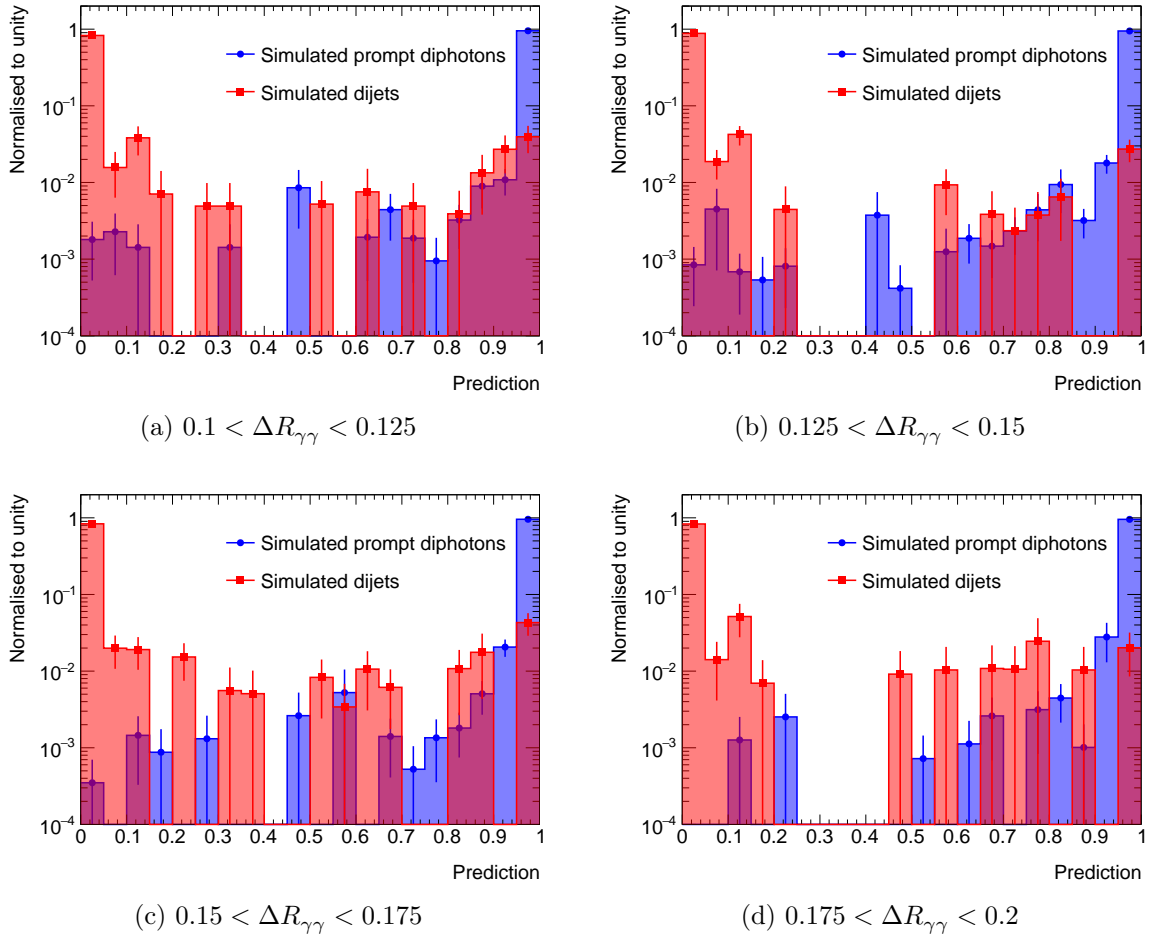


Figure 6.11: Distribution of predictions for prompt and fake diphotons.

Figure 6.12 and Figure 6.13 show the distributions of predictions per bin of $\Delta R_{\gamma\gamma}$ and $m_{\gamma\gamma}$, respectively. The ROC curve, showing the selection efficiency of prompt diphotons against fake ones, is presented in Figure 6.14.

If the statistics were sufficient, it would be possible to calculate the threshold that maximises the prompt diphoton efficiency and the dijet rejection rate. However, the number of dijet events is too low, making the distribution of predictions too fluctuating. The threshold to classify an event as a prompt diphoton has therefore been fixed at 0.5.

The efficiency per bin of $\Delta R_{\gamma\gamma}$ and per bin of $m_{\gamma\gamma}$ is shown in Figure 6.15. The prompt diphoton efficiency exceeds 95 % across all mass and angular separation bins. The rejection rate is above 90 % for all $\Delta R_{\gamma\gamma}$ bins. It ranges between 75 % and 95 % for masses between 5 GeV and 35 GeV.

Figure 6.12: Distributions of predictions for prompt and fake diphotons per $\Delta R_{\gamma\gamma}$ range.

6.3.4 Conclusion

In conclusion, the diphoton identification using a fully connected neural network developed in this work, based on photon shower shape variables and their correlations as a function of $\Delta R_{\gamma\gamma}$, is highly efficient. The prompt diphoton efficiency is close to one, while the fake efficiency is very low, below 0.25, and even lower at low masses.

Further developments are ongoing or remain to be carried out. A study of the correlations between shower shape variables is in progress. Systematic uncertainties must also be estimated. In particular, the shower shape variable distributions in the simulations do not exactly match those observed in the data. To correct for this discrepancy, correction factors are applied to the simulation. These corrections were not applied to the simulated samples used in this study and may represent one of the main sources of systematic uncertainty. Preliminary tests suggest that applying the standard ATLAS correction factors has only a limited impact on the performance, indicating that the changes in correlations between the shower shape variables are similar with or without these correction factors, and that the network can still learn from them.

Moreover, for the JZ samples, which simulate fake photons, only a few thousand events pass the selections defined for this work, out of the two billion generated. The events used to train the neural network lie in the tail of the diphoton distribution and may not be representative of actual fake diphotons.

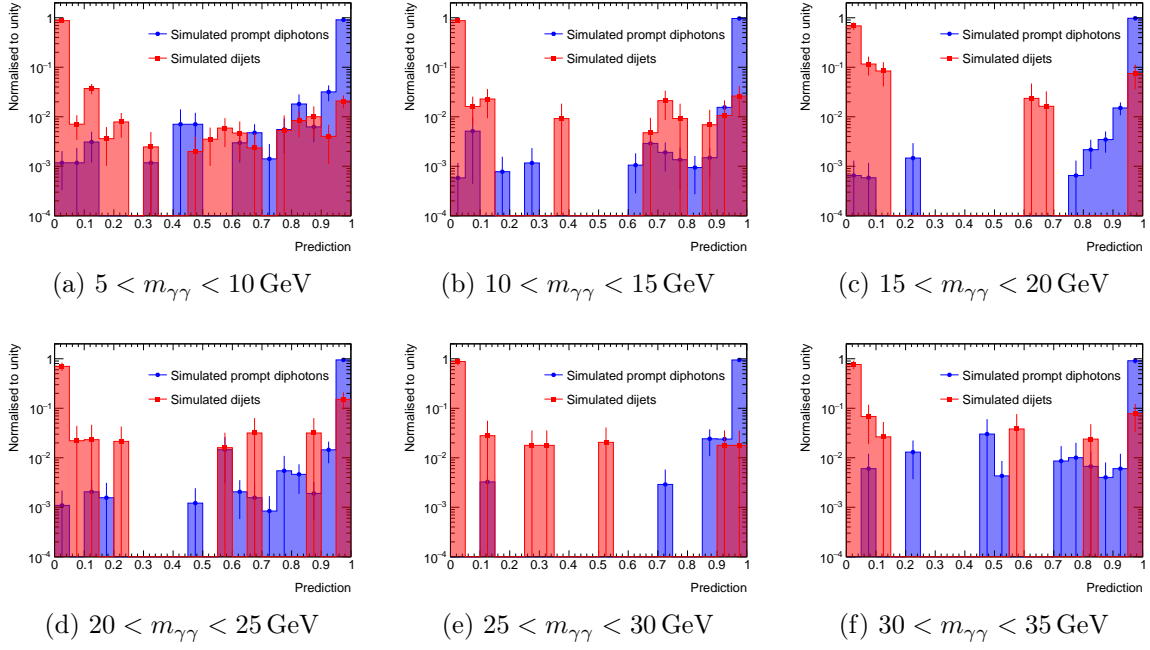
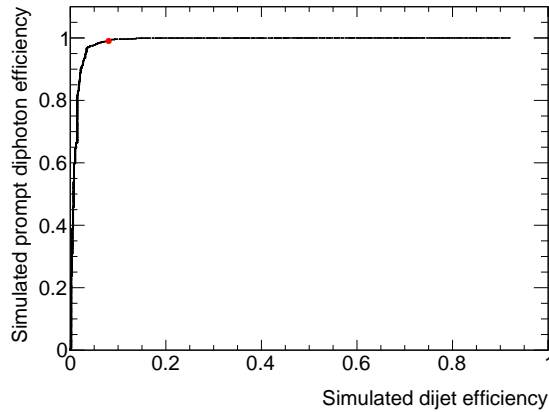
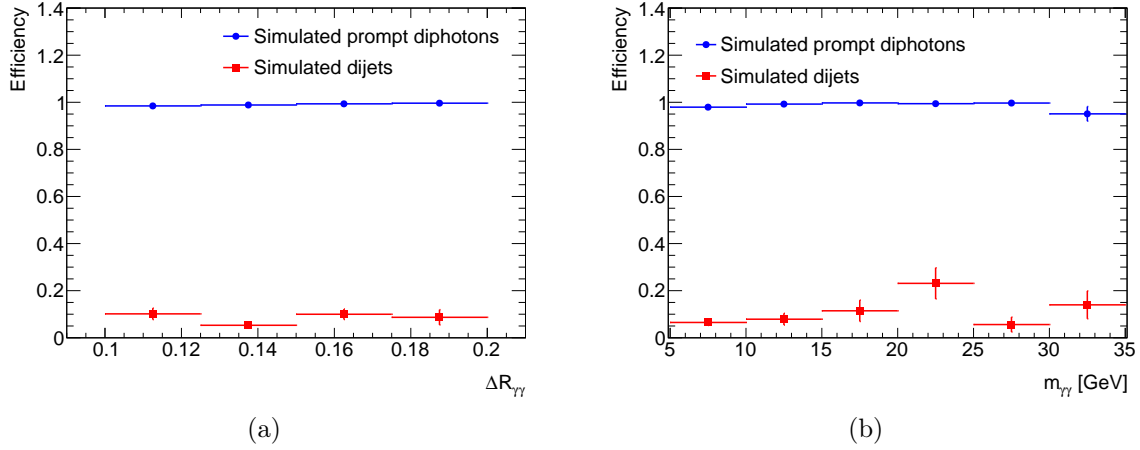
Figure 6.13: Distributions of predictions for prompt and fake diphotons per $m_{\gamma\gamma}$ range.

Figure 6.14: ROC curve. The red dot indicates the threshold position (0.5).

Despite the limited JZ statistics, preliminary tests suggest that the shower shape variables of the fake diphotons are slightly different between the control region in data and the fake diphotons in the simulated JZ sample. The impact of applying a factor similar to the standard ones should be evaluated. $Z \rightarrow ee\gamma$ events could be used, as they provide a very clear signature: the invariant mass of the system composed of the two electrons and the photon must match the mass of the Z boson. Unfortunately, the available simulation statistics for this process are very low.

Figure 6.15: Efficiency per bin of (a) $\Delta R_{\gamma\gamma}$ and (b) $m_{\gamma\gamma}$.

6.4 New diphoton isolation criterion

The standard photon calorimetric isolation applies p_T -dependent cuts on the total energy of clusters located within a fixed-size cone centred on the photon candidate, after pile-up removal and subtraction of the photon candidate's energy. In closely collimated topologies, these criteria are affected by the nearby photon, which falls within the isolation cone of the photon under consideration, thereby reducing the efficiency of the standard photon calorimetric isolation criterion to zero.

In this work, a different approach is proposed. All clusters in the vicinity of the diphoton candidate, excluding those originating from the two photons, are collected and passed through a neural network.

A fixed-size ring centred on the diphoton system is used to select clusters, independently of $\Delta R_{\gamma\gamma}$. The centre of the diphoton system is defined as the mean position of the two photons in the $\eta - \phi$ plane (geometrical centre). The ring's inner and outer radii are set to 0.15 and 0.4, respectively, to exclude the two photon clusters for all diphoton candidates with $0.1 < \Delta R_{\gamma\gamma} < 0.2$.

The number of clusters varies between events. In contrast to the diphoton identification previously described, a simple fully connected neural network cannot handle variable-size inputs. Due to this limitation, three more flexible neural network topologies have been tested: a CNN, a GNN, and a Set Transformer. Their performances, along with their inherent advantages and drawbacks, are presented below.

The diphoton isolation neural networks make use of cluster-level variables: the cluster's transverse energy E_T , its relative position with respect to the geometrical centre of the diphoton system ($\Delta\eta_{\text{cl}-\gamma\gamma}$ and $\Delta\phi_{\text{cl}-\gamma\gamma}$), and its depth λ in the calorimeter. The depth is defined as the longitudinal distance of the cluster centre from the front face of the calorimeter, measured along the shower axis. The cluster centre itself is taken as the energy-weighted 3D position of the cluster, computed as:

$$x_c = \frac{1}{E_{\text{norm}}} \times \sum_{\{\text{cell}|E_{\text{cell}}>0\}} E_{\text{cell}} x_{\text{cell}}, \quad (6.4)$$

$$\text{with } E_{\text{norm}} = \sum_{\{\text{cell}|E_{\text{cell}}>0\}} E_{\text{cell}}. \quad (6.5)$$

Here, x_c denotes the energy-weighted position of the cluster along the x -axis (analogous expressions hold for y_c and z_c). The sum runs over all calorimeter cells with positive energy, each located at $(x_{\text{cell}}, y_{\text{cell}}, z_{\text{cell}})$.

The neural networks also use event-level information: the mean number of interactions per bunch crossing, $\langle\mu\rangle$. The variables used by the neural networks are summarised in Table 6.5. These variables have been normalised by subtracting their mean and dividing by their standard deviation across all training and validation inputs. The test dataset is normalised using the mean and standard deviation computed from the training and validation inputs. The distributions of E_T and λ for simulated prompt diphotons and fake diphotons from data are shown in Figure 6.16.

<i>Variable</i>	<i>Notation</i>
Cluster relative η position with respect to the diphoton	$\Delta\eta_{\text{cl}-\gamma\gamma}$
Cluster relative ϕ position with respect to the diphoton	$\Delta\phi_{\text{cl}-\gamma\gamma}$
Cluster depth	λ
Cluster transverse energy	E_T
Mean number of interactions per bunch crossing	$\langle\mu\rangle$

Table 6.5: Variables used by the diphoton isolation neural networks.

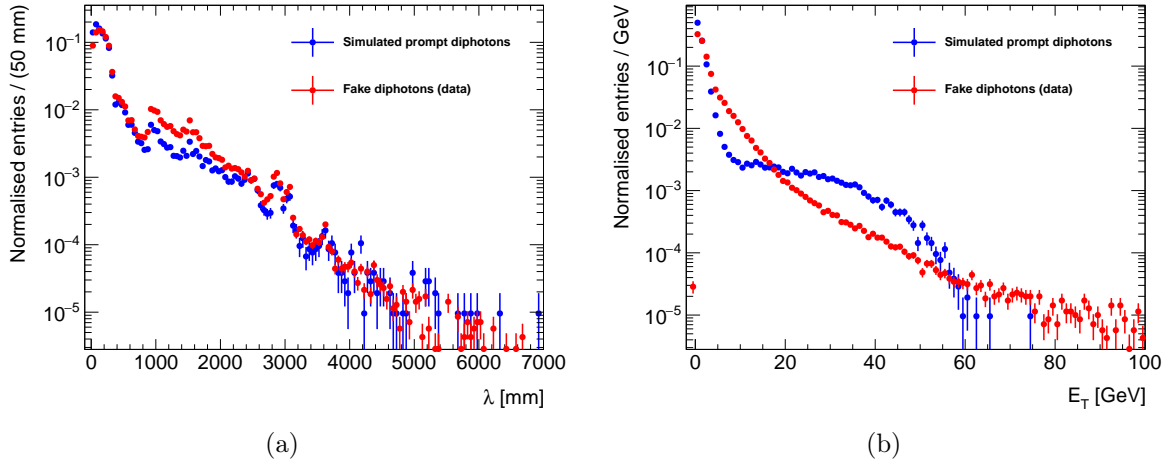


Figure 6.16: Distributions of (a) λ and (b) E_T , for simulated prompt diphotons and fake diphotons (from data).

Due to the limited size of the simulated fake diphoton dataset, the fake-enriched data sample is used, along with the same simulated prompt diphoton sample as for the identification task. In addition to the selections in Table 6.1 and Table 6.3, only events with $\langle\mu\rangle$ between 30 and 64 are retained, since the available pile-up profiles differ between the two samples. To further prevent the neural networks from learning directly from $\langle\mu\rangle$, the inputs are reweighted to flatten the $\langle\mu\rangle$ distributions, in the same way as for $\Delta R_{\gamma\gamma}$ in the identification study. Furthermore, only events with $5 < m_{\gamma\gamma} < 20$ GeV are retained, to optimise the training on low-mass diphoton candidates. As will be shown by the estimated background rates in the next chapter, almost no prompt diphotons with masses above 20 GeV pass the selections, making the 5 – 20 GeV range of particular interest.

A total of 22 127 prompt and 102 483 fake diphotons pass the selections. For each neural network presented in the following, 15 % of the available inputs are set aside to evaluate the model’s performance after training. A total of 50 independent training runs have been performed, each stopped if the loss did not decrease for 40 consecutive epochs. The training and validation datasets, corresponding to 70 % and 15 % of all inputs, respectively, are randomly reshuffled before each training run. Inputs are also reweighted to account for the differing amounts of each class. The total weights, which account for the $\langle\mu\rangle$ distributions and the imbalanced number of inputs in each category, are used in the loss function, as in Eq. 6.3.

6.4.1 Diphoton isolation with a CNN

6.4.1.1 Input preparation and model architecture

A Convolutional Neural Network (CNN) takes “images” as inputs and applies convolution operations to them (see Section 5.6.1). The cluster distribution around a diphoton candidate is represented as such an image, centred on the diphoton candidate. Each cluster is located in one “pixel”, which corresponds to a cell of the second layer of the calorimeter. The cluster spread is not taken into account; instead, each cluster is placed at the position of its barycentre. The image has dimensions of 32×32 cells, to include all clusters within $0.15 < \Delta R < 0.4$ around the diphoton candidate, a L2 calorimeter cell being of size 0.025×0.025 in the η - ϕ plane. Among the variables listed in Table 6.5, λ , E_T , and $\langle\mu\rangle$ are associated with the clusters, each forming a separate input channel. $\Delta\eta_{\text{cl}-\gamma\gamma}$ and $\Delta\phi_{\text{cl}-\gamma\gamma}$ define the position of the cluster in the image. An example of an input image is shown in Figure 6.17, for the normalised transverse energy.

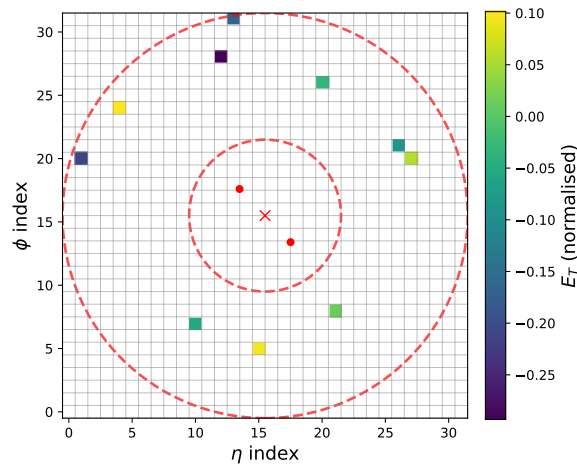


Figure 6.17: Example of an input image used in the CNN for an event from the fake-enriched data sample, with $\langle\mu\rangle = 40.6$. The grid represents 32×32 cells of the second layer of the calorimeter, centred on the photon pair candidate, shown as two red dots with a red cross at their geometric centre. Clusters are selected within $0.15 < \Delta R < 0.4$ from the diphoton centre. These limits are indicated by red dotted circles. Coloured cells contain a cluster, with the colour representing the normalised cluster transverse energy. Similar images are generated for the other input channels—cluster depth and pile-up.

The model comprises two-dimensional convolutional layers, two-dimensional max pooling, and ReLU activation functions, followed by linear transformations, with dropout ap-

plied and a sigmoid activation in the final layer. These components were selected for their good performance.

The complete model is shown in Figure 6.18.

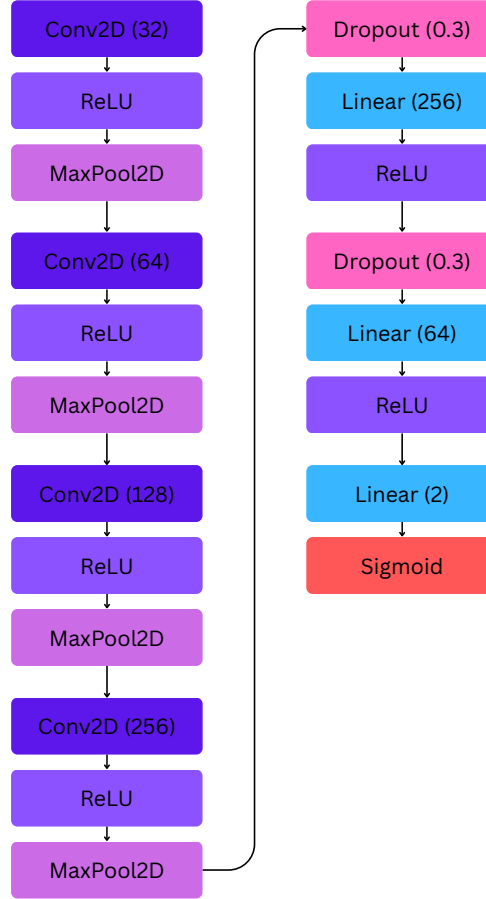


Figure 6.18: Architecture of the CNN used for diphoton isolation. The inputs have three channels: the cluster’s transverse energy E_T , depth λ and the mean number of interactions $\langle\mu\rangle$. The cluster relative position with respect to the diphoton, $(\eta_{\text{cl}-\gamma\gamma}, \phi_{\text{cl}-\gamma\gamma})$, is encoded in the image pixels: each cluster is placed in the pixel corresponding to its position relative to the diphoton, which is located at the centre of the image. The output dimensions of the layers are indicated in brackets. The convolutional layers use a kernel of size 3×3 with padding of 1. The pooling operations use a kernel of size 2×2 with a stride of 2.

Binary cross-entropy (BCE) is used as the loss function, and stochastic gradient descent (SGD) is employed as the optimiser. The main hyperparameters of the model are listed in Table 6.6.

<i>Parameter</i>	<i>Value</i>
Batch size	64
Learning rate	0.001
Momentum	0.9

Table 6.6: Hyperparameters of the CNN.

6.4.1.2 Performance

The training and validation losses for the best model out of the 50 training runs are shown in Figure 6.19. The distribution of predictions evaluated on the test dataset is shown in Figure 6.20. Results are also presented as a function of $m_{\gamma\gamma}$ and $\Delta R_{\gamma\gamma}$ in Figure 6.21 and Figure 6.22. The ROC curve is shown in Figure 6.23. It shows the efficiency for prompt diphotons as a function of the efficiency for fake diphotons.

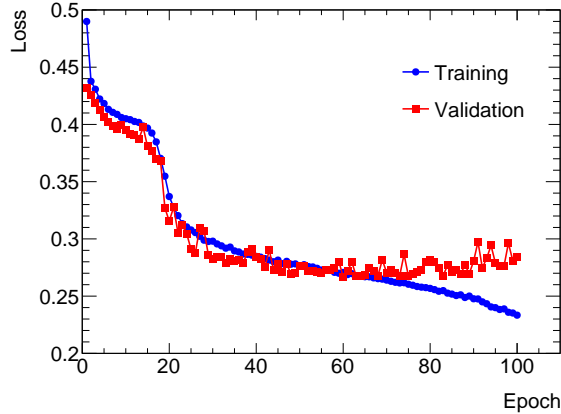


Figure 6.19: Training and validation losses vs. epochs for the best training run, with the CNN.

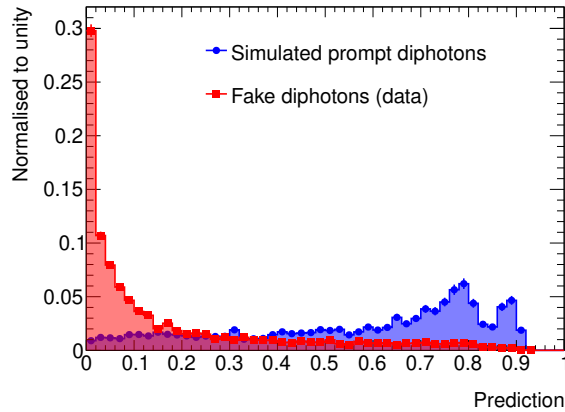


Figure 6.20: Distribution of CNN predictions for prompt and fake diphotons.

The selection efficiencies are shown in Figure 6.24 as a function of $m_{\gamma\gamma}$ and $\Delta R_{\gamma\gamma}$. Due to insufficient statistics to compute the threshold that maximises sensitivity, a fixed threshold of 0.25 was chosen, which offers good performance for all $m_{\gamma\gamma}$ and $\Delta R_{\gamma\gamma}$ bins.

The CNN performs reasonably well, but this model presents significant limitations for this task. First, the exact position of each cluster is lost, being replaced by an approximate value determined by the cell sizes. Although the clusters' η and ϕ coordinates could be added as features, this would not constitute an optimal use of the available spatial information.

Moreover, a 3D CNN would, in principle, be better suited to this problem. However, the calorimeter layers depend on $|\eta|$, and the cell sizes vary from one layer to another, making it difficult to construct a consistent three-dimensional image. In addition, the

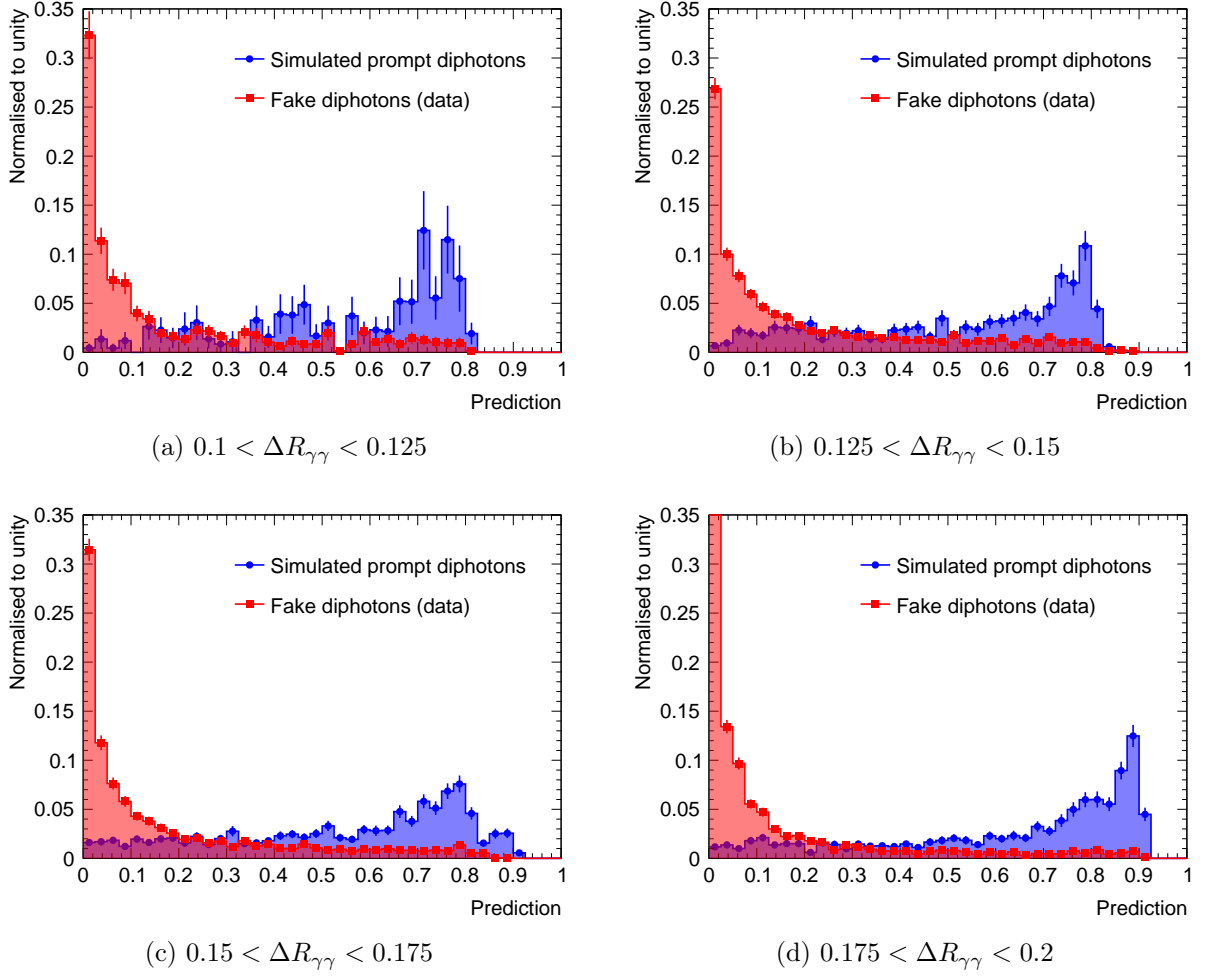


Figure 6.21: Distribution of CNN predictions for prompt and fake diphotons per $\Delta R_{\gamma\gamma}$ range.

input images are mostly sparse, resulting in high memory usage with limited informational content.

For these reasons, a Graph Neural Network (GNN) has also been employed, as it does not rely on fixed-size cells and can better exploit the variable and irregular structure of the data.

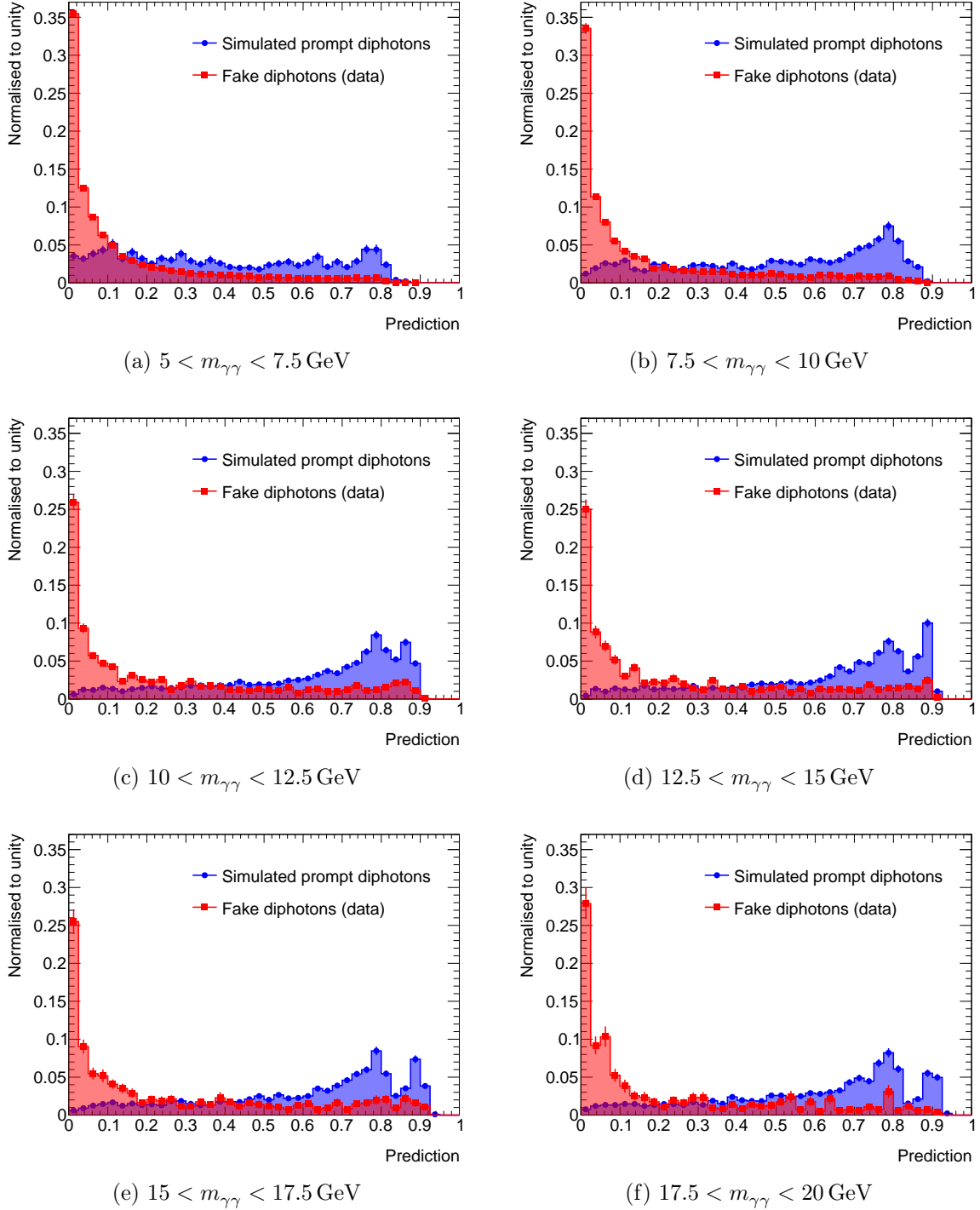


Figure 6.22: Distributions of CNN predictions for prompt and fake diphotons per $m_{\gamma\gamma}$ range.

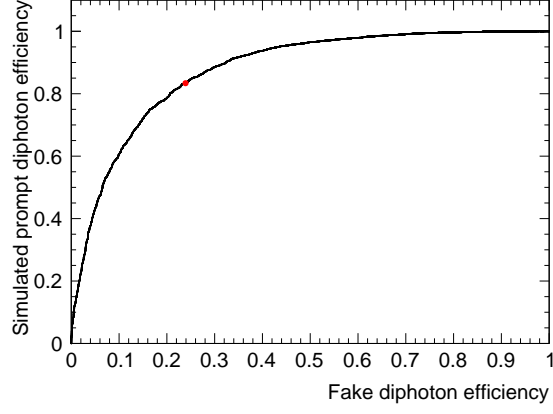


Figure 6.23: ROC curve for the CNN. The red dot indicates the threshold position (0.25).

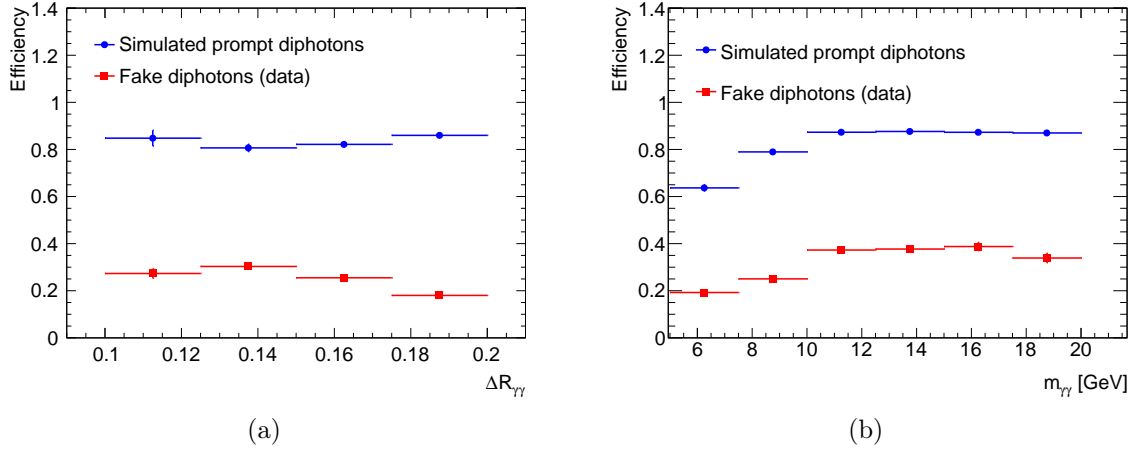


Figure 6.24: Efficiency per bin of (a) $\Delta R_{\gamma\gamma}$ and (b) $m_{\gamma\gamma}$, with the CNN.

6.4.2 Diphoton isolation with a GNN

6.4.2.1 Input preparation and model architecture

As detailed in Chapter 5, a Graph Neural Network (GNN) can perform various tasks, including graph classification and node prediction. Diphoton isolation corresponds to the former. The cluster distribution around diphoton candidates is therefore modelled as a graph, where each node represents a cluster. As in the case of the CNN, the normalised variables listed in Table 6.5 are included in the nodes: the three-dimensional cluster position, with respect to the diphoton, ($\Delta\eta_{\text{cl}-\gamma\gamma}$, $\Delta\phi_{\text{cl}-\gamma\gamma}$, and depth λ), cluster E_T , and the mean number of interactions per bunch crossing, $\langle\mu\rangle$. Since the network cannot process graphs without any nodes, a single cluster was added to inputs that contained no clusters, using its corresponding $\langle\mu\rangle$ value, while all other variables were set to zero.

There is no obvious way to define the connections between clusters. Several configurations have been tested, including connecting all clusters or none. Different edge weights were also explored, based on either the distance between clusters or their energy differences. The results shown are based on the ParticleNet architecture [161]: each cluster is connected to its five nearest neighbours, based on their angular distance. No weights are applied. An example of a graph input is shown in Figure 6.25.

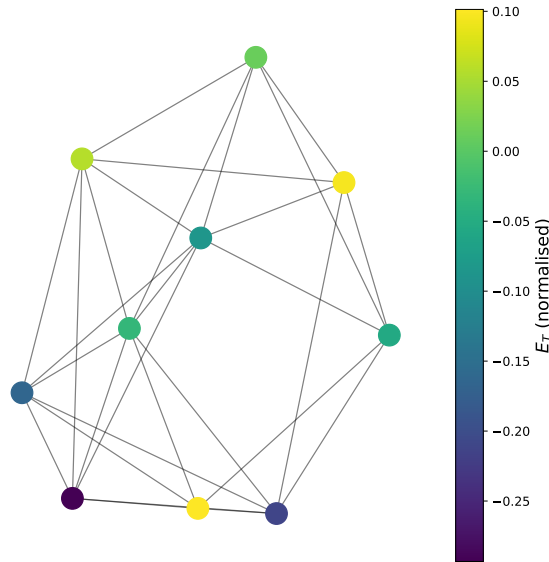


Figure 6.25: Example of a graph input used in the GNN, corresponding to the same event as in Figure 6.17. The nodes represent clusters, and the colour indicates the cluster normalised E_T . In this example, each node is connected to its five nearest neighbours, and no weights are applied to the edges, but other configurations of connexions were tested. Similar representations can be produced for the other cluster variables.

The network consists of edge convolution (“EdgeConv”) operations [162], which takes the form, for each cluster x_i :

$$\mathbf{x}'_i = \frac{1}{5} \sum_{j=1}^5 \mathbf{h}_{\Theta}(\mathbf{x}_i, \mathbf{x}_{i_j} - \mathbf{x}_i), \quad (6.6)$$

where $\mathbf{x}_i \in \mathbb{R}^5$ is the feature vector of the cluster x_i , $\{i_1, \dots, i_5\}$ are the indices of the five nearest clusters of the cluster x_i . \mathbf{h}_{Θ} is a function parametrised by a set of

learnable parameters Θ , shared for all clusters, which substitutes the feature vectors of the neighbours, \mathbf{x}_{i_j} , by their difference from the cluster \mathbf{x}_i . It can be implemented as an MLP. Batch normalisation [163] is used to stabilise the training and improve the performance. Several EdgeConv operations can be stacked. The structure of the EdgeConv block and the model architecture are shown in Figure 6.26.

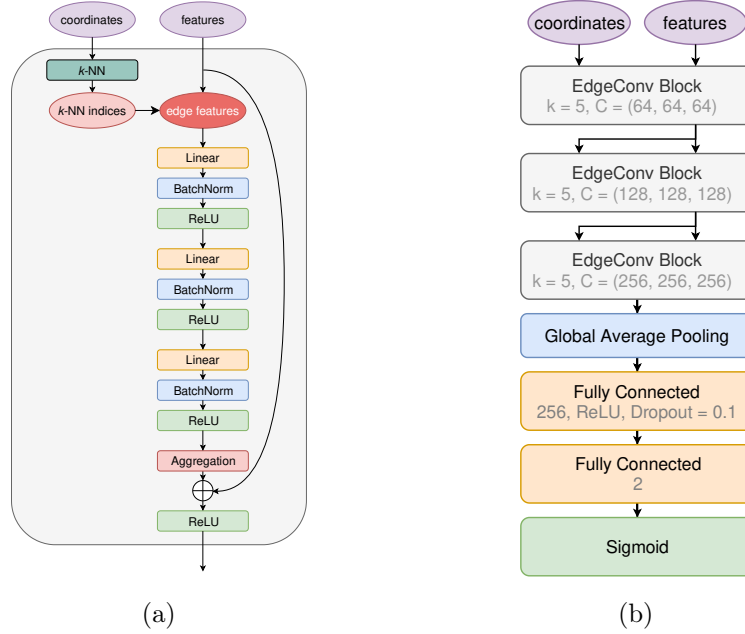


Figure 6.26: (a) Structure of the EdgeConv block. (b) Architecture of the GNN, based on ParticleNet. Images adapted from Ref. [161].

Binary cross-entropy (BCE) is used as the loss function, along with the Adam optimiser. The main hyperparameters of the model are listed in Table 6.7.

<i>Parameter</i>	<i>Value</i>
Batch size	64
Learning rate	0.001

Table 6.7: Hyperparameters of the GNN.

6.4.2.2 Performance

The training and validation losses for the best model out of the 50 training runs are shown in Figure 6.27. The distribution of predictions evaluated on the test dataset is shown in Figure 6.28. Results are also presented as a function of $m_{\gamma\gamma}$ and $\Delta R_{\gamma\gamma}$ in Figure 6.29 and Figure 6.30. The ROC curves for the CNN and the GNN are shown in Figure 6.31. The GNN performs slightly better than the CNN.

The selection efficiencies are shown in Figure 6.32 as a function of $m_{\gamma\gamma}$ and $\Delta R_{\gamma\gamma}$. The same samples with insufficient statistics are used, making it difficult to choose the optimal threshold. A threshold of 0.5 was retained.

Despite offering a more flexible representation of the cluster distribution compared to the CNN, the GNN faces a fundamental limitation: there is no principled way to define

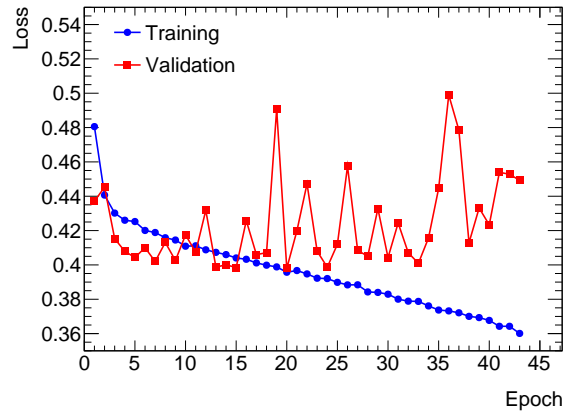


Figure 6.27: Training and validation losses vs. epochs for the best training run with the GNN.

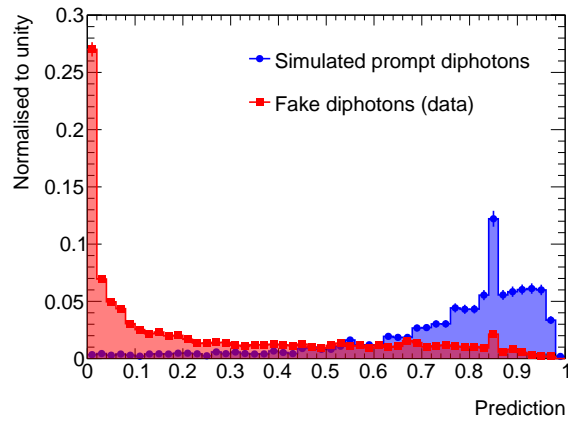


Figure 6.28: Distribution of GNN predictions for prompt and fake diphotons.

connections between clusters. Connecting clusters with their nearest neighbours based on the angular distance provides an imperfect solution. Clusters originating from pile-up do not share a common physical origin and are uniformly distributed in the $\eta - \phi$ plane. Fake clusters may have a common origin within jets, but their energies and positions in the $\eta - \phi$ plane do not necessarily reflect this. Avoiding the use of edges prevents the introduction of handcrafted information but does not fully exploit the graph architecture. Faced with these challenges, a third architecture has been employed: a Set Transformer.

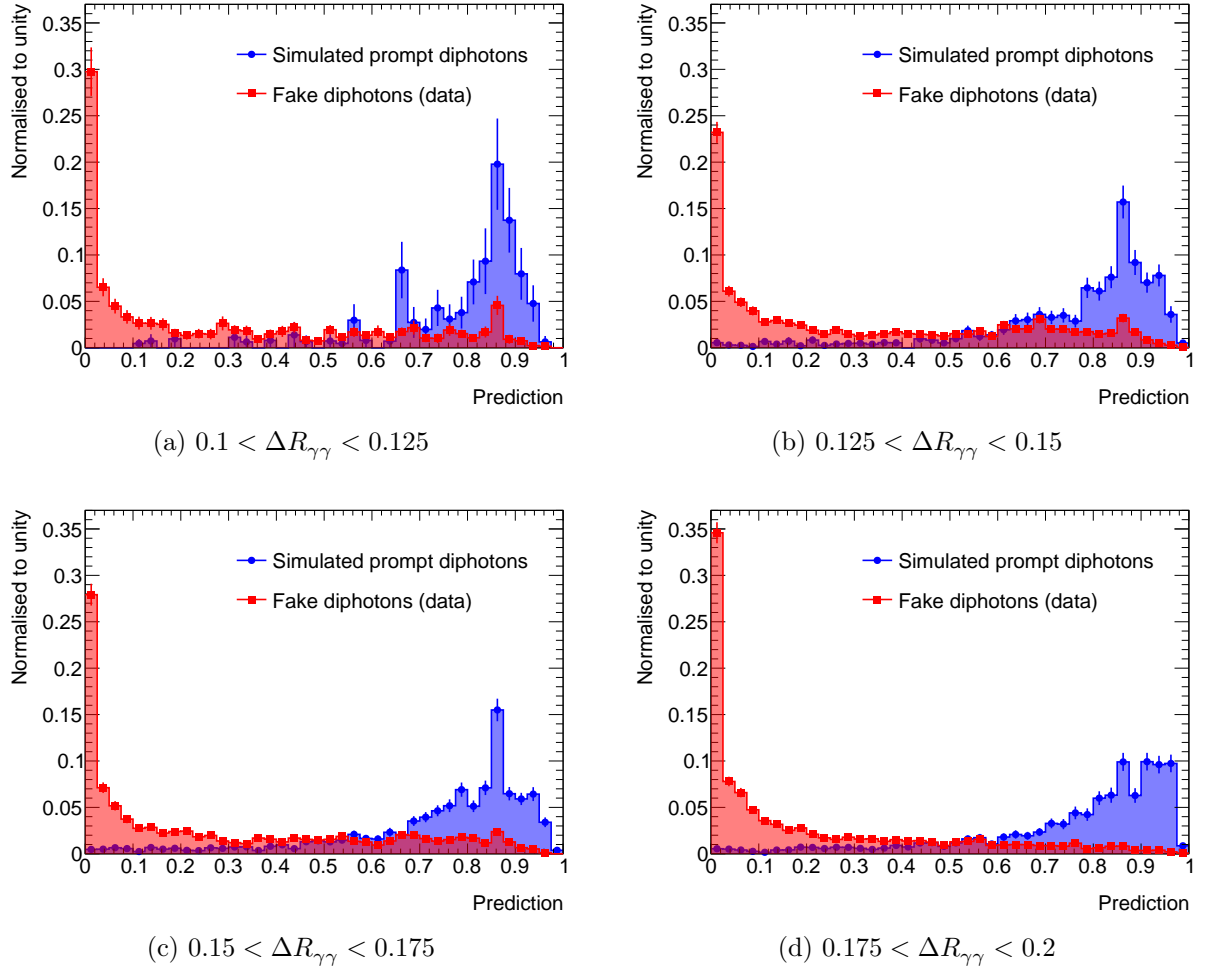


Figure 6.29: Distribution of GNN predictions for prompt and fake diphotons per $\Delta R_{\gamma\gamma}$ range.

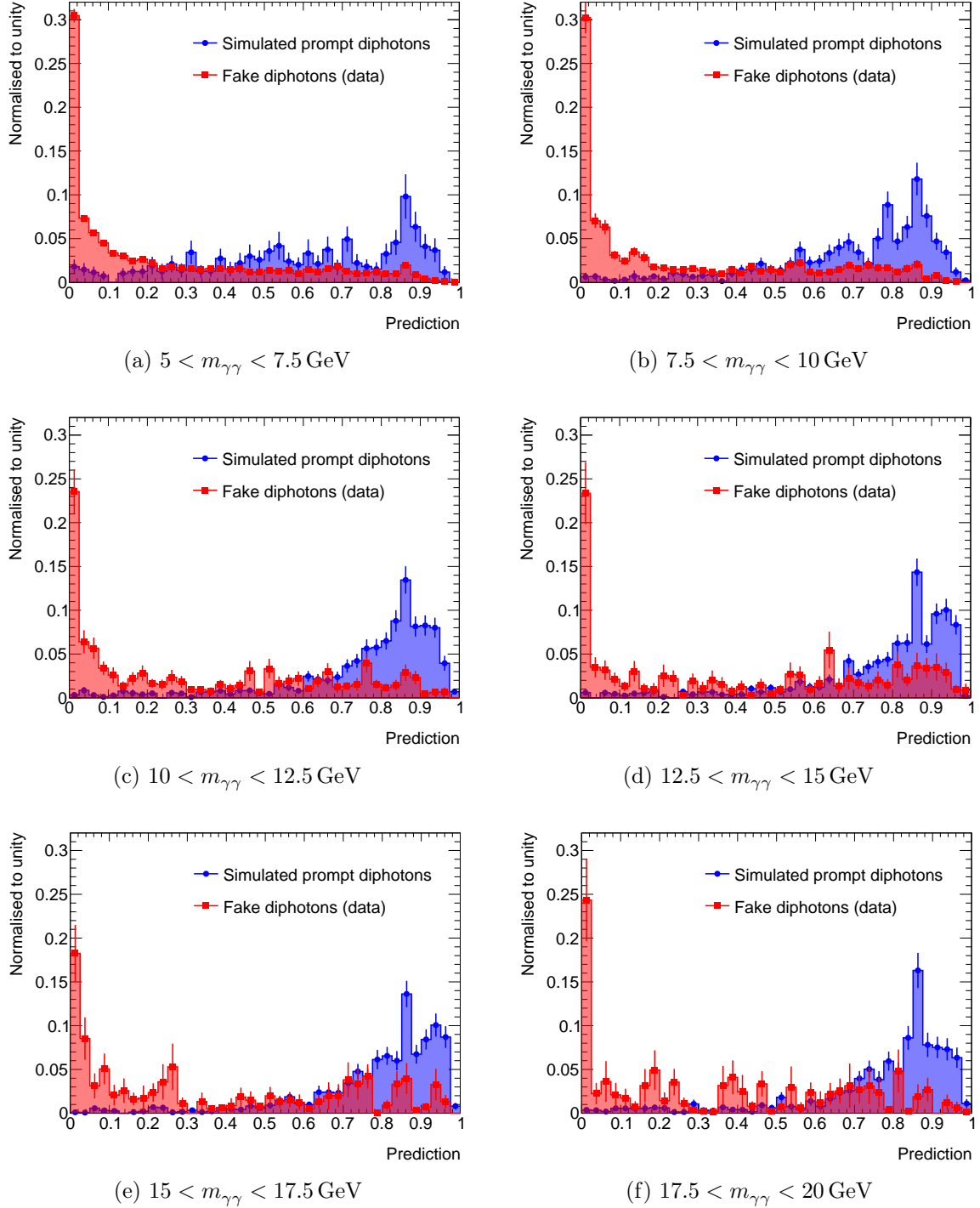


Figure 6.30: Distributions of GNN predictions for prompt and fake diphotons per $m_{\gamma\gamma}$ range.

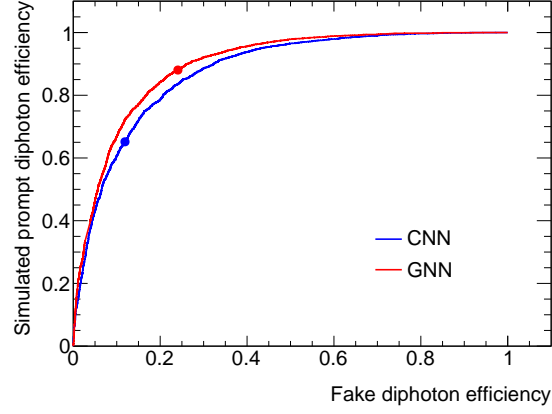


Figure 6.31: ROC curves for the CNN and the GNN. The dots indicates the threshold positions, 0.25 for the CNN and 0.5 for the GNN.

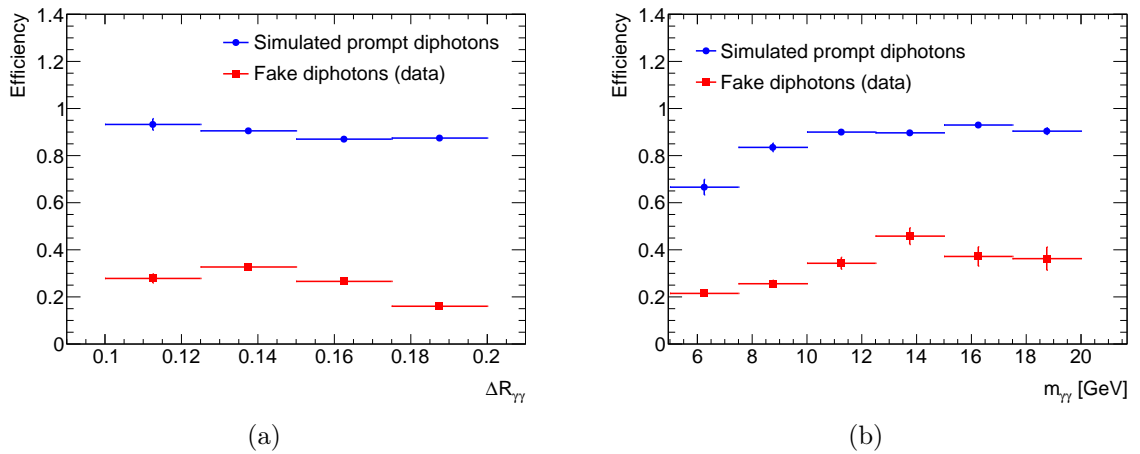


Figure 6.32: Efficiency per bin of (a) $\Delta R_{\gamma\gamma}$ and (b) $m_{\gamma\gamma}$ with the GNN.

6.4.3 Diphoton isolation with a Set Transformer

6.4.3.1 Input preparation and model architecture

As described in Section 5.6.3, the Set Transformer is a neural network architecture capable of processing permutation-invariant and variable-size inputs. It relies on attention mechanisms and does not require manually connecting clusters.

In this model, the input is the list of clusters surrounding a diphoton candidate. The same variables as in the CNN and GNN cases, listed in Table 6.5, are used. They are normalised in the same way, by subtracting their mean and dividing by their standard deviation over all inputs. Similarly to the GNN, a cluster is added to inputs that contain no clusters, with all its variables set to zero except for the pile-up value.

The model architecture employs the modules introduced earlier. It begins with an encoder:

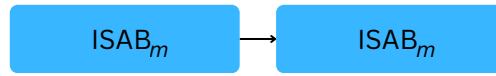


Figure 6.33: Set Transformer encoder.

followed by a decoder:

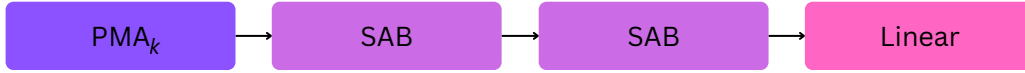


Figure 6.34: Set Transformer decoder.

Finally, a sigmoid function is applied:



Figure 6.35: Set Transformer final sigmoid function.

Layer normalisation is applied [164]. Using the same notations as in the previous chapter, the dimensions of the different modules are listed in Table 6.8.

<i>Parameter</i>	<i>Value</i>
Input dimension (d)	5
Number of inducing points (m)	4
Hidden layers dimension	128
Number of heads (h)	8
Number of vectors in pooling (k)	1
Output dimension	1

Table 6.8: Architecture details of the Set Transformer.

BCE loss is used, along with the Adam optimiser. The hyperparameters of the model are listed in Table 6.9:

<i>Parameter</i>	<i>Value</i>
Batch size	64
Learning rate	0.0001

Table 6.9: Hyperparameters of the Set Transformer.

6.4.3.2 Performance

The training and validation losses for the best model out of the 50 training runs are shown in Figure 6.36. The distribution of predictions evaluated on the test dataset is shown in Figure 6.37. Results are also presented as a function of $m_{\gamma\gamma}$ and $\Delta R_{\gamma\gamma}$ in Figure 6.38 and Figure 6.39. The ROC curves of the three models are shown in Figure 6.40. The performances of the CNN and the Set Transformer are almost identical. The GNN performs slightly better.

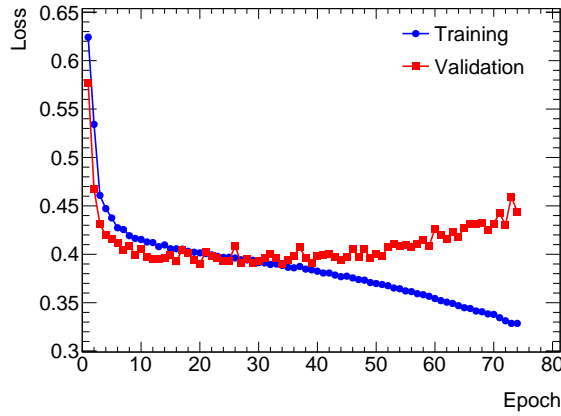


Figure 6.36: Training and validation losses vs. epochs for the best training run with the Set Transformer.

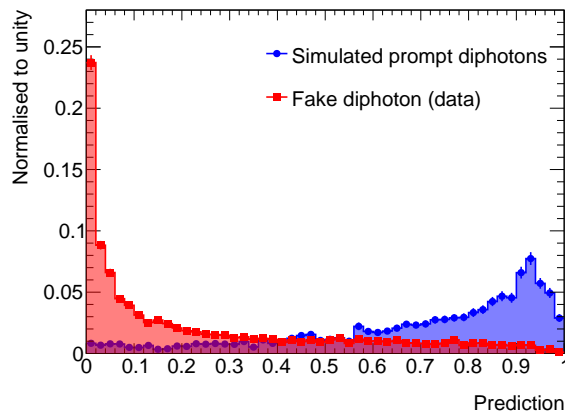


Figure 6.37: Distribution of Set Transformer predictions for prompt and fake diphotons.

The selection efficiencies are shown in Figure 6.41 as a function of $m_{\gamma\gamma}$ and $\Delta R_{\gamma\gamma}$. Due to insufficient statistics to compute the threshold that maximises sensitivity, a fixed threshold of 0.5 was chosen.

This network architecture offers several advantages over the CNN and GNN. It can handle inputs with a variable number of clusters, and its structure is inherently

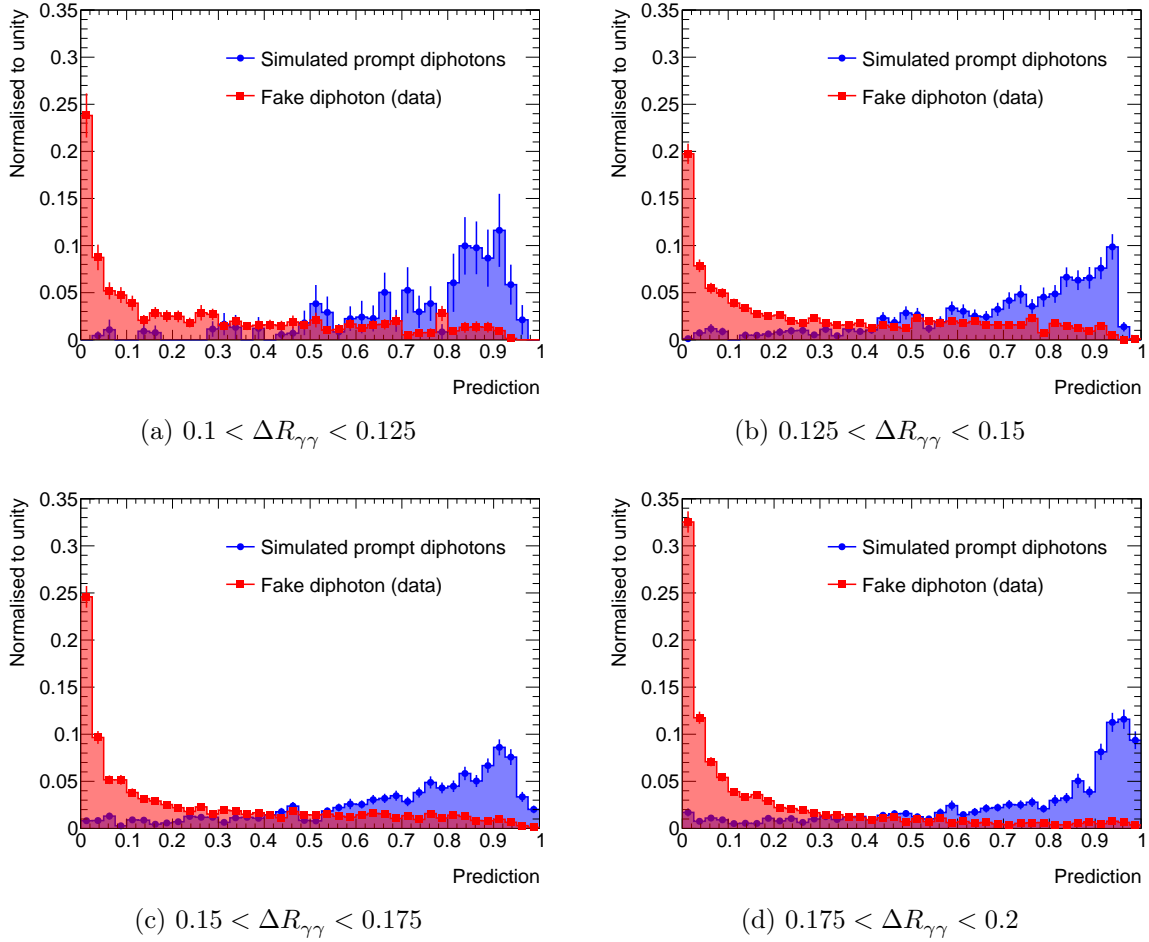


Figure 6.38: Distribution of Set Transformer predictions for prompt and fake diphotons per $\Delta R_{\gamma\gamma}$ range.

permutation-invariant, which is well-suited to the fact that cluster order is irrelevant. Its use of attention mechanisms allows it to learn relationships between clusters without the need to connect them manually. However, all inputs within a batch must have the same size, making zero-padding necessary. Moreover, as with the GNN, the model cannot process inputs without clusters, requiring the addition of a virtual zero-cluster.

6.4.4 Conclusion

In conclusion, the GNN provides the best performance for diphoton isolation. The prompt diphoton efficiency is higher than 65 %, and exceeds 90 % above 10 GeV. The fake efficiency remains below 50 %, and drops even lower at low masses, down to about 20 %.

Additional cluster variables, such as time, could be considered in future studies. Time was not included in this work, as it is poorly modelled in the simulations. More detailed depth information could not be used, such as the deposited energy per layer, but it would be useful. This would allow a pile-up-dependent five-dimensional (3D position, time, and energy) isolation.

A mass-dependent threshold could be employed to improve the model's performance. Such a threshold would adapt the selection criteria according to the mass, helping to prevent the drop in efficiency observed at low masses. This approach would allow for a

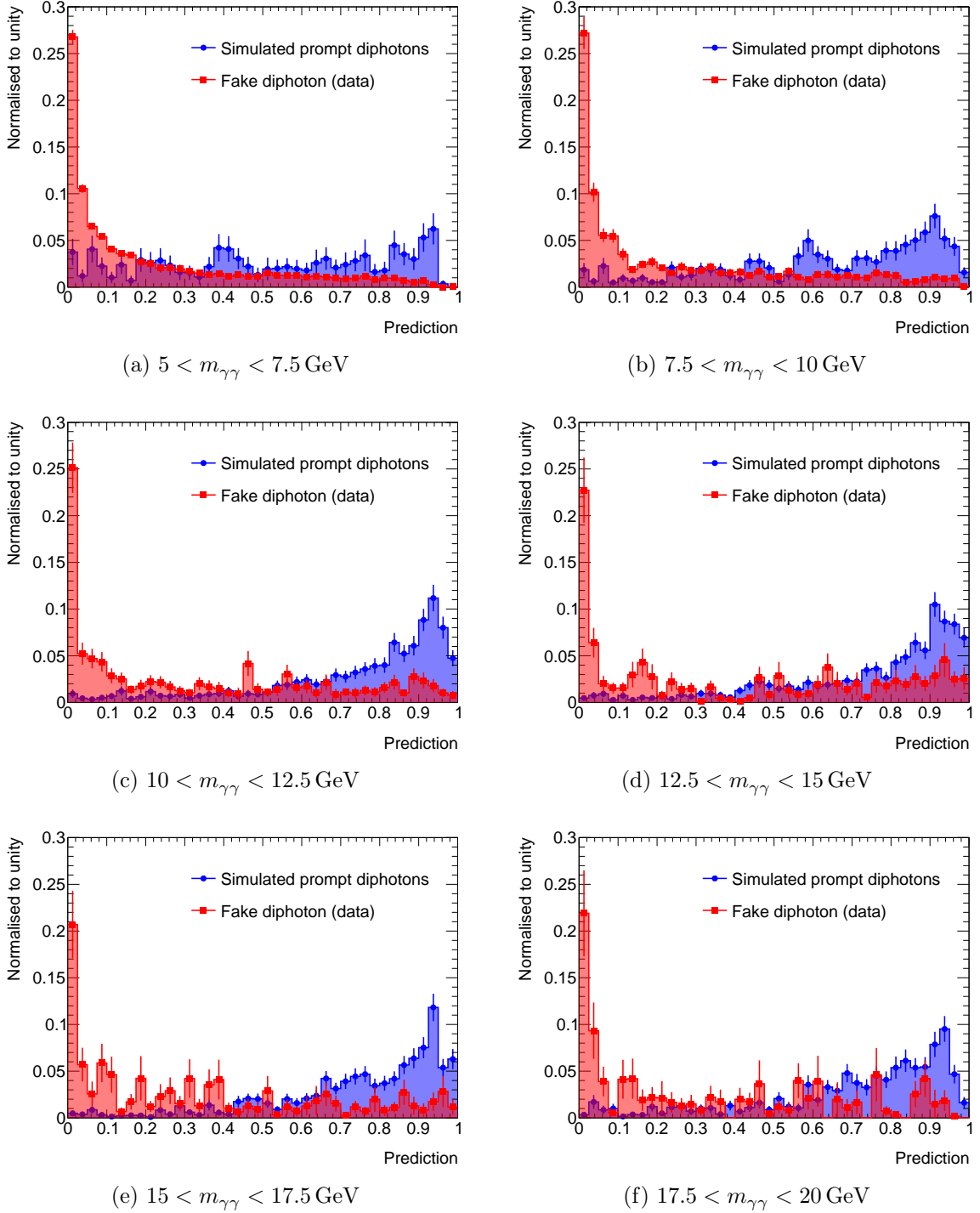


Figure 6.39: Distributions of Set Transformer predictions for prompt and fake diphotons per $m_{\gamma\gamma}$ range.

more uniform performance across the entire mass spectrum.

Some models have inherent limitations: the CNN offers a discrete representation of space, and the GNN lacks a straightforward method for defining connections between clusters.

The fake-enriched sample in the data was constructed using a cut on track isolation, which is correlated with the calorimetric isolation studied in this work. This correlation

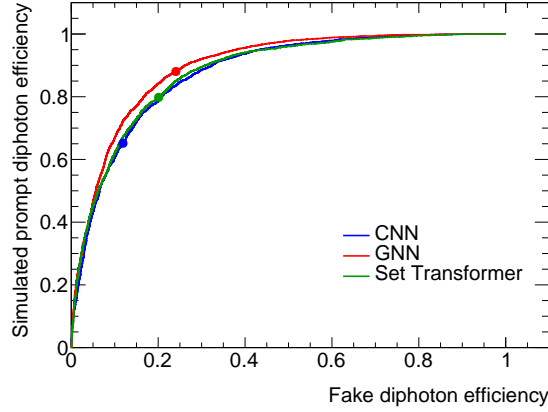


Figure 6.40: ROC curves for the CNN, the GNN and the Set Transformer. The dot indicates the thresholds positions (0.25 for the CNN, 0.5 for the GNN and 0.4 for the Set Transformer).

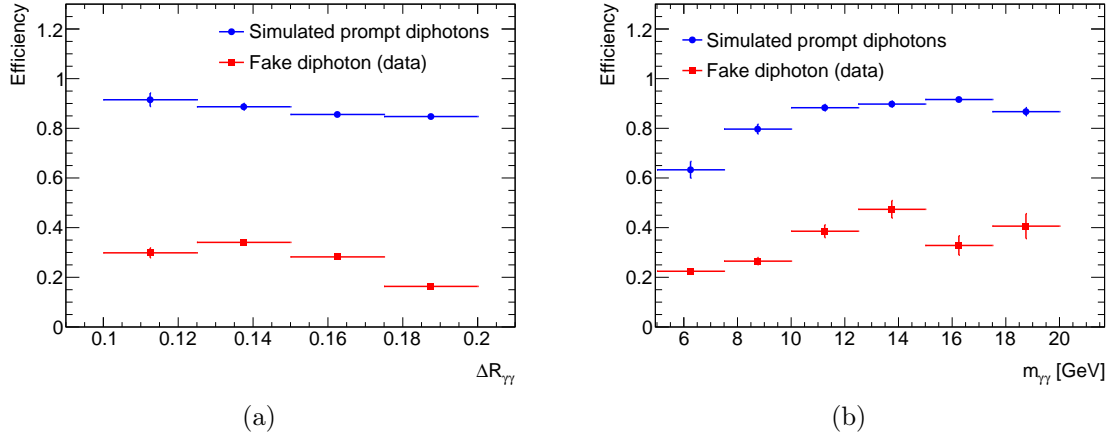


Figure 6.41: Efficiency per bin of (a) $\Delta R_{\gamma\gamma}$ and (b) $m_{\gamma\gamma}$ with the Set Transformer.

should be investigated in future analyses. Moreover, events in this fake-enriched region in data can differ slightly from simulated fake diphotons. A way to assess the systematic uncertainty arising from the use of real or simulated samples could be to measure the model's performance with closely matching data and simulated samples. $Z \rightarrow ee\gamma$ events could be used, thanks to the very clear signature they provide. Unfortunately, as previously mentioned, the available simulation statistics are very limited.

Another interesting test consists of passing the simulated fake events used for the identification task through the model trained on the fake-enriched sample from data.

Chapter 7

Sensitivity to a new diphoton resonance

Contents

7.1	Background components	162
7.2	Mass resolution	169
7.3	Effective background and expected limit	171

This chapter is dedicated to the search for a new resonance in the diphoton channel, with masses down to 5 GeV, which can be interpreted within the framework of axion-like particles. The neural networks developed in the previous chapter are used to estimate the event rates of the reducible and irreducible backgrounds, as well as the effective background as a function of the resonance mass. This work has benefited from the contributions of Anne-Gabrielle Gibeili, an intern I co-supervised.

7.1 Background components

In a search for a closely collimated diphoton resonance, four background components must be considered: prompt photon pairs of QCD origin, fake photon pairs from jets, and the $Z \rightarrow e^+e^-\gamma$ and $W^\pm \rightarrow e^\pm\nu_e\gamma$ processes, in which an electron is reconstructed as a photon and an electron-photon pair passes the diphoton selections.

The $Z \rightarrow e^+e^-\gamma$ and $W^\pm \rightarrow e^\pm\nu_e\gamma$ processes have been estimated using PYTHIA 8 simulated samples. Applying the selections in Table 6.2, except for jet matching, the selection efficiency is 4.0×10^{-11} for the radiative Z decay, and 1.5×10^{-11} and 5.3×10^{-11} for the W^+ and W^- decays, respectively. The mass distributions of the events passing the selections are shown in Figure 7.1. For an integrated luminosity of 500 fb^{-1} , this corresponds to a few dozen events at low mass, to be compared with the thousands of prompt diphotons of QCD origin, as shown in the following. The radiative W^\pm and Z decays do not contribute significantly to the background. Moreover, as will be shown in Figure 7.5, their spectra appear to be sufficiently similar to the ones from prompt and fake diphotons, so that it does not need a specific treatment.

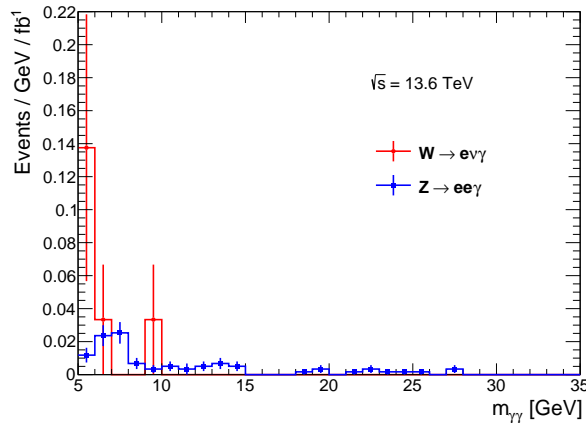


Figure 7.1: Mass distributions of diphoton events passing the selections, originating from radiative Z and W^\pm decays.

To estimate the fake and prompt diphoton backgrounds, a blind and data-driven estimation is performed by extrapolating from a control region (CR) to the signal region (SR). The regions are defined to maximise the prompt diphoton rate and minimise the fake rate in the signal region. The selections applied are orthogonal, which allows the control region to be examined without unblinding the signal region. The control region in data is defined using the track isolation variable p_T^{cone30} , introduced in Section 4.2. By definition, only tracks with transverse momentum larger than 1 GeV are included in the track isolation computation. The p_T^{cone30} distribution thus exhibits a peak at 0, corresponding to events with zero tracks in the region, a gap between 0 and 1 GeV, and

is continuous above 1 GeV. The signal region is defined by events having at least one isolated photon, i.e.:

$$\min(p_{\text{T,leading}}^{\text{cone30}}, p_{\text{T,subleading}}^{\text{cone30}}) = 0, \quad (7.1)$$

which is very tight. To simplify notation, this quantity is denoted $p_{\text{T,min}}^{\text{cone30}}$ in the following.

The control region is defined by events with:

$$1 < p_{\text{T,min}}^{\text{cone30}} < 10 \text{ GeV}. \quad (7.2)$$

The PG simulated prompt diphoton sample shows that the track isolation for prompt photons decreases sharply, as shown in Figure 7.2, where the distributions have been fitted with two exponential functions between 1 and 10 GeV, in bins of $m_{\gamma\gamma}$. The mass bins have been chosen sufficiently large compared to the resolution so as not to reveal a potential new resonance. The mass resolution is studied in the next section.

Moreover, these distributions can be used to estimate the prompt diphoton component in the signal region: the fraction of prompt diphoton events in the control region is assumed to be the same in the PG sample and in the data. These fractions, computed from the PG sample, are given in Table 7.1, per mass bin.

<i>Mass bin</i>	<i>Fraction</i>
$5 < m_{\gamma\gamma} < 7.5 \text{ GeV}$	0.162 ± 0.007
$7.5 < m_{\gamma\gamma} < 10 \text{ GeV}$	0.144 ± 0.005
$10 < m_{\gamma\gamma} < 15 \text{ GeV}$	0.154 ± 0.003
$15 < m_{\gamma\gamma} < 25 \text{ GeV}$	0.157 ± 0.003
$25 < m_{\gamma\gamma} < 35 \text{ GeV}$	0.153 ± 0.005

Table 7.1: Fraction of prompt diphotons with $1 < p_{\text{T,min}}^{\text{cone30}} < 10 \text{ GeV}$ relative to the total number of prompt diphotons with $p_{\text{T,min}}^{\text{cone30}} < 10 \text{ GeV}$, computed on the PG sample.

The JZ samples have too few statistics to model the fake component of the track isolation properly. However, the distributions of $p_{\text{T,min}}^{\text{cone30}}$ for the JZ slices with non-negligible statistics show that the track isolation is roughly constant in the control region for the fake diphoton component. These distributions are shown in Figure 7.3 in separate slices to avoid applying weights.

These distributions justify modelling the prompt diphoton component with an exponential function and the fake component with a constant. Since almost all prompt diphotons have $p_{\text{T,min}}^{\text{cone30}} < 2 \text{ GeV}$, a simple exponential function is sufficient to fit the prompt component in the data, and there is no sensitivity to the contribution of the second exponential.

A fit was subsequently performed on the data control region using a combined model of an exponential function and a constant:

$$f(p_{\text{T,min}}^{\text{cone30}}) = A \exp[c(p_{\text{T,min}}^{\text{cone30}} - 1 \text{ GeV})] + b \quad (7.3)$$

The distributions, along with their fits, are shown in Figure 7.4.

Compared to the track isolation in the simulated PG sample modelled in Figure 7.2, the exponential component decreases more sharply in the data.

The estimated numbers of prompt diphotons in the control region are extrapolated to the signal region using the fractions in Table 7.1. The fake component cannot be

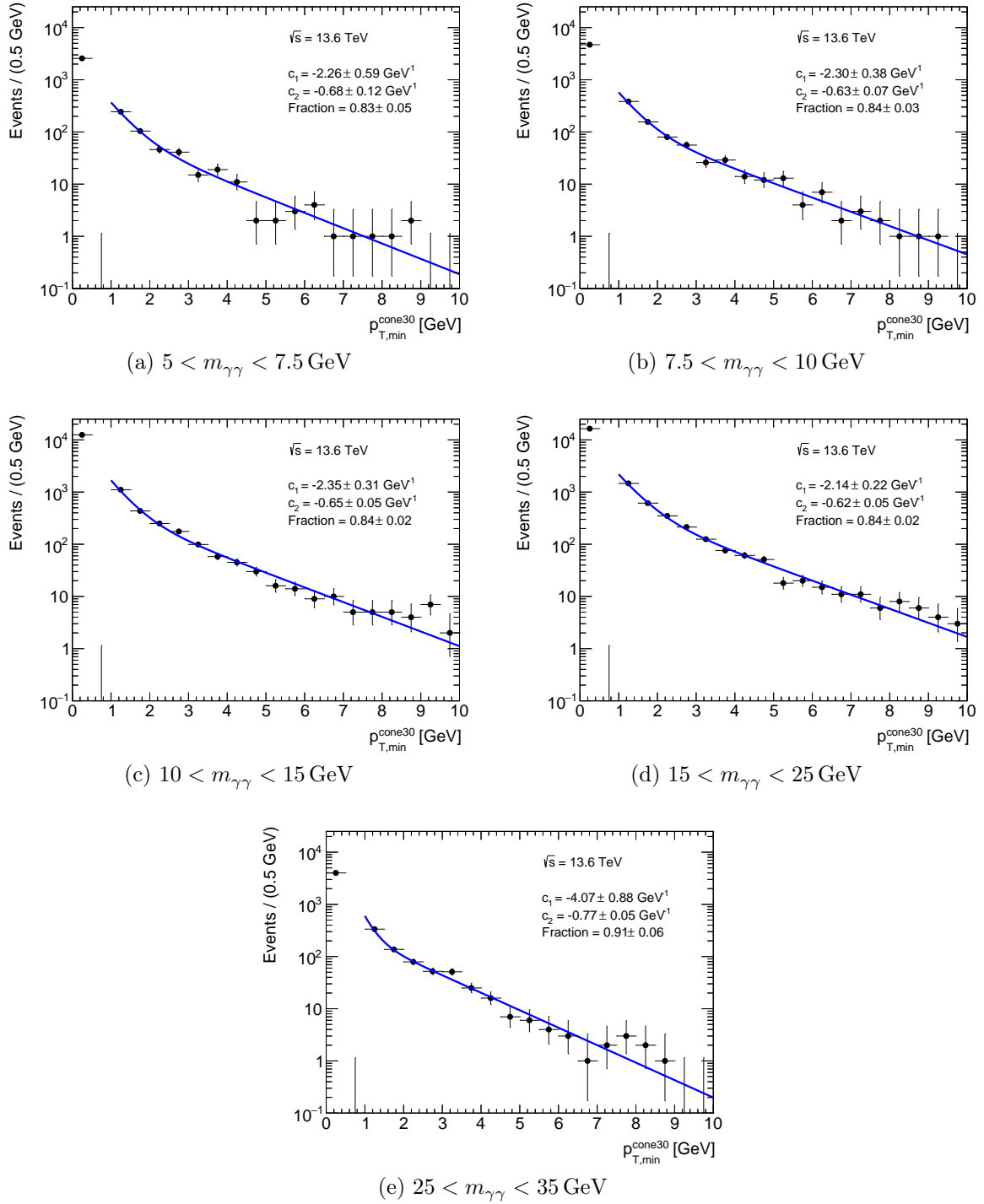


Figure 7.2: Fits of the $p_{T,\min}^{\text{cone30}}$ distributions in the PG simulated prompt diphoton sample, shown in bins of $m_{\gamma\gamma}$. The exponential coefficients and the fraction of the first exponential in the total model are shown in the plots.

extrapolated, as the JZ samples are not reliable. It is calculated as the difference between the number of events in the signal region and the estimated number of prompt diphoton events in the signal region:

$$N_{\text{fakes}}^{\text{signal}} = N_{\text{total}}^{\text{signal}} - N_{\gamma\gamma}^{\text{signal}}, \quad (7.4)$$

where $N_{\text{fakes}}^{\text{signal}}$ is the estimated number of fake diphotons in the signal region, $N_{\text{total}}^{\text{signal}}$ is the measured total number of events in the signal region, and $N_{\gamma\gamma}^{\text{signal}}$ is the estimated number of prompt diphoton events in the signal region. The number of signal events is assumed to be negligible compared to the background (before applying diphoton identification and isolation).

The background rates are shown in Figure 7.5. The prompt diphoton yield is low at large masses, justifying the choice to train the neural networks with events having masses up to 20 GeV.

The diphoton identification and isolation efficiencies from the neural networks developed in Chapter 6 can be applied to the background rates. They are shown in Figure 7.6. The irreducible background then becomes largely dominant. The total background is modelled using the following function:

$$f(m_{\gamma\gamma}) = A \exp[c(m_{\gamma\gamma} - 5 \text{ GeV})]. \quad (7.5)$$

The fitted parameters, along with the χ^2 divided by the number of degrees of freedom, are indicated in the figure.

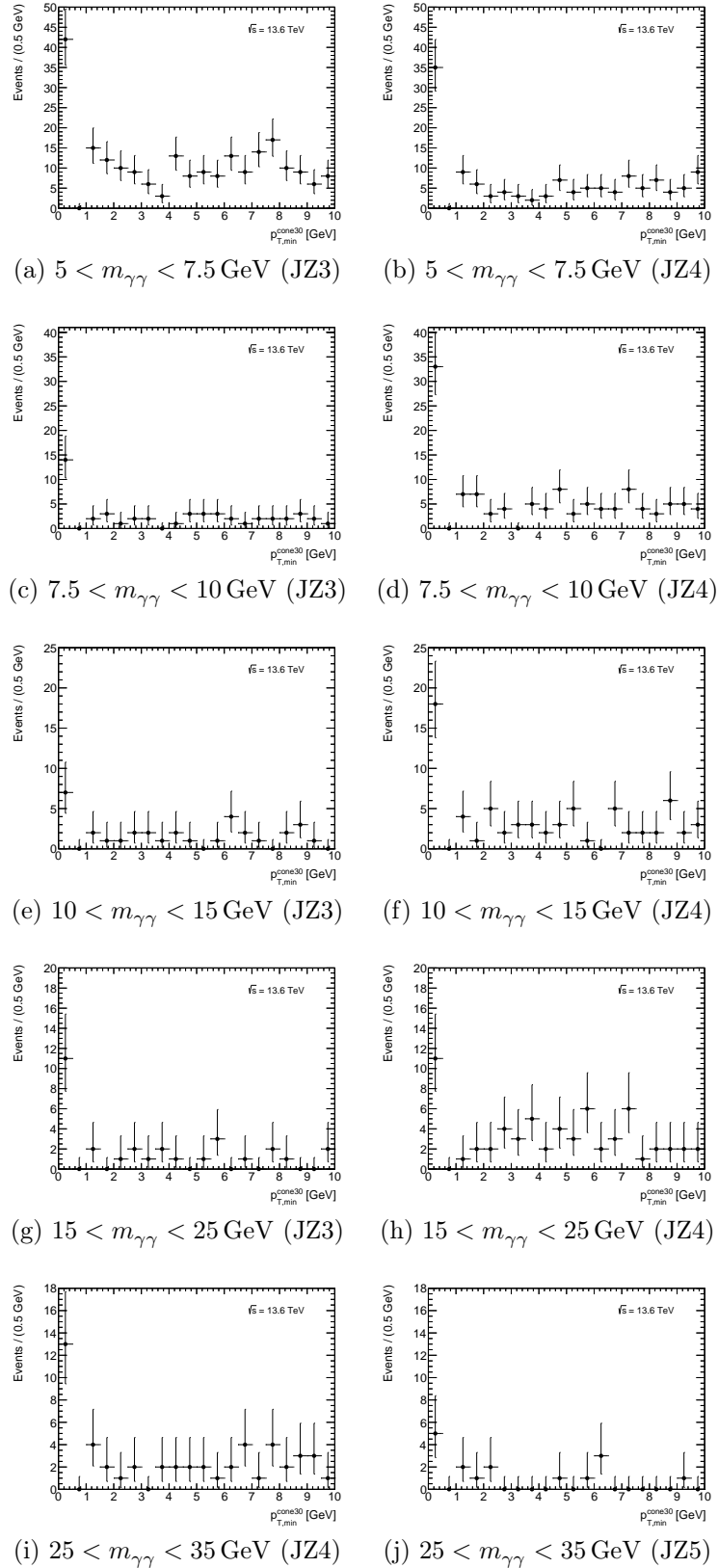


Figure 7.3: $p_{T,\min}^{\text{cone30}}$ distributions by mass bin in the JZ samples (fake diphotons), shown in separate slices. JZ3, JZ4 and JZ5 correspond to different simulated dijet samples, each covering a distinct range of jet transverse momentum p_T^{jet} and filtered by parton-level momentum p_T^{hat} (see Section 6.2.2). Only slices with a non-negligible number of events are shown. For JZ3, $p_T^{\text{hat}} > 50$ GeV and $160 < p_T^{\text{jet}} < 400$ GeV; for JZ4, $p_T^{\text{hat}} > 150$ GeV and $400 < p_T^{\text{jet}} < 800$ GeV; for JZ5, $p_T^{\text{hat}} > 350$ GeV and $800 < p_T^{\text{jet}} < 1300$ GeV.

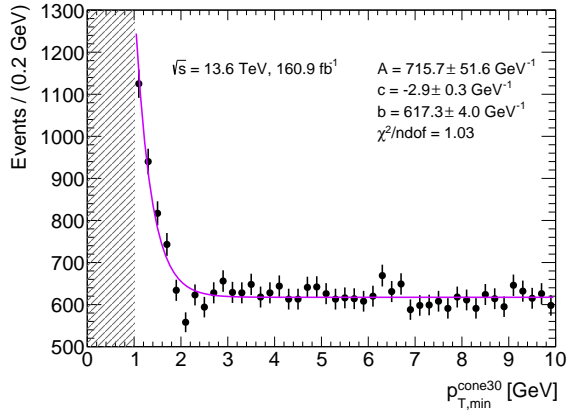
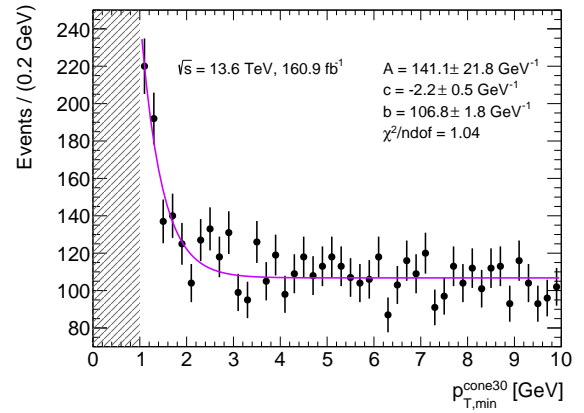
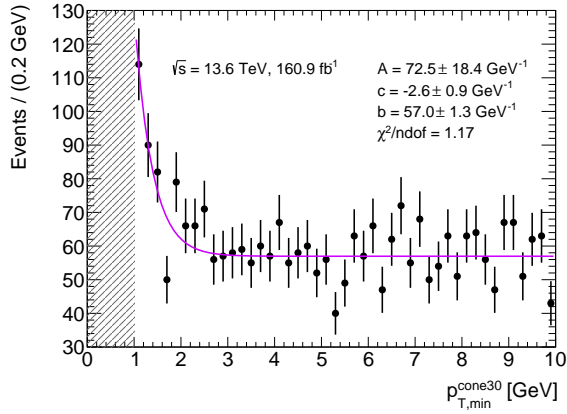
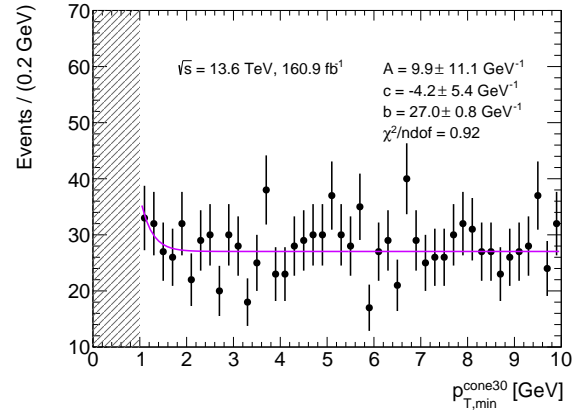
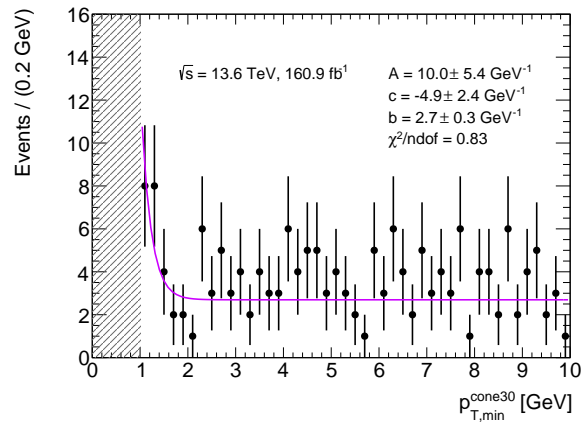
(a) $5 < m_{\gamma\gamma} < 7.5$ GeV(b) $7.5 < m_{\gamma\gamma} < 10$ GeV(c) $10 < m_{\gamma\gamma} < 15$ GeV(d) $15 < m_{\gamma\gamma} < 25$ GeV(e) $25 < m_{\gamma\gamma} < 35$ GeV

Figure 7.4: Fit to $p_{T,min}^{cone30}$ on data, with an exponential function and a constant: $f(p_{T,min}^{cone30}) = A \exp[c(p_{T,min}^{cone30} - 1 \text{ GeV})] + b$. The dashed band hides the isolated events in the signal region.

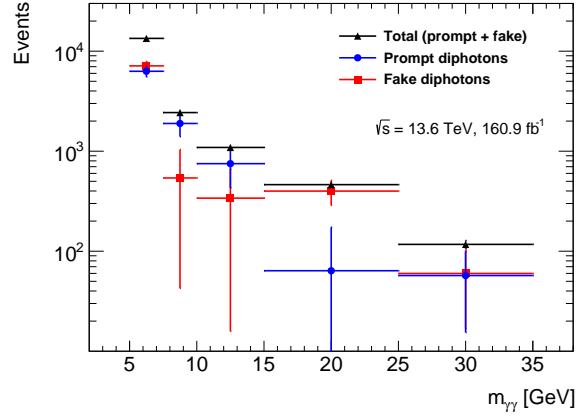


Figure 7.5: Estimated prompt and fake diphoton yields in the signal region, for an integrated luminosity of 160.9 fb^{-1} . The distributions are shown unstacked, and the total distribution corresponds to the sum of the prompt and fake diphoton contributions.

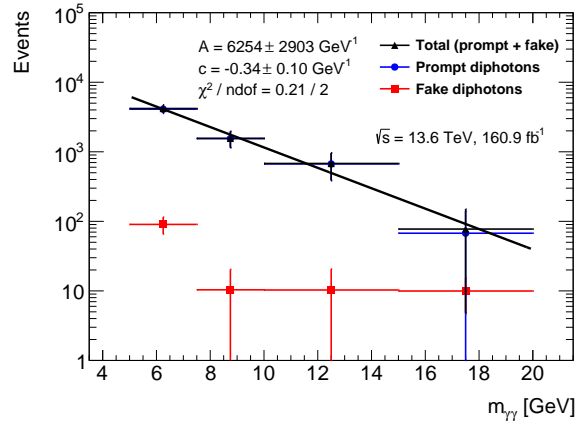


Figure 7.6: Estimated prompt and fake diphoton yields in the signal region, after applying diphoton identification and isolation efficiencies, for an integrated luminosity of 160.9 fb^{-1} . The fit parameters and the χ^2 divided by the number of degrees of freedom are indicated.

7.2 Mass resolution

The mass resolution is modelled using the PG simulated prompt diphoton sample. The intrinsic resonance width is assumed to be negligible compared to the detector resolution, the resonance shape therefore being determined by the detector's mass resolution.

In the PG simulated sample, the diphoton primary vertex cannot be determined for two reasons. First, there are no other tracks or jets coming from the process producing the ALP. Moreover, the two photons, coming from the ALP, are closely collimated, having an unfavourable lever arm, so that pointing does not help. Consequently, the reconstructed η does not take into account the z position of the primary vertex. To correctly estimate the mass resolution, it is assumed that a primary vertex is properly reconstructed: the true photon η and ϕ , along with the reconstructed energy, are used to compute the reconstructed diphoton mass. The mass resolution therefore accounts for the energy resolution only.

Another difficulty with this sample arises from the fact that the p_T distributions of the two photons are not realistic, especially that of the subleading one, as shown in Figure 7.7. There are two effects: the PG is too soft at very low p_T , and too hard at high p_T . To obtain a realistic mass resolution, events are reweighted so that the subleading p_T spectrum resembles that observed for track-isolated events in data.

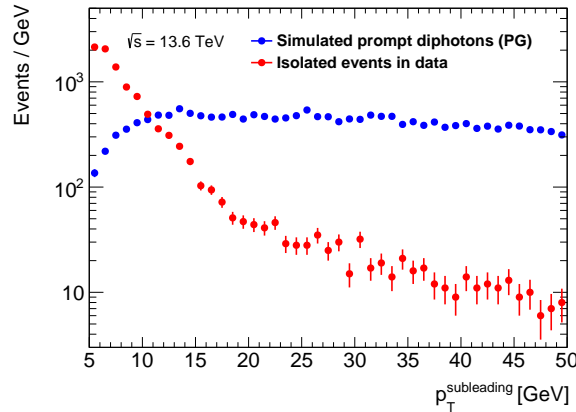


Figure 7.7: p_T of the subleading photon in the PG-simulated prompt diphoton sample and in isolated data. Isolated events in data refer to those in the signal region (with $p_{T,\min}^{\text{cone30}} = 0$) and not to diphoton isolation (which has not been applied here).

Finally, the mass resolution, $m_{\gamma\gamma}^{\text{reco}} - m_{\gamma\gamma}^{\text{truth}}$, has been estimated in bins as narrow as permitted by the available statistics. It was fitted with a Gaussian in a range centred on the peak position μ_p : $[\mu_p - 1.5\sigma, \mu_p + 1.5\sigma]$. Several fits are shown in Figure 7.8.

The width σ is shown as a function of the mass in Figure 7.9. It has been parametrised with a fourth-degree polynomial of the form:

$$\sigma(m_{\gamma\gamma}^{\text{truth}}) = \sigma_0 + p_2 \left(\frac{m_{\gamma\gamma}^{\text{truth}} - 5 \text{ GeV}}{25 \text{ GeV}} \right)^2 + p_4 \left(\frac{m_{\gamma\gamma}^{\text{truth}} - 5 \text{ GeV}}{25 \text{ GeV}} \right)^4, \quad (7.6)$$

where σ_0 , p_2 , and p_4 are fit parameters.

The behaviour of the resolution as a function of mass, which is almost constant below 20 GeV, differs from that observed in other diphoton analyses. This reflects the specifics

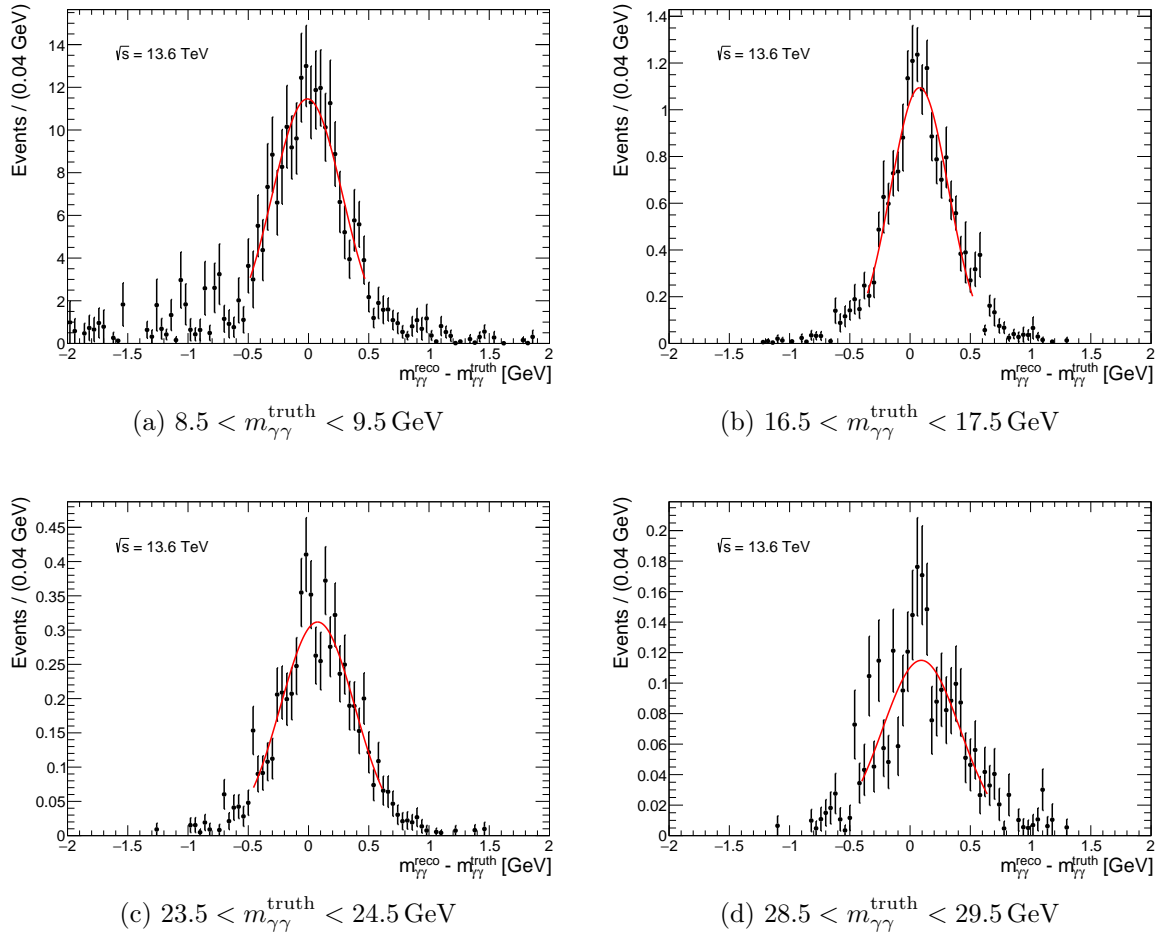


Figure 7.8: Fits to the mass resolution using a Gaussian, over a range centred on the peak position: $[\mu_p - 1.5\sigma, \mu_p + 1.5\sigma]$.

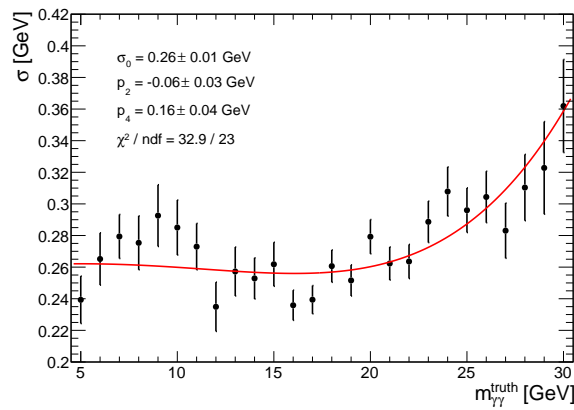


Figure 7.9: Fit to the mass resolution as a function of the mass. The fit parameters and the χ^2 divided by the number of degrees of freedom are indicated.

of the selection used in this study, with a hard leading photon and a very soft subleading one.

7.3 Effective background and expected limit

The effective background is defined as the background under the resonance peak. It can be calculated using the background spectrum determined in Section 7.1 and the mass resolution computed in Section 7.2.

Background toy datasets are generated following the background estimation performed in Section 7.1. For different hypothetical resonance masses m_X , the background datasets are fitted with a combined model consisting of a signal, given by the resolution computed in Section 7.2, and a background component:

$$\text{model}(m_{\gamma\gamma}; m_X) = N_S \times \text{PDF}_S(m_{\gamma\gamma}; m_X) + N_B \times \text{PDF}_B(m_{\gamma\gamma}), \quad (7.7)$$

where $\text{model}(m_{\gamma\gamma}; m_X)$ is the total model for a resonance mass m_X , $\text{PDF}_S(m_{\gamma\gamma}; m_X)$ the signal PDF for a resonance mass m_X (given by the Gaussian calorimeter resolution computed in Section 7.2), $\text{PDF}_B(m_{\gamma\gamma})$ the background PDF, and N_S and N_B the numbers of signal and background events. Several fits are shown in Figure 7.10.

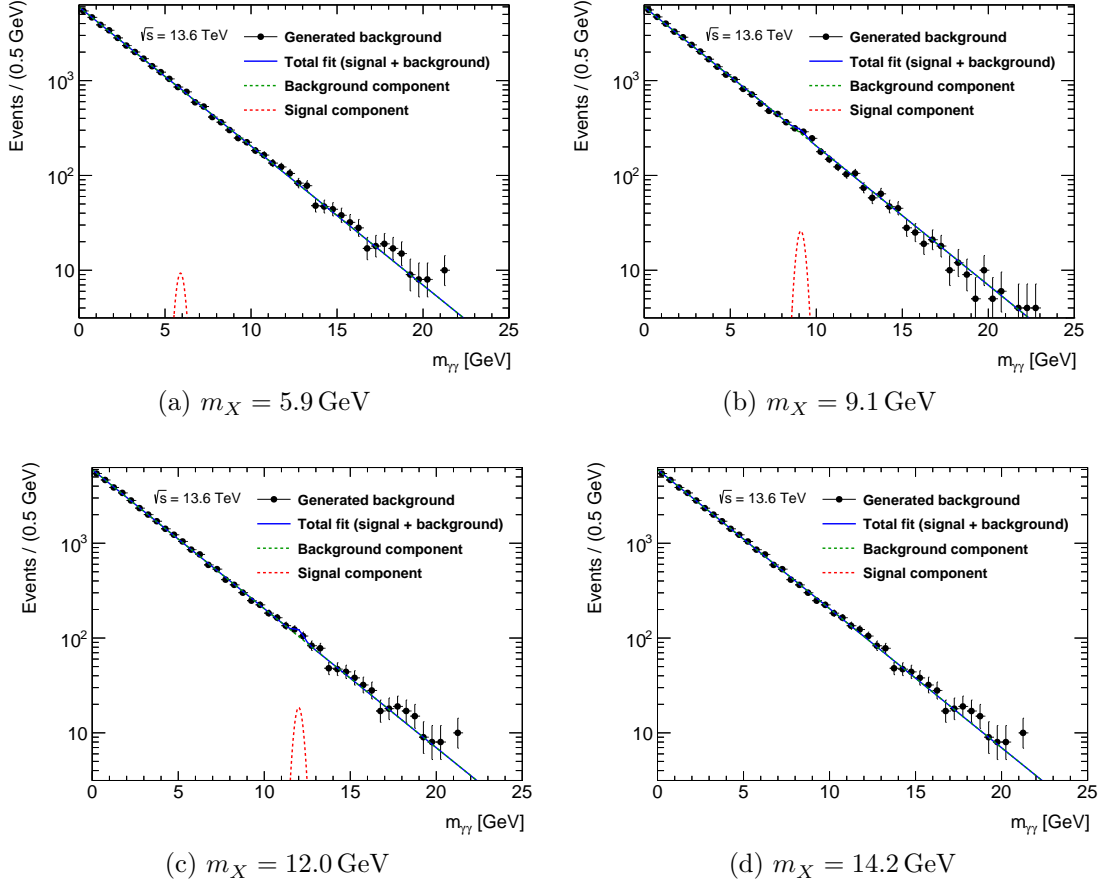


Figure 7.10: Background toy datasets, along with their fits. The signal shape is fixed to the resonance mass. The shape and number of background events are fixed to the estimation using the control region.

For a given resonance mass m_X , the shapes of the signal and background PDFs are kept fixed, while N_S and N_B are treated as free parameters. Since no signal events have been generated, the error on the fit parameter N_S is given by:

$$\sigma(N_S) = \sqrt{N_B^{\text{eff}}}, \quad (7.8)$$

where N_B^{eff} is the effective background under the resonance peak.

1000 background toy datasets were generated and fitted per resonance mass, for 150 mass points uniformly distributed between 5 and 20 GeV. For each mass point, the mean effective background is computed from the 1000 fits.

This effective background can be described as an exponential. It is shown in Figure 7.11. There is no sensitivity above 20 GeV.

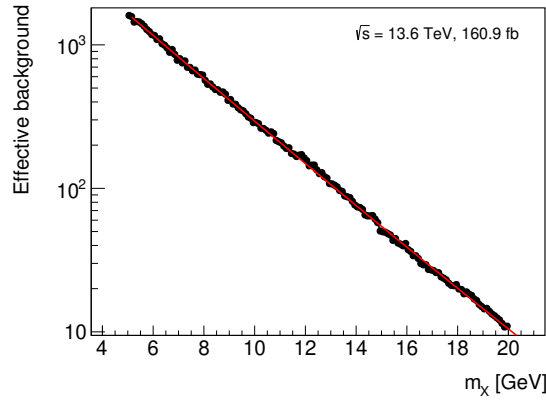


Figure 7.11: Effective background as a function of the resonance mass for an integrated luminosity of 160.9 fb^{-1} . At this stage, one input is still missing to allow for an exclusion calculation in terms of the KSVZ prediction. This parameter is essential to interpret the results within that theoretical framework fully. It is expected that this missing information will become available soon, allowing the analysis to be completed accordingly.

Given the effective background as a function of the mass, a limit can be established on the total cross section of the $pp \rightarrow X \rightarrow \gamma\gamma$ process, defined as:

$$\sigma_X \cdot \mathcal{B}(X \rightarrow \gamma\gamma) = \frac{N_{\text{sig}}^{\text{reco}}}{C_X \cdot A_X \cdot \mathcal{L}_{\text{int}}}, \quad (7.9)$$

where \mathcal{B} is the branching ratio, $N_{\text{sig}}^{\text{reco}}$ the measured signal yield, \mathcal{L}_{int} the integrated luminosity, and C_X and A_X the factors introduced in Figure 6.1. A_X represents the acceptance of the fiducial volume. It corresponds to the selections given in Table 7.2, at truth level.

<i>Object</i>	<i>Fiducial definition</i>
Photons	$ \eta^\gamma < 1.37$
Diphotons	$\Delta R_{\gamma\gamma} < 0.2$
	$p_{\text{T}}^{\gamma\gamma} > 140 \text{ GeV}$

Table 7.2: Particle-level selections defining the fiducial volume.

The C_X correction factor accounts for detector effects on events within the fiducial volume, as well as for diphoton identification and isolation efficiencies obtained from the trained neural networks. The selections are summarised in Table 7.3.

<i>Object</i>	<i>Selections</i>
Photons	Loose $p_T^\gamma > 5 \text{ GeV}$ Unconverted
Diphotons	HLT_g140_loose $\left(\frac{\Delta\eta_{\gamma\gamma}}{0.1}\right)^2 + \left(\frac{\Delta\phi_{\gamma\gamma}}{0.15}\right)^2 > 1$ Diphoton identification Diphoton isolation

Table 7.3: Reconstruction-level selections used to compute C_X .

The acceptance is strongly model-dependent. A fiducial cross section can also be defined:

$$\sigma_{X,\text{fid}} \cdot \mathcal{B}(X \rightarrow \gamma\gamma) = \frac{N_{\text{sig}}^{\text{reco}}}{C_X \cdot \mathcal{L}_{\text{int}}}, \quad (7.10)$$

where A_X does not appear.

The efficiencies of each selection listed in Table 7.3 are shown in Figure 7.12, along with the total efficiency, i.e., the C_X factor. The latter is modelled with the following function:

$$C_X(m_X) = \frac{a_0 - a_1 \times \exp(-m_X/a_2)}{1 + a_3 \times \exp(-m_X/a_4)}, \quad (7.11)$$

where a_0 , a_1 , a_2 , a_3 and a_4 are fit parameters.

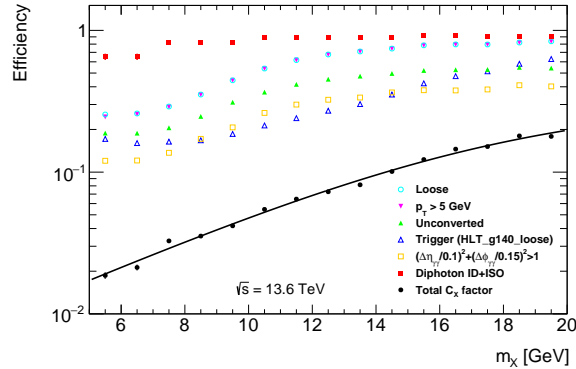


Figure 7.12: Efficiencies of the different selections, considered independently, and the total efficiency (C_X factor) with its fit, computed using the PG simulated prompt-diphoton sample.

The expected upper limit at 95 % confidence level on the number of signal events is computed from signal-plus-background fits to background-only Asimov datasets, for masses between 5 and 20 GeV, using asymptotic formulas (see Appendix A). It is then converted to a limit on the fiducial cross section of a hypothetical resonance times its branching ratio, $\sigma_{\text{fid}} \times \text{BR}$, as shown in Figure 7.13.

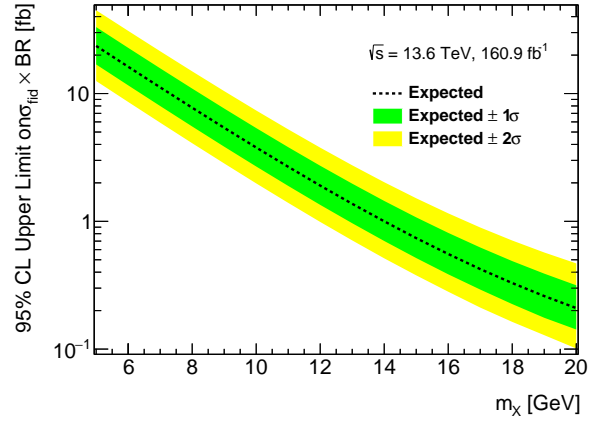


Figure 7.13: Expected upper limit at 95 % confidence level on the fiducial cross section times the branching ratio of a hypothetical diphoton resonance, for an integrated luminosity of 160.9 fb^{-1} . The expected upper limit is shown as a dashed black line, with its ± 1 and ± 2 standard deviation bands.

Chapter 8

Carbon footprint of research

Contents

8.1	Global situation	176
8.1.1	Increase in greenhouse gas emissions	176
8.1.2	Global warming consequences	176
8.1.3	Scenarios for the future	179
8.2	Greenhouse gas emissions from research	179
8.2.1	French public research	179
8.2.2	CERN emissions	183
8.3	LPNHE's emissions	185
8.3.1	Laboratory presentation	185
8.3.2	GHG emissions assessment	187
8.3.3	Reduction measures	190

Climate change is, by now, scientifically well established [165]. Its consequences for agriculture, health, and living conditions are already noticeable and will become more severe if no action is taken to slow global warming in the coming years. The research sector, particularly international collaborations involving large infrastructures such as CERN, contributes to greenhouse gas (GHG) emissions. This chapter provides an overview of the current global situation, with a focus on the French research sector. It then presents the work carried out at the LPNHE to measure and reduce the laboratory’s GHG emissions, in which I took part.

8.1 Global situation

8.1.1 Increase in greenhouse gas emissions

The Intergovernmental Panel on Climate Change (IPCC) is an intergovernmental organisation established in 1988, tasked with assessing the extent, causes, and consequences of global warming [166]. It regularly reviews the scientific literature on climate change and provides summaries for policymakers¹. Its conclusions are unequivocal: between 2011 and 2020, the global surface temperature was 1.1 °C higher than during the 1850–1900 period, as shown in Figure 8.1. Global warming is a result of human activities and is primarily driven by emissions of carbon dioxide (CO₂) and methane (CH₄). Their concentrations have increased by 47 % and 156 %, respectively, since 1750—far exceeding the natural multi-millennial fluctuations observed between glacial and interglacial periods over at least the past 800,000 years. Warming from GHGs is partially offset by cooling due to aerosols. Figure 8.1 shows the evolution of the concentrations of the three main GHGs since 1850. GHG emissions from human activities continue to rise (see Figure 8.1).

The consequences of global warming are already visible and affect our lives in numerous ways: food and water security, health, economies, and associated losses and damages to both nature and people. Moreover, communities that have historically contributed the least to climate change are often the most vulnerable to its impacts.

Despite their profound impact on our lives, GHG emissions continue to increase. The average annual emissions during 2010–2019 were higher than in any previous decade. However, the rate of growth has slowed: 1.3 % per year during 2010–2019, compared to 2.1 % per year between 2000 and 2009.

Climate change is the direct result of over a century of GHG emissions from energy use, land use, lifestyles, and patterns of consumption and production. Figure 8.2 presents the assessment of observed changes in large-scale indicators of mean climate across various climate system components and their attribution to human influence.

8.1.2 Global warming consequences

Global warming is already affecting our lives in many ways.

First, irreversible damage has been caused to terrestrial and ocean ecosystems. Extreme atmospheric phenomena, such as heatwaves, heavy precipitation, droughts, and tropical cyclones, have become more intense and more frequent. This has direct consequences for food and water security, putting entire populations at risk.

¹This section is mainly based on Ref. [165]

Human activities are responsible for global warming

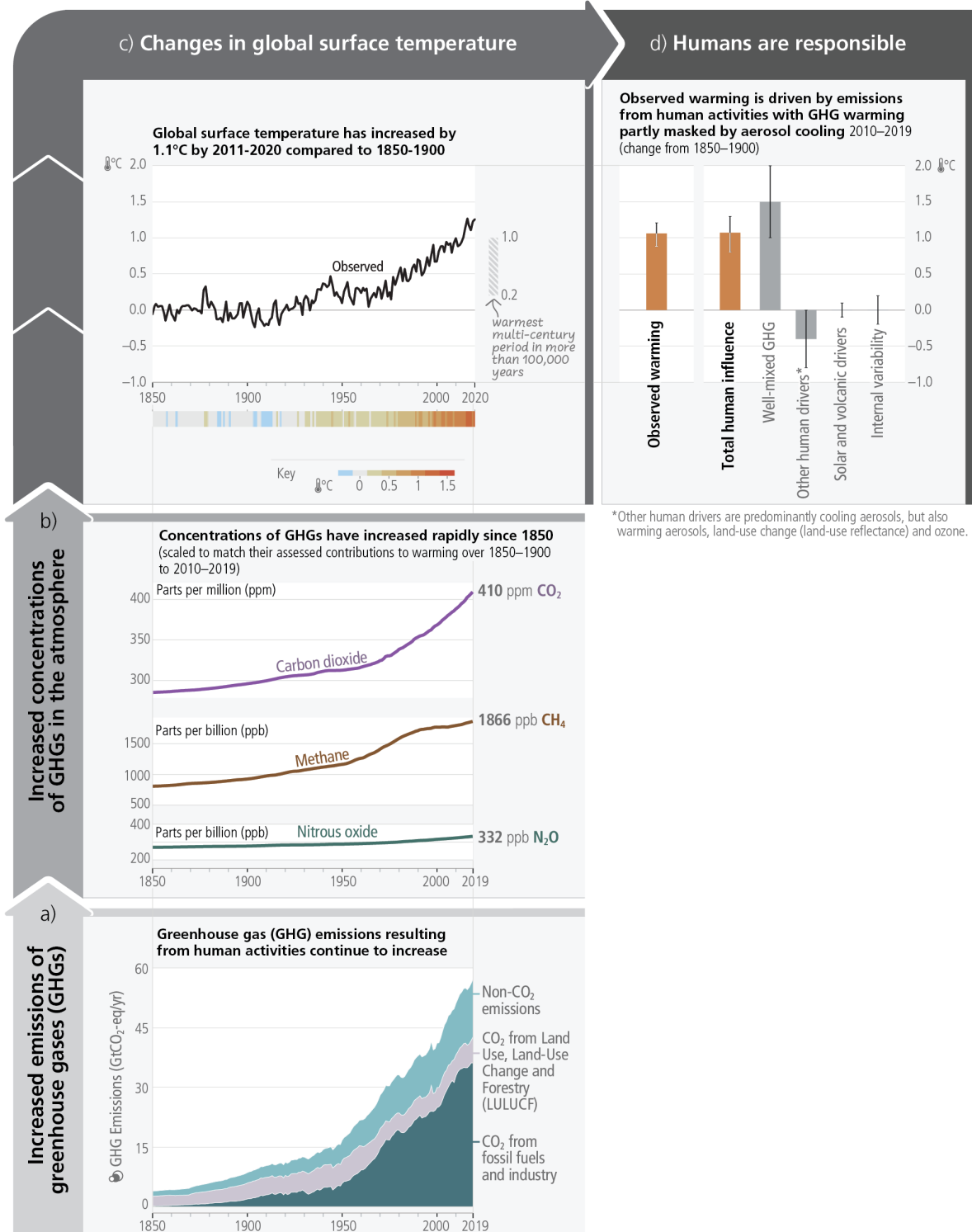


Figure 8.1: The causal chain from emissions to the resulting warming of the climate system. **a)** Global net anthropogenic GHG emissions. **b)** Concentrations of GHGs in the atmosphere. **c)** Changes in global surface temperature, shown as annual anomalies from a 1850–1900 baseline. **d)** Origins of temperature change. Figure from Ref. [165].

In urban settings, human health, livelihoods, and key infrastructure are affected by

Change in indicator		Observed change assessment	Human contribution assessment
Atmosphere and water cycle	Warming of global mean surface air temperature since 1850-1900		<i>likely</i> range of human contribution ((0.8-1.3 °C) encompasses the very likely range of observed warming ((0.9-1.2 °C))
	Warming of the troposphere since 1979		Main driver
	Cooling of the lower stratosphere since the mid-20th century		Main driver 1979 - mid-1990s
	Large-scale precipitation and upper troposphere humidity changes since 1979		
	Expansion of the zonal mean Hadley Circulation since the 1980s		Southern Hemisphere
Ocean	Ocean heat content increase since the 1970s		Main driver
	Salinity changes since the mid-20th century		
	Global mean sea level rise since 1970		Main driver
Cryosphere	Arctic sea ice loss since 1979		Main driver
	Reduction in Northern Hemisphere springtime snow cover since 1950		
	Greenland ice sheet mass loss since 1990s		
	Antarctic ice sheet mass loss since 1990s		<i>Limited evidence & medium agreement</i>
	Retreat of glaciers		Main driver
Carbon cycle	Increased amplitude of the seasonal cycle of atmospheric CO ₂ since the early 1960s		Main driver
	Acidification of the global surface ocean		Main driver
Land climate	Mean surface air temperature over land (about 40% larger than global mean warming)		Main driver
Synthesis	Warming of the global climate system since preindustrial times		

Key

medium confidence	likely / high confidence	very likely	extremely likely	virtually certain	fact

Figure 8.2: Assessment of observed changes in large-scale indicators of mean climate across climate system components, and their attribution to human influence. The colour indicates the assessed confidence in the observed change and the human contribution as a driver or main driver. Figure from Ref. [165].

heatwaves, peaks in air pollution, and the reduced functioning of essential services. Urban infrastructure, including transportation, water, sanitation, and energy systems, can be compromised.

Physical and mental health are also impacted. Extreme heat events contribute to increased mortality and morbidity. Animal and human diseases, including zoonoses, are appearing in new regions. Mental health issues are associated with rising temperatures, as well as trauma resulting from extreme events. Humanitarian crises involving population displacement have also become more frequent.

Compound extreme events are expected to increase with climate change, including combinations such as heatwaves and droughts, fire weather, and flooding. These events pose risks to health, ecosystems, infrastructure, livelihoods, and food systems.

Climate change also affects the economy. Sectors directly exposed to climate risks—such as agriculture, forestry, energy, and tourism—are especially vulnerable. Individual livelihoods are impacted through decreased agricultural productivity, poorer human health and

food security, destruction of homes and infrastructure, and loss of property and income. Extreme weather events, such as cyclones, have short-term negative effects on economic growth. Other phenomena, like heavy precipitation and wildfires, disrupt economic activity and public health.

Through damage to nature—unevenly distributed across the globe—climate change also leads to cultural loss, affecting cultural practices, sense of belonging, identity, and connection to place. For example, in many Arctic regions, changes in snow cover, lake and river ice, and permafrost are damaging the livelihoods and cultural identity of Arctic residents.

The most vulnerable people are generally the most affected by climate change. Regions facing development constraints are more exposed to climatic hazards. Vulnerability is also greater in areas affected by poverty, governance challenges, limited access to basic services and resources, and violent conflicts.

8.1.3 Scenarios for the future

The actions already undertaken to mitigate climate change are not sufficient. There is a gap between global ambitions and the sum of national commitments, and another between national ambitions and their actual implementation. GHG emissions projected according to Nationally Determined Contributions (NDCs), required under the Paris Agreement, are likely to lead to a warming that exceeds $+1.5^{\circ}\text{C}$ and will make it more difficult to limit warming to below $+2^{\circ}\text{C}$. Current initiatives tend to prioritise short-term risks and do not sufficiently address the structural changes needed to reduce emissions. Systemic barriers, such as insufficient funding and limited knowledge, hinder the implementation of climate actions. Figure 8.3 shows global GHG emissions under modelled pathways and projected outcomes from near-term policy assessments for 2030.

8.2 Greenhouse gas emissions from research

8.2.1 French public research

The research sector is not exempt from GHG emissions. The research group (*groupement de recherche*) Labos 1point5 [167] was created in 2019 to measure the carbon footprint of French research activities and to propose ways to reduce GHG emissions. It is organised around two main axes: an analysis axis, which measures and characterises the carbon footprint of French public research; and a transition axis, which supports laboratories and provides them with tools to reduce their GHG emissions.

The group has developed a standardised tool for measuring the carbon footprint of research that can be applied to any department, laboratory, or team worldwide [168]. It takes into account the main sources of emissions, such as those related to buildings, purchases, and travel. Based on preliminary data from over a hundred laboratories in France, they found that average emissions are $479\text{ t CO}_2\text{e}$ per research laboratory and $3.6\text{ t CO}_2\text{e}$ per average laboratory member for their research activities [168]. The carbon dioxide equivalent (CO_2e) of a given quantity of gas is the amount of CO_2 that would cause the same warming effect on Earth as that quantity of gas, based on its global warming potential (GWP).

According to the internationally recognised methodology of the Greenhouse Gas Protocol [169], GHG emissions are categorised into three groups, known as scopes:

Projected global GHG emissions from NDCs announced prior to COP26 would make it *likely* that warming will exceed 1.5°C and also make it harder after 2030 to limit warming to below 2°C

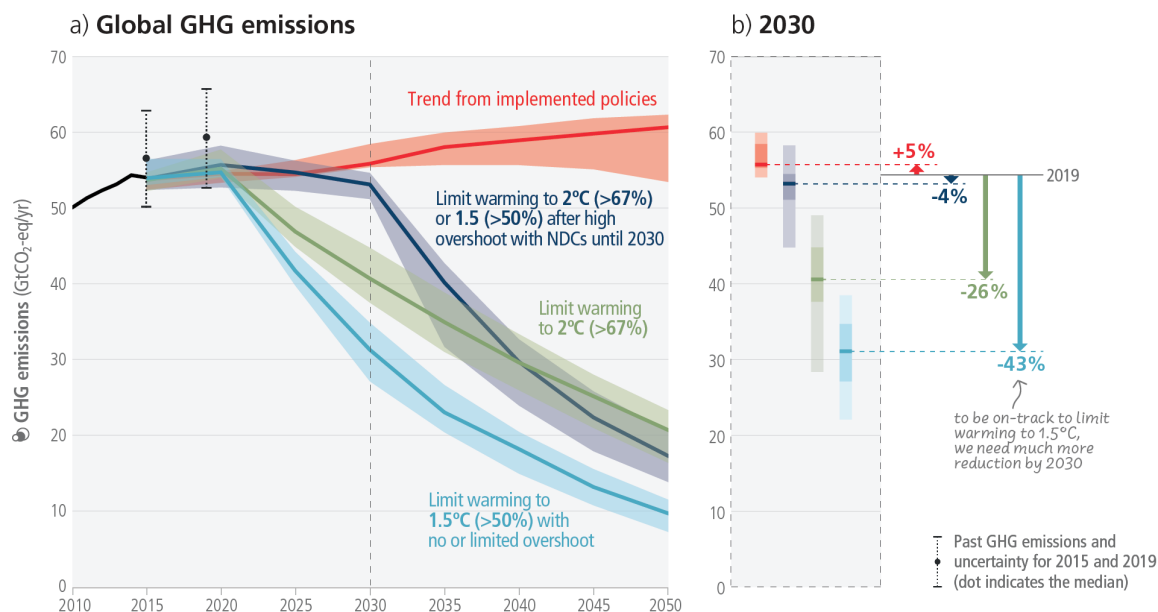


Figure 8.3: Global GHG emissions under modelled pathways and projected outcomes from near-term policy assessments for 2030. **a)** Global GHG emissions over 2015–2050 for four types of assessed modelled global pathways. **b)** Snapshot of the GHG emission ranges of the modelled pathways in 2030 and projected emissions outcomes from near-term policy assessments in 2030. Figure from Ref. [165].

- **Scope 1** includes the organisation’s direct emissions, such as those from facilities and vehicles.
- **Scope 2** covers the organisation’s indirect emissions from electricity consumption, heating, or cooling purchased for its own use.
- **Scope 3** includes all other indirect emissions, such as business travel, commuting, catering, and procurement.

Scope 3 emissions are the most difficult to measure, although they represent the largest share of laboratory emissions, sometimes up to 80 % of total emissions [170]. In particular, purchases dominate emissions, as shown in Figure 8.4. These emissions can be estimated from monetary flows. Labos 1point5 used NACRES codes (Nomenclature Achats Recherche Enseignement Supérieur), the French system for classifying purchases in the research and higher education sector, and assigned monetary emission factors to each category, in kgCO₂e/€.

Data from 700 laboratories in the Labos 1point5 network, including over 20 000 staff members from all disciplines, showed that purchases represent more than 50 % of emissions, with a median of 2.7 t CO₂e/person [170]. This is three to four times higher than the separate contributions from travel, commuting, and heating. Furthermore, the results vary significantly between laboratories and are linearly correlated with budget, with an average carbon intensity of 0.31 ± 0.07 kgCO₂e/€. Large disparities are observed between research domains (see Figure 8.5).

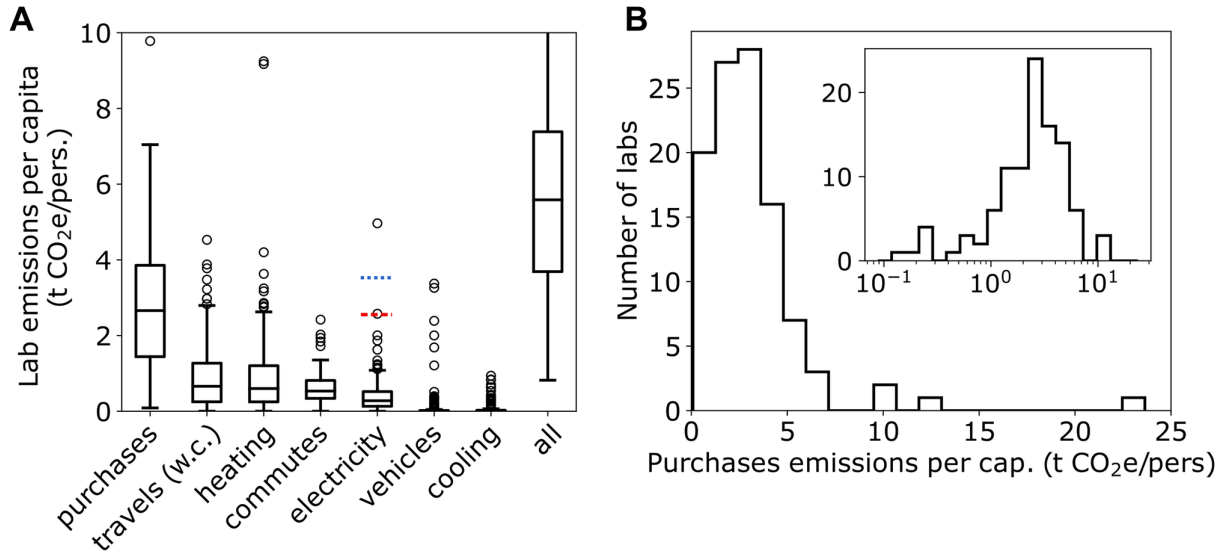


Figure 8.4: **A)** Box plot of laboratory emissions per capita per emission source. The box extends from the 1st to 3rd quartile (Q1, Q3), with a line at the median, the whiskers extend to $1.5 \times (Q3 - Q1)$, and the circles represent all data outside the whiskers. Electricity emissions are calculated for different energy mix: French mix (black box plot), world mix (median as dashed red line), high-carbon mix (median as dotted blue line). **B)** Distribution of purchases emissions per capita. The inset shows the same data in log scale. Figure from Ref. [170].

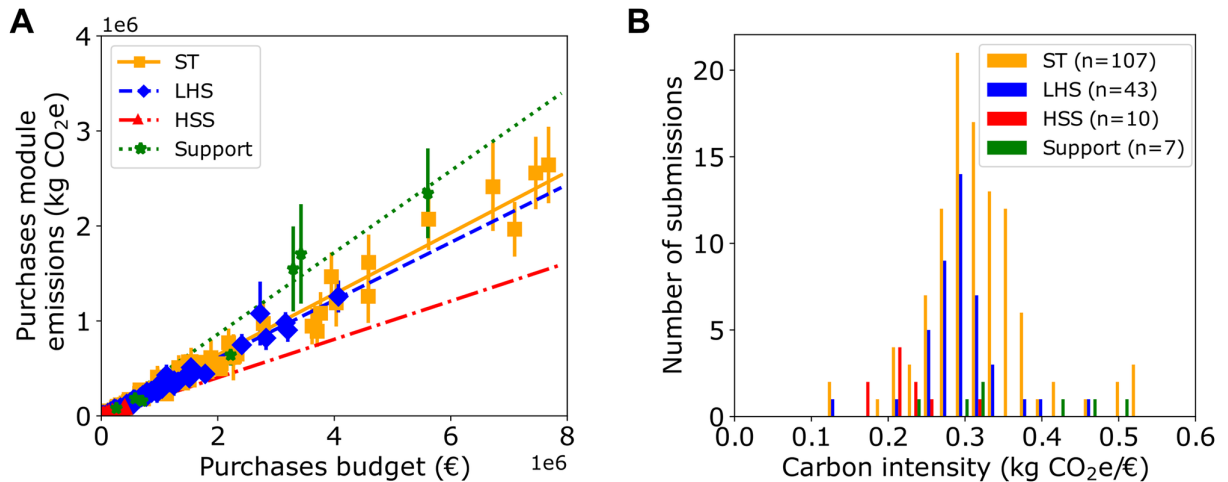


Figure 8.5: **A)** Purchases module emissions vs. budget for all GHG laboratory footprints in the GES 1point5 lab emissions database. Error bars correspond to one standard deviation. **B)** Histogram of purchases module carbon intensities for different scientific domains. HSS: Human and Social Sciences, LHS: Life and Health Sciences, ST: Science and Technology. Support laboratories: large experimental platforms providing analysis services. Figure from Ref. [170].

Travel also contributes significantly to the laboratory's GHG emissions [171]. Air travel is by far the main contributor, as shown in Figure 8.6, representing more than 96 % of research travel GHG emissions. Among these, intercontinental flights are less frequent (less than 10 % of all plane travel) but dominate travel-related GHG emissions, accounting for 64 % of the total. Domestic and continental flights are much more common,

but their mitigation potential is limited: trips under 1000 km account for only 15 % of total emissions.

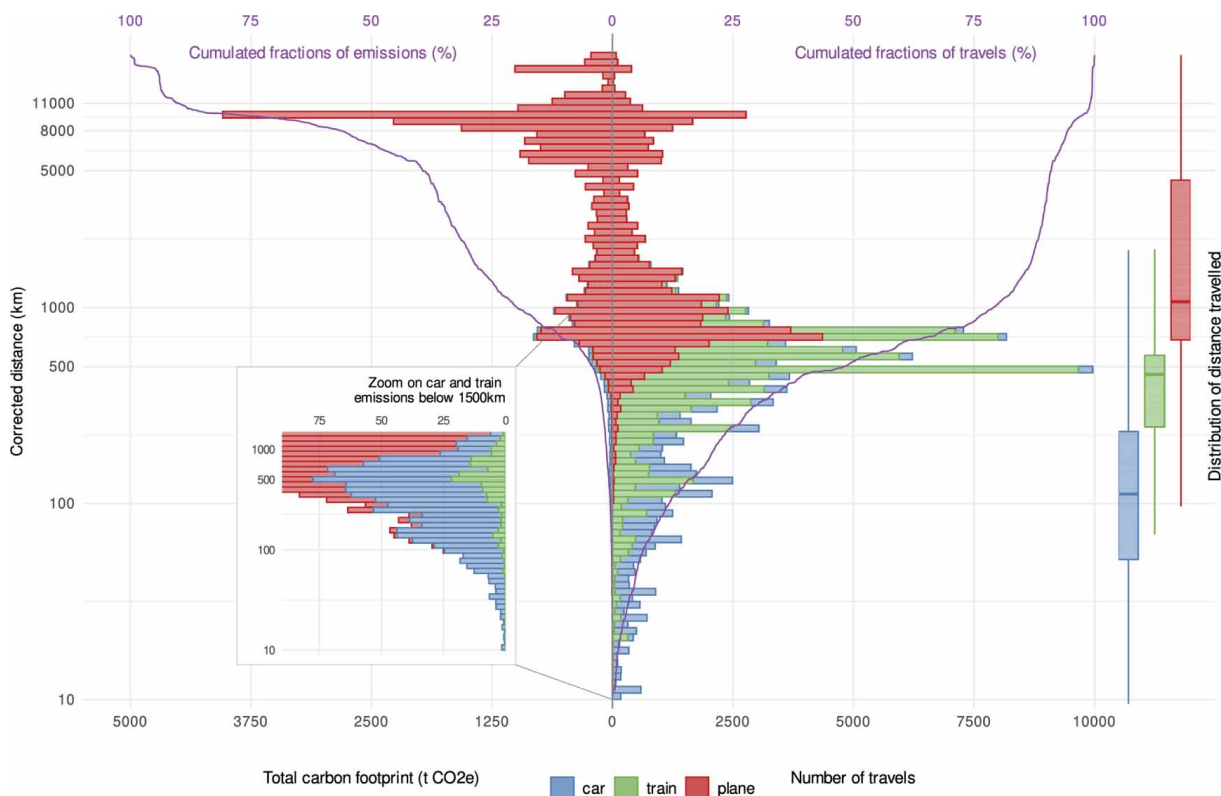


Figure 8.6: Histogram of travelled distance (right) and associated GHG emissions (left) for car (blue), train (green), and plane (red) travel, based on 137 081 academic trips from 159 research units in France in 2019 (Labos 1point5 database). Box plots extend from the 1st to the 3rd quartile (Q1 to Q3), and whiskers to $1.5 \times (Q3 - Q1)$. The zoom focuses on car and train travel for distances below 1500 km. Bold purple lines show the cumulative percentages of the number of trips (right) and GHG emissions (left). Figure from Ref. [172].

It is important to note that air travel concerns only a minority of the global population. In 2018, only 11 % of the world’s population took a flight, and only 4 % took an international one. This contrasts with results from a survey of over 6000 respondents representative of French research staff, showing that 58 % of PhD-level respondents (including PhD students) travelled by plane for professional reasons in 2019. Figure 8.7 shows the disparities in travelled distances by plane between research disciplines. Large disparities are observed, with a factor greater than 3 between health and medical research and astronomy. Figure 8.8 presents the average distance travelled by plane in 2019 by respondent status. Large disparities are also observed, with senior researchers taking almost four times as many flights as fully funded PhD students.

Finally, other sources also contribute significantly to laboratory carbon footprints, such as commuting, use of office space, catering, and heating and electricity consumption. In some disciplines, scientific equipment represents the largest share of emissions—for instance, in astronomy, due to the use of space-based facilities (such as telescopes) and energy-intensive supercomputers.

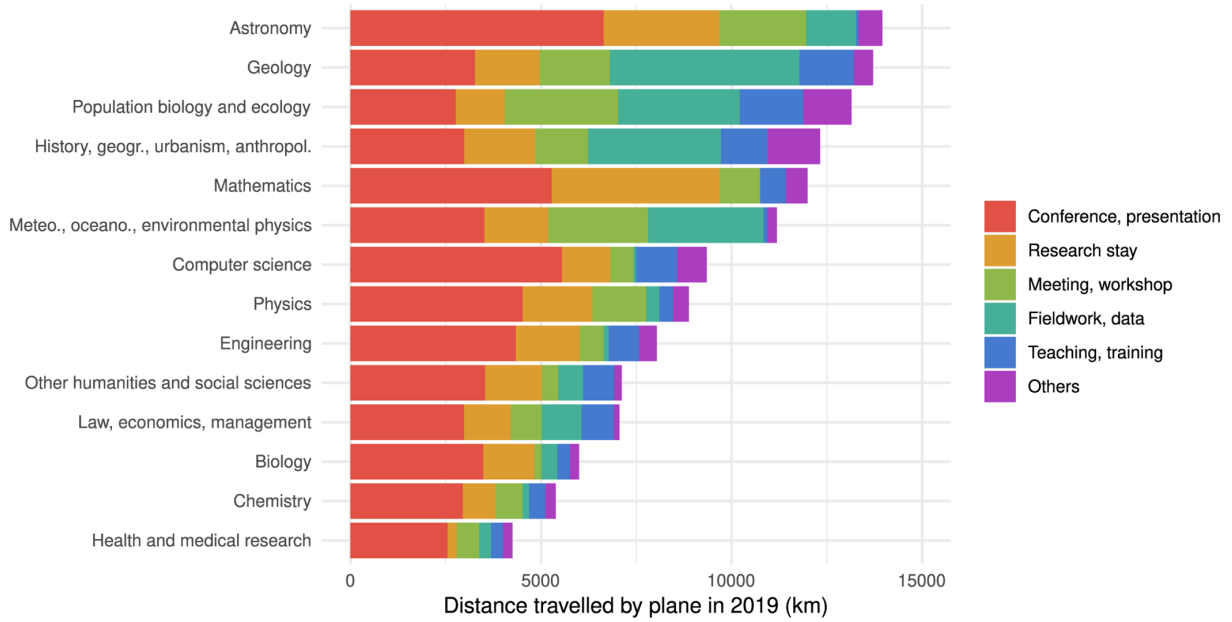


Figure 8.7: Average distance travelled by plane in 2019 by discipline of respondents. Figure from Ref. [168].

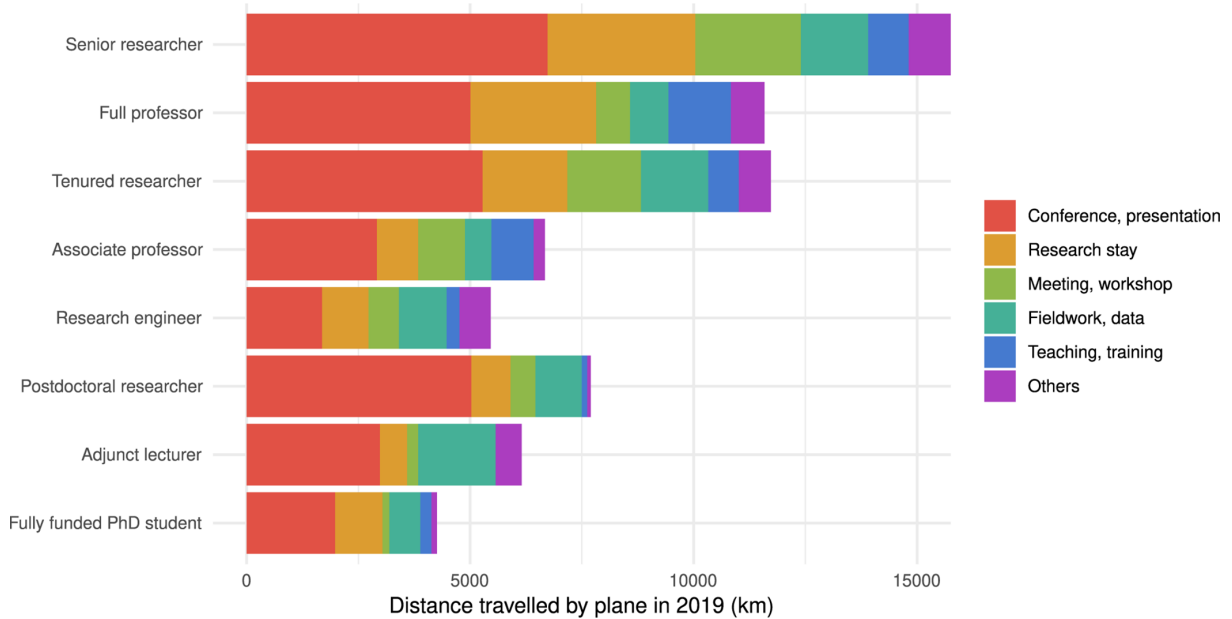


Figure 8.8: Average distance travelled by plane in 2019 by status of respondents. Figure from Ref. [168].

8.2.2 CERN emissions

The work presented in this thesis is part of the CERN organisation, whose research activities are highly energy-intensive and GHG-emitting. CERN publishes regular environmental reports, which include its GHG emissions. The latest report covers the years 2021 and 2022 [173]. Following the internationally recognised Greenhouse Gas Protocol methodology, CERN evaluates its emissions across the three scopes introduced above. For CERN, scope 1 emissions dominate, followed by scopes 3 and 2 (see Figure 8.9).

Scope 1 covers the Laboratory's industrial infrastructure and on-site activities. About

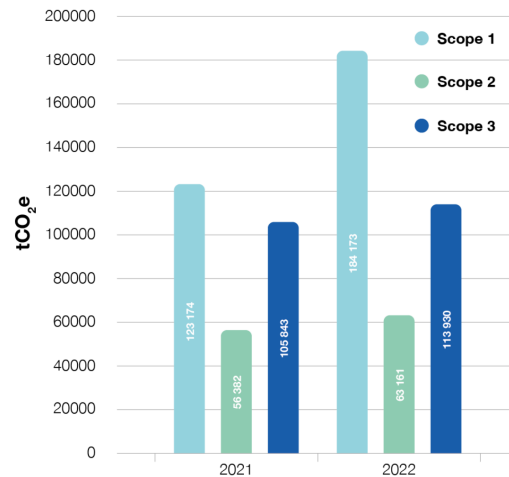


Figure 8.9: CERN’s total scope 1, 2 and 3 2021-2022 emissions. Figure from Ref. [173].

90 % of these emissions originate from experiments. This is mainly due to the gases used for particle detection and detector cooling, which leak into the atmosphere due to the extreme thinness of certain detector components. The main gases used are fluorinated gases, which have a high global warming potential: FCs, PFCs, and SF₆ for particle detection; PFCs and HFCs for detector cooling; and HFOs/HFCs for standard air conditioning systems. SF₆ is also used for electrical insulation in power supply systems. These gases have a very high global warming potential. CERN’s total scope 1 emissions amounted to 123 174 t CO₂e in 2021 and 184 173 t CO₂e in 2022. Figure 8.10 shows emissions by category from 2017 to 2022. CERN aims to reduce these emissions by 28 % by the end of Run 3, relative to 2018 levels. To achieve this, gas recirculation and recovery systems are being developed, along with research into more environmentally friendly alternatives to the current gases.

Scope 2 corresponds to CERN’s electricity consumption. GHG emissions amounted to 56 382 t CO₂e in 2021 and 63 161 t CO₂e in 2022. CERN’s main electricity supplier, EDF, provides electricity that is largely of nuclear origin and therefore mostly low-carbon. Figure 8.11 shows CERN’s scope 2 emissions between 2017 and 2022.

Scope 3 includes business travel, commuting, catering, waste treatment, and water purification. In total, excluding procurement (evaluated separately), scope 3 GHG emissions amounted to 78 13 t CO₂e in 2021 and 89 56 t CO₂e in 2022. Figure 8.12 details CERN’s scope 3 emissions, excluding procurement, from 2019 to 2022.

Business travel data in the environmental report only includes personnel on CERN’s payroll and does not account for the wider CERN user community. As CERN users are numerous, their travel-related emissions are likely several times higher than those of CERN staff. For staff, travel-related emissions were 151 t CO₂e in 2021 and 827 t CO₂e in 2022. These emissions are largely dominated by air travel, particularly long-haul flights. It is important to note that travel volumes were still affected by the COVID-19 pandemic.

Commuting is the main contributor to CERN’s scope 3 emissions (excluding procurement), amounting to 5 443 t CO₂e in 2021 and 5 507 t CO₂e in 2022. As with business travel, only staff on CERN’s payroll are included in this figure.

Finally, procurement is the largest contributor to scope 3 emissions, but also the most difficult to quantify. CERN spent approximately 471 MCHF in 2021 and 462 MCHF in 2022 on purchases, including goods, services, and supplies. Emissions related to purchased

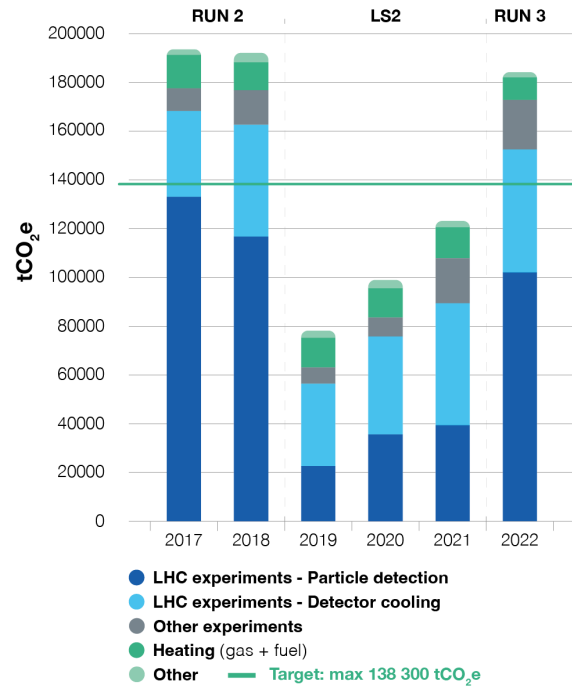


Figure 8.10: CERN scope 1 emissions for 2017-2022 by category. Figure from Ref. [173].

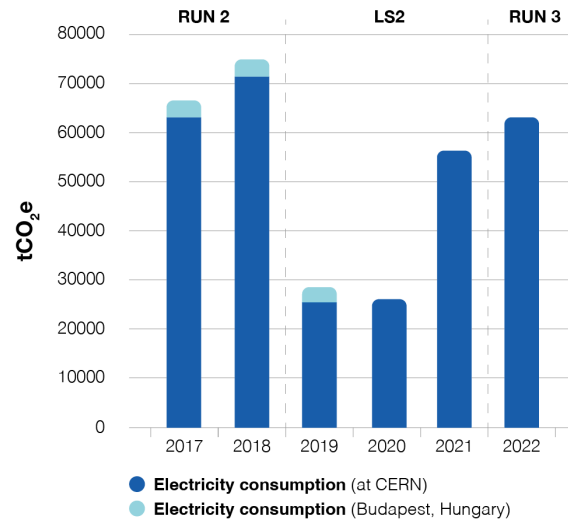


Figure 8.11: CERN scope 2 emissions for 2017-2022. Figure from Ref. [173].

goods, services, and capital goods amounted to 98 030 t CO₂e in 2021 and 104 974 t CO₂e in 2022. Figure 8.13 shows emissions by procurement category.

8.3 LPNHE's emissions

8.3.1 Laboratory presentation

The LPNHE (Laboratoire de Physique Nucléaire et de Hautes Énergies) is a French laboratory affiliated with IN2P3 (Institut National de Physique Nucléaire et de Physique

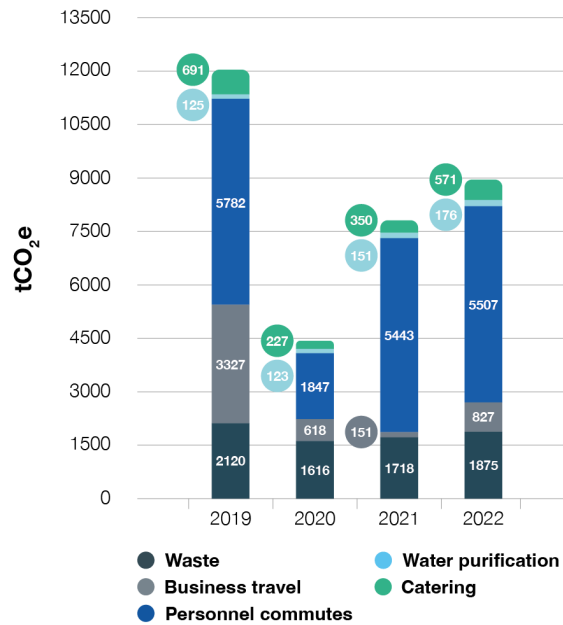


Figure 8.12: CERN scope 3 emissions for 2019-2022 (excluding procurement). Figure from Ref. [173].

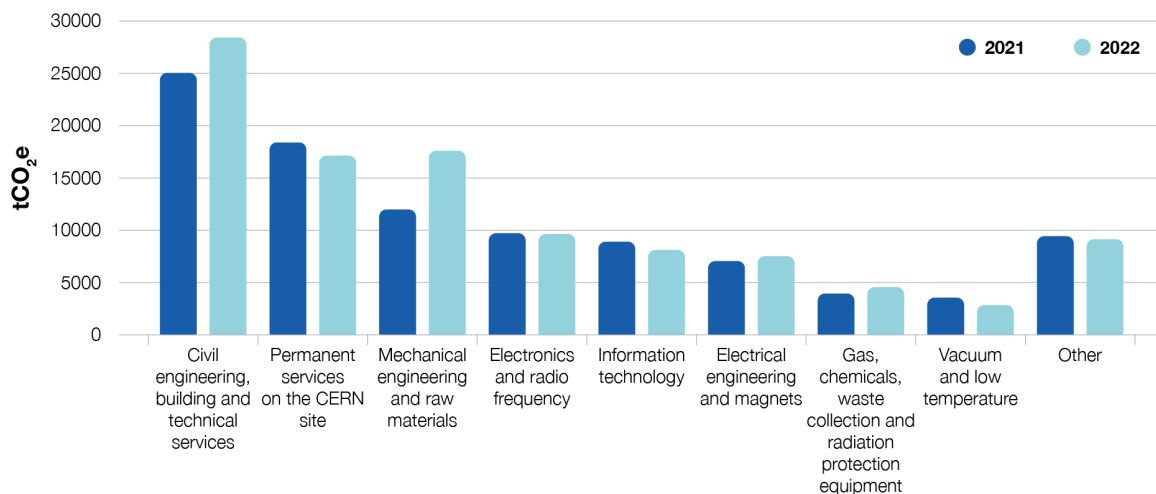


Figure 8.13: CERN emissions by procurement category for 2021-2022. Figure from Ref. [173].

des Particules), an institute of CNRS (Centre National de la Recherche Scientifique), as well as with Sorbonne Université and Université Paris Cité [174]. It comprises research groups, technical services (informatics, mechanics, and electronics), and support services (administration and logistics).

The laboratory is involved in several major international experiments, such as ATLAS (including ITk and HGTD) and FCC in the context of research on mass and fundamental interactions; LHCb, T2K, and COMET (J-PARC) for the study of matter-antimatter asymmetry; CTA, H.E.S.S., DAMIC-M, GRAND, and XENON for the study of cosmic rays and dark matter; and Rubin Observatory (LSST), DESI, Subaru, Hubble, and ZTF in the fields of cosmology and dark energy. The laboratory also contributes to the devel-

opment of advanced technologies in areas such as photometry, mechatronics, and silicon sensors. A large part of its expertise also lies in data science.

Thanks to the tools developed by Labos 1point5, it is possible to estimate the laboratory's GHG emissions. In March 2022, a task force dedicated to sustainable development, the Sustainable Group, was established at the LPNHE to evaluate the laboratory's carbon footprint and to implement a strategy aimed at reducing GHG emissions, in line with ministerial objectives: a 40 % reduction by 2030, and the achievement of carbon neutrality by 2050 [175].

8.3.2 GHG emissions assessment

Using the Labos 1point5 tools, the laboratory's carbon footprint has been established for the years 2019 to 2024, excluding 2020 and 2021, which are not representative due to the COVID-19 pandemic. These are represented in Figure 8.14. The year 2023 can be compared by removing the contribution from scientific activities, which were not included in previous years (due to unavailable data) and are not yet available for subsequent years. The year 2023 is therefore particularly representative, and its corresponding carbon footprint is detailed below.

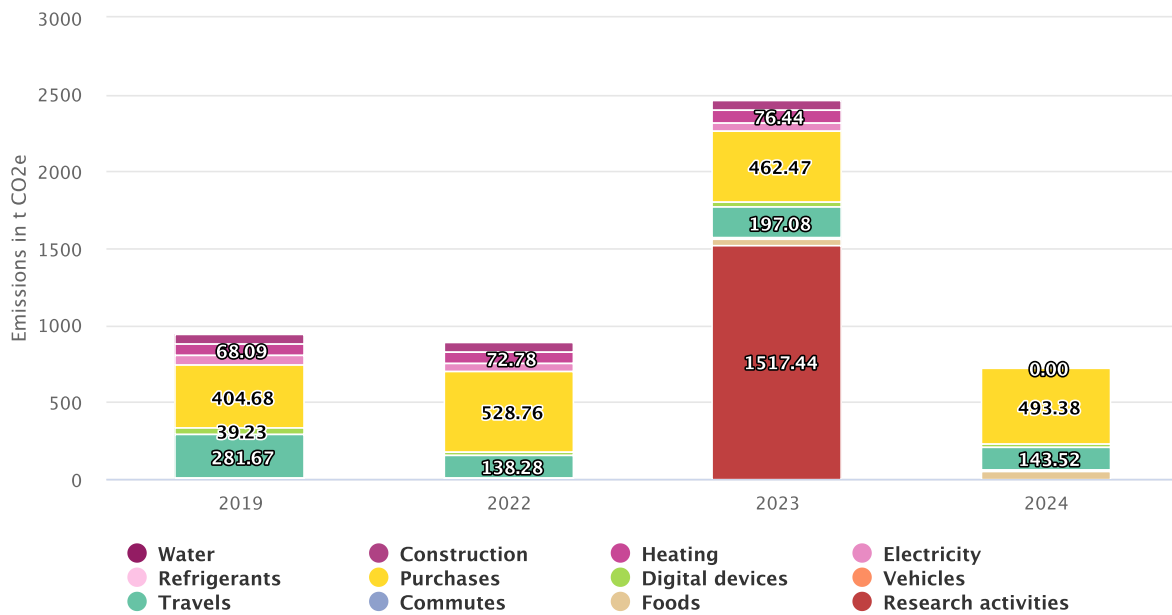


Figure 8.14: LPNHE carbon footprint for 2019, 2022, 2023 and 2024. Figure from Labos 1point5.

In 2023, the laboratory had 147 members: 36 researchers, 20 professors, 46 ITA (engineers, technicians, and administrative personnel), and 46 PhD students and postdocs.

Total GHG emissions for the year amounted to $2466.38 \pm 532.52 \text{ t CO}_2\text{e}$. The carbon footprint per capita was $16\,877 \pm 3643 \text{ kg CO}_2\text{e}$. Assistant and full professors were counted as 50 % members of the laboratory, regardless of the actual time they spent there, to account for half of their work devoted to teaching outside the laboratory. The laboratory's carbon intensity was $1444 \pm 312 \text{ gCO}_2\text{e/e}$.

Research activities were the largest contributor to GHG emissions (61.5 %). A simple cross-multiplication was performed using the carbon footprints provided by major exper-

iments (e.g. CERN) and the number of laboratory users involved in these experiments. The total carbon footprint, $1517.44 \pm 511.81 \text{ t CO}_2\text{e}$, was largely dominated by CERN (96%). Figure 8.15 shows the respective impacts of LHC and non-LHC experiments, as well as astronomy experiments.

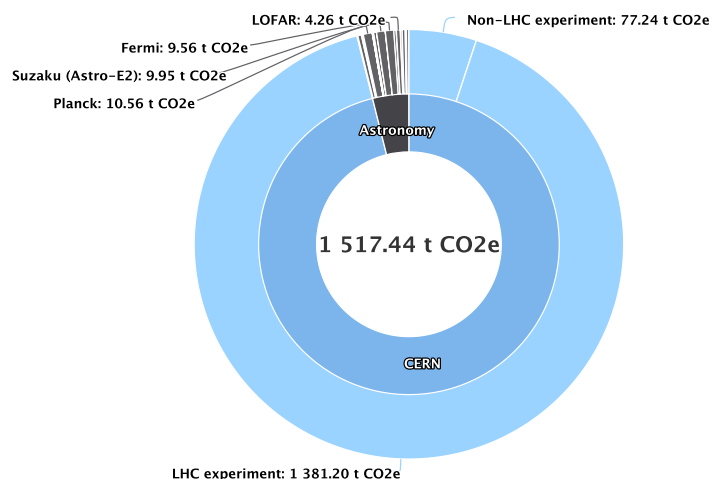


Figure 8.15: Carbon footprint of LPNHE research activities by type (2023). Figure from Labos 1point5.

The second source of emissions is due to purchases: $462.47 \pm 121.89 \text{ t CO}_2\text{e}$, representing 19% of the total carbon footprint. Laboratory instruments and equipment are the main contributors, as shown in Figure 8.16. These emissions were estimated based on NACRES codes, to which monetary emission factors are associated.

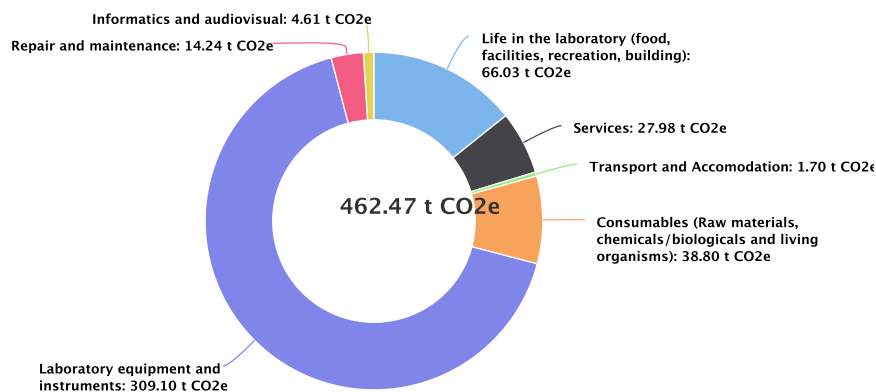


Figure 8.16: Carbon footprint of LPNHE purchases by category (2023). Figure from Labos 1point5.

The third most significant contributor to the LPNHE carbon footprint is travel: $207.23 \pm 71.09 \text{ t CO}_2\text{e}$, accounting for 8%. Professional air travel is by far the dominant component, as shown in Figure 8.17. Commuting is negligible ($8.58 \pm 1.40 \text{ t CO}_2\text{e}$), which is likely due to the laboratory's location in the centre of Paris, benefiting from a wide range of public transport options, especially electric modes such as trains and subways, as well as bike commuting. This is certainly different in cities lacking public transport,

where car commuting can contribute significantly to the carbon footprint. Information about commuting was collected through an anonymous survey sent to all laboratory personnel, with a response rate of 50 %. Results are shown in Figure 8.18. The retained carbon footprint is based only on respondents and constitutes a lower bound on total commuting-related GHG emissions.

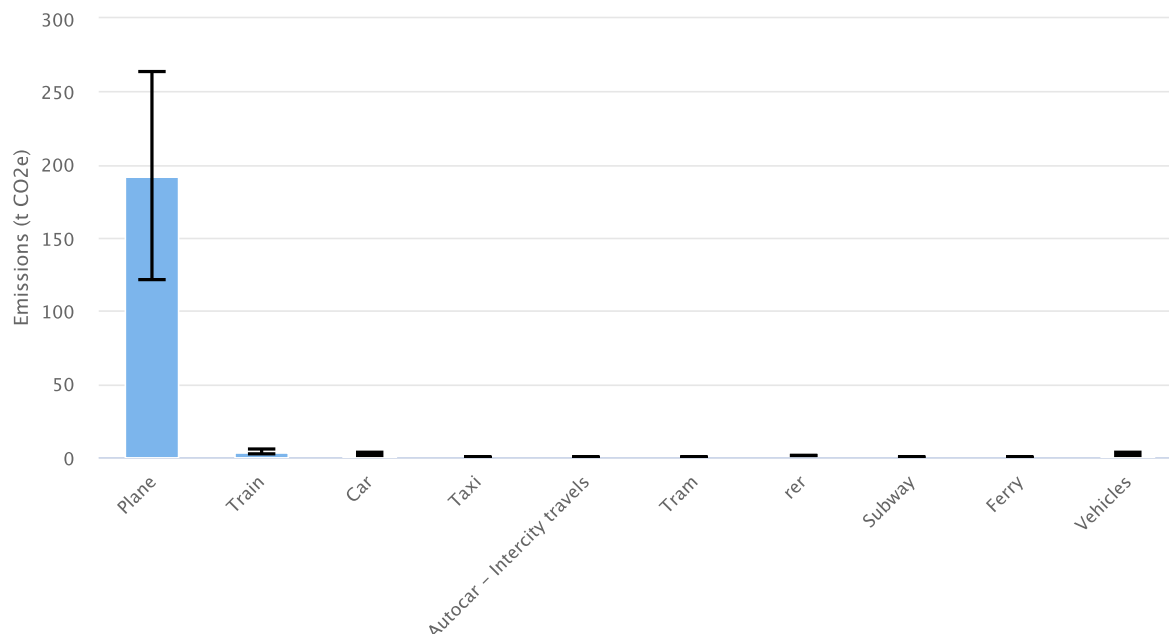


Figure 8.17: Carbon footprint of LPNHE travels (2023). Figure from Labos 1point5.

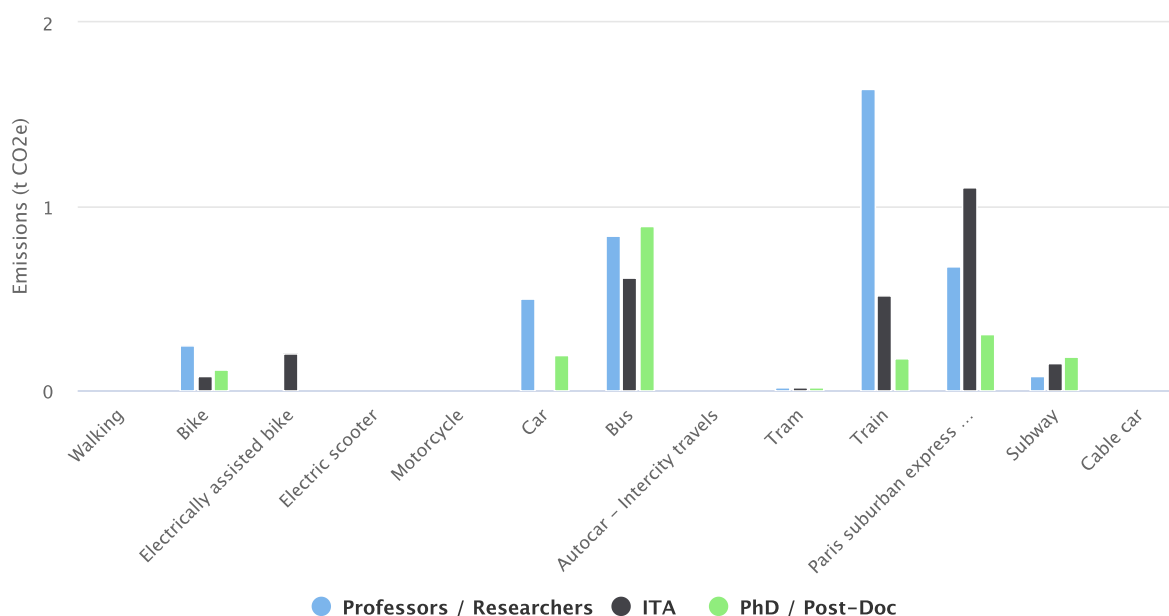


Figure 8.18: Carbon footprint of LPNHE commuting (2023). Figure from Labos 1point5.

The carbon footprint related to the building was estimated based on the data provided by Sorbonne Université for the Pierre et Marie Curie campus. It was not possible to make a more precise measurement than a rough estimate based on the proportion of the building

surface occupied by the laboratory. The building's carbon footprint is divided between its use (mainly heating and electricity) and the amortisation of its construction, for a total of $199.74 \pm 40.80 \text{ t CO}_2\text{e}$, representing 8 % of total emissions.

Finally, the remaining emissions are divided between catering and digital devices. The catering carbon footprint was estimated according to meal type: vegan, vegetarian, with white meat, with red meat, or with fish. An anonymous survey was sent to the entire laboratory to collect information on meal consumption, with a response rate of 46 %. Results are shown in Figure 8.19.

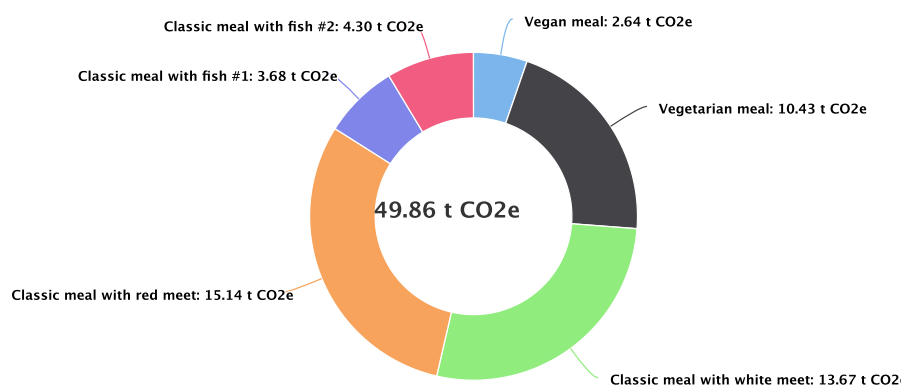


Figure 8.19: Carbon footprint of LPNHE catering (2023). Figure from Labos 1point5.

It is important to note that the content of a meal has a great impact on its carbon footprint, as shown in Figure 8.20. This should be considered in relation to the GHG emissions of an average French citizen (Figure 8.21), for whom food represents one of the largest contributors: the easiest way to reduce GHG emissions at the individual level is to avoid animal products.

8.3.3 Reduction measures

8.3.3.1 Opinions

The survey created by Labos 1point5 mentioned above [171] also assessed the perception of research personnel regarding climate change and their willingness to change practices. First, research staff are largely aware of climate change: 99 % believe that “the climate of the planet is changing,” and 95 % believe that human activities play a major role in, or are the only cause of, climate change. In addition, research staff are highly concerned about climate change: 99 % of respondents expressed concern, and 72 % were very or extremely concerned. 90 % agreed with the statement, “If things continue on their present course, we will soon experience a major ecological catastrophe.”

These concerns are accompanied by a strong desire for change, as illustrated in Figure 8.22 and Figure 8.23. 88 % of respondents agreed with the statement, “Climate urgency calls for profound changes in the practice of our professions.” Opinions on climate change vary little between disciplines. Although PhD-level staff are more concerned about climate change than support staff, the latter are more willing to change their practices.

The conclusion of this survey is the following: the scientific community is largely aware of climate change and wants to implement change, but its attitudes and practices are not

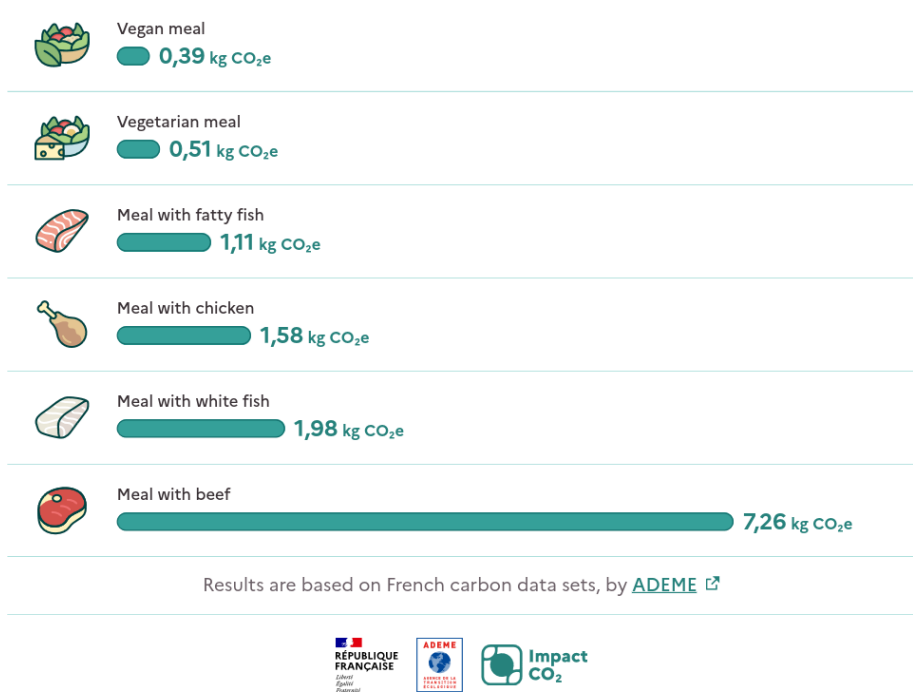


Figure 8.20: Typical carbon footprint of different meals, based on French carbon data sets, by ADEME. Figure from ADEME [176].

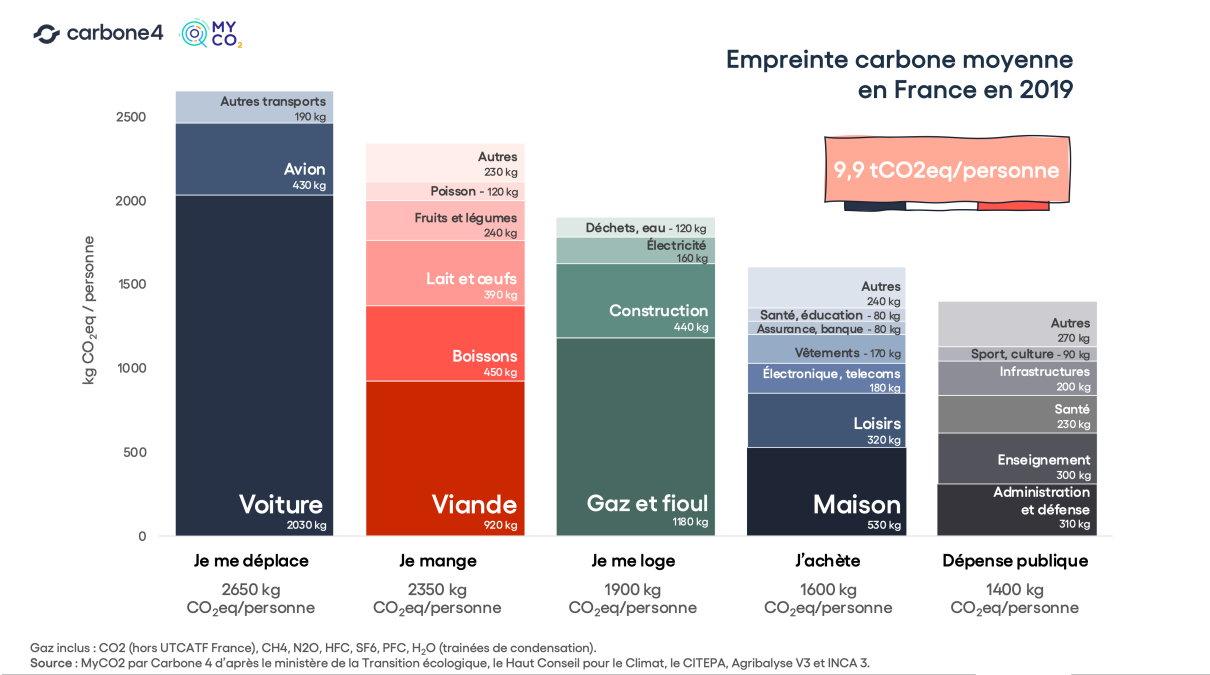


Figure 8.21: Average carbon footprint in France in 2019. Figure from Carbon4 [177].

aligned, and it continues to emit large quantities of GHGs.

This can be explained by several factors, including the importance of travel for career advancement, through participation in conferences, meetings, and similar events. In a survey of over 6000 respondents randomly selected across all research disciplines [178], air travel was associated with a higher publication rate and *h*-index. In particular, the effects

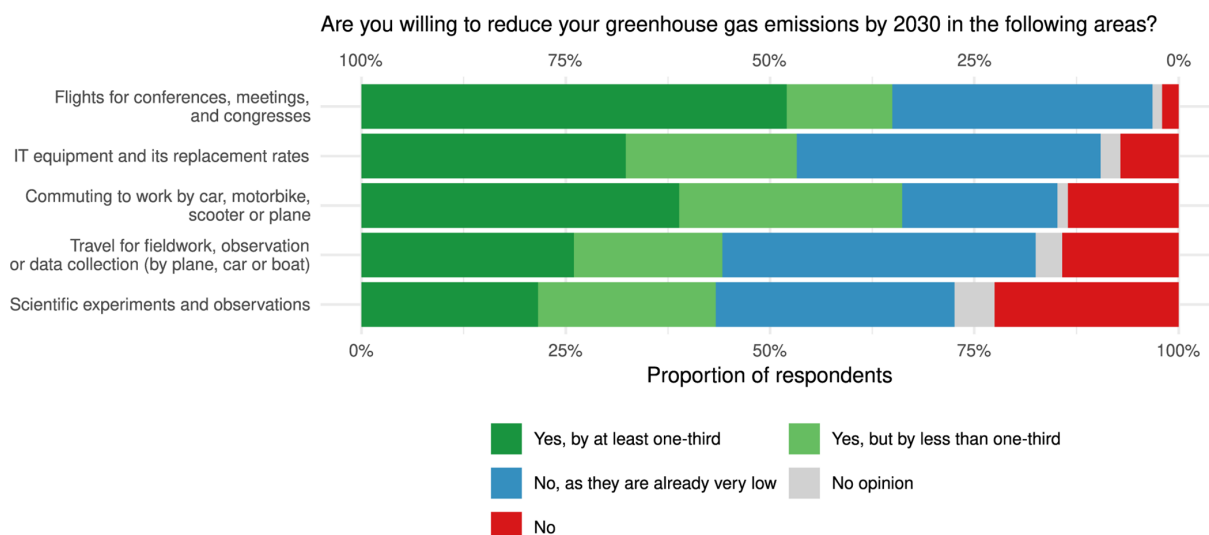


Figure 8.22: Willingness to reduce GHG emissions by 2030 in various areas. Figure from Ref. [171].

of gender, career stage, and discipline were considered. The study suggests that air travel enables early-career scientists to gain visibility and senior scientists to maintain theirs.

8.3.3.2 Travels

Air travel accounts for almost all (96 %) GHG emissions from professional travel, two-thirds of which come from intercontinental flights. In a study based on the Labos 1point5 database, including 130 000 travel segments by car, train, and plane in 159 research entities across a wide range of disciplines in France [172], different mitigation strategies have been explored to reduce the GHG emissions related to research travel. The strategies studied include modal shifts, consisting of replacing air travel with low-carbon alternatives, and moderating air mileage, which involves reducing travel distance or the number of flights.

First, the study showed that very focused recommendations from official guidelines are insufficient. For example, institutional recommendations to take trains for trips under 4 h only allow a reduction of 3 % in total emissions from air travel. Several institutional policies propose switching the vehicle fleets owned by research units to electric or hybrid vehicles. However, in the GHG 1point5 database, vehicle emissions account for only about 2.1 % of total travel-related GHG emissions. Replacing the entire vehicle fleet with electric cars would reduce travel-related GHG emissions by just 1.1 %.

Moderation options have also been studied. Several strategies have been tested. For instance, decreasing the flown distance by 20 % for each research unit would reduce travel-related emissions by 20 %. Applying a threshold below which travel must be undertaken by train can have a significant effect: with a threshold of 5780 km/person/year, GHG emissions are reduced by 38 %, and half of the units are affected. The emission reduction reaches 47 % if the threshold is set at 4500 km/person/year, with 61 % of research units affected. Limits on the number of plane trips per capita allow GHG emissions to decrease by between 20 % and 60 %. In addition, quotas depending on travel purpose allow carbon footprint reductions of 8 % when reducing air travel to conferences by 20 %, 19 % by halving conference attendance, or 41 % by enforcing a fully virtual format for all conferences and seminars. Figure 8.24 shows the percentage of total travel-induced GHG reductions from single or combined options.

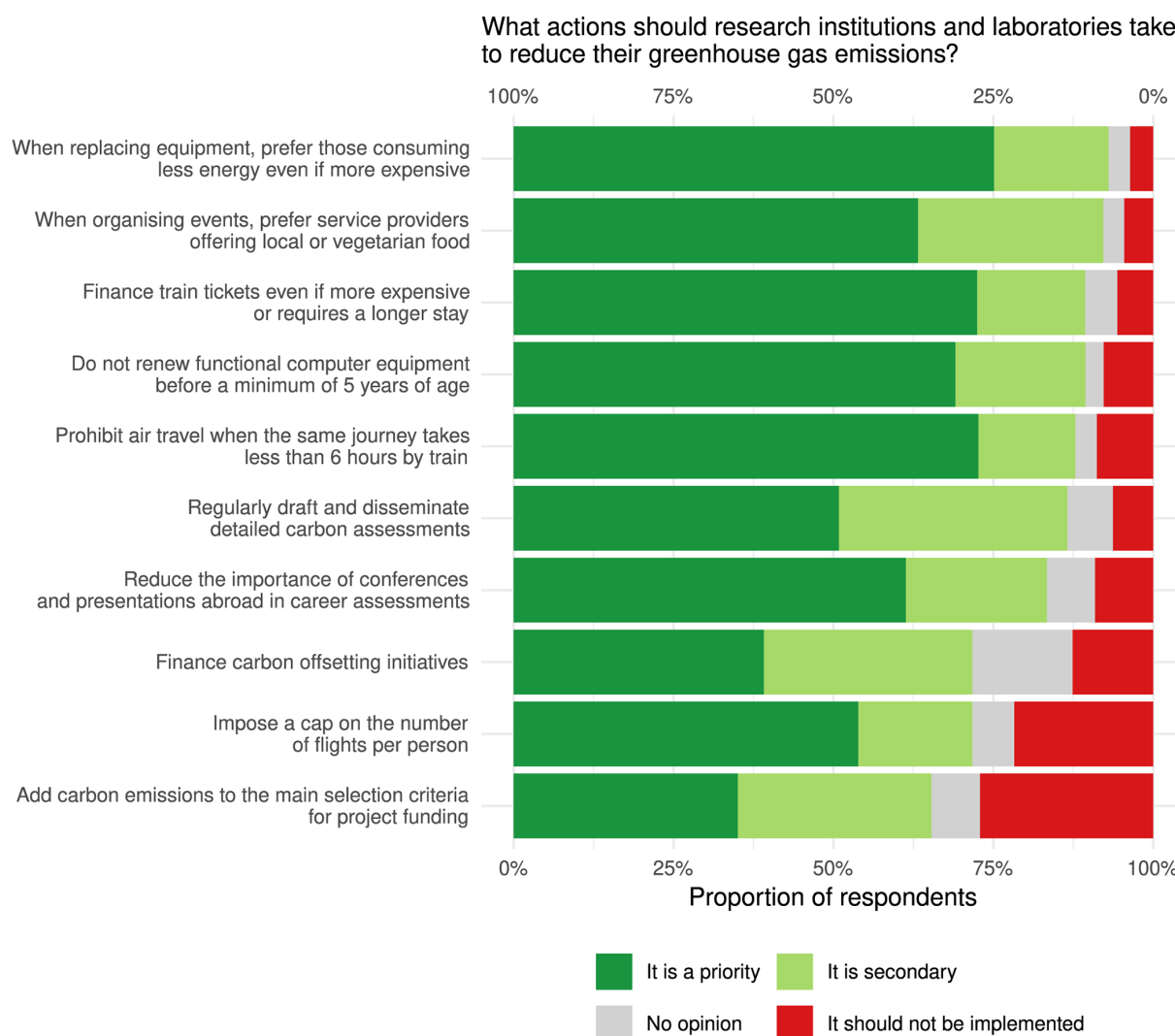


Figure 8.23: Support to institutional actions to reduce GHG emissions. Figure from Ref. [171].

To achieve the European Community's ambition—reducing GHG emissions by 55 % by 2030 compared to 1990 and reaching climate neutrality by 2050—it is useful to combine mitigation options. The study shows, for example, that a 50 % reduction can be achieved or even exceeded by halving the number of flights or the overall distance travelled.

Moreover, imposing a limit on air-travel mileage or on the number of long-haul trips at the per-unit median level allows for a 50 % reduction in travel emissions, if combined with a minimum allowable distance of 1500 km for taking a plane.

Shifting from plane to train for journeys up to 10 h, or 1500 km, has a limited impact, allowing only a 20 % reduction in travel emissions.

Finally, policies based on flight quotas are simpler to implement and can lead to significant results if they effectively reduce the number of long-haul flights.

8.3.3.3 At the LPNHE

Up to now, the LPNHE Sustainable Group has focused on estimating the laboratory's carbon footprint. The transition phase is just beginning.

A charter has been written by the group and adopted by the laboratory council [179].

		Minimum Allowable Distance (in distance or approximate duration) for Air Travel Clearance								
		No modal shift policy	375km (~2h30)	600km (~4h)	In mainland France	900km (~6h)	1000km (~6h40)	1200km (~8h)	1500km (~10h)	
Air Mileage Moderation in Quantity or Distance	No moderation policy	0	0.3	3	8	12	15	18	21	
	Reduce air travel number for conferences	20% fewer trips	8	8	10	13	17	20	23*	26*
		50% fewer trips	19	19	21	24	27	30	32	34
		1 r. trip/3 years	14	13	14	17	18*	19*	21**	22**
		1 r. trip/4 years	18	18	19	22	22	23*	24*	23**
	Reduce long-haul air travel number	20% fewer trips	13**	14**	16**	21**	26**	28**	32**	35**
		50% fewer trips	32**	33**	35**	40**	44**	47**	51**	53**
		1 r. trip/4 years	28**	28**	31**	35**	40**	43**	46**	49**
		1 r. trip/6 years	37**	38**	40**	45**	50**	52**	56**	59**
	Reduce air travel mileage	20% decrease	20***	20***	22***	26***	30***	32***	34***	37***
		50% decrease	48***	49***	50***	52***	54***	56***	57***	59***
		5800km/year	38***	38***	39***	41***	42***	44***	45***	46***
		4500km/year	47***	47***	48***	50***	51***	52***	52***	53***
	Reduce air travel number	20% fewer trips	19	20*	22*	25*	29*	31*	34*	36**
		50% fewer trips	48	48	49	52	54	55	57*	58*
		1 r. trip/year	36*	36*	33	32*	29*	28**	27**	27***
		1 r. trip/2 years	61	61	58	57	51*	50*	46*	44**

Figure 8.24: Percentage of total (plane, train, car) travel-induced GHG reductions from single or combined options. Colours represent the intensity in GHG mitigation potential (yellow corresponds to higher potential). Figure from Ref. [172].

It recalls the climate objectives and states that the laboratory is committed to a sustainable development strategy and will reduce its GHG emissions to meet the ministry's recommendations. The charter is reproduced below, translated into English:

Charter of Commitment to Sustainable Development at the LPNHE (UMR 7585)

Preamble

a) The laboratory is committed to the sustainable development* of its research and research support activities. With this in mind, the following charter was drawn up and unanimously approved by the Laboratory Council on 12 November 2024.

b) This commitment is based on the LPNHE's Sustainable Development Task Force (SDTF), which has existed since 2022. It is made up of members of the laboratory, regardless of rank, position or supervisory authority, who have voluntarily committed to thinking about solutions to reduce the laboratory's greenhouse gas emissions. The SDTF is supported by the laboratory's management. It will draw up the GHG balance sheet (Greenhouse Gas Balance Sheet) for 2019, based on available data, and propose concrete actions (see below).

Commitment

1. The laboratory is making a long-term, collective commitment to implementing measures to reduce greenhouse gas emissions. Each member of the lab is invited to share their proposals with the members of the SDTF.

2. The proposals received by the SDTF are discussed by all members of the laboratory in a collegial manner, for example during a presentation at the Friday meeting. The proposals are first submitted to the laboratory management for approval. If necessary, the Scientific Advisory Board is consulted, particularly when the proposed actions are likely to have an impact on scientific projects. They are then presented, discussed, and voted on by the Laboratory Council. Once adopted, these proposals form the building blocks of the Sustainable Development Action Plan (SDAP).

3. The aim of this SDAP is to reduce the laboratory's carbon footprint by 10 % per year, starting in 2024, in order to reach the target of 50 % by 2030, in line with the recommendations of the French Ministry of Ecological Transition**.

4. The SDAP defines the actions to be taken in concrete terms. It is not set in stone and is intended to include, on an ongoing basis, any new action proposals validated as provided for in point 2. The annual GHG balance sheet, the SDAP, and its results are available on the laboratory's website.

* Sustainable development: development that is economically efficient, socially equitable and ecologically sustainable (Earth Summit, UN, 1992)

** <https://www.ecologie.gouv.fr/politiques-publiques/strategie-nationale-bas-carbone-snbc>

A survey has been sent to all laboratory members to collect ideas on how to reduce the laboratory's carbon footprint. To date, few responses have been received.

Additionally, purchase points of contact have been appointed in the various laboratory services (administration, electronics, IT, mechanics, and maintenance) to propose sustainable purchasing solutions, such as material reuse, resource and equipment sharing between laboratories on the Pierre et Marie Curie campus or within the network of laboratories in which the LPNHE is involved, and the identification of suppliers offering sustainable options.

Conclusions

The first part of this thesis presents a method for correcting the calorimetric isolation energy of photons in MC simulations, whose modelling shows a discrepancy with experimental data. The methodology developed is based on all Run 2 data, corresponding to an integrated luminosity of 139 fb^{-1} , and uses MC simulations to estimate the fraction of true photons in the data. A continuous correction is provided as a function of the photons' transverse energy and the average number of interactions per beam crossing, for each bin of detector region and photon conversion type, thus offering greater precision than previous corrections.

The second part of this work involves developing new criteria for identifying and isolating photons using machine learning tools, specifically designed for photon pairs rather than individual photons. This approach makes it possible to identify diphotons with an angular separation of $0.1 < \Delta R_{\gamma\gamma} < 0.2$, thus enabling access to photon pairs with a mass below 10 GeV. Diphoton identification exploits correlations between the angular separation of the two photons and shower shape variables, which describe the shape of the electromagnetic shower in the calorimeter. A multilayer perceptron (MLP) has been developed, achieving excellent performance, with a selection efficiency greater than 95 % for true diphotons and a rejection rate greater than 90 % for fake diphotons. Diphoton isolation uses the energy deposits in the calorimeter around the diphoton candidate, exploiting their transverse energy and three-dimensional position, as well as the average number of interactions per beam crossing. Since the inputs are of variable size, more flexible architectures than an MLP have been studied: a convolutional neural network (CNN), a graph neural network (GNN), and a Set Transformer. The GNN achieves the best performance, with an efficiency for selecting true diphotons greater than 65 % (and up to 90 % above 10 GeV), as well as a fake diphoton rejection rate greater than 50 %, reaching 80 % at low mass.

A search for a new diphoton resonance in the mass range between 5 and 20 GeV is then presented. The background components are estimated from a control region defined in the data using track isolation. The previously developed diphoton identification and isolation methods significantly reduce the reducible background, dominated by jets, making true QCD diphotons the main residual component. The expected shape of a possible signal is derived from the detector's mass resolution. The effective background, defined as the contribution below the peak of a hypothetical resonance, is then calculated. The limit on the number of signal events is determined and converted into the expected limit on the fiducial cross section multiplied by the branching ratio of the potential resonance, at a confidence level of 95 %. The tools developed in this work thus make it possible to extend the scope of the search to previously unexplored masses below 10 GeV.

Finally, in the current context of global warming, scientific research plays a significant role. Its impact on climate change must be accurately assessed, and concrete measures to reduce emissions must be implemented. An overview of greenhouse gas emissions from

French public research, from CERN, a collaboration in which this thesis is involved, and from LPNHE, the host laboratory for this work, is presented. The actions taken within LPNHE to reduce the laboratory's carbon footprint are also detailed.

Appendix A

Statistical tools

A.1 Probability interpretation

Statistics is one of the fundamental tools in experimental and theoretical sciences¹. They enable us to quantify the probability of an experiment's outcome given a theoretical prediction, or to assess the likelihood of a theory based on experimental data.

They are based on the concept of probability. Formally, given a set Ω and a σ -algebra A on Ω (the collection of all subsets of Ω , called *events*), a probability P is a map from A to \mathbb{R} such that:

- $\forall X \in A, P(X) \geq 0$
- $P(\Omega) = 1$
- For any countable sequence of disjoint sets X_1, X_2, \dots : $P\left(\bigcup_{i=1}^{\infty} X_i\right) = \sum_{i=1}^{\infty} P(X_i)$.

These conditions are referred to as the Kolmogorov axioms [182].

Two schools of thought offer different interpretations of probability. On one side, frequentists consider the parameters of a theory as constants of Nature, to which probability cannot be assigned. Probabilities are applied to experiments and are defined as the long-term relative frequency of an event, assuming the experiment can be repeated infinitely many times:

$$P(X) = \lim_{n \rightarrow \infty} \frac{\text{number of occurrences of event } X}{n}, \quad (\text{A.1})$$

where n is the number of experiments.

Probabilities are assigned to experimental results, assuming a fixed underlying theory, without any prior assumption about the theory itself:

$$P(\text{data} \mid \text{theory}). \quad (\text{A.2})$$

Bayesians, on the other hand, consider theory parameters as random variables. Probability is defined as a degree of belief:

$$P(X) = \text{degree of belief that proposition } X \text{ is true.} \quad (\text{A.3})$$

¹This appendix is primarily based on my readings of Ref. [180] and Ref. [181].

In this framework, probabilities are assigned not to experiments, but to theories:

$$P(\text{theory} \mid \text{data}). \quad (\text{A.4})$$

The fundamental theorem used in Bayesian statistics is Bayes' theorem:

$$P(\text{theory} \mid \text{data})P(\text{data}) = P(\text{data} \mid \text{theory})P(\text{theory}), \quad (\text{A.5})$$

which requires a prior assumption on the theory, $P(\text{theory})$.

While the frequentist formalism proposes an objective definition of probability based on the limit of repeated experiments, it does not offer the possibility of assigning a probability to theories. On the contrary, the Bayesian formalism assigns probabilities to theories, but at the cost of subjective assumptions about them. This duality is well summarised by Louis Lyons [183]:

Bayesians address the question everyone is interested in by using assumptions no-one believes, while Frequentists use impeccable logic to deal with an issue of no interest to anyone.

In experimental particle physics, the frequentist formalism is primarily used.

A.2 Probability densities and likelihood functions

Let's first consider an observable x , which can be any measured quantity in an experiment, such as the diphoton invariant mass. Repeating the experiment many times will result in different values of x . The distribution of the observable is called a probability density function (PDF) and is denoted by $f(x)$. It must be normalised to unity:

$$\int f(x)dx = 1. \quad (\text{A.6})$$

Often, the PDF depends on other parameters $\boldsymbol{\eta}$, giving rise to a parametric family of PDFs $f(x \mid \boldsymbol{\eta})$, called a probability model, or simply a model. In the following, bold notation is used for vector quantities. Usually, the parameters can be separated into two categories: those of interest, called *parameters of interest* (usually denoted $\boldsymbol{\mu}$), and the remaining ones, known as *nuisance parameters* (denoted $\boldsymbol{\theta}$).

Parameters are estimated using functions of the data called estimators. They are denoted with a "hat". Some examples of estimators, considering a sample x_1, \dots, x_n , are the mean:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (\text{A.7})$$

and the variance:

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{\mu})^2. \quad (\text{A.8})$$

A good estimator must satisfy several properties. It must be:

- Consistent: $\lim_{n \rightarrow \infty} \hat{\theta} = \theta$

- Unbiased: $E[\hat{\theta}] = \theta$, where $E[\hat{\theta}]$ is the expectation value of $\hat{\theta}$, defined as $E[\hat{\theta}] = \int \hat{\theta}(x)f(x)dx$
- Efficient: its variance $\sigma_{\hat{\theta}}^2$ must be as low as possible
- Robust: a single x_i cannot arbitrarily affect the estimation. For instance, the median is a robust measure of central tendency, whereas the mean is not.

$f(x)$ describes the probability density for the observable x for a single event. For a dataset with many independent events, $\mathbf{x} = (x_1, \dots, x_n)$, the probability density for observing this set of events is given by:

$$\mathbf{f}(\mathbf{x} \mid \nu, \boldsymbol{\eta}) = \text{Pois}(n \mid \nu) \prod_{i=1}^n f(x_i \mid \boldsymbol{\eta}), \quad (\text{A.9})$$

where $\text{Pois}(n \mid \nu)$ is the Poisson probability of observing n events, given ν expected.

This allows for the construction of another fundamental quantity, called the *likelihood function*. It is numerically equivalent to $\mathbf{f}(\mathbf{x} \mid \boldsymbol{\eta})$ with \mathbf{x} fixed. It is not a probability density, but a function of the parameters, given the data: $L(\boldsymbol{\eta} \mid \mathbf{x})$.

The likelihood function encodes the compatibility of the data with a model, and provides a way to estimate the parameter values: the best estimates of the parameters $\boldsymbol{\eta}$ are those that maximise the likelihood function. These estimators are known as maximum likelihood estimators (MLEs). In practice, for numerical stability, the negative log-likelihood $-\ln L(\boldsymbol{\eta} \mid \mathbf{x})$ is minimised.

For a multi-parameter likelihood function $L(\boldsymbol{\eta})$ (with the dependence on the data being implicit), the maximum likelihood estimate $\hat{\boldsymbol{\eta}}$ is well defined. The components $\hat{\eta}_p$ ($1 \leq p \leq n$) are called *unconditional maximum likelihood estimates*. It can also be useful to define the maximum likelihood estimate of η_p when the other parameters are fixed to a value $\boldsymbol{\eta}_0$ (here, $\boldsymbol{\eta}_0$ is a vector of $n - 1$ components). This is referred to as the *conditional maximum likelihood estimate* and is denoted $\hat{\eta}_p(\boldsymbol{\eta}_0)$.

A.3 Hypothesis testing

In frequentist statistics, a hypothesis is tested using the formalism of a *test statistic* T , which is a map from the data to real values: $T(\mathbf{x}) \in \mathbb{R}$.

First, a null and an alternative hypothesis, H_0 and H_1 , are defined. For example, in a BSM search, the null hypothesis is “SM only” and the alternative hypothesis is “SM + BSM”. Then, a test statistic T is defined, for example, the number of observed events. The distributions of the test statistic under both hypotheses are derived: $f(T \mid H_0)$ and $f(T \mid H_1)$. An acceptance region is defined such that the null hypothesis is accepted if the test statistic falls within it:

$$T(\mathbf{x}) < k_{\alpha}, \quad (\text{A.10})$$

where k_{α} is the upper limit on $T(\mathbf{x})$ such that \mathbf{x} lies within the chosen acceptance region.

The size of the test, α , is defined as the probability of rejecting the null hypothesis H_0 when it is true (also called a Type-I error), i.e., the probability that the test statistic falls outside the acceptance region when H_0 is true:

$$\alpha = P(T(\mathbf{x}) \geq k_\alpha \mid H_0). \quad (\text{A.11})$$

The probability of accepting the null hypothesis when the alternative is true (Type-II error) is denoted β and given by:

$$\beta = P(T(\mathbf{x}) < k_\alpha \mid H_1). \quad (\text{A.12})$$

$1 - \beta$ is called the *power* of the test.

Using the experimental data, the observed value of the test statistic, T_{obs} (which is a function of the data \mathbf{x}), is used to calculate the *p*-value:

$$\begin{aligned} p &= \int_{T_{\text{obs}}}^{\infty} f(T \mid H_0) dT, \\ &= P(T \geq T_{\text{obs}} \mid H_0). \end{aligned} \quad (\text{A.13})$$

The *p*-value is the probability of obtaining this value for the test statistic, or a more extreme one, if the null hypothesis H_0 is true. If it is small enough, i.e., lower than the size of the test, the data are considered too improbable under H_0 , and the null hypothesis is rejected.

The choice of the test size is purely conventional. In experimental sciences, a common choice is $\alpha = 0.05$. In particle physics, for discoveries, $\alpha = 2.87 \times 10^{-7}$ is usually chosen, corresponding to 5σ of a unit Gaussian (see Section A.4.2). An $\alpha = 1.35 \times 10^{-3}$ (3σ) is regarded as evidence.

For a fixed test size α , the test with the highest power is desired. This is given by the Neyman–Pearson lemma. For a fixed size α , the test that maximises the power is the ratio of the likelihoods:

$$T_{NP} = \frac{L(H_0 \mid \mathbf{x})}{L(H_1 \mid \mathbf{x})}. \quad (\text{A.14})$$

A.4 LHC statistical procedures

A.4.1 Test statistics

In LHC studies, it is common to introduce the signal strength μ , which scales the total rate of signal events:

$$\mu = \frac{\sigma_{\text{measured}}}{\sigma_{\text{theory}}}, \quad (\text{A.15})$$

where σ_{measured} is the measured cross section (or cross section times branching ratio), and σ_{theory} is the cross section predicted by the theory, e.g., from the Standard Model. Thus, $\mu = 0$ corresponds to the background-only hypothesis, and $\mu > 0$ corresponds to the signal+background hypothesis. It is then useful to separate the full list of parameters into the parameter of interest μ and the nuisance parameters $\boldsymbol{\theta}$:

$$\boldsymbol{\eta} = (\mu, \boldsymbol{\theta}). \quad (\text{A.16})$$

To obtain a continuum of hypotheses, the *profile likelihood ratio* is defined:

$$\lambda(\mu) = \frac{L(\mu, \hat{\boldsymbol{\theta}}(\mu))}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})}, \quad (\text{A.17})$$

where $\hat{\boldsymbol{\theta}}(\mu)$ is the “profiled value of $\boldsymbol{\theta}$ ”, i.e., the value of $\boldsymbol{\theta}$ that maximises the likelihood for the given μ . This ratio depends implicitly on the data.

In any physical theory, the signal rate must be positive, imposing $\mu \geq 0$. However, it is convenient to allow $\mu < 0$. In particular, $\hat{\mu} < 0$ indicates a deficit in signal-like events with respect to the background-only hypothesis. Consequently, it is helpful to allow $\mu < 0$ and to impose the constraint within the test statistic itself:

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{L(\mu, \hat{\boldsymbol{\theta}}(\mu))}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})} & \text{if } \hat{\mu} \geq 0, \\ \frac{L(\mu, \hat{\boldsymbol{\theta}}(\mu))}{L(0, \hat{\boldsymbol{\theta}}(0))} & \text{if } \hat{\mu} < 0. \end{cases} \quad (\text{A.18})$$

For discovery, the test statistic q_0 is used to differentiate the background-only hypothesis ($\mu = 0$) from the alternative hypothesis ($\mu > 0$):

$$q_0 = \begin{cases} -2 \ln \frac{L(\mu, \hat{\boldsymbol{\theta}}(\mu))}{L(0, \hat{\boldsymbol{\theta}}(0))} & \text{if } \hat{\mu} > 0, \\ 0 & \text{if } \hat{\mu} \leq 0. \end{cases} \quad (\text{A.19})$$

q_0 is a test statistic for a one-sided alternative.

For limit setting, the test statistic \tilde{q}_μ is used to differentiate a signal produced at a rate μ from a signal produced at a lower rate $\mu' < \mu$:

$$\tilde{q}_\mu = \begin{cases} -2 \ln \tilde{\lambda}(\mu) & \text{if } \hat{\mu} \leq \mu \\ 0 & \text{if } \hat{\mu} > \mu \end{cases} = \begin{cases} -2 \ln \frac{L(\mu, \hat{\boldsymbol{\theta}}(\mu))}{L(0, \hat{\boldsymbol{\theta}}(0))} & \text{if } \hat{\mu} < 0, \\ -2 \ln \frac{L(\mu, \hat{\boldsymbol{\theta}}(\mu))}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})} & \text{if } 0 \leq \hat{\mu} \leq \mu, \\ 0 & \text{if } \hat{\mu} > \mu. \end{cases} \quad (\text{A.20})$$

This is a test statistic for a one-sided alternative (one-sided upper limit).

Finally, the test statistic \tilde{t}_μ is used to differentiate a signal produced at a rate μ a signal produced at a lower or higher rate $\mu' \neq \mu$ (alternative hypothesis):

$$\tilde{t}_\mu = -2 \ln \tilde{\lambda}(\mu). \quad (\text{A.21})$$

This is a test statistic for a two-sided alternative.

A.4.2 p -values

The test statistic distribution depends simultaneously on the parameter of interest μ and the nuisance parameters $\boldsymbol{\theta}$: $f(\tilde{q}_\mu \mid \mu, \boldsymbol{\theta})$. Similar expressions are obtained for the other test statistics.

The p -value also depends on both:

$$p_{\mu, \boldsymbol{\theta}} = \int_{\tilde{q}_{\mu, \text{obs}}}^{\infty} f(\tilde{q}_{\mu} \mid \mu, \boldsymbol{\theta}) d\tilde{q}_{\mu}, \quad (\text{A.22})$$

where $\tilde{q}_{\mu, \text{obs}}$ is the observed value of the test statistic, and thus depends on the outcome of the experiment.

Since the null hypothesis is rejected for a sufficiently small p -value, it must be rejected if the p -value is below the chosen threshold for all values of $\boldsymbol{\theta}$. The supremum p -value is subsequently considered to determine whether to reject the null hypothesis:

$$p_{\mu}^{\text{sup}} = \sup_{\boldsymbol{\theta}} p_{\mu, \boldsymbol{\theta}}. \quad (\text{A.23})$$

One important result stemming from Wilks' theorem [184] is that, asymptotically (i.e., for a large number of events), the profile likelihood ratio becomes independent of the nuisance parameters. As a result, in this limit:

$$\forall \boldsymbol{\theta}, \quad p_{\mu}^{\text{sup}} = p_{\mu}^{\boldsymbol{\theta}}. \quad (\text{A.24})$$

When using toy MC techniques to calculate p -values, some residual dependence on $\boldsymbol{\theta}$ may remain. Therefore, the profiled value of $\boldsymbol{\theta}$, which yields the highest p -value, and thus the best agreement with the null hypothesis, is used in the calculation:

$$p_{\mu, \boldsymbol{\theta}} = \int_{\tilde{q}_{\mu, \text{obs}}}^{\infty} f(\tilde{q}_{\mu} \mid \mu, \hat{\boldsymbol{\theta}}(\mu, \text{obs})) d\tilde{q}_{\mu}. \quad (\text{A.25})$$

A standard 95 % confidence-level, one-sided frequentist confidence upper limit is obtained by solving $p_{\mu_{\text{up}}} = 0.05$ for μ_{up} .

For downward fluctuations, the upper limit of the confidence interval can become arbitrarily small, even smaller than the actual sensitivity of the experiment. To avoid claiming constraints on arbitrarily small signal rates, a modified frequentist method, called CL_s [185, 186], is introduced. A ratio p'_{μ} of p -values is defined:

$$p'_{\mu} = \frac{p_{\mu}}{1 - p_b}, \quad (\text{A.26})$$

where p_b is the p -value from the same test statistic but under the background-only hypothesis:

$$p_b = \int_{\tilde{q}_{\mu, \text{obs}}}^{\infty} f(\tilde{q}_{\mu} \mid 0, \hat{\boldsymbol{\theta}}(\mu = 0, \text{obs})) d\tilde{q}_{\mu}. \quad (\text{A.27})$$

This ratio of probabilities, p'_{μ} , is not itself a probability.

For discoveries, the \tilde{q}_0 test statistic, based on the background-only hypothesis, is used to compute the p -value:

$$p_0 = \int_{\tilde{q}_0, \text{obs}}^{\infty} f(\tilde{q}_0 \mid 0, \hat{\boldsymbol{\theta}}(\mu = 0, \text{obs})) d\tilde{q}_0. \quad (\text{A.28})$$

This p -value is usually converted to the quantile (or “sigma”) of a unit Gaussian corresponding to the same p -value. This is purely conventional and does not assume that p_0 is Gaussian-distributed. The conversion is given by:

$$Z = \Phi^{-1}(1 - p_0), \quad (\text{A.29})$$

where Φ^{-1} is the inverse of the cumulative distribution function of a unit Gaussian.

The result is generally expressed as “ $Z\sigma$ ”, and a discovery is claimed for $Z > 5$ ($p_0 = 2.87 \times 10^{-7}$).

A.4.3 Expected sensitivity

The expected upper limit on a parameter (typically, the cross section times branching ratio) is defined as the upper limit that would be obtained on average (or, more precisely, as the median value) if the background-only hypothesis were true. If a higher limit is observed with the data, this indicates an excess of events compared to the background (a signal, or just a fluctuation).

The expected significance is the significance of the observation based on the signal rate predicted by the Standard Model.

The expected upper limit is computed using the distribution $f(\mu_{\text{up}} | \mu = 0, \theta)$, while the expected significance uses $f(p_0 | \mu = 1, \theta)$ or $f(Z | \mu = 1, \theta)$. The median value of the distribution is used instead of the mean, since it is independent of the use of p_0 or Z .

These distributions depend on the values of the nuisance parameters. By convention, the profiled values based on the observed data are used: $f(\mu_{\text{up}} | \mu = 0, \hat{\theta}(\mu = 0, \text{obs}))$ and $f(p_0 | \mu = 1, \hat{\theta}(\mu = 1, \text{obs}))$.

Additionally, bands are usually drawn around the median upper limit: a dark green band, corresponding to 68 % of $f(\mu_{\text{up}} | \mu = 0, \hat{\theta}(\mu = 0, \text{obs}))$ (equivalent to $\pm 1\sigma$ for a unit Gaussian):

$$\int_0^{\mu_{\pm 1}} f(\mu_{\text{up}} | 0, \hat{\theta}(\mu = 0, \text{obs})) d\mu_{\text{up}} = \Phi(\pm 1); \quad (\text{A.30})$$

and a light yellow band, corresponding to 95 % of the distribution (equivalent to $\pm 2\sigma$ for a unit Gaussian):

$$\int_0^{\mu_{\pm 2}} f(\mu_{\text{up}} | 0, \hat{\theta}(\mu = 0, \text{obs})) d\mu_{\text{up}} = \Phi(\pm 2). \quad (\text{A.31})$$

To perform these integrals, a “toy Monte Carlo” technique can be used. Many pseudo-experiments are generated, and μ_{up} is calculated for each of them, yielding its distribution. The values of $\mu_{\pm 1}$ and $\mu_{\pm 2}$ can then be deduced.

A.4.4 Asymptotic formulas

In the large-sample limit, asymptotic formulas can be used [187], avoiding the generation of a large number of pseudo-experiments. The distribution $f(\tilde{q}_\mu | \mu')$ can be obtained from Wald’s theorem [188], which shows that, for a single parameter of interest, \tilde{q}_μ can be approximated by:

$$\tilde{q}_\mu = \begin{cases} \frac{\mu^2}{\sigma^2} - \frac{2\mu\hat{\mu}}{\sigma^2} & \text{if } \hat{\mu} < 0, \\ \frac{(\mu - \hat{\mu})^2}{\sigma^2} & \text{if } 0 \leq \hat{\mu} \leq \mu, \\ 0 & \text{if } \hat{\mu} > \mu, \end{cases} \quad (\text{A.32})$$

where $\hat{\mu}$ is Gaussian-distributed, with mean μ' and standard deviation σ .

It follows from this equation that $f(\tilde{q}_\mu | \mu')$ satisfies:

$$\begin{aligned}
 f(\tilde{q}_\mu \mid \mu') &= \Phi\left(\frac{\mu' - \mu}{\sigma}\right) \delta(\tilde{q}_\mu) \\
 &+ \begin{cases} \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\tilde{q}_\mu}} \exp\left[-\frac{1}{2} \left(\sqrt{\tilde{q}_\mu} - \frac{\mu - \mu'}{\sigma}\right)^2\right] & \text{if } 0 < \tilde{q}_\mu \leq \mu^2/\sigma^2, \\ \frac{1}{\sqrt{2\pi}(2\mu/\sigma)} \exp\left[-\frac{1}{2} \frac{(\tilde{q}_\mu - (\mu^2 - 2\mu\mu')/\sigma^2)^2}{(2\mu/\sigma)^2}\right] & \text{if } \tilde{q}_\mu > \mu^2/\sigma^2. \end{cases} \quad (\text{A.33})
 \end{aligned}$$

Asymptotic formulas are also available for \tilde{q}_0 and \tilde{t}_μ .

These expressions depend on σ , the variance of the estimator $\hat{\mu}$. It can be estimated using an ‘‘Asimov dataset’’², where all simulated samples are replaced by a single representative one [187]. In an Asimov binned dataset, each bin is filled with its exact predicted number of events. Using this dataset, the variance σ can be estimated as:

$$\sigma \sim \frac{\mu}{\sqrt{\tilde{q}_{\mu,A}}}, \quad (\text{A.34})$$

where $\tilde{q}_{\mu,A}$ is the test statistic evaluated on the background-only Asimov dataset, i.e., with $\mu = 0$.

²The name comes from the story *Franchise* by Isaac Asimov, in which elections are held by replacing all voters with a single one, representative of the entire electorate [189].

List of Figures

1.1	Feynman diagrams at order λ^2 in the scalar ϕ^3 theory.	11
1.2	One-loop in scalar ϕ^3 theory.	11
1.3	Summary of measurements of α_s as a function of the energy scale Q [3]. . .	13
1.4	Illustration of an effective field theory, where a heavy mediator is replaced by a point-like interaction at low energy.	13
1.5	Fundamental particles of the Standard Model.	17
1.6	$xf(x, \mu^2)$ as a function of x , the fraction of momentum carried by the parton, with f the parton distribution functions (PDFs) of the proton obtained in the NNLO NNPDF3.1 global analysis, for two energy scales, (a) $\mu_F^2 = 10 \text{ GeV}^2$ and (b) $\mu_F^2 = 10^4 \text{ GeV}^2$ [12]. The valence quarks are denoted with the letter v . The sea quarks are also represented.	20
1.7	Prompt photon production at leading order in pp collisions. (a) Quark-gluon Compton scattering (direct). (b) Quark-antiquark annihilation (direct). (c) Bremsstrahlung photon radiated by a final state quark (fragmentation).	21
1.8	Diagrams of main processes of diphoton production at the LHC. The upper diagrams depict prompt diphoton processes: (a) Born, (b) box and (c) Bremsstrahlung processes. The bottom row ((d) and (e)) represents jet and dijet fragmentation with a diphoton-like final state. The ‘ γ ’ denotes a final state with a photon-like signature, such as the decay of a π^0 into two collimated photons. Hadron decays are represented by large grey circles. . .	23
1.9	Feynman diagram of ALP production via gluon fusion and decay into photons. The dashed circles represent the effective couplings of the ALP to gluons and to photons.	27
1.10	Left: constraints on the parameter space spanned by the ALP mass and ALP-photon coupling [65]. Right: enlarged display of the constraints from collider searches: LEP (light blue and blue), CDF (purple), LHC from associated production and Z decays (orange), LHC from photon fusion (light orange), and from heavy-ion collisions at the LHC (green).	28
2.1	The first web server: this NeXT machine was used by Tim Berners-Lee in 1990 to develop and run the first WWW server, multimedia browser and web editor [74].	32
2.2	The CERN accelerator complex [78].	34
2.3	(a) Total integrated luminosity and data quality in 2015-2018 (Run 2) [84]. (b) Total integrated luminosity in Run 3 (13.6 TeV pp data only) [85]. . .	36
2.4	Recorded pile-up by ATLAS in (a) Run 2 (2015-2018) [86] and (b) Run 3 (2022-2024) [87].	36

2.5	LHC and HL-LHC plan [89].	37
2.6	Run 3 configuration of the ATLAS detector [90].	38
2.7	Diagram of particle paths in the detector [91].	39
2.8	ATLAS Inner Detector [90].	40
2.9	The structure of the Inner Detector, with the Pixel Detector, including the Insertable B-Layer, the Semiconductor Tracker, and the Transition Radiation Tracker [93]. The trajectory of a hypothetical particle is represented (red line).	41
2.10	Lateral cut-off of the Inner Detector, showing the barrel and end-cap components [79]. The trajectories of two hypothetical charged particles are represented (red lines). The TRT barrel and the IBL are not represented.	42
2.11	Layout of the Inner Detector in the Radius- z plan, showing the major elements and their active dimensions [90].	42
2.12	(a) Fractional energy loss per radiation length in lead as a function of electron or positron energy [97]. The value at which the energy losses by ionisation and by Bremsstrahlung are equal is called the critical energy (E_c). (b) Photon total cross sections as a function of energy in lead, showing the contributions of different processes [97]. At high energies, pair production dominates (κ_{nuc} stands for pair production in the nuclear field and κ_e in the electric field). At lower energies, Compton effect (σ_{Compton}) and photoelectric effect ($\sigma_{\text{p.e.}}$) are dominant.	44
2.13	Energy deposit of electrons with energies between 1 GeV and 1 TeV as a function of the depth in a block of copper [98].	45
2.14	The ATLAS calorimeter systems [90].	48
2.15	Cumulative amounts of material, in units of radiation length X_0 and as a function of $ \eta $, in the electromagnetic calorimeters [79]. The two figures show, in contrast, separately for (a) the barrel and (b) end-cap (right), the thicknesses of each accordion layer as well as the amount of material in front of the accordion.	48
2.16	(a) Accordion structure of the barrel EM calorimeter [101]. (b) Sketch of a barrel module [79]. The different layers and the granularity in η and ϕ are visible.	49
2.17	Cut-away view of an end-cap cryostat showing the positions of the three end-cap calorimeters [79].	49
2.18	Sketch of the mechanical assembly, with the optical readout of the Tile calorimeter [79]. The various components of the optical readout (the tiles, the fibres and the photomultipliers) are shown.	51
2.19	Representation of the muon system [90].	52
2.20	ATLAS magnet system [79].	52
2.21	ATLAS solenoid magnetic field map at $\phi = 20\pi/16$ [79].	53
3.1	Overall architecture of the LAr readout electronics, for the EM calorimeters [106]. The LAr detectors are positioned at the bottom, and the signal proceeds upwards, through the Front End crates, then the Back End system.	57
3.2	FEB architecture [107]. The data flow is depicted for four of the 128 channels per board.	57
3.3	LAr current pulse before and after the shaper [106]. The dots denote ideal positions of samples separated by 25 ns.	58

3.4	Total reconstruction efficiency for simulated electrons as a function of the true generated energy [111]. The efficiencies for the different steps of the reconstruction are depicted, in red for the seed cluster. Since the cluster reconstruction needs uncalibrated cluster seeds with $E_T > 2.5$ GeV, the total reconstruction efficiency is less than 65 % below 4.5 GeV (red dashed line). The seed cluster efficiency ranges from 65 % at $E_T = 4.5$ GeV, to 96 % at $E_T = 7$ GeV, to more than 99 % above $E_T = 15$ GeV.	61
3.5	Representation of a supercluster showing a seed electron cluster and a satellite photon cluster [114].	62
3.6	(a) f_{EM} distribution and (b) reconstruction efficiency as a function of the f_{EM} selection cut for simulated true electron (black) and pile-up (red) clusters [115].	63
3.7	A schematic illustration of the path of an electron through the detector [111]. The red line shows the hypothetical path of an electron passing through the tracking system (pixel detectors, SCT, and TRT) and then entering the electromagnetic calorimeter. The dashed red line shows the trajectory of a photon produced by the interaction of the electron with the tracking system material.	64
3.8	(a) Converted photon reconstruction efficiency as a function of E_T^{true} , averaged over $\langle\mu\rangle$ for a uniform $\langle\mu\rangle$ distribution between 0 and 60, (b) converted photon reconstruction efficiency as a function of $\langle\mu\rangle$, and (c) probability for an unconverted photon to be mistakenly reconstructed as a converted photon [115]. The contributions of the different conversion types are shown in each figure.	66
3.9	Diagram of the superclustering algorithm for photons and electrons [114]. Seed clusters are depicted in red and satellite clusters in blue.	67
3.10	Ambiguity resolution for particles initially reconstructed both as electrons and photons. An “innermost hit” is a hit in the functioning pixel nearest to the beam-line along the track trajectory, E/p is the ratio of the supercluster energy to the measured momentum of the matched track, R_{conv} is the radial position of the conversion vertex, and $R_{firstHit}$ is the smallest radial position of a hit in the track or tracks that make a conversion vertex [115].	68
3.11	The cluster, track, cluster and track, and electron reconstruction efficiency, against the generated electron transverse energy E_T [115].	68
3.12	Overview of the electron and photon energy calibration [123].	69
3.13	(a) Relative calibration scale factor α_{12} between the first and second layers of the EM calorimeter as a function of $ \eta $ [123]. Open blue squares represent results derived from the study of muon energy deposits in $Z \rightarrow \mu\mu$ events, combining the <i>truncated mean</i> and <i>most probable value</i> methods. Open red circles correspond to the combined values extracted from the dependence of the dielectron invariant mass m_{ee} and the E/p ratio on E_1/E_2 in $Z \rightarrow ee$ events. The final scale factors, obtained by combining the results from both electron and muon studies, are depicted as solid black lines. (b) Corresponding total uncertainties, taking into account rescaling in the case of large χ^2	71

3.14	(a) Energy scale calibration factors α_i^{acc} and $\alpha_i^{\text{closure}}$, and (b) the additional constant term c_i , as a function of η [123]. The statistical uncertainties are depicted with the shaded areas. The bottom panels show the differences between (a) α_i^{acc} and (b) c_i measured in a given data-taking period and the measurement using 2018 only.	72
3.15	Comparison of the invariant mass distributions of electron pairs from selected $Z \rightarrow ee$ candidates in data and simulation, following the application of calibration and resolution corrections [123]. The simulation is normalised to match the total event count in the data. The bottom panel displays the data-to-simulation ratio, with the uncertainty band reflecting the effects of uncertainties in the calibration and resolution correction factors.	72
3.16	The difference between the leakage fractions in data and simulation, in (a) the barrel and (b) the end-caps as a function of E_T , for electrons, converted photons and unconverted photons [123]. A variable bin size is used, indicated by the dotted lines. The last bin corresponds to the range $E_T > 40$ GeV. The markers are placed near the average E_T value of the corresponding bin.	74
3.17	Residual energy scale as a function of (a) η and (b) E_T , as measured with $J/\psi \rightarrow ee$ events [123]. The data points and uncertainty bands are shown for both the pre- and post-linearity-fit energy scale models. The uncertainty bands correspond to the energy calibration uncertainty for the energy range of the $J/\psi \rightarrow ee$ decays.	75
3.18	Variation of the residual energy scale for (a, b) unconverted and (c, d) converted photons as a function of (a, c) $ \eta $ and (b, d) E_T , as measured with $Z \rightarrow \ell\ell\gamma$ events [123]. The data points and uncertainty bands are shown for both the pre- and post-linearity-fit energy scale models. The uncertainty bands correspond to the energy calibration uncertainty for photons from $Z \rightarrow \ell\ell\gamma$ decays.	76
3.19	Efficiency of Tight identification and data-to-MC ratio for unconverted photons, preselected with a Loose isolation requirement, as a function of E_T , for (a) $0.0 < \eta < 0.6$, (b) $0.6 < \eta < 1.37$, (c) $1.52 < \eta < 1.81$ and (d) $1.81 < \eta < 2.37$ [115]. The combined scale factor, derived as a weighted average of the individual scale factors from different measurement methods, is also presented. The band represents the total uncertainty. . . .	80
3.20	Efficiency of Tight identification and data-to-MC ratio for converted photons, preselected with a Loose isolation requirement, as a function of E_T , for (a) $0.0 < \eta < 0.6$, (b) $0.6 < \eta < 1.37$, (c) $1.52 < \eta < 1.81$ and (d) $1.81 < \eta < 2.37$ [115]. The combined scale factor, derived as a weighted average of the individual scale factors from different measurement methods, is also presented. The band represents the total uncertainty.	81

4.1	Schema of the calorimetric isolation energy variable [128]. The grid represents the second layer of the calorimeter in the $\eta - \phi$ plane. The two blue circles are the two cones of 0.2 and 0.4 opening used to compute the isolation energy. All topo-clusters, represented in orange, whose barycentre falls within the cone, are added in the calculation. The photon candidate is at the centre, with its hottest cell denoted with a red square. The 5×7 fixed-size mask (in EM-middle-layer units) used to subtract the photon energy is represented, with energy leaking outside. The two dotted circles represent the first and second Molière radius.	85
4.2	Fit parameters $f_{i_{\eta\text{bin}}}$ and $\zeta(\eta)$ for E_T^{cone20} and E_T^{cone40} versus $ \eta $ [132]. The statistical uncertainties on the fit parameters $f_{i_{\eta\text{bin}}}$ and the propagated uncertainties on $\zeta(\eta)$ are shown. The range where the reference value of $\rho_{\text{median}}^{\text{central}}$ is measured on an event-by-event basis is shown as a dashed line.	86
4.3	Schema of the track isolation variable. All good tracks (brown) within the cone (blue) centred on the photon candidate are selected. The dotted line shows a track outside the cone, which is not selected. The p_T of selected tracks is summed to compute the track isolation variable, and the core energy is subtracted. Image adapted from Ref. [128].	87
4.4	Isolation energy distributions for data and MC, with their fit, for a “toy” dataset, to illustrate the concept of the calculation. The data-driven shift is the difference in the peak positions of the distributions of Tight Data and Tight MC.	88
4.5	Efficiencies of HLT_g120_loose and HLT_g140_loose triggers with respect to photons passing Tight identification criteria as a function of the photon transverse energy, excluding the crack region ($1.37 < \eta < 1.52$) [133]. The efficiency is calculated using events recorded with a L1 trigger requiring an EM cluster with $E_T > 7$ GeV, and no background subtraction is applied.	90
4.6	An ACB with $N = 1000, \mu_P = 0 \text{ GeV}, \sigma_L = 2 \text{ GeV}, \sigma_R = 3 \text{ GeV}, \alpha = 1.3, n = 10$	92
4.7	p_T distribution for 2018 data, Tight photons, passing HLT_g120_loose only. The turn-on is visible between 120 GeV and 125 GeV. A “turn-down” is also visible above 135 GeV, corresponding to events firing HLT_g140_loose.	92
4.8	E_T^γ spectrum of inclusive photons for the full Run 2 dataset. Discontinuities in the distribution arise from trigger prescales, while smooth transitions are observed in the turn-on regions corresponding to trigger thresholds. The p_T thresholds of the triggers are listed in Table 4.2.	93
4.9	Fits to MC simulated calorimetric isolation energy distributions, with a cone size of $\Delta R = 0.4$, for $1.81 < \eta < 2.37$ and unconverted photons, shown for (a) Tight direct, (b) Loose’4 direct, (c) Tight Bremsstrahlung and (d) Loose’4 Bremsstrahlung photons. The p_T and $\langle \mu \rangle$ bins are indicated on the plots.	94
4.10	Fits to MC simulated calorimetric isolation energy distributions for a cone size of $\Delta R = 0.4$, in the region $1.81 < \eta < 2.37$, for converted photons. Results are shown for (a) Tight direct, (b) Loose’4 direct, (c) Tight Bremsstrahlung and (d) Loose’4 Bremsstrahlung photons. The p_T and $\langle \mu \rangle$ bins are indicated on the plots.	95

4.11	Fits to σ_R against p_T and $\langle\mu\rangle$, for Tight direct converted photons, for (a) $0 < \eta < 0.60$ and (c) $1.81 < \eta < 2.37$, along with (b), (d) their pull distributions. An isolation cone size of $\Delta R = 0.4$ was used. Since the widths of the pull distributions are slightly larger than one, the uncertainties on σ_R can be increased by approximately 27 % and 37 %, respectively, following the procedure described in Ref. [3].	96
4.12	Fits to σ_L against p_T and $\langle\mu\rangle$, for Tight direct converted photons, for (a) $0 < \eta < 0.60$ and (c) $1.81 < \eta < 2.37$, along with (b), (d) their pull distributions. An isolation cone size of $\Delta R = 0.4$ was used. Since the widths of the pull distributions are slightly larger than one, the uncertainties on σ_L can be increased by approximately 32 % and 35 %, respectively, following the procedure described in Ref. [3].	97
4.13	Fits to μ_P against p_T and $\langle\mu\rangle$, for $0 < \eta < 0.60$ and unconverted photons, for (a) Tight direct, (b) Loose'4 direct, (c) Tight Bremsstrahlung and (d) Loose'4 Bremsstrahlung photons.	98
4.14	Fit to $\frac{\sigma_{\text{direct}}^{\text{filter}} \cdot \epsilon_{\text{direct}}^{\text{filter}}}{\sigma_{\text{Brem}} \cdot \epsilon_{\text{Brem}}}$, with $f(p_T) = a \times \left(\frac{p_T}{\text{GeV}}\right)^b$. Fitted parameters: $a = 0.149 \pm 0.015$ and $b = 0.411 \pm 0.017$. $\chi^2/\text{ndf} = 4.59/9$	99
4.15	(a) $c_{\text{direct}}^{\text{Tight}}$ and (b) $c_{\text{direct}}^{\text{Loose'4}}$ fits against p_T and $\langle\mu\rangle$ for unconverted photons, $0 < \eta < 0.6$	100
4.16	Fit to $N_{\text{Tight}}/(N_{\text{Tight}} + N_{\text{Loose'4}})$ for unconverted photons, and $0 < \eta < 0.6$, with a series expansion.	101
4.17	Fits to data for unconverted photons, $0 < \eta < 0.6$, for several p_T and $\langle\mu\rangle$ bins. The signal (blue line) and the background (red line) components are shown.	102
4.18	(a) Fit to data-driven shift against p_T and $\langle\mu\rangle$ and (b) pull distribution, for unconverted photons in $ \eta < 0.6$	103
4.19	Former data-driven shifts [128].	103
5.1	Fully connected neural network representation.	107
5.2	Representation of a perceptron.	108
5.3	Usual activation functions, including (a) the Heavyside step function, (b) the sigmoid function, (c) the hyperbolic tangent, (d) the Rectified Linear Unit, (e), the leaky Rectified Linear Unit (with $\alpha = 0.1$), and (f) the Exponential Linear Unit (with $\alpha = 1$).	109
5.4	A CNN representation, from Ref. [145].	115
5.5	A convolution example, from Ref. [145].	115
5.6	Representation of a convolutional layer with a 3D kernel, from Ref. [145].	116
5.7	(a) Stride representation and (b) a convolutional kernel sliding over an image with a stride of 2×2 . Images from Ref. [145].	116
5.8	(a) Padding representation and (b) a convolutional kernel sliding over an image with a padding of 1. Images from Ref. [145].	116
5.9	An oriented and weighted graph.	118
5.10	Transformer architecture, from Ref. [149].	120
5.11	(a) Scaled dot-product attention, and (b) multi-head attention, consisting of several attention layers running in parallel. Images from Ref. [149].	122
5.12	Diagrams of the attention-based set operations of the Set Transformer, from Ref. [151].	123

6.1	Selection efficiency in a search for boosted diphoton resonances in the 10 to 70 GeV mass range, using the full Run 2 dataset, with the ATLAS detector [156]. Events with at least two photon candidates, with transverse energies above 22 GeV, $ \eta < 2.37$, and satisfying the Tight identification criteria were selected. The correction factor, C_X , is defined as the ratio between the number of reconstructed signal events passing the analysis cuts and the number of signal events at the particle level generated within the fiducial volume, and the acceptance correction factor, A_X , is defined as the fraction of diphoton resonances satisfying the fiducial acceptance at the particle level. The diphoton selection efficiency is the product of these two quantities. The fiducial volume is defined by requiring two photons at particle level with $E_T > 22$ GeV, $ \eta < 2.37$ and $p_T^{\gamma\gamma} > 50$ GeV.	129
6.2	Distributions at generator level from the prompt diphoton sample, generated with ParticleGun: (a) leading vs subleading photon p_T and (b) diphoton angular separation $\Delta R_{\gamma\gamma}$	131
6.3	$\Delta\eta_{\gamma\gamma}$ vs $\Delta\phi_{\gamma\gamma}$ for 2022 to 2024 data, showing two clusters of events at $\Delta\phi \sim \pm 0.1$	132
6.4	Mean shower shape variables as a function of the diphoton angular separation, shown for the leading photon, separately for prompt and fake diphoton events.	134
6.5	Mean shower shape variables as a function of the diphoton angular separation, shown for the subleading photon, separately for prompt and fake diphoton events.	135
6.6	Correlations among pairs of shower shape variables as a function of the diphoton angular separation, shown separately for prompt and fake diphotons.	136
6.7	$\Delta R_{\gamma\gamma}$ distributions, for (a) simulated prompt diphotons and (b) simulated dijets, before and after reweighting.	136
6.8	Reweighted and normalised distributions of f_1 and w_{s3} for the leading (left) and the subleading (right) photons.	137
6.9	Invariant mass distributions, for (a) simulated prompt diphotons and (b) simulated dijets, before and after reweighting. The first two bins in the prompt diphoton distribution are due to the large weights applied to low- $\Delta R_{\gamma\gamma}$ events (see Figure 6.7).	137
6.10	Training and validation losses vs. epochs for the best training run.	139
6.11	Distribution of predictions for prompt and fake diphotons.	139
6.12	Distributions of predictions for prompt and fake diphotons per $\Delta R_{\gamma\gamma}$ range.	140
6.13	Distributions of predictions for prompt and fake diphotons per $m_{\gamma\gamma}$ range.	141
6.14	ROC curve. The red dot indicates the threshold position (0.5).	141
6.15	Efficiency per bin of (a) $\Delta R_{\gamma\gamma}$ and (b) $m_{\gamma\gamma}$	142
6.16	Distributions of (a) λ and (b) E_T , for simulated prompt diphotons and fake diphotons (from data).	143

6.17	Example of an input image used in the CNN for an event from the fake-enriched data sample, with $\langle\mu\rangle = 40.6$. The grid represents 32×32 cells of the second layer of the calorimeter, centred on the photon pair candidate, shown as two red dots with a red cross at their geometric centre. Clusters are selected within $0.15 < \Delta R < 0.4$ from the diphoton centre. These limits are indicated by red dotted circles. Coloured cells contain a cluster, with the colour representing the normalised cluster transverse energy. Similar images are generated for the other input channels—cluster depth and pile-up.	144
6.18	Architecture of the CNN used for diphoton isolation. The inputs have three channels: the cluster's transverse energy E_T , depth λ and the mean number of interactions $\langle\mu\rangle$. The cluster relative position with respect to the diphoton, $(\eta_{\text{cl-}\gamma\gamma}, \phi_{\text{cl-}\gamma\gamma})$, is encoded in the image pixels: each cluster is placed in the pixel corresponding to its position relative to the diphoton, which is located at the centre of the image. The output dimensions of the layers are indicated in brackets. The convolutional layers use a kernel of size 3×3 with padding of 1. The pooling operations use a kernel of size 2×2 with a stride of 2.	145
6.19	Training and validation losses vs. epochs for the best training run, with the CNN.	146
6.20	Distribution of CNN predictions for prompt and fake diphotons.	146
6.21	Distribution of CNN predictions for prompt and fake diphotons per $\Delta R_{\gamma\gamma}$ range.	147
6.22	Distributions of CNN predictions for prompt and fake diphotons per $m_{\gamma\gamma}$ range.	148
6.23	ROC curve for the CNN. The red dot indicates the threshold position (0.25).	149
6.24	Efficiency per bin of (a) $\Delta R_{\gamma\gamma}$ and (b) $m_{\gamma\gamma}$, with the CNN.	149
6.25	Example of a graph input used in the GNN, corresponding to the same event as in Figure 6.17. The nodes represent clusters, and the colour indicates the cluster normalised E_T . In this example, each node is connected to its five nearest neighbours, and no weights are applied to the edges, but other configurations of connexions were tested. Similar representations can be produced for the other cluster variables.	150
6.26	(a) Structure of the EdgeConv block. (b) Architecture of the GNN, based on ParticleNet. Images adapted from Ref. [161].	151
6.27	Training and validation losses vs. epochs for the best training run with the GNN.	152
6.28	Distribution of GNN predictions for prompt and fake diphotons.	152
6.29	Distribution of GNN predictions for prompt and fake diphotons per $\Delta R_{\gamma\gamma}$ range.	153
6.30	Distributions of GNN predictions for prompt and fake diphotons per $m_{\gamma\gamma}$ range.	154
6.31	ROC curves for the CNN and the GNN. The dots indicates the threshold positions, 0.25 for the CNN and 0.5 for the GNN.	155
6.32	Efficiency per bin of (a) $\Delta R_{\gamma\gamma}$ and (b) $m_{\gamma\gamma}$ with the GNN.	155
6.33	Set Transformer encoder.	156
6.34	Set Transformer decoder.	156
6.35	Set Transformer final sigmoid function.	156

6.36	Training and validation losses vs. epochs for the best training run with the Set Transformer.	157
6.37	Distribution of Set Transformer predictions for prompt and fake diphotons.	157
6.38	Distribution of Set Transformer predictions for prompt and fake diphotons per $\Delta R_{\gamma\gamma}$ range.	158
6.39	Distributions of Set Transformer predictions for prompt and fake diphotons per $m_{\gamma\gamma}$ range.	159
6.40	ROC curves for the CNN, the GNN and the Set Transformer. The dot indicates the thresholds positions (0.25 for the CNN, 0.5 for the GNN and 0.4 for the Set Transformer).	160
6.41	Efficiency per bin of (a) $\Delta R_{\gamma\gamma}$ and (b) $m_{\gamma\gamma}$ with the Set Transformer.	160
7.1	Mass distributions of diphoton events passing the selections, originating from radiative Z and W^\pm decays.	162
7.2	Fits of the $p_{T,\min}^{\text{cone30}}$ distributions in the PG simulated prompt diphoton sample, shown in bins of $m_{\gamma\gamma}$. The exponential coefficients and the fraction of the first exponential in the total model are shown in the plots.	164
7.3	$p_{T,\min}^{\text{cone30}}$ distributions by mass bin in the JZ samples (fake diphotons), shown in separate slices. JZ3, JZ4 and JZ5 correspond to different simulated dijet samples, each covering a distinct range of jet transverse momentum p_T^{jet} and filtered by parton-level momentum p_T^{hat} (see Section 6.2.2). Only slices with a non-negligible number of events are shown. For JZ3, $p_T^{\text{hat}} > 50$ GeV and $160 < p_T^{\text{jet}} < 400$ GeV; for JZ4, $p_T^{\text{hat}} > 150$ GeV and $400 < p_T^{\text{jet}} < 800$ GeV; for JZ5, $p_T^{\text{hat}} > 350$ GeV and $800 < p_T^{\text{jet}} < 1300$ GeV.	166
7.4	Fit to $p_{T,\min}^{\text{cone30}}$ on data, with an exponential function and a constant: $f(p_{T,\min}^{\text{cone30}}) = A \exp[c(p_{T,\min}^{\text{cone30}} - 1 \text{ GeV})] + b$. The dashed band hides the isolated events in the signal region.	167
7.5	Estimated prompt and fake diphoton yields in the signal region, for an integrated luminosity of 160.9 fb^{-1} . The distributions are shown unstacked, and the total distribution corresponds to the sum of the prompt and fake diphoton contributions.	168
7.6	Estimated prompt and fake diphoton yields in the signal region, after applying diphoton identification and isolation efficiencies, for an integrated luminosity of 160.9 fb^{-1} . The fit parameters and the χ^2 divided by the number of degrees of freedom are indicated.	168
7.7	p_T of the subleading photon in the PG-simulated prompt diphoton sample and in isolated data. Isolated events in data refer to those in the signal region (with $p_{T,\min}^{\text{cone30}} = 0$) and not to diphoton isolation (which has not been applied here).	169
7.8	Fits to the mass resolution using a Gaussian, over a range centred on the peak position: $[\mu_p - 1.5\sigma, \mu_p + 1.5\sigma]$	170
7.9	Fit to the mass resolution as a function of the mass. The fit parameters and the χ^2 divided by the number of degrees of freedom are indicated.	170
7.10	Background toy datasets, along with their fits. The signal shape is fixed to the resonance mass. The shape and number of background events are fixed to the estimation using the control region.	171

7.11	Effective background as a function of the resonance mass for an integrated luminosity of 160.9 fb^{-1} . At this stage, one input is still missing to allow for an exclusion calculation in terms of the KSVZ prediction. This parameter is essential to interpret the results within that theoretical framework fully. It is expected that this missing information will become available soon, allowing the analysis to be completed accordingly.	172
7.12	Efficiencies of the different selections, considered independently, and the total efficiency (C_X factor) with its fit, computed using the PG simulated prompt-diphoton sample.	173
7.13	Expected upper limit at 95 % confidence level on the fiducial cross section times the branching ratio of a hypothetical diphoton resonance, for an integrated luminosity of 160.9 fb^{-1} . The expected upper limit is shown as a dashed black line, with its ± 1 and ± 2 standard deviation bands.	174
8.1	The causal chain from emissions to the resulting warming of the climate system. a) Global net anthropogenic GHG emissions. b) Concentrations of GHGs in the atmosphere. c) Changes in global surface temperature, shown as annual anomalies from a 1850–1900 baseline. d) Origins of temperature change. Figure from Ref. [165].	177
8.2	Assessment of observed changes in large-scale indicators of mean climate across climate system components, and their attribution to human influence. The colour indicates the assessed confidence in the observed change and the human contribution as a driver or main driver. Figure from Ref. [165].	178
8.3	Global GHG emissions under modelled pathways and projected outcomes from near-term policy assessments for 2030. a) Global GHG emissions over 2015-2050 for four types of assessed modelled global pathways. b) Snapshot of the GHG emission ranges of the modelled pathways in 2030 and projected emissions outcomes from near-term policy assessments in 2030. Figure from Ref. [165].	180
8.4	A) Box plot of laboratory emissions per capita per emission source. The box extends from the 1st to 3rd quartile (Q1, Q3), with a line at the median, the whiskers extend to $1.5 \times (Q3 - Q1)$, and the circles represent all data outside the whiskers. Electricity emissions are calculated for different energy mix: French mix (black box plot), world mix (median as dashed red line), high-carbon mix (median as dotted blue line). B) Distribution of purchases emissions per capita. The inset shows the same data in log scale. Figure from Ref. [170].	181
8.5	A) Purchases module emissions vs. budget for all GHG laboratory footprints in the GES 1point5 lab emissions database. Error bars correspond to one standard deviation. B) Histogram of purchases module carbon intensities for different scientific domains. HSS: Human and Social Sciences, LHS: Life and Health Sciences, ST: Science and Technology. Support laboratories: large experimental platforms providing analysis services. Figure from Ref. [170].	181

8.6	Histogram of travelled distance (right) and associated GHG emissions (left) for car (blue), train (green), and plane (red) travel, based on 137 081 academic trips from 159 research units in France in 2019 (Labos 1point5 database). Box plots extend from the 1st to the 3rd quartile (Q1 to Q3), and whiskers to $1.5 \times (Q3 - Q1)$. The zoom focuses on car and train travel for distances below 1500 km. Bold purple lines show the cumulative percentages of the number of trips (right) and GHG emissions (left). Figure from Ref. [172].	182
8.7	Average distance travelled by plane in 2019 by discipline of respondents. Figure from Ref. [168].	183
8.8	Average distance travelled by plane in 2019 by status of respondents. Figure from Ref. [168].	183
8.9	CERN's total scope 1, 2 and 3 2021-2022 emissions. Figure from Ref. [173].	184
8.10	CERN scope 1 emissions for 2017-2022 by category. Figure from Ref. [173].	185
8.11	CERN scope 2 emissions for 2017-2022. Figure from Ref. [173].	185
8.12	CERN scope 3 emissions for 2019-2022 (excluding procurement). Figure from Ref. [173].	186
8.13	CERN emissions by procurement category for 2021-2022. Figure from Ref. [173].	186
8.14	LPNHE carbon footprint for 2019, 2022, 2023 and 2024. Figure from Labos 1point5.	187
8.15	Carbon footprint of LPNHE research activities by type (2023). Figure from Labos 1point5.	188
8.16	Carbon footprint of LPNHE purchases by category (2023). Figure from Labos 1point5.	188
8.17	Carbon footprint of LPNHE travels (2023). Figure from Labos 1point5.	189
8.18	Carbon footprint of LPNHE commuting (2023). Figure from Labos 1point5.	189
8.19	Carbon footprint of LPNHE catering (2023). Figure from Labos 1point5.	190
8.20	Typical carbon footprint of different meals, based on French carbon data sets, by ADEME. Figure from ADEME [176].	191
8.21	Average carbon footprint in France in 2019. Figure from Carbon4 [177].	191
8.22	Willingness to reduce GHG emissions by 2030 in various areas. Figure from Ref. [171].	192
8.23	Support to institutional actions to reduce GHG emissions. Figure from Ref. [171].	193
8.24	Percentage of total (plane, train, car) travel-induced GHG reductions from single or combined options. Colours represent the intensity in GHG mitigation potential (yellow corresponds to higher potential). Figure from Ref. [172].	194

List of Tables

2.1	η coverage of the ATLAS sub-detectors.	40
2.2	Lead and Liquid-Argon properties [95, 97].	46
3.1	Summary of the cells included in the cluster. Column 1 gives the order the layers are processed. Columns 3 and 4 give the size $\Delta\eta_{\text{cl}} \times \Delta\phi_{\text{cl}}$ of the window centred on the position given in column 5, in which the cells are included in the cluster.	60
3.2	Cluster size $N_{\eta}^{\text{cluster}} \times N_{\phi}^{\text{cluster}}$ for different hypothesized particles and regions of the detector.	60
3.3	Shower-shape variables used for photon identification.	78
3.4	Working points definitions. Loose'4 uses the same shower-shape variables as the ones used by the Tight working point, except for w_{s3} , f_{side} , ΔE_s , and E_{ratio}	79
4.1	Definition of the photon isolation working points.	88
4.2	List of single photon triggers, along with their p_{T} thresholds.	89
4.3	Binning used to fit the calorimetric isolation energy.	93
6.1	Selection criteria applied to diphoton candidates in the simulated prompt diphoton sample.	132
6.2	Selection criteria applied to diphoton candidates in the simulated dijet sample.	133
6.3	Selection criteria applied to diphoton candidates in the data sample.	133
6.4	Diphoton identification model characteristics.	138
6.5	Variables used by the diphoton isolation neural networks.	143
6.6	Hyperparameters of the CNN.	145
6.7	Hyperparameters of the GNN.	151
6.8	Architecture details of the Set Transformer.	156
6.9	Hyperparameters of the Set Transformer.	157
7.1	Fraction of prompt diphotons with $1 < p_{\text{T,min}}^{\text{cone30}} < 10 \text{ GeV}$ relative to the total number of prompt diphotons with $p_{\text{T,min}}^{\text{cone30}} < 10 \text{ GeV}$, computed on the PG sample.	163
7.2	Particle-level selections defining the fiducial volume.	172
7.3	Reconstruction-level selections used to compute C_X	173

Bibliography

- [1] Matthew D Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge University Press, 2014 (cit. on pp. 6, 13).
- [2] David Tong. *Lectures on Quantum Field Theory*. <https://www.damtp.cam.ac.uk/user/tong/qft.html>. Accessed 03/05/25. 2006 (cit. on p. 6).
- [3] S. Navas et al. “Review of Particle Physics”. In: *Phys. Rev. D* 110 (3 Aug. 2024), p. 030001. DOI: [10.1103/PhysRevD.110.030001](https://doi.org/10.1103/PhysRevD.110.030001). URL: <https://link.aps.org/doi/10.1103/PhysRevD.110.030001> (cit. on pp. 13, 27, 96, 97).
- [4] David Tong. *Lectures on the Standard Model*. <http://www.damtp.cam.ac.uk/user/tong/standardmodel.html>. Accessed 03/05/25. 2019 (cit. on p. 13).
- [5] R. P. Feynman. “The behavior of hadron collisions at extreme energies”. In: *Conf. Proc. C* 690905 (1969), pp. 237–258 (cit. on p. 19).
- [6] 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](https://doi.org/10.1103/PhysRev.185.1975). URL: <https://link.aps.org/doi/10.1103/PhysRev.185.1975> (cit. on p. 19).
- [7] Jon Butterworth et al. “PDF4LHC recommendations for LHC Run II”. In: *Journal of Physics G: Nuclear and Particle Physics* 43.2 (Feb. 2016), p. 023001. ISSN: 0954-3899, 1361-6471. DOI: [10.1088/0954-3899/43/2/023001](https://doi.org/10.1088/0954-3899/43/2/023001) (cit. on p. 19).
- [8] Richard D Ball et al. “The PDF4LHC21 combination of global PDF fits for the LHC Run III*”. en. In: *Journal of Physics G: Nuclear and Particle Physics* 49.8 (Aug. 2022), p. 080501. ISSN: 0954-3899, 1361-6471. DOI: [10.1088/1361-6471/ac7216](https://doi.org/10.1088/1361-6471/ac7216) (cit. on p. 19).
- [9] 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).
- [10] 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).
- [11] 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](https://doi.org/10.1016/0550-3213(77)90384-4) (cit. on p. 19).
- [12] Richard D. Ball et al. “Parton distributions from high-precision collider data: NNPDF Collaboration”. In: *The European Physical Journal C* 77.10 (Oct. 2017). ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-017-5199-5](https://doi.org/10.1140/epjc/s10052-017-5199-5). URL: <http://dx.doi.org/10.1140/epjc/s10052-017-5199-5> (cit. on p. 20).

- [13] John C. Collins, Davison E. Soper, and George F. Sterman. “Factorization of Hard Processes in QCD”. In: *Adv. Ser. Direct. High Energy Phys.* 5 (1989), pp. 1–91. DOI: [10.1142/9789814503266_0001](https://doi.org/10.1142/9789814503266_0001). arXiv: [hep-ph/0409313](https://arxiv.org/abs/hep-ph/0409313) (cit. on p. 19).
- [14] 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 (Jan. 2007), pp. 89–193. ISSN: 0034-4885, 1361-6633. DOI: [10.1088/0034-4885/70/1/R02](https://doi.org/10.1088/0034-4885/70/1/R02) (cit. on p. 19).
- [15] Stefano Catani et al. “Cross section of isolated prompt photons in hadron-hadron collisions”. In: *Journal of High Energy Physics* 2002.05 (May 2002), pp. 028–028. ISSN: 1029-8479. DOI: [10.1088/1126-6708/2002/05/028](https://doi.org/10.1088/1126-6708/2002/05/028). URL: <http://dx.doi.org/10.1088/1126-6708/2002/05/028> (cit. on p. 21).
- [16] *Measurement of the production cross-section of isolated photon pairs in pp collisions at 13 TeV with the ATLAS detector*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2020-024>. Geneva: CERN, 2020. URL: <http://cds.cern.ch/record/2725552> (cit. on p. 22).
- [17] ATLAS Collaboration. *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. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2019-029>. Geneva: CERN, 2019. URL: <http://cds.cern.ch/record/2682800> (cit. on p. 22).
- [18] S. Frixione. “Isolated photons in perturbative QCD”. en. In: *Physics Letters B* 429.3–4 (June 1998). arXiv:hep-ph/9801442, pp. 369–374. ISSN: 03702693. DOI: [10.1016/S0370-2693\(98\)00454-7](https://doi.org/10.1016/S0370-2693(98)00454-7) (cit. on p. 22).
- [19] Eleanor Hall and Jesse Thaler. “Photon isolation and jet substructure”. en. In: *Journal of High Energy Physics* 2018.9 (Sept. 2018). arXiv:1805.11622 [hep-ph], p. 164. ISSN: 1029-8479. DOI: [10.1007/JHEP09\(2018\)164](https://doi.org/10.1007/JHEP09(2018)164) (cit. on p. 22).
- [20] E. W. N. Glover and A. G. Morgan. “Measuring the photon fragmentation function at LEP”. en. In: *Zeitschrift für Physik C Particles and Fields* 62.2 (June 1994), pp. 311–321. ISSN: 0170-9739, 1434-6052. DOI: [10.1007/BF01560245](https://doi.org/10.1007/BF01560245) (cit. on p. 22).
- [21] Stefano Catani et al. “Diphoton Production at Hadron Colliders: A Fully Differential QCD Calculation at Next-to-Next-to-Leading Order”. In: *Physical Review Letters* 108.7 (Feb. 2012). ISSN: 1079-7114. DOI: [10.1103/physrevlett.108.072001](https://doi.org/10.1103/physrevlett.108.072001). URL: <http://dx.doi.org/10.1103/PhysRevLett.108.072001> (cit. on p. 22).
- [22] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. “The anti-ktjet clustering algorithm”. In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. ISSN: 1029-8479. DOI: [10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063). URL: <http://dx.doi.org/10.1088/1126-6708/2008/04/063> (cit. on p. 22).
- [23] *Standard Model Summary Plots June 2024*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2024-011>. Geneva: CERN, 2024. URL: <https://cds.cern.ch/record/2903866> (cit. on p. 22).

- [24] ATLAS Collaboration. “Measurement of the production cross section of pairs of isolated photons in pp collisions at 13 TeV with the ATLAS detector”. In: *Journal of High Energy Physics* 2021.11 (Nov. 2021). ISSN: 1029-8479. DOI: [10.1007/JHEP11\(2021\)169](https://doi.org/10.1007/JHEP11(2021)169). URL: [http://dx.doi.org/10.1007/JHEP11\(2021\)169](http://dx.doi.org/10.1007/JHEP11(2021)169) (cit. on p. 23).
- [25] Jonathan L. Rosner. “CP violation—a brief review”. In: *AIP Conference Proceedings*. Vol. 540. AIP, 2000, pp. 283–304. DOI: [10.1063/1.1328890](https://doi.org/10.1063/1.1328890). URL: <http://dx.doi.org/10.1063/1.1328890> (cit. on p. 23).
- [26] M. Ahmadvand. “The Higgs boson as a self-similar system: Towards a new solution to the hierarchy problem”. en. In: arXiv:2204.09916 (May 2024). arXiv:2204.09916 [hep-ph]. DOI: [10.48550/arXiv.2204.09916](https://doi.org/10.48550/arXiv.2204.09916). URL: <http://arxiv.org/abs/2204.09916> (cit. on p. 23).
- [27] V. C. Rubin, N. Thonnard, and Jr Ford W. K. “Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc”. eng. In: *The Astrophysical journal* 238 (1980), pp. 471–. ISSN: 0004-637X (cit. on p. 24).
- [28] 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), p. 728. DOI: [10.1086/517498](https://doi.org/10.1086/517498) (cit. on p. 24).
- [29] S. Perlmutter et al. “Measurements of Ω and Λ from 42 High-Redshift Supernovae”. In: *The Astrophysical Journal* 517.2 (June 1999), pp. 565–586. ISSN: 1538-4357. DOI: [10.1086/307221](https://doi.org/10.1086/307221). URL: <http://dx.doi.org/10.1086/307221> (cit. on p. 24).
- [30] Adam G. Riess et al. “The Rise Time of Nearby Type Ia Supernovae”. In: *The Astronomical Journal* 118.6 (Dec. 1999), pp. 2675–2688. ISSN: 0004-6256. DOI: [10.1086/301143](https://doi.org/10.1086/301143). URL: <http://dx.doi.org/10.1086/301143> (cit. on p. 24).
- [31] 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: *The Astrophysical Journal* 607.2 (June 2004), pp. 665–687. ISSN: 1538-4357. DOI: [10.1086/383612](https://doi.org/10.1086/383612). URL: <http://dx.doi.org/10.1086/383612> (cit. on p. 24).
- [32] R. D. Peccei and Helen R. Quinn. “CP Conservation in the Presence of Pseudoparticles”. en. In: *Physical Review Letters* 38.25 (June 1977), pp. 1440–1443. ISSN: 0031-9007. DOI: [10.1103/PhysRevLett.38.1440](https://doi.org/10.1103/PhysRevLett.38.1440) (cit. on p. 24).
- [33] A. Ringwald. “Axions and Axion-Like Particles”. en. In: arXiv:1407.0546 (July 2014). arXiv:1407.0546 [hep-ph]. DOI: [10.48550/arXiv.1407.0546](https://doi.org/10.48550/arXiv.1407.0546). URL: <http://arxiv.org/abs/1407.0546> (cit. on pp. 24, 25).
- [34] Alberto Mariotti et al. “New LHC bound on low-mass diphoton resonances”. en. In: *Physics Letters B* 783 (Aug. 2018). arXiv:1710.01743 [hep-ph], pp. 13–18. ISSN: 0370-2693. DOI: [10.1016/j.physletb.2018.06.039](https://doi.org/10.1016/j.physletb.2018.06.039) (cit. on pp. 25, 26).
- [35] John Preskill, Mark B. Wise, and Frank Wilczek. “Cosmology of the invisible axion”. In: *Physics Letters B* 120.1 (1983), pp. 127–132. ISSN: 0370-2693. DOI: [https://doi.org/10.1016/0370-2693\(83\)90637-8](https://doi.org/10.1016/0370-2693(83)90637-8) (cit. on p. 25).
- [36] Jihn E. Kim. “Weak-Interaction Singlet and Strong CP Invariance”. In: *Phys. Rev. Lett.* 43 (2 July 1979), pp. 103–107. DOI: [10.1103/PhysRevLett.43.103](https://doi.org/10.1103/PhysRevLett.43.103). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.43.103> (cit. on p. 25).

- [37] M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov. “Can confinement ensure natural CP invariance of strong interactions?” In: *Nuclear Physics B* 166.3 (1980), pp. 493–506. ISSN: 0550-3213. DOI: [https://doi.org/10.1016/0550-3213\(80\)90209-6](https://doi.org/10.1016/0550-3213(80)90209-6) (cit. on p. 25).
- [38] ATLAS Collaboration. “Search for boosted diphoton resonances in the 10 to 70 GeV mass range using 138 fb⁻¹ of 13 TeV *pp* collisions with the ATLAS detector”. en. In: *Journal of High Energy Physics* 2023.7 (July 2023). arXiv:2211.04172 [hep-ex], p. 155. ISSN: 1029-8479. DOI: [10.1007/JHEP07\(2023\)155](https://doi.org/10.1007/JHEP07(2023)155) (cit. on pp. 25, 27).
- [39] Anson Hook et al. “High Quality QCD Axion and the LHC”. In: *Phys. Rev. Lett.* 124 (22 June 2020), p. 221801. DOI: [10.1103/PhysRevLett.124.221801](https://doi.org/10.1103/PhysRevLett.124.221801). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.124.221801> (cit. on p. 26).
- [40] David I. Dunskey, Lawrence J. Hall, and Keisuke Harigaya. “A Heavy QCD Axion and the Mirror World”. In: *JHEP* 2402 (2024). 36 pages, 9 figures, p. 212. DOI: [10.1007/JHEP02\(2024\)212](https://doi.org/10.1007/JHEP02(2024)212). arXiv: [2302.04274](https://arxiv.org/abs/2302.04274). URL: <https://cds.cern.ch/record/2856521> (cit. on p. 26).
- [41] Raymond T. Co, Tony Gherghetta, and Keisuke Harigaya. “Axiogenesis with a heavy QCD axion”. In: *Journal of High Energy Physics* 2022.10 (Oct. 2022). ISSN: 1029-8479. DOI: [10.1007/jhep10\(2022\)121](https://doi.org/10.1007/jhep10(2022)121). URL: [http://dx.doi.org/10.1007/JHEP10\(2022\)121](http://dx.doi.org/10.1007/JHEP10(2022)121) (cit. on p. 26).
- [42] Ann E. Nelson and Nathan Seiberg. “R-symmetry breaking versus supersymmetry breaking”. In: *Nuclear Physics B* 416.1 (Mar. 1994). ISSN: 0550-3213. DOI: [10.1016/0550-3213\(94\)90577-0](https://doi.org/10.1016/0550-3213(94)90577-0). URL: [http://dx.doi.org/10.1016/0550-3213\(94\)90577-0](http://dx.doi.org/10.1016/0550-3213(94)90577-0) (cit. on p. 26).
- [43] Jonathan Bagger, Erich Poppitz, and Lisa Randall. “The R-axion from dynamical supersymmetry breaking”. In: *Nuclear Physics B* 426.1 (Sept. 1994), pp. 3–18. ISSN: 0550-3213. DOI: [10.1016/0550-3213\(94\)90123-6](https://doi.org/10.1016/0550-3213(94)90123-6). URL: [http://dx.doi.org/10.1016/0550-3213\(94\)90123-6](http://dx.doi.org/10.1016/0550-3213(94)90123-6) (cit. on p. 26).
- [44] James Unwin and Tom Yildirim. *A QCD R-Axion*. 2024. arXiv: [2407.17557](https://arxiv.org/abs/2407.17557) [hep-ph]. URL: <https://arxiv.org/abs/2407.17557> (cit. on p. 26).
- [45] Gabriele Ferretti. “Gauge theories of partial compositeness: scenarios for Run-II of the LHC”. In: *Journal of High Energy Physics* 2016.6 (June 2016). ISSN: 1029-8479. DOI: [10.1007/jhep06\(2016\)107](https://doi.org/10.1007/jhep06(2016)107). URL: [http://dx.doi.org/10.1007/JHEP06\(2016\)107](http://dx.doi.org/10.1007/JHEP06(2016)107) (cit. on p. 26).
- [46] Tony Gherghetta and Minh D. Nguyen. “A composite Higgs with a heavy composite axion”. In: *Journal of High Energy Physics* 2020.12 (Dec. 2020). ISSN: 1029-8479. DOI: [10.1007/jhep12\(2020\)094](https://doi.org/10.1007/jhep12(2020)094). URL: [http://dx.doi.org/10.1007/JHEP12\(2020\)094](http://dx.doi.org/10.1007/JHEP12(2020)094) (cit. on p. 26).
- [47] Jared A. Evans, Stefania Gori, and Jessie Shelton. “Looking for the WIMP next door”. In: *Journal of High Energy Physics* 2018.2 (Feb. 2018). ISSN: 1029-8479. DOI: [10.1007/jhep02\(2018\)100](https://doi.org/10.1007/jhep02(2018)100). URL: [http://dx.doi.org/10.1007/JHEP02\(2018\)100](http://dx.doi.org/10.1007/JHEP02(2018)100) (cit. on p. 26).
- [48] I.G. Irastorza et al. “Towards a new generation axion helioscope”. In: *Journal of Cosmology and Astroparticle Physics* 2011.06 (June 2011), p. 013. DOI: [10.1088/1475-7516/2011/06/013](https://doi.org/10.1088/1475-7516/2011/06/013). URL: <https://dx.doi.org/10.1088/1475-7516/2011/06/013> (cit. on p. 26).

- [49] Igor G. Irastorza and Javier Redondo. “New experimental approaches in the search for axion-like particles”. en. In: *Progress in Particle and Nuclear Physics* 102 (Sept. 2018). arXiv:1801.08127 [hep-ph], pp. 89–159. ISSN: 01466410. DOI: [10.1016/j.pnpnp.2018.05.003](https://doi.org/10.1016/j.pnpnp.2018.05.003) (cit. on p. 26).
- [50] 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](https://doi.org/10.1103/PhysRev.81.899). URL: <https://link.aps.org/doi/10.1103/PhysRev.81.899> (cit. on p. 26).
- [51] 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](https://doi.org/10.1103/PhysRevD.38.3375). URL: <https://link.aps.org/doi/10.1103/PhysRevD.38.3375> (cit. on p. 27).
- [52] R. Khatiwada et al. “Axion Dark Matter Experiment: Detailed design and operations”. In: *Review of Scientific Instruments* 92.12 (Dec. 2021). ISSN: 1089-7623. DOI: [10.1063/5.0037857](https://doi.org/10.1063/5.0037857). URL: <http://dx.doi.org/10.1063/5.0037857> (cit. on p. 27).
- [53] P. Brun et al. “A new experimental approach to probe QCD axion dark matter in the mass range above 40 μeV ”. In: *The European Physical Journal C* 79.3 (Mar. 2019). ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-019-6683-x](https://doi.org/10.1140/epjc/s10052-019-6683-x). URL: <http://dx.doi.org/10.1140/epjc/s10052-019-6683-x> (cit. on p. 27).
- [54] O. Adriani et al. “Isolated hard photon emission in hadronic Z^0 decays”. In: *Physics Letters B* 292.3 (1992), pp. 472–484. ISSN: 0370-2693. DOI: [https://doi.org/10.1016/0370-2693\(92\)91205-N](https://doi.org/10.1016/0370-2693(92)91205-N). URL: <https://www.sciencedirect.com/science/article/pii/037026939291205N> (cit. on p. 27).
- [55] G. Abbiendi et al. “Multi-photon production in e^+e^- collisions at $s = 189 \text{ GeV}$ ”. In: *Physics Letters B* 465.1 (1999), pp. 303–314. ISSN: 0370-2693. DOI: [https://doi.org/10.1016/S0370-2693\(99\)01092-8](https://doi.org/10.1016/S0370-2693(99)01092-8). URL: <https://www.sciencedirect.com/science/article/pii/S0370269399010928> (cit. on p. 27).
- [56] Gautam Rupak and Elizabeth H. Simmons. “Limits on pseudoscalar bosons from rare Z decays at LEP”. In: *Physics Letters B* 362.1–4 (Nov. 1995), pp. 155–163. ISSN: 0370-2693. DOI: [10.1016/0370-2693\(95\)01152-g](https://doi.org/10.1016/0370-2693(95)01152-g). URL: [http://dx.doi.org/10.1016/0370-2693\(95\)01152-G](http://dx.doi.org/10.1016/0370-2693(95)01152-G) (cit. on p. 27).
- [57] The ATLAS collaboration. “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: *Journal of High Energy Physics* 2021.3 (Mar. 2021), p. 243. ISSN: 1029-8479. DOI: [10.1007/JHEP03\(2021\)243](https://doi.org/10.1007/JHEP03(2021)243) (cit. on p. 27).
- [58] CMS Collaboration. “Evidence for light-by-light scattering and searches for axion-like particles in ultraperipheral PbPb collisions at $\sqrt{s_{\text{NN}}}=5.02\text{TeV}$ ”. In: *Physics Letters B* 797 (2019), p. 134826. ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2019.134826>. URL: <https://www.sciencedirect.com/science/article/pii/S0370269319305404> (cit. on p. 27).
- [59] ATLAS Collaboration. “Search for Scalar Diphoton Resonances in the Mass Range 65–600 GeV with the ATLAS Detector in pp Collision Data at $\sqrt{s} = 8 \text{ TeV}$ ”. In: *Phys. Rev. Lett.* 113 (17 Oct. 2014), p. 171801. DOI: [10.1103/PhysRevLett.113.171801](https://doi.org/10.1103/PhysRevLett.113.171801). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.113.171801> (cit. on p. 27).

- [60] Aram Hayrapetyan et al. “Search for new physics in high-mass diphoton events from proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *JHEP* 08 (2024), p. 215. DOI: [10.1007/JHEP08\(2024\)215](https://doi.org/10.1007/JHEP08(2024)215). arXiv: [2405.09320](https://arxiv.org/abs/2405.09320) [hep-ex] (cit. on p. 27).
- [61] ATLAS Collaboration. “Search for resonances decaying to photon pairs in 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Physics Letters B* 822 (Nov. 2021), p. 136651. ISSN: 0370-2693. DOI: [10.1016/j.physletb.2021.136651](https://doi.org/10.1016/j.physletb.2021.136651). URL: <http://dx.doi.org/10.1016/j.physletb.2021.136651> (cit. on p. 27).
- [62] Georges Aad et al. “Search for diphoton resonances in the 66 to 110 GeV mass range using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *JHEP* 01 (2025), p. 053. DOI: [10.1007/JHEP01\(2025\)053](https://doi.org/10.1007/JHEP01(2025)053). arXiv: [2407.07546](https://arxiv.org/abs/2407.07546) [hep-ex] (cit. on p. 27).
- [63] Aram Hayrapetyan et al. “Search for a standard model-like Higgs boson in the mass range between 70 and 110 GeV in the diphoton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *Phys. Lett. B* 860 (2025), p. 139067. DOI: [10.1016/j.physletb.2024.139067](https://doi.org/10.1016/j.physletb.2024.139067). arXiv: [2405.18149](https://arxiv.org/abs/2405.18149) [hep-ex] (cit. on p. 27).
- [64] CMS Collaboration. *Search for a very low mass diphoton resonance using 2018 data*. Tech. rep. Geneva: CERN, 2025. URL: <https://cds.cern.ch/record/2927383> (cit. on pp. 27, 128).
- [65] Martin Bauer et al. “Axion-like particles at future colliders”. In: *The European Physical Journal C* 79.1 (Jan. 2019). ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-019-6587-9](https://doi.org/10.1140/epjc/s10052-019-6587-9). URL: <http://dx.doi.org/10.1140/epjc/s10052-019-6587-9> (cit. on p. 28).
- [66] Michael H. Seymour and Marilyn Marx. “Monte Carlo Event Generators”. en. In: arXiv:1304.6677 (Apr. 2013). arXiv:1304.6677 [hep-ph]. DOI: [10.48550/arXiv.1304.6677](https://doi.org/10.48550/arXiv.1304.6677). URL: <http://arxiv.org/abs/1304.6677> (cit. on p. 28).
- [67] Christian Bierlich et al. “A comprehensive guide to the physics and usage of PYTHIA 8.3”. en. In: arXiv:2203.11601 (Mar. 2022). arXiv:2203.11601 [hep-ph]. DOI: [10.48550/arXiv.2203.11601](https://doi.org/10.48550/arXiv.2203.11601). URL: <http://arxiv.org/abs/2203.11601> (cit. on p. 28).
- [68] Enrico Bothmann et al. “Event Generation with Sherpa 2.2”. en. In: *SciPost Physics* 7.3 (Sept. 2019). arXiv:1905.09127 [hep-ph], p. 034. ISSN: 2542-4653. DOI: [10.21468/SciPostPhys.7.3.034](https://doi.org/10.21468/SciPostPhys.7.3.034) (cit. on p. 29).
- [69] S. Agostinelli et al. “Geant4—a simulation toolkit”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8). URL: <https://www.sciencedirect.com/science/article/pii/S0168900203013688> (cit. on pp. 29, 90).
- [70] Coll ATLAS et al. *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2010-013>. Geneva: CERN, 2010. URL: <https://cds.cern.ch/record/1300517> (cit. on p. 29).
- [71] CERN. *The history of CERN*. URL: <https://home.cern/about/who-we-are/our-history> (visited on 06/19/2024) (cit. on p. 32).

- [72] CERN. *Convention for the Establishment of a European Organization for Nuclear Research*. URL: <https://council.web.cern.ch/en/content/convention-establishment-european-organization-nuclear-research> (visited on 06/19/2024) (cit. on p. 32).
- [73] Timothy J Berners-Lee and Robert Cailliau. “World-Wide Web”. In: (1992). DOI: [10.5170/CERN-1992-007.69](https://cds.cern.ch/record/245440). URL: <https://cds.cern.ch/record/245440> (cit. on p. 32).
- [74] Patrice Loiez. “The first web server: this NeXT machine was used by Tim Berners-Lee in 1990 to develop and run the first WWW server, multimedia browser and web editor.. Le premier serveur WWW: en 1990, sur cette machine NeXT Tim Berners-Lee developpa le premier serveur WWW, le navigateur multimédia et l’éditeur web.” 1990. URL: <https://cds.cern.ch/record/42413> (cit. on p. 32).
- [75] CERN. *Our People / CERN*. URL: <https://home.cern/about/who-we-are/our-people> (visited on 06/20/2024) (cit. on p. 33).
- [76] Lyndon Evans and Philip Bryant. “LHC Machine”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08001. DOI: [10.1088/1748-0221/3/08/S08001](https://dx.doi.org/10.1088/1748-0221/3/08/S08001). URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08001> (cit. on p. 33).
- [77] Maurizio Vretenar et al. *Linac4 design report*. Vol. 6. CERN Yellow Reports: Monographs. Geneva: CERN, 2020. DOI: [10.23731/CYRM-2020-006](https://cds.cern.ch/record/2736208). URL: <https://cds.cern.ch/record/2736208> (cit. on p. 33).
- [78] Esma Mobs. “The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018”. In: (2018). General Photo. URL: <https://cds.cern.ch/record/2636343> (cit. on p. 34).
- [79] ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08003. DOI: [10.1088/1748-0221/3/08/S08003](https://dx.doi.org/10.1088/1748-0221/3/08/S08003). URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08003> (cit. on pp. 33, 36, 42, 48, 49, 51–53).
- [80] CMS Collaboration. “The CMS experiment at the CERN LHC”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08004. DOI: [10.1088/1748-0221/3/08/S08004](https://dx.doi.org/10.1088/1748-0221/3/08/S08004). URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08004> (cit. on p. 33).
- [81] ALICE Collaboration. “The ALICE experiment at the CERN LHC”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08002. DOI: [10.1088/1748-0221/3/08/S08002](https://dx.doi.org/10.1088/1748-0221/3/08/S08002). URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08002> (cit. on p. 33).
- [82] LHCb Collaboration. “The LHCb Detector at the LHC”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08005. DOI: [10.1088/1748-0221/3/08/S08005](https://dx.doi.org/10.1088/1748-0221/3/08/S08005). URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08005> (cit. on p. 33).
- [83] Werner Herr and B Muratori. “Concept of luminosity”. In: (2006). DOI: [10.5170/CERN-2006-002.361](https://cds.cern.ch/record/941318). URL: <https://cds.cern.ch/record/941318> (cit. on p. 35).
- [84] ATLAS Collaboration. *Public ATLAS Luminosity Results for Run-2 of the LHC. Total Integrated Luminosity and Data Quality in 2015-2018*. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2> (visited on 02/14/2025) (cit. on p. 36).
- [85] ATLAS Collaboration. *Public ATLAS Luminosity Results for Run-3 of the LHC. Total Integrated Luminosity in Run 3 (13.6 TeV p-p data only)*. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3> (visited on 02/14/2025) (cit. on p. 36).

- [86] ATLAS Collaboration. *Public ATLAS Luminosity Results for Run-2 of the LHC. Number of Interactions per Crossing*. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2> (visited on 02/14/2025) (cit. on p. 36).
- [87] ATLAS Collaboration. *Public ATLAS Luminosity Results for Run-3 of the LHC. Interactions per Crossing 2022-2024*. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3> (visited on 02/14/2025) (cit. on p. 36).
- [88] O. Aberle et al. *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2020. DOI: [10.23731/CYRM-2020-0010](https://doi.org/10.23731/CYRM-2020-0010). URL: <https://cds.cern.ch/record/2749422> (cit. on p. 35).
- [89] *The HL-LHC project. LHC/ HL-LHC Plan (last update October 2024)*. URL: <https://hilumilhc.web.cern.ch/content/hl-lhc-project> (visited on 02/14/2025) (cit. on p. 37).
- [90] ATLAS Collaboration. “The ATLAS experiment at the CERN Large Hadron Collider: a description of the detector configuration for Run 3”. In: *Journal of Instrumentation* 19.05 (May 2024), P05063. DOI: [10.1088/1748-0221/19/05/P05063](https://doi.org/10.1088/1748-0221/19/05/P05063). URL: <https://doi.org/10.1088/1748-0221/19/05/P05063> (cit. on pp. 38, 40, 42, 48, 52).
- [91] Joao Pequenaio and Paul Schaffner. “How ATLAS detects particles: diagram of particle paths in the detector”. 2013. URL: <https://cds.cern.ch/record/1505342> (cit. on p. 39).
- [92] *ATLAS inner detector: Technical Design Report, 1*. Technical design report. ATLAS. Geneva: CERN, 1997. URL: <https://cds.cern.ch/record/331063> (cit. on p. 39).
- [93] ATLAS Collaboration. “Experiment Briefing: Keeping the ATLAS Inner Detector in perfect alignment”. General Photo. 2020. URL: <https://cds.cern.ch/record/2723878> (cit. on p. 41).
- [94] M Capeans et al. *ATLAS Insertable B-Layer Technical Design Report*. Tech. rep. 2010. URL: <https://cds.cern.ch/record/1291633> (cit. on p. 40).
- [95] *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical design report. ATLAS. Geneva: CERN, 1996. DOI: [10.17181/CERN.FWRW.FOOQ](https://doi.org/10.17181/CERN.FWRW.FOOQ). URL: <https://cds.cern.ch/record/331061> (cit. on pp. 43, 46).
- [96] H. Bethe. “Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie”. In: *Annalen der Physik* 397.3 (1930), pp. 325–400. DOI: <https://doi.org/10.1002/andp.19303970303>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.19303970303>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19303970303> (cit. on p. 43).
- [97] S. Navas et al. “Review of particle physics”. In: *Phys. Rev. D* 110.3 (2024), p. 030001. DOI: [10.1103/PhysRevD.110.030001](https://doi.org/10.1103/PhysRevD.110.030001) (cit. on pp. 44–46).
- [98] I. Wingerter-Seez. “Particle Physics Instrumentation”. In: (2017). DOI: [10.23730/CYRSP-2017-002.295](https://doi.org/10.23730/CYRSP-2017-002.295). arXiv: [1804.11246](https://arxiv.org/abs/1804.11246). URL: <https://cds.cern.ch/record/2315747> (cit. on p. 45).
- [99] Christian Wolfgang Fabjan and F Gianotti. “Calorimetry for Particle Physics”. In: *Rev. Mod. Phys.* 75 (2003), pp. 1243–1286. DOI: [10.1103/RevModPhys.75.1243](https://doi.org/10.1103/RevModPhys.75.1243). URL: <https://cds.cern.ch/record/692252> (cit. on p. 46).

- [100] ATLAS Collaboration. “Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data”. In: *The European Physical Journal C* 74.10 (Oct. 2014), p. 3071. ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-014-3071-4](https://doi.org/10.1140/epjc/s10052-014-3071-4) (cit. on p. 47).
- [101] C Gabaldon. *Drift Time measurement in the ATLAS Liquid Argon electromagnetic calorimeter using cosmic muons*. Tech. rep. Geneva: CERN, 2010. URL: <https://cds.cern.ch/record/1301524> (cit. on p. 49).
- [102] *ATLAS muon spectrometer: Technical Design Report*. Technical design report. ATLAS. Geneva: CERN, 1997. URL: <https://cds.cern.ch/record/331068> (cit. on p. 50).
- [103] ATLAS Collaboration. “The ATLAS Trigger System for LHC Run 3 and Trigger performance in 2022”. In: *JINST* 19 (2024). All figures including auxiliary figures are available at <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/TRIG-2022-01>, P06029. DOI: [10.1088/1748-0221/19/06/P06029](https://doi.org/10.1088/1748-0221/19/06/P06029). arXiv: 2401.06630. URL: <https://cds.cern.ch/record/2886407> (cit. on p. 53).
- [104] ATLAS Collaboration. “Performance of electron and photon triggers in ATLAS during LHC Run 2”. In: *Eur. Phys. J. C* 80.1 (2020). 56 pages in total, author list starting page 40, 26 figures, 10 tables, published in EPJC. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/TRIG-2018-05>, p. 47. DOI: [10.1140/epjc/s10052-019-7500-2](https://doi.org/10.1140/epjc/s10052-019-7500-2). arXiv: 1909.00761. URL: <https://cds.cern.ch/record/2688248> (cit. on p. 54).
- [105] ATLAS Collaboration. “Performance of the Electronic Readout of the ATLAS Liquid Argon Calorimeters”. In: *JINST* 5 (2010), P09003. DOI: [10.1088/1748-0221/5/09/P09003](https://doi.org/10.1088/1748-0221/5/09/P09003). URL: <https://cds.cern.ch/record/1303004> (cit. on p. 56).
- [106] M (CERN) Aleksa et al. *ATLAS Liquid Argon Calorimeter Phase-I Upgrade: Technical Design Report*. Tech. rep. Final version presented to December 2013 LHCC. 2013. URL: <https://cds.cern.ch/record/1602230> (cit. on pp. 57, 58).
- [107] *Technical Design Report for the Phase-II Upgrade of the ATLAS TDAQ System*. Tech. rep. Geneva: CERN, 2017. DOI: [10.17181/CERN.2LBB.4IAL](https://doi.org/10.17181/CERN.2LBB.4IAL). URL: <https://cds.cern.ch/record/2285584> (cit. on p. 57).
- [108] W.E. Cleland and E.G. Stern. “Signal processing considerations for liquid ionization calorimeters in a high rate environment”. en. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 338.2–3 (Jan. 1994), pp. 467–497. ISSN: 01689002. DOI: [10.1016/0168-9002\(94\)91332-3](https://doi.org/10.1016/0168-9002(94)91332-3) (cit. on p. 58).
- [109] M. Aharrouché 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> (cit. on p. 58).
- [110] W Lampl et al. *Calorimeter Clustering Algorithms: Description and Performance*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-LARG-PUB-2008-002>. Geneva: CERN, 2008. URL: <http://cds.cern.ch/record/1099735> (cit. on p. 59).

- [111] ATLAS Collaboration. “Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} = 13$ TeV”. en. In: *The European Physical Journal C* 79.8 (Aug. 2019), p. 639. ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-019-7140-6](https://doi.org/10.1140/epjc/s10052-019-7140-6) (cit. on pp. 61, 63, 64).
- [112] ATLAS Collaboration. “Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data”. en. In: *The European Physical Journal C* 74.10 (Oct. 2014), p. 3071. ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-014-3071-4](https://doi.org/10.1140/epjc/s10052-014-3071-4) (cit. on p. 60).
- [113] ATLAS Collaboration. “Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1”. In: *Eur. Phys. J. C* 77 (2017). Comments: 64 pages plus author list + cover page (87 pages in total), 41 figures, 3 tables, submitted to EPJC. All figures including auxiliary figures are available at <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/PERF-2014-07/>, p. 490. DOI: [10.1140/epjc/s10052-017-5004-5](https://doi.org/10.1140/epjc/s10052-017-5004-5). arXiv: 1603.02934. URL: <https://cds.cern.ch/record/2138166> (cit. on pp. 60, 85).
- [114] *Electron and photon reconstruction and performance in ATLAS using a dynamical, topological cell clustering-based approach*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-022>. Geneva: CERN, 2017. URL: <https://cds.cern.ch/record/2298955> (cit. on pp. 61, 62, 67).
- [115] ATLAS Collaboration. “Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data”. In: *Journal of Instrumentation* 14.12 (Dec. 2019), P12006. DOI: [10.1088/1748-0221/14/12/P12006](https://doi.org/10.1088/1748-0221/14/12/P12006). URL: <https://dx.doi.org/10.1088/1748-0221/14/12/P12006> (cit. on pp. 61, 63–66, 68, 75, 80, 81).
- [116] Yu-ting Shen et al. *Electron isolation efficiencies with 2015 data*. Tech. rep. Geneva: CERN, 2015. URL: <https://cds.cern.ch/record/2112167> (cit. on p. 62).
- [117] Nadezda Proklova et al. *Photon isolation performance plots using 2017 data*. Tech. rep. Geneva: CERN, 2017. URL: <https://cds.cern.ch/record/2282730> (cit. on p. 62).
- [118] ATLAS Collaboration. “Particle-flow reconstruction and global event description with the CMS detector”. In: *Journal of Instrumentation* 12.10 (Oct. 2017), P10003–P10003. ISSN: 1748-0221. DOI: [10.1088/1748-0221/12/10/p10003](https://doi.org/10.1088/1748-0221/12/10/p10003). URL: <http://dx.doi.org/10.1088/1748-0221/12/10/P10003> (cit. on p. 62).
- [119] ATLAS Collaboration. “Jet reconstruction and performance using particle flow with the ATLAS Detector”. In: *The European Physical Journal C* 77.7 (July 2017). ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-017-5031-2](https://doi.org/10.1140/epjc/s10052-017-5031-2). URL: <http://dx.doi.org/10.1140/epjc/s10052-017-5031-2> (cit. on p. 62).
- [120] T Cornelissen et al. “The new ATLAS track reconstruction (NEWT)”. In: *J. Phys.: Conf. Ser.* 119 (2008), p. 032014. DOI: [10.1088/1742-6596/119/3/032014](https://doi.org/10.1088/1742-6596/119/3/032014). URL: <https://cds.cern.ch/record/1176900> (cit. on p. 63).
- [121] T G Cornelissen et al. “The global χ^2 track fitter in ATLAS”. In: *J. Phys.: Conf. Ser.* 119 (2008), p. 032013. DOI: [10.1088/1742-6596/119/3/032013](https://doi.org/10.1088/1742-6596/119/3/032013). URL: <https://cds.cern.ch/record/1176901> (cit. on p. 64).

- [122] ATLAS Collaboration. “Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties”. In: *Journal of Instrumentation* 12.05 (May 2017), P05002–P05002. ISSN: 1748-0221. DOI: [10.1088/1748-0221/12/05/p05002](https://doi.org/10.1088/1748-0221/12/05/p05002). URL: <http://dx.doi.org/10.1088/1748-0221/12/05/P05002> (cit. on p. 65).
- [123] ATLAS Collaboration. “Electron and photon energy calibration with the ATLAS detector using LHC Run 2 data”. In: *Journal of Instrumentation* 19.02 (Feb. 2024), P02009. DOI: [10.1088/1748-0221/19/02/P02009](https://doi.org/10.1088/1748-0221/19/02/P02009). URL: <https://dx.doi.org/10.1088/1748-0221/19/02/P02009> (cit. on pp. 68, 69, 71, 72, 74–76).
- [124] ATLAS Collaboration. “Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton-proton collision data”. In: *Journal of Instrumentation* 14.03 (Mar. 2019), P03017. DOI: [10.1088/1748-0221/14/03/P03017](https://doi.org/10.1088/1748-0221/14/03/P03017). URL: <https://dx.doi.org/10.1088/1748-0221/14/03/P03017> (cit. on pp. 68, 70).
- [125] Andrea Valassi. “Combining correlated measurements of several different physical quantities”. en. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 500.1–3 (Mar. 2003), pp. 391–405. ISSN: 01689002. DOI: [10.1016/S0168-9002\(03\)00329-2](https://doi.org/10.1016/S0168-9002(03)00329-2) (cit. on p. 70).
- [126] ATLAS Collaboration. “Electron and photon performance measurements with the ATLAS detector using the 2015-2017 LHC proton-proton collision data”. In: *JINST* 14.12 (2019). 31 figures, 3 tables. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/EGAM-2018-01>, P12006. DOI: [10.1088/1748-0221/14/12/P12006](https://doi.org/10.1088/1748-0221/14/12/P12006). arXiv: 1908.00005. URL: <https://cds.cern.ch/record/2684552> (cit. on pp. 84, 87, 89).
- [127] ATLAS Collaboration. “Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV”. In: *Eur. Phys. J. C* 79.8 (2019). 63 pages in total, author list starting page 47, 16 figures, 4 tables, final version published in EPJC. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/PERF-2017-01>, p. 639. DOI: [10.1140/epjc/s10052-019-7140-6](https://doi.org/10.1140/epjc/s10052-019-7140-6). arXiv: 1902.04655. URL: <https://cds.cern.ch/record/2657964> (cit. on pp. 84, 87).
- [128] Luis Pascual Dominguez. “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. 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”. Presented 29 Jan 2021. Paris U., 2021. URL: <https://cds.cern.ch/record/2754121> (cit. on pp. 85, 87, 103).
- [129] Matteo Cacciari and Gavin P. Salam. “Pileup subtraction using jet areas”. en. In: *Physics Letters B* 659.1–2 (Jan. 2008), pp. 119–126. ISSN: 03702693. DOI: [10.1016/j.physletb.2007.09.077](https://doi.org/10.1016/j.physletb.2007.09.077) (cit. on p. 84).
- [130] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “FastJet user manual: (for version 3.0.2)”. In: *The European Physical Journal C* 72.3 (Mar. 2012). ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-012-1896-2](https://doi.org/10.1140/epjc/s10052-012-1896-2). URL: <http://dx.doi.org/10.1140/epjc/s10052-012-1896-2> (cit. on p. 84).

- [131] 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](https://doi.org/10.1103/PhysRevD.48.3160). URL: <https://link.aps.org/doi/10.1103/PhysRevD.48.3160> (cit. on p. 84).
- [132] ATLAS Collaboration. “Electron and photon efficiencies in LHC Run 2 with the ATLAS experiment”. en. In: *Journal of High Energy Physics* 2024.5 (May 2024). arXiv:2308.13362 [hep-ex]. DOI: [10.1007/JHEP05\(2024\)162](https://doi.org/10.1007/JHEP05(2024)162) (cit. on pp. 85, 86).
- [133] Gabriella Pasztor et al. *Photon trigger performance in 2015 ATLAS data*. Tech. rep. For upcoming conferences: EPS, DPF and LISHEP. Geneva: CERN, 2015. URL: <https://cds.cern.ch/record/2034866> (cit. on p. 90).
- [134] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. “A brief introduction to PYTHIA 8.1”. In: *Computer Physics Communications* 178.11 (June 2008), pp. 852–867. ISSN: 0010-4655. DOI: [10.1016/j.cpc.2008.01.036](https://doi.org/10.1016/j.cpc.2008.01.036) (cit. on p. 90).
- [135] T. Gleisberg et al. “Event generation with SHERPA 1.1”. In: *Journal of High Energy Physics* 2009.02 (Feb. 2009), p. 007. DOI: [10.1088/1126-6708/2009/02/007](https://doi.org/10.1088/1126-6708/2009/02/007). URL: <https://dx.doi.org/10.1088/1126-6708/2009/02/007> (cit. on p. 90).
- [136] Hung-Liang Lai et al. “New parton distributions for collider physics”. In: *Phys. Rev. D* 82 (7 Oct. 2010), p. 074024. DOI: [10.1103/PhysRevD.82.074024](https://doi.org/10.1103/PhysRevD.82.074024). URL: <https://link.aps.org/doi/10.1103/PhysRevD.82.074024> (cit. on p. 90).
- [137] ATLAS Collaboration. “The ATLAS Simulation Infrastructure”. In: *The European Physical Journal C* 70.3 (Dec. 2010), pp. 823–874. ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-010-1429-9](https://doi.org/10.1140/epjc/s10052-010-1429-9) (cit. on p. 90).
- [138] *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie-Landshoff diffractive model*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2016-017>. Geneva: CERN, 2016. URL: <https://cds.cern.ch/record/2206965> (cit. on p. 90).
- [139] Simon S. Haykin. *Neural networks and learning machines*. en. 3rd ed. New York: Prentice Hall, 2009. ISBN: 978-0-13-147139-9 (cit. on p. 106).
- [140] F. Rosenblatt. “The perceptron: A probabilistic model for information storage and organization in the brain.” en. In: *Psychological Review* 65.6 (1958), pp. 386–408. ISSN: 1939-1471, 0033-295X. DOI: [10.1037/h0042519](https://doi.org/10.1037/h0042519) (cit. on p. 107).
- [141] K. L. Chung. “On a Stochastic Approximation Method”. In: *The Annals of Mathematical Statistics* 25.3 (1954), pp. 463–483. ISSN: 00034851. URL: <http://www.jstor.org/stable/2236830> (visited on 05/01/2025) (cit. on p. 111).
- [142] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: [1412.6980](https://arxiv.org/abs/1412.6980) [cs.LG]. URL: <https://arxiv.org/abs/1412.6980> (cit. on p. 111).
- [143] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. eng. Adaptive computation and machine learning. Cambridge, Massachusetts London, England: The MIT Press, 2016. ISBN: 978-0-262-03561-3 (cit. on p. 114).

- [144] Yann LeCun et al. “Handwritten Digit Recognition with a Back-Propagation Network”. In: *Advances in Neural Information Processing Systems*. Ed. by D. Touretzky. Vol. 2. Morgan-Kaufmann, 1989. URL: https://proceedings.neurips.cc/paper_files/paper/1989/file/53c3bce66e43be4f209556518c2fcb54-Paper.pdf (cit. on p. 114).
- [145] CNRS MIAI Université Grenoble Alpes. *Formation d’Introduction au Deep Learning (FIDLE)*. Portée par l’Institut pluridisciplinaire en intelligence artificielle (MIAI), l’Université Grenoble Alpes (UGA) et le Centre National de la Recherche Scientifique (CNRS). Accédé le 25 mars 2025. 2024. URL: <https://fidle.cnrs.fr/w3/> (cit. on pp. 115, 116).
- [146] Franco Scarselli et al. “The Graph Neural Network Model”. In: *IEEE Transactions on Neural Networks* 20.1 (Jan. 2009), pp. 61–80. ISSN: 1941-0093. DOI: [10.1109/TNN.2008.2005605](https://doi.org/10.1109/TNN.2008.2005605) (cit. on p. 117).
- [147] Ines Chami et al. “Machine Learning on Graphs: A Model and Comprehensive Taxonomy”. en. In: arXiv:2005.03675 (Apr. 2022). arXiv:2005.03675 [cs]. DOI: [10.48550/arXiv.2005.03675](https://doi.org/10.48550/arXiv.2005.03675). URL: <http://arxiv.org/abs/2005.03675> (cit. on pp. 117, 118).
- [148] Yao Ma and Jiliang Tang. *Deep Learning on Graphs*. Cambridge: Cambridge University Press, 2021. ISBN: 978-1-108-83174-1. DOI: [10.1017/9781108924184](https://doi.org/10.1017/9781108924184) (cit. on pp. 118, 119).
- [149] Ashish Vaswani et al. “Attention is All you Need”. In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc., 2017. URL: <https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html> (cit. on pp. 119, 120, 122, 123).
- [150] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. “Neural Machine Translation by Jointly Learning to Align and Translate”. In: arXiv:1409.0473 (May 2016). arXiv:1409.0473 [cs]. DOI: [10.48550/arXiv.1409.0473](https://doi.org/10.48550/arXiv.1409.0473). URL: <http://arxiv.org/abs/1409.0473> (cit. on p. 120).
- [151] Juho Lee et al. “Set Transformer: A Framework for Attention-based Permutation-Invariant Neural Networks”. In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, June 2019, pp. 3744–3753. URL: <https://proceedings.mlr.press/v97/lee19d.html> (cit. on pp. 122, 123).
- [152] ATLAS Collaboration. “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 (2014). Comments: 5 pages plus author list + cover page (18 pages total), 4 figures, 2 tables, submitted to Physical Review Letters. All figures including auxiliary figures are available at <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/HIGG-2014-04/>, p. 171801. DOI: [10.1103/PhysRevLett.113.171801](https://doi.org/10.1103/PhysRevLett.113.171801). arXiv: [1407.6583](https://arxiv.org/abs/1407.6583). URL: <https://cds.cern.ch/record/1745602> (cit. on p. 128).

- [153] ATLAS Collaboration. “Search for resonances decaying into photon pairs in 139 fb⁻¹ of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.” In: *Phys. Lett. B* 822 (2021). 33 pages in total, author list starting page 17, 5 figures, 2 tables, submitted to Physics Letters B. All figures including auxiliary figures are available at <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/HIGG-2018-27>, p. 136651. DOI: [10.1016/j.physletb.2021.136651](https://doi.org/10.1016/j.physletb.2021.136651). arXiv: [2102.13405](https://arxiv.org/abs/2102.13405). URL: <https://cds.cern.ch/record/2753130> (cit. on p. 128).
- [154] CMS Collaboration. “Search for diphoton resonances in the mass range from 150 to 850 GeV in pp collisions at $\sqrt{s} = 8$ TeV”. In: *Phys. Lett. B* 750 (2015). Replaced with published version. Added journal reference, pp. 494–519. DOI: [10.1016/j.physletb.2015.09.062](https://doi.org/10.1016/j.physletb.2015.09.062). arXiv: [1506.02301](https://arxiv.org/abs/1506.02301). URL: <https://cds.cern.ch/record/2022620> (cit. on p. 128).
- [155] CMS Collaboration. “Search for physics beyond the standard model in high-mass diphoton events from proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *Physical Review D* 98.9 (Nov. 2018). ISSN: 2470-0029. DOI: [10.1103/physrevd.98.092001](https://doi.org/10.1103/physrevd.98.092001). URL: <http://dx.doi.org/10.1103/PhysRevD.98.092001> (cit. on p. 128).
- [156] ATLAS Collaboration. “Search for boosted diphoton resonances in the 10 to 70 GeV mass range using 138 fb⁻¹ of 13 TeV pp collisions with the ATLAS detector”. In: *JHEP* 2307 (2023). 40 pages in total, 8 figures, 1 table, author list starting in page 23, published in JHEP. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/HIGG-2019-23>, p. 155. DOI: [10.1007/JHEP07\(2023\)155](https://doi.org/10.1007/JHEP07(2023)155). arXiv: [2211.04172](https://arxiv.org/abs/2211.04172). URL: <https://cds.cern.ch/record/2839865> (cit. on pp. 128, 129).
- [157] *Search for a standard model-like Higgs boson in the mass range between 70 and 110 GeV in the diphoton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV*. Tech. rep. Geneva: CERN, 2023. URL: <https://cds.cern.ch/record/2852907> (cit. on p. 128).
- [158] Adam Paszke et al. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. 2019. arXiv: [1912.01703](https://arxiv.org/abs/1912.01703) [cs.LG]. URL: <https://arxiv.org/abs/1912.01703> (cit. on p. 129).
- [159] ATLAS Collaboration. “Search for short- and long-lived axion-like particles in $H \rightarrow aa \rightarrow 4\gamma$ decays with the ATLAS experiment at the LHC”. In: *The European Physical Journal C* 84.7 (July 2024). ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-024-12979-0](https://doi.org/10.1140/epjc/s10052-024-12979-0). URL: <http://dx.doi.org/10.1140/epjc/s10052-024-12979-0> (cit. on p. 130).
- [160] Luis Pascual Dominguez. “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”. Theses. Université de Paris / Université Paris Diderot (Paris 7), Jan. 2021. URL: <https://hal.science/tel-03409306> (cit. on p. 131).
- [161] Huilin Qu and Loukas Gouskos. “ParticleNet: Jet Tagging via Particle Clouds”. en. In: *Physical Review D* 101.5 (Mar. 2020). arXiv:1902.08570 [hep-ph]. ISSN: 2470-0010, 2470-0029. DOI: [10.1103/PhysRevD.101.056019](https://doi.org/10.1103/PhysRevD.101.056019). URL: <http://arxiv.org/abs/1902.08570> (cit. on pp. 150, 151).
- [162] Yue Wang et al. *Dynamic Graph CNN for Learning on Point Clouds*. 2019. arXiv: [1801.07829](https://arxiv.org/abs/1801.07829) [cs.CV]. URL: <https://arxiv.org/abs/1801.07829> (cit. on p. 150).

- [163] Sergey Ioffe and Christian Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. 2015. arXiv: [1502.03167](https://arxiv.org/abs/1502.03167) [cs.LG]. URL: <https://arxiv.org/abs/1502.03167> (cit. on p. 151).
- [164] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. “Layer Normalization”. In: arXiv:1607.06450 (July 2016). arXiv:1607.06450 [stat]. DOI: [10.48550/arXiv.1607.06450](https://doi.org/10.48550/arXiv.1607.06450). URL: <http://arxiv.org/abs/1607.06450> (cit. on p. 156).
- [165] IPCC. “Sections”. In: *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by Core Writing Team, H. Lee, and J. Romero. Geneva, Switzerland: IPCC, 2023, pp. 35–115. DOI: [10.59327/IPCC/AR6-9789291691647](https://doi.org/10.59327/IPCC/AR6-9789291691647) (cit. on pp. 176–178, 180).
- [166] IPCC. *About the IPCC*. <https://www.ipcc.ch/about/>. Accessed: 25/05/2025. 2025 (cit. on p. 176).
- [167] Labos 1point5. *Le Groupement De Recherche*. <https://labos1point5.org/le-gdr>. Accessed: 27/05/2025. 2025 (cit. on p. 179).
- [168] Jérôme Mariette et al. “An open-source tool to assess the carbon footprint of research”. en. In: *Environmental Research: Infrastructure and Sustainability* 2.3 (Sept. 2022), p. 035008. ISSN: 2634-4505. DOI: [10.1088/2634-4505/ac84a4](https://doi.org/10.1088/2634-4505/ac84a4) (cit. on pp. 179, 183).
- [169] Greenhouse Gas Protocol. *Homepage / GHG Protocol*. <https://ghgprotocol.org/>. Accessed: 27/05/2025. 2025 (cit. on p. 179).
- [170] Marianne De Paepe et al. “Purchases dominate the carbon footprint of research laboratories”. en. In: *PLOS Sustainability and Transformation* 3.7 (July 2024). Ed. by Lian Pin Koh, e0000116. ISSN: 2767-3197. DOI: [10.1371/journal.pstr.0000116](https://doi.org/10.1371/journal.pstr.0000116) (cit. on pp. 180, 181).
- [171] Marianne Blanchard et al. “Concerned yet polluting: A survey on French research personnel and climate change”. en. In: *PLOS Climate* 1.9 (Sept. 2022). Ed. by Malcolm Fairbrother, e0000070. ISSN: 2767-3200. DOI: [10.1371/journal.pclm.0000070](https://doi.org/10.1371/journal.pclm.0000070) (cit. on pp. 181, 190, 192, 193).
- [172] Tamara Ben-Ari et al. “Flight quotas outperform focused mitigation strategies in reducing the carbon footprint of academic travel”. en. In: *Environmental Research Letters* 19.5 (May 2024), p. 054008. ISSN: 1748-9326. DOI: [10.1088/1748-9326/ad30a6](https://doi.org/10.1088/1748-9326/ad30a6) (cit. on pp. 182, 192, 194).
- [173] CERN *Environment Report—Rapport sur l’environnement 2021–2022*. Report Vol. 3. CERN, Dec. 2023. DOI: [10.25325/CERN-Environment-2023-003](https://doi.org/10.25325/CERN-Environment-2023-003). URL: https://e-publishing.cern.ch/index.php/CERN_Environment_Report/issue/view/156 (cit. on pp. 183–186).
- [174] *Rapport d’activité 2022-2023*. Laboratoire de Physique Nucléaire et de Hautes Énergies (LPNHE), Apr. 2024. URL: https://lpnhe.in2p3.fr/IMG/pdf/rapport_d_activite_2022-2023_hd.pdf?3630/2965836c68cdef6f08df14c617271b3a8afcb01d (cit. on p. 186).
- [175] *Stratégie nationale bas-carbone*. Consulted in May 2025. Ministère de la Transition écologique et solidaire, Mar. 2020. URL: https://www.ecologie.gouv.fr/sites/default/files/documents/2020-03-25_MTES_SNBC2.pdf (cit. on p. 187).

- [176] Agence de la transition écologique (ADEME). *L'Agence*. <https://www.ademe.fr/lagence/>. Consulté en mai 2025 (cit. on p. 191).
- [177] Carbone 4. *À propos de Carbone 4*. <https://www.carbone4.com/about>. Consulté en mai 2025 (cit. on p. 191).
- [178] Olivier Berné et al. “The carbon footprint of scientific visibility”. en. In: *Environmental Research Letters* 17.12 (Dec. 2022), p. 124008. ISSN: 1748-9326. DOI: [10.1088/1748-9326/ac9b51](https://doi.org/10.1088/1748-9326/ac9b51) (cit. on p. 191).
- [179] Laboratoire de Physique Nucléaire et de Hautes Énergies (LPNHE). *Charte de développement durable du LPNHE*. https://lpnhe.in2p3.fr/IMG/pdf/charte_dd_lpnhe_v27012025.pdf?3672/071aaaf5bdcae4f70a190f4e96a0231cce5711a3. Version du 27 janvier 2025. Jan. 2025 (cit. on p. 193).
- [180] K Cranmer. “Practical Statistics for the LHC”. en. In: (2015). DOI: [10.5170/CERN-2015-001.247](https://doi.org/10.5170/CERN-2015-001.247). URL: <https://cds.cern.ch/record/2004587> (cit. on p. 199).
- [181] Robert Les. *Conceptual Introduction to Statistical Methods*. ATLAS Lecture Series, CERN. Accessed via Indico. Nov. 2024. URL: <https://indico.cern.ch/event/1455306/> (cit. on p. 199).
- [182] Andrey N. Kolmogorov. *Foundations of the Theory of Probability*. Translated from the second Russian edition. New York: Chelsea Publishing Company, 1950 (cit. on p. 199).
- [183] L. Lyons. “A Particle Physicist’s Perspective on Astrostatistics”. In: *Statistical Challenges in Modern Astronomy IV*. Ed. by G. J. Babu and E. D. Feigelson. Vol. 371. Astronomical Society of the Pacific Conference Series. Nov. 2007, p. 361 (cit. on p. 200).
- [184] S. S. Wilks. “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses”. In: *The Annals of Mathematical Statistics* 9.1 (1938), pp. 60–62. ISSN: 00034851, 21688990. URL: <http://www.jstor.org/stable/2957648> (visited on 08/08/2025) (cit. on p. 204).
- [185] A L Read. “Modified frequentist analysis of search results (the CL_s method)”. In: (2000). DOI: [10.5170/CERN-2000-005.81](https://doi.org/10.5170/CERN-2000-005.81). URL: <https://cds.cern.ch/record/451614> (cit. on p. 204).
- [186] Alexander L. Read. “Presentation of search results: The CL_s technique”. In: *J. Phys. G* 28 (2002). Ed. by M. R. Whalley and L. Lyons, pp. 2693–2704. DOI: [10.1088/0954-3899/28/10/313](https://doi.org/10.1088/0954-3899/28/10/313) (cit. on p. 204).
- [187] Glen Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. en. In: *The European Physical Journal C* 71.2 (Feb. 2011). arXiv:1007.1727 [physics]. ISSN: 1434-6044, 1434-6052. DOI: [10.1140/epjc/s10052-011-1554-0](https://doi.org/10.1140/epjc/s10052-011-1554-0). URL: <http://arxiv.org/abs/1007.1727> (cit. on pp. 205, 206).
- [188] Abraham Wald. “Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large”. In: *Transactions of the American Mathematical Society* 54.3 (1943), pp. 426–482. ISSN: 00029947, 10886850. URL: <http://www.jstor.org/stable/1990256> (visited on 07/27/2025) (cit. on p. 205).
- [189] Isaac Asimov. “Franchise”. In: *Isaac Asimov: The Complete Stories, Vol. 1*. Originally published in 1955. Broadway Books, 1990 (cit. on p. 206).